



Flash Memory Summit

Scaling and Cost Forecast for DRAM, NAND, and Emerging Memories

Mark Webb

MKW Ventures Consulting

www.mkwventures.com

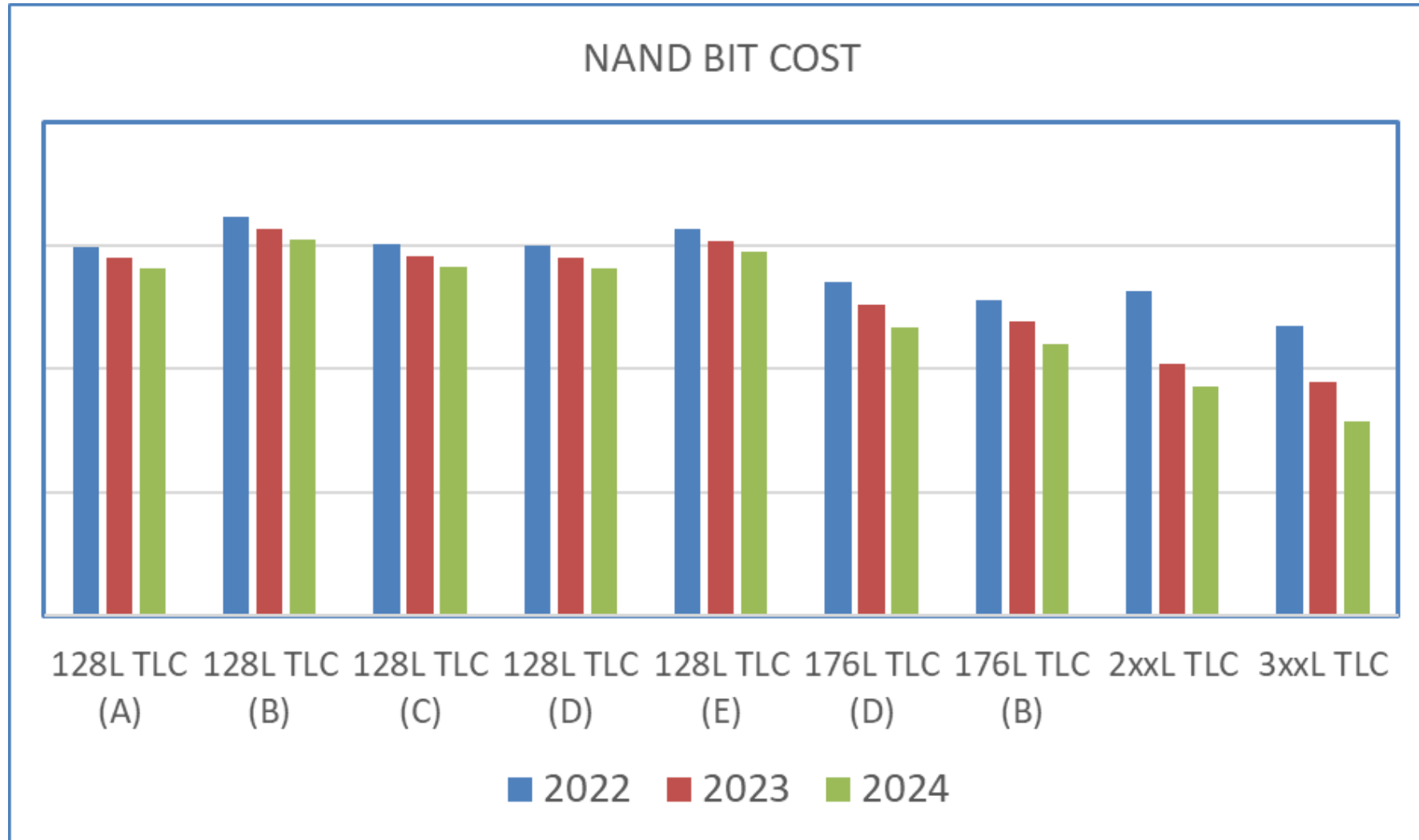
- Memory Scorecard
- NAND Scaling and cost
 - Bit growth and revenue
- DRAM scaling and cost
 - Bit growth and revenue
- MRAM/ReRAM (1T1R) scaling and cost
- PCM/ReRAM (crosspoint) scaling and cost
- Impact in next 5 years, 10 years

Memory Technologies Reviewed

whats new?

	Latency	Density	Cost	HVM ready
DRAM	*****	***	***	*****
NAND	*	*****	*****	*****
MRAM 1T1R	*****	*	*	***
RRAM 1T1R	*****	*	*	***
3DXP/Optane	***	*****	*****	*****
NRAM	***	**	***	*
FE RAM	***	**	***	*
Other	***	**	**	*

NAND BIT Costs By Technology

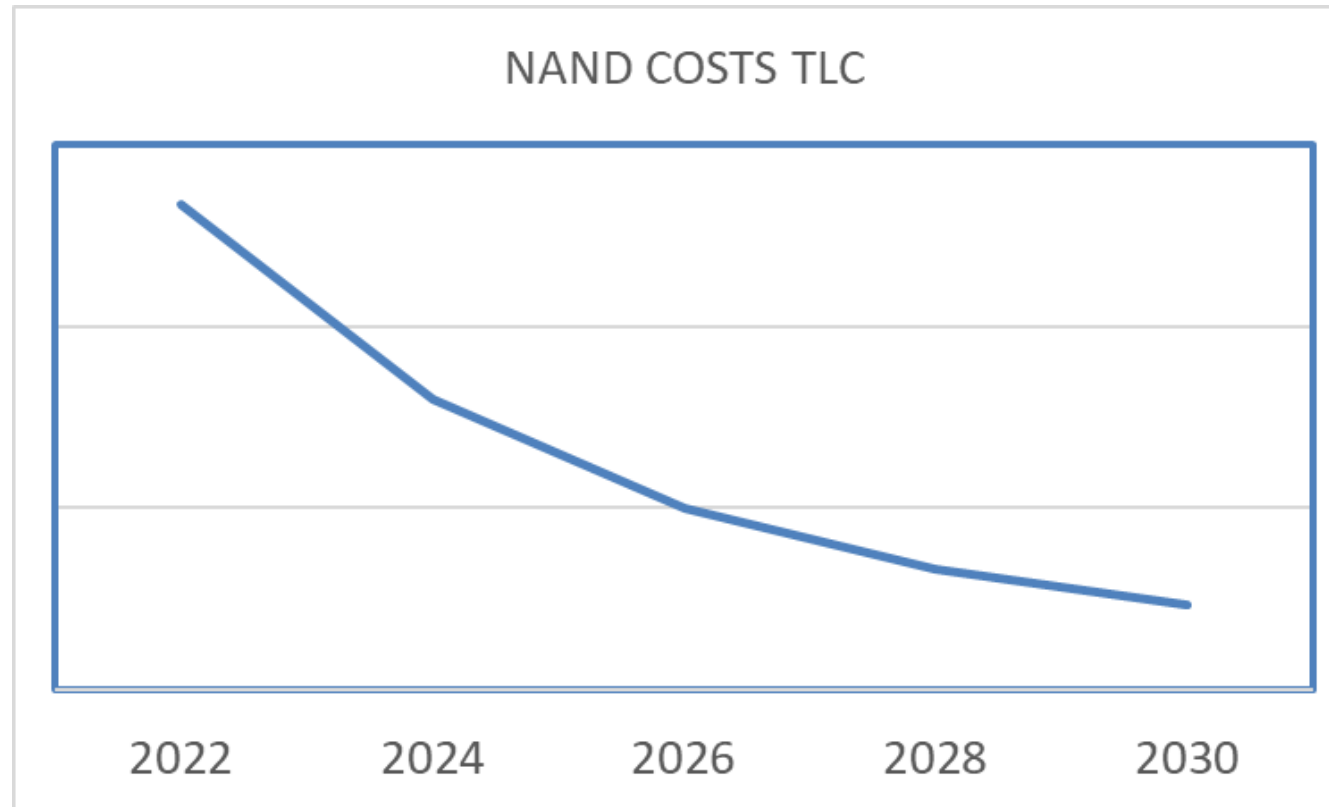




NAND Model and Takeaway

- NAND is pragmatic about adding layers. Choice of layers is based on what makes sense for your technology, tool set and cost.
 - Expect strange layer counts, and more layers not always being cheaper
 - Expect a 30-40% add in layers per tech. If 20% was easy, that could happen. If 10% Capex allows 60% more layers ...that could work
- Adding layers adds some smaller amount of wafer cost.
 - Historically, add 15% for adding 50% more layers. Each gen is slightly different.
- But each year Fabs become about 5% more efficient with existing technologies. Inflation has an unknown impact at this time
- Cost reductions are not dramatic, but by choosing layer count, we can be confident in steady, continuous, cost reduction (Assuming yield ramp)
- QLC is ~20% cheaper in model. 5% inefficiency and test time cost modeled in.
- There is no brick wall. Things get complicated with string stacks/tiers and potential for bonding options... but scaling is good through 500+ layers.
- Lots of details with all of these options we can discuss offline

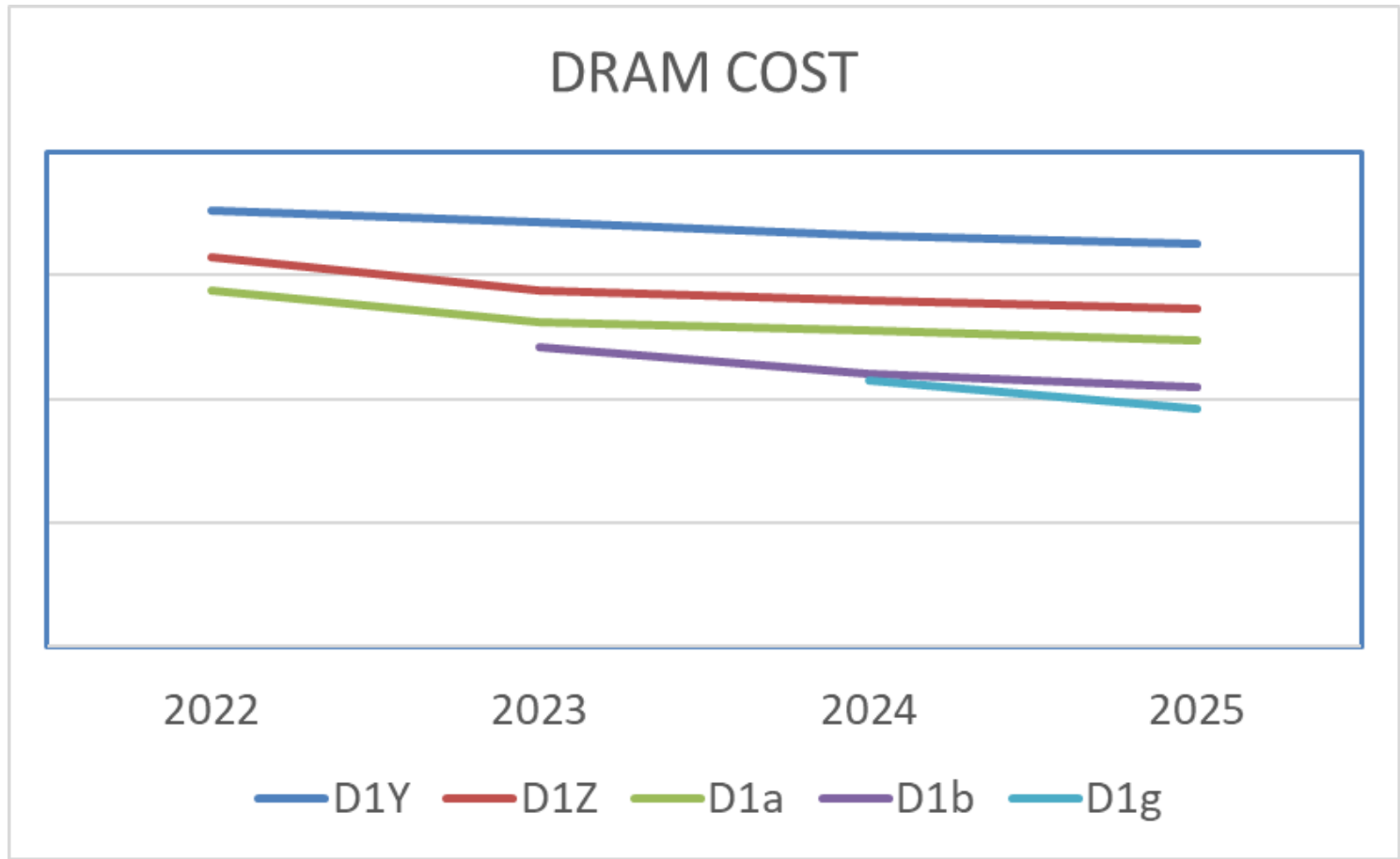
NAND Bit Cost Extrapolation to 2030



Costs Continue to Drop for next 8-10 years. Average will be 15-20% per year

DRAM Costs Over Time

(across all suppliers, each company different)

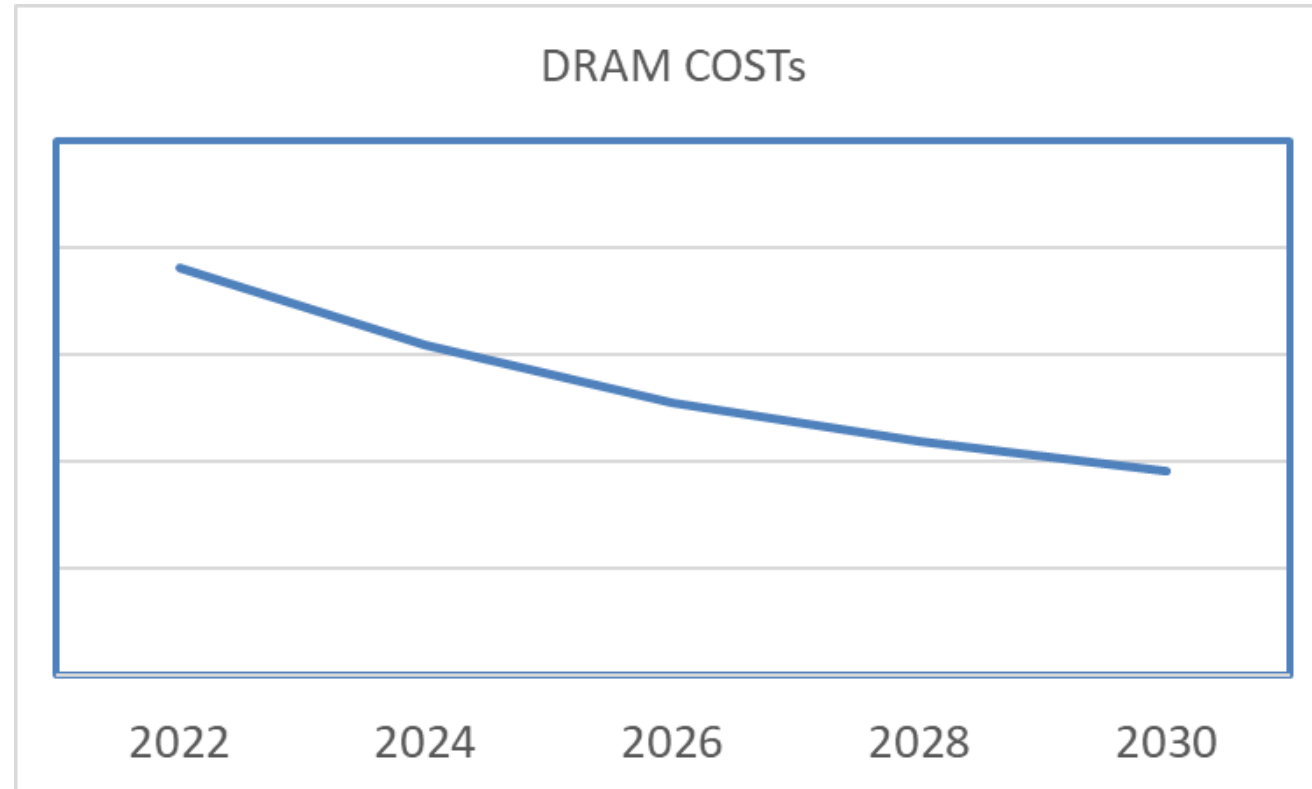


- DRAM Cell size and Die size reduction is slowed since D2x time frame.
- More like 15-20% area reduction per node vs 40% historical
- Wafer costs per node are increasing. Our model is ~10-15% cost increase for added tools and reduced outs per tool per node
- But each year Fabs become about 5% more efficient with existing technologies. Again, no inflation incorporated
- The cost reduction is so tight, that node ramps are delayed/less predictable based on yields and wafer cost improvements.
- Companies could end up running 4 nodes Not all for all products

DRAM COSTS AND MODELING

- DRAM can continue to scale in current architecture and will use “tricks” to enable scaling past 1g (see Techinsights).
- Due to small cost reduction, new nodes will only ramp when stable and providing steady cost reduction (New Normal)
 - May be one product type or density at start
 - PRAGMATIC decisions. Decisions based on performance requirements
 - 10-15% cost reduction per node... with some ups and downs based on incremental or large process/equipment changes
 - With 25% cost reduction, missing on yield or wafer cost is OK. Not now.
 - One company's product on 1a could be cheaper than another company's product on 1b
 - Look at die size but also complexity (EUV, Process integration).
- Key is that cost reduction and density increases will continue, slower pace

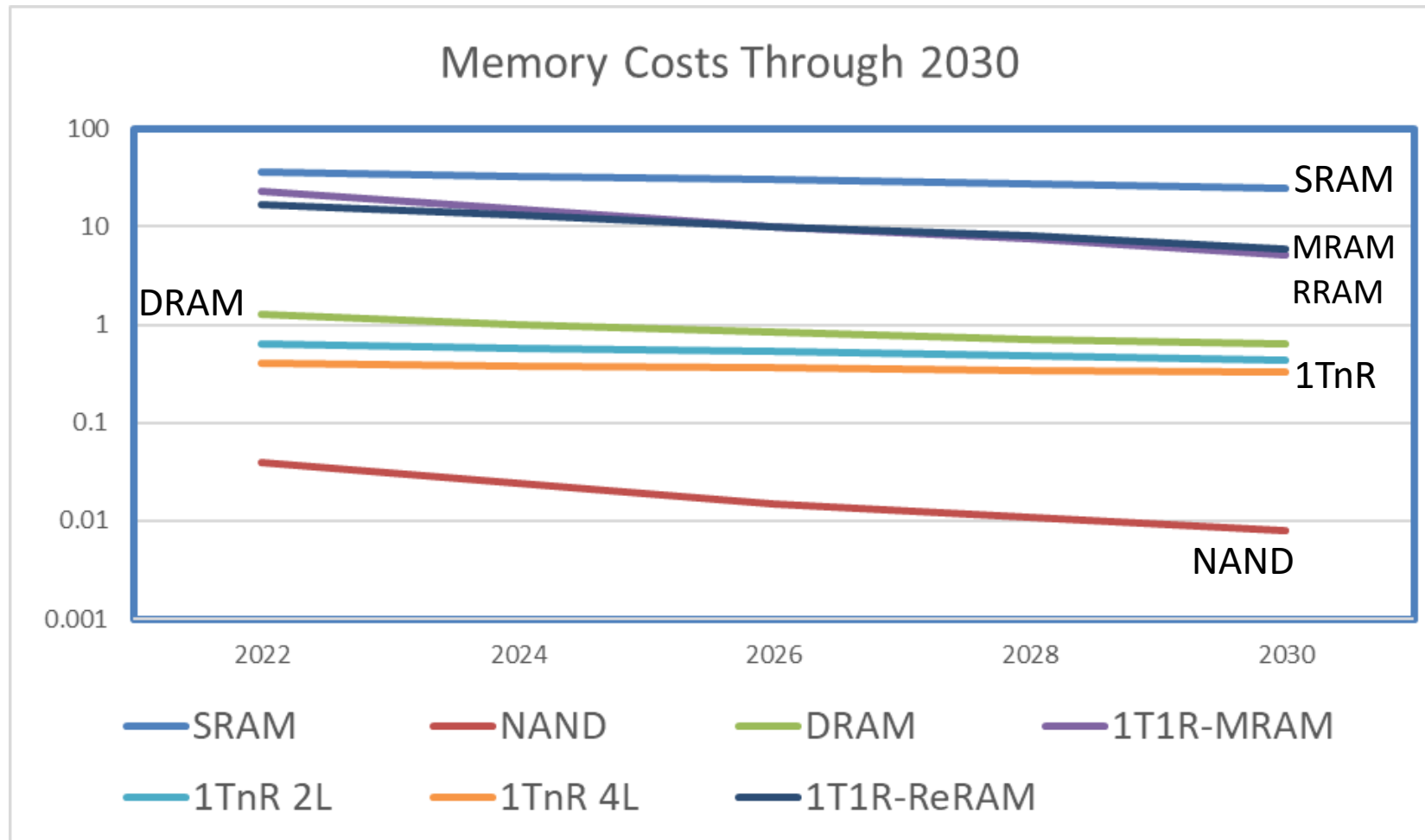
DRAM Bit Cost Extrapolation to 2030



Takeaway: Costs continue to decrease for next 8-10 years at a modest rate of 6-10% per year.

- **1T1R Technologies are usually limited by drive transistor in size**
 - Technology can improve this but for embedded and niche applications it is not worth risk from reliability issues
 - Optimizing for retention or speed vs cell size is happening now on MRAM.
 - Fairly large tradeoffs have been shown for MRAM Scaling ((shows maturity)
 - Cost are still very useful for embedded technologies and SRAM Replacement
 - eFLASH should be replaced with embedded MRAM/ReRAM across the board
- **Only 1TnR in production is Optane. Other Crosspoint technologies would follow**
 - Optane not modeled to have any future cost reduction or scaling as no new technology has been forecast by Intel. All PMEM is 2 Layers, SSDs are 4 Layers (and now EOL is planned)
 - If a 1TnR ReRAM/other cell is introduced, then one would expect it to scale with lithography and adding layers.
 - This will potentially limit speed/reliability
 - Cost scaling is shown for lithography only. Doubling layers (2 to 4) is modeled to add 20% to wafer cost

Cost of Memories Over Time



Conclusions



Flash Memory Summit

- **NAND will continue to scale with cost reduction at 15%+ per year**
 - NAND will always be cheapest and highest density technology
 - Pragmatic choices for layer addition based on each companies fabs
- **DRAM will continue to scale with cost reduction at 6%+ per year**
 - DRAM will continue to be dominant main/random memory
 - Pragmatic, slow ramps of new nodes
- **1T1R MRAM and ReRAM have modest cost reduction**
 - well suited/cost effective for embedded and lower density products.
 - Able to replace SRAM in LLC applications
- **1TnR costs are shown. Optane will not continue but other technologies can look at crosspoint designs (or potentially 3D NAND type designs)**

Thank You



Flash Memory Summit

- Mark Webb
- MKW Ventures Consulting LLC
- www.mkwventures.com

Mark@mkwventures.com

505-681-7614