

Development of the magnetic tunnel junction MRAM at IBM: From first junctions to a 16-Mb MRAM demonstrator chip

W. J. Gallagher
S. S. P. Parkin

This paper reviews the remarkable developments of the magnetic tunnel junction over the last decade and in particular, work aimed at demonstrating its potential for a dense, fast, and nonvolatile random access memory. The initial focus is on the technological roots of the magnetic tunnel junction, and then on the recent progress made with engineered materials for this device. Following that, we discuss the development of the magnetic random access memory (MRAM) technology, in which the magnetic tunnel junction serves as both the storage device and the storage sensing device. The emphasis is on work at IBM, including demonstrations of basic capabilities of the technology and work on a 16-Mb “product demonstrator” design in 180-nm node technology, which was targeted to be a realistic test bed for the MRAM technology. Performance and cost are compared with those of competing technologies. The paper also serves as an introduction to more specialized papers in this issue on MRAM device physics, magnetic tunnel junction materials and device characterization, MRAM processing, and MRAM design.

Introduction

Magnetism has played important roles in the evolution of computer memory and storage technology. Most apparent are the enduring roles played in secondary or permanent computer system storage, evolving more or less continuously through approximately ten orders of magnitude in density from tape and drums in the late 1940s to the disk and tape storage systems of today. However, it should be remembered that magnetism in the form of the core memory element was the overwhelmingly predominant active computer memory device used from the mid-1950s into the early 1970s. Throughout this time, magnetic core devices maintained the dominant active memory role because they were both the most reliable and the fastest memory devices available. For instance, in 1968, core memories had a 4,096-bit capacity and an access time of ~ 120 ns. Semiconductor memory would not match that performance for several more years [1]; by then, magnetic memory in IBM computers had improved more than five orders of magnitude in density. Further

evolution of core memory would have been possible, but by the late 1960s it had become evident that semiconductor bipolar and FET (DRAM) memories would undergo even more rapid advances in density and performance, and at lower power. Higher-performance plated-wire and thin-film magnetic memories that were being developed at the time never saw widespread use. Further improvements in magnetic memory density, though without performance or cost benefits, were demonstrated in the 1970s by yet another type of magnetic memory technology, bubble memory, although none competed economically with semiconductor memories and were never pursued in IBM products. For a review of IBM memory research, development, and products—from magnetic core and thin-film memories to semiconductor memories—see the paper by Pugh et al. [1].

For specialized components for certain military and aerospace systems, core memories in plated-wire form continued to be the preferred solution for many

©Copyright 2006 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/06/\$5.00 © 2006 IBM

years [2]. This market, addressed by Honeywell, valued nonvolatility and radiation hardness more than density or performance, which by the 1980s were better for silicon-based memories. Magnetic memories with read operations based on the use of anisotropic magnetoresistance rather than magnetic induction were also pursued for this market beginning in the mid-1980s [3]. Later, the discovery of giant magnetoresistance in magnetic multilayers and sandwiches and the invention of the spin valve device [4, 5], which showed great promise for read-head sensors for hard disk drives, led Tang et al. of IBM [6] to propose the use of a spin-valve-based MRAM, and development interest shifted to this type of magnetic RAM for the military and aerospace market [7].

A new magnetic device, the magnetic tunnel junction, began to emerge in the mid-1990s with the potential to enable a new “universal” memory that could simultaneously achieve high speed, high density, and nonvolatility. This broader range of applicability has led to widespread research and development activity aimed at demonstrating magnetic tunnel junction random access memory with commercial market potential. This paper reviews the history of the magnetic tunnel junction and its emergence as an attractive device for a computer memory. The review begins in the next section of this paper, in which we cover the emergence and evolution of the device itself. Following that, the properties of the device that make it attractive for memory are described, along with memory architectures for which it is suitable. We next review the evolution of magnetic tunnel junction MRAM demonstrations from first bits to the 16-Mb chip level, which is the highest MRAM capacity level demonstrated to date. We then compare MRAM to other potential embedded technologies and finally conclude with an outlook for the future development of MRAM. Selected aspects of these developments are described in greater detail in other papers in this issue of the *IBM Journal of Research and Development*.

Evolution of the magnetic tunnel junction

Magnetic tunnel junctions owe their properties to two physical phenomena that were established in the early 1960s and 1970s by work at low temperatures. Giaever at GE Research demonstrated the phenomenon of electron tunneling through thin insulators in his Nobel-prize-winning studies, reported in 1960, of tunneling between superconductors and normal metals [8]. Ten years later, Tedrow and Meservey at MIT carried out experiments on tunneling between ferromagnetic metals and superconductors which demonstrated that spin was conserved in tunneling and that the tunneling conductance depended on the degree of the spin polarization in the magnetic electrodes [9]. (For a review, see [10].) Shortly thereafter, in 1975, Julliere

demonstrated, for the first time, tunneling between two ferromagnetic films—in an Fe–Ge–Co tunnel junction [11]. The 14% effect he observed at low voltages and helium temperature is reproduced in **Figure 1(a)**, which shows the difference in conductance G between the anti-parallel and parallel alignment states of the Fe and Co films as a function of junction voltage V . Julliere analyzed his results in terms of spin conservation during tunneling and compared his 14% effect to tunneling, which he expected to be 26% on the basis of the tunneling experiments by Tedrow and Meservey. The rapid falloff with a few millivolts of applied voltage was attributed to spin-flip scattering.

At just about the same time as Julliere was doing this first demonstration of the magnetic tunnel junction, Slonczewski of IBM Research proposed three types of magnetic sensors based upon tunneling between ferromagnetic metals (later published as [12–14]). Slonczewski estimated that the tunneling magnetoresistance (TMR) values should be of the order of 40% for devices fabricated with common ferromagnetic metals. Moreover, since nothing was particularly temperature-dependent in tunneling, this value should have been obtainable at higher temperatures.

The field of tunneling between ferromagnets developed very slowly after that, primarily because of fabrication difficulties. Tedrow and Meservey continued their superconductor–ferromagnet tunneling studies, but these usually involved challenging efforts to find ways of growing “natural” tunnel barriers on different ferromagnetic metals. Some attempts by others to perform similar experiments with superconductors and ferromagnetic metals were not successful.¹ Slonczewski attempted to stimulate more work on ferromagnetic/insulator/ferromagnetic tunneling, but progress was slow. In the early 1980s, for instance, Maekawa and Gafvert, two visiting scientists at IBM Research, studied the effect in Ni–NiO–Co tunnel junctions [15], but the largest effect they could observe was 2% at 4.2 K. Their results, plotted in terms of tunneling magnetoresistance, are reproduced in **Figure 1(b)**. By the late 1980s, Slonczewski had analyzed the effect of an abrupt change in potential at the tunnel barrier interface [16] in an attempt to explain why the tunneling magnetoresistance effect was much smaller than he had earlier predicted.

In the intervening years, from the early 1970s to the mid-1980s, vacuum and tunnel junction technology advanced, particularly for Josephson tunnel junctions. IBM Research, Bell Laboratories, and others pioneered the fabrication of digital circuits based on the use of such junctions, and ultimately IBM pursued an associated

¹ S. von Molnar, Florida State University, Tallahassee (private communication while he was a Research Staff Member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1978).

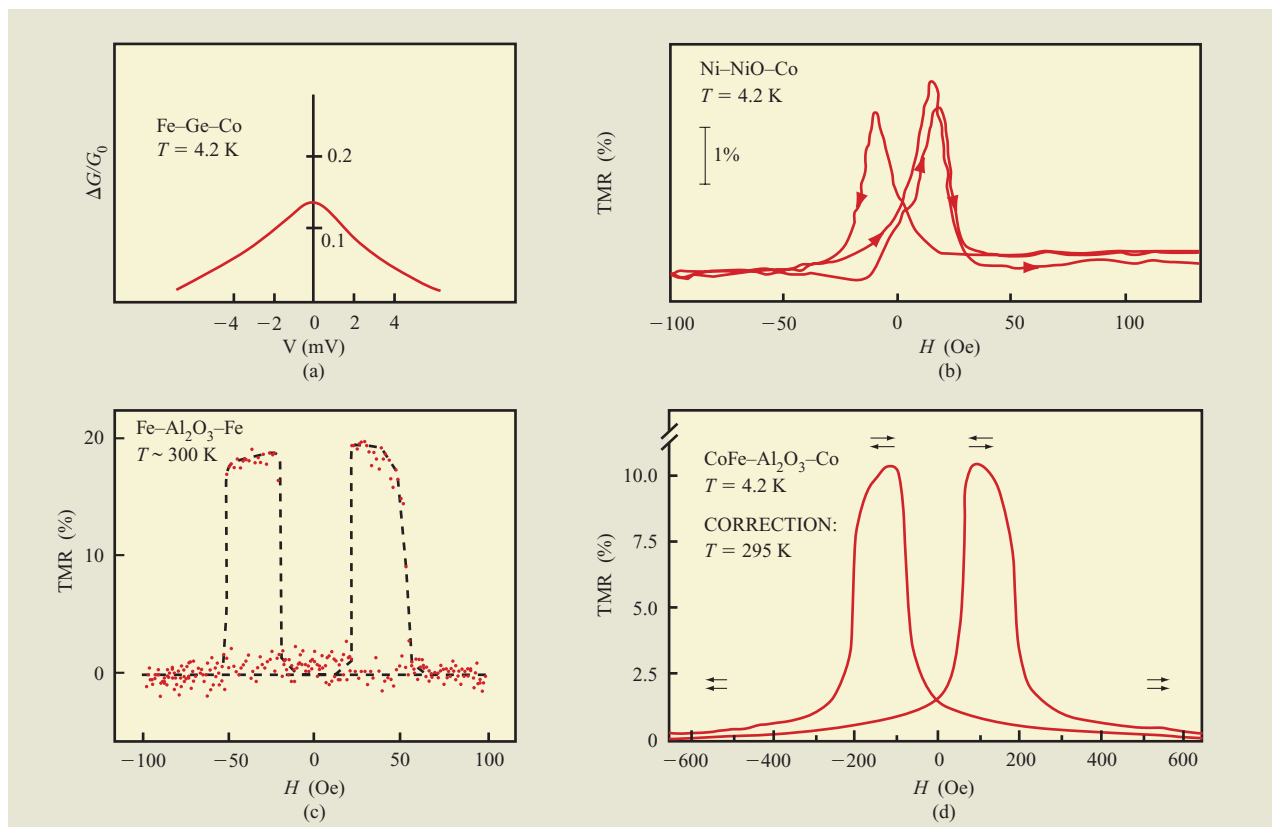


Figure 1

Early demonstrations of tunneling between ferromagnetic metals. (a) First demonstration of such tunneling — by Julliere in 1975 of tunneling at 4.2 K in an Fe-Ge-Co junction. Adapted from [11]; ©1975, with permission from Elsevier. (b) Demonstration by Maekawa and Gafvert in 1982 of tunneling at 4.2 K in a Ni-NiO-Co tunnel junction. Adapted from [15], with permission; ©1982 IEEE. (c) Demonstration by Miyazaki and Tezuka in 1995 of tunneling at 300 K in an Fe-Al₂O₃-Fe junction. Adapted from [26], with permission. (d) Demonstration by Moodera et al. in 1995 of tunneling at 295 K in a CoFe-Al₂O₃-Co junction. Adapted from [27], with permission; ©1995 American Physical Society.

processor cross section [17–19]. The junctions were fabricated primarily by oxidizing patterned Pb-alloy and Nb superconductors, and early difficulties in cycling from room temperature to liquid helium temperature were overcome. In the early 1980s, Rowell, Gurvitch, and Geerk at Bell Laboratories discovered that very thin layers of Al would wet Nb surfaces and enable very-high-quality “artificial” Al₂O₃ tunnel barriers to be grown on Nb base electrodes [20, 21]. A 20-to-40-Å-thick Al layer was sputter-deposited and oxidized to produce a high-quality amorphous Al₂O₃ layer which served as a tunnel barrier above a residual thin Al layer. Success with the artificial tunnel barrier relied upon the superconducting proximity effect, in which superconductivity extends over a relatively long “coherence length” distance into suitable adjacent normal metals, of which Al was an excellent example, having a coherence length of the order of 1 μm.

This meant that the residual layer of Al left after the oxidation would not degrade device properties. A tunnel junction was produced over a full wafer and subsequently subtractively patterned. This “full-wafer” tunnel junction fabrication technique was robust and has been widely emulated around the world for fabricating Josephson tunneling circuits up to the present time (see for example [22].)

Attempts to apply this artificial tunnel barrier technique to magnetic tunnel junctions, though initially slow to succeed, eventually resulted in demonstrations at room temperature of an effect of the size earlier expected. First, in 1991, Miyazaki and co-workers at Tohoku University reported NiFe/Al-Al₂O₃/Co junctions with TMRs of 2.7% at room temperature [23, 24]. This was a significantly higher percentage than earlier room-temperature results, but far lower than expectations.

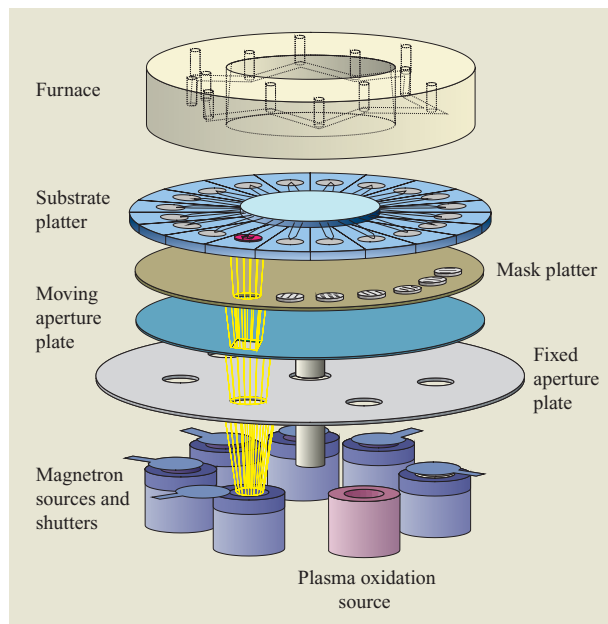


Figure 2

Second-generation IBM Almaden multilayer sputter-deposition system, incorporating plasma oxidation source and metal masking capability and used for engineering of magnetic tunnel junction structures.

Nowak and Rauluszhkiewicz [25] tried oxidizing evaporated Al for artificial barriers at about the same time, but without success. Finally, in 1994 two groups working independently, both using artificial tunnel barriers of Al_2O_3 , obtained magnetoresistance values at room temperature that began to approach the magnitude expected. Miyazaki and Tezuka, experimenting with $\text{Fe-Al}_2\text{O}_3\text{-Fe}$ structures, obtained an 18% effect at room temperature [26], while Moodera et al. at MIT, working with $\text{CoFe-Al}_2\text{O}_3\text{-Co}$, produced effects of up to 11.8% at room temperature [27]. As indicated in **Figures 1(c)** and **1(d)**, peaks were found for the tunneling magnetoresistance of the indicated size as the magnetic field is swept from zero field past the coercive field of one magnetic layer, remaining below the coercive field of the other magnetic layer.

IBM was in a unique position to follow up on these results. In particular, one of us (S. S. P. P.), had pioneered the development of multilayer sputtered films [28] for fabricating magnetic devices [5] and was just beginning to make use of a second-generation exploratory sputter-deposition system for multilayer magnetic materials (**Figure 2**) that was readily adaptable for magnetic tunnel junctions. The other author (W. J. G.) was actively engaged in tunnel junction device and microfabrication technology development using superconducting materials

[22, 29, 30] that could be readily adapted for fabricating magnetic tunnel junctions. The second-generation multilayer magnetic film system was designed to be applied for studies of perpendicular transport through multilayer metallic “spin-valve” [4] structures. The system was designed for the deposition of layers onto as many as 20 substrates through a series of metallic shadow masks via computer-controlled placement of the substrates and masks. The perpendicular transport studies would have been challenging, requiring either sensitive low-voltage measurements using superconducting sensors to very low voltages through large metal-mask-defined structures [31] or the production of extremely fine-dimensional structures defined by e-beam lithography to increase the resistance of the perpendicular transport [32]. However, by replacing one of the deposition sources with a plasma oxidation source, it was possible to use this system to fabricate magnetic tunnel junctions. Moreover, the specific resistance (resistance–area product) of magnetic tunnel junctions is orders of magnitude higher than that of all-metal spin-valve structures; hence, intricate low-voltage characterization techniques would not be required.

These growth and fabrication capabilities enabled many advances in magnetic tunnel junction structures. TMR values were increased, first to $\sim 25\%$ [33] and then to more than 40% [34]. Device resistance was shown to scale with area over many orders of magnitude, and deep-submicron devices ($\sim 125 \text{ nm} \times 250 \text{ nm}$) with high TMR values were demonstrated [33, 35]. The resistance–area product was shown to be tunable from at least $60 \Omega\text{-}\mu\text{m}^2$ to more than $10^7 \Omega\text{-}\mu\text{m}^2$ by varying the thickness of the Al_2O_3 tunnel barrier [36].

Of perhaps even greater significance than these achievements of higher TMR values, exchange biasing materials and techniques developed for spin-valve materials (see [34] for a review) were shown to be applicable for engineering the response of magnetic tunnel junctions [28]. **Figure 3** illustrates structures used to engineer the response of magnetic tunnel junctions in ways beneficial for memory applications. **Figure 3(a)** shows the basic magnetic tunnel junction structure, such as that used for the early studies, with specific characteristics illustrated in **Figures 1(b)** to **1(d)**. While in principle this structure could be made to work for a memory if the coercivity of one of the layers, a “reference” layer, were much higher than that of the other, difficulties would arise with this approach. First, field excursion would have to be restricted to being lower than a maximum value so that the high-coercivity layer would never be disturbed. Even so, it is possible that repeated low-field excursions could reverse small domains in the higher-coercivity reference layer that have no way of returning to their original state [37]. The possibility of

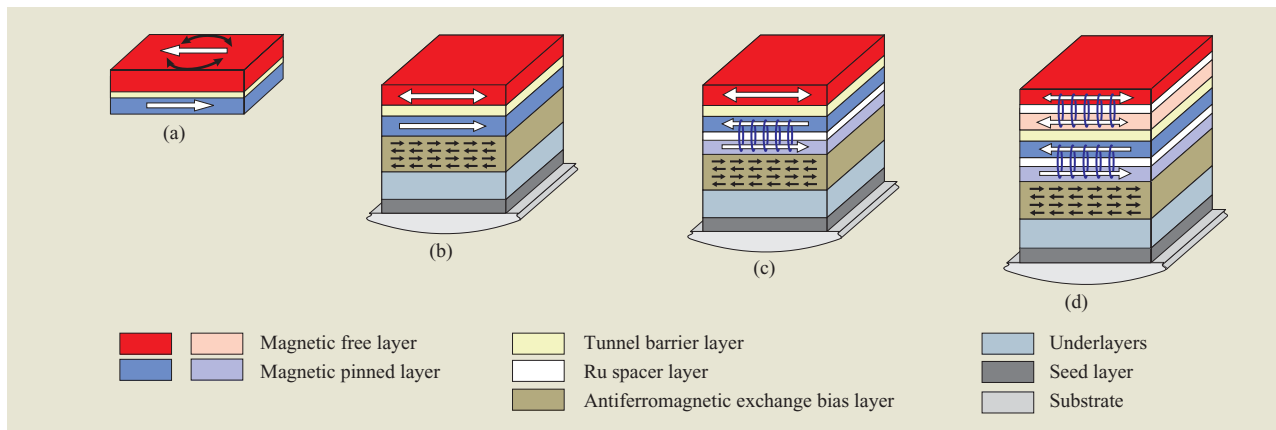


Figure 3

Evolution of tunnel junctions engineered for MRAM applications. (a) Basic magnetic tunnel junction structure consisting of two ferromagnetic metals separated by a thin insulating layer. With the same anisotropy direction for both magnetic film layers, the junction has a hysteretic TMR response characteristic like that shown in Figure 1(b, c, d). (b) By exchange-coupling one of the magnetic layers to an antiferromagnetic layer — i.e., by “pinning” the layer — the TMR response reflects the hysteresis of the other so-called “free” layer and has a response curve more suitable for memory (as shown in each of the parts of the next figure). (c) The magnetic offset caused by fields emanating from the pinned layer can be avoided by replacing a simple pinned layer with a synthetic antiferromagnetic pinned layer, which consists of a pair of ferromagnetic layers antiferromagnetically coupled through a ruthenium (Ru) spacer layer. The lower layer in this artificial antiferromagnet is pinned via exchange bias, as shown in (b). This flux closure increases the magnetic stability of the pinned layer and reduces coupling to the free layer. (d) Structure in which both the pinned and free elements consist of antiferromagnetically coupled pairs. Adapted from [28], with permission.

upsetting the reference layer could be avoided by pinning one of the magnetic electrodes via exchange-coupling to an adjacent antiferromagnet, as illustrated in Figure 3(b). Only the other “free layer” electrode responds to the field. The low-field electrical response of such a structure would very directly reflect the memory function of the magnetic hysteresis of the free layer [33]. In subsequent structures, such as the one illustrated in Figure 3(c), the antiferromagnetic material was replaced by a synthetic antiferromagnet (SAF) sandwich comprising, for example, CoFe/Ru/CoFe, with the Ru thickness being 7–8 Å [38]. In this thickness range, the Ru exchange-couples the moments of the two ferromagnetic layers in opposite directions. Thus, a “fixed” reference layer, the SAF, could be produced with no net magnetic bias on the other, “free,” magnetic layers, which is important if a suitable response is to be obtained for a magnetic memory in the absence of a magnetic-field bias.

The resistance vs. magnetic field characteristics for a small array of devices with pinned SAF layers are shown in **Figure 4(a)**. In this case it can be seen that the control of resistance spread σ_R is adequate to separately identify the direction in which the free-layer moment is pointing from reading the zero field resistance (basic memory functionality). Also, because of the balanced

SAF structure, the positive and negative write fields are similar in magnitude, with margin for overdriving the field. This is in contrast to Figures 1(b–d), in which the two magnetic electrodes of the tunnel junction have reversal fields of similar magnitudes.

Early “simple pinned” structures such as that in Figure 3(a) employed FeMn as the antiferromagnetic material. This material is antiferromagnetic if it is grown in a bias field on a magnetic seed layer, but it is not very stable during subsequent annealing. For this reason, the use of other antiferromagnetic materials, under development for magnetic disk read heads, was investigated. As a result, the FeMn was replaced by IrMn, resulting in thermal stability up to ~230°C to 300°C [36, 39]. Later the thermal stability was further advanced with the use of PtMn as the antiferromagnetic pinning material. With PtMn, a relatively high TMR was achievable even after several hours of annealing at 400°C for devices in which use was made of a 200-Å-thick CoFe free-layer electrode [40].

Advances in Al₂O₃-based MTJs also occurred with changes in the material used for the free-layer electrode. Kano et al. introduced the use of amorphous CoFeB for that purpose and showed that this made it possible to achieve a TMR of ~60% [41]. One of us (S. S. P. P.) showed that the use of a bilayer consisting of CoFe with

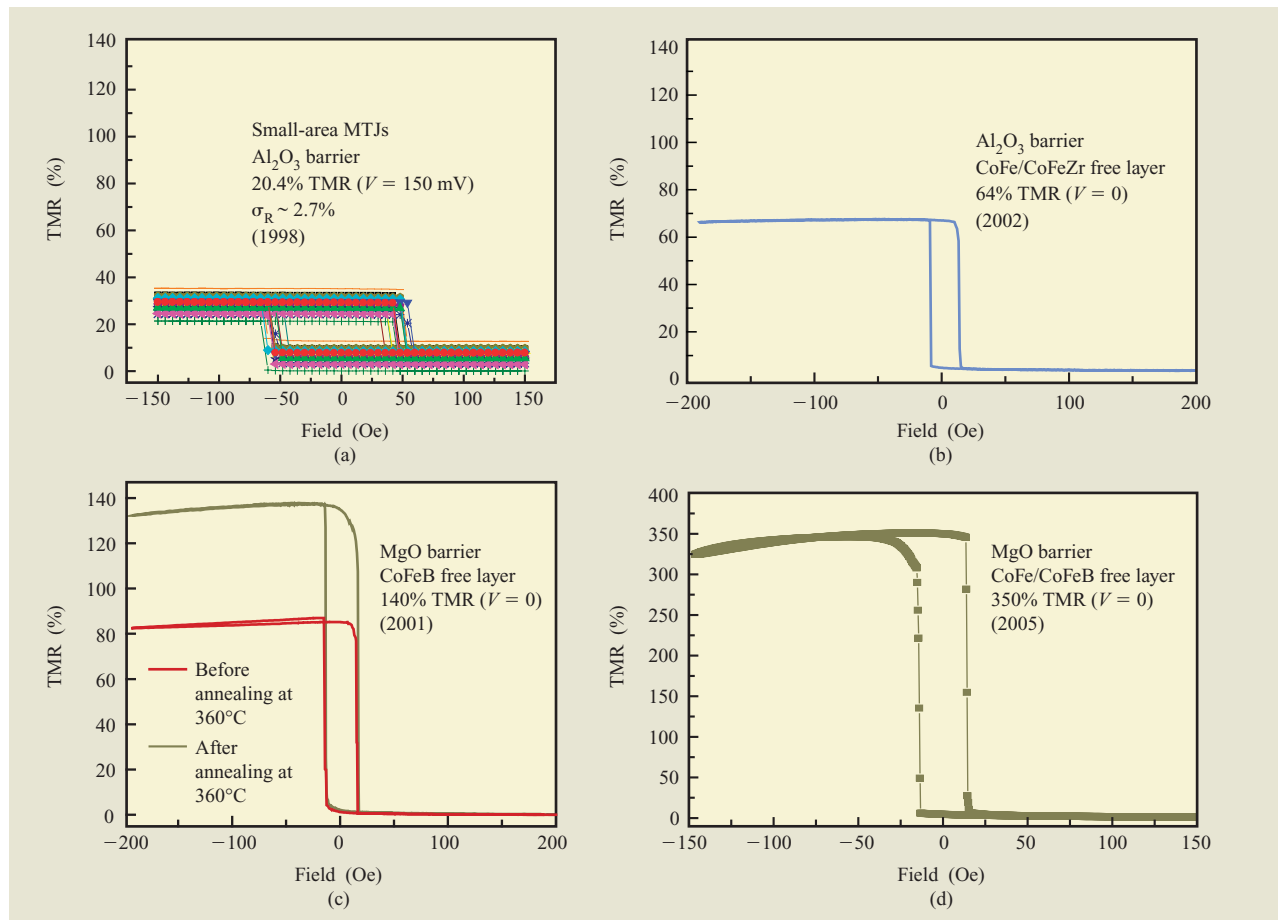


Figure 4

Results of materials advances on TMR vs. field characteristics (a) for an array of 24 small-area MTJs with exchange-biased synthetic antiferromagnetic pinned layers, (b) for a large-area MTJ with an amorphous free layer containing CoFeZr, (c) for a large-area MTJ with an MgO barrier and a CoFeB free layer before and after annealing, and (d) for such an MTJ following optimization and after annealing. Part (c) from [46], with permission; ©2004 Nature Publishing Group. The dates shown are the dates at which results were obtained. All characteristics were obtained at room temperature except for that shown in (d), which was obtained at 290 K.

amorphous CoFeZr could achieve a TMR of $\sim 65\%$, as shown in **Figure 4(b)**.² Later, magnetic tunnel junctions with Al_2O_3 barriers and CoFeB free-layer electrodes were further optimized by others to produce TMR values of $\sim 70\%$ [42, 43]. At the time of this writing, those values remain the highest reported for room-temperature magnetic tunnel junctions with Al_2O_3 barriers.

Meanwhile, the same author (S. S. P. P.) and his collaborators had been investigating predictions of much higher TMR values that had been made independently by Butler et al. [44] and by Mathon and Umerski [45] for epitaxial [001] Fe/MgO/Fe tunnel junctions. Beginning in mid-2001, as indicated in **Figure 5**, much higher values of TMR began to be observed. **Figure 4(c)** reproduces one

of the subsequently published TMR characteristics for an epitaxial tunnel junction with the nominal structure of $100\text{TaN}/250\text{IrMn}/8\text{Co}_{84}\text{Fe}_{16}/30\text{Co}_{70}\text{Fe}_{30}/31\text{MgO}/150\text{Co}_{84}\text{Fe}_{16}/125\text{TaN}$, where the numbers before alloy composition indicate the layer thicknesses in Å [46]. Companion spin-polarized tunneling studies of similar MgO tunnel junctions having one magnetic electrode and one superconducting Al(Cu) electrode indicated very high degrees of spin-polarized transport for structures with both Fe and CoFe base electrodes, with higher values achieved for CoFe electrodes [46]. To obtain such values successfully, the growth technique that had worked so well for Al- Al_2O_3 had to be modified. For both Al_2O_3 and MgO barriers, it was generally advantageous to deposit a very thin metallic wetting layer on the

² S. S. P. Parkin, unpublished results, September 2002.

Table 1 Magnetic tunnel junction properties for read operation.

High magnetoresistance (MR)	advantage
High resistance	advantage
Controllable resistance	advantage
Weak temperature dependence	advantage
Scalable to small sizes with high MR	advantage
MR falls off with increasing voltage	disadvantage

base electrode layer. However, high-quality MgO tunnel barrier junctions were not produced when a metallic Mg layer was oxidized; high quality was achieved when the MgO tunnel barrier was subsequently reactively deposited on Mg. It is believed that the reason for this is that metallic Mg contracts by about 20% upon oxidation, in contrast to metallic Al, which *expands* approximately 27% during oxidation. The contraction of Mg upon oxidation very likely results in the formation of pinholes in grain-boundary regions. In addition, as indicated by the before and after annealing curves in Figure 4(c), TMR increased substantially upon annealing up to temperatures in the range 360–380°C, most likely reflecting an increase in epitaxial junction quality. A further increase was achieved by process optimization, as indicated in **Figure 4(d)**.³ A similar structure but with a CoFeB top electrode, with TMR characterized by the current-in-plane-tunneling technique [47, 48], gave a TMR value of 220% ± 10% after annealing to 350°C [46]. This device had the nominal structure 100TaN/150IrMn/35Co₇₀Fe₃₀/20MgO/75(Co₇₀Fe₃₀)₈₀B₂₀/75Ru, possibly with some crystallization occurring in the CoFeB, at least near the MgO interface.

As indicated in Figure 5, public awareness that high TMR is obtainable with MgO tunnel barriers came about three years after the unpublished IBM results by Parkin et al. Published demonstrations of epitaxial structures with MgO initially gave TMR values substantially less than the maximum range obtained with Al₂O₃ barriers. Eventually the use of MBE techniques resulted in values in the range of 67% to 88% [49, 50] and ultimately 188% [51]. Very significantly, because they were using a Canon ANELVA commercial 200-mm production sputter-deposition system, Djayaprawira et al. [52] then obtained a TMR value of 230% for MgO-based tunnel junctions with nominal structures containing PtMn/CoFeB/MgO/CoFeB layers. Recently this group reported an even higher TMR value of 268% [53], and one of us (S. S. P. P.) and his collaborators have recently obtained a TMR

³ S. S. P. Parkin, unpublished results, July 2005.

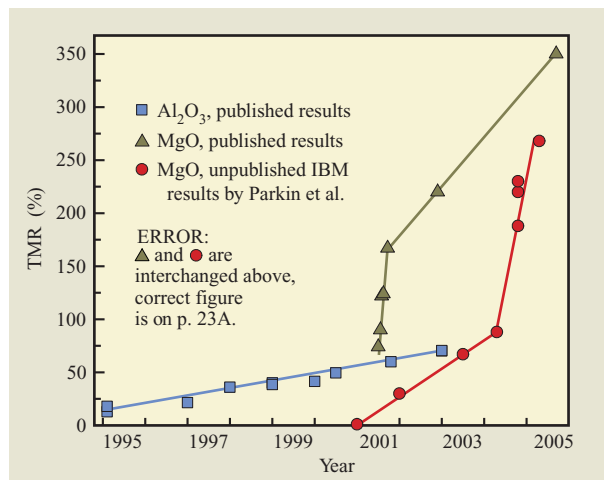


Figure 5

Maximum demonstrated TMR values over time for magnetic tunnel junctions with Al₂O₃ and with MgO tunnel barriers.

value of 350% [see Figure 4(d)]. At the time of this writing, to the best knowledge of the authors, that value is the highest yet reported. Finally, to provide some perspective, it is worth noting that the theoretical models [44, 45] do not appear to place a limit on the TMR, so perhaps higher values yet will soon be reported, although the models are based on the assumptions of perfect crystalline lattices, epitaxy, and stoichiometry.

From magnetic tunnel junctions to early MTJ MRAM demonstrations

Following the initial demonstration of high TMR at room temperature, it was quickly recognized at IBM that the magnetic tunnel junction had a number of properties that would make it very attractive for the read operation in a very dense memory cell (see **Table 1**). The high magnetoresistance is of course important for obtaining a large signal from a memory cell, especially if a fast read operation is desired. Equally important, and perhaps not as readily recognized, the high resistance of a magnetic tunnel junction, compared with that of earlier all-metal magnetic devices, is compatible with high-speed sensing using CMOS circuitry. The high resistance is introduced in a very compact structure, basically in the space of an electrical via connection between two wiring layers. The resistance depends exponentially on the thickness of the very thin tunnel barrier, of the order of 10 to 20 Å. The extremely small thicknesses involved might appear to make the resistance difficult to control, but on the basis of earlier experiences at IBM with Josephson technology, it was confidently assumed that it would be possible to

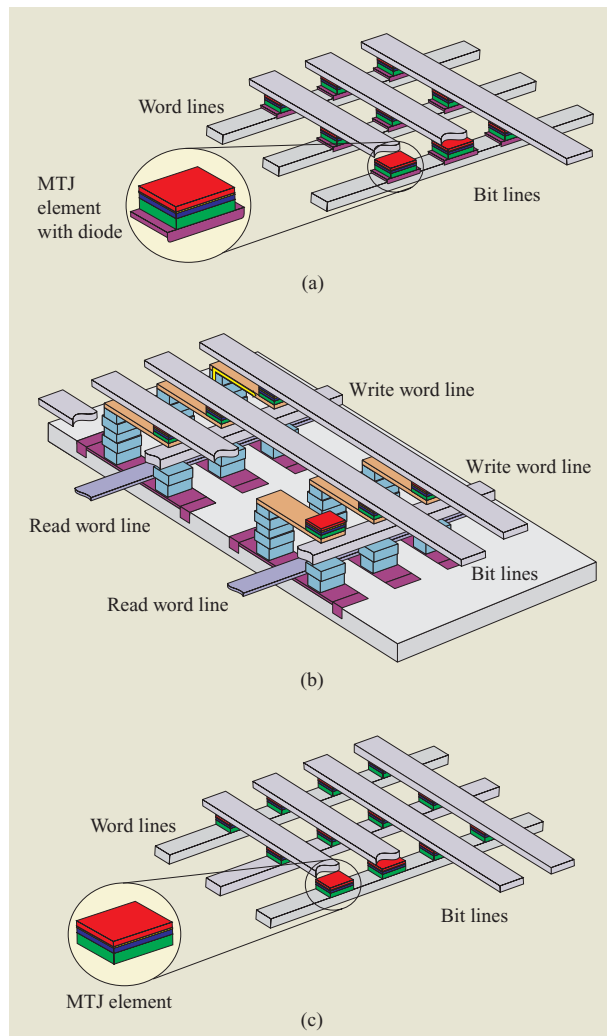


Figure 6

Alternate MRAM architectures. (a) Cross-point architecture with thin-film diode in the tunnel junction stack. Adapted from [54], with permission; ©1998 IEEE. (b) FET architecture; from [28], with permission. (c) Switchless cross-point architecture.

control the resistance. (What was not known at the time, however, was how reliable thin magnetic tunnel junctions could be if operated for long periods of time at room temperature.) Another essential attribute of magnetic tunnel junctions for read operations in a memory is that their electrical properties vary only slowly with temperature. The important scale for temperature variations is set by the Curie temperature of the magnetic electrodes, which is of the order of 500°C or greater in commonly used alloys of Ni, Fe, and Co. As discussed above and important for future scaling, the magnetic tunnel junction could be scaled to very small sizes while its magnetoresistance value remained unchanged. The

sole negative aspect of the properties of the magnetic tunnel junction from a read perspective is that the magnitude of the magnetoresistance effect falls off with increasing voltage. While Julliere's first demonstration [Figure 1(a)] showed this to be a very rapid falloff in a few tens of millivolts [11], in 1995 Moodera showed (along with his initial high TMR results) that the voltage at which the TMR falls to half of its maximum value could be 200 mV [27]. This range of falloff is slow enough to result in a sufficiently large signal for sensing by memory readout circuitry; with improvements in MTJ quality, typical half voltages currently range from 500 mV to more than 700 mV.

MRAM architectures enabled by the magnetic tunnel junction

As first discussed by Scheuerlein [54], the magnetic tunnel junction device, with its large magnetoresistance signal, its CMOS-compatible resistance value, and its compact structure, has led to the consideration of several MRAM architecture possibilities. The small, low sheet resistance available from early anisotropic magnetoresistance (AMR) and spin-valve devices pursued for MRAMs meant that the magnetic devices had to be laid out as relatively long-aspect-ratio structures in two dimensions, making impossible the achievement of small cells. The small signal also required that the magnetic storage state of those devices be disturbed in some manner (destructively or nondestructively) during the read operation in order to provide a self-reference signal. The small signal and associated complex referencing and sensing operations inevitably result in a slow read operation. The TMR device, on the other hand, could be laid out compactly and placed in series with a switch element (either an FET transistor or a diode) and sensed directly, assuming that tracking variations of the resistance value between different TMR elements are smaller than the magnetoresistance ratio.

Three architectures that should be feasible because of the large signal of the TMR element are illustrated in **Figure 6**. Figure 6(a) shows a particularly compact "cross-point" arrangement in which each magnetic tunnel junction stack also contains a thin-film diode. The diode serves to block the sneak current paths in the matrix arrangement of the cells. During a read operation, a cell is selected by grounding one word line while all of the other word lines are biased as high as the sense line. Then, just one device along a bit line will be forward-biased, and the current that flows through it will be detected. As discussed in detail by Scheuerlein, the requirements for high-speed operation of this type of array are 1) that the diode carry a relatively high forward current so that the voltage drop across it when forward biased is less than that across the magnetic tunnel junction and 2) that the

diode have a high rectification ratio in order to limit sneak paths in large arrays. For the compact cross-point structure, the diode must be formed above a wire that can carry substantial current (several milliamps). Unfortunately, the forward conductance of thin-film diodes that can be formed on high-conductivity metal wires is insufficient for high-speed cells with small areas. It has thus not been possible to realize this very attractive architecture. A second architecture, shown in Figure 6(b), utilizes an FET switch in the substrate to eliminate the sneak paths during the read operation. The cell area is roughly doubled because it is necessary to connect to the FET switch in the substrate through a via to the side of the MTJ stack. However, fast operation is obtained, since there is then generally enough silicon area in the larger cell to implement a high-conductance FET switch. In principle, a second type of cross-point architecture, shown in Figure 6(c), is also possible with the difficult-to-implement thin-film diode eliminated [55–57]. In this case, to avoid sneak paths during write operations, the resistance value must be large. Then, during the read operation, all lines in the entire array are carefully biased so that current flows preferentially through just one device. The signal is weak because of the large resistance and sneak current paths, so sensing is slow, but the very high density of the cross-point arrangement is maintained. Moreover, since the substrate is not involved in the cells, it is conceptually possible to stack layers of cells.

In all cases, the MRAM write operation is done by a coincidence of x - and y -currents, and these are easiest to control if the lines supplying these are isolated, as for the architectures of Figures 6(a) and 6(b). In conventional MRAM, the write operation makes use of the stability boundary curve shown in part (b) of **Figure 7**. The boundary astroid curve illustrated is that calculated for a single-domain model of the switching free layer, and is expressed as

$$H_x^{2/3} + H_y^{2/3} = H_i^{2/3},$$

where H_x and H_y are respectively the x and y fields, and H_i is the anisotropy field of the assumed single-domain magnetic element. This functional form is that of a mathematical curve called an *astroid*, but in the MRAM field the term has generally taken on the meaning of the experimentally obtained switching boundary, which in the best cases only approximates a true mathematical astroid curve. The magnetic element of Figure 7 is assumed to have a left–right uniaxial anisotropy, which usually arises from an intrinsic anisotropy established during film growth, annealing in a bias magnetic field, and shape anisotropy of the storage layer of the element. For field excursions that remain inside the astroid curve,

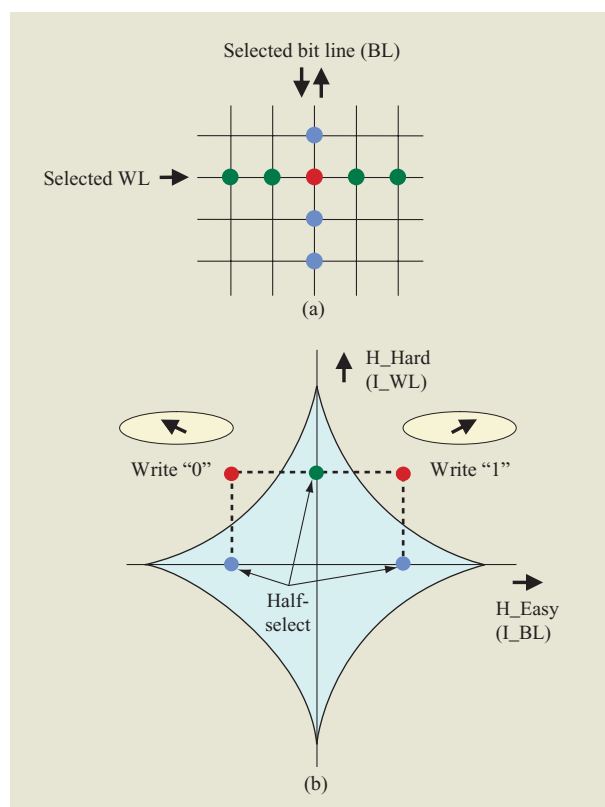


Figure 7

Coincident field selection for writing a magnetic element of an MRAM. (a) Schematic of array with colored dots indicating the selected (red) and half-selected (blue and green) bits. (b) Switching threshold (astroid) curve with appropriate field values indicated for the write operation and with magnetization orientation fields. Part (b) adapted from [58], with permission; ©2005 IEEE. The ellipses show the direction of the shape anisotropy of the free magnetic layers of the bits; the arrows inside the ellipses indicate the orientations of the free-layer magnetization.

the element is magnetically stable, pointing either to the left or to the right. For field excursions that go outside the astroid curve, the element is written to one definite state.

Many of the largest challenges of MRAM are associated with controlling the write operation, particularly with ensuring that none of the write operations cause an undesired switching event in any of the half-selected bits [see part (a) of Figure 7]. Without a carefully balanced pinned layer structure, such as that of Figure 3(d), the astroids can be offset to the right or left. In addition, the element shapes are not perfectly uniform in size, causing the astroid sizes and shapes to vary. Finally, it turns out that the energy barrier that separates the two stable states vanishes as the write field approaches the astroid boundary. Thus, there is a non-negligible

probability of the magnetization spontaneously switching by a thermally activated process as this occurs [59]. A more advanced “toggle” MRAM write architecture was later introduced by Motorola for a bilayer storage film, as depicted in Figure 3(d), which considerably lessens the write-control challenges [60]. This architecture is briefly introduced later on in this paper, and is analyzed in detail in another paper in this issue [61].

First magnetic tunnel junction MRAM demonstrations and early MRAM technology development

While several of us at IBM quickly recognized the potential of the MTJ device as the basis for an attractive and competitive RAM memory technology, the serious development of such a technology was also greatly accelerated by the serendipity of interest by the U.S. Department of Defense Advanced Research Projects Agency (DARPA) in establishing a major program to foster RAM development based on magnetic devices. The paper by Wolf et al. in this issue reviews the DARPA Spintronics program [62], under which IBM received partial support for its MTJ device and MRAM efforts from approximately mid-1997 until the end of 2001, thus being able to undertake a more rapid development program than would otherwise have been possible. While materials and device developments came quickly and several array architectures looked attractive, it took approximately five years from the time of the TMR breakthrough to demonstrate working MTJ MRAM arrays [63]. The demonstration of exchange-biased magnetic tunnel junctions [Figure 3(c)], particularly with a synthetic antiferromagnetic pinned layer [Figure 3(d)], showed how a switching characteristic suitable for an MRAM could be obtained. The antiferromagnetic material for the exchange bias evolved from FeMn in the early days to IrMn, which showed improved thermal stability, although processing temperatures continued to be restricted to about 100 degrees below the 350–400°C range which is typical for fabricating CMOS “back-end-of-line” (BEOL) wiring layers.

A major stumbling block was the limited availability of processing equipment suitable for CMOS integration. In 1995 leading-edge CMOS development and manufacturing was on 200-mm-diameter silicon substrates, and it was clear that this would be moving to 300-mm-diameter substrates before long. In principle, the deposition systems used for MTJ read heads could be adapted for MRAM, but there was no standardization of substrate size in the head industry, and that industry had limited motivation to go to the 200-mm, much less the 300-mm level. (The smaller volume of material that must be manufactured each year for read heads compared with semiconductor material can be estimated by comparing

the area of the magnetic read heads in a typical computer to the area of all of the silicon chips—a difference of two to three orders of magnitude.)

With virtually all of the IBM semiconductor capability in the mid-1990s on 200-mm-diameter substrates, and with the leading edge expected to move to 300 mm in the early 2000s, it was decided that a first demonstration would be done in a hybrid mode, by cutting up pieces of 200-mm-diameter silicon wafers and finishing them with a few layers of special processing in research systems in order to add the MRAM devices. The IBM technology CMOS 6sf was chosen for the demonstration, since it was relatively advanced and yet mature enough so that most of the significant new processing efforts could be focused on the magnetic elements. CMOS 6sf featured 0.25- μm front-end ground rules and Al(Cu) wiring with 0.4- μm ground rules for the second level of metallization and above. Two special adjustments were made to the normal 200-mm process flow. First, dopant implantations were adjusted so that diodes could be implanted in the silicon substrates in order to enable simulation of the diode cross-point architecture by using a via to the side cell and diodes in the substrates. (It turned out that the FET cells performed better than the diode cells, so the substrate diode cells were not pursued for long.) Second, the interlayer dielectric between the second and third metallization levels was specially thinned by a CMP operation to enable the MTJ devices to be located closer to the underlying wiring layer. BEOL processing was stopped after this special thinning, the substrates were diced into one-inch squares, and then the MRAM stack was deposited utilizing the system depicted in Figure 2. Following the “setting anneal” for the antiferromagnetic layer, all subsequent processes could be carried out at ambient temperatures. A practical challenge was achieving acceptable lithography, compatible with 0.25- μm technology, on the one-inch diced squares. In the beginning, e-beam lithography was used for all four levels because of its greater flexibility for focusing and alignment. Eventually this was followed by the use of modified optical stepper lithography.

Figure 8(a) shows a scanning electron microscope cross section of a portion of the demonstration array—a 1-Kb CMOS MRAM twin-cell array—the first one fabricated within the IBM MRAM program [63]. Four main processing modules were required after dicing and MTJ stack deposition. First the local interconnects to the vias were subtractively patterned by lithography and ion milling. Then a photoresist stencil was applied to enable the patterning of the tunnel junction shapes by ion milling. The stencil also served as a self-aligned liftoff mask for an insulation layer that was deposited after the ion milling. Next, a via-hole pattern was defined in resist, and vias were etched in the periphery of the array to make

it possible to form contacts between the second and third levels of metallization (Al–Cu). Finally, a photoresist stencil for the third level of metallization was applied and patterned, and an Al–Cu film was deposited and then patterned by means of stencil liftoff. The size of a twin cell was $2\ \mu\text{m} \times 2.3\ \mu\text{m} = 4.6\ \mu\text{m}^2$. Each half cell occupied about 27 minimum-lithographic-area squares.

The performance of this first MRAM array matched high expectations. **Figure 8(b)** shows the read responses for a bit in the array. Nominal bit transitions from “address in” on chip pads to “data out” on other pads were 2.6 and 3.0 ns for the two transitions. That was much faster than any previous MRAM and comparable to or better than the performance of all memory types except for the fastest SRAMs. Although high bit yields were not expected from the demonstration, the results obtained highlighted areas in which further work would be needed.

180-nm ground-rule MRAM technology development within the MRAM Development Alliance

On the basis of the results cited above, which demonstrated the functionality and performance potential of MRAM, IBM teamed up with Infineon Technologies Corporation for a full 200-mm-wafer-scale technology development in 180-nm CMOS in a joint effort known as the MRAM Development Alliance (MDA). The mission of the MDA was to develop a competitive and scalable MRAM technology. The 200-mm-technology work took place primarily at two sites, the IBM Advanced Silicon Technology Center in East Fishkill, New York, where “standard” 200-mm tooling was located, and the IBM Thomas J. Watson Research Center in Yorktown Heights, New York, where specialized tooling for processing the 200-mm-wafer-scale magnetic tunnel junction layers was located. The MDA also undertook basic magnetic materials and device research, centered at the IBM Almaden Research Center in California, and had design and functional characterization activities at the IBM Essex Junction, Vermont, site. In a number of cases, IBM and Infineon worked with vendors to develop first-of-a-kind 200-mm MTJ processing equipment (e.g., 200-mm MTJ stack deposition and 200-mm magnetic annealing equipment) and quick-turnaround characterization equipment. Some of the unique characterization equipment and techniques developed are described in another paper in this issue [48].

The MDA work spanned a period of three and a half years from late 2000 to mid-2004. The two prime targets of the MDA development activity were the initial bringup of a complete 200-mm process for MRAM fabrication based on the IBM 180-nm-node CMOS 7sf technology and the development of a 16-Mb product

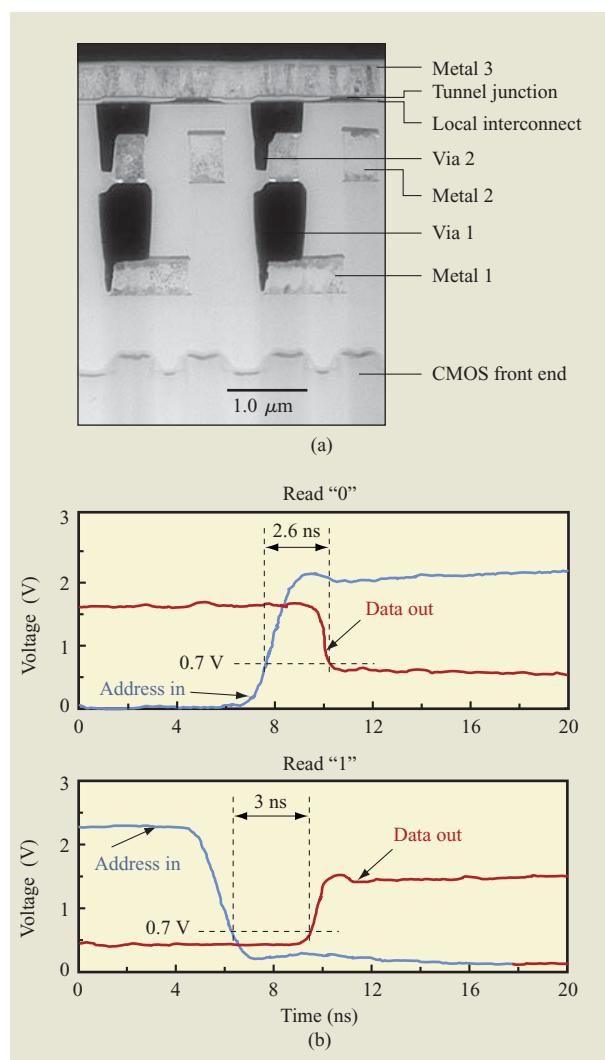


Figure 8

(a) SEM cross section of twin-cell structure of first CMOS MRAM array fabricated at IBM — a 1-Kb array containing FET twin cells fabricated in a 0.25-μm technology. (b) Time traces of input and output signals during read access operation. Both parts from [58], with permission; ©2005 IEEE.

demonstrator chip. The initial technology developed was described in a 2003 VLSI Technology Symposium paper by Sitaram et al. [64]. At the time, the development was based on the most advanced node used for fabricating an MRAM chip. The paper reported good control of MRAM device resistance distributions for read yield, sufficient write threshold control for a checkerboard write yield demonstration on a 2-Kb array, and testing-limited endurance of $>6 \times 10^8$ write cycles. Additionally, the technology was used to fabricate a 128-Kb core which was carefully characterized for performance [65] and

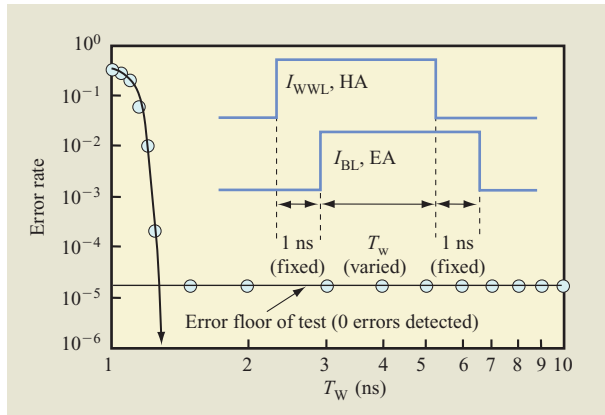


Figure 9

Write error rate vs. overlap time for 128-Kb core of 16-Mb product demonstrator chip. Inset shows the timing for the hard-axis (HA) and easy-axis (EA) pulses. Other measurements showed the error floor to be less than 10^{-8} . From [58], with permission; ©2005 IEEE.

served as the basis for the later product demonstrator design. **Figure 9** shows the write error rate as a function of the overlap time between x - and y -write pulses, which were staggered by 1 ns. It can be seen that by the time the overlap T_w was 1.5 ns (for a total write pulse time of 3.5 ns), the error rate reached the noise floor. Thus, writing was demonstrated to be as fast as the nominal 3-ns read speed reported above. Fundamentally (ignoring circuit limitations), writing much faster than this, at speeds approaching one nanosecond, would be expected to be influenced by ferromagnetic resonance phenomena occurring within the MRAM devices at frequencies of a few GHz. It is thus difficult to conceive of writing much faster than one nanosecond without explicitly harnessing the resonance phenomena, which, from a precise timing perspective, would be expected to be very challenging to implement in large arrays.

One of the final activities of the MDA was the transfer of the MRAM technology into a manufacturing site, the jointly IBM–Infineon-owned Altis Semiconductor Corporation in Corbeil-Essonnes, France. Following the end of the MRAM Development Alliance in mid-2004, Infineon continued its MRAM development activity at the Altis site. After the transfer to Altis, IBM refocused its MRAM effort into a more exploratory program comprising a basic yield demonstration component for 180-nm technology development and a significant exploratory materials and device component for demonstrating MRAM scalability at advanced CMOS process nodes. This is consistent with the desired IBM product vision for MRAM as an embedded memory

at an advanced CMOS node. Some of these efforts are described in this issue in the papers by Gaidis et al. on a two-mask-level fabrication route for quick process learning [66]; by Worledge on spin-flop switching in toggle MRAM; [61], by Sun on spin angular momentum transfer devices [67], and by Jiang et al. on tunnel spin injectors for semiconductors [68].

Based on the learning from the 180-nm process bringup and the 128-Kb test site, another of the final projects of the MDA was the design and fabrication of the 16-Mb demonstrator chip [69] mentioned above. **Figure 10** shows

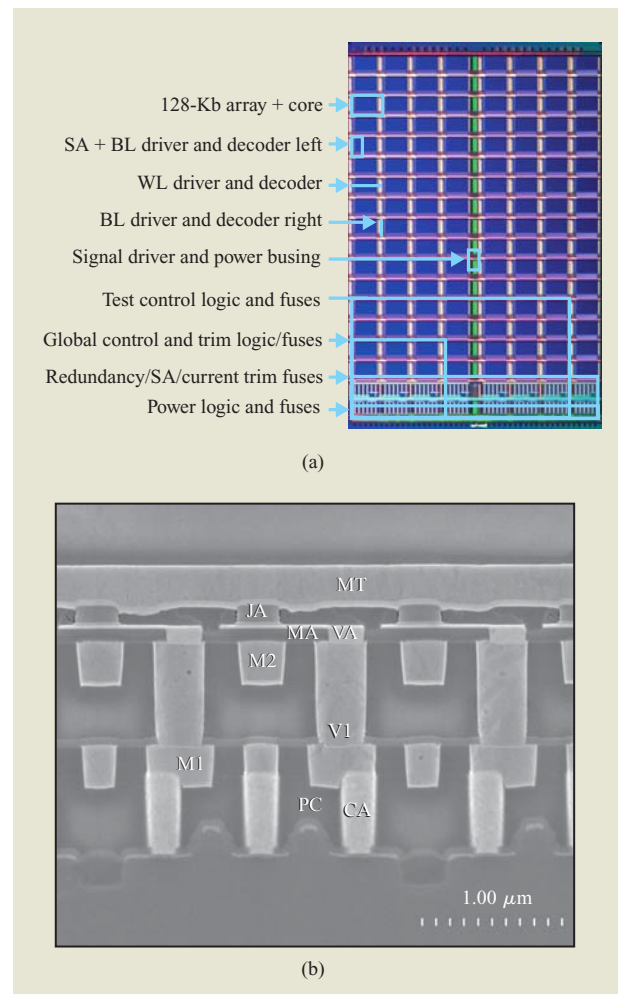


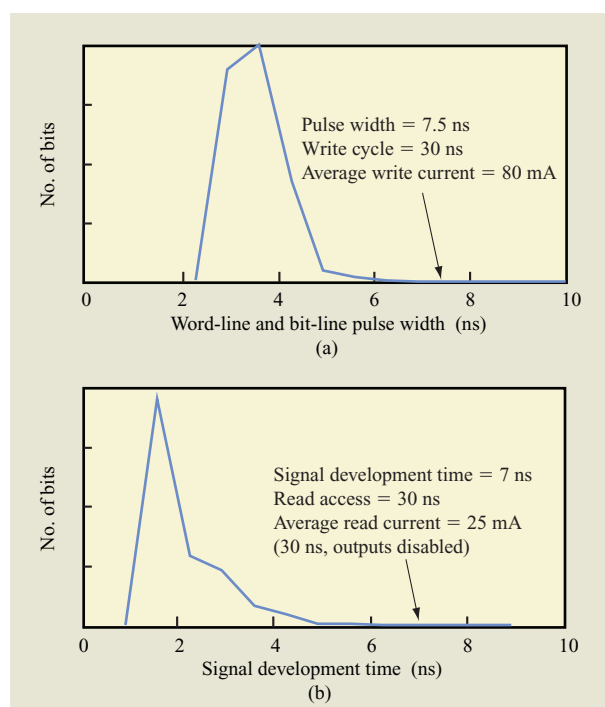
Figure 10

(a) Photograph of 16-Mb product demonstrator chip, with indications of main functional subunits. From [58], with permission; ©2005 IEEE. (b) SEM cross section showing three cells: M1, M2, and M3 are segments of three global wiring layers; VA, V1, and CA are segments of three via layers; PC is a segment of the transistor gate layer; JA is a segment of the tunnel-junction definition layer; and MA is a segment of the local wiring layer.

Table 2 Design parameters for 16-Mb MRAM.

Interface	16-Mb asynchronous SRAM (1 Mb $\times$ 16)
Technology	CMOS 7sf (180 nm), 1-poly, 3-Cu metal
MRAM specific layers	Three-level MRAM adder
Chip size	7.9 mm $\times$ 10 mm
Cell size	1.42 μm^2 (efficiency $\sim$ 30%)
Package	48 FBGA (fine ball grid array), 0.75-mm ball pitch
Package size	$\sim$ 10 mm $\times$ 12 mm
V_{dd}	2.3–3.3 V or 1.8 V
Access/cycle times	Read: 30 ns, Write: 30 ns
Active current (read)	25 mA @ 30 ns
Active current (write)	80 mA @ 30 ns
Standby current	32 μA
Deep power-down current	5 μA , no loss of data

a photograph of the 16-Mb chip and an SEM cross section of three cells. The chip was designed as a 1M $\times$ 16 SRAM with an asynchronous interface and had a 30-ns read and write performance. It consisted of two 8-Mb subunits, each fabricated from 64 cores of 128 Kb each. The core design was an optimized version of that in the earlier 128-Kb core demonstration described above. At the time of this publication, the MDA 16-Mb MRAM is the largest MRAM yet fabricated, with the smallest cell size (1.42 μm^2) among multi-megabit MRAM designs. The main features of the design are summarized in **Table 2**. It made possible a statistical demonstration of the read and write performance capability of the 180-nm MRAM technology. **Figure 11** shows histograms of the distribution of write (a) and read (b) times. In Figure 11(a) it can be seen that most of the bits of the 16-Mb chip are successfully written with a pulse width of less than 5 ns. A pulse width of 7.5 ns captures the entire histogram and corresponds to a write-cycle time of 30 ns. The histogram in Figure 11(b) shows the minimum signal development time required for a successful read. Signal development time is measured from the time the sense amplifier (SA) equalize is turned on until there is data at the SA output node. The peak of the distribution, corresponding to nominal bit performance, is at 2–3 ns, and a signal development time of 7 ns captures the distribution. The 7-ns time for the distribution is consistent with the budget for this operation within the 30-ns read access time of the chip design.

**Figure 11**

Performance verification of 16-Mb MRAM product demonstration chip. (a) Write performance: Histogram of minimum word-line and bit-line write-current pulse widths. (b) Read performance: Histogram of minimum signal development time (from sense amplifier equalize to sense amplifier output). Both parts from [58], with permission; ©2005 IEEE.

These results are intended to be illustrative of the potential performance of MRAM. The peaks of the distributions correspond reasonably well to the nominal performance expectations (above), and the distributions reflect the results after only a limited time for process optimization. Another paper in this issue, by Maffitt et al. [70], describes the design learning and tradeoffs in greater detail.

Performance projection and MRAM comparison with other memory technologies for embedded applications

The learning from the work described in the preceding section has made it possible to use reasonable projections of device parameters from the recent high-TMR materials advances to estimate the performance of an optimized macro design in the 180-nm technology [58] for which use would be made of those materials. **Figure 12** shows the estimated read-signal development time for a 128-Kb core as a function of device TMR at 300 mV. The curves were generated by optimizing the MTJ resistance and SA

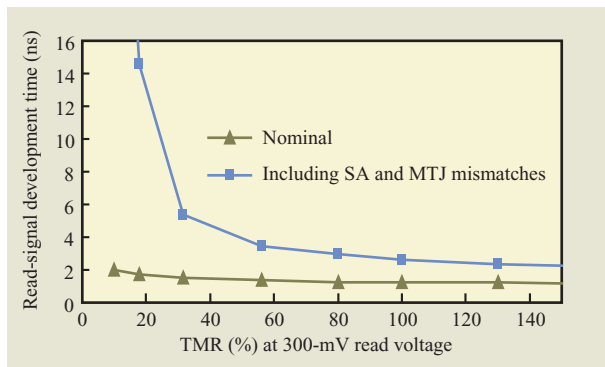


Figure 12

Estimated decrease in read-signal development time for 128-Kb core if TMR were to be increased based upon circuit simulations including margins. From [58], with permission; ©2005 IEEE.

device matching as the TMR was assumed to increase. The results indicate that although nominal device performance would improve only slightly, a very significant improvement should occur for the tails of the distributions: A yieldable array read access time of 3 ns should be achievable with MR ~100% at a 300-mV MTJ bias voltage. Since there are also non-array portions to add to the signal development time to estimate the total access time, it is reasonable to expect a net 5–6-ns read operation for a performance-optimized embedded macro.

By utilizing these and other similar estimates for power and performance parameters for MRAM and other 180-nm embedded memory technologies, we have constructed **Table 3**, which compares embedded SRAM, DRAM, and Flash (NOR Flash) with estimated embedded MRAM for cost, performance, power, and write endurance. We used 180-nm technologies for the comparison, because more solid numbers are available at that node. However, the general trends and conclusions are expected to hold for several future technology generations. The figures of merit for cost are cell area and the percentage of extra processing cost compared to a base logic process. From the table it can be seen that embedded Flash and DRAM have the smallest cell areas, followed by MRAM and then SRAM. In terms of process cost, all memory technologies are approximately 25% more expensive than embedded SRAM because of the varying cost of additional masks and process steps that are required. Next, the table shows a performance figure of merit as read access and write cycle time. As can be seen, embedded SRAM has the fastest read access, followed by closely matched performance for embedded Flash, DRAM, and MRAM. For write-cycle time, embedded SRAM is again the fastest, followed by similar performance for embedded

DRAM and MRAM, with embedded Flash slower by more than two orders of magnitude. Next, data-retention, standby, and read-active and write-active power are considered. Embedded Flash and MRAM have zero data-retention power, followed by embedded SRAM and DRAM. Read-active power is lowest for embedded DRAM and MRAM, followed by embedded SRAM and Flash. Finally, for write-active power, embedded DRAM is the lowest, followed by embedded SRAM, MRAM, and Flash, which is higher by about three orders of magnitude. The final figure of merit shown is write endurance. Only embedded Flash has limited write endurance.

In summary, among the best attributes of MRAM are zero data-retention power, very low standby power, unlimited write endurance, and relatively high speed. Low standby power can be achieved because MRAM has no array leakages (all voltages are zero in an MRAM array in standby), few reference currents, and no pumped supplies that must be maintained. Read-active power is good and is comparable to that for embedded DRAM. Write-active power is high compared to that for embedded SRAM and DRAM, but much lower than for embedded Flash. For MRAM, the write power is dominated by bit-line current, since the word-line current is shared by many bits. The charge/written bit for MRAM is one to four times higher than for embedded SRAM and three to twelve times higher than for embedded DRAM. Read performance is comparable to that of embedded Flash and DRAM, but slower than for embedded SRAM. And finally, write-cycle time is comparable to that of embedded DRAM and slower than that of embedded SRAM, but much faster than that of embedded Flash.

The choice of which embedded memory technology to use depends on the intended application. In some cases the application attributes dictate clear choices. An application would select embedded MRAM if it most valued nonvolatility and low standby power; otherwise, embedded SRAM and DRAM could be used. Furthermore, if the application valued the best possible write performance, write power, and write endurance above other attributes, embedded MRAM would be the best choice; otherwise, embedded Flash could be used. Within IBM, it is of particular interest to compare embedded DRAM and MRAM vs. embedded SRAM for a high-performance application, such as on-chip cache memory for a processor. Although embedded SRAM is faster, for the L2 and L3 caches, the five-to-seven-times-greater densities of embedded MRAM and DRAM can be traded off against performance. Indeed, embedded DRAM, which is already developed and available in high-volume manufacturing, is finding its way onto chip caches, such as the L3 cache of the Blue Gene*/L

Table 3 Typical or estimated parameter values for a variety of embedded memory technologies at the 180-nm technology node. Entries in color indicate areas in which embedded MRAM is expected to have a significant advantage.

Category	Parameter	SRAM	DRAM	NOR Flash	MRAM
Cost	Cell area	3.7 μm^2	0.56 μm^2	0.5 μm^2	0.7–1.4 μm^2
	Process cost adder	0%	25%	25%	25%
Performance	Read access	3.3 ns	13 ns	13 ns	5–20 ns
	Write cycle	3.4 ns	20 ns	5 μs	5–20 ns
Power	Data retention	0.6 nA per bit at 85°C	0.2 nA per bit at 85°C	0	0
	Read active	15 pC per bit	7 pC per bit	30 pC per bit	7 pC per bit
	Write active	15 pC per bit	7 pC per bit	30 nC per bit	45 pC per bit
Miscellaneous	Write endurance	Unlimited	Unlimited	10 ⁵ cycles	Unlimited

processor chip [71] in 130-nm technology. If embedded MRAM were at a comparable state of maturity, it would present a choice that could avoid the overhead of embedded DRAM refreshes and offer the tradeoff of a larger write power vs. lower power for data retention and standby operation. Depending on the application, this could be an attractive tradeoff. The issue for MRAM is that it is not yet at a comparable state of maturity that allows these comparisons to be real options.

Other significant MRAM technology developments

The emphasis in this paper has been on MRAM developments as they have happened or have been viewed from within IBM, in conjunction with IBM funding and IBM development partners. To be sure, other companies have had active and productive MRAM programs with significant developments. Freescale Semiconductor (formerly the Motorola Semiconductor Products Sector) in particular has had an active MRAM development program since before the original 1995 breakthrough of high-TMR magnetic tunnel junctions, and it appears to have the largest effort at this time. Within the DARPA program, Freescale initially pursued both GMR- and TMR-based MRAM, although for many years now they have been focused on the latter. Two significant advances Freescale has pursued have been the use of a magnetic liner around the write wires [72] to enable a two-to-threefold more efficient conversion of current to write field and the use of a novel toggle write architecture [60] for a cell with a flux-closed free layer [the magnetic bit structure is similar to that of Figure 3(d)] that basically eliminates the half-select write issue. The paper in this issue by Worledge [61] analyzes the toggle MRAM device in detail. On the basis of this analysis, IBM designed its 16-Mb chip with a toggle write-mode option. Freescale

has also consistently moved toward an MRAM product and has been making available samples of a 4-Mb standalone toggle MRAM chip in 180-nm technology. Together with their development partners, the Philips Corporation and STMicroelectronics Corporation, they have also recently reported a 90-nm toggle MRAM technology demonstration [73].

Other areas in which significant technological developments were required were in characterization and the equipment for processing. Characterization is covered in part later in this issue (see the paper in this issue by Abraham et al. [48]). For processing, work by companies specializing in tool development was needed. On a Research scale, IBM had equipment like that shown in Figure 2, which has continued to evolve over the years. However, on the scale of even prototyping equipment relevant to the current semiconductor industry, nothing comparable existed. Multilayer tools were available for GMR read-head manufacturing, but these did not address tunnel junction growth. At a basic level, wafer size was generally limited to a diameter of 150 mm or less, and sometimes systems were designed for square substrates. Furthermore, throughput and cleanliness requirements for read-head manufacturing were much less than would be required for semiconductor development. The primary unique processing tools needed were for the formation of the MTJ tunnel junctions and for the annealing of the magnetic layers. Ion milling or reactive ion etching also required some attention in order to address requirements specific to MRAM. IBM worked with a number of tool vendors to define and in some cases jointly address the needs.

One of the most significant developments over the last ten years has been the emergence of reliable and reproducible MTJ deposition equipment capable of handling substrates up to 300 mm in diameter.

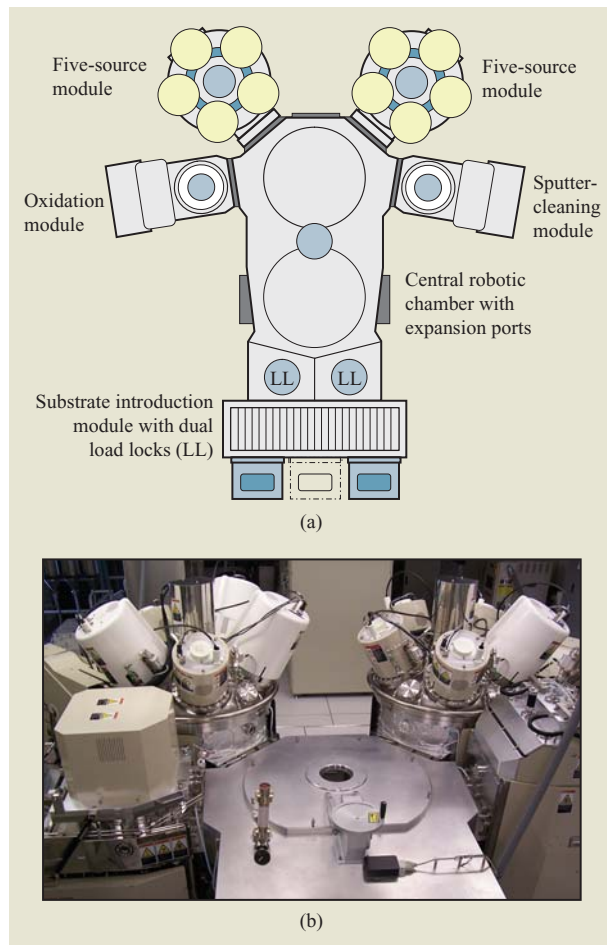


Figure 13

(a) Diagram of Canon ANELVA system suitable for fabricating MTJs on 200-mm or 300-mm-diameter wafers (courtesy of Canon ANELVA); (b) photograph of system installed at IBM, showing portion of central robotic chamber in the foreground, two five-source modules in the background (left and right), and a cleaning (etching) module at the lower left.

Figure 13(a) is a diagram of a system developed for that purpose by Tsunoda et al. [74] of the Canon ANELVA Corporation (Tokyo, Japan). A photograph of such a system, installed at IBM, is shown in **Figure 13(b)**. The system is capable of uniform and reproducible formation of MTJ structures on 200- or 300-mm-diameter substrates and bears some similarity to that of Figure 2. Single chambers contain multiple sputter-deposition sources, and the substrates can be quickly cycled under the sources for multiple sequential thin-layer deposition. A load lock chamber is used to introduce wafers one at a time from a cassette holder. The main deposition chambers, of which there are two on the system, each contain five sputtering

sources. In addition, separate chambers are available for oxidation to form the tunnel barriers and for sputter-cleaning.

Summary and outlook

This paper has described the remarkably rapid development of the magnetic tunnel junction over the last decade as well as the promising MRAM technology based on that device. Over the last decade, many major hurdles for MRAM product development have been surmounted, while technology breakthroughs have continued to occur at a brisk pace. A challenge faced by MRAM, like other new memory technologies, will be finding initial markets that can support the technology and give it time to mature and be developed for more advanced CMOS nodes, which in turn should enable its markets to expand. Outside the scope of this paper have been the equally impressive development of the magnetic tunnel junction as a read-head sensor in disk-drive products [75]. Such sensors have already penetrated a significant part of the disk drive market. In addition, several exciting new proposals have been made—for thermally assisted MRAM [76]; for spin-momentum-transfer MRAM [67, 77], which looks particularly attractive for MRAM at advanced process nodes; and for very-high-density serial magnetic storage [78]. Thus, it is clear that magnetic devices will continue to play multiple evolving roles in computing technology, as has now been the case for fifty years.

Acknowledgments

The authors are grateful to John DeBrosse for compiling Table 3 with the help of John Barth, Chung Lam and Ron Piro. In addition, the authors would like to express their gratitude to their colleagues in magnetic tunnel junction and MRAM development within IBM and also at Infineon Technologies, with whom they worked within the MRAM Development Alliance. In addition, partial financial support from DARPA is acknowledged for some of the developments reviewed in the paper.

*Trademark or registered trademark of International Business Machines Corporation.

References and notes

1. E. W. Pugh, R. A. Henle, D. L. Critchlow, and L. A. Russell, "Solid-State Memory Development in IBM," *IBM J. Res. & Dev.* **25**, No. 5, 585–602 (1981).
2. T. S. Crutcher and J. A. Sartell, "Plated Wire Memory: Its Evolution for Aerospace Utilization," *Honeywell Computer J.* **6**, 15–22 (1972).
3. J. M. Daughton, "Magnetoresistive Memory Technology," *Thin Solid Films* **216**, No. 1, 162–168 (1992).
4. B. Dieny, V. S. Speriosu, S. Metin, S. S. P. Parkin, B. A. Gurney, P. Baumgart, and D. R. Wilhoit, "Magnetotransport Properties of Magnetically Soft Spin-Valve Structures," *J. Appl. Phys.* **69**, No. 8, 4774–4779 (1991); 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**, 2304–2307 (1990); S. S. P. Parkin, Z. G. Li, and D. J. Smith, "Giant Magnetoresistance in Antiferromagnetic Co/Cu Multilayers," *Appl. Phys. Lett.* **58**, 2710 (1991); S. S. P. Parkin, R. Bhadra, and K. P. Roche, "Oscillatory Magnetic Exchange Coupling Through Thin Copper Layers," *Phys. Rev. Lett.* **66**, 2152 (1991).
5. For reviews of the development of spin-valve devices, see S. S. P. Parkin, "Giant Magnetoresistance in Magnetic Nanostructures," *Annual Review of Materials Science*, Vol. 25, B. W. Wessels, Ed., Annual Reviews, Inc., Palo Alto, CA, 1995, pp. 357–388; and S. S. P. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche, and M. Samant, "Magnetically Engineered Spintronic Sensors and Memory," *Proc. IEEE* **91**, 661–680 (2003).
6. D. D. Tang, P. K. Wang, V. S. Speriosu, S. Le, and K. K. Kung, "Spin-Valve RAM Cell," *IEEE Trans. Magn.* **31**, 3206–3208 (1995).
7. R. R. Katti, "Giant Magnetoresistive Random-Access Memories Based on Current-in-Plane Devices," *Proc. IEEE* **91**, No. 5, 687–702 (2003).
8. I. Giaever, "Energy Gap in Superconductors Measured by Electron Tunneling," *Phys. Rev. Lett.* **5**, 147 (1960).
9. P. M. Tedrow and R. Meservey, "Spin-Dependent Tunneling into Ferromagnetic Nickel," *Phys. Rev. Lett.* **26**, No. 4, 192–195 (1971).
10. R. Meservey and P. M. Tedrow, "Spin-Polarized Electron Tunneling," *Phys. Rep.* **238**, No. 4, 173–243 (1994).
11. M. Julliere, "Tunneling Between Ferromagnetic Films," *Phys. Lett. A* **54**, No. 3, 225–226 (1975).
12. J. C. Slonczewski, "Magnetic Bubble Tunnel Detector," *IBM Tech. Disclosure Bull.* **19**, No. 6, 2328–2330 (1976).
13. J. C. Slonczewski, "Magnetic-Field Tunnel-Sensor," *IBM Tech. Disclosure Bull.* **19**, No. 6, 2331–2332 (1976).
14. J. C. Slonczewski, "Magnetic-Barrier Current Amplifier," *IBM Tech. Disclosure Bull.* **19**, No. 6, 2333–2336 (1976).
15. S. Maekawa and U. Gafvert, "Electron Tunneling Between Ferromagnetic Films," *IEEE Trans. Magn.* **MAG-18**, 707–708 (1982).
16. J. C. Slonczewski, "Conductance and Exchange Coupling of Two Ferromagnets Separated by a Tunneling Barrier," *Phys. Rev. B (Condens. Matter)* **39**, No. 10, 6995–7002 (1989).
17. J. Matisoo, "Tunneling Cryotron Using the Josephson Effect," *IEEE Trans. Magn.* **MAG-4**, No. 3, 324 (1968).
18. W. Anacker, "Josephson Computer Technology: An IBM Research Project," *IBM J. Res. & Dev.* **24**, No. 2, 107–112 (1980); see also the other papers in this issue.
19. M. B. Ketchen, D. J. Herrell, and C. J. Anderson, "Josephson Cross-Sectional Model Experiment," *J. Appl. Phys.* **57**, 2550–2573 (1985).
20. J. M. Rowell, M. Gurvitch, and J. Geerk, "Modification of Tunneling Barriers on Nb by a Few Monolayers of Al," *Phys. Rev. B (Condens. Matter)* **24**, No. 4, 2278–2281 (1981).
21. M. Gurvitch, J. M. Rowell, H. A. Huggins, M. A. Washington, and T. A. Fulton, "Nb Josephson Tunnel Junctions with Thin Layers of Al Near the Barrier," *IEDM Tech. Digest*, pp. 115–117 (1981).
22. M. B. Ketchen, D. Pearson, A. W. Kleinsasser, C. K. Hu, M. Smyth, J. Logan, K. Stawiasz, E. Baran, M. Jaso, T. Ross, K. Petrillo, M. Manny, S. Basavaiah, S. Brodsky, S. B. Kaplan, W. J. Gallagher, and M. Bhushan, "Sub- μm , Planarized, Nb–AlO–Nb Josephson Process for 125 mm Wafers Developed in Partnership with Si Technology," *Appl. Phys. Lett.* **59**, No. 20, 2609–2611 (1991).
23. T. Miyazaki, T. Yaoi, and S. Ishio, "Large Magnetoresistance Effect in 82Ni–Fe/Al–Al₂O₃/Co Magnetic Tunneling Junction," *J. Magn. Magn. Mater.* **98**, No. 1–2, L7–L9 (1991).
24. T. Yaoi, S. Ishio, and T. Miyazaki, "Magnetoresistance in 82Ni–Fe/Al–Al₂O₃/Co Junction—Dependence of the Tunneling Conductance on the Angle Between the Magnetizations of Two Ferromagnetic Layers," *IEEE Trans. J. Magn. Jpn.* **8**, No. 8, 498–504 (1993).
25. J. Nowak and J. Rauluszkiwicz, "Spin Dependent Electron Tunneling Between Ferromagnetic Films," *J. Magn. Magn. Mater.* **109**, 79–90 (1992).
26. T. Miyazaki and N. Tezuka, "Giant Magnetic Tunneling Effect in Fe/Al₂O₃/Fe Junction," *J. Magn. Magn. Mater.* **139**, No. 3, L231–L234 (1995).
27. J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions," *Phys. Rev. Lett.* **74**, No. 16, 3273–3276 (1995).
28. S. S. P. Parkin, "Giant Magnetoresistance and Oscillatory Interlayer Coupling in Polycrystalline Transition Metal Multilayers," *Ultrathin Magnetic Structures*, Vol. II, B. Heinrich and J. A. C. Bland, Eds., Springer-Verlag, Berlin, 1994, p. 148.
29. W. J. Gallagher, "Three-Terminal Superconducting Devices," *IEEE Trans. Magn.* **MAG-21**, No. 2, 709–716 (1985).
30. R. L. Sandstrom, A. W. Kleinsasser, W. J. Gallagher, and S. I. Raider, "Josephson Integrated Circuit Process for Scientific Applications," *IEEE Trans. Magn.* **MAG-23**, No. 2, 1484–1488 (1987).
31. W. P. Pratt, Jr., S.-F. Lee, J. Slaughter, R. Loloe, P. A. Schroeder, and J. Bass, "Perpendicular Giant Magnetoresistance of Ag/Co Multilayers," *Phys. Rev. Lett.* **66**, 3060 (1991).
32. M. A. M. Gijs, S. K. J. Lenczowski, and J. B. Giesbers, "Perpendicular Giant Magnetoresistance of Microstructured Fe/Cr Magnetic Multilayers from 4.2 to 300 K," *Phys. Rev. Lett.* **70**, 3343 (1993).
33. W. J. Gallagher, S. S. P. Parkin, L. Yu, X. P. Bian, A. Marley, R. A. Altman, S. A. Rishton, K. P. Roche, C. Jahnes, T. M. Shaw, and X. Gang, "Microstructured Magnetic Tunnel Junctions," *J. Appl. Phys.* **81**, No. 8, 3741–3746 (1997); S. S. P. Parkin, R. E. Fontana, and A. C. Marley, "Low Field Magnetoresistance in Magnetic Tunnel Junctions Prepared by Contact Masks and Lithography: 25% Magnetoresistance at 295 K in Mega Ohm Micron Sized Junctions," *J. Appl. Phys.* **81**, 5521 (1997).
34. S. S. P. Parkin, K. P. Roche, M. G. Samant, P. M. Rice, R. B. Beyers, R. E. Scheuerlein, E. J. O'Sullivan, S. L. Brown, J. Buccignano, D. W. Abraham, Y. Lu, M. Rooks, P. L. Trouilloud, R. A. Wanner, and W. J. Gallagher, "Exchange-Biased Magnetic Tunnel Junctions and Application to Nonvolatile Magnetic Random Access Memory," *J. Appl. Phys.* **85**, No. 8, 5828–5833 (1999).
35. S. A. Rishton, Y. Lu, R. A. Altman, A. C. Marley, X. P. Bian, C. Jahnes, R. Viswanathan, G. Xiao, W. J. Gallagher, and S. S. P. Parkin, "Magnetic Tunnel Junctions Fabricated at Tenth-Micron Dimensions by Electron Beam Lithography," *Microelectron. Eng.* **35**, No. 1–4, 249–252 (1997).
36. S. S. P. Parkin, K. S. Moon, K. E. Pettit, D. J. Smith, R. E. Dunin-Borkowski, and M. R. McCartney, "Magnetic Tunnel Junctions Thermally Stable to Above 300 Degrees C," *Appl. Phys. Lett.* **75**, No. 4, 543–545 (1999).
37. S. Gider, B. U. Runge, A. C. Marley, and S. S. P. Parkin, "The Magnetic Stability of Spin-Dependent Tunneling Devices," *Science* **281**, No. 5378, 797–799 (1998).
38. 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**, 2304–2307 (1990).
39. M. G. Samant, J. A. Luning, J. Stohr, and S. S. P. Parkin, "Thermal Stability of IrMn and MnFe Exchange-Biased Magnetic Tunnel Junctions," *Appl. Phys. Lett.* **76**, No. 21, 3097–3099 (2000).
40. W. J. Gallagher and S. S. P. Parkin, "High-Speed 128Kbit MRAM Core in a 0.18 μm CMOS Technology Utilizing PtMn-Based Magnetic Tunnel Junctions," *Proceedings of the Nonvolatile Memory Technology Symposium*, November 2003,

- D. Willmott, M. Wood, W. J. Gallagher, G. Mueller, and A. R. Sitaram, "A 16Mb MRAM Featuring Bootstrapped Write Drivers," *Digest of Technical Papers, Symposium on VLSI Circuits (IEEE Cat. No. 04CH37525)*, 2004, pp. 454–457.
70. 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**, No. 1, 25–39 (2006, this issue).
 71. 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).
 72. M. Durlam, P. Naji, A. Omair, M. DeHerrera, J. Calder, J. M. Slaughter, B. Engel, N. Rizzo, G. Grynkeiwich, B. Butcher, C. Tracy, K. Smith, K. Kyler, J. Ren, J. Molla, B. Feil, R. Williams, and S. Tehrani, "A Low Power 1 Mbit MRAM Based on 1T1MTJ Bit Cell Integrated with Copper Interconnects," *Digest of Technical Papers, Symposium on VLSI Circuits (IEEE Cat. No. 02CH37302)*, 2002, pp. 158–161.
 73. M. Durlam, T. Andre, P. Brown, J. Calder, J. Chan, R. Cuppens, R. W. Dave, T. Ditewig, M. DeHerrera, B. N. Engel, B. Feil, C. Frey, D. Galpin, G. Garni, G. Grynkeiwich, J. Janesky, G. Kerszykowski, M. Lien, J. Martin, J. Nahas, K. Nagel, K. Smith, C. Subramanian, J. J. Sun, J. Tamim, R. Williams, L. Wise, S. Zoll, F. List, R. Fournel, B. Martino, and S. Tehrani, "90nm Toggle MRAM Array with $0.29\mu\text{m}^2$ Cells," *Digest of Technical Papers, Symposium on VLSI Technology*, 2005, pp. 186–187.
 74. T. Tsunoda, M. Kosuda, K. Tsunekawa, and N. Watanabe, "A Novel PVD Tool for Magnetoresistive Random Access Memory (MRAM) Mass Production," *Semicond. Manuf.* **4**, 154–165 (2003).
 75. S. Mao, J. Nowak, D. Song, P. Kolbo, L. Wang, E. Linville, D. Sanders, E. Murdock, and P. Ryan, "Spin Tunneling Heads above 20Gb/in^2 ," *IEEE Trans. Magn.* **38**, 78 (2002).
 76. 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. Magn.* **40**, 2625 (2004); O. Redon, M. Kerekes, R. Sousa, L. Prejbeanu, H. Sibuet, F. Ponthonnier, A. Persico, and J. P. Nozieres, "Thermo Assisted MRAM for Low Power Applications" *Proceedings of the First International Conference on Memory Technology and Design (ICMTD-2005)*, Giens, France, May 21–24, 2005, pp. 113–114.
 77. 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, N. C. Emley, I. N. Krivorotov, P. M. Braganca, E. M. Ryan, S. I. Kiselev, J. C. Sankey, D. C. Ralph, R. A. Buhrman, and J. A. Katine, "Spin Transfer Effects in Nanoscale Magnetic Tunnel Junctions," *Appl. Phys. Lett.* **85**, 1205 (2004).
 78. S. S. 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.

Received June 14, 2005; accepted for publication October 4, 2005; Internet publication January 24, 2006

William J. Gallagher *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (glgr@us.ibm.com)*. Dr. Gallagher joined the IBM Research Division in 1978 after receiving his B.S. degree in physics (summa cum laude) from Creighton University in 1974 and his Ph.D. degree in physics from MIT. He worked for five years at IBM on scientific and engineering aspects of Josephson computer technology and then, for six years, managed the IBM Exploratory Cryogenics Research Group. In 1989, he participated in the formation of the IBM–AT&T–MIT Consortium for Superconducting Electronics (CSE). He served as a director of the CSE from 1989 until 1995. Since 1995, he has led an effort to explore the use of magnetic tunnel junctions for a nonvolatile random access memory, MRAM, including serving from 2000 to 2004 as the IBM project manager in the MRAM Development Alliance with Infineon. Currently Dr. Gallagher is Senior Manager of the Exploratory Nonvolatile Memories program at the IBM Thomas J. Watson Research Center. He is a Fellow of the American Physical Society and of the Institute of Electrical and Electronics Engineers.

Stuart S. P. Parkin *IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (parkin@almaden.ibm.com)*. Dr. Parkin joined IBM Research in San Jose in 1982 as a World Trade Postdoctoral Fellow, becoming a permanent member of the staff the following year. His current work involves the study of magnetic tunnel junctions and the development of an advanced nonvolatile magnetic random access memory based on magnetic tunnel junction storage cells. His earlier research interests have included organic superconductors, ceramic high-temperature superconductors, and, most recently, the study of magnetic thin-film structures and nanostructures exhibiting giant magnetoresistance (GMR). In 1991, he discovered oscillations in the magnitude of the interlayer exchange coupling in transition-metal magnetic multilayered GMR systems. For this and related work, Dr. Parkin shared both the American Physical Society International New Materials Prize (1994) and the European Physical Society Hewlett-Packard Europhysics Prize (1997). He has received other awards, including the Materials Research Society Outstanding Young Investigator Award (1991), the Charles Vernon Boys' Prize from the Institute of Physics, London (1991), the 1999–2000 American Institute of Physics Prize for Industrial Applications of Physics, and several awards from IBM. In 2001 he was named R&D Magazine's first Innovator of the Year. A native of the United Kingdom, Dr. Parkin received his B.A. degree in 1977; he was elected a Research Fellow in 1979 at Trinity College in Cambridge, England, and was awarded his Ph.D. degree in 1980 at the Cavendish Laboratory, also in Cambridge. He was elected a Fellow of the Royal Society in 2000; he is also a Fellow of the American Physical Society and the Institute of Physics (London). In 1997, Dr. Parkin was elected a member of the IBM Academy of Technology and named an IBM Research Master Inventor. In 1999 he was appointed an IBM Fellow, IBM's highest technical honor.

Erratum

In the paper “Development of the magnetic tunnel junction MRAM at IBM: From first junctions to a 16-Mb MRAM demonstrator chip” by W. J. Gallagher and S. S. P. Parkin in the *IBM Journal of Research and Development*, Volume 50, No. 1, January 2006, the legend in Figure 1(d) should read $T = 295$ K instead of 4.2 K. In the legend in Figure 5, the triangle and circle symbols were erroneously interchanged. The corrected figure appears below.

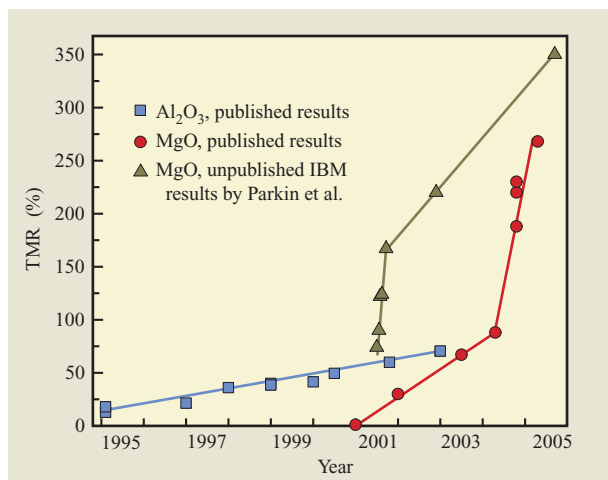


Figure 5

Maximum demonstrated TMR values over time for magnetic tunnel junctions with Al₂O₃ and with MgO tunnel barriers.