AI HARDWARE SUMMIT

Counterintuitive Intelligence: Doing More to Cost Less A TCO Story

SEPTEMBER 17, 2019

eSilicon®

Collaborate. Differentiate. Win.

Carlos Macián, Ph.D.

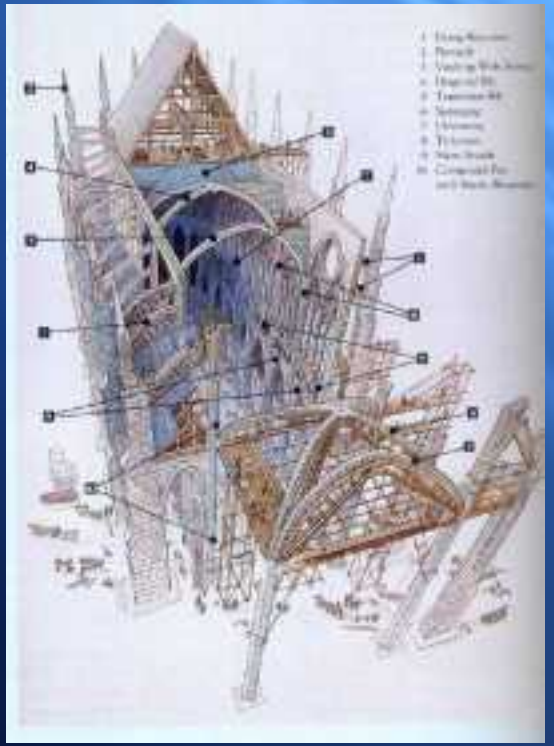
SR DIRECTOR AI STRATEGY & PRODUCTS

Counterintuitive Intelligence: A TCO Story

Agenda

- Of architects and masons
- AI challenges in the data center
- Maturing enabling technologies
- 3D IC applied to memory overlays
- Conclusions

Of Architects and Masons

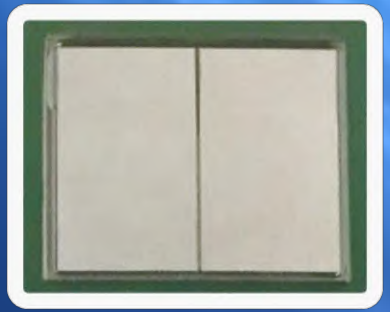


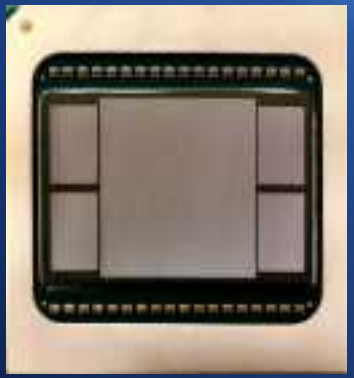
Of Architects and Masons

AI in the Data Center

AI challenges are networking/HPC's déjà vu all over again…

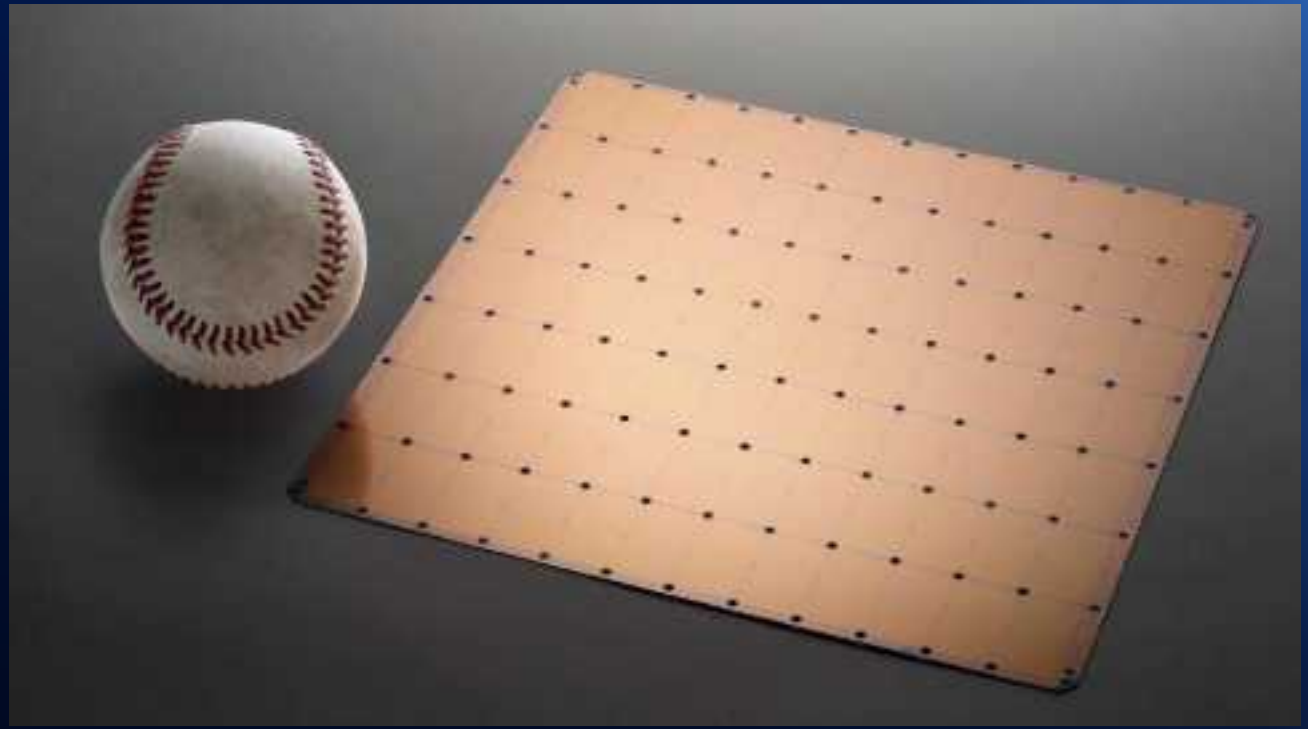
- Real estate and yield
- Memory size, latency and bandwidth
- Chip-to-chip interface density and latency
- Power and energy efficiency
- … times hyperscale!

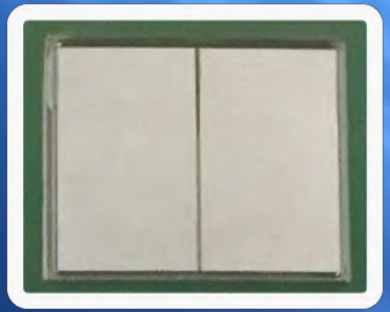


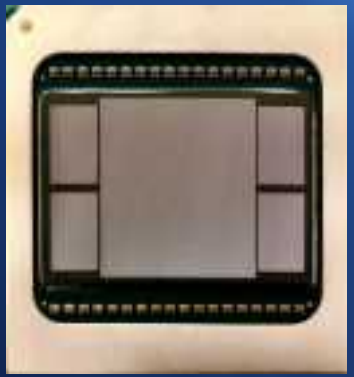


AI in the Data Center

AI challenges are networking/HPC's déjà vu all over again…

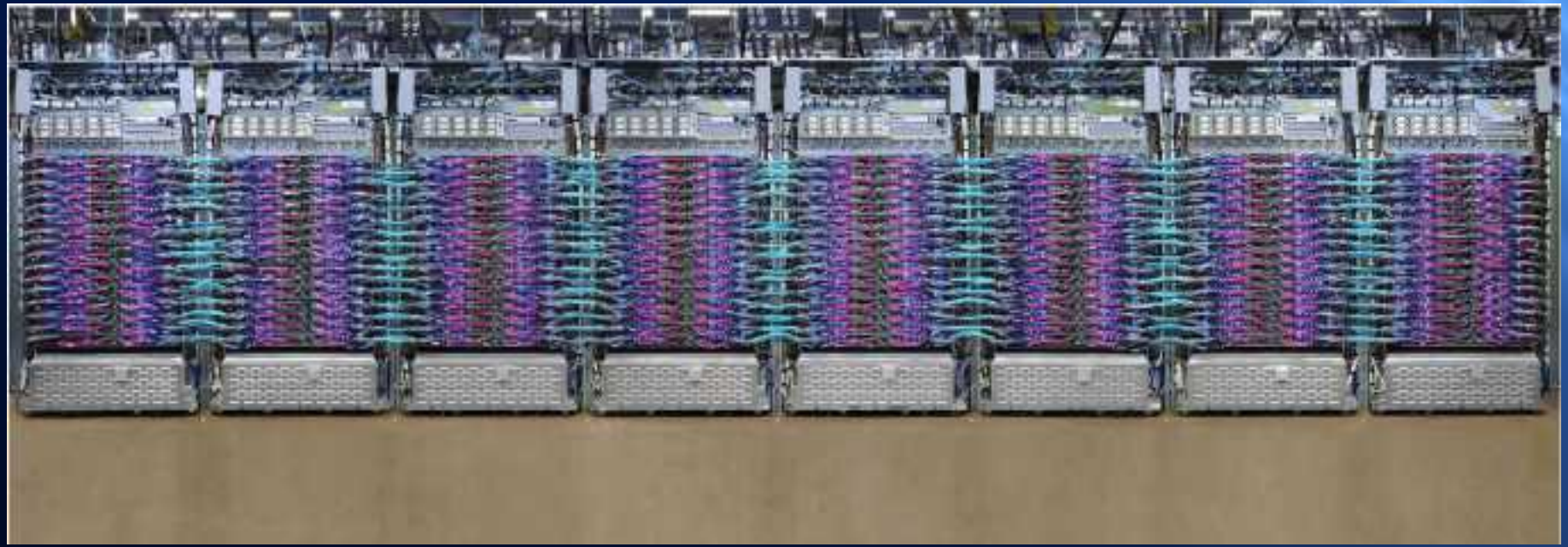






AI in the Data Center

AI challenges are networking/HPC's déjà vu all over again…



Challenges vs. Enabling Technologies

Challenges	Enabling technologies
Real estate	Multi-die, chiplets, 3D stacks
Memory size, latency and bandwidth	3D memory overlays New memory technologies
Chip-to-chip (C2C) and die-to-die (D2D) interface density and latency	XSR on substrate AIB/HBI+ (high-bandwidth interface) on EMIB/HDMI
Power and energy efficiency	In- and near-memory computing

ASIC, Enabling Tech and Hyperscale: A TCO story

Targeted performance with energy efficiency

Push the Pareto curve outwards

Avalanche effect on TCO

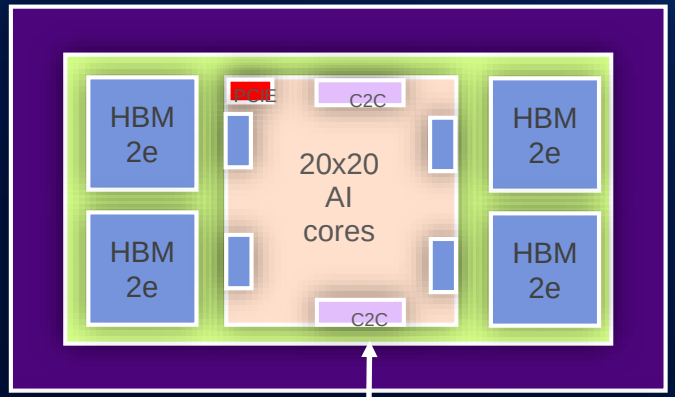
A TCO Story

How is this going to cost less?

- Higher yield
- Heterogenous processes
- Cheaper packaging, assembly and manufacturing
- Minimize overhead
- Higher energy efficiency at all levels of the system

Build highly efficient, cost-effective ASICs to optimize your AI architecture

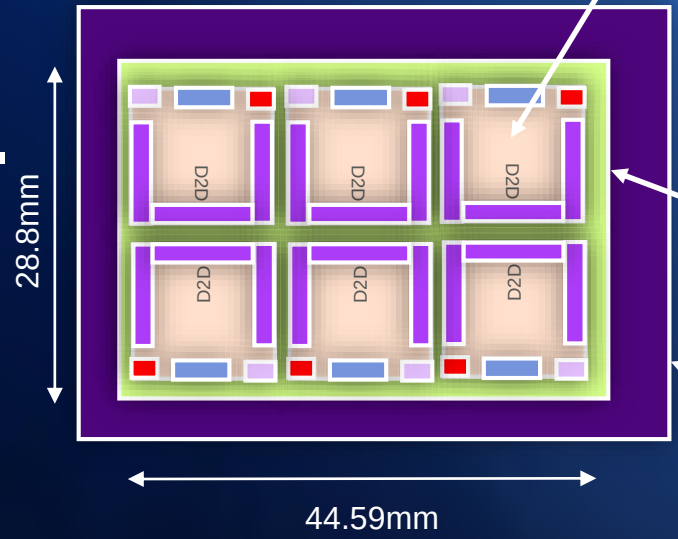
A Better Way of Building AI chips



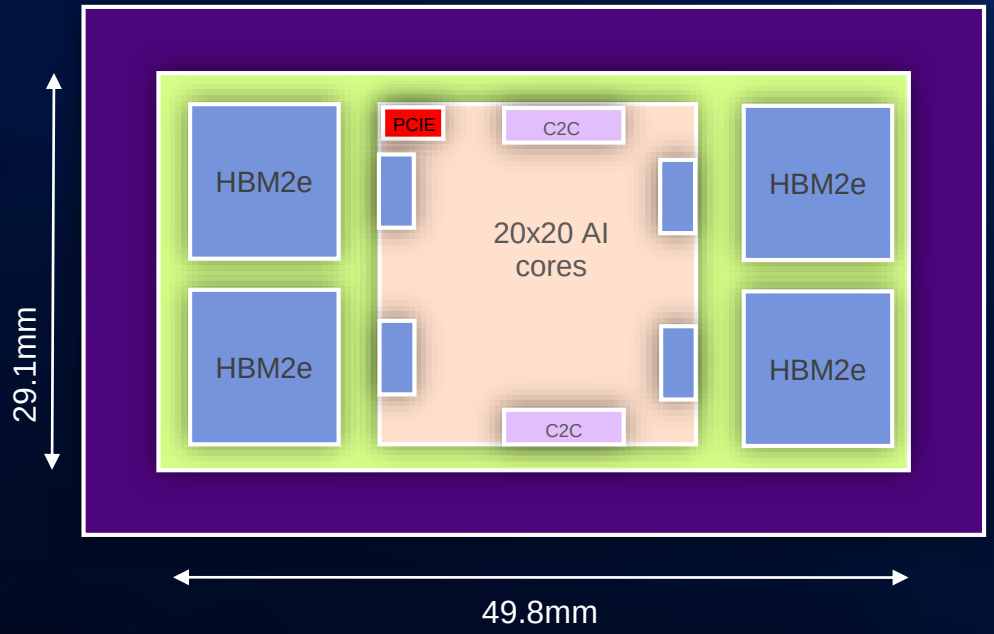
1x

vs.

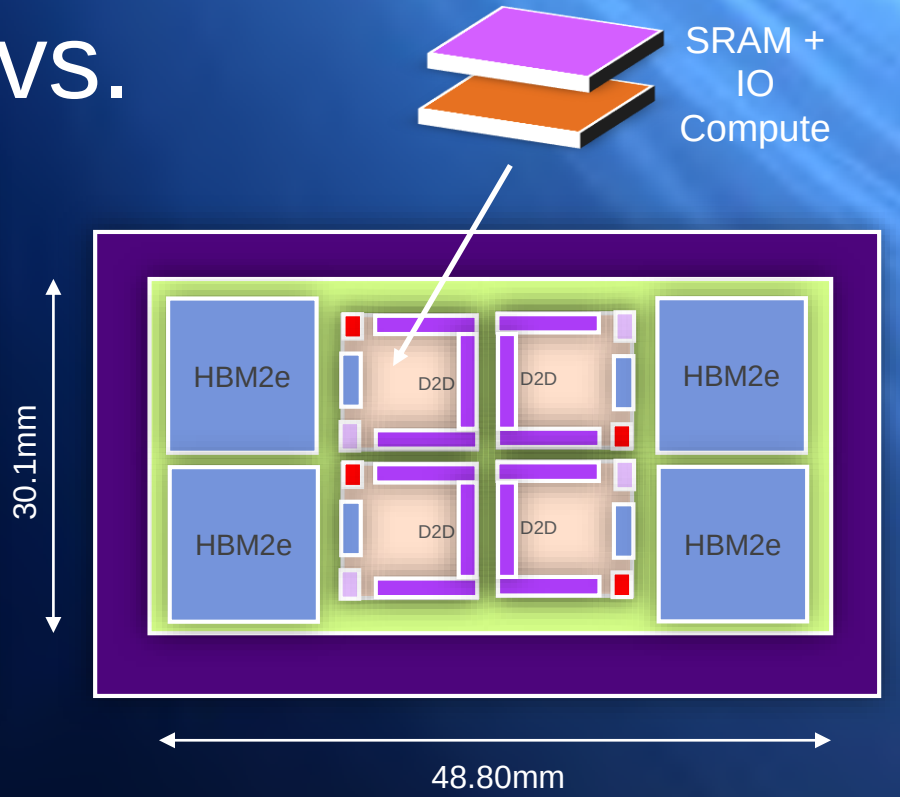
1x



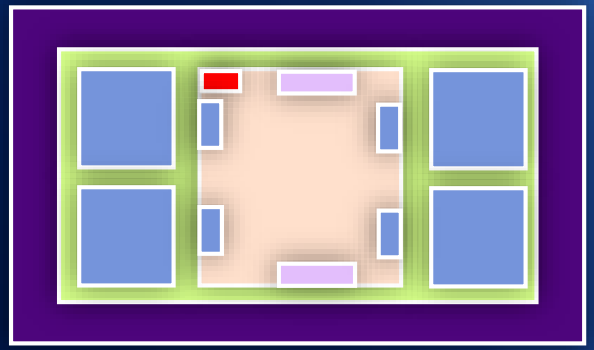
A Better Way of Building AI chips

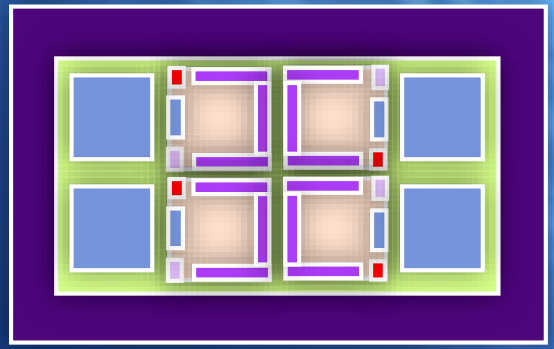


VS.



A Better, Cheaper Way of Building AI Chips





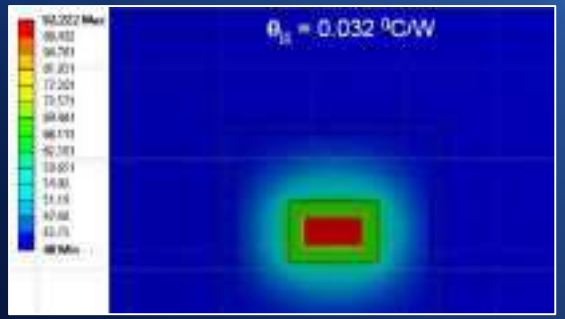
Metric per die	2.5D	3D + 2.5D
Size @ N7	789mm^2	197mm^2 x4 x2
SRAM	>4Gbit	>4Gbit
AIB+/HBI+ bandwidth	N/A	12Tbps
LR SERDES @ 56G	32 lanes	32 lanes
Yield	15.7%	68.6%
Total KGD + interposer cost (*)	$1,294.6	$673.4

(*) Excluding DRAM

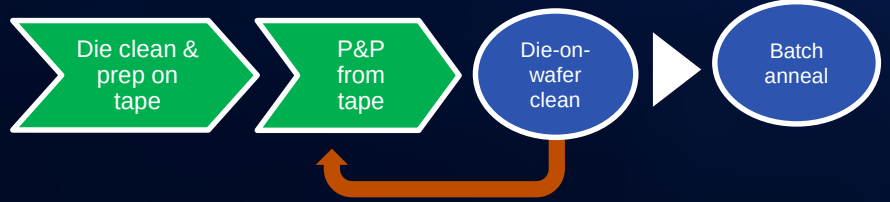
From Feasibility to Maturity

What are the (new) challenges?

Challenge	Response
Die bonding technology (WoW, DoW)	• Xperi's DBI • TSMC's Hybrid Bond
Vertical signal density	2-5um DBI via pitch
Thermal density	Worst case with cold plate: • Tmax= 92.2ºC • Thetajc=0.032ºC/W
Combined yield	Established DoW KGD assembly flow for mass production



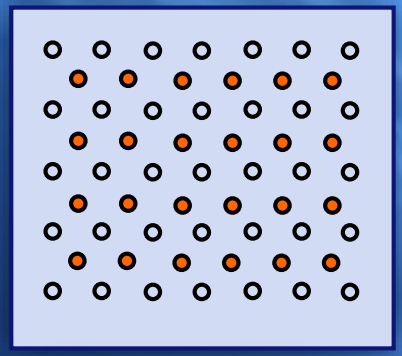
Production Ready Flow

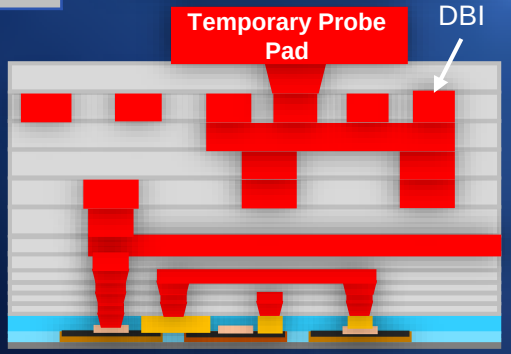


From Feasibility to Maturity

What are the (new) challenges? (II)

Challenge	Response
Wafer probing	Temporary probe pads
Power delivery	Independent power meshes + TSV
TSV-aware placement & routing	• Place PDN TSV matrix first • TSV pitch 55um – 388um for 0.5% IR drop • Bulk of memories on top die



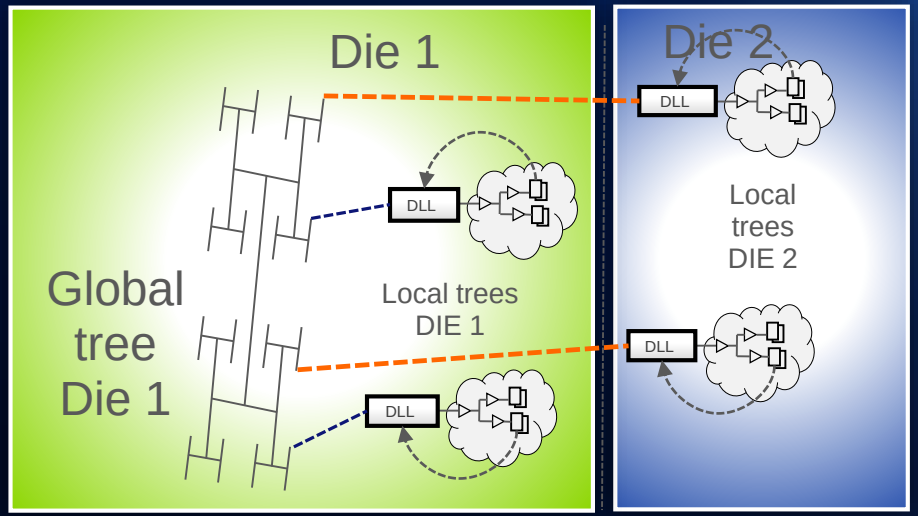


Array	pitch (um)	x (um)	y (um)	area of TSV (um2)	Open window (um)
1x1	55.56	3	3	9	52.56
2x2	166.67	6	6	36	160.67
3x3	277.78	9	9	81	268.78
4x4	388.89	12	12	144	376.89

From Feasibility to Maturity

What are the (new) challenges? (III)

Challenge	Response
CTS	DLL-based CTS for block-level trees
Multi-die signoff	Same as single die, if DLL-based CTS used



From Feasibility to Maturity

What are the (new) challenges? (IV)

Challenge	Response
C2C interconnect medium	• Interposer: 52x52mm in 2020 • Interposer-less: Xperi's EOI, TSMC's InFO-MSL
Warpage	• Stiffener ring-based design; advanced materials; dummy dies
Supply chain readiness	• Supply chain enabled • In mass production today for image sensors • First SoC commercial products hitting the market now

Conclusions

Words from a techno-mason

- Hyperscale dominates the TCO discussion
- AI is all about efficiency
- Efficiency is all about ASIC
- Powerful new enabling technologies are coming of age
  - 3D IC
  - New memory technologies
  - Memory overlays
  - AI-specific IP
- Key to a reduced TCO

Collaborate. Differentiate. Win.