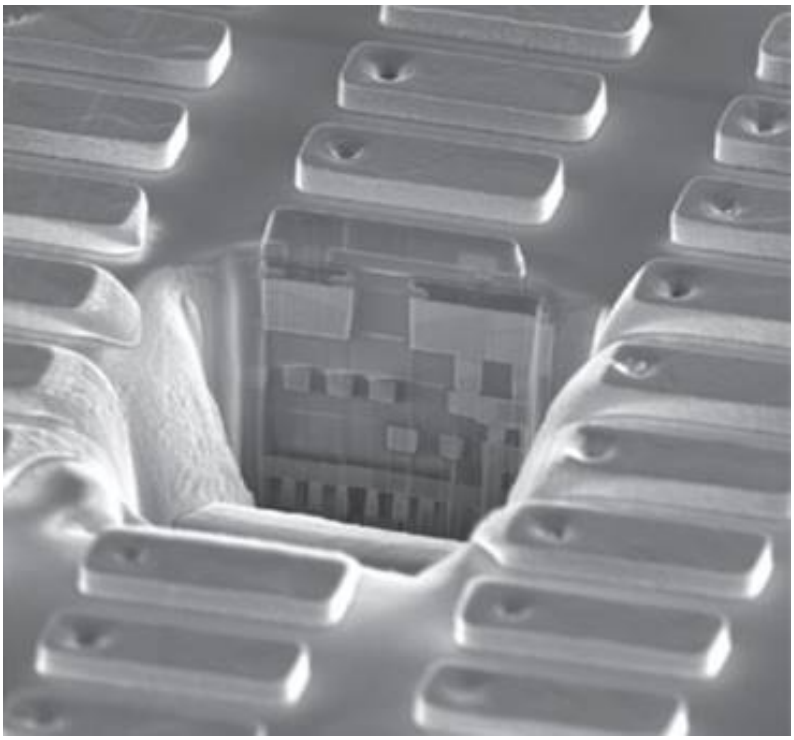
A DEEP LOOK AT NEW MEMORIES



COUGHLIN ASSOCIATES & OBJECTIVE ANALYSIS

October 2024

Cover Art from Leti, used with permission

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

Scaling limits have become very real for flash memory (NAND and NOR), DRAM, SRAM, and other popular memory technologies are facing technology limits to their continued improvement. With this facing them, many researchers are devoting significant resources to develop new memory technologies. The most promising of these new technologies are nonvolatile memories, which can be used for long-term storage or to provide a memory that retains information when powered down. Nonvolatile memories offer energy savings that are appealing to designers of battery and ambient powered devices and also for energy savings, and performance improvements in data centers.

This report covers PCM, ReRAM, FRAM, Toggle MRAM, and STT MRAM, as well as a variety of less mainstream technologies such as III-V devices and carbon nanotubes. At the moment, resistive RAM (ReRAM) technologies are viewed as a potential replacement for flash memory, but future NAND flash memory is anticipated to support several more layer-count increases and other approaches for higher storage density, to push costs much lower than today, allowing many years to pass before a replacement is required. Thus, a transition from NAND flash to some other technology is not expected to occur until the next decade, at the earliest.

In this market, the economies of scale reign supreme. We saw that with Intel's 2022 termination of its Optane 3D XPoint campaign and had seen the same phenomenon play out before with NAND flash, whose high volume drove its costs below that of DRAM. The report delves deeply into the impact of the economies of scale, and concludes that both the success of NAND flash and the failure of Optane proved that wafer volume must approach 10% of the volume of a competing technology in order to reach cost parity.

Embedded NOR flash has become the first technology to convert in significant volume to an emerging memory technology, yielding to toggle-mode magnetic RAM (MRAM) and spin transfer torque RAM (STT MRAM), and in some cases SRAM is also being replaced with MRAM. Within the foreseeable future an emerging technology could displace some DRAM, well before ReRAM replaces NAND flash memory.

The rate of emerging memory technology development and increasing production volume will gradually result in lower prices, and low prices are the single most important driver to the acceptance of a new memory technology. Systems architects are interested in replacing volatile memory with high-speed, high-endurance nonvolatile memory and that bolsters the popularity these technologies will gain as their volume increases to reduce production costs (and thus purchase prices).

Ferroelectric RAM (FRAM) and some ReRAM technologies already ship to certain niche applications, despite the use of materials that are poorly managed in a production environment. HfO FRAM offers promise by using a material that is well-understood in today's FinFET fabs, and this might allow the number of niche markets available for FRAM to increase.

These new memory types enable new system configurations that have so far been unavailable. A nonvolatile main memory and cache memory in a computer will reduce power usage directly as well as enable new power saving modes, provide faster recovery

from power off, and enable more stable computer architectures that retain their state even when power is off. Spintronic technology, that uses spin rather than current for logic functions, could be used to make future microprocessors. A combination of spin-based logic and spin-based memory would be only natural on a single wafer, and could team up to enable very efficient in-memory computing. Many of the emerging memory technologies profiled in this report are being used as the basis of neuromorphic computing chips, which may become the basis for future AI-enabled endpoint devices.

The first widespread use of emerging memories is for embedded memory on a CMOS logic chip, replacing NOR flash, which has reached its scaling limit at 28nm, and is today frequently being replaced with MRAM and ReRAM. Single-transistor MRAM cells are now competing with multi-transistor SRAM, to dramatically reduce the number of memory transistors on a chip in order to provide a lower-cost, higher-density solution. Several enterprise and consumer devices currently use MRAM as an embedded memory, and this trend will continue.

MRAM processes have already been developed on conventional CMOS logic processes, allowing them to be built directly on top of CMOS logic wafers, using fewer additional mask steps than a more conventional flash memory. The use of a nonvolatile emerging memory can provide significant power savings in comparison to SRAM. As emerging memory cost per gigabyte ($/GB) approaches that of SRAM, this replacement could lead to significant market expansion.

It is projected that total baseline emerging memory annual shipping capacity will rise from an estimated 340TB in 2023 to 8.46EB in 2034. Total emerging memory baseline revenues are expected to increase from $421M in 2023 to about $71.7B by 2034. The bulk of this rapid revenue growth will be supported by emerging memories' displacement of SRAM, NOR flash and some DRAM.

Emerging memory demand should drive incremental demand for the capital equipment needed to manufacture these devices. This report models this requirement by focusing on the MRAM market. While MRAM can be built on standard CMOS circuits supplied by large semiconductor fabricators, specialized fabrication equipment is required for the MRAM layers. This additional equipment is similar to or the same as that used in manufacturing the magnetic read sensors in hard disk drives. Other emerging memory technologies will require their own complement of additional equipment, most of it identical to tools that are already in widespread use.

Growing demand for emerging memory technologies will increase total manufacturing equipment revenues by an estimated $53.4M in 2023 to between $434M to $3.8B by 2034 with a baseline projected spending of $2.4B.

INTRODUCTION

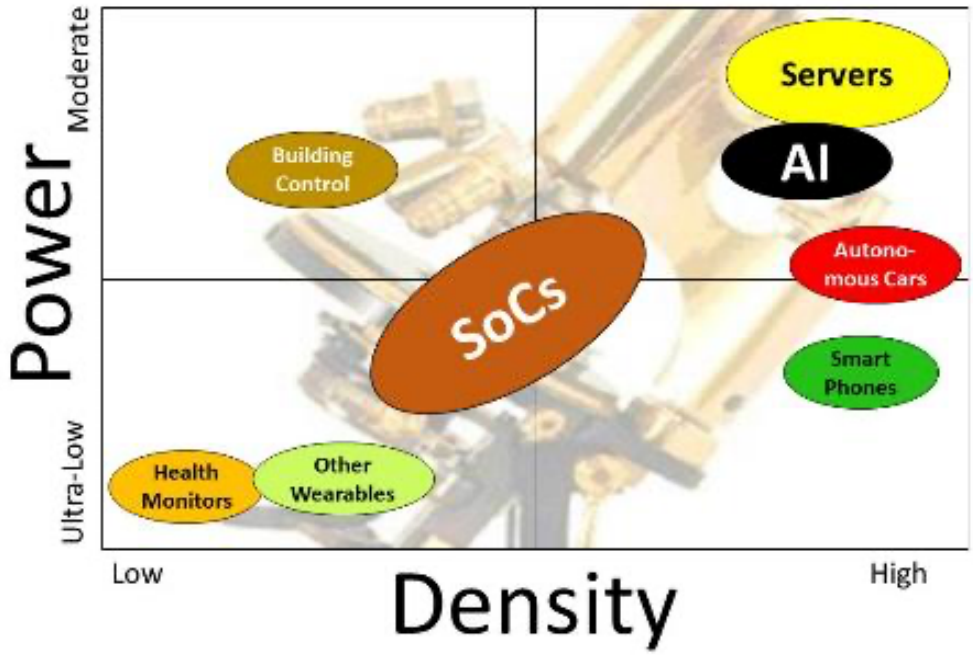
As the Internet of Things builds out, and as a growing number of devices make new measurements of data that was previously unmeasured, the world's data processing needs will grow exponentially.

This growth will not be matched with increases in communication bandwidth, despite the adoption of new wireless standards like 5G. In response to this mismatch, an increasing amount of processing will be performed at the edge, particularly in single-chip devices whose power consumption must be kept in check. This is the application that promises to benefit the most from today's nascent adoption of emerging memory technologies.

IoT endpoints need to process and interpret this data, often in near real time. Combine this with the amount of data that must be stored and processed to train various types or artificial intelligence (AI) models, and the result is a great increase in demand for memory and storage.

Figure 1 provides a rough idea of the memory density and power consumption needs of a number of modern application types.

Figure 1. Memory Density and Power Requirements by Application Category



Source: Objective Analysis, 2018

In order to make this information useful, our ability to process and interpret this data must also increase—the use of fast nonvolatile memory devices will be a key element in making this vast amount of information useful. Improved processing performance requires a good match of processor speed with communication bus speeds as well as memory data rates. It also requires latencies that are adequate for the application being processed.

More advanced semiconductor processing technologies will also play a key role in the transition to emerging memories, as established memory technologies like NOR flash and SRAM fail to keep pace with the relentless march of Moore's Law.

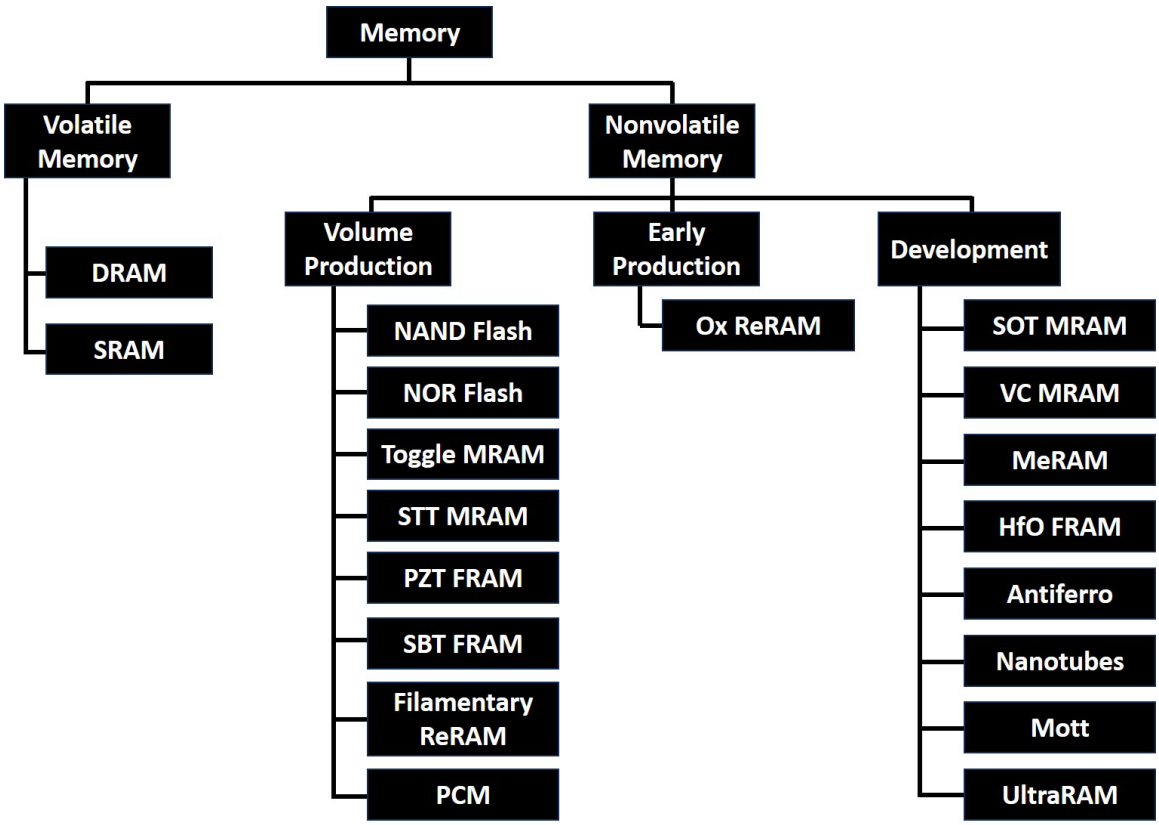
This report investigates the technologies associated with solid-state nonvolatile memories and evaluates each from the viewpoint of competition with current volatile memories, magnetic hard disk drives and flash memories.

There are numerous emerging solid-state nonvolatile memory technologies described in the literature, some rather old, while others are more recent. Some products such as ferroelectric random-access memory (FRAM), phase change RAM (PCM) and toggle and STT magnetic RAM (MRAM) are available as low to moderate density memory or storage products. Other technologies including spin-orbit torque MRAM and resistive RAM (ReRAM) are in earlier stages of pilot or production ramps. Technologies such as carbon nanotube RAM (NRAM) are still primarily laboratory demonstrations.

The move to emerging technologies is not governed simply by technical issues. Intel's recent foray into the emerging memory business with its Optane brand of PCM "3D XPoint" memory has proven that the economies of scale play a role in a new technology's adoption that cannot be ignored.

This report discusses the leading solid-state memory and storage technologies shown in Figure 2.

Figure 2. Solid State Memory/Storage Technologies



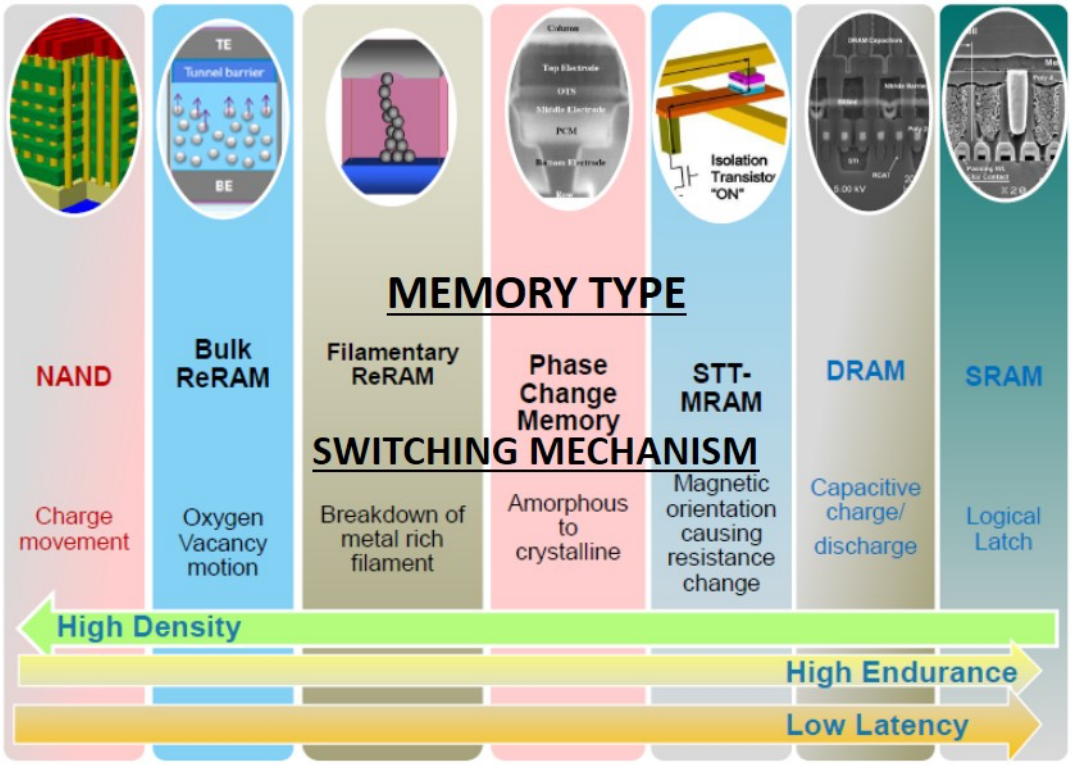
Source: Objective Analysis, 2024

The following emerging memory technologies are discussed in this report:

1. MRAM: Memories that use a magnetic effect
2. PCM: Memories based on phase changes through thermal effects
3. ReRAM: Memories that use ionic effects
4. FRAM: Electric field driven memories
5. New nonvolatile technologies such as Carbon Nanotube memories

The chart of Figure 3 provides an elementary overview of the first three of these technologies against established memory technologies: NAND flash, DRAM, and SRAM. This report will study all of the technologies mentioned above in considerable depth.

Figure 3. Rough Comparison of Established and Emerging Memory Types



Source: IBM. Used with Permission

This report addresses the characteristics of each nonvolatile technology from the following properties:

1. Compatibility with CMOS process and structure
2. Present commercial state of the technology
3. Reliability and/or endurance
4. Cost
5. Density
6. Performance
7. Ultimate limits

Two fundamental application types of emerging memories are examined in this report. The first is a standalone memory chip. Flash memory chips are an example of such a technology, as are DRAM chips. The second is adding an emerging memory within a logic chip like a microcontroller, ASIC, or SoC (system-on-a-Chip). These are often called embedded applications.

Note that the use of NAND flash memory as an embedded memory within a logic chip necessitates a significant variation to a standard CMOS process, which can multiply its cost. As a consequence, NAND flash is not used for internal memories in logic devices. NOR flash memory can be compatible with CMOS for an on-chip embedded memory. Similarly, DRAM is far less commonly found embedded in logic than is SRAM, its logic-friendly counterpart.

The report provides some details on the manufacture of MRAM products which are ramping up in volume to compete with SRAM, NOR and DRAM as a nonvolatile microprocessor memory. We will also point out what drives memory adoption and discuss why one memory can come to displace another.

The report then presents the Coughlin Associates analysis of MRAM production tool technologies and demand for these tools under a variety of scenarios, and the characteristics of capital equipment used for MRAM manufacturing. This demand estimate is based on the best-case of the above scenarios for MRAM market acceptance. We use this to project the capital equipment demand to produce MRAM from 2023-2034. We also discuss the implications for other nonvolatile technologies such as FRAM and ReRAM.

The report finishes with a listing of over 100 companies involved in various emerging memory technology development, production, and capital equipment.

The report presents data collected through careful interviews participants on all sides of the emerging memory market: Developers, vendors, licensors and licensees, tool makers, foundries, etc. Data was collected by Coughlin Associates and Objective Analysis via field interviews and at trade conferences and seminars. Other historical data was gathered from a wide variety of sources by the authors. Forecast data has been derived from primary and secondary sources, along with Objective Analysis' highly-regarded trend analysis of memory chip pricing and supply/demand dynamics. Capital spending reported here is an estimate since this is an emerging market. Conference attendance and extensive reading of widespread sources provided insight in technical areas. In certain cases, analyst judgment was called upon to expand upon the data that was available through these sources.

During reviews we were directed to lectures, interviews, and articles to obtain information concerning forecasts, expected technology directions, and future requirements. The authors wish to express our thanks to all colleagues who contributed their knowledge and insights to us as a valuable service to the storage and memory industry.

Not all of the detailed data that were gathered were included. As usual, if further questions or additional information is required, please contact the authors. The authors are also available to provide or develop additional information, on a project basis, (with due regard for the confidentiality of our sources.)

AUTHORS

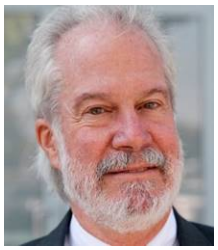
Tom Coughlin

Tom Coughlin has worked for over 40 years in the data storage industry and is President of Coughlin Associates, Inc. He has over 800 publications and six patents and is a frequent public speaker on storage and memory topics. Tom is active with the IEEE, SMPTE, SNIA, and other professional organizations. Dr. Coughlin is an IEEE Fellow and HKN member. He is co-chair of the iNEMI Mass Storage Technical Working Group, Education Chair for SNIA CSMI, Past-President of IEEE-USA and a board member of the IEEE Consultants Network of Silicon Valley. Tom is the 2024 IEEE President.

His publications include the ***Digital Storage Technology Newsletter***, ***Media and Entertainment Storage Report***, and ***The Emerging Memory Storage Report***. Tom is the author of ***Digital Storage in Consumer Electronics: The Essential Guide***, now in its second edition with Springer. He has a regular Forbes.com blog called ***Storage Bytes*** and had a long running regular digital storage column for the ***IEEE Consumer Electronics Magazine***. Tom is currently editing a column on digital storage and memory for the ***IEEE Computer Magazine***.

He was the founder and organizer of the ***Storage Visions Conferences*** as well as the ***Creative Storage Conferences***. He was general Chairman of the annual Flash Memory Summit for 10 years and was the program chair for 2 years. Coughlin Associates provides market and technology analysis as well as data storage technical and market consulting. For more information go to www.TomCoughlin.com

Jim Handy



Jim Handy, a widely recognized semiconductor analyst, comes to Objective Analysis with over 30 years in the electronics industry including over 20 years as an industry analyst for Dataquest (now Gartner), Semico Research, and Objective Analysis. His background includes marketing and design positions at market-leading suppliers including Intel, National Semiconductor, and Infineon.

Mr. Handy is a member of the Storage Networking Industry Association (SNIA) Compute, Memory, and Storage Initiative (SSSI). He is also a Leader in the GLG Councils of Advisors, and serves on the Advisory Boards of the Flash Memory Summit. He has served as a member of the Mass Storage Technical Working Group of the International Electronics Manufacturing Initiative (iNEMI). He is the author of three blogs covering memory chips (www.TheMemoryGuy.com), SSDs (www.TheSSDguy.com), and semiconductors for the investor (www.Smartkarma.com and formerly blogs.Forbes.com/JimHandy). He contributes from time to time to a number of other blogs and professional journals.

A frequent presenter at trade shows, Mr. Handy is known for his widespread industry presence and volume of publication. He has written hundreds of articles for trade journals, Dataquest, Semico, and others, and is frequently interviewed and quoted in the electronics trade press and other media.

Mr. Handy has a strong technical leaning, with a Bachelor's degree in Electrical Engineering from Georgia Tech, and is a patent holder in the field of cache memory design. He is the author of ***The Cache Memory Book*** (Harcourt Brace, 1993), the leading reference in the field. Handy also holds an MBA degree from the University of Phoenix. He has performed rigorous technical analysis on the economics of memory manufacturing and sales, discrediting some widely held theories while unveiling other true motivators of market behavior.

Mr. Handy may be contacted at Jim.Handy@Objective-Analysis.com, or by telephone at +1 (408) 356-2549.