nature reviews electrical engineering

https://doi.org/10.1038/s44287-024-00111-z

Review article

Check for updates

Spin-transfer torque magnetoresistive random access memory technology status and future directions

Daniel C. Worledge[1] & Guohan Hu[2]

Abstract

Spin-transfer torque magnetoresistive random access memory (STT-MRAM) is a non-volatile memory technology with a unique combination of speed, endurance, density and ease of fabrication, which has enabled it to recently replace embedded Flash as the embedded non-volatile memory of choice for advanced applications, including automotive microcontroller units. In this Review, we describe the working principles of STT-MRAM, and provide a brief history of its development. We then discuss the requirements, product status and outlook for four key STT-MRAM applications: stand-alone, embedded non-volatile memory, non-volatile working memory and last-level cache. Finally, we review potential future directions beyond STT-MRAM, including spin–orbit torque MRAM (SOT-MRAM) and voltage control of magnetic anisotropy MRAM (VCMA-MRAM), with an emphasis on their technological potential.

Sections

Introduction

STT-MRAM operation

Applications

Future directions

Outlook

[1]IBM Almaden Research Center, San Jose, CA, USA. [2]IBM T. J. Watson Research Center, Yorktown Heights, NY, USA.
e-mail: worledge@us.ibm.com

Key points

- All advanced foundries now offer embedded spin-transfer torque magnetoresistive random access memory (STT-MRAM) as a replacement for embedded Flash below the 28 nm node, where embedded Flash does not exist.
- Embedded STT-MRAM is planned to be used in the next generation of automotive microcontroller units.
- In the near term, STT-MRAM is being developed for use as a non-volatile working memory for ultra-low-power, low-performance edge and Internet of Things applications, replacing both eFlash and SRAM.
- Longer term, STT-MRAM research is focused on reducing the write current to enable last-level cache.
- Areas of research to improve MRAM beyond STT include spin–orbit torque and voltage control of magnetic anisotropy.

Introduction

Spin-transfer torque magnetoresistive random access memory (STT-MRAM) is an emerging memory technology that stores information in a magnetic tunnel junction (MTJ)[1]. STT-MRAM provides a unique combination of non-volatility, high write endurance and high speed. Furthermore, this technology can be easily integrated into a standard semiconductor back-end-of-line process flow, enabling a wide range of applications.

STT-MRAM has a long history in research and development since the invention of the MTJ in the 1970s (refs. 2,3) (see Box 1). An earlier version of magnetic field-switched MRAM, Toggle MRAM[4], was successfully commercialized but has a limited market, due to the lack of scalability and high cost. Despite promising research results, until recently it was not clear whether STT-MRAM would be a commercial success. However, in recent years, STT-MRAM has replaced embedded Flash as an embedded non-volatile memory (eNVM), laying a solid foundation for the future of STT-MRAM technology.

Samsung started selling its first embedded STT-MRAM (eMRAM) products in 2019. All advanced semiconductor foundries, including TSMC, GlobalFoundries and Samsung, have announced their plans to replace embedded Flash with eMRAM beyond the 28 nm node, to reduce cost and complexity. eMRAM is poised to make a major impact on the world, for example, in automotive microcontroller units, driven by the adoption of hybrid, electric and self-driving vehicles. Having emerged from research and development into manufacturing, this use of STT-MRAM as eNVM is expected to grow steadily as new microcontroller unit circuit designs are migrated to advanced nodes in the next few years. Furthermore, a wide range of research directions promise even more advanced applications in the future, such as non-volatile working memory and last-level cache.

In this Review, we elaborate on the basic device physics behind STT-MRAM operation and review the history of its development. We discuss the different types of STT-MRAM applications, as well as current products and near-term directions towards more advanced STT-MRAM. Finally, we summarize the latest research trends and potential future directions, with a critical eye for those topics most likely to be of practical use.

STT-MRAM operation

STT-MRAM stores information in the magnetization direction of a free layer in an MTJ (Fig. 1a). The information is read out using the magnetoresistance of the tunnel junction, with the resistance of the junction being several times higher when the free and reference layer magnetizations are antiparallel compared with parallel (Fig. 1b). Reading is performed at a lower voltage than that used for writing, because stochastic read disturb errors can occur if the read voltage is too high. Information is written to the junction via STT (Box 2), by driving an electric current either up or down through the junction, to write ‘1’ or ‘0’. The electrons are spin-polarized by the reference layer, and as they traverse the tunnel barrier, spin angular momentum is transferred to the free layer, causing its magnetization to switch to the reverse direction[5]. Data retention is ensured by an energy barrier, E_b (Fig. 1c), separating the parallel and antiparallel states, created by perpendicular magnetic anisotropy (Box 3). Accurately measuring E_b requires bake retention measurements on arrays of bits at elevated temperatures[6]. Retaining data at higher temperatures requires higher E_b, which in turn increases the switching current (Fig. 1d).

STT switching is thermally activated for long write pulses at lower currents, where the effective E_b is partly reduced (Fig. 1e). At short write pulses with larger currents, where the effective E_b is reduced to zero, the switching current is inversely proportional to the pulse length, due to conservation of angular momentum[7]. In this Review, we use the single-domain model for approximating switching currents and activation energies. The threshold switching current, I_{c0}, is defined as the thermally activated write current extrapolated back to the thermal activation attempt time $\tau_0 = 1$ ns. All bits have an intrinsic write-error rate, caused by the STT vanishing when the free and reference layers are parallel or antiparallel (Box 2). Therefore, a sufficiently large thermal fluctuation is required to initiate every write. Hence, bits must be engineered carefully to achieve steep write-error rate slopes free of anomalies (Fig. 1f). Accurately evaluating the write-error rate requires measuring down to an error floor of 10^{-6} errors per write or below, ideally on hundreds of junctions. Data reported on a linear scale instead of a logarithmic scale are not sufficient to evaluate the write-error rate. Both the write-error rate slope (the slope of the curve in Fig. 1f) and bit-to-bit distributions (the variation in switching current from bit to bit) determine I_{write}, the current which will reliably write all the bits in the memory (Fig. 1f). Write endurance is limited by MgO tunnel barrier reliability; although there is no magnetic wear-out mechanism, the write voltage across the MgO barrier can move oxygen ions over time within the barrier and into the neighbouring metal layers, reducing the MgO, leading to shorted bits. Endurance is evaluated using time-dependent dielectric breakdown measurements on arrays of devices at high write voltage to accelerate the failure mechanism[8]. Measurements of individual bits surviving 10^{10} write pulses are not sufficient to evaluate endurance.

Each memory cell contains one MTJ and one access transistor, which is used to select the junction for reading and writing (Fig. 2a). STT-MRAM contains an array of junctions, with the word lines orthogonal to the bit lines, to allow selection of a single bit at the intersection of the selected word and bit lines (Fig. 2b). The read or write current is returned through source lines, which can be oriented parallel to either the bit lines (to maintain constant bit line plus source line resistance for all bits in the array, with the source line on the same metal level as the bit line) or word lines (for maximum density) (Fig. 2b,c). Cell density (Fig. 2d) is a critical consideration for STT-MRAM technology. Today’s STT-MRAM cell area is not limited by the size of the MTJ but, rather, by

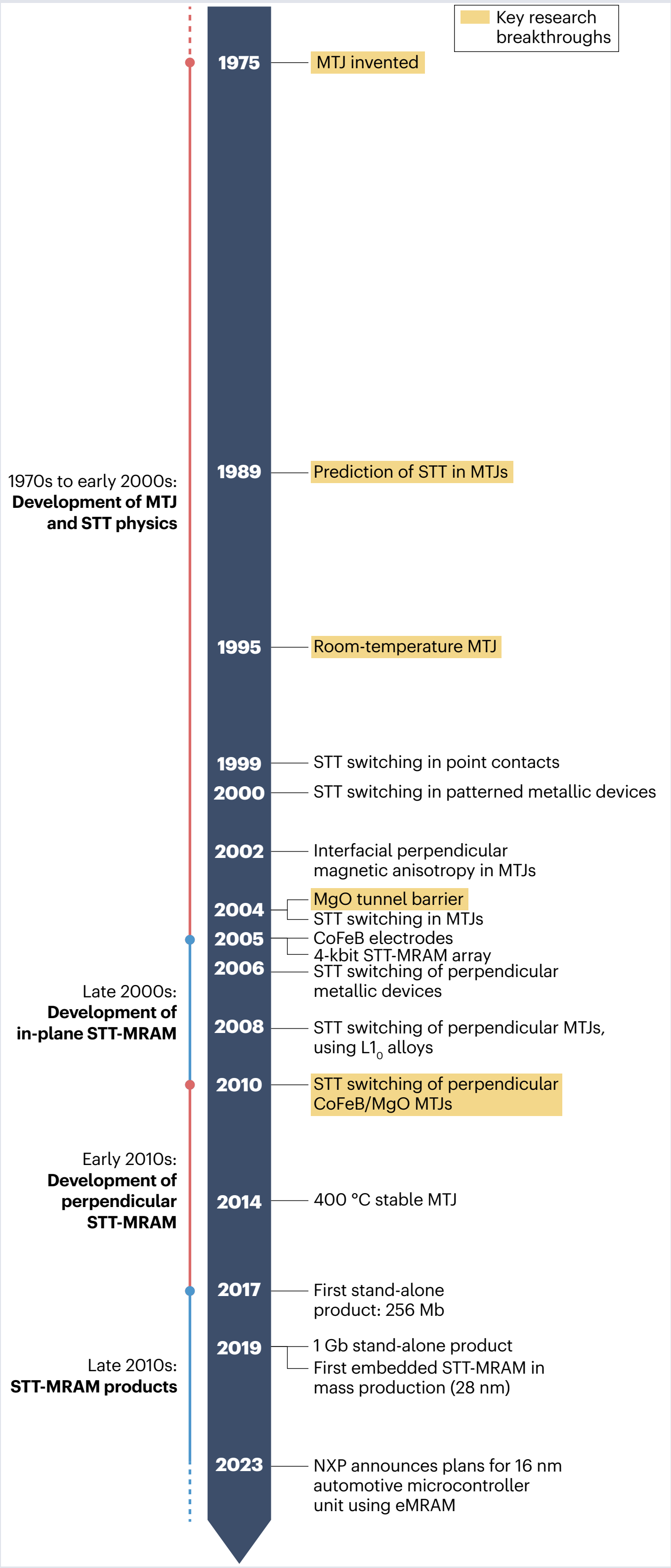
Box 1 | Early history

The magnetic tunnel junction (MTJ) was invented independently by Julliere[2] and Slonczewski[3], with the first experimental results demonstrated at low temperature in 1975 (see the figure)[2]. Room-temperature MTJs were first demonstrated in 1995 (ref. 130), which launched substantial worldwide interest in developing MTJs for use in magnetic memory and as read sensors in hard disk drives. Early results used amorphous AlO$_x$ tunnel barriers, which limited the magnetoresistance, $(R_{\mathrm{high}}-R_{\mathrm{low}})/R_{\mathrm{low}}$, to about 70% (ref. 131). Initially, hard disk drive read heads and Toggle magnetoresistive random access memory (MRAM) used AlO$_x$ tunnel barriers (or TiO$_2$, for read heads). Parkin[132] and Yuasa[133] independently discovered high magnetoresistance using crystalline MgO tunnel barriers[134,135] and CoFe or Fe electrodes, reporting magnetoresistance as high as 220%. Subsequently, easier to grow CoFeB electrodes with 230% magnetoresistance were introduced[9]. These milestones have had a major impact on the world by enabling dense hard disk drives, and laid the foundation for MRAM.

In 1989, spin-transfer torque (STT) was theoretically predicted in MTJs[136]. Independently, Luc Berger theoretically explored the torque that an electric current applies to a domain wall in a single magnetic film[137], which was eventually recognized to be the same STT physics[138,139]. Although these theoretical predictions presented a major breakthrough, early experimental demonstrations of STT[140,141] were challenging, requiring junctions with diameter below 100 nm to separate a small STT effect from a larger effect of the Oersted magnetic field simultaneously generated by the current.

The first demonstration of controllable STT switching was reported in 1999 in point contact junctions[142], and in 2000 in patterned junctions[143], both using metallic giant magnetoresistive multilayers. Despite the STT switching concept demonstration, applications require the use of MTJs, owing to the low resistance of metallic devices (<1 Ω), too small to be easily sensed when placed in series with an access transistor (2–10 kΩ). STT switching in MTJs was demonstrated in 2004 (refs. 144,145), and one year later Sony demonstrated a 4 kbit STT-MRAM array with basic read, write and storage functionality[146]. This and subsequent work used in-plane MTJs, which resulted in large switching currents and unreliable switching. It had been known since Slonczewski's original work that perpendicularly magnetized junctions would enable lower switching currents[7], and in 2006 STT switching of perpendicular magnetization was demonstrated in metallic giant magnetoresistive multilayers[147,148]. However, it was experimentally challenging to simultaneously obtain magnetoresistance and perpendicular magnetization in tunnel junctions. Although Toshiba demonstrated initial results on individual perpendicular junctions using L1$_0$ ordered alloys[149], this new class of materials did not prove technologically useful.

In 2010, Tohoku University[150] and IBM[151] independently published the first demonstration of STT switching in perpendicularly magnetized CoFeB-based tunnel junctions, using the perpendicular magnetic anisotropy at the CoFeB/MgO interface[152,153] (see Box 3). In addition to the expected benefits of the lower switching current and scaling of junctions, the perpendicular MTJ stack[154] also solved the problem of unreliable switching[151]. These papers set off a flurry of activity in industry to develop STT-MRAM products that, at present, are all based on perpendicularly magnetized CoFeB-based tunnel junctions. eMRAM, embedded STT-MRAM.


Key research breakthroughs
1975
MTJ invented
1970s to early 2000s: Development of MTJ and STT physics
1989
Prediction of STT in MTJs
1995
Room-temperature MTJ
1999
STT switching in point contacts
2000
STT switching in patterned metallic devices
2002
Interfacial perpendicular magnetic anisotropy in MTJs
2004
MgO tunnel barrier
STT switching in MTJs
2005
CoFeB electrodes
4-kbit STT-MRAM array
2006
STT switching of perpendicular metallic devices
Late 2000s: Development of in-plane STT-MRAM
2008
STT switching of perpendicular MTJs, using L1$_0$ alloys
2010
STT switching of perpendicular CoFeB/MgO MTJs
Early 2010s: Development of perpendicular STT-MRAM
2014
400 °C stable MTJ
2017
First stand-alone product: 256 Mb
Late 2010s: STT-MRAM products
2019
1 Gb stand-alone product
First embedded STT-MRAM in mass production (28 nm)
2023
NXP announces plans for 16 nm automotive microcontroller unit using eMRAM

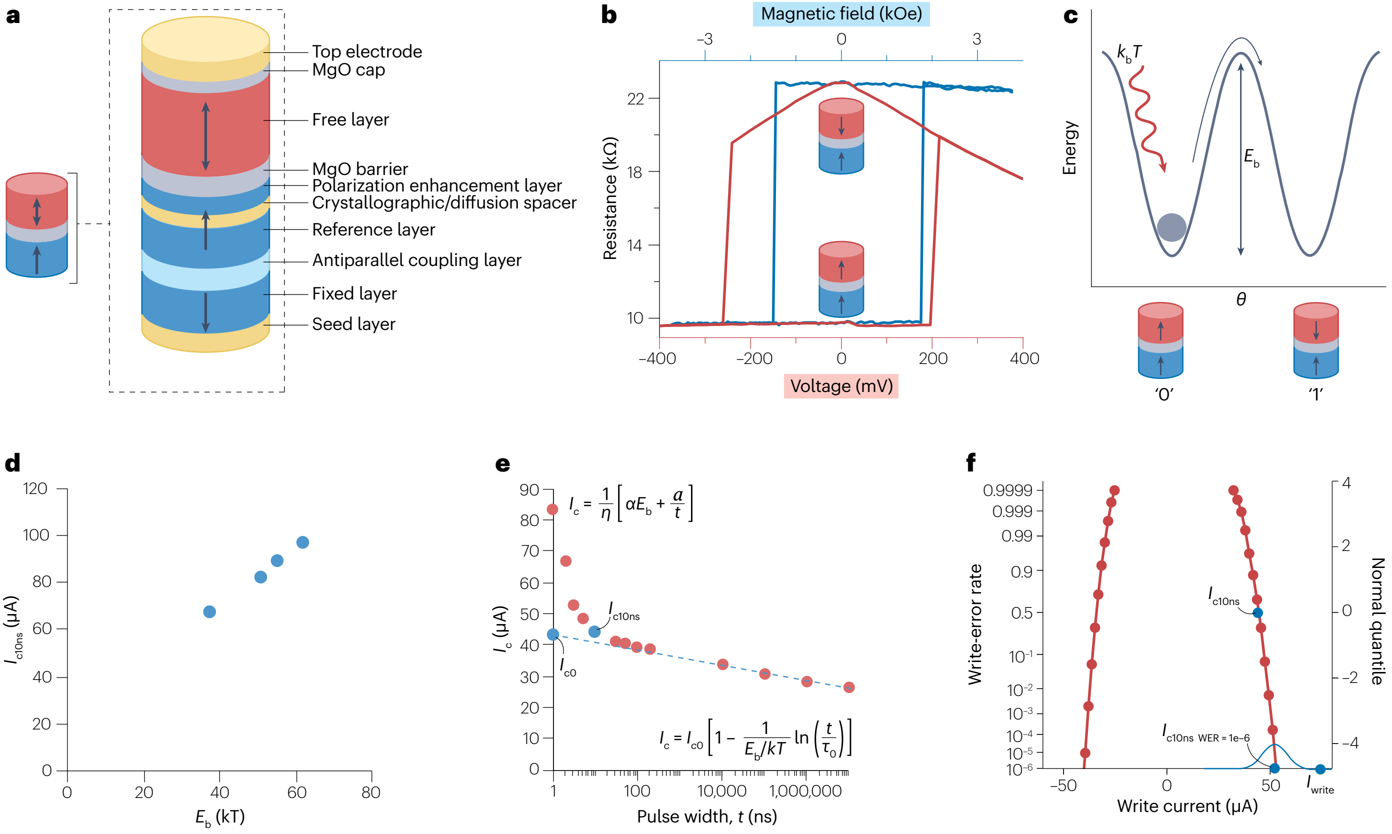


Fig. 1 | Magnetic tunnel junctions. a, Magnetic tunnel junction (MTJ) film stack. The reference layer contains two antiparallel coupled layers; for simplicity, only the top reference layer is shown in many of the figures in this article. **b**, The free layer can be switched by an applied voltage using spin-transfer torque (STT), instead of an applied magnetic field. The resistance of the MgO barrier depends on the relative orientations of the free and reference magnetic layers. **c**, Data retention is determined by the energy barrier, E_b, caused by perpendicular magnetic anisotropy (see Box 3). **d**, Higher E_b junctions require larger switching currents, increasing memory size and decreasing write endurance. **e**, Switching current, I_c, is thermally activated for long write pulses, and inversely proportional to write pulse width for short write pulses. In the top formula, η is proportional to the reference layer spin polarization, α is the magnetic damping constant and a depends on the free-layer material but is independent of pulse width, t. **f**, Every bit has a finite write-error rate, defined as the fraction of attempts, when a write pulse is applied, that the bit does not write. The commonly used normal quantile scale allows the read disturb rate (1 − write-error rate) at low currents to be analysed. The current at which all bits can be reliably written, I_{write}, is determined by write-error rate and bit-to-bit distributions. I_{write} must be kept below the threshold for breakdown. I_{c0}, threshold switching current.

the size of the access transistor (Fig. 2c), which must be designed to be large enough to deliver the required write current. Hence, reducing the switching current, I_c, is critical for advanced MRAM. With the current progress in transistor scaling, the STT-MRAM cell size can be reduced by moving to a more advanced technology node, because more advanced transistors can deliver the required current in a smaller area. For example, as Everspin's products moved from the 40 nm to 28 nm node, the cell area was reduced from 0.156 μm^2 to 0.041 μm^2. Most of this approximately four times area decrease is due to improved transistor performance. In addition to the array, the STT-MRAM macro contains significant peripheral circuitry (Fig. 2e), including many sense amplifiers enabling reading of the small read signals.

Applications

STT-MRAM applications can be divided into four broad categories: stand-alone, eNVM, non-volatile working memory and last-level cache. Stand-alone memory is a general-purpose memory chip that contains only memory, whereas embedded memory means that the STT-MRAM is fabricated on a custom-designed chip along with logic and, potentially, other functions (such as wireless communications, analogue circuitry and sensors) at a foundry. Non-volatile working memory and last-level cache are also examples of embedded applications.

Regardless of the type of application, common STT-MRAM technology challenges, associated, for example, with MTJ deposition and ion beam etching, must be addressed. The deposition tool vendors Applied Materials, TEL, Canon-Anelva and Singulus have overcome numerous fundamental product challenges to enable reliable deposition of the MTJ[9–11]. Maintaining stability of the resistance–area product of the MgO tunnel barrier in the deposition process is challenging, and even small fluctuations of a few per cent from run to run can reduce manufacturing yields. MgO deposited by depositing magnesium metal followed by oxidation in an O_2 ambient tends to produce a more stable resistance–area product and fewer particles than MgO deposited by radio-frequency sputtering. However, radio-frequency sputtering tends to give higher magnetoresistance and tighter resistance distributions from bit to bit. Further progress in both methods will be important for future products. The target purity requirements for some materials have pushed vendors to develop high purity targets, with purity of MgO reaching 99.9999%.

High deposition throughput has always been challenging for MRAM stacks, due to multiple individual layers (Fig. 1a). Moreover, some of the deposition steps involve cryo-cooling to obtain smoother layers or heating to improve crystalline quality, both of which take additional time to reach the deposition temperature, and in some cases return to room temperature for subsequent layers[11]. Typical deposition throughput, lower than five wafers per hour, is still low compared with standard complementary metal oxide semiconductor (CMOS) process steps. A related challenge is to arrange the targets and chambers so that the wafer does not pass through the same chamber twice, a basic contamination requirement in modern semiconductor fabrication. This typically involves the use of two chambers with MgO targets, one for the MgO barrier and the other for the MgO cap; each chamber may have multiple MgO targets to increase throughput, especially important due to the very low sputter rate of MgO, <0.01 nm s^{-1} for a single target.

For years, etching MTJs was a major roadblock to advanced products. Reactive ion etching of cobalt, iron and many other elements in the complex MTJ stack is difficult, due to the challenge of forming volatile etch products without chemically attacking the remaining junction material[12]. The most successful reactive ion etching system, methanol, oxidizes the edges of the MTJs and so cannot be used in junctions with diameters below about 100 nm. This edge damage shows up as an increase of device resistance–area product and switching voltage at smaller diameter. In well-prepared junctions, both should be constant with the device diameter. Ion beam etching has been developed as an effective solution for MTJ etching in advanced MRAM, and is universally used in advanced products. Here the cobalt, iron and other hard to etch materials are physically sputtered away using argon instead of chemically

Box 2 | Spin-transfer torque mechanism

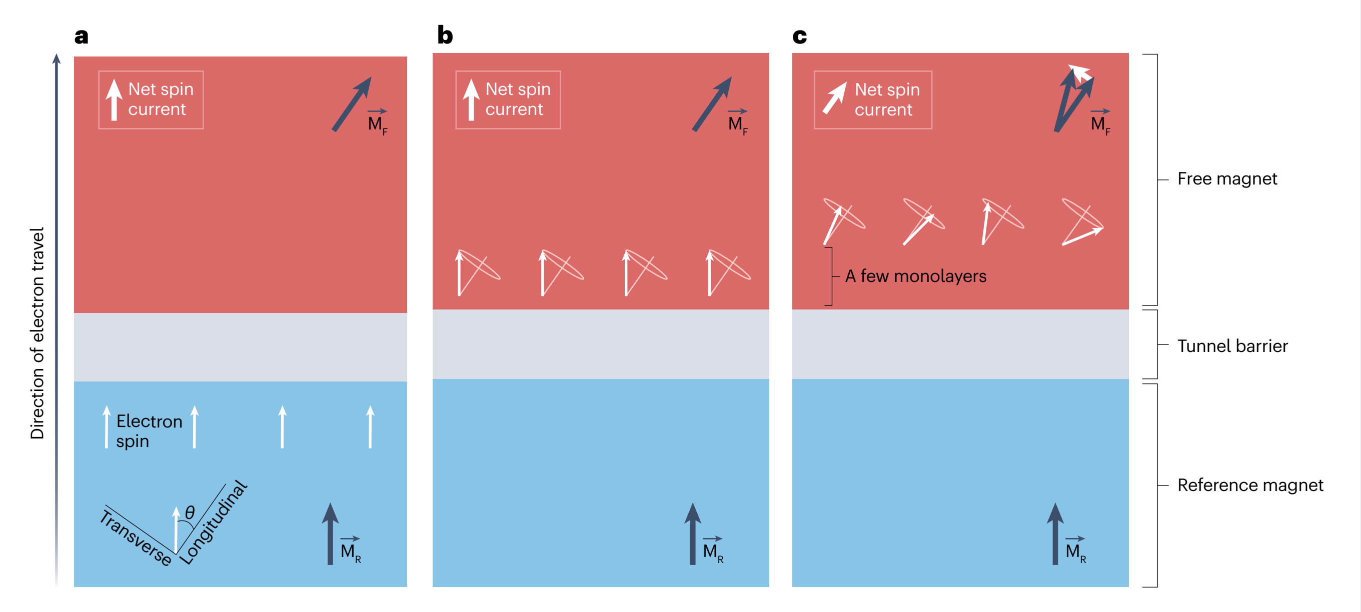
Consider a current of electrons travelling up, from a reference magnet into a free magnet, with magnetizations at a relative angle θ. Assume that the magnets are 100% spin-polarized, for simplicity (see the figure).

Initially, the net spin current is parallel to the reference magnetization (see the figure, panel **a**). After tunnelling through the barrier, we can consider the spins classically as magnetic moments precessing in the exchange field of the free magnet (see the figure, panel **b**). As the exchange field is very large (hundreds of Tesla), the precession frequency is very high (in the terahertz). Quantum mechanically, the precession frequency depends on the momentum of each electron, so each electron precesses at a different frequency.

After traversing only a few monolayers, the spins have precessed many times, and at different frequencies, resulting in each spin pointing in a different direction (see the figure, panel **c**). This spin dephasing is the fundamental cause of spin-transfer torque (STT). After a few monolayers, the net spin current is purely longitudinal; the transverse spin has been averaged out by the dephasing. Conservation of angular momentum requires that this transverse spin is conserved — it is absorbed by the free magnetization, rotating it towards the reference magnetization.

When the current is reversed, spins travelling from the free magnet to the reference magnet are partially reflected at the tunnel barrier. The reflected spin component then dephases in the same way, rotating the free magnetization antiparallel to the reference magnetization.

The STT is proportional to the transverse spin, and hence to sin(θ). The torque therefore vanishes when θ=0 or π; thermal fluctuations large enough to move θ significantly away from zero or π are required to initiate STT in practice. A write error is caused if a sufficiently large thermal fluctuation does not arrive during the write pulse (see Fig. 1f).



Box 3 | Perpendicular magnetic anisotropy

Perpendicularly magnetized junctions have lower switching voltage and current compared with in-plane magnetized junctions, due to the form of their magnetic anisotropy (see the figure, left panel).

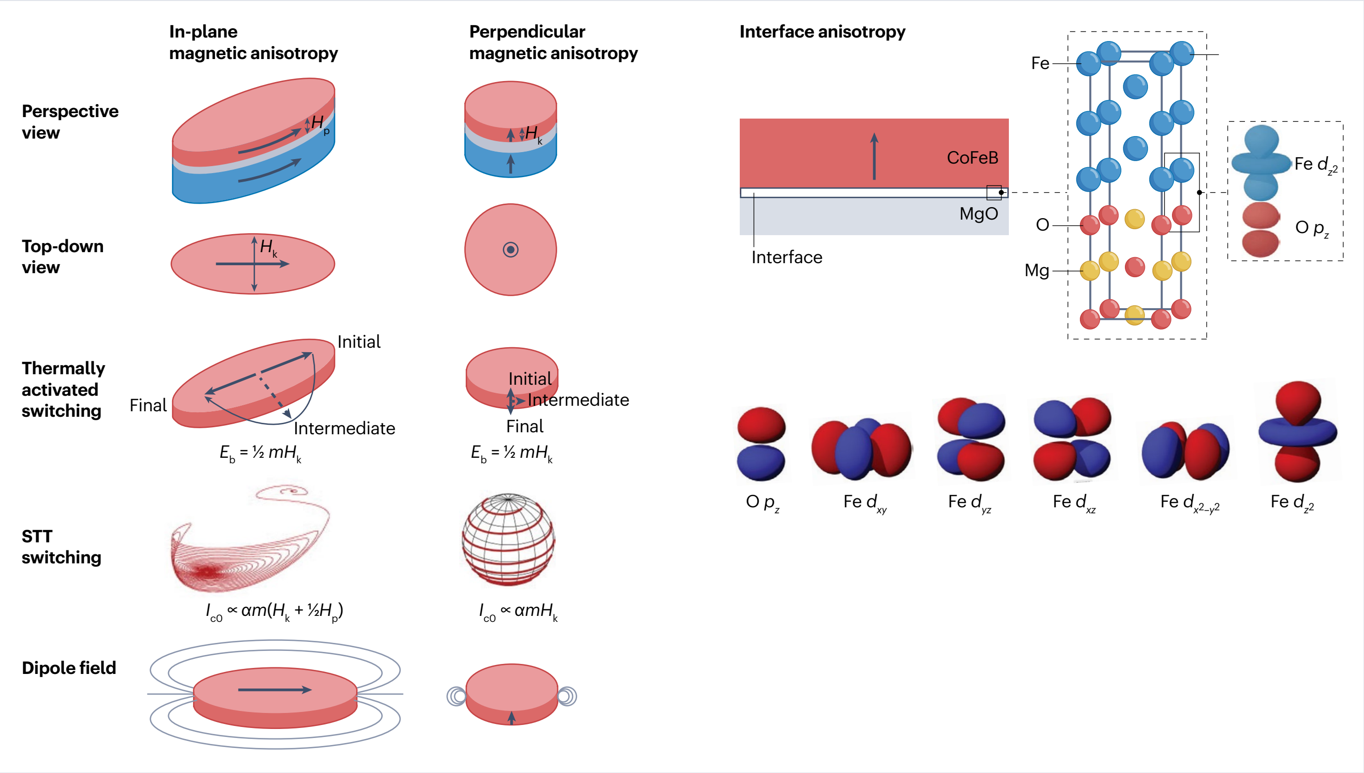
Magnetic shape anisotropy, caused by demagnetizing fields, increases magnetostatic energy when the magnetization lies along shorter dimensions. In-plane free-layer magnetizations lie along the long axis of the ellipse in equilibrium, and require a large field equal to the easy-plane anisotropy field H_p~10 kOe to rotate out of plane, and field equal to the uniaxial anisotropy field H_k~300 Oe to rotate to the short axis of the ellipse.

In perpendicular junctions, perpendicular magnetic anisotropy at the CoFeB/MgO interface (see the figure, right panel) overcomes H_p, resulting in a net anisotropy H_k~5 kOe out of plane. This is due to spin–orbit coupling inducing electronic hybridization between oxygen p_z and iron d_{z^2}, d_{xz} and d_{yz} orbitals, resulting in those states being lower in energy when the magnetization is out of plane versus in plane.

In-plane junction data retention is determined by thermal activation over the smaller in-plane energy barrier determined by H_k, not the larger out-of-plane barrier determined by H_p (see the figure, left panel).

The key to spin-transfer torque (STT) switching is that the small STT steadily amplifies the precession of the magnetization over many precessional orbits (see the figure, left panel). During STT switching of in-plane junctions, the torque must overcome H_k at all points along a precessional orbit, and H_p during the top and bottom parts of the orbit, resulting in a large switching current, $I_c \propto \alpha m(H_k + ½H_p)$, where α is the magnetic damping and m is the free-layer moment. In comparison, perpendicular junctions need only overcome their singular anisotropy field, resulting in $I_c \propto \alpha m H_k$. At the same energy barrier, E_b, in-plane junctions have a substantially higher switching current than perpendicular junctions, due to paying a switching penalty for H_p for which there is no benefit in data retention.

In addition, in-plane junctions are not possible to pack as densely as perpendicular junctions, as dipole fields (which could disturb neighbouring bits) fall off on a length scale related to the distance between magnetic poles (see the figure, left panel). In-plane junctions also have a worse write-error rate and longer switching times, due to larger moments (needed to compensate for smaller H_k, to maintain the same E_b). I_{c0}, threshold switching current.



aggressive etch gases. Care must be taken to avoid redeposition of this material on the sidewall of the tunnel junction, where it can cause an electrical short if it bridges the MgO tunnel barrier. Typically, a sequence of high-angle (near-normal) etches to remove material in the field and low-angle etches to remove redeposited material on the junction sidewall are used to achieve short rates of less than 1 bit per million. LAM[13] and Canon-Anelva[14] offer 300 mm ion beam etch tools, using ion beam etch technology acquired from Veeco and Nordiko, respectively.

Stand-alone applications

Today, two companies offer stand-alone STT-MRAM products: Everspin and Avalanche. Everspin started selling the first STT-MRAM product, a 256 Mb, 40 nm node DDR3 chip, in 2017 (ref. 15), and a 1 Gb, 28 nm node DDR4 chip in 2019 with a 220 nm × 180 nm memory cell[16,17], both manufactured by GlobalFoundries. The specifications of the 1 Gb chip are shown in Table 1. Data retention of 3 months at 70 °C is sufficient for normally-on data centre applications (disaster recovery efforts typically conclude within 3 months, if the data are not destroyed in the disaster). Recently, Everspin started selling 8–64 Mb xSPI STT-MRAM chips with more than 10^{14} write endurance and magnetic field immunity during read/write of 300 Oe (ref. 18). The higher endurance is achieved by greatly reducing the bit density; the total number of bits on the die is increased by roughly ten times over the nominal bit count, and presumably a combination of error correction code and active redundancy is used to handle endurance fails. Also, the resistance–area product of the tunnel barrier is reduced to increase the window between writing and breakdown, and the resulting increase in time-zero shorts is handled by the increase in redundancy[19]. The xSPI chips offer 10-year data retention, at temperatures up to 105 °C. These chips serve as a high-performance replacement for stand-alone NOR Flash, offering 500–1,000 times faster write, 10,000 times lower energy write and 10–100 times higher endurance[18]. Avalanche introduced their highest capacity 32 Mb, 40 nm node chip in 2021, using a 324 nm × 204 nm memory cell with 10^{14} write endurance, 10-year data retention and magnetic field immunity during read/write of 350 Oe (ref. 17). This chip also serves as a NOR Flash replacement, is manufactured by Sony and is also sold under the Renesas brand. Here, the high endurance is also achieved by greatly reducing the bit density.

In general, stand-alone STT-MRAM represents a small market today, due to the low bit density of STT-MRAM. For example, as of October 2023, the largest capacity commercially available stand-alone STT-MRAM is Everspin's 1 Gb (priced around US$100), compared with 128 Gb for DRAM (around US$75) and 2 Tb for NAND Flash (around US$80; prices from www.digikey.com in October 2023), making STT-MRAM 128 times lower capacity than DRAM and 2,000 times lower capacity than NAND Flash at the package level (the DRAM and NAND

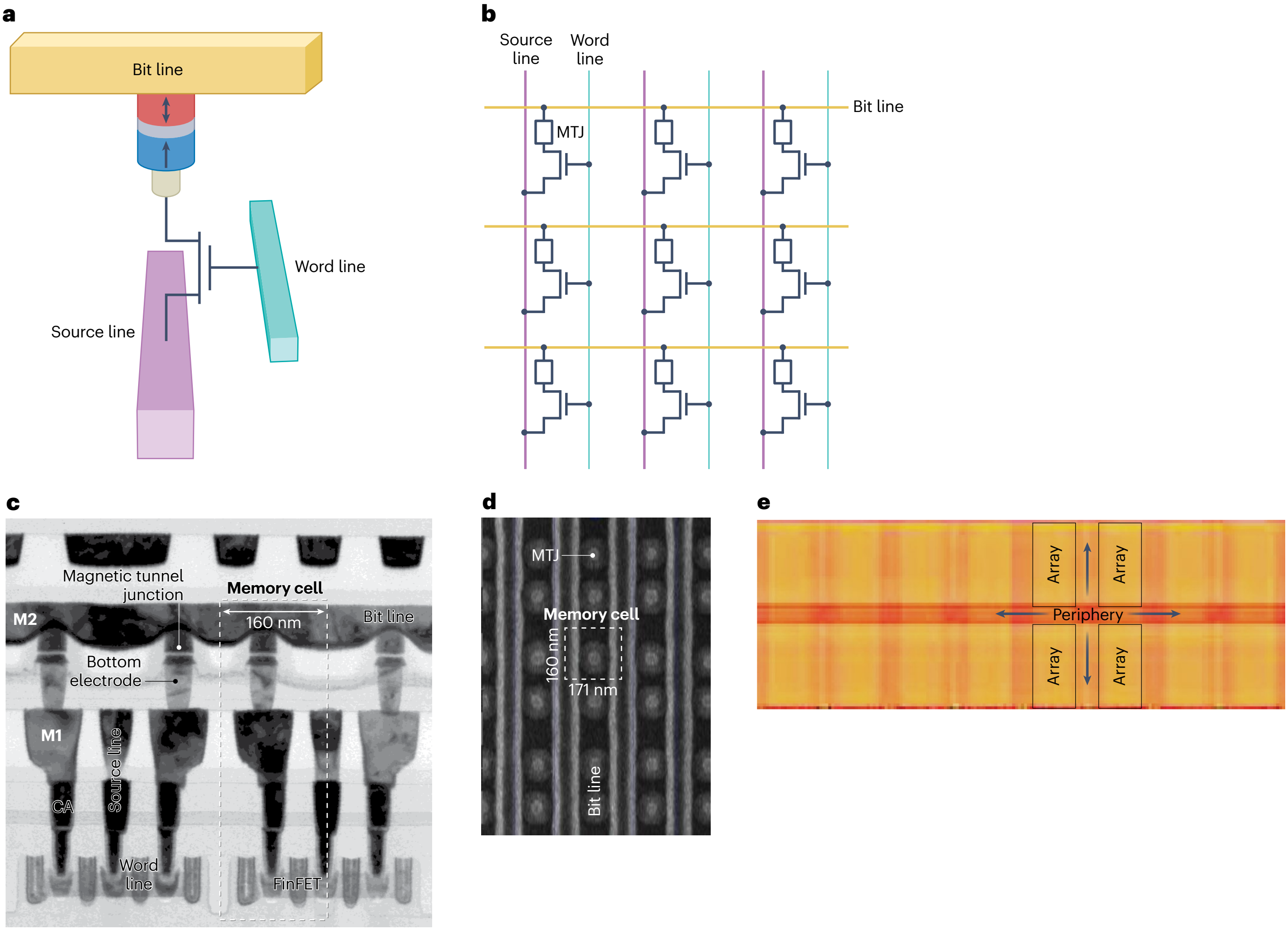


Fig. 2 | The STT-MRAM memory cell and array. a, Each memory cell contains one magnetic tunnel junction (MTJ) and one access transistor. **b**, An array of memory cells has the bit lines orthogonal to the word lines, to allow one bit to be selected for reading or writing. The source lines can run in either direction. **c**, Scanning electron microscope (SEM) cross-section of a 14 nm node spin-transfer torque magnetoresistive random access memory (STT-MRAM) array. In this chip, the source lines run into the page, to provide a dense cell, and the MTJ is embedded between lowest metal levels M1 and M2 to reduce parasitic capacitance. MRAM products today typically have the MTJ on a higher metal level. **d**, SEM top-down view of the 14 nm node STT-MRAM. **e**, In addition to the array, an MRAM chip contains substantial peripheral circuitry, including circuitry for reading and writing.

Table 1 | Requirements and typical specifications for the four main STT-MRAM applications

	Stand-alone	Embedded non-volatile memory	Non-volatile working memory	Last-level cache
Capacity/density	1→4 Gb (9.5→38 Mb mm^{-2})	15→30 Mb mm^{-2}	20 Mb mm^{-2}	2×SRAM (~60 Mb mm^{-2})
Write endurance	10^{10}→10^{12}	10^{6}→10^{8}	10^{12}	10^{17}
Read cycle	135→40 ns	30→10 ns	10 ns	2 ns
Write pulse	30→10 ns	200→50 ns	10 ns	2 ns
Operating temperature	0 to 85 °C→−40 to 105 °C	−40 to 150 °C→−55 to 180 °C	0–85 °C	0–85 °C
Data retention	3 months, 70 °C→10 years, 105 °C	90 s, 260 °C≈10 years, 200 °C→no change	10 years, 85 °C	None needed
Magnetic field immunity	25→500 Oe	300→1,000 Oe	25 Oe	25 Oe

For stand-alone and embedded non-volatile memory applications, the first entry in each cell represents what is commercially available today. The other data in the table are predictions for what will be commercially available in the next 5 years, except for those in the last-level cache column, which represent the long-term goal. STT-MRAM, spin-transfer torque magnetoresistive random access memory.

Flash packages contain multiple die). For most stand-alone applications, the other advantages of STT-MRAMs – such as non-volatility compared with DRAM, or more than 10^4 times faster write speed and more than 10^6 times higher write endurance compared with NAND Flash – are not enough to justify the use of STT-MRAM at the system level. However, IBM used Everspin's 1 Gb STT-MRAM chips in the IBM FlashCore Module, an enterprise-grade solid-state drive, to provide high reliability. The STT-MRAM is used as a write buffer to store data as they are being encrypted and compressed, before being written to NAND Flash. If the power is interrupted, in-progress STT-MRAM writes can be completed quickly using power stored in a small number of tantalum capacitors, without the need for large-area and unreliable super-capacitors or batteries that would be required if a DRAM-based write buffer was used. Another use case is space applications, where the natural radiation hardness of the MTJ enables radiation-hard STT-MRAM[20], with the use of standard radiation-hard CMOS circuit design. To this end, Avalanche proposed an 8 Gb MRAM package, composed of eight 1 Gb chips, with sufficient radiation hardness for use in low earth orbit applications, and potential use in geostationary orbit applications[21].

Short-term future directions for stand-alone STT-MRAM include improving write endurance by 10–100 times (without decreasing density) to relax system wear-levelling requirements, reducing the bit error rate to enable less system-level error correction code overhead and reducing cost. Additional input/output interfaces using the same underlying MTJ technology are another potential direction, for example Everspin's recent expansion beyond DDR4 to xSPI[18], and Avalanche's offering of serial D-QSPI and parallel interfaces. STT-MRAM with improved read schemes has also been proposed. For example, Everspin's 256 Mb and 1 Gb STT-MRAM use a self-reference read scheme, wherein bits are read, then written to the '1' state, read again and if their resistance changed, written back to the '0' state. This change in resistance is used to determine the initial state. This self-reference read relaxes requirements on the magnetoresistance and distribution of resistance from bit to bit. However, it comes at the cost of a longer read cycle, higher read power, limited read endurance (limited by the write endurance) and volatility if the power is interrupted during the read cycle before the bits are written back to the '0' state. In contrast, in a mid-point reference read scheme, the bits are simply read once and their resistance is compared with a reference, set near the mid-point of low and high resistances. The smaller capacity Everspin 64 Mb and Avalanche 32 Mb stand-alone chips already use a mid-point reference read, and future high-capacity chips may adopt it. Samsung demonstrated a 40 ns mid-point reference read, 160 ns write and 10^{14} write endurance on a 28 nm node stand-alone memory chip designed by Netsol, with initial data at the 14 nm node showing 15 ns read and plans for 50 ns write[22].

Although it is unlikely that STT-MRAM will replace DRAM, owing to the continued density scaling of DRAM, multiple research and development programmes aim at developing denser stand-alone STT-MRAM. Kioxia and Hynix have had a long-standing partnership to develop a 4 Gb stand-alone STT-MRAM, using a 90 nm × 90 nm (9 F^2) memory cell[23]. They demonstrated fewer than five errors in a write/read test of the 4 Gb chip, after redundancy repair and with 1-bit error correction code turned on. Recently, Kioxia demonstrated 14 nm diameter bits with good data retention of 10 years at 90 °C, with initial data showing 5 ns switching on individual bits, using a CoFeB/CoPt free layer[24].

Western Digital recently launched a project to develop STT-MRAM for storage-class memory applications. Dense storage-class memory requires storing multiple bits per cell, which could potentially be achieved in STT-MRAM by using multiple layers of MTJs and two-terminal back-end selectors. This is a challenging approach, because the back-end selector must not limit the endurance or write speed of the STT-MRAM, which could prevent the use of selectors involving atomic motion during switching. For example, Avalanche reported initial results using a doped-HfO$_x$ selector[25] with an on/off ratio of 10^7. Kioxia and Hynix reported STT-MRAM arrays with 4 F^2 cells at 45 nm pitch and 20 nm diameter MTJs, using an arsenic-doped SiO$_2$ back-end selector[26]. A single cell was written 1,000 times with a write pulse in the range 30–200 ns. A major challenge at these tight pitches is to etch away the material between the junctions without redepositing it on the sidewall of the junctions. Using ion beam etch, Western Digital demonstrated 22 nm diameter junctions on a 50 nm pitch, with electrical measurements showing that individual junctions were not shorted[27].

Embedded non-volatile memory

STT-MRAM for eNVM has recently replaced embedded Flash at all advanced nodes in foundries. This type of memory, used to store code and data in microcontroller units, is ubiquitous in electronic equipment, ranging from automobiles and factory robots to appliances, and everything in between. Traditionally, embedded NOR Flash (eFlash) was used for eNVM, but it became too expensive to develop and manufacture, and so all advanced foundries developed eMRAM as an alternative starting at the 28 nm node, where they offered both eMRAM and eFlash. eFlash required expensive development of specialized floating gate transistors, in addition to the standard logic transistors, at every new node, plus the use of high voltages for writing. Manufacturing the floating gate transistors required more than 12 additional masks, with even more masks predicted at future nodes. By comparison, once the MTJ for eMRAM is developed, it can be used at any node, can be fabricated with only two or three additional masks and can be written with standard transistors and voltages. For this reason, eFlash will not be offered below the 28 nm node at any foundry, and all advanced applications will use eMRAM.

Samsung was the first to enter volume manufacturing of eMRAM, in early 2019 at the 28 nm node, demonstrating a technology operating in the range −40 °C to 125 °C, with write endurance of 10^6 writes, solder reflow retention at 260 °C for 90 s and magnetic field immunity during write of 550 Oe (refs. 28,29). Commercial applications include a Sony global positioning system receiver chip containing 8 Mb of eMRAM with a 190 nm × 190 nm memory cell used in smartwatches, including the Huawei GT2 (ref. 17). The ultra-low power of the STT-MRAM helps the smartwatch have a 2-week battery life, compared with a few days for previous models. Samsung demonstrated a denser 28 nm node eMRAM (13.9 Mb mm^{-2}) for use as a frame-buffer memory in CMOS image sensor applications, a very different application to eNVM, with 10^{10} write endurance over the operating range −20 °C to 85 °C, and a fast write of 50 ns relative to 200 ns for eNVM for standard eFlash replacement. The trade-off is that only 1 s retention at 85 °C is needed for this application[30]. Sony has also demonstrated a 40 nm node eMRAM for use as a buffer memory for CMOS image sensor applications[31]. Samsung is now developing eMRAM for eNVM at the 14 nm node, with a focus on automotive applications[32]. Samsung has demonstrated more than 90% yield at the 14 nm node, with an operating temperature range of −40 °C to 125 °C and 10^6 write endurance, and promising results for both 160 °C operation (11 ns read, 200 ns write) and a sub-10 nm node technology[32]. Samsung also demonstrated 18.1 Mb mm^{-2} density at the 14 nm node, with an operating range of −40 °C to 150 °C (ref. 33). Although Samsung offered both eFlash and eMRAM at the 28 nm node, at the 14 nm node and below eFlash will no longer be offered, and only eMRAM will be offered. Samsung recently announced plans for eMRAM manufacturing at the 14 nm node in 2024, 8 nm in 2026, and 5 nm in 2027.

TSMC manufactures eMRAM at the 22 nm node[34], using a 220 nm × 210 nm memory cell, for use in Ambiq's Apollo4 Blue system-on-chip for low-power applications, used in the Fitbit Luxe[17]. Available with up to 16 Mb of eMRAM, the Apollo4 provides low-power compute for battery-powered edge devices, including controlling the sensors on the Fitbit Luxe. Renesas used TSMC's 22 nm process to demonstrate 5.9 ns read access time at 150 °C in a 32 Mb eMRAM macro for use in microcontroller units, using a 500 ns write/verify scheme in which bits are first written with a 250 ns write pulse, then read to verify their state and then written a second time with a higher voltage 250 ns write pulse if needed[35]. NXP has announced that TSMC will manufacture their latest automotive microcontroller unit, the S32 automotive processor, using eMRAM at the 16 nm node[36]. The S32 is a substantial processor with four cores that will be used as a zonal controller to enable software-defined vehicles requiring regular over-the-air software updates. This function was not possible with eFlash, due to the long write times (~100 μs) resulting in unacceptably long software update times. With eMRAM, software updates will download in only a few seconds, due to the faster (5.5 μs) write cycle, which uses a write/verify scheme[37]. TSMC will not offer eFlash below the 28 nm node. In addition to eMRAM, TSMC is offering embedded RRAM at advanced nodes, as a lower cost, lower performance alternative to eMRAM. The embedded RRAM has six times lower bandwidth and ten times lower write endurance than eMRAM.

GlobalFoundries has also developed eMRAM at the 22 nm node, and does not plan to offer eFlash below the 28 nm node. GlobalFoundries demonstrated a 40 Mb, 22 nm node eMRAM macro operating in the range −40 °C to 125 °C, with 10^6 write endurance and 260 °C, 90 s solder reflow compatibility, plus promising results for future qualification at 150 °C (ref. 38). GlobalFoundries manufactures an ultra-low-power Sony global navigation satellite system receiver, containing 16 Mb of 22 nm node eMRAM using a 224 nm × 208 nm memory cell, for use in wearables and vehicle tracking. Using their 22 nm eMRAM technology, GlobalFoundries demonstrated 10-year standby magnetic immunity (while not reading or writing) to more than 1,500 Oe external magnetic fields[39]. In addition to eMRAM, GlobalFoundries is also developing eRRAM, licensed from Renesas, as a lower cost but lower endurance and lower bandwidth alternative to eMRAM at advanced nodes.

Intel demonstrated a 7.2 Mb, 22 nm node eMRAM with a 216 nm × 225 nm memory cell, 10^6 write endurance, 10 ns read and operation up to 105 °C (ref. 40). UMC has partnered with Avalanche to develop eMRAM at the 28 nm node.

eNVM requires several features beyond what are required for stand-alone MRAM. As the eNVM macro is fabricated on the same chip as other circuit blocks using the same process flow for the metal levels, no customization of the metal levels in the MRAM array is allowed. This means that the MTJ must fit vertically in between two standard metal levels. At advanced nodes below 14 nm this presents a challenge, so care must be taken to minimize the thickness of the MTJ stack and the metal hard mask used to etch it. One option at very advanced nodes is to skip a metal level inside the MRAM array, and have the MTJ fit vertically between, say, metal levels 1 and 3. In addition, all embedded applications also require that the MTJs are compatible with the standard CMOS back-end-of-line processing temperature of 400 °C, with exposure for at least an hour. This requirement was first satisfied in the demonstration by TDK[41]. Finally, in comparison with stand-alone applications, the yield requirements for embedded applications are much stricter, as the MRAM circuit is only a small fraction of the entire chip area, and potentially several other foundry technologies are also included in the same chip. Hence, all of the technologies must individually have high yields, so that the overall yield is sufficient. This places strict requirements on both time-zero fails, the fraction of bits in the array that fail when the chip is tested at the beginning of its life, and reliability fails during the lifetime of the chip. Time-zero fails are typically shorts, either due to pinholes in the MgO barrier or redeposition during ion beam etching of the junction. Foundries have improved time-zero fails to sub-parts per million levels by careful optimization of both the MgO deposition process and the junction etch process[32,36,39].

The majority of eNVM applications require solder reflow retention, meaning that the eMRAM must retain data when the packaged chip is soldered to a board, typically at 260 °C for 30 s, with up to three solder attempts. Developing eMRAM that stores data for 90 s at 260 °C requires thicker free layers, larger-area junctions and larger write currents. The resulting eMRAM also reliably stores data for 10 years at temperatures around 200 °C. Solder reflow retention allows data (for example, chip ID, redundancy addresses, trim settings and program code) to be stored in the eMRAM inexpensively at wafer-level test, without the need for tracking each chip during dicing and packaging, and expensive programming at board level. When higher temperature retention is needed in a small number of bits, individual MTJs can intentionally be electrically broken down by applying ~ 2 V to permanently create a short in the MgO tunnel barrier, enabling a one-time programmable memory[42].

eNVM applications require a wide operating temperature range, for example from −40 °C to 125 °C or 150 °C for automotive applications. At high temperatures, reading becomes a significant challenge due to the drop in magnetoresistance. This means that tighter resistance distributions, higher magnetoresistance and a lower temperature-dependence of magnetoresistance are required (mid-point reference read is used for eNVM applications, to provide fast read times similar to eFlash). Writing, on the other hand, is more challenging at lower temperatures because the switching voltage increases slightly faster than the breakdown voltage. However, even considering the higher E_b required for retention, writing is significantly easier than for stand-alone applications, due to greatly relaxed requirements on write endurance and write pulse width for eFlash replacement.

A large externally applied magnetic field can erase information in an STT-MRAM chip, or prevent it from being accurately read or written. Magnetic field immunity, or magnetic immunity, refers to the largest field the chip can reliably withstand. This metric is typically reported for operating conditions (active reading and writing) and is considerably higher on standby[43]. Despite some user concerns, magnetic immunity is not a practical limitation for almost all applications. This is because magnetic fields encountered in common situations are small and drop off rapidly with distance. For example, the magnetic field 1 mm from the tip of a magnetic screwdriver is well below 250 Oe (1 mm represents the thickness of a package separating the eMRAM chip from the screwdriver tip); the field 2 cm away from the end of a fairly large cylindrical NdFeB magnet (1 cm diameter and 4 cm long) is less than 200 Oe; and the field 1 cm away from a wire carrying 100 A is 20 Oe. Although Toggle MRAM[4] has a magnetic field immunity of only 100–150 Oe, it has established itself as a highly reliable product for more than 17 years in various harsh environments in industrial, military and space applications. Recent STT-MRAM products have even higher magnetic field immunity. Although Everspin's 256 Mb and 1 Gb parts have 25 Oe magnetic immunity, the Everspin xSPI and Avalanche parts have 300 Oe and 350 Oe active immunity, respectively. eMRAM for eNVM is offered with an active magnetic immunity of more than 300 Oe, and magnetic shielding in the package can substantially increase the immunity[44–46]. Another strategy for enhancing magnetic immunity is to develop new free layers with larger magnetic anisotropy than the typical 0.5 T used today. For example, perpendicularly magnetized Heusler alloys[47] can have magnetic anisotropy of the free layer $H_k > 7$ T. Samsung[28] and TSMC[44] have already demonstrated 550 Oe active magnetic immunity, and this is expected to improve further over the next few years. However, many applications do not require high magnetic immunity. For example, for automotive applications, the eMRAM chip can be easily placed a few centimetres away from any sources of large magnetic fields and the exterior of the car. The fact that hard disk drives have been widely used for more than 60 years with typical magnetic field immunity of only 5 Oe shows that concerns over STT-MRAM magnetic field immunity are not technically substantiated.

Non-volatile working memory

Non-volatile working memory is a potential killer application for STT-MRAM where it can replace both SRAM and eFlash in embedded applications[48–50]. Initially, 5–10 ns reads and writes may be used in extremely power-sensitive, normally off applications, where performance is not critical. SRAM and eFlash are not suitable for an ultra-low-power, normally off device if data must be frequently written to the eFlash, owing to its high write energy resulting from the high voltage and long write pulses. In the case of SRAM and eFlash, on wake-up the device would need to write data into the SRAM, complete its operating task and then write data back to the eFlash before powering down. In contrast, with an eMRAM working memory, the device can simply wake up, operate and then power down, without the need for writing data back and forth between cache and storage. Furthermore, eMRAM offers greatly reduced standby power dissipation compared with SRAM, affected only by the leakage of peripheral circuitry. This type of non-volatile working memory opens entirely new types of ultra-low-power applications and devices which were not possible before. Applications may include wearables, implantables, co-processors, Internet of Things and edge devices – where low power is the dominant factor over high performance. For example, a cell phone could have a low-power eMRAM-based co-processor, which runs when the phone is in a sleep mode, in addition to its usual high-performance and power-hungry processor. The co-processor monitors various inputs, including listening for audio instructions, which can prompt it to wake up the main processor, thus greatly extending battery life.

Although no commercial non-volatile working memory is available today, the specifications are almost within reach using existing CoFeB-based materials, and initial products may be expected in the next few years. Very few data showing reliable writing at and below 10 ns have been published to date. Intel explored STT-MRAM for cache applications and demonstrated a good write-error rate curve of a single bit down to 1×10^{-4} errors per write using 10 ns write pulses[51]. Avalanche demonstrated initial results of array-level writing using 10 ns write pulses on a 22 nm node test vehicle fabricated at UMC[52]. Steep write-error rate curves on a single bit using 10 ns write pulses, down to a write-error rate floor of 10^{-6}, were also reported[53]. Kioxia demonstrated 5 ns switching of a 14 nm diameter bit, with the write-error rate plotted on a linear scale[24]. IMEC published data on a thousand junctions written with 5 ns pulses[54], with many junctions reaching the write-error rate floor of 10^{-4}. Toshiba demonstrated reliable writing with 3 ns pulses on a single junction[55] down to a write-error rate of 10^{-4}. TDK showed an 8 Mb array written once without errors using 3 ns write pulses[56], corresponding to an array-level write-error rate floor of 1.2×10^{-7}. The IBM–Samsung MRAM Alliance presented reliable writing with 2 ns write pulses on hundreds of junctions[57], with steep write-error rate slopes and tight bit-to-bit distributions, down to a write-error rate floor of 10^{-6}.

Last-level cache

The most challenging goal for STT-MRAM is to make it fast and dense enough for use as last-level cache in high-performance applications, in place of SRAM. Here, last level means L3 cache, or L4 cache in some high-performance systems. eDRAM has been used as last-level cache in high-performance systems since the 90 nm node; however, the challenge of scaling the deep trench capacitor at advanced nodes has resulted in discontinuation of its use below the 14 nm node. In today's advanced processor chips, if data cannot be found in the SRAM cache, the system must go to off-chip stand-alone DRAM to retrieve the data, a round-trip cost of roughly 50 ns. If, instead, some of the SRAM cache was replaced with eMRAM which would be twice as dense, fewer uses of the stand-alone DRAM would be required. Even at eMRAM's read and write cycle times of 2–3 ns, slower than sub-nanosecond SRAM times, the advantage at the system level would still be significant.

Despite the stringent requirements for last-level cache (Table 1), great progress has been made over the last 5 years in demonstrating fast and highly reliable switching, as well as fast reading. Fast reading down to 4 ns has already been demonstrated[51]. Further improvements can be made by increasing magnetoresistance, decreasing resistance distributions and designing smaller arrays to reduce the resistor-capacitor time constant for charging read lines. As mainstream cache applications are read-intensive in nature, reading at least as fast as writing is required.

Additionally, STT-MRAM must be at least twice as dense as SRAM at the macro level. The macro is composed of the array of bits and the peripheral circuitry (Fig. 2e). STT-MRAM entails more peripheral circuitry compared with SRAM, due to the small read signal, which necessitates the use of multiple sense amplifiers. Taking this into account, the STT-MRAM cell must be at least three times denser than that of SRAM to provide the desired two times density improvement at the macro level. For that, the STT-MRAM switching current must be reduced by about two times. In addition to solving the density problem, this will also solve the write endurance problem, as reducing the write current automatically reduces the write voltage.

The IBM–Samsung MRAM Alliance has demonstrated reliable 500 ps STT-MRAM switching in a hundred junctions with tight bit-to-bit distributions down to a write-error rate of 10^{-6}, and initial results at 250 ps (ref. 58). This was achieved using a double spin-torque MTJ that had a second reference layer above the free layer, separated by a non-magnetic, low-resistance spacer, so that the free layer received STT from both top and bottom interfaces. By doubling the torque, the switching current was reduced by a factor of two. Due to the use of a low-resistance spacer instead of a second tunnel barrier, the magnetoresistance was not diluted. Intriguingly, theory predicts that a factor of three or four reduction in switching current may be realistically possible with slightly higher spin polarization[59]. With further improvements in switching efficiency, activation energy and magnetoresistance, the double spin-torque MTJ may be a viable path to last-level cache.

Future directions

Spintronics devices utilizing switching mechanisms other than STT have been explored extensively for memory applications and beyond[60–64]. In this section we discuss the device and technology aspects of these approaches, targeting last-level cache applications.

Spin–orbit torque MRAM

Among the various alternative spintronics device concepts, spin–orbit torque (SOT) MRAM (Fig. 3) is the most studied and mature. In an SOT-MRAM device, the MTJ is written by passing an electrical current through a metal SOT wire underneath the free layer (Fig. 3a). The magnetization of the free layer is manipulated by STT originating from the spin–orbit interactions in the adjacent SOT material[65,66]. The read operation is performed by passing the current through the MTJ, as in STT-MRAM.

Compared with STT-MRAM, SOT-MRAM devices are expected to show improved endurance, read disturb, switching speed and switching energy. In SOT-MRAM devices, the read and write paths are separated, so that the endurance issue of STT-MRAM, related to MgO barrier breakdown, is mitigated (Fig. 3a,b). This type of device is expected to operate at a higher speed with a larger write current, compared with STT-MRAM. The three-terminal structure also addresses the read disturb problem found in STT-MRAM, when the devices can be accidentally written during the read process.

For STT and SOT-MRAM devices, the free-layer magnetization is switched by the spin current generated from the charge (electrical) current. A key metric that determines the device switching efficiency is the charge current to spin current conversion efficiency. In an STT device, the same electrons in the charge current passing through the MTJ form the spin current; the conversion efficiency is determined by the spin polarization of the reference layer material, P_{Ref}, which is no greater than 100%:

$$I_{\mathrm{spin}} = \frac{\hbar}{2e} P_{\mathrm{Ref}}\, I_{\mathrm{charge}}. \tag{1}$$

Here, $\hbar/2e$ simply converts units from charge to spin. In an SOT device, the spin current density flowing vertically into the free layer is proportional to the charge current density flowing laterally through the SOT wire:

$$J_{\mathrm{spin}} = \frac{\hbar}{2e} T_{\mathrm{int}} \Theta_{\mathrm{SH}}\, J_{\mathrm{charge}} = \frac{\hbar}{2e} \xi\, J_{\mathrm{charge}}, \tag{2}$$

where Θ_{SH} is the spin Hall ratio (also called the spin Hall angle), $T_{\mathrm{int}} \leq 1$ is the interfacial spin transparency, which takes into account spins that are reflected or flipped at the interface between the SOT wire and the free layer, and $\xi = T_{\mathrm{int}}\Theta_{\mathrm{SH}}$ is the SOT efficiency. There are two factors that can increase the charge current to spin current conversion efficiency of SOT devices with respect to STT devices. First, when converting the current densities (Eq. 2) to currents, as the charge and spin currents flow through different areas, SOT devices pick up a geometric advantage of d/t, where d is the lateral dimension of the free layer and t is the thickness of the SOT wire, with d on the order of 50 nm and t on the order of 5 nm. Hence, each electron in the charge current can contribute to the spin current multiple times as it passes laterally under the free layer. Second, Θ_{SH}, and hence ξ, can be larger than one, due to intrinsic band-structure effects. For example, $\xi > 2$ was demonstrated at room temperature in the topological insulators $\mathrm{Bi_2Se_3}$, $\mathrm{Bi_xSe_{1-x}}$ and BiSb[67–69], and $\xi > 100$ was reported at low temperature in a topological insulator bilayer[70]. These two effects make SOT-MRAM a promising candidate to achieve more energy-efficient switching than STT-MRAM.

However, to date there has been no clean demonstration of the superiority of SOT-MRAM over STT-MRAM in terms of switching speed or switching efficiency[58]. Today's SOT devices suffer from impractical materials, large switching currents due to inefficient spin polarization direction, large memory cells due to multiple and large transistors, and processing challenges.

Large SOT efficiency has only been observed in unconventional materials, including topological insulators and two-dimensional materials, which have high resistivity and are not compatible with the standard CMOS process. Technology-related demonstrations have been limited to conventional heavy metal SOT materials, including platinum, tantalum and tungsten, where ξ is in the order of 0.1–0.4 (ref. 71).

Conventional polycrystalline SOT materials can only generate in-plane spin polarization (Fig. 3d,e), constrained by symmetry (Box 4), which limits the switching efficiency. For SOT devices with in-plane magnetized tunnel junctions (Fig. 3d), the polarization of the spin current is collinear with the magnetization of the free layer, and so the spin torque only needs to overcome the damping torque to induce switching, as in STT[72,73]. However, due to the in-plane magnetic anisotropy, the critical switching current is proportional to $\alpha(H_{\mathrm{k}} + \tfrac{1}{2}H_{\mathrm{p}})$, where α is the magnetic damping constant with a typical value of 0.01, H_{k} is

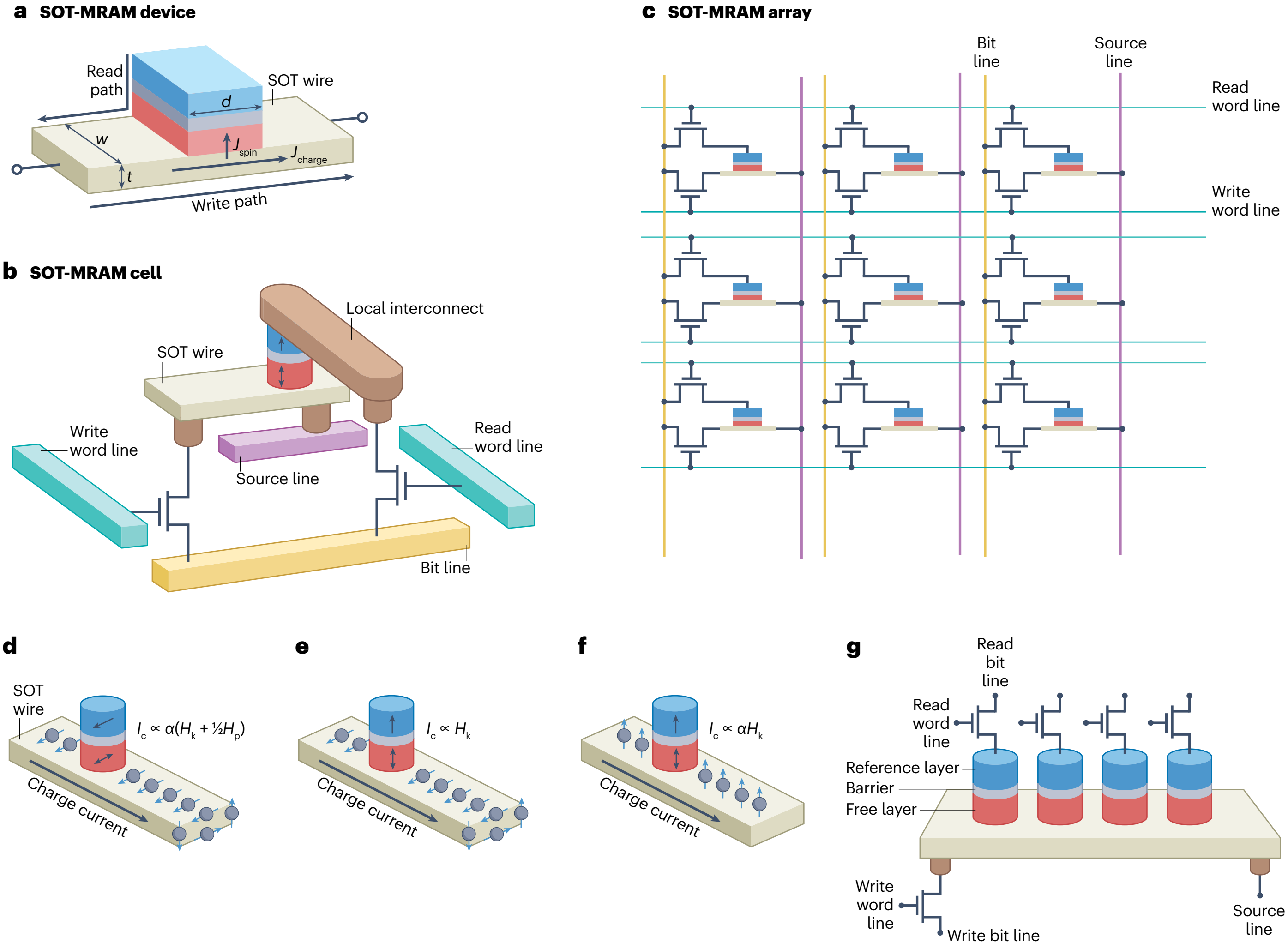


Fig. 3 | The SOT-MRAM memory cell and array. a, Three-terminal spin–orbit torque magnetoresistive random access memory (SOT-MRAM) device, simplified so that the free layer has the same width, *w*, as the SOT wire. A lateral charge current generates a vertical spin current. **b**, Example of an SOT-MRAM memory cell with one magnetic tunnel junction (MTJ) and two access transistors. **c**, Example of an array of SOT-MRAM memory cells with the bit lines orthogonal to the word lines, to allow one bit to be selected for reading or writing. **d**, In-plane magnetized SOT-MRAM device with conventional SOT material generating in-plane polarized spins. The switching current is large, as it is proportional to $\alpha(H_k + ½H_p)$. **e**, Perpendicularly magnetized SOT-MRAM device with conventional SOT material generating in-plane polarized spins. The switching current is large, as it is proportional to H_k. **f**, Perpendicularly magnetized SOT-MRAM device with unconventional SOT material generating perpendicularly polarized spins. The switching current, I_c, is small, as it is proportional to αH_k. **g**, Multiple SOT-MRAM devices on a shared SOT wire, switched using a combination of SOT plus either spin-transfer torque (STT) or voltage-controlled magnetic anisotropy (VCMA). α, magnetic damping constant; d, lateral dimension of the free layer; H_k, magnetic anisotropy of the free layer; H_p, easy-plane anisotropy field; t, thickness of the SOT wire.

the magnetic anisotropy of the free layer and H_p is the large easy-plane anisotropy field (see Box 3). For perpendicularly magnetized tunnel junctions (Fig. 3e), the polarization of the incoming spins is orthogonal to the magnetization of the free layer. As such, the spin torque must overcome the full H_k value of the free layer to induce switching. Therefore, the critical switching current is proportional to H_k, making the switching current large[73]. If asymmetric materials are used to generate perpendicular spin polarization (see Box 4) in perpendicularly magnetized tunnel junctions, then the polarization of the spin current is collinear with the magnetization of the free layer (Fig. 3f), and the spin torque only needs to overcome the damping torque. Hence, the switching current is proportional to αH_k, resulting in efficient switching. To be technologically useful, SOT devices have to be sufficiently dense, which will likely require perpendicular spin polarization and perpendicularly magnetized tunnel junctions (Fig. 3f).

Both in-plane SOT-MRAM and STT-MRAM devices have the same disadvantages in terms of switching efficiency, density and scaling potential compared with perpendicular devices (Box 3). Due to the limitations of SOT-MRAM devices with in-plane magnetization, perpendicular SOT devices have been studied extensively, despite the large switching current needed to overcome the full H_k value of the free layer. For perpendicular SOT devices, an external in-plane field is required

Box 4 | Symmetry for spin–orbit torques

A simple symmetry argument

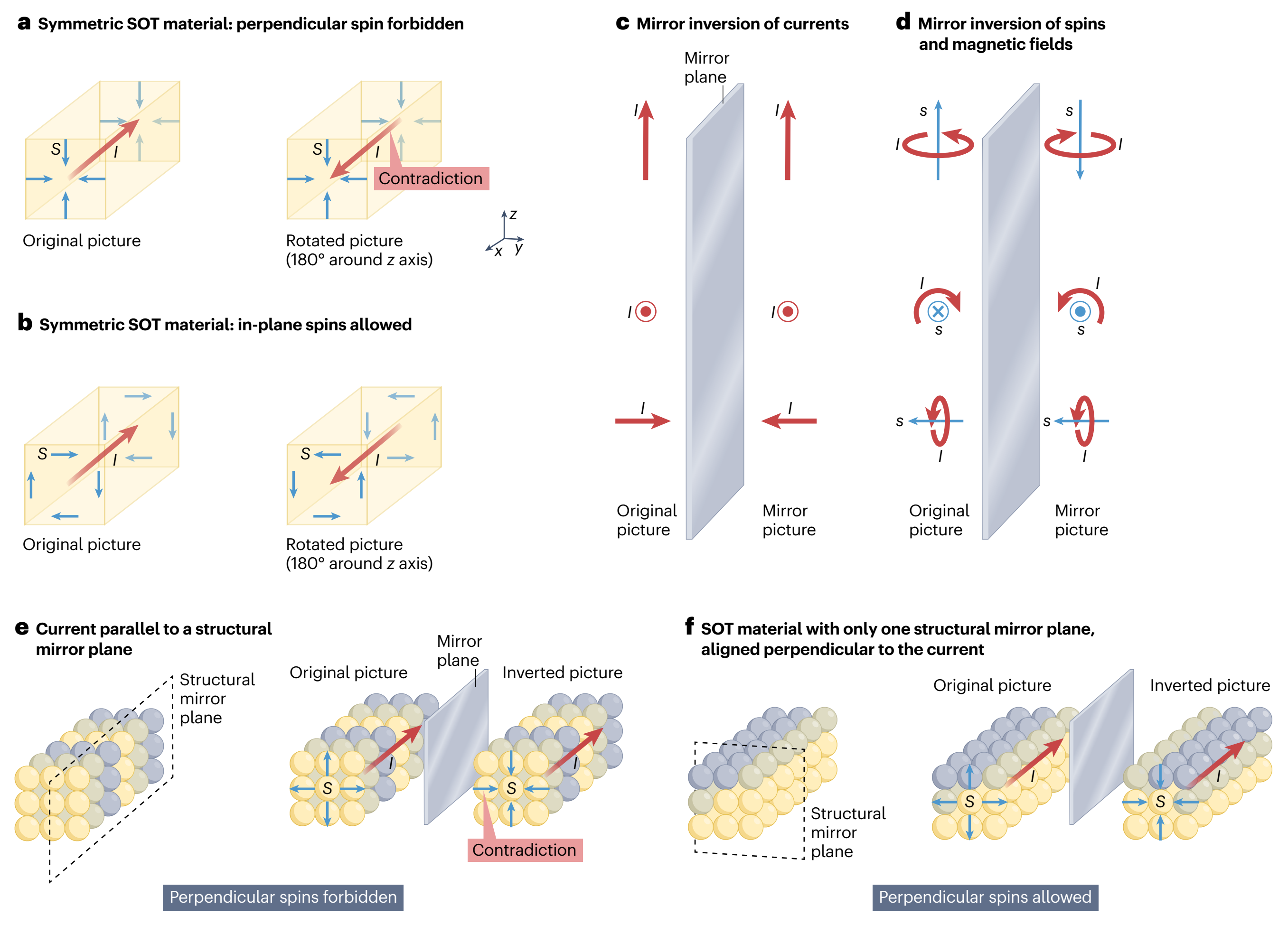
Symmetry can be a powerful tool for ruling out potential physical effects. All of physics is obviously invariant under rotations (consider walking around to the other side of the laboratory bench and observing the same experiment). Therefore, if a picture of a hypothesized effect and a picture of the same effect from a rotated viewpoint contradict each other, the effect is forbidden. (Note that symmetry cannot be used to prove an effect will exist or be large).

Spin–orbit torque (SOT) materials with high structural symmetry (including amorphous or fine-grain polycrystalline materials) cannot generate perpendicular spins at the top interface, where the magnetic tunnel junction (MTJ) is located (see the figure, panel **a**). This can be seen by observing that the original picture and the rotated picture predict the same directions for those spins, for oppositely directed currents (here we assume the spins must reverse when the current reverses). In-plane spins at the top interface are not forbidden by rotation symmetry (see the figure, panel **b**).

More powerful symmetry arguments

All physics except that involving the weak nuclear force is invariant under parity inversion, defined as the inversion of the coordinates of all particles: (*x*,*y*,*z*)→(−*x*,−*y*,−*z*). Hence, if we draw a picture of a hypothesized spintronic effect and a parity inverted picture, physics must work the same way in both pictures. Parity inversion is equivalent to mirror inversion through a plane followed by a 180° rotation in that plane, for example, mirror inversion in the *xz* plane, (*x*,*y*,*z*)→(*x*,−*y*,*z*), followed by a 180° rotation around the *y* axis, (*x*,*y*,*z*)→(−*x*,*y*,−*z*). As physics is invariant under rotations, we only need to check the mirror-inverted picture for a contradiction.

Whereas the impact of mirror inversion on a current is straightforward (see the figure, panel **c**), to understand the impact of mirror inversion on a magnetic field, consider the electrons in a hypothetical current loop that creates the magnetic field (see the figure, panel **d**). If the field is parallel (perpendicular) to the mirror, then the movement of the electrons in the current loop, and hence the magnetic field, are (are not) reversed in the mirror. The same mirror inversion rules apply to magnetic moments and spins.



a Symmetric SOT material: perpendicular spin forbidden

b Symmetric SOT material: in-plane spins allowed

c Mirror inversion of currents

d Mirror inversion of spins and magnetic fields

e Current parallel to a structural mirror plane

f SOT material with only one structural mirror plane, aligned perpendicular to the current

(continued from previous page)

Mirror inversion symmetry shows that perpendicular spins are forbidden if the current is applied parallel to a structural mirror plane: the spin directions in the original and inverted pictures contradict each other (see the figure, panel **e**).

However, perpendicular spins are not forbidden in sufficiently asymmetric SOT materials, for example, materials possessing only one structural mirror plane, aligned orthogonal to the current direction (see the figure, panel **f**). The inverted and original pictures can be seen to be consistent: rotate the entire inverted picture 180° around the z axis to recover the original crystal structure, and then reverse the current direction, which reverses the spin directions; the picture is then identical to the original picture. Asymmetry is essential for SOT to create perpendicular spins.

The Curie principle provides an even more powerful symmetry argument, based on any symmetry of the system[155].

to break the symmetry and enable deterministic switching. This adds new challenges to the technology development of perpendicular SOT devices. To this end, field-free switching of perpendicular SOT devices was achieved through incorporation of exchange bias in the SOT stack[74], utilizing a magnetic hard mask[75] and combining STT with SOT[76–78]. Despite the demonstration of deterministic field-free switching in perpendicular SOT devices, all approaches add complexity and, potentially, new failure mechanisms to device operation. Although both in-plane SOT[79–82] and perpendicular SOT[83–85] devices are being actively pursued, it is unlikely for in-plane devices to be technologically useful, due to inefficient switching (owing to large H_p) and large dipole fields that prevent them from being densely packed (see Box 3). Ultimately, combining perpendicularly spin-polarized spin current with perpendicularly magnetized SOT devices is the most promising path to harness the potential of SOT-MRAM. This has led to great efforts in pursuing spin current generation with perpendicular spins. The challenge is that fundamental symmetry considerations forbid perpendicular spin polarization in symmetric materials (see Box 4). Some form of asymmetry is required. There have been demonstrations of spin current with perpendicular spins in several types of asymmetric SOT materials systems, including WTe_2 and $MoTe_2$ with low crystal symmetry[86–88], epitaxial design of moderate crystal symmetry IrO_2 (ref. 89), magnetic asymmetry in antiferromagnetic RuO_2 (ref. 90) and magnetic asymmetry in a ferromagnet[91–93].

For memory applications, the biggest disadvantage of three-terminal SOT-MRAM devices is their low density relative to two-terminal STT-MRAM devices. Based on a design technology co-optimization study from IMEC[94], the optimized bit cell area of SOT-MRAM is roughly twice the size of STT-MRAM at the 5 nm node, due to the second transistor and an additional via in each memory cell connecting the top of each MTJ with the read transistor. Approaches to address the density challenge of SOT-MRAM devices include two-terminal SOT devices that combine the STT and SOT effects[95], and SOT devices with a shared SOT wire, to amortize the write transistor over multiple junctions[96] (Fig. 3g). The results from both approaches are preliminary at this point and more work is needed to demonstrate their technological potential.

The fabrication process of SOT-MRAM is also more challenging than that of STT-MRAM, due to stringent requirements on the MTJ etch. The etch stop needs to be controlled within the SOT layer, with a typical thickness of ~5 nm. Under-etched devices will have an extended free-layer area with compromised switching performance, whereas over-etched devices will have high-resistance SOT wires. To date, most publications on SOT-MRAM focus on single-device performance without addressing the fabrication challenges of large arrays across a wafer. To date, only Intel and TSMC have reported yield learning data on SOT-MRAM arrays[77,79].

For SOT-MRAM to become a competitive technology for last-level cache applications, the gaps need to be addressed, including generation of perpendicular spins through materials and device structure innovation with $\xi > 0.3$ (ref. 97), demonstration of comparable density with two-terminal STT-MRAM and demonstration of array-level performance with tight distribution and decent yield.

Voltage-controlled magnetic anisotropy MRAM

Voltage control of magnetism has been realized in many materials systems with different characteristics and mechanisms[62,63,98,99]. Among them, the discovery of voltage-controlled magnetic anisotropy (VCMA) in 3d metal (Fe, CoFe, CoFeB)|MgO systems has had the most impact on memory applications[100–102]. Here, we focus on MRAM devices that utilize the VCMA effect in CoFe(B)|MgO-based MTJs.

VCMA-MRAM adopts the same memory cell architecture as STT-MRAM (Fig. 2a) but with high resistance–area product (>100 Ω·μm^2) MTJs. The read process is the same as that of STT-MRAM. When a large voltage is applied during the write process it modifies the magnetic anisotropy of the free layer and induces switching, with minimal electrical current flowing through the MTJ. In VCMA-MRAM, when a voltage is applied across the MTJ, the electric field across the MgO tunnel barrier modulates the orbital occupancy of the electrons at the CoFeB|MgO interface, thus modulating the interface perpendicular magnetic anisotropy[63,101] (Box 3). Two opposite voltage polarities either enhance the interface anisotropy or weaken it, allowing for magnetization to be switched from perpendicular to in plane, but not from up to down. The VCMA coefficient is defined as the change in the magnetic anisotropy energy areal density per unit electric field, in femtojoules per volt-metre. Typical VCMA coefficients of the sputtered 3d metal|MgO system are on the order of tens of femtojoules per volt-metre up to hundreds of femtojoules per volt-metre (refs. 102–105). Substantially larger VCMA coefficients on the order of 1,000 fJ V^{-1}m^{-1} have been reported for some 3d transition metal|oxide systems, where the VCMA effect was dominated by charge trapping and/or ionic migration[106,107]. However, in this type of device, the response time is too slow for last-level cache applications.

VCMA-MRAM has the potential to operate at high speed and extremely low energy[108–110]. Two types of VCMA-MRAM devices have been pursued in the field: precessional-switching VCMA-MRAM[102,103,111,112] and VCMA-assisted SOT/STT-MRAM[113,114]. Precessional-switching VCMA-MRAM utilizes the VCMA effect solely in the presence of an external in-plane magnetic field. During the write process, the energy barrier (Fig. 1c) is lowered to close to zero under the applied voltage, and the free-layer magnetization precesses around the external in-plane field. By applying voltage pulses with duration at the half precession period of the free layer, the device can be toggled between the high and low resistance states in both directions using the same polarity and amplitude voltage pulse (a pre-read before write is required, to determine whether writing is needed, as the writing is not directional). For VCMA-MRAM devices, the precession frequency of the magnetization can be on the order of gigahertz, and sub-nanosecond switching has been demonstrated experimentally[103,112,115]. One major challenge of this device concept is the reliability of the write process, where a precisely controlled voltage pulse duration, tuned for the precession frequency of the device, is required to achieve successful switching. This requires extremely tight control of the precession frequency and its device to device distribution. To date, the best write-error rate demonstrated in single devices is about 10^{-6} errors per write[116] and is expected to worsen substantially for an array, considering the device to device variation. Therefore, write reliability for precessional-switching VCMA-MRAM has to be significantly improved to meet the last-level cache application requirement.

The second type of VCMA-MRAM device relies on the combination of VCMA with STT or SOT switching[113,117,118], where the VCMA effect is utilized to lower the energy barrier and assist the STT or SOT switching. VCMA-STT switching devices showed an improved switching speed compared with STT-only devices, and an improved switching reliability compared with VCMA-only devices[113]. The VCMA effect was also utilized to assist SOT switching with four to eight MTJs sitting on a shared SOT wire, with VCMA applied individually to each junction through its own transistor[117,118]. A 25% SOT switching current reduction was demonstrated at a write pulse width of 0.4 ns with conventional CoFeB|MgO-based MTJs and a tungsten SOT wire[118], limited by the small VCMA coefficient of the free-layer material (~15 fJ $\mathrm{V}^{-1}\mathrm{m}^{-1}$). Although a large VCMA coefficient, up to 1,000 fJ $\mathrm{V}^{-1}\mathrm{m}^{-1}$, was reported in strained CoFe grown on single crystal MgO substrates[119], practical free-layer materials grown on standard substrates with a much improved VCMA coefficient (300–800 fJ $\mathrm{V}^{-1}\mathrm{m}^{-1}$) are needed to make VCMA-MRAM a competitive candidate for last-level cache applications[118].

Outlook

Several exploratory scientific ideas have been proposed to further improve MRAM in the future. For example, thermal magnons could generate STT ten times more efficiently than electrically driven STT[120,121]. Ultrafast sub-picosecond optical switching has been demonstrated using laser pulses[122], but optics are challenging to integrate into dense memory technology[123]. Racetrack memory, using STT to move multiple domain walls in three-dimensional racetracks, has been proposed to improve density; however, domain walls are hard to manipulate in practical applications[124]. Chiral materials have been predicted to generate spin polarization parallel to the direction of current[125], a potential path to reducing the switching current. Remarkably, antiferromagnetic tunnel junctions have been demonstrated to produce magnetoresistance[126,127] and can be switched by SOT[128,129], which may enable faster switching. Overall, spintronics remains an active field generating a steady stream of new and innovative ideas for future STT-MRAM technologies.

Published online: 6 November 2024

References

1. Kent, A. & Worledge, D. A new spin on magnetic memories. *Nat. Nanotech* **10**, 187–191 (2015).
2. Julliere, M. Tunneling between ferromagnetic films. *Phys. Lett. A* **54**, 225–226 (1975).
3. Slonczewski, J. C. Magnetic bubble tunnel detector. *IBM Tech. Discl. Bull.* **19**, 2328–2330 (1976).
4. Engel, B. N. et al. A 4-Mb Toggle MRAM based on a novel bit and switching method. *IEEE Trans. Magn.* **41**, 132 (2005).
5. Slonczewski, J. in *Handbook of Magnetism and Advanced Magnetic Materials* Vol. 5 (eds Kronmuller, H. & Parkin, S.) 2648 (Wiley, 2007).
This article explains the theory of STT, from the original inventor.
6. Thomas, L. et al. Solving the paradox of the inconsistent size dependence of thermal stability at device and chip-level in perpendicular STT-MRAM. In *2015 IEEE Int. Electron Devices Meeting (IEDM)* 26.4.1–26.4.4 (IEEE, 2015).
7. Sun, J. Z. Spin–current interaction with a monodomain magnetic body: a model study. *Phys. Rev. B* **62**, 570 (2000).
This is a classic article on theoretical prediction of the switching current.
8. Naik V. B. et al. Extended MTJ TDDB model,and improved STT-MRAM reliability with reduced circuit and process variabilities. In *IEEE Int. Reliability Physics Symp. (IRPS)* 6B.3-1 (IEEE, 2022).
9. Djayaprawira, D. D. et al. 230% room-temperature magnetoresistance in CoFeB/MgO/CoFeB magnetic tunnel junctions. *Appl. Phys. Lett.* **86**, 092502 (2005).
10. Xue, L. et al. Process optimization of perpendicular magnetic tunnel junction arrays for last-level cache beyond 7 nm node. In *2018 IEEE Symp. VLSI Technology* 117–118 (IEEE, 2018).
11. Ichinose, T. et al. Cryogenic temperature deposition of high-performance CoFeB/MgO/cofeb magnetic tunnel junctions on φ300 mm wafers. *ACS Appl. Electron. Mater.* **5**, 2178 (2023).
12. Islam, R., Cui, B. & Miao, G. X. Dry etching strategy of spin-transfer-torque magnetic random access memory: a review. *J. Vac. Sci. Technol. B* **1**, 050801 (2020).
13. Ip, V. et al. Ion beam patterning of high-density STT-RAM devices. *IEEE Trans. Magn.* **53**, 1–4 (2017).
14. Park, H. et al. High reliability CoFeB/MgO/CoFeB magnetic tunnel junction fabrication using low-damage ion beam etching. *Jpn. J. Appl. Phys.* **59**, SGGB05 (2020).
15. Slaughter, J. M. et al. Technology for reliable spin-torque MRAM products. In *2016 IEEE Int. Electron Devices Meeting (IEDM)* 21.5.1–21.5.4 (IEEE, 2016).
16. Aggarwal, S. et al. Demonstration of a reliable 1 Gb standalone spin-transfer torque MRAM for industrial applications. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 2.1.1–2.1.4 (IEEE, 2019).
17. Choe, J. Recent technology insights on STT-MRAM: structure, materials, and process integration. In *2023 IEEE Int. Memory Workshop (IMW)* 1–4 (IEEE, 2023).
18. Alam S. M. et al. Persistent xSPI STT-MRAM with up to 400MB/s read and write throughput. In *2022 IEEE Int. Memory Workshop (IMW)* 1–4 (IEEE, 2022).
19. Ikegawa, S. et al. High-speed (400MB/s) and low-BER STT-MRAM technology for industrial applications. In *2022 Int. Electron Devices Meeting (IEDM)* 10.4.1–10.4.4 (IEEE, 2022).
20. Vartanian, S. et al. Total ionizing dose and reliability evaluation of the ST-DDR4 spin-transfer torque magnetoresistive random access memory (STT-MRAM). In *2022 IEEE Radiation Effects Data Workshop (REDW) (in conjunction with 2022 NSREC)* 1–5 (IEEE, 2022).
21. Wang, Z. et al. Dual QSPI 8 Gb STT-MRAM for space applications. In *2023 Int. Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2023).
22. Lee, T. Y. et al. World-most energy-efficient MRAM technology for non-volatile RAM applications. In *2022 Int. Electron Devices Meeting (IEDM)* 10.7.1–10.7.4 (IEEE, 2022).
23. Chung, S. W. et al. 4Gbit density STT-MRAM using perpendicular MTJ realized with compact cell structure. In *2016 IEEE Int. Electron Devices Meeting (IEDM)* 27.1.1–27.1.4 (IEEE, 2016).
24. Nakayama, M. et al. 14 nm high-performance MTJ with accelerated STT-switching and high-retention doped Co-Pt alloy storage layer for 1Znm MRAM. In *2023 Int. Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2023).
25. Huai, Y. et al. High density 3D cross-point STT-MRAM. In *2018 IEEE Int. Memory Workshop (IMW)* 1–4 (IEEE, 2018).
26. Seo, S. M. et al. First demonstration of full integration and characterization of 4F^2 1S1M cells with 45 nm of pitch and 20 nm of MTJ size. In *2022 Int. Electron Devices Meeting (IEDM)* 10.1.1–10.1.4 (IEEE, 2022).
27. Wan, L. et al. Fabrication and individual addressing of STT-MRAM bit array with 50 nm full pitch. *IEEE Trans. Magn.* **58**, 1–6 (2022).
28. Lee, Y. K. et al. Embedded STT-MRAM in 28-nm FDSOI logic process for industrial MCU/IoT application. In *2018 IEEE Symp. VLSI Technology* 181–182 (IEEE, 2018).
29. Song, Y. J. et al. Demonstration of highly manufacturable STT-MRAM embedded in 28 nm logic. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2018).
30. Lee, K. et al. 28 nm CIS-compatible embedded STT-MRAM for frame buffer memory. In *2021 IEEE Int. Electron Devices Meeting (IEDM)* 2.1.1–2.1.4 (IEEE, 2021).
31. Oka, M. et al. 3D stacked CIS compatible 40 nm embedded STT-MRAM for buffer memory. In *2021 Symp. VLSI Technology* 1–2 (IEEE, 2021).
32. Ko, S. et al. Highly reliable and manufacturable MRAM embedded in 14 nm FinFET node. In *2023 IEEE Symp. VLSI Technology and Circuits* 1–2 (IEEE, 2023).
33. Kang, G. et al. A 14 nm 128 Mb embedded MRAM macro achieved the best figure-of-merit with 80 MHz read operation and 18.1 Mb/mm^2 implementation at 0.64 V. In *2023 IEEE Symp. VLSI Technology and Circuits* 1–2 (IEEE, 2023).
34. Gallagher, W. J. et al. Recent progress and next directions for embedded MRAM technology. In *2019 Symp. VLSI Technology* T190–T191 (IEEE, 2019).
35. Shimoi, T. et al. A 22 nm 32 Mb embedded STT-MRAM macro achieving 5.9 ns random read access and 5.8 MB/s write throughput at up to Tj of 150 °C. In *2022 IEEE Symp. VLSI Technology and Circuits* 134-135 (IEEE, 2022).
36. Shih, Y. -C. et al. A reflow-capable, embedded 8 Mb STT-MRAM macro with 9 nS read access time in 16 nm FinFET logic CMOS process. In *2020 IEEE Int. Electron Devices Meeting (IEDM)* 11.4.1–11.4.4 (IEEE, 2020).

37. Chih, Y. -D. et al. Design challenges and solutions of emerging nonvolatile memory for embedded applications. In *2021 IEEE Int. Electron Devices Meeting (IEDM)* 2.4.1–2.4.4 (IEEE, 2021).

38. Naik, V. B. et al. JEDEC-qualified highly reliable 22nm FD-SOI embedded MRAM for low-power industrial-grade, and extended performance towards automotive-grade-1 applications. In *2020 IEEE Int. Electron Devices Meeting (IEDM)* 11.3.1–11.3.4 (IEEE, 2020).

39. Naik, V. B. et al. STT-MRAM: a robust embedded non-volatile memory with superior reliability and immunity to external magnetic field and RF sources. In *2021 Symp. VLSI Technology* 1–2 (IEEE, 2021).

40. Golonzka, O. et al. MRAM as embedded non-volatile memory solution for 22FFL FinFET technology. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2018).

41. Thomas, L. et al. Perpendicular spin transfer torque magnetic random access memories with high spin torque efficiency and thermal stability for embedded applications (invited). *J. Appl. Phys.* **115**, 172615 (2014).

42. Jan. G. et al. Demonstration of an MgO based anti-fuse OTP design integrated with a fully functional STT-MRAM at the Mbit level. In *2015 Symp. VLSI Technology* T164–T166 (IEEE, 2015).

43. Lee, T. Y. et al. Magnetic immunity guideline for embedded MRAM reliability to realize mass production. In *2020 IEEE Int. Reliability Physics Symp. (IRPS)*. 1-4 (IEEE, 2020).

44. Chen, C. -H. et al. Reliability and magnetic immunity of reflow-capable embedded STT-MRAM in 16 nm FinFET CMOS process. In *2021 Symp. VLSI Technology* 1–2 (IEEE, 2021).

45. Wang, C. -Y. et al. Reliability demonstration of reflow qualified 22nm STT-MRAM for embedded memory applications. In *2020 IEEE Symp. VLSI Technology* 1–2 (IEEE, 2020).

46. Bhushan, B. et al. Enhancing magnetic immunity of STT-MRAM with magnetic shielding. In *2018 IEEE Int. Memory Workshop (IMW)* 1–4 (IEEE, 2018).

47. Jeong, J. et al. Termination layer compensated tunnelling magnetoresistance in ferrimagnetic Heusler compounds with high perpendicular magnetic anisotropy. *Nat. Commun.* **7**, 10276 (2016).

48. K. Lee, K., Kan, J. J. & Kang, J. S. Unified embedded non-volatile memory for emerging mobile markets. In *2014 IEEE/ACM Int. Symp. Low Power Electronics and Design (ISLPED)* 131–136 (IEEE, 2014).

49. Lu, Y. et al. Fully functional perpendicular STT-MRAM macro embedded in 40 nm logic for energy-efficient IOT applications. In *2015 IEEE Int. Electron Devices Meeting (IEDM)* 26.1.1–26.1.4 (IEEE, 2015).

50. Hwang, W. et al. Energy efficient computing with high-density, field-free STT-assisted SOT-MRAM (SAS-MRAM). *IEEE Trans. Magn.* **59**, 1–6 (2023).

51. Alzate, J. G. et al. 2 MB array-level demonstration of STT-MRAM process and performance towards L4 cache applications. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 2.4.1–2.4.4 (IEEE, 2019).

52. Wang, Z. et al. 22 nm embedded STT-MRAM macro with 10 ns switching and >10^{14} endurance for last level cache applications. In *2021 Symp. VLSI Technology* 1–2 (IEEE, 2021).

53. Miura, S. et al. Scalability of quad interface p-MTJ for 1× nm STT-MRAM with 10-ns low power write operation, 10 years retention and endurance >10^{11}. *IEEE Trans. Electron. Devices* **67**, 5368–5373 (2020).

54. Sakhare, S. et al. JSW of 5.5 MA/cm2 and RA of 5.2-Ω·μm^2 STT-MRAM technology for LLC application. *IEEE Trans. Electron. Devices* **67**, 3618–3625 (2020).

55. Saida, D. et al. 1×- to 2×-nm perpendicular MTJ switching at sub-3-ns pulses below 100 μA for high-performance embedded STT-MRAM for sub-20-nm CMOS. *IEEE Trans. Electron. Devices* **64**, 427–431 (2017).

56. Jan, G. et al. Achieving sub-ns switching of STT-MRAM for future embedded LLC applications through improvement of nucleation and propagation switching mechanisms. In *2016 IEEE Symp. VLSI Technology* 1–2 (IEEE, 2016).

57. Hu, G. et al. Spin-transfer torque MRAM with reliable 2 ns writing for last level cache applications. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 2.6.1–2.6.4 (IEEE, 2019).

58. Safranski, C. et al. Reliable sub-nanosecond switching in magnetic tunnel junctions for MRAM applications. *IEEE Trans. Electron. Devices* **69**, 7180–7183 (2022).

59. Worledge, D. C. Theory of spin torque switching current for the double magnetic tunnel junction. *IEEE Magn. Lett.* **8**, 4306505 (2017).

60. Shao, Q. et al. Roadmap of spin–orbit torques. *IEEE Trans. Magn.* **57**, 800439 (2021).

61. Han, X., Wang, X., Wan, C., Yu, G. & Lv, X. Spin–orbit torques: materials, physics, and devices. *App. Phys. Lett.* **118**, 120502 (2021).

62. Rana, B. & Otani, Y. Towards magnonic devices based on voltage-controlled magnetic anisotropy. *Commun. Phys.* **2**, 90 (2019).

63. Nozaki, T. et al. Recent progress in the voltage-controlled magnetic anisotropy effect and the challenges faced in developing voltage-torque MRAM. *Micromachines 2019* **10**, 327 (2019).

64. Chen, B. et al. Spintronic devices for high-density memory and neuromorphic computing—a review. *Mater. Today* **70**, 193–217 (2023).

65. Miron, I. et al. Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. *Nature* **476**, 189–193 (2011).

66. Liu, L. et al. Spin-torque switching with the giant spin Hall effect of tantalum. *Science* **336**, 555–558 (2012).
Together with Miron et al. (2011), these articles show the first SOT switching of a ferromagnet.

67. Mellnik, A. et al. Spin-transfer torque generated by a topological insulator. *Nature* **511**, 449–451 (2014).

68. Khang, N. H. D., Ueda, Y. & Hai, P. N. A conductive topological insulator with large spin Hall effect for ultralow power spin–orbit torque switching. *Nat. Mater.* **17**, 808–813 (2018).

69. DC, M. et al. Room-temperature high spin–orbit torque due to quantum confinement in sputtered Bi_xSe_{1-x} films. *Nat. Mater.* **17**, 800–807 (2018).

70. Fan, Y. et al. Magnetization switching through giant spin–orbit torque in a magnetically doped topological insulator heterostructure. *Nat. Mater.* **13**, 699–704 (2014).

71. Hibino, Y. et al. Highly energy-efficient spin–orbit-torque magnetoresistive memory with amorphous W–Ta–B alloys. *Adv. Electron. Mater.* **10**, 2300581 (2024).

72. Sun, J. Z. Spin-transfer torque switched magnetic tunnel junctions in magnetic random access memory. *Proc. SPIE* **9931**, 112–124 (2016).

73. Fukami, S. et al. A spin–orbit torque switching scheme with collinear magnetic easy axis and current configuration. *Nat. Nanotech* **11**, 621–625 (2016).

74. Zhu, D. Q. et al. First demonstration of three terminal MRAM devices with immunity to magnetic fields and 10 ns field free switching by electrical manipulation of exchange bias. In *2021 IEEE Int. Electron Devices Meeting (IEDM)* 17.5.1–17.5.4 (IEEE, 2021).

75. Garello, K. et al. Manufacturable 300 mm platform solution for field-free switching SOT-MRAM. In *2019 IEEE Symp. VLSI Technology* T194–T195 (IEEE, 2019).

76. Grimaldi, E. et al. Single-shot dynamics of spin–orbit torque and spin transfer torque switching in three-terminal magnetic tunnel junctions. *Nat. Nanotechnol.* **15**, 111–117 (2020).

77. Sato, N. et al. CMOS compatible process integration of SOT-MRAM with heavy-metal bi-layer bottom electrode and 10ns field-free SOT switching with STT assist. In *2020 IEEE Symp. VLSI Technology* 1–2 (IEEE, 2020).

78. Tsou, Y. -J. et al. First demonstration of interface-enhanced SAF enabling 400 °C-robust 42 nm p-SOT-MTJ cells with STT-assisted field-free switching and composite channels. In *2021 IEEE Symp. VLSI Technology* 1–2 (IEEE, 2021).

79. Song, M. Y. et al. High speed (1ns) and low voltage (1.5V) demonstration of 8 Kb SOT-MRAM array. In *2022 IEEE Symp. VLSI Technology and Circuits* 377–378 (IEEE, 2022).

80. Rahaman, S. Z. et al. Structure and performance co-optimization for the development of highly reliable spin-orbit torque magnetic random access memory. In *2023 International VLSI Symp Technology, Systems and Applications (VLSI-TSA/VLSI-DAT)* 1–2 (IEEE, 2023).

81. Honjo, H. et al. First demonstration of field-free SOT-MRAM with 0.35 ns write speed and 70 thermal stability under 400 °C thermal tolerance by canted SOT structure and its advanced patterning/SOT channel technology. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 28.5.1–28.5.4 (IEEE, 2019).

82. Yoda, H., Yakushiji, K. & Fukushima, A. Proposal & demonstration of low current SOT-MRAM based on brand new mechanism for retention energy of strain-induced magnetic anisotropy. In *2022 Int. Symp. VLSI Technology, Systems and Applications (VLSI-TSA)* 1–2 (IEEE, 2022).

83. Garello, K. et al. SOT-MRAM 300 mm integration for low power and ultrafast embedded memories. In *2018 IEEE Symp. VLSI Circuits* 81–82 (IEEE, 2018).

84. Couet, S. et al. BEOL compatible high retention perpendicular SOT-MRAM device for SRAM replacement and machine learning. In *2021 Symp. VLSI Technology* 1–2 (IEEE, 2021).

85. Shao, Q. et al. Room temperature highly efficient topological insulator/Mo/CoFeB spin-orbit torque memory with perpendicular magnetic anisotropy. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 36.3.1–36.3.4 (IEEE, 2018).

86. MacNeill, D. et al. Control of spin–orbit torques through crystal symmetry in WTe_2/ferromagnet bilayers. *Nat. Phys.* **13**, 300–305 (2017).

87. MacNeill, D. et al. Thickness dependence of spin-orbit torques generated by WTe_2. *Phys. Rev. B* **96**, 054450 (2017).

88. Stiehl, G. M. et al. Layer-dependent spin-orbit torques generated by the centrosymmetric transition metal dichalcogenide β-$MoTe_2$. *Phys. Rev. B* **100**, 184402 (2019).

89. Patton, M. et al. Symmetry control of unconventional spin–orbit torques in IrO_2. *Adv. Mater.* **35**, 2301608 (2023).

90. Bose, A. et al. Tilted spin current generated by the collinear antiferromagnet ruthenium dioxide. *Nat. Electron.* **5**, 267–274 (2022).

91. Safranski, C., Sun, J. Z., Jun-Wen Xu, J.-W. & Kent, A. D. Planar Hall driven torque in a ferromagnet/nonmagnet/ferromagnet system. *Phys. Rev. Lett.* **124**, 197204 (2020).

92. Hibino, Y. et al. Giant charge-to-spin conversion in ferromagnet via spin–orbit coupling. *Nat. Commun.* **12**, 6254 (2021).

93. Ryu, J. et al. Efficient spin–orbit torque in magnetic trilayers using all three polarizations of a spin current. *Nat. Electron.* **5**, 217–223 (2022).

94. Gupta, M. et al. High-density SOT-MRAM technology and design specifications for the embedded domain at 5 nm node. In *2020 IEEE Int. Electron Devices Meeting (IEDM)* 24.5.1–24.5.4 (IEEE, 2020).

95. Sato, N. et al. Two-terminal spin–orbit torque magnetoresistive random access memory. *Nat. Electron.* **1**, 508–511 (2018).

96. Cai, K. et al. Selective operations of multi-pillar SOT-MRAM for high density and low power embedded memories. In *2022 IEEE Symp. VLSI Technology and Circuits* 375–376 (IEEE, 2022).

97. Liao, Y. -C. et al. Spin-orbit-torque material exploration for maximum array-level read/write performance. In *2020 IEEE Int. Electron Devices Meeting (IEDM)* 13.6.1–13.6.4 (IEEE, 2020).

98. Weisheit, M. et al. Electric field-induced modification of magnetism in thin-film ferromagnets. *Science* **315**, 349–351 (2007).

99. Song, C. et al. Recent progress in voltage control of magnetism: materials, mechanisms, and performance. *Prog. Mater. Sci.* **87**, 33–82 (2017).
100. Shiota, Y. et al. Voltage-assisted magnetization switching in ultrathin Fe$_{80}$Co$_{20}$ alloy layers. *Appl. Phys. Express* **2**, 063001 (2009).
101. Maruyama, T. et al. Large voltage-induced magnetic anisotropy change in a few atomic layers of iron. *Nat. Nanotech.* **4**, 158–161 (2009).
This article is the first to show voltage control of magnetic anisotropy in Fe/MgO.
102. Kanai, S. et al. Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction. *App. Phys. Lett.* **101**, 122403 (2012).
103. Wu, C. Y. et. al. Deterministic and field-free voltage-controlled MRAM for high performance and low power applications. In *2020 IEEE Symp. VLSI Technol* 1–2 (IEEE, 2020).
104. Carpenter, R. et al. Atomistic simulations enabling BEOL compatible VCMA-MRAM with a coefficient ≥100fJ/Vm. In *2021 IEEE Int. Electron Devices Meeting (IEDM)* 17.6.1–17.6.4 (IEEE, 2021).
105. Nozaki, T. et al. Large voltage-controlled magnetic anisotropy effect in magnetic tunnel junctions prepared by deposition at cryogenic temperatures. *APL. Mater.* **11**, 121106 (2023).
106. Bauer, U. et al. Magneto-ionic control of interfacial magnetism. *Nat. Mater.* **14**, 174–181 (2015).
107. Rajanikanth, A. et al. Magnetic anisotropy modified by electric field in V/Fe/MgO(001)/Fe epitaxial magnetic tunnel junction. *Appl. Phys. Lett.* **103**, 062402 (2013).
108. Wang, W. G. et al. Electric-field-assisted switching in magnetic tunnel junctions. *Nat. Mater.* **11**, 64–68 (2012).
109. Grezes, C. et al. Ultra-low switching energy and scaling in electric-field-controlled nanoscale magnetic tunnel junctions with high resistance-area product. *Appl. Phys. Lett.* **108**, 012403 (2016).
110. Kang, W., Chang, L., Zhang, Y. & Zhao, W. Voltage-controlled MRAM for working memory: perspectives and challenges. In *2017 Design, Automation and Test in Europe (DATE)* 542–547 (2017).
111. Shiota, Y. et al. Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses. *Nat. Mater.* **11**, 39–43 (2012).
112. Noguchi, H. et al. Novel voltage controlled MRAM (VCM) with fast read/write circuits for ultra large last level cache. In *2016 IEEE Int. Electron Devices Meeting (IEDM)* 27.5.1–27.5.4 (IEEE, 2016).
113. Kanai, S. et al. Magnetization switching in a CoFeB/MgO magnetic tunnel junction by combining spin-transfer torque and electric field effect. *Appl. Phys. Lett.* **104**, 212406 (2014).
114. Peng, S. Z. et al. Field-free switching of perpendicular magnetization through voltage-gated spin-orbit torque. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 28.6.1–28.6.4 (IEEE, 2019).
115. Shao, Y. et al. Sub-volt switching of nanoscale voltage-controlled perpendicular magnetic tunnel junctions. *Commun. Mater.* **3**, 87 (2022).
116. Yamamoto, T. et al. Improvement of write error rate in voltage driven magnetization switching. *J. Phys. D: Appl. Phys.* **52**, 164001 (2019).
117. Yoda, H. et al. Voltage-Control Spintronics Memory (VoCSM) having potentials of ultra-low energy-consumption and high-density. In *2016 IEEE Int. Electron Devices Meeting (IEDM)* 27.6.1–26.6.4 (IEEE, 2016).
118. Wu, C. Y. et al. Voltage-gate-assisted spin–orbit-torque magnetic random-access memory for high-density and low-power embedded applications. *Phys. Rev. Appl.* **15**, 064015 (2021).
119. Kato, Y. et al. Giant voltage-controlled magnetic anisotropy effect in a crystallographically strained CoFe system. *Appl. Phys. Express* **11**, 053007 (2018).
120. Slonczewski, J. C. Initiation of spin-transfer torque by thermal transport from magnons. *Phys. Rev. B* **82**, 054403 (2010).
121. Mojumder, N. N., Abraham, D. W., Roy, K. & Worledge, D. C. Magnonic spin-transfer torque MRAM with low power, high speed, and error-free switching. *IEEE Trans. Magn.* **48**, 2016–2024 (2012).
122. Stanciu, C. D. et al. All-optical magnetic recording with circularly polarized Light. *Phys. Rev. Lett.* **99**, 047601 (2007).
123. Kimel, A. V. & Li, M. Writing magnetic memory with ultrashort light pulses. *Nat. Rev. Mater.* **4**, 189–200 (2019).
124. Bläsingm, R. et al. Magnetic racetrack memory: from physics to the cusp of applications within a decade. *Proc. IEEE* **108**, 1303 (2020).
125. Chang, G. et al. Topological quantum properties of chiral crystals. *Nat. Mater.* **17**, 978–985 (2018).
126. Qin, P. et al. Room-temperature magnetoresistance in an all-antiferromagnetic tunnel junction. *Nature* **613**, 485–489 (2023).
127. Chen, X. et al. Octupole-driven magnetoresistance in an antiferromagnetic tunnel junction. *Nature* **613**, 490–495 (2023).
128. Wadley, P. et al. Electrical switching of an antiferromagnet. *Science* **351**, 587 (2016).
129. Deng, Y. et al. All-electrical switching of a topological non-collinear antiferromagnet at room temperature. *Natl Sci. Rev.* **10**, nwac154 (2023).
130. Moodera, J. S., Kinder, L. R., Wong, T. M. & Meservey, R. Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.* **74**, 3273 (1995).
This is a classic article on the first demonstration of large magnetoresistance at room temperature.
131. Wang, D., Nordman, C., Daughton, J. M., Qian, Z. & Fink, J. 70% TMR at room temperature for SDT sandwich junctions with CoFeB as free and reference layers. *IEEE Trans. Magn.* **40**, 2269–2271 (2004).
132. Parkin, S. et al. Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. *Nat. Mater.* **3**, 862–867 (2004).
133. Yuasa, S. et al. Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. *Nat. Mater.* **3**, 868–871 (2004).
Together with Parkin et al. (2004), these articles present the discovery of high magnetoresistance using MgO tunnel barriers.
134. Butler, W. H., Zhang, X.-G., Schulthess, T. C. & MacLaren, J. M. Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches. *Phys. Rev. B* **63**, 054416 (2001).
135. Mathon, J. Theory of spin-dependent tunnelling in magnetic junctions. *J. Phys. D: Appl. Phys.* **35**, 2437 (2002).
136. Slonczewski, J. C. Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier. *Phys. Rev. B* **39**, 6995 (1989).
This is a classic article predicting STT in tunnel junctions.
137. Berger, L. Low-field magnetoresistance and domain drag in ferromagnets. *J. Appl. Phys.* **49**, 2156–2161 (1978).
138. Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353 (1996).
139. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
This is a classic article predicting STT in magnetic multilayers.
140. Tsoi, M. et al. Excitation of a magnetic multilayer by an electric current. *Phys. Rev. Lett.* **80**, 4281 (1998).
141. Sun, J. Z. Current-driven magnetic switching in manganite trilayer junctions. *J. Magn. Magn. Mater.* **202**, 157 (1999).
142. Myers, E. B. et al. Current-induced switching of domains in magnetic multilayer devices. *Science* **285**, 867 (1999).
143. Katine, J. A., Albert, F. J., Buhrman, R. A., Myers, E. B. & Ralph, D. C. Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars. *Phys. Rev. Lett.* **84**, 3149 (2000).
144. Huai, Y., Albert, F., Nguyen, P., Pakala, M. & Valet, T. Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions. *Appl. Phys. Lett.* **84**, 3118–3120 (2004).
145. Fuchs, G. D. et al. Spin-transfer effects in nanoscale magnetic tunnel junctions. *Appl. Phys. Lett.* **85**, 1205–1207 (2004).
146. Hosomi, M. et al. A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram. In *2005 Int. Electron Devices Meeting (IEDM)* 459–462 (2005).
147. Mangin, S. et al. Current-induced magnetization reversal in nanopillars with perpendicular anisotropy. *Nat. Mater.* **5**, 210–215 (2006).
148. Meng, H. & Wang, J.-P. Spin transfer in nanomagnetic devices with perpendicular anisotropy. *Appl. Phys. Lett.* **88**, 172506 (2006).
149. Kishi, T. et al. Lower-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM. In *2008 Int. Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2008).
150. Ikeda, S. et al. A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction. *Nat. Mater.* **9**, 721–724 (2010).
This article is the first to show perpendicular CoFeB/MgO MTJs.
151. Worledge, D. C. et al. Switching distributions and write reliability of perpendicular spin torque MRAM. In *2010 Int. Electron Devices Meeting (IEDM)* 12.5.1–12.5.4 (IEEE, 2010).
This article is the first to show reliable perpendicular STT-MRAM (using CoFeB/MgO).
152. Dieny, B. & Chshiev, M. Perpendicular magnetic anisotropy at transition metal/oxide interfaces and applications. *Rev. Mod. Phys.* **89**, 025008 (2017).
This article is an excellent review of interfacial perpendicular magnetic anisotropy.
153. Monso, S. et al. Crossover from in-plane to perpendicular anisotropy in Pt/CoFe/AlO$_x$ sandwiches as a function of Al oxidation: a very accurate control of the oxidation of tunnel barriers. *Appl. Phys. Lett.* **80**, 4157–4159 (2002).
154. Worledge, D. C. et al. Spin torque switching of perpendicular Ta|CoFeB|MgO-based magnetic tunnel junctions. *Appl. Phys. Lett.* **98**, 022501 (2011).
This article is the first to show a complete perpendicular CoFeB/MgO MTJ stack.
155. Davidson, A., Amin, V. P., Aljuaid, W. S., Haney, P. M. & Fan, X. Perspectives of electrically generated spin currents in ferromagnetic materials. *Phys. Lett. A* **384**, 11 (2020).

Author contributions

All authors contributed equally to the preparation of this manuscript. All authors researched the data, substantially contributed to discussion of content, and wrote and reviewed/edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Electrical Engineering* thanks Andrew D. Kent, Shan X. Wang, Kyung-Jin Lee, Shinji Yuasa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Related links

eMRAM: https://www.eenewseurope.com/en/nxp-tsmc-bring-embedded-mram-to-automotive-mcus/
Future directions: https://www.flashmemorysummit.com/Proceedings2019/08-05-Monday/20190805_MRAMDD_App_Briefs_Yardley.pdf
Hard disk: http://www.ibmfiles.com/ibmfiles/powerpc/hdd_spec_scsi.pdf
NXP: https://www.nxp.com/company/about-nxp/nxp-and-tsmc-to-deliver-industrys-first-automotive-16-nm-finfet-embedded-mram:NW-NXP-AND-TSMC-DELIVER-FIRST16NM-FINFET-MRAM
Prices from www.digikey.com in October 2023: https://www.digikey.com/
Samsung: https://news.samsung.com/global/samsung-electronics-starts-commercial-shipment-of-emram-product-based-on-28nm-fd-soi-process
Samsung: https://news.samsung.com/global/samsung-electronics-unveils-automotive-process-strategy-at-samsung-foundry-forum-2023-eu
Sony: https://www.techinsights.com/blog/product-sony-cxd5610gf-globalfoundries-22fdx-emram-telit-gnss-receiver-advanced-memory
UMC: https://www.avalanche-technology.com/umc-and-avalanche-technology-partner-for-mram-development-and-28nm-production/
Western Digital: https://www.westerndigital.com/company/innovation/non-volatile-memory