Thermally assisted MRAM

2007 J. Phys.: Condens. Matter 19 165218

(http://iopscience.iop.org/0953-8984/19/16/165218)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **19** (2007) 165218 (23pp) doi:10.1088/0953-8984/19/16/165218

Thermally assisted MRAM

I L Prejbeanu[1,2], M Kerekes[1], R C Sousa[1], H Sibuet[3], O Redon[3], B Dieny[1] and J P Nozières[2]

[1] Spintec, URA 2512 CEA/CNRS, CEA-Grenoble, 17 r. des Martyrs, 38054 Grenoble Cedex 9, France
[2] Crocus Technology, 5 Place Robert Schumann, BP 1510, 38025 Grenoble Cedex 1, France
[3] LIMN, DRT/Léti/DIHS, CEA-Grenoble, 17 r. des Martyrs, 38054 Grenoble Cedex 9, France

E-mail: rsousa@cea.fr

Received 4 October 2006, in final form 17 November 2006
Published 6 April 2007
Online at stacks.iop.org/JPhysCM/19/165218

Abstract

Magnetic random access memories (MRAMs) are a new non-volatile memory technology trying establish itself as a mainstream technology. MRAM cell operation using a thermally assisted writing scheme (TA-MRAM) is described in this review as well as its main design challenges. This approach is compared to conventional MRAM, highlighting the improvements in write selectivity, power consumption and thermal stability. The TA-MRAM writing was tested and validated in the dynamic regime down to 500 ps write pulses. The heating process was investigated for the influence of the voltage pulse width, junction area and lead volume looking at the required write power density. The possibilities to control and reduce the write power density are described. The most promising solution to optimize the heating process and reduce the power consumption is to insert two thermal barrier layers at both ends of the MTJ layer stack, between the junction and the electrical leads, using low thermal conductivity materials. This minimizes the heating losses and improves the heating efficiency.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Magnetic random access memories (MRAMs) integrate a magnetoresistive device with a silicon-based selection matrix. The key attributes of MRAM are non-volatility, low voltage operation and unlimited read and write endurance combined with fast read and write operation. These characteristics give MRAM the potential to replace current memory types in specific applications. The interest in MRAM started some 20 years ago with a cross-tie RAM concept [1], followed by the use of the anisotropic magnetoresistance (AMR) materials [2], later replaced by higher sensitivity giant magnetoresistance (GMR) devices [3]

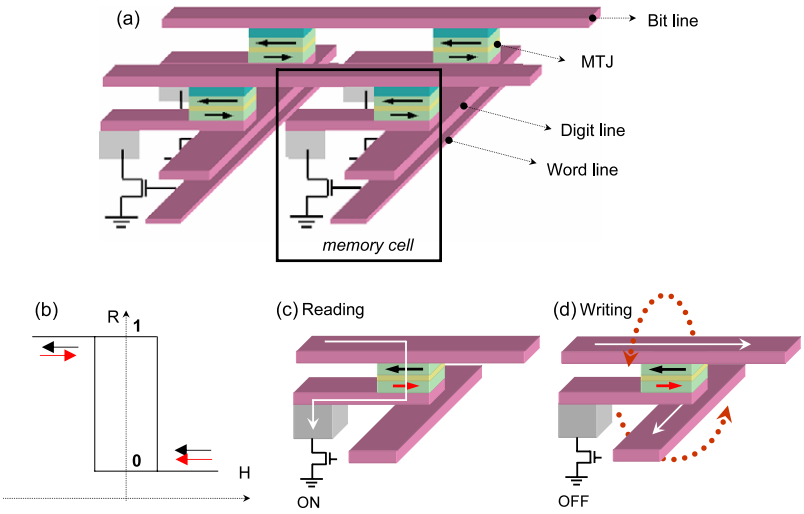

Figure 1. (a) Conventional architecture used in the first MRAM generation containing MTJ cells at the intersection of orthogonal writing lines and on top of a selection transistor. (b) Schematic view of a minor hysteresis loop showing the reversal of the storage layer and two corresponding resistance levels: high '1' and low '0'. Reading (c) and writing schemes (d) used in the conventional MRAM architecture. At read, the selection transistor is closed (on) and a small electrical current can pass through the MTJ cell, allowing the measure of its resistance. At write, the selection transistor is open (off) and the combination of two orthogonal magnetic fields ensures the selectivity.

and more recently using spin dependent tunnel junctions [4–6]. The MRAM development effort was renewed after the first successful attempts in fabricating spin dependent tunnel junctions showing significant magnetoresistance at room temperature using amorphous AlOx barriers [7, 8]. Recent studies using crystalline MgO barriers have shown that it is possible to increase the available magnetoresistive signal to more than 200% resistance change at room temperature [9, 10]. In its most simple implementation an MRAM cell is composed by a magnetic tunnel junction (MTJ) connected to a selection transistor [11]. The resistance of the memory bit is either low or high depending on the magnetization orientation of the free layer relative to the pinned reference layer, parallel or antiparallel. To read one bit, the selection transistor is turned on, and a read current flows through the tunnel junction as shown in figure 1.

The value of the junction resistance is then compared to a reference resistance half-way between the high and low resistance values. Initially the bit writing process proposed for MRAM relied on two orthogonal magnetic fields being applied to a cell. This switching approach is based on a coherent magnetization reversal, Stoner–Wohlfarth model. The magnitude of the fields required for switching is defined by an astroid equation:

$$H_{\mathrm{h}}^{2/3} + H_{\mathrm{e}}^{2/3} \geqq \left[\frac{2K}{M_{\mathrm{s}}}\right]^{2/3} \tag{1}$$

where H_{e} and H_{h} are the magnetic fields applied along the easy and hard axis directions, K is an effective anisotropy constant and M_{s} the saturation magnetization. The magnetic fields are generated by a current flow in two perpendicular running bit and digit lines as shown in figure 1. Only the bit at the intersection of these two lines will be subjected to the field combination necessary for the free layer switching. This coherent reversal can be achieved in a few nanoseconds (3–7 ns) [12, 13]. The magnetization direction of the reference pinned layer is fixed by direct exchange coupling with an antiferromagnetic material.

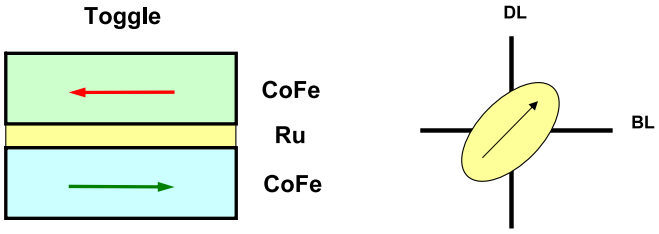

Figure 2. Typical storage layer used for the toggle–Savtchenko writing approach. The memory cell is generally elliptical and oriented at 45° with respect to the writing lines.

1.1. Selectivity and switching field distribution

The scalability of MRAM to smaller bit sizes is faced with challenges. Reducing the size of the tunnel junction below 0.1 μm^2 creates issues related to the device uniformity, control of switching field, power consumption and long term bit stability. At small dimensions, the magnitude of the switching field is critically governed by the precise shape and material anisotropy of the single bit. Upon decreasing the size and the bit aspect ratio, the width of the switching field distribution is enlarged. This is due to deviations from the nominal bit geometry or edge roughness that appear during the nanofabrication. Finding a set of fields which can be used to program all cells becomes difficult or unusable due to the very narrow operating window [14]. Ultimately the switching field depends on the dimensional control of the lithography and patterning process, which becomes less precise as the bit dimensions become closer to the minimum feature size. This becomes an important scaling limitation as the field write margin is reduced by the larger switching field distribution.

To overcome the selectivity issue the Stoner–Wohlfarth writing approach was replaced by a 'toggle' switching proposed by Savtchenko [15] at Freescale/Motorola. The usual storage layer used in this switching approach is a synthetic ferrimagnet (SF) free layer as shown in figure 2. The three-layer synthetic antiferromagnet structure is a magnetically rigid system consisting of two ferromagnetic layers separated by a thin Ru spacer layer. The coupled system responds to an applied magnetic field in the following manner: having some net anisotropy H_u in each layer, there exists a critical spin flop field H_{sf} at which the two antiparallel layer magnetizations will rotate (flop) to be orthogonal to the applied field H and with each other, scissoring in the direction of H (see step t_1 in figure 3). Detailed analytical treatment of the system can be found in recent publications [16, 17].

For fields $H \geqslant H_{sf}$ the SF can lower its total magnetic energy by decreasing its dipole energy with a flop and scissor. This can be used patterning the bit with its magnetic easy axis oriented at 45° to the orthogonal current lines. The programming pulse sequence and resulting magnetic behaviour are shown in figure 3. The arrows represent the magnetic moments of the two ferromagnetic layers in the free SF layer, the green arrow being the magnetization of the layer that is adjacent to the tunnel barrier which determines the junction resistance. To toggle the bit from one magnetic state to another, two pulses are used and the reversal is achieved in four steps (t_1, t_2, t_3 and t_4). In the first step (t_1) a magnetic field H_1 is applied at an angle of 45° to the bit easy axis. In step 2 (t_2) a second field H_2 is applied simultaneously with H_1. In step 3 (t_3) H_1 is turned off and only H_2 is acting on the magnetization of the SF. The applied field magnitudes are such that the total resulting field is higher than the flop field. When H_2 is turned off in step 4 (t_4) the magnetization relaxes to its easy-axis direction and the green arrow has a direction opposite to the initial state. This approach also improves the long term retention of the information, since the energy barrier between stable states of 'half-selected' bits subject

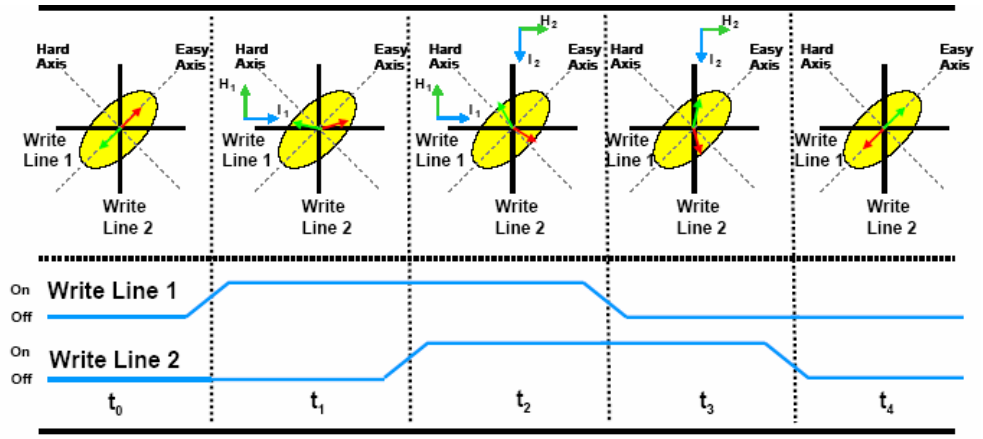

Figure 3. Schematic diagram of the toggle switching. Pulses are applied in a sequence that rotates the SAF by 180° to the opposite magnetization. This determines a change of the magnetoresistance, from [18].

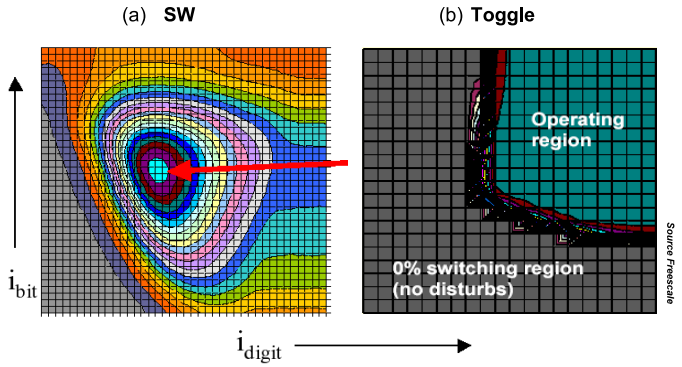

Figure 4. A switching map with the switching operating regions as a function of the current applied in the digit and in the bit line, for (a) a Stoner–Wohlfarth and (b) a toggle writing approach, from [18].

to only one field direction is increased, as opposed to the Stoner–Wohlfarth switching process. The percentage of successful writes as a function of the current in the digit and bit lines is shown in figure 4 for Stoner–Wohlfarth (SW) and toggle approaches. The operating region in toggle MRAM writing is considerably larger than the SW operating window. The toggle switching approach solves the bit selectivity issue, but the scalability of the cell size to deep sub-micron dimensions still remains. For sizes below 100 nm the problems of long term data retention and high power consumption still subsist in toggle switching.

1.2. Thermal stability and power consumption

To guarantee the bit information for long time periods it is necessary to take into account thermally activated switching processes, the so called superparamagnetic limit. The energy barrier E_b between two stable magnetic configurations is proportional to the volume of the magnetic material, to the shape and uniaxial anisotropies and the saturation magnetization M_s. As the bit size is reduced, this energy barrier may become comparable to the thermal

energy and unwanted thermally activated switching can occur, leading to data loss. The energy barrier is usually quoted as a dimensionless ratio to the thermal energy $\alpha = \Delta E/k_{\mathrm{b}}T$. To guarantee thermal stability required for 10 years data retention a minimum energy barrier of $\alpha \approx 65$ is required [18]. This creates a necessary compromise between write power consumption and bit stability. The magnetic volume reduction associated with scaling must be compensated by higher anisotropy, leading to higher switching fields and an increase in write power consumption.

2. Thermally assisted switching (TA-MRAM)

Recently a new write approach, called thermally assisted switching (TA-MRAM), was proposed to improve the thermal stability, write selectivity and power consumption in MRAM applications [19–21]. Heating had already been previously proposed to write the bit state of tunnel junctions by circulating a current in the write lines [22–24] and heating the magnetic layers in the MRAM cell above their magnetic ordering temperature, greatly reducing the write fields. The first designs proposed use a low Curie point ferromagnetic material as the storage layer [23, 22]. The bit is heated close to the Curie temperature and a small magnetic field generated by the digit line sets the write direction. Another design employs the exchange coupling between an antiferromagnet (AF) and a ferromagnetic (FM) layer as the storage mechanism [23]. The two films are heated above the Néel temperature of the AF layer and written by cooling down in the presence of a magnetic field. In this design, two orthogonal current lines are used to heat the cell and generate the write field. The coincident heating from two lines can be used to select one cell for thermal writing. The main disadvantage of such write architectures is that the heating is indirect and relies on heat diffusion, meaning high power consumption and long write cycles. A better approach to TA-MRAM [24, 25] is to heat directly with the current flow through the MTJ, after turning the selection transistor ON during the write procedure—figure 5(b). Since the heat is generated locally inside the junction this is a more efficient solution than the indirect diffusive heating.

In the scheme proposed by Spintec [25], the junction structure is modified inserting a second antiferromagnetic layer which exchange biases the storage layer. The write procedure implies heating the junction above the storage layer blocking temperature and cooling down in the presence of a magnetic field. The reference and the storage layer must be exchange biased at different blocking temperatures. This has multiple advantages and solves the limitations of the conventional MRAM architecture.

(i) As the write selection is temperature driven, a combination of magnetic field and heating current is needed to select a junction, virtually eliminating any addressing errors.

(ii) One single magnetic field is required to write, which leads to a reduced power consumption even in the presence of an additional heating current. The write power can be further reduced, by using circular elements with no shape anisotropy reducing the switching field. At the write temperature, the exchange energy is cancelled and the barrier height is reduced to the magneto-crystalline anisotropy energy, which is generally very low. The bit orientation is then defined by the field setting the direction of the storage layer during cooling [26].

(iii) The exchange bias anisotropy of the storage layer ensures the thermal stability and data retention for long time periods at the working temperature. In the TA-MRAM cell, the energy barrier ΔE at the working temperature is essentially determined by the exchange energy J_{eb}.

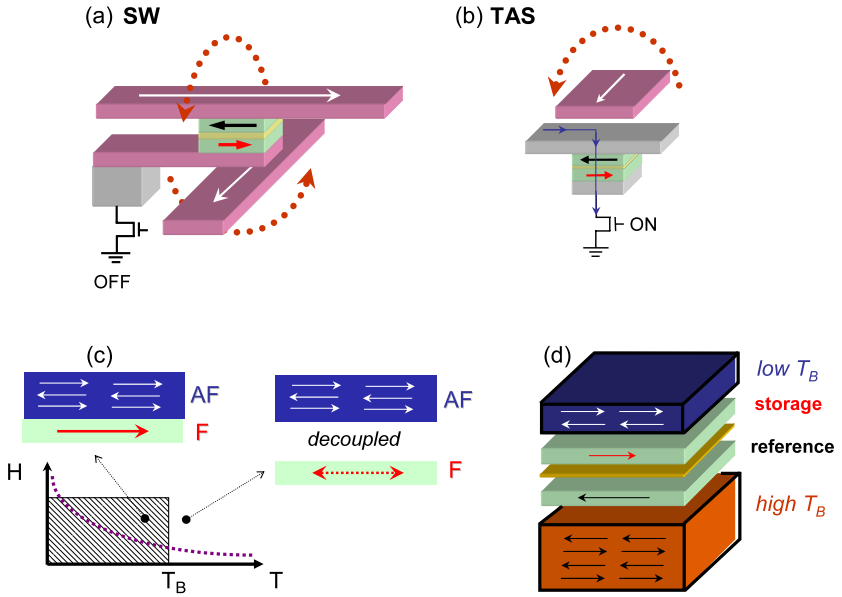

Figure 5. The writing procedure in a conventional MRAM architecture (a) and in the TA-MRAM architecture (b). In the TA-MRAM writing approach (c), the writing of the memory cell is possible only after overcoming the blocking temperature of the antiferromagnet which pins the storage layer (when the two layers are magnetically decoupled). (d) General view of a TA-MRAM stack with both the storage and the reference layers pinned with antiferromagnetic layers.

The total energy E of an elliptical bit can be written as the sum of the magneto-crystalline, shape and exchange bias anisotropies as

$$E \approx K + (AR - 1) \times \left(\frac{t}{L}\right) \times M_s^2 + \frac{(J_{eb} \times M_s^2)}{t} \times \left(1 - \frac{T}{T_b}\right) \tag{2}$$

where K is the magneto-crystalline energy, AR the aspect ratio, t is the thickness of the free layer, L is the long axis dimension of the bit, J_{eb} is the exchange energy, M_s the saturation magnetization, T the actual temperature and T_b the storage layer blocking temperature. The energy barrier of the elliptical MRAM bit, being inversely proportional to the bit length L, increases significantly as the cell size is reduced. Therefore the required writing field increases rapidly and consequently also the write current. The calculated write current of a conventional MRAM cell as a function of bit length L is shown in figure 6 (open circles). These values are compared to the writing currents needed to heat and switch a circular TA-MRAM cell (yellow dots). TA-MRAM has an advantage for bit lengths lower than 850 nm. However, large heating currents (red triangles) also require a large selection transistor, which limits the bit density. The region where the TA-MRAM cell is the most competitive even for high bit densities is for junction sizes smaller than 250 nm.

2.1. TA-MRAM writing sequence

The initial orientation of storage layer exchange biased by an IrMn AF layer (leftmost drawing in figure 7(a)) sets the bit in a low resistance state '0'. The storage layer loop is clearly biased at a positive field. The reversal of the storage layer bias is achieved by heating the IrMn layer above its blocking temperature with a current pulse and applying simultaneously an external

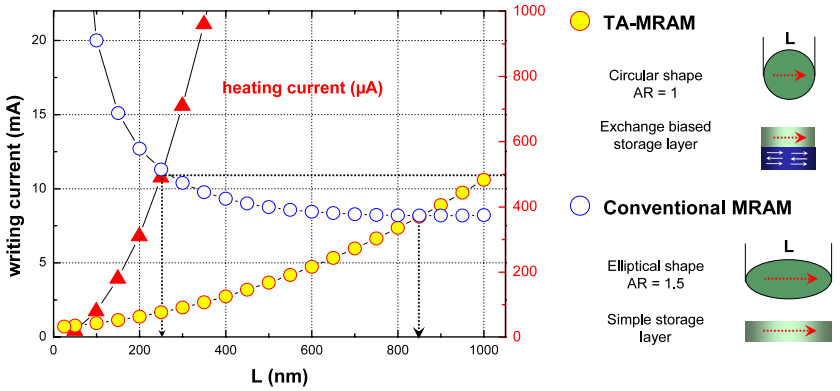

Figure 6. Variation of the total writing current for a memory cell versus the bit length L for a conventional Stoner–Wohlfarth MRAM architecture (white circles) and a TA-MRAM architecture (yellow/grey circles) left axis. For the conventional SW-MRAM the assumption of an elliptic structure with a normal (simple) storage layer was made, while for the TA-MRAM the cell was assumed to be of circular shape with an exchange biased storage layer. The total current represents the sum of the currents needed to produce two orthogonal magnetic fields for SW-MRAM and the sum of the current generating the magnetic field and the heating current in the TA-MRAM approach. The heating current is shown by red triangles (right axis). The dashed line delimitates the region in which the TA-MRAM junction is more favourable from the bit density point of view. Less current is needed to write the memory cell for the TA-MRAM approach below a bit length of 850 nm. The limiting parameter is the current which can pass through the selection transistor.

magnetic field H_{sw} larger than the coercive field of the storage layer—figure 7(b). The field is applied in a direction that favours the anti-parallel alignment of the storage and reference layers. The current pulse is terminated and the system is cooled in a magnetic field. This maintains the storage layer orientation and prevents freezing a vortex-like configuration induced by the circular symmetry of the Ampere field generated by the heating current. The result is a reversal of the pinning orientation of the storage layer and a bit state change to a high resistance '1' shown in the rightmost drawing of figure 7(c). As a result, the storage layer loop is now shifted towards negative values.

3. Heating in a magnetic tunnel junction

The principle behind the heat generation in a magnetic tunnel junction is both the Joule heating in the metallic layers and heating due to tunnelling electrons. Joule heating produces energy per unit time equal to ρj^2, where ρ is the electrical resistivity and j is the current density. Electrons tunnelling at the insulating barrier will arrive in the second ferromagnet as hot electrons. Their excess energy will be lost through inelastic scattering in the arrival electrode. This generates phonons and magnons [27] that macroscopically translate into a temperature increase ΔT. The energy carried by each electron is proportional to the voltage drop V and the number of electrons proportional to the current density j. The total heat generated Q can be expressed as

$$Q = \frac{jV}{l} \exp\left[-\frac{x}{\lambda_{\text{inel}}}\right] \tag{3}$$

where x is the stack position, V the applied bias voltage, and λ_{inel} the inelastic scattering mean free path of the electrons (see figure 8). With these assumptions the one-dimensional (1D) heat

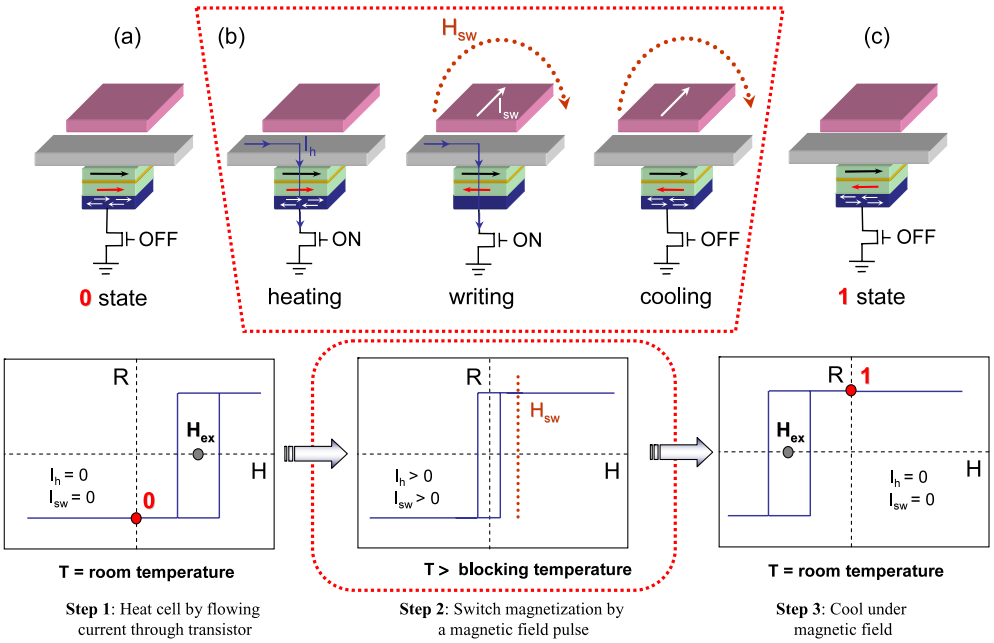

Figure 7. The writing steps of the thermally assisted switching approach (as described in the text).

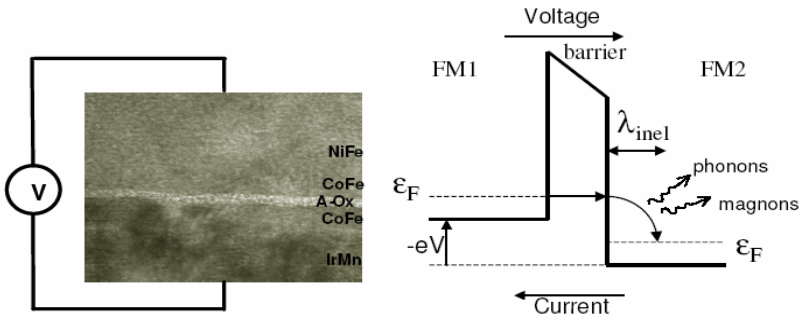

Figure 8. Heating process with a current injected through the junction barrier.

equation can be written as

$$c_p d\frac{\partial T}{\partial t} - K\frac{\partial^2 T}{\partial^2 x} = \rho j^2 + \frac{jV}{\lambda_{\text{inel}}}\exp\left[-\frac{x}{\lambda_{\text{inel}}}\right]. \tag{4}$$

The symbols have the following correspondence: c_p heat capacity, d mass density, T temperature, t time, K heat conductivity. This equation can be solved for the 1D case using a numerical finite difference method with boundary conditions. These simulations can be used to optimize the junction stack and look at the temperature distribution inside the multi-layer stack. However, the analysis of the heating process in a TA-MRAM cell requires solving a 3D geometry to fully take into account the heating occurring at the electrical contact leads.

The temperature increase in the TA-MRAM cell is proportional to the dissipated power density P_{d}. This proportionality between the temperature increase and the power density is shown in figures 9(a), (b). It can be seen that the exchange-bias field varies linearly

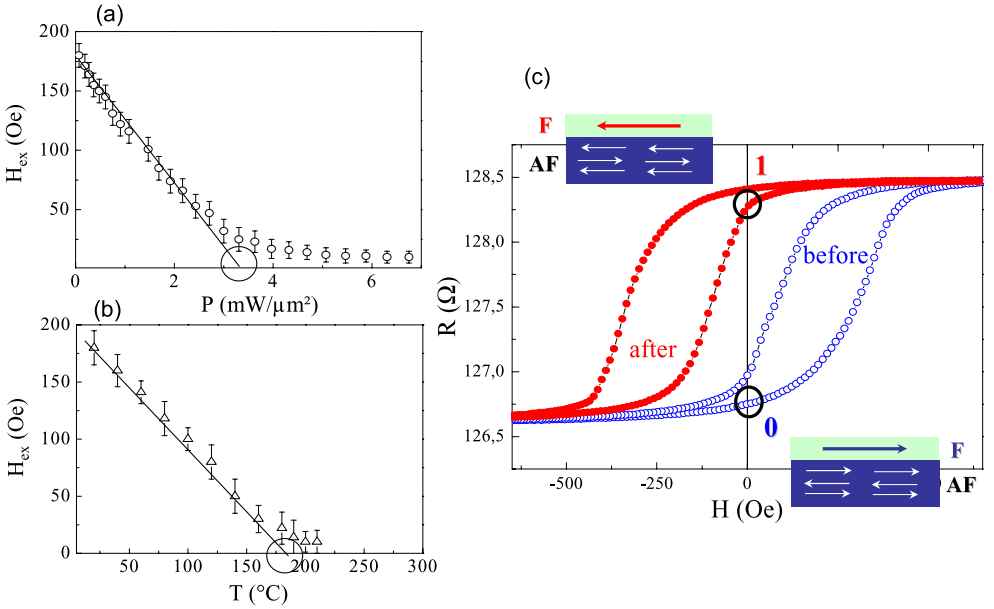

Figure 9. (a) Variation of the exchange field with the current passing through the junction and (b) upon heating the whole sample to a given temperature. (c) Demonstration of writing using DC current. TMR minor loops obtained prior to (open dots) and after (solid dots) the writing procedure.

both with the total heating power density and with temperature. The total power density is essentially determined by the $R \times A$ product of the junction and the current density j as $P_d = (R \times A)j^2$. The maximum power density is limited by the junction breakdown voltage as $j_{max} = V_{bd}/(R \times A)$. Maintaining a high V_{bd} at low $R \times A$ is essential for this application. Therefore, there is great interest in developing low $R \times A$ barriers with high breakdown voltage and high MR ratios for high current density MRAM applications [28, 29].

4. Demonstration of TA-MRAM write operation

The first demonstrations of TA-MRAM were realized in micron size junctions ($2 \times 2\ \mu m^2$) for DC currents [19], shown in figure 9(c), and current pulses down to 10 ns [20]. For the DC current experiments, the storage layer loop is shifted towards positive fields before writing and the junction is in a low resistance state at zero field. To write the junction a heating current corresponding to a DC power density of 3.6 mW μm^{-2} was applied in the presence of a 60 Oe field. The result is the reversal of the storage layer pinning direction, and a high resistance remanent state—point '1' in figure 9(c).

A large temperature difference in blocking temperatures for the reference and storage layers can be obtained selecting either a thick IrMn or a PtMn antiferromagnet, with blocking temperatures $T_b > 250\,°C$ for the reference electrode. Figure 10(a) shows the exchange field of the storage and reference electrodes. The storage layer is exchange biased by a thin 50 Å IrMn antiferromagnetic layer characterized by a relatively low T_b (150 °C). The pinning properties of the IrMn layer can be tuned by varying its thickness, as shown in figure 10(b). A maximum exchange field of about 250 Oe is obtained for a 50–60 Å thick IrMn layer.

The dynamic writing demonstration of TA-MRAM was realized on two different types of junction structures patterned to sub-micron dimensions using e-beam lithography and ion beam

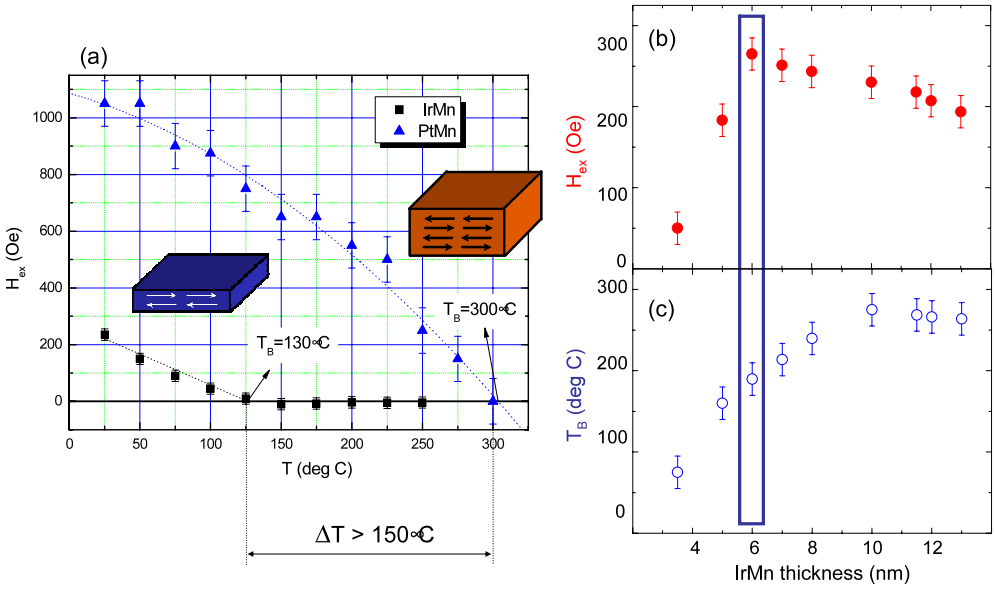

Figure 10. (a) Variation of the exchange bias field for the storage and reference layers pinned respectively by IrMn and PtMn antiferromagnets. A low blocking temperature of 150 °C characterizes the IrMn layer while the blocking temperature of the PtMn layer is much larger, about 300 °C. Plots of the exchange bias field (b) and the blocking temperature (c) as a function of the IrMn thickness.

etching. The structure stack and cell layout are shown in figure 11. The stack of the junction structure type 1 is Ta 30 Å/NiFe 70 Å/IrMn 60 Å/NiFe 30 Å/AlOx 6 Å/CoFe 35 Å/Ru 7 Å/CoFe 25 Å/IrMn 250 Å/Ru 30 Å/TiW(N) 150 Å. The 6 Å Al layer was oxidized by natural oxidation (1 Torr partial pressure of O_2 for 5 min), leading to an $R \times A$ of 35 Ω μm^2 and 10% TMR. An IrMn antiferromagnetic layer was used to pin both the storage and the reference layer, using different thicknesses to obtain different blocking temperature values. The measured storage and reference layer blocking temperatures were 160 °C and 210 °C, respectively. Junction areas ranged from 0.015 to 2.4 μm^2. The top and bottom connecting leads to the junction are 3 and 6 μm wide (figure 11(a)).

The stack of the junction structure 'type 2' is Ta 30 Å/CuN 700 Å/Ta 30 Å/PtMn 200 Å/CoFe 25 Å/Ru 8 Å/ CoFe 30 Å/Al 5 Å + Ox/CoFe 15 Å/NiFe 25 Å/IrMn 50 Å/Ta 1250 Å. The tunnel barrier is a 5 Å Al layer plasma oxidized for 15 s, leading to an $R \times A$ product of 17 Ω μm^2 and 2–4% TMR. The low TMR of these samples resulted from re-deposition during processing. The measured storage and reference layer blocking temperatures were 160 and 300 °C, respectively. Junctions were then patterned to sizes from 0.049 to 4 μm^2, with lead dimensions of 0.8 and 0.5 μm or 1 and 4 μm (figure 11(b)).

Transfer curves before and after writing a parallel (P) state are shown in figure 12 for the two junction types. The initial black curves are characterized by a shift of the storage layer loop towards positive fields. The parallel (P) state is set applying a positive external magnetic field and a heating current, shifting the storage layer loop towards negative fields (blue curves), while the reference layer loop remains unchanged. The storage layer loop of sample type 1 (figure 12(a)) shows a 560 Oe exchange bias field and 280 Oe coercive field. After writing the P state the loop is 'hidden', because the negative exchange bias fields of the reference and storage layers overlap. The dotted line shows the expected loop of the storage layer. The traces

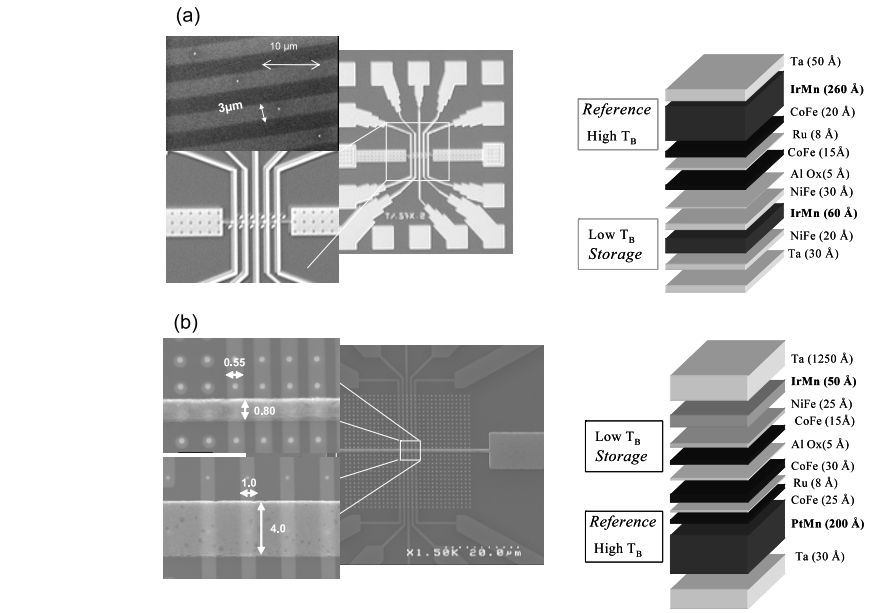

Figure 11. Different junction structures and stacks of the test structures used for the validation of the TA-MRAM approach: type 1 (a) and type 2 (b), described in more detail in the text.

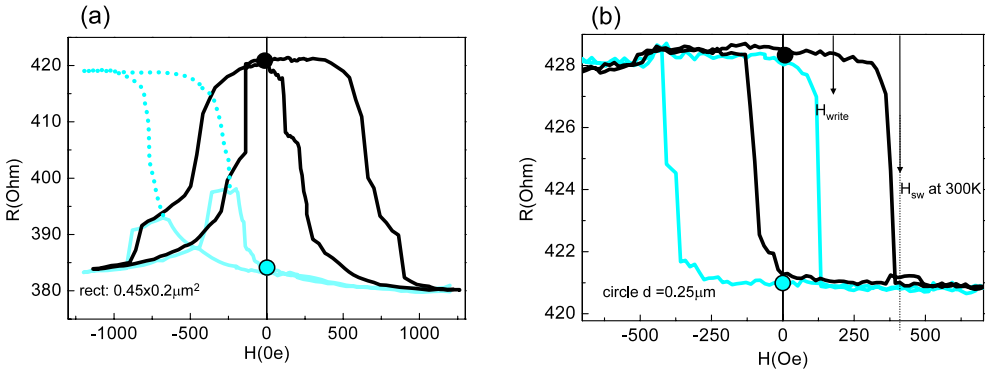

Figure 12. Transfer curves for samples type 1 (a) and type 2 (b) before writing, in parallel configuration at zero field (blue/light grey curve), and after writing the antiparallel configuration (black curve) at zero field.

show that for both P and AP written states there is only one defined resistance state at zero applied magnetic field. The condition for this is that for the storage layer loop $H_{eb} \geqslant H_c$. The advantage is that the P and AP resistances will remain unchanged even if the MTJ is subjected to external magnetic field perturbations.

A transfer curve corresponding to sample type 2 for a circular junction with diameter 0.25 μm is shown in figure 12(b). The exchange bias field of the reference layer is higher than the maximum 1400 Oe set-up available field. The 195 Oe exchange field (H_{eb}) is lower than the 225 Oe coercive field (H_c). Therefore, after writing P or AP states, with the corresponding resistance change and exchange field sign change, there are two possible resistance states at zero applied field. This means that sample type 2 is less stable against magnetic perturbations,

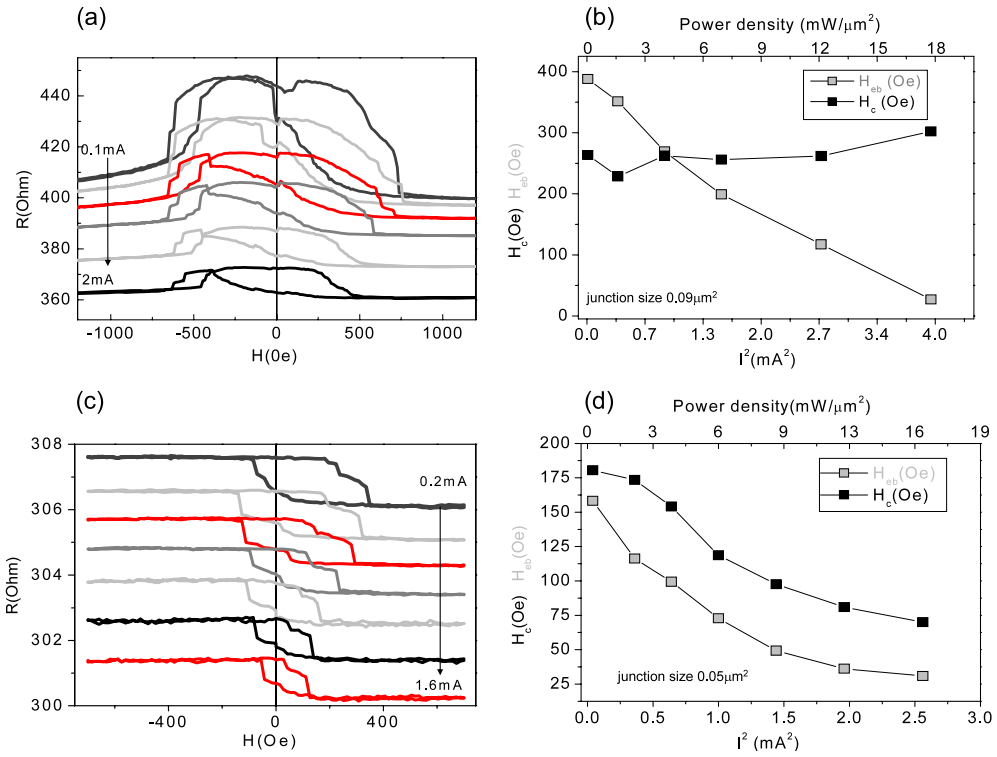

Figure 13. (a) Evolution of the TMR loop with increasing bias current and (b) H_{eb} and H_b as a function of I^2, which is proportional to the power density, for sample type 1. (c) Evolution of the TMR loop with increasing bias current and (d) H_{eb} and H_b as a function of I^2 for sample type 2.

since an external write field such that $H_{write} > H_{sw}$ could reverse the storage layer direction, leading to information loss even in the absence of a heating pulse.

The influence of a DC heating current on the TMR loop shape is shown in figure 13. For sample type 1 (figures 13(a), (b)), for a current of 0.1 mA the storage layer exchange bias is 400 Oe and the coercive field 280 Oe. The values of the exchange bias and coercive field are plotted in figure 13(b) as a function of the square of the injected current I^2, which is proportional to the total power density $P_d = \frac{R \times I^2}{A}$. The exchange field decreases gradually with increasing current. For a DC current of 2 mA, corresponding to a current density of 22 mA μm^{-2} and a power density of 17.6 mW μm^{-2}, the exchange bias is lost. The coercive field increases continuously from 260 to 305 Oe, while the reference layer exchange bias is reduced by only 40 Oe. For sample type 2 the initial exchange field is 160 Oe and the coercive field 180 Oe (see figure 13(d)) and both H_{eb} and H_c decrease with increasing current. For a DC current of 1.6 mA, corresponding to a power density of 17 mW μm^{-2}, H_{eb} and H_c decrease to 25 Oe (80% of the initial value) and 45 Oe (60% of the initial value), respectively.

5. TA-MRAM heating dynamics

In conventional MRAM architecture the writing time required for the alignment of the storage layer along the applied field is about 3–7 ns [12, 13]. The TA-MRAM write cycle should match these values to be competitive compared to conventional magnetic field induced switching

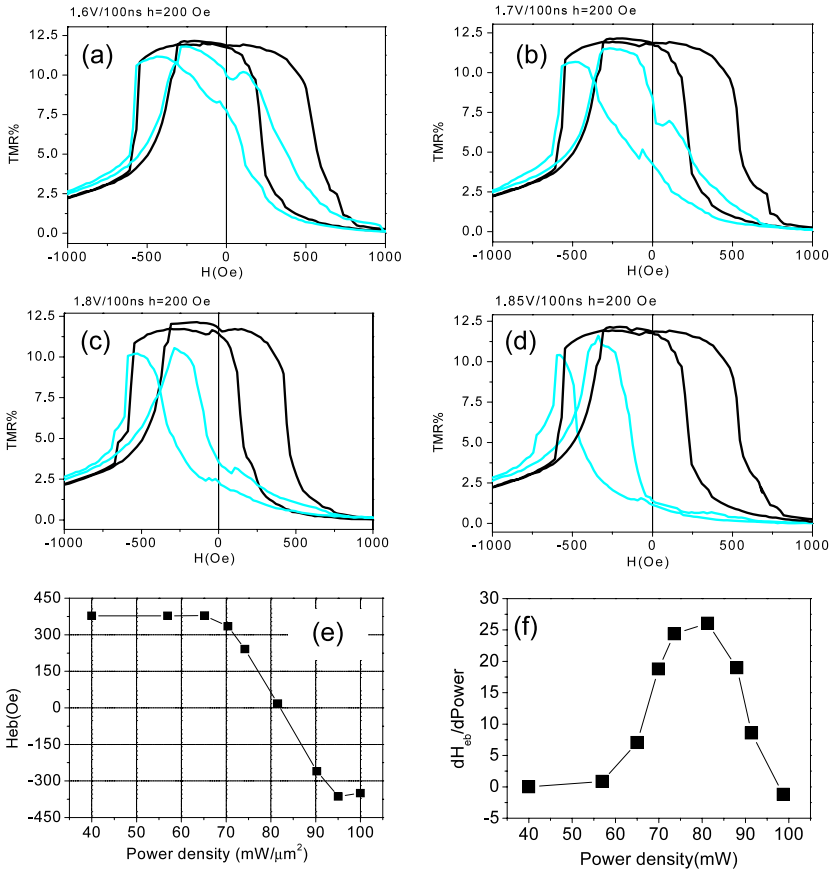

Figure 14. Evolution of the characteristic transfer curve for a 0.1 μm^2 MTJ cell of type 1 for an applied external field of 200 Oe, a pulse width of 100 ns and different pulse amplitudes: (a)–(c) incompletely written and (d) completely written (e) state H_{eb} versus power density (P_d) and (f) the derivative of H_{eb} as a function of P_d ($\frac{\partial H_{eb}}{\partial P_d}$) used to estimate the blocking temperature distribution.

MRAMs. The tunnel junction temperature can be monitored experimentally by looking at the exchange bias field of the pinned storage layer. The dynamic characterization of the heating process was realized using the storage layer writing to establish whether T_b was reached during the current pulse. The power needed to write for a certain pulse width was measured using the following procedure. A current pulse was applied to the junction simultaneously with a static magnetic field (200 Oe) opposing the initial direction of the magnetization of the storage layer. After each pulse application the transfer curve was measured and the magnetic configuration of the junction restored to its initial AP (high resistance) state. For a constant pulse width the heating pulse voltage was increased until complete switching of the storage layer was achieved. A typical sequence is shown in figures 14(a)–(c) for varying pulse amplitudes and constant 100 ns pulse width. A pulse amplitude of 1.6–1.8 V is not enough to heat the junction above the blocking temperature and the storage layer loop is not completely shifted. For a writing pulse of 1.85 V (corresponding to a heating power density of 96 mW μm^{-2}) the junction is considered as completely written since the H_{eb} of the storage layer loop after writing is completely shifted towards negative fields as shown in figure 14(d). These intermediate states can be observed

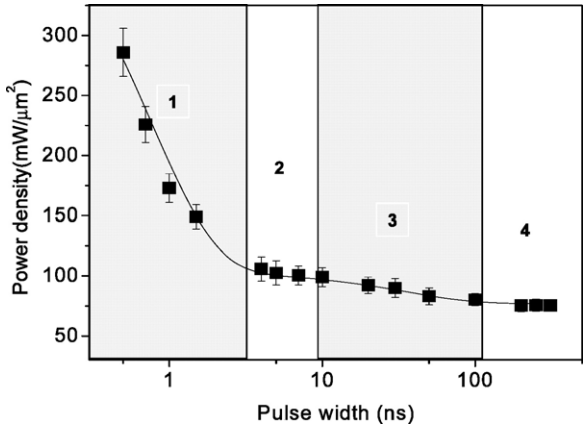

Figure 15. The variation of the minimum heating power density required to write a 0.1 μm^2 junction of sample type 1 as a function of heating pulse width.

since there is a distribution of T_b. This distribution can be estimated by plotting the H_{eb} values of the storage layer as a function of temperature and taking the derivative $\partial H_{eb}/\partial T$. Since $\Delta T \propto P_d$, H_{eb} was plotted as a function of the power density P_d in figure 14(e). The derivative $\partial H_{eb}/\partial T$ yields a dispersion of P_d of about 21 mW μm^{-2} for the junction presented above—figure 14(f). Considering that $\Delta T \propto P_d$ and that 96 mW μm^{-2} corresponds to a temperature increase of about 160 °C, 21 mW μm^{-2} results in ΔT_b dispersion equal to 35 °C.

5.1. Dependence of the heating power density on the heating pulse width

The writing process has been characterized by studying the variation of the minimum heating power density for a large range of heating pulse widths (from 500 ps to 250 ns). For each pulse width, the pulse amplitude was varied in order to find the minimum power density needed to heat the junction above the blocking temperature of the storage layer. The write voltage has to be increased from 1.7 to 1.9 V upon reducing the pulse width from 200 ns down to 10 ns. Accordingly, the associated heating power density $P_d = V^2/(R \times A)$ increases from ≈80 to ≈100 mW μm^{-2}. The dependence of the minimum heating power density on the heating pulse width is shown in more detail in figure 15.

There are mainly two remarks which can be made. First, the shortest pulse for which complete writing was experimentally observed was 500 ps. This means that in 500 ps the junction temperature was increased above $T_b \approx 160$ °C. For shorter pulse widths, the heating pulse voltage would have to be higher than the breakdown voltage, and the junction would be irreversibly damaged. Second, four different heating regions can be distinguished on this curve—a sharp increase of the power density for pulses shorter than 4 ns in region 1, region 2 between 5 and 10 ns, where the power density is almost constant, and a slow decrease of the power density in region 3 between 20 and 100 ns, followed by almost constant power density in region 4, for pulses larger than 150 ns. In order to understand this variation of the power density with the pulse width and the physical origin of these regions, we performed numerical simulations of the time resolved heat equation. Generally, if at $t = 0$ a power density P_d is delivered to a tunnel junction, the temperature within the junction evolves according to the following equation:

$$T = T_0 + P_d \frac{A}{K}\left[1 - \exp\left(-\frac{t}{\tau_0}\right)\right] \tag{5}$$

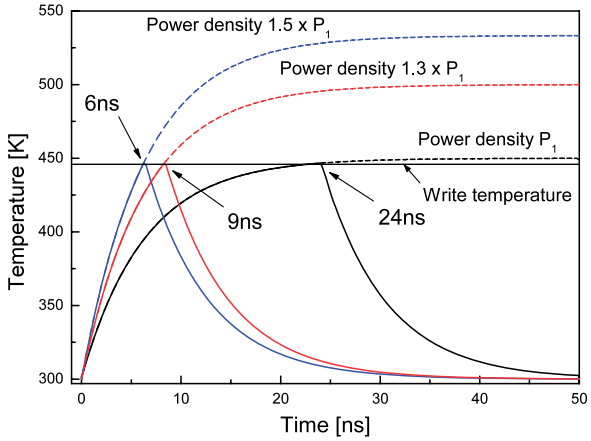

Figure 16. Temperature evolution in time for an MTJ using different heating conditions (power density, pulse width).

where T_0 is the temperature at $t = 0$, A the junction area, K the proportionality constant between power density and temperature increase of the system and τ_0 the characteristic heating time. Using the substitution

$$\Delta T = \frac{P_\mathrm{d} \times A}{K} \tag{6}$$

where P_d is the power density, equation (5) can be written as

$$T = T_0 + \Delta T \left[1 - \exp\left(-\frac{t}{\tau_0}\right)\right]. \tag{7}$$

For $t > 5\tau_0$ the temperature of the system reaches its equilibrium temperature, which for our writing condition is equal to the blocking temperature, $T \approx T_0 + P_\mathrm{d}\frac{A}{K} = T_\mathrm{b}$. When the pulse width is reduced so that $t_{\text{pulse width}} < 5\tau_0$, P_d has to be increased in order to reach the write temperature T_b within the pulse duration time.

Figure 16 shows the general temperature evolution of a system in general for three different heating power densities.

The first condition is a 24 ns heating pulse with a heating power density P_1, corresponding to the black curve in figure 16. The characteristic heating time of the system is taken to be $\tau_0 = 7$ ns. In this case P_1 is sufficient to increase the temperature of the junction above 448 K, considered the write threshold temperature. As the pulse width is sufficiently long (almost $5\tau_0$), the system reaches its equilibrium temperature T_b at the end of the pulse application (in 24 ns). If we want to achieve the same temperature T_b during a shorter heating pulse, e.g. 9 ns—red curve, we have to increase the power density needed by a factor of 1.3 with respect to the precedent case ($1.3 \times P_1$). Decreasing furthermore the pulse width to 6 ns (blue curve), the write temperature can be reached only with $1.5 \times P_1$.

The results obtained experimentally were confirmed by thermal simulations realized with a commercial finite element heat equation solver FEMLAB in a full 3D approach (see figure 17). It was assumed that there is no sidewall heat diffusion. The geometry and the dimensions used in the simulations are displayed in figure 17(a). The top lead is an Al wire of length $L = 9$ μm, width $w = 6$ μm and thickness $t = 0.3$ μm. For the bottom lead, the length has the same value, the width $w = 3$ μm and the thickness $t = 0.15$ μm. The leads are thermalized at the sidewalls and at the bottom.

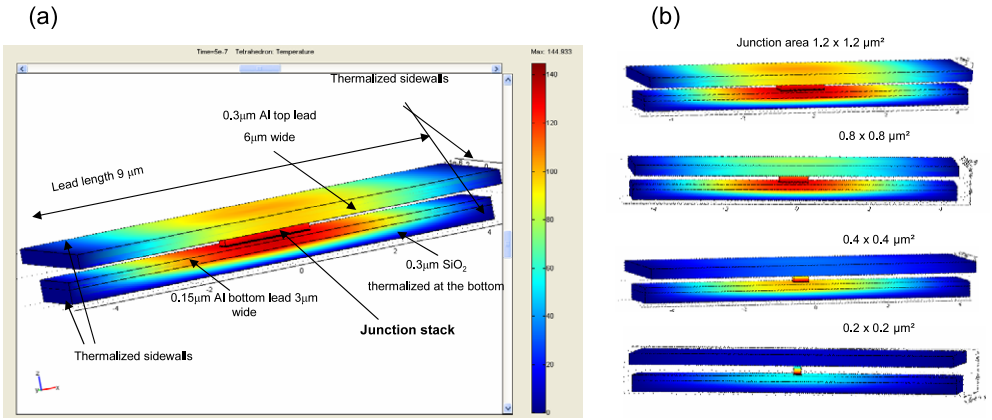

Figure 17. (a) The geometry of the leads used in 3D simulations; (b) 3D thermal simulation with FEMLAB for different junction areas using a constant power density of 40 mW μm^{-2} showing the temperature distribution in the system.

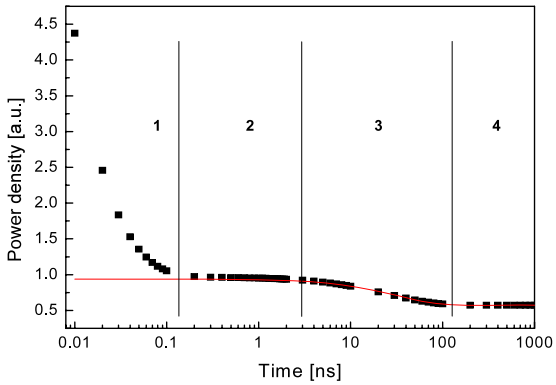

Figure 18. The calculated variation of the power density as a function of heating time.

Equation (5) was solved to calculate the power density required for a given temperature increase at a given time t. This allows an easy comparison of the simulation with the experimental data. Thermal simulation results plotting the power density values necessary to reach a 120 K temperature increase as a function of the heating time are shown in figure 18.

The simulation confirms the existence of the four distinct regions observed experimentally. Region 1 is associated with the intrinsic heating of the tunnel junction in an adiabatic process, as the generated heat is essentially confined to the junction barrier region. The intrinsic equilibrium temperature of the MTJ is reached in region 2 followed by the temperature increase of the leads in region 3, as the heat generated in the tunnel junction dissipates into the connecting leads. In the power density plot this translates into a lower power density necessary to reach the same temperature threshold. Finally in region 4 the equilibrium temperature is reached in the whole system junction and leads. The solid line shows a fit of the power density using the expression

$$P_{\text{write}} = P_{\text{ad}} - \Delta P_{\text{d}} \left[1 - \exp\left(-\frac{t}{\tau_1} \right) \right] \tag{8}$$

where P_{write} is the required write power density, P_{ad} is the intrinsic write power density, P_{d} the

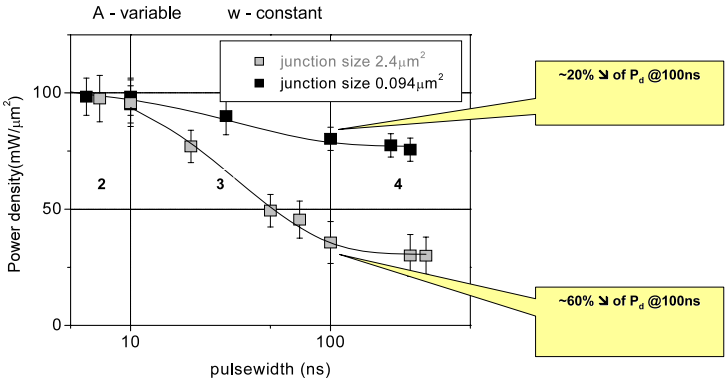

Figure 19. Measured power density needed to heat the MTJ cell for different heating pulse widths. The junction area is varied to 0.094 and 2.4 μm^2; it is connected to leads of 4 μm at the bottom and 6 μm at the top.

power density decrease due to the heating contribution from the leads and τ_1 the characteristic lead heating time equal to 40 ns in our simulation. The intrinsic tunnel junction heating in region 1 can be fitted with a similar expression:

$$P_{\text{write}} = \frac{P_{\text{ad}}}{1 - \exp(-\frac{t}{\tau_0})} \tag{9}$$

with a characteristic heating time τ_0 of 40 ps deduced from the fitting. As a conclusion, the heating of a junction is a relatively fast process: a temperature increase of 120 °C can be achieved in less than a nanosecond. If the junction is to be heated in a time comparable to the intrinsic heating time constant τ_0, the heating power density has to be increased. However, the maximum power density that can be injected experimentally is limited by the breakdown voltage of the tunnel junction.

5.2. Influence of the junction area and lead widths on the heating process

To study experimentally the influence of the leads on the heating process, two types of configurations were measured where either the width w of the lead or the junction area A was varied.

The first case corresponds to junctions with w constant and A variable. Figure 19 shows results on the variation of the heating power density as a function of the pulse-width. The measurements have been made on two different junction sizes, one with an area of 0.094 μm^2, and the second with an area of 2.4 μm^2. The leads had in both cases the same width (4 μm for the bottom and 6 μm for the top lead). First, it can be noticed that the same power density is required for both samples in region 2 (pulses shorter than 10 ns), before the onset of the lead heating. In the case of the 2.4 μm^2 surface area junction, data for pulse widths shorter than 5 ns could not be measured because higher power densities implied voltages beyond the barrier breakdown. Second, the power density values obtained for long pulses (in regions 3 and 4) are lower for the larger area junction. For 100 ns pulses, P_{write} decreases by 60% with respect to the intrinsic power density value P_{ad}, while the decrease is only 20% for the smaller junction. This decrease is more pronounced for longer pulses around 250 ns when the entire system, including the leads and the junction, achieves its equilibrium temperature.

The results obtained experimentally were confirmed by thermal simulations realized considering a layer stack of Ta 100 Å/IrMn 60 Å/NiFe 40 Å/AlOx 10 Å/NiFe 100 Å/Al

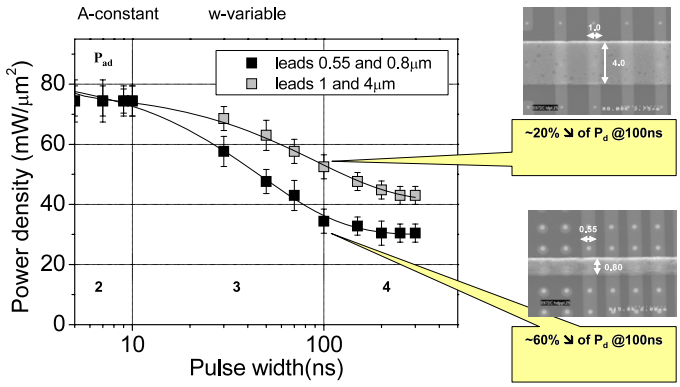

Figure 20. Measured power densities needed to write the junction for different heating pulse widths. The junction size is 0.05 μm^2, connected to μm sized (1 and 4 μm) or sub-μm sized (0.55 and 0.8 μm) leads.

2800 Å. The temperature increase was calculated for a power density of 40 mW μm^{-2} in the interval ranging from 1 ps up to 1000 ns. Figure 17(b) illustrates the temperature distributions corresponding to four different junction areas: 1.2×1.2 μm^2, 0.8×0.8 μm^2, 0.4×0.4 μm^2 and 0.2×0.2 μm^2. It can be observed that the lead temperature has greatly increased in the large area junction and only slightly in the small area junction. The heating of the electrical leads elevates the base temperature of the junction. Since in all cases the lead volume is the same, the lead temperature increase is higher for the larger junctions. This results in a lower write power density, in agreement with the results obtained experimentally.

The second case corresponds to w variable and A constant. A similar effect was observed by keeping the junction size constant and varying the width of the leads, as shown in figure 20. Here junctions of 0.05 μm^2 surface area were connected both to sub-micron sized Cu leads (0.55 and 0.8 μm) and to larger micron sized Cu leads (1 and 4 μm). The main characteristics of the observed behaviour are a required P_{write} for narrow leads that was only 32 mW μm^{-2} for 100 ns pulse width, much smaller than what is obtained for a sample connected to large leads, which required 58 mW μm^{-2}. Also, the characteristic heating time constant of the whole system changes from 90 to 45 ns. This can be explained by the larger volume of the μm size leads. Finally, for 100 ns pulses P_{write} decreases by 20 and 60% with respect to the adiabatic values. This decrease of the P_{write} could be eventually be used for the minimization of the write power density in TA-MRAM cells using sub-μm sized lines and sub-μm sized junctions. The percentage decrease of P_{write} for long pulses is related to the ratio between the junction area and lead width as well as to the thermal conductivity of the lead material. When the values of w and A are comparable, the decrease is significant (about 60% of the P_{ad} value). For lead dimensions much larger than the junction area, this decrease of the P_{write} is less significant. The minimum heating power density values obtained as a function of the junction area are summarized in figure 21.

Upon decreasing the junction area, the minimum power needed to write decreases linearly for the small junctions with areas below 0.1 μm^2. The corresponding power density value is found to be constant and independent of the junction area. This is an important advantage for TA-MRAM scalability. The dependence of the heating power density as a function of the surface area can be fitted using a constant power density threshold P_{ad}, obtained for small junction areas. The straight lines define constant power density levels of 200 mW μm^{-2}, 100 mW μm^{-2}, 30 mW μm^{-2} and 20 mW μm^{-2}. For micron sized junction areas, the power

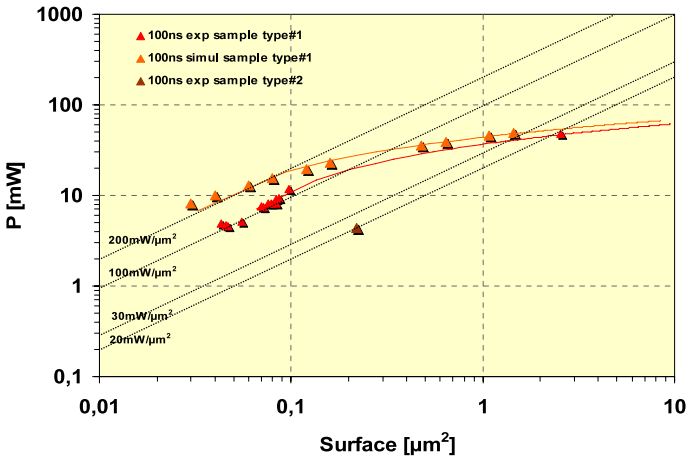

Figure 21. Measured and simulated power needed to heat junctions of different areas for a writing time of 100 ns. The continuous line fits for 100 ns pulse widths are a linear fit for sample sizes below 0.3 μm^2 and of the form $P = P_0 + \frac{cte}{A}$ for larger sample sizes.

density required to heat the junction above the blocking temperature deviates from the constant power density model. This can be understood by the heating occurring in the leads as the heat flows away from the junction into the leads. As a consequence, the power density required to achieve the same ΔT will decrease. However, the total absolute power required to elevate the junction temperature by ΔT continues to decrease with the junction area. The power densities needed for dynamic writing are still relatively high between 20 and 75 mW μm^{-2}. The minimum heating power for longer pulses (100 ns) can be reduced by making the dimensions of the junction area and the leads comparable. However, this power decrease is only observed for long pulses and in this case there is a risk of thermal crosstalk in the memory due to the line heating.

5.3. Minimization of the heating power by using thermal barriers

Another way to reduce the power density is to insert two layers of low thermal conductivity materials at both ends of the magnetic tunnel junction stack, serving as thermal barriers between the junction and the electrical leads (for example BiTe thermal barrier with 1.5 W (m K)$^{-1}$ thermal conductivity). In this way the thermal crosstalk is significantly decreased. In figure 22, the experimental results obtained for a sample without thermal barrier are compared with the simulation obtained with and without thermal barrier.

Simulation shows that the use of thermal barriers minimizes significantly the thermal losses and power densities lower than 20 mW μm^{-2} are achievable. Furthermore, thermal barriers also provide better scalability as the write power density is constant and independent of the junction size. The thermal barriers also minimize the risk of thermal cross-talk since lead heating is reduced. Electrical and magnetic properties of a tunnel junction with the integrated thermal barrier have already been tested at INESC-MN [30, 31]. The authors obtained 20% TMR and a $R \times A$ of 1.2 k Ω μm^2 per barrier for the double barrier junctions with a GeSbTe thermal barrier. The lowest heating power density obtained was 0.3 mW μm^{-2} for a DC writing current and μm sized junctions.

It is also worth mentioning that the optimum operating region as far as the heating is concerned is region 2 (between 10 and 4 ns). This region is the most favourable from a

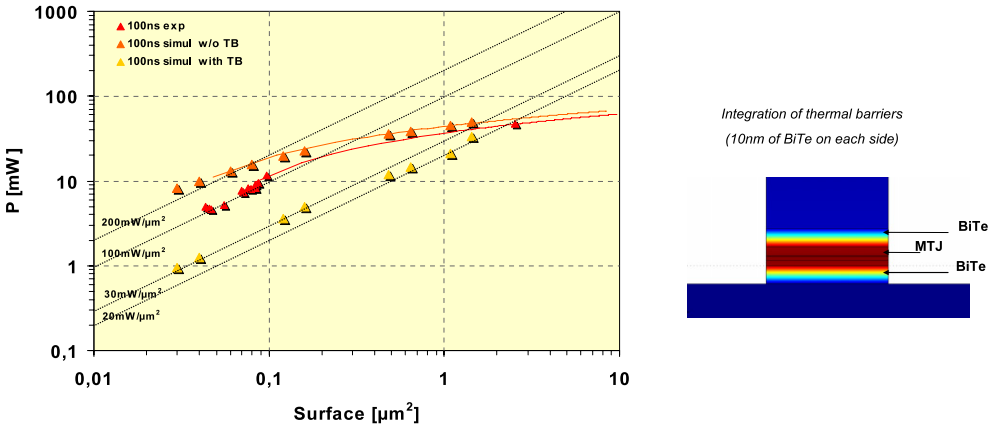

Figure 22. Measured and simulated minimum power needed to heat up to the write temperature as a function of the junction area. Junctions with and without thermal barriers are compared for heating pulse widths of 100 ns. To the right is a scheme of an MTJ with integrated thermal barriers.

consumption point of view. In this region the power density is independent of junction size and there are no visible effects of lead heating, therefore no risk of thermal crosstalk. The origin of the variations in P_{ad} for the different samples cannot be explained by lead heating. Usually there is a thick metallic layer deposited between the actual active junction and the connecting leads. This metallic material of ≈100 nm thickness is used as a hard mask in the patterning process. The energy loss which has been observed in regions 1 and 2 can be attributed to the heating of this embedded hard mask material. When materials of different thermal conductivity are used, as with the copper and tantalum hard masks used ($K_{Cu} = 401$ W (m K)$^{-1}$, $K_{Ta} = 58$ W (m K)$^{-1}$), the corresponding energy loss varies and consequently the intrinsic power density P_{ad}. We compared results obtained for a sample with a thick (125 nm) Cu hard mask and one with a thick 150 nm Ta layer. The measured P_{ad} values are 100 mW μm^{-2} and 75 mW μm^{-2} respectively (see figure 23). The simulations and the experimental results show that P_{ad} increases if the thermal conductivity of the material is higher. Moreover, the heating power density can also be reduced in region 2 by using materials with lower thermal conductivity, as is the case of thermal barriers with very low thermal conductivity like BiTe ($K_{BiTe} = 1.5$ W (m K)$^{-1}$) and GeSbTe ($K_{GST} = 0.5$ W (m K)$^{-1}$). As can be seen from the simulations P_{ad} can be reduced to a quarter of P_{ad} obtained for the case of Ta.

6. Write selectivity and protection against stray fields

One of the important advantages of the TA-MRAM writing approach is that the layer stack used already has some protection against stray magnetic field erasure. This means that, due to the exchange biasing of the storage layer, the P and AP resistances will remain unchanged, even if the MTJ is subject to magnetic field perturbations. Here, the protection against magnetic erasure of the thermo-magnetic write scheme is demonstrated at room temperature for a 0.45 × 0.2 μm^2 size junction. In figure 24(a) the resistances of the P and AP states are shown upon cycling, using a magnetic perturbation field of ±400 Oe. The resistances are measured each time in zero magnetic field. Without a heating pulse, both P and AP states are stable and insensitive to magnetic field perturbations. This unique feature of the TA-MRAM approach offers a high protection against stray magnetic fields even in the absence of any

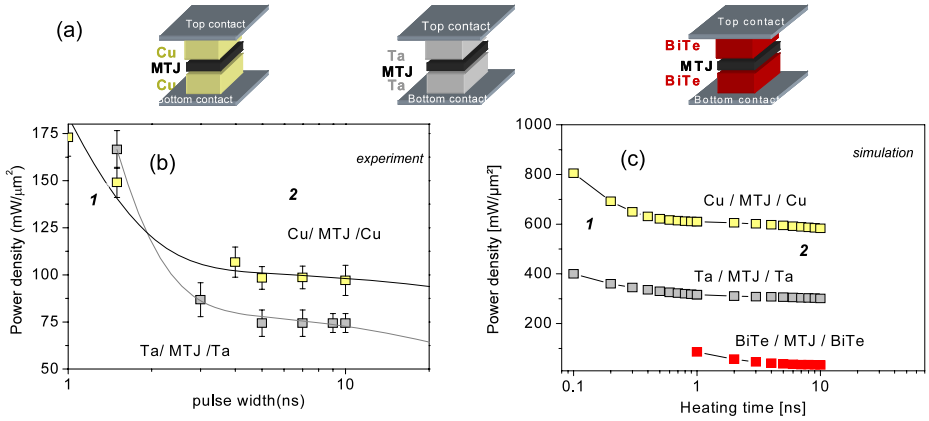

Figure 23. (a) Measured power densities which are needed to write the junction in regions 1 and 2 for the junctions embedded with Cu (sample type 1 yellow/light grey squares) or with Ta (sample type 2 grey squares). (b) Calculated power densities compared for junctions with Cu (yellow/light grey squares) or Ta (grey squares) hard masks or with a thermal barrier of BiTe (red squares).

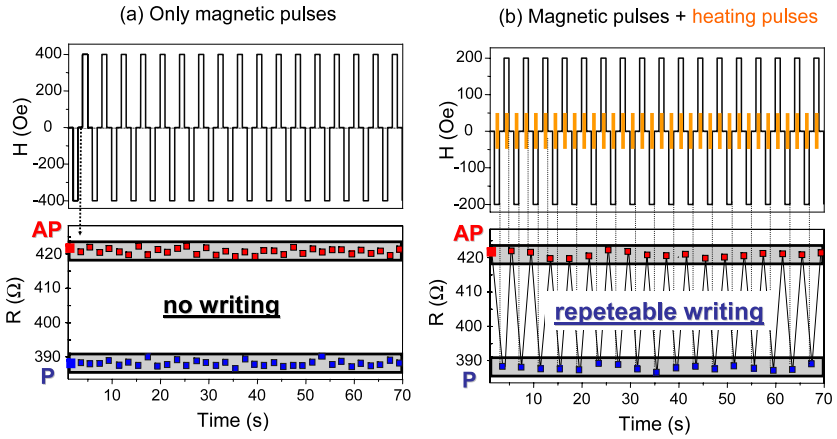

Figure 24. Demonstration of the selectivity and protection against stray magnetic fields in the TA-MRAM writing scheme. Only with simultaneous current pulse heating is it possible to change the resistance state. (a) Measurement of P and AP state stability under the application of a magnetic perturbation of ±400 Oe. Plot (a) shows the zero field values of the resistance in the P or AP states without pulse application. (b) Measured zero field values of the junction resistance and (bottom) sequence of applied magnetic fields and the position of the 20 ns heating pulses of 2 V.

magnetic shielding. Selective writing of the memory cell is achieved, applying simultaneously a current pulse and a ±200 Oe magnetic field. This is demonstrated in figure 24(b), applying a heating pulse of 2 V for 20 ns. Each time a heating pulse and magnetic field pulse are applied to the junction, the temperature in the storage layer exceeds T_b of the 60 Å IrMn antiferromagnetic layer and the magnetization is switched. The memory cell is correctly written as the resistance jumps each time between two well defined resistance levels corresponding to the P and AP states. This measurement highlights one important feature of the thermo-magnetic write scheme: excellent selectivity as only heated junctions can be magnetically written.

7. Conclusion

The concept of thermo-magnetic writing (TA-MRAM) in the dynamic regime was tested and validated. The submicron sized MTJs can be written using 500 ps heating pulses, proving that the TA-MRAM thermo-magnetic approach is competitive with the conventional scheme in terms of write speed. The influence of the voltage pulse width, junction area or lead volume on the required write power density was evidenced. The power densities needed for short heating pulses (shorter than 10 ns) are still relatively high. One way to reduce the power density is by reducing the heat flow away from the junction by making lead lines comparable in size with the junction area which causes lead heating and an increase of the junction base temperature. The other possibility is to optimize the tunnel junction heating process by inserting two layers of low thermal conductivity materials at both ends of the MTJ layer stack, serving as thermal barriers between the junction and the electrical leads (for example BiTe or GeSbTe thermal barriers). The thermally assisted approach offers a promising solution for the next generation of MRAM as it can solve most of the current issues: write selectivity, power consumption and thermal stability, whilst offering full scalability to the 65 nm node and beyond. The integration of thermal barriers is the last technological hurdle before a real implementation on large CMOS arrays.

Acknowledgments

This work was supported by the IST project NEXT, contract number 2001-37334. The authors would like to acknowledge P P Freitas and B Ocker for fruitful discussions, comments and shared data. This paper is dedicated to the late Marta Kerekes.

References

[1] Schwee L J, Hunter P E, Restorff K A and Shephard M T 1982 *J. Appl. Phys.* **53** 2762
[2] Pohm A, Daughton J, Comstock C, Yoo H and Hur J 1987 *IEEE Trans. Magn.* **23** 2575
[3] Tehrani S *et al* 1999 *J. Appl. Phys.* **85** 5822
[4] Daughton J and Chen Y 1993 *IEEE Trans. Magn.* **29** 2705
[5] Naji P K, Durlam M, Tehrani S, Calder J and DeHerrera M F 2001 *ISSCC Digest of Technical Papers* p 122
[6] Scheuerlein R 2000 *ISSCC Digest of Technical Papers* p 128
[7] Moodera J S, Kinder L R, Wong T M and Meservey R 1995 *Phys. Rev. Lett.* **74** 3273
[8] Parkin S S P, Fontana R E and Marley A C 1997 *J. Appl. Phys.* **81** 5521
[9] Parkin S S *et al* 2004 *Nat. Mater.* **3** 862
[10] Yuasa S, Nagahama T, Fukushima A, Suzuki Y and Ando K 2004 *Nat. Mater.* **3** 868
[11] Durlam M, Tehrani S, Calder J, DeHerrera M F and Naji P K 2000 *ISSCC Digest of Technical Papers* p 130
[12] Koch R H *et al* 1998 *Phys. Rev. Lett.* **81** 4512
[13] Yamamoto T *et al* 2005 *J. Appl. Phys.* **97** 10P503
[14] Yoshikawa M *et al* 2005 *J. Appl. Phys.* **97** 10P508
[15] Savtchenko L, Engel B N, Rizzo N D, DeHerrera M and Janesky J 2003 Method of writing to scalable magnetoresistance random access memory element *US Patent Specification* 6,545,906 B1
[16] Worledge D C 2004 *Appl. Phys. Lett.* **84** 4559
[17] Fujiwara H, Wang S-Y and Sun M 2005 *J. Appl. Phys.* **97** 10P507
[18] Akerman J *et al* 2004 *Magnetoelectronics* (Amsterdam: Elsevier)
[19] Prejbeanu I *et al* 2004 *IEEE Trans. Magn.* **40** 2625
[20] Wang J and Freitas P P 2004 *Appl. Phys. Lett.* **84** 945
[21] Sousa R C *et al* 2004 *J. Appl. Phys.* **95** 6783
[22] Beech R S, Anderson J A, Pohm A V and Daughton J M 2000 *J. Appl. Phys.* **87** 6403
[23] Daughton J M and Pohm A V 2003 *J. Appl. Phys.* **93** 7304
[24] Abraham D W and Trouilloud P L 2001 Thermally-assisted magnetic random access memory (mram) *US Patent Specification* 6,385,082 B1

[25] Dieny B and Redon O 2003 Magnetic device with magnetic tunnel junction, memory array and read/write methods using same *Patent Specification* 2003043017
[26] Nozières J-P, Dieny B, Redon O, Sousa R C and Prejbeanu I L 2004 Magnetic memory with a magnetic tunnel junction written in a thermally assisted manner and method for writing the same *Patent Specification* 2005086171
[27] Zhang S, Levy P M, Marley A C and Parkin S S P 1997 *Phys. Rev. Lett.* **79** 3744
[28] Jiang Y *et al* 2004 *Phys. Rev. Lett.* **92** 167204
[29] Meng H, Wang J, Diao Z and Wang J-P 2005 *J. Appl. Phys.* **97** 10C926
[30] Cardoso S *et al* 2006 *J. Appl. Phys.* **99** 08N901
[31] Ferreira R, Cardoso S and Freitas P P 2006 Dynamic thermomagnetic writing in tunnel junction cells incorporating two GeSbTe thermal barriers *IEEE Trans. Magn.* **42** 2718–20