Design considerations and strategies for high-reliable STT-MRAM

W.S. Zhao\textsuperscript{a,b,*}, T. Devolder\textsuperscript{a,b} Y. Lakys\textsuperscript{a,b}, J.O. Klein\textsuperscript{a,b}, C. Chappert\textsuperscript{a,b}, P. Mazoyer\textsuperscript{c}

\textsuperscript{a}IEF, Univ. Paris-Sud 11, Orsay 91405, France
\textsuperscript{b}UMR8522, CNRS, Orsay 91405, France
\textsuperscript{c}STMicroelectronics, 850 Rue Jean Monnet Crolles, Grenoble 38026, France

1. Introduction

Magneto-resistive random access memory (MRAM) is regarded as a promising non-volatile memory candidate and it features high speed, infinite endurance and great scalability etc. \cite{1}. The first generation of MRAM based on field induced magnetic switching (FIMS) has been commercialized since 2006 and it plays a more and more important role for high reliability applications, particularly in the aerospace, defence, nuclear and transport sectors \cite{2}. Nevertheless, FIMS limits greatly the interest of MRAM for wide applications, as two high currents (>10 mA) are required for the magnetic field generation \cite{3}. This causes high programming power, low storage density and difficult integration with CMOS process. Spin Transfer Torque (STT) is the switching method to build up the second generation of MRAM and overcome all the drawbacks of FIMS \cite{1,4,5}; it allows MRAM to be used in both standalone and embedded memory applications. Only a bi-directional small current (<150 μA, 65 nm node) can switch the state of MTJ.

A number of pre-industrial prototypes have been demonstrated since 2005 \cite{7,8} and it is expected to be commercialized in the next few years. However, the STT approach causes reliability degradation compared to FIMS such as erroneous write by read-current \cite{9} and high error rate of reading circuits due to low Resistance Area (RA) value etc. They become major obstacles to the use of STT-MRAM. In this paper, we address these issues from circuit design point of view and present some efficient solutions. By using a STT-MTJ compact model (see Fig. 2) for reliability prediction \cite{10} and CMOS 65 nm design kit \cite{11}, mixed transient and statistical simulations have been performed.

2. Reliability of MTJ switched by STT

2.1. Theoretical analysis of erroneous sensing

As shown in Fig. 1, a low current \((I_{\text{switch}})\) can switch the state of MTJ as it is higher than the threshold current \(I_c\) which can be approximately calculated by the Slonczewski theoretical model \cite{5}. However, according to thermal activation \cite{12} (Eqs. (1)–(3)), a small current lower than \(I_c\) can reverse erroneously the state of MTJ.

\begin{equation}
\tau_1 = \tau_0 \exp \left( \frac{\Delta E}{k_B T} \left( 1 - \frac{I}{I_c} \right) \right)
\end{equation}

\begin{equation}
\Delta E = \frac{\mu_0 M_s \times V \times H_C}{2}
\end{equation}

\begin{equation}
Pr = 1 - \exp \left( -\frac{\text{Duration}}{\tau_1} \right)
\end{equation}
than the practical coercive field = 25 mT for cell size ~100 nm², and Duration is the read-current pulse duration.

Fig. 3 shows the read disturb margin related to characteristic energy barrier $\Delta E$ (see Fig. 3) storing the binary data in free layer with partial perpendicular anisotropy or exchange bias [14–16] instead of planar anisotropy, however they degrade other performances like the sensing margin or the switching speed.

From circuit design point of view, we can avoid read disturb by reducing the read-current value and its duration. Pre-charge dynamic sensing is an efficient method (see Fig. 4), in which the sensing operation is driven by the control signal “PC” and dynamic current $I_{\text{read}}$ passes through the MTJ during ~200 ps per operation [17] (see Fig. 5). This method can reduce greatly the total cumulated read time and then lower the erroneously reading rate.

In the conventional STT-MRAM cell, there is a word selection transistor per MTJ, which controls both the sensing and writing operations [7–9]. It should be a large transistor (e.g. 4 F²) to pass through enough current for state switching and this explains the scaling difficulty for current STT-MRAM proposals. Another unexpected effect of this transistor is to pass relatively high $I_{\text{read}}$, 29 $\mu$A or 0.19 $I_c$ in 65 nm node, which is superior to the disturb margin 0.18 $I_c$ for $\Delta E = 40 k_B T$ shown in Fig. 3 and could cause erroneous reverse issue. In order to reduce $I_{\text{read}}$, we propose to use two word selection transistors (Word_1R, Word_1W) per MTJ cell for respectively reading and switching operations.

The reading transistor can be thus designed with minimum size (1 F²) and this reduces systematically $I_{\text{read}}$ down to ~100 $\mu$A or 0.06 $I_c$ (see Fig. 5), which is far below the disturb margin. Compared with the conventional STT-MRAM cell design based on 1T + 1MTJ structure [7–9], the adding of a minimum sized

2.2. Reliability improvement through dynamic sensing

In order to keep the data up to 10 years without read disturb issue, a number of solutions are under investigation. From device point of view, the most efficient way to improve the characteristic energy barrier $\Delta E$ (see Fig. 3) are storing the binary data in free layer with partial perpendicular anisotropy or exchange bias [14–16] instead of planar anisotropy, however they degrade other performances like the sensing margin or the switching speed.

From circuit design point of view, we can avoid read disturb by reducing the read-current value and its duration. Pre-charge dynamic sensing is an efficient method (see Fig. 4), in which the sensing operation is driven by the control signal “PC” and dynamic current $I_{\text{read}}$ passes through the MTJ during ~200 ps per operation [17] (see Fig. 5). This method can reduce greatly the total cumulated read time and then lower the erroneously reading rate.

In the conventional STT-MRAM cell, there is a word selection transistor per MTJ, which controls both the sensing and writing operations [7–9]. It should be a large transistor (e.g. 4 F²) to pass through enough current for state switching and this explains the scaling difficulty for current STT-MRAM proposals. Another unexpected effect of this transistor is to pass relatively high $I_{\text{read}}$, 29 $\mu$A or 0.19 $I_c$ in 65 nm node, which is superior to the disturb margin 0.18 $I_c$ for $\Delta E = 40 k_B T$ shown in Fig. 3 and could cause erroneous reverse issue. In order to reduce $I_{\text{read}}$, we propose to use two word selection transistors (Word_1R, Word_1W) per MTJ cell for respectively reading and switching operations.

The reading transistor can be thus designed with minimum size (1 F²) and this reduces systematically $I_{\text{read}}$ down to ~100 $\mu$A or 0.06 $I_c$ (see Fig. 5), which is far below the disturb margin. Compared with the conventional STT-MRAM cell design based on 1T + 1MTJ structure [7–9], the adding of a minimum sized

Fig. 1. MTJ is mainly composed of three thin films (e.g. CoFeB/MgO/CoFeB): two ferromagnetic layers and one oxide barrier. It presents two resistance values depending on the relative magnetization of two ferromagnetic layers. Through the STT mechanism, a low bi-directional current ($I_{\text{switch}}$) can switch the MTJ between two states.

Fig. 2. STT-MRAM compact model for write/read circuit and reliability simulation; ‘a’ and ‘b’ are respectively the length and width of the elliptical MTJ.

Fig. 3. Disturb read margin of STT-MRAM ($Pr = 10^{-3}$) depends on the read-current value and its duration. $\tau_{\text{cumulated}}$ is the cumulated read-current pulse duration.

Fig. 4. Pre-Charge Sense Amplifier (PCS A) circuit for STT-MRAM. Two word selection transistor per MTJ. $I_{\text{read}}$ is designed from free layer to reference layer and MTJ_ref is in anti-parallel state and with larger size than storage cells.
transistor dedicated for sensing will lead to area overhead, which is about 25% for 65 nm node. However, it can be ignorable for cross point memory design [18] where the selection transistors are shared by a number of MTJ in the same word (see Fig. 10).

2.3. Design considerations related to asymmetric switching and resistance variation

Design considerations should be addressed also to the asymmetric switching of STT between the two states. As the threshold current $I_{c}$ from parallel to anti-parallel is higher than the reversing one [12] (see Fig. 2), the direction of $I_{\text{read}}$ should be designed from free layer to reference layer (see Fig. 4) to create a wider margin between $I_{c}$ and $I_{\text{read}}$.

The fabrication and integration of MTJ at the back-end process of CMOS drive high variation of its characteristics like resistance value in the wafer level. Thereby it is preferable to use a MTJ device as reference instead of a constant resistance to overcome this issue. This MTJ can be set in anti-parallel state to obtain the best sensing margin as its state will not be changed accidently by $I_{\text{ref}}$, which passes from free layer to reference layer. Its value can be obtained by an appropriate size with Eqs. (4), (5) and we can find that this method promises higher density than using parallel state as reference, where the minimum fabrication node is set to MTJ instead of the MTJ for storage. Nevertheless the reference resistance should be calculated correctly taking into account the bias voltage dependence to avoid the shrinking of sense margin [10].

$$R_{\text{ref}}(V_{\text{bias-sense}}) = R_{\text{parallel}} + R_{\text{anti-parallel}}(V_{\text{bias-sense}})$$

$$R_{\text{MTJ}} = \frac{223.76 \times \varphi^{1/2} \times \text{surface}}{\text{tox}} \times \exp(1.025 \times \text{tox} \times \varphi^{3/2})$$

(4)

where $\varphi$ is the potential barrier height of MgO = 0.4, tox is the height of oxide barrier = 0.85 nm, surface is the area of MTJ and $V_{\text{bias-sense}}$ is the bias voltage of MTJ during the sensing.

3. Reliability of Sense Amplifier (SA)

For the MTJ switched by STT approach, a low RA value (e.g. 10 Ohm $\mu$m$^2$) is often expected to protect the oxide barrier for the period of switching, which cannot sustain high voltage (e.g. 1 V for 1.3 nm MgO barrier). However this low RA will degrade greatly the sensing hardness of SA to mismatch variation of transistors as the $I_{\text{read}}$ difference between two branches becomes smaller. Furthermore, the Tunnel Magneto-Resistance ratio (TMR) = $(R_{\text{para}} - R_{\text{anti}})/R_{\text{para}}$ is also reduced due to the low RA and it is difficult to achieve an expected value like 300% from the technological point of view [6]. This leads to worse margin between $I_{\text{read}}$ and $I_{\text{ref}}$ (see Fig. 4).
Figs. 6 and 7 (rectangular dot line) show respectively the reading error rate of PCSA according to different RA value and TMR ratio through Monte-Carlo statistical simulation [19]. For example, as RA is as low as 10 Ohm \( \mu \text{m}^2 \) and TMR ratio is reduced down to 150%, 16 errors have been observed for 100 run simulation. This confirms the poor performance of MTJ sensing reliability caused by low values of RA and serious design considerations should be taken into account. Two solutions have been found to overcome efficiently this reliability issue.

3.1. Triple Modular Redundancy (TMR)

The first one is to implement TMR logic block [20] at the output of the SA. This technique is basically used to mitigate Single Event Effects (SEE) (i.e. soft errors) as the majority vote eliminates errors occurring on one of triplicate outputs. TMR technique allows also the improvement of the sensing reliability as shown in Figs. 6 and 7 (circular dot line). As RA is 10 Ohm \( \mu \text{m}^2 \) and TMR ratio is 150%, zero error can be achieved during 100 runs. However this approach costs important CMOS overhead as TMR logic block includes two additional SAs and a voting circuit (see Figs. 8 and 4), which represents about quadruple die area of PCSA circuit composed of only minimum transistors.

3.2. Transistors sizing

The second method is to increase directly the size of active transistors (MN0-1, MP0-1) for current amplification in the PCSA circuit. Figs. 6 and 7 (triangle dot line) confirm that the sensing error due to mismatch variations can be decreased greatly with the larger transistors in PCSA. Nevertheless the sensing circuit area should be triple of the minimum size of PCSA to ensure error-free reading for 10 Ohm \( \mu \text{m}^2 \) RA and 150% TMR ratio.

Despite the fact that the CMOS overhead is more important, TMR technique is preferred as it improves the hardness of STT-MRAM to both soft error (radiation or harsh environment) and hard error (device mismatch). As mentioned above (see Figs. 4 and 8), a PCSA can be shared by a number of MTJs in STT-MRAM.
and this mitigates the CMOS overhead for reliability improvement. For instance, if there are 128 MTJs per PCSA, the TMR logic block represents only 13.3% overhead.

3.3. Symmetric layout implementation

Another important element related to the sensing reliability of SA is the parasite capacitances due to the asymmetric implementation of the two branches (see Fig. 4). This is more critical for pre-charge sensing approach, which depends on the discharge mechanism to detect the state of MTJ [17]. To avoid these parasite capacitances, symmetric layout implementation is required to connect the PCSA and MTJ cells. Fig. 9 shows an example of symmetric implementation, where even and odd words are connected in the two sides of PCSA and two reference MTJs are used to sense them separately. Fig. 10 shows the layout implementation of this circuit. For example, to read the state of MTJ1, we activate the transistors Word_1R and EN_EW at the same time and close all the other transistors in the two sides, the currents \( I_{\text{read}} \) and \( I_{\text{ref}} \) allow the MTJ1 to be detected after the pre-charge operation. As there are already the same number of closed transistors in the two sides, the parasite capacitances can be neglected.

4. A high reliable cross-point STT-MRAM array

Based on the considerations and strategies shown in the above sections, we developed a cross-point memory array performing high reliability (see Fig. 10) [21]. This new structure presents good trade off among the density, speed and power beyond its high reliability. It operates parallel sensing/switching for the MTJs in the same word and series operations for different words in the memory array. The cross point array [22] allows the CMOS overhead for reliability enhancement to be well mitigated.

5. Conclusion and perspectives

In this paper, we presented design considerations and strategies to address different reliability issues of STT-MRAM. New MTJ/CMOS interface structures have been firstly designed to reduce the erroneous switching caused by the reading current. A number of methods like transistor sizing, Triple Modular Redundancy and symmetric physical implementation have been presented and compared to relax the reading errors caused by the transistor mismatch variation and radiation. Benifiting from these studies, we investigated a new STT-MRAM array based on cross point architecture, which is under prototyping in our lab. Mixed transient and statistical simulations using a STT-MTJ compact model and CMOS 65 nm design kit confirm the effectiveness of our solutions.

Acknowledgements

The authors wish to acknowledge support Nano2012 project with STMicroelectronics and the French national project NANOIN-NOV-SPIN. We thank G. PRENAT and B. DIENY from SPINTEC, K.TORKI and G. DIPENDINA from CMP for decisive inputs scientific discussions and crucial help.

References