PART 1

Information Storage and the State of the Art of Electronic Memories

1

General Issues Related to Data Storage and Analysis Classification of Memories and Related Perspectives

Over the past few years, the massive increase in the volume of information has encouraged the search for ways to resolve issues related to data storage and processing. The technical means for resolving these problems go hand in hand with the improvement of read–write–erase speeds, a reduction in energy consumption of electronic memory devices as well as an increase in their storage capacity. A short description of the characteristics of the different types of memories, volatile or non-volatile, reveals the existence of a technological gap between the performance of extremely rapid volatile memories and non-volatile memories, the latter being used for storage, but too slow for handling large volumes of data. The search for universal memories capable of combining storage and processing capabilities for data-use is a new field of research in the industry. In the years to come, this is expected to lead to the progressive replacement of current systems by new generations of memories with qualification characteristics equivalent to those of "Storage Class Memories" (SCMs).

1.1. Issues arising from the flow of digital information

Information storage and the continuous increase in the volume of information circulating in the world are topics that preoccupy large bodies responsible for political anticipation and decision-making. A recent report by the International Data Corporation (IDC), "The Digital Universe in 2020" [GAN 12], indicates that the volume of digital information in the world decupled between 2006 and 2011, increasing from 180 to 1,800 Exabytes (EBs)[1], and should reach an astounding 40,000 EBs by 2020, the equivalent of 5,200 Gigabytes (GBs) per human being.

This trend is not expected to slow down – the authors of the report estimate that the volume of digital information will double every two years from 2012 to 2020 – thus leading *de facto* to the consideration of issues in energy consumption inherent in the running of large servers that, for obvious reasons, are now preferably located close to centers of electrical energy production and distribution, a tendency emphasized by the recent boom in "Cloud Computing" [MEL 11].

Since the 1990s, digital technologies have taken over from analog technologies. In 2007, 99.9% of telecommunications were carried-out digitally. As of the early 2000s, the majority of information was also stored in digital mode, representing 94% of all stored information in 2007 [HIL 11].

This irreversible and ultra-fast increase in the global flux of information with, in addition to this, a strong demand for increasingly powerful computers, requires an extreme miniaturization of electronic components and memories. The end of the applicability of Moore's Law[2] is considered as imminent, and new solutions must be found to resolve issues related to information storage.

This question has already been considered by many bodies, and a very general prospective has been developed over the years with a view to the possible replacement of current components (transistors and memories), essentially founded on silicon-based technology, by

---

[1] An EB represents 10^{18} bytes or 1 billion Gigabytes

[2] Moore's Law, laid down during the boom in computing (1965), stipulates that the number of transistors implanted on a "chip" approximately doubles every two years, which means that the width of the printed circuit lines decrease by a factor of 0.7 for each generation of transistors.

new systems based on materials capable of increasing the integration density of components in the electronic circuits so as to improve energy efficiency while promoting high operational reliability.

1.2. Current electronic memories and their classification

Computers and information storage currently depend on the use of two kinds of memories: volatile and non-volatile (Figure 1.1).

Volatile memories (essentially Dynamic Random Access Memory (DRAM) and Static Random Access Memory (SRAM)), which are used to run computers, have very short execution times but, unfortunately, the conservation of data with time (retention) requires either periodical refreshing (DRAMs) or a constant power supply (SRAMs), both of which are costly in terms of energy.

Non-volatile memories, consisting mainly of hard drives (Hard Disk Drives (HDDs)) and more recently Flash memories (NOR and NAND) have retention times that are convenient for the requirements of information storage. At rest, they do not require a power supply but have read–write and erase times that are too long for logic operations. They are used for storage and are classified as Read Only Memory (ROM) memories. Magnetic memories (MRAM) are also non-volatile and very fast, and can be addressed in random access.

Among volatile memories, DRAMs and SRAMs are the main memories used to run logic operations. DRAMs have very short retention times (at the ms level or less) and therefore require periodical refreshing. SRAMs conserve information when they are connected to a power supply, and lose this information when the power is off. This last type of memory, for which write, read and erase tasks are very rapid (a few nanoseconds), is mostly used in computers as cache memory[3].

---

[3] Cache memory is temporarily saved data that is extracted from the main memory. It is a data and instruction "reserve" used to run repetitive operations, and has the advantage of being very rapidly accessible, the effect of which is to shorten execution times of certain computing operations.

Non-volatile memories, which have very high retention times, must be considered as peripheral components of the computer that do not take part in logic functions. They are used for reading information that is archived, and hence are referred to as ROM.

Within this first and very general definition, a distinction is made between memories used to store information that is considered as programmed just once, and can be read without any possible modification – One Time Programmable Read Only Memory (OTPROM) or Write One Time Read Many (WORM) – and those where the information can also be indefinitely conserved, but this time, with the possibility of modifying it when required – EPROM memories, i.e. ROM memories that can be erased and reprogrammed.

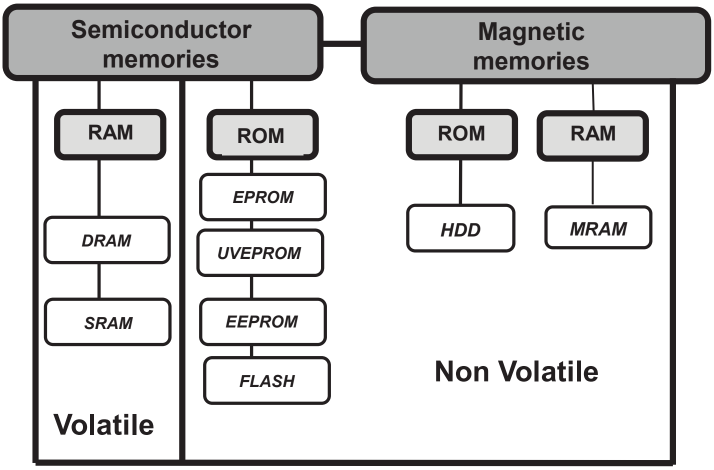

Figure 1.1. *Classification of the main current types of volatile and non-volatile memories. Adapted from Jeong et al. [JEO 12]*

The first memories of the latter type, known as UV-EPROM, which appeared on the market in the 1970s, could be erased by exposing the entire device to prolonged UV irradiation. In the 1980s the first memories appeared that could be written and erased electronically (Electrically Erasing PROM (EEPROM)) but that could also conserve the information indefinitely, therefore providing an advantageous replacement for UV-EPROM.

Flash memories, first produced in the 1980s by Toshiba, and a few years later by INTEL (1988), were derived from EEPROM memories. These memories are in fact an assembly of EEPROMs that, depending on their connection mode (parallel or series), lead to NOR Flash and NAND Flash memories. In the past few years, these memories have been the object of considerable industrial development and are considered as future storage memories, capable of competing with magnetic hard drives.

Their common feature is the local appearance or disappearance of an electrical charge trapped in a "floating" electrode, designated as a storage "node", and corresponding to processes involving "charge storage nodes" [ZHI 12].

The magnetic storage of information is without doubt the oldest procedure[4]. Magnetic hard drives (HDD) rely on a process in which the memory effect is due to the recognition of magnetic micro-domains that can be reversibly created and erased. They constitute exceptional non-volatile memories that have the advantage of allowing periodical and almost indefinite writing and erasure of data, and are able to conserve it for as long as the rotating disk is functional. The most significant event that can lead to the loss of data is a mechanical incident that crashes the read–write head onto the rotating disk, which unfortunately, like any catastrophe, occurs without warning and not infrequently. Contrary to HDDs, for which access to data is sequential, MRAM magnetic memories operate by random access and no longer have any mechanical parts, but require greater space, due to the number of leads necessary for the magnetic field, and have a higher energy consumption, for which reasons they are restricted to specific applications (see Chapter 2, section 2.4.5).

---

[4] The first device at the origin of current hard drives was produced by IBM in 1954 (350-Random Access Method of Accounting and Control (RAMAC)). It weighed over a ton, consisted of 50–60 cm-diameter disks, had a storage capacity of just 5 megabytes. In 1956 it cost more than 50,000 dollars [IBM 57].

1.3. Memories of the future

For a long time the electronic memory industry was dominated by the production of DRAMs and HDDs, and it is only recently, with the beginning and the intrusion into everyday life of portable devices of all kinds, that the share in the market of Flash memories (Solid State Drives (SSD)) has considerably increased compared to that of HDDs.

It is admitted, however, that the scale reduction of floating gate Flash memories beyond the 16 nm scale will be difficult to achieve without a significant increase in their manufacturing costs [BUR 13].

The ambition to develop ever more powerful yet less power-hungry calculators, while remaining within reasonable costs, implies that either great progress has to be made in the conception of storage hard drives or that new approaches for data storage must be considered.

It is this research effort toward new technologies that has led the industry to select a limited group of systems capable of combining the dynamic characteristics of DRAMs with storage characteristics close to those of HDDs.

The International Technology Roadmap for Semiconductors[5] (ITRS) in its 2010 editions (Emerging Research Materials, [HUT 10, ITR 11]) suggests new paths in research for the elaboration of

---

[5] The ITRS is an international body created by the Semiconductor Industry Association (SIA) that ever since its formation in 1977 has guided the majority of the major American electrical component manufacturers by proposing important R&D directions, the latter discussed and established by a committee of researchers and engineers, which has led to the birth of the National Technology Roadmap for Semiconductors (NTRS). This committee rapidly grew on the international scale and, in 1998, several countries associated themselves with the SIA to found ITRS, a body that is now sponsored by Europe, Japan, South Korea, Taiwan and the USA. Since 1999 this new international body has published an important report every year, the fruit of a widespread consensus between manufacturers, whose goal is to guide the semiconductor industry in its choice of R&D programs with a realistic vision of "emerging" systems for the 15 years to come. It is also at the origin of "More than Moore" (MtM), a concept for the development of new technologies, implying the creation of devices that combine different functionalities on the same electronic chip.

electronic memories so as to reduce energy consumption but also increase the density of chip memories. The evolution toward two-contact memories, which remains compatible with current Complementary Metal Oxide Semiconductor (CMOS) technology, is considered as being the best solution [ITR 11].

In order to better understand the challenges involved in the development of information, with the related problems, the necessity for shorter and shorter access times and for increasingly large storage capacities (in terms of time and quantity), we show in Table 1.1 a few essential properties of the most widely used types of memories that make up the major part of the market.

	SRAM	**DRAM**	**Flash (NAND)**	**HDD**
Reciprocal density *(F^2)*	**140**	**6-12**	**1-4**	**2/3**
Energy per bit *(pJ)*	**0.0005**	**0.005**	**0.00002 (a)**	**5x10^3- 10^4 (b)**
Read time *(ns)*	**0.1-0.3**	**10**	**100 000**	**5-8x10^6**
Write time *(ns)*	**0.1-0.3**	**10**	**100 000**	**5-8x10^6**
Retention	**as long as V applied**	**<<second**	**years**	**years**
Endurance *(cycles)*	**> 10^{16}**	**> 10^{16}**	**10^4**	**10^4 (c)**

Table 1.1. *Comparative study of operational characteristics of the main commercialized volatile (SRAMs and DRAMs) and non-volatile (NAND Flash and HDD) memories*

Comments on Table 1.1.– *(a) The energy required to write a bit with an EEPROM cell is of the order of 100 picojoules (pJ). The mode whereby a very large number of cells are addressed simultaneously explains why the energy is significantly reduced in Flash memories. (b) The energy per bit for an HDD is estimated from the power used for a given number of write cycles. The 5 × 10^3 and 10^4 pJ values have been calculated for a number of write cycles of 28 and 16 megabytes (MB) per watt for Western Digital (RED WD30EFRX and Black WD4001FAEX) hard drives. (c) This number of cycles (small)* *indicates the average reliability of the electromechanical system. All other data are extracted from Yang et al. [YAN 13], Jeong et al. [JEO 12] and the ITRS report [ITR 11].*

We have selected the space requirement criterion, given by the number of F^2 units taken by the memory[6], the energy in picojoules (pJ) required to write one bit of information, the time (in nanoseconds) required to write and read a bit, the retention time of the information, and endurance to cycling. Obviously, of all the memories (SRAM, DRAM, Flash (NAND) and HDD), there is not one that combines the two essential qualities: fast write and read operations at the nanosecond scale and data retention over several years.

Table 1.1 presents two groups of memories, that are not only distinct from one another in that some are volatile (DRAMs and SRAMs) and others non-volatile with retention times of several years (Flash and HDDs), but also in that the former have read–write times on the nanosecond scale, whereas those of the second group (Flash and HDDs) are considerably longer (100 μs for NAND Flash and between 5 and 8 ms for HDDs).

From the energy point of view, we note that HDDs are the most power-hungry (between 5×10^3 and 10^4 pJ, as against 2×10^{-5} pJ for NAND Flash), one of the main reasons being that the disks rotate constantly, thus considerably increasing the energy costs[7].

From the point of view of space requirements, it is SRAMs that, due to their complexity (memory consisting of 6 transistors), take up the most space ($140F^2$), followed by DRAMs (between 6 and $10F^2$).

---

[6] Integrated circuit engineers are used to compare the space required by different electronic components in a relative manner. For this, they adopt a unit of measurement F in line with currently existing technology and that corresponds to the minimum component size. As a result, in the case of "20 nm" technology, (F = 20 nm) a cell occupying an area of 40×40 nm^2 will be considered as occupying an area equal to $4F^2$.

[7] In large data archiving centers relying on HDDs the data may be split over two types of servers, those for which the reading demand is frequent and others where the data is considered as archived and rarely accessed. In the second case this distinction allows the magnetic hard drives to be switched off and therefore power consumption to be reduced.

Their high manufacturing costs and their considerable size have limited their use compared to DRAMs, which have a simpler structure and higher integration density.

All the above considerations show that there is a technological gap between the characteristics of DRAM and SRAM memories and those of Flash and HDD memories. This conclusion justifies the search for new memory devices with characteristics approaching those of DRAMs, for their high read–write speeds, and those of NAND Flash and HDDs, for their high data storage and retention capabilities.

This research is therefore directed toward the elaboration of "polyvalent" memories capable of fulfilling the logic functions of DRAMs and SRAMs as well as the archiving functions of Flash and HDD memories. These properties should lead to a group of new generation memories, qualified by Freitas and Wilcke [FRE 08] as SCM.

Hutchby and Garner [HUT 10], in the 2010 ITRS evaluation report concerning emerging materials and devices (Emerging Research Devices and Emerging Research Materials), recommend several new systems that appear compatible with the objectives of future electronic memories, on which they consider research must focus and intensify so as to allow commercial use in five to ten years.

For this, it must be demonstrated that the devices perform well, the way they work will have to be clearly established and, finally, they will have to be compatible with 16 nm[8] technologies and beyond, thus allowing a high integration density, with space requirements below $4F^2$.

As a result, systems considered as potentially viable for future use are as follows:

1) ferroelectric memories. This essentially concerns MOSFET-type (Metal Oxide Semiconductor Field Effect Transistor) and DRAM memories where the dielectric of the MOSFET and that of the

---

[8] In its 2011 edition (Executive Summary, p. 78), the ITRS expects 16 nm technology to emerge by 2015 and to be applicable for the manufacture of Flash memories.

capacitor associated with the DRAM have been replaced by a ferroelectric to make FeFETs and FeRAMs, respectively. The specific properties of these memories have already led to their commercialization for certain applications;

2) last-generation RAM magnetic memories, based on spin transfer (Spin Torque Transfer RAM or (STT-RAM));

3) nanoelectromechanical RAMs, that operate on the basis of the opening and closing of a circuit by a nanometric electromechanical commutator (Nano Electromechanical Switch – NEMS);

4) phase-change memories (PCRAM). These are the most advanced systems in terms of their commercialization, due to the fact that the materials used benefit from technological progress in the field of optical storage;

5) systems based on the use of the resistive properties of ionic oxides and solid electrolytes (RRAM). These are classified as emerging systems, but the exact nature of the mechanisms responsible for the commutative properties of metal oxides is not clear and is the object of intensive research;

6) polymer and organic molecule-based memories. The great advantage of these systems lies in the use of preparation techniques based on the handling of solutions that are therefore far less costly than those used in microelectronics. A further advantage is the great variety of organic products capable of leading to a memory effect. Finally, the possibility of reaching a resolution close to the nanometer provides perspectives with which previous systems cannot compete.

In conclusion, we can divide the previous list into two memory categories. The first ((1) to (2)) is related to new devices which can be considered as the result of improvements on existing systems; the second concerns systems based on entirely new concepts ((3) to (4)).

In line with this classification, we have structured this work into two parts: the first is dedicated to the description of current memories and their evolution, while the second focuses on new concepts of memory. The execution speed and information storage properties of some of these qualify them as SCMs, according to ITRS estimates.