Annual Update on Emerging Memories 2019

Mark Webb

MKW Ventures Consulting, LLC

www.mkwventures.com

mark@mkwventures.com

Topics

- Why New Memories
  - Scaling (new memories or new architecture)
  - The search for universal memory, New applications
- Historical and Recent Technologies compared
- Memory Product Lifecycle
- PCM/3D Xpoint, MRAM, ReRAM
- Additional Technologies
- Equipment/Market development challenges and solutions

Memory Technologies Reviewed
whats new?

	Latency	Density	Cost	HVM ready
DRAM	*****	***	***	*****
NAND	*	*****	*****	*****
MRAM	*****	*	*	***
3DXP	***	****	****	****
ReRAM	***	****	****	**
NRAM	***	**	**	*
Other	***	**	**	*

Why New Memories? (the "Pull")

- Existing Memories will stop scaling
  - Or will they?
- Existing memories are not meeting needs
  - Speed
  - Quality
- We want universal memory
  - NAND and DRAM are far from model techs

Why New Memories? NAND Scaling

- NAND flash has line of sight to 128L+ and QLC.
  - And no….. 256L does not take 6 months
- This takes us to 2023+ timeframe with a ~75% bit cost reduction
- We can do a 1Tbit Chip today, 1-2TB in a package
  - Sweet spot is 256Gb so density is not issue
- String stacking, CMOS under array allows a lot of circuit options
- If cost goes down, people will make the trade offs because it benefits them.
- NAND is not done anytime soon. And no one is even close on costs

DRAM Scaling

- DRAM scaling has slowed down
- This slow scaling WAS providing record margins, revenue and well managed growth
  - Price crashed and margins dropped. Now the model is being re--evaluated
- Plan is to get to 10nm region <u>very slowly</u> only if ROI is there
  - Most companies are reporting modest incremental changes that will take us to 10nm
- New architectures could work… if needed
- DRAM is fine for next 5 years

Why New Memory:
Not Meeting Needs

- DRAM and NAND are not meeting needs
- NAND is slow, block oriented, wears out
  - But nothing can approach it for price
- DRAM is volatile, density is growing slowly, it is less than ideal
  - But at least its RAM

Summary

- NAND Scaling is going to continue
  - With unfortunate quality and speed trade offs
- DRAM scaling is going to continue
  - Much slower than we wish!
  - No incentive to risk anything
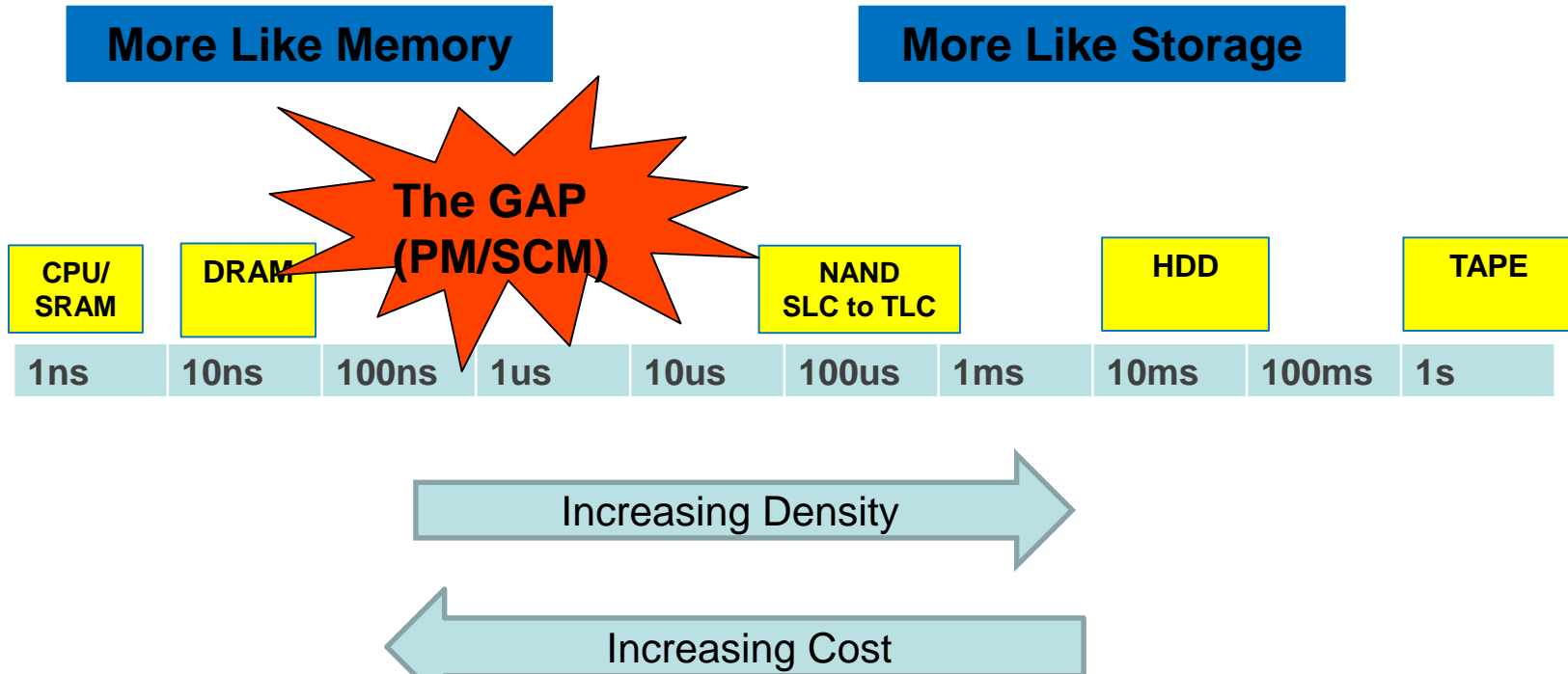- <u>Not meeting needs IS the really compelling reason for new memory</u>

What is Needed in New Technology

- Ideal Universal Memory
  - Fast as DRAM, Non-Volatile, Infinite cycles, Cheaper than NAND
  - This doesn't exist. Nothing is replacing DRAM or NAND
- Reality of what will happen
  - Combination of tradeoffs will need to be made
  - Compute architecture designed around what is available
    - Phone, PC, Server designed around DRAM+SCM or SCM+ cloud
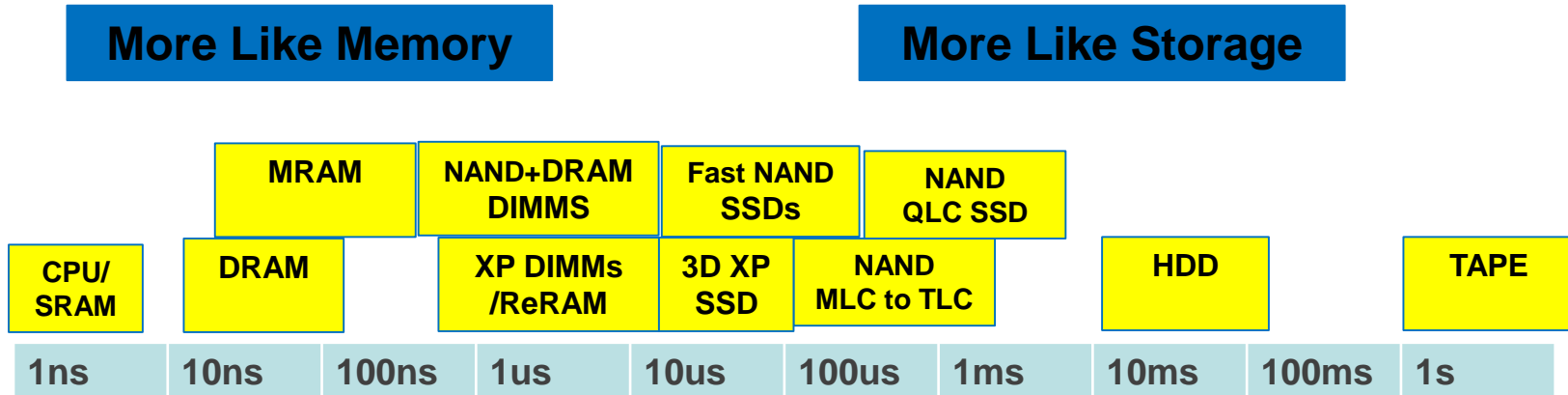- We are making REAL progress here!

Mark Webb, MKW Ventures Consulting

The Latency Spectrum and Gaps ~2015

More Like Memory

More Like Storage

The GAP (PM/SCM)

| CPU/ SRAM | DRAM | | NAND SLC to TLC | HDD | TAPE |

| 1ns | 10ns | 100ns | 1us | 10us | 100us | 1ms | 10ms | 100ms | 1s |

Increasing Density →

← Increasing Cost

Mark Webb, MKW Ventures Consulting

The Latency Spectrum and Gaps
Future

More Like Memory

More Like Storage

	MRAM	NAND+DRAM DIMMS	Fast NAND SSDs	NAND QLC SSD		
CPU/ SRAM	DRAM	XP DIMMs /ReRAM	3D XP SSD	NAND MLC to TLC	HDD	TAPE

| 1ns | 10ns | 100ns | 1us | 10us | 100us | 1ms | 10ms | 100ms | 1s |

Increasing Density →

← Increasing Cost

Mark Webb, MKW Ventures Consulting

Today is VERY Different than 2014

- We now have a volume PCM (3D Xpoint) memory ramping with 100s of Millions in revenue in multiple markets
  - More revenue than all previous "new NVM"
- We have ReRAM technologies you can pursue for your product at a foundry
- We have MRAM being used in actual products
- We have multiple companies with <u>great financial resources</u> who are developing MRAM for embedded
- We still have new memories being proposed (5+ years)
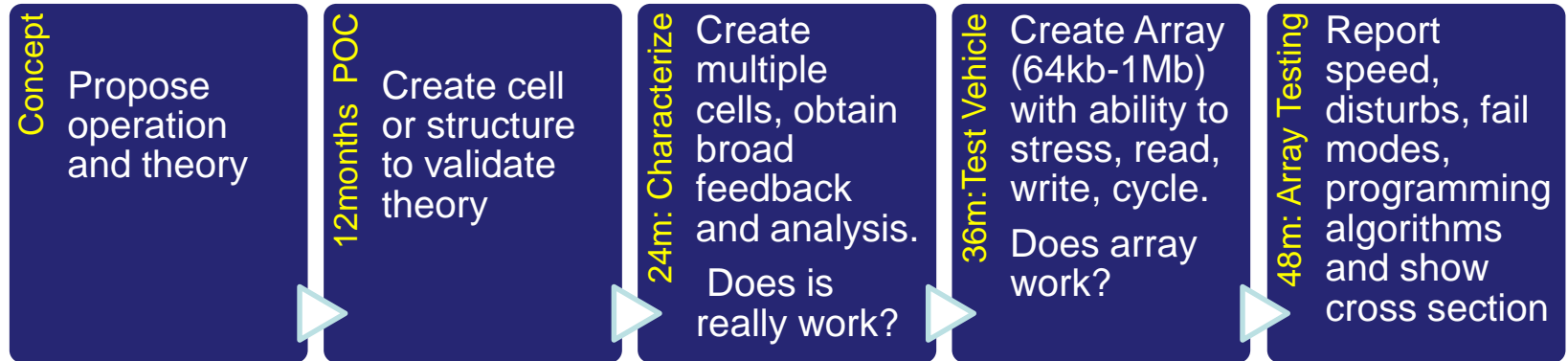- This is VERY different than past

NEW MEMORY PRODUCT LIFECYCLE

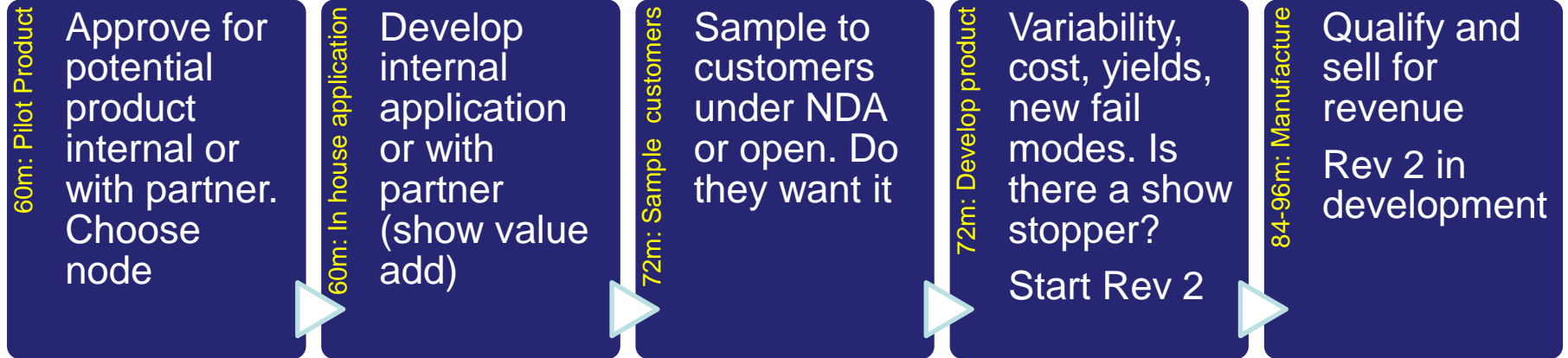# NEW: Product Lifecycle for Memory Part 1: Technical Proof/Open Communication

Concept	12months POC	24m: Characterize	36m:Test Vehicle	48m: Array Testing
Propose operation and theory	Create cell or structure to validate theory	Create multiple cells, obtain broad feedback and analysis. Does is really work?	Create Array (64kb-1Mb) with ability to stress, read, write, cycle. Does array work?	Report speed, disturbs, fail modes, programming algorithms and show cross section

Summary: " A Novel xRAM technology built in 90nm Lithography
With pictures! (4 years after concept)
Estimate: 90% of technologies fails to ramp AFTER achieving this

Product Lifecycle for Memory Part 2: Usually Stealth
"We are going to build a product"

60m: Pilot Product
Approve for potential product internal or with partner. Choose node

60m: In house application
Develop internal application or with partner (show value add)

72m: Sample customers
Sample to customers under NDA or open. Do they want it

72m: Develop product
Variability, cost, yields, new fail modes. Is there a show stopper? Start Rev 2

84-96m: Manufacture
Qualify and sell for revenue Rev 2 in development

7-8 Years assuming no restarts or resource limitations from concept to revenue
Intel-Micron 3D Xpoint was demonstrated in an array 8 years ago.
MRAM was proposed 20 years ago and starting manufacturing 10 years ago

Problem with Hype/New tech

- Jump to page two before completing page one. Fine to speed flow, but we know that 90% of technologies don't make it.

- It rarely makes sense to be stealth in part one
  - You want to show research, You need funding
  - The work is either patented or is not novel
  - You need to get feedback

- Open and closed companies follow this (FinFet, 3D NAND, PCM, MRAM, ReRAM, etc

- RECOMMENDATION: Evaluate all technologies and chance for success based on where they are on lifecycle.
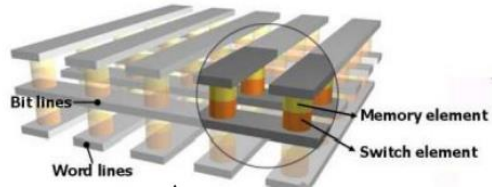
3D Xpoint

What is 3D Xpoint

- Micron/Intel Technology announced in July 2015
- Thanks to great analysis last 2 years, we know
  - PCM technology , 20nm lithography
  - Cross point array, Selector in stack
  - We have cross sections and details
  - 10+ years in the making, multiple publications by Intel, Micron, Numonyx, IP providers, etc

Simple Architecture Overview
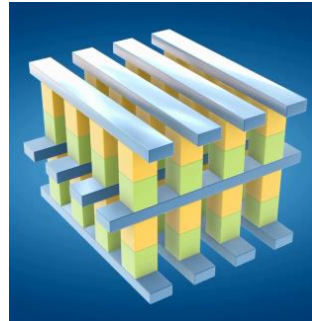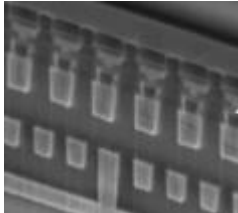
Source: Intel

Speculated details on Technology based on:
- Memory and SSD modules sold
- EE-Times/Techinsights
- The register/Ron Neale/Chris Mellor Feb 1 2016
- ISS 2016 (Jan 12)
- Dave Eggleston FMS2015
- Plus multiple FMS presentations from Techinsights

Current and Emerging Memory Technology Landscape
Atwood FMS 2011
Note: 8 Years ago.
Papers in 2001, 2005, 2009

3D Xpoint Technology Update 2019

- Products
  - Fastest SSD, with best SSD endurance, selling millions of units
  - DIMMs are quite delayed with multiple launches but <u>are selling today</u>.
  - More Optane/3D Xpoint bits sold than all other emerging memories combined in 2019

- Technology
  - Technology shipping today is based on Gen 1
  - Endurance increased significantly over past 2 years
  - Intel/Micron Plan Gen 2 for release in 2020. <u>Estimate</u> is 4 stacks, similar lithography, SLC, 35% lower bit cost. Nominally 256Gbit
  - Expect Gen 2 measurable volume in late 2020

3D Xpoint Technology Numbers

- A model for what 3D Xpoint…
  - 128Gbit Chip with >10% overprovisioning on the chip itself
    - DIMMS are overprovisions by about 25% and have DRAM onboard
  - Read Latency (Chip): ~125ns, Write Latency: "higher"
    - DIMM is spec'd 350ns read, estimated at 650ns+ Write
  - Chip Endurance: ~200K+ cycles spec with management techniques
    - DIMMS do not have endurance limitations in actual use
  - Great persistent memory, but not replacing NAND or DRAM

Optane Persistent Memory (DIMM)

- Model: Optane persistent memory won't replace DRAM, It supplements it.
- Server configurations with Optane DIMMS have above average DRAM capacity installed as well. Intel recommends Xpoint:DRAM ratio of ~5:1
- Memory controller manages endurance and performance by moving data between Xpoint and DRAM.
- If >90% of reads and writes are to DRAM, occasional R/W to slower Xpoint is not impactful
- CPU/memory controller is optimized to work with this configuration
- DRAM+NVM Solution is the OPTIMAL persistent memory solution

3D Xpoint Competition

.

- Competing PCM technologies: All companies have worked on PCM, Many showed information at recent conferences.
    - We expect Crosspoint PCM announcements from multiple companies
    - Other companies have expertise and IP from previous work
    - Top issues are selector, and more importantly 10 years of optimization
- MRAM, ReRam and others will be discussed later
- Fast NAND: Fast NAND+DRAM solutions from NAND/NVDIMM companies will compete in both DIMM and SSD formats.

3D Xpoint Capacity/Manufacturing

- 3D Xpoint is manufactured only in IMFT facility in Lehi, Utah
  - Was a JV with Intel, Micron has executed its option to take 100% ownership around October 2019
  - Micron is <u>required</u> to provide capacity to Intel for another year and <u>may</u> provide capacity long after.
- Intel has moved 3D Xpoint (Intel Optane Memory Media) development to Rio Rancho, NM facility. 2nd Gen+ development will continue there.
- Without additional tools, We estimate IMFT can support Intel volume and Micron initial shipments. Factory could ramp 2-3x from these levels if needed

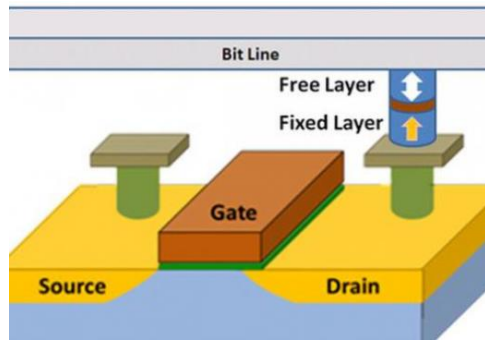|	Output Q4 2019	Output Q4 2020
Gen 1	45M GB/Month	45M GB/Month
Gen 2	5M GB/Month	25M GB/Month

New Memories

MRAM

- Storage in Magnetic Element
- Toggle is historical tech
- STT-MRAM provides scalability
- One of the oldest "new NVM"
- 1T1R. Cross point options are not clear
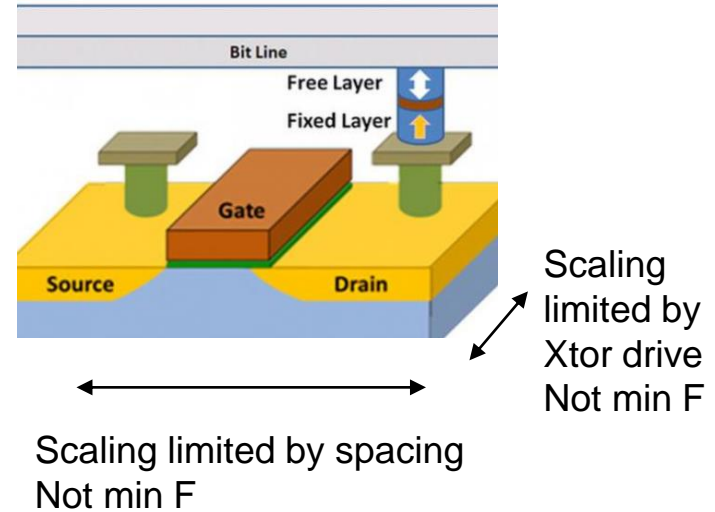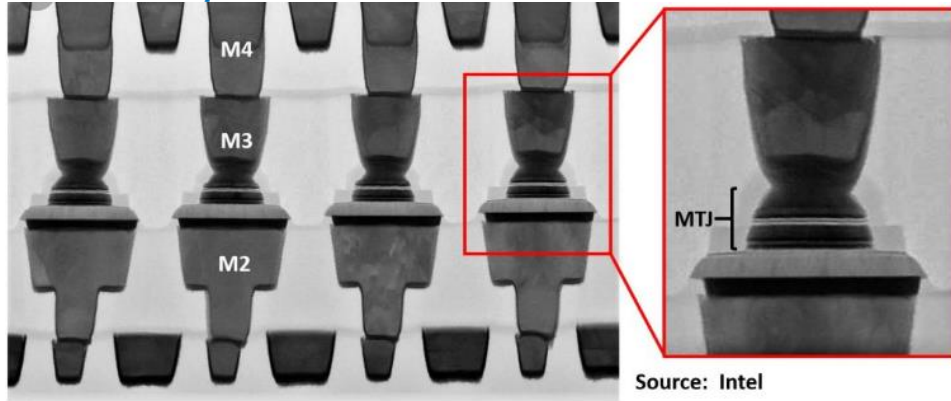
MRAM Compared to Other Technologies

- **MRAM Advantages**
  - Speed. It is the fastest new NVM
  - Endurance is higher than competition but not infinite
  - Maturity compared to other "new NVM"
  - Multiple manufacturers joining development
    – IP providers, Foundries, Memory Companies
    – NEW: Samsung, Intel, TSMC, Global are all developing MRAM and have committed published technologies
    – Equipment suppliers engaged and spending money on development

Mark Webb, MKW Ventures Consulting

MRAM Compared to Other Technologies

- **Disadvantages**
  - Cost/Density
    - $8F^2$ theory, $10\text{-}20F^2$ planned, $50\text{-}100F^2$ is more average today
    - 22nm technologies have cell size of .048u2 (Intel, Global, TSMC)
  - STT is less mature than Toggle
    - Toggle ships today but doesn't lead to cost effective applications
    - STT is the focus for all future applications
  - Even with STT, density cannot match ReRAM, 3D Xpoint
    - MRAM current needs limit ability as a 3D technology

Mark Webb, MKW Ventures Consulting

MRAM Scaling Challenges

Source: Intel

Scaling limited by Xtor drive Not min F

Scaling limited by spacing Not min F

End Result: Cell area ends up at 50-100 times "F^2" where F is minimal half pitch or features size

Technology Status
Widespread Availability

- Companies shipping measurable volumes of MRAM
  - Toggle today as it is a mature technology with sales
  - STT-MRAM provides higher density and is the future
- Everspin partnering with Global Foundries to ship stand alone and embedded
  - 28nm MRAM standalone being planned (40nm Shipping now)
  - 256Mb Shipping, Plan to 28nm Ship 1Gb 2019
  - 22nm embedded MRAM available in upcoming GF designs
- Multiple Companies licensing IP to improve performance and reliability
- Multiple major logic companies and foundries are committed to MRAM
- IMO: MRAM future growth confidence is a "Tale of Two Markets"

Mark Webb, MKW Ventures Consulting

A Tale of Two Markets

- Embedded: MRAM is ideal for market (revenue not measurable)
  - Potential to replace NOR, SRAM, DRAM applications
  - Ability to integrate (metal stacks), density (Mbit), performance (DRAM) match embedded needs well
  - Endurance work needed for full RAM replacement
  - Looks like Embedded "Universal Memory"… Why isn't it here already?
- Discrete: Targets and Market growth uncertain (Rev <$100M today)
  - Target applications requires specific density, speed, with NVM requirement.
  - Small markets exist, but they are vulnerable to attack on all sides.
  - No measurable NAND replacement market (Too small, expensive)
  - Performance/cost/density ratio not on track to match DRAM
  - Needs to dominate Niche or have "Killer App"

Embedded MRAM Updates

- MRAM is potential replacement for eFlash and SRAM given their issues (Scaling)

- Array size is optimal for embedded 8-64Mbit, 10Mb/mm2

- Major updates over the past year

- Multiple announcements and demonstrations
  - TSMC 22nm
  - Intel 22nm
  - Samsung 28nm
  - Global (not new, more details now)

- This is Extremely important to meet the financial challenges of new memories

MRAM Stand Alone Memory

- **Standalone MRAM memory has more challenges than embedded**
  - No short term path to MRAM being able to match DRAM on cost or density
  - NOR flash is a very mature execute-in-place NVM in 256MB and below
  - Higher density NVM (>1Gbit) will use NAND due to extremely low cost
- **So MRAM does not have a path to replace stand alone NAND or DRAM on broad basis**
- **Therefore MRAM is best applied to 256-1Gbit where DRAM like speed is desired, NVM needed, and cost is not a major issue**
  - MRAM Replaces battery/capacitor backed DRAM, Low density DRAM, NOR applications.

Stand Alone MRAM (cont)

- **Everspin shipping 256Mb for many quarters**
- **Announced 1Gb availability**
  - Long delayed chip on 28nm
  - "First Chip to Start with a G"
- **Allows multiple GBs in DIMM/SSD format**
  - This is breakthrough for SSD caching, logging, low density NVM

Side note: Thoughts on Endurance

- Endurance metrics are FLAWED when looking at new technologies. Single cell theoretical claims (10^6, 10^8, 10^{12})

- MLC NAND is listed as "5000 cycles"
  - Most bits probably last >50,000 cycles (don't get me started on UBER and RBER
  - Today's TLC is 3000 cycles, QLC might be 1000 cycles
  - Endurance management can detect and predict many failures. Wear leveling and overprovisioning can multiply endurance >2X

- MRAM is often listed as 10^{10} to 10^{15}
  - Typically these are average numbers

- Look for numbers that match what volume companies are sharing
  - Specification, at a certain fail rate, with ECC. 100K cycles at 1000DPM

ReRAM
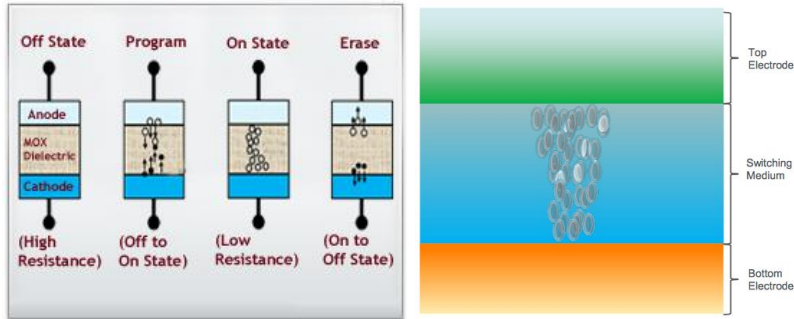
Storage by changing resistance of a element

Example from Filament ReRAM

High resistance State
Bias applied: Causes filament to form (low resistance)
Reset to eliminate filament to return to high bias state

What We Know About ReRAM

- Lots of technology papers on storage and select elements
- Historic Revenue has come from embedded applications in Kbits or Mbits
  - Unity (Rambus), Adesto, Panasonic, etc
  - Higher Density Needed
- 2018: Crossbar,Inc ReRAM is available today as embedded memory (1T1R) from Foundry. Standalone memory planned
- 2017: 4DS announced functionality on 1000+ cells, DRAM like speeds. Now partnering with IMEC for development
  - Interface switching is different from historical filament base
- Multiple Memory companies are investigating ReRAM

Updates

- Intel Announced embedded ReRam at ISSCC
  - 10Mb/mm2
  - 5ns access
  - 22nm technology
- 4DS is reporting data on 300mm wafers that show progress as they move to a 1Mbit test chip in futures
- ReRAM is considered for many neuromorphic and AI applications
- Applied Materials showed strong support for ReRAM and other new memories with announcements in July
- Adesto is in full production Stage 10 at very low density, Crossbar is in stage 7 (Part 2) of development, 4DS is in Stage 4 (part 1) of PLC

Why ReRAM

- Scalable, Crosspoint solutions are low cost ($4F^2$) relatively simple and allow foundry utilization

- No significant new Fab tools needed (model)

- Selector elements allow for high density 1TnR

- Latency is far better than NAND, 100nS range

- Cycling past 1M cycles on cells is achievable

- All of these are in same category as 3D Xpoint.
  - Faster than NAND, slower than DRAM
  - Cheaper than DRAM, More expensive than NAND

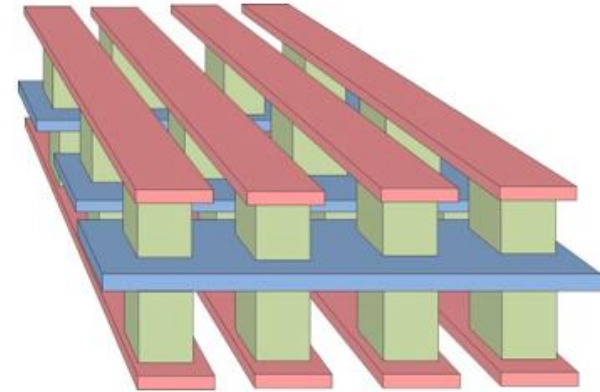- Scalable, stackable, high density, lots of research

Previous Technologies Reviewed

Device	COST $/GB
2017 DRAM COST	1X
27nm	1.1x
20nm	0.6x
14nm	0.4x
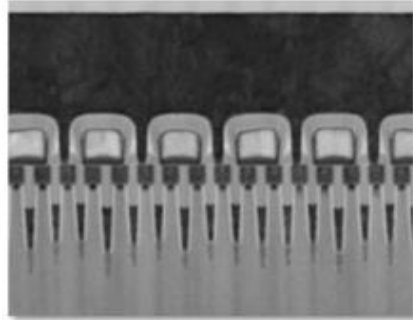
Advanced 27nm RRAM Cell

Device	COST $/GB
2017 DRAM COST	1X
24nm	0.6x
14nm	0.3x
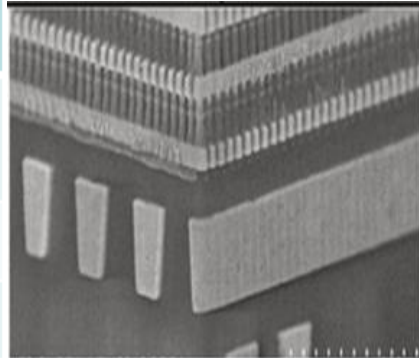
- Multiple ReRAM technologies reviewed
- **Stacked Crosspoint (~4F^2) is optimum for next 5 years**
- Allows memory optimized or foundry
- 3D NAND like vertical later

Additional Memories

- 3D Xpoint, MRAM, ReRAM are in some stage of availability today.

- All have been in development for years

- Previous versions of PCM memory from Intel/Micron were published in detail with 64M arrays almost 10 years ago
  - MODEL: IT IS 10 years from tested array to revenue
  - That is with 100s of Millions in investment

- Other memories show promise but have limited chip level data so ramp in next 5 years is not clear

NRAM
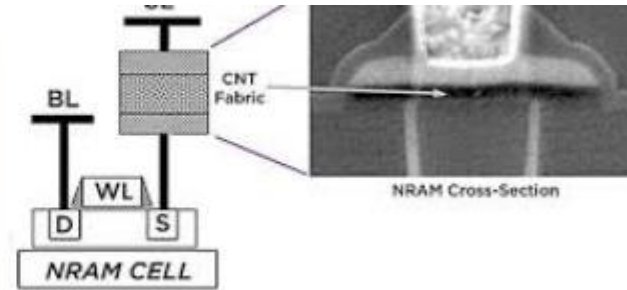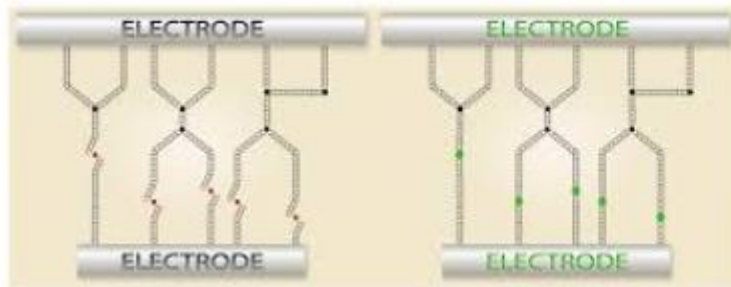
- Carbon Nanotube Memory. Nanotubes aligned to conduct or not based on bias

- Reported to be 10ns latency, infinite endurance, scalable to 16Gb DDR4 memory and beyond

- Nantero announcements: Low density at Fujitsu 2019, high density samples 2019, production 2020. Nantero partners will do more announcements

FRAM
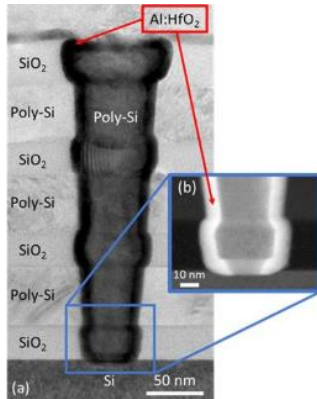
- FRAM have been used in low density (Kb) for years
- Recent IMEC focus has targeted higher density.
- Potential for speed and endurance is driver, tradeoffs TBD
- Advantages: Low write current, DRAM cell size, high speed
- Potential "DRAM Replacement" with supporters
- Timing: middle of part 1 (>5 years from volume)

Mark Webb, MKW Ventures Consulting

Memory Costs

NAND Costs

We track memory costs for each <u>technology</u> at each <u>company</u> and publish updates with reasons for changes for each company

Key takeaway:

25% cost reduction per year

Moving from 64-128+ layers continues to reduce cost

Changes to market and technology will affect 128 cost reductions and leadership

2018/2019/2020 NAND Die Costs
Does not include packaging, Final test, module costs

Categories: 64L TLC (A), 64L TLC (E), 64L TLC (B), 64L TLC (C), 64L TLC (D), 64L QLC (B), 96L TLC (A), 96L TLC (B), 96L TLC (E), 96L TLC (C), 96L TLC (D), 96L QLC (B), 96L*QLC(D), 128L TLC, 128L QLC

Legend: ■ Cost 2018* ■ Cost 2019 ■ Cost 2020

Mark Webb, MKW Ventures Consulting
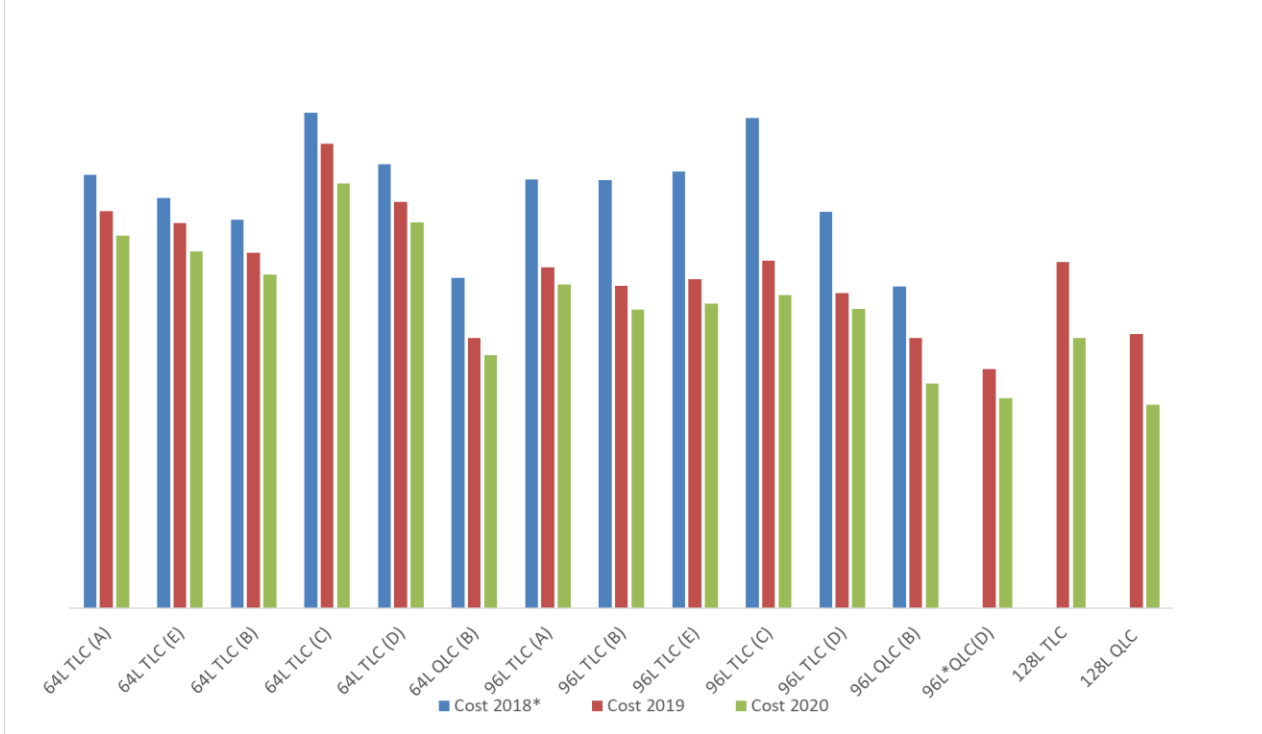
Cost for Multi-Gbit Technologies

New NVM Costs Compared
Cost/GB

DRAM   SLC NAND   Micron 3DXP Actual   STT MRAM Example   Crosspoint ReRAM

2018
2019
2020
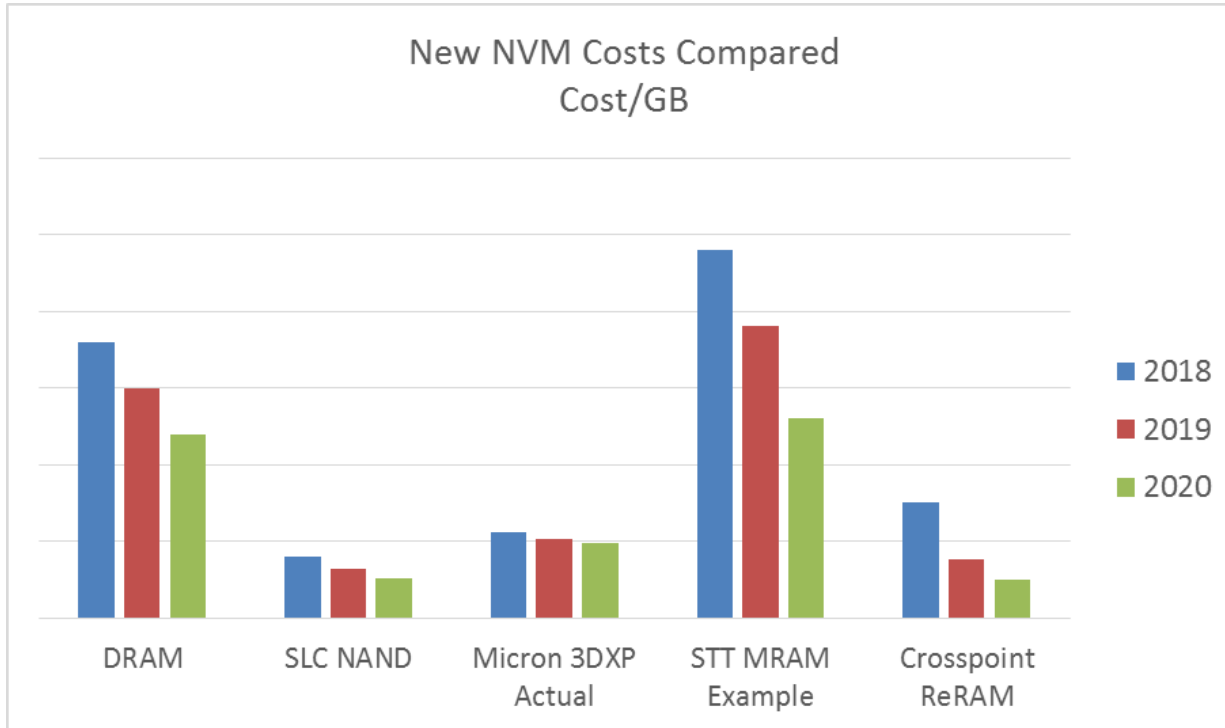
Emerging and Other Memory Cost

- Detailed reports available on each … below is a ROUGH SUMMARY (order of magnitude)

Memory/Storage	Cost/GB (high density)	Notes
DRAM	180c/GB	High volume
3D XPoint	70c/GB	Shipping today. Cost assumes full ramp
QLC NAND	6c/GB	NAND Cost Leader today
Low latency NAND	30c/GB	Being sampled today
ReRam	~150c/GB	Target No, could match 3DXP long term
MRAM	~1000c/GB	Large cell size
SRAM	~900c/GB	Depends on logic, speed needed. Embedded in logic
NOR	~1000c/GB	Often sold in low density at wildly varying prices
HDD	2c/GB	Average cost for >1TB capacity
Tape	1c/GB	Lowest cost for archival

Summary of Emerging Memory

	Pros	**Cons**	**Summary**
3D XPoint	Fast SCM, in production, relatively high density and low cost	Can't cycle like DRAM. Slower than DRAM. Scaling	Highest Volume new NVM
MRAM	Fastest NVM… Multiple products in production, embedded apps	high cost, relatively low density	In volume with multiple applications
ReRAM	High density, low cost, Fast SCM, available today	Can't cycle like DRAM. Slower than DRAM. Low volume	Will compete with 3DXP apps with volume release
NRAM	Fast, high density, no cycling issues, stable	No manufacturing data, limited reports	Planning sample in 2019, volume in 2020
FRAM	Fast, potential for endurance	5+ years	Need data
Other	In Concept mode	10+ years	Need data

Mark Webb, MKW Ventures Consulting

Memory Product Lifecycle

Concept — Propose operation and theory	**12months POC** — Create cell or structure to validate theory	**24m: Characterize** — Create multiple cells, obtain broad feedback and analysis. Does is really work?	**36m:Test Vehicle** — Create Array (64kb-1Mb) with ability to stress, read, write, cycle. Does array work?	**48m: Array Testing** — Report speed, disturbs, fail modes, programming algorithms and show cross section
1	2	3	4	5

6	7	8	9	10
60m: Pilot Product — Approve for potential product internal or with partner. Choose node	**60m: In house application** — Develop internal application or with partner (show value add)	**72m: Sample customers** — Sample to customers under NDA or open. Do they want it	**72m: Develop product** — Variability, cost, yields, new fail modes. Is there a show stopper? Start Rev 2	**84-96m: Manufacture** — Qualify and sell for revenue Rev 2 in development

Equipment/Investment Challenges and Solution

- It takes 100s of millions to properly develop a new memory to a volume market
  - The revenue while in development is an order of magnitude less
  - Equipment vendors only support high volume products
- Samsung and Intel both overcame this, Can others?
- Solution options
  - Large Memory manufacturer needs to see continuation of its memory business or extension to fund investment (Billions)
  - Foundries allow multiple uses of equipment and fund embedded. Eventually manufacture stand alone

Future NVM Scenario 2019

- NAND and DRAM stay through 2025. But we have options!
- PCM/3D Xpoint will initially dominate the NAND/DRAM gap. Ramp in 2019-2020
- Competing PCM chips proposed by other companies with similar architecture to Xpoint will be released in 2020 timeframe
- ReRAM develops to address similar market with similar performance and cost to 3D Xpoint
- MRAM continues to grown quickly in embedded, and in lower density, higher speed applications
- Chip Samples/technical evaluations on NRAM, FRAM reported in 2020. based on results, revenue products are proposed

Summary

- NAND has 5+ years and 3+ generations to go.

- DRAM will slowly scale for next 5 years

- There is no QLC NAND replacement… ultracheap memory

- We have solid new memories that are actually available and shipping

- These will grow and augment, not replace DRAM and NAND

THANKS!

- ## Questions

- ## Mark Webb

- ## www.mkwventures.com