Magnetoresistive Random Access Memory

Michael C. Gaidis
IBM T.J. Watson Research Center

Abstract

Magnetoresistive Random Access Memory (MRAM) offers the potential of a universal memory – it can be simultaneously fast, nonvolatile, dense, and high-endurance. MRAM differs from earlier incarnations of magnetic memory in that MRAM tightly couples electronic readout with magnetic storage in a compact device structure competitive with state-of-the-art semiconductor memories. Small-scale demonstrations have realized much of the potential of MRAM, but shrinking the cell size or embedding the memory with logic circuitry creates difficult challenges. This chapter provides an overview of the basic MRAM magnetic structure, including an explanation of the functions of various elements in the complex multilayer magnetic film stack. Principles of MRAM circuit design and operation are covered, with a detailed look at the design of a single-bit cell. Also included is a discussion of the techniques developed to fabricate large arrays of MRAM devices with high yield. MRAM reliability issues and the potential for scaling MRAM for future device generations concludes the chapter.

Contents

1. Magnetoresistive Random Access Memory (MRAM)

Through the merging of magnetics (spin) and electronics, the burgeoning field of "spintronics" has created MRAM memory with characteristics of non-volatility, high density, high endurance, radiation hardness, high speed operation, and inexpensive CMOS integration. MRAM is unique in combining all the above qualities, but is not necessarily the best memory technology for any single characteristic. For example, SRAM is faster, flash is more dense, and DRAM is less expensive. Stand-alone memories are generally valued for one particular characteristic: speed, density, or economy. MRAM therefore faces difficult odds in competing against the aforementioned memories in a stand-alone application. However, embedded memory for application-specific integrated circuits or microprocessor caching often demands flexibility over narrow performance optimization. This is where MRAM excels. It can be called the "handyman of memories" for its ability to flexibly perform a variety of tasks for a relatively low cost.[1] While one may hire a specialist to rewire an entire house's electrical circuitry or install entirely new plumbing, a handyman with a flexible toolbox is a much more reasonable option for repairing a single electrical outlet or leaky sink. And, the handyman may be able to repair a defective electrical circuit discovered while in the process of repairing leaky plumbing.

A semiconductor fabrication facility that has MRAM in its toolbox is more likely to tailor circuit designs to a customer's individual needs for optimal performance at reasonable cost. Table 1 shows how the characteristics of MRAM compare to other embedded memory technologies at the relatively conservative 180nm node. The remainder of the present chapter will review the state of the art in MRAM technology: how it works, how its memory circuits are designed, how it is fabricated, potential pitfalls, and an outlook for future use of MRAM as devices scale smaller.

		eSRAM	**eDRAM**	**eFlash**	**eMRAM**
Size	Cell Area (μm^2)	3.7	0.6	0.5	1.2
Size	Array Effic.	65%	40%	30%	40%
Cost	Add'l Process	0	20% (4 msk)	25% (8 msk)	20% (3 msk)
Speed	Read Access	3.3 nsec	13 nsec	13 nsec	15 nsec
Speed	Write Cycle	3.4 nsec	20 nsec	5000 nsec	15 nsec
Power	Data Retention	400 μA	5000 μA	0	0
Power	Active Read	15 pC/b	5.4 pC/b	28 pC/b	6.3 pC/b
Power	Active Write	15 pC/b	5.4 pC/b	31,000 pC/b	44 pC/b
Endur.	Write	Unlimited	Unlimited	1e5 cycles	Unlimited
Rad Hard		Average	Poor	Average	Excellent

Table 1: Embedded memory comparison at the 180nm node. Shaded cells indicate where MRAM has a distinct advantage. Relative comparisons should hold through scaling to the 65nm node.[2]

2. Basic MRAM

MRAM (magnetoresistive RAM) differs from earlier incarnations of magnetic memory (magnetic RAM) in that MRAM tightly couples electronic readout with magnetic storage in a compact device structure. In the early second half of the 20th century, the most widely used RAM was a type of magnetic RAM called ferrite core memory. These memories utilized tiny ferrite rings threaded by multiple wires used to generate fields to write or to sense the switching of the magnetic polarity in the rings.[3] Highly valued for its speed, reliability, and radiation hardness, approximately 400kB of this core memory was used in early IBM model AP-101B computers on the space shuttle. With the advent of compact, reliable, and inexpensive semiconductor memory, the $1mm^2$ cell size of the core memory could no longer compete – and, in 1990, the space shuttle converted to battery-backed semiconductor memory with around 1MB capacity.[4]

For magnetic memory to compete again in the RAM arena, miniaturization on the scale of semiconductor integrated circuitry had to be implemented. This was stimulated by the discovery in 1988 of giant magnetoresistance (GMR) structures which provided an elegant means of coupling a magnetic storage (spin) state with an electronic readout, and created the field of spintronics.[5] It relies on the phenomenon wherein electrons in certain ferromagnetic materials will align their spins with the magnetization in the ferromagnet. In essence, this is a result of a greater electron density of states at the Fermi level for electrons with spin aligned parallel to the magnetization in the ferromagnet. While passing current along two ferromagnetic films in close proximity, one can influence the transport of the electrons by adjusting the relative orientation of the two films' magnetization. As illustrated in Figure 1, for parallel orientation, electrons are less likely to suffer resistive spin-flip scattering events, but for antiparallel orientation, electrons will exhibit a stronger preference for scattering and thus an increase in resistance will be apparent. The different resistance values for the high resistance state (R_{high}) and the low resistance state (R_{low}) can be used to define a magnetoresistance ratio (MR) as in equation 1:

$$MR = \frac{\left(R_{high} - R_{low}\right)}{R_{low}} \quad . \qquad (1)$$

MR values for GMR devices are in the 5 – 10% range for room temperature operation.

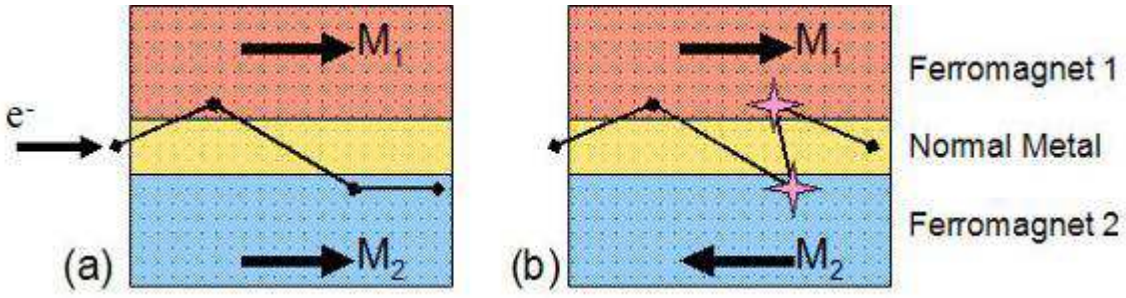


Figure 1: Illustration of the GMR principle. For parallel alignment (a), electron flow is subject to fewer resistive spin-flip scattering events than for antiparallel alignment (b).

By choosing different coercive fields for the two ferromagnets, one can create a so-called *spin-valve* MRAM structure with configuration similar to that shown in Figure 1. For example, ferromagnet 1 can be chosen to have a high coercivity, thus fixing its magnetization in a certain direction. Ferromagnet 2 can be chosen with a lower coercivity, allowing its magnetization direction to fluctuate. For a magnetic field sensor such as used in disk drive read heads, small changes in the magnetization angle of ferromagnet 2 induced by an external magnetic field can be sensed as changes in resistance of the spin valve. Because the spin-valve sensitivity to external fields can be substantially better than inductive pickup, such devices have enabled dramatic shrinkage of the bit size in modern hard drives. An alternative use for the spin-valve structure is found if one designs it to utilize just two well-defined magnetization states of ferromagnet 2 (e.g., parallel or antiparallel to ferromagnet 1). Such spin-valve designs serve as a binary memory device, and have found application in rad-hard nonvolatile memories as large as 1Mb.[6] Drawbacks of this type of memory are

- relatively low magnetoresistance, providing only low signal amplitudes and thus longer read times,
- low device resistance, making for difficult integration with resistive CMOS transistor channels,
- in-plane device formation which is more difficult to scale to small dimensions than devices formed perpendicular to the plane.

Solutions to these problems can all be found in the magnetic tunnel junction (MTJ) MRAM. The MTJ structure is similar to the GMR spin-valve in that it uses the property of electron spins aligning with the magnetic moment inside a ferromagnet. Instead of passing current in-plane through a normal metal between ferromagnets, however, the MTJ passes current perpendicular to the plane, through an insulating barrier separating two ferromagnets. Figure 2 shows an MTJ structure in its simplest form where one can envision the electric current impinging first on a ferromagnet which acts as a spin polarizer, then passing through the tunnel barrier and into a second ferromagnet which acts as a spin filter. The separation of polarizing and filtering functions is enabled by the physical thickness of the tunnel barrier, noting that the tunneling process preserves electron spin. The tunneling conductance will be proportional to the product of electron densities of states on each side of the barrier, and in general for ferromagnets there will be a larger density of states near the Fermi level for electrons polarized parallel to the magnetization of the ferromagnet as opposed to electrons polarized antiparallel to the magnetization of the ferromagnet. For polarizer and filter magnetizations aligned in the same direction, the density of states for spin-polarized electrons is large on both sides of the barrier, and the conductance of the structure is relatively high. For anti-parallel alignment of the polarizer and filter, the density of states available for spin-polarized electrons to tunnel into is somewhat reduced, and the conductance of the structure is relatively low. Proposed around 1974 [7], the first demonstrations of MTJs used Fe/Ge/Co multilayer stacks, but only showed appreciable MR (14%) at 4 K temperatures.[8] It was not until 1995 that improvements in materials processing techniques and the use of robust aluminum oxide tunnel barriers began to show reasonably large MR (18%) for MTJ devices at room temperature.[9] This breakthrough

brought about huge investments from numerous companies and ushered in a new era in the field of spintronics.

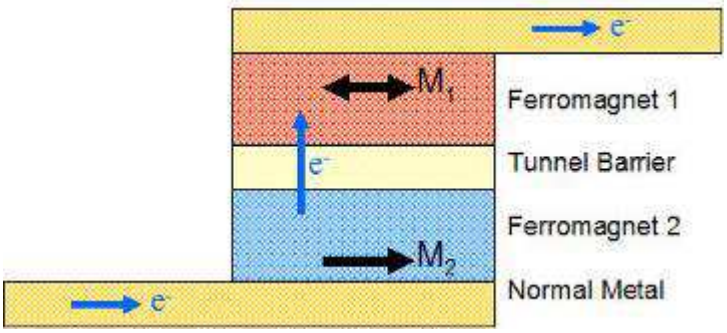


Figure 2: Simple magnetic tunnel junction structure. Ferromagnet 2 acts as an electron spin polarizer, and ferromagnet 1 acts as an electron spin filter, with magnetization either parallel or anti-parallel to the magnetization of ferromagnet 2. Parallel magnetizations generally result in lower device resistance than anti-parallel magnetizations.

3. MTJ MRAM

The structure illustrated in Figure 2 can store binary information in the direction of magnetization within ferromagnet 1 (the "free layer"), provided the magnetization within ferromagnet 2 (the "pinned layer") remains fixed in a predetermined direction. An asymmetry induced in the structure from device shape or intrinsic magnetic anisotropy can stabilize preferred orientations for the free layer to be one of either parallel to or anti-parallel to the pinned layer, thus maximizing MR. A straightforward way to enable switching in the free layer without switching of the pinned layer is through the use of a material with low coercive field H_c for the free layer and a material with high H_c for the pinned layer. Figure 3a illustrates this technique with the hysteresis loops of a soft (low H_c) free layer and a hard (high H_c) pinned layer in isolation (i.e., not in the integrated MTJ stack structure). For operation at applied magnetic fields within the bounds set by H_c of the pinned layer, only the free layer will switch direction of magnetization. The hysteresis curve for the free layer demonstrates the necessary memory effect when the applied field is reduced to zero.

With the integrated multilayer structure of Figure 2, however, the hysteresis curves of the free and pinned layers in isolation are not straightforward predictors of the resistance states of the MTJ device. Because the pinned layer will maintain a remanence in zero applied field, there will be an offset imparted to the hysteresis loop of the free layer. (Note that there will be a similar offset of the pinned layer hysteresis loop imparted by the free layer's remanence, but for large enough H_{c2} there will be no affect on the device operation.) Figure 3b illustrates the effect of the pinned layer remanence on the magnetoresistive hysteresis loop R versus the applied field. For large remanence M_{r2}, the loop may shift so much that there is no longer a bistable memory for zero applied field. In principle, such an offset in memory product chips could be compensated by an external field applied from a permanent magnet incorporated into the chip packaging. This is somewhat impractical, however, due both to packaging cost and to stringent requirements of across-chip uniformity.

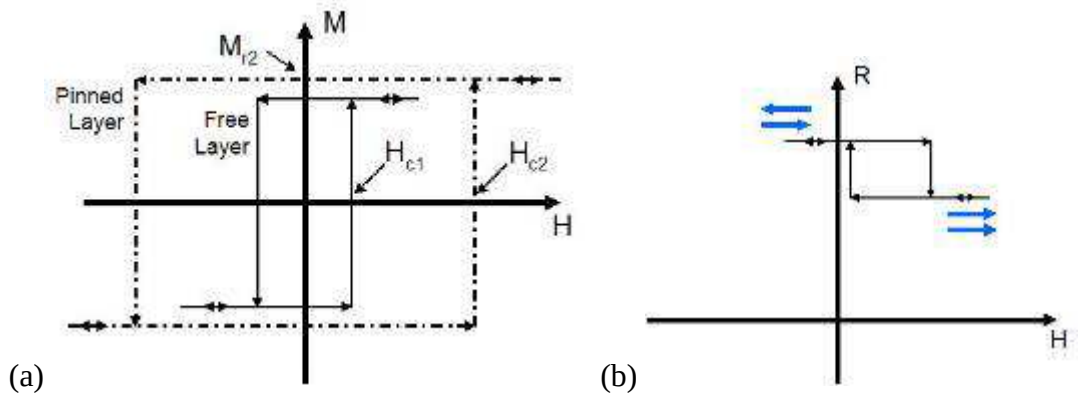


Figure 3: (a) Representative hysteresis curves of magnetization M versus applied field H, for a soft ferromagnet free layer and for a hard ferromagnet pinned layer in isolation. Coercive field H_{c2} is chosen large enough to keep the orientation of the pinned layer from switching while the free layer is being switched. M_{r2} represents the remanence from the pinned layer at zero applied field. (b) Resultant hysteresis of the MTJ resistance shown as a function of applied magnetic field. Due to the remaining magnetization from the pinned layer, the resistance loop is offset from zero applied field, and (as shown) can even result in but a single stable resistance at zero applied field. The blue arrows represent the magnetization state of the MTJ structure (anti-parallel or parallel).

Fortunately, clever manipulation of film properties has driven the evolution of several generations of magnetic tunnel junction structures, overcoming issues such as the offset field described above. Two such advances are illustrated in Figure 4. In Figure 4a, an antiferromagnet is exchange coupled to the pinned layer, providing a much larger effective coercive field for the pinned, or "reference" side of the tunnel junction.[10] With exceptional care to maintain a clean, smooth interface between the antiferromagnet and the pinned layer above it, one can obtain the strong exchange coupling between these films that is necessary to resist field switching. At least 1 to 1.5nm of ferromagnetic pinned layer must still remain in the stack to act as an electron spin polarizer, but when coupled to the antiferromagnet it can be extremely well pinned even if the ferromagnet has low H_c. By removing the need for a high H_c ferromagnet in the pinned layer, this structure allows some additional flexibility in the choice of ferromagnet pinned layer material. One can optimize for maximum electron spin polarization for best magnetoresistance, and one can choose film qualities for low remanence and thus less offset of the R vs. H hysteresis curve. Correspondingly, Figure 4b illustrates a representative improvement in offset, for comparison with Figure 3b from the simpler stack structure.

Although there is much benefit in using the simple antiferromagnet (AF)-pinned structure of Figure 4a, best device operation often calls for reducing the R vs. H hysteresis offset to an even smaller value. In this case, the flux-closed AF-pinned structure shown in Figure 4c can be tailored to give arbitrarily small offset fields. Here a synthetic antiferromagnet (SAF) is formed from two ferromagnets separated by a thin spacer layer. For common spacer layers of 0.6 to 1.0nm of Ru, one can obtain a strong antiparallel coupling between the two ferromagnets.[11] For reasonable external fields, this coupling forces them to be antiparallel and thus the thicknesses of the two ferromagnets can be balanced such that the external magnetic flux is negligible. Pinning

of one of these ferromagnet layers with an antiferromagnet gives a high effective H_c while at the same time causing negligible offset to the R vs. H hysteresis loop (Figure 4d).

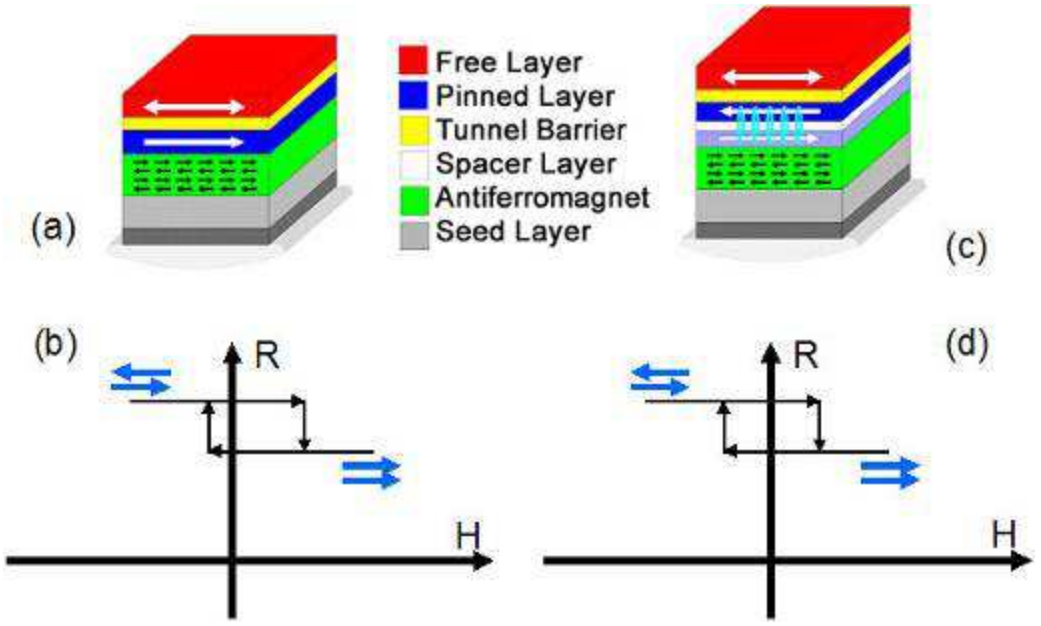


Figure 4: (a) Antiferromagnet-pinned reference layer structure with corresponding R vs. H hysteresis loop (b). Also shown (c) is a flux-closed antiferromagnet-pinned reference layer structure with corresponding R vs. H hysteresis loop (d).[2]

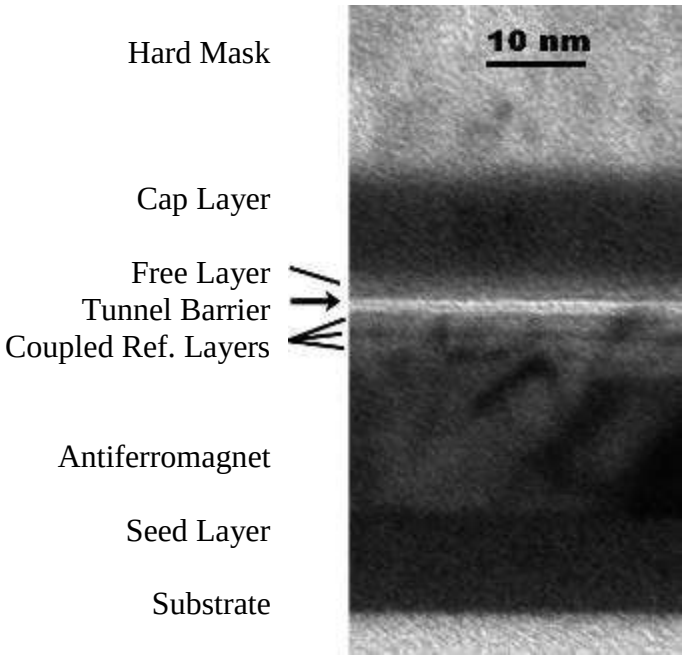


Figure 5: A TEM high resolution cross-section image of a MTJ stack with flux-closed, antiferromagnet-pinned reference layers.

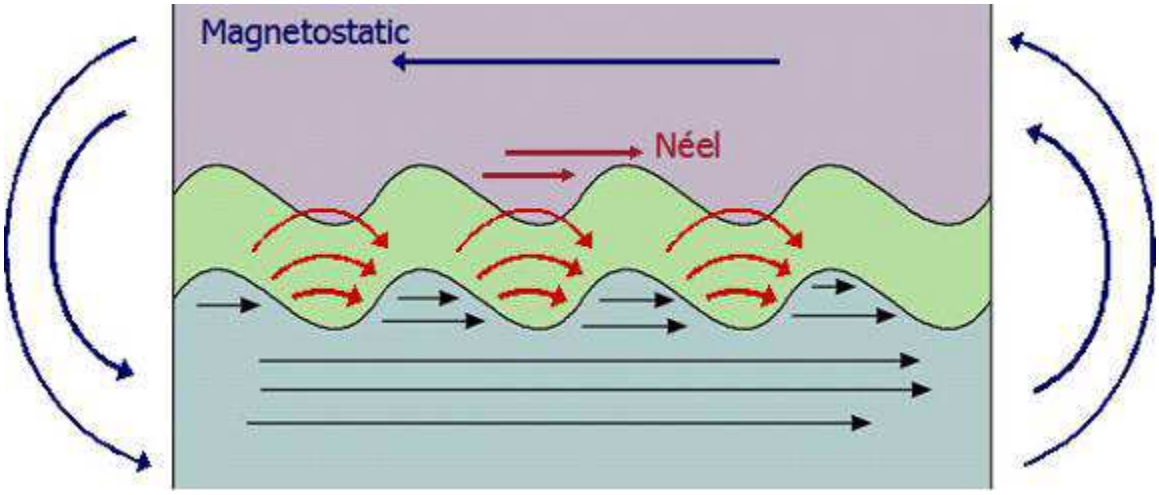


Figure 6: A schematic description of Néel coupling and how it relates to magnetostatic coupling. The rough-topped bottom film represents the pinned layer of Figure 4. Although exaggerated in the figure for clarity, actual roughness greater than one atomic monolayer is cause for concern. The green intermediate layer represents the tunnel barrier, and the layer above is the free layer. Black arrows in the bottom film represent the internal magnetization of the pinned layer, but due to the rough surface, the magnetic poles are uncompensated in the region of the tunnel barrier. The resultant field from these poles creates a Néel field which favors parallel orientation of the free and pinned layers. The magnetostatic demagnetization field from the ends of the pinned layer favors antiparallel orientation of the free and pinned layers, but since this is non-local, it is less important in breaking the symmetry of devices with multiple layers. [12]

Flux closing the reference layer ferromagnet works remarkably well in practice, particularly with recent advances in materials deposition tooling which enables tight control over film thicknesses for multilayer film structures covering entire 200 to 300mm wafers.[13] Figure 5 shows a cross-section TEM image of such a flux-closed reference layer MTJ stack. Some interesting features of the magnetics-related elements can be discerned from the TEM image and are discussed below.

3.1 Antiferromagnet

The antiferromagnet is generally a polycrystalline material like FeMn, PtMn, or IrMn. It is chosen and grown with several characteristics in mind:

- Interface roughness of the antiferromagnet must be sufficiently small so one can neglect Néel coupling (Figure 6) and ensure a smooth, pinhole-free tunnel barrier.
- Pinning strength must be large compared to the fields used to switch the free layer between its binary memory states.
- The blocking temperature of the antiferromagnet must be in a suitable range. To obtain ideal pinning of the ferromagnet reference layer, one anneals the antiferromagnet/ferromagnet bilayer above the blocking temperature T_B at which the exchange coupling between the films is zero. An applied magnetic field fixes the orientation of the ferromagnet, and then the bilayer is cooled. During cooling, the surface magnetization of the antiferromagnet aligns with the field-imposed ferromagnet magnetization. After cooling and removal of the field, exchange coupling across this interface keeps the ferromagnet pinned. One must choose an antiferromagnet with blocking temperature T_B below approximately 300°C to

minimize material diffusion and tunnel barrier degradation. In addition, T_B must be sufficiently above device operating temperatures, around 125°C.

- The antiferromagnet must be able to withstand process temperatures of the ensuing circuit integration. Roughly, this translates into saying that the components of the antiferromagnet should not dissociate and diffuse out of the layer for process temperatures below about 250°C.

3.2 Reference Layer

The reference layer closest to the tunnel barrier must act as an effective spin polarizer, so must be of thickness at least of order the electron spin-flip scattering length. This implies 1 to 1.5 nm is the minimum thickness of the layer closest to the tunnel barrier. For best flux closure and minimal offset to the free layer, the reference layer adjacent to the antiferromagnet will be of a similar thickness, although perfect zero free-layer offset may dictate small differences in the thicknesses. An upper limit to the thickness is set by the additional surface roughening and resultant Néel coupling that thicker films will generate. Reference layer materials are chosen for best spin polarization properties and compatibility with device processing techniques (e.g., minimal corrosion and thermal stability). Films of CoFe of order 2 nm thick are typically used, separated by the 0.6 – 1.0 nm exchange-coupling Ru layer.

3.3 Tunnel Barrier

Aside from the requirement of reasonable magnetoresistive properties, the tunnel barrier is chosen primarily for robustness. It must be extremely thin to ensure spin polarization is maintained during electron transit across the barrier, and the barrier must be able to survive under billions of cycles of electrical bias in its lifetime without developing pinholes or any substantial shift in resistance. Aluminum oxide has proven an extremely suitable candidate for such tunnel barriers, and is known to offer reasonable magnetoresistance for suitable magnetic pinned and free layers. Recent developments in tunnel barrier engineering show that magnesium oxide tunnel barriers can offer MR near 500% at room temperature, although MgO-barrier devices are not yet proven as robust, manufacturable layers in large arrays with good magnetic switching characteristics.[14] Aluminum oxide barrier devices can display MR near 100%, but tradeoffs in choice of magnetic materials for best switching characteristics, and in choice of operating point for best CMOS integration generally result in an MR less than 50%. Such MR is suitable for maintaining distinct resistance groupings of millions of devices in modern MRAM arrays, and increasing the MR is advantageous primarily in that it can reduce the necessary signal integration time to read the state of a device. Such a reduction is not a terribly strong driver at this time, as array read time is set as much by circuit overhead as by device signal-to-noise ratio. Increasing MR to 500% would likely result in only a 10-20% reduction in read duration. One place MgO barriers may soon establish a strong foothold is in the formation of highly transparent tunnel barriers. As device sizes shrink, the lower resistance-area product afforded by MgO will enable the best match to CMOS drive transistors and thus highest speed operation. Even more highly transparent tunnel barriers are under development for a class of devices using electron spin current to switch the device state.

3.4 Free Layer

The free layer shown in the TEM is reasonably thin, like the underlying pinned layers. It does have a minimum thickness limit set by the spin filtering characteristics: for thickness less than the approximate electron spin-flip scattering length, the magnetoresistance will begin to drop. This again sets the thickness at around 1.5 nm or more. Thicker free layers require additional energy to switch, so are undesirable for low-power operation. Of critical importance in the characteristics of the free layer is the need for well-defined magnetic states and well-behaved magnetic switching. As one cannot tailor the read or write circuitry to every individual device in megabit arrays of MRAM devices, it is critical that each device behave very much like all others in the array. Ill-defined magnetization states such as vortices, S-shapes, C-shapes, and multiple domains will add variability to the resistance measured by the circuitry, because electron spin polarization filtering may not be strictly parallel or antiparallel to the spin polarization imparted by the pinned layer. In addition, sensitivity of the film switching behavior to tunnel barrier and cap materials, or to device edge roughness or chemistry can impart variability to the write operation of the individual bits in megabit arrays. NiFe alloys are preferred for good magnetic behavior with reasonable corrosion resistance. Addition of Co or Fe to the NiFe, or dusting with Co or Fe between the tunnel barrier and NiFe layer can help to adjust magnetic anisotropy and improve MR. Layer thicknesses are typically in the 2 to 6 nm range for best low-power operation with good switching characteristics.

Several additional nonmagnetic elements are visible in the TEM image and discussed below.

3.5 Substrate

An ultra-smooth substrate is required as the starting point for smooth, uniform, and reliable tunnel barriers. Rough interfaces also result in increased Néel coupling, which is detrimental to device performance. Representative materials for the substrate are thermally oxidized silicon, or chemical-mechanical planarized (CMP) polished dielectrics such as silicon nitride, silicon oxide, or silicon carbide.

3.6 Seed Layer

An appropriate seed layer is required to obtain good growth conditions for the antiferromagnet, both to ensure a smooth top surface and to ensure good magnetic pinning strength. Given the high stress in some of the films in the MTJ stack, this seed layer is also critical for ensuring good adhesion to the substrate. It may be formed from tantalum nitride or permalloy (NiFe), for example.

3.7 Cap Layer

Proper choice of a cap layer is necessary to protect the free layer during further device fabrication processing. It is essential as a barrier or getter for contaminants, keeping the free layer clean and magnetically well-behaved. Often-used materials for this layer include ruthenium, tantalum, and aluminum. Choice of this material can also depend on its effect on the magnetic behavior of the free layer: certain cap materials can discourage smooth switching between free layer states, and can result in substantial “dead layers” which must be compensated for by a thicker free layer.

3.8 Hard Mask

A hard mask (as opposed to a "soft" photoresist mask) is used to enable patterning of the MTJ with industry-standard etch techniques. It also eases the integration with surrounding circuitry by providing a contact layer to connect the MTJ to wiring levels above. The hard mask material is largely chosen for its compatibility with subsequent processing in the fabrication route, and can be chosen from any number of metallic or dielectric materials.

The processing of MTJ structures to integrate them with CMOS circuitry is discussed in more detail later in the chapter.

4. MRAM Cell Structure and Circuit Design

4.1 Writing the Bits

The mechanism for switching the state of the free layer in MRAM lends itself well to array layout with conventional planar semiconductor design and fabrication. Figure 7 illustrates a typical rectangular MTJ array layout, with word lines arrayed beneath the devices and bit lines arrayed atop the devices. Current driven along the word lines or bit lines generates a magnetic field which imparts a torque on the magnetization of the device. In normal operation, the superposition of properly-sized "write" fields from both word line and bit line will enable a switching event to occur in the free layer of the device at the intersection of the two lines. The write fields are chosen small enough so as not to exceed the coercivity of the pinned layer. Potential pitfalls from this scheme include write errors from half-selected devices (i.e., those subjected to only a word line or a bit line field, but not both) and worse, write errors from near-neighbor half-selected devices (those subjected to a half select field, but only one row or column away from another active line).

The diagrams in Figure 8 give more details about the superposition of magnetic fields used to switch the active device. With the flux-closed antiferromagnet-pinned reference layer structure (Figure 4c) forming the MTJ, the single-layer free layer is switched with characteristics first described by Stoner and Wohlfarth.[15] A simple case is that of an elliptical-shaped MTJ with shape anisotropy defining an easy axis (the major axis of the ellipse) and a hard axis (the minor axis of the ellipse). To switch the magnetization, a hard-axis field is applied to tilt the free layer magnetization away from the easy axis energy minimum, and an easy-axis field is applied to "set" the magnetization of the device in the desired easy-axis direction – parallel or antiparallel to the pinned layer. With this Stoner-Wohlfarth (S-W) switching, relatively small operating margins are illustrated by the closeness of the green and purple dots to the S-W boundary in Figure 8b. In addition to accounting for spreads in the switching characteristics between devices, one must also budget in extra operating window for thermal activation errors and the disturb effects of half-selects and near-neighbor field interaction. Circuit designers will try to tailor the operating window for at least 10 years of error-free operation. Without use of error-correction techniques, one generally aims for operating margins to keep the activation energy for a bit error to greater than 60 k_BT, where k_B is the Boltzmann constant and T is the temperature. This imposes extremely tight

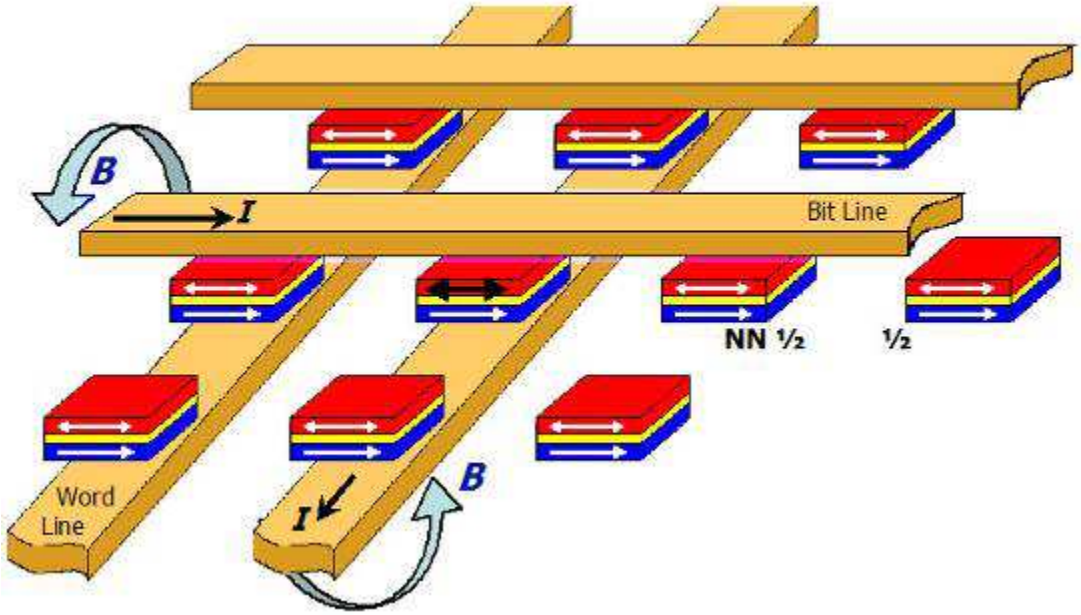


Figure 7: Illustration of a rectangular array of MTJ devices, with bit line and word line circuitry for writing the bits. Current-generated magnetic fields from a given bit line and word line are sufficient only to switch the device at the intersection of the two wires. Write errors are typically worse for devices in the half-select state (MTJs labeled "½" in the figure), where a word line or a bit line is active, but not both. The situation is even worse for near-neighbor half-selected devices ("NN ½" in the figure) where, for example, the device is in the column adjacent to the active word line, but is half-selected by the active bit line.

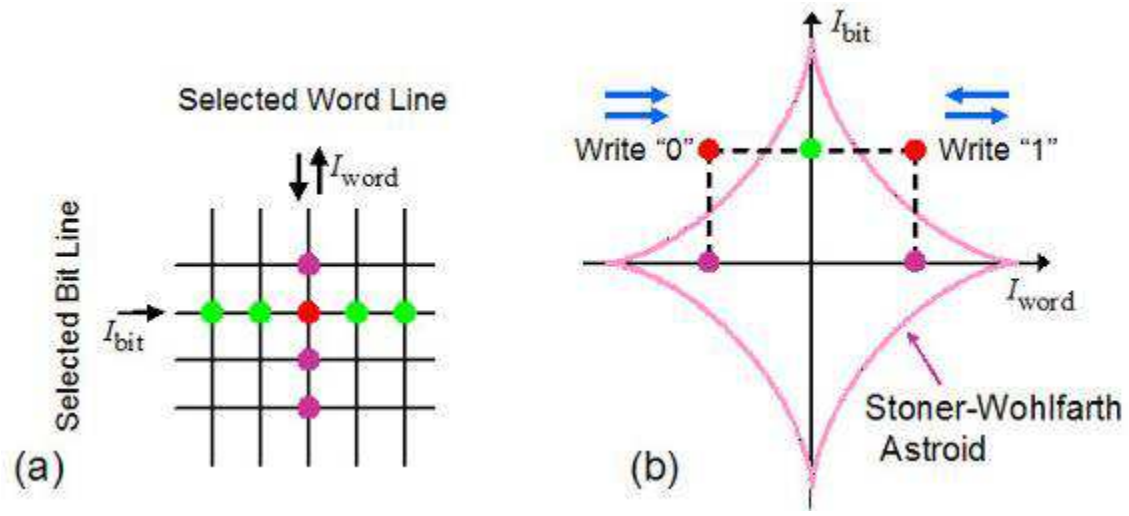


Figure 8: (a) Top-down schematic view of an MTJ array with rows and columns of bit lines and word lines with fields superposed to switch the device represented by a red dot. Devices shown as green and purple dots are half-selected devices. Those green and purple devices adjacent to the red device are near-neighbor half-selected devices. (b) A graph showing necessary bit line and word line current values needed to switch a desired device. The colored dots on the plot correspond to the devices represented by colored dots in Figure 8a. For suitable choice of word line and bit line currents, one can ensure switching of the desired device without switching half-selected devices.

requirements on how uniform the array must be in terms of switching, described in equation form by the array quality factor (AQF):

$$AQF = \frac{H_{sw}}{\sigma_{Hsw}} \quad . \qquad (2)$$

Here, H_{sw} is the average switching field of the devices and σ_{Hsw} is the standard deviation of the switching field distribution of all elements in the array. Roughly, the AQF must be larger than about 30 to ensure lifetime of 10 years, although some relief can be gained through the use of error correction techniques.

Toggle MRAM was invented to circumvent the difficulties faced by S-W MRAM in terms of operating margin for half-selected bits.[16] As illustrated in Figure 9a, the structure has taken the flux-closed antiferromagnet-pinned reference layer structure (Figure 4c) a step further by also flux-closing the ferromagnetic free layer. This is done by depositing a spacer layer atop the free layer ferromagnet, followed by a second ferromagnet. The spacer can be chosen (as in the pinned layer) with a spacer that enhances antiparallel coupling, or the spacer can be chosen with zero or even with some parallel coupling characteristics to decrease the write field needed to switch the bit. The magnetizations of two ferromagnets in the free layer will point in opposite directions, and their balance and proximity will flux-close the layers so there is little field seen emanating from the structure at a distance. The write operation of this toggle-mode structure is illustrated in Figure 9b. Noting the colors assigned to represent the magnetization of the free layers in Figure 9a (green for the top layer, and red for the bottom layer), the plots at the top of Figure 9b show the relative orientation of the two magnetizations. Note that the initial state is such that the magnetization of the MTJ has easy (preferred) axis at 45 degrees to the word and bit lines, rather than aligned parallel to one of them. Figure 9b illustrates the need for staggered timing of word line and bit line write-field pulses. To switch the state of the free layers, a first magnetic field is applied from the word line along the positive *y*-direction. This magnetic field cants the magnetizations of both free layers as they try to align to the field. The antiparallel nature of the magnetic coupling between the free layers prevents the magnetizations from both fully lining up with the applied word field, so long as the field is not too large to overwhelm this antiparallel state. Once the magnetizations are canted sufficiently, there is a net magnetic moment to the free layers, and this moment can be grabbed like a handle by the field now imparted by the bit line. The bit line applies a field in the positive *x*-direction and the net moment of the two free layers follows this bit line field. The word line field is then shut off, and the net moment continues to rotate around towards the applied bit line field. As the bit line field is shut off, the free layer magnetizations relax into their energetically-favorable antiparallel configuration, but now with magnetizations exactly opposite to those at the start.

The name “toggle-mode device” comes from the characteristic that cycling the word and bit lines in this manner will always switch the state of the device. To set a bit in a particular state, a read operation must be performed to determine if a write “toggle” operation is required. Aside from this drawback and the additional complexity of the magnetic stack, there are several advantages to the toggle-mode structure:

- As alluded to above, the write operating margins can be substantially larger than for devices with S-W switching. Rather than a S-W astroid boundary, the toggle-mode devices exhibit an L-shaped boundary that does not approach the word or bit line axes. The potential for half-select errors is dramatically reduced, and the requirement on AQF is approximately halved.
- In principle, shape anisotropy is not required to ensure the bit has only two preferred states for binary memory. One can utilize intrinsic anisotropy of the ferromagnetic free layers to define two such states. This allows one to use circular MTJ devices for smallest memory cell size.
- The flux-closed nature of the free layers greatly reduces dipole fields emanating from the free layer. Such fields can affect the energetics of nearby devices, resulting in variability of switching characteristics, depending on the states of nearby devices. Thus, with flux-closed free layers, nearby devices can be packed in closer proximity for improved scaling.

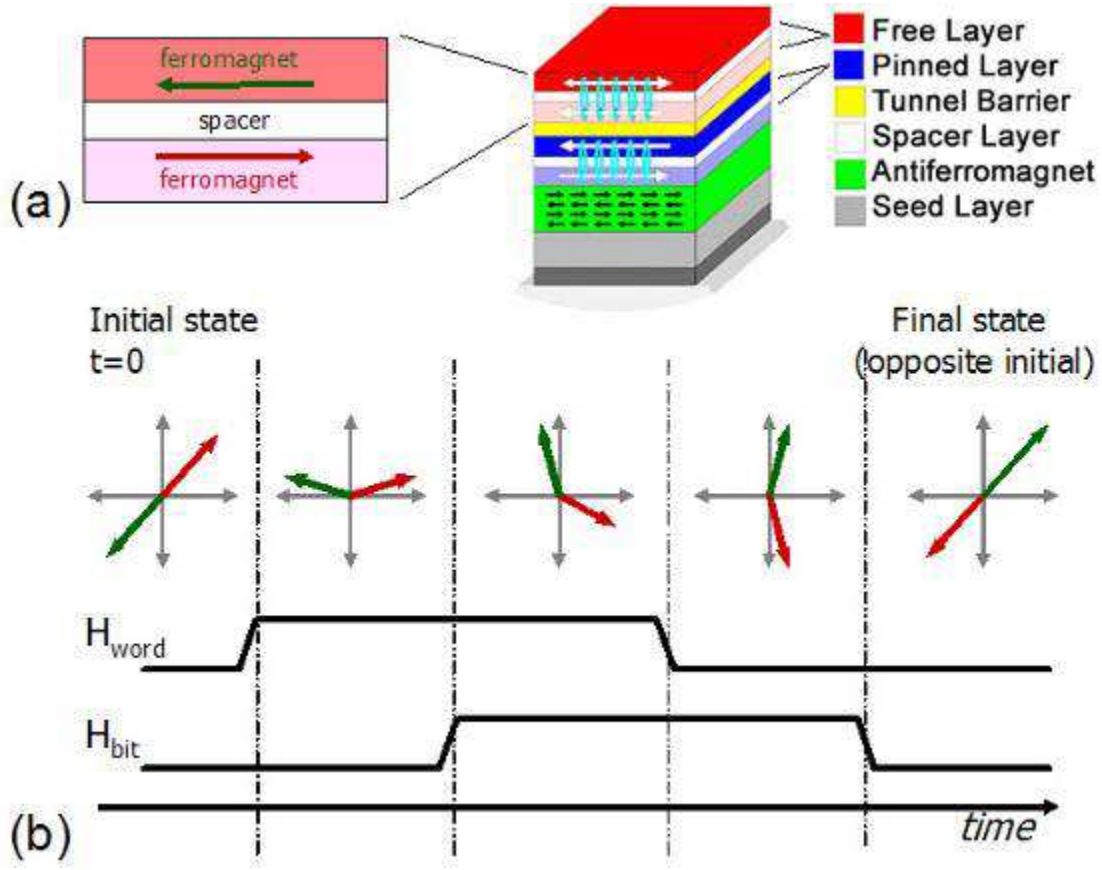


Figure 9: (a) Structure of the toggle-mode MTJ stack, and (b) time evolution of the free layer switching. See the text for discussion.

4.2 Reading the Bits

The array structure illustrated in Figure 7 is often termed a "cross-point cell" (XPC) structure. More specifically, XPC refers to the case where the MTJ devices are

located at the cross points of the bit and word lines, and are directly connected to the bit and word lines above and below the MTJ stack. This structure offers extremely high packing density for the lowest cost memory. The write mechanism is reasonably straightforward as described above, so long as the MTJ resistance is not so low that it shunts the write currents. More troublesome is that the read mechanism suffers from reduced signal-to-noise ratio in this XPC structure. To read the resistance state of a XPC bit, a bias is applied between a desired bit line and word line, and the resistance is measured. However, due to the interconnected nature of the XPC structure, not only the resistance of the cross-point device is measured – there are parallel contributions of resistance from many other devices along "sneak paths" that include traversing additional sections of bit and word line. Due to the resulting loss of signal, the device must be read much more slowly to allow for integration to improve the signal-to-noise ratio. Device read times can be substantially longer for such XPC structures, making this type of memory far less desirable than one which could be read as fast as DRAM, for example.

The solution to the problem of sneak paths is to insert an isolation mechanism that ensures read currents will only traverse a single MTJ device. For example, this can be achieved by placing a diode in series with each MTJ. Although this seems simple when drawn as a circuit schematic on paper, it is actually more straightforward to put a FET in series with each MTJ, and assign a second word line to control the read operation. Figure 10 illustrates the "FET cell" circuit structure, with separate word lines for the write and the read operations. The bit line is used for both read and write operations.

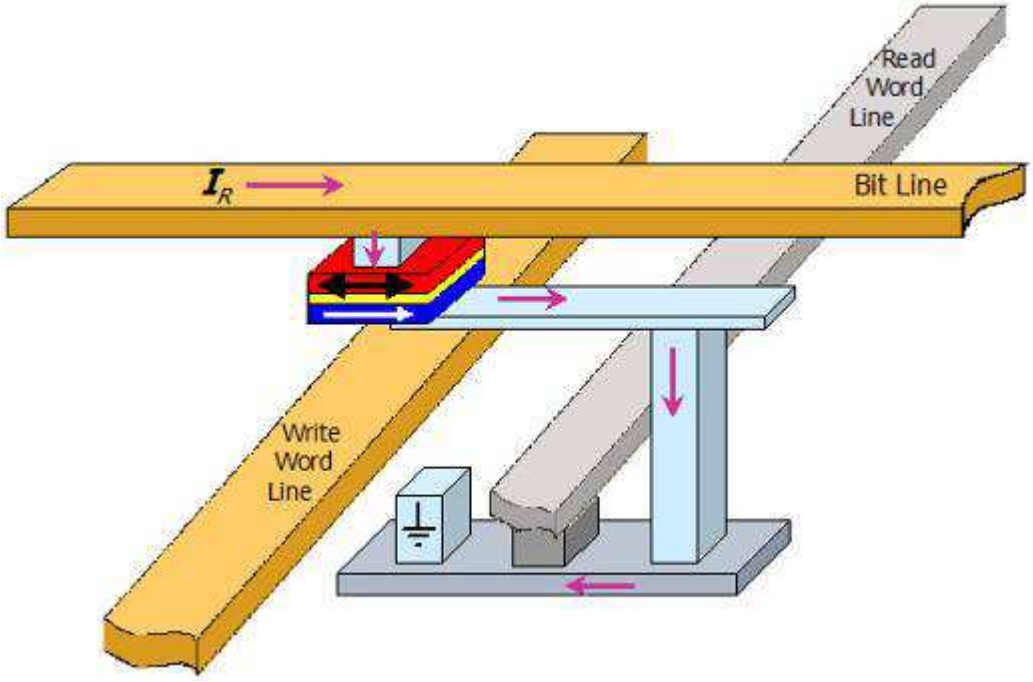


Figure 10: FET cell circuit topology, showing individual word lines for reading and for writing. A FET located in the silicon beneath the MTJ is used to switch on only the device being read, thus preventing leakage of read currents (purple arrows) through nearby MTJ devices. Additional conductor elements in this structure (compared to Figure 7) include a contact between the bit line and the top of the MTJ, a local metal strap connecting the base of the MTJ with a via chain that connects to the underlying FET.

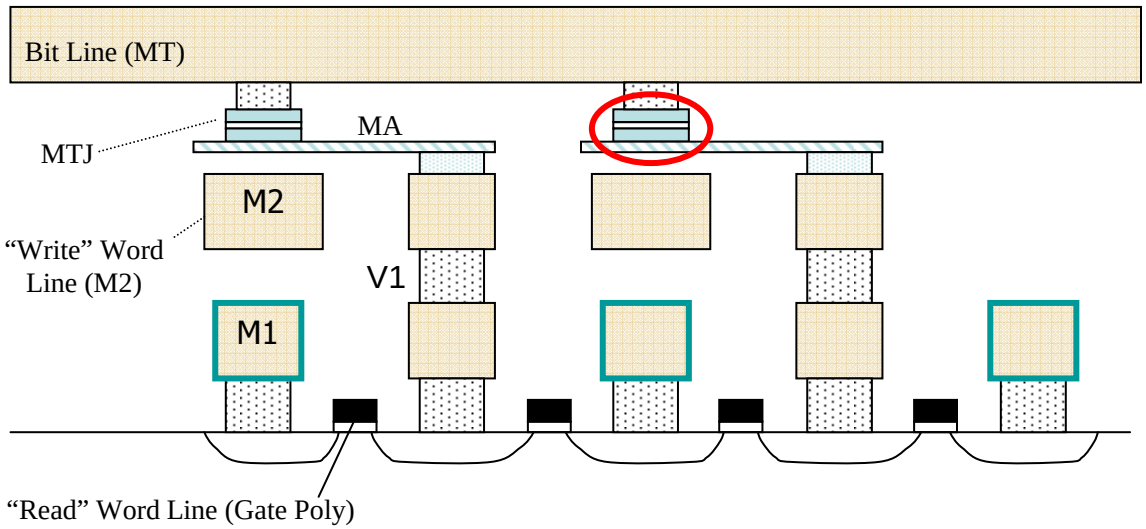


Figure 11: A cross-section of the FET cell topology, with two adjacent cells shown atop the silicon CMOS front-end of line (FEOL) structure. The red oval encloses the critical components for MRAM implementation. As cell size is determined primarily by the MTJ and via chain above the via V1, two FETs can be used for each MTJ in order to achieve lower resistance and some redundancy. Thus, the FET gates on either side of a V1 via chain will be connected to the same "read" word line. Wires formed in the first level of metallization (M1) (outlined in green) form a grid at a reference potential. M2 denotes the second level of metallization. The reader is referred to Reohr *et al.*[17] for a more details on such structures.

Figure 11 illustrates the implementation of the Figure 10 circuit structure suitable for a densely-packed array of MTJs. Structural additions to standard CMOS circuitry include:

- the via contact VJ between the bit line and the top of the MTJ stack,
- the MTJ device
- the local metal strap MA between the bottom of the MTJ stack and the via to M2
- the via VA between the MA strap and the M2 wiring, which serves to isolate the MTJ from the write word line while providing connection to the underlying FET structure for reading.

Slightly higher packing density may be achieved with a mirror-cell design, where adjacent bits mirror each other. The simple unmirrored design of Figure 11 is preferable to minimize across-array nonuniformity due to inter-level misalignment and inter-cell magnetic interference. Megabit and larger MRAM memories are formed from multiple subarrays with size determined largely by the resistance of the bit and word lines. There is desire to keep applied voltage low, for CMOS compatibility and best array efficiency. The required current to generate the necessary MTJ switching fields then sets a maximum length on the bit or word line depending on the resistive voltage drop. Bootstrapped write drivers can be used to allow smaller write drivers with improved write current control.[18] A 16Mb MRAM under development at IBM utilizes 128Kb subarrays as shown in Figure 12, with 512 word lines and 256 bit lines of active memory elements.

The read operation is performed with sense amplifiers that compare the desired bit to a reference cell. The reference cell uses two adjacent MTJs fixed in opposite states in a configuration that acts like an ideal mid-point reference between the R_{high} and the R_{low} states.[18] Four bit lines are activated in a given cycle and are uniformly spaced along

the height of the array to reduce magnetic interference between activated bit lines during the write operation, and to minimize distance from the activated bit lines to the sense amps during a read operation. Additional reference bit lines are located within the array, with one set shared by sense amplifiers 0 and 1, and one set shared by sense amplifiers 2 and 3.

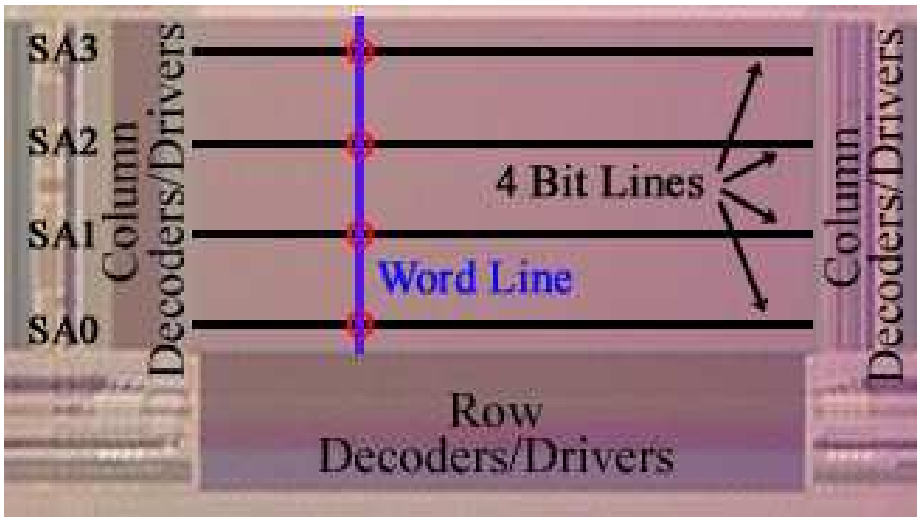


Figure 12: Photograph of a 128Kb subarray, showing locations of the sense amplifiers (SA), the row and column decoders and drivers, and the concurrent activation of four bit lines with one word line for an x4 organization of the block. A single MTJ cell is indicated by a circle at the intersection of a word line (WL) and bit line (BL).

The array driving circuitry for MRAM memories is commonly standardized to an asynchronous SRAM-like interface for easy interchangeability in battery-backed SRAM applications. The IBM 16Mb chip uses an x16 architecture that is prevalent in mobile and handheld applications with packaging intended for simple direct replacement of SRAM chips. Shown in Figure 13, the 16Mb chip measures 79mm^2 with individual memory cells of 1.42μm^2, for an array efficiency of nearly 30%. Array efficiency can be improved by using more metal layers and by eliminating some of the developmental test mode structures used in this chip. Reduction of standby current for power-critical applications is achieved through extensive use of high threshold, long channel FET devices and careful grounding of inactive terminals in the arrays and in the write driver devices.[18]

Redundant elements are included in the chip to allow correction of defective array elements. Such redundancy is implemented with fuse latches and address comparators in a manner consistent with industry-standard memory products. The CMOS base technology is quite mature so the focus of the redundancy is on the MRAM features. Single-cell failures or partial word line fails (from MRAM reference cell defects) are considered the most likely defects. The redundancy architecture favors replacement of word lines to capture the partial word line fails from MRAM reference cell defects. Redundancy domains are implemented at a high level in the block hierarchy so as to span several blocks and be able to effectively fix random defects.[18]

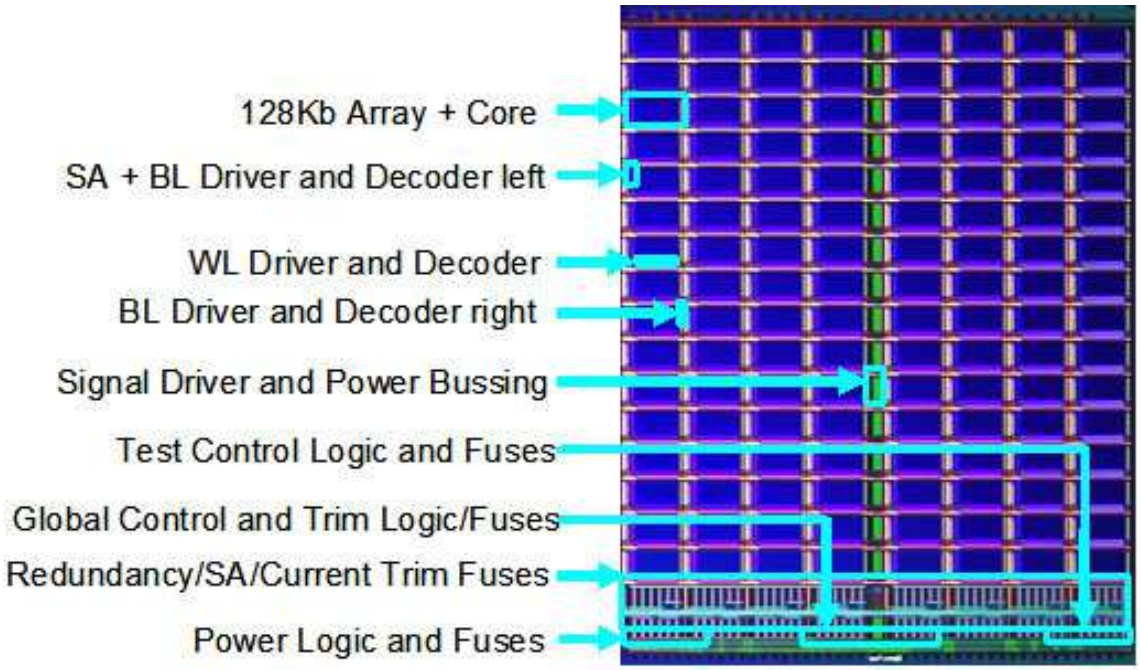


Figure 13: Photograph of a 16Mb MRAM chip, showing locations of 128Kb array cores (8 columns of 16 rows) and support circuitry.[18]

4.3 MRAM Processing Technology and Integration

The implementation of MRAM hinges on complex magnetic film stacks and several critical steps in back-end-of-line (BEOL) processing. Cell size is presently limited by the size of the MTJ devices and driving wires, and older, mature CMOS FEOL technology can be used without limiting performance. Fabrication of the FET-cell circuit, from the CMOS FEOL through the MRAM BEOL, can encompass several hundred process steps, resulting in the fully-functional structure shown in Figure 14. The MRAM-critical portion of the circuit is a relatively small part of the entire configuration. After the last standard CMOS step (the M2 wire completion), there remains the need to pattern the shallow vias, the MTJs, the local interconnects, and at least one level of wiring with contact to MTJs and the functional circuitry below. Even for simple functional circuits, five or more photomask levels are required to complete the MRAM-centric portion of the structure.

Process Steps

In conjunction with the steps outlined in Figure 14, below is a discussion of the important considerations for the process steps in the fabrication of the MRAM-specific levels.

1,2. VA Contact Via and ILD: The VA via provides a path for read current to flow from the local (MA) metal strap down through a via chain to the underlying read transistor. The most critical aspect of this module is that it must form a substrate smooth enough for good magnetic stack growth.

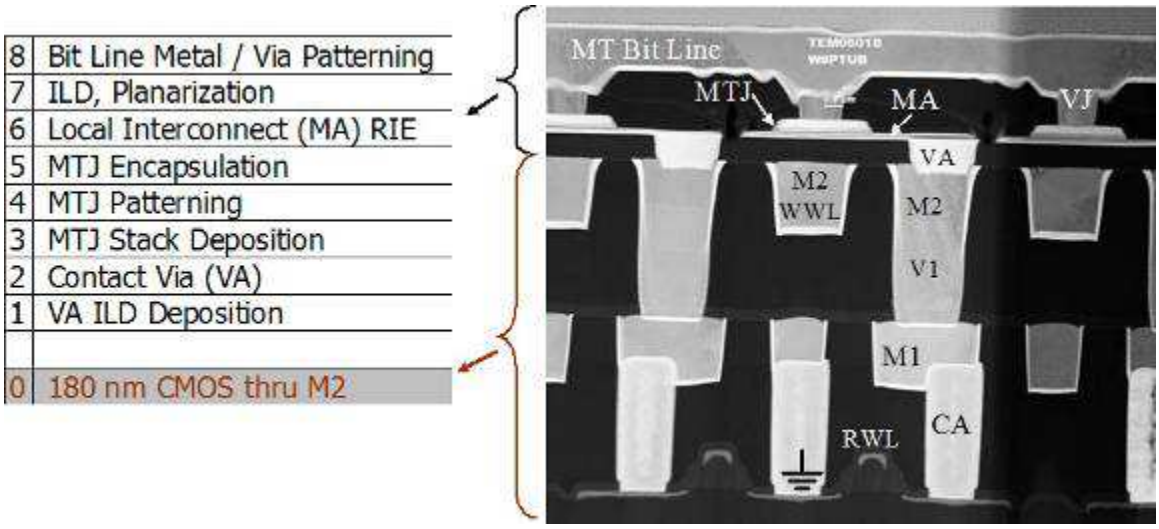
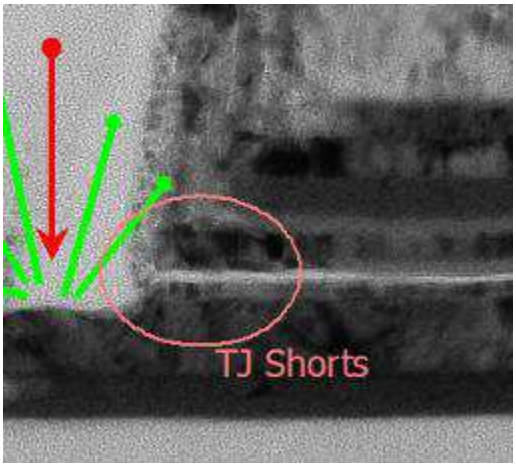


Figure 14: Cross-section of a product cell, showing the integration of MRAM with CMOS, and the process steps used for the MRAM-specific layers. "ILD" refers to interlayer dielectric.

3. Magnetic film stack deposition: Arguably the most essential technological advance in enabling MTJ MRAM was the development of tooling for large area deposition of extremely uniform films with well-controlled thickness. Such tooling has proven suitable for deposition of magnetic, spacer, and tunnel barrier films with sub-Ångström uniformity across 200mm and even 300mm wafers.[13] The critical aluminum oxide tunnel barrier is generally formed by depositing a thin aluminum layer, followed by exposure to an oxidizing plasma.[19]

4. Tunnel junction patterning: A commonly-used and straightforward approach to patterning the MTJs is with the use of a conducting hard mask. The conducting mask is later utilized as a self-aligned stud bridging the conductive MT wiring to the active magnetic films in the device. A thick hard mask, however, introduces additional difficulty in that it can shadow the etch being used to pattern the magnetic devices. Such shadowing can add an element of variability into the size of the devices, and can also result in metal redeposits on the sidewall of the device structure. As illustrated in Figure 15, sidewall redeposits are particularly troublesome for commonly used MRAM stack materials because the materials do not readily form volatile RIE byproducts that give some isotropic character to the etch. Directional physical sputtering is the main mechanism for etching of the stack materials.[20] Because the difficulties in etching the magnetic stack materials often outweigh the benefits of a simpler process integration scheme, it is often preferred to use a thinner hard mask for less etch shadowing, and an additional via level (VJ in Figure 14) to connect the top of the MTJ with the bit line wiring.

(a) 

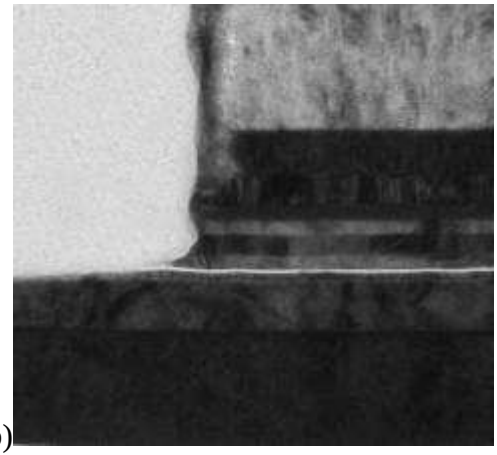
(b) 

Figure 15: (a) Shown is a TEM image of the edge of a MTJ after etch to define the free magnetic layers. The etch has progressed to a depth just past the oxide tunnel barrier (the lightest contrast film in the stack). Consequences of sputter-etch redeposits on the sidewall of a MTJ device can be seen as a short-circuited tunnel barrier and a poorly defined edge with thick redeposits. (b) Improvement in the etch conditions can result in a much cleaner sidewall and elimination of residues that would short-circuit the tunnel junction.

5. MTJ encapsulation: Silicon nitride and similar compounds are desirable for their adhesion to the MA and MTJ metal surfaces, and for strong interfacial bonds that inhibit migration of metal atoms along the dielectric / metal interfaces. Such metal migration is one well-documented cause of MTJ thermal degradation, and can limit processing temperatures in patterned MTJ devices to below 300º C.[21] The use of TEOS (tetra ethyl ortho silicate) as a precursor in the deposition of silicon oxide films is known to [22] offer the benefits of a relatively inert depositing species which can readily diffuse into spaces adjacent to high-aspect ratio structures, even at temperatures below 250º C.

6. MA Patterning: For suitable thickness of seed and reference layers, the series resistance of layers remaining after MTJ etch is small enough to impart negligible dilution of the MTJ MR signal. This simplifies processing as a dedicated film need not be created for the MA strap, and the reference or seed layers of the magnetic stack can perform double duty. Like in the MTJ etch, the MA etch may be subject to the problem of non-volatile etch byproducts redepositing along the hard mask sidewalls.

7,8. ILD Deposition and Planarization and Wiring: After the MA metal strap is patterned, an interlayer dielectric is deposited in which to house the counterelectrode wiring layers VJ and MT. The counterelectrode wiring is formed with well-established semiconductor-industry Damascene techniques.

As alluded to in Figure 11, the MRAM-specific elements form but a small portion of the entire integrated circuit. For rapid characterization of these MRAM-specific elements, one need not perform a fully CMOS-integrated wafer build, but instead can make do with a subset of the process integration steps to focus only on critical magnetics issues.[23]

5. MRAM Reliability

One of the strong selling points of MRAM is its reliability: write endurance is expected to be essentially infinite, the magnetics are intrinsically rad-hard, and its nonvolatile memory storage can eliminate soft errors in many applications. As in any new technology struggling for successful commercialization, there are certain aspects of the new technology that are unproven and require demonstration of reliability. Areas of potential reliability risk include: [24]

5.1 Electromigration in the write word and bit lines, resulting from high write current density. Current pulses of 10mA are typical for conservative wire cross-sections of 0.2 μm^2, corresponding to a current density of 5 MA/cm^2. This alone represents a serious challenge to the reliability in the array, and can potentially be worsened by local disruptions to the quality and thickness of wire material. The VJ vias of Figure 14, or direct connection between the bit line and the MTJ hard mask in the thick hard mask integration scheme discussed above, can impact the bit line wiring electromigration resistance.

Electromigration issues can potentially be improved through the use of bidirectional switching currents, which fit neatly into toggle-mode MRAM operation, but cost in array efficiency. One promising method for reducing electromigration stress is through the use of ferromagnetic liners in a U-shape around the bit and word lines. These liners serve to focus the magnetic field onto the MTJs in the desired row or column, and can increase the effective field by as much as a factor of 2 for a given current.[25] Figure 16 illustrates the use of ferromagnetic liners around the bit line. Similar, but inverted,

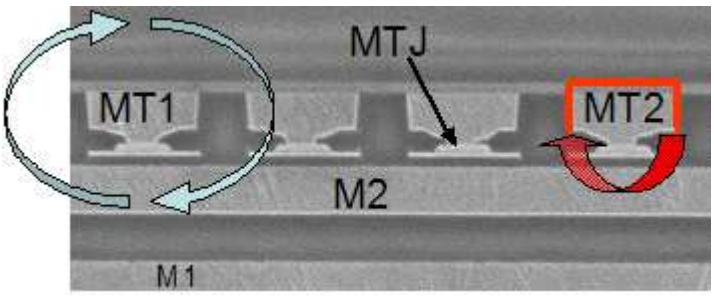


Figure 16: A cross-section image of a product array, from viewpoint perpendicular to that of Figure 14. Blue arrows around bit line wire MT1 suggest the magnetic field configuration generated by a current through wire MT1 – it is loosely contained, with only moderate magnitude at the MTJ free layers. Conversely, the wire MT2 exhibits an enhanced field magnitude due to its localization by the ferromagnetic film (red lines) surrounding the copper MT2 wire.

structures can be formed around the write word line (M2 in Figure 16) to enhance the field from that wire. The potential reduction in necessary current to obtain a required switching field can dramatically reduce electromigration issues. Not only do ferromagnetic liners offer potential reduction in current density, but they also improve electromigration performance relative to conventional copper processes. By reducing the interface diffusion of copper atoms, ferromagnetic cladding on the top surface of the MT wire enhances electromigration reliability to an extent similar to that seen in the industry by advanced Ta/TaN or CoWP capping processes.[26]

One added benefit of the ferromagnetic liner field focusing is the reduction of near-neighbor disturb effects. Because the field is better focused on devices along the

desired word line and bit line, adjacent devices are less likely to be switched by near-neighbor fields, or the combination of near-neighbor fields and thermal activation.

5.2 Tunnel Barrier Dielectrics are subject to reliability concerns because of the extremely thin nature of the barrier and related susceptibility to pinholes or dielectric breakdown. Aluminum oxide tunnel barriers have so far proven quite robust. Time dependent dielectric breakdown (TDDB) and time dependent resistance drift (TDRD) have been examined in 4 Mb arrays and found to exceed requirements for 10 year lifetime.[27] The voltage stresses on the tunnel barrier are relatively modest, as the read operation takes place at 100 – 300 mV because MR is higher for lower voltage. The write operation is performed with one side of the MTJ floating, so there is no significant voltage stress on the MTJ during the higher-power write pulse.

5.3 BEOL Thermal Budget for MRAM devices (< 250 – 300°C) is significantly lower than for conventional semiconductor fabrication processes (~ 400°C), to prevent degradation to the MTJs. This can affect the intrinsic quality of dielectrics being used in the BEOL, and can worsen seam and void formation around topographical features being encapsulated. Low thermal budget also prevents the use of certain post-processing passivation anneals, and packaging materials and processes. The move to lead-free solder with increased solder reflow temperatures is a further challenge for MRAM.

5.4 Film Adhesion is a serious concern with the multiple new materials being introduced into the integrated process. Novel etch and passivation techniques being used also may leave behind poorly-adherent layers which cannot be subjected to harsh wet cleans without MTJ exposure and degradation. Delamination risks must be mitigated through specially-developed dry and wet cleans, the use of materials with tuned stress, and choice of materials with compatible thermal expansion.

6. The Future of MRAM

As of July, 2006, MRAM products like the 4Mb memory shown in Figure 17 have been available from Freescale Semiconductor.[28] The market space targeted by Freescale includes networking, security, data storage, gaming, and printer data logging and configuration storage. From a customer viewpoint, this product means fewer part counts, a higher level of performance, higher reliability, environmentally friendlier, and lower cost solution than their current approaches, such as battery backed SRAM.

Progressing downwards from the available 180nm technology, future generations of MRAM are expected to utilize the same magnetic infrastructure with only evolutionary improvements, to below the 90nm node. Constraining the scaling are the following concerns:

- *Near-neighbor interactions*. In packing the devices closer together, magnetic fields emanating from a given device can affect the switching behavior of devices nearby – and can be dependent on the given device's free layer state. In addition, the write wires for switching a given device will perturb neighboring devices to a greater extent as the neighboring devices come closer. It remains unclear how well one can suppress these effects with the use of flux-closed MTJ layers and ferromagnetic

cladding of write wires. Additional techniques such as enhanced-permeability dielectric (EPD) encapsulating films may be required.[29]

- *Increased switching fields*. As devices are scaled to smaller volumes, the anisotropy field must be increased to compensate and maintain activation energy greater than 60 k_BT.[30] Write fields will scale to be of similar magnitude to the anisotropy field, and will increase superlinearly with inverse device size. Like the MTJs, the write wires must scale to smaller footprint, making it more difficult to accommodate the increasing switching fields. In addition, ferromagnetic cladding of the wires becomes less effective because of the bending energy of the flux inside the cladding as the wire corner radius sharpens. EPD device encapsulations will help in this regard.
- *Device-to-Device Variability*. Process-induced line-edge roughness will become a more substantial fraction of the total device width, so that edge irregularities may become more effective at pinning the domains so they do not switch smoothly. Total device area and aspect ratio will also exhibit larger spreads, both from line-edge roughness and from variability in lithography. Reduced aspect ratio for tighter packing density will also decrease AQF, as anisotropy field is more sensitive to shape for devices with smaller aspect ratio.[30]

Each of these concerns is not a fundamental limitation, but rather a practical limitation that can likely be overcome with sufficient, but perhaps prohibitively expensive, investment in materials development and processing techniques. Hard physical limits do not appear to set in until superparamagnetism becomes important – for device sizes below 20nm.[30] This is substantially below the limits suggested by the aforementioned practical issues.

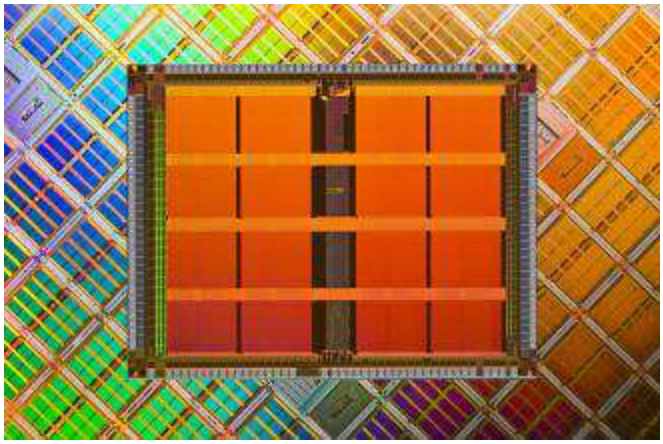


Figure 17: Photograph of a MR2A16A 4Mb MRAM chip atop a wafer filled with such chips, presently available from Freescale semiconductor. (Courtesy of G. Grynkewich and Freescale Semiconductor)

Even with the practical limits to scaling conventional MRAM, one can expect to see revolutionary modifications to standard MRAM cell such that magnetic random access memory will be available with far greater densities, lower cost, and faster operation. Outside the scope of this chapter are impressive developments and exciting new proposals in the areas of:

- thermally assisted MRAM for reduced power requirements, [31]

- spin-momentum transfer (SMT) MRAM for scaling to advanced process nodes and extremely small active memory devices, [32]
- domain-wall memory for very high density serial storage, and [33]
- embedded MRAM as a replacement for embedded flash and low-density on-chip SRAM, for high-performance microprocessor cache memory and other ASIC applications. [34]

This chapter has provided an overview of the rapid development in MRAM technology over the past decade. Many major hurdles for MRAM product development have been surmounted in the face of funding limits set by competition with the huge silicon industry. Now that MRAM devices have found a toe-hold in the marketplace, new applications will be found and MRAM development will proceed at an even faster pace over the next decade. Perhaps soon we will see magnetic RAM in spacecraft again.

7. Acknowledgements

W.J. Gallagher for figures and editorial assistance; IBM's Materials Research Laboratory (MRL) for process development and fabrication; P. Rice, T. Topuria, E. Delenia, and B. Herbst for TEM imaging; contributions from J. DeBrosse, T. Maffitt, C. Jessen, R. Robertazzi, E. O'Sullivan, D.W. Abraham, E. Joseph, J. Nowak, Y. Lu, S. Kanakasabapathy, P. Trouilloud, D. Worledge, S. Assefa, G. Wright, B. Hughes, S.S.P. Parkin, C. Tyberg, S.L. Brown, J.J. Connolly, R. Allen, E. Galligan; and M. Lofaro for advances in CMP. We also wish to acknowledge that the work summarized here was supported in part by the Defense Microelectronics Activity (DMEA) and built on prior work done with Infineon (now Qimonda) within the MRAM Development Alliance and on earlier MRAM work at IBM that was supported in part by DARPA.

References

[1] J. DeBrosse, personal communication
[2] W.J. Gallagher and S.S.P. Parkin, "Development of the magnetic tunnel junction MRAM at IBM: From first junctions to a 16-Mb MRAM demonstrator chip," *IBM J. Res. & Dev.* **50**, pp. 5-23 (2006). Also, sincere thanks to John DeBrosse and the help of John Barth, Chung Lam and Ron Piro.
[3] J. Jones, "Coincident Current Ferrite Core Memories," *Byte*, **1**, pp 6-22 (July, 1976)
[4] J.F. Hanaway and R.W. Moorehead, "Space Shuttle Avionics System," NASA SP-504 (1989), available at http://klabs.org/DEI/Processor/shuttle/sp-504/sp-504.htm
[5] G. Binasch et al., PRB vol. 39, pp. 4828-4830 (1989), and M.N. Baibich et al., "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices," Phys. Rev. Lett. vol. 61, p. 2472 (1988). For further details, see C. Heiliger, P. Zahn, and I. Mertig, "Microscopic Origin of Magnetoresistance," *Materials Today*, **9**, pp. 46-54 (2006).
[6] H. Kaakani, "Radiation Hardened Memory Development at Honeywell," IEEE Aerospace Conf., Mar. 10-17, 2001, Big Sky, MT, vol. 5, pp. 2273-2279, and R.R. Katti, *J. Appl. Phys.* **91**, p. 7245 (2002)
[7] J. Slonczewski, IBM internal
[8] 1975 Juliere (CNR-France) - first MTJ demonstration Fe/Ge/Co, DR/R~14% at 4.2 K. M. Julliere, "Tunneling Between Ferromagnetic Films," *Phys. Lett. A* **54**, pp. 225–226 (1975).

[9] T. Miyazaki and N. Tezuka, "Giant Magnetic Tunneling Effect in Fe/Al_2O_3/Fe Junction," *J. Magn. Magn. Mater.* **139**, No. 3, L231–L234 (1995).
[10] Review articles can be found in A.E. Berkowitz and K. Takano, "Exchange Anisotropy – A Review," *J. Magn. Magn. Mater.* **200**, pp. 552-570 (1999), and J. Nogues and I.K. Schuller, "Exchange Bias," *J. Magn. Magn. Mater.* **192**, pp. 203-232 (1999).
[11] M.D. Stiles, "Exchange Coupling in Magnetic Multilayers," *Nanomagnetism: Ultrathin Films, Multilayers and Nanostructures (Contemporary Concepts of Condensed Matter Science, Vol 1)*, eds. D. Mills and J. A. C. Bland, New York: Elsevier, 2006 pp 51-77.
S.S.P. Parkin and M.G. Samant, "Magnetic Random Access Memory with Thermally Stable Magnetic Tunnel Junction Cells," U.S. Patent 6,518,588 (2003).
S.S.P. Parkin, N. More, and K.P. Roche, "Oscillations in Exchange Coupling and Magnetoresistance in Metallic Superlattice Structures: Co/Ru, Co/Cr, and Fe/Cr," *Phys. Rev. Lett.* **64**, 2304-2306 (1990).
J.M. Slaughter, R.W. Dave, M. DeHerrera, M. Durlam, B.N. Engel, J. Janesky, N.D. Rizzo, and S. Tehrani, "Fundamentals of MRAM Technology," *J. Superconductivity: Incorporating Novel Magnetism* **15**, 19-25 (2002).
S.S.P. Parkin et al., "Exchange-Biased Magnetic Tunnel Junctions and Application to Nonvolatile Magnetic Random Access Memory," *J. Appl. Phys.* **85**, 5828-5833 (1999).
[12] B.D. Schrag et al., "Néel 'orange-peel' coupling in magnetic tunneling junction devices," *Appl. Phys. Lett.*, **77**, pp. 2373-2375 (2000).
[13] C-7100EX Sputter Deposition Tool from Canon Anelva Corporation, Tokyo, Japan (see http://www.canon-anelva.co.jp/english), and Timaris Sputter Deposition Tool from Singulus Technologies, Kahl, Germany (see http://www.singulus.de)
[14] S.S.P. Parkin et al., "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nature Materials* **3**, pp. 862-867 (2004).
J. Hayakawa, S. Ikeda, Y.M. Lee, F. Matsukura, and H. Ohno, "Effect of High Annealing Temperature on Giant Tunnel Magnetoresistance Ratio of CoFeB/MgO/CoFeB Tunnel Junctions," *Appl. Phys. Lett*., **89**, pp. 232510-232512 (2006).
[15] E.C. Stoner and E.P. Wohlfarth, "A Mechanism of Magnetic Hysteresis in Heterogenous Alloys" *Phil. Trans. Roy. Soc*. **A240**, pp. 599–642 (1948).
[16] L. Savtchenko, B.N. Engel, N.D. Rizzo, M.F. DeHerrera, and J.A. Janesky, "Method of Writing to Scalable Magnetoresistance Random Access Memory Element," U.S. Patent 6,545,906 (2003).
M. Durlam et al., "A 0.18 um 4Mb toggling MRAM," in IEDM Tech. Dig., 2003, p. 995.
D. Worledge, "Spin Flop Switching for Magnetic Random Access Memory," *Appl. Phys. Lett.* **84**, 4559-4561 (2004).
D. C. Worledge, "Single-Domain Model for Toggle MRAM," *IBM J. Res. & Dev.* **50**, pp. 69–79 (2006).
[17] W. Reohr, H. Hoenigschmid, R. Robertazzi, D. Gogl, F. Pesavenot, S. Lammers, K. Lewis, C. Arndt, Y. Lu, H. Viehmann, R. Scheuerlein, L.-K. Wang, P. Trouilloud, S. Parkin, W. Gallagher, and G. Mueller, "Memories of Tomorrow," *IEEE Circuits & Devices* **18,** 17–27 (2002).
[18] T. M. Maffitt, J. K. DeBrosse, J. A. Gabric, E. T. Gow, M. C. Lamorey, J. S. Parenteau, D. R. Willmott, M. A. Wood, and W. J. Gallagher, "Design Considerations for MRAM," *IBM J. Res. & Dev.* **50**, pp. 25–39 (2006)

D. Gogl et al., “A 16Mb MRAM Featuring Bootstrapped Write Drivers,” *IEEE J. Solid State Circuits* **40**, pp. 902-908 (2005).
[19] J.G. Zhu and C. Park, “Magnetic Tunnel Junctions,” *Materials Today*, **9**, pp. 36-45 (2006).
[20] S.J. Pearton et al., “Dry Etching of MRAM Structures,” *Mat. Res. Soc. Symp. Proc.* **614**, F10.2.1 – F10.2.11 (2000).
[21] M.G. Samant, J. Luning, J. Stohr, and S.S.P. Parkin, “Thermal Stability of IrMn and MnFe Exchange-Biased Magnetic Tunnel Junctions,” *Appl. Phys. Lett.* **76**, 3097-3099 (2000).
[22] J. Crowell, L. Tedder, H. Cho, F. Cascarano, and M. Logan, “Model Studies of Dielectric Thin Film Growth: Chemical Vapor Deposition of SiO2,” *J. Vac. Sci. Technol.* **A8**, 1864-1870 (1990).
[23] M. C. Gaidis, E. J. O'Sullivan, J. J. Nowak, Y. Lu, S. Kanakasabapathy, P. L. Trouilloud, D. C. Worledge, S. Assefa, K. R. Milkove, G. P. Wright, and W. J. Gallagher, “Two-Level BEOL Processing for Rapid Iteration in MRAM Development,” *IBM J. Res. & Dev.* **50**, pp. 41–54 (2006).
[24] B. Hughes, “Magnetoresistive Random Access Memory (MRAM) and Reliability,” *Proc. IEEE 42nd Annual Int’l Reliability Physics Symp., Phoenix, AZ*, pp. 194-199 (2004).
[25] M. Durlam et al., “A 1-Mbit MRAM Based on 1T1MTJ Bit Cell Integrated with Copper Interconnects,” *IEEE J. Solid-State Circuits* **38,** pp. 769-773 (2003).
[26] D.A. Gajewski, T. Meixner, B. Feil, M. Lien, J. Walls, ”Electromigration of MRAM-Customized Cu Interconnects with Cladding Barriers and Top Cap,” *IEEE Integrated Reliability Workshop Final Report*, pp. 90-92 (2004).
J. Walls et al., “Improved Electromigration Resistance of Copper Interconnects using Multiple Cladding Layers,” unpublished; see www.IP.com, Document ID IPCOM000009315D (2004).
[27] J. Åkerman et al., “Demonstrated Reliability of 4-Mb MRAM,” *IEEE Trans. Device Materials Reliability* **4**, pp. 428-435 (2004).
J. Åkerman et al., “Reliability of 4-Mbit Toggle MRAM,” *Mater. Res. Soc. Symp. Proc.*, **830**, pp. 191-200 (2005).
[28] B.N. Engel, J. Åkerman, B. Butcher, R.W. Dave, M. DeHerrera, M. Durlam, G. Grynkewich, J. Janesky, S.V. Pietambaram, N.D. Rizzo, J.M. Slaughter, K. Smith, J.J. Sun, and S. Tehrani, “A 4-Mbit Toggle MRAM Based on a Novel Bit and Switching Method,” *IEEE Trans. Mag.* **41**, 132 (2005).
http://www.freescale.com/mram
[29] S.V. Pietambaram, N.D. Rizzo, R.W. Dave, J. Goggin, K. Smith, J.M. Slaughter, and S. Tehrani, “Low-Power Switching in Magnetoresistive Random Access Memory Bits using Enhanced Permeability Dielectric Films,” *Appl. Phys. Lett.* **90**, 143510 (2007).
[30] R.P. Cowburn, “Superparamagnetism and the Future of Magnetic Random Access Memory,” *J. Appl. Phys.* **93**, 9310 (2003).
[31] D. W. Abraham and P. L. Trouilloud, “Thermally Assisted Magnetic Random Access Memory,” U. S. Patent 6,385,082, May 7, 2002.
I. L. Prejbeanu, W. Kula, K. Oundadjela, R. C. Sousa, O. Redon, B. Dieny, and J.-P. Nozieres, “Thermally Assisted Switching in Exchange-Biased Storage Layer Magnetic Tunnel Junctions,” *IEEE Trans. Mag.* **40**, 2625 (2004).

O. Redon, M. Kerekes, R. Sousa, L. Prejbeanu, H. Sibuet, F. Ponthennier, A. Persico, and J. P. Nozières, "Thermo Assisted MRAM for Low Power Applications", *Proc. 1st Int'l Conf. on Memory Technology and Design* (ICMTD-2005), Giens, France, May 21–24, 2005, pp. 113–114.
[32] M.D. Stiles and J. Miltat, "Spin Transfer Torque and Dynamics," in *Spin Dynamics in Confined Magnetic Structures III: Topics in Applied Physics 101*, eds. B. Hillebrands and A. Thiaville, Berlin: Springer, 2006 pp 225-308.
Y. Huai, F. Albert, P. Nguyen, M. Pakala, and T. Valet, "Observation of Spin-Transfer Switching in Deep Submicron-Sized and Low-Resistance Magnetic Tunnel Junctions," *Appl. Phys. Lett*. **84**, 3118 (2004)
G. D. Fuchs et al., "Spin Transfer Effects in Nanoscale Magnetic Tunnel Junctions," *Appl. Phys. Lett.* **85**, 1205 (2004).
[33] S.S.P. Parkin, "Shiftable Magnetic Shift Register and Method of Using the Same," U.S. Patent 6,834,005, 2004;."System and Method for Writing to a Magnetic Shift Register," U.S. Patent 6,898,132, 2005.
[34] S. S. Iyer, J. E. Barth, Jr., P. C. Parries, J. P. Norum, J. P. Rice, L. R. Logan, and D. Hoyniak, "Embedded DRAM: Technology Platform for the Blue Gene/L chip," *IBM J. Res. & Dev*. **49**, No. 2/3, 333–350 (2005).