White Paper

NRAM as Storage Class Memory

Memory: The Next Big Opportunity for Differentiated Mobile Devices

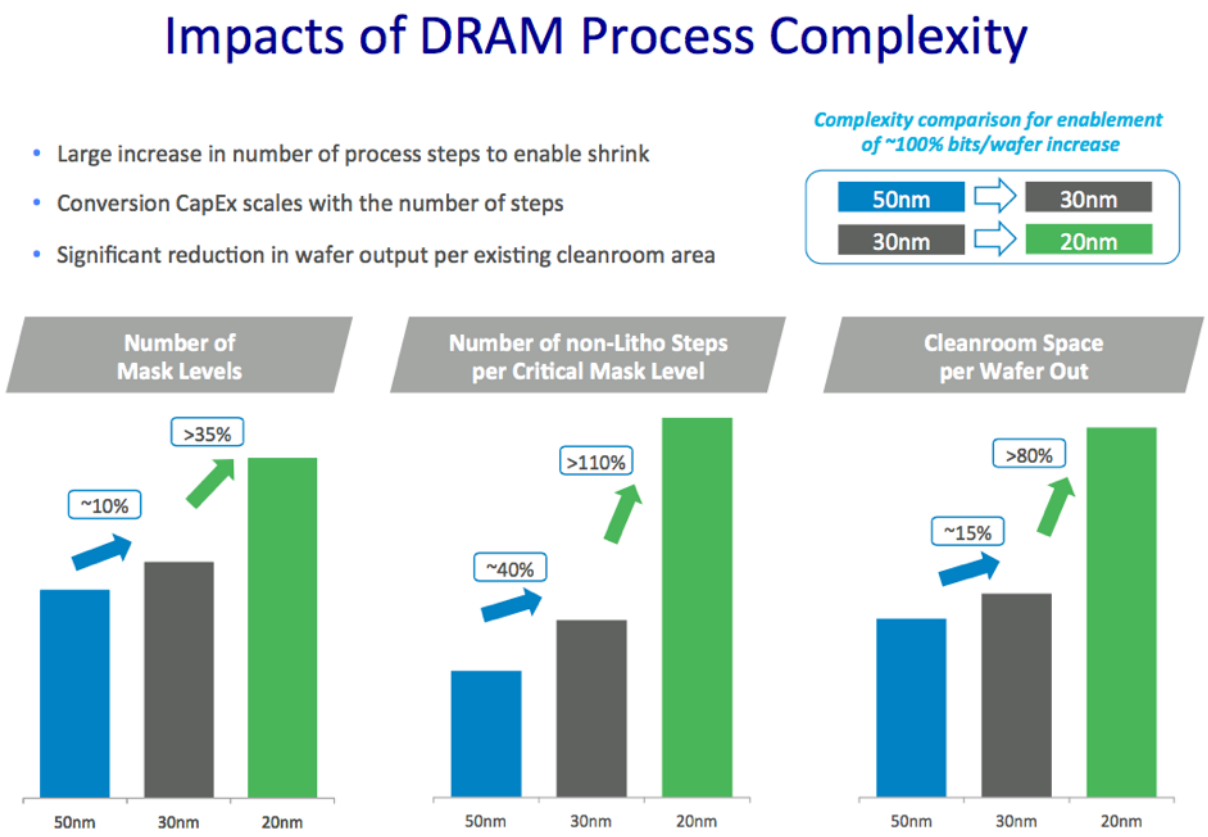
By Ed Doller

Abstract: Storage class memory is a term first used by IBM in the late 2000s. It has become synonymous with the majority of next-generation memory technologies given the fact that they all fall somewhere between NAND flash memory and Dynamic Random Access Memory (DRAM) in just about every attribute ranging from cost to performance. Significant investments are underway already to develop storage class memory, due to the enormous customer demand and the fact that both DRAM and NAND are falling off Moore's law. This white paper will discuss the compelling value proposition of storage class memory in mobile applications (i.e., those that run on battery) and discuss why NRAM® – based on Nantero's carbon nanotube technology (CNT) - is well suited to provide such value.

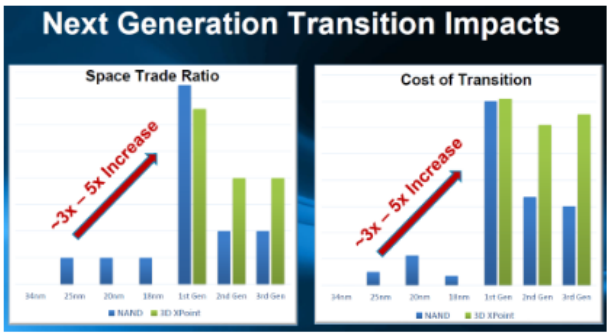
About the Author: Mr. Doller is a semiconductor memory & storage industry expert with over 31 years experience. He was most recently a corporate officer at Micron Technology where he held the role of VP & Chief Strategist of the NAND Solutions Group. Prior to that, he served with Micron as VP & GM Enterprise Storage and also as VP & Chief Memory Systems Architect. Mr. Doller joined Micron in 2010 via the Numonyx acquisition where he served as VP & Chief Technology Officer after its formation in 2008. Before Numonyx, Mr. Doller spent 15 years at Intel in the Flash memory group where he was appointed its Chief Technology Officer in 2004. Prior to Intel, he spent 9 years at IBM in East Fishkill, N.Y. and held several key positions all in advanced semiconductor memories. Mr. Doller earned a Bachelor of Science degree in computer engineering from Purdue University. He holds multiple patents, is a co-author of the IEEE floating gate standard, and is a frequent keynote speaker at memory conferences. He is a member of Nantero's Advisory Board.

Introduction

It is well known that the entire memory industry is investing heavily in next-generation memory technologies, as it now is very clear that customer demand is enormous and still unmet. On the technical side, the reason for the increase in R&D is that despite the market hype otherwise, both DRAM and NAND are falling off Moore's law, which states that the number of transistors in a dense integrated circuit double approximately every two years. In memory, this implies a 2X capacity at the same cost every two years. Furthermore, while engineers figure out how to get the next-generation memory technology to be 40-50% smaller, the finance teams are looking at the costs associated with the new technologies, such as 3D NAND, and are very concerned with their ROIs given the high cost of technology conversions. It should be noted that while NAND has a kink in the Moore's law curve given the 2D to 3D transition, once transitioned, the technology could be back on Moore's law for a while. Below are two excerpts showing this for both DRAM and NAND, both presented by Micron Technology in 2015.



Source: Micron Analyst Presentation, February '15



Source: EETimes , January '16

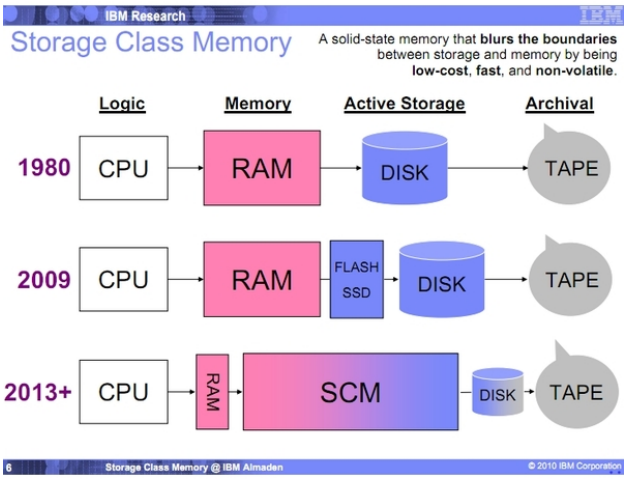
If we look at the two main memory technologies in the industry today, they are dynamic random access memory (DRAM) and NAND flash memory. Roughly speaking, they represent a $45B and $35B market respectively. DRAM is connected directly to a CPU and NAND is typically "abstracted" behind a controller that in turn is connected to the CPU. Because NAND is an abstracted memory technology, it's VERY simple to compete with architecturally. That is, all one needs to do is develop a technology that is lower cost than NAND, abstract it behind a controller and compete. Much easier said than done obviously, but easy from a hierarchy standpoint .

If a new technology is NOT lower cost than NAND, then the other memory technology it must compete with is DRAM. While DRAM is a much easier target from a cost standpoint, architecturally the bar is high, as the memory technology must "talk" directly to the CPU. If one can develop a "drop-in DRAM-compatible memory" that is lower cost, one can compete directly with DRAM as no changes to the CPU will be required. Again, much easier said than done as the new memory must be fast enough and low power enough to be compatible with the DDR standards. Because NAND is a highly-cost-optimized, high volume technology, it is extremely difficult, if not impossible, for any new technology to be lower cost than NAND right out of the gate. To become cost-competitive with NAND requires selling substantial volumes to get down the cost curve, but customers will not buy if it is more expensive, leading to a deadly catch-22 if the new technology is trying to compete with NAND right away. Therefore, any new memory technology must initially compete with DRAM in order to succeed. Only one with the right balance of cost, performance, and reliability can ultimately win out.

What is rarely discussed is the value proposition of a new memory technology and where it will likely fit in the memory hierarchy. This white paper will take a view of the compelling value proposition for storage class memory replacing DRAM + NAND, specifically in mobile applications.

Memory Hierarchy

In the late 2000s, IBM coined the term "storage class memory" (SCM). It was used to generically describe a class of memory that sat, hierarchically, between DRAM and storage. Bellow is a diagram showing their depiction.



As you can see, in 1980 memory hierarchy consisted of RAM and disk. DRAM was volatile and called "working memory" and storage was non-volatile. With the introduction of solid-state drives (SSDs) in the late 2000s, the depiction shows SSDs slipping in between RAM and Disk. This location was based on price and performance and is an accurate depiction from a cost / performance / price standpoint. The $/GB of SSDs are an order of magnitude lower than RAM while being an order of magnitude higher than traditional storage, namely HDDs.

The 2013 view shows SCM sitting between RAM and disk. While the depiction is admirable and correct from a cost & performance standpoint, there are several issues. First, the advent of SCM does not negate the need for SSD unless the SCM technology is lower cost than NAND flash memory, which is highly unlikely for quite a while for the reasons discussed above. Second, if it isn't lower cost than SSD, in order to get value it cannot serve as mass storage, it needs to either displace some DRAM or directly augment DRAM as another level of cache. The amount of DRAM it can displace will be a function of the attributes highlighted above such as cost, power, performance and endurance. In order to displace DRAM, it must therefore have a direct connection to the CPU.

Historical View of Mobile Architecture

Despite the SCM concept being introduced by IBM in the mid 2000s, one doesn't need to go back that far to get a view of "SCM" in action. Let's take a historical look at mobile architectures, focusing mainly on the cellular phone in the 1990s to understand the value proposition of SCM. During this time frame, the predominant memory technology used in cellular phones was NOR. At its peak, it was an ~$8B market and could easily fit the definition of a storage class memory. That is, it was directly connected to the CPU, lower cost per bit than DRAM + NAND – at the desired capacities, and was able to execute code and store data with non-volatile characteristics. The key to any SCM is the requirement for it to be able to execute code AND store information with some level of permanence. NOR did all of these well with its architecture coined "XIP" or execute in place. This stems from the ability for the storage device, the NOR, to store the code and data AND execute from it directly to the CPU. However, an alternative architecture ultimately won out, which was called store and download or "SND". This architecture consisted of DRAM and NAND whereby the code and data were stored in NAND, copied to DRAM upon power up, and executed to the CPU out of DRAM. The reason NOR ultimately lost to DRAM + NAND simply because it could no longer scale which translated into higher costs and lower capacities than DRAM, so for this reason the SND architecture prevailed.

Mobile compute platforms, such as laptops, have always utilized SND architecture. This is primarily due to the increased performance requirements placed on the DRAM and the amount of DRAM used. While some have argued for XIP in mobile laptops, the technologies, architectures, and software have so far failed to align. The goal is that a next-generation memory will finally provide the performance, cost point, and value proposition to make the architectural conversion to XIP a reality, providing the benefits to the end-user in terms of increased speed and reduced power consumption that XIP has always promised.

Requirements for Next-Generation Memory

As with any memory technology, each has its own unique value set. Below is a chart that summarizes the incumbent technology values:

Technology	Read	Write	Power	Endurance	Cost
DRAM	Fast	Fast	High	High	High
NAND	Slow	Slow	Low	Low	Low

As you can see these two technologies are almost the opposite of each other and paint the two ends of the spectrum. All next-generation technologies fit somewhere in the middle.

In order to understand the potential value in mobile products, these five attributes need to be examined.

Performance: NOR had asymmetrical read write characteristics. That is, most NOR could be read in sub 10ons but had write times in the micro-second range and required erase before re-write. The reality is, writes are far less important than reads - given the direct connection to the CPU - and can be managed. Because NOR could act as a code device and a data device, it was very easy to size the amount of NOR required. Performance was high with almost instant on qualities. With the SND architecture, performance was lower simply because the code and data needed to be copied from NAND to RAM prior to execution. Furthermore, performance was directly related to the amount of RAM in the system. If the goal was to maximize performance, more RAM rather than less would be required. The downside impact of "more RAM" was cost and power (more on that in a bit). Anyone that has used a PC, which employs a SND architecture, knows that increasing RAM generally results in higher performance as there's less actual "store and download" going on because it's already resident in DRAM. In a mobile laptop, read performance requirements are more critical than in a mobile handset. Today, high end mobile phones are shipping with LPDDR DRAM and have DDR rates in the 1600 MT/s range vs. 2133 MT/s for laptops utilizing DDR3 DRAM. It wasn't that long ago when mobile phones shipped with LPDDR1 which has DDR rates of 533 MT/s. The architectures and performance requirements for phones and laptops seem to be converging. XIP architecture provides much better performance than SND due to its direct connection to the CPU and the lack of need to fetch information from another memory device as in the case of SND. The impact of SCM to the end user is two-fold: 1) "instant on" performance and 2) code not consuming your storage GB's.

Key message #1: Having a next-generation memory technology that provides DRAM-like performance is key.

Power: Non-volatile memories will draw less power than DRAM while in stand-by simply because a DRAM needs to have its contents refreshed. Despite advances in LPDDR, any non-volatile memory will burn much less power in stand-by. Power is critical when talking about mobile devices. The iphone 4 used 512MB of LPDDR DRAM while its laptop counterpart used 4-8GB on average. That's an 8-16X increase. Because DRAM is a large consumer of power in stand-by, one could easily see that the impact of a low power executable next-generation memory would be bigger in a laptop given this ratio. The iphone 6s is shipping with 2GB of LPDDR DRAM. This is an 8X increase over the iphone 4 and converging on the laptop requirements. Having a non-volatile executable memory will increase battery life by a noticeable amount—exactly how much is highly dependent on the amount of memory in the system as well as other power hungry components and usage models. In a patent application by Intel (Patent 20130283079), they stated that replacing DRAM in a laptop with an SCM has

the potential to increase standby power **from days to a few hundred days**. Thus, significant power saving can be achieved.

Key message #2: Having a non-volatile next-generation technology has the ability to lower stand-by power dramatically.

Endurance: As noted above, DRAM has "infinite" endurance while NAND is fairly limited. 2D NAND is in the 1-3K write cycle range while 3D promises to deliver 40-50K write cycles. Having a next-generation memory technology that is infinite, like DRAM, is highly desired. While the desire is to have infinite endurance, simply because not thinking about it is always easier, lower endurance levels can certainly be managed around for many applications. The closer a technology gets to infinite, the easier that management becomes.

Key Message #3: While having a next-generation technology that has infinite write capability is desired, management techniques can be employed to deal with <infinite, and those management techniques are often what differentiates one systems provider or device manufacturer from another.

Cost: As stated, if there's a "drop-in" compatible technology that's lower cost than DRAM, it will win. In the height of NOR, the reality was NOR was lower cost than NAND + DRAM. You may be asking "why is it a NOR vs. NAND + DRAM" and the answer is quite simple. If the requirement for a phone was 2Gb of code, then a SND architecture required 2Gb of non-volatile storage to store the code (e.g., the NAND) and 2Gb of RAM to execute the code. Paging architectures have been employed to minimize the RAM for the reasons stated above but nevertheless NOR competed quite nicely. It took a tremendous amount of education in the industry to make the point that NOR NEVER competed with NAND alone, it competed with DRAM + NAND. This is an extremely important concept, as it will become center stage once again as the industry moves to next-generation memory solutions and influences the choice of solution.

Key Message #4: Having a next-generation technology that is lower cost than DRAM + non-volatile storage is imperative.

Despite all of this, given the past conversion of XIP to SND in cellular handsets, it's easy to see that cost is likely a major driving factor and cannot be ignored. On an equal cost playing field, performance, power and ease of conversion will be highly valued.

NRAM®

Let's examine the key attributes of the technology given the attributes discussed above.

Technology	Read	Write	Power	Endurance	Cost
DRAM	Fast	Fast	High	High	High
NAND	Slow	Slow	Low	Low	Low
NRAM	Fast	Fast	Low	High	<DRAM

Performance: With a DDR interface, latency of ~50ns, and a bandwidth of 2133MT/s, NRAM is by far the highest performance next-generation memory technology today. No other technology proposed offers this level of DRAM-like performance.

Power: It is nonvolatile so refresh is not required like DRAM hence standby power is much, much lower.

Endurance: While it is believed to fundamentally have infinite write capability, like DRAM, its first instantiation will offer 10^9 which is more than enough for many demanding applications.

Cost: Less than DRAM.

Conclusion

The time for thinking about next-generation memory is here today, as they are emerging and users of memory need to start planning how they will incorporate it into their architectures. Changes to both architecture and software will be required to take full advantage of a new memory technology. The balance of cost, power, reliability, and ease of use will need to be balanced between the engineering team and the procurement teams to ensure the right choices are made. For end consumers, the expectation of higher performance and MUCH less stand-by power on their mobile devices should be realized, as well as true instant-on for laptops. This will create a competitive advantage for OEMs as they can now use memory technologies to differentiate their products. Nantero's NRAM has the highest balance of performance, power, and reliability in the industry.