Microelectronics Reliability 126 (2021) 114231

Contents lists available at ScienceDirect

Microelectronics Reliability

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

Improving failure rates in pulsed SOT-MRAM switching by reinforcement learning

J. Ender [a,b,*], R.L. de Orio [b], S. Fiorentini [a], S. Selberherr [b], W. Goes [c], V. Sverdlov [a,b]

[a] Christian Doppler Laboratory for Nonvolatile Magnetoresistive Memory and Logic at the Institute for Microelectronics, 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

Finding and optimizing robust schemes for field-free switching remains a challenging problem in spin-orbit torque magnetoresistive random access memories. In this work reinforcement learning is employed for the optimization of switching schemes for such memory cells. A cell is switched purely electrically by applying pulses to two orthogonal metal wires. It is shown that a neural network model trained on a fixed material parameter set is suitable to determine optimal pulse sequences for reliable switching in the presence of thermal fluctuations, material parameter variations and reduction of the current to a sub-critical value. Multiple realizations of switching by means of simulation prove the reliability of magnetization reversal based on the pulse sequences found via reinforcement learning and show that the failure rate due to material parameter variations in these memory devices can be significantly reduced.

1. Introduction

Standard charge-based static random access memory (SRAM) cells are volatile by design and the progressive down-scaling of the CMOS technology utilized for their fabrication has led to an increase in standby power consumption. A possible solution to this problem is to use adequate 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 particularly suitable for caches, where currently CMOS-based SRAM is 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].

In this work we extend the previously published proof-of-concept [4], which showed that RL can find switching pulse sequences for an SOT-MRAM cell, but where some manual intervention was still necessary. We demonstrate that RL can be used to autonomously improve the switching efficiency of SOT-MRAM cells by learning how to apply pulses to achieve fast reversal of the magnetization in the memory cell. Most importantly, a model trained for a specific parameter set performs excellently on a broad distribution of varying materials and parameters and can even cope with a reduction of the switching current to below the critical value.

2. Spin-orbit torque memory

At the heart of MRAM devices lies a magnetic tunnel junction (MTJ), consisting of two ferromagnetic layers sandwiching a non-magnetic tunnel barrier. In SOT-MRAM devices, switching is achieved by passing a current through a heavy metal wire attached to the magnetic free layer (FL). The heavy metal wire exhibits a large spin Hall angle, which translates the charge current into a transverse spin current interacting with the ferromagnetic FL. In contrast to spin-transfer torque MRAM, the read and write paths are separated in SOT devices, leading to increased reliability, as no oxide degradation occurs in the MTJ and no accidental writing during a read operation can happen. This read-write-path separation also leads to a more energy-efficient operation. Although for the write current densities in the range of ~200 MA cm^{-2} high endurance is

* Corresponding author at: Christian Doppler Laboratory for Nonvolatile Magnetoresistive Memory and Logic at the Institute for Microelectronics, 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.2021.114231
Received 21 May 2021; Accepted 27 June 2021
Available online 11 October 2021

still achieved, to reduce the stress on the surrounding circuitry, further decreasing the required currents is essential [5,6]. An issue with perpendicular SOT-MRAM is the need for an external magnetic field to reliably switch the FL [7]. Besides solutions which circumvent this problem by breaking the mirror symmetry [8,9], a recently proposed scheme allows purely electrical switching by adding a second heavy metal wire orthogonal to the first one, only partially overlapping the FL (c.f. Fig. 1) [10]. The right part, consisting of NM1, FL and NM2, is responsible for the writing of the memory cell, while the left part deals with the read-out of the stored information. A sequence of current pulses through the two heavy metal wires, NM1 and NM2, is able to reverse the perpendicularly magnetized FL. It has been shown that the critical current, which is required to reliably reverse the magnetization in the memory cell, depends on the value of the anisotropy constant as well as on the saturation magnetization [11].

3. Reinforcement learning

The general reinforcement learning setup consists of an agent and an environment. The agent repeatedly interacts with the environment by performing certain actions, making the environment transition from one state to another. After every transition, the environment returns the new state, as well as a reward to the agent. The basis for the decision-making in so-called value-based learning algorithms, like Q-learning [2], is the action-value function, defined as

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

which describes the expected cumulative discounted reward for taking action a in state s at time t following a policy π, with γ being the discount factor that defines how strongly future rewards influence the estimate at time t. In the described experiments we employed the deep Q-network (DQN) algorithm [12], a version of the Q-learning algorithm using a neural network (NN) as function approximator. Due to the repeated interaction with the environment during a learning phase, the agent adjusts the weights of the neural network representing the action-value function to improve its approximation. Having an estimate of the quality of the state-action pairs, the agent can either make a greedy decision and take the action which promises the highest cumulative reward and exploit its current knowledge, 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. In order to make the best possible decision, the agent must have a good estimate of Eq. (1). Thus, during the learning phase, it is important to thoroughly explore the state-action space, such that as many state-action pairs as possible are represented in the Q-function approximation. This can be influenced by the exploration probability ε. Each time an action can be taken, an explorative, random action is taken with probability ε, and a greedy action is taken with probability 1-ε. As learning progresses, the initial value of ε is gradually reduced to a small value to allow the action-value function to converge.

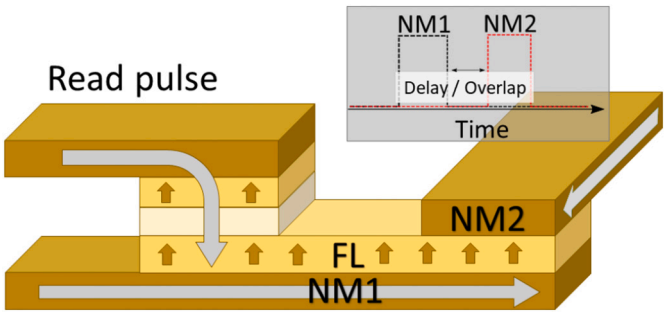

Fig. 1. SOT-MRAM cell for switching based on two orthogonal current pulses. The pulses are sent through the structure via two non-magnetic heavy metal wires, of which one is fully overlapping the FL (NM1) and one only partially (NM2).

4. RL for SOT switching

The general setup of the pulsed switching cell in an RL setting can be seen in Fig. 2. For the basic RL functionality an existing Python RL library was used [13]. The single components will be described in the following.

4.1. Agent

For the DQN agent, the implementation of [13] is used. To approximate the Q-function, the DQN algorithm uses a neural network. Apart from the parameters given in Table 1, the default configuration parameters were used, which empirically delivered the best results.

4.2. Environment

The environment contains a simulation of the two-pulse switching memory cell. For this purpose, an in-house developed simulator [14] was applied. This finite difference simulator solves the Landau-Lifshitz-Gilbert equation which describes the magnetization dynamics:

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

Here, $\mathbf{m}$ 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 $\mathbf{H}_{\text{eff}}$ includes the exchange field, the uniaxial perpendicular anisotropy field, the demagnetizing field, the current-induced field, and a stochastic thermal field at 300 K. The SOT, which acts on the memory cell and is generated by the currents through the NM1 and NM2 wires, is described by the latter two terms on the right-hand side. e is the elementary charge, $\hbar$ is the reduced Plank constant, θ_{SH} is the effective Hall angle, $j_{1,2}$ are the current densities in the two wires, d is the thickness of the FL and $\theta_{1,2}$ are functions describing when the pulses in NM1 and NM2 are active. $\mathbf{x}$ and $\mathbf{y}$ are the unit vectors pointing into the direction of the two heavy metal wires. The parameters used in the simulation are given in Table 2.

4.3. State

A crucial part for deciding which action to take depends on 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

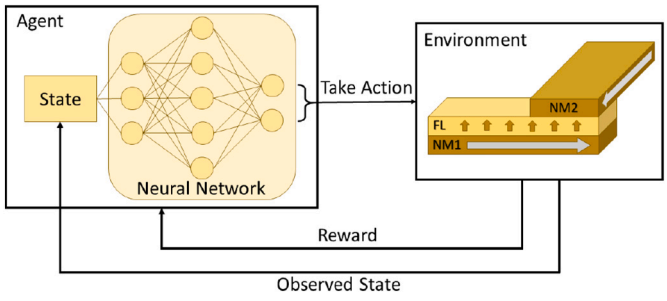

Fig. 2. General setup of the reinforcement learning approach: A simulation of the SOT-MRAM cell acts as environment which an agent interacts with to build up a policy based on a neural network.

Table 1
DQN parameters.

Parameter	Value
Size of NN layers	$11 \times 150 \times 100 \times 4$
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

Table 2
Simulation parameters. Assuming β-tungsten metal wires, CoFeB magnetic layers and an MgO tunnel barrier.

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	20 nm × 3 nm

information for the agent to make a decision. The state vector used for the experiments consists of 11 variables:

- The average x/y/z magnetization components,
- the average x/y/z effective field components,
- the difference of the magnetization's average x/y/z components to the previous time step, and
- two variables indicating whether the pulses are currently settable.

While the importance of the average magnetization components is apparent, as they are the state variables we ultimately want to change, it is also 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.

4.4. Actions

The action space of the RL agent is restricted to four actions, namely having both pulses off or both pulses on, as well as switching them on individually. The current value of the two pulses is fixed to 130 μA for the NM1 wire and 100 μA for the NM2 wire, and the minimum time between pulse state changes is 100 ps.

4.5. 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 has been 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:

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

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 towards the target value rapidly, in order to reduce the overall accumulated negative reward.

5. Results

By employing this RL approach the RL agent learns how to reverse the magnetization of an SOT-MRAM cell. For 10^6 training simulation steps, which correspond to 50 switching simulations, the agent refined its action-selection policy and was able to successfully reverse the magnetization. The trained model can subsequently be used to carry out switching simulations in which the model decides when to apply current pulses. To check the switching reliability of the best-performing neural network model found during the learning phase, 50 realizations under thermal fluctuations were subsequently performed with it. The results are shown in Fig. 3. The slight transparency of the single trajectories is intended to show paths that are taken more often and appear more solid, and those that are taken less often, which are only faintly visible. Up until 1 ns, the applied pulses as well as the trajectories of the z-component of the magnetization are basically identical for every realization. Only afterwards, when the thermal field leads to a slight divergence of the magnetization between the runs, the neural network model applies further NM2 pulses whose exact positions vary depending on the respective trajectory of the magnetization. Nevertheless, in all realizations the z-component of the magnetization is deterministically reversed from +1 to −1.

To further study the reliability of the learned model, experiments with varying material parameters were performed. The anisotropy constant K as well as the saturation magnetization M_S were varied individually up to ±5%. The pulses applied by the model and the trajectories of the z-component of the magnetization can be seen in Figs. 4 and 5, respectively. The two figures do not provide specific information about single switching simulations, but due to the slight transparency of the single plot lines, they deliver a good overview of the behavior of the learned neural network model. Compared to the results for fixed material parameters (Fig. 3), varying the material parameters also creates more variation in the applied pulses as well as in the magnetization trajectories. This indicates that the model indeed makes decisions which depend on the state of the system and does not simply apply a static set of pulses. In the simulated time window of 2 ns, ~75% of the trajectories are still successfully brought below the threshold of −0.9, at which we considered the cell to be switched. However, there are material parameter combinations for which the magnetization cannot even be brought below the xy-plane. 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

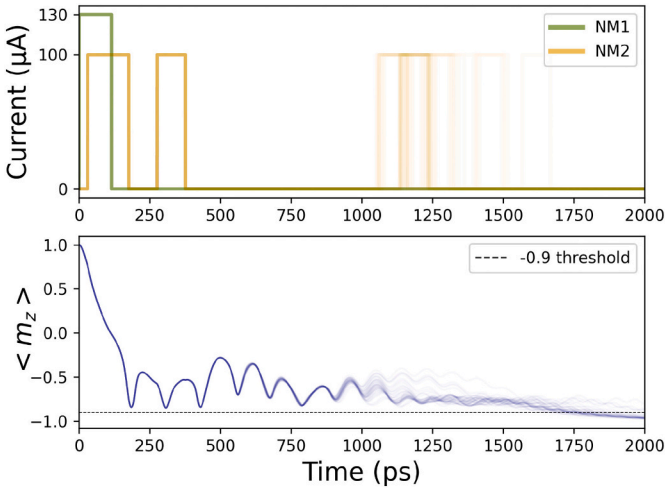

Fig. 3. Results of 50 realizations for fixed material parameters and an NM1 current value of 130 μA using the learned neural network model. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

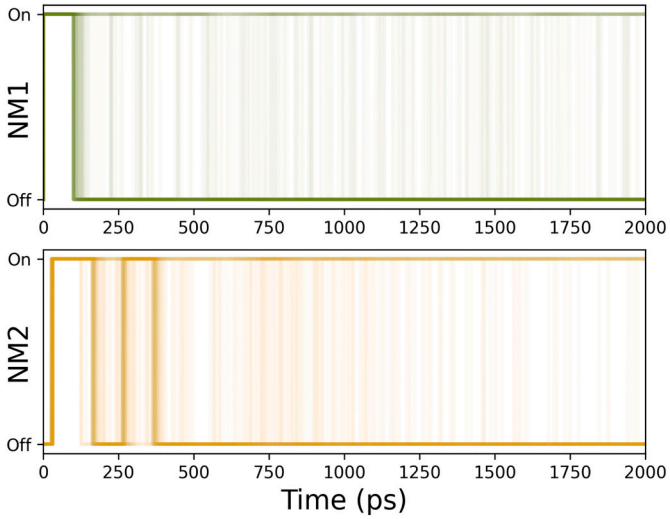

Fig. 4. Pulses applied to NM1 and NM2 during 121 realizations with varying material parameters and an NM1 current value of 130 μA. The results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

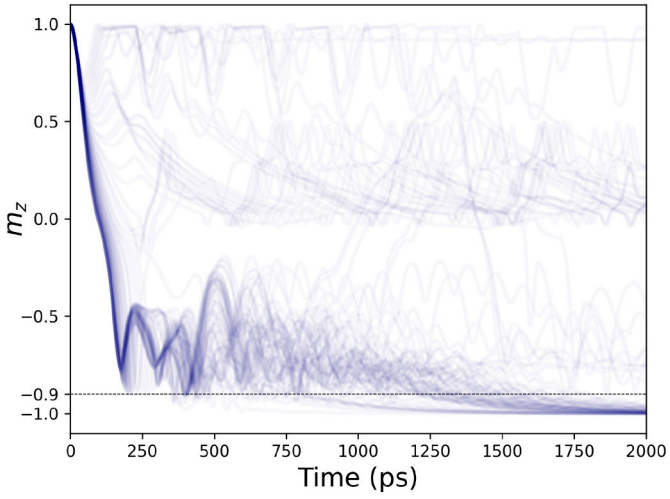

Fig. 5. Average z-component of the magnetization for 121 realizations with varying material parameters and an NM1 current value of 130 μA. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

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, however, is consistent with results published in [11], which indicate that in this range of the two material parameters, a higher current is required to deterministically switch the memory cell. Seeing how good the switching performance across this wide range of parameter variations is, further experiments were performed with a reduction of the NM1 current to 110 μA, which lies below the critical value of 120 μA [11]. First, again 50 realizations under the influence of a thermal field and with fixed material parameters were carried out, resulting in the trajectories shown in Fig. 7. Interestingly, setting the current value of the NM1 wire to below the critical one, it seems to be easier for the model to reverse the magnetization. Due to the reduced slope of the decreasing z-component of the magnetization, the NM1 pulse and the first NM2 pulse are kept on slightly longer. After the

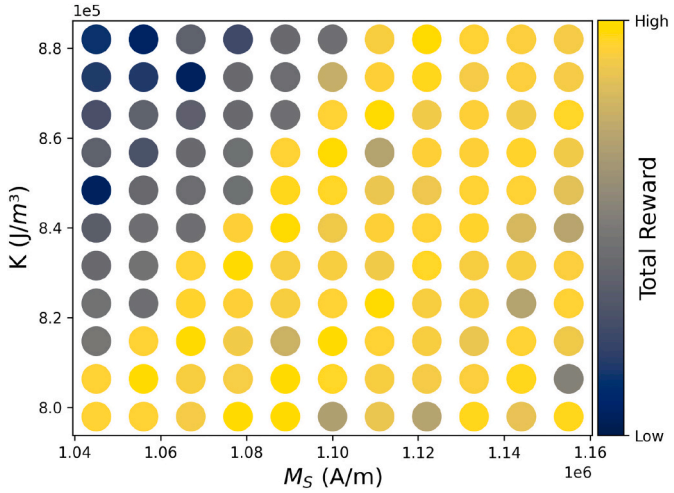

Fig. 6. Accumulated reward achieved for anisotropy constant K and saturation magnetization M_S varied by ±5% and an NM1 current value of 130 μA. Results are shown for a total of 121 realizations.

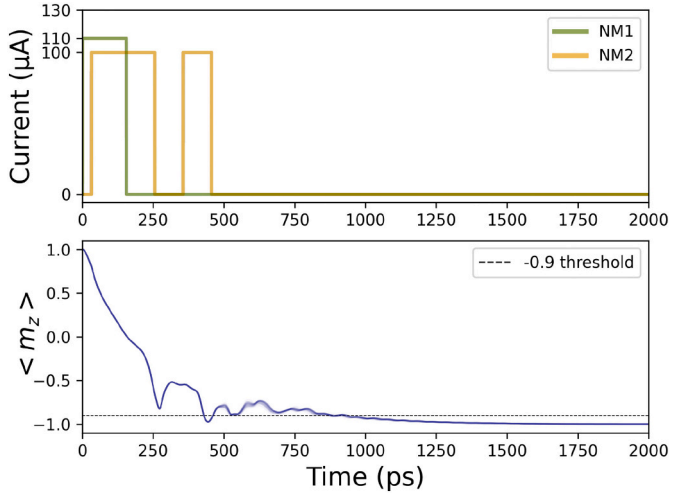

Fig. 7. Results of 50 realizations for fixed material parameters and an NM1 current value of 110 μA using the learned neural network model. Results of the single runs are plotted slightly transparent, such that regions where multiple lines overlap appear more solid.

initial two pulses on the NM2 wire, no further pulses are required. Looking at the magnetization trajectories, one can see why no further NM2 pulses were needed. There is less variation between the realizations and the −0.9 threshold is reached ~800 ps earlier than with the higher current. Again, also for this reduced-current scenario, the model trained on fixed material parameters was confronted with the variations of the anisotropy constant and the saturation magnetization of ±5%. The overview of the accumulated reward is presented in Fig. 8. The line separating the higher-performing runs from the lower-performing ones has shifted slightly towards the bottom right corner. The model though is still capable of reversing the magnetization in a large portion of the parameter variation space and the number of trajectories with successful switching has only reduced to ~59%.

6. Conclusion

We demonstrated that reinforcement learning is a promising technique to guarantee reliable switching of SOT-MRAM cells. An optimal pulse scheme for deterministic switching in the presence of thermal

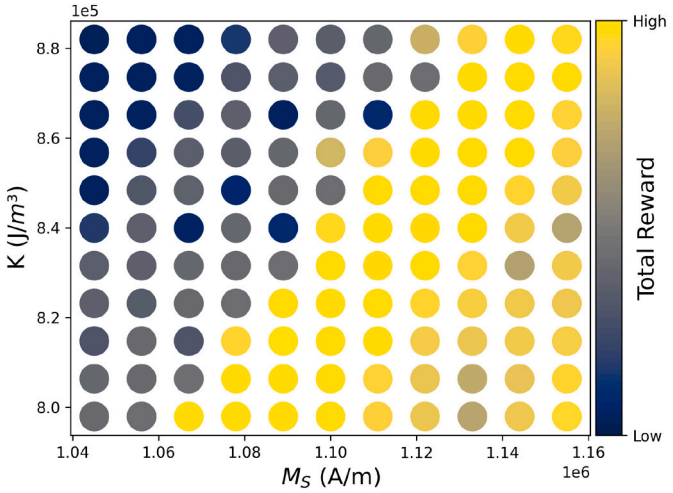

Fig. 8. Accumulated reward achieved for anisotropy constant K and saturation magnetization M_S varied by ±5% and an NM1 current value of 110 µA. Results are shown for a total of 121 realizations.

fluctuations and parameter variations is achieved after training the neural network model to maximize its received reward during the learning phase for a fixed material parameter set. Using the trained model afterwards to perform simulations, we could not only show that the model is flexible and can cope with varying material parameters, but as well deal with sub-critical current values. However, a further reduction of the switching current is still desirable to reduce stress in the overall circuitry of the memory cells.

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.

Acknowledgements

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.

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. 91 (4) (Dec. 2019), 045002, https://doi.org/10.1103/RevModPhys.91.045002.
[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] 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.
[5] K. Garello, F. Yasin, S. Couet, L. Souriau, J. Swerts, S. Rao, SOT-MRAM 300mm integration for low power and ultrafast embedded memories, in: Proc. IEEE Symp. VLSI Circuits, Oct. 2018, pp. 81–82, https://doi.org/10.1109/VLSIC.2018.8502269.
[6] M. Gupta, M. Perumkunnil, K. Garello, S. Rao, F. Yasin, G.S. Kar, High-density SOT-MRAM technology and design specifications for the embedded domain at 5nm node, in: Proc. of the IEDM, Dec. 2020, https://doi.org/10.1109/IEDM13553.2020.9372068, pp. 24.5.2-24.5.4.
[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] H. Wu, S.A. Razavi, Q. Shao, X. Li, K.L. Wong, Y. Liu, et al., Spin-orbit torque from a ferromagnetic metal, Phys. Rev. B 99 (May 2019), 184403, https://doi.org/10.1103/PhysRevB.99.184403.
[10] 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.
[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 on 12 May 2021).
[14] A. Makarov, Modeling of Emerging Resistive Switching Based Memory Cells, Institute for Microelectronics, TU Wien, Vienna, 2014, https://doi.org/10.13140/RG.2.2.11456.74242. Ph.D. Thesis.