35 years providing complete ASIC & COT Solutions

Chiplet Era – The Future of Semiconductor Integration

04.12.2025 **Pavel Vilk, GM, Head of Engineering**

This is Avnet ASIC

We provide a spec to mass production ASIC solutions

35+
Years of Experience

2nm
TSMC Projects

Full Turnkey Solutions

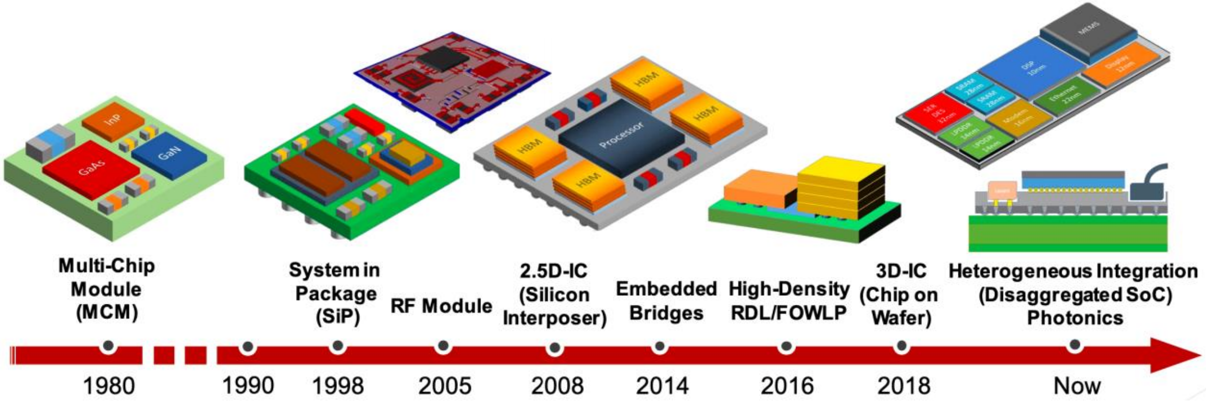
350
Successful Projects

TSMC VCA

All Disciplines Under One Roof

Division of Avnet Silica

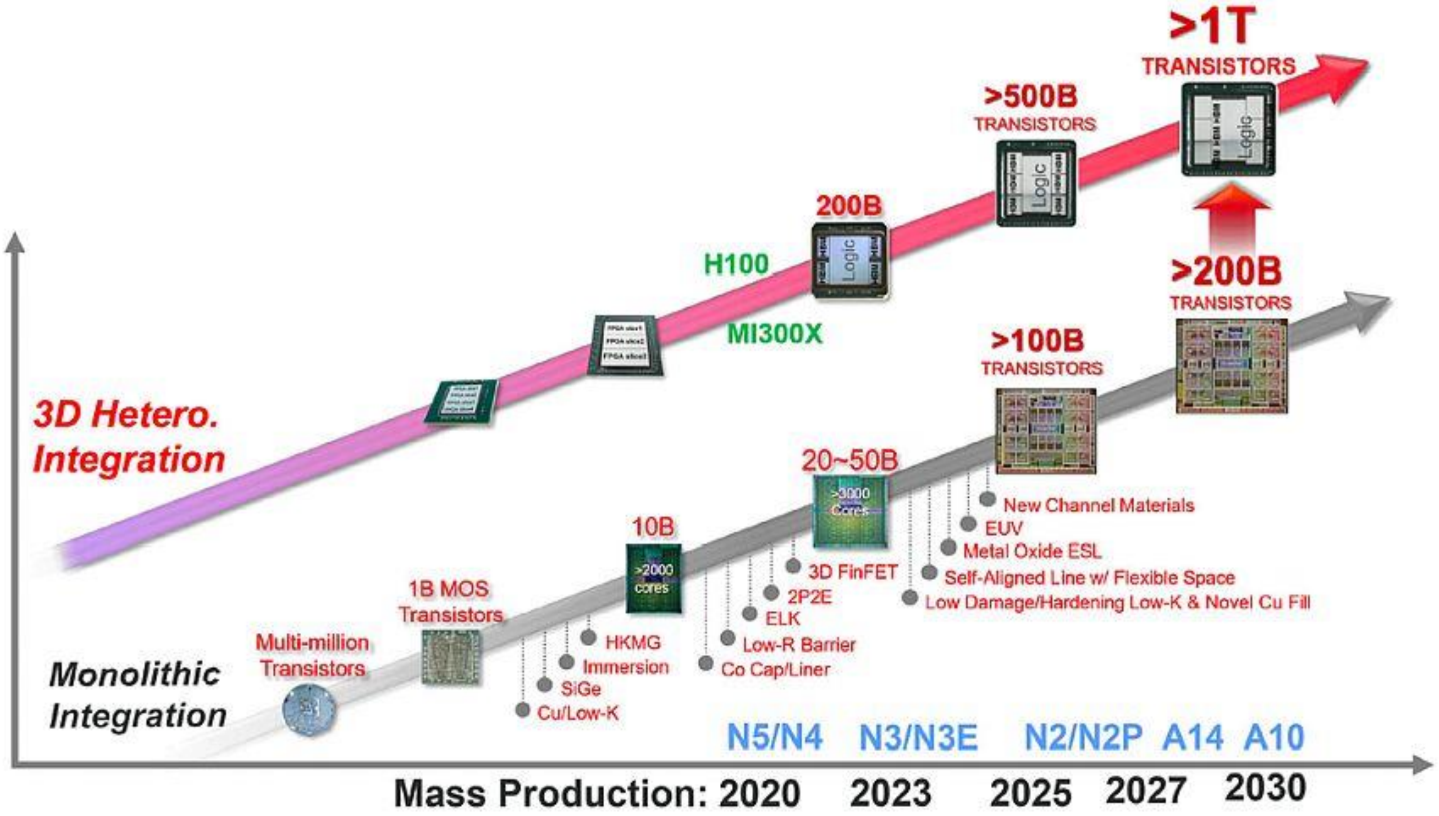
Evolution of the Semiconductor Integration



Era / Concept	*Description*
MCM (Multi-Chip Module)	*Emerged in the 1980s–90s: multiple **bare dies** placed on a substrate to work as a system. Focus was on **integration density** and reducing board area.*
SiP (System-in-Package)	*Broader and more flexible than MCM — integrates **heterogeneous dies** (logic, memory, analog, RF, sensors) in one package.*
Chiplet Architecture	*Takes SiP further by using **standardized, modular dies** that are **designed to interoperate** — not just be co-packaged. Chiplets are a **design methodology**, not just a packaging technique.*

Trends & Projections For Number Of Transistors Inside System

Single die maximum size ~830 mm

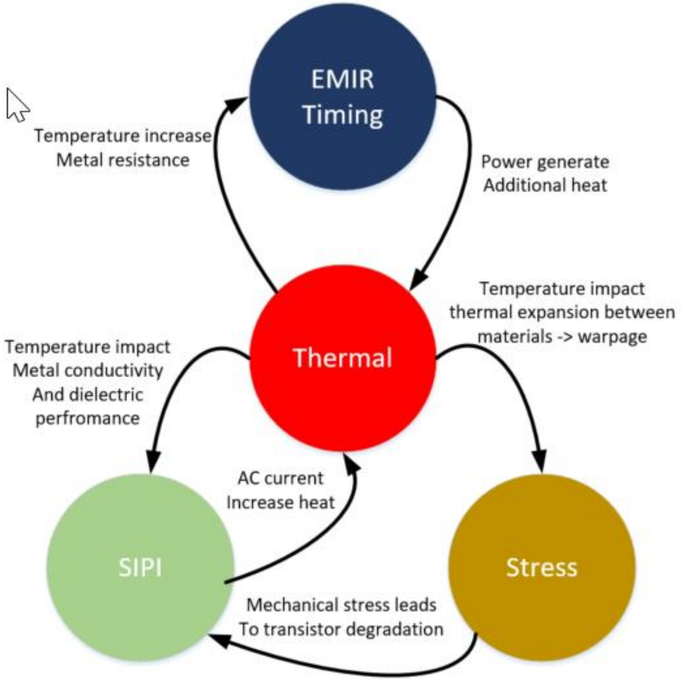


Why Chiplets?

Benefit vs monolithic die	Reasoning
Improved yield	*Smaller chips have fewer defects, improving wafer yield and reducing cost.*
Heterogeneous integration	*Combine logic, memory, analog, I/O, or accelerators built on different process nodes.*
Design reuse	*The same Chiplet can be reused across multiple products or platforms.*
Faster time-to-market	*Teams can work on different Chiplets in parallel and integrate them later.*
Scalability	*Enables modular upgrades (e.g., scale CPU/GPU/AI cores or memory bandwidth).*

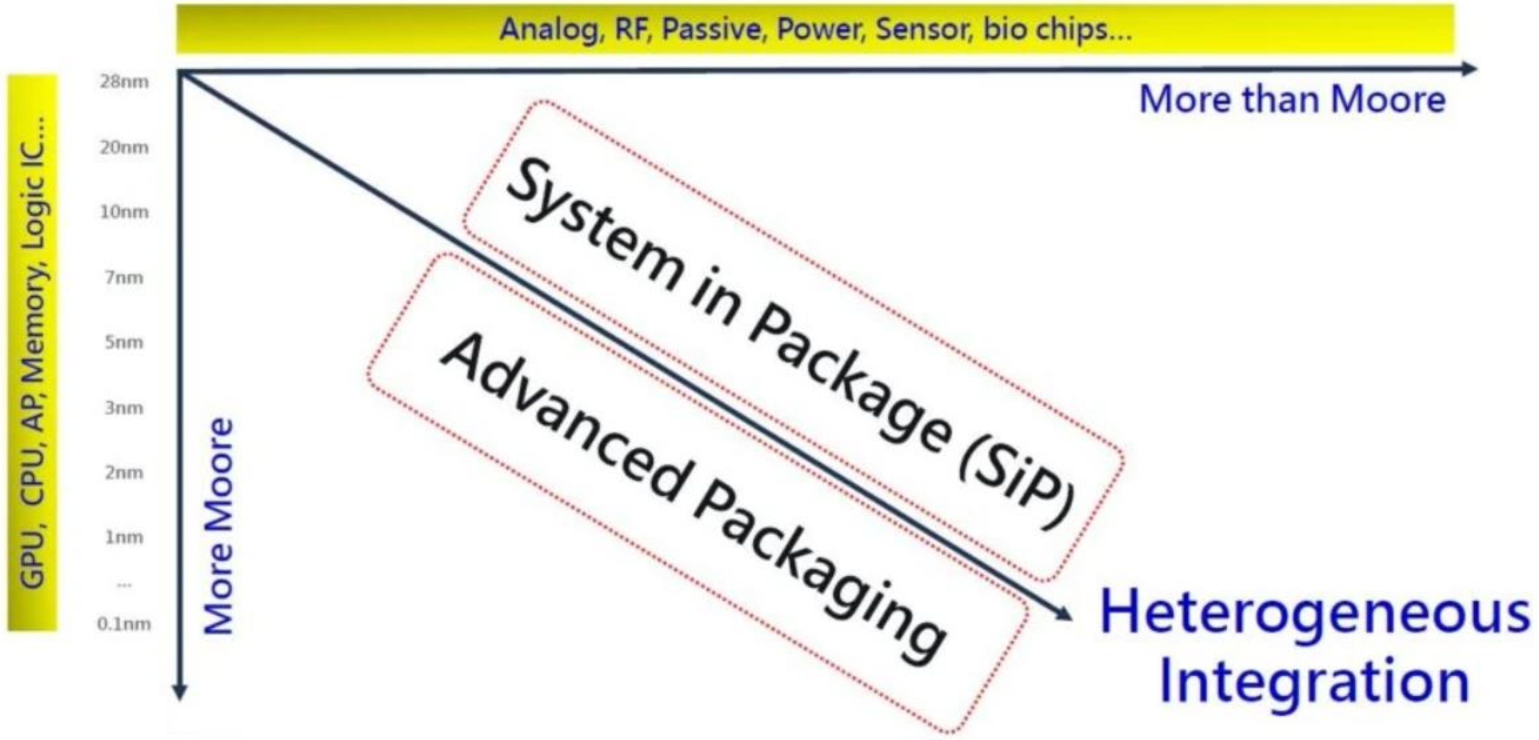
Challenges Of The Heterogenous Integration

Challenge	Root Cause
Package Warpage	Large interposer or organic substrate area with many dies
Die Shift / Misalignment	Uneven thermal expansion (CTE mismatch), bonding pressure
Stacked Die Hotspots	3D stacking or dense die proximity (logic + memory + accelerator)
Inadequate Heat Dissipation	Lack of direct contact to heat sink (middle layers in 3D)
IR Drop / PDN Collapse	Shared PDN across Chiplets, long redistribution paths, TSV resistance
Uneven Power Delivery	Different Chiplet power profiles, asymmetric loading
Lossy Die-to-Die Interfaces	Long traces, interposer metal losses, dielectric properties
Crosstalk / Skew	Dense bump patterns, insufficient shielding, variation in trace impedance



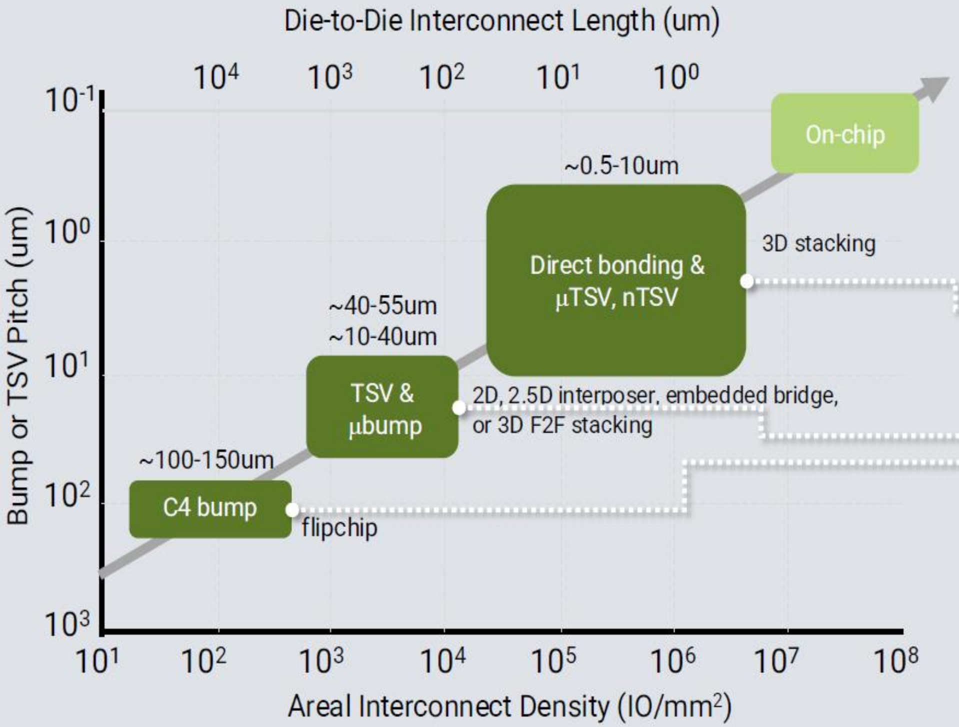
Cross impact between the factors requires delicate balancing

What is the right integration approach?

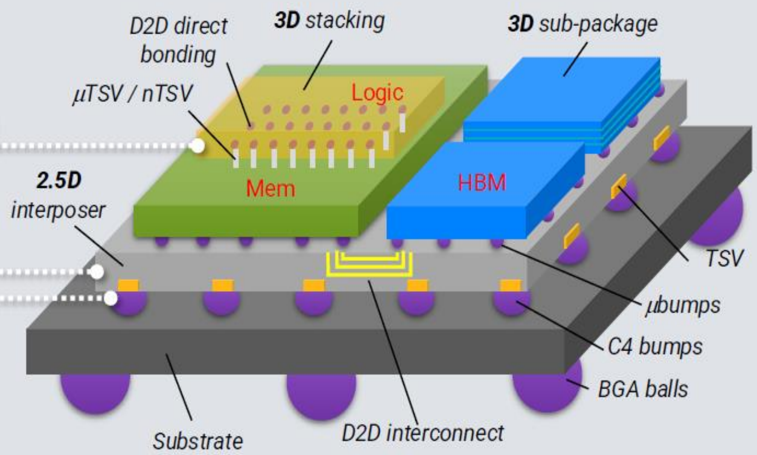


Analog, RF, Passive, Power, Sensor, bio chips...
GPU, CPU, AP, Memory, Logic IC...
28nm
20nm
10nm
7nm
5nm
3nm
2nm
1nm
...
0.1nm
More Moore
More than Moore
System in Package (SiP)
Advanced Packaging
Heterogeneous Integration

Interconnect Physical Dimensions



Multi-scale interconnects may co-exist in a package

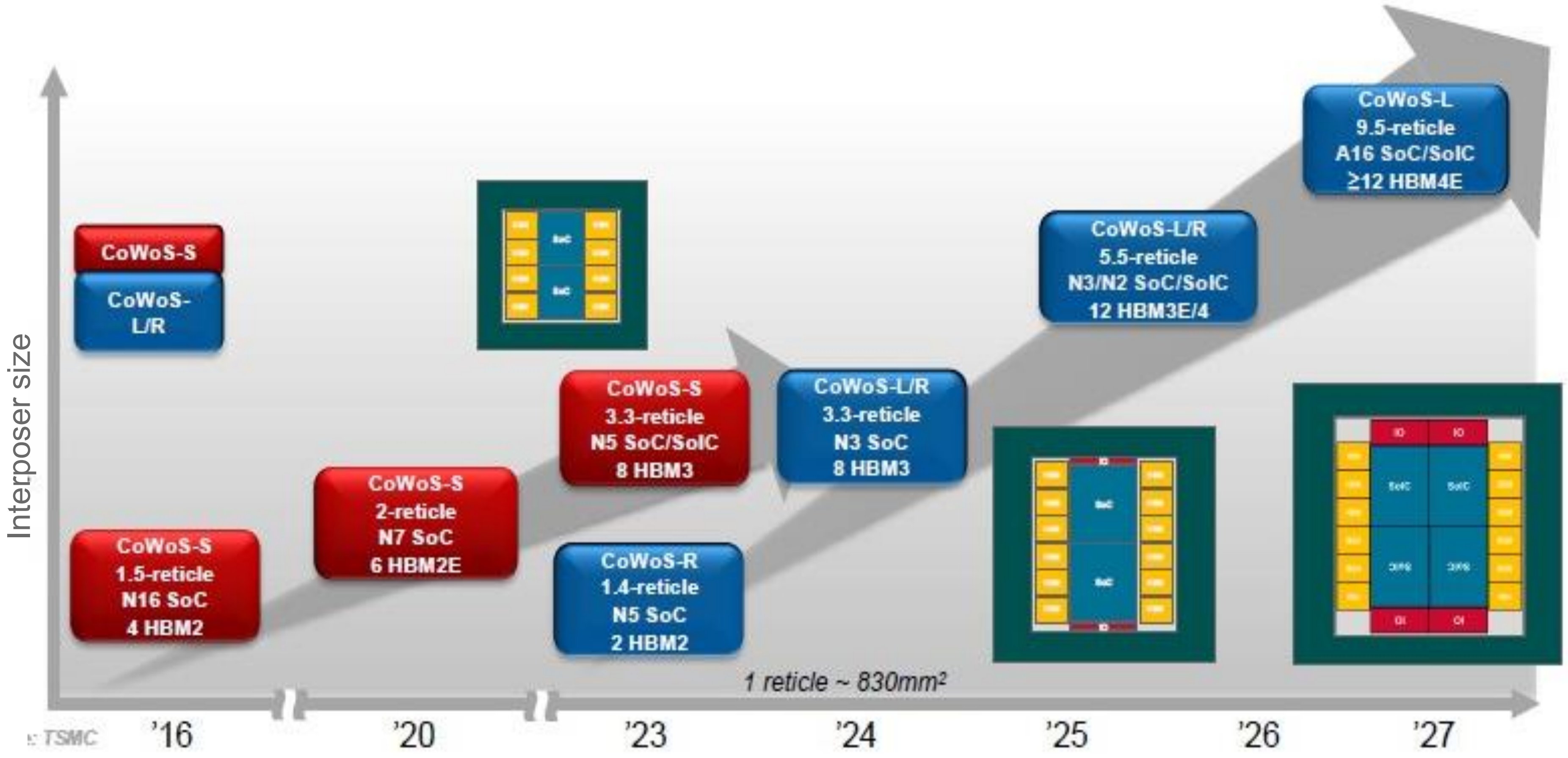


Source: Xi-Wei Lin et al., IEDM 2021

Different Connectivity Types

Parameter	Silicon Interposer (2.5D)	RDL Organic Interposer	Hybrid Interposer (Si + RDL)	Laminate Substrate	3D Stacking (Hybrid Bonding)
Bump Pitch (µm)	40–65	40–65	40–65	90–150	1.6–9
I/O Density (IO/mm^2)	High (500–1000)	High (500–1000)	High (500–1000)	Low (50–200)	Ultra High (>10,000)
Trace Width/Spacing (µm)	0.4–2 / 0.4–2	2–6 / 2–6	0.4–6 / 2–6	8–20 / 8–20	0.5–2 / 0.5–2
Energy per Transmit (pJ/bit)	Low (~0.1–0.5)	Medium (~0.5–1.0)	Medium–Low (~0.1–0.8)	Medium–High (1–3)	Very Low (~0.05–0.2)
Relative Cost	High	Medium	Medium–High	Very Low	Extra High

2.5 Interposer Evolution



CoWoS-S
1.5-reticle
N16 SoC
4 HBM2

CoWoS-S
2-reticle
N7 SoC
6 HBM2E

CoWoS-S
3.3-reticle
N5 SoC/SoIC
8 HBM3

CoWoS-R
1.4-reticle
N5 SoC
2 HBM2

CoWoS-L/R
3.3-reticle
N3 SoC
8 HBM3

CoWoS-L/R
5.5-reticle
N3/N2 SoC/SoIC
12 HBM3E/4

CoWoS-L
9.5-reticle
A16 SoC/SoIC
≥12 HBM4E

1 reticle ~ 830mm²

TSMC

From Monolithic SoC to Logic–I/O–Memory Chiplet Integration

Logic Die

CPU/GPU/AI compute tiles interconnected in a scalable mesh

D2D Interface

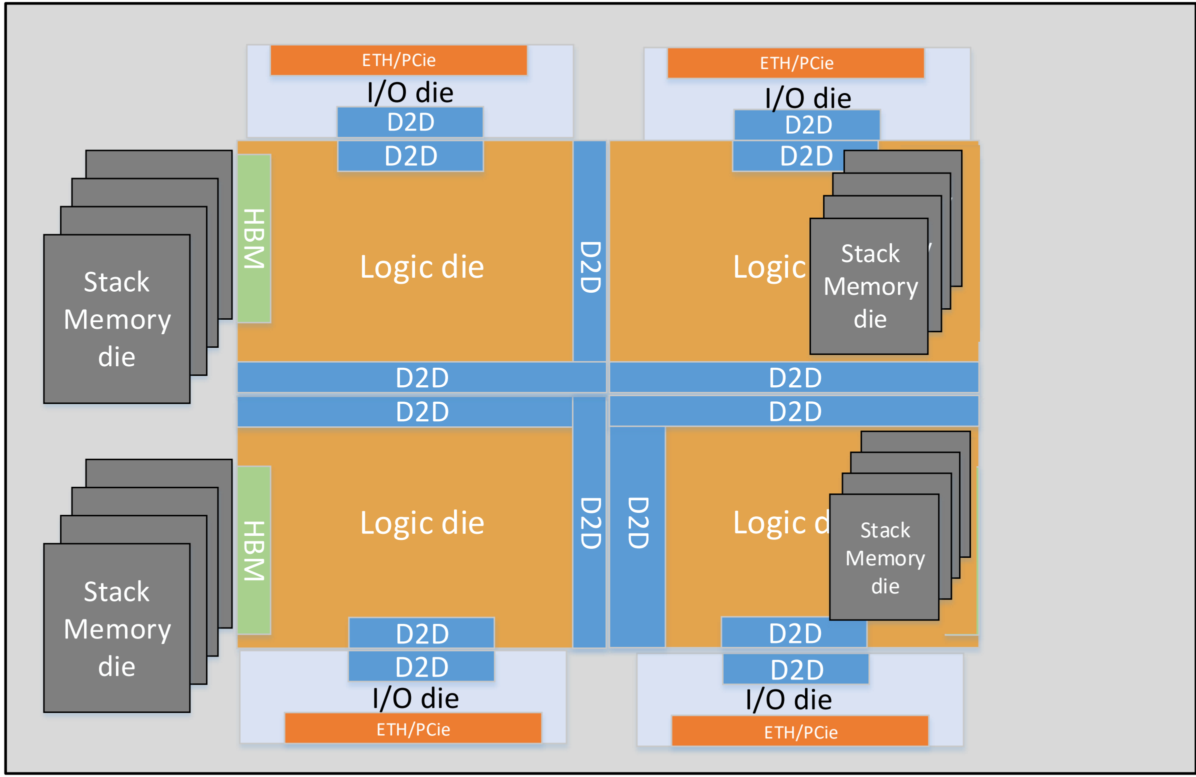
High-bandwidth, low-latency links multiplying total compute performance

I/O Die

Bridges external PCIe/Ethernet with internal Chiplet D2D

Memory Die

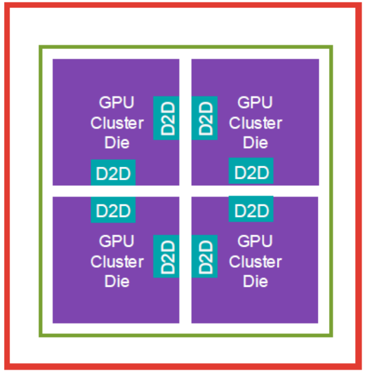
HBM2/2E/3/3E/CHBM stacks providing ultra-high bandwidth



Applications that benefit from high bandwidth Die-2-Die

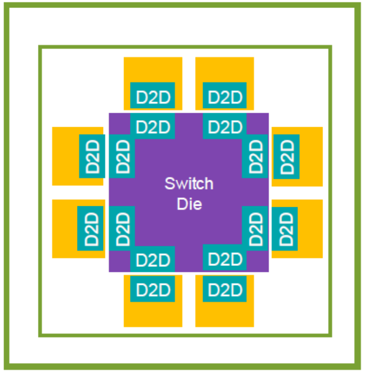
Very Large AI Training SoCs

Area & power efficiency



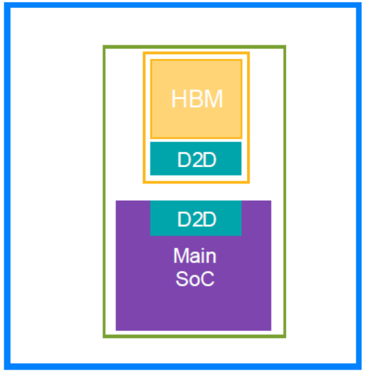
100T Switch SoC Optical & Electrical

Edge occupancy



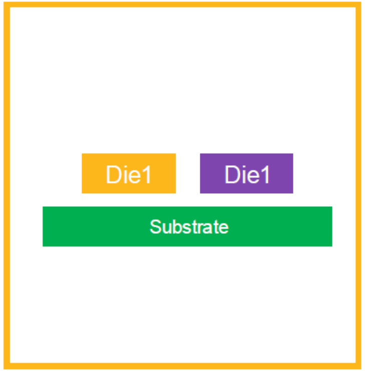
Custom HBM Base Die

Edge BW efficiency

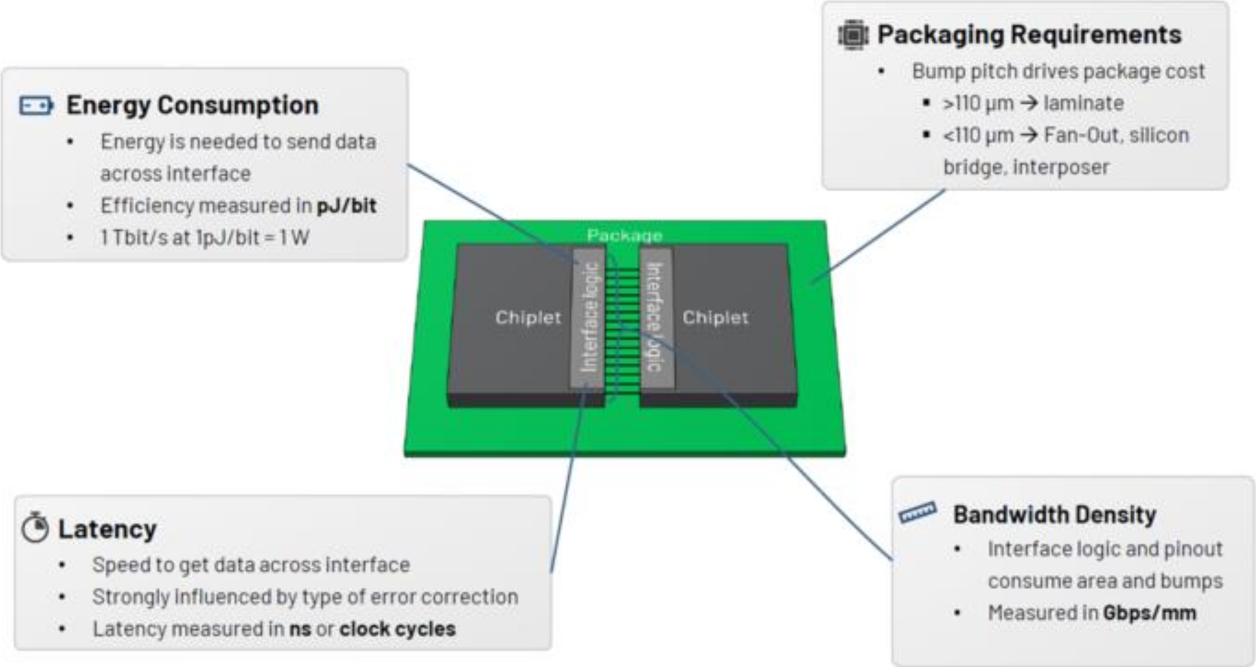


High BW D2D over Substrate

Package cost reduction



Die-to-Die Interface Selection Considerations



Energy Consumption

- Energy is needed to send data across interface
- Efficiency measured in **pJ/bit**
- 1 Tbit/s at 1pJ/bit = 1 W

Packaging Requirements

- Bump pitch drives package cost
  - >110 µm → laminate
  - <110 µm → Fan-Out, silicon bridge, interposer

Latency

- Speed to get data across interface
- Strongly influenced by type of error correction
- Latency measured in **ns** or **clock cycles**

Bandwidth Density

- Interface logic and pinout consume area and bumps
- Measured in **Gbps/mm**

Interface Protocol Comparison

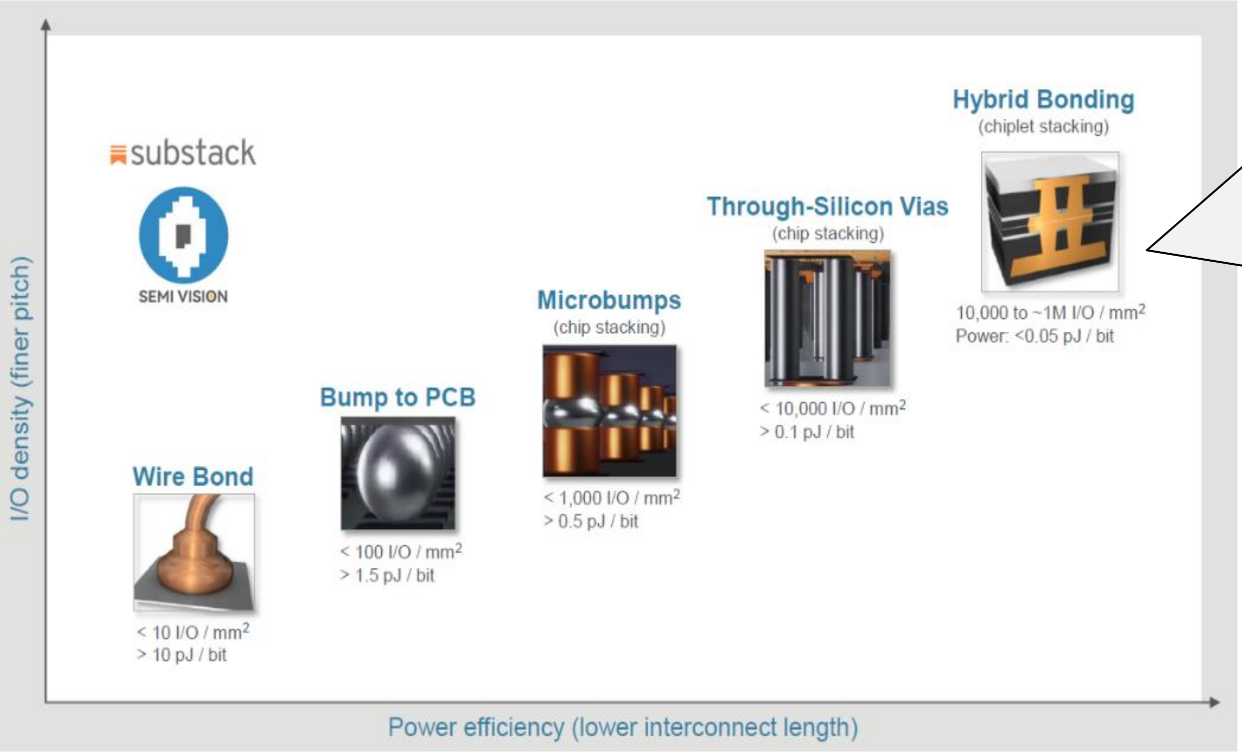
Protocol	***Bandwidth Density***	***Latency***	***Power Efficiency***	***Medium***	***Protocols Supported***
UCIe (Universal Chiplet Interconnect Express)	~1–1.3 Tbps/mm (SR mode), up to 5+ Tbps/mm with advanced HB/3D	~2–5 ns	~0.3–0.6 pJ/bit	Silicon interposer, RDL, hybrid-bonding	PCIe, CXL, Streaming/Custom
BoW (Bunch of Wires, ODSA/OCP)	0.5–1.5 Tbps/mm	~3–8 ns	~0.4–1.2 pJ/bit	Organic substrate or RDL	Streaming / custom
OpenHBI (Open High Bandwidth Interface)	1–2 Tbps/mm (HBM-like parallel PHY)	~2–3 ns	~0.3–0.5 pJ/bit	Silicon interposer	HBM-like PHY, high-throughput streaming
Proprietary D2D (NVIDIA, AMD, Apple)	2–6 Tbps/mm (varies, often > UCIe 1.0)	~1–3 ns	0.2–0.5 pJ/bit	3D stacking, hybrid bonding, silicon interposer	Custom NoC, coherence, dataflow
UALink (GPU/Accelerator Fabric)	Tens to hundreds of GB/s per link (PCIe/CXL class PHY)	~100–500 ns	Few pJ/bit–tens pJ/bit	PCB, copper cables	Fabric protocol for GPU clusters
Ethernet 224G (future AI clusters)	Multi-Tbps per port	~1000 ns+	High (tens pJ/bit)	PCB/optical	Networking fabric
On-Package Silicon Photonics (PoP optical)	~1–4 Tbps per waveguide	~5–20 ns	0.1–0.3 pJ	Integrated SiPh waveguides	Custom SerDes → optical

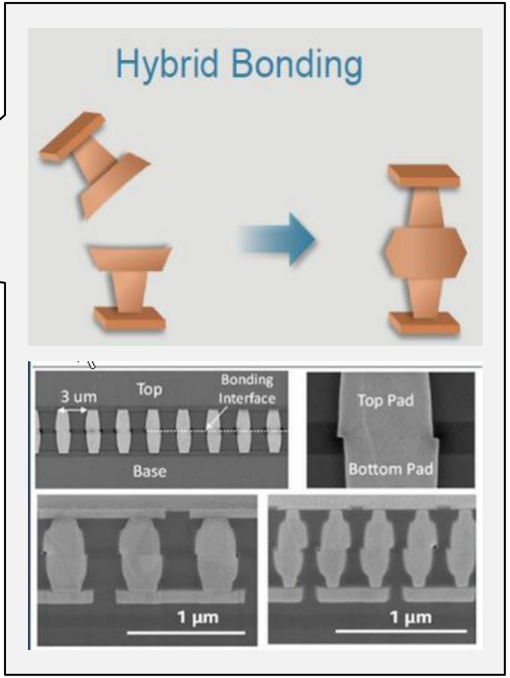
Dominant Die-2-Die is UCIE

Key Parameters	UCIe-S (2D)	UCIe-A (2.5D)	UCIe 3D
Data Rate (GT/s)	*4, 8, 12, 16, 24, 32*		*Up to 4*
Width (each cluster)	*16*	*64*	*80*
Bump Pitch (µm)	*100 – 130*	*25 – 55*	*1.6-9*
Channel Reach (mm)	*≤ 25*	*≤ 2*	*3D vertical Hybrid Bonding stacking*
BW Density (GB/s/mm²)	*22 – 125*	*188 – 1350*	*4,000 – 300,000 (4TB/s/mm² @9µm, ~12TB/s/mm² @5µm, ~35TB/s/mm²*
Power Efficiency Target (pJ/b)	*0.5*	*0.25*	*<0.05 at 9µm → 0.01 at 1µm*
Latency	*< 2ns*	*<0.5 ns*	*<0.1ns*

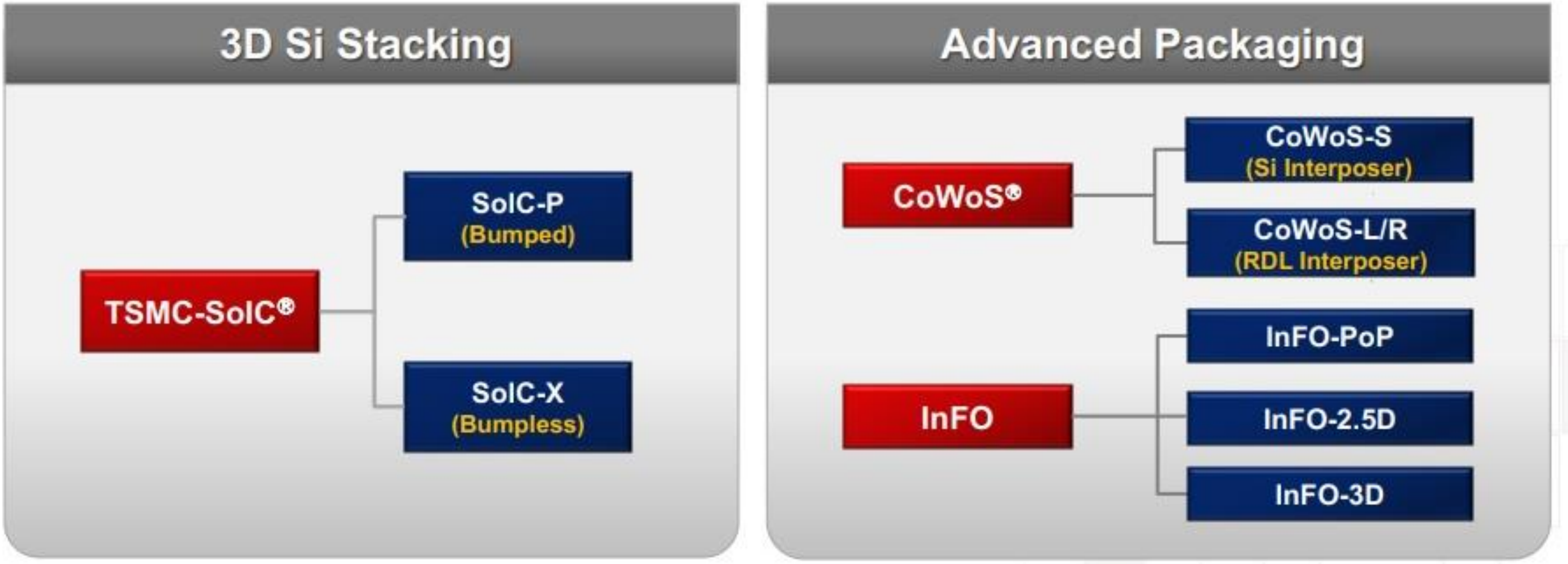
芯耀辉 AkroStar
Alibaba Group
alchip
Amkor Technology
ANALOG DEVICES
apmemory
AsteraLabs
芯云凌 Atrontech
AVERY design systems
AyarLabs
blaize
BLUE CHEETAH ANALOG DESIGN
BROADCOM
BlueOcean Smart
Brite semiconductor
cadence
cxmt
DECA
etopus
GUC The Advanced ASIC Leader
InPsytech
INNOSILICON
JCET
kiwi moore
KNIULINK
M² SEMI 超摩科技
MEDIATEK
Micron
MICROCHIP
摩尔精英 MooreElite.com
nepes
NEURON IP
openfive
oppo
proteanTecs
SOFICS
SmartDV
SPIL 矽品精密 Siliconware
希姆计算 STREAM COMPUTING
SYNOPSYS Silicon to Software
Tachyum
UniversaL Instruments
WILDER TECHNOLOGIES
VeriSilicon
XPEEDIC

The Next Generation of 3D Silicon Stacking





TSMC 3DFabric Technology Portfolio



3D Si Stacking

TSMC-SoIC®

- SoIC-P (Bumped)
- SoIC-X (Bumpless)

Advanced Packaging

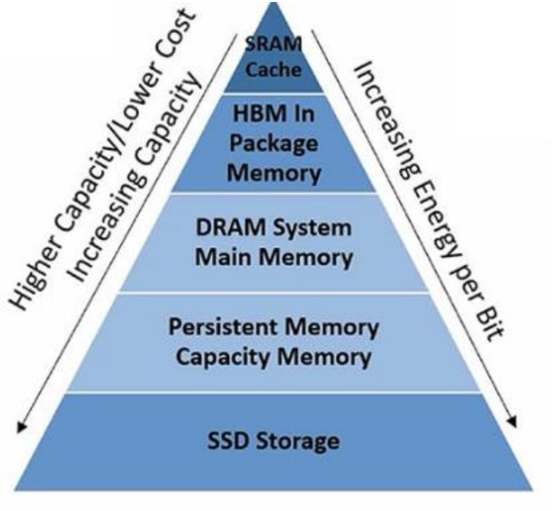
CoWoS®

- CoWoS-S (Si Interposer)
- CoWoS-L/R (RDL Interposer)

InFO

- InFO-PoP
- InFO-2.5D
- InFO-3D

Addressing Memory BW and Data Movement Power



Higher Capacity/Lower Cost

Increasing Capacity

Increasing Energy per Bit

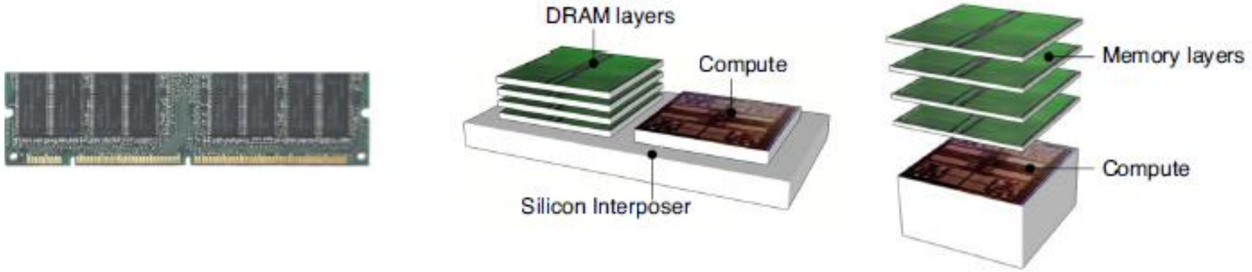
SRAM Cache

HBM In Package Memory

DRAM System Main Memory

Persistent Memory Capacity Memory

SSD Storage



Integration Enables Higher Bandwidth at Lower Power

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

Key Parameters Across Different Memory and Integration Types

Parameter	LPDDR5X on organic substrate	HBM3E on interposer	3D-stacked DRAM on logic (SoIC/Hybrid bonding)
Physical integration	Package on package (PoP) or BGA near SoC	TSV DRAM stack on silicon interposer beside logic die	DRAM die(s) vertically stacked directly on logic die
Typical distance from compute	Few mm to ~1–3 cm routing on PCB/package	Sub-mm vertical + few mm on interposer	Tens of microns (vertical)
Signaling style	Low-power parallel, reduced swing	Very wide, relatively low-swing parallel interface	Short-reach, very low-swing vertical links
Nominal data rate per pin/lane	6.4–10.7 Gb/s per pin	6.4–9.0+ Gb/s per pin	Custom buses; very high effective per-pin rate
Aggregate bandwidth (per device)	50–80 GB/s per 16–32 GB package	0.8–1.3 TB/s per stack	Tens to low hundreds of GB/s per stack (design-dependent)
Latency (relative)	High but better than DDR DIMM	Medium (better than LPDDR/DDR, still off-die)	Lowest (best)
Energy per bit (pJ/bit, order of mag)	~5–10 pJ/bit	~2–4 pJ/bit	~0.5–1.5 pJ/bit
Capacity per device	8–32 GB per package (higher with 3D stacks)	24–36 GB per stack (HBM3E)	Low: typically, 1–8 GB per stack today
Scalability (bandwidth density)	Moderate: narrow bus (64-bit), many channels needed	Excellent: TB/s-class from few stacks	Very high locally; limited total cross-section
Cost per bit	Very good (near DDR5)	Higher than DDR/LPDDR, but good per bandwidth	Very high per bit (custom 3D, low volume)
Integration complexity	Low–medium (mobile packaging)	High (2.5D, TSV, interposer, advanced package)	Very high (3D integration, hybrid bonding, yield compounding)
Main use cases	Mobile SoCs, edge AI, low-power servers	AI training/inference accelerators, high-end GPUs, networking ASICs	On-package cache/near memory, latency-critical accelerators

Key Presentation Takeaways

- Shift from monolithic dies to heterogeneous Chiplets
- Chiplets enable best-fit integration of logic, I/O, memory, and accelerators
- Advanced 2.5D/3D packaging boosts bandwidth, performance, and scalability
- Multiple Die-to-Die interfaces exist to match different system needs

and ...

some exciting news before we finish

EnICS – AVNET Innovation Center for Maximum Impact

EAIC

ENICS – AVNET

INNOVATION CENTER

Start Operation in April, 2026

ΛVNET
Reach Further

EAIC as a Driver of Innovation in Chiplet and Beyond

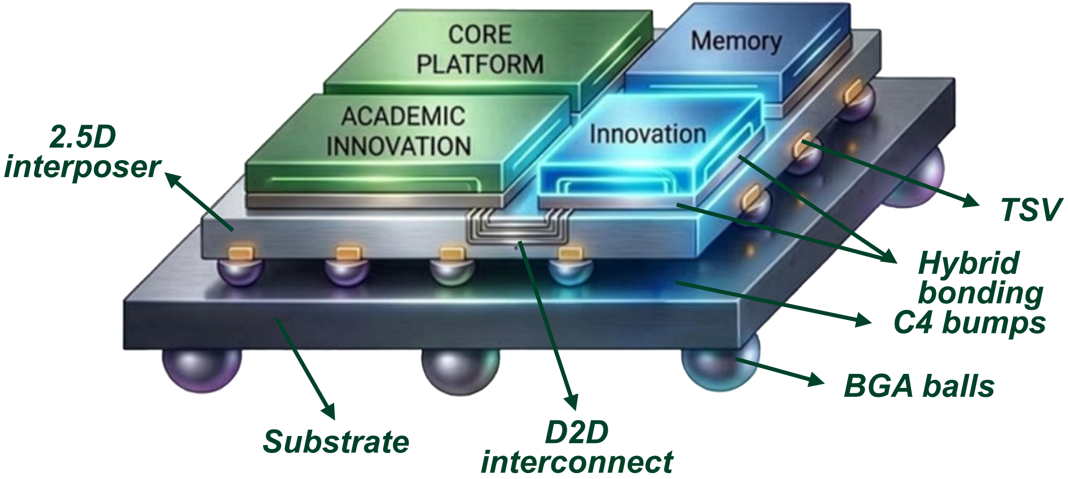
Development of a Platform to support next-generation 2.5D/3D integration solutions in TSMC technology

A collaborative environment where Avnet engineers and Bar-Ilan researchers work side by side

Operation as **a physical hub within Avnet ASIC's** development facility

Combination of Avnet's advanced node and packaging **experience** with unique academic **innovation**

Bridging academic innovation to future industry needs



ENICS - AVNET INNOVATION CENTER

We Know The Way
We Are AVNET

Thank you