Quo Vadis IC System Design?

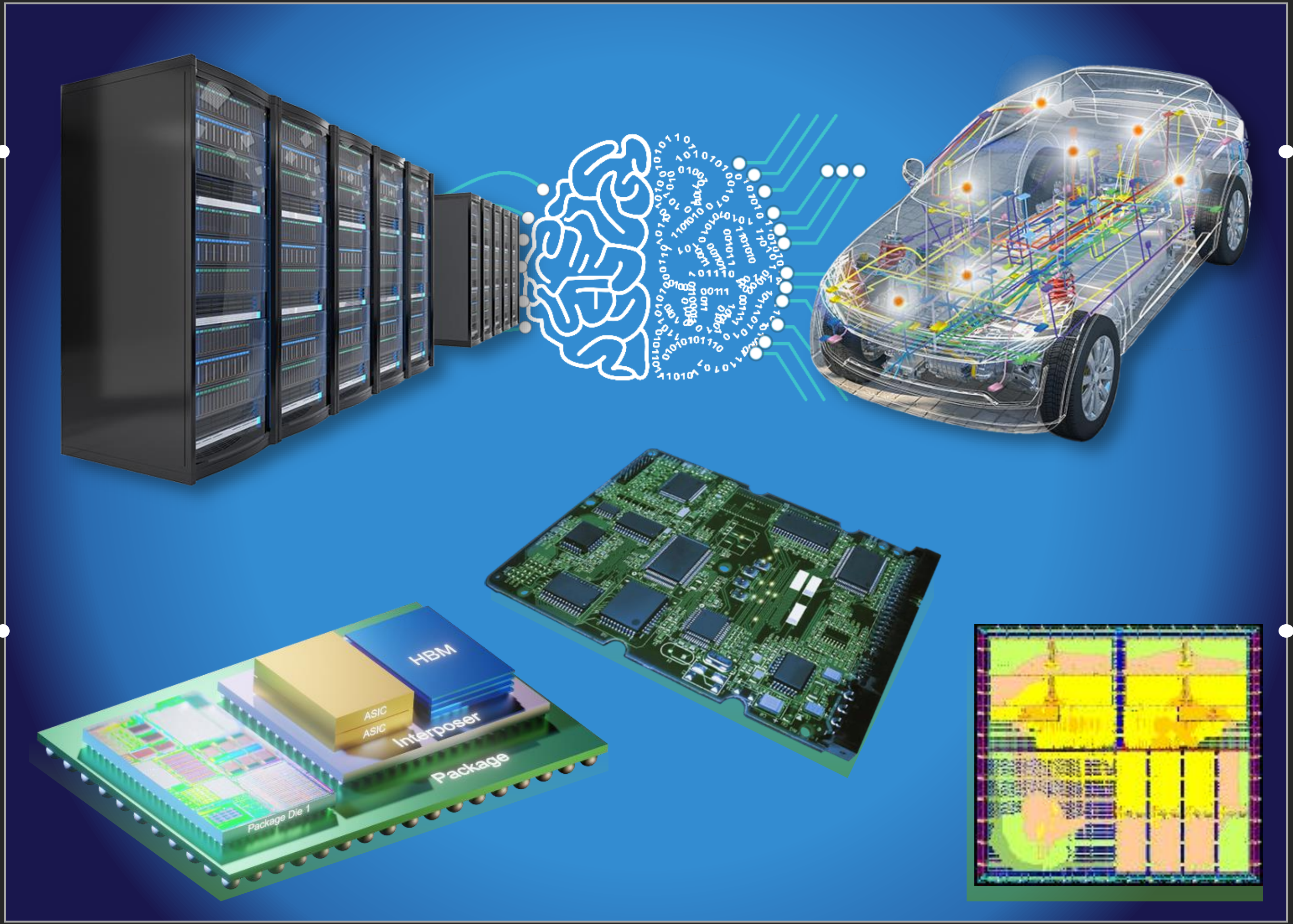
Alberto Sangiovanni-Vincentelli
University of California, Berkeley

SHINE Centre partnership launch & inaugural technical workshop
1 March 2023

Intelligent System Design: Exponential Complexity

Electronics and Software Design

3D-IC System-in-Package Design



Multiphysics Design

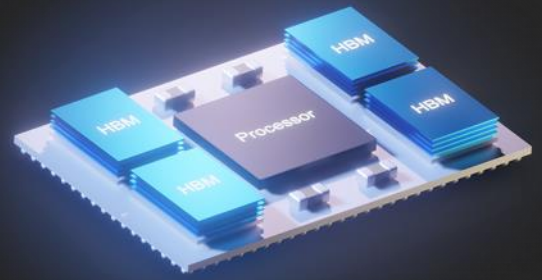
Advanced-Node Chip Design

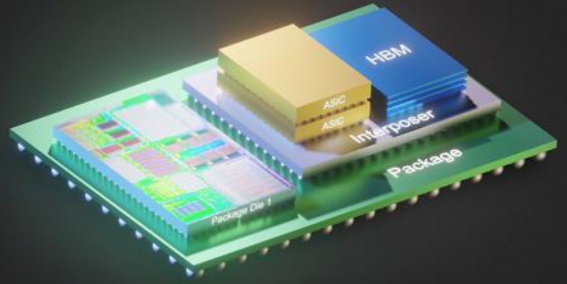
"Our ability to integrate has outpaced our ability to comprehend"

Matt Graham, Cadence

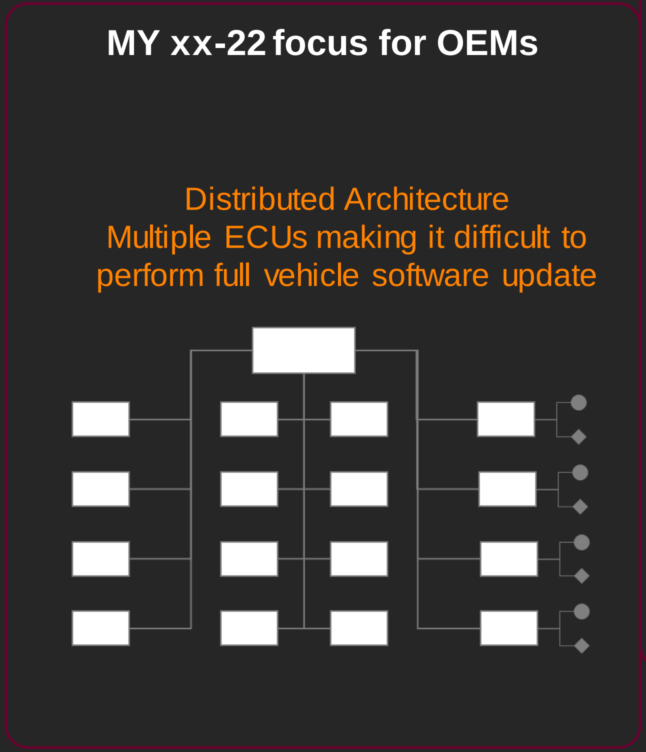
Agenda

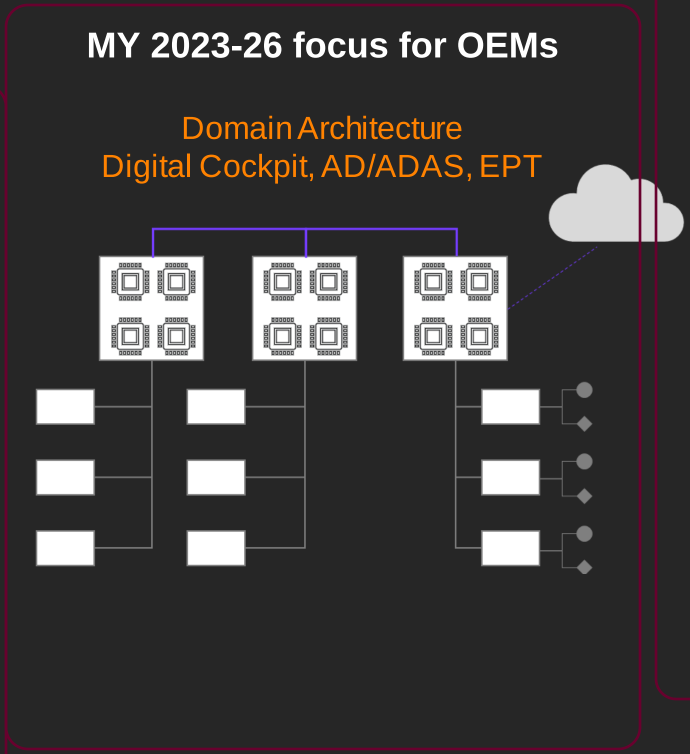
- Single chip or multiple chiplets?
- 3D Packaging and 3D Integration
- Design flow

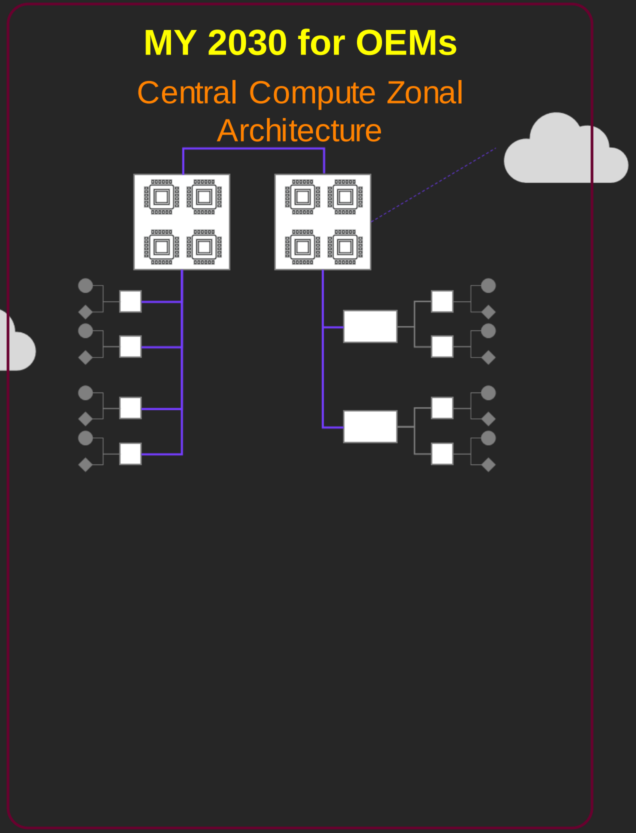




System Architecture Revolution in Automotive: Tesla Single Chip



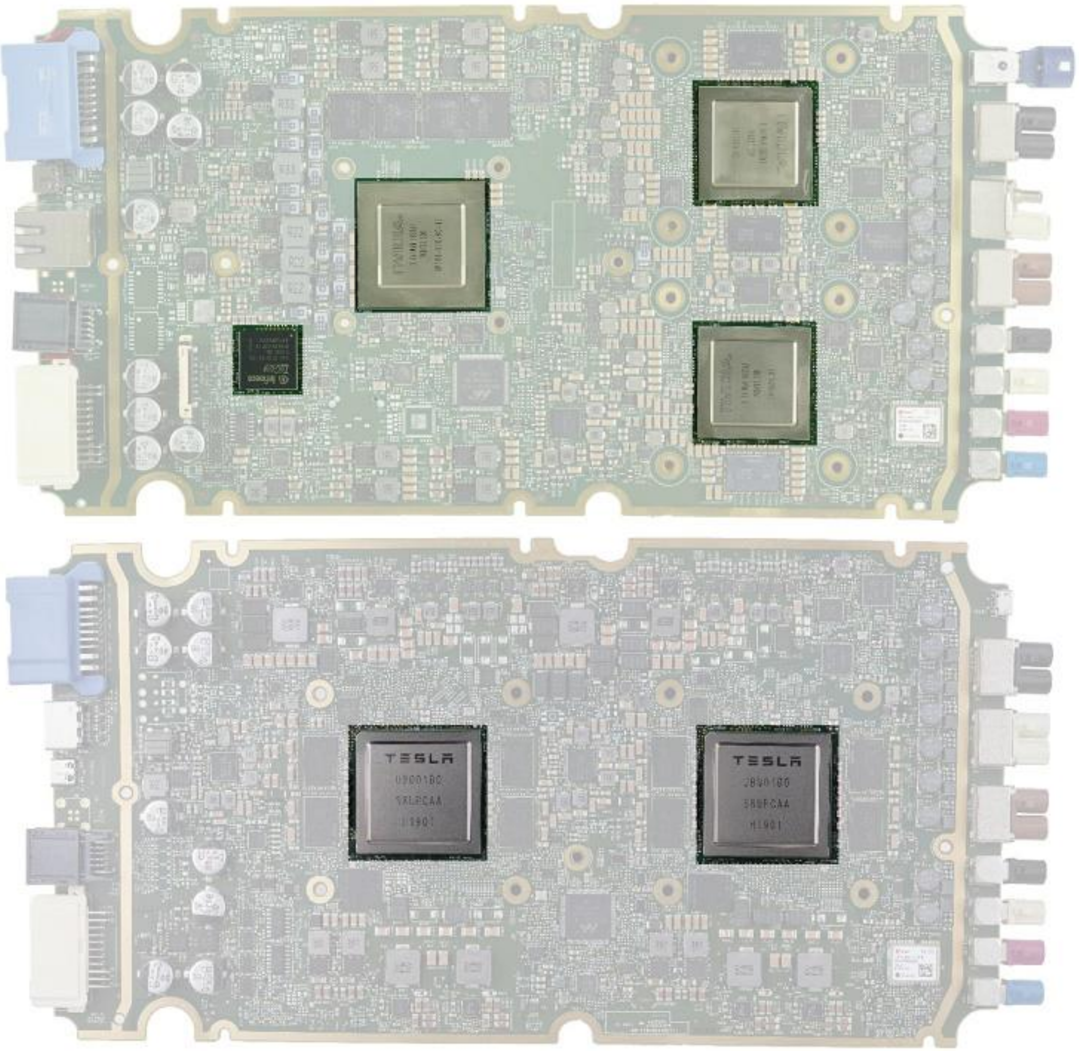


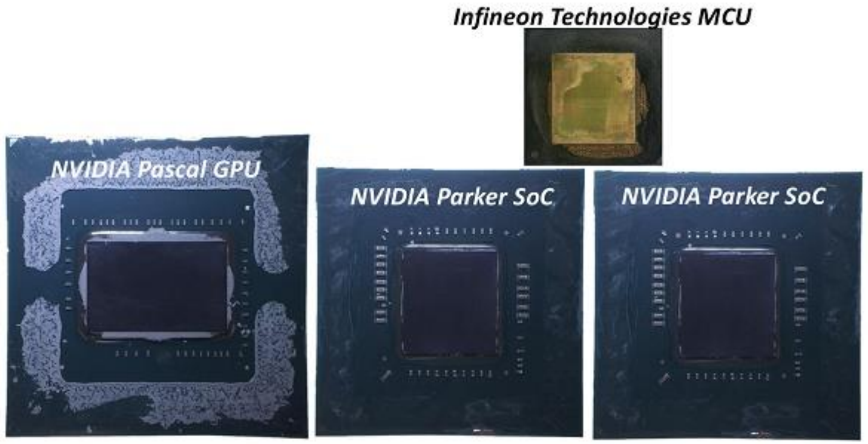


Courtesy: Anup Gupta, KPIT

Tesla Model 3 - Autopilot processors evolution

(Source: Automotive Teardown Track, System Plus Consulting, 2020)





HW3.0 vs HW2.5

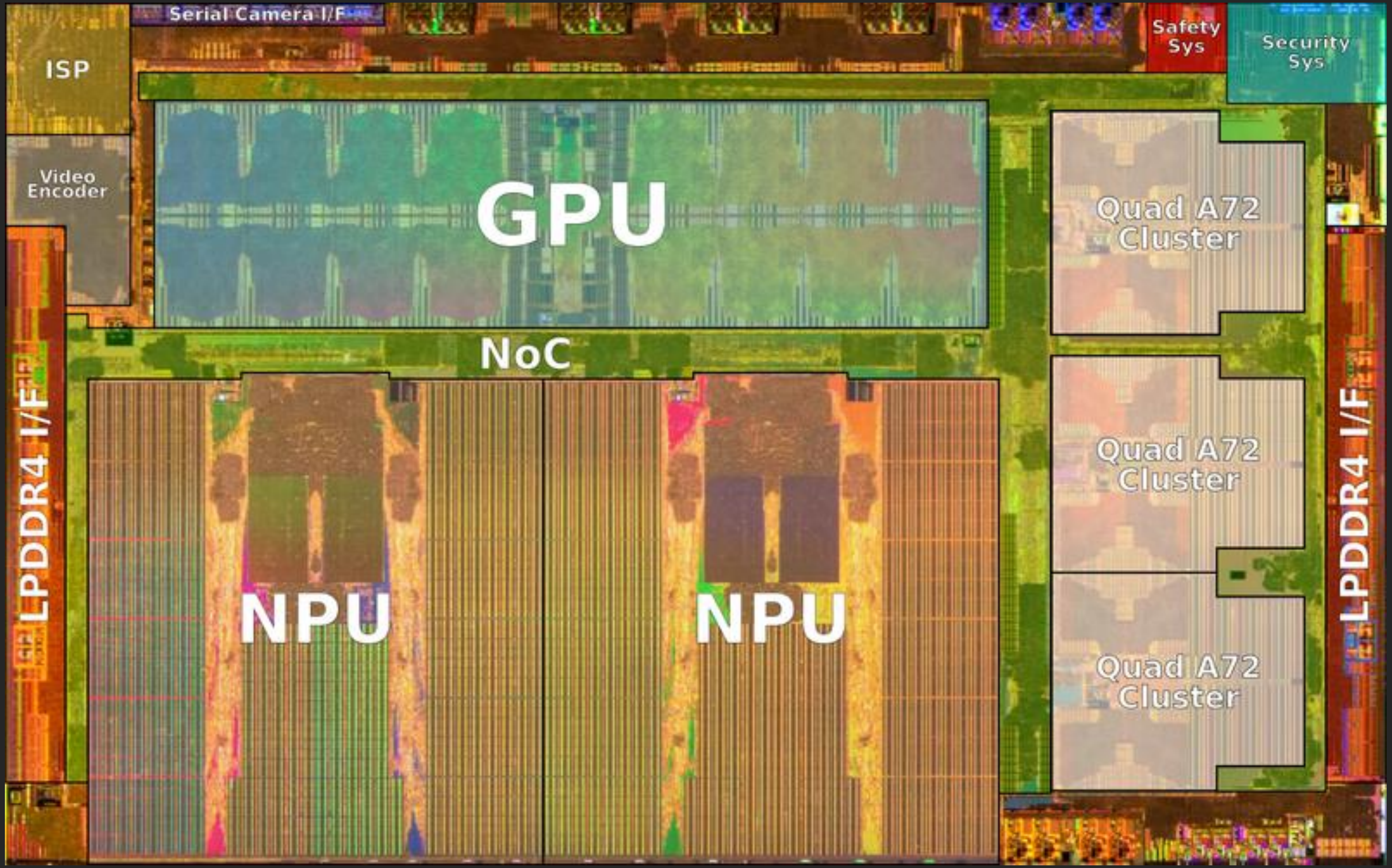
Processors count: 2 (Tesla) vs 4 (NVIDIA, Infineon)

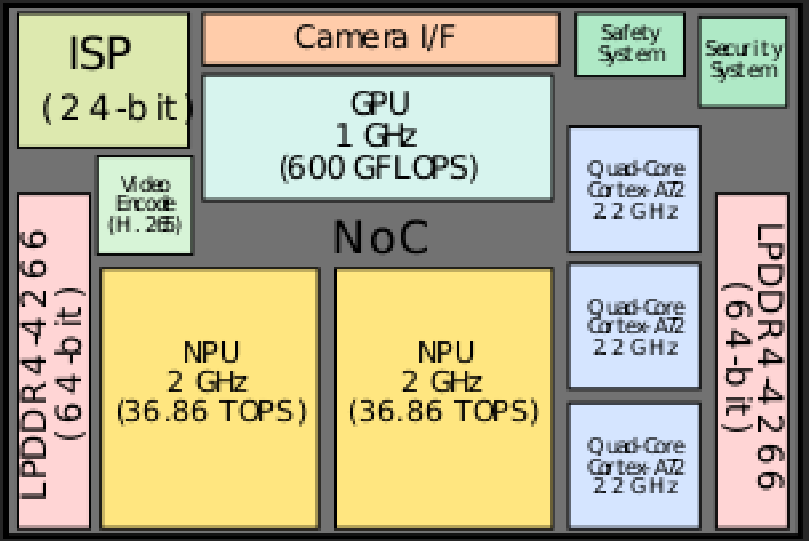
Processors footprint: 4,180mm² vs 2,812mm²

Processors dies area: 611mm² vs 512mm²

Technology Node: 14nm vs 16nm (excepted for Infineon Technologies MCU)

TESLA FSD SoC

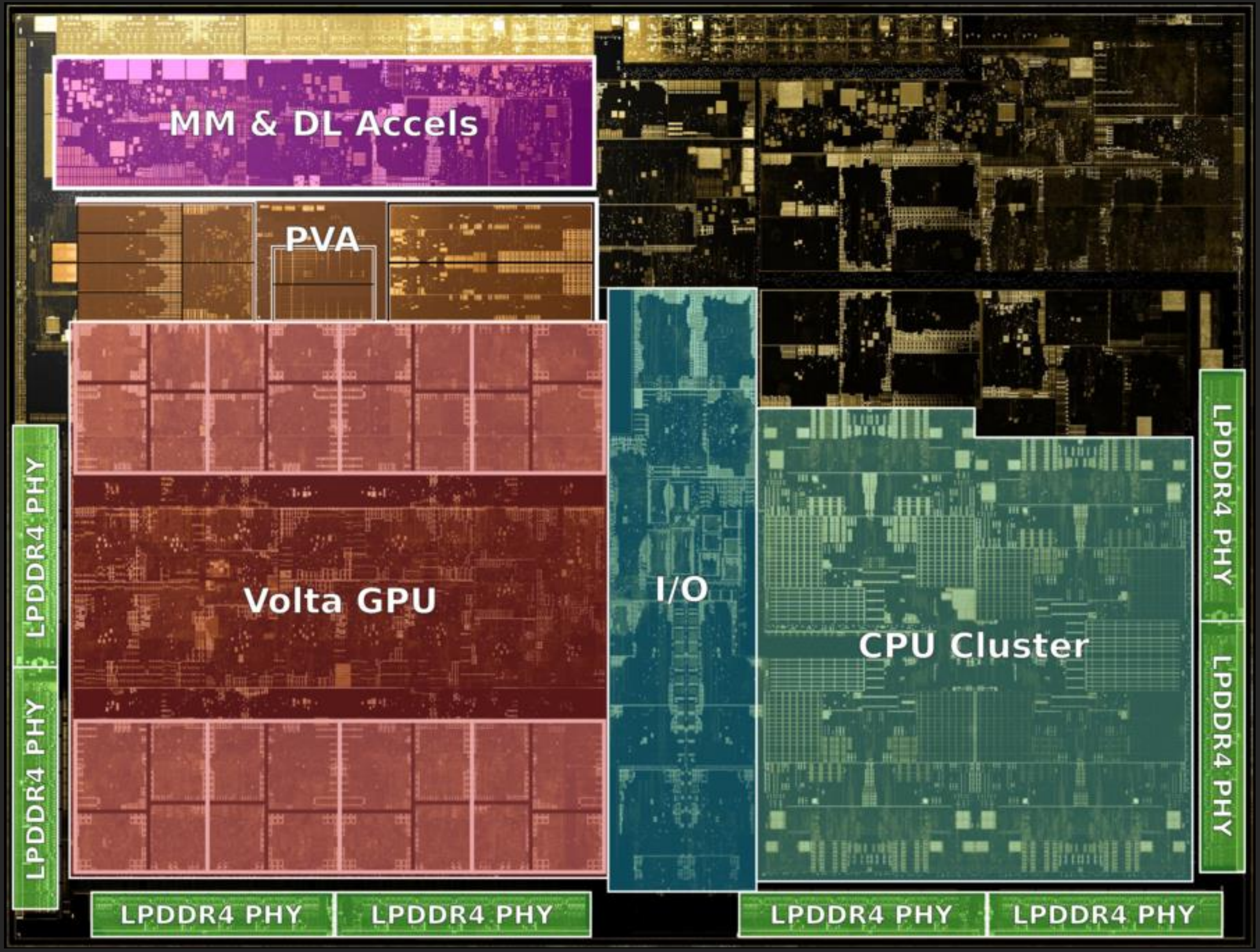


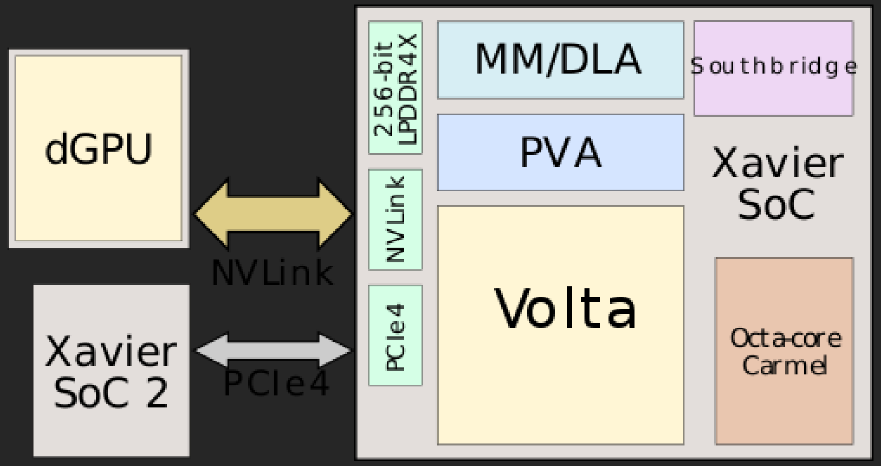


- NoC – Network on chip
- ISP – Image Signal processing
- Safety Sys – Lock step for ISO26262
- Security – only TESLA certified software

Chip focused on Automotive L5 use case for Deep learning

Another Massive Single Chip: NVIDIA Xavier SoC



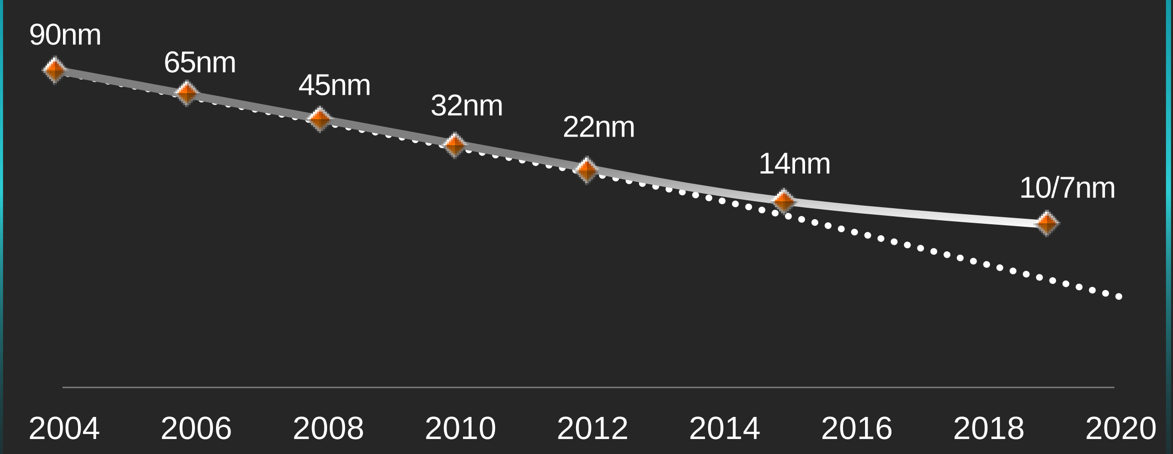


PVA - Programmable Vision Accelerator
MM – Multi media (Stereo and optical flow) engine
DLA – Deep learning accelerator

Chip focused on gaming (graphics and high quality audio) more than Automotive L5 use case for Deep learning

However, Moore's Law Slowing and Die Size Reaching Lithography's Reticle Limit

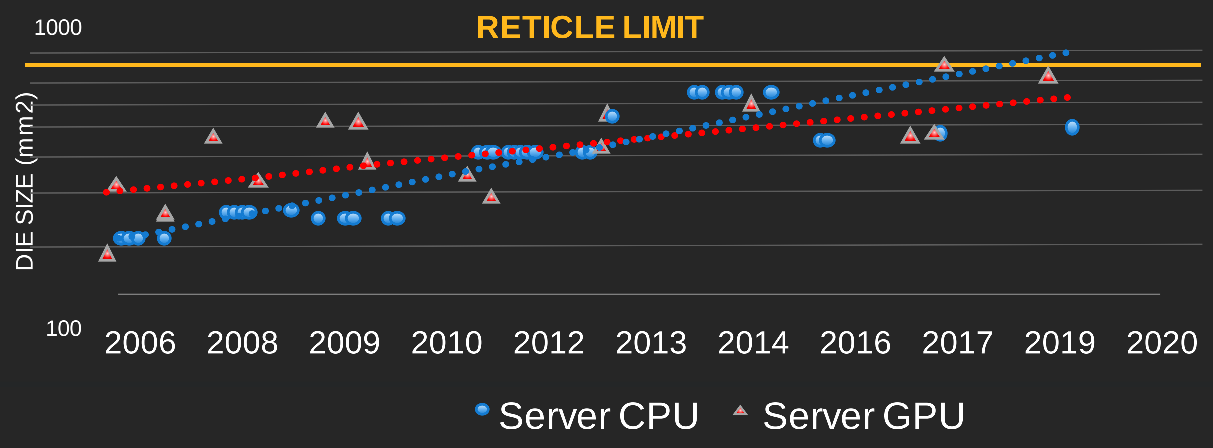
Moore's Law Slowdown



Advanced Nodes Reaching Physical Transistor Size Limit

Die Size Trend

Die Size Increases Over Time in Server CPUs and GPUs



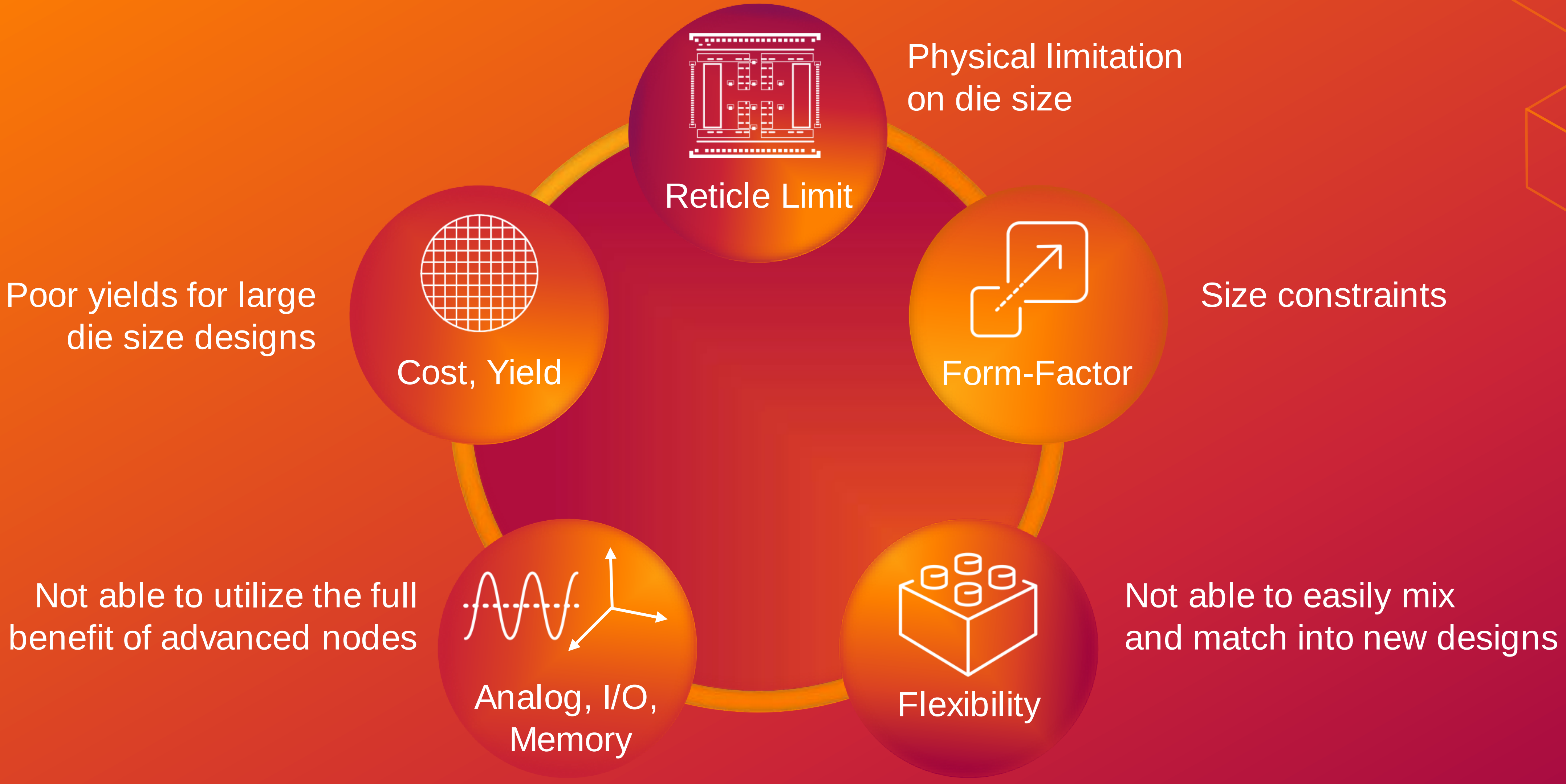
Die Sizes Increasing at an Unsustainable Rate

Source: Delivering the Future of High-Performance Computing, AMD keynote, Dr. Lisa Su, Hotchips, August 2019

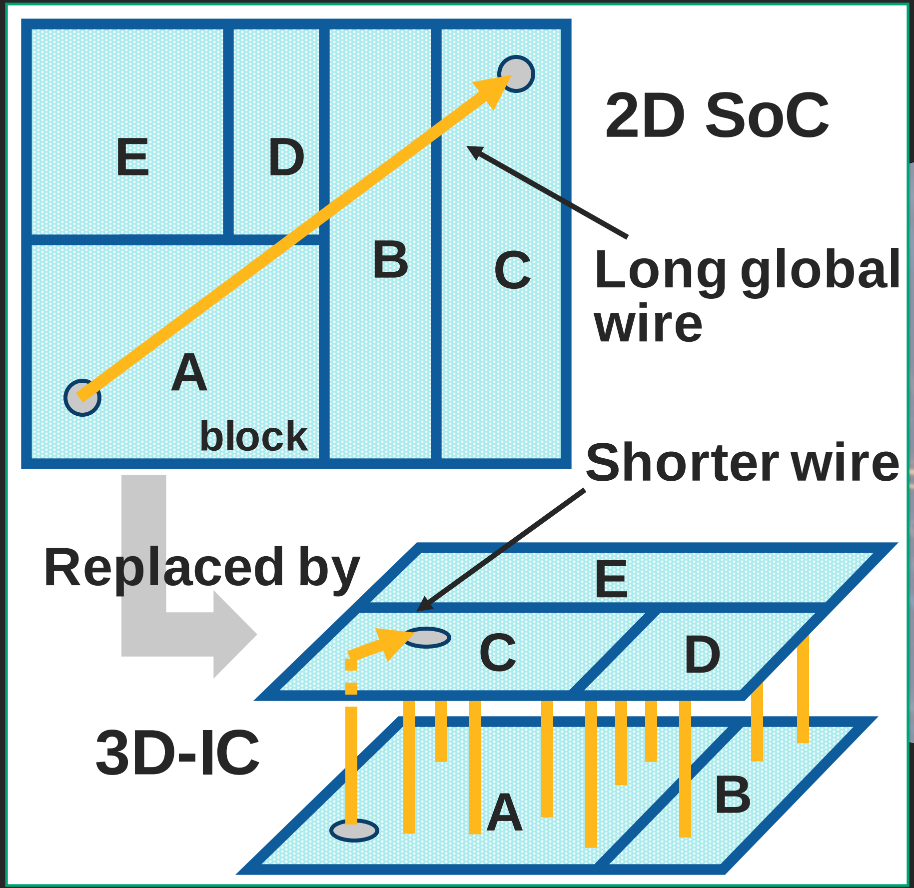
Larger Die Sizes Problematic

Lisa Su, AMD (IEDM 2017)

Monolithic 2D-IC Design Limitations



Silicon stacking (3D-IC)



- Shorter Wire
- Less Power
- Higher Performance
- Higher Bandwidth
- Smaller Profile
- Better Yield

3D-IC: Opportunities and Challenges

100X+

TFLOPS increase over the next decade

20X

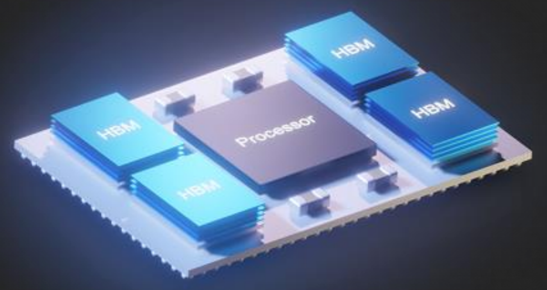
Memory capacity increase

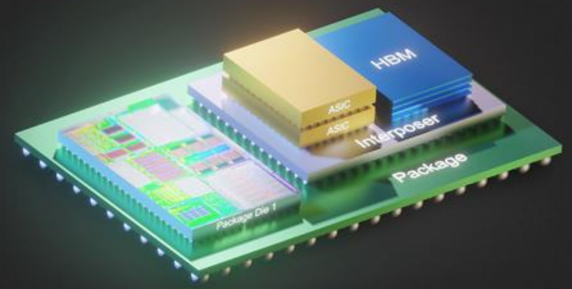
But…Exponential Increase in Design Complexity

Courtesy: A. Devgan, Cadence

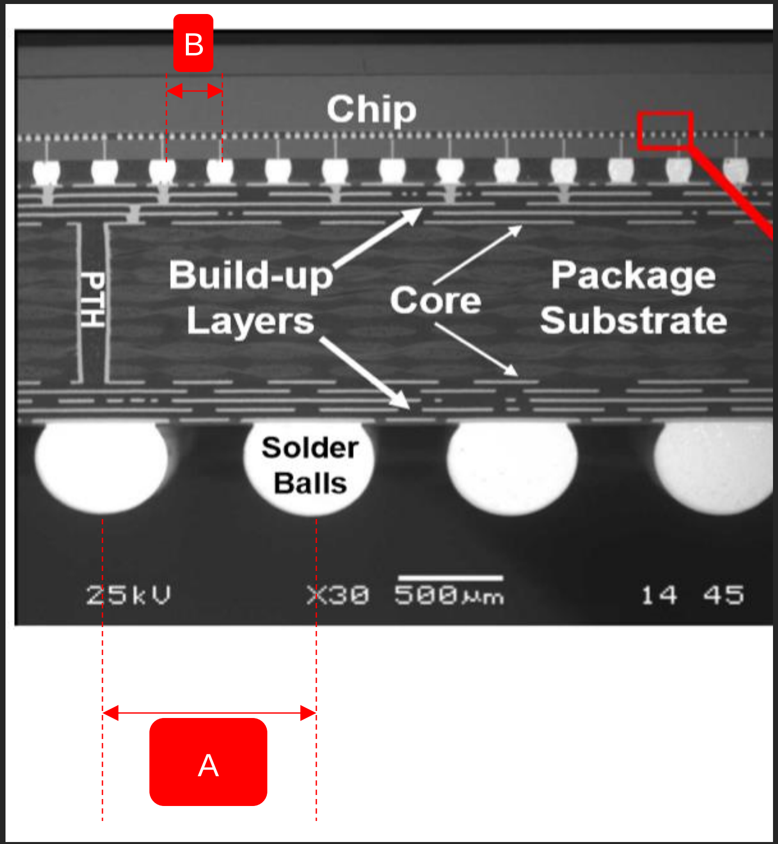
Agenda

- Single chip or multiple chiplets?
- 3D Packaging and 3D Integration
- Design flow





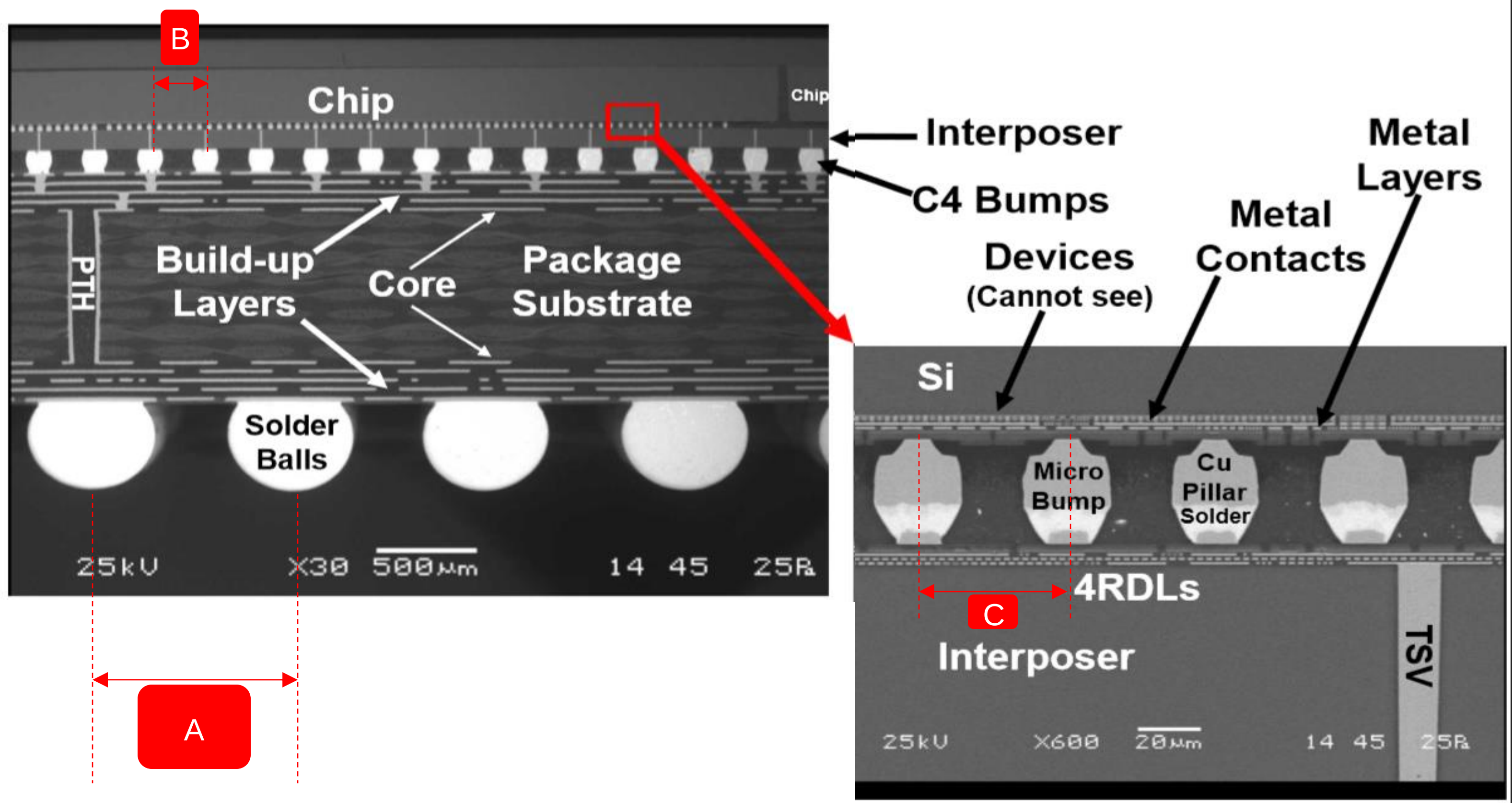
Foundry Trends : Chip-to-Chip Connection advancements – Package Level



Connection Type	Solder Ball	C4 Bump
Typical Pitch of Connection	A >1000um	B 250um
# of Connection (in $1mm^2$ area)	<1	16

资料来源：**Semiconductor Advanced Packaging**, John H. Lau

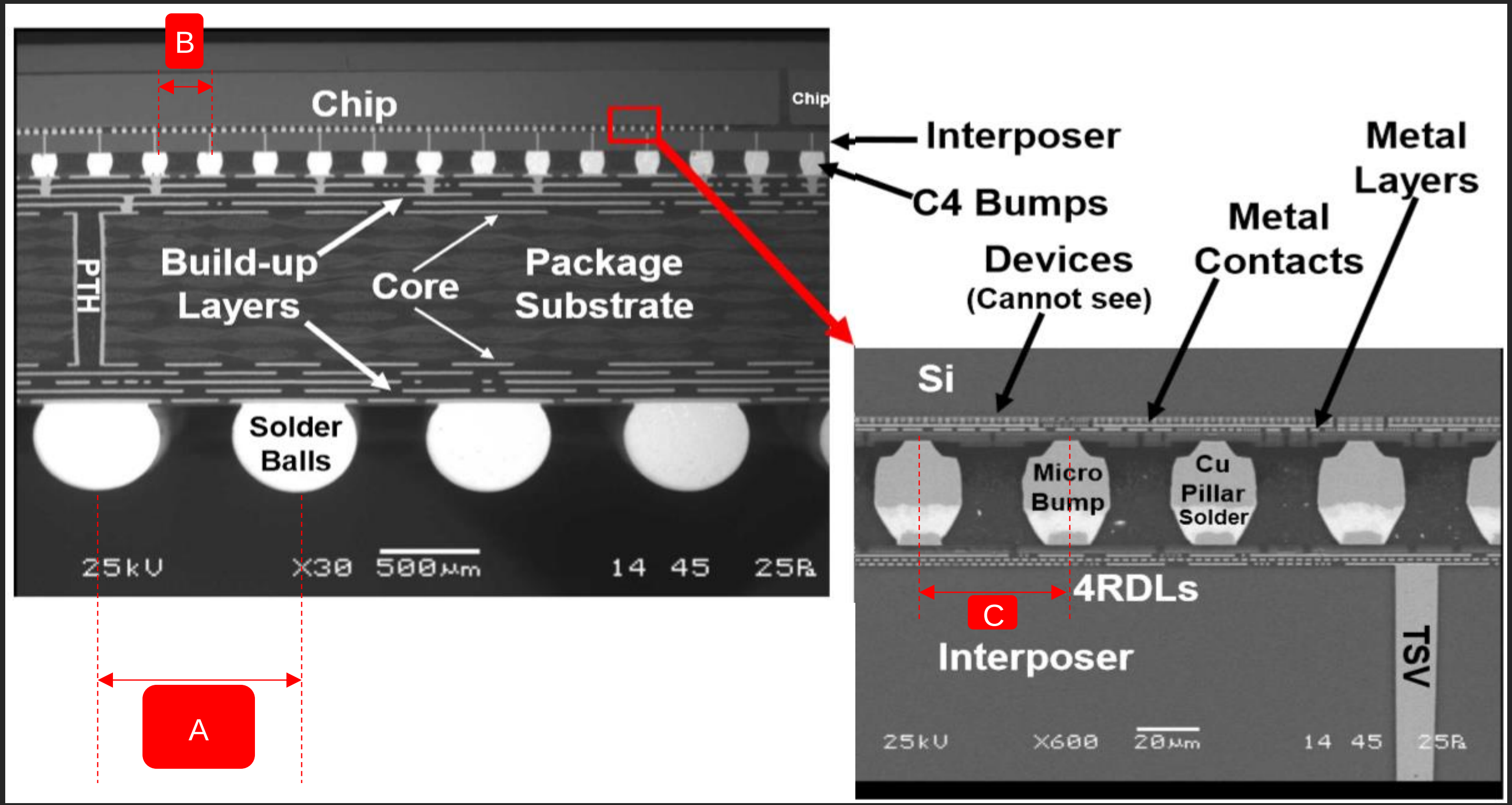
Chip-to-Chip Connection Roadmap – Interposer Level

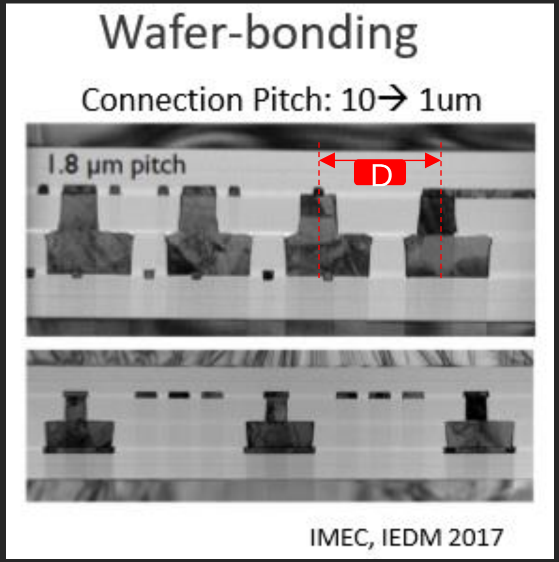


Connection Type	Solder Ball	C4 Bump	Micro Bump
Typical Pitch of Connection	A >1000um	B 250um	C 50um
# of Connection (in $1mm^2$ area)	<1	16	400

资料来源：**Semiconductor Advanced Packaging**, John H. Lau

Chip-to-Chip Connection Roadmap – Wafer Level





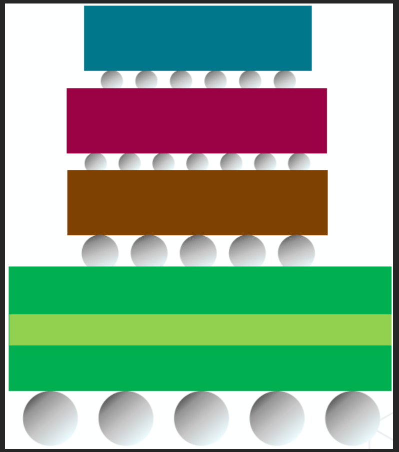
Connection Type	Solder Ball	C4 Bump	Micro Bump	Hybrid Bond
Typical Pitch of Connection	A >1000um	B 250um	C 50um	D <10um
# of Connection (in 1mm^2 area)	<1	16	400	>10000

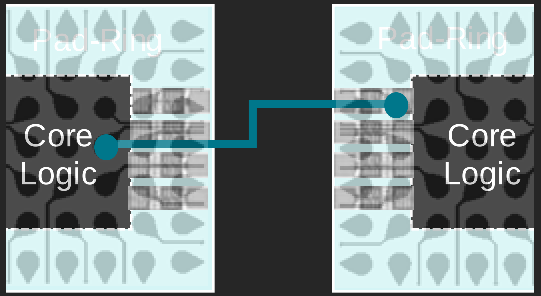
资料来源：**Semiconductor Advanced Packaging**, John H. Lau

3D Packaging Versus Silicon Stacking (3DHI)

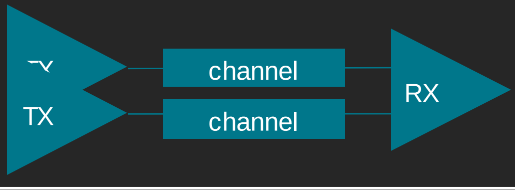
3D Packaging

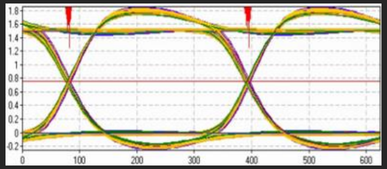
- Solder-based connections (>25um)
- Each die designed independently
  - Black-box abstracts used for layout
- I/O buffer to I/O buffer signaling





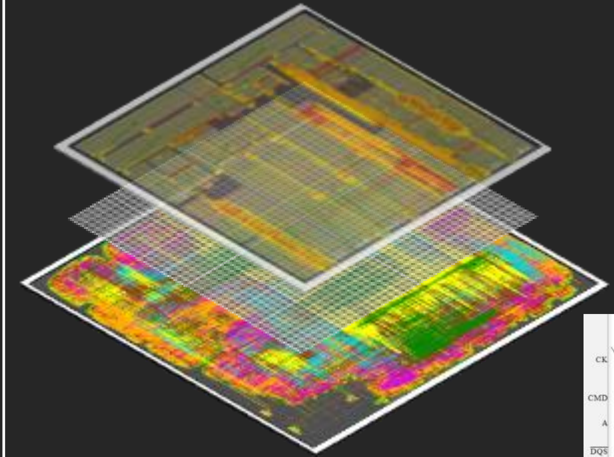
I/O buffer to I/O buffer
(AIB, OpenHBI, BoW, UCIe, etc)

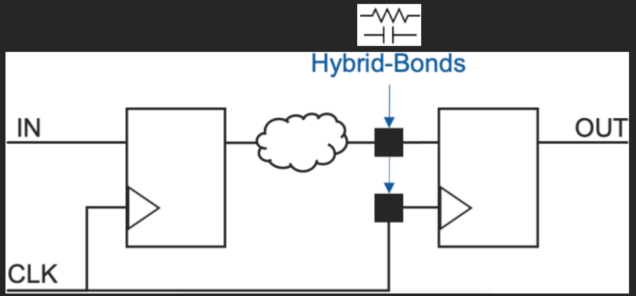




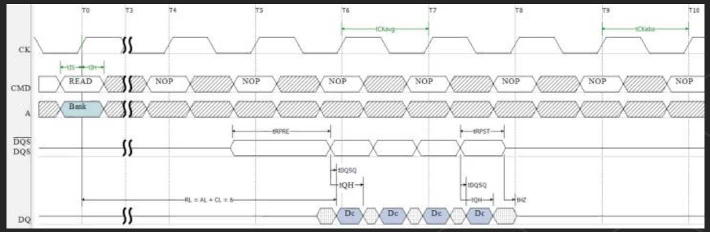
Silicon Stacking

- Solder-free connections (<10um)
- Single RTL partitioned at implementation
  - Full detail of IC required for layout
- DBI, hybrid-bond, cu-to-cu, direct connection

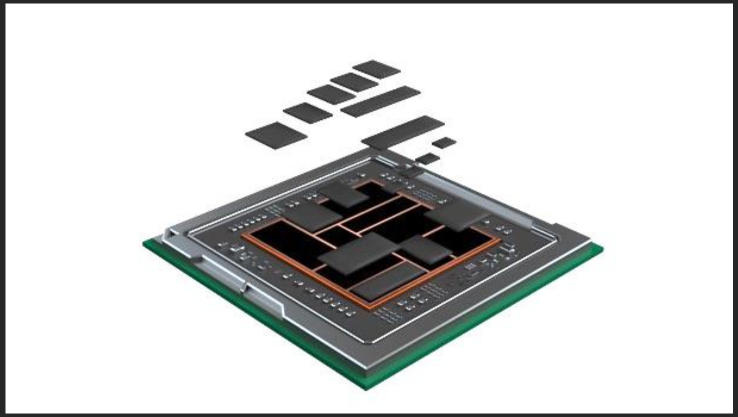


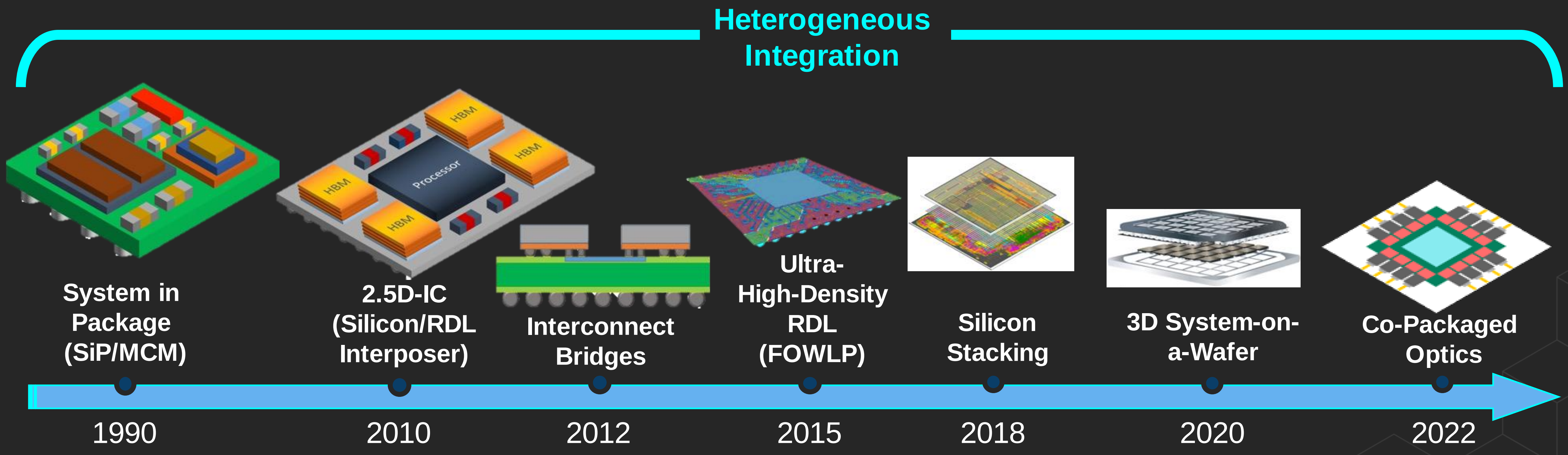


Interconnect through hybrid bond



Heterogenous Integration:
Multiple Packaging Technologies





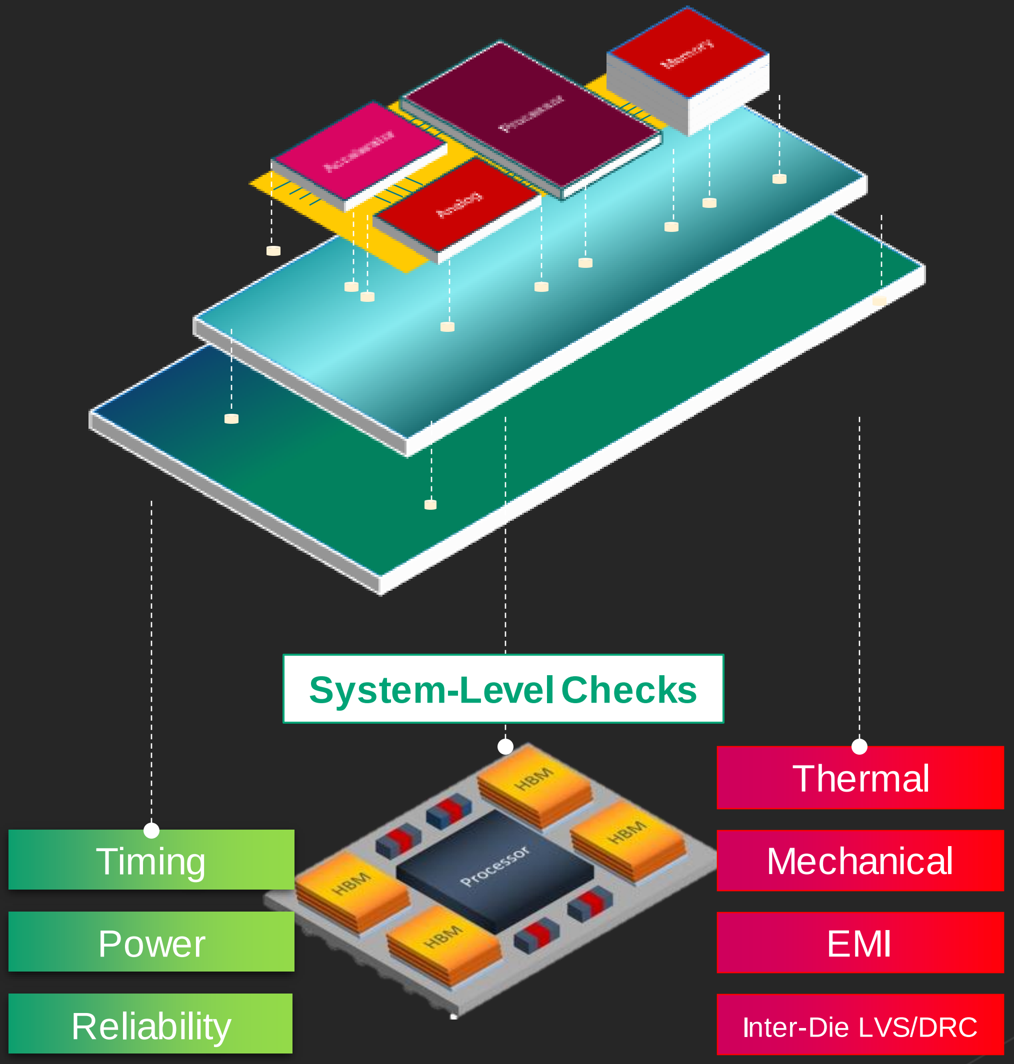
Multi-Chiplet 3D Flow Challenges

- 3D-IC design aggregation and management
  - Die placement and bump planning
  - SoC and packaging teams works in silos
  - No single database to represent multiple technologies
- Additional system-level verification
  - Thermal analysis from across chip(lets) and package
  - 3D STA with explosion of corners for signoff
  - Inter-die connectivity validation at the system level

☒ Current industry solutions: Disjointed, point solution-based

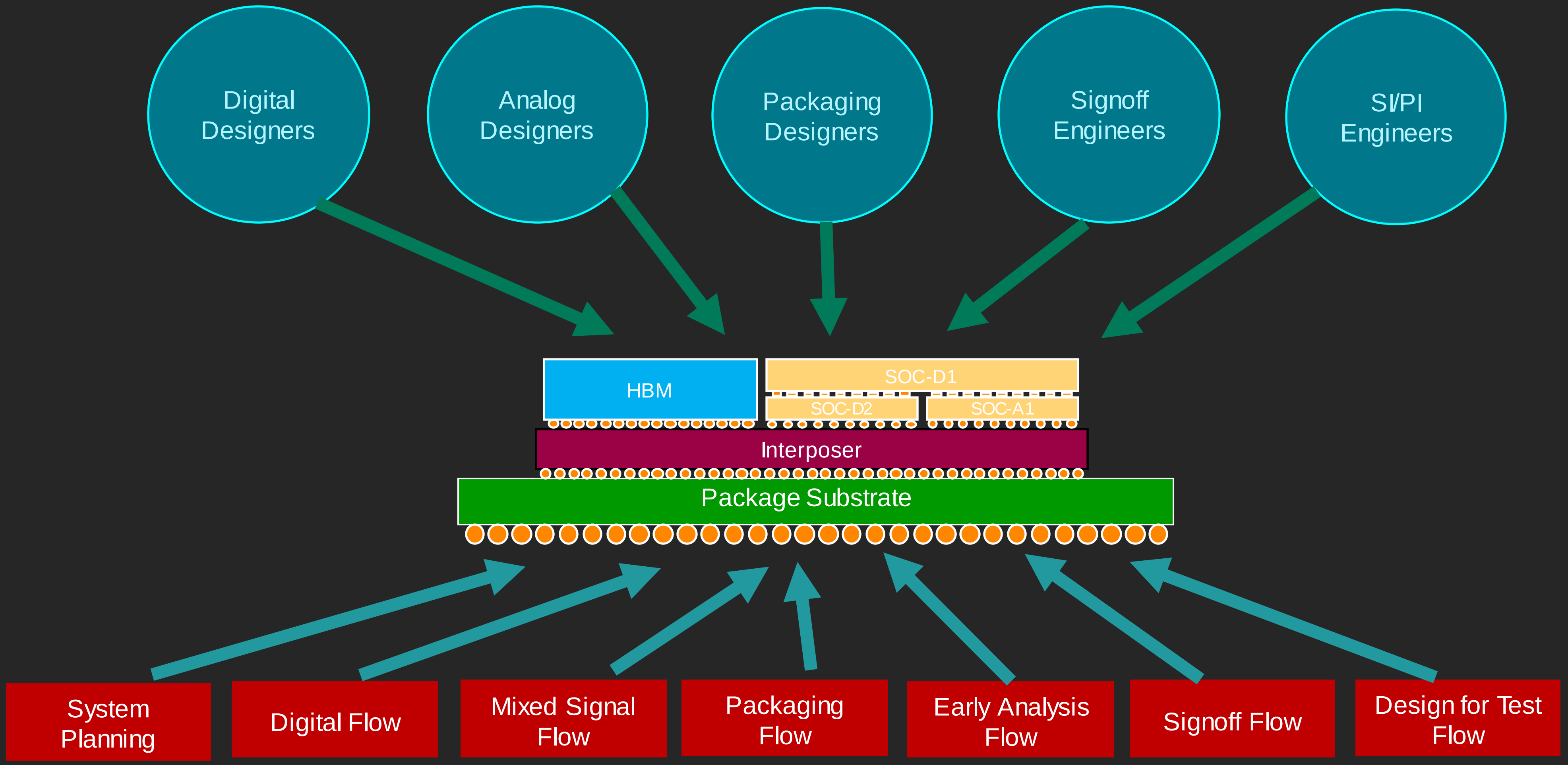
☒ No way to do **exploration/early feedback**

☒ **Causes costly overdesign of individual dies in a stack**



Chiplet 3D-IC Design Flow – Different Requirements based on users

Requires a Platform for full solution yet modular for different teams

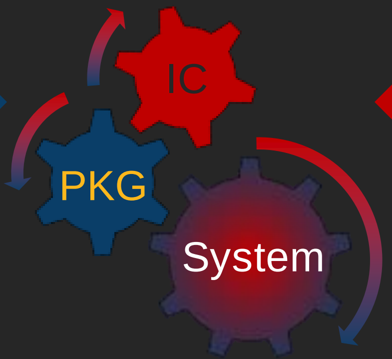


3D-IC Flow is a multi-engineering discipline methodology
Requires tight collaboration across all teams.
The cross-over is between SoC, Chiplet and Packaging Design Methodology

The Needs of IC and Systems Designers are Converging

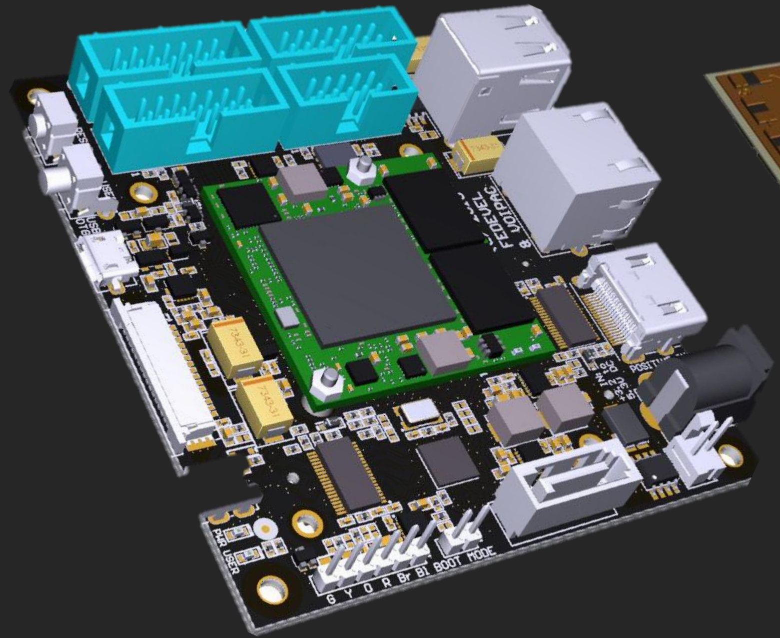
OSATs (SWaP)

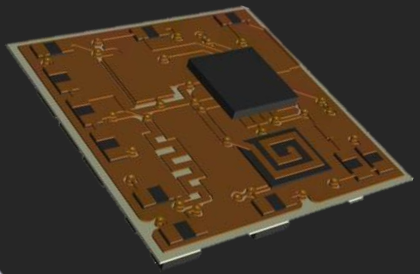
1980-2010

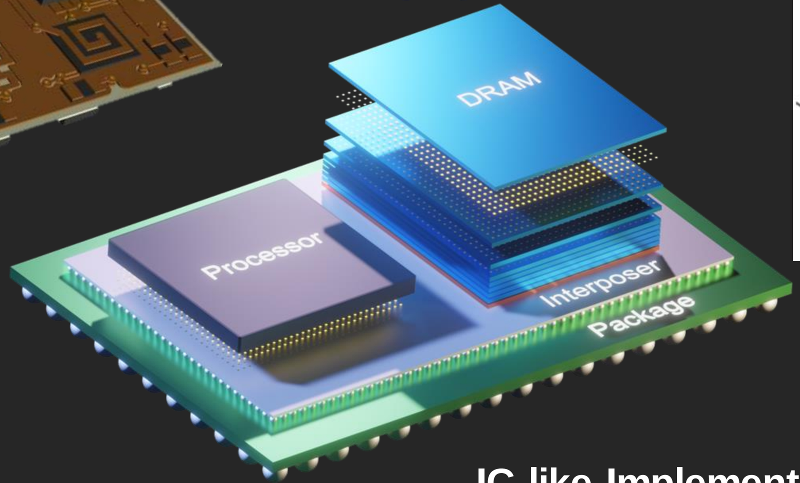


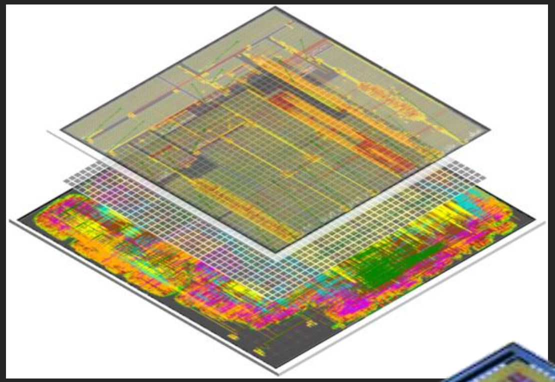
Foundries (PPA)

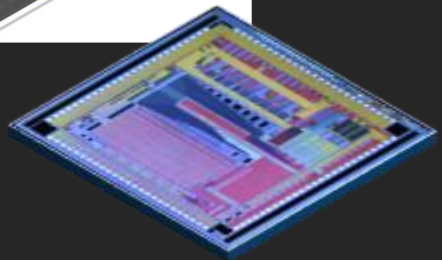
2011-Now











PCB Layout Flow
System-Level Analysis

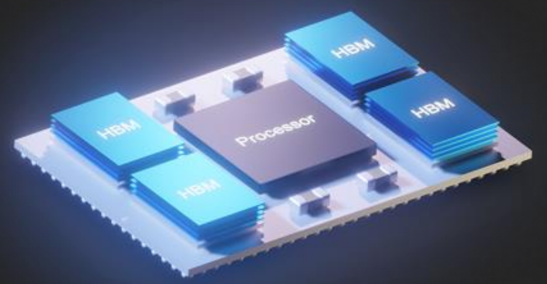
PCB-like Flow
System-Level Analysis

IC-like Implementation Flow
IC signoff Methodology and
System-Level Analysis

IC Flow
IC signoff Methodology

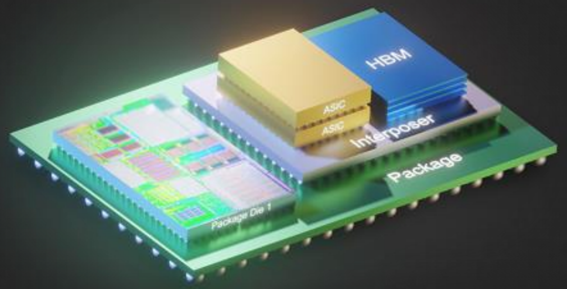
Agenda

- Single chip or multiple chiplets?



- 3D Packaging and 3D Integration

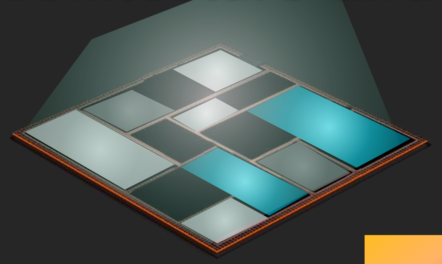
- Design flow with Cadence Integrity Example



Design Flow: 3D-IC Design and Optimization

3D-IC Chiplet and Package Implementation

Architecture and IP



3D-IC Architect Cockpit

3D-IC System Planning and Optimization

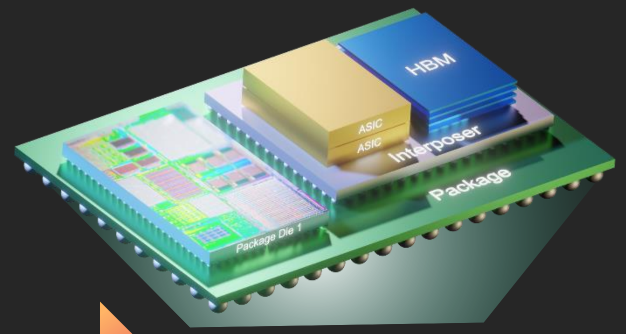
Chiplet Implementation

3D Analysis/Signoff

- Electromagnetic
- Thermal
- Timing
- Power

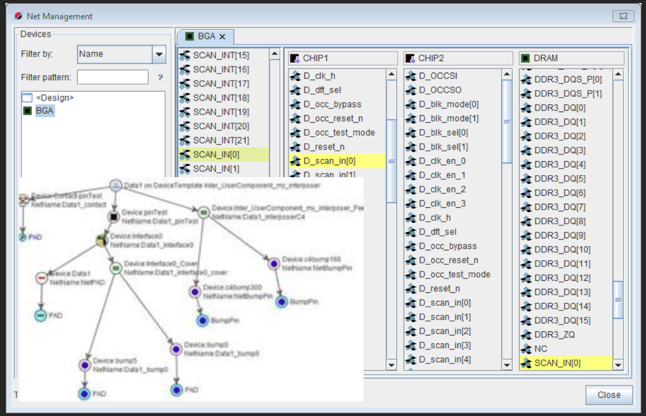
Analog and Package

- Analog and Custom IC Design
- Package Design

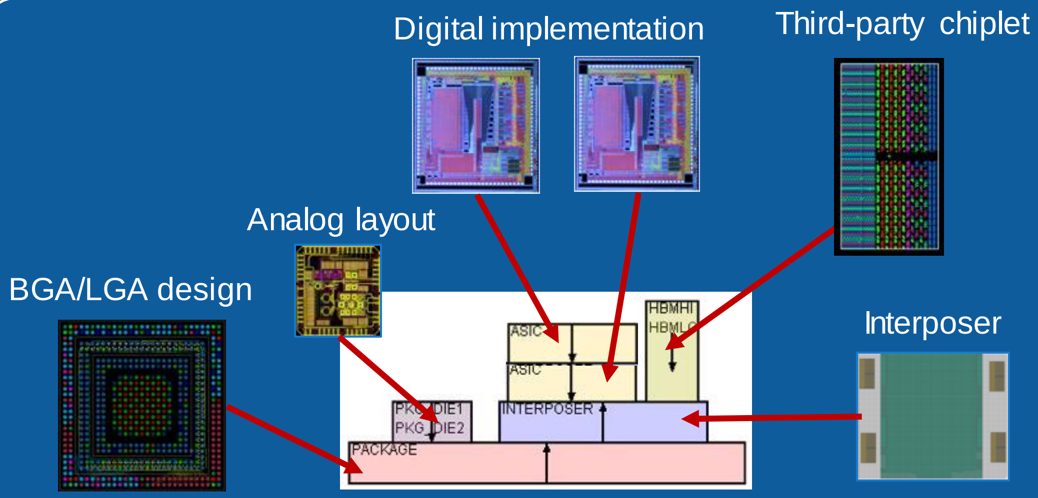


Optimized Chiplets and Design Package

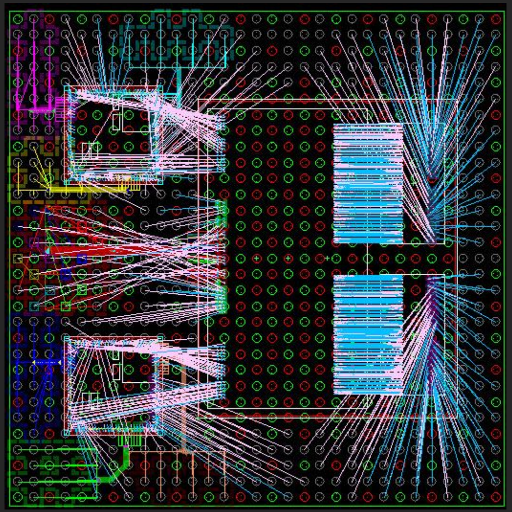
Top-Level 3D-IC Design Management & Optimization



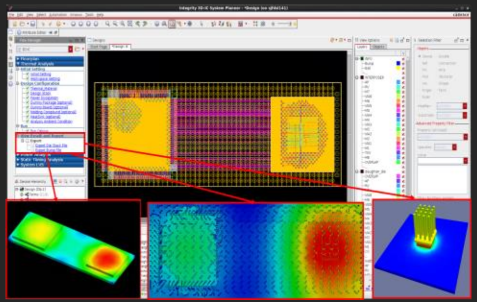
Chip(let)-chip(let)-package-board signal-mapping



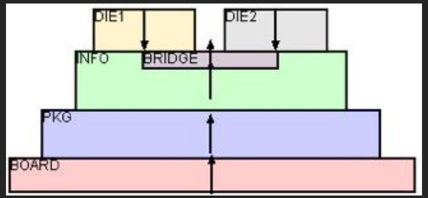
Hierarchical Planning and Optimization of System-level Design and Connectivity



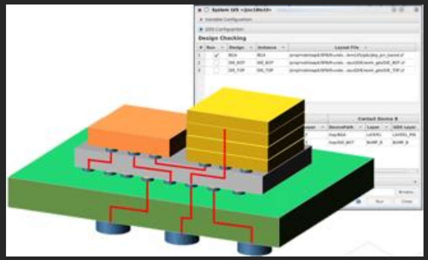
Top-level connectivity optimization



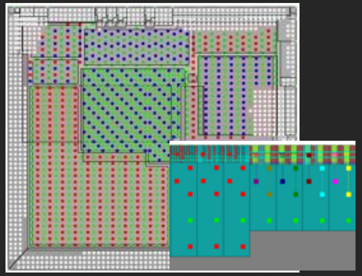
Early-Stage Thermal/Power Analysis



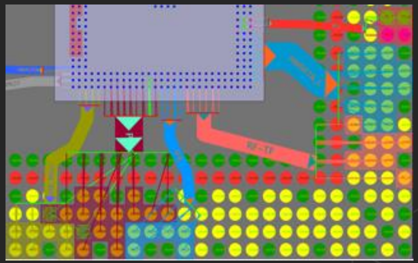
Stack management



System-level connectivity/stack alignment verification (SystemLVS)

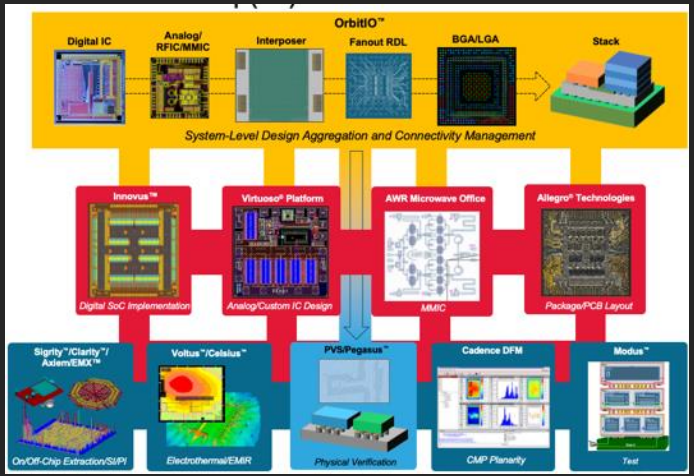


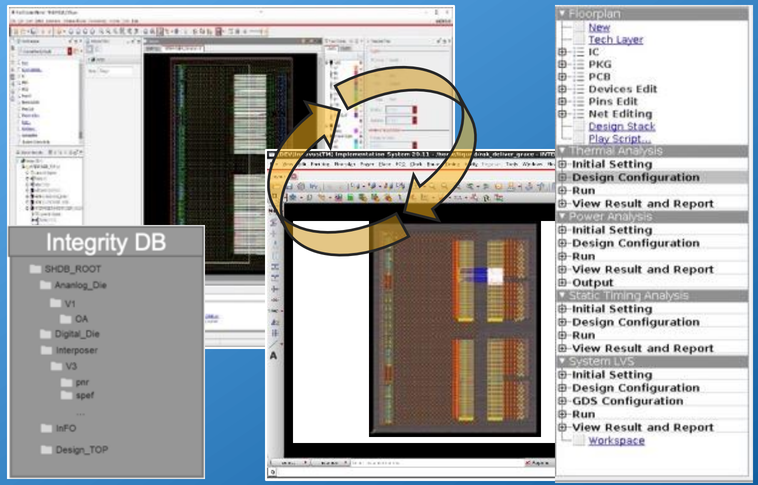
Advanced bump/TSV planning



Bundle/bus-driven pin optimization

The Cadence Integrity 3D-IC Platform Evolution





2019

- Complete 2D and 2.5D-/3D-IC packaging flows
- Stand-alone tools
- Complex work-flows

Point-Tool Centric

2020-2021

- Integrated multi-die planning and implementation
- Allegro co-design
- 3D-IC signoff & system analysis flows
- Integrity database

Direct Integration

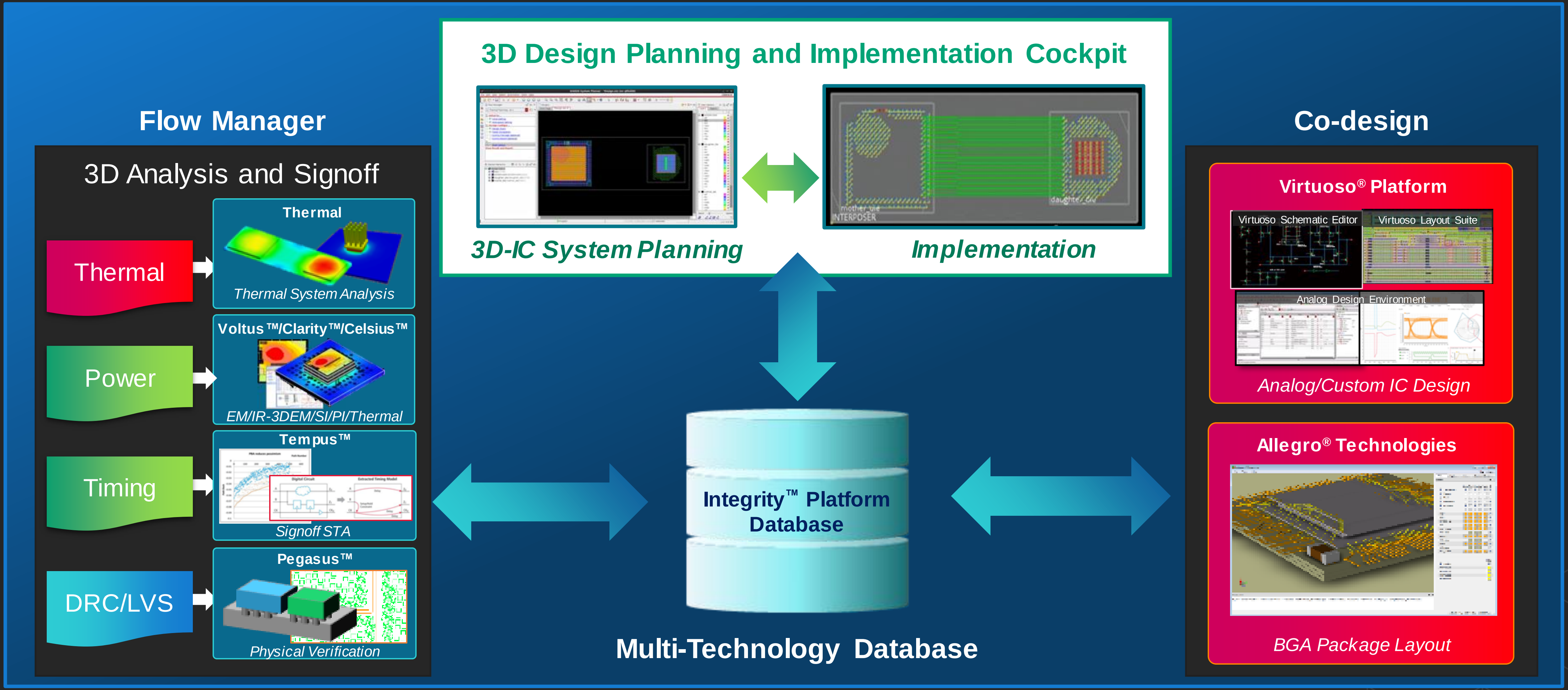
Q4 2021

- Cross-die optimization
- System-driven PPA considering thermal and power
- Virtuoso co-design
- Advanced multi-chip(let) STA

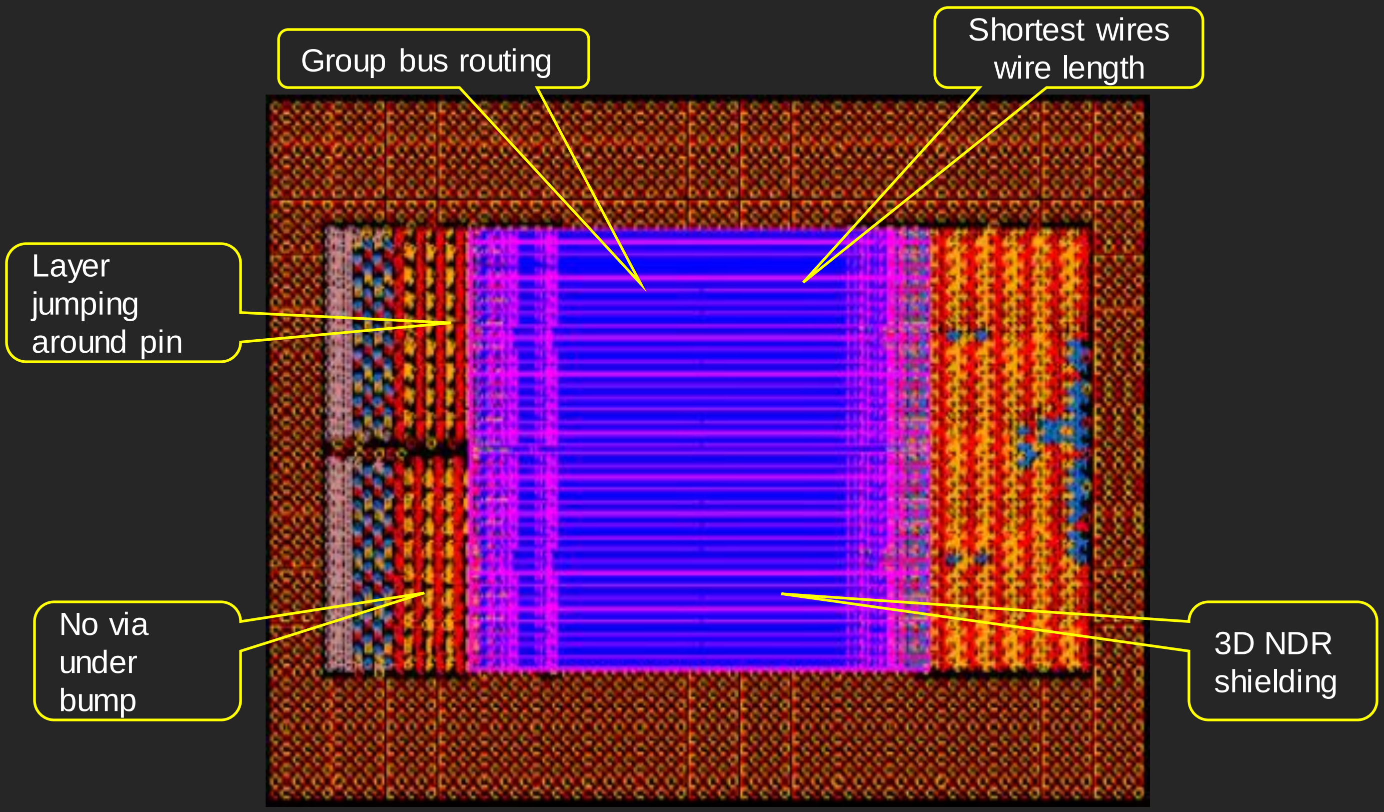
Platform

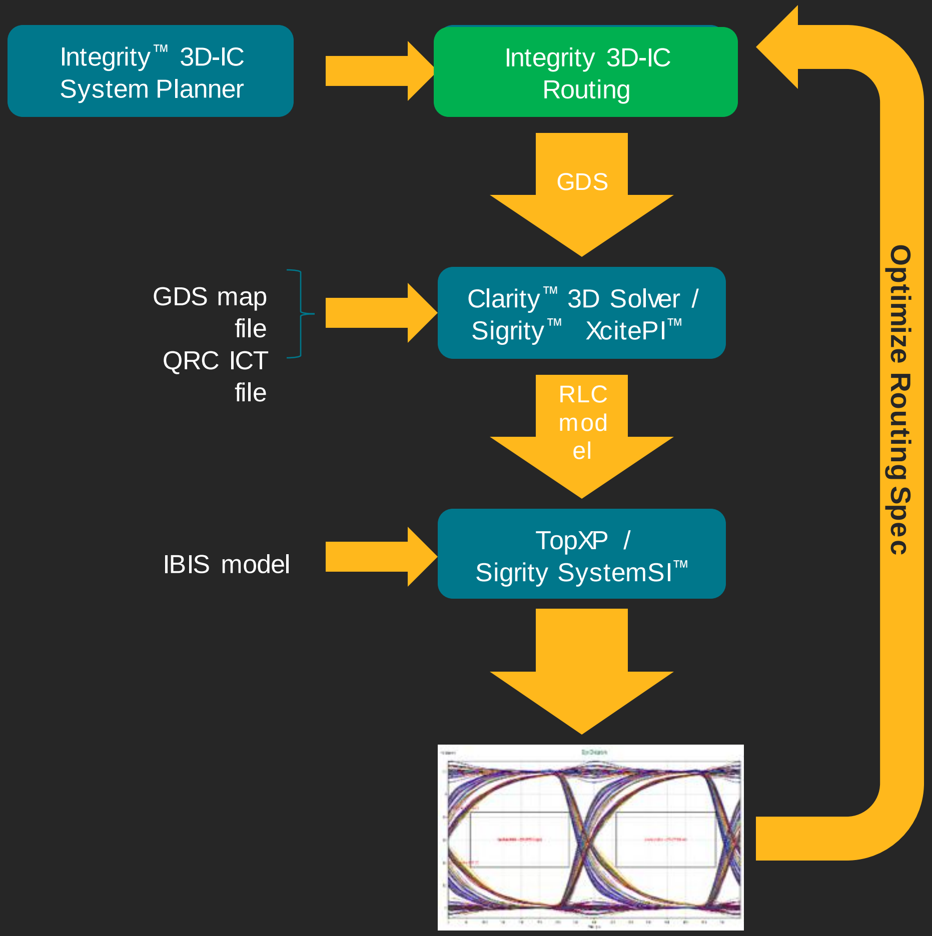
Cadence Integrity 3D-IC Platform Overview

Industry's first integrated, high-capacity 3D-IC platform that enables 3D design planning, implementation and system analysis in a single, unified cockpit



Interposer Implementation and Routing Flow

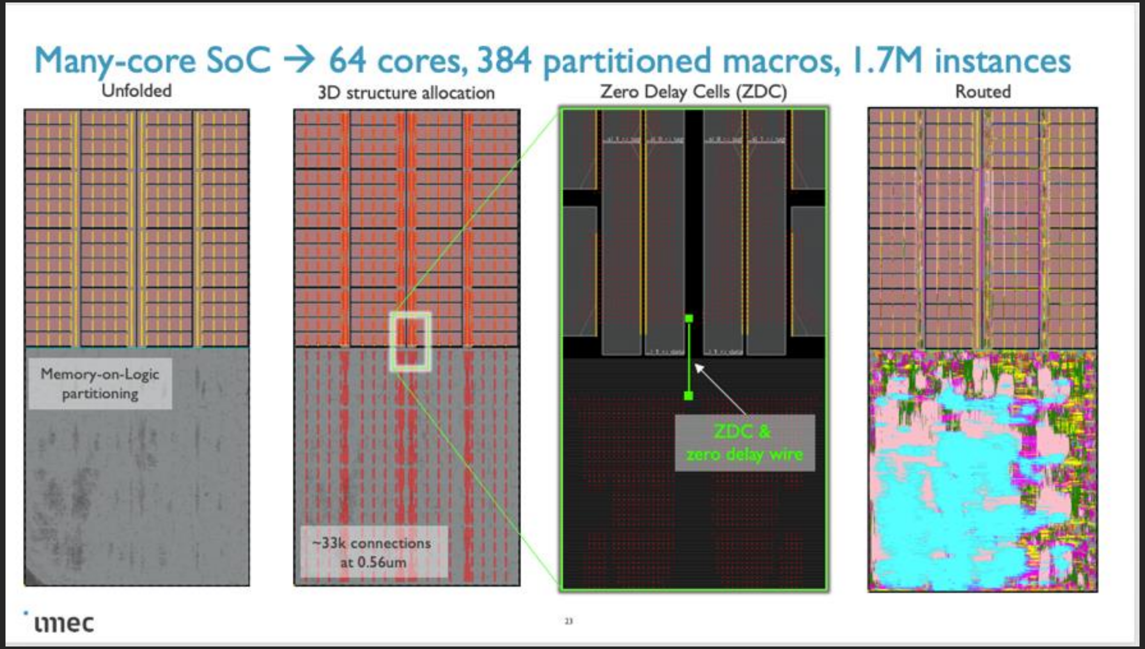


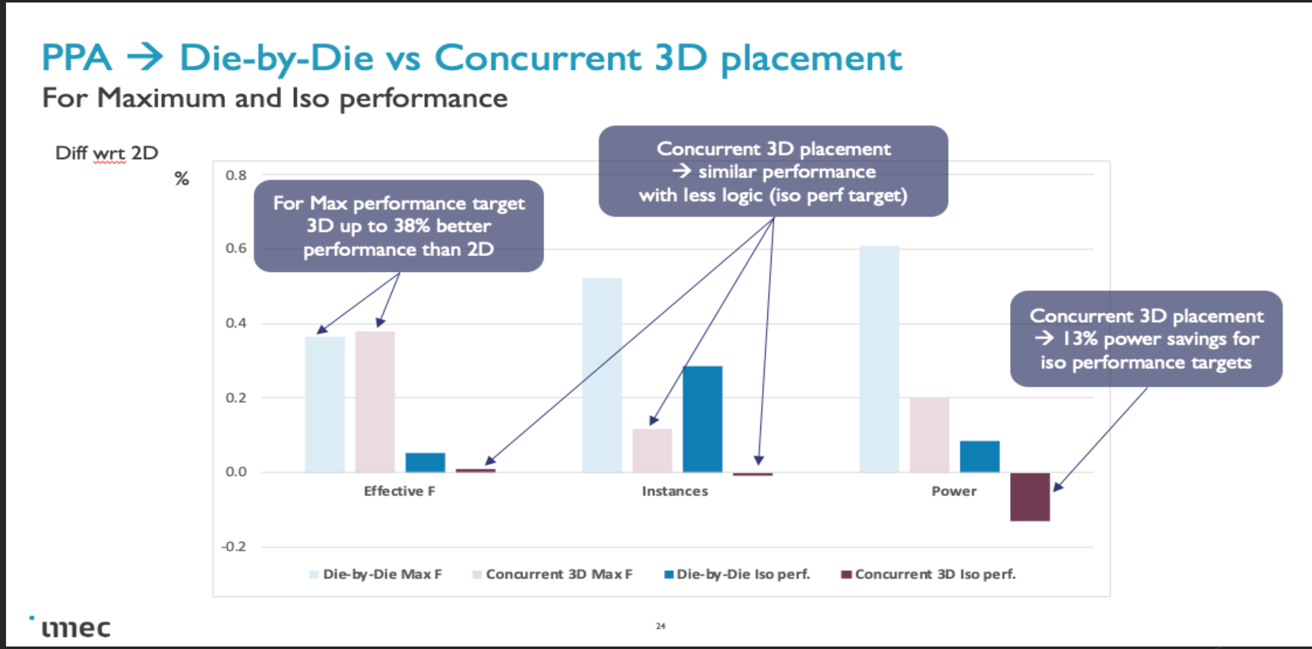


Case Study : Imec Experience

"With 3D-IC design continuing to gain momentum, there is an increased need to automate the planning and partitioning of a 3D stack die system more efficiently. "

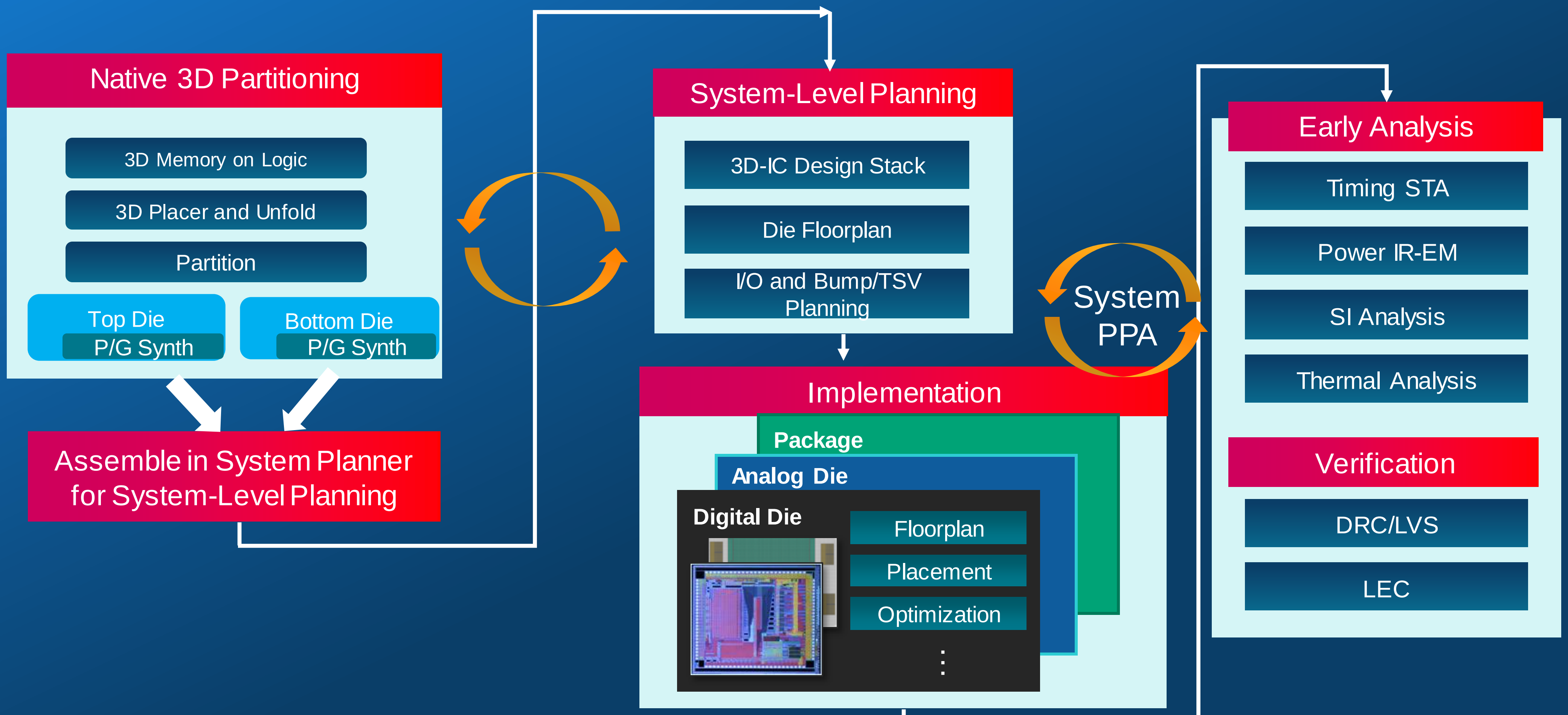
***Eric Beyne,** senior fellow and program director, 3D System Integration, imec*





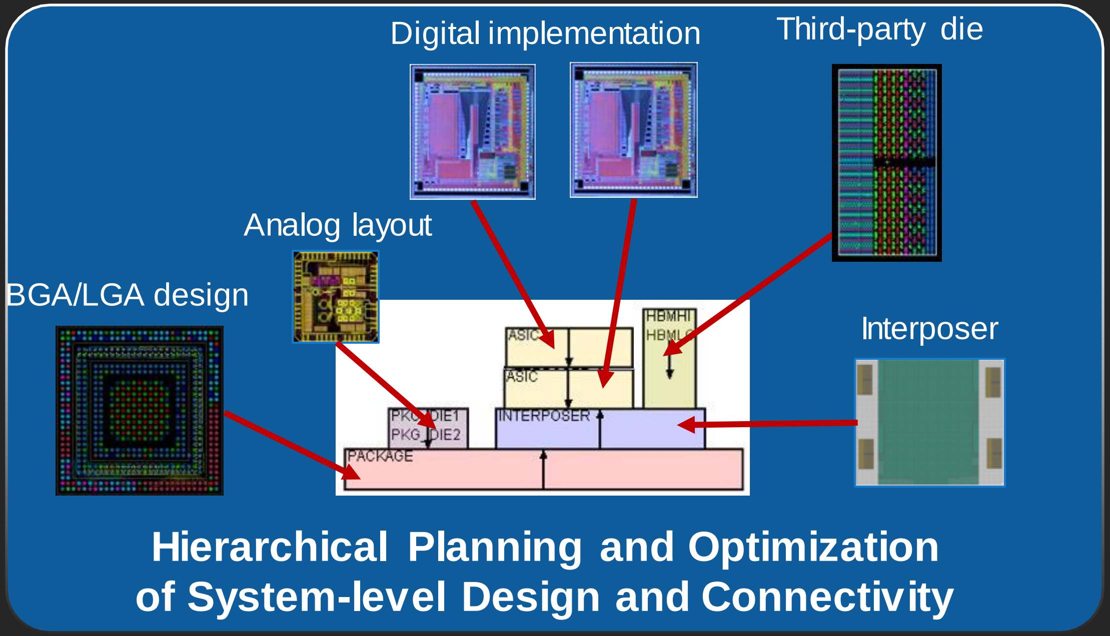
Source : CadenceLive Europe 2021

Summary : System-Driven PPA



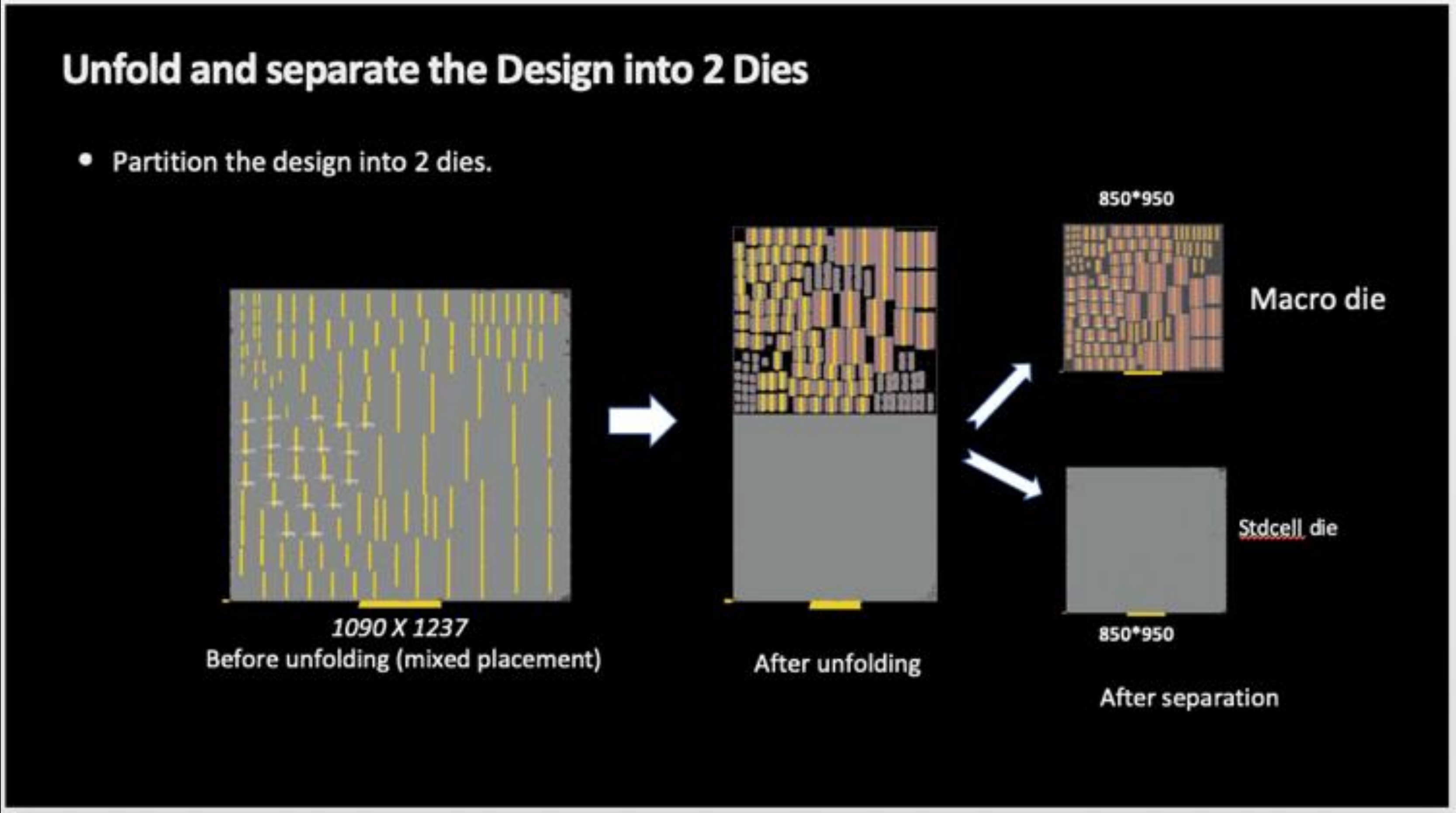
Comprehensive, yet modular

Goal of our research

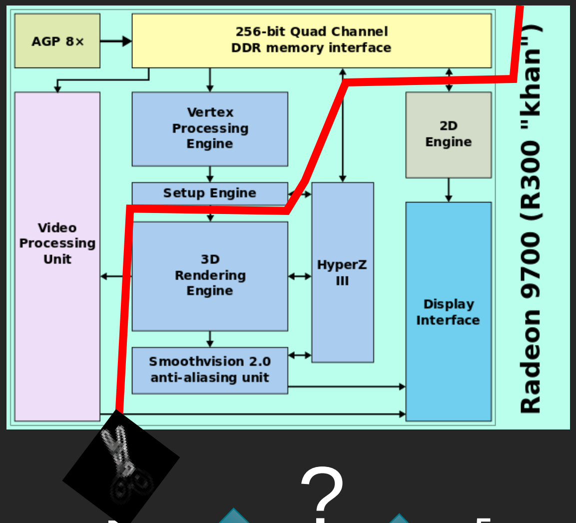


Identifying the partition of original design into components to be implemented as chiplets for optimal performance/cost/re-use potential

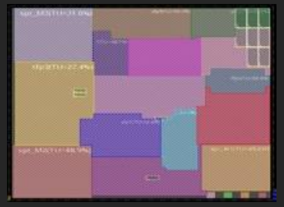
Partitioning



Module-Level 3D-IC Partitioning



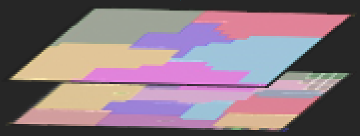
7nm ? 5nm



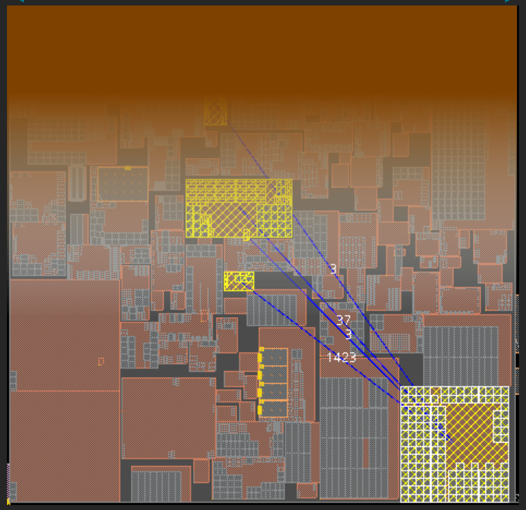
Which blocks should go to 5nm and which blocks should go to 7nm

7nm

5nm



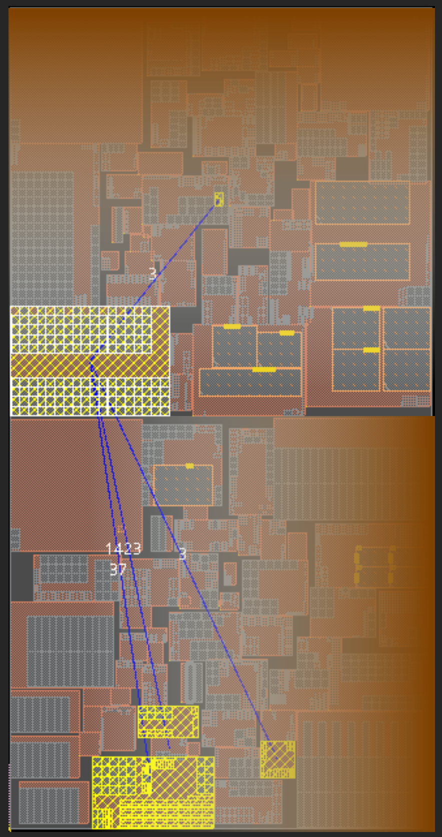
1.0



2D module placement

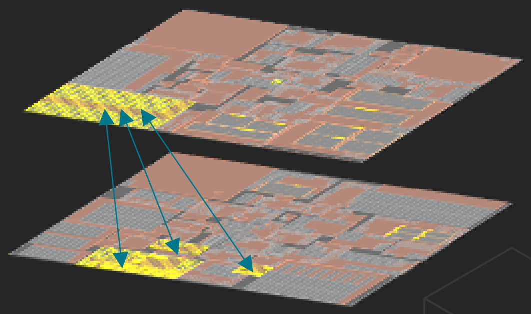
vs.

0.7



Shorter wire length in 3D

- Multi-die partitioning and hierarchical floorplan synthesis in one unified algorithm (EFS)
- Consider wire length, bump interface penalty, different process nodes and module utilization on the functional module level



ACKNOWLEDGEMENTS

- Cadence Design Systems (Vinay Patwardhan, Venkat Thanvantri, Michael Jackson, Anirudh Devgan, Chris Hook, Yoon Kim)
- KPIT (Anup Gupta)
- Zheng Liang

33