Chapter 1

Introduction

Semiconductor Flash memory is an indispensable component of modern electronic systems which has gained a strategic position in the last decades due to the progressive shift from computing to consumer (and particularly mobile) products, as revenue drivers for IC (integrated circuits) companies. Today, memories are used in personal computers (PCs), cellular phones, MP3 players, digital cameras, televisions, wireless handsets, smart-media, networks, automotive systems and global positioning systems. At the core of the successful development of memories has been the steady increase of capacity, strictly linked to the shrink ability of the IC technologies.

For decades now, semiconductor scientists and engineers have been looking for a universal memory, combining low-cost, speed and capacity, non-volatility, compatibility with current complementary-metal-oxide-semiconductor (CMOS) manufacturing techniques, and scalability. Today, it has been finally recognized that different applications have different requirements, and no one solution could be optimal for a wide range of conflicting demands. Nevertheless, in the past few years, the research and development into new types of memory have dramatically accelerated. There are several reasons for this: first, the "end" of Moore's Law. Traditional semiconductor technology will not easily scale much further. How far it can be pushed is a matter of conjecture (32 nm node or some smaller nodes), but it is clear that we are close to the end of CMOS scaling. With 45 nm-32 nm semiconductor devices in volume production and 22 nm devices in the design stage, alternative technologies are the subject of a great deal of

attention worldwide. This consideration obviously also applies to microprocessors and logic. Second, there is the need for better memories. Conventional memory technology has inherent limitations. DRAM (Dynamic RAM) is cheap but volatile and requires continual refresh. SRAM (Static RAM) is faster but expensive and prone to soft errors – and it is still volatile. Flash is non-volatile but slow. Both magnetic and optical storage are high capacity, but are running up against the limits of current technology. The growing role of mobile computing and communications has led both semiconductor firms, such as Intel, and OEMs (Original Equipment Manufacturers), such as Motorola, to speculate about a future ubiquitous computing paradigm in which most of the customer equipment is untethered. There is a burgeoning demand for low cost, low power consuming non-volatile memory technologies, with a considerable improvement on today's Flash in terms of capacity and speed, since the new mobile paradigm for computing and communications will require that these new memories will need to provide rapid write and read capabilities for very large amounts of data (including video data). The third element will be memory for "disposable" electronics products, such as smart cards, RFIDs (radio-frequency identification), electronic tickets, toys, electronic greeting cards, etc., which are beginning to proliferate, as a result of the commercialization of printable and organic electronics. While capacity, speed and most of the other requirements for memory in such applications are not essential requirements, a significant cost reduction would be welcome in this market.

In this book, based on the knowledge that we have acquired over several years of experimental and theoretical research in the field of non-volatile memories, we will try to give a general overview of the different technological approaches currently studied worldwide to fulfill the requirements of future memory applications.

Although a critical economic analysis of the semiconductor industry is beyond the scope of this dissertation, as scientists working in this field, we cannot ignore the fact that we are currently living in a period of strong changes. All people working today in microelectronics (in companies, research laboratories or universities) live in a climate of challenging contradictions and constant urgency. In Chapter 2, based on historical data, we will show that semiconductor manufacturing is a highly competitive industry, where high technology, device cost and speed are the major economic driving forces. Rapid obsolescence of products (becoming obsolete in less than a year) and dramatic price decline are the main

characteristics. We will talk about the volatility of the semiconductor markets, the constant price wars among IC companies and the impact of accelerated device scaling on technology costs. Particular attention will be given to the evolution of CMOS memory technologies over the last few decades. Using this analysis, we will try to understand the main economic factors at the basis of the current consolidation of the IC industry. In fact, the semiconductor industry is currently continuing its transformation from a relatively young, high-growth industry to a more mature one, facing many of the same issues that other established industries have already faced, as their markets became more saturated and geographical patterns of supply and demand shifted. Chip-makers are redefining business strategies to compensate for a variety of difficult economic pressures, including the increasing costs of developing new technology and adding new high yield manufacturing capacities. In addition to these fundamental changes, there is an unmistakable shift of capital and research and development (R&D) investments away from Europe (and the USA) to regions which are increasingly important as an end-market for semiconductors, i.e.: South Korea, Taiwan, China, Singapore and Malaysia. In the longer term, the migration of manufacturing and R&D could mean that regions which initiated the industry in terms of innovation and capital investment risk losing their status as centers of the semiconductor sector, along with the current economic leverage that status commands. This is the reality of the new global economy facing high technology-based companies, as well as applied research laboratories, in Europe today. We are crossing a threshold where the pace of disruptive shifts is no longer inter-generational with a meaningful impact over the span of careers and future of our students. In this context, in Chapter 2, we will try to identify and understand the main factors at the origin of these phenomena, looking at the semiconductor industry and market trends from an economic perspective.

In Chapter 3, we will shortly introduce the main features and scaling limits of current Flash memory technologies. Then, the main strategy of the innovative research in this field will be presented. Today, two main research paths can be identified. To extend the classical floating gate technology to the 22 nm and possibly lower nodes, different “evolutionary paths”, essentially based on the use of new materials and of new transistor structures, can be investigated. On the other hand, to address smaller IC generations, “disruptive paths”, based on new storage mechanisms and new technologies, are envisaged. In this chapter, we will focus on the

"evolutionary approaches", paying particular attention to the results obtained in our laboratory (LETI, CEA-Grenoble/France) in recent years, in the framework of research funded by internal projects, the French government, European institutions and industrial partnerships. Note that a crucial point in the definition of the research plans has always been maintaining a good equilibrium between short-term (made in collaboration with IC companies) and long-term (developed in collaboration with fundamental research laboratories, universities) solutions. Given the large variety of technologies currently invoked as potential replacements for conventional Flash, one of the hardest tasks for a scientist working in this field, at least concerning near-middle term research (i.e. time to be in production < 10 years), is to identify the right framework of study (for example, embedded or stand-alone environment) for the different technologies, in order to be able to assess the main advantages and disadvantages, and thus to prospect future applications. The "evolutionary approaches" include new modules (i.e. discrete trap memories, and more specifically silicon nanocrystal memories), new materials (high-k materials for the interpoly layer of Flash) and innovative architectures (as FinFlash memories). Moreover, obviously targeting a longer term application, hybrid approaches, which make use of organic molecules – grafted on silicon substrates – as storage sites, have been developed. Finally, the main theoretical limits of charge storage memories (i.e. reliability issues linked to certain few electron phenomena) have been identified, opening up the path to the introduction of disruptive memory technologies based on new storage mechanisms.

In Chapter 4, we will try to define the perspective of the accomplished and main paths of research for the next few years. As discussed previously, in order to reduce bit cost and increase bit density, the shrinkage of Flash memories has been aggressively driven by reducing the cell size and introducing multi level technologies. Nevertheless, the linear scaling down of MLC NAND Flash memories is approaching its critical physical, electrical and reliability limits. Conventional cost-reduction approaches, notably smaller design rules, are having less effect, and new approaches should be considered. According to the main IC companies, the short-term way to circumvent these barriers is stacking memory cells on a single Si wafer. In this context, innovative integration paths, suitable for three-dimensional integration of IC memory circuits, and new design/system solutions are introduced. On the other hand, "new breakthrough memory technologies", such as phase-change memories, resistive RAMs, insulator

and organic polymer cross-bar memories, are also considered as possible candidates for future memory applications. In the last part of this chapter, we will give a general overview of their main advantages and disadvantages in view of future memory applications.

In the conclusion, we will summarize the main findings of the previous chapters and make some general considerations on innovation paths in the field of memories and on the future of microelectronics technologies as a whole.