10 FLASH MEMORY TECHNOLOGY

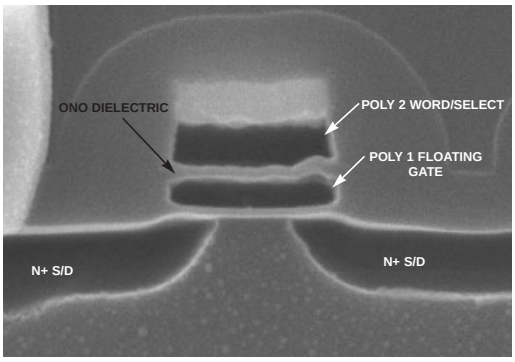
OVERVIEW

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.

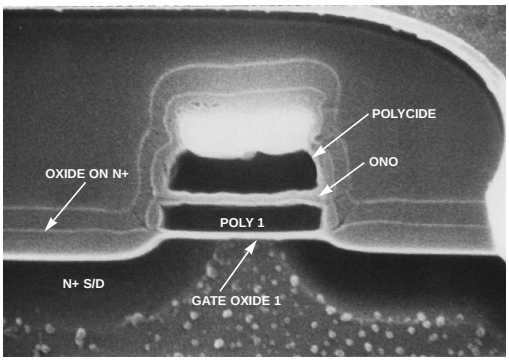
Flash memory technology is today a mature technology. It is a strong competitor to other non-volatile memories such as EPROMs and EEPROMs, and to some DRAM applications.

HOW THE DEVICE WORKS

The more common 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. Source and drain diffusions are also different. These differences allow the flash device to be programmed and erased electrically. Figures 10-1 and 10-2 show a comparison between a flash memory cell and an EPROM cell from a same manufacturer (AMD) with the same technology complexity. The cells look similar since the gate oxide thickness and the source/drain diffusion differences are not visible in the photographs.



EPROM MEMORY CELL



FLASH MEMORY CELL

Photos by ICE, "Memory 1997" 22482

Figure 10-1. AMD EPROM Versus AMD Flash Memory Cells

Type	Density	Date Code	Cell Size	Cell Gate Length
Flash	4Mbit	9406	6μm^2	0.7μm
EPROM	1Mbit	9634	5.52μm^2	0.7μm

Source: ICE, "Memory 1997" 22483

Figure 10-2. EPROM Versus Flash Cell (AMD)

Other flash cell concepts are based upon EEPROM technology. Figure 10-3 shows a split-gate cell and Figure 10-4 shows a transistor with the tunnel oxide in only a part of the oxide under the floating gate. These cells are larger than the conventional one-transistor cell, but are far smaller than the conventional two-transistor EEPROM cell.

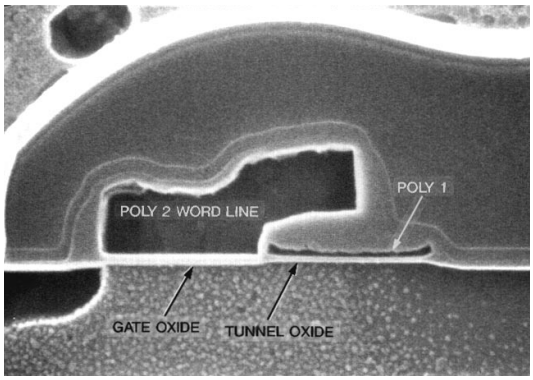


Photo by ICE, "Memory 1997" 22480

Figure 10-3. Split Gate Flash Cell

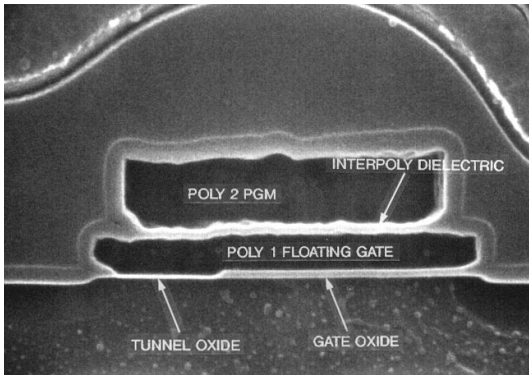


Photo by ICE, "Memory 1997" 22481

Figure 10-4. Tunnel Window Flash Cell

The electrical functionality of the flash memory cell is similar to that of an EPROM or EEPROM. Electrons are trapped onto the floating gate (see a detailed description in Section 9). These electrons modify the threshold voltage 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 as with the EEPROM. Figure 10-5 summarizes the different modes of flash programming.

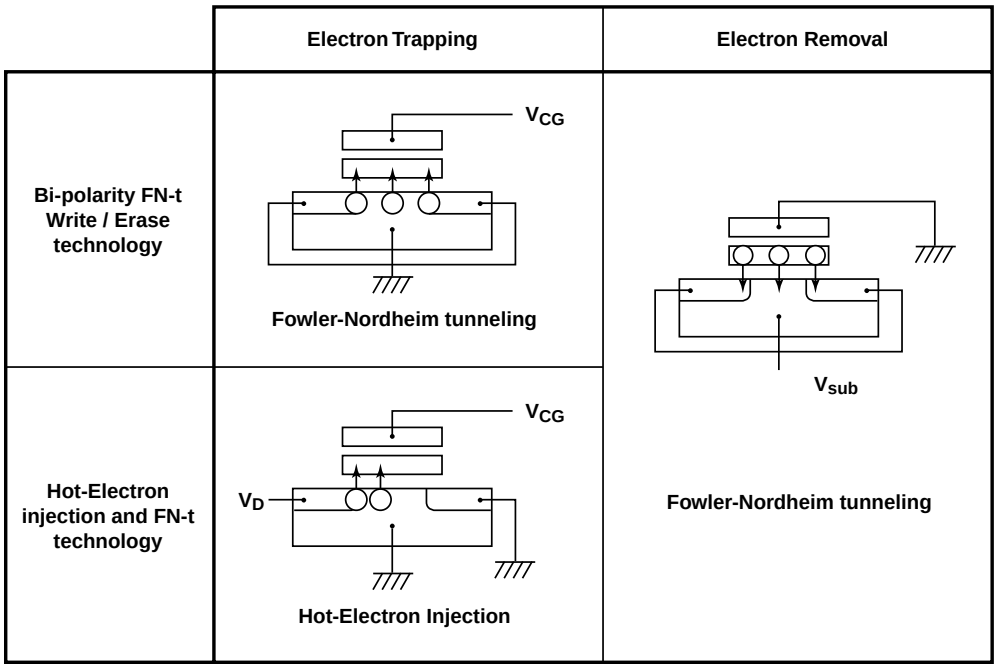


Figure 10-5. Comparison Between the Different Types of Flash Programming

Figure 10-6 summarizes chip and cell sizes of some of the flash memories analyzed by ICE's laboratory. Most of these are date coded 1994 but give a good idea of what is widely used in 1997. All these memories use the NOR flash architecture. A photo of SanDisk's 32Mbit flash cell (used on its CompactFlash cards) featuring a cell size of 1.8μm^2 is shown Figure 10-7.

ARCHITECTURE

As with other semiconductors, 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.

	Density	Date Code	Cell Size (μm^2)	Cell Type	Gate Length (μm)	Die Size (mm^2)
SST	1Mbit	9417	10.2	Split Gate	0.95	29.0
AMD	2Mbit	9325	8.0	1T	0.80	51.2
AMD	4Mbit	9406	6.0	1T	0.70	49.8
ATMEL	4Mbit	9411	16.6	Tunnel Window	N/A	107.0
INTEL	16Mbit	1993	3.3	1T	0.75	123.6
AMD/FUJITSU	16Mbit	9436	2.7	1T	0.60	87.0

Source: ICE, "Memory 1997" 22479

Figure 10-6. Flash Chip and Cell Size Comparison

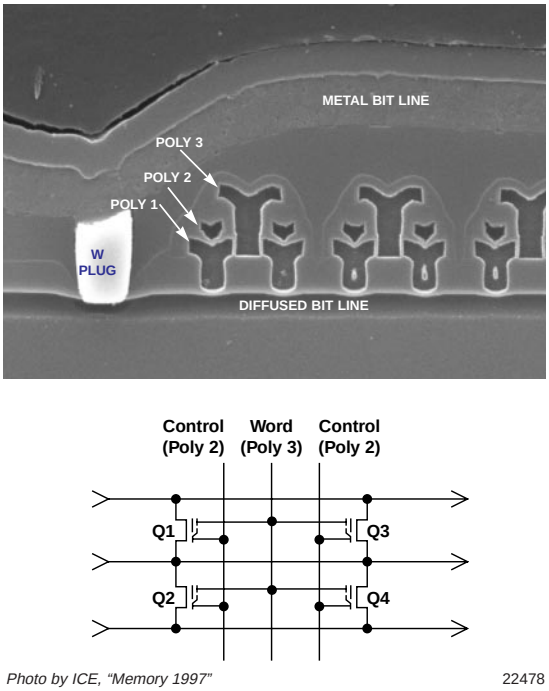


Photo by ICE, "Memory 1997" 22478

Figure 10-7. SanDisk Flash Cell

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 architecture alternatives. Electron trapping in the floating gate is done by hot-electron injection. Electrons are removed by Fowler-Nordheim tunneling. The world's leading manufacturers of flash devices (Intel, AMD) use NOR cell configurations.

NAND Cell

To reduce cell area, the NAND configuration was developed. Figure 10-8 shows the layouts of NOR and NAND configurations for the same feature size. The NAND structure is considerably more compact.

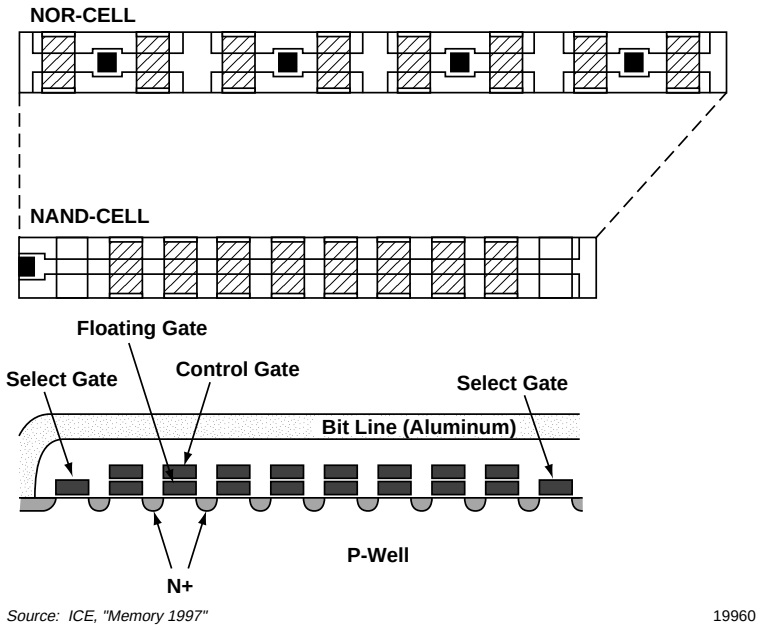


Figure 10-8. Comparison of NOR and NAND Architectures

A drawback to the NAND configuration is that when a cell is read, the sense amplifier sees a weaker signal than that on a NOR configuration since several transistors are in series. Figures 10-9 and 10-10 describe the NAND architecture from Toshiba. 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 10-11 shows a speed comparison of NOR and NAND devices.

DINOR Cell

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

Photo by ICE, "Memory 1997" 22476

Figure 10-9. Toshiba Flash NAND Cell

Architecture	NAND
Date Code	9528
Cell Size	1.3μm^2
Die Size	103mm^2
Min Feature Size (Gate)	Cell: 0.25μm Periphery: 0.5μm

Source: ICE, "Memory 1997" 22475

Figure 10-10. ToshibaÕs 32Mbit Flash Characteristcs

Architecture	NOR	NAND
Random Access Time	80ns	20μs
Serial Access Time	—	80ns

Source: ICE, "Memory 1997" 19961

Figure 10-11. NOR Versus NAND Access Times

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 10-12 shows the DINOR architecture.

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. HitachiÕs device is specified to operate with a 50ns high-speed serial access time.

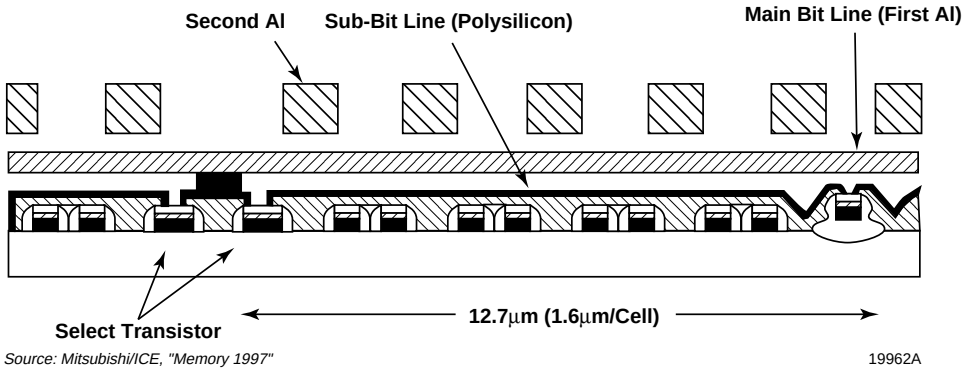


Figure 10-12. DINOR Architecture

Figure 10-13 presents a review of the different flash architectures. Figure 10-14 shows a cell size comparison between DRAM, NAND, and NOR flash architectures. The NOR flash one-transistor cell has roughly the same size as a DRAM cell for the same process generation.

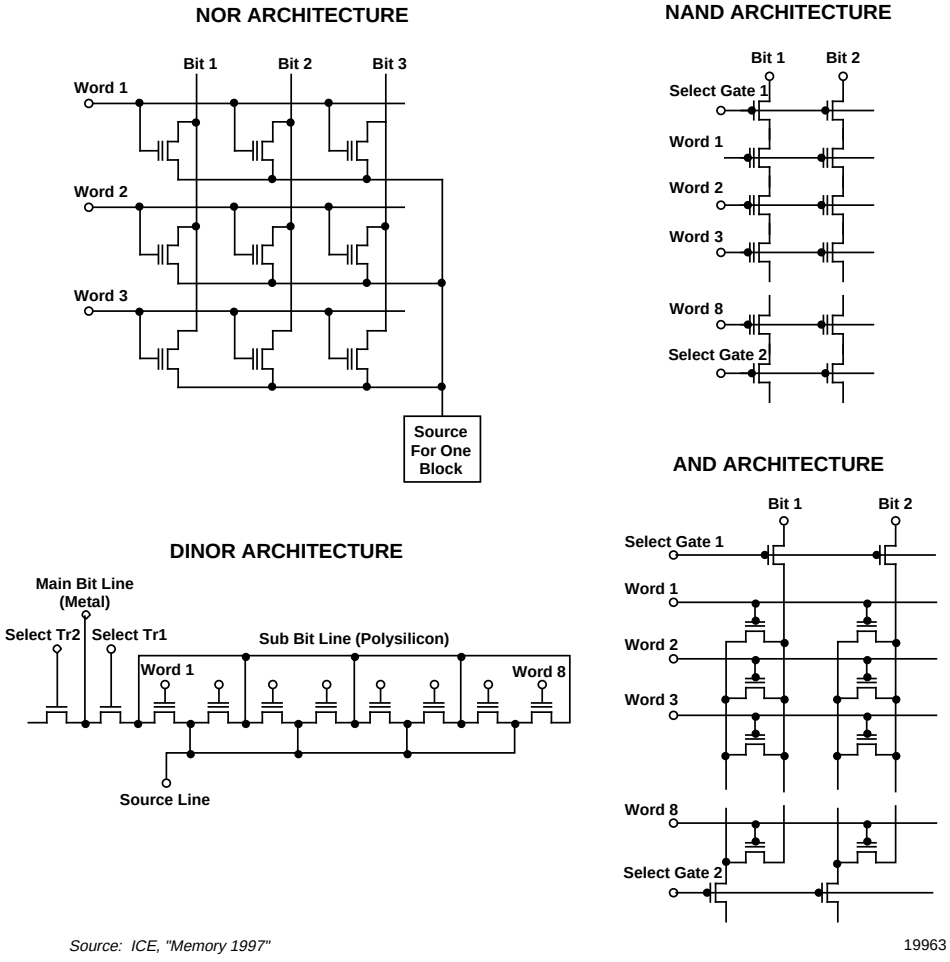


Figure 10-13. Flash Architectures

Technology (Gate Length)	NAND Flash Cell	NOR Flash Cell	DRAM Cell
0.6μm	3μm^2	6μm^2	6μm^2
0.4μm	1.3μm^2	2.5μm^2	2.5μm^2

Source: ICE, "Memory 1997" 22474

Figure 10-14. Flash and DRAM Cell Size Comparison

Several companies strongly support one type of flash architecture. However, to hedge their bets and to offer products for several different end uses, many firms have elected to build flash devices using more than one type of architecture. Figure 10-15 shows vendors' support of flash memory architectures.

NOR	NAND	AND	DINOR
Intel	National	Hitachi	Mitsubishi
AMD	Samsung	Mitsubishi	Hitachi
Atmel	Toshiba		Motorola
Fujitsu	Fujitsu		
TI	AMD		
Micron			
SGS-Thomson			
Macronix			
UMC			
Mitsubishi			
Samsung			
Toshiba			

Winbond uses its proprietary "split-gate" architecture.

Source: ICE, "Memory 1997" 20080C

Figure 10-15. Vendors' Support of Flash Memory Architectures

Audio NAND Flash

Toshiba, Samsung, and National Semiconductor each 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 uses serial access to face speed problems. Moreover, audio NAND devices are cheaper than standard NAND flash since they contain fewer functions. Sometimes audio flash devices may contain some bad cells. Even though those faulty cells would not affect the audio applications, the product would sell for less money.

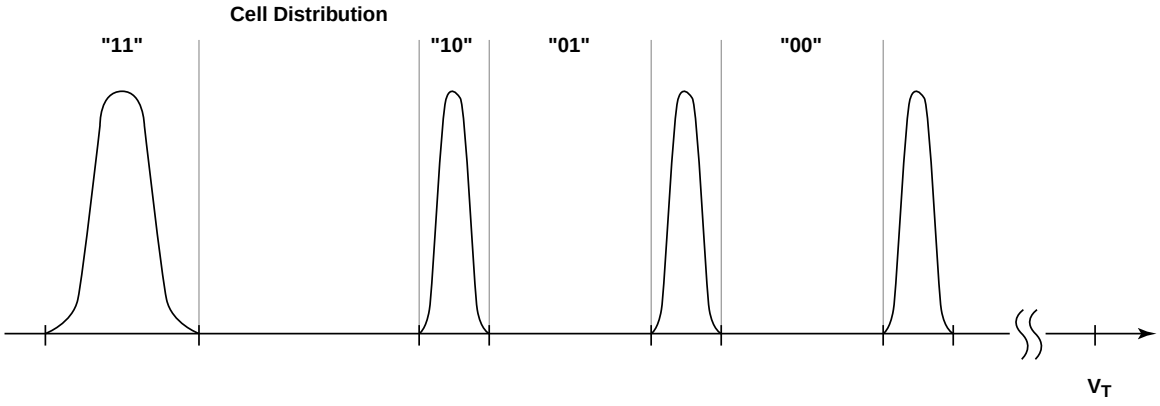
MULTI-LEVEL STORAGE CELL (MLC)

Four-Level Storage Cell

Most of the major flash companies are working to develop their version of a multi-level cell flash device. The goal of this device is to store information in several different levels inside the same memory cell. The most common developments are those that store information on four different levels in the same cell.

In multi-level cell, there are two difficult issues that must be addressed by manufacturers. The first is to tightly control the program cycle that gives four different levels of charge. The second difficulty is to accurately recognize, during the read cycle, the four different threshold voltages of the programmed transistor.

Flash devices must be reliable even in worst case conditions. External parameters (power supplies, temperatures, etc.) may vary from the time the flash device is programmed to the time it is read. Figure 10-16 shows an example of threshold voltage distribution for four stages stored on the same transistor.



Source: ICE, "Memory 1997" 20805

Figure 10-16. Threshold Voltage Distribution for Four States

Different companies are working intensively on this issue. During each of the past several years, papers were presented by most of the major flash manufacturers regarding multi-level cell technology. Intel presented a paper on its four-level storage work at the 1995 ISSCC conference. At the 1996 ISSCC conference, two papers were presented on this concept. Samsung presented a 128Mbit four-level NAND flash cell and NEC presented a 64Mbit four-level NOR flash cell. At the 1995 Symposium on VLSI Circuits, Toshiba presented a development for future high density

MLC NAND flash memories. At the December 1996 IEDM Conference, SGS-Thomson presented a study on MLC for the different flash architectures and their trade-offs. Highlights of this study are presented in Figure 10-17.

Array Architecture	Cell Size*	Advantage as Single-Bit Concept	Disadvantage as Single-Bit Concept	Advantage as Multi-Bit Concept	Disadvantage as Multi-Bit Concept
Common Ground	9-11F^2	1. General purpose applications and most understood array and technology	1. Relatively large cell size	1. Minimum interaction between neighbors 2. CHEI for programming	1. Closely coupled metal bitline 2. V_t distribution affected by neighbor data
DINOR	7.5F^2	1. Reduced cell size while preserving the common ground array	1. Requires triple poly	1. Reduction in BL-BL coupling	1. Tunneling during programming 2. Source resistance
AND	8F^2	1. Good combination of CG and DINOR 2. Drain contact every 32-128 cells	—	1. Reduction in BL-BL coupling	1. Tunneling during programming 2. Source resistance
NOR Virtual Ground - AMG	6F^2	1. Small cell size 2. Low current programming	—	1. CHE programming 2. Reduction in BL-BL coupling	1. Resistive diffusion bitlines 2. Neighbor interaction affecting V_t distribution
NOR Virtual Ground - Split Gate Poly-Poly Erase	7.5F^2	1. Overerase not an issue	1. Requires triple poly	1. CHEI programming 2. Disturb reduction due to poly-poly erase	1. Neighbor interaction affecting V_t distribution 2. Low read current and high erase voltages
NAND	6F^2	1. Small cell size	1. Read thru stack of 15 cells 2. High read and programming voltages	—	1. Programming by tunneling in the channel

***F is the technology feature size**

Source: SGS-Thomson/ICE, "Memory 1997" 22595

Figure 10-17. Trade-Off of MLC Using Different Flash Architectures

During the first half of 1997, Intel announced that it sampled 64Mbit MLC parts. SanDisk, along with manufacturing partner Matsushita, used the technology to boost single-chip capacity to 64Mbit. It refers to its multi-level cell technology as "Double Density" or "D^2". SanDisk claims that the 64Mbit die is only 10 percent larger than the company's 32Mbit die. Meanwhile, the company is also working on a 256Mbit Double Density flash device.

Multi-Level Storage Cell for Audio Applications

Development of MLC cell takes 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) proposed non-volatile memories that are able to store 256 different levels on the same transistor. ISD's product family is called ChipCorder and enables a single chip solution for voice recording and playback. It currently has a chip with up to four minutes of voice storage capacity.

POWER SUPPLY

Currently, flash power supplies range from 5V/12V down to 2V. Flash memory power supplies vary widely from vendor to vendor. There are two main reasons for this variation. First, flash cells need high voltage for programming. With different types of flash architectures and designs, different program/erase techniques (Fowler-Nordheim tunneling or hot-electron injection) exist. These architectures do not share the same voltage requirements. For example, high voltage with no current can be generated internally with a voltage pump. However the source/drain current of hot-electron injection requires an external power supply.

The second reason for wide power supply variation is that there are many applications that currently require different power supply levels. Some applications may require low-voltage flash devices while others operate well using flash device with high-voltage characteristics. Manufacturers can propose different types of power supplies that best fit a specific application.

SmartVoltage

SmartVoltage is an Intel concept. However, other manufacturers including Sharp and Micron have signed on to license the technology. SmartVoltage parts can be used for several power supplies. Read voltage may be 2.7V, 3.3V or 5.5V and programming voltage may be 3.3V, 5V or 12V.

Flash memories are used in a wide variety of applications as illustrated Figure 10-18. All these applications allow vendors to offer several flash solutions. Using the NAND flash architecture for serial access applications is one example. Figure 10-19 shows the diversity of the flash memory types.

Focus Segment	Application
Auto	Engine Control
PC	BIOS
HDD	Disc Control
Wireless	Analog/SSM
Networking	Hub Control

Source: TI/ICE, "Memory 1997" 22596

Figure 10-18. Flash Target Segments

Core Architecture	**NOR, NAND, DiNOR, AND**
Cell Architecture	**1 Transistor, Split Gate, Others**
Storage	**1 Level, Multi-Level Cell (MLC)**
Voltage (Read/Program)	**5V/12V, 5V/5V, 3V/5V, 3V/3V, 2.7V/2.7V, 2.2V/2.2V, Smart Voltage**
Configurations	**Random Access, Serial Access, Others**
Applications	**Audio, PC, Wireless**

Source: ICE, "Memory 1997" 22473

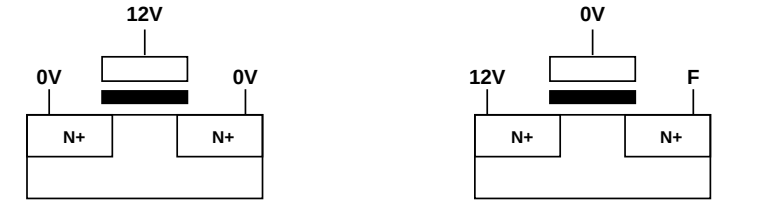
Figure 10-19. Flash Diversity

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 erase / program, flash ICs that use hot electron injection for trapping electrons in the floating gate are programmed (data equal to 0) by capturing electrons in the floating gate, as with 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 with an EEPROM. The reliability concern is to either over program or over erase as shown in Figure 10-20.



V_T SHIFT HIGH

- **FN tunneling for program**
- **Low program/erase current**
- **Over program problem**
- **Slow program time**

V_T SHIFT LOW

- **Hot electron injection to program**
- **Lower program disturb**
- **Over erase problem**
- **Fast program time**

Source: Motorola/ICE, "Memory 1997" 20841

Figure 10-20. Erased Threshold Voltage Shift for Flash Memory Cell

PCMCIA

Magnetic memory storage and flash memory devices 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.

Hitachi proposed a 75Mbyte ATA PC Card using a mostly good memory (MGM) production technique. The chip must have a minimum of 98 percent of its memory cell sectors free of defect and have all logic circuits 100 percent functional. Figure 10-21 illustrates an ATA Card using the MGM technology.

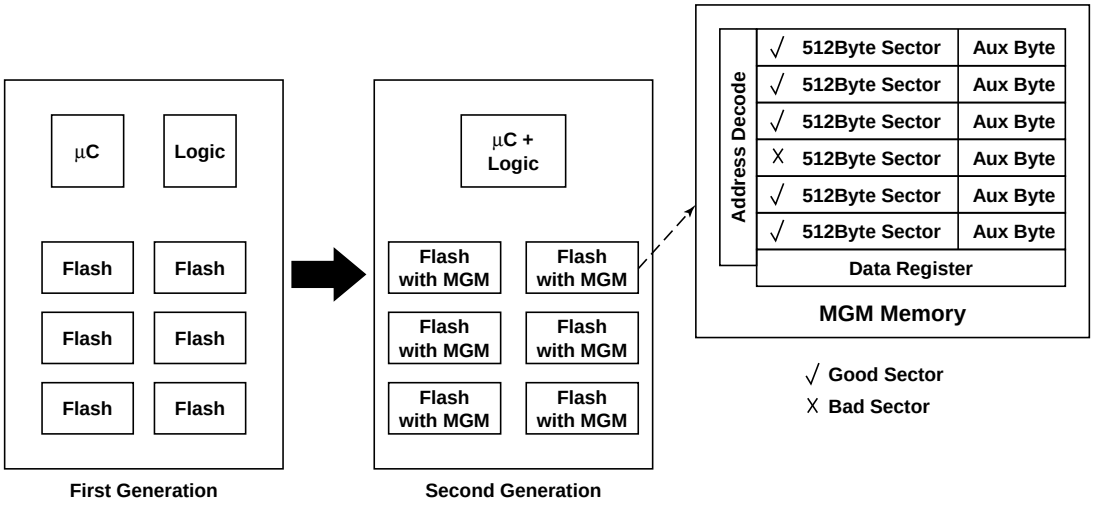


Figure 10-21. ATA Card Evolution

SMALL FLASH-MEMORY MODULES

Small flash-memory modules 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.

Three developments—CompactFlash, Miniature Card and Solid State Floppy Disk Card (SSFDC)—are similar in size but employ substantially different electrical interface schemes. Figure 10-22 presents the three miniature flash card solutions.

	CompactFlash	Miniature Card	SSFDC
Original Developers	SanDisk	Intel/AMD	Toshiba
Industry Alliance	CampactFlash Association	Miniature Card Implementers Forum	SSFDC Forum
Module Dimensions	43 x 36 x 3.3mm	38 x 33 x 3.5mm	45 x 37 x 0.76mm
Memory Type	NOR Flash	NOR Flash, DRAM, SRAM, OTP, ROM	NAND Flash
Capacity	2 to 15Mbytes	2, 4Mbytes	2, 4Mbytes
Connector Type	50-Pin subset of PCMCIA	40-Pad Elastomeric	68-Pin PCMCIA With Adapter
Number of Contacts	Circular Pins	Flat-Edge Contacts	Flat-Surface Contacts
Software Interface	ATA	FTL (Flash Translation Layer)	Host-Based Controller
Built-In Controller?	Yes	No	No

Source: ICE, "Memory 1997" 22598

Figure 10-22. Standards for Small Flash-Memory Modules

CompactFlash

CompactFlash was developed by SanDisk Corporation, Sunnyvale, California, in 1994. The CompactFlash Association (CFA) was established in October, 1995, to promote and encourage the worldwide adoption of CompactFash technology as an open industry standard. More than 40 companies have joined the CFA.

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.

Miniature Card

The Miniature Card, originally developed by Intel, is supported by the Miniature Card Implementers Forum (MCIF). The Miniature Card 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. Miniature Cards are cheaper than CompactFlash cards but need that additional software. Figure 10-23 shows the ATA configuration versus the linear configuration. Intel developed its Miniature Card for high-volume consumer applications and will not support CompactFlash.

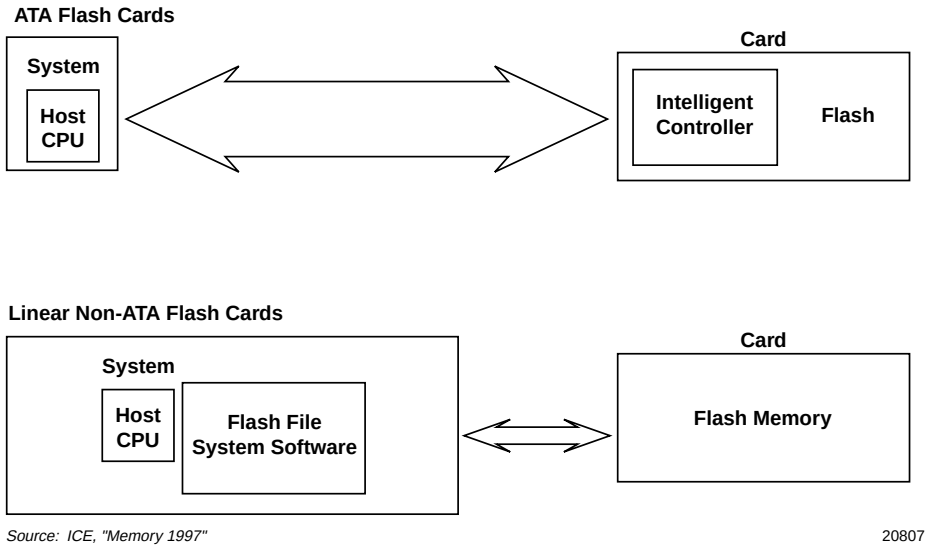


Figure 10-23. ATA Versus Linear Flash Card

Solid State Floppy Disk Card (SSFDC)

Toshiba's Solid State Floppy Disk Card is based on its flash NAND cell technology. This card was announced in late 1995. 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. An SSFDC Forum was held in April 1996 in Japan to agree on an industry standard for a super-small data storage medium. More than 40 companies, including Samsung Electronics, have joined the SSFDC Forum.

The SSFDC is the size of a credit card, and is much thinner than any of the other small-form factor memory cards. Used with an ATA PC card adapter, SSFDC can be used as a standard PC card.