Energy Barrier Engineering in Nanomagnetic Devices for Unconventional Computing

---

A

Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Md Golam Morshed

August 2024

© 2024 Md Golam Morshed

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Md Golam Morshed

This Dissertation has been read and approved by the examing committee:

Advisor: Avik W. Ghosh

Advisor:

Committee Member: Mircea R. Stan

Committee Member: S. Joseph Poon

Committee Member: Prasanna V. Balachandran

Committee Member: Andrew D. Kent

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

August 2024

To my family, whose unwavering support and selfless sacrifices made this possible

Abstract

The burgeoning big data era has been pushing our limits of computing and processing. While hardware miniaturization and Moore's law have driven innovation in digital electronics for several decades, scaling hardware has recently become challenging due to the high energy cost of computing and the increasingly steep memory wall caused by the segregation of the memory and processing units. One alternative is to utilize spins of electrons instead of charge to store and process information, popularly known as *spintronics*. Nanomagnetic devices are the building blocks in spintronics and exhibit non-volatility, which is attributed to the energy barrier between different spin states. While nanomagnets are now integrated onto silicon as commercial non-volatile memory elements, size scaling them for compactness reduces their energy barrier and turns them volatile. This dissertation addresses this considerable challenge by exploring different modes of energy barrier engineering in nanomagnetic devices that leverage topological invariants and traditional nanomagnets for unconventional computing. First, we utilize ultrasmall ($\sim 10\ nm$) and ultrafast ($\sim 1000\ m/s$) topological excitations — *magnetic skyrmions* with topologically protected barriers for temporal computing. We engineer the topological barrier by tuning the interfacial Dzyaloshinskii-Moriya interaction (DMI) to control the size and stability of skyrmions. We model a racetrack with periodic notches to produce and tune the energy barrier to hold skyrmions in place in the presence of thermal jitter and further employ machine learning (ML) to automate the process. Second, while topology (skyrmions) is one way to artificially increase the barrier of a tiny magnet, an alternate approach is to utilize instead the truly random white noise in low barrier

magnets (LBMs) to accelerate fast optimization algorithms. This concept of probabilistic or p-bits, in effect, a binary stochastic neuron (BSN), relies, however, on magnets of perfect circular symmetry and is hyper-sensitive to process variations. We study the reliability of the computational networks built from LBMs and utilize their inherent stochasticity for inferencing tasks. Additionally, we model more realistic, commercially viable medium barrier magnets (MBMs) actuated by short current pulses for energy-efficient and robust probabilistic computing. Third, we turn to potentially more energy-efficient strain-gated dynamical barrier lowering of a nanomagnet in a heterogeneous piezoelectric/magnet/topological insulator (TI)/magnet stack that naturally encompasses logic and memory in an in-memory computing architecture with a minuscule energy cost and negligible footprint to circumvent the memory wall bottleneck. We solve a coupled stochastic Landau-Lifshitz-Gilbert (LLG) equation for the heterostructure, analyze the phase space of the working device, estimate the device level energy cost, and project it to the architecture level for in-memory logic operations. This dissertation may provide a comprehensive approach to energy barrier engineering in nanomagnetic devices, potentially leading to alternative technologies beyond complementary metal-oxide-semiconductor (CMOS) and facilitating more than Moore era.

Acknowledgments

I am deeply grateful to my advisor, Professor Avik Ghosh, for his invaluable guidance and support throughout my Ph.D. journey. He constantly pushed me to the limit, and his encouragement to step out of my comfort zone and think creatively was instrumental in my growth as a researcher. Apart from the research, I appreciate his empathy and patience during my health and family crises.

I am thankful to my committee members, Professor Mircea Stan, Professor Andy Kent, Professor Joseph Poon, and Professor Prasanna Balachandran, for helping and guiding me with constructive feedback on research and career – a special thanks to Professor Andy Kent for always being my support. I am fortunate to work closely with you on multiple projects.

I am grateful to my brilliant colleagues and collaborators – Hamed Vakili, Samiran Ganguly, Yunkun Xie, Jianhua Ma, Faiyaz Mullik, Nazmus Sakib, Laura Rehm, Yassine Quessab, Shruba Gangopadhyay, and Amy Khoo – for helping me with my research. I acknowledge the support and friendship of all the present and past Virginia Nano Computing (VINO) research group members.

I thank all the staff members from the Electrical and Computer Engineering (ECE) department at the University of Virginia (UVA), especially Beth, for her constant support. I also acknowledge the support from the UVA Graduate Affairs.

I acknowledge the financial support from the Defense Advanced Research Projects Agency (DARPA), NSF I/UCRC on Multi-functional Integrated System Technology (MIST) Center, ECE department, and the computational resource support from the UVA Research Computing.

I want to express my heartfelt gratitude to my friends. There's the saying, "We are the average of the five people we spend the most time with." I was fortunate to be surrounded by brilliant minds and good human beings in all stages of life. Asif Iqbal Arnab and Ashikuzzaman Rasel top the list, for sure. Their influence and support have always been pivotal, from the beginning of my undergrad until today. Perhaps I could not have made it thus far if they were not there to look up to. I want to mention a few more names – Mohammad Rafat, Zubair Ebne Rafique, Alam Mahmud, Shafiq Islam, Sabbir Ibn Akhter Sifat, and Manik Hossain for their support anytime. I am grateful to all my friends from school, colleges, and friends at UVA.

Above all, I thank my precious family for their unconditional love and sacrifices. Being the youngest child, I was showered with love and care. They always tried to give me the best possible conditions, even if the situation was unfavorable. I am forever grateful to my father, Mazibar Rahman, for his vision that only education can uplift the next generation, encouragement to dream big, and lifelong sacrifices. My mother, Hazera Begum, is the pillar of my strength, without whom I can not stand where I am today. She always believed in me, which gave me courage and made me resilient. I thank my elder brother, Mamunur Rashid, for guiding me in the right direction during my school and college days and encouraging me to get into engineering; my eldest sister, Momotaz Rezvin, for mentoring and raising me during my school days; my sister, Mamataz Jesmin, for enduring all my tantrums and for her selflessness; and my sister-in-law, Sabina Yeasmin, for her support. To my nephew Mugdho and nieces Mrittika and Muntaha – thank you for your love and affection; you motivate me to do better and be your role model. I am thankful to my in-laws for their prayers. I am grateful to Mahabubun Nayeema Aunty for her motherly care throughout my undergrad.

Lastly, and most importantly, to my loving wife, Tasnia Sharmeen – your sacrifice and unwavering support have made this journey significantly less difficult and stressful. Without you, this achievement would not have been possible. I love you! < 3

Contents

Contents **i**
List of Tables . . . l
List of Figures . . . m
List of Abbreviations . . . u

1 Introduction **1**
1.1 Background . . . 1
1.2 Dissertation organization . . . 4

2 Energy barrier engineering in skyrmionics for temporal computing **6**
2.1 Engineering topological barrier of skyrmions through tuning Dzyaloshinskii-Moriya interaction . . . 7
2.1.1 Computational details . . . 9
2.1.2 Tuning DMI through W composition in the cap layer . . . 10
2.1.3 Change in SOC energy dictates the DMI trend . . . 14
2.1.4 Tuning DMI through cap layer heavy metal . . . 15
2.1.5 Conclusion . . . 18
2.2 Engineering energy barrier via notches for skyrmions' positional stability . . . 19
2.2.1 Computational details . . . 21
2.2.2 Tuning E_b through material and physical parameters . . . 22
2.2.3 Deriving empirical equations for estimating E_b . . . 24
2.2.4 Alternate pinning sites: local variation of material params . . . 28
2.2.5 Estimating E_b for rare-earth Mn_4N . . . 28
2.2.6 Alternate notch shape: triangular notch . . . 30
2.2.7 Large energy barrier yet low unpinning current . . . 30
2.2.8 Prediction of E_b using surrogate ML models . . . 31
A. Materials informatics workflow . . . 32
B. Performance of the ML models . . . 35
2.2.9 Conclusion . . . 39

3 Energy barrier engineering in nanomagnets for probabilistic computing **40**
3.1 Reliability of the computational networks built from low-barrier magnets . . . 41
3.1.1 Building 'p-bits' using LBMs . . . 43
3.1.2 Computational details . . . 45
3.1.3 Impact of variability on the "p-bit" performance . . . 46

A. Characteristics distortion . . . 47
B. Energy barrier variability . . . 48
3.1.4 Sampling vs. Simulated annealing . . . 50
3.1.5 Conclusion . . . 51
3.2 Employing stochasticity for inferencing tasks . . . 53
3.2.1 Brief overview on neuron models . . . 56
3.2.2 Computational details . . . 58
3.2.3 Binary vs. Analog: inference errors . . . 60
3.2.4 Deterministic vs. Stochastic: generalizability and robustness . . . 65
3.2.5 Synaptic weights dynamic range: hardware implementability . . . 65
3.2.6 Memory capacity . . . 68
3.2.7 Conclusion . . . 69
3.3 Modeling short pulse-actuated medium barrier magnet-based robust and energy-efficient TRNG units . . . 72
3.3.1 Computational details . . . 74
3.3.2 Short-pulse actuated STT switching for $P_{\mathrm{sw}} = 0.5$ . . . 75
3.3.3 Impact of pulse amplitude and duration on $P_{\mathrm{sw}} = 0.5$ . . . 77
3.3.4 Impact of temperature on $P_{\mathrm{sw}} = 0.5$ . . . 78
3.3.5 Impact of process variations on $P_{\mathrm{sw}} = 0.5$ . . . 79
3.3.6 Energy cost of switching . . . 80
3.3.7 Conclusion . . . 83

4 Energy barrier lowering in piezoelectric/magnet/topological insulator/magnet stacks for in-memory computing **84**
4.1 Computational details . . . 87
4.2 Four-layer vertical stack for PiM crossbar . . . 89
4.3 Functionality of the device . . . 90
4.4 Gating and writing mechanism requirements . . . 94
4.5 Energy cost . . . 96
4.6 Energy cost projection for in-memory AND and OR operations . . . 97
4.7 Impact of material parameters . . . 98
4.8 Challenges and opportunities . . . 100
4.9 Conclusion . . . 103

5 Conclusion and outlook **104**

Appendices **107**

A Dissipation in SMART MTJ devices **108**
A.1 Energy dissipation . . . 108

B PiM sensing scheme **110**

C Publications **112**
C.1 Journal articles (published) 112
C.2 Journal articles (under review & in-preparation) 113
C.3 Conferences proceedings 114
C.4 Conference abstracts & presentations 114

Bibliography **117**

List of Tables

2.1 The micromagnetic parameters used for Mn_4N [1–4]. These parameters are used in the mumax^3 calculation and treated as constants. For comparison, the micromagnetic parameters used by Ref. [5] for GdCo are also given. 38
2.2 The search space of Mn_4N. Min and Max indicates the lower and upper boundary, respectively. The Resolution means the distance between each sampled value. 38
2.3 Candidate functions used for fitting PDPs. "a" indicates the fitting parameter with an initial value of $a = 1$. 39

3.1 Average NMSE data extracted from the ASN and BSN models ($b = 5\%$) for various reservoir sizes. The form of the input signal is, $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t) + C[rand(1,t) - 0.5]$. 70
3.2 Robustness vs. accuracy trade-off ($N = 20$). The label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions described earlier, respectively. 71
3.3 Linear memory capacity (MC) for different neuron models. 71
3.4 Material parameters for MTJ free layer. 83

4.1 Material parameters 100
4.2 Energy consumption and area for in-memory AND and OR operations. E_{write} and E_{read} are the combined energy for both the bit cells, while $E_{\text{sense amp.}}$ represents the energy associated with a single sense amplifier. Area is the total area of the bit cells and the read access transistor (W/L = 10:1). 101

List of Figures

1.1 Pictorial synopsis of the topics covered in this dissertation. (a) (Top) Energy barrier of skyrmions modulated by DMI. (Bottom) Energy barrier in a notched racetrack. (b) Energy barrier in nanomagnets for HBM, LBM, and MBM actuated by a short pulse. (c) TI topological surface states modulation via magnetization orientation of a gating magnet. The orange arrows indicate the logical connection between different research fronts. 5

2.1 Schematic of Pt(2)/GdCo(2)/$Pt_{1-x}W_x$(2) structure (number in parentheses denoting the number of monolayers) corresponding to $x = 12.5\%$ for (a) CW, and (b) ACW spin configurations. The red arrows in the figure show the spin orientations for non-collinear calculations. L1, ..., L6 denote the layer number while numbers in circles label atomic positions. 10

2.2 The DMI as a function of W composition (x) in Pt/GdCo/$Pt_{1-x}W_x$. (a) DMI variation with respect to W positions while the Gd atoms are fixed at $(1, 7)$ positions. (b) Total spectrum of the DMI as both Gd and W positions are varied. For a specific W composition, each of the different colors represents the variation of Gd atomic positions, and the scattered points within the same color represent different W positions for that particular Gd arrangement in the structure (Fig. 2.1). (c) Surface DMI in comparison with experimentally observed DMI [6]. The numbers followed by the symbols Gd and W represent the positions of the respective atoms in the structure shown in Fig. 2.1. . . . 12

2.3 (a) Change in SOC energy at the interfacial HM layers (L2 & L5) as a result of changing spin chirality of the magnetic layers (L3 & L4) from CW to ACW. All the color bars on the left (right) side represents the SOC energy change at L2 (L5) for different W compositions. (b) Layer resolved DMI for structures having W composition, $x = 0\% - 50\%$. 14

2.4 (a) Calculated DMI in Pt/GdCo/X, where X=Ta, W, Ir. (b) Layer resolved DMI. 15

2.5 Experimentally measured spin Hall angle in Pt/GdCo/X, where X=Ta, W, Ir, showing a similar qualitative trend as the calculated DMI. The figure is reprinted from [7] with permission. 16

2.6 Projected density of states (p-DOS) showing $3d$-$5d$ band overlap between Co (black) and X (colored) in Pt/GdCo/X. (a) X=Ta, (b) X=W, and (c) X= Ir. The red up (down) arrow represents the spin-up (spin-down) channel. 16

2.7 (a) Simulated racetrack with notched geometry. Each of the figures represents different snapshots of the skyrmion trajectory along the racetrack, referred to as 'image index'. (b) The energy vs image index for the optimal trajectory of a skyrmion with varying DMIs in a 800 $nm \times 200$ nm racetrack with notch radius, $R_N = 100$ nm. The left valley, peak, and right valley correspond to the image index shown in the top, middle, and bottom panels of Fig. 2.7(a) respectively. (c) Change in skyrmion radius, ΔR_{sk} when the skyrmion passes over the notch. Higher DMIs initiate larger initial skyrmions that undergo bigger shrinkage and larger energy costs when forced through the constriction. 22

2.8 Material parameters dependence of the energy barrier, E_b. Effect of notch size (radius R_N) on (a) E_b, and (b) ΔR_{sk}, both for varying DMIs. We see that a larger R_N leads to a larger $|\Delta R_{sk}|$ that corresponds to a higher barrier. (c) Thickness dependence of E_b in a racetrack ($R_N = 100$ nm). The inset shows that E_b increases linearly as a function of racetrack thickness, t_F for any specific D. The linear variation of E_b vs t_F is consistent with the overall uniform cylindrical shape of the skyrmion at ultrathin limit. The color code to represent DMI variations in (a), (b), and inset of (c) are the same. 23

2.9 (a) Fitting of E_b (normalized by t_F) with respect to R_N. (b), (c) The relation between the fitting constants obtained from (a) and D. The red colored texts in each graph represent the fitting function. 25

2.10 (a) Energy landscape of skyrmions in a racetrack ($R_N = 100$ nm) from the MEP simulations (scatter circles) and analytical equations (solid curves) [8, 9]. The analytical equations include the energetics of 2π skyrmions on an infinite plane plus a phenomenological confinement correction. (b) Explanation of the mismatch between simulated and analytical (calculated using Eq. 2.4) data in a racetrack ($R_N = 100$ nm). A skyrmion confined above the pinning site has the energy of a skyrmion confined in a circle of diameter $W - R_N$ above the notch minus that of an unconfined infinite plane skyrmion. Including only one fitting term can describe E_b with good accuracy for varying D values. The black circular region above the notch in the inset schematic represents the simulation geometry for the confined case. 25

2.11 (a) Energy barrier for different pinning sites in a 5 nm thick racetrack. For all the cases, the pinning site is a semi-circular region of 100 nm radius. K_u and A_{ex} of the pinning site are 2 times higher, and D and M_s are 10 times lower than the rest of the track region. We choose the ratio that gives the highest E_b. (b) Change in skyrmion radius corresponding to the E_b in (a), shows a proportional relation between E_b and $|\Delta R_{sk}|$. 29

2.12 Energy barrier of a triangular notch with 75 nm (black) and 100 nm (blue) depth in comparison with semi-circular notch. 29

2.13 The critical current J_c (black) to unpin the skyrmion and the corresponding E_b (blue) for a 5 nm thick racetrack as a function of R_{sk}. We see a low unpinning current with a fairly large energy barrier. The arrows point each colored data to the corresponding y-axis. 31

2.14 Schematic representation illustrating the integration of closed-loop adaptive learning workflow with post-hoc model explainable techniques and symbolic regression. We start by enumerating all possible combinations of the three inputs ($\mathbf{X}$) $\{D, R_N, t_F\}$, which defines our large search space ($\mathcal{S}$). A subset of 20 data points were selected from this search space using the space-filling Latin hypercube sampling (LHS) method. We then calculated E_b (our output, $\mathbf{y_t}$) for these 20 data points using mumax^3. We now have a training data set ($\mathcal{T}$), $\mathcal{T} = (\mathbf{X_t}, \mathbf{y_t})$. The next step is the closed-loop adaptive learning framework. We used an ensemble of support vector regression (eSVR) to map a quantitative relationship between $\mathbf{X_t}$ and $\mathbf{y_t}$. The trained eSVR model was then used to predict the E_b for the data points in the virtual set $\mathcal{V}$, where $\mathcal{V} = \mathcal{S} - \mathbf{X_t}$. We then rank each data point in $\mathcal{V}$ using two independent acquisition functions: the non-parametric ICAL and parametric SD. The top ranked candidate is recommended to mumax^3 for the calculation of E_b. We then augment $\mathcal{T}$ with the new data point and iterate. The entire adaptive learning loop was setup to run autonomously. The next step is the post-hoc analysis of the trained black-box eSVR models based on global feature importance and partial dependence plots, which helps us in uncovering the functional relationships between each input descriptor in $\mathbf{X_t}$ and the predicted E_b as learned by the eSVR models. The outcomes from post-hoc model explanation methods inform the symbolic regression. We employed the grammatical evolution method to accomplish this task. The outcome from grammatical evolution is an interpretable mathematical model to predict $E_b = f(D, R_N, t_F)$. 34

2.15 Comparison of the predictive power of the trained eSVR models on their training set (red diamond) and unseen test set (black circle) that was not used for training the models. Parity plot of (a) eSVR-SD model with $R^2 = 0.986$ for the training set and $R^2 = 0.989$ for the test set, (b) eSVR-ICAL model with $R^2 = 0.987$ for the training set and $R^2 = 0.970$ for the unseen test set. In both (a) and (b), X-axis is the mumax^3-calculated E_b and Y-axis is the ML predicted E_b. Although the overall performances of both models are good, differences are evident especially in (b) when $E_b > 60\ k_BT$. 36

2.16 Partial dependence plots for Mn_4N. (a), (b), and (c) uses the trained eSVR-SD model to calculate the marginal dependencies; (d), (e), and (f) uses the trained eSVR-ICAL model to calculate the marginal dependencies. The marginal effects of individual features D, R_N and t_F on E_b as learned by the trained ML models are captured by the plots. Descriptors D and R_N show a non-linear relationship with E_b, whereas t_F shows a linear relationship with E_b. Despite the differences in the sampling between SD and ICAL acquisition functions, both models have captured similar trends. 37

2.17 X-axis is the E_b calculated from micromagnetic simulations (mumax^3) for Mn_4N and Y-axis is the predicted E_b from $\alpha D^8 R_N^2 t_F$ model given in Eq. (2.8) (filled black diamond) and eSVR-ICAL model (filled red circles). Error bars for Eq. (2.8) are computed from Bayesian inference. In the case of eSVR-ICAL model, error bars represent the ensemble standard deviation of model predictions. 38

3.1 (a) Illustrative schematic of an embedded RBM, an energy-based optimization and learning algorithm, in a dual-stacked feedback cross-bar structure with neurons (the compute units) at the edges (large circles), while the synaptic weights (the program) loaded in memristors located at the cross-points of the core cross-bar structure (small circles). The active neurons and synapses are colored bold (red and yellow), while inactive units are greyed out. (b) The RBM network that gets embedded in (a). The bidirectional blue lines represent the synaptic connections between the neurons (red circles). The yellow circles used in (a) are not shown here for simplicity. (c) The design of an LBM MTJ-based p-bit unit. (d) Ideal characteristics of a p-bit device. (e) Schematics of different characteristics distortions. (f) Illustration of energy barrier variation in a nanomagnet. Symbols (diamond, square, etc.) in (e) and (f) represent different variabilities henceforth. 42

3.2 EPP from horizontal shifting for (a) EMOA and (b) PGA with different network sizes (size of the J matrix). Fig. 3.2 and all the subsequent figures show the mean EPP and the error bar represents the standard deviation. The mean and standard deviation are calculated from 100 simulations. The figures use different colors to represent various network sizes, and different symbols represent different distortions introduced in Fig. 3.1(e). Inset in (a) shows the schematic of a 3×3 EMOA network, while inset in (b) shows the schematic of an 8×8 PGA Bayesian network representing a family tree (GF: Grandfather, GM: Grandmother, etc.). 46

3.3 EPP from vertical shifting for (a) EMOA and (b) PGA with different network sizes. 47

3.4 EPP from horizontal scaling for (a) EMOA and (b) PGA with different network sizes. 49

3.5 EPP from vertical scaling for (a) EMOA and (b) PGA with different network sizes. 49

3.6 EPP arising from different characteristics distortions ($N = 100$). The green, light green, light red, and red colors represent the saturation, sub-linear, linear, and super-linear trends of EPP, respectively, as a function of characteristics distortion. 50

3.7 EPP from energy barrier variability for (a) EMOA and (b) PGA with different network sizes. 51

3.8 EPP calculated using sampling technique (dashed line) vs. simulated annealing technique (solid line) for EMOA for (a) horizontal shifting, (b) vertical shifting, (c) horizontal scaling, and (d) vertical scaling. 52

3.9 (a) Schematic of different types of widely used neuron models with their output characteristics. In the bottom panel, all the red curves represent the deterministic neurons' output characteristics. In the top panel, the blue curves represent the actual stochastic output characteristics while the red is the corresponding deterministic/expected value of the output ($< stochastic\ output >$) characteristics. Spiking neurons (SpN and SSpN) can be considered in between the two limits of purely binary vs. purely analog neurons. Please note that we only analyze the analog and binary neurons (including their stochastic counterparts) in this work, as indicated by the purple-colored bold font labels. (b) Schematic of a reservoir setup using neurons connected with each other bidirectionally with random weights. 57

3.10 Comparison of NMSE for an analog time-series prediction task between (a) ASN and (b) BSN models as a function of reservoir size with 5% stochasticity incorporated in both the neuron models for a clean input signal. The form of the clean input signal is $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t)$, where $A = 1$, $B = 2$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. ASN performs better than BSN for the entire range of reservoir size as indicated by the average (μ) NMSE (cyan dashed-dotted line). ASN shows a decreasing trend in NMSE as a function of reservoir size while BSN results remain almost unchanged. The NMSE data for every reservoir size is obtained from five different reservoir topologies and 1000 simulation runs (different random 'seed') within each topology (total sample size is 5000). The color bar represents the frequency of the NMSE data. Note that in some cases, our model fails to generate a meaningful NMSE as the reservoir output blows up. We get meaningful output from $\sim 90\% - 100\%$ cases depending on the reservoir sizes, and those data are plotted here and used to estimate the average NMSE. The bottom panel is the zoomed version of the top panel and the magenta dashed-dotted lines are the guide to the eye that shows the data distribution in the range of $\mu \pm \sigma$. The color codes to represent the μ and σ are the same for the subsequent figures henceforth. . . 61

3.11 Evolution of NMSE for different degrees of stochasticity (noise percentages) associated with the (a) ASN and (b) BSN models. ASN performs better than the BSN model for analog time-series prediction tasks throughout the ranges of the degree of stochasticity as indicated by the average NMSE shown in (c) and (d) for ASN and BSN, respectively. The characteristics of the average NMSE as a function of reservoir size i.e., the decreasing trend for ASN while almost no change for BSN holds throughout the range of b. 63

3.12 Evolution of NMSE for different degrees of stochasticity for (a) ASN and (b) BSN models for a distorted input signal. Random white noise is added to the clean input signal to introduce distortion and the form of the distorted signal is $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t) + C[rand(1,t) - 0.5]$, where $A = 1$, $B = 2$, $C = 1$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. ASN performs better than BSN for the distorted input, as indicated by the average NMSE shown in (c) and (d) for ASN and BSN, respectively, which dictates the robustness of the ASN model in terms of performance irrespective of the input signals. 66

3.13 Comparison of NMSE for time-series prediction task between ASN and BSN models for various input functions for a reservoir size of (a) N = 20 and (b) N =30. The degree of stochasticity incorporated in both neuron models is 5%. The label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions, respectively. ANS performance is better than BSN in terms of NMSE for different input functions. 67

3.14 Dynamic range of the learned synaptic weights, W_{out} for all the neuron models ($N = 20$). 5% stochasticity is considered in the ASN and BSN models. ASN model shows the smallest dynamic range that leads to better hardware implementability. The label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions, respectively. 68

3.15 (a) Probability density for an MBM ($\Delta \sim 35$) activated by a 1 ns pulse (top panel). The pulse is turned ON at $t = 0$ ns. A bimodal distribution emerges after the pulsing, which leads to a 50% switching probability. The bottom panel shows the time evolution of the probability density (colormap) during the pulsing. (b) Probability of switching (colormap) as a function of pulse amplitude and duration. The dotted blue overlaid curve represents a 50% switching probability. 74

3.16 Variation in midpoint switching probability with respect to (a) pulse amplitude and (b) pulse duration for various pulse durations. 78

3.17 (a) Variation in midpoint switching probability as a function of temperature for various pulse durations. The black dash-dotted lines show the linear fit. (b) Variation in midpoint switching probability with respect to temperature for a free layer with different thermal stability factors. The black and red dash-dotted lines in (b) represent the macrospin approximation in the ballistic (short pulse) and diffusive (long pulse) limits, respectively, and the texts represent the corresponding equation. 78

3.18 Variation in midpoint switching probability with respect to (a) free layer diameter, (b) free layer thickness, (c) saturation magnetization, (d) anisotropy field, (e) magnetic damping coefficient, and (f) spin polarization efficiency factor for various pulse durations. For all the variations, the change in the midpoint switching probability is lower for the short pulse limit than the longer pulse limit, leading to robust TRNG operations. 81

3.19 (a) Average junction conductance, (b) average energy dissipation, and (c) average energy-resistance product for 50% switching for various pulse durations. In (a), the inset shows the zoomed view of $\langle G \rangle$. In all the figures, the error bars represent the standard deviation in our ensemble. 81

4.1 (a) Schematic of a compact, low-power strained topological insulator SOTRAM (STI-SOTRAM). An out-of-plane gating (selector) magnet gaps the bottom surface states of the TI and places the upper magnet in storage mode (left fig.). The strain generated by the piezoelectric rotates the selector magnet from out-of-plane to in-plane, restoring the bottom layer surface states and activating the TI. The drain polarity V_D sets the storage bit with $m_{2\mathrm{y}}$ (bit '0') or $-m_{2\mathrm{y}}$ (bit '1') using high SHA SOT (middle and right figs, respectively). The fixed layer of the MTJ reads the state as '0' (parallel) or '1' (antiparallel). (b) Schematic of a PiM crossbar architecture where each activated stack (bit cell) can be selected by activating the row column. Two selected bit cells can feed to a sense amplifier that processes the local inputs by comparing with a reference voltage/current and performing Boolean logic operations (e.g., AND, OR, etc.), thereby processing local data from the magnetic memory. (c) Baising condition for read-write operations. WWL: Write Word Line, RWL: Read Word Line, WBL: Write Bit Line, and RBL: Read Bit Line. 89

4.2 Coupled LLG simulations showing gating magnet switching with strain (a), turning ON TSS (closing band gap) and hence delivering surface current (b), and switching MTJ free layer with SOT (c) without the thermal field. (d)-(e) Magnetization dynamics for the gating magnet and the MTJ free layer, respectively, influenced by the stochastic thermal field (50 simulations). (f) Histogram of switching time from 1000 stochastic simulations to estimate the switching time corresponding to WER $= 10^{-9}$. The switching time is calculated as $t_{\mathrm{sw}} = \mu + 6\mathrm{SD}$ from the histogram. The red curve is the Gaussian fit. The results are generated using $I_{0,\mathrm{surf}} = 6I_{\mathrm{c,surf}}$. 91

4.3 (a) Gate voltage requirement in the PZT to generate electrical strain (stress). The capacity to generate strain in the PZT limits the maximum achievable stress. (b) Gating magnet's uniaxial anisotropy vs. stress phase-space for the switching probability (colorplot) of the MTJ free layer. The probability of switching is calculated from 10^5 stochastic simulations. The dash-dotted line represents the TbCo uniaxial anisotropy ($K_{\mathrm{u1}} = 64$ kJ/m^3). 92

4.4 Supplied drain voltage to the TI for writing operation. We use $I_{\mathrm{surf}}/I_{\mathrm{tot}} = 15\%$. 93

4.5 (a) Drain voltage requirement for high speed switching regime operation ($2-10$ ns). (b) Energy dissipation in the TI during the writing process. (c) Comparison of energy dissipation between traditional HM- and TI-based SOT mechanism. TI consumes orders of magnitude less energy than the HM. . . . 95

4.6 Variation in switching time with respect to (a) the thermal stability factor of the gating magnet, (b) the damping coefficient of the gating magnet, and (c) the exchange strength between the TI and the gating magnet. (d)-(f) Histograms of the switching time from 1000 stochastic LLG simulations corresponding to (a)-(c), respectively. We use $I_{0,\mathrm{surf}} = 6I_{\mathrm{c,surf}}$ while calculating the t_{sw}. 99

B.1 (a) Equivalent circuit representation when two bit cells are selected simultaneously and fed to a sense amplifier for in-memory computing. The voltage level for AND and OR operation are shown at the bottom. (b)-(e) Various possible combinations of resistance used in our study. Fig. (a) is reproduced from [10] with permission. 110

List of Abbreviations

ASN	analog stochastic neuron
BSN	binary stochastic neuron
CMOS	complementary metal-oxide-semiconductor
DFT	density functional theory
DMI	Dzyaloshinskii-Moriya interaction
DRAM	dynamic random access memory
FM	ferromagnetic
HBMs	high barrier magnets
HM	heavy metal
LBMs	low barrier magnets
LLG	Landau-Lifshitz-Gilbert
MBMs	medium barrier magnets
ML	machine learning
MRAMs	magnetic random access memories
MTJ	magnetic tunnel junction
NMSE	normalized means squared error
PiM	processing in memory
SHA	spin Hall angle
SOC	spin-orbit coupling
SOT	spin-orbit torque
SRAM	static random access memory
STT	spin-tranfer torque
TI	topological insulator
TRNG	true random number generator
TSS	topological surface states

Chapter 1

Introduction

1.1 Background

The advancement of modern high-performance computing has primarily revolved around the Boolean computing paradigm, leveraging complementary metal-oxide-semiconductor (CMOS) hardware. Over the years, digital electronics have been driven by sustained hardware scaling and Moore's law. With the recent slowdown in transistor scaling and rapid growth of software, along with the migration from cloud computing toward edge devices, there is a pressing need to re-examine the limits of computing. At the device level, transistor miniaturization faces limitations imposed by thermodynamics (the *Boltzmann tyranny*). Meanwhile, at the architecture level, in the conventional Von Neumann computer architecture, the transfer of data is hindered by the latency caused by the separation of memory and processing units (*memory wall problem*).

An alternative approach involves leveraging the spin degrees of freedom of elections — popularly known as *spintronics*, which relies on the active control and manipulation of spin within magnetic devices [11–13]. Nanomagnetic devices are at the core of the spintronics technology. Nanomagnet-based nonvolatile memory has a rich history, with field-switched magnetic random access memories (MRAMs) and spin-tranfer torque (STT)-based MRAMs

(STT-MRAMs) now being commercialized [14, 15]. With both the integration density of dynamic random access memory (DRAM) and comparable performance to static random access memory (SRAM), STT-MRAMs benefit further from their compatibility with CMOS processes and voltages, zero standby leakage, scalability, high endurance, and high retention time. However, one of the significant challenges with the widespread adoption of STT-MRAMs as a universal memory technology is the high write energy and low read speed along with reliability issues coming from a shared read-write path [16, 17]. Structures with orthogonal read–write paths, such as those using spin-orbit torque (SOT) MRAMs (SOT-MRAMs) provide a possible way out — a metallic path used to write information, separate from an orthogonal read path [18].

In a nanomagnet-based non-volatile memory, information is stored in two distinct spin states — spin up (bit '1') and spin down (bit '0'), which are separated by an energy barrier (Fig. 1.1(b) top left). This energy barrier restricts the spontaneous fluctuation of magnetization between these two states, which ensures non-volatility. While magnets are now integrated onto silicon as commercial non-volatile memory elements [19, 20], size scaling them for compactness reduces their energy barrier and turns them volatile [21].

This brings us to the evaluation of bits encoded by ultra-small topological (solitonic) excitations in thin magnetic films, such as magnetic skyrmions [22]. Magnetic skyrmions are spin vortices that act as tiny mobile magnets ($\sim 10\ nm$) and can encode information in ultrasmall volumes below the thermal superparamagnetic limit that constrains regular nanomagnets [23]. The solitons can be driven at high speeds ($\sim 1000\ m/s$) by modest currents along magnetic racetracks with low energy dissipation in heavy metal (HM) underlayers, generating unique device applications [23, 24]. The high density of ultrasmall skyrmions stabilized somewhat by their topological barriers (Fig. 1.1(a) top), as well as their quasi-ballistic, tunable, and linear dynamics are particular attributes that make skyrmionic devices potentially useful in a variety of unconventional applications, for instance, temporal computing [25]. Moreover, introducing notches in a racetrack can produce a large energy barrier (Fig. 1.1(a) bottom),

which restricts the diffusive displacement and ensure a long lifetime of skyrmions.

Alternatively, the volatility (stochasticity) of small magnets in the superparamagnetic limit can be utilized for emerging technology such as probabilistic/stochastic computing [26–28]. The energy barrier of the magnets used in this paradigm of computing is in the order of thermal energy ($\sim k_BT$, k_B is the Boltzmann constant and T is the temperature) as opposed to high barrier magnets (HBMs) with energy barrier $> 40\ k_BT$ that are typically used for non-volatile MRAM (Fig. 1.1(b) top left) and are referred to as low barrier magnets (LBMs). The magnetization of the LBM fluctuates randomly between the spin-up and spin-down states (Fig. 1.1(b) top right), giving rise to probabilistic bits ('p-bits') and can be harnessed for accelerating optimization algorithms where the p-bit devices can be utilized as a tunable true random number generator (TRNG) hardware unit [26]. The concept of p-bits naturally connects to the binary stochastic neuron (BSN) concept in the area of machine learning (ML) [29] and an analog version of this is the analog stochastic neuron (ASN) [30]; both of them can be utilized for inferencing tasks. However, LBMs are extremely sensitive to process variations and hard to build in practice, which motivates the exploration of more realistic medium barrier magnets (MBMs) actuated by short current pulses. MBMs have energy barriers in the range of $20 - 40\ k_BT$ and their actuation with a short current pulse makes them an energy-efficient and robust TRNG unit for probabilistic and stochastic computing (Fig. 1.1(b) bottom) [31].

Moreover, there is considerable interest in exploring emerging material stacks that can be heterogeneously integrated with high-quality interface control to enable new energy-efficient and ultrafast actuating mechanisms for low-power data storage and computing. One such direction combines piezoelectric, magnet, and topological insulator (TI). A piezoelectric-induced strain can strategically switch a magnet from out-of-plane to in-plane by lowering its energy barrier dynamically [32]. A magnet atop TI can modulate TI's surface conductivity through its magnetization orientation (Fig. 1.1(c)) [33, 34]. TI offers unique properties such as spin-momentum-locked topological surface states with charge-to-spin conversion efficiency

much greater than unity, resulting in energy-efficient SOT-based switching of a storage magnetic tunnel junction (MTJ) – making the stack suitable for compact and energy-efficient processing in memory (PiM) [34] applications.

1.2 Dissertation organization

This dissertation explores different modes of energy barrier engineering in nanomagnetic devices that leverage topological invariants and traditional nanomagnets. In Chapter 2, we look into energy barrier engineering for magnetic skyrmions for temporal computing. Using first-principles density functional theory (DFT) calculations, we tune interfacial Dzyaloshinskii-Moriya interaction (DMI), which controls the topological barrier of the skyrmions and eventually the size and stability (Section 2.1). Furthermore, we employ micromagnetic simulations to design a notched racetrack for producing and tuning the energy barrier to hold skyrmions in place against thermal jitter and automate the energy barrier predictions through ML-based surrogate model (Section 2.2). Chapter 3 discusses how we can utilize the inherent stochasticity of low and medium-barrier magnets for probabilistic, stochastic, and neuromorphic computing. In particular, we study the LBM-based computational networks and quantify the error in such networks, as the LBMs are extremely sensitive to process variations (Section 3.1). Additionally, we analyze and compare the BSN model that connects to LBM-based p-bit characteristics and its analog counterpart ASN model for signal prediction tasks (Section 3.2). Furthermore, we model and analyze MBM-based MTJ actuated by a short current pulse employing our in-house Fokker–Planck solver for robust and energy-efficient TRNG units for probabilistic and stochastic computing (Section 3.3). In Chapter 4, we look into another energy-efficient actuating mechanism, such as strain-based switching of a magnet in a piezoelectric/magnet/TI/magnet (MTJ) stack to gate the TI surface current, which writes information in the MTJ, through lowering gating magnet's barrier dynamically from out-of-plane to in-plane. This four-layer stack naturally suits a compact device in a crossbar

architecture for in-memory computing.

In a nutshell, topics covered in this dissertation can be categorized as follows, and a pictorial synopsis is shown in Fig. 1.1.

- **Research Front 1:** Energy barrier engineering in skyrmionics for temporal computing
- **Research Front 2:** Energy barrier engineering (low to medium barrier) in nanomagnets for probabilistic computing
- **Research Front 3:** Energy barrier lowering in piezoelectric/magnet/topological insulator/magnet stacks for in-memory computing

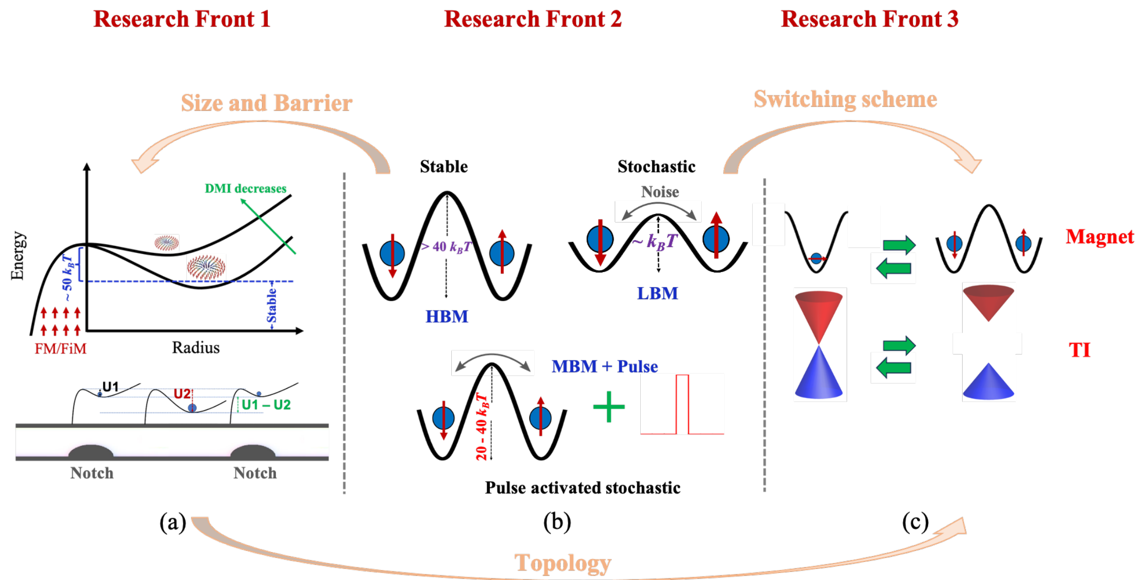


Figure 1.1: Pictorial synopsis of the topics covered in this dissertation. (a) (Top) Energy barrier of skyrmions modulated by DMI. (Bottom) Energy barrier in a notched racetrack. (b) Energy barrier in nanomagnets for HBM, LBM, and MBM actuated by a short pulse. (c) TI topological surface states modulation via magnetization orientation of a gating magnet. The orange arrows indicate the logical connection between different research fronts.

Chapter 2

Energy barrier engineering in skyrmionics for temporal computing

Magnetic skyrmions are topologically protected ultra-small and ultra-fast spin textures and are attractive for next-generation spintronic applications, such as racetrack memory and logic devices for temporal computing [22, 25, 35–38]. In magnetic materials, skyrmions emerged from the complex interplay between various energy terms, such as exchange interaction, anisotropy energy, Zeeman energy, demagnetization energy, and DMI [22, 39]. The delicate balance between these energy contributions governs the energy barrier associated with the skyrmions and determines the size and stability of skyrmions. Additionally, for data retention in skyrmion-based memory applications, skyrmions must be locally held in place for a long time. This chapter discusses the energy barrier engineering in skyrmions through DMI for tuning the size and stability of skyrmions (Section 2.1) and through artificially created constrictions for holding them in place (Section 2.2). This chapter is reprinted from Ref. [40] with permission from APS coauthored with KH. Khoo, Y. Quessab, J-W. Xu, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh and from Ref. [5] with permission from APS coauthored with H. Vakili, and A. W. Ghosh.

2.1 Engineering topological barrier of skyrmions through tuning Dzyaloshinskii-Moriya interaction

The interfacial DMI, an antisymmetric exchange originating from the strong spin-orbit coupling (SOC) in systems with broken inversion symmetry [41, 42], is one of the key ingredients in the formation of skyrmions in magnetic multilayers [43–45]. Controlling the DMI offers the possibility to manipulate skyrmion properties, i.e., size and stability [46, 47].

Over the past few years, the underlying DMI physics and overall skyrmion dynamics have been studied extensively for ferromagnetic (FM) systems [45, 48–53]. Ref. [48] reported interfacial DMI in Pt/Co bilayer with varying thickness of Pt and Co layers. A wide range of heavy metals (HMs) with large SOC in proximity with FM layers demonstrated the role of band overlap around the Fermi level on the sign and magnitude of the DMI [49, 51]. Moreover, another common approach to produce a net effective DMI is to insert a FM layer between different SOC metals [45, 54]. Although both HM/FM bilayers and HM/FM/HM sandwiched structures have been explored, most of the reported results are based on ideal interfaces. Indeed, very few studies focus on the role of disorder on DMI [55]. Furthermore, ferrimagnetic materials have drawn attention due to their low saturation magnetization, low stray fields, reduced sensitivity to external magnetic fields, and fast spin dynamics, all of which favor ultra-fast and ultra-small skyrmions [23, 56–59]. One prototypical example is GdCo thin film. GdCo is an amorphous alloy with antiferromagnetically coupled two sublattices (Gd and Co). At low temperatures, two different sublattices have different magnetic moments, i.e., Gd has a higher magnetic moment, and Co has a much lower magnetic moment. The net magnetization can be expressed as $M = |M(Co, 0K)(1 - T/Tc)^a - M(Gd, 0K)(1 - T/Tc)^b|$, where T_c is the Curie temperature of GdCo alloy. However, the magnetization of GdCo is a function of temperature. With increasing temperature, the magnetic moment the Gd decreases faster than Co. The net magnetization of GdCo goes to zero at a temperature, which is commonly known as magnetic moment compensation temperature, and makes it

compensated. Very recently, Quessab *et al.* have experimentally studied the interfacial DMI in amorphous Pt/GdCo thin films (magnetic compensation of GdCo occurs $\sim$ 150 K), and shown a strong tunability of the DMI by varying the thickness of the GdCo alloy and cap layer composition [6]. However, a detailed understanding of DMI, including the impact of two-sublattice ferrimagnetism, as well as the role of an experimentally realistic, chemically disordered interface are both missing.

In this section, we present a systematic theoretical analysis of the DMI in a compensated ferrimagnetic alloy using first principles calculations. In particular, we explore the variation of the DMI in Pt/GdCo/$Pt_{1-x}W_x$ (Fig. 2.1) and find a strong tunability from 0 to 4.42 mJ/m^2 with variation in the W composition (Fig. 2.2). We studied the influence of atom placement and observed that the DMI is sensitive to structural variations such as the GdCo configuration in the thin magnetic film, and the PtW configuration at the interface. This is important to consider because, in reality, we have an amorphous alloy and the interfaces in deposited films are not perfect. We find a spectrum of DMI values that show an overall saturating trend, as seen in the experimental data [6]. We argue that the change in SOC energy in the interfacial HM layers, especially the constancy of the SOC energy at the bottom layer and reduction of it in the cap layer, generates the observed saturating trend in the DMI with percentage of W incorporated (Fig. 2.3). Additionally, we theoretically predict the variation of the DMI depending on the cap layer material, specifically for Pt/GdCo/X, where X $=$ Ta, W, Ir (Fig. 2.4). We find that the DMI is highest for W in the cap layer and lowest for Ir, a trend that correlates with $3d$-$5d$ Co-X band overlap at the cap layer interface (Fig. 2.6). Furthermore, the observed DMI trend agrees well with the experimentally measured spin Hall angle (Fig. 2.5), indicating an energy-efficient application with W in the cap layer. Our results identify the chemical and geometric factors responsible for interfacial DMI, and provide a potential path forward towards the engineering of material properties towards next generation skyrmion based spintronic applications.

2.1.1 Computational details

We use the technique of constraining the magnetic moments in a supercell to calculate the DMI within the DFT framework [48]. The Vienna *ab initio* simulation package (VASP) is used for the DFT calculations [60]. We use the projector augmented wave (PAW) potential to describe the core-electron interaction [61,62]. The Perdew-Burke-Ernzerhof (PBE) functional form of the generalized gradient approximation (GGA) is used for the exchange-correlation functional [63]. In order to treat the on-site Coulomb interaction of Gd $4f$-electrons, we use the GGA+U method [64] with an effective value of $U = 6$ eV for Gd, as reported in previous studies for both bulk and slab calculations [65–67]. We also validate the effective U for our GdCo alloy by taking a range of U values from $1 - 7$ eV, and confirming a stable ferrimagnetic ground state configuration of GdCo at $U = 6$ eV. A $4 \times 1 \times 1$ supercell of Pt(2)/GdCo(2)/$Pt_{1-x}W_x$(2) (numbers in the parenthesis represent the number of monolayers) is used in all our calculations. While creating the GdCo alloy by replacing Gd atoms in the hcp Co(0001) slab, a 25% Gd composition is maintained, which is the closest to the experimental proportion (22% Gd [6]) achievable within our structural arrangement. The trilayers are formed by aligning fcc(111) and hcp(0001) planes. The in-plane lattice constant of the slab structure is set to 2.81 Å, equal to the calculated nearest neighbor distance of bulk Pt, and the supercells are separated by a vacuum layer of 10 Å in the [001] direction. The cutoff energy is set to 500 eV, and a $4 \times 16 \times 1$ Monkhorst-pack k-grid is used for all the calculations. We verify the convergence of our calculations with cutoff energy, number of k-points, and the thickness of the vacuum layer.

The three-step DMI calculation procedure starts with ionic relaxation along the atomic z-coordinate to mimic a thin film, until the forces become less than 0.01 eV/Å and, the energy difference between two ionic relaxation steps becomes smaller than 10^{-6} eV. Next, in the absence of SOC, the non-spin polarized Kohn-Sham equations are solved to find an initial charge density. Finally, SOC is included, and the total energy of the system is calculated self-consistently for clockwise (CW) and anticlockwise (ACW) spin configurations (Fig. 2.1)

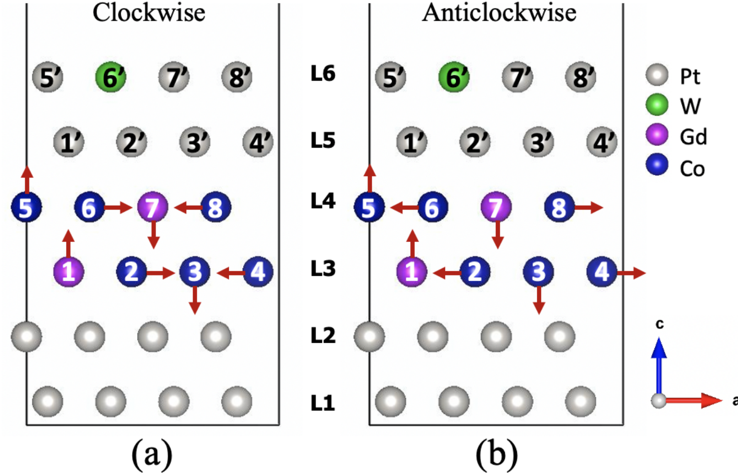


Figure 2.1: Schematic of Pt(2)/GdCo(2)/Pt$_{1-x}$W$_x$(2) structure (number in parentheses denoting the number of monolayers) corresponding to $x = 12.5\%$ for (a) CW, and (b) ACW spin configurations. The red arrows in the figure show the spin orientations for non-collinear calculations. L1, ..., L6 denote the layer number while numbers in circles label atomic positions.

until the energy difference between two consecutive steps becomes smaller than 10^{-6} eV. We consider antiferromagnetic coupling between Co and Gd moment while performing the relaxation and ground-state calculations.

2.1.2 Tuning DMI through W composition in the cap layer

The DMI energy (E_{DMI}) can be defined as

$$E_{DMI} = \sum_{\langle i,j \rangle} \mathbf{d}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j) \tag{2.1}$$

where $\mathbf{S}_i$, $\mathbf{S}_j$ are the nearest neighboring normalized atomic spins and $\mathbf{d}_{ij}$ is the corresponding DMI vector. The total DMI strength, d^{tot}, defined by the summation of the DMI coefficient of each layer, to a first approximation, is calculated by the energy difference between the CW and ACW spin configurations [48], and expressed as $d^{tot} = (E_{CW} - E_{ACW})/12$. The relation between the d^{tot} that relates the DMI energy to spin configurations and the micromagnetic DMI, D that relates the micromagnetic energy per unit volume to magnetization $\boldsymbol{m}$ via the

equation $E = D(m_z\partial_x m_x - m_x\partial_x m_z + m_z\partial_y m_y - m_y\partial_y m_z)$ [22] is given by $D = 3\sqrt{2}d^{tot}/N_F a^2$, where N_F and a represent the number of magnetic layers and the fcc lattice constant respectively [48]. We take into account both the Co-Co and Co-Gd bond while doing the DMI calculations, and report the average DMI per bond.

Before presenting the numerical results, it is worth mentioning that we can only investigate a limited subset of the structures for our calculations, as exploring all combinatorial possibilities is not feasible in terms of time and computational resources. We consider two separate alloy configurations: (i) Gd alloying in the magnetic layers, and (ii) W alloying in the cap layers.

In case (i), we first fix the position of the Gd atoms in the GdCo alloy. We maintain 25% Gd composition separately in each magnetic layer, arguing that steric repulsion implies two Gd atoms are energetically unlikely to sit in the same layer, as assumed in previous studies [68]. The Gd atoms can thus arrange themselves in $\binom{4}{1} \times \binom{4}{1} = 16$ ways. These sixteen combinations can be grouped into just four distinct sets because of their translational symmetry. In Fig. 2.1, looking at positions $(1-8)$ in magnetic layers (L3 & L4), it can be seen that Gd in $(1,7), (2,8), (3,5)$, and $(4,6)$ positions represent equivalent structures once the unit cell is periodically extended. Similarly, the other three groups are $[(1,8), (2,5), (3,6), (4,7)]$, $[(1,6), (2,7), (3,8), (4,5)]$, and $[(1,5), (2,6), (3,7), (4,8)]$. We confirmed this equivalence by calculating the energy of the Pt/GdCo stack by varying all the Gd positions and indeed find equal energy for the four structures within the same group. For case (ii), we choose one representative from each of the above four groups and proceed with W positional variations in the cap layer. While exploring W alloy configurations, for lower composition $(12.5\% - 50\%)$, W is only incorporated in layer L6. Finally, we vary all the possible W positions and calculate the DMI for a total of 76 structures.

Figure 2.2(a) shows the calculated DMI, D for Pt/GdCo/$Pt_{1-x}W_x$, as a function of W composition. At $x = 0\%$, the DMI vanishes as expected because, for a perfectly symmetric trilayer structure, the contributions from the bottom and top interfaces are equal and opposite.

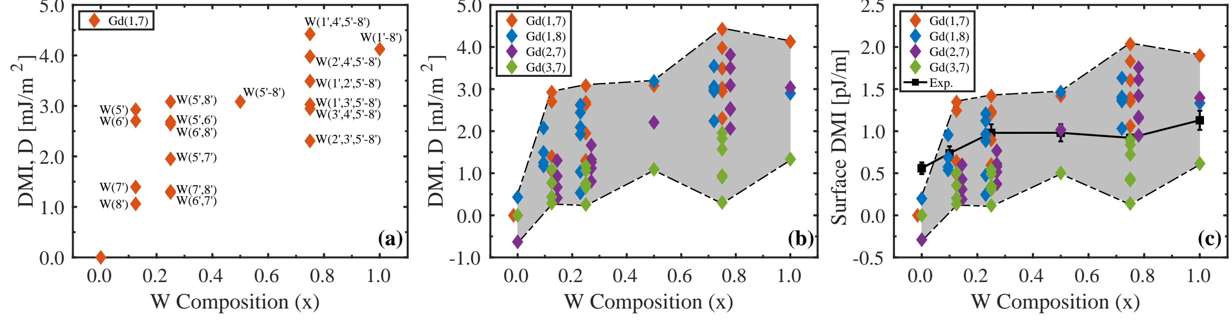


Figure 2.2: The DMI as a function of W composition (x) in Pt/GdCo/$Pt_{1-x}W_x$. (a) DMI variation with respect to W positions while the Gd atoms are fixed at $(1, 7)$ positions. (b) Total spectrum of the DMI as both Gd and W positions are varied. For a specific W composition, each of the different colors represents the variation of Gd atomic positions, and the scattered points within the same color represent different W positions for that particular Gd arrangement in the structure (Fig. 2.1). (c) Surface DMI in comparison with experimentally observed DMI [6]. The numbers followed by the symbols Gd and W represent the positions of the respective atoms in the structure shown in Fig. 2.1.

As the W composition increases from 0% to 12.5%, we find a maximum DMI of 2.93 mJ/m^2. The underlying mechanism behind this non-zero DMI is the symmetry breaking of the Pt/GdCo/Pt structure by the insertion of W atoms in the cap layer. When a W atom is included in the cap layer, the overall symmetry of the stack is broken, but it still leaves intact the mirror symmetry with the normal of the symmetry plane parallel to the b-axis (Fig. 2.1), and hence by the Moriya rules [42], we get the D vector parallel to b-axis. We find that a small amount of W (12.5%) gives a large DMI change, and subsequent to that initial rise, with increasing W content, the DMI saturates. As the composition of W increases, we find a maximum DMI of 4.42 mJ/m^2 corresponding to 75% W composition.

We find that the DMI is very sensitive to the structural details, specifically the positions of the Gd and W atoms. Figure 2.2(a) shows the variation of D as the position of the W atoms changes. In Fig. 2.2(a), for all cases, Gd atoms are fixed at the $(1, 7)$ positions. We show the variation of W positions for the structures with 12.5%, 25%, and 75% compositions because for the other three cases there is only one combination possible in terms of W positions. Figure 2.2(b) shows the total spectrum of the DMI variation while varying both the Gd and W positions in the structure. Interestingly, for all the cases, the increasing trend of the DMI is

very similar. We conjecture that changing the position of the atoms within the small unit cell will change the nature of the interface that gives variations in the DMI. For example, in the case of Pt/GdCo/W, when Gd atoms placed at position (1,7), the SOC energy change in the interfacial Pt layer is higher than that of position (3,7), which translates to the corresponding DMI as well.

To validate our results against the recent experiment [6], we calculate the surface DMI (in units of pJ/m) by multiplying the calculated DMI, D with the thickness of the magnetic layers. In our calculations, we use the thickness as $N_F a/\sqrt{3} = 4.6$ Å for the magnetic layers, while the experimental thickness is 5 nm. Figure 2.2(c) shows the surface DMI from both the DFT calculation and the experiment, scaled by their respective thickness. In the experiment, a non-zero DMI of 0.56 pJ/m (solid black line) is found for the Pt/GdCo/Pt structure because of the asymmetry in the bottom and the top interfaces due to the difference of interface roughness and intermixing [6]. On the contrary in our DFT model, we use a perfect crystal structure that gives a near zero DMI for the symmetric cases (a small non-zero DMI might arise from intrinsic asymmetry within a thin crystalline GdCo film modeled here). We find an overall matching trend between the DFT and experimental data for the rest of the compositions. An exact quantitative agreement between the DFT results and the experiment is difficult to achieve because we use a crystal structure for our model, whereas, in the experiment, amorphous or polycrystalline materials are used. Additionally, the magnetization also differs between our model and the experiment as the thickness and the dimensions of the structure are different. However, we argue that the structural imperfections in the experiment amount to an ensemble averaging over the various configurations we theoretically explore, so that the experimental data falls in the middle of the spectrum (gray shaded area) of our DFT data.

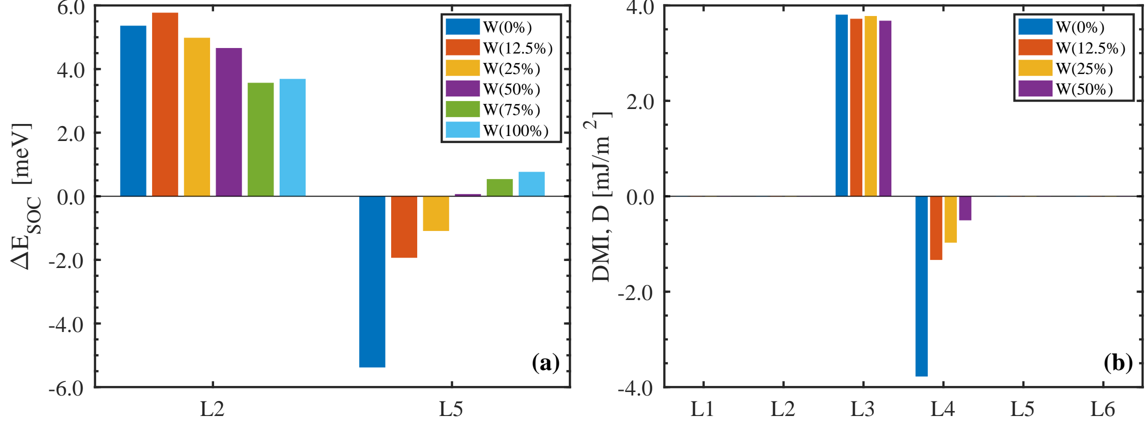


Figure 2.3: (a) Change in SOC energy at the interfacial HM layers (L2 & L5) as a result of changing spin chirality of the magnetic layers (L3 & L4) from CW to ACW. All the color bars on the left (right) side represents the SOC energy change at L2 (L5) for different W compositions. (b) Layer resolved DMI for structures having W composition, $x = 0\% - 50\%$.

2.1.3 Change in SOC energy dictates the DMI trend

In Fig. 2.2, the DMI increases in a non-linear fashion as a function of W composition as opposed to a linear increase one may expect. This non-linear trend can be explained by the change of spin-orbit coupling energy, ΔE_{SOC} in the HM layers adjacent to the magnetic layers in Fig. 2.1. The ΔE_{SOC} is defined by the change in SOC energy as the spin orientation changes from CW to ACW at the magnetic layers (L3 & L4 in Fig. 2.1). In Fig. 2.3(a), we show the ΔE_{SOC} per atom in L2 (adjacent to the bottom magnetic layer) and L5 (adjacent to the top magnetic layer) for all W compositions ($0\% - 100\%$). We find that ΔE_{SOC} in L5 changes drastically as W composition changes from 0% to 12.5%, slowing down thereafter. On the other hand, distributions of ΔE_{SOC} in L2 are not very sensitive to the W composition. Although we find a relatively lower ΔE_{SOC} at L2 for 75% and 100% W compositions, the corresponding ΔE_{SOC}s at L5 are positive. In trilayer structures, the DMIs of the bottom and top interface are additive [45, 54], so that the sum arising from L2, and L5 accounts for the observed non-linear change of DMI in Fig. 2.2. From our findings, we conjecture that the symmetry breaking plays a vital role on the DMI while the effect of W composition is not that prominent, in agreement with the recent experiment [6].

To corroborate our analysis, we calculate the layer resolved DMI contribution from a single

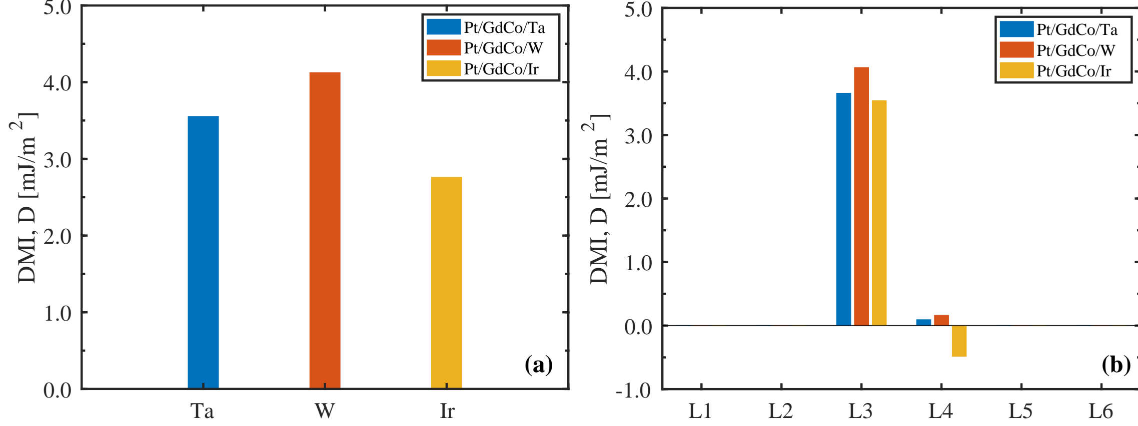


Figure 2.4: (a) Calculated DMI in Pt/GdCo/X, where X=Ta, W, Ir. (b) Layer resolved DMI.

layer by changing the spin orientation of that particular layer of interest from CW to ACW while the spins of all the other layers remain in the original ground state configuration [48]. For instance, when we calculate the DMI contribution of L3 (Fig. 2.1), we only change the spin orientation of L3 from CW to ACW (along x and z), and we constraints the spins of all other layers along y, keeping the antiferromagnetic ground-state coupling between Gd and Co magnetic moments. Figure 2.3(b) shows the layer resolved contribution of the DMI for the structures with $0\% - 50\%$ W composition. The results show that the DMI comes only from the interfacial magnetic layers. We can see that the change in the DMI contribution from the top interfacial layer (L4) with increasing W is small, generating a similar trend as ΔE_{SOC} shown in Fig. 2.3(a). Additionally, the contribution from the bottom interfacial layer (L3) remains almost the same throughout the range of W compositions. The addition of the DMI from the bottom and the top interfaces produces a saturation in the overall DMI curve.

2.1.4 Tuning DMI through cap layer heavy metal

Finally, our theoretical model allows us to explore the tuning of DMI in ferrimagnetic systems with different cap layer compositions, which could be critical in designing suitable materials for hosting ultrasmall high-speed skyrmions. Furthermore, for applications, skyrmions can be driven by current-induced SOT [69]. Changing the cap layer HM offers the ability to

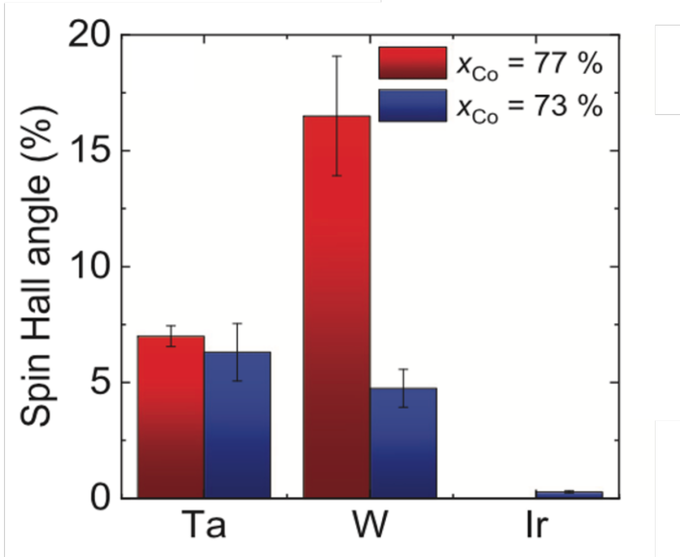


Figure 2.5: Experimentally measured spin Hall angle in Pt/GdCo/X, where X=Ta, W, Ir, showing a similar qualitative trend as the calculated DMI. The figure is reprinted from [7] with permission.

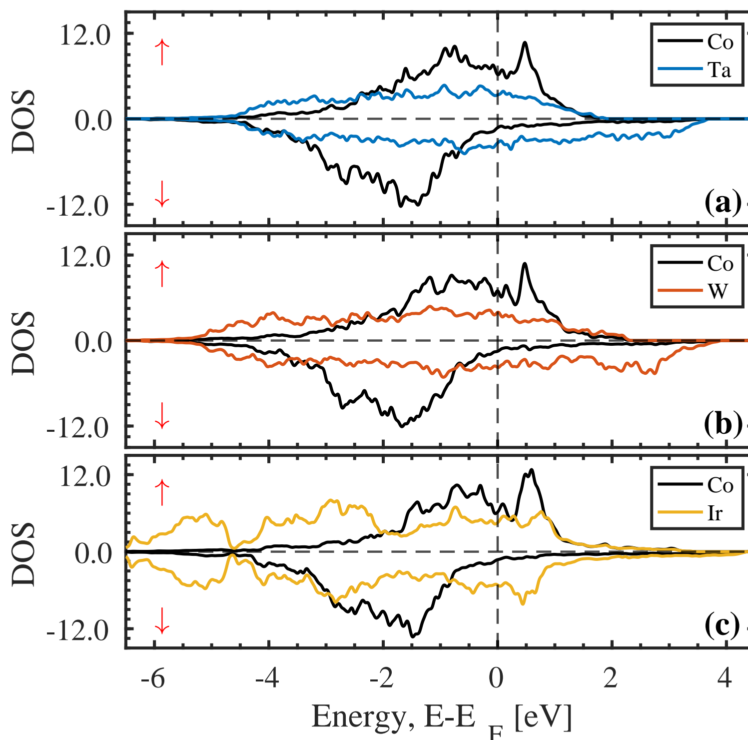


Figure 2.6: Projected density of states (p-DOS) showing 3d-5d band overlap between Co (black) and X (colored) in Pt/GdCo/X. (a) X=Ta, (b) X=W, and (c) X= Ir. The red up (down) arrow represents the spin-up (spin-down) channel.

tune the SOT efficiency and DMI simultaneously. We report the DMI of Pt/GdCo/X where X = Ta, W, Ir, to demonstrate the effect of cap layer $5d$ transition HM on the DMI in Fig. 2.4(a). We choose the cap layer as X = Ta, W, Ir because we want to maximize the SOT for fast and energy-efficient dynamics in addition to tune DMI. W and Ta are known for their giant spin-Hall angle [70, 71], and previous studies have shown an additive DMI for a ferromagnet sandwiched between Pt and Ir [44, 54], which motivates us to explore these structures and see which one of them has the largest DMI. We find that W in the cap layer favors higher DMI than Ta and Ir. The trend of the calculated DMI is in agreement with the trend of experimentally measured spin Hall angle (SHA) via spin-torque ferromagnetic resonance (ST-FMR) as shown in Fig. 2.5. A higher SHA indicates a greater SOT efficiency, suggesting that the Pt/GdCo/W is a suitable structure for hosting small and stable skyrmions for energy-efficient applications, facilitating the interplay between SOT and DMI [7]. To explain the DMI trend, we calculate the layer resolved DMI contribution from bottom and top interfaces, as shown in Fig. 2.4(b). From Fig. 2.4(b), we can observe that the DMI contribution from the top interface (L4) is large when Ir is used as a cap layer material while the DMI contributions are smaller for the cases of W and Ta. The observed trend of the DMI can be explained qualitatively by the Co $3d$-X $5d$ band overlap, which controls the corresponding orbital hybridization. Figure 2.6 shows the projected density of states (p-DOS) of Co-$3d$ and HM-$5d$ orbitals. Clearly, in Co/Ir, the band overlap around the Fermi level is higher than that of Co/W and Co/Ta, which in turn produce larger DMI contributions from L4 for Ir over W and Ta. The band overlap of Co/W and Co/Ta are close to each other. However, we note that the sign of the DMI contribution from the top interface is different for Ir than Ta and W. By analyzing the atom resolved orbital projected density of states of the cap layer HM, we find that for Ta and W, $d_{xy}, d_{x^2-y^2}$ are the major contributors near the Fermi-level while d_{xz}, d_{yz} have smallest contributions. On the other hand, for Ir, d_{xz}, d_{yz} has significant contributions near the Fermi-level. We correlate this behavior with the DMI contribution from L4 (Fig. 2.4(b)), i.e, the Ir 5d states near Fermi-level contribute differently

than Ta and W, which translate qualitatively into the DMI behavior from the magnetic layer adjacent to the cap layers. Moreover, the variation of the DMI sign depending on the adjacent HM has previously been seen in both theoretical and experimental studies [49, 72]. Finally, adding the DMI contribution from both the interfaces (Fig. 2.4(b)) gives a smaller overall DMI for Pt/GdCo/Ir because of the large negative contribution from the top interface.

2.1.5 Conclusion

In summary, we demonstrate the impact of W composition in the cap layer of Pt/GdCo/$Pt_{1-x}W_x$ trilayer structures using first principles calculations. We find excellent tunability of the DMI that shows a tendency of saturation with increasing W composition. The saturating trend of the DMI is attributed to the change of SOC energy at the top and the bottom intefacial HM layers as a function of W composition. Moreover, we find DMI sensitivity to the structural variation. We also demonstrate the DMI variation in Pt/GdCo/(Ta, W or Ir). We find W in the cap layer provides a higher DMI than Ta and Ir, due to the varying degree of orbital hybridization controlled by the band overlap between $3d$-$5d$ orbitals at the cap layer interface. Pt/GdCo/W also requires less current to move skyrmions, which makes the system suitable for energy-efficient applications. Our results provide critical insights to the control mechanism of DMI in ferrimagnetic GdCo based systems, providing a path towards manipulating skyrmion properties for spintronic applications.

2.2 Engineering energy barrier via notches for skyrmions' positional stability

Conventional memory technology is bottlenecked by the delay in fetching instruction sets between the logic cores and the memory elements that are slower to scale. Magnetic skyrmions [43,73,74] bear non-trivial properties, such as topological protection [75], ultra-small size [23], high-speed [76], current-induced motion including nucleation and annihilation [47,77]. Therefore, they have emerged as potential candidates for digital and analog information-carrying bits in low-power high-density, fast, all-electronic non-volatile memory, and logic applications [9,22,25,78–80]. Skyrmions originate and stabilize by the DMI [41,42] in systems lacking inversion symmetry while tuning the DMI [6,7,40,48] controls skyrmion size [46] and overall stability.

Racetrack memory [81] is one of the common platforms studied in the context of skyrmionics [25,35,82,83]. In a skyrmion-based Boolean racetrack memory, information is encoded by the presence (bit '1') and absence (bit '0') of skyrmions at a particular position. For an analog domain utilization of a skyrmion racetrack, such as a native temporal memory for race logic [25], the information is encoded directly into the spatial coordinates of the skyrmions that can be translated back into the timing information of wavefront duty cycles carrying out the race logic operations [84]. The positional stability of skyrmions is a critical issue for both of these applications because a randomly displaced skyrmion can alter the bit sequence in Boolean memory applications and change the spatial coordinates hence the encoded analog timings in race logic applications. For reliability, it is essential to guarantee the positional stability of skyrmions for a certain amount of time. For instance, for long-term memory applications, it requires positional stability of years, while for cache memory, hours or even minutes would be sufficient. In an ideal racetrack, skyrmions are susceptible to thermal fluctuations and exhibit Brownian motion leading to diffusive displacement [85,86]. Moreover, skyrmions show inertia-driven drift shortly after removing a current pulse rather

than stopping immediately. One way to control such undesirable motion is by engineering confinement barriers such as defects created by local variations of material parameters and notches etched into the racetrack, which ensures the pinning of skyrmions [47,87]. Notches can provide a controlled localization of the skyrmion, and using multiple of them in a racetrack allows digitizing the positional information of the skyrmion. It would ensure a deterministic behavior for any skyrmionic device that relies on the position of the skyrmion in the racetrack, ranging from traditional long-term memory applications to temporal computing applications where the temporal data are stored in the spatial coordinates of the skyrmion.

The interactions and dynamics of skyrmions with defects and other pinning sites such as notches have been studied over the past few years [87–90]. References [47, 91] reported skyrmions displacement by the current induced spin-torque with the presence of defects (defects were realized by notches with varying anisotropy). Nucleation of skyrmions in a constricted geometry has been discussed in ref. [92]. Reference [93] used notch to suppress the clogging of skyrmions bit in a racetrack while ref. [94] incorporated notch to do logic operations. Reference [95] discussed the coupling between mobility and breathing mode of skyrmions in a racetrack where periodic notches with varying DMI are placed opposite to each other. Few of the studies have discussed the energy barrier associated with the pinning sites peripherally [85, 96, 97]. Recently, notches have been used to achieve positional stability in skyrmion and domain wall-based artificial synapses [98–101]. Nonetheless, what is missing is a systematic analysis of a notched racetrack, the mechanics of the energy barrier, and its impact on skyrmion mobility, stability, and unintended nucleation and annihilation which sets its operating limits. The combination of required positional stability and operational current range defines an optimal 'Goldilocks' regime in parameter phase space, which is the focus of this work.

In this section, we systematically analyze our ability to produce and tune the energy barrier in a skyrmion-based racetrack with notched geometry (Fig. 2.7) using micromagnetic simulations to achieve positional stability of skyrmions for traditional long-term memory

applications as well as the unique temporal computing applications. In particular, we vary the material parameters DMI (varying skyrmion sizes), the geometry of the notches (Fig. 2.8a), and thickness of the racetrack (Fig. 2.8c) to achieve a high tunability of the energy barrier for long-term positional stability of skyrmions. We demonstrate that the energy barrier is attributed to the constriction in the skyrmion sizes arising from the notch created in the racetrack. Additionally, we come up with an empirical equation based on our simulations (Fig. 2.9). Furthermore, we explain and compare our simulated data with the analytical energy equations of skyrmions on an unconfined infinite plane plus a phenomenological confinement correction that shows an excellent match (Fig. 2.10a). The quantitative difference between the energy of skyrmions on an unconfined infinite plane and our simulated data is attributed to the different geometric boundary conditions (Fig. 2.10b). We also explore other pinning sites, such as local variations of material parameters to put the notched geometry into perspective with other types of defects (Fig. 2.11). Moreover, we show the energy barrier dependence on the shape of the notch (Fig. 2.12). Besides, we find that the required unpinning current is small enough for skyrmion-based devices to be integrated with electrical circuits (Fig. 2.13). Finally, we predict the energy barrier and construct mathematical expressions utilizing ML-based surrogate models (Figs. 2.15–2.17). Our results provide a path forward towards practical, reliable skyrmion-based racetrack memory applications.

2.2.1 Computational details

We perform the simulations using mumax^3 [102], a micromagnetic simulator that solves the Landau-Lifshitz-Gilbert (LLG) equation. The dimensions of the racetrack are length $L = 800\ nm$, width $W = 200\ nm$, and thickness $t_F = 5\ nm$. The simulation mesh is divided into $400 \times 100 \times 5$ grids with a cell size of $2\ nm \times 2\ nm \times 1\ nm$ without considering periodic boundary condition. We note that the choice of cell size is much lower than the exchange length l_{ex} ($\sim 33\ nm$). We use GdCo material parameters such as exchange stiffness $A_{ex} = 7\ pJ/m$, anisotropy $K_u = 50\ kJ/m^3$, saturation magnetization $M_s = 100\ kA/m$ throughout all the

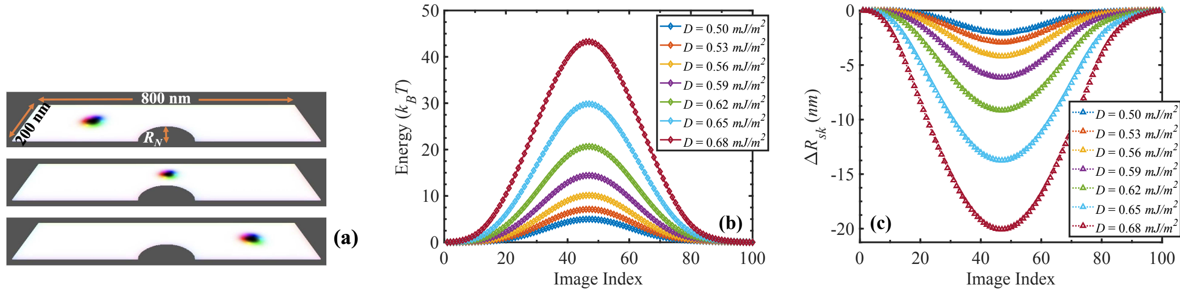


Figure 2.7: (a) Simulated racetrack with notched geometry. Each of the figures represents different snapshots of the skyrmion trajectory along the racetrack, referred to as 'image index'. (b) The energy vs image index for the optimal trajectory of a skyrmion with varying DMIs in a 800 $nm \times 200$ nm racetrack with notch radius, $R_N = 100$ nm. The left valley, peak, and right valley correspond to the image index shown in the top, middle, and bottom panels of Fig. 2.7(a) respectively. (c) Change in skyrmion radius, ΔR_{sk} when the skyrmion passes over the notch. Higher DMIs initiate larger initial skyrmions that undergo bigger shrinkage and larger energy costs when forced through the constriction.

calculations unless otherwise specified [8, 9, 23]. We use varying interfacial DMIs to control the size of skyrmions because the DMI can be easily tuned by interface engineering [6, 40]. We calculate the minimum energy path (MEP) using the String method [103, 104]. The basic idea of the string method is to find the transition path by evolving a curve (string) connecting two endpoints along the energy landscape and the reparametrization of the string by interpolation [103, 104]. It is an iterative method that continues until the path converges to the MEP with the desired accuracy. In our simulations, we use 100 iterations to calculate the MEP.

2.2.2 Tuning E_b through material and physical parameters

Figure 2.7(a) shows the schematic of a racetrack with notched geometry. We create the semi-circular notch (radius R_N) by removing materials from the racetrack. The snapshots represent different positions of the skyrmion trajectory (referred to as 'image index') along the racetrack as it moves from one side to the other side of the notch. Figure 2.7(b) shows the total energy obtained from the MEP calculations for a racetrack with notch radius 100 nm and thickness 5 nm (recall the racetrack width is 200 nm). The zero for energy

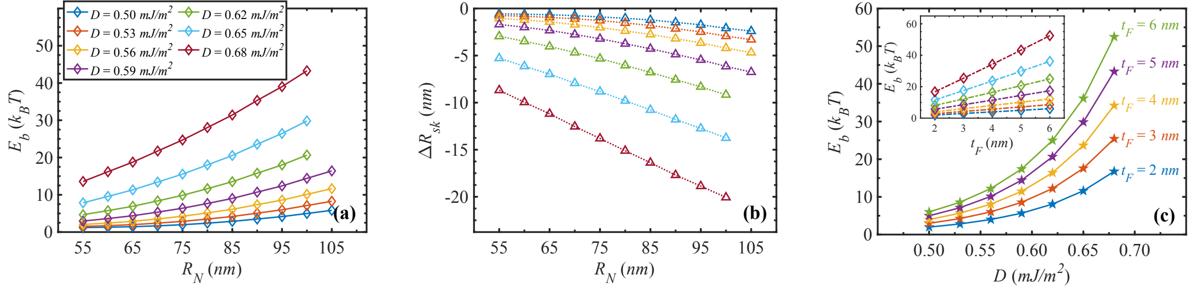


Figure 2.8: Material parameters dependence of the energy barrier, E_b. Effect of notch size (radius R_N) on (a) E_b, and (b) ΔR_{sk}, both for varying DMIs. We see that a larger R_N leads to a larger $|\Delta R_{sk}|$ that corresponds to a higher barrier. (c) Thickness dependence of E_b in a racetrack ($R_N = 100\ nm$). The inset shows that E_b increases linearly as a function of racetrack thickness, t_F for any specific D. The linear variation of E_b vs t_F is consistent with the overall uniform cylindrical shape of the skyrmion at ultrathin limit. The color code to represent DMI variations in (a), (b), and inset of (c) are the same.

is set as the energy of the first image index in the simulation domain shown in Fig. 2.7(a) top. We vary the interfacial DMI, D from 0.50 mJ/m^2 to 0.68 mJ/m^2 and find energy barriers range from $\sim$ 5 to 45 k_BT. For the used exchange and anisotropy, the critical DMI, $D_c = 4\sqrt{A_{ex}K}/\pi = 0.71\ mJ/m^2$, which gives the ratio D/D_c from 0.70 to 0.96 where the effective anisotropy, $K = K_u - \frac{1}{2}\mu_0 M_s^2$ [105,106]. We find that the energy barrier results from the change in the skyrmion radius, ΔR_{sk} that arises from the interaction between skyrmion and the edge of the notch [96,107] as the skyrmion passes past the notch. We conjecture that the reduction of the skyrmion size in squeezing through the constriction produces the energy barrier. Figure 2.7(c) shows a series of ΔR_{sk} corresponding to the energy plots shown in Fig. 2.7(b). As the D increases, for a specific exchange and anisotropy, the skyrmion size gets bigger, making it harder to squeeze through and generating a higher energy barrier. The positional lifetime of the skyrmion is often described using an Arrhenius form $\tau = f_0^{-1} e^{E_b/k_BT}$ where f_0 is the attempt frequency, and E_b is the height of the gaussian energy profile (energy barrier). Approximately an E_b of 30 k_BT (35 k_BT) will provide positional lifetime in seconds (days) for $f_0 = 10^{10}\ Hz$. 45 k_BT energy barrier will give a lifetime in years. We note that estimating the exact attempt frequency is still an open question, and the value can be higher due to the entropy effect [108–110]. However, we use $10^{10}\ Hz$ because the most commonly

accepted estimate for the attempt frequency is in the range of $10^9 - 10^{10}$ Hz for magnetic materials [111–114]. We calculate the MEP for various notch sizes and demonstrate the impact on the energy barrier in Fig. 2.8(a) for $R_N = 55 - 105$ nm. We find that for all the DMIs, the energy barrier increases as R_N increases, by reducing the size of the skyrmion, consistent with Fig. 2.8(b). As R_N increases, the skyrmion size shrinks more, $|\Delta R_{sk}|$ gets larger, and the energy barrier increases proportionally. However, if we continue to increase R_N, at some point, the skyrmion starts annihilating as the region over the notch is insufficient to pass through and the skyrmion touches the notch boundary and the edge of the racetrack. For instance, from Fig. 2.8(a), we can see that throughout the range of the DMIs ($D = 0.50 - 0.68$ mJ/m^2), skyrmions pass through without annihilation up to $R_N = 100$ nm. For a larger notch, for example, $R_N = 105$ nm, the skyrmion gets annihilated when D is greater than 0.59 mJ/m^2. We also vary the thickness t_F of the racetrack for several DMIs, and find an increase in energy barrier height for a thicker racetrack for a specific D. Figure 2.8(c) shows the thickness dependence of the energy barrier in a racetrack with $R_N = 100$ nm. We get an energy barrier of ~ 45 k_BT for a ~ 45 nm skyrmion ($D = 0.68$ mJ/m^2) in a moderately thick (5 nm) racetrack, which ensures years long lifetime that makes the device suitable for storage class memory applications. The inset shows the linearity of E_b as a function of t_F, which dictates that we can increase the energy barrier even further by increasing the thickness of the racetrack. Clearly, we can get a large enough energy barrier for smaller skyrmions as well in a thicker racetrack.

2.2.3 Deriving empirical equations for estimating E_b

To quantify E_b, combining the data we get by varying the D, R_N, and t_F, we come up with a fitted empirical equation, normalized by t_F. Fig. 2.9(a) shows the simulated E_b vs R_N data are perfectly described by the fitted curves of the form $y = a(e^{bx} - 1)$. We find that a is an exponential function of D, while b is a quadratic function of D, as shown in Fig. 2.9(b) and

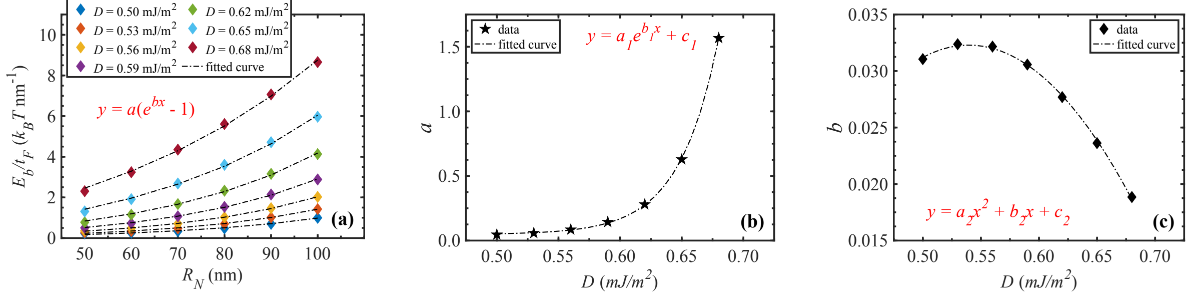


Figure 2.9: (a) Fitting of E_b (normalized by t_F) with respect to R_N. (b), (c) The relation between the fitting constants obtained from (a) and D. The red colored texts in each graph represent the fitting function.

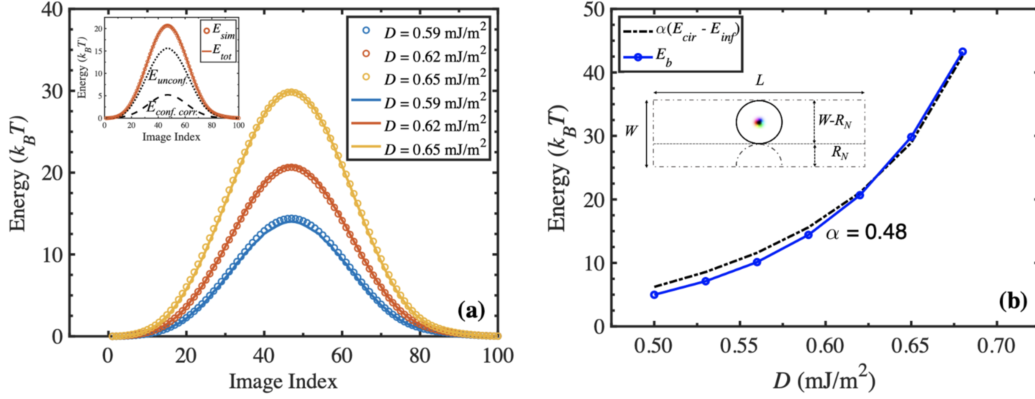


Figure 2.10: (a) Energy landscape of skyrmions in a racetrack ($R_N = 100\ nm$) from the MEP simulations (scatter circles) and analytical equations (solid curves) [8,9]. The analytical equations include the energetics of 2π skyrmions on an infinite plane plus a phenomenological confinement correction. (b) Explanation of the mismatch between simulated and analytical (calculated using Eq. 2.4) data in a racetrack ($R_N = 100\ nm$). A skyrmion confined above the pinning site has the energy of a skyrmion confined in a circle of diameter $W - R_N$ above the notch minus that of an unconfined infinite plane skyrmion. Including only one fitting term can describe E_b with good accuracy for varying D values. The black circular region above the notch in the inset schematic represents the simulation geometry for the confined case.

2.9(c) respectively. The final form of the equation is

$$E_b/t_F = a(e^{bR_N} - 1) \tag{2.2}$$

where a and b are related to D as follows

$$\begin{aligned} a &= a_1 e^{b_1 D} + c_1 \\ b &= a_2 D^2 + b_2 D + c_2 \end{aligned} \tag{2.3}$$

where the prefactor constants are material specific. For GdCo, $[a_1,\ b_1,\ c_1] = [6.874 \times 10^{-10},\ 31.64,\ 0.04783$, and $[a_2,\ b_2,\ c_2] = [-0.6922,\ 0.7471\ , -0.1693]$. The units of E_b, t_F, D, and R_N are in k_BT, nm, mJ/m^2, and nm respectively. It is worth mentioning that the form of the Eq. (2.2) is physically meaningful as it gives $E_b = 0$ when $R_N = 0$, which is expected. As energy of skyrmion increases linearly with t_F for the limit of uniform cylindrical shape of skyrmion (inset of Fig. 2.8(c)) at the limit of ultrathin films ($\lesssim 10\ nm$), the t_F is simply a scaling factor.

We compare our simulated data with the analytical equation derived for skyrmions [8,9]. The different energy terms that give total energy equation of the skyrmions on an infinite plane are derived as [8,9]

$$\begin{aligned} E_{\text{ex}} &= (2\pi A_{ex} t_F)\left(\frac{2R_{sk}}{\Delta} + \frac{2\Delta}{R_{sk}} N_{sk}^2\right) f_{\text{ex}}(\rho) \\ E_{\text{DMI}} &= -\left(2\pi R_{sk} t_F\right) \pi D f_{\text{DMI}}(\rho) \\ E_{\text{ani}} &= \left(4\pi K_u t_F\right) R_{sk} \Delta f_{\text{ani}}(\rho) \end{aligned} \tag{2.4}$$

where $R_{sk},\ \Delta,\ N_{sk}$ are skyrmion radius, domain wall width, and skyrmion winding number respectively. The form factors for small size, obtained by fitting numerical simulations, are given by [8,9]

$$\begin{aligned} f_{\text{ex}}(\rho) &\approx \left[1 + 1.93 \frac{\rho\,(\rho - 0.65)}{\rho^2 + 1} e^{-1.48(\rho - 0.65)}\right] \\ f_{\text{ani}}(\rho) &\approx \left[1 - \frac{1}{6\rho} e^{-\rho/\sqrt{2}}\right] \\ f_{\text{DMI}}(\rho) &\approx \left[N_{sk} + \frac{1}{2\pi\rho} e^{-\rho}\right] \end{aligned} \tag{2.5}$$

where $\rho = R_{sk}/\Delta$. Figure 2.10(a) shows that our simulated energy profiles (scattered circles) are perfectly matched with the analytical equation (solid curves) that includes the energetics of skyrmions within 2π model on an infinite plane $E_{unconf.} = E_{ex} + E_{DMI} + E_{ani}$, plus a phenomenological confinement correction $E_{conf.\ corr.}$ (see the dotted and dashed black curves in the inset of Fig. 2.10(a) for $E_{unconf.}$ and $E_{conf.\ corr.}$ respectively; adding up these two curves give the E_{tot}). $E_{conf.\ corr.}$ can be expressed as

$$E_{conf.\ corr.} = Ae^{-(\frac{x-B}{C})^2} \tag{2.6}$$

where A is the magnitude, $B = 47$, $C = 20$, and x is a placeholder for 'image index'. A is a quadratic function of R_N and D, and can be expressed as

$$\begin{aligned} A &= pR_N^2 + qR_N + r \\ p &= p_2D^2 + p_1D + p_0 \\ q &= q_2D^2 + q_1D + q_0 \\ r &= r_2D^2 + r_1D + r_0 \end{aligned} \tag{2.7}$$

where $[p_2,\ p_1,\ p_0] = [-0.0661,\ 0.0648,\ -0.0142]$, $[q_2,\ q_1,\ q_0] = [11.8989,\ -11.8985,\ 2.7966]$, and $[r_2,\ r_1,\ r_0] = [-134.0478,\ 100.5548,\ -10.6283]$. We note that we use $N_{sk} = 1$ and the value of R_{sk} and Δ obtained from simulations while calculating the energy from the Eq. (2.4). However, Eq. (2.4) alone fails to capture the simulated energy profiles because it assumes an unconfined planar geometry, while in our simulations, we use a confined geometry. We conjecture that our simulated energy barrier will be somewhere in between the skyrmion energy on an unconfined infinite plane and that for a confined circular region around the notch. To verify, we calculate the static energy of skyrmions both for a circle (E_{cir}) above the notch region with diameter $W - R_N$ (see the inset of Fig. 2.10(b)), and for an unconfined infinite plane (E_{inf}). In Fig. 2.10(b), we show the energy difference between E_{cir} and E_{inf} for

a racetrack with $R_N = 100\ nm$. We find that our simulated E_b matches with $\alpha(E_{cir} - E_{inf})$ for the entire range of D, where α is a prefactor dependent on the racetrack geometry. We compare $\alpha(E_{cir} - E_{inf})$ and E_b for other notch radii and find overall agreement while α varies as a function of notch radius.

2.2.4 Alternate pinning sites: local variation of material params

We explore alternate pinning mechanisms to compare the energy barrier among them. One common approach to introduce pinning sites is a local variation of the material parameters in a specific region in the racetrack [47, 87, 115–117]. In practical systems, the variation of material parameters can be achieved by naturally occurring and intentional defects, grain boundaries, composition, thickness gradient in the thin films, voltage gating, modulating the heavy metal layer, etc. We create the pinning sites by locally varying K_u, A_{ex}, D, and M_s. We vary one parameter at a time while the other parameters remain constant throughout the racetrack. Figure 2.11(a) shows the energy barrier for different pinning sites, including the fully notched geometry for a 5 nm thick racetrack having a semi-circular pining site of 100 nm radius. It appears that a racetrack with a fully notched geometry produces the highest energy barrier compared to the rest, which attributes to the largest change in skyrmion radius while passing over the notch as shown in Fig. 2.11(b). Additionally, notches are easier to create experimentally than controlling local variations of material parameters.

2.2.5 Estimating E_b for rare-earth Mn_4N

We also calculate E_b for another promising material Mn_4N for skyrmion-based spintronics applications [118–121]. Mn_4N is a ferrimagnet and attractive for hosting small and speedy skyrmions [121]. In a 5 nm thick Mn_4N racetrack ($R_N = 75\ nm$), we find an E_b of $\sim 45\ k_BT$ for a $\sim 40\ nm$ skyrmion, while for GdCo with identical R_N, t_F, and R_{sk}, the E_b is $\sim 22\ k_BT$. The used parameters for Mn_4N are $A_{ex} = 15\ pJ/m$, $K_u = 110\ kJ/m^3$, $M_s = 105\ kA/m$ [118]. Our finding suggests that Mn_4N offers a higher E_b than GdCo, which is mainly because of

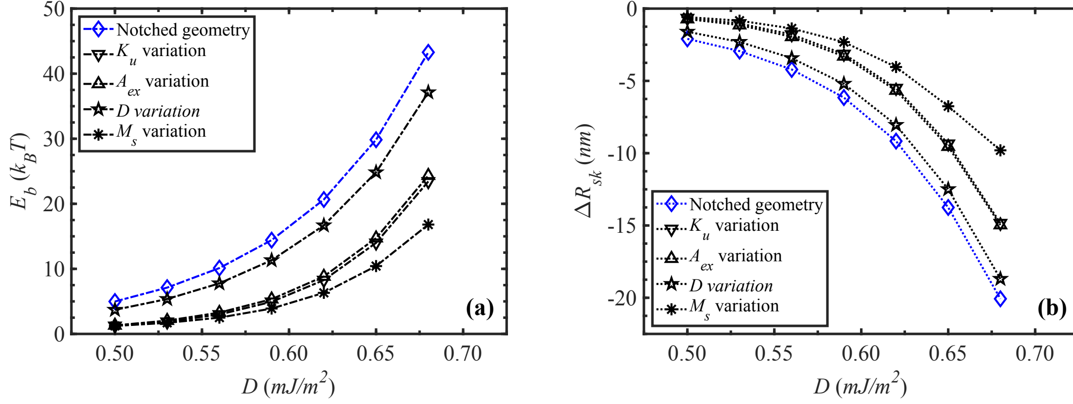


Figure 2.11: (a) Energy barrier for different pinning sites in a 5 nm thick racetrack. For all the cases, the pinning site is a semi-circular region of 100 nm radius. K_u and A_{ex} of the pinning site are 2 times higher, and D and M_s are 10 times lower than the rest of the track region. We choose the ratio that gives the highest E_b. (b) Change in skyrmion radius corresponding to the E_b in (a), shows a proportional relation between E_b and $|\Delta R_{sk}|$.

the higher exchange stiffness of Mn_4N. Needless to say that the E_b can be further increased by tuning the R_N and t_F of the racetrack.

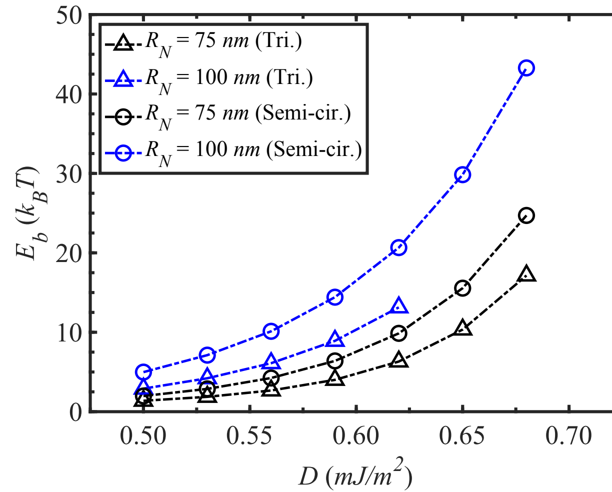


Figure 2.12: Energy barrier of a triangular notch with 75 nm (black) and 100 nm (blue) depth in comparison with semi-circular notch.

2.2.6 Alternate notch shape: triangular notch

We have focused on semi-circular notches so far, however, the shape of the notch is important for the dynamics of the skyrmions, including their nucleation and annihilation [91–94]. As a case study, we show the energy barrier associated with a triangular notch with 75 nm and 100 nm depth in comparison with the semi-circular notch of the same radii in Fig. 2.12 for a 5 nm thick racetrack. We see that for the same skyrmion size, a triangular notch produces a lower E_b than the semi-circular one. Moreover, annihilation occurs for smaller skyrmion size in the case of a triangular notch. For instance, for 100 nm notch depth, skyrmion gets annihilated when D is greater than 0.62 mJ/m^2 as shown in Fig. 2.12. These findings show that a semi-circular notch is more favorable to achieving a high energy barrier. Nonetheless, one important thing to note is that the trend of the energy barrier is qualitatively similar irrespective of the notch shape. Therefore, our derived quasi-analytical equations hold for other notch shapes as well with different prefactor constants. Similarly, our explanation of simulated E_b in terms of $\alpha(E_{cir} - E_{inf})$ also works for triangular shaped notch for a different fitting parameter α. For instance, in a 5 nm thick racetrack with 75 nm notch, $\alpha = 0.37$ for semi-circular shape while for triangular shape $\alpha = 0.25$.

2.2.7 Large energy barrier yet low unpinning current

While a barrier is needed to hold the skyrmion in place, it is equally important to ensure that the critical current to depin the skyrmions is not too large, as that would cause unacceptable energy dissipation when integrated with the peripheral circuitry, not to mention random skyrmion annihilation, and even occasional unintended nucleation [122]. The energy barrier can be tuned by varying various knobs such as materials parameters and notch geometries. However, we need to optimize it to get a high enough hold time for the skyrmions yet require a moderate unpinning current.

Figure 2.13 shows the unpinning current of racetracks with 95 nm and 100 nm notch radii, and 5 nm thickness. The current density distribution for the racetrack with the notch

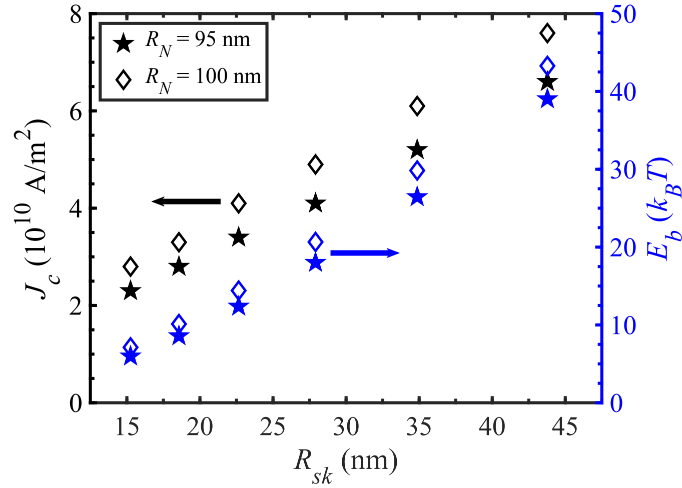


Figure 2.13: The critical current J_c (black) to unpin the skyrmion and the corresponding E_b (blue) for a 5 nm thick racetrack as a function of R_{sk}. We see a low unpinning current with a fairly large energy barrier. The arrows point each colored data to the corresponding y-axis.

is calculated using COMSOL Multiphysics® [123]. We employ SOT with a SHA $\theta_{sh} = 0.15$ while estimating the unpinning current. We use current pulses ranging from 8 ns to 25 ns to unpin the skyrmions. We find that bigger skyrmions need a shorter pulse and the critical current increases as E_b increases. The energy barrier increases faster with radius than the critical current, which would help us to get a large enough barrier and a small enough unpinning current. We find moderate critical currents for large energy barriers. For instance, a ~ 45 nm skyrmion can be unpinned with currents of 6.6×10^{10} A/m^2 and 7.6×10^{10} A/m^2 while the corresponding energy barriers are ~ 40 k_BT and ~ 45 k_BT respectively, which are orders of magnitude smaller than the critical current required to unpin domain walls [81, 119]. Moreover, the obtained critical currents are significantly lower than the nucleation current ($> 10^{12}$ A/m^2) of the skyrmions in a constricted geometry [92, 122], which prevents any unintended nucleation of skyrmions during the unpinning process.

2.2.8 Prediction of E_b using surrogate ML models

As discussed, the energy barrier E_b associated with a notch in a racetrack that provides the positional stability of skyrmions depends on various physical and material parameters, i.e.,

D, R_N, t_F etc. Finding a Goldilocks regime for the energy barrier and unpinning current requires optimization of multiple variables. However, these micromagnetic simulations are computationally expensive. For example, one $\{D, R_N, t_F\}$ parameter combination requires a run time of $\sim$ 6 hours in a graphical processing unit (GPU). *A priori*, the optimal $\{D, R_N, t_F\}$ combination that will give rise to long positional lifetime τ for every skyrmion host material is not known. High-throughput strategy to sweep the entire parameter space in a brute force manner is not efficient. We develop a closed-loop, adaptive ML framework for building surrogate models[1] for micromagnetic simulations that enable rapid prediction of E_b (output) as a function of $\{D, R_N, t_F\}$ (inputs) using a previously unexplored acquisition function in the materials science literature, referred to as information condensing active learning (ICAL) [124]. We compare its performance with the traditional parametric standard deviation (SD) method. Moreover, we uncover insights from the trained black-box models using global-level post-hoc model explainable methods, namely partial dependency plots (PDP) and global feature importance (GFI), and finally, construct mathematical expressions based on grammatical evolution (a symbolic regression technique) for predicting E_b as a function of $\{D, R_N, t_F\}$ [125]. The grammatical evolution method uses the data set constructed from the adaptive learning framework and leverages the insights gleaned from PDPs to build the analytical expression. The results from grammatical evolution are also compared with the well-known genetic programming method [126–129]. We demonstrate the efficacy of this approach using the magnetic parameters of the ferrimagnetic Mn_4N/HM system.

A. Materials informatics workflow

Our overarching computational workflow is shown in Fig. 2.14. There are four components in this workflow: (*i*) Defining the search space ($\mathcal{S}$) and selecting the training data for building ML models. We enumerated a total of 330 unique combinations of input descriptors $\{D, R_N, t_F\}$ to define $\mathcal{S}$. The initial selection of the data points for training the ML models can be

[1]ML model is developed by Prof. Prasanna Balachandran's group (MSE, UVA)

chosen either randomly, use of a space-filling design method such as the Latin hypercube sampling (LHS) [130], recommendation from the domain experts, or based on the available data from surveying the published literature. We used the LHS method to select 20 data points from $\mathcal{S}$. We calculated the E_b for the 20 data points using the protocol discussed earlier based on micromagnetic simulations as implemented in the mumax^3 package [102]. (*ii*) Closed-loop adaptive learning to build surrogate ML models for the micromagnetic simulations [131,132]. Here, the goal is to build a surrogate model $f : \mathbf{X} \mapsto \mathbf{y}$ that will rapidly learn a mapping, which is representative of the entire search space. We approximate f using the ensemble of support vector regression (eSVR) models. It is common in the literature to use the parametric SD as the acquisition function for such problem formulation [133], where a data point with the largest eSVR predicted SD in the unexplored search space (referred to as virtual space $\mathcal{V}$) will be recommended for validation and feedback. In addition to SD, we also explored a new acquisition function referred to as the ICAL, which was first introduced by Jain et al. for batch or pool-based active learning [124]. Unlike the parametric SD, the ICAL acquisition function does not make any distribution assumption. We implemented two independent adaptive learning loops to compare SD and ICAL acquisition functions. As a result, we built two independent surrogate models: eSVR-SD and eSVR-ICAL for the micromagnetic simulations. The adaptive learning loop was terminated after 50 iterations. (*iii*) Post-hoc model explanation of the trained surrogate models. To uncover insights about the trained eSVR models, we used the GFI analysis [134] and PDPs [135]. In addition, we use the PDPs to extract candidate expressions that best fit the $E_b = f(X_i)$ relationship, where X_i is an input descriptor. (*iv*) Extracting mathematical expressions using grammatical evolution [125]. We combine the candidate expressions obtained from each PDP to inform the grammatical evolution algorithm in a novel way that is not explored in the literature, which in turn returns a final expression that best captures the variance in the data. The performance of grammatical evolution is also compared with the well-known symbolic regression methods such as genetic programming [126, 127].

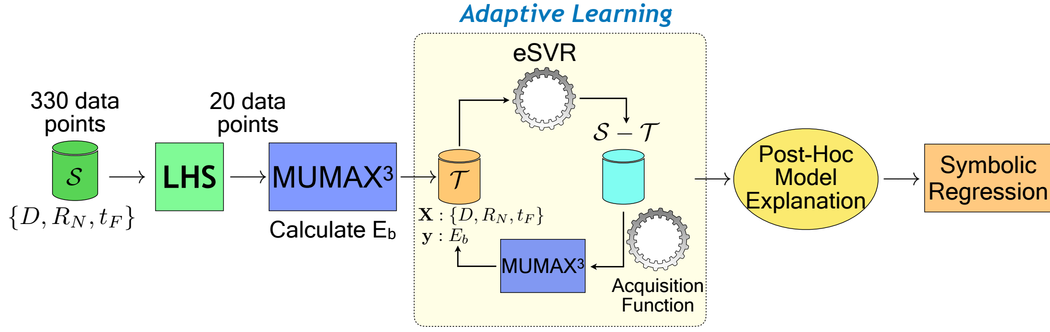


Figure 2.14: Schematic representation illustrating the integration of closed-loop adaptive learning workflow with post-hoc model explainable techniques and symbolic regression. We start by enumerating all possible combinations of the three inputs ($\mathbf{X}$) $\{D,\ R_N,\ t_F\}$, which defines our large search space ($\mathcal{S}$). A subset of 20 data points were selected from this search space using the space-filling Latin hypercube sampling (LHS) method. We then calculated E_b (our output, $\mathbf{y_t}$) for these 20 data points using mumax^3. We now have a training data set ($\mathcal{T}$), $\mathcal{T} = (\mathbf{X_t}, \mathbf{y_t})$. The next step is the closed-loop adaptive learning framework. We used an ensemble of support vector regression (eSVR) to map a quantitative relationship between $\mathbf{X_t}$ and $\mathbf{y_t}$. The trained eSVR model was then used to predict the E_b for the data points in the virtual set $\mathcal{V}$, where $\mathcal{V} = \mathcal{S} - \mathbf{X_t}$. We then rank each data point in $\mathcal{V}$ using two independent acquisition functions: the non-parametric ICAL and parametric SD. The top ranked candidate is recommended to mumax^3 for the calculation of E_b. We then augment $\mathcal{T}$ with the new data point and iterate. The entire adaptive learning loop was setup to run autonomously. The next step is the post-hoc analysis of the trained black-box eSVR models based on global feature importance and partial dependence plots, which helps us in uncovering the functional relationships between each input descriptor in $\mathbf{X_t}$ and the predicted E_b as learned by the eSVR models. The outcomes from post-hoc model explanation methods inform the symbolic regression. We employed the grammatical evolution method to accomplish this task. The outcome from grammatical evolution is an interpretable mathematical model to predict $E_b = f(D, R_N, t_F)$.

B. Performance of the ML models

The mumax^3 simulation parameters [1–4] and the Mn_4N search space is given in Table 2.1 and Table 2.2, respectively. Figure 2.15 shows the predictive performance of the trained eSVR models on training and previously unseen test data. Both models have high R^2 (> 0.95) on the unseen test data, indicating good prediction performance. However, Fig. 2.15(b) indicates that the eSVR-ICAL model underestimates the E_b value when $E_b > 60\ k_BT$, while the eSVR-SD model (Fig. 2.15(a)) does not show such a trend. One of the reasons for observing this trend is attributed to the way in which the SD and ICAL acquisition functions have recommended new data points in the adaptive learning loop. It is important to note that our input descriptors span a three-dimensional space. Our analysis reveals that the SD acquisition function predominantly sampled the edges, faces, and corners of the search space at the rate of 62%, 24%, and 14%, respectively. Intriguingly, the SD acquisition function did not sample a single data point in the interior of the three-dimensional discrete input descriptor space. In contrast, the ICAL acquisition function sampled the edge, face, and corner of the search space at the rate of 14%, 46%, and 2%, respectively. Thus, the ICAL acquisition function also sampled the interior of the search space 38% of the time. Despite the similar R^2 score, we notice the key difference in the sampling strategy between the two acquisition functions. The impact of sampling is also seen in the GFI analysis. The final eSVR-ICAL model ranked D as the most important descriptor followed by R_N and t_F. Although the eSVR-SD model also ranked D as the most important descriptor, both R_N and t_F were ranked as approximately equal.

In Fig. 2.16, we compare the PDPs for the trained eSVR-SD and eSVR-ICAL models. The PDPs capture the marginal effect that an input descriptor has on the predicted outcome of a ML model. Since we have an ensemble of SVR models, we calculated the marginal dependency for each input descriptor within each SVR model in the ensemble. Figure 2.16 shows the averaged marginal effect for an input descriptor that is representative of the entire ensemble of models and the error bars reflect the associated standard deviations. However, the size of

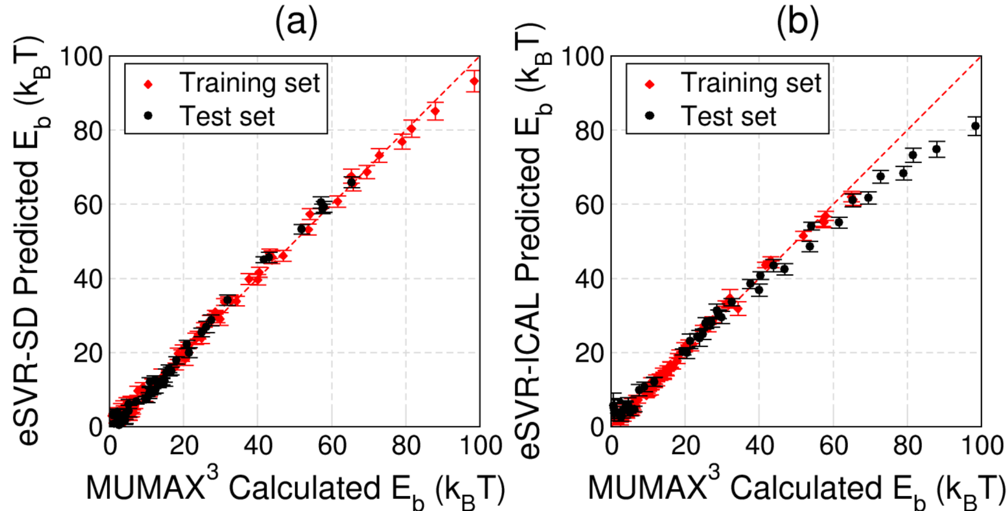


Figure 2.15: Comparison of the predictive power of the trained eSVR models on their training set (red diamond) and unseen test set (black circle) that was not used for training the models. Parity plot of (a) eSVR-SD model with $R^2 = 0.986$ for the training set and $R^2 = 0.989$ for the test set, (b) eSVR-ICAL model with $R^2 = 0.987$ for the training set and $R^2 = 0.970$ for the unseen test set. In both (a) and (b), X-axis is the mumax^3-calculated E_b and Y-axis is the ML predicted E_b. Although the overall performances of both models are good, differences are evident especially in (b) when $E_b > 60\ k_BT$.

the error bar is smaller than the data points. Despite being trained on different datasets, both eSVR-SD and eSVR-ICAL models have captured similar trends: D and R_N show non-linear relationship with E_b, whereas t_F has a linear relationship with E_b. Intriguingly, a similar conclusion was made based on GdCo data discussed earlier without using any ML approach [5].

The next step involves transforming the data shown in PDPs (Fig. 2.16) into mathematical expressions. To accomplish this, we used a non-linear least squares fitting procedure. We selected a total of nine commonly used mathematical expressions (see Table 2.3) and individually fit those expressions to the data shown in Fig. 2.16. In principle, this list of expressions can be extended to any arbitrary number. In this step, we only focused on D and R_N descriptors because of their non-linear relationship with E_b. For the two input descriptors (D and R_N), the top five best fit expressions are selected as candidates to inform the symbolic regression using the grammatical evolution method. The goal of grammatical evolution is to

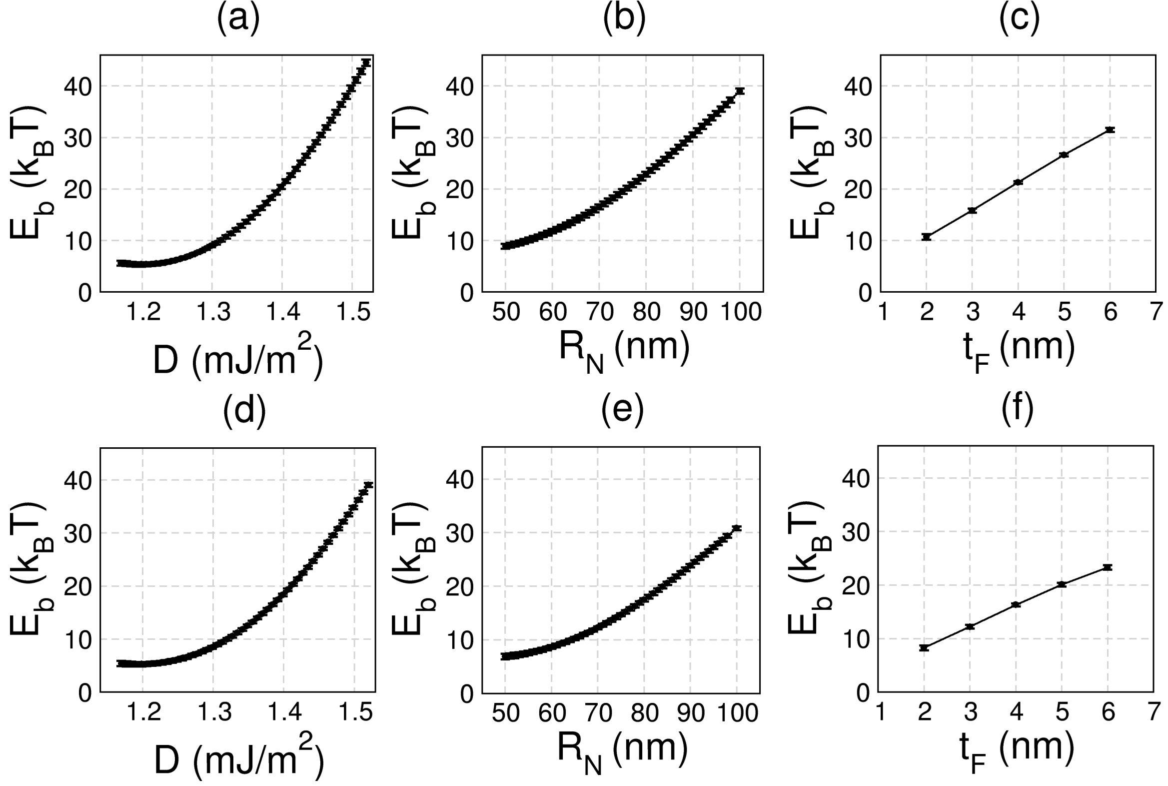


Figure 2.16: Partial dependence plots for Mn_4N. (a), (b), and (c) uses the trained eSVR-SD model to calculate the marginal dependencies; (d), (e), and (f) uses the trained eSVR-ICAL model to calculate the marginal dependencies. The marginal effects of individual features D, R_N and t_F on E_b as learned by the trained ML models are captured by the plots. Descriptors D and R_N show a non-linear relationship with E_b, whereas t_F shows a linear relationship with E_b. Despite the differences in the sampling between SD and ICAL acquisition functions, both models have captured similar trends.

leverage these candidate expressions derived from Fig. 2.16 to construct a final expression that correlates well with the E_b. We designate Eq. (2.8) as the final expression for E_b from grammatical evolution, which can be written in the multivariate power law functional form as follows:

$$\frac{E_b}{t_F} \propto D^8 \times R_N^2 \Rightarrow \alpha \times D^8 \times R_N^2 \tag{2.8}$$

where $\alpha = 5.34 \times 10^{-5} \pm 3.67 \times 10^{-7}$ is the pre-factor optimized by Bayesian inference. The performance of Eq. (2.8) was compared with the eSVR-ICAL model. The results are shown on Fig. 2.17, where the Eq. (2.8) had an R^2 score of 0.989. Thus, using PDP-informed

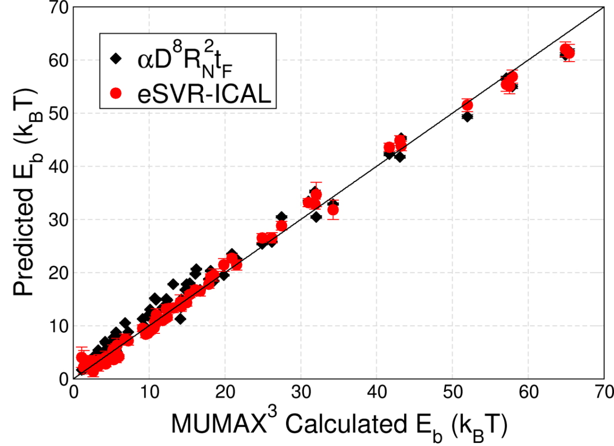


Figure 2.17: X-axis is the E_b calculated from micromagnetic simulations (mumax^3) for Mn_4N and Y-axis is the predicted E_b from $\alpha D^8 R_N^2 t_F$ model given in Eq. (2.8) (filled black diamond) and eSVR-ICAL model (filled red circles). Error bars for Eq. (2.8) are computed from Bayesian inference. In the case of eSVR-ICAL model, error bars represent the ensemble standard deviation of model predictions.

Table 2.1: The micromagnetic parameters used for Mn_4N [1–4]. These parameters are used in the mumax^3 calculation and treated as constants. For comparison, the micromagnetic parameters used by Ref. [5] for GdCo are also given.

Material	**GdCo**	**Mn_4N**
Exchange, Aex (pJ/m)	7	15
Anisotropy, Ku (kJ/m^3)	50	110
Saturation Magnetization, Ms (kA/m)	100	105

Table 2.2: The search space of Mn_4N. Min and Max indicates the lower and upper boundary, respectively. The Resolution means the distance between each sampled value.

Property	**Min**	**Max**	**Resolution**
DMI D (mJ/m^2)	1.17	1.52	0.07
Notch radius R_N (nm)	50	100	5
Thickness t_F (nm)	2	6	1

grammatical evolution, we can also formulate a fitted equation that captures the E_b trend, without sacrificing the accuracy; albeit explanation of the physical meaning behind the D^8 term needs further work.

Table 2.3: Candidate functions used for fitting PDPs. "a" indicates the fitting parameter with an initial value of $a = 1$.

Function list	**Function name**	**Fitting expression**
1	linear_model	$y = a * x$
2	x_2	$y = a * x^2$
3	x_3	$y = a * x^3$
4	x_4	$y = a * x^4$
5	log_func	$y = log(a * x)$
6	exp_func	$y = exp(a * x)$
7	xlnx	$y = a * x * log(x)$
8	lnx2_func	$y = a * (lnx)^2$
9	double_well	$y = a * (x^4 - x^2)$

2.2.9 Conclusion

In summary, we demonstrated that skyrmion positional stability is achievable by creating notches along a racetrack. We presented quantitative analyses, backed by analytical equations for various material parameters, notch geometries, and racetrack thicknesses. An optimal combination of skyrmion size, notch radius, and thickness of the racetrack provides a large enough energy barrier ($\sim 45\ k_B T$) to achieve a positional lifetime of years for long-term memory applications. We found a moderately low minimum critical current to unpin the skyrmion ($\sim 10^{10}\ A/m^2$), which is an essential aspect for low-power operations. Furthermore, we demonstrated a materials informatics workflow that synergistically integrates adaptive learning with post-hoc model explanation and symbolic regression methods to accelerate the prediction of energy barrier. These results provide critical design insights on skyrmionic racetracks, and potentially an argument for reliable, long-term skyrmion-based memory applications.

Chapter 3

Energy barrier engineering in nanomagnets for probabilistic computing

In chapter 2, we have seen that we can utilize the exotic properties of skyrmions for encoding information for abacus-type applications and how we can tune their size, stability, and positional lifetime. However, skyrmions-based technology is still in its early stages, and the material systems are not yet mature enough for industry-level production. Moreover, one of the known challenges is the electrical readout of skyrmions because of their small fill factor under a reading MTJ. Additionally, if we make skyrmions very small, stability becomes a concern, similar to that of traditional nanomagnets. Now, we turn to the traditional nanomagnets, whose energy barrier is proportional to the volume of the magnets. Therefore, downscaling a magnet makes it volatile as the energy barrier becomes small. Although magnets with low barriers are unsuitable for information storage, they can be used for probabilistic, stochastic, and neuromorphic computing [26, 27]. This chapter explores using LBMs in stochastic applications, their reliability issues (Section 3.1), their application in inferencing tasks (Section 3.2), and engineering MBMs to circumvent some issues associated with LMBs (Section 3.3). This chapter is reprinted from Ref. [136] with permission from IEEE coauthored with S. Ganguly, and A. W. Ghosh, from Ref. [137] with permission from

Frontiers coauthored with S. Ganguly, and A. W. Ghosh, and from Ref. [138] coauthored with L. Rehm, A. Shukla, Y. Xie, S. Ganguly, S. Rakheja, A. D. Kent, and A. W. Ghosh.

3.1 Reliability of the computational networks built from low-barrier magnets

Low energy barrier magnet technology, which utilizes nanomagnets with barrier height in the order of thermal energy, has recently been proposed as a potential candidate for hardware accelerators for probabilistic computing and stochastic sampling [29, 139]. These accelerators may be broadly considered as hardware Markov chain Monte Carlo implementation that utilizes the built-in stochasticity provided by the dynamics of the LBM, which results in highly compact devices with true stochasticity, as compared to linear feedback-shift register (LFSR) based pseudo-random number generators (pRNGs) [26]. The magnetization component m_z of the LBM randomly fluctuates between two stable states ($\uparrow$, $\downarrow$) under the influence of the thermal noise, and the probability of getting any one of the two stable states can be driven via an external current [27]. There are a handful of applications ranging from probabilistic computing to machine learning and artificial intelligence that leverage the intrinsic stochastic nature of LBMs [27, 30, 140–142]. The prototype hardware building blocks are the BSNs, popularly known as "p-bits" with programmable weights in a recurrent configuration. An illustrative example of a dual-stacked feedback cross-bar structure is shown in Fig. 3.1(a). The synaptic weights or the "program" is loaded in memristors located at the cross-points of the core cross-bar structure, whereas the neurons are at the peripheries. Using a dual cross-bar structure, it is possible to build recurrent networks, including a Restricted Boltzmann Machine (RBM, Fig. 3.1(b)), an example application area of this accelerator. The RBM is embedded in the computing fabric by enabling certain neurons and synaptic connections while disabling the rest.

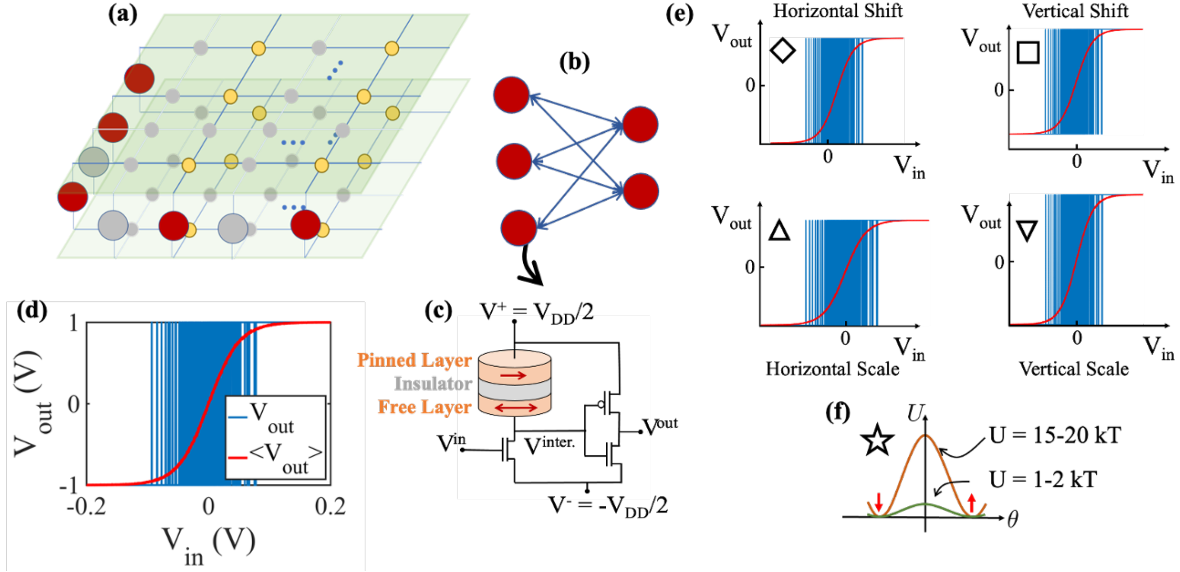


Figure 3.1: (a) Illustrative schematic of an embedded RBM, an energy-based optimization and learning algorithm, in a dual-stacked feedback cross-bar structure with neurons (the compute units) at the edges (large circles), while the synaptic weights (the program) loaded in memristors located at the cross-points of the core cross-bar structure (small circles). The active neurons and synapses are colored bold (red and yellow), while inactive units are greyed out. (b) The RBM network that gets embedded in (a). The bidirectional blue lines represent the synaptic connections between the neurons (red circles). The yellow circles used in (a) are not shown here for simplicity. (c) The design of an LBM MTJ-based p-bit unit. (d) Ideal characteristics of a p-bit device. (e) Schematics of different characteristics distortions. (f) Illustration of energy barrier variation in a nanomagnet. Symbols (diamond, square, etc.) in (e) and (f) represent different variabilities henceforth.

Although BSN-based non-Boolean probabilistic applications are inherently more error resilient than conventional nanomagnet switches used for deterministic Boolean memory and logic applications, the computational reliability of these accelerators that employ LBMs as their hardware RNG, needs to be carefully assessed. Recently, several studies have discussed the impact of geometric, structural, and process variation from device-to-device that can create ignorable to high variability in the characteristics of LBMs depending on the degree of variation [143–145], however, the resulting impact of these "non-idealities" on the computational networks is still largely not understood.

In this section, we discuss the issues of variability in the context of circuits and networks built from LBM-based BSN devices. We categorize the variability into a few broad classes,

namely shifting and scaling of the device characteristics from the ideal as expected from the mathematical model, and the variability of the barrier heights for two broad classes of algorithms that can be solved using p-bits, such as energy minimization-based optimization algorithm (EMOA) and probabilistic graphical algorithm (PGA). EMOA includes problems such as Ising model and RBMs, which seek to define a problem in terms of a thermodynamically definable "energy-landscape" with the embedding of the desired optimal result in the ground/vacuum energy, while PGA includes Bayesian decision diagrams, which do not have an inherent notion of energy and thermodynamics. In terms of network connectivity (using the spectral theorem of linear systems) [146], this implies that the EMOA networks have symmetric or undirected connections, resulting in eigenstates that are real-valued and reachable via real-space computation, whereas PGA networks are asymmetric or directed, resulting in non-real or complex eigenstates not reachable via real-space computation.

We estimate the error per p-bit (EPP) to quantify the performance deviation from the ideal devices. We find the EPP shows a sub-linear saturation for EMOA, while in the PGA, the error grows linearly to super-linearly. Moreover, the networks are found to be more prone to shifting variability than scaling. Additionally, for EMOA, larger networks are less affected by the variability, while for PGA, the trend is the opposite. Our findings may provide a potential path forward toward designing reliable LBM-based hardware accelerators.

3.1.1 Building 'p-bits' using LBMs

The building blocks of MRAM technology are MTJs. MTJs consist of a thin insulator sandwiched between two ferromagnetic layers - a "pinned layer" whose magnetization is fixed and a "free layer" whose magnetization can be reoriented by a spin current (Fig. 3.1(c)). The free layer exhibits a double potential well corresponding to the two easy points (Fig. 3.1(f)). The height of the barrier determines the expected state retention time using the Arrhenius relation given by:

$$\tau = \tau_0 e^{U/k_B T} \tag{3.1}$$

In the above equation, $U(=\mu_0 M_s H_k \Omega/2)$ is the energy barrier, where the symbols respectively stand for permeability of free space, saturation magnetization, magnetic anisotropy field strength, and volume. For a conventional storage class memory, U is set to $40-60\ k_BT$, which yields a decade-long state retention time τ depending on the τ_0, the inverse of attempt frequency that ranges from $0.1-1\ ns$ [147]. However, if the magnet is ultra-scaled by reducing the volume Ω or its profile is made circular, which reduces the H_k by removing the shape anisotropy, the retention time can be scaled down to near τ_0 [148]. In this case, the free layer's magnetization fluctuates between the two easy points under the influence of the thermal noise, which is able to "kick" the magnetization over the barrier with ease, at near GHz frequencies. The LBM-based MTJ typically utilizes CoFeB as the ferromagnetic material and MgO as the insulator layer [139]; however, it's an active field of research. MTJ structure allows this fluctuation to be translated into an equivalent fluctuation in the resistance of the device, i.e., low (high) resistance depending on the parallel (anti-parallel) alignment between the pinned and free layer magnetization, which can be used for building useful devices that can harvest true randomness from the environment.

One such device is the "p-bit", which is a binary stochastic neuron with a compact model given by:

$$V_i^{out} = \mathrm{sgn}[\tanh\left(\beta \mathrm{V_i^{in}}\right) + \alpha \cdot \mathrm{rnd}(-1,+1)]\mathrm{V_{DD}}/2 \tag{3.2}$$

In this device, the output swings between $-V_{DD}/2$ to $V_{DD}/2$ corresponding to -1 and $+1$ state labels of m_z, however, the ratio of these states is controllable by an input signal, which imposes a tanh-like probability distribution. rnd is a uniform random distribution. The parameters β and α represent the transfer gain of the unit and the relative contribution of the stochasticity to the characteristics, respectively. For large scale correlated networks, V^{in} can be represented as:

$$V_i^{in} = \kappa[h_i + \sum_j J_{ij} V_j^{out}/(V_{DD}/2)] \tag{3.3}$$

where j stands for the index over all input devices connected to the particular i-th device, h is

the bias vector, and J is the synaptic matrix. Different functionalities correspond to different choices of h and J. κ is a coupling coefficient representing the inverse of the "temperature" of the system.

3.1.2 Computational details

We implement the compact model of p-bit networks described by (3.2) and (3.3) in MATLAB according to the methodology discussed by Camsari et al. [27]. The MATLAB model is a parameterized version of the compact modeling simulation performed in SPICE [27, 149]. In MATLAB implementation, we use $\alpha = 1$, $\beta = 1$ (for ideal case), $\kappa = 0.8$, and $V_{DD} = 2\ V$ throughout the calculation unless otherwise specified.

We use computational networks constructed from p-bits of varying sizes. For EMOA, we use AND gate and full-adder having J matrices sized 3×3 and 14×14, respectively [27]. We construct an arbitrary symmetric J matrix of 50×50 for a large network. For PGA, we use Bayesian networks (BNs) constructed from 8, 20, and 50 p-bits (J matrices are asymmetric in these cases). For EMOA, EPP is computed by taking the summation of the absolute difference between the output probability distribution of ideal and non-ideal cases, divided by the number of p-bit units in the network. However, for PGA, we calculate the EPP from the difference in the correlation matrix ($\sigma(i,j) = \frac{1}{T}\int_0^T V_i^{out} V_j^{out}\, dt$) between the ideal and non-ideal cases. Normalization through the number of p-bit units allows us to examine how the average error changes with an increase in the number of p-bit units in the network. Note that throughout the paper, EPP represents the average absolute output error introduced by each p-bit unit in the network. For both algorithms, we use $T = 10^6$ simulation steps to get to the V^{out}. If the sample generation time is 2 ns, this is equivalent to 2 ms of compute time. The mean and standard deviation of the EPP are calculated from $N = 100$ simulations.

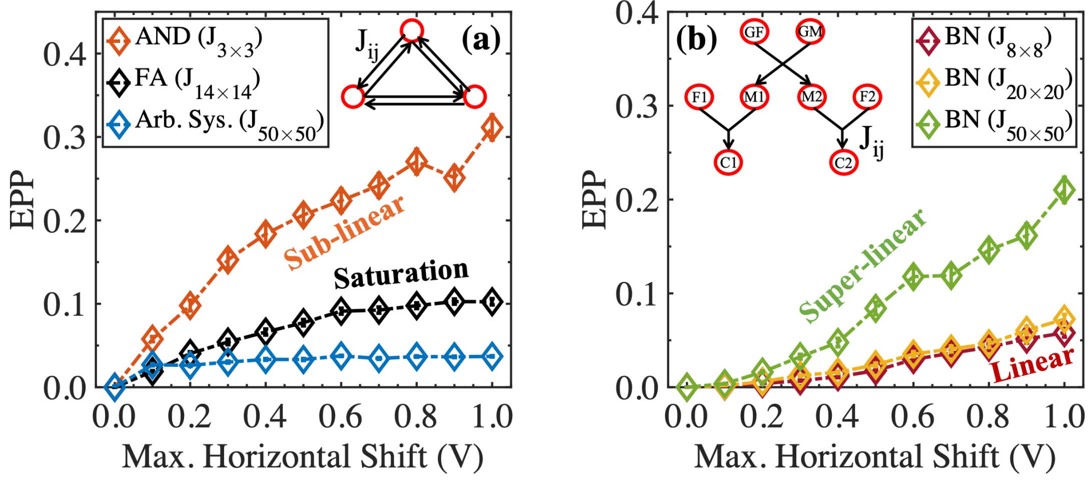


Figure 3.2: EPP from horizontal shifting for (a) EMOA and (b) PGA with different network sizes (size of the J matrix). Fig. 3.2 and all the subsequent figures show the mean EPP and the error bar represents the standard deviation. The mean and standard deviation are calculated from 100 simulations. The figures use different colors to represent various network sizes, and different symbols represent different distortions introduced in Fig. 3.1(e). Inset in (a) shows the schematic of a 3×3 EMOA network, while inset in (b) shows the schematic of an 8×8 PGA Bayesian network representing a family tree (GF: Grandfather, GM: Grandmother, etc.).

3.1.3 Impact of variability on the "p-bit" performance

LBM devices are hybrids of silicon CMOS, which is a highly mature technology, and spintronics/magnetics, which is a relatively new technology. While they have been successfully integrated into the context of high energy barrier storage class MRAM technology by several commercial vendors, its LBM variant comes with lithographic challenges that may require a long process of technological developments to perfect. These lithographic challenges mainly concern the quality of magnetic films and the precision control over their geometry. Ref. [143] studied the impact of geometrical irregularities, such as dimples, holes, shape variance, etc. on the characteristic correlation times of LBMs and found that the distribution of correlation times can be large. These kinds of variations can have implications that are beyond the intrinsic behavior of the free-layer magnet of the MTJ itself. In particular, two critical sets of variations are discussed next. Note that these variations become relevant in the context of circuits and networks built from these devices.

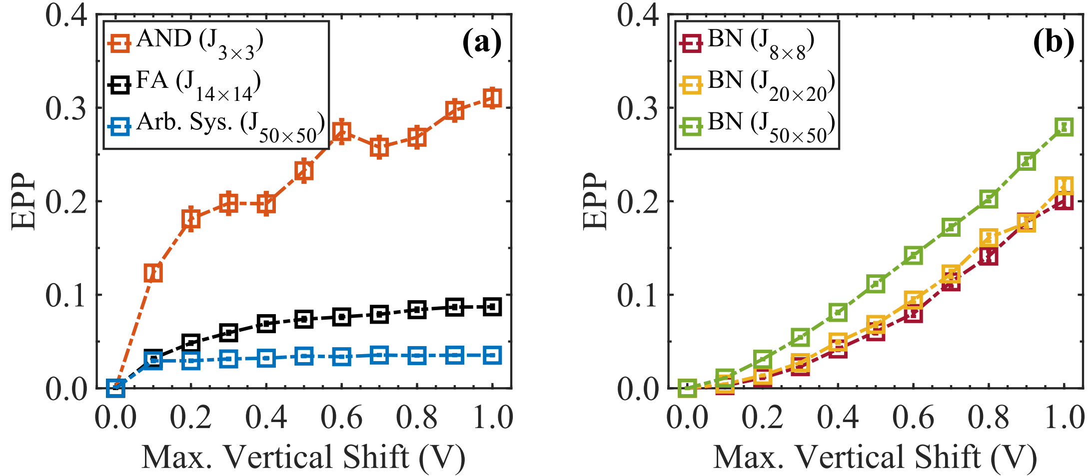


Figure 3.3: EPP from vertical shifting for (a) EMOA and (b) PGA with different network sizes.

A. Characteristics distortion

Fig. 3.1(c) and 3.1(d) shows the proposed device and its ideal output characteristics, respectively. The characteristics of the device depend on the swing that is generated by the NMOS transistor turning on or turning off, balanced around the MTJ's characteristic resistance, i.e., the resistance of the transistor in the linear intermediate mode should match the MTJ's average resistance. In the linear mode of operation, the tanh shape shows up as an interplay between the MTJ's average and transistor's intermediate resistance as it swings from on to off, while the MTJ's magnetization flipping adds the fluctuation on the characteristics. A mismatch between these two can lead to a deviation from the "ideal" model presented in (3.2). Moreover, variations from the ideal can occur as a result of limitations of the CMOS components within the device, especially due to the limited transconductance gain. We categorize the variations into four categories that broadly cover the phase space of such distortions (shown in Fig. 3.1(e)): 1. horizontal shift; 2. vertical shift; 3. horizontal scale; 4. vertical scale. Non-uniformity in the circular magnet often leads to asymmetry in parallel and anti-parallel MTJ states. It tends to bias toward one state over another thermodynamically, which might lead to a horizontal shift in the p-bit characteristics. Additionally, resistance

mismatch between the MTJ and the NMOS and read disturbance can lead to a vertical shift. Variation in the gain β can lead to horizontal scale, while vertical scale might arise from loading effects from follow-on p-bits that the output stage may not be able to handle adequately because of the weak buffer and large fan out. Fig. 3.2 shows the EPP that emerged from horizontal shifting in the networks for both the EMOA and PGA problem classes. We vary the maximum voltage shift from 0 V to 1 V. From Fig. 3.2(a), we find that for AND gate, the error increases rapidly up to a $\sim$ 20% horizontal voltage shift and slows down afterward. However, the error has an overall increasing trend. For larger networks, the error starts saturating at $\sim$ 10% voltage shift. For AND gate, we find a maximum of $\sim$ 30% EPP corresponding to a horizontal voltage shift of 1 V. We see that as the network size increases, the error percentage decreases for EMOA. On the other hand, for PGA, from Fig. 3.2(b), we can see that the error increases almost linearly as a function of horizontal voltage shift. However, the relation between the error and the network size is opposite to that of EMOA. Figs. 3.3 – 3.5 show the EPP that emerged from vertical shifting, horizontal scaling, and vertical scaling, respectively, for both EMOA and PGA. The increasing trend of the EPP is similar for different types of distortion; however, the error percentage varies depending on the problem class, distortion type, and network size. We list the maximum error arising from different types of distortion in Fig. 3.6, where different colors represent the overall trend of EPP. It is important to note that we vary only one type of distortion at a time.

B. Energy barrier variability

It is clear from (3.1) that a small variation in the energy barrier U can lead to a large variation in the expected state retention time τ. This translates to a circuit encountering widely different time scales or a large dynamic range of operation within its individual components. This can lead to significant issues with the operational viability of a circuit built from p-bits. We, therefore, analyze the effect of energy barrier variation on the performance of the networks. As a result of the energy barrier variation, the magnetic states of different

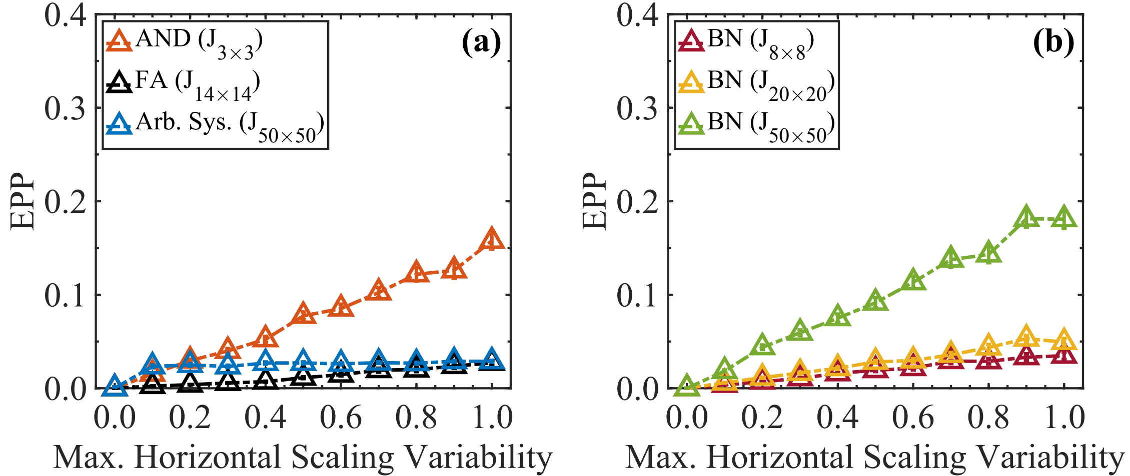


Figure 3.4: EPP from horizontal scaling for (a) EMOA and (b) PGA with different network sizes.

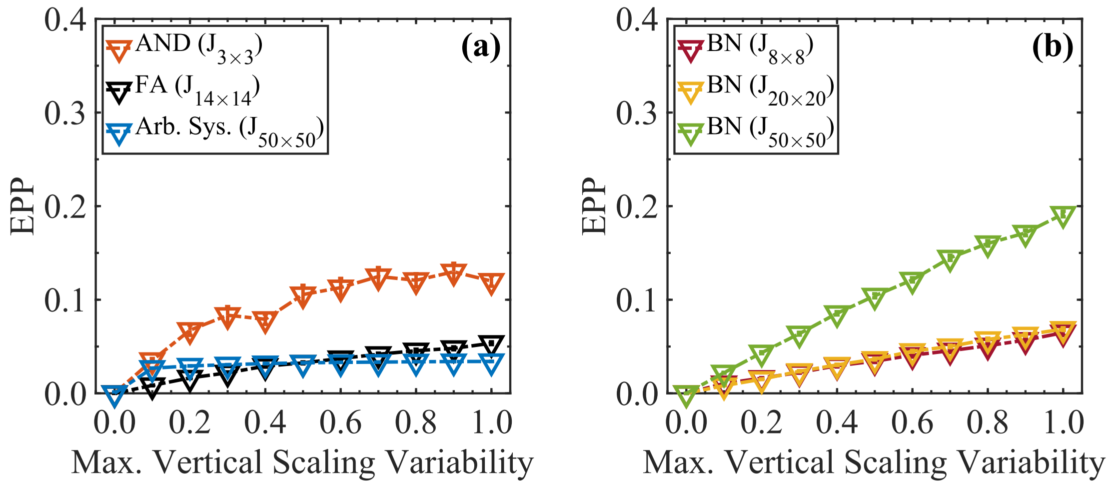


Figure 3.5: EPP from vertical scaling for (a) EMOA and (b) PGA with different network sizes.

nanomagnets update at different times than the ideal case (assuming 0 k_BT energy barrier), leading to an overall error in the output quantity. Fig. 3.7 shows the EPP for EMOA and PGA arising from energy barrier variability. We find that for both classes of problems, the error percentage is small (within $\sim$ 10%) up to an energy barrier variation of $\sim$ 10 k_BT. For EMOA, the impact of a high energy barrier variation in a small network is severe ($\sim$ 40% error), while the large network seems more forgiving in terms of error ($\sim$ 4% error). On the contrary, the trend is the opposite in the case of PGA. We find a maximum of $\sim$ 50% error for a large-sized

Problem Class	Network	EPP (%)			
		H. Sh. (1 V)	V. Sh. (1 V)	H. Sc. V. (1)	V. Sc. V. (1)
EMOA	AND (J_{3x3})	31.1 ± 1.5	31.0 ± 1.0	15.8 ± 0.9	12.0 ± 0.5
	FA (J_{14x14})	10.3 ± 0.4	8.7 ± 0.2	2.7 ± 0.3	5.4 ± 0.1
	Arb. Sys. (J_{50x50})	3.7 ± 0.1	3.5 ± 0.05	2.9 ± 0.1	3.4 ± 0.01
PGA	BN (J_{8x8})	5.8 ± 0.5	20.1 ± 0.8	3.5 ± 0.4	6.5 ± 0.3
	BN (J_{20x20})	7.3 ± 0.5	21.7 ± 0.6	4.9 ± 0.4	6.9 ± 0.1
	BN (J_{50x50})	21.0 ± 1.4	27.9 ± 0.3	10.0 ± 1.0	19.1 ± 0.3

Figure 3.6: EPP arising from different characteristics distortions ($N = 100$). The green, light green, light red, and red colors represent the saturation, sub-linear, linear, and super-linear trends of EPP, respectively, as a function of characteristics distortion.

BN. Note that the sample generation time is determined by the fastest magnet. Relying on the slowest magnet would result in reduced computational throughput and extended computation time. Therefore, we continue the computation by updating the magnetic state of the fastest magnet while the slower magnets update their states intermittently on average. This results in temporary freezing out of certain magnets during computation, leading to an increase in the EPP. This suggests a trade-off between computational speed and accuracy, which requires further investigation in future studies. Also, note that the characteristics distortions are not included while taking into account the energy barrier variability.

3.1.4 Sampling vs. Simulated annealing

The results discussed above for EMOA are calculated using the sampling technique, based on a fixed interaction strength κ (pseudo-inverse temperature) throughout the simulation, and run the simulation long enough time (10^6 steps) so that the p-bits visit primarily the low-energy state. Fig. 3.8 shows the EPP using simulated annealing in comparison with the sampling technique. We vary κ from 0.5 to 5 after every 2×10^5 steps while calculating the output using the simulated annealing technique. We find that the error percentage is slightly

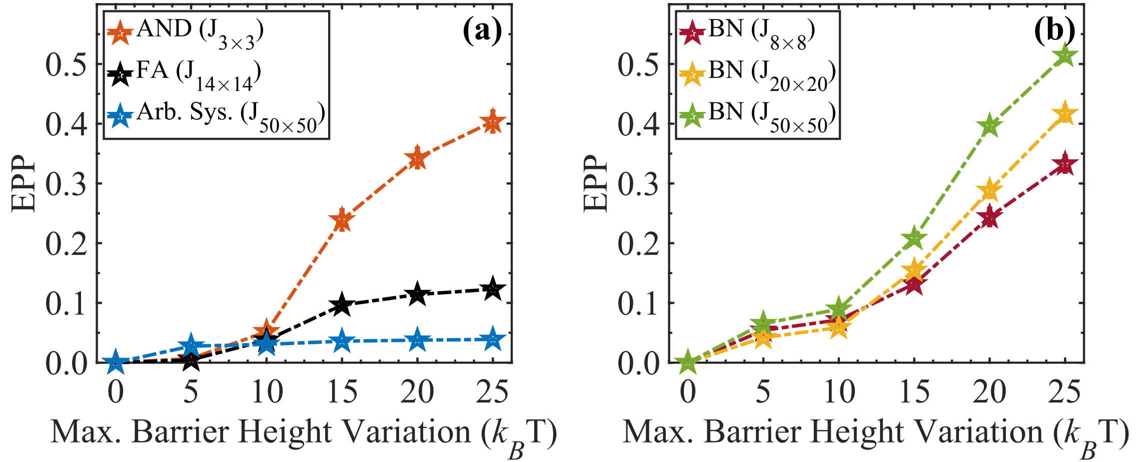


Figure 3.7: EPP from energy barrier variability for (a) EMOA and (b) PGA with different network sizes.

higher for all types of characteristics distortions for the simulated annealing technique. We conjecture that this is because the sampling method, when run long enough, can cover the system's phase space better ergodically than a linear simulated annealing schedule, which is in essence a guided importance sampling for a shorter time, may not be able to sample the phase space as comprehensively to discover the true ground state. This may be improved by more complex annealing schedules, which we do not discuss further.

3.1.5 Conclusion

In summary, we quantify the impact of non-idealities in computational networks built from LBM-based BSNs using two different techniques. In all the possible variances studied in this work, the error shows a sub-linear saturation at the extremal device variability points for EMOA, while in the PGA, the error grows linearly to super-linearly. We conjecture that this is because, in EMOA, the system tries to seek a single thermodynamically favorable fixed point in a finite phase space, which limits the growth of error, whereas, in PGA, there is no similar principle that can check the growth of the error. Additionally, running multiple samples of the same problem with different random seeds (thereby simulating the "real world") helps in reducing the variance of the error, but not its mean value. This suggests that for

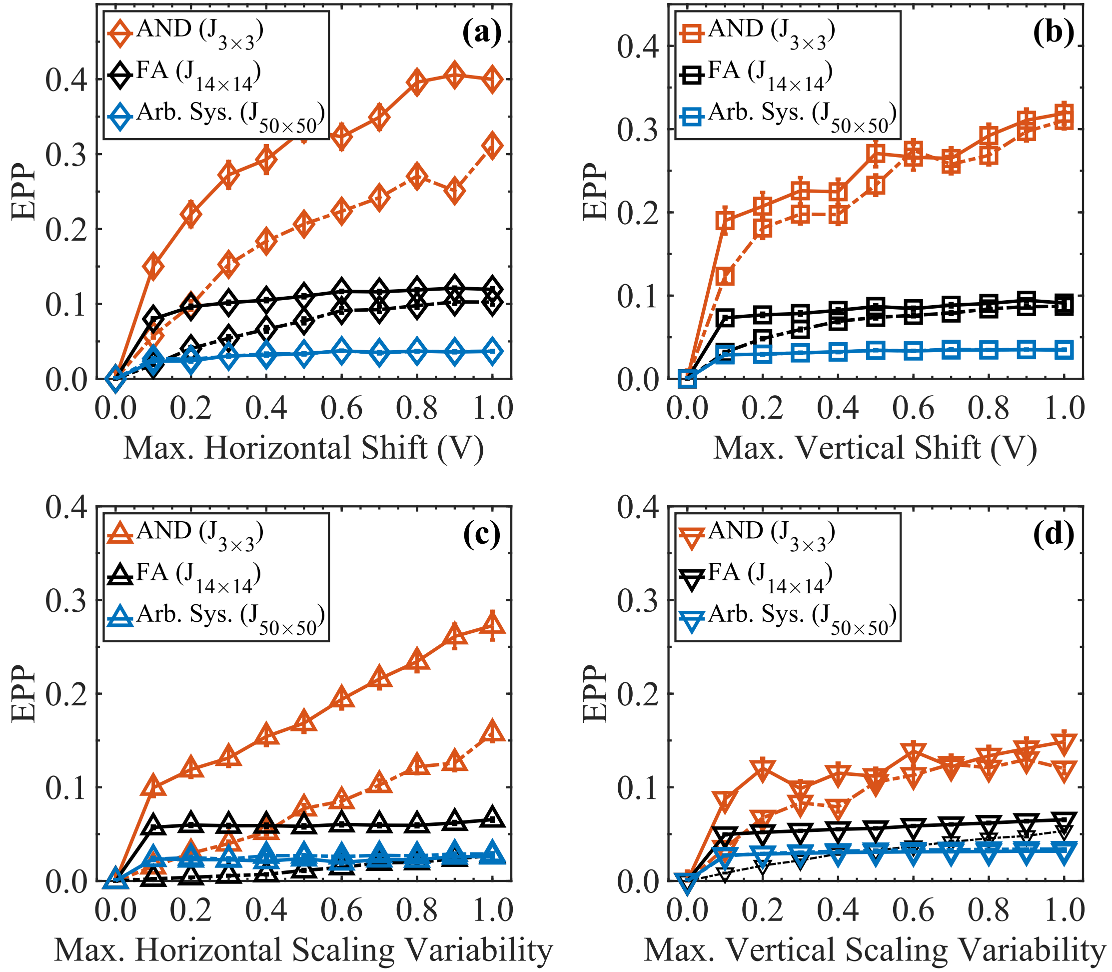


Figure 3.8: EPP calculated using sampling technique (dashed line) vs. simulated annealing technique (solid line) for EMOA for (a) horizontal shifting, (b) vertical shifting, (c) horizontal scaling, and (d) vertical scaling.

a certain amount of device variability, the average error is fixed, which may be estimated or characterized beforehand, and the results are certified accordingly. These findings may provide critical design insights for building suitable LBM-based hardware accelerators.

3.2 Employing stochasticity for inferencing tasks

High-performance computing has historically developed around the Boolean computing paradigm, executed on silicon (Si) CMOS hardware. In fact, software has for decades been developed around the CMOS fabric that has singularly dictated our choice of materials, devices, circuits, and architecture – leading to the dominant processor design paradigm: von Neumann architecture that separates memory and processing units. Over the last decade, however, Moore's law for hardware scaling has significantly slowed down, primarily due to the prohibitive energy cost of computing and an increasingly steep memory wall. At the same time, software development has significantly evolved around "Big Data" paradigm, with ML and artificial intelligence (AI) dominating the roost. Additionally, the push towards the internet of things (IoT) edge devices has prompted an intensive search for energy-efficient and compact hardware systems for on-chip data processing [150].

One such direction is neuromorphic computing, which uses the concept of mimicking a human brain architecture to design circuits and systems that can perform highly energy-efficient computations [151–155]. A human brain is primarily composed of two functional elemental units - synapses and neurons. Neurons are interconnected through synapses with different connection strengths (commonly known as synaptic weights), which provide the learning and memory capabilities of the brain. A neuron receives synaptic inputs from other neurons, generates output in the form of action potentials, and distributes the output to the subsequent neurons. A human brain has $\sim 10^{11}$ neurons and $\sim 10^{15}$ synapses and consumes $\sim 1-10$ fJ per synaptic event [156–158].

To emulate the organization and functionality of a human brain, there are many proposals for physical neuromorphic computing systems using memristors [159–161], spintronics [162–164], charge-density-wave (CDW) devices [165], photonics [166, 167], etc. In recent years, there has been significant progress in the development of physical neuromorphic hardware, both in academia and industry. The hierarchy of neuromorphic hardware implementation spans from the system level to the device level and all the way down to the level of the

material. At the system level, various large-scale neuromorphic computers utilize different approaches - for instance, IBM's TrueNorth [168], Intel's Loihi [169], SpiNNaker [170], BrainScaleS [171], Tianjic chip [172], Neurogrid [173], etc. They support a broad class of problems ranging from complex to more general computations. At the device level, the most commonly used component is the memristor which can be utilized in synapse and neuron implementations [174–177]. Memristor crossbars are frequently used to represent synapses in neuromorphic systems [178, 179]. Memristor can also provide stochasticity in the neuron model [180]. Another emerging class of devices for neuromorphic computing is spintronics devices [162]. Spintronics devices can be implemented with low energy and high density and are compatible with existing CMOS technology [181]. The spintronics devices utilized in neuromorphic computing include spin-torque devices [182–184], magnetic domain walls [185–187], and skyrmions [188, 189]. Optical or photonics devices are also implemented for neurons and synapses in recent years [166, 190, 191]. The field is very new and many novel forms of neuron and synaptic devices can be designed to match the mathematical model of neural networks (NNs). Physical neuromorphic computing can implement these functionalities directly in their physical characteristics (I-I, V-V, I-V), which results in highly compact devices that are well-suited for scalable and energy-efficient neuromorphic systems [27, 30, 149, 192]. This is critical as current NN-based computing is highly centralized (resident-on and accessed-via cloud) and is energy inefficient because the underlying volatile, often von Neumann, digital Boolean-based system design unit has to emulate inherently analog, mostly non-volatile distributed computing model of neural systems, even if at a simple abstraction level [168]. Recent advances in custom design such as FPGAs [193] and more experimental Si FPNAs [194] have demonstrated that a new form of device design rather than emulation is the way to go, and physical neuromorphic computing based on emerging technology can go a long way to achieve this [195].

There is an increased use of noise-as-a-feature rather than a nuisance in NN models [196–198], and physical neuromorphic computing can provide natural stochasticity, with

various noise colors depending on the device physics [199, 200]. Some prominent areas where stochasticity and noise have been used include training generalizability [201], stochastic sampling [202], and recently proposed and coming into prominence, diffusion-based generative models [203]. In all these models, noise plays a fundamental role, i.e., these algorithms do not work without inherent noise.

It is therefore critical to study and analyze the kinds of devices that will be useful to implement physical neuromorphic computing. We understand from neurobiology that there is a large degree of neuron design customization that has developed through evolution to obtain high task-based performance. Similarly, a variety of mathematical models of neurons have been designed in NN literature as well [30, 152, 204]. It is quite likely that the area of physical neuromorphics will use a variety of device designs rather than the uniformity of NAND gate-based design commonly seen in Boolean-based design, to achieve the true benefits of energy efficiency and scalability brought forth by this paradigm of system design.

In this section, we study a subset of this wide variety of neuron designs that are well-represented and easily available from many proposed physical neuromorphic platforms to understand and analyze their task specialization. In particular, we analyze analog and binary neuron models, including stochasticity in the model, for analog temporal inferencing tasks, and evaluate and compare their performances. We numerically estimate the performance metric normalized means squared error (NMSE), discuss the effect of stochasticity on prediction accuracy vs. robustness, and show the hardware implementability of the models. Furthermore, we estimate the memory capacity for different neuron models. Our results suggest that analog stochastic neurons perform better for analog temporal inferencing tasks both in terms of prediction accuracy and hardware implementability. Additionally, analog neurons show larger memory capacity. Our findings may provide a potential path forward toward efficient neuromorphic computing.

3.2.1 Brief overview on neuron models

An essential function of a neuron in a NN is processing the weighted synaptic inputs and generating an output response. A single biological neuron itself is a complex dynamical system [205]. Proposed artificial neurons in most implementations of NNs (either software or hardware) are significantly simpler unless they specifically attempt to mimic the biological neuron [152,206,207]. As such their mathematical representations are cheaper and a significant amount of computational capabilities derive from the network itself. However, a NN is an interplay of the neurons, the synapses, and the network structure itself, and therefore the neuron model itself may provide certain capabilities that can help make a more efficient NN, in the context of the application specialization [208].

The set of behavior over which such neurons can be classified and analyzed is vast and may include spiking vs. non-spiking behavior with associated data representation, deterministic vs. stochastic output response function, discrete (or binary) vs. continuous (or analog) output response function, the particular mathematical model of the output response function itself (e.g., sigmoid, tanh, ReLU), presence or absence of memory states with a neuron, etc [209–211]. In the software NN world, specialization of certain neural models and connectivity are well appreciated, as an example sparse vs. dense vs. convolutional layers, or the use of ReLU neurons in the hidden layers vs. sigmoidal, softmax layers at outputs employed in many computer vision tasks [212–214]. Figure 3.9(a) schematically shows the output characteristics of different types of widely used neuron models.

In this work, we have focused on two particular behaviors of neural models that we believe can capture a significant application space, particularly in the domain of lightweight real-time signal processing tasks, and are readily built from emerging materials technology. We specifically look at binary vs. analog and deterministic vs. stochastic neuron output response functions (purple-colored bold font labels in Fig. 3.9(a)). We also use them in a reservoir computing (RC)-like context for signal processing tasks for our analysis. Reservoir computing uses the dynamics of a recurrently connected network of neurons to project an

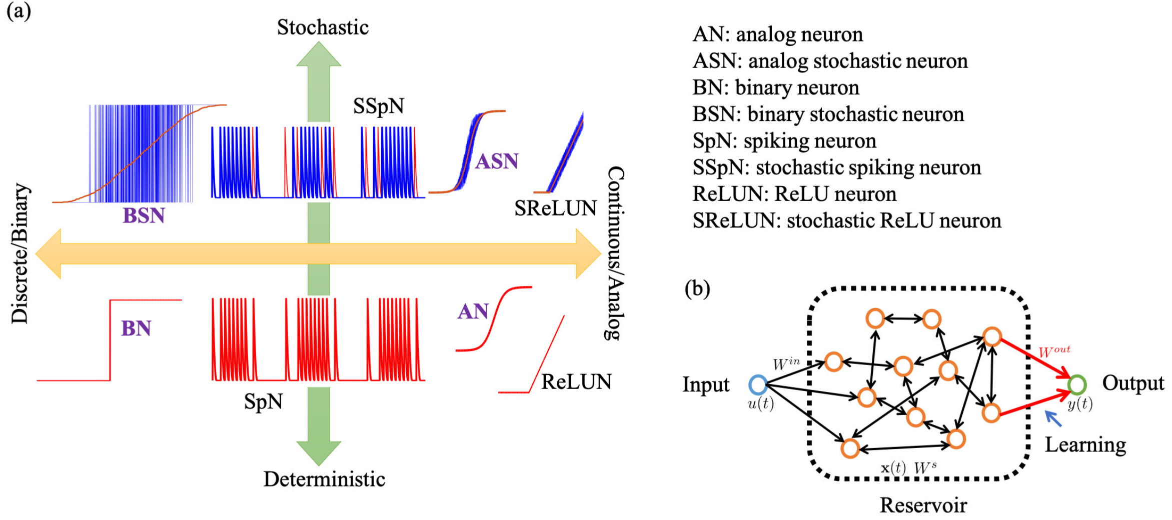


Figure 3.9: (a) Schematic of different types of widely used neuron models with their output characteristics. In the bottom panel, all the red curves represent the deterministic neurons' output characteristics. In the top panel, the blue curves represent the actual stochastic output characteristics while the red is the corresponding deterministic/expected value of the output ($<$ *stochastic output* $>$) characteristics. Spiking neurons (SpN and SSpN) can be considered in between the two limits of purely binary vs. purely analog neurons. Please note that we only analyze the analog and binary neurons (including their stochastic counterparts) in this work, as indicated by the purple-colored bold font labels. (b) Schematic of a reservoir setup using neurons connected with each other bidirectionally with random weights.

input (spatio-)temporal signal onto a high dimensional phase space, which forms the basis of inference, typically via a shallow 1-layer linear transform or a multi-layer feedforward network [161, 215–218]. A schematic of a reservoir is shown in Fig. 3.9(b) where the neurons are connected with each other bidirectionally with random weights. Multiple reservoirs may be connected hierarchically for more complex deep RC architecture. RC may be considered as an ML analog of an extended Kalman filter where the state space and the observation models are learned and not designed a priori [215].

Our choice of evaluating these specific behavior differences on an RC-based NN reflects the prominent use-case that is made out for many emerging nano-materials technology-based neuron and synaptic devices, viz. energy-efficient learning, and inference at the edge. These tasks often end up involving temporal or spatio-temporal data processing to extract relevant and actionable information, some examples being anomaly detection [219], feature

tracking [220], optimal control [221], and event prediction [222], all of which are well-suited for an RC-based NN. Therefore this testbench forms a great intersection for our analysis.

It should be noted that we do not include spiking neurons in this particular analysis. Spiking neurons have significantly different data encoding (level vs. rate or inter-spike interval encoding) and learning mechanisms (back-propagation or regression vs. spike-time dependent plasticity) that it is hard to disentangle the neuron model itself from demonstrated tasks, therefore we leave such a contrasting analysis of spiking neuron devices with non-spiking variants for a future study.

The neurons are modeled in the following way:

$$\mathbf{y} = f_N(\sum w^T\mathbf{x}) + r_N \tag{3.4}$$

Here, the symbols have the usual meaning, i.e., $\mathbf{y}$ is the output activation of the neuron, f_N is the activation function, which is a sigmoidal or hyperbolic tangent for most non-spiking hardware neurons, and r_N is a random sample drawn from a random uniform distribution to represent stochasticity. It is possible to use a ReLU-like activation function or some other distribution for sampling stochasticity, particularly if the hardware neuron shows colored noise behavior, we do not particularize for such details and keep the analysis confined to the most common hardware neuron variants. Therefore, in our analysis, the r_N term is weighed down by an arbitrary factor to mimic the degree of stochasticity displayed by the neuron, and the f_N is either a continuous $tanh()$ for analog neuron or a $sgn(tanh())$ for a binary neuron ($sgn()$ being the signum function).

3.2.2 Computational details

As discussed previously, the neuron models are analyzed in the context of a reservoir computer, specifically an echo-state network (ESN). An ESN is composed of a collection of recurrently connected neurons, with randomly distributed weights of the interconnects within this

collection [223, 224]. This forms the "reservoir", which is activated by an incoming signal, and whose output is read by an output layer trained via linear regression.

We employ different neuron models in this work, such as analog and binary neurons (with and without stochasticity in the model), which makes a total of four models at our disposal, namely, analog neuron (AN), analog stochastic neuron (ASN), binary neuron (BN), and binary stochastic neuron (BSN). The dynamical equations of the reservoirs built using different neuron models are described as follows [30]:

$$\begin{aligned}
&\text{AN} : \mathbf{x}[t+1] = (1-a) * \mathbf{x}[t] + a * tanh(\mathbf{z}[t+1]) \\
&\text{ASN} : \mathbf{x}[t+1] = (1-a) * \mathbf{x}[t] + a * tanh(\mathbf{z}[t+1]) + b * r_N[t] \\
&\text{BN} : \mathbf{x}[t+1] = (1-a) * \mathbf{x}[t] + sgn(a * tanh(\mathbf{z}[t+1])) \\
&\text{BSN} : \mathbf{x}[t+1] = (1-a) * \mathbf{x}[t] + sgn(a * tanh(\mathbf{z}[t+1]) + b * r_N[t])
\end{aligned} \tag{3.5}$$

where $\mathbf{z}[t+1] = W^{in}\mathbf{u}[t+1] + W^s\mathbf{x}[t]$. Here, $\mathbf{u}$ is the input vector, $\mathbf{x}[t]$ represents the reservoir state vector at the time t, a is the reservoir leaking rate (assumed to be the constant for all the neurons), b is the neuron noise scaling parameter to include stochasticity in the neuron model, r_N is a uniform random distribution, and W^{in} and W^s are the random weight matrices of input-reservoir and reservoir-reservoir connections, respectively. We use the same leaking rate across all models to ensure a fair comparison among the neuron models on an equal footing. It can be challenging to compare models that have different parameters as it can introduce biases. One of the unique features of reservoir computing is having random weight matrices [215] and we consider five different network topologies by creating five sets of W^s using random 'seed' for various reservoir sizes, which makes our analysis unbiased to any particular network topology. The W^s elements are normalized using the spectral radius. We perform 1000 simulations within each network topology making the total sample size 5000 for every reservoir size within each neuron model. The output vector $\mathbf{y}$ is obtained as:

$$\mathbf{y} = W^{out}\mathbf{x} \tag{3.6}$$

where W^{out} represents the reservoir-output weight matrix. We consider two different types of training methods, i.e, 'offline' and 'online' training. In the case of 'offline' training, we extract the output weight matrix, W^{out} once at the end of the training cycle and use that static W^{out} for the testing cycle. In contrast, for 'online' training, W^{out} is periodically updated throughout the testing cycle. The entire testing cycle is divided into 40 segments. The first segment uses the W^{out} extracted from the initial training cycle. We calculate a new W^{out} after the first segment of the testing cycle. Then, we update the W^{out} such that the elements are composed of 90% from the older version and 10% from the new one. The updated W^{out} is used for the second segment and the procedure keeps going on throughout the testing cycle. This stabilizes the learning at the cost of higher error rates as the learning evolution slowly evolves to a new configuration. This is akin to the successive over-relaxation methods used in many self-consistent numerical algorithms for improved convergence.

3.2.3 Binary vs. Analog: inference errors

We implement the temporal inferencing task, specifically, the time-series prediction task to test and compare the performance of different neuron models. We consider an input signal of the form $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t)$, which we referred to as a clean input. We use $A = 1$, $B = 2$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. Although we choose the magnitude and frequency of the input arbitrarily, we further investigate other combinations of these variables (Table 3.1) to ensure that our analysis remains independent of them. We train the neuron models using the clean input signal and test the models on a test signal from the same generator. The neuron models learn to reproduce the test signal from its previously self-generated output. The performance of the neuron models for time-series prediction tasks is usually measured by the NMSE, which is the metric that indicates how accurately the

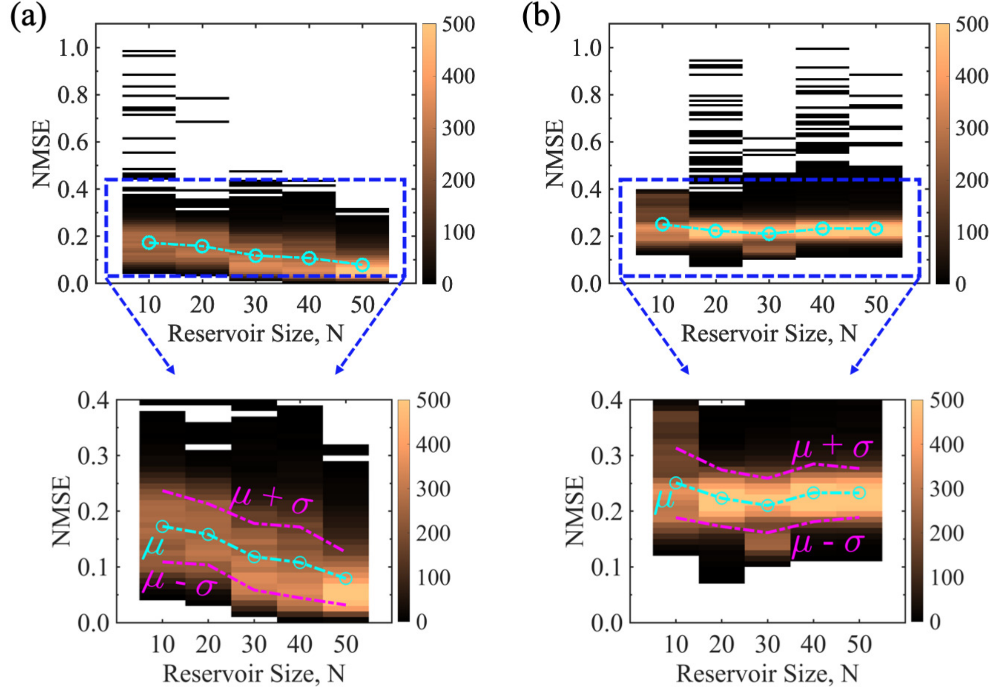


Figure 3.10: Comparison of NMSE for an analog time-series prediction task between (a) ASN and (b) BSN models as a function of reservoir size with 5% stochasticity incorporated in both the neuron models for a clean input signal. The form of the clean input signal is $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t)$, where $A = 1$, $B = 2$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. ASN performs better than BSN for the entire range of reservoir size as indicated by the average (μ) NMSE (cyan dashed-dotted line). ASN shows a decreasing trend in NMSE as a function of reservoir size while BSN results remain almost unchanged. The NMSE data for every reservoir size is obtained from five different reservoir topologies and 1000 simulation runs (different random 'seed') within each topology (total sample size is 5000). The color bar represents the frequency of the NMSE data. Note that in some cases, our model fails to generate a meaningful NMSE as the reservoir output blows up. We get meaningful output from $\sim 90\% - 100\%$ cases depending on the reservoir sizes, and those data are plotted here and used to estimate the average NMSE. The bottom panel is the zoomed version of the top panel and the magenta dashed-dotted lines are the guide to the eye that shows the data distribution in the range of $\mu \pm \sigma$. The color codes to represent the μ and σ are the same for the subsequent figures henceforth.

models can predict the test signal. If y_{tar} is the target output and y_{pre} is the actual predicted output, for N_T time steps, we define NMSE as:

$$NMSE = \frac{1}{N_T(y_{tar}^{max} - y_{tar}^{min})} \sum_{i=1}^{i=N_T} (y_{tar}(i) - y_{pre}(i))^2 \tag{3.7}$$

Figure 3.10(a) and (b) show the NMSE for ASN and BSN, respectively for the time-series prediction task for various reservoir sizes. We generate the results using the 'offline' training as discussed in the method section, for a clean input signal. We incorporate the stochasticity by adding 5% white noise in both neuron models ($b = 0.05$). The total sample size is 5000 for a specific reservoir size, however, it is worth mentioning that we do not get valid NMSE for all the 5000 cases because the network fails to predict the input signal and blows up for some cases. We get $\sim 90\% - 100\%$ successful cases depending on the reservoir sizes. Only valid data points are included in Fig. 3.10 and all the subsequent figures. We find ASN performs better than BSN for all the reservoir sizes indicated by the average NMSE (cyan dashed-dotted line). Overall the NMSE is less scattered for ASN than BSN, so is their standard deviation, (magenta dashed-dotted line) as shown in the bottom panel of Fig. 3.10. For ASN, we find that the average NMSE has a decreasing trend as the reservoir size increases, which indicates larger size networks can predict better. This happens because of the substantially richer dynamics and phase-space volume possible in a large network. In contrast, for BSN, the average NMSE is almost unchanged as the reservoir size increases.

We vary the stochasticity incorporated in the neuron models. Figure 3.11(a) and (b) show the distribution of the NMSE for different percentages of stochasticity, b for ASN and BSN models, respectively. We find that ASN performs better than its BSN counterpart throughout the ranges of b as indicated by the average NMSE. For ASN, the average NMSE shows a sub-linear trend as a function of b (Fig. 3.11(c)) for various reservoir sizes, while for BSN, the average NMSE remains unchanged (Fig. 3.11(d)). For pure analog neuron ($b = 0\%$), the NMSE is not much spread out, and also, for larger reservoir size, the average NMSE is smaller than the neuron model with stochasticity, however, having a neuron model with zero stochasticity is not practical. Moreover, stochasticity helps to make the system stable and reliable as discussed in the next sub-section. Although the average NMSE increases with increasing b, we conjecture that $b = 2 - 5\%$ would be optimal.

The aforementioned results are based on a clean input signal. We tested the models for

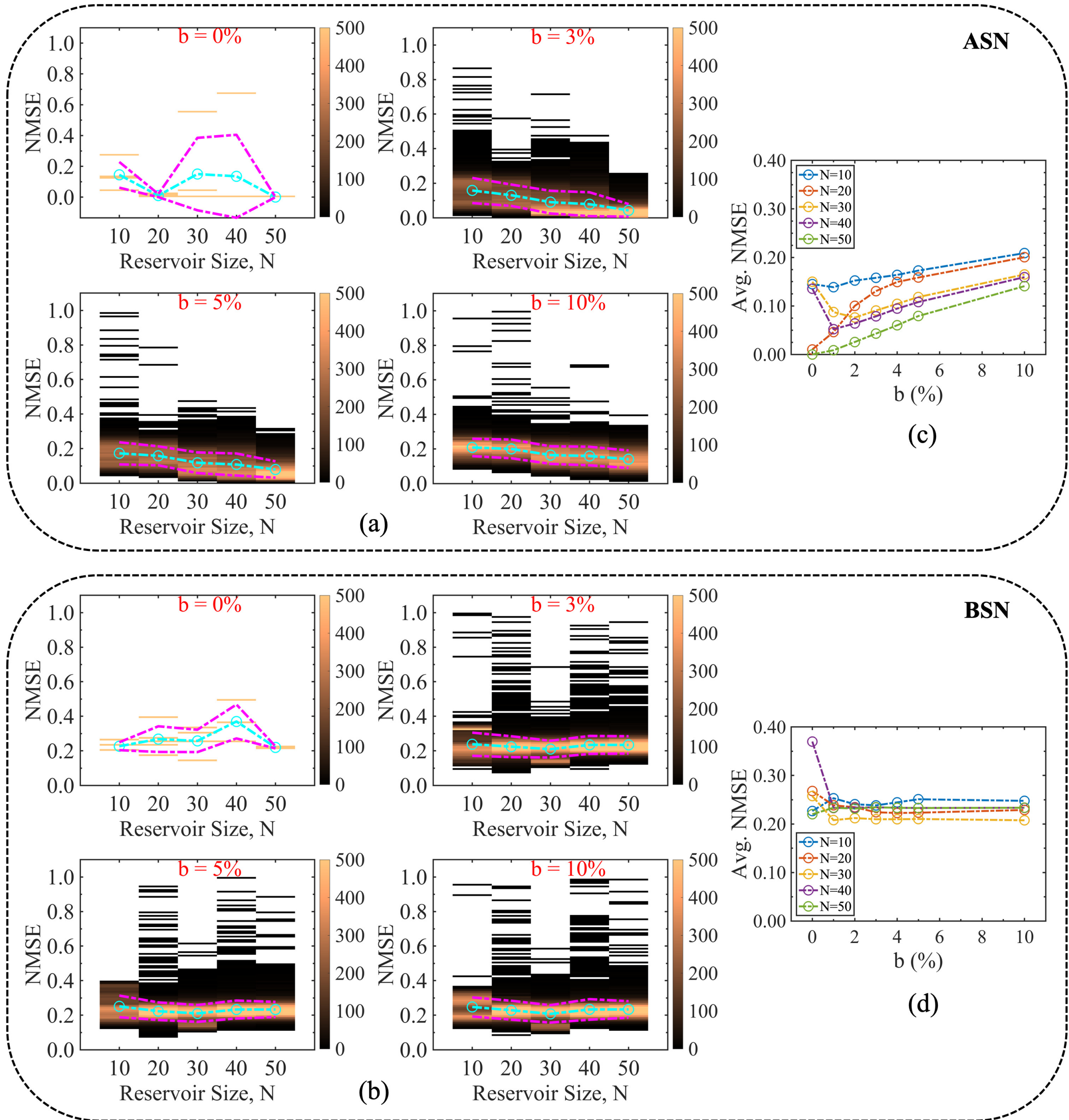


Figure 3.11: Evolution of NMSE for different degrees of stochasticity (noise percentages) associated with the (a) ASN and (b) BSN models. ASN performs better than the BSN model for analog time-series prediction tasks throughout the ranges of the degree of stochasticity as indicated by the average NMSE shown in (c) and (d) for ASN and BSN, respectively. The characteristics of the average NMSE as a function of reservoir size i.e., the decreasing trend for ASN while almost no change for BSN holds throughout the range of b.

distorted input as well. For the distorted case, we add a white noise in the clean input and the form of the distorted input signal is $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t) + C[rand(1,t) - 0.5]$.

The white noise is uniformly distributed for all t values, both in the positive and negative half of the sinusoidal input. The degree of noise has been chosen arbitrarily. Again, we show various degrees of noise (Table 3.1) to make the analysis independent of a specific value of the noise margin. The NMSE results shown in Fig. 3.12(a) and (b) are calculated using $A = 1$, $B = 2$, $C = 1$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. We find a better performance for ASN than that of BSN for the distorted input as well. It appears that for ASN, with a distorted input signal, the spectrum of NMSE is smaller, which reduces the standard deviation. The characteristics of the average NMSE are similar for the clean and distorted input for both ASN (Fig. 3.12(c)) and BSN (Fig. 3.12(d)) models. However, the average NMSE is slightly lower for the distorted input for both types of neuron models. Furthermore, we use different combinations of signal magnitude, frequency, and the weight of noise in the input signal. We list the average NMSE for various reservoir sizes in Table 3.1. Additionally, we explore other input functions beyond the simple sinusoidal input used in the aforementioned results. In particular, we use a sinusoidal with higher harmonic terms, a sawtooth input function, and a square input function. The used form of the functions are $u(t) = \frac{4}{\pi}\sum_{i=1}^{15}\frac{1}{n}\sin 2\pi n f_1 t$ (odd n), $u(t) = A\ sawtooth(2\pi f_1 t) + B\ sawtooth(2\pi f_2 t)$, $u(t) = A\ square(2\pi f_1 t) + B\ square(2\pi f_2 t)$, respectively. In the case of sinusoidal with higher harmonic terms, we use the fundamental frequency $f_1 = 0.1\ Hz$. For the sawtooth and square inputs, the magnitude and frequency remain the same as of the original sinusoidal clean input. The results are summarized in Fig. 3.13, where the label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions, respectively. Fig. 3.13 shows that for all the different inputs, ANS performance is better than BSN in terms of NMSE. Comparing all the cases, we conjecture that ASN performs better than BSN for the temporal inferencing task.

3.2.4 Deterministic vs. Stochastic: generalizability and robustness

One important aspect of any NN implementation is the generalizability and robustness of the learning. A model trained to a very specific data distribution will fail when it is running on a distribution that differs from the trained model. This is particularly true if a generative model guides its own subsequent learning, which is the example we have used in our online learning scenario. In this case, the underlying distribution is varied slowly while the network evolves its internal generative model to match the output of distribution, i.e., it works as a dynamically evolving temporal auto-encoder.

The stochasticity of the neuron response will add errors to the generated output as we see in the previous cases, however, we find that after a few iterations of the online learning cycle, the ability of this online learning blows up, i.e., the linear regression-based learning cannot keep up with the test distribution evolution and the error builds up (we call it blowup) and the whole training needs to be fully reset or reinitiated and cannot merely evolve from previous learning. This blowup occurs 100% for deterministic analog neurons, and the rate reduces as the degree of stochasticity increases (parameter b).

This is shown in Table 3.2 for various input functions. It should be noted that at very high stochasticity while the training is more robust, the errors will be high, therefore a minimal amount of stochasticity is useful as a trade-off between these ends. The degree to which the trade-off can be performed depends on the application scenario. If full retraining is too expensive or not acceptable, then a relatively higher degree of stochasticity in the neuron is necessary, but if it is cheap and acceptable to retrain the whole network frequently, a near-deterministic neuron will be better suited to meet the requirements

3.2.5 Synaptic weights dynamic range: hardware implementability

One critical aspect of hardware implementability of neuromorphic computing is the ability to modulate the weights and the dynamic range or the order of magnitude to which weights may be distributed. It can be shown that a 30-bit weight resolution represents about a 100 dB

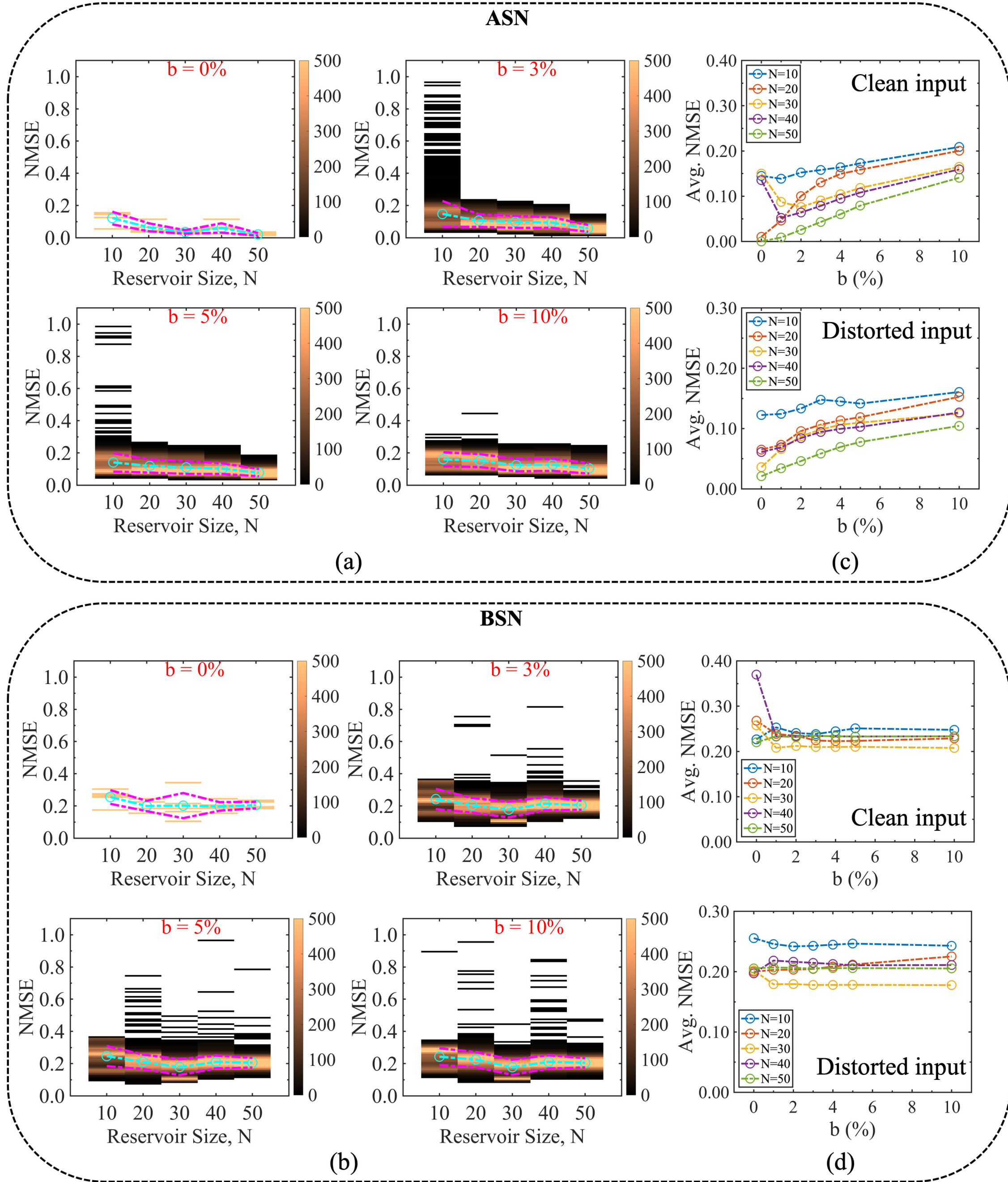


Figure 3.12: Evolution of NMSE for different degrees of stochasticity for (a) ASN and (b) BSN models for a distorted input signal. Random white noise is added to the clean input signal to introduce distortion and the form of the distorted signal is $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t) + C[rand(1,t) - 0.5]$, where $A = 1$, $B = 2$, $C = 1$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. ASN performs better than BSN for the distorted input, as indicated by the average NMSE shown in (c) and (d) for ASN and BSN, respectively, which dictates the robustness of the ASN model in terms of performance irrespective of the input signals.

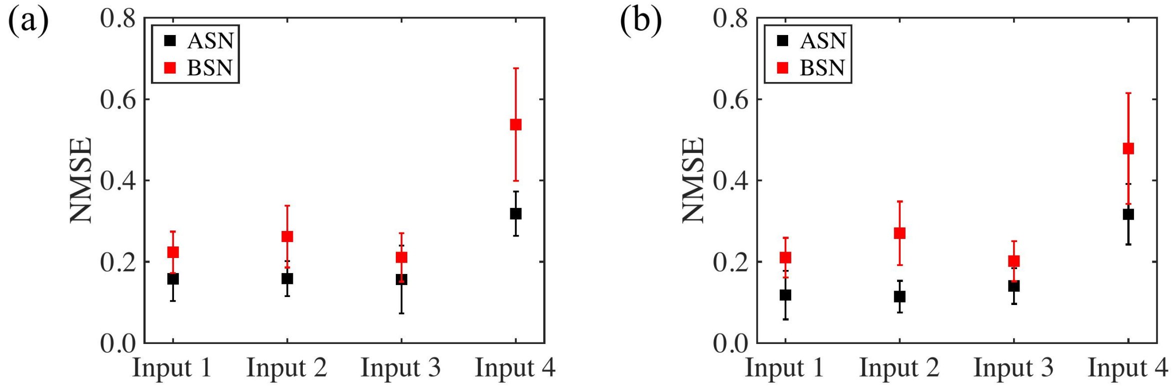


Figure 3.13: Comparison of NMSE for time-series prediction task between ASN and BSN models for various input functions for a reservoir size of (a) N = 20 and (b) N =30. The degree of stochasticity incorporated in both neuron models is 5%. The label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions, respectively. ANS performance is better than BSN in terms of NMSE for different input functions.

dynamic range. While such ranges might be comparatively easily implemented in software, it is significantly difficult to implement such a high dynamic range in physical hardware. While some memristive materials may show multi-steps, it is hard to achieve much more than one order of magnitude change in the weights. Please note that we do not mean the change in the physical characteristics (typically the resistance) used to represent the weights themselves, but rather the number of steps that the weight can be implemented as.

We compare the dynamic range of the learned synaptic weights that need to be implemented in the reservoir networks (in the trained output readout layer) for various input functions and find that the ASN networks show the smallest dynamic range for all the cases (Fig. 3.14) and suggest the easiest path to hardware implementability of physical neuromorphic computing. It is important to note that the hardware implementation of neuromorphic computing is an open question and the dynamic range of the synaptic weights is one of the important factors when it comes to the physical deployment of neuromorphic computing as discussed above. ASN networks show better performance in terms of the dynamic range of learned synaptic weights compared to other models, which suggests that networks that employed ASN models might have better hardware implementability; however, it requires more analysis in terms of

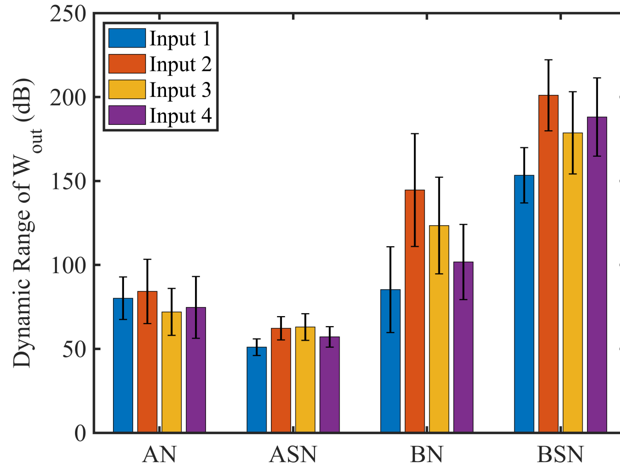


Figure 3.14: Dynamic range of the learned synaptic weights, W_{out} for all the neuron models ($N = 20$). 5% stochasticity is considered in the ASN and BSN models. ASN model shows the smallest dynamic range that leads to better hardware implementability. The label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions, respectively.

energy cost, scalability, and reconfigurability, which we leave as a future study.

3.2.6 Memory capacity

The performance of reservoir computing is often described by memory capacity (MC) [225–227]. It measures how much information from previous input is present in the current output state of the reservoir. The task is to reproduce the delayed version of the input signal. For a certain time delay k, we measure how well the current state of the reservoir $y_k(t)$ can recall the input u at time $t - k$. The linear MC is defined as:

$$MC = \sum_k \frac{cov^2(u(t-k), y_k(t))}{\sigma^2(u(t-k)\sigma^2(y_k(t))} \tag{3.8}$$

where $u(t - k)$ is the delayed version of the input signal, which is the target output, and $y_k(t)$ is the output of the reservoir unit trained on the delay k. cov and σ^2 denote *covariance* and *variance*, respectively.

Table 3.3 shows the linear MC for different neuron models for the distorted input $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t) + C[rand(1,t) - 0.5]$, where $A = 1$, $B = 2$, $C = 1$, $f_1 = 0.10\ Hz$, and $f_2 = 0.02\ Hz$. We consider the delayed signal over 1 to 50 timesteps, meaning k spans from 1 to 50. We find that analog neurons have significantly larger linear MC than binary neurons. For analog neurons, linear MC increases as the reservoir size increases, which is expected because a larger dynamical system can retain more information from the past [225]. Additionally, including stochasticity in the analog neuron model degrades the linear MC as reported previously [225]. In contrast, binary neurons fail to produce substantial differences in linear MC when reservoir size is varied and stochasticity is included in the model.

Besides the previously mentioned properties, physical neuromorphic computing exhibits chaos or edge-of-chaos property, which has been shown to enhance the performance of complex learning tasks [228–230]. The edge-of-chaos property refers to the transition point between ordered and chaotic behavior in a system. In the discussed models, it may be possible to achieve the edge-of-chaos state by introducing increasing amounts of noise to the models, resulting in chaotic behavior that could potentially improve network performance. We find that with an increased degree of stochasticity in the neuron models, the learning process becomes more robust, which could be a signature of the performance improvement by including the edge-of-chaos property. However, the prediction accuracy and the linear MC tend to decrease with a higher degree of stochasticity, so the trade-off needs to be considered. It should be noted that a more comprehensive analysis is required to fully understand the impact of edge-of-chaos behavior on the discussed neuron models, which is beyond the scope of this paper and will be explored in future studies.

3.2.7 Conclusion

In summary, we studied different neuron models for the analog signal inferencing (time-series prediction) task in the context of reservoir computing and evaluate their performances for various input functions. We show that the performance metrics are better for ASN than BSN

Table 3.1: Average NMSE data extracted from the ASN and BSN models ($b = 5\%$) for various reservoir sizes. The form of the input signal is, $u(t) = A\cos(2\pi f_1 t) + B\sin(2\pi f_2 t) + C[rand(1,t) - 0.5]$.

Model	Reservoir Size	Avg. NMSE for different input signals		
		$\{A,\ B,\ C\}$ $= \{0.5,\ 1.0,\ 0.0\}$ $\{f_1,\ f_2\}$ $= \{0.20,\ 0.04\}\ Hz$	$\{A,\ B,\ C\}$ $= \{1.0,\ 2.0,\ 0.5\}$ $\{f_1,\ f_2\}$ $= \{0.10,\ 0.02\}\ Hz$	$\{A,\ B,\ C\}$ $= \{1.0,\ 2.0,\ 1.5\}$ $\{f_1,\ f_2\}$ $= \{0.10,\ 0.02\}\ Hz$
ASN	$N = 10$	0.1729	0.1453	0.1501
	$N = 20$	0.1585	0.1199	0.1161
	$N = 30$	0.1183	0.0960	0.0984
	$N = 40$	0.1080	0.0775	0.1001
	$N = 50$	0.0791	0.0605	0.0816
BSN	$N = 10$	0.2510	0.2396	0.2546
	$N = 20$	0.2233	0.2102	0.2184
	$N = 30$	0.2103	0.1895	0.2028
	$N = 40$	0.2331	0.2156	0.2040
	$N = 50$	0.2329	0.2142	0.2173

for both clean and distorted input signals. We find that the increasing degree of stochasticity makes the models more robust, however, decreases the prediction accuracy. This introduces a trade-off between accuracy and robustness depending on the application requirements and specifications. Furthermore, the ASN model turns out to be the suitable one for hardware implementation, which attributes to the smallest dynamics range of the learned synaptic weights, although other aspects, i.e., energy requirement, scalability, and reconfigurability need to be assessed. Additionally, we estimate the linear memory capacity for different neuron models, which suggests that analog neurons have a higher ability to reconstruct the past input signal from the present reservoir state. These findings may provide critical insights for choosing suitable neuron models for real-time signal-processing tasks and pave the way

Table 3.2: Robustness vs. accuracy trade-off ($N = 20$). The label Input 1, Input 2, Input 3, and Input 4 correspond to the sinusoidal clean input, sinusoidal with higher harmonic terms, sawtooth, and square input functions described earlier, respectively.

Model	b (%)	Blowup (%)				Avg. NMSE			
		Input 1	Input 2	Input 3	Input 4	Input 1	Input 2	Input 3	Input 4
AN	0	100	100	100	100	–	–	–	–
ASN	1	74.7	81.3	98.5	98.6	0.3175	0.2759	0.4947	0.5475
	2	66.4	79.3	92.0	92.9	0.2921	0.3225	0.3947	0.5537
	3	60.7	78.7	85.9	88.9	0.2854	0.3301	0.3744	0.5591
	4	56.2	77.0	81.0	84.3	0.2782	0.3534	0.3572	0.5515
	5	53.9	76.3	76.4	80.7	0.2778	0.3597	0.3636	0.5358
	10	49.1	71.6	66.5	71.4	0.2849	0.3903	0.3398	0.5316
	15	48.8	69.3	59.7	67.3	0.3019	0.4266	0.3557	0.5412

Table 3.3: Linear memory capacity (MC) for different neuron models.

Model	Reservoir Size	MC	
		$b = 0\%$	$b = 5\%$
Analog	$N = 40$	39.0	32.5
	$N = 50$	45.2	36.2
Binary	$N = 40$	2.7	2.8
	$N = 50$	3.4	3.2

toward building energy-efficient neuromorphic computing platforms.

3.3 Modeling short pulse-actuated medium barrier magnet-based robust and energy-efficient TRNG units

TRNGs are employed in many applications, including cryptography [231], Monte Carlo simulations [232], neuromorphic computing [137, 233], and probabilistic [27] and stochastic computing [234]. Conventional software algorithm-based and CMOS-based random number generators, *e.g.*, LFSR do not serve the purpose of TRNGs because they produce pseudorandom bitstreams that are correlated and can be predetermined if the initial seed is known [235–237]. In contrast, TRNGs utilize physical phenomena that are inherently random, such as thermal noise, and radioactive decay [238–241]. Existing CMOS-based implementations of TRNGs use thermal jitter for generating true random numbers; however, they have large footprints and are energy-hungry [242–244].

Spintronic TRNGs provide an opportunity in this regard [245–247]. MTJs constitute a fundamental building block for spintronic devices and manifest CMOS compatibility [20, 248]. MTJs consist of two ferromagnetic layers — a "pinned layer" whose magnetization is fixed and a "free layer" whose magnetization can be reoriented by a spin current — separated by an insulator. The alignment between the magnetization of the pinned layer and the free layer creates parallel (P) and anti-parallel (AP) states. The MTJ free layer exhibits a double potential well with two low energy states along the easy axis, separated by an energy barrier E_b. The free layer magnetization can be switched from P to AP and vice versa by applying a current/voltage pulse utilizing STT [249–251]. Such STT-driven MTJs show a prominent stochastic switching behavior in the presence of a thermal field [199, 252] that will form the basis of our analysis.

In the past, MTJs consisting of HBMs ($E_b > 40\ k_BT$) were frequently advocated as TRNGs; however, they suffer from high energy costs and low throughput [245, 253, 254]. At the opposite end of the spectrum, superparamagnetic tunnel junctions employing LBMs ($E_b \sim k_BT$) have also been advocated as probabilistic bits. These LBMs allow the magnetization states to

randomly fluctuate between P and AP, under the influence of the thermal field [26, 255, 256]. Although the process is very energy-efficient, LBMs suffer from slow dynamics and are rather sensitive to process and temperature variations that degrade the quality of the random bitstreams (meaning they are not "fair coins") [26, 31]. The rate of fluctuations in passively fluctuating magnets is very sensitive to temperature [26, 257]. Accounting for these variations would require more circuit overhead, *e.g.*, more MTJ devices, and XOR operations to pass the NIST test for TRNG [26, 31]. Besides, they require near-perfect circular cross-sections and are thus hard to build in practice [136, 143, 258]. Stochastic magnetic actuated random transducer (SMART) devices based on perpendicular MTJs with MBMs ($E_b \sim 20 - 40\ k_BT$) seem like a good compromise between these two extremes, for building energy-efficient and robust TRNGs [31, 259]. However, a systematic analysis of their energy-delay-reliability-variability trade-off has not yet been undertaken.

In this section, we present a comprehensive analysis of STT-driven SMART TRNGs. We numerically solve the Fokker–Planck (FP) equation to calculate the 50% switching probability across a wide range of pulse durations (Fig. 3.15) and analyze the effect of different kinds of variations on this probability. Specifically, our study investigates the influence of pulse amplitude and duration (Fig. 3.16), temperature (Fig. 3.17), and geometric and material parameters (Fig. 3.18) on the 50% switching probability. We find that SMART devices exhibit relatively low sensitivity to the process-voltage-temperature (PVT) variations but greater sensitivity to pulse duration variation, especially when operated under a short-pulse limit. We estimate the energy dissipated during stochastic switching (Fig. 3.19) and find that short pulse-activated switching consumes less energy than the same device operated with longer pulses, which suggests that SMART devices operating in the short-pulse limit can achieve robustness and energy efficiency concurrently. Our results provide a potential pathway toward the realization of fast, energy-efficient, and robust TRNG operations.

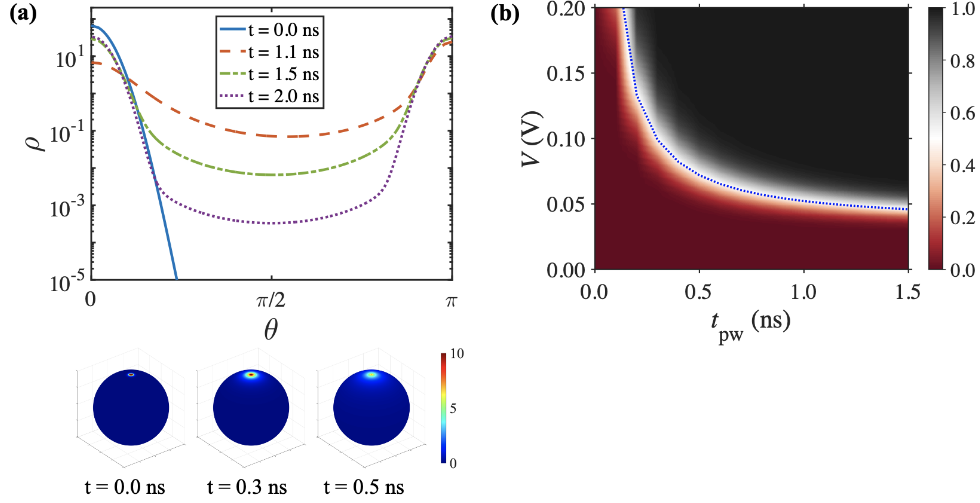


Figure 3.15: (a) Probability density for an MBM ($\Delta \sim 35$) activated by a 1 ns pulse (top panel). The pulse is turned ON at $t = 0$ ns. A bimodal distribution emerges after the pulsing, which leads to a 50% switching probability. The bottom panel shows the time evolution of the probability density (colormap) during the pulsing. (b) Probability of switching (colormap) as a function of pulse amplitude and duration. The dotted blue overlaid curve represents a 50% switching probability.

3.3.1 Computational details

The magnetization dynamics is commonly described by the stochastic LLG equation, given by:

$$\begin{aligned}\frac{1+\alpha^2}{\gamma}\cdot\frac{\partial \mathbf{m}}{\partial t} &= -\mu_0\cdot(\mathbf{m}\times\mathbf{H}_{\text{eff}}) - \alpha\mu_0\cdot\mathbf{m}\times(\mathbf{m}\times\mathbf{H}_{\text{eff}}) \\ &\quad -\frac{\hbar}{2e}\cdot\frac{\eta I}{M_s\Omega}\cdot\mathbf{m}\times(\mathbf{m}\times\mathbf{m}_p),\end{aligned} \tag{3.9}$$

where $\mathbf{m} = \mathbf{M}/M_s$ is the normalized magnetization and M_s is the saturation magnetization. I is the applied charge current and $\mathbf{m}_p$ is the unit vector along the spin polarization direction. α, μ_0, γ, η, $\hbar$, e, and Ω are the magnetic damping coefficient, permeability of free space, gyromagnetic ratio, spin polarization efficiency factor, reduced Plank constant, elementary charge, and MTJ free layer volume, respectively. Generally, without external field, the effective field $\mathbf{H}_{\text{eff}}$ includes anisotropy, demagnetization, dipolar effects, Dzyaloshinskii–Moriya interaction, exchange coupling, and thermal effects. Nonetheless, a common practice for thin

perpendicular anisotropy magnets is to include the effective anisotropy field $\mathbf{H}_k$ (including uniaxial anisotropy and demagnetization), and the thermal field $\mathbf{H}_{\text{th}}$ ($\mathbf{H}_{\text{eff}} = \mathbf{H}_k + \mathbf{H}_{\text{th}}$) [249, 259, 260], which we also adopt in our study. We consider a macrospin model in which the spins are assumed to be strongly exchange coupled. The $\mathbf{H}_{\text{th}}$ provides a random stochastic field, which can be incorporated in a Monte Carlo solution of the Eq. (3.9).

Alternatively and more efficiently, by solving a FP equation, we can quantify the statistical nature of magnetization switching under thermal fluctuations [252, 261–263]. We numerically solve the 1-D differential equation form of the general FP equation:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\mathbf{L}\rho) + D_i \nabla^2 \rho, \tag{3.10}$$

where $\rho(\theta; t)$ is the probability density of the magnetization at time t, θ being the magnetization angle to the easy axis (z-axis). $\mathbf{L}$ is the sum of all the effective torques and D_i is the effective diffusive constant that accounts for the thermal fluctuations and is defined as:

$$D_i = \frac{\alpha \gamma k_B T}{(1+\alpha^2)\mu_0 M_s \Omega}. \tag{3.11}$$

The details of the numerical methods can be found in Ref. [263].

The probability of switching can be estimated from the probability density as follows:

$$P_{\text{sw}} = \int_{\pi/2}^{\pi} \rho(\theta; t) d\theta = 1 - \int_{0}^{\pi/2} \rho(\theta; t) d\theta. \tag{3.12}$$

3.3.2 Short-pulse actuated STT switching for $P_{\text{sw}} = 0.5$

STT-driven magnetization switching under applied current or voltage pulses can generally be categorized into two limits — ballistic and diffusive. Ballistic switching refers to the magnetization dynamics under short-duration pulses, while longer-duration pulses dictate the diffusive limit. In the ballistic limit, the short pulse transfers spin-angular momentum to the free layer, and there is little effect of thermal fluctuation during the pulsing. The switching

probability in the ballistic limit for a macrospin model can be expressed as [31,264]:

$$P_{\text{sw}}^{\text{ballistic}} = \exp\left[-\frac{\pi^2\Delta}{4}\exp\left\{-\left(\frac{V}{V_{c0}}-1\right)\frac{2t_{\text{pw}}}{\tau_D}\right\}\right], \quad (3.13)$$

where thermal stability factor, $\Delta = E_b/k_BT = \mu_0 H_k M_s \Omega/2k_BT$ and critical voltage for switching, $V_{c0} = 2\alpha e\mu_0 H_k M_s \Omega R_P/\eta\hbar$, R_P is the MTJ junction resistance in the P state. $\tau_D = (1+\alpha^2)/\alpha\gamma\mu_0 H_k$ is the intrinsic time scale for the dynamics. $H_k = 2K_u/\mu_0 M_s - M_s$ is the effective anisotropy field, where K_u is uniaxial anisotropy constant. V and t_{pw} are the applied pulse amplitude and duration, respectively. Note that Eq. (3.13) is valid for high barrier ($\Delta \gg 1$) and strong drive voltage ($V \gg V_{c0}$) and its residual error is estimated of the order $\exp\{-\Delta\}$ [265–267]. However, the numerically solved FP works both in the supercritical ($V \gg V_{c0}$) and sub-critical ($V \ll V_{c0}$) regimes [261,263]. As the Δ and V/V_{c0} become smaller, the difference between the numerical FP and Eq. (3.13) increases because the latter assumes no thermal fluctuations during the pulse. In our numerical FP simulations, we use an MBM having $\Delta \sim 35$ unless otherwise specified. The parameters used in the simulations are listed in Table 3.4.

Figure 3.15(a) shows the probability density of an MBM ($\Delta \sim 35$) activated by a 1 ns pulse. The initial quasi-equilibrium Boltzmann distribution confined near the $\theta = 0$ well (θ is the angle between the magnetization and the $z-$axis) as shown by the blue curve in the top panel of Fig. 3.15(a). We turn ON the pulse at $t = 0$ ns. Immediately after pulsing at $t = 1.1$ ns, STT from the short duration pulse drives the probability density to the $\theta = \pi$ well, creating a bimodal distribution. If we relax the system for some time, we can clearly see the bimodal distribution leading to a 50% switching probability. It takes ~ 1 ns to reach the quasi-equilibrium distribution after turning OFF the pulse ($t = 2$ ns curve). The bottom panel of Fig. 3.15(a) shows the evolution of the probability density during the pulse duration. At the beginning of pulsing, the probability density is confined to the north pole of the unit Bloch sphere ($\theta = 0$) and starts spreading with time towards the south pole ($\theta = \pi$).

Figure 3.15(b) shows a colorplot of the probability of switching P_{sw} as a function of pulse amplitude V, and pulse duration t_{pw}. As expected, the pulse amplitude required for the magnetization switching is inversely proportional to the pulse duration [251]. The dotted blue overlaid curve shows the 50% switching probability, which is the ideal value of the TRNG operation. We aim to operate the device near this value.

3.3.3 Impact of pulse amplitude and duration on $P_{\mathrm{sw}} = 0.5$

The probability of switching is tunable through pulse amplitude V, and pulse duration t_{pw} as shown in Fig. 3.15(b). For a specific t_{pw}, we set V such that the probability of switching is 50% ($P_{\mathrm{sw}} = 0.5$). We denote the voltage required for $P_{\mathrm{sw}} = 0.5$ as $V_{1/2}$. From Fig. 3.15(b), it is clear that $V_{1/2}$ will decrease as the t_{pw} increases and vice versa. However, both the $V_{1/2}$ and t_{pw} are subject to variation because, in reality, it is not feasible to apply an absolutely precise pulse amplitude and duration. We show the impact of such variations on the switching probability around the 50% midpoint (referred to as 'midpoint switching probability' hereafter) in Fig. 3.16. Figure 3.16(a) shows the change in midpoint switching probability as V is varied to be $\pm 10\%$ of $V_{1/2}$ for various t_{pw}. We find the change of midpoint switching probability with respect to V, dP_{sw}/dV is lower for the short-pulse limit than for the longer pulse limit. On the other hand, Fig. 3.16(b) shows the change in midpoint switching probability with respect to t_{pw}, $dP_{\mathrm{sw}}/dt_{\mathrm{pw}}$ for $\pm 10\%$ variations in t_{pw}. We get an opposite trend for the sensitivity to t_{pw}. From Eq. (3.13), it can be shown that dP_{sw}/dV is proportional to t_{pw} while $dP_{\mathrm{sw}}/dt_{\mathrm{pw}}$ is proportional to $(V/V_{c0} - 1)$ around $P_{\mathrm{sw}} = 0.5$. Therefore, in the short-pulse limit, dP_{sw}/dV is lower while $dP_{\mathrm{sw}}/dt_{\mathrm{pw}}$ is higher because short pulses require larger pulse amplitudes. Note that t_{pw} is kept fixed while we vary V. Similarly, V remains fixed at the corresponding $V_{1/2}$ value during t_{pw} variations. Also, note that for Fig. 3.16(b), we show data up to 20 ns because $dP_{\mathrm{sw}}/dt_{\mathrm{pw}}$ becomes vanishingly small for longer pulses.

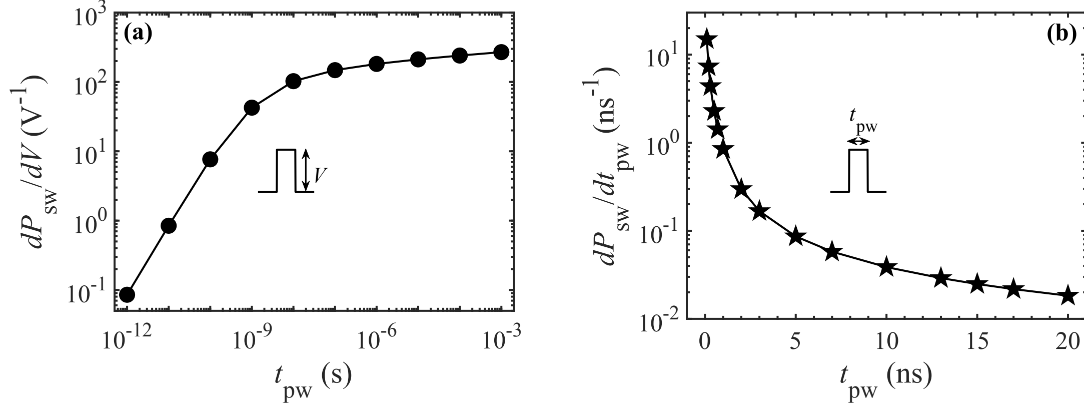


Figure 3.16: Variation in midpoint switching probability with respect to (a) pulse amplitude and (b) pulse duration for various pulse durations.

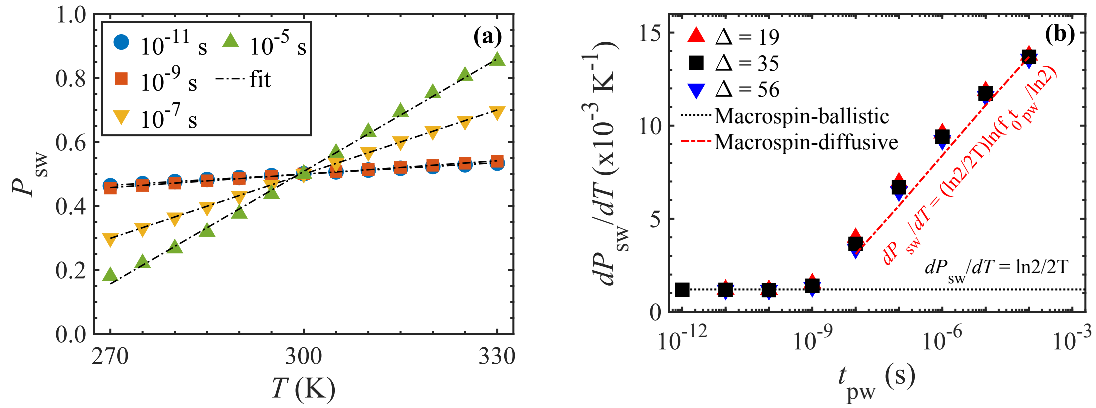


Figure 3.17: (a) Variation in midpoint switching probability as a function of temperature for various pulse durations. The black dash-dotted lines show the linear fit. (b) Variation in midpoint switching probability with respect to temperature for a free layer with different thermal stability factors. The black and red dash-dotted lines in (b) represent the macrospin approximation in the ballistic (short pulse) and diffusive (long pulse) limits, respectively, and the texts represent the corresponding equation.

3.3.4 Impact of temperature on $P_{\text{sw}} = 0.5$

Temperature plays a critical role in STT-driven MTJ switching, as it directly impacts the functionality and reliability of the device [268, 269]. During the writing process, Joule heating can increase the junction temperature, which affects the device performance [270]. The impact of thermal fluctuations mainly affects the initial magnetization distribution and the thermal stability factor ($\Delta = E_b/k_B T$) [263, 269, 271]. When T increases from room temperature (300 K), Δ decreases. Therefore, for a specific t_{pw}, with a fixed $V_{1/2}$, P_{sw} would be greater

than 0.5. Similarly, P_{sw} would be less than 0.5 when T decreases below 300 K as it increases Δ. As expected, Fig. 3.17(a) shows a linear relationship between the P_{sw} and T for a $\pm 10\%$ change in T from room temperature. Figure 3.17(b) shows the dP_{sw}/dT for various t_{pw}. We find that dP_{sw}/dT in the short-pulse limit is lower than for the longer-pulse limit. This higher sensitivity of dP_{sw}/dT in the diffusive limit arises from the double exponential dependence of P_{sw} on the energy barrier and temperature,

$$P_{\text{sw}}^{\text{diffusive}} = 1 - \exp\{-f_0 t_{\text{pw}} \exp(-E_b/k_B T)\}, \tag{3.14}$$

where f_0 is the attempt frequency. Interestingly, our FP-based result agrees well with the macrospin approximation in both ballistic and diffusive limits (Eqs. (3.13), (3.14)). Specifically, in the ballistic limit, dP_{sw}/dT can be estimated as $dP_{\text{sw}}^{\text{ballistic}}/dT = \ln 2/2T$ while for the diffusive limit, $dP_{\text{sw}}^{\text{diffusive}}/dT = (\ln 2/2T)\ln(f_0 t_{\text{pw}}/\ln 2)$ around the $P_{\text{sw}} = 50\%$ value. We use $f_0 = 1$ GHz, which is a commonly accepted value for magnetic materials [147]. We also show the dP_{sw}/dT for the MTJ free layer with lower and higher Δ values than that of MBM ($\Delta \sim 35$). We find that dP_{sw}/dT are similar regardless of the Δ values in the ballistic limit, while there is a slight variation in the dP_{sw}/dT in the diffusive limit, where a lower Δ value results in a higher temperature sensitivity. We understand that temperature change will change the material parameters, and the initial effect of these changes can be accounted for in our model. While we expect qualitatively similar results, considering the temperature effect on all parameters simultaneously requires a detailed analysis, which we leave as a future study.

3.3.5 Impact of process variations on $P_{\text{sw}} = 0.5$

We now move on to the impact of geometric and material parameter variation on the midpoint switching probability. Equation (3.13) suggests that for a specific V and t_{pw}, the probability of switching characteristic is driven by three quantities, namely Δ, V_{c0}, and τ_D. These

quantities are all dependent on the geometric and material parameters of the free layer, and in practice are also susceptible to temperature and process variations during fabrication. It is thus critical to analyze the impact of parameter variation on the switching probability [272].

We show the impact of variation in free layer diameter D and thickness t_F and material parameters M_s, H_k, α, and η on the midpoint switching probability in Fig. 3.18. Among these parameters, D and t_F affect Δ and V_{c0} proportionally through volume, while α and H_k act like physical forces that oppose switching, affecting Δ and V_{c0} proportionally and τ_D inversely. For a fixed H_k, the parameter M_s has a similar effect on Δ and V_{c0}, while η only affects V_{c0} inversely. Overall, from Fig. 3.18, we find that for all kinds of parameter variations, the variation in the midpoint switching probability is weaker in the short-pulse limit than for longer pulses. This attribute indicates the robustness of the TRNG operation against process variations in short-pulse-activated SMART devices. While assessing parameter sensitivity, we keep $V_{1/2}$ fixed to its ideal value for a specific t_{pw} and vary only one parameter at a time. The percent variation for D, t_F, Ms, H_k, α, and η are $\pm 2.5\%$, $\pm 5\%$, $\pm 5\%$, $\pm 5\%$, $\pm 5\%$, $\pm 10\%$, respectively. We select these variations in ranges that each produce a linear fit with P_{sw}. Also, note that for H_k and α variations (Figs. 3.18(d) and 3.18(e)), we exclude very low pulse duration because at such small t_{pw} values, the interplay between Δ, V_{c0}, and τ_D changes P_{sw} in such a way that we are unable to get a linear fit to the P_{sw} data with respect to H_k and α.

3.3.6 Energy cost of switching

The above discussions on the 50% switching probability of the short-pulse driven MBM focus on robustness. However, the other important metric during switching is energy dissipation. Moreover, from Fig. 3.16, we can see that dP_{sw}/dV and $dP_{\mathrm{sw}}/dt_{\mathrm{pw}}$ show opposite trends as a function of t_{pw}. The energy dissipation metric can set the pulse limits for the device to achieve energy efficiency and robustness simultaneously. The STT-driven switching in the presence of a thermal field is stochastic and the junction conductance varies in time in a stochastic way. For a constant applied voltage, we can estimate the ensemble-averaged

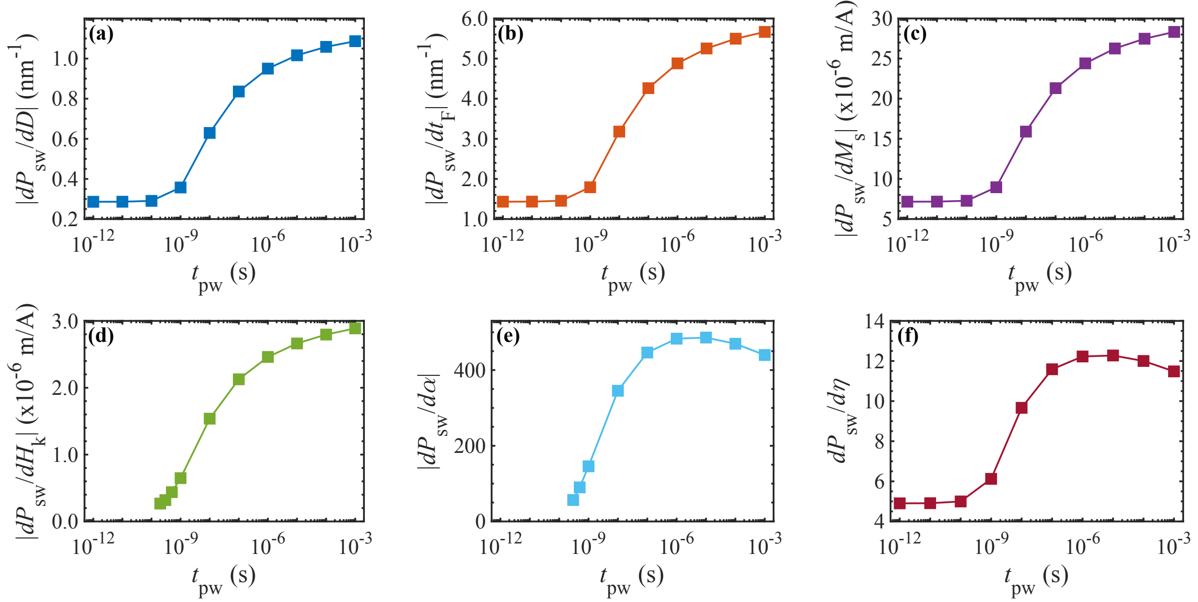


Figure 3.18: Variation in midpoint switching probability with respect to (a) free layer diameter, (b) free layer thickness, (c) saturation magnetization, (d) anisotropy field, (e) magnetic damping coefficient, and (f) spin polarization efficiency factor for various pulse durations. For all the variations, the change in the midpoint switching probability is lower for the short pulse limit than the longer pulse limit, leading to robust TRNG operations.

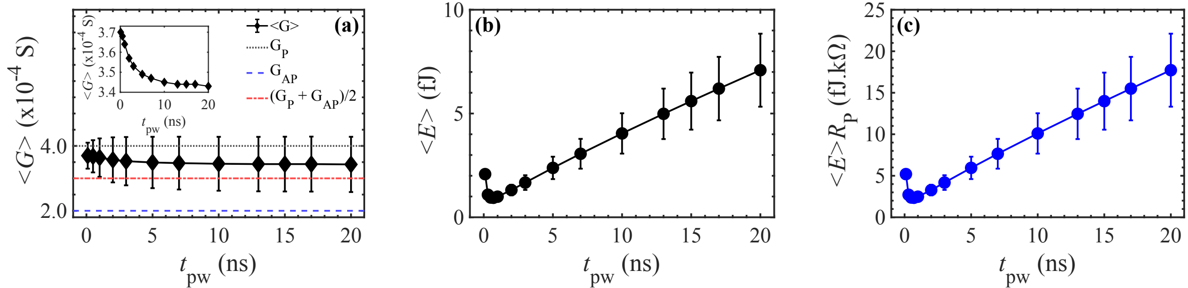


Figure 3.19: (a) Average junction conductance, (b) average energy dissipation, and (c) average energy-resistance product for 50% switching for various pulse durations. In (a), the inset shows the zoomed view of $\langle G \rangle$. In all the figures, the error bars represent the standard deviation in our ensemble.

energy dissipation $\langle E \rangle = V^2 \langle G \rangle t_{\mathrm{pw}}$, where $\langle G \rangle$ is the ensemble-averaged junction conductance during the pulse duration. One might assume $\langle G \rangle = (G_{\mathrm{P}} + G_{\mathrm{AP}})/2$, considering half the time the magnetization is in the P state, while in the other half, it is in the AP state (G_{P} and G_{AP} are the junction conductance in the P and AP state, respectively); however, it is not guaranteed that the magnetization will spend equal time in P and AP states. An

accurate way is to employ the probability density obtained by solving the FP equation, $\langle G\rangle = \frac{1}{t_{\mathrm{pw}}}\int_0^{t_{\mathrm{pw}}} dt \int_0^{\pi} d\theta \rho(\theta;t) G(\theta)$, where $G(\theta) = \frac{1}{2}(G_{\mathrm{P}} + G_{\mathrm{AP}}) + \frac{1}{2}(G_{\mathrm{P}} - G_{\mathrm{AP}})\cos\theta$.

Figure 3.19(a) shows the $\langle G\rangle$ for various t_{pw}. For short t_{pw}, $\langle G\rangle$ is close to the G_{P}, and as the t_{pw} increases, $\langle G\rangle$ decreases because the magnetization spends more time in the AP states (see inset for zoomed view). However, $\langle G\rangle$ starts to saturate for longer pulses. We conjecture that for longer pulses, the thermal energy started taking over, which limits the evolution of probability density to the AP states regardless of the pulse duration. If we consider $\langle G\rangle = (1-k)G_{\mathrm{P}} + kG_{\mathrm{AP}}$, k varies from $\sim 15-30\%$ as a function of t_{pw}. It is noteworthy that $\langle G\rangle$ is greater than $(G_{\mathrm{P}} + G_{\mathrm{AP}})/2$ throughout the range of the t_{pw}. From $\langle G\rangle$ we calculate the ensemble-averaged energy dissipation $\langle E\rangle$ in Fig. 3.19(b). We find a lower $\langle E\rangle$ for the short-pulse limit over the longer-pulse limit. There is an optimal pulse duration that minimizes the energy dissipation [259] because the $V_{1/2}$ increases significantly as t_{pw} decreases, which results in a high $\langle E\rangle$ for very short pulse durations (picoseconds). In the short-pulse limit ($\lesssim 1$ ns), the energy dissipation is in the range of only a few femtojoules, which is orders of magnitude lower than the CMOS-based TRNGs (usually in the picojoules range) [243, 273]. It should be noted, that we are only considering the random bit write (activation) energy here. Finally, we quote $\langle E\rangle R_{\mathrm{P}}$ (in fJ.kΩ unit) because it is relatively easy to vary the R_{P} in the experiment ($R_{\mathrm{P}} = 1/G_{\mathrm{P}}$ is the junction resistance in the P state). Using the tunnel magnetoresistance (TMR) relation, TMR $= (G_{\mathrm{P}} - G_{\mathrm{AP}})/G_{\mathrm{AP}}$ in the $G(\theta)$ equation, from straightforward algebra, it can be shown that the quantity $\langle E\rangle R_{\mathrm{P}}$ depends only on the applied voltage, TMR, and the probability density obtained from the FP equation. In the short-pulse limit, for 100% TMR, $\langle E\rangle R_{\mathrm{P}}$ is ~ 5 fJ.kΩ or lower. Note that in Fig. 3.19, we show data up to 20 ns because the energy dissipation is very high for longer pulses and is unsuitable for comparison with the short-pulse limit.

Table 3.4: Material parameters for MTJ free layer.

Symbol	**Definition**	**Value**
D (nm)	Diameter	15
t_F (nm)	Thickness	1.5
M_s (kA/m)	Saturation magnetization	300 [260]
K_u (kJ/m^3)	Uniaxial anisotropy	600 [260]
α	Damping coefficient	0.01 [260]
TMR	Tunnel magnetoresistance	100% [274]
R_{P} (R_{AP}) (kΩ)	Resistance in P (AP) state	2.5 (5.0)
η	Spin polarization efficiency	0.433

3.3.7 Conclusion

In summary, we demonstrate the suitability of MBM-based SMART devices for TRNG operations for a wide range of pulse durations. We studied the impact of various kinds of variations around the 50% percent switching probability. Furthermore, we evaluate the energy consumption associated with the stochastic switching process. Our results show that the SMART devices operating in the short-pulse limit ($\lesssim 1$ ns) can achieve both robustness and energy efficiency. Our findings offer insights into the development of fast, energy-efficient, and reliable TRNGs for various applications.

Chapter 4

Energy barrier lowering in piezoelectric/magnet/topological insulator/magnet stacks for in-memory computing

The future of semiconductors in electronics, especially in the Edge Computing era, requires confronting multiple fundamental scientific challenges with a holistic approach – encompassing new material combinations with emergent functionalities, devices explicitly designed around those properties, and circuits built on those devices to reduce overall size, power, delay, and error [275, 276]. To this end, the intersection of magnetism, topology, and strain offers a suite of symmetries and tunable properties that can be exploited strategically to synthesize components beyond the reach of conventional semiconductor technology [34]. For instance, a gated piezoelectric can strain a magnet to alter its energy landscape [277] and rotate its magnetization from in to out-of-plane [278]. An out-of-plane magnet atop a 3D TI can gap its topological surface states (TSS) by breaking time-reversal symmetry and turn it semiconducting, modulating its spin conductivity [33, 279, 280]. Finally, the spin-momentum

locked TSS can write information onto an MTJ with a large SOT, its SHA exceeding unity [281–284]. This heterogeneous stack naturally encompasses logic and memory in a PiM architecture with a minuscule energy cost and negligible footprint [34].

PiM is an emerging architectural design that integrates memory and logic simultaneously, leading to energy-efficient and faster information processing [285–288]. In traditional Von Neumann architecture, the physical separation between memory and processing units results in significant latency due to the back-and-forth data transfer [289, 290]. In contrast, in PiM architectures, some processing tasks are performed locally within the memory, allowing the "processed" data to be transferred to the primary processing unit with reduced latency [291]. The individual memory cells in a PiM architecture (bit cells) are organized in a crossbar layout. Selectors control each row and column of the crossbar grid, enabling the bit cells for read or write operations. Utilizing sense amplifiers, the entire row of the crossbar can be read by comparing the state of the bit cells with a known reference voltage or current to perform logic operations within the memory [292–294].

MRAM is a leading contender for a PiM bit cell due to its non-volatility, high speed, low power consumption, high endurance, scalability, and excellent compatibility with CMOS process technology [10, 288, 295]. The fundamental building block for MRAM devices is the MTJ, which consists of a thin insulator sandwiched between two magnetic layers—a "pinned layer" whose magnetization is fixed and a "free layer" whose magnetization can be reoriented by a spin current. Recently, current-induced SOT mechanism, resulting from either the spin Hall effect [71,296] or the Rashba effect [297], has emerged as a promising approach for energy-efficient switching of the MTJ free layer. SOT-based switching addresses several limitations of its counterpart, STT-based switching, such as the need for high write currents, shared read-write paths, and wear of the insulating layer [298–300]. SOT-based MRAM (SOTRAM) is a three-terminal device featuring a decoupled read-write path and in-plane charge current flow in the non-magnetic layer underneath the MTJ free layer with a high charge-to-spin conversion efficiency, yielding a low write current and infinite endurance [71, 296]. SOTRAM

is attractive as a last level embedded cache [301, 302], with distinct advantages for PiM-like edge-based motion detection and cross-modality object recognition [10, 303]. Traditional SOT devices typically utilize HMs such as Pt [296, 304], Ta [71], and W [70] as spin current injectors; however, these HMs exhibit low SHA (Pt: 0.08, Ta: 0.15, and W: 0.4 [305]), leading to low charge-to-spin conversion efficiency [284]. Achieving higher charge-to-spin conversion efficiency (large SHA) is essential for energy-efficient applications.

Recently, TIs have garnered attention as an attractive spin source for SOT switching [281–283, 306, 307]. TIs are characterized by their unique features, including spin-momentum-locked TSS, insulating bulk states, and SHA greater than unity [308–310]. The conducting TSS arise from band inversion at the surface, mediated by strong SOC, and are topologically protected by time-reversal symmetry [311–313]. These attributes make TIs highly efficient spin current injectors for SOT switching, facilitating energy-efficient information writing [314]. Conversely, a magnet's ability to modulate the spin conductivity of the TSS of a TI, when in proximity with the TI, based on its magnetization direction, offers a pathway for an intrinsic gating mechanism to control the TI surface current [33, 34]. Switching the magnet (gating magnet hereafter) from in-plane to out-of-plane orientations, gaps in the TSS can be opened, thereby modulating the spin conductivity [279, 280]. This 90° switching of the gating magnet can be achieved through various mechanisms, such as strain [315–318], voltage control magnetic anisotropy [319], and changing the anisotropy by an applied voltage [33, 320]. Nonetheless, strain-induced 90° switching of magnets is very energy-efficient [32, 277, 317, 318], where a piezoelectric material induces electrical strain in response to an applied gate voltage to facilitate switching from in-plane to out-of-plane and vice versa.

In this chapter, we design and model a potentially compact and energy-efficient four-layer piezoelectric/magnet/TI/magnet (MTJ) stack suitable for a bit cell in in-memory computing architecture. We employ a coupled stochastic LLG simulation to study and analyze the device dynamics. We present the device's functionality, the required phase space for stress generated by the piezo-induced strain and the gating magnet's anisotropy, and the writing condition's

phase space (writing voltage vs. switching time/delay) of the MTJ. We estimate the energy cost for gating and writing mechanisms and find that energy dissipation is significantly reduced compared to HM-based SOT switching, which indicates the benefits of utilizing TI as a spin source and an intrinsic gating mechanism as opposed to an access transistor traditionally used. Furthermore, we project the energy cost for 2-bit AND and OR operations, which shows lower energy costs than traditional HM-based SOT switching. Finally, we show the impact of various material parameters on the device metric and discuss some of the challenges of implementing the devices.

4.1 Computational details

To characterize the magnetization dynamics in our device, we solve a coupled stochastic LLG equation in the macrospin limit using the fourth order-Runge Kutta Method. As explained in section 4.2, the device consists of a piezoelectric/gating magnet/TI/MTJ heterostructure. We simultaneously solve the magnetization dynamics of the gating magnet and the MTJ free layer. In the case of the gating magnet, the LLG equation is described as:

$$\begin{aligned} \frac{1+\alpha_1^2}{\gamma} \cdot \frac{\partial \mathbf{m_1}}{\partial t} &= -\mu_0 \cdot (\mathbf{m_1} \times \mathbf{H}_{\text{eff1}}) \\ &\quad - \alpha_1 \mu_0 \cdot \mathbf{m_1} \times (\mathbf{m_1} \times \mathbf{H}_{\text{eff1}}), \end{aligned} \tag{4.1}$$

where $\mathbf{m_1} = \mathbf{M_1}/M_{\text{s1}}$ is the normalized magnetization and M_{s1} is the saturation magnetization of the gating magnet. α_1, μ_0, and γ are magnetic damping coefficient, permeability of free space, and gyromagnetic ratio, respectively. We consider $\mathbf{H}_{\text{eff1}} = \mathbf{H}_{\text{k1}} + \mathbf{H}_{\text{stress}} + \mathbf{H}_{\text{th}}$, where $\mathbf{H}_{\text{k1}}$ and $\mathbf{H}_{\text{stress}}$ are the effective anisotropy field and stress field of the gating magnet, respectively [321, 322]. $H_{\text{k1}} = \frac{2K_{\text{u1}}}{\mu_0 M_{\text{s1}}} - M_{\text{s1}}$ (K_{u1} is the uniaxial anisotropy of the gating magnet). $H_{\text{stress}} = \frac{3\lambda_s \sigma_s}{\mu_0 M_{\text{s1}}}$ (λ_s is the magnetostriction coefficient of the gating magnet, and σ_s is the stress generated by the electrical strain induced by the piezoelectric).

For the MTJ free layer, the switching is facilitated by the SOT arising from the TI, and the LLG equation takes the form:

$$\begin{aligned}\frac{1+\alpha_2^2}{\gamma}\cdot\frac{\partial\mathbf{m_2}}{\partial t} &= -\mu_0\cdot(\mathbf{m_2}\times\mathbf{H}_{\text{eff2}}) \\ &\quad -\alpha_2\mu_0\cdot\mathbf{m_2}\times(\mathbf{m_2}\times\mathbf{H}_{\text{eff2}}) \\ &\quad -\frac{\hbar}{2e}\cdot\frac{\theta_{\text{sh}}^{\text{eff}}J}{M_{\text{s2}}t_{\text{f2}}}\cdot\mathbf{m_2}\times(\mathbf{m_2}\times\boldsymbol{\sigma}_{\text{p}}),\end{aligned} \tag{4.2}$$

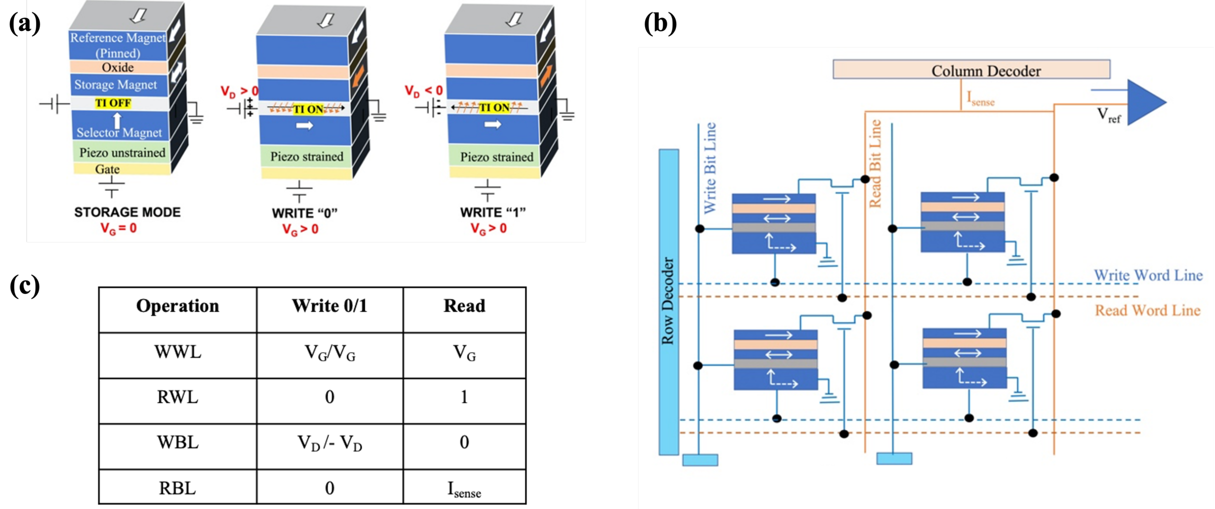
where e is the elementary charge, $\hbar$ is the reduced Plank constant, $\theta_{\text{sh}}^{\text{eff}}$ is the effective spin Hall angle of the TI, t_{f2} is the thickness of the MTJ free layer, J is the surface current density of the TI, and $\boldsymbol{\sigma}_{\text{p}}$ is the unit vector along the spin polarization direction, which is in $y-$direction for $x-$directed current. The other variables and constants have the same meaning as defined in Eq. (4.1), and a subscript of '2' represents the parameter for the MTJ free layer. $\theta_{\text{sh}}^{\text{eff}} = \theta_{\text{sh}}(1-\text{sech}(t_{\text{TI}}/\lambda))$, where θ_{sh}, t_{TI}, and λ are the SHA, thickness, and spin diffusion length of the TI. In Eq. (4.2), $\mathbf{H}_{\text{eff2}} = \mathbf{H}_{\text{d}} + \mathbf{H}_{\text{th}}$. $\mathbf{H_d}$ is the demagnetization field of the in-plane MTJ free layer and expressed as $H_{\text{d}} = -M_{\text{s2}}[N_{\text{dx}}\ N_{\text{dy}}\ N_{\text{dz}}]$, where N_{dx}, N_{dy}, N_{dz} are the demagnetization factors of the free layer along x, y, z axis, respectively.

In both Eqs. (4.1) and (4.2), $\mathbf{H}_{\text{th}}$ is a random thermal field with zero mean ($\mu = 0$) and standard deviation,

$$\text{SD} = \sqrt{\frac{2\alpha k_B T}{\mu_0^2\gamma M_{\text{s}} V\Delta t}}, \tag{4.3}$$

where α, M_{s}, and V are the damping coefficient, saturation magnetization, and volume of the respective magnets. Δt is the simulation time step. μ_0, γ, k_B, and T have their usual meanings.

Throughout our study, we use TbCo as both the gating magnet and the MTJ free layer magnet, Bi_2Se_3 as the TI, and $Pb(Zr,Ti)O_3$ (PZT) as the piezoelectric materials unless otherwise specified. The used parameters are listed in Table 4.1.



Operation	Write 0/1	Read
WWL	V_G/V_G	V_G
RWL	0	1
WBL	V_D/- V_D	0
RBL	0	I_{sense}

Figure 4.1: (a) Schematic of a compact, low-power strained topological insulator SOTRAM (STI-SOTRAM). An out-of-plane gating (selector) magnet gaps the bottom surface states of the TI and places the upper magnet in storage mode (left fig.). The strain generated by the piezoelectric rotates the selector magnet from out-of-plane to in-plane, restoring the bottom layer surface states and activating the TI. The drain polarity V_D sets the storage bit with m_{2y} (bit '0') or $-m_{2y}$ (bit '1') using high SHA SOT (middle and right figs, respectively). The fixed layer of the MTJ reads the state as '0' (parallel) or '1' (antiparallel). (b) Schematic of a PiM crossbar architecture where each activated stack (bit cell) can be selected by activating the row column. Two selected bit cells can feed to a sense amplifier that processes the local inputs by comparing with a reference voltage/current and performing Boolean logic operations (e.g., AND, OR, etc.), thereby processing local data from the magnetic memory. (c) Baising condition for read-write operations. WWL: Write Word Line, RWL: Read Word Line, WBL: Write Bit Line, and RBL: Read Bit Line.

4.2 Four-layer vertical stack for PiM crossbar

Figure 4.1(a) depicts the schematic of the compact four-layer structure of the device comprising a piezoelectric/gating magnet/TI/MTJ stack in the vertical direction. The device contains a gating (selector) magnet at the bottom whose magnetization can be switched in and out of the plane of the TI in nanoseconds with a voltage-gated piezoelectric at a low energy cost [32, 278, 323]. An out-of-plane selector magnet gaps the bottom surface states of the TI and places the upper MTJ free layer in storage mode. Activating the TI bottom surface with applied voltage drives the top MTJ free layer into one of three modes – (i) altering its magnetization with drain bias for data writing, (ii) discharging its stored magnetization state

for data reading, the output set by its low (parallel)/high (antiparallel) resistance relative to the MTJ pinned layer; (*iii*) or execute a logic operation (e.g., bitwise AND, OR, etc.) using a sense amplifier. Since selector and storage magnets are co-located in a vertical geometry, the structure is scalable and naturally suited for a PiM architecture that pre-processes stored data locally, all with the same vertically integrated, compact bit cell as shown in Fig. 4.1(b). Figure 4.1(c) shows the biasing condition for various operations.

4.3 Functionality of the device

The first functional block of the entire stack is the piezoelectric/gating magnet heterostructure. Applying a voltage to the piezoelectric (PZT), electrical strain is generated and transferred to the adjacent gating magnet (TbCo) via the magnetostriction effect. This strain is then converted into stress, multiplied by the elastic modulus of the gating magnet, which counteracts the uniaxial anisotropy of the gating magnet. When the stress is sufficient, it can switch the gating magnet from out-of-plane to in-plane with minuscule energy consumption in the order of $\sim$ 10 aJ [32, 278, 323] (energy cost estimation for our device is shown later). The strain-induced switching of the gating magnet facilitates the opening and closing of the band gap in the TSS of the TI (Bi_2Se_3), depending on the out-of-plane and in-plane magnetization orientation of the gating magnet, respectively.

Figure 4.2(a) shows the magnetization dynamics of the gating magnet under the influence of the stress induced by the piezoelectric strain. We started with an out-of-plane magnetization ($m_{1z} = +1$) for the gating magnet, which corresponds to the OFF state of the device as it opens a band gap in the TSS and, hence, no current flow. Under the influence of stress, it takes a very short time, $\sim$ 1 ns for the 90° switching ($m_{1z} = 0$) of the gating magnet for a stress σ_s = 100 MPa. The switching delay depends on the strength of σ_s and the material parameters of the gating magnet. The out-of-plane magnetization component (m_{1z}) is fed to the next block (TI) dynamically, which modulates the band gap in the TSS and,

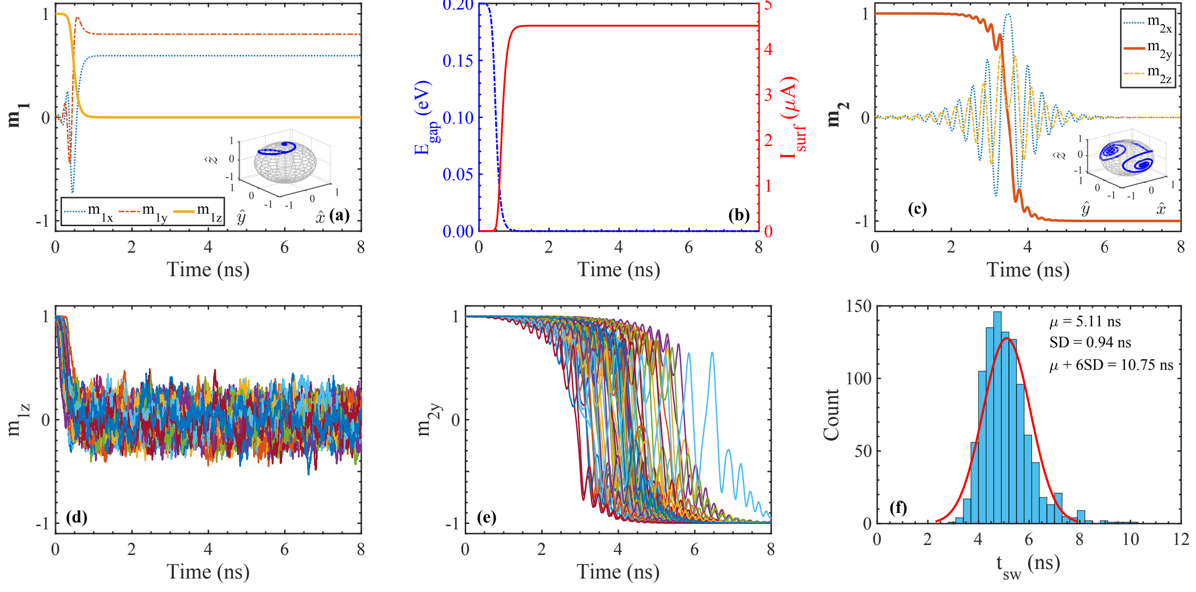


Figure 4.2: Coupled LLG simulations showing gating magnet switching with strain (a), turning ON TSS (closing band gap) and hence delivering surface current (b), and switching MTJ free layer with SOT (c) without the thermal field. (d)-(e) Magnetization dynamics for the gating magnet and the MTJ free layer, respectively, influenced by the stochastic thermal field (50 simulations). (f) Histogram of switching time from 1000 stochastic simulations to estimate the switching time corresponding to WER $= 10^{-9}$. The switching time is calculated as $t_{\mathrm{sw}} = \mu + 6\mathrm{SD}$ from the histogram. The red curve is the Gaussian fit. The results are generated using $I_{0,\mathrm{surf}} = 6I_{\mathrm{c,surf}}$.

hence, the surface current. We consider a Dirac Hamiltonian form to model the surface state of the TI. In the presence of a magnet in proximity to the TI, the Hamiltonian can be expressed as $H = \hbar v_F(\boldsymbol{\sigma} \times \mathbf{k}) \cdot \hat{z} + M_0\mathbf{M_1} \cdot \boldsymbol{\sigma}$ [33, 324], where v_F is the Fermi velocity, $\mathbf{k}$ is the wavevector, $\boldsymbol{\sigma}$ is the Pauli spin matrices, M_0 is the exchange strength between the gating magnet and TI, and $\mathbf{M_1}$ is magnetization of the gating magnet. From simple algebra, for the case of an out-of-plane magnetization orientation of the gating magnet, we can show the energy dispersion takes the form $E = \pm\sqrt{\hbar^2 v_F^2 |k|^2 + (M_0 m_{1\mathrm{z}})^2}$, where $|k| = \sqrt{k_x^2 + k_y^2}$. This gives rise to a band gap $E_{\mathrm{gap}} = \sqrt{(2M_0 m_{1\mathrm{z}})^2}$ at $k = 0$. The band gap will modulate the surface current of the TI as $I_{\mathrm{surf}} = I_{0,\mathrm{surf}} e^{-E_{\mathrm{gap}}/k_B T}$, where $I_{0,\mathrm{surf}}$ represents the surface current needed to switch the MTJ free layer for a specific switching delay. Figure 4.2(b) shows the evolution of the E_{gap} and the I_{surf} with time in response to the magnetization dynamics of the gating magnet (Fig. 4.2(a)). Initially, when $m_{1\mathrm{z}} = +1$, a band gap of $2M_0$

opened, and we get vanishing surface current. As the gating magnet switches ($m_{1z} = 0$) by the stress, the conductivity of the TSS is restored since the band gap is closed. We use $M_0 = 0.1$ eV, a typical value for Bi_2Se_3 [33, 34]. Note that we cannot open an infinite band gap as it is limited by the bulk band gap of the TI (0.3 eV for Bi_2Se_3 [325, 326]). We use $I_{0,\mathrm{surf}} = 6I_{c,\mathrm{surf}}$ while generating Fig. 4.2(b), where $I_{c,\mathrm{surf}}$ is the critical surface current needed to switch the MTJ free layer.

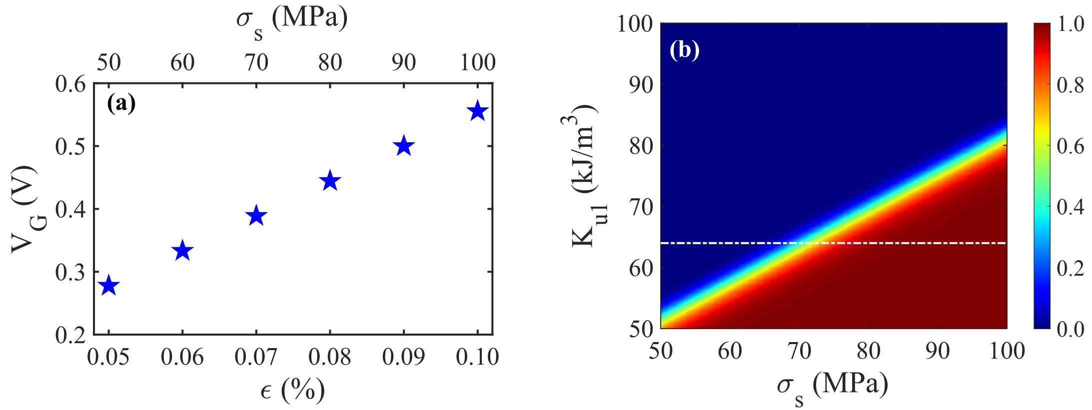


Figure 4.3: (a) Gate voltage requirement in the PZT to generate electrical strain (stress). The capacity to generate strain in the PZT limits the maximum achievable stress. (b) Gating magnet's uniaxial anisotropy vs. stress phase-space for the switching probability (colorplot) of the MTJ free layer. The probability of switching is calculated from 10^5 stochastic simulations. The dash-dotted line represents the TbCo uniaxial anisotropy ($K_{u1} = 64$ kJ/m^3).

We utilize an in-plane magnet-based MTJ for our device for primarily two reasons: (*i*) SOT can provide deterministic switching without an external field assist only for an in-plane magnet because of in-plane spin polarization (orthogonal to the charge current) [71, 281, 327], and (*ii*) we are limited by the interaction between the TI and the MTJ free layer as such an out-of-plane free layer will open gap in the TSS and hence no current conduction [33, 279, 280]. While utilizing an in-plane magnet occupies more area than a perpendicular magnetic anisotropy (PMA) magnet, our design integrates the gating and storage magnets within the same stack. This configuration offers both area and energy efficiency advantages, along with the added benefit of field-free switching. Extensive research is underway for field-free switching of PMA magnets. However, these require a graded structure or an added magnet

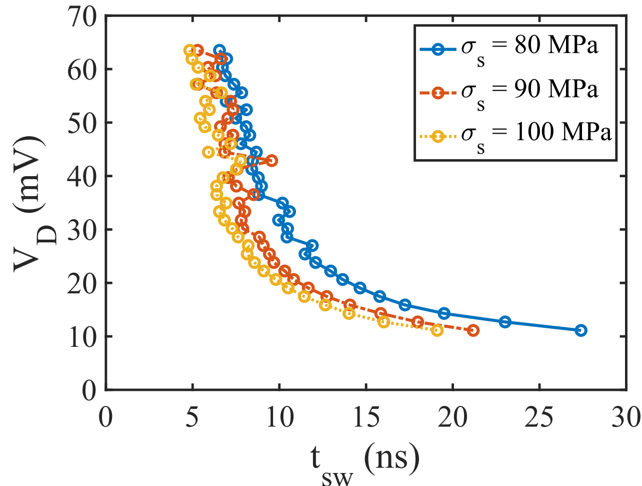


Figure 4.4: Supplied drain voltage to the TI for writing operation. We use $I_{\mathrm{surf}}/I_{\mathrm{tot}} = 15\%$.

that provides the symmetry-breaking field and adds structural complexity [328, 329]. We use an in-plane magnet of "type y" (easy axis in y-direction). The critical current density for switching J_c for this type of MTJ is expressed as [327, 330, 331]:

$$J_c = \frac{2e\alpha_2\mu_0 M_{\mathrm{s2}} t_{\mathrm{f2}}}{\hbar\theta_{\mathrm{sh}}^{\mathrm{eff}}}\left(H_{\mathrm{in}} + \frac{H_{\mathrm{out}}}{2}\right) \tag{4.4}$$

H_{in} is the in-plane shape anisotropy of the free layer, and H_{out} is the out-of-plane demagnetization component. The other variables and constants have the same meaning defined in section 4.1. Using the parameters listed in Table 4.1, we find $J_c = 1.88 \times 10^{10}$ A/m^2, which is at least one order of magnitude less than conventional heavy-metal-based SOT switching [305, 327] and laid the foundation of energy-efficient switching. This amount of critical current density needs to be supplied by the top surface states of TI for the switching of MTJ free layer, from which we can estimate the critical surface current $I_{\mathrm{c,surf}} = J_c W t_{\mathrm{surf}} = 0.75\ \mu$A ($W$ is the device width, $t_{\mathrm{surf}} = 1$ nm is the thickness of the TI top surface [306]).

Figure 4.2(c) shows the magnetization dynamics of the MTJ free layer in response to the SOT generated by the TI surface current. We can clearly see the magnetization switching ($m_{2\mathrm{y}} = +1$ to $m_{2\mathrm{y}} = -1$) with a switching delay of ~ 4 ns for a surface current of $6I_{\mathrm{c,surf}}$. Note that Figs. 4.2(a)-(c) are generated in the absence of a thermal field. Figures 4.2(d)

and (e) show the magnetization dynamics of the gating magnet and the MTJ free layer, respectively, under the influence of the thermal field for 50 simulation runs for the same current as without thermal case, and we can see the stochastic behavior of the magnetization dynamics. Figure 4.2(f) shows the histogram from 1000 simulations runs, and we find a switching time of 10.75 ns, calculated by $t_{\mathrm{sw}} = \mu + 6\mathrm{SD}$, corresponding to a write error rate WER $= 10^{-9}$, which is the industry standard for memory applications [299]. In the subsequent part, we calculate t_{sw} similarly unless otherwise specified.

4.4 Gating and writing mechanism requirements

Figure 4.2 demonstrates the functionality of the device. We will now delve into the details and constraints of the individual blocks. Beginning from the bottom to up, the piezoelectric/gating magnet stack enables the strain-driven out-plane to in-plane switching. As discussed earlier, the switching is governed by the competition between the uniaxial anisotropy energy and the stress energy induced by the strain. PZT typically can generate an electrical strain of 0.05% − 0.1% (500 − 1000 ppm) in response to an applied voltage [32], which sets the achievable stress energy limit to counteract the anisotropy energy. The relation between the applied voltage is $V_G/t_{\mathrm{piezo}} = \epsilon/d_{31}$ [32, 277], where V_G, t_{piezo}, ϵ, d_{31} are the applied gate voltage, thickness of piezoelectric, strain, and piezoelectric constant, respectively. The typical d_{31} constant for PZT is 1.8×10^{-10} [277]. TbCo has Young's modulus of 100 GPa [332] and $\lambda_s = 400 \times 10^{-6}$ [333], which gives a range of stress σ_s from 50 to 100 MPa. Figure 4.3(a) shows the gate voltage required for generating the strain (stress) for a 100 nm thick PZT. The critical question is whether this stress is sufficient to overcome the anisotropy barrier. TbCo has a low uniaxial anisotropy K_{u1} of 64 kJ/m^3 [283, 334]. Considering the demagnetization effect, the effecting anisotropy becomes 38.9 kJ/m^3 ($K_{\mathrm{eff}} = K_{\mathrm{u1}} - \frac{1}{2}\mu_0 M_{\mathrm{s1}}^2$). Conversely, the stress energy $E_{\mathrm{stress}} = \frac{3}{2}\lambda_s\sigma_s$ [321, 322] ranges from 30 kJ/m^3 to 60 kJ/m^3 depending on the amount of strain (stress) shown in Fig. 4.3(a). Figure 4.3(b) presents the phase space for

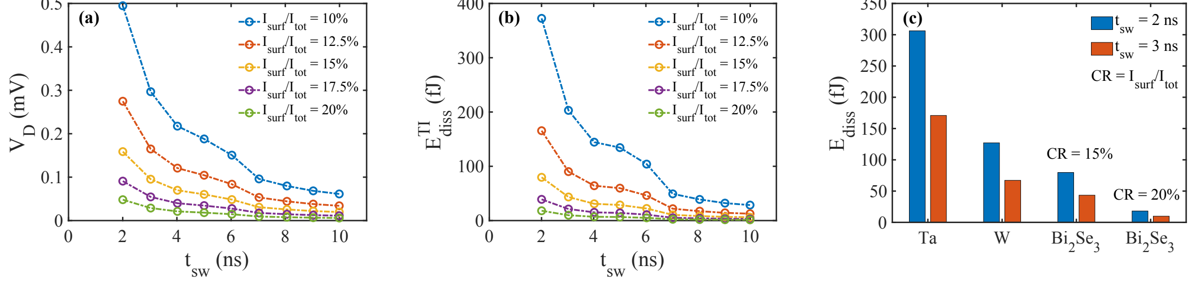


Figure 4.5: (a) Drain voltage requirement for high speed switching regime operation (2−10 ns). (b) Energy dissipation in the TI during the writing process. (c) Comparison of energy dissipation between traditional HM- and TI-based SOT mechanism. TI consumes orders of magnitude less energy than the HM.

anisotropy and stress for the switching probability of the MTJ-free layer (colormap). The P_{sw} is calculated from 10^5 stochastic LLG runs [259] for 10 ns with a surface current of $6I_{\text{c,surf}}$. The figure illustrates the reciprocal relation between the K_{u1} and σ_s, and the range of K_{u1} and σ_s for a working device. This phase space is important because K_{u1} can be tuned by adjusting the composition of TbCo and alloying [335, 336], while σ_s is adjustable through applied gate voltage.

Next, we consider the voltage requirement for the MTJ write operation (switching the MTJ free layer). For the write operation, three important quantities are the switching time/delay, switching current/voltage, and WER. We calculate the switching voltage vs. switching time for various stress values in Fig. 4.4, showing the typical reciprocal relation for spin-torque switching [138, 251, 337]. The switching voltage shown in Fig. 4.4(a) correspond to WER $= 10^{-9}$. As σ_s increases, the out-to-in-plane switching time of the gating magnet is reduced, and it requires less time for the band gap to close, which is attributed to the lower overall switching time of the device. We vary the surface current from 3.5 to 20 times the critical surface current and let the magnetization evolve until it reaches 95% of its final value ($m_{2\text{y}} = -0.95$) [338]. We denote the time as the switching time for the individual switching event, and from the switching time distribution of 1000 simulations, we calculate the t_{sw} corresponding to WER $= 10^{-9}$. While calculating the total current requirement for the required surface current, we consider a $I_{\text{surf}}/I_{\text{tot}} = 15\%$ (see more discussion later).

4.5 Energy cost

Furthermore, we estimate the energy consumption for the switching process. The energy has two components: the gating energy and the switching energy. The energy associated with the gating mechanism comes from the applied voltage in the piezoelectric and can be estimated as $E_{\text{piezo}} = \frac{1}{2}C_pV_G^2$ [32, 277], where C_p is the capacitance of the PZT, and V_G is the applied gate voltage. For the case of 0.1% strain, we get $V_G = 0.56$ V and $C_p = \epsilon_r\epsilon_0 WL/t_{\text{piezo}} = 0.071 \times 10^{-15}$ F (L is the length of the device, ϵ_0 and $\epsilon_r = 1000$ [277] are the permittivity of free space and the relative permittivity of PZT, respectively). We obtain $E_{\text{piezo}} = 11.13$ aJ using these values, which is minuscule and consistent with previous studies [32, 278, 323]. The other part of the energy is the I^2R loss in the TI. TI is typically modeled as parallel surface and bulk channels (bulk channel accounts for the shunting through the bulk states) [306, 326]. The bulk resistance is $R_{\text{bulk}} = L/\sigma_c W t_{\text{bulk}} = 1.46$ kΩ, where σ_c is the average conductivity of the TI ($\sigma_c = 5.7 \times 10^4\ \Omega^{-1}m^{-1}$ for Bi_2Se_3 [305]), and t_{bulk} is the thickness of the bulk state of TI. We use $t_{\text{bulk}} = 6$ nm since the TI is 8 nm thick and the thickness of the top and the bottom surface is 1 nm each [306]. We estimate R_{surf} based on the current distribution in the top surface channel, which is found to be 30% of the total current as previously reported in both theory [326] and experiments [306]. This makes $R_{\text{surf}} = 1.94$ kΩ. Nonetheless, the bottom surface needs to be ground through the gating magnet for the gating mechanism to work for our vertical structure. Therefore, we split the bottom surface resistance equally with a ground in the middle and estimate the equivalent resistance $R_{\text{eq}} = 633.5\ \Omega$, and eventually, we achieve 15% current through the top surface channel of the TI for our resistance configuration. Shunting through the bulk in TI is an open question, and it requires materials with a higher band gap to get more current in the surface [326]. For high-speed applications, we want the switching time in single-digit nanoseconds [299, 302, 339]. We estimate the energy used in this operation regime. Figure 4.5(a) shows the voltage requirement for t_{sw} for $2 - 10$ ns (WER $= 10^{-9}$), while Fig. 4.5(b) shows the energy consumption for the writing operation for a range of $I_{\text{surf}}/I_{\text{tot}}$.

$I_{\text{surf}}/I_{\text{tot}}$ can be varied by tuning the conductivity, bulk band gap, and thickness of the TI [306]. Our analysis suggests energy consumption in the TI is in the order of < 100 fJ, even with a low current ratio of 15% in the top surface, which is energy efficient compared to HM-based SOT switching (Fig. 4.5(c)), consistent with previous studies [283, 305]. If we increase the top surface current, the energy consumption will be decreased further. Fig 4.5(c) shows the energy consumption comparison between HM and TI-based switching, which shows an order of magnitude reduction in energy consumption for TI. Although the conductivity of HM is very high, the current dictates the I^2R loss, and TI requires a significantly low current because of its higher SHA [305].

The total energy consumption $E_{\text{write}} = E_{\text{piezo}} + E_{\text{TI}}$, which is dominated by the E_{TI} because of the minuscule energy in the gating mechanism. This suggests the advantage of the stain-based intrinsic gating mechanism because it eliminates the need for an access transistor for the write operation, costs negligible energy, and has a reduced footprint compared to a CMOS access transistor. The energy consumption in TI will be much lower if we incorporate other TIs with higher SHAs. For example, for BiSb (SHA = 10 and $\sigma_c = 1.5 \times 10^5\ \Omega^{-1}\text{m}^{-1}$ [305]), the energy consumption is ~ 0.9 fJ and ~ 0.2 fJ for a current ratio of 15% and 20%, respectively, which is $\sim 100\times$ lower than Bi_2Se_3 and $\sim 1000\times$ lower than traditional HM.

4.6 Energy cost projection for in-memory AND and OR operations

The above-estimated write energy is at the bit cell level. Now, we project the energy cost for a PiM array from the bit cell energy. In particular, we show the projection for a 2−bit Boolean AND and OR operation using the PiM crossbar array (Fig. 4.1(b)). Before doing the projection, we need to estimate the read energy cost. For reading operation, we use a small sense current $I_{\text{sense}} = 1\ \mu$A, significantly less than the write current, eliminating the risk of an accidental switch. We assume a resistance area product for the MTJ $R_{\text{P}}A = 2\ \Omega\mu\text{m}^2$ and

a TMR = 100% [31], which makes $R_{\mathrm{P}} = 2.5$ kΩ and $R_{\mathrm{AP}} = 5$ kΩ, respectively (R_{P} and R_{AP} are the resistance of MTJ in parallel and anti-parallel state, respectively). We consider a 16 nm PTM HP model for the read access transistor with a $R_{\mathrm{ON}} = 5$ kΩ. We need to keep the gate voltage ON for the read operation so that the TI surface state is conductive, adding the gating energy during the read operation. Considering this, we found a read energy of ~ 45 aJ and ~ 55 aJ for parallel (bit '0') and antiparallel (bit '1') states at a read speed of 4 ns [340]. For in-memory 2−bit AND and OR operations, we need to select two bit cells simultaneously, and the PiM operation is done by comparing the sense voltage from the bit cells with a reference voltage V_{ref} through a sense amplifier (see Fig. 4.1(b)), where V_{ref} is set to a value depending on the operation we want to perform. For two bit cells, we will have $\{R_{\mathrm{AP}},\ R_{\mathrm{AP}}\}$, $\{R_{\mathrm{AP}},\ R_{\mathrm{P}}\}$, $\{R_{\mathrm{P}},\ R_{\mathrm{AP}}\}$, $\{R_{\mathrm{P}},\ R_{\mathrm{P}}\}$ combinations in the resistance states of the MTJ, corresponding to sense voltages $\{V_{\mathrm{AP}},\ V_{\mathrm{AP}}\}$, $\{V_{\mathrm{AP}},\ V_{\mathrm{P}}\}$, $\{V_{\mathrm{P}},\ V_{\mathrm{AP}}\}$, $\{V_{\mathrm{P}},\ V_{\mathrm{P}}\}$. Setting a $V_{\mathrm{ref}} = (V_{\mathrm{AP,AP}} + V_{\mathrm{AP,P}})/2$ ensures the sense amplifier will produce an output of '1' only when both the bit cells are in antiparallel state (logic AND). Similarly, for logic OR, $V_{\mathrm{ref}} = (V_{\mathrm{AP,P}} + V_{\mathrm{P,P}})/2$ will serve the purpose. We can set the reference voltage by tuning the reference resistance of the sense amplifier, and therefore, we can perform re-configurable in-memory computing by tuning the reference voltage. The energy cost for a full cycle (write-read) in-memory Boolean computing and area are summarized in Table 4.2. Note that we assume a symmetric write operation [341] for both bits '0' and '1', and the energy associated with the sense amplifier is estimated as $\frac{1}{2}C\Delta V^2$, where C is the capacitance of the sense amplifier (typically 1 pF [342, 343]) and ΔV is the difference between the sense voltage and the reference voltage.

4.7 Impact of material parameters

Finally, we present the effect of various key parameters on the device switching time t_{sw}. We show the effect of the thermal stability factor Δ_1, exchange constant M_0, and damping

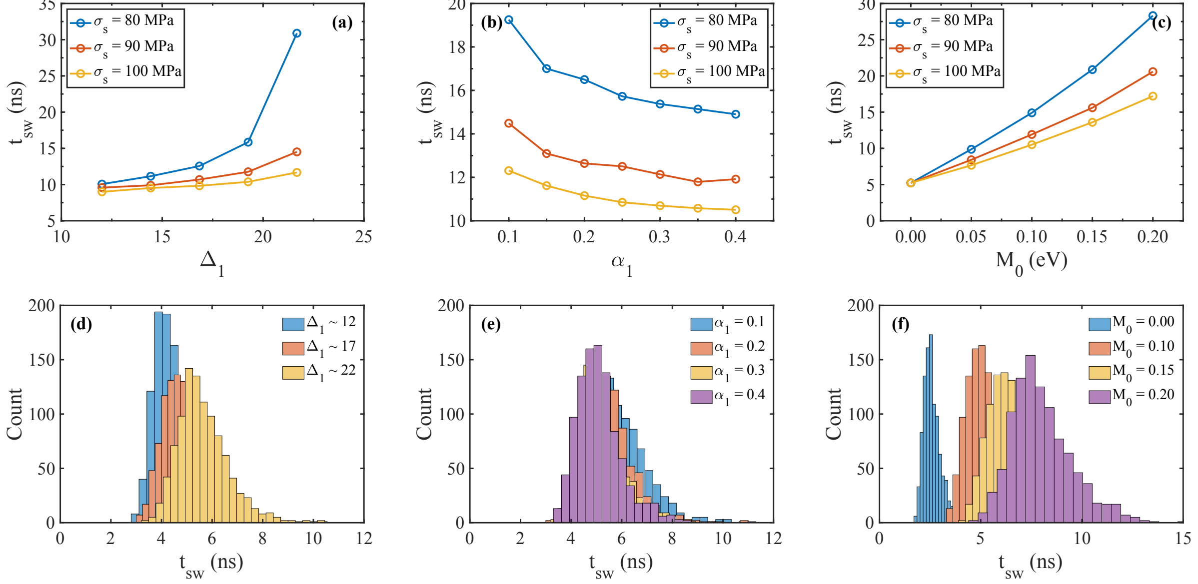


Figure 4.6: Variation in switching time with respect to (a) the thermal stability factor of the gating magnet, (b) the damping coefficient of the gating magnet, and (c) the exchange strength between the TI and the gating magnet. (d)-(f) Histograms of the switching time from 1000 stochastic LLG simulations corresponding to (a)-(c), respectively. We use $I_{0,\mathrm{surf}} = 6I_{\mathrm{c,surf}}$ while calculating the t_{sw}.

coefficient α_1 of the gating magnet on t_{sw} for various σ_s values. In Fig. 4.6, the top panel shows the t_{sw} while the bottom panel shows the histogram of t_{sw} from 1000 simulations. For the case of varying Δ_1, as it increases, the switching time of the gating magnet will increase for a specific σ_s, which increases the overall switching time of the device. While for a larger α_1, t_{sw} reduces as the gating magnet switching time reduces with a large damping coefficient because the gating magnet dynamics is governed by precessional and damping torques. Lastly, if we increase M_0, t_{sw} increases because the gap opening is larger, and it takes longer to provide sufficient surface state current in the TI. Note that even if we get a large M_0, the TSS band gap cannot be infinitely large as we are limited by the bulk bandgap of the TI. In contrast, we need a larger value of M_0 for a high ON/OFF ratio, which suggests the need for a TI with larger bulk bandgap [34].

Table 4.1: Material parameters

Dimension		
Length	L	20 nm
Width	W	40 nm
Gating Magnet: TbCo		
Thickness	t_{f1}	2.5 nm
Saturation magnetization	M_{s1}	200×10^3 A/m [283, 334, 336]
Damping	α_1	0.4
Uniaxial anisotropy	K_{u1}	64×10^3 J/m^3 [283, 334]
Magnetostrictive coefficient	λ_s	400×10^{-6} [333]
Young's modulus	Y	100×10^9 Pa [332]
MTJ Free Layer: TbCo		
Thickness	t_{f2}	12.5 nm
Saturation magnetization	M_{s2}	400×10^3 A/m [336, 344]
Damping	α_2	0.01
Topological Insulator: Bi_2Se_3		
Thickness	t_{TI}	8 nm [283, 306]
Spin Hall angle	θ_{sh}	3.5 [305]
Spin diffusion length	λ	6.2 nm [345]
Conductivity	σ_c	5.7×10^4 Ω^{-1}m^{-1} [305]
Piezoelctric: PZT		
Thickness	t_{piezo}	100 nm
Piezoelectric constant	d_{31}	1.8×10^{-10} m/V [277]
Max. strain	ϵ	1000 ppm (0.1%) [32]
Relative dielectric constant	ϵ_r	1000 [277]

4.8 Challenges and opportunities

The above results show excellent promise for utilizing the STI-SOTRAM for an energy-efficient bit-storage device. However, it is worth discussing the challenges associated with

Table 4.2: Energy consumption and area for in-memory AND and OR operations. E_{write} and E_{read} are the combined energy for both the bit cells, while $E_{\text{sense amp.}}$ represents the energy associated with a single sense amplifier. Area is the total area of the bit cells and the read access transistor (W/L = 10:1).

AND operation					
Bit cell 1 state	Bit cell 2 state	E_{write} (fJ)	E_{read} (fJ)	$E_{\text{sense amp.}}$ (fJ)	E_{total} (fJ)
0	0	159.455	0.09	4.01×10^{-4}	159.55
0	1	159.455	0.10	6.30×10^{-5}	159.56
1	0	159.455	0.10	6.30×10^{-5}	159.56
1	1	159.455	0.11	6.30×10^{-5}	159.57
OR operation					
0	0	159.455	0.09	3.65×10^{-5}	159.55
0	1	159.455	0.10	3.65×10^{-5}	159.56
1	0	159.455	0.10	3.65×10^{-5}	159.56
1	1	159.455	0.11	4.80×10^{-4}	159.57
Area = 6720 nm^2					

implementing STI-SOTRAM devices. There, of course, exist material issues [346] since we need three separate interfacial processes with input-output isolation – (*i*) strain rotating the bottom selector magnet, (*ii*) modulating the TSS with such rotation, and (*iii*) TSS writing information onto the top storage magnet. For (*i*), magnetocrystalline anisotropy density K_u is often larger than stress energy; however, sputtered magnets such as amorphous TbCo can be compositionally tuned [335, 336, 347], while epitaxially grown CoFe film thickness can be tailored to approach a small K_u near the out-of-plane to in-plane crossover occurs [348]. It allows for achieving low K_u, which can be overcome through stress energy while large enough to avoid spontaneous thermal fluctuations. The condition for switching is $(3/2)\lambda_s\sigma_s > K_u$. These magnets exhibit high $\lambda_s \sim 200 - 400 \times 10^{-6}$ [333, 349] with elastic moduli $\sim$ 100 GPa [332], yielding the switching of the magnetization with low stress.

For (*ii*), theory [280] and experiments [279] show that an interfacing out-of-plane magnet opens a TSS bandgap, leaving only chiral quantum anomalous edge states with significantly diminished spin surface current. A 3D tight binding-based study suggests the same, even accounting for bands along the patterned TI side walls [34]. Critically, the magnet must adjoin the TI, providing at least 50 mV exchange coupling. The g-factor for the Zeeman gap

is highly material dependent (~ 18 for Bi_2Se_3, ~ 6 for Sb_2Te_2Se [350]), while that depends on bulk bandgap due to the magnetic proximity effect (~ 20 meV for EuS/Bi_2Se_3 [351], ~ 100 meV for $MnBi_2Se_4/Bi_2Se_3$ [352]).

For (*iii*), SHA, even in sputtered TI, can be large ($\sim 10-20$, about $100\times$ greater than conventional HMs, e.g., Pt), switching magnets very efficiently [305, 353]. TIs have small bulk bandgaps $\sim 100-300$ meV that shunt current into the bulk and sidewalls. However, previous studies show that even for a modest bandgap TI, an ON-OFF ratio of ~ 10 can flip the spins of a magnet with a very low WER of 10^{-9} [34]. This self-correction happens through internal anisotropy fields [354], an increased bandgap with thin film quantization [305, 355], and reduced charge current with gap opening. To avoid current shunting into the storage magnet, a thin layer of insulating NiO [356] or MgO [357] can be grown that transmits magnon torque between the magnet/TI, or else use an insulating $BaFe_{12}O_{19}$ as the storage magnet's free layer [358].

Now, the important question arises: can we grow and pattern the heterogeneous stack? Recently, our collaborators reported sputtered growth of amorphous/nanograined Bi_2Se_3 directly onto a magnet on a piezo with compatible processing temperatures (personal communication). Moreover, previous experiments show amorphous TI retains bandgap, topological protection, and high SHA [305, 359, 360], even after patterning with wet etching [361]. Therefore, the usual back-end-of-the-line (BOEL) metallization steps should work. Besides mechanical decoupling, the stacks are also electrically decoupled, as the strains on a common piezo substrate stay localized around the contacts [278, 362].

The other important concern is endurance. While HM-based SOTRAM endurance is $\sim 10^{15}$, piezo fatigue and TI endurance limit our STI-SOTRAM device to $\sim 5\times10^{6}$ cycles [363, 364]. This suffices for hyperdense, nonvolatile persistent cells, as the probability of data overwrites per cell is very low ($\sim 10^{-3}$, easily error corrected [365]).

4.9 Conclusion

In summary, we demonstrated a novel four-layer piezoelectric/magnet/TI/MTJ bit cell design with reduced energy cost and footprint for in-memory computing. Eliminating an access transistor with a built-in strain-based gating mechanism has proved energy-efficient, yielding an overall reduced energy cost. High-speed operation is achieved with significantly less energy than the traditional SOT metals at the device and array level. Our results suggest this heterogeneous stack may provide a compact and energy-efficient design for low-power, high-speed in-situ applications.

Chapter 5

Conclusion and outlook

The primary objective of this dissertation is to explore different modes of energy barrier engineering in nanomagnetic devices used in spintronics applications, aiming to address some of the bottlenecks of the existing computing technology. Nanomagnetic devices have the potential to become a universal memory due to their non-volatility, scalability, and ultra-fast operation, which could meet density and speed requirements simultaneously while offering the extra feature of non-volatility. However, challenges arise as the size of the magnets is reduced. In ultra-small dimensions, there is a trade-off between density and non-volatility. The dichotomy is that smaller devices allow for high density but become volatile, losing the long data retention. Conversely, if we want to store information for a long time, we must use larger devices because the non-volatility is proportional to the volume of the magnet. This dissertation aims to assess and workaround these conflicting issues associated with nanomagnetic devices.

Specifically, we explore three facets of it. First, we examine ultra-small and ultra-fast skyrmions as an alternative to traditional nanomagnets to leverage their small size while retaining a high barrier and engineer their topological barrier by controlling DMI. We unveil the underlying mechanism of the observed DMI behavior in a ferrimagnetic system with a composition gradient in the capping layer and predict the energy-efficient cap layer materials.

Additionally, we explore barrier engineering by placing notches in a magnetic racetrack, which can provide a longer lifetime of skyrmions, supplemented by empirical equations and ML-based automated prediction of the energy barrier. Next, we delve into the feasibility of utilizing the smaller magnets with low barriers for probabilistic computing, stochastic, and neuromorphic computing, capitalizing on the inherent stochasticity exhibited by this class of magnets. We investigate the influence of different variations on the computational networks built from LBM-based hardware units and quantify the error associated with the output. We also analyze the performance of various neuron models that connect to the stochasticity of the LBM for signal inferencing tasks. Additionally, we model more practical and experimentally realistic MBMs that can potentially be energy-efficient, fast, and robust TRNG units for probabilistic and stochastic computing. Finally, we combine new materials to build a heterogeneous stack with an energy-efficient gating mechanism, which dynamically engineers the barrier of a magnet for out-to-in-plane and vice versa. We deploy a piezoelectric/magnet/TI/magnet stack to utilize the highly energy-efficient strain-based gating to strategically modulate the TSS of a TI for energy-efficient data writing. This heterogeneous stack naturally encompasses logic and memory in a PiM architecture with a minuscule energy cost and negligible footprint. This dissertation may provide a path forward to energy barrier engineering in nanomagnetic devices, potentially leading to next-generation technologies, and here is an incomplete list of possible extensions of this work:

Storing multi-bit information in a notched racetrack

Multi-bit operations would require multiple notches in a racetrack. One of the future studies in this direction could be to analyze the energy barrier and dynamics of skyrmions in a racetrack with multiple notches to find out the optimal separation between the notches and the relation between notch separation and skyrmion diameters. Other topics could include the interaction between the skyrmion and notches while changing positions, the feasibility

of using multiple skyrmions, and the effect of temperature on the metrics, i.e., the energy barrier and unpinning current.

Ferroelectric "footprint" skyrmions

A known challenge in skyrmionics is the electrical reading of skyrmions, as the fill factor of skyrmions compared to the reading MTJ is very low, which produces a low TMR and makes it very difficult to read the resistance change based on the presence and absence of skyrmions. Although emerging efforts are ongoing for all electrical detection of skyrmions, one promising direction could be to use multiferroics materials as an abstraction layer to have an electrical footprint of magnetic skyrmions (especially Néel skyrmions for practical applications) from a magnetic layer so that the skyrmions can be detected capacitively.

Weyl semimetals (WSM) for field-free SOT switching

TI is an excellent spin source with a high SHA. However, the spin polarization is in-plane to the TI. Therefore, the SOT from TI can only provide deterministic switching for in-plane magnets. Here, low-symmetry materials, such as WSM, are potential candidates for field-free switching to a PMA magnet because they can provide out-of-plane spin polarization and hence out-of-plane SOT components. One future study worth pursuing is to use a piezoelectric/magnet/WSM/PMA-magnet stack and control the conductivity of WSM surface states through the strain from the piezoelectric to achieve field-free switching of the PMA magnet. The feasibility of combining a TI and WSM as the spin source to achieve field-free switching can also be explored.

Appendices

Appendix A

Dissipation in SMART MTJ devices

For stochastic magnetic actuated random transducer (SMART) devices, we assume (and to a good approximation) that the junction conductance depends on the angle between the magnetizations of the free and fixed layer θ as follows:

$$G(\theta) = \frac{1}{2}(G_{\mathrm{P}} + G_{\mathrm{AP}}) + \frac{1}{2}(G_{\mathrm{P}} - G_{\mathrm{AP}}) \cos\theta. \tag{A.1}$$

The tunneling magnetoresistance (TMR) is thus the fraction:

$$T = \frac{G_{\mathrm{P}} - G_{\mathrm{AP}}}{G_{\mathrm{AP}}}. \tag{A.2}$$

A.1 Energy dissipation

We now consider the energy dissipated during a voltage pulse of duration t_{pw}:

$$E = \int_0^{t_{\mathrm{pw}}} \frac{V^2}{R(t)} dt = \int_0^{t_{\mathrm{pw}}} V^2 G(t) dt = V^2 \int_0^{t_{\mathrm{pw}}} G(t) dt. \tag{A.3}$$

As we take the direction of the magnetization of the fixed layer to be up (in the z-direction), θ is the polar angle of the free layer magnetization. The ensemble average energy dissipated

is thus:

$$\langle E \rangle = V^2 \int_0^{t_{\mathrm{pw}}} G(\theta) P(\theta, t) dt, \tag{A.4}$$

where $P(\theta, t)$ is the probability of the magnetization angle θ at time t computed with the Fokker-Plank equation (FPE).

Using Eq. (A.2), $G_{\mathrm{P}} = (T+1)G_{\mathrm{AP}}$ and we can rewrite Eq. (A.3) as follows:

$$G(\theta) = \left[\frac{T}{2}(1 + \cos\theta) + 1\right] G_{\mathrm{AP}}, \tag{A.5}$$

$$= \left[T \cos^2 \frac{\theta}{2} + 1\right] G_{\mathrm{AP}}, \tag{A.6}$$

or, in terms of the parallel state conductance G_{P}:

$$G(\theta) = \frac{G_{\mathrm{P}}}{1+T}\left[T \cos^2 \frac{\theta}{2} + 1\right]. \tag{A.7}$$

Finally, the average energy dissipated to create a random bit can be written:

$$\langle E \rangle = \frac{V^2 G_{\mathrm{P}}}{1+T} \int_0^{t_{\mathrm{pw}}} \left(T \cos^2 \frac{\theta}{2} + 1\right) P(\theta, t) dt. \tag{A.8}$$

So, we see that the energy dissipated depends on the parallel state conductance and the TMR. The larger the conductance and TMR, the greater the energy that will be dissipated and one can write:

$$\langle E \rangle R_{\mathrm{P}} = \frac{V^2}{1+T} \int_0^{t_{\mathrm{pw}}} \left(T \cos^2 \frac{\theta}{2} + 1\right) P(\theta, t) dt, \tag{A.9}$$

where $R_{\mathrm{P}} = 1/G_{\mathrm{P}}$ is the junction resistance in the parallel state. In other words, $\langle E \rangle R_{\mathrm{P}}$ only depends on the voltage amplitude, TMR, and FPE results, $P(\theta, t)$.

Appendix B

PiM sensing scheme

The working principle for in-memory computing is based on a simple sensing scheme. Figure B.1(a) shows the equivalent circuit for a 2−bit operation where two bit cells are simultaneously selected through the row-column selectors. R_1 and R_2 are the resistance of the read access transistors, and R_{M1} and R_{M2} are the resistance of the MTJ. Depending on the state of the MTJ, we can get either $\{R_{\mathrm{AP}}, R_{\mathrm{AP}}\}$; $\{R_{\mathrm{AP}}, R_{\mathrm{P}}\}$; $\{R_{\mathrm{P}}, R_{\mathrm{AP}}\}$; $\{R_{\mathrm{P}}, R_{\mathrm{P}}\}$ values for R_{M1} and R_{M2}, respectively, where R_{P} is the resistance when the MTJ is in parallel state while R_{AP} for anti-parallel state. Figures B.1(b)-(e) depict the resistance combinations used in our study (see Section 4.6).

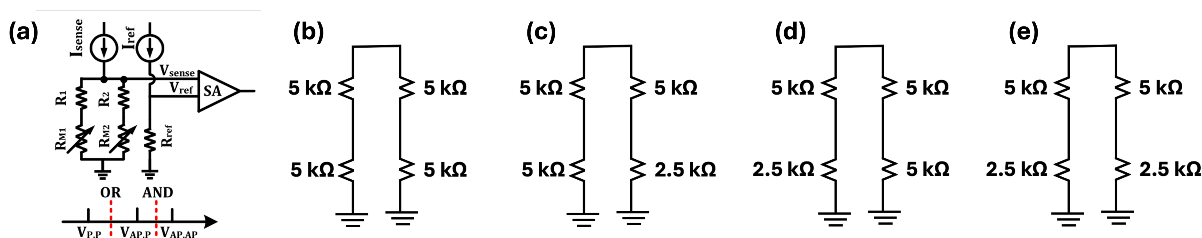


Figure B.1: (a) Equivalent circuit representation when two bit cells are selected simultaneously and fed to a sense amplifier for in-memory computing. The voltage level for AND and OR operation are shown at the bottom. (b)-(e) Various possible combinations of resistance used in our study. Fig. (a) is reproduced from [10] with permission.

Now, we can calculate the V_{sense} by multiplying the equivalent resistance with the sense current, and that is compared with a reference voltage that is set depending on the desired

operation. For example, for Boolean AND operations, we need to set the V_{ref} as $\frac{V_{\text{AP,AP}}+V_{\text{AP,P}}}{2}$, so that the sense amplifier will give an output of logical '1' only if both the MTJs are in AP state. Similarly, for OR operations, the V_{ref} would be $\frac{V_{\text{AP,P}}+V_{\text{P,P}}}{2}$.

For the mean specific resistance and sense current ($I_{\text{sense}} = 1\ \mu$A) used in our study, we get the following values for various voltages are: $V_{\text{AP,AP}} = 5$ mV; $V_{\text{AP,P}} = 4.29$ mV; $V_{\text{P,P}} = 3.75$ mV; $V_{\text{ref,AND}} = 4.65$ mV; $V_{\text{ref,OR}} = 4.02$ mV.

Appendix C

Publications

C.1 Journal articles (published)

1. L. Rehm, **MG. Morshed**, S. Misra, A. Shukla, S. Rakheja, M. Pinarbasi, A. D. Kent, and A. W. Ghosh, "Temperature-Resilient True Random Number Generation with Stochastic Actuated Magnetic Tunnel Junction Devices," *Appl. Phys. Lett.* 124, 052401 (2024).

2. **MG. Morshed**, S. Ganguly, and A. W. Ghosh, "A Deep Dive into the Computational Fidelity of High Variability Low Energy Barrier Magnet Technology for Accelerating Optimization and Bayesian Problems," *IEEE Magnetics Letters*, vol. 14, pp. 1-5 (2023).

3. **MG. Morshed**, S. Ganguly, and A. W. Ghosh, "Choose your tools carefully: A Comparative Evaluation of Deterministic vs. Stochastic and Binary vs. Analog Neuron models for Implementing Emerging Computing Paradigms," *Front. Nanotechnol.* 5:1146852 (2023).

4. **MG. Morshed**, H. Vakili, and A. W. Ghosh, "Positional Stability of Skyrmions in a Racetrack with Notched Geometry," *Phys. Rev. Applied* 17, 064019 (2022).

5. **MG. Morshed**, KH. Khoo, Y. Quessab, J-W. Xu, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh, "Tuning Dzyaloshinskii-Moriya Interaction in Ferrimagnetic GdCo: A First Principles Approach," *Phys. Rev. B* 103, 174414 (2021).

6. Y. Quessab, J-W. Xu, **MG. Morshed**, A. W. Ghosh, and A. D. Kent, "Interplay between Spin-Orbit Torques and Dzyaloshinskii-Moriya Interactions in Ferrimagnetic Amorphous Alloys," *Advanced Science* 8, 2100481 (2021).

7. H. Vakili, J-W. Xu, W. Zhou, MN. Sakib, **MG. Morshed**, T. Q. Hartnett, Y. Quessab, K. Litzius, C. T. Ma, S. Ganguly, M. Stan, P. V. Balachandran, G. S. D. Beach, S. J. Poon, A. D. Kent, and A. W. Ghosh, "Skyrmionics - Computing and Memory Technologies based on Topological Excitations in Magnets," *J. Appl. Phys.* 130, 070908 (2021).

8. C. T. Ma, S. Kittiwatnakul, A. Sittipongpittaya, Y. Wang, **MG. Morshed**, A. W. Ghosh, and S. J. Poon, "Phase Change Induced Magnetic Switching through Insulator-metal Transition in VO_2/TbFeCo Films," *Nanomaterials*, 13(21), 2848 (2023).

C.2 Journal articles (under review & in-preparation)

1. A. Biswas, S. Liu, S. Garg, **MG. Morshed**, H. Vakili, A. W. Ghosh, and P. V. Balachandran, "Integrating Adaptive Learning with Post-Hoc Model Explanation and Symbolic Regression to Build Interpretable Surrogate Models," (2024). [*under review*]

2. **MG. Morshed**, L. Rehm, A. Shukla, Y. Xie, S. Ganguly, S. Rakheja, A. D. Kent, and A. W. Ghosh, "Reduced sensitivity to process, voltage and temperature variations in activated perpendicular magnetic tunnel junctions based stochastic devices," *arXiv*:2310.18781 (2023). [*under review*]

MG. Morshed, H. Vakili, MN. Sakib, S. Ganguly, M. R. Stan, and A. W. Ghosh, "Strained topological insulator spin-orbit torque random access memory (STI-SOTRAM) bit cell for energy-efficient in-memory computing," (2024). [*in-preparation*]

C.3 Conferences proceedings

1. R. Sreekumar, F. E. Mullick, **MG. Morshed**, A. W. Ghosh, and M. Stan, "Exploring Noise-Resilient Spiking Neural Encoding using $\Sigma\Delta\Sigma$ neuron," *2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Springfield, MA, USA, 2024.

2. F. E. Mullick, R. Sreekumar, **MG. Morshed**, S. Ganguly, M. Stan, and A. W. Ghosh, "Double magnetic tunnel junction based $\Sigma\Delta\Sigma$ hardware neuron," *2024 IEEE 82nd Device Research Conference (DRC)*, College Park, MD, USA, 2024.

3. A. Shukla, L. Heller, **MG. Morshed**, L. Rehm, A. W. Ghosh, A. D. Kent, S. Rakheja, "A True Random Number Generator for Probabilistic Computing using Stochastic Magnetic Actuated Random Transducer Devices," *24th International Symposium on Quality Electronic Design (ISQED)*, San Francisco, CA, USA, 2023, pp. 1-10.

C.4 Conference abstracts & presentations

1. **MG. Morshed**, L. Rehm, A. Shukla, S. Ganguly, S. Rakheja, A. D. Kent, A. W. Ghosh, "Reduced sensitivity to process-voltage-temperature variations in stochastic magnetic actuated random transducer devices," The 68th Annual Conference on Magnetism and Magnetic Materials (MMM), Oct. 30 - Nov. 3, 2023, Dallas, TX, USA [Oral].

2. **MG. Morshed**, L. Rehm, A. Shukla, S. Ganguly, S. Rakheja, A. D. Kent, A. W. Ghosh, "Pulse-activated Probabilistic Switching of Medium Barrier Magnet-based Perpendicular

Magnetic Tunnel Junctions," The 2023 Around-the-Clock Around-the-Globe (AtC-AtG) Magnetics Conference, August 24, 2023, Virtual Conference [Oral].

3. **MG. Morshed**, A. Shukla, L. Rehm, L. Heller, Y. Xie, S. Ganguly, S. Rakheja, A. D. Kent, A. W. Ghosh, "Probabilistic spin-torque switching of perpendicular magnetic tunnel junctions under short current pulses," APS March Meeting 2023, March 5 - 10, 2023, Las Vegas, NV, USA [Oral].

4. **MG. Morshed**, H. Vakili, P. V. Balachandran, and A. W. Ghosh, "Stabilizing skyrmions deterministically for racetrack memory applications," The 67th Annual Conference on Magnetism and Magnetic Materials (MMM), Oct. 31 - Nov. 4, 2022, Minneapolis, MN, USA [Oral].

5. **MG. Morshed**, H. Vakili, and A. W. Ghosh, "Positional Stability of Skyrmions in a Racetrack Memory," Sol-SkyMag 2022, June 27 - July 1, 2022, San Sebastian, Spain [Oral].

6. **MG. Morshed**, H. Vakili, and A. W. Ghosh, "Positional Stability of Skyrmions in a Racetrack for Memory Applications," The 2021 Around-the-Clock Around-the-Globe (AtC-AtG) Magnetics Conference, August 24, 2021, Virtual Conference [Oral].

7. **MG. Morshed**, H. Vakili, and A. W. Ghosh, "Positional Stability of Skyrmions via Pinning Sites in a Racetrack Memory," INTERMAG 2021, April. 26 - 30, 2021, Virtual Conference [Oral].

8. **MG. Morshed**, KH. Khoo, Y. Quessab, J-W. Xu, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh, "Tunable Dzyaloshinskii-Moriya Interaction in Ferrimagnetic GdCo alloy through Interface Engineering," APS March Meeting 2021, March 15 - 19, 2021, Virtual Conference [Oral].

9. S. Ganguly, **MG. Morshed**, and A. W. Ghosh, "Impact of non-idealities on computational networks built from low energy-barrier magnet based stochastic neurons," APS

March Meeting 2021, March 15 - 19, 2021, Virtual Conference [Oral, presented by S. Ganguly].

10. H. Vakili, **MG. Morshed**, and A. W. Ghosh, "Challenges and opportunities for skymionic racetrack devices," APS March Meeting 2021, March 15 - 19, 2021, Virtual Conference [Oral, presented by H. Vakili].

11. **MG. Morshed**, KH. Khoo, Y. Quessab, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh, "Controlling Dzyaloshinskii-Moriya Interaction in Ferrimagnetic GdCo - a First Principles Calculation,", MRS Fall Meeting 2020, Nov. 28 - Dec. 4, 2020, Virtual Conference [Oral].

12. **MG. Morshed**, KH. Khoo, Y. Quessab, J-W. Xu, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh, "Interface Engineering to Control the Dzyaloshinskii-Moriya Interaction in Ferrimagnetic GdCo:ab initio Calculations," The 65th Annual Conference on Magnetism and Magnetic Materials (MMM 2020), Nov. 2 - 6, 2020, Virtual Conference [Oral].

13. **MG. Morshed**, KH. Khoo, Y. Quessab, J-W. Xu, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh, "Effective Tailoring of the Dzyaloshinskii-Moriya Interaction in Ferrimagnetic GdCo through Capping Layer Engineering," The 2020 Around-the-Clock Around-the-Globe (AtC-AtG) Magnetics Conference, August 27, 2020, Virtual Conference [Oral].

Bibliography

[1] K. Kabara and M. Tsunoda, “Perpendicular magnetic anisotropy of Mn_4N films fabricated by reactive sputtering method,” *Journal of Applied Physics*, vol. 117, no. 17, 2015.

[2] W. Zhou, C. Ma, T. Hartnett, P. Balachandran, and S. Poon, “Rare-earth-free ferrimagnetic Mn_4N sub-20 nm thin films as potential high-temperature spintronic material,” *AIP Advances*, vol. 11, no. 1, 2021.

[3] C. T. Ma, T. Q. Hartnett, W. Zhou, P. V. Balachandran, and S. J. Poon, “Tunable magnetic skyrmions in ferrimagnetic Mn_4N,” *Applied Physics Letters*, vol. 119, no. 19, 2021.

[4] W. Zhou, C. T. Ma, and S. J. Poon, “Measurement of the Dzyaloshinskii–Moriya interaction in Mn_4N films that host skyrmions,” *Nanomaterials*, vol. 13, no. 10, p. 1672, 2023.

[5] M. G. Morshed, H. Vakili, and A. W. Ghosh, “Positional Stability of Skyrmions in a Racetrack Memory with Notched Geometry,” *Phys. Rev. Appl.*, vol. 17, p. 064019, Jun 2022.

[6] Y. Quessab, J.-W. Xu, C. T. Ma, W. Zhou, G. A. Riley, J. M. Shaw, H. T. Nembach, S. J. Poon, and A. D. Kent, “Tuning interfacial Dzyaloshinskii-Moriya interactions in thin amorphous ferrimagnetic alloys,” *Sci. Rep.*, vol. 10, pp. 1–8, May 2020.

[7] Y. Quessab, J.-W. Xu, M. G. Morshed, A. W. Ghosh, and A. D. Kent, “Interplay between Spin-Orbit Torques and Dzyaloshinskii-Moriya Interactions in Ferrimagnetic Amorphous Alloys,” *Adv. Sci.*, vol. 8, p. 2100481, Sep 2021.

[8] F. Büttner, I. Lemesh, and G. S. D. Beach, “Theory of isolated magnetic skyrmions: From fundamentals to room temperature applications - Scientific Reports,” *Sci. Rep.*, vol. 8, pp. 1–12, Mar 2018.

[9] H. Vakili, J.-W. Xu, W. Zhou, M. N. Sakib, M. G. Morshed, T. Hartnett, Y. Quessab, K. Litzius, C. T. Ma, S. Ganguly, M. R. Stan, P. V. Balachandran, G. S. D. Beach, S. J. Poon, A. D. Kent, and A. W. Ghosh, “Skyrmionics—Computing and memory technologies based on topological excitations in magnets,” *J. Appl. Phys.*, vol. 130, p. 070908, Aug 2021.

[10] Z. He, Y. Zhang, S. Angizi, B. Gong, and D. Fan, "Exploring a SOT-MRAM Based In-Memory Computing for Data Processing," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 4, pp. 676–685, May 2018.

[11] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molna'r, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, "Spintronics: A Spin-Based Electronics Vision for the Future," *Science*, vol. 294, pp. 1488–1495, Nov. 2001.

[12] I. Žutić, J. Fabian, and S. Das Sarma, "Spintronics: Fundamentals and applications," *Rev. Mod. Phys.*, vol. 76, pp. 323–410, Apr. 2004.

[13] S. A. Wolf, A. Y. Chtchelkanova, and D. M. Treger, "Spintronics—A retrospective and perspective," *IBM J. Res. Dev.*, vol. 50, pp. 101–110, Jan. 2006.

[14] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nat. Nanotechnol.*, vol. 10, pp. 187–191, Mar. 2015.

[15] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. N. Piramanayagam, "Spintronics based random access memory: a review," *Mater. Today*, vol. 20, pp. 530–548, Nov. 2017.

[16] Y. Huai *et al.*, "Spin-transfer torque mram (stt-mram): Challenges and prospects," *AAPPS bulletin*, vol. 18, no. 6, pp. 33–40, 2008.

[17] W. S. Zhao, Y. Zhang, T. Devolder, J. O. Klein, D. Ravelosona, C. Chappert, and P. Mazoyer, "Failure and reliability analysis of STT-MRAM," *Microelectron. Reliab.*, vol. 52, pp. 1848–1852, Sept. 2012.

[18] Q. Shao, P. Li, L. Liu, H. Yang, S. Fukami, A. Razavi, H. Wu, K. Wang, F. Freimuth, Y. Mokrousov, M. D. Stiles, S. Emori, A. Hoffmann, J. Åkerman, K. Roy, J.-P. Wang, S.-H. Yang, K. Garello, and W. Zhang, "Roadmap of Spin–Orbit Torques," *IEEE Trans. Magn.*, vol. 57, pp. 1–39, May 2021.

[19] S. Matsunaga, J. Hayakawa, S. Ikeda, K. Miura, H. Hasegawa, T. Endoh, H. Ohno, and T. Hanyu, "Fabrication of a Nonvolatile Full Adder Based on Logic-in-Memory Architecture Using Magnetic Tunnel Junctions," *Appl. Phys. Express*, vol. 1, p. 091301, Aug. 2008.

[20] P. Barla, V. K. Joshi, and S. Bhat, "A Novel Low Power and Reduced Transistor Count Magnetic Arithmetic Logic Unit Using Hybrid STT-MTJ/CMOS Circuit," *IEEE Access*, vol. 8, pp. 6876–6889, Jan. 2020.

[21] C. Safranski, J. Kaiser, P. Trouilloud, P. Hashemi, G. Hu, and J. Z. Sun, "Demonstration of Nanosecond Operation in Stochastic Magnetic Tunnel Junctions," *Nano Lett.*, vol. 21, pp. 2040–2045, Mar. 2021.

[22] A. Fert, N. Reyren, and V. Cros, "Magnetic skyrmions: advances in physics and potential applications - Nature Reviews Materials," *Nat. Rev. Mater.*, vol. 2, pp. 1–15, Jun 2017.

[23] L. Caretta, M. Mann, F. Büttner, K. Ueda, B. Pfau, C. M. Günther, P. Hessing, A. Churikova, C. Klose, M. Schneider, D. Engel, C. Marcus, D. Bono, K. Bagschik, S. Eisebitt, and G. S. D. Beach, "Fast current-driven domain walls and small skyrmions in a compensated ferrimagnet - Nature Nanotechnology," *Nat. Nanotechnol.*, vol. 13, pp. 1154–1160, Dec 2018.

[24] Y. Quessab, J.-W. Xu, E. Cogulu, S. Finizio, J. Raabe, and A. D. Kent, "Zero-Field Nucleation and Fast Motion of Skyrmions Induced by Nanosecond Current Pulses in a Ferrimagnetic Thin Film," *Nano Lett.*, vol. 22, pp. 6091–6097, Aug. 2022.

[25] H. Vakili, M. N. Sakib, S. Ganguly, M. Stan, M. W. Daniels, A. Madhavan, M. D. Stiles, and A. W. Ghosh, "Temporal Memory With Magnetic Racetracks," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 6, pp. 107–115, Sep 2020.

[26] D. Vodenicarevic, N. Locatelli, A. Mizrahi, J. S. Friedman, A. F. Vincent, M. Romera, A. Fukushima, K. Yakushiji, H. Kubota, S. Yuasa, S. Tiwari, J. Grollier, and D. Querlioz, "Low-Energy Truly Random Number Generation with Superparamagnetic Tunnel Junctions for Unconventional Computing," *Phys. Rev. Appl.*, vol. 8, p. 054045, Nov. 2017.

[27] K. Y. Camsari, R. Faria, B. M. Sutton, and S. Datta, "Stochastic p-Bits for Invertible Logic," *Phys. Rev. X*, vol. 7, p. 031014, Jul 2017.

[28] K. Hayakawa, S. Kanai, T. Funatsu, J. Igarashi, B. Jinnai, W. A. Borders, H. Ohno, and S. Fukami, "Nanosecond Random Telegraph Noise in In-Plane Magnetic Tunnel Junctions," *Phys. Rev. Lett.*, vol. 126, p. 117202, Mar. 2021.

[29] K. Y. Camsari, B. M. Sutton, and S. Datta, "p-bits for probabilistic spin logic," *Appl. Phys. Rev.*, vol. 6, Mar. 2019.

[30] S. Ganguly, K. Y. Camsari, and A. W. Ghosh, "Analog Signal Processing Using Stochastic Magnets," *IEEE Access*, vol. 9, pp. 92640–92650, Apr. 2021.

[31] L. Rehm, C. C. M. Capriata, S. Misra, J. D. Smith, M. Pinarbasi, B. G. Malm, and A. D. Kent, "Stochastic Magnetic Actuated Random Transducer Devices Based on Perpendicular Magnetic Tunnel Junctions," *Phys. Rev. Appl.*, vol. 19, p. 024035, Feb. 2023.

[32] K. Roy, S. Bandyopadhyay, and J. Atulasimha, "Energy dissipation and switching delay in stress-induced switching of multiferroic nanomagnets in the presence of thermal fluctuations," *J. Appl. Phys.*, vol. 112, July 2012.

[33] X. Duan, X.-L. Li, X. Li, Y. G. Semenov, and K. W. Kim, "Highly efficient conductance control in a topological insulator based magnetoelectric transistor," *J. Appl. Phys.*, vol. 118, Dec. 2015.

[34] H. Vakili, S. Ganguly, G. J. de Coster, M. R. Neupane, and A. W. Ghosh, "Low Power In-Memory Computation with Reciprocal Ferromagnet/Topological Insulator Heterostructures," *ACS Nano*, vol. 16, pp. 20222–20228, Dec. 2022.

[35] A. Fert, V. Cros, and J. Sampaio, "Skyrmions on the track," *Nat. Nanotechnol.*, vol. 8, pp. 152–156, Mar 2013.

[36] W. Koshibae, Y. Kaneko, J. Iwasaki, M. Kawasaki, Y. Tokura, and N. Nagaosa, "Memory functions of magnetic skyrmions," *Jpn. J. Appl. Phys.*, vol. 54, p. 053001, Apr 2015.

[37] X. Zhang, M. Ezawa, and Y. Zhou, "Magnetic skyrmion logic gates: conversion, duplication and merging of skyrmions," *Sci. Rep.*, vol. 5, pp. 1–8, Mar 2015.

[38] M. N. Sakib, H. Vakili, S. Ganguly, S. Mosanu, A. W. Ghosh, and M. Stan, "Magnetic skyrmion-based programmable hardware," in *Spintronics XIII*, vol. 11470, pp. 129 – 140, SPIE, 2020.

[39] H. Vakili, J.-W. Xu, W. Zhou, M. N. Sakib, M. G. Morshed, T. Hartnett, Y. Quessab, K. Litzius, C. T. Ma, S. Ganguly, *et al.*, "Skyrmionics–Computing and memory technologies based on topological excitations in magnets," *Journal of Applied Physics*, vol. 130, no. 7, p. 070908, 2021.

[40] M. G. Morshed, K. H. Khoo, Y. Quessab, J.-W. Xu, R. Laskowski, P. V. Balachandran, A. D. Kent, and A. W. Ghosh, "Tuning Dzyaloshinskii-Moriya interaction in ferrimagnetic GdCo: A first-principles approach," *Phys. Rev. B*, vol. 103, p. 174414, May 2021.

[41] I. Dzyaloshinsky, "A thermodynamic theory of "weak" ferromagnetism of antiferromagnetics," *J. Phys. Chem. Solids*, vol. 4, pp. 241–255, Jan 1958.

[42] T. Moriya, "Anisotropic Superexchange Interaction and Weak Ferromagnetism," *Phys. Rev.*, vol. 120, pp. 91–98, Oct 1960.

[43] S. Heinze, K. von Bergmann, M. Menzel, J. Brede, A. Kubetzka, R. Wiesendanger, G. Bihlmayer, and S. Blügel, "Spontaneous atomic-scale magnetic skyrmion lattice in two dimensions," *Nat. Phys.*, vol. 7, pp. 713–718, Jul 2011.

[44] C. Moreau-Luchaire, C. Moutafis, N. Reyren, J. Sampaio, C. A. F. Vaz, N. Van Horne, K. Bouzehouane, K. Garcia, C. Deranlot, P. Warnicke, P. Wohlhüter, J.-M. George, M. Weigand, J. Raabe, V. Cros, and A. Fert, "Additive interfacial chiral interaction in multilayers for stabilization of small individual skyrmions at room temperature," *Nat. Nanotechnol.*, vol. 11, pp. 444–448, May 2016.

[45] A. Soumyanarayanan, M. Raju, A. L. Gonzalez Oyarce, A. K. C. Tan, M.-Y. Im, A. P. Petrović, P. Ho, K. H. Khoo, M. Tran, C. K. Gan, F. Ernult, and C. Panagopoulos, "Tunable room-temperature magnetic skyrmions in Ir/Fe/Co/Pt multilayers," *Nat. Mater.*, vol. 16, pp. 898–904, Jul 2017.

[46] X. S. Wang, H. Y. Yuan, and X. R. Wang, "A theory on skyrmion size," *Commun. Phys.*, vol. 1, pp. 1–7, Jul 2018.

[47] J. Sampaio, V. Cros, S. Rohart, A. Thiaville, and A. Fert, "Nucleation, stability and current-induced motion of isolated magnetic skyrmions in nanostructures - Nature Nanotechnology," *Nat. Nanotechnol.*, vol. 8, pp. 839–844, Nov 2013.

[48] H. Yang, A. Thiaville, S. Rohart, A. Fert, and M. Chshiev, "Anatomy of dzyaloshinskii-moriya interaction at Co/Pt interfaces," *Phys. Rev. Lett.*, vol. 115, p. 267210, Dec 2015.

[49] A. Belabbes, G. Bihlmayer, F. Bechstedt, S. Blügel, and A. Manchon, "Hund's Rule-Driven Dzyaloshinskii-Moriya Interaction at $3d-5d$ Interfaces," *Phys. Rev. Lett.*, vol. 117, p. 247202, Dec 2016.

[50] V. Kashid, T. Schena, B. Zimmermann, Y. Mokrousov, S. Blügel, V. Shah, and H. G. Salunke, "Dzyaloshinskii-Moriya interaction and chiral magnetism in $3d - 5d$ zigzag chains: Tight-binding model and ab initio calculations," *Phys. Rev. B*, vol. 90, p. 054412, Aug 2014.

[51] P. Jadaun, L. F. Register, and S. K. Banerjee, "The microscopic origin of DMI in magnetic bilayers and prediction of giant DMI in new bilayers," *npj Comput. Mater.*, vol. 6, pp. 1–6, Jul 2020.

[52] O. Boulle, J. Vogel, H. Yang, S. Pizzini, D. de Souza Chaves, A. Locatelli, T. O. Menteş, A. Sala, L. D. Buda-Prejbeanu, O. Klein, M. Belmeguenai, Y. Roussigné, A. Stashkevich, S. M. Chérif, L. Aballe, M. Foerster, M. Chshiev, S. Auffret, I. M. Miron, and G. Gaudin, "Room-temperature chiral magnetic skyrmions in ultrathin magnetic nanostructures," *Nat. Nanotechnol.*, vol. 11, pp. 449–454, Jan 2016.

[53] S. Tacchi, R. E. Troncoso, M. Ahlberg, G. Gubbiotti, M. Madami, J. Åkerman, and P. Landeros, "Interfacial Dzyaloshinskii-Moriya Interaction in Pt/CoFeB Films: Effect of the Heavy-Metal Thickness," *Phys. Rev. Lett.*, vol. 118, p. 147201, Apr 2017.

[54] H. Yang, O. Boulle, V. Cros, A. Fert, and M. Chshiev, "Controlling Dzyaloshinskii-Moriya Interaction via Chirality Dependent Atomic-Layer Stacking, Insulator Capping and Electric Field," *Sci. Rep.*, vol. 8, pp. 1–7, Aug 2018.

[55] B. Zimmermann, W. Legrand, D. Maccariello, N. Reyren, V. Cros, S. Blügel, and A. Fert, "Dzyaloshinskii-Moriya interaction at disordered interfaces from ab initio theory: Robustness against intermixing and tunability through dusting," *Appl. Phys. Lett.*, vol. 113, p. 232403, Dec 2018.

[56] S. A. Siddiqui, J. Han, J. T. Finley, C. A. Ross, and L. Liu, "Current-Induced Domain Wall Motion in a Compensated Ferrimagnet," *Phys. Rev. Lett.*, vol. 121, p. 057701, Jul 2018.

[57] C. T. Ma, Y. Xie, H. Sheng, A. W. Ghosh, and S. J. Poon, "Robust Formation of Ultrasmall Room-Temperature Neél Skyrmions in Amorphous Ferrimagnets from Atomistic Simulations," *Sci. Rep.*, vol. 9, pp. 1–10, Jul 2019.

[58] D.-H. Kim, M. Haruta, H.-W. Ko, G. Go, H.-J. Park, T. Nishimura, D.-Y. Kim, T. Okuno, Y. Hirata, Y. Futakawa, H. Yoshikawa, W. Ham, S. Kim, H. Kurata, A. Tsukamoto, Y. Shiota, T. Moriyama, S.-B. Choe, K.-J. Lee, and T. Ono, "Bulk Dzyaloshinskii–Moriya interaction in amorphous ferrimagnetic alloys," *Nat. Mater.*, vol. 18, pp. 685–690, May 2019.

[59] S. J. Poon and C. T. Ma, "Amorphous Ferrimagnets: an Ideal Host for Ultra-Small Skyrmions," *J. Supercond. Novel Magn.*, vol. 33, pp. 269–273, Jan 2020.

[60] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Phys. Rev. B*, vol. 54, pp. 11169–11186, Oct 1996.

[61] P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B*, vol. 50, pp. 17953–17979, Dec 1994.

[62] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Phys. Rev. B*, vol. 59, pp. 1758–1775, Jan 1999.

[63] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct 1996.

[64] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, "First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA+ U method," *J. Phys.: Condens. Matter*, vol. 9, pp. 767–808, Jan 1997.

[65] A. B. Shick, W. E. Pickett, and C. S. Fadley, "Electron correlation effects and magnetic ordering at the Gd(0001) surface," *Phys. Rev. B*, vol. 61, pp. R9213–R9216, Apr 2000.

[66] P. Kurz, G. Bihlmayer, and S. Blügel, "Magnetism and electronic structure of hcp Gd and the Gd(0001) surface," *J. Phys.: Condens. Matter*, vol. 14, pp. 6353–6371, Jun 2002.

[67] M. Petersen, J. Hafner, and M. Marsman, "Structural, electronic and magnetic properties of Gd investigated by DFT+U methods: bulk, clean and H-covered (0001) surfaces," *J. Phys.: Condens. Matter*, vol. 18, pp. 7021–7043, Jul 2006.

[68] T. Nozaki, A. Kozioł-Rachwał, M. Tsujikawa, Y. Shiota, X. Xu, T. Ohkubo, T. Tsukahara, S. Miwa, M. Suzuki, S. Tamaru, H. Kubota, A. Fukushima, K. Hono, M. Shirai, Y. Suzuki, and S. Yuasa, "Highly efficient voltage control of spin and enhanced interfacial perpendicular magnetic anisotropy in iridium-doped Fe/MgO magnetic tunnel junctions," *NPG Asia Mater.*, vol. 9, Dec 2017.

[69] S. Woo, K. M. Song, H.-S. Han, M.-S. Jung, M.-Y. Im, K.-S. Lee, K. S. Song, P. Fischer, J.-I. Hong, J. W. Choi, B.-C. Min, H. C. Koo, and J. Chang, "Spin-orbit torque-driven skyrmion dynamics revealed by time-resolved X-ray microscopy," *Nat. Commun.*, vol. 8, pp. 1–8, May 2017.

[70] C.-F. Pai, L. Liu, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin transfer torque devices utilizing the giant spin Hall effect of tungsten," *Appl. Phys. Lett.*, vol. 101, Sept. 2012.

[71] L. Liu, C.-F. Pai, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin-Torque Switching with the Giant Spin Hall Effect of Tantalum," *Science*, vol. 336, pp. 555–558, May 2012.

[72] X. Ma, G. Yu, C. Tang, X. Li, C. He, J. Shi, K. L. Wang, and X. Li, "Interfacial Dzyaloshinskii-Moriya Interaction: Effect of 5d Band Filling and Correlation with Spin Mixing Conductance," *Phys. Rev. Lett.*, vol. 120, p. 157204, Apr 2018.

[73] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii, and P. Böni, "Skyrmion Lattice in a Chiral Magnet," *Science*, vol. 323, pp. 915–919, Feb 2009.

[74] H. Vakili, Y. Xie, and A. W. Ghosh, "Self-focusing hybrid skyrmions in spatially varying canted ferromagnetic systems," *Phys. Rev. B*, vol. 102, p. 174420, Nov 2020.

[75] S.-G. Je, H.-S. Han, S. K. Kim, S. A. Montoya, W. Chao, I.-S. Hong, E. E. Fullerton, K.-S. Lee, K.-J. Lee, M.-Y. Im, and J.-I. Hong, "Direct Demonstration of Topological Stability of Magnetic Skyrmions via Topology Manipulation," *ACS Nano*, vol. 14, pp. 3251–3258, Mar 2020.

[76] Y. Liu, N. Lei, C. Wang, X. Zhang, W. Kang, D. Zhu, Y. Zhou, X. Liu, Y. Zhang, and W. Zhao, "Voltage-Driven High-Speed Skyrmion Motion in a Skyrmion-Shift Device," *Phys. Rev. Appl.*, vol. 11, p. 014004, Jan 2019.

[77] N. Romming, C. Hanneken, M. Menzel, J. E. Bickel, B. Wolter, K. von Bergmann, A. Kubetzka, and R. Wiesendanger, "Writing and Deleting Single Magnetic Skyrmions," *Science*, vol. 341, pp. 636–639, Aug 2013.

[78] S. Luo, M. Song, X. Li, Y. Zhang, J. Hong, X. Yang, X. Zou, N. Xu, and L. You, "Reconfigurable Skyrmion Logic Gates," *Nano Lett.*, vol. 18, pp. 1180–1184, Feb 2018.

[79] M. G. Mankalale, Z. Zhao, J.-P. Wang, and S. S. Sapatnekar, "SkyLogic—A Proposal for a Skyrmion-Based Logic Device," *IEEE Trans. Electron Devices*, vol. 66, pp. 1990–1996, Mar 2019.

[80] S. Luo and L. You, "Skyrmion devices for memory and logic applications," *APL Mater.*, vol. 9, p. 050901, May 2021.

[81] S. S. P. Parkin, M. Hayashi, and L. Thomas, "Magnetic Domain-Wall Racetrack Memory," *Science*, vol. 320, pp. 190–194, Apr 2008.

[82] R. Tomasello, E. Martinez, R. Zivieri, L. Torres, M. Carpentieri, and G. Finocchio, "A strategy for the design of skyrmion racetrack memories - Scientific Reports," *Sci. Rep.*, vol. 4, pp. 1–7, Oct 2014.

[83] B. Göbel and I. Mertig, "Skyrmion ratchet propagation: utilizing the skyrmion Hall effect in AC racetrack storage devices," *Sci. Rep.*, vol. 11, pp. 1–11, Feb 2021.

[84] A. Madhavan, T. Sherwood, and D. Strukov, "Race logic: a hardware acceleration for dynamic programming algorithms," in *ACM SIGARCH Computer Architecture News*, vol. 42, pp. 517–528, New York, NY, USA: Association for Computing Machinery, Jun 2014.

[85] D. Cortés-Ortuño, W. Wang, M. Beg, R. A. Pepper, M.-A. Bisotti, R. Carey, M. Vousden, T. Kluyver, O. Hovorka, and H. Fangohr, "Thermal stability and topological protection of skyrmions in nanotracks," *Scientific Reports*, vol. 7, p. 4060, Jun 2017.

[86] L. Zhao, Z. Wang, X. Zhang, X. Liang, J. Xia, K. Wu, H.-A. Zhou, Y. Dong, G. Yu, K. L. Wang, X. Liu, Y. Zhou, and W. Jiang, "Topology-Dependent Brownian Gyromotion of a Single Skyrmion," *Phys. Rev. Lett.*, vol. 125, p. 027206, Jul 2020.

[87] Y.-H. Liu and Y.-Q. Li, "A mechanism to pin skyrmions in chiral magnets," *J. Phys.: Condens. Matter*, vol. 25, p. 076005, Jan 2013.

[88] J. Müller and A. Rosch, "Capturing of a magnetic skyrmion with a hole," *Phys. Rev. B*, vol. 91, p. 054410, Feb 2015.

[89] H. C. Choi, S.-Z. Lin, and J.-X. Zhu, "Density functional theory study of skyrmion pinning by atomic defects in MnSi," *Phys. Rev. B*, vol. 93, p. 115112, Mar 2016.

[90] C. Reichhardt, C. J. O. Reichhardt, and M. V. Milosevic, "Statics and Dynamics of Skyrmions Interacting with Pinning: A Review," *arXiv*, Feb 2021.

[91] C. Jin, C. Song, J. Wang, and Q. Liu, "Dynamics of antiferromagnetic skyrmion driven by the spin Hall effect," *Appl. Phys. Lett.*, vol. 109, p. 182404, Oct 2016.

[92] J. Iwasaki, M. Mochizuki, and N. Nagaosa, "Current-induced skyrmion dynamics in constricted geometries - Nature Nanotechnology," *Nat. Nanotechnol.*, vol. 8, pp. 742–747, Oct 2013.

[93] X. Zhang, G. P. Zhao, H. Fangohr, J. P. Liu, W. X. Xia, J. Xia, and F. J. Morvan, "Skyrmion-skyrmion and skyrmion-edge repulsions in skyrmion-based racetrack memory," *Sci. Rep.*, vol. 5, pp. 1–6, Jan 2015.

[94] H. Xia, C. Jin, C. Song, J. Wang, J. Wang, and Q. Liu, "Control and manipulation of antiferromagnetic skyrmions in racetrack," *J. Phys. D: Appl. Phys.*, vol. 50, p. 505005, Nov 2017.

[95] J. Leliaert, P. Gypens, M. V. Milošević, B. Van Waeyenberge, and J. Mulkers, "Coupling of the skyrmion velocity to its breathing mode in periodically notched nanotracks," *J. Phys. D: Appl. Phys.*, vol. 52, p. 024003, Nov 2018.

[96] V. M. Uzdin, M. N. Potkina, I. S. Lobanov, P. F. Bessarab, and H. Jónsson, "The effect of confinement and defects on the thermal stability of skyrmions," *Physica B*, vol. 549, pp. 6–9, Nov 2018.

[97] D. Suess, C. Vogler, F. Bruckner, P. Heistracher, F. Slanovc, and C. Abert, "Spin Torque Efficiency and Analytic Error Rate Estimates of Skyrmion Racetrack Memory," *Sci. Rep.*, vol. 9, pp. 1–17, Mar 2019.

[98] Y. Huang, W. Kang, X. Zhang, Y. Zhou, and W. Zhao, "Magnetic skyrmion-based synaptic devices," *Nanotechnology*, vol. 28, p. 08LT02, Jan 2017.

[99] T. Bhattacharya, S. Li, Y. Huang, W. Kang, W. Zhao, and M. Suri, "Low-Power (1T1N) Skyrmionic Synapses for Spiking Neuromorphic Systems," *IEEE Access*, vol. 7, pp. 5034–5044, Jan 2019.

[100] R. Chen, C. Li, Y. Li, J. J. Miles, G. Indiveri, S. Furber, V. F. Pavlidis, and C. Moutafis, "Nanoscale Room-Temperature Multilayer Skyrmionic Synapse for Deep Spiking Neural Networks," *Phys. Rev. Appl.*, vol. 14, p. 014096, Jul 2020.

[101] S. Liu, T. P. Xiao, C. Cui, J. A. C. Incorvia, C. H. Bennett, and M. J. Marinella, "A domain wall-magnetic tunnel junction artificial synapse with notched geometry for accurate and efficient training of deep neural networks," *Appl. Phys. Lett.*, vol. 118, p. 202405, May 2021.

[102] A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, and B. Van Waeyenberge, "The design and verification of MuMax3," *AIP Adv.*, vol. 4, p. 107133, Oct 2014.

[103] W. E, W. Ren, and E. Vanden-Eijnden, "String method for the study of rare events," *Phys. Rev. B*, vol. 66, p. 052301, Aug 2002.

[104] W. E, W. Ren, and E. Vanden-Eijnden, "Simplified and improved string method for computing the minimum energy paths in barrier-crossing events," *J. Chem. Phys.*, vol. 126, p. 164103, Apr 2007.

[105] V. P. Kravchuk, D. D. Sheka, U. K. Rößler, J. van den Brink, and Y. Gaididei, "Spin eigenmodes of magnetic skyrmions and the problem of the effective skyrmion mass," *Phys. Rev. B*, vol. 97, p. 064403, Feb 2018.

[106] B. F. McKeever, D. R. Rodrigues, D. Pinna, Ar. Abanov, J. Sinova, and K. Everschor-Sitte, "Characterizing breathing dynamics of magnetic skyrmions and antiskyrmions within the Hamiltonian formalism," *Phys. Rev. B*, vol. 99, p. 054430, Feb 2019.

[107] S. Rohart and A. Thiaville, "Skyrmion confinement in ultrathin film nanostructures in the presence of Dzyaloshinskii-Moriya interaction," *Phys. Rev. B*, vol. 88, p. 184422, Nov 2013.

[108] J. Wild, T. N. G. Meier, S. Pöllath, M. Kronseder, A. Bauer, A. Chacon, M. Halder, M. Schowalter, A. Rosenauer, J. Zweck, J. Müller, A. Rosch, C. Pfleiderer, and C. H. Back, "Entropy-limited topological protection of skyrmions," *Sci. Adv.*, vol. 3, p. e1701704, Sep 2017.

[109] L. Desplat, D. Suess, J.-V. Kim, and R. L. Stamps, "Thermal stability of metastable magnetic skyrmions: Entropic narrowing and significance of internal eigenmodes," *Phys. Rev. B*, vol. 98, p. 134407, Oct 2018.

[110] L. Desplat and J.-V. Kim, "Entropy-reduced Retention Times in Magnetic Memory Elements: A Case of the Meyer-Neldel Compensation Rule," *Phys. Rev. Lett.*, vol. 125, p. 107201, Sep 2020.

[111] D. Weller and A. Moser, "Thermal effect limits in ultrahigh-density magnetic recording," *IEEE Trans. Magn.*, vol. 35, pp. 4423–4439, Nov 1999.

[112] E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S. Wang, S. A. Wolf, A. W. Ghosh, J. W. Lu, S. J. Poon, M. Stan, W. H. Butler, S. Gupta, C. K. A. Mewes, T. Mewes, and P. B. Visscher, "Advances and Future Prospects of Spin-Transfer Torque Random Access Memory," *IEEE Trans. Magn.*, vol. 46, pp. 1873–1878, May 2010.

[113] L. Rózsa, E. Simon, K. Palotás, L. Udvardi, and L. Szunyogh, "Complex magnetic phase diagram and skyrmion lifetime in an ultrathin film from atomistic simulations," *Phys. Rev. B*, vol. 93, p. 024417, Jan 2016.

[114] S. Rohart, J. Miltat, and A. Thiaville, "Path to collapse for an isolated N\'eel skyrmion," *Phys. Rev. B*, vol. 93, p. 214412, Jun 2016.

[115] W. Kang, Y. Huang, C. Zheng, W. Lv, N. Lei, Y. Zhang, X. Zhang, Y. Zhou, and W. Zhao, "Voltage Controlled Magnetic Skyrmion Motion for Racetrack Memory," *Sci. Rep.*, vol. 6, pp. 1–11, Mar 2016.

[116] H. T. Fook, W. L. Gan, and W. S. Lew, "Gateable Skyrmion Transport via Field-induced Potential Barrier Modulation," *Sci. Rep.*, vol. 6, pp. 1–8, Feb 2016.

[117] R. Gruber, J. Zázvorka, M. A. Brems, D. R. Rodrigues, T. Dohi, N. Kerber, B. Seng, K. Everschor-Sitte, P. Virnau, and M. Kläui, "Skyrmion pinning energetics in thin film systems," *arXiv*, Jan 2022.

[118] T. Gushi, L. Vila, O. Fruchart, A. Marty, S. Pizzini, J. Vogel, F. Takata, A. Anzai, K. Toko, T. Suemasu, and J.-P. Attané, "Millimeter-sized magnetic domains in perpendicularly magnetized ferrimagnetic Mn4N thin films grown on SrTiO3," *Jpn. J. Appl. Phys.*, vol. 57, p. 120310, Oct 2018.

[119] T. Gushi, M. Jovičević Klug, J. Peña Garcia, S. Ghosh, J.-P. Attané, H. Okuno, O. Fruchart, J. Vogel, T. Suemasu, S. Pizzini, and L. Vila, "Large Current Driven Domain Wall Mobility and Gate Tuning of Coercivity in Ferrimagnetic Mn4N Thin Films," *Nano Lett.*, vol. 19, pp. 8716–8723, Dec 2019.

[120] W. Zhou, C. T. Ma, T. Q. Hartnett, P. V. Balachandran, and S. J. Poon, "Rare-earth-free ferrimagnetic Mn4N sub-20 nm thin films as potential high-temperature spintronic material," *AIP Adv.*, vol. 11, p. 015334, Jan 2021.

[121] C. T. Ma, T. Q. Hartnett, W. Zhou, P. V. Balachandran, and S. J. Poon, “Tunable magnetic skyrmions in ferrimagnetic Mn4N,” *Appl. Phys. Lett.*, vol. 119, p. 192406, Nov 2021.

[122] F. Büttner, I. Lemesh, M. Schneider, B. Pfau, C. M. Günther, P. Hessing, J. Geilhufe, L. Caretta, D. Engel, B. Krüger, J. Viefhaus, S. Eisebitt, and G. S. D. Beach, “Field-free deterministic ultrafast creation of magnetic skyrmions by spin–orbit torques,” *Nat. Nanotechnol.*, vol. 12, pp. 1040–1044, Nov 2017.

[123] COMSOL Multiphysics® v. 5.6. www.comsol.com. COMSOL AB, Stockholm, Sweden.

[124] S. Jain, G. Liu, and D. Gifford, “Information Condensing Active Learning,” *arXiv preprint arXiv:2002.07916*, 2020.

[125] M. O’Neill and C. Ryan, “Grammatical evolution,” *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 4, pp. 349–358, 2001.

[126] J. R. Koza, “Genetic programming as a means for programming computers by natural selection,” *Statistics and computing*, vol. 4, pp. 87–112, 1994.

[127] B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, and W.-J. Yin, “Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts,” *Nature Communications*, vol. 11, p. 3513, Jul 2020.

[128] Y. Wang, N. Wagner, and J. M. Rondinelli, “Symbolic regression in materials science,” *MRS Communications*, vol. 9, no. 3, pp. 793–805, 2019.

[129] S. Sun, R. Ouyang, B. Zhang, and T.-Y. Zhang, “Data-driven discovery of formulas by symbolic regression,” *MRS Bulletin*, vol. 44, no. 7, pp. 559–564, 2019.

[130] R. Carnell, *lhs: Latin Hypercube Samples*, 2022. R package version 1.1.6.

[131] P. V. Balachandran, D. Xue, J. Theiler, J. Hogden, and T. Lookman, “Adaptive strategies for materials design using uncertainties,” *Scientific reports*, vol. 6, no. 1, p. 19660, 2016.

[132] P. V. Balachandran, “Adaptive machine learning for efficient materials design,” *MRS Bulletin*, vol. 45, pp. 579–586, 2020.

[133] A. Castillo and S. R. Kalidindi, “A Bayesian framework for the estimation of the single crystal elastic parameters from spherical indentation stress-strain measurements,” *Frontiers in Materials*, vol. 6, p. 136, 2019.

[134] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, Oct 2001.

[135] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189 – 1232, 2001.

[136] M. G. Morshed, S. Ganguly, and A. W. Ghosh, "A Deep Dive Into the Computational Fidelity of High-Variability Low Energy Barrier Magnet Technology for Accelerating Optimization and Bayesian Problems," *IEEE Magn. Lett.*, vol. 14, p. ArticleSequenceNumber:6100405, May 2023.

[137] M. G. Morshed, S. Ganguly, and A. W. Ghosh, "Choose your tools carefully: a comparative evaluation of deterministic vs. stochastic and binary vs. analog neuron models for implementing emerging computing paradigms," *Front. Nanotechnol.*, vol. 5, p. 1146852, May 2023.

[138] M. G. Morshed, L. Rehm, A. Shukla, Y. Xie, S. Ganguly, S. Rakheja, A. D. Kent, and A. W. Ghosh, "Reduced sensitivity to process, voltage and temperature variations in activated perpendicular magnetic tunnel junctions based stochastic devices," *arXiv*, Oct. 2023.

[139] B. Parks, M. Bapna, J. Igbokwe, H. Almasi, W. Wang, and S. A. Majetich, "Superparamagnetic perpendicular magnetic tunnel junctions for true random number generators," *AIP Adv.*, vol. 8, May 2018.

[140] R. Faria, K. Y. Camsari, and S. Datta, "Low-Barrier Nanomagnets as p-Bits for Spin Logic," *IEEE Magn. Lett.*, vol. 8, p. ArticleSequenceNumber:4105305, Mar. 2017.

[141] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, "Intrinsic optimization using stochastic nanomagnets," *Sci. Rep.*, vol. 7, pp. 1–9, Mar. 2017.

[142] O. Hassan, K. Y. Camsari, and S. Datta, "Voltage-Driven Building Block for Hardware Belief Networks," *IEEE Des. Test*, vol. 36, pp. 15–21, Feb. 2019.

[143] Md. A. Abeed and S. Bandyopadhyay, "Low Energy Barrier Nanomagnet Design for Binary Stochastic Neurons: Design Challenges for Real Nanomagnets With Fabrication Defects," *IEEE Magn. Lett.*, vol. 10, p. ArticleSequenceNumber:4504405, July 2019.

[144] M. A. Abeed and S. Bandyopadhyay, "Sensitivity of the Power Spectra of Thermal Magnetization Fluctuations in Low Barrier Nanomagnets Proposed for Stochastic Computing to In-Plane Barrier Height Variations and Structural Defects," *SPIN*, vol. 10, p. 2050001, Sept. 2019.

[145] J. L. Drobitch and S. Bandyopadhyay, "Reliability and Scalability of p-Bits Implemented With Low Energy Barrier Nanomagnets," *IEEE Magn. Lett.*, vol. 10, p. ArticleSequenceNumber:4510404, Dec. 2019.

[146] G. Strang, *Introduction to Linear Algebra.* Wellesley, MA, USA: Wellesley-Cambridge Press, 2016.

[147] L. Lopez-Diaz, L. Torres, and E. Moro, "Transition from ferromagnetism to superparamagnetism on the nanosecond time scale," *Phys. Rev. B*, vol. 65, p. 224406, May 2002.

[148] P. Debashis, R. Faria, K. Y. Camsari, J. Appenzeller, S. Datta, and Z. Chen, "Experimental demonstration of nanomagnet networks as hardware for ising computing," in *2016 IEEE International Electron Devices Meeting (IEDM)*, pp. 34.3.1–34.3.4, 2016.

[149] K. Y. Camsari, S. Salahuddin, and S. Datta, "Implementing p-bits With Embedded MTJ," *IEEE Electron Device Lett.*, vol. 38, pp. 1767–1770, Oct 2017.

[150] "Big data needs a hardware revolution," Feb. 2018. [Online; accessed 28. Dec. 2022].

[151] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1629–1636, Oct. 1990.

[152] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, "A Survey of Neuromorphic Computing and Neural Networks in Hardware," *arXiv*, May 2017.

[153] D. Marković, A. Mizrahi, D. Querlioz, and J. Grollier, "Physics for neuromorphic computing," *Nat. Rev. Phys.*, vol. 2, pp. 499–510, Sept. 2020.

[154] D. V. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M. Le Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel, I. Valov, G. Milano, C. Ricciardi, S.-J. Liang, F. Miao, M. Lanza, T. J. Quill, S. T. Keene, A. Salleo, J. Grollier, D. Marković, A. Mizrahi, P. Yao, J. J. Yang, G. Indiveri, J. P. Strachan, S. Datta, E. Vianello, A. Valentian, J. Feldmann, X. Li, W. H. P. Pernice, H. Bhaskaran, S. Furber, E. Neftci, F. Scherr, W. Maass, S. Ramaswamy, J. Tapson, P. Panda, Y. Kim, G. Tanaka, S. Thorpe, C. Bartolozzi, T. A. Cleland, C. Posch, S. Liu, G. Panuccio, M. Mahmud, A. N. Mazumder, M. Hosseini, T. Mohsenin, E. Donati, S. Tolu, R. Galeazzi, M. E. Christensen, S. Holm, D. Ielmini, and N. Pryds, "2022 roadmap on neuromorphic computing and engineering," *Neuromorph. Comput. Eng.*, vol. 2, p. 022501, May 2022.

[155] D. Kireev, S. Liu, H. Jin, T. Patrick Xiao, C. H. Bennett, D. Akinwande, and J. A. C. Incorvia, "Metaplastic and energy-efficient biocompatible graphene artificial synaptic transistors for enhanced accuracy neuromorphic computing," *Nat. Commun.*, vol. 13, pp. 1–11, July 2022.

[156] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, A. J. Hudspeth, S. Mack, *et al.*, *Principles of neural science*, vol. 4. McGraw-hill New York, 2000.

[157] L. Squire, D. Berg, F. E. Bloom, S. Du Lac, A. Ghosh, and N. C. Spitzer, *Fundamental neuroscience*. Academic press, 2012.

[158] N. K. Upadhyay, S. Joshi, and J. J. Yang, "Synaptic electronics and neuromorphic computing," *Sci. China Inf. Sci.*, vol. 59, pp. 061404–26, May 2016.

[159] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, "Fully hardware-implemented memristor convolutional neural network," *Nature*, vol. 577, pp. 641–646, Jan. 2020.

[160] Q. Duan, Z. Jing, X. Zou, Y. Wang, K. Yang, T. Zhang, S. Wu, R. Huang, and Y. Yang, "Spiking neurons with spatiotemporal dynamics and gain modulation for monolithically integrated memristive neural networks," *Nat. Commun.*, vol. 11, pp. 1–13, July 2020.

[161] J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, and W. D. Lu, "Temporal data classification and forecasting using a memristor-based reservoir computing system," *Nat. Electron.*, vol. 2, pp. 480–487, Oct. 2019.

[162] J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, "Neuromorphic spintronics," *Nat. Electron.*, vol. 3, pp. 360–370, July 2020.

[163] N. Locatelli, V. Cros, and J. Grollier, "Spin-torque building blocks," *Nat. Mater.*, vol. 13, pp. 11–20, Jan. 2014.

[164] W. Lv, J. Cai, H. Tu, L. Zhang, R. Li, Z. Yuan, G. Finocchio, S. Li, X. Sun, L. Bian, B. Zhang, R. Xiong, and Z. Zeng, "Stochastic artificial synapses based on nanoscale magnetic tunnel junction for neuromorphic applications," *Appl. Phys. Lett.*, vol. 121, p. 232406, Dec. 2022.

[165] H. Liu, T. Wu, X. Yan, J. Wu, N. Wang, Z. Du, H. Yang, B. Chen, Z. Zhang, F. Liu, W. Wu, J. Guo, and H. Wang, "A Tantalum Disulfide Charge-Density-Wave Stochastic Artificial Neuron for Emulating Neural Statistical Properties," *Nano Lett.*, vol. 21, pp. 3465–3472, Apr. 2021.

[166] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," *Nat. Photonics*, vol. 15, pp. 102–114, Feb. 2021.

[167] J. M. Shainline, S. M. Buckley, R. P. Mirin, and S. W. Nam, "Superconducting Optoelectronic Circuits for Neuromorphic Computing," *Phys. Rev. Appl.*, vol. 7, p. 034013, Mar. 2017.

[168] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, pp. 668–673, Aug. 2014.

[169] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, vol. 38, pp. 82–99, Jan. 2018.

[170] S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, "The SpiNNaker Project," *Proc. IEEE*, vol. 102, pp. 652–665, Feb. 2014.

[171] J. Schemmel, D. Brüderle, A. Grübl, M. Hock, K. Meier, and S. Millner, "A wafer-scale neuromorphic hardware system for large-scale neural modeling," in *2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1947–1950, IEEE, May 2010.

[172] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, and L. Shi, "Towards artificial general intelligence with hybrid Tianjic chip architecture," *Nature*, vol. 572, pp. 106–111, Aug. 2019.

[173] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations," *Proc. IEEE*, vol. 102, pp. 699–716, Apr. 2014.

[174] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale Memristor Device as Synapse in Neuromorphic Systems," *Nano Lett.*, vol. 10, pp. 1297–1301, Apr. 2010.

[175] A. Serb, A. Corna, R. George, A. Khiat, F. Rocchi, M. Reato, M. Maschietto, C. Mayr, G. Indiveri, S. Vassanelli, and T. Prodromakis, "Memristive synapses connect brain and silicon spiking neurons," *Sci. Rep.*, vol. 10, pp. 1–7, Feb. 2020.

[176] G. Innocenti, M. Di Marco, A. Tesi, and M. Forti, "Memristor Circuits for Simulating Neuron Spiking and Burst Phenomena," *Front. Neurosci.*, vol. 15, June 2021.

[177] A. Mehonic and A. J. Kenyon, "Emulating the Electrical Activity of the Neuron Using a Silicon Oxide RRAM Cell," *Front. Neurosci.*, vol. 10, Feb. 2016.

[178] G. C. Adam, B. D. Hoskins, M. Prezioso, F. Merrikh-Bayat, B. Chakrabarti, and D. B. Strukov, "3-D Memristor Crossbars for Analog and Neuromorphic Computing Applications," *IEEE Trans. Electron Devices*, vol. 64, pp. 312–318, Dec. 2016.

[179] M. Hu, H. Li, Y. Chen, Q. Wu, G. S. Rose, and R. W. Linderman, "Memristor Crossbar-Based Neuromorphic Computing System: A Case Study," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 25, pp. 1864–1878, Jan. 2014.

[180] M. Suri, V. Parmar, A. Kumar, D. Querlioz, and F. Alibart, "Neuromorphic hybrid RRAM-CMOS RBM architecture," in *2015 15th Non-Volatile Memory Technology Symposium (NVMTS)*, pp. 1–6, IEEE, Oct. 2015.

[181] A. Sengupta, K. Yogendra, and K. Roy, "Spintronic devices for ultra-low power neuromorphic computation (Special session paper)," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 922–925, IEEE, May 2016.

[182] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, pp. 428–431, July 2017.

[183] K. Roy, M. Sharad, D. Fan, and K. Yogendra, "Brain-inspired computing with spin torque devices," in *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1–6, IEEE, Mar. 2014.

[184] A. Sengupta, P. Panda, A. Raghunathan, and K. Roy, "Neuromorphic Computing Enabled by Spin-Transfer Torque Devices," in *2016 29th International Conference on VLSI Design and 2016 15th International Conference on Embedded Systems (VLSID)*, pp. 32–37, IEEE, Jan. 2016.

[185] S. A. Siddiqui, S. Dutta, A. Tang, L. Liu, C. A. Ross, and M. A. Baldo, "Magnetic Domain Wall Based Synaptic and Activation Function Generator for Neuromorphic Accelerators," *Nano Lett.*, vol. 20, pp. 1033–1040, Feb. 2020.

[186] T. Leonard, S. Liu, M. Alamdar, H. Jin, C. Cui, O. G. Akinola, L. Xue, T. P. Xiao, J. S. Friedman, M. J. Marinella, C. H. Bennett, and J. A. C. Incorvia, "Shape-Dependent Multi-Weight Magnetic Artificial Synapses for Neuromorphic Computing," *Adv. Electron. Mater.*, vol. 8, p. 2200563, Dec. 2022.

[187] W. H. Brigner, N. Hassan, X. Hu, C. H. Bennett, F. Garcia-Sanchez, C. Cui, A. Velasquez, M. J. Marinella, J. A. C. Incorvia, and J. S. Friedman, "Domain Wall Leaky Integrate-and-Fire Neurons With Shape-Based Configurable Activation Functions," *IEEE Trans. Electron Devices*, vol. 69, pp. 2353–2359, Mar. 2022.

[188] P. Jadaun, C. Cui, S. Liu, and J. A. C. Incorvia, "Adaptive cognition implemented with a context-aware and flexible neuron for next-generation artificial intelligence," *PNAS Nexus*, vol. 1, p. pgac206, Nov. 2022.

[189] K. M. Song, J.-S. Jeong, B. Pan, X. Zhang, J. Xia, S. Cha, T.-E. Park, K. Kim, S. Finizio, J. Raabe, J. Chang, Y. Zhou, W. Zhao, W. Kang, H. Ju, and S. Woo, "Skyrmion-based artificial synapses for neuromorphic computing," *Nat. Electron.*, vol. 3, pp. 148–155, Mar. 2020.

[190] B. Romeira, R. Avó, J. M. L. Figueiredo, S. Barland, and J. Javaloyes, "Regenerative memory in time-delayed neuromorphic photonic resonators," *Sci. Rep.*, vol. 6, pp. 1–12, Jan. 2016.

[191] X. Guo, J. Xiang, Y. Zhang, and Y. Su, "Integrated Neuromorphic Photonics: Synapses, Neurons, and Neural Networks," *Adv. Photonics Res.*, vol. 2, p. 2000212, June 2021.

[192] J. J. Yang, D. B. Strukov, and D. R. Stewart, "Memristive devices for computing," *Nat. Nanotechnol.*, vol. 8, pp. 13–24, Jan. 2013.

[193] R. M. Wang, C. S. Thakur, and A. van Schaik, "An FPGA-Based Massively Parallel Neuromorphic Cortex Simulator," *Front. Neurosci.*, vol. 12, Apr. 2018.

[194] E. Farquhar, C. Gordon, and P. Hasler, "A field programmable neural array," in *2006 IEEE International Symposium on Circuits and Systems*, pp. 4pp.–4117, IEEE, May 2006.

[195] B. Rajendran and F. Alibart, "Neuromorphic Computing Based on Emerging Memory Technologies," *IEEE J. Emerging Sel. Top. Circuits Syst.*, vol. 6, pp. 198–211, Mar. 2016.

[196] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nat. Rev. Neurosci.*, vol. 9, pp. 292–303, Apr. 2008.

[197] C. Baldassi, F. Gerace, H. J. Kappen, C. Lucibello, L. Saglietti, E. Tartaglione, and R. Zecchina, "Role of Synaptic Stochasticity in Training Low-Precision Neural Networks," *Phys. Rev. Lett.*, vol. 120, p. 268103, June 2018.

[198] J. Goldberger and E. Ben-Reuven, "Training deep neural-networks using a noise adaptation layer," in *International Conference on Learning Representations*, 2017.

[199] A. F. Vincent, J. Larroque, N. Locatelli, N. B. Romdhane, O. Bichler, C. Gamrat, W. S. Zhao, J.-O. Klein, S. Galdin-Retailleau, and D. Querlioz, "Spin-Transfer Torque Magnetic Memory as a Stochastic Memristive Synapse for Neuromorphic Systems," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, pp. 166–174, Apr. 2015.

[200] S. D. Brown, G. Chakma, M. M. Adnan, M. S. Hasan, and G. S. Rose, "Stochasticity in Neuromorphic Computing: Evaluating Randomness for Improved Performance," in *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 454–457, IEEE, Nov. 2019.

[201] K.-C. Jim, C. L. Giles, and B. G. Horne, "An analysis of noise in recurrent neural networks: convergence and generalization," *IEEE Trans. Neural Networks*, vol. 7, pp. 1424–1438, Nov. 1996.

[202] R. L. Cook, "Stochastic sampling in computer graphics," *ACM Trans. Graphics*, vol. 5, pp. 51–72, Jan. 1986.

[203] C.-W. Huang, J. H. Lim, and A. C. Courville, "A Variational Perspective on Diffusion-Based Generative Models and Score Matching," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22863–22876, Dec. 2021.

[204] A. N. Burkitt, "A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input," *Biol. Cybern.*, vol. 95, pp. 1–19, July 2006.

[205] C. Bick, M. Goodfellow, C. R. Laing, and E. A. Martens, "Understanding the dynamics of biological and neural oscillator networks through exact mean-field reductions: a review," *J. Math. Neurosci.*, vol. 10, pp. 1–43, Dec. 2020.

[206] L. D. Harmon, "Artificial Neuron," *Science*, vol. 129, pp. 962–963, Apr. 1959.

[207] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, "Opportunities for neuromorphic computing algorithms and applications," *Nat. Comput. Sci.*, vol. 2, pp. 10–19, Jan. 2022.

[208] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, p. e00938, Nov. 2018.

[209] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT press, 2016.

[210] S. Davidson and S. B. Furber, "Comparison of Artificial and Spiking Neural Networks on Digital Hardware," *Front. Neurosci.*, vol. 15, Apr. 2021.

[211] G. Barna and K. Kaski, "Stochastic vs. Deterministic Neural Networks for Pattern Recognition," *Phys. Scr.*, vol. 1990, p. 110, Jan. 1990.

[212] T. Szandała, "Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks," in *Bio-inspired Neurocomputing*, pp. 203–224, Singapore: Springer, July 2020.

[213] C. Zhang and P. C. Woodland, "Parameterised sigmoid and relu hidden activation functions for dnn acoustic modelling," in *Interspeech*, 2015.

[214] E. Oostwal, M. Straat, and M. Biehl, "Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation," *Physica A*, vol. 564, p. 125517, Feb. 2021.

[215] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, July 2019.

[216] F. Triefenbach, A. Jalalvand, B. Schrauwen, and J.-p. Martens, "Phoneme recognition with large hierarchical reservoirs," in *Advances in Neural Information Processing Systems* (J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, eds.), vol. 23, Curran Associates, Inc., 2010.

[217] A. Jalalvand, G. Van Wallendael, and R. Van De Walle, "Real-Time Reservoir Computing Network-Based Systems for Detection Tasks on Visual Contents," in *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, pp. 146–151, IEEE, June 2015.

[218] S. Ganguly, Y. Gu, M. R. Stan, and A. W. Ghosh, "Hardware based spatio-temporal neural processing backend for imaging sensors: towards a smart camera," in *Proceedings Volume 10656, Image Sensing Technologies: Materials, Devices, Systems, and Applications V*, vol. 10656, pp. 135–145, SPIE, May 2018.

[219] J. Kato, G. Tanaka, R. Nakane, and A. Hirose, "Proposal of Reconstructive Reservoir Computing to Detect Anomaly in Time-series Signals," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, IEEE, July 2022.

[220] F. Abreu Araujo, M. Riou, J. Torrejon, S. Tsunegi, D. Querlioz, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, "Role of non-linear data processing on speech recognition task in the framework of reservoir computing," *Sci. Rep.*, vol. 10, pp. 1–11, Jan. 2020.

[221] I. Engedy and G. Horváth, "Optimal control with reinforcement learning using reservoir computing and gaussian mixture," in *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1062–1066, IEEE, 2012.

[222] V. Pyragas and K. Pyragas, "Using reservoir computer to predict and prevent extreme events," *Phys. Lett. A*, vol. 384, p. 126591, Aug. 2020.

[223] M. Lukoševičius, "A Practical Guide to Applying Echo State Networks," in *Neural Networks: Tricks of the Trade: Second Edition*, pp. 659–686, Berlin, Germany: Springer, 2012.

[224] D. Li, M. Han, and J. Wang, "Chaotic Time Series Prediction Based on a Novel Robust Echo State Network," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, pp. 787–799, Apr. 2012.

[225] H. Jaeger, "Short term memory in echo state networks. gmd-report 152," *GMD-German National Research Institute for Computer Science (2002)*, 2002.

[226] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Networks*, vol. 20, pp. 391–403, Apr. 2007.

[227] M. Inubushi and K. Yoshimura, "Reservoir Computing Beyond Memory-Nonlinearity Trade-off," *Sci. Rep.*, vol. 7, pp. 1–10, Aug. 2017.

[228] S. Kumar, J. P. Strachan, and R. S. Williams, "Chaotic dynamics in nanoscale NbO2 Mott memristors for analogue computing," *Nature*, vol. 548, pp. 318–321, Aug. 2017.

[229] J. Hochstetter, R. Zhu, A. Loeffler, A. Diaz-Alvarez, T. Nakayama, and Z. Kuncic, "Avalanches and edge-of-chaos learning in neuromorphic nanowire networks," *Nat. Commun.*, vol. 12, pp. 1–13, June 2021.

[230] D. Nishioka, T. Tsuchiya, W. Namiki, M. Takayanagi, M. Imura, Y. Koide, T. Higuchi, and K. Terabe, "Edge-of-chaos learning achieved by ion-electron–coupled dynamics in an ion-gating reservoir," *Sci. Adv.*, vol. 8, p. eade1156, Dec. 2022.

[231] B. Sunar, "True Random Number Generators for Cryptography," in *Cryptographic Engineering*, pp. 55–73, Boston, MA, USA: Springer, Boston, MA, 2009.

[232] H. Bauke and S. Mertens, "Random numbers for large-scale distributed Monte Carlo simulations," *Phys. Rev. E*, vol. 75, p. 066701, June 2007.

[233] S. Misra, L. C. Bland, S. G. Cardwell, J. A. C. Incorvia, C. D. James, A. D. Kent, C. D. Schuman, J. D. Smith, and J. B. Aimone, "Probabilistic Neural Computing with Stochastic Devices," *Adv. Mater.*, vol. 35, p. 2204569, Sept. 2023.

[234] M. W. Daniels, A. Madhavan, P. Talatchian, A. Mizrahi, and M. D. Stiles, "Energy-Efficient Stochastic Computing with Superparamagnetic Tunnel Junctions," *Phys. Rev. Appl.*, vol. 13, p. 034016, Mar. 2020.

[235] R. M. D'Souza, Y. Bar-Yam, and M. Kardar, "Sensitivity of ballistic deposition to pseudorandom number generators," *Phys. Rev. E*, vol. 57, pp. 5044–5052, May 1998.

[236] J. F. Fernández and C. Criado, "Algorithm for normal random numbers," *Phys. Rev. E*, vol. 60, pp. 3361–3365, Sept. 1999.

[237] A. Poorghanad, A. Sadr, and A. Kashanipour, "Generating high quality pseudo random number using evolutionary methods," in *2008 International Conference on Computational Intelligence and Security*, vol. 1, pp. 331–335, 2008.

[238] C. S. Petrie and J. A. Connelly, "A noise-based IC random number generator for applications in cryptography," *IEEE Trans. Circuits Syst. I*, vol. 47, pp. 615–621, May 2000.

[239] R. Brederlow, R. Prakash, C. Paulus, and R. Thewes, "A low-power true random number generator using random telegraph noise of single oxide-traps," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, pp. 1666–1675, 2006.

[240] A. Alkassar, T. Nicolay, and M. Rohe, "Obtaining True-Random Binary Numbers from a Weak Radioactive Source," in *Computational Science and Its Applications – ICCSA 2005*, pp. 634–646, Berlin, Germany: Springer, 2005.

[241] D. Ruschen, M. Schrey, J. Freese, and I. Heisterklaus, "Generation of true random numbers based on radioactive decay," *power*, vol. 3, p. 3V, 2017.

[242] N. Liu, N. Pinckney, S. Hanson, D. Sylvester, and D. Blaauw, "A true random number generator using time-dependent dielectric breakdown," in *2011 Symposium on VLSI Circuits - Digest of Technical Papers*, pp. 216–217, 2011.

[243] S. K. Mathew, S. Srinivasan, M. A. Anders, H. Kaul, S. K. Hsu, F. Sheikh, A. Agarwal, S. Satpathy, and R. K. Krishnamurthy, "2.4 Gbps, 7 mW All-Digital PVT-Variation Tolerant True Random Number Generator for 45 nm CMOS High-Performance Microprocessors," *IEEE J. Solid-State Circuits*, vol. 47, pp. 2807–2821, Oct. 2012.

[244] S. K. Mathew, D. Johnston, S. Satpathy, V. Suresh, P. Newman, M. A. Anders, H. Kaul, A. Agarwal, S. K. Hsu, G. Chen, and R. K. Krishnamurthy, "μ," *IEEE J. Solid-State Circuits*, vol. 51, pp. 1695–1704, May 2016.

[245] A. Fukushima, T. Seki, K. Yakushiji, H. Kubota, H. Imamura, S. Yuasa, and K. Ando, "Spin dice: A scalable truly random number generator based on spintronics," *Appl. Phys. Express*, vol. 7, p. 083001, July 2014.

[246] A. Fukushima, T. Yamamoto, T. Nozaki, K. Yakushiji, H. Kubota, and S. Yuasa, "Recent progress in random number generator using voltage pulse-induced switching of nano-magnet: A perspective," *APL Mater.*, vol. 9, Mar. 2021.

[247] Z. Fu, Y. Tang, X. Zhao, K. Lu, Y. Dong, A. Shukla, Z. Zhu, and Y. Yang, "An Overview of Spintronic True Random Number Generator," *Front. Phys.*, vol. 9, p. 638207, Apr. 2021.

[248] S. Matsunaga, J. Hayakawa, S. Ikeda, K. Miura, T. Endoh, H. Ohno, and T. Hanyu, "Mtj-based nonvolatile logic-in-memory circuit, future prospects and issues," in *2009 Design, Automation & Test in Europe Conference & Exhibition*, pp. 433–435, 2009.

[249] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Phys. Rev. B*, vol. 62, pp. 570–578, July 2000.

[250] Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchula, E. Chen, L.-C. Wang, and Y. Huai, "Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque," *J. Phys.: Condens. Matter*, vol. 19, p. 165209, Apr. 2007.

[251] D. Bedau, H. Liu, J.-J. Bouzaglou, A. D. Kent, J. Z. Sun, J. A. Katine, E. E. Fullerton, and S. Mangin, "Ultrafast spin-transfer switching in spin valve nanopillars with perpendicular anisotropy," *Appl. Phys. Lett.*, vol. 96, Jan. 2010.

[252] W. F. Brown, "Thermal Fluctuations of a Single-Domain Particle," *Phys. Rev.*, vol. 130, pp. 1677–1686, June 1963.

[253] N. Rangarajan, A. Parthasarathy, and S. Rakheja, "A spin-based true random number generator exploiting the stochastic precessional switching of nanomagnets," *J. Appl. Phys.*, vol. 121, June 2017.

[254] Y. Qu, B. F. Cockburn, Z. Huang, H. Cai, Y. Zhang, W. Zhao, and J. Han, "Variation-Resilient True Random Number Generators Based on Multiple STT-MTJs," *IEEE Trans. Nanotechnol.*, vol. 17, pp. 1270–1281, Oct. 2018.

[255] D. Vodenicarevic, N. Locatelli, A. Mizrahi, T. Hirtzlin, J. S. Friedman, J. Grollier, and D. Querlioz, "Circuit-level evaluation of the generation of truly random bits with superparamagnetic tunnel junctions," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–4, 2018.

[256] L. Schnitzspan, M. Kläui, and G. Jakob, "Nanosecond true-random-number generation with superparamagnetic tunnel junctions: Identification of joule heating and spin-transfer-torque effects," *Phys. Rev. Appl.*, vol. 20, p. 024002, Aug 2023.

[257] P. Debashis and Z. Chen, "Tunable random number generation using single superparamagnet with perpendicular magnetic anisotropy," in *2018 76th Device Research Conference (DRC)*, pp. 1–2, 2018.

[258] R. Rahman and S. Bandyopadhyay, "The Strong Sensitivity of the Characteristics of Binary Stochastic Neurons Employing Low Barrier Nanomagnets to Small Geometrical Variations," *IEEE Trans. Nanotechnol.*, vol. 22, pp. 112–119, Feb. 2023.

[259] A. Shukla, L. Heller, M. G. Morshed, L. Rehm, A. W. Ghosh, A. D. Kent, and S. Rakheja, "A true random number generator for probabilistic computing using stochastic magnetic actuated random transducer devices," in *2023 24th International Symposium on Quality Electronic Design (ISQED)*, pp. 1–10, 2023.

[260] N. Kani, S. Rakheja, and A. Naeemi, "A Probability-Density Function Approach to Capture the Stochastic Dynamics of the Nanomagnet and Impact on Circuit Performance," *IEEE Trans. Electron Devices*, vol. 63, pp. 4119–4126, Aug. 2016.

[261] W. H. Butler, T. Mewes, C. K. A. Mewes, P. B. Visscher, W. H. Rippard, S. E. Russek, and R. Heindl, "Switching Distributions for Perpendicular Spin-Torque Devices Within the Macrospin Approximation," *IEEE Trans. Magn.*, vol. 48, pp. 4684–4700, July 2012.

[262] H. Liu, D. Bedau, J. Z. Sun, S. Mangin, E. E. Fullerton, J. A. Katine, and A. D. Kent, "Dynamics of spin torque switching in all-perpendicular spin valve nanopillars," *J. Magn. Magn. Mater.*, vol. 358-359, pp. 233–258, May 2014.

[263] Y. Xie, B. Behin-Aein, and A. W. Ghosh, "Fokker—Planck Study of Parameter Dependence on Write Error Slope in Spin-Torque Switching," *IEEE Trans. Electron Devices*, vol. 64, pp. 319–324, Dec. 2016.

[264] J. Z. Sun, "Spin angular momentum transfer in current-perpendicular nanomagnetic junctions," *IBM J. Res. Dev.*, vol. 50, pp. 81–100, Jan. 2006.

[265] 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, "Effect of subvolume excitation and spin-torque efficiency on magnetic switching," *Phys. Rev. B*, vol. 84, p. 064413, Aug. 2011.

[266] J. Z. Sun, "Spin-transfer torque switched magnetic tunnel junctions in magnetic random access memory," in *Proceedings Volume 9931, Spintronics IX*, vol. 9931, pp. 112–124, SPIE, Sept. 2016.

[267] J. Z. Sun, "Spin-transfer torque switched magnetic tunnel junction for memory technologies," *J. Magn. Magn. Mater.*, vol. 559, p. 169479, Oct. 2022.

[268] J. Cao, Y. Liu, Y. Ren, F. Wei, and P. P. Freitas, "Effect of annealing temperature on formation of superparamagnetism in CoFeB/MgO/CoFeB magnetic tunnel junctions," *Appl. Surf. Sci.*, vol. 314, pp. 443–446, Sept. 2014.

[269] B. Teso, S. Kravenkit, K. Sorn-in, A. Kaewrawang, A. Kruesubthaworn, A. Siritaratiwat, T. Mewes, C. K. A. Mewes, and C. Surawanitkun, "Temperature dependence of magnetic properties on switching energy in magnetic tunnel junction devices with tilted magnetization," *Appl. Surf. Sci.*, vol. 472, pp. 36–39, Apr. 2019.

[270] D. H. Lee and S. H. Lim, "Increase of temperature due to Joule heating during current-induced magnetization switching of an MgO-based magnetic tunnel junction," *Appl. Phys. Lett.*, vol. 92, June 2008.

[271] N. N. Mojumder, D. W. Abraham, K. Roy, and D. C. Worledge, "Magnonic Spin-Transfer Torque MRAM With Low Power, High Speed, and Error-Free Switching," *IEEE Trans. Magn.*, vol. 48, pp. 2016–2024, Dec. 2011.

[272] J. Song, H. Dixit, B. Behin-Aein, C. H. Kim, and W. Taylor, "Impact of Process Variability on Write Error Rate and Read Disturbance in STT-MRAM Devices," *IEEE Trans. Magn.*, vol. 56, p. ArticleSequenceNumber:3400411, Oct. 2020.

[273] K. Yang, D. Fick, M. B. Henry, Y. Lee, D. Blaauw, and D. Sylvester, "16.3 a 23mb/s 23pj/b fully synthesized true-random-number generator in 28nm and 65nm cmos," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 280–281, 2014.

[274] S. Ota, M. Ono, H. Matsumoto, A. Ando, T. Sekitani, R. Kohno, S. Iguchi, T. Koyama, and D. Chiba, "CoFeB/MgO-based magnetic tunnel junction directly formed on a flexible substrate," *Appl. Phys. Express*, vol. 12, p. 053001, Apr. 2019.

[275] A. Roohi, S. Angizi, and D. Fan, "Enabling Edge Computing Using Emerging Memory Technologies: From Device to Architecture," in *Frontiers of Quality Electronic Design (QED): AI, IoT and Hardware Security*, pp. 415–464, Cham, Switzerland: Springer, Sept. 2022.

[276] "Maximizing Edge Intelligence Requires More Than Computing | SEMI," June 2024. [Online; accessed 7. Jun. 2024].

[277] K. Roy, S. Bandyopadhyay, and J. Atulasimha, "Hybrid spintronics and straintronics: A magnetic technology for ultra low energy computing and signal processing," *Appl. Phys. Lett.*, vol. 99, Aug. 2011.

[278] A. K. Biswas, H. Ahmad, J. Atulasimha, and S. Bandyopadhyay, "Experimental Demonstration of Complete 180° Reversal of Magnetization in Isolated Co Nanomagnets on a PMN–PT Substrate with Voltage Generated Strain," *Nano Lett.*, vol. 17, pp. 3478–3484, June 2017.

[279] J.-C. Rojas-Sánchez, S. Oyarzún, Y. Fu, A. Marty, C. Vergnaud, S. Gambarelli, L. Vila, M. Jamet, Y. Ohtsubo, A. Taleb-Ibrahimi, P. Le Fèvre, F. Bertran, N. Reyren, J.-M. George, and A. Fert, "Spin to Charge Conversion at Room Temperature by Spin Pumping into a New Type of Topological Insulator: α-Sn Films," *Phys. Rev. Lett.*, vol. 116, p. 096602, Mar. 2016.

[280] J. M. Marmolejo-Tejada, K. Dolui, P. Lazić, P.-H. Chang, S. Smidstrup, D. Stradi, K. Stokbro, and B. K. Nikolić, "Proximity Band Structure and Spin Textures on Both Sides of Topological-Insulator/Ferromagnetic-Metal Interface and Their Charge Transport Probes," *Nano Lett.*, vol. 17, pp. 5626–5633, Sept. 2017.

[281] A. R. Mellnik, J. S. Lee, A. Richardella, J. L. Grab, P. J. Mintun, M. H. Fischer, A. Vaezi, A. Manchon, E.-A. Kim, N. Samarth, and D. C. Ralph, "Spin-transfer torque generated by a topological insulator," *Nature*, vol. 511, pp. 449–451, July 2014.

[282] Y. Fan, P. Upadhyaya, X. Kou, M. Lang, S. Takei, Z. Wang, J. Tang, L. He, L.-T. Chang, M. Montazeri, G. Yu, W. Jiang, T. Nie, R. N. Schwartz, Y. Tserkovnyak, and K. L. Wang, "Magnetization switching through giant spin–orbit torque in a magnetically doped topological insulator heterostructure," *Nat. Mater.*, vol. 13, pp. 699–704, July 2014.

[283] J. Han, A. Richardella, S. A. Siddiqui, J. Finley, N. Samarth, and L. Liu, "Room-Temperature Spin-Orbit Torque Switching Induced by a Topological Insulator," *Phys. Rev. Lett.*, vol. 119, p. 077702, Aug. 2017.

[284] N. H. D. Khang, Y. Ueda, and P. N. Hai, "A conductive topological insulator with large spin Hall effect for ultralow power spin–orbit torque switching," *Nat. Mater.*, vol. 17, pp. 808–813, Sept. 2018.

[285] D. Patterson, T. Anderson, N. Cardwell, R. Fromm, K. Keeton, C. Kozyrakis, R. Thomas, and K. Yelick, "A case for intelligent RAM," *IEEE Micro*, vol. 17, pp. 34–44, Mar. 1997.

[286] S. Ghose, A. Boroumand, J. S. Kim, J. Gómez-Luna, and O. Mutlu, "Processing-in-memory: A workload-driven perspective," *IBM J. Res. Dev.*, vol. 63, pp. 1–3, Aug. 2019.

[287] N. Verma, H. Jia, H. Valavi, Y. Tang, M. Ozatay, L.-Y. Chen, B. Zhang, and P. Deaville, "In-Memory Computing: Advances and Prospects," *IEEE Solid-State Circuits Mag.*, vol. 11, pp. 43–55, Aug. 2019.

[288] P. Mannocci, M. Farronato, N. Lepri, L. Cattaneo, A. Glukhov, Z. Sun, and D. Ielmini, "In-memory computing with emerging memory devices: Status and outlook," *APL Mach. Learn.*, vol. 1, Mar. 2023.

[289] W. A. Wulf and S. A. McKee, "Hitting the memory wall: Implications of the obvious," *ACM SIGARCH computer architecture news*, vol. 23, no. 1, pp. 20–24, 1995.

[290] X. Zou, S. Xu, X. Chen, L. Yan, and Y. Han, "Breaking the von neumann bottleneck: architecture-level processing-in-memory technology," *Science China Information Sciences*, vol. 64, no. 6, p. 160404, 2021.

[291] O. Mutlu, S. Ghose, J. Gómez-Luna, and R. Ausavarungnirun, "Processing data where it makes sense: Enabling in-memory computation," *Microprocess. Microsyst.*, vol. 67, pp. 28–41, June 2019.

[292] F. Parveen, S. Angizi, Z. He, and D. Fan, "Low power in-memory computing based on dual-mode sot-mram," in *2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 1–6, 2017.

[293] J. Lee, B.-G. Park, and Y. Kim, "Implementation of Boolean Logic Functions in Charge Trap Flash for In-Memory Computing," *IEEE Electron Device Lett.*, vol. 40, pp. 1358–1361, July 2019.

[294] X. Xie, Z. Liang, P. Gu, A. Basak, L. Deng, L. Liang, X. Hu, and Y. Xie, "Spacea: Sparse matrix vector multiplication on processing-in-memory accelerator," in *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pp. 570–583, 2021.

[295] D. Fan, S. Angizi, and Z. He, "In-memory computing with spintronic devices," in *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pp. 683–688, 2017.

[296] L. Liu, O. J. Lee, T. J. Gudmundsen, D. C. Ralph, and R. A. Buhrman, "Current-Induced Switching of Perpendicularly Magnetized Magnetic Layers Using Spin Torque from the Spin Hall Effect," *Phys. Rev. Lett.*, vol. 109, p. 096602, Aug. 2012.

[297] I. M. Miron, K. Garello, G. Gaudin, P.-J. Zermatten, M. V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, and P. Gambardella, "Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection," *Nature*, vol. 476, pp. 189–193, Aug. 2011.

[298] G. Prenat, K. Jabeur, G. Di Pendina, O. Boulle, and G. Gaudin, "Beyond STT-MRAM, Spin Orbit Torque RAM SOT-MRAM for High Speed and High Reliability Applications," in *Spintronics-based Computing*, pp. 145–157, Cham, Switzerland: Springer, 2015.

[299] I. Ahmed, Z. Zhao, M. G. Mankalale, S. S. Sapatnekar, J.-P. Wang, and C. H. Kim, "A Comparative Study Between Spin-Transfer-Torque and Spin-Hall-Effect Switching Mechanisms in PMTJ Using SPICE," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 3, pp. 74–82, Oct. 2017.

[300] Q. Shao, P. Li, L. Liu, H. Yang, S. Fukami, A. Razavi, H. Wu, K. Wang, F. Freimuth, Y. Mokrousov, M. D. Stiles, S. Emori, A. Hoffmann, J. Åkerman, K. Roy, J.-P. Wang, S.-H. Yang, K. Garello, and W. Zhang, "Roadmap of Spin–Orbit Torques," *IEEE Trans. Magn.*, vol. 57, p. ArticleSequenceNumber:800439, May 2021.

[301] K. Garello, F. Yasin, S. Couet, L. Souriau, J. Swerts, S. Rao, S. Van Beek, W. Kim, E. Liu, S. Kundu, D. Tsvetanova, K. Croes, N. Jossart, E. Grimaldi, M. Baumgartner, D. Crotti, A. Fumémont, P. Gambardella, and G. Kar, "Sot-mram 300mm integration for low power and ultrafast embedded memories," in *2018 IEEE Symposium on VLSI Circuits*, pp. 81–82, 2018.

[302] F. Oboril, R. Bishnoi, M. Ebrahimi, and M. B. Tahoori, "Evaluation of hybrid memory technologies using sot-mram for on-chip cache hierarchy," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 3, pp. 367–380, 2015.

[303] T. Kim, Y. Jang, M.-G. Kang, B.-G. Park, K.-J. Lee, and J. Park, "Sot-mram digital pim architecture with extended parallelism in matrix multiplication," *IEEE Transactions on Computers*, vol. 71, no. 11, pp. 2816–2828, 2022.

[304] V. Ostwal, A. Penumatcha, Y.-M. Hung, A. D. Kent, and J. Appenzeller, "Spin-orbit torque based magnetization switching in Pt/Cu/[Co/Ni]5 multilayer structures," *J. Appl. Phys.*, vol. 122, Dec. 2017.

[305] T. Fan, N. H. D. Khang, S. Nakano, and P. N. Hai, “Ultrahigh efficient spin orbit torque magnetization switching in fully sputtered topological insulator and ferromagnet multilayers,” *Sci. Rep.*, vol. 12, pp. 1–8, Feb. 2022.

[306] Y. Wang, D. Zhu, Y. Wu, Y. Yang, J. Yu, R. Ramaswamy, R. Mishra, S. Shi, M. Elyasi, K.-L. Teo, Y. Wu, and H. Yang, “Room temperature magnetization switching in topological insulator-ferromagnet heterostructures by spin-orbit torques,” *Nat. Commun.*, vol. 8, pp. 1–6, Nov. 2017.

[307] M. Dc, R. Grassi, J.-Y. Chen, M. Jamali, D. Reifsnyder Hickey, D. Zhang, Z. Zhao, H. Li, P. Quarterman, Y. Lv, M. Li, A. Manchon, K. A. Mkhoyan, T. Low, and J.-P. Wang, “Room-temperature high spin–orbit torque due to quantum confinement in sputtered BixSe(1–x) films,” *Nat. Mater.*, vol. 17, pp. 800–807, Sept. 2018.

[308] M. Z. Hasan and C. L. Kane, “Colloquium: Topological insulators,” *Rev. Mod. Phys.*, vol. 82, pp. 3045–3067, Nov. 2010.

[309] Y. Ando, “Topological Insulator Materials,” *J. Phys. Soc. Jpn.*, vol. 82, p. 102001, Sept. 2013.

[310] Y. Tokura, K. Yasuda, and A. Tsukazaki, “Magnetic topological insulators,” *Nat. Rev. Phys.*, vol. 1, pp. 126–143, Feb. 2019.

[311] M. S. Bahramy, P. D. C. King, A. de la Torre, J. Chang, M. Shi, L. Patthey, G. Balakrishnan, Ph. Hofmann, R. Arita, N. Nagaosa, and F. Baumberger, “Emergent quantum confinement at topological insulator surfaces,” *Nat. Commun.*, vol. 3, pp. 1–7, Oct. 2012.

[312] S. Barua, K. P. Rajeev, and A. K. Gupta, “Evidence for topological surface states in metallic single crystals of Bi2Te3,” *J. Phys.: Condens. Matter*, vol. 27, p. 015601, Dec. 2014.

[313] Y. Wang, P. Deorani, K. Banerjee, N. Koirala, M. Brahlek, S. Oh, and H. Yang, “Topological Surface States Originated Spin-Orbit Torques in Bi_2Se_3,” *Phys. Rev. Lett.*, vol. 114, p. 257202, June 2015.

[314] H. Wu, A. Chen, P. Zhang, H. He, J. Nance, C. Guo, J. Sasaki, T. Shirokura, P. N. Hai, B. Fang, S. A. Razavi, K. Wong, Y. Wen, Y. Ma, G. Yu, G. P. Carman, X. Han, X. Zhang, and K. L. Wang, “Magnetic memory driven by topological insulators,” *Nat. Commun.*, vol. 12, pp. 1–7, Oct. 2021.

[315] J. Atulasimha and S. Bandyopadhyay, “Bennett clocking of nanomagnetic logic using multiferroic single-domain nanomagnets,” *Appl. Phys. Lett.*, vol. 97, Oct. 2010.

[316] K. Roy, S. Bandyopadhyay, and J. Atulasimha, “Switching dynamics of a magnetostrictive single-domain nanomagnet subjected to stress,” *Phys. Rev. B*, vol. 83, p. 224412, June 2011.

[317] M. Trassin, "Low energy consumption spintronics using multiferroic heterostructures," *J. Phys.: Condens. Matter*, vol. 28, p. 033001, Dec. 2015.

[318] S. Bandyopadhyay, J. Atulasimha, and A. Barman, "Magnetic straintronics: Manipulating the magnetization of magnetostrictive nanomagnets with strain for energy-efficient applications," *Appl. Phys. Rev.*, vol. 8, Dec. 2021.

[319] S. Manipatruni, D. E. Nikonov, C.-C. Lin, B. Prasad, Y.-L. Huang, A. R. Damodaran, Z. Chen, R. Ramesh, and I. A. Young, "Voltage control of unidirectional anisotropy in ferromagnet-multiferroic system," *Sci. Adv.*, vol. 4, Nov. 2018.

[320] Y. G. Semenov, X. Duan, and K. W. Kim, "Electrically controlled magnetization in ferromagnet-topological insulator heterostructures," *Phys. Rev. B*, vol. 86, p. 161406, Oct. 2012.

[321] D. Winters, M. A. Abeed, S. Sahoo, A. Barman, and S. Bandyopadhyay, "Reliability of Magnetoelastic Switching of Nonideal Nanomagnets with Defects: A Case Study for the Viability of Straintronic Logic and Memory," *Phys. Rev. Appl.*, vol. 12, p. 034010, Sept. 2019.

[322] P. K. Mishra, N. Halavath, and S. Bhuktare, "Strain-mediated voltage controlled magnetic anisotropy based switching for magnetic memory applications," *J. Appl. Phys.*, vol. 134, Sept. 2023.

[323] M. S. Fashami, K. Roy, J. Atulasimha, and S. Bandyopadhyay, "Magnetization dynamics, Bennett clocking and associated energy dissipation in multiferroiclogic," *Nanotechnology*, vol. 22, p. 155201, Mar. 2011.

[324] R. Yu, W. Zhang, H.-J. Zhang, S.-C. Zhang, X. Dai, and Z. Fang, "Quantized Anomalous Hall Effect in Magnetic Topological Insulators," *Science*, vol. 329, pp. 61–64, July 2010.

[325] S. Cho, N. P. Butch, J. Paglione, and M. S. Fuhrer, "Insulating Behavior in Ultrathin Bismuth Selenide Field Effect Transistors," *Nano Lett.*, vol. 11, pp. 1925–1927, May 2011.

[326] Y. Xie, H. Vakili, S. Ganguly, and A. W. Ghosh, "Anatomy of nanomagnetic switching at a 3D topological insulator PN junction," *Sci. Rep.*, vol. 13, pp. 1–10, June 2023.

[327] S. Fukami, T. Anekawa, C. Zhang, and H. Ohno, "A spin–orbit torque switching scheme with collinear magnetic easy axis and current configuration," *Nat. Nanotechnol.*, vol. 11, pp. 621–625, July 2016.

[328] G. Yu, P. Upadhyaya, Y. Fan, J. G. Alzate, W. Jiang, K. L. Wong, S. Takei, S. A. Bender, L.-T. Chang, Y. Jiang, M. Lang, J. Tang, Y. Wang, Y. Tserkovnyak, P. K. Amiri, and K. L. Wang, "Switching of perpendicular magnetization by spin–orbit torques in the absence of external magnetic fields," *Nat. Nanotechnol.*, vol. 9, pp. 548–554, July 2014.

[329] G. Yu, L.-T. Chang, M. Akyol, P. Upadhyaya, C. He, X. Li, K. L. Wong, P. K. Amiri, and K. L. Wang, "Current-driven perpendicular magnetization switching in Ta/CoFeB/[TaOx or MgO/TaOx] films with lateral structural asymmetry," *Appl. Phys. Lett.*, vol. 105, Sept. 2014.

[330] J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Myers, and D. C. Ralph, "Current-Driven Magnetization Reversal and Spin-Wave Excitations in Co /Cu /Co Pillars," *Phys. Rev. Lett.*, vol. 84, pp. 3149–3152, Apr. 2000.

[331] K.-S. Lee, S.-W. Lee, B.-C. Min, and K.-J. Lee, "Threshold current for switching of a perpendicular magnetic layer induced by spin Hall effect," *Appl. Phys. Lett.*, vol. 102, Mar. 2013.

[332] S. Yoshino, H. Takagi, S. Tsunashima, M. Masuda, and S. Uchiyama, "Perpendicular Magnetic Anisotropy of TbCo Films," *Jpn. J. Appl. Phys.*, vol. 23, p. 188, Feb. 1984.

[333] J. Betz, K. Mackay, and D. Givord, "Magnetic and magnetostrictive properties of amorphous Tb(1-x)Cox thin films," *J. Magn. Magn. Mater.*, vol. 207, pp. 180–187, Dec. 1999.

[334] F. Büttner, I. Lemesh, and G. S. D. Beach, "Theory of isolated magnetic skyrmions: From fundamentals to room temperature applications," *Sci. Rep.*, vol. 8, pp. 1–12, Mar. 2018.

[335] Z. Zhao, M. Jamali, A. K. Smith, and J.-P. Wang, "Spin Hall switching of the magnetization in Ta/TbFeCo structures with bulk perpendicular anisotropy," *Appl. Phys. Lett.*, vol. 106, Mar. 2015.

[336] J. Finley and L. Liu, "Spin-Orbit-Torque Efficiency in Compensated Ferrimagnetic Cobalt-Terbium Alloys," *Phys. Rev. Appl.*, vol. 6, p. 054001, Nov. 2016.

[337] Q. Wang, J. Domann, G. Yu, A. Barra, K. L. Wang, and G. P. Carman, "Strain-Mediated Spin-Orbit-Torque Switching for Magnetic Memory," *Phys. Rev. Appl.*, vol. 10, p. 034052, Sept. 2018.

[338] U. Roy, T. Pramanik, L. F. Register, and S. K. Banerjee, "Write Error Rate of Spin-Transfer-Torque Random Access Memory Including Micromagnetic Effects Using Rare Event Enhancement," *IEEE Trans. Magn.*, vol. 52, p. ArticleSequenceNumber:3402106, June 2016.

[339] S. Yu and P.-Y. Chen, "Emerging Memory Technologies: Recent Trends and Prospects," *IEEE Solid-State Circuits Mag.*, vol. 8, pp. 43–56, June 2016.

[340] L. Wei, J. G. Alzate, U. Arslan, J. Brockman, N. Das, K. Fischer, T. Ghani, O. Golonzka, P. Hentges, R. Jahan, P. Jain, B. Lin, M. Meterelliyoz, J. O'Donnell, C. Puls, P. Quintero, T. Sahu, M. Sekhar, A. Vangapaty, C. Wiegand, and F. Hamzaoglu, "13.3 a 7mb stt-mram in 22ffl finfet technology with 4ns read sensing time at 0.9v using write-verify-write scheme and offset-cancellation sensing technique," in *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, pp. 214–216, 2019.

[341] M. Y. Song, C. M. Lee, S. Y. Yang, G. L. Chen, K. M. Chen, I. J. Wang, Y. C. Hsin, K. T. Chang, C. F. Hsu, S. H. Li, J. H. Wei, T. Y. Lee, M. F. Chang, X. Y. Bao, C. H. Diaz, and S. J. Lin, "High speed (1ns) and low voltage (1.5v) demonstration of 8kb sot-mram array," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, pp. 377–378, 2022.

[342] R. C. Tozer, "A low input capacitance sense amplifier for impedance imaging applications," *Meas. Sci. Technol.*, vol. 3, p. 508, May 1992.

[343] A. Conte, G. L. Giudice, G. Palumbo, and A. Signorello, "A high-performance very low-voltage current sense amplifier for nonvolatile memories," *IEEE J. Solid-State Circuits*, vol. 40, pp. 507–514, Jan. 2005.

[344] S. Honda, M. Nawate, M. Yoshiyama, and T. Kusuda, "Magnetization properties of bias modulated TbCo films with short periodic modulation," *IEEE Trans. Magn.*, vol. 23, pp. 2952–2954, Sept. 1987.

[345] P. Deorani, J. Son, K. Banerjee, N. Koirala, M. Brahlek, S. Oh, and H. Yang, "Observation of inverse spin Hall effect in bismuth selenide," *Phys. Rev. B*, vol. 90, p. 094403, Sept. 2014.

[346] T. Schenk, M. Pešić, S. Slesazeck, U. Schroeder, and T. Mikolajick, "Memory technology—a primer for material scientists," *Rep. Prog. Phys.*, vol. 83, p. 086501, June 2020.

[347] N. Anuniwat, M. Ding, S. J. Poon, S. A. Wolf, and J. Lu, "Strain-induced enhancement of coercivity in amorphous TbFeCo films," *J. Appl. Phys.*, vol. 113, Jan. 2013.

[348] R. Mandal, J. W. Jung, K. Masuda, Y. K. Takahashi, Y. Sakuraba, S. Kasai, Y. Miura, T. Ohkubo, and K. Hono, "Investigation of Gilbert damping of a tetragonally distorted ultrathin Fe0.5Co0.5 epitaxial film with high magnetic anisotropy," *Appl. Phys. Lett.*, vol. 113, Dec. 2018.

[349] K. Serizawa, M. Ohtake, T. Kawai, M. Futamoto, F. Kirino, and N. Inaba, "Magnetostriction Behaviors of Fe100–xCox Alloy Epitaxial Thin Films under Rotating Magnetic Field," *J. Magn. Soc. Jpn.*, vol. 43, pp. 50–58, May 2019.

[350] Y.-S. Fu, T. Hanaguri, K. Igarashi, M. Kawamura, M. S. Bahramy, and T. Sasagawa, "Observation of Zeeman effect in topological surface state with distinct material dependence," *Nat. Commun.*, vol. 7, pp. 1–6, Feb. 2016.

[351] J. Wang, T. Wang, M. Ozerov, Z. Zhang, J. Bermejo-Ortiz, S.-K. Bac, H. Trinh, M. Zhukovskyi, T. Orlova, H. Ambaye, J. Keum, L.-A. de Vaulchier, Y. Guldner, D. Smirnov, V. Lauter, X. Liu, and B. A. Assaf, "Energy gap of topological surface states in proximity to a magnetic insulator," *Commun. Phys.*, vol. 6, pp. 1–9, Aug. 2023.

[352] A. K. Kaveev, S. M. Suturin, V. A. Golyashov, K. A. Kokh, S. V. Eremeev, D. A. Estyunin, A. M. Shikin, A. V. Okotrub, A. N. Lavrov, E. F. Schwier, and O. E. Tereshchenko, “Band gap opening in the $BiSbTeSe_2$ topological surface state induced by ferromagnetic surface reordering,” *Phys. Rev. Mater.*, vol. 5, p. 124204, Dec. 2021.

[353] N. H. D. Khang, T. Shirokura, T. Fan, M. Takahashi, N. Nakatani, D. Kato, Y. Miyamoto, and P. N. Hai, “Nanosecond ultralow power spin orbit torque magnetization switching driven by BiSb topological insulator,” *Appl. Phys. Lett.*, vol. 120, Apr. 2022.

[354] L. Rehm, M. G. Morshed, S. Misra, A. Shukla, S. Rakheja, M. Pinarbasi, A. W. Ghosh, and A. D. Kent, “Temperature-resilient random number generation with stochastic actuated magnetic tunnel junction devices,” *Appl. Phys. Lett.*, vol. 124, Jan. 2024.

[355] J. Chang, L. F. Register, and S. K. Banerjee, “Topological insulator Bi2Se3 thin films as an alternative channel material in metal-oxide-semiconductor field-effect transistors,” *J. Appl. Phys.*, vol. 112, Dec. 2012.

[356] Y. Wang, D. Zhu, Y. Yang, K. Lee, R. Mishra, G. Go, S.-H. Oh, D.-H. Kim, K. Cai, E. Liu, S. D. Pollard, S. Shi, J. Lee, K. L. Teo, Y. Wu, K.-J. Lee, and H. Yang, “Magnetization switching by magnon-mediated spin torque through an antiferromagnetic insulator,” *Science*, vol. 366, pp. 1125–1128, Nov. 2019.

[357] P. Khanal, B. Zhou, M. Andrade, Y. Dang, A. Davydov, A. Habiboglu, J. Saidian, A. Laurie, J.-P. Wang, D. B. Gopman, and W. Wang, “Perpendicular magnetic tunnel junctions with multi-interface free layer,” *Appl. Phys. Lett.*, vol. 119, Dec. 2021.

[358] P. Li, J. Kally, S. S.-L. Zhang, T. Pillsbury, J. Ding, G. Csaba, J. Ding, J. S. Jiang, Y. Liu, R. Sinclair, C. Bi, A. DeMann, G. Rimal, W. Zhang, S. B. Field, J. Tang, W. Wang, O. G. Heinonen, V. Novosad, A. Hoffmann, N. Samarth, and M. Wu, “Magnetization switching using topological surface states,” *Sci. Adv.*, vol. 5, Aug. 2019.

[359] M. Dc, J.-Y. Chen, T. Peterson, P. Sahu, B. Ma, N. Mousavi, R. Harjani, and J.-P. Wang, “Observation of High Spin-to-Charge Conversion by Sputtered Bismuth Selenide Thin Films at Room Temperature,” *Nano Lett.*, vol. 19, pp. 4836–4844, Aug. 2019.

[360] Q. Lu, P. Li, Z. Guo, G. Dong, B. Peng, X. Zha, T. Min, Z. Zhou, and M. Liu, “Giant tunable spin Hall angle in sputtered Bi2Se3 controlled by an electric field,” *Nat. Commun.*, vol. 13, pp. 1–8, Mar. 2022.

[361] A. T. Barton, L. A. Walsh, C. M. Smyth, X. Qin, R. Addou, C. Cormier, P. K. Hurley, R. M. Wallace, and C. L. Hinkle, “Impact of Etch Processes on the Chemistry and Surface States of the Topological Insulator Bi2Se3,” *ACS Appl. Mater. Interfaces*, vol. 11, pp. 32144–32150, Sept. 2019.

[362] S. Karki, J. Kwon, J. Davies, R. Fabiha, V. Rogers, T. Leonard, S. Bandyopadhyay, and J. A. C. Incorvia, “Low voltage local strain enhanced switching of magnetic tunnel junctions,” *arXiv*, Nov. 2023.

[363] H. Zhu, C. A. Richter, E. Zhao, J. E. Bonevich, W. A. Kimes, H.-J. Jang, H. Yuan, H. Li, A. Arab, O. Kirillov, J. E. Maslar, D. E. Ioannou, and Q. Li, "Topological Insulator Bi2Se3 Nanowire High Performance Field-Effect Transistors," *Sci. Rep.*, vol. 3, pp. 1–5, Apr. 2013.

[364] J. Jang, G.-T. Hwang, Y. Min, J.-W. Kim, C.-W. Ahn, J.-J. Choi, B.-D. Hahn, J.-H. Choi, D.-S. Park, Y. Jung, and W.-H. Yoon, "Fatigue study and durability improvement of piezoelectric single crystal macro-fiber composite energy harvester," *J. Korean Ceram. Soc.*, vol. 57, pp. 645–650, Nov. 2020.

[365] T. Hey, *Feynman Lectures on Computation:Anniversary Edition*. Andover, England, UK: Taylor & Francis, May 2023.