

Successful MRAM Production Requires Good Magnetic Test Equipment

Tom Coughlin, Coughlin Associates, Inc.

3/12/20

Introduction

Memory plays a key role in both training and implementation of artificial intelligence solutions, such as machine learning. It is also a requirement for the creation of advanced network technologies, such as 5G, which will require processing and memory at the edge of networks as well as at the endpoints to implement IoT and other applications.

Today, most high-performance memory is volatile, that means, when power is removed from the device, whatever was stored in memory, is lost. However, memory can consume a lot of power, especially DRAM, which requires regular and frequent refreshing of the data in memory. Many new applications, such as IoT, require the placement of connected sensors and other devices in energy constrained situations, running on batteries or using energy harvesting. In the data center, memories, such as DRAM, consume a significant percentage of total power. There are many applications that would benefit from the use of non-volatile memories that retain their data, even when the power is turned off.

For many IoT applications, including AI inference, the IoT system is not always fully active, so parts of the system can be turned off until needed. Using non-volatile memories makes it easier to turn the power off for these memories when they are not needed. Because they are non-volatile they can be turned off more often and both power-down and power-up require less time, resulting in better system performance, while minimizing power consumption.

In addition to the need for non-volatile memory in many embedded and stand-alone applications, some emerging non-volatile memories can scale to higher densities than many conventional memories. For instance, NOR flash appears to be effectively constrained to 22nm or greater lithographic features, planar NAND flash is limited to 15nm features or greater (this is why high capacity NAND is moving to 3D structures), and of course SRAM, with its use of 5-6 transistors, takes up a lot of real estate on a semiconductor die. DRAM is also limited in its density scaling by the size of its transistors.

For these reasons, many of the major semiconductor foundries are offering MRAM as non-volatile memory for embedded applications. MRAM can replace NOR or SRAM to provide higher density non-volatile memory on a device. Replacing SRAM with MRAM could allow AI inference engines with more memory (and non-volatile memory to boot) for storing trained models.

Important keys for successful manufacturing of non-volatile memory products are good design, including testing and verification of device properties and once in manufacturing, quality control testing at the wafer and device level. We will look at the manufacturing and testing of emerging non-volatile memory technologies that will enable IoT, AI and advanced network

development and commercialization, but first, let's look at characteristics of current memory and emerging non-volatile memory technologies and see why MRAM stands out.

Comparison of Non-Volatile Memory Technologies

The table below compares some important characteristics of various emerging non-memory technologies compared with established memories (SRAM, DRAM, NOR and NAND Flash)¹.

Comparison of Established and Emerging Memories

	Established Memory Types				Emerging Memory Types				
	SRAM	DRAM	NOR Flash	NAND Flash	FRAM	ReRAM	Toggle MRAM	STT	PCM
Nonvolatile?	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cell Size	50-120f ²	6-10f ²	10f ²	4-5f ²	15-32f ²	4-8f ²	16-32f ²	6-8f ²	2-8f ²
Read Time	1-100ns	30ns	10ns	50ns	20-50ns	10-20ns	3-20ns	3-15ns	5-20ns
Write Time	1-100ns	50ns	10 ⁵ -10 ⁷ ns	10 ⁶ -10 ⁸ ns	50ns	10-60μs	10-20ns	3-15ns	>30ns
Endurance	∞	∞	10 ⁵	10 ³	10 ¹²	10 ⁵⁻⁶	10 ¹⁵	10 ¹⁵	10 ¹²
Write Energy	Low	Low	High	Med	Low	Low	Somewhat High	Low	Low
Write Voltage	None	2	6-8	1.8	2-3	1.2	3	1.5	1.5-3

Phase Change Memory (PCM), such as Intel's Optane memory, is smaller than DRAM (and thus more dense), although not as dense as 3D NAND flash, while it has performance between that of DRAM and NAND flash. This is why Optane is being sold in NVMe SSDs and in DIMMs in computer memory buses to provide either a higher performance storage tier compared to NAND flash, or an augmentation to DRAM memory in computing applications.

Ferroelectric RAM (FRAM) and Resistive RAM (RRAM), like standalone MRAM have been used for special niche applications for many years. FRAM standalone memories have been low density products but recent discoveries about using a special phase of hafnium oxide as a ferroelectric memory give some hope that this technology could have a wider use. Resistive memory (which encompasses many different resistive technologies) show promise to continue scaling to very small memory sizes and may be used for embedded memory applications.

However, probably the greatest promise for embedded as well as standalone non-volatile emerging memory is MRAM.

Development of MRAM Technology

¹ Emerging Memories Ramp Up, Coughlin Associates and Objective Analysis, 2019, <https://tomcoughlin.com/product/emerging-memory-report/>

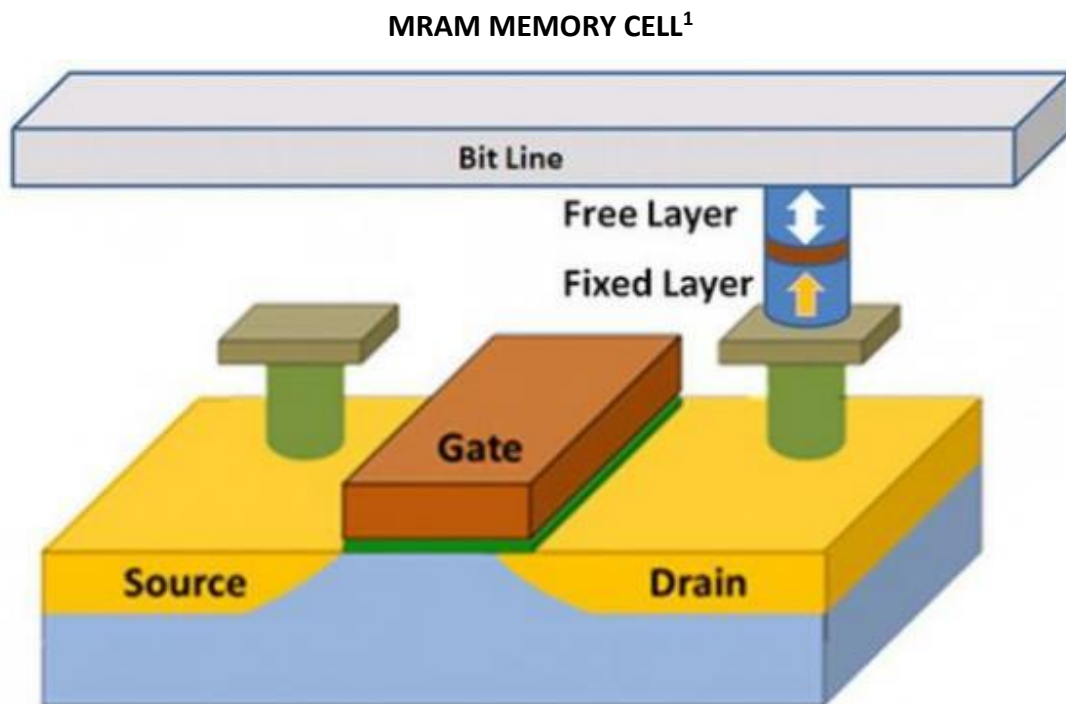
Toggle (or field driven) MRAM comprises the bulk of shipped standalone MRAM devices. However, toggle MRAM will not scale enough to replace most other memories. Spin Tunnel Torque (STT) MRAM products will scale to much higher densities and require lower energy to write than Toggle MRAM. In 2019, Everspin, the company that has shipped most or all of the standalone MRAM memory, began to ship STT MRAM with up to 1 Gb chip capacities, a memory density that makes these devices of greater interest for many applications.

Major embedded semiconductor foundries are offering MRAM non-volatile memory options for embedded products used in industry as well as consumer applications. These foundries include Global Foundries, Intel, Samsung and TSMC.

STT MRAM has high performance, but it is not as fast as the fastest SRAM. However, an MRAM technology called Spin Orbit Torque (SOT) holds potential to match SRAM performance. Currently SOT MRAM is being developed in laboratories throughout the world, but with the decreasing costs of STT MRAM products, both embedded and standalone, SOT MRAM could be the means for MRAM to replace even the fastest SRAM applications, providing higher non-volatile memory densities that could enable very low power IoT and AI applications.

MRAM Production and Testing

Magnetic RAM memories (field or spin devices) use a select transistor for the memory cell, see the figure below. An MRAM cell might have a dedicated transistor or this transistor may be shared between two memory cells. Specific select transistor designs must be created on a wafer prior to its use for MRAM production.

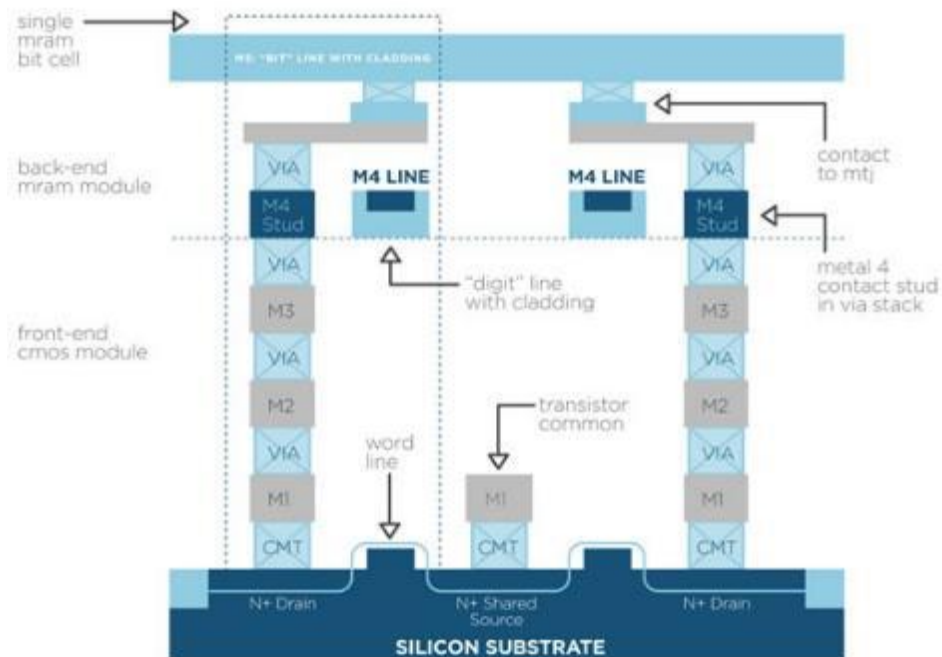


MRAM fabrication today can be split between CMOS wafer and magnetic cell fabrication steps (done at the back end of the CMOS wafer process). The CMOS wafers may be supplied by a semiconductor foundry, and the magnetic cell fabrication is accomplished in a specialized fab using tools that are not generally available in semiconductor foundries, although this may be changing as MRAM becomes more common in foundries.

The CMOS wafer fabrication today is done in the front end of line (FEOL) and the magnetic layer fabrications is done at the back end of line (BEOL). This is common practice, and it helps standalone MRAM companies reduce equipment costs by allowing them to outsource the CMOS fabrication. It also keeps materials that are used in the magnetic layers, which might otherwise be considered contaminants, out of the CMOS foundry.

The CMOS wafers supplied by the foundry are fabricated with metal studs in order to interface the conductors, MTJ elements, and the surface of the CMOS wafer is smoothed by chemical mechanical polishing (CMP). Typically, the magnetic facility deposits the metal write/read/selection conductors, MTJ material, top electrodes, and any required magnetic shielding. Patterning of the magnetic layers is usually performed by the magnetic back-end facility, since the materials used in the magnetic layers may not be compatible with CMOS processing. The figure below shows the basic construction of the CMOS and magnetic layers.

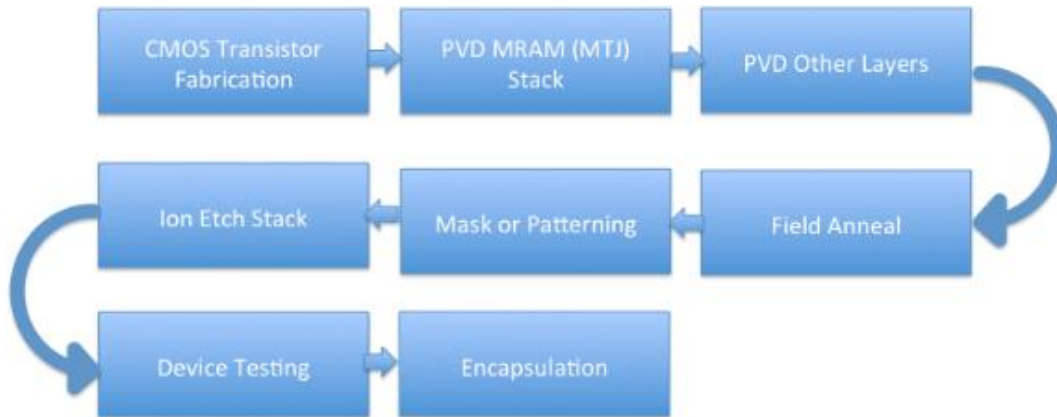
CROSS-SECTION OF AN MRAM DEVICE SHOWING DETAILS OF THE CMOS SUBSTRATE AND MAGNETIC LAYERS FABRICATED ON TOP OF THE SUBSTRATE¹



Source: Everspin

The figure below shows a representative MRAM manufacturing process flow on an existing CMOS wafer.

MRAM MANUFACTURING PROCESS FLOW¹



Source: Coughlin Associates, 2019

Measurement and characterization of devices and pre-device steps are important to control the manufacturing processes of MRAM memories. There are numerous measurements that are very common to regular semiconductors using common analytical tools such as SEMs, TEMs, various depth profiling as well as material and structural characterization. However, the production of MRAM memories also depends upon on measurement tools that are specialized for MRAM and STT MRAM measurement.

In particular, quasi-static testing of magnetoresistive materials is an important measurement in making successful MRAM devices. This testing, usually done at the wafer level after blanket coating with MRAM magnetic layers, can help in the successful design of MRAM memory, but is also useful for quality control measurements to certify a production process and as a troubleshooting tool when something goes amiss in an MRAM factory. Magnetic testing is a crucial requirement for achieving optimal MRAM products, whether for stand-alone or embedded memory applications.

ISI's Wafer Level MRAM Testers

Integral Solutions Inc (ISI) makes Quasi-Static testers for hard disk drive magnetic tunnel junction reader production as well as MRAM development and production. They have been in the business of magnetic tunnel junction (MTJ) testing for hard disk drive heads for over 24 years, with current installations in the head production factories of the major hard disk drive (HDD) companies.

Besides the specialized products used for production testing of HDD heads, ISI was the first company to make wafer level testers for the MRAM industry and probably has the largest installed base of STT-MRAM wafer level testers. Their longtime experience in the HDD head industry gives them unsurpassed experience in MTJ testing that they bring to the emerging MRAM testing industry.

Their WLA-3000 STT-MRAM Wafer Level Analyzer, the Gen3 Pulser, matched with its proprietary probe card interface, produces programmable pulses as low as 5nS, with in-situ

ability to perform ultra-fast measurements on the MTJs after pulsing. Their 4 point and 2 point I-V/R-V tester can providing pulse and DC measurements with up to 3K Oersted in plane and 5K Oersted perpendicular to the plane. Their machine is capable of 360-degree field angles and temperatures up to 150 degrees C, with their hot chuck option. [The upcoming Gen4 system includes even further enhancements to these specifications, including pulse widths down to 2nS and perpendicular field as high as 15K Oersted](#)

In addition, they can provide film reliability tests including breakdown voltage, endurance and write probability and read disturb measurements and their thermal ferromagnetic resonance (FMR) characterization (using an RF preamp connected to their RF probe assembly) can be used for process control as well as device development. The figure below shows the ISI Wafer Level Quasi-Static Tester.

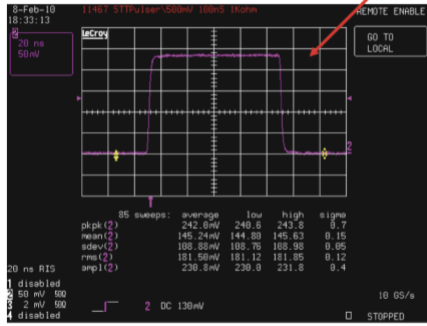
ISI WLA 3000 WAFER LEVEL QUASI-STATIC TESTER



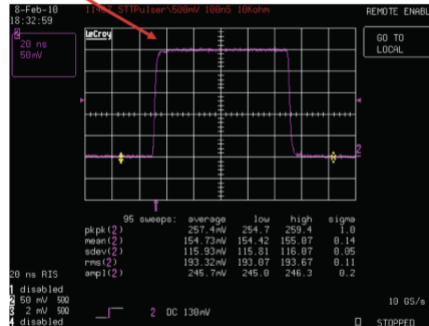
ISI can set up their testing with EG, TEL, and Accretech Probers (shown above) and they can configure for virtually any other prober. Their electronics and test software are independent from the wafer prober and their test head can be installed on any prober. A simple cantilevered probe card can be used to provide multi-device testing capability on 200 or 300 mm wafers.

The ISI electronics are best of breed. ISI's methodology has been to develop their own proprietary frontend circuitry, to perform both pulsing and measurement operations on MTJs. The figure below shows pulse samples of the proprietary ISI pulsing channel with virtually no overshoot.

Virtually no Overshoot

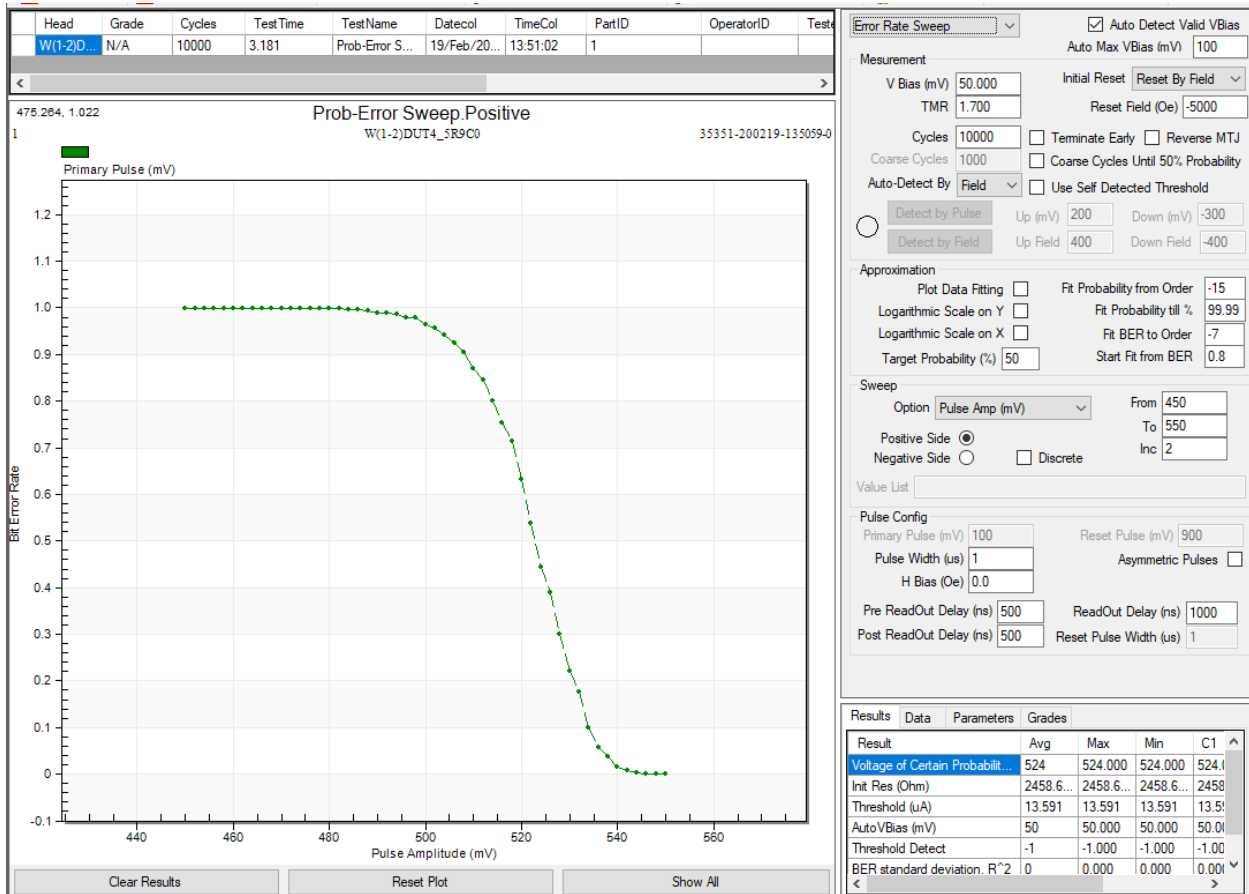


100nS 500mV Pulse with
1 kOhm DUT



100nS 500mV Pulse with
10 kOhm DUT

This proprietary design not only provides specific test functionality for the MRAM industry, but unique to the industry, ISI's fully-integrated frontend offers extremely fast write/read cycling measurements, allowing the performance of extremely fast measurements such as Error Rates. The figure below shows the probability of error with pulse amplitude. For each data point on the curve these tests were configured to run 10^5 write/reset cycles, yet the overall test time for the entire curve was still under 3.2s.



ISI has an aggressive roadmap for future MRAM testing and other production tools. They are working on their fourth generation ISI pulser module, and they are developing external pulser integration for next generation spin-orbit torque (SOT) devices that could replace fast SRAM memory. They are also developing a 2.5-3.0 Tesla fully-automated wafer initializer.

Conclusions

MRAM is enabling the next generation of embedded devices supporting AI, IoT and advanced networking technology; in the data center, at the edge and in the endpoints. In addition, standalone MRAM has become an important non-volatile cache and buffer for many applications. Providing MRAM for all these applications requires good MTJ design and testing in the production environment.

ISI has long time experience in the design and factory testing of MTJ devices and can meet your MRAM development and production testing needs. They are also creating new testing and production tools to help your MRAM program to be a success. Talk with ISI and see why they should be your MRAM testing partner.

ISI is located in Santa Clara, CA. They are ready to meet your design and production testing MRAM needs.

3000 Olcott St., Santa Clara, CA

Phone: 408-653-0300

Email: sales@us-isi.com

Website: www.us-isi.com