THE FUTURE OF LOGIC

Gate-All-Around Is Here, Now What?

Bala Haran

Vice President, Integrated Materials Solutions

Semiconductor Products Group

APPLIED MATERIALS®

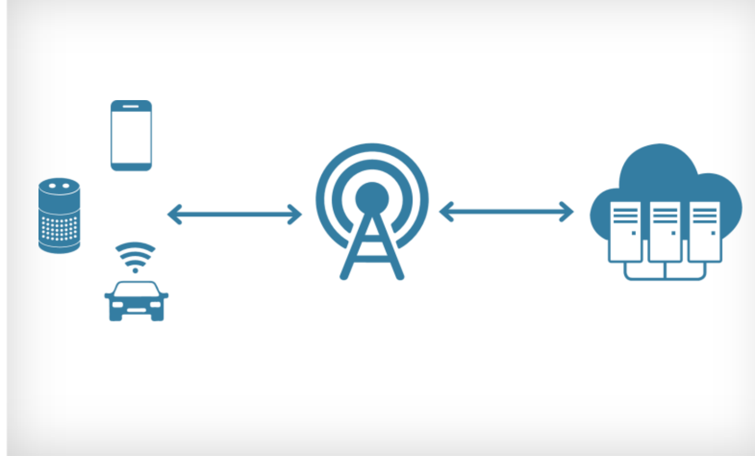
2019 Panel Takeaways

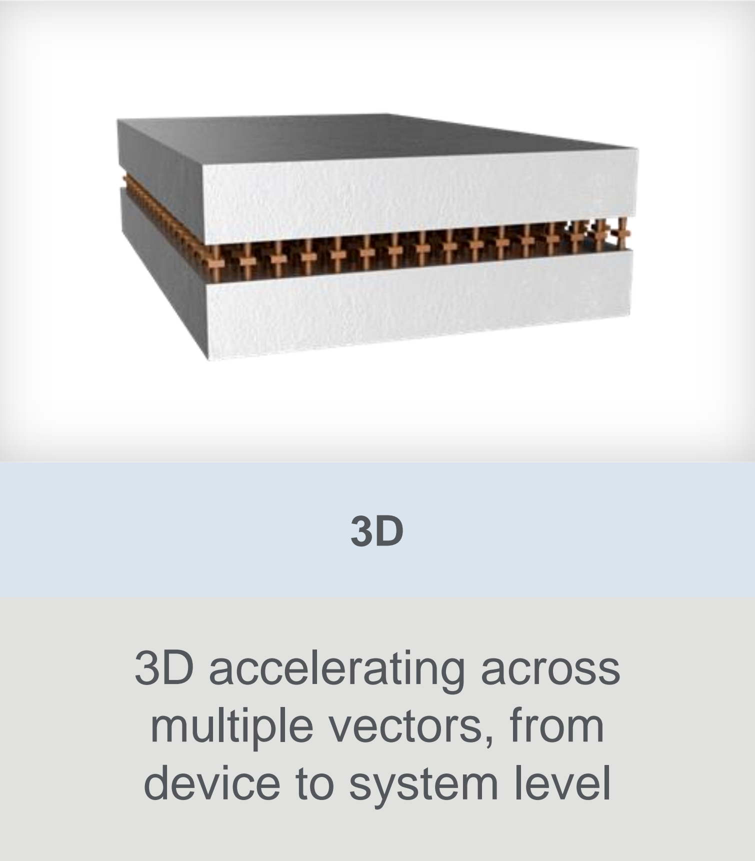
EUV Is Here, Now What?

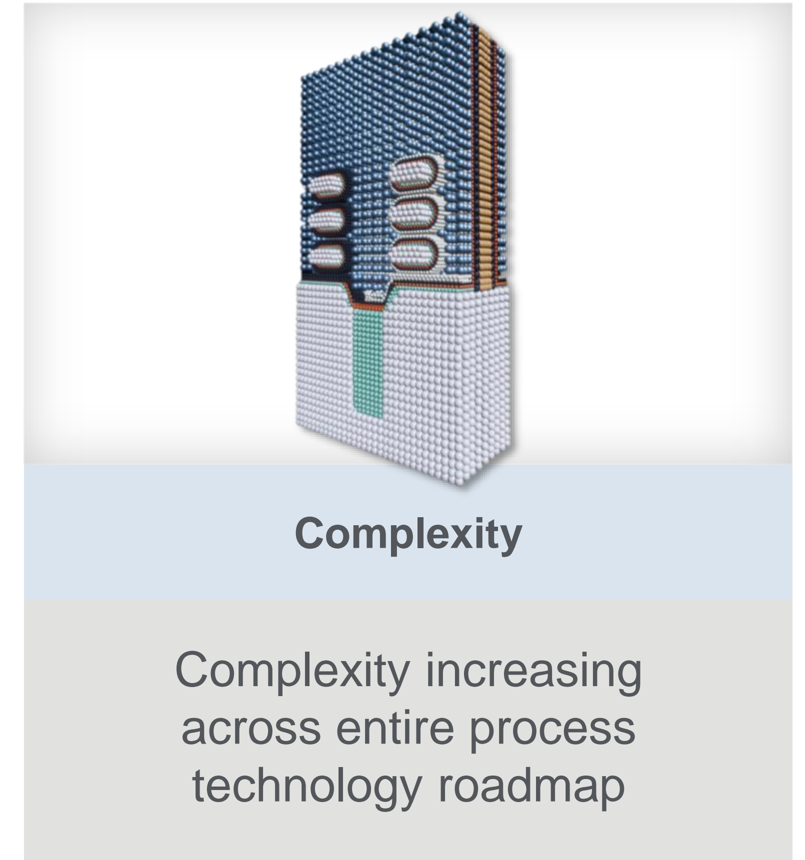
- AI beginning to shape the technology roadmap
- New approaches needed to drive innovation
- Requires co-optimization from materials to systems

A new playbook for collaborative innovation

What's Changed Since 2019?







IoT +AI	3D	Complexity
Cloud and edge AI driving new wave of growth and innovation	3D accelerating across multiple vectors, from device to system level	Complexity increasing across entire process technology roadmap

GAA Is Here, Now What?

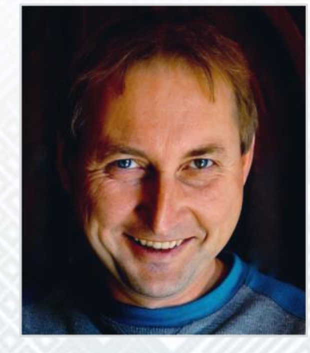
Nick Yu
Google
Custom Silicon Technologist

PR Chidambaram
Qualcomm
VP of Engineering

Kevin Fischer
Intel
VP, Interconnect and Memory Integration

Dong-Won Kim
Samsung
Fellow, Logic Technology Development

Geoffrey Yeap
TSMC
VP, Research & Development



Victor Moroz
Synopsys
Fellow

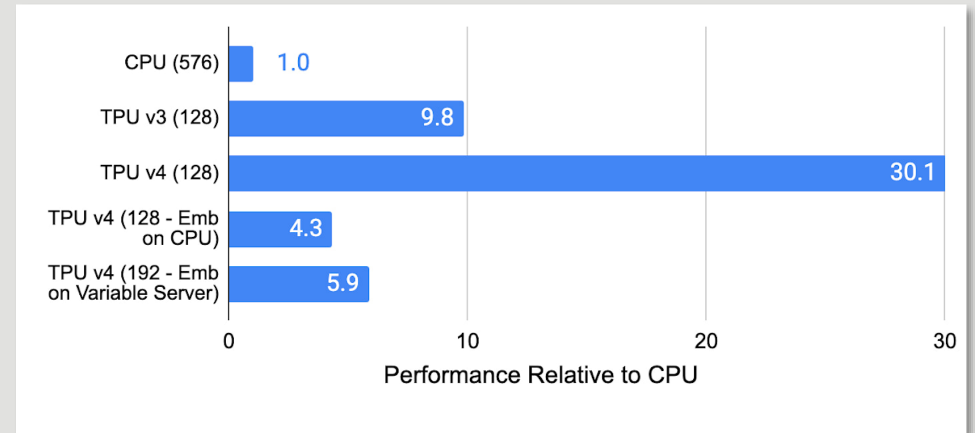
Paul Lindner
EV Group
Executive Technology Director

APPLIED MATERIALS

Google’s Cloud TPU v4

Google’s Cloud TPU v4 provides exaFLOPS-scale ML with industry-leading efficiency

April 5, 2023

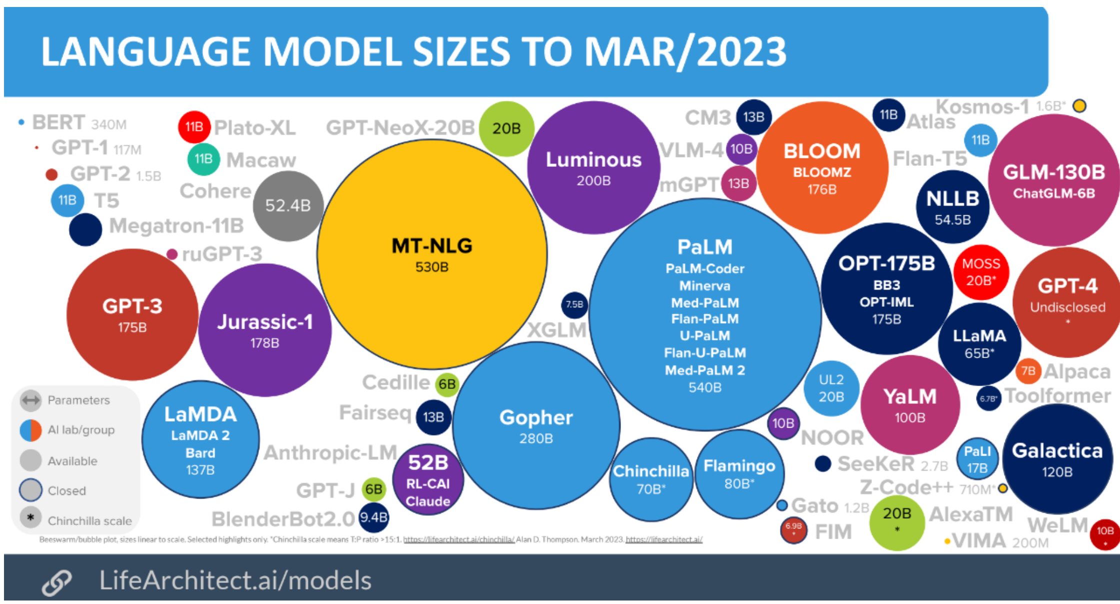


Google’s Cloud TPUv4 enabled:

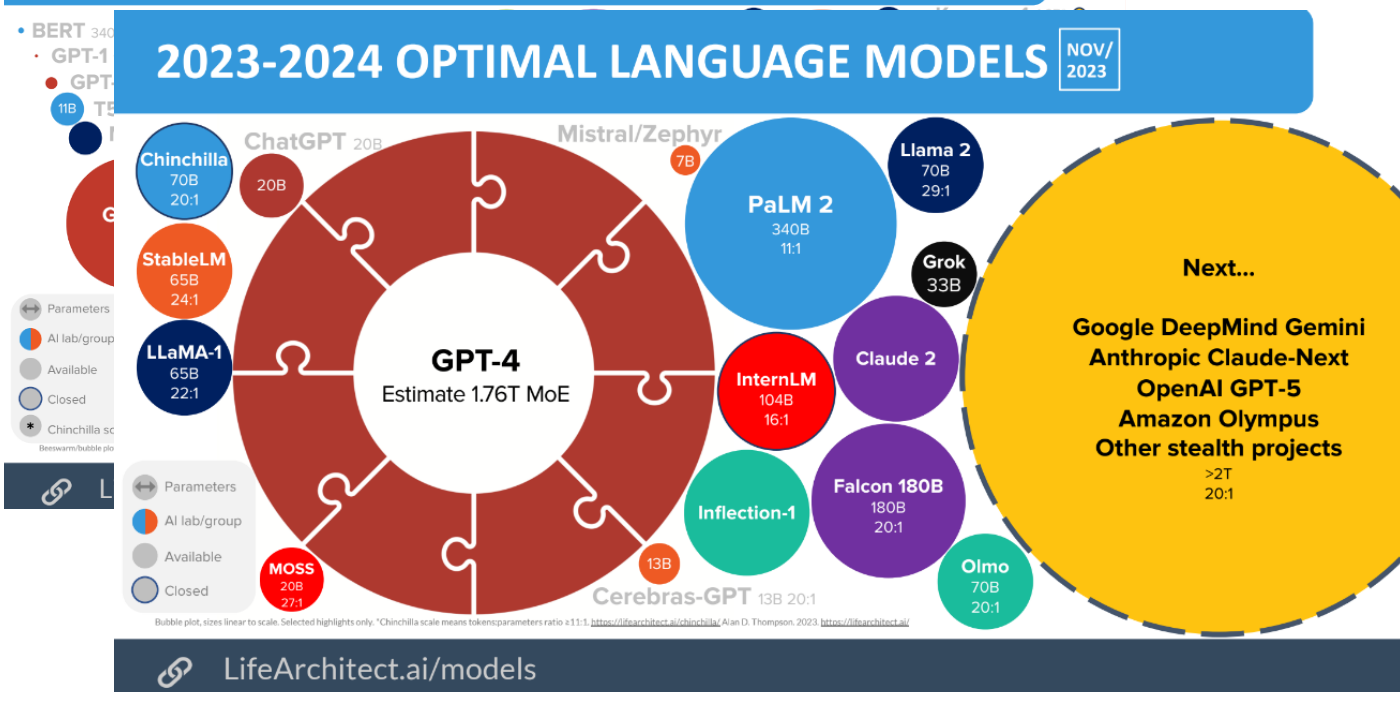
- a nearly 10x leap forward in scaling ML system performance over TPU v3
- boosting energy efficiency ~2-3x compared to contemporary ML DSAs, and
- reducing CO2e as much as ~20x over these DSAs in typical on-premise data centers

TPU v4 provides exascale ML performance, with 4096 chips interconnected by an internally-developed industry-leading optical circuit switch (OCS).

LLMs own Moore's Law?



LLMs own Moore's Law?



LLMs own Moore’s Law?

LANGUAGE MODEL SIZES TO MAR/2023

2023-2024 OPTIMAL LANGUAGE MODELS NOV/ 2023

Chinchilla 70B 20:1

StableLM 65B 24:1

LLaMA-1 65B 22:1

Parameters
AI lab/group
Available
Closed

https://huggingface.co/blog/large-language-models

Model Size (in billions of parameters)

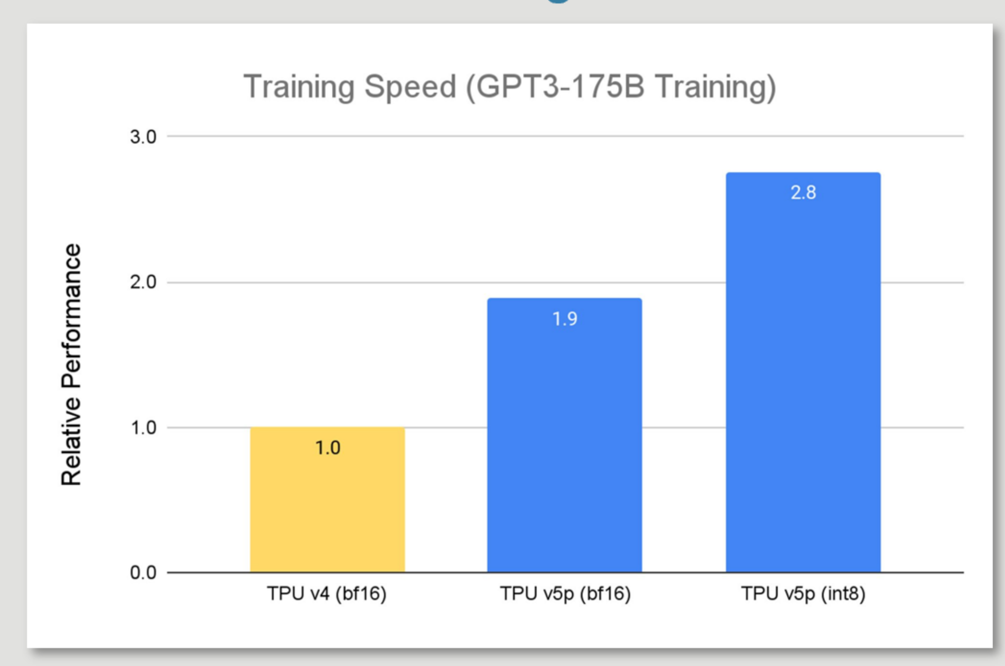
1000
100
10
1
0.1
0.01

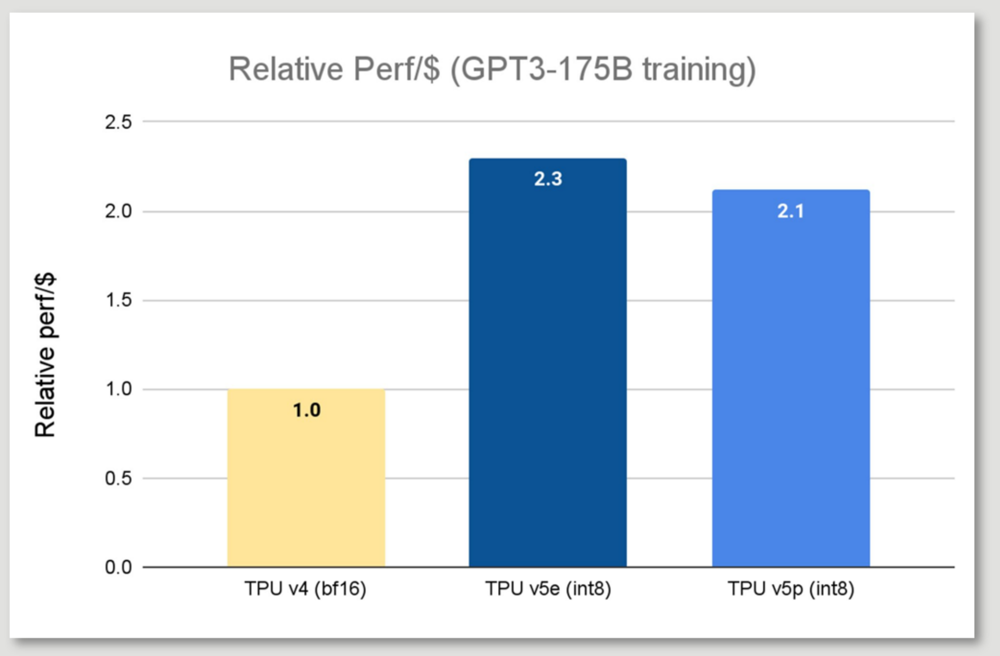
2018
2019
2020
2021
2022

ELMo (94M)
BERT-Large (340M)
GPT-2 (1.5B)
Megatron-LM (8.3B)
T5 (11B)
Turing-NLG (17.2B)
GPT-3 (175B)
Megatron-Turing NLG (530B)

LLMs own Moore's Law?

Google's Cloud TPU v5 *[Announced 6th Dec 2023]*





Will GAA/Nanosheet trajectory keep up with LLM's growth?

Gen AI to Accelerate Innovations in Our Industry

Key System Themes

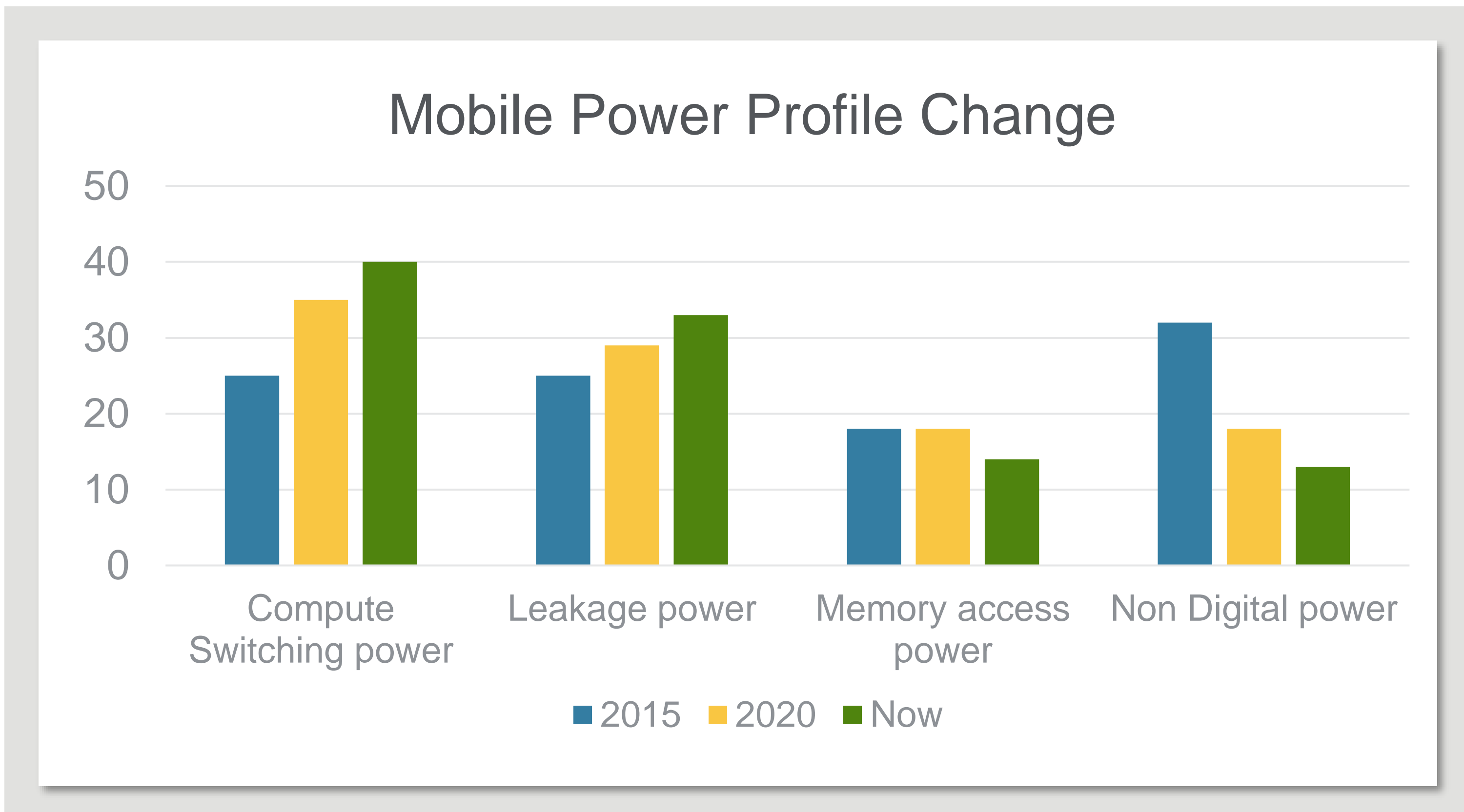
- GPUs vs DSAs
- Memory (more HBM, what's next)
- Chiplets (what's next)
- Interconnect (optical)
- Thermal
- Security
- Technology guardrails

Digging Deeper

- Die size scaling, intrinsic area scaling
  - Computation and Power density
- New memory - overcome logic-to-memory bottleneck
- Perf/TCO(OpEx) vs Perf/CapEx
  - Thermal/Power
- Silent Data Corruption eliminate/early-detect soft defects?

Digital Switching and Leakage Power Will Limit Industry

Qualcomm



- Product power reductions slowing for lack of true universal curve gain
- Multi-die/Chiplet mostly help with memory access power
- DTCO based power gains advertised hard to realize on product
- Vdd and leakage reduction essential

2023 ~~2019~~: “The Future of Logic. Gate All Around ~~EUV~~ is Here, Now What?”

- The State of Logic is STRONG
  (~~5nm mass production in 1H ’20~~)
  N3/N4/N5 in mass production
- The Future of Logic is BRIGHT
  (~~3nm and 2nm in development~~)
  2nm MP in 2025
- Adv. Logic Tech:
  1) ~~Mobile~~ AI mobile
  2) ~~High Performance Computing~~ AI data center scale computing
  3) ~~Automotive~~ AI driving
  4) ~~IOT~~ AI everywhere

- PPACt: traditional scaling ➔ DTCO scaling ➔ system scaling/integration
- Power supply scaling and low Vdd operation
- More energy-efficient and high-performance transistors
- Houston, we have ~~a problem: Interconnect Resistances~~ BSPDN

System Scaling Innovation

Enabled by TSMC Adv Logic + 3DFabric™ (CoWoS, SoIC, and InFO) Technologies

>1T TRANSISTORS

>500B TRANSISTORS

200B

H100

MI300X

>200B TRANSISTORS

>100B TRANSISTORS

3D Hetero. Integration

20~50B

>3000 Cores

10B

>2000 cores

1B MOS Transistors

Multi-million Transistors

Monolithic Integration

New Channel Materials
EUV
Metal Oxide ESL
Self-Aligned Line w/ Flexible Space
Low Damage/Hardening Low-K & Novel Cu Fill

3D FinFET
2P2E
ELK
Low-R Barrier
Co Cap/Liner

HKMG
Immersion
SiGe
Cu/Low-K

Mass Production: 2020 2023 2025 2027 2030

AI Revolution Driven by Semiconductors

Perf/Watt/mm^2

> 130X Improvement

N28 N16 N10 N7 N5 N3 N2

Cu
Strained Si
HK / MG
Immersion
FinFET
InFO
CoWoS
3D stacked DRAM
EUV
Chiplets
SoIC
3D stacked SRAM
Nanosheet

DeepBlue
MNIST
LSTM
ImageNet
Watson
AlexNet
Attention
AlphaGo
Transformers
BERT
GPT-3
AlphaFold
DALL-E
ChatGPT
GPT-4
AlphaCode

1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022 2024

0.13 µm 45 nm 28 nm 4 nm

- More Bandwidth
- More Memory
- More Compute

Intel Roadmap: Next Step in 5 Nodes in 4 Years

Performance Per Watt

Intel 90nm — Strained Silicon

Intel 65nm — Enhanced Strain

Intel 45nm — Hi-K Metal Gate

Intel 32nm — Enhanced HKMG

Intel 22nm — First FinFET

Intel 14nm — Enhanced FinFET

Intel 10nm — COAG

Intel 10nm SuperFin — Super MIM Capacitor; Enhanced FinFET

Intel 7 — Transistor optimization for performance; Metal stack enhancements

Intel 4 — EUV Litho

Intel 3 — Denser design libraries; Increased transistor drive current; Reduced via resistance; Increased use of EUV

Angstrom era

Intel 20A — RibbonFET & PowerVia

Intel 18A — Ribbon innovation increased performance; Continued metal linewidth reduction

Intel NEXT

Time

PowerVia and RibbonFET on Intel 20A and Intel 18A

- Improved perf. per watt
- Improved density
- Reduced node over node cost increases
- Improved design flexibility

Learn more at www.Intel.com/ProcessInnovation

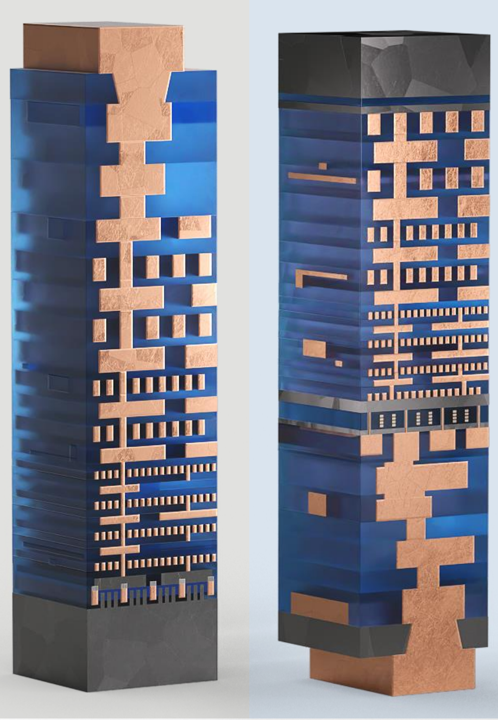
Frontside Power Delivery vs PowerVia

Frontside Power Delivery

Signal wires and power wires compete for the same resources at every metal layer

Requires aggressive scaling of metal layer pitches:

- High cost
- Higher voltage droop
- Higher RC delay



PowerVia Backside Power Delivery

Signal wires and power wires are decoupled and optimized separately

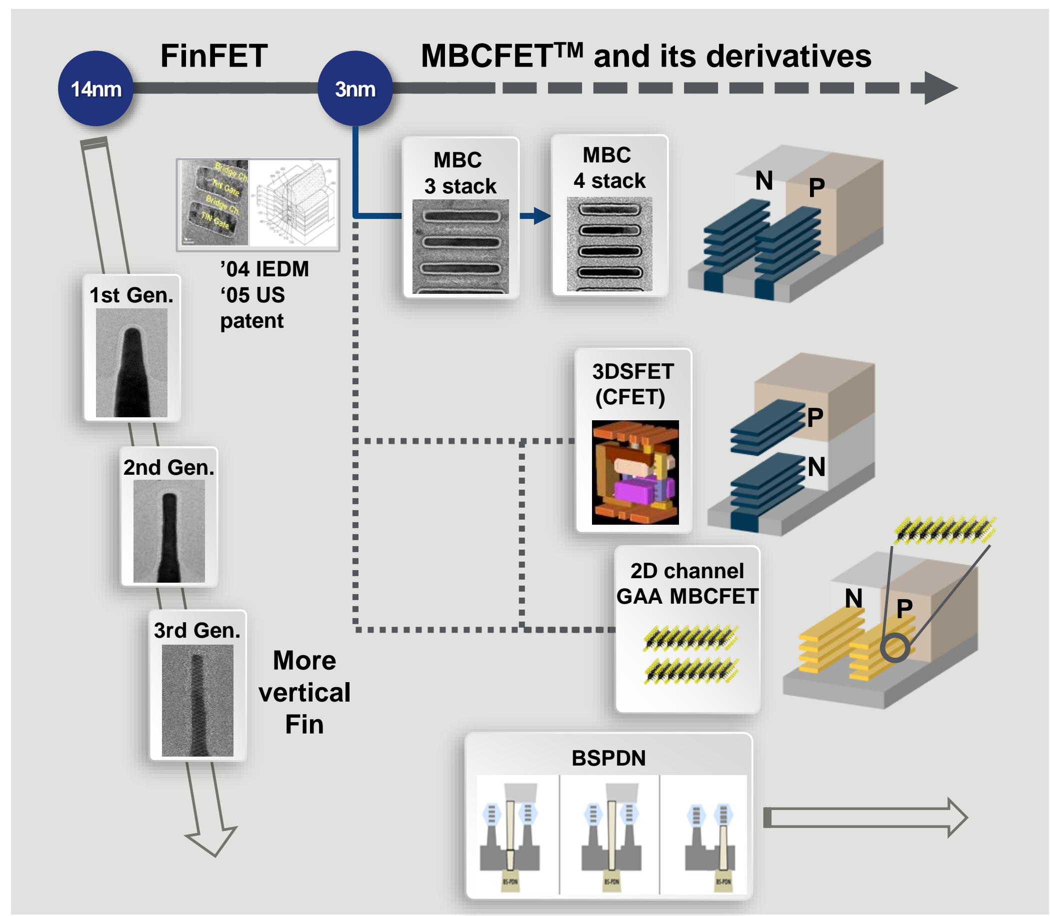
Value Proposition:

- Better Performance
- Lower signal RC
- Reduces voltage droop
- Lower Cost
- More Design Flexibility

Intel Innovation 2023

MBCFET™ Learning and Future

SAMSUNG

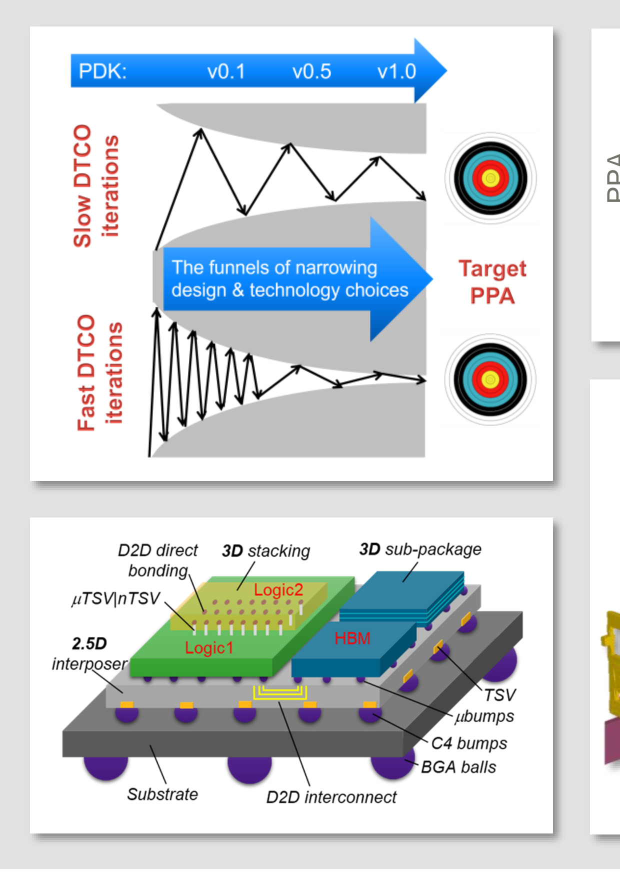


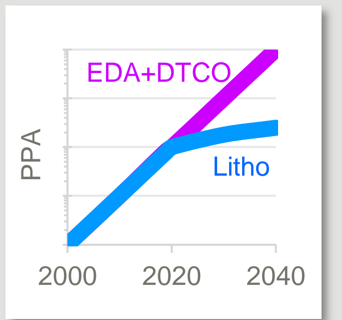
MBCFET (Multi-Bridge-Channel FET)

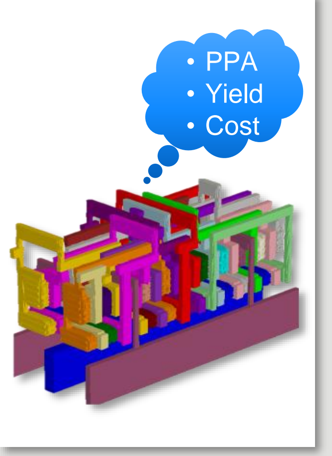
Key Messages

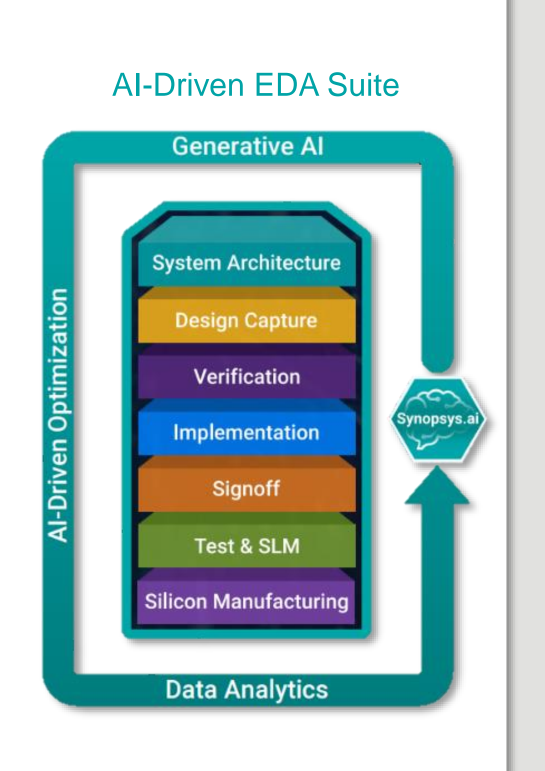
- Learnings from the 3nm MBCFET ramp
  - » Metrology & Inspection
  - » Limited spacing for metal gate
  - » Co-optimization of devices with wide/narrow widths
- View of the roadmap for upcoming inflections
  - » Extending the MBCFET era
    - – Additional nanosheets for enhanced drive current and cell scaling
    - – Introduction of CFET and/or BSPDN for STD cell scaling extension
    - – Implementation of 2D Materials in GAA MBCFET architecture
  - » Implementation of backside power delivery
    - – Improved IR drop and area gain corresponding to frontside PDN
    - – Additional performance/power gain and increased design freedom with less congestion
    - – Increased the difficulty of DTCO and STCO to utilize BSPN

EDA Role in Advanced CMOS Eco-System





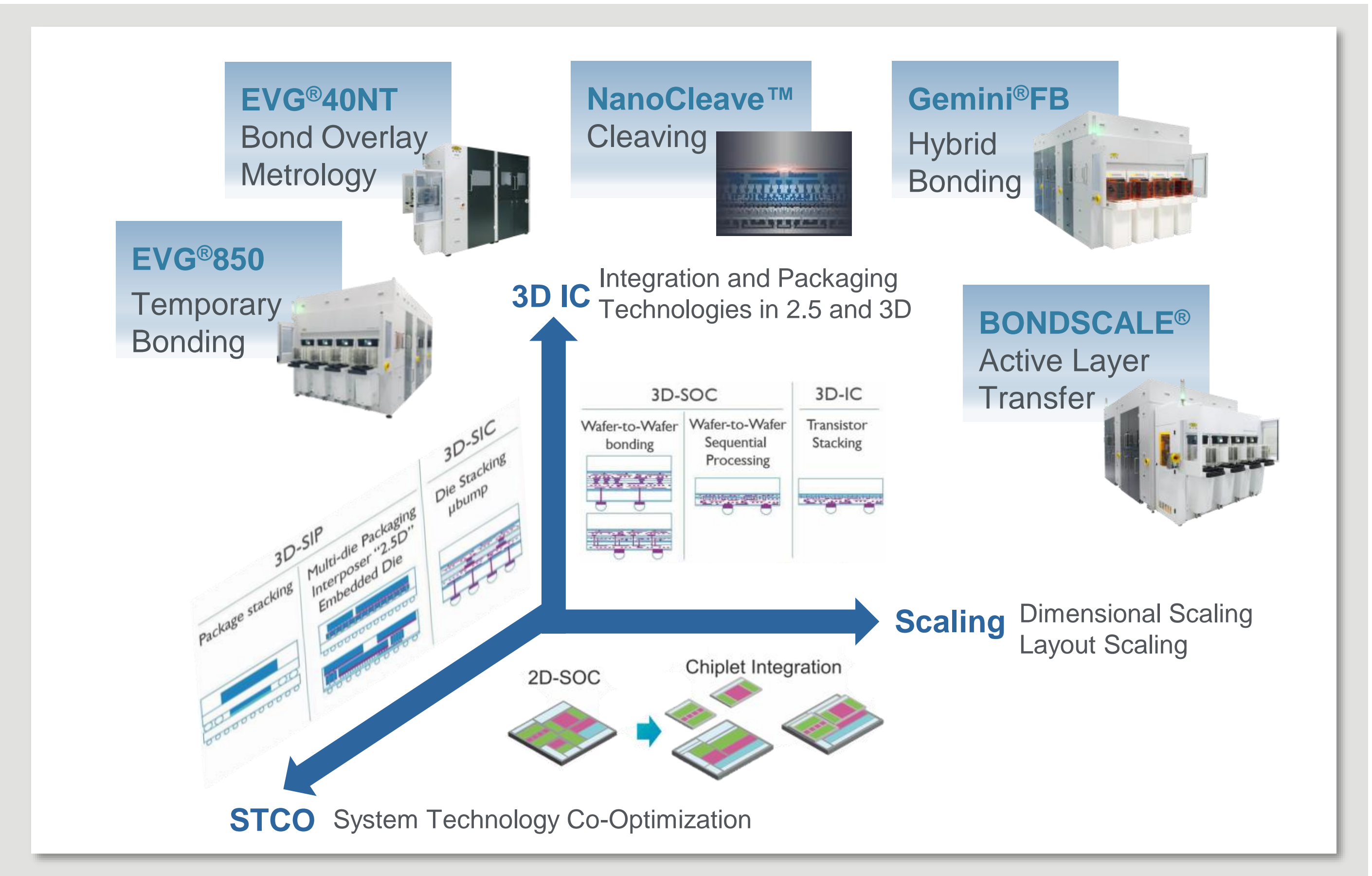




Key Messages

- Collaboration with Foundries & Fabless to Optimize EDA tools for
  - GAA & CFET (Si, SiGe, stress)
  - BS PDN (power, clock, signal)
  - Chiplets (hybrid bonding, I/O devices)
  - Library optimization (track height mix)
- Fast DTCO exploration to boost PPA for specific chip applications
  - 3 days turn around time
- Collaborative and Generative AI
  - DSO.ai (Design Space Optimization)
  - Synopsys.ai Copilot (full EDA stack)

Wafer Bonding is a Scaling Booster



- Wafer Bonding
  - » Complements the traditional dimensional scaling
- Wafer-to-Wafer Hybrid Bonding:
  - » Enables higher functional density at the wafer level back end
- Active Layer Transfer Fusion Bonding:
  - » Exploits the wafer back side and enables front end transistor level stacking

GAA Is Here, Now What?

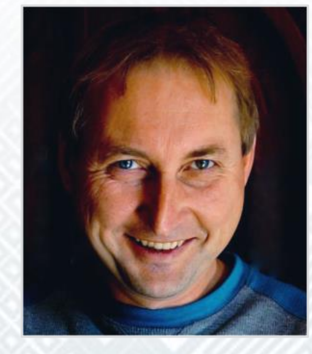
Nick Yu
Google
Custom Silicon Technologist

PR Chidambaram
Qualcomm
VP of Engineering

Kevin Fischer
Intel
VP, Interconnect and Memory Integration

Dong-Won Kim
Samsung
Fellow, Logic Technology Development

Geoffrey Yeap
TSMC
VP, Research & Development



Victor Moroz
Synopsys
Fellow

Paul Lindner
EV Group
Executive Technology Director

APPLIED MATERIALS

Thank You for Joining Us!