

The Non-Volatile Memory Industry – A Personal Journey

Eli Harari

Abstract— In this keynote address, the author recalls his personal involvement as a device physicist from the early days of the non-volatile memory industry, then through the 22 years that he served as CEO of SanDisk Corp., a company he co-founded in 1988. (Abstract)

Keywords— flash memory, SanDisk, Eli Harari

I. INTRODUCTION

I am honored to address you today at the non-volatile memory industry's premier technical conference. I recently retired from my position as Chairman and CEO of SanDisk, so I am here today as a private citizen representing my own personal opinions rather than as a representative of SanDisk.

A. The Early Years

When I started my career in semiconductors in the early 1970's, the non-volatile memory industry was in its infancy, and none of us could have imagined what a phenomenal impact it would make over the coming several decades. In my talk today I will first discuss my research work in the 1970's on thin films of SiO₂, which became, I believe an important element of future EEPROM and Flash EEPROM. I will then focus on the early years at SanDisk, sharing the initial challenges that we faced and the radical solutions that we invented, which helped establish the Flash data store industry. It is impossible to do justice to 40 years of non-volatile memory in just 30 minutes, therefore please excuse the superficial coverage and the omission of the many key contributors that made our industry such a disruptive force.

In 1969, just ten years after Bob Noyce and Jack Kilby invented the integrated circuit, I came to America to do my graduate work in Solid State Sciences at Princeton University. The title of my 1973 PhD thesis was "Charge Trapping Effects in Thin Films of Al₂O₃ and SiO₂". I could not have hoped for a better way to prepare myself for my lifelong career in non-volatile semiconductor memories.

My second lucky break came at my first job as a technologist at Hughes Microelectronics. In 1975, I was exploring ways to electrically erase the ultraviolet erasable EPROM invented by Frohman Benckowski at Intel, a device which became indispensable for storing the microcode in early microprocessors and microcontrollers. In those early MOS days, the SiO₂ gate oxide thickness was in the range of 700-1000 Angstroms. By exploring much thinner SiO₂ films in the 100 Angstrom range I was able to induce, under high electric field conditions, electron conduction through these films. The

responsible phenomenon was indirect tunneling from the silicon conduction band to the SiO₂ conduction band, a quantum mechanical effect known as Fowler-Nordheim Tunneling, first discovered by Fowler and Nordheim quite a few years earlier. This mode of conduction was important because 100 Angstrom films of SiO₂ would support long-term retention of electrons trapped on a floating gate, in sharp contrast to the 25-30 Angstrom films of SiO₂ that were employed in those days in MNOS/SONOS for direct tunneling of electrons from the silicon conduction band into interface traps, but were generally unreliable and incapable of long-term retention of trapped electrons. Not much was known in 1975 about SiO₂ films in the 100 Angstroms thickness range. They were thought to be unreliable, and suffered from leakage and dielectric breakdown under high electric field.

Early results on floating gate capacitors looked very good, and over the next two years at Hughes, we undertook a comprehensive research that encompassed more than ten thousand samples of thermally grown SiO₂ in the thickness range 33-160 Angstroms. This work was published in the Journal of Applied Physics in 1978 [1]. The new data showed that SiO₂ films in the 100 Angstroms thickness range could be thermally grown with high quality and were highly reliable as insulators. When high electric fields were applied across these films, they exhibited excellent electron conduction, however they consistently exhibited catastrophic breakdown after a certain amount of charge was passed through them. We showed that this was intrinsic oxide breakdown, unrelated to pinholes or other imperfections or contamination. Breakdown was invariably induced by electrons trapped in cumulatively generated new oxide traps (Fig.1) under sufficiently high applied field conditions.

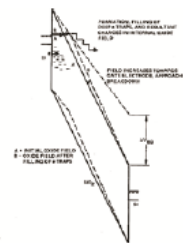


Figure 1. Fowler-Nordheim electronic conduction and trapping in thin film SiO₂.

In the course of this work, we developed new experimental techniques to characterize and optimize these thin SiO₂ films.

Applying constant current conditions, we established two key parameters: TBD, the time to breakdown under given constant current conditions, and ΔV_{BD} , the voltage shift V_{FB} required to maintain the current constant up until breakdown, and which is representative of the cumulative trapped charge, QBD at the onset of breakdown. Fig. 2 shows typical sharp distributions of these parameters.

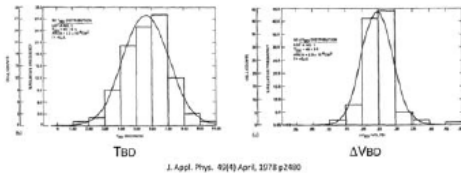


Figure 2. Typical SiO₂ film (45±5 Angstrom) breakdown distributions.

These findings enabled the development at Hughes of the first practical floating gate EEPROM, US Patent 4,115,914, filed Mar. 26, issued Sept. 26, 1978, (Fig. 3) which relied on a thin SiO₂ layer 59 to permit writing and erasing of floating gate 56 through Fowler-Nordheim tunneling under appropriate voltage conditions applied between control gate 62 and substrate 48.

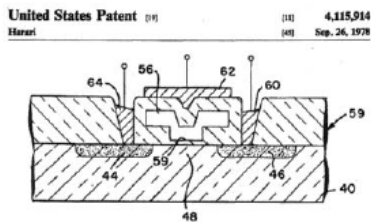


Figure 3. Floating gate EEPROM employing Fowler-Nordheim tunnel erase.

In the mid-1980's, Toshiba's Masuoka invented NAND Flash, which to this day, almost 25 years later, relies on SiO₂ films in the thickness range 70-85 Angstrom to write and erase each NAND cell through the very same Fowler-Nordheim tunneling mechanism.

II. SANDISK – A NEW BEGINNING

A. Engineering Milestones

Fast forward ten years to June 1988 when SanDisk (then SunDisk) was founded in Santa Clara. The three founders were Jack Yuan- our process engineer, Sanjay Mehrotra- our memory chip designer, and myself- our device physicist and CEO. We were joined by Bob Norman- our system architect. Over the next decade, the talented, small and highly dedicated team that we put together broke new grounds and created some of the fundamental data store Flash architecture and techniques that became the foundation of today's Flash industry.

Many of SanDisk's early patents were transformative: US patent 5,095,344, filed June, 8, 1988, issued Mar. 10, 1992, titled "Highly Compact EPROM and Flash EEPROM devices",

taught a Flash EEPROM having a very high storage density (each cell storing more than one bit of information) and a long life (adaptive program and erase voltages to minimize oxide stress during cycling), "making it particularly useful as a solid state memory in place of magnetic disk storage devices in computer systems."

US Patent 5,172,338, filed April 13, 1989, issued Dec. 15, 1992, titled "Multistate read and write circuits and techniques", taught massively parallel programming of flash cells by applying stepped programming to designated multilevels while inhibiting further programming from correctly verified cells. These massively parallel program-inhibit circuit techniques became particularly powerful for overcoming write-speed bottlenecks. Without these techniques, practical multistate NAND (two, three and four bits per cell) would not be possible today.

B. A Radically New Concept – System Flash

US patent 5,297,148, filed Apr. 13, 1989, issued Mar. 22, 1994, titled "Flash EEPROM System" (Fig.4), teaches "A system of Flash memory chips with controlling circuits serves as a non-volatile memory such as that provided by magnetic disk drives", one of whose characteristics is the "ability to remap and replace defective cells with substitute cells. The remapping is performed automatically as soon as a defective cell is detected. When the number of defective cells becomes large, the whole sector is remapped." This represented a radically new concept, which we called "system Flash", characterized by a new Flash memory chip architecture, working cooperatively with a dedicated intelligent hardware/firmware controller. Our rationale was that Flash memory was inherently unreliable, and therefore to make it 100 percent reliable to the user, it needed to have a closed-loop controller to constantly monitor and repair the memory arrays for the entire operating lifetime of the storage device.

A key feature of the '148 patent was a new physical partitioning of the flash memory arrays, resembling the way disk drives are partitioned. Here, Flash memory cells were arranged together in erasable sectors, with each sector having a partition dedicated to storing user data and a "header" partition provided for use solely by the controller for managing all memory cells and sectors and keeping track of the ever-changing file addresses and links during the lifetime of the device.

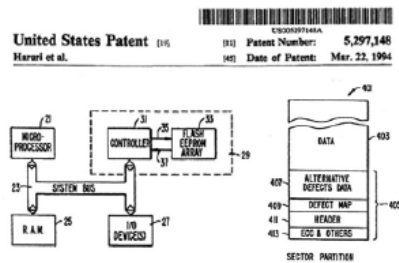


Figure 4. SanDisk's Flash EEPROM system.

III. SANDISK'S STRATEGY

A. Controller Technology

From day one, SanDisk's strategy was to develop and sell complete Flash system solutions (Fig.5) consisting of proprietary Flash memory chips with physical sectors and headers, managed by a dedicated intelligent controller, and emulating rotating memory that was optimized for storing data and content rather than microcode.

The controller's role was to format the memory arrays, randomly access sectors and headers, implement on the fly ECC, apply dynamic defect management, enhance Flash system endurance by minimizing program/erase voltages based on the device's life cycle, level out the wear-out by tracking sector level "hot count", write and verify multilevel cells, provide programmable reference cells, and numerous other proprietary tricks.

The controller also provided file management and standard I/O interface that rendered it plug and play compatible, for example with IDE Hard Disk Drives. A major side benefit was that our frequent introduction of new generations of Flash memory, or transitioning from single bit cells to multibit cells, were easily accommodated by controller firmware changes that were transparent to the user.



Figure 5. Data flash is born (1988) – used in a customer presentation.

B. The First Product

Our first product was developed for IBM. It was a 20 megabyte (MB) Solid State Drive (SSD) (Fig.6) that sold for \$50 per MB in 1991 and was a plug and play replacement for a 20MB ATA Hard Disk Drive made by Conner Peripherals. We sold ten thousand of these to IBM for their first ThinkPad tablet PC, but their product was ahead of its time and they killed it. Today we pack an entire 64 gigabyte (GB) SSD into a single postage stamp-sized package weighing one gram and selling for less than \$2 per GB.

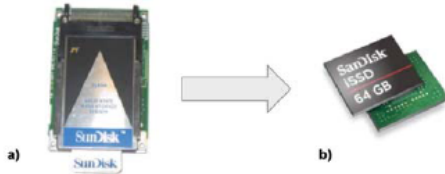


Figure 6. a) World's first Flash SSD, a 20MB SSD that sold for \$50/MB; b) New 64GB SanDisk integrated solid state drive (iSSD), selling for less than \$2/GB.

C. The Competition

In our early years, SanDisk's competition came from well-entrenched competitors who were selling stand-alone Flash chips and promoting software controllers residing on a host (Fig.7). These solutions, although initially cheaper, proved unwieldy because of incompatibility between different hosts as well as between the various Flash chip suppliers.

- Intel+ Microsoft: Linear Flash (with Msystems FFS/TFFS)
- Intel/Conner: 2.5 inch ATA SSD JV
- Intel/AMD: Miniature Card (with Olympus)
- Toshiba: Smartmedia (with Fujii)
- Sony: First Memorystick

- We overcame competition by:
 - Create and license open industry Standard Formats
 - Partner with industry leaders, e.g Canon/Kodak (CF), Matsushita/Toshiba (SD, microSD)
 - Relentlessly drive down cost of system overhead

Figure 7. Facing competition: alternative Flash solutions.

For our part, we sought to license our patented concepts to any and all competitors, believing that good competition would accelerate the creation of new Flash markets. However, for at least our first five to seven years in business, our concepts were by and large ridiculed by our competitors as well as venture capital investors who saw the high cost of a hardware controller as a non-starter. Nonetheless, we firmly believed that by relentlessly driving Moore's Law to both our proprietary flash memory as well as to our hardware controller, we would bring the system cost down to the point that it would no longer be at a cost disadvantage. We were correct, and eventually our Flash system concepts prevailed and were widely adopted, particularly with the advent of NAND flash.

D. Controller Investments Pay Off

SanDisk's patient investments in proprietary controllers paid off handsomely, as shown in Fig.8. I believe that sophisticated controllers will become ever more critical for future scaling of NAND flash to 20 nanometer and below.

- ~1990: 2.5 in. ATA SSD: 12 chips, cost ~\$70, write ~35KB/s
- ~1993: PC Card: 3 chips, cost ~\$ 5, write ~.5 MB/s
- ~1995: CompactFlash: 1 chip, cost ~\$ 3, write ~1 MB/s
- Current: SD, USB: 1 chip, cost < \$.3, write > 10MB/s

- In past ~20 years our controllers:
 - Cost down ~250X
 - Performance up ~300X
- In coming decade, controller will become even more critical to further scaling

Figure 8. SanDisk's Flash system controller evolution.

E. Moore's Law Applied to Flash Scaling

Low cost is the most important requirement for mass consumer markets to take off, and Moore's law as applied to Flash scaling was executed with great passion and zeal at SanDisk. In the past 20 years we were able to introduce 15 new generations of flash technology together with our technology and manufacturing partners, with each generation doubling the number of memory bits per chip. These cost reductions were

accelerated through the successful commercialization of 2 bits per cell (X2) and 3 bits per cell (X3) technology. Today the vast majority of NAND flash memory bits sold by all suppliers relies on SanDisk's patented X2 and X3 technology. We achieved a cumulative 25,000X cost and price reduction over 20 years (Fig. 9) meaning that a Flash storage product that we sold in 1990 for \$250 would sell today for about one cent.

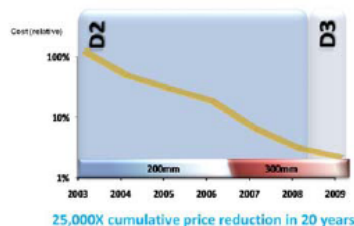


Figure 9. Moore's law drove dramatic flash cost reductions.

IV. ABANDONING NOR, ADOPTING NAND

By 1998, we were using every trick in the book to scale our proprietary triple poly 256Mbit serial NOR flash below 0.4 micron. We concluded that we had just one more generation before the end of the runway, so we set out to identify and develop a more scalable Flash technology.

A. Strategic Manufacturing Partnerships

Eventually we concluded that NAND was the most attractive candidate for further scaling, and we approached Toshiba, the original NAND inventor. Toshiba had lost its NAND partner, Samsung, and was equally eager to acquire our MLC patents and know-how. In 1999 we signed our first 50-50 NAND joint venture with Toshiba. By 2001 we were fully converted to NAND flash, including transforming 100 percent of our products and controllers to this new technology. In 2002, SanDisk and Toshiba introduced the world's first 1Gbit NAND MLC chip. Today we are engaged in three joint venture mega-fabs that are 100 percent dedicated to NAND manufacturing, and we currently jointly manufacture 64Gbit X3 chips. This is one of the longest lasting and most successful partnerships in the history of the semiconductor industry.

B. Why NAND Won Over NOR

NAND Flash was initially considered to be a niche technology. It was serial, with very limited applications such as voice recorders. However, NAND proved to be ideally suited for data store applications primarily because of its inherently superior scalability. This scalability came primarily from two sources: first, it offered highly regular memory arrays with highly efficient contacts and therefore almost a cross-point $4F^2$ cell. Second, it relied on highly efficient, extremely low current Fowler-Nordheim tunneling for both write and erase. This allowed the implementation of massive parallelisms in write and erase operations, which dramatically improved performance compared to NOR flash, yet was easy on battery life for portable applications. NAND flash quickly embraced MLC, which represents approximately 90 percent of the

NAND bits shipped today. NAND flash today is by far the dominant flash storage technology.

C. Disruptive Flash Technology

In just the past 15 years, flash film cards completely replaced the beloved 35 mm roll of film based on superb Silver Halide technology perfected by Kodak and Fuji over 80 years. USB Flash drives completely replaced floppy disk, Zip drives and magnetic tape in consumer electronics. Flash enabled consumers to store all kinds of personal and premium content in mobile phones, smartphones and tablets, including music, video, games, maps and apps. (Fig.10). Flash is currently making serious inroads in replacing optical disk, notebook HDD, and enterprise SSD.



Figure 10. Inventing new markets for flash.

V. FUTURE CHALLENGES AND OPPORTUNITIES

Our industry now faces significant manufacturing and device challenges. In the years ahead, NAND scaling below 20nm is believed to be feasible, but will not come as easy as in past years, and this may slow the rate of future cost reductions.

Extreme Ultra Violet (EUV) lithography is a new, highly complex and costly manufacturing technology that will be crucial for continued memory scaling as we push farther below 20 nanometer.

While existing NAND fabs will likely continue to flourish throughout this decade, every competitor in the field is exploring various candidates for a successor technology to NAND. 3D Read/Write and various flavors of vertical NAND are potential successors, but there is currently no clear heir apparent that can seriously challenge 20 nanometer NAND MLC.

The good news for this conference is that for the foreseeable several years the Flash storage industry is pretty much set to continue to grow at a hefty pace, driven by insatiable demand from consumers everywhere, who want their personal information in their pocket, accessible at any and all times.

"In the coming decade Flash will be bigger than you think."

Thank you.

REFERENCES

- [1] E. Harari, "Dielectric breakdown in electrically stressed thin films of thermal SiO₂," Appl. Phys. 49(4), pp.2478-2489, April 1978