

Magnetic Tunnel Junctions as Gradient-based Neurons and Multi-state Synapses for Neuromorphic Computing

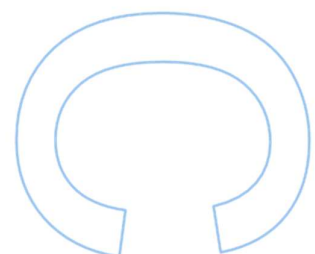
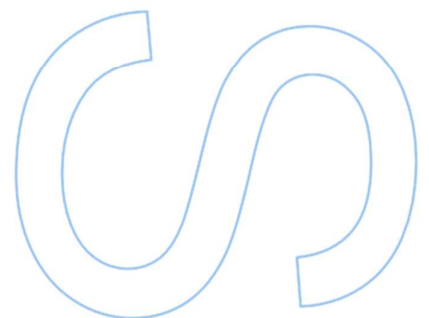
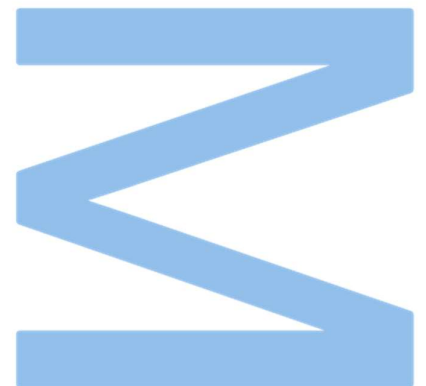
Catarina da Silva Miranda Almendra Pereira

Master in Engineering Physics

Department of Physics and Astronomy

Faculty of Sciences of University of Porto and Faculty of Engineering of University of Porto

2025



Magnetic Tunnel Junctions as Gradient-based Neurons and Multi-state Synapses for Neuromorphic Computing

Catarina da Silva Miranda Almendra Pereira

Dissertation carried out as part of the Master in Engineering
Physics

Department of Physics and Astronomy
2025

Supervisor

Alex Jenkins, Research Scientist, Iberian International Nanotechnology Laboratory

Co-supervisor

Tim Boehnert, Research Scientist, Iberian International Nanotechnology Laboratory

João Ventura, Research Scientist, Faculty of Sciences of University of Porto



“O chão da revolução não é a morte da diferença, nem o bom riso está à flor da mão. O chão da revolução é a morte do valor da diferença, de todas as diferenças”

Novas Cartas Portuguesas de
Maria Isabel Barreno, Maria Teresa Horta e Maria Velho da Costa

Acknowledgements

First of all, I want to thank my supervisors, Alex Jenkins, Tim Boehnert, and João Ventura, as well as Ricardo Ferreira, leader of the spintronics group, for allowing me to carry out this thesis at INL and for supporting me throughout the process from the very beginning. Without this opportunity, I wouldn't have been able to participate in two distinct but very interesting projects, both of which taught me a great deal and provided me with a strong foundation in these topics.

To Alex, I'm truly grateful for your orientation and patience with all my doubts and mistakes. Your kindness and sense of humor made everything easier, and I really appreciate the trust and opportunities you gave me.

To Tim, I thank you for all your support and generosity. Even though the topic was new to me, you were always available to help me understand and to think through solutions together.

To my supervisor from the University of Porto, João Ventura, I am grateful for always being present, even from a distance, and for the valuable suggestions whenever I needed guidance.

I am very grateful to the entire spintronics team for welcoming me so warmly. In particular, I thank Luana Benetti and Subhajit Roy for preparing the samples used in this work and for clarifying all my questions about the fabrication processes, and Ensieh Iranmehr for her patience with my doubts and her valuable feedback on the machine learning part. I also want to thank Davi Rodrigues from Politecnico di Bari, who came up with the idea, further developed it in the first part of this work, and was a constant presence throughout the process. Finally, I thank Xabier Castro, whose collaboration over a few weeks made the experience more productive and enjoyable.

To all my friends, who have always been there during these university years, helping me relax and unwind in the stressful moments. A special thanks to Ana, who has been by my side since childhood, and to André, who was a great support during this thesis.

And last but not least, to my family for their unconditional support. To my parents, for shaping who I am today, for always believing in me, and for encouraging me in everything I do. And finally, to my cat Kika, who has been with me since I was little and kept me company through countless long study sessions.

Resumo

Spintrónica é um campo dentro da física da matéria condensada que estuda o movimento do spin do eletrão, em vez de sua carga, permitindo o desenvolvimento de componentes capazes de operar com menor consumo de energia. Um exemplo que ganhou atenção significativa é a junção de túnel magnético (MTJ). Esta tese centra-se no desenvolvimento de MTJs para computação neuromórfica, uma área inspirada no cérebro humano e assente em dois elementos fundamentais: neurónios e sinapses artificiais. Ao contrário das arquiteturas de hardware tradicionais, a computação neuromórfica integra memória e capacidades de processamento no mesmo dispositivo, reduzindo as transferências constantes de dados entre unidades separadas de memória e lógica que caracterizam a arquitetura de von Neumann. Esta co-localização oferece benefícios claros para aplicações intensivas em dados, permitindo um processamento mais rápido e, potencialmente, menor consumo de energia. Assim, a combinação destas duas áreas constitui um caminho promissor para sistemas escaláveis e energeticamente eficientes. O trabalho desenvolvido foi realizado no Laboratório Ibérico Internacional de Nanotecnologia (INL) e está dividido em duas partes principais.

A primeira parte deste trabalho investigou neurónios artificiais através da análise da não-linearidade das curvas I–V em MTJs como funções de ativação e do desenvolvimento de um método experimental de diferenças finitas para o cálculo do seu gradiente. Dado que grande parte dos algoritmos de treino de redes neuronais depende do método gradiente descendente, a determinação precisa do gradiente é essencial. Deste modo, como o método das diferenças finitas precisa de dois dispositivos com respostas elétricas semelhantes, diferentes configurações de MTJs foram analisadas para obter funções de ativação não-lineares e estáveis. Assim, foram avaliados aspetos como a não linearidade, a assimetria e a reprodutibilidade, tendo-se verificado que, entre as amostras estudadas, a com maior concentração de nitrogénio na barreira apresentou o comportamento mais favorável. As curvas I–V e os gradientes das funções de ativação desta amostra foram implementados numa rede neuronal, tanto em simulação como em testes experimentais. Simulações em larga escala com o conjunto de dados MNIST, utilizando uma rede com duas camadas escondidas (128 e 64 neurónios), atingiram uma precisão de 96.8%. Estes resultados demonstram que a função de ativação desenvolvida e o método de cálculo do gradiente permitem alcançar elevada precisão no treino de redes neuronais mesmo em arquiteturas mais complexas e em multicamada.

A segunda parte deste trabalho centrou-se nos Quantizados Nano-Osciladores por Transferência de Spin (WSTNOs), que também exploram a dinâmica não linear dos MTJs para funcionarem como neurónios artificiais. As oscilações foram obtidas através de um estado de vórtice acima de uma determinada corrente limite, demonstrando o seu potencial como elementos de tipo neuronal, caracterizados pela relação não-linear entre a potência emitida pela oscilação e a current aplicada. Além disso, os WSTNOs são explorados para funcionalidades sinápticas. Ao alternar entre dois estados de resistência distintos, os MTJs fornecem a não volatilidade necessária para implementar pesos quantizados. As sinapses artificiais foram demonstradas pela ligação em série de cinco MTJs com diferentes dimensões, resultando em 32 estados de resistência distintos (2^5). Os testes de estabilidade térmica demonstraram que a configuração em série manteve um comportamento estável em todos os estados programados, sem alterações, com estabilidade a durar pelo menos 17 horas em alguns casos, destacando o seu potencial para aplicações em MRAM. Embora a variabilidade entre dispositivos continue a ser um desafio, estes resultados evidenciam um caminho promissor para a integração de WSTNOs com dupla funcionalidade de neurónio e sinapse.

Palavras-chave: spintrónica, junções de túnel magnéticas, nano-osciladores por transferência de spin, redes neuronais artificiais, neurónios artificiais, sinapses artificiais

Abstract

Spintronics is a field within condensed matter physics that focuses on manipulating the electron's spin, rather than its charge, enabling the development of components capable of operating at lower power. One example that has gained significant attention is the magnetic tunnel junction (MTJ). This thesis focuses on investigating MTJs for neuromorphic computing, an area inspired by the human brain and centered on two key elements: artificial neurons and synapses. Unlike traditional hardware architectures, neuromorphic computing integrates memory and processing capabilities within the same device, alleviating the constant data transfer between separate memory and logic units that characterizes the von Neumann architecture. This co-localization offers clear benefits for data-intensive applications, enabling faster processing and potentially lower energy overhead. The combination of these two areas provides a promising route toward scalable, energy-efficient systems. The work developed was carried out at the International Iberian Nanotechnology Laboratory (INL), and is divided into two main parts.

The first part of this work focused on the study of artificial neurons by analyzing the nonlinearity of I–V curves in MTJs and developing a finite difference method for experimentally calculating their gradient. Since most of neural network training algorithms relies on gradient descent, accurate gradient computation is essential. As the finite difference method requires two devices with similar electrical responses, different MTJ configurations were analyzed to obtain nonlinear and stable activation functions. Thus, aspects such as nonlinearity, asymmetry, and reproducibility were evaluated, and it was found that, among the samples studied, the one with the highest nitrogen concentration in the barrier exhibited the most favorable behavior. Its I–V curves and corresponding activation function gradients were implemented in a neural network, both in simulation and experimental tests. In large-scale simulations using the MNIST dataset and a network with two hidden layers (128 and 64 neurons), an accuracy of 96.8% was achieved. These results highlight that the developed activation function and gradient calculation method are capable of achieving high accuracy even when training multilayer neural networks.

The second part of this work focused on Weighted Spin-Torque Nano-Oscillators (WSTNOs), which also exploit the nonlinear dynamics of MTJs to operate as artificial neurons. Oscillations were obtained through a vortex state above a specific bias current, demonstrating their potential as neuron-like elements, characterized by a nonlinear relationship between the power emitted by the oscillation and the bias current. WSTNOs are explored

for synaptic functionalities. By switching between two distinct resistance states, MTJs provide the non-volatility required to implement quantized weights. Artificial synapses were demonstrated by connecting five MTJs of different dimensions in series, yielding 32 distinct resistance states (2^5). Thermal stability tests demonstrated that the series configuration maintained stable behavior across all programmed states without switching, with stability lasting at least 17 hours in some cases, highlighting its potential for MRAM applications. Although variability between devices remains a challenge, these results demonstrate a promising pathway for integrating WSTNOs with dual neuron-synapse functionality.

Keywords: spintronics, magnetic tunnel junctions, spin-torque nano-oscillators, artificial neural networks, artificial neurons, artificial synapses

Table of Contents

List of Tables	xi
List of Figures	xiii
List of Abbreviations	xvii
1. Introduction	1
1.1. Thesis Structure	2
2. State of the Art	5
2.1. Spintronics	5
2.1.1. Tunnel Magnetoresistance Effect	5
2.1.1.1. Simmons' Tunneling Model	6
2.1.1.2. Jullière Model and Spin-Dependent Tunneling	7
2.1.2. Magnetization Dynamics	10
2.1.3. Gibbs Free Energy	12
2.1.3.1. Zeeman Energy	12
2.1.3.2. Exchange Energy	13
2.1.3.3. Demagnetizing Energy	13
2.1.3.4. Dipolar Field Energy	14
2.1.3.5. Anisotropy Energy	14
2.1.3.6. Ruderman-Kittel-Kasuya-Yoshida (RKKY) energy	15
2.1.3.7. Exchange Bias Energy	15
2.1.4. MTJ structure	17
2.1.5. MTJ Thermal Stability	18
2.1.6. Magnetic Vortices	20
2.1.7. Spin-Transfer Torque	21
2.1.8. STNO with STT-induced Dynamics in Magnetic Vortices	23
2.2. Artificial Neural Networks	24
2.2.1. Network Architecture	26
2.2.2. Training Neural Networks	27
2.2.3. Backpropagation Algorithm	28
2.3. Spintronics for Neuromorphic Computing	30
2.3.1. Artificial Neurons	31
2.3.2. Artificial Synapses	32
3. Experimental Methods	35
3.1. Fabrication Techniques	35
3.1.1. Magnetron Sputtering	35

3.1.2. Etching-Ion beam Milling.....	37
3.1.3. Lithography	38
3.1.3.1. Direct Write Lithography.....	38
3.1.3.2. E-beam Lithography	38
3.2. Electrical Measurement Techniques	39
3.2.1. Current–Voltage and Transfer Curve Characterization	39
3.2.2. RF measurement	40
4. Spintronic Neurons and Finite Analog Difference Method.....	43
4.1. Spintronic Neurons	43
4.2. Analog Finite-Difference Gradient Method	44
4.3. I-V and Gradient Responses	46
4.3.1. Nitrogen Incorporation in the Barrier	47
4.3.2. Nitrogen Incorporation in the Free Layer	50
4.3.3. Comparison between Configurations	53
4.4. Neural Network Implementation: Experimental and Simulation Approach.....	56
5. Weighted Spin-Torque-Nano-Oscillators	65
5.1. Multi-State Devices	66
5.1.1. Time Retention Analysis	71
5.2. STNOs	73
6. Conclusions and Future Work.....	77
Bibliography	78

List of Tables

4.1. Values of r_{XX} and r_{YY} for different MTJ positions.	48
4.2. Maximum accuracy achieved by different neural network models with constant and decaying learning rates.	62
5.1. Device size, area, step height and symmetry analysis.	69
5.2. Expected vs experimental linear fit for $R = m(state \#) + b$. Values are estimate $\pm$ standard error.	72
1. Minor-loop sequences and resulting states.	88

List of Figures

2.1. Schematic of electron tunneling through a thin barrier. (Top) Wave view: an incident wave decays inside the barrier but partially reemerges on the other side if the barrier is thin enough. (Bottom) Particle view: the transmitted-to-incident amplitude ratio gives the tunneling probability [16].	5
2.2. Schematic representation of the tunneling barrier in an MTJ. On the left, the energy band diagram shows the applied bias voltage V , the barrier thickness t_b , and the barrier height ϕ . On the right, present the structural view of the MTJ, where two ferromagnetic electrodes are separated by an insulating barrier of thickness t_b under an applied voltage V [16].	6
2.3. Simulations of Simmons' model showing the typical relation of current density J as a function of voltage. a) Variation of the barrier height ϕ_0 while keeping the barrier thickness t_b constant. b) Variation of the barrier thickness t_b while keeping the barrier height ϕ_0 constant.	7
2.4. Spin-dependent tunneling in an MTJ, showing a schematic density of states for both magnetic electrodes in the parallel and antiparallel configurations of magnetizations.	8
2.5. Evolution of TMR values at room temperature for AlO_x and MgO barriers [21].	9
2.6. TEM images of MTJs with AlO_x and MgO barriers. a) $\text{Fe}/\text{AlO}_x/\text{NiFe}$ junction with AlO_x formed by plasma oxidation. b) $\text{CoFeB}/\text{MgO}/\text{CoFeB}$ MTJ after annealing, showing crystalline CoFeB templated by MgO [28].	10
2.7. Precession of magnetization in a magnetic field, where ω_L is the resonance frequency [29].	11
2.8. Oscillatory behavior of J_{RKKY} as a function of the distance between localized magnetic moments.	15
2.9. Shifted loop of partially oxidized cobalt particles, measured after cooling to 77 K in a 1 T field [32].	16
2.10 Stack of an MTJ structure with and with out a SAF [33].	18
2.11a) Schematic representation of a vortex magnetization configuration. Representation of the two main properties: b) polarity and (c) chirality [42].	20
2.12 Magnetic vortex states. a) Magnetic phase diagram of cylindrical dots with radius R and thickness L . Three magnetization states are stable: vortex (upper part), in-plane single domain (small thickness), and perpendicular single domain (small radii). Solid lines show equilibrium conditions, dashed line corresponds to the stability border of the vortex state, and the dot-dashed line indicates the equilibrium vortex core radius ($L_E \approx 18$ nm for permalloy). b) Scheme of vortex magnetization reversal in a cylindrical dot under an in-plane magnetic field H . The remanent state corresponds to a centered vortex [43].	21
2.13a) Spin-transfer torque promotes parallel alignment when current flows from the pinned to the free layer. b) Antiparallel alignment results when current flows in the reverse direction due to reflected spin-polarized electrons. Adapted from [29].	22
2.14 Magnetization dynamics under spin-transfer torque (STT). a) Precession under STT; b) relaxation when STT reinforces damping; c) steady-state precession when STT and damping balance; d) switching when STT overcomes damping. Adapted from [29].	23
2.15 Schematic of the different forces acting on the vortex core [46].	24
2.16 Comparison between a) biological and b) artificial neural networks [50].	25

2.17	Comparison between a artificial neural network (ANN) and a deep neural network (DNN).	26
2.18	a) Integrated output power and b) resonance frequencies as a function of the current bias. c) Neural Network architecture with one hidden layer for the MNIST dataset. d) Recognition accuracy of the inference dataset in function of the training epochs obtained from the NN constructed using the STO-activation function and a conventional ReLu function. Adapted from [57].	31
2.19	a) Measured $R(V)$ curve of seven identical MTJs connected in series. b) MNIST handwritten digit dataset used for testing. c) Neural network architecture with two hidden layers. d) Classification error rate on the test dataset as a function of the number of neurons per hidden layer, for different numbers of MTJ pairs used per multi-state synaptic cell [61].	33
2.20	d) Measured oscillation power (P) of the WSTNO as a function of the input voltages (V_1 and V_2) for each possible combination of the non-volatile states of the MRAMs: parallel and antiparallel. The contour of $1 \mu\text{W}$ output power is shown in red. e–h Analytical prediction of the WSTNO behavior as a function of the two input voltages for each state combination [62].	34
3.1.	a) Magnetron sputtering scheme adapted from [64] and b) MTM system used at INL for MTJ deposition [65].	36
3.2.	a) Schematic of the ion milling process and b) Nordiko ion milling system available at INL [65].	37
3.3.	a) DWL lithography system and b) e-beam lithography system at INL cleanroom [66].	39
3.4.	Experimental setup: a) Sample, probes, magnet, optical microscope. b) Setup scheme for I–V and $R(H)$ curve measurements.	40
3.5.	Schematic representation of the experimental setup for the RF measurement.	41
4.1.	Schematic of the analog finite difference method used to estimate the gradient of the MTJ I-V characteristic. a) shows schematically the proposed architecture using a pair of nominally identical MTJs that are simultaneously biased with slightly different input currents, I and $I + \Delta I$, where $\Delta I = 100 \mu\text{A}$. b) Corresponding measured voltages (V) in function of current (I). c) Comparison between the experimental (red) and numerical (green) gradients of the I-V characteristic in function of current (I).	45
4.2.	I-V curve for S_A with a radius of 20000 nm showing that above $\pm 1 \text{ V}$ the MTJ start to exhibit memristive effects.	47
4.3.	Schematic representation of the MTJ stacks used for I-V measurements. a) Stack A (S_A), b) Stack B (S_B), c) Stack C (S_C), and d) Stack D (S_D). The MgO barrier is shown in grey, while nitrogen doping in S_A and S_B is highlighted in yellow. Schematics are not in scale.	47
4.4.	Dependence of TMR on RA for Stack S_A (green) and Stack S_B (orange).	49
4.5.	Device-to-device variability analysis of S_A and S_B . a) Standard deviation of voltage for each diameter according to the designation of Table 4.1. b) Average of standard deviations across all geometries.	49
4.6.	Measured I-V curves (blue), experimental gradient (red), and numerical gradient (green) for the two magnetization states of the device: a) parallel (P) and b) antiparallel (AP).	50
4.7.	Dependence of TMR on RA for the Stack C (S_C) and Stack D (S_D).	51
4.8.	Device-to-device variability analysis of S_C and S_D . a) Standard deviation of voltage for each diameter according to the designation of Table 4.1. b) Average of standard deviations across all geometries.	52
4.9.	Dependence of TMR on RA for S_D with different annealing temperatures.	52

4.10	Measured I–V curves (blue), experimental gradient (red), and numerical gradient (green) for the two magnetization states of the device: parallel (P) in panels (a, c, e) and antiparallel (AP) in panels (b, d, f). Results are shown for Stack D annealed at 330 °C (a, b), 250 °C (c, d), and 200 °C (e, f).	53
4.11	Comparison of asymmetry errors for S_A and S_D at different annealing temperatures.	55
4.12	Comparison of nonlinearity metrics ($\overline{\text{MSE}}$ and $\overline{\text{MSE}}^{\text{max}}$) for S_A and S_D at different annealing temperatures.	56
4.13	Training simulation of neural network using MTJ-based neurons from S_A and S_D for Iris classification. a) Architecture with one hidden layer. b) Activation functions obtained from the I-V curves for S_A (green), S_D (dark blue), and S_A restricted to the low-bias region (grey). c), d) Mean validation accuracy and loss, respectively, obtained via 5-fold cross-validation over 500 epochs. Shaded areas indicate the standard deviation across folds. Green curves correspond to networks using neurons derived from S_A , blue curves correspond to networks using neurons derived from S_D , and grey curves correspond to networks derived from S_A but restricted to the low-bias region.	58
4.14	Inputs distribution of the activation functions before (blue) and after (orange) training.	59
4.15	Implemented protocol for the experimental and simulated approach.	60
4.16	Experimental and simulated training of a neural network using MTJ-based neurons from S_A for Iris classification. a) Architecture with a single hidden layer (HL 1) of five distinct MTJ neurons represented in blue, b) Architecture with two hidden layers represented in orange (HL 1 and HL 2), where the second layer is identical to the first. (c, d) Mean validation accuracy obtained via 5-fold cross-validation. Shaded areas indicate the standard deviation across folds. Blue curves correspond to networks with one hidden layer (light = experimental, dark = simulation), while orange curves correspond to two hidden layers (simulation only). The Insets represent the validation loss over the epochs. c) Training with a constant learning rate of 0.002 over 300 epochs. d) Training with a linearly decaying learning rate from 0.01 to 0.0088 over 120 epochs.	61
4.17	Training simulation of a neural network using MTJ-based neurons and tanh activation for MNIST classification. a) Architecture with two hidden layers with 128 and 64 neurons, respectively. b) Mean validation accuracy obtained via 5-fold cross-validation over 50 epochs. Shaded areas indicate the standard deviation across folds. The Inset represents the validation loss over the epochs. Orange curves correspond to results using MTJ neurons, while green curves correspond to results using the tanh activation function.	62
5.1.	Layout of the configuration tested with two series of 5 MRAMs and a STNO.	65
5.2.	Comparison of the transfer curves for a serie of 5 MRAMs for a) S_1 and b) S_2	67
5.3.	Transfer curve of a series of 5 MRAMs for S_1 , where the red dots indicate the reversal fields, and the values next to each jump represent their height.	68
5.4.	Resistance values as a function of decimal state, with expected values shown in blue and experimental data in red. Dashed lines indicate the linear fit for each results.....	71
5.5.	Time retention analysis. The blue curve represents the transfer curve of a series of five MRAM devices under a constant bias of 10 μA . The resistance stability over time for a programmed state is shown in green and red: the red curve corresponds to a bias of 10 μA , while the green curve corresponds to a bias of 1 mA.....	73
5.6.	$R(H)$ for a vortex configuration with a diameter of 350 nm under a bias current of 4 mA.	74
5.7.	$R(H)$ for a vortex configuration with a diameter of 1000 nm under a bias current of a) 4 mA, b) 8 mA and c) 12 mA.	74

5.8. Power Spectral Density (PSD) in function of frequency represented $\mu\text{W Hz}^{-1}$ a) and dBm Hz^{-1} b) for a measured spectrum with 17 mA of bias current.	76
5.9. STNO characterization. a) Integrated power, b) resonance frequency and c) linewidth as a function of the bias current, obtained from a Lorentz fit.	76
1. I-V curves and corresponding gradients for S_A used in the neural network implementation, with (a-e) representing P1 to P5.	87
2. Magnetic field minor loops to set the chain of 5 MTJs in the 32 resistance states.	87

List of Abbreviations

AFM	Antiferromagnetic
AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Antiparallel
FM	Ferromagnetic
MRAM	Magnetoresistive Random-Access Memory
MTJ	Magnetic Tunnel Junction
P	Parallel
RKKY	Ruderman-Kittel-Kasuya-Yoshida
STNO	Spin-Torque Nano-Oscillator
STT	Spin-Transfer Torque
TMR	Tunnel Magnetoresistance
WSTNO	Weighted Spin-Torque Nano-Oscillator

1. Introduction

In recent years, the demand for artificial intelligence (AI) systems has grown significantly, driven by their broad range of applications such as image recognition, natural language processing, and autonomous systems. Artificial neural networks (ANNs) have been at the center of this progress due to their capacity to model complex patterns and decision-making processes, reaching high values of accuracy [1]. ANNs are composed of artificial neurons that consist on computational units that process input signals and synapses, which represent weighted connections that modulate how information is transmitted between neurons.

Despite their advantages, ANNs impose significant computational and memory demands. In traditional von Neumann architectures [2], the physical separation between memory and processing units creates the well-known von Neumann bottleneck, which severely limits efficiency [3]. As neural networks continue to scale in size and complexity to address more advanced tasks, both energy consumption and latency increase exponentially. At the same time, the continued miniaturization of transistors is approaching its physical limits, approaching to the end of Moore's Law.

To overcome these challenges, analog computing is an alternative particularly promising because it leverages the intrinsic physical properties of devices to perform computations directly in hardware, thereby reducing the reliance on energy-intensive digital operations [4–6]. Within this context, neuromorphic computing has emerged as a powerful approach inspired by the brain's architecture and operating principles. Among the various candidate technologies, spintronics stands out due to its inherent non-volatility, low power consumption, scalability, and compatibility with existing CMOS technology [7, 8]. Spintronic devices, such as magnetic tunnel junctions (MTJs), naturally exhibit nonlinear behavior and operate effectively at the nanoscale, making them well-suited for neuromorphic applications [9, 10].

However, analog devices are still under active development, as their behavior is affected by complex and non-ideal physical dynamics, like stochasticity, drift, and variability across devices. This mismatch frequently limits performance and adaptability. Additionally, most training approaches focus on the algorithmic level, thereby neglecting the potential for in situ or adaptive learning directly in hardware [11–14].

This thesis aims to address some of these issues by exploring the use of MTJs in two distinct neuromorphic roles: as artificial neurons and as artificial synapses. First, the non-linear current–voltage (I-V) response of MTJs was investigated as a basis for activation functions, which are essential for the expressiveness and accuracy of neural networks. Since gradient calculation is a critical component of learning algorithms, a finite-difference approach to approximate gradients efficiently was proposed. This method enables the integration of complex hardware-based nonlinearities into neural networks without requiring explicit analytical gradients, thereby reducing computational and energy costs.

Second, the use of MTJs as multi-bit memories functioning as synapses was investigated, along with their potential integration with an spin-torque nano-oscillator (STNO) acting as an artificial neuron. In this context, MTJs can serve as magnetic random-access memory (MRAM), where their non-volatility stems from distinct magnetization states. A single device can, thus, encode a binary weight, corresponding to the parallel or antiparallel configuration. In this work, the connection of multiple MTJs in series will be explored, where the number of achievable weight states increases exponentially, enabling 2^N distinct values for N devices. Furthermore, these MTJs are also fabricated to exhibit vortex states, which can be exploited in the form of STNOs to emulate neuronal behavior through their tunable oscillatory dynamics.

Thus, by experimentally exploring the properties of these spintronic devices, different approaches to artificial neurons and synaptic functionalities are investigated. These findings aim to help narrow the gap between the physical hardware implementation of ANNs simulated on traditional computing hardware, highlighting the potential of spintronic architectures for scalable and energy-efficient neuromorphic systems.

1.1. Thesis Structure

This section describes the structure of the thesis to provide a more comprehensive overview of how it is organized.

Chapter 2 presents a theoretical overview of the two main fields studied: Spintronics and ANNs. In the Spintronics section, the tunnel magnetoresistance effect is explained, with special attention to MTJs, including their structure, magnetization dynamics, and spin-transfer torque, particularly in the context of vortex states. The ANN section provides an

overview of network architectures and training methods, with emphasis on backpropagation. The chapter concludes with examples of applications of spintronic MTJs in neuromorphic computing relevant to this study.

Chapter 3 explains the main experimental techniques used at Iberian Nanotechnology Laboratory (INL) for the fabrication and characterization of the MTJs.

Chapter 4 presents the results of the first part of this work, which introduces the finite difference calculation method and explores the nonlinear I–V characteristics of different MTJ configurations. The results obtained when implementing it into different neural network architectures, through simulations and experimental tests, are also reported.

Chapter 5 shows the results and analysis of multi-state devices obtained from a series connection of five MTJs, as well as a brief discussion on the characterization of STNOs.

Chapter 6 summarizes the main conclusions drawn from this thesis and outlines possible directions for future research to further develop this work.

2. State of the Art

2.1. Spintronics

Spintronics is a branch of physics that explores the spin degree of freedom as a means to control electrical current and broaden the applications of electronic devices. The field emerged in the 1960s, pioneered by a research group at IBM led by Leo Esaki, a Japanese physicist who later received the Nobel Prize in Physics in 1973 for his discovery of electronic tunneling [15]. This phenomenon, along with other fundamental principles, will be addressed in the following chapter.

2.1.1 Tunnel Magnetoresistance Effect

The tunnel magnetoresistance (TMR) effect is a quantum mechanical phenomenon that occurs in magnetic tunnel junctions (MTJs). An MTJ consists of two ferromagnetic (FM) layers separated by a thin insulating barrier, typically aluminum oxide (AlO_x) or magnesium oxide (MgO). Although classically insulating, such barriers permit electron transport through quantum tunneling when the barrier thickness is sufficiently small (on the order of 1–2 nm).

Figure 2.1 illustrates this tunneling process. From a wave perspective, the incident electron wave decays exponentially inside the barrier, but if the barrier is thin enough, part of the wavefunction reaches the opposite side, resulting in a finite tunneling probability.

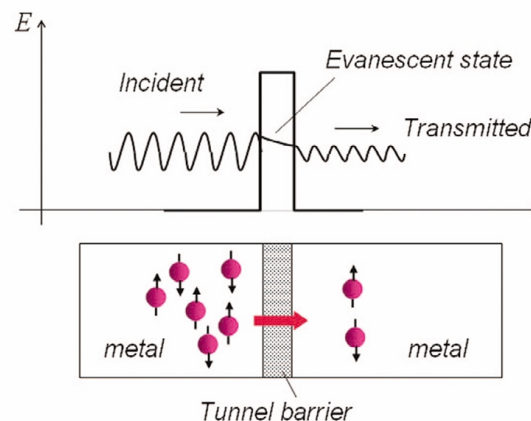


Figure 2.1: Schematic of electron tunneling through a thin barrier. (Top) Wave view: an incident wave decays inside the barrier but partially reemerges on the other side if the barrier is thin enough. (Bottom) Particle view: the transmitted-to-incident amplitude ratio gives the tunneling probability [16].

Applying a voltage V across the MTJ shifts the Fermi level of one electrode relative to the other by eV , e being the electron charge, enabling current flow (I) due to tunneling between occupied states on one side and unoccupied states on the other, as represented in Figure 2.2. Using Fermi's golden rule, the tunneling current can be written as

$$I(V) \propto |T|^2 \int_{-\infty}^{+\infty} D_1(E) D_2(E + eV) [f(E) - f(E + eV)] dE, \quad (2.1)$$

where $D_{1,2}(E)$ are the energy-dependent densities of states (DOS) of electrodes 1 and 2, $f(E)$ the Fermi-Dirac distribution, and $|T|^2$ is the transmission probability. At low temperature and bias ($V \rightarrow 0$, $T \rightarrow 0$), the conductance (G) becomes

$$G = \left. \frac{dI}{dV} \right|_{V \rightarrow 0} \propto |T|^2 D_1(E_F) D_2(E_F), \quad (2.2)$$

showing that the conductance depends directly on the product of the DOS at the Fermi energy level (E_F) of both electrodes.

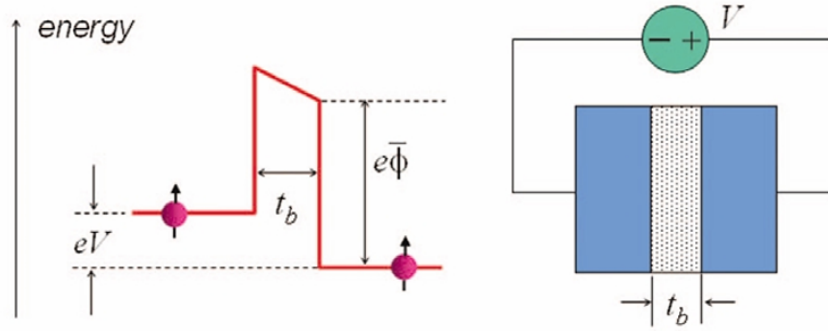


Figure 2.2: Schematic representation of the tunneling barrier in an MTJ. On the left, the energy band diagram shows the applied bias voltage V , the barrier thickness t_b , and the barrier height ϕ . On the right, present the structural view of the MTJ, where two ferromagnetic electrodes are separated by an insulating barrier of thickness t_b under an applied voltage V [16].

2.1.1.1 Simmons' Tunneling Model

To quantitatively model electron tunneling, the insulating barrier is often approximated as a rectangular or trapezoidal potential. In 1963, Simmons proposed a model for a symmetric rectangular barrier, where the current–voltage (I – V) relation is given by [17]:

$$I(V) = A \left\{ k_1 \left(\phi_0 - \frac{eV}{2} \right) \exp \left[-k_2 \sqrt{\phi_0 - \frac{eV}{2}} \right] - k_1 \left(\phi_0 + \frac{eV}{2} \right) \exp \left[-k_2 \sqrt{\phi_0 + \frac{eV}{2}} \right] \right\}, \quad (2.3)$$

where A is the area, ϕ_0 is the barrier height, t_b the barrier thickness, and

$$k_1 = \frac{e}{2\pi\hbar t_b^2}, \quad (2.4)$$

$$k_2 = -\frac{4\pi t_b}{\hbar} \sqrt{2m_e}, \quad (2.5)$$

with $\hbar$ Planck's constant and m_e the electron mass.

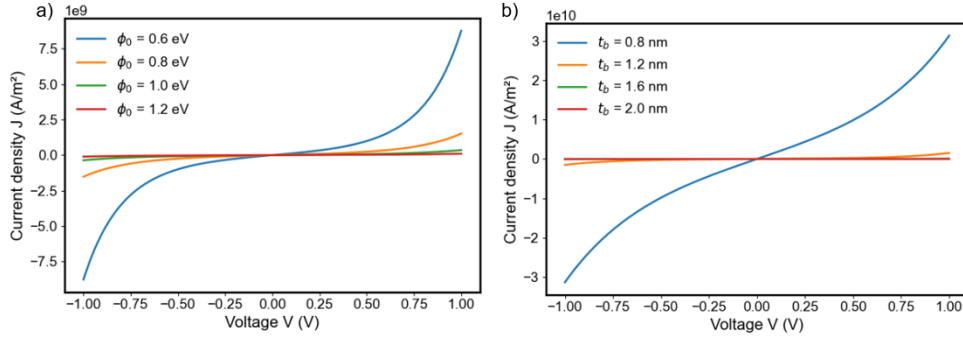


Figure 2.3: Simulations of Simmons' model showing the typical relation of current density J as a function of voltage. a) Variation of the barrier height ϕ_0 while keeping the barrier thickness t_b constant. b) Variation of the barrier thickness t_b while keeping the barrier height ϕ_0 constant.

Figure 2.3a and Figure 2.3b present simulations of Simmons' model of the current density J as a function of voltage. In Figure 2.3a, the barrier height ϕ_0 is varied while the thickness t_b is kept constant ($t_b = 1.2$ nm), whereas in Figure 2.3b the thickness t_b is varied while the barrier height ϕ_0 remains fixed ($\phi_0 = 0.8$ eV). The curves clearly exhibit a pronounced nonlinear behavior, which arises from the exponential dependence of the tunneling current. For asymmetric barriers, the Brinkman–Dynes–Rowell model [18] generalizes Simmons' approach.

2.1.1.2 Jullière Model and Spin-Dependent Tunneling

While Simmons' model describes charge tunneling, it does not include the spin degree of freedom. The spin dependence of tunneling in MTJs was first introduced by Jullière in 1975 [19]. The key assumption is that electron spin is conserved during tunneling: spin-up ($\uparrow$) electrons in one electrode tunnel into spin-up states in the other, and likewise for spin-down ($\downarrow$). In ferromagnets, exchange splitting makes the DOS at E_F asymmetric, with majority spins having higher DOS than minority spins [20]. This leads to two possible magnetic configurations in MTJs: parallel (P), when the magnetizations of the two ferromagnetic electrodes are aligned, and antiparallel (AP), when they are opposed. According to Fermi's golden rule (Equation 2.1), the tunneling current for a given spin channel is proportional to the product of the DOS of the initial and final ferromagnetic

electrodes. Jullière's model formalizes this relationship as [19]:

$$J_P \propto D_1^\uparrow(E_F)D_2^\uparrow(E_F) + D_1^\downarrow(E_F)D_2^\downarrow(E_F), \quad (2.6)$$

$$J_{AP} \propto D_1^\uparrow(E_F)D_2^\downarrow(E_F) + D_1^\downarrow(E_F)D_2^\uparrow(E_F), \quad (2.7)$$

where $D_i^{\uparrow\downarrow}(E_F)$ is the spin-resolved DOS at the Fermi level of electrode i .

Figure 2.4 schematically depicts this spin-dependent tunneling process. In the P configuration, majority-spin states in both electrodes overlap at the Fermi energy, enabling a large tunneling current, while minority-spin overlap contributes only weakly. In the AP configuration, the majority spins in one electrode face the minority spins in the other, reducing the overlap at E_F for both spin channels. As a result, the total tunneling conductance is higher in the P state than in the AP state, because electrons more easily find available final states for tunneling when the magnetizations are aligned.

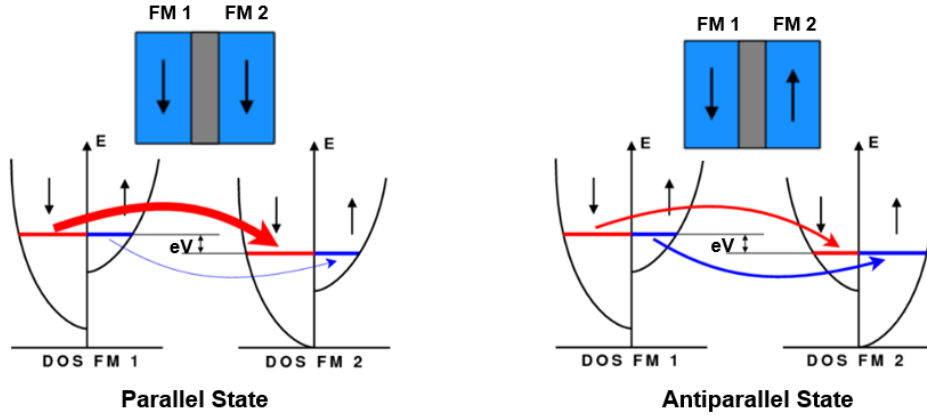


Figure 2.4: Spin-dependent tunneling in an MTJ, showing a schematic density of states for both magnetic electrodes in the parallel and antiparallel configurations of magnetizations.

In this way, the spin polarization of electrode i is defined as

$$P_i = \frac{D_i^\uparrow(E_F) - D_i^\downarrow(E_F)}{D_i^\uparrow(E_F) + D_i^\downarrow(E_F)}. \quad (2.8)$$

From this, the tunnel magnetoresistance ratio is given by

$$\text{TMR} = \frac{R_{AP} - R_P}{R_P} = \frac{G_P - G_{AP}}{G_{AP}}. \quad (2.9)$$

Finally, assuming spin polarizations P_1 and P_2 for the two electrodes, Jullière's model predicts:

$$\text{TMR} = \frac{2P_1P_2}{1 - P_1P_2}. \quad (2.10)$$

This expression highlights the central role of spin polarization in determining the TMR effect.

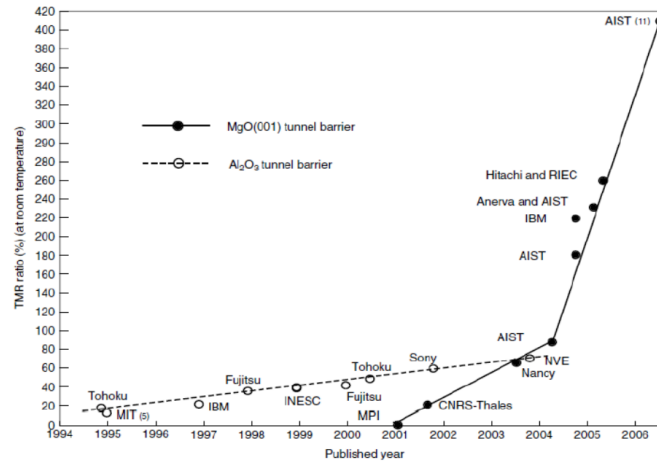


Figure 2.5: Evolution of TMR values at room temperature for AlO_x and MgO barriers [21].

In 1975, Jullière found a TMR ratio of 14% at 4 K in a $\text{Fe}/\text{Ge}/\text{Co}$ junction [19]. Two decades later, significant advances in MTJs emerged through the use of amorphous Al_2O_3 (AlO_x) tunnel barriers. These structures enabled the observation of TMR effects at room temperature, with $\text{Co}/\text{AlO}_x/\text{Co}$ junctions exhibiting TMR values exceeding 15% [22]. Over the following decade, intensive efforts focused on improving the TMR ratio by optimizing electrode materials—such as polycrystalline CoFe and amorphous CoFeB —and refining the quality of the tunnel barrier via enhanced deposition methods and post-deposition annealing [23]. Such improvements led to substantial increases in TMR values, reaching up to 70% at room temperature in $\text{CoFeB}/\text{AlO}_x/\text{CoFeB}$ junctions fabricated using standard sputtering techniques. For instance, with a CoFeB composition of $\text{Co}_{60}\text{Fe}_{20}\text{B}_{20}$, a spin polarization factor of $P = 0.61$ was deduced from the measured TMR using Jullière’s equation.

The focus of TMR enhancement shifted significantly in the early 2000s when theoretical studies predicted remarkably high TMR ratios in MTJs with crystalline MgO tunnel barriers [24, 25]. These predictions were based on the coherent tunneling model, wherein the symmetry of the electronic wavefunctions in Fe electrodes matches the evanescent states within a crystalline MgO (001) barrier. Specifically, the tunneling of Δ_1 symmetry states was shown to be highly spin-selective due to their slow decay within the MgO barrier, compared to other symmetries such as Δ_2 and Δ_5 .

Experimental verification followed swiftly in 2004, when two independent studies—one

utilizing molecular beam epitaxy (MBE) and the other employing sputtering with under-layer texturing—reported room-temperature TMR ratios of over 180% and 220%, respectively [23, 26]. These discoveries catalyzed a surge of research into MgO-based MTJs, establishing them as the benchmark for high-performance spintronic devices.

Further advances combined the use of amorphous CoFeB with crystalline MgO barriers, exploiting the ability of CoFeB to crystallize during post-deposition annealing and align with the MgO (001) texture. This combination led to ever-increasing TMR values, with a record value of 604% measured at room temperature [27]. Theoretical studies explained this behavior by showing that only the Δ_1 band has a 100% spin polarization at the Fermi level, while other bands contribute with lower polarizations due to the presence of both majority and minority spin states.

Despite theoretical predictions of TMR ratios exceeding 1000% — based on ideal crystalline structures and thick tunnel barriers — the practical realization of such values remains elusive due to structural imperfections and the contribution of less spin-polarized bands as the barrier thickness decreases. Nevertheless, the TMR effect remains the most utilized spin-dependent transport phenomenon, particularly in applications such as magnetoresistive random-access memory (MRAM), because of its compatibility with CMOS electronics and its large signal amplitude at room temperature.

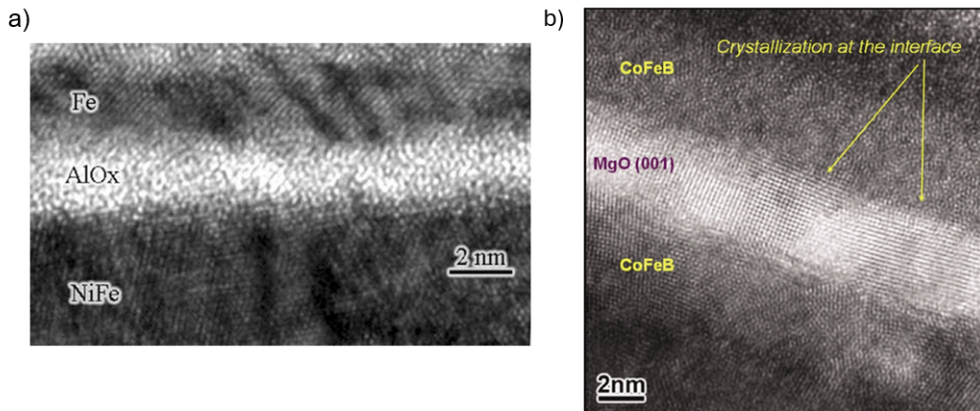


Figure 2.6: TEM images of MTJs with AlOx and MgO barriers. a) Fe/AlOx/NiFe junction with AlOx formed by plasma oxidation. b) CoFeB/MgO/CoFeB MTJ after annealing, showing crystalline CoFeB templated by MgO [28].

2.1.2 Magnetization Dynamics

Applying a magnetic field to an MTJ can induce switching between the antiparallel and parallel magnetic configurations. By fixing the magnetization of one ferromagnetic layer

(the pinned layer) and applying a magnetic field within a specific range, it is possible to switch the magnetization of the other layer (the free layer).

This switching mechanism is governed by the torque exerted on a magnetic moment $\mathbf{m}$ when placed in a magnetic field $\mathbf{H}$ that is not aligned with it. The torque is given by

$$\boldsymbol{\tau} = \mu_0 \mathbf{m} \times \mathbf{H},$$

where μ_0 is the vacuum magnetic permeability. As a result, the magnetization undergoes precession around the direction of the magnetic field (Fig. 2.7).

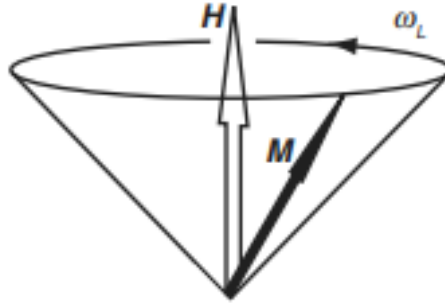


Figure 2.7: Precession of magnetization in a magnetic field, where ω_L is the resonance frequency [29].

The $\mathbf{M}$ is the magnetization vector, that is, the magnetic moment per unit of volume, and can be defined as $\mathbf{M} = M_S \mathbf{m}$, where M_S is the magnetization saturation (expressed in A/m). The time evolution of the magnetization vector $\mathbf{M}$ is described by the Larmor equation [29]:

$$\frac{d\mathbf{M}}{dt} = \gamma \mu_0 \mathbf{M} \times \mathbf{H},$$

where $\gamma = \frac{g\mu_B}{\hbar}$ is the gyromagnetic ratio. Here, g is the Landé g -factor (dimensionless), $\mu_B = \frac{e\hbar}{2m_e}$ is the Bohr magneton, and $\hbar$ is the reduced Planck constant. The magnetic field $\mathbf{H}$ includes contributions from external sources, exchange interactions, magnetic anisotropy, and demagnetizing effects; therefore, it is normally defined as the effective magnetic field $\mathbf{H}_{\text{eff}}$.

However, this conservative equation neglects energy dissipation. In real magnetic systems, precessional motion is accompanied by damping, which is necessary to align the magnetization vector along the direction of the effective field. This dissipation is accounted for in the Landau–Lifshitz–Gilbert (LLG) [29] equation which introduces a phenomenological damping parameter α :

$$\frac{d\mathbf{M}}{dt} = -\gamma \mu_0 \mathbf{M} \times \mathbf{H}_{\text{eff}} + \frac{\alpha}{M_s} \mathbf{M} \times \frac{d\mathbf{M}}{dt}, \quad (2.11)$$

where α is the dimensionless Gilbert damping constant. The first term governs the precessional motion, while the second term causes the magnetization to spiral inward and eventually align with $\mathbf{H}_{\text{eff}}$.

2.1.3 Gibbs Free Energy

The effective magnetic field $\mathbf{H}_{\text{eff}}$ is derived from the total Gibbs free energy E of the magnetic system. The effective field is then given by the functional derivative of the total energy with respect to the magnetization:

$$\mathbf{H}_{\text{eff}} = -\frac{1}{\mu_0 M_s} \frac{\delta E}{\delta \mathbf{m}}. \quad (2.12)$$

This formulation reflects the tendency of the system to evolve toward a local or global energy minimum. This energy encompasses several physical contributions, depending on the specific material structure and environment. The total energy can be expressed as [30]:

$$E = E_Z + E_{\text{dem}} + E_{\text{an}} + E_{\text{bias}} + E_{\text{dip}} + E_{\text{RKKY}} + E_{\text{ex}}, \quad (2.13)$$

where E_Z is the Zeeman energy generated by the interaction with an external magnetic field, E_{an} is the magnetocrystalline or anisotropy energy, E_{dem} the demagnetizing field energy, E_{bias} the exchange bias energy, E_{dip} the dipolar field energy, E_{RKKY} the energy arising from the Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction and E_{ex} the exchange energy.

2.1.3.1 Zeeman Energy

The Zeeman energy describes the interaction between the magnetization of a material and an external magnetic field, $\mathbf{H}$. This contribution is minimized when the magnetization aligns with the direction of the applied field and can be written as a volume integral:

$$E_{\text{Zeeman}} = -\mu_0 \int_V \mathbf{m} \cdot \mathbf{H} dV \quad (2.14)$$

2.1.3.2 Exchange Energy

The exchange energy is a quantum-mechanical interaction that arises from the symmetry requirements of the wavefunction for indistinguishable particles. It originates from a combination of the Pauli exclusion principle and Coulomb repulsion, which together favor specific spin alignments to minimize the system's total energy. The exchange energy between two adjacent spins S_1 and S_2 can be expressed as:

$$E_{ex} = J_{ex} \mathbf{S}_1 \cdot \mathbf{S}_2 \quad (2.15)$$

where J_{ex} is the exchange constant. The sign of J_{ex} indicates if the spins are parallel or antiparallel aligned, and its magnitude quantifies the energy required to disrupt the magnetic ordering. That is associated with the Curie temperature for a ferromagnetic materials or to the Néel temperature in antiferromagnetics, through the thermal energy scale $k_B T$, where k_B is the Boltzmann constant and T is the temperature.

2.1.3.3 Demagnetizing Energy

When a magnetic material is magnetized, it generates a demagnetizing field ($\mathbf{H}_{dem}$) that extends into the surrounding space. Importantly, this self-generated field also interacts with the material itself. This interaction gives rise to a competition. On one hand, the exchange interaction (discussed in the section 2.1.3.2) is a short-range quantum-mechanical effect that favors parallel alignment of neighboring spins, and, thus, promotes a uniform magnetization across small regions of the material. On the other hand, the demagnetizing field is a long-range effect of magnetostatic origin, which acts to reduce the overall magnetization, depending on the geometry and spatial distribution of magnetic moments within the material.

The interplay between these effects leads to the formation of magnetic domains—regions in which the magnetization is nearly uniform. These domains are separated by domain walls, where the direction of the magnetization gradually varies in space in order to minimize the total magnetic energy.

The $\mathbf{H}_{dem}$ is the internal field responsible for reducing the net magnetization, and is given by:

$$\mathbf{H}_{dem} = -\mathbf{N} \cdot \mathbf{M}, \quad (2.16)$$

where $\mathbf{N}$ is the demagnetizing tensor, which encapsulates the effect of the sample's geometry. For uniformly magnetized ellipsoids, $\mathbf{N}$ is diagonal and satisfies the condition:

$$N_{xx} + N_{yy} + N_{zz} = 1.$$

The energy associated with the demagnetizing field is computed as a volume integral over the magnetic body:

$$E_{\text{dem}} = -\frac{1}{2}\mu_0 \int_V \mathbf{m} \cdot \mathbf{H}_{\text{dem}} dV. \quad (2.17)$$

2.1.3.4 Dipolar Field Energy

The dipolar field energy shares the same magnetostatic origin as the demagnetizing field energy. However, while the demagnetizing field $\mathbf{H}_{\text{dem}}$ arises from the interaction of a magnetic layer with its own magnetization, the dipolar field $\mathbf{H}_{\text{dip}}$ refers to the magnetic field generated by one magnetic layer that acts on a different magnetic layer.

In this case, the interaction can be described using the same formalism as the Zeeman energy, as presented in sub-section 2.1.3.1.

2.1.3.5 Anisotropy Energy

The anisotropy energy originates from the interaction between the electron orbitals and the crystal field, which tends to stabilize specific electronic configurations. Due to the spin-orbit coupling, this orbital stabilization translates into a preferred direction for the magnetic moment within the crystal. As a result, the magnetization tends to align along certain crystallographic directions, known as easy axes. Correspondingly, the directions of highest energy are called hard axis.

In many magnetic materials, particularly those with hexagonal or tetragonal symmetry, the anisotropy energy landscape is dominated by a single preferred direction. This is referred to as uniaxial anisotropy, and the corresponding energy contribution can be expressed as a volume integral:

$$E_{\text{an}} = \int_V K_u \left[1 - (\mathbf{u} \cdot \mathbf{m})^2 \right] dV, \quad (2.18)$$

where K_u is the uniaxial anisotropy constant (in units of energy density), $\mathbf{u}$ is the unit vector defining the easy axis direction.

This expression can be approximated by a second-order term:

$$E_{an} \approx K_u \sin^2 \theta, \quad (2.19)$$

where θ is the angle between the magnetization vector $\mathbf{M}$ and the easy axis $\mathbf{u}$.

2.1.3.6 Ruderman-Kittel-Kasuya-Yoshida (RKKY) energy

The Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction describes an indirect exchange mechanism between two localized magnetic moments in a solid, mediated by conduction electrons [31]. When a conduction electron interacts with a localized magnetic moment, it becomes polarized, and this polarization is then transferred to other localized magnetic moments through conduction electrons. The RKKY interaction energy is given by:

$$E_{RKKY} = -J_{RKKY} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (2.20)$$

where J_{RKKY} represents the effective coupling strength between magnetic impurities located at sites i and j . Depending on the spatial separation between the localized moments, the RKKY interaction can be ferromagnetic, antiferromagnetic or zero, exhibiting an oscillatory dependence on the inter-spin distance, as illustrated in Fig. 2.8.

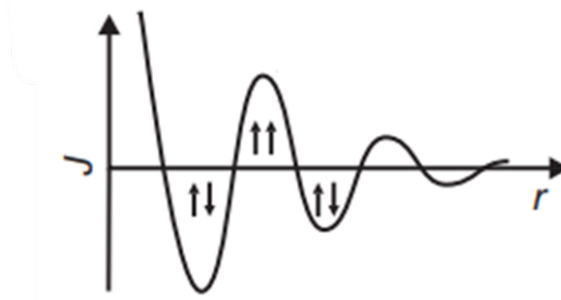


Figure 2.8: Oscillatory behavior of J_{RKKY} as a function of the distance between localized magnetic moments.

2.1.3.7 Exchange Bias Energy

The exchange bias energy originates from the direct coupling at the interface between antiferromagnetic (AFM) and ferromagnetic materials. This phenomenon was first observed by Meiklejohn and Bean in 1956 [32] in cobalt (Co) nanoparticles (with Curie temperature $T_C = 1390$ K) coated with cobalt oxide (CoO), which has a Néel temperature $T_N = 291$ K.

The exchange bias is established by cooling the FM/AFM bilayer in the presence of a magnetic field strong enough to saturate the FM layer. This process becomes effective when the Curie temperature of the FM layer exceeds the Néel temperature of the AFM layer ($T_C > T_N$). During cooling, exchange interactions at the FM/AFM interface induce a preferential alignment of the FM magnetic moments. These moments are subsequently pinned by the ordered magnetic structure of the AFM, resulting in a unidirectional anisotropy.

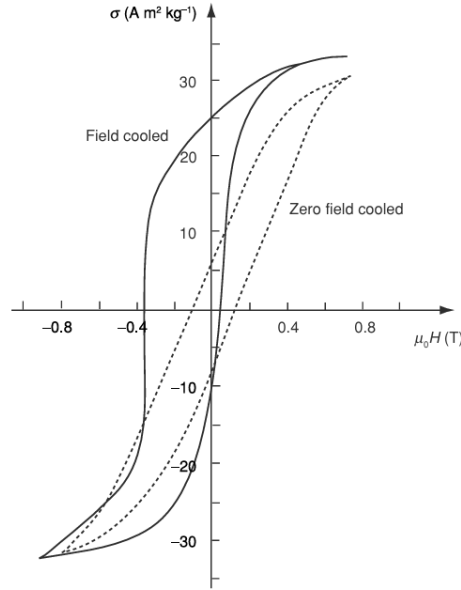


Figure 2.9: Shifted loop of partially oxidized cobalt particles, measured after cooling to 77 K in a 1 T field [32].

This phenomenon is typically observed as a horizontal shift in the ferromagnetic hysteresis loop, as illustrated in Figure 2.9, based on the experimental results of Meiklejohn and Bean. The total magnetic energy of a bilayer system, when the FM magnetization vector $\mathbf{M}_p$ forms an angle ϕ with an in-plane magnetic field H_x applied along the x -axis, is given by:

$$E_x = -\mu_0 M_p H_x \cos \phi - K_{ex} \cos \phi, \quad (2.21)$$

where K_{ex} is the exchange anisotropy energy per unit volume and the shift in the hysteresis loop is attributed to the unidirectional anisotropy term $K_{ex} \cos \phi$. Thus, the effective magnetic field acting on the pinned FM layer is:

$$H_{\text{eff}} = H + H_{\text{ex}}, \quad \text{where} \quad H_{\text{ex}} = \frac{K_{\text{ex}}}{\mu_0 M_p}. \quad (2.22)$$

The exchange anisotropy constant K_{ex} is related to the interfacial exchange energy density σ_{ex} and the thickness of the pinned ferromagnetic layer t_p by:

$$K_{\text{ex}} = \frac{\sigma_{\text{ex}}}{t_p}. \quad (2.23)$$

2.1.4 MTJ structure

Understanding the various contributions to the Gibbs free energy, and how they affect the individual layers and materials, is essential to comprehending the structure and function of an MTJ. As previously discussed, an MTJ consists of two ferromagnetic layers separated by a thin insulating barrier, typically an oxide. One of the ferromagnetic layers—referred to as the pinned layer—has its magnetization stabilized through exchange bias coupling with an adjacent antiferromagnetic layer, while the other, called the free layer (FL), remains free to switch its magnetization under an external stimulus.

However, to enhance the thermal stability of the MTJ, reduce stray magnetic fields, and increase the effective exchange field, the pinned FM layer is often not connected directly to the tunnel barrier. Instead, a synthetic antiferromagnetic (SAF) structure is used. A SAF is a trilayer system composed of two FM layers separated by an ultra-thin nonmagnetic spacer (typically ruthenium, Ru). The two FM layers in the SAF are coupled via the RKKY interaction, resulting in a more stable and rigid reference layer. Also, an MTJ usually presents a bottom buffer layer that promotes the desired crystallographic orientation for the subsequent layers and, at the top, a capping layer that protects the device from mechanical and chemical damage. A simplified representation of the MTJ configuration with and without a SAF layer is showed in Figure 2.10.

The annealing process of MTJs is a critical post-fabrication step that significantly enhances device performance. One of the primary objectives of annealing is to promote crystallization of both the MgO tunnel barrier and the adjacent ferromagnetic layers, typically composed of amorphous FeCoB. The MgO layer, which is usually deposited with a (001) crystalline texture, acts as a template for the crystallization of the neighboring FeCoB layers upon annealing [34, 35]. This crystallographic ordering is essential for achieving high TMR, as it enables coherent spin-polarized tunneling across the junction [36].

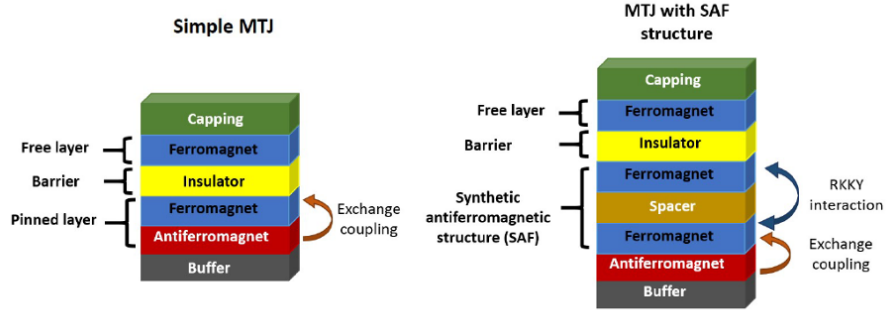


Figure 2.10: Stack of an MTJ structure with and with out a SAF [33].

To facilitate this process and optimize magnetic properties, a layer of amorphous Ta is often inserted adjacent to the FeCoB. This Ta layer serves two key roles: it acts as a boron sink, absorbing B atoms expelled from FeCoB during annealing [37], and it interrupts the propagation of the crystalline texture, which can be detrimental to other magnetic or structural properties in the stack.

In addition to crystallization, annealing plays a crucial role in establishing exchange bias between the AFM and FM layers. This is typically achieved by annealing the stack above the Néel temperature of the AFM material in the presence of a strong and uniform magnetic field, generally on the order of several hundred Oersted [38]. The applied field sets a preferred magnetic orientation for the pinned ferromagnetic layer, which remains stable unless subjected to an external magnetic field exceeding the exchange bias field. This fixed alignment is fundamental for SAF layers.

2.1.5 MTJ Thermal Stability

The thermal stability factor (Δ) is a critical parameter in evaluating the reliability and performance of MTJs, especially for memory applications. It is defined by the ratio between the energy barrier that is required to switch the FL magnetization between the equilibrium states (P and AP states) and the thermal energy available at a given temperature, and is given by:

$$\Delta = \frac{E_b}{k_B T} \quad (2.24)$$

where E_b is the energy barrier.

In the absence of external influences (such as magnetic fields, voltage, or current), the energy barrier is governed by the anisotropy energy of the free layer. For a uniformly magnetized system, the energy barrier can be expressed as (according to the Equation 2.18):

$$E_b = K_u V \quad (2.25)$$

where K_u is the uniaxial anisotropy constant and V is the volume of the free layer.

Given the relation $K_u = \frac{1}{2} H_K M_S$, where H_K is the anisotropy field and M_S the saturation magnetization, the thermal stability can also be written as:

$$\Delta = \frac{H_K M_S V}{2 k_B T}. \quad (2.26)$$

For in-plane anisotropy (IPA) MTJs, which will be the subject of this work, the anisotropy arises primarily from the shape of the device, particularly its aspect ratio (AR), defined as the length-to-width ratio. In these structures, the magnetization lies within the plane, and shape anisotropy plays a dominant role in determining the energy barrier. Geometries with high aspect ratios promote magnetization alignment along the long axis, thereby increasing the energy barrier for the short axis.

The anisotropy field H_K for an elliptical in-plane structure can be approximated as [39]:

$$H_K = 2 \frac{4\pi M_S t_f (AR - 1)}{w AR} \quad (2.27)$$

where t_f is the thickness and w the width of the free layer. Substituting into the expression for Δ , the thermal stability factor for in-plane MTJs becomes:

$$\Delta_{IP} = \frac{\pi^2 (M_S t_f)^2 w (AR - 1)}{k_B T} \quad (2.28)$$

This expression shows that Δ_{IP} scales with the width w , aspect ratio AR , and the square of the specific magnetic moment per unit area $M_S t_f$.

At finite temperature, thermal fluctuations can lead to spontaneous switching of the magnetization direction. The average time between two such thermally activated reversals is

known as the Néel relaxation time (τ), given by the Néel–Arrhenius relation:

$$\tau = \tau_0 \exp(\Delta) \quad (2.29)$$

Here, τ_0 is the attempt time, which although material dependent, is typically assumed to be on the order of 1 ns [40]. The exponential dependence of τ on Δ underscores the importance of thermal stability in ensuring long-term data retention in MTJ-based devices.

2.1.6 Magnetic Vortices

A magnetic vortex is characterized by an in-plane curling of magnetic moments around a central core, where the magnetization gradually points out of plane [41]. The vortex state is defined by two main parameters: chirality ($c = \pm 1$), which specifies the clockwise or counterclockwise sense of in-plane rotation, and polarity ($P = \pm 1$), which indicates the upward or downward orientation of the core magnetization, as represented in Figure 2.11. The formation of a vortex state depends critically on the ratio between the disc

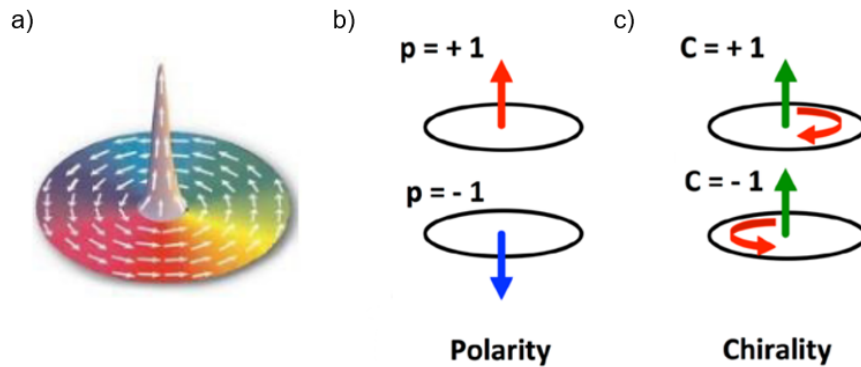


Figure 2.11: a) Schematic representation of a vortex magnetization configuration. Representation of the two main properties: b) polarity and (c) chirality [42].

radius (R) and thickness (L) and requires a ferromagnet with weak magnetocrystalline anisotropy. Figure 2.12a presents the phase diagram for nanoscale discs of a soft ferromagnet (e.g., NiFe), where L and R are normalized to the exchange length (L_E) of permalloy, a nickel–iron alloy widely used for its very low coercivity and high magnetic permeability. When $L \gg R$, the demagnetizing field favors an out-of-plane magnetization. Conversely, when $R \gg L$, the exchange interaction is more effective, maintaining in-plane magnetization. In an intermediate regime of R/L , the system can stabilize into a vortex configuration, balancing exchange and demagnetizing energies. The red area shows the region of metastable vortex dimensions.

Figure 2.12b illustrates the magnetization reversal of a vortex under an applied in-plane magnetic field. The stability of the vortex is characterized by two critical fields: the annihilation field (H_{an}), at which the vortex is destroyed, and the nucleation field (H_n), at which it re-forms. As the in-plane field increases, the vortex core is displaced toward the disc edge until it is annihilated, leading to irreversible switching into a saturated state. Upon reducing the field, the vortex reappears once the nucleation field is reached.

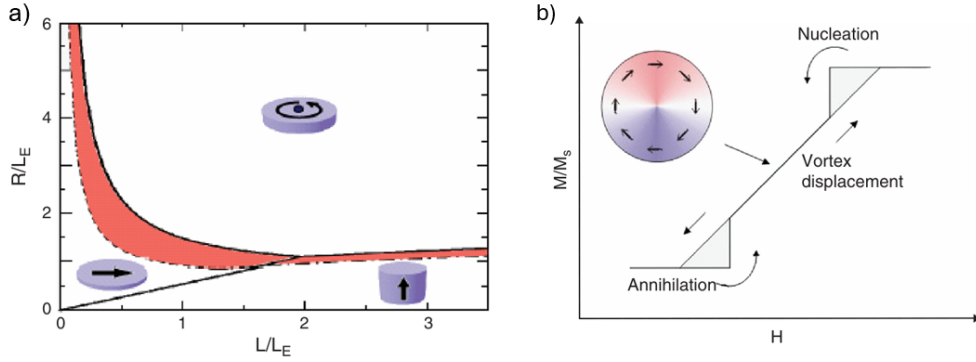


Figure 2.12: Magnetic vortex states. a) Magnetic phase diagram of cylindrical dots with radius R and thickness L . Three magnetization states are stable: vortex (upper part), in-plane single domain (small thickness), and perpendicular single domain (small radii). Solid lines show equilibrium conditions, dashed line corresponds to the stability border of the vortex state, and the dot-dashed line indicates the equilibrium vortex core radius ($L_E \approx 18$ nm for permalloy). b) Scheme of vortex magnetization reversal in a cylindrical dot under an in-plane magnetic field H . The remanent state corresponds to a centered vortex [43].

2.1.7 Spin-Transfer Torque

Previously, it was discussed how the application of a magnetic field can induce the switching of magnetic layers. However, this is not the only mechanism capable of controlling their magnetization. Another key phenomenon is spin-transfer torque (STT), which arises from the interplay between spin-polarized electric currents and the magnetization of ferromagnetic materials [44]. Although the electric current is conserved, the angular momentum of the electrons can be transferred to the magnetic lattice through spin-dependent scattering and other interactions. This transfer generates a torque on the magnetic moments, corresponding to the rate of change of angular momentum in the lattice. STT becomes particularly efficient at the nanoscale, where such interactions are enhanced.

Consider an MTJ composed of pinned and free magnetic layers, with their magnetizations initially non-collinear, as illustrated in Fig. 2.13, where F1 is the pinned layer with fixed magnetization, and F2 is the free layer. In Fig. 2.13a, the current flows from the pinned layer to the free layer. As the initially unpolarized electron current enters the pinned layer, it becomes spin-polarized along its magnetization direction. Upon reaching the free layer,

the transverse component of this spin polarization (relative to the magnetization of the free layer) induces a torque through a precessional mechanism. This spin-transfer torque tends to rotate the free layer's magnetization toward alignment with the spin polarization of the current, favoring a parallel configuration. In contrast, Fig. 2.13b depicts the case where the current flows from the free to the pinned layer. The electrons are spin-polarized according to the free layer's magnetization, but ineffectually exert torque on the pinned layer. However, the electrons spin-polarized in the opposite direction of the pinned layer will be reflected at the interface. These reflected electrons return to the free layer, where their transverse spin component is absorbed, exerting a torque that drives the free layer's magnetization toward an antiparallel alignment.

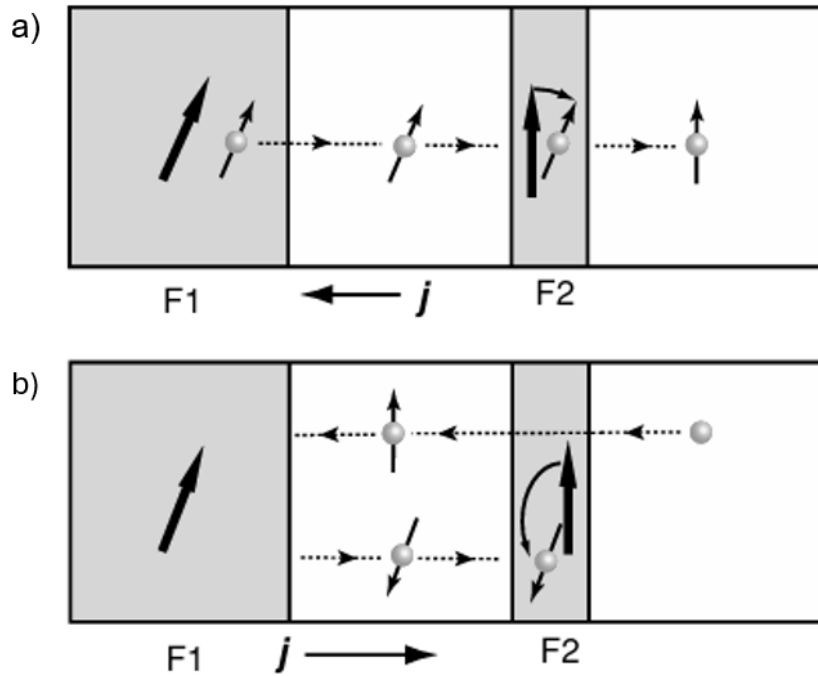


Figure 2.13: a) Spin-transfer torque promotes parallel alignment when current flows from the pinned to the free layer. b) Antiparallel alignment results when current flows in the reverse direction due to reflected spin-polarized electrons. Adapted from [29].

Thus, the magnetization dynamics defined by the LLG equation (Equation 2.11) must take into account this precession phenomenon, and an extra damping term must be added to account for the spin transfer torque.

Figure 2.14a illustrates the magnetization precession induced by STT. Although often referred to as antidamping torque, the STT vector can either reinforce or oppose the intrinsic damping, depending on the direction of the current and the magnetic configuration. When STT acts in the same direction as the damping torque, it accelerates the relaxation of the magnetization toward equilibrium (Fig. 2.14b). This occurs when the current flows

in a direction that stabilizes the current magnetic configuration. Conversely, when the STT opposes the damping, it can destabilize the equilibrium state. This leads to a precessional motion where the magnetization spirals away from its initial orientation. If the STT is strong enough, the magnetization may eventually reverse, leading to full switching (Fig. 2.14d).

An important intermediate case occurs when the damping and STT torques are equal in magnitude but opposite in direction. Under this condition, the magnetization enters a regime of steady-state precession (Fig. 2.14c), which is the operating principle behind spin-torque nano-oscillators (STNOs). These devices are particularly promising in the field of radio-frequency (RF) applications because of their ability to generate tunable microwave signals in the GHz range. Beyond oscillators, the STT effect can also have other applications such as the spin-transfer torque magnetic random-access memory (STT-MRAM), where magnetization switching is driven purely by electric current. STT is thus a scalable solution for both memory and advanced applications in RF and neuromorphic computing.

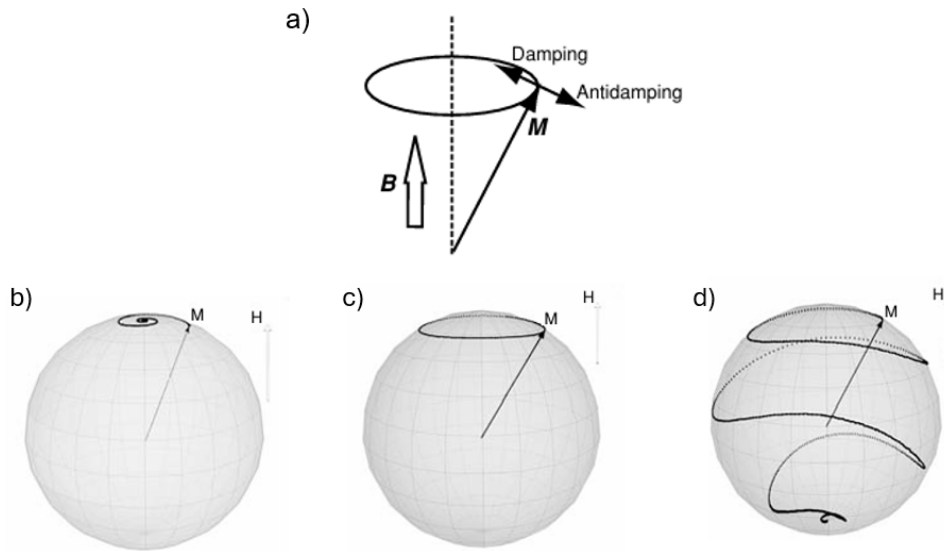


Figure 2.14: Magnetization dynamics under spin-transfer torque (STT). a) Precession under STT; b) relaxation when STT reinforces damping; c) steady-state precession when STT and damping balance; d) switching when STT overcomes damping. Adapted from [29].

2.1.8 STNO with STT-induced Dynamics in Magnetic Vortices

As mentioned, STNO rely on STT to generate a RF electrical signal. This effect is typically realized in MTJs with various magnetic configurations. In this work, the focus was on STNO with a magnetic vortex.

The vortex dynamics can be described by the Thiele equation [45]. At equilibrium, the vortex core resides at the center of the free layer. However, external magnetic fields or the STT effect can displace it from equilibrium, giving rise to gyrotropic motion:

$$\mathbf{G} \times \frac{d\mathbf{X}_C}{dt} - \frac{\delta \mathbf{W}}{\delta \mathbf{X}_C} - \mathbf{D} \times \frac{d\mathbf{X}_C}{dt} + \mathbf{F}_{\text{STT}} = 0, \quad (2.30)$$

where $\mathbf{X}_C$ denotes the coordinates of the vortex core, $\mathbf{G}$ is the gyrovector defined as $\mathbf{G} = -\frac{2pt_f M_s}{\gamma} \hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ is the unit perpendicular to the surface, p is the polarity, M_s the saturation magnetization γ the gyromagnetic ratio and t_f the thickness of the FL. The term $\mathbf{W}$ is the potential energy of the shifted vortex, responsible for the restoring force, $\mathbf{D}$ accounts for the damping factor, and $\mathbf{F}_{\text{STT}}$ is the vector force exerted from STT. These

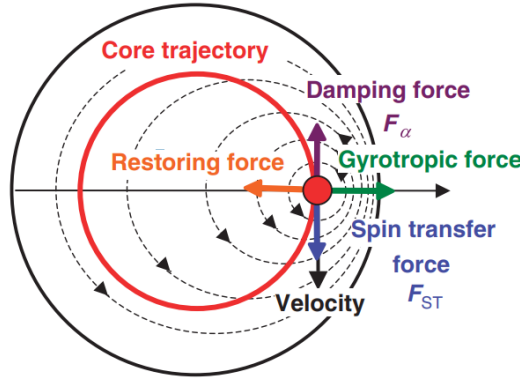


Figure 2.15: Schematic of the different forces acting on the vortex core [46].

contributions are represented in Figure 2.15. The terms related with the gyrotropic and the restoring forces are radial, defining the frequency motion, while the terms relative with damping and the STT forces are tangent to the trajectory, defining the amplitude of the orbit [46].

In general, spin-torque vortex oscillations (STVOs) exhibit oscillation frequencies typically below 1 GHz. Their main advantage lies in achieving higher output power and improved spectral purity compared to uniform-mode STNOs. With optimized device design, STVOs can deliver oscillation amplitudes in the microwatt range, while maintaining linewidths below 1 MHz [47].

2.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) were developed to address complex cognitive tasks, such as data prediction, language processing, and content generation, that traditional

logic-based algorithms struggle to solve [1]. These limitations arise because such algorithms are inherently sequential and rule-based, making them unsuitable for problems that require pattern recognition or contextual understanding. On the other hand, ANNs are inspired by the structure and functioning of the human brain. Biological neural networks consist of highly complex systems, composed of various types of neurons and synapses, interconnected in intricate ways. Their behavior is governed by dynamic processes involving both electrical impulses and chemical interactions. Moreover, other types of cells in the brain also contribute to cognitive functions, adding further layers of complexity.

As simplified computational models inspired by biological neural systems, ANNs consist of interconnected processing units, called artificial neurons, which receive inputs from other neurons through weighted connections, known as artificial synapses. Figure 2.16 illustrates a comparison between biological and artificial neural networks.

Two main paradigms of ANNs have emerged. The most widely used are non-spiking ANNs, where information is represented by continuous activation values (e.g., sigmoid, ReLU, or tanh functions). In contrast, Spiking Neural Networks (SNNs) encode and transmit information using discrete electrical events, or spikes, thereby more closely resembling the temporal dynamics of biological neurons [48]. Non-spiking ANNs currently dominate machine learning and AI applications, and they will be the focus of this study. These networks are typically designed to perform specific tasks. For example, in a classification problem, the network receives a set of input features and generates an output corresponding to a label or category. The entire process is defined by mathematical formulations, enabling training through optimization algorithms such as gradient descent and backpropagation [49], as this chapter will explore. Although SNNs are being investigated for their potential energy efficiency and biological plausibility, they remain outside the scope of this work.

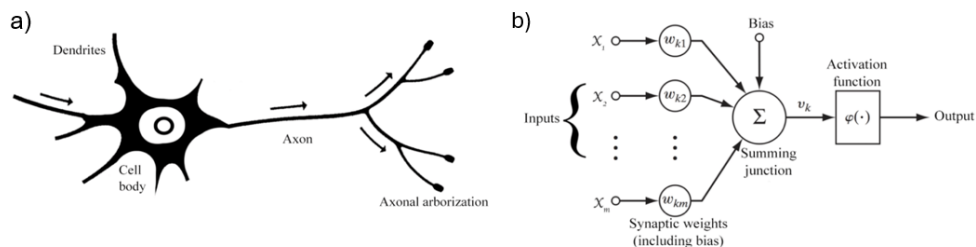


Figure 2.16: Comparison between a) biological and b) artificial neural networks [50].

2.2.1 Network Architecture

An artificial neural network (ANN) is composed of multiple layers of interconnected computational units, commonly referred to as artificial neurons or nodes. These neurons are organized in successive layers: an input layer, one or more hidden layers, and an output layer. A neural network that contains two or more hidden layers is typically referred to as a deep neural network (DNN), as illustrated in Fig. 2.17.

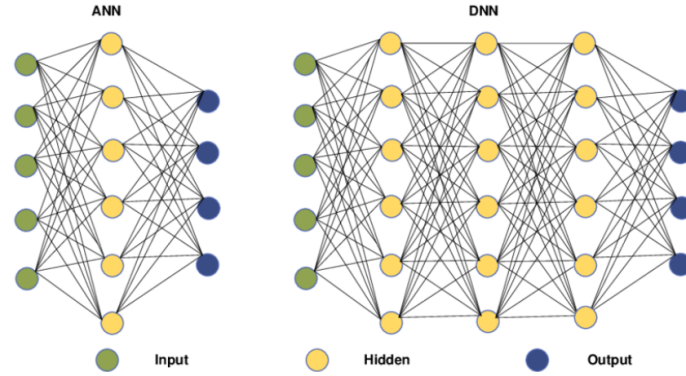


Figure 2.17: Comparison between an artificial neural network (ANN) and a deep neural network (DNN).

Each neuron receives signals from other neurons through weighted connections, with each connection assigned a numerical weight that determines the strength of the transmitted signal. These weights, along with a bias term, are the trainable parameters of the network and are adjusted during the training process. Their role is to emphasize relevant features and suppress irrelevant ones, allowing the network to learn meaningful patterns.

A fundamental operation performed by each neuron is the multiply-and-accumulate (MAC) operation, where the neuron computes a weighted sum of its inputs and adds a bias term. This operation is defined as:

$$z_j = \sum_{i=1}^N w_{ij}x_i + b_j, \quad (2.31)$$

where z_j is the pre-activation value (logit) of neuron j , N is the number of input connections, x_i are the input signals, w_{ij} are the synaptic weights, and b_j is the bias associated with neuron j .

After computing the weighted sum, the neuron applies a nonlinear activation function $a(\cdot)$ to produce its final output:

$$y_j = a(z_j), \quad (2.32)$$

The nonlinearity introduced by $a(\cdot)$ is essential, it enables neural networks to approximate complex, non-linear functions. In ANNs conventional analytic activation functions such as sigmoid [51] and the hyperbolic tangent [52] are commonly used. As the complexity of the activation function increases, the network's capacity to learn intricate patterns and achieve higher accuracy typically improves. In a fully connected, feedforward network, this MAC and activation sequence is applied to all neurons layer by layer, with the outputs of one layer serving as the inputs to the next, ultimately generating the final prediction at the output layer.

2.2.2 Training Neural Networks

For a neural network to perform effectively, it must undergo a training process in which its parameters—weights and biases—are optimized based on a given task. This optimization is guided by a loss function, which quantitatively measures the discrepancy between the network's output and the target output. This approach, known as supervised learning, requires a labeled dataset consisting of input samples paired with their corresponding expected outputs. The objective of training is to minimize the loss by updating the model parameters using, for example, a gradient-based optimization. Functions like the sigmoid or hyperbolic tangent are frequently used because their smooth, continuous derivatives that provide high resolution for gradient-based learning. A common update rule for weights in this context is:

$$w \leftarrow w - \eta \frac{\partial L}{\partial w}, \quad (2.33)$$

and similarly, for biases:

$$b \leftarrow b - \eta \frac{\partial L}{\partial b}, \quad (2.34)$$

where η is the learning rate, and $\frac{\partial L}{\partial w}$, $\frac{\partial L}{\partial b}$ are the gradients of the loss L with respect to the weight and bias, respectively.

The choice of loss function depends on the task. For regression problems, the Mean Squared Error (MSE) is commonly used. For multi-class classification, the Cross-Entropy

Loss is more appropriate, defined as:

$$L = - \sum_{i=1}^K \hat{y}_i \log(p_i), \quad (2.35)$$

where K is the number of classes, $\hat{y}_i$ is the true class label (typically 1 for the correct class and 0 for others), and p_i is the predicted probability for class i .

The output of the final layer is usually a set of logits z_i , which are unnormalized scores. To transform these into probabilities, the SoftMax function is applied:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \quad (2.36)$$

where z_i is the logit corresponding to class i . The SoftMax function converts logits into a valid probability distribution over the K classes. Combining SoftMax with the cross-entropy loss yields a convenient expression for the derivative of the loss with respect to each logit:

$$\frac{\partial L}{\partial z_i} = p_i - \hat{y}_i, \quad (2.37)$$

which serves as the error signal at the output layer.

2.2.3 Backpropagation Algorithm

Training a neural network involves computing gradients of the loss function with respect to all trainable parameters. This can be achieved efficiently using the backpropagation algorithm, which applies the chain rule of calculus to propagate gradients backward from the output layer through each layer of the network. This method is presented because it is a very common approach and will be adopted in this work.

For a weight $w_{ij}^{(l)}$ connecting neuron i in layer $l - 1$ to neuron j in layer l , the gradient of the loss is:

$$\frac{\partial L}{\partial w_{ij}^{(l)}} = \frac{\partial L}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial w_{ij}^{(l)}}. \quad (2.38)$$

Applying the full chain rule:

$$\frac{\partial L}{\partial w_{ij}^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot \frac{\partial a_j^{(l)}}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial w_{ij}^{(l)}}, \quad (2.39)$$

where $a_j^{(l)} = a(z_j^{(l)})$ is the output of the activation function.

From the definition of $z_j^{(l)} = \sum_i w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}$, it follows that:

$$\frac{\partial z_j^{(l)}}{\partial w_{ij}^{(l)}} = a_i^{(l-1)}, \quad (2.40)$$

which leads to:

$$\frac{\partial L}{\partial w_{ij}^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot a'(z_j^{(l)}) \cdot a_i^{(l-1)}. \quad (2.41)$$

The error term can be defined for the neuron j in layer l as:

$$\delta_j^{(l)} = \frac{\partial L}{\partial z_j^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot a'(z_j^{(l)}), \quad (2.42)$$

where $a'(\cdot)$ is the derivative of the activation function. The weight update rule then simplifies to:

$$\frac{\partial L}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} \cdot a_i^{(l-1)}. \quad (2.43)$$

To compute $\frac{\partial L}{\partial a_j^{(l)}}$, the recursive backward rule is used:

$$\frac{\partial L}{\partial a_j^{(l)}} = \sum_k \frac{\partial L}{\partial z_k^{(l+1)}} \cdot \frac{\partial z_k^{(l+1)}}{\partial a_j^{(l)}} = \sum_k \delta_k^{(l+1)} \cdot w_{jk}^{(l+1)}, \quad (2.44)$$

Thus, the recursive definition of the error term becomes:

$$\delta_j^{(l)} = \left(\sum_k \delta_k^{(l+1)} \cdot w_{jk}^{(l+1)} \right) \cdot a'(z_j^{(l)}). \quad (2.45)$$

where the sum is over all neurons k in the next layer $l + 1$ that receive input from the neuron j .

Analogously, the gradient of the loss function with respect to the bias, $b_j^{(l)}$, is given by:

$$\frac{\partial L}{\partial b_j^{(l)}} = \frac{\partial L}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial b_j^{(l)}}, \quad (2.46)$$

that simplifies to

$$\frac{\partial L}{\partial b_j^{(l)}} = \frac{\partial L}{\partial z_j^{(l)}} = \delta_j^{(l)}. \quad (2.47)$$

In this way, it was demonstrated mathematically how to obtain the gradients of the trainable parameters using the backpropagation algorithm. Nevertheless, alternative training approaches, such as equilibrium propagation or direct feedback alignment, also rely on gradient calculation. By analyzing this process, it was possible to understand how classical ANNs operate and how their training is performed, which is very useful to establish the foundations that will be applied throughout this study.

2.3. Spintronics for Neuromorphic Computing

The previous section provided a simplified overview of ANNs. However, the growing energy demands of artificial intelligence (AI) systems highlight the limitations of conventional CMOS-based, von Neumann architectures, particularly as Moore's Law slows down. Implementing ANNs in hardware using CMOS technology requires a vast number of transistors, resulting in significant area and power overheads. Moreover, the inherently two-dimensional nature of CMOS circuits constrains efficient scaling and dense connectivity [53].

In recent years, neuromorphic computing has gained significant attention as a computing paradigm that mimics the organizational and functional principles of the human brain. Within this field, spintronic technologies have attracted particular interest due to their ability to combine energy efficiency, non-volatility, and nanoscale integration. MTJs are especially promising, as they naturally exhibit nonlinear and non-volatile behavior and can be easily integrated with conventional CMOS circuitry, making them suitable candidates for emulating both neuronal and synaptic functions in hardware [7, 8]. The following sections discuss their use in this dual role, highlighting ongoing research that is relevant to the context of this work. This thesis is therefore structured around two distinct investigations, examining the use of MTJs as artificial neurons and as synapses.

2.3.1 Artificial Neurons

MTJs enable the implementation of different types of artificial neurons depending on the network architecture. In ANNs, neurons require differentiable and nonlinear activation functions, since higher nonlinearity typically improves learning accuracy. MTJs are particularly advantageous in this context because of their intrinsic nonlinear magnetic dynamics, such as domain wall motion, which can be exploited as hardware-based activation functions in neural networks [54]. Using device-level nonlinearities is beneficial because they generate rich activation functions directly in hardware, thereby reducing the computational overhead associated with implementing complex nonlinearities in software. Another promising approach to exploit MTJ nonlinearity is through STNOs. STNOs exhibit long lifetimes and low energy consumption, while converting direct current into microwave power. With the growing use of RF signals in information and wireless communication, STNOs represent an appealing alternative [55, 56]. Figure 2.18a shows the output power as a function of the bias current, reaching the nanowatt range, while the oscillation frequency (Fig. 2.18b) extends into the gigahertz range, demonstrating their potential for computation using microwave signals. To evaluate their performance, the experimentally

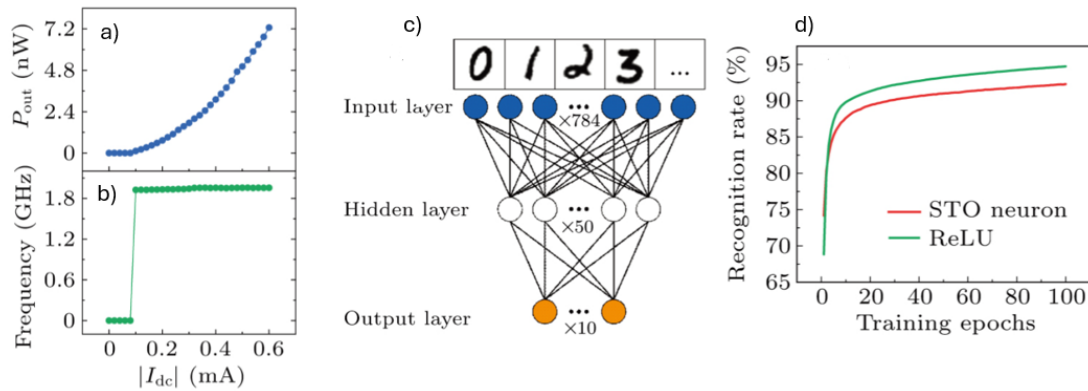


Figure 2.18: a) Integrated output power and b) resonance frequencies as a function of the current bias. c) Neural Network architecture with one hidden layer for the MNIST dataset. d) Recognition accuracy of the inference dataset in function of the training epochs obtained from the NN constructed using the STO-activation function and a conventional ReLU function. Adapted from [57].

obtained STNO activation function (from Fig. 2.18a) was applied to a recognition task using the MNIST handwritten digit dataset. An ANN represented in Figure 2.18c with two hidden layers (100 and 10 neurons) was implemented, where the STNO-based function was used as the activation function, and the results were compared against a conventional ANN using ReLU. As shown in Figure 2.18d, the STNO-based ANN achieved an accuracy of 92.28%, while the ReLU-based ANN reached 94.72%. Although the ReLU

performed slightly better, the results demonstrate the potential of device-level nonlinear functions for neuromorphic computation. However, these approaches still face limitations. During training, gradient calculations are performed in software, which can become increasingly computationally demanding as the activation function grows in complexity. In this thesis, a new method will be explored that calculates the gradient of a proposed I–V activation function directly from experimental measurements. This hardware-oriented approach aims to simplify training and enable the use of more complex nonlinear functions in practice.

2.3.2 Artificial Synapses

MTJ-based devices can also be employed as artificial synapses in neuromorphic systems. Their non-volatility stems from the controlled orientation of the ferromagnetic layers adjacent to the tunnel barrier, which enables them to retain information even in the absence of power, exploring the binary switching behavior between the P and AP magnetic states. This phenomenon can be driven by STT or spin-orbit-torque (SOT) [58–60].

In particular, in the Ref. [61], the authors demonstrate the use of STT-MRAMs to implement quantized synaptic weights. They connect seven identical devices in series, in order to obtain $N + 1$ resistance states, a configuration known as the thermometer scale. In this scheme, since all devices share the same dimensions, the P and AP resistances are the same for each MTJ, and the number of accessible states depends only on how many devices are in the P or AP state. Consequently, one state corresponds to all MTJs in the P state, with a total resistance of $N \times R_P$; another corresponds to all MTJs in the AP state, with a resistance of $N \times R_{AP}$. For intermediate cases, such as when k devices are in the P state and $(N - k)$ are in the AP state, the total resistance becomes $k \times R_P + (N - k) \times R_{AP}$. Thus, the resulting resistance depends solely on the number of devices in a specific state. Figure 2.19a presents the measured $R(V)$ curve for the seven MTJs, clearly showing the eight distinct resistance states. This approach was further evaluated on the MNIST dataset (Figure 2.19b) using a neural network with two hidden layers (Figure 2.19c). They then applied this system to neuromorphic computing by mapping high-precision floating-point weights from a trained software model onto the discrete conductance levels of multi-state MTJ cells. Since inputs to the neurons are voltages, weights were implemented using conductance values and, to allow both positive and negative weights, each effective synaptic weight was encoded as the difference between two paired multi-cell conductances.

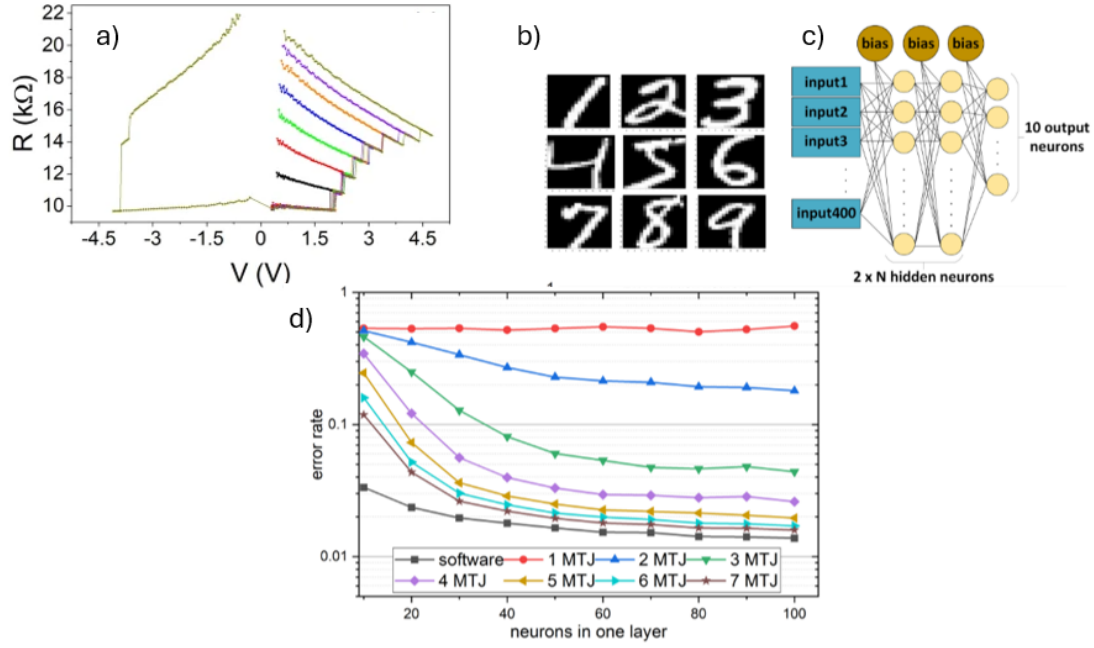


Figure 2.19: a) Measured $R(V)$ curve of seven identical MTJs connected in series. b) MNIST handwritten digit dataset used for testing. c) Neural network architecture with two hidden layers. d) Classification error rate on the test dataset as a function of the number of neurons per hidden layer, for different numbers of MTJ pairs used per multi-state synaptic cell [61].

The network performance is shown in Figure 2.19d, which reports the error rate as a function of the number of neurons per layer for different numbers of MTJs pairs used per multi-state synaptic cell. The results indicate that, when using at least three MTJs per cell, the performance of the MTJ-based network closely matches that of the original floating-point implementation.

Another recent example was presented in Ref. [62], where a weighted spin-torque nano-oscillator (WSTNO) system was demonstrated for neuromorphic computing. In this design, two MTJs were used as programmable MRAM-based weights, both operating in a bi-stable configuration and connected in series. Then, these weights were used to modulate the excitation current of another WSTNO acting as the output neuron. This architecture allowed the oscillator's output (frequency and power) to be modulated based on the synaptic weight as shown in Figure 2.20, enabling a tunable, non-linear neuron activation function. The use of non-volatile MTJ weights combined with a dynamic STNO output provides a compact, efficient mechanism for implementing both synaptic storage and neuron computation.

As a future perspective, and as will be explored in this thesis, this architecture can be extended by using MRAM elements with different physical dimensions in series, enabling the

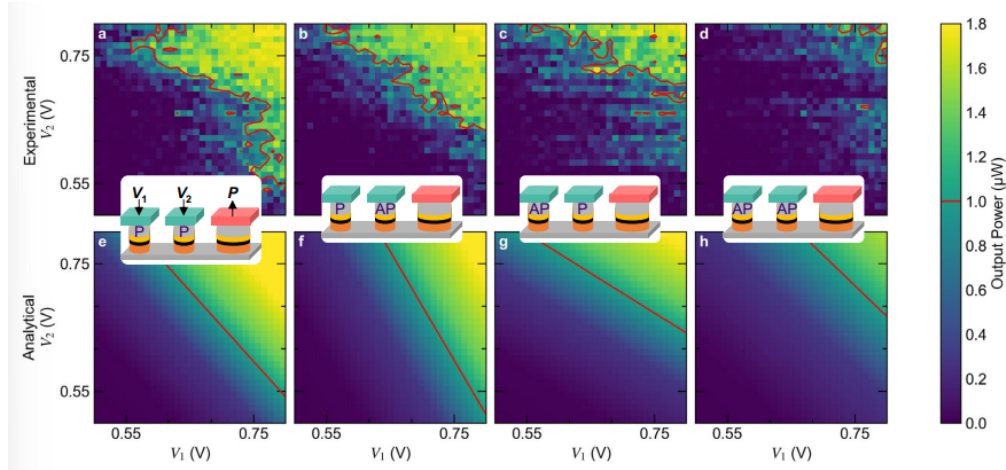


Figure 2.20: d Measured oscillation power (P) of the WSTNO as a function of the input voltages (V_1 and V_2) for each possible combination of the non-volatile states of the MRAMs: parallel and antiparallel. The contour of 1 μW output power is shown in red. e–h Analytical prediction of the WSTNO behavior as a function of the two input voltages for each state combination [62].

implementation of more resistance levels, as will be discussed in Chapter 5. This would allow for higher resolution weight encoding and more precise control over the synaptic input to STNO neurons, paving the way for more advanced neuromorphic hardware systems.

3. Experimental Methods

This chapter explains the fabrication techniques used to produce the devices and covers the characterization methods, where all measurements were carried out at INL. The samples were also fabricated at INL cleanroom by the Spintronics team.

3.1. Fabrication Techniques

The samples used in this thesis were fabricated through three main processes. First, the multilayer stack was deposited via physical vapor deposition (PVD) using magnetron sputtering. Multiple lithography steps were then employed to define both the MTJ pillars and electrical contacts. For microscale devices, direct laser writing was used for all patterning, while for nanoscale devices, DWL was used for contact definition and electron-beam (e-beam) lithography for the MTJ pillars. Pattern transfer was achieved via ion milling, a physical etching technique. Finally, after fabrication, the samples were annealed in the presence of a magnetic field to ensure that the pinned layer's magnetic orientation is fixed and to promote the crystallization of the MgO and CoFeB layers.

3.1.1 Magnetron Sputtering

Sputtering is a PVD technique for depositing thin films onto substrates. The process is carried out in a vacuum chamber filled with an inert gas, typically argon. Within the chamber, one or more solid targets that are composed with the desired deposition material are subjected to a high voltage, generating a plasma. In this plasma, argon ions are accelerated toward the negatively biased target, ejecting atoms from its surface through momentum transfer. These atoms then travel through the low-pressure environment and condense on the substrate, forming a uniform film. Depending on the nature of the target material, different sputtering modes are employed. DC sputtering is typically used for conductive targets. For insulating materials, RF sputtering is preferred, otherwise the insulator becomes charged.

There, a magnetron sputtering is used, where a magnetic field is present near the target. This field confines the electrons close to the target surface via the Lorentz force, increasing the ionization probability and plasma density, improving deposition rates and film uniformity (see Fig. 3.1a) [63].

At INL, deposition of MTJ multilayer stacks is performed using a Singulus Timaris Multi-Target Module (MTM) system (Fig. 3.1b). This platform includes a wafer cassette, transport module, soft/etch oxidation unit, and a UHV-compatible multi-target deposition chamber. The MTM houses up to 11 different targets, allowing the complete MTJ stack to be deposited sequentially without breaking vacuum. The substrate undergoes dynamic deposition, moving linearly back and forth across the targets. The soft/etch oxidation unit is used to clean the wafer surface by removing contaminants such as native oxides and adsorbed water before deposition.

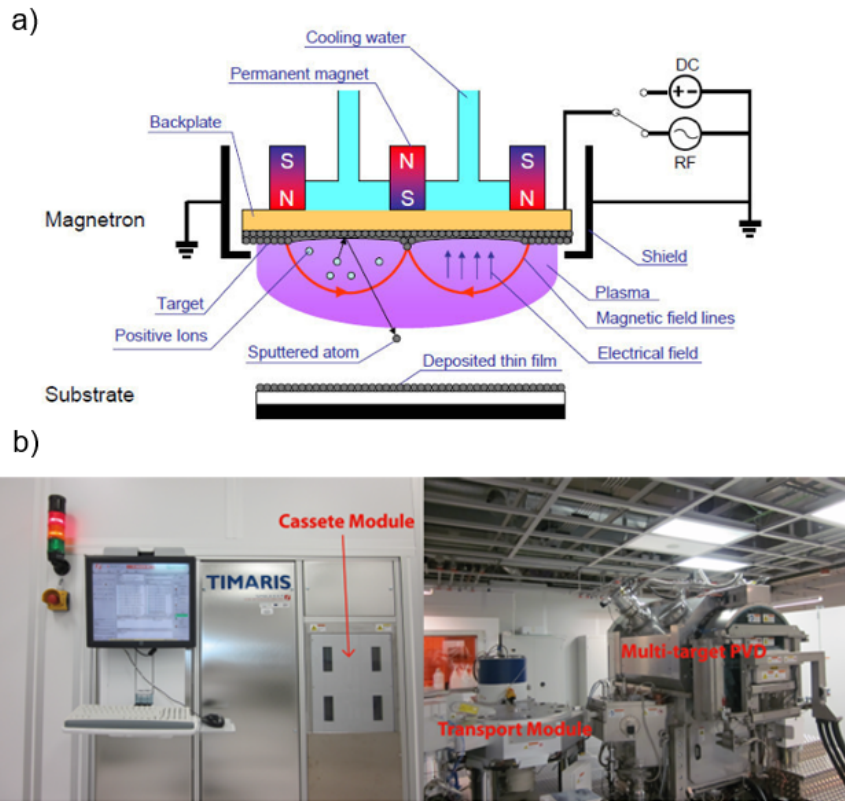


Figure 3.1: a) Magnetron sputtering scheme adapted from [64] and b) MTM system used at INL for MTJ deposition [65].

Unlike conventional systems, where deposition time determines thickness, the MTM controls deposition through the substrate velocity and the number of horizontal passes. Additional parameters, such as plasma power and working pressure, also allow fine-tuning of the film thickness and properties. For the samples used in this work, different deposition parameters were employed; however, the magnetron current was kept constant at -40 A. The insulating barrier was deposited using a power of 3 kW, with an argon flow of 600 sccm for the deposition of MgO. In some samples, nitrogen doping of the barrier was introduced by mixing 420 sccm of Ar with 200 sccm of N_2 . It should be noted that the total gas flow during nitrogen doping was 620 sccm, higher than the standard 600 sccm. This

increase arises from the lower molecular mass of nitrogen, which requires a higher total flow to maintain the same deposition pressure of 2.75×10^{-3} mbar.

3.1.2 Etching-Ion beam Milling

Ion milling is a dry etching technique based on the physical removal of material by bombardment with an accelerated ion beam, typically composed of argon ions (Ar^+). The ions strike the surface of the sample at a defined angle, removing the material without relying on chemical reactions (Fig. 3.2a). At INL, the machine used is a Nordiko 7500 system, which consists of a cassette module, a transfer module, and a processing chamber that aggregates both the ion beam source and the working chamber as shown in Figure 3.2b.

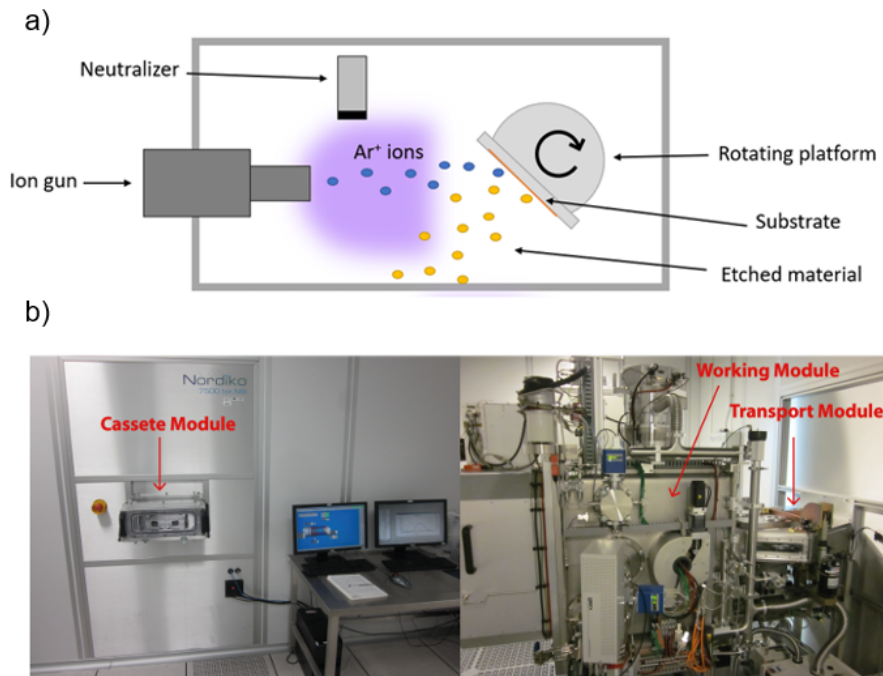


Figure 3.2: a) Schematic of the ion milling process and b) Nordiko ion milling system available at INL [65].

To produce the ion beam, inert argon gas is introduced into a discharge chamber, where an RF power source ionizes the gas, creating a plasma. From this plasma, positively charged argon ions (Ar^+) are extracted through a system of three electrostatic grids with different apertures, which control the ion energy. Optimizing the potentials applied between these grids is crucial to achieve the desired etching rate and uniformity. In this work, the operating parameters were: RF power of 300 W, positive grid voltage of 400 V, negative grid voltage of -2800 V, and an ion beam current of 240 mA. Once accelerated, the Ar^+ ions, supplied at a flow of 600 sccm, strike the sample surface, which is rotated at 30 rpm. The collisions transfer momentum to the surface atoms, ejecting them from

the material. The ion beam was directed at an incidence angle of 50° for the definition of the pillars, and at 6° during the cleaning step of the pillar sidewalls to remove redeposited material that could cause short circuits across the barrier. This process is essentially the same as sputtering, with the key difference that in ion milling the removed atoms do not redeposit elsewhere in the system, but instead leave the chamber. The working chamber in this machine also includes two electron neutralizers, which emit electrons to neutralize the positive ion beam. This is essential to prevent charge accumulation on the sample surface, which can cause variation in etch rate due to ion deceleration by surface charges; risk of electrostatic discharge in dielectric or resist layers, and distortion of the beam profile due to ion repulsion, leading to reduced etching uniformity.

3.1.3 Lithography

3.1.3.1 Direct Write Lithography

Different lithography techniques are required depending on the feature size and design specifications of the device. For the microscale samples used in this thesis, Direct Write Lithography (DWL) was employed for all patterning steps. DWL is a maskless lithography technique that offers high flexibility, rapid prototyping, and ease of use. In this process, a digital mask is created and written directly onto the photoresist (PR) using a focused laser beam. Before exposure, the substrate surface is prepared via vapor priming, and the photoresist is spin-coated onto it. Depending on whether a positive or negative resist is used, the laser selectively exposes the regions to be removed or retained, respectively, according to the designed pattern. At INL, the system used is the DWL 2000 tool by Heidelberg Instruments (Fig.3.3a), which employs a 405 nm diode laser and achieves a minimum feature size of approximately $1\ \mu\text{m}$. After exposure, the pattern is developed by removing either the exposed or unexposed regions of the PR, depending on the resist tone, typically using plasma etching, thus revealing the final pattern.

3.1.3.2 E-beam Lithography

As mentioned, the fabrication of MTJs for nanometer-scale resolution, such as those required for STNO applications, cannot be achieved using conventional optical lithography. This is because the feature size, d , of a lithographic system is given by:

$$d = k \frac{\lambda}{\text{NA}} \quad (3.1)$$

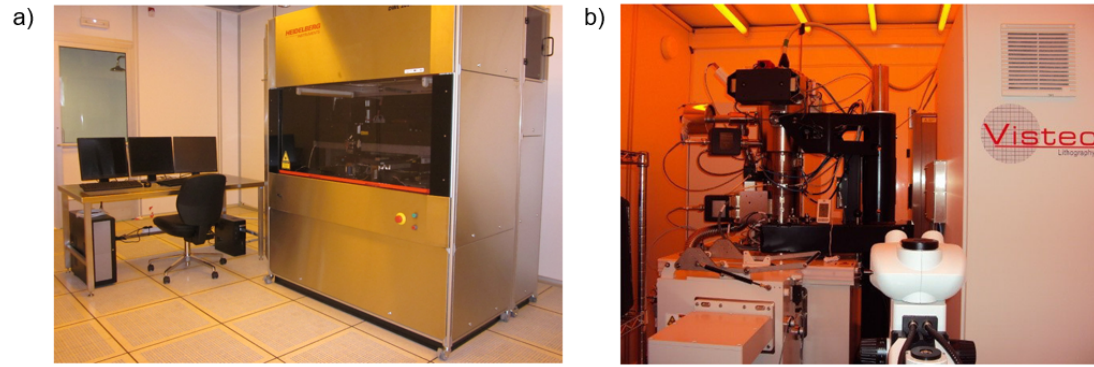


Figure 3.3: a) DWL lithography system and b) e-beam lithography system at INL cleanroom [66].

where k is a process-dependent coefficient, λ is the wavelength of the beam, and NA is the numerical aperture as seen from the wafer. From the Equation 3.1, it is evident that the resolution can be improved by either increasing the numerical aperture, i.e., using a more focused beam, or by decreasing the wavelength.

Since with UV beams is hard to achieve sub-100 nm features—due to the relatively long wavelength, e-beam lithography is employed instead.

E-beam lithography consists of scanning a highly focused electron beam across a resist-coated substrate, following a computer-controlled pattern. The electron beam locally alters the solubility of the resist (typically a polymer like PMMA), either by breaking molecular chains (positive resist) or by cross-linking them (negative resist). After exposure, the resist is developed to remove either the exposed or unexposed regions, depending on the resist type. Although e-beam lithography provides extremely high resolution, it is a relatively slow and not well-suited for large-scale production. Nevertheless, for research purposes and for the fabrication of nanoscale devices, it remains an indispensable tool. At INL a Vistec system is used, as shown in Figure 3.3b.

3.2. Electrical Measurement Techniques

In this section, the measurement techniques employed in this thesis are presented. The experimental work focused on three main approaches: current-voltage (I-V) characterization, transfer curve (R(H)) analysis, and RF measurements. These methods were applied to different configurations of MTJs, which will be further examined in the following chapters.

3.2.1 Current–Voltage and Transfer Curve Characterization

The setup used to characterize the I-V curves of the samples is illustrated in Figure 3.4.

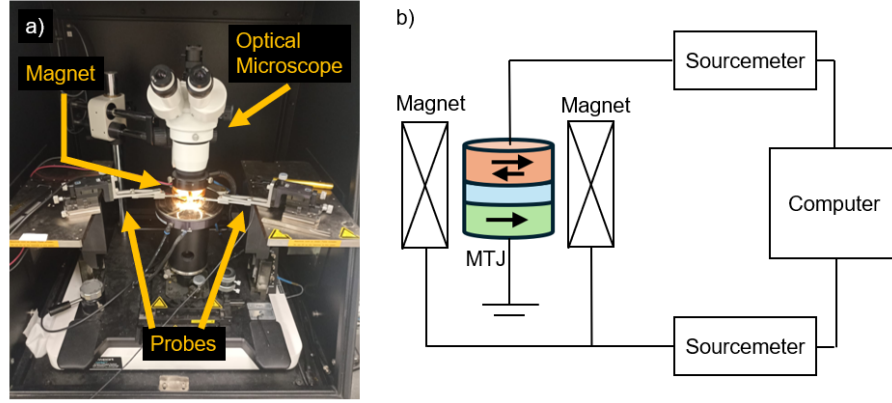


Figure 3.4: Experimental setup: a) Sample, probes, magnet, optical microscope. b) Setup scheme for I–V and R(H) curve measurements.

The sample is connected to a sourcemeter Keithley 2400 via two DC probes, which are attached to the top and bottom electrodes of the MTJ to ensure vertical current flow through the junction. The sourcemeter applies a current and simultaneously measures the resulting voltage across the device. A magnetic field, aligned with the MTJ orientation, is generated using a magnet and controlled by a second sourcemeter. All instruments are interfaced with LabVIEW via a GPIB connection, enabling precise control of experimental parameters and real-time data acquisition and visualization.

The same setup was employed for transfer curve characterization. In this case, however, a constant current was applied while the external magnetic field was swept. The voltage across the MTJ was measured as a function of the applied field, from which the resistance of the device was extracted.

3.2.2 RF measurement

To analyze the oscillation profile of an MTJ-STNO, the measurement setup shown in Fig. 3.5 was used. The MTJ is contacted using Cascade Microtech RF probes, which provide low contact resistance, a characteristic impedance of 50Ω , and support RF measurements up to 67 GHz. The probes connect the device to a bias tee, a three-port component that separates the DC and RF signals. One port is connected to the MTJ, carrying both the DC bias and the RF output. The second port, containing an inductor, passes the DC current (I_{bias}) while blocking the RF component. The third port, containing a capacitor that blocks DC but allows the RF signal to pass through, is connected to a spectrum analyzer that measures the voltage as a function of frequency $V(f)$, for a given bias current I_{bias} . The working principle of the second port is based on the increase of

the inductive reactance with frequency ($X_L = 2\pi fL$) so that at high frequencies, it behaves like an open circuit, preventing RF leakage into the DC line while still passing the DC bias. Conversely, the third port is based on the fact that the capacitive reactance decreases with frequency ($X_C = \frac{1}{2\pi fC}$) so that the capacitor behaves like a short circuit, efficiently transmitting the RF signal while blocking DC. The bias tee is designed for a characteristic impedance of $Z_0 = 50 \Omega$, matching the rest of the RF system. To ensure proper operation, the capacitor impedance at the measurement frequency must be much smaller than Z_0 to avoid insertion loss, while the inductor impedance must be much larger than Z_0 to isolate the RF path from the DC source effectively.

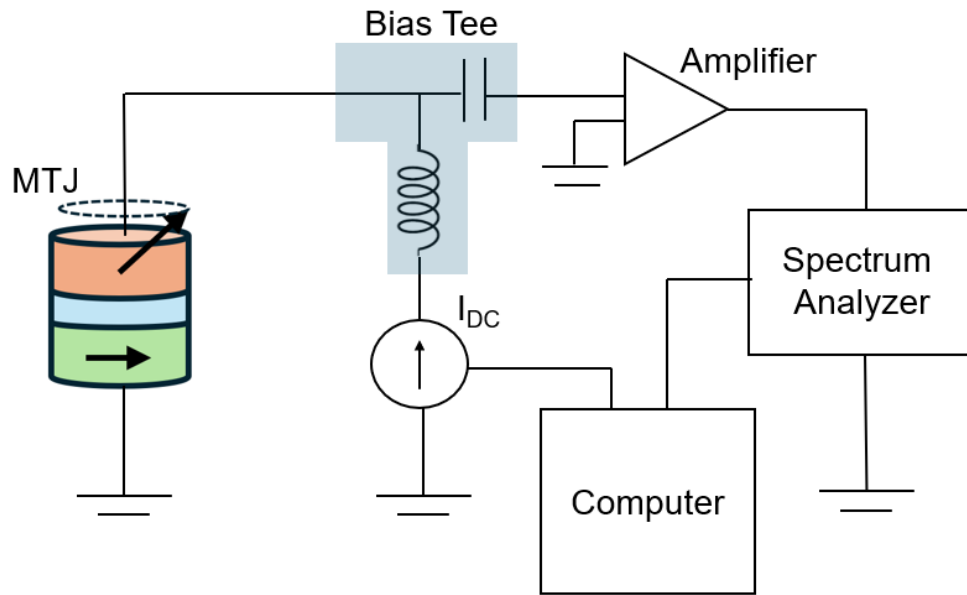


Figure 3.5: Schematic representation of the experimental setup for the RF measurement.

4. Spintronic Neurons and Finite Analog Difference Method

MTJs offer several compelling advantages for use as neurons. A key property is that these systems exhibit nonlinear responses, which is essential for training ANNs relying on nonlinear activation functions. However, training analog neuromorphic systems can be a challenging task. The behavior of analog devices is governed by complex, often non-ideal physical dynamics, including stochasticity, drift, and variability between devices. As discussed, performing network training requires knowledge of the activation function's gradient, which is essential for algorithms such as backpropagation. Therefore, this work aims to present a method to experimentally determine the gradient of the activation function using an analog finite-difference approach.

Recent efforts in digital computing have explored finite-difference methods to incorporate complex nonlinearities into neural networks without relying on analytical gradients [67, 68]. While effective in principle, these methods require multiple function evaluations per parameter update, effectively doubling computational costs and exacerbating the memory and latency challenges already present in digital systems [12, 69]. In order to mitigate these problems, this work will explore the nonlinearity of the I-V response of MTJs, and its gradient generation through an analog finite difference method (in Section 4.1 and Section 4.2). The study also examines different MTJ configurations to identify which are best suited for operation as spintronic neurons (Section 4.3). Finally, the performance of these MTJ-based neurons is evaluated within neural networks (Section 4.4), demonstrating strong generalization capabilities even when scaled to deeper architectures.

4.1. Spintronic Neurons

In this section, it will be demonstrated how the intrinsic analog I-V response from MTJs can serve as activation functions within gradient-trained networks. The I-V curves were chosen for their simplicity in measurement and their inherent nonlinearity, as illustrated in Figure 2.3. Similarly with the general case given by Equation 2.31, to model the behavior of MTJ-based neurons within a layered neural network, it is considered the input current to neuron j in layer l as:

$$I_j^{(l)} = \sum_i w_{ij}^{(l)} V_i^{(l-1)} + b_j^{(l)}, \quad (4.1)$$

where $w_{ij}^{(l)}$ is the synaptic weight connecting neuron i in layer $l - 1$ to neuron j in layer l , $V_i^{(l-1)}$ is the output from neuron i in the previous layer, and $b_j^{(l)}$ is a bias term. The output of neuron j is then computed as:

$$V_j^{(l)} = f(I_j^{(l)}), \quad (4.2)$$

where $f(\cdot)$ represents the activation function realized by the MTJ device. Unlike conventional analytic activation functions such as ReLU or sigmoid, $f(\cdot)$ is experimentally extracted from the voltage response of the MTJ. This allows us to incorporate the intrinsic device-level nonlinearities, asymmetries, and saturation effects directly into the network's computation (as presented in Figure 4.1a). This makes it possible to leverage the spintronic neurons for training a neural network using the backpropagation algorithm; however, access to the function's gradient is still required.

4.2. Analog Finite-Difference Gradient Method

To obtain the gradient of nonlinear functions without relying on digital emulation and while maintaining low computational overhead, an analog finite difference method was developed. This method consist on employing a pair of nominally identical devices (MTJ_1 and MTJ_2) characterized by closely matching I-V curves. In Figure 4.1a, the schematic illustrates the idea where MTJ_1 receives a reference input current I , while MTJ_2 is biased with a slightly increased current $I + \Delta I$. Then, the gradient is estimated by measuring the voltage difference between the two devices. Assuming V_1 and V_2 are the voltage outputs corresponding to the two current inputs, the finite difference approximation of the gradient is given by:

$$\left. \frac{dV}{di} \right|_{i=I+\frac{\Delta I}{2}} \approx \frac{V_2(I + \Delta I) - V_1(I)}{\Delta I} \quad (4.3)$$

Since the experimental gradient is obtained by applying a current increment (ΔI) to one of the MTJs in a device pair, the resulting estimate corresponds to an average slope over the interval $[I, I + \Delta I]$, rather than being centered at the reference current, I . To align the experimental gradient with the ideal numerical one, it is necessary to account for this offset. According with the Mean Value Theorem, if the function is sufficiently smooth, the gradient should be evaluated at the midpoint current $I + \frac{\Delta I}{2}$. Applying this

correction ensures that the experimentally measured gradient is properly centered and directly comparable to its numerical counterpart.

In Figure 4.1b, the I-V curves of two nominally identical devices (MTJ_1 and MTJ_2) are shown, biased over the range $I = [-800, 800] \mu A$, with MTJ_2 offset by a current of $\Delta I = 100 \mu A$. By subtracting the corresponding voltages, $\Delta V = V_2 - V_1$, and dividing by ΔI , the experimental gradient is obtained. Then, the experimental gradient was calculated according to Equation 4.3, where each point is evaluated at the current value of $I + \frac{\Delta I}{2}$, as shown in red in Figure 4.1c. This ensures that the gradient is aligned correctly along the current axis; otherwise, it would appear offset when compared to the numerical derivative of the I-V response of MTJ_1 , which is shown in green.

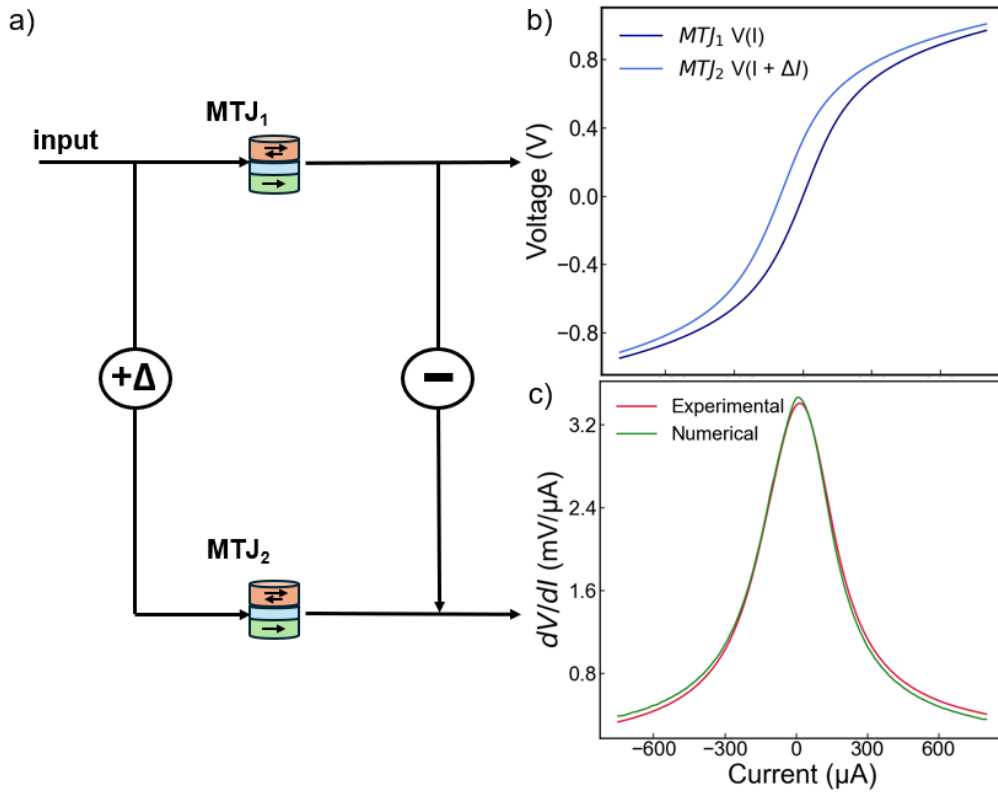


Figure 4.1: Schematic of the analog finite difference method used to estimate the gradient of the MTJ I-V characteristic. a) shows schematically the proposed architecture using a pair of nominally identical MTJs that are simultaneously biased with slightly different input currents, I and $I + \Delta I$, where $\Delta I = 100 \mu A$. b) Corresponding measured voltages (V) in function of current (I). c) Comparison between the experimental (red) and numerical (green) gradients of the I-V characteristic in function of current (I).

The close agreement between the experimental and numerical curves demonstrates the accuracy and reliability of the analog finite-difference gradient method, showing its potential for practical implementation in neural networks. This simple approach relies on using two devices with identical nonlinear responses, a principle that could also be extended to

other classes of nonlinear behaviors. However, doubling the number of devices leads to a larger area footprint and higher energy consumption. In this study, the devices tested are at the microscale, but modern spintronics systems can already be fabricated at very small scales with low power requirements, while still preserving strong nonlinear characteristics, thereby reducing the overall energy consumption. Another important advantage of this method is that, although the device count is doubled, the computational cost remains independent of the complexity of the nonlinear activation function. Unlike digital implementations, where the cost of evaluating an activation function increases with its mathematical complexity, here the cost scales only polynomially with the number of neurons in the network. Furthermore, this cost can be further reduced by leveraging the fact that highly nonlinear activation functions increase the expressive power of the network, thus reducing the total number of neurons required to achieve a given task.

4.3. I-V and Gradient Responses

This section presents an analysis of spintronic devices with varying material and thermal configurations, aiming to identify candidates capable of reliably functioning as artificial neurons. The primary objective of this analysis is to identify devices with consistent nonlinear I-V behavior across identical geometries, allowing reliable experimental extraction of the activation function and its corresponding gradient. To ensure this consistency, each device was evaluated over a range of input currents to determine the region in which the I-V response remains stable and reversible. Through this analysis, it was observed that the devices maintained stable and reproducible behavior within the voltage range of approximately ± 1 V. When the input current exceeds this range, the devices begin to exhibit soft breakdown [70], as illustrated in Figure 4.2, where a voltage value of approximately ± 1.5 V was reached. These effects are characterized by small hysteresis in the I-V curves, meaning the forward and reverse sweeps no longer overlap. This introduces unwanted variability in the activation function and hinders the reproducibility required for accurate gradient estimation.

Four different configurations were studied, distinguished by the incorporation of nitrogen in different parts of the stack. In the first stage, two configurations were evaluated with varying amounts of nitrogen doping in the insulating barrier. In the second stage, two additional configurations were analyzed in which nitrogen was introduced into the free layer. This approach was performed to identify nonlinear activation functions and to explore

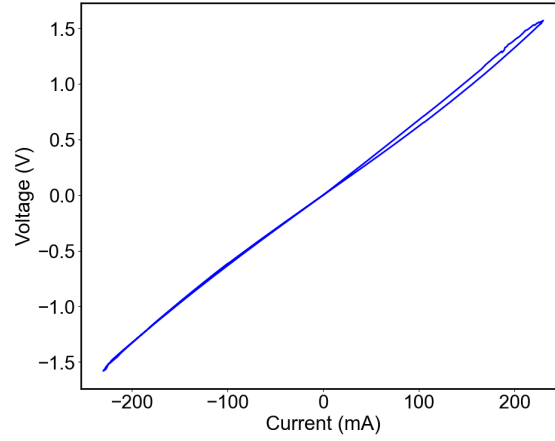


Figure 4.2: I-V curve for S_A with a radius of 20000 nm showing that above ± 1 V the MTJ start to exhibit memristive effects.

asymmetry and more complex behaviors that are less commonly observed in conventional numerical activation functions. To this end, a study was conducted on the incorporation of nitrogen and its impact. These four designs are summarized schematically in Figure 4.3. Each configuration included multiple devices with circular pillar shapes and varying diameters, as detailed in Table 4.1.

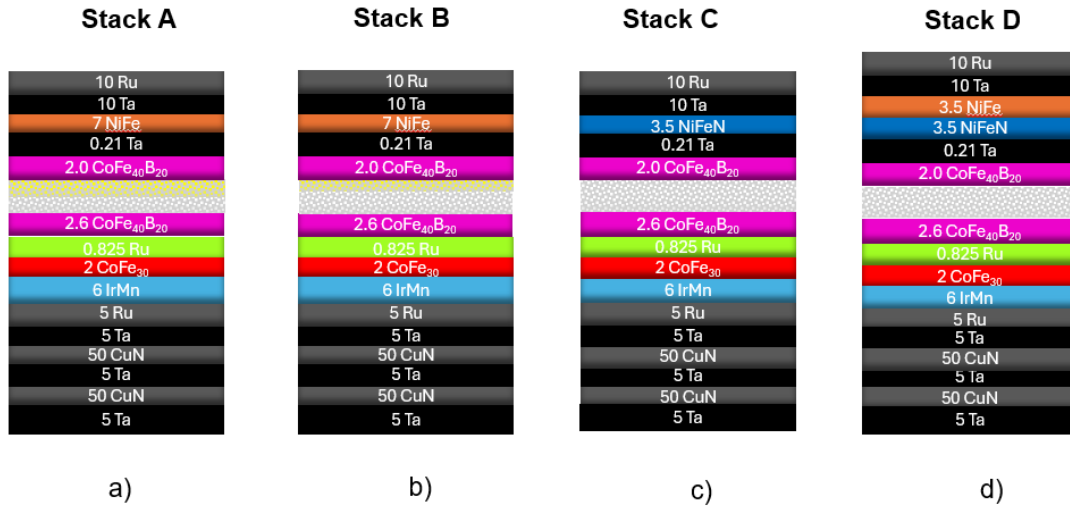


Figure 4.3: Schematic representation of the MTJ stacks used for I-V measurements. a) Stack A (S_A), b) Stack B (S_B), c) Stack C (S_C), and d) Stack D (S_D). The MgO barrier is shown in grey, while nitrogen doping in S_A and S_B is highlighted in yellow. Schematics are not in scale.

4.3.1 Nitrogen Incorporation in the Barrier

Two main configurations were evaluated, which differ in the quantity of nitrogen doping of the insulating barrier. Both structures share the same pinned reference layer (6 nm IrMn / 2 nm CoFe₃₀ / 0.83 nm Ru / 2 nm Co₄₀Fe₄₀B₂₀) and the same free layer composition (2 nm

Table 4.1: Values of r_{XX} and r_{YY} for different MTJ positions.

Position	r_{XX} [nm]	r_{YY} [nm]
1	2500	2500
2	4000	4000
3	5000	5000
4	7500	7500
5	10000	10000
6	12500	12500
7	15000	15000
8	17500	17500
9	20000	20000
10	25000	25000

$\text{Co}_{40}\text{Fe}_{40}\text{B}_{20}$ / 0.21 nmTa / 7 nm NiFe). For clarity, these configurations will be referred to as Stack A (S_A) and Stack B (S_B).

In S_A , the MgO barrier was deposited with four passes under a pure Ar atmosphere, followed by four passes under a mixed Ar/N₂ atmosphere. In contrast, S_B was deposited with six passes of MgO layers under pure Ar and only two passes under mixed Ar/N₂, resulting in a lower nitrogen proportion compared to S_A (see Chapter 3 for detailed deposition parameters). Figure 4.3a and Figure 4.3b show the schematic layer compositions of S_A and S_B , respectively.

Figure 4.4 shows the dependence of TMR on RA for S_A (green) and S_B (orange). Although the difference in TMR is not very pronounced, Stack S_B (with a lower nitrogen concentration in the MgO barrier) reaches slightly higher TMR values. This behavior can be explained by the fact that excessive nitrogen incorporation may degrade the crystalline quality of MgO, weakening the coherent tunneling of Δ_1 symmetry states and thereby reducing spin filtering efficiency. In contrast, S_A exhibits a significantly larger RA. This increase is attributed to the higher nitrogen concentration in its MgO barrier, which raises the effective barrier height, thus enhancing resistance.

To further assess device performance, the variability of the I–V characteristics was analyzed. The standard deviation of the measured voltage was calculated across devices of different diameters (Figure 4.5a). Figure 4.5b presents the average standard deviation across all geometries, providing a global comparison between configurations.

From these results, it is evident that Stack S_A exhibits significantly lower variability compared to S_B . This observation aligns with reports in the literature that nitrogen incorporation into oxide tunnel barriers enhances their reliability by reducing the density of defect

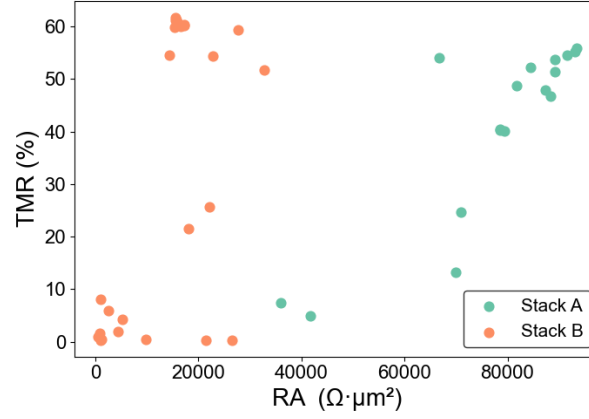


Figure 4.4: Dependence of TMR on RA for Stack S_A (green) and Stack S_B (orange).

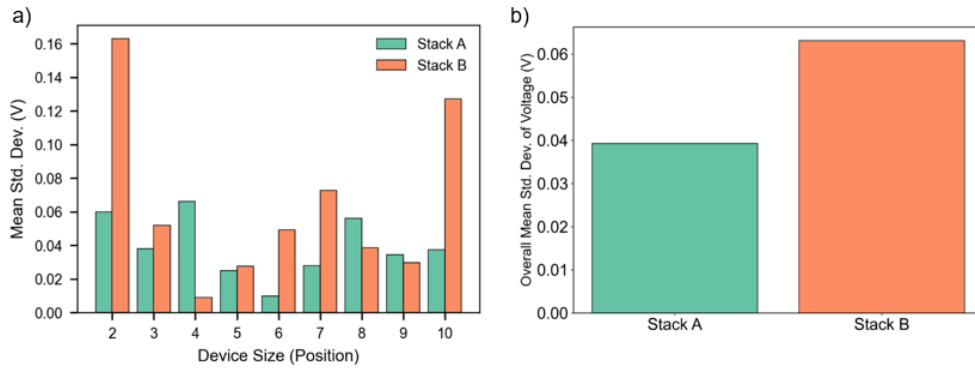


Figure 4.5: Device-to-device variability analysis of S_A and S_B . a) Standard deviation of voltage for each diameter according to the designation of Table 4.1. b) Average of standard deviations across all geometries.

trap sites [71]. The proposed mechanism is that nitrogen atoms form bonds with oxygen vacancies and trap sites within the barrier, thereby suppressing trap-assisted tunneling and improving barrier uniformity. As a result, nitrogen incorporation introduces a trade-off: while higher nitrogen concentrations lead to increased RA and reduced TMR due to disruption of coherent tunneling, they also improve device reliability by stabilizing the barrier.

Given these findings, and since reduced variability and consistent behavior are crucial for reliable gradient estimation, the configuration with higher nitrogen doping (S_A) was selected for further investigation. In this way, each device provides two activation functions, corresponding to the two possible magnetization orientations of the free layer with respect to the reference layer (parallel and antiparallel). This behavior is illustrated in Figure 4.6, where, for both magnetic configurations, the I-V curves are shown in blue, the experimental gradient in red, and the numerical gradient in green.

In Figure 4.6a, the device is in the parallel state, while Figure 4.6b corresponds to the

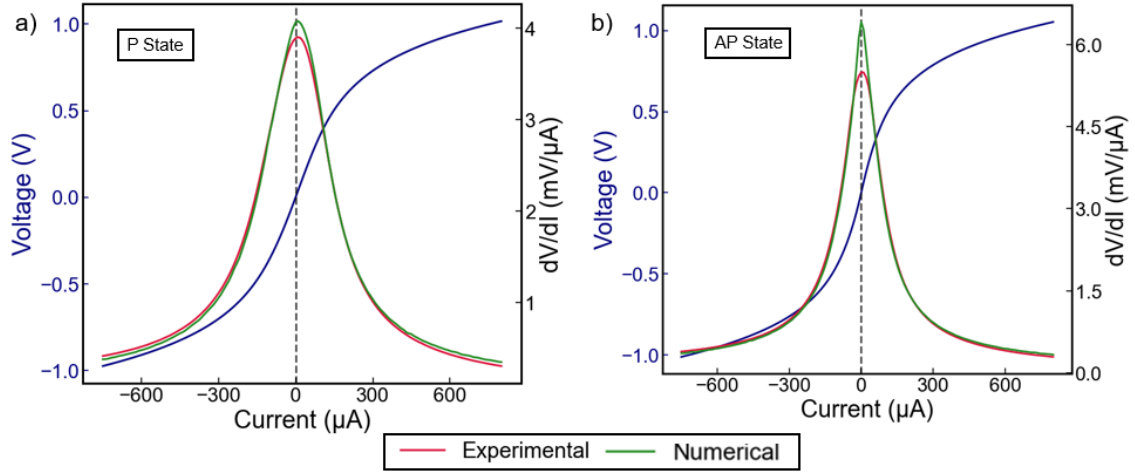


Figure 4.6: Measured I-V curves (blue), experimental gradient (red), and numerical gradient (green) for the two magnetization states of the device: a) parallel (P) and b) antiparallel (AP).

antiparallel state. As expected, the AP configuration presents higher voltage values due to its larger resistance. Overall, the experimental results display a very good agreement with the numerical calculation, with only a small deviation observed near zero current. This discrepancy can be attributed to the difficulty of extracting precise derivative values in that region, as the measurement resolution is limited and the number of sampling points around $I \approx 0$ A may be insufficient. However, the overall shape of the activation functions, and consequently their gradients, remains qualitatively similar for both magnetize states.

4.3.2 Nitrogen Incorporation in the Free Layer

In order to identify devices exhibiting greater asymmetry in their electrical response, two additional configurations were investigated. These alternative designs aim to enhance the asymmetry in activation functions, thereby potentially increasing their richness and variability. Furthermore, this study also assessed whether the proposed gradient calculation method remains effective when applied to more complex activation function responses.

The two additional configurations are referred to as Stack C (S_C) and Stack D (S_D). Both share the same pinned reference layer as S_A and S_B , and in both cases the MgO tunnel barrier is undoped (i.e., no nitrogen incorporation). The key distinction lies in the free layer composition, which incorporates nitrogen through a NiFeN sublayer. Figure 4.3c and Figure 4.3d illustrate the schematic structures of S_C and S_D , respectively. The free layer of S_C consists of CoFeB (2 nm) / Ta (0.21 nm) / NiFeN (3.5 nm). In S_D , an additional undoped NiFe layer (3.5 nm) is introduced on top of the NiFeN, resulting in the following

free layer structure: CoFeB (2 nm) / Ta (0.21 nm) / NiFeN (3.5 nm) / NiFe (3.5 nm). The lateral dimensions of the devices are identical to those listed in Table 4.1.

At standard annealing conditions (330 °C for 2 h under 1 T), the performance of the two configurations was compared. Figure 4.7 shows the dependence of TMR on RA for both configurations. The S_C configuration exhibits a much lower RA ($\approx 10 \text{ k}\Omega \mu\text{m}^2$) compared to S_D ($\approx 200 \text{ k}\Omega \mu\text{m}^2$). In addition, S_C reaches a TMR of approximately 80%, while S_D is limited to about 25%. The difference in electrical performance can be explained by the strong affinity of tantalum (Ta) for nitrogen. Ta readily forms stable nitrides such as TaN and Ta₃N₅ [72, 73]. In configuration S_D , the additional NiFe layer acts as a buffer between NiFeN and Ta, restricting nitrogen diffusion toward the Ta. As a consequence, more nitrogen remains available to diffuse into the MgO barrier, which explains the higher resistance observed in S_D . This interpretation is consistent with the comparison of the devices from S_A and S_B , which only differed in the nitrogen doping of the barrier.

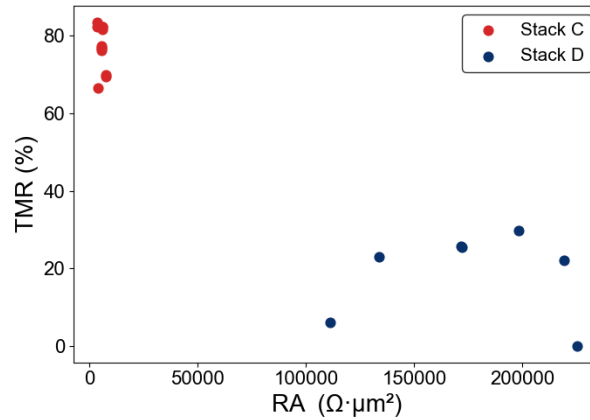


Figure 4.7: Dependence of TMR on RA for the Stack C (S_C) and Stack D (S_D).

The variability analysis shown in Figure 4.8 further supports this conclusion. Although Stack S_D was compared against S_C using devices of different dimensions, as shown in Figure 4.8a, a less reliable metric, S_D consistently exhibits lower variability. While smaller devices (e.g., position P2) generally show higher variability, the mean of standard deviations (Fig. 4.8b) demonstrates that S_D delivers more stable I–V responses, as reflected by the lower variability values, particularly at positions P4 and P5 where a direct comparison between the two stacks is possible. This observation is consistent with the earlier finding that higher nitrogen incorporation into the MgO barrier (leading to higher resistance) improves uniformity across devices.

To further investigate S_D , devices annealed at different temperatures (200 °C, 250 °C and

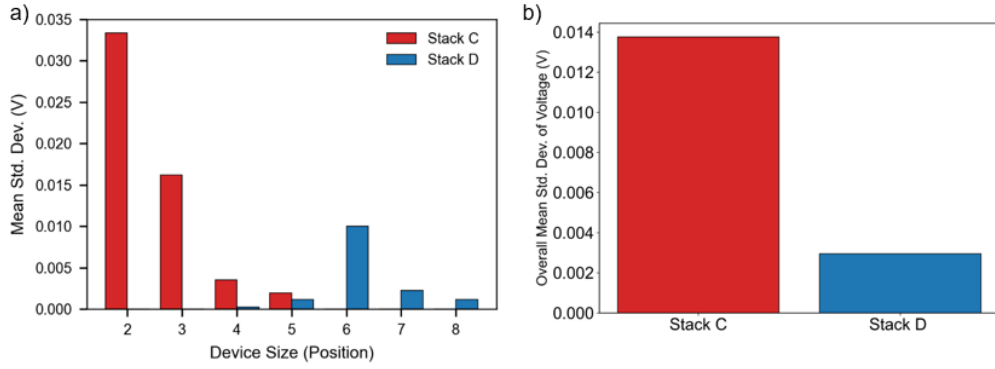


Figure 4.8: Device-to-device variability analysis of S_C and S_D . a) Standard deviation of voltage for each diameter according to the designation of Table 4.1. b) Average of standard deviations across all geometries.

330 °C) were tested. As shown in Figure 4.9, RA increases and TMR decreases with annealing temperature. This behavior can be attributed to enhanced nitrogen diffusion into the MgO barrier at higher temperatures, reinforcing the hypothesis that nitrogen incorporation plays a key role in tuning the electrical properties of the devices.

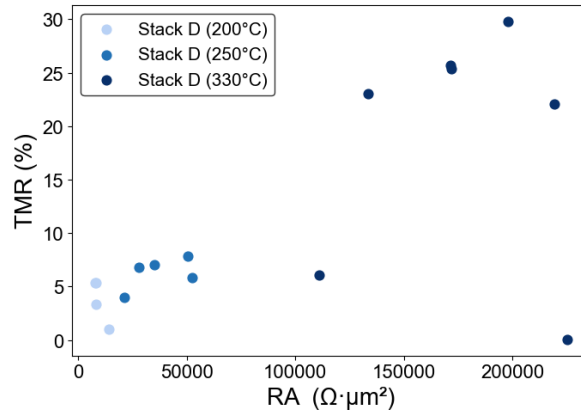


Figure 4.9: Dependence of TMR on RA for S_D with different annealing temperatures.

The I–V curves of S_D at different annealing temperatures were also analyzed, with the gradient calculated for both the parallel and antiparallel magnetization states. Figure 4.10 shows that the experimental gradients (red) are in excellent agreement with the numerical gradients (green), validating the effectiveness of the gradient calculation method. Moreover, the results reveal asymmetric gradients, as the device behavior differs for positive and negative current bias ($I = 0$ A). The ability to obtain gradients of more complex activation functions is particularly important, as it demonstrates that this method can be applied to functions that conventional numerical methods may struggle to handle.

Another important observation is the distinct behavior between the P and AP states. This difference could be leveraged to exploit the intrinsic non-volatility of MTJs, potentially

enabling the realization of distinct activation functions based on the magnetization state. Such a feature could be valuable for neuromorphic applications, where the availability of multiple activation functions and the use of non-volatile states can improve learning performance [74].

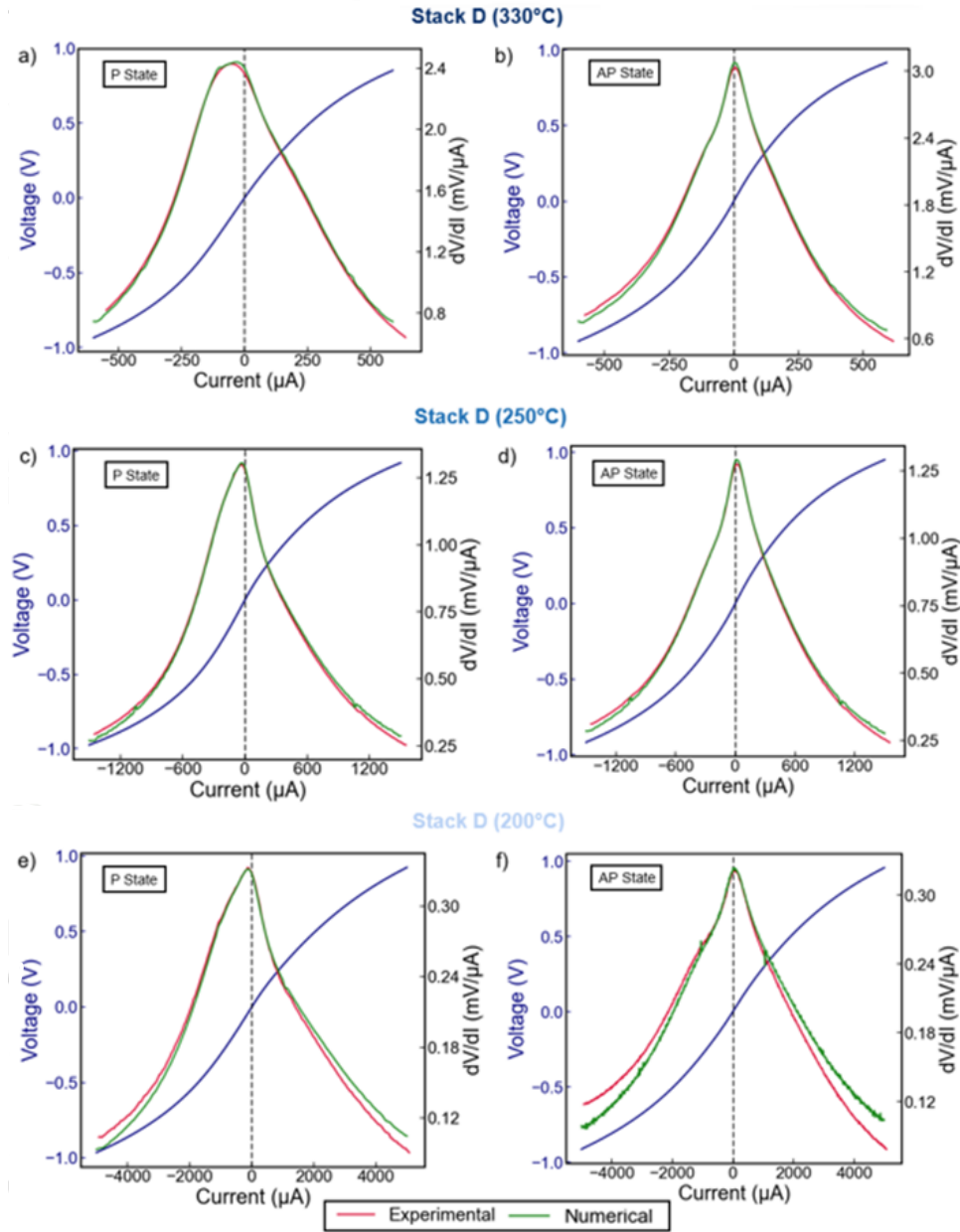


Figure 4.10: Measured I-V curves (blue), experimental gradient (red), and numerical gradient (green) for the two magnetization states of the device: parallel (P) in panels (a, c, e) and antiparallel (AP) in panels (b, d, f). Results are shown for Stack D annealed at 330 °C (a, b), 250 °C (c, d), and 200 °C (e, f).

4.3.3 Comparison between Configurations

After selecting the best sample for the main two different configurations, i.e. S_A and S_D , their electrical responses were analyzed with respect to the asymmetry and nonlinearity

of the activation functions. Starting with the asymmetry analysis, the asymmetry error was defined as:

$$\epsilon_V(I) = \left| \frac{V(-I)}{V(I)} - 1 \right| \times 100\%, \quad (4.4)$$

which measures the relative deviation of the voltage values for symmetric current points.

For each curve with N points, two metrics were calculated:

- **Mean asymmetry error per curve:**

$$\overline{\epsilon}_V = \frac{1}{N} \sum_{j=1}^N \epsilon_{V,j}, \quad (4.5)$$

- **Maximum asymmetry error per curve:**

$$\epsilon_V^{\max} = \max_{1 \leq j \leq N} \epsilon_{V,j}. \quad (4.6)$$

Since multiple curves were obtained for each configuration, the average values of these metrics across L curves were computed as:

- **Mean of the mean asymmetry error:**

$$\overline{\overline{\epsilon}}_V = \frac{1}{L} \sum_{i=1}^L \overline{\epsilon}_{V,i}, \quad (4.7)$$

- **Mean of the maximum asymmetry error:**

$$\overline{\epsilon}_V^{\max} = \frac{1}{L} \sum_{i=1}^L \epsilon_{V,i}^{\max}. \quad (4.8)$$

In Figure 4.11, S_A annealed at 200 °C exhibited the lowest asymmetry error, with a mean value of 4.1% and a maximum of 7.7%. In contrast, S_D consistently presented higher asymmetry errors across all annealing conditions. The largest discrepancy was observed at 200 °C, where the $\overline{\overline{\epsilon}}_V$ reached 8.4% and the $\overline{\epsilon}_V^{\max}$ 13.6%.

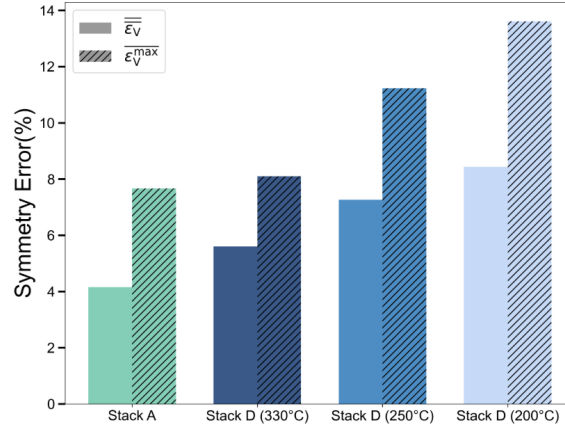


Figure 4.11: Comparison of asymmetry errors for S_A and S_D at different annealing temperatures.

Then, the nonlinearity was quantified in terms of the mean squared error (MSE), defined as:

$$MSE(\%) = \frac{1}{N} \sum_{k=1}^N (V_k - V_{fit}(I_k))^2 \times 100\%, \quad (4.9)$$

where the fitted voltage is expressed as $V_{fit}(I) = mI + b$. This metric evaluates the deviation of the measured I–V curve from its corresponding linear fit. Since not all curves extend exactly to ± 1 V, the analysis was restricted to the interval $V \in [-0.9, 0.9]$ V. For each curve with N points, the following metrics were defined:

- **Mean nonlinearity per curve:**

$$\overline{MSE}_S = \frac{1}{N} \sum_{j=1}^N MSE_j, \quad (4.10)$$

- **Maximum nonlinearity per curve:**

$$MSE_S^{\max} = \max_{1 \leq j \leq N} MSE_j. \quad (4.11)$$

For the L curves obtained per configuration, the overall averages were then computed as:

- **Mean of the mean nonlinearity:**

$$\overline{\overline{MSE}} = \frac{1}{L} \sum_{i=1}^L \overline{MSE}_{S_i}, \quad (4.12)$$

- **Mean of the maximum nonlinearity:**

$$\overline{MSE}^{\max} = \frac{1}{L} \sum_{i=1}^L \max MSE_{S_i}. \quad (4.13)$$

In Figure 4.12, the values of $\overline{\overline{\text{MSE}}}$ and $\overline{\text{MSE}}^{\text{max}}$ are presented for both configurations, S_A and S_D . The results indicate that nonlinearity is significantly more pronounced in S_A , reaching 5.1% for $\overline{\overline{\text{MSE}}}$ and 10.4% for $\overline{\text{MSE}}^{\text{max}}$. In contrast, S_D exhibits considerably lower nonlinearity, with $\overline{\overline{\text{MSE}}}$ remaining below 3% across all annealing temperatures, and the highest $\overline{\text{MSE}}^{\text{max}}$ limited to 4.9%.

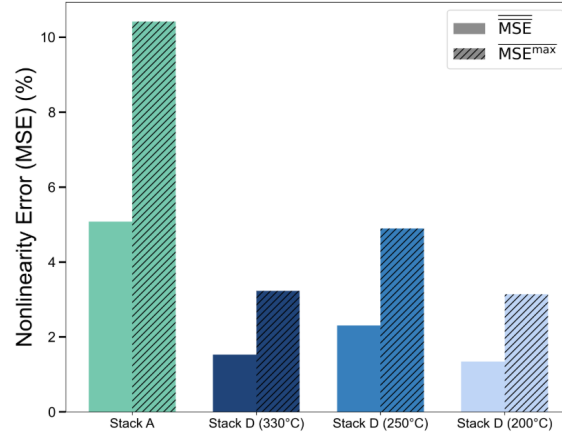


Figure 4.12: Comparison of nonlinearity metrics ($\overline{\overline{\text{MSE}}}$ and $\overline{\text{MSE}}^{\text{max}}$) for S_A and S_D at different annealing temperatures.

Therefore, while S_A shows a more symmetric behavior, it also demonstrates greater nonlinearity, compared to S_D . The following section will analyze the performance of these activation functions and their evaluated gradients within a neural network.

4.4. Neural Network Implementation: Experimental and Simulation Approach

To evaluate the performance and reliability of these MTJ-neurons it was carried out a classification task based on the Iris dataset. This dataset requires distinguishing between three classes of iris flowers (Iris Versicolor, Iris Setosa and Iris Virginica) based on four input features: the length and width of the petals and sepals. The task was implemented in a device-in-the-loop configuration, where at each training step, both the I-V derived activation functions and their experimentally measured gradients were incorporated. The training was performed using the backpropagation algorithm with a 5-fold cross-validation scheme, which divides the dataset into five subsets, iteratively using one for validation while the remaining four are used for training. The use of cross-validation provides a more robust evaluation by reducing bias introduced by a single data partition and providing more consistent results across different data splits. In this way, a simulation-based analysis was conducted. Full I–V curves and their associated experimentally measured gradients were

used to construct a Look-Up Table (LUT), enabling emulation of the complete training process (detailed simulation protocol in Fig. 4.15).

First, the activation functions from S_A and S_D are compared. Figure 4.13b shows the I-V curves (activation functions) obtained from S_A (green), S_D (blue), and S_A restricted to a narrower range (grey), where the behavior is more linear. The motivation for this restriction will be explained in the following paragraphs. The neural network architecture consisted of a single hidden layer (Figure 4.13a), where three different configurations were tested, each using activation functions from the different curves represented in Figure 4.13b. Figure 4.13c and Figure 4.13d present, respectively, the mean validation accuracy and the loss across the five folds, with the shaded regions indicating the corresponding standard deviation.

The results obtained with neurons from S_A are shown in green, while those using neurons from S_D are shown in dark blue. For S_D , the network initially converged to a relatively high accuracy but began to deteriorate after approximately 300 epochs: accuracy decreased while loss increased. Even before this point, the performance was unstable. In contrast, S_A exhibited stable convergence, consistently reaching a high accuracy of 95%. This difference can be attributed to the lower nonlinearity of the S_D activation functions, which limits network performance.

To further investigate this effect, an additional simulation was conducted using the same activation functions from S_A , but restricted to the low-bias region, where the response is nearly linear. This case, shown in grey in Figure 4.13b, where the nonlinearity becomes more pronounced at higher bias.

The results shown in Figure 4.13c and Figure 4.13d demonstrate that this quasi-linear case reproduced the same behavior observed for S_D , i.e., the accuracy decreased after a certain number of epochs, confirming the importance of nonlinearity in achieving high network performance, and indicating that the asymmetry of these curves has a minor impact. Therefore, the analysis of different activation functions arising from the non-volatile states was not carried out, since the observed asymmetry was not relevant to this study.

It is worth noting that all simulations in this section were rescaled on the current axis to the range -5 to 5 . This range was chosen to ensure that, after multiplication by the weights, the dataset inputs would fall within the activation function region. This is because the original current values were in the μA range (thus being very small) and varied depending

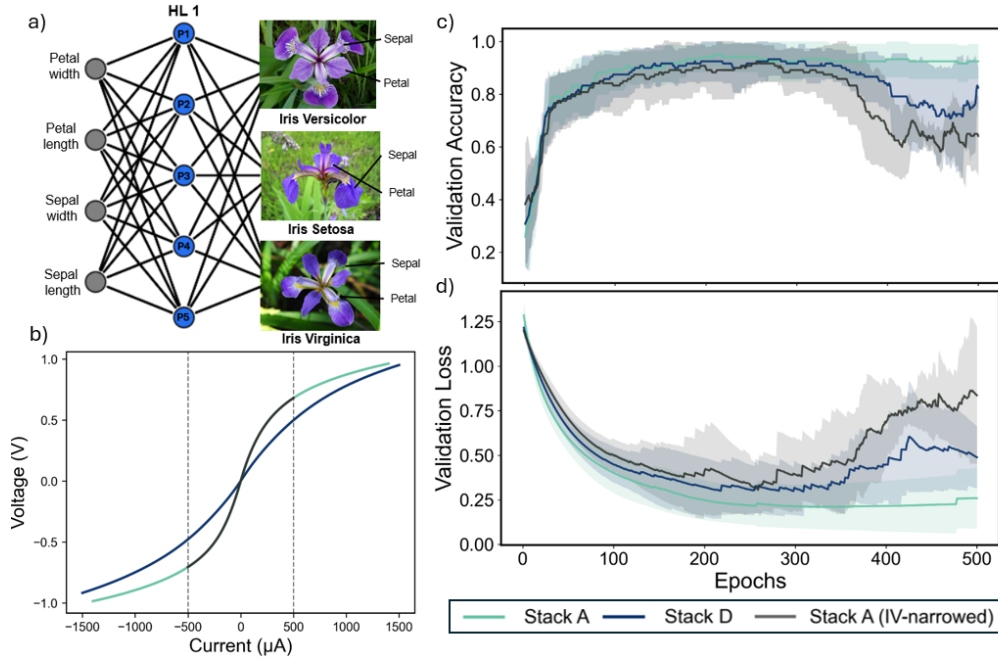


Figure 4.13: Training simulation of neural network using MTJ-based neurons from S_A and S_D for Iris classification. a) Architecture with one hidden layer. b) Activation functions obtained from the I-V curves for S_A (green), S_D (dark blue), and S_A restricted to the low-bias region (grey). c), d) Mean validation accuracy and loss, respectively, obtained via 5-fold cross-validation over 500 epochs. Shaded areas indicate the standard deviation across folds. Green curves correspond to networks using neurons derived from S_A , blue curves correspond to networks using neurons derived from S_D , and grey curves correspond to networks derived from S_A but restricted to the low-bias region.

on device size, leading to poorly optimized activation functions. Figure 4.14 illustrates a representative input distribution for fold 5 using MTJ-neurons from S_A , where the inputs to the activation function are shown in blue before training and in orange after training. Before training, nearly all inputs fall within the central portion of the activation function (approximately -2 to 2). After training, the distribution broadens (approximately -6 to 6) while retaining the highest density near zero, which is important because it ensures that most inputs stay in the useful nonlinear region, avoiding pushing neurons into pure saturation (flat regions) where the gradient vanishes. In case the inputs of the activation function fell outside the -5 to 5 range, it was clipped to the corresponding boundary value: inputs below -5 were assigned the voltage of the lowest measured point, while inputs above $+5$ were assigned the voltage of the highest measured point. Accordingly, when the I-V curve of S_A was narrowed, it was also rescaled to the same -5 to 5 range to guarantee a fair comparison.

Thus, given the superior results obtained with S_A , a fully experimental device-in-the-loop training process was carried out. In this approach, the real behavior of five MTJs was

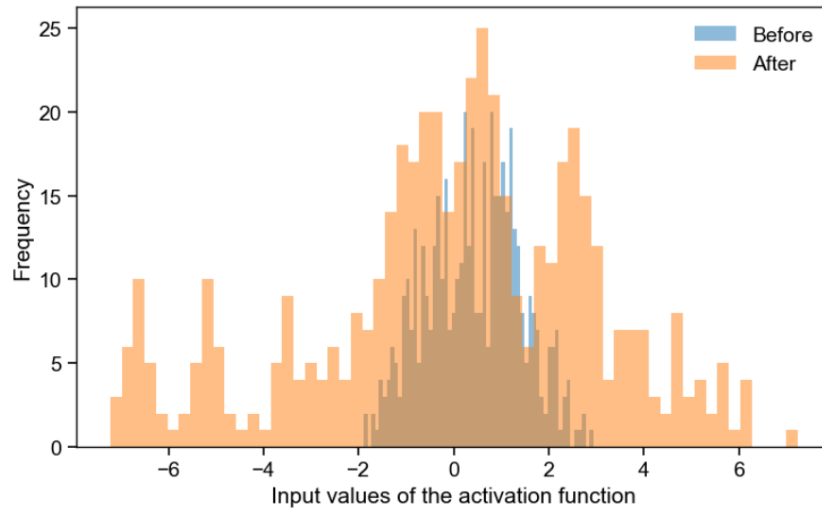


Figure 4.14: Inputs distribution of the activation functions before (blue) and after (orange) training.

directly used during both forward and backward passes. During training, activation functions were measured in real time from the devices, while the gradients measured beforehand through the analog finite method were incorporated into a developed LabVIEW training environment, which executed the backpropagation algorithm and calculated the accuracy and loss using the 5-fold cross-validation technique. In this setup, five different MTJs were connected to a separate sourcemeter to apply the current and measure the corresponding voltage directly.

To provide a comparison, simulations were performed under the same initial conditions as the experimental setup. Figure 4.15 illustrates the steps involved in both the experimental and simulated protocols, providing a more comprehensive overview of the two approaches implemented. The MTJ pillar diameters varied across the five devices: three devices had diameters of 4000 nm, one had a diameter of 5000 nm, and one had a diameter of 7500 nm.

Experimental training was then conducted to a network with one hidden layer as shown in Figure 4.16a. The devices are labeled as P1, P2, P3, P4, and P5 in the figures for clarity and the respective I-Vs and experimental gradients are represented in Figure 1 of the attachments. The experimental results for the validation accuracy and loss are represented by the light-blue curves in Figure 4.16c and Figure 4.16d. While two network architectures were tested in the simulation approach: one with a single hidden layer, matching the experimental case (Figure 4.16a), and another with two hidden layers, each having the same composition as the single layer in the first architecture, as represented in

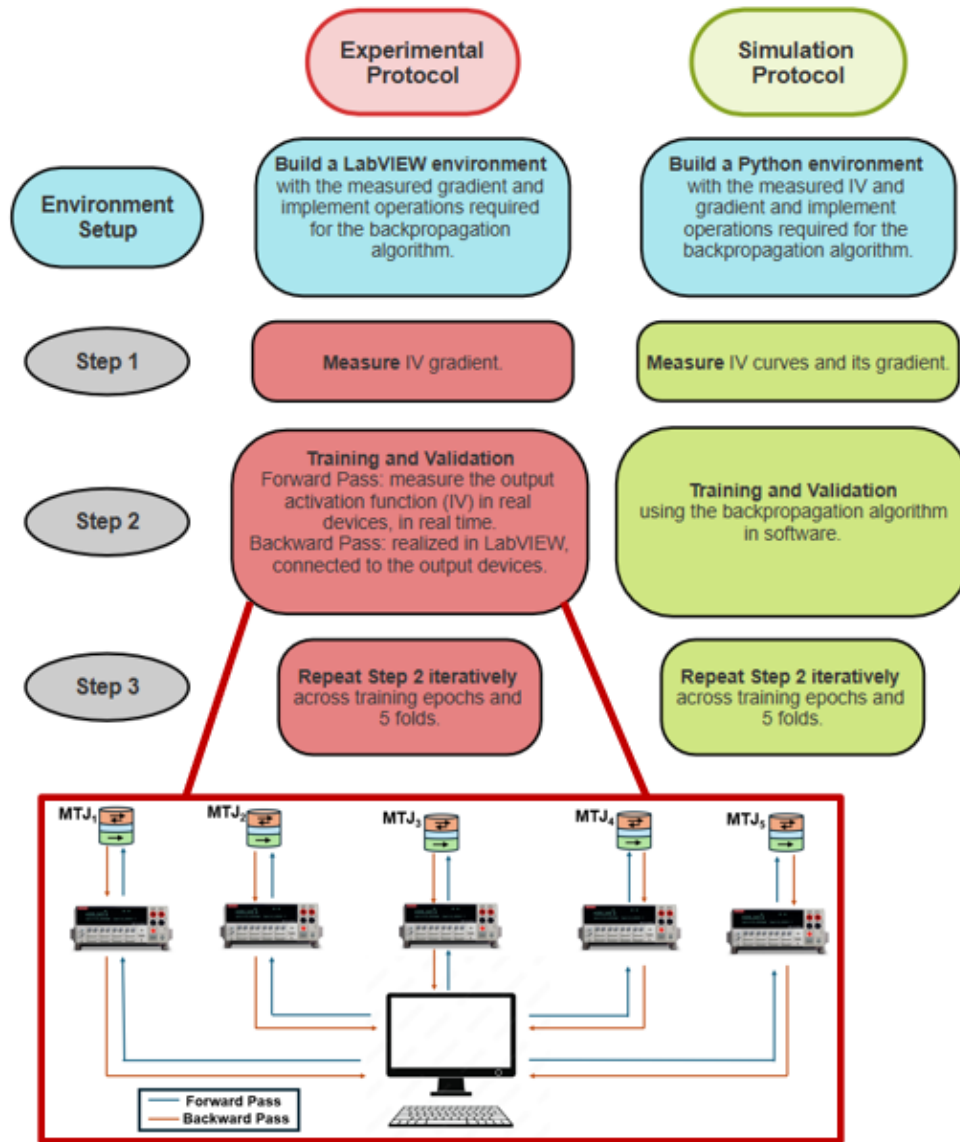


Figure 4.15: Implemented protocol for the experimental and simulated approach.

Figure 4.16b. The simulation results are shown in Figure 4.16c and Figure 4.16d, where the dark blue curves correspond to the simulation results with one hidden layer, while the orange curves represent the simulation results with two hidden layers.

In both the experimental and simulation-based approaches, training was conducted using two learning rate (LR) schedules: a constant rate of 0.002 over 300 epochs (Figure 4.16c), and a linear decay schedule decreasing from 0.01 to 0.0088 over 120 epochs (Figure 4.16d). In both cases, convergence was consistent and no overfitting was observed, where accuracy increased smoothly across epochs, confirming the stability and robustness of the training procedure. However, the linear decay schedule systematically outperformed the constant learning rate. In the experimental case, the linear decay schedule achieved a validation accuracy of 93.3%, compared to 89.2% for the constant

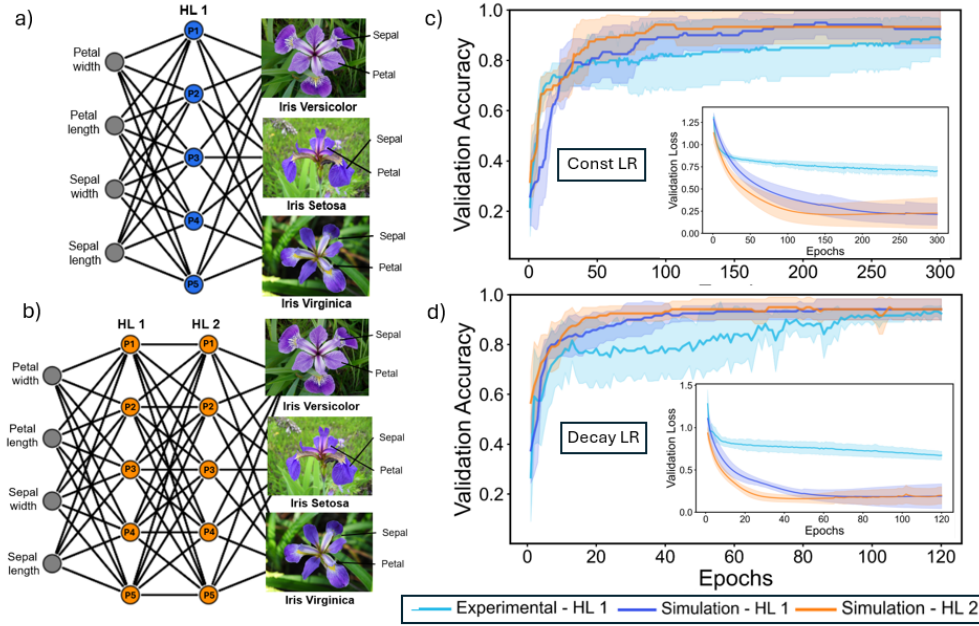


Figure 4.16: Experimental and simulated training of a neural network using MTJ-based neurons from S_A for Iris classification. a) Architecture with a single hidden layer (HL 1) of five distinct MTJ neurons represented in blue, b) Architecture with two hidden layers represented in orange (HL 1 and HL 2), where the second layer is identical to the first. (c, d) Mean validation accuracy obtained via 5-fold cross-validation. Shaded areas indicate the standard deviation across folds. Blue curves correspond to networks with one hidden layer (light = experimental, dark = simulation), while orange curves correspond to two hidden layers (simulation only). The Insets represent the validation loss over the epochs. c) Training with a constant learning rate of 0.002 over 300 epochs. d) Training with a linearly decaying learning rate from 0.01 to 0.0088 over 120 epochs.

rate. In the simulations with one hidden layer, slightly higher performance was obtained compared to the experimental implementation, achieving a 94.2% validation accuracy for the linear decay rate and 92.5% for the constant rate case. Despite this small difference, the experimental results demonstrated that device-in-the-loop training is effective: the network successfully learned the underlying data distribution using real MTJ devices, and the observed trends closely matched those from simulation.

The simulations of a two-hidden-layer network confirmed the feasibility of multilayer training with device-based nonlinearities. Validation accuracy reached 94.2% with a constant learning rate and 95% with the linear decay schedule, surpassing the single-layer case. These results highlight the scalability and effectiveness of using device-generated nonlinearities in deeper neural networks, addressing a limitation of previous approaches that were often restricted to a single hidden layer due to the difficulty of analytically modeling device-specific gradients [75].

The Table 4.2 summarizes the results obtained for the Iris task.

Table 4.2: Maximum accuracy achieved by different neural network models with constant and decaying learning rates.

NN model	Constant LR (300 epochs)	Decay LR (120 epochs)
Experimental - 1 Layer	89.2%	93.3%
Simulation - 1 Layer	92.5%	94.2%
Simulation - 2 Layers	94.2%	95.0%

To study scalability, a two-layer neural network was evaluated on a simulation level for the MNIST handwritten digit classification task. Figure 4.17a shows the architecture of the network with two hidden layers, where the first one has 128 neurons and the second has 64 neurons. As in the previous cases, the simulations were performed using activation functions and gradients extracted from previously characterized pairs of distinct devices. In this case, the training was conducted with 15 different devices randomly assigned across the two layers. Also, to compare the performance of the MTJ-based neurons, a simulation was conducted using a conventional hyperbolic tangent ($\tanh$) function. The results are summarized in Figure 4.17b shows the validation accuracy over 50 epochs, with an inset displaying the validation loss across the training epochs. The orange curve corresponds to the MTJ-based neurons, while the green curve represents the standard $\tanh$ function.

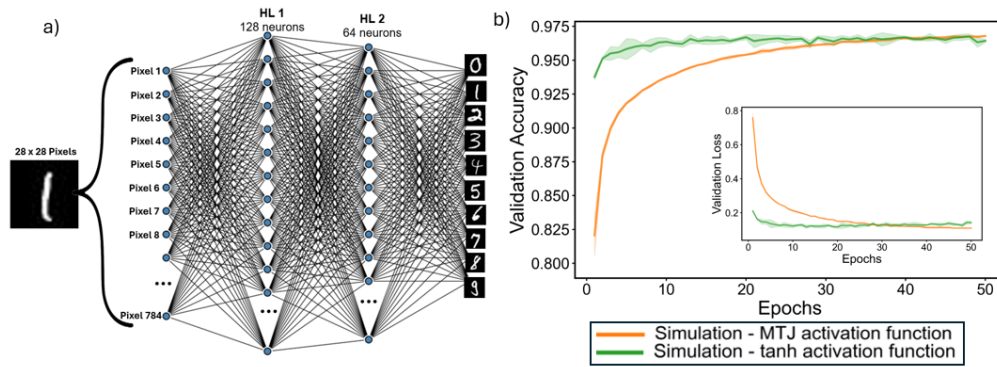


Figure 4.17: Training simulation of a neural network using MTJ-based neurons and tanh activation for MNIST classification. a) Architecture with two hidden layers with 128 and 64 neurons, respectively. b) Mean validation accuracy obtained via 5-fold cross-validation over 50 epochs. Shaded areas indicate the standard deviation across folds. The Inset represents the validation loss over the epochs. Orange curves correspond to results using MTJ neurons, while green curves correspond to results using the tanh activation function.

The final validation accuracy achieved was 96.8% for the MTJ-based system and 96.7% for the conventional digital neural network. Although the results from the tanh function

demonstrates faster convergence, the final accuracy after 50 epochs is quite similar between the two approaches. The MTJ-based neurons model achieves slightly higher accuracy (by less than 0.1%), indicating that the MTJ-based neurons can closely match the performance of conventional functions, even when applied to more complex tasks. This section has demonstrated impressive results when applying MTJ-based neurons to both experimental and simulated approaches, showing that the gradient accuracy achieved through the analog method is sufficient to support backpropagation across multiple layers, even in the context of more complex tasks.

5. Weighted Spin-Torque-Nano-Oscillators

This chapter focuses on an approach that combines artificial synapses and artificial neurons within the same configuration. MTJs function as MRAM devices due to their non-volatility, which is derived from their P and AP states. This characteristic makes them suitable for emulating synaptic weights (Section 5.1). Additionally, MTJs can leverage the nonlinearity of STNOs while oscillating, functioning as artificial neurons (Section 5.2).

For the MRAM implementation, five devices with different diameters were connected in series. Since the resistance values for the P and AP states differ among the devices, each MTJ can switch independently, giving rise to 32 (2^5) distinct states. Figure 5.1 illustrates the layout of the configuration used to study the five MTJs in series, with different dimensions. In this figure, two series of five MRAMs are represented, where both are connected to an STNO. The figure also shows the top and bottom contacts for each of the devices.

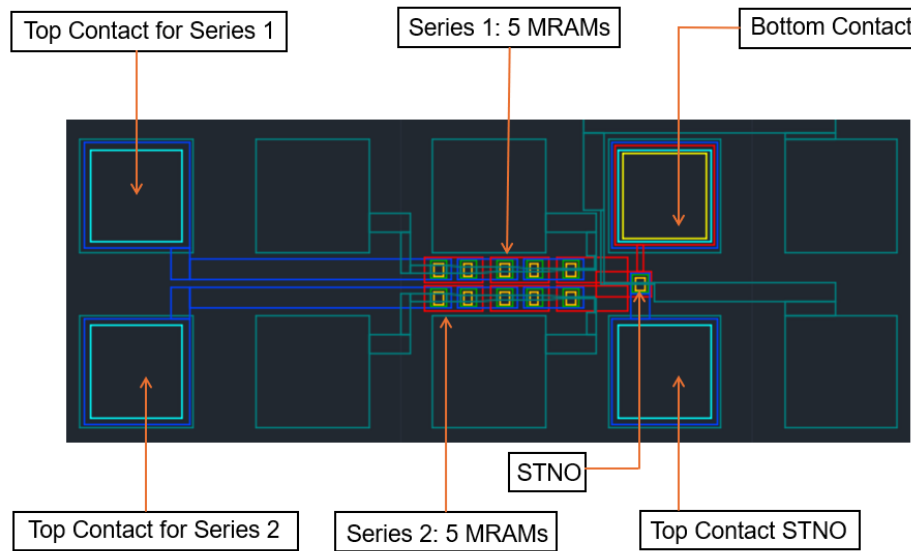


Figure 5.1: Layout of the configuration tested with two series of 5 MRAMs and a STNO.

The final goal of this configuration is to use the vortex STNO as a neuron, generating a nonlinear output based on the weighted sum of the inputs provided by the series of MTJs. The five devices were fabricated using the same configuration as the STNO, requiring a design compromise that enabled both proper STNO operation and reliable weighted MRAM functionality. Consequently, these devices can be referred to as weighted spin-torque nano-oscillators (WSTNOs). To realize a vortex-based STNO, the stack design

incorporated a thick free layer, as indicated in Figure 2.12a, with a circular nanopillar geometry.

For WSTNOs designed with memory functionality, elliptical nanopillars were employed to achieve a uniformly in-plane magnetized free layer. The additional shape anisotropy introduced by this geometry stabilizes a bi-stable magnetization state, making the devices suitable for memory applications. The integration of both functionalities within a single fabrication process is highly advantageous, as it reduces the overall footprint, ensures compatibility with CMOS technology, and provides a promising pathway toward the development of more complex neuromorphic computing systems.

5.1. Multi-State Devices

The configuration presented to analyse the MTJs with different diameters connected in series, can be defined as binary scaling: if the resistance of each MTJ is designed such that one device has twice the resistance of the previous, then for N number of MRAMs, there are 2^N distinct states, since each combination corresponds to a unique total resistance. This scaling can be implemented by reducing by half the device area successively while keeping the same RA product. However, this approach becomes challenging for larger N , as it requires devices to cover a very wide range of areas.

The evaluation of the artificial synapses began with the measurement of the hysteresis loop of the series. Since each device has a different size, it also exhibits a different coercive field, i.e., the magnetic field strength required to switch between the P and AP states will be different. The objective of this characterization is to observe five distinct resistance jumps on each branch of the curve. The devices were fabricated with a target RA of $10 \Omega \mu\text{m}^2$, and a TMR of approximately 100%. Therefore, devices with larger area, have lower resistance and a smaller difference between the R_P and R_{AP} , while also requiring a smaller coercive field to switch states. However, another important factor to consider is the influence of the aspect-ratio (AR) of the elliptical shape. Elliptical devices with a longer axis aligned along the easy axis of the magnetic material enhance anisotropy in that direction, thereby increasing the coercive field. This enhancement grows with increasing AR. Taken together, these two factors indicate that small devices with high aspect ratio ellipses maximize the effective coercive field of the MTJs.

Two configurations were analyzed, denoted as stacks S_1 and S_2 , where both have the same pinned reference layer (6 nm IrMn / 2 nm CoFe₃₀ / 0.825 nm Ru / 2.6 nm CoFe₄₀B₂₀)

and insulating barrier (≈ 1 nm MgO). The primary distinction between these configurations lies in the free layer; specifically, in S_1 the free layer is 1.25 nm $\text{CoFe}_{40}\text{B}_{20}$ / 0.21 nm Ta / 2.5 nm CoFeSiB , whereas in S_2 it is 2 nm $\text{CoFe}_{40}\text{B}_{20}$ / 0.21 nm Ta / 2.5 nm CoFeSiB . Thus, the difference is in the thickness of the $\text{CoFe}_{40}\text{B}_{20}$ layer, which is smaller in S_1 than in S_2 . Figure 5.2a and Figure 5.2b illustrate the transfer curves for S_1 and S_2 , respectively, revealing that the second configuration exhibits significantly greater stability, characterized by more consistent jumps.

In contrast, the first configuration demonstrates a reduction in resistance at both the highest and lowest endpoints. This phenomenon may be attributed to the increased thickness of the free layer in S_2 , which enhances the energy barrier, as the barrier is dependent on the volume of the layer according to Equation 2.25. The increase of the energy barrier leads to an exponential increase in the Néel relaxation time (Equation 2.29), thereby resulting in a more thermally stable system that facilitates a prolonged switching duration.

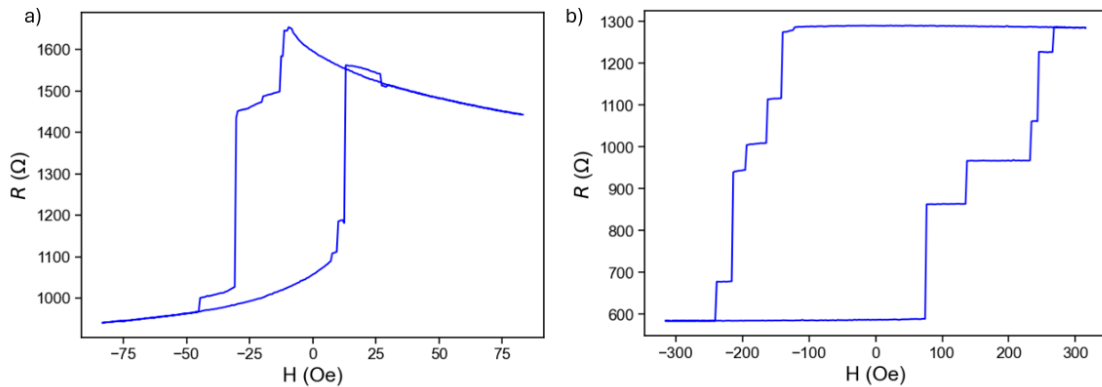


Figure 5.2: Comparison of the transfer curves for a series of 5 MRAMs for a) S_1 and b) S_2 .

Stack S_2 was further investigated, and Figure 5.3 presents a representative transfer curve in which five distinct resistance jumps are clearly observed. These jumps correspond to the transitions between the parallel and antiparallel magnetic states. The step heights associated with each transition are also shown in the figure. The overall trend is that smaller devices, such as those with a larger resistance jump, exhibit a smaller coercive field. It is worth noting from Table 5.1 that, as the device size decreases, the AR also becomes smaller. As previously explained, these two factors act in opposite directions: while reducing device size tends to increase the coercive field, lowering the AR tends to reduce it. This interplay may explain the overall trend observed in the measurements.

Nevertheless, the devices were fabricated with these dimensions because, on one hand, very small devices with high ARs can exhibit excessively large coercive fields, requiring stronger external magnetic fields to induce switching and, on the other hand, devices with larger dimensions but low AR tend to form multidomain states rather than remaining single-domain. As lateral size increases and AR decreases, the shape anisotropy weakens, reducing the preferential alignment direction (see Section 2.1.3.5). The magneto-static energy, which scales with volume (Section 2.1.3.3) outweighs the cost of creating domain walls, given by $\sigma_w \approx \sqrt{AK_{eff}}$ per unit area, where A is the exchange stiffness. Since K_{eff} is small for low AR devices, domain walls are relatively inexpensive, making multidomain formation energetically favorable.

Analysing Figure 5.3, the three smallest jumps also appear to be interchanged, as one would ideally expect a monotonic behaviour. This could be attributed to dipolar interactions between closely spaced devices, which can slightly shift the coercive fields. Since the smallest jumps exhibit coercive fields that are very close to each other, the effect becomes more pronounced in these cases.

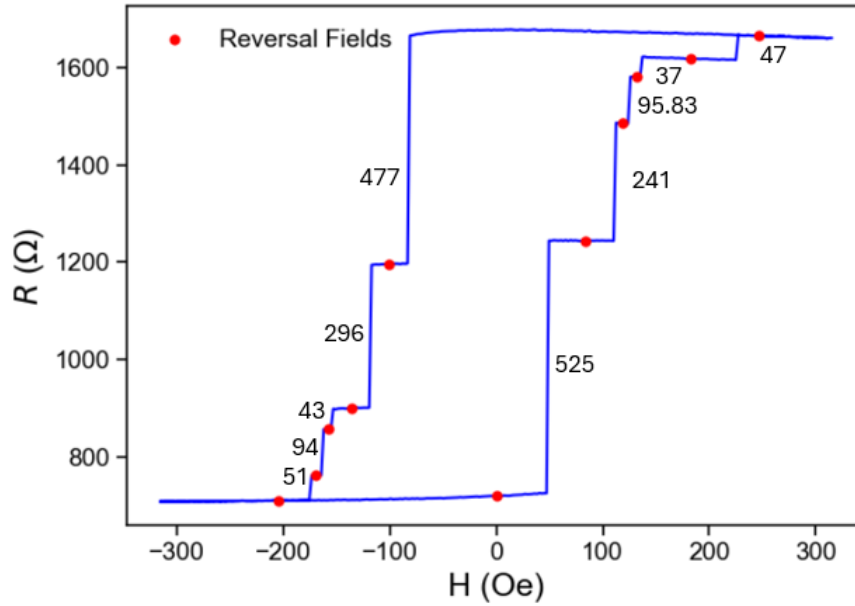


Figure 5.3: Transfer curve of a series of 5 MRAMs for S_1 , where the red dots indicate the reversal fields, and the values next to each jump represent their height.

Table 5.1 reports the size and area of the five devices under study. Since these devices were fabricated with a target RA of $10 \Omega \mu\text{m}^2$, the expected parallel resistance values for each configuration could be derived. By combining this with the TMR, the theoretical step heights were quantified. Based on the major loop, which shows a minimum resistance of 707Ω and a maximum of 1664Ω , the overall TMR of the stack was estimated to be 135%.

This value was adopted as the average TMR across the five devices. Ideally, each device should be characterized individually in order to determine its specific properties; however, individual devices with these dimensions were not available.

Table 5.1: Device size, area, step height and symmetry analysis.

MM pos.	x (μm)	y (μm)	A (μm ²)	AR	Step height (Ω) (TMR ≈ 135%)	Step height (Ω) (+ side)	Error (%)	Step height (Ω) (- side)	Error (%)	Symmetry (%) (+ vs -)
1	0.1	0.3	0.02	3	572.9	525.4	8.3	477.3	16.7	10.1
2	0.131	0.458	0.05	3.50	286.5	240.9	15.9	296.0	3.3	18.6
3	0.173	0.693	0.09	4.00	143.4	95.8	33.1	94.8	33.9	1.1
4	0.231	1.039	0.19	4.50	71.6	47.3	33.9	51.0	28.8	7.3
5	0.31	1.545	0.38	5.00	35.9	37.4	4.2	43.9	22.3	14.8

Table 5.1 also presents the experimental step heights obtained from the transfer curves, for both the positive and negative branches, along with the associated errors relative to the theoretical values. These discrepancies can be due to device-to-device variability. Although the nominal RA considered was the same, small fabrication variations, such as in device area, barrier thickness, or edge roughness, can lead to measurable differences in resistance. Furthermore, while a uniform TMR value was assumed for all devices, this assumption does not fully capture the intrinsic variability obtained in practice. Another important observation is the asymmetry between the positive and negative sides of the hysteresis loop. This asymmetry is quantified in the table as a percentage symmetry error, calculated by comparing the step heights between the two branches for each device. Such differences can arise from variations in the switching processes along the two branches. For example, these WSTNO devices may exhibit distinct reversal mechanisms, influenced by dipolar fields arising from neighboring patterned devices, as well as by intergranular variations in the free layer, like differences in grain size and crystalline anisotropy, that can further promote incoherent and nonuniform magnetization reversal, rather than the ideal coherent rotation [76].

After recording the transfer curve, minor loops were performed in order to selectively switch the devices into different resistance states. Unlike the transfer curve (major loop), where the magnetic field is swept all the way to saturation, minor loops restrict the field excursion so that only a specific device is switched while the others remain unchanged. In Figure 5.3, the red dots indicate the ten reversal fields chosen for this purpose: -204.2 , -169.0 , -157.7 , -136.3 , -101.2 , 83.6 , 118.7 , 131.3 , 182.8 and 246.9 Oe. Each device can be flipped by applying one of these reversal fields. A positive field induces a transition to the AP state, while a negative field switches the device to the P state. The assignment of reversal fields to each device is as follows:

- Device 5 (largest size): -204.2 Oe, $+246.9$ Oe
- Device 4: -169.0 Oe, $+182.8$ Oe
- Device 3: -157.7 Oe, $+131.3$ Oe
- Device 2: -136.3 Oe, $+118.7$ Oe
- Device 1 (smallest size): -101.2 Oe, $+83.6$ Oe

To represent these states systematically, an analogy with binary encoding is established, where the P state corresponds to 0, and the AP state corresponds to 1. This mapping allows us to represent any resistance configuration as a 5-bit binary word, yielding $2^5 = 32$ distinct states, where the least significant bit (LSB) corresponds to the largest device, while the most significant bit (MSB) corresponds to the smallest one.

In Attachments, Table 1 lists all 32 combinations of minor-loop sequences, the corresponding binary codes, and the measured final resistance values and Figure 2 shows the representation of these minor loops. As an illustration, let us consider state 5 from Table 1, obtained using the following minor-loop sequence:

$$0.63 \rightarrow 246.9 \rightarrow -169.0 \rightarrow 131.3 \rightarrow -136.3 \rightarrow 0.63.$$

The corresponding binary transitions are:

1. Start at 0.63 Oe: initial state.
2. Apply $+246.9$ Oe: all devices switch to AP $\rightarrow$ binary = 11111.
3. Reduce to -169.0 Oe: the four MSBs flip to P, while the LSB remains unchanged $\rightarrow$ binary = 00001.
4. Increase to $+131.3$ Oe: the three MSBs flip back to AP $\rightarrow$ binary = 11101.
5. Reduce to -136.3 Oe: the two MSBs flip to P $\rightarrow$ binary = 00101.

Thus, the final configuration is binary 00101, which corresponds to decimal state number 5.

Figure 5.4 shows the resistance values as a function of the 32 decimal states obtained from the minor loops. The experimental results are represented in red, while the expected

resistance values, calculated from the derived R_P and R_{AP} values of Table 5.1, are shown in blue. Dashed lines represent the corresponding linear fits. Although the expected and experimental results both follow a clear linear trend, some experimental states exhibit deviations from the ideal values. This could be due to the fact that the three larger devices are switched in the two branches of the major loop, resulting in the overlap of some of the 32 resistance states.

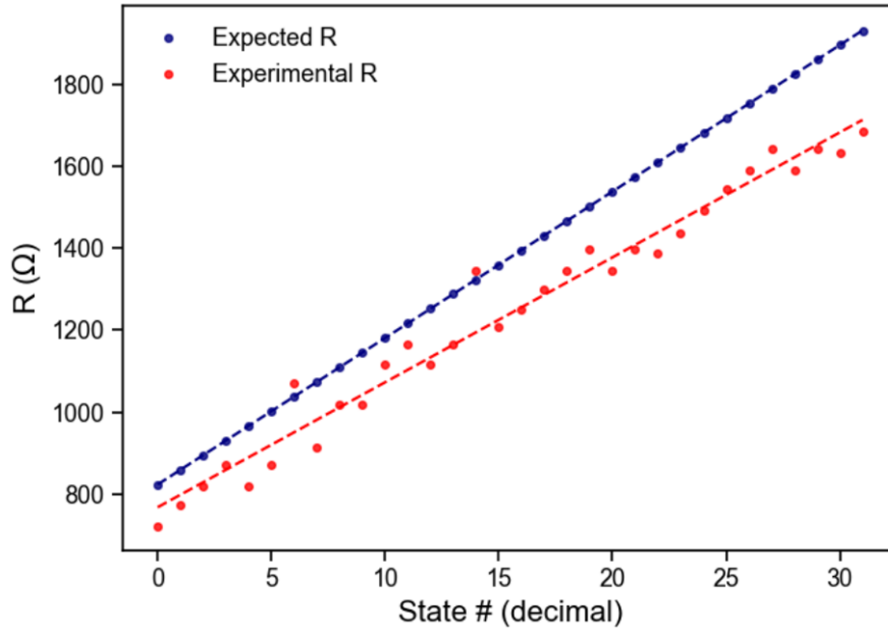


Figure 5.4: Resistance values as a function of decimal state, with expected values shown in blue and experimental data in red. Dashed lines indicate the linear fit for each results.

Table 5.2 summarizes the linear fit parameters for both datasets. For the expected values, the slope is 35.812 ± 0.001 and the intercept is 822.54 ± 0.03 , with a perfect correlation ($R^2 = 1.00$) and a very small standard error of the estimate ($S_{ey} = 0.06$). In contrast, the experimental fit gives a smaller slope of 30.551 ± 0.957 , corresponding to a relative error of 14.7%, and a lower intercept of 766.85 ± 17.26 , with a relative error of 6.8%. The correlation coefficient is $R^2 = 0.97$, and the standard error of the estimate increases substantially to $S_{ey} = 49.98$. These results show that, although the experimental data exhibit a larger spread and greater uncertainty compared to the ideal predictions, the overall linear dependence of resistance on state index is preserved, as evidenced by the correlation remaining close to unity.

5.1.1 Time Retention Analysis

Then, to study the stability of these devices, a time retention evaluation was performed. This involves first placing the device in a specific resistance state, similar to the minor loop

Table 5.2: Expected vs experimental linear fit for $R = m(\text{state \#}) + b$. Values are estimate $\pm$ standard error.

	m (slope)	b (intercept)	R^2	S_{ey}
Expected	35.812 ± 0.001	822.54 ± 0.03	1.00	0.06
Experimental	30.551 ± 0.957	766.85 ± 17.26	0.97	49.98
Error (%)	14.69	6.77	—	—

technique, and then continuously applying a bias current, measuring the resistance of the device every minute. This was done to analyze the thermal stability of the configuration and to determine if thermal effects could reach the thermal stability factor (Δ) and cause spontaneous switching of the magnetization direction in some devices, thereby changing the resistance state.

In Figure 5.5, the blue curve shows the transfer characteristic for the configuration used in this analysis, obtained with a bias current of $10 \mu\text{A}$. Unlike the transfer curve in Figure 5.3, this case is less smooth and well-defined, making the five switching events harder to distinguish and in a different order. Although this is not ideal to clearly identify the 32 resistance states, it was still suitable for performing the time-retention evaluation. The green curves represent three different initial states under a constant $10 \mu\text{A}$ bias: the parallel state, the antiparallel state, and an intermediate state. The figure presents the resistance over several hours, showing no evidence of switching during the entire measurement period. The devices remained stable for at least 20 h in the parallel state, 17.5 h in the intermediate state, and 3 h in the antiparallel state. It is therefore reasonable to expect that the actual stability times are significantly longer than reported here, particularly for the antiparallel state, where the measurement was limited to only 3 h. Longer-term studies are required to determine when switching would eventually occur. The experiment was also repeated using a higher bias current of 1 mA to further investigate the influence of thermal effects caused by Joule heating. The red curves correspond to the antiparallel and intermediate states under this higher bias. As expected, the absolute resistance values were lower than with $10 \mu\text{A}$, but the devices remained very stable and showed no unwanted switching. These results demonstrate that the devices can operate reliably as memory elements, even under long-term current stress.

These results demonstrate the potential of weighted synapses using non-volatile, thermally stable technology, although identifying a configuration with well-defined resistance states proved to be a challenge and required careful optimization. Therefore, as future

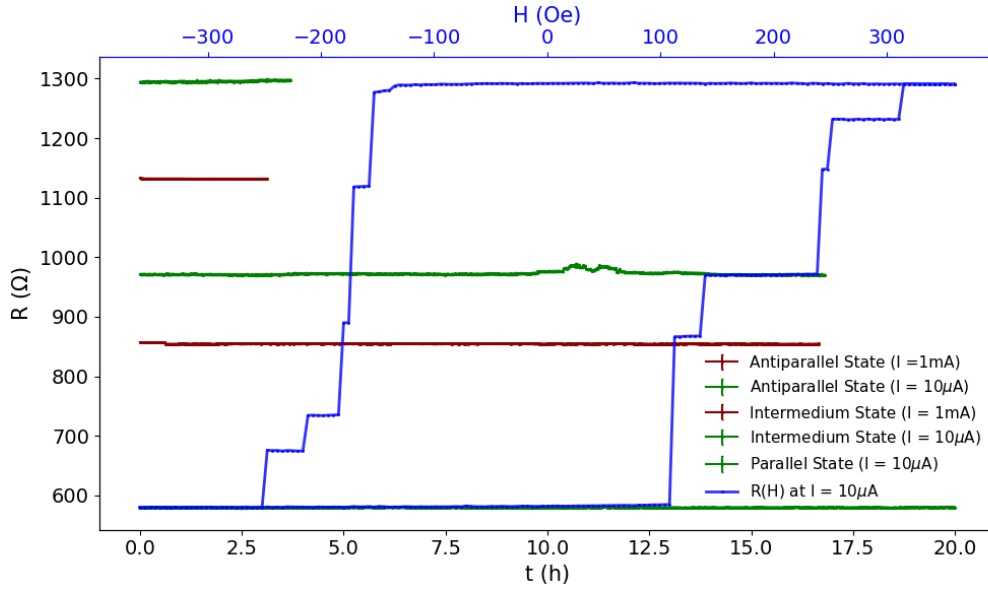


Figure 5.5: Time retention analysis. The blue curve represents the transfer curve of a series of five MRAM devices under a constant bias of $10\ \mu\text{A}$. The resistance stability over time for a programmed state is shown in green and red: the red curve corresponds to a bias of $10\ \mu\text{A}$, while the green curve corresponds to a bias of $1\ \text{mA}$.

work, it will be important to develop devices that can improve reliability and reduce device-to-device variation. However, it should be noted that as devices are manufactured at the nanoscale, the manufacturing process inherently involves variability.

5.2. STNOs

To analyze the best configuration of the vortex STNO, devices were studied separately on a different wafer with varying diameters of circular nanopillars ranging from $100\ \text{nm}$ to $1000\ \text{nm}$. In particular, Figure 5.6 presents the $R(H)$ measurement for a $350\ \text{nm}$ device under a $4\ \text{mA}$ bias current, where a clear vortex configuration was confirmed.

After identifying the vortex state through the $R(H)$ curves, preliminary tests were carried out to probe the excitation of the vortex core. Typically, an external out-of-plane magnetic field is applied to stimulate the gyrotropic motion of the vortex. In this study, however, no external magnetic field was applied. Instead, the oscillation behavior can be attributed to a small intrinsic out-of-plane component, possibly originated from inhomogeneities in the magnetization of the fixed layer. This, combined with the Oersted field generated by the bias current, acted as the driving force. These contributions can tilt the magnetization sufficiently to displace the vortex core from its equilibrium position [77]. Nevertheless, no oscillations were detected in the $350\ \text{nm}$ device. In contrast, a $1000\ \text{nm}$ nanopillar exhibited pronounced zero-field oscillations. Figure 5.7 shows three $R(H)$ curves for this device under increasing bias currents of $4\ \text{mA}$ (a), $8\ \text{mA}$ (b), and $12\ \text{mA}$ (c). With increasing

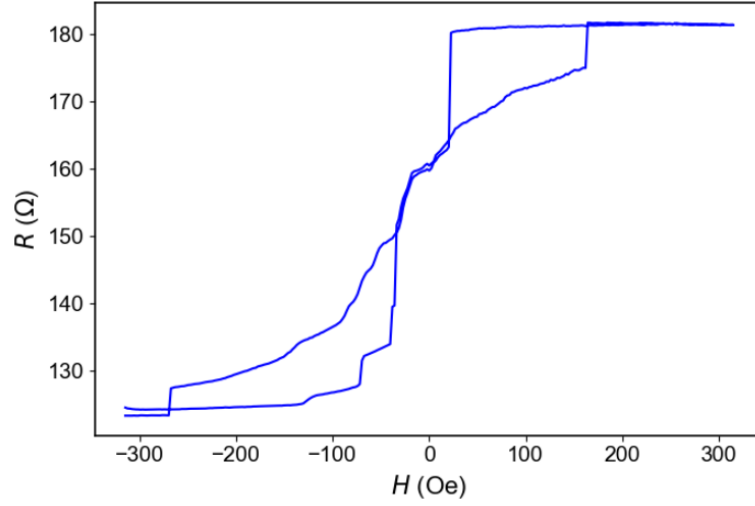


Figure 5.6: $R(H)$ for a vortex configuration with a diameter of 350 nm under a bias current of 4 mA.

current, the vortex state becomes more stable due to the Oersted field, which favors the curling configuration of the free layer. The direction of the current further sets the chirality of the vortex, since the Oersted field circulation coincides with the vortex circulation.

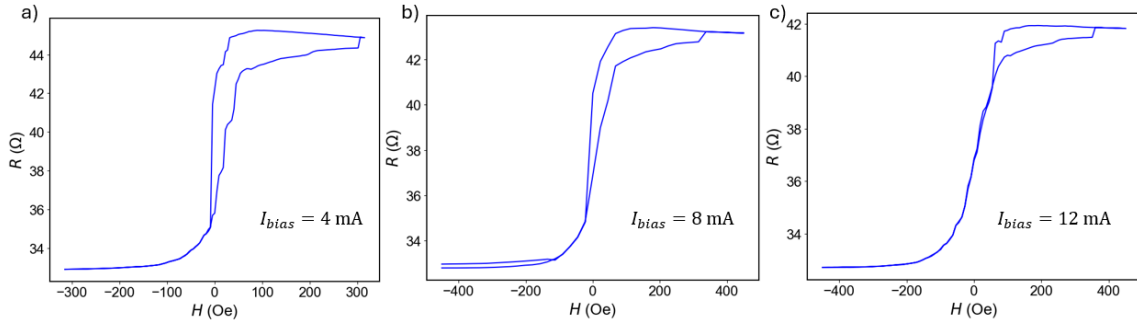


Figure 5.7: $R(H)$ for a vortex configuration with a diameter of 1000 nm under a bias current of a) 4 mA, b) 8 mA and c) 12 mA.

This allowed the acquisition of $V(f)$ spectra for different I_{bias} values. Each measured $V(f)$ spectrum was fitted with a Lorentzian function to extract the oscillation frequency f_0 , the linewidth (LW) Δf , and the peak power (V_0):

$$V(f) = V_0 \frac{\left(\frac{\Delta f}{2}\right)^2}{(f - f_0)^2 + \left(\frac{\Delta f}{2}\right)^2}. \quad (5.1)$$

The displayed power at each frequency bin of the spectrum analyzer is reported in dBm, a logarithmic unit defined as:

$$P_{dBm} = 10 \log_{10} \left(\frac{P}{1 \text{ mW}} \right), \quad (5.2)$$

which expresses power relative to 1 mW. The use of a logarithmic scale is convenient to represent very large or very small power values, but it also compresses large variations in power.

In the measurement, the reported value at each frequency corresponds to the total power integrated over the resolution bandwidth (RBW). The RBW defines the smallest frequency difference between two signals that the analyzer can distinguish separately. A smaller RBW provides higher frequency resolution, but at the cost of slower sweep speeds.

For quantitative analysis, the measured power values were converted from dBm to W according to:

$$P_W = 10^{\frac{P_{\text{dBm}} - 30}{10}} \quad (5.3)$$

and then normalized by the RBW to obtain the power spectral density (PSD), which represents how the total power is distributed across frequency, expressed in W/Hz (or more conveniently in $\mu\text{W Hz}^{-1}$).

Figure 5.8a and Figure 5.8b show Lorentzian fits for a spectrum measured at a bias current of 17 mA using a RBW of 43 kHz, plotted in $\mu\text{W Hz}^{-1}$ and dBm Hz^{-1} , respectively. In the first scale (Figure 5.8a), the resonance peak appears more pronounced because the exponential conversion from dBm to W preserves the full dynamic range. This enhances the visibility of the main peak while making the secondary peaks negligible due to their much smaller amplitudes. From the Lorentz fit, it is possible to extract the resonance frequency, linewidth, and the total emitted power of each spectrum for the different bias currents applied. Figure 5.9a presents the oscillation power as a function of the bias current over a range of [5, 21] mA. The integrated power exhibits a non-linear increase with current, and the oscillation threshold can be identified at approximately $I_{\text{th}} \approx 8$ mA. Above this threshold, the maximum measured power reaches about 0.06 μW . The oscillation frequency as a function of bias current above the threshold is shown in Figure 5.9b, with the linewidth displayed in the inset of this figure. The device exhibits a frequency that increases from about 100 to 130 MHz, and then remains relatively stable above 15 mA. This increase of the oscillation frequency as a function of the bias current is supported with the analytical model proposed in [77]. The linewidth varies between 1 and 2.5 MHz. In an ideal oscillator, the linewidth would approach zero, corresponding to a Dirac delta function in frequency space. However, in real STNOs, a finite linewidth always appears due to thermal fluctuations and phase noise, which limits the coherence of the magnetization precession.

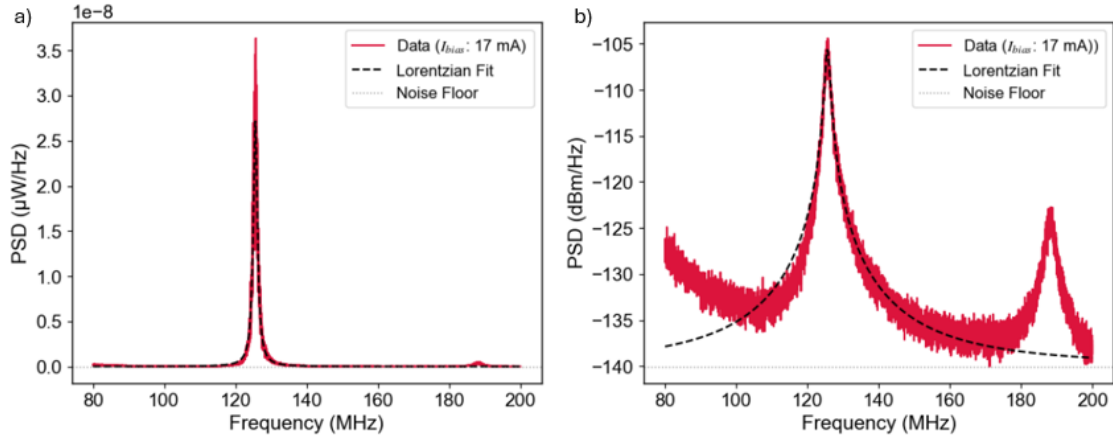


Figure 5.8: Power Spectral Density (PSD) in function of frequency represented $\mu\text{W Hz}^{-1}$ a) and dBm Hz^{-1} b) for a measured spectrum with 17 mA of bias current.

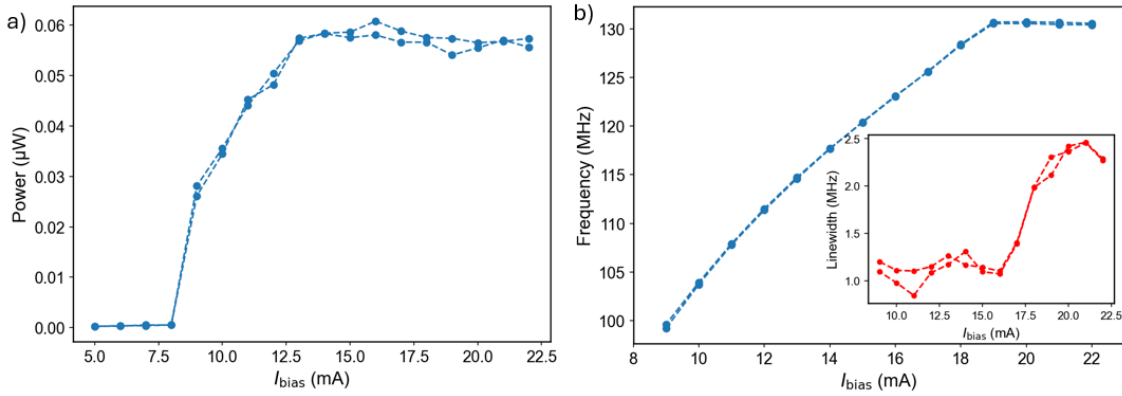


Figure 5.9: STNO characterization. a) Integrated power, b) resonance frequency and c) linewidth as a function of the bias current, obtained from a Lorentz fit.

These results demonstrate the successful generation of auto-oscillations in STNOs and reveal the nonlinear dependence of output power on the applied current bias. This behavior supports their use as artificial neurons through activation functions, achieved here without the need for an out-of-plane field, which is an important advantage for practical applications. However, time constraints limited further exploration in this study. Future work should focus on a more detailed investigation of the characteristics of these STNOs, together with optimization of the nanopillar geometry, in order to balance reliable MRAM functionality with efficient artificial neuron operation. In addition, exploring the oscillation properties of STNOs with the goal of increasing output power and reducing linewidth would be valuable for their use as CMOS-compatible oscillators in a broader range of applications. An important step forward will also be the integration of the two concepts of WSTNO explored in this chapter into a neural network, thereby demonstrating their combined potential for neuromorphic computing.

6. Conclusions and Future Work

This thesis has shown that MTJs play a valuable role in neuromorphic computation, where their non-linearity, non-volatility, stability, and energy efficiency can be leveraged across different applications.

In the first part of the thesis, the finite difference method is introduced for calculating the gradient of an I-V response that serves as an activation function. This approach demonstrated simplicity in implementation and yielded significant results when trained with various network architectures through simulations and experimental tests. Two tasks were used to validate the approach. The first, based on the iris dataset, involved experimental implementation and achieved an accuracy of 93.3%, which was corroborated by simulations reaching 94.2%. The second, based on the MNIST dataset, involved only simulated implementation, using two hidden layers with 128 and 64 neurons, respectively, and achieved 96.8%. An additional advantage of this approach lies in its inherent tolerance to device-to-device variability. As fabrication at the nanoscale makes it difficult to produce perfectly identical devices, the method only requires that the two devices forming a pair behave similarly. This property enhances robustness to manufacturing imperfections and natural fluctuations. Future research should extend this method to smaller devices, explore more complex functional responses, and investigate the possibility of treating the activation function itself as a trainable parameter. The tunability of spintronic devices and the study of more complex activation functions could allow activation responses to be tailored for specific tasks, reducing the number of devices needed in a neural network and thus lowering both area footprint and energy consumption.

The second part of the thesis explored MTJs configured to provide quantized synaptic weights through chains of devices, as well as the use of STNOs as nonlinear activation elements. A chain of five MTJs functioning as MRAMs produced 32 distinct resistance states, suitable for use as quantized weights. This configuration also demonstrated excellent stability, with states maintained for at least 17 h under continuous current application, underscoring its potential for reliable operation. Additionally, a brief study of STNOs revealed vortex-induced oscillations in a 1000 nm device once a threshold current of 8 mA was reached. While further investigation is needed to improve performance and reliability, these findings reinforce the potential of STNOs as artificial neurons due to their inherent nonlinearity. Looking ahead, an important direction is to integrate the two approaches.

This can first be explored in simulations using Quantization-Aware Training (QAT) with the Straight-Through Estimator (STE), allowing the network to adapt to quantized MRAM levels while retaining gradient-based optimization. Experimentally, the trained weights could then be directly programmed into MRAM devices. Remaining challenges include the need to represent both positive and negative weights, as MRAMs inherently provide only non-negative states. This can be addressed through rescaling operations in the trained weights before inference, or by using paired MRAM devices as demonstrated in Ref. [61], though this will double device count. Another critical issue is the current mismatch between MRAMs, which operate in the μA range, and STNOs, that operate in the mA-scale, likely necessitating amplifiers for practical integration.

In conclusion, this work contributes to the field of spintronic neuromorphic computing by exploring both MTJ-based activation functions and quantized synaptic weights. While further research is required to overcome practical challenges, the results presented here strengthen the case for spintronics-based analog neuromorphic systems as a promising and competitive platform for real-world machine learning applications.

References

- [1] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, “State-of-the-art in artificial neural network applications: A survey,” *Heliyon*, vol. 4, no. 11, 2018. [Cited on pages 1 and 25.]
- [2] M. D. Godfrey, “Introduction to “the first draft report on the edvac” by john von neumann,” *Annals of the History of Computing*, vol. 15, no. 1, pp. 11–21, 1993. [Cited on page 1.]
- [3] X. Zou, S. Xu, X. Chen, L. Yan, and Y. Han, “Breaking the von neumann bottleneck: architecture-level processing-in-memory technology,” *Science China Information Sciences*, vol. 64, no. 6, p. 160404, 2021. [Cited on page 1.]
- [4] C. Mead, “Neuromorphic electronic systems,” *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990. [Cited on page 1.]
- [5] G. Finocchio, J. A. C. Incorvia, J. S. Friedman, Q. Yang, A. Giordano, J. Grollier, H. Yang, F. Ciubotaru, A. V. Chumak, A. J. Naeemi *et al.*, “Roadmap for unconventional computing with nanotechnology,” *Nano Futures*, vol. 8, no. 1, p. 012001, 2024.
- [6] W. Haensch, T. Gokmen, and R. Puri, “The next generation of deep learning hardware: Analog computing,” *Proceedings of the IEEE*, vol. 107, no. 1, pp. 108–122, 2018. [Cited on page 1.]
- [7] J. Grollier, D. Querlioz, and M. D. Stiles, “Spintronic nanodevices for bioinspired computing,” *Proceedings of the IEEE*, vol. 104, no. 10, pp. 2024–2039, 2016. [Cited on pages 1 and 30.]
- [8] N. Locatelli, V. Cros, and J. Grollier, “Spintronic devices as key elements for energy-efficient neuroinspired architectures,” *Nature Materials*, vol. 13, pp. 11–20, 2014. [Cited on pages 1 and 30.]
- [9] D. Marković, A. Mizrahi, D. Querlioz, and J. Grollier, “Physics for neuromorphic computing,” *Nature Reviews Physics*, vol. 2, no. 9, pp. 499–510, 2020. [Cited on page 1.]
- [10] K. Roy, S. Bandyopadhyay, and J. Atulasimha, “Brain-inspired computing with spin torque devices,” *Nature Electronics*, vol. 2, no. 1, pp. 2–6, 2019. [Cited on page 1.]

- [11] T. Wu, Y. Li, L. Ge, and L. Feng, “Field-programmable photonic nonlinearity,” *Nature Photonics*, pp. 1–8, 2025. [Cited on page 1.]
- [12] S. Schmitt, J. Klähn, G. Bellec, A. Grübl, M. Guettler, A. Hartel, S. Hartmann, D. Husmann, K. Husmann, S. Jeltsch *et al.*, “Neuromorphic hardware in the loop: Training a deep spiking network on the brainscales wafer-scale system,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 2227–2234. [Cited on page 43.]
- [13] B. Cramer, S. Billaudelle, S. Kanya, A. Leibfried, A. Grübl, V. Karasenko, C. Pehle, K. Schreiber, Y. Stradmann, J. Weis *et al.*, “Surrogate gradients for analog neuromorphic computing,” *Proceedings of the National Academy of Sciences*, vol. 119, no. 4, p. e2109194119, 2022.
- [14] A. Ross, N. Leroux, A. De Riz, D. Marković, D. Sanz-Hernández, J. Trastoy, P. Borlototti, D. Querlioz, L. Martins, L. Benetti *et al.*, “Multilayer spintronic neural networks with radiofrequency connections,” *Nature Nanotechnology*, vol. 18, no. 11, pp. 1273–1280, 2023. [Cited on page 1.]
- [15] L. Esaki, “Long journey into tunneling,” *Science*, vol. 183, no. 4130, pp. 1149–1155, 1974. [Cited on page 5.]
- [16] J.-G. J. Zhu and C. Park, “Magnetic tunnel junctions,” *Materials today*, vol. 9, no. 11, pp. 36–45, 2006. [Cited on pages xiii, 5, and 6.]
- [17] J. G. Simmons, “Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film,” *Journal of applied physics*, vol. 34, no. 6, pp. 1793–1803, 1963. [Cited on page 6.]
- [18] W. Brinkman, R. Dynes, and J. Rowell, “Tunneling conductance of asymmetrical barriers,” *Journal of applied physics*, vol. 41, no. 5, pp. 1915–1921, 1970. [Cited on page 7.]
- [19] M. Jullière, “Tunneling between ferromagnetic films,” *Physics Letters A*, vol. 54, no. 3, pp. 225–226, 1975. [Cited on pages 7, 8, and 9.]
- [20] S. Maekawa, *Concepts in spin electronics*. Oxford University Press, 2006, vol. 13. [Cited on page 7.]
- [21] M. SÉRGIO, “Optimization of cofeb-alox thin films for integration in advanced magnetoresistive devices,” 2014. [Cited on pages xiii and 9.]

- [22] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions," *Phys. Rev. Lett.*, vol. 74, no. 16, pp. 3273–3276, 1995. [Cited on page 9.]
- [23] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, "Giant room-temperature magnetoresistance in single-crystal fe/mgo/fe magnetic tunnel junctions," *Nature Materials*, vol. 3, pp. 868–871, 2004. [Cited on pages 9 and 10.]
- [24] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, "Spin-dependent tunneling conductance of fe|mgo|fe sandwiches," *Phys. Rev. B*, vol. 63, no. 5, p. 054416, 2001. [Cited on page 9.]
- [25] J. Mathon and A. Umerski, "Theory of tunneling magnetoresistance of an epitaxial fe/mgo/fe junction," *Phys. Rev. B*, vol. 63, no. 22, p. 220403(R), 2001. [Cited on page 9.]
- [26] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, "Giant tunnelling magnetoresistance at room temperature with mgo (100) tunnel barriers," *Nature Materials*, vol. 3, pp. 862–867, 2004. [Cited on page 10.]
- [27] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300 k by suppression of ta diffusion in cofeb/mgo/cofeb pseudo-spin-valves annealed at high temperature," *Applied Physics Letters*, vol. 93, no. 8, p. 082508, 2008. [Cited on page 10.]
- [28] C. Park, J.-G. Zhu, M. T. Moneck, Y. Peng, and D. E. Laughlin, "Annealing effects on structural and transport properties of rf-sputtered cofeb/ mgo/ cofeb magnetic tunnel junctions," *Journal of applied physics*, vol. 99, no. 8, 2006. [Cited on pages xiii and 10.]
- [29] J. M. Coey, *Magnetism and magnetic materials*. Cambridge university press, 2010. [Cited on pages xiii, 11, 22, and 23.]
- [30] P. Dey and J. N. Roy, *Spintronics*. Springer, 2021. [Cited on page 12.]
- [31] M. A. Ruderman and C. Kittel, "Indirect exchange coupling of nuclear magnetic moments by conduction electrons," *Physical Review*, vol. 96, no. 1, p. 99, 1954. [Cited on page 15.]

- [32] W. H. Meiklejohn and C. P. Bean, “New magnetic anisotropy,” *Physical Review*, vol. 105, no. 3, p. 904, 1957. [Cited on pages xiii, 15, and 16.]
- [33] J. Valadeiro, “Magnetoresistive sensors with pico-tesla sensitivities,” Master’s thesis, Instituto Superior Técnico, Sep. 2014. [Cited on pages xiii and 18.]
- [34] S. Ikeda, J. Hayakawa, Y. Ashizawa *et al.*, “Tunnel magnetoresistance of 604% at 300 k by suppression of ta diffusion in cofeb/mgo/cofeb pseudo-spin-valves annealed at high temperature,” *Applied Physics Letters*, vol. 93, no. 8, p. 082508, 2008. [Cited on page 17.]
- [35] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, “Giant room-temperature magnetoresistance in single-crystal fe/mgo/fe magnetic tunnel junctions,” *Nature Materials*, vol. 3, no. 12, pp. 868–871, 2004. [Cited on page 17.]
- [36] S. S. P. Parkin, C. Kaiser, A. Panchula *et al.*, “Giant tunnelling magnetoresistance at room temperature with mgo (100) tunnel barriers,” *Nature Materials*, vol. 3, no. 12, pp. 862–867, 2004. [Cited on page 17.]
- [37] H. Sato, M. Yamanouchi, S. Ikeda *et al.*, “Properties of mgo/cofeb/ta and mgo/cofeb/mgo magnetic tunnel junctions with perpendicular magnetic anisotropy,” *Applied Physics Letters*, vol. 101, no. 2, p. 022414, 2012. [Cited on page 18.]
- [38] J. Nogués and I. K. Schuller, “Exchange bias,” *Journal of Magnetism and Magnetic Materials*, vol. 192, pp. 203–232, 1999. [Cited on page 18.]
- [39] I. Prejbeanu, M. Kerekes, R. C. Sousa, H. Sibuet, O. Redon, B. Dieny, and J. Nozières, “Thermally assisted mram,” *Journal of Physics: Condensed Matter*, vol. 19, no. 16, p. 165218, 2007. [Cited on page 19.]
- [40] L. Lopez-Diaz, L. Torres, and E. Moro, “Transition from ferromagnetism to superparamagnetism on the nanosecond time scale,” *Physical Review B*, vol. 65, no. 22, p. 224406, 2002. [Cited on page 20.]
- [41] J. M. Kosterlitz and D. J. Thouless, “Ordering, metastability and phase transitions in two-dimensional systems,” *Journal of Physics C: Solid State Physics*, vol. 6, no. 7, p. 1181, 1973. [Cited on page 20.]
- [42] G. Hrkac, P. S. Keatley, M. T. Bryan, and K. Butler, “Magnetic vortex oscillators,” *Journal of Physics D: Applied Physics*, vol. 48, no. 45, p. 453001, 2015. [Cited on pages xiii and 20.]

- [43] K. Y. Guslienko, “Magnetic vortex state stability, reversal and dynamics in restricted geometries,” *Journal of nanoscience and nanotechnology*, vol. 8, no. 6, pp. 2745–2760, 2008. [Cited on pages xiii and 21.]
- [44] J. C. Slonczewski, “Current-driven excitation of magnetic multilayers,” *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1-2, pp. L1–L7, 1996. [Cited on page 21.]
- [45] A. Thiele, “Steady-state motion of magnetic domains,” *Physical Review Letters*, vol. 30, no. 6, p. 230, 1973. [Cited on page 24.]
- [46] A. Dussaux, B. Georges, J. Grollier, V. Cros, A. Khvalkovskiy, A. Fukushima, M. Konoto, H. Kubota, K. Yakushiji, S. Yuasa *et al.*, “Large microwave generation from current-driven magnetic vortex oscillators in magnetic tunnel junctions,” *Nature communications*, vol. 1, no. 1, p. 8, 2010. [Cited on pages xiii and 24.]
- [47] S. Tsunegi, K. Yakushiji, A. Fukushima, S. Yuasa, and H. Kubota, “Microwave emission power exceeding 10 μ W in spin torque vortex oscillator,” *Applied Physics Letters*, vol. 109, no. 25, 2016. [Cited on page 24.]
- [48] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, “Deep learning in spiking neural networks,” *Neural networks*, vol. 111, pp. 47–63, 2019. [Cited on page 25.]
- [49] R. Rojas, “The backpropagation algorithm,” in *Neural networks: a systematic introduction*. Springer, 1996, pp. 149–182. [Cited on page 25.]
- [50] E. Akgün and M. Demir, “Modeling course achievements of elementary education teacher candidates with artificial neural networks,” *International Journal of Assessment Tools in Education*, vol. 5, no. 3, pp. 491–509, 2018. [Cited on pages xiii and 25.]
- [51] C.-H. Tsai, Y.-T. Chih, W. H. Wong, and C.-Y. Lee, “A hardware-efficient sigmoid function with adjustable precision for a neural network system,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 11, pp. 1073–1077, 2015. [Cited on page 27.]
- [52] D. Baptista and F. Morgado-Dias, “Low-resource hardware implementation of the hyperbolic tangent for artificial neural networks,” *Neural Computing and Applications*, vol. 23, no. 3, pp. 601–607, 2013. [Cited on page 27.]

- [53] S. Moradi and R. Manohar, “The impact of on-chip communication on memory technologies for neuromorphic systems,” *Journal of Physics D: Applied Physics*, vol. 52, no. 1, p. 014003, 2018. [Cited on page 30.]
- [54] S. Deb, T. Vatsani, A. Chattopadhyay, A. Basu, and X. Fong, “Domain wall motion-based xor-like activation unit with a programmable threshold,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8. [Cited on page 31.]
- [55] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima *et al.*, “Neuromorphic computing with nanoscale spintronic oscillators,” *Nature*, vol. 547, no. 7664, pp. 428–431, 2017. [Cited on page 31.]
- [56] N. Leroux, A. Mizrahi, D. Marković, D. Sanz-Hernández, J. Trastoy, P. Bortolotti, L. Martins, A. Jenkins, R. Ferreira, and J. Grollier, “Hardware realization of the multiply and accumulate operation on radio-frequency signals with magnetic tunnel junctions,” *Neuromorphic Computing and Engineering*, vol. 1, no. 1, p. 011001, 2021. [Cited on page 31.]
- [57] H. Tu, Y. Luo, K. Zeng, Y. Wu, L. Zhang, B. Zhang, and Z. Zeng, “Spin torque oscillator based on magnetic tunnel junction with mgo cap layer for radio-frequency-oriented neuromorphic computing,” *Chinese Physics B*, vol. 32, no. 10, p. 107504, 2023. [Cited on pages xiv and 31.]
- [58] A. Kumar, D. J. Lin, D. Das, L. Huang, S. L. Yap, H. R. Tan, H. K. Tan, R. J. Lim, Y. T. Toh, S. Chen *et al.*, “Multistate compound magnetic tunnel junction synapses for digital recognition,” *ACS Applied Materials & Interfaces*, vol. 16, no. 8, pp. 10 335–10 343, 2024. [Cited on page 32.]
- [59] K. Zhang, X. Jia, K. Cao, J. Wang, Y. Zhang, K. Lin, L. Chen, X. Feng, Z. Zheng, Z. Zhang *et al.*, “High on/off ratio spintronic multi-level memory unit for deep neural network,” *Advanced Science*, vol. 9, no. 13, p. 2103357, 2022.
- [60] Y. Kim, X. Fong, K.-W. Kwon, M.-C. Chen, and K. Roy, “Multilevel spin-orbit torque mrams,” *IEEE Transactions on Electron Devices*, vol. 62, no. 2, pp. 561–568, 2014. [Cited on page 32.]
- [61] P. Rzeszut, J. Chłociński, I. Brzozowski, S. Ziutek, W. Skowroński, and T. Stobiecki, “Multi-state mram cells for hardware neuromorphic computing,” *Scientific reports*, vol. 12, no. 1, p. 7178, 2022. [Cited on pages xiv, 32, 33, and 78.]

- [62] T. Böhnert, Y. Rezaeiyan, M. Claro, L. Benetti, A. Jenkins, H. Farkhani, F. Moradi, and R. Ferreira, “Weighted spin torque nano-oscillator system for neuromorphic computing,” *Communications Engineering*, vol. 2, no. 1, p. 65, 2023. [Cited on pages xiv, 33, and 34.]
- [63] P. J. Kelly and R. D. Arnell, “Magnetron sputtering: a review of recent developments and applications,” *Vacuum*, vol. 56, no. 3, pp. 159–172, 2000. [Cited on page 35.]
- [64] R. Ferreira, “Ion beam deposited magnetic spin tunnel junctions targeting hdd read heads, non-volatile memories and magnetic field sensor applications,” *Universidade Técnica de Lisboa*, 2008. [Cited on pages xiv and 36.]
- [65] L. J. F. Martins, “Cofeb/mgo/cofeb-based magnetic tunnel junctions for memory applications: optimization of the fabrication process and magnetic properties,” Master’s thesis, Universidade do Porto (Portugal), 2016. [Cited on pages xiv, 36, and 37.]
- [66] J. D. T. B. Costa, “Spin-transfer-torque nano-oscillators: Fabrication, characterization and dynamics,” Ph.D. dissertation, Universidade do Porto (Portugal), 2017. [Cited on pages xiv and 39.]
- [67] J. C. Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation,” *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332–341, 2002. [Cited on page 43.]
- [68] P. Ramuhalli, L. Udpa, and S. S. Udpa, “Finite-element neural networks for solving differential equations,” *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1381–1392, 2005. [Cited on page 43.]
- [69] L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, and P. L. McMahon, “Deep physical neural networks trained with backpropagation,” *Nature*, vol. 601, no. 7894, pp. 549–555, 2022. [Cited on page 43.]
- [70] D. Dimitrov, Z. Gao, X. Wang, W. Jung, X. Lou, and O. G. Heinonen, “Dielectric breakdown of mgo magnetic tunnel junctions,” *Applied Physics Letters*, vol. 94, no. 12, 2009. [Cited on page 46.]
- [71] K.-S. Kim, K.-S. Lee, C. Nam, and B. K. Cho, “Effect of nitrogen incorporation to oxidation process on the reliability of magnetic tunnel junctions,” *IEEE transactions on magnetics*, vol. 42, no. 1, pp. 2–4, 2006. [Cited on page 49.]

- [72] N. Dalili, Q. Liu, and D. G. Ivey, “Thermal and electrical stability of tan x diffusion barriers for cu metallization,” *Journal of materials science*, vol. 48, no. 1, pp. 489–501, 2013. [Cited on page 51.]
- [73] S. Tsukimoto, M. Moriyama, and M. Murakami, “Microstructure of amorphous tantalum nitride thin films,” *Thin Solid Films*, vol. 460, no. 1-2, pp. 222–226, 2004. [Cited on page 51.]
- [74] M. G. Duman, S. Koparal, N. Ömür, A. Ertürk, and E. Aptoula, “Adlu: Adaptive double parametric activation functions,” *Digital Signal Processing*, p. 105579, 2025. [Cited on page 53.]
- [75] T. Wu, Y. Li, L. Ge, and L. Feng, “Field-programmable photonic nonlinearity,” *Nature Photonics*, pp. 1–8, 2025. [Cited on page 61.]
- [76] M. Heigl, C. Vogler, A.-O. Mandru, X. Zhao, H. J. Hug, D. Suess, and M. Albrecht, “Microscopic origin of magnetization reversal in nanoscale exchange-coupled ferri-/ferromagnetic bilayers: implications for high energy density permanent magnets and spintronic devices,” *ACS applied nano materials*, vol. 3, no. 9, pp. 9218–9225, 2020. [Cited on page 69.]
- [77] A. Dussaux, A. Khvalkovskiy, P. Bortolotti, J. Grollier, V. Cros, and A. Fert, “Field dependence of spin-transfer-induced vortex dynamics in the nonlinear regime,” *Physical Review B—Condensed Matter and Materials Physics*, vol. 86, no. 1, p. 014402, 2012. [Cited on pages 73 and 75.]

Attachments

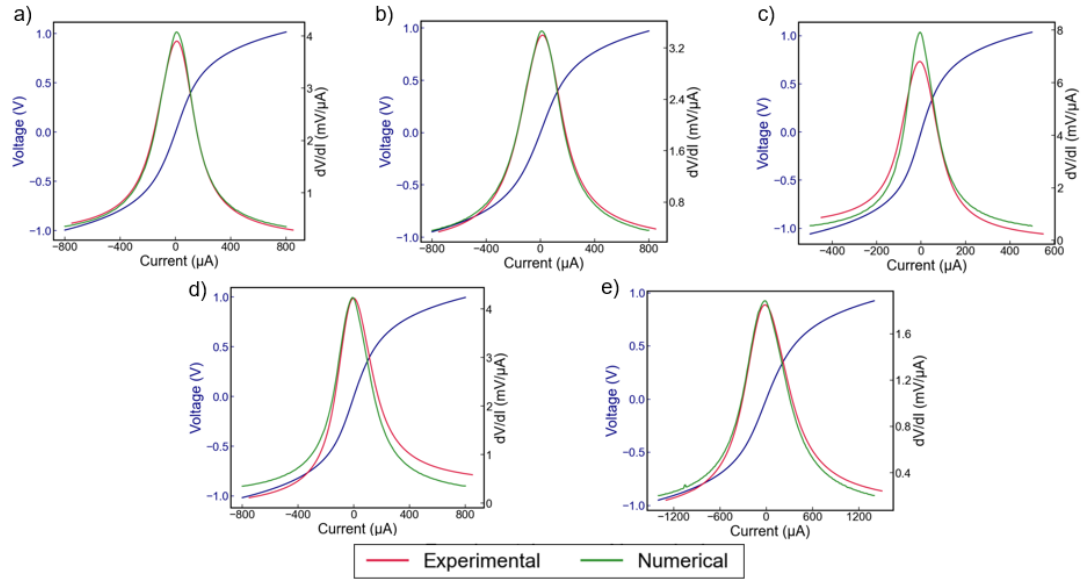


Figure 1: I-V curves and corresponding gradients for S_A used in the neural network implementation, with (a-e) representing P1 to P5.

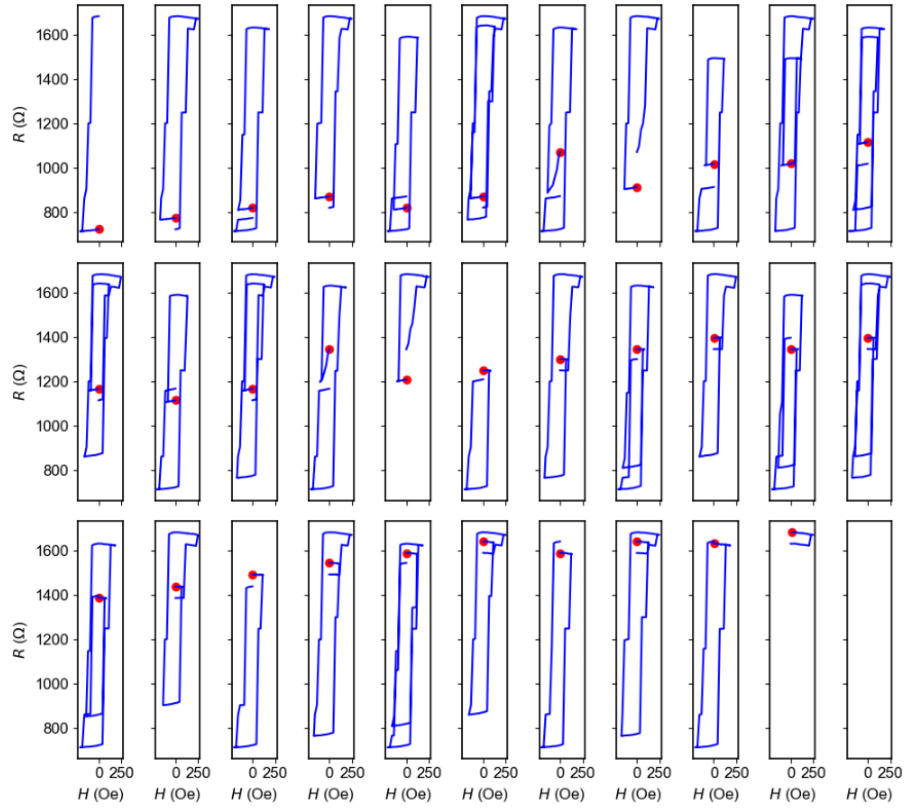


Figure 2: Magnetic field minor loops to set the chain of 5 MTJs in the 32 resistance states.

Table 1: Minor-loop sequences and resulting states.

Minor loops Field Order	Final state (Binary)	Final R (Ω)
0 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 0.63	00000	721.3
1 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 0.63	00001	772.5
2 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ -157.7 $\rightarrow$ 0.63	00010	819.3
3 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -157.7 $\rightarrow$ 0.63	00011	869.6
4 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 131.3 $\rightarrow$ -136.3 $\rightarrow$ 0.63	00100	819.5
5 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 131.3 $\rightarrow$ -136.3 $\rightarrow$ 0.63	00101	871.1
6 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ -136.3 $\rightarrow$ 0.63	00110	1069.6
7 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -136.3 $\rightarrow$ 0.63	00111	912.6
8 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 118.7 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01000	1017.5
9 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 118.7 $\rightarrow$ 101.16 $\rightarrow$ 0.63	01001	1017.7
10 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ -157.7 $\rightarrow$ -118.73 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01010	1115.0
11 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -157.7 $\rightarrow$ 118.7 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01011	1166.1
12 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 131.3 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01100	1114.8
13 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 131.3 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01101	1166.3
14 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01110	1344.5
15 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -101.2 $\rightarrow$ 0.63	01111	1208.3
16 : 0.63 $\rightarrow$ -204.19 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10000	1249.1
17 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10001	1300.7
18 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.84 $\rightarrow$ -157.7 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10010	1346.0
19 : 0.63 $\rightarrow$ 26.9 $\rightarrow$ -157.7 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10011	1397.3
20 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 131.3 $\rightarrow$ -136.3 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10100	1346.4
21 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 131.3 $\rightarrow$ -136.3 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10101	1397.4
22 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ -136.3 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10110	1388.5
23 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -136.3 $\rightarrow$ 83.6 $\rightarrow$ 0.63	10111	1438.6
24 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 118.7 $\rightarrow$ 0.63	11000	1492.8
25 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 118.7 $\rightarrow$ 0.63	11001	1545.7
26 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ -157.7 $\rightarrow$ 118.7 $\rightarrow$ 0.63	11010	1590.5
27 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -157.7 $\rightarrow$ 118.7 $\rightarrow$ 0.63	11011	1642.7
28 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 131.3 $\rightarrow$ 0.63	11100	1590.7
29 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ -169.0 $\rightarrow$ 131.3 $\rightarrow$ 0.63	11101	1642.2
30 : 0.63 $\rightarrow$ -204.2 $\rightarrow$ 182.8 $\rightarrow$ 0.63	11110	1633.9
31 : 0.63 $\rightarrow$ 246.9 $\rightarrow$ 16.1	11111	1684.5