Chapter 1

INTRODUCTION TO DATA STORAGE

TECHNOLOGIES

There are two broad classifications of the memories.

1. Volatile (Programmable)

Category 1 generally deals with the Programmable devices such as Hard Disk drive, RAM, whereas Category 2 invariably consists of ROM.

We begin with a brief overview of the existing technologies in these domains.

1.1 MAGNETIC STORAGE

Magnetic storage and magnetic recording are terms from engineering referring to the storage of data on a magnetized medium. Magnetic storage uses different patterns of magnetization in a magnetizable material to store data and is a form of non-volatile memory. The information is accessed using one or more read/write heads. As of 2009, magnetic storage media, primarily hard disks, are widely used to store computer data as well as audio and video signals.

1.1.1 Analog recording

Analog recording is based on the fact that remnant magnetization of a given material depends on the magnitude of the applied field. The magnetic material is normally in the form of tape, with the tape in its blank form being initially demagnetized. When recording, the tape runs at a constant speed. The writing head magnetizes the tape with current proportional to the signal. A magnetization distribution is achieved along the magnetic tape. Finally, the distribution of the magnetization can be read out, reproducing the original signal. The magnetic tape is typically made by embedding magnetic particles in a plastic binder on polyester film tape. The commonly used magnetic particles are Iron...
oxide particles or Chromium oxide and metal particles with size of 0.5 micrometers. Analog recording was very popular in audio and video recording. In the past 20 years, however, tape recording has been gradually replaced by digital recording.

1.1.2 Digital recording

Instead of creating a magnetization distribution in analog recording, digital recording only need two stable magnetic states, which are the +Ms and -Ms on the hysteresis loop. Examples of digital recording are floppy disks and HDDs.

1.1.3 Access method

Magnetic storage media can be classified as either sequential access memory or random access memory although in some cases the distinction is not perfectly clear. In the case of magnetic wire, the read/write head only covers a very small part of the recording surface at any given time. Accessing different parts of the wire involves winding the wire forward or backward until the point of interest is found. The time to access this point depends on how far away it is from the starting point. The case of ferrite-core memory is the opposite. Every core location is immediately accessible at any given time.

Hard disks and modern linear serpentine tape drives do not precisely fit into either category. Both have many parallel tracks across the width of the media and the read/write heads take time to switch between tracks and to scan within tracks. Different spots on the storage media take different amounts of time to access. For a hard disk this time is typically less than 10 ms, but tapes might take as much as 100 s.
1.2 DRAM

Dynamic random access memory (DRAM) is a type of random access memory that stores each bit of data in a separate capacitor within an integrated circuit. Since real capacitors leak charge, the information eventually fades unless the capacitor charge is refreshed periodically. Because of this refresh requirement, it is a dynamic memory as opposed to SRAM and other static memory.

The advantage of DRAM is its structural simplicity: only one transistor and a capacitor are required per bit, compared to four transistors in SRAM. This allows DRAM to reach very high density. Unlike flash memory, it is volatile memory (cf. non-volatile memory), since it loses its data when the power supply is removed.

1.2.1 BASIC OPERATION

DRAM is usually arranged in a square array of one capacitor and transistor per cell. The long lines connecting each row are known as word lines. Each column is actually composed of two bit lines, each one connected to every other storage cell in the column. They are generally known as the + and − bit lines. A sense amplifier is essentially a pair of cross-connected inverters between the bit lines. That is, the first inverter is connected from the + bit line to the − bit line, and the second is connected from the − bit line to the + bit line.
Figure 2. Simple 4X4 DRAM read write ckt.
1.3 SRAM

Static Random Access Memory (SRAM) is a type of semiconductor memory where the word static indicates that, unlike dynamic RAM (DRAM), it does not need to be periodically refreshed, as SRAM uses bistable latching circuitry to store each bit. SRAM exhibits data remanence, but is still volatile in the conventional sense that data is eventually lost when the memory is not powered.

![A 6 Transistor CMOS SRAM cell](image)

Figure 3. A 6 Transistor CMOS SRAM cell

1.4 SDRAM

Synchronous dynamic random access memory (SDRAM) is dynamic random access memory (DRAM) that has a synchronous interface. Traditionally, dynamic random access memory (DRAM) has an asynchronous interface which means that it responds as quickly as possible to changes in control inputs. SDRAM has a synchronous interface, meaning that it waits for a clock signal before responding to control inputs and is therefore synchronized with the computer's system bus. The clock is used to drive an internal finite state machine that pipelines incoming instructions. This allows the chip to have a more complex pattern of operation than asynchronous DRAM which does not have a synchronized interface.
PROGRAMMABLE ROM (PROM)

A programmable read-only memory (PROM) or field programmable read-only memory (FPROM) or one-time programmable non-volatile memory (OTP NVM) is a form of digital memory where the setting of each bit is locked by a fuse or antifuse. Such PROMs are used to store programs permanently. The key difference from a strict ROM is that the programming is applied after the device is constructed. These types of memories are frequently seen in video game consoles, mobile phones, radio-frequency identification (RFID) tags, implantable medical devices, high-definition multimedia interfaces (HDMI) and in many other consumer and automotive electronics products.

ERASABLE PROGRAMMABLE ROM (EPROM)

An EPROM, or erasable programmable read only memory, is a type of memory chip that retains its data when its power supply is switched off. In other words, it is non-volatile. It is an array of floating-gate transistors individually programmed by an electronic device that supplies higher voltages than those normally used in digital circuits. Once programmed, an EPROM can be erased only by exposing it to strong ultraviolet light. That UV light usually has a wavelength of 253.7nm (for optimum erasure time) and belongs to the UVC range of UV light. EPROMs are easily recognizable by the transparent fused quartz window in the top of the package, through which the silicon chip is visible, and which permits exposure to UV light during erasing.
1.6.1 OPERATION

Each storage location of an EPROM consists of a single field-effect transistor. Each field-effect transistor consists of a channel in the semiconductor body of the device. Source and drain contacts are made to regions at the end of the channel. An insulating layer of oxide is grown over the channel, then a conductive (silicon or aluminum) gate electrode is deposited, and a further thick layer of oxide is deposited over the gate electrode. The floating gate electrode has no connections to other parts of the integrated circuit and it is completely insulated by the surrounding layers of oxide. A control gate electrode is deposited and further oxide covers it.

To retrieve data from the EPROM, the address represented by the values at the address pins of the EPROM is decoded and used to connect one word (usually an 8-bit byte) of storage to the output buffer amplifiers. Each bit of the word is a 1 or 0, depending on the storage transistor being switched on or off, conducting or non-conducting.
The switching state of the field-effect transistor is controlled by the voltage on the control gate of the transistor. Presence of a voltage on this gate creates a conductive channel in the transistor, switching it on. In effect, the stored charge on the floating gate allows the threshold voltage of the transistor to be programmed.

Storing data in the memory requires selecting a given address and applying a higher voltage to the transistors. This creates an avalanche discharge of electrons, which have enough energy to pass through the insulating oxide layer and accumulate on the gate electrode. When the high voltage is removed, the electrons are trapped on the electrode. Because of the high insulation value of the silicon oxide surrounding the gate, the stored charge cannot readily leak away and the data can be retained for decades.

Figure 6. Cross Section of a Conventional MOS Transistor and a Floating-Gate MOS Transistor

Figure 7. Cross-section of a floating gate transistor
**EPROM has two significant limitations:**

1. It has to be taken out of the circuit and put under an ultraviolet light source to erase it.
2. Higher (non-TTL) voltages are generally required to program it. These voltages are normally not available in regular computer circuit boards.
3. It facilitates complete erasure even if a small block of memory needs changing.

![Figure 8. A 32Kb EPROM](image)
1.7 **Electrically Erasable Programmable Read-Only Memory (EEPROM)**

EPROM stands for Electrically Erasable Programmable Read-Only Memory. An EEPROM is like an EPROM chip since it can be written in or programmed more than once. Unlike the EPROM chip, however, an EEPROM chip need not be taken out of the computer or electronic device of which it is part when a new program or data needs to be written on it.

Selective programming can be done to an EEPROM chip. The user can alter the value of certain cells without needing to erase the programming on other cells. Thus, sections of data can be erased and replaced without needing to alter the rest of the chip's programming.

Data stored in an EEPROM chip is permanent, at least until the user decides to erase and replace the information it contains. Furthermore, the data stored in an EEPROM chip is not lost even when power is turned off.

### 1.7.1 EEPROM STRUCTURE

The EEPROM chip is physically similar to the EPROM chip. It is also composed of cells with two transistors. The floating gate is separated from the control gate by a thin oxide layer. Unlike the EPROM chip, however, the EEPROM chip's oxide layer is much thinner. In EEPROM chips, the insulating layer is only around 1 nanometer thick whereas in EPROM chips, the oxide layer is around 3 nanometers thick. The thinner oxide layer means lower voltage requirements for initiating changes in cell value.

Tunneling the electrons of the floating gate towards the oxide layer separating the floating gate and the control gate is still the method of changing a bit's value from 1 to 0. To erase EEPROM programming, the electron barrier still has to be overcome by the application of enough programming voltage.
1.7.2 COMPARISON BETWEEN EPROM AND EEPROM

The difference between EPROM and EEPROM lies in the way that the memory programs and erases. EEPROM can be programmed and erased electrically using field electron emission (more commonly known in the industry as "Fowler–Nordheim tunneling").

EPROMs can't be erased electrically, and are programmed via hot carrier injection onto the floating gate. Erase is via an ultraviolet light source, although in practice many EPROMs are encapsulated in plastic that is opaque to UV light, and are "one-time programmable"

1.7.3 EEPROM memory failure modes

One of the main problems with EEPROM technology is its overall reliability. There are two ways in which these memory devices can fail:

1. **Endurance:** It is found that during the rewrite operations of the EEPROM memory, the gate oxide in the floating-gate transistors of the memory cell gradually accumulate trapped electrons. The electric field associated with these trapped electrons combines with that of the wanted electrons in the floating gate. As a result the state where there are no electrons in the floating gate still has a residual field, and as this rises as more electrons become trapped, a condition eventually rises when it is not possible to differentiate between the thresholds for the zero state cannot be detected and the cell is stuck in programmed state. The manufacturers usually specify minimal number of rewrite cycles being 10 million or more

2. **Data retention time:** The data retention time is also very important, especially if the EEPROM contains software that is required for the operation of an item of electronics equipment. The data retention period is limited. This results from the fact that during storage, the electrons injected into the floating gate may drift through the insulator as it is not a perfect insulator, especially at increased temperature. This causes any charge being stored in the floating gate to be lost and the memory cell will revert to its erased state.
The time taken for this to happen is very long, and manufacturers usually guarantee data retention of 10 years or more for most devices.

1.8  FLASH MEMORY

Flash memory is a non-volatile computer storage that can be electrically erased and reprogrammed. It is a technology that is primarily used in memory cards and USB flash drives for general storage and transfer of data between computers and other digital products. It is a specific type of EEPROM (Electrically Erasable Programmable Read-Only Memory) that is erased and programmed in large blocks; in early flash the entire chip had to be erased at once. Flash memory costs far less than byte-programmable EEPROM and therefore has become the dominant technology wherever a significant amount of non-volatile, solid state storage is needed. Example applications include PDAs (personal digital assistants), laptop computers, digital audio players, digital cameras and mobile phones. It has also gained popularity in console video game hardware, where it is often used instead of EEPROMs or battery-powered static RAM (SRAM) for game save data.

Since flash memory is non-volatile, no power is needed to maintain the information stored in the chip. In addition, flash memory offers fast read access times (although not as fast as volatile DRAM memory used for main memory in PCs) and better kinetic shock resistance than hard disks. These characteristics explain the popularity of flash memory in portable devices. Another feature of flash memory is that when packaged in a "memory card," it is enormously durable, being able to withstand intense pressure, extremes of temperature, and even immersion in water.

Although technically a type of EEPROM, the term "EEPROM" is generally used to refer specifically to non-flash EEPROM which is erasable in small blocks, typically bytes. Because erase cycles are slow, the large block sizes used in flash memory erasing give it a significant speed advantage over old-style EEPROM when writing large amounts of data.
1.8.1 OPERATION

Flash memory is a type of EEPROM chip. It has a grid of columns and rows with a cell that has two transistors at each intersection (see image below). The two transistors are separated from each other by a thin oxide layer. One of the transistors is known as a...
floating gate, and the other one is the control gate. The floating gate's only link to the row, or wordline, is through the control gate. As long as this link is in place, the cell has a value of 1. To change the value to a 0 requires a curious process called Fowler-Nordheim tunneling.

Tunneling is used to alter the placement of electrons in the floating gate. An electrical charge, usually 10 to 13 volts, is applied to the floating gate. The charge comes from the column, or bitline, enters the floating gate and drains to a ground.

This charge causes the floating-gate transistor to act like an electron gun. The excited electrons are pushed through and trapped on other side of the thin oxide layer, giving it a negative charge. These negatively charged electrons act as a barrier between the control gate and the floating gate. A special device called a cell sensor monitors the level of the charge passing through the floating gate. If the flow through the gate is greater than 50 percent of the charge, it has a value of 1. When the charge passing through drops below the 50-percent threshold, the value changes to 0. A blank EEPROM has all of the gates fully open, giving each cell a value of 1.

The electrons in the cells of a Flash-memory chip can be returned to normal ("1") by the application of an electric field, a higher-voltage charge. Flash memory uses in-circuit wiring to apply the electric field either to the entire chip or to predetermined sections known as blocks. This erases the targeted area of the chip, which can then be rewritten. Flash memory works much faster than traditional EEPROMs because instead of erasing one byte at a time, it erases a block or the entire chip, and then rewrites it.

1.8.2 LIMITATIONS

Block erasure

One limitation of flash memory is that although it can be read or programmed a byte or a word at a time in a random access fashion, it must be erased a "block" at a time. This generally sets all bits in the block to 1. Starting with a freshly erased block, any location within that block can be programmed. However, once a bit has been set to 0, only by erasing the entire block can it be changed back to 1. In other words, flash memory
(specifically NOR flash) offers random-access read and programming operations, but cannot offer arbitrary random-access rewrite or erase operations. A location can, however, be rewritten as long as the new value's 0 bits are a superset of the over-written value's. For example, a nibble value may be erased to 1111, and then written as 1110. Successive writes to that nibble can change it to 1010, then 0010, and finally 0000. File systems built on NOR flash made use of this capability to represent sector metadata.

Although data structures in flash memory cannot be updated in completely general ways, this allows members to be "removed" by marking them as invalid. This technique may need to be modified for Multi-level Cell devices, where one memory cell holds more than one bit.

Memory wear

Another limitation is that flash memory has a finite number of erase-write cycles. Most commercially available flash products are guaranteed to withstand around 100,000 write-erase-cycles, before the wear begins to deteriorate the integrity of the storage. Micron Technology and Sun Microsystems announced an SLC flash memory chip rated for 1,000,000 write-erase-cycles on December 17, 2008.

The guaranteed cycle count may apply only to block zero (as is the case with TSOP NAND parts), or to all blocks (as in NOR). This effect is partially offset in some chip firmware or file system drivers by counting the writes and dynamically remapping blocks in order to spread write operations between sectors; this technique is called wear leveling. Another approach is to perform write verification and remapping to spare sectors in case of write failure, a technique called Bad Block Management (BBM). For portable consumer devices, these wear out management techniques typically extend the life of the flash memory beyond the life of the device itself, and some data loss may be acceptable in these applications. For high reliability data storage, however, it is not advisable to use flash memory that would have to go through a large number of programming cycles. This limitation is meaningless for 'read-only' applications such as thin clients and routers, which are only programmed once or at most a few times during their lifetime.
1.8.3 FLASH SCALABILITY

Figure 12. Scalability of flash devices over years, Courtesy Intel Corp.

The aggressive trend of process design rule shrinks in NAND Flash memory technology effectively accelerates Moore’s Law.

Due to its relatively simple structure and high demand for higher capacity, NAND flash memory is the most aggressively scaled technology among electronic devices. The heavy competition among the top few manufacturers only adds to the aggression. Current projections show the technology to reach approximately 20 nm by around 2010. While the expected shrink timeline is a factor of two every three years per original version of Moore’s law, this has recently been accelerated in the case of NAND flash to a factor of two every two years.
1.9 MOORE’S LAW

Moore's Law describes a long-term trend in the history of computing hardware, in which the number of transistors that can be placed inexpensively on an integrated circuit has doubled approximately every two years. The law is named for Intel co-founder Gordon E. Moore, who introduced it in a 1965 paper. It has since been used in the semiconductor industry to guide long term planning and to set targets for research and development. Rather than being a naturally-occurring "law" that cannot be controlled, however, Moore's Law is effectively a business practice in which the advancement of transistor counts occurs at a fixed rate.

The capabilities of many digital electronic devices are strongly linked to Moore's law: processing speed, memory capacity, sensors and even the number and size of pixels in digital cameras. All of these are improving at (roughly) exponential rates as well. This has dramatically increased the usefulness of digital electronics in nearly every segment of the world economy. Moore's law precisely describes a driving force of technological and social change in the late 20th and early 21st centuries.
Looking forward, maintaining the pace of Moore's law scaling is increasingly difficult. The challenging areas include electrical, physical and reliability characteristics of the memory cell. However, through innovative device design, introduction of new materials and memory error management, we anticipate that floating gate flash memories are economically viable beyond 45 nm to 32 nm.

It is difficult to maintain the Moore’s Law curve using the transistor technology alone. As the process complexity increase to address the fundamental limits of Physics, Moore’s law can only continue through innovations, in terms of new memory structures and new materials.

******
CHAPTER 2

PHASE CHANGE MEMORIES (PCM): ANSWER TO MOORE’S LAW PROBLEM

Most nonvolatile memory devices are Flash memory chips, so-called because of the ability to write them individually while erasing them in chunks. This type of device is ubiquitous in today’s cell phones, digital cameras, media cards, etc. But Flash memory suffers from several shortcomings that limit its market potential. Primarily, writing data to a Flash memory is too slow for Flash to rival its DRAM cousins. Secondly, Flash memories can only be reprogrammed a limited number of times, typically on the order of a million. While this may be enough for certain applications, programming limitations make Flash memory ill-suited for general computing or intense data manipulation applications. Perhaps most importantly, the nonvolatile memory industry recognizes that current flash memory technology will soon face a crisis: along with the rest of computer technology, flash chips must shrink to remain profitable, but the physics of the barrier oxide used in current flash technology prevents memory cells from being shrunk much further.

As a consequence, a number of different nonvolatile memory technologies are emerging as potential alternatives to replace Flash, most prominently Ferro-electric RAM (FRAM or FeRAM), Magneto-resistive RAM (MRAM), Chalcogenide RAM (CRAM, but also called Ovonic Unified Memory (OUM), or Phase-Change RAM (PRAM)). These devices have little in common, except that they can be reprogrammed a near unlimited number of times and be programmed in nanoseconds rather than microseconds.

Phase Change Memory (PCM) is a term used to describe a class of non-volatile memory devices that employ a reversible phase change in materials to store information. Matter can exist in various phases such as solid, liquid, gas, condensate and plasma. PCM exploits differences in the electrical resistivity of a material in different phases.

The PCM technology being developed by Intel uses a class of materials known as chalcogenides (“kal-koj–uh-nyde”). Chalcogenides are alloys that contain an element in the Oxygen/Sulphur family of the Periodic Table (Group 16 in the new style or Group
VIa in the old style Periodic Table). Numonyx PCM is using an alloy of Germanium, Antimony and Tellurium (Ge2Sb2Te5), known more commonly as “GST”. Most companies performing research and development in PCM today are using GST or closely related alloys.

### 2.1 THEORY OF OPERATION

Phase change chalcogenides exhibit a reversible phase change phenomenon when changed from the amorphous phase to the crystalline phase. As shown in Figure 1, in the amorphous phase, the material is highly disordered—there is an absence of regular order to the crystal-line lattice. In this phase, the material demonstrates high resistivity and high reflectivity. In contrast, in the polycrystalline phase, the material has a regular crystalline structure and exhibits low reflectivity and low resistivity.

![Figure 1](image)

**Figure 1.** Comparison of the states of a PCM, Courtesy Numonyx whitepaper

In PCM, we are exploiting the difference in resistivity between the two phases of the material. This phase change is induced in the material through intense localized Joule heating caused by current injection. The end phase of the material is modulated by the magnitude of the injected current, the applied voltage, and the time of the operation. Figure 2 shows a graphical representation of a basic PCM storage element. As shown on
the left, a layer of chalcogenide is sandwiched between a top electrode and a bottom electrode. A resistive heating element extends from the bottom electrode and contacts a layer of the chalcogenide material. Current injected into the junction of the chalcogenide and the heater induces the phase change through Joule heating. At right is the actual implementation of the concept, showing an amorphous bit formed in a layer of polycrystalline chalcogenide. Because of the change in reflectivity, the amorphous bit appears as a mushroom cap shaped structure in the layer of polycrystalline chalcogenide.

2.2 PCM ATTRIBUTES AND CAPABILITIES

Phase Change Memory blends the attributes commonly associated with NOR-type flash, memory NAND-type flash memory, and RAM or EEPROM. These attributes are summarized in the chart in the figure below.

![PCM Attributes Chart](image)

Figure 15. PCM Attributes: This new class of non-volatile memory brings together the best attributes of NOR, NAND and RAM.

**Bit-alterable**

Like RAM or EEPROM, PCM is bit alterable. Flash technology requires a separate erase step in order to change information. Information stored in bit-alterable memory can be switched from a one to zero or zero to a one without a separate erase step.
Non-volatile

Just like NOR flash and NAND flash, PCM is nonvolatile. RAM, of course, requires a constant power supply, such as a battery backup system, to retain information. DRAM technologies also suffer from susceptibility to so-called “soft errors” or random bit corruption caused by alpha particles or cosmic radiation. Early testing results conducted by Intel on multimegabit PCM arrays for long term data retention show excellent results.

Read speed

Like RAM and NOR-type flash, the technology features fast random access times. This enables the execution of code directly from the memory, without an intermediate copy to RAM. The read latency of PCM is comparable to single bit per cell NOR flash, while the read bandwidth can match DRAM. In contrast, NAND flash suffers from long random access times on the order of 10s of microseconds that prevent direct code execution.

Write/erase speed

PCM is capable of achieving write speeds like NAND, but with lower latency and with no separate erase step required. NOR flash features moderate write speeds but long erase times. As with RAM, no separate erase step is required with PCM, but the write speed (bandwidth and latency) does not match the capability of RAM today. The capability of PCM is expected, however, however, to improve with each process generation as the PCM cell area decreases.

Scaling

Scaling is the fifth area where PCM will offer a difference. Both NOR and NAND rely on memory structures which are difficult to shrink at small lithos. This is due to gate thickness remaining constant and the need for operation voltage of more than 10V while the operation of CMOS logic has been scaled to 1V or even less. This scaling effect is often referred to as Moore’s Law, where memory densities double with each smaller generation. With PCM, as the memory cell shrinks, the volume of GST material shrinks as well, providing a truly scalable solution.
2.3 THE SOLUTION

Phase Change Memory is a promising memory technology that has recently experienced a resurgence of interest. PCM employs a reversible phase change phenomenon to store information through a resistance change in different phases of a material. PCM offers a combination of some of the best attributes of NOR flash, AND flash, EEPROM and RAM in a single memory device. These capabilities uniquely combined with the potential for lower memory subsystem costs could potentially create new applications and memory architectures in a wide range of systems.
CHAPTER 3

OVONICS UNIFIED MEMORY (OUM)

3.1 INTRODUCTION

The use of phase-change chalcogenide alloy films to store data electrically and optically was first reported in 1968 and in 1972, respectively. Early phase-change memory devices used tellurium-rich, multi-component chalcogenide alloys with a typical composition of $\text{Te}_{81}\text{Ge}_{15}\text{Sb}_{2}\text{S}_{2}$. Both the optical and electrical memory devices were programmed by application of an energy pulse of appropriate magnitude and duration. A short pulse of energy was used to melt the material, which was then allowed to cool quickly enough to “freeze in” the glassy, structurally disordered state. To reverse the process, a somewhat lower-amplitude, longer-duration pulse was used to heat a previously vitrified region of the alloy to a temperature below the melting point, at which crystallization could occur rapidly. Differences in electrical resistivity and the optical constants between the amorphous and polycrystalline phases were used to store data.

During the 1970s and 1980s, significant research efforts by many industrial and academic groups were focused on understanding the fundamental properties of chalcogenide alloy amorphous semiconductors. Prototype optical memory disks and electronic memory device arrays also were announced, beginning in the early 1970s. Rapidly crystallizing chalcogenide alloys were later reported by several optical memory research groups. These new material compositions, derived from the Ge-Te-Sb ternary system, did not phase-segregate upon crystallization like the earlier Te-rich alloys, but instead exhibited congruent crystallization with no large-scale atomic motion.

In the 1990s, researchers at Energy Conversion Devices Inc. and Ovonyx Inc. developed new, thermally optimized phase-change memory device structures that exploited rapidly crystallizing chalcogenide alloy materials to achieve increased programming speed and reduced programming current.
These devices could be programmed in 20 ns—about six orders of magnitude faster than the early phase-change memory cells, and their much lower programming current requirements permitted the design of memory arrays using memory bit access devices (transistors or diodes) fabricated at minimum litho-graphic dimensions. Ovonyx is now commercializing its phase-change memory technology called Ovonics Unified Memory (OUM) through a number of license agreements and joint development programs with semiconductor device manufacturers.

Figure 16. Comparison of Memory Technologies
3.2 CHALCOGENIDES

Figure 17. The highlighted elements form the basis of Chalcogenides

The crystalline and amorphous states of chalcogenide glass have dramatically different electrical resistivity, and this forms the basis by which data are stored. The amorphous, high resistance state is used to represent a binary 0, and the crystalline, low resistance state represents a 1. Chalcogenide is the same material used in re-writable optical media (such as CD-RW and DVD-RW). In those instances, the material's optical properties are manipulated, rather than its electrical resistivity, as chalcogenide's refractive index also changes with the state of the material.
The term “chalcogen” refers to the Group VI elements of the periodic table. “Chalcogenide” refers to alloys containing at least one Group VI element such as the alloy of germanium, antimony, and tellurium discussed here. Energy Conversion Devices, Inc. has used this particular alloy to develop a phase-change memory technology used in commercially available re-writeable CD and DVD disks. This phase-change technology uses a thermally activated, rapid, reversible change in the structure of the alloy to store data. Since the binary information is represented by two different phases of material it is inherently non-volatile, requiring no energy to keep the material in either of its two stable structural states.

Used in a binary mode, the two structural states of the chalcogenide alloy, as shown in Figure, are an amorphous state (no long-range order of atoms) and a polycrystalline state (composed of many crystals, each having atoms placed in a repetitive order). Relative to the amorphous state, the polycrystalline state shows a dramatic increase in free electron density (similar to a metal). This difference in free electron density gives rise to a difference in reflectivity and, more importantly, resistivity. In the case of the rewriteable CD and DVD disk technology, this difference in reflectivity is used to read the state of each memory bit by directing a low-power laser at the material and detecting the amount of light reflected.

Figure 18. Experimental OUM Chip, Courtesy Ovonyx
3.3 **Storage Mechanism: The Amorphous–Crystalline Structural Phase Change**

Glassy materials are produced by rapidly super cooling a liquid below its melting point to a temperature at which the atomic motion necessary for crystallization cannot readily occur. Chalcogenide alloys - materials containing one or more elements from Group VI of the periodic table—are typically good glass formers, in large part because the Group VI elements form predominantly twofold-coordinated covalent chemical bonds that can produce linear, tangled, polymer like clusters in the melt. This increases the viscosity of the liquid, inhibiting the atomic motion necessary for crystallization. Many amorphous chalcogenide alloys have been reported in the literature. The Ge2Sb2Te5 (GST 225) chalcogenide alloy currently used in OUM memory devices melts at approximately 610°C and has a glass-transition temperature of 350°C. In order to crystallize an amorphous region of GST 225, the material must be heated to a temperature somewhat below the melting point and held at this temperature for a time sufficient to allow the crystallization to occur. The compositional dependence of crystallization kinetics in the GeSbTe ternary system has been extensively studied and reported in the literature. OUM cells based on GST 225 that can be programmed (crystallized) to the “SET” state in <20 ns have been reported.

![Figure 19. Schematic temperature–time relationship during programming, in a phase-change rewriteable memory device. Tm and Tx are the amorphization and crystallization temperatures, respectively. The SET and RESET states of the memory correspond to a stored binary 1 or binary 0.](image-url)
For R/W CD’s and DVD’s heat is supplied by use of a laser

For integrated circuits heat is supplied by resistors

### 3.4 ELECTRONIC PROPERTIES OF CRYSTALLINE AND AMORPHOUS GST ALLOYS

Two special electronic properties of chalcogenide amorphous semiconductor alloys are required for the operation of OUM memory—the strong dependence of electrical resistivity on the structural state of the material and the high-field threshold switching phenomenon. Polycrystalline GST 225 alloy has a resistivity of ~25 mΩcm, while resistivity in the vitreous state is three orders of magnitude higher—sufficient to enable good memory read capability. Both structural states of the alloy are semiconductors with comparable energy band gaps. The band gap $E_g$ is 0.7 eV in the amorphous state and 0.5 eV in the poly-crystalline state. The conductivity activation energy $E_a$ is ~0.3 eV for the amorphous state and 0.02 eV for the polycrystalline state. In addition, the amorphous phase exhibits a very low, trap-limited hole mobility of $\sim 2 \times 10^{-5} \text{cm}^2/\text{V s}$, while the polycrystalline phase shows band-type mobility of $\sim 10 \text{cm}^2/\text{V s}$. These large differences come about because of disorder-induced localized electronic states as originally described by Mott and by Cohen, Fritzche, and Ovshinsky (CFO) and later by Kastner, Adler, and Fritzche. When chalcogenide alloy semiconductors are amorphized, electronic energy levels originating in the valence and conduction bands are pulled into the what was originally the empty energy band gap of the crystalline material. As described by Mott–CFO, these new gap states are localized spatially and do not extend throughout the material. Consequently, carriers move through the amorphous material either by hopping among the localized states or by being successively thermally excited to spatially extended band states and then being trapped into localized states. This gives rise to a mobility gap—a range of energy between the valence and conduction bands in which carriers have small, trap-limited mobility. The later work by Kastner, Adler, and Fritzche explained the observation that $E_a \approx E_g/2$ in terms of a large density of special negatively and positively charged traps that also result from structural disorder in amorphous chalcogenide alloys. Kastner et al argued that charged traps (valence alternation pairs) act
like compensating dopant levels in a conventional crystalline semiconductor, effectively forcing the Fermi level to lie near the mid gap between the energy levels of the two types of traps.

In the polycrystalline state, crystal vacancies are proposed to give rise to acceptor-like states that move the Fermi level close to the valence-band edge. This Fermi level position, plus the loss of the disorder produced trapping states, gives rise to the nearly degenerate p-type high conductivity of the polycrystalline state. Thus, the phase-change memory cell uses a reversible change in long-range atomic order (the amorphous-to-crystalline phase change) to modulate both the Fermi level position in the chalcogenide alloy and the carrier mobility to change the cell’s resistance.

![Figure 20. Transmission Electron Microscope images of the two phases of a GeSbTe alloy](image)

![Figure 21. Corresponding Electron Diffraction patterns](image)
The Electron diffraction pattern reveals the following insights about both phases:

<table>
<thead>
<tr>
<th>AMORPHOUS</th>
<th>CRystalline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Range atomic order</td>
<td>Long-range atomic order</td>
</tr>
<tr>
<td>Low free electron density</td>
<td>High free electron density</td>
</tr>
<tr>
<td>High activation energy</td>
<td>Low activation energy</td>
</tr>
<tr>
<td>High resistivity</td>
<td>Low resistivity</td>
</tr>
</tbody>
</table>

Table 1. Comparison of properties of both states

3.5 THRESHOLD SWITCHING

Threshold switching, also first reported by Ovshinsky in 1968, is the other property of chalcogenide alloys exploited in phase-change memory, and it is also a consequence of disorder-induced localized states. In the amorphous phase, the chalcogenide alloy material has a high electrical resistance at low electric fields, as seen above. With increasing voltage, conductivity is initially ohmic, but it begins to grow exponentially when the field exceeds 105V/cm. When a particular threshold voltage, Vth, is exceeded, the material switches rapidly into a highly conductive “dynamic ON state.” The dynamic ON state is maintained so long as a sufficient holding current is passing through the device. This transient high-conductivity state is electronic in origin and does not involve a structural transformation from the amorphous to the low-resistance crystalline state, since it also has been observed in molten chalcogenide semiconductors.

Threshold switching has been explained in terms of electric-field-induced filling of the charged valence alternation pair traps, which alters the recombination kinetics. In the phase-change memory, threshold switching provides a means to deliver the required programming current needed to program a bit in the high-resistance state at low voltage. From a high-resistance (“RESET”) state, an OUM bit is programmed into a low-resistance (“SET”) state by applying programming.
voltage in excess of $V_{th}$, allowing the bit to enter the dynamic ON state. Current then is allowed to flow for a length of time sufficient to ensure crystallization. The device can then be programmed to the RESET state by applying a short, somewhat larger current pulse to a bit in the polycrystalline state. The reset pulse only needs to be of sufficient magnitude and duration to melt the programmed volume of chalcogenide alloy and to have a fast enough falling edge to permit the molten programmed volume of material to cool fast enough to vitrify. The duration of the reset pulse can be short, since the material in the programmed volume can be heated to the melting point in a few nanoseconds.

![I-V characteristics of OUM cell](image)

**Figure 22. I-V characteristics of OUM cell**

The diagram above depicts the Current–voltage characteristics for an Ovonic Unified Memory (OUM) cell element in both the RESET (amorphous, high-resistance) and SET (crystalline, low-resistance) states, showing key device parameters: Read/SET/RESET regimes and SET and RESET states. $V_h$ is the holding voltage, and $V_{th}$ is the switching threshold voltage.
Figure 23. Basic Threshold crossover mechanism

Figure 24. Application of threshold crossing in OUM memory device
As it is desirable for practical applications that the temperature variation should be reduced to as much low as possible

This is achieved by implementing differential alloying schemes as depicted by the Group IV-V-VI Ternary Phase Diagram below:

![Ternary Phase Diagram](image)

**Figure 25. Congruent Crystallization in a Ge\textsubscript{x}-Sb\textsubscript{y}-Te\textsubscript{z} system**

Rapid, reversible changes between the disordered and ordered atomic structure can be made to happen for compositions along the pseudo binary line shown above.

*******
CHAPTER 4

OUM MEMORY STRUCTURE

Figure 26. Block diagram of a OUM cell

Chalcogenide alloys (alloys that contain elements such as Se and Te from Group VI of the Periodic Table) exhibit electronic threshold switching. This phenomenon allows Ge-Sb-Te – based OUM cells to be programmed at low voltage whether they are in the resistive or conductive state.

- The OUM cell is programmed by application of a current pulse at a voltage above the switching threshold.
- The programming pulse drives the memory cell into a high or low resistance state, depending on current magnitude.
- Information stored in the cell is read out by measurement of the cell’s resistance.
- OUM devices are programmed by electrically altering the structure (amorphous or crystalline) of a small volume of chalcogenide alloy.
4.1 Ovonic Information Storage/Retrieval and Display By Structural Transformation

The energy barrier shown above can be reduced by any of the following applied singly or in combination:

- Light
- Heat
- Electric Field (used in the case of OUM)
- Chemical catalyst
- Stress-tension pressure

Also the transformations in Amorphous material due to the application of above mentioned stress factors produce changes in the following:

- Resistance
- Capacitance
- Dielectric constant
- Charge retention
- Index of refraction

Figure 27. Graphical depiction of the Energy states of both the phases
• Surface reflection
• Light absorption, transmission and scattering.
• Differential wetting and sorption
• Others, including magnetic susceptibility

4.2 BASIC MEMORY CELL SCHEMATIC

The following photograph shows us the actual implementation

Figure 28. An actual OUM device and its comparison with block diagram to identify various parts

Figure 29. OUM cell Electrical Schematic
• **RESET State**: High current pulse applied for a short time, which melts the material. Material is subsequently quenched, with no time allowed for crystallization. Amorphous state is obtained with high electrical resistance.

• **SET State**: Medium current applied for longer time. Temperature is below melting point, and material crystallizes, having low electrical resistance.

A memory array is designed using the above cell recursively.

![Figure 30. A basic cell array design for OUM implementation](image)

4.3 **CELL PROGRAMMING R-I CHARACTERISTICS**

![Figure 31. Graph depicting the variation of resistivity as phase changes](image)
4.4 RETENTION AND ENDURANCE

The retention offered by OUM is excellent and can be seen in Figure 32 below showing very small deviation even after $10^{12}$ cycles. This extremely stable nature of its operation makes it the best choice for a long term memory solution.

The demonstrated Endurance of OUM is $10^{13}$, whereas that of conventional FLASH memory is of the order of $10^6$. Excellent data retention has been reported on large arrays.

The Figure below shows the Temperature dependence of programming characteristics.

The high resistance, “Reset,” state shows activated, semiconductor like behavior while the low resistance, “Set,” state shows essentially temperature independent metallic behavior.

Figure 32. Temperature dependence of resistance

The high resistance, “Reset,” state shows activated, semiconductor like behavior while the low resistance, “Set,” state shows essentially temperature independent metallic behavior.
The graph above demonstrates the almost independent nature of the resistance in SET as well as RESET state as number of cycles of operation increase.

4.5 TECHNOLOGY CAPABILITIES

Direct write capability (no erase before write) as well as byte function (no block flash erase) makes it RAM like, easing significantly system implementation

- For flash, changing a byte involves saving the current data, erasing a whole block (>100 mSec) and writing back old data + new byte (total ~1 sec)
- For PCM, changing a byte involves writing the new data: (total < 100 nSec, can be less than 50 nSec with new alloy)
- Demonstrated endurance of $10^{13}$ cycles

With read current > 10 μA, read speed is expected to be comparable to NOR and D RAM
The following graph shows the relationship between the period of operation and the percentage probability of device failure (on Y Axis) accrued over time (on X Axis). The point worth having a glance is that the unit of time is in year, and as it is visible from graph, the device has an excellent life as failure occurrence is negligible.

![Graph showing failure percentage over time](image)

Figure 34. Graph shows the predicted failure %

4.6 **OUM: A MEMORY FOR EVERYBODY**

Even though Phase Change Memory (OUM) is not an universal memory in all respects, it comes close to an universal memory on how it can be used.

- For embedded with the smallest amount of process change, one can get a high cycle EEPROM equivalent at lower cost.
- By using the same process but changing the alloy, one can get a high temperature memory suitable for the most demanding automotive application.
- For dedicated high density memory, one can get a memory cell that is smaller than DRAM and has multi-level cell capability.
- Using high performance alloys, programming speed similar to DRAM are available.
- Using special switched selectors, one can get multi-layer memory than can rival NAND memory in cost.
CHAPTER 5

DISTINCTIVE ADVANTAGES, LIMITATIONS, RISK FACTORS & COMPARITIVE STUDY

5.1 ADVANTAGES

Cost/Bit reduction

- Small active storage medium
- Small cell size – small die size
- Simple manufacturing process – low step count
- Simple planar device structure
- Low voltage – single supply
- Reduced assembly and test costs

Near Ideal Memory Quality

- Non-volatile
- High endurance – >1013 demonstrated
- Long data retention – >10 years
- Static – no refresh overhead penalty
- Random accessible – read and write
- Direct overwrite capability
- Low standby current (<1µA)
- Large dynamic range for data (>40X)
- Actively driven digit-line during read

- Good array efficiency expected
- No memory SER – RAD hard
• No charge loss failure mechanisms
• High switching speed
• Non-destructive read

**Highly Scalable**

• Performance improves with scaling
• Only lithography limited
• Low voltage operation
• Multi-state demonstrated
• 3D multi-layer potential with thin films
• Small storage active medium
• Higher Density, Ease of Integration.

**Logic Process Compatible**

• Late low-temperature processing– Doesn’t compromise P-channel devices
• Adds 2 to 4 mask steps to conventional CMOS logic process with low topography
• Low-voltage operation
• Enables economic merged memory/logic
• Enables realistic System-On-a-Chip (SOC) products: – Logic/Non-volatile memory/Data memory/Linear

**RADIATION IMMUNITY**

One very important advantage offered by OUM over other conventional memory types, is irradiation immunity.

FLASH memory stores the bits in the form of connections between the floating gate and the control gate. FLASH which is otherwise nonvolatile, can loose its data in the presence of radiations which are very common considering Medical and Space applications. The
memory element is exposed to ionizing radiation which have the ability to eject an electron thereby causing disruption of the links which form the basis of memory storage in the flash memory.

OUM however offers the distinctive advantage over all other memory types owing to its Radiation Immunity characteristics. As the bits are not stored in form of charge or links but in form of two definitive phases, Crystalline and amorphous, which are seldom affected by the presence of High energy radiations. Hence OUM can be safely employed in all Medical and Space application.

3-D NATURE (similar to polymer memory)

Another important characteristic of OUM is that like its counterpart Polymer memory, the OUM can also be grown in 3 dimensions thereby offering even high density using a small area.

5.2 RISK FACTORS

- Reset current < min W switch current
- Standard CMOS process integration
- Alloy optimization for robust high temp operation and speed
- Cycle life endurance consistency
- Endurance testing to $10^{14}$–DRAM
- Defect density and failure mechanisms

5.3 TECHNOLOGY CHALLENGES

- Reduction of programming current for lower voltage and lower power operation.
- Increased set/reset resistance and decreased read current/set current margin with scaling → impact on read performance/margin.
- Management of proximity heating with declining cell space → disturb risk.

### 5.4 COMPARITIVE STUDY

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DRAM</th>
<th>SRAM</th>
<th>FLASH</th>
<th>OUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatile/Non-volatile</td>
<td>Volatile</td>
<td>Volatile</td>
<td>Non-Volatile</td>
<td>Non-Volatile</td>
</tr>
<tr>
<td>Endurance</td>
<td>∞/∞</td>
<td>∞/∞</td>
<td>10⁶/∞</td>
<td>10¹³/∞</td>
</tr>
<tr>
<td>Read/Write</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read</td>
<td>Destructive</td>
<td>Partial-</td>
<td>Non-</td>
<td>Non-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Destructive</td>
<td>Destructive</td>
<td>Destructive</td>
</tr>
<tr>
<td>Direct</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Overwrite</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write/Erase(bit/byte wise)</td>
<td>Yes</td>
<td>Yes</td>
<td>Blockwise</td>
<td>Yes</td>
</tr>
<tr>
<td>Read Dynamic Range (Margin)</td>
<td>100-200mV</td>
<td>100-200mV</td>
<td>Delta Current</td>
<td>10X-100X R</td>
</tr>
<tr>
<td>Programming</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write/Erase/Read time</td>
<td>50ns/50ns/50ns</td>
<td>8ns/8ns/8ns</td>
<td>1us/1-100us(block)/60ns</td>
<td>10ns/50ns/20ns</td>
</tr>
<tr>
<td>Transistors</td>
<td>Low Performance</td>
<td>High Performance</td>
<td>High Voltage</td>
<td>High Performance</td>
</tr>
<tr>
<td>CMOS Logic Compatibility</td>
<td>Bad</td>
<td>Good</td>
<td>OK, but high voltage required</td>
<td>Good</td>
</tr>
<tr>
<td>Scalability Limits</td>
<td>Capacitor</td>
<td>6T</td>
<td>Tunnel Oxide</td>
<td>Litho</td>
</tr>
<tr>
<td>Multi-bit Storage</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>3-D Potential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Susceptibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Cost per bit</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
</tbody>
</table>

Table 2. Comparative Study of different Memory technologies
SUMMARY

Non volatile OUM with fast read and write speeds, high endurance, low voltage/low energy operation, ease of integration and competitive cost structure is suitable for Ultra High density Standalone and embedded memory applications. These attributes make OUM an attractive alternative and potentially competitive with volatile memory technologies.

Apart from being highly scalable, this technology also offers the potential of easy addition of nonvolatile memory to a standard CMOS process, and is also low cost.

It can hence be concluded that OUM offers a very promising future for Non-Volatile Memory and other allied devices.
REFERENCES

- White paper- Phase Change Memory Technology – Numonyx.
- J. Maimon, E. Spall, R. Quinn, S Schnur, “Chalcogenide-Based Non-Volatile Memory Technology” NVMTS (Nonvolatile Memory Technology Symposium, 2000).
LIST OF ACRONYMS USED

- OUM- Ovonics Unified Memory
- EPROM- Erasable Programmable Read only Memory
- EEPROM- Electronically Erasable Programmable Read only memory
- DRAM- Dynamic Random Access Memory
- SRAM- Static Random Access memory
- SDRAM- Synchronous Dynamic Random Access Memory
- CMOS- Complementary Metal Oxide Semiconductor
- TTL- Transistor Transistor logic
- PCM- Phase Change Memory
- FRAM- Ferroelectric Random Access Memory
- MRAM- Magneto Resistive Random Access Memory
- CRAM- Chalcogenide Random Access Memory
- G-S-T – Germanium- Antimony – Tellurium
- ETOX- EPROM Tunnel Oxide
- SER- Soft Error Rate