Microelectronics Reliability 135 (2022) 114570

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel

Reinforcement learning to reduce failures in SOT-MRAM switching

Johannes Ender [a,b,*], Roberto Lacerda de Orio [b], Simone Fiorentini [a,b], Siegfried Selberherr [b], Wolfgang Goes [c], Viktor Sverdlov [a,b]

[a] Christian Doppler Laboratory for Nonvolatile Magnetoresistive Memory and Logic, TU Wien, Gußhausstraße 27-29/E360, 1040 Vienna, Austria
[b] Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, 1040 Vienna, Austria
[c] Silvaco Europe Ltd., Cambridge, United Kingdom

ARTICLE INFO

Keywords:
Reinforcement learning
Spin-orbit torque memory
Magnetic field-free
Switching reliability

ABSTRACT

A reinforcement learning strategy is applied to find a reliable switching scheme for deterministic switching of a perpendicularly magnetized spin-orbit torque magnetoresistive memory cell. Current pulses sent along orthogonal metal wires allow the field-free reversal of the magnetization. The current pulses are optimized such that reliable switching can be achieved over a wide range of material parameters. Micromagnetic simulations confirm the reliability of the presented approach.

1. Introduction

The continuous downscaling of CMOS devices has reached a stage at which leakage currents become increasingly problematic, as they lead to an increase in standby power consumption. Spin-orbit torque magnetoresistive random access memory (SOT-MRAM), being nonvolatile, as well as possessing high endurance and fast operation speed, is a promising candidate to replace CMOS-based devices in high-level caches [1].

The downscaling of semiconductor devices has at the same time led to an increase in available computational power, which, among other fields, has enabled scientific simulations to generate ever more and better data. The adoption of machine learning (ML) has allowed not only to handle this large amount of information but has also helped in making scientific advancements. Reinforcement learning [2], a sub-branch of ML, most well-known for its application to games like chess or Go [3], was already successfully applied in scientific research, e.g. [4].

This work is a follow-up of [5] and an extension to the results previously published in [6]. It presents the application of RL to SOT-MRAM switching in order to achieve more reliable switching, even under the unavoidable variability of material properties due to the imperfection of fabrication processes.

2. Spin-orbit torque memory

The core element of MRAM devices is the magnetic tunnel junction (MTJ), which is formed by two ferromagnetic layers sandwiching a nonmagnetic tunnel barrier. The orientation of the magnetization is free to move only in one of the ferromagnetic layers, enabling a parallel and anti-parallel configuration of the two layers. The difference in electrical resistance between these two states can be sensed by an electric current flowing through the structure, and represents the binary information stored in the memory cell. The two main types of MRAM devices are spin-transfer torque MRAM (STT-MRAM) and SOT-MRAM.

In SOT-MRAM, the layer in which the magnetization is free to move, the free layer (FL), is placed on top of a heavy metal wire exhibiting a large spin Hall angle. A charge current passing through this heavy metal wire produces a transverse spin current which enters the FL and results in a torque being exerted on the magnetization. If current and torque are large enough, the precessional reversal of the magnetization is initiated. Thus, contrary to STT-MRAM, the read and write paths are separated in SOT-MRAM, reducing the susceptibility to oxide reliability issues due to high write currents through the MTJ.

Perpendicularly magnetized SOT-MRAM, however, still faces challenges for deterministic magnetization reversal. The torque generated by a current pulse is enough to bring the magnetization to an in-plane orientation. For deterministic switching, an external magnetic field is still needed [7]. In the last years, several approaches to achieve field-free switching have been proposed [8], of which we want to focus on a solution presented in [9]. Here, in addition to the heavy metal wire below the FL, another heavy metal wire, orthogonal to the first one, is placed on top of the FL (cf. Fig. 1). It was shown that by sending current pulses through the two metal wires, the perpendicularly magnetized FL can

* Corresponding author at: Christian Doppler Laboratory for Nonvolatile Magnetoresistive Memory and Logic, TU Wien, Gußhausstraße 27-29/E360, 1040 Vienna, Austria.
E-mail address: ender@iue.tuwien.ac.at (J. Ender).

https://doi.org/10.1016/j.microrel.2022.114570
Received 9 February 2022; Received in revised form 19 May 2022; Accepted 20 May 2022
Available online 2 June 2022

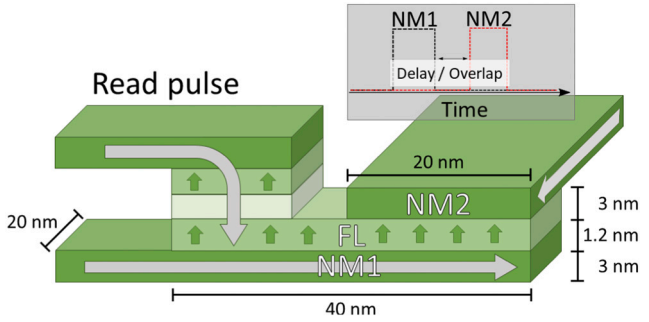

Fig. 1. SOT-MRAM cell for switching based on two orthogonal current pulses. The pulses are sent through the structure via two non-magnetic heavy metal wires, of which one is fully overlapping the FL (NM1) and one only partially (NM2).

reliably and deterministically be switched.

The dynamics of the magnetization in this memory cell is described by the following extended version of the Landau-Lifshitz-Gilbert equation:

$$\begin{aligned}\frac{\partial \mathbf{m}}{\partial t} &= -\gamma\mu_0 \mathbf{m}\times\mathbf{H}_{\text{eff}} + \alpha\mathbf{m}\times\frac{\partial \mathbf{m}}{\partial t}\\ &-\gamma\frac{\hbar}{2e}\frac{\theta_{SH} j_1}{M_S d}[\mathbf{m}\times(\mathbf{m}\times\mathbf{y})]f_1(t)\\ &+\gamma\frac{\hbar}{2e}\frac{\theta_{SH} j_2}{M_S d}[\mathbf{m}\times(\mathbf{m}\times\mathbf{x})]f_2(t)\end{aligned} \quad (1)$$

$\mathbf{m}$ is the normalized magnetization, γ is the gyromagnetic ratio, μ_0 is the vacuum permeability, α is the Gilbert damping factor, and M_S is the saturation magnetization. There are several contributions to the effective field $\mathbf{H}_{\text{eff}}$, namely the exchange field, the uniaxial perpendicular anisotropy field, the demagnetizing field, the current-induced field, and a stochastic thermal field at 300 K. $\hbar$ is the Planck constant, e is the electron charge, θ_{SH} is the spin Hall angle, and $j_{1,\ 2}$are the current densities of the pulses in the NM1 and NM2 wires along the directions $\mathbf{x}$ and $\mathbf{y}$, respectively. The two functions f_1 and f_2 define when the NM1 pulse and the NM2 pulse are active. The proper choice of f_1 and f_2 is of paramount importance as they determine the switching performance and reliability of the memory cell. However, an additional uncertainty factor is the underlying variability of MRAM manufacturing processes. Material parameters can vary up to ±10% [10], impacting the critical current of the memory cell. A first investigation of the influence of varying material parameters on the pulsed SOT-MRAM cell was already performed in [11].

3. Reinforcement learning

Contrary to supervised and unsupervised learning algorithms, the data used for RL is generated during the training phase. Through repeated interaction of a so-called RL agent with an environment the agent shall learn how best to apply actions in this environment in order to achieve a certain objective. After taking an action in the environment, the agent receives a reward signal as well as information about the new state the environment has transitioned to. Value-based RL algorithms, like Q-learning [2], use the information gathered during training to build up an approximation of the so-called action-value function Q_π:

$$Q_\pi(s,a) = \mathbb{E}\left[\sum_{t=0}^{T}\gamma^t R_t \mid S_t = s, A_t = a\right] \quad (2)$$

The action value function Q_π thus is the expected value of the reward R_t, discounted by the discount factor γ^t, accumulated over the length of the task, T, given that the current state S_t is s and the current action A_t is a. A good approximation of this function allows making the optimal action choice in every state. The more different state-action combinations the RL agent encounters during training, the more refined its action-value function gets. An important advantage of RL compared to conventional optimization methods is that the result is a trained model which was exposed to many different environment states and can dynamically adjust its actions based on the given situation. Conventional optimization methods optimize for a specific scenario and have to be re-run, if the scenario changes.

In the here-described experiments, the deep Q-network (DQN) algorithm [12] was employed, which is an advancement of the original Q-learning algorithm using a neural network to approximate the action-value function.

4. RL for SOT switching

The main components of the RL setup presented in Fig. 2 were implemented in Python using the RL library *Stable Baselines 3* [13]. The environment is based on an in-house developed finite differences simulator [14]. The parameters used in the micromagnetic simulations are given in Table I.

4.1. State

The number of input nodes of the agent's neural network is determined by the state vector. This consists of 11 variables which are returned from the environment after each iteration. The returned variables are the average values of the vector components of the magnetization, $m_{x/y/z}$, the difference of these values to the previous iteration, $\Delta m_{x/y/z}$, the average vector components of the effective magnetic field, $H_{eff,x/y/z}$, and two values indicating whether the two pulses can currently be turned on.

4.2. Actions

The DQN agent is allowed to perform four different actions, i.e., turning both pulses on, both pulses off, NM1 pulse on and NM2 pulse off, and NM1 pulse off and NM2 pulse on. Thus, the output layer of the DQN agent's neural network contains four nodes. In order to prevent arbitrarily fast switching of the two pulses, a minimum pulse period of 100 ps was defined. Previous publications have shown that the critical current for this memory cell is 120 µA [11]. The current value of the NM1 pulse was thus set a little higher to 130 µA. In [11] it was also shown that the current value of the NM2 pulse can be reduced without compromising the switching reliability and therefore in this work the NM2 current value was set to 100 µA.

4.3. Reward

The rewards returned by the environment after every iteration are calculated based on the following formula:

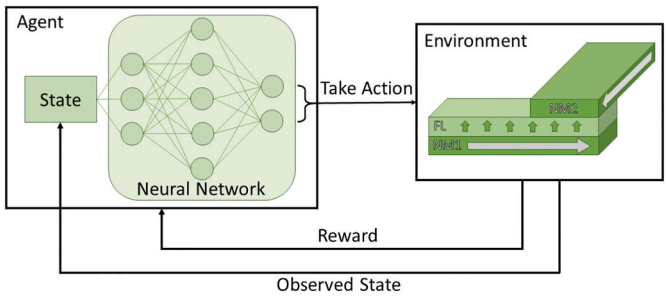

Fig. 2. General setup of the reinforcement learning simulation: A simulation of the SOT-MRAM cell acts as the environment which an agent interacts with to build up a policy based on a neural network.

Table I
SOT cell simulation parameters.

Parameter	Value
Saturation magnetization, M_S	1.1×10^6 A/m
Perpendicular anisotropy, K	8.4×10^5 J/m^3
Exchange constant, A	1.0×10^{11} J/m
Gilbert damping factor, α	0.035
Spin Hall angle, θ_{SH}	0.3

$$r = m_{z,target} - m_{z,current} \tag{3}$$

$m_{z,\ target}$ corresponds to the target value the z-component of the magnetization should reach and is set to -1. $m_{z,\ current}$ is the current value of the average z-component of the magnetization. The reward is always negative, unless the magnetization is perfectly aligned with the negative z-axis, at which point it would be zero, and it is more negative the farther away the current value of the magnetization is from the target value. As the DQN agent tries to maximize its overall accumulated reward over a switching simulation, this rewarding scheme ensures that the agent tries to bring the magnetization towards the target value as fast as possible to not accumulate more negative reward.

5. Results

Using the approach depicted in Fig. 2, which combines a finite differences micromagnetic simulation with an RL framework, a DQN agent was trained with the objective to learn fast switching of an SOT-MRAM cell. After completing a training period, this same setup can be used to perform simulations, in which the DQN agent applies pulses according to the policy it has learned. Fig. 3 shows a set of 50 realizations in which the agent at simulation time decided, when to apply pulses. The slight transparency of each plot line allows getting a qualitative idea of when pulses were applied in the various realizations. If the plot appears more solid, there are more lines overlapping, meaning that more realizations followed this trajectory. It can be seen that the first NM1 pulse, as well as the first two NM2 pulses are very solid lines and thus have been applied by the agent in all of the realizations. At around 1 ns, however, some variation can be observed: Due to the slight differences in the magnetization trajectories, the time at which the pulses were applied varies. The DQN agent successfully learned to apply the two pulses to switch the memory cell, as in all the realizations the magnetization crosses the threshold of -0.9, at which point the memory cell is considered switched, and remains below it.

To test the flexibility of the trained agent, a set of 441 simulations were performed in which the saturation magnetization M_S and the anisotropy constant K were varied individually by $\pm 10\%$. Figs. 4 and 5a show the pulses applied by the trained model and the resulting magnetization trajectories, respectively. In Fig. 4 one can see that the time at which pulses were applied varied a lot, but within the first 500 ps there is a cluster of overlapping pulses, suggesting that in many simulations the pulses were applied at the same time. The same holds for the magnetization trajectories shown in Fig. 5a: Many of them coincide in the first 500 ps, but diverge later. Some of the simulations never even reach the threshold value of -0.9 at the end of the 2 ns simulation time. Fig. 5b presents the distribution of final m_z values reached in the simulations. A large part of the trajectories (42%) reaches a final value below -0.9. 14.5%, however, remain close to the initial position of $+1$. In 11.8% of the simulations the agent is able to bring the trajectories in the xy-plane.

A different quantitative view of the results of this experiment is presented in Fig. 6. As the setup with the simulator and RL agent in the loop was employed, at the end of each such simulation an overall accumulated reward is returned, which correlates with how good the DQN agent could bring the magnetization towards the target value. In Fig. 6 this information is represented by the color-coding. It can be observed that a large part of the performed simulations achieved a high overall reward (yellow/light green). Towards the upper-left and bottom-right corners, though, the achieved reward declines. The behavior in the upper-left corner was already observed in [11] and is due to the fact that the critical current in this regime is higher. Overall, in 42% of the simulations the magnetization is successfully switched. For the same number of trajectories, 42%, it was not possible to bring them below the xy-plane, and 16% of the trajectories reached a final value between 0 and -0.9.

In [15] it was shown that, using the data presented in Fig. 6, a static pulse sequence can be extracted, which performs equally well over the parameter variation range. Said static pulse sequence can be seen in Fig. 7: It consists of a single NM1 pulse which is applied for 100 ps and two NM2 pulses, one started shortly after the NM1 pulse with a duration of 143 ps and another one starting 100 ps later, which is turned off again after 106 ps. We note that for demonstration purposes the static pulse scheme in Fig. 7 is defined with a very precise timing. Although we expect the scheme to also work, when the precise requirements are relaxed, additional studies beyond the scope of this manuscript incorporating rise and fall times as well as inexactness of timing and their influence on the reliability should be performed. Further investigating

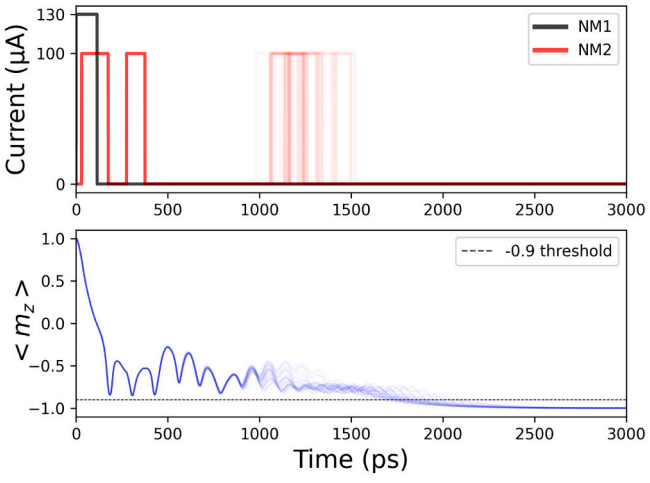

Fig. 3. Results of 50 realizations for fixed material parameters using the trained neural network model. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

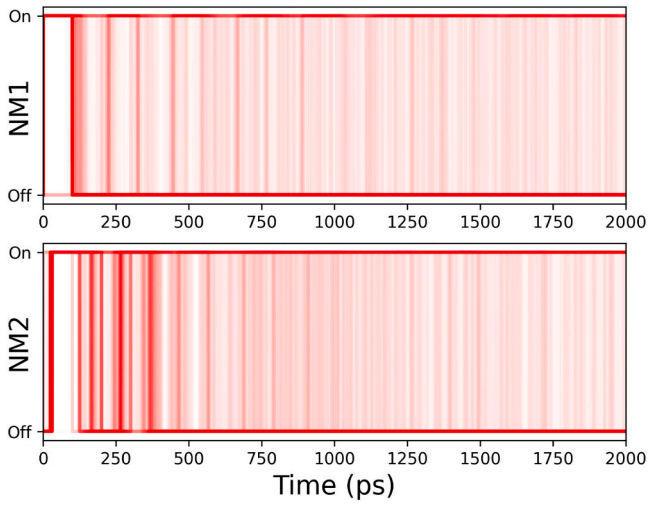

Fig. 4. Pulses applied to NM1 and NM2 during 441 realizations with varying material parameters. The results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

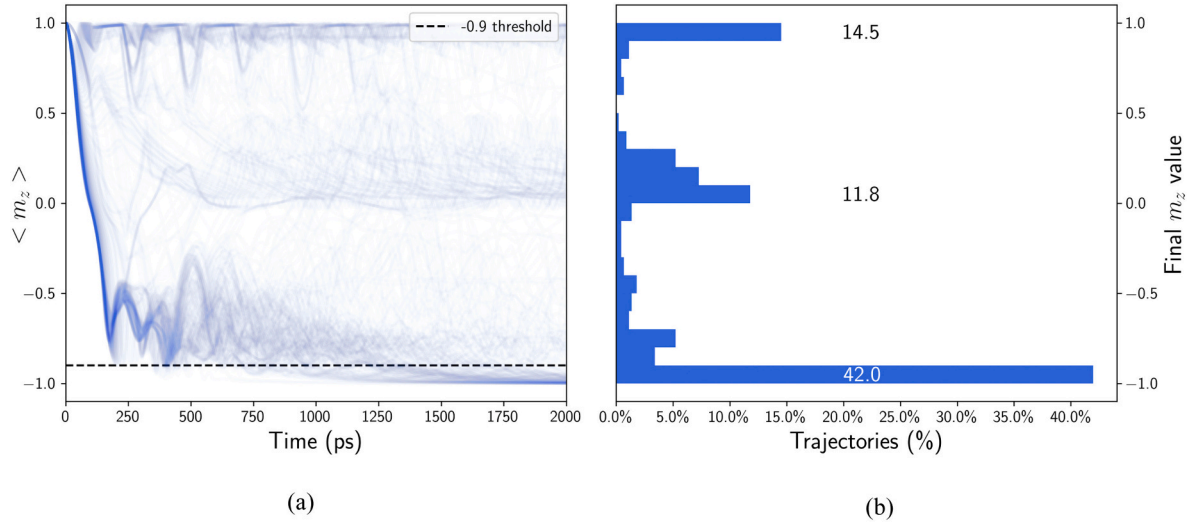

Fig. 5. (a) Average z-component of the magnetization for 441 realizations with varying material parameters. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid. (b) Percentage of trajectories, which reaches a certain final m_z value.

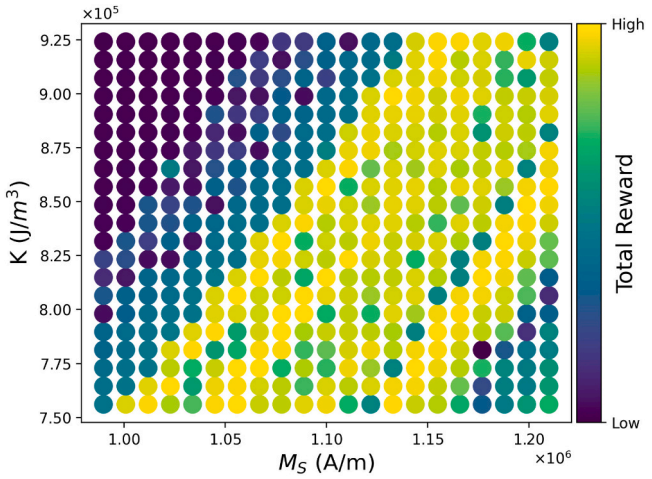

Fig. 6. Accumulated reward achieved for anisotropy constant K and saturation magnetization M_S varied by ±10%. Results are shown for a total of 441 realizations.

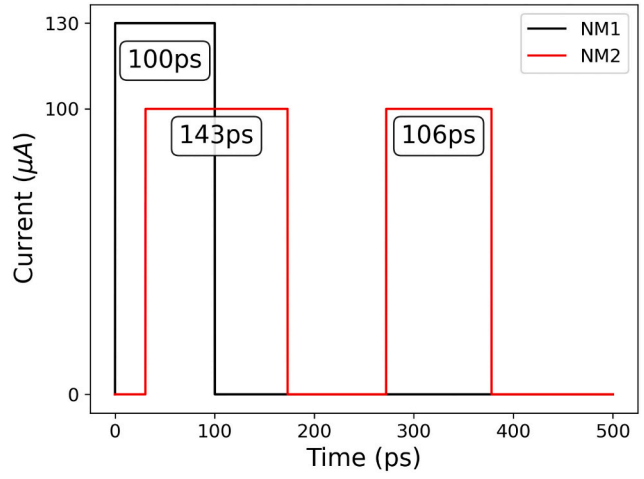

Fig. 7. Static pulse sequence extracted from successfully switched simulation runs with varying anisotropy constant and saturation magnetization.

the robustness of this static pulse sequence, we performed experiments with a ±10% variation of the FL thickness. Fig. 8 shows the achieved switching times depending on the different FL thicknesses. For each FL thickness, five simulations were performed. The color-coding indicates whether switching was achieved. There appears to be a sweet spot between 1.13 nm and 1.19 nm, for which switching is achieved significantly faster than above and below. For FL thicknesses above 1.26 nm switching could not be achieved anymore.

For reliable device operation, another aspect of the given SOT-MRAM cell in combination with the static pulse sequence has to be investigated. In a crossbar arrangement of memory cells, as depicted in Fig. 9, multiple cells are attached to a single heavy metal wire NM2, as the memory cell is selected using the NM1 wire, while the NM2 wire completes the switching of the cell. It has to be ensured, however, that the magnetization in the memory cells not intended to be switched is not accidentally reversed. We investigated the impact on the other memory cells attached to the NM2 wire by only applying the NM2 pulses and leaving the NM1 pulse turned off. The results of these experiments can be seen in Fig. 10. There is almost no observable difference between the 50 performed realizations, and the largest deviation from the stable state only reaches a value of ~0.95. The impact of the NM2 current pulses on the neighboring memory cells can thus be considered to be not critical.

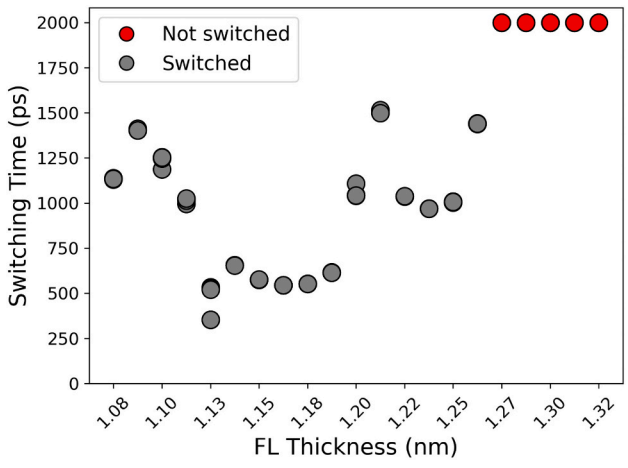

Fig. 8. Switching times achieved by applying the derived static pulse sequence to FL thickness variations of up to ±10%.

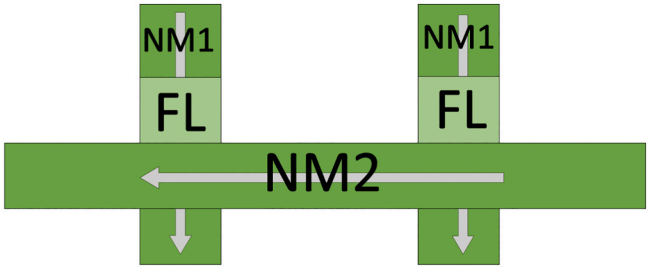

Fig. 9. Top-down view of two memory cells in a crossbar architecture sharing an NM2 wire.

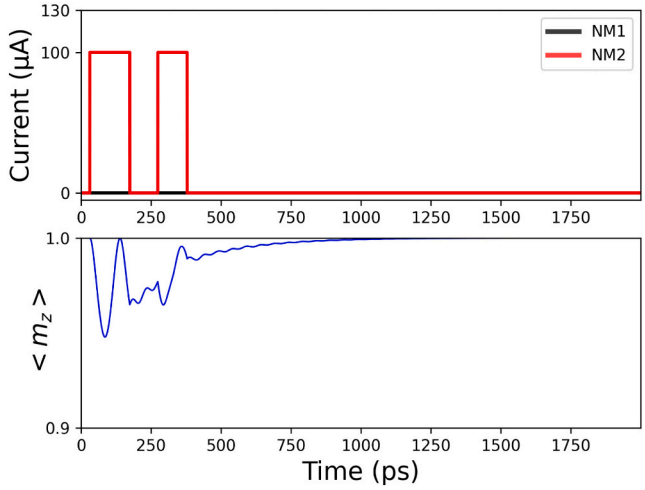

Fig. 10. Impact of NM2 pulses on neighboring memory cells in a crossbar architecture.

6. Conclusion

We demonstrated that RL can successfully be used to increase the reliability of a field-free SOT-MRAM switching scheme by optimizing the applied sequence of current pulses. We showed how an RL agent trained on fixed material parameters is able to generalize and find effective pulse sequences for switching over a wide material parameter range. Variations of saturation magnetization, anisotropy constant as well as FL thickness were investigated. Finally, the impact on neighboring memory cells in a crossbar architecture was studied, showing that only the memory cells selected by the first pulse are affected by the second pulse, proving the reliability of the scheme.

CRediT authorship contribution statement

Johannes Ender: Conceptualization, Software, Visualization.
Roberto Lacerda de Orio: Software, Methodology, Supervision.
Simone Fiorentini: Software, Writing – Review & Editing.
Siegfried Selberherr: Supervision, Writing – Review & Editing.
Wolfgang Goes: Writing – Review & Editing.
Viktor Sverdlov: Supervision, Writing – Review & Editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the Christian Doppler Association is gratefully acknowledged. The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

References

[1] A. Manchon, S. Zhang, Theory of spin torque due to spin-orbit coupling, Phys. Rev. B, Condens. Matter 79 (Mar. 2009) 094422, https://doi.org/10.1103/PhysRevB.79.094422.
[2] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT press, Cambridge, MA, USA, 1998.
[3] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, et al., A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, Science 362 (6419) (Dec. 2018) 1140–1144, https://doi.org/10.1126/science.aar6404.
[4] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, et al., Machine learning and the physical sciences, Rev. Mod. Phys. 91 (4) (Dec. 2019), 045002, https://doi.org/10.1103/RevModPhys.91.045002.
[5] J. Ender, R. Orio, S. Fiorentini, S. Selberherr, W. Goes, V. Sverdlov, Reinforcement learning to reduce failures in SOT-MRAM switching, in: Proc. of the IPFA, Nov. 2021, pp. 1–4, https://doi.org/10.1109/IPFA53173.2021.9617362.
[6] R.L. de Orio, J. Ender, S. Fiorentini, W. Goes, S. Selberherr, V. Sverdlov, Optimization of a spin-orbit torque switching scheme based on micromagnetic simulations and reinforcement learning, Micromachines 12 (4) (Apr. 2021) 443, https://doi.org/10.3390/mi12040443.
[7] S. Fukami, T. Anekawa, C. Zhang, H. Ohno, A spin-orbit torque switching scheme with collinear magnetic easy axis and current configuration, Nat. Nanotechnol. 11 (Mar. 2016) 621–626, https://doi.org/10.1038/nnano.2016.29.
[8] S. Fukami, C. Zhang, S. DuttaGupta, A. Kurenkov, H. Ohno, Magnetization switching by spin-orbit torque in an antiferromagnet-ferromagnet bilayer system, Nat. Mater. 15 (Feb. 2016) 535–541, https://doi.org/10.1038/nmat4566.
[9] V. Sverdlov, A. Makarov, S. Selberherr, Two-pulse sub-ns switching scheme for advanced spin-orbit torque MRAM, Solid State Electron. 155 (Mar. 2019) 49–56, https://doi.org/10.1016/j.sse.2019.03.010.
[10] J. Song, H. Dixit, B. Behin-Aein, C.H. Kim, W. Taylor, Impact of process variability on write error rate and read disturbance in STT-MRAM devices, IEEE Trans. Magn. 56 (12) (Dec. 2020) 1–11, https://doi.org/10.1109/TMAG.2020.3028045.
[11] R.L. de Orio, J. Ender, S. Fiorentini, W. Goes, S. Selberherr, V. Sverdlov, Numerical analysis of deterministic switching of a perpendicularly magnetized spin-orbit torque memory cell, IEEE J. Electron Devices Soc. 9 (Nov. 2020) 61–67, https://doi.org/10.1109/JEDS.2020.3039544.
[12] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, et al., Human-level control through deep reinforcement learning, Nature 518 (7540) (Feb. 2015) 529–533, https://doi.org/10.1038/nature14236.
[13] A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, N. Dormann, Stable baselines 3, Available online, https://github.com/DLR-RM/stable-baselines3. (Accessed 8 May 2022).
[14] A. Makarov, Modeling of Emerging Resistive Switching Based Memory Cells, Ph.D. Thesis, Institute for Microelectronics, TU Wien, Vienna, 2014, https://doi.org/10.13140/RG.2.2.11456.74242.
[15] J. Ender, R. Orio, S. Fiorentini, S. Selberherr, W. Goes, V. Sverdlov, Reinforcement learning approach for deterministic SOT-MRAM switching, in: Proc. of the SPIE 11805, Aug. 2021, pp. 1180519-1–1180518-8, https://doi.org/10.1117/12.2593937.