The Increasing Verification Horizon in the Era of AI-Driven Pervasive Intelligence

Vikas Gautam, VP Engineering, Synopsys
Subrangshu Das, Director Design, Google Silicon

September 2024

New Opportunities Driven by AI

Pervasive Intelligence enabling our lives ahead

A video from Swaayatt Robots shows the company's self-driving car navigating hectic streets. Swaayatt

A self-driving car from the startup Minus Zero relies on an end-to-end model of autonomous driving. Minus Zero

THE LATEST DEVELOPMENTS IN TECHNOLOGY, ENGINEERING, AND SCIENCE

News

The Bhopal-based startup Swaayatt Robots has been testing its autonomous driving algorithms on Indian roads, facing unpredictable driving conditions with sometimes adversarial drivers, roaming animals, and unruly scooters.

TRANSPORTATION

Startups Say India Is Ideal for Testing Self-Driving Cars ›
Unruly traffic forces innovative approaches to autonomy

BY EDD GENT

After experiencing New Delhi's chaotic roads, former Uber CEO Travis Kalanick famously said that India will be the last place in the world to get self-driving cars. But a handful of startups think the country could be the perfect testbed for creating autonomous vehicles that can handle anything.

A recent video from Bhopal-based startup Swaayatt Robots suggests they're making progress. In the 6-minute clip, a sensor-laden SUV weaves through narrow unmarked streets, dodging pedestrians, dogs, cows, slow-moving tractors, and a constant stream of scooters overtaking other vehicles, cutting across lanes, and even driving on the wrong side of the road.

Swaayatt CEO Sanjeev Sharma says the video highlights the two major characteristics of Indian traffic that make it so challenging. It is both stochastic and adversarial. In simpler terms that means road conditions and driver behavior are almost entirely unpredictable and that other road users are more likely to play chicken than give way. (In fact, the stochastic quality of driving has led the company to probabilistically model no small part of its traffic simulations.)

While self-driving cars developed by Western technology companies have already begun commercial operations, this rollout has been made possible only by training on millions of miles of driving data painstakingly gathered over many years. And despite all that training, these companies are still bedeviled by the "long-tail problem"—the idea that no matter how many scenarios you train on, you will

LEFT: SWAAYATT ROBOTS; RIGHT: ISTOCK

6 SPECTRUM.IEEE.ORG JUNE 2024

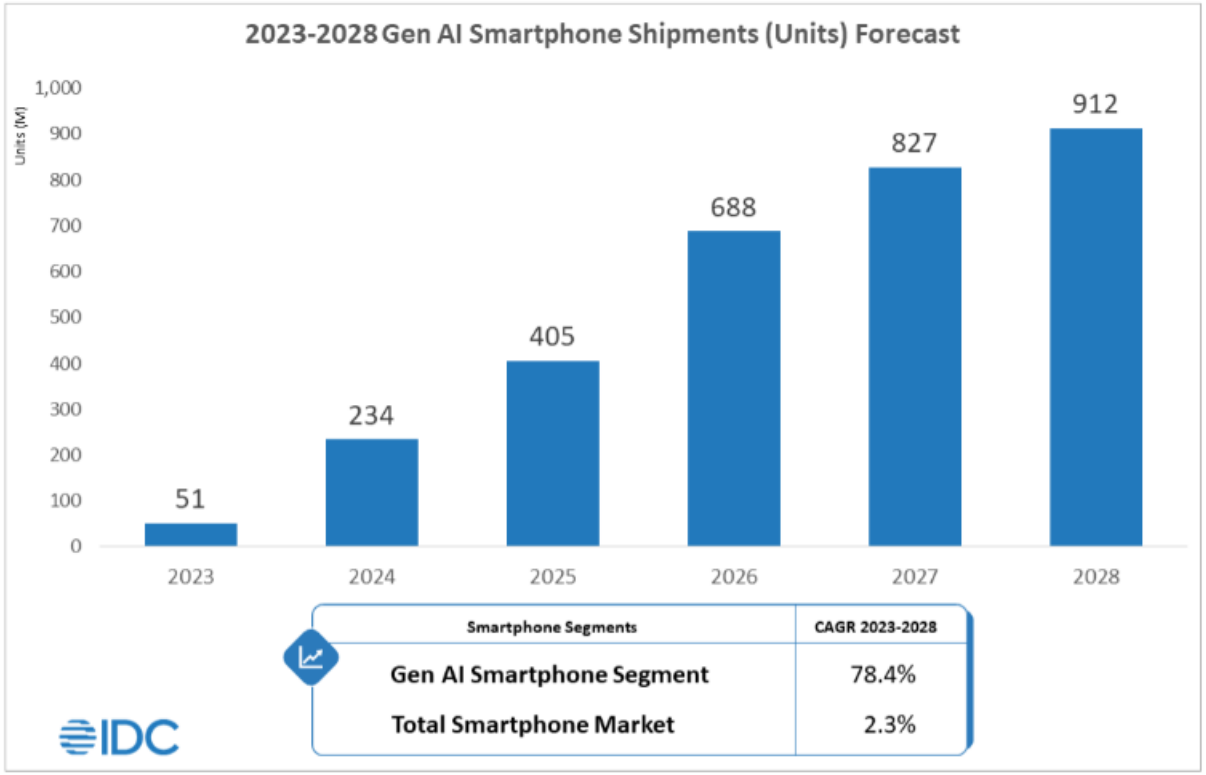
Source: IEEE Spectrum, June 2024, https://spectrum.ieee.org/india-self-driving-car

AI Reshaping the Mobile Experience

Pervasive Intelligence enabling our lives ahead

- Pixel 9 phones also have new camera features
  - **Add Me**, which lets you add yourself to group photos,
  - Super Res Zoom Video, which lets you zoom in on videos without losing quality

- Gemini is the new AI powered personal assistant in Android phones

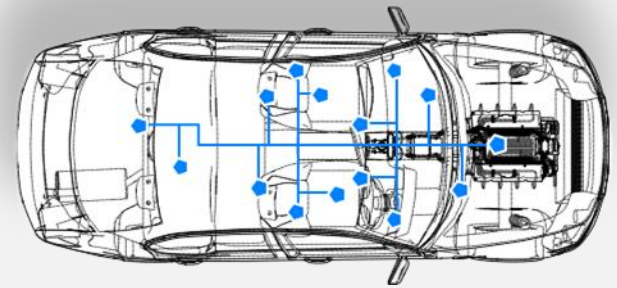


Source: https://blogs.idc.com/2024/07/05/the-rise-of-gen-ai-smartphones/

Source: https://blog.google/products/pixel/google-pixel-9-new-ai-features/#pixel9phones

Pervasive Intelligence Driven By Increasing Silicon and Software Content In Systems Products

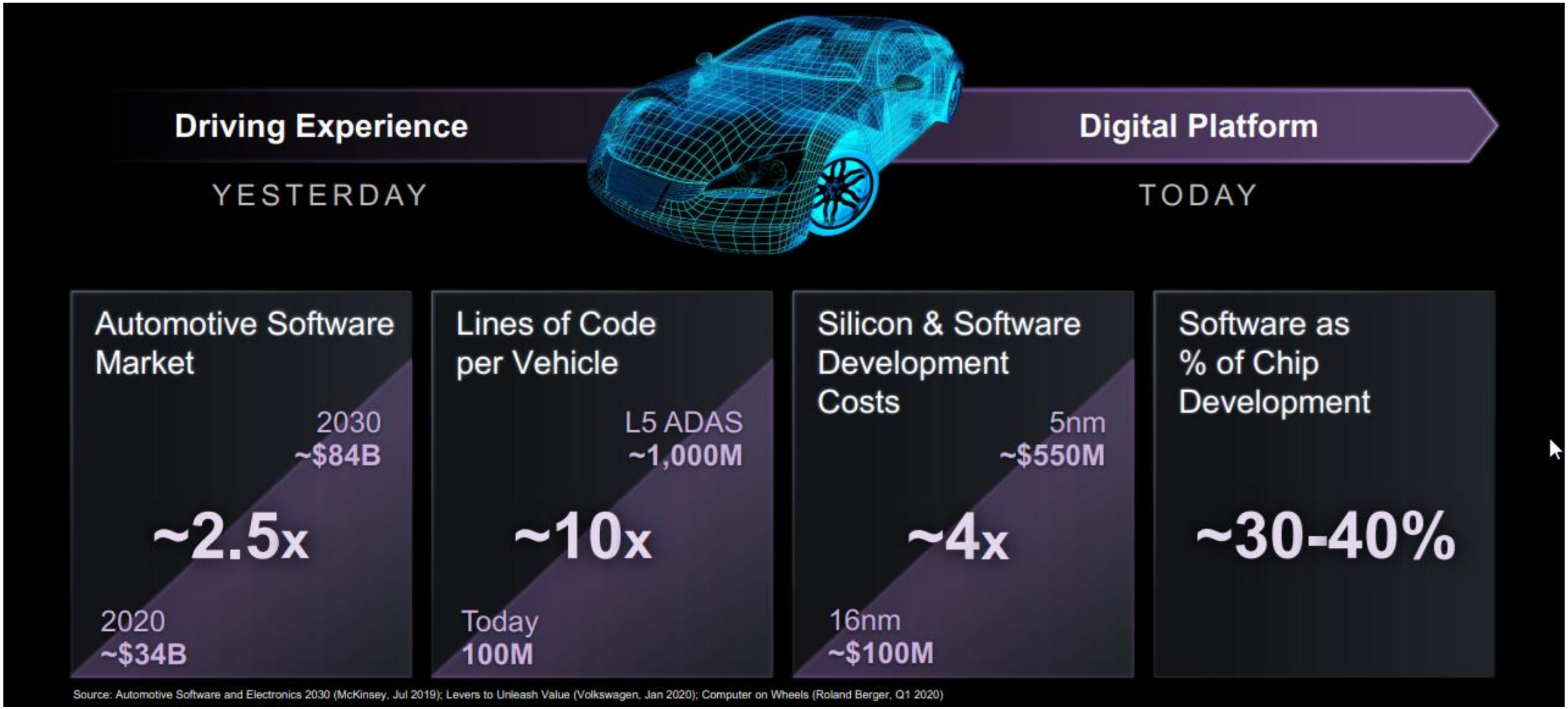
Autonomy and software-defined systems reshaping industries



Systems companies re-architecting products, business models and development processes

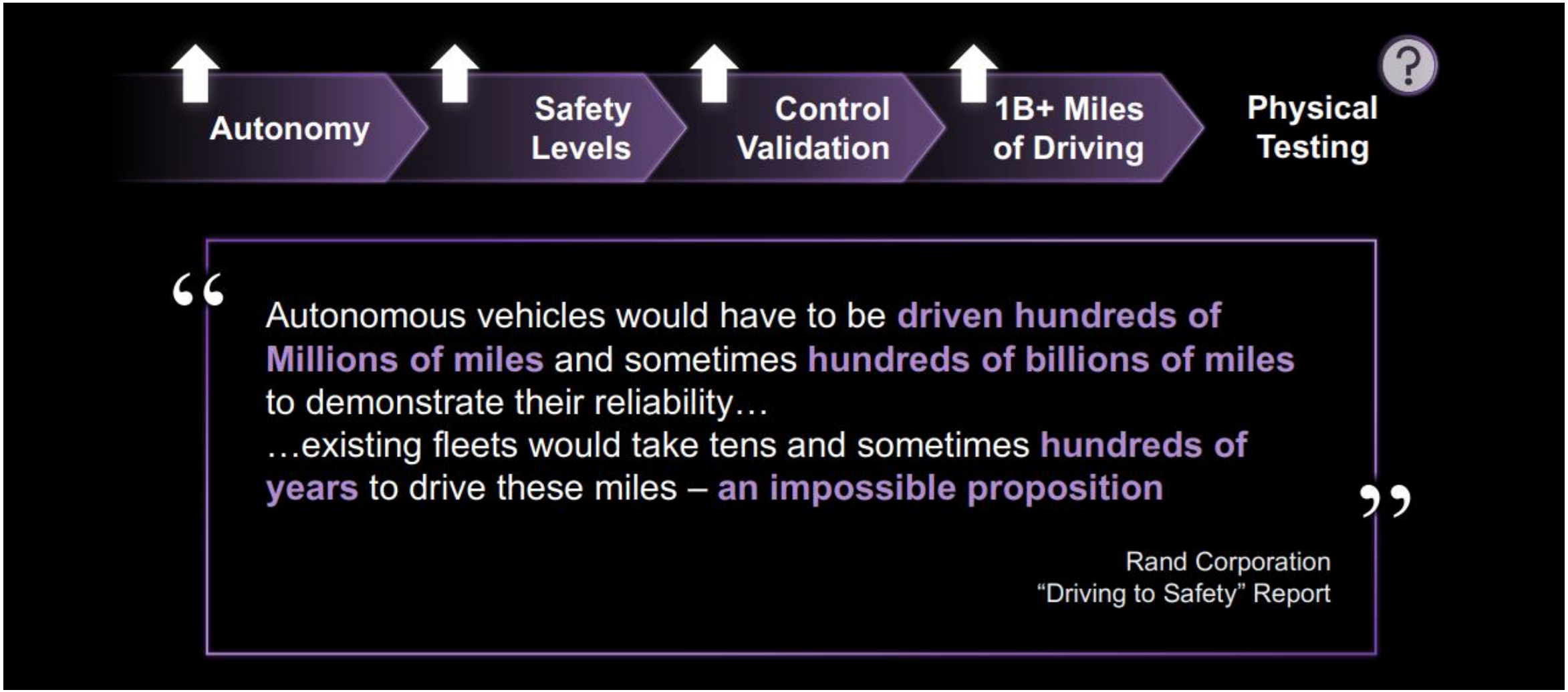
Driving demand for massive compute, both at the edge and the data center

Massive Challenges Facing Auto Industry Forcing Change



Source: Synopsys Investor Day

…and Previous Product Testing Methods Are Unsustainable



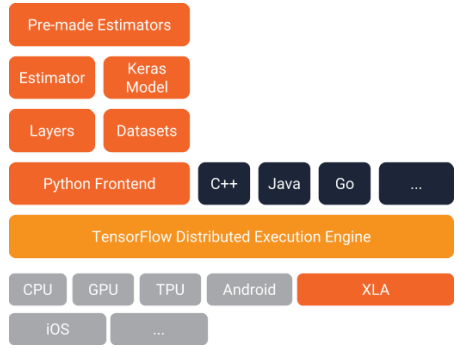
Source: Synopsys Investor Day

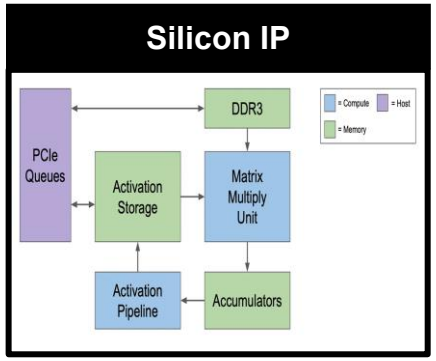
Data Center: Google Tensor Inference & Training Accelerator

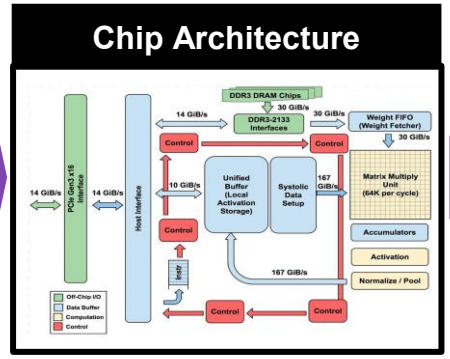
From **Silicon IP** to **System Hardware** that runs **Software**

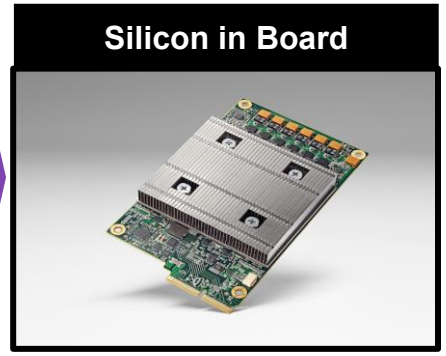
"With TPU v5p and AI Hypercomputer efficiencies, we achieved a whopping **2.5X** increase in training speed! The 6th generation of Trillium TPUs are incredible with a **4.7X** increased compute performance per chip and **2X HBM** Capacity and Bandwidth improvement over the previous generation." |

Yoav HaCohen, PhD, Core Generative AI Research Team Lead, Lightricks









Sources: What is Google's Tensor chip? Everything you need to know
Google's First Tensor Processing Unit : Architecture
Introduction to TensorFlow Datasets and Estimators
Announcing Trillium, the sixth generation of Google Cloud TPU

Google

Verification Use Cases in the Era of Pervasive Intelligence

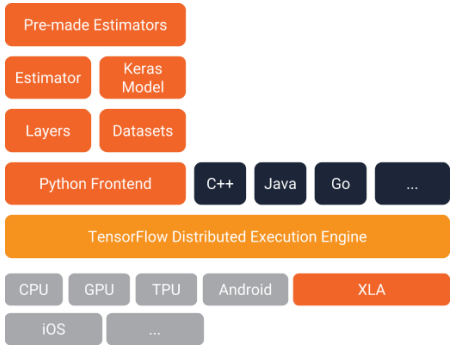
From **Silicon IP** to **System Hardware** that runs **Software**

Early Software Bring-up, Hardware/Software Validation

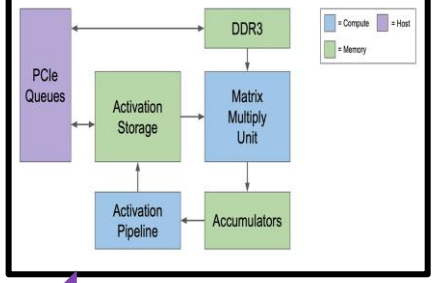
Functional Verification, RTL Regressions

IP in System Validation, Performance Validation, Compliance

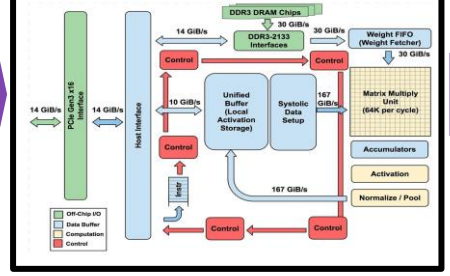
Workloads



Silicon IP

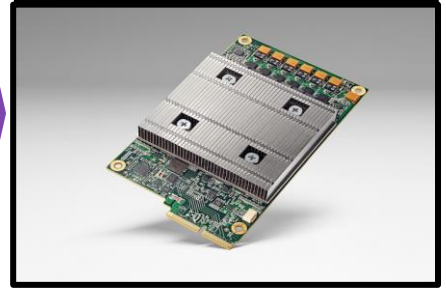


Chip Architecture



Silicon

Silicon in Board



Accelerator System

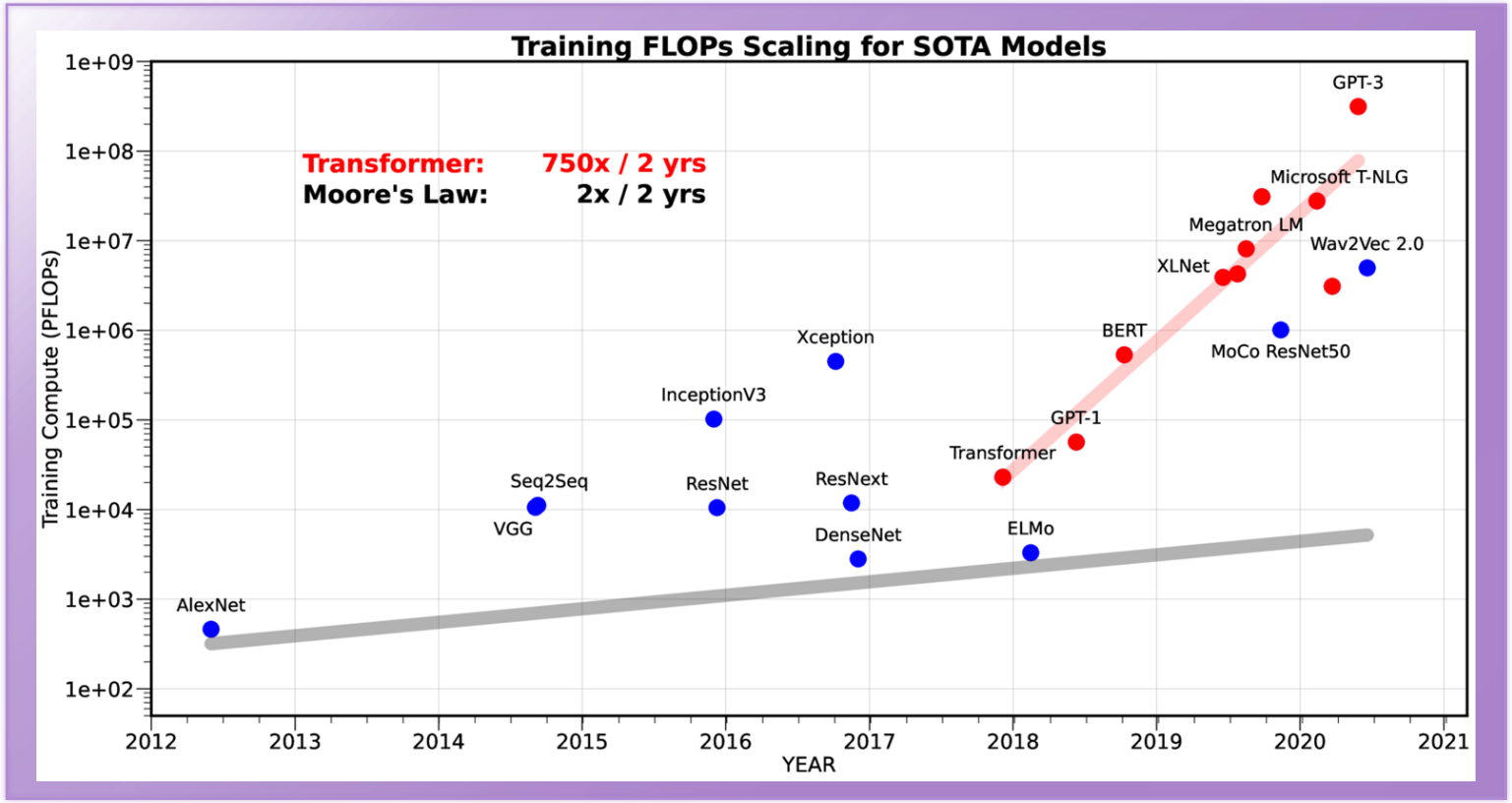
Early Performance and Low Power Analysis, Test Generation

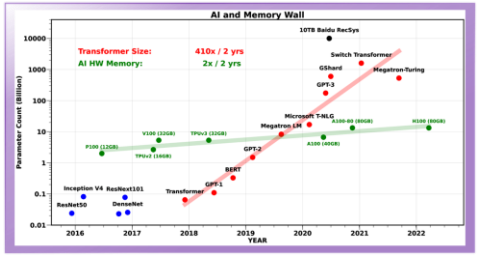
Safety, Reliability, Security, Design for Test, Silicon Lifecycle Management

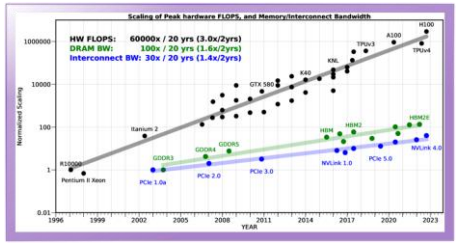
AI Requirements Drive New Design Paradigms

Key Driver: The Need for More Compute

The Processing Wall





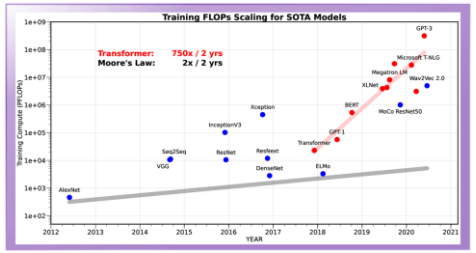


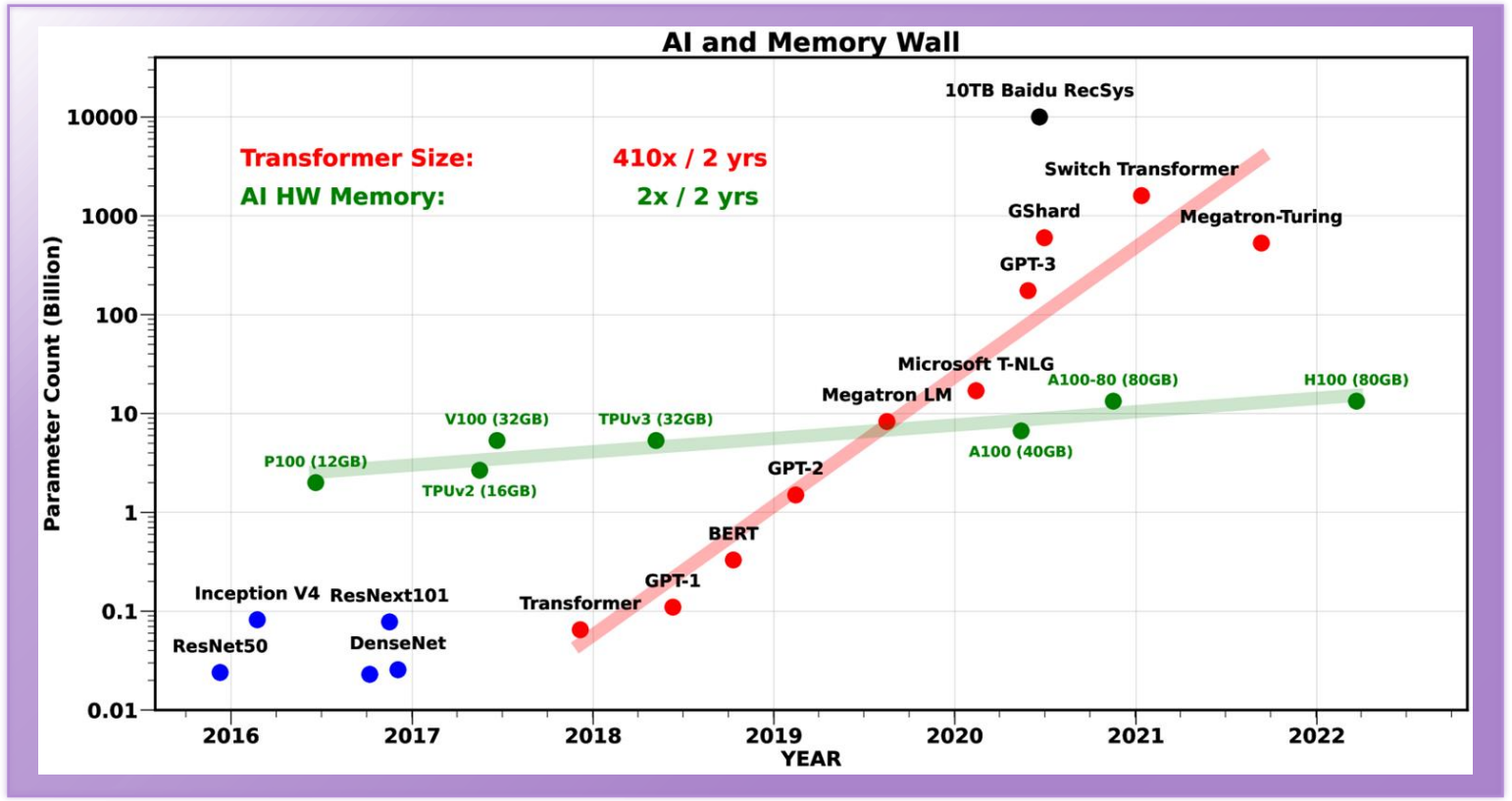
The “Processing Wall”

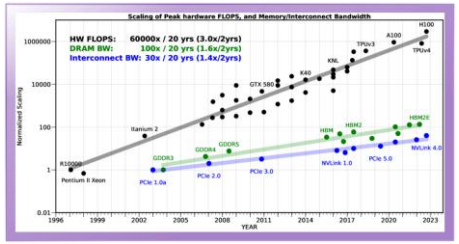
Source: AI and Memory Wall: 2403.14123 (arxiv.org), Synopsys

Key Driver: The Need for More Compute

The Memory Wall





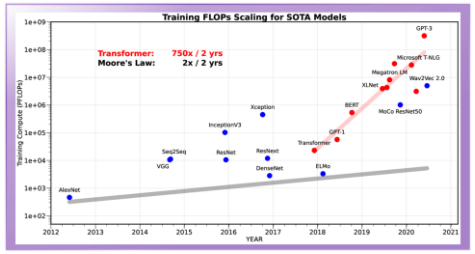


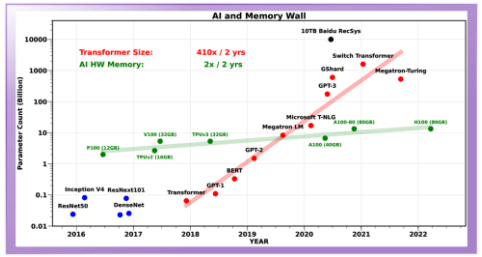
The "Memory Wall"

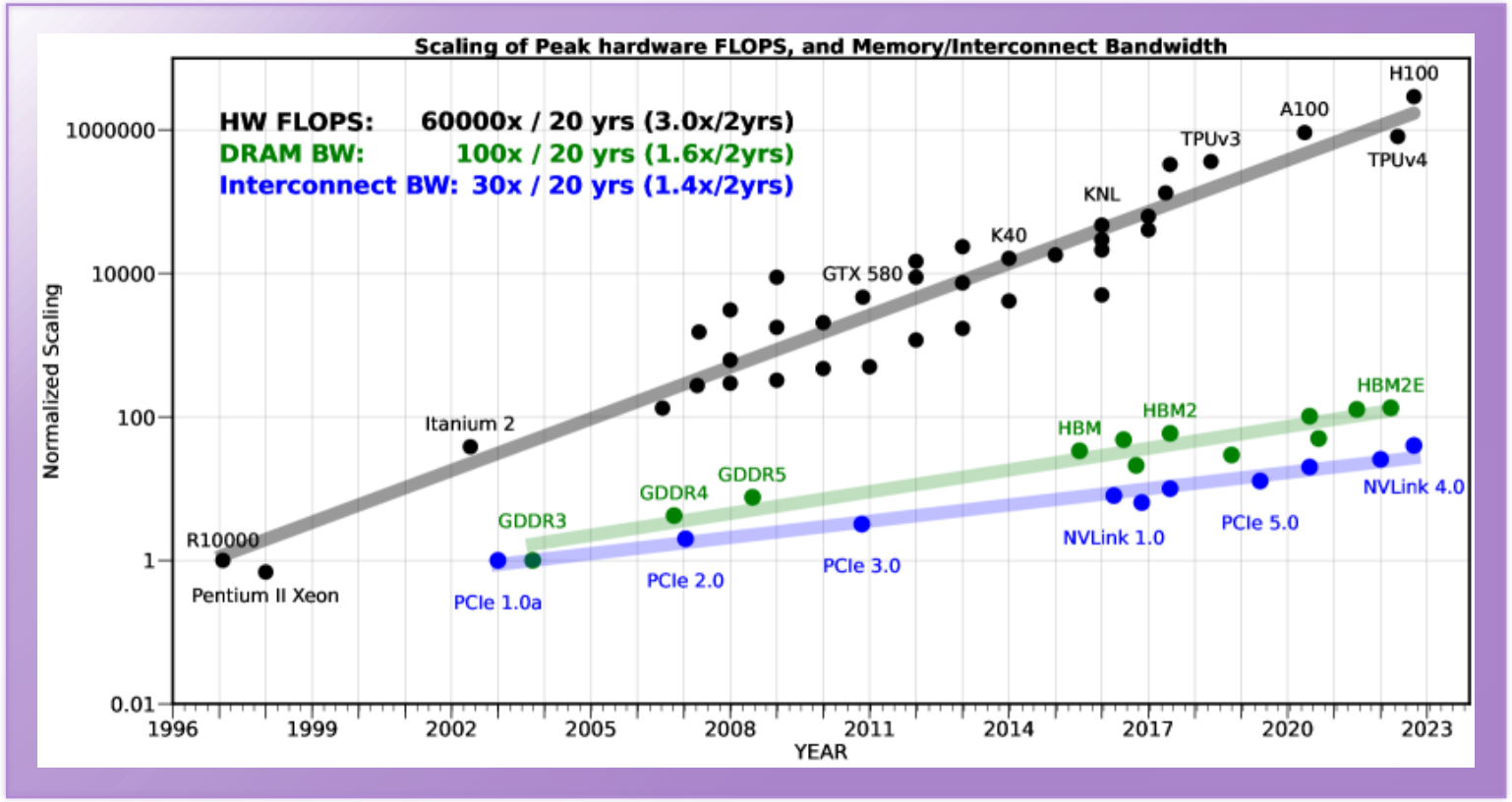
Source: AI and Memory Wall: 2403.14123 (arxiv.org), Synopsys

Key Driver: The Need for More Compute

The Connectivity Wall







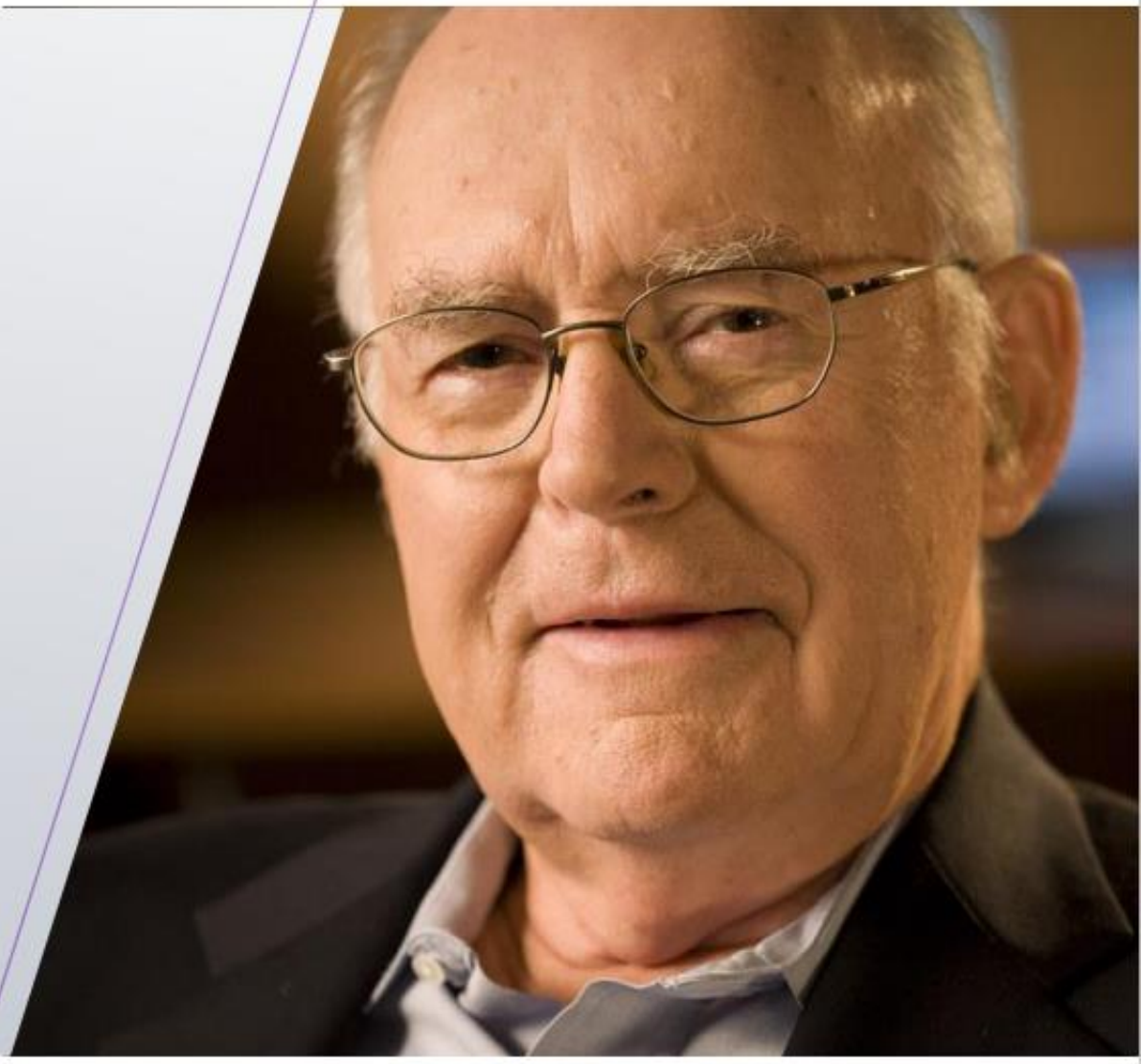
The “Connectivity Wall”

Source: AI and Memory Wall: 2403.14123 (arxiv.org), Synopsys

“Day of Reckoning”

Predicted by Gordon Moore

“It may prove to be more economical to build large systems out of smaller functions, which are separately packaged and interconnected.”



Courtesy Intel
Source: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf

Multi-Die is Enabling Truly Transformative Products

Some Commercial Products

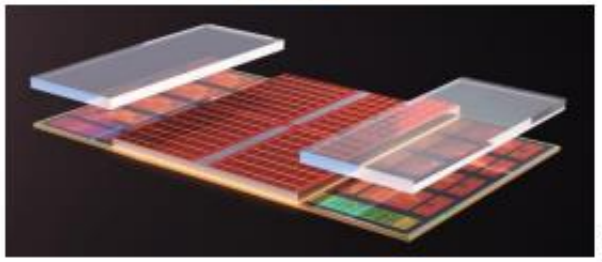
Disaggregate

Split

Aggregate

Scale-Up

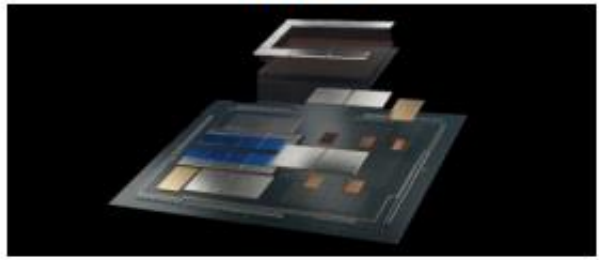
AMD



3D V-Cache: Hybrid Bonded

3x Energy Efficiency, 15x Interconnect Density (v. µbumps)

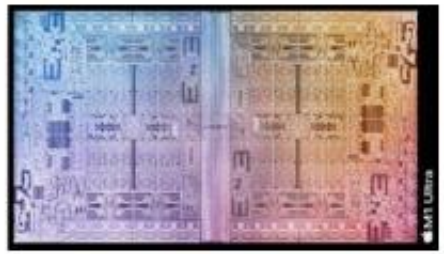
Intel



Exascale Computing / AI

100B+ Tr's, 47 Active Tiles, 5 Process Nodes, EMIB/Foveros

Apple



Personal Computing

2x Dies, 114B Transistors, 2.5TB/s D2D BW, Silicon Connected

Nvidia

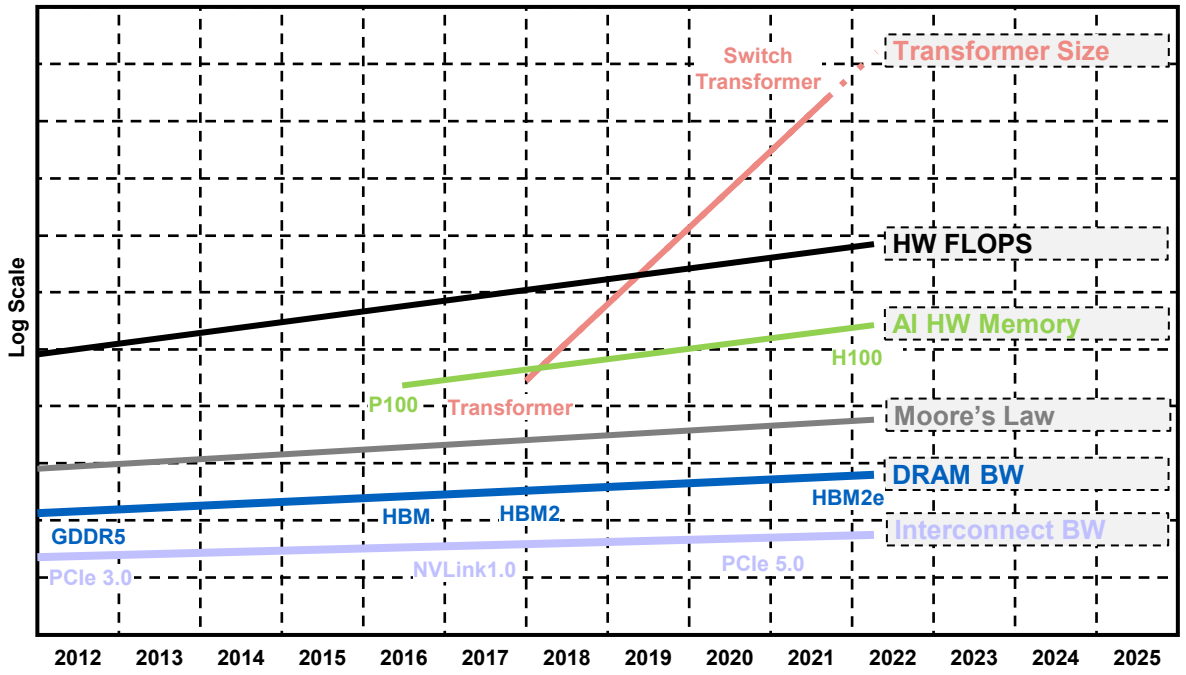
Superchip: Gen AI Computing

2x Dies, 200B+ Transistors, 3.2TB/s Fabric

Verification for the Era of Pervasive Intelligence

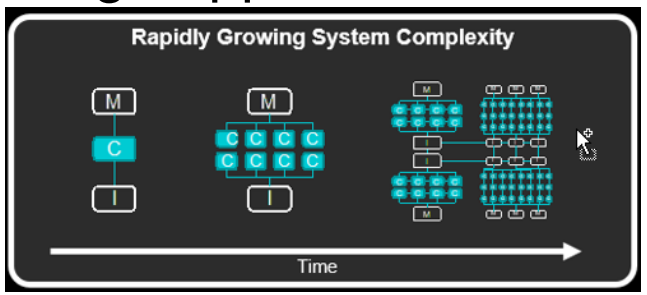
AI Drives New Design Paradigms

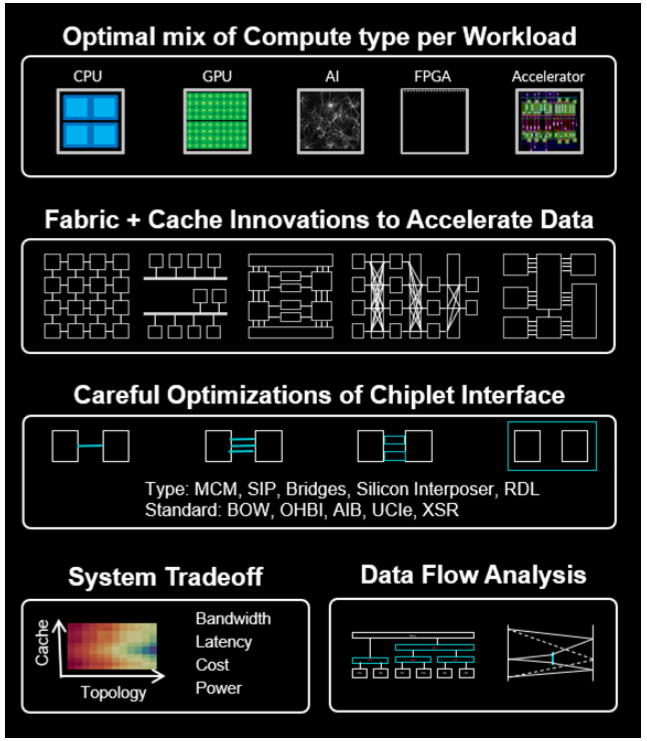
The need for hardware scaling leads to new architectures and design approaches



Transformer Size: 410x / 2yrs
HW FLOPS: 3.0x / 2 yrs, 60000x / 20 yrs
AI HW Memory: 2x /2 yrs
Moore's Law: 2x / 2 yrs
DRAM Bandwidth: 1.6x / 2 yrs
Interconnect Bandwidth: 1.4x / 2 yrs

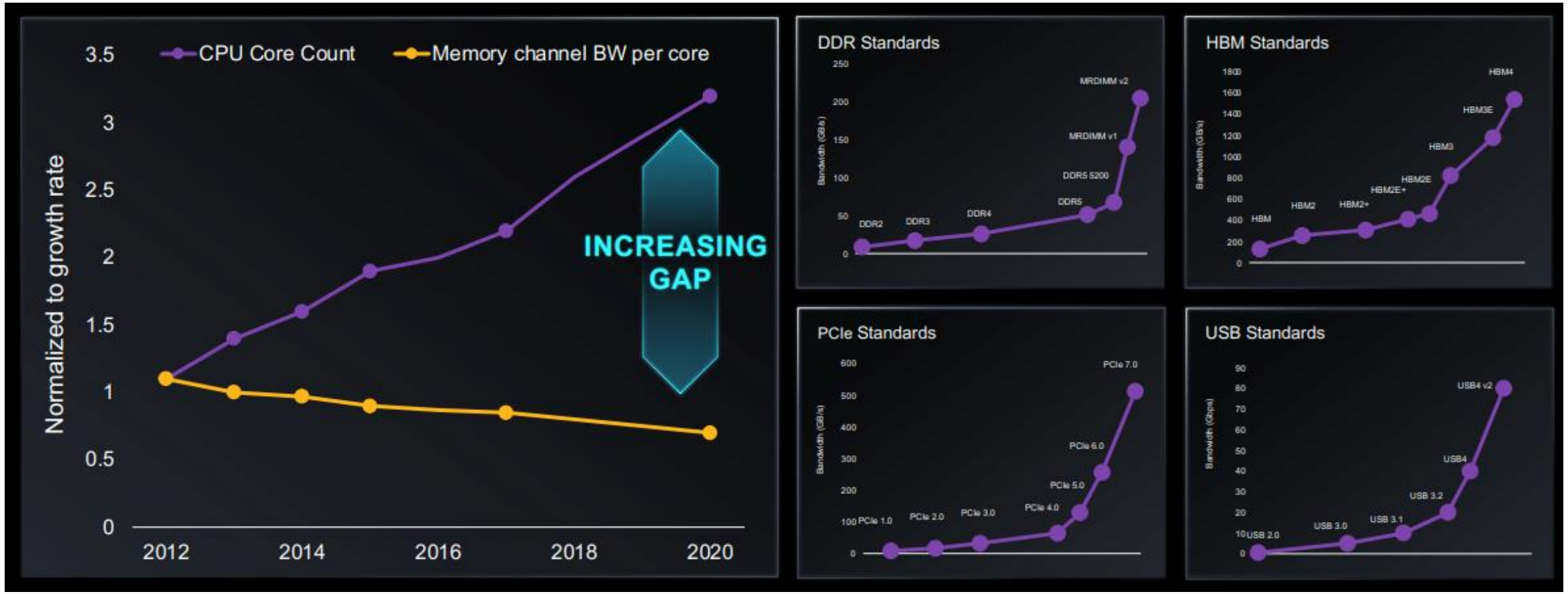
Wide range of optimizations & innovations





Source: AI and Memory Wall: 2403.14123 (arxiv.org), Baya Systems @ EETimes Chiplet Summit

Memory and Protocol Standards Change Faster Than Ever

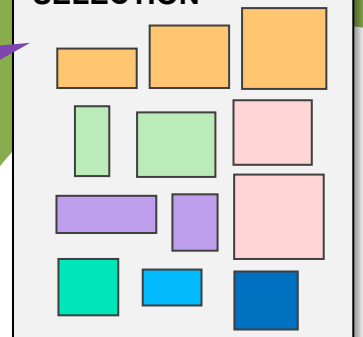


Source: Synopsys Investor Day

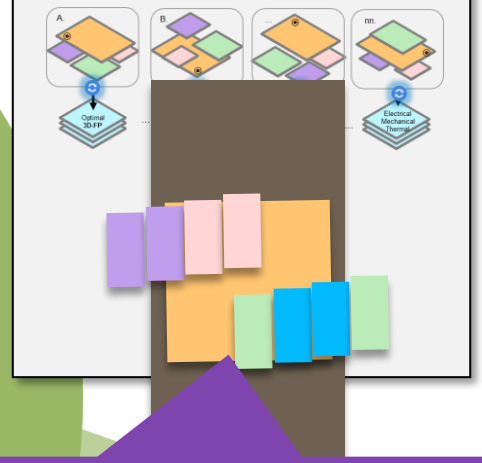
Analyze Impact of Multi-Die Architecture on System KPIs

- Translate functional requirements into physical constraints, e.g. technology, area, aspect ratios

PARTITIONING AND TECHNOLOGY SELECTION

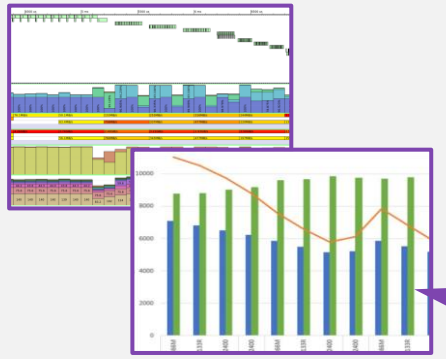


MULTI-DIE SYSTEM PROTOTYPING



- Partition system into chiplets
- Mapping of workload and data structures determines traffic across chiplet boundaries
- Larger on-chip memories and caches help to reduce die-to-die traffic

Performance and Power Analysis



- Analyze resulting impact of multi-die system partitioning on performance and power KPIs

WORKLOAD DEFINITION

Application Requirements:
End-to-end latency, throughput, …

Functional Blocks

Sensing	Comms
Inference	Apps
OS	…

ARCHITECTURE DEFINITION

HW Requirements:
Performance, power

HW Components

Chiplet	IP
SRAM	NoC
HBM	…

- Capture component-level and system-level application and architecture requirements

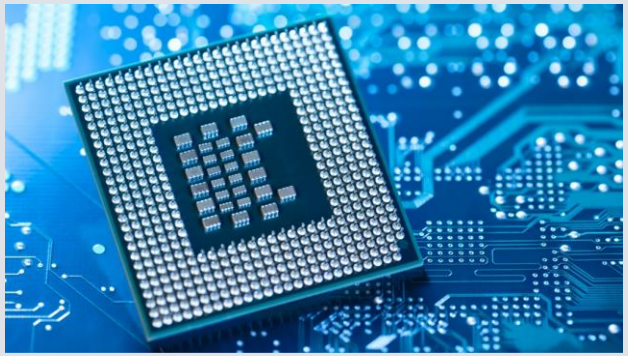
Market Needs Driving Verification Architecture Evolution

Growing SW Complexity

Verification Platforms

Performance	Scalability
Total cost of ownership	Flexibility

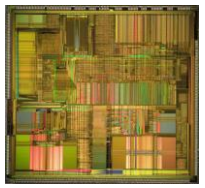
Growing HW Complexity



The Era of Pervasive Intelligence Demands Innovation

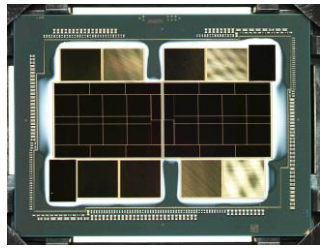
Verification & Validation Dramatically Changed As We Moved to System-Level Complexity

1994 • Pentium P54C
3 million transistors



5 orders of magnitude

2021 • Ponte Vecchio
100 billion transistors



Petacycles (10^{15})

Teracycles (10^{12})

Gigacycles (10^{9})

Megacycles (10^{6})

Kilocycles (10^{3})

Test Requirement

Workload performance Silicon bug escapes

Test Generation Productivity

Test Generation Productivity Gap

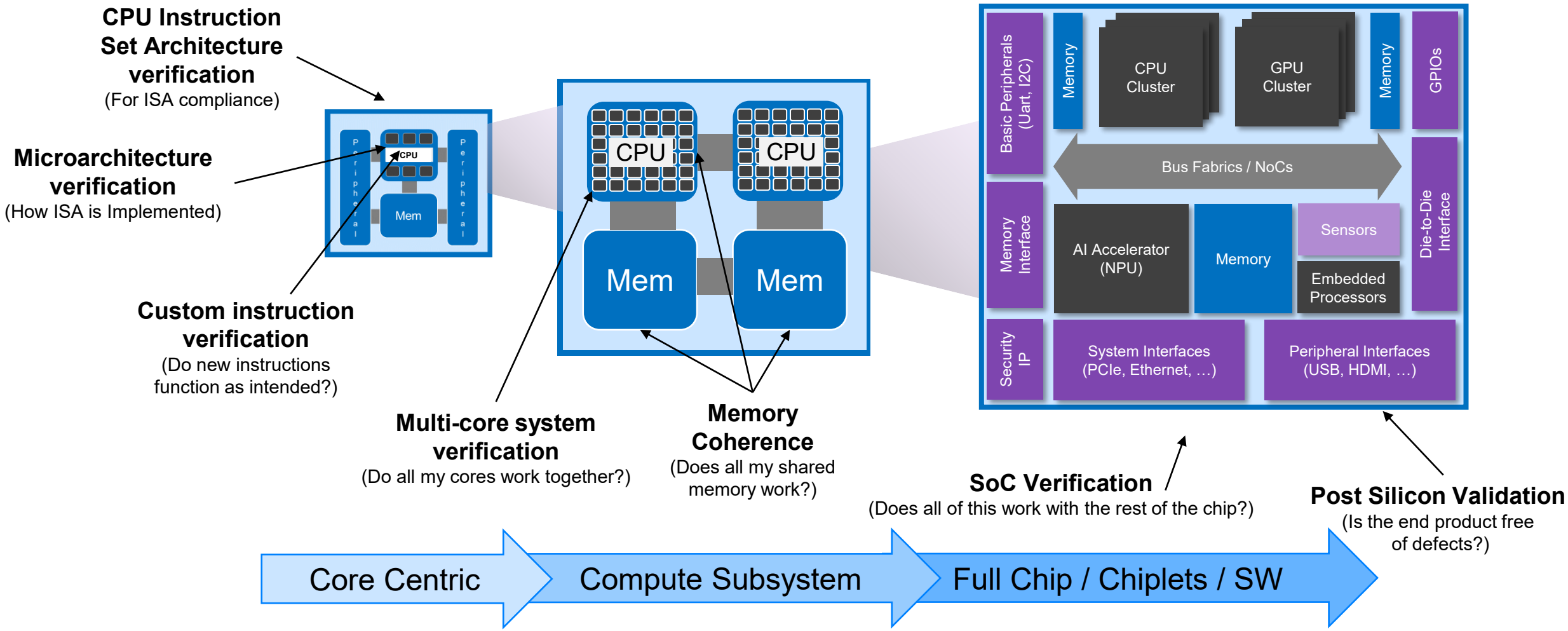
Test Generation Productivity Gap

Verification HW Market

1980s - 1990s	1990s - 2000s	2000s-2010s	2010s-2020s	2020s-2030s
Manual	**Innovation**	**Standardization**	**Manual**	**Innovation Needed**
Ad-hoc, HW directed test	*Constrained-random, assertions, coverage*	*SystemVerilog/UVM*	*Ad-hoc, SW directed test*	*System-level, Software-driven*

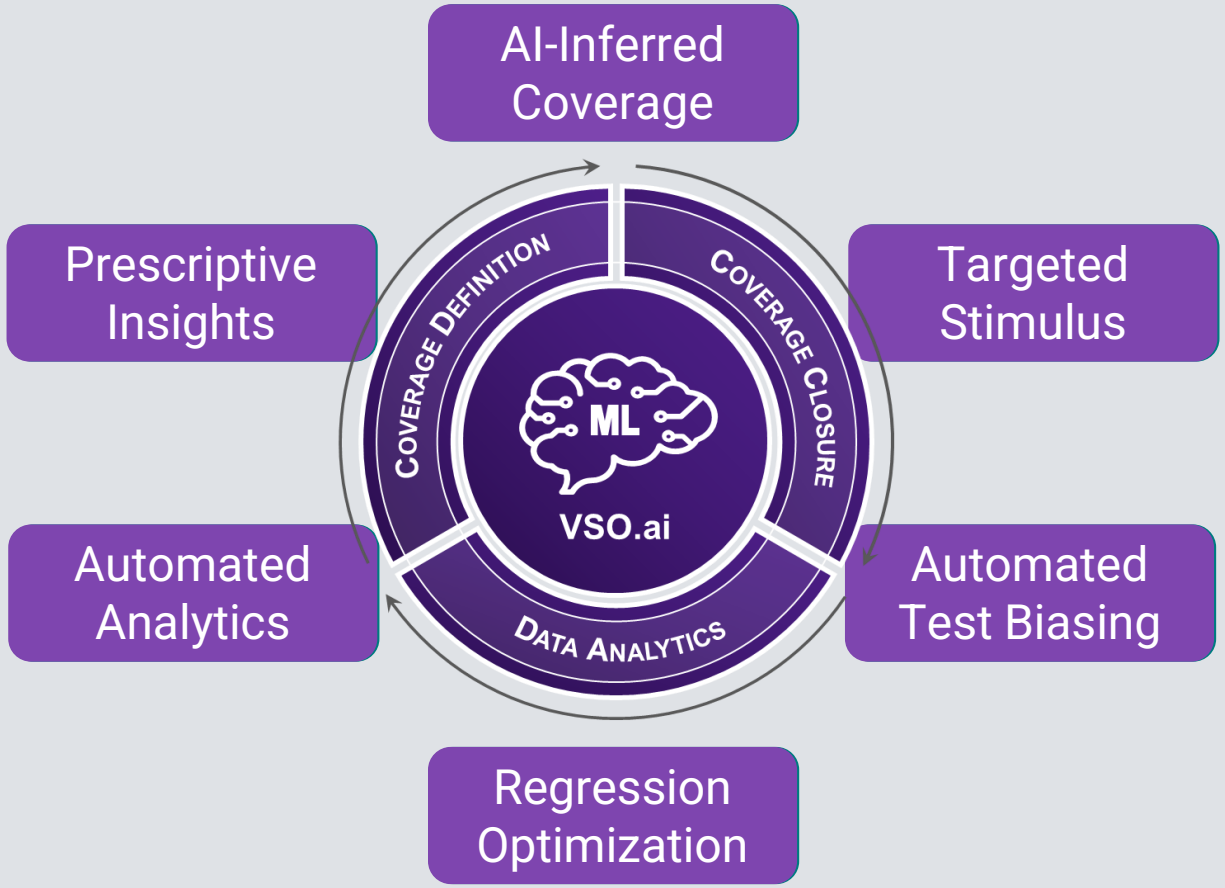
Test Generation Challenges from IP to Systems

From ISA through CPU, arrays of CPUs and Systems on Chips / of Chiplets



AI Driven Verification Space Optimization

Faster, Higher Coverage Closure & Analytics



Better Coverage: *Coverage Inference Engine Helps to Define Coverage from Stimulus and RTL*

Productivity Boost: *Connectivity Engines & ML-Based Solver Target Hard to Hit Coverage*

Improved HW Utilization: *Regression Optimizer Ensures Highest ROI Tests Run First*

Higher Resource Efficiency: *Advanced Root Cause Analysis to Identify Unreachable Coverage*

Recap

- AI Driving New Exciting Opportunities and User Experiences
- Increasing SW and Silicon content driving Pervasive Intelligence
- AI Requirements driving new design paradigms
- Testing Methods and Verification Architecture Need to evolve
- Type of SW Workloads defining new use-cases

SYNOPSYS®

Thank You