



HAL
open science

Plasticity in MTJ-based synapses

Salah Daddinounou, Elena Ioana Vatajelu, Giorgio Di Natale

► **To cite this version:**

Salah Daddinounou, Elena Ioana Vatajelu, Giorgio Di Natale. Plasticity in MTJ-based synapses. TIMA Scientific Days 2021: RISC-V in Research and Education, TIMA / LCIS / VERIMAG Laboratories, Jul 2021, Grenoble (FR), France. hal-04358393

HAL Id: hal-04358393

<https://hal.science/hal-04358393>

Submitted on 28 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Plasticity in MTJ-based synapses

Salah DADDINOUNOU
TIMA laboratory
Université Grenoble Alpes
Grenoble, France

salah.daddinounou@univ-grenoble-alpes.fr

Elena Ioana VATAJELU
CNRS, TIMA laboratory
Université Grenoble Alpes
Grenoble, France

ioana.vatajelu@univ-grenoble-alpes.fr

Giorgio Di Natale
CNRS, TIMA laboratory
Université Grenoble Alpes
Grenoble, France

giorgio.di-natale@univ-grenoble-alpes.fr

Abstract—Spiking Neural Networks (SNNs) are being widely studied to mimic the intelligence of biological brains. In hardware level, multiple emerging non-volatile memories (NVM) are proposed to model the neural synapses, one of them being Spin-transfer Torque Magnetic Random-Access Memory (STT-MRAM). This device can be used as a building block to build multilevel conducting synapses. In this paper we describe the simulation results of a compound magnetic synapse in nominal situation, and then present the effect of process variability on its performance.

Keywords—emerging NVM, reliability, STT-MRAM, spiking neural networks, synapse.

I. CONTEXT

In the era of big data and internet of things, the search for more energy efficient computing paradigms than classic Von Neuman architecture is in continuous growth. The use of neural networks for image recognition is known for decades now in the field of artificial intelligence (AI). To take the inspiration from nature one step further, spiking neural networks were proposed simultaneously with the rise of new devices, which will ultimately allow the emulation of the most powerful and yet energy efficient computing system: The Human brain.

STT-MRAM uses a magnetic tunnel junction (MTJ) which consists of a stack of magnetic layers separated by a thin insulating layer. This device having two stable resistances is used as a NV memory with high resistance (logic 1) and low resistance (logic 0) [1]. The desired state can be written on the MTJ by controlling the direction and the magnitude of the current through it, whereas the reading operation consists of probing the resistance of the MTJ with a one-direction low current.

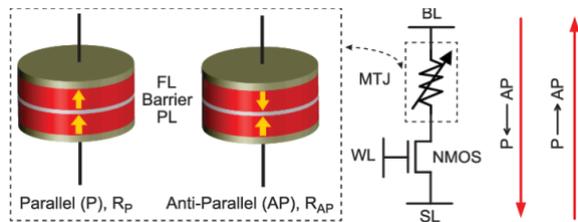


Fig. 1. The MTJ is a non-volatile memory, one bit is stored as a resistance. when the two magnetic layers are parallel, this is a low resistance and vice versa. The magnetization of one layer is pinned (PL) and the other (free layer FL) is free to flip when enough polarized spin current makes it do so. Resulting in parallel and anti-parallel states (low and high resistances respectively). from [1].

The MTJ device has an intrinsic stochasticity due to the sensitivity of magnetization to thermal fluctuations, this property which is not desired for memories is rather an advantage when using the MTJ to build synapses.

In previous work we looked to the use of MTJ as a memory cell, we investigated the range of voltages and pulse widths to achieve a successful writing attempt. Moreover, resistive defects which occur during fabrication were looked into as well.

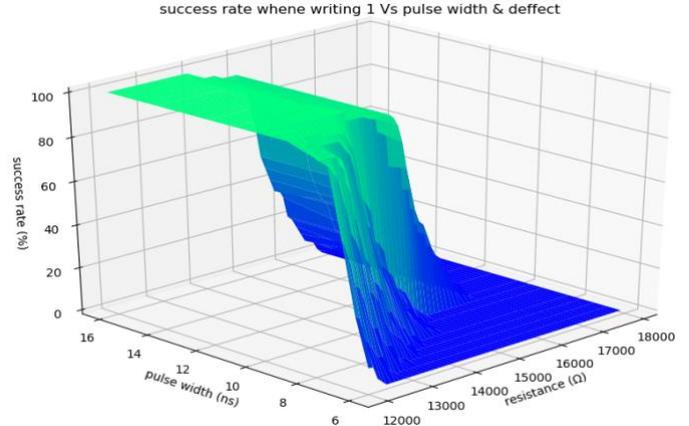


Fig. 2. in this figure, probability of writing '1' is given as function of the pulse duration and the resistive defects implemented in series with the MTJ, higher writing probability is achieved with higher voltages and within smaller defects conditions.

II. MOTIVATIONS

multiple research papers showed the feasibility of implementing spiking neural networks for image recognition to mimic the visual cortex of the brain. The most suitable algorithm for hardware implemented SNN is spike timing dependent plasticity (STDP). The advantage of this neural implementation and algorithm over other artificial neural networks is the very low energy consumption, as low as 20W in case of the human brain.

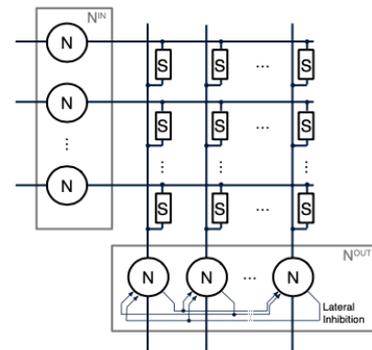


Fig. 3. The targeted neural network takes the shape of a crossbar array, with synapses between input and output neurons in an all-to-all connection fashion. from [2].

In this paper we only focus on the synapse study. To map the biological neural plasticity, four MTJs are put in parallel to stand as a single synapse. This allows a multilevel (5levels) conductance of the of the synapse. As shown in figure3, every input neuron is connected to all the output neurons, but not all connections are equally important, indeed some connections should be more relevant than others. This relative connection importance called the weight is determined during learning, once the right connections are set, the weights are fixed, and the network is ready to be used for inference.

the STDP algorithm used for learning is quite simple, the relevance of the connection between two neurons is determined by the time delay between the activations of these respective neurons. If an output neuron spikes shortly after an input neuron, this means that these two ones are strongly connected, therefore their connection is increased (Potentiation). Any inverting in the chronological spiking order results in penalizing the connection by decreasing its weight (depression).

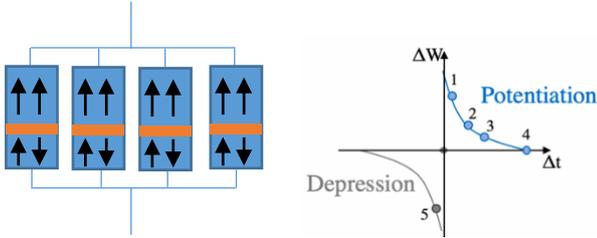


Fig. 4. left: the 4-MTJ synapse resulting in 5 levels of conductance. Right: The weight dynamics, spiking causality (input before output) between two neurons is rewarded by increasing their weight, depression takes place when this causality is not respected.

III. RESULTS

Our simulations are conducted in Cadence® Virtuoso® System Design Platform. We use an MTJ model [8] which allows multiple functionalities, stochasticity parameter to start with, is important to make the MTJs of one synapse slightly different from each other in order to switch state at different times, leading to the multi-step conductance profile of the synapse.

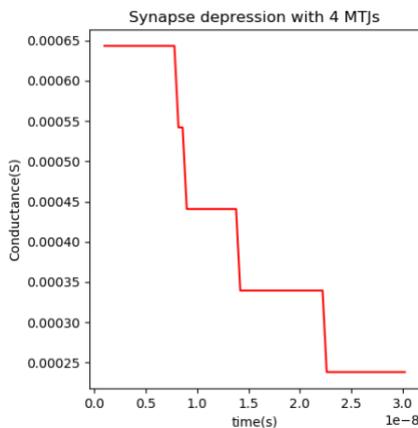


Fig. 5. depression: weight decrease of a synapse, four MTJ synapse lead to discrete 5 levels of conductance.

The figure5 shows the depression of one synapse when a voltage pulse of 210 mV during 30ns is applied. Using more MTJs for one synapse results in smoother continuity of the conductance which will have more levels, the choice of the number of MTJ will depend on the complexity of the features that the network wants to learn. The spiking time of the neuron should not be long enough to make all the MTJs in one synapse change state, we rather want it to be probabilistic by setting the pulse width on the example above at 15ns for example.

For more realistic simulations, process variability should be considered. We varied the thickness of the oxide layer separating the two magnetic layers inside the MTJ, the thickness of the free magnetic layer, and the TMR ratio. We ran a Monte Carlo simulation and varied these parameters according to a gaussian distribution.

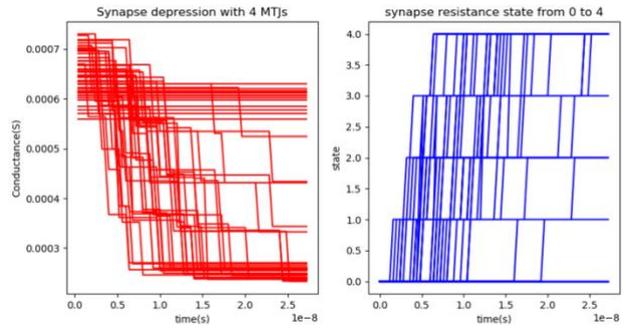


Fig. 6. Monte Carlo simulation with 50 iterations. Left: the conductance of the synapse depression with process variability. Right: at the same conditions, number of MTJs which have switched state from 0 to higher state.

Figure6 shows the effect of process variability on the synapse conductance, not only the switching time is affected but also conductance value, nevertheless we notice that the conductances are grouped in 5 distinct ranges, which correspond to the number of the MTJs that switched state at a certain time. On the left are shown these distinct groups.

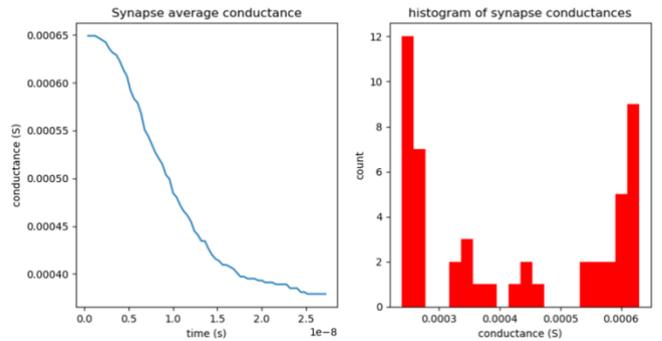


Fig. 7. left: averaging the conductance over the 50 MC simulations. Right: grouping the conductances in a histogram.

Figure7 shows the synapse average conductance at any time, this gives an insight on the continuous depression function. We retrieve the theoretical exponential dependence of the resistance as a function of time. The histogram plot on the right groups the conductances at $t=15$ ns into four major classes, some states are more preferred than others and this depends on the voltage and time. These results should be considered when initializing the weights for learning.

IV. CONCLUSION

We presented in this paper the possibility of using the magnetic tunnel junction devices as a synapse with a good plasticity. The intrinsic stochastic behavior of the MTJ is the key for such plasticity when multiple MTJs are put in parallel, the equivalent conductance maps for the neurotransmitters flow between neurons in biological systems. We restricted our study on depression mechanism which occurs when lowering the connection importance between two neurons during learning. we conducted Monte Carlo simulations to investigate the effect of process variability, the result will allow us to better choose the spiking signals of neurons in terms of voltage and pulse width. Next work will first focus on potentiation characterization, and then implanting the neurons with MTJ as well, finally implementing the STDP rules and build a proof-of-concept network.

REFERENCES

- [1] Wang Kang-Liuyang Zhang-Jacques-Olivier Klein-Youguang Zhang-Dafine Ravelosona-Weisheng Zhao -“**Reconfigurable Codesign of STT-MRAM Under Process Variations in Deeply Scaled Technology**” IEEE Transactions on Electron Devices - 2015
- [2] Elena-Ioana Vatajelu-Giorgio Natale-Lorena Anghel – “**Special Session: Reliability of Hardware-Implemented Spiking Neural Networks (SNN)**” 2019 IEEE 37th VLSI Test Symposium (VTS) - 2019
- [3] Chun, K. C., Zhao, H., Harms, J. D., Kim, T.-H., Wang, J.-P., & Kim, C. H. (2013). ‘**A Scaling Roadmap and Performance Evaluation of In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory**’. *IEEE Journal of Solid-State Circuits*, 48(2), 598–610. <https://doi.org/10.1109/jssc.2012.2224256>
- [4] Patrick Girard, Yuanqing Cheng, Arnaud Virazel, Weisheng Zhao, Rajendra Bishnoi, et al. ‘**A Survey of Test and Reliability Solutions for Magnetic Random Access Memories**’. Proceedings of the IEEE, Institute of Electrical and Electronics Engineers, In press, pp.1-21. 10.1109/JPROC.2020.3029600. lirmm-03031646
- [5] Nowak, J. J., Robertazzi, R. P., Sun, J. Z., Hu, G., Abraham, D. W., Trouilloud, P. L., Brown, S., Gaidis, M. C., O’Sullivan, E. J., Gallagher, W. J., & Worledge, D. C. (2011). ‘**Demonstration of ultralow bit error rates for spin-torque magnetic random-access memory with perpendicular magnetic anisotropy**’. *IEEE Magnetics Letters*, 2, 3000204. <https://doi.org/10.1109/lmag.2011.2155625>
- [6] Li, J., Ndai, P., Goel, A., Salahuddin, S., & Roy, K. (2010). **Design Paradigm for Robust Spin-Torque Transfer Magnetic RAM (STT MRAM) From Circuit/Architecture Perspective**. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 18(12), 1710–1723. <https://doi.org/10.1109/tvlsi.2009.2027907>
- [7] J.W. Kwak, A. Marshall, H. Stiegler, ‘**28nm STT-MRAM Array and Sense Amplifier**’ 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST), 978-1-7281-1184-1/19 IEEE2019
- [8] Y. Wang, Y. Zhang, W.S. Zhao, L. A. B. Naviner, H. Cai and J-O. Klein, ‘Manual of compact models for Perpendicular Magnetic Anisotropy (PMA) Magnetic Tunnel Junction (MTJ)