Semiconductor Memories

Xiang Yu, Department of Electrical and Computer Engineering

Abstract—A brief overview on memory architecture and processes for several conventional types within the MOS family. Current trends and limitations would be discussed before leading to some insight on the next generation of memory products.

Index Terms—Random Access Memories, SRAM, DRAM, Flash memory, FRAM, MRAM, PRAM.

I. INTRODUCTION

The term memory usually refers to storage space or chips capable of holding data. Semiconductor memory is computer memory on an integrated circuit or chip. The present day MOS (metal oxide semiconductor) family now consists of SRAMs (static random access memories), DRAMs (dynamic random access memories), ROMs (read only memories), PROMs (programmable read-only memories), EPROMs (erasable programmable read-only memories), EEPROMs (electrically erasable programmable read-only memories), and flash memory [1].

For the past three and a half decades in existence, the family of semiconductor memories has expanded greatly and achieved higher densities, higher speeds, lower power, more functionality, and lower costs [2,3]. At the same time, some of the limitations within each type of memory are also becoming more realized. As such, there are several emerging technologies aiming to go beyond those limitations and potentially replace all or most of the existing semiconductor memory technologies to become a Universal Semiconductor Memory (USM). The quest for such a USM device has long been an objective for technical and research communities. In addition, the rewards for achieving such a device would be to gain control of an enormous market which has expanded from computer applications to all of consumer electronic products as shown from Fig. 1 [3,4]. Some of those emerging technologies include FRAM (Ferroelectric RAM), MRAM (Magnetic RAM), and PRAM (Phase-change RAM).

II. CONVENTIONAL SEMICONDUCTOR MEMORIES

Semiconductor memories started growing when Intel released the 1103 chip, the first commercially available 1kb DRAM chip, back in 1970. The technology has evolved successfully along, following the trend characterized by Moore’s law.

Currently with SRAM having the highest performance in speed, they get to occupy the immediate memory space in computers, communicating with the central processing unit (CPU) as cache memory. However, a single SRAM cell array consists of six transistors (6T) in order to hold one bit as shown in Figures 2 and 3 (a). This becomes a big physical limiting factor in terms of scaling optimization. In addition, there is the issue with data stability with the faster speeds and having leakage standby current [3].

Manuscript received December 12, 2006; revised December 18, 2006.
X. Yu is with the Electrical & Computer Engineering Department, University of Rochester, Rochester, NY 14627 USA (e-mail: althalus13@gmail.com).

Fig. 1: Diversified memory demands from various groups of users

Fig. 2: Cost-Performance map of conventional semiconductor memories including hard disk drives

Next, with DRAM having lower speed performance and costs, they serve predominantly as main memory within computers. A single cell DRAM array is much denser with only one pass transistor and a capacitor (1T1C) as shown in Fig. 3 (b). Being a form of volatile memory, the most critical
limiting factor for DRAM would be the data retention rate after each refresh cycle. The 1T DRAM is also approaching the limit of its cell size.

Flash memory is slightly different, because it is a form of non-volatile memory (NVRAM) unlike the first two types. Flash memories are especially slow to write, making them not cost efficient as compared to DRAM as shown in Fig. 2. Each cell of flash memory consists of just one floating gate transistor, meaning it is similar to a standard MOSFET except there are two gates. One is the control gate (CG) and the other is a floating gate (FG) surrounded by an insulating oxide layer between the CG and the substrate as shown in Fig. 4.

Development in the recent years was spurred on due to interest in its non-volatility properties and portability. In addition, consumers’ demand for applications like digital cameras, cellular phones, and flash drives have also helped drive this growth. In terms of limiting factors, the insulating oxide layers around the floating gate allows for extensive date retention time, it also forces the WRITE and ERASE functions to operate at higher voltage levels than conventional CMOS to overcome the barriers. On the other hand, by make the oxide layers were any thinner would lead to capacitance coupling from adjacent cells due to its already compact layout [2,3]. As such, there is a huge performance drop when compared with SRAM. In addition, flash memories are limited in the number of WRITE-ERASE cycles [3].

Lastly, for long term storage space, hard disk drives meet the requirements with its low costs and performances as compared to the rest as shown in Fig. 2 [3].

III. EMERGING SEMICONDUCTOR MEMORIES

To improve on some of the limitations as encountered by conventional memory, there are some emerging candidates that show some promise. With the focus being on non-volatility and portability first, there are three new types of memory, each possessing nearly ideal properties, i.e., fast random access, relatively low voltages, and much higher performance.

First, among its primary competitors, there is the FRAM or FeRAM, which uses a similar topology to DRAM cells with one transistor and one ferroelectric capacitor (1T1C) as shown in Fig. 5 (a). The material for the capacitor is typically lead zirconate titanate or PZT. By using the ferroelectric characteristic of the capacitor, binary data can be stored depending on the polarity of the cells. Ferroelectric materials have a nonlinear relationship between the applied electric field and the apparent stored charge, or more specifically, it is characterized by the form of a hysteresis loop as shown in Fig. 6. Due to this nature, the WRITE process is accomplished by applying a field across the capacitor plates and the polarization of the atoms can occur extremely fast. However, for the READ process, an initial state is first assumed. If the state is correct, then nothing happens, but if the assumed state is incorrect, another WRITE cycle has to occur after the READ cycle. While there is a high limit to the number of WRITE cycles, the possible requirement for a WRITE cycle after each READ cycle is a potential limiting factor [5]. In addition, there could also be polarization degradation within the ferroelectric capacitor over time [3]. Finally, current day FRAMs are still trying to achieve higher densities.

Another type of memory is the MRAM, using the principle of magnetism and the development of the Magnetic Tunneling Junction (MTJ) yielding some high speed operations comparable to SRAM. The cell array consists of one transistor and one MTJ (1T1MTJ) as shown in Fig. 7 (a). The MTJ is formed from two ferromagnetic plates separated by a thin insulating layer. Each of the plates can hold a magnetic field, with one set to a particular polarity and the other plate will change accordingly, depending on the external field. The most critical limitation with this technology is the large writing current generated to overcome the existing fields between the two plates. There have been modifications based upon the MJT concept, such as using a multiple layered cell
structure and a newer technique called spin-torque transfer (STT) [6]. The idea behind STT is that if the electrons in a layer have to change their spin, they will develop a torque that can be transferred to the next closest layer. In this manner, the overall amount of current need to write to the cells is lowered. Overall, the MRAM technology has similar speeds to SRAM, similar density to DRAM, much lower power consumption than DRAM, and suffers no degradation over time in comparison to flash memory.

The third type of memory is PRAM, which uses the unique behavior of chalcogenide glass, which can change between two states, either crystalline or amorphous, with the addition of heat. The two states have dramatically different electrical resistivity values, and this forms the basis by which data is stored. The amorphous state is a high resistance state and is used to represent a binary 0, while the crystalline state is a low resistance state and is used to represents a binary 1. Nearly all prototype PRAM devices make use of a chalcogenide alloy of germanium, antimony and tellurium (GeSbTe) called GST. It gets heated to over 600°C, at which point the chalcogenide becomes a liquid. Once cooled, it is frozen into an amorphous glass-like state and its electrical resistance is high. By heating the chalcogenide to a temperature above its crystallization point, but below the melting point, it will transform into a crystalline state with a much lower resistance. This phase transition process can be completed in as quickly as 5 nanoseconds, making it comparable to the speeds of DRAM [7]. The cell array for a PRAM is once again similar to DRAM with a single transistor and the variable resistance from the chalcogenide glass as shown in Fig. 7 (b). There are variations on this cell array where a diode can be used instead of a transistor to minimize sizing. Some of the limitations involve using large amounts of current to heat to chalcogenide glass.

**IV. CONCLUSION**

Despite its limitations, the field of conventional semiconductor memories would continue to flourish. All three emerging types of memory technology show promising results. Table I shows a comparison between the various technologies. In time and with some further improvements, they will attempt to gain acceptance as an alternative to flash memory, which currently dominates the consumer market. With each type having its own uniqueness, the need from other applications and future consumer demands will determine its later existences.

**REFERENCES**


**TABLE I**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>SRAM</th>
<th>DRAM</th>
<th>Flash</th>
<th>FRAM</th>
<th>MRAM</th>
<th>PRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonvolatile</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Refresh</td>
<td>No</td>
<td>ms</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cell Size</td>
<td>6T</td>
<td>1TIC</td>
<td>1T</td>
<td>1TIC</td>
<td>1TMTJ</td>
<td>1TIR</td>
</tr>
<tr>
<td>Read time</td>
<td>2 ns</td>
<td>10 ns</td>
<td>70 ns</td>
<td>10 ns</td>
<td>10 ns</td>
<td>10 ns</td>
</tr>
<tr>
<td>Write time</td>
<td>2 ns</td>
<td>10 ns</td>
<td>10 μs</td>
<td>20 ns</td>
<td>5 ns</td>
<td>5 ns</td>
</tr>
<tr>
<td>Write cycles</td>
<td>&gt; 10¹⁵</td>
<td>&gt; 10¹⁵</td>
<td>10⁷</td>
<td>&gt; 10¹²</td>
<td>&gt; 10¹²</td>
<td>&gt; 10¹²</td>
</tr>
</tbody>
</table>