



OBJECTIVE ANALYSIS

Semiconductor Market Research

WHY WE NEED A NEW MEMORY TECHNOLOGY

White Paper

Any investor wishing to profit from a new memory technology must first understand the dynamics of the memory market to appreciate why any new technology would be likely to succeed or fail in tomorrow's memory business. This white paper describes the "Hows" and "Whys" of today's memory market, explains why memory chip makers are interested in alternative new technologies, and attempts to rate the prevailing new technologies against each other.

Background: Basic Memory Market Facts

Today there are five main memory types in production: DRAM, NAND flash, NOR flash, SRAM, and EEPROM. Although other types exist, these five make up over 95% of all memories sold.

No one memory solves all the needs of the system designer. As Table 1 illustrates, each of these memories has a key attribute, and each has a key weakness. The volatile memories, DRAM and SRAM, offer high speed but lose their contents when powered down, so they must be reloaded every time power is restored. The nonvolatile technologies: NAND and NOR flash and EEPROM, do not suffer from this data loss but take over 1,000 times longer to store new data. The market continues to await the development of a single technology that can solve both problems – a nonvolatile memory with a fast write speed.

Table 1. Comparison of Today's Memory Technologies

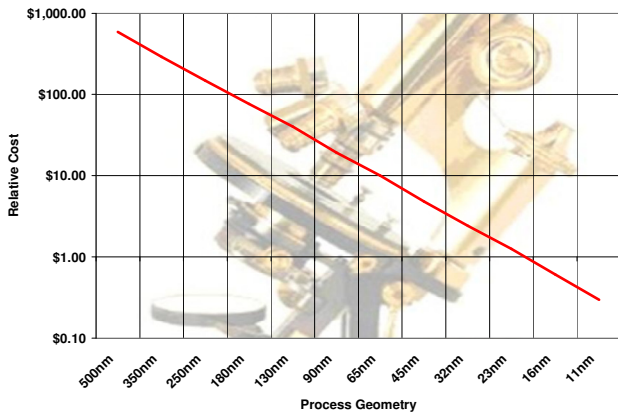
Technology	Volatility	Read Speed	Write Speed	Price/GB
DRAM	Volatile	300MB/s	300MB/s	\$20
SRAM	Volatile	400MB/s	400MB/s	\$2,000
NAND	Nonvolatile	20MB/s	3MB/s	\$5
NOR	Nonvolatile	50MB/s	0.15MB/s	\$50
EEPROM	Nonvolatile	2MB/s	0.006MB/s	\$10,000

Source: Objective Analysis January 2008.

Within each of these technologies there is very little product differentiation. Chips from Samsung are similar to those of Micron, Elpida, Infineon, or many others. This means that competition is price-based rather than feature-based. This in turn drives memory makers to compete to produce the lowest manufacturing cost and fuels the chip-shrinking race which has characterized the semiconductor business for the past four decades.

The cost of a semiconductor chip is determined by two factors: The cost to produce a silicon wafer and the number of chips that can be squeezed onto that wafer. Over history semiconductor manufacturers have migrated to ever-shrinking processes in order to drive down the costs out of their chips. This is conceptually illustrated in Figure 1. Assuming the manufacturing cost of a silicon wafer is relatively constant (which is generally true at roughly \$1,000 for a 200mm wafer and \$1,600 for a 300mm wafer), the cost of a chip declines in proportion to the process used to make it. Figure 1 shows that as the process shrinks (across the bottom axis of the chart) the cost of the chip should decrease in proportion (the vertical axis).

Figure 1. Chip Cost Depends on Process Geometry

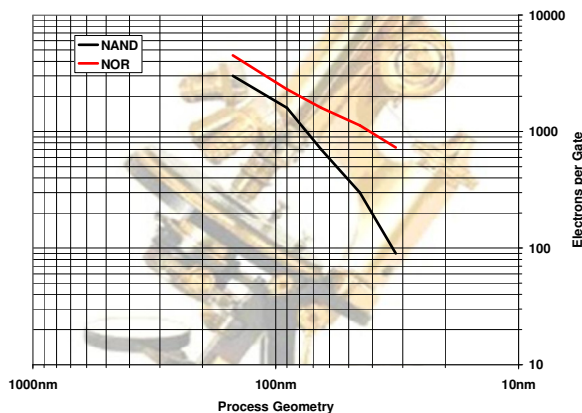


Source: Objective Analysis January 2008.

This chart has to be used with care since it leads to the mistaken assumption that costs will continue to decline as they always have in the past. A prime example of this is the two flash memory technologies, both NAND and NOR.

The flash memory industry accepts that there is a limit to how small a flash chip can be shrunk. This is called the “Scaling Limit”, and is determined by the number of electrons that can be stored on a flash gate (using today’s flash technology) at various processes. Figure 2 shows that the number of electrons becomes so small at finer processes that the job of determining whether or not these electrons are actually there becomes extraordinarily difficult.

Figure 2. As Flash Chips Shrink Electrons per Gate Drop



Sources: NOR figures: Intel Corp., NAND figures: Samsung Electronics

Once a flash chip shrinks beyond a certain point the size of the chip no longer reduces in proportion to the process and the die size (and cost) of the flash chip remains the same even if the process continues to diminish.

As recently as 2003 Intel publicly announced that the limit for flash was going to be 65nm. The company carefully mapped out their plans to use an alternative technology for processes smaller than 65nm. By 2004 Intel encountered a breakthrough that allowed them to revise this stance, and they projected devices down to the 35nm node, anticipating further shrinks below that level. By the end of 2007 Toshiba researchers showed a way to manufacture basic flash technology at the 11nm geometry, showing that flash technology could be with us for another few years.

Even though breakthroughs have been made, NAND and NOR makers continue to research alternative technologies, since they realize that the scaling limit of flash technology, although postponed, is inevitable. Today the leading contender to replace flash is PRAM or OUM technology.

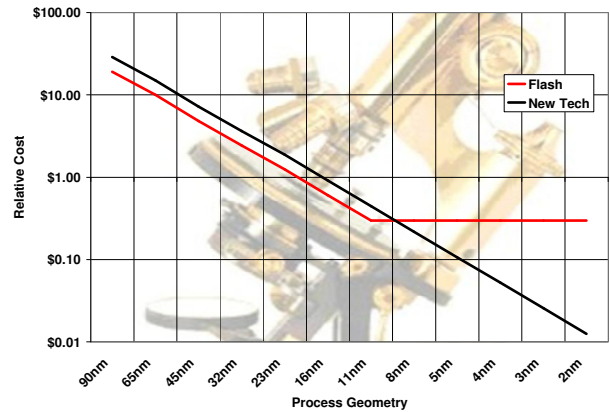
All of these new technologies involve the addition of a new material on top of the silicon foundation of the chip. PRAM or OUM requires the addition of an amorphous layer called chalcogenide glass. MRAM requires a ferromagnetic layer. FRAM, or ferroelectric memory, despite its name, does not use a ferroelectric layer, but employs a material called a Perovskite crystal. All of these materials react badly with the underlying silicon substrate, so they all present the tricky problem of using a relatively inert boundary layer to isolate the two materials. To assure that contamination will not shut down the fab, new manufacturing flows will have to be developed to allow such materials to be used in a cost-

effective manner. To date such flows do not exist.

Before flash reaches the end of its life it will be a preferred technology for a number of good reasons. First of all, the cost of processing a silicon wafer will be lower than the cost of processing any wafer with a new material. The reason is simple: Over the history of the silicon chip a cumulative \$200 billion or more of R&D funds have been spent to understand this material. Engineers know more about how to make silicon do their bidding than they do of any other semiconductor material. This allows them to push the cost of processing a pure silicon wafer significantly below the cost of processing a wafer with a new material. Next we have the economies of scale. Any technology a company manufactures in volume will be much less costly than a new technology until that new technology reaches production volume. It can be very expensive to adopt a new technology before it becomes absolutely necessary. Finally, the costs of processing a new technology can be higher than that of a standard silicon process. All of the alternative technologies mentioned above will require special manufacturing flows to assure that the underlying silicon avoids contamination, and these flows will be more expensive than those currently used in a standard flash processes, at least for the near term.

From an economics standpoint how will an alternative technology take over the nonvolatile memory market? Figure 3 expands upon the chart of Figure 1 to illustrate relative costs of flash vs. an alternative technology. For this particular chart we have arbitrarily set the wafer cost of the alternative to be 50% higher than that of standard flash. We also assume that the scaling limit for flash technology is 10nm in keeping with the December 2007 Toshiba announcement mentioned above.

Figure 3. When Flash Scaling Stops Other Technologies Can Succeed



Source: Objective Analysis January 2008.

We can see that flash will lose its cost edge over the alternative technology once flash loses its ability to scale. The new technology, even with a higher wafer cost, will be able to cost-reduce below flash's costs to become the preferred technology for future generations.

The timing of this crossover depends on two factors: The actual scaling limit of a pure silicon flash technology, and the wafer processing cost difference between a pure silicon and silicon with the alternative technology.

Alternative Technologies

An interesting point about the new technologies being evaluated today is that they all offer better features than to the current technologies outlined in Table 1. Where current devices are either volatile or nonvolatile, or have fast or slow read and write speeds, or are more or less costly than the others, or even suffer from wear-out mechanisms, the new technologies all offer nonvolatility, fast read and write (at very low power levels), and are anticipated to have lower costs when they reach production volume. This last point is the main reason that none of these technologies have displaced current technologies even though some have been trying to for over 20 years!

Most of these technologies uses a different storage approach than flash employs, that of storing a charge to determine a bit's state (a 1 or a 0). This moves these technologies past the concerns expressed in Figure 2.

Let's explore the leading technologies that are being evaluated today.

PRAM or OUM

The PRAM (phase-change RAM) or OUM (Ovonic unified memory) is a chalcogenide glass memory type. A layer of this glass is deposited on top of a standard silicon logic chip, and carefully-controlled heat is used to turn the glass into a crystalline conductor or an amorphous insulator. Transistors sense the conductance of this area to tell whether a bit is a 1 or a 0.

Intel published a paper on their PRAM technology efforts in 1970. This technology has taken a very long time to reach a sampling stage which it did in 2007.

Since the technology does not use a stored charge to determine the bit's state, there is no question of the number of electrons stored on a bit, allowing the bit to be shrunk considerably smaller than is anticipated for standard flash today. This is important, since the PRAM bit currently in use is $15f^2$ compared to flash's $5f^2$, leading to a much larger device at the same process geometry. (We have heard that future PRAMs may be able to shrink to $5.5f^2$.) Between the larger bit size and the higher cost of processing a PRAM wafer, the technology's advantage is not likely to allow it to take over the NOR or NAND market until flash's process migration has been stalled for two process nodes.

PRAM suffers from write-induced wear like NAND and NOR flash, but proponents explain that PRAMs can endure

significantly more writes than can flash technologies.

PRAM is now being sampled to select customers by Intel, STMicroelectronics, and Samsung, but none of these companies has outlined any road map to production volume.

Ferroelectrics or FRAM

FRAM is another long-lived alternative memory technologies. This technology, first announced in the late 1980s, relies on a layer of Perovskite crystals being deposited on top of a standard silicon substrate. At its announcement FRAM pioneer Ramtron disclosed that the simple addition of a ferroelectric layer would convert any standard DRAM design to a nonvolatile memory with the speed and low power of an SRAM at the cost of a DRAM. At that time flash memory didn't exist and DRAM was the least costly memory available. In that context FRAM would clearly be capable of replacing all memory types then available.

FRAM works by a unique characteristic of Perovskite crystals. When a current is passed through the crystal one atom moves from one side of the crystal to the other. That atom can be moved back by reversing the current. These crystals are read by trying to move the atom – if the atom moves a high current is consumed, but if the atom is already in the position you are trying to move it to, the consumed current is significantly lower.

Perovskite crystals contain elements that can interfere with silicon transistors, so a barrier layer is generally used to isolate this layer from the silicon below. This adds to the device's cost by increasing the cost of processing the wafer.

Like PRAM, FRAM has a wear-out problem, but designers have found ways to make the FRAM's wear less of a factor by increasing the cell size. This

works against the technology's cost structure, and has pushed it behind PRAM from that perspective.

Today FRAMs are in volume production at very low densities where they have found favor in certain niche applications. It is not clear that this technology will ever find a way to catch up with other alternative memory types, especially since its cell size is very large, at approximately 30f^2 and may only shrink as small as 12f^2 .

MRAM

MRAM, the magnetic RAM, uses the magnetic principals employed for over a century in magnetic recording tape and hard disk drives: If a portion of a magnetic layer points north then it signifies a 1, and if it points south it is a 0.

It was not possible to harness magnetic technology on a silicon chip until the discovery of giant magnetoresistance (GMR) in 1988. Through GMR the direction of a magnetic field can be sensed without the magnetic media having to move across the sensing area of a reading head. This allows the motionless silicon chip to read the magnetic layer.

MRAM is a relatively new technology so it has not undergone the succession of refinements that PRAM and FRAM have enjoyed. Because of this there are still difficulties that confront MRAM. The most recent of these was the finding that current densities would become impossibly high with shrinking process geometries. This problem has been addressed first by the employment of cladding layers to confine weaker fields, then by the development of "spintronics" in which a more elegant device structure that takes advantage of previously-untapped quantum effects inherent in GMR devices. This has not dissuaded numerous firms from investing in this technology, including Freescale, Grandis,

Hitachi, Honeywell, NEC, NVE, Renesas, Sony, and Toshiba.

One advantage that MRAM has is that production volumes of this material have been made in the manufacture of certain read/write heads for hard disk drives. This will help developers ramp this technology to production with potentially less fuss than will occur with PRAM and FRAM.

MRAM has found a home in the Military/Aerospace market since magnetic storage has proven to be insensitive to the increased radiation levels typical of high altitudes, where the atmosphere has not attenuated the flow of atomic particles from our sun's solar wind. These particles tend to tear electrons out of the floating gate used in flash memories, and as Figure 2 shows, these electrons are rather rare.

A clear disadvantage of MRAM is that its cell size is the largest of any of the alternative technologies at 35f^2 . Since today's flash has a cell size of only 5f^2 MRAM will probably have to scale three or more process steps beyond flash's scaling limit to become cost-competitive with any flash technology. This may be accelerated with the adoption of a new technology, STT-MRAM, with a theoretical minimum cell size of 6f^2 .

Carbon Nanotubes

Carbon nanotubes are just now moving from the realm of scientific curiosity to a manufacturable technology. With that comes a novel approach to making semiconductor memories out of carbon nanotubes being spearheaded by a company called Nantero.

In this technology a shallow trench is spanned by one or more nanotubes. These tubes are attracted to the bottom of the trench by an electric field. When the field is removed van der Waals forces hold these tubes to the bottom un-

til another field is applied forcing them apart. The cell is read by measuring whether the tube shorts to the bottom of the trench.

In theory nanotube memories can reach a density similar to that of DRAM, which has a cell size of $6f^2$ - That is, each bit is six times the square of the chip's process geometry.

Although it shows a lot of potential, this technology seems a little farther from production than any of the three above. There is some question of whether it can be brought into a volume production quickly enough to contend in the race to displace flash. This, of course, depends in turn on how long flash can forestall its inevitable scaling limit.

Molecular Memory

Another very new technology is the recently-announced "molecular" memory. A molecular memory consists of a special molecule that attaches itself to etched silicon columns in the form of a sphere at the top. The technology is similar to flash since it stores an electrical charge, but the shape of the spheres at the tops of columns allows significantly more charge to be stored in the same silicon area, supporting further scaling than is possible with the flat structures currently used in pure silicon.

As the technology is new, it faces more hurdles on its way to production than do the more mature advanced technologies. The company is well funded, but to date it is very far removed from production-volume manufacturing.

Comparing the Technologies

In Table 2 we compare some of the salient attributes of the technologies outlined above.

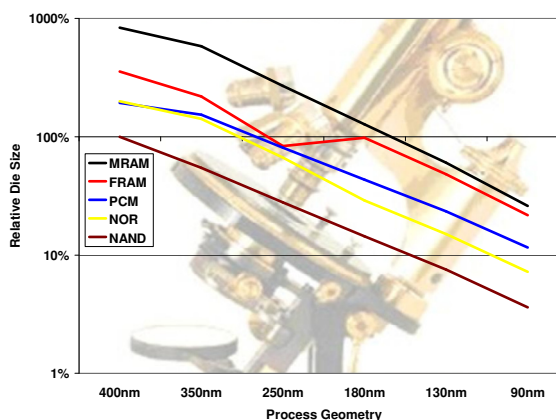
Table 2. Attributes of Various New Memory Technologies

Technology	Strengths	Weaknesses	Cell Size	Status
PRAM/ OUM	Scalable	Large die, Endurance	$16f^2$	Limited samples
FRAM	High speed	Large die, Endurance	$30f^2$	Low volume production
MRAM	High speed	Large die, Scaling difficulties	$35f^2$	Some military production
Nano-tubes	Scalable	Very new	$\sim 6f^2$	Research
Molecular	High density	Charge-based	TBD	Research

Source: Objective Analysis January 2008.

Figure 4 estimates relative die sizes for each technology based upon a history of the process geometry and the cell size available for each memory type.

Figure 4. Cost Comparison: NAND and NOR vs. Leading Alternative Memories



Source: Objective Analysis January 2008.

The numbers that drive this chart have been derived from actual prototypes or production chips that were presented at memory chip technology conferences. The chart compares the costs of five key technologies at a range of processes to the cost of a NAND chip made using a 400nm process.

This chart clarifies why NAND and NOR should continue to be the economical alternative until the point is reached where the technology no longer scales with process.

Readers will note that the relative cost of FRAM jumps at the 180nm node. This is the point at which the technology con-

verted from a one-transistor cell with endurance issues to a two-transistor cell with unlimited endurance.

From this chart alone PCM looks to be a lower-cost contender to MRAM and FRAM.

Conclusion

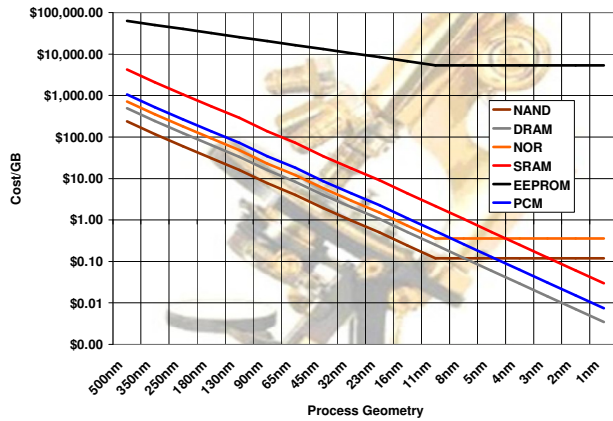
There is no disagreement that a new memory technology will become necessary in the foreseeable future, although there is little agreement as to when that juncture will be reached. When it is finally reached then there are many alternative memory technologies vying to take over where NAND and NOR are unable to go.

Of these alternatives only a few stand out. PRAM has advantages of significant backing and a small die size. MRAM and FRAM are at a disadvantage to this technology.

How will this takeover fall into place? Certain other memory technologies like SRAM and EEPROM stand threatened before NOR yields its market to an alternative technology, since alternative memory technologies offer similar attributes to these parts at prices that promise to be significantly lower once production volumes are reached.

We illustrate this in Figure 5. This chart shows costs of various memory technologies across a wide range of processes. We have used PRAM as the one alternative technology because it has the lowest cost structure of any currently-sampling alternative memory technology. We should explain that we once again assume that flash stops scaling at 10nm but that SRAM and DRAM continue to scale. If this were not to occur, then these technologies would be threatened earlier than is shown in the chart.

Figure 5. Cost of Various Memories as a Function of Process



Source: Objective Analysis January 2008.

PRAM, the blue line, is clearly less expensive than both EEPROM and SRAM, and could threaten these markets today. At the 8nm process PRAM becomes as cost effective as NOR, dropping below NOR's cost for all future generations. This implies that NOR would be the next memory type threatened by such an alternative technology. NAND undergoes a similar fate at 5nm, yet DRAM maintains a slight cost advantage over PRAM for the long term.

Based on this last graph it appears likely that any new technology will be able to overtake the EEPROM and SRAM markets in preparation for the day that NOR flash hits its scaling limit and succumbs to an alternative memory technology.

Jim Handy, January 2008