FastFLASH: A New Electrically Erasable CPLD Technology

XBRF010 January, 1997 (Version 1.0) Application Brief

Summary

The Xilinx FastFLASH technology, used in the XC9500 family, provides key advantages in reliability, density, and performance. This overview describes the FastFLASH process technology and compares it with EEPROM technology.

Xilinx Family

XC9500

Introduction

The Xilinx FastFLASH technology is a double-polysilicon, 2-layer metal CMOS flash technology for CPLDs. This technology allows 5 V in-system programming. FastFLASH technology is capable of producing the high cell density needed for pin-locking, with a high endurance level of 10,000 program/erase cycles. Compared to typical CPLD EEPROM technology, FastFLASH technology provides advantages in reliability, density, and performance. The compatibility of the technology with industry-leading flash processes ensures the scalability of the basic process and the availability of foundry capacity.

Background

CPLDs are programmable by designers to implement desired logic functions. As non-volatile devices, CPLDs retain the programmed information even after removing power. The underlying non-volatile process technology supports this ability.

The process technology for CPLDs traditionally follows non-volatile memory (NVM) technologies. EPROM memory technology offers excellent memory cell density at a low process cost, although it is not electrically erasable. EEPROM technology offers electrical erasability at a reasonable process cost. However, it requires a relatively large memory cell size and offers limited endurance. Flash technology is an electrically erasable extension of EPROM technology. FastFLASH technology offers the best long term technology solution by providing the cell density of EPROM, the electrical erasability of EEPROM, excellent endurance characteristics, and long term process cost benefits.

Compared to EEPROM technology, the proprietary FastFLASH technology supports the needs of CPLDs by providing:

- high-performance logic capability
- high memory cell density
- electrical erasability
- 5 V program and erase
- high reliability and endurance
- process scalability
- fast device programming times

Figure 1 shows the relative cell sizes for FastFLASH and conventional EEPROM technologies.

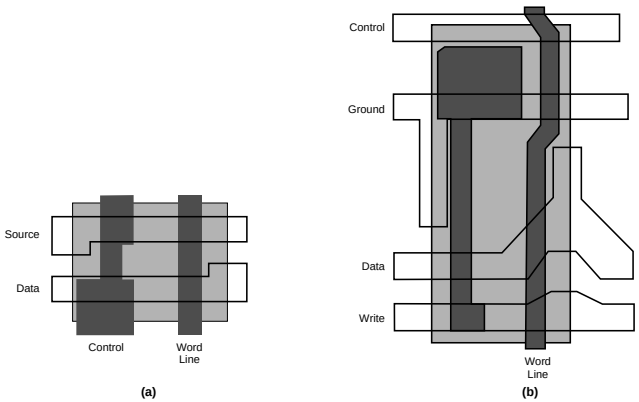

Figure 1: Layout Comparison of a FastFLASH Cell (a) and an EEPROM Cell (b)

EEPROM Technology

EEPROM technology was the first electrically erasable technology used for CPLDs. The programmable element is a special thin oxide capacitor that conducts a small current when a sufficient voltage is applied across the oxide. The tunnel oxide, approximately 80 Angstroms thick, is used to inject or extract charge from a floating gate via Fowler-Nordheim (FN) tunneling. The floating gate is connected to the gate of a sense transistor in order to sense the programming state. In addition to the tunnel oxide capacitor and sense transistor, two more transistors and an additional control capacitor are required to create a single EEPROM cell that can be programmed and erased in a CPLD application.

Figure 2 shows the schematic and cross-section of an EEPROM cell used in CPLDs. The tunnel oxide capacitor transports charge to and from the floating gate, which controls the sense transistor. Two additional transistors are used for the program and read operations. A control gate capacitor transfers voltage to the floating node for program and erase operations. Compared to standard CMOS logic processes, three additional device structures are created for the EEPROM cell: the tunnel oxide capacitor, the control gate capacitor and the high-voltage transistor. The resulting process complexity makes the scalability of the process and the EEPROM cell more difficult in future technology generations.

The EEPROM cell density is inferior because each cell is a circuit consisting of five separate device structures. The EEPROM cell area in a typical 0.6 micron technology is 75 to 100 square microns. The inadequate cell density directly affects the pin-locking ability of the architecture because more routing switches significantly increases die cost. The large cell size also introduces parasitic capacitances with the potential to limit the overall performance.

In general, EEPROM technology is developed and enhanced by individual companies without the benefit of an industry-wide development interest. This further increases the difficulty of long-term process migration.

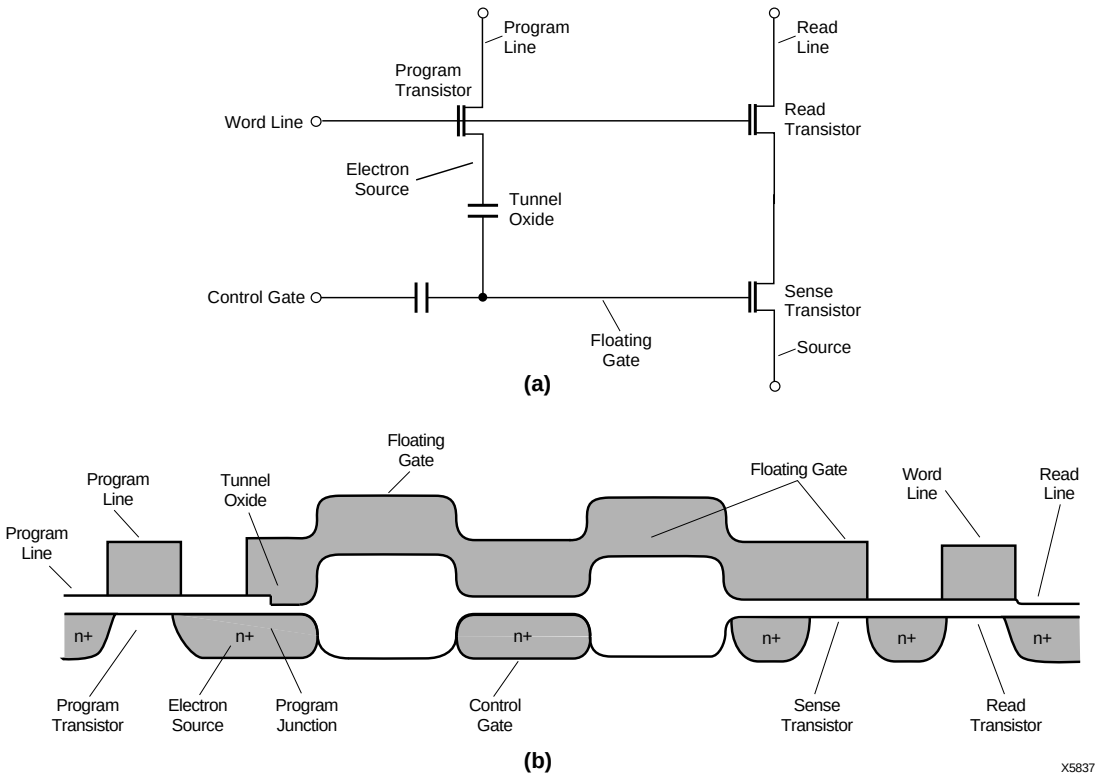

Figure 2: Schematic and Cross Section of an EEPROM Cell

FastFLASH Technology

FastFLASH technology is compatible with the industry-leading flash processes. The fundamental programmable element is the flash transistor. It incorporates the floating gate into the device structure for improved cell area as shown in Figure 3. By the addition of an NMOS transistor in series, as shown in Figure 4, the flash transistor is incorporated into the FastFLASH cell.

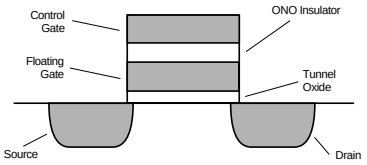

Figure 3: FastFLASH Transistor

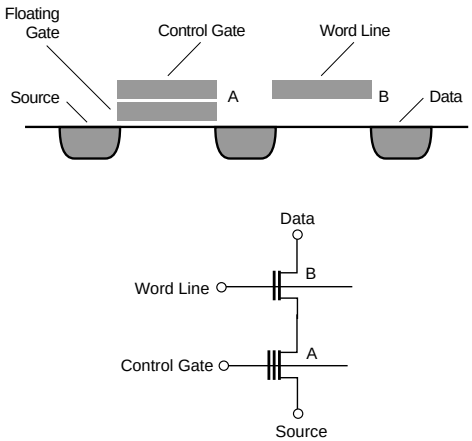

Figure 4: FastFLASH Cell with Series Transistor

The behavior of an individual flash transistor is changed with a program or an erase operation. When a flash transistor is in the erased state, the threshold voltage (V_{ta}) is approximately 1 V. During programming, the threshold voltage (V_{tb}) increases sufficiently above 5.5 V so the transistor does not turn on for a logic operation, as shown in Figure 5.

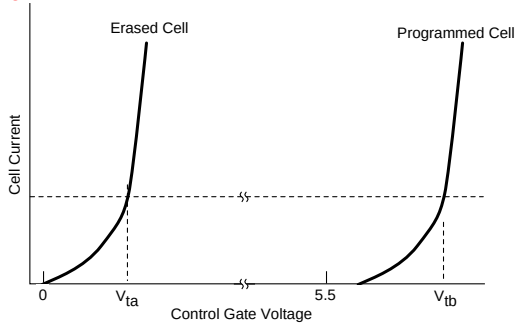

Figure 5: Cell Characteristics

The physical structure of the flash transistor includes a floating gate polysilicon layer that is isolated from the silicon substrate by a thin oxide approximately 100 Angstroms thick. Above the floating gate is the control gate polysilicon layer, with an insulating oxide-nitride-oxide layer between them. The control gate is driven by internal logic circuits, while the floating gate is unconnected. When the flash transistor is in the erased state, there is no net charge on the floating gate. By modifying the net electrical charge on the floating gate, the threshold voltage may be increased to 6 V or more.

The flash transistor is programmed by applying approximately 12 volts to the control gate, 5.5 volts to the drain, and 0 volts to the source as shown in Figure 6. The voltages are supplied by internal voltage pumps or externally by a device programmer. During the programming operation, channel hot electrons (CHE) are created near the pinch-off region. Some CHEs have sufficient thermal energy to pass through the thin oxide and remain on the floating gate. The collected electrons create a net negative voltage on the floating gate that opposes the electric field emanating from the control gate. The result is a net increase in the threshold voltage.

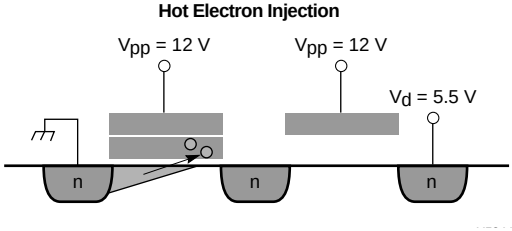

Figure 6: Programming a FastFLASH Cell

The flash transistor is erased by applying 0 volts to the control gate and approximately 10 volts to the source with the drain left floating. In Figure 7, the electric field between the floating gate and the source node is increased to the point where Fowler-Nordheim tunneling takes place. Excess electrons are transported from the floating gate to the source. The transistor is designed to make the erase process self-limiting. The electric field decreases as electrons are removed from the floating gate. FN tunneling effectively stops when the floating gate is electrically neutral.

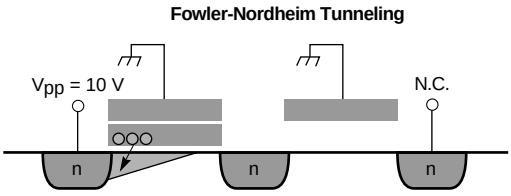

Figure 7: Erasing a FastFLASH Cell

Reliability and Endurance

The reliability of the programmable cell relates to data retention and endurance. Data retention is how long the cell can keep its programmed and erased states. Endurance is the number of times a cell can be programmed and erased without error.

When a cell is programmed and electrons are injected into the floating cell, the net charge is expected to remain indefinitely. In practice, charge leaks off -- typically tens of years under normal storage and operating conditions. The charge transport can occur due to direct tunneling and thermal leakage. Direct tunneling is exponentially dependent on electric field. Generally, the direct tunneling leakage current is very small and largely negligible. Thermal leakage is the dominant leakage component. The leakage characteristics are thoroughly characterized and modeled.

Endurance is determined by the magnitude of the applied electric field and the quality of the thin oxide used for the program and erase operations. When program and erase operations are performed in FastFLASH cells, electrons are transported across the oxide in the direction of the applied electric field.

Endurance failures occur in the alteration of the tunnel oxide characteristics where charge is repeatedly transported across the oxide. In FastFLASH technology, the maximum electric field is approximately 10 MV/cm. Consequently, the major cause of endurance failure is electron trapping in the tunnel oxide. After a large number of reprogramming cycles (typically over 100,000 and up to 1 million cycles), traps can be created within the oxide. When a sufficient number of electrons become trapped in the oxide, the localized electric field is distorted. The result is reduced program and erase efficiency, and reduced voltage margins between programmed and erased states. The FastFLASH technology endurance failure is a gradual degradation rather than an abrupt failure.

In EEPROM cells, the program and erase fields are 15 to 20 MV/cm. The substantially higher field strength makes the EEPROM cell susceptible to destructive tunnel oxide breakdown in addition to electron trapping in the oxide. Therefore, the oxide is physically ruptured and the cell is permanently damaged. The EEPROM endurance failure mechanism is a hard failure that is probabilistic in nature and difficult to screen.

In both flash and EEPROM technologies, oxide quality control becomes important. In typical CPLD EEPROM technologies, the tunnel oxide is approximately 80 Angstroms, among the thinnest tunneling oxides in use today. In contrast, FastFLASH technology uses industry-standard process steps to achieve a high-quality tunnel oxide of approximately 100 Angstroms.

There is a difference between the endurance of a single cell, and the overall endurance of the CPLD product using 100,000 cells or more. For the CPLD, failure statistics of a large number of cells suggest that the product endurance is typically 1.5 orders of magnitude lower than the endurance of an individual cell.

By using a significantly thicker tunnel oxide and lower electric fields than EEPROM technologies, the FastFLASH technology offers an endurance level of 10,000 cycles today, with plans for 1,000,000 cycles in the future. Comparatively, EEPROM CPLD technologies offer 10,000 cycles for the older, 0.8 micron, geometries and only 100 cycles for leading edge 0.6 micron (and smaller) technologies.

Cell Density

The FastFLASH cell, with its flash transistor and series read transistor, is considerably smaller than the comparable EEPROM cell with five device structures. Figure 1 shows the relative layouts of the FastFLASH cell and an EEPROM cell. The FastFLASH cell is approximately 25 square microns in 0.6 micron technology. In contrast, the EEPROM cell is approximately 75 to 100 square microns, or three to four times larger. Therefore, FastFLASH offers three or more times cell density than EEPROM technology.

Cell density is important in CPLD technology, because a programmable cell controls each routing switch. The routability and pin-locking capability are directly impacted by EEPROM-based CPLD architectures using fewer programmable cells for routing. These architectures typically devote 10 to 20% of the chip area to programmable cells. Therefore, increasing cell count by a factor of three would substantially increase the production costs in these devices.

Program and Erase Times

The CHE mechanism used in programming the FastFLASH cell is very fast, typically on the order of 20 microseconds for 8 bits. This is an important advantage for production device programming where devices can be programmed from the erased state in as few as two seconds using automatic test equipment. In contrast, FN tunneling used in EEPROM cells is several orders of magnitude slower, at approximately 40 milliseconds for several hundred cells. The net result is slower device programming.

For both technologies, FN tunneling is used in device erasure. In either case, the device is erased one sector at a time, with a typical erase time of several hundred milliseconds. In a production environment the devices are received in an erased state and therefore the erase time is not a cost factor in production programming.

Process Scalability

Process scalability is defined as the ease of shrinking the process and chips into future generations of technology with smaller chip sizes, lower chip costs, and faster chips. It is important to ensure that a migration path for cost and speed improvements, as well as continued availability of the current process to meet production requirements.

Factors that contribute to process scalability include:

- Number of different device structures
- Scalability of the programmable cell
- Compatibility with main-stream memory processes

The leading flash memory technologies are double-polysilicon, stacked gate technologies with CHE injection for programming and FN tunneling for erase. These technologies offer superior process scaling and continued development improvements in both cost and speed. Since the FastFLASH technology is compatible with these process technologies, it will continue to enjoy continued process migration into tighter geometries.

In contrast, the EEPROM technology used in CPLDs is not used in high-volume memory technologies. Combined with a complex cell structure, EEPROM technologies are substantially more difficult to migrate to tighter geometries.

Conclusion

FastFLASH technology offers significant advantages in reliability, density, and performance over comparable EEPROM technologies. Utilizing a flash transistor structure that is compatible with the leading non-volatile memory technology ensures process scalability and future foundry availability.

References

1. C.Hu, “Nonvolatile Semiconductor Memories: Technology, Design, and applications” IEEE press, New York, 1991.
2. C.Hu, “Thin oxide reliability”, IEDM Tech. Digest pp 368-371, 1985.
3. S.K.Lai, V.K.Dham and D.Guterman, “Comparison and trends in today’s dominant E^2 technologies”, IEDM Tec. Dig., pp 580-583, 1986.
4. B.Dipert and L.Hebert, “Flash memory goes mainstream”, IEEE spectrum, pp 48-52, October 1993.