Advances in modeling emerging magnetoresistive random access memories: from finite element methods to machine learning approaches

Ender, Johannes, Fiorentini, Simone, L. de Orio, Roberto, Hadámek, Tomáš, Bendra, Mario, et al.

Johannes Ender, Simone Fiorentini, Roberto L. de Orio, Tomáš Hadámek, Mario Bendra, Wolfgang Goes, Siegfried Selberherr, Viktor Sverdlov, "Advances in modeling emerging magnetoresistive random access memories: from finite element methods to machine learning approaches," Proc. SPIE 12157, International Conference on Micro- and Nano-Electronics 2021, 1215708 (30 January 2022); doi: 10.1117/12.2624595

SPIE.

Event: International Conference on Micro- and Nano-Electronics 2021, 2021, Zvenigorod, Russian Federation

Advances in Modeling Emerging Magnetoresistive Random Access Memories: From Finite Element Methods to Machine Learning Approaches

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

ABSTRACT

Emerging spin transfer torque magnetoresistive random access memories (STT MRAM) are nonvolatile and offer high speed and endurance. They are promising for stand-alone and embedded applications in the automotive industry, microcontrollers, Internet of Things, frame buffer memory, and slow SRAM. The MRAM cell usually includes a CoFeB fixed reference layer and a free magnetic layer (FL), separated by a tunnel barrier. To design ultra-scaled MRAM cells it is necessary to accurately model the torques acting on the magnetization in composite magnetic layers with one or several nonmagnetic inclusions between the ferromagnetic parts. The magnetization dynamics is governed by the Landau-Lifshitz-Gilbert (LLG) equation supplemented with the corresponding torques. The torques depend on nonequilibrium spin accumulation generated by an electric current. The electric current and the spin accumulation also depend on the magnetization. Therefore, the LLG and the spin-charge transport equations are coupled and must be solved simultaneously.

We apply the finite element method (FEM) to numerically solve this coupled system of partial differential equations. To develop an open source solver, we use well-developed C++ FEM libraries. The computationally most expensive part is the demagnetizing field calculation. It is performed by a hybrid finite element-boundary element method. This confines the simulation domain for the field evaluation to the ferromagnets only. Advanced compression algorithms for large, dense matrices are used to optimize the performance of the demagnetizing field calculation in complex structures. To evaluate the torques acting on the magnetization, a coupled spin and charge transport approach is implemented. For the computation of the torques acting in a magnetic tunnel junction (MTJ), a magnetization-dependent resistivity of the tunnel barrier is introduced. A fully three-dimensional solution of the equations is performed to accurately model the torques acting on the magnetization. The use of a unique set of equations for the whole memory cell including the FL, fixed layer, contacts, and nonmagnetic spacers is one of the advantages of our approach. To incorporate the temperature increase due to the electric current, we solve the heat transport equation coupled to the electron, spin, and magnetization dynamics, and we demonstrate that the FL temperature is highly inhomogeneous due to a non-uniform magnetization of the FL during switching.

Spin-orbit torque (SOT) MRAM is fast-switching and thus well suitable for caches. By means of micromagnetic simulations, we demonstrate the purely electrical switching of a perpendicular FL by the SOTs due to two orthogonal short current pulses. To further optimize the pulse sequence, a machine learning approach based on reinforcement learning is employed. Importantly, a neural network trained on a fixed material parameter set achieves switching for a wide range of material parameter variations.

Keywords: magnetoresistive random access memories, spin-transfer torque MRAM, spin-orbit torque MRAM

1. INTRODUCTION

A viable option to mitigate the unfavorable trend of increasing power dissipation in modern integrated circuits is to introduce nonvolatile elements capable of preserving the state of a variable used in computations, even when the supply

* http://www.iue.tuwien.ac.at, Email: (Ender|Fiorentini|Orio|Hadamek|Bendra|Selberherr|Sverdlov}@iue.tuwien.ac.at

International Conference on Micro- and Nano-Electronics 2021, edited by Vladimir F. Lukichev, Konstanten V. Rudenko, Proc. of SPIE Vol. 12157, 1215708 · © 2022 SPIE · 0277-786X · doi: 10.1117/12.2624595

power is turned off. Nonvolatility is crucial for eliminating the leakage power dissipation. In addition, nonvolatility allows to make circuits with almost zero power consumption in a stand-by mode. Apart from stand-alone applications, e.g. critical program and data storage devices, it is extremely attractive to use nonvolatility in the main computer memory as replacement of conventional volatile CMOS-based dynamic random-access memory (DRAM) [1], [2] which requires refreshing of the charge to maintain the gradually dissipating stored information. Embedding nonvolatile memory on a chip close to a processing unit potentially removes the data exchange bottleneck of the traditional von Neumann architecture between the central processing unit (CPU) and the distant memory. In modern multicore processors, an extensive energy consumption takes place in the hierarchical multi-level cache memory structure. Replacing the caches with nonvolatile memory reduces the power consumption and optimizes the memory cell size compared to static random-access memory (SRAM) [1], [2].

The development of a novel electrically addressable nonvolatile memory is essential to mitigate the increase of the stand-by power and to introduce instant-on architectures. To be competitive with the traditional CMOS-based memory and nonvolatile flash, emerging nonvolatile memories must offer fast switching, high integration density, long retention, high endurance, and low power consumption. At the same time, they must possess a simple structure to reduce fabrication costs. An important requirement is that emerging memories must be compatible with CMOS processes to benefit from the advantages provided by the well-developed CMOS fabrication technology.

Applications driven by magnetic moments and fields played an important role in magnetic hard disk drives. However, the coupling between the electric currents and magnetic moments is weak, which results in low efficiency and high energy supply costs. An efficient coupling between the electrical and the magnetic degree of freedom was discovered at a quantum mechanical level as a phenomenon called the giant magnetoresistance effect. This facilitated a reliable, purely electrical read operation of the information encoded by the magnetization orientation. Based on this principle, hard disk drives with extremely high density appeared on the market. The next generation of magnetic storage devices with even higher density and without moving parts was based on the unique properties of magnetic tunnel junctions (MTJs). The tunneling current through the MTJ structure strongly depends on the relative polarization of the ferromagnetic contacts and the tunneling magnetoresistance ratio (TMR) can reach several hundred percent at room temperature [3]. Therefore, using MTJs as data storage elements provides an efficient way of converting the spin (magnetization) degree of freedom into a charge (current) employed by CMOS devices. Memories employing the dependence of the resistance on the relative magnetization orientations are called magnetoresistive random access memories (MRAM). As their MTJ resistances are similar to the resistances of MOSFETs, MRAM is electrically compatible with CMOS devices without intermediate circuitry needed.

To write the data in a MRAM, a way to convert the charge information into magnetic moments orientation is required. The employment of magnetic fields created by electrical currents in coils is highly inefficient. In addition, although fast switching of a specially designed free magnetic layer has been achieved by applying carefully engineered orthogonal current pulses [4], MRAM with field induced switching cannot be scaled below the 90nm technology node [5]. In contrast, the spin-transfer torque effect (STT) [6], [7] has been proven to be particularly useful for purely electrical magnetic moment manipulation. The torque is generated by passing a current through a MTJ. When electrons travel through the fixed ferromagnetic layer, their spins become aligned with the magnetization. When these spin-polarized electrons enter the free magnetic layer, they become aligned with the magnetization of the free layer (FL) within a transition layer of a few angstroms. The change in the spin of these electrons is transferred to the FL magnetization due to the conservation of the momentum. Therefore, the spin-polarized current exerts a torque on the magnetization of the FL. If the current is sufficiently strong, this torque causes magnetization switching. By inverting the current direction the sign of the torque and the magnetization switching direction is also inverted.

STT MRAM is considered a perfect candidate for future universal memory applications. STT MRAM is fast (10ns), possesses high endurance (10^{12}), and has a simple structure. It is compatible with CMOS technology and can be straightforwardly embedded in circuits [8]. It is particularly promising to employ nonvolatility in Internet of Things (IoT) and automotive applications, as well as a replacement of conventional volatile CMOS-based DRAM and nonvolatile flash memory. High-density STT MRAM arrays with 4Gbit capacities have been already demonstrated [9]. For embedded applications, a successful implementation of 8 Mb 1T-1MTJ STT MRAM on a 28 nm CMOS logic platform [10] was demonstrated. Recently, 128 Mb embedded MRAM with 14 ns write speed was reported [11]. An embedded MRAM solution compatible with 22FFL FinFET technology is available [12].

Regardless of the first commercial STT MRAM-based products for stand-alone applications being available, one critical aspect of the currently used STT MRAM technology is a relatively high switching current. The problem of high

switching current and large active writing energy may jeopardize the advantages provided by nonvolatility, such as zero stand-by power, no data refreshment, and no data recovery. At the same time, the energy barrier separating the two stable memory states must be large, at least sixty-eight times the thermal energy [8], to negate the errors due to erroneous thermally agitated magnetization switching in big (~1 Gb) memory arrays at long (~10 years) retention time. Since the same barrier must be overcome for memory writing, a careful MRAM cell optimization to maintain the desired balance between the long retention and low switching currents as well as fast and reliable cell information reading without destroying the data is needed.

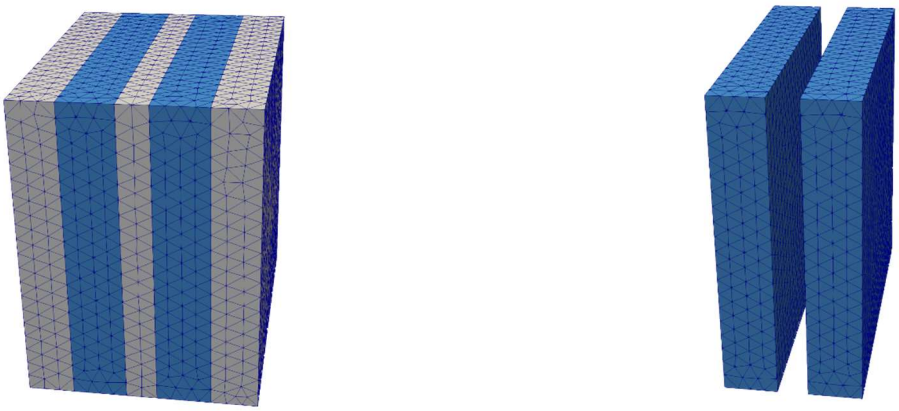
In this review, we briefly describe our finite element-method (FEM) based modeling and simulation approach aiming at facilitating the development and fabrication of modern and emerging micro- and nanoelectronic magnetoresistive devices. The key simulation capabilities to evaluate efficiently (i) the effective field (including the demagnetization and stray field contributions in a complex disconnected geometry), (ii) coupled three-dimensional charge and spin transport, and (iii) the corresponding spin-transfer torques driving (iv) the magnetization dynamics and switching required for modeling MRAM are being developed and implemented in a prototype of a parallel simulation environment devised in C++, which utilizes well supported FEM libraries under open-source licenses.

Although the use of STT MRAM in last-level caches is conceivable [13], the switching current density for operating in high-level caches at sub-nanosecond speed becomes so large that it can damage the tunnel barrier. A new technological solution is thus required to use MRAM in L1 and L2 caches. Among the newly discovered physical phenomena suitable for next-generation MRAM is the spin-orbit torque (SOT) assisted switching of a free ferromagnetic layer in a MTJ grown on top of a heavy metal [13], [14], [15]. The SOT acting on the free magnetic layer is generated by the current passing through the heavy metal material due to the spin Hall effect. The switching current is injected along the heavy metal/ferromagnet and does not flow through the MTJ. In order to read the information, a much smaller read current is applied through the MTJ. Therefore, a SOT MRAM cell is a three-terminal device, where the read and write current paths are decoupled. To reduce the cell size and to increase the integration density, structures with large perpendicular magnetic anisotropy are required. As IMEC recently demonstrated the feasibility of a perpendicular CoFeB/MgO/CoFeB stack grown on beta-tungsten on a 30 mm CMOS wafer by only CMOS compatible processes [16], there is a cautious confidence that fast, low-power nonvolatile magnetoresistive memory suitable for processor caches will be available soon.

Despite the undisputable progress in developing SOT MRAM, there is one important issue which has not been convincingly resolved yet. Namely, a static in-plane magnetic field is required to guarantee deterministic switching [17] of a perpendicularly magnetized FL. Several paths to achieve deterministic switching without magnetic fields are based on the mirror symmetry breaking of the magnetic cell with respect to the plane defined by the vectors of the magnetization and the current direction. Biasing the FL by employing an exchange coupling to an antiferromagnet [18] as well as the use of peculiar tapered sample shapes [19] were proposed as methods to break the mirror symmetry, however, these approaches complicate the large-scale integration.

Recently, a dynamic approach to deterministic switching of a perpendicularly magnetized layer was proposed [20]. An application of two perpendicular current pulses results in fast, reliable, and magnetic field free perpendicular magnetization reversal of a rectangular FL. While the first current pulse forces the magnetization in-plane perpendicular to the long rectangle edge, the second pulse deviates the magnetization towards one of the elongated rectangle edges. The magnetization experiences the shape anisotropy field which plays the role of a dynamically generated external field and makes the switching deterministic. The speed of switching is faster when the second current is applied to a part of the FL. In this case, the FL part subject to the SOT of the second current is quickly rotated in-plane parallel to the long edges of the rectangle. The stray field generated from this part makes the rest of the FL rotate about it, moving the magnetization deterministically from its in-plane orientation, thus completing the switching. We demonstrate that this dynamic scheme does neither require a perfect current pulse synchronization nor precise patterning of the second current line [21].

To optimize the pulse sequence for a faster switching, a machine learning approach based on reinforcement learning is employed. Initially, a neural network is trained on a fixed material parameter set. Importantly, this trained neural network is able to switch the memory cell over a broad material parameter range variation around the values at which it was trained.



(a) (b)

Figure 1. (a) STT MRAM cell. Nonmagnetic contacts and tunnel barrier are shown in gray, ferromagnetic layers are blue. (b) Calculation of the demagnetizing and stray fields is performed only on the magnetic domain.

2. FINITE ELEMENT MODELING OF SPIN-TRANSFER TORQUE MRAM

A typical spin transfer torque (STT) MRAM cell is shown in Fig.1a. It consists of two ferromagnetic layers separated by a thin tunnel barrier. The ferromagnetic layers are connected to nonmagnetic electrodes delivering the electric current which runs through the structure. The magnetization of one ferromagnetic layer is fixed. This layer serves as a spin polarizer, as the electric current passing through it carries a non-zero spin along its magnetization orientation. After passing through the tunnel barrier the spin-polarized current impinges the second ferromagnetic layer. The magnetization of this FL can be altered by the spin-transfer torque. The magnetization dynamics in the FL of a MRAM cell is described by the solution of the Landau-Lifshitz-Gilbert (LLG) equation

$$\frac{\partial \mathbf{m}}{\partial t} = -\gamma \mu_0 \mathbf{m} \times \mathbf{H}_{\text{eff}} + \alpha \mathbf{m} \times \frac{\partial \mathbf{m}}{\partial t} + \frac{1}{M_s} \mathbf{T}_S, \tag{1}$$

supplemented with the spin-transfer torque $\mathbf{T}_S$ exerted by a spin-polarized current flowing through the memory cell. $\mathbf{m}=\mathbf{M}/M_S$ is the position-dependent magnetization normalized by the saturation magnetization M_S, γ is the gyromagnetic ratio, α is the Gilbert damping constant, $\mathbf{H}_{\text{eff}}$ is the effective magnetic field which includes the external field contribution, the magnetic anisotropy field, the exchange field, the Ampere field as well as the demagnetizing field which includes the magnetostatic coupling between the free and the fixed layers.

2.1 Demagnetizing field calculations

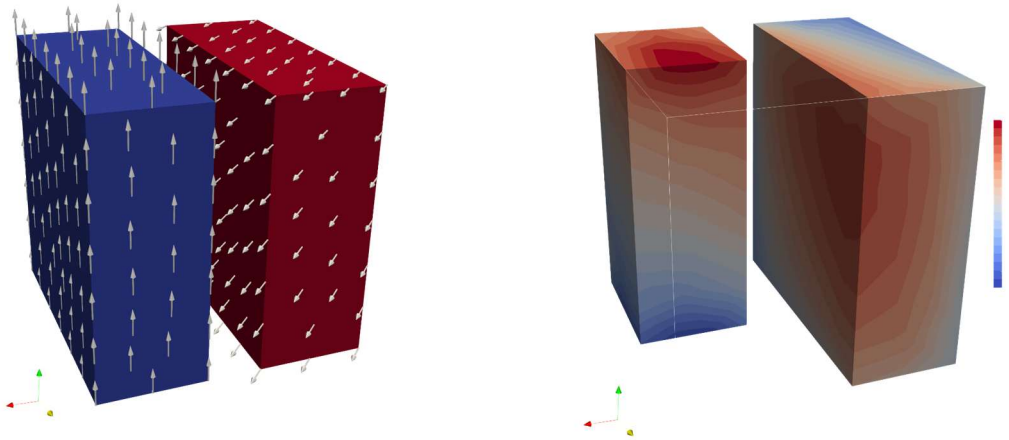
In a finite element method discretization (FEM) of the LLG equation, the computationally most demanding contribution is the demagnetizing field $\mathbf{H}_d$. As this field arises due to the long-range dipole interaction of the magnetic moments, it is conveniently described by a gradient of the scalar magnetic potential u,

$$\mathbf{H}_d = -\nabla u\,, \tag{2}$$

with u being the solution of the following Poisson equation:

$$\nabla^2 u = \nabla \cdot \mathbf{M}. \tag{3}$$

As the potential $u(\mathbf{x})$ slowly decays to zero as $|\mathbf{x}| \to \infty$, a large computational domain surrounding the magnetic material is needed in a plain FEM simulation to accurately describe this behavior. The truncation of the external computational domain surrounding the magnetic material at a certain distance is usually called truncation approach.



(a) (b)

Figure 2. (a) Magnetization of the fixed (red) and free (blue) layers. (b) Magnetic potential for disconnected geometry.

To achieve high accuracy and reduce the computational costs, a hybrid approach, where the finite element method is coupled to the boundary element method (hybrid FEM-BEM), is employed [22]. Here, the potential is split into two parts u_1 and u_2, calculated by solving a Poisson and a Laplace equation with appropriate boundary conditions. The potentials u_1 and u_2 are related by the integral over the boundary $\partial\Omega$ of the magnetic domain [22]

$$u_2 = \int_{\partial\Omega} u_1 \frac{\partial}{\partial \mathbf{n}} \frac{1}{|\mathbf{x}-\mathbf{x}'|} d^2\mathbf{x}\,, \tag{4}$$

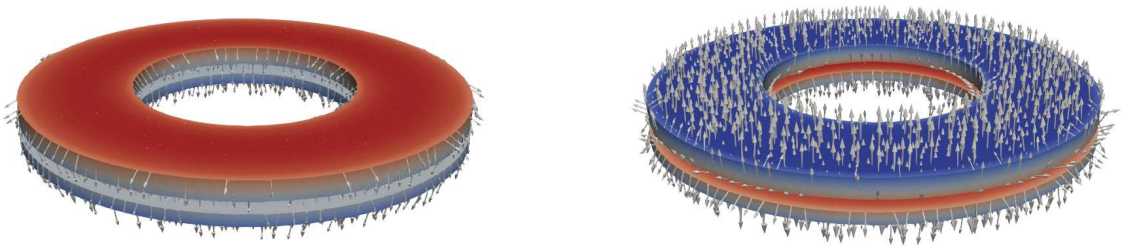
which in its discretized form reduces to the following matrix-vector multiplication:

$$u_2 = \mathbf{B} \cdot u_1 \tag{5}$$

These equations allow for the evaluation of the demagnetizing field to be performed only on the magnetic domain (Fig.1b). However, despite the benefit of the hybrid FEM-BEM method to confine the simulation domain to the ferromagnetic subdomains, it comes at the cost of dealing with dense matrices, arising from the discretization of the boundary operator **B** in (5). To reduce the memory demands, matrix compression algorithms can be applied [23]. The use of such compression algorithms considerably reduces the memory requirements of the simulation.

The demagnetizing field calculations are integrated into a micromagnetic simulation environment written in C++, which uses the finite element method to solve (1) in STT MRAM devices. To enhance the performance and accuracy, the open-source third party dependencies of the tool used are the FEM library MFEM [24] and H2Lib [25]. The latter library contains matrix compression algorithms and provides basic BEM functionality. Using a set of micromagnetic standard problems [26], the correctness of the results for the demagnetizing field calculations was verified. We confirm that the hybrid FEM-BEM approach is more accurate compared to the brute-force truncation approach.

Simulating magnetization dynamics in a STT MRAM cell requires the calculation of the demagnetizing field of complex magnetic structures. These consist of several disconnected magnetic layers whose stray fields act on each other. An example of the evaluation of the scalar magnetic potential u in a MTJ with non-collinear ferromagnetic layers (Fig.2a) is shown in Fig.2b. Fig.3 illustrates the demagnetizing field for a doughnut-like MTJ with a hole in the center for the parallel (Fig.3a) and the anti-parallel (Fig.3b) relative magnetization orientation. It is important that the demagnetizing field calculations are restricted to the magnetic sub-domains only.



(a) (b)

Figure 3. Demagnetizing field in a MTJ of a doughnut shape. The magnetization of the ferromagnetic layers is (a) parallel and (b) anti-parallel.

3. EVALUATING SPIN-TRANSFER TORQUE

The spin-transfer torque entering the LLG equation (1) is expressed via the nonequilibrium spin accumulation **S** as [27], [28]

$$\mathbf{T_S} = -\frac{D_e}{\lambda_J^2}\mathbf{m}\times\mathbf{S} - \frac{D_e}{\lambda_\varphi^2}\mathbf{m}\times(\mathbf{m}\times\mathbf{S}), \tag{6}$$

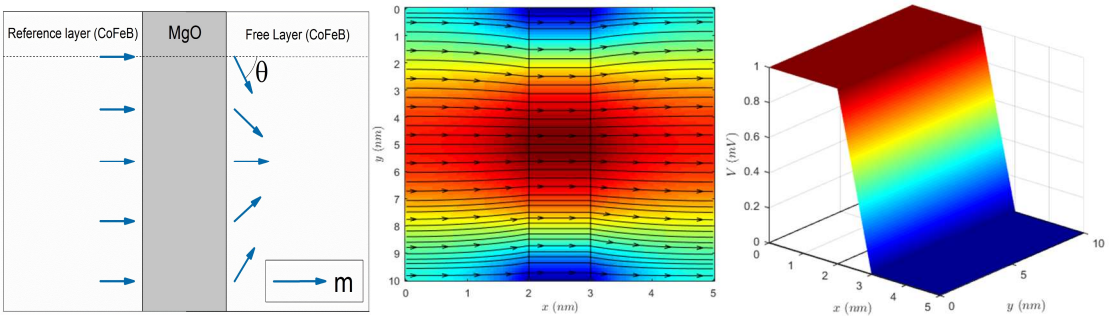
where λ_J is the exchange length, λ_φ is the spin dephasing length, and D_e is the electron diffusion constant in the ferromagnetic layers.

The spin accumulation **S** is induced by an electrical current running through the structure. In order to obtain **S**, coupled spin and charge transport must be resolved. The main element of a STT MRAM cell is a MTJ representing a sandwich of two metallic ferromagnets separated by a tunnel barrier. The calculation of the respective spin accumulation requires modeling the transport through magnetic valves which can also contain tunnel barriers. The tunnel barrier is essential to achieve a high tunneling magnetoresistance ratio (TMR), related to the large difference in the resistances R_P and R_{AP} in the parallel/anti-parallel MTJ configuration.

$$\mathrm{TMR} = \frac{R_{AP}\text{-}R_P}{R_P} \tag{7}$$

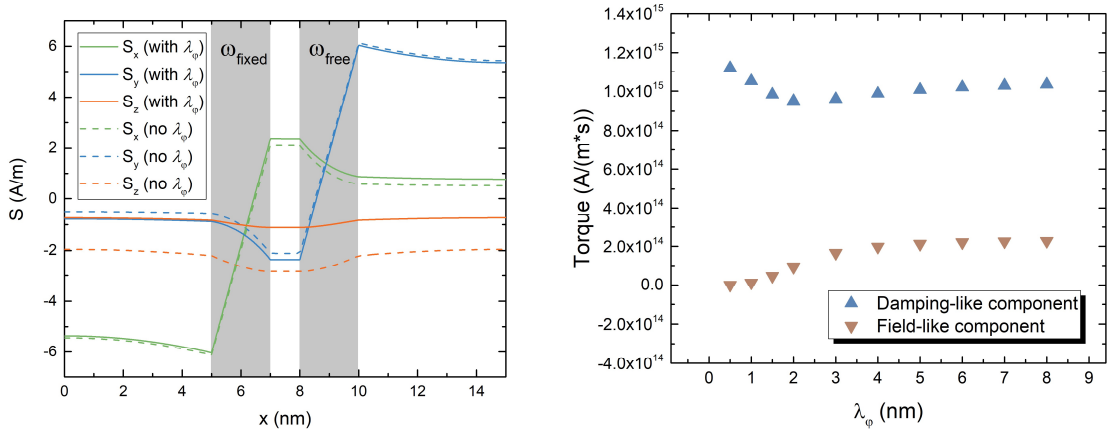
The resistance of the tunnel barrier is much larger than that of the contacts. The tunnel resistance defines the current through the MTJ. To compute the current through the structure, the tunnel barrier can be modeled as a poor conductor whose low conductivity depends on the relative orientation of the magnetization in the ferromagnetic layers as

$$\sigma(\theta) = \sigma_0\left(1+\left(\frac{\mathrm{TMR}}{2+\mathrm{TMR}}\right)\cos(\theta)\right), \tag{8}$$



(a) (b) (c)

Figure 4. (a) Nonuniform magnetization in a FL of a MTJ. (b) The current density distribution. More current flows through the middle of the structure due to the highest conductivity (8) for parallel magnetization orientation. (c) Potential drop at the tunnel barrier.



(a) (b)

Figure 5. Spin accumulation components in a MTJ with the magnetization of the left fixed layer along the X-axis and the right FL along the Y-axis. Dashed lines show results without spin dephasing. (b) Damping- and field-like torques' dependencies on the spin dephasing length.

where σ_0 is the average conductivity value between the conductivities in the parallel (P) and antiparallel (AP) state, and θ is the local angle between the magnetic vectors in the free and reference layer. For non-uniform relative magnetization orientation (Fig.4a), characteristic to switching, the resistance and the current through the MTJ depend strongly on the position. This results in a highly non-uniform current density (Fig.4b), provided the voltage at the contacts is fixed (Fig.4c).

Obtaining a good approximation of the current density $\boldsymbol{J}_C$ is not sufficient to model spin transport in MTJs. We could assume that the spin accumulation is preserved across the barrier in the absence of spin flips. Once the current $\boldsymbol{J}_C$ is known, the spin accumulation and the spin current density $\boldsymbol{J}_S$ are found as [28], [29]

$$\mathbf{J_S} = \frac{\mu_B}{e}\beta_\sigma\left(-\mathbf{J_C} + \beta_D D_S \frac{e}{\mu_B}[(\boldsymbol{\nabla}\mathbf{S})\mathbf{m}]\right) \otimes \mathbf{m} - D_S\boldsymbol{\nabla}\mathbf{S}, \tag{9a}$$

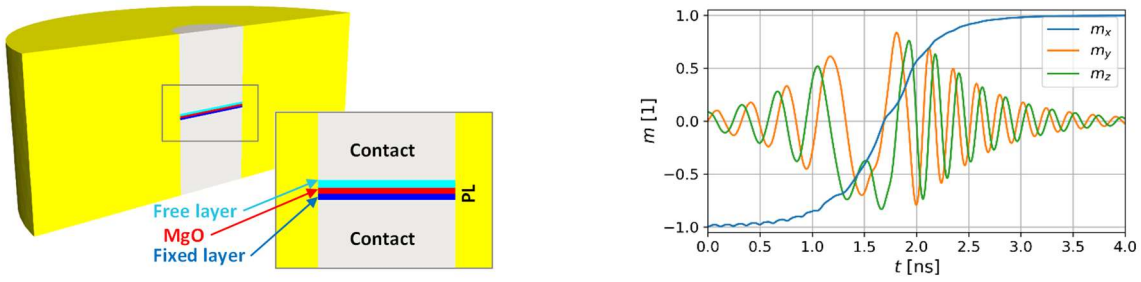
$$\frac{\partial \mathbf{S}}{\partial t} = -\boldsymbol{\nabla}\mathbf{J_S} - D_S\left(\frac{\mathbf{S}}{\lambda_{sf}^2} + \frac{\mathbf{S}\times\mathbf{m}}{\lambda_j^2} + \frac{\mathbf{m}\times(\mathbf{S}\times\mathbf{m})}{\lambda_\varphi^2}\right), \tag{9b}$$

where μ_B is the Bohr magneton, e is the electron charge, β_σ and β_D are polarization parameters, with D_S we indicate the spin diffusion constant in the barrier region, λ_{sf} is the spin-flip length, and ⊗ stands for the tensor product.

To model the torques, we evaluate the spin accumulation in the FL. As there is no spin flip within the tunnel barrier, it is assumed that all relaxation lengths are infinite in this region. However, this appears not to be sufficient to preserve the spin accumulation across the barrier. The spin diffusion coefficient in the barrier has also to be chosen considerably larger compared to the electron diffusion coefficients in the ferromagnetic layers, as only in this case the spin accumulation across the barrier is preserved. This is not a limitation, however, as we only need to mimic the correct behavior of the spin and charge transport through the tunnel barrier, we are free to choose an appropriate value for the diffusion constant in the barrier. The spin accumulation in a MTJ with the fixed (left) layer along the X-axis and the FL (right) along the Y-axis is show in Fig.5a. The role of the spin dephasing length λ_φ on the spin accumulation is also shown. The term with λ_φ in (6) suppresses the out of-plane component of the spin accumulation normal to both magnetizations in the fixed and FL (Fig.5a). As the main contribution to the field-like term comes exactly from the term with λ_φ, a strong spin dephasing suppresses the field-like torque component, while the effect on the damping-like torque is less pronounced resulting in about 20% increase of this torque (Fig.5b).

4. TEMPERATURE MODELING

The current density flowing through the MRAM cell is relatively high and results in a temperature increase, which mediates the switching of the FL magnetization. On the other hand, the increased FL temperature, caused by self-heating, can result in an information loss, as it compromises the thermal stability. To preserve the data, the temperature must rapidly relax after writing. Besides the relatively high current density in a MTJ, it is also strongly inhomogeneous due to non-uniform



(a) (b)

Figure 6. (a) Complete structure of a MRAM cell including the MTJ stack (blue, cyan, red) and contacts (grey), coated with a 200 nm passivation layer (yellow). (b) Averages of magnetization components of the FL during switching from anti-parallel to parallel magnetization.

magnetization (Fig.5). Therefore, to evaluate the temperature during the MRAM switching, it must be modeled together with the charge, spin, and magnetization dynamics in a fully three-dimensional (3D) model.

To model the temperature T in the system at position $\boldsymbol{r}$ and time t we employ the heat flow equation

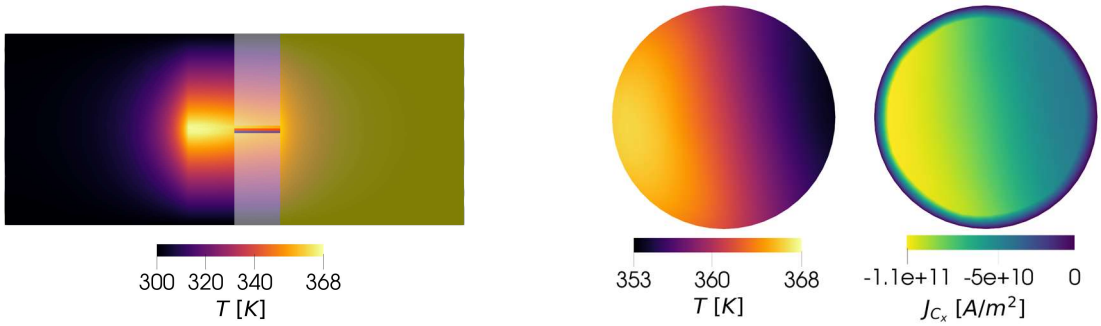
$$c_v \rho \frac{\partial T(\boldsymbol{r},t)}{\partial t} - \boldsymbol{\nabla} \cdot [\kappa \, \boldsymbol{\nabla} T(\boldsymbol{r},t)] = q(\boldsymbol{r},t). \tag{10}$$

c_v, ρ, and κ stand for the heat capacity, mass density, and heat conductivity of the material. q represents the heat source term which can further be split into two separate parts, representing the two main heating contributions [30].

The first source term is attributed to Joule heating $q_r(\boldsymbol{r},t) = \boldsymbol{J_C}^2(\boldsymbol{r},t)\rho_{\mathrm{E}}$ in the contacts and the magnetic layers. Here ρ_{E} and $\boldsymbol{J_C}$ stand for material resistivity and current density, respectively. The second type of heating $q_t(\boldsymbol{r},t)$ is attributed to electrons tunneling through the barrier and can be expressed by [31]

$$q_t(\boldsymbol{r},t) = \left(1 \pm \alpha(\Delta U)\right) \frac{J_{C_x}(y,z,t)\Delta U(y,z,t)}{2\lambda} \exp\left(-\frac{|x - x_{F/P}|}{\lambda}\right), \tag{11}$$

considering the main axis of the MRAM and the X-axis to coincide. Here $J_{C_x}(y,z,t)$ and $\Delta U(y,z,t)$ are the X-component of the current density and a potential drop across the barrier at position (y,z), λ is a characteristic length at which hot electrons/holes lose their energy, and x_F and x_P stand for the free layer/tunnel barrier-interface and pinned layer/tunnel barrier-interface positions. The asymmetry coefficient $\alpha(\Delta U)$ accounts for the heat production imbalance between the source and receiver side.



(a) (b)

Figure 7. (a) Averages of the magnetization components of the FL during switching from anti-parallel to parallel magnetization. (b) Temperature of the structure at 1.75 ns. On the left side the colors are to the temperature scale shown below, on the right the temperature field is displayed transparently over the MRAM. The hot area is confined around the MTJ stack.

Fig.7a shows a cylindrical MRAM structure with a diameter of 40 nm. The MTJ stack consisting of CoFeB(1 nm)/MgO(1 nm)/CoFeB(1.2 nm) is connected to 40 nm long contacts. The contact ends are kept at the constant temperature 300 K. The structure is coated with a 200 nm SiO_2 passivation layer (PL). No heat transfer is allowed across the outer shell of the passivation layer. When a potential is applied between the contacts, the current starts to flow, exerting a STT on the FL which eventually leads to the magnetization flipping. The switching from the initial anti-parallel to the final parallel magnetization state is displayed in Fig 7b. The switching simulations in Fig.7b are the results of numerically solving (1) with the ingredients (2-9) on the finite element mesh of the structure shown in Fig.7a. The averages of the normalized magnetization components are displayed and show three different switching phases: Initial magnetization oscillations in the Y-Z-plane, fast X-component of magnetization m_x change, and a final magnetization oscillation in the Y-Z-plane.

During switching the local magnetization alignment is time dependent, resulting in a rapidly changing electric current density and the corresponding heat sources in (10) and in particular in (11). In Fig.7a, the temperature of the memory cell at 1.75 ns is shown. The left side shows the temperature colors in the true scale (shown below). On the right side the temperature is made transparent to show the MRAM structure behind for convenience. It follows from Fig.7a that the warmest region is strongly confined around the MTJ stack. Fig.7b shows a temperature profile (left) and the current density distribution (right) in the FL at 1.75 ns. The temperature in the FL is inhomogeneous and the warm (cold) region corelates well with the region of the high (low) absolute value of the current density due to non-uniform magnetization of the FL.

5. DYNAMIC DETERMINISTIC SWITCHING OF SOT MRAM

Spin-orbit torque three terminal memory devices are attractive to introduce nonvolatility in high level caches and replace SRAM. To increase the integration density, devices with perpendicular magnetization are required, as they allow to reduce the FL size and the total footprint of the memory cell. However, an external in-plane magnetic field is required to guarantee deterministic switching of the perpendicular FL. To achieve magnetic field-free switching, a scheme based on two sub-nanosecond orthogonal current pulses was proposed [20]. This scheme is suitable for integration in a cross-bar architecture [21].

Fig.8 shows a schematic of the structure of a memory cell for the two-pulse SOT switching scheme. The cell is formed by growing a perpendicularly magnetized FL on top of a heavy metal wire (NM1), where a first current pulse is applied to generate the initial SOT on the FL. On the right part of the cell, a second, orthogonal heavy metal wire (NM2) is placed on top of the FL, through which a second current pulse is applied. The SOT generated due to this second pulse acts on the FL to complete the magnetization switching of the memory cell. The NM1/FL/NM2 stack composes the structural part used for the writing operation of the memory cell. In the left part of the cell, next to the SOT writing stack, a MTJ (not shown) is grown on top of the FL, which is required for reading the state of the memory cell. The reading operation is carried out by applying a low current pulse through the MTJ and sensing the corresponding tunneling magnetoresistance ratio. The reading current through the MTJ is typically much smaller than the writing currents, which prevents reliability problems with the MTJ tunnel barrier.

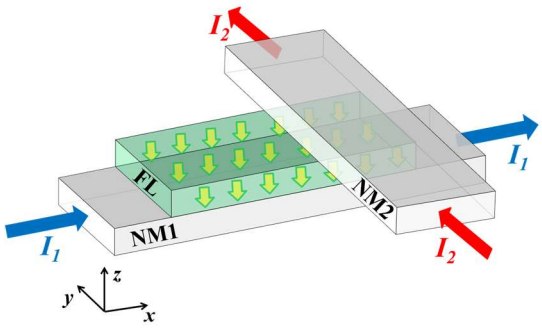


Figure 8. SOT-MRAM cell for switching based on two orthogonal current pulses. First, a current pulse (I_1) is applied to the NM1 wire, which is followed by a second current pulse (I_2) applied through the NM2 wire. The SOT generated by these current pulses acts on the magnetization of the FL and guarantees deterministic switching.

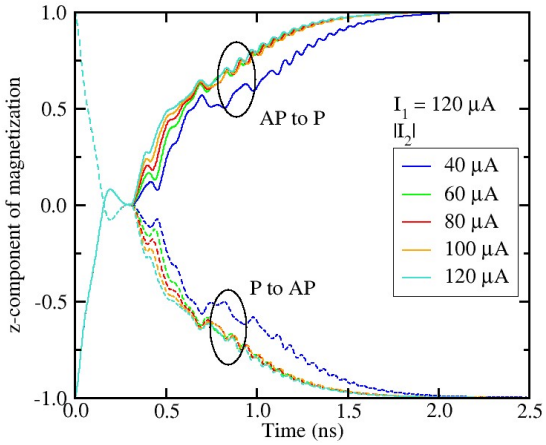


Figure 9. Magnetization (Z-component) of the FL as a function of time from switching simulations from parallel to anti-parallel (P to AP) and anti-parallel to parallel (AP to P) for various amplitudes of the second current pulse. Each curve represents an average of fifty switching realizations, corresponding to 100% switching probability. Despite the large current reduction, the switching remains deterministic and fast.

Based on this scheme, the FL magnetization is flipped purely electrically, i.e. without an external magnetic field, due to the sequence of applied current pulses. Considering the schematic of Fig.8, the first current pulse applied to the NM1 wire in the x direction brings the magnetization, initially oriented to the -z direction, to the plane of the FL pointing to the y direction. Upon the application of the second current pulse through the NM2 wire in the y direction, the magnetization of the FL under this wire, which overlaps just part of the FL, is rotated towards the X-axis. The resulting magnetization configuration creates an in-plane dipolar magnetic field which acts on the rest of the FL and generates a torque on the magnetization to complete the switching deterministically, similar to an applied external field. Thus, the magnetic field creation which drives deterministic switching, is the result of a dynamic process driven by the two consecutively applied current pulses.

Considering the parameters in Table 1, Fig.9 shows the magnetization switching for several magnitudes of the second current pulse. Results for both parallel to anti-parallel (P to AP) switching and vice-versa (AP to P) are presented. The magnitude of the first current pulse is set to 120 µA, corresponding to a current density of $2\text{x}10^{12}$ A/m^2, which is the critical current required for the SOT to be strong enough to put the magnetization in the plane of the FL, as previously mentioned. In turn, the second current amplitude can be significantly reduced. Our results have indicated that the current of the second pulse can be decreased to about one-third of the current of the first pulse, while deterministic switching is still guaranteed, as can be verified from the switching simulation shown in Fig.9. Furthermore, the switching dynamics and time are not significantly affected even for such a large current reduction. It should be pointed out that the polarity of the second current pulse has to be inverted depending on the switching direction (AP to P or P to AP). In our implementation, a positive

Table 1. Simulation parameters.

Parameter	Value
Saturation magnetization, M_S	1.1×10^6 A/m
Exchange constant, A	1.0×10^{-11} J/m
Perpendicular anisotropy, K	8.4×10^5 J/m^3
Gilbert damping factor, α	0.035
Spin Hall angle, θ_{SH}	0.3
Thermal stability factor, Δ	45
Free layer dimensions	40 nm × 20 nm × 1.2 nm
NM1: $w_1 \times l$	20 nm × 3 nm
NM2: $w_2 \times l$	20 nm × 3 nm

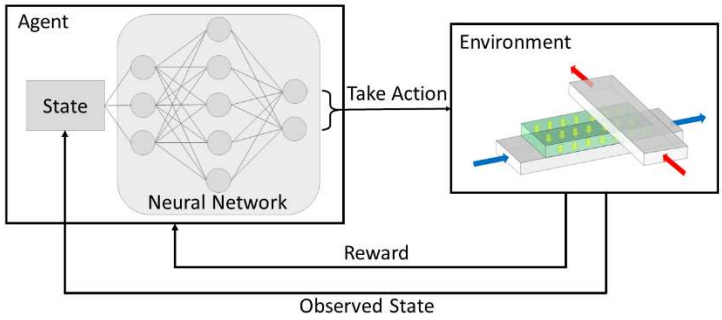


Figure 10. General setup of the reinforcement learning simulation: a simulation of the SOT MRAM cell acts as environment which an agent interacts with to build up a policy based on a neural network.

current flowing through the NM2 wire, i.e. along the +y direction, is required for an AP to P switching, while a negative current (along the –y direction) is required for a P to AP switching. Each curve in Figure 9 represents an average of fifty switching realizations, which corresponds to 100% switching probability, taking into account a stochastic thermal field at room temperature.

Reducing the second current amplitude has an important and advantageous impact for the two-pulse switching scheme. As the required writing power is proportional to the square of the current, lowering the current immediately leads to a decrease in the consumed amount of power. The reduction of the second current from 120 μA to 40 μA reported in Fig.9 yields a decrease of about 45% of the required writing power for our SOT MRAM cell based on the two-pulse switching scheme. Importantly, such a significant decrease in power consumption is achieved while preserving fast and deterministic switching.

6. REINFORCEMENT LEARNING FOR DETERMINISTIC SOT MRAM SWITCHING

In the SOT switching scheme presented in the previous section, the switching behavior of the memory cell is highly dependent on the exact placement of the current pulses, turning switching of the magnetization into a decision-making problem. Reinforcement learning (RL), a branch of machine learning, deals exactly with this kind of problem, in which a certain objective is to be optimized [32]. The typical RL setting consists of an agent and an environment. While the environment represents the system from an initial state into a target state, the agent is the entity which tries to learn the best decision-making policy by repeated interaction with that environment. In every loop of this iterative process, the environment transitions from the current state to a new state, reacting to the actions taken by the agent. The new state, as well as a reward indicating how good or bad the previous action was with respect to the objective, are returned to the agent. Repeatedly making these observations, the agent refines its policy over many iterations.

The setup in which we apply RL in order to learn how to optimally switch a SOT MRAM cell is given in Fig.10. The agent uses the deep Q-network (DQN) algorithm [33] to improve its decision-making and the inputs and outputs of the RL agent, as well as the environment are explained in the following.

6.1 State

The state information observed by the agent at every iteration consists of 11 variables, namely 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, $H_{ef\ x,y,z}$, and two Boolean variables indicating whether the two pulses can be currently set.

6.2 Actions

A set of four different actions can be taken by the agent. The two current pulses can be turned on individually, i.e. both pulses can be turned on, both pulses can be turned off, only the first pulse can be turned on, or only the second pulse can be turned on. Based on results in [21], [34], the current amplitudes of the pulses were chosen to be 130 μA and 100 μA, respectively. A minimum on/off period of 100 ps was enforced to prevent the agent from turning the current pulses on and off arbitrarily fast.

6.3 Environment

An in-house developed finite difference simulation of the pulsed SOT MRAM cell presented in Fig.10 is used as the environment [35].

6.4 Reward

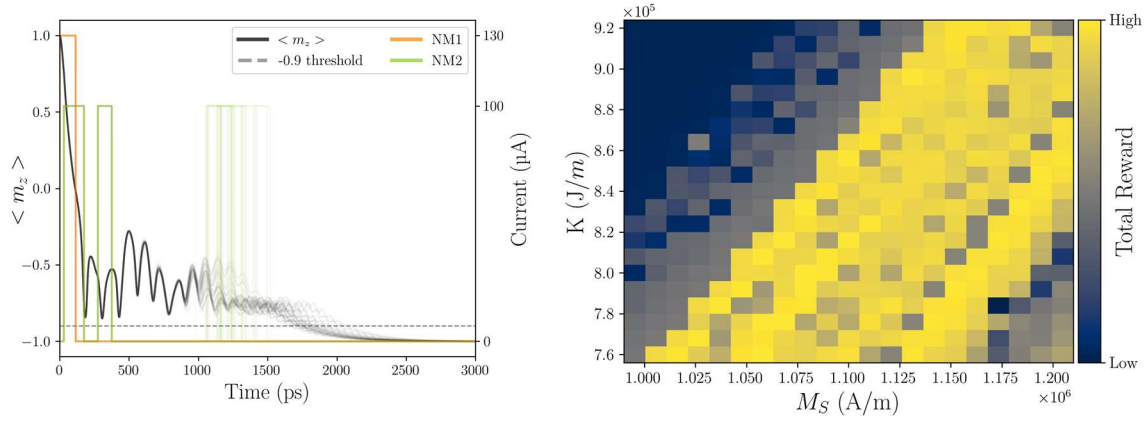
The second part contributing to the agent's observation is the reward. The reward function encodes the objective, which in this case is a fast reversal of the Z-component of the magnetization from +1 to -1. The following formula achieves exactly this:

$$r = m_{z,target} - m_{z,current} \tag{12}$$

Due to the target value of -1, the reward is always negative, but the more negative the further away the current value of the average Z-component of the magnetization is from the target value. As the RL agent tries to maximize the accumulated overall reward over one switching simulation, (12) ensures both, that the agent tries to align the current value of the average Z-component of the magnetization with the target value, and that this transition occurs fast.

During the training phase in which the agent refines its policy, many consecutive switching simulations are performed from which the agent accumulates experience. Over the course of this period, the material parameters were set to M_S=1.1×10^6 A/m and K=8.4×10^5 J/m^3. After the training phase, the same setup as depicted in Fig.10 can be used to perform simulations and assess the performance of the trained neural network model. Due to the stochastic influence of the thermal field on the effective magnetic field, 50 switching realizations were performed to assess the result of the training phase, with the results presented in Fig.11a. The pulses applied by the trained neural network model, as well as the trajectories of the average Z-component of the magnetization of all 50 realizations, are shown. The slight transparency of the single plot lines reveals trajectories followed more often, as those appear more solid. The results show that in the beginning of the realizations (<500 ps), the same NM1 and NM2 pulses are applied in all of them. Only later, at around 1 ns, the behavior starts to differ slightly. Depending on the exact trajectory of the Z-component of the magnetization, the time when further NM2 pulses are applied varies. In all the realizations, the trained neural network model is able to apply pulses such that the magnetization is reliably reversed from +1 to -1.

By performing simulations under variation of the anisotropy constant K and the saturation magnetization M_S of up to 10%, the performance of the model trained on fixed material parameters was evaluated. The results of these simulations are shown in Fig. 11b. As the same setup was used as in the training phase, at the end of each switching simulation an accumulated reward is returned, corresponding to the performance of the neural network model. Here, a high overall reward corresponds to the Z-component of the magnetization being brought faster and closer to the target value. Looking at the results in Fig.11b, one can see that the neural network model is performing well over a wide range of material parameter



(a) (b)

Figure 11. (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.

combinations. In the upper left corner, with higher values of the anisotropy constant and lower values of the saturation magnetization, the performance starts to worsen. This, however, is consistent with previously published results, which have shown that for this regime of parameter combinations, a higher current is required [21]. Overall, around 42% of the realizations reach the threshold value of -0.9, at which we consider the memory cell to be switched.

7. CONCLUSION

Magnetoresistive nonvolatile devices based on magnetic tunnel junctions possess a TMR suitable for practical applications. STT MRAM is positioned as a successor not only to flash memory but is also capable to penetrate CMOS-based main computer memory applications and even to slow last-level caches. STT MRAM is aiming at stand-alone as well as data-intensive embedded and low-power mobile, automotive, and Internet of Things applications. While designing STT MRAM cells, it is necessary to satisfy several important requirements, namely fast data writing, reliable data reading, large endurance, and long retention time. As improving one of the characteristics (speed) may result in a deterioration of others (endurance and retention), a careful optimization of the cell is paramount. Computer-aided design helps improving the devices, thus shortening the time for research and development and accelerating the introduction of this disruptive technology to the market. We presented high-performance finite element-method based modeling methods and simulation approaches in the area of nonvolatile magnetoresistive random access memories to facilitate development and fabrication of modern and emerging micro- and nanoelectronic magnetoresistive devices [1-6] utilizing the electron spin as a computational degree of freedom. In particular, we demonstrated three-dimensional self-consistent simulations of spin accumulation, torques, and magnetization dynamics in magnetic structures including tunnel junctions. The key simulation capabilities to evaluate efficiently (i) the effective field (including the demagnetization and stray field contributions in a disconnected geometry), (ii) coupled three-dimensional charge and spin transport, and (iii) the corresponding spin-transfer torques driving (iv) the magnetization dynamics and switching required for modeling MRAM were discussed. An implementation is devised in C++ by utilizing well supported FEM libraries under open-source licenses.

Because of large switching currents and insufficient speed, STT MRAM is unlikely to replace SRAM. A SOT MRAM cell with a perpendicular FL is a good candidate as it is sufficiently fast to serve in high-level caches. We demonstrated that the perpendicular FL can be switched rapidly and deterministically without an external magnetic field by using a dynamic scheme based on orthogonal current pulses. The optimal pulse sequence can be determined by applying machine learning algorithms. The successful adoption of nonvolatility in microelectronic systems by developing various logic, logic-in-memory architectures, and in-memory processing will inevitably result in increasing dissemination of this technology to other applications such as ultralow-power electronics, high-performance computing, Big Data analysis, automotive electronics, and the Internet of Things and is expected to have an enormous impact on neuromorphic computing and artificial intelligence.

ACKNOWLEDGMENT

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] H. Ohno, M.D. Stiles, B. Dieny: Spintronics, Proc. of the IEEE 104, 1782-1786 (2016). DOI: 109/JPROC.2016.260116301163

[2] https://irds.ieee.org/editions/2021

[3] S. Iwasaki: Perpendicular Magnetic Recording-Its Development and Realization, Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 85, 37-54 (2009). DOI: 10.2183/pjab.85.37

[4] L. Savtchenko, B. Engel, N. Rizzo et al.: Method of Writing to Scalable Magnetoresistance Random Access Memory Element, US Patent 6545906 B1 (2003)

[5] R. Sbiaa, H. Meng, S.N. Piramanayagam: Materials with Perpendicular Magnetic Anisotropy for Magnetic Random Access Memory, Phys. Stat. Solidi (RRL) – Rapid Research Letters 5, 413-419 (2011). DOI: 10.1002/pssr.201105420

[6] J.C. Slonczewski: Current-driven Excitation of Magnetic Multilayers, J. Magn. Magn. Mater. 159, L1-L7 (1996). DOI: 10.1016/0304-8853(96)00062-5

[7] L. Berger: Emission of Spin Waves by a Magnetic Multilayer Traversed by a Current, Phys. Rev.B 54, 9353-9358 (1996). DOI: 10.1103/PhysRevB.54.9353

[8] D. Apalkov, B. Dieny, and J.M. Slaughter: Magnetoresistive Random Access Memory, Proc. of the IEEE 104, 1796-1830 (2016). DOI: 10.1109/JPROC.2016.2590142

[9] S.-W. Chung, T. Kishi, J.W. Park et al.: 4Gbit Density STT-MRAM Using Perpendicular MTJ Realized with Compact Cell Structure, Proc. International Electron Devices Meeting (IEDM), 659-662 (2016). DOI: 10.1109/IEDM.2016.7838490

[10] Y.J. Song, J.H. Lee, H.C. Shin et al.: Highly Functional and Reliable 8Mb STT-MRAM Embedded in 28nm Logic, Proc. International Electron Devices Meeting (IEDM), 663-666 (2016). DOI: 10.1109/IEDM.2016.7838491

[11] H. Sato, H. Honjo, T. Watanabe et al.: 14ns Write Speed 128Mb Density Embedded STT-MRAM with Endurance >1010 and 10yrs Retention @85°C Using Novel Low Damage MTJ Integration Process, Proc. International Electron Devices Meeting (IEDM), 608-611 (2016). DOI: 10.1109/IEDM.2018.8614606

[12] O. Golonzka, J.-G. Alzate, U. Arslan et al.: MRAM as Embedded Non-volatile Memory Solution for 22FFL FinFET Technology, Proc. International Electron Devices Meeting (IEDM), 412-415 (2018). DOI: 10.1109/IEDM.2018.8614620

[13] I.M. Miron, K. Garello, G. Gaudin et al.: Perpendicular Switching of a Single Ferromagnetic Layer Induced by In-plane Current Injection, Nature 476, 189 (2011). DOI: 10.1038/nature10309

[14] L. Liu, J. Lee, T.J. Gudmundsen et al.: Current-induced Switching of Perpendicularly Magnetized Magnetic Layers Using Spin Torque from the Spin Hall Effect, Phys. Rev. Lett. 109, 096602 (2012). DOI: 10.1103/PhysRevLett.109.096602

[15] S.-W. Lee, K.-J. Lee: Emerging Three-terminal Magnetic Memory Devices, Proc. of the IEEE 104, 1831-1843 (2016). DOI: 10.1109/JPROC.2016.2543782

[16] K. Garello, F. Yasin, S. Couet et al.: SOT-MRAM 300mm Integration for Low Power and Ultrafast Embedded Memories, Proc. VLSI Symp. Technology and Circuits, C8-2 (2018). DOI: 10.1109/VLSIC.2018.8502269

[17] S. Fukami, T. Anekawa, C. Zhan, H. Ohno: A Spin–orbit Torque Switching Scheme with Collinear Magnetic Easy Axis and Current Configuration, Nature Nanotechnology 11, 621-625 (2016). DOI: 10.1109/IEDM.2012.6479128

[18] Y.-C. Lau, D. Betto, K. Rode et al.: Spin–orbit Torque Switching without an External Field using Interlayer Exchange Coupling, Nature Nanotechnology 11, 758-762 (2016). DOI: 10.1038/nnano.2016.84

[19] C.K. Safeer, E. Jué, A. Lopez et al.: Spin-Orbit Torque Magnetization Switching Controlled by Geometry, Nature Nanotechnology 11 143-146 (2015). DOI: 10.1038/nnano.2015.252

[20] V. Sverdlov, A. Makarov, S. Selberherr: Two-pulse Sub-ns Switching Scheme for Advanced Spin-Orbit Torque MRAM, Solid-State Electronics 155, 49-56 (2019). DOI: 10.1016/j.sse.2019.03.010

[21] R.L. Orio, J. Ender, S. Fiorentini et al.: 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

[22] D.R. Fredkin, T.R. Koehler: Hybrid Method for Computing Demagnetizing Fields, IEEE Transactions on Magnetics 26(2), 415-417 (1990). DOI: 10.1109/20.106342

[23] N. Popović, D. Praetorius: Applications of H-Matrix Techniques in Micromagnetics, Computing 74(3), 177-204 (2005). DOI: 10.1007/s00607-004-0098-7

[24] T. Kolev, V. Dobrev: MFEM: Modular Finite Element Methods Library, http://mfem.org, (2010)

[25] N. Albrecht, C. Börst, D. Boysen et al.: H2Lib, http://www.h2lib.org, (2016)

[26] C. Abert, L. Exl, G. Selke: Numerical Methods for the Stray-Field Calculation: A Comparison of Recently Developed Algorithms, Journal of Magnetism and Magnetic Materials 326, 176-185 (2013). DOI: 10.1016/j.jmmm.2012.08.041

[27] C. Abert, M. Ruggeri, F. Bruckner et al.:A Three-dimensional Spin-diffusion Model for Micromagnetics, Scientific Reports 5, 14855 (2015). DOI: 10.1038/srep14855

[28] S. Lepadatu: Unified Treatment of Spin Torques Using a Coupled Magnetisation Dynamics and Three-dimensional Spin Current Solver, Scientific Reports 7, 12937 (2017). DOI: 10.1038/s41598-017-13181-x

[29] S. Fiorentini, J. Ender, S. Selberherr et al.: Coupled Spin and Charge Drift-Diffusion Approach Applied to Magnetic Tunnel Junctions, Solid-State Electronics 186, 108103 (2021). DOI:10.1016/j.sse.2021.108103

[30] I.L. Prejbeanu, M. Kerekes, R.C. Sousa et al.: Thermally assisted MRAM, Journal of Physics: Condensed Matter 19 (16), 165218 (2007). DOI:10.1088/0953-8984/19/16/165218.

[31] T. Hadámek, S. Selberherr, W. Goes, V. Sverdlov: Heating Asymmetry in Magnetoresistive Random Access Memories, Proceedings of the World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI), 63–66 (2021), ISBN: 978-1-950492-55-8

[32] R.S. Sutton and A.G. Barto,:Reinforcement Learning: An Introduction,Cambridge, MA, USA: MIT press (1998)

[33] D. Silver, T. Hubert, J. Schrittwieser et al.: A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-play, Science 362, no. 6419, 1140-1144 (2018). DOI: 10.1126/science.aar6404

[34] R.L. Orio, J. Ender, S. Fiorentini et al.: Optimization of a Spin-Orbit Torque Switching Scheme Based on Micromagnetic Simulations and Reinforcement Learning, Micromachines 12, 443 (2021). DOI: 10.3390/mi12040443

[35] 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