Semiconductor Memory Storage (popular types)

- **Volatile**
  - RAM
    - DRAM
    - SRAM
  - Floating Gate
    - Ferro-electric
  - Nitride
    - Magnetic
  - Emerging
    - Phase Change

- **Non-Volatile**
  - ROM & Fuse
    - Polymer

“NV Ram”
“Unified Mem”
Non-Volatile Memory History

- Bipolar ROMS/PROMS → Late 60’s
- EPROM 1971
- EEPROM 1980
- NOR Flash 1988
- AND, DiNOR Flash 1990’s
- NAND Flash 1985
- MLC NAND Flash 1996
- MLC NOR Flash 1995
- Ferro-electric 1988
- Nitride Storage 2002
- MRAM, Phase Change, Polymer 2000’s

MLC = Multi-Level Cell
Technology Comparison

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>Research</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOR Flash</td>
<td>NAND Flash</td>
</tr>
<tr>
<td>Cost</td>
<td>$10\lambda^2$</td>
<td>$6\lambda^2$</td>
</tr>
<tr>
<td>Cell Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read Characteristics</td>
<td>10’s ns 100’s cells</td>
<td>10’s us 1000’s cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write Characteristics</td>
<td>100’s ns 10’s cells</td>
<td>100’s us 1000’s cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $\lambda$ represents minimum feature size for any technology
- Feature size=process lithography capability
- Example: 0.12u lithography, $10\lambda^2$ cell size yields a cell area of $0.144u^2$
Trending Storage Technologies

• **Tunneling Magnetic Junction RAM (TMJ-RAM):**
  – Speed of SRAM, density of DRAM, non-volatile (no refresh)
  – New field called “Spintronics”: combination of quantum spin and electronics
  – Same technology used in high-density disk-drives

• **MEMs storage devices:**
  – Large magnetic “sled” floating on top of lots of little read/write heads
  – Micromechanical actuators move the sled back and forth over the heads

**PRAM – phase change devices**
- is a type of non-volatile computer memory. PRAM uses the unique behavior of chalcogenide glass, which can be "switched" between two states, crystalline and amorphous, with the application of heat.
Tunneling Magnetic Junction

Magnetic Memory Cell Size <1μm

Write Line

Read Line

H

Current
History

In 1984 Drs. Arthur Pohm and Jim Daughton, both employed at that time by Honeywell, conceived of a new class of magnetoresistance memory devices which offered promise for high density, random access, nonvolatile memory. In 1989 Dr. Daughton left Honeywell to form Nonvolatile Electronics, Inc.

Today there is a tremendous potential for MRAM as a nonvolatile, solid state memory to replace flash memory and EEPROM where fast writing or high write endurance is required, and in the longer term as a general purpose read/write random access memory. NVE has a substantial patent portfolio containing 10 MRAM patents, and is willing to license these, along with 12 Honeywell MRAM patents, to companies interested in manufacturing MRAM. In addition, NVE is considering internal production of certain niche MRAM products over the next several years.
These early memories (mostly magnetic core memories) used inductive signals for determining the storage state (“1” or “0”). A magnetic field (current) was used to “interrogate” the memory element, and the polarity of induced voltages in a sensing circuit depended on whether a “1” or “0” was stored.
In the mid 1980’s an MRAM concept was developed at Honeywell which has some common features with most modern versions.

- Writing using magnetic hysteresis
- Reading using magnetoresistance of the same body where data is stored
- Memory cells integrated on an integrated circuit chip

Figure 3. Earliest MRAM Concept.
Magnetoresistance

Reading of this cell depended on the differential resistance of the cell when a sense current was passed through it. Because the sense current creates a magnetic field which opposes the magnetization in one storage state, but is in the same direction in the other state, the angle of rotation was different for a “1” or “0”. The magnetic material used was a cobalt-permalloy alloy with a normal anisotropic magnetoresistance (AMR) ratio of about 2%. Despite improvements in reading methods [6], the maximum differential resistance of the cell between a “1” and a “0” when it was read was about ¼ of the 2% magnetoresistance, or about 0.5%. In real arrays with practical sense currents, this gave differential sense signals of 0.5 to 1.0 mV. These sense signals allowed 16K bit integrated MRAM chips to operate with a read access time of about 250 ns [7]. Write times for the MRAM was 100 ns, and could have been faster if needed.

Figure 4. 256 K MRAM Chip (Courtesy of Honeywell).
Giant Magnetoresistance

The discovery of Giant Magnetoresistance (GMR) materials in 1989 [8,9] gave hope for higher signals and faster read access time. In 1991 magnetic films sandwiching a copper layer and etched into stripes showed a magnetoresistance ratio of about 6%. This magnetic configuration fit the aforementioned MRAM cell with little modification. Since the read access times tend to improve as the square of the signal, normal scaling would indicate that the improvement of a factor of 3 in magnetoresistance would lead to a 9 times improvement in read access time. Read access times of under 50 ns were achieved for MRAM with GMR materials [10].

Even with GMR materials this cell had serious limitations. The competition – semiconductor memory – was still faster because of the low MRAM sense signal. Worse, there was a limit to the reduction of cell size because the cell would not work with sense lines narrower than about 1 micron. This was due to magnetization curling from the edges of the stripe, where the magnetization is pinned along the stripe.
Freescale’s magnetoresistive random access memory (MRAM) products combine magnetic storage elements with a standard complementary-metal-oxide-semiconductor (CMOS) logic process to obtain the benefits of high density, low cost, non-volatility, SRAM speed and unlimited read/write endurance, a combination not found in other existing volatile or non-volatile memory technologies.

Freescale MRAM products employ a one transistor, one magnetic tunnel junction (MTJ) memory cell (Figure 1). MRAM products are currently fabricated using a 180 nm CMOS process using five levels of metal, including program lines clad with highly permeable material for magnetic flux concentration. The patented Freescale architecture, bit structure and toggle mode storage technique deliver nvRAM products with the best price/performance and highest reliability.

Figure 1. Schematic of a 1-transistor, 1-MTJ memory cell showing the write lines above and below the bit and the read current path.
MRAM Description

MRAM is based on magnetic storage elements integrated with CMOS processing. Each storage element uses a magnetic tunnel junction (MTJ) device. The MTJ is composed of a fixed magnetic layer, a thin dielectric tunnel barrier and a free magnetic layer. When a bias is applied to the MTJ, electrons that are spin polarized by the magnetic layers traverse the dielectric barrier through a process known as tunneling. The MTJ device has a low resistance when the magnetic moment of the free layer is parallel to the fixed layer and a high resistance when the free layer moment is oriented antiparallel to the fixed layer moment. This change in resistance with the magnetic state of the device is an effect known as magnetoresistance, hence the name “magnetoresistive” RAM.

Unlike most other semiconductor memory technologies, the data is stored as a magnetic state rather than a charge and sensed by measuring the resistance without disturbing the magnetic state. Using a magnetic state for storage has two main benefits. First, the magnetic polarization does not leak away with time like charge does, so the information is stored even when the power is turned off. And second, switching the magnetic polarization between the two states does not involve actual movement of electrons or atoms and thus no known wear-out mechanism exists. The magnetoresistive device used in MRAM is very similar to the device used for the read head of magnetic hard disk drives.

During the write operation, current pulses are passed through a digit line and a bit line, writing only the bit at the cross point of those two lines. During the read operation, the target bit’s isolation transistor is turned on to bias the MTJ and the resulting current is compared to a reference to determine if the resistance state is low or high.
Comparing MRAM with other memory technologies suggests that it can be competitive in overall performance. Since MRAM is nonvolatile, it retains the data when completely turned off. Since background refreshing is not required, MRAM can be shut down when inactive, significantly reducing system power consumption when compared to DRAM. The straightforward integration scheme used for MRAM also makes it easier to embed. Compared to SRAM, MRAM is more cost effective due to its smaller cell size. It is also nonvolatile, which is only available for SRAM in more complex and expensive battery backup solutions. MRAM achieves much better write performance than flash since no high-voltage tunneling mode is required. The MRAM write cycle is much faster and consumes much less energy because the energy per bit is several orders of magnitude lower than flash. In addition, MRAM has unlimited endurance, with no known deterioration mechanism, while typical flash endurance is 10^5 read/write cycles.
MRAM products are designed and specified to prevent changes to stored data for externally applied external magnetic fields of <15 Gauss (Oersted). As a comparison, the earth’s magnetic field at its surface is less than 0.5 Gauss and the Occupational Safety and Health Administration (OSHA) require a posted warning for areas around instruments that exceed 5 Gauss. As is evident, the specification for MRAM products exceeds these values by a wide margin. There are two main sources of magnetic fields—current-carrying wires and permanent magnetic materials. In each case, the source’s geometry determines the extent of the magnetic field and its magnitude, field at considerable distances from the field’s source.
A schematic cross-sectional view of Freescale’s integrated MRAM cell for a 1T-1MTJ cell architecture is shown in Figure 7. The MRAM process module is integrated between the last two layers of metal in an otherwise standard semiconductor process flow. The MRAM module is termed a “backend” module because it is inserted after all of the associated CMOS circuitry has been fabricated. This integration scheme requires no alteration to the front-end CMOS process flow. This back-end approach separates the specialized magnetic materials processing from the standard CMOS process. This integration scheme lends itself to embedded application where the memory core is part of a non-memory circuit such as a processor or controller. For example, a processor may need to have some fast memory and some nonvolatile memory on board—MRAM can provide both capabilities. Because the MRAM module is independent of the front-end CMOS, the MRAM capability can be added without perturbing the CMOS logic process. This approach provides cost and performance advantages in many system-on-chip applications.

Integration of Magnetic Devices with CMOS
Current Status

Several large companies currently have R&D programs on MRAM technology, and Honeywell has announced working MRAM components. With numerous competitors in the field, there has been a reluctance to publish results.

But it is clear that MRAM has the potential to be as fast and dense as DRAM with the additional advantage of nonvolatility. Compared with flash and EEPROMs, MRAM writes much faster and does not deteriorate with millions of write cycles.

2011
March – PTB, Germany, announces having achieved a below 500 ps (2GBit/s) write cycle

2012
November - Chandler, AZ, USA, Everspin debuts 64Mb ST-MRAM
Latest product

Everspin Technologies was formed in June 2008 as a spin out of Freescale Semiconductor

Standard asynchronous, parallel MRAM products are available from 256Kb to 16Mb densities. Serial MRAMs are available in densities from 256Kb to 4Mb.

Non-volatile 64Mb DDR3

- DDR3-1600 ST-MRAM
- 16Mbx4, 8Mbx8, and 4Mbx16 configurations
- Supports Standard DDR3 SDRAM Features
- No refresh required
- Burst length: 8 (programmable Burst Chop of 4)
- DDR3 SDRAM Standard FBGA Package
  Pinout:
  - VDD = 1.5V +/-0.075V
  - On-device termination
  - 20 year data retention

Presently, MRAM is approx. 50x the cost of FLASH per GB
Latest product

Spin-Torque – Next Generation MRAM

Current Toggle MRAM uses a magnetic field for switching

Limits scaling due to constant magnetic field

Next generation MRAM enables scaling to Gb densities

Everspin on track to deliver industry’s first ST-MRAM
MEMS-based Storage

- Magnetic “sled” floats on array of read/write heads
  - Approx 250 Gbit/in²
  - Data rates:
    IBM: 250 MB/s w 1000 heads
    CMU: 3.1 MB/s w 400 heads
- Electrostatic actuators move media around to align it with heads
  - Sweep sled ±50µm in < 0.5µs
- Capacity estimated to be in the 1-10GB in 10cm²

See Ganger et all: http://www.lcs.ece.cmu.edu/research/MEMS
IBM Millipede

The "millipede" project
A nanomechanical AFM-based data storage system

Project overview

The "millipede" high-density data storage system is based on micromechanical (MEMS*) components borrowed from atomic force microscopy (AFM). Tiny depressions created with an AFM tip in a polymer medium represent stored data bits that can then be read back by the same tip. Data written in this way can also be erased using the same tip, and the polymer medium can be reused thousands of times. This thermomechanical storage technique is capable of achieving data densities exceeding 1 Tb/in², well beyond the expected limits of magnetic recording.

Although the read-back rate of an individual probe is limited, high data rates can still be achieved by making use of massive parallelism. In our "millipede" system concept, an array consisting of thousands of thermomechanical probes operate in a highly parallel manner, with each individual probe capable of reading, writing and erasing data in its own small storage field. The read- and write-array can be fabricated in large quantities using well-established, low-cost semiconductor microfabrication techniques.

*MEMS: Micro-electrical-mechanical system.
Nanochip produced Millipede memory devices until May 2009.
A resistor-based approach has been developed as a basis for a new nonvolatile memory that is potentially denser, faster, and easier to make than Dynamic RAM. It relies on phase transitions induced by nanosecond-scale heating and cooling of small volumes of chalcogenide films within the memory cell. Initial target markets include FLASH memory, embedded memory, and DRAM.
As early as the 1950s, the semiconducting properties of a range of crystalline and amorphous chalcogenide alloys were investigated. In the early 1960s, new reversible phase change materials and electrically and optically programmable devices were reported, and these devices were proposed for use in digital computers as non-volatile memory. 650 MByte PD and CD-RW disks and 5.2 GByte DVD-RAM optical memory disks using a laser-induced structural phase change in a chalcogenide alloy are now in production. A number of recent developments have made it possible to use phase-change semiconductor memory devices as practical high-performance memory elements.

- Alloys developed for use in optical disk applications have been shown to be well suited for use in semiconductor memory applications.
- A more detailed understanding of the device behavior has developed that permits the engineering of practical memory devices.
- Lithographic scaling and processing improvements have reduced the programming requirements of new cell designs, which can now be driven by minimum geometry MOS transistors.

It is now believed that chalcogenide technology can provide a lower cost and higher performance memory than EEPROM and DRAM. Furthermore, embedded memory applications can readily be achieved due to the simple, planar structure of the storage element.
The phase conversion is accomplished by heating and cooling the material. When melted it loses all crystalline structure, and rapid cooling below glass transition temperature causes the material to be locked into its amorphous phase. This phase is very stable near room temperature, but the rate of nucleation and growth of crystallites increases exponentially as the melting temperature is approached. To switch the memory element back to its conductive state, the material is heated to a temperature between the glass transition temperature and the melting temperature, causing nucleation and crystal growth to rapidly occur over a period of several nanoseconds.

The crystal nucleation and growth rate depends on the chalcogenide alloy composition and can vary by more than 20 orders of magnitude among various materials. Changes in the electrical properties of these alloys are far more dramatic, with variations in the electrical conductivity by up to 6 orders of magnitude between phases.
To create an electronic memory from these materials, an array of access transistors must each be capable of providing sufficient power to a memory element to melt a portion of the chalcogenide alloy. Thermal isolation of the memory element itself from the heat-sinking substrate and metallization is a crucial aspect of the memory element design.

Three modes of operation –

• **Read** - the electric field is limited by applying a low voltage. A small current will pass if the material is in the amorphous state; in the crystalline phase, the applied voltage and the resistance of the contact will limit the current through the device.

• **Set** - the voltage must be high enough to ensure that the alloy will switch into a low impedance state. An intermediate current level will heat the material but not melt it.

• **Reset** - the voltage must be high enough to ensure that the alloy will switch into a low impedance state with sufficient current to heat a portion of the material above its melting temperature. When the current is removed, the small volume of material that has melted will rapidly quench into the amorphous state.
One structure currently being investigated is shown above. The bottom electrode contacts the chalcogenide alloy material, forming a ring-shaped contact area. The chalcogenide alloy is deposited in the highly conductive crystalline phase. The region where phase transition occurs is limited to the chalcogenide material immediately adjacent to the lower electrode. This reduction in the volume of material being melted reduces the power requirement sufficiently to allow a minimum-feature-size MOS transistor to easily control the device.

Detailed modeling of the electrical, thermal, and phase-transition behavior of these device structures has suggested numerous alternative structures that will have improved reliability and that will integrate well into a standard CMOS logic or memory process.
Chalcogenide-alloy, phase-change memories can provide competitive alternatives to established semiconductor memory products. Additionally, chalcogenide-based memories offer nonvolatile operation, DRAM speeds, no practical cycling limitations, higher integration density, and lower manufacturing cost than any established technology. Inherent radiation hardness of the basic material makes this technology an ideal candidate for space electronics applications.

These memory elements are fundamentally different from other semiconductor memories; information storage is achieved through changes in electrical resistivity rather than through manipulation of exceedingly small amounts of charge.

Efforts to date have focused on reducing the power needed to switch the memory element to allow operation by a single, minimum-feature-size MOS transistor, thereby allowing maximum memory density integration. Extensive modeling supports the achievement of this goal. New devices are currently being fabricated to demonstrate the feasibility of integrating planar resistive elements with MOS transistors. Device modeling suggests that these devices will meet the goal of low power switching and simple lithographic-dependent scalability.
Latest Product

July 2012: Micron announces availability of Phase-Change Memory for mobile devices - the first PRAM solution in volume production

Micron makes two types of PCM – serial and parallel

Serial densities range from 32 MB to 128 MB
Parallel densities are fixed at 128 MB
January 2014 - Micron has removed “128-Mbit 90nm serial and parallel NOR pin-out PCM devices from the products listed on its website".

PCM isn’t addressed in blocks and is faster than NAND. It is positioned as a potential post-NAND successor when flash process technology cannot be shrunk further, roughly beyond 15-10nm.

Micron's latest 3D NAND project, however, is seemingly ready for rollout. The technology aims to extend the life of 2D or planar NAND technology by placing multiple layers of it on a single chip, creating a 3D structure. It hopes to start production sampling of 3D NAND chips in the second quarter of this year, with general availability happening some time in 2015.

Using the 3D NAND process means the firm does not have to undergo the expense of a wholesale move to PCM production technology and testing equipment. This may be the underlying reason for the (temporary) withdrawal from PCM tech by Micron.

IBM demonstrates next-gen phase-change memory that’s up to 275 times faster than your SSD

PCMs, built on 90nm CMOS and at extremely low density (modern NAND flash is now available in 512Gbit sizes compared to 128Mbit for PCM) is a full order of magnitude faster than commercial NAND, with vastly superior write performance and data longevity.

IBM makes a point of noting that its PSS solution uses 90nm memory produced by Micron. The only problem? Micron gave notice earlier this year that it was cancelling all of its PCM production and pulling out of the industry.

Right now, PCM is the most promising next-generation memory technology on the market — but if no one steps forward to manufacture it, it’s going to be a tough sell.
A memristor is a non-linear passive two-terminal electrical component relating electric charge and magnetic flux linkage. It was envisioned, and its name coined, in 1971 by circuit theorist Leon Chua. According to the characterizing mathematical relations, the memristor would operate in the following way:

- The memristor's electrical resistance is not constant but depends on the history of current that had previously flowed through the device, i.e., its present resistance depends on how much electric charge has flowed in what direction through it in the past; the device remembers its history — the so-called non-volatility property. When the electric power supply is turned off, the memristor remembers its most recent resistance until it is turned on again.
- In 2008, a team at HP Labs claimed to have found Chua’s missing memristor based on an analysis of a thin film of titanium dioxide thus connecting the operation of RRAM devices to the memristor concept.
- These devices are intended for applications in nanoelectronic memories, computer logic and neuromorphic/neuromemristive computer architectures.
- In March 2012, a team of researchers from HRL Laboratories and the University of Michigan announced the first functioning memristor array built on a CMOS chip.

The memristor definition is based solely on the fundamental circuit variables of current and voltage and their time-integrals, just like the resistor, capacitor and inductor. Unlike those three elements however, which are allowed in linear time-invariant or LTI system theory, memristors of interest have a dynamic function with memory and may be described as some function of net charge. There is no such thing as a standard memristor. Instead, each device implements a particular function, wherein the integral of voltage determines the integral of current, and vice versa. A linear time-invariant memristor, with a constant value for $M$, is simply a conventional resistor. Manufactured devices are never purely memristors (ideal memristor), but also exhibit some capacitance and resistance.