Switching Current Reduction in Advanced Spin-Orbit Torque MRAM

Viktor Sverdlov, Alexander Makarov, and Siegfried Selberherr
Institute for Microelectronics, TU Wien, Gußhausstraße 27-29, A-1040 Wien, Austria
e-mail: {Sverdlov|Makarov|Selberherr}@iue.tuwien.ac.at

Abstract— **The steady increase in performance and speed of modern integrated circuits is continuously supported by constant miniaturization of complementary metal-oxide semiconductor (CMOS) devices. However, a rapid growth of the dynamic and stand-by power due to transistor leakages becomes a pressing issue. A promising way to slow down this trend is to introduce non-volatility. The development of an electrically addressable non-volatile memory combining high speed and high endurance is essential to achieve these goals. It is particularly promising to employ non-volatility in IoT and automotive applications, as well as in the main computer memory as a replacement of conventional volatile CMOS-based DRAM. To further reduce the energy consumption, it is essential to replace SRAM in modern hierarchical multi-level processor memory structure with a non-volatile memory technology. The spin-orbit torque magnetic random access memory (SOT-MRAM) combines non-volatility, high speed, and high endurance and is thus suitable for applications in caches. However, its development is still hindered by relatively high switching currents. Several paths to reduce the switching current in an in-plane SOT-MRAM structure are analyzed. The switching by means of two orthogonal current pulses complemented with an interface-induced perpendicular magnetic anisotropy allows reducing the switching current significantly for achieving sub-500ps switching.**

Keywords— Spin-orbit torque; MRAM; switching; in-plane MTJ, perpendicular magnetic anisotropy

I. INTRODUCTION

Continuous miniaturization of complementary metal-oxide semiconductor (CMOS) devices is one of the driving forces for the unprecedented increase of speed and performance delivered by modern integrated circuits However, a rapid increase of the stand-by power due to transistor leakages and the need to refresh the data in dynamic random access memory (DRAM) is becoming a pressing issue. The microelectronics industry is facing major challenges related to power dissipation and energy consumption, and the microprocessors' scaling will hit a power wall soon. An attractive path to mitigate the unfavorable trend of increasing power at stand-by is to introduce non-volatility in the circuits. The development of an electrically addressable non-volatile memory which combines fast operation, simple structure, and high endurance is essential to mitigate the increase of the stand-by power and to introduce instant-on architectures without the need of data initialization when going from a stand-by to an operation regime. Oxide-based resistive RAM (RRAM) possesses filamentary switching between the on/off states and is thus intrinsically prone to significant resistance fluctuations in both states. In addition, the endurance reported is only slightly higher than that in flash memory. For this reason, it is premature to consider RRAM at its current stage of development for digital applications. As continuous conductance modulation is suitable for implementing analog synaptic weights, both filamentary and non-filamentary switching RRAM types are currently intensively investigated, particularly for neuromorphic applications [1].

Spin-transfer torque magnetic RAM (STT-MRAM) is fast (10ns), possesses high endurance (10^{12}), and has a simple structure. It is compatible with CMOS and can be straightforwardly embedded in circuits, e.g. [2]. High-density STT-MRAM arrays with 4Gbit capacities have been already demonstrated [3]. The availability of high-capacity non-volatile memory close to high-performance CMOS circuits allows exploring conceptually new logic-in-memory [4] and computing-in-memory architectures for future artificial intelligence and cognitive computing.

Although the use of STT-MRAM in last-level caches is conceivable [5], the switching current for operating at a speed faster than 10ns is rapidly growing. This results in large dimensions of the access transistor. The need of even larger switching currents for faster operation in higher-level caches potentially prevents STT-MRAM from entering in L2 and L1 caches currently mastered by static RAM (SRAM). In addition, rapidly increasing critical currents required for operating STT-MRAM at 5ns result in large current densities running through magnetic tunnel junctions. This leads to oxide reliability issues, which in turn reduce the MRAM endurance to that of the flash memory, thus negating one of the important MRAM advantages over flash.

The development of an electrically addressable non-volatile memory combining high speed (sub-ns operation) and high endurance is essential for replacing SRAM in higher-level caches of hierarchical multi-level processor memory structures with a non-volatile memory [4]. Among the newly discovered physical phenomena suitable for next-generation MRAM is the spin-orbit torque (SOT) assisted switching at room temperature in heavy metal/ferromagnetic [6-12] or topological insulator/ferromagnetic [13,14] bilayers. In this memory cell the MTJ's free layer is grown on a material with a large spin Hall angle. By passing the current through the material the SOT acting on the free layer is generated. The large switching current is injected in-plane along the heavy metal/ferromagnetic bilayer and does not flow through the MTJ. This results in a three-terminal configuration, in which

978-1-5386-4811-7/18/$31.00 ©2018 IEEE

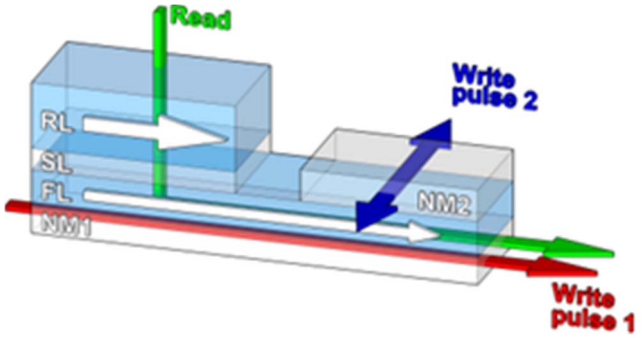

Fig. 1. Schematic of the in-plane SOT-MRAM cell. The free layer of an MTJ is grown above the heavy metal wire NM1, through which the current pulse is applied. The wire NM2 serves to conduct the perpendicular current "Write pulse 2" of the in-plane SOT-MRAM cell.

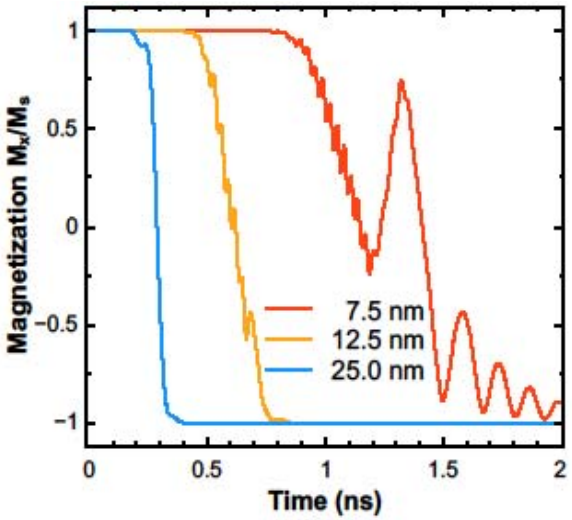

Fig. 2. Time dependent magnetization evolution for several values of the second wire NM2 overlap. The current is scaled with NM2 and is equal to 180μA for NM2=25nm, 90μA for NM2=12.5nm, and 54μA for NM2=7.5nm.

the read and write current paths are decoupled. Since the large write current does not flow through the oxide in the MTJ, this prevents the tunnel barrier from damage. It improves device reliability by eliminating correlations between the switching current and the retention time. Therefore, three-terminal MRAM cells are promising candidates for future generations of non-volatile memory for fast sub-ns switching [15].

SOT-MRAM is an electrically addressable non-volatile memory combining high speed and high endurance and is thus suitable for applications in caches [15]. Although the high switching current is not flowing through a magnetic tunnel junction but rather through a heavy metal wire under it, the current is still high, and its reduction is the pressing issue in the field of SOT-MRAM development.

II. Method

The investigated memory cell is shown in Fig.1. It consists of an in-plane magnetized MTJ with its free layer lying on top of a heavy metal wire NM1 of 3nm thickness. The dimensions of the free layer are $52.5\times12.5\times2\text{nm}^3$. Another heavy metal wire NM2 with an overlap from the right side lesser or equal to the total free layer width 52.5nm serves to apply the second perpendicular current pulse and the spin-orbit torque associated with it. The magnetization dynamics of the magnetic system due to the spin current densities and the spin accumulations is well described by the Landau-Lifshitz-Gilbert equation [15,16].

$$\frac{\partial \mathbf{m}}{\partial t} = -\gamma \mathbf{m} \times \mathbf{H}_{\text{eff}} + \alpha \mathbf{m} \times \frac{\partial \mathbf{m}}{\partial t} + \gamma \frac{\hbar}{2e} \frac{\Theta_{SH} J}{M_0 d} [\mathbf{m} \times (\mathbf{m} \times \mathbf{y})] \quad (1)$$

m is the position-dependent magnetization **M** normalized by the saturation magnetization M_S: $\mathbf{m} = \mathbf{M}/M_S$, γ is the gyromagnetic ratio, α is the Gilbert damping parameter, e is the electron charge, ħ is the reduced Plank constant, and Θ_{SH} is an effective Hall angle relating the strength of the spin current density to the charge current density J. If the charge current flows along direction **x** in the heavy metal wire, the spin current floats perpendicularly along the **z** direction from the heavy metal to the free ferromagnetic layer of the thickness d. The polarization of the spin current is pointing out along the direction ***y*** perpendicular to both the spin and charge current directions. The spin polarization entering into the free ferromagnetic layer is becoming quickly aligned with the local magnetization **M** generating the torque [17], which acts on the magnetization, as described by the last term in (1). This spin-orbit torque together with the magnetic field $\mathbf{H}_{\text{eff}}$ describes the damped magnetization dynamics (1). The magnetic field $\mathbf{H}_{\text{eff}}$ includes the external field as well as the contributions due to bulk and interface-induced magnetic anisotropies, exchange field, and demagnetization field. Thermal effects on the magnetization dynamics are incorporated by means of a random magnetic field added to $\mathbf{H}_{\text{eff}}$. The strength of the thermal field fluctuations is proportional to temperature [18].

III. Results

We apply the two-pulse switching scheme previously proposed for efficient switching of an in-plane structure [19]. The SOT due to the first pulse through the nonmagnetic heavy metal wire NM1 tilts the magnetization of the free layer in-plane perpendicular to the direction of the "Write pulse 1" (Fig.1). We note that the "Write pulse 1" does not guarantee the deterministic magnetization switching, and an external magnetic field is required to break the mirror symmetry and complete the switching [20].

In contrast, the SOT of the second consecutive pulse through the wire NM2 is able to switch the magnetization deterministically even if it is applied alone [10,20]. The "Write pulse 2" current is applied along the **y** axis, which results in a torque term similar to the one in (1), where +(-)**y** is replaced with –(+)**x**. However, in order to switch the magnetization fast, large current densities and thus large currents are required. This

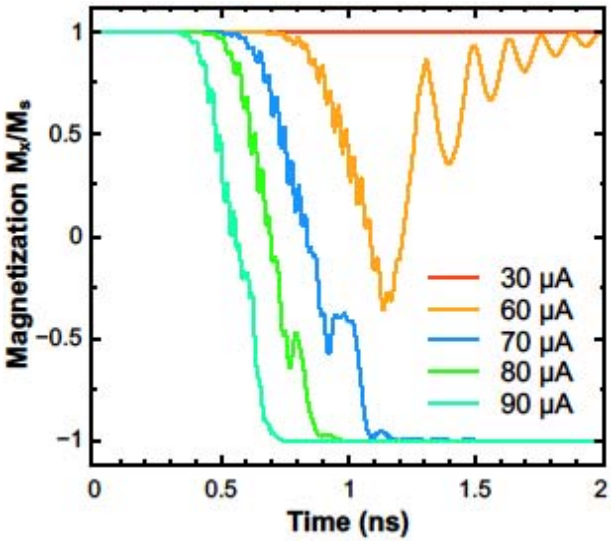

Fig. 3. Time evolution of the magnetization for several values of the current through wire NM2 with the width 12.5nm. The current through wire NM1 is zero.

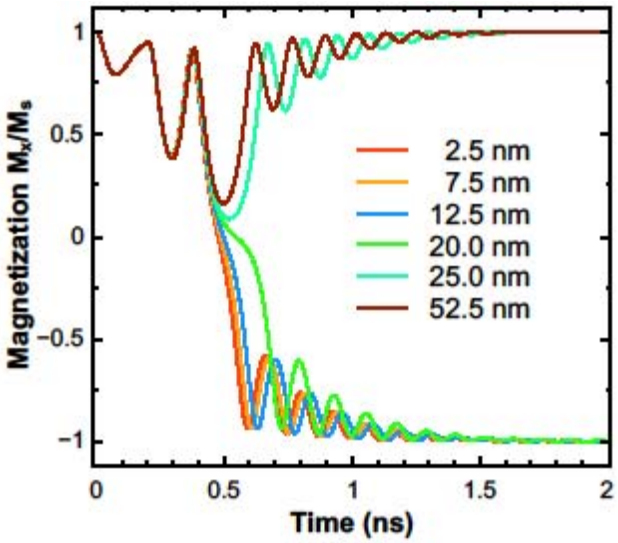

Fig. 4. The magnetization dynamics for several overlap lengths of wire NM2. Two consecutive perpendicular pulses with 30μA current and 200ps duration are applied.

results in large dimensions of the control transistor needed to supply the current. Therefore the switching current reduction becomes necessary in order to reduce the MRAM cell size.

The reduction of the current can be achieved by supplying the "Write pulse 2" through the NM2 wire with a reduced overlap to the free layer, provided the current density is fixed. A simple reduction of the overlap results in a smaller area of the free layer where the SOT is active. The reduction of the switching time is only a factor of two in the case when the NM2 is reduced from 25nm to 12.5nm. This scaling of the switching time with the current is attractive for applications as compared to the much weaker, logarithmic scaling of the switching time decrease with the current increase observed in STT-MRAM at precessional fast switching [2]. However, when the overlap is further decreased to 7.5nm, the switching time increases more rapidly compared to that expected from a geometrical scaling. This is due to the fact that the torque magnitude acting only on a small part of the free layer becomes insufficient to switch the whole layer efficiently, if the current pulse is 1ns.

The overlap of 12.5nm of the NM2 wire, which is equal to the width of the NM1 wire, is sufficient to provide sub-ns switching at a current value of 90μA. Fig.3 shows the magnetization switching if the current value is further decreased, while the pulse duration is fixed at 1ns. It appears that a switching current of 70μA still switches the magnetization, while 60μA is not sufficient to invert the magnetization within 1ns. We also note that the current of 30μA applied within 1ns, which is about a factor of two weaker than the critical switching current, does not manage to develop any visible dynamics.

One of the reasons why the magnetization does not show any substantial dynamics at a current value of 30μA is that the SOT due to the "Write pulse 2" is not efficient at the beginning when the magnetization is along the **x** direction parallel to the spin accumulation. Indeed, the last term in (1) is close to zero.

Therefore, although the maximal torque amplitude described by the coefficient in front of the cross product in the last term in (1) is sufficient to overcome the Gilbert damping and to invert the magnetization; the switching is characterized by a very long incubation time [20] necessary to develop a substantial deviation of the magnetization from the **x** axis. The initial spatial magnetization distribution is created during a 200ps thermalization process under the influence of the random thermal magnetic field (Fig.2, Fig.3). Although at every point the magnetization is not perfectly aligned with the **x** axis and contains projections on the other axes **y** and **z**, for which the SOT is efficient, the projections on these axes averaged over the whole sample are close to zero due to the random character of the thermal field. As a development of a substantial deviation of the averaged projections from zero is more probable in a smaller sample area, the SOT for NM2=12.5nm, albeit weaker than that for NM2=25nm, is more efficient from the beginning due to this deviation. This explains the better scaling of the switching time with the current in Fig.1 as compared to the one at STT switching and demonstrates the importance of making the SOT efficient from the beginning of switching for the reduction of the switching time.

In order to create an initial deviation of the average magnetization, we apply the current pulse through the wire NM1 just before the second, switching, pulse through the wire NM2 [19]. We assume equivalent pulses of 200ps duration. The first pulse through NM1 creates an initial deviation of the magnetization from its equilibrium direction and tends to put the magnetization along the **y** axis. This makes the torque from the second pulse efficient from the beginning thus removing the incubation period. The two-pulse scheme allows reducing the switching current by a factor of 3 as compared to single pulse switching of a similar duration. Fig.4 demonstrates that the optimal width of NM2 is around 12.5nm as its further reduction results in a high current density. In Fig.4 the magnitude of the currents of both pulses is fixed at 30μA.

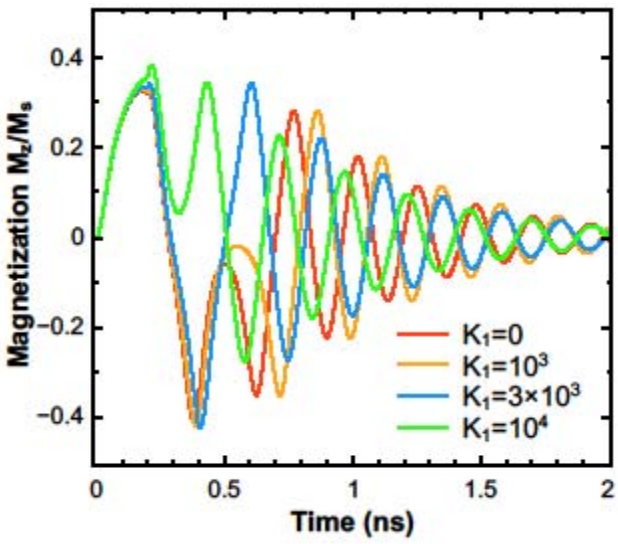

Fig. 5. Time dependence of the magnetization projection on the perpendicular **z** axis, for several values of the interface-induced perpendicular anisotropy.

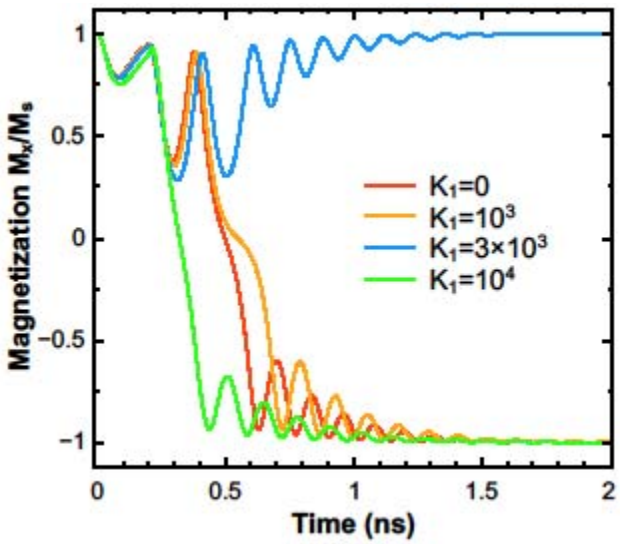

Fig. 6. Adding perpendicular magnetization if a two-pulse scheme allows to reduce the current by a factor of 6 for sub-0.5ns switching.

To further decrease the switching current, a perpendicular magnetic anisotropy typically developed at the MgO/CoFeB interface is introduced [21]. The perpendicular anisotropy compensates the large demagnetizing contribution when the magnetization is out of plane. However, the perpendicular anisotropy may compromise the temperature stability of the in-plane structure, when its value K_1 is too large. Fig.5 shows the time evolution of the magnetization projection on the **z** direction perpendicular to the structure, for several K_1, for a current fixed at 15µA, while Fig.6 displays a typical dependence of the in-plane magnetization along the easy axis.

IV. Conclusion

In conclusion, the introduction of a perpendicular anisotropy K_1 into the two-pulse switching scheme allows reducing the switching current from 180µA to 15µA, while achieving fast sub-0.5ns writing suitable for cache applications.

References

[1] J G.W. Burr, R.M. Shelby, A.Sebastian et al., "Neuromorphic Computing Using Non-volatile Memory", Advances in Physics:X, vol.2, 89 (2017).

[2] D. Apalkov, B. Dieny, and J.M. Slaughter, "Magnetoresistive Random Access Memory", Proceedings of the IEEE,vol.104, 1796 (2016).

[3] S.-W. Chung, T. Kishi, J. W. Park et al.," 4Gbit Density STT-MRAM Using Perpendicular MTJ Realized with Compact Cell Structure", IEDM 2016 Techn. Digest, 660 (2016).

[4] T. Hanyu, T. Endoh, D. Suzuki et al., "Standby-Power-Free Integrated Circuits Using MTJ-Based VLSI Computing", Proceedings of the IEEE, vol.104, 1844 (2016).

[5] G.Jan, L.Thomas, S.Le, et al.,"Achieving Sub□ns Switching of STT□ MRAM for Future Embedded LLC Applications through Improvement of Nucleation and Propagation Switching Mechanisms", Symp. VLSI Technology, 18 (2016).

[6] T.Taniguchi, J.Grollier, and M.D.Stiles, "Spin-Transfer Torques Generated by the Anomalous Hall Effect and Anisotropic Magnetoresistance", Phys.Rev.Appl.,vol.3, 044001 (2015).

[7] I.M.Miron, G.Gaudin, S.Auffret et al., "Current-driven Spin Torque Induced by the Rashba Effect in a Ferromagnetic Metal Layer", Nature Materials, vol.9, 230 (2010).

[8] I.M.Miron, K.Garello, G.Gaudin et al., Perpendicular Switching of a Single Ferromagnetic Layer Induced by In-plane Current Injection", Nature, vol.476, 189 (2011).

[9] 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., vol.109, 096602 (2012).

[10] L.Liu, C.-F.Pai,, Y.Li et al., "Spin-Torque Switching with the Giant Spin Hall Effect of Tantalum", Science, vol.336, 555 (2012).

[11] A.Brataas and K. M.D.Hals, "Spin–orbit Torques in Action", Nature Nanotechnoogy, vol.9, 86 (2014).

[12] D.MacNeil, G.M.Stiehl, M.H.D.Guimaraes et al.,"Control of Spin–orbit Torques through Crystal Symmetry in WTe_2/Ferromagnet Bilayers", Nature. Physics, vol.13, 300 (2017).

[13] J.Han, A.Richardella, S.A.Siddiqui et al., "Room-Temperature Spin-Orbit Torque Switching Induced by a Topological Insulator", Phys.Rev.Lett., vol.119, 077702 (2017).

[14] Y.Wang, D.Zhu, Y.Wu et al., "Room Temperature Magnetization Switching in Topological Insulator-ferromagnet Heterostructures by Spin-orbit Torques", Nature Communications, vol.8, 1364 (2017).

[15] S.-W. Lee and K.-J. Lee, "Emerging Three-Terminal Magnetic Memory Devices", Proceedings of the IEEE, vol.104, 1831 (2016).

[16] T.Gilbert, "A Phenomenological Theory of Damping in Ferromagnetic Materials", IEEE Transactions on Magnetics, vol.40, 3443 (2004).

[17] A.D.Kent, B.Ozylimaz, and E.del Barco,"Spin-transfer-induced Precessional Magnetization Reversal, Appl.Phys.Lett.,vol.84, 3897 (2004).

[18] G.Finocchio, M.Carpentieri, B.Azzerboni et al., "Micromagnetic Simulations of Nanosecond Magnetization Reversal Processes in Magnetic Nanopillar", J.Appl.Phys.,vol.99, 08G522 (2006).

[19] A. Makarov, T. Windbacher, V. Sverdlov, and S. Selberherr, "CMOS-compatible Spintronic Devices: A Review", Semiconductor Science and Technology, vol.31, 113006 (2016).

[20] S.Fukami, T.Anekawa, C.Zhan, and H.Ohno, "A Spin–orbit Torque Switching Scheme with Collinear Magnetic Easy Axis and Current Configuration", Nature Nanotechnology, vol.11, 621 (2016).

[21] L.Liu, T.Moriyama, D. C.Ralph, and R.A Buhrman, "Reduction of the Spin-torque Critical Current by Partially Canceling the Free Layer Demagnetization Field", Appl.Phys.Lett., vol. 94, 122508 (2009).