CROCUS Technology
Blossoming future

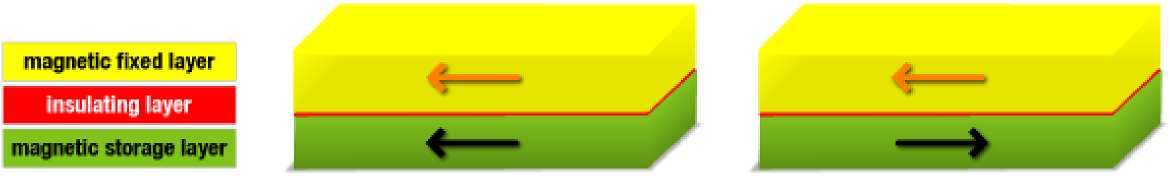
# The Emergence of Practical MRAM

Barry Hoberman
Crocus Technologies

The Nobel Prize for physics in 2007 was awarded to the discoverers of something called Giant Magneto-Resistance, or GMR for short. GMR, together with its cousin TMR (Tunnel Magneto-Resistance), is an effect that occurs in ultra-thin (i.e. a few nanometers) multi-layers of magnetic materials separated by a metallic (for GMR) or insulating (for TMR) film. This new effect utilizes the familiar electrical charge on electrons as well as the not-so-familiar 'spin', and combines them to create a new class of electrical devices, frequently categorized as 'spintronics'. The economic impact of GMR since its fundamental discovery in Europe in the late 1980's has been huge, as it is now widely used in hard disk drive read/write heads and automotive position sensors. TMR today stands poised to potentially reshape the semiconductor memory market as well.

Semiconductor engineers have been hunting for a way to deploy TMR in high-density memory, leading to the emergence of MRAM, or magnetic RAM. Many established chip companies have been involved in this research and development, but to date there has been only limited success. While disk drive heads use only a single TMR element, called an MTJ (for magnetic tunnel junction), MRAM memories require the use of millions of MTJ's in a regular array structure (i.e. typically one MTJ for each bit of stored data). This is one of the challenges that makes MRAM harder to implement than disk drive heads.

Figure 1



MTJ Storing a "1"

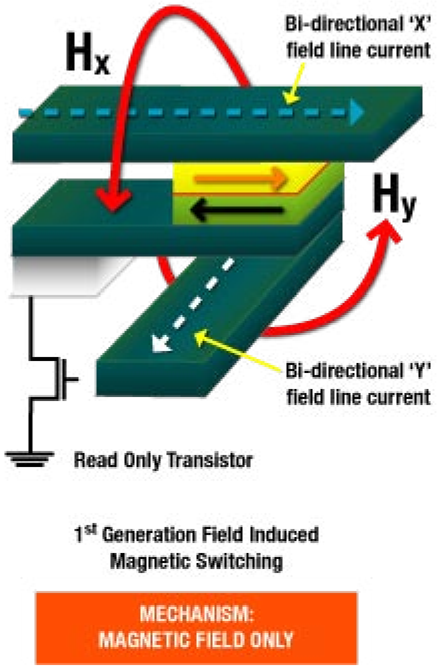
MTJ Storing a "0"

Stored data follows the magnetization direction (parallel or anti-parallel) of magnetic layers in the MTJ.

MRAM offers some very special benefits, and has motivated more than a decade's effort in the laboratory. It holds the promise of non-volatility, infinite endurance, high-speed reading and writing, random access to data, and low cost - a combination of characteristics not offered by any of today's popular memory types: DRAM, Flash, and SRAM. With MRAM, whole new approaches to system level memory semantics become attractive. Also, SOC's with easily integrated embedded non-volatile memory and lower cost SRAM (much denser than 6-transistor SRAM) fall into grasp. Many semiconductor companies have been publicly involved in MRAM research and development, including Cypress, Hitachi, Hynix, IBM, Motorola (subsequently Freescale, now Everspin), NEC, (now Renesas), Samsung, and Toshiba.

The first generation of MRAM technology has seen difficulties coming to market. Technical challenges hindering manufacturability center around three factors: stability, selectivity, and scalability. First generation MRAM technologies have primarily depended on writing the memory bit with a magnetic field produced in metal lines on the chip. Specifically, driving strong electrical currents down both 'x' and 'y' metal lines produces a 'threshold' magnetic field at the cross-point of the 'x' and 'y' lines that can write data into the bit cell. All the other neighboring bit cells exposed directly to the 'x' and 'y' line see a little over half the nominal 'threshold' field, and as a result are at risk of unwanted overwriting subject to the statistics of the parametric control of the manufacturing process. Thus said, the 'x'/'y' writing technique has inadequate selectivity caused by this well known 'half select' phenomenon. The stability and scalability challenges are intertwined. Data stored in an MRAM bit (like any memory technology) is subject to both statistical data loss and other forms of parasitic field disturbances. In MRAM's case, thermal agitation sets up a small but finite probability of random data 'flipping'. The standard solution to the data loss problem is to raise the switching threshold (i.e. coercive field) of the magnetic material in the MTJ. However, raising the coercive field demands a larger applied magnetic field for writing, leading to larger 'write' currents in the 'x' and 'y' field lines. Process feature scaling drives this problem into a roadblock, because as feature sizes shrink (think 180nm to 130nm to 90nm to 65nm to 45nm), the thermal stability of the MTJ degrades as the spatial volume of the MTJ shrinks. This worsening thermal instability requires ever increasing switching thresholds and field line write currents, in opposition to general scaling principles. At some point, the field line currents, stability issues, and feature size just can't be overcome. Experts in the MRAM community say that this barrier arises at about 90nm or so.

Figure 2

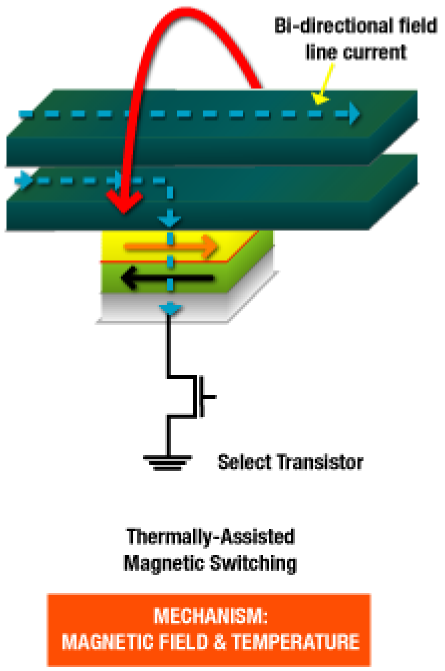

$H_x$
Bi-directional 'X' field line current
$H_y$
Bi-directional 'Y' field line current
Read Only Transistor
1st Generation Field Induced Magnetic Switching
MECHANISM: MAGNETIC FIELD ONLY

Only one first-generation MRAM technology has ever made it to market. This was a recent optimization dubbed 'Toggle', which was pioneered by Everspin, and evades the selectivity issue with an optimization that trades increased power for enhanced selectivity. Available MRAM's in the market are limited in size to 4Mbit, in 180nm technology, with costs per bit of the order of 1E-4 cents/bit. This is in comparison to DRAM, with costs on the order of 1E-6 cents/bit.

Today, several technical approaches to second generation MRAM are in the works to address all these limitations and allow scaling to 65nm and below. Second generation MRAM development is being done at both established semiconductor companies and in venture capital sponsored start-ups. The leading candidates for workable second generation MRAM technologies are called Thermal Assisted Switching (TAS) and Spin Torque. TAS is being developed by Crocus Technology, and represents a strong step into second generation MRAM physics while getting strong leverage from first generation processes and materials, making it a prime candidate for quick time to market. Spin Torque is being developed by multiple teams worldwide, but still has challenges in basic physics and materials to overcome in order to achieve market readiness.

TAS is a mature idea already developed for disk drives and poised for volume production. The patented adaptation of TAS to the MRAM application is the heart of Crocus Technology's innovation in MRAM technology. With an innovative modification of first generation MRAM technology, TAS enables a breakthrough MRAM structure that completely solves the first-generation selectivity and stability problems, enabling a cost-effective and scalable memory technology to at least the 32nm node.

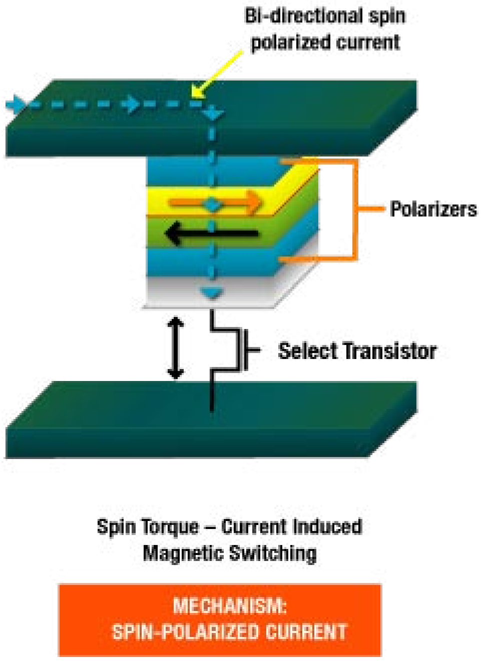
Figure 3



TAS operates by introducing the concept of 'Blocking Temperature' (or Tb) in a specially designed multilayered magnetic stack from which the MTJ is built. While idling and reading, the MTJ of the bit cell sits at a temperature well below Tb. During writing, the MTJ to be written has its temperature raised above Tb. The 'magic' of TAS is that when the MTJ's temperature is below Tb, the data in the bit cell is orders of magnitude more stable than when the MTJ temperature is above Tb. Since the MTJ is heated by current in a unique bit cell transistor, the selectivity problem of the first generation is gone. And since the relative stability of the MTJ when hot (i.e. above Tb) versus cold (i.e. below Tb) can be effectively engineered by the materials selection applied in the MTJ, the stability versus scaling obstacle of the first generation MRAM technology vanishes. Because the Blocking Temperature can also be engineered to be close to the normal operating temperature range of the chip (in contrast, for example, to the very high temperatures required in phase-change memory), the job of heating and cooling the MTJ in the bit cell can be accomplished with small currents in just a few nanoseconds. TAS can be used with field line writing, as in the first generation MRAM, or with Spin Torque in the longer run.

Spin Torque MRAM operation introduces a very significant paradigm change in the writing process. Spin Torque depends on a recently discovered effect in which the magnetization of nano-elements is flipped back and forth by an electrical current (no applied magnetic field, no ‘x’ or ‘y’ metal field line) provided that such current is heavily ‘spin polarized’. This ‘spin polarization’ is achieved by passing the current through a thin magnetic layer (polarizer) that is added to the MTJ, out of which only one type of spin flows. The polarized current interacts with the other layers of the MTJ in such a way as to affect the stored magnetization, therein allowing writing in one direction or the other depending on the current polarity. When idling (i.e. holding data), the bit cell MTJ has no electrical current moving through it. During reading, a low current moves through it. When writing, a much higher (spin polarized) current moves through the MTJ.

Figure 4


Bi-directional spin polarized current
Polarizers
Select Transistor
Spin Torque – Current Induced Magnetic Switching
MECHANISM:
SPIN-POLARIZED CURRENT

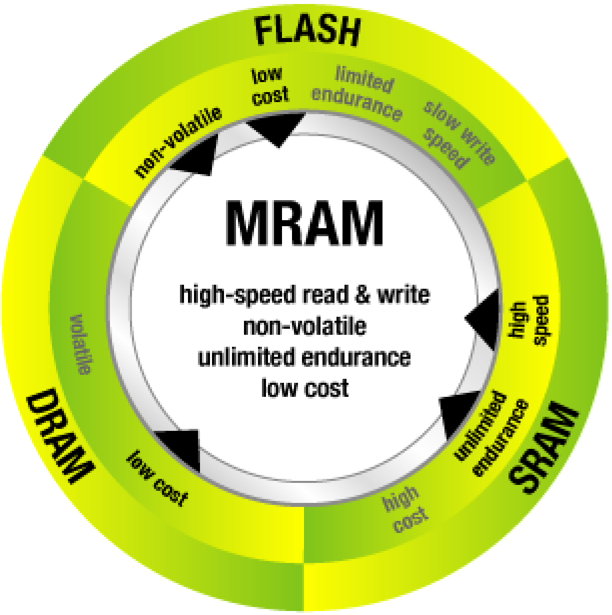
The challenge facing the near-term deployment of Spin Torque is that the best-case, state-of-the-art write current densities are still almost an order of magnitude too high for building practical memories. These current densities are defined by the basic material properties and physics of the MTJ, and will be reduced only with significant development and innovation. It is important to note that when TAS is used in combination with Spin Torque, it provides the same benefits of robust stability with practical write currents as when used with field-induced writing.

For the electronics industry, the impact of finding a highly manufacturable and cost effective MRAM technology will be significant, because it may significantly erode market share that currently belongs to DRAM, embedded SRAM and embedded Flash.

First, let's look at the cost parameters. One way to evaluate memory technologies is to compare the bit cell areas. When doing this, it is important to compare 'like-feature' processes (i.e. 65nm to 65nm) in order to make an apples-to-apples comparison. It is common in the industry to measure bit-cell sizes in 'f-squared' terms. For example, in a 65nm technology, the minimum lithographic feature is, obviously, 65nm, and the unit of comparison is how many 65nmx65nm squares are used in the layout of a memory bit cell. In this way, one can consider how a memory technology will scale during process shrinks, as well as compare bit cells at different geometry nodes. In today's world, SRAM bit cells (for embedded applications)are typically about 100-150 $f^2$, DRAM bit cells are typically 6-10 $f^2$, and Flash bit cells are typically 4-10 $f^2$. First generation MRAM typically sizes at 30-40 $f^2$, making it smaller only than SRAM. But this size advantage over SRAM is very important, because most SOC's include one or several very large embedded SRAMs, typically consuming 25-50% of the chip area. Even first generation MRAM looks attractive as a replacement for embedded SRAM, with the prospect of reducing the embedded SRAM component of the chip area by a factor of 3-4, while maintaining convenient RAM memory organization and speeds high enough for general computation. The second generation MRAM bit cells are likely to further improve the density situation, starting at 20-30 $f^2$, and progressing to as low as $10f^2$. Even further down the road, MRAM technologists talk about multi-bit per cell, leading to cell sizes well below $10f^2$.

Embedded MRAM can replace embedded Flash technology in SOC, in addition to its SRAM replacement potential. Because MRAM typically only adds 3 or 4 masks to a conventional process flow (inserted in the metal-via backend process), it offers the advantage of 6-9 fewer mask steps than an embedded Flash process. It has the added benefits of low-voltage writing and infinite endurance, which Flash cannot provide. Although MRAM bit cells are quite a bit larger than embedded Flash, this advantage of fewer masks makes economic sense for all but the very largest embedded Flash instances.

Figure 5



MRAM combines the best characteristics of FLASH, SRAM and DRAM.

Even more exciting for MRAM is the prospect of providing system architects with a new set of semantics for memory design. The combination of non-volatility, infinite endurance, high-speed read/write, and low-cost open a new door for system architects.  The likely applications are in storage, mobile communications, networking, automotive, and high-performance computation. Some tantalizing application prospects are instant-on PC's and smart-phones, much faster communication systems, and power-fault-proof configurations of network backbones.

Finally, the holy grail of MRAM research is the ultimate replacement of DRAM. The Spin Torque technologies in the laboratory today provide a path to bit cells in the 10 $f^2$ range. As MRAMs approach this bit cell density, applications that use DRAM will look to MRAM as a higher functionality, cost-effective alternative to conventional DRAM. Helping to spur this possible transition are concerns in the DRAM industry that the storage capacitor technologies used in so many previous generations of DRAM will start running into serious scaling problems below 45nm. These advanced MRAM technologies threaten a serious shift in the roughly $50 billion in yearly DRAM revenue.

In summary, SOC's may be possible in the near term, with embedded MRAM that enables both non-volatile capability and an SRAM replacement that is three to four times denser than current 6-transistor memory technology, with only modest changes to the underlying CMOS processes.  At a reasonable level of maturity, MRAM will become a cost effective alternative to DRAM. It thus has the potential to displace tens of billions of dollars of DRAM business, while at the same time providing system designers exciting new capabilities leveraging their high speed and infinite-endurance non-volatility.  The physics and process research that is underway today pushes the MRAM cost roadmap into intersection with the DRAM roadmap. The successful development of these second generation MRAM technologies will touch businesses and consumers everywhere.