Energy Efficient Spintronic Device for Neuromorphic Computation

Md Ali Azam

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

Part of the Computer and Systems Architecture Commons, Electrical and Electronics Commons, Electronic Devices and Semiconductor Manufacturing Commons, and the Nanotechnology Fabrication Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/6036

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.
Energy Efficient Spintronic Device for Neuromorphic Computation

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Mechanical and Nuclear Engineering at the Virginia Commonwealth University.

by

MD ALI AZAM
Bachelor of Science in Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, 2013

Supervisor: Jayasimha Atulasimha, Ph.D.
Professor, Department of Mechanical and Nuclear Engineering

Virginia Commonwealth University
Richmond, Virginia
September, 2019
Acknowledgment

Firstly, I would like to express my sincerest gratitude to my advisor Professor Jayasimha Atulasimha for his continuous support of my studies and research, for his motivation, patience, guidance and immense knowledge. It would have been impossible for me to come this far and accomplish what I did during my time here in Virginia Commonwealth University without his mentorship. It is an honor to have known him in person. I couldn’t have imagined a better advisor.

I whole heartedly thank my thesis committee members Professor Supriyo Bandyopadhyay and Professor Ravi L. Hadimani for taking interest in my work and making time for the evaluation of the dissertation despite having a very busy schedule.

I sincerely thank my fellow labmates Dr. Vimal Sampath, Dr. Mamun Al-Rashid, Pallabi Sutradhar, Austin R. Roe, Walid Al-Misbah, Mahadi Rajib and my friend Dr. Ahsanul Abeed for the stimulating discussions and guidance, for the sleepless nights we were working together before deadlines, and for making my stay here at VCU a pleasant and friendly experience.

A thank my labmate and long-term friend Dhritiman Bhattacharya from the bottom of my heart for all his inputs and suggestions in finishing a well-organized and informative manuscript of the dissertation. It would have been impossible for me to finish it in timely and orderly manner without his involvement.

Last but not the least, I would like to thank my family: my father, who, although is no longer with us, my mother, my brother and my sister for their unparalleled love, encouragement and support.

My research was supported in part by the National Science Foundation NSF CAREER grant CCF-1253370 (original grant and the NCS Core +Supplement award) and SHF Small grant #CCF-1815033 (the supplement to this grant funded by 2-month internship at National Institute of Standards and Technology, Gaithersburg in summer 2019 with Dr. Daniel Gopman on material characterization.

Some of the group’s research activity related to by research topics also was supported through the Virginia Microelectronics Seed Grant, Virginia Quest Commercialization Grant and Collaborative/Educational Seed Project with Manufacturing Techniques, Inc. (MTEQ) Prime Contract with the US Army CSISR Night Vision and Electronic Sensors Directorate (NVESD) and Virginia Microelectronics Consortium (VMEC), grant #MTEQ-19-028 TO 54_FP1032.
# Table of Contents

Acknowledgement ................................................................................................................. iii

List of Figures ........................................................................................................................ v

List of Tables ........................................................................................................................ vii

Abstract ................................................................................................................................ viii

Chapter 1. Introduction ........................................................................................................... 1
  1.1. Nanomagnetic memory, computing and Neuromorphic Devices ......................... 1
  1.2. Neuromorphic Computing ...................................................................................... 8
  1.3. Outline of the Dissertation ....................................................................................... 11

Chapter 2. Voltage Control of Domain Walls in Magnetic Nanowires for Energy Efficient
  Neuromorphic Devices ........................................................................................................ 13
  2.1. Introduction ................................................................................................................ 13
  2.2. Device Working Principle and Micromagnetic Modeling Approach .................... 15
  2.3 Discussion of Modeling Results .................................................................................. 20
  2.4. Energy Efficiency of Deep Neural Networks (DNNs) with Voltage Control of Domain
      Walls (VC-DW) ............................................................................................................... 26
  2.5. Conclusion .................................................................................................................. 30

Chapter 3. Resonate and Fire Neuron with Fixed Magnetic Skyrmion ......................... 31
  3.1. Introduction ................................................................................................................ 31
  3.2. Device .......................................................................................................................... 33
  3.3. Methods ...................................................................................................................... 36
  3.4. Results ........................................................................................................................ 38
  2.5. Conclusion .................................................................................................................. 46

Chapter 4. Summary and Future Work ............................................................................. 48

References ............................................................................................................................... 49
List of Figures

Figure 1.1. Spin current generation and STT for nanomagnet switching in MTJ……………………3

Figure 1.2. Switching of nanomagnet with STT via Giant Spin Hall Effect .........................4

Figure 1.3. Current flows along the surface of a topological insulator, net current flow creates a spin-polarized surface. The magnetization of nanomagnets are coupled to the surface spins .................................................................5

Figure 2.1. Schematic of the spintronic DW-synapse. (a) Top view showing a magnetic tunnel junction (MTJ) stack placed adjacent to an electrode on a piezoelectric substrate. The inset shows the PMA as a function of distance along the free layer when a voltage is applied to the strain electrode. (b) Front view of the stack (top panel). In the bottom panel, shows just the heavy metal/free layer, the electrons on the top surface of the Pt are spin polarized into the plane of the figure electrons at the bottom are spin polarized out of the plane of the figure due to the spin Hall effect. ..................................................................................................16

Figure 2.2. A schematic showing the strain gradient produced on the DW MTJ when a voltage is applied to the top strain electrode. The in-plane tensile strain decreases from the right end of the DW MTJ to the left producing a PMA gradient..........................................................18

Figure 2.3. (a) Micromagnetic model showing the final DW positions in the free layer of the MTJ along with the PMA gradients applied for a synergistic PMA gradient and current driven operation of the synaptic weight adjustment. A clear monotonically increasing PMA gradient profile is required to translate the DW from notch ① to notch ②, ③, ④ or ⑤. Both current and DW motion are left to right. NOTE: Supplementary Video 1 a, b, c and d shows a movie of the DW dynamics for each of the above PMA gradients. (b) Stochastic behaviour of the DW position due to thermal noise depicted by the widely dispersed distribution of the final DW position. The strain profiles and PMA variations are idealized assumptions and not calculated with detailed finite element analysis. They are estimated based on [30] and scaling arguments discussed later in section 2.4..........................24

Figure 2.4. Schematic of the spintronic DW-synapse. (a) Top view showing a magnetic tunnel junction (MTJ) stack placed adjacent to the electrodes on a piezoelectric substrate for SOT driven DW with PMA modulation at regular
intervals due to voltage applied to the piezoelectric locally by the electrodes. (b) Different position of the DWs achieved shown in the bottom figure by different PMA profiles created as shown in the upper panel. Initial position of the DW was 180 nm from left (not shown here). The strain profiles and PMA variations are assumptions that were estimated based on [30] and scaling arguments discussed in section 2.4.

Figure 2.5. Left: Schematic of the spintronic DW-neuron implemented with SOT. Right: Hybrid neuron synapse.

Figure 3.1. (a) A core-up skyrmion, color code on the right shows the direction of the spin, (b) Proposed device structure operated with voltage control of magnetic anisotropy (VCMA) (c) MTJ structure stacked on PZT layer for strain control of magnetic anisotropy. Note that, CMOS buffer is driven by the MTJ resistance. Therefore, fixed layer magnetization needs to be opposite to that of skyrmion core. The resonate and fire behavior of this device is described by the simulations in Figure 3.3.

Figure 3.2. (a) Damped oscillation of a fixed skyrmion’s core due to stimulation with a single pulse [Red color line: Input spike, Blue color line: Output average magnetization along the perpendicular direction (z-axis) (b) Modulation of breathing frequency by varying the interfacial parameters.

Figure 3.3. Resonant behavior: (a) Spike input (b) sinusoidal input

Figure 3.4. Phase detection: As the phase difference increases the amplitude of the input decreases thus making it harder for the magnetization to reach the threshold limit for firing.

Figure 3.5. Frequency synchronization detection
List of Tables

Table 2.1. Material parameters used for the CoFe soft layer in the Pt/CoFe/MgO heterostructure as compiled from previously published works [9,76,87-89]. A higher value of Gilbert damping is chosen for the simulation.

Table 2.2. PMA profile for achieving different positions of the DW. The DW starts at the 1st notch. The second column gives the change in PMA over the entire length of the device required to stop the DW at different notches. The third column shows the gradient in PMA for a 1 µm long device.

Table 2.3. PMA profile for arresting the DW at different electrode position with SOT clocking, showing the PMA decrease at the rightmost electrode. The decrease in the leftmost electrode is 1/3 of this value, and intermediate electrodes change the PMA in a linear fashion as shown in Figure 2.4(b).

Table 3.1. Magnetic parameters used.

Table 3.2. Performance comparison of proposed hybrid nanomagnet-CMOS vs. all CMOS resonate and fire neuron [14].
Abstract

Energy Efficient Spintronic Device for Neuromorphic Computation

By Md Ali Azam, M. Sc.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mechanical and Nuclear Engineering at Virginia Commonwealth University.

Virginia Commonwealth University, 2019.

Major Director: Jayasimha Atulasimha, Ph.D.

Professor,
Department of Mechanical and Nuclear Engineering

Future computing will require significant development in new computing device paradigms. This is motivated by CMOS devices reaching their technological limits, the need for non-Von Neumann architectures as well as the energy constraints of wearable technologies and embedded processors. The first device proposal, an energy-efficient voltage-controlled domain wall device for implementing an artificial neuron and synapse is analyzed using micromagnetic modeling. By controlling the domain wall motion utilizing spin transfer or spin orbit torques in association with voltage generated strain control of perpendicular magnetic anisotropy in the presence of Dzyaloshinskii-Moriya interaction (DMI), different positions of the domain wall are realized in the free layer of a magnetic tunnel junction to program different synaptic weights. Additionally, an artificial neuron can be realized by combining this DW device with a CMOS buffer. The second neuromorphic device proposal is inspired by the brain. Membrane potential of many neurons oscillate in a sub-threshold damped fashion and fire when excited by an input frequency that nearly equals their Eigen frequency. We investigate theoretical implementation of such “resonate-and-fire” neurons by utilizing the magnetization dynamics of a fixed magnetic skyrmion based free layer of a magnetic tunnel junction (MTJ). Voltage control of magnetic anisotropy or voltage generated strain
results in expansion and shrinking of a skyrmion core that mimics the subthreshold oscillation. Finally, we show that such resonate and fire neurons have potential application in coupled nanomagnetic oscillator based associative memory arrays.
Chapter 1

Introduction

1.1. Nanomagnetic memory, computing and neuromorphic devices

The increasing demand for computational needs is putting a constant strain on the current CMOS technology, which has been delivering the necessary improvements in performance for a long time. However, this is about to come to an end. One of the primary reasons for the saturation in performance in CMOS technology is the ever-decreasing size of the transistor, which continued for a long time as Moore’s law of miniaturization [1] was holding good. However, transistor technology will soon reach its physical limits. One of the primary concerns is the heat management of in electronic chips as size of the transistor continues to decrease further. Furthermore, as wearable technology and embedded processors becomes more ubiquitous, energy consumed in computation becomes a real challenge. Moreover, CMOS devices are very limited in their function as they are best suited to function as a binary switch. Thus, it is very difficult to apply it as a basic component in non Von Neumann computation with enough efficiency. Keeping all these restrains in mind, a possible solution on the device level needs to have at least two basic characteristics. Either it should be able perform the same task as the CMOS transistor with smaller size and energy efficiency or it needs to have more functionality that can compensate for space and energy taken up by the device. The size of the devices is already restricted due to the physical limitations. That leaves us with the second option. Non-Boolean architectures with devices that are energy efficient and provide more functionality when compared to CMOS devices are a good fit.
A key fundamental aspect that makes spintronic devices such a promising candidate for low energy usage is the coherent switching of electronic spin [2]. Due to the exchange coupling [3, 4] all spins in a single domain acts like one giant spin. Thus, in spintronics application where single domains are used, the number information carriers are lower due to the exchange coupling of magnetic spin. In a CMOS device, individual electrons act independently making the number of information carrier same as the number of electrons involved in the switching mechanism. The energy required for device operation goes down with the number of information carriers [5].

The simplest conceivable spintronic device is a single domain nanomagnet. Just like the transistor, the nanomagnet can be used as a switch. But the scope of nanomagnetic devices can go beyond the switching. The orientation of the magnetic domain holds the key information when it acts like a switch, i.e. the information is “saved” even when power is switched off. Consider, an elliptical nanomagnet with in plane anisotropy. The magnetization direction will always lie along the longer axis of the ellipse. If we were to use this as a switch the energy consumption in question should arise from the mechanism that we are using to switch it from one orientation to the other. This reasoning applies to all sorts of spintronic devices as the magnetization direction is the key information holder whether it is applied as a memory device or as a processing unit. All such schemes for manipulating the magnetization dynamics can be broadly divided into: Current controlled and voltage-controlled mechanisms. We focus on the current controlled mechanisms first and then the voltage-controlled mechanisms with each method briefly discussed as reported in Ref. [2].

Earliest attempts of manipulating magnetization in magnetic materials involve utilizing a magnetic field. Magnetic Field generated by induction method via current flowing through a coil can be controlled very precisely with electronic circuits. This feature though made it a very popular candidate
for magnetic field generation on larger scales however is not well suited for nanoscale device applications due to the high energy requirements (i.e. the $I^2R$ loss in the coil). Other constrains include the limitation on focusing the magnetic field generated in the coil, bulky design and need of mechanism to move the coil over the required location of the device (else it would need as many coils as the number of individual magnets on the chip).

![Diagram](image.png)

**Figure 1.1. Spin current generation and STT for nanomagnet switching in MTJ**

Switching a nanomagnet with spin-transfer-torque (STT) can be obtained by a spin-polarized current injected into the magnetic material. This can eliminate the chance of effecting neighboring devices while switching any particular device as electrical isolation is very easily achievable even at nanoscales. The magnetization direction in the material will switch to the direction of the spin of the spin polarized current due to spin angular momentum transfer to the resident electron spins in the nanomagnet [6-11]. Spin polarized current can be generated in various ways. Most common is the Magnetic Tunnel Junction (MTJ) where the electron passes through hard layer first which has very high anisotropy separated from the free layer via a non-magnetic spacer as shown in Figure 1.1. There is a minimum current required to switch the magnetization of the free layer which depends primarily on the degree of polarization in the hard layer, energy barrier of the free layer
between desired magnetic states. This method though provides isolation but has poor performance in terms of energy as the study shows a minimum of 100 fJ [12] is achievable which is way more compared to the CMOS device we have discussed earlier.

![Diagram of nanomagnet with STT via Giant Spin Hall Effect](image)

*Figure 1.2. Switching of nanomagnet with STT via Giant Spin Hall Effect [2]*

Efforts to reduce the minimal current requirement in STT switching has led to the generation of spin polarized current via the Giant Spin Hall Effect (GSHE). The generation of Spin current via GSHE is graphically demonstrated in Figure 1.2. As the current passes through the heavy metal slab below, electrons go through spin-dependent scattering due to strong spin orbit interaction which leads to accumulation of spin of particular direction on top surface of the slab and opposite spins on the bottom surface. The spin direction depends on the current direction thus it can be manipulated by changing the current polarity. The accumulated spin diffuses through the magnetic layer providing torque for the desired switching. Other mechanism of producing spin torque in a ferromagnet is the Rashba–Edelstein effect in the ferromagnet [13,14]. In ferromagnet with Rashba spin–orbit interaction [15], passing the current can cause spin polarization in a particular direction and switch the magnet’s magnetization to that direction. Also, the spin–orbit interaction acts like an effective magnetic field [16] and that field will assist the switching of the magnetization by delivering a torque. Energy required to produce the desired switching in this method can be reduced
further as the spin current is not a direct product of the charge current. The ratio of the spin current to the charge current can be increased by modifying the geometry and material with desired spin characteristics [12,17]. Experiments has shown as little as 1.6 fJ of energy in switching via the GSHE in low loss magnetic materials like CoFeB. Further scaling down with required energy as low as 100 aJ is achievable [18].

Figure 1.3. Current flows along the surface of topological insulator, net velocity in the direction of current flow creates a spin-polarized surface. Circular nanomagnet spins are coupled to the surface spins

Topological Insulators (TI) are a class of materials that act as insulators in the bulk but has conducting states on the surface [19]. The origin of such properties is credited to non-trivial symmetry-protected topological order [20, 21]. Topological insulators (TI) [19] can be used in generating spin accumulation. As the bulk of the material does not conduct thus requiring less current for generating similar effect when compared to GSHE in regular materials with large spin-orbit interaction. The interface between the TI and the nanomagnet contains two-dimensional spin electrons with Rashba Spin-orbit interaction [22]. The spin accumulation due to the GSHE and the Rashba spin-orbit interactions are given by $\langle S_y \rangle_{GSHE} = -\frac{h}{2eV_F}J_x$ and $\langle S_y \rangle_{Rashba} = \frac{\hbar m^*\alpha_R}{4eV_F}J_x$ respectively where $\hbar$ is reduced Planck’s constant, $J_x$ is the surface current density, $e$ is the charge of electron, $V_F$ is the
Fermi velocity on the surface, $E_F$ is the Fermi energy, $m^*$ is the effective mass of the electron at the interface and $\alpha_R$ is the Rashba constant. The total spin accumulation is the sum of these two terms. Exchange coupling forces the spin in the TI-nanomagnet interface to be parallel with each other. As a result, the spin current diffuses into the magnet (into the -z direction of Figure 1.3). One drawback for such scheme is that if the magnetic material is conductive it will shunt any surface current form TI-nanomagnet interface thus making the whole mechanism less effective. In order to have best efficiency it is therefore necessary to have an insulating nanomagnet on top of the TI [23].

Switching of nanomagnets with interface spin-orbit torques [24,25] have been reported. Rashba spin-orbit interaction arises in solids with structural inversion asymmetry. This produces a non-zero slope of the conduction band creating an effective electric field which gives rise to the spin-orbit interaction [15]. The spin accumulation produced by the Rashba effect can effectively switch magnets with PMA in presence of an in-plane magnetic field. Structural inversion has been achieved in a scheme called vertical structural asymmetry by sandwiching the magnetic layer between materials of different composition [25]. Elimination of the need of in plane field for switching nanomagnets with PMA has been demonstrated via lateral structural asymmetry [26]. Magnetization switching using SOT via domain wall motion has also been successfully demonstrated [27, 28]

One of the voltage-controlled energy efficient mechanisms used to manipulate the magnetization in a nanomagnet is “Straintronics”. This particular mechanism works with two phase (piezoelectric/magnetostrictive) multiferroic nanomagnet. Such switching involves applying a strain on the nano-magnet via the piezoelectric substrate. Due to magnetostriction the anisotropy profile of the
magnet can be manipulated by changing the magnitude, direction and nature (tensile or compressive) of the applied stress causing the switching or change in the magnetization direction of the magnet [29]. Such schemes have been demonstrated successfully with promising energy efficiency and possible application in making computational and memory devices [2, 30-32].

Voltage control of magnetic anisotropy (VCMA) is another energy efficient method of controlling the magnetization in nanomagnetic devices. When magnetic materials are interfaced with non-magnetic material the interface may become sensitive to electric field i.e. it may become magnetoelectric (ME) [33-35] depending on the electrical properties (i.e. possible interaction of spin and orbital momentum) of the material. Specially in case of oxides such as MgO when interfaced with metallic ferromagnets exhibits large perpendicular magnetic anisotropy (PMA) [36-38] and a greater ME is observed. This effect has been attributed to spin dependent charge screening and electric field induced modulation of the relative occupancy of the d-orbital at the interface [33-35,39]. Due to the interfacial nature of the origin of the ME effect only thin film metallic nanomagnets exhibit VCMA effects. The inherent energy efficiency is demonstrated in switching of MTJs [40-42] is largely due to the enhanced areal density and the high resistance of the MTJ due to the use of dielectric tunnel barrier.

We have discussed some of the methods for manipulating the magnetization dynamics that are supported by significant works to be as effective as described. However, in our discussion so far, the primary focus was on the energy efficiency of spintronic devices. There are however other key advantages of spintronic devices over conventional CMOS transistors and pure electrical equivalents. As the current computational trend is moving towards the non-Boolean realm devices with non-linearity as well as many stable states will become a suitable replacement for the 2-state CMOS
device. Neuromorphic computation which is gaining momentum now, due to its primitive nature relies more on analog nature of a device thus making it harder to implement such architecture at least on hardware with devices that can only provide discreet logic states. Artificial neural networks require modifiable nonlinear functionality for neurons as well as synapses with near analog nature to provide significant accuracy to be useful in real life applications. It is however undeniable that the CMOS transistors will always remain the best choice for Boolean logic architecture as the spintronic devices can hardly be implemented without incorporating certain degree of error in switching [43, 44].

1.2. Neuromorphic Computing

Neuromorphic computing as concept was developed in early eighties by Carver Mead when developing very-large-scale integration (VLSI) systems to implement neuro-biological architecture [45]. In general, it refers to computer architecture that mimics biological brains to provide outstanding performance and efficiency in areas where traditional Von Neumann architecture is either too energy inefficient or does not provide the required functionality. The human brain, when compared to traditional CMOS based processors, has some clear advantages like energy efficiency and showing great accuracy at certain tasks which do not involve repeated fast calculation to be performed accurately. Tasks like image and pattern recognition, language processing falls into this category. Physical devices on the other hand are free from various limitations that human brain is entailed with. Fatigue, poor multi-tasking abilities, slower speed, bias and inconsistency are some of the limitations all biological brains have in common. By combining the fast acting, relentless and reliable workability of physical devices with the energy efficient and uniqueness of human brain, neuromorphic computation holds greater promises in overcoming many challenges faced by traditional
computational methods. However, it has been a major challenge to mimic the functionality of brain using physical devices. Among the various possible solutions, we briefly explore two very promising ones here namely Deep Neural Networks and Oscillator Neuromorphic Computing based on energy efficient voltage controlled spintronic/nanomagnetic devices.

**a. Deep Neural Networks:**

Deep neural networks are special classes of artificial neural network where the neurons are arranged on multiple layers. Each layer progressively extracts higher-level features from the raw input data provided. As the complexity of the data starts to buildup, it becomes more and more difficult to linearly separate the solution space as a result the general approach of problem-solving neural networks with few neurons and few layers start to fail. Due to the nature of the network it is required to have a tremendous number of synapses as well as neuronal functionality that supports the exact training algorithm that needs to be implemented. Implementing such a scheme on device level thus poses two fundamental challenge to begin with. One is the implementing the synapse with a certain degree of accuracy and the other is the finding a suitable way of implementing the neuronal functionality. Recent efforts in developing dedicated CMOS neuromorphic processors [46] have been successful but cannot perform onboard (real-time) learning/training. Developing energy efficient multistate (preferably continuous and therefore analog) but non-volatile synapses would be a key enabler in realizing such real time learning for medical processors and sensor networks. In such applications energy efficiency is at a premium and synaptic weights of limited accuracy are enough [47]. We propose an analog non-volatile synapse using Domains Walls (DWs) clocked with Spin Orbit Torque (SOT) and arrested at specific points by voltage induced strain during the learning phase of a Deep Neural Network (DNN) [48]. This is discussed in Chapter 2.
b. Oscillator Neuromorphic Computing

One of the reasons for human/biological brains to be so energy efficient other than the fact that the biological neurons are extremely efficient electrochemical components is that major portion of the neurons are spiking in nature. That means they only fire periodically and firing also depends on fulfilment of certain conditions created by the actual network (firing of other neurons and synaptic properties). A large portion of the brain therefore is basically a spiking network where the actual conduction time is much smaller when compared to a network where continuous signal flow is prerequisite to fulfill the information processing. Other advantage of having such a system is that the information now can be decoded into many aspects of the firing pattern namely frequency, phase, duration of firing, amplitude, inter-spike interval, modulation of frequency any so on. Spiking neural networks are artificial neural networks that are based on this fundamental nature of biological neurons [49]. In that sense the spiking neural networks are the closest to the human brain in terms of working principal when compared to all other types of artificial neural networks. First biological model of spiking neuron was proposed by Alan Lloyd Hodgkin and Andrew Huxley in 1952 [50] for which they received the Nobel Prize in Physiology in 1963. The complex nature of biological neurons has led to many other models but for implementation of spiking neural network much simpler versions are enough as it would be unimaginably complex to find device equivalents of biological neurons. In recent years several hardware implementations have been successfully done including Intel’s LOIHI [51] and IBM’s Truenorth [46] which shows the great potential this architecture holds for the future. In any case the challenge for the hardware implementation lies in choosing the encoding mechanism whether it’s frequency, phase, or the timing interval or any other aspect and finding the proper devices to implement the network.
Recent works on spintronic oscillators [52-54] shows great promises and we present our effort in implementing a resonate and fire spiking neuron using fixed magnetic skyrmion in chapter 3.

1.3. Outline of this dissertation

The thesis is divided in two parts based on two published papers (one posted on arXiv [Chapter 2] and one published in the special issue on neuromorphic computing in the Journal of Applied Physics in 2018 [Chapter 3]).

The first part focuses on our work on implementing domain wall device as an artificial neuron as well as a synapse. The current state of the art device level implementation of artificial neural network still has some major hurdles to overcome in terms of energy efficiency and on-board/real-time learning. In order to overcome these issues, we come up with energy efficient neurons as well as non-volatile synapses with the ability to be program weights in a deterministic manner. While artificial neurons and synapses have been proposed using current-controlled nanomagnets and memristors that are potentially more energy efficient than pure CMOS implementations, there is still room for increasing the energy efficiency. Therefore, we proposed and comprehensively model the Domain Wall (DW) dynamics in magnetostrictive CoFe nanowires that are clocked with current in conjunction with an anisotropy gradient. As the domain wall moves controllably due to the current and anisotropy gradient, we ensure it gets arrested at discreet locations with the help of notches placed in the nanowire. This discreet positioning allows deterministic control of the weights of the synapse. We examine the magnetization dynamics of the chiral Néel domain walls in a thin nanowire of CoFe with PMA that is modulated by a voltage-induced strain. The CoFe experiences a Dzyaloshinskii-Moriya Interaction (DMI) that stabilizes chiral walls, which are translated by STT
or SOT to implement energy efficient non-volatile artificial synapses. We also describe the way this device, in combination with a CMOS buffer, can also function as a neuron and implement Deep Neural Networks (DNNs).

The second part comprises of our work on implementing an artificial Resonate and Fire Neuron with Fixed Magnetic Skyrmions as the free layer of a Magnetic Tunnel Junction. Magnetic skyrmions have been studied as a potential candidate for memory application [55-57]. The primary motivation in implementing skyrmions for making an artificial neuron was the inherent energy efficiency in manipulating the magnetization dynamics with voltage and also similarity with some biological neurons. The dynamics of the magnetization of the skyrmion when subjected to a changing PMA mimics the behavior of biological neurons with subthreshold damped oscillation. As a result, voltage spikes can be used in the same manner as an action potential, which acts as an input for a biological neuron. We propose to employ voltage control of magnetic anisotropy or voltage generated strain as an input (spike or sinusoidal) signal, which modulates the perpendicular magnetic anisotropy (PMA). We backup our claim by rigorous micro magnetic simulation performed using Mumax©.
Chapter 2

Voltage control of domain walls in magnetic nanowires for energy efficient neuromorphic devices

2.1. Introduction

There has been considerable recent progress in the development of dedicated CMOS processors for neuromorphic computing such as IBM’s TrueNorth that can implement 1 million spiking neurons and 256 million configurable synapses [46] while consuming ~70mW power. However, these neuromorphic processors have drawbacks such as lack of onboard (real-time) learning/training. More importantly, they have poor energy efficiency in comparison to the human brain, which has ~ 100 billion neurons and ~ 500 trillion synapses and consumes a mere ~ 20 watts of power [47]. Thus, a key challenge for hardware implementation of artificial neural networks lies in finding energy efficient hardware implementations of neurons and non-volatile synapses whose weights can be changed easily and deterministically with very little energy as the network learns from data in real time. While artificial neurons and synapses have been proposed using current-controlled nanomagnets [48,58-62] and memristors [63-68] that are potentially more energy efficient than pure CMOS implementations, there is still room for increasing the energy efficiency.

We propose implementing energy efficient artificial synapse using a magnetic tunnel junction (MTJ). The free layer of the MTJ comprises a magnetostrictive nanowire racetrack made of CoFe or CoFeB for example. Here we model the magnetization dynamics of the domain walls in the racetrack. The wall is driven through the racetrack clocked with current passing through the racetrack exerting a spin transfer torque (STT) [69-72], or by spin orbit torque (SOT) due to current flowing in a heavy metal layer directly underneath the racetrack [24,73-75]. The heavy metal layer
leads to a perpendicular magnetic anisotropy (PMA) in the CoFeB or CoFe layer, and a Dzyaloshinskii-Moriya interaction (DMI) which stabilizes chiral domain walls. In order to achieve controlled positioning of the DW we propose current clocked DW motion in conjunction with a gradient in the PMA [76,77]. Notches are placed at regular intervals to arrest the DW at different locations of the racetrack. For the SOT-clocked DW motion [27,28,75,78] instead of using a notched race track and PMA gradient, a racetrack of uniform width is used along with modulation of the PMA at regular intervals. This modulation (PMA reduction) creates a barrier to the motion of the DW, arresting the DW at different locations depending on the voltage applied. We also describe the manner in which this device, in combination with a CMOS buffer, can also function as a neuron and implement Deep Neural Networks (DNNs). Such an implementation is important in applications where energy efficiency is at a premium, such as medical processors and sensor networks that need to learn from data in real time rather than be trained offline, and where synaptic weights of limited accuracy are sufficient [79].

Section 2.2 describes the working principle of the device and our micromagnetic modeling approach. Section 2.3 presents and discusses simulation of the DW dynamics in the presence of PMA gradients and STT/SOT while Section 2.4 compares the energy efficiency of this approach with other spintronic and memristor approaches.

### 2.2. Device working principle and micromagnetic modeling approach

The working principle of the device (Figure 2.1) is explained in terms of the DW dynamics within the magnetic free layer of a magnetic tunnel junction (MTJ). The resistance of the MTJ, which
consists of a free layer, a tunnel junction and a fixed layer pinned by a synthetic antiferromagnet (SAF), varies with the location of the DW in the free layer. Therefore, the DW position determines the non-volatile resistance states of the spintronic synapse and can be programmed by a voltage, as described below. As the Deep Neural Network (DNN) learns from data in real time, a backpropagation algorithm [80] implemented on a CMOS application-specific co-processor can calculate the new weights for different synapses and output these as specific programming voltages (not addressed in this paper). These voltages should be able to reprogram the resistance states of the synapses to update their resistance values, as described in this work.

**Clocking:** Consider a perpendicular magnetic anisotropy racetrack consisting of a heavy metal/ferromagnet bilayer that could be deposited a piezoelectric film to realize our proposed device as shown in Figure 2.1. Such a bilayer (e.g. Pt/CoFe) derives its PMA from interfacial effects and exhibits significant DMI that stabilizes the formation of chiral Néel domain walls [75].

**SOT clock:** SOT acting on the magnetization is generated when current flows in the heavy metal layer. The damping like field (DL-field) thus produced is responsible for translating the Néel domain wall in the ferromagnetic layer [75]. Reversing the direction of the current in the Pt layer reverses the direction of domain wall motion, resets the domain wall position, and hence resets the resistance of the DW MTJ device.
STT clock: Alternatively, for current clocking by spin transfer torque, electrons passing through the domain wall are preferentially polarized along the magnetization orientation of the region through which they pass and exert a torque on the magnetization of the subsequent region they enter [69-72]. This causes the spins within the wall to rotate thus initiating a domain wall motion in the direction of the electron flow.

SOT vs. STT clock: When CoFe is used with a heavy metal (Pt) underlayer that leads to PMA and DMI, the ratio of the current flowing through the CoFe that leads to STT and to the current flowing through the Pt that leads to SOT depends inversely on the ratio of their resistances. Rather than consider the case of mixed STT and SOT, we consider the two extreme cases: pure SOT and pure STT in our simulations to understand the domain wall motion with these two clocking mechanisms.
also study the manner in which the DW can be arrested in a specific region of the racetrack by applying a voltage-induced strain under these two different clocking mechanisms.

**Voltage control of domain wall position:** Stopping the domain wall at a specific position along the racetrack is accomplished by applying a voltage to the side electrode (Figure 2.1(a)) while the DW is being “clocked” by SOT or STT. Consider a domain wall that has been “reset” to one end of the racetrack and is moved along the racetrack towards the other end in Figure 2.1 (a) by SOT from a current in the adjacent Pt layer or by STT from a charge current through the free layer. Application of a voltage between the side electrode and the bottom contact of the piezoelectric layer produces an electric field through the piezoelectric thickness, which in turn produces an in-plane stress in the manner described in Ref [30]. This leads to a local strain gradient in the piezoelectric, which is transferred to the ferromagnetic layer, altering its PMA as shown in Figure 2.1(a). A modified scheme as shown in Figure 2.4 can also be used as discussed later.

The mechanism of generation of the strain gradient is explained in detail in Figure 2.2. When a voltage is applied to the top electrode, a local electric field is generated through the thickness of the piezoelectric between the area directly underneath the top electrode area and the bottom electrode. This causes an out of plane expansion of the piezoelectric and consequently an in-plane contraction (due to Poisson’s ratio) of the area below the top electrode. This produces a tensile in-plane strain in the region of the piezoelectric immediately adjoining the electrode, with a magnitude decreasing with distance away from the electrode. This creates a strain gradient as shown in Figure 2.2, upper schematic. While a similar strain gradient is created in the in-plane direction orthogonal to that shown in the figure, we are only concerned with the strain gradient along the DW MTJ device. Furthermore, if the piezoelectric is deposited on a stiff substrate, the bottom of the piezoelectric is clamped but the top part of the piezoelectric can experience the in-plane strain gradient.
The benefit of this scheme is that the piezoelectric film develops a strain gradient even though it is not patterned provided the in-plane dimension of the electrode is approximately equal to the thickness of the film [30]. The strain gradient will be most significant within a distance of one to two times the piezoelectric film thickness [30]. This in-plane strain in turn modulates the perpendicular anisotropy of the soft layer and provides a spatial variation of the energy landscape of the Néel DW in the racetrack.

Thus, the device relies on stress generated by the electrode to arrest the SOT/STT-induced motion of the DW, leaving the DW pinned at a notch or a specific location where there is a PMA barrier (Figure 2.4, discussed later). The strain and therefore modulation of PMA is largest at the left end, and minimum at the right end (Figure 2.2). This PMA gradient in conjunction with the torque on the DW and notches patterned in the wire determines the position where the DW is arrested.

**Micromagnetic model:** Mumax [81] was used to perform simulations of the domain wall dynamics using the Landau-Lifshitz-Gilbert (LLG) equation in the presence of thermal noise at room temperature. The time rate of change of magnetization in a volume element of the magnetic material is given by:
\[ \frac{\partial \vec{m}}{\partial t} = \vec{\tau} = \left( \frac{\gamma}{1 + \alpha^2} \right) \left( -\vec{m} \times \vec{H}_{\text{eff}} + \alpha \left( \vec{m} \times (\vec{m} \times \vec{H}_{\text{eff}}) \right) \right) \]  

(2.1)

where \( \vec{m} \) is the reduced magnetization (\( \vec{M}/M_{\text{sat}} \)), \( M_{\text{sat}} \) is the saturation magnetization, \( \gamma \) is the gyromagnetic ratio and \( \alpha \) is the Gilbert damping coefficient. The quantity \( \vec{H}_{\text{eff}} \) is the effective magnetic field, which is given by:

\[ \vec{H}_{\text{eff}} = \vec{H}_{\text{demag}} + \vec{H}_{\text{exchange}} + \vec{H}_{\text{anisotropy}} + \vec{H}_{\text{thermal}} \]  

(2.2)

Here, \( \vec{H}_{\text{demag}} \) is the demagnetizing field produced by all the other volume elements of the magnetic material, and \( \vec{H}_{\text{exchange}} \) is the effective field due to Heisenberg exchange coupling and DMI [82, 83]. The DMI contribution to the effective exchange field is given by:

\[ H_{DM} = \frac{2D}{\mu_0 M_s} [(\vec{\nabla} \cdot \vec{m}) \hat{z} - \vec{v} m_z] \]  

(2.3)

where \( m_z \) is the z-component of magnetization and \( D \) is the effective DMI constant. The effective field due to the perpendicular anisotropy is:

\[ \vec{H}_{\text{anisotropy}} = \frac{2K_{u1}}{\mu_0 M_{\text{sat}}} (\vec{u} \cdot \vec{m}) \vec{u} + \frac{4K_{u2}}{\mu_0 M_{\text{sat}}} (\vec{u} \cdot \vec{m})^3 \vec{u} \]  

(2.4)

where, \( K_{u1} \) and \( K_{u2} \) are first and second order uniaxial anisotropy constants respectively and \( \vec{u} \) is the unit vector in the out-of-plane direction. Strain effectively modulates the anisotropy energy and is incorporated by modulating \( K_{u1} \) according to Eq. (2.2). We assume \( K_{u2} = 0 \), as we are not dealing with textured or single crystal materials [84].

Thermal noise is modeled by a random, effective magnetic field (\( \vec{H}_{\text{thermal}} \)) applied in the manner described in [85, 86] within the micromagnetic framework [81]. Furthermore, the field-like and damping-like SOTs and the STT due to charge current are modeled with the appropriate terms [9, 81] assuming polarization=1, Slonczewski parameter \( \Lambda=1 \) and secondary spin-torque parameter \( \dot{\varepsilon}=0 \). In our simulations, we do not consider STT and SOT at the same time. Instead, we present two
cases: one in which only STT is considered and the other in which only SOT is considered. This leads to an understanding of the clocking of DW motion and its arrest at a specific position using voltage induced strain for these two different clocking mechanisms.

The discretization cell sizes used for the simulations were 4nm×4nm×1nm and the material parameters used for CoFe (soft layer) of the Pt/CoFe/MgO heterostructure is summarized in Table 2.1. CoFe has sufficient magnetostriction to produce a PMA gradient that can arrest the DW at specific positions in the model. While Gilbert damping is ~0.01 to 0.03 in these materials, we used a higher value (0.1) so the DW exhibits more stable dynamics. In practice, defects and edge roughness are likely to impede the DW giving dynamics characteristic of the higher damping.

**Table 2.1.** Material parameters used for the CoFe soft layer in the Pt/CoFe/MgO heterostructure as compiled from previously published works [9,76, 87-89]. A higher value of Gilbert damping is chosen for the simulation.

<table>
<thead>
<tr>
<th>Parameters used in simulation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturation Magnetization ($M_{sat}$)</td>
<td>$1 \times 10^6 \text{ Am}^{-1}$</td>
</tr>
<tr>
<td>Exchange Constant ($A_{ex}$)</td>
<td>$2 \times 10^{-11} \text{ Jm}^{-1}$</td>
</tr>
<tr>
<td>Perpendicular Anisotropy Constant ($K_{u1}$)</td>
<td>$7.5 \times 10^5 \text{ Jm}^{-3}$</td>
</tr>
<tr>
<td>Gilbert Damping ($\alpha$)</td>
<td>0.1</td>
</tr>
<tr>
<td>DMI Constant ($D$)</td>
<td>0.001 Jm$^{-2}$</td>
</tr>
<tr>
<td>Saturation Magnetostriction ($\lambda_s$)</td>
<td>250 p.p.m</td>
</tr>
</tbody>
</table>

**2.3. Discussion of modeling results**

We discuss the modeling results for nanowires of length 1000 nm and width 100 nm patterned with five notches as shown in Figure 2.3. A gradient in PMA caused by the strain gradient drives the DW towards the lower PMA region in order to reduce the DW energy, i.e. motion is induced in the direction of the negative PMA gradient. We consider the case where the PMA gradient and STT
from the charge current drive the DW in the same direction, and then we perform simulation in the presence of thermal noise to understand if these DWs can be arrested deterministically at room temperature at specific notches. We subsequently perform another SOT-driven simulation in which the DW is arrested in regions of varying PMA produced by an alternative electrode design (Figure 2.4) and no notches are used.

a. PMA gradient assists STT driven DW motion

In a MTJ racetrack with CoFe soft layer, the DW motion was assumed to be initiated by STT assisted by a PMA gradient (without including the effect of thermal noise). Given that the current pulse acts as the clocking signal, the ON time for current is fixed at 6 ns. This is considered the “write time” for reprogramming the synaptic weight and, based on the analysis of the DW motion at different currents and PMA gradients, it is sufficient to translate the DW to any possible location within this STT clocked device.

Table 2.2 shows the PMA gradients required to drive the DW from notch 1 to any of the other notches in Figure 2.3, if the current density of the clock is kept constant at $8.7 \times 10^{12} \text{A m}^{-2}$, which is just below the critical current needed to de-pin and initiate the DW motion. For a CoFe layer of 1 nm thickness and width 100 nm this corresponds to a current of ~1 mA (we do not consider any current through the Pt layer and do not consider SOT for this case). We do not account for increase in current density at the notches in this simulation. Furthermore, from the point of view of programming the synapse, the voltage required to achieve a certain DW position need not be a linear function of position, as this is envisaged to be a pre-calculated analog voltage that is output by the co-processor that implements the learning algorithm.
Table 2.2. PMA profile for achieving different positions of the DW. The DW starts at the 1st notch. The second column gives the change in PMA over the entire length of the device required to stop the DW at different notches. The third column shows the gradient in PMA for a 1 µm long device.

<table>
<thead>
<tr>
<th>Final Position of the DW</th>
<th>Required ∆PMA (Jm(^{-3}) over device length)</th>
<th>PMA gradient that assists current (Jm(^{-3})/nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2(^{nd}) Notch</td>
<td>0.2 \times 10^4</td>
<td>2.00</td>
</tr>
<tr>
<td>3(^{rd}) Notch</td>
<td>1.6 \times 10^4</td>
<td>16.00</td>
</tr>
<tr>
<td>4(^{th}) Notch</td>
<td>3 \times 10^4</td>
<td>30.00</td>
</tr>
<tr>
<td>5(^{th}) Notch</td>
<td>3.7 \times 10^4</td>
<td>37.00</td>
</tr>
</tbody>
</table>

These results show that at this current density, the current alone cannot drive the DW out of the first notch, but a combination of the current and PMA gradient can. The PMA gradient reduces the current density required to initiate and sustain the DW motion along the racetrack when compared to the case of current only. The DW motion is retarded as it moves into the region of lower PMA, helping to arrest the DW at the notches. The synergistic effects of PMA gradient and current lead to a lower energy operation. Finally, we confirm through simulation (Figure 2.3) that each of the different positions of the DW is achievable deterministically, because at a certain PMA gradient the combined strength of the current and PMA gradient is sufficient to move the DW to the desired notch during the current ON period. Therefore, for different strengths of the PMA gradient, the DW travels different distances even though the application time of the current and PMA gradient is kept constant.

**Thermal noise effect at room temperature and scaling issues:** The simulations were repeated in the presence of thermal noise at 300 K. With thermal noise, the minimum current required to initiate the motion of the DW has a lower value when compared to the zero thermal noise case. Moreover, there is a reduction in the effectiveness of the notch in arresting the motion of the DW in presence of thermal noise. Thus, a slight reduction of current is also required to regain the effectiveness of
the notches while keeping it high enough so that DW motion can be initiated. This determined our choice of a current density of $8.4 \times 10^{12} \, \text{A} \, \text{m}^{-2}$ to explore the effect of thermal noise in arresting the DW motion at notch 3 with a PMA gradient of $10 \, \text{Jm}^{-3}/\text{nm}$. This value of the PMA gradient is chosen between the PMA gradient required for arresting the DW at the 2nd notch ($2 \, \text{Jm}^{-3}/\text{nm}$) and the 3rd notch ($16 \, \text{Jm}^{-3}/\text{nm}$). The probability distribution of the final position of the DW is shown in Figure 2.3(b). The DW was most likely to be found in notches 3 or 4, but there was a significant probability of its being in notches 2 or 5.

This can be attributed to the relatively small change in energy due to PMA modulation through strain compared to that of thermal energy ($k_b T = 4.14 \times 10^{-21} \, \text{J}$). To illustrate the energy scales, considering a $10 \, \text{Jm}^{-3}/\text{nm}$ PMA gradient (i.e. a total PMA modulation of $10 \, \text{kJm}^{-3}$ over the length of $1 \, \mu\text{m}$ nanowire), this energy change, $\Delta E = (\Delta \text{PMA}/L) \times \text{notch spacing} \times \text{Volume} = (10 \, \text{Jm}^{-3}/\text{nm}) \times 167 \, \text{nm} \times 16700 \, \text{nm}^3 = 26 \times 10^{-21} \, \text{J} \approx 6.5 \, k_b T$. Here $(\Delta \text{PMA}/L)$ is the PMA gradient ($\Delta \text{PMA}$ over 1 micron length), the volume corresponds to the volume of free layer between two notches, $k_b$ is the Boltzmann constant and $T$ is room temperature in Kelvin. This shows that the change in PMA is modest and hence the PMA gradient does not have deterministic control in positioning the DW in the presence of thermal noise. To circumvent this issue, a higher PMA gradient could be used, though this would require a greater strain gradient, or a greater thickness free layer could be used with its PMA derived from bulk effects (e.g. magnetoelastic) instead of interfacial anisotropy. This analysis has not considered edge roughness [90], which can provide additional pinning sites and even remove the need for lithographic patterning of notches.
Figure 2.3 (a) Micromagnetic model showing the final DW positions in the free layer of the MTJ along with the PMA gradients applied for a synergistic PMA gradient and current driven operation of the synaptic weight adjustment. A clear monotonically increasing PMA gradient profile is required to translate the DW from notch 1 to notch 2, 3, 4 or 5. Both current and DW motion are left to right. NOTE: Supplementary Video 1 a, b, c and d shows a movie of the DW dynamics for each of the above PMA gradients.

(b) Stochastic behaviour of the DW position due to thermal noise depicted by the widely dispersed distribution of the final DW position. The strain profiles and PMA variations are idealized assumptions and not calculated with detailed finite element analysis. They are estimated based on [30] and scaling arguments discussed later in section 2.4.

b. SOT-driven DW motion with pinning of the DW achieved by spatial PMA modulation

We also simulated DW motion clocked by SOT without including the effect of thermal noise in an MTJ racetrack with CoFe soft layer and no notches. Here we assume no current flows in the CoFe layer and hence there is no STT. We found that discrete PMA variation as shown in Figure 2.4 was more suitable to control DW in this case with SOT, instead of using notches and a uniform PMA gradient as in the previous case with STT. The electrode arrangement shown in Figure 2.4 alters the PMA at specific regions of the racetrack between the electrodes enabling creation of regions of
25

Moreover, due to the different spacing between the pairs of electrodes, the one with smaller gap will create a larger decrease in PMA due to higher stress. However, the PMA profiles in Figure 2.4(b) are ideal representation of the actual PMA profile based on scaling arguments presented in section 2.4 and prior work [30]. Determination of the real stress and PMA profile with detailed finite element analysis is beyond the scope of this paper.

The SOT clock was simulated with a current density \( J_c \sim 0.7 \times 10^{11} \text{ Am}^{-2} \) through the Pt layer of length 1000 nm, width 100 nm and thickness 1 nm for a clocking period of 20 ns. For lower voltage applied to the strain electrode the PMA reduction is small at the left end. Therefore, the DW experiences smaller barriers and can be translated further to the right. With increasing voltage, barriers at the left increase, thus arresting the DW closer to the left end as shown in Figure 2.4(b). The
geometric arrangement of the electrodes will ensure that the PMA decrease due to adjacent electrodes varies in a linear fashion. In this case, the ratio between the maximum and minimum PMA change between right most and left most electrode position is kept at 3. This will ensure by only varying the voltage the DW can be arrested at all electrode location without requiring a very large stress to be applied to the piezoelectric. The maximum PMA change for each case of the DW position is shown in Table 2.3. The DW would eventually drift to the center of the racetrack to minimize magnetostatic energy once the applied voltage is removed. However, with realistic edge roughness the DW is pinned [90] and hence implements a non-volatile synapse.

Table 2.3. PMA profile for arresting the DW at different electrode position with SOT clocking, showing the PMA decrease at the rightmost electrode. The decrease in the leftmost electrode is 1/3 of this value, and intermediate electrodes change the PMA in a linear fashion as shown in Figure 2.4(b)

<table>
<thead>
<tr>
<th>Final Position of the DW (Electrode position starting from left)</th>
<th>Maximum ΔPMA (Jm⁻³) at the rightmost electrode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ⁿᵈ</td>
<td>0.36 × 10⁵</td>
</tr>
<tr>
<td>3ʳᵈ</td>
<td>0.30 × 10⁵</td>
</tr>
<tr>
<td>4ᵗʰ</td>
<td>0.24 × 10⁵</td>
</tr>
<tr>
<td>5ᵗʰ</td>
<td>0.18 × 10⁵</td>
</tr>
</tbody>
</table>

2.4. Energy Efficiency of Deep Neural Networks (DNNs) with Voltage Control of Domain Walls (VC-DW)

While we have discussed the DW dynamics and operation of the non-volatile voltage programmed synapse in detail, Figure 2.5(a) shows the manner in which this device can be adapted to form a hybrid DW-CMOS neuron. The CMOS buffer implements the threshold functionality of a neuron (Figure 2.5 (a)) as well as the ability of the neuron output of one stage to drive inputs to various neurons of the next stage (high fan-out of the CMOS stage) via synapses. In order to reset the
neuron, the current/SOT clock is used with current flowing in the opposite direction. Thereafter, the clock is used to synchronize the information flow from one state of the DNN to the next. Figure 2.5 describes the manner in which the outputs of one set of neurons can be multiplied by the synaptic weights and input to a neuron at the next stage.

**Figure 2.5.** Left: Schematic of the spintronic DW-neuron implemented with SOT. Right: Hybrid neuron synapse.

**Energy efficiency and area-density of the voltage control DW DNN implementation vs. other implementation schemes**

The energy dissipation in the device can be divided into two parts. One part consists of charging the piezoelectric layer for stress generation, which is essentially the energy lost in charging the capacitor $\frac{1}{2}CV^2$, $C =$ capacitance of the piezoelectric layer between the metal contacts, $V =$ voltage applied). The other part is the $I^2R$ loss of the clocking current through the magnetic layer of the racetrack or current through the platinum layer.

For the device simulated in the previous sections, the maximum $\Delta PMA$ was $0.37 \times 10^5 Jm^{-3}$ across the length of the device. The stress required to obtain this may be estimated as $\sigma = \frac{\Delta PMA}{3/2 \lambda}$.
where \( \lambda \) is the magnetostriction. For CoFeB, with \( \lambda \sim 30 \times 10^{-6} \), an unreasonably high stress of \(~800\) MPa would be required, but for CoFe with \( \lambda \sim 250 \times 10^{-6} \) a stress of order \(~100\) MPa can produce the needed \( \Delta \)PMA. From the Young’s modulus of CoFe of \(~2.0 \times 10^5\) MPa, a strain of \(~5 \times 10^{-4}\) is required (or more specifically a strain gradient of \(5 \times 10^{-4}\) over 1 micron distance). Reference [30] shows that with a 500 \( \mu \)m thick lead zirconate titanate (PZT) substrate and electrodes of side 600 \( \mu \)m, a strain gradient of \(~10^{-3}\) over a distance of 500 \( \mu \)m is feasible with application of 1.5 kV. This suggests that a strain gradient of \(~10^{-3}\) (\(~5 \times 10^{-4}\) if a single electrode is considered) over 1000 nm distance is possible with application of 3 V. The effective capacitance, \( C = \sim 26\) fF, assuming relative permittivity of 3000. Hence, the energy dissipated is \( \frac{1}{2} CV^2 \sim 117\) fJ for the electrical control of a scaled DW nanowire device with a footprint of 1000 nm \( \times \) 100 nm and electrode of footprint 1000 nm \( \times \) 1000 nm. In the SOT-clocked case, considering 70 nm square electrodes and 100 nm thick PZT, the total capacitance is \( C \sim 18\) fF while voltage needed for strain generation is 0.3 V following Ref [30]. This results in an energy dissipation of \(~1\) fJ.

For the charge current through the CoFe layer with resistivity of \(280\) \( \Omega \).nm [91] and a current density of \( J_c \sim 8.7 \times 10^{12}\) Am\(^2\) with a dimension of 1000 nm length, 100 nm width and 1 nm thickness for a clocking period of 6 ns the energy dissipated due to \( I^2R \) loss is \(~15\) pJ. For the SOT scheme, considering the resistivity of Pt to be \(100\) \( \Omega \).nm and charge current density \( J \sim 0.7 \times 10^{11}\) Am\(^2\) through the Pt layer of length 1000 nm, width 100 nm and thickness 1 nm for a clocking period of 20 ns, the \( I^2R \) loss is \(~100\) fJ. Therefore, energy consumption in the device is dominated by the current that produces the SOT. This can be further reduced if low damping materials such as iron garnets are used [92], which have the further advantage of avoiding current shunting through the magnetic layer. The difficulty will lie in optimizing the edge roughness and geometrical design of the racetrack to provide controllability at such low Gilbert damping. In summary, this clocked
domain wall device concept provides a pathway to realize novel energy efficient DW neuromorphic devices where reprogramming of synaptic weights can be performed at \(\sim100\) fJ per synapse during the learning phase and similarly small \(\sim100\) fJ per neuron during the inference phase of the neural network. In fact, during the inference phase, a neuron implemented with only CMOS devices (not a hybrid DW–CMOS device) would only need \(\sim\) few fJ per neuron and the synapses would consume no energy as they are non-volatile.

It is interesting to compare these numbers with alternative implementations of artificial neurons and synapses. The most important benefit of our approach for artificial synapses is the large reduction in energy consumption. The use of voltage control in conjunction with SOT drastically reduces the energy requirements versus purely spin torque domain wall-based devices [93]. Non-spintronic nanodevices can also provide multilevel synapses, such as oxide-based memristors [94] and phase change memories [79]. Programming such devices requires the physical motion of atoms in order to create or dissolve conductive filaments (oxide-based memristor) or to crystalize amorphous volumes of chalcogenide materials (phase change memory), which has an inherent energy cost, usually higher than picojoules even in highly scaled devices. On the other hand, these alternative technologies may provide more compact synapses than spintronic ones. Our solution therefore offers an extremely energy efficient approach to potentially implement real time learning-capable systems.

In contrast, the benefits of our approach for implementing artificial neurons is reduction in area (density). As neurons do not require non-volatility, CMOS-based solutions are typically used for neurons and have comparable energy consumption with ours. On the other hand, they typically require multiple transistors and several micrometre square of area [95].
## 2.5. Conclusion

The feasibility of an energy efficient voltage-controlled DW implementation of an artificial neuron and synapse was demonstrated using micromagnetic simulations. In this approach, modulation of perpendicular anisotropy with stress in combination with SOT or STT is used to program different synaptic weights as well as to mimic a neuron. Scaling this device to smaller dimensions (for example, ~500 nm × 50 nm × 1 nm) could result in much lower energy dissipation as well as high densities for comparable energy dissipation (for implementing neurons) compared to competing approaches. However, to avoid loss of controllability in deterministic positioning of the DW in the presence of thermal noise, careful optimization of material and device geometry are necessary. In summary, this work provides a pathway to the realization of energy efficient voltage controlled artificial neuron networks with real time learning capability and could stimulate more experimental work in this direction.
Chapter 3

Resonate and Fire Neuron with Fixed Magnetic Skyrmion

3.1. Introduction

Following the pioneering vision of Carver Mead [45], neuromorphic computing has garnered considerable interest in recent times due to its potential advantage in dealing with computational problems with ill-conditioned input data, adaptive nature of these systems to mitigate the effect of component failure and efficiency compared to fully Boolean logic based computation [52, 46, 96, 97]. Due to the complex and mixed analog-digital nature of the brain, a major hurdle towards developing neuromorphic computation platforms has been finding materials and devices to mimic brain-like behavior and developing architectures based on such systems. Current hardware artificial neural networks are mostly implemented with purely CMOS circuits and require large number of components to ensure robustness [46, 98]. This poses a major challenge for the scaling and energy efficiency of neuromorphic computation.

Nanomagnetic devices are one of the promising alternatives to implement neuromorphic computing and other non-von-Neumann like architectures due to their low energy consumption, nonlinear dynamics, and non-volatility. Many nanomagnetic devices that can potentially form the building blocks of neuromorphic computing: artificial neurons and synapses, have been proposed [52, 62, 99-105]. Among artificial neurons, most emulate the behavior of (leaky) integrate and fire type neurons [46,101]. In an integrate and fire type neuron, the membrane potential increases in response to an input spike and fires if it reaches a certain threshold [106]. Therefore, the firing frequency depends only on the strength of the stimulus in such integrate and fire neurons. However, in the brain, many neurons also feature
damped or sustained subthreshold oscillation [107-110] of membrane potential. Such neurons therefore show sensitivity towards the timing of stimulus. Consequently, a strong stimulus may not produce a spiking output if the incoming stimulus is not in phase with the oscillation of membrane potential, thus providing an inhibitory function. It also leads to many interesting phenomena such as fluctuation of spiking probability and selective communication [111]. Such “resonate-and-fire” neurons could also be useful in different neural networks where computation involves synchronized oscillation of several spin torque nano-oscillators (SNTOs) for pattern recognition [112]. Such networks come in different versions: for example, in ref. [112] patterns are encoded by frequency-shift keying (FSK) whereas in most other work [113-115] patterns are encoded with phase shift keying (PSK), but all of them could benefit from circuits able to detect synchrony through resonance.

In this work, we investigate the implementation of an artificial resonate-and-fire neuron by utilizing the magnetization dynamics of a fixed magnetic skyrmion in the free layer of a magnetic tunnel junction. Magnetic skyrmions (Figure 3.1(a)) are topologically protected spiral spin textures [116,117], which can be translated by applying small current [118,119] or reversed (in patterned dots) using a small voltage that controls the magnetic anisotropy [120,121]. Until now, this behavior has been leveraged to propose logic and memory devices based on magnetic skyrmions [57,122-125].

While nanomagnetic device based integrate and fire type neurons have been studied extensively, the resonate and fire type neuron proposed using nanomagnetic devices in this work, is unique to this paper and would be essential to compare the synchronization in arrays of nanomagnetic oscillators as described above for applications such as pattern recognition. While various schemes mimicking neuron and synapse activities have also been proposed utilizing current induced motion of skyrmions [126-129], neuromorphic devices based on moving skyrmions could have a large foot
print and are dissipative as they use current to move the skyrmions. We previously proposed nano-
magnetic memory devices utilizing voltage control of fixed magnetic skyrmions in the free layer of
a MTJ structure [120,130,131] that can alleviate these issues. In this paper, we use such a voltage
control of a fixed skyrmion based scheme to achieve the functionality of resonate and fire neurons.

The next section describes two “resonate and fire” neuron devices based on the novel mechanism
of resonant oscillations of a skyrmion core due to voltage control of anisotropy: direct voltage con-
trol of magnetic anisotropy (VCMA) and strain mediated voltage control of anisotropy in magne-
tostrictive materials. This is followed by a section explaining the modeling of voltage induced mag-
etization dynamics, followed by a discussion of skyrmion core oscillation, resonant behavior and
application of the “resonate and fire” functionality for detection of phase and frequency synchroni-
ization.

3.2. Device

Our proposed device is an MTJ structure in which the circular free layer hosts a fixed skyrmion.
We propose two different devices where application of a voltage modulates the perpendicular ani-
sotropy of the free layer through two different physical mechanisms. The anisotropy can be modu-
lated via voltage control of magnetic anisotropy [39,56,132,133] in the device shown in Figure
3.1(b) and voltage generated strain [32,134] in the device shown in Figure 3.1(c). Modulation of
perpendicular anisotropy in the system induces breathing of skyrmions. In other words, the skyrm-
ion core increases and decreases in size. This mimics the subthreshold damped oscillations of
resonate and fire neurons. The electrical resistance of the MTJ layer (R₂) depends on the magneti-
zation orientation of the free layer (i.e. the size of the skyrmion core) relative to that of the fixed
layer. For the sake of explanation, let us assume that the orientation of the magnetization of the fixed layer is antiparallel with respect to the one of the skyrmion core. As the skyrmion core expands during the breathing, more spins in the free layer will be antiparallel with respect to the fixed layer spins. Therefore, the resistance of the MTJ structure will increase. We thus propose to use a voltage divider consisting of a fixed resistor and the voltage controlled MTJ resistance to drive a CMOS buffer from OFF to ON state as shown in Figure 3.1.

Figure 3.1. (a) A core-up skyrmion, color code on the right shows the direction of the spin, (b) Proposed device structure operated with voltage control of magnetic anisotropy (VCMA) (c) MTJ structure stacked on PZT layer for strain control of magnetic anisotropy. Note that, CMOS buffer is driven by the MTJ resistance. Therefore, fixed layer magnetization needs to be opposite to that of skyrmion core. The resonate and fire behavior of this device is described by the simulations in Figure 3.3.

If the resistance of the MTJ stack increases during skyrmion expansion, potential drop across the MTJ resistance ($R_2$) will be higher. We can choose the ratio between $R_1$ and $R_2$ such that, at a given threshold this increase causes the transistor T1 to be turned on and generate a firing pulse. (NOTE: A similar behavior can be achieved by choosing the fixed layer magnetization orientation to be parallel with respect to the magnetization orientation of the skyrmion core and driving the CMOS
buffer by the potential drop across resistance \( R_1 \)). In short, if the skyrmion core size increases beyond a threshold, potential drop across the MTJ stack will produce an input voltage to the CMOS circuit that will exceed the threshold voltage, causing the buffer to “fire”, i.e. produce a high output. In ref. [135] resistance-area product of MTJ was found to be in the range of 225-650 \( \Omega \mu m^2 \) and typical tunnel magneto-resistance ratio between parallel and antiparallel configuration is 100% [135]. However, in this case, magnetization is oscillating between a skyrmionic state and a quasi-antiparallel state. Typical CMOS buffer have gating voltage in the range of 1 V. Hence, we can design a bias voltage \( V_{bias} \) and appropriate ratio for \( R_1 \) to \( R_2 \). It would be preferable to maximize \( R_1 \) and \( R_2 \) (to minimize standby power dissipation due to \( V_{bias} \)) while ensuring reasonable RC time constant for resonant operation of the device. In this work, for the sake of simplicity, we do not model the magnetoresistance change due to skyrmion breathing and the circuit dynamics of the CMOS buffer. Furthermore, at the scale of integrated circuits envisioned here, inductive effects are weak and not a concern for frequencies below 10 GHz and are therefore not modeled. In other words, in Figure 3.1 (corresponding to both device implementations for the resonate and fire neuron), we only model the magnetization dynamics of the skyrmions and set a threshold value of average magnetization along the z-axis (\( m_{z,\text{threshold}} \approx 0.8 \), magnetization is almost antiparallel to the free layer). For \( m_{z,\text{free}} > m_{z,\text{threshold}} \), we consider the CMOS buffer to be in the ‘ON’ or “high” state and “OFF” or “low” otherwise. This naturally gives rise to a firing output.
3.3. Methods

Micromagnetic simulation software MuMax3 [81] was used to perform the simulations where the magnetization dynamics is found by solving the Landau-Lifshitz-Gilbert (LLG) equation,

\[
\frac{\partial \vec{m}}{\partial t} = \vec{\tau} = \left(\gamma \frac{1}{1+\alpha^2} \right) \left( \vec{m} \times \vec{H}_{eff} + \alpha \left( \vec{m} \times \left( \vec{m} \times \vec{H}_{eff} \right) \right) \right)
\]  

where \( \vec{m} \) is the reduced magnetization (\( \vec{M}/M_s \)), \( M_s \) is the saturation magnetization, \( \gamma \) is the gyromagnetic ratio and \( \alpha \) is the Gilbert damping coefficient. The effective magnetic field \( \vec{H}_{eff} \) is given by,

\[
\vec{H}_{eff} = \vec{H}_{demag} + \vec{H}_{exchange} + \vec{H}_{anis} + \vec{H}_{thermal}
\]

Here, \( \vec{H}_{demag} \) is the effective field due to demagnetization energy, \( \vec{H}_{exchange} \) is the effective field due to Heisenberg exchange coupling and Dzyaloshinskii-Moriya Interaction (DMI). The DMI contribution to \( \vec{H}_{exchange} \) is given by [81]:

\[
\vec{H}_{DM} = \frac{2D}{\mu_0 M_s} \left[ \frac{\partial m_z}{\partial x} \hat{u}_y - \frac{\partial m_z}{\partial y} \hat{u}_x - \frac{\partial m_x}{\partial y} \hat{u}_y + \frac{\partial m_y}{\partial x} \hat{u}_x \right]
\]  

(3.3)

where \( m_z \) is the \( z \)-component of magnetization and \( D \) is the effective DMI constant.

The effective field due to the perpendicular anisotropy, \( \vec{H}_{anis} \), is expressed as [81],

\[
\vec{H}_{anis} = \frac{2K_{u1}}{\mu_0 M_s} (\hat{u} \cdot \vec{m}) \hat{u} + \frac{4K_{u2}}{\mu_0 M_s} (\hat{u} \cdot \vec{m})^3 \hat{u}
\]  

(3.4)

where \( K_{u1} \) and \( K_{u2} \) are first and second order uniaxial anisotropy constants and \( \hat{u} \) is the unit vector in the direction of the anisotropy (i.e. perpendicular anisotropy in this case). VCMA/strain effectively modulates the anisotropy energy density. The resultant change in uniaxial anisotropy due to VCMA/strain is incorporated by modulating \( K_{u1} \), while keeping \( K_{u2} = 0 \). For VCMA, this change is
given by \( \Delta k_{u1} = \Delta \text{PMA} = aE \). Here, \( a \) and \( E \) are respectively the coefficient of electric field control of magnetic anisotropy and the applied electric field. On the other hand, for strain,

![Table 3.1. Magnetic parameters used](image)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturation Magnetization (( M_{sat} ))</td>
<td>( 1 \times 10^6 ) A/m</td>
</tr>
<tr>
<td>Exchange Constant (( A_{ex} ))</td>
<td>( 2 \times 10^{-11} ) J/m</td>
</tr>
<tr>
<td>Perpendicular Anisotropy Constant (( K_{u1} ))</td>
<td>( 6 \times 10^5 ) J/m³</td>
</tr>
<tr>
<td>Gilbert Damping (( \alpha ))</td>
<td>0.03</td>
</tr>
<tr>
<td>DMI Constant (( D ))</td>
<td>3 mJ/m²</td>
</tr>
</tbody>
</table>

data given by \( \Delta k_{u1} = \Delta \text{PMA} = \frac{3}{2} \lambda \sigma \), where \( \lambda \) and \( \sigma \) are respectively the magnetostriction coefficient and the applied stress.

In order to reduce the effect of VCMA/strain on the fixed layer, the thickness of the fixed layer can be made lower compared to that of the free layer. This lower thickness ensures a high perpendicular magnetic anisotropy. Materials with low VCMA/magnetostriction co-efficient can be chosen for the fixed layer so that effect of voltage application is minimal. Additionally, one can use a synthetic antiferromagnetic [136] layer to increase magnetic stability of the fixed layer. These ensure the magnetization of the fixed magnetic layer does not rotate much due to VCMA or due to strain (if any) transferred to it. Therefore, we only simulate the magnetization dynamics of the free layer. The synthetic antiferromagnetic layer also offsets the dipolar interaction between the fixed and the free layer. Hence, we ignore anti-symmetric modification effects due to dipolar effects in our model. Exchange interaction and DMI can be modulated when an electric field is applied. However, these effects are minimal [131] and will only result in a small change in the breathing frequency and will not change the key results of our work significantly.
3.4. Results

a. Damped Oscillatory Behavior of Skyrmions

We simulated the magnetization dynamics in a 100 nm diameter nanodisk with thickness of 1 nm. Our geometry was discretized into $1 \times 1 \times 1 \text{ nm}^3$ cells. Using the parameter values in Table 3.1, the ground magnetization state was found to be a skyrmion. A triangular input spike of $\Delta \text{PMA}=1 \times 10^5 \text{ J/m}^3$ was applied with 50 ps rise and 50 ps fall time (NOTE: We use fast rise and fall time in the triangular pulse to simulate response to a near ideal pulse whereas sinusoidal inputs are used later for more realistic device simulations). Furthermore, we are mostly interested in frequency or phase synchronization of sinusoidal waveforms, but nevertheless choose triangular spikes initially due to similarity to actual spike like stimulus available in real neurons (though time scales for biological and artificial skyrmions resonate and fire neurons are vastly different). The momentary change in anisotropy causes the core of the skyrmions to expand and oscillate about the equilibrium state.

Figure 3.2. (a) Damped oscillation of a fixed skyrmion’s core due to stimulation with a single pulse [Red color line: Input spike, Blue color line: Output average magnetization along the perpendicular direction (z-axis)] (b) Modulation of breathing frequency by varying the interfacial parameters.
The oscillatory behavior can be seen from the net magnetization curve in Figure 3.2 (a). This imitates the subthreshold neuron oscillation of a resonant neuron. From this magnetization dynamics, the breathing frequency of the skyrmion can be determined. This is analogous to the eigen frequency (i.e. damped oscillation frequency) of the neuron. This information is important as an input spike train or sinusoid should have a frequency that nearly equals the eigen frequency to cause a neuron to resonate and spike. This breathing frequency strongly depends on the parameters of the system. Here, we determine the breathing frequency as a function of interfacial parameters PMA and DMI (Figure 3.2 (b)). This frequency can be easily tuned in the range of 1.8 GHz to 5.75 GHz.

In addition to interfacial parameters, one can use a DC bias voltage to change the PMA which will subsequently tune the frequency about which the skyrmion oscillates but this is not discussed here as it is beyond the scope of this paper.

b. Resonant behavior of Skyrmions

The skyrmion breathing frequency estimated in the last sub-section is now utilized to drive the skyrmion into resonance and show the resonate and fire behavior is very sensitive to this excitation frequency. We again start with triangular input pulses for reasons mentioned in the prior sub-section. Triangular pulses of $\Delta \text{PMA}=1.5 \times 10^5 \text{J/m}^3$ of time interval in a range of 0.35-0.50 ns was applied to the system. At the PMA chosen (in the absence of voltage applied) the core had a breathing frequency of approximately 2.86 GHz ($T=0.35$ ns). Skyrmion breathing of significant amplitude was observed when two input spikes were separated by an interval that falls in the range 0.43 ns to 0.46 ns. Breathing with diminishing amplitude was observed in other cases. The example in Figure 3.3(a) shows cases for 3 different time intervals between two successive input spikes: 0.35 ns, 0.45 ns and 0.50 ns.
Considering $M_z_{\text{threshold}} = 0.8$, a spiking output can be found for time interval of 0.45 ns, while no output spikes are found when the time interval was 0.35 ns and 0.50 ns. Other than the dependency on time interval the skyrmion core resonance is significantly sensitive to the amplitude of the input impulse. Lowering $\Delta P_{\text{MA}}$ to $1.4 \times 10^5 \text{ J/m}^3$ failed to produce any output as expected due to sub-

Figure 3.3. Resonant behavior: (a) Spike input (b) Sinusoidal input
threshold oscillation. However, increasing the ΔPMA to 1.75×10⁵ J/m³ lowered (instead of increasing) the firing rate from 4 for first 10ns to just 2. This is because we consider M_z_threshold = 0.8, which occurs when ~80% of the spins point upwards. As the core size is very large, peripheral (boundary) effects strongly influence the breathing dynamics which makes the behavior strongly nonlinear. Due to this, the correlation between the change of input magnitude and the spiking behavior is hard to predict at these limits. Hence, for triangular input with T=0.45ns, ΔPMA of 1.5×10⁵ J/m³ resulted in the best firing behavior.

While triangular spikes were used to illustrate the spiking behavior, a sinusoidal input pulse is more useful for many practical applications. Appropriate frequency sinusoidal inputs can also result in firing due to the same principle, i.e. a sinusoid of given amplitude, whose frequency is resonant with the eigen frequency produces the strongest spiking behavior. Sinusoids of different frequencies with peak to peak ΔPMA=0.96×10⁵ J/m³ were used as inputs. Strongest firing (4 spikes over 6 ns) was found around 2.86 GHz (time period of 0.35 ns) input frequency. Higher frequency (3 GHz or time period of 0.33 ns) and lower frequency (2.5 GHz or time period of 0.4 ns) resulted in weaker spiking behavior (less than 4 spikes over the same 6 ns). Further deviation in frequency from resonance: 3.3 GHz (time period of 0.3 ns) and 2.38 GHz (time period of 0.42 ns) resulted in no spiking behavior at all.

We note that the eigen frequency (for single triangular pulse) and resonant frequency for triangular and sinusoidal inputs all appear to be different. This is because the ΔPMA produced by the input voltage leads to a variation in the net PMA experienced by the breathing skyrmion, which in turn changes its frequency.

The change in perpendicular anisotropy is given by ΔPMA = αE. The VCMA co-efficient was found to be as large as 290 fJ/Vm experimentally [137] and 1800 fJ/Vm theoretically[138]. Using
a= 100 fJ/Vm, peak to peak $\Delta PMA=0.96 \times 10^5$ J/m$^3$ can be achieved using a peak voltage of 0.48 V, considering the MgO layer to be of 1 nm thickness. Energy dissipation per cycle will be $\frac{1}{2} CV^2$ considering energy dissipation is dominated by the energy required to charge the capacitive MgO layer. The capacitance of the MgO layer $C= 0.487$ fF for relative permittivity of 7 [139]). This translates into an energy dissipation of 56 aJ.

On the other hand, change in PMA achieved via strain is given by $\Delta PMA = \frac{3}{2} \lambda \sigma$. Magnetostrictive coefficient $\lambda$ is 37 ppm for CoFeB [140]. Stress cycles with magnitude ~1 GPa will be needed to drive this system to resonance, which is not practical. Materials with higher magnetostrictive coefficients exist. For example, FeGa has a coefficient of 300 ppm [141] while $\lambda$ can be as high as 1000 ppm for Terfenol-D [142,143]. The requirement of stress will be correspondingly lower (respectively ~167 MPa and ~50MPa) assuming that such material systems with such highly magnetostrictive materials also exhibit DMI (has not been studied so far). To generate 50 MPa stress, required voltage is 83.375 mV. We again consider energy dissipation is dominated by the energy required to charge the capacitive piezoelectric layer. The relative permittivity of the piezoelectric layer is taken to be 1000. Considering 100 nm thick PZT layer, capacitance $C=0.695$ fF. This gives rise to an energy dissipation of mere 2.4 aJ.

We note that, energy dissipation ~ femto-Joules (fJ) in the resistive elements (due to $V_{bias}$) will dominate energy dissipated in the scaled MTJ (~10-100 aJ) as well as the CMOS buffer (each CMOS device typically require ~100 aJ per switching event [17]). Thus, the total energy requirement will be ~ femto-Joule/spiking event.

These values are highly attractive in comparison to a purely CMOS implementation of the resonate and fire neuron. In reference [144], CMOS implementation of a resonate-and-fire neuron involves
capacitors approaching pico-Farad, leading to an energy consumption per firing event in the range of pico-Joules, an area of many micrometer square and resonant frequency of a few 10s of Hz. In fact, the proposed hybrid skyrmion-MTJ and CMOS buffer implementation of the resonate and fire neuron, is capable of resonant frequencies ~few GHz and is potentially 3 orders of magnitude more energy efficient/spiking event and potentially has 2 orders of magnitude higher density than that the all CMOS implementation [144] as shown in Table 3.2.

**Table 3.2.** Performance comparison of proposed hybrid nanomagnet-CMOS vs. all CMOS resonate and fire neuron [144].

<table>
<thead>
<tr>
<th>Performance metric</th>
<th>Hybrid fixed skyrmion-MTJ and CMOS buffer</th>
<th>All CMOS [144]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy dissipation/spiking event</td>
<td>~ femto-Joule</td>
<td>~pico-Joule</td>
</tr>
<tr>
<td>Density (area per device)</td>
<td>~0.01 micron²</td>
<td>~micron²</td>
</tr>
<tr>
<td>Resonance frequency</td>
<td>~ GHz</td>
<td>~10s Hz (can be designed to be much faster)</td>
</tr>
</tbody>
</table>

c. **Frequency and Phase synchronization detection of STNO oscillators**

Frequency and phase synchronization detection of coupled spin torque nano-oscillators (STNO) is an important component in neuromorphic computing schemes that implement associative memory[113-115]. However, in this work we do not investigate the mechanism of synchronization of such a system of oscillators which has been explored elsewhere (see for example, Ref [145-147]). Here, we show that our proposed device (single voltage controlled oscillator) can be used to detect the relative degree of phase and frequency synchronization of the outputs of two coupled STNOs (and, in general, any coupled oscillators). Thus, this work is focused on detection of phase locking and frequency synchronization of STNOs rather than exploring the mechanisms by which the synchronization is achieved. In this subsection, we consider the outputs of two STNOs (or two coupled oscillators) in general have been added together as:
Here $V_1$ ($V_2$), $f_1$ ($f_2$) and $\Phi_1$ ($\Phi_2$) are respectively the voltage, frequency and phase of the first and second oscillator output.

**Case I: Phase differs, frequency synchronized**

![Diagram](image)

**Figure 3.4. Phase detection:** As the phase difference increases the amplitude of the input decreases thus making it harder for the magnetization to reach the threshold limit for firing.

Here the two signals have no difference in frequency but have a phase difference ($\Delta \Phi$) of $\pi/18$ ($10^o$), $\pi/4$ ($45^o$), $\pi/3$ ($60^o$). We also include a random phase noise as follows:

$$\Phi_1 = \Phi_{10} + \Phi_{random,1}$$

$$\Phi_2 = \Phi_{20} + \Phi_{random,2}$$

$$\Delta \Phi = \Phi_1 - \Phi_2$$
The random phase noise added here is white noise. In addition to phase noise, coupling between the radial and in-plane modes of titling of spins in the skyrmions and additionally boundary effects (if the core oscillations have a large amplitude) could also lead to amplitude noise. This is intrinsically accounted for in our micromagnetic modeling of the skyrmion oscillations in the presence of thermal noise. However, for inputs to our “resonate and skyrmion” device we would need to include amplitude noise (in addition to phase noise we already include), if we are comparing nano-oscillators that are based on skyrmion oscillations. However, investigating this is beyond the scope of this paper.

In Figure 3.4 we show that output spikes several times when the phase difference is below a certain limit (e.g. 45° and below) and when the phase difference is larger (60°) the output fails to spike.

Case II: Frequency differs, (we assume that at t=0, Δ∅ =0)

![Figure 3.5. Frequency synchronization detection](image_url)
Here the two signals have different frequency but have no phase difference at \( t=0 \). Both signals are subjected to phase noise and the spiking output is analyzed over 10 ns. A signal with a frequency of 2.7 GHz (\( T=0.37\)ns) is chosen as the base signal. This frequency is slightly lower than the actual resonance frequency (2.86GHz) and intentionally chosen so to demonstrate robustness of the frequency synchronization detection to frequencies that are slightly off resonance. Successive signals added to it have frequencies of 5 GHz, 3.33 GHz, 2.7 GHz, 2.22 GHz and 1.82 GHz. When both frequencies are equal (2.7 GHz) 4 spikes are produced in 10 ns; when mismatched by \( \sim20\% \) (e.g. the 3.33 GHz and 2.22 GHz cases), less than 4 spikes are produced in 10 ns and finally with significant deviation (e.g. 5 GHz and 1.82 GHz) no output spike is produced. This suggests that further investigation into the skyrmion magnetization dynamics may reveal an appropriate input amplitude (and other conditions) where the number of output spikes over a given time window can provide an estimate of the degree of synchronization.

### 3.5. Conclusion

In this work, we studied novel nonlinear resonant dynamics of the core of a fixed skyrmion and showed that it has potential to lead to an energy efficient hybrid voltage controlled nanomagnetic device – CMOS device based circuit that can implement a “resonate and fire” neuron whereas prior work studied magnetic tunnel junctions (MTJs) hosting skyrmions primary for memory applications [120,121,132,148]. The energy dissipation of such a device per spiking event can potentially be \( \sim\)femto-Joules, which is 3 orders of magnitude (1000 times) less than an all CMOS implementation [144]. It can scale to much higher densities (\( \sim100 \) times less area) than an all CMOS implementation [144], while being able to exhibit resonance frequencies in the range of a few Giga-Hertz.
Furthermore, future work on combining such voltage controlled nanomagnetic frequency and phase synchronization detectors with voltage controlled nanomagnetic oscillators (not discussed in detail this paper) can lead to all voltage controlled-nanomagnetic devices (with some CMOS devices) based neuromorphic circuits that are potentially very energy efficient, dense and fast.
Chapter 4

Summary and Future Work

In summary, we have proposed and simulated two energy efficient voltage controlled spintronic/nanomagnetic device concepts that can (i) Implement Deep Neural Networks (DNNs) capable of real-time learning and (ii) Form building blocks for oscillatory neuromorphic computing.

This could stimulate future experimental work in fabrication and demonstration of such spintronic neuromorphic devices and evaluation of such proof-of-concept devices for practical applications.
References


