Reinforcement Learning to Reduce Failures in SOT-MRAM Switching

Johannes Ender[1,2], Roberto Lacerda de Orio[2], Simone Fiorentini[1,2], Siegfried Selberherr[2], Wolfgang Goes[3], Viktor Sverdlov[1,2]
[1]Christian Doppler Laboratory for Nonvolatile Magnetoresistive Memory and Logic at the
[2]Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, 1040 Vienna, Austria
[3] Silvaco Europe Ltd., Cambridge, United Kingdom
Phone : +43 1 58801-36022, Email : ender@iue.tuwien.ac.at

2021 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA) | 978-1-6654-3988-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IPFA53173.2021.9617362

***Abstract*—We demonstrate the use of reinforcement learning for achieving efficient switching schemes for a field-free operation of spin-orbit torque magnetoresistive random access memory cells. A cell is switched purely electrically by applying two orthogonal current pulses. It is shown that using a reinforcement learning approach, a neural network model can be trained on a fixed material parameter set for finding optimal switching pulse sequences. This model is not only suitable to switch a memory cell in the presence of thermal fluctuations, but also for varied cell material parameters.**

Keywords—reinforcement learning, spin-orbit torque memory, magnetic field-free, switching reliability

I. Introduction

The charge-based SRAM memory cells broadly in use today are volatile by design. The progressive down-scaling of these devices comes at the price of an increase in standby power consumption. A possible solution to this problem is to use adequately fast nonvolatile memory devices. Spin-orbit torque magnetoresistive random access memory (SOT-MRAM) is one of the most promising variants. SOT-MRAM devices exhibit large endurance and very fast operation, which makes them suitable for use in high-level caches, where currently CMOS-based devices are predominant. Another technology development entering various scientific fields is machine learning (ML). Its ability to handle huge data sets and infer knowledge from them has enabled many scientific advances [1]. The ML sub-branch of reinforcement learning (RL) [2] is based on the imitation of the way humans learn, with impressive demonstrations of superior performance in chess or Go [3].

This work is an advancement to the previous introduction of RL for SOT-MRAM switching presented in [4]. In this work we show that RL can be used to find pulse sequences for reliable SOT-MRAM cell switching. Most importantly, a model trained for a specific parameter set performs excellently on a broad distribution of varying materials and parameters and can help to find more reliable pulse sequences.

II. Spin-Orbit Torque Memory

In MRAM devices, the information is stored as the relative orientation of the magnetization in two ferromagnetic layers, which, together with a tunnel barrier which lies between them, form a magnetic tunnel junction (MTJ). The orientation of the magnetization changes only in one of the two ferromagnetic layers, the free layer (FL). The other one is called reference layer and its magnetization orientation is kept constant. The two biggest contenders in the nonvolatile memory field are spin-transfer torque MRAM (STT-MRAM) and SOT-MRAM. The reading procedure for stored information is the same in these two types of devices, but they differ in how the information is written. In STT-MRAM a spin-polarized current is sent through the MTJ, initiating the precessional motion of the magnetization, which eventually leads to switching. For SOT-MRAM on the other hand, the switching current does not flow through the MTJ, but through a heavy metal wire attached to the bottom of the FL, which exhibits a large spin Hall angle. Due to the spin Hall effect, the charge current is converted into a transverse spin current, exerting a torque on the magnetization in the FL, again initiating the precession and reversal of the magnetization. This separation of the read and the write path has positive effects on the reliability of these devices, as the current flowing through the MTJ cannot only lead to degradation of the thin tunnel oxide, but also to erroneously written information, when a read operation is performed. However, to deterministically switch an SOT-MRAM cell, an external magnetic field is still needed [5]. Many solutions have been proposed, e.g. [6], out of which one was introduced only recently and solves the problem by adding a second heavy metal wire, orthogonal to the first one, but only partially overlapping the FL (cf. Fig. 1) [7]. By applying pulses to both wires, the magnetization can be reversed reliably without the need of an external field. However, pulses applied to both metal wires are needed for reliable switching. The dynamics of the magnetization in the free layer are described by an extended version of the Landau-Lifshitz-Gilbert equation

$$\begin{aligned}\frac{\partial \mathbf{m}}{\partial t} &= -\gamma\mu_0 \mathbf{m}\times\mathbf{H_{eff}} + \alpha\mathbf{m}\times\frac{\partial \mathbf{m}}{\partial t}\\ &-\gamma\frac{\hbar}{2e}\frac{\theta_{SH} j_1}{M_S d}[\mathbf{m}\times(\mathbf{m}\times\mathbf{y})]\theta_1(t)\\ &+\gamma\frac{\hbar}{2e}\frac{\theta_{SH} j_2}{M_S d}[\mathbf{m}\times(\mathbf{m}\times\mathbf{x})]\theta_2(t)\,,\end{aligned} \tag{1}$$

in which $\mathbf{m}$ 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. The effective field $\mathbf{H_{eff}}$ consists of several contributions, namely the exchange field, the uniaxial perpendicular anisotropy field, the demagnetizing field, the current-induced field, and a stochastic thermal field at 300 K.

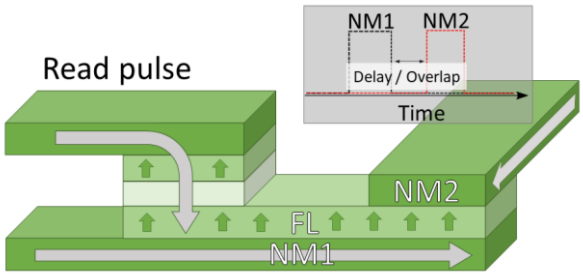

Fig. 1: SOT-MRAM cell for switching based on two orthogonal current pulses. The pulses are sent through the structure via two non-magnetic heavy metal wires, of which one is fully overlapping the FL (NM1) and one only partially (NM2).

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.

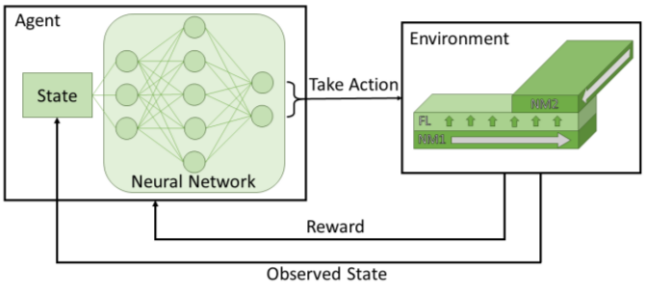

Fig. 2: General setup of the reinforcement learning simulation: A simulation of the SOT-MRAM cell acts as environment which an agent interacts with to build up a policy based on a neural network.

Θ_1 and Θ_2 are functions defining when the NM1 pulse and the NM2 pulse are active. While previous publications on this SOT memory cell have investigated the switching reliability under variation of the pulse duration, pulse current values, material parameter variation, and the delay of an NM2 pulse following an NM1 pulse, decisions on the exact placement of the pulses were based on intuition. An automated way for finding optimal pulse sequences is highly desirable. What complicates the challenge of finding good pulse sequences for switching a memory cell, is the fact that MRAM fabrication processes underlie variability. Thus, material parameters like the saturation magnetization and the magnetic anisotropy can have variations of $\pm 10\%$ [8]. The impact of variations of these parameters on the critical current of the pulsed SOT-MRAM cell was already investigated in [9].

III. Reinforcement Learning

In reinforcement learning, an agent repeatedly interacts with an environment by performing certain actions. In this iterative process, after every action of the agent, the environment returns information about the new state it has transitioned to and a reward, indicating how good or bad the action was. This information is used by the agent to refine its policy, which is a mapping from states to actions. The basis for the decision-making in so-called value-based reinforcement learning algorithms, like Q-learning [2], is the action-value function.

$$Q_\pi(s,a) = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_t \mid S_t = s, A_t = a\right] \quad (1)$$

The action-value function is defined with respect to a policy π. It is the expectation value of the cumulated reward R_t, discounted by the factor γ, given that the action at time t, A_t, is a and that the state at time t, S_t, is s. The discount factor γ defines how strongly future rewards influence the action-value estimate at time t. Other solution methods for approximate optimization, like genetic algorithms, simulated annealing or evolution strategies do not estimate value functions. They rather mimic biological evolution with a survival of the fittest behavior. As they do not make use of state or action information, learning is often less efficient than in RL [2]. If an RL agent encounters the same state multiple times, it can refine its estimate of the action-value function by either making a greedy decision and taking the action which promises the highest reward, or it can decide to further explore the state-action space by performing a - from the current point of view - sub-optimal action, with the possibility of discovering a new, better policy. As the action-value function is the basis for decision-making in Q-learning, having a good estimate is important and the state-action space should be explored thoroughly to have a good action-value approximation. The trade-off between exploration and exploitation of existing knowledge can be controlled with the exploration probability ε. The action with the highest estimate of the action-value is taken with a probability 1-ε and an explorative choice is made with a probability ε. For our experiments we employed the deep Q-network (DQN) algorithm [10]. This advancement of the original Q-learning algorithm uses a neural network as function approximator.

IV. RL For SOT Switching

For applying RL to the pulsed SOT-MRAM cell, the approach depicted in Fig. 2 was set up. Based on an existing Python RL library [11], a custom RL environment was created, whose core consists of an in-house developed finite difference simulator of the previously described memory cell [12]. The parameters used for the micromagnetic simulations are given in Table I. The implementation of [11] was used for the DQN functionality with the RL agent being configured with the default parameters of the RL framework, apart from the given settings in Table II. With these settings the best learning performance was observed.

A. State

A crucial part for deciding which action to perform is the state vector returned to the RL agent at every time step. It has to be ensured that ambiguities are avoided and that the state delivers sufficient information. It is thus important, that data about the dynamics of the magnetization are included, because it would not be possible to decide on the best action without knowing in which direction the magnetization components are moving. The state vector used for the experiments consists of

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
Free layer dimensions	40 nm × 20 nm × 1.2 nm
NM1: w1 × l	20 nm × 3 nm
NM2: w2 × l	20nm × 3 nm

TABLE II. DQN Parameters

Parameter	Value
Neural network size	11 × 150 × 100 × 4
Discount factor, γ	0.9997
Learning rate	7.5×10^{-4}
Exploration fraction	0.2
Final exploration probability, ε	0.01
Replay buffer size	3×10^5
Batch size	512

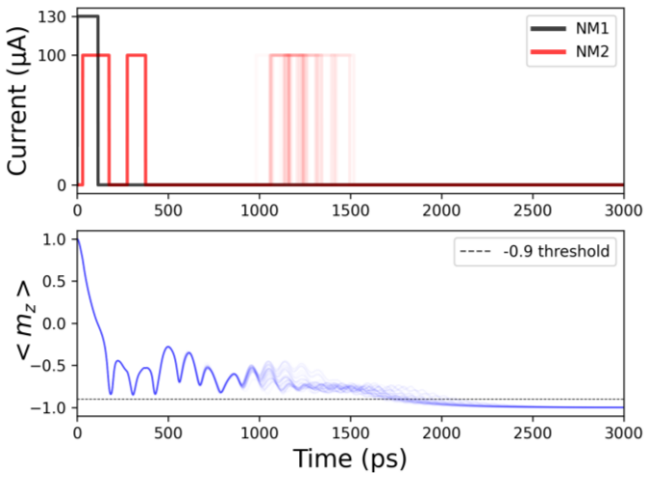

Fig. 3: Results of 50 realizations for fixed material parameters using the trained neural network model. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

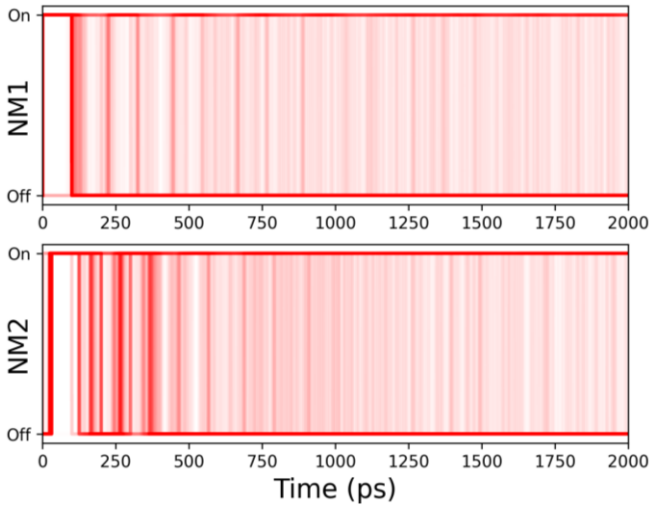

Fig. 4: Pulses applied to NM1 and NM2 during 441 realizations with varying material parameters. The results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

11 variables: the averages of the magnetization vector components (m_x, m_y, m_z), the average effective field vector components ($H_{eff,x}$, $H_{eff,y}$, $H_{eff,z}$), the difference of the magnetization's vector components to the previous time step (Δm_x, Δm_y, Δm_z), and two variables indicating whether the NM1 and NM2 pulse are currently settable. These 11 state variables lead to a policy network input layer size of 11.

B. Actions

The agent is allowed to perform four different actions. With the restriction of a minimum amount of time between pulse changes of 100 picoseconds, the NM1 pulse and the NM2 pulse can be turned on and off individually. The output layer of the policy neural network is thus of size 4. The output of each of the output layer's neurons corresponds to an estimate of the action-value function for a specific state-action pair. Depending on whether a greedy action or an explorative action should be taken, either an action promising the highest reward or a random one is returned. The returned action is an integer value from 0 to 3, which subsequently is used to select the settings of the pulses in the environment code. Based on the critical current value of 120 µA [9], the current value for the NM1 pulse was fixed to a value slightly above, i.e. 130 µA. As in [9] it was shown that the current value of the NM2 pulse can be below the critical current, a current value of 100 µA was chosen.

C. Reward

The rewarding scheme is what leads the learning algorithm in the right direction and thus has to be designed carefully. The objective of the experiments is to achieve a fast transition of the average z-component of the magnetization from $+1$ to -1. For every simulation step, the agent receives a negative reward whose exact value depends on the distance between the current position of the average z-component $m_{z,current}$ and the target value $m_{z,target}$ and is defined as:

$$r = m_{z,target} - m_{z,current} \tag{2}$$

Thus, with $m_{z,target} = -1$, the further away the magnetization is from the target value, the more negative the reward is. This also ensures that the agent tries to get the z-component down rapidly, in order to reduce the overall accumulated negative reward.

V. Results

The RL approach as seen in Fig. 2 was employed to train a policy neural network to reverse the FL magnetization in the SOT memory cell from $+1$ to -1, whereas we consider the reversal to be successful when the z-component of the magnetization reaches -0.9. The training process consisted of repeated, independent switching simulations, each with a maximum simulated time of 2 nanoseconds. After 1 picosecond, the new state information as well as the reward for the previously selected action were returned to the agent, amounting to a total of 10^6 state-action pairs with their rewards for the agent to learn from.

Among the performed training processes, the best-performing neural network model was further scrutinized by again performing switching simulations, but without any further adjustment of the neural network weights. Like this, the switching reliability over 50 realizations under a fluctuating thermal field was investigated. The results are shown in Fig. 3. The applied pulses and the magnetization trajectories of the single realizations are almost undistinguishable up until 1 nanosecond. Only afterwards the thermal field leads to a slight divergence of the magnetization between the realizations. Then the neural network model applies further NM2 pulses, of which the exact positions vary depending on the respective trajectory of the magnetization. Nevertheless, within a time window of ~2 nanoseconds, the z-component of all realizations is deterministically brought from $+1$ to -1.

As material parameter variations of $\pm 10\%$ can occur in MRAM fabrication processes [8], we further studied how reliable the trained model can reverse the magnetization in the presence of variations of the saturation magnetization M_S and the anisotropy constant K. For every combination of the parameters a simulation was performed and the trained model decided when to apply NM1 and NM2 pulses. The resulting trajectories of the current pulses and the magnetization can be seen in Fig. 4 and Fig. 5, respectively. Comparing the results

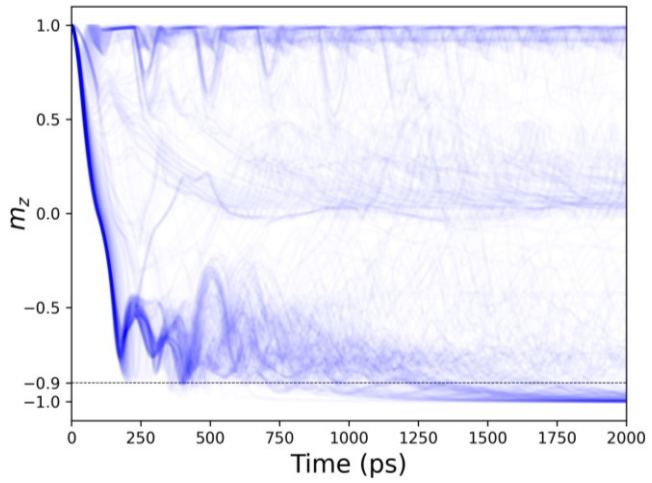

Fig. 5: Average z-component of the magnetization for 441 realizations with varying material parameters. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

with the ones presented in Fig. 3, one can see that there is more variability in the applied pulses as well as the magnetization trajectories. Out of the 441 realizations, ~42 % reach the -0.9 threshold at the end of the 2-nanosecond simulation time. An equal number of trajectories, ~42 %, could not even be brought below the xy-plane, leaving ~16 % of the trajectories between 0 and -0.9.

For a clearer picture of the performance of the model in this varied-parameter scenario, Fig. 6 gives an overview of the achieved accumulated reward for all the examined variation combinations. Most apparent is the upper left corner, for which the model accumulates more negative rewards, i.e. struggles to bring the z-component closer to -1. These low-performing runs correspond to the magnetization trajectories whose z-components stay positive throughout the simulation. This is consistent with results published in [9], which indicate that in this region of the two material parameters, a higher current is required to deterministically switch the memory cell. The lower right corner suggests that the applied current can also be too high for certain parameter combinations, as the achieved reward in this region starts to

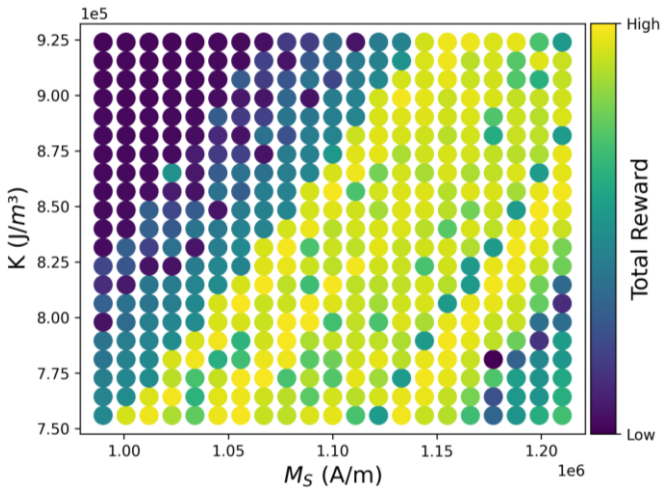

Fig. 6: Accumulated reward achieved for anisotropy constant K and saturation magnetization M_S varied by $\pm 10\%$. Results are shown for a total of 441 realizations.

decrease and it is known from [9] that for a high saturation magnetization and a low anisotropy constant, the critical current is well below the used 130 µA. Nevertheless, a broad parameter range can still successfully be switched.

VI. Conclusion

We demonstrated that reinforcement learning is a promising technique to optimize the switching of SOT-MRAM cells. It is shown that, by training the neural network model to maximize its received reward during the learning phase for a fixed parameter set, an optimal pulse scheme for deterministic switching in the presence of thermal fluctuations and parameter variations is achieved. Pulse sequences derived from this study can considerably reduce the failure rate of SOT-MRAM write operations and increase their reliability. By incorporating material parameter variations into the learning phase of the neural network, there is potential for improving the switching reliability even more.

References

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, *et al.*, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, no. 4, p. 045002, Dec. 2019. DOI: 10.1103/RevModPhys.91.045002

[2] R. S. Sutton, and A. G. Barto, "Reinforcement learning: An introduction," Cambridge, MA, USA: MIT press, 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*, vol. 362, no. 6419, pp. 1140-1144, Dec. 2018. DOI: 10.1126/science.aar6404

[4] R. L. de Orio, J. Ender, S. Fiorentini, W. Goes, S. Selberherr, and V. Sverdlov, "Optimization of a spin-orbit torque switching scheme based on micromagnetic simulations and reinforcement learning," *Micromachines*, vol. 12, no. 4, p. 443, Apr. 2021. DOI: 10.3390/mi12040443

[5] S. Fukami, T. Anekawa, C. Zhang, and H. Ohno, "A spin-orbit torque switching scheme with collinear magnetic easy axis and current configuration," *Nat. Nanotechnol.*, vol. 11, pp. 621–626, Mar. 2016. DOI: 10.1038/nnano.2016.29

[6] S. Fukami, C. Zhang, S. DuttaGupta, A. Kurenkov, and H. Ohno, "Magnetization switching by spin-orbit torque in an antiferromagnet-ferromagnet bilayer system," *Nat. Mater.,* vol. 15, pp. 535–541, Feb. 2016. DOI:10.1038/nmat4566

[7] V. Sverdlov, A. Makarov, and S. Selberherr, "Two-pulse sub-ns switching scheme for advanced spin-orbit torque MRAM," *Solid-State Electron.*, vol. 155, pp. 49–56, Mar. 2019. DOI: 10.1016/j.sse.2019.03.010

[8] J. Song, H. Dixit, B. Behin-Aein, C. H. Kim, and W. Taylor, "Impact of process variability on write error rate and read disturbance in STT-MRAM devices," *IEEE Trans. Magn.*, vol. 56, no. 12, pp. 1-11, Dec. 2020. DOI: 10.1109/TMAG.2020.3028045

[9] R. L. de Orio, J. Ender, S. Fiorentini, W. Goes, S. Selberherr, and V. Sverdlov, "Numerical analysis of deterministic switching of a perpendicularly magnetized spin-orbit torque memory cell," *IEEE J. Electron Devices Soc.*, vol. 9, pp. 61-67, Nov. 2020. DOI: 10.1109/JEDS.2020.3039544

[10] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, Feb. 2015. DOI: 10.1038/nature14236

[11] A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, and N. Dormann, Stable Baselines 3. Available online: https://github.com/DLR-RM/stable-baselines3 (accessed on 12 May 2021).

[12] A. Makarov, "Modeling of emerging resistive switching based memory cells," Ph.D. Thesis, Institute for Microelectronics, TU Wien, Vienna, 2014. DOI: 10.13140/RG.2.2.11456.74242