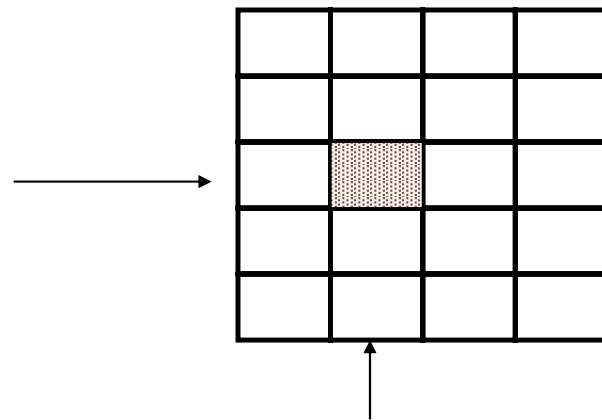


Memory Scaling: A Look Inside

MARK HOROWITZ
STANFORD UNIVERSITY

Why Memories Are Great

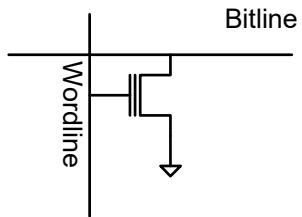
Ever wonder why memories often dominate VLSI systems?



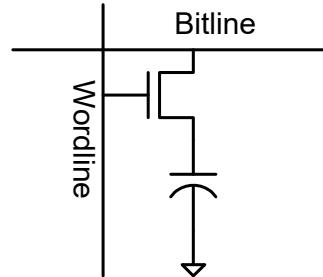
Very limited communication
&
Low activity

- Think about performance metrics
 - › Power, Performance, and Area/Cost (PPA)
 - › Can't really do better

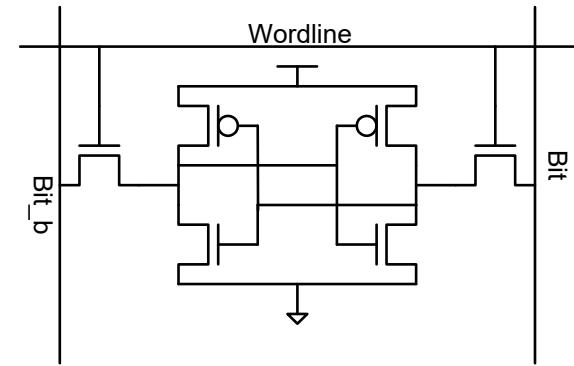
Memories: Simple



FLASH



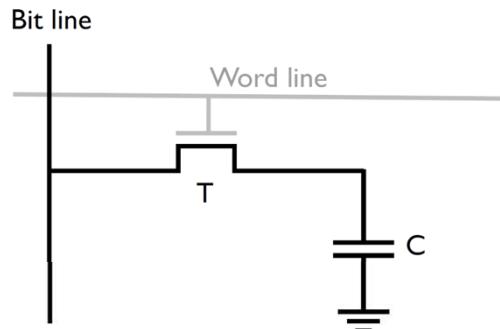
DRAM



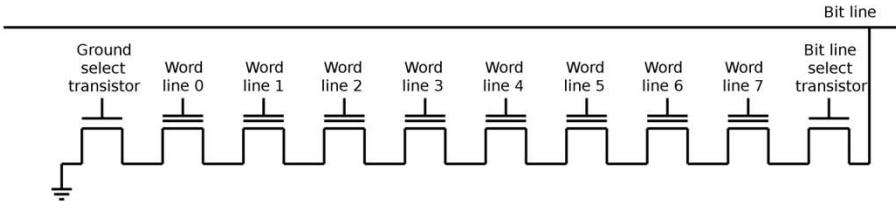
SRAM

And Complex

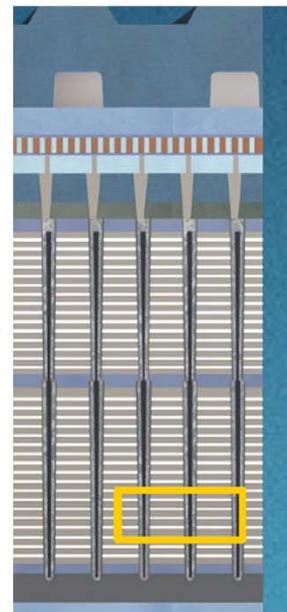
- We think



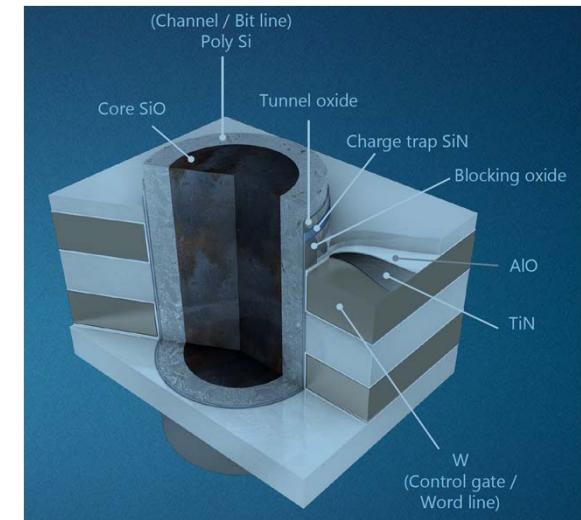
- Flash is worse



We build

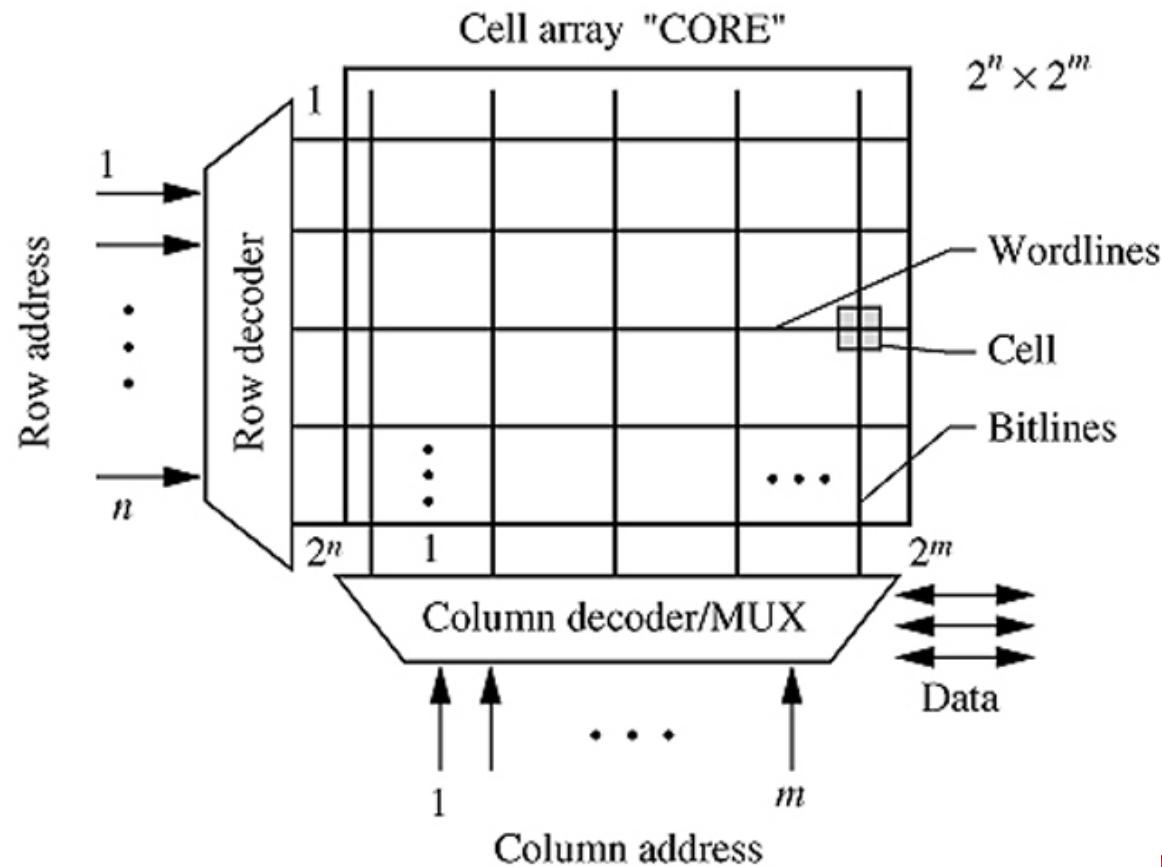


(and that is just the transistors)



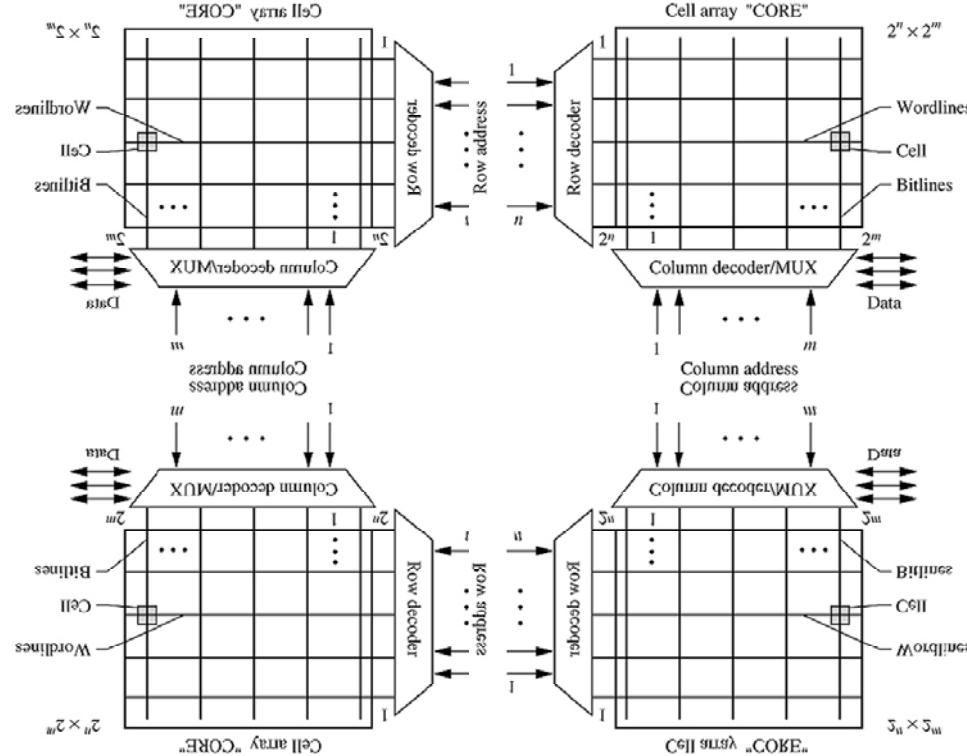
Stanford University

Organization of All Memories



- Recurse

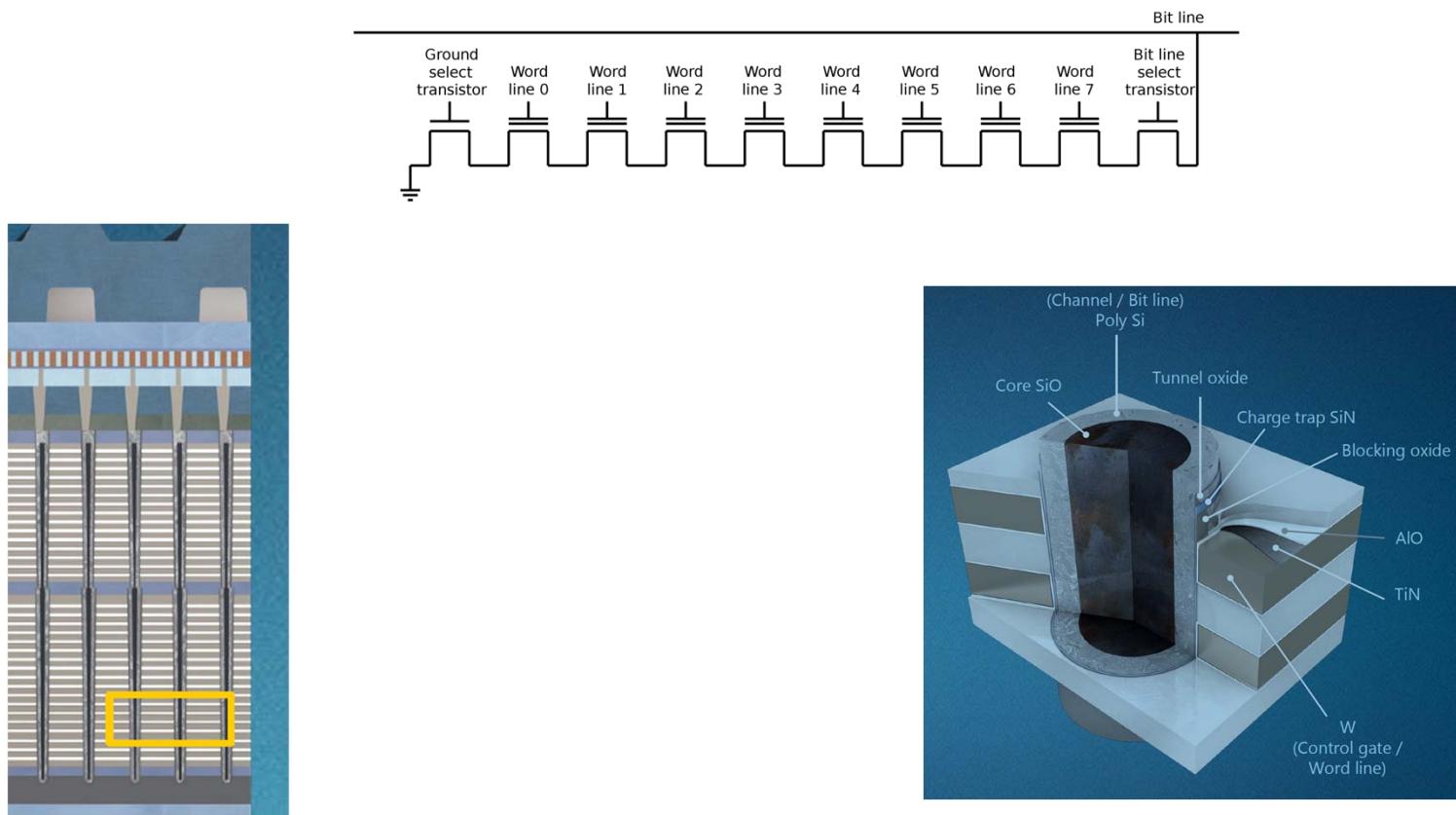
Organization of All Memories



■ Recurse

NAND FLASH

NAND FLASH

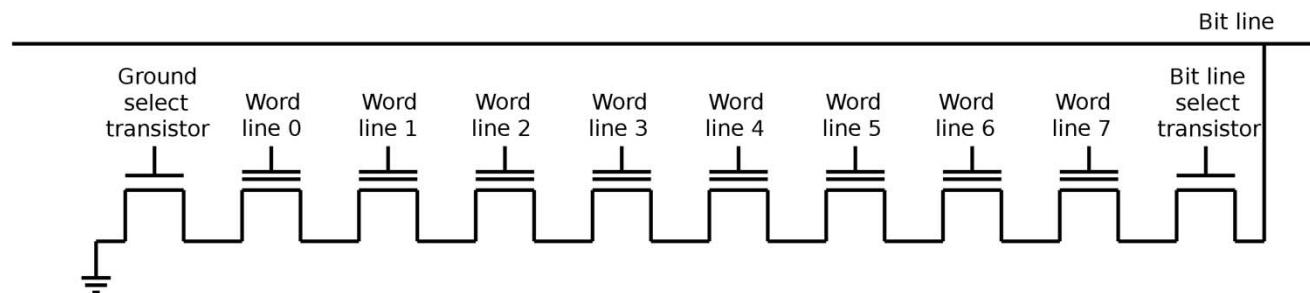
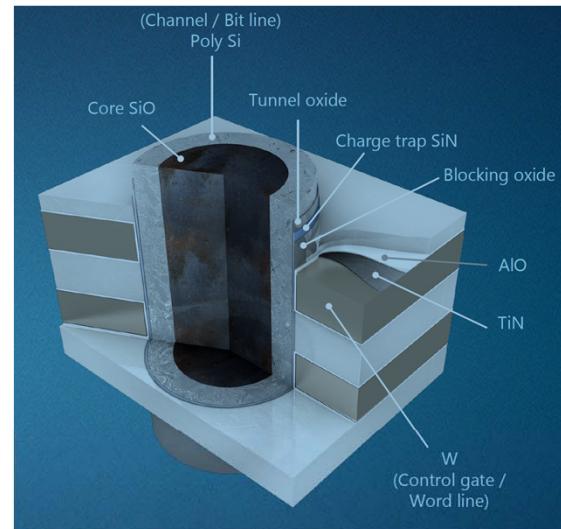


<https://ir.appliedmaterials.com/static-files/0b3c41c5-566d-4460-9e34-a01f09405c9c>

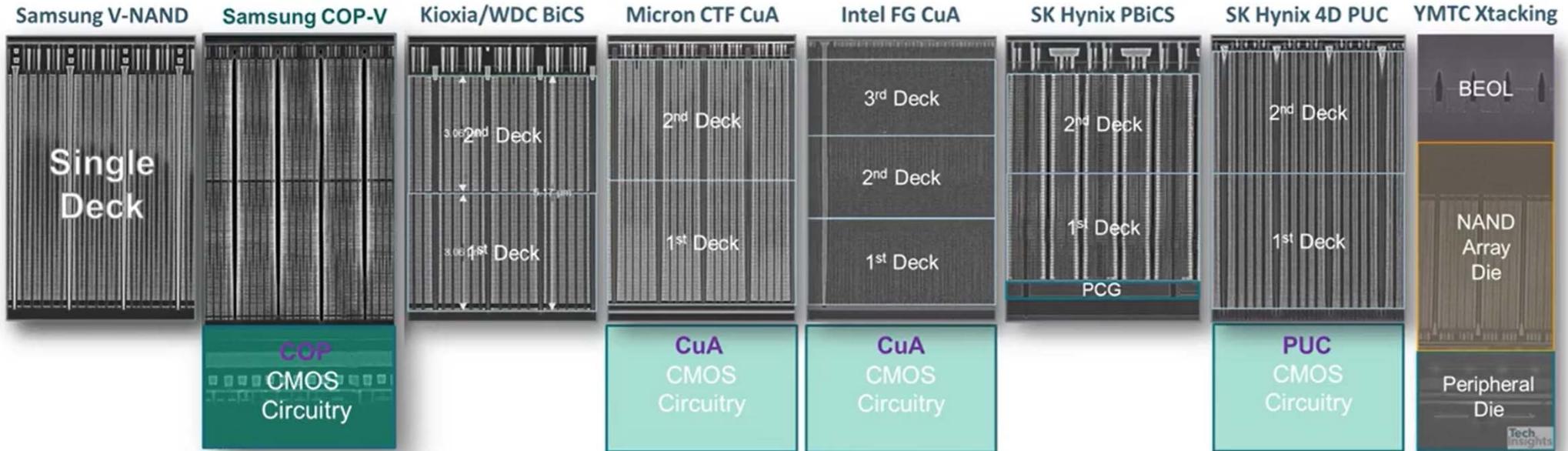
Stanford University

NAND FLASH Speed

- Not going to be fast
 - › Transistors are poly-silicon
 - › You have large number in series
 - › Reads are low current!



Memory Array Bonded on CMOS Circuits

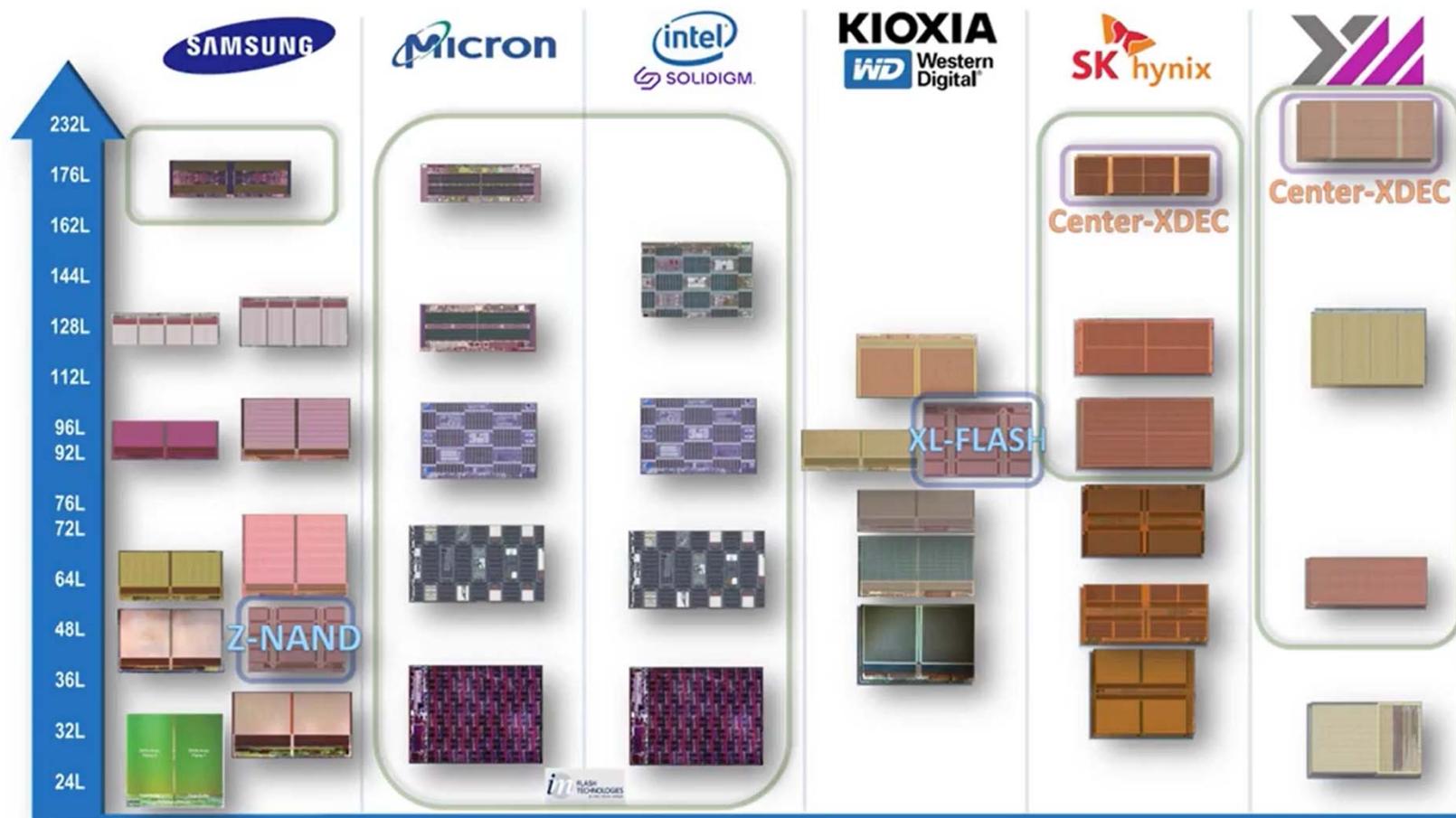


<https://doi.org/10.1117/12.2658765>

- *CuA: CMOS Under Array*
- *PUC: Periphery Under Cell array*
- *COP: Cell array On Periphery*

3D NAND Die Floorplan Trend

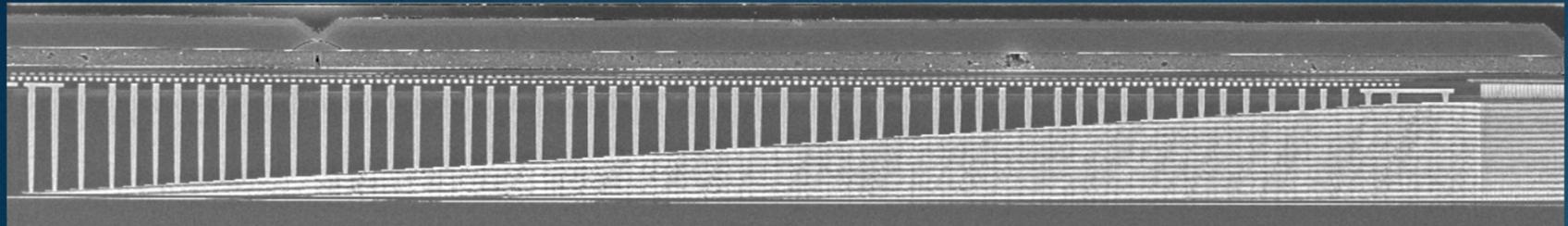
Cell-On-Periphery



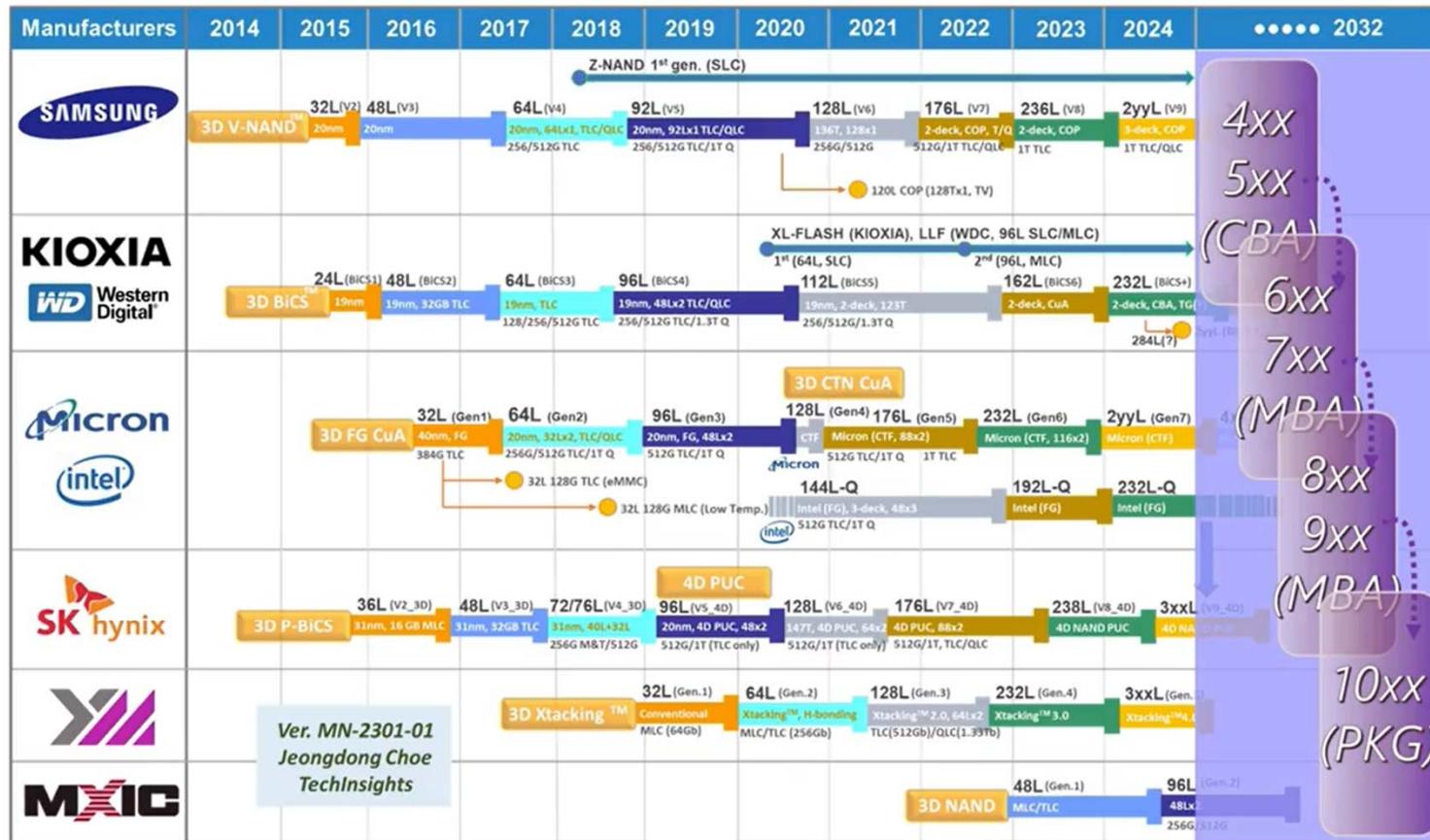
<https://doi.org/10.1117/12.2658765>

Problem with Stacking

Traditional Staircase – Single Flight

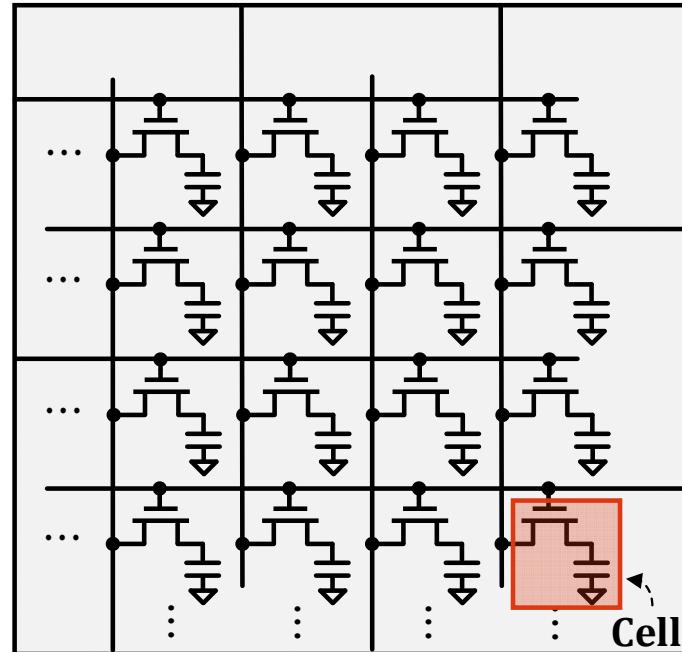


NAND Roadmap



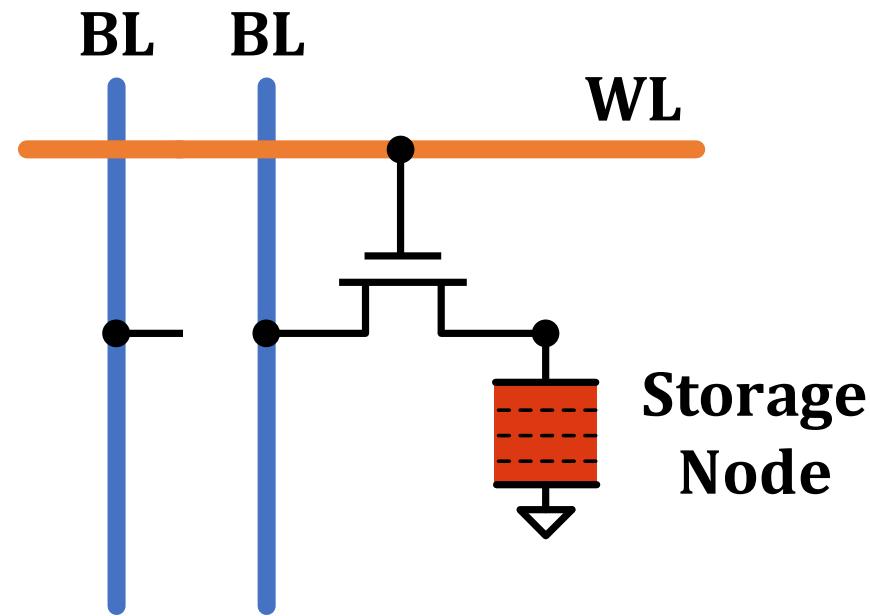
DRAM

Hierarchical DRAM Structure

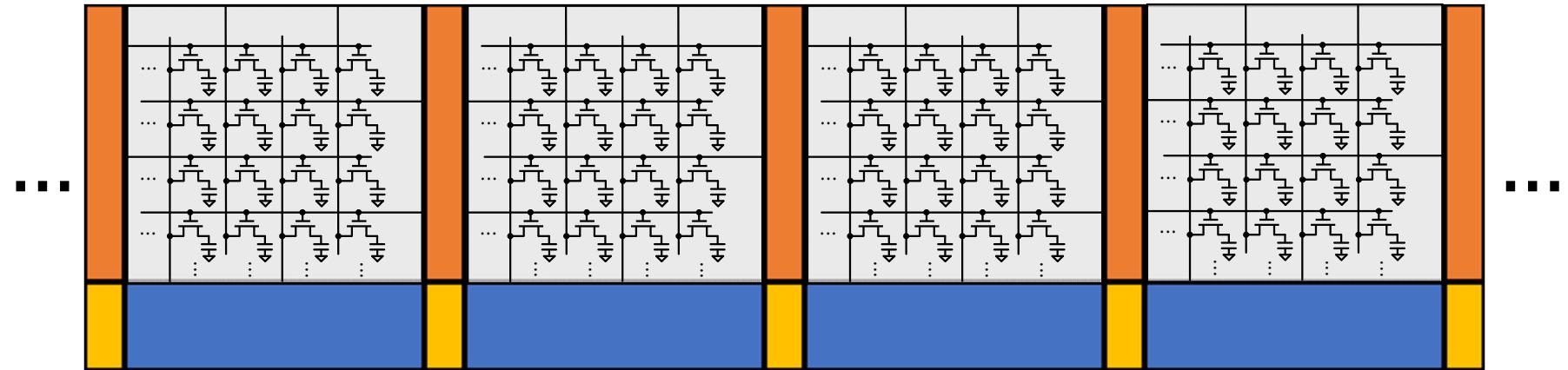


2D-matrix of cells form a *MAT*

DRAM Sensing

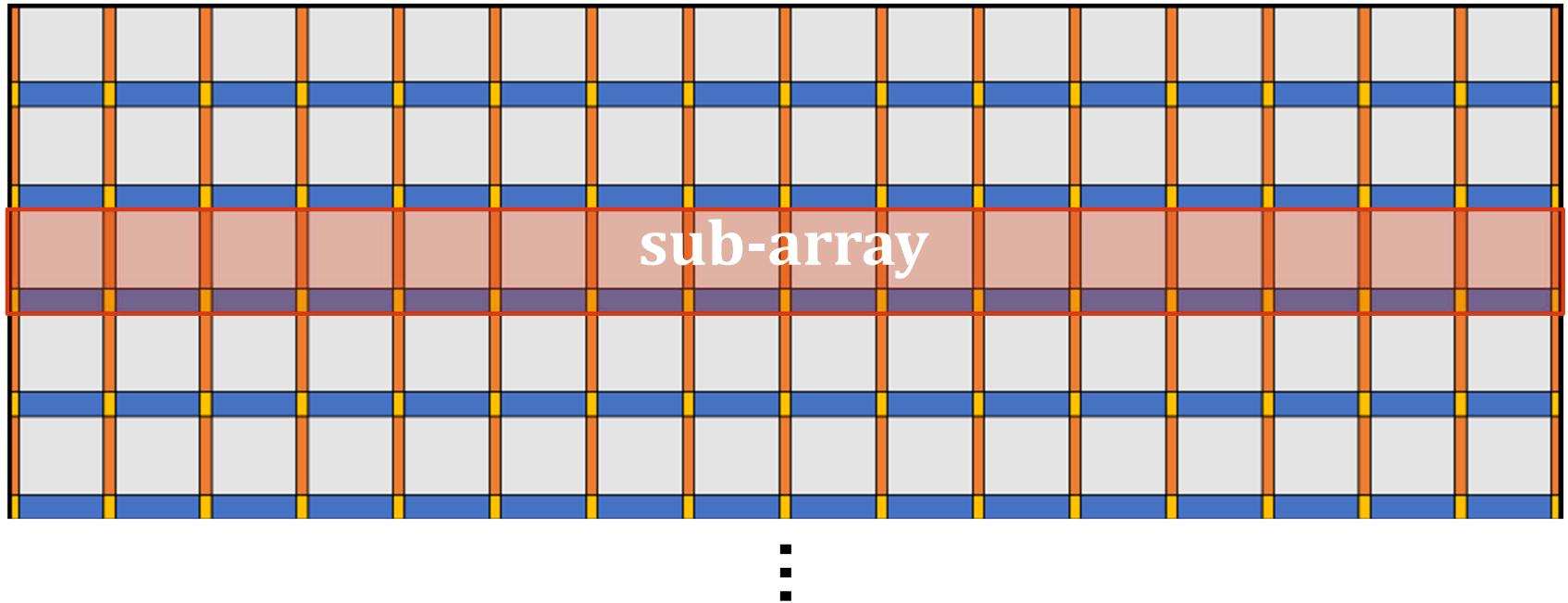


Hierarchical DRAM Structure



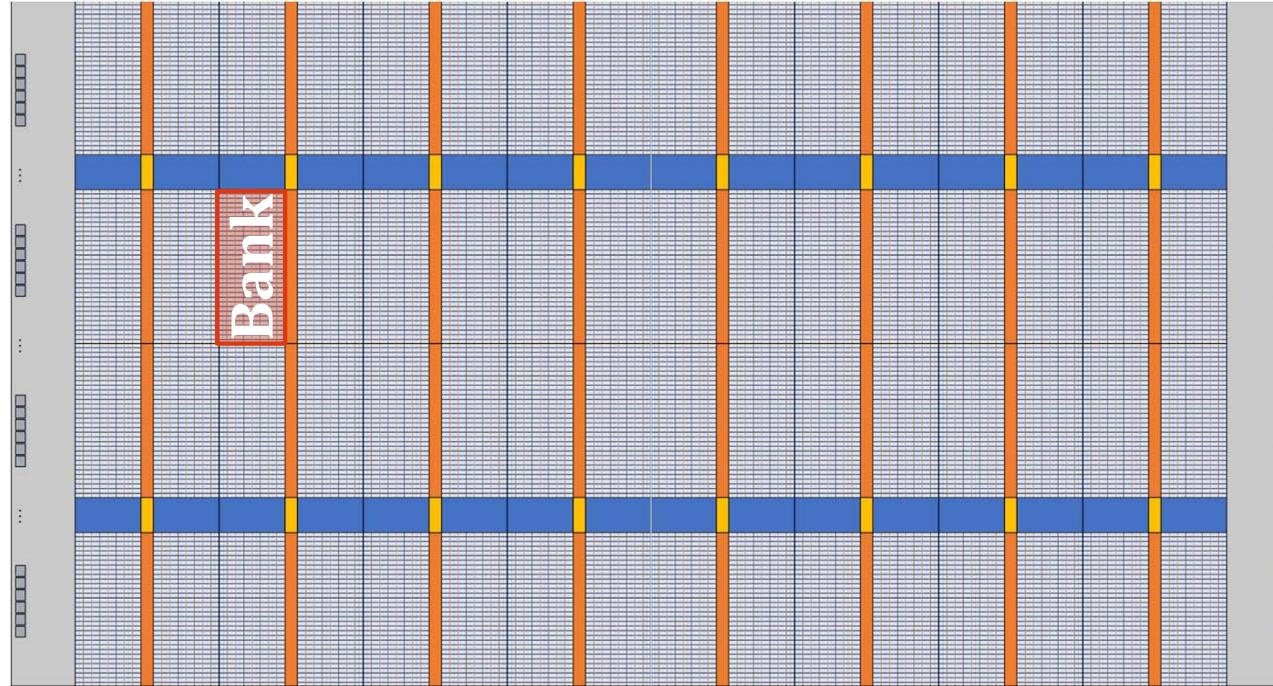
Multiple MATs are grouped into a *sub-array*

Hierarchical DRAM Structure



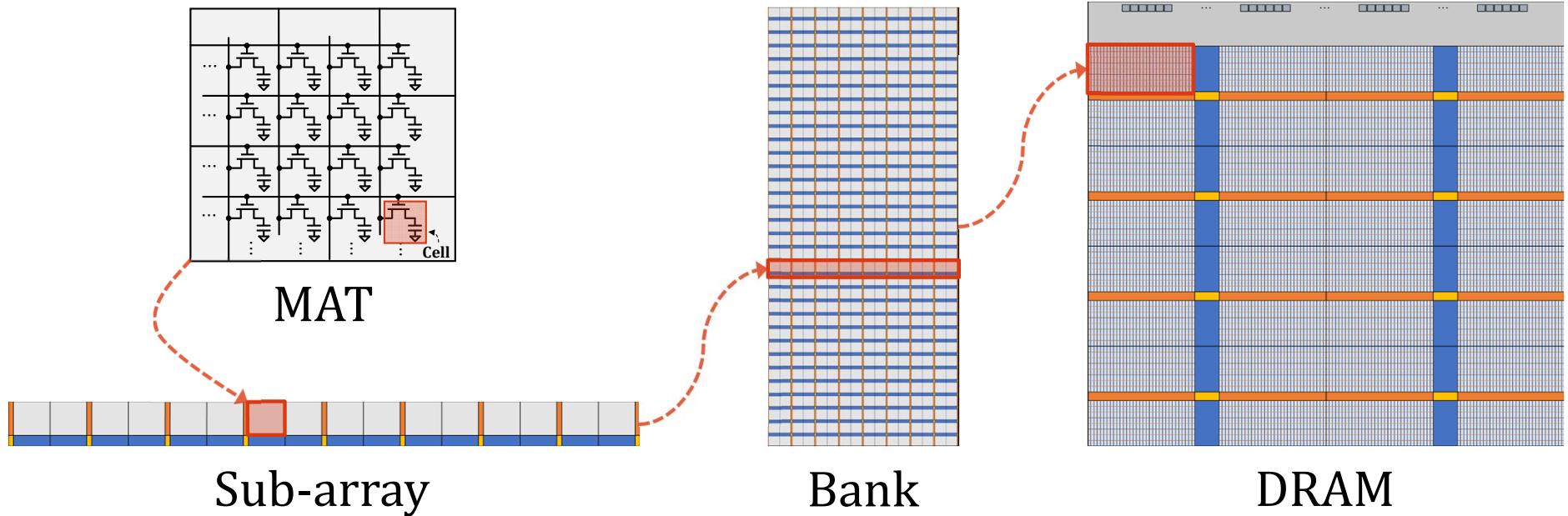
Sub-arrays are stacked on top of each other to form a ***bank***

Hierarchical DRAM Structure



Many banks are stitched with periphery circuits to form a **DRAM**

Hierarchical DRAM Structure



2y nm 8Gb LPDDR4 Modeling Result

Area Error: 1.13%

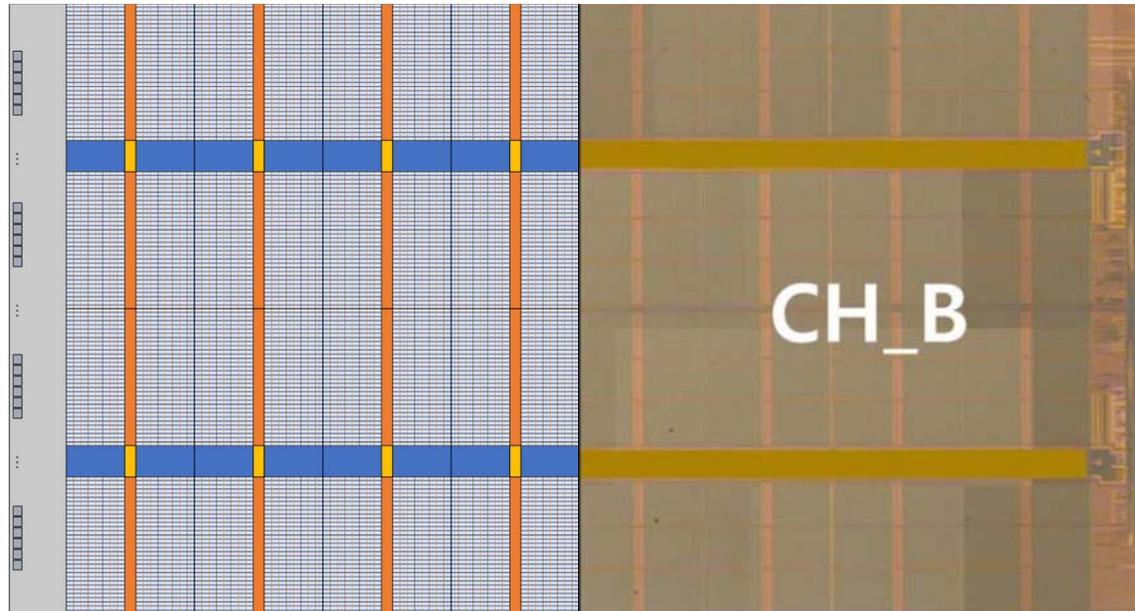


Image generated by DramDSE

Half Die photo of 2y nm 8Gb LPDDR4 by SK Hynix†

Power Validation Result

	VDD2	VDD1	Total
Row	0.08%	0.98%	1.06%
Read	2.03%	0.03%	2.06%
Write	1.50%	0.03%	1.53%
Refresh	0.50%	0.22%	0.72%

2x nm 2Gb HBM Modeling Result

Area Error: 4.74%

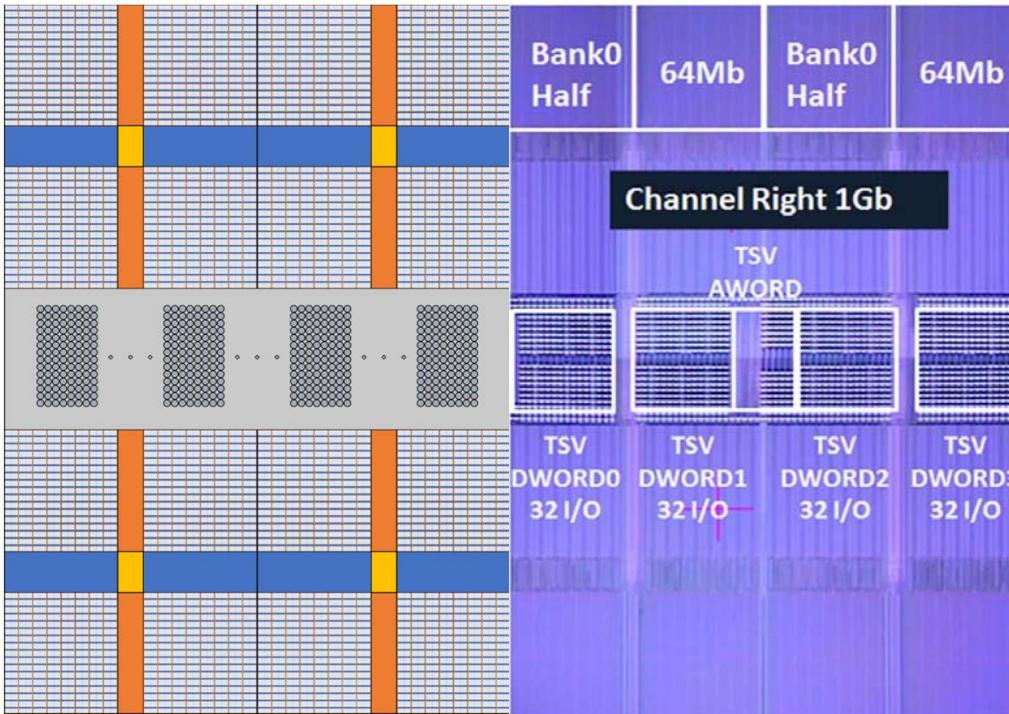


Image generated by DramDSE

Half Die photo of 2x nm 2Gb HBM by SK Hynix†

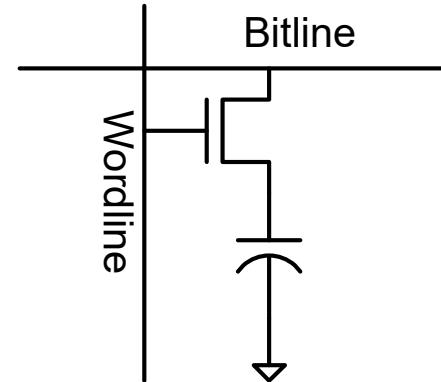
Power Validation Result

	VDD2	VDD1	Total
Row	5.61%	0.54%	6.15%
Read	0.30%	0.06%	0.36%
Write	0.39%	0.07%	0.46%
Refresh	3.61%	0.38%	3.99%

- Scaled technology parameters of 2y nm LPDDR4.

Scaling DRAMs

- Need to build:

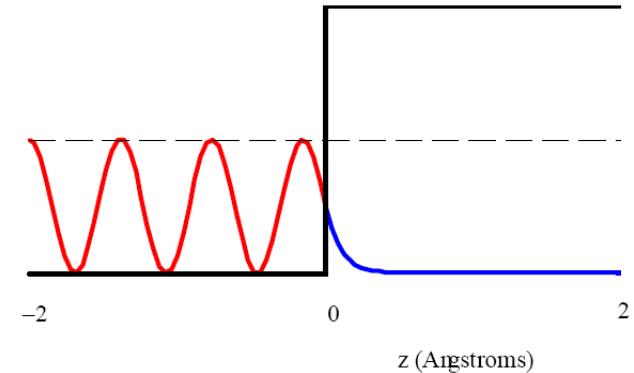


- Requirements
 - › The capacitance value to remain constant
 - › Very low leakage currents
- Scaling make both worse

Leakage

- Tunneling
 - › Limits how thin the insulators can be
 - › And how thin depletion regions can be

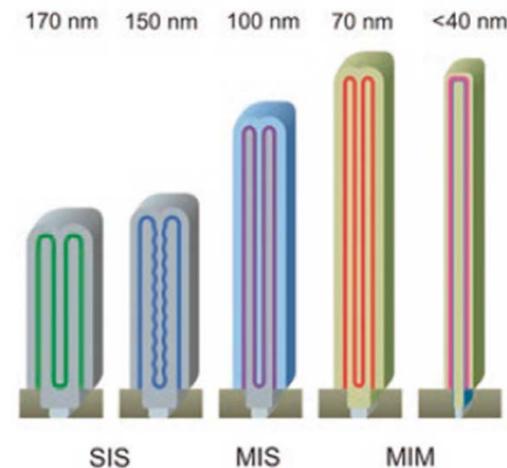
- Sub threshold leakage
 - › Need high thresholds
 - › Long channel lengths



Capacitance

$$C = \frac{\epsilon A}{t_{ox}}$$

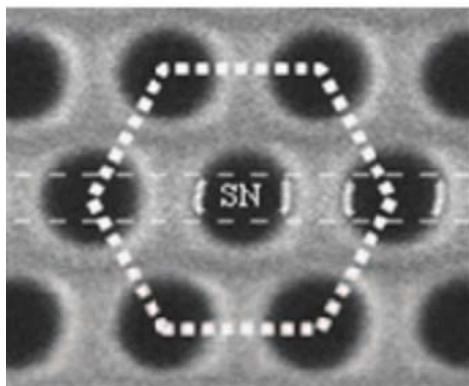
- Can't scale t_{ox}
 - › So you don't want to scale A
 - › But Z is scaling, so grow Z



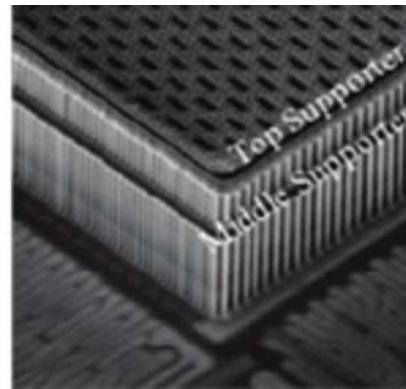
DOI: 10.1557/mrs.2018.95

Modern DRAM Capacitors

- Have aspect ratios of about 30
 - › 30nm size, 1 μ m tall



(a) A Plane view of honeycomb Structure



(b) A birds' eye view of honeycomb cell capacitors

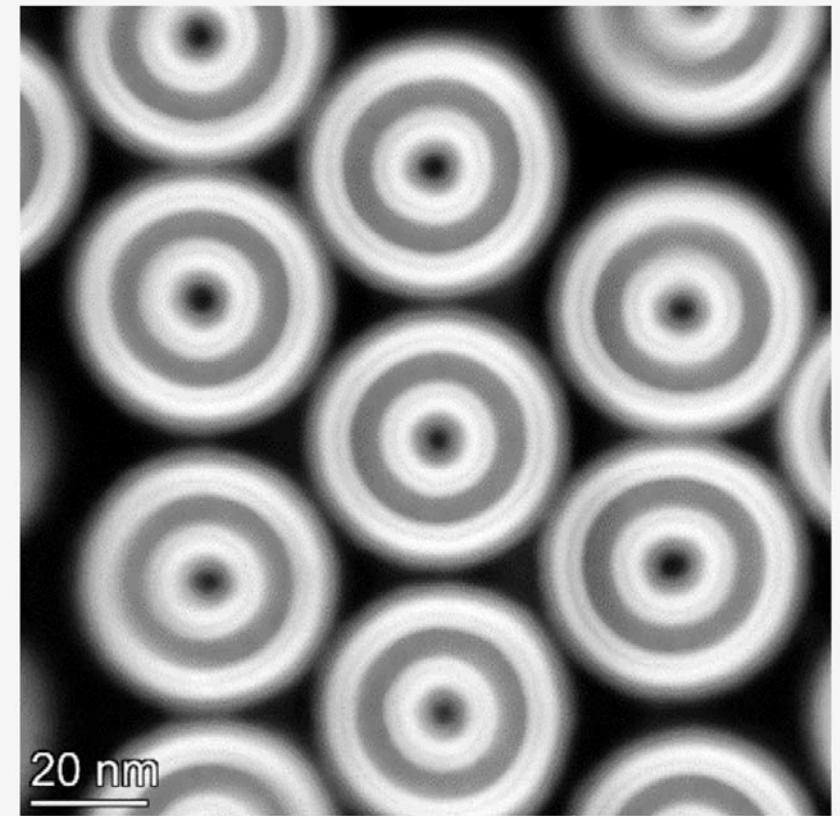
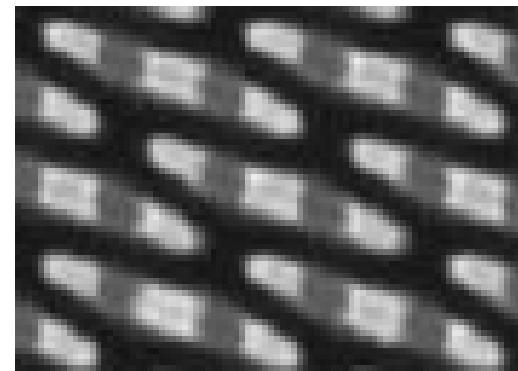
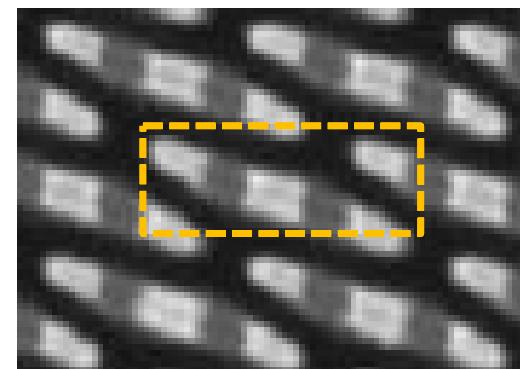


Figure 2. Scanning transmission electron microscopy image showing closely packed DRAM capacitors with a concentric layer structure.

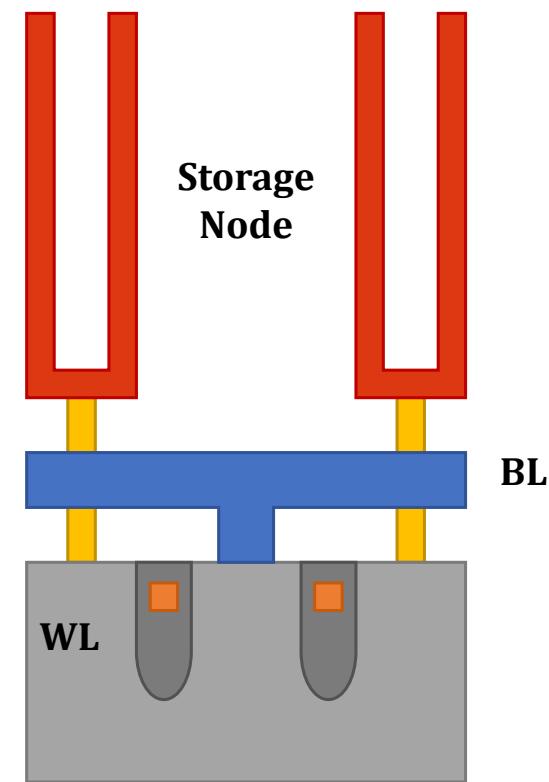
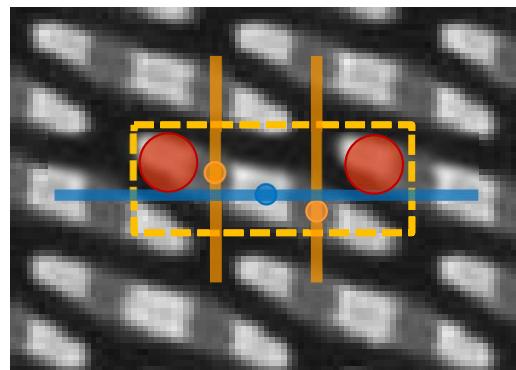
DRAM Cell Array



DRAM Cell Array

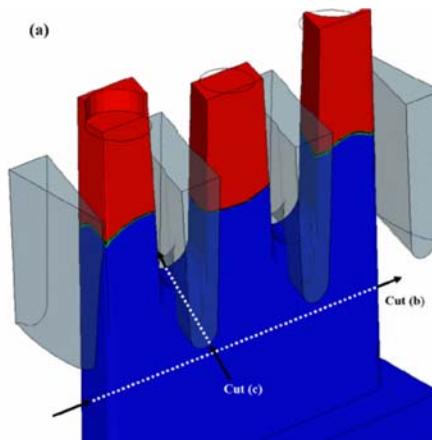


DRAM Cell Array

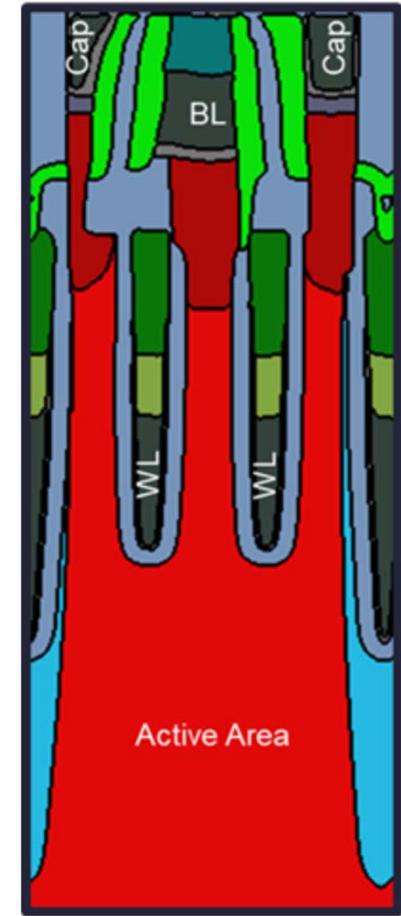


DRAM Transistors

- How to get a long transistor to shrink in size?
 - › Go into the silicon
- But you need better current control



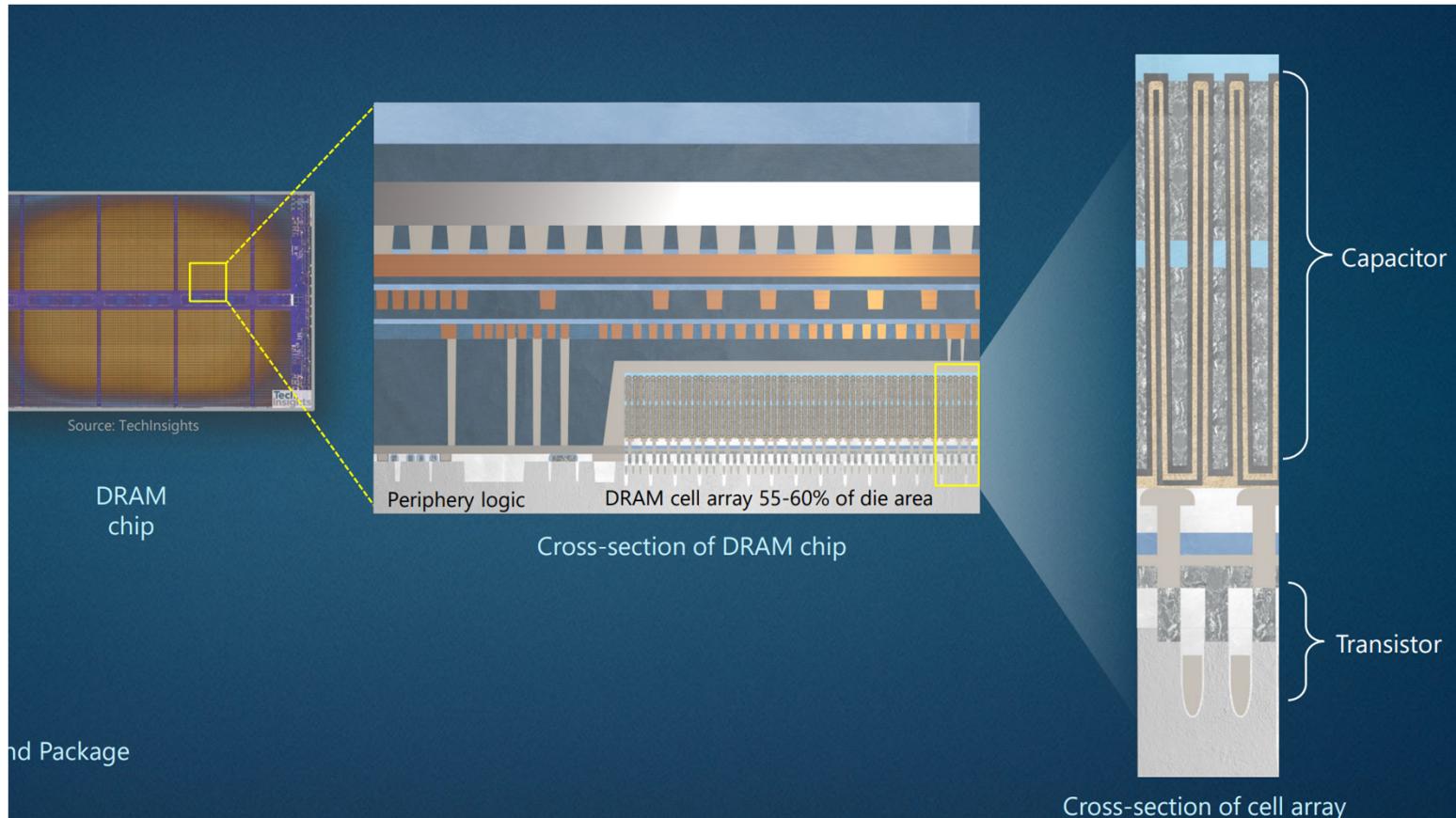
<https://ieeexplore.ieee.org/document/7150305>



<https://newsroom.lamresearch.com/3D-DRAM-architecture-proposal>

Stanford University

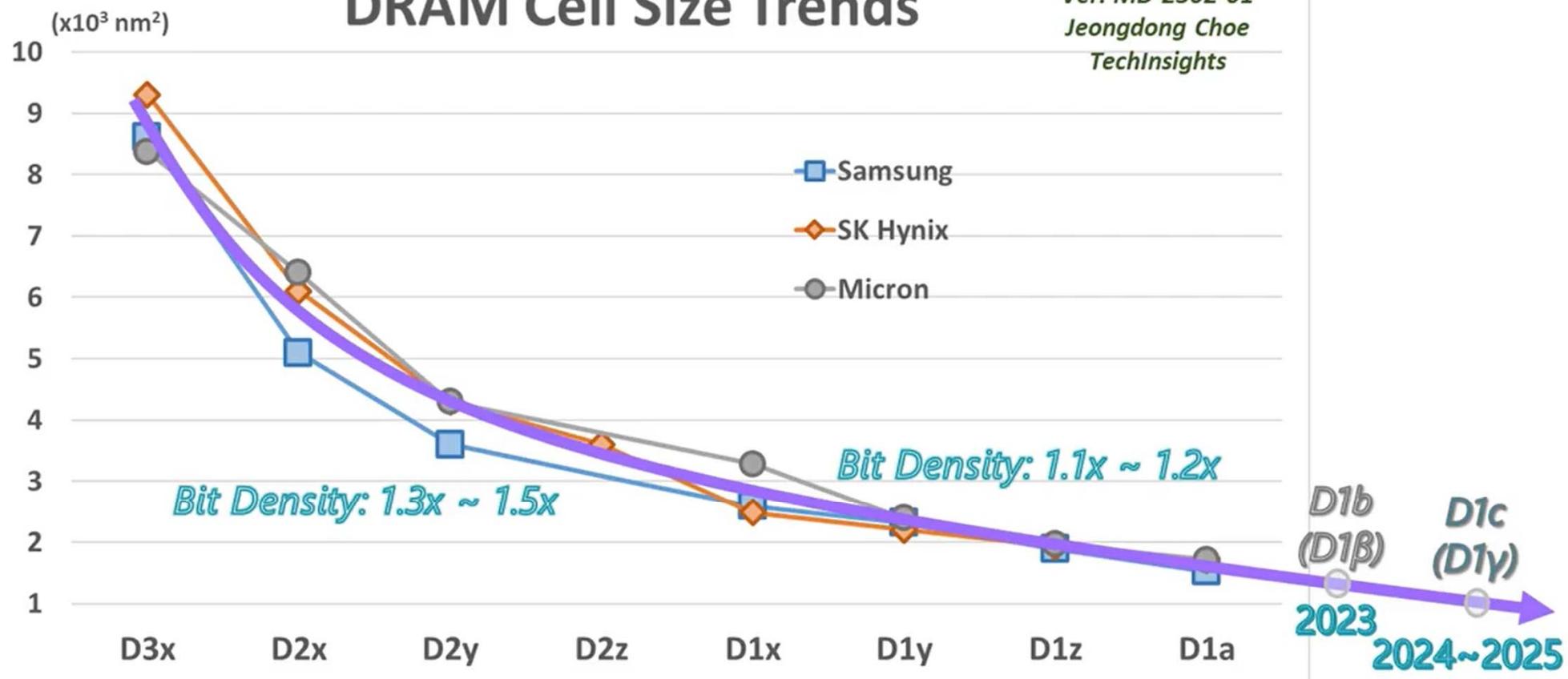
DRAM Topology



DRAM Scaling

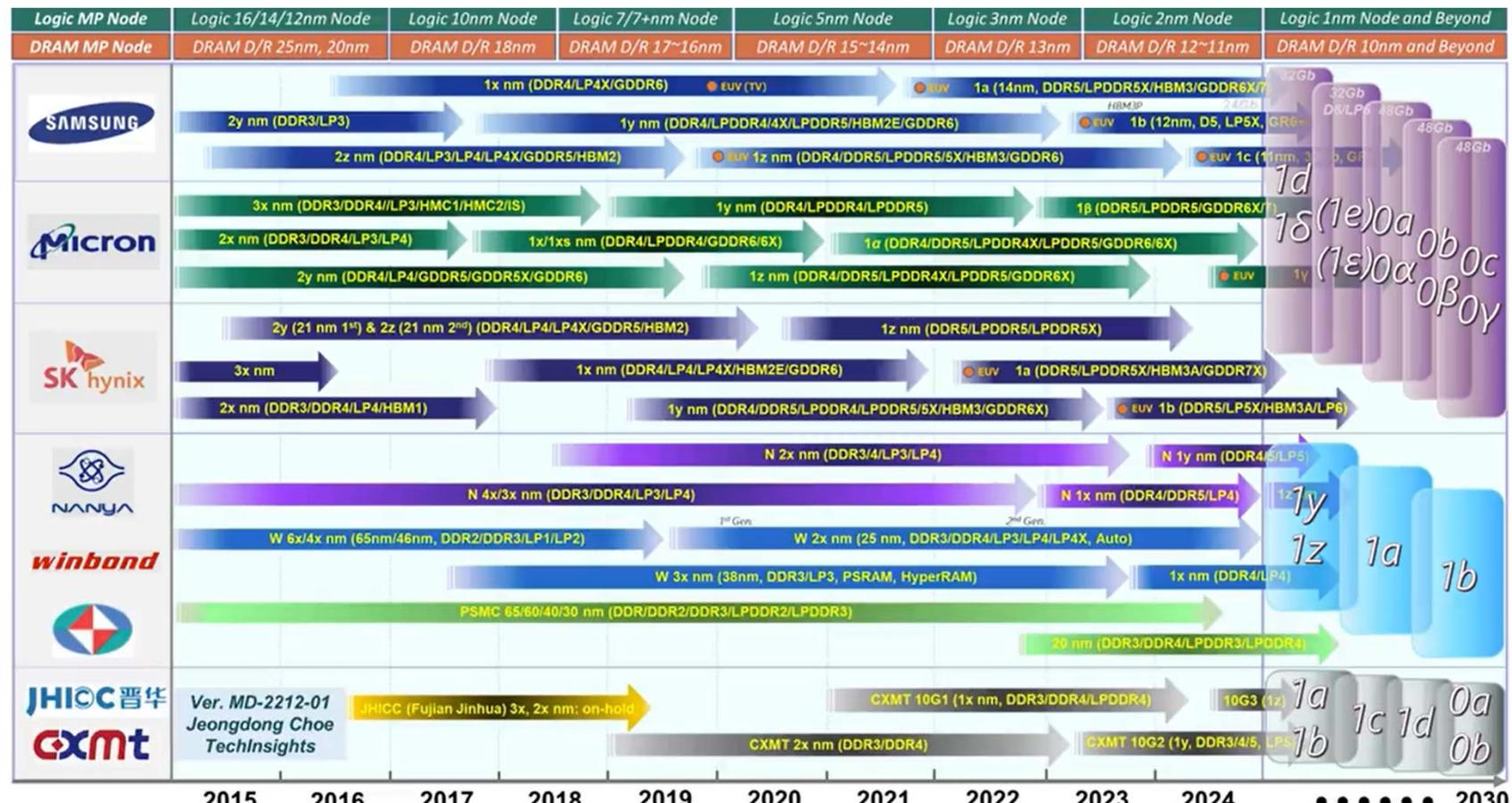
DRAM Cell Size Trends

Ver. MD-2302-01
Jeongdong Choe
TechInsights



<https://doi.org/10.11117/12.2658765>

DRAM Roadmap

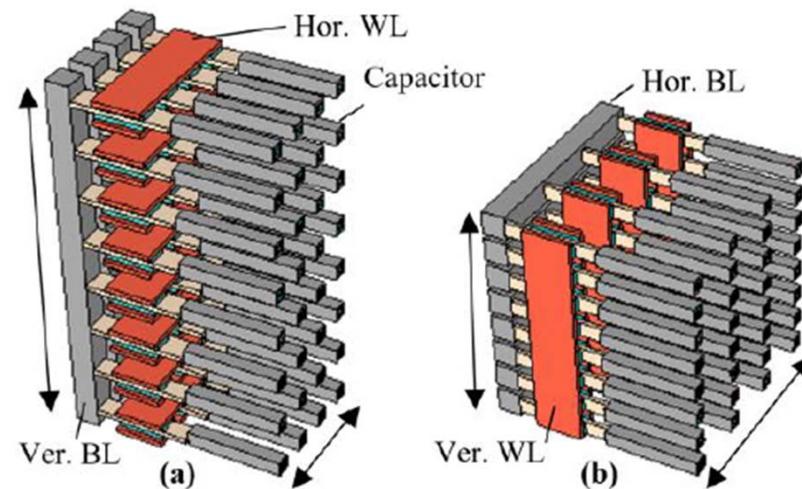


2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 ••••• 2030

<https://doi.org/10.11117/12.2658765>

3D DRAM

- People are discussing it
- But:
 - › Cell density drops by >30x
 - › Transistors need very low leakage



<https://ieeexplore.ieee.org/document/10185290>

Opportunities: Use Memories in New Ways

- Embedded DRAM is not feasible
 - › But wafer attached DRAM is
 - › What DRAM chiplet would be the right building block?
- FLASH costs are still dropping
 - › Much cheaper than DRAM
- And keep watching for new memory devices