Dominant growth directions in integrated ciruit technology: 3D device architectures, 3D heterogeneous integration and Silicon Carbide

Andrzej J. Strójwąs

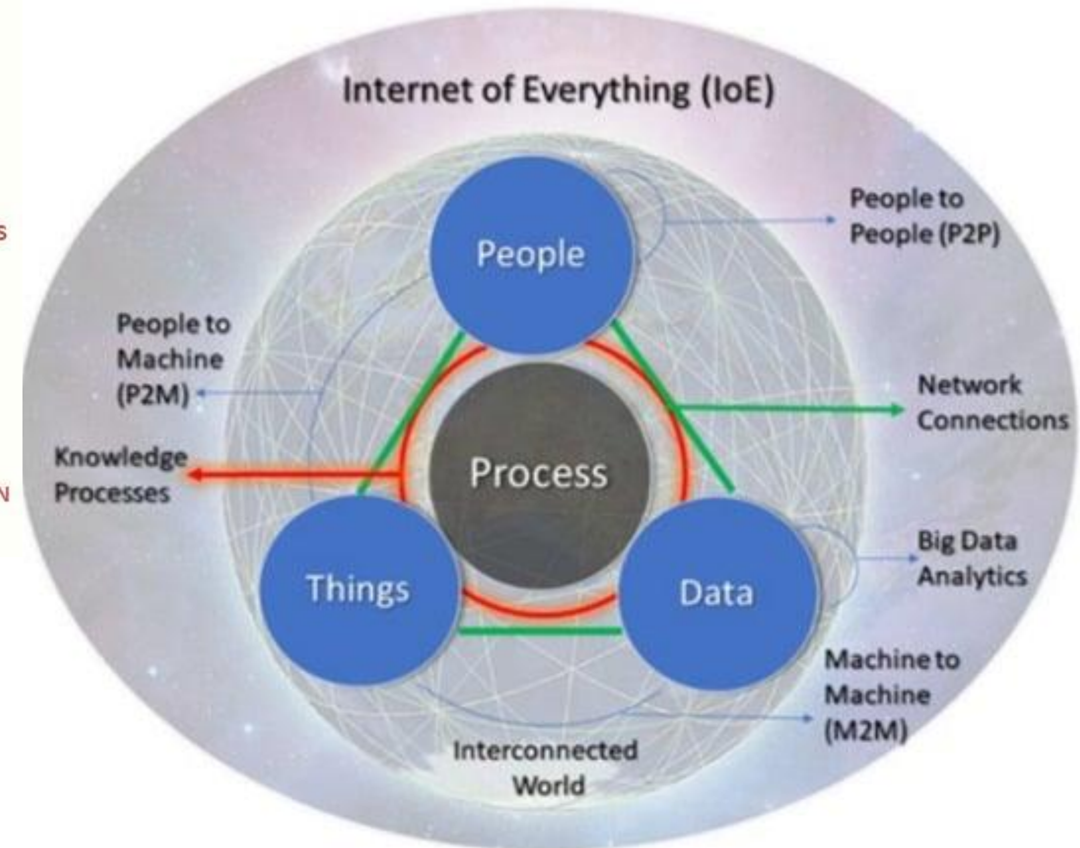
PDF Solutions Inc. and Carnegie Mellon University

ELTE 2023, Ryn, 19 kwietnia 2023

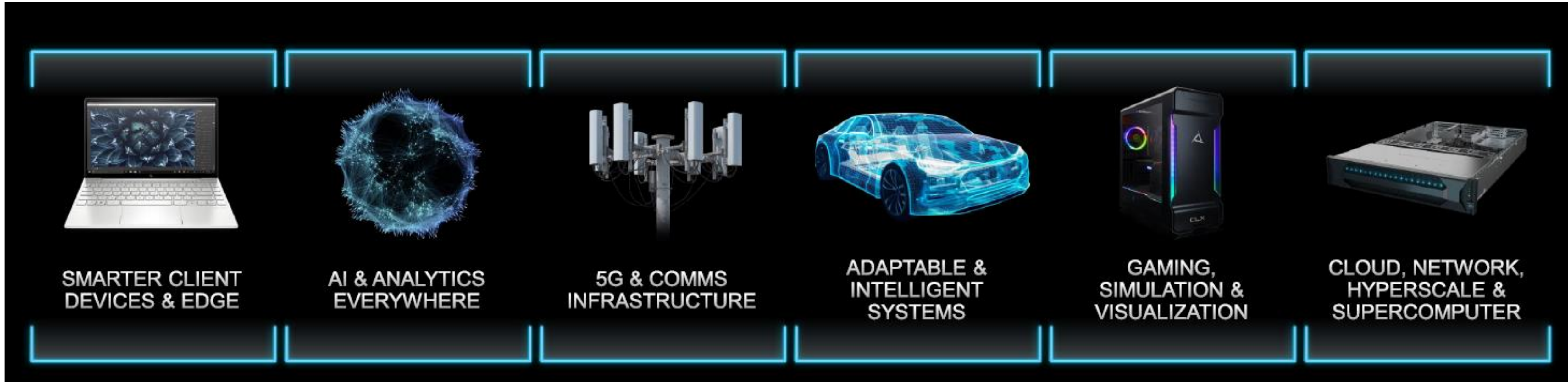
Outline

- ➢ Current technology landscape and future growth directions
- ❖ 3D device structures from FinFET to nanosheet architectures and beyond Silicon
- ❖ 3D heterogeneous integration: chiplets and neuromorphic computing
- ❖ Electrical car driven SiC market explosion

From Inetrnet of Things to Internet of Everything



Infrastructure Requirements

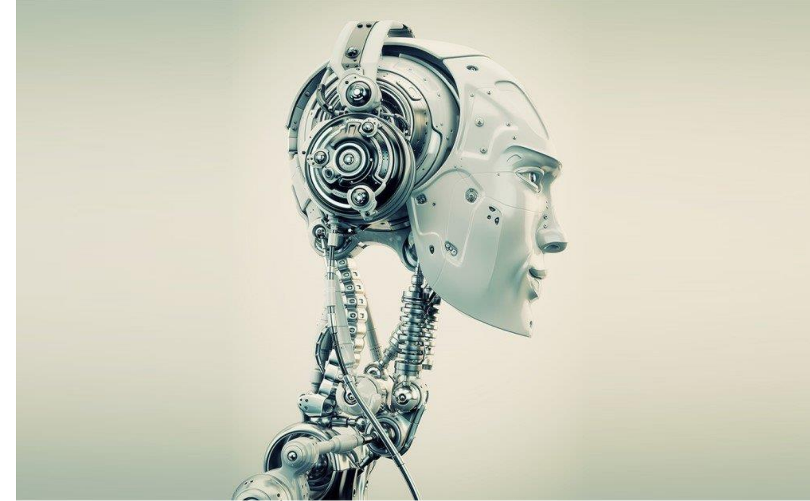


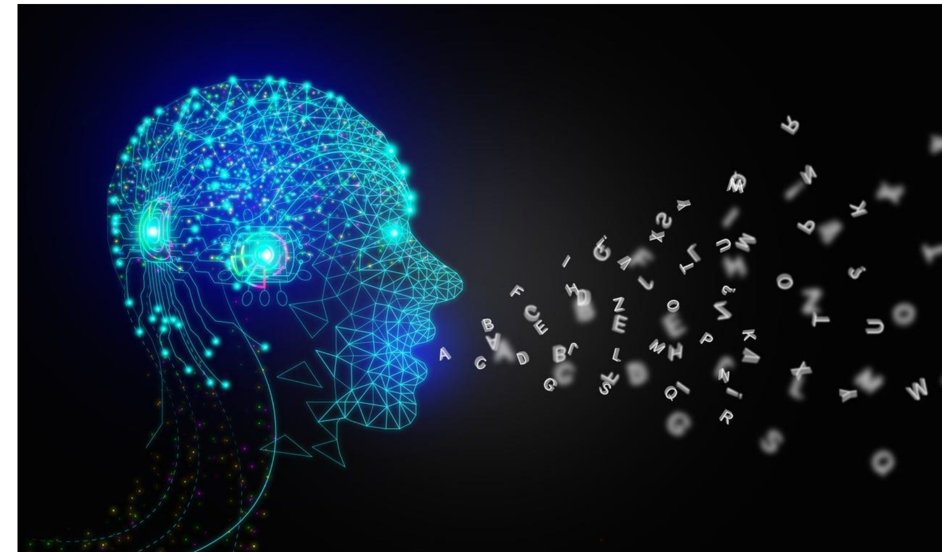
- **Data Availability/Connectivity (5G/6G)**
- **Artificial Intelligence Algorithms**
- **Computing Power and Massive Data Storage**
- **Data Analytics**
- **Intelligent System Implementation (cars, cities,...)**

Source: M.Mayberry,_Intel,_VLSI 2020

Artificial Intelligence Methods

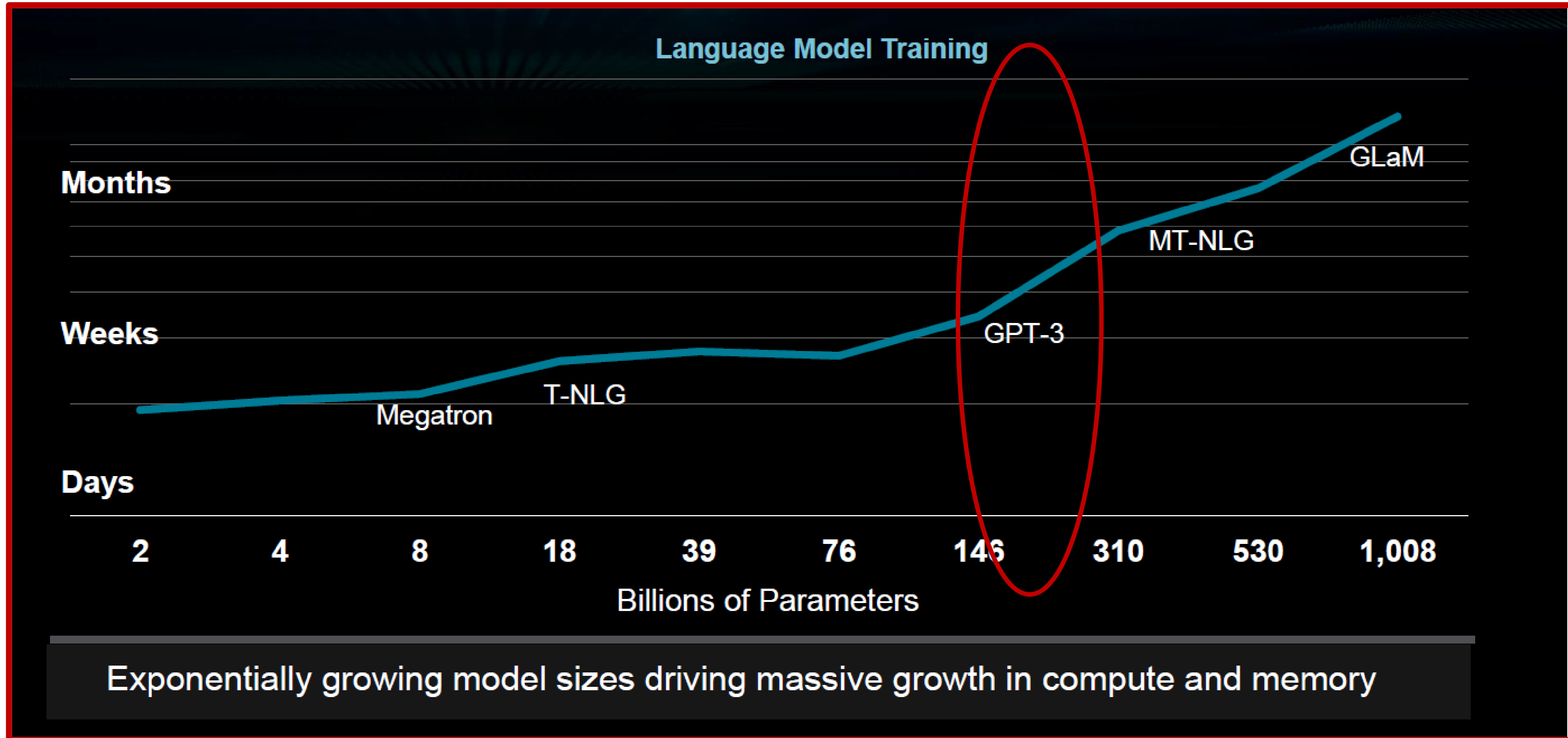
- **Machine Learning – Without Human Intervention**
- **Deep Learning – Deployment of neural networks on huge data sets**
- **Generative AI – Creation of new data, text, audio, video. Text is generated in natural language using Large Language Models (LLM)**
- **ChatGPT – Most popular system nowadays. Can answer any questions, pass any exam, write a poem, compose a song, create an impressive painting, etc.**





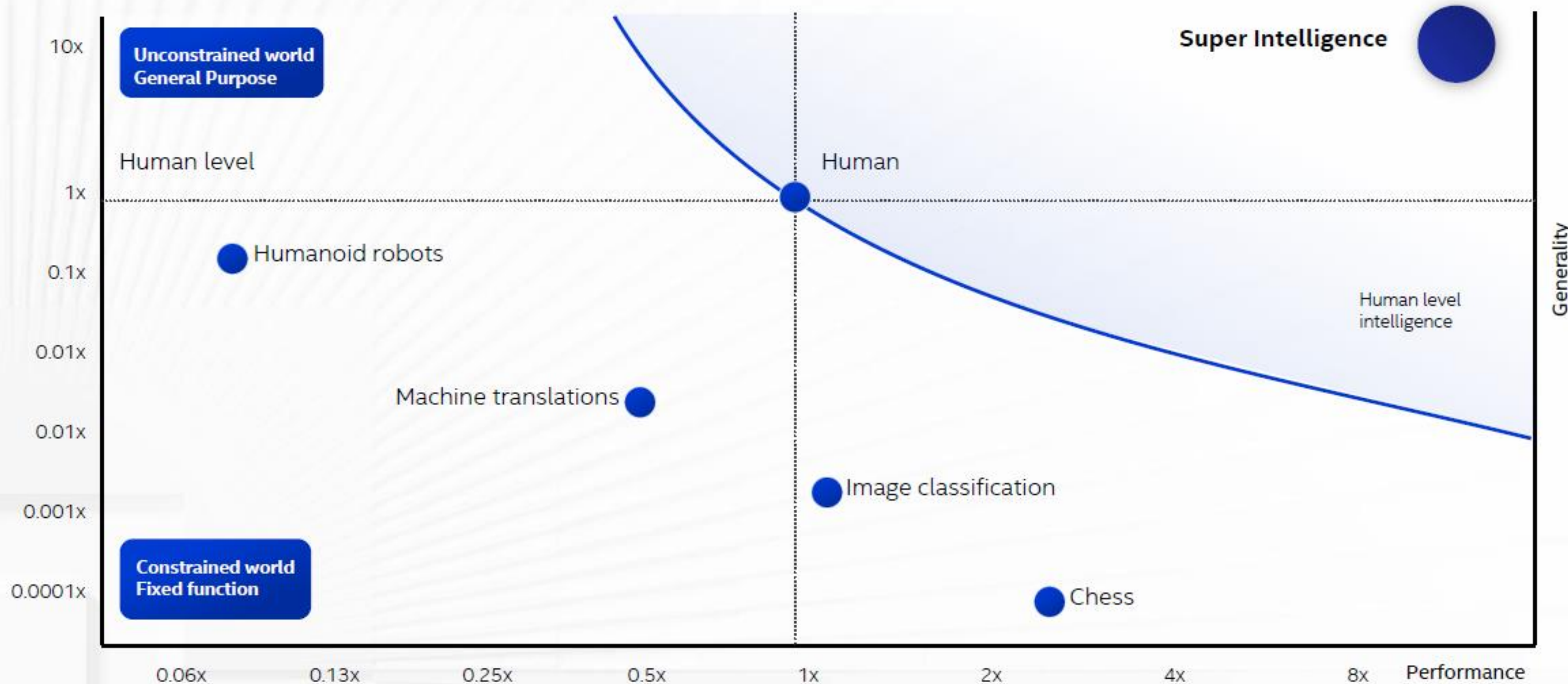
But… What are the computing and storage needs to train such a system?

Training Needs of the Large Language Models



Source: L. Su AMD, ISSCC 2023

Artificial Intelligence Evolution

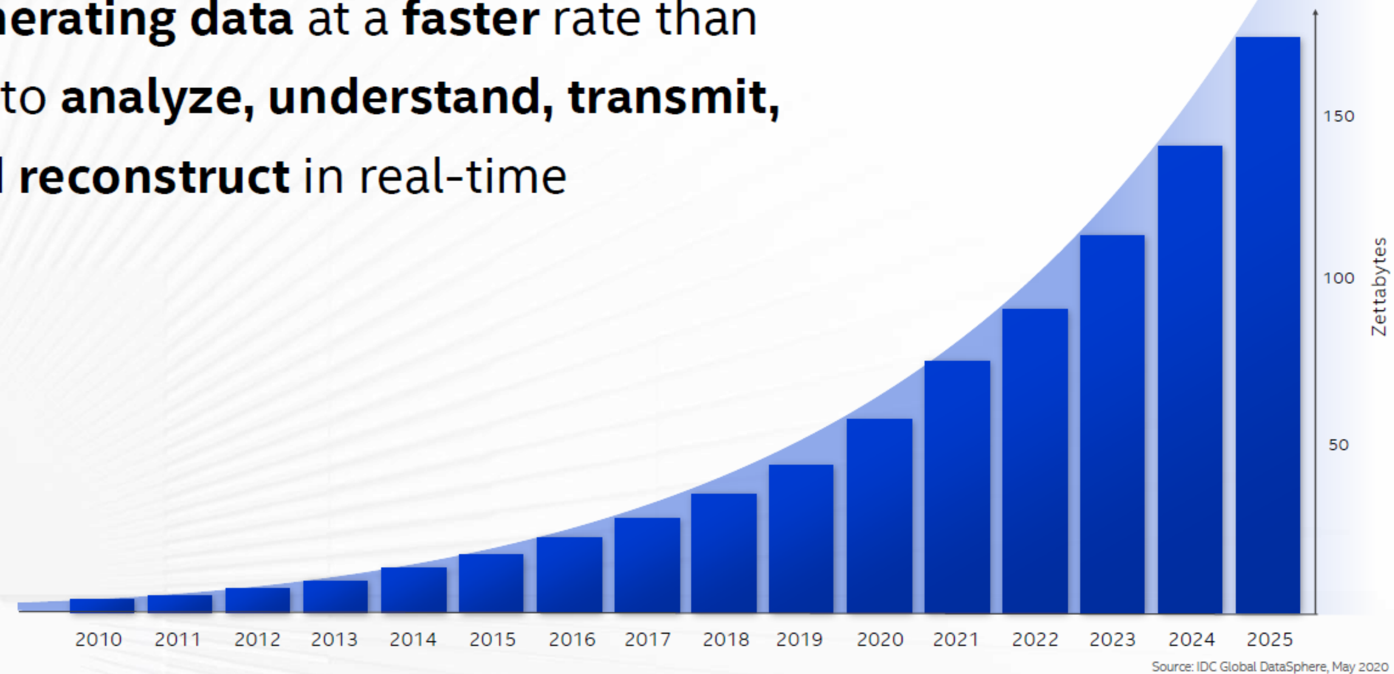


How do we approach human brain capabilities?

- So far successes only for the application specific systems (robotics, image classification, chess,...
- Achieved with huge compute and storage resources with huge energy consumption
- Human brain: almost 100 billion neurons, up to 1,000 trillion synapses, but...
- Power consumption below 20 W

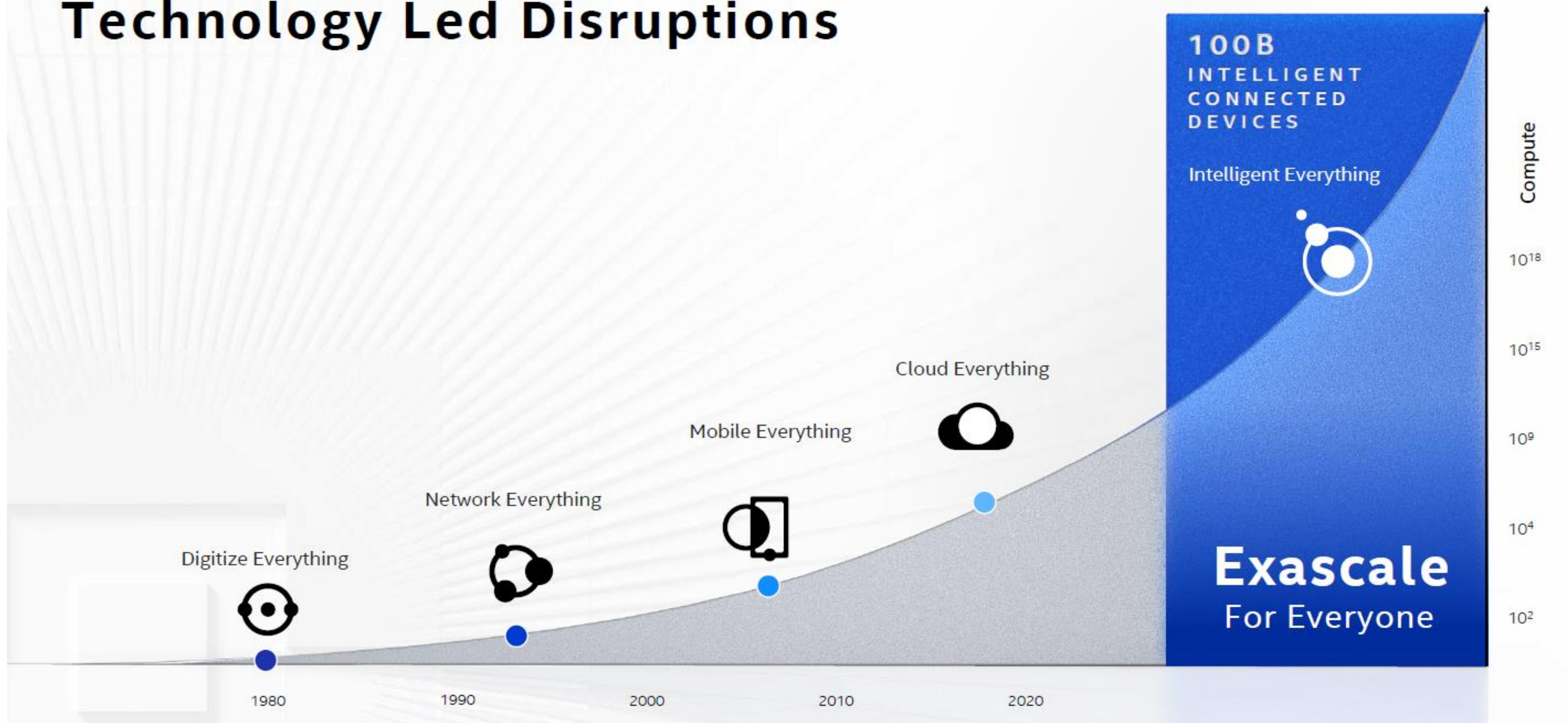
Source: R. Koduri, Intel, Hot Chips 2020

Evolution of Market Drivers: Data Explosion



Source: R. Koduri, Intel, Hot Chips 2020

Evolution of Market Drivers



Source: R. Koduri, Intel, Hot Chips 2020

Distributed Compute Era

2000's-2010's Bringing Data to the Compute

2020's → Bringing Compute to the Data

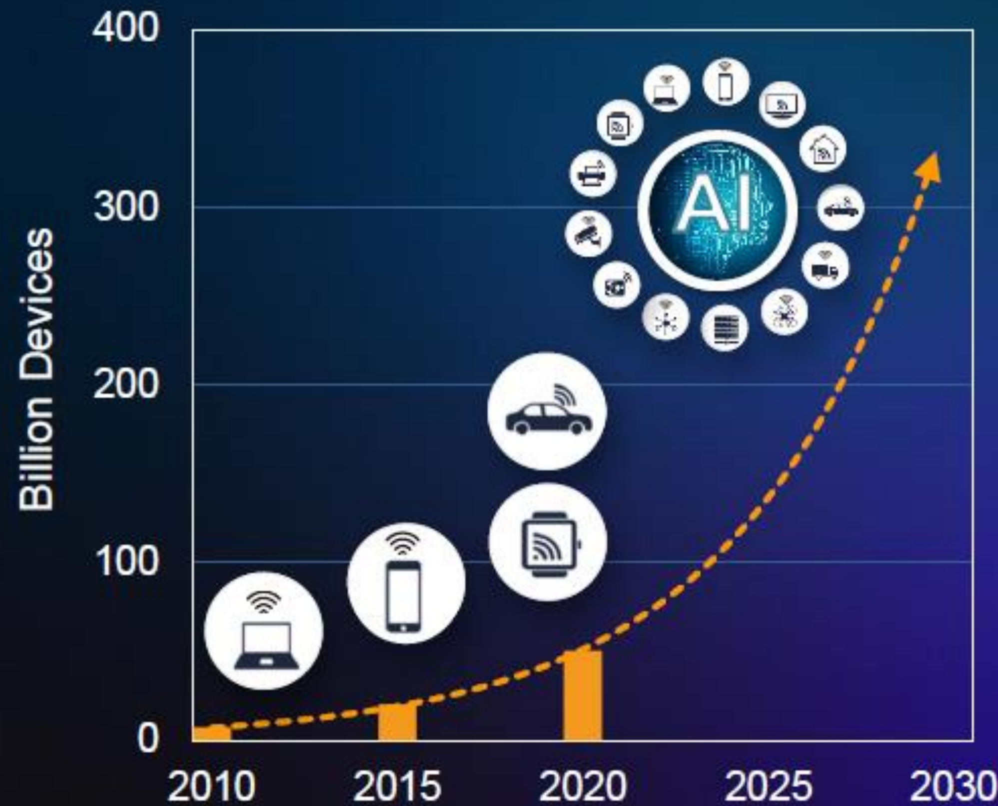
Distributed connected compute brings compute to data and adds QoS and Security to data lifecycle

Source: M.Mayberry,_Intel,_VLSI 2020

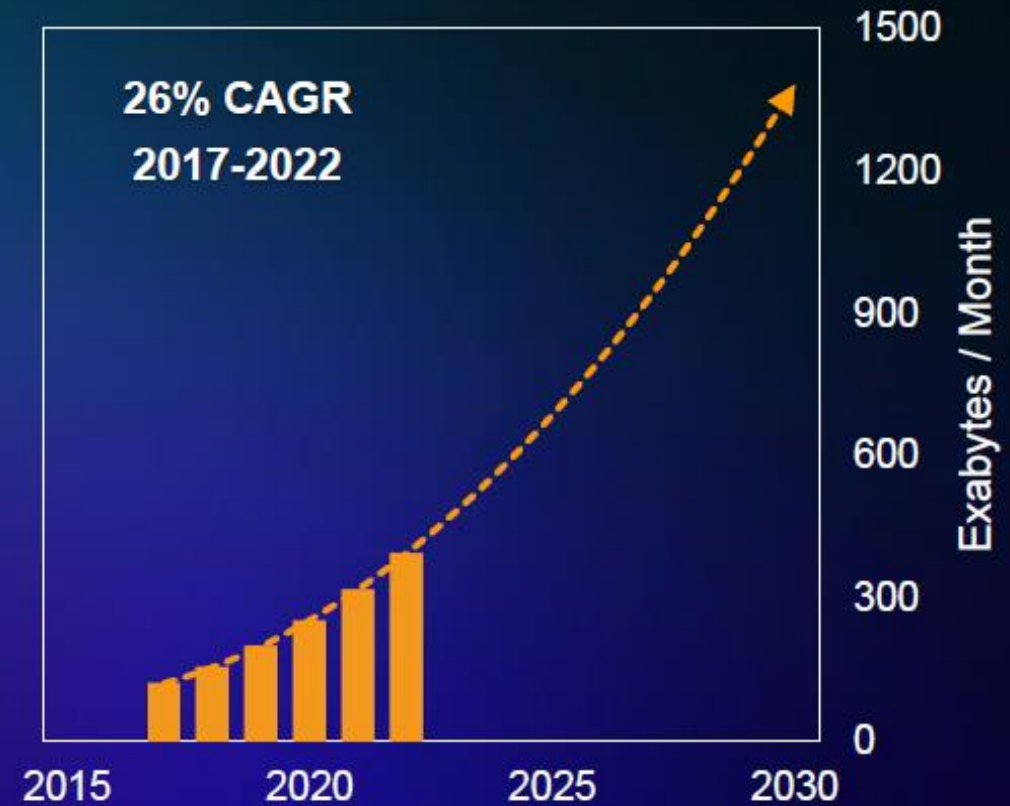
Edge AI Device Trajectory

>350 Billion AIoT (AI+IoT) Devices in Year 2030

Worldwide Devices

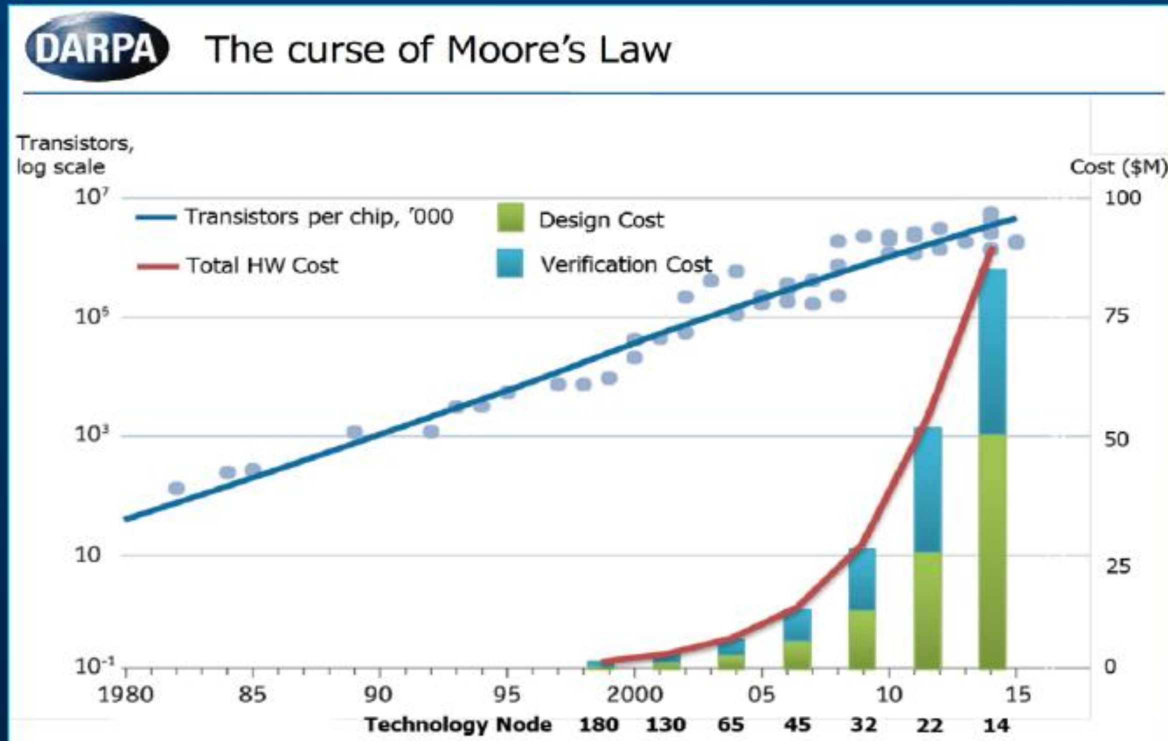


Global Internet Traffic



Source: K-H Loh, MediaTek, ISSCC, 2020

Compute Barriers: Complexity and Cost



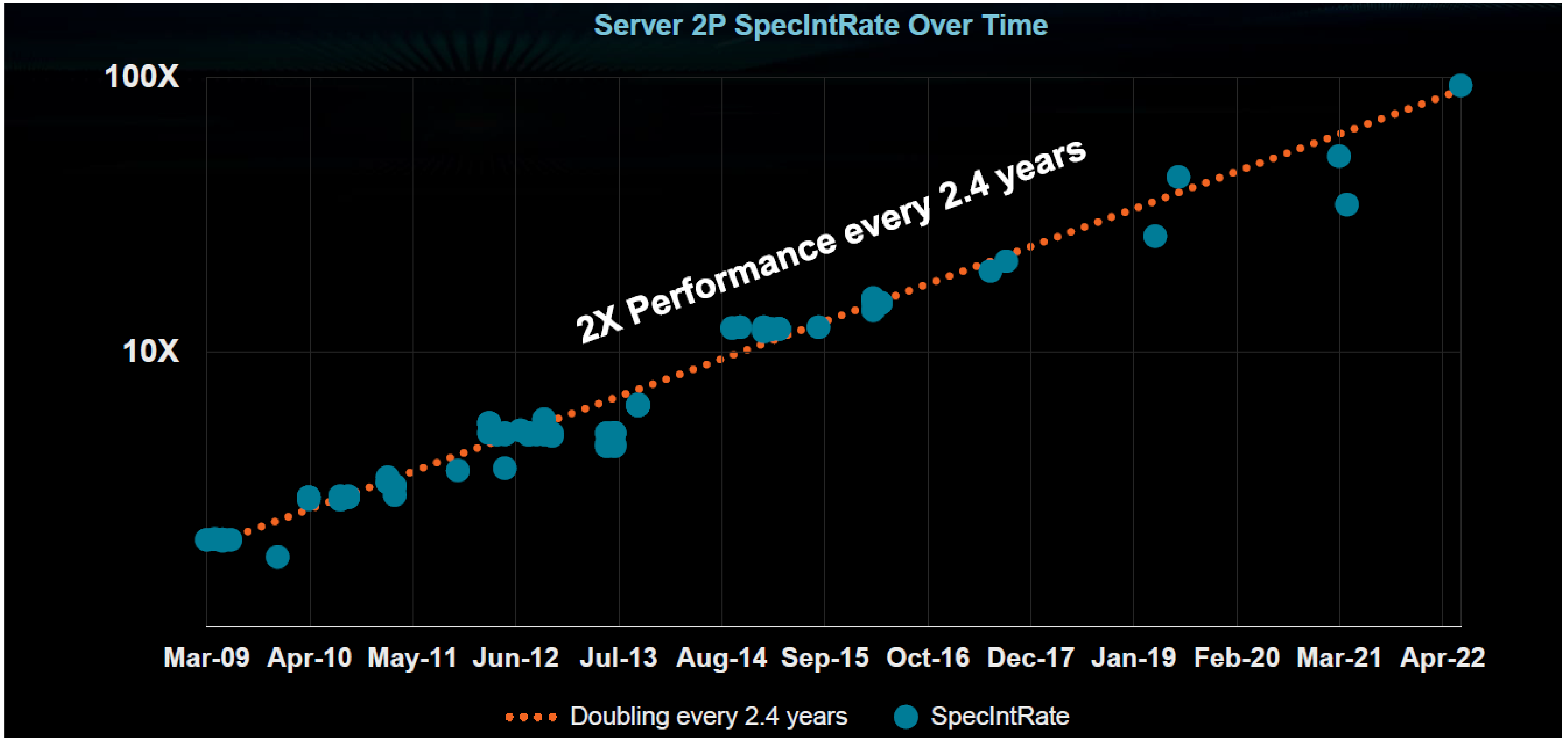
Ref: DARPA IDEA

Increasing cost of HW design and verification

Increasing programming complexity

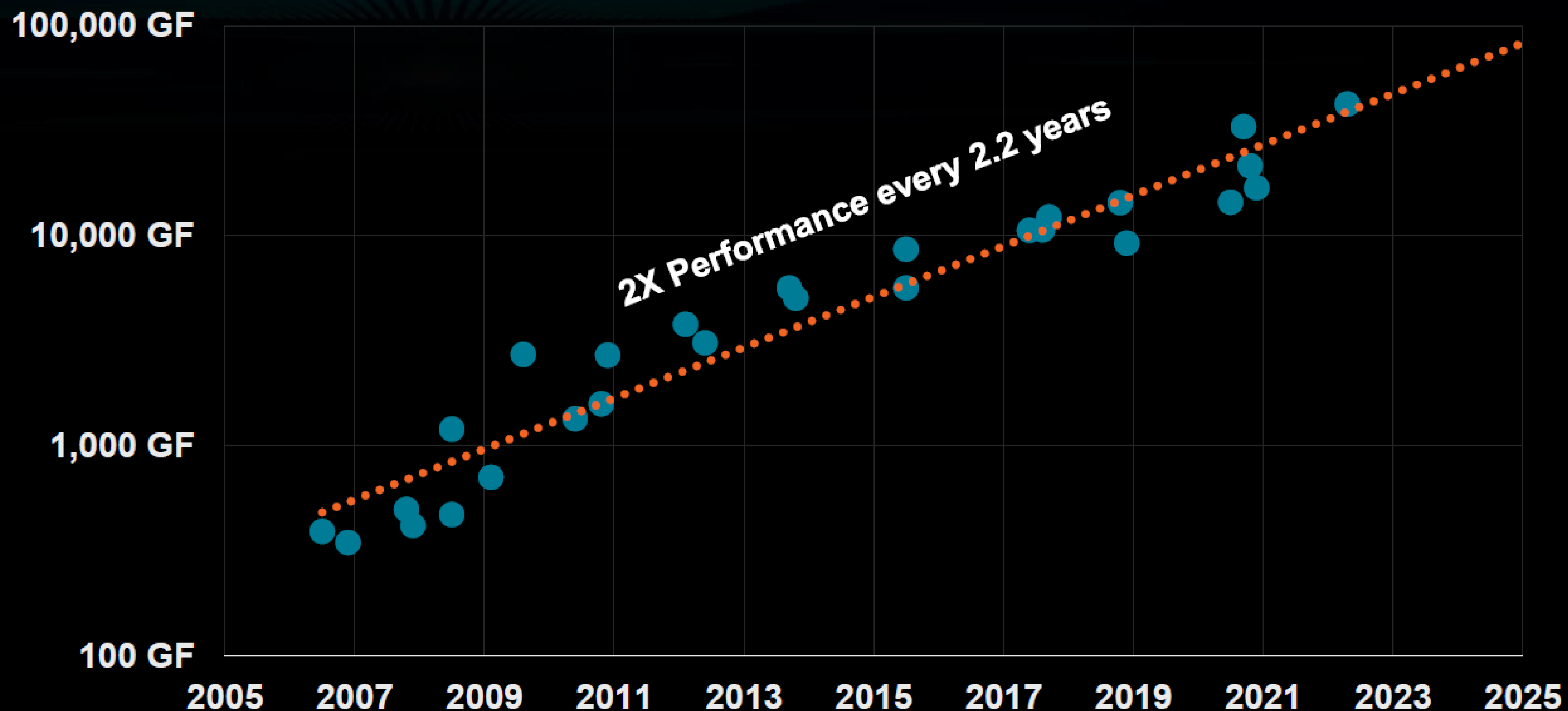
Source: M.Mayberry, Intel, VLSI 2020

Data Server Performance Evolution



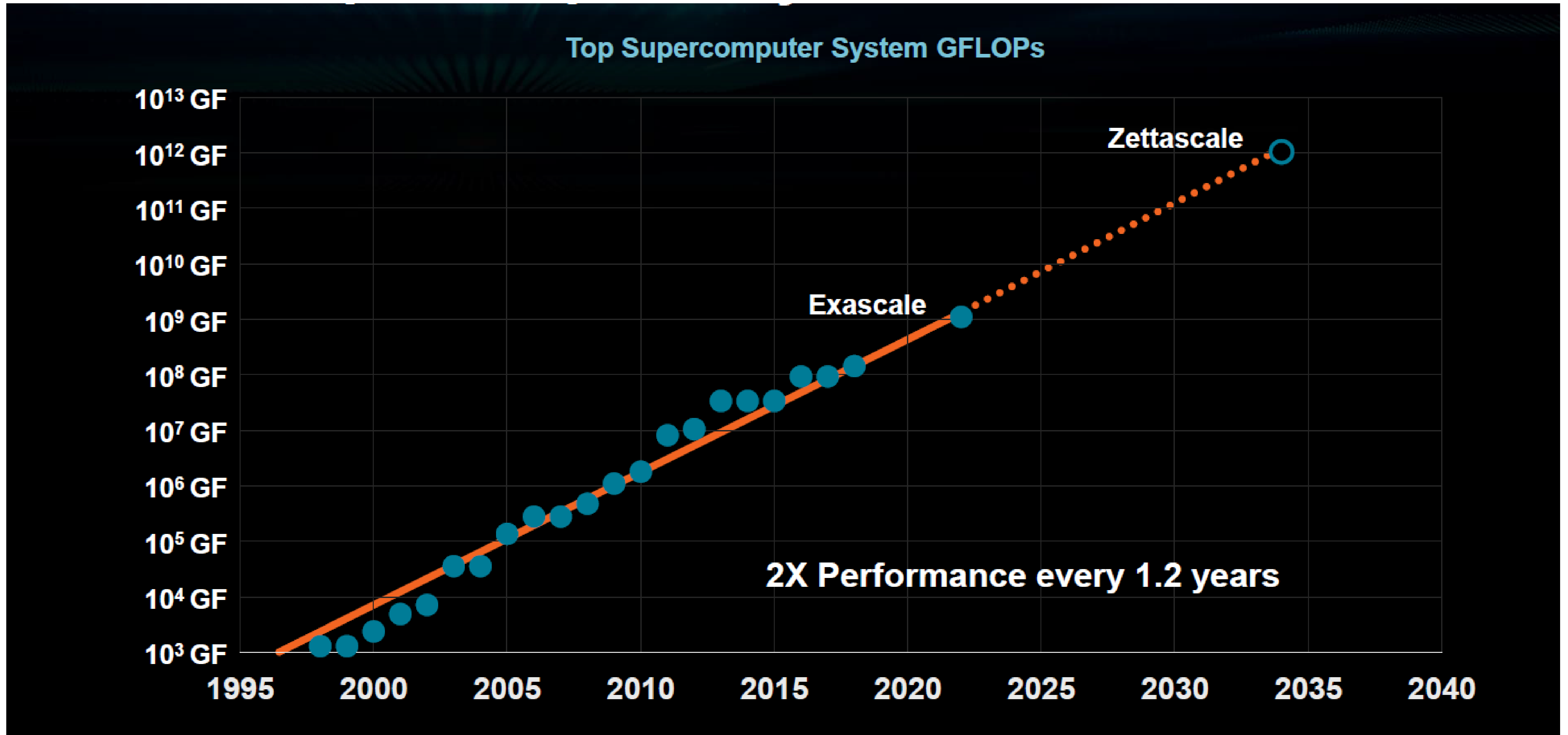
Source: L. Su AMD, ISSCC 2023

Graphical Processor (GPU) Performance Evolution



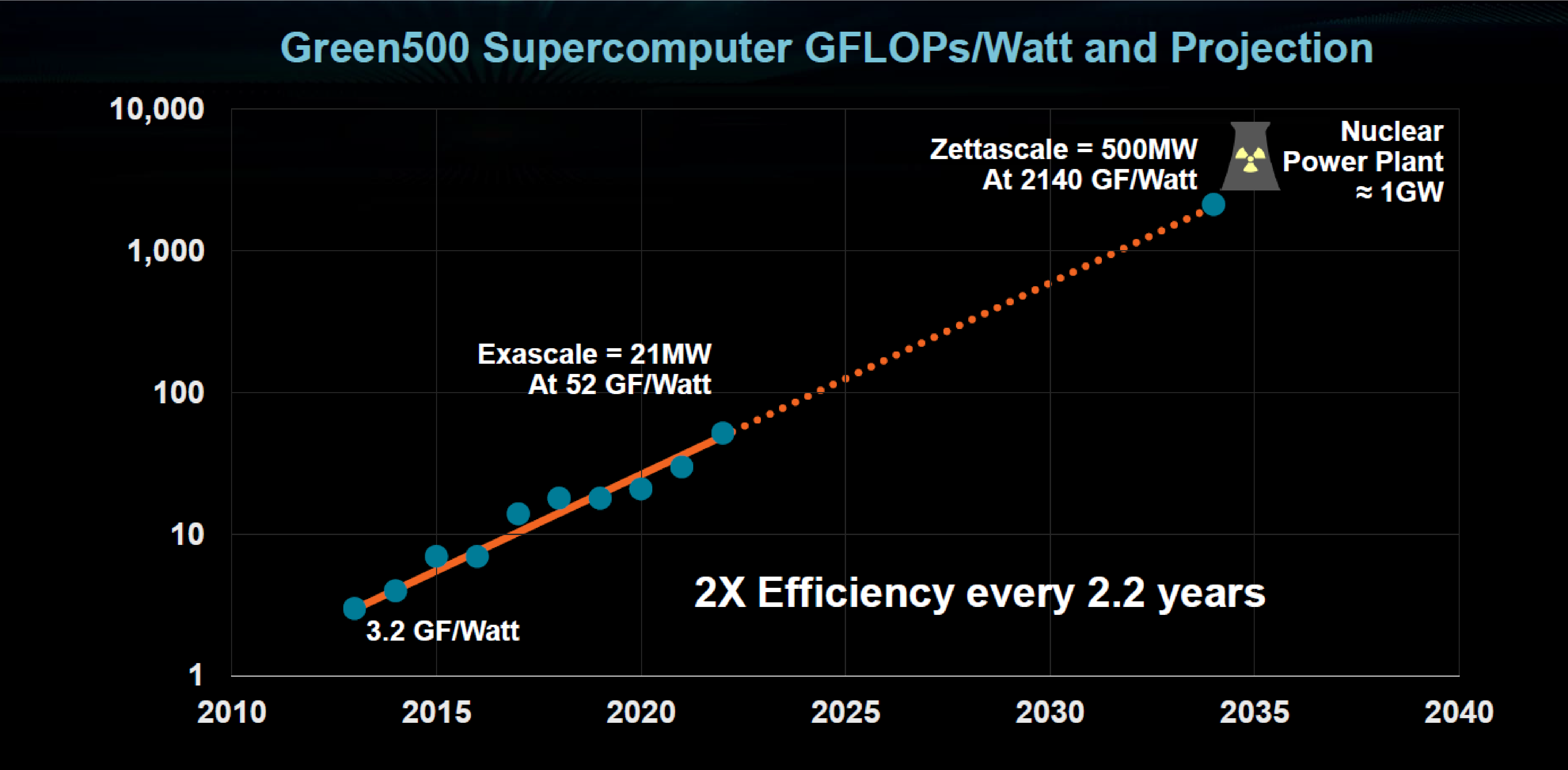
Source: L. Su AMD, ISSCC 2023

Supercomputer Performance Evolution



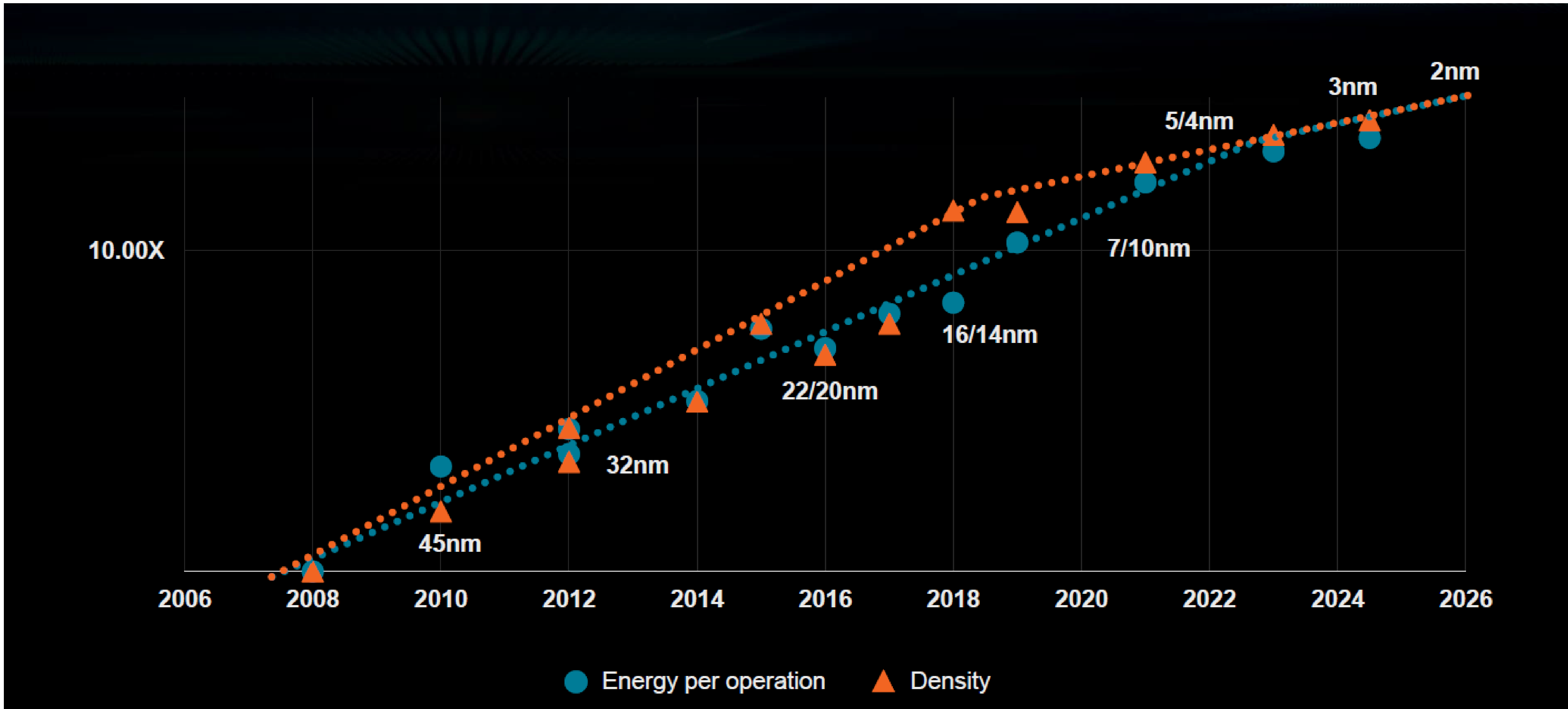
Source: L. Su AMD, ISSCC 2023

Performance/Power Consumption Evolution



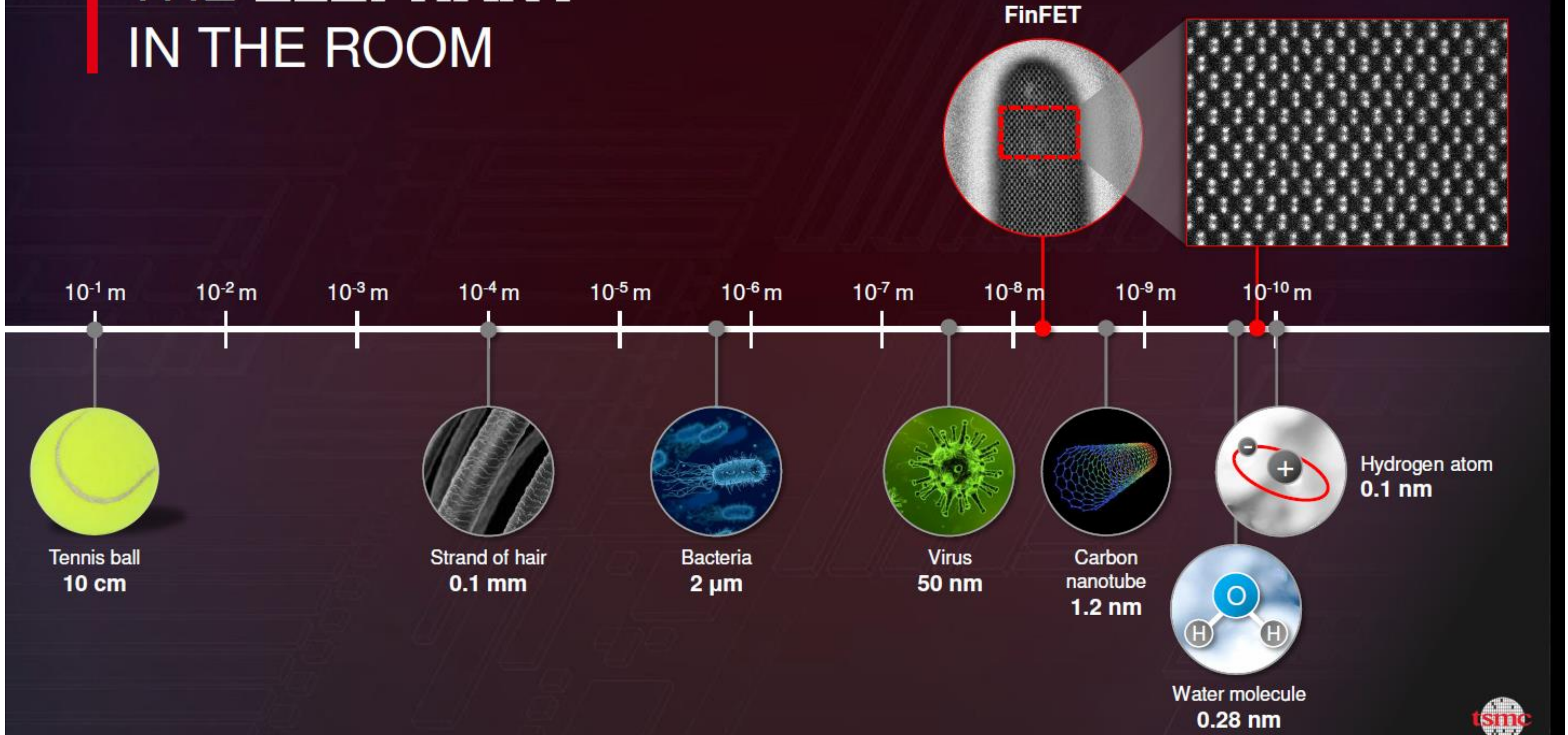
Source: L. Su AMD, ISSCC 2023

ULSI Technology Miniaturization



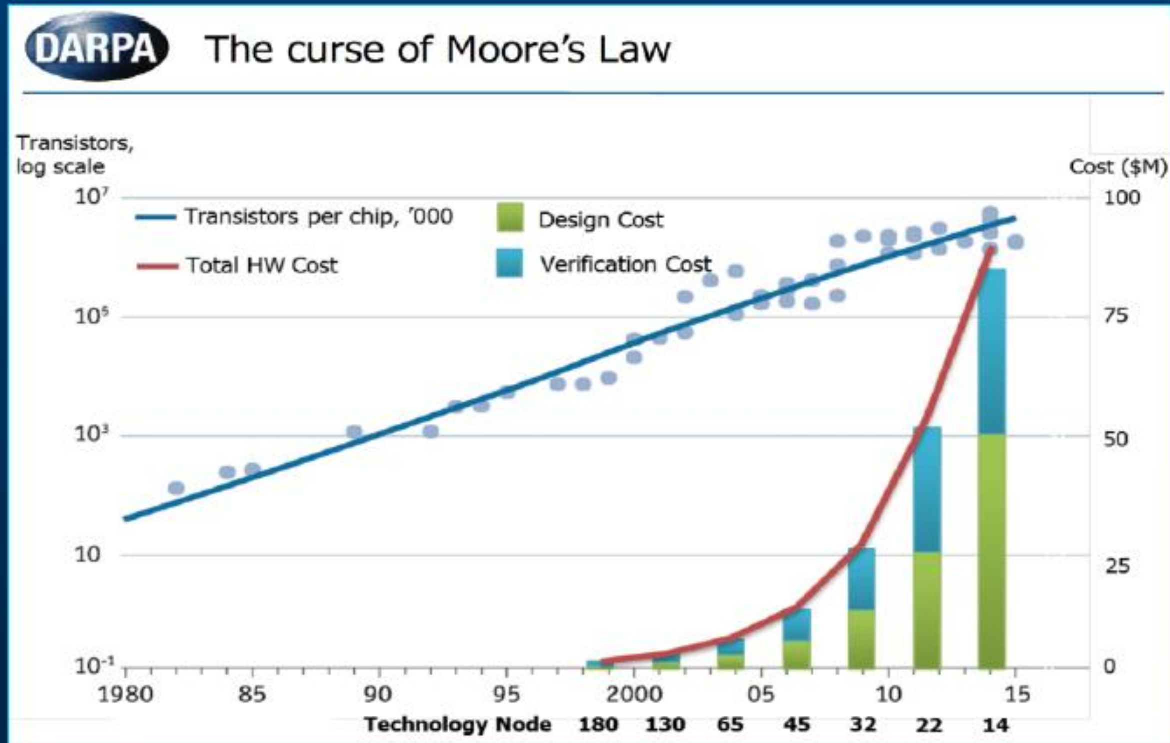
Source: L. Su AMD, ISSCC 2023

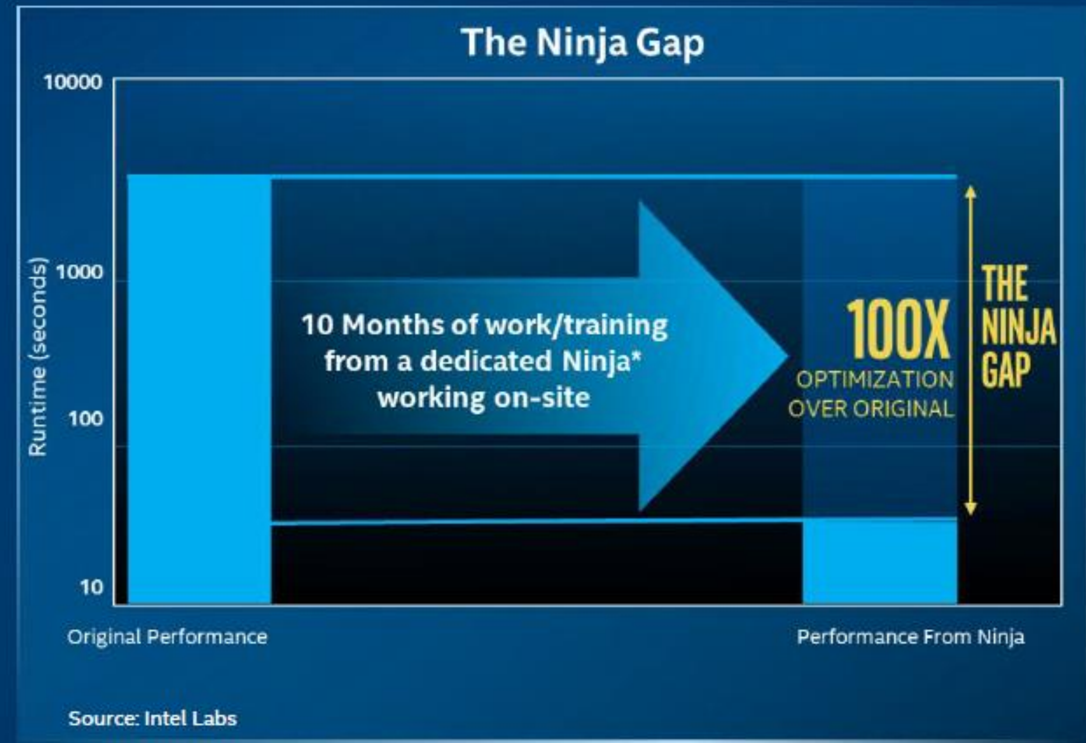
Miniaturization Scale



Source: P. Wong, Stanford/tsmc, Hot Chips 2019

Key Barriers: Complexity and Costs





Ref: DARPA IDEA

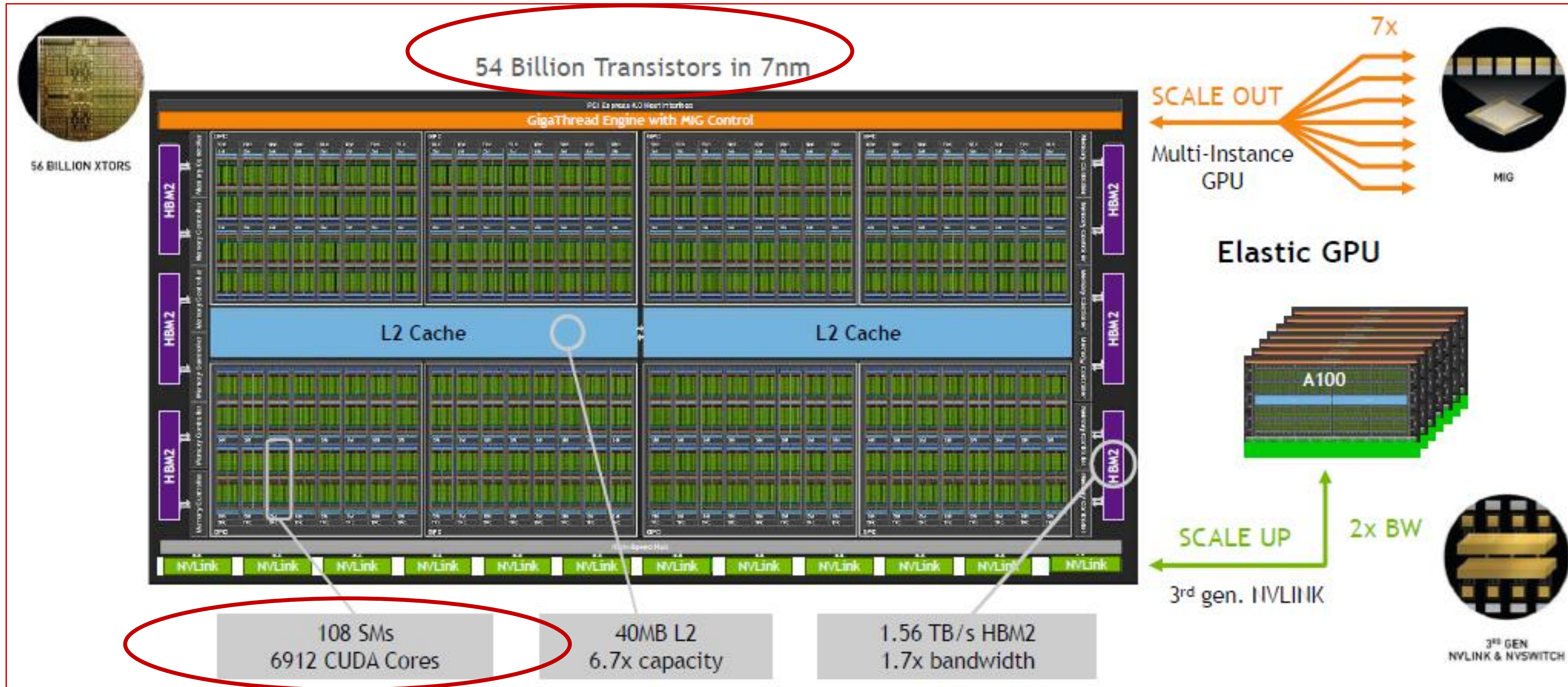
Increasing cost of HW design and verification

Increasing programming complexity

Source: M.Mayberry, Intel, VLSI 2020

Solution: Massively Parallel Architecture

NVIDIA: A100 Chip



Source: J. Choquette, NVIDIA, Hot Chips,2020

Solution: Wafer Scale Integration?

Cerebras Wafer-Scale Engine (WSE-2)

The Largest Chip in the World

850,000 cores optimized for sparse linear algebra

46,225 mm^2 silicon

2.6 trillion transistors

40 gigabytes of on-chip memory

20 PByte/s memory bandwidth

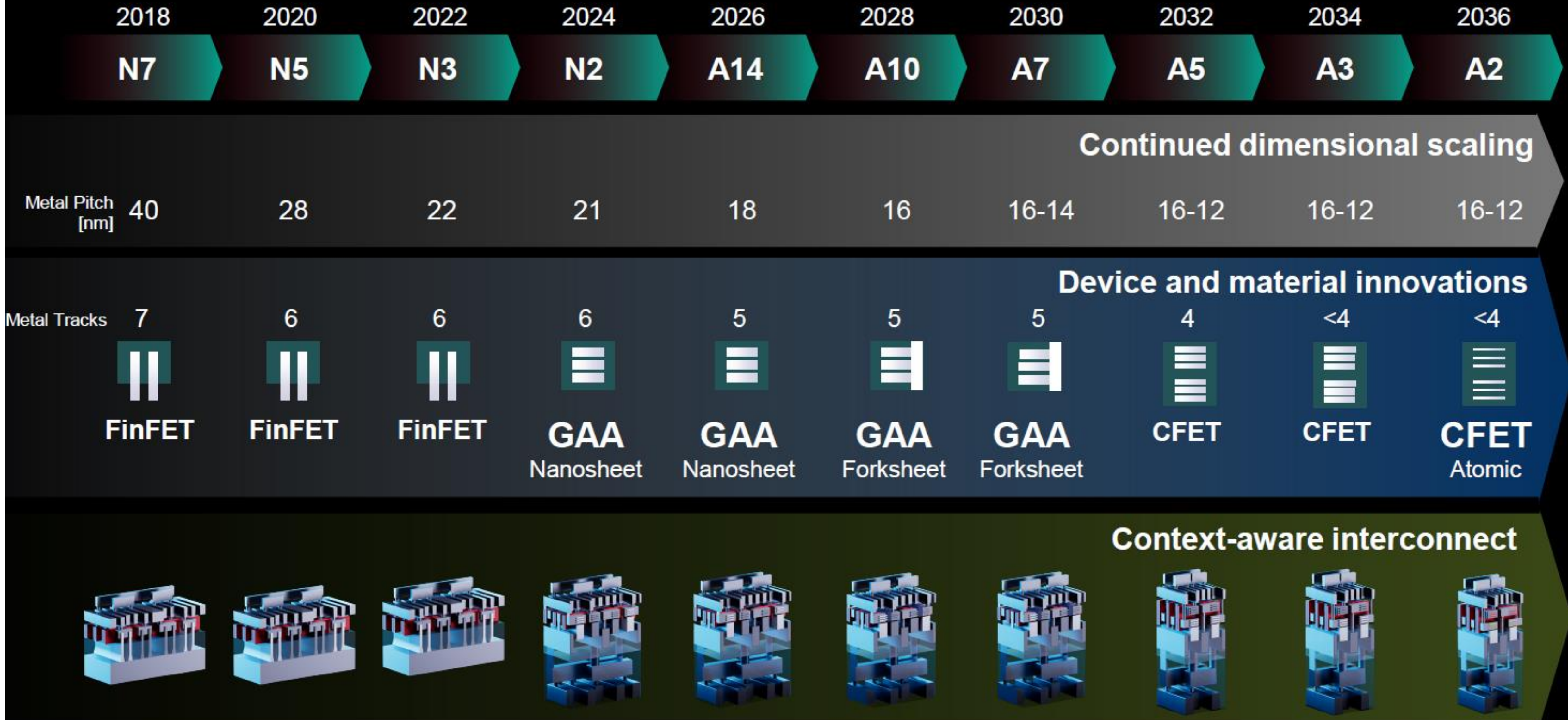
220 Pbit/s fabric bandwidth

7nm process technology

56x larger than largest GPU

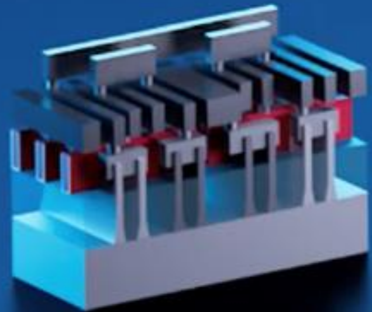
Outline

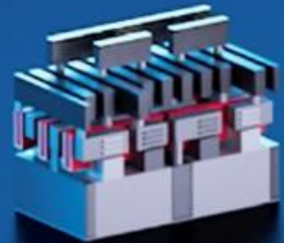
- Current technology landscape and future growth directions
- 3D device structures from FinFET to nanosheet architectures and beyond Silicon
- 3D heterogeneous integration: chiplets and neuromorphic computing
- Electrical car driven SiC market explosion



Source: imec, 2023

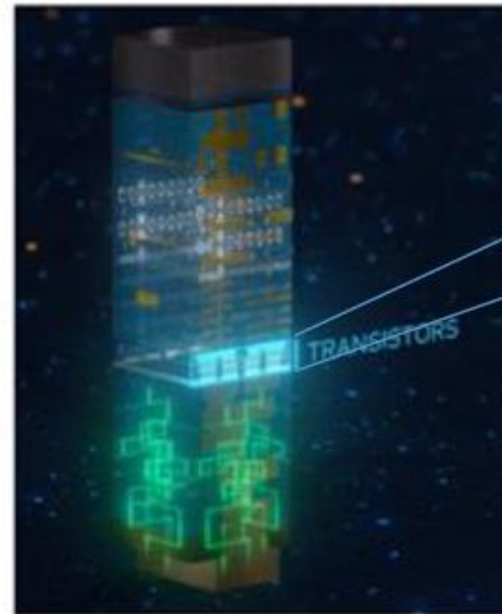
Disruptive architecture innovations





Source: imec, 2023

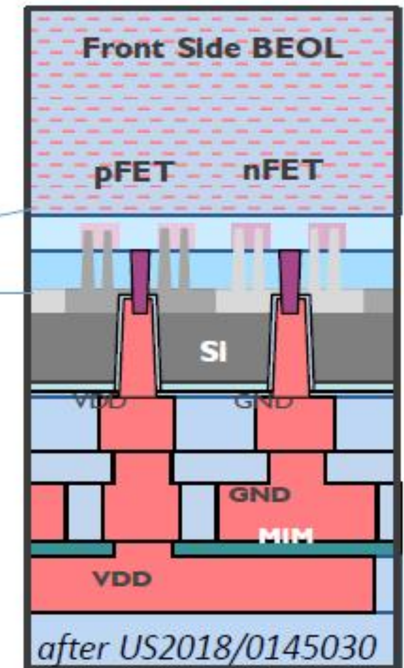
Next step in differentiation: backside power delivery



signal connections above

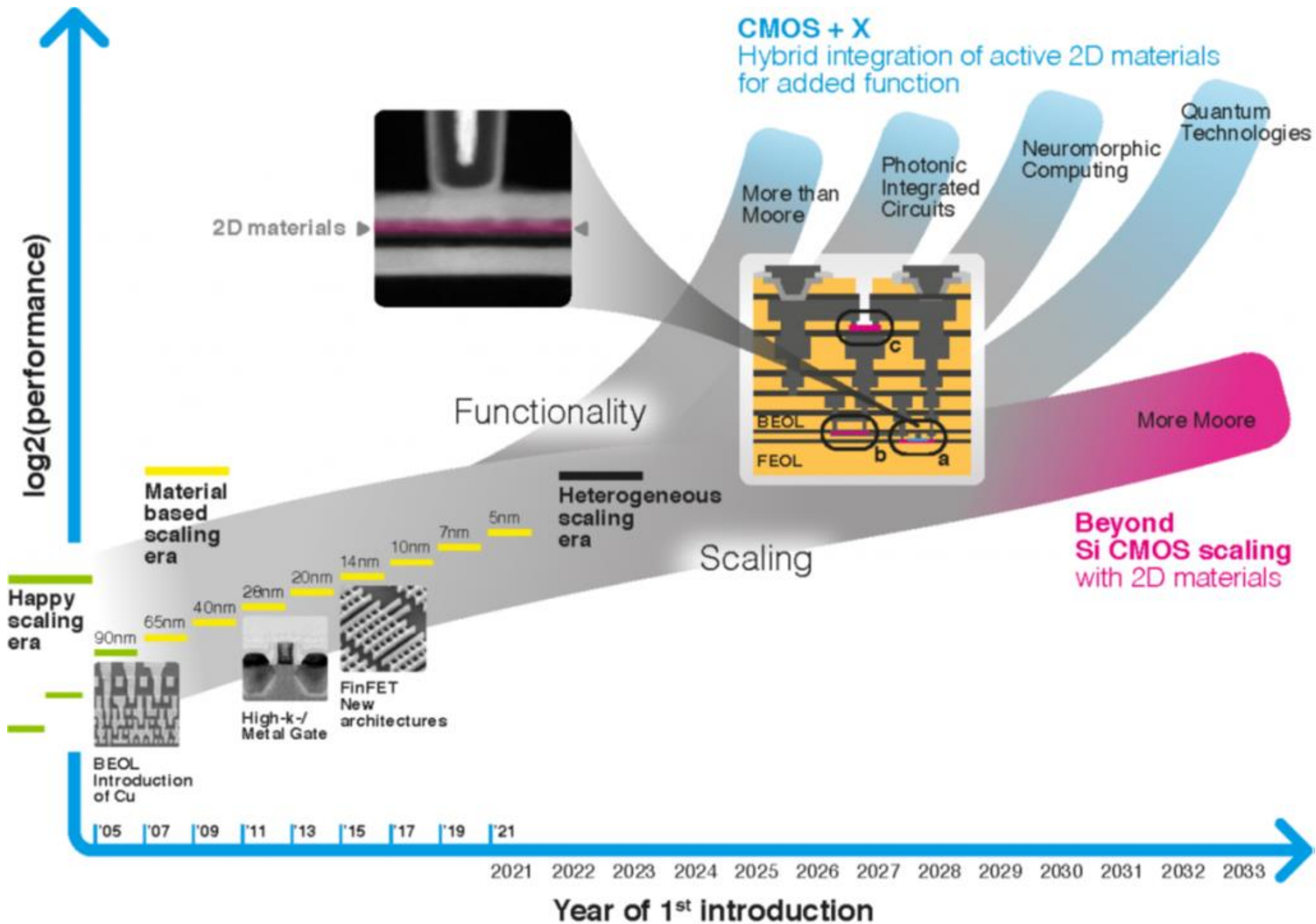
FETs

power delivery below



- backside power delivery allows next-level differentiation of signal wiring and power delivery
  ➔ big Integration challenge

Source: imec, 2023

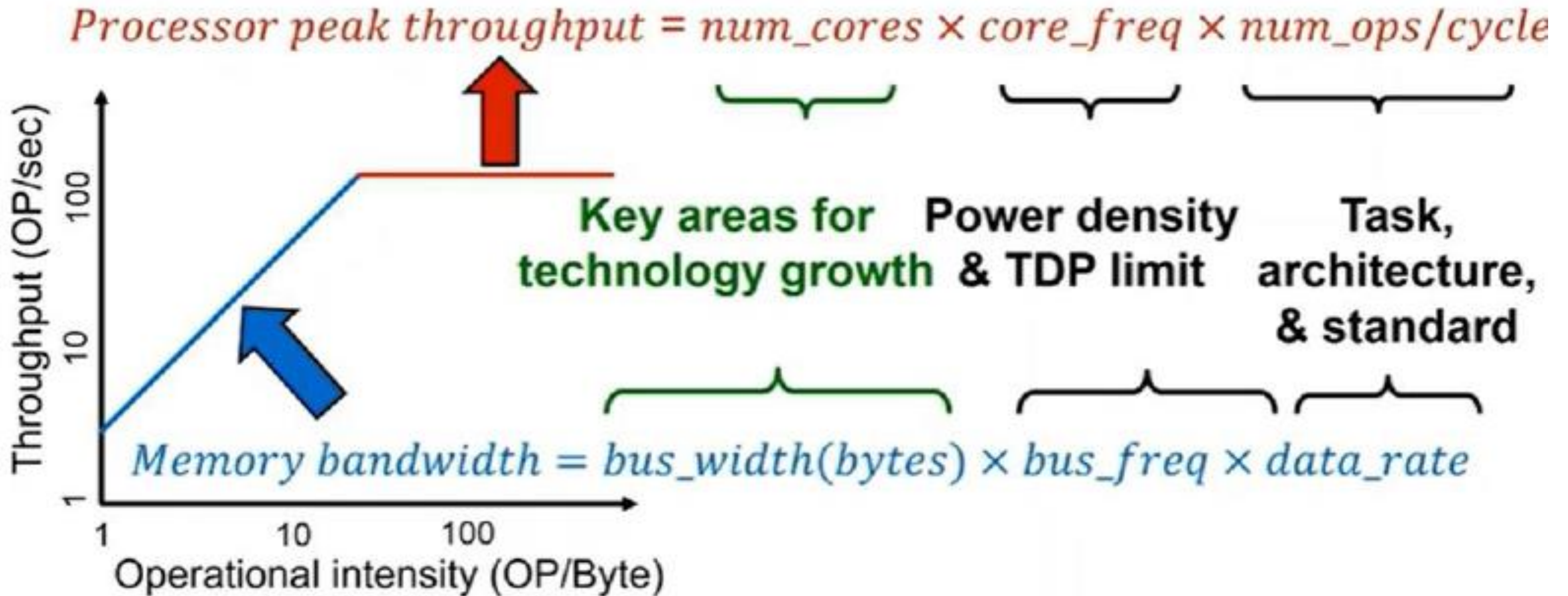


CMOS + X
Hybrid integration of active 2D materials
for added function
2D materials
More than
Moore
Photonic
Integrated
Circuits
Neuromorphic
Computing
Quantum
Technologies
log2(performance)
Functionality
BEOL
FEOL
c
b
a
More Moore
Material
based
scaling
era
Heterogeneous
scaling
era
Scaling
Beyond
Si CMOS scaling
with 2D materials
Happy
scaling
era
90nm
65nm
40nm
28nm
20nm
14nm
10nm
7nm
5nm
BEOL
Introduction
of Cu
High-k-/
Metal Gate
FinFET
New
architectures
'05
'07
'09
'11
'13
'15
'17
'19
'21
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
Year of 1st introduction

Outline

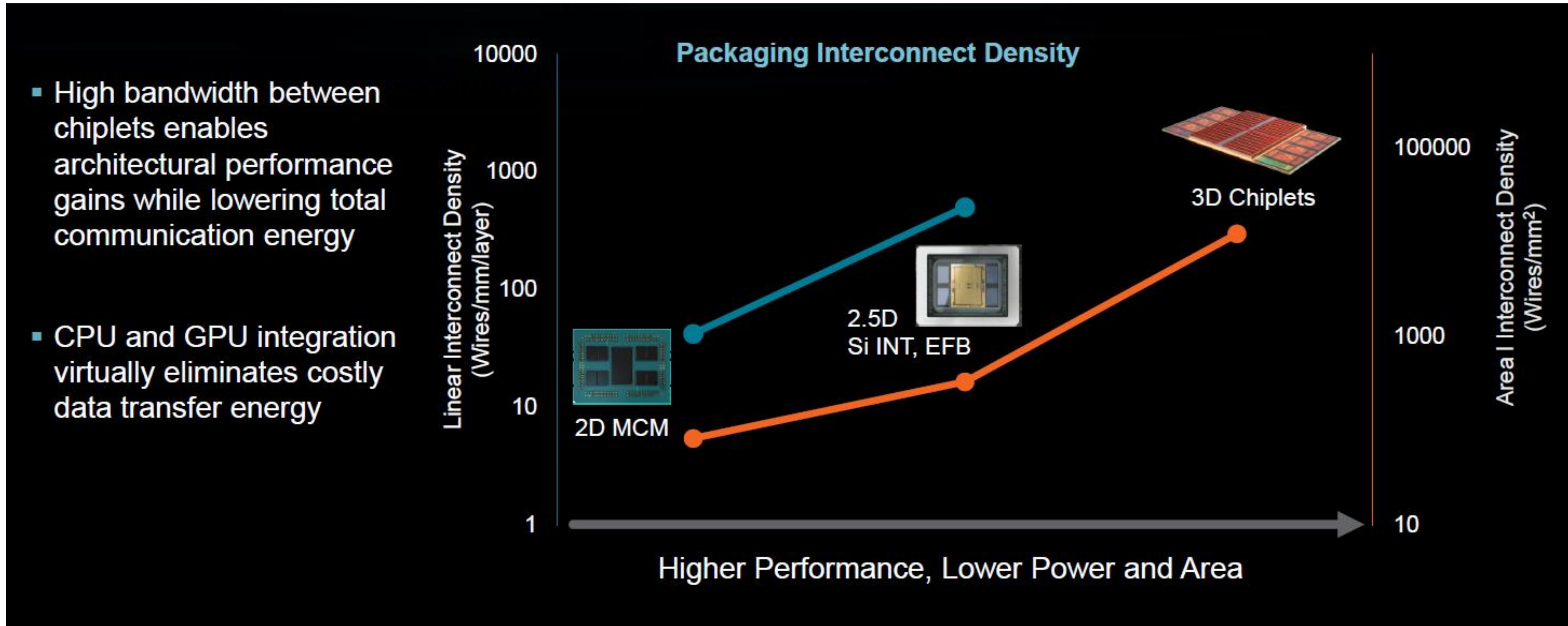
- Current technology landscape and future growth directions
- 3D device structures from FinFET to nanosheet architectures and beyond Silicon
- 3D heterogeneous integration: chiplets and neuromorphic computing
- Electrical car driven SiC market explosion

System Performance Roofline Plot



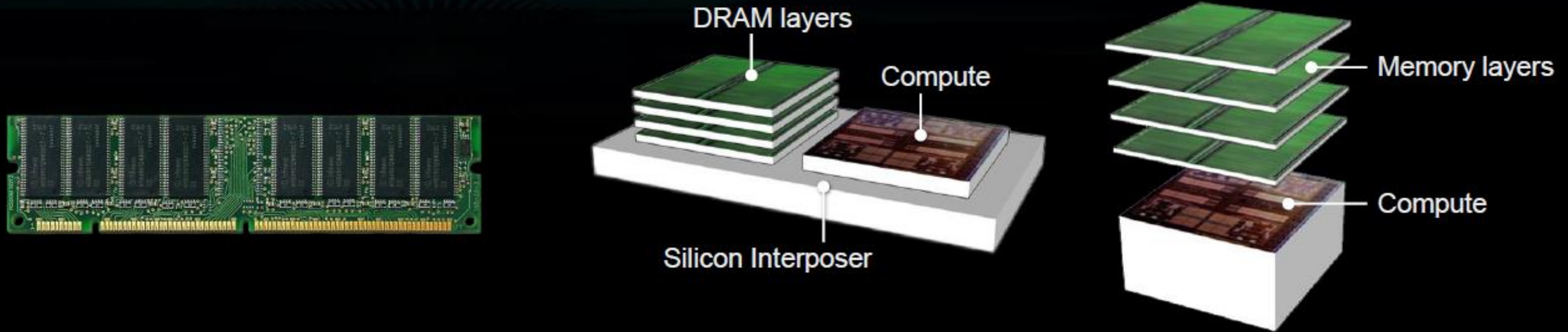
Source: P. Wong, Stanford/tsmc, DARPA ERI Summit,2020

How Do We further Increase System Efficiency?



- **2.5 D & 3D Integration**
  - **More than 90% of Power Consumption for Processor-Memory Communication**
- **Near Memory and In-Memory Computing next on roadmap**

Energy Benefits of 2.5D & 3D Integration



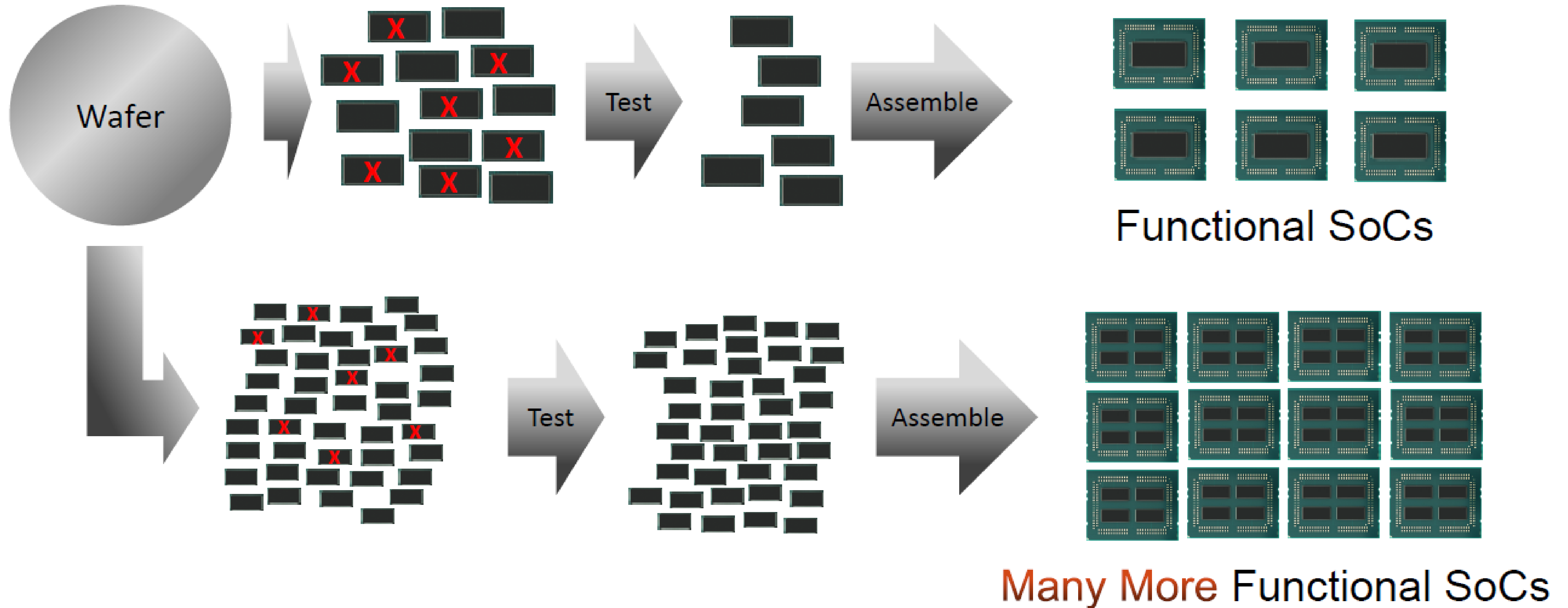
Integration Enables Higher Bandwidth at Lower Power

	DIMMS	2.5D Micro-bumps (HBM)	3D Hybrid Bond
pj/bit	~12	~3.5	~0.2

Amount of energy to send data from the processor to DRAM

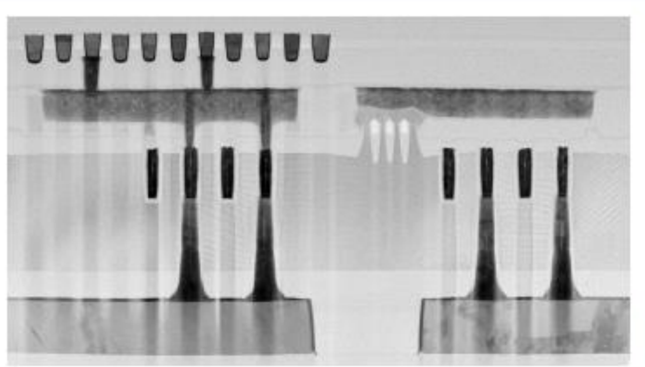
- **2D - 12pJ/bit** **2.5D - 3.5 pJ/bit** **3D - .2pJ/bit**

Additional Benefits to Chiplets: Higher Yield

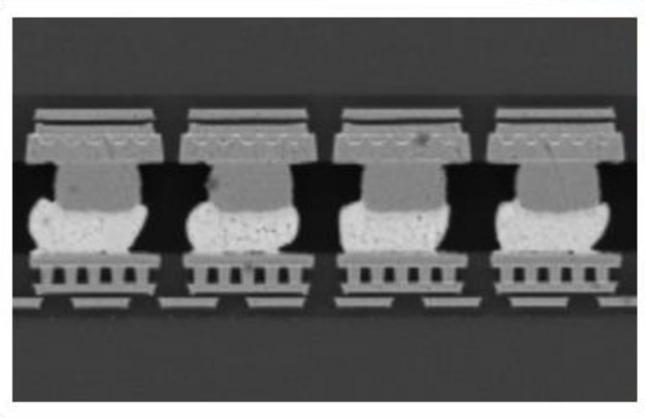


Source: S. Naffziger , AMD, VLSI Technology Symposium, 2020

3D Integration Examples

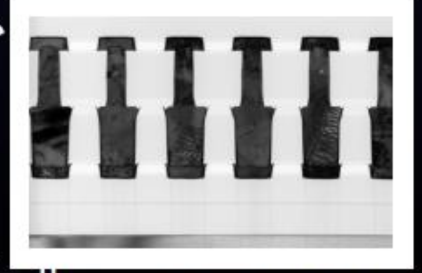


BACK-SIDE POWER DISTRIBUTION

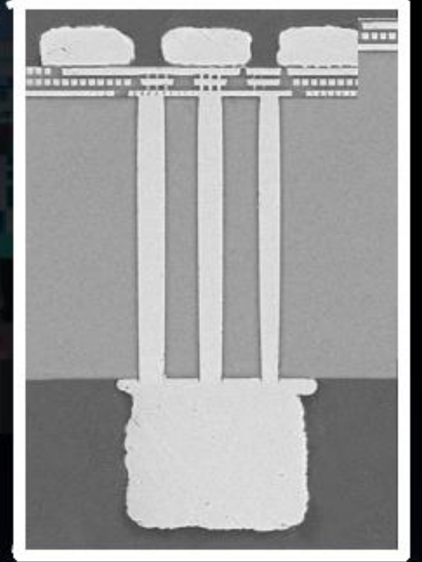


DIE TO WAFER MICRO-BUMPS

3D stacked core processor SOC



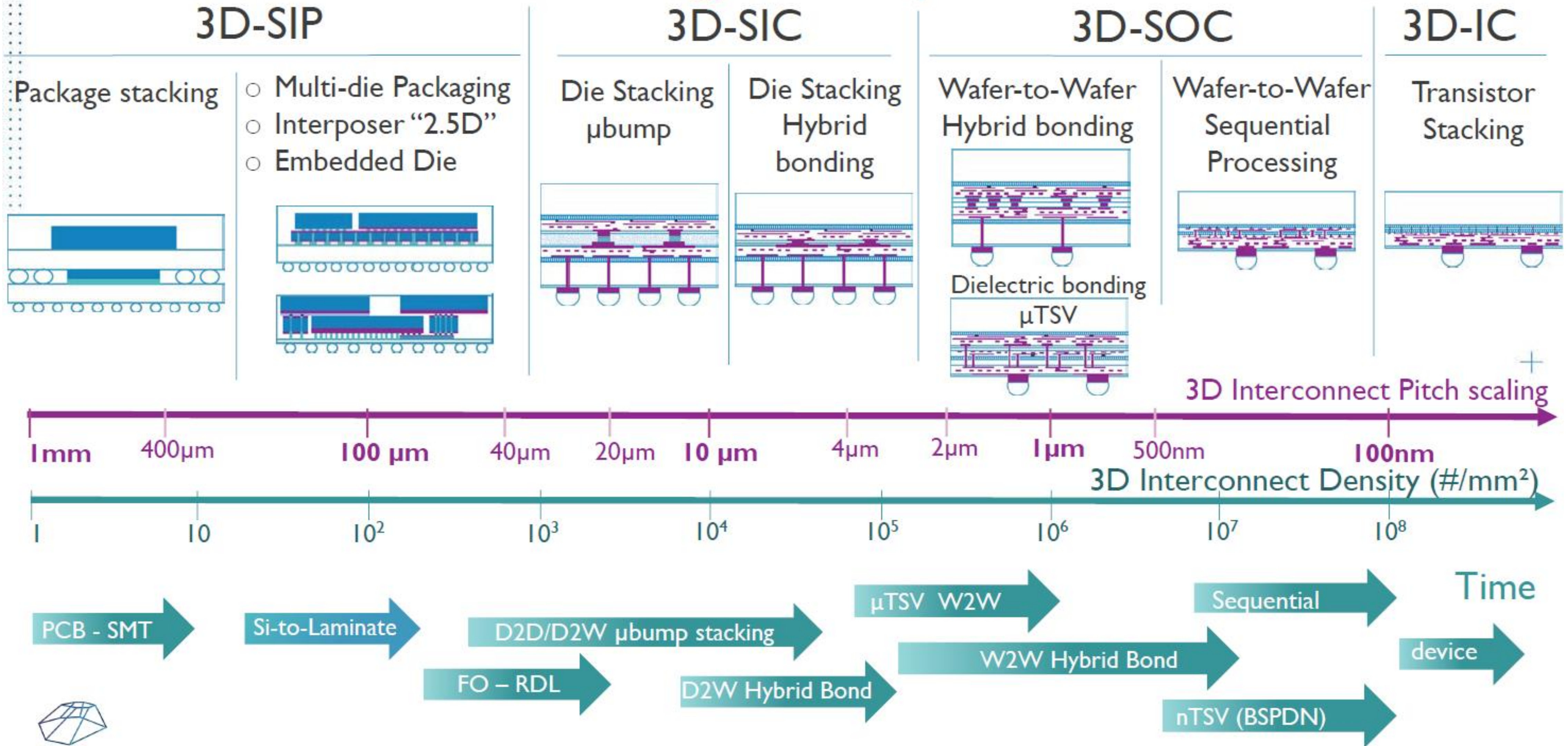
Cu-Cu HYBRID BONDING



THROUGH-Si VIA

Source: imec, 2022

3D Integration Roadmap



Source: imec, 2022

3D Landscape and 3D Interconnect Roadmap

1E+08
1E+07
1E+06
1E+05
1E+04
1E+03
1E+02

3D Interconnect Density ($\#/mm^2$)

Wafer-to-Wafer

Die-to-Silicon

W2W HB

D2W HB

D2W µbump

0.1
1
10
100

3D interconnect pitch (µm)

2005 2010 2015 2020 2025 2030

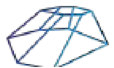
Year

Wafer-to-Wafer Hybrid bonding

Die-to-Wafer Hybrid bonding

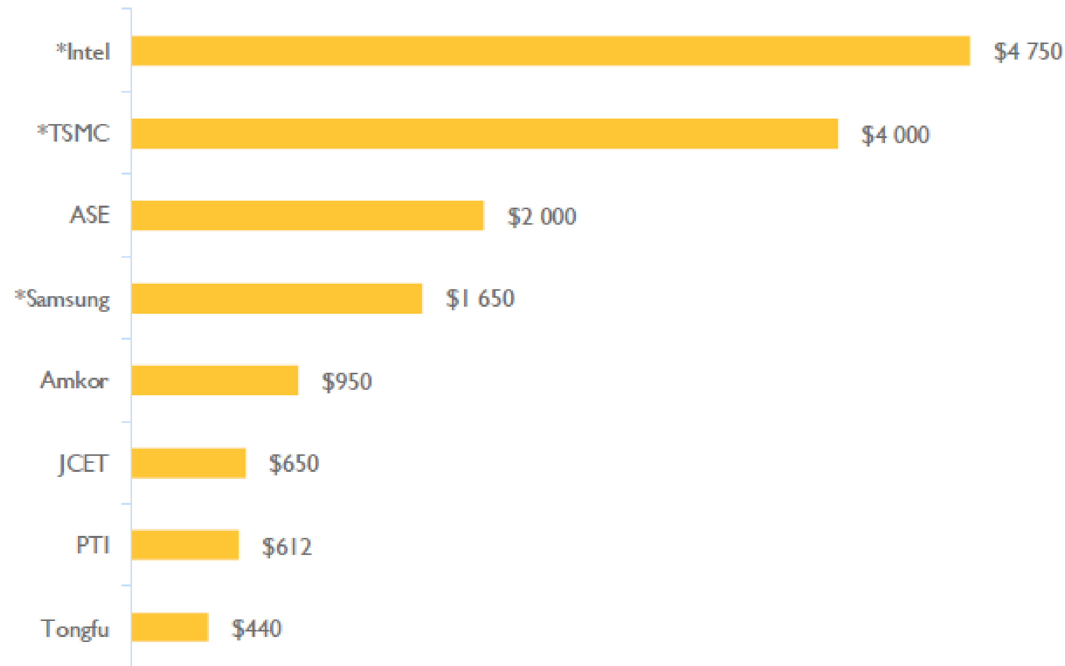
Die-to-Wafer Flip-chip µBump

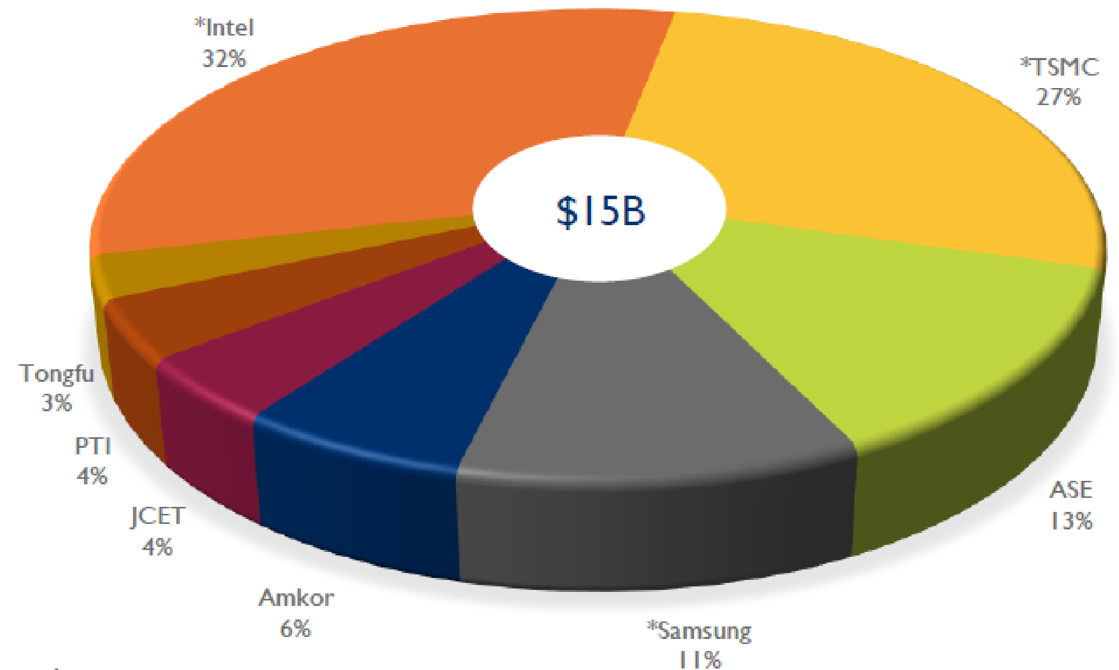
D2W= Die-to-wafer bonding
W2W= Wafer-to-Wafer bonding
HB= Hybrid bonding



Source: imec, 2022

2022 CAPEX HIGHLIGHTS FOR ADVANCED PACKAGING PLAYERS

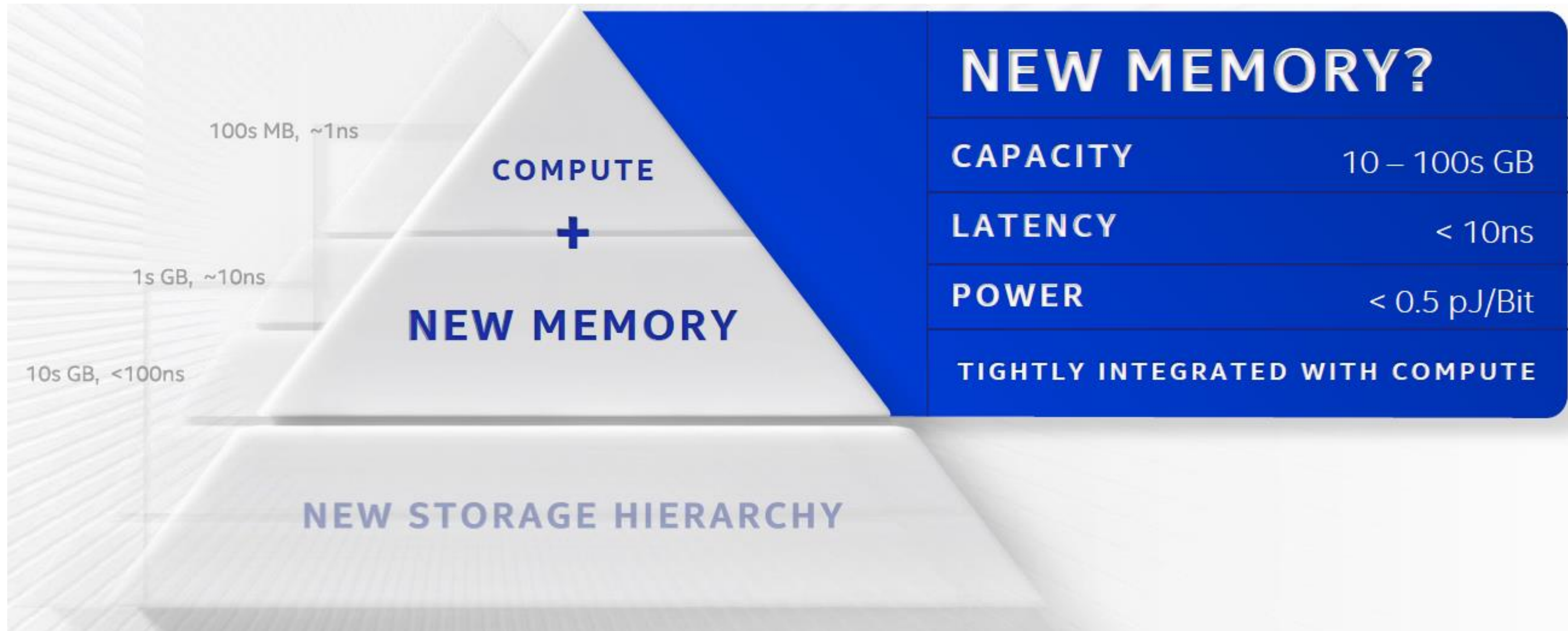




*Intel, TSMC and Samsung advanced packaging CapEx were estimated based on earnings calls statements and recent announcements of investments, since packaging is not their main business focus

All CapEx data is estimated based on information gathered during Q1 2022

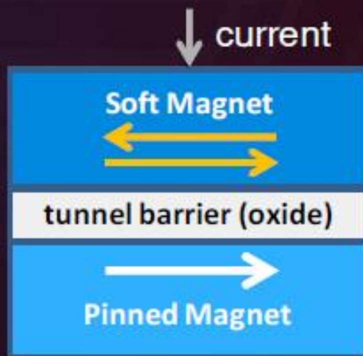
Memory/Storage Hierarchy: New Memory Needs



Source: R. Koduri, Intel, Hot Chips 2020

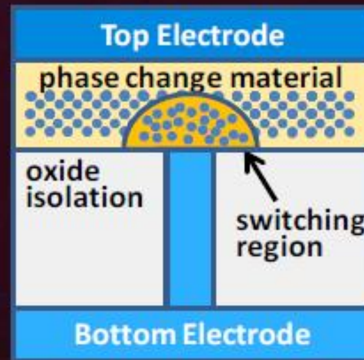
Emerging Non-Volatile Memories

Random access, non-volatile, no erase before write, on-chip integration



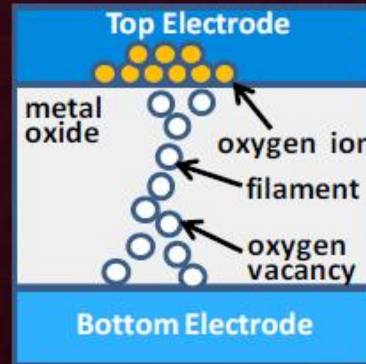
STT-MRAM

Spin torque transfer magnetic random access memory



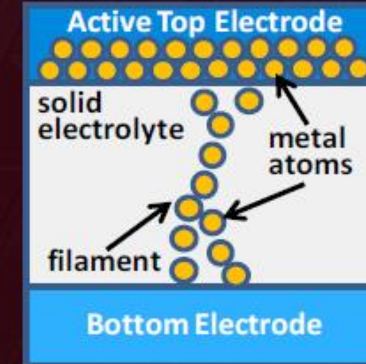
PCM

Phase change memory



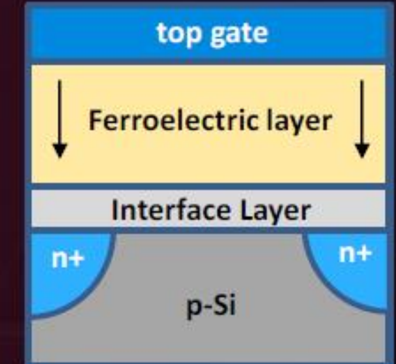
RRAM

Resistive switching random access memory



CBRAM

Conductive bridge random access memory



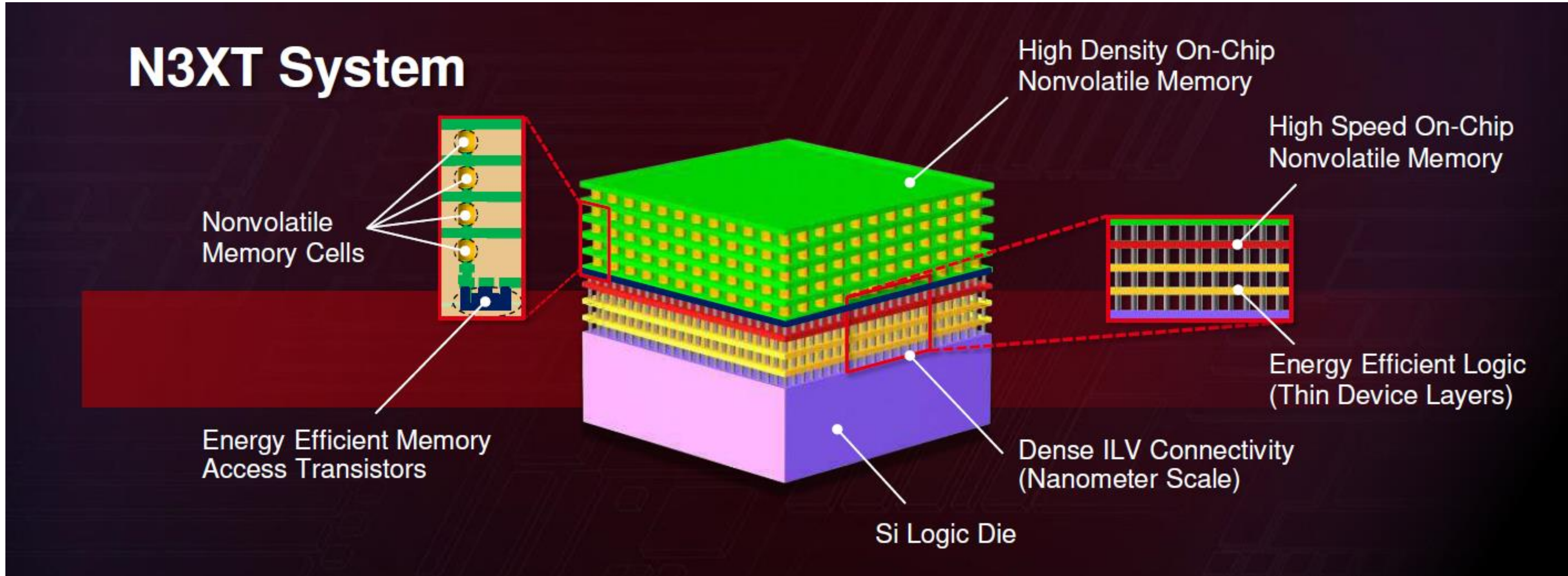
FERAM

Ferro-electric random access memory

Source: H.-S. P. Wong, S. Salahuddin, Nature Nanotech (2015)

Source: P. Wong, Stanford/tsmc, Hot Chips 2019

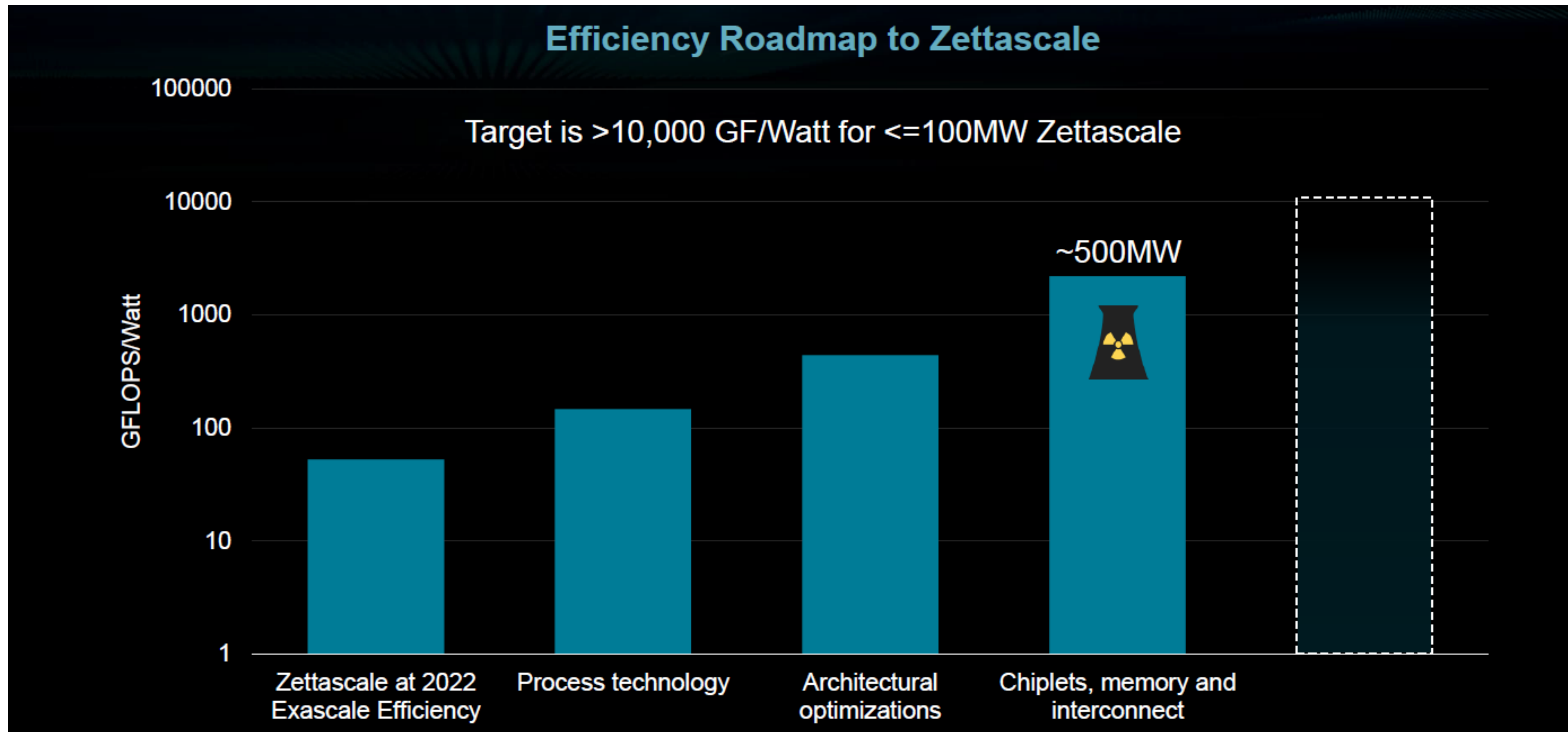
Net Steps on Roadmap: Neuromorphic Computing



Goal: Approach human brain neuron density with acceptable power consumption

Source: P. Wong, Stanford/tsmc, Hot Chips 2019

Conclusions: System Efficiency Roadmap



Dominant Directions:

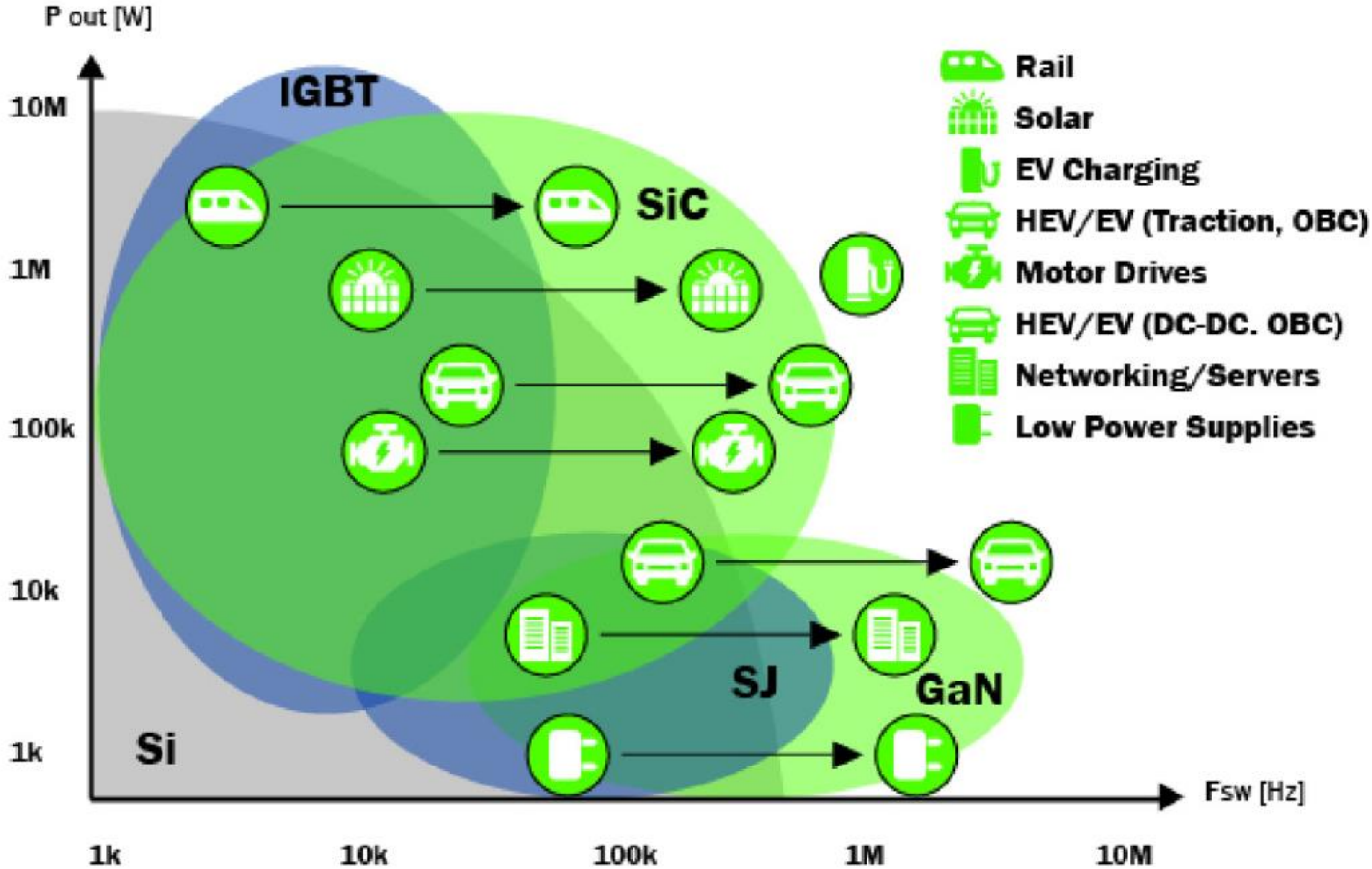
- **Further technological advances: device architectures, backside power delivery, emerging memories, 2D material, Carbon Nanotubes**
- **System Architecture Optimization (software-hardware co-design, neuromorphic computing)**
- **3D Heterogeneous Integration including Photonics**

Source: L. Su AMD, ISSCC 2023

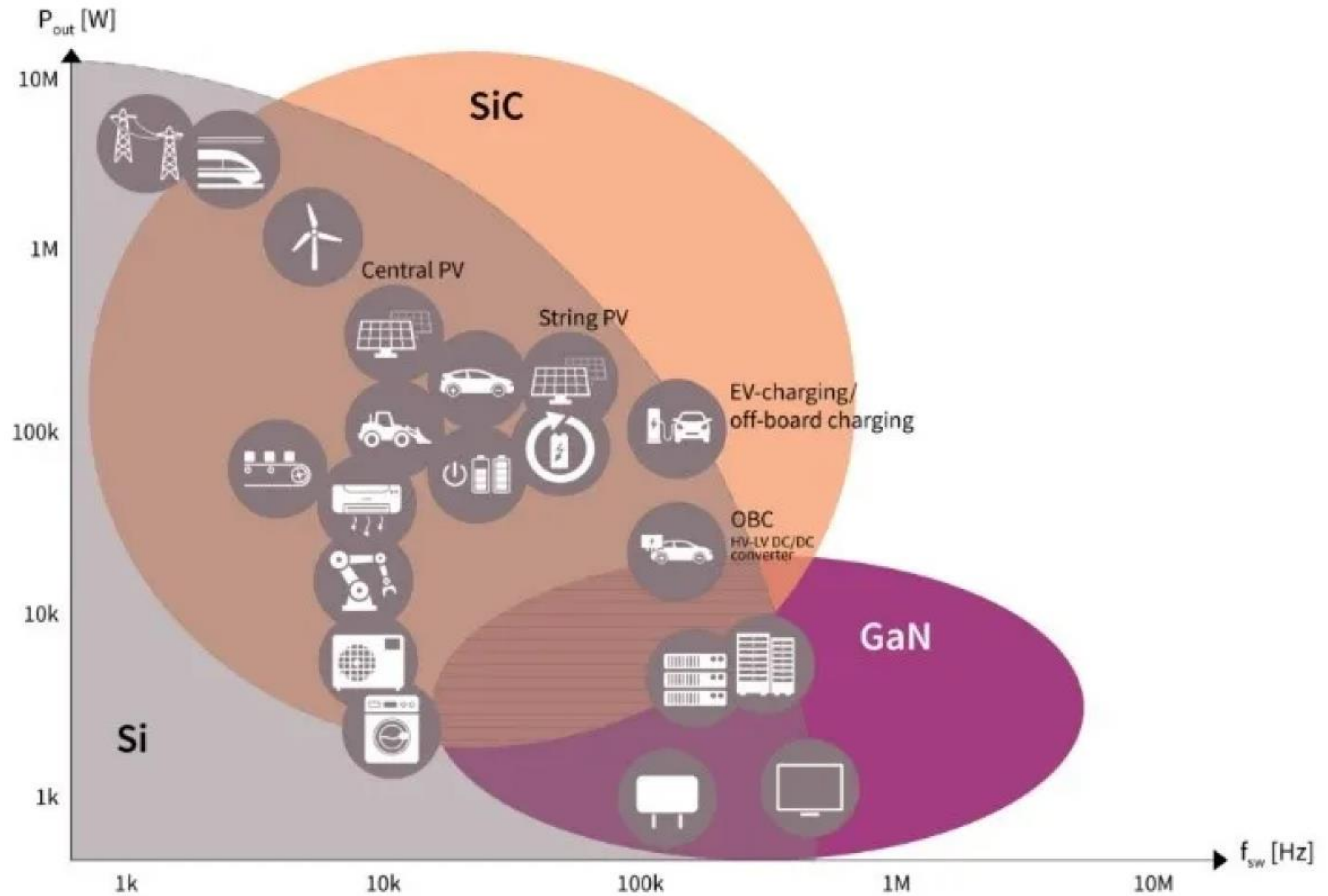
Outline

- Current technology landscape and future growth directions
- 3D device structures from FinFET to nanosheet architectures and beyond Silicon
- 3D heterogeneous integration: chiplets and neuromorphic computing
- Electrical car driven SiC market explosion

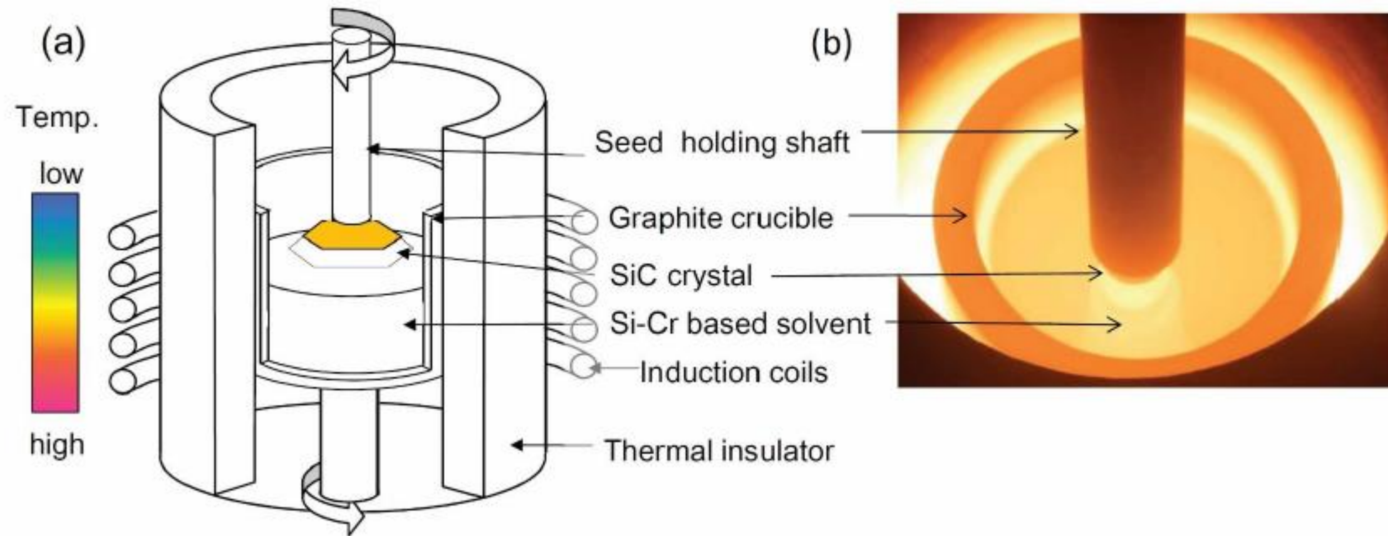
Motivation for SiC



SiC vs. GaN

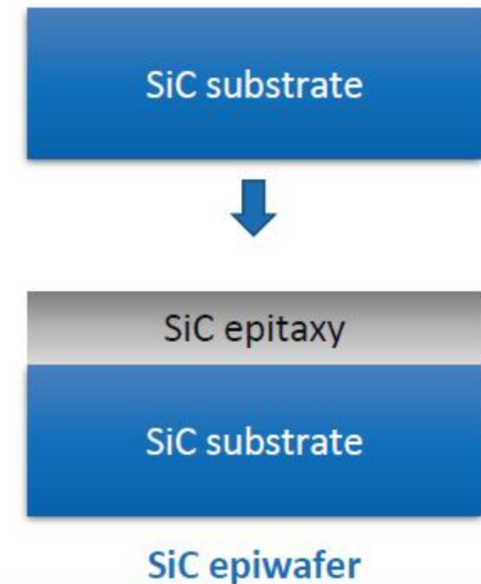


SiC Crystal Growth



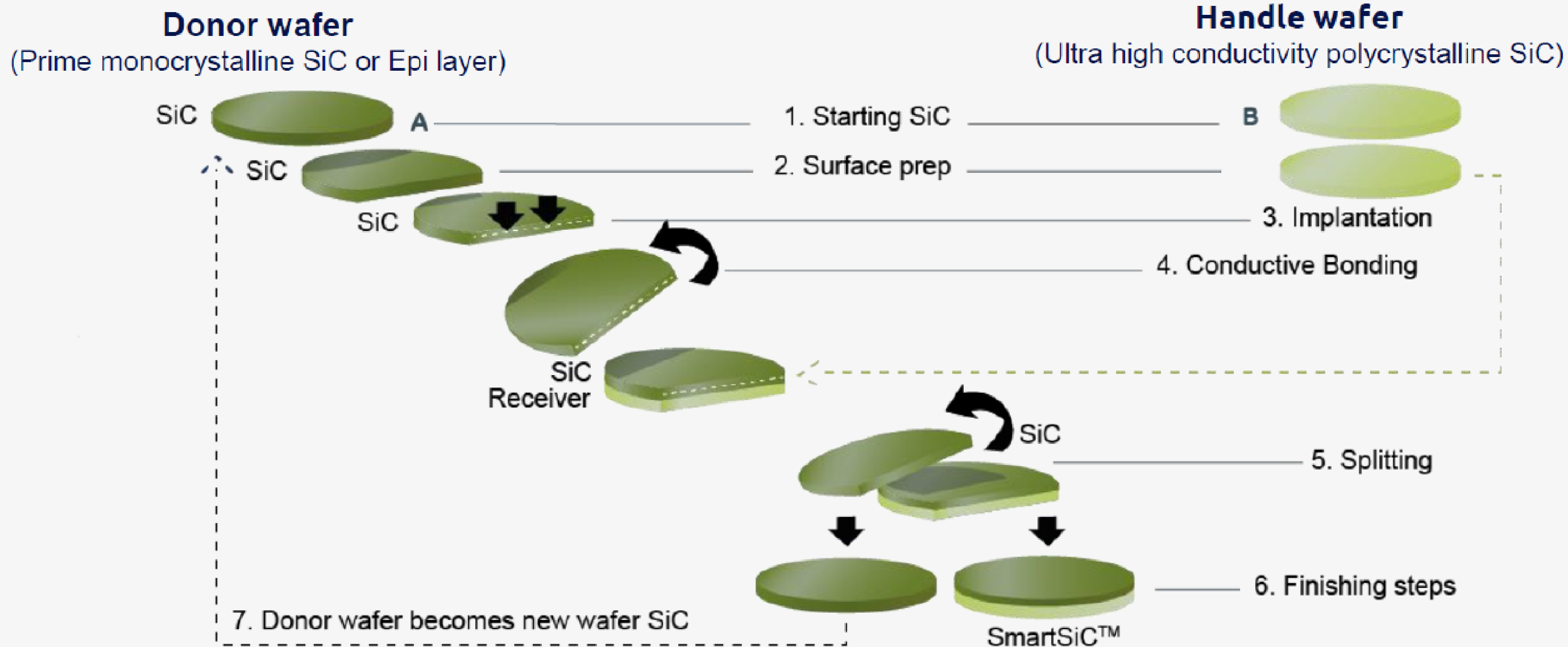
SiC Growth - Source: CS [Website]

- **Substrate**: Raw, Epi-ready wafer, ready to enter to epitaxy reactor.
- **Epiwafer**: Front End-ready wafer.



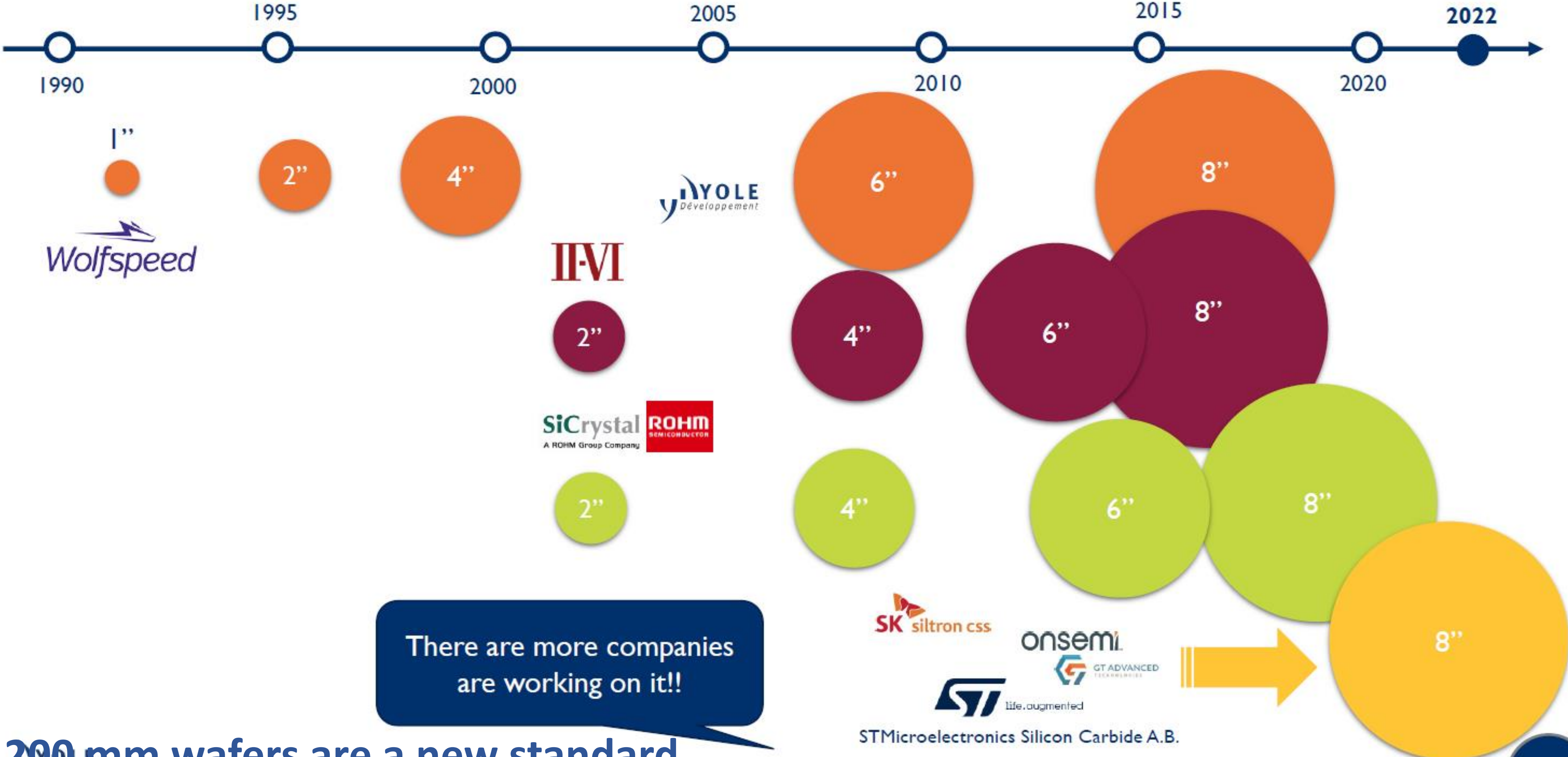
- **SiC Boule Growth Process:**
  - **Very slow (> 2 weeks)**
  - **Crystal defects are still a problem**

SOITEC Revolutionary Breakthrough



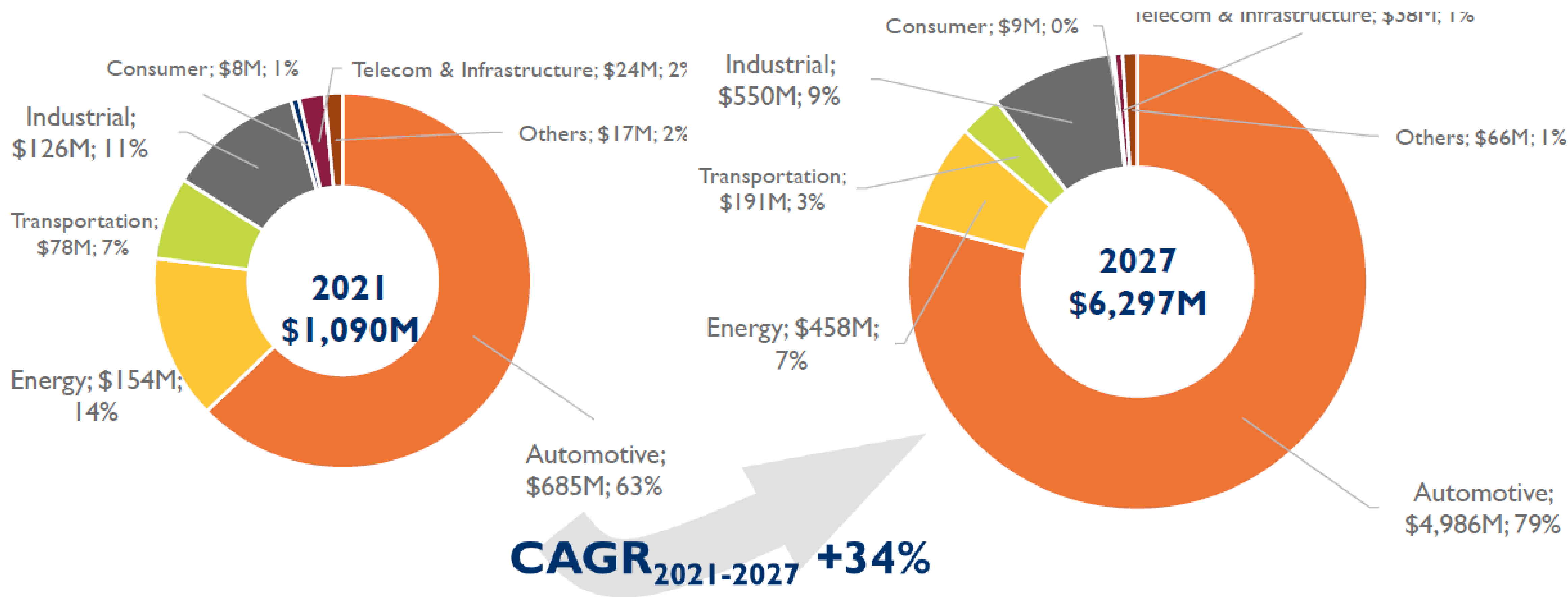
- Adaptation of SmartCut to SiC wafer production:
  - The single crystal SiC donor wafer can be reused up to 20 times

SiC Wafer Size Evolution



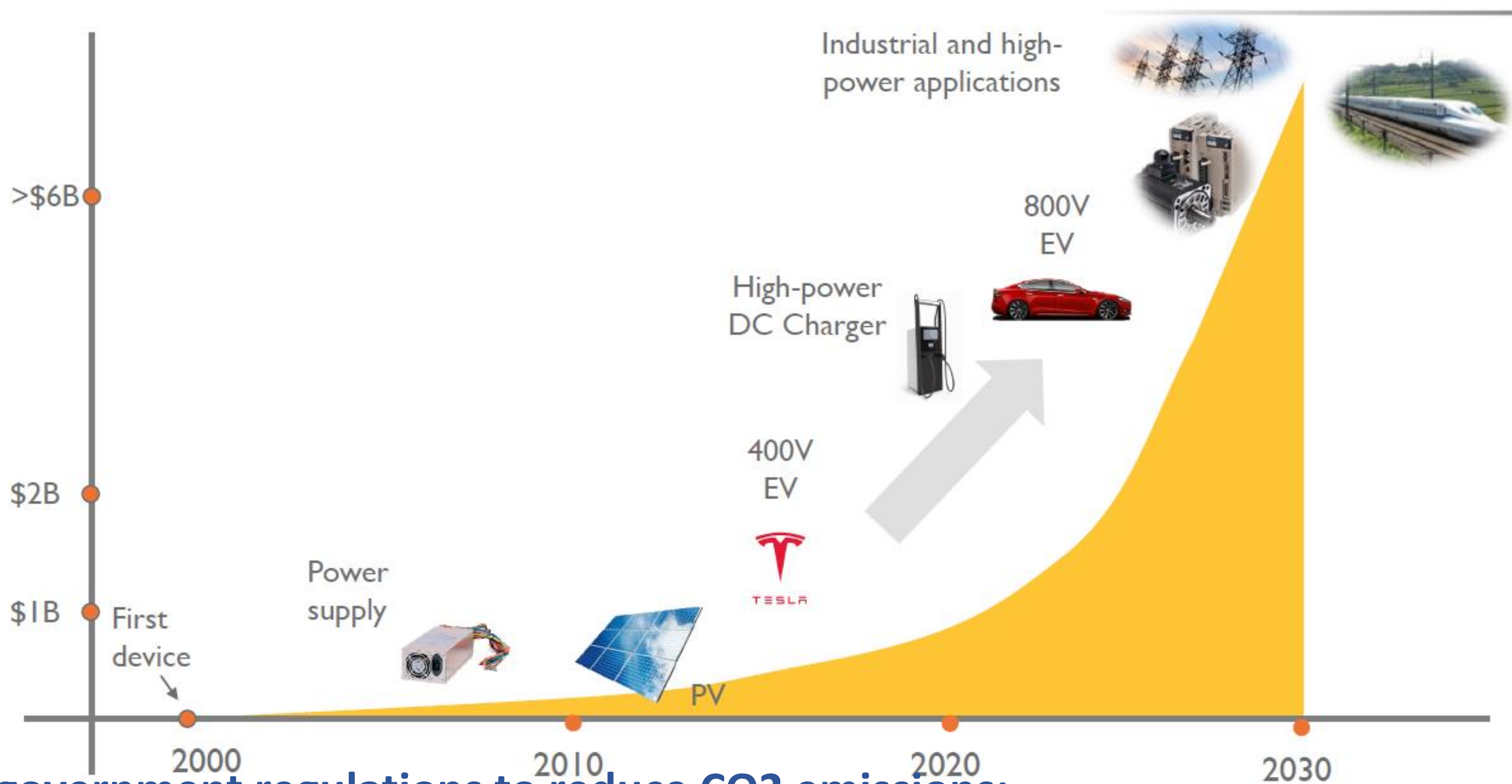
- ➢ 200 mm wafers are a new standard
- ➢ Vertical integration is a new trend
  - ➢ From SiC boule growth to the SiC Module Production (e.g., ST)

SiC Market Growth Prediction



- **Fastest growth rate in semiconductor industry:**
  - **Demand driven by Battery Electric Vehicles**

SiC Market Size Explosion



- New government regulations to reduce CO_2 emissions:
  - By 2035 the demand for 200mm SiC wafers will exceed the current worldwide production of all TSMC fabs