

STT-MRAM for embedded memory applications from eNVM to Last Level Cache.

Luc Thomas, Guenole Jan, Son Le, Santiago Serrano-Guisan, Yuan-Jen Lee, Huanlong Liu, Jian Zhu, Jodi Iwata-Harms, Ru-Ying Tong, Sahil Patel, Vignesh Sundar, Dongna Shen, Yi Yang, Renren He, Jesmin Haq, Zhongjian Teng, Vinh Lam, Paul Liu, Yu-Jen Wang, Tom Zhong, and Po-Kang Wang
 TDK- Headway Technologies, Milpitas CA, USA.
 email: luc.thomas@headway.com

Abstract—*Spin-Transfer-Torque Magnetic Random Access Memory (STT-MRAM) is emerging as a leading candidate for a variety of embedded memory applications ranging from embedded NVM to working memory and last level cache. In this paper, we review recent breakthroughs that have brought perpendicular STT-MRAM to the cusp of mass production.*

Keywords—*STT-MRAM, memory technology, embedded memory, data retention.*

I. INTRODUCTION

Perpendicular Spin-Transfer-Torque Magnetic Random Access Memories (pSTT-MRAMs) combine fast read/write, low voltage operation, low power consumption, non-volatility and quasi-infinite endurance [1]. Because of the physics of STT used for writing bits [2-4], Magnetic Tunnel Junction (MTJ) devices at the heart of the technology can be tailored to emphasize low current/high speed, data retention and/or high operation temperature, depending on the specific application. This versatility makes pSTT-MRAM an ideal candidate for next generation “universal” embedded memory, potentially capable of replacing technologies spanning from embedded Flash to SRAM [5]. However, the emergence of pSTT-MRAM as universal embedded memory has been hindered by several major technical challenges. First and foremost, the MTJ devices must withstand temperatures used in CMOS backend processing without degradation of their magnetic properties. Second, to become a competitive alternative to e-flash, pSTT-MRAM chips must retain data during the reflow soldering operation used to package the chips. Third, in order to compete with SRAM at advanced nodes, pSTT-MRAM technology must demonstrate reliable switching at array level at nanosecond time scale. In this paper, we review the recent advances in our group at TDK-Headway that have allowed us to overcome these hurdles [6-14]. All the MTJ film stack development, device fabrication, and 10MB test chip integration has been done in our backend semiconductor facility.

II. MAGNETIC TUNNEL JUNCTION STACK

MTJ stacks consist of two magnetic layers separated by an oxide tunnel barrier (Fig. 1). Readout depends on the relative orientation of the magnetization of the two magnetic

electrodes, owing to the Tunnel Magnetoresistance (TMR) effect. Parallel alignment leads to a low resistance state (logic 0), whereas antiparallel alignment gives rise to a high resistance state (logic 1). The on/off ratio of today’s state-of-the-art MTJ stacks using crystalline MgO tunnel barriers is typically between 2 and 3, although values up to 7 have been achieved in specially designed structures [15]. The two magnetic electrodes are designed to serve as storage (“free”) and reference (“pinned”) electrodes, respectively. The free layer is written by taking advantage of the Spin Transfer Torque (STT) mechanism arising from current pulses flowing directly across the MTJ device.

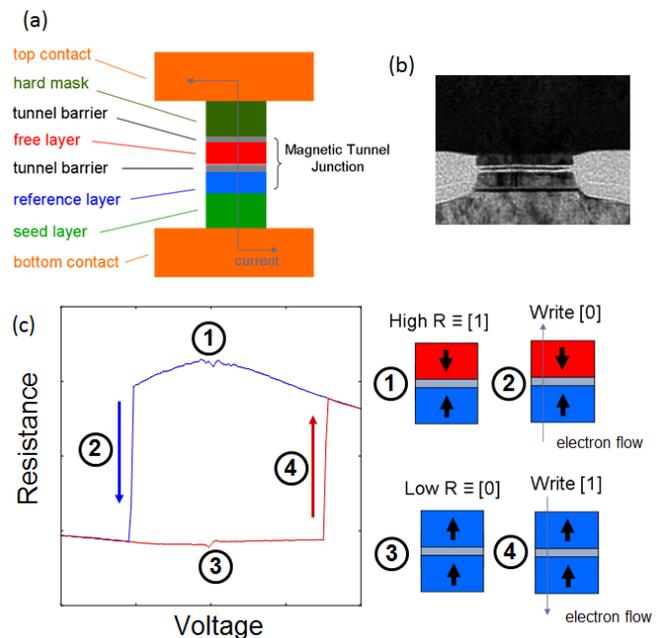


Fig. 1. (a) Schematic description of the Magnetic Tunnel Junction device. (b) Example of a Transmission Electron Micrograph of a typical MTJ device. (c) Example of the typical resistance vs voltage characteristics of a MTJ device. Current passing through the device is used to both read (1 and 3) and write (2 and 4). Writing direction (0 to 1 or 1 to 0) is controlled by the current polarity.

Both data retention and writing efficiency are much improved for MTJs stacks magnetized perpendicular to the plane of the layers (i.e. pMTJs) [16]. Today’s industry

standard for pMTJ stacks is based on CoFeB/MgO bilayers. As was discovered a few years ago, there is a strong interfacial anisotropy at the CoFeB / MgO interface, which allows the CoFeB layer's magnetization to be magnetized perpendicular to the plane of the layer when its thickness is small enough (typically less than 1 nm) [17,18]. In order to maintain perpendicular orientation for thicker CoFeB layers, so as to improve thermal stability of the storage layer, MTJ stacks comprising two interfaces have been developed (Fig. 1b) [6,7,19]. Since it is challenging to control the growth and oxidation of such MgO/CoFeB/MgO structures, more complex composite CoFeB/M/CoFeB trilayers have been developed, in which M is a non-magnetic or weakly magnetic metal such as Ta [20] or a Ta-based alloy [21]. The reference layer of the pMTJ stack also requires extensive material engineering to achieve adequate electrical and magnetic properties, e.g. small stray fields, high stability, high TMR, high spin polarization. Combining all these features typically requires using materials having different crystal structures in the same stack, while maintaining low roughness and high crystalline quality. Consequently, pMTJ stacks suitable for STT-MRAM have become complex. For example, the full stack reported by Tohoku University at the 2014 IEDM comprises more than 30 layers [22].

III. MTJ STACKS COMPATIBLE WITH 400°C BEOL PROCESSES

Standard backend-of-line (BEOL) CMOS processes such as low-k dielectric deposition or forming gas annealing are performed at 400°C. Thus, for pSTT-MRAM to become a viable option for embedded applications, the MTJ stack must withstand this temperature for an extended period of time. Exactly how long depends on the metal layer at which pSTT-MRAM is fabricated, relative to the total number of metal layers. For high-density applications, it is preferable to position the embedded memory just above CMOS, where lithographic features are the smallest. In this case, the total time at 400°C may exceed 3 hours.

The challenges of submitting a multilayered MTJ stack comprising many atomically thin layers to such a thermal treatment can be easily understood: high temperature annealing leads to recrystallization and grain growth, which can impact the roughness and microstructure of the stack. Mitigating the effect of solid state diffusion of elements within each layer and between layers is crucial to achieve the required thermal tolerance. Moreover, it is also important to note that BEOL 400°C processes must take place after patterning the MTJ stack. Thus, sidewall damage, etch residues and encapsulation materials also play a major role. Controlling these edge effects will become even more important at and below 1X lithography nodes, for which device diameters smaller than 30nm will likely be needed.

Solving these problems requires thorough engineering of not only the MTJ stack constituting layers, but also of the top and bottom electrodes and encapsulation materials, as well as process conditions. Since we gave the first demonstration of full 400°C compatibility at the 2013 MMM conference [7], other groups have reported such results [23]. Fig. 2 shows the

resistance versus magnetic field hysteresis loop of a 30nm device fabricated in our fab and annealed for more than 2.5 hours at 400°C after patterning. This device exhibits a TMR ratio of 175% for a resistance area product of 8.5 Ohm.μm². The hysteresis loop is square, exhibiting well-defined high and low resistance levels. Moreover, the device's switching field exceeds 3000 Oe, with negligible asymmetry.

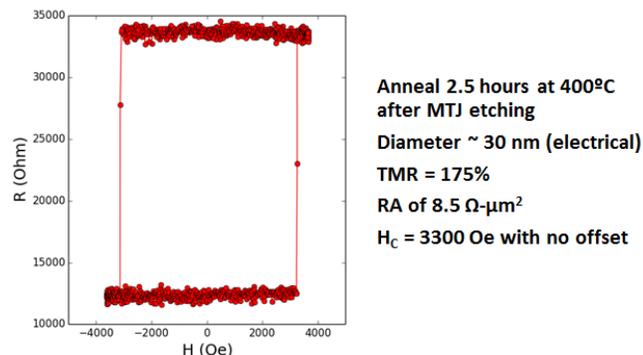


Fig. 2. Resistance versus field hysteresis loop of a 30 nm device which has been annealed at 400°C for 2.5 hours. Device diameter is calculated from its resistance and the MTJ stack's resistance-area product RA.

IV. STT-MRAM QUALIFIED FOR REFLOW SOLDERING

STT-MRAM is an attractive alternative to e-flash at and below 40nm lithography node. Indeed, pSTT-MRAM is faster and operates at lower voltage than e-flash, while at the same time meeting e-flash data retention requirements. Moreover, pSTT-MRAM is much cheaper to fabricate than e-flash. It only requires 2 to 3 additional masking layers, compared to more than 10 layers for e-flash. Thus pSTT-MRAM has a significant cost advantage, in particular for small memory sizes and/or small volumes. Many e-flash customers require the program code to be loaded in the chips before the chips are packaged and soldered to a printed circuit board. The soldering process uses reflow soldering for a total of 90 seconds at 260°C. In order to target this market, pSTT-MRAM must retain data during solder reflow with a low error rate.

The challenge facing pSTT-MRAM for reflow soldering qualification can be understood by considering the expression of the thermal stability factor Δ , which determines data retention (the error rate decreases exponentially with increasing Δ): $\Delta = E_B/k_B T$ where k_B is the Boltzmann constant, T the absolute temperature (in K) and E_B the energy barrier separating parallel 0 and antiparallel 1 states. In order to reach error rates below 10 ppm after 90 seconds, Δ must be larger than 36.7 at 260°C. However, since E_B is a function of the perpendicular anisotropy of the free layer, which is strongly temperature dependent, increasing temperature from room temperature up to 260°C has a twofold effect. First, the increase in temperature at the denominator reduces Δ by 46%. Second, E_B is also reduced. We have reported that this combined effect leads to an almost linear decrease of Δ over a wide range of temperatures, which can be approximated by a variation coefficient between 0.2 and 0.45 per degree,

depending on device diameter and MTJ stack properties [9, 11]. In the latter case, passing reflow requirements at 260°C would require Δ exceeding 150 at room temperature. Not only would such a high value be very difficult to achieve, but it would also mean that devices would be extremely hard to write by STT. To solve this challenge and qualify pSTT-MRAM for reflow soldering, we have designed a MTJ stack with low Δ temperature coefficient [14]. Fig. 3a shows the error rate of 30 chips after simulated reflow procedure. In this test, the two halves of 10Mb test chips were initialized as logical 0 and 1, and both 1 to 0 and 0 to 1 errors were recorded after baking the chips for 90 seconds at 260°C. All 30 chips show error rates below the 10 ppm level, low enough for to be handled by Error Correction Code (ECC). The write error rate at room temperature of one of these 10 Mb chips is shown in Fig. 3b for 250 ns long write pulses. Without ECC, error-free writing is achieved with significant margin. This demonstrates that reflow qualification is achieved without compromising the chips performance.

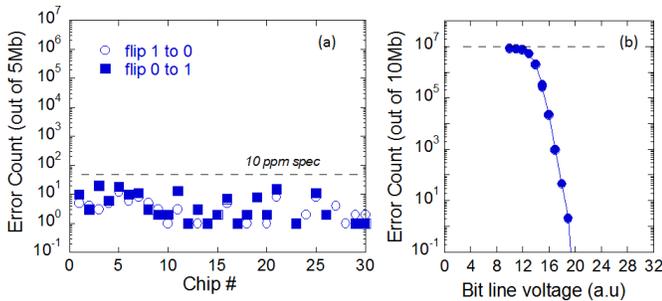


Fig. 3. (a) Bit error count measured on 30 different 10Mb chips fabricated in our backend facility, after simulated reflow soldering procedure (90 seconds bake at 260°C). The chips are initialized in two 5 Mb blocks of logic 0 and 1. Solid and open symbols show bit flips from 0 to 1 and 1 to 0, respectively. (b) Example of bit error count as a function of bit line voltage for one of the chips shown in (a). Data are measured without ECC. All bit can be written without any error. Write pulses are 250 ns long.

V. STT-MRAM FOR LLC APPLICATIONS

In this section, we show that pSTT-MRAM has also the potential to replace SRAM at advanced nodes for Last Level Cache (LLC) applications. Since the standard SRAM cell consists of 6 transistors, cell scaling has become increasingly challenging at advanced nodes. By contrast, the pSTT-MRAM cell is more compact, with only one transistor and one MTJ (1T-1MTJ). Besides, SRAM leakage issues could be alleviated by taking advantage of pSTT-MRAM non-volatility.

Even though L1 and L2 cache memories operate as rates beyond today's pSTT-MRAM capabilities, LLC cache is within reach if write speeds below 10 ns can be achieved. We have recently reported that this is indeed possible [13]. As shown in Fig. 4, full array switching is achieved on a 10 Mb test chip without ECC for write pulses as short as 3 ns. It should be emphasized that pSTT-MRAM is the only emerging non-volatile memory capable of such nanosecond write speed.

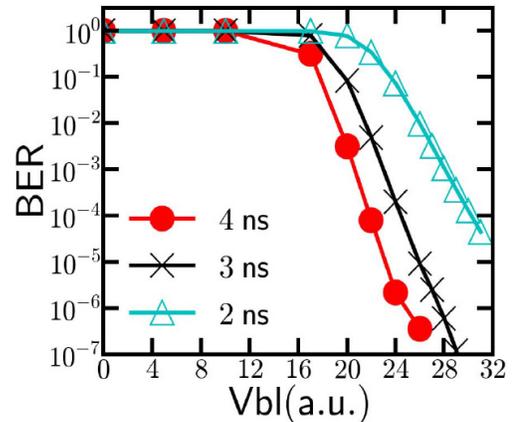


Fig. 4. Bit error rate as a function of bit line voltage V_{bl} for a 10 Mb chip, using write pulse lengths of 4, 3 and 2 ns. Data are taken without ECC. Error-free writing of then entire array is achieved for 3ns long pulses (reproduced from [13]).

CONCLUSION

Recent advances in embedded pSTT-MRAM have solved major technological hurdles and demonstrated the viability of the technology, opening the way to mass production. Major foundries have reported their progress at technical conferences in the past year [14, 23, 24], and have all announced production schedules beginning in 2018. Owing to its versatility, simple cell structure and compatibility with CMOS logic, pSTT-MRAM is well positioned to become a key component of next generation electronic devices.

REFERENCES

- [1] A. D Kent and D. C. Worledge, "A new spin on magnetic memories", *Nature Nanotech.* 10, p187, (2015).
- [2] J. Slonczewski, "Current-driven excitation of magnetic multilayers", *J. Magn. Magn. Mater.* 159, L1 (1996).
- [3] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a Current", *Phys. Rev. B* 54, 9353 (1996).
- [4] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study", *Phys. Rev. B* 62, 570 (2000).
- [5] S. H. Kang, "Embedded STT-MRAM for Energy-efficient and Cost-effective Mobile Systems", *Dig. Tech. Pap., Symp. VLSI Technology*, 2014, p36.
- [6] G. Jan et al., "High Spin Torque Efficiency of Magnetic Tunnel Junctions with MgO/CoFeB/MgO Free Layer", *Appl. Phys. Express* 5, 093008 (2012).
- [7] L. Thomas et al., "Perpendicular spin transfer torque magnetic random access memories with high spin torque efficiency and thermal stability for embedded applications", *J. Appl. Phys.* 115, 172615 (2014).
- [8] G. Jan et al., "Demonstration of fully functional 8Mb perpendicular STT-MRAM chips with sub-5ns writing for non-volatile embedded memories", *Dig. Tech. Pap., Symp. VLSI Technology*, 2014, p. 42.
- [9] L. Thomas, G. Jan, S. Le and P.-K. Wang, "Quantifying data retention of perpendicular spin-transfer-torque magnetic random access memory chips using an effective thermal stability factor method", *App. Phys. Lett.* 106, 162402 (2015).
- [10] G. Jan et al., " Demonstration of an MgO Based Anti-Fuse OTP Design Integrated With a Fully Functional STT-MRAM at the Mbit Level", *Dig. Tech. Pap., Symp. VLSI Technology*, 2015, p. 164.

- [11] L. Thomas *et al.*, "Solving the Paradox of the Inconsistent Size Dependence of Thermal Stability at Device and Chip-level in Perpendicular STT-MRAM", Tech. Dig. - Int. Electron Devices Meet. 2015, p 672.
- [12] Y. Lu *et al.*, "Fully Functional Perpendicular STT-MRAM Macro Embedded in 40 nm Logic for Energy-efficient IOT Applications", Tech. Dig. - Int. Electron Devices Meet. 2015, p 660.
- [13] G. Jan *et al.*, "Achieving sub-ns switching of STT-MRAM for future embedded LLC applications through improvement of nucleation and propagation switching mechanisms", Dig. Tech. Pap., Symp. VLSI Technology, 2016, p. 18
- [14] M. C. Shih *et al.*, "Reliability Study of perpendicular STT-MRAM as emerging embedded memory qualified for reflow soldering at 260°C", Dig. Tech. Pap., Symp. VLSI Technology, 2016, p. 114
- [15] S. Ikeda *et al.*, "Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeB/MgO/CoFeB/ pseudo-spin-valves annealed at high temperature", Appl. Phys. Lett. 93, 082508 (2008).
- [16] Worledge, D. C. *et al.*, "Switching distributions and write reliability of perpendicular spin torque MRAM", Tech. Dig. - Int. Electron Devices Meet. 2010, p296.
- [17] S. Ikeda *et al.*, "A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction", Nature Mater. 9, 721 (2010).
- [18] D. C. Worledge *et al.*, "Spin torque switching of perpendicular TaCoFeBMgO-based magnetic tunnel junctions", Appl. Phys. Lett. 98, 022501 (2011).
- [19] J. -H. Park *et al.*, "Enhancement of data retention and write current scaling for sub-20nm STT-MRAM by utilizing dual interfaces for perpendicular magnetic anisotropy", Dig. Tech. Pap., Symp. VLSI Technology, 2012, p. 57.
- [20] H. Sato *et al.*, "Perpendicular-anisotropy CoFeB-MgO magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure," Appl. Phys. Lett., vol. 101, p. 022414, 2012.
- [21] K. Tsunoda *et al.*, "A Novel MTJ for STT-MRAM with a Dummy Free Layer and Dual Tunnel Junctions", Tech. Dig. - Int. Electron Devices Meet. 2012, p. 665.
- [22] S. Ikeda *et al.*, "Perpendicular-anisotropy CoFeB-MgO based magnetic tunnel junctions scaling down to 1X nm", Tech. Dig. - Int. Electron Devices Meet. 2014, p 796.
- [23] D. Shum, *et al.*, "CMOS-embedded STT-MRAM Arrays in 2x nm Nodes for GP-MCU applications", Dig. Tech. Pap., Symp. VLSI Technology, 2017, p. 18.
- [24] Y. J. Song *et al.*, "Highly Functional and Reliable 8Mb STT-MRAM Embedded in 28nm Logic", Tech. Dig. - Int. Electron Devices Meet. 2016, p 663.