Enabling the Semiconductor Roadmap from a Multi-Angled Approach

June 13th, 2019

Steven Welch

Senior Director of Strategy, Advanced Product and Technology Development, Applied Materials

Introduction

Steven Welch is currently Senior Director of Strategy for Applied Materials' Advanced Product and Technology Development group. In this role, he is responsible for identifying and synthesizing key inflections in the longer-term Semiconductor technology roadmap, as well as enabling disruptive technology development and business growth strategies to intercept these emerging opportunities. With over 20 years in the industry, Steven started his career in 1998 as a photolithography engineer at IBM's Storage Systems Division in San Jose, California. Since then, he has held marketing and strategy leadership positions in KLA-Tencor's wafer inspection, Applied Materials' etch, and ASML's holistic lithography businesses. Steven holds a B.S. in Chemistry from Harvey Mudd College and MA & MBA degrees from the University of Pennsylvania's Wharton School of Business.

AGENDA

Semiconductor Scaling and Enablers

Logic Scaling Advancements

Memory Advancements to Scale A.I. Big Data

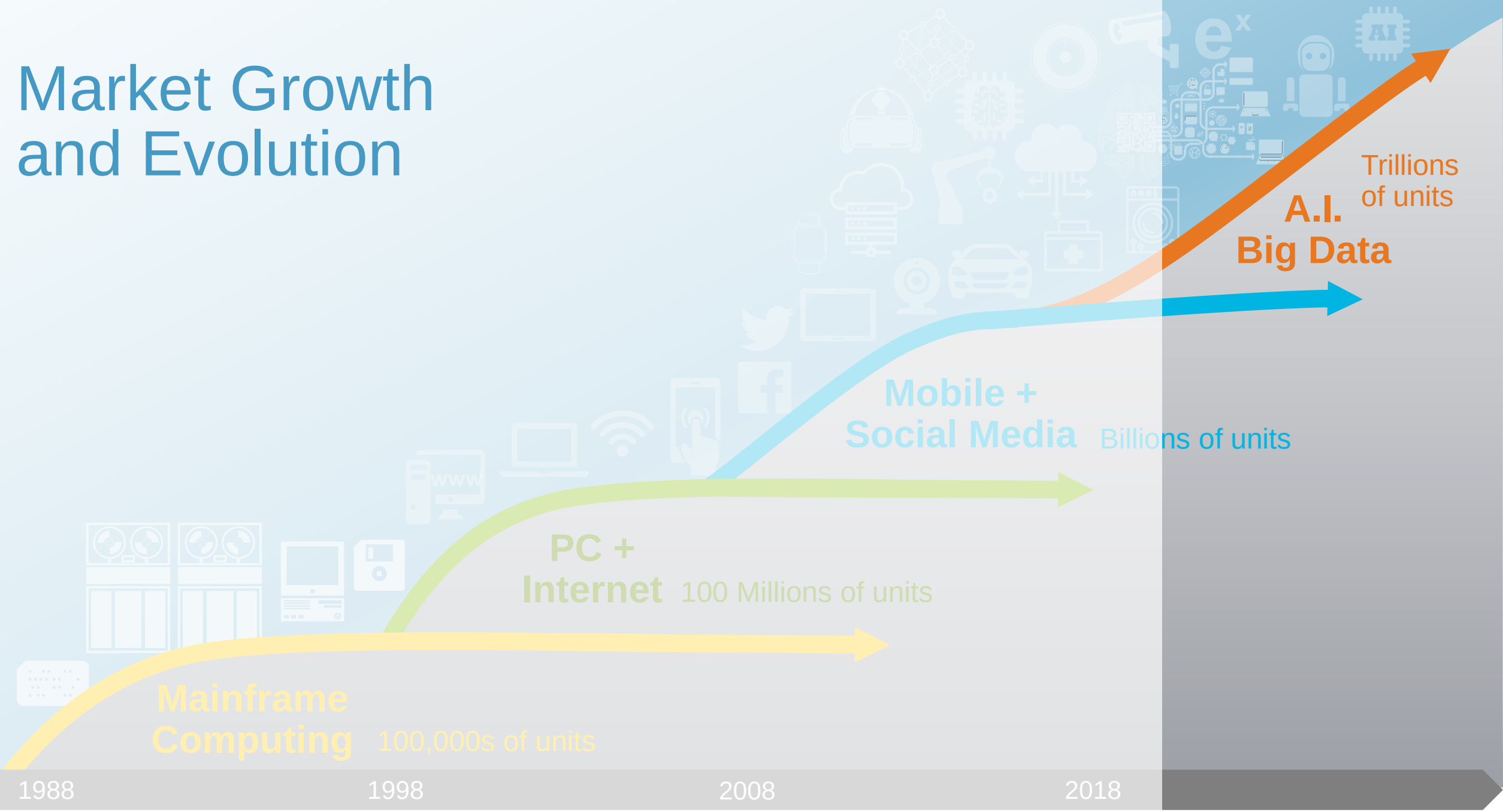
AGENDA

Semiconductor Scaling and Enablers

Logic Scaling Advancements

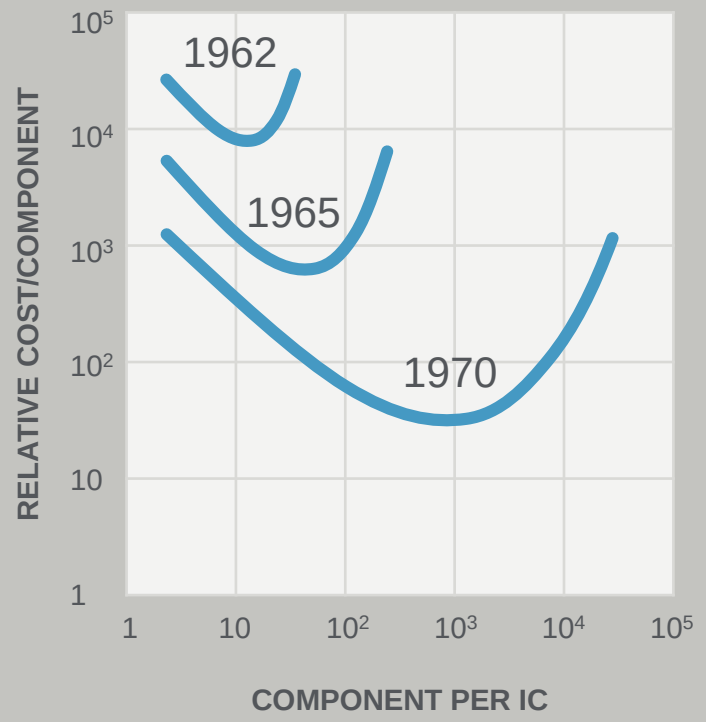
Memory Advancements to Scale A.I. Big Data

Market Growth and Evolution



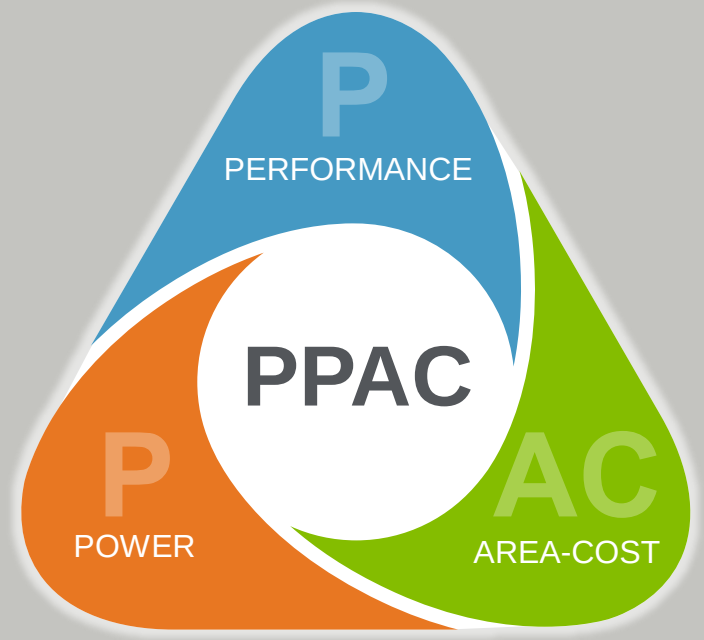
Industry's Old Playbook…

MOORE'S LAW



Now: Apple A12X Bionic
>10 billion transistors

PPAC



ENABLED BY

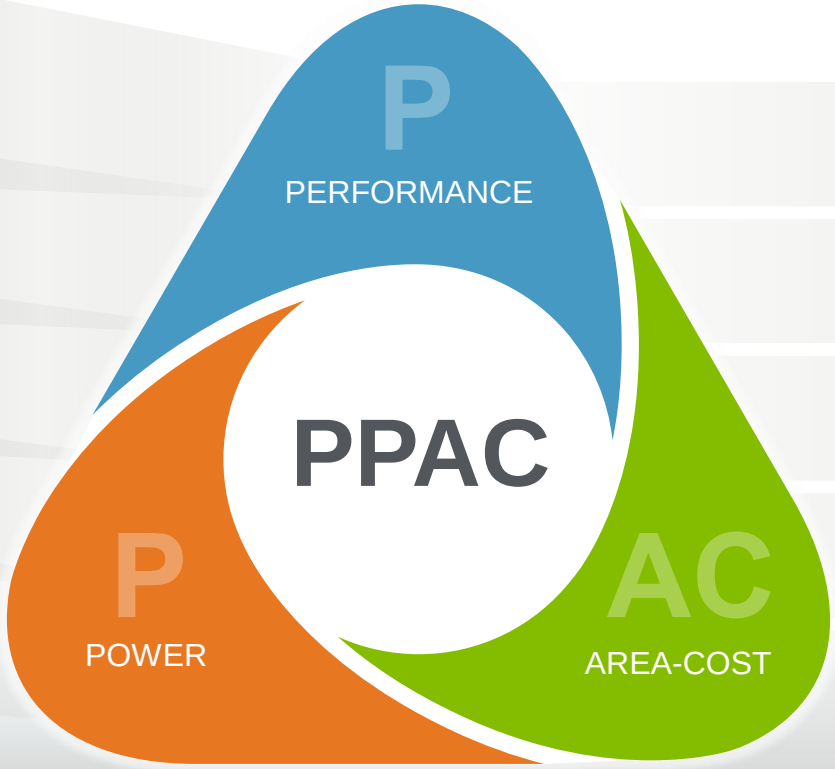
"Classic" 2D feature shrinking

+

materials engineering to drive power and performance

https://wccftech.com/apple-a12x-10-billion-transistors-performance/

In the Future…

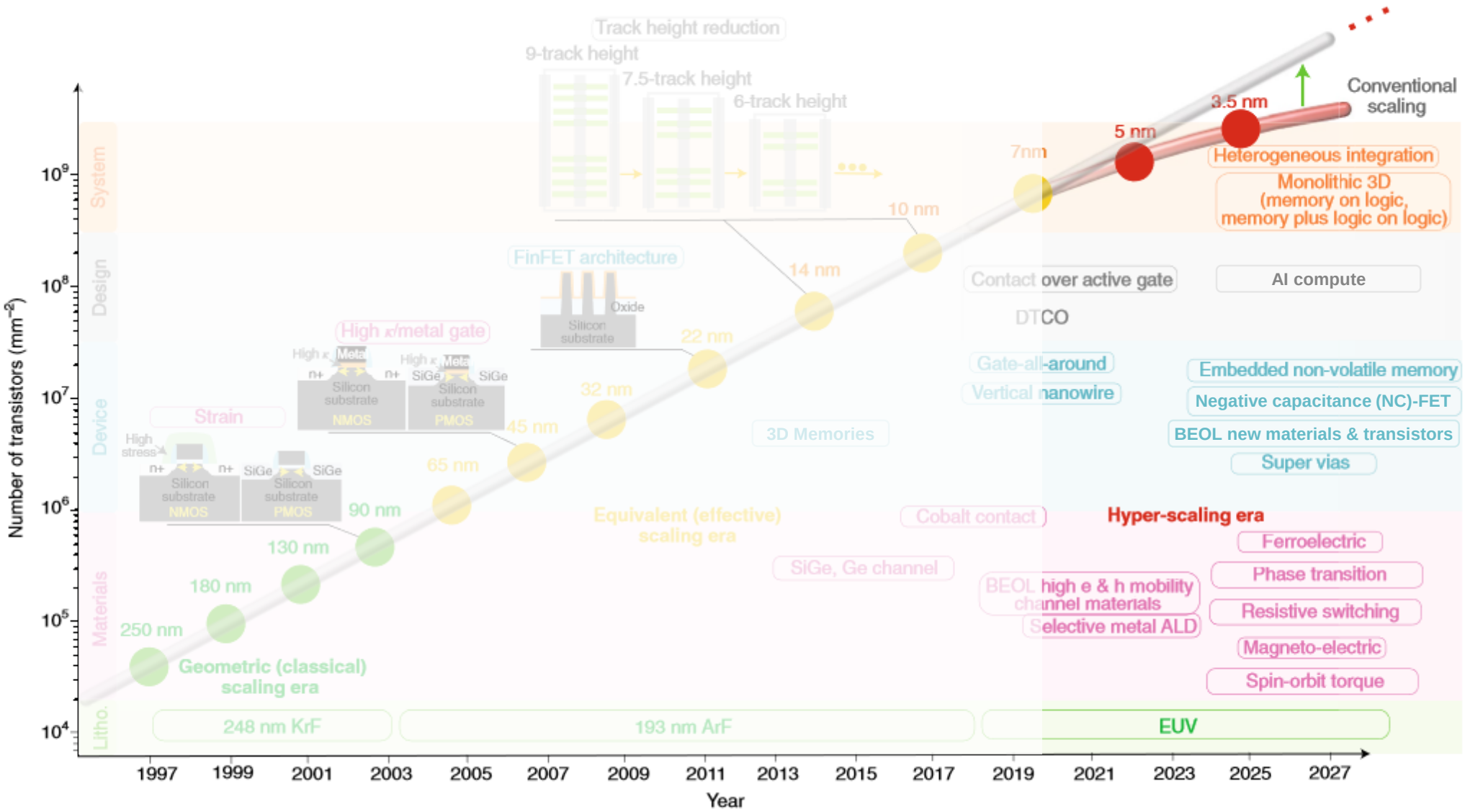


ENABLED BY

- New **architectures**
- New **structures / 3D**
- New **materials**
- New ways to **shrink**
- Advanced **packaging**

… New industry playbook needed to drive PPAC

Moore's Law beyond 5nm Node?



Adapted from Nature Electronics, V1, August' 2018, 442–450

AGENDA

Semiconductor Scaling and Enablers

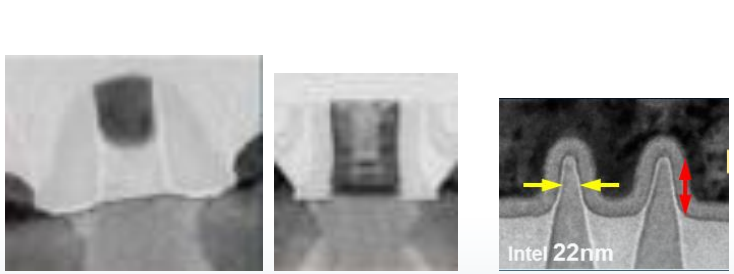
Logic Scaling Advancements

Memory Advancements to Scale A.I. Big Data

CMOS Scaling and Enablers

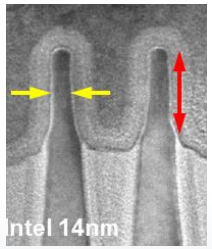
	65nm	45nm	22/14nm	10/7nm	5/3nm	3nm
Material	Strain		High-k Metal Gate		New Metal Fill	SiGe-Ch.
Architecture			FinFET → taller/ narrower & fewer fins			GAA-FET

Source: AMAT, SMC Korea 2019 conference

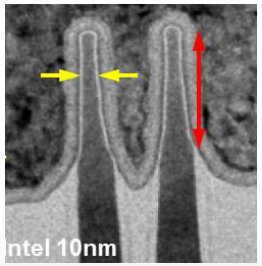


Intel, 65nm
Intel, 45nm
Research@Intel press event 2011

Intel, VLSI 2012



Intel, IEDM 2014



Intel, IEDM 2017

IBM, VLSI 2017

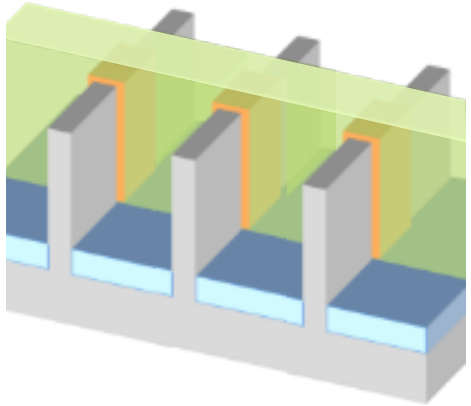
New Materials: New metal gate and new contact fill

New Architecture: FinFET to **hGAA**

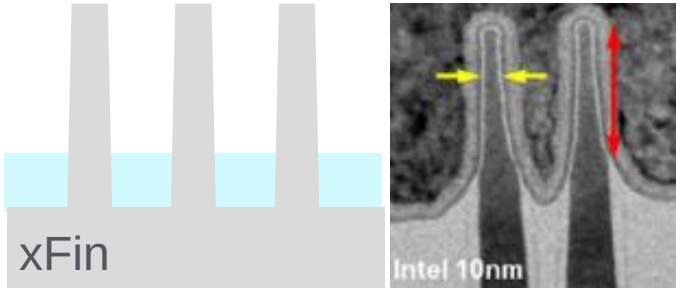
Samsung announced GAA 3nm node processor will ship in 2021

Why *Horizontal*-GAA ➔ Evolutional Change

FinFET

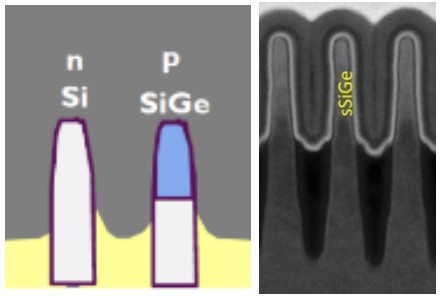


Thinner & Taller Fins
better gate control & higher drive current



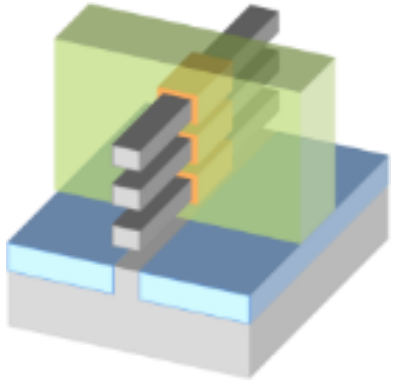
Intel, IEDM 2017

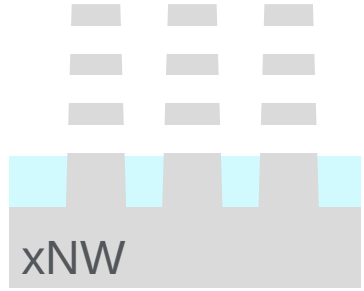
SiGe p-Channel
mobility boosting

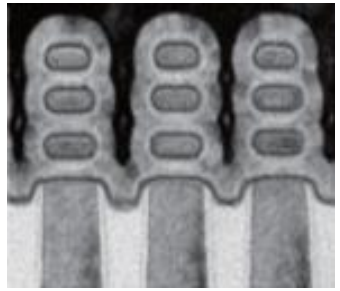


GLOBALFOUNDRIES/IBM, IEDM 2016

hGAA





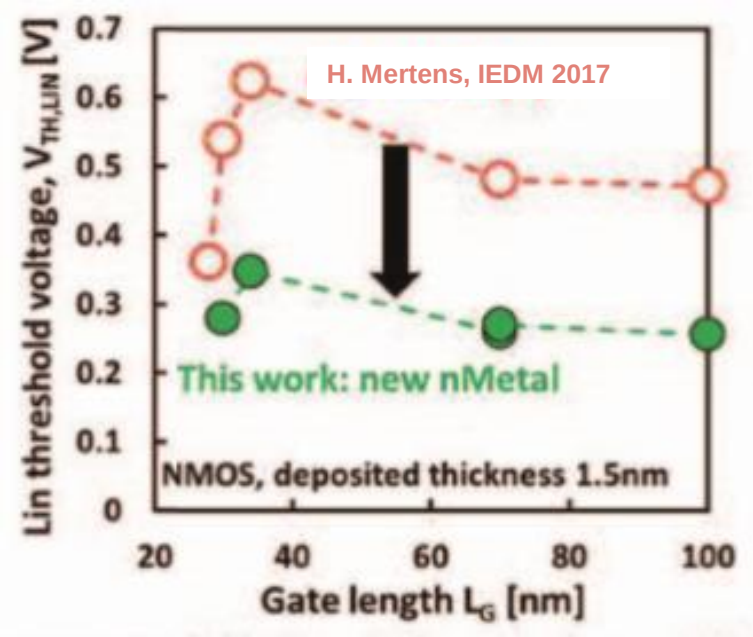


IBM, VLSI 2017

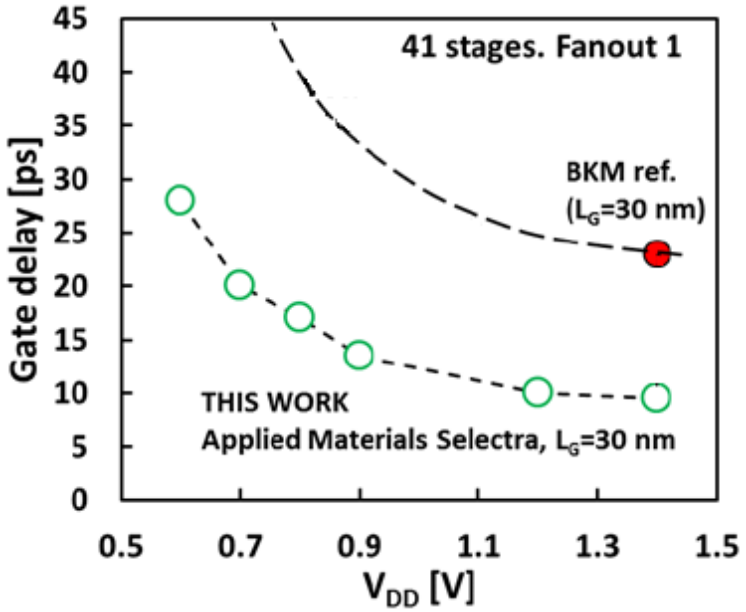
Vertically stackable channel (NW/NS) for increased current per area
Similar structures & flows: hGAA vs. FinFET

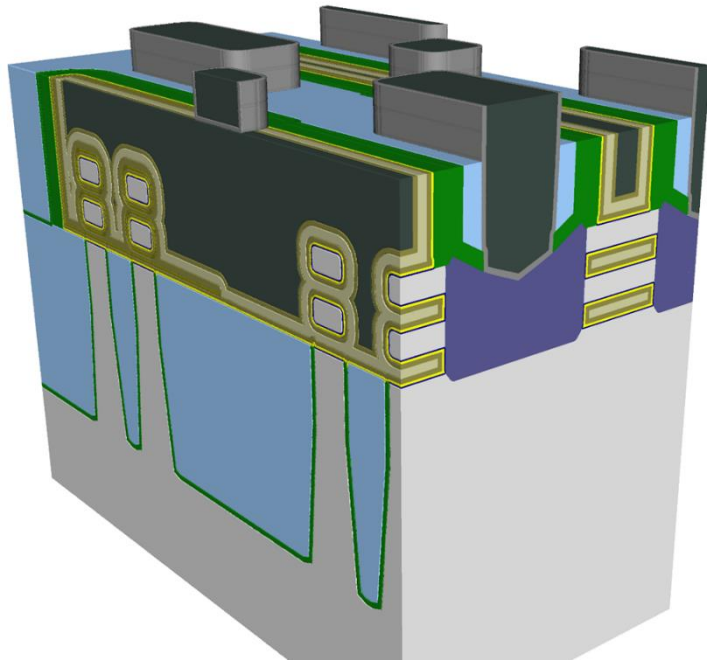
hGAA CMOS Initial Demonstrations

Transfer Characteristics



Ring Oscillator

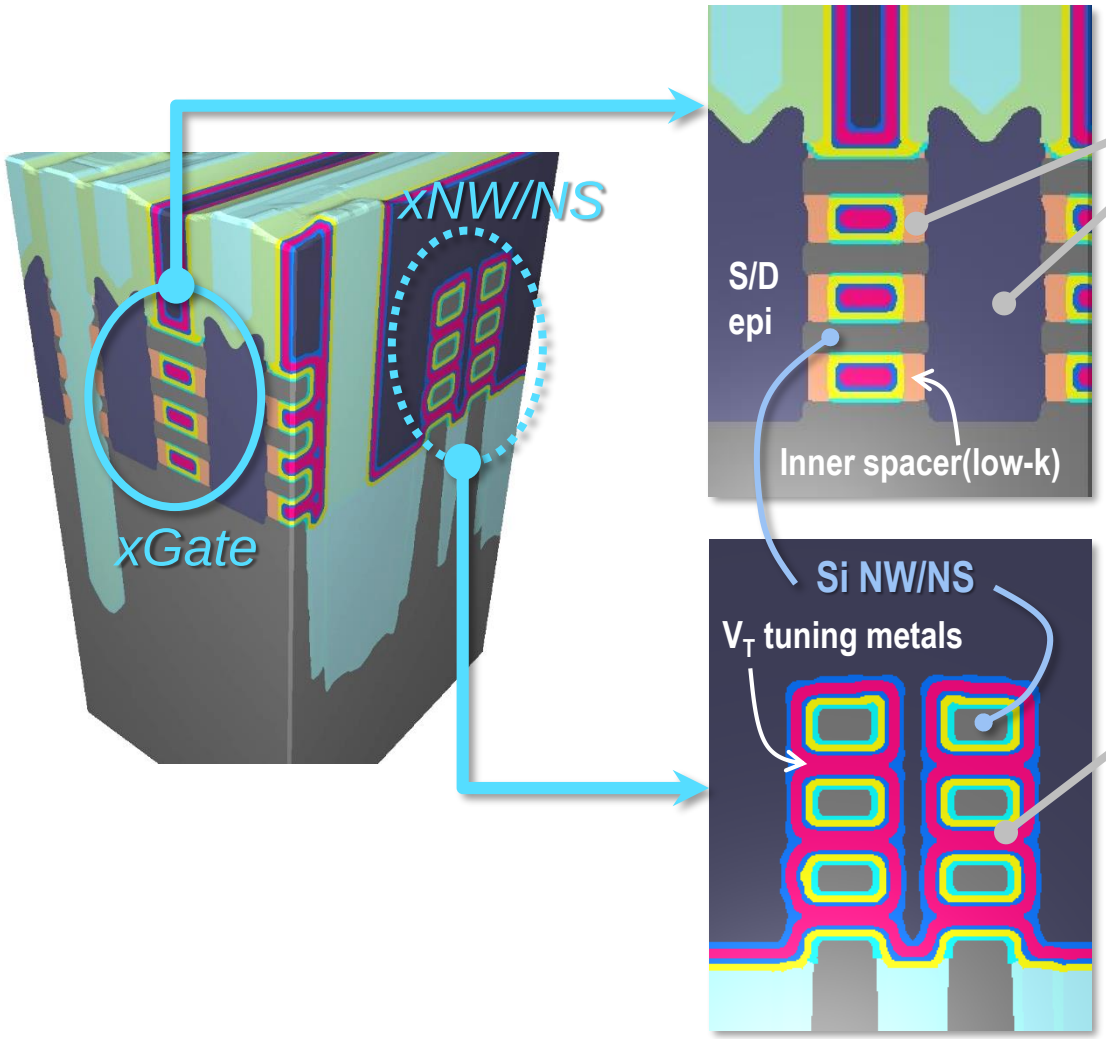




Source: IMEC/ Applied Materials, IEDM 2018

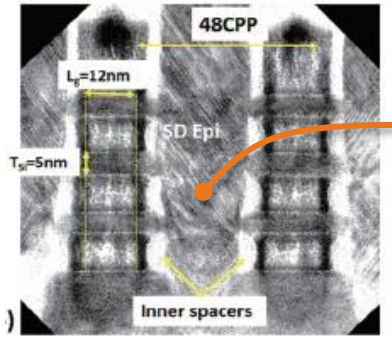
Functional CMOS demonstrated – getting ready for ≤3nm-node

hGAA CMOS – Key Challenges



Source: AMAT, SMC Korea 2019 conference

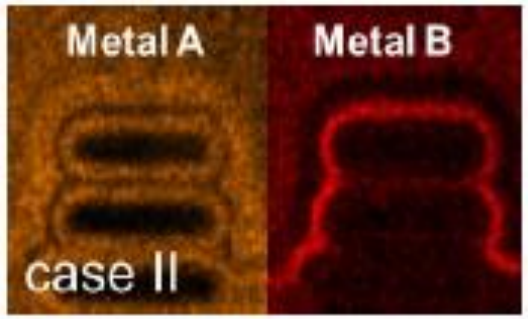
Parasitic capacitance reduction between sheets



Epi defects

IBM VLSI-T 2017

Metal-gate Vt tuning in narrow gaps



IBM, IEDM 2017

Beyond Conventional CMOS (≤2nm)

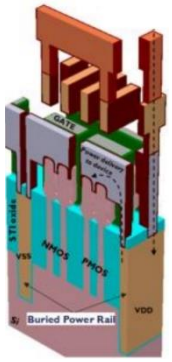
More new materials

Ge channel

Boosters

SiGe PFET

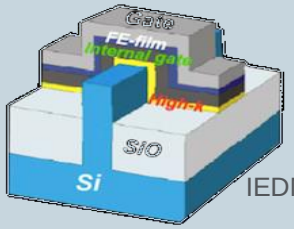
Buried Power Rail (Ru)



IMEC, IEDM, 2018

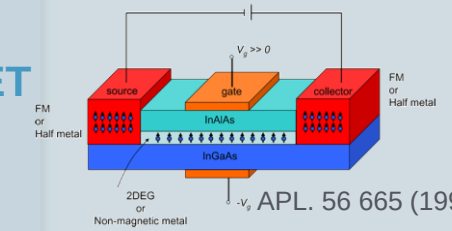
New devices

NCFET/Ferro (HZO)



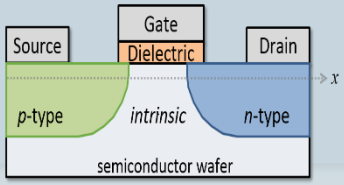
IEDM 2011

SpinFET



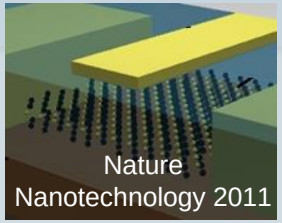
APL. 56 665 (1990)

Tunnel FET

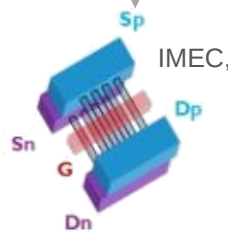


2016 ICEEOT

2D TMD (MoS_2)



Complementary FET

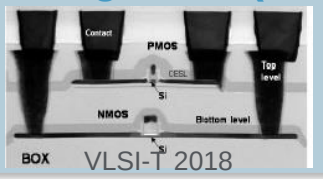


IMEC, IEDM, 2018

New scaling

Co-integration (VFET, TFT)

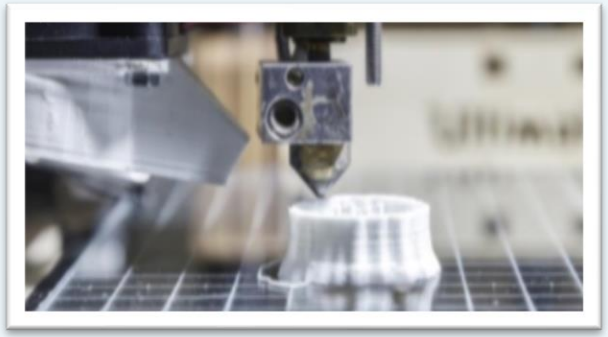
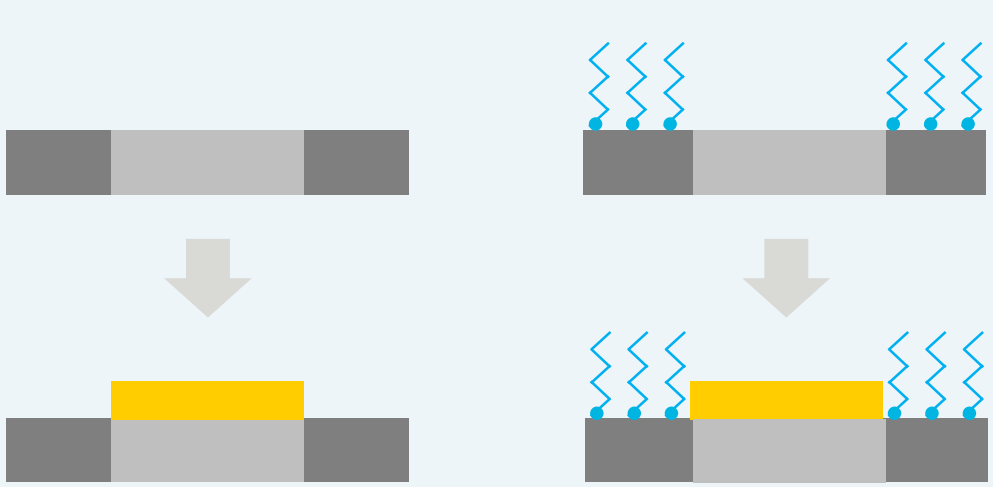
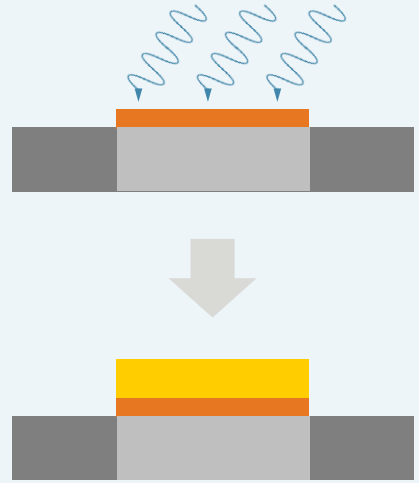
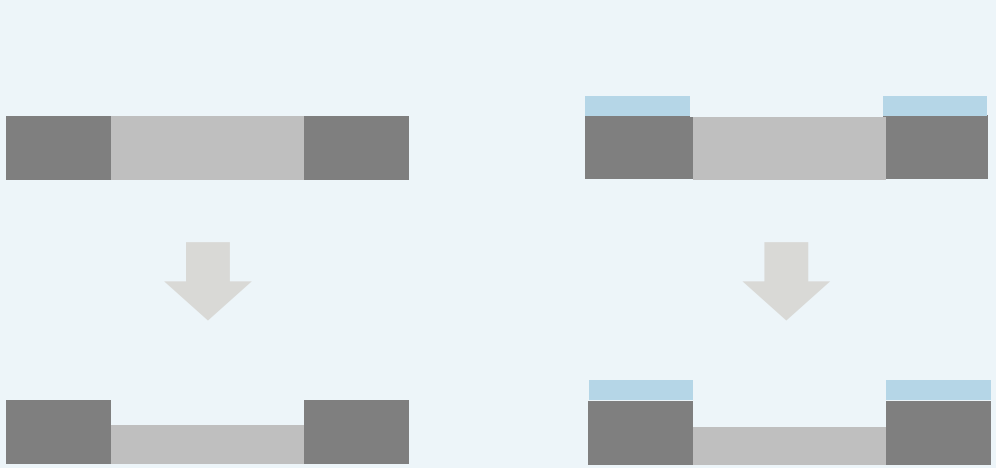
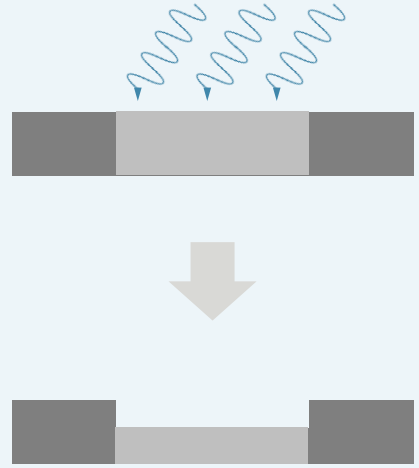
3D integration (M3D)



VLSI-T 2018

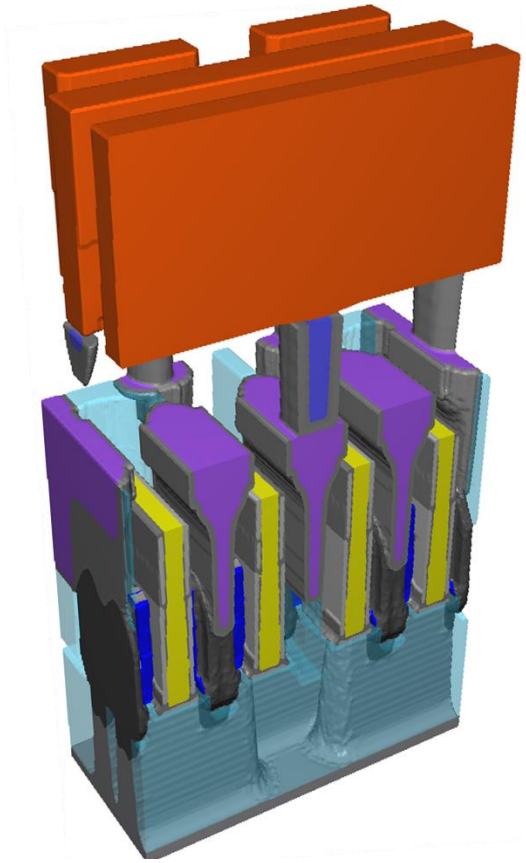
Source: AMAT, SMC Korea 2019 conference

Selective Processes – Enabling New Ways to Build Chips

	Inherent Selective	Selective by Deactivation	Selective by Activation
Selective Deposition	 ANALOGY: 3D Printing – putting material only where you want it (materials-enabled not litho)		
Selective Removal	 ANALOGY: Removing a single weed without damaging adjacent grass or leaving residue		

T. Faraz et al, *ECS J. Solid State Sci. Technol.* 4(6), N5023-N5032 (2015)

Scaling Opportunities in the Back-end-of-line Interconnect



Source: AMAT, IEDM 2018 conference

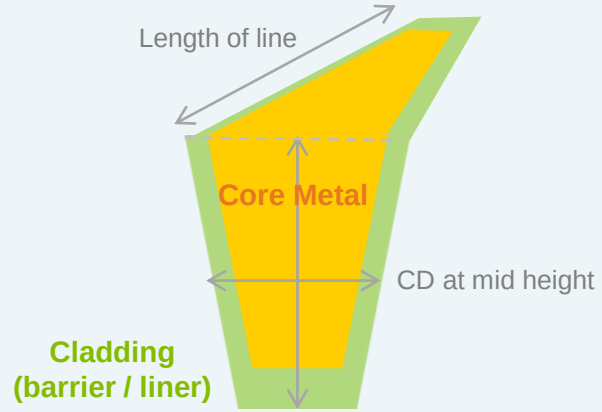
BEOL Cu wire resistance (R) increasing at smaller geometries due to:

- Inherent bulk conductor resistivity characteristics
- Trade-off fill volume of high-conducting vs. cladding materials
- Scattering at surfaces and grain boundaries

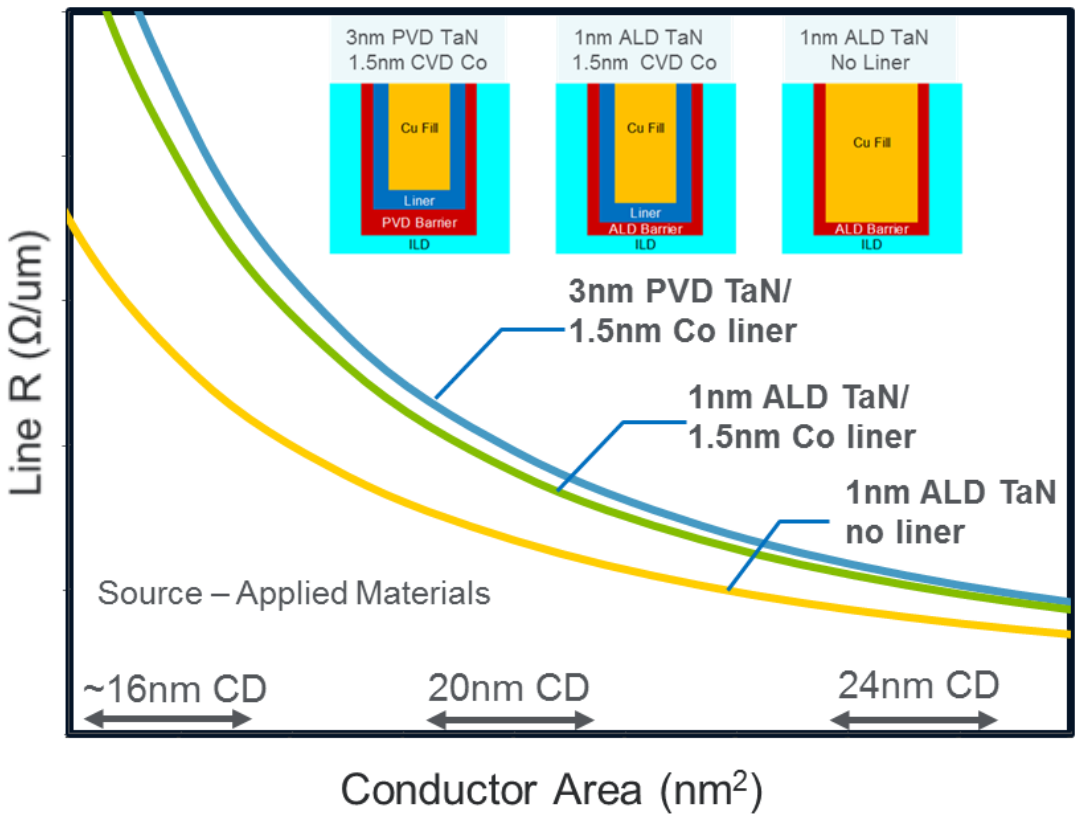
…resulting in slower performance and higher power consumption

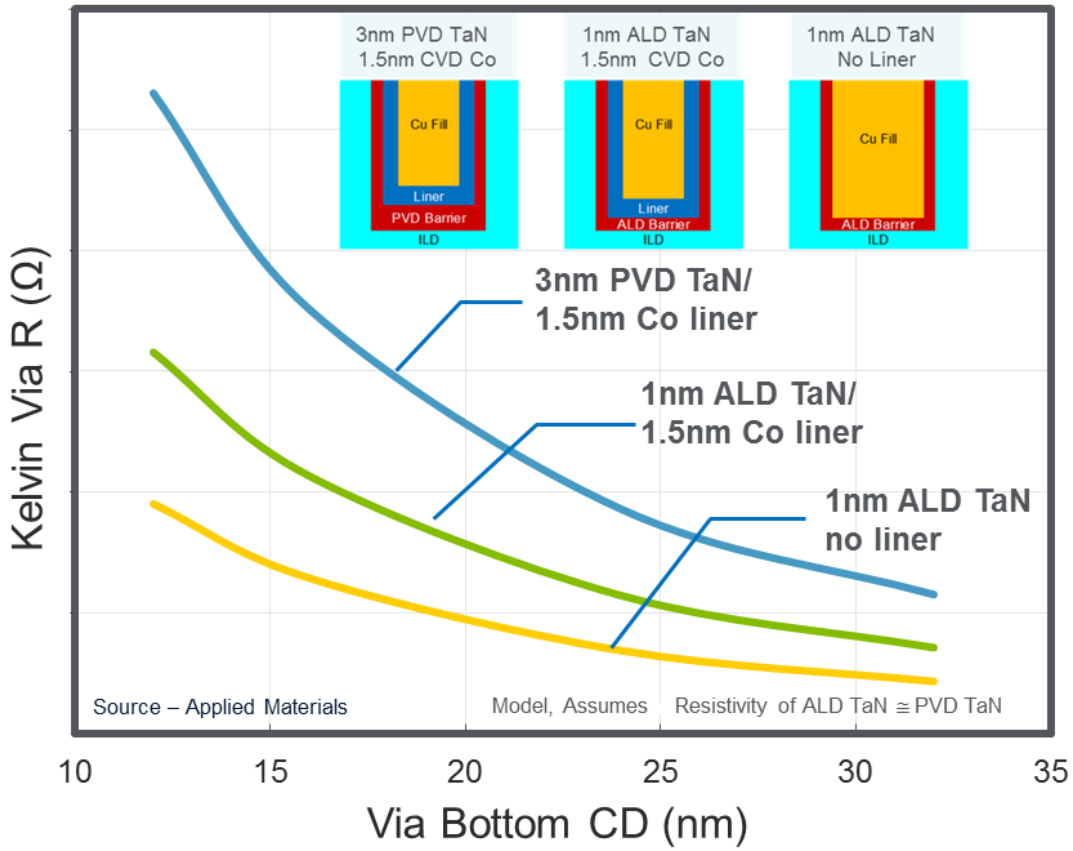
Materials-Engineering Solutions:

- New materials (lowest R at CD)
- Full-volume metal fills
- Interface management



Copper Extension Scenarios

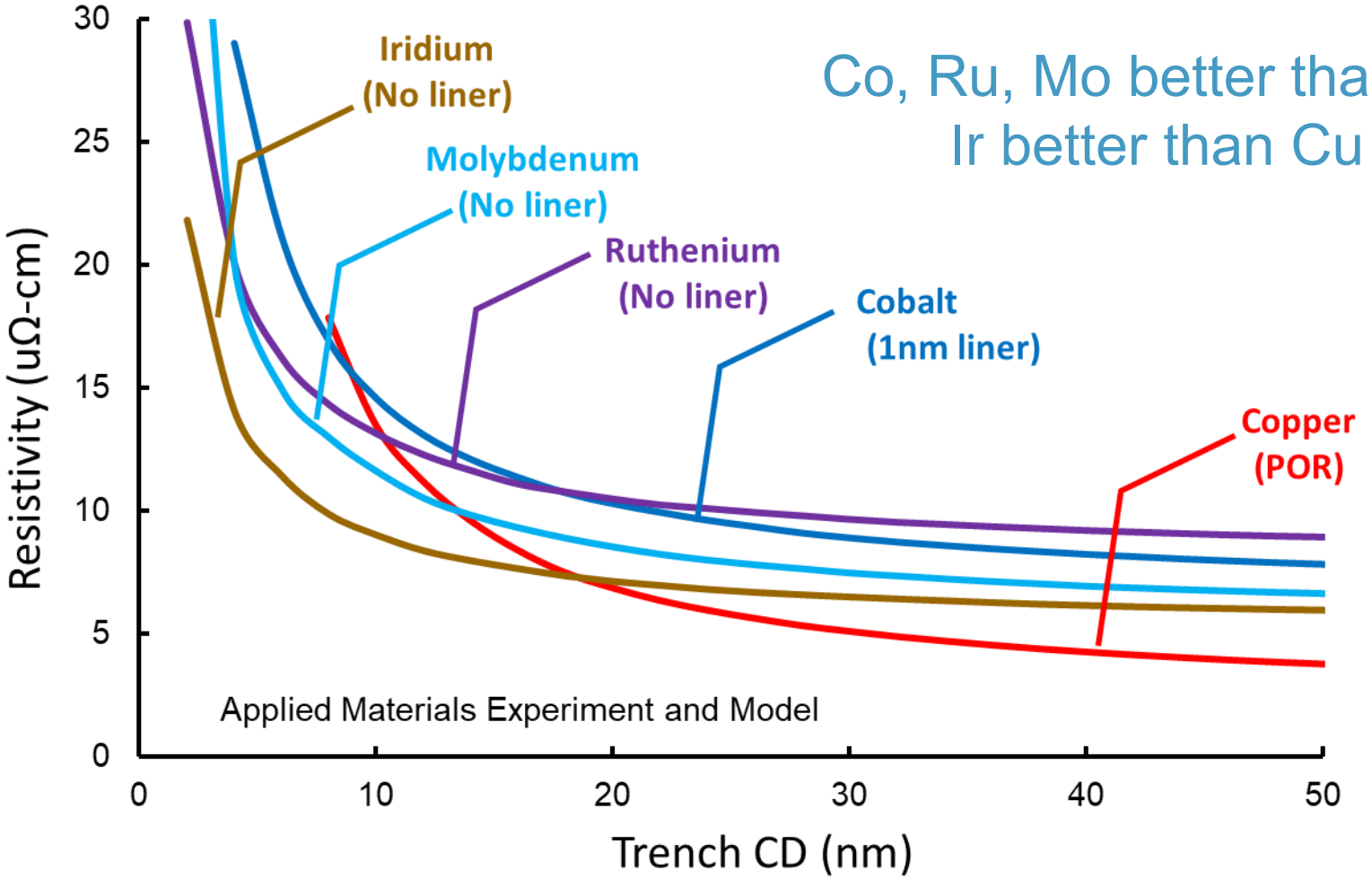




Co-optimization of ALD barriers with thinner liners and new fill technology is key to maximize conductor volume (low line R) and minimize interface resistance (low via R)

Source: AMAT, IEDM 2018 conference

New Metals for Resistance Scaling



Co, Ru, Mo better than Cu at CD's between 10-15nm
Ir better than Cu at CD's between 15-20nm

Is a barrier or liner required?

How to fill?

Can it be integrated?

Source: AMAT, IEDM 2018 conference

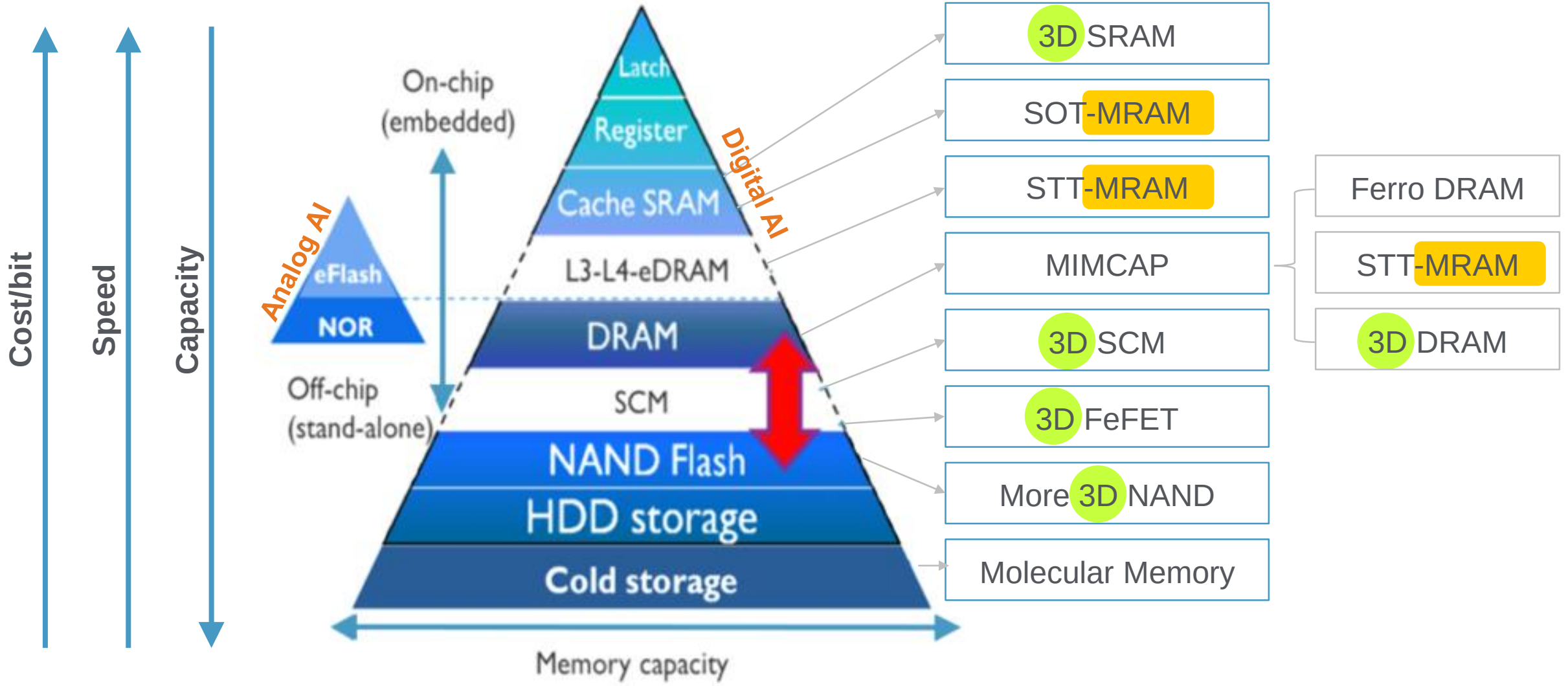
AGENDA

Semiconductor Scaling and Enablers

Logic Scaling Advancements

Memory Advancements to Scale A.I. Big Data

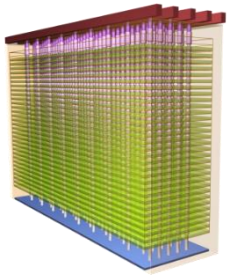
Emerging Memories for Pervasive Data and Compute



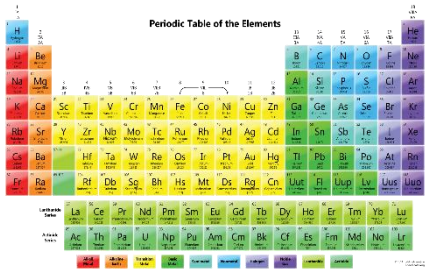
Source: http://www.imec.be from techspot.com (February 21, 2019)

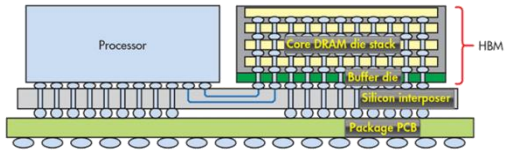
Architectural Advancements Leveraging Memory & Design

3D Architecture → **Material Innovation** → **DRAM & Packaging** → **Distributed Cache** → **Analog Computing**

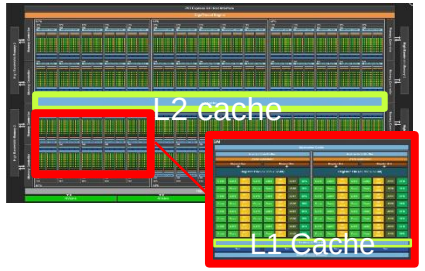


Micron.com

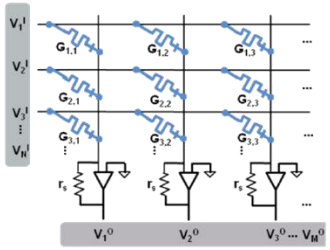




ElectronicDesign.com (Samsung)



NVIDIA Tesla P100 white paper

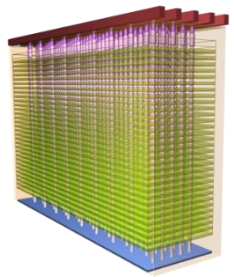


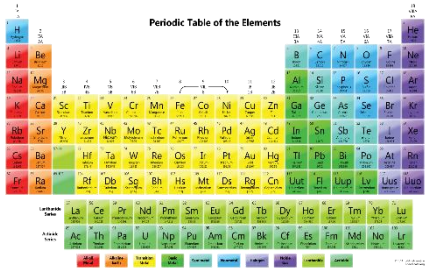
HPE, ACM 2016

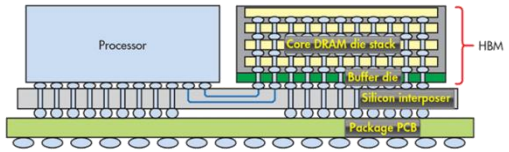
Significant investments in memory to enable better bit scaling and performance

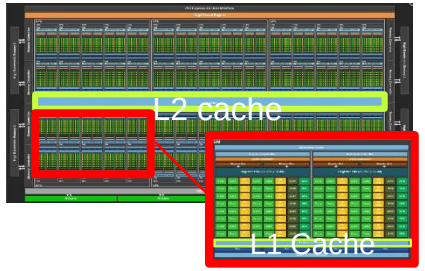
3D Architectures

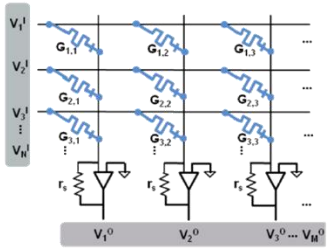
3D Architecture → **Material Innovation** → **DRAM & Packaging** → **Distributed Cache** → **Analog Computing**



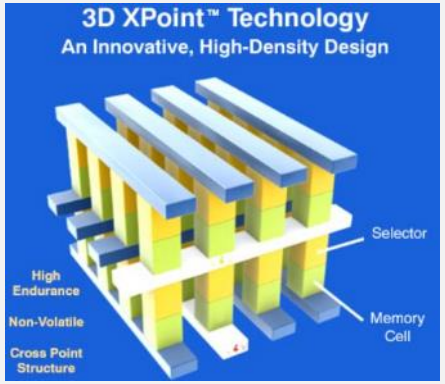




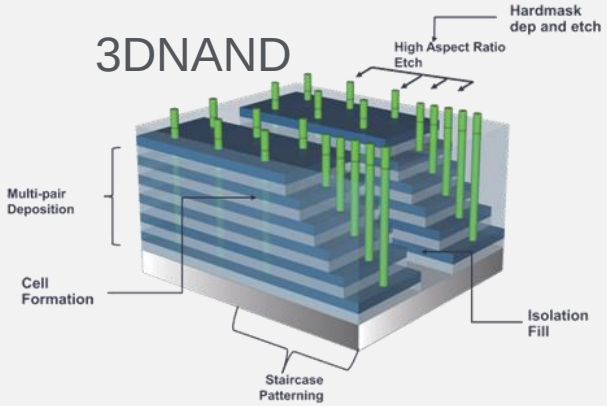




3D architectures extend memory densities



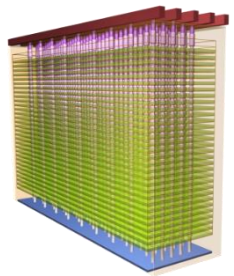
Source: Intel 2015

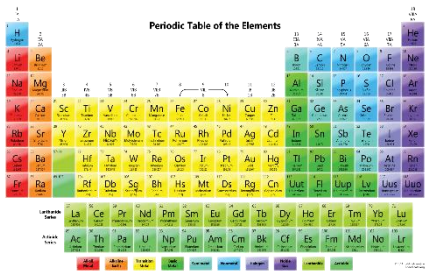


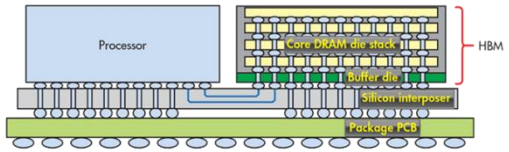
S. Kang, IMW2018 tutorial

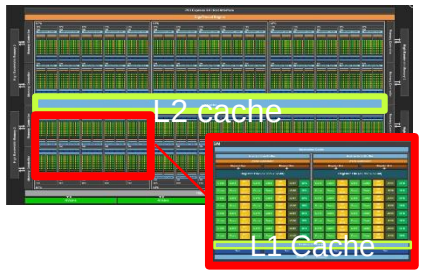
3D Architectures

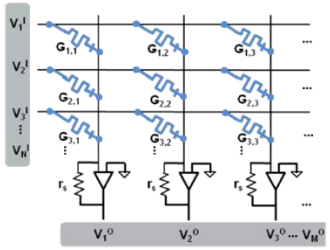
3D Architecture → **Material Innovation** → **DRAM & Packaging** → **Distributed Cache** → **Analog Computing**











Materials and Interfaces

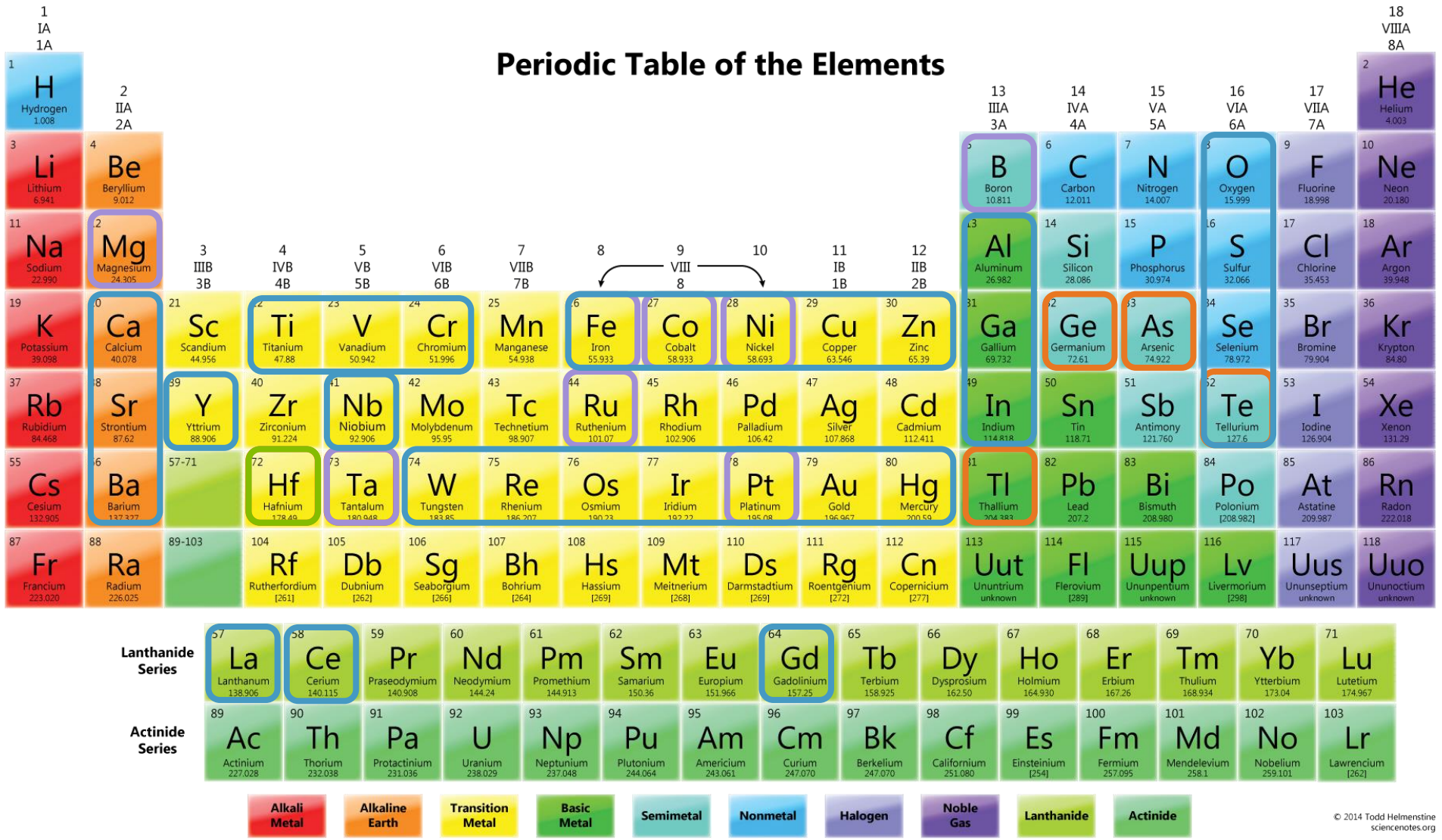
- Complex stacks with multiple exotic materials are required to enable new memories
- Deposition, etch and CMP are critical process development areas

Deposition

Etch

Planarization

Example of Elements to Enable New Memories



PCM Selector

- M-K. Lee, IEDM 2012
- G.H. Kim, APL 2012
- L. Zhang, IEDM 2014

STT-MRAM

- K. Ando, JAP 2014

FE-FET

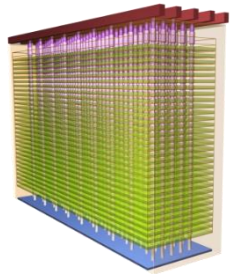
- X. Tian, APL 2018
- A. Pal, APL 2017

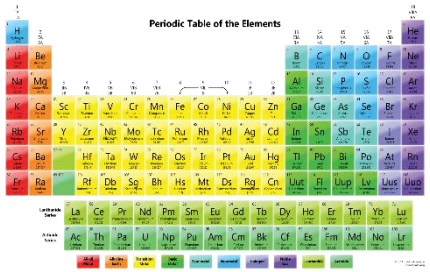
CBRAM

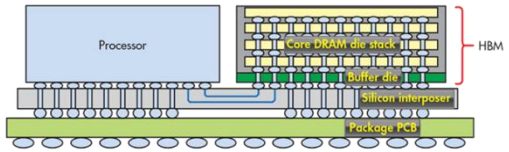
- Adesto Technologies, IEEE 2013 talk

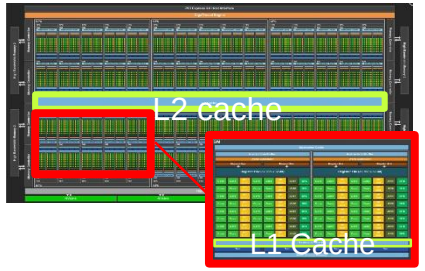
DRAM and Packaging

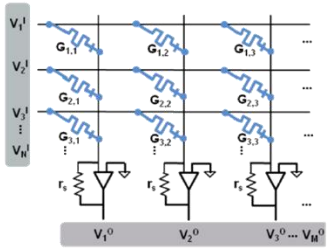
3D Architecture → Material Innovation → DRAM & Packaging → Distributed Cache → Analog Computing



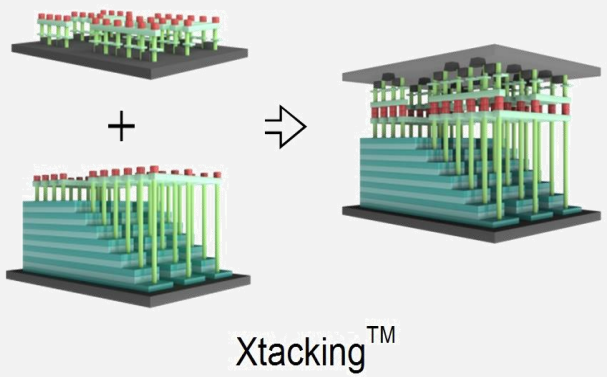




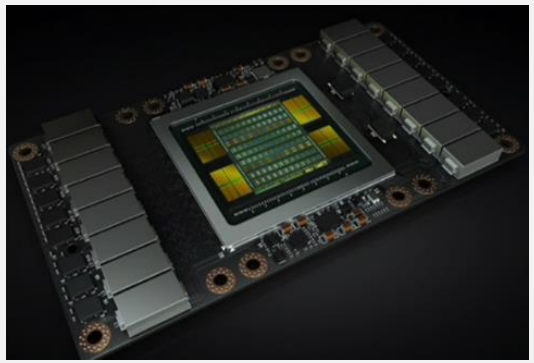




Packaging enables higher performance to support more complex workloads



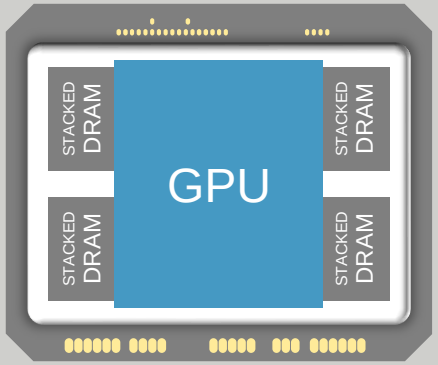
YMTC.com (2018)



Nvdia V100 source: techreport.com 2017

Beyond Chip Scaling: Advanced Packaging

DRAM ON PCB to STACKED DRAM IN PACKAGE

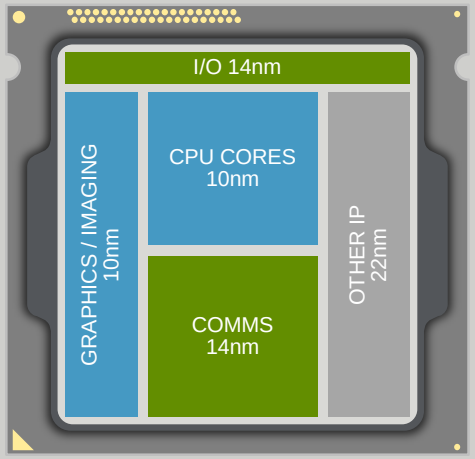


SOURCES: AMD, NVIDIA

3x
Logic ←→ DRAM bandwidth performance

50%
Power savings per bit

HETEROGENEOUS INTEGRATION



SOURCES: Intel, GLOBALFOUNDRIES

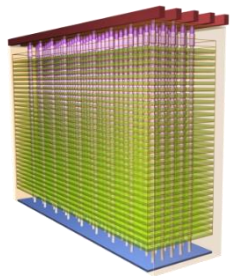
System on Chip to System on Package

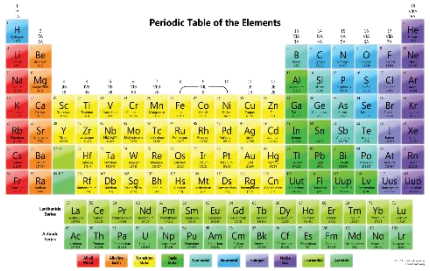
Integration of chiplets provides **time, cost** and **yield** benefits

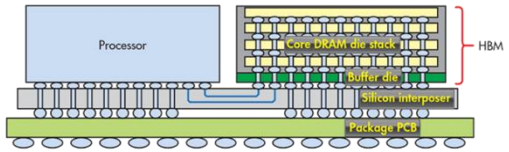
Connecting chips together in new ways using advanced packaging

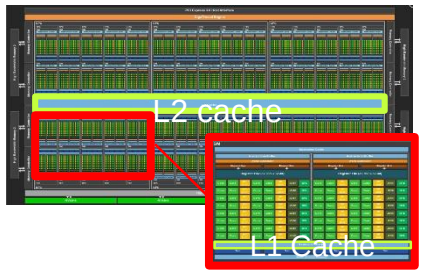
Distributed Cache

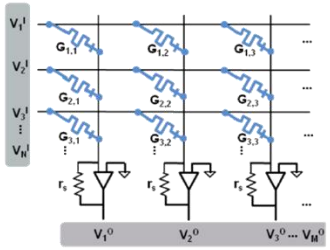
3D Architecture → Material Innovation → DRAM & Packaging → Distributed Cache → Analog Computing



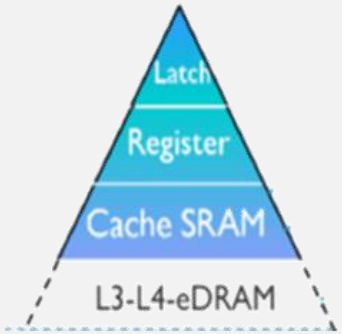




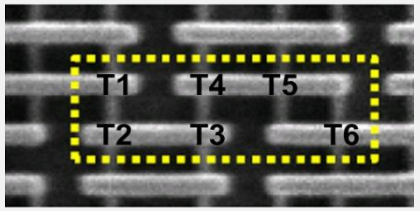




- Distributed computing further reduces memory bottlenecks → SRAM
- MRAM alternative to improve area + energy efficiency

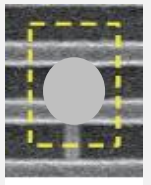


Source: techspot.com



6T SRAM

Source: Intel, IEDM, 2012



1T, 1 STT-MRAM

Source: AMAT

Multiple MRAM Milestone Announcements

8/2017

Everspin announced sampling of the world's first 1Gb MRAM product

9/2017

Global Foundries announced availability of embedded MRAM on leading 22FDX FD-SOI platform

7/2018

TSMC "Over 50 customers and 140 tapeout in N22ULP, spanning many application areas… connectivity, digital TV/STB, application processors."

11/2018

Spin Memory & **Applied Materials** commercial agreement to create a comprehensive embedded MRAM solution

12/2018

Gyrfalcon Technology announced the commercial availability of its AI ASIC that include **TSMC**'s 22nm eMRAM

12/2018 IEDM conference

Intel described the first FinFET-based MRAM technology (22nm) is production ready

Samsung showed MRAM design technology co-optimization for hardware neural networks

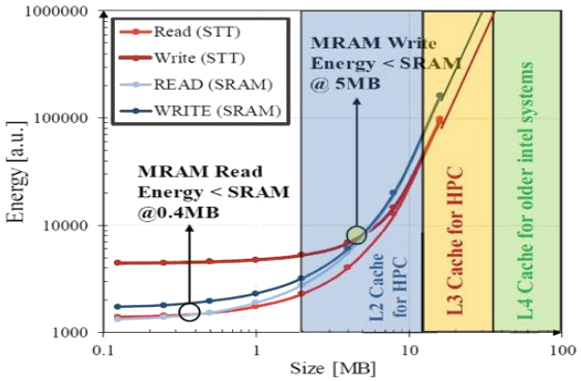
Global Foundries showed 22nm embedded 40 Mb MRAM for low-power automotive MCU application

3/2019

Samsung shipped the first 28nm eMRAM

MRAM Benefits

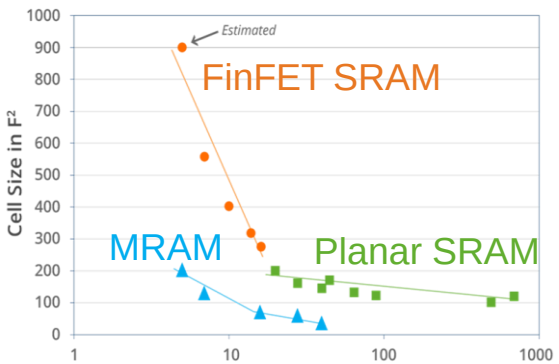
Low Power & Non-Volatile



- IMEC DTCO studied:
  - MRAM write energy < SRAM @ 5MB
  - MRAM read energy < SRAM @ 0.4MB

Source: IMEC, IEDM 2018

High Density

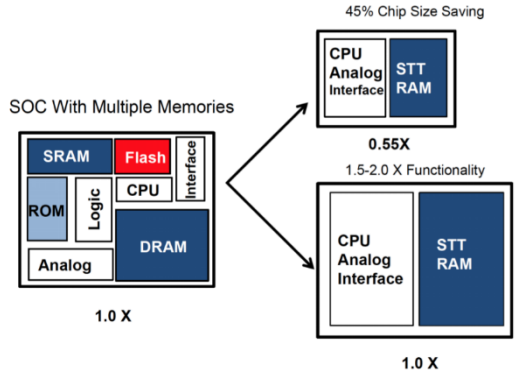


CMOS Technology Node (nm)

- MRAM is with > 3x cell density than SRAM
- Greater benefit in advanced node

Source: SpinMemory.com 2019

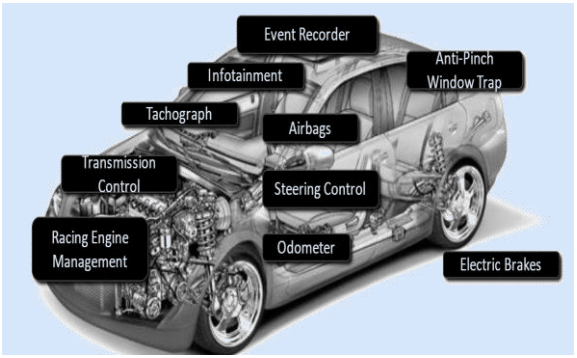
More Functionality



- MRAM implementation in MCU / SoC:
  - 45% chip size saving (~ one advanced technology node advantage) **OR**
  - 1.5-2x more processors at same chip size

Source: Spintec ORaP Forum 2015

Robust

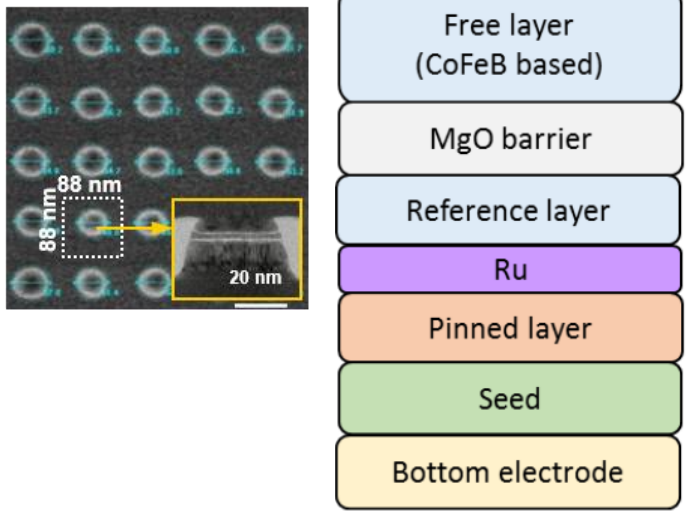


Source: Everspin.com

- MRAM for high temperature, low-power operations:
  - 22nm embedded 40 Mb MRAM could be used between -40 to 150 degrees C with good read and write characteristics and immunity to 500 Oersted magnetic fields.

Source: GF, IEDM 2018

MRAM Capability Demonstrated by Optimization of Complex Stack



Sources: Applied Materials

Systematic Optimization of 1 Gbit Perpendicular Magnetic Tunnel Junction Arrays for 28 nm Embedded STT- MRAM and Beyond

IEDM, 2015

C. Park[1,a], J.J. Kan[1], C. Ching[2], Hassan[2]

[1]Qualco
[2]Silicon Systems
[a]Tel: 1-858-658-1890, [a]cha

Systematic Validation of 2x nm Diameter Perpendicular MTJ Arrays and MgO Barrier for Sub-10 nm Embedded STT-MRAM with Practically Unlimited Endurance

J.J. Kan[1,a], C. Park[1], C. Ching[2], J. Ahn[2], L. Xue[2], R. Wang[2], A. Kontos[2], S. Liang[2], M. Bangar[2], H.Chen[2], S. Hassan[2], S. Kim[1], M. Pakala[2,b], and S. H. Kang[1]

IEDM, 2016

[1]Qualcomm Technologies, Inc., San Diego, California 92121, USA. E-mail: [a]jkan@qti.qualcomm.com
endra_pakala@amat.com

PENDICULAR MTJ DEVICES

Low RA Magnetic Tunnel Junction Arrays in Conjunction with Low Switching Current and High Breakdown Voltage for STT-MRAM at 10 nm and Beyond

C. Park[1,a], H. Lee[1], C. Chi

[1] Corporate Research and Development, Qualcomm
[2]Silicon Systems Group, App

VLSI, 2018

Abstracts

The scaling of STT-MRAM for deeply scaled nod
10 nm CMOS) requires low resistance-area-pr
magnetic tunnel junctions (MTJs) to contai
voltage (V_c) and to assure high endurance. In
various reports, we demonstrate systematic eng
low-RA MTJs without trading off key device at
remarkably, with higher barrier reliability. The MT
an ultra-thin synthetic antiferromagnetic layer (t
Co/Pt pseudo-alloy pinned layer. By reducing RA
5 Ωμm^2, significantly reduced V_c and reliable sw
ns have been achieved. Furthermore, the breakd
(V_{BD}) has been improved. The results sugge
tunability of MTJ is extended to sub-10 nm CM
performance and high-reliability MRAM.

Process Optimization of Perpendicular Magnetic Tunnel Junction Arrays for Last-Level Cache beyond 7 nm Node

Lin Xue [a], Chi Ching, Alex Kontos, Jaesoo Ahn, Xiaodong Wang, Renu Whig, Hsin-wei Tseng, James Howarth, Sajjad Hassan, Hao Chen, Mangesh Bangar, Shurong Liang, Rongjun Wang, Mahendra Pakala [b]

Applied Materials, Inc., Sunnyvale, California 94085, USA
[a] Lin_Xue@amat.com, [b] Mahendra_Pakala@amat.com

VLSI, 2018

Abstract

This paper demonstrates systematic process optimization of perpendicular magnetic tunnel junction (pMTJ) by hardware, unit-process, and material stack design. TMR of 200% at RA 5 Ohm·μm^2, H_{SAF} ~ 8 kOe, and 10-time tunability of Hc were achieved at the film level. After patterning, 10^{-6} write error rate was reached at 0.4 pJ, V_{BD} was as high as 1600 mV at 20 ns pulse width, and excellent device stability against 400°C BEOL baking was demonstrated. The device performance along with the process capability to make MTJ array at 88 nm pitch provides opportunities for LLC applications.

Introduction

STT-MRAM has become one of the leading new memory technologies as it demonstrated excellent scalability and low

1~30 were deposited, and their TMR values were measured. Although lower polarization at lower RA limits TMR, TMR of 200% was achieved at RA of 5 Ohm·μm^2, relevant to LLC applications. TMR ~ 100% was measured at RA ~ 1 Ohm·μm^2. The results showed excellent RA scalability of MTJ films.

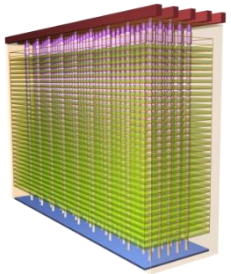
The magnetic properties of the MTJ stack were improved by testing different materials for capping and SAF coupling. Fig. 4 shows films of four different materials of capping with Hc from 8 to 80 Oe. No obvious change of RA or TMR was measured. High SAF coupling (H_{SAF}) of the MTJ film allows low write error rate. By optimizing SAF coupling material as well as magnetic material thickness within the reference layer, H_{SAF} increased from 4 to 8 kOe (Fig. 5).

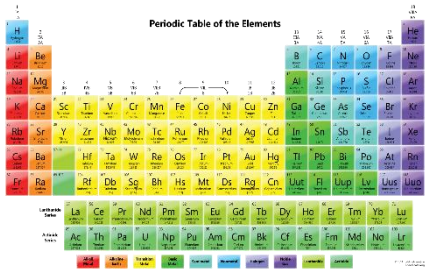
Device and Array Performance

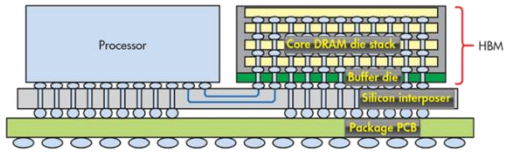
Roadmap for additional density and performance scaling in development

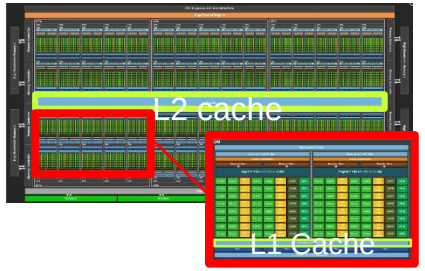
Digital to Analog Computing

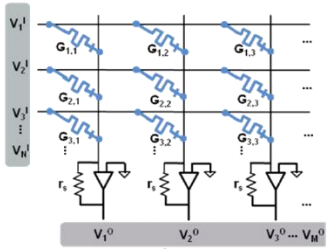
3D Architecture → **Material Innovation** → **DRAM & Packaging** → **Distributed Cache** → **Analog Computing**



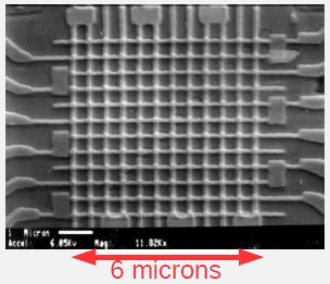




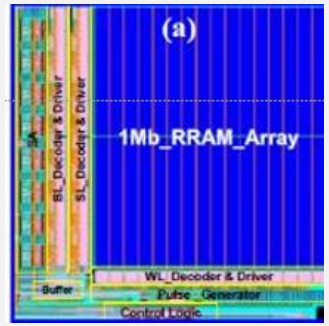




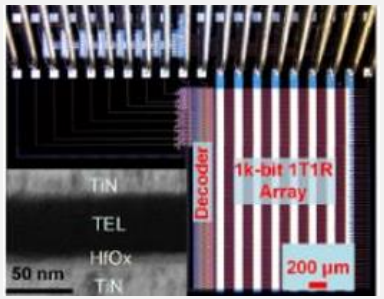
Analog vector-matrix multiplier: reduced complexity and power; will require further advances in process variability



12x12 array, Bell Labs 1986
cited by LeCuin, ISSCC 2019



CAS, IEDM 2017



Tsinghua, IEDM 2017

Emerging Memory for Machine Learning Accelerators

1. Digital Accelerator

SRAM → MRAM available in Foundry

gyrfalcon technology

tsmc

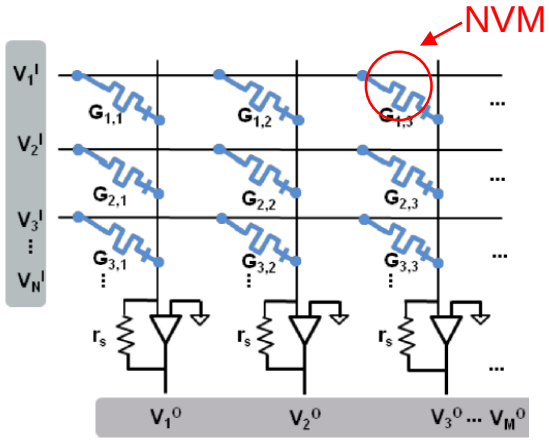
"The 2802M ASIC has 40MB of eMRAM memory, which can support large AI models or multiple AI models within a single chip." – gyrfalcontech.ai 2019

- MRAM can enable area benefit at lower power versus traditional distributed SRAM techniques

2. Analog Accelerator

eFlash → RERAM, FEFET, PCRAM in Development

In-memory Compute Vector-Matrix Multiplier



Weights: Conductance of NVM

Matrix multiplication in hardware (Ohms, Kirchhoff law)

"1,000 to 10,000 better speed-energy efficiency product than digital ASICs" - HPE, ACM 2016

- New analog memories, once performance at scale is robust, can enable additional performance boosts @ density and new ways of compute

PPAC Optimization Beyond the Chip

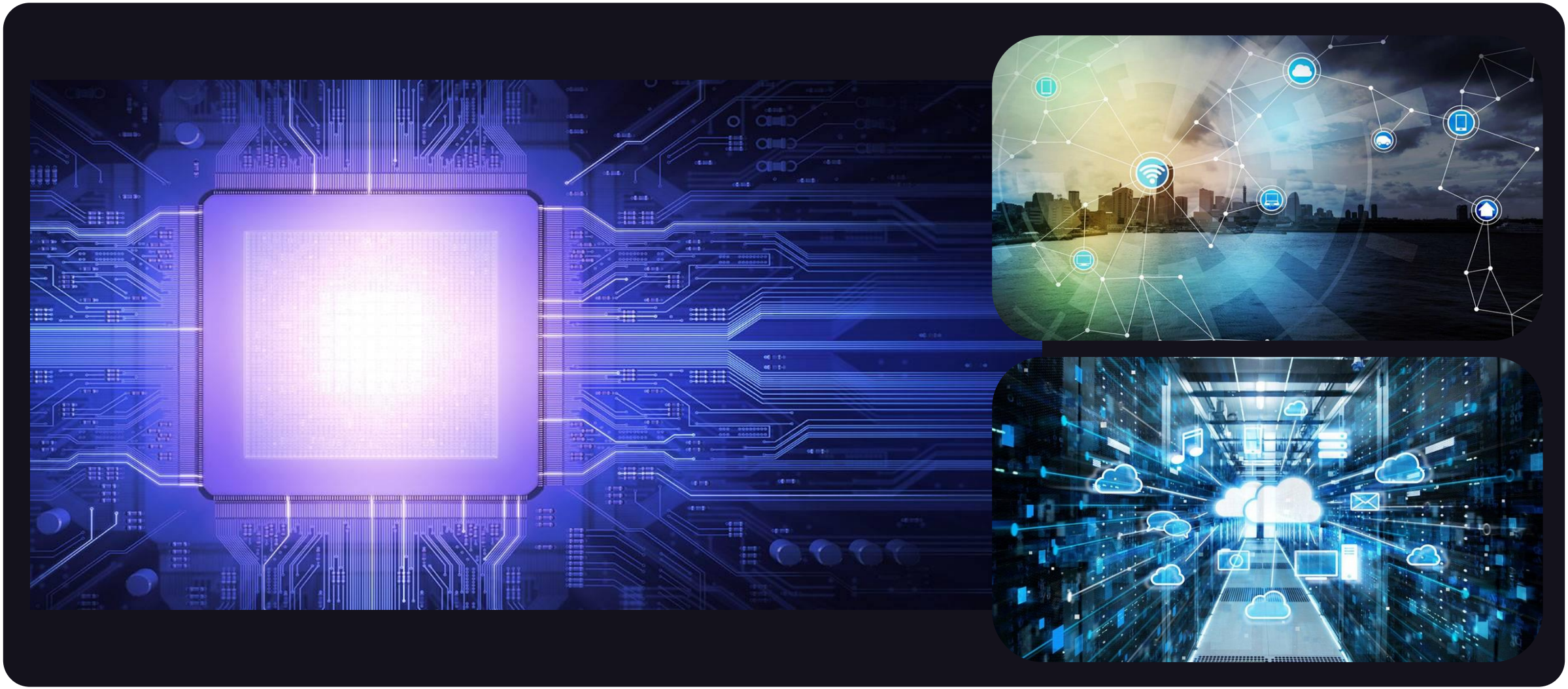


Image sources: Applied Materials Blog


APPLIED
MATERIALS®
make possible