

The Era of Chiplets and its Impact to Future System Designs

Tobias Webel

IBM Senior Technical Staff Member (STSM)
Hardware Development
webel@de.ibm.com



Tobias Webel

- Manageability and Boot Security
Hardware Architect for IBM zSystems
and IBM POWER Systems
- Manageability
 - Reliability, Availability and Serviceability (RAS)
 - IBM zSystems High availability: 7 nines
 - Designed for 99.99999 % up time
- Boot Security
 - Selfboot
 - Secure Boot
 - Quantum Safe Secure Boot
- IBM zSystems
 - Data serving and transaction processing platform
 - Designed for large-scale enterprise computing with a focus on RAS
- IBM POWER Systems
 - Designed for high-performance general-purpose computing workloads

IBM zSystems

Single frame 42U machine (z16)
could go up to 4 frames



Market Segments

45 of the world's top 50 banks
8 of the top 10 insurers
4 of the top 5 airlines
7 of the top 10 global retailers
8 of the top 10 telco's

IBM POWER Systems

Single frame 42U machine (E1080)



2U rack (S1022)



Agenda

- Chiplets
 - Introduction and Motivation
- State of the Art
 - IBM, INTEL, AMD
- Universal Chiplet Interconnect Express (UCIe) 1.0 Specification
- Usage Models for UCIe
- Future of Chiplets

Lecture: The Era of Chiplets and its Impact to Future System Designs

Tobias Webel (IBM)

Moore's Law, which predicts that the number of transistors on a microchip will double every 18 to 24 months, is still a guiding principle in the design and manufacture of computer hardware. While some experts have predicted that Moore's Law may eventually reach its limits due to the physical constraints of semiconductor technology, many others believe that new innovations and approaches will continue to push the boundaries of what is possible.

One innovation area which is currently heavily pursued by key Hardware companies is the so called chiplet approach. Chiplets are small, modular components that can be combined to create larger, more complex systems. They allow for greater flexibility and customization in the design of integrated circuits, and can also help to reduce costs and increase efficiency. Chiplets have become increasingly popular in recent years, and are expected to play a major role in the development of future computing systems.

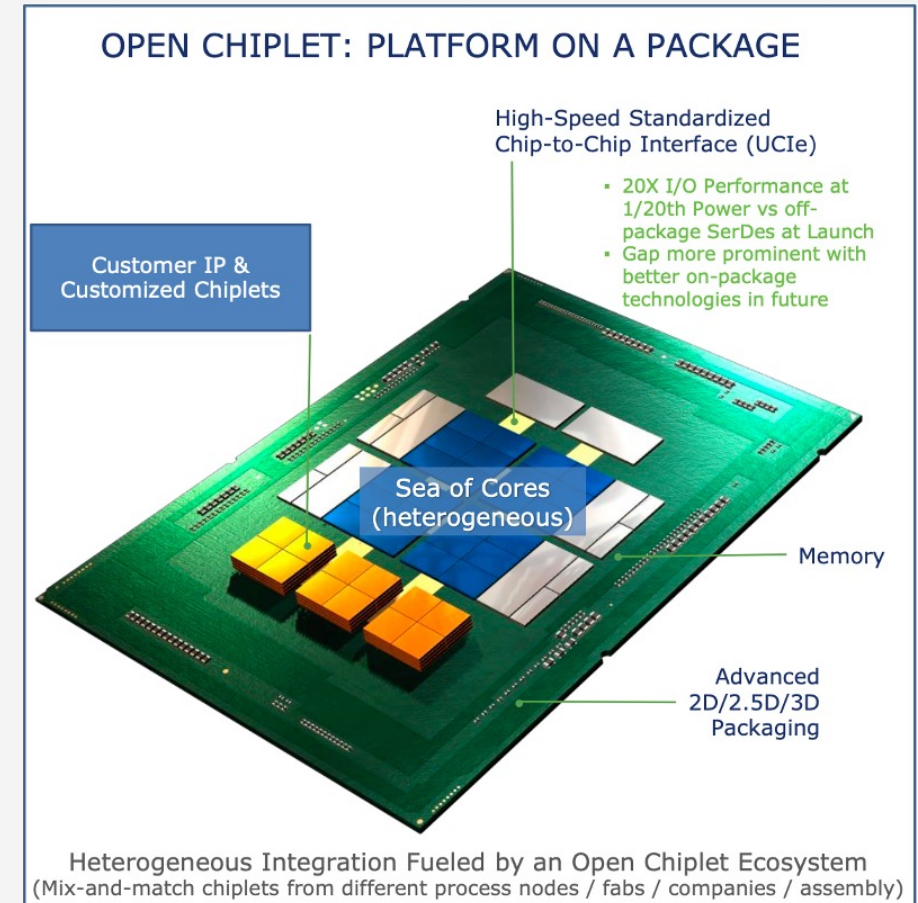
Intel CEO Pat Gelsinger promised at company event that "Moore's law is alive and well," adding that "we are predicting that we will maintain or even go faster than Moore's law for the next decade.

This lecture will introduce the chiplet concept, addresses opportunities and challenges of the chiplet approach and will share insights in how the industry moves forward with an openness and standardization.

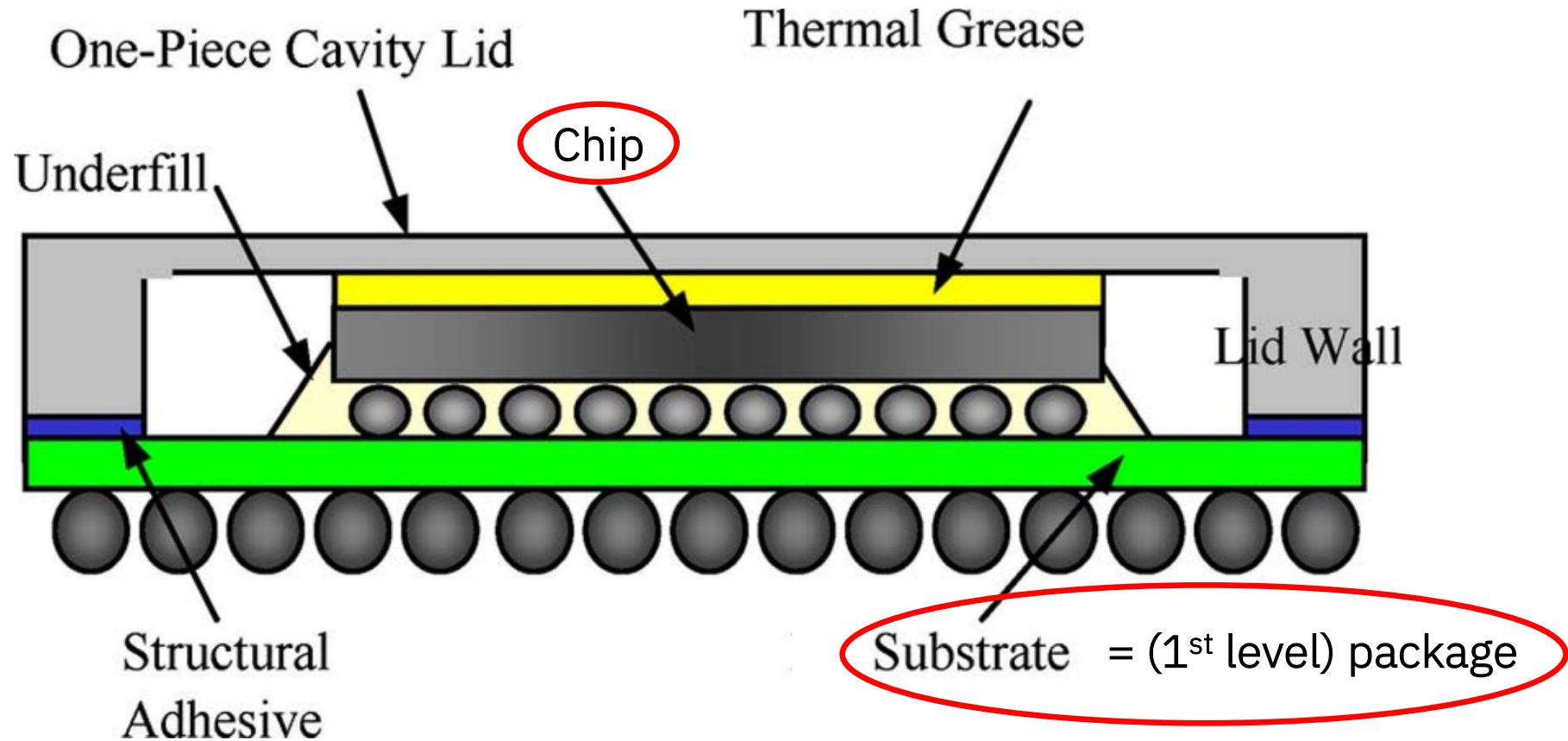
Tobias Webel (IBM)



Chiplets

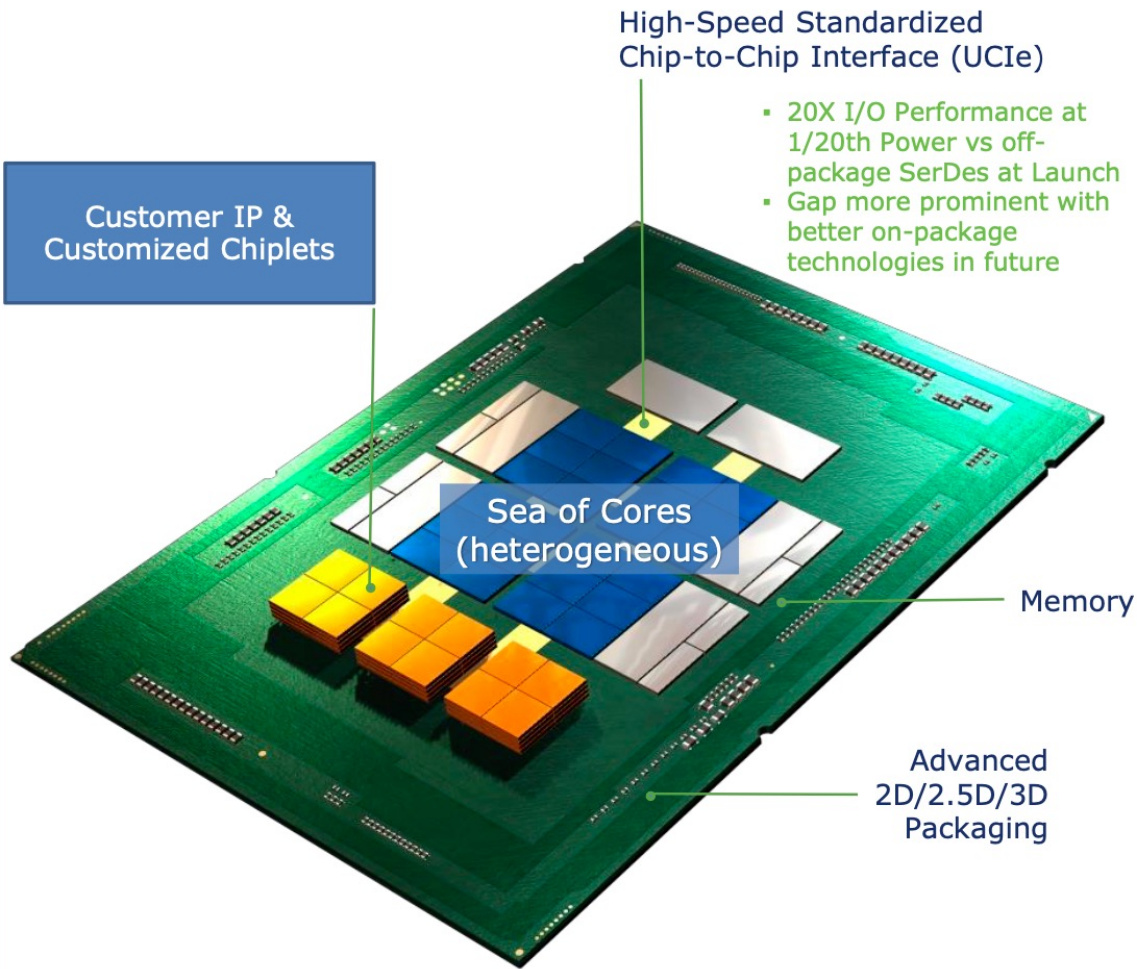


High Performance Single Chip Module



Motivation

OPEN CHIPLET: PLATFORM ON A PACKAGE



Heterogeneous Integration Fueled by an Open Chiplet Ecosystem
(Mix-and-match chiplets from different process nodes / fabs / companies / assembly)

Align Industry around an open platform to enable chiplet based solutions

- Enables construction of SoCs that exceed maximum reticle size
 - Package becomes new System-on-a-Chip (SoC) with same dies (Scale Up)
- Reduces time-to-solution (e.g., enables die reuse)
- Lowers portfolio cost (product & project)
 - Enables optimal process technologies
 - Smaller (better yield)
 - Reduces IP porting costs
 - Lowers product SKU cost
- Enables a customizable, standard-based product for specific use cases (bespoke solutions)
- Scales innovation (manufacturing/ process locked IPs)

SKU: Stock Keeping Unit

Major Chiplet Industry Initiatives

Universal Chiplet Interconnect Express (UCIe)

- Founded: 2022
- Goal: Heterogeneous integration of silicon by an open chiplet ecosystem
- Die-to-Die interface: UCIe



UCIe™
Universal Chiplet
Interconnect Express

80+
members

UCIe™ Board Members





UCIe™ Contributor Members



Property of Universal Chiplet Interconnect Express™ (UCIe™) as of September 2022

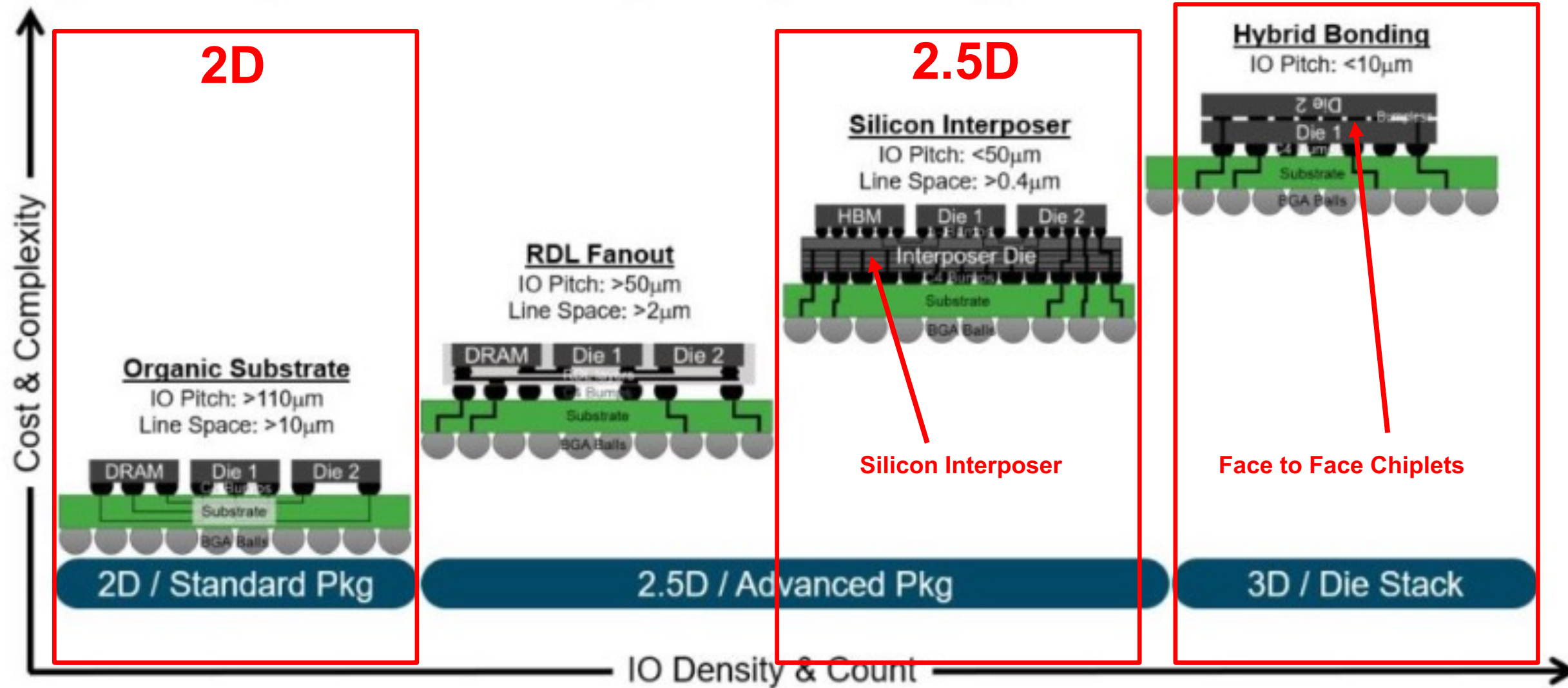
Open Compute Project (OCP)

- Founded: 2011
- Open Domain Specific Architecture (ODSA) WG
- Goal: Integrate best of class chiplets from multi vendors thru open interfaces
- Die-to-Die interface: Bunch Of Wires (BoW)

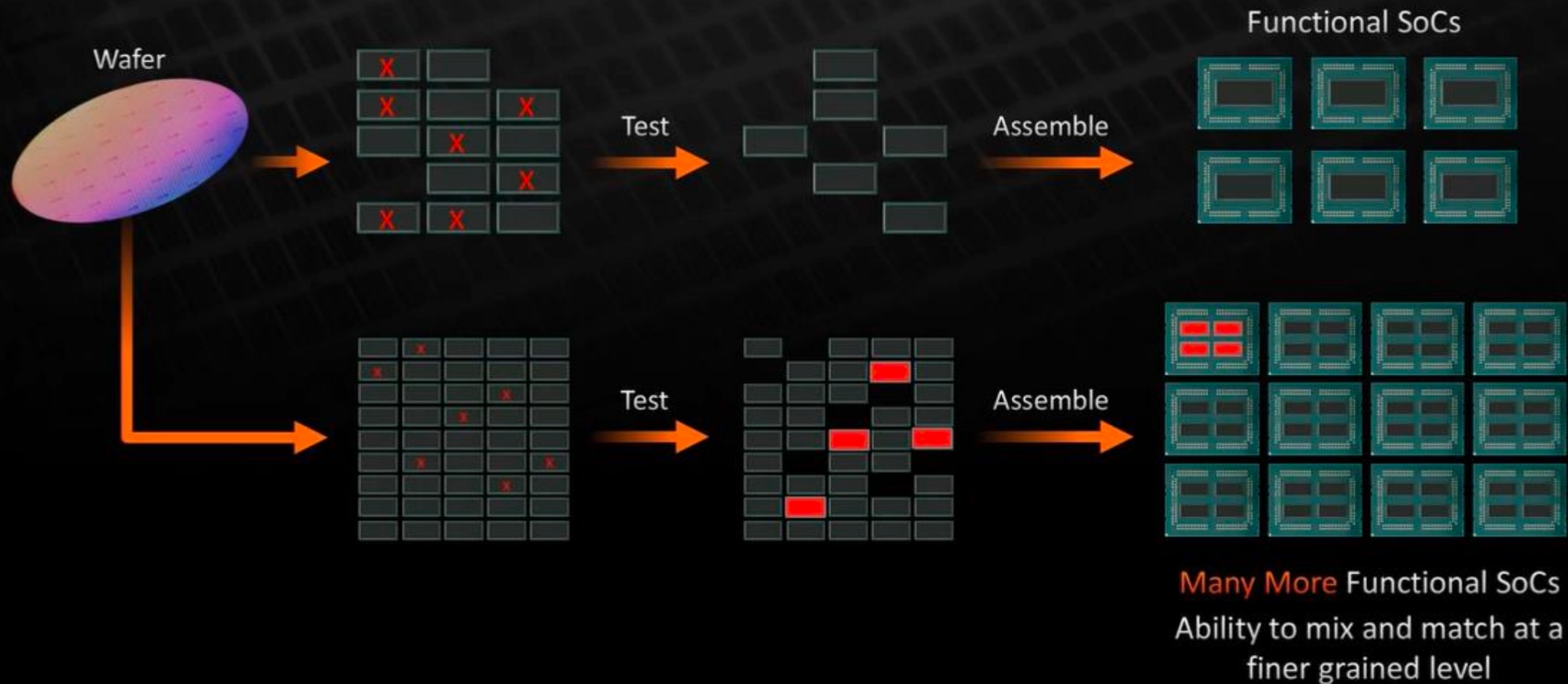
Alibaba Group (since 2017)	AMD (since 2012)	AMI (since 2018)	Ampere Computing (since 2019)
Alibaba Group	AMD	ami	AMPERE
Arista Networks (since 2019)	ARM (since 2018)	ASUS (since 2019)	AT&T (since 2015)
ARISTA	arm	ASUS	AT&T
Baidu (since 2019)	ByteDance Technology (since 2021)	Cisco (since 2014)	Delta Electronics (since 2016)
Baidu	ByteDance	CISCO	DELTA
Edgework Networks (since 2016)	Google (since 2015)	HPE (since 2015)	Huawei (since 2021)
Edgework Networks	Google	Hewlett Packard Enterprise	HUAWEI
Hyve Solutions (since 2012)	IBM (since 2013)	Inspur (since 2016)	Intel (since 2011)
hyve solutions	IBM	inspur	intel
Meta (since 2011)	Microsoft (since 2014)	MITAC (since 2017)	Nokia (since 2015)
Meta	Microsoft	MITAC	NOKIA
NVIDIA (since 2017)	Quanta Cloud Technology (since 2012)	Rackspace (since 2011)	Ragile Networks (since 2020)
NVIDIA	QCT	rackspace	RAGILE
Rittal (since 2017)	Samsung Electronics (since 2019)	Schneider (since 2014)	Shell (since 2020)
RITTAL	SAMSUNG	Life is On Schneider	Shell
Sims Lifecycle Services (since 2020)	Submer (since 2018)	Target (since 2021)	Tencent (since 2018)
SIMS LIFECYCLE SERVICES	submer	target	Tencent 腾讯
UL Inc. (since 2021)	Vertiv (since 2019)	Wiyynn (since 2014)	
UL Solutions	VERTIV	wiwyynn	

Advanced Packaging Technology Summary

Integration of up to 4x reticle limit by 2024 (~2.5x today)

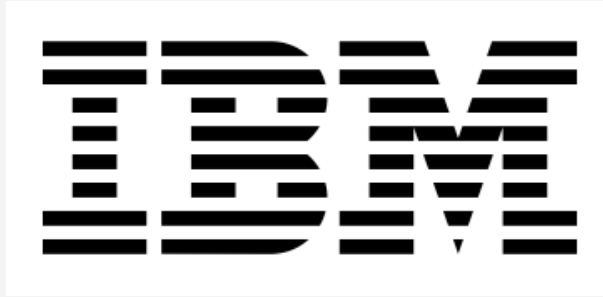


HIGH-LEVEL APPROACH TO CHIPLETS



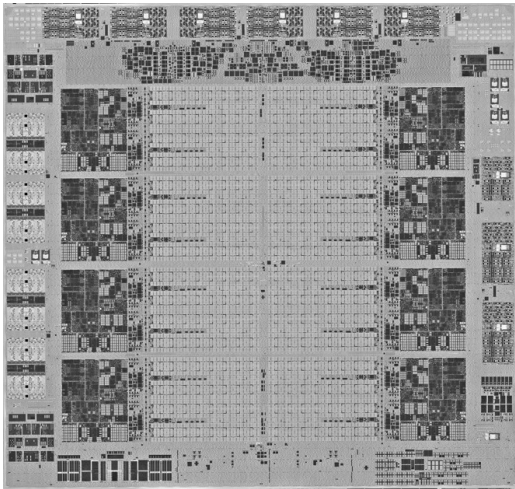
State of the Art

IBM, INTEL, AMD



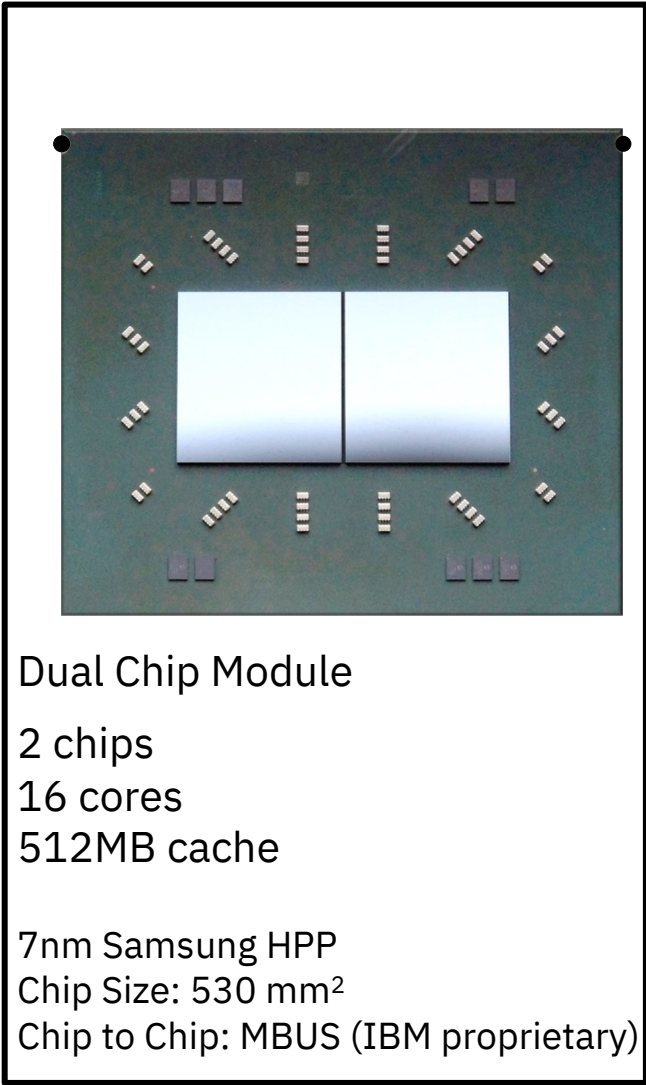
State of the Art

IBM zSystems, Machine Generation z16, GA: 2Q22



Single Chip

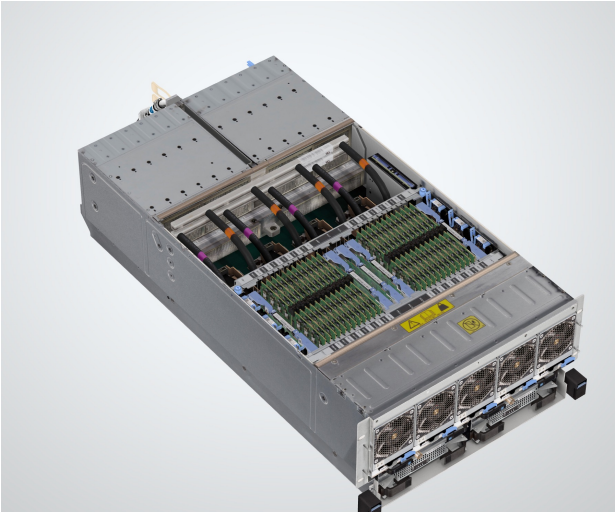
- 1 chip
- 8 cores
- 256MB cache



Dual Chip Module

- 2 chips
- 16 cores
- 512MB cache

7nm Samsung HPP
Chip Size: 530 mm²
Chip to Chip: MBUS (IBM proprietary)



4-Socket Drawer

- 8 chips
- 64 cores
- 2GB cache



4-drawer system

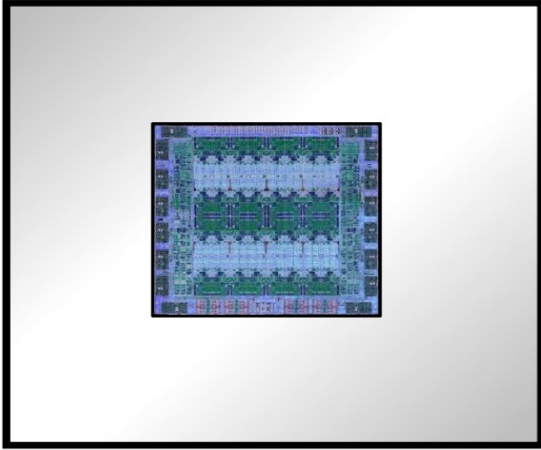
- 32 chips
- 256 cores (*)
- 8GB cache

(*) up to 200 client configurable cores in max system

State of the Art

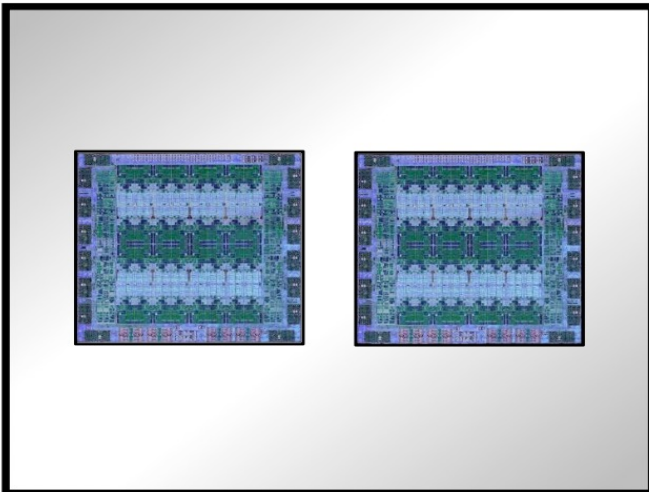
IBM POWER, Machine generation POWER10, GA: 3Q21

Socket Composability: **SCM & DCM**



Single-Chip Module Focus:

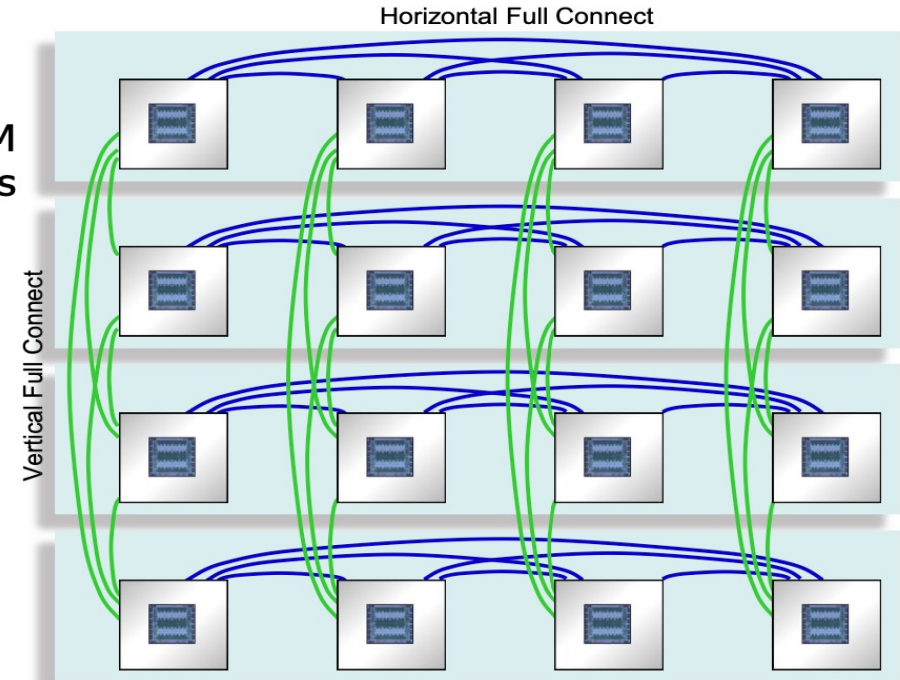
- 602mm² 7nm (18B devices)
- **Core/thread Strength**
 - Up to 15 SMT8 Cores (4+ GHz)
- **Capacity & Bandwidth / Compute**
 - Memory: x128 @ 32 GT/s
 - SMP/Cluster/Accel: x128 @ 32 GT/s
 - I/O: x32 PCIe G5
- **System Scale (Broad Range)**
 - 1 to 16 sockets



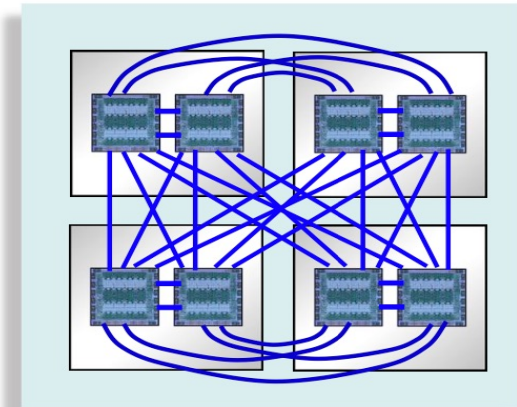
Dual-Chip Module Focus:

- 1204mm² 7nm (36B devices)
- **Throughput / Socket**
 - Up to 30 SMT8 Cores (3.5+ GHz)
- **Compute & I/O Density**
 - Memory: x128 @ 32 GT/s
 - SMP/Cluster/Accel: x192 @ 32 GT/s
 - I/O: x64 PCIe G5
- 1 to 4 sockets

Up to
16 SCM
Sockets



Up to
4 DCM
Sockets



IBM POWER10

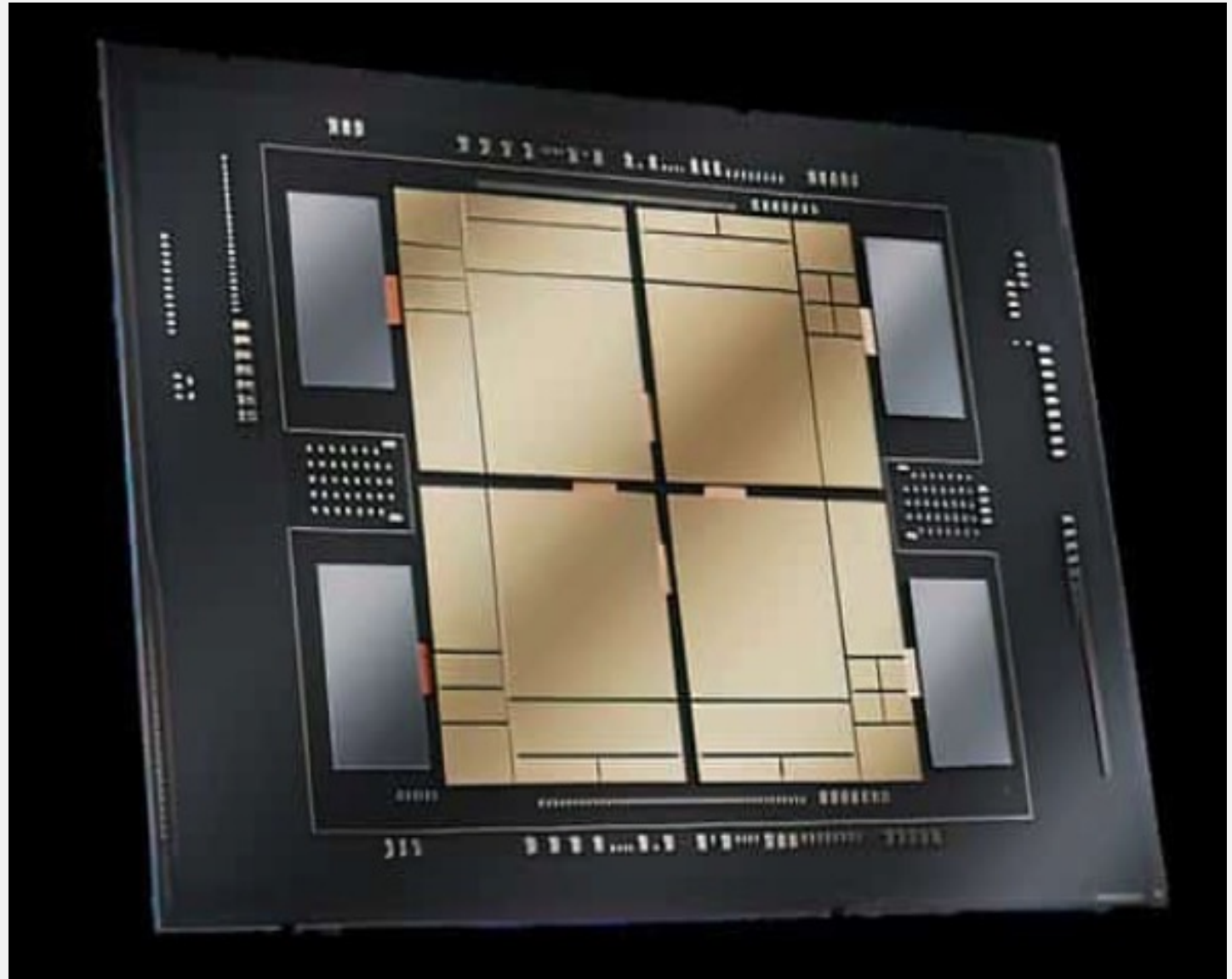
(Multi-socket configurations show processor capability only, and do not imply system product offerings)

INTEL State of the Art (CPU)

Server CPU

INTEL Sapphire Rapids

- GA: January 2023
- 56 cores
- 4x Processor Chiplets
 - 7nm Intel
 - 400 mm²
- 2.5D Packaging (EMIB)
- 100 billion transistors
- 4x HBM2e (64GB)
- CXL 1.1+
- Chiplet to Chiplet Interface:
 - Modular Die Interface (MDF)
 - Intel proprietary

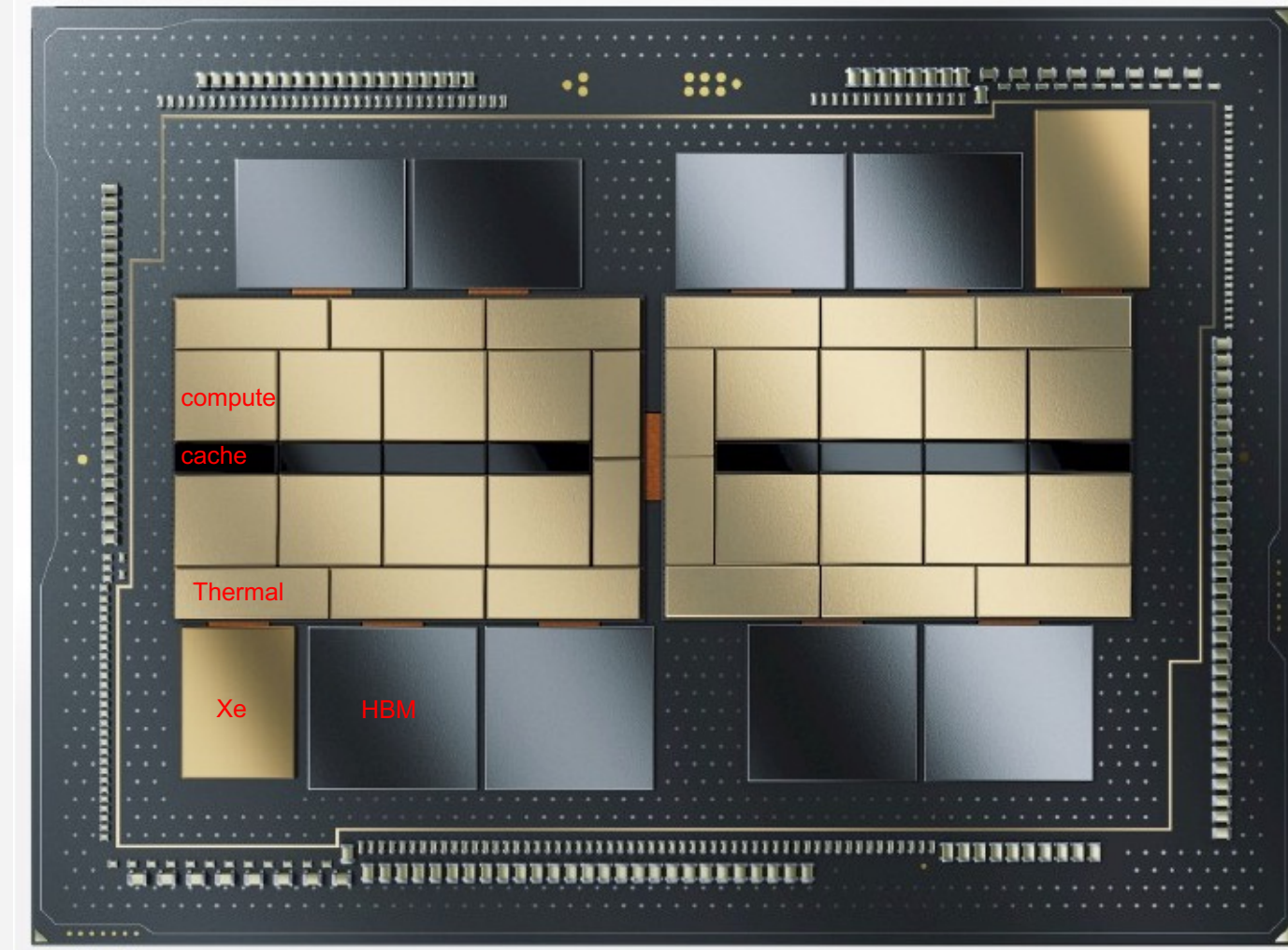


INTEL State of the Art (GPU)

GPU

INTEL Ponte Vecchio

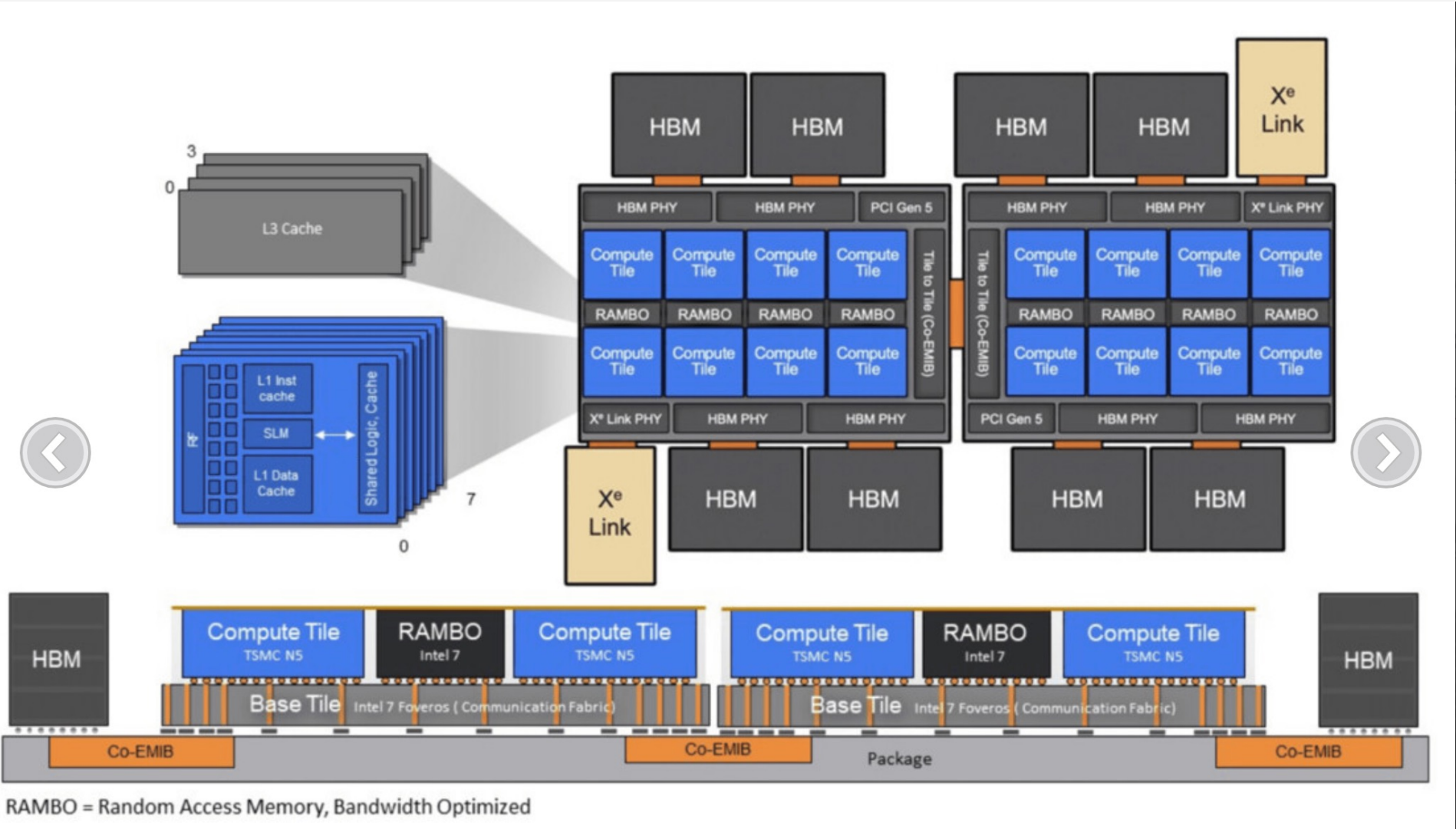
- GA: January 2023
- 16x compute chiplets
 - 5nm TSMC
- 8x Rambo Cache chiplets
 - 7nm Intel
- 2x Xe-Link chiplets
 - 7nm TSMC
- 2x Base chiplets (MEM, PCIe, CXL)
 - 7nm Intel
 - 646 mm²
- 8x HBM (128 GB)
- 11x EMIB chiplets
- 16 thermal chiplets
- 2.5D (EMIB) + 3D (Foveros) Packaging
- Silicon area = 2330 mm²
- 100 billion transistors
- CXL 1.1+



- Exascale supercomputer, Aurora, sponsored by the United States Department of Energy (DOE), expected completion is 2023
- Each node consists of
 - two CPUs (Intel Sapphire Rapids)
 - Six GPUs (**Intel Ponte Vecchio**)

INTEL State of the Art (GPU)

GPU INTEL Ponte Vecchio

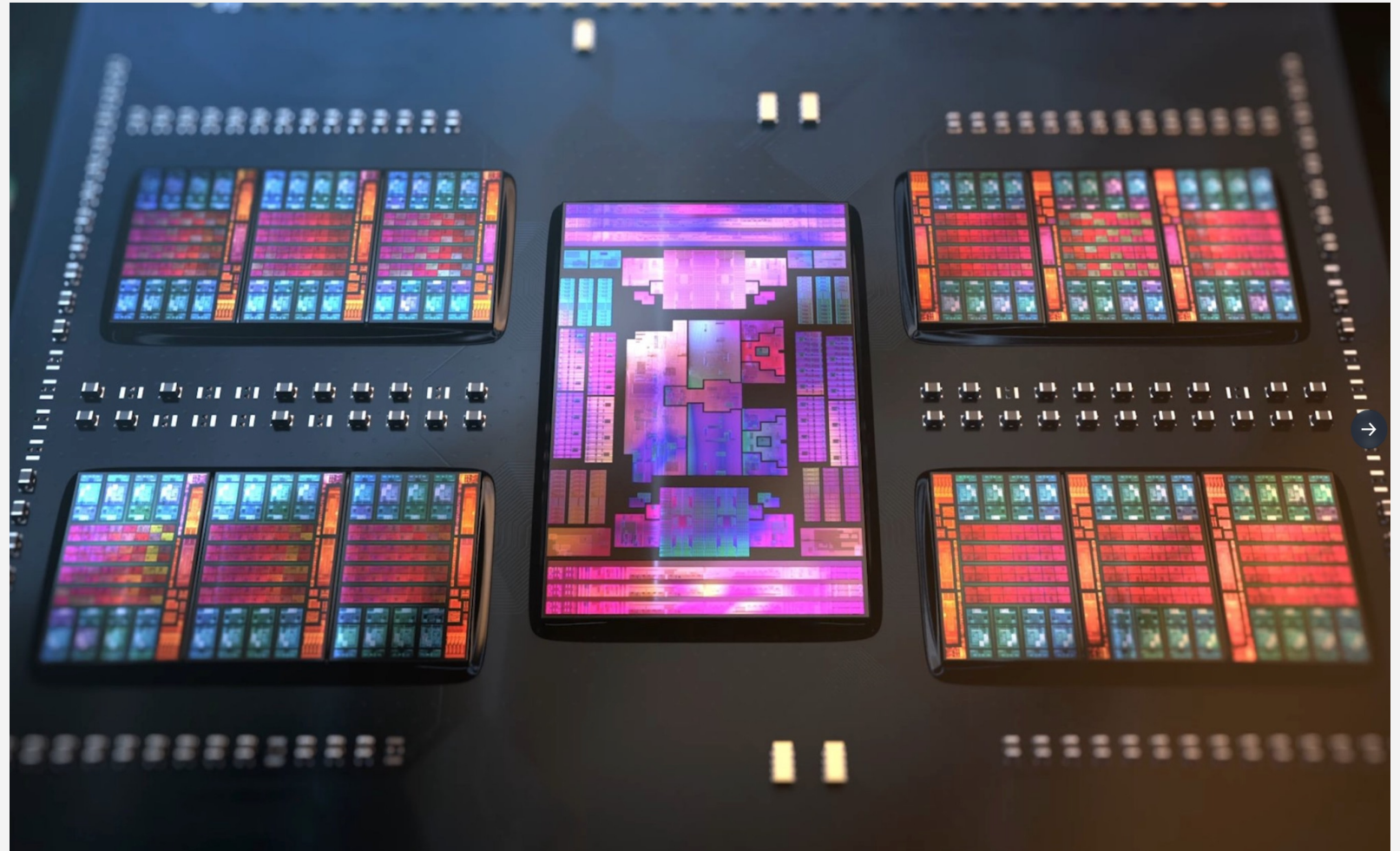


AMD State of the Art (CPU)

Server CPU

AMD Epyc Genoa 4th Generation (9654P)

- GA: November 2022
- 96 cores
- 12x CCD (Core-Cache Dies) Chiplets
 - 5nm TSMC
 - 72 mm²
- 1x IO Chiplet
 - 6nm TSMC
 - 397 mm²
- 2.5D Packaging
- 90 billion transistors
- CXL 1.1+
- Chiplet to Chiplet Interface:
 - Infinity Fabric On package (IFOP)
 - AMD proprietary

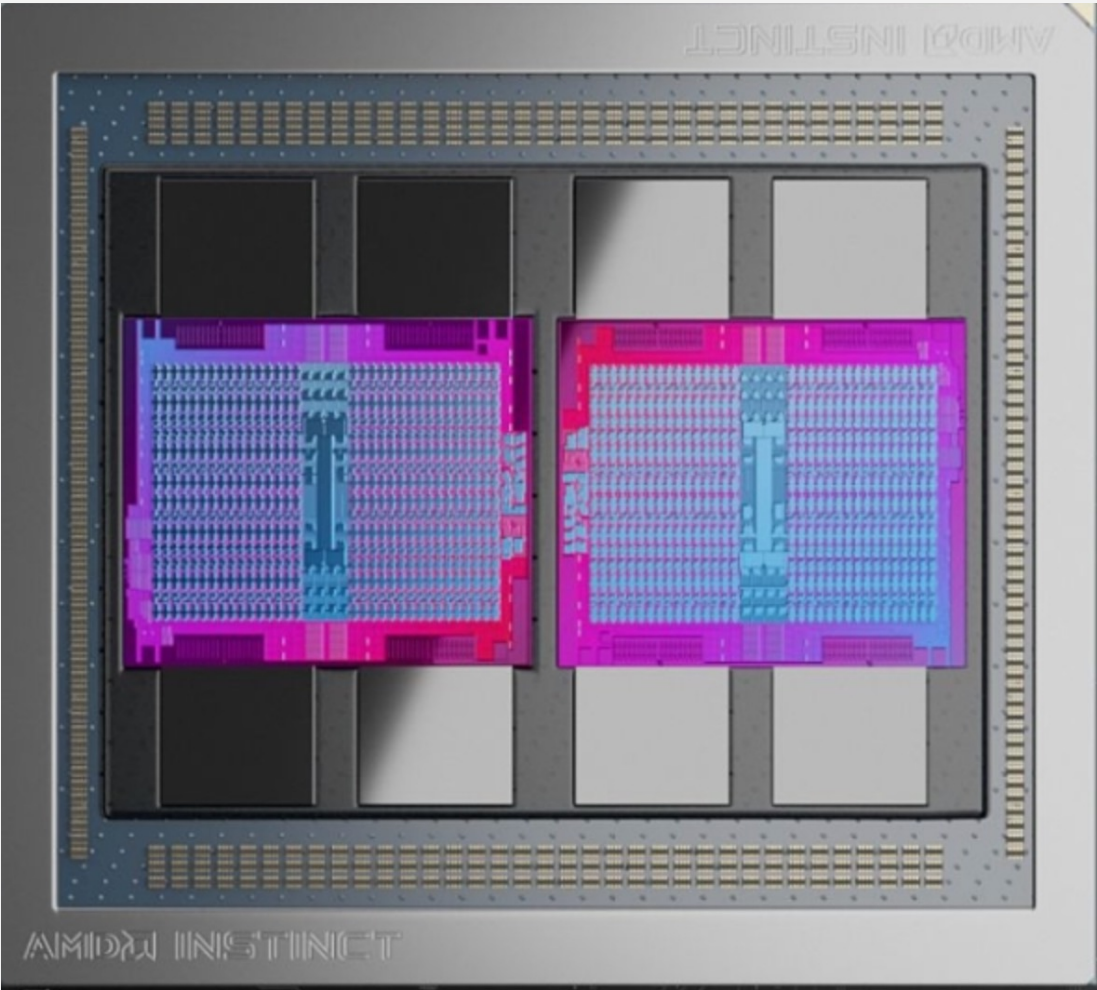


AMD State of the Art (GPU)

GPU

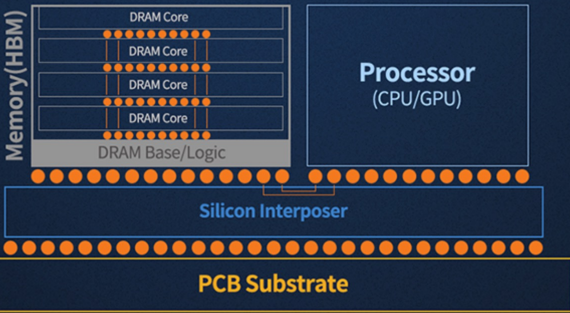
AMD Instinct MI250X

- GA: November 2021
- 14080 stream processors (CDNA 2.0)
- 2x Processor Chiplets
 - 6nm TSMC
 - 724 mm²
- 8x HBM2e (128 GB)
- 2.5D + 3D Packaging
- 116 billion transistors

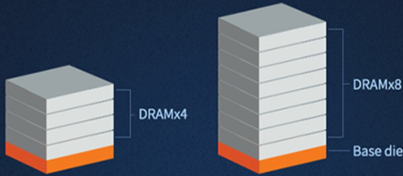


- Exascale supercomputer, Frontier, hosted at the Oak Ridge Leadership Computing Facility (OLCF) in Tennessee, United States and first operational in 2022
- Each node consists of
 - one CPU (AMD Epyc 3rd Generation 7453s "Trento" 64 core)
 - four GPUs (**Instinct MI250X**)

HBM in SiP (System-In-Package)



HBM TECHNOLOGY ROADMAP



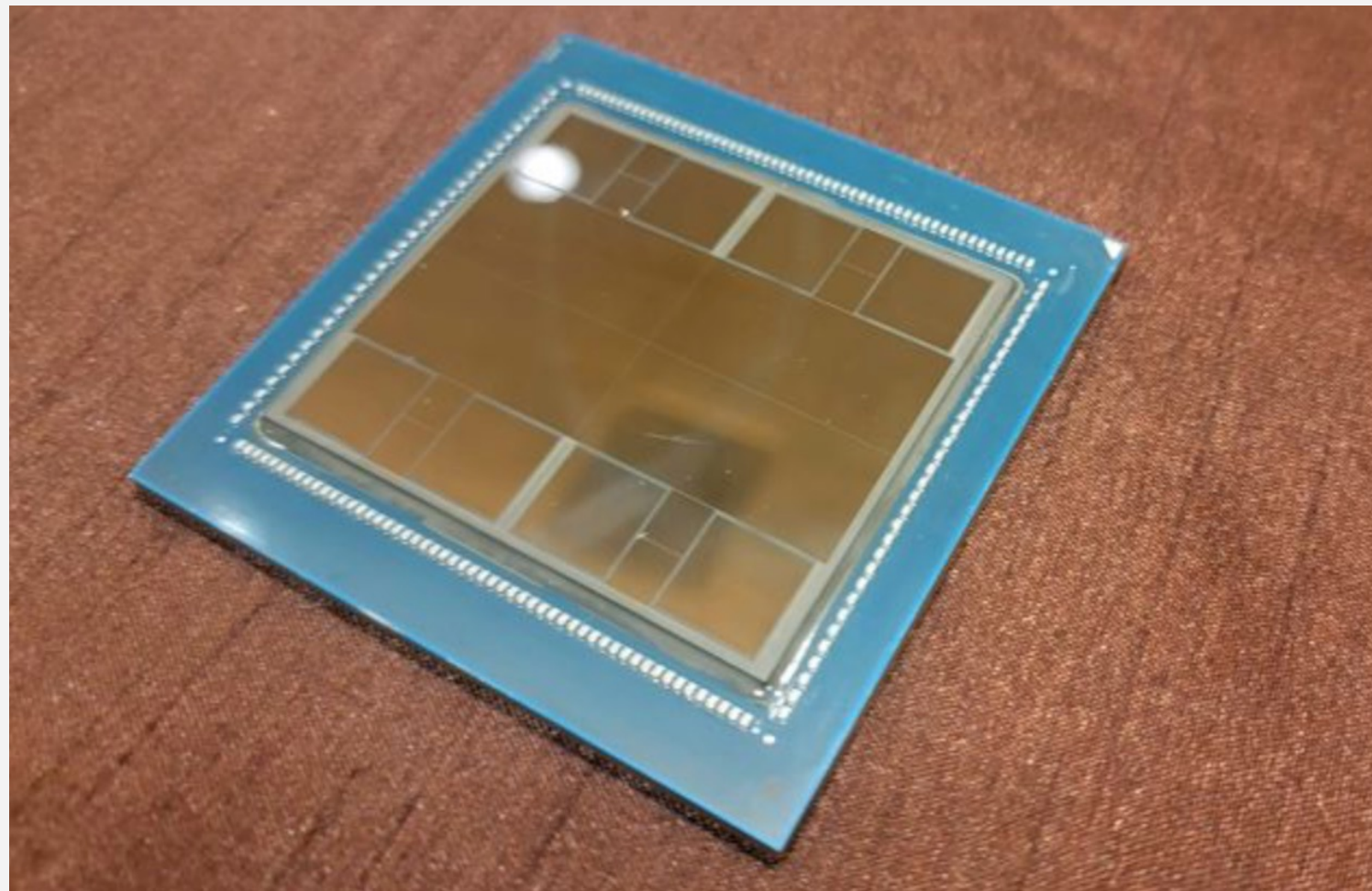
	HBM1	HBM2	HBM2E
Die Density (Gb)	2Gb	8Gb	16Gb
Stack Height	4Hi	4Hi / 8Hi	4Hi / 8Hi
Capacity (GB)	1GB	4GB / 8GB	8GB / 16GB
Bandwidth(GB/s)	128GB/s (1Gbps)	307GB/s (2.4Gbps)	410GB/s (3.2Gbps)

AMD State of the Art (APU = CPU + GPU)

APU

AMD Instinct MI300 APU (Accelerated Processing Unit)

- GA: 2H 2023
 - 24 zen4 cores
 - TBD stream processors (CDNA 3.0)
 - 13 chiplets
 - 9 chiplets 5nm TSMC (CPU, GPU ?)
 - 4 chiplets 6nm TSMC (MEM, IO, .. ?)
 - 5nm chiplets sit on top of 6nm chiplet
 - 8x HBM3 (128 GB)
 - 2.5D + 3D Packaging
 - 146 billion transistors
 - CXL 3.0
-
- MI300 announced by AMD, Consumer Electronic Show (CES), January 5th, 2023
-
-
-
-
-
-
-
-
-
-
- Exascale supercomputer, **El Capitan**, hosted at the Lawrence Livermore National Laboratory, United States and projected to become operational in 2023
 - El Capitan has been announced to use an unknown number of [AMD Instinct](#) MI300 CPUs



3D CPU+GPU Integration for Next-Level Efficiency

CDNA: Compute DNA

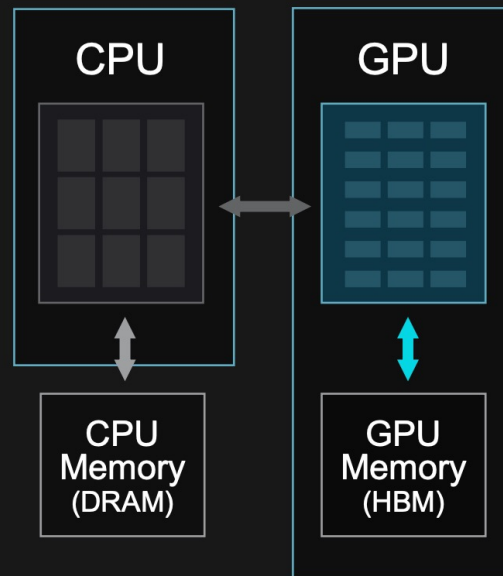
AMD CDNA™ 2 Coherent Memory Architecture



AMD CDNA™ 3 Unified Memory APU Architecture

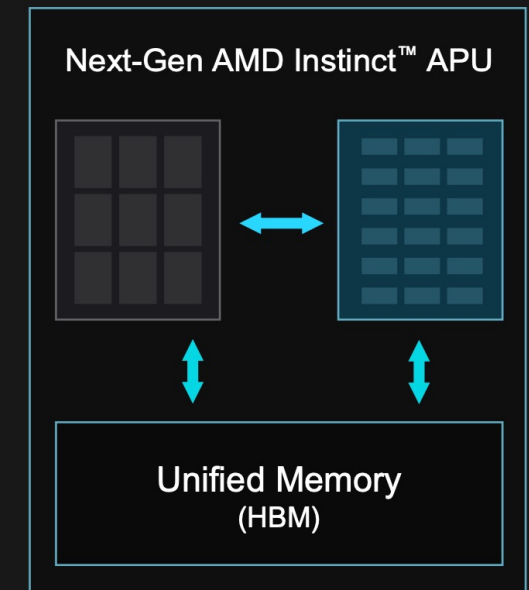
MI250 Accelerator

- Simplifies programming
- Low overhead 3rd Gen Infinity interconnect
- Industry standard modular design



MI300 Accelerator

- Eliminates redundant memory copies
- High bandwidth, low latency communication
- Low TCO with unified memory APU package



Universal Chiplet Interconnect Express (UCIe) 1.0 Specification



May 2022

Think of UCle being the PCIe but for chips

The first incarnation of UCle will be roughly four times faster than PCI-Express over the same distance

UCle has the potential to be well over 10 to 20 times faster than PCI Express

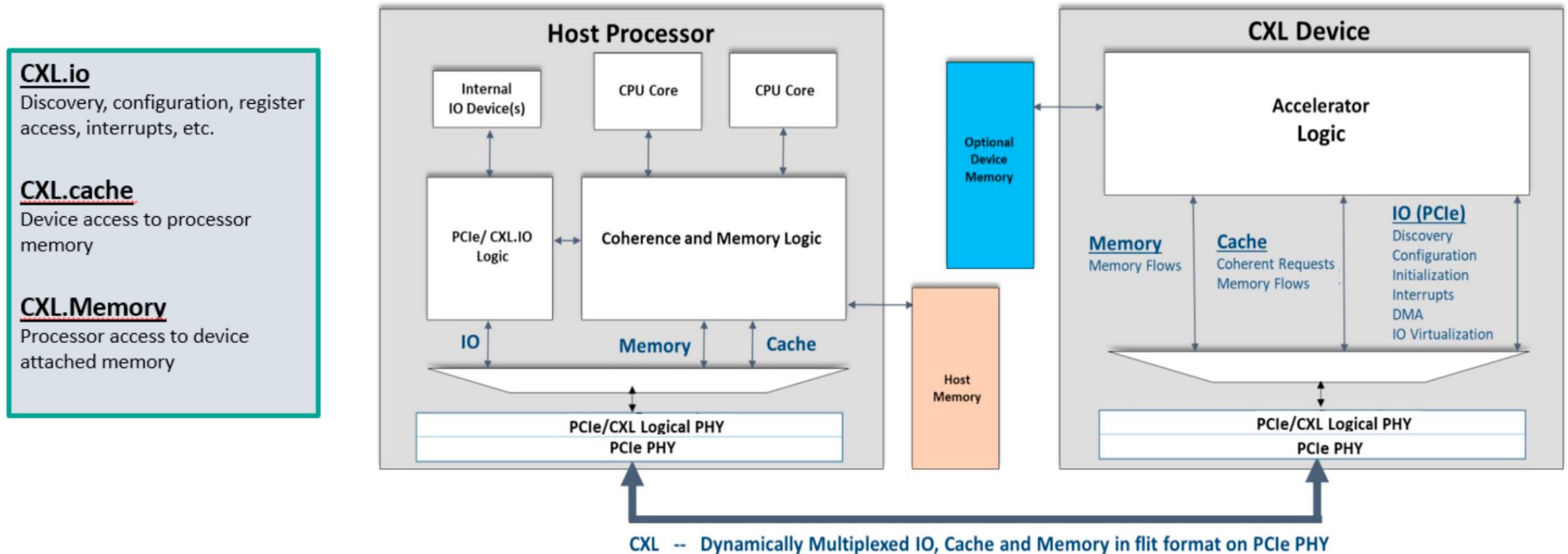
Bob Brennan, INTEL

Vice president and General Manager for customer solution engineering of Intel Foundry Services

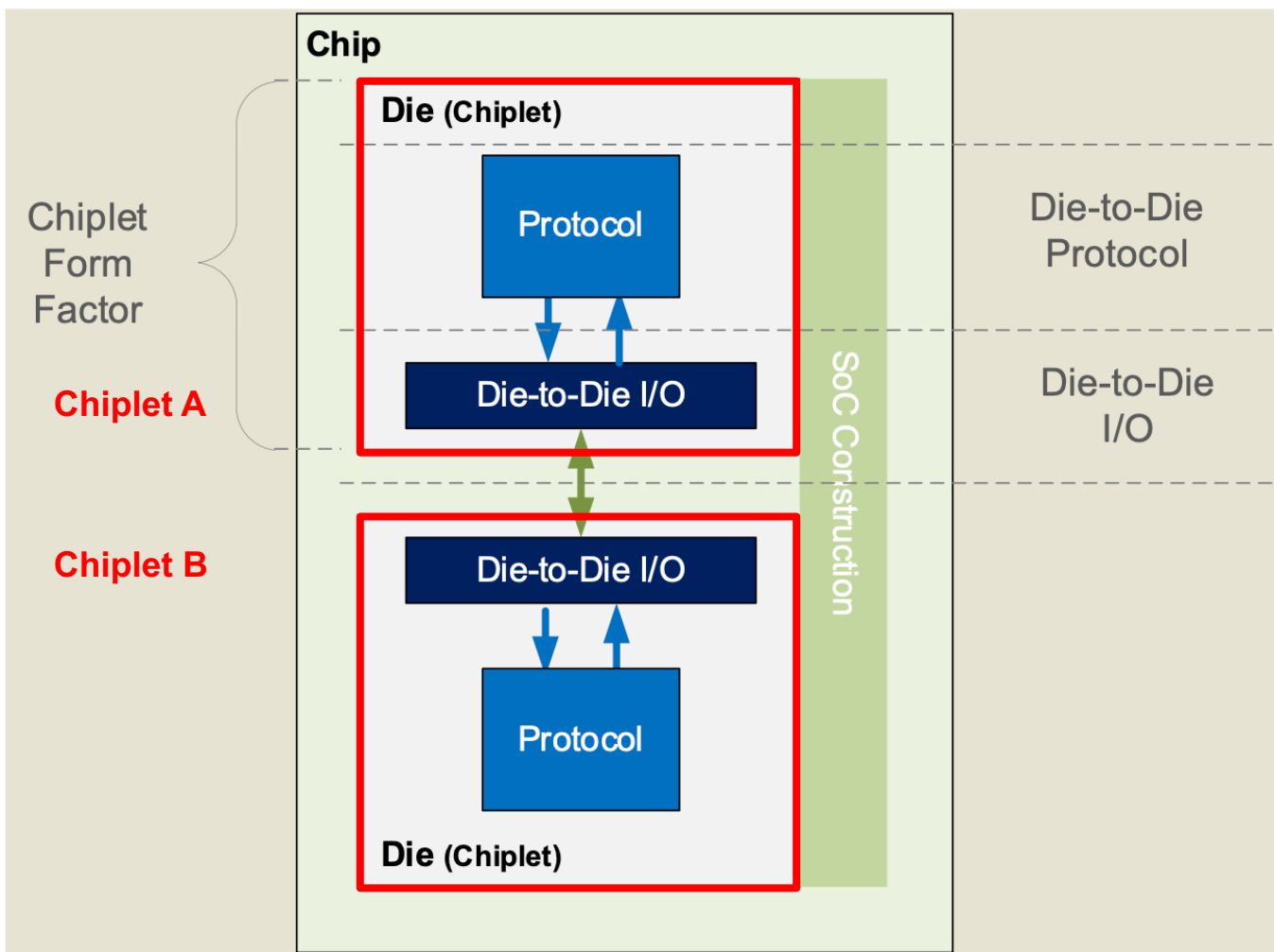


Source: <https://www.hpcwire.com/2022/05/11/intel-says-ucle-to-outpace-pcie-in-speed-race/>
May 2022

- The CXL transaction layer is comprised of three dynamically multiplexed sub-protocols on a single link:



Components of Chiplet Interoperability

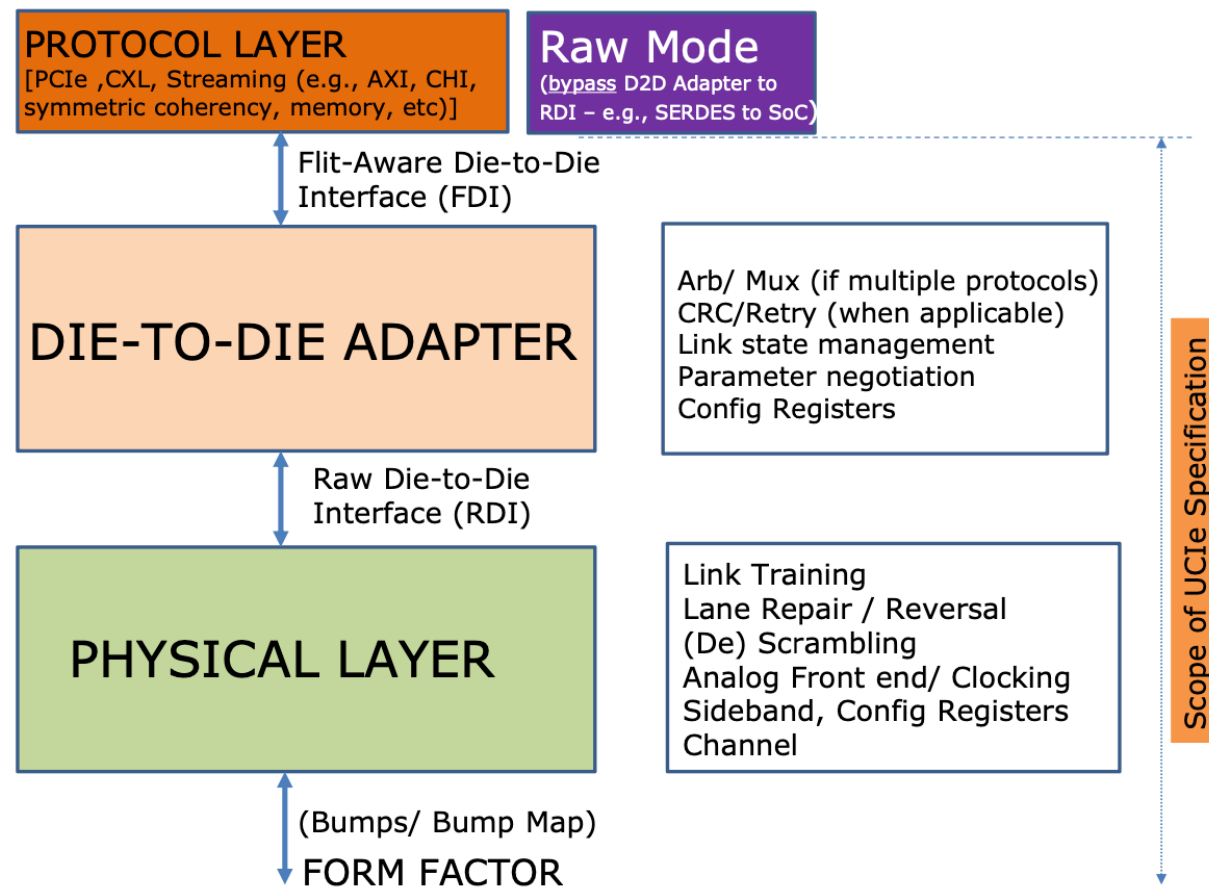


(Example SoC showing two chiplets only)

- **Chiplet Form Factor**
 - Die Size / bump location
 - Power delivery
- **SoC Construction (Application Layer)**
 - Reset and Initialization
 - Register access
 - Security
- **Die-to-Die Protocols (Data Link to Transaction Layer)**
 - PCIe/ CXL/ Streaming
 - Plug and play IPs
- **Die-to-Die I/O (Physical Layer)**
 - Electrical, bump arrangement, channel, reset, initialization, power, latency, test repair, technology transition

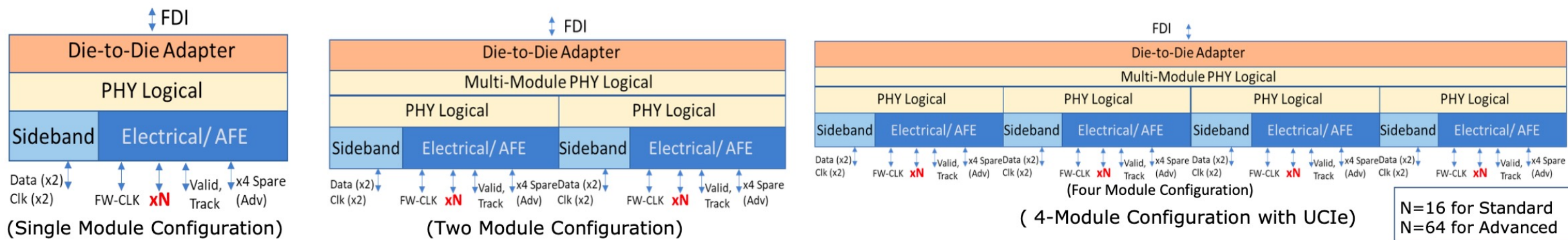
UCIe 1.0 Specification

- **Layered Approach with industry-leading KPIs**
- **Physical Layer:** Die-to-Die I/O
- **Die to Die Adapter:** Reliable delivery
 - Support for multiple protocols: bypassed in raw mode
- **Protocol:** CXL/PCIe and Streaming
 - **CXL[™]/PCIe[®] for volume attach and plug-and-play**
 - SoC construction issues are addressed w/ CXL/PCIe
 - CXL/PCIe addresses common use cases
 - I/O attach, Memory, Accelerator
 - **Streaming for other protocols**
 - Scale-up (e.g., CPU/ GP-GPU/Switch from smaller dies)
 - Protocol can be anything (e.g., AXI/CHI/SFI/CPI/ etc)
- **Well defined specification:** interoperability and future evolution
 - Configuration register for discovery and run-time
 - control and status reporting in each layer
 - transparent to existing drivers
 - Form-factor and Management
 - Compliance for interoperability
 - Plug-and-play IPs with RDI/ FDI interface



Physical Layer

- Unit is One Module: uni-directional: 1, 2, or 4 modules form a Link
 - 16 (64) SE Lanes for Std (Adv)
 - 1 SE Lane of valid
 - 1 differential pair of forwarded clock
 - 1 lane (SE) calibration - Track
 - Lane reversal on Transmit side
 - Reliability: Spare Lanes in Adv; degradation in Std
 - Supported frequencies: 4, 8, 12, 16, 24, 32 GHz
 - A component must support all data rates up to its advertised maximum data rate for interoperability
 - B/W per module/ dir: 64 GB/s Std, 256 GB/s Adv: Two module gets 2X, 4-module gets 4X
- Sideband: always on; 2 Lanes/ direction @ 800 MHz – data and clock
 - Used for training, debug, management, etc; Leverages depopulated bumps to ensure no extra shore-line
- Valid used for effective dynamic power management



UCIe 1.0: Characteristics and Key Metrics

CHARACTERISTICS	STANDARD PACKAGE	ADVANCED PACKAGE	COMMENTS
Data Rate (GT/s)	4, 8, 12, 16, 24, 32		Lower speeds must be supported -interop (e.g., 4, 8, 12 for 12G device)
Width (each cluster)	16	64	Width degradation in Standard, spare lanes in Advanced
Bump Pitch (um)	100 – 130	25 - 55	Interoperate across bump pitches in each package type across nodes
Channel Reach (mm)	<= 25	<=2	

KPIs / TARGET FOR KEY METRICS	STANDARD PACKAGE	ADVANCED PACKAGE	COMMENTS
B/W Shoreline (GB/s/mm)	28 – 224	165 – 1317	Conservatively estimated: AP: 45u; Standard: 110u; Proportionate to data rate (4G – 32G)
B/W Density (GB/s/mm ²)	22-125	188-1350	
Power Efficiency target (pJ/b)	0.5	0.25	
Low-power entry/exit latency	0.5ns <=16G, 0.5-1ns >=24G		Power savings estimated at >= 85%
Latency (Tx + Rx)	< 2ns		Includes D2D Adapter and PHY (FDI to bump and back)
Reliability (FIT)	0 < FIT (Failure In Time) << 1		FIT: #failures in a billion hours (expecting ~1E-10) w/ UCIe Flit Mode

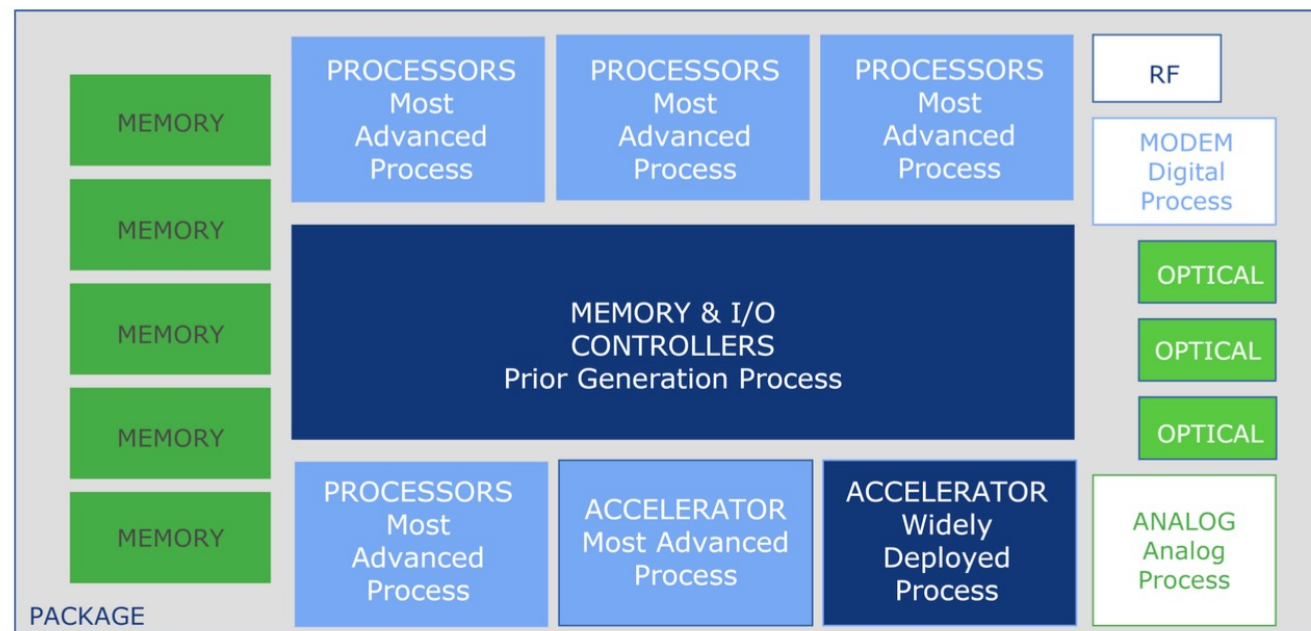
UCIe 1.0 delivers the best KPIs while meeting the projected needs for the next 5-6 years across the compute continuum.

Usage Models for UCle



Usage Models for UCIE: SoC at Package level

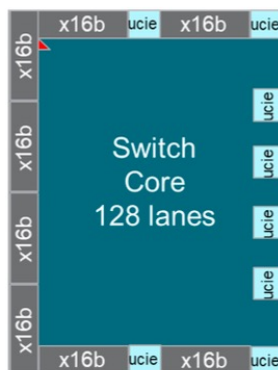
- SoC as a Package level construct
 - Standard and/ or Advanced package
 - Homogeneous and/or heterogeneous chiplets
 - Mix and match chiplets from multiple suppliers
- Across segments: Hand-held, Client, Server, Workstation, Comms, HPC, Automotive, IoT, etc
- UCIE PHY and D2D adapter common
 - PCIe/CXL protocol for plug-and-play
 - Streaming for others (similar to board level connectivity today where scale-up systems are on PCIe PHY)
 - Similar to PCIe/ CXL at board level



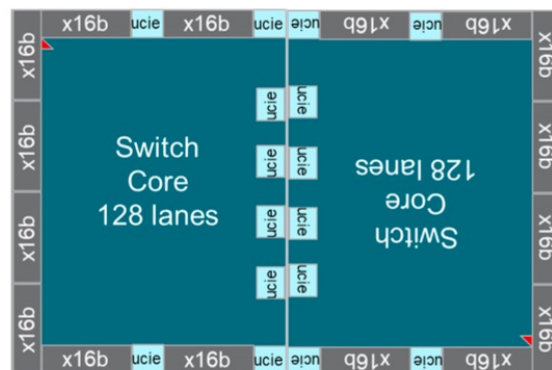
Processors: symmetric coherency protocol mapped on UCIE through FDI
Memory: CXL.Mem mapped on UCIE through FDI
Accelerators: PCIe/ CXL mapped on UCIE through FDI
Modem/ RF/ Optical: Raw mode on UCIE

Example Scale-up SoC from homogeneous dies: Large Switch with on-die protocol as streaming over UCIe

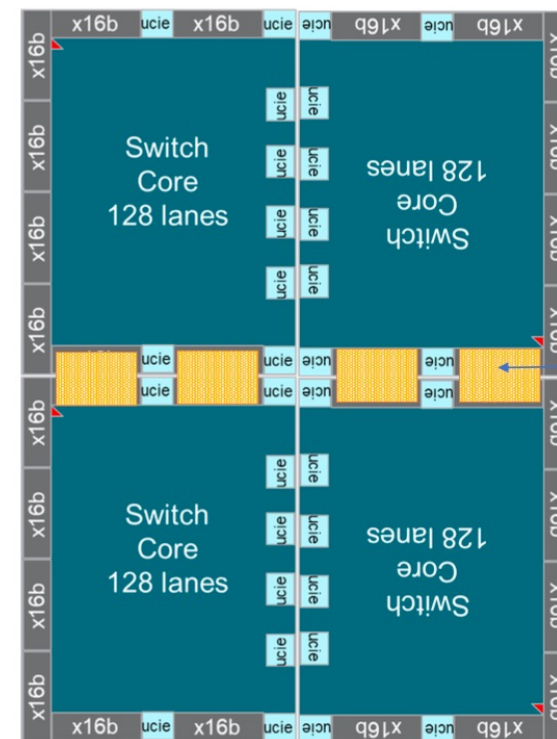
- Need large radix CXL switches – challenges: reticle limit, cost, etc.
- UCIe based Chiplets should help with scalable products
- 64G Gen6 x16b CXL links
- UCIe as d2d interconnect – while this is a scale-up CXL switch, a switch vendor may prefer to have their on-die interconnect protocol be transported over UCIe rather than create a hierarchy of switches which will not work for CXL 2.0 tree-based topology



Small CXL Switch (128 lanes)



Medium-sized CXL Switch (256 lanes)



Large CXL switch (512 lanes)

Unused x16 ports
(2 per die)

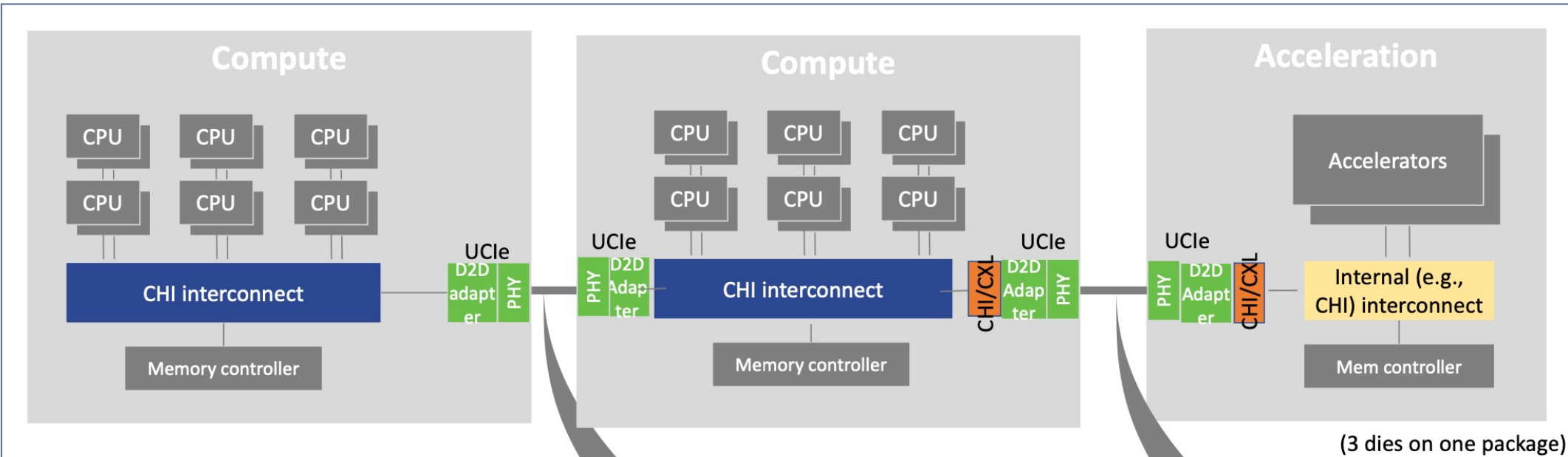
One can construct CPUs (low, medium, large core-count CPUs) from smaller dies connected through UCIe using the same principle
Here the UCIe PHY and D2D adapter will carry the packetized version of internal CPU interconnect fabric

Source: UCIe Webinar, February 21, 2023

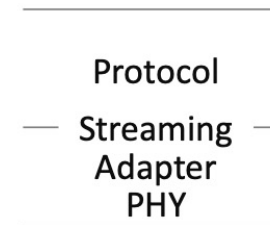
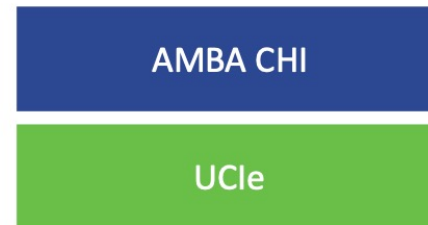
Ack: Nathan Kalyanasundaram

Example Scale-up Package using Streaming and open-plug-in using PCIe/ CXL

AMBA:
Advanced Microcontroller Bus
Architecture by ARM
CHI:
Coherent Hub Interface



- Transporting the same on-chip protocol allows seamless use of architecture specific features without protocol conversion
- Streaming interface with additional flit formats provide link robustness using UCle defined data-link CRC and retry

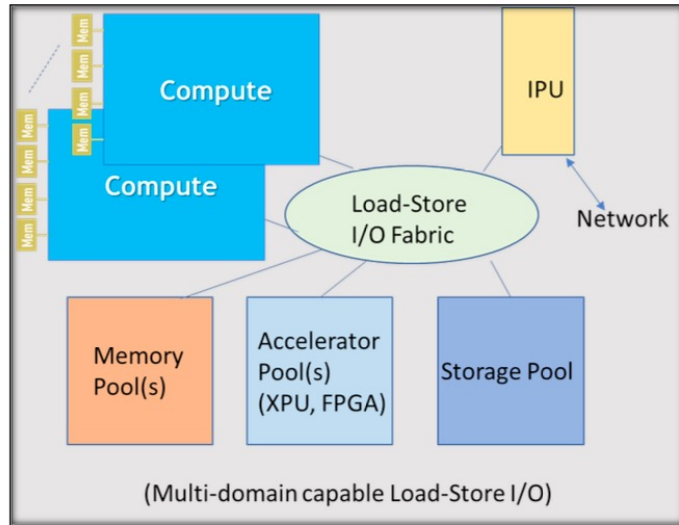


Not drawn to scale

Protocol
Streaming Adapter PHY

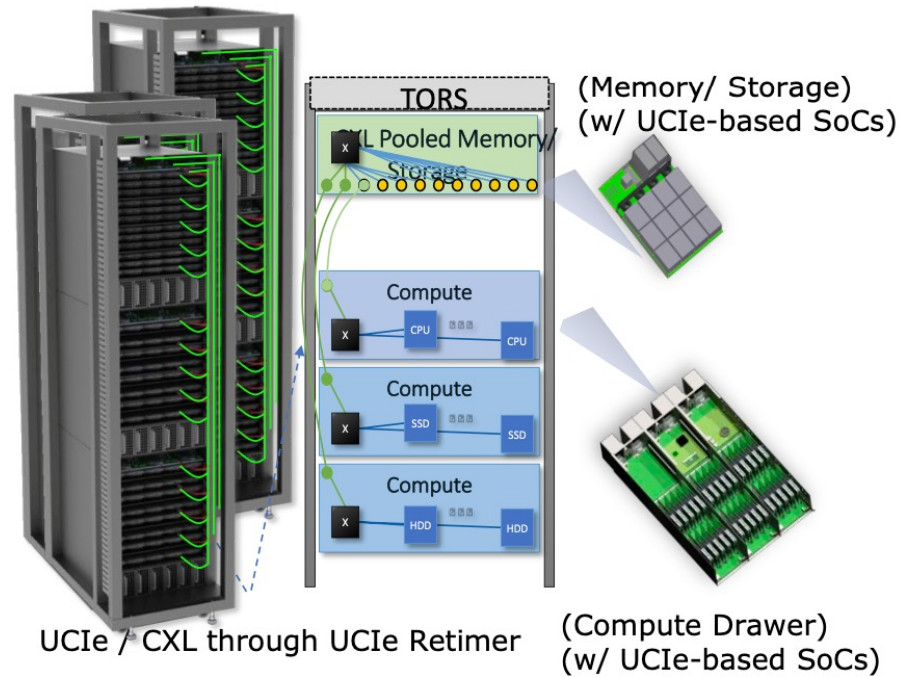
- Any device type in this open plug-in slot with CXL (or CHI if both support it)

UCIe Usage: Off-package connectivity w/ Retimers



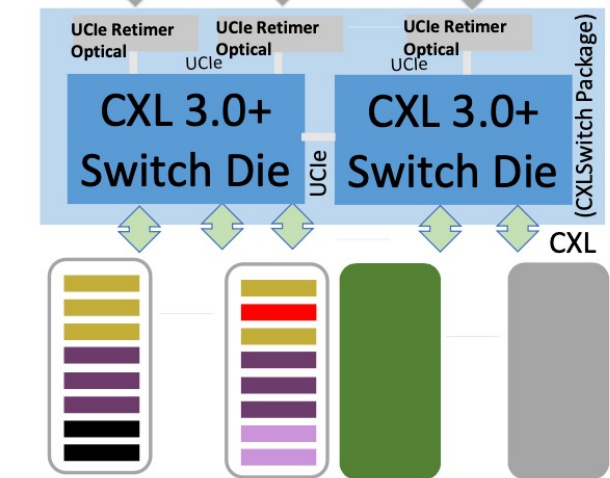
(Use Case: Load-Store I/O (CXL) as the fabric across the Pod providing low-latency and high bandwidth resource pooling/ sharing as well as message passing)

(Another example can be multi-terabit networking switches Constructed from UCIe-based co-packaged optics and partitionable networking switch dies connected through UCIe on package)



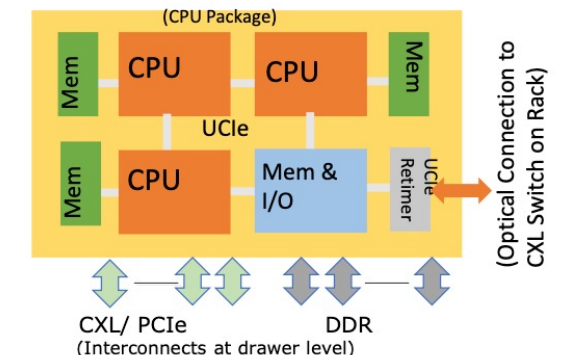
Provision to extend off-package with UCIe Retimers connecting to other media (e.g., optics)

(Optical connections: Intra-Rack and Pod)



(Pooled/ Shared Memory) (Pooled Accelerator)

(Switch dies connected through UCIe PHY + Adapter Running a proprietary switch internal protocol)

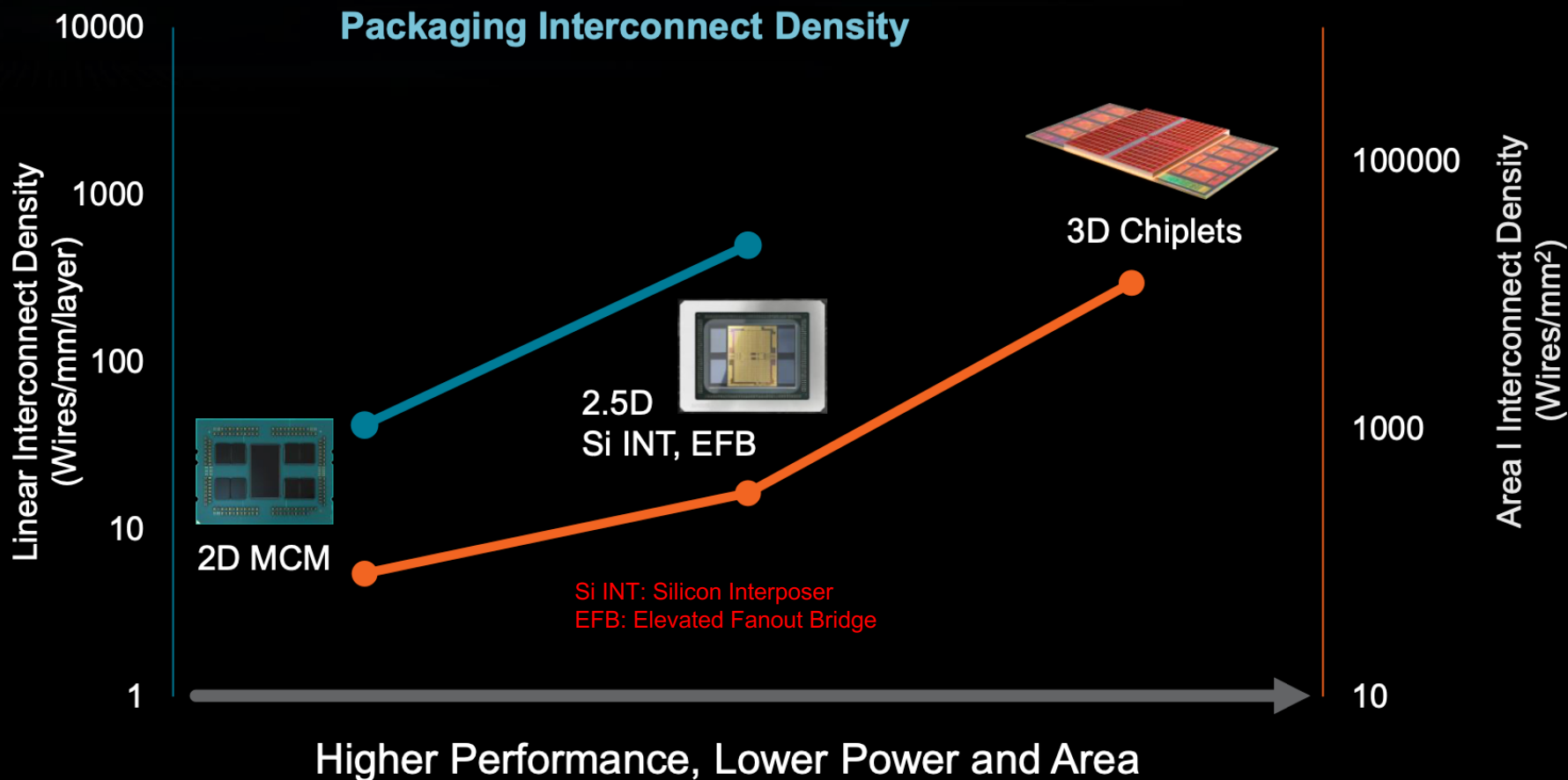


Future of Chiplets



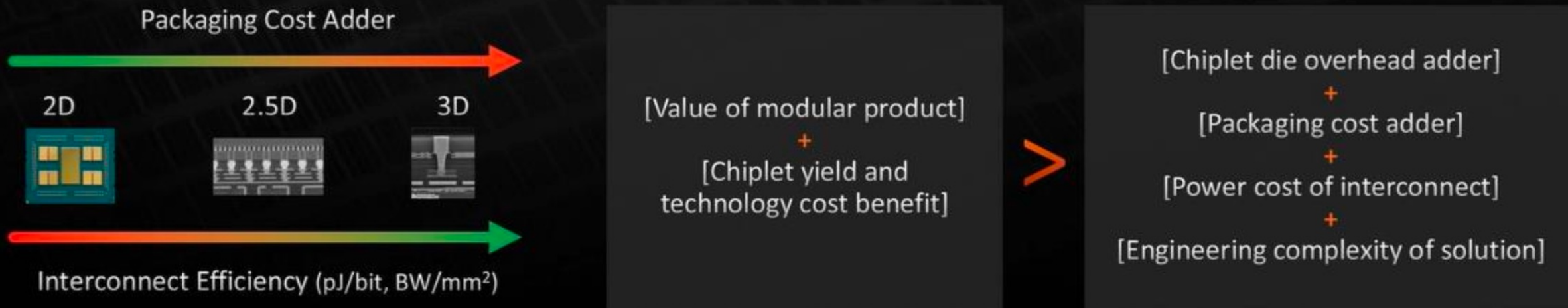
Advanced Packaging Enables Significant Gains in Performance and Efficiency

- High bandwidth between chiplets enables architectural performance gains while lowering total communication energy
- CPU and GPU integration virtually eliminates costly data transfer energy



FINDING THE OPTIMAL SOLUTION

Chiplet package architecture selection requires balancing a complex equation...

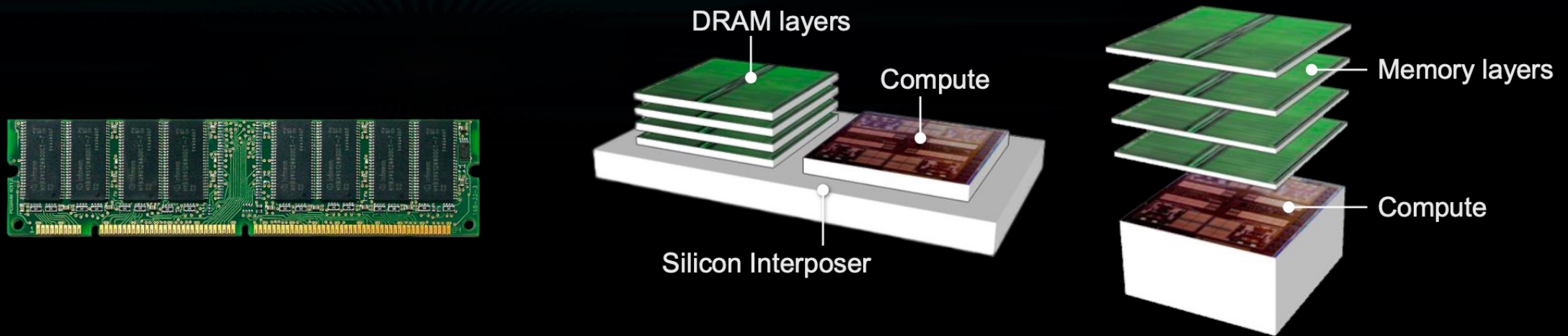


Architectural need for bandwidth, die partition options and package technology create a multi-disciplinary optimization equation

FUTURE OF 3D STACKING



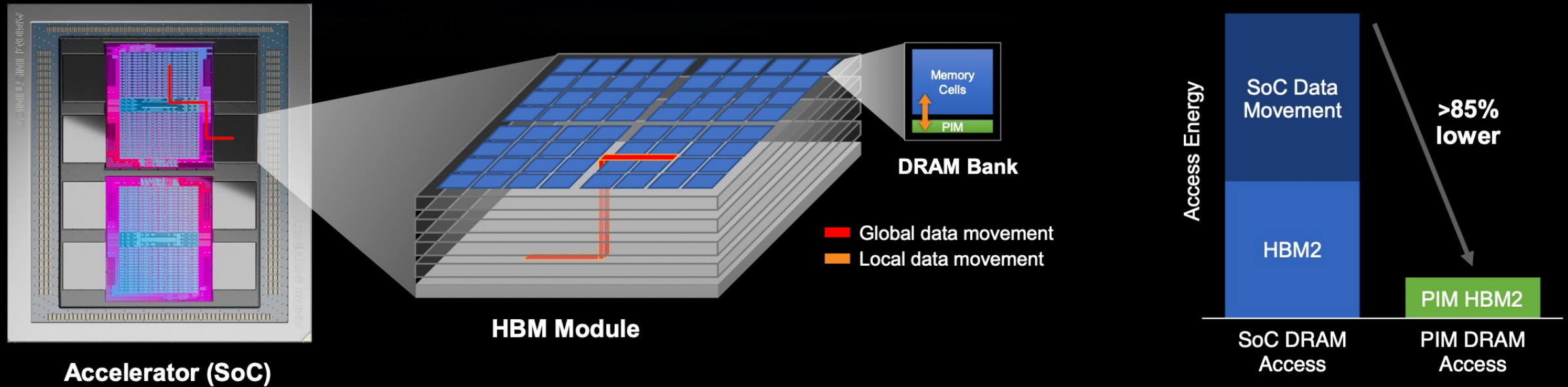
Even Tighter Integration of Compute and Memory



Integration Enables Higher Bandwidth at Lower Power

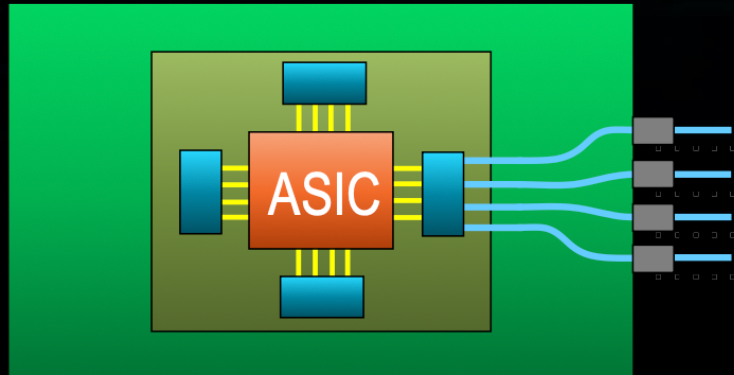
	DIMMS	2.5D Micro-bumps (HBM)	3D Hybrid Bond
pj/bit	~12	~3.5	~0.2

Processing-in-Memory

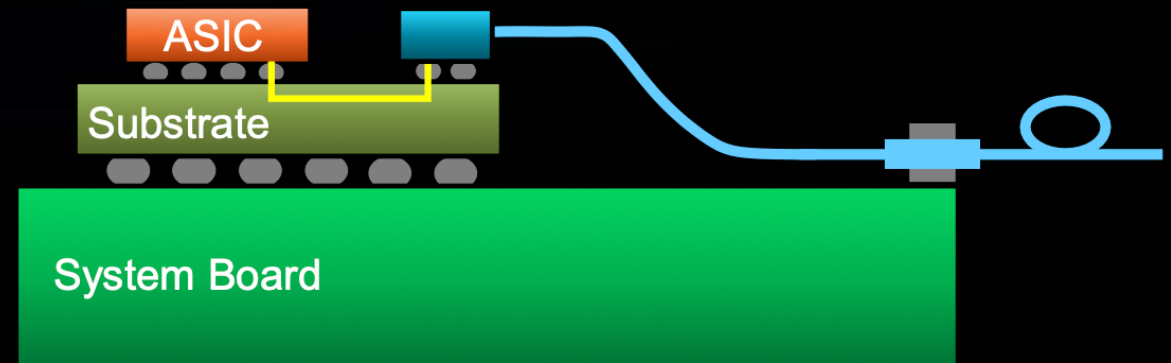


Key algorithmic kernels can be executed directly in memory, saving precious communication energy

Optical Communication for Energy Efficient Long Reach



Top View



Side View

As demonstrated in paper 12.1 of this conference, co-packaged optics provide a path forward

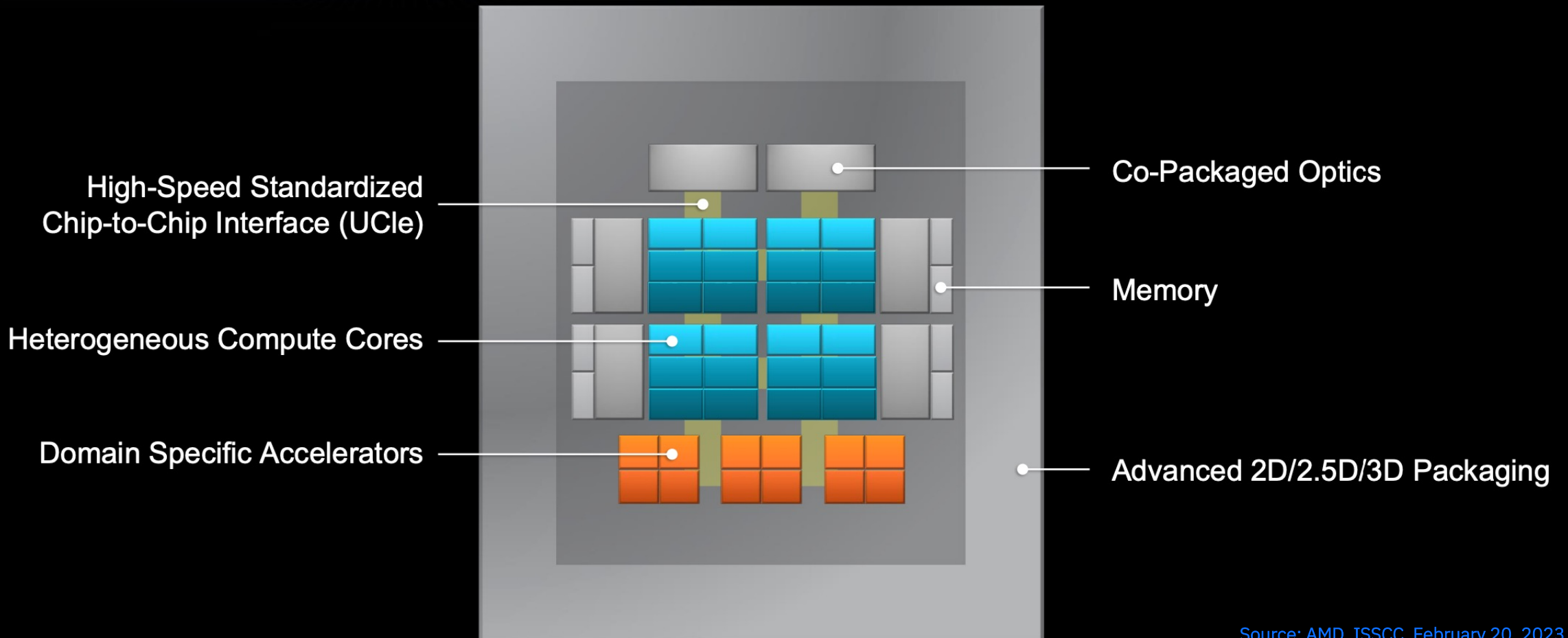
Single mode, enabling 10m up to 2km reach

Energy efficient at $< 1\text{pJ/bit}$ receive energy

Tight integration of optical transceivers to compute die is the key to efficiency

Future System-in-Package Architecture

- Advanced packaging enables maximally efficient integration of compute elements and memory
- System level communication accomplished with low-power, high-bandwidth optical



Intel Roadmap



INTEL decommitted
Rialto Bridge
March 03, 2023

Source:
<https://www.nextplatform.com/2022/06/08/the-increasingly-graphic-nature-of-intel-datacenter-compute/>
 June 8, 2022

GPU only

Next Gen Flexible Architecture
Codenamed
Falcon Shores

XPU

Significant improvement across

Performance
/ Watt

Compute
density
in x86
socket

Memory
Capacity &
Bandwidth

Manufactured with **IDM 2.0**

CPU + GPU

CPU only

GPU

x86

CPU

Under Embargo until November 9th, 2022, 6am PT

Thank You

Lecture: The Era of Chiplets and its Impact to Future System Designs

Tobias Webel (IBM)

Moore's Law, which predicts that the number of transistors on a microchip will double every 18 to 24 months, is still a guiding principle in the design and manufacture of computer hardware. While some experts have predicted that Moore's Law may eventually reach its limits due to the physical constraints of semiconductor technology, many others believe that new innovations and approaches will continue to push the boundaries of what is possible.

One innovation area which is currently heavily pursued by key Hardware companies is the so called chiplet approach. Chiplets are small, modular components that can be combined to create larger, more complex systems. They allow for greater flexibility and customization in the design of integrated circuits, and can also help to reduce costs and increase efficiency. Chiplets have become increasingly popular in recent years, and are expected to play a major role in the development of future computing systems.

Intel CEO Pat Gelsinger promised at company event that "Moore's law is alive and well," adding that "we are predicting that we will maintain or even go faster than Moore's law for the next decade.

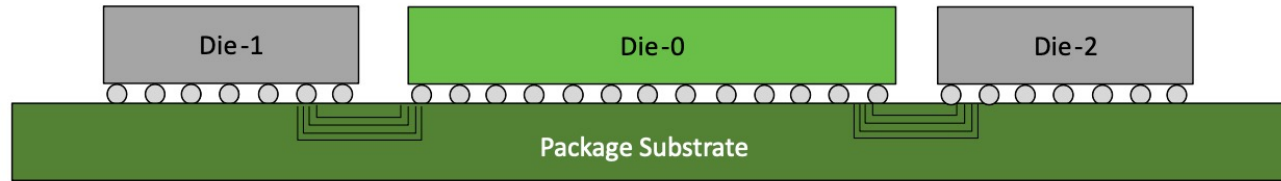
This lecture will introduce the chiplet concept, addresses opportunities and challenges of the chiplet approach and will share insights in how the industry moves forward with an openness and standardization.

Tobias Webel (IBM)



Backup

UCIe 1.0: Supports Standard and Advanced Packages

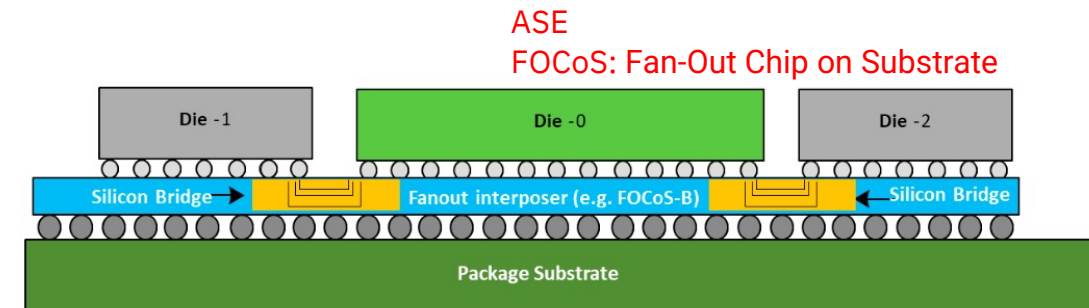
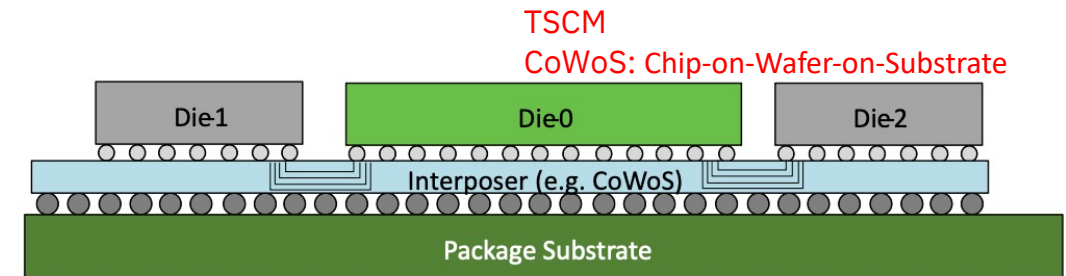
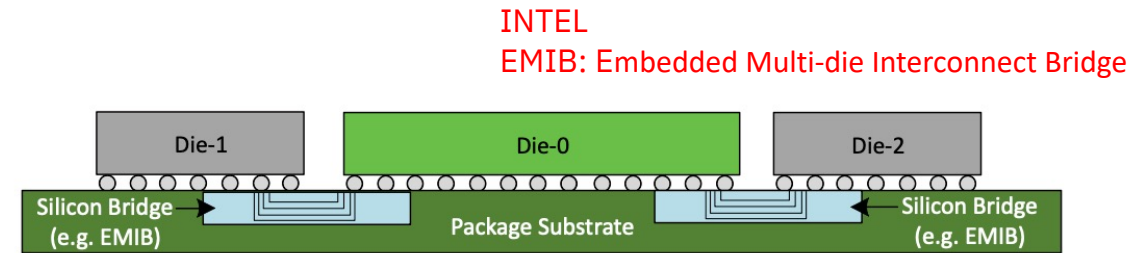


(Standard Package)

Standard Package: 2D – cost effective, longer distance

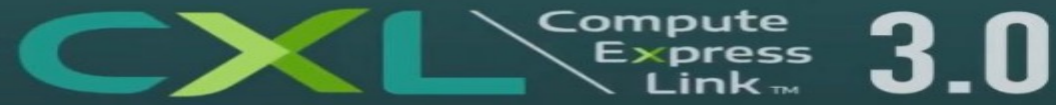
Advanced Package: 2.5D – power-efficient, high bandwidth density

Dies can be manufactured anywhere and assembled anywhere – can mix 2D and 2.5D in same package – Flexibility for SoC designer



(Multiple Advanced Package Choices)

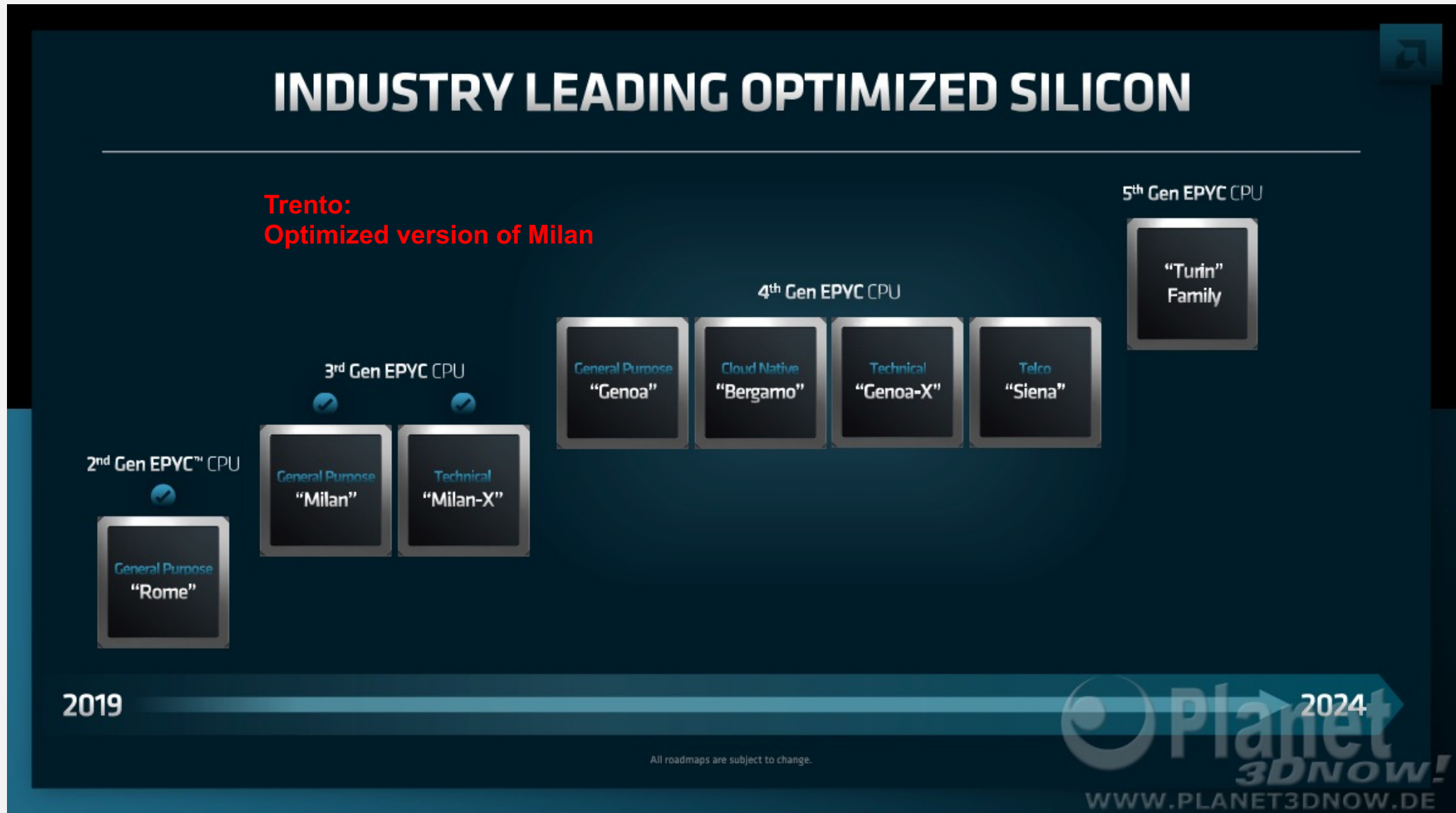
One UCIe 1.0 spec supports different flavors of packaging options to build an open ecosystem



the next generation of the CXL specification

Features	CXL 1.0 / 1.1	CXL 2.0	CXL 3.0
Release date	2019	2020	1H 2022
Max link rate	32GTs	32GTs	64GTs
Flit 68 byte (up to 32 GTs)	✓	✓	✓
Flit 256 byte (up to 64 GTs)			✓
Type 1, Type 2 and Type 3 Devices	✓	✓	✓
Memory Pooling w/ MLDs		✓	✓
Global Persistent Flush		✓	✓
CXL IDE		✓	✓
Switching (Single-level)		✓	✓
Switching (Multi-level)			✓
Direct memory access for peer-to-peer			✓
Enhanced coherency (256 byte flit)			✓
Memory sharing (256 byte flit)			✓
Multiple Type 1/Type 2 devices per root port			✓
Fabric capabilities (256 byte flit)			✓

AMD EPYC Roadmap



AMD State of the Art (CPU)

Gaming CPU

AMD Ryzen Genoa 4th Generation (7900X3D)

- GA: February 2023
- 16 cores
- 2x CCD (Core-Cache Dies) Chiplets
 - 5nm TSMC
 - 71 mm²
- 1x 3D-V Cache Chiplet
 - 5nm TSMC
 - Just one CCD gets a 3D-V Cache Chiplet (64MB)
- 1x IO Chiplet
 - 6nm TSMC
 - 122 mm²
- 2.5D + 3D Packaging
- 13 billion transistors

