

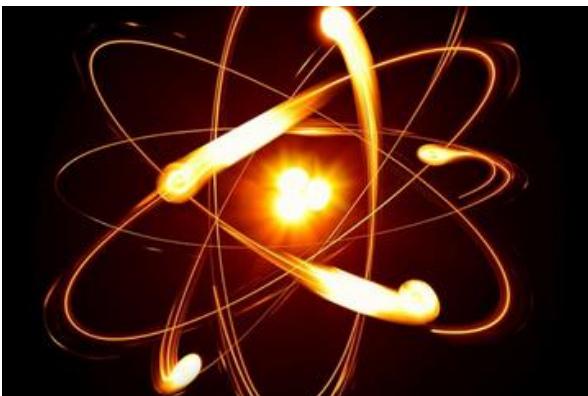
VLSI Design Methodology

**Introduction, Overview,
and Semiconductor Sector Trends**

Roman Genov
Sept 2025

Electronics

- A science about **uses and effects of electrons** (Webster Dict.)



Livescience.com



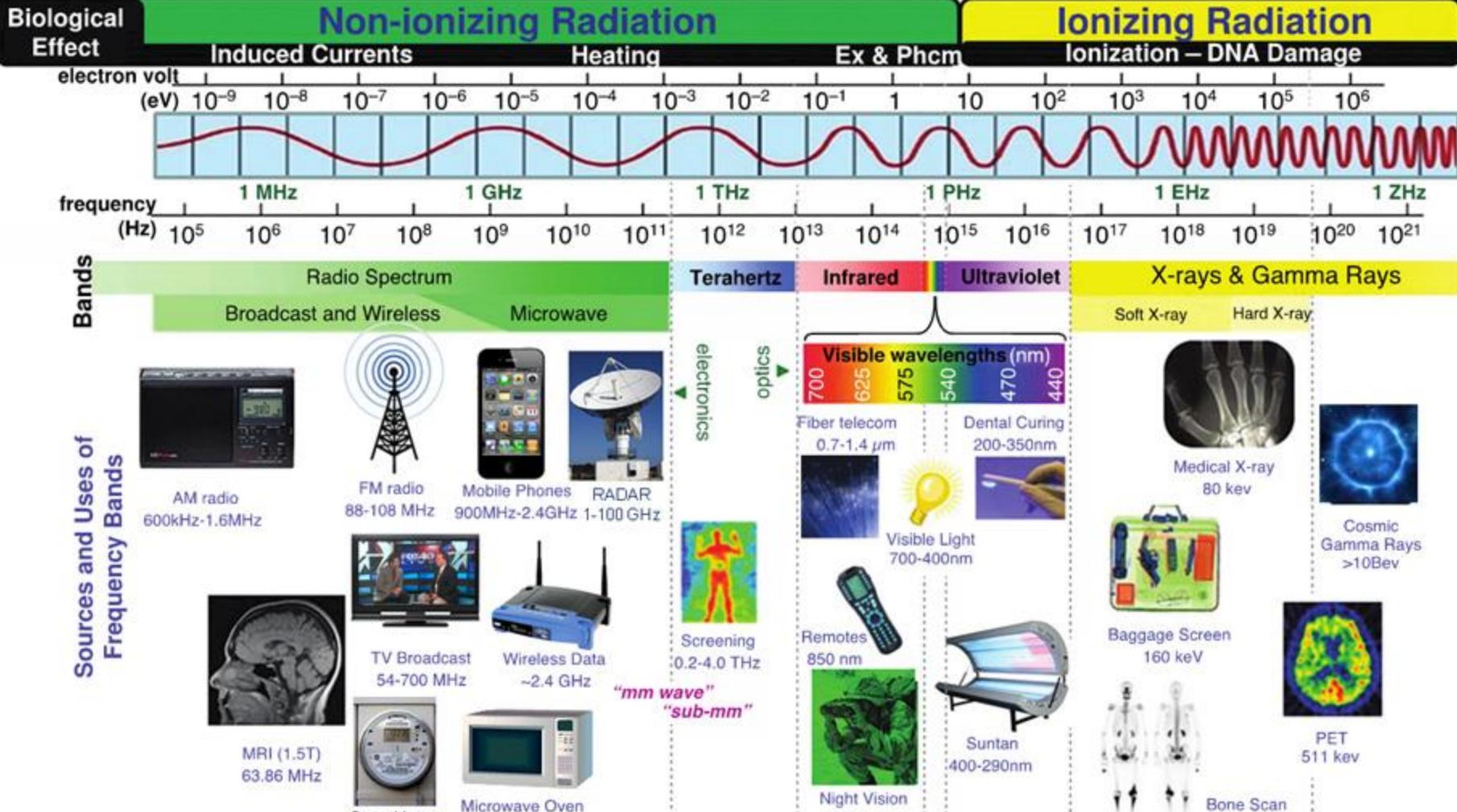
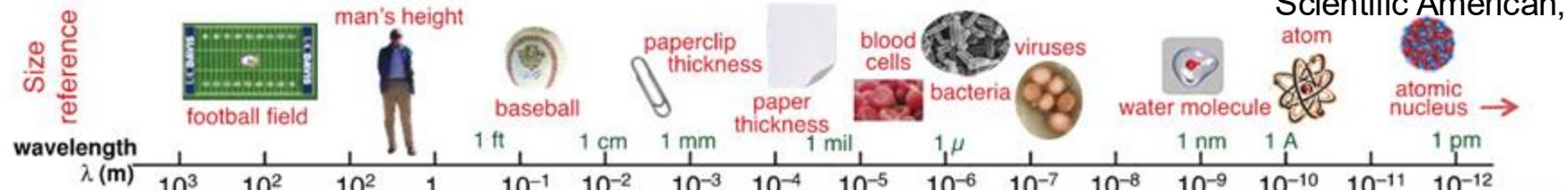
Apple iPhone microprocessor

- A science of how to control electric energy (Wikipedia)

- **electrons** have a fundamental role in electrical circuits
- such circuits involve **active electrical components**
 - **transistors**, diodes and
 - **integrated circuits (ICs)**
 - **primarily** or exclusively active **semiconductors**
 - + passive electrical components and interconnection technologies

ELECTROMAGNETIC RADIATION SPECTRUM

Scientific American, 2011



ELECTRONICS IS HERE

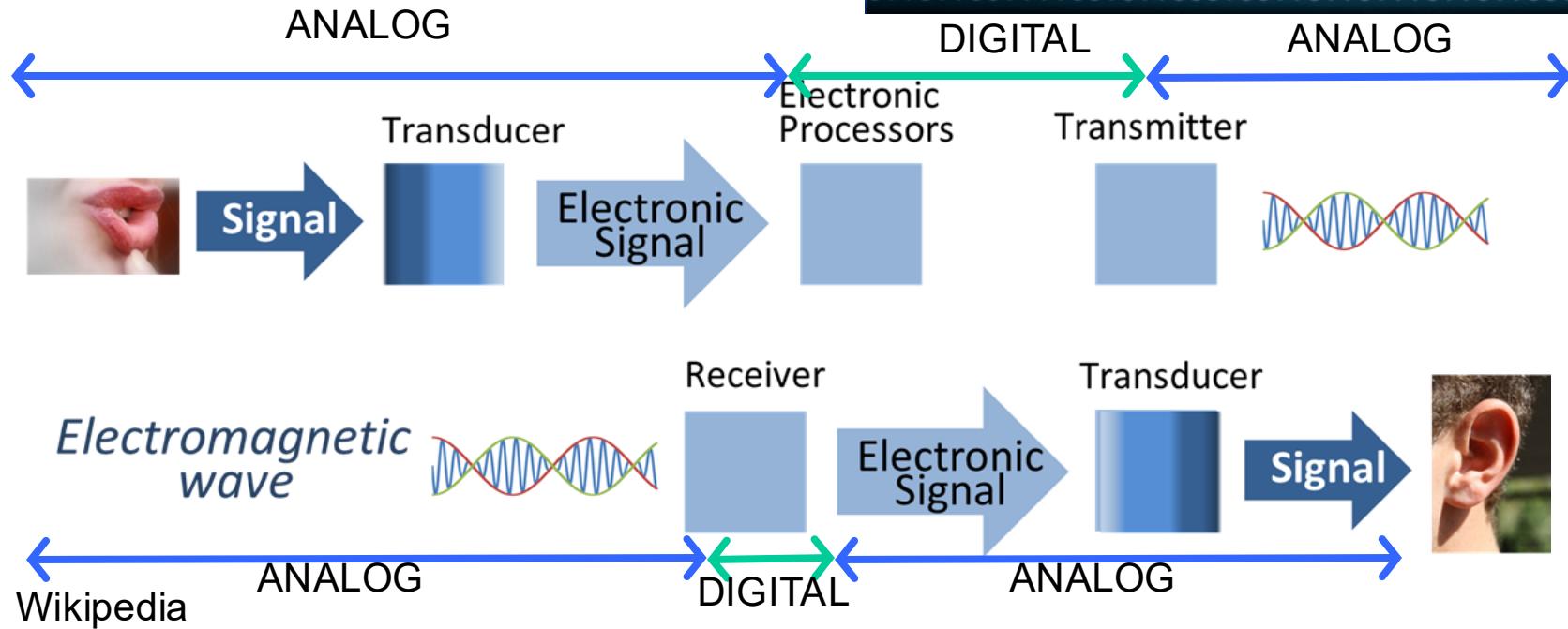
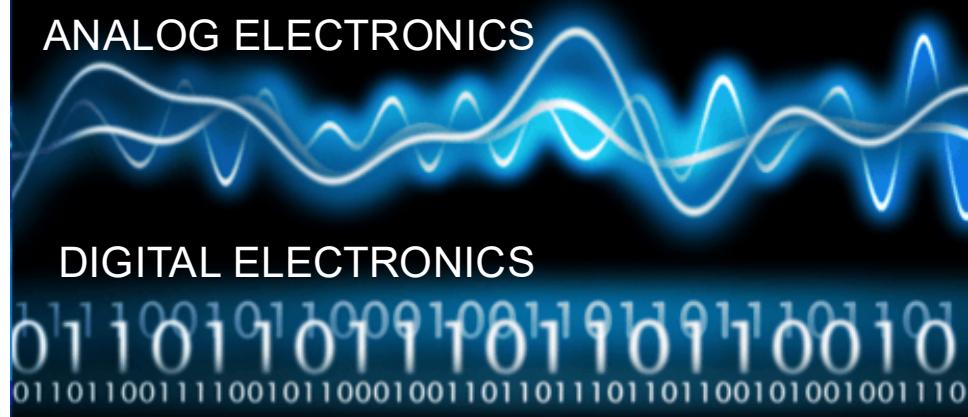
ELECTRONICS IS ALSO USED HERE

Analog versus Digital Electronics

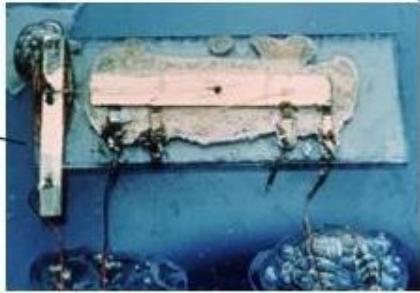
- Electronics is used to process signals (e.g., CMOS)

Coursera

- Reliable / manufacturable
- High-performance
- Versatile
- Low-cost



History of Integrated Circuits



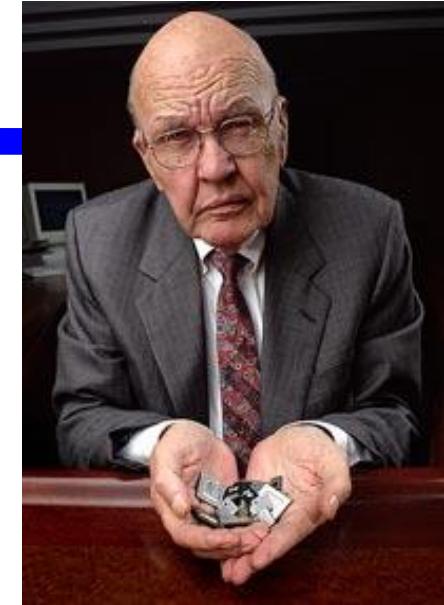
Germanium IC
Jack Kilby
Texas Instruments

1 Transistor
3 Resistors
1 Capacitor

1958



Silicon IC
Robert Noyce
Fairchild
Semiconductor
(then Intel)

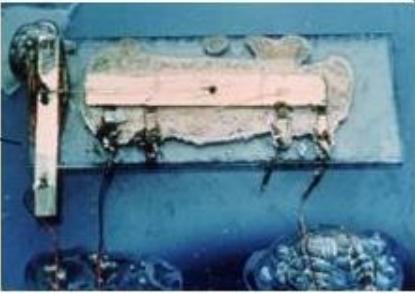


2000 Physics Nobel Prize



Flip-flop (aluminium wires) “The Mayor of Silicon Valley,” co-founded Intel with Gordon Moore

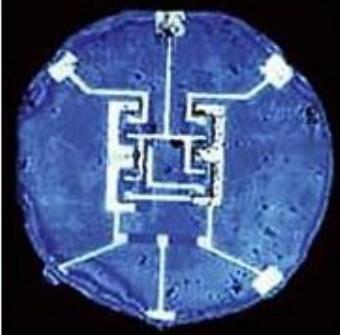
History of Integrated Circuits



Germanium IC
Jack Kilby
Texas Instruments

1 Transistor
3 Resistors
1 Capacitor
2000 Physics Nobel Prize

1958



Silicon IC
Robert Noyce
Fairchild
Semiconductor
(then Intel)
Flip-flop (aluminium wires)

■ Progress driven primarily by microprocessor technology

WALTER ISAACSON

author of *Steve Jobs*

"A stirring reminder of what Americans are capable of doing when they think big, risk failure, and work together."
—*The Atlantic*

the innovators

how a group of
hackers,
geniuses,
and geeks
created the
digital revolution



Intel: 12-Gen Processor and GPU/AI Chip (7nm)



DESKTOP:

Intel Core 12th Gen Adler

Lake Processor

(Raptor Lake - Fall 2022)

~25B Transistors

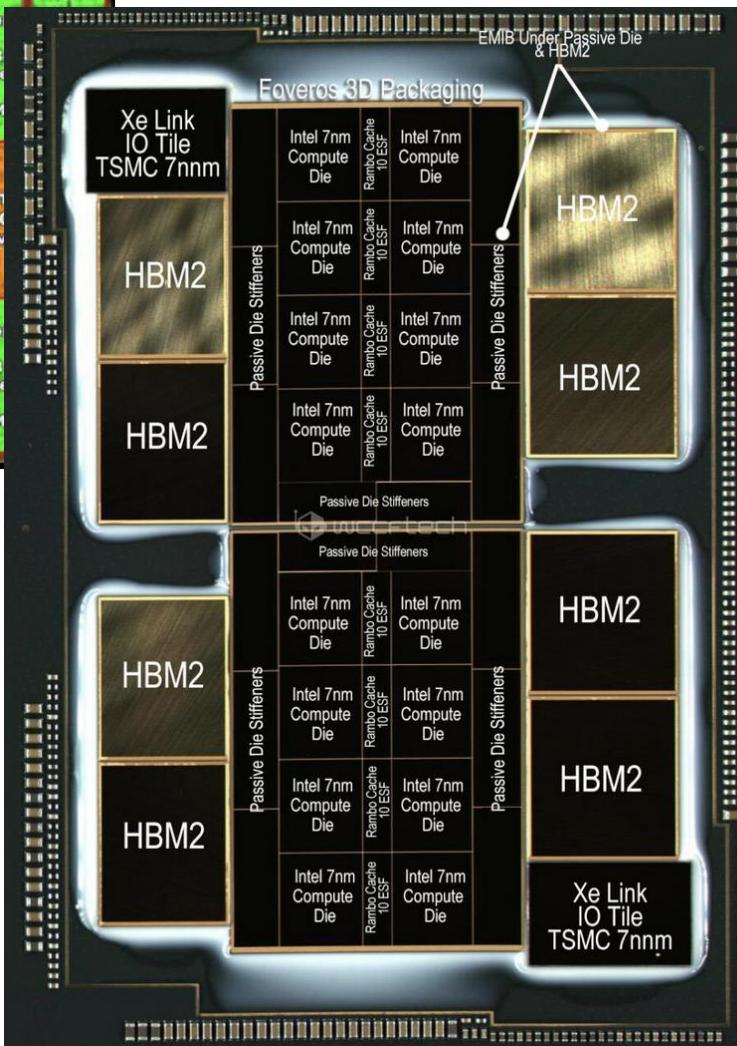
SERVER:

Pointe Veccio GPU/AI Chip!
47 chiplets on interposer, 600 W!

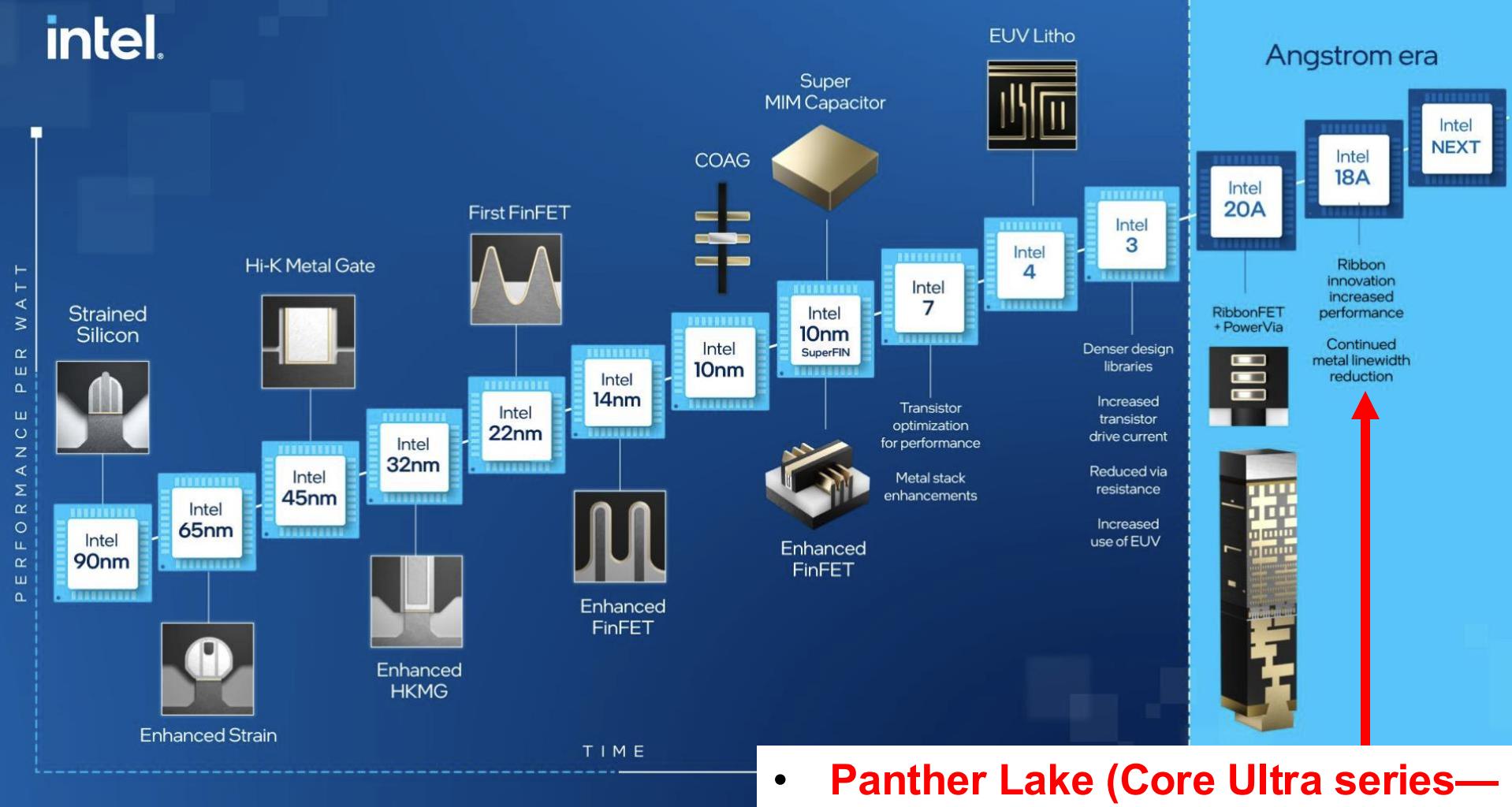
(Realto Bridge - Fall 2022)

100 Billion Transistors!

(Intel)



Intel: What is Next?

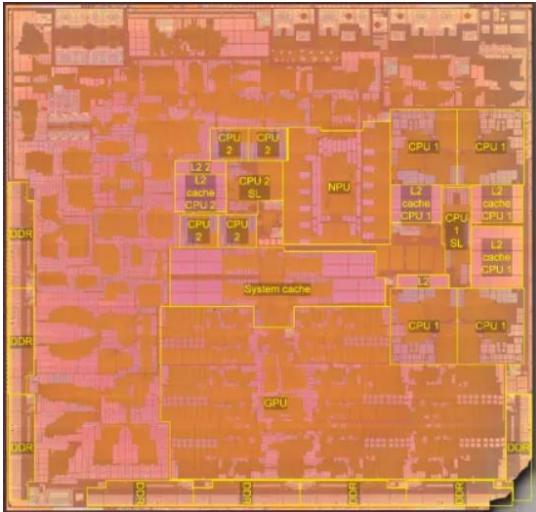


New naming convention for Intel technology nodes
- no longer directly refers to nm

- **Panther Lake (Core Ultra series—desktop & mobile) Fall 2025**
- **Clearwater Forest (Xeon Server Processor) Announced Aug 2025**

Apple 2024: M4 (3nm, 28B Ts)

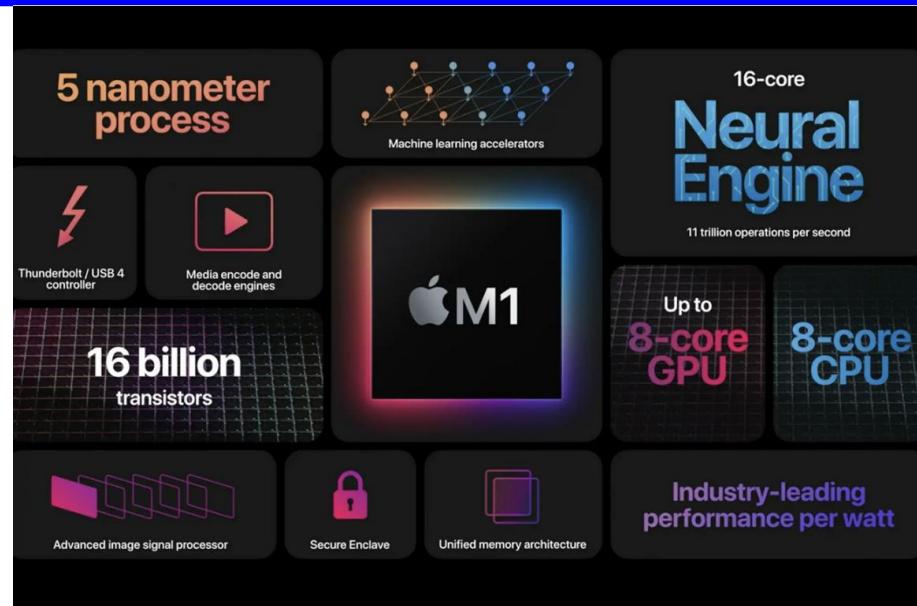
M1:
2021



M4:
2024

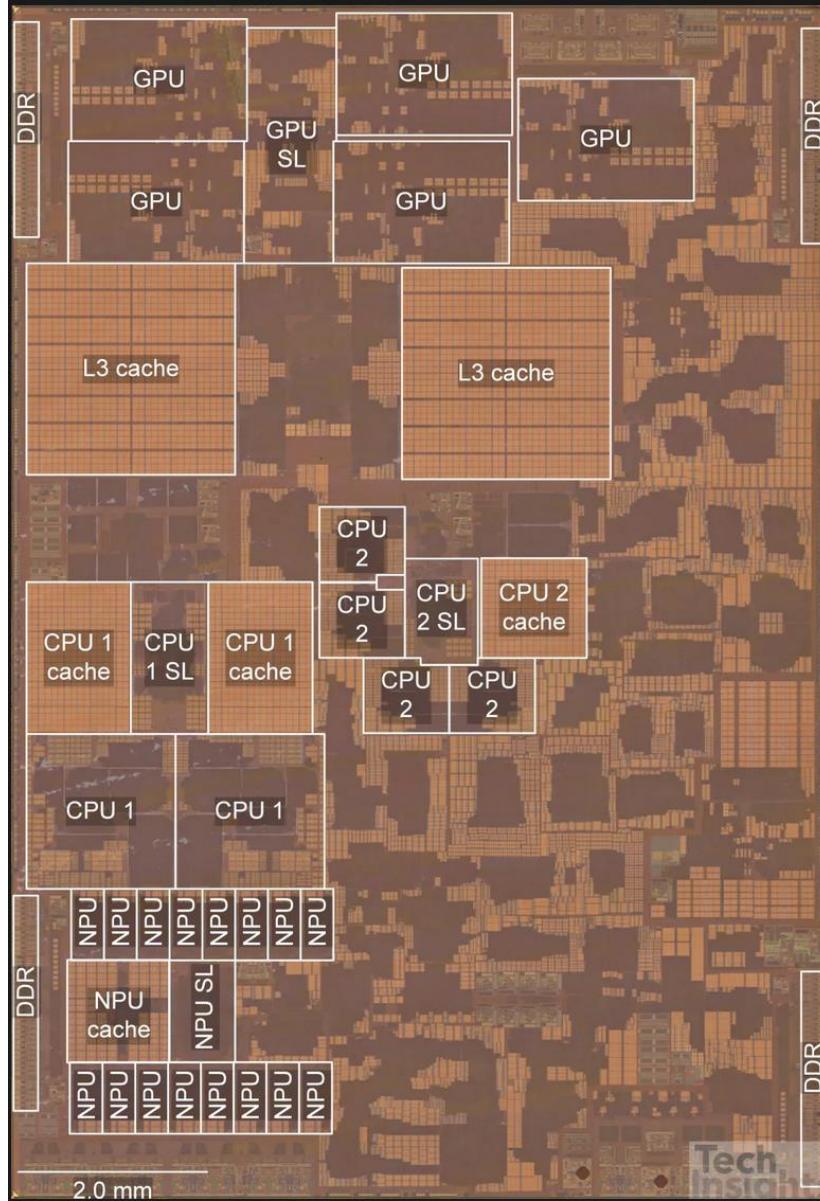
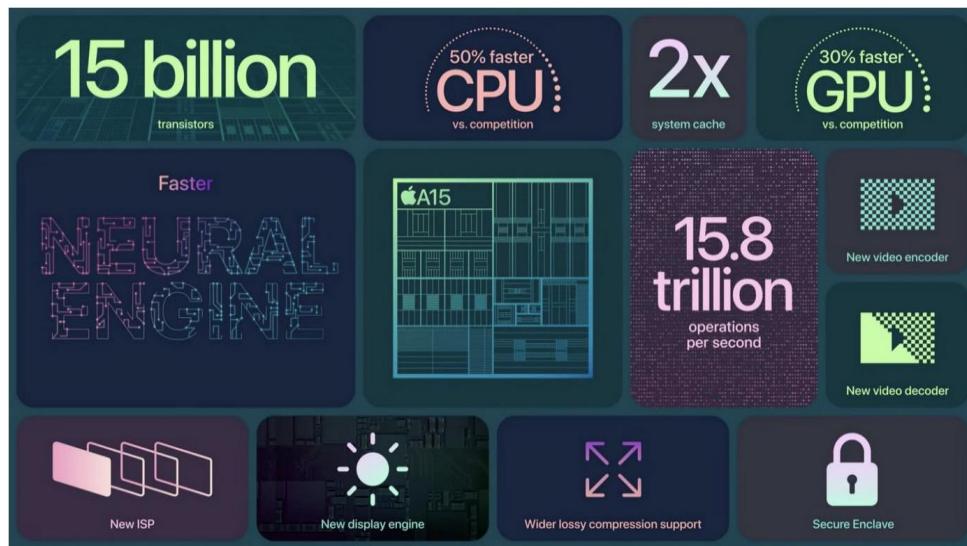


(Apple)



Complex Systems-on-Chip: Apple iPhone

- A15 - main processor in iPhone 14
 - teardown (from techinsight, chipworks)



Technology Scaling

Moore's law

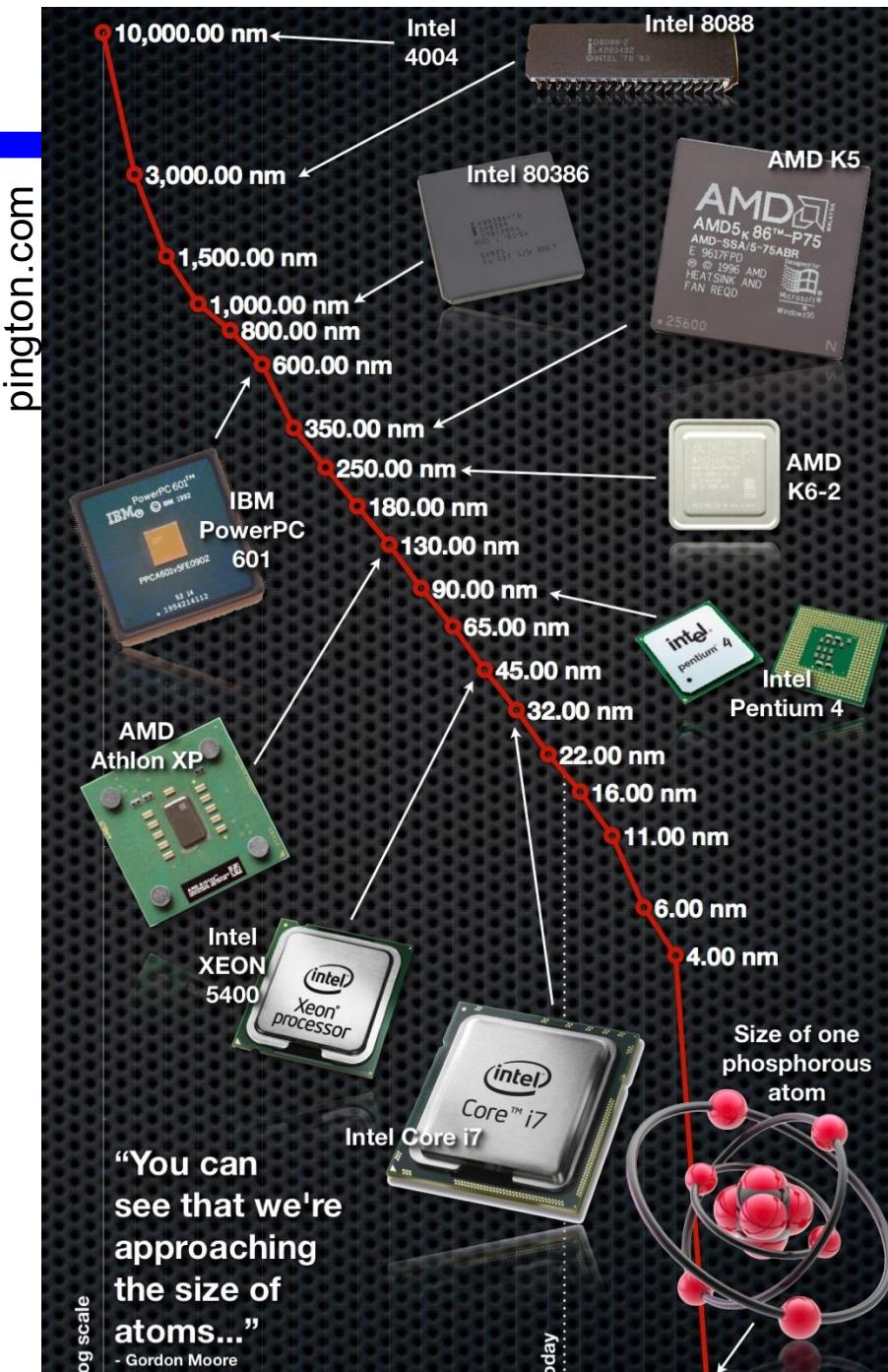
- # of transistors doubles every two years
- Now at 1-2nm
- What is next?

“More Moore”

- Beyond CMOS
- 3D stacking

“More than Moore”

- ICs applications besides μPs
- Diversification: new applications
 - IoT, AI, health, etc
 - interactive (robots etc)



Technology Scaling

Moore's law

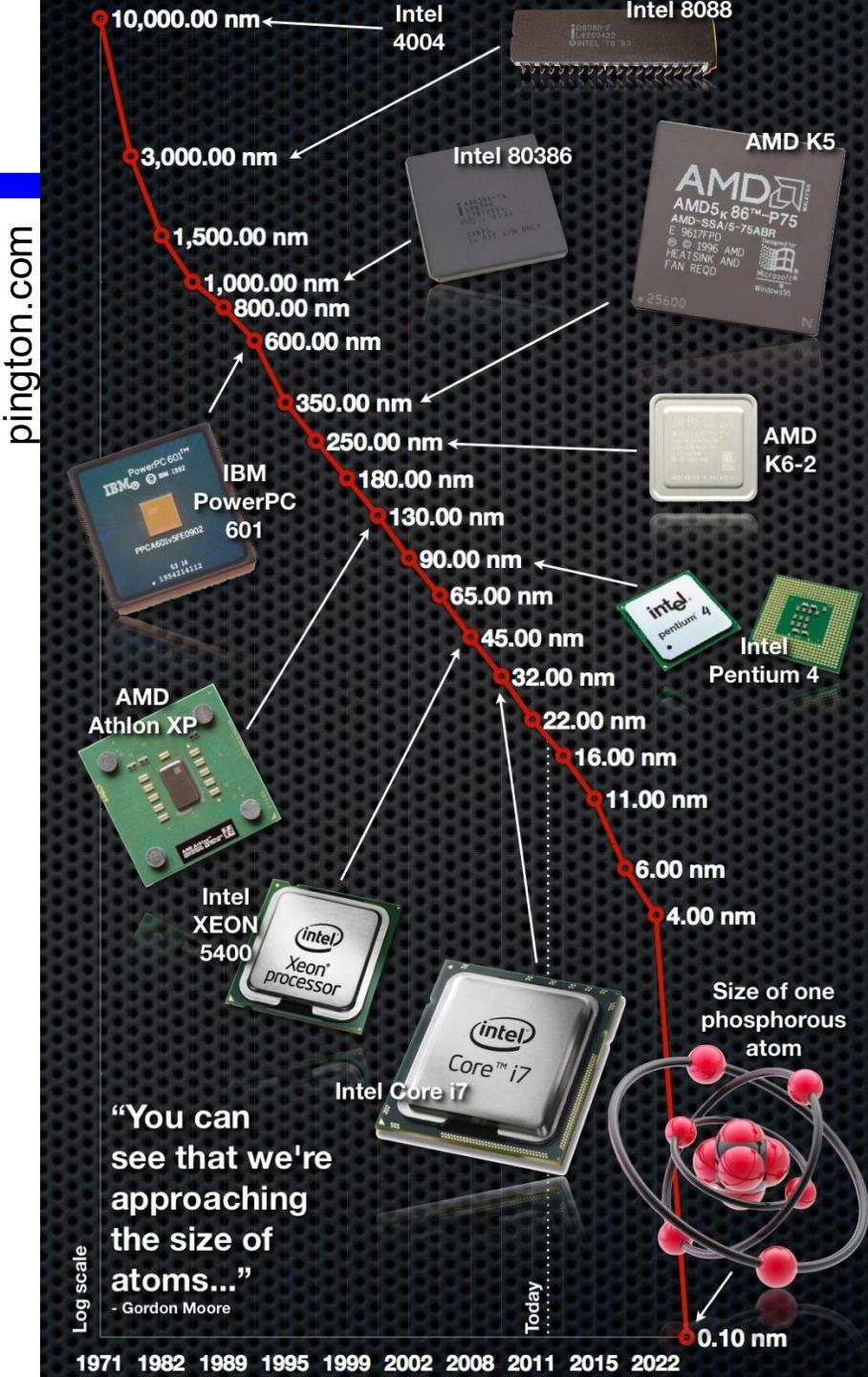
- # of transistors doubles every two years
- Now at 1-2nm
- What is next?

→ “More Moore”

- Beyond CMOS
- 3D stacking

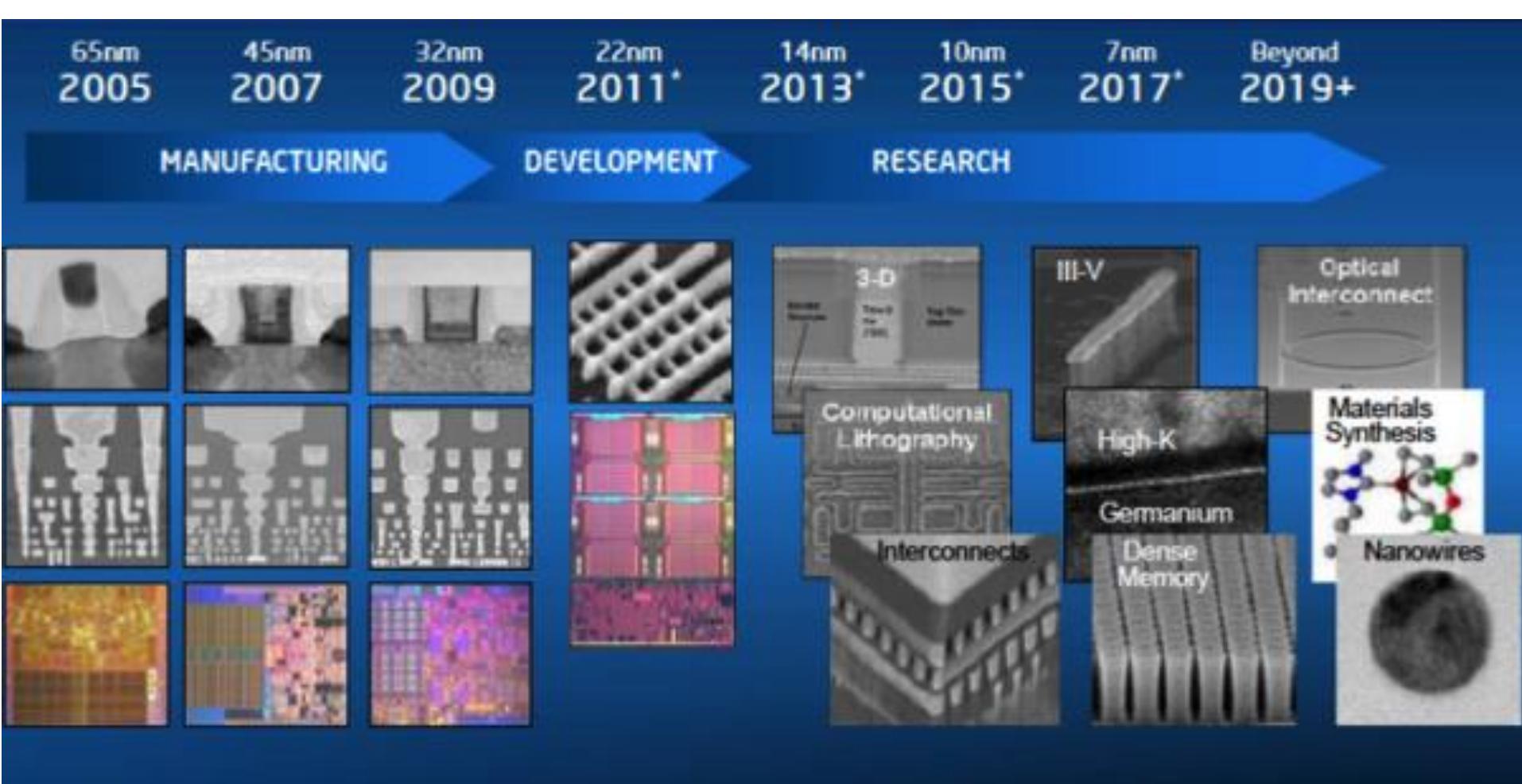
“More than Moore”

- ICs applications besides μPs
- Diversification: new applications
 - IoT, AI, health, etc
 - interactive (robots etc)



Semiconductor Devices: CMOS and Beyond

Intel



(23)
CMOS

3D Stacking

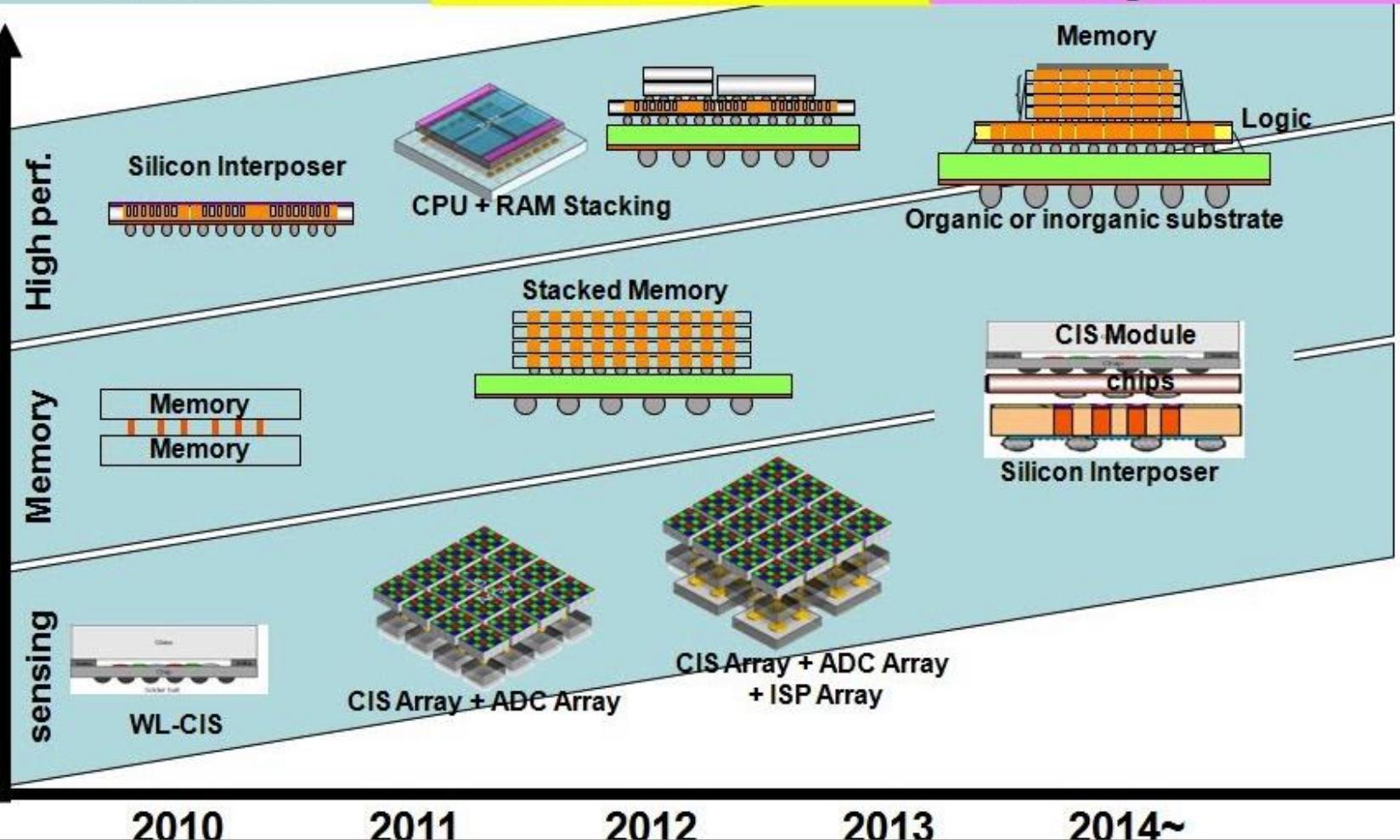
ITRI

Performance/complexity

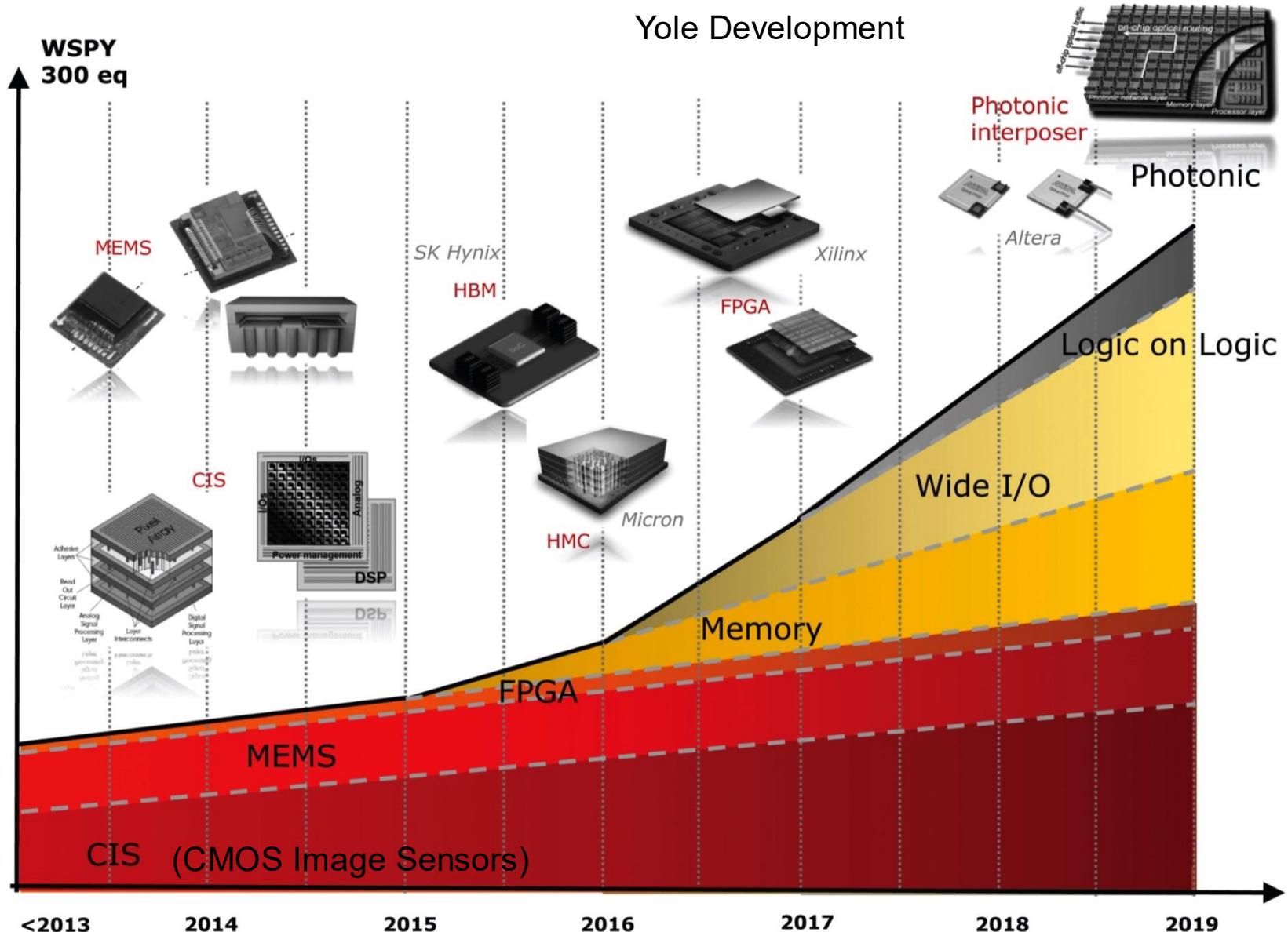
Key : Reliable baseline
Tech : Via last/middle
C2C/C2W

Capacity, DDR4, C/P Opti.
Via last/middle/first
C2W/W2W

High BW & speed (1TB/s)
Via last/middle/first
Cubic integration



3D Stacking Growth



Technology Scaling

Moore's law

- # of transistors doubles every two years
- Now at 1-2nm
- What is next?

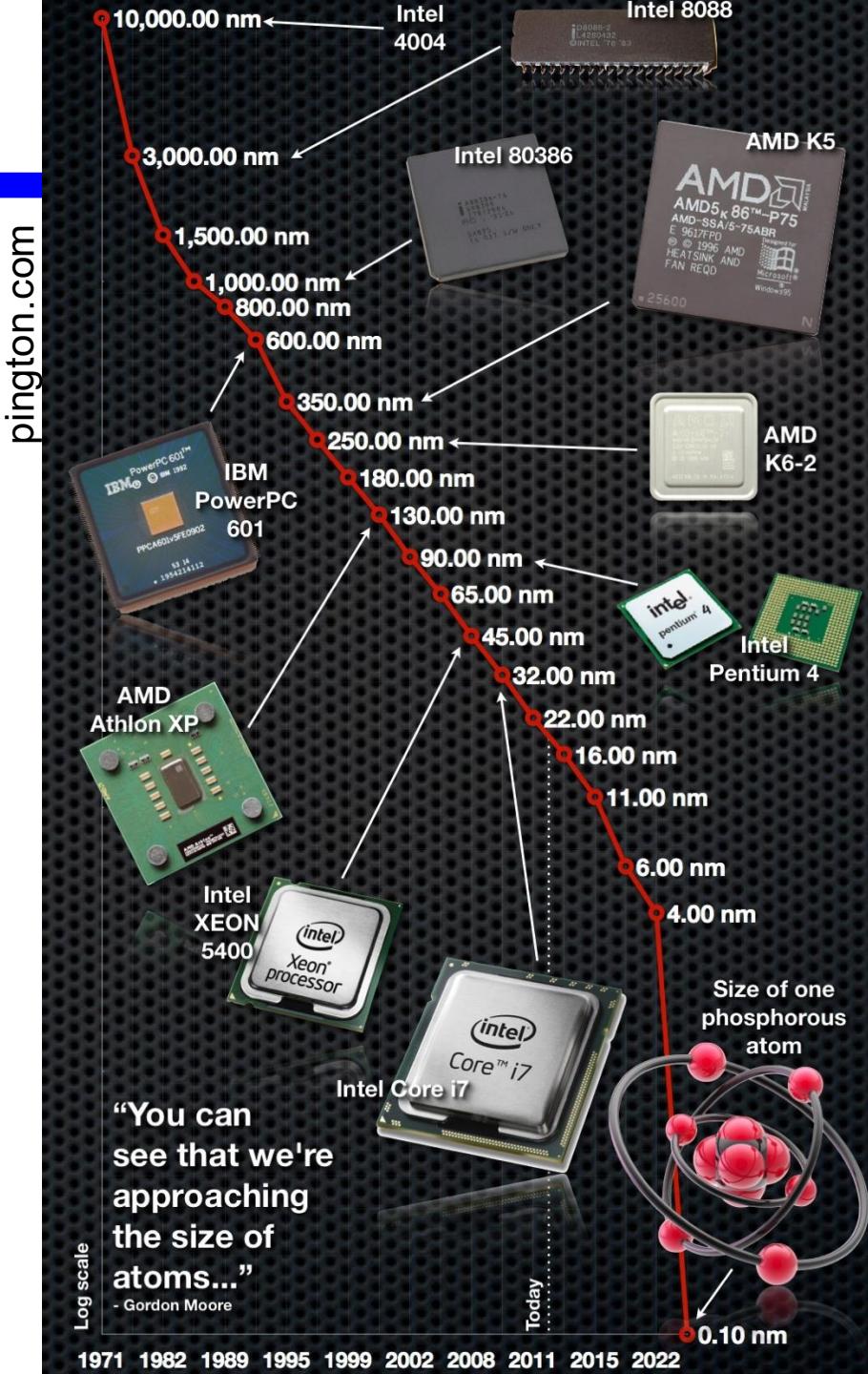
“More Moore”

- Beyond CMOS
- 3D stacking



“More than Moore”

- ICs applications besides μPs
- Diversification: new applications
 - IoT, AI, health, etc
 - interactive (robots etc)



Technology Scaling

Moore's law

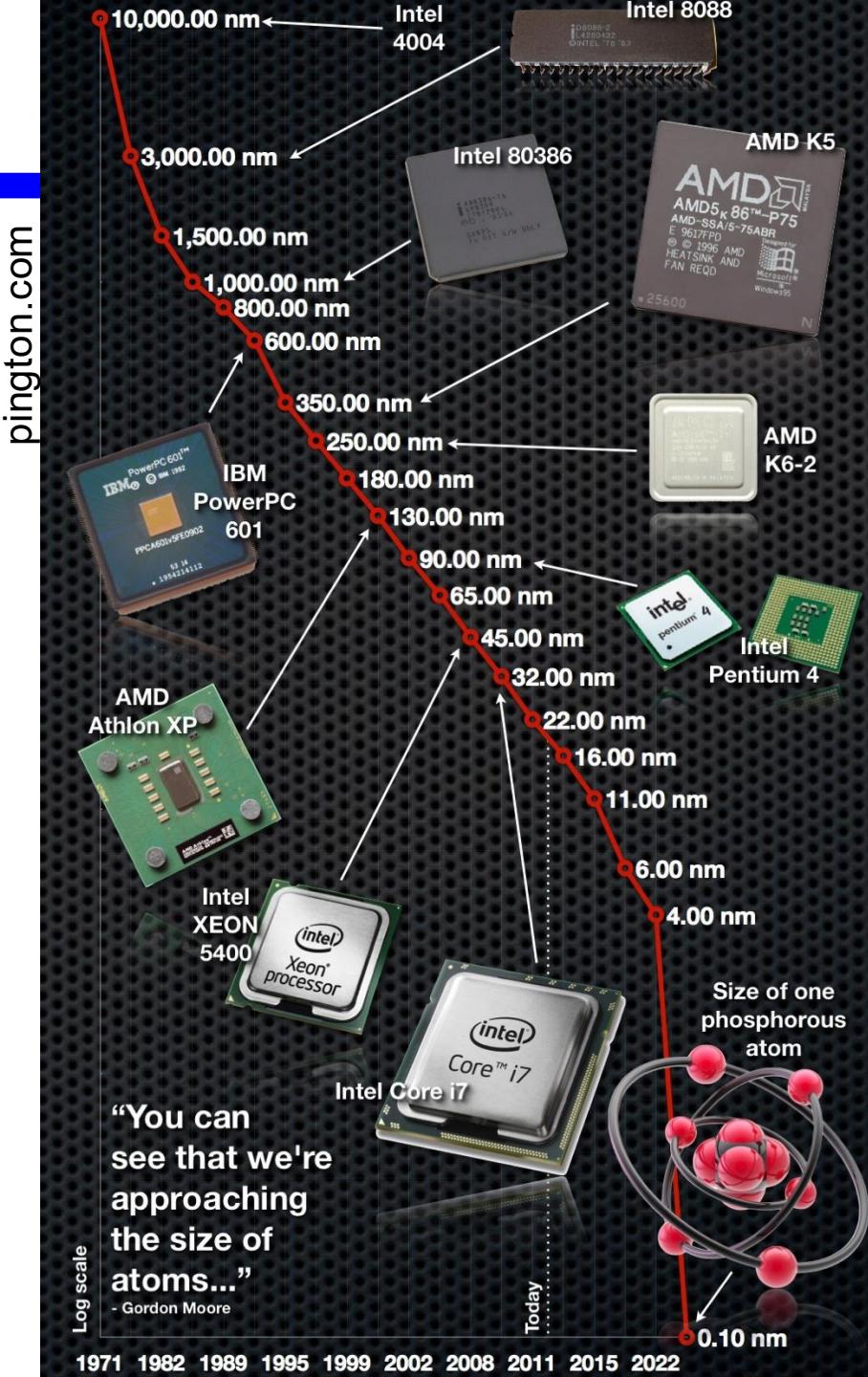
- # of transistors doubles every two years
- Now at 5-10nm
- What is next?

“More Moore”

- Beyond CMOS
- 3D stacking

“More than Moore”

- ICs applications besides μPs
- Diversification: new applications
 - IoT, AI, health, etc
 - interactive (robots etc)



Integrated Circuits Applications (besides μPs)

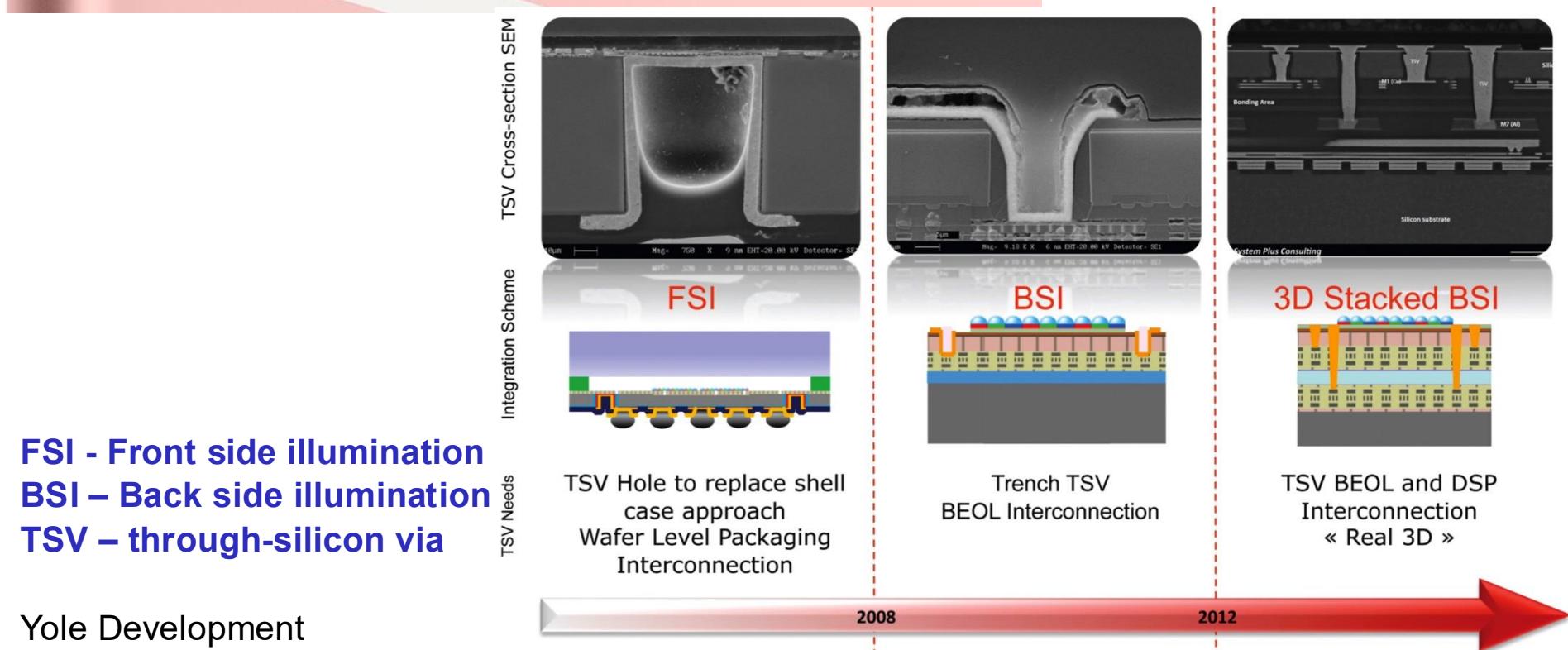
- Sensors / actuators
 - Image sensors
 - MEMS
- Communications
 - Wireless
 - Wireline
- Memories
 - Volatile (SRAM, DRAM)
 - Non-volatile (Flash)
- Emerging applications: interacting with people/environment
 - Acceleration of Artificial Intelligence / Machine learning algorithms
 - Biochips / medical / wellness (implantable, wearable, disposable)
 - New analog / RF
 - New sensors / actuators (MEMS, terahertz, ultrasound)



CMOS Image Sensors (CIS)



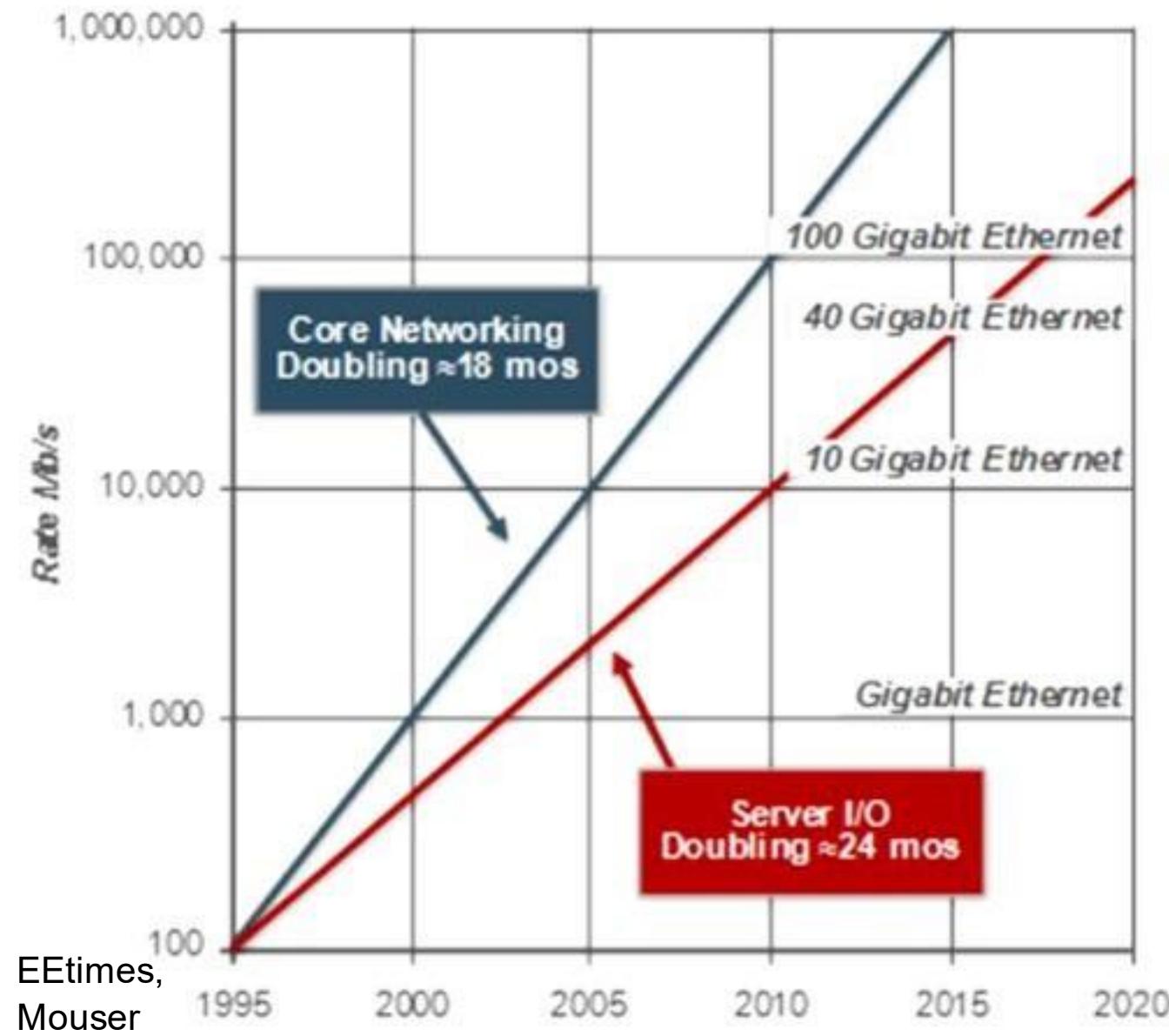
Apple



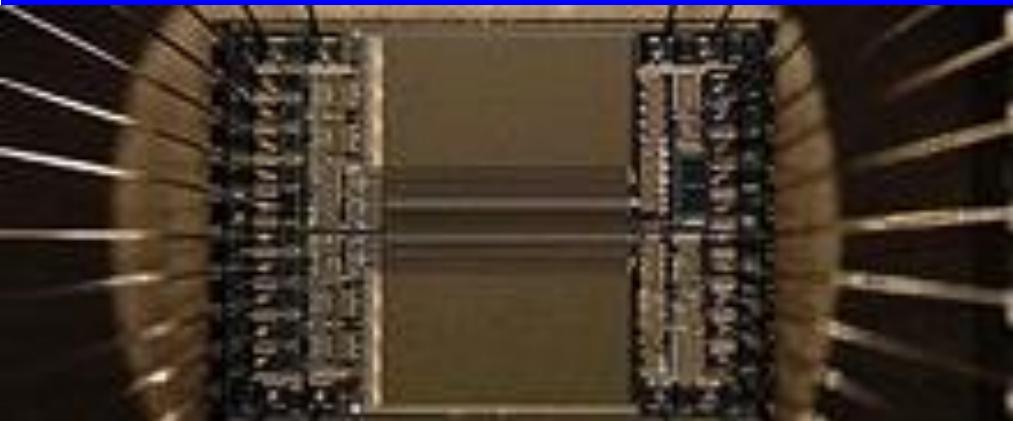
Wireless Communications



Wireline Communications (Cloud)

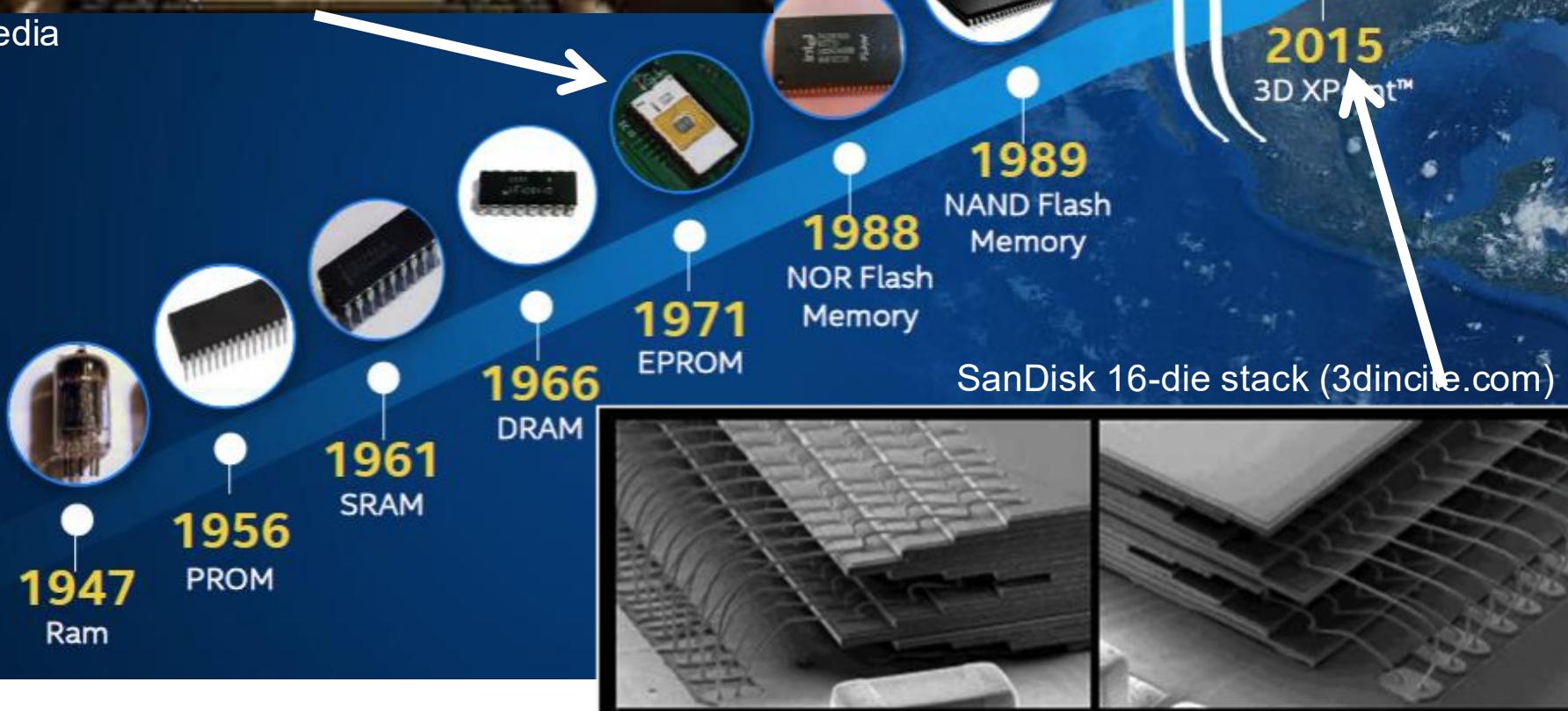


Semiconductor Memories



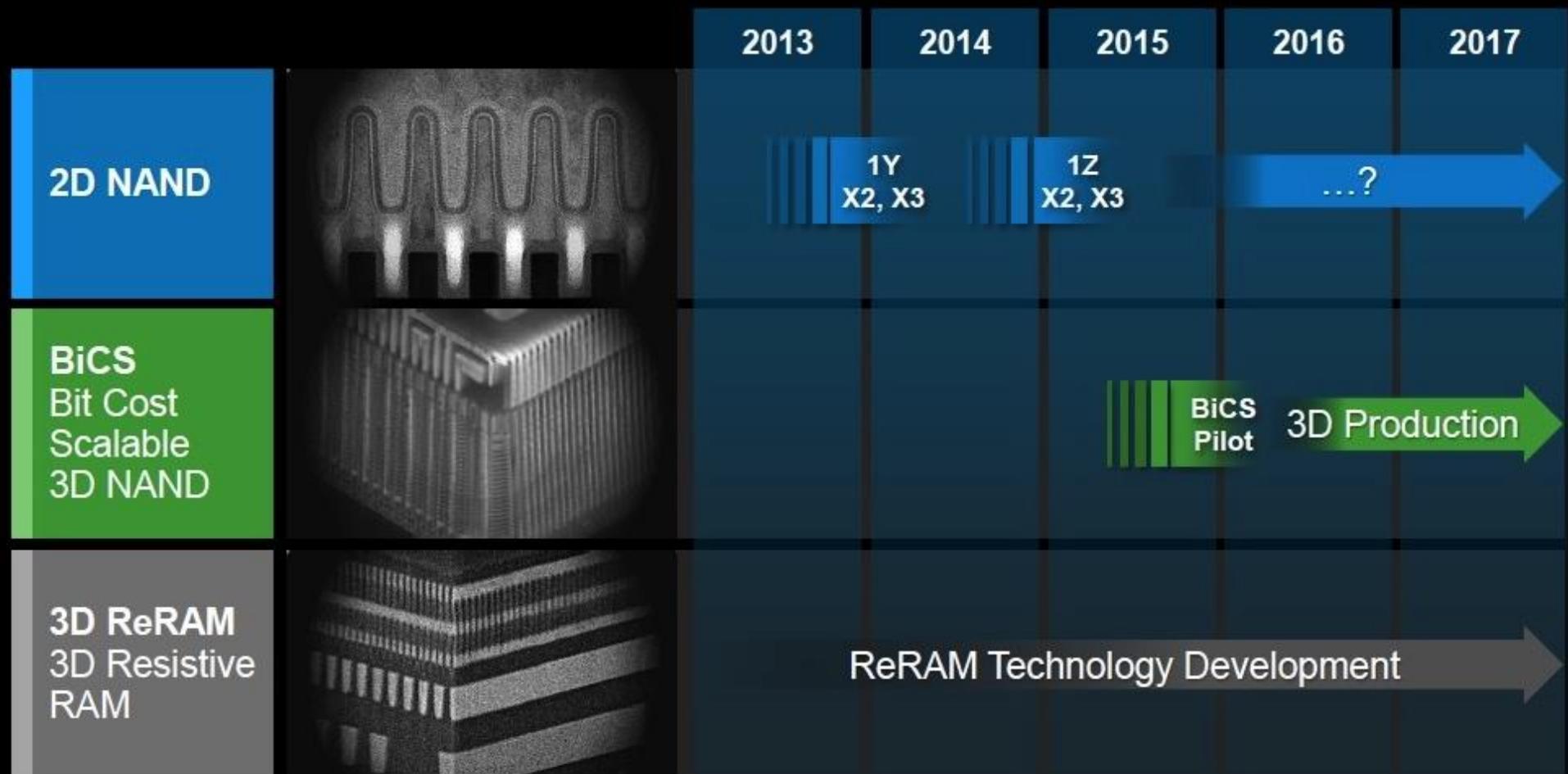
thememoryguy.com

Wikipedia



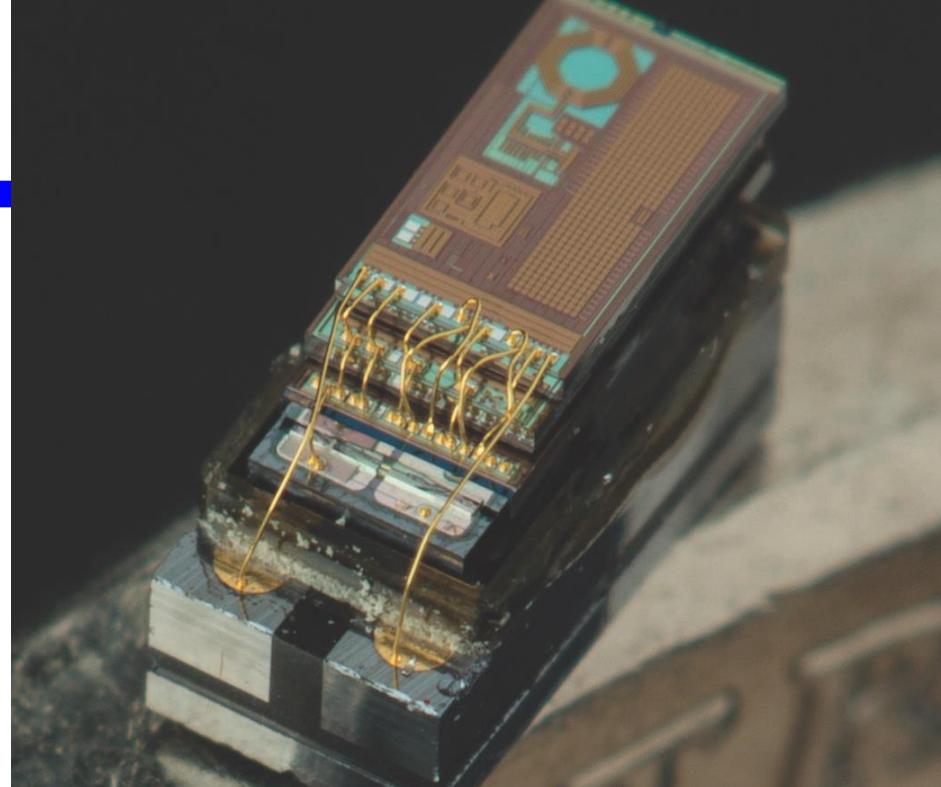
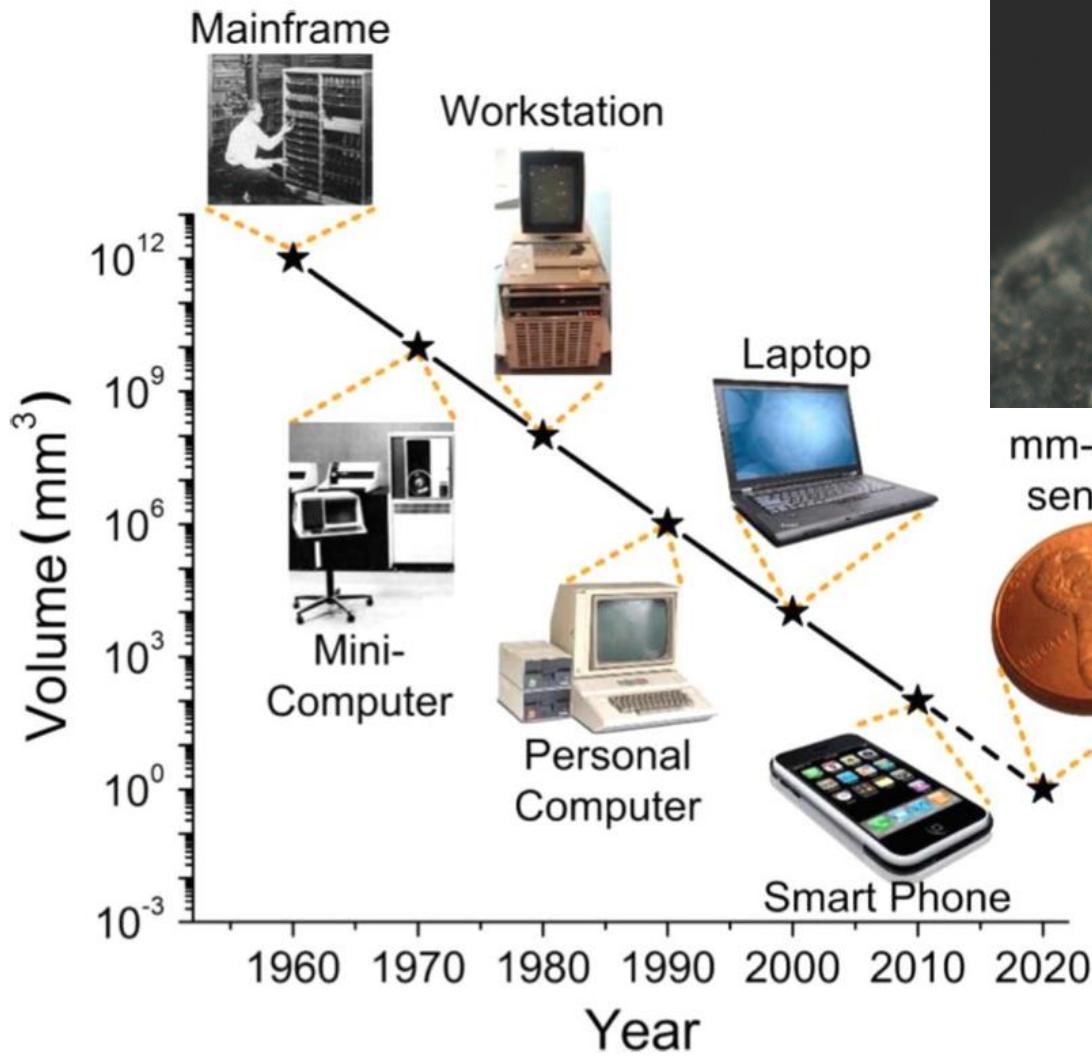
Example: Flash Memories

SanDisk Technology Roadmap

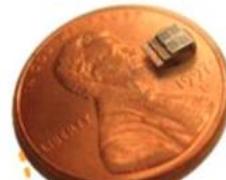


Bell's Law

- A new computer class forms about every decade



mm-scale
sensors

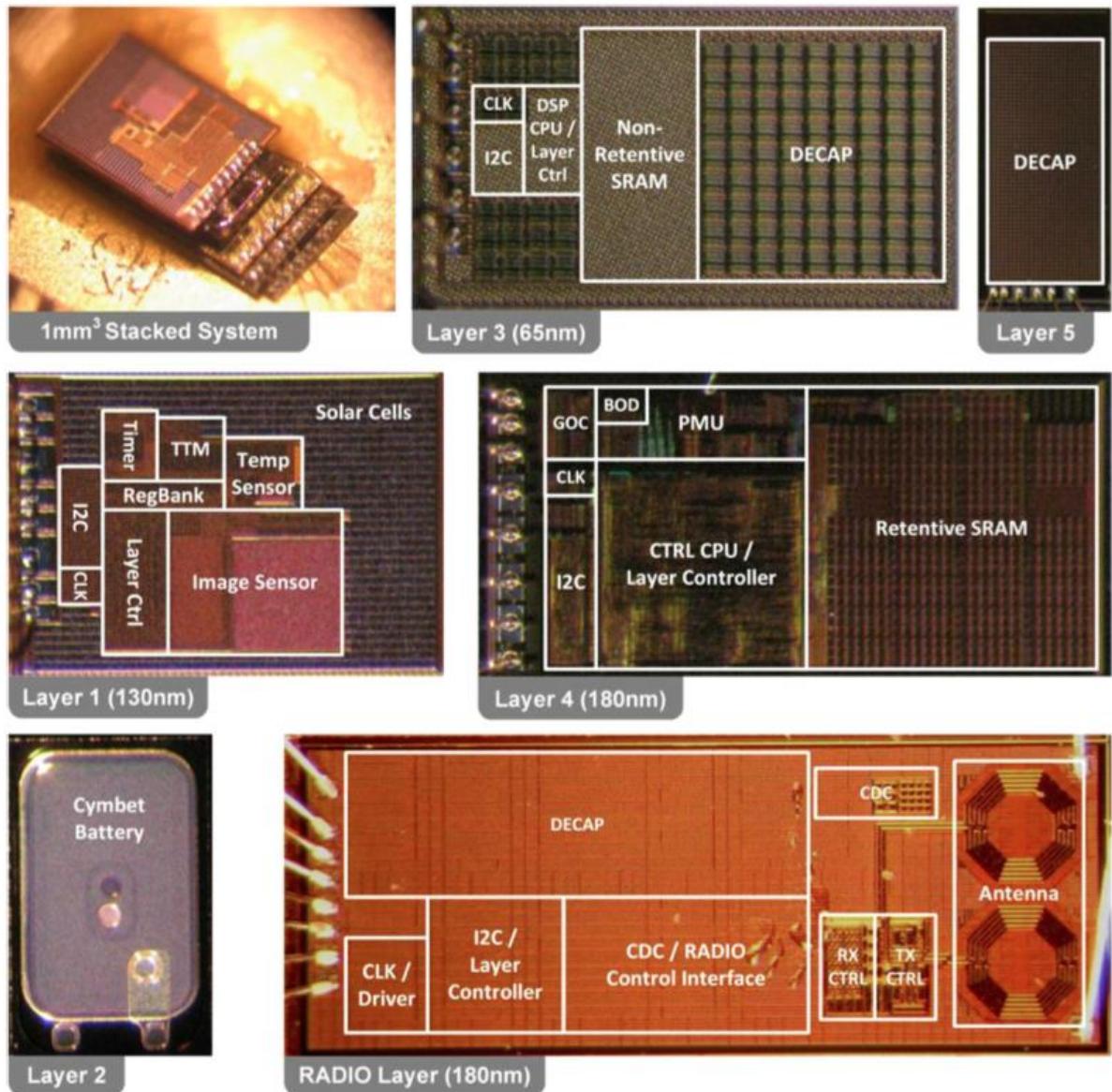
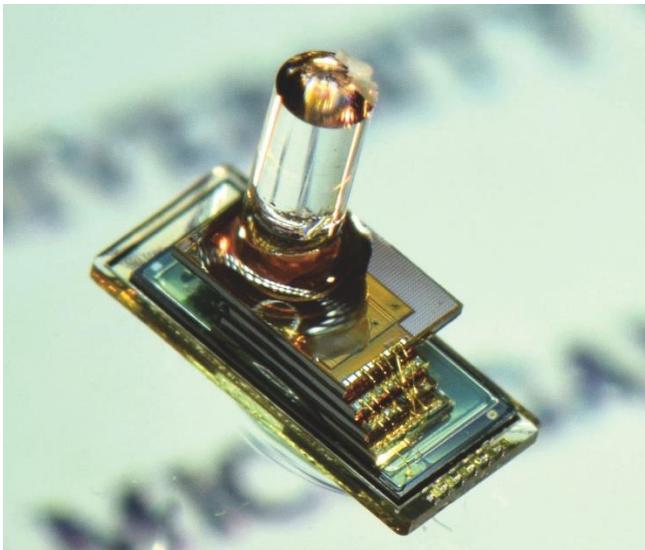


1mm³ temperature
sensor Profs. Blauuw
and Sylvester, U
Michigan

MICROSYSTEMS

Microsystems: Michigan Micro Mote

- Image Sensor
 - 1mm³



Technology Scaling

Moore's law

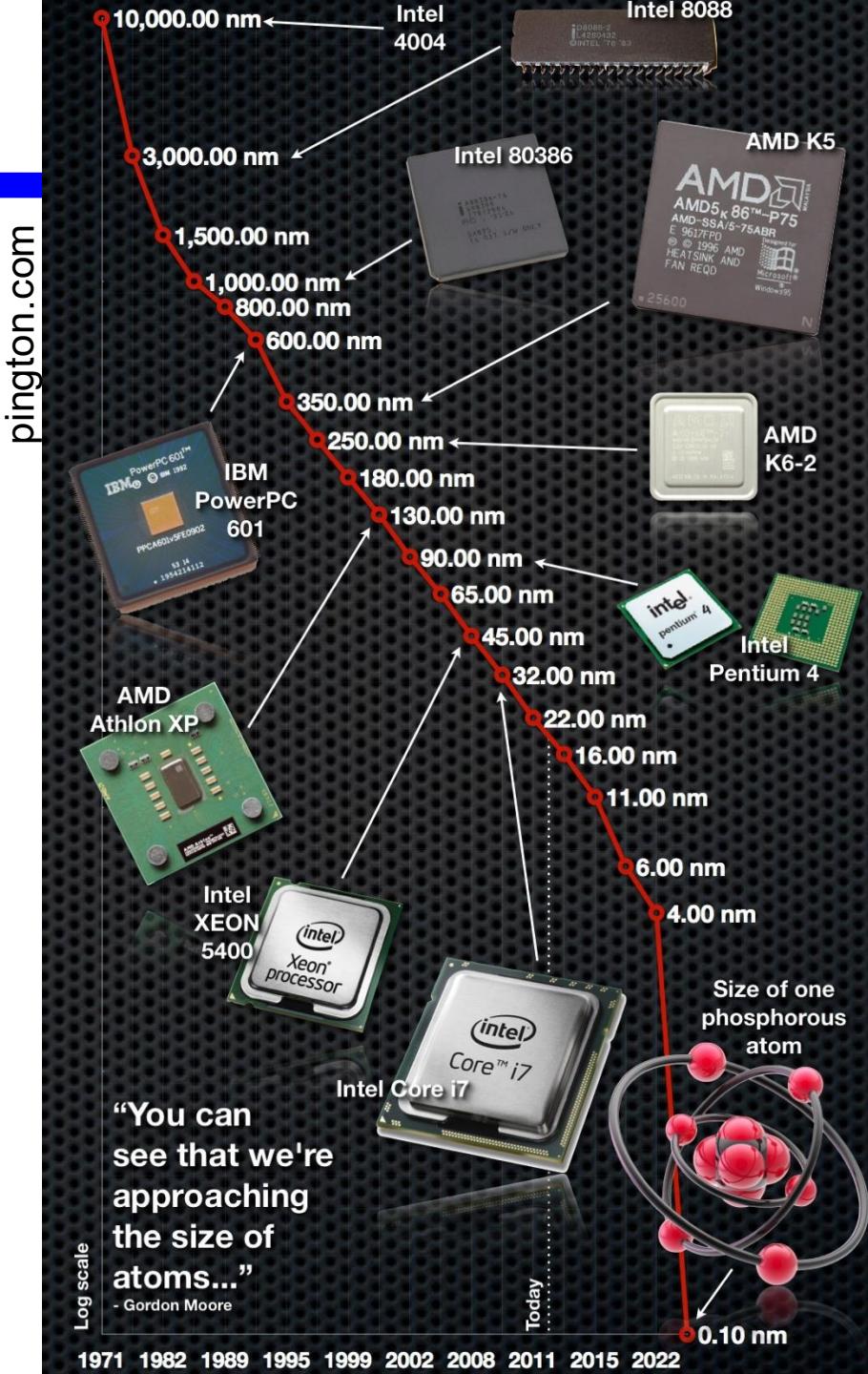
- # of transistors doubles every two years
- Now at 1-2nm
- What is next?

“More Moore”

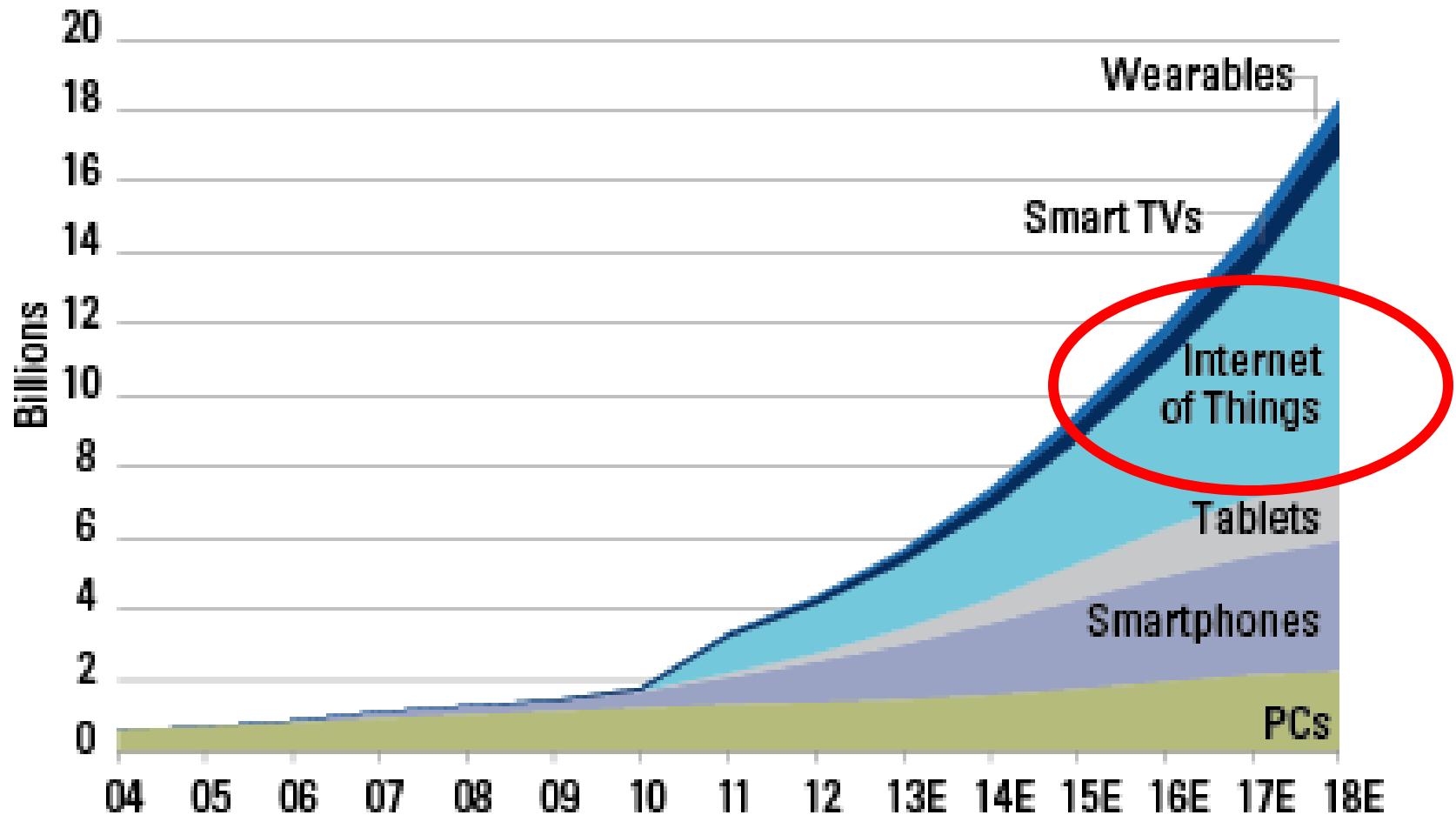
- Beyond CMOS
- 3D stacking

“More than Moore”

- ICs applications besides μPs
- Diversification: new applications
 - IoT, AI, health, etc
 - interactive (robots etc)

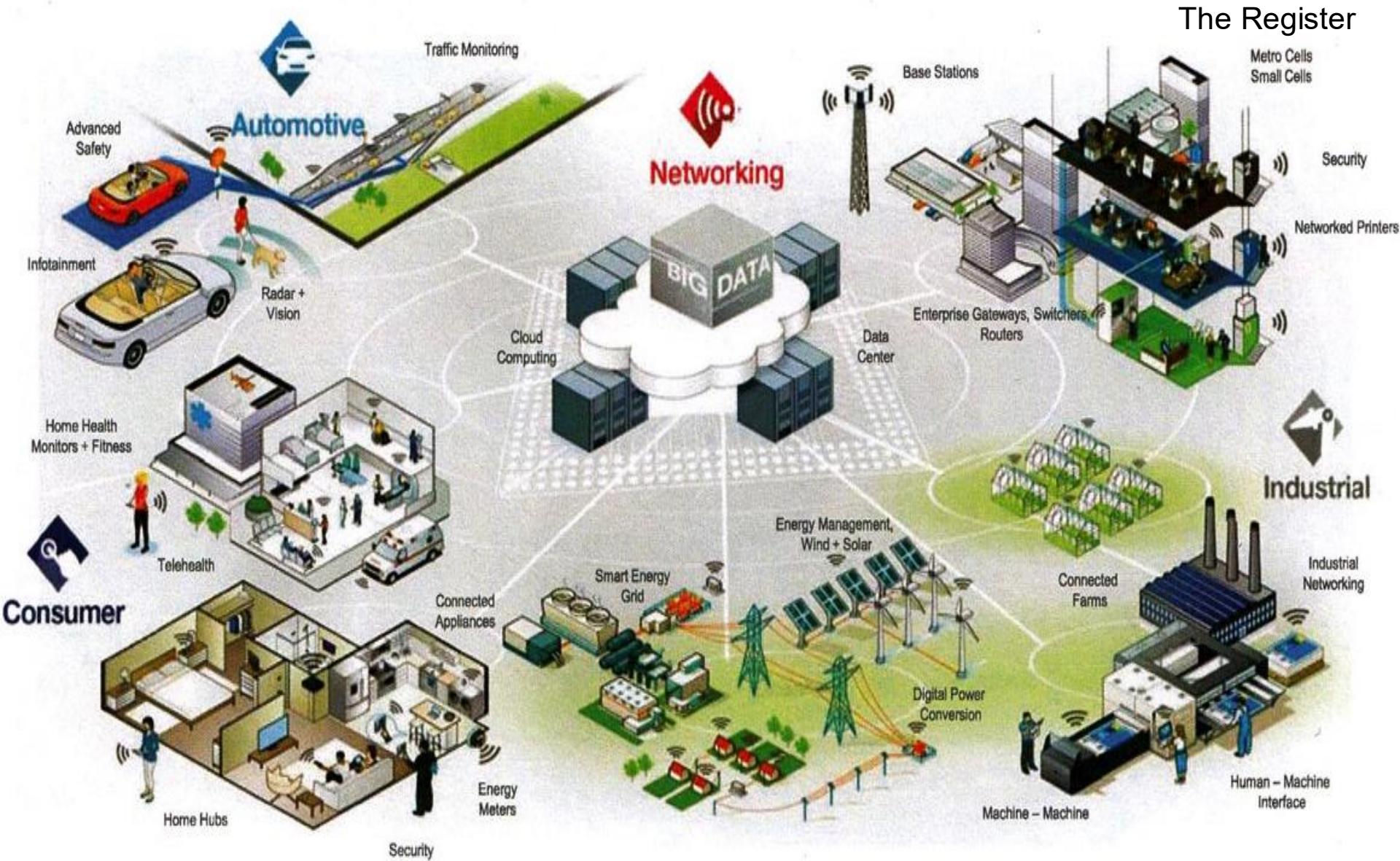


Future Markets: Internet-Connected Devices

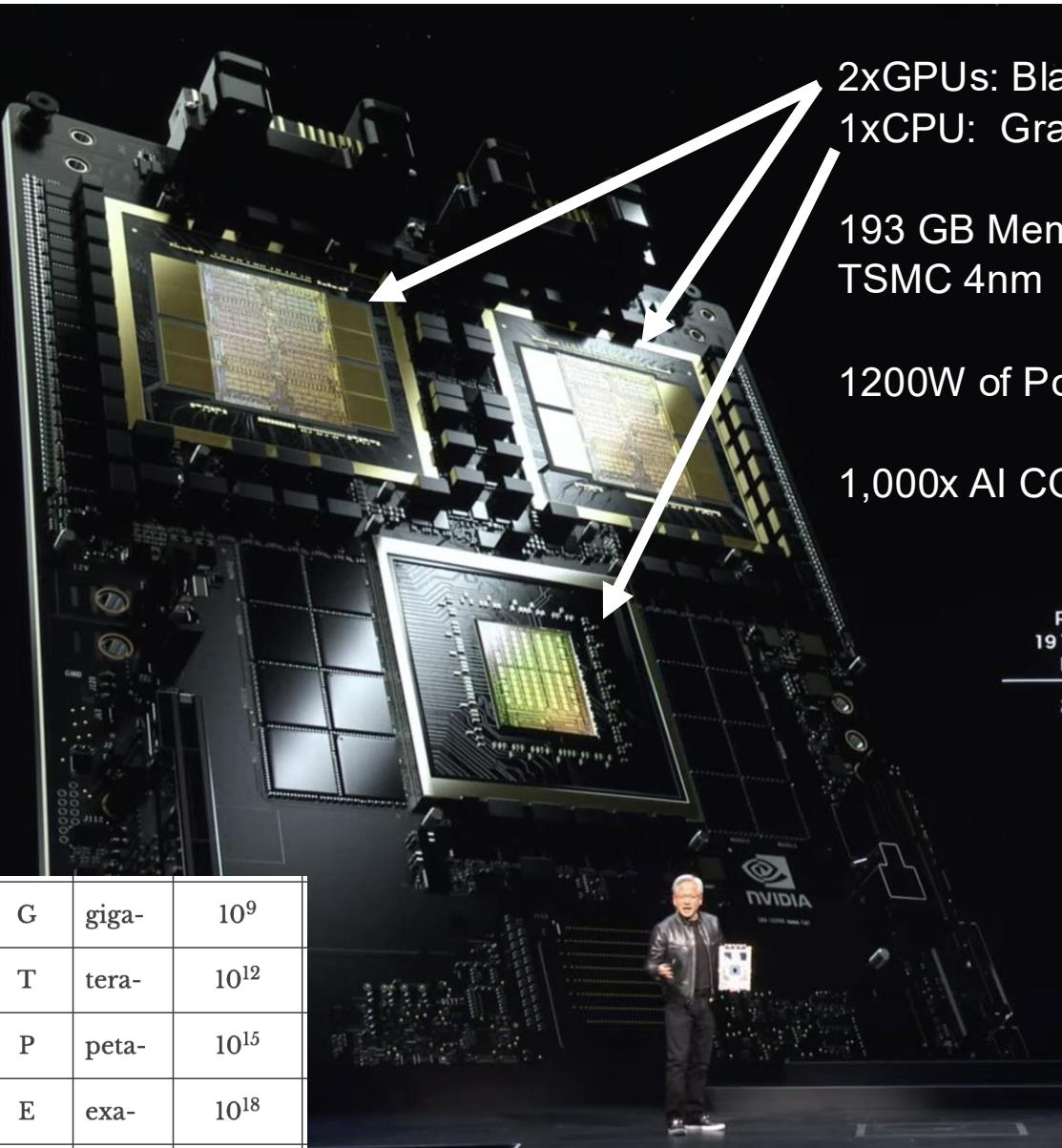


Sources: Gartner, IDC, Strategy Analytics, Machina Research, company filings, BII estimates

Future Markets: Internet of Things



NVIDIA: AI Chip GB200 Far Outpaces Moore's Law



2xGPUs: Blackwell 208 Billion Ts

1xCPU: Grace CPU

193 GB Memory, 4-bit floating point (FP4)
TSMC 4nm

1200W of Power → Liquid Cooling

1,000x AI COMPUTE IN 8 YEARS

Pascal 19 TFLOPS FP16

2016

Volta 130 TFLOPS FP16

2018

Ampere 620 TFLOPS FP16

2020

Hopper 4,000 TFLOPS FP8

2022

Blackwell 20,000 TFLOPS FP4

2024

72x GPUs Rack

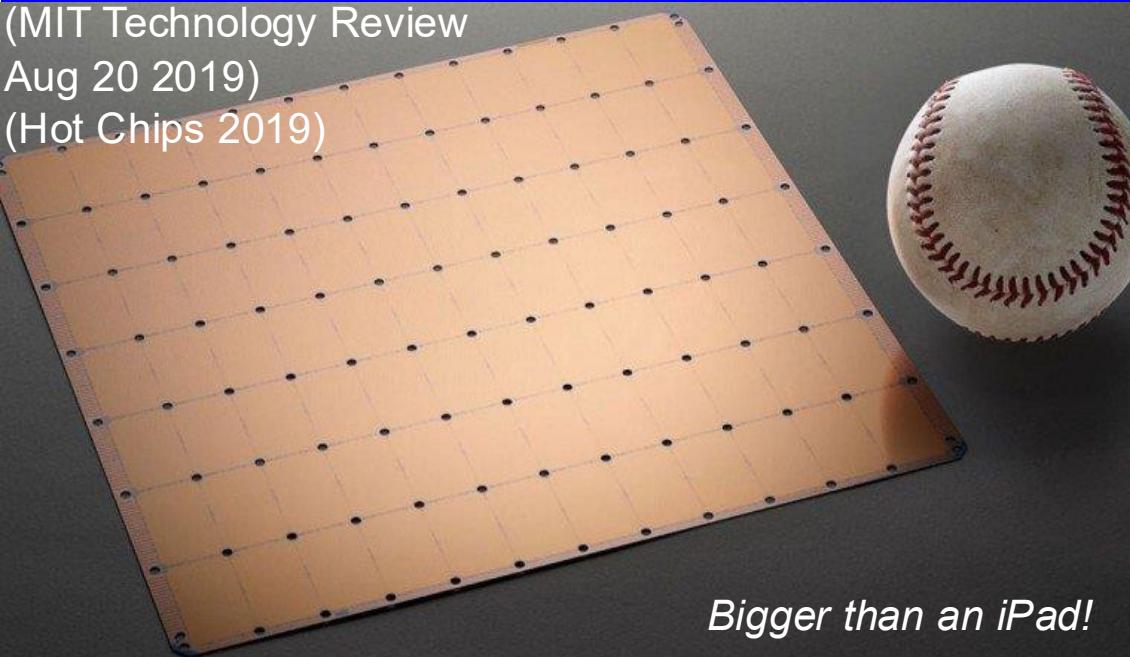
1,440 petaFLOPS
576 TB/s of GPU BW
10 trillion parameters

(NVIDIA)

G	giga-	10^9
T	tera-	10^{12}
P	peta-	10^{15}
E	exa-	10^{18}

World's Biggest Chip - Will Help Train AI

(MIT Technology Review
Aug 20 2019)
(Hot Chips 2019)



Cerebras Systems' Wafer-Scale Engine (16nm)

- 1.2 trillion transistors, 215 x 215mm², 400k cores
 - 40 TFLOPs (10¹²)
- Compared to Nvidia's largest GPU:
 - 57x bigger
 - 3,000x on-chip memory
- Saves power
 - by avoiding chip-to-chip communication
- Claim proprietary ways to improve yield
 - overcome defects / impurities

Compare to: world's fastest super-computer as of 2020:

- Fugaku supercomputer by FUJITSU
- 415 PFLOPs (10¹⁵)



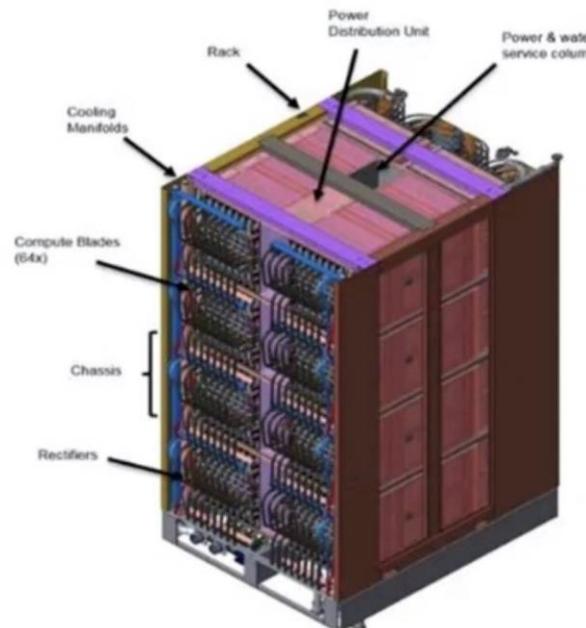
World's Fastest Supercomputer as of 2024

Frontier supercomputer at Oak Ridge National Laboratory



System

- 2 EF Peak DP FLOPS
- 74 compute racks
- 29 MW Power Consumption
- 9,408 nodes
- 9.2 PB memory
(4.6 PB HBM, 4.6 PB DDR4)
- Cray Slingshot network with dragonfly topology
- 37 PB Node Local Storage
- 716 PB Center-wide storage
- 4000 ft² foot print



Olympus rack

- 128 AMD nodes
- 8,000 lbs
- Supports 400 KW

All water cooled, even DIMMs and NICs

G	giga-	10^9
T	tera-	10^{12}
P	peta-	10^{15}
E	exa-	10^{18}

AMD node

- 1 AMD "Trento" CPU
- 4 AMD MI250X GPUs
- 512 GiB DDR4 memory on CPU
- 512 GiB HBM2e total per node
(128 GiB HBM per GPU)
- Coherent memory across the node
- 4 TB NVM
- GPUs & CPU fully connected with AMD Infinity Fabric
- 4 Cassini NICs, 100 GB/s network BW

Compute blade

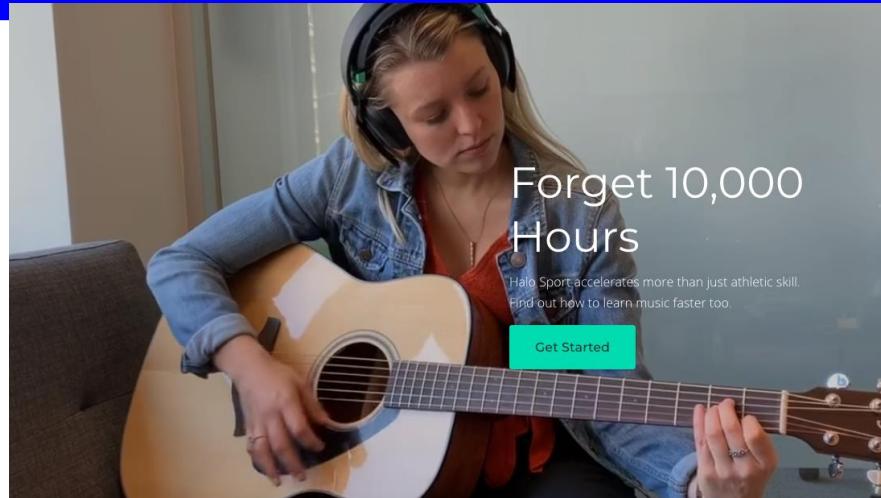
- 2 AMD nodes



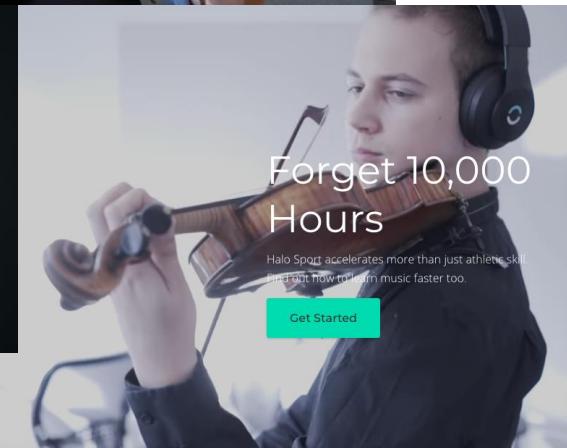
Microelectronics in Medicine and Wellness



- **FLOW**, startup in Sweden
- Headset to treat depression
 - Transcranial brain stimulation
 - Medication free



(MIT Technology
Review 2016)



- **Halo Neuroscience**, startup in Boston
- Headset for performance improvement
 - Transcranial brain stimulation
- Headphones could make you a better
 - Athlete, musician ...?

Technology Scaling

Moore's law

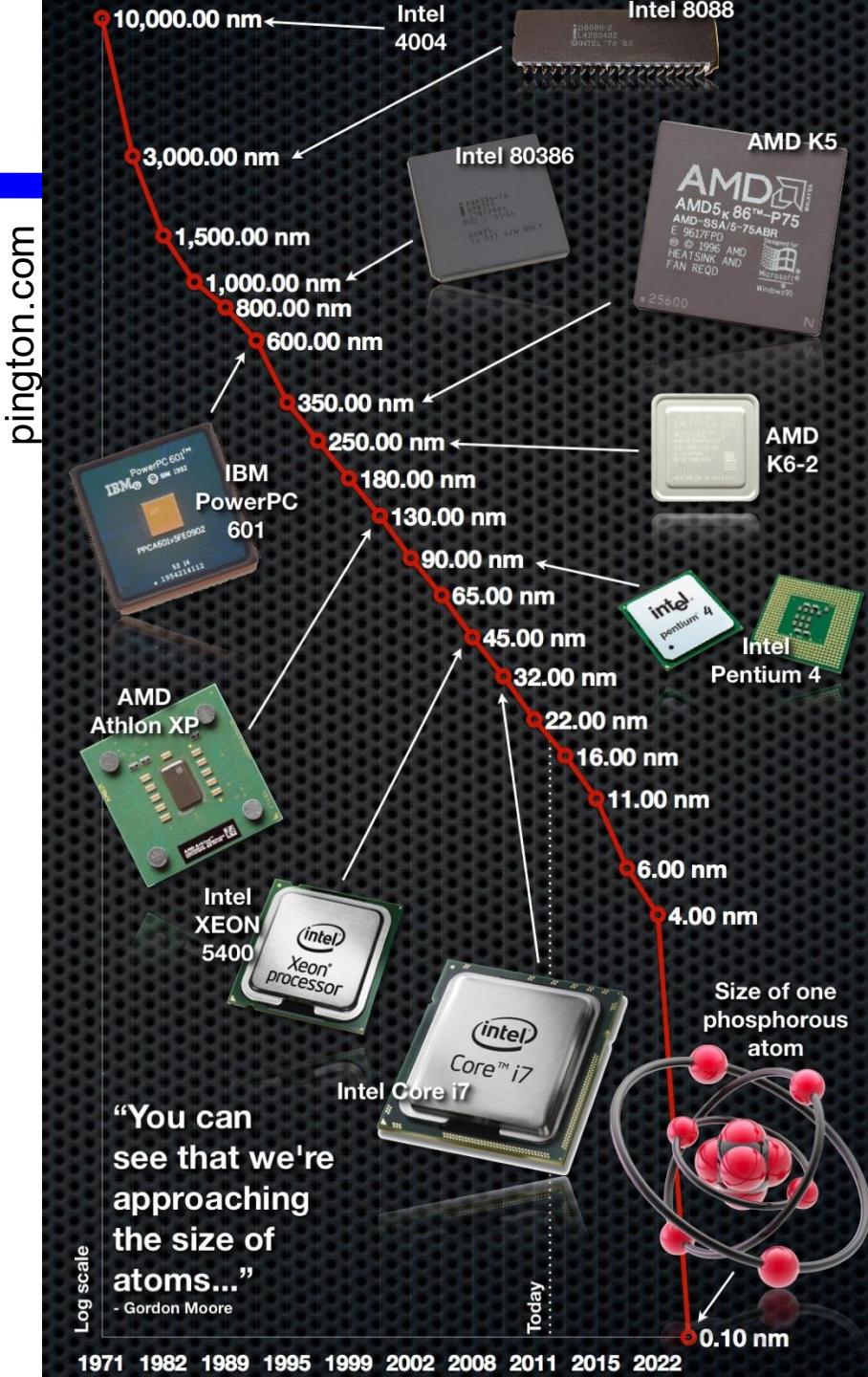
- # of transistors doubles every two years
- Now at 1-2nm
- What is next?

"More Moore"

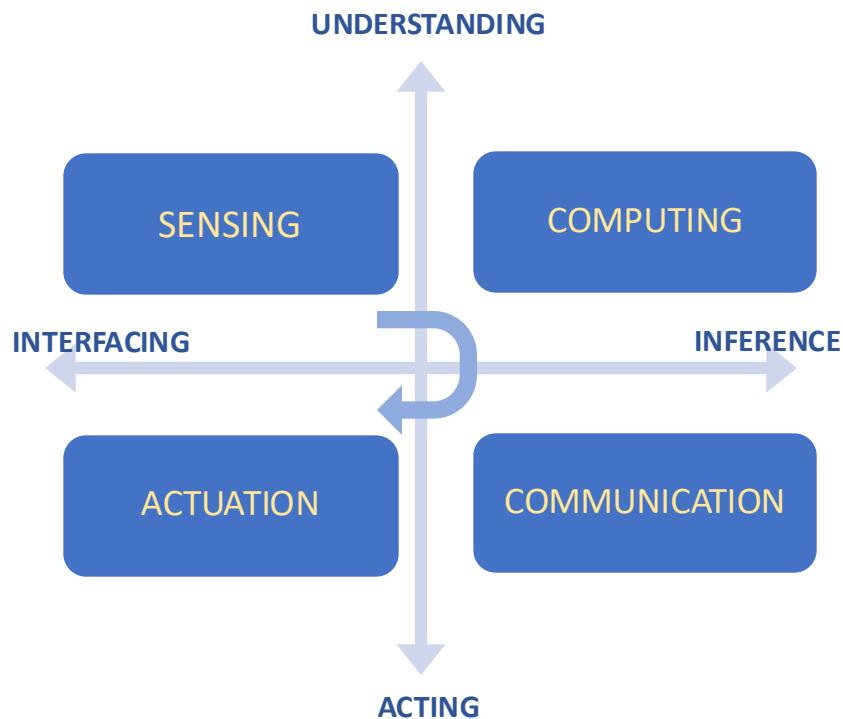
- Beyond CMOS
- 3D stacking

"More than Moore"

- ICs applications besides μPs
- Diversification: new applications
 - IoT, AI, health, etc
 - interactive (robots etc)



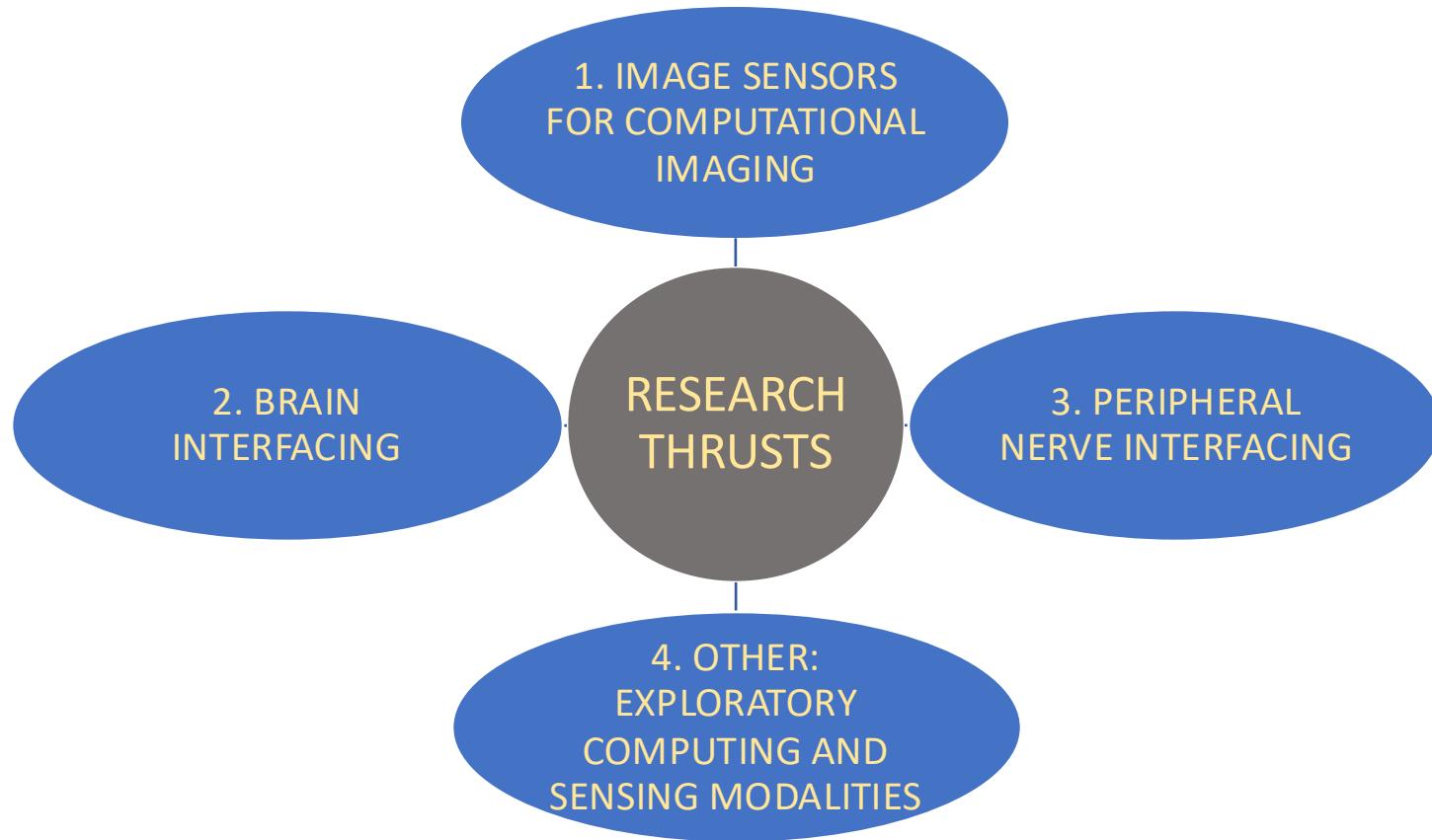
Genov Lab - Intelligent Sensory Microsystems Lab: Research Scope



NOVEL IC SOLUTIONS THAT ARE:

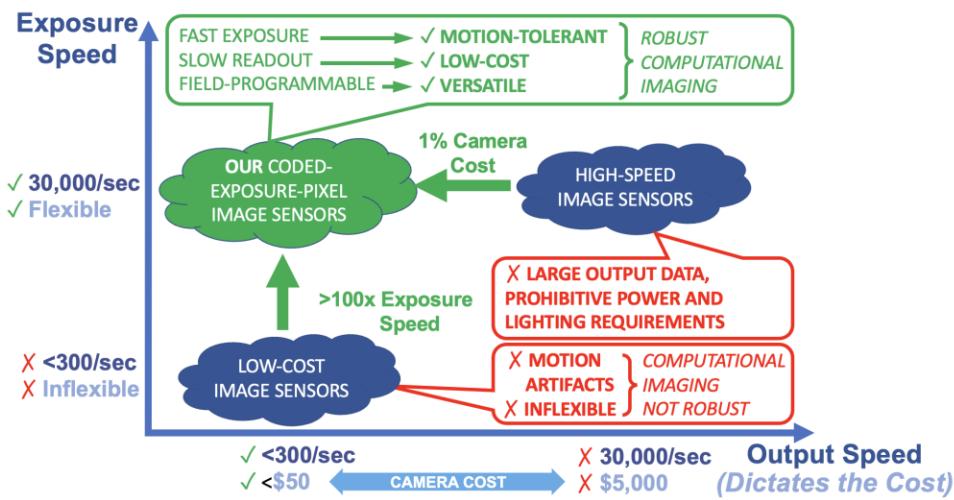
- **SYNERGISTIC**
 - Better than the sum of elements
- **ENABLING**
 - Energy-/cost-efficient, chip-scale
- **DISRUPTIVE**
 - New applications/markets

Genov Lab - Intelligent Sensory Microsystems Lab: Research Thrusts



Thrust 1: Fast Software-Defined Image Sensors for Versatile Computational Imaging - the ‘FPGAs’ of the Image Sensors World (ISSCC 19, 23; Symp. on VLSI 22, 24)

Motivation



Recent Sensor and Camera Prototypes

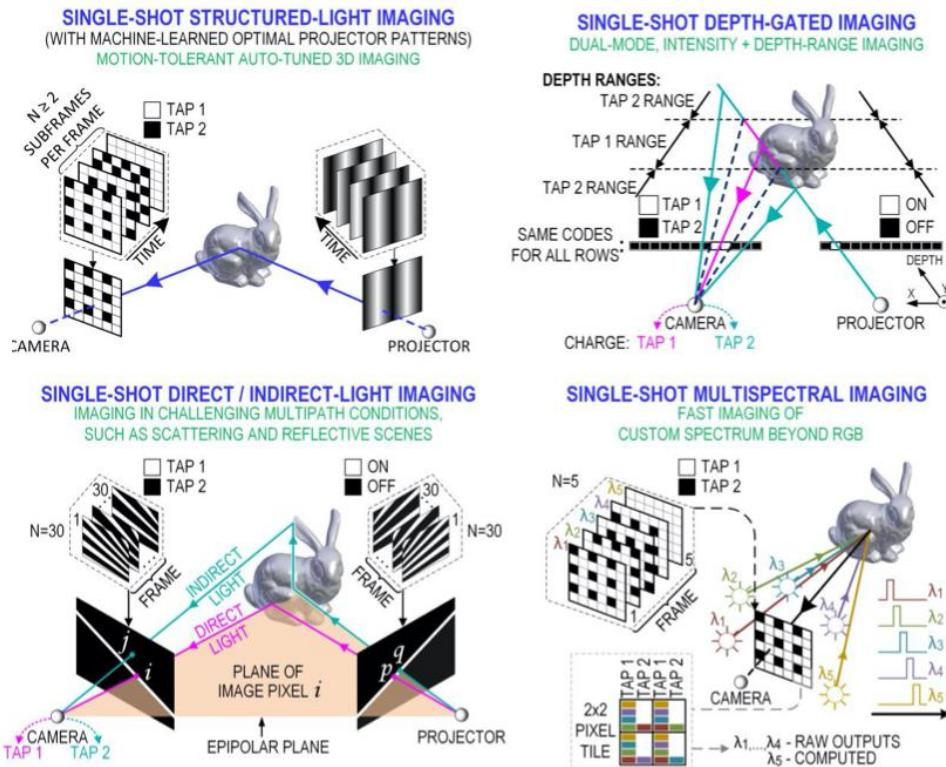
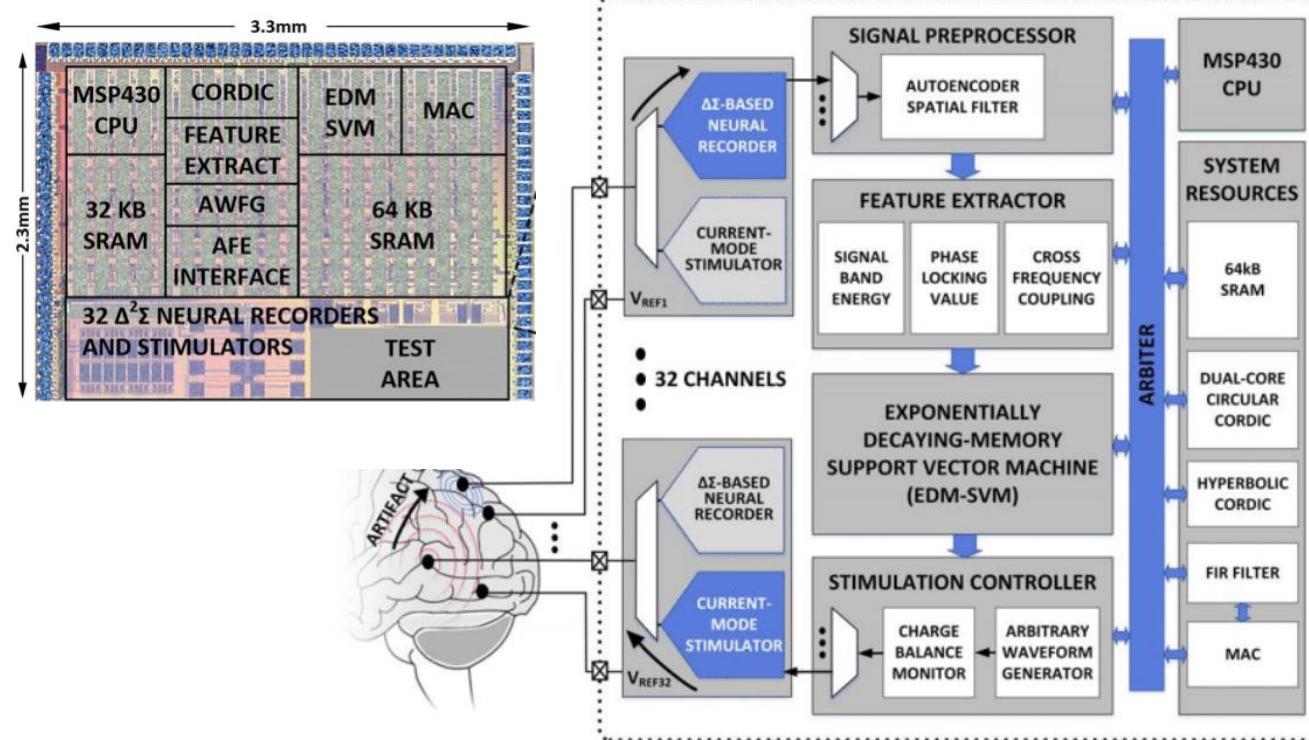
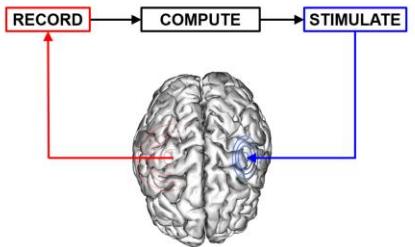


Image sensors with dynamically reconfigurable exposure time for every pixel: ie **coded-exposure pixels**

- Multiple fast image captures and in-pixel charge-based processing → reduced motion artifacts
- Single slow read out at video rate → power dissipation/data rate comparable to conventional imagers

Thrust 2: Artificially Intelligent Brain Implants for Automated Therapy (ISSCC 2017, 2018 x2, 2019, 2020)



Implantable brain implants to treat neurological disorders (eg epilepsy)

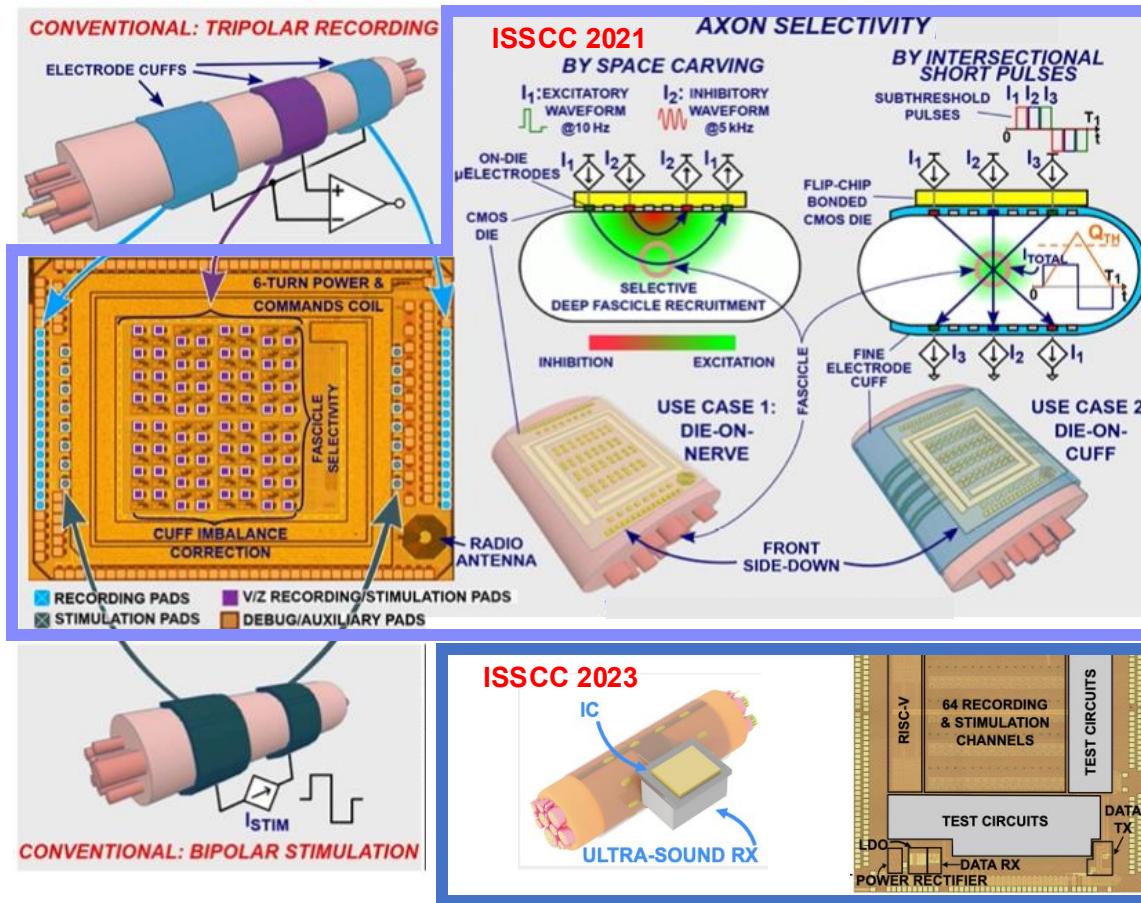
Brain-interfacing ICs:

- monitor brainwaves
- identify brain pathological brain states (such as epileptic seizures)
- trigger electrical stimulation to abort such states
- for automated neuromodulation therapy (e.g., epilepsy)

The designs includes:

- low-noise analog front-end
- wireless comm / power
- on-chip digital ML accelerator (SVM, decision trees)

Thrust 3: Smart Interfaces with Human Organs via *Peripheral Nerves* (ISSCC 2021, 2023)



Nerve-interfacing ICs:

- listen to PNS nerves
- interpret signals and
- alter brain-organ communication
- for automated neuromodulation therapy (e.g., spinal cord injury, chronic pain, autoimmune disorder, etc)

These designs include:

- low-noise analog front-end
- wireless comm / power
- wearable digital ML accelerator (CNN)

Research Opportunities



- **Participation in a research-oriented final project in ECE1388**
 - One or more such projects from my lab are available
 - Participation in a full-cycle IC design and prototyping, including tape-out
 - Can lead to a subsequent or concurrent MEng project ECE2500Y
- **ECE2500Y MEng projects are available** (worth 3 grad courses)
 - Apply by email any time - info @ <https://www.eecg.utoronto.ca/~roman/>