JULY 9–13, 2023

MOSCONE WEST CENTER

SAN FRANCISCO, CA, USA

GUC's Chiplet Solutions from 2.5D to 3D

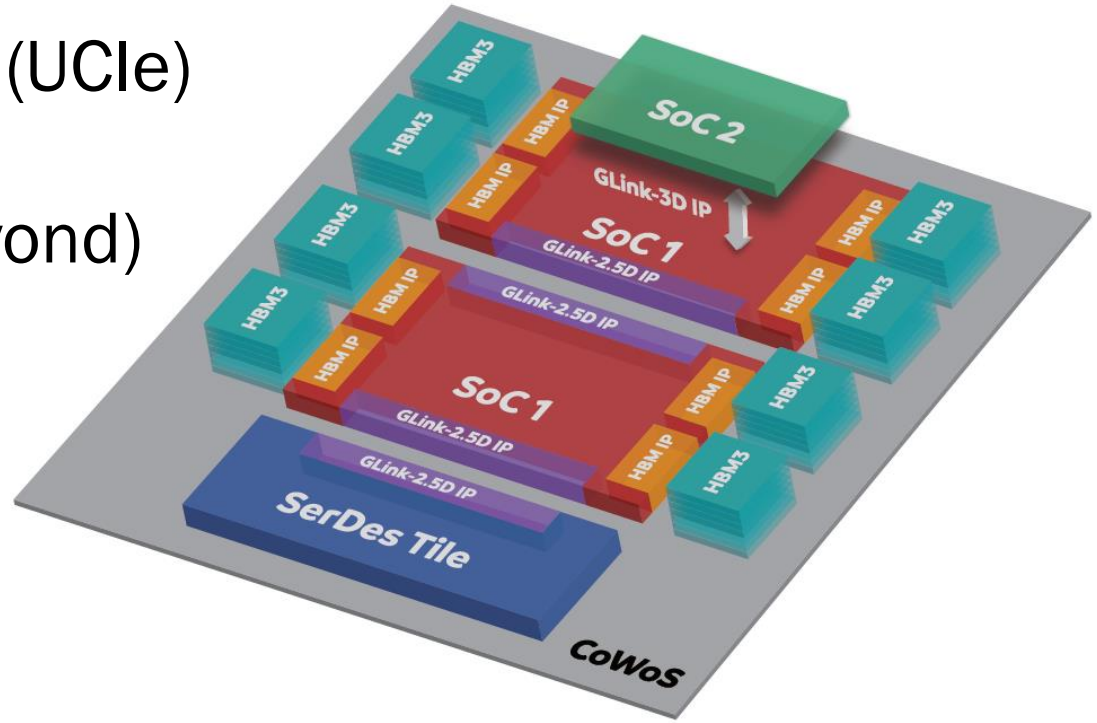
By Igor Elkanovich, GUC CTO

GUC

The Advanced ASIC Leader

Agenda

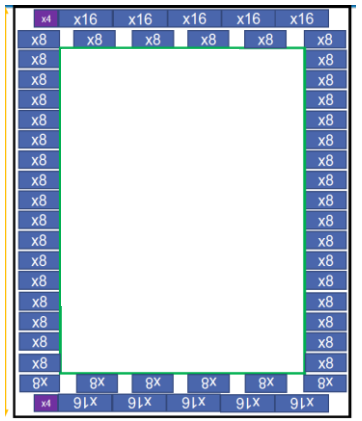
- Chiplets-based architectures: advantages
- Challenge #1: interfaces standardization (UCIe)
- Challenge #2: memory (HBM3/4 and beyond)
- Challenge #3: density and power (3DIC)



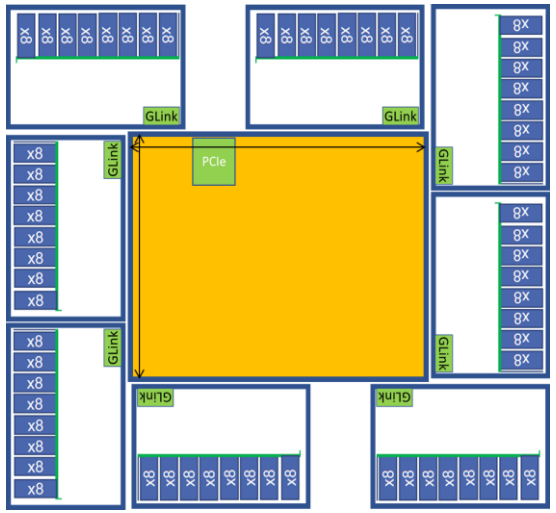
51.2Tbps Switch Monolithic vs. Chiplet-based

- 51.2Tbps Switch can be implemented in two ways: Monolithic and Chiplet-based
- There are many pro/cons but today I'd like to focus on cost comparison only
- Chiplet-based solution is 2.2x cheaper than Monolithic
- Key reason is much higher yield of chiplets
- 2nd TO NRE is breakeven after 8 Ku

Monolithic Die



Chiplet-based



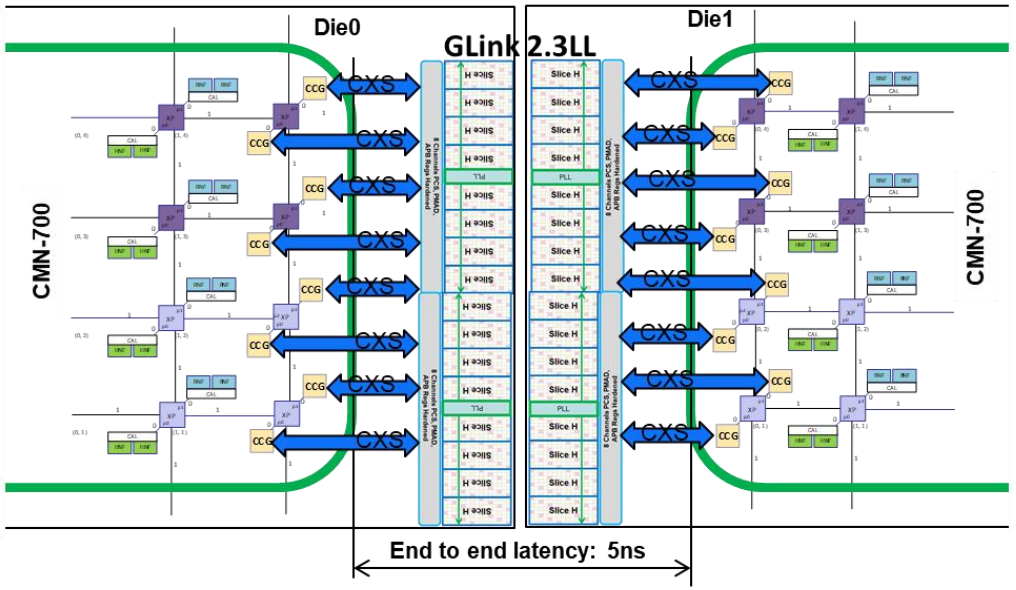
Cost Comparison

	Monolithic	Chiplet-based	
	Main die	Main chiplet	Serdes chiplet
Yield	15%	40%	80%
Total cost	2.2x $X	$X	

Next Gen 2.5D Chiplets Processors

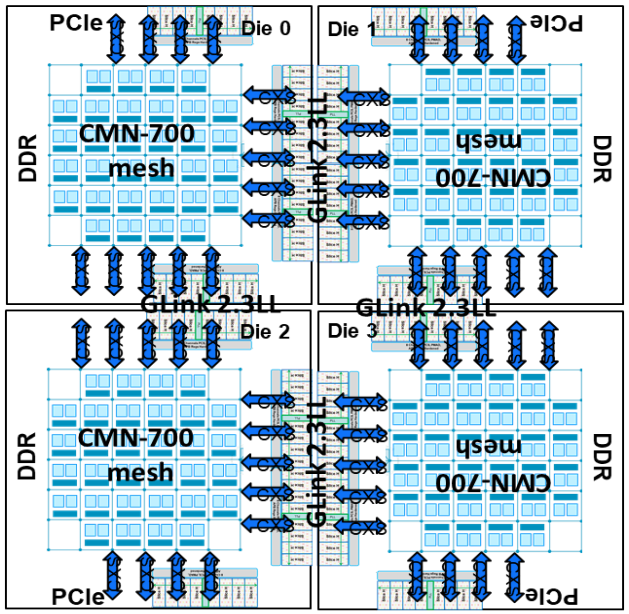
- 2.5D packaging for dense, low latency, low power interconnect
- All mesh boundary nodes of both dies are connected by low latency wide buses
- Performance : like one huge mesh of cores

Connecting all Mesh Boundary Nodes



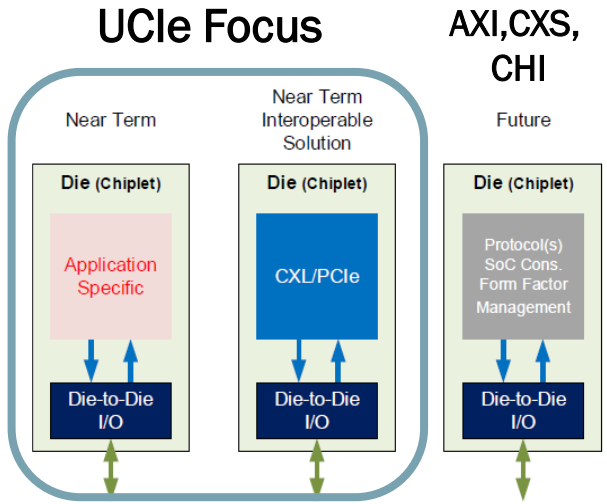
Quad-die ARM CPU

Bigger than reticle

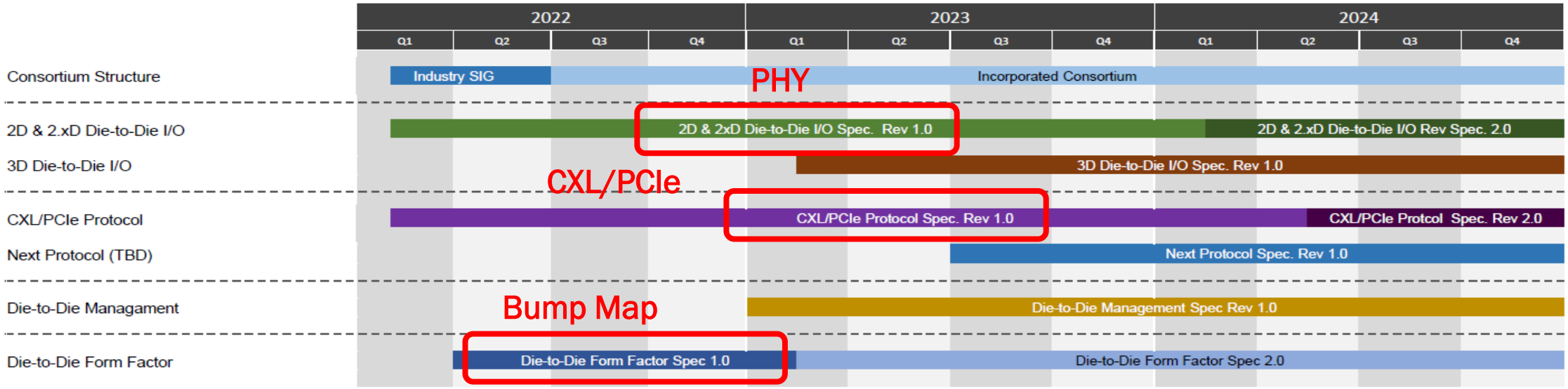


UCIe Standard Status

- 1 year since UCIe-1.0 was published (24/Feb/2022)
- Working on UCIe-1.1, target Q2/2023:
  - Full backward compatibility to UCIe-1.0
  - Minor changes, improvements, clarifications, small additions
  - Main Intel focus: PCIe/CXL protocol and PHY layer
  - Streaming/Raw mode the most common use case

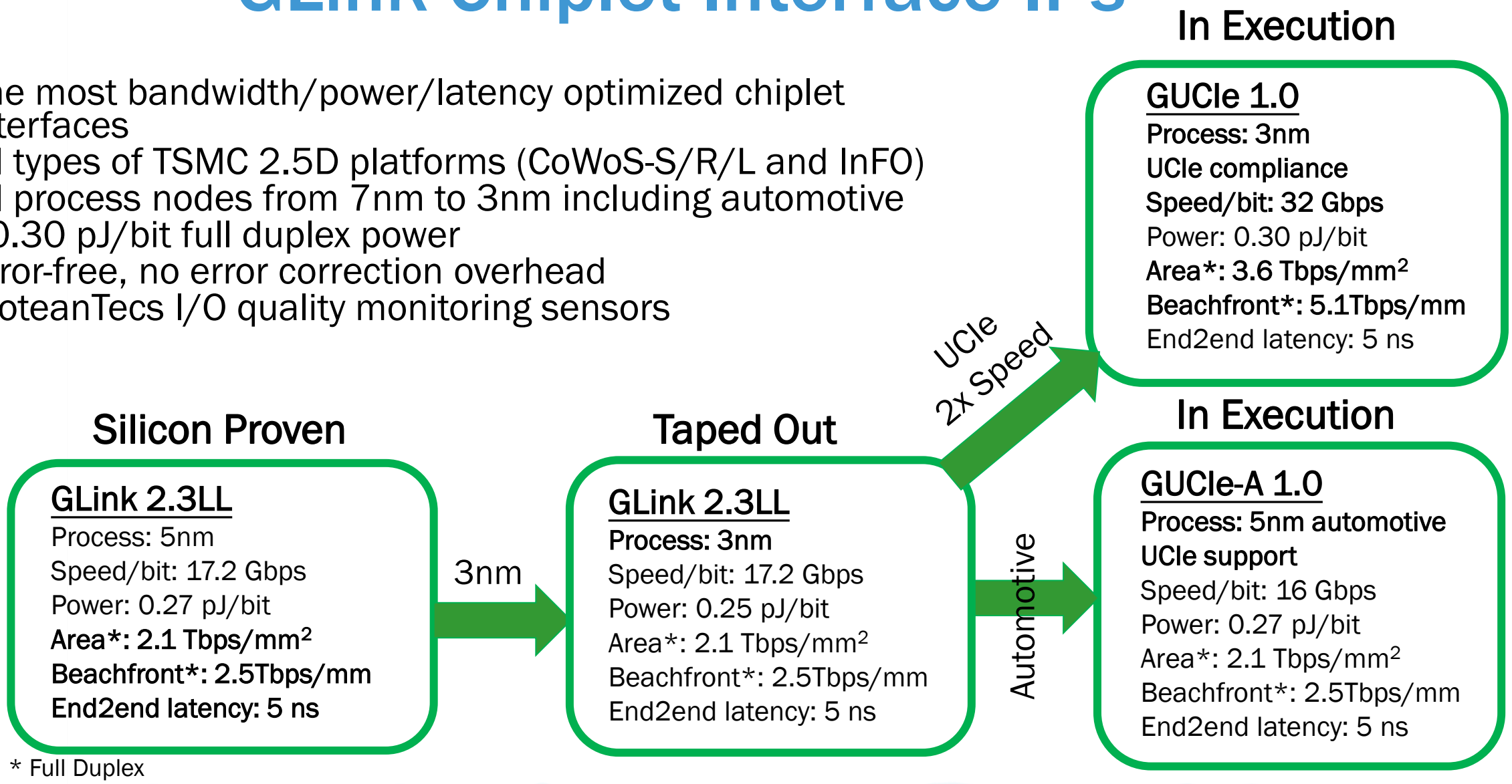


UCIe Schedule



GLink Chiplet Interface IPs

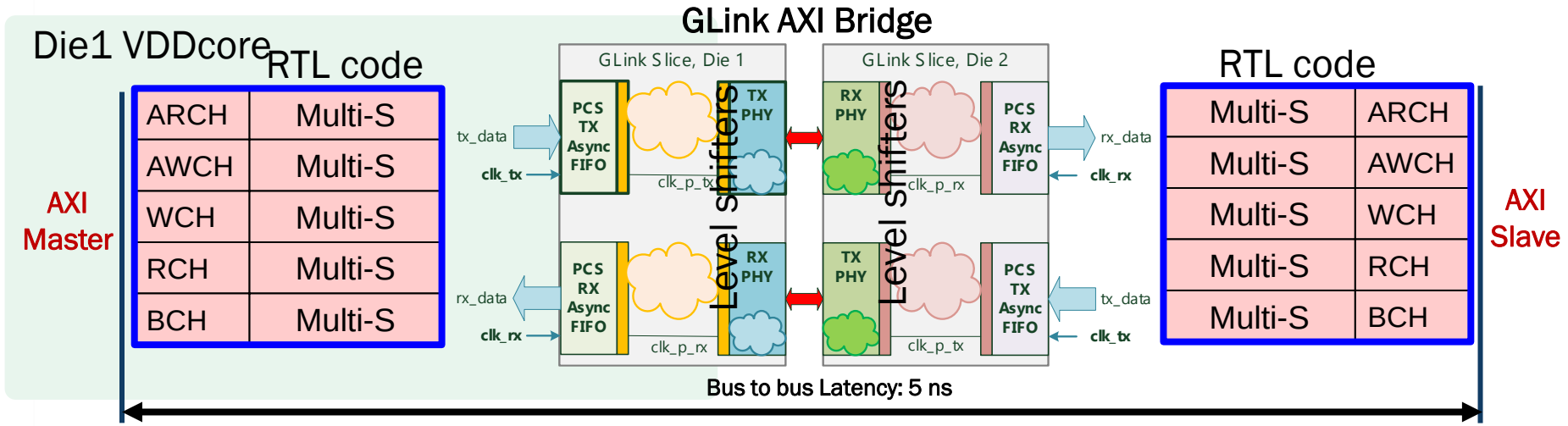
- The most bandwidth/power/latency optimized chiplet interfaces
- All types of TSMC 2.5D platforms (CoWoS-S/R/L and InFO)
- All process nodes from 7nm to 3nm including automotive
- <0.30 pJ/bit full duplex power
- Error-free, no error correction overhead
- proteanTecs I/O quality monitoring sensors



* Full Duplex

AXI/CXS/CHI Bus Bridges

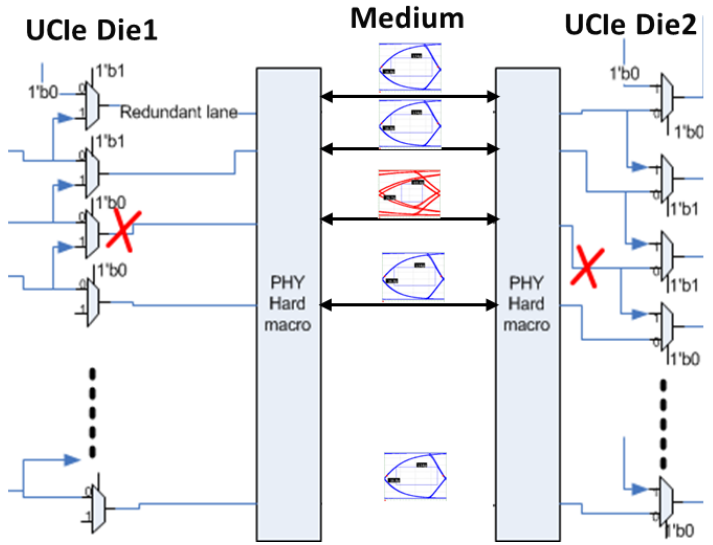
- GUC provides die to die interface bridges for popular buses: AXI, CXS and CHI
- Bus parameters (width, etc..) are configurable
- GLink allows error-free data transfer and efficient flow control
- DVFS: chiplets' voltages and frequencies can be changed on the fly while keeping data transmission



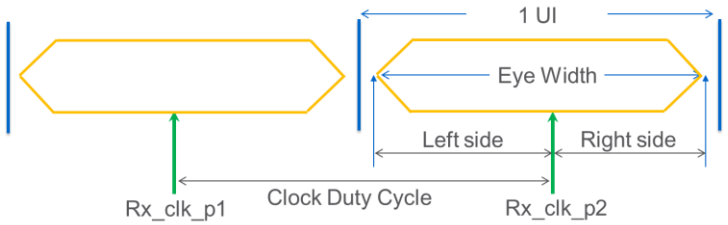
Preventive Monitoring Proposal to UCIe

- Continuous monitoring of UCIe signals was proposed by proteanTecs, GUC and Mercedes-Benz
- Based on proteanTecs silicon proven signal quality monitoring in series of GUC's HBM and GLink testchips
- Signal quality is monitored in mission mode, during data transfer
- Every signal lane is monitored and reported separately
- Monitoring is done continuously, power and signal integrity events are detected and their time is reported
- Bumps and traces defects are identified before they cause interface fail
- Repair algorithm can replace marginal I/Os with redundant ones
- System operation failure is prevented, chip lifetime is extended

Monitoring System



Monitored Parameters



From HBM3 to HBM4

- HBM vendors keep aggressive roadmap increasing throughput and memory size, amount of banks
- Fundamental DRAM timing parameters don't change
- HBM Controller is getting more sophisticated to allow full bus utilization at random access

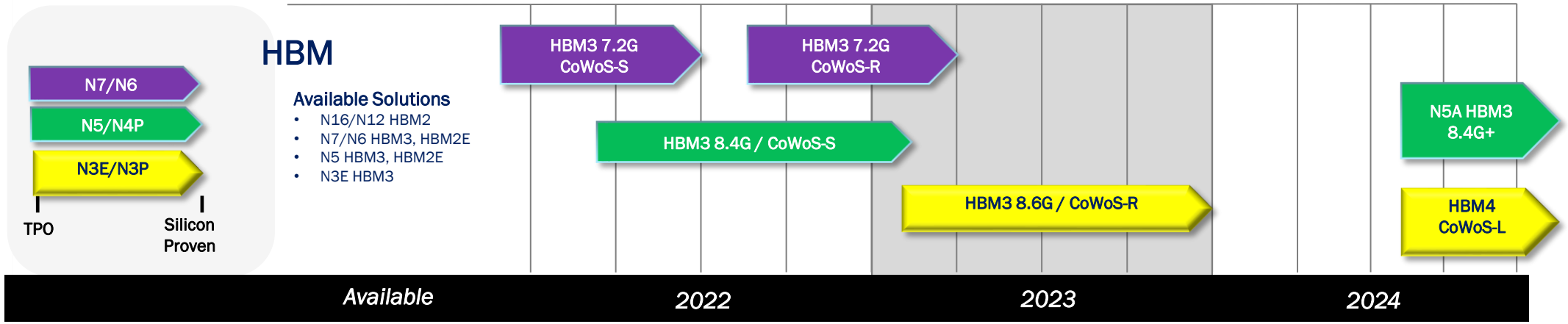
HBM Roadmap

	HBM3	HBM3P/A	HBM4
I/O bus width	1024 bit	1024 bit	2048 bit
Speed	6.4 Gbps	8.0 Gbps	6-8 Gbps
Memory size	16 GB	24-36 GB	24-48 GB
Availability	2022	2023-2024	2025

GUC's HBM IP and CoWoS Production Experience

- GUC's HBM/CoWoS production experience
  - 6 AI/HPC products in production, 6 AI/Networking products at design stage
- The most power and area efficient HBM3 PHY IP
  - In all nodes from 7nm to 3nm including N5A automotive
- Best-in-class HBM3 Controller : ~90% bus utilization rate at random access
- GUC's proprietary interposer router for best SI/PI/Xtalk
- In-mission mode signal performance and health monitoring by proteanTecs

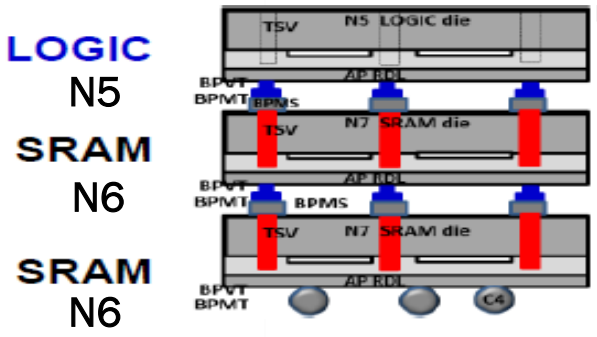
GUC's Controller and PHY Roadmap



GLink-3D Die to Die Interface

- GLink-3D supports all types of TSMC 3D SoIC platforms:
  - Chip on Wafer (CoW) and Wafer on Wafer (WoW)
  - Face to Face and Face to Back
- First GLink-3D testchip is silicon tested
- Next GLink-3D generations:
  - Increasing bandwidth
  - Power and latency reduction

GLink-3D 1.0 Testchip:
Processor on SRAM



Silicon Tested

GLink-3D 1.0
Process: N7/N6 & N5
Architecture: DDR
Speed: 5.0 Gbps/bond
Power: <0.20 pJ/bit
Area*: 9 Tbps/mm2
Bond/TSV pitch: 9 um
Latency: 1-2ns

Density Increase
Power reduction

Latency reduction
Power reduction

In Development

GLink-3D 2.0
Process: N7/N6 & N5
Architecture: DDR
Speed: 5.0 Gbps/bond
Power: <0.10 pJ/bit
Area*: 20 Tbps/mm2
Bond/TSV pitch: 6 um
Latency: 1ns

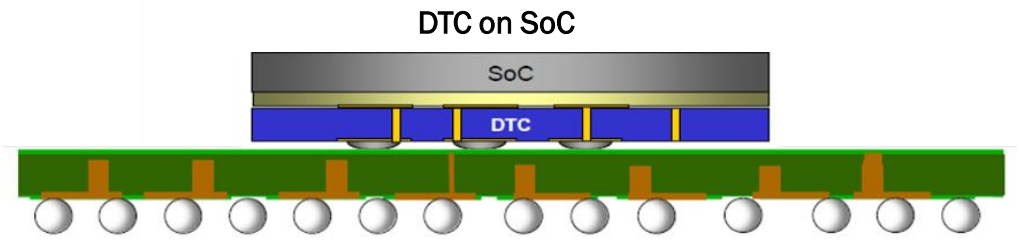
In Development

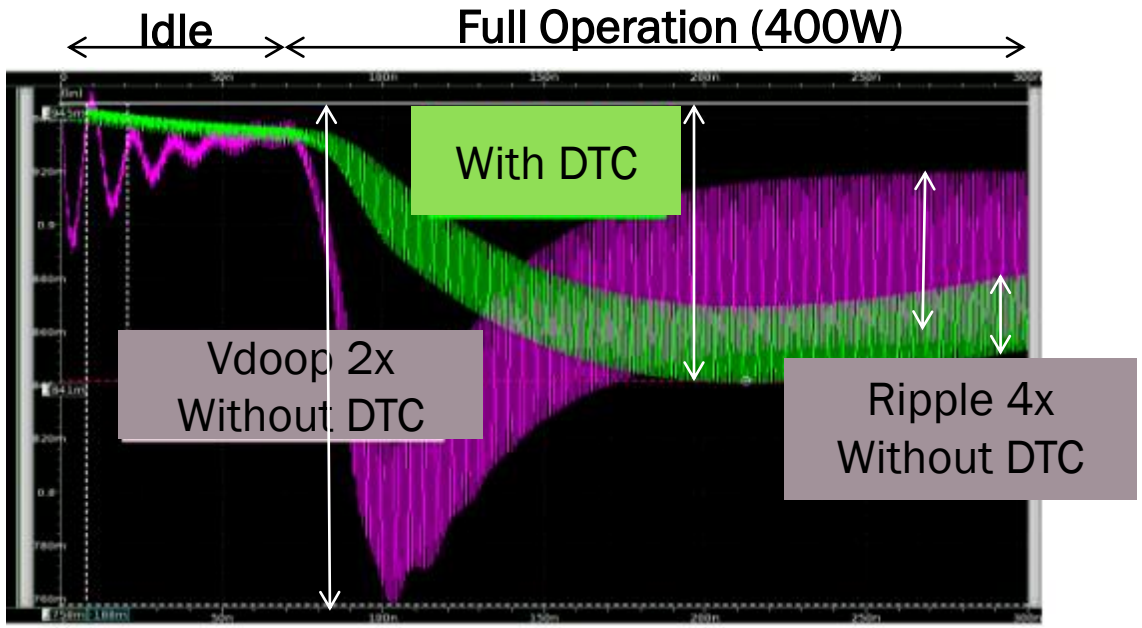
GLink-3D 2.0LL
Process: N7/N6 & N5
Architecture: SDR
Speed: 2.0 Gbps/bond
Power: <0.05 pJ/bit
Area*: 8 Tbps/mm2
Bond/TSV pitch: 6 um
Latency: <500 ps

* Full duplex bandwidth

3D Assembly of DTC on SoC

- DTC allows 100x higher capacitance efficient at high frequencies
- Vdroop is reduced 2x when switching from Idle to full operation
- Voltage ripple at steady state is reduced 4x





Summary

- Chiplets enable modular, bigger-than-reticle, high yield products
  - GLink-2.5D is the most bandwidth, power and latency optimized chiplet interface
  - UCIe enables chiplets interoperability
- HBM3 PHY and Controller are silicon proven
  - CoWoS-R flow is silicon-correlated with better signal and power integrity
- GLink-3D: order of magnitude higher bandwidth, lower power and latency

Comparison of Chiplet Interfaces

	112G-XSR (2D)	GLink-2.5D	GLink-3D
Connectivity	Substrate	CoWoS/InFO	SoIC
Bit Error Rate	1E-7…1E-9	Error-free	Error-free
Power efficiency	1.5 pJ/bit	<0.30 pJ/bit	0.05 pJ/bit
Beachfront efficiency	0.8 Tbps/mm	2.5 Tbps/mm	NA
Area efficiency	0.7 Tbps/mm^2	2.3 Tbps/mm^2	20 Tbps/mm^2
End to end latency	20-30ns	5ns	<500ps