10  FLASH TECHNOLOGY

Overview

Flash memory technology is a mix of EPROM and EEPROM technologies. The term “flash” was chosen because a large chunk of memory could be erased at one time. The name, therefore, distinguishes flash devices from EEPROMs, where each byte is erased individually.

Flash memory technology is today a mature technology. Flash memory is a strong competitor to other memories such as EPROMs, EEPROMs, and to some DRAM applications. Figure 10-1 shows the density comparison of a flash versus other memories.

How the Device Works

The elementary flash cell consists of one transistor with a floating gate, similar to an EPROM cell. However, technology and geometry differences between flash devices and EPROMs exist. In particular, the gate oxide between the silicon and the floating gate is thinner for flash technology. It is similar to the tunnel oxide of an EEPROM. Source and
drain diffusions are also different. Figure 10-2 shows a comparison between a flash cell and an EPROM cell with the same technology complexity. Due to thinner gate oxide, the flash device will be more difficult to process.

![Flash Memory Cell Versus EPROM Cell](image)

The electrical functionality of the flash memory cell is similar to that of an EPROM or EEPROM. Electrons are trapped onto the floating gate. These electrons modify the threshold of the storage transistor. Electrons are trapped in the floating gate using Fowler-Nordheim tunneling (as with the EEPROM) or hot electron injection (as with the EPROM). Electrons are removed from the floating gate using Fowler-Nordheim tunneling. Figure 10-3 summarizes the different modes of programming.

**Construction**

As with other memories, the flash memory chip size is the major contributor to the cost of the device. For this reason, designers have developed alternative memory array architectures, yielding a trade-off between die size and speed. NOR, NAND, DINOR, and AND are the main architectures developed for flash memories.

**NOR Cell**

The NOR architecture is currently the most popular flash architecture. It is commonly used in EPROM and EEPROM designs. Aside from active transistors, the largest contributor to area in the cell array is the metal to diffusion contacts. NOR architecture
requires one contact per two cells, which consumes the most area of all the flash configuration alternatives. Electron trapping in the floating gate is done by hot-electron injection. Electrons are removed by Fowler-Nordheim tunneling. The world’s two leading manufacturers of flash devices (Intel and AMD) use NOR cell configurations.

<table>
<thead>
<tr>
<th>Electron Trapping</th>
<th>Electron Removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-polarity FN-t Write / Erase technology</td>
<td>Fowler-Nordheim tunneling</td>
</tr>
<tr>
<td>Fowler-Nordheim tunneling</td>
<td></td>
</tr>
<tr>
<td>Hot-Electron injection and FN-t technology</td>
<td>Fowler-Nordheim tunneling</td>
</tr>
<tr>
<td>Hot-Electron Injection</td>
<td></td>
</tr>
</tbody>
</table>

Source: ICE, "Memory 1996"

Figure 10-3. Comparison Between the Different Types of Flash Programming

**NAND Cell**

To reduce cell area, the NAND configuration was developed. Figure 10-4 shows the layouts of NOR and NAND configurations for the same feature size. The NAND structure is considerably more compact.

A drawback to the NAND configuration is that when a cell is written, the sense amplifier sees a signal eight times weaker than for a NOR configuration. This is because eight transistors are in series. The weak signal slows down the speed of the read circuitry, which can be overcome by operating in serial access mode. This memory will not be competitive for random access applications. Figure 10-5 shows a speed comparison of NOR and NAND devices.
DINOR Cell

DINOR (Divided bit-line NOR) and AND architectures are two other configurations that attempt to reduce die area compared to the conventional NOR configuration. Both architectures were co-developed by Hitachi and Mitsubishi.

The DINOR design uses sub-bit lines in polysilicon. Mitsubishi states that its device shows low power dissipation, sector erase, fast access time, high data transfer rate, and 3V operation. Its device uses a complex manufacturing process involving a 0.5µm CMOS triple well, triple-level polysilicon, tungsten plugs, and two layers of metal. Figure 10-6 shows the DINOR architecture.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>NOR</th>
<th>NAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Access Time</td>
<td>80ns</td>
<td>20µs</td>
</tr>
<tr>
<td>Serial Access Time</td>
<td>—</td>
<td>80ns</td>
</tr>
</tbody>
</table>

Source: ICE, "Memory 1996"
AND Cell

With AND architecture, the metal bit line is replaced by an embedded diffusion line. This provides a reduction in cell size. The 32Mbit AND-based flash memory device proposed by Hitachi needs a single 3V power supply. In random access mode, the device is slower than a NOR-based device. Hitachi’s device is specified to operate with a 50ns high-speed serial access time. Figure 10-7 presents a review of the different flash architectures.
Audio NAND Flash

Toshiba, Samsung, and National Semiconductor have introduced 4Mbit serial audio NAND flash devices. Their devices used the NAND cell configuration. These parts, used for telephone answering machines or other audio data storage, have started to replace audio DRAMs. Based on the small NAND cell, audio NAND flash use the serial access to face speed problems. Moreover, audio NAND devices are cheaper than standard NAND flash since they contain fewer functions and may contain some bad cells without affecting the audio applications.

Four-Level Storage Cell

Several companies including Intel, NEC, and Samsung, have developed a cell that is able to store four different storage levels on the same transistor. Figure 10-8 shows an example of threshold voltage distribution for four stages stored on the same transistor.

![Cell Distribution](image)

**Figure 10-8.** Threshold Voltage Distribution for Four States

Intel presented a paper on its four-level storage work at the 1995 ISSCC conference. Prototype devices have been manufactured, but Intel does not expect to commercialize these parts before 1997. In the 1996 ISSCC conference, two papers were presented based on this concept. Samsung presented a 128Mbit four-level NAND flash cell and NEC presented a 64Mbit four-level NOR flash cell.
Multi-Level Storage Cell

Development of a four-level storage cell may take considerable time because digital storage needs to be reliable. The data needs to stay valid in worst-case conditions. For audio applications, however, tolerances allow for some error. For this reason Information Storage Devices (ISD) has proposed non-volatile memories that are able to store 256 different levels on the same transistor.

Reliability Concerns

There are three primary reliability concerns of a flash memory IC. They are data retention, thin oxide stress, and over or under erasing/programming.

Regarding the erase program concern, flash ICs that use hot electron injection for trapping electrons in the floating gate are programmed (data equal to 0) by trapping electrons in the floating gate, as an EPROM.

Flash ICs that use Folwer-Nordheim tunneling for trapping electrons in the floating gate will be programmed (data equal to 0) by removing the electrons from the floating gate as an EEPROM. The reliability concern is to either over program or over erase as shown in Figure 10-9.

Applications

There are two distinct markets for flash devices: replacement of other ROM devices, and replacement of rotating media in portable system applications. Figure 10-10 shows flash memory applications.
The advantage of flash technology in the ROM market lies in flash’s in-circuit erasing capability, combined with its non-volatility. For those historically labeled ROM applications, where data is read and not changed, the trade-offs are price versus flexibility. Mask ROMs are the least expensive, but require long lead times, stable program code, and large volumes. Conventional EEPROMs are more flexible than flash because they can be reprogrammed in very small increments, but are more costly.

Flash ICs are found in systems requiring larger amounts of memory storage, and often where there is some other form of memory in the system. Here, the flash memory is “semi-permanently” storing data that is temporarily in DRAM. Flash memory is also used when a program of some reasonable size periodically requires updating.

**PCMCIA**

Magnetics and flash memory will co-exist. Magnetic memory will continue to dominate in ultra-high capacity, low cost/Mbyte applications where power, weight/size, and mechanical ruggedness are not a consideration. Flash-based mass storage will become pervasive in small, low power, portable electronic platforms, providing low power, small
size and unparalleled ruggedness/reliability and offering lowest entry cost of any mass storage. PCMCIA (Personal Computer Memory Card International Association) cards were developed for this flash mass storage application.

**Miniature Flash Card**

Miniature flash cards were developed for applications where PCMCIA storage cards will not physically fit. The main applications are for equipment needing small size storage such as PDAs, cameras, and digital audio recorders.

There are three developments, CompactFlash, Minicard and Solid State Floppy Disk Card (SSFDC) which are similar in size but employ substantially different electrical interface schemes. Figure 10-11 presents the three miniature flash card solutions.

<table>
<thead>
<tr>
<th>Proponent</th>
<th>Size (mm)</th>
<th>Memory Type</th>
<th>Capacity</th>
<th>Connector Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CompactFlash</td>
<td>36 x 43 x 3.3</td>
<td>NOR flash</td>
<td>2, 4, 12, 15Mbytes</td>
<td>50-pin subset of PCMCIA</td>
</tr>
<tr>
<td>Minicard</td>
<td>35 x 33 x 3.5</td>
<td>NOR flash, DRAM, SRAM, OTP, ROM</td>
<td>64Kbytes* to 128Mbytes</td>
<td>40-pad elastomeric</td>
</tr>
<tr>
<td>Solid State Floppy Disk</td>
<td>45 x 37 x 0.76</td>
<td>NAND flash</td>
<td>2Mbytes</td>
<td>68-pin PCMCIA with adapter</td>
</tr>
</tbody>
</table>

*Full range not available at launch

Source: EE Times/ICE, “Memory 1996” 20420A

**Figure 10-11. Miniature Flash Cards**

**CompactFlash**

CompactFlash was developed by SanDisk Corporation. The CompactFlash Association (CFA) was established in October, 1995, to promote and encourage the worldwide adoption of CompactFlash technology as an open industry standard. The CFA board members include Apple Computer, Canon, Eastman Kodak, Hewlett-Packard, LG Semicon, Matsushita, Motorola, NEC, Polaroid, Sandisk, Seagate Technology and Seiko Epson.

The CompactFlash design incorporates the ATA (AT-Attachment) interface standard, that uses the same electrical signals as PCMCIA/ATA flash cards. The first product that employed CompactFlash technology was IBM’s Palm Top PC110, which was introduced in September, 1995.
MiniCard

The Intel MiniCard incorporates a linear-addressed format like PCMCIA flash cards. This card needs host-based software to be read. This software is called Flash Translation Layer (FTL) and was developed by M Systems. MiniCards are cheaper than CompactFlash cards but need that additional software. Figure 10-12 shows the ATA configuration versus the linear configuration. Intel developed its MiniCard for high-volume consumer applications and will not support CompactFlash.

![Figure 10-12. ATA Versus Linear Flash Card](image_url)

Solid State Floppy Disk Card (SSFDC)

Toshiba’s Solid State Floppy Disk Card is based on its flash NAND cell technology. With its small die size, the NAND technology is more cost effective. Like the CompactCard this card includes an adapter to be compatible with the PCMCIA Type II cards.