Preprint: J. Magnetic Society of Japan, v.5, p. 171 (2010)

Toggle and Spin-Torque MRAM: Status and Outlook

J. M. Slaughter, N. D. Rizzo, F. B. Mancoff, R. Whig, K. Smith, S. Aggarwal and S. Tehrani
Everspin Technologies, Inc., *1300 N. Alma School Road, Chandler, Arizona 85224, USA*

We review the development of MRAM technology at Everspin, focusing on both toggle MRAM, which is used in our current commercial products, and spin-torque MRAM, which offers significant potential for cell size reduction that could enable higher memory densities. Toggle MRAM uses magnetic fields for programming the bits with a particular free layer structure, bit orientation, and write pulse sequence to avoid the half-select disturbs present in other field-writing techniques. We find excellent read, write, and reliability performance with toggle MRAM and the potential for continued scaling beyond our current 16Mb memory size. Next, we describe spin-torque MRAM, its development challenges and expected performance attributes, and the results of our sub-100 nm spin-torque bits, with both CoFeB and NiFe free layers, integrated into 16kb CMOS arrays. We measure an intrinsic array separation between the spin-torque switching and breakdown voltage distributions $>12\sigma$ for bits with a magnetic stability against thermal disturbs $E_b / k_bT \approx 52$ sufficient for a 10 year data retention lifetime.

Key words: toggle MRAM, spin-torque MRAM, spin-transfer, magnetic tunnel junction, nonvolatile memory

1. Introduction*

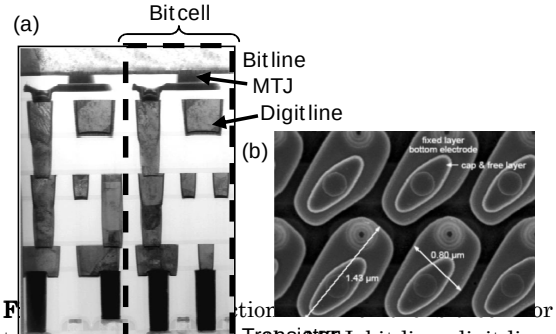
Magnetoresistive random access memory (MRAM) employs ferromagnetic storage devices integrated with semiconductor circuitry to provide a nonvolatile random access memory with fast read and write as well as virtually unlimited read and write cycles. Each MRAM bit contains a magnetic tunnel junction (MTJ) consisting of two ferromagnetic layers separated by a thin (~ 1 nm) insulating layer. Data is stored in terms of the magnetization direction of one of the ferromagnets (the free layer) either parallel or antiparallel to the other (the fixed layer). A tunnel magnetoresistance (*MR*) that is low (high) for parallel (antiparallel) magnetizations provides the read-out signal. The first MRAM product used toggle-mode writing as a form of magnetic-field induced switching and became commercially available in 2006 (the 4Mb Everspin Technologies MR2A16A, then produced by Freescale Semiconductor). This part was followed by a family of related parts, the most recent being a 16Mb circuit. Spin-torque MRAM is currently under development and uses electric current-induced switching as an alternative writing scheme with potentially higher density, lower power operation.

2. Toggle MRAM: Operation and attributes

A toggle-mode MRAM is shown by cross-section transmission electron microscopy (TEM) in Figure 1(a). The memory uses one transistor and one MTJ (1T/1MTJ) for each bit cell. The transistor provides the current flow through the MTJ needed for reading the *MR*. Each MTJ is located at the intersection of two orthogonal conductive lines, the bit line above the MTJ and the digit line below it, which are energized to switch the bit. The selected bit experiences magnetic fields from both lines, whereas other bits along the energized lines, called half-selected bits, experience only one of the two fields. One of the fundamental problems of field-switched MRAM is designing the bit and the field pulses so that the selected bit always switches, and the half-selected bits never switch.

Toggle MRAM avoids the half-select disturb problem through the use of a switching mode, dubbed "Savtchenko switching" after its late inventor[1], which employs a free layer synthetic antiferromagnet (SAF) and a bit orientation rotated 45° with respect to the bit and digit lines (as shown by top-down SEM in Figure 1(b)) combined with a specific current pulse sequence. The selectivity using this mode is greatly enhanced because a single current line pulse alone does not lower the energy barrier to reversal, and even large single-line fields cannot switch the bit. This unique behavior results in a wide operating region with a threshold onset for switching[2].

The programming pulse sequence and resulting



Ftion or to ng the MTJ, bit line, digit line, and underlying CMOS circuitry. **(b)** Top-down SEM of toggle MRAM MTJ bits and their underlying bottom electrode.

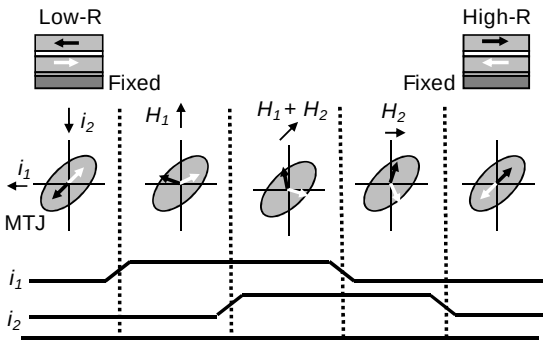


Fig. 2: Schematic of a toggle MRAM bit with the field sequence used to switch the free layer from one state to the other. The fields H_1, $H_1 + H_2$, and H_2 are produced by passing currents, i_1 and i_2, through the write lines. (From B. N. Engel *et al.*, *IEEE Trans. Magn.* **41**, 132 (2005). © 2005 IEEE).

magnetic behavior are depicted in Figure 2. The black and white arrows represent the magnetic moment of the two sublayers in the SAF free layer, where the white represents the layer that is in contact with the tunnel barrier. To toggle the bit from an initial "0" to a final "1", the currents i_1 and i_2 are pulsed with the phase relationship shown in the figure such that the vector sum of the two magnetic fields begins 45° from the easy axis and then effectively rotates through 90°. The SAF responds by orienting nominally orthogonal to this field and rotating with it such that when i_2 is turned off, the moments relax to their easy-axis with each oriented 180° from its initial state. Because the *MR* only depends on the direction of the layer that is in contact with the tunnel barrier (white), the resistance has switched from the low to the high state, i.e., from "0" to "1".

Since the two magnetic sublayers are identical, if the process is repeated with the same polarity pulses, the free layer will reverse again in the same manner. Because the same pulse sequence makes the bit alternate between the "0" and "1" state, it is said to toggle between the two states. This toggle-write mode requires that the memory operates with a decision write scheme, where the bit state is read first and only toggled if the new data differ from the existing. This approach has benefits in limiting the overall power consumption, and the unipolar current allows the use of smaller transistors, thereby improving array efficiency.

If only a single-line current is applied (half-selected bits), the 45° field angle cannot switch the state. In fact, the single-line field raises the switching energy barrier of those bits so that they are stabilized against reversal during the field pulse. This is in marked contrast to the conventional field-switching approach, where all of the half-selected bits have their switching energy reduced and are, therefore, more susceptible to disturbance. The phase relationship of the pulses is, therefore, required for switching resulting in significantly improved bit selectivity in an array.

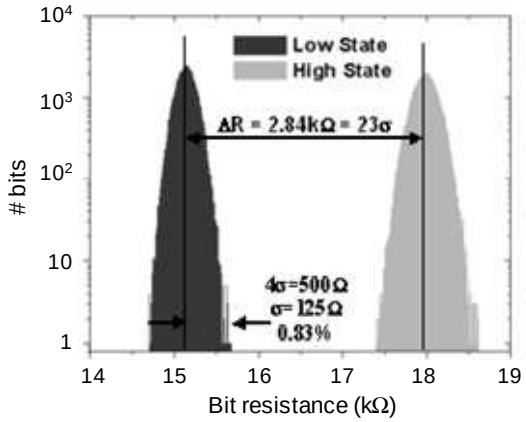


Fig. 3: Low- and high-state resistance distributions measured at operating bias and with series transistor resistance. Data are from a 64kb sample of a 4Mb device. (From B. N. Engel *et al.*, *IEEE Trans. Magn.* **41**, 132 (2005). © 2005 IEEE).

3. Toggle MRAM: Read, write, and reliability

Proper uniformity control of both the MTJ *MR* and the resistance-area product (*RA*) is important for functional read-out from a memory array. The exponential thickness dependence of *RA* on the very thin barrier creates a challenge for producing MTJ material that is repeatable and uniform over the 200 mm diameter wafers used in production. However, 1σ resistance uniformity of 5% and *MR* uniformity of 1% over 200 mm wafers have been achieved[3]. These submicron patterned bits have average values of *MR* = 45% and *RA* = 10.2 kΩ–μm^2. The metal layers were formed by sputter deposition, and an oxygen plasma was used to oxidize a thin Al layer to form the AlOx tunnel barrier. This uniformity is achieved through a deposition process with a 1σ Al thickness uniformity of 0.5%. The AlOx tunnel barrier can be engineered by optimizing the Al thickness and oxidation time to have *MR* = 40% - 50% for the range of *RA* > 200 Ω–μm^2 [4].

A number of factors can contribute to the resistance distribution. Bit area variations, e.g., owing to lithography or etch variations, will directly cause variations in bit resistance. Process damage or veils created during the etch process also may contribute. However, we have found that with an optimized patterning process, the quality of the MTJ material itself plays a major role in determining the resistance distributions. Figure 3 shows measured array distributions for a 4Mb circuit, with MTJ material optimized for MRAM, having 23σ separation between the mean high and low resistances[5], significantly greater than the targeted 12σ for functionality. The main factors that led to this result were the high-quality patterning process to minimize area variation and optimized MTJ materials and processes to make the tunnel barrier more uniform.

For the write performance of a 4Mb toggle MRAM, the large operating region is seen in Figure 4. Within

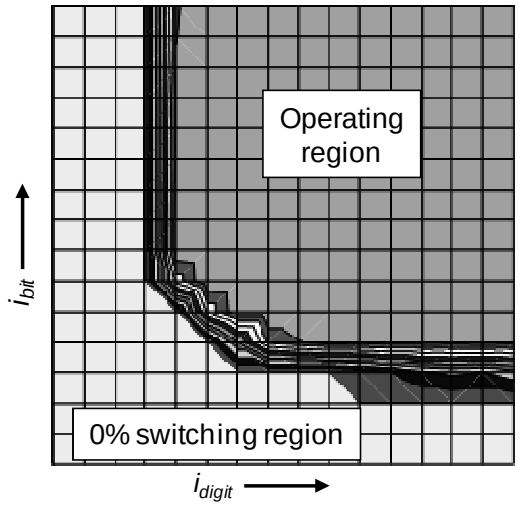


Fig. 4: Switching map for a 4Mb toggle MRAM array. (From B. N. Engel *et al.*, *IEEE Trans. Magn.* **41**, 132 (2005). © 2005 IEEE).

the operating region, 100% of the bits are successfully written by the combined bit and digit line fields. Outside the operating region at lower fields, 0% of the bits are written, thus avoiding half-select disturbs. A narrow but finite transition width between the two regions is apparent due to the variation of the switching field for different bits in the array. Nevertheless, the operating region is greatly increased compared to the much narrower operating point of conventional field-switched MRAM that results from the distribution of switching fields and the related half-select disturbs.

For product reliability, the two MTJ barrier failure modes that must be controlled are time-dependent dielectric breakdown (TDDB) and resistance drift. Dielectric breakdown is detected as an abrupt increase of junction current owing to a short forming through the tunneling barrier. Resistance drift is a gradual reduction of the junction resistance over time that can eventually lead to reduced read margin and increased error rate. The amount of drift that can be tolerated is determined by a combination of the *MR*, the original resistance distribution and the capability of the read-out circuit to handle small signals. Both failure modes are strongly accelerated by voltage bias and temperature (Figure 5). Their effect on lifetime at operating conditions can be deduced using stress testing methods[6]. In addition, the resistance drift was found to have a strong favorable frequency dependence[7], greatly increasing the lifetime at the operating frequency of the part. Rigorous reliability analysis done for the Everspin 4Mb MRAM product has shown that both lifetimes far exceed the reliability requirements of industrial and automotive products[8)-9)].

Scalability is also crucial for continued MRAM competitiveness. Write scalability is influenced by the increased switching field caused by a decreasing bit area and the increasing free layer thickness needed to

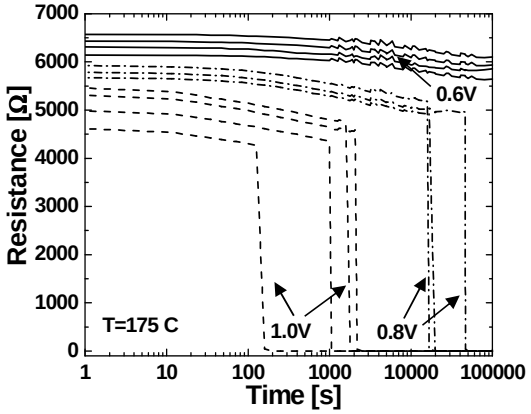


Fig. 5: Time-dependent dielectric breakdown (TDDB) and resistance drift failure modes accelerated by voltage bias (0.6 V solid line, 0.8 V dash-dot line, 1 V dashed line) and temperature (175°C).

maintain 10 year stability against thermal disturbs. However, as the feature size scales, the bit line and digit line widths decrease as well as the interlayer dielectric spacing between them and the bit. Thus, the field experienced by the bit can increase along with the switching field, while sufficient thermal stability is generated for operation down to the 45 nm node. In addition, the mean switching field at these nodes may be decreased by alternative toggle configurations[10)-11)]. For scaling the read performance, the higher *MR* available with MgO compared to AlOx can enable faster read and help overcome the increasing resistance variability inherent in scaling interconnects and devices to smaller sizes, as demonstrated in an MRAM array using MgO integrated with 90 nm CMOS in an 8kb array[12)].

4. Spin-torque MRAM: Operation and expected attributes

The spin-torque effect[13)-14)] is a result of conservation of angular momentum in layered magnetic devices (for an excellent review[15)-17)]). Electrons flowing through the device acquire a spin polarization from a first layer, referred to as the fixed layer, which then exerts a spin-torque on a second layer, the free layer. At large current densities (~10^6 - 10^7 A/cm^2), the spin-torque can cause the free layer to reverse its magnetization, forming the basis for writing bits in spin-torque MRAM (ST-MRAM). For ST-MRAM, the simplest free layer consists of a single ferromagnetic layer, as opposed to the SAF used for toggle MRAM. Spin torque switching has been observed in MTJ's[18)-19)] and is under examination at a number of industrial companies for the purpose of programming MRAM bits[20)-29)]. The critical current density for spin torque switching in zero field can be shown to be approximately[13)]:

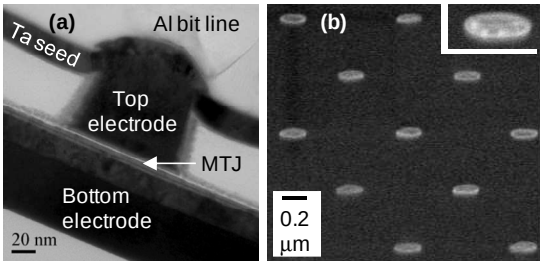


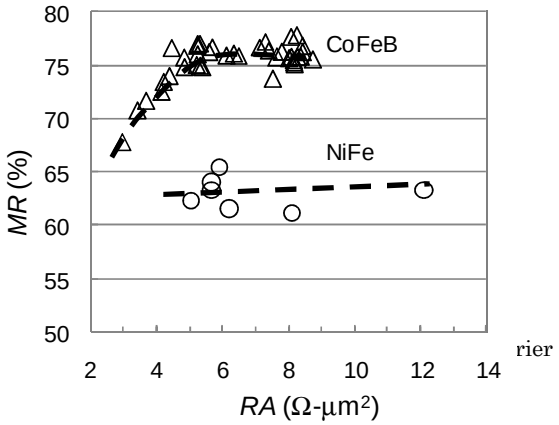
Fig. 6: Cross-section TEM **(a)** and top down SEM **(b)** images of MTJ's for ST-MRAM in a short flow process.

where $\hbar$ is Planck's constant, e is the electron charge, α is the Gilbert damping constant of the free layer, M_s is the saturation magnetization, t is the free layer thickness, η is the spin torque efficiency, and H_k is the intrinsic anisotropy of the free layer. For typical spin-torque devices, $2\pi M_s >> H_k$.

Due to the simple cell structure, ST-MRAM has the potential to reduce the MRAM cell size and thus extend MRAM technology to higher memory densities. For ST-MRAM, a single conductive line makes electrical contact to the top of the bit, and a single pass transistor beneath the bit is used for programming and reading. The elimination of the digit line, used to write toggle MRAM, will result in a smaller cell size if the pass transistor can be made small enough. A typical logic transistor can pass approximately 600 μA per μm of transistor width, almost independent of technology node. In general, the goal is to reduce the switching current I_{sw} so that a minimum-size transistor can be used at each technology node, thus achieving the highest possible memory density. For example, it would be desirable to achieve J_{sw} ~ 1 MA/cm^2 to enable I_{sw} ~ 100 μA for a 90 nm wide bit. The decreasing spin torque write current required for decreasing free layer volume is beneficial for scaling the bit for future CMOS generations.

Further areas must be addressed to bring ST-MRAM from development to production. The spin torque switching current density J_{sw} needs to be significantly reduced from the current state-of-the-art for two reasons. First, the required switching current is directly proportional to the size of the CMOS pass transistor below the bit, so reduced J_{sw} is needed to reach the desired high memory density for commercial production. Second, low J_{sw} is needed to prevent tunnel barrier damage or breakdown due to the write voltage bias across the bit and to ensure memory reliability.

From Equation (1), one path to lower J_{sw} is to reduce the free layer thickness t or magnetization M_s. This path is limited for several reasons. First, t cannot be thinner than ≈ 2 nm or $J_{sw} = \left(\frac{2e\alpha M_s t}{\hbar}\right) \cdot \frac{(H_k + 2\pi M_s)}{\eta}$ (1) the magnetic properties of the film degrade. Second, reducing M_s or t increases the likelihood of data loss due to thermal fluctuations. The energy barrier to magnetization reversal caused by thermal fluctuations is $E_b \propto N_d M_s^2 A \cdot t$, where A is the bit area and N_d is the effective demagnetizing factor of the free layer. For a memory with a typical 10 year nonvolatility requirement,



$E_b \geq 50 k_b T$, where k_b is Boltzmann's constant and T is the temperature.

5. Spin-torque MRAM: Device and array properties

We have fabricated MTJ's designed for ST-MRAM using a naturally oxidized MgO tunnel barrier, for which a metallic Mg precursor layer is exposed to an O_2 ambient, and a single free layer of either CoFeB or NiFe (with composition close to $Ni_{80}Fe_{20}$). The MTJ bits are patterned across 200 mm diameter Si wafers using standard optical lithography followed by reactive ion etching. Figure 6 shows examples of these MTJ's as integrated in a short-flow test vehicle, in which only the MTJ layer itself and the top electrical contact line are patterned in order to provide a rapid turnaround time for fabrication. Figure 6(a) shows a cross-section TEM image. The MTJ, with its brightly colored MgO tunnel barrier, is sandwiched between top and bottom electrodes. Figure 6(b) shows a top-down SEM view of the MTJ's following the reactive ion etch. The bit size is roughly 0.09 μ m × 0.21 μm. We also fabricated similar MTJ's into 16kb sized arrays integrated with a 1T/1MTJ CMOS test vehicle.

Figure 7 shows the measured MR vs. RA for MTJ's in the short-flow process. Each data point in Figure 7 represents the median values over a single wafer, and the variation in RA from 3 – 12 Ω–μm^2 was produced by varying the MgO barrier oxidation time for different wafers. The dashed lines are guides to the eye. The circles show data for a NiFe free layer, and the triangles show data for a CoFeB free layer with a magnetization close to the NiFe. The MR increases only slowly with RA in this resistance range and reaches a value as high as 78% for CoFeB and 65% for the NiFe. Even higher MR up to 135% was also observed in the same RA range for a somewhat higher magnetization CoFeB alloy.

According to Equation (1), the CoFeB might at first be expected to show lower J_{sw} than NiFe due to its

higher MR and resulting large spin-torque efficiency η.

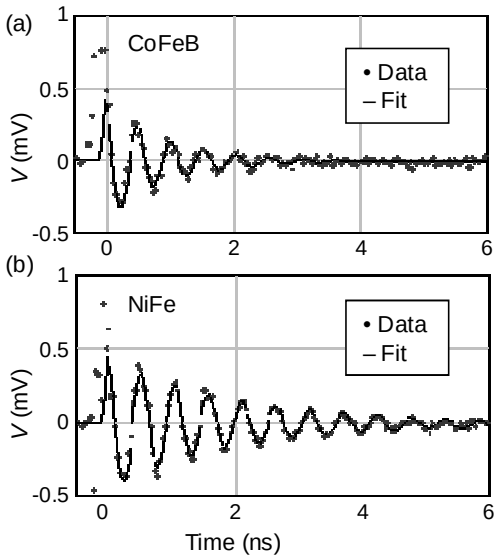


Fig. 8: Voltage output signal vs. time from pulsed inductive microwave magnetometer. The solid line fit determines the magnetic damping parameter for NiFe and CoFeB free layers.

However, another important consideration for J_{sw} in Equation (1) is given by the magnetic damping parameter α. We measured α in unpatterned free layer films of NiFe and the CoFeB with comparable magnetization by using a pulsed inductive microwave magnetometer[30], as shown in Figures 8(a) and 8(b) respectively. In this measurement, an unpatterned film of the free layer is exposed to a fast magnetic field pulse generated by a coplanar waveguide, and the subsequent ringing of the free layer magnetic moment in response to the field is measured in the time domain according to the voltage that the moment induces in the waveguide. The dots in Figures 8(a) and 8(b) show the measured voltage signal, and the solid black lines are fits to the data using a Landau-Lifshitz-Gilbert form for the magnetization ringing in response to a fast field pulse. We can intuitively see that the damping is lower for NiFe in Figure 8(b) since the ringing of its magnetization persists for a longer time than for the CoFeB. From the fits, we determine values of $\alpha \approx 0.007$ for NiFe and $\alpha \approx 0.013$ for CoFeB.

Based on this data from the unpatterned films and the short flow MTJ's, we integrated both the CoFeB and NiFe free layers into CMOS arrays. Figure 9 shows measurements of the voltage required for spin-torque switching V_{sw} from antiparallel to parallel (AP → P) free and fixed layer magnetizations as a function of the voltage pulse time t_p ranging from 20 ns up to 100 ms. Data for CoFeB (NiFe) is given by the triangles (circles). Each data point in Figure 9 represents the median V_{sw} at a particular t_p measured for all the CMOS arrays across a single representative wafer. For both CoFeB and NiFe, V_{sw} increases monotonically as a function of

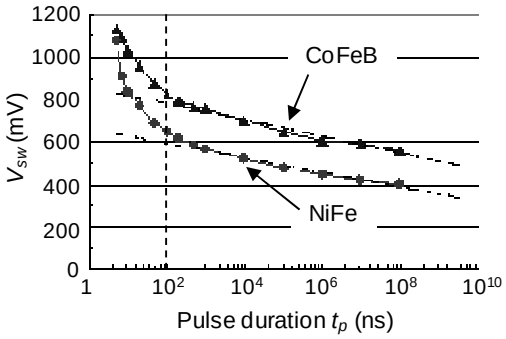


Fig. 9: Switching voltage V_{sw} vs. pulse time t_p for NiFe and CoFeB free layer MTJ's in the ST-MRAM CMOS circuit.

decreasing t_p, with two different regimes that cross over at a pulse time of roughly 100 ns. For long t_p, V_{sw} increases relatively slowly with decreasing t_p and can be well fit by the theoretical form from a thermal activation model[31)-32)]:

$$V_{sw} = V_{sw}^{0}\left[1 - \frac{k_b T}{E_b}\ln\left(\frac{t_p}{t_0}\right)\right] \quad (2)$$

V^0_{sw} represents the intrinsic switching voltage extrapolated to a short time $t_0 = 1$ ns. The black dashed lines in Figure 9 give the fits to this form and match the data well for $t_p > 100$ ns. From the fits, we determine $E_b / k_b T \approx 52$ for both CoFeB and NiFe, which is sufficient to meet a 10 year thermal stability requirement of $E_b / k_b T > 50$. Also, we see that the required switching voltage V^0_{sw} is lower for the NiFe free layers than for the CoFeB, as determined by Equation (1) using the MR and magnetic damping α measured in Figs. 7-8, along with the other free layer parameters. Finally, for $t_p < 100$ ns, V_{sw} increased more rapidly as a function of decreasing t_p, as the spin-torque switching transitioned from a thermally activated mode to a magnetodynamical regime which requires increased V_{sw} to drive the switching more rapidly. V_{sw} increases more quickly with t_p for NiFe in this range and begins to converge with CoFeB.

For adequate electrical reliability of the ST-MRAM tunnel barrier, a large separation of the switching voltage V_{sw} distribution from the breakdown voltage V_{bd} distribution is critical. The V_{sw} and V_{bd} distributions in the AP → P bias direction for a CMOS array using a CoFeB free layer are shown in Figure 10 at a pulse time t_p = 100 ns. The plot shows the probability for either switching (diamonds) or breaking down (squares) the many bits across the array as a function of the applied voltage. The median switching voltage in the array (given by 50% probability) is well separated from the median breakdown voltage by roughly 700 mV in this case. Fitting the data in Figure 10 to error functions,

we find relative standard deviations for the spin-torque switching and breakdown distributions of $\sigma_{sw}^{rel} \approx 4\%$ and $\sigma_{bd}^{rel} \approx 4\%$. Defining the separation of switching and breakdown as $S = (V_{bd} - V_{sw})/(\sigma_{sw} + \sigma_{bd})/2$, we find $S > 12\sigma$ for this data (where σ is the average of the absolute standard deviations σ_{sw} and σ_{bd}), as needed for a fully

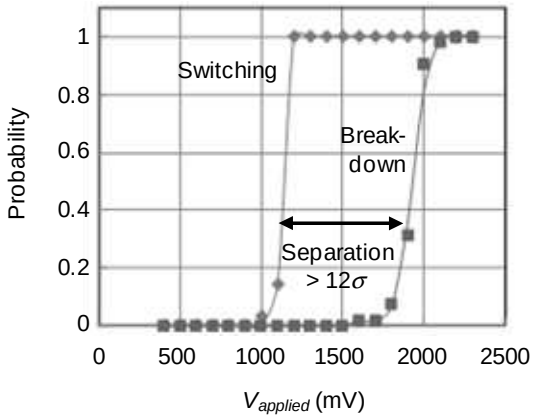


Fig. 10: Spin-torque switching and breakdown distributions for CoFeB free layers in the ST-MRAM CMOS circuit.

functioning memory. This data provides an important proof of concept for ST-MRAM reliability.

6. Conclusions

Toggle mode writing has enabled commercial production at Everspin of a family of MRAM products, as large as 16Mb so far, with excellent read, write, and reliability performance. Continued toggle MRAM development will advance the technology to greater memory sizes and a further expanded market. At the same time, MRAM with spin-torque switching also is under development and offers potential for higher density MRAM circuits. We integrated arrays of MTJ bits designed for spin-torque MRAM into a 16kb CMOS test vehicle using both CoFeB and NiFe free layers, demonstrating good separation (>12σ) between the switching and breakdown distributions.

[*]Portions quoted from J. M. Slaughter: *Annual Review of Materials Research*, **39,** 277 (2009).

References

1) L. Savtchenko, B. N. Engel, N. D. Rizzo, M. F. DeHerrera, and J. A. Janesky: U.S. Patent No. 6545906 B1 (2003).
2) B. N. Engel *et al.*: *IEEE Trans. Magn.*, **41**, 132 (2005).
3) J. Åkerman *et al.*: Magnetoelectronics, ed. M. Johnson, p. 231, (Elsevier, Amsterdam, 2004).
4) S. S. P. Parkin *et al.*: *Appl. Phys. Lett.*, **75**, 543 (1999).
5) M. Durlam *et al.*: *IEDM Tech. Digest,* **34**, 1 (2003).
6) J. Åkerman *et al.*: *IEEE Int. Reliability Phys. Symp.*, **163** (2005).
7) S. K. Pon and Y. Chung: *IEEE Int. Reliability Phys. Symp.*, **437** (2006).
8) S. Tehrani: *IEDM Tech. Digest*, 1, (2006).
9) J. Åkerman *et al.*: *IEEE Trans. Magn.*, **42**, 2661 (2006).
10) D. W. Abraham and D. C. Worledge*: Appl. Phys. Lett.*, **88**, 262505 (2006).
11) Y. Fukumoto, T. Suzuki, and S. Tahara: *Appl. Phys. Lett.* **89**, 061909 (2006).
12) J. M. Slaughter *et al.*: *IEDM Tech. Digest*, 873 (2005).
13) J. Slonczewski: *J. Magn. Magn. Mater.*, **159**, L1 (1996).
14) L. Berger: *Phys. Rev. B*, **54**, 9353 (1996).
15) J. A. Katine and E. E. Fullerton*: J. Magn. Magn. Mater.*, **320**, 1217 (2008).
16) J. Z. Sun and D. C. Ralph*: J. Magn. Magn. Mater.*, **320**, 1227 (2008).
17) C. Chappert, A. Fert, and F. Nguyen Van Dau: *Nature Materials*, **6**, 813 (2007).
18) Y. Huai *et al.*: *Appl. Phys. Lett.*, **84**, 3118 (2004).
19) G. D. Fuchs *et al.*: *Appl. Phys. Lett.*, **85**, 1205 (2004).
20) M. Hosomi *et al.*: *IEDM Tech. Digest*, 459 (2005).
21) T. Kawahara *et al.*: *IEEE J. Solid-State Circuits*, **43**, 109 (2008).
22) T. Kishi *et al.*: *IEDM Tech. Digest*, 1 (2008).
23) B. Beach *et al.*: *IEDM Tech. Digest* , 305 (2008).
24) Z. Diao *et al.*: *J. Phys.: Condens. Matter* , **19**, 1 (2007).
25) D. V. Dimitrov *et al.*: *Appl. Phys. Lett.*, **94**, 123110 (2009).
26) J. M. Slaughter *et al.*: *IEEE Intermag Tech. Digest*, CF-02 (2009).
27) C. J. Lin *et al.*: *IEDM Tech. Digest*, 279 (2009).
28) S.-C. Oh *et al.*: *Nature Physics,* **5**, 898 (2009).
29) S. Kuami *et al.*: *Symp. VLSI Tech.*, 230 (2009).
30) A. B. Kos, T. J. Silva, and P. Kabos: *Rev. Sci. Instrum.*, **73**, 3563 (2002).
31) R. H. Koch, J. A. Katine, and J. Z. Sun: *Phys. Rev. Lett.*, **92**, 088302 (2004).
32) Z. Li and S. Zhang*: Phys. Rev. B,* **69**, 134416 (2004).

Received Jan. 08, 2010