Intel StrataFlash™ Memory Technology Overview

Greg Atwood, Flash Technology and Manufacturing, Santa Clara, CA, Intel Corp.
Al Fazio, Flash Technology and Manufacturing, Santa Clara, CA, Intel Corp.
Duane Mills, Memory Components Division, Folsom, CA, Intel Corp.
Bill Reaves, Memory Components Division, Folsom, CA, Intel Corp.

Index words: StrataFlash, MLC, flash, memory

Abstract

The Intel StrataFlash™ memory technology represents a cost breakthrough for flash memory devices by enabling the storage of two bits of data in a single flash memory transistor. This paper will discuss the evolution of the two bit/cell technology from conception to production.

The flash memory business has grown from about $50M in 1987 to roughly $2.5B in 1997 due to its unique mix of functionality and cost. Flash memory devices are now found in virtually every PC and cellular phone and are one of the key components of the emerging digital imaging and audio markets.

Cost per bit reduction of flash memory devices has been traditionally achieved by aggressive scaling of the memory cell transistor using silicon process-scaling techniques such as photolithography line width reduction. In an attempt to accelerate the rate of cost reduction beyond that achieved by process scaling, a research program was started in 1992 to develop methods for the reliable storage of multiple bits of data in a single flash memory cell. The Intel StrataFlash two bit/cell memory technology is the first output of the multi-bit per cell storage effort. By storing two bits in a single memory transistor, the memory cell area is effectively cut in half allowing the storage of twice as much data in the same area as the standard single bit per cell technology.

This paper provides insight into the Intel StrataFlash memory technology development effort. It discusses the evolution of the two bit/cell capability from conception to production and the challenges that were successfully overcome to produce a high-quality product compatible with the standard single bit per cell devices. This paper also presents examples that showcase the benefits of the current Intel StrataFlash memory devices and discusses some of the driving forces for high density flash memory.

Introduction

History has shown that as the price of memory drops and the density increases, the application usage and demand for that memory will increase. The cost for semiconductor memories (i.e., DRAM, SRAM, ROM, and flash) is largely determined by the amount of silicon area it takes to store a data bit of information. As with other semiconductor memories, flash memory, which retains its data even when the power is removed, achieves higher density and lower cost through traditional silicon process scaling techniques, such as feature size reduction. To build on process scaling, a concept called Multi-Level-Cell (M.L.C.) technology was introduced. This technology lowers the cost by enabling the storage of multiple bits of data per memory cell thereby reducing the consumption of silicon area. The two bit/cell Intel StrataFlash memory technology provides a cost structure equivalent to the next generation of process technology while using the current generation of process technology equipment. Figure 1 illustrates the substantial acceleration of the rate of cost reduction possible with M.L.C.

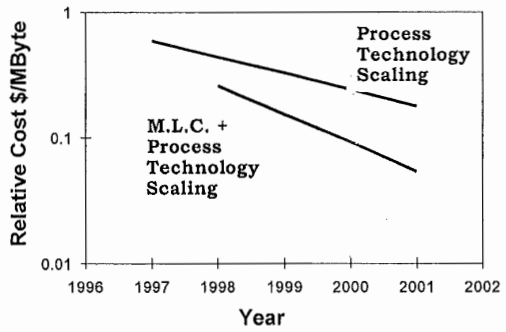

Figure 1: Accelerated cost reduction using M.L.C.

Introduction to Flash Memory

A discussion of the Intel StrataFlash memory and technology first requires a brief overview of the standard ETOX™ flash memory technology and its use. Flash memory is a member of the non-volatile class of memory devices, storage devices that maintain their data in the absence of applied power. The ETOX technology is the predominate flash technology, representing over 70% of flash memory shipments. Data is entered into the flash memory on a bit, byte, word, or page boundary through an operation called programming. Once data is entered into the device it will remain, regardless of the presence or absence of power. Data is cleared from the flash memory with an erase operation. The contents of the flash memory are erased on a block boundary, where a block size can be anywhere from 8Kbits to 1Mbit depending on the product design.

The ETOX flash memory storage element, or memory cell, shown in Figure 2, is a single transistor with the addition of an electrically isolated polysilicon floating gate capable of storing charge (electrons). The amount of stored charge modifies the behavior of the memory cell transistor. This change in transistor behavior is translated into stored data: the presence of charge is interpreted as data "0;" the absence of charge is interpreted as data "1." The single transistor memory cell results in a small cell size, and thus a small amount of silicon area is consumed for the storage of one bit of data, resulting in low cost.

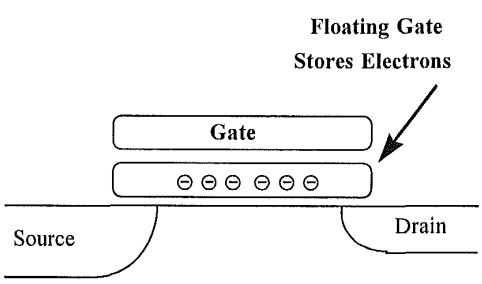

Figure 2: Single transistor flash memory cell

Flash Memory Market

The combination of non-volatility, electrical alterability, and low cost is attractive to small systems that do not have access to a continuous power source such as battery-powered devices. For example, almost every cellular phone sold today contains a flash memory device. This device stores the program that the cellular phone uses to communicate over the wireless network and interact with the phone user. In some cases, the memory device also stores incoming messages, much like a cellular digital answering machine. The non-volatility of flash ensures that, when you remove the battery from the phone, it will not forget how to communicate nor will it loose any of your messages.

The unique combination of features provided by flash memory has enabled the market for these devices to grow from less than $50M in 1987 to over $2.5B in 1997. Flash memory devices are found in over 90% of PC's, over 90% of cellular phones, and over 50% of modems. Applications are as diverse as airline flight recorders, medical recording equipment, digital answering machines, arcade games, printers, and network routers. Flash memory is a key component of the emerging digital imaging and audio markets where it serves as the digital "film" or digital "tape."

The Pursuit of Lower Cost Memory

The growth of the flash memory market has been driven by a continual increase in density and reduction in cost, enabling new applications to emerge and further fuel the demand for more flash. Figure 3 illustrates the rapid increase in the flash market size driven by the reduction in memory price. As the price of memory was reduced, new applications for flash memory emerged (some examples are shown in Figure 3) fueling further market growth.

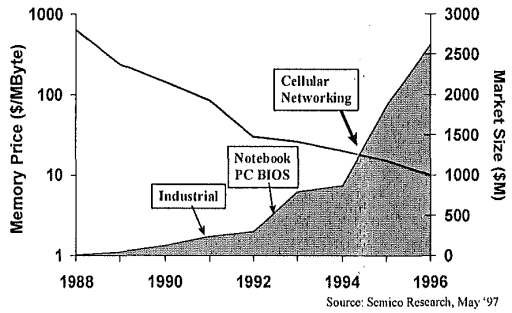

Figure 3: Flash memory price and market size

Traditionally, cost reduction and density increase for flash memory has been driven by process scaling in the same way as other semiconductor memory devices such as DRAMs and SRAMs. As the ability of the semiconductor manufacturing process equipment improves, smaller features can be resolved on the silicon wafer resulting in a smaller memory cell and thus more bits in a given amount of silicon area. More bits in a given silicon area result in higher density memories and lower cost per bit. Using the technique of process

technology scaling, the flash memory cell size has been reduced by 18 times in the past 10 years as shown in Figure 4. The reduced cell area combined with increases in the size of the memory product (brought about by improved manufacturing techniques and yields) has resulted in a product density increase of over 100 times in the same 10-year period.

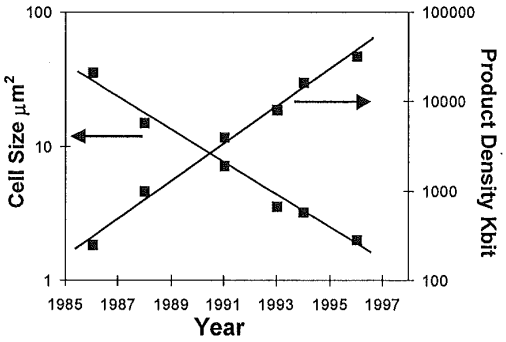

Figure 4: Cell area and product density in time

The flash memory cell is a single transistor; one bit of data is stored in one transistor. By comparison, a SRAM memory cell requires six transistors (or four transistors and two resistors), a DRAM memory cell requires one transistor and one capacitor, and an E^2PROM cell requires two transistors.

A single transistor has been generally considered the smallest practical unit for the storage of a bit of data. In 1992, the Intel flash development team began a research effort to reduce the amount of silicon required to store a bit of data to a fraction of a transistor through the storage of more than one bit in a single memory cell transistor. The Intel StrataFlash two bit/cell memory technology recently introduced is the first achievement of this multi-bit per cell storage effort. It provides the cost structure of the next-generation process technology while using the current generation process technology equipment (see Figure 5).

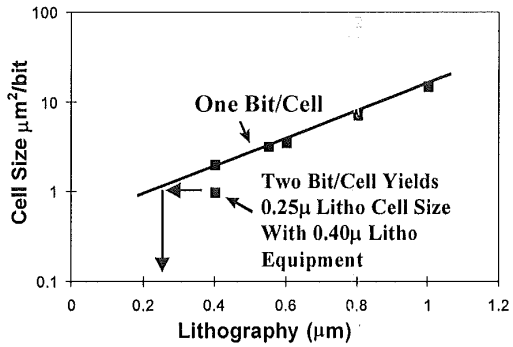

Figure 5: Cell area as a function of lithography

The Multi-Bit Storage Breakthrough: Intel StrataFlash Technology

As discussed earlier, the flash memory device is a single transistor that includes an isolated floating gate. The floating gate is capable of storing electrons. The behavior of the transistor is altered depending on the amount of charge stored on the floating gate. Charge is placed on the floating gate through a technique called programming. The programming operation generates hot electrons in the channel region of the memory cell transistor. A fraction of these hot electrons gain enough energy to surmount the 3.2eV barrier of the Si-SiO_2 interface and become trapped on the floating gate. For single bit per cell devices, the transistor either has little charge (<5,000 electrons) on the floating gate and thus stores a "1" or it has a lot of charge (>30,000 electrons) on the floating gate and thus stores a "0." When the memory cell is read, the presence or absence of charge is determined by sensing the change in the behavior of the memory transistor due to the stored charge. The stored charge is manifested as a change in the threshold voltage of the memory cell transistor. Figure 6 illustrates the threshold voltage distributions for a half million cell (1/2Mc) array block. After erasure or programming, the threshold voltage of every memory cell transistor in the 1/2Mc block is measured, and a histogram of the results is presented. Erased cells (data 1) have threshold voltages less than 3.1v, while programmed cells (data 0) have threshold voltages greater than 5v.

PGR2024-00048/PHI EX. 1038 Page 3 of 8

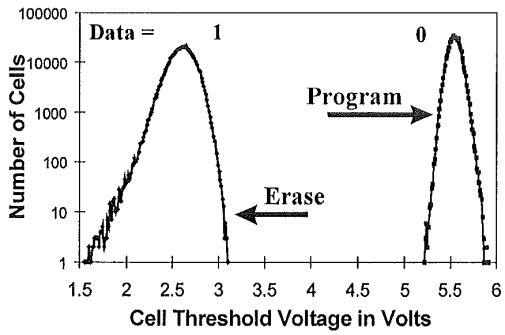

Figure 6: Single bit/cell array threshold voltage histogram

The charge storage ability of the flash memory cell is a key to the storage of multiple bits in a single cell. The flash cell is an analog storage device not a digital storage device. It stores charge (quantized at a single electron) not bits. By using a controlled programming technique, it is possible to place a precise amount of charge on the floating gate. If charge can be accurately placed to one of four charge states (or ranges), then the cell can be said to store two bits. Each of the four charge states is associated with a two-bit data pattern. Figure 7 illustrates the threshold voltage distributions for a 1/2Mc block for two bit per cell storage. After erasure or precise programming to one of three program states, the threshold of each of the 1/2Mc is measured and plotted as a histogram. Notice the precise control of the center two states, each of which is approximately 0.3v (or 3,000) electrons in width.

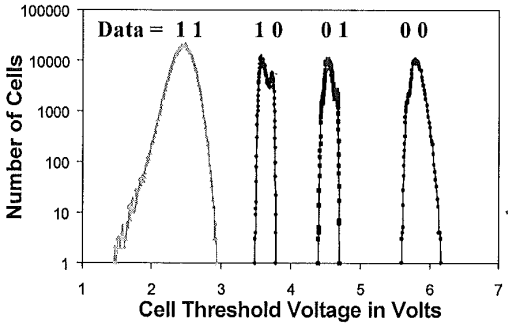

Figure 7: Two bit/cell array threshold voltage histogram

Higher bit per cell densities are possible by even more precise charge placement control. Three bits per cell requires eight distinct charge states; four bits per cell requires sixteen distinct charge states. In general, the number of states required is equal to 2^N where N is the desired number of bits.

The ability to precisely place charge on the floating gate and at some later time sense the amount of charge that was stored has required substantial innovations and extensive characterization and understanding of cell device physics, memory design, and memory test. These innovations are discussed in detail in the paper entitled "Intel StrataFlash Memory Technology Development and Implementation" also published in this issue of the Intel Technology Journal.

Evolution of the Intel StrataFlash Memory Technology Development

This section will outline the development of the Intel StrataFlash memory technology from conception in 1992 to productization in 1997, highlighting the key innovations along the way. The 64Mbit product recently introduced differs markedly from the 1992 view of what a two bit/cell product might look like. The learning that has occurred over the past four years has enabled the development of a two bit/cell memory device that functionally looks almost identical to a one bit/cell device, far exceeding the capability that was considered possible when the development program started. Figure 8 shows the timeline of the major Intel StrataFlash memory technology development milestones.

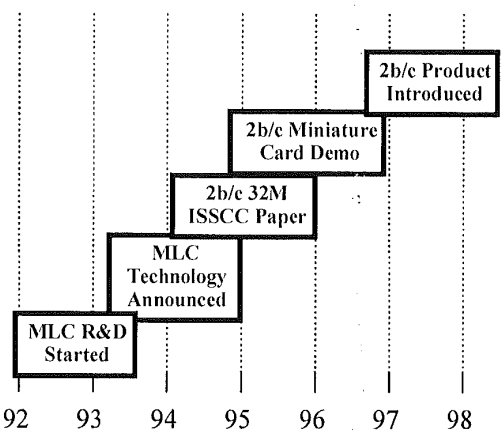

Figure 8: Intel StrataFlash development program

The Multi-Level-Cell (M.L.C.) Concept

Storage of analog data in a floating gate memory device is not a new concept. It was suggested as early as 1971 for EPROM devices[1] and was implemented on E^2PROM devices for use in neural networks, voice recorders, and toys as early as 1982. These analog storage applications can tolerate a high error rate and thus do not place stringent requirements on the memory reliability or accuracy. Neural networks are, by their nature, fault tolerant. Voice storage and simple talking toys can tolerate a few lost bits without any audible impact. These high error rate, lossy memories are generally not usable for mainstream digital storage and thus have had limited acceptance. The goal of the M.L.C. program was to produce a two bit/cell digital storage technology capable of penetrating the larger non-volatile memory market, enabling the growth of new digital flash memory applications.

The 1992 View of M.L.C.

In the early 1990s, flash memory was considered as a potential replacement for hard disks at lower densities for applications that require small, rugged, and low-power storage. One of the main issues for use of flash in this application was the high cost of the flash memory as compared to that of magnetic storage. A lower cost flash memory was required. The hard disk requirements are much relaxed over silicon memory due to the inclusion of error correction in the hard disk subsystem, the block transfer of data (no byte access), and the relative low read performance. Multi-level cell technology appeared to be an ideal solution for the solid state disk, addressing the lower cost through two bit/cell (and later three or four bit/cell) technology. The use of error correction, and the large block transfer of data in the solid state disk would address any reliability issues with multi-level storage. The Intel M.L.C. program was thus started with a goal of a high-density, low-cost, solid-state disk.

The basic techniques for accurate charge placement and sensing were developed in the lab and implemented into a 32Mbit silicon test chip. During this time frame, the three major challenges for multi-bit storage were identified:

- *Precise Charge Placement*: The flash memory cell programming must be very accurately controlled, requiring a detailed understanding of the physics of programming as well as the control and timing of the voltages applied to the cell.
- *Precise Charge Sensing*: The read operation of a M.L.C. memory is basically an analog to digital conversion of the analog charge stored in the memory cell to digital data- a concept new to memory devices.
- *Stable Charge Storage*: Meeting the data retention goals would require the stored charge to be stable with a leakage rate of less than one electron per day.

The 32Mbit test chip clearly demonstrated the ability to store multiple bits in a single cell. Based on the functionality of this device, the M.L.C. technology was announced in 1994.

The 'First' M.L.C. 'Product'

With the knowledge gained from the 32Mbit test chip, the first attempt at a two bit/cell storage product was started. This device was aimed at the solid-state disk goal. The solid-state disk system would include error correction and would generate non-standard voltages to interface to the two bit/cell memory device. A special DC to DC voltage converter was commissioned that would generate 12v$\pm$1% and 5.5$\pm$1% from a 3v$\pm$10% external supply. The M.L.C. part required these precise supply voltages to perform the accurate program and read operations. An error corrector was designed to be integrated with the other control logic of the solid state disk. A paper based on this 32Mbit M.L.C. memory was presented at the prestigious International Solid State Circuits Conference (ISSCC) in 1995[2], winning the best paper of the conference award. The 32Mbit device became the workhorse for the M.L.C. technology development effort, demonstrating the ability of M.L.C. to meet Intel's stringent reliability requirements and to produce yield equivalent to single bit/cell flash memories. It was also used to develop the M.L.C. testing and to debug the manufacturing process for test and packaging.

The Question of Reliability

The primary concern for M.L.C. was the reliability of the storage of the multiple charge states. Charge states would be separated by a few thousand electrons in an M.L.C. device, and a loss of one electron per day from the floating gate could result in a bit error after ten years of storage. To understand the detailed physics of charge storage, a large experiment was started to monitor the charge storage behavior of 200 billion cells (2×10^{11} cells). This massive experiment could resolve changes in the stored charge of as small as 100 electrons on all of the cells under evaluation. The rate of charge loss was accelerated through the use of elevated temperatures. This experiment, which was started in early 1994, is still running today with an accumulated high temperature stress time of over three years, representing over 50 years at normal operating temperatures. The knowledge gained and models developed based on this experiment have resulted in changes to the design of both the product and the process, allowing removal of the error correction

requirement for two bit/cells. This data fundamentally changed the direction of the multi-bit storage program.

Removing the Constraints

Toward the end of 1995, the M.L.C. project had grown from a small research effort to a full blown program. Almost two years worth of reliability data was showing excellent performance indicating that the error corrector was not required. The 32Mbit device had demonstrated the viability of the circuit techniques and the device physics used for the precision program and read operations. Moreover, the yield was looking excellent, and the manufacturing issues were understood. Test circuits had demonstrated the ability to provide the required voltages and voltage regulation on the memory chip, eliminating the need for the external DC to DC converter. It became clear that the project could accomplish much more than the initial vision of a solid state disk. The team believed that it was possible to remove the two major requirements initially envisioned for M.L.C.: error correction and precision external power supplies. The solid state disk market, while developing, had not reached the desired volume levels. The decision was made to not take the 32Mbit device to production and focus on the design of an M.L.C. two bit/cell part with functionality substantially equivalent to the standard one bit/cell products.

The 1997 View of M.L.C.

The first two bit/cell Intel StrataFlash memory device was introduced in September of 1997, a 64Mbit device. This device has functionality that is largely equivalent to the standard one bit/cell flash products. A highlight comparison of the Intel StrataFlash memory features to an Intel 16Mbit single bit/cell product is shown in Figure 9.

	1b/c Flash Memory	Intel StrataFlash™ 2b/c Memory	
Density	16Mbit	32Mbit	64Mbit
Read Speed	100 ns	120 ns	150 ns
Block Size	64KByte	128KByte	
Architecture	x8	x8 / x16	
V_{cc} Power Supply (+/-10%)	5V	5V	
V_{pp} (Program/Erase Voltage)	5V or 12V	5V	
Effective Write Speed	6 µS/Byte	6 µs/Byte	
I_{ccr} (Read Current)	35 mA	55 mA	
I_{ppw} + I_{ccw} (Write Current)	75 mA	90 mA	
Endurance	100,000 Cycles	10,000 Cycles	
Operating Temperature	Extended	Comercial	

Figure 9: Comparison of 1b/c and 2b/c product features

Read performance is in line with expectations for memories of 32Mbit and 64Mbit densities with about a 20% increase in read access time for a doubling of memory density. Two bits/cell doubles the erase block size as compared to one bit/cell since each cell now stores twice as much data. The power supply is maintained at the 5v industry standard. The two bit/cell write performance is maintained equivalent to one/bit cell, even with the more complex (and slower) precision write algorithm, through the use of an eight-byte write buffer and a higher write bandwidth into the array. The 10,000 erase/write endurance specification is more than acceptable for virtually all flash applications and easily justified by the reduced cost.

The 64Mbit device integrates all of the knowledge gained from the two previous test vehicles and advances beyond them with the introduction of precision internal voltage regulation and internal test capability. The first silicon wafer out of the manufacturing line was fully functional, and the program is on track for volume shipments. The 64Mbit two bit/cell Intel StrataFlash memory is just 5% larger than the 32Mbit one bit/cell device on the 0.4µ ETOX flash memory process, delivering on the promise of 2x the bits in 1x the space and setting a new cost paradigm for flash memory devices. A photomicrograph of the 64Mbit Intel StrataFlash memory is shown in Figure 10.

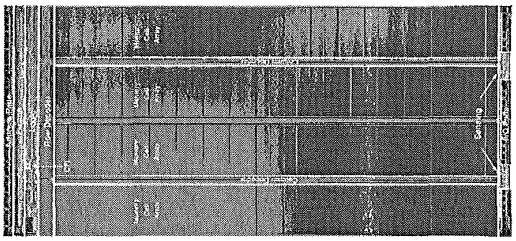

Figure 10: The Intel StrataFlash 64Mbit memory

Examples of Intel StrataFlash Memory Usage

Intel StrataFlash memory is finding acceptance in a wide variety of applications that all share a common need for high density and low cost. Applications evaluating the use of Intel StrataFlash memory include small office voice-mail PBX systems to store incoming messages, network routers to store operating programs, digital cameras for digital image "film," digital voice recorders for digital audio "tape," Windows* CE hand-held computers for storage of programs and data, and set-top boxes for storage of programs and data. As for the goal of replacing hard disks in small, low-power, or rugged systems, Intel StrataFlash memory will enable solid state

disks to be cheaper than hard disks for capacities of 32MByte and lower.

The Future of M.L.C.

With two bits per cell well along the way to mainstream production, attention turns to the prospects of even higher bit per cell densities. It is now recognized that the Intel ETOX flash memory cell and array structure is ideally suited to the storage of multiple bits per cell due to its simple direct access array architecture and proven high volume manufacturability. The direct access memory array allows the precise control of the cell voltages required to reliably and reproducibly place precise amounts of charge on the floating gate. The well controlled, high yielding ETOX process technology ensures that sufficient process margins exist for the more stringent M.L.C. requirements. Three bits per cell storage has been demonstrated in the laboratory, achieving state widths of less than 0.15v on a 48Mbit test chip.

Conclusion

The concept of multi-level storage using an ETOX flash memory cell transistor has been demonstrated. This concept builds on traditional semiconductor process scaling, providing the cost structure of the next-generation silicon process technology with the current generation of silicon process equipment. The evolution of the multi-level-cell development from concept in 1992 to production in 1997 has required many innovations in the areas of device physics, circuit design, and product test. The combination of precisely controlled on-chip voltages and timing, the direct access ETOX flash memory array architecture, and a highly manufacturable silicon process technology has resulted in a two bit/cell memory device largely identical to the industry standard one bit/cell flash memory. The Intel StrataFlash memory sets a new cost paradigm for the flash industry.

*All other brands and names are the property of their respective owners.

Acknowledgments

A program of this scope requires many dedicated hours by many creative individuals. The authors would like to thank the members of the Intel StrataFlash memory development team for the effort leading to the world's first two bit/cell ETOX flash memory product.

References

[1]Frohman-Bentchowsky, Floating Gate Solid State Storage Device and Methodology for Charging and Discharging Same, U.S. Patent #3,755,721, Aug. 28, 1973.

[2] Bauer, M., et. al., "A Multilevel-Cell 32Mb Flash Memory," Technical Digest IEEE International Solid State Circuits Conference, 1995, pp. 132,133.

Authors' Biographies

Greg Atwood is an Intel Fellow and is the director of Flash Memory Architecture. He received an M.Sc. degree in Physics from Purdue University in 1979 and joined Intel the same year. He has worked on numerous technology development programs including Logic, SRAM, EPROM, E^2PROM, Flash, and Multi-Level Flash. He was the program manager for the Intel StrataFlash memory development program responsible for all aspects of the project. He holds 20 patents covering a wide range of technical topics. Greg is presently responsible for next-generation Flash and Intel StrataFlash memory technologies. His e-mail address is greg_atwood@ccm.sc.intel.com.

Al Fazio is a principal engineer in Flash Technology development. He received a B.Sc. in Physics from the State University of New York at Stony Brook in 1982 and joined Intel the same year. He has been involved in development programs such as SRAM, EPROM, E^2PROM, NVRAM, and Flash Memories. He was responsible for the technology development of the Intel StrataFlash memory. He holds more than a dozen patents and has authored or co-authored several technical papers. He is presently responsible for Intel's Multi-Level-Cell and Advanced Flash Memory Cell development and currently serves as general chairman of the IEEE Non-Volatile Semiconductor Memory Workshop. His e-mail address is al_fazio@ccm.sc.intel.com

Duane Mills is the Intel StrataFlash Memory Design Engineering Manager. He received a B.Sc. in Electrical Engineering from Purdue University in 1984 and joined Intel the same year. He has worked on numerous EPROM and Flash memory designs during the past 13 years. Mr. Mills was the design manager for the first Intel StrataFlash memory product and is currently responsible for the definition and design of the next-generation Intel StrataFlash product. His e-mail address is duane_r_mills@ccm.fm.intel.com.

Bill Reaves is the Intel StrataFlash Memory Product Engineering Manager. He received a B.Sc. in Electrical

and Computer Engineering from U.C. Santa Barbara in 1984 and joined Intel the same year. He has been involved in numerous non-volatile memory projects, including the test development and manufacturing start-up of Intel's first flash memories. He is currently the Product Engineering manager for the Intel StrataFlash Memory program and the project manager for the initial product line. His e-mail address is bill_t_reaves@ccm.fm.intel.com.