PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Reinforcement learning approach for deterministic SOT-MRAM switching

Ender, Johannes, L. de Orio, Roberto, Fiorentini, Simone, Selberherr, Siegfried, Goes, Wolfgang, et al.

Johannes Ender, Roberto L. de Orio, Simone Fiorentini, Siegfried Selberherr, Wolfgang Goes, Viktor Sverdlov, "Reinforcement learning approach for deterministic SOT-MRAM switching," Proc. SPIE 11805, Spintronics XIV, 1180519 (1 August 2021); doi: 10.1117/12.2593937

Event: SPIE Nanoscience + Engineering, 2021, San Diego, California, United States

SPIE.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Reinforcement Learning Approach for Deterministic SOT-MRAM Switching

Johannes Ender *[a,b], Roberto L. de Orio[b], Simone Fiorentini[a], Siegfried Selberherr[b],
Wolfgang Goes[c], and Viktor Sverdlov[a]
[a]Christian Doppler Laboratory for Nonvolatile Magnetoresistive Memory and Logic at the
[b]Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, 1040 Vienna, Austria
[c]Silvaco Europ Ltd., Cambridge, United Kingdom

ABSTRACT

We employ a reinforcement learning strategy for finding switching schemes for deterministic switching of a spin-orbit torque magnetoresistive random access memory cell. The free layer of the memory cell is perpendicularly magnetized, and the spin-orbit torques are generated by currents through two orthogonal heavy metal wires. A rewarding scheme for the reinforcement learning approach is defined such that the objective of the algorithm is to find a pulse sequence that leads to fast deterministic field-free switching of the memory cell. The reliability of the found switching scheme is tested by performing micromagnetic simulations. The results show that a neural network model trained on fixed material parameters is able to reverse the memory cell magnetization for a wide range of material parameters and can be used to derive a writing pulse sequence for fast and deterministic spin-orbit torque switching of a perpendicular free layer.

Keywords: Spin-orbit torque, reinforcement learning, field-free switching, SOT-MRAM

1. INTRODUCTION

Nonvolatile memories become an important solution to reducing the energy consumption of modern integrated circuits, as traditional semiconductor device scaling results in a high energy consumption due to increased leakage currents. Spin-orbit torque magnetoresistive random access memories (SOT-MRAM) are an attractive alternative to their charge-based counterparts for applications like high-density static RAM in registers, as they exhibit nonvolatility, large endurance, and high-speed operation.[1-3] However, the deterministic switching of most promising SOT MRAM cells with a perpendicularly magnetized free layer requires an external magnetic field.[3, 4] The two-pulse switching scheme introduced in Sverdlov *et al.*[5] is purely electrical and field-free. However, it requires an optimization of the applied pulse sequence.[6]

The increase in available computational power has enabled scientific simulations to generate ever more data, but at the same time allowed the field of machine learning to flourish and its use in science has become widespread.[7] Machine learning algorithms can easily handle large amounts of data and infer knowledge from it. The machine learning sub-field of reinforcement learning (RL)[8], which tries to mimic the way humans learn, was successfully applied in various fields, e.g. Fösel *et al.*[9], after initially having made breakthrough advances in learning how to play games like chess or Go.[10]

Based on the results previously published in Orio *et al.*[11], which introduced the application of RL for SOT-MRAM switching, this work demonstrates a strong advancement in the level of maturity of the RL approach, but also its practical usability for finding pulse sequences for reliable SOT-MRAM cell switching.

*ender@iue.tuwien.ac.at; phone +43 1 58801 36022

Spintronics XIV, edited by Henri-Jean M. Drouhin, Jean-Eric Wegrowe,
Manijeh Razeghi, Proc. of SPIE Vol. 11805, 1180519 · © 2021 SPIE
CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2593937

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

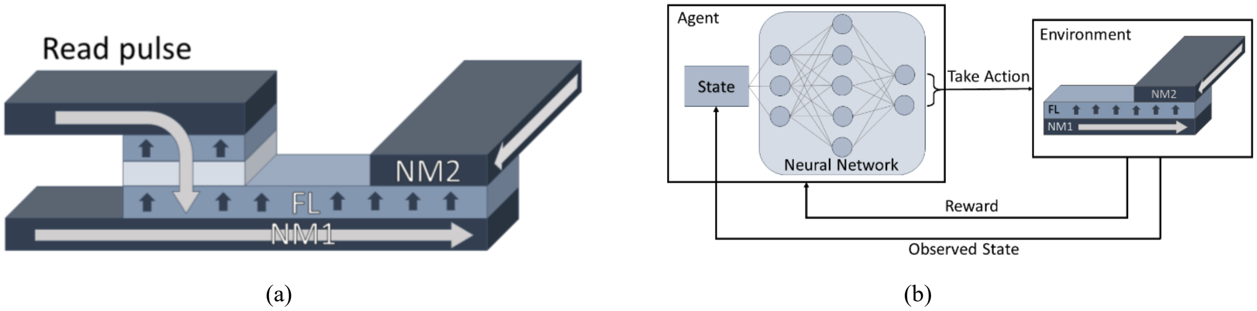


Figure 1. (a) SOT-MRAM cell for switching based on two orthogonal current pulses. The pulses are sent through the structure via two non-magnetic heavy metal wires, of which one is fully overlapping the FL (NM1) and one only partially (NM2) (b) General setup of the reinforcement learning simulation: A simulation of the SOT-MRAM cell acts as environment which an agent interacts with to build up a policy based on a neural network.

2. SPIN-ORBIT TORQUE MEMORY

SOT-MRAM, like spin-transfer torque MRAM (STT-MRAM), uses magnetic tunnel junctions (MTJ) as means for storing information. MTJs are structures consisting of a thin tunnel barrier which is sandwiched between two ferromagnetic layers. In one of the two layers, the reference layer, the magnetization is fixed, while in the other layer, the free layer, the magnetization can change its orientation, such that the relative orientation of the magnetization in the MTJ can be in two stable states, anti-parallel and parallel. Due to the tunneling magnetoresistance effect, these two states possess a different electrical resistance and can be used to store binary information. The writing of the information differentiates SOT-MRAM and STT-MRAM. In SOT-MRAM a heavy metal wire NM1 with a large spin Hall angle is attached to the free layer. When sending a charge current through the metal wire, a transverse spin-current is created which exerts a torque on the magnetization of the free layer and initiates a precessional motion, eventually switching the memory cell. For deterministic switching, however, an external magnetic field is still needed.[4] Among several proposed solutions to this problem (e.g. Fukami *et al.*[12]), we focus on an approach presented in Sverdlov *et al.*[5]. Here, a second heavy metal wire NM2 is attached to the free layer (cf. Fig. 1(a)). By sending pulses through these two orthogonal wires, it was shown that the memory cell can be switched reliably without the need of an external magnetic field.

The dynamics of the magnetization in the free layer of this pulsed SOT-MRAM cell are described by the following extended Landau-Lifshitz-Gilbert equation.

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

$\mathbf{m}$ 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. 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. f_1 and f_2 are functions defining when the NM1 pulse and the NM2 pulse are active. Until now, coming up with pulse sequences which lead to switching was a heuristic approach, based on the intuitive understanding of the behavior of the memory cell, but having an automated way for arriving at pulse sequences for optimal behavior is highly desirable.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3. REINFORCEMENT LEARNING

The general idea of reinforcement learning (RL) is to mimic the way humans learn, by repeated interaction with the environment and gathering experiences in the form of positively and negatively reinforcing signals, which allow building up a strategy to achieve a certain goal. In RL terms, the learning entity is the agent which interacts with an environment by performing actions and receives a reward and information about the current state of the environment. This iterative process gathers experience about which actions are good in certain states and bad in others, allowing the agent to assign a value to state-action pairs, expressed as the following action-value function.[8]

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

Following the policy π, the action-value function is defined as the expectation value of the cumulated reward R_t, discounted by the factor γ. This factor defines the problem horizon, that is, how far into the future rewards are considered for the estimation of the expected cumulative reward at time t. R_t is the reward at time t, given that the action A_t is a and that the state at time t, S_t, is s. During a training period one tries to get an approximation of the action-value function as good as possible, such that one always knows the action which leads to the highest expected reward in a certain state. Value-based algorithms like Q-learning can be used to approximate the action-value function.[8]

4. RL FOR SOT-MRAM SWITCHING

The RL setup shown in Fig. 1(b) is implemented using an open-source Python library which provides implementations for several popular RL algorithms and allows to easily create custom environments.[13] We employed the deep Q-network (DQN) algorithm implementation of Raffin *et al.*[13] for the agent. The DQN algorithm is a derivative of the Q-learning algorithm, which uses a neural network for the approximation of the action-value function.[14] In the following, the other components of the RL setup are described.

4.1 State

Besides the reward, the state information is the main source of information for the agent. For our experiments, we used a state vector consisting of 11 variables. These are the average vector components of the magnetization, $m_{x,y,z}$, the difference of the average magnetization vector components to the last iteration, $\delta m_{x,y,z}$, the average vector components of the effective magnetic field, $H_{eff_{x,y,z}}$, and two Boolean variables indicating whether the two pulses can be currently set.

4.2 Actions

The agent can turn both pulses on and off individually, which results in four different states the NM1 and the NM2 wire can be in. As shown in Orio *et al.*[15], the critical current value of the given SOT-MRAM cell is 120 µA, with the possibility of reducing the NM2 pulse current to below the critical current, while maintaining deterministic switching. Thus, the current values for NM1 and NM2 were chosen to be 130 µA and 100 µA, respectively.

4.3 Environment

The environment consists of a simulation of the SOT-MRAM cell presented in Fig. 1(a). The in-house developed simulator written in C++ uses the finite difference method to solve Eq. (1).[16]

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

4.4 Reward

The second input of the agent besides the state information is the reward. The reward is what defines the objective of the agent, which is to obtain fast reversal of the z-component of the magnetization from +1 to -1. The following simple reward formula achieves exactly this goal:

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

If the reward is more negative, the farther away is the current value of the magnetization from the target value, which encourages the agent to bring them closer. In addition, as the reward is always negative, in order to accumulate less negative reward and maximize the overall reward, the agent will try to reverse the magnetization quickly.

5. RESULTS

During a training phase, the presented RL setup was used to let the agent interact with the SOT-MRAM cell environment. By adjusting the weights of its neural network approximation of the policy, the strategy to apply pulses in order to switch the SOT-MRAM cell as fast as possible was refined. The material parameters used for training were M_S=1.1×10^6 A/m and K=8.4×10^5 J/m^3. By using the policy network generated during the training phase, but without further adjustment of the weights, switching simulations were performed with the same setup as depicted in Fig 1(b). Because of the thermal field, there is a stochastic element influencing the behavior of the magnetization, therefore 50 switching realizations were performed. The resulting trajectories of the z-component of the magnetization as well as the pulses applied by the neural network model are shown in Fig. 2. The transparency of the single plot line indicates the frequency this path was visited: Paths taken more often appear more solid, while paths taken less often appear more transparent. Fig. 2 shows that a single NM1 pulse is applied right in the beginning, while the NM2 pulse is turned on several times. The first two NM2 pulses shown in solid red were applied in all simulations; the later pulses between 1000 ps and 1500 ps indicate other applied NM2 pulses depending on the particular realization. This is due to the fact that the random thermal field included in the simulations leads to slightly different trajectories of the magnetization dynamics in every particular realization. Nevertheless, all the realizations reliably reach the threshold of -0.9, at which we consider the memory cell switched, and reverse the z-component of the magnetization from +1 to -1.

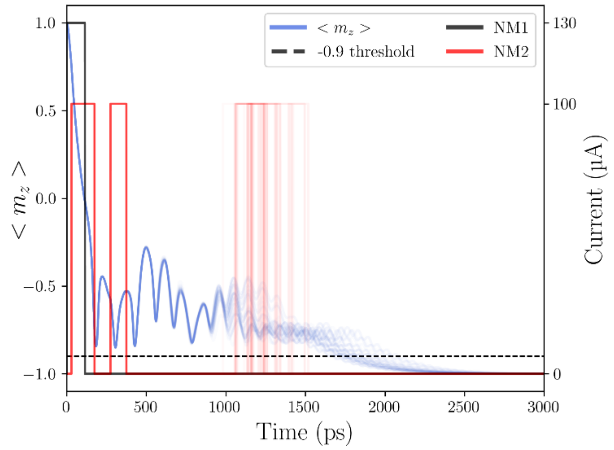
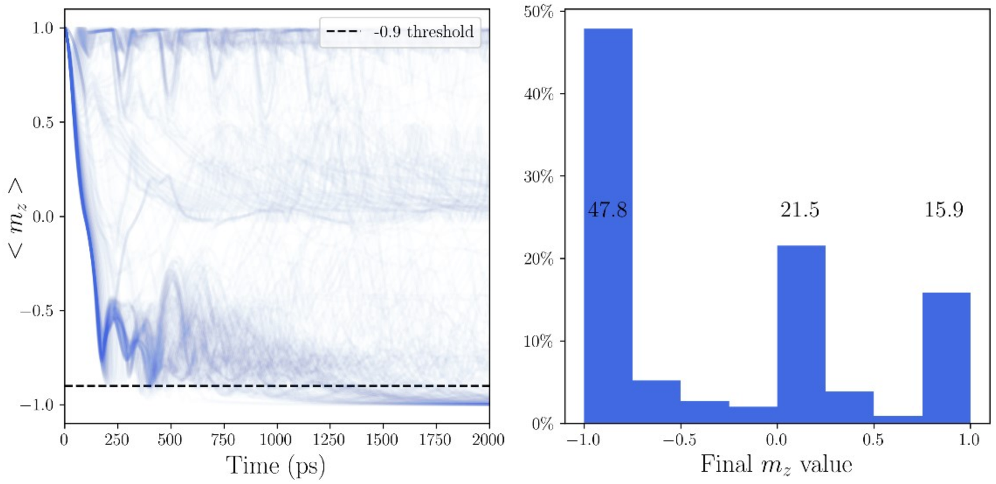


Figure 2. Trajectories of the z-component of the magnetization as well as applied NM1 and NM2 pulses of 50 realizations. Lines are plotted slightly transparent, such that regions with more overlapping plot lines appear more solid than regions with less overlapping lines.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



(a) (b)

Figure 3. (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) Histogram showing the percentage of realizations that reached a certain final m_z value.

To further investigate the performance of the trained model, switching simulations with varying material parameters were performed. The saturation magnetization M_S as well as the anisotropy constant K were varied individually by up to ±10%, which corresponds to the variability of MRAM manufacturing processes.[3] The resulting trajectories of the z-component of the magnetization are shown in Fig. 3(a). Each trajectory corresponds to a specific M_S-K combination. Again, the single trajectory lines were plotted slightly transparent, so the thickness of a line gives a qualitative measure of where most of the trajectories lie. Variation of the material parameters has a strong influence on the ability of the trained model to successfully reverse the magnetization. One can see that many trajectories can still be brought towards the target level of -1, but there are also material parameter combinations for which the z-component of the magnetization undulates close to the xy-plane or deviates only slightly from the initial position. Fig 3(b) gives a more quantitative view of the results and shows the percentage of trajectories which reach a certain final magnetization value. In 15.9% of the realizations, the magnetization hardly moves away from the initial state and stays between m_z values of 0.75 and 1. For 21.5% of the realizations, the trained model brings m_z closer to the target value, but the m_z values remain in a range between 0 and 0.25. The largest fraction of the trajectories, however, still can be brought close to the target value of -1: 47.8% of the realizations lie between -1 and -0.75.

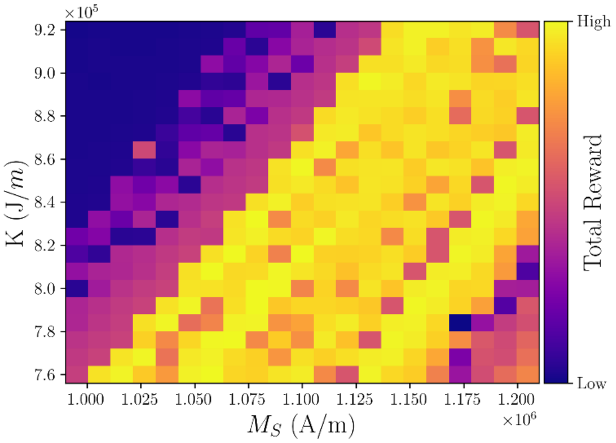
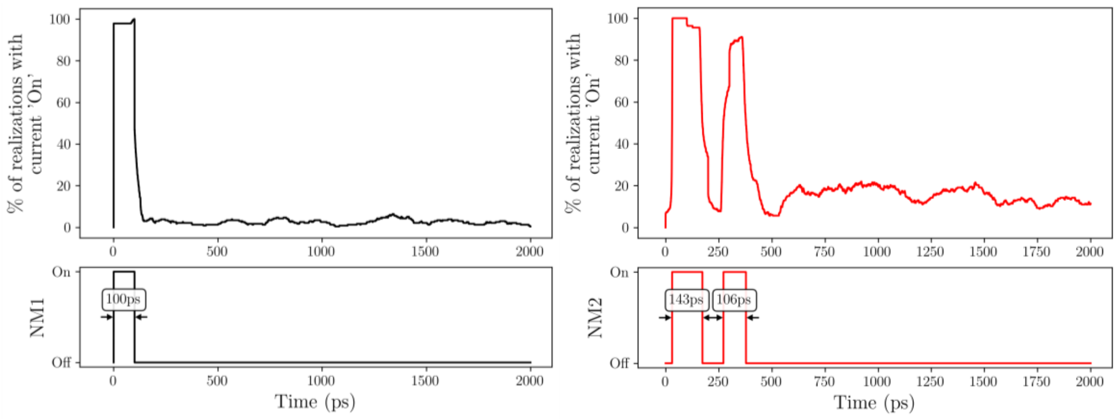


Figure 4. Accumulated reward achieved for anisotropy constant K and saturation magnetization M_S varied by ±10%. Results are shown for a total of 441 realizations.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

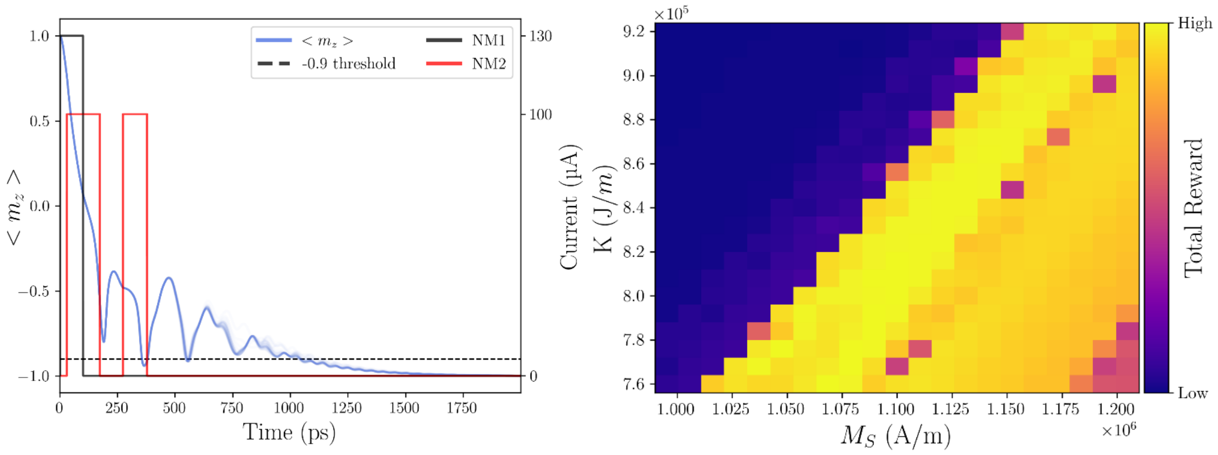


(a) (b)

Figure 5. The top panels show the percentage of the switching realizations which had the respective pulse turned on at the respective times of the simulation. The bottom panels show the derived static pulse sequences, in which pulses are turned on, if more than 50% of the realizations had the pulse turned on. Each shown for NM1 (a) and NM2 (b).

Fig. 4 highlights the switching performance of the model trained on a particular parameter set for different material parameter variations. A higher total reward correlates with the value of the z-component of the magnetization brought closer to the target value. It can be seen that a broad band of realizations from the bottom left to the top right performs well. The realizations in the top left corner of Fig. 4, however, perform worse and only achieve a low total reward. These runs correspond to the trajectories in Fig. 3(a), which remain close to the initial magnetization state. One can observe the general trend that the switching performance gets worse towards the top left corner of Fig. 4 and the trained model does not manage to accumulate a high total reward. Results published in Orio *et al.*[15] confirm this observation. In order to be able to switch the magnetization in this region of material parameter combinations, larger currents are required as the critical current for these material parameters is higher. Approximately 42% of the 441 realizations reach the threshold of -0.9.

The trained model reacts dynamically to changes in the behavior of the magnetization and applies additional pulses if needed, trying to bring the magnetization to -1. The practical usability of the neural network model is limited, as a static pulse sequence is required, not a dynamic one. We therefore took the 42% of the realizations, which were able to switch



(a) (b)

Figure 6. (a) Trajectories of the z-component of the magnetization as well as applied NM1 and NM2 pulses of 50 realizations. Lines are plotted slightly transparent, such that regions with more overlapping plot lines appear more solid than regions with less overlapping lines. (b) Accumulated reward achieved for anisotropy constant K and saturation magnetization M_S varied by ±10%. Results are shown for a total of 441 realizations.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

the memory cell under variation of the material parameters and analyzed how the pulses were applied by the neural network. These realizations were chosen to arrive at a static pulse sequence, which performs well under varying parameter conditions. For every time step of the 2 ns simulation time, we determined the percentage of realizations in which the respective pulses were active (top panels of Fig. 5). Based on these results, the static pulse was then constructed in such a way that it is turned on, when at least 50% of the realizations had the respective pulse ‘On’, and it is turned off, when 50% of the realizations had the respective pulse in the ‘Off’ state (bottom panels of Fig. 5). With this approach, a static pulse sequence was extracted with a single NM1 pulse of 100 ps length, which is turned on right in the beginning. The NM2 pulse is turned on two times, once just after the beginning of the switching simulation for a duration of 143 ps and a second time shortly after for a duration of 106 ps.

To confirm the choice of the static pulse sequence, 50 realizations with the extracted pulse sequence were performed to evaluate the reliability of switching with fixed material parameters (M_S=1.1×10^6 A/m, K=8.4×10^5 J/m^3). The results presented in Fig. 6(a), show that the magnetization is reversed reliably and deterministically. The switching threshold of -0.9 is reached after ~1000 ps in every realization. Fig. 6(b) shows the results of applying the static pulse sequence to the SOT-MRAM cell with varying parameters. The overall qualitative behavior is similar to the one shown in Fig. 4. However, the transition from good performing realizations to bad performing realizations in the top left corner is now more pronounced. This can be explained by the fact, that with the static pulse sequence no further pulses are applied, while the neural network model tries to switch the z-component by dynamically turning on more pulses if needed.

6. CONCLUSION

We presented an RL approach which allows to train a neural network such that it determines an optimal policy for applying pulses to an SOT-MRAM cell in order to reverse its magnetization reliably and as fast as possible. The results show, that after training the neural network on a fixed set of material parameters, the trained model is capable of reliably switching the z-component of the magnetization, even under material parameter variations. For the variations typical for the MRAM fabrication process, the switching performance remains reliable. For practical usability, a static pulse sequence was extracted from data of successfully switched simulations. This fixed pulse sequence achieves reliable and deterministic switching for fixed material parameters but also performs well over a wide parameter range.

ACKNOWLEDGEMENT

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 Research Association is gratefully acknowledged.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

REFERENCES

[1] Gupta, M., Perumkunnil, M., Garello, K., Rao, S., Yasin, F., Kar, G. S., *et al.*, "High-density SOT-MRAM technology and design specifications for the embedded domain at 5nm node," Proc. IEDM, 24.5.1–24.5.4 (2020). DOI: 10.1109/IEDM13553.2020.9372068

[2] Honjo, H., Nguyen, T. V. A., Watanabe., T., Nasuno, T., Zhang, C., Tanigawa, T., *et al.*, "First demonstration of field-free SOT-MRAM with 0.35 ns write speed and 70 thermal stability under 400°C thermal tolerance by canted SOT structure and its advanced patterning/SOT channel technology," Proc. IEDM, 28.5.1–28.5.4 (2019). DOI: 10.1109/IEDM19573.2019.8993443

[3] Garello, K., Yasin, F., Couet, S., Souriau, L., Swerts, J., Rao, S., *et al.*, "SOT-MRAM 300mm integration for low power and ultrafast embedded memories," Proc. IEEE Symp. VLSI Circuits, 81–82 (2018). DOI: 10.1109/VLSIC.2018.8502269

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

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

[6] Orio, R. L., Makarov, A., Selberherr, S., Goes, W., Ender, J., Fiorentini, S., Sverdlov, V., "Robust magnetic field-free switching of a perpendicularly magnetized free layer for SOT-MRAM," Solid-State Electron., 168, 107730 (2020). DOI: 10.1016/j.sse.2019.10773

[7] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., *et al.*, "Machine learning and the physical sciences," Rev. Mod. Phys. 91(4), 045002 (2019). DOI: 10.1103/RevModPhys.91.045002

[8] Sutton, R. S., and Barto, A. G., "Reinforcement learning: An introduction," Cambridge, MA, USA: MIT press, 1998.

[9] Fösel, T., Tighineanu, P., Weiss, T., and Marquardt, F., "Reinforcement learning with neural networks for quantum feedback," Phys. Rev. X 8(3), 031084 (2018). DOI: 10.1103/PhysRevX.8.031084

[10] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," Science 362(6419), 1140-1144 (2018). DOI: 10.1126/science.aar6404

[11] Orio, R. L., Ender, J., Fiorentini, S., Goes, W., Selberherr, S., and Sverdlov, V., "Optimization of a spin-orbit torque switching scheme based on micromagnetic simulations and reinforcement learning," Micromachines, 12(4), 443 (2021). DOI: 10.3390/mi12040443

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

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

[14] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., *et al.*, "Human-level control through deep reinforcement learning," Nature 518(7540), 529–533 (2015). DOI: 10.1038/nature14236

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

[16] Makarov, A., "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

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 09 Aug 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use