ARTERIS IP

Integrating AI Accelerators into Automotive SoCs with Functional Safety

ARM TECH SYMPOSIA 2019

arm Tech Symposia

Beijing, October 23
Shanghai, October 25

Arteris IP – The World's #1 Interconnect IP Company

FOUNDED 2003; HEADQUARTERS IN SILICON VALLEY

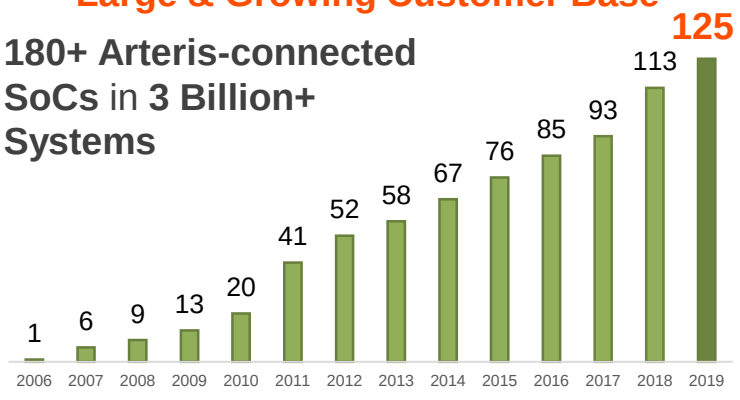
Data is current as of 15 October 2019

Production-Ready Products

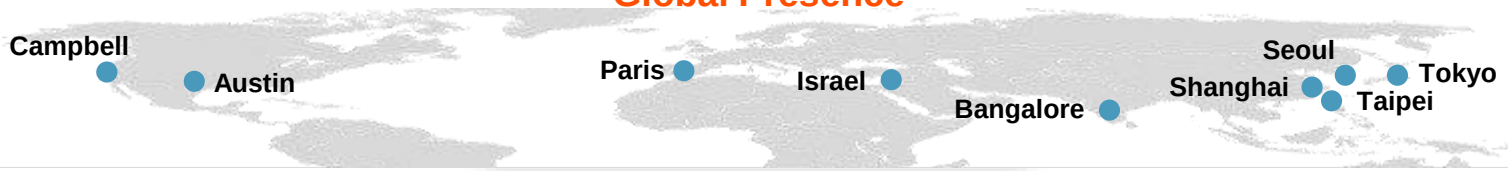
FlexNoC®	2010	Main interconnect, 2nd gen
FlexWay™	2010	IP subsystem interconnect
FlexPSI	2013	All-digital interchip link
FlexNoC Resilience	2014	Resilience for ISO 26262
FlexNoC Physical™	2015	Links to physical SP&R
Ncore®	2016	Cache coherent interconnect
PIANO®	2017	Automated timing closure
Ncore 2	2017	Resilient interconnect, LL$
CodaCache®	2018	Independent last level cache
AI Package™	2018	Machine learning w/ FlexNoC 4

Large & Growing Customer Base

180+ Arteris-connected SoCs in 3 Billion+ Systems



Global Presence



Top Semis use Arteris IP

Publicly Disclosed Customers

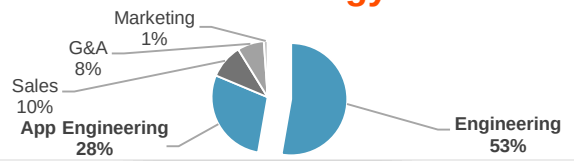
SAMSUNG, Intel, AMD, Baidu, Hyundai, HiSilicon, LG, Spreadtrum, Movidius, NXP, ZTE, SK hynix, ST, TI, arbe, Toshiba, Renesas, nextchip, achronix, Rockchip, MegaChips, applied micro, Kyocera, Hisense, NPT Electronics, SCKIPIO, Schneider Electric, GCT, GUC, pixelworks, SCIUANS, Open-Silicon, Montage Technology, octasic, SEMIDRIVE, VeriSilicon, BLU WIRELESS, Lynxi, Canaan, Enflame, Cambricon, CHX, RDA, Silicon Mobility, WAVE COMPUTING, BITMAIN, NETINT, SiFive, UHNDER, Siengine, AutoChips, BLACK SESAME, DisplayLink

In Leading Systems

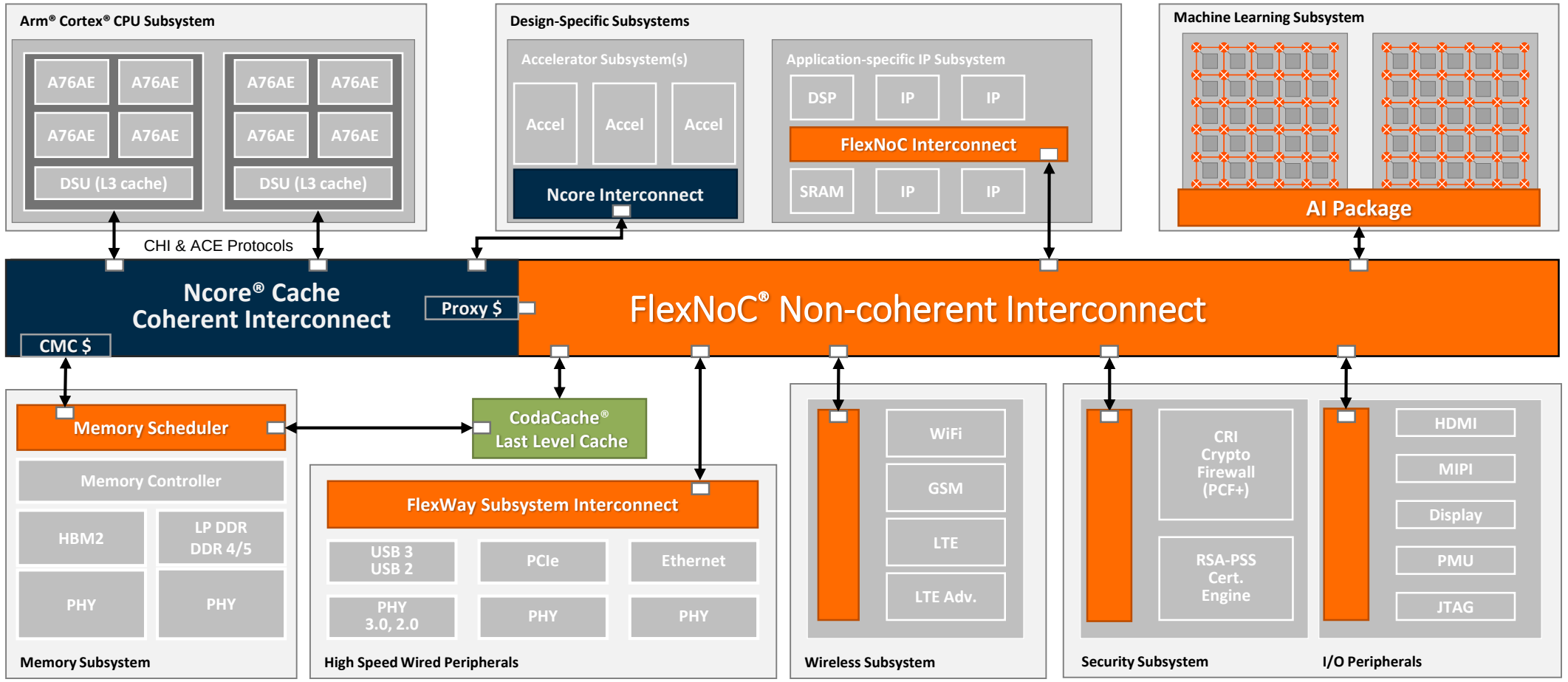
Connected by Arteris Ecosystem

arm, SYNOPSYS, tsmc, SAMSUNG, cadence, ISO 26262, exida, RESILTECH, CXL, CCIX

Interconnect Technology Think Tank



Arteris IP, the Data Highway of the SoC



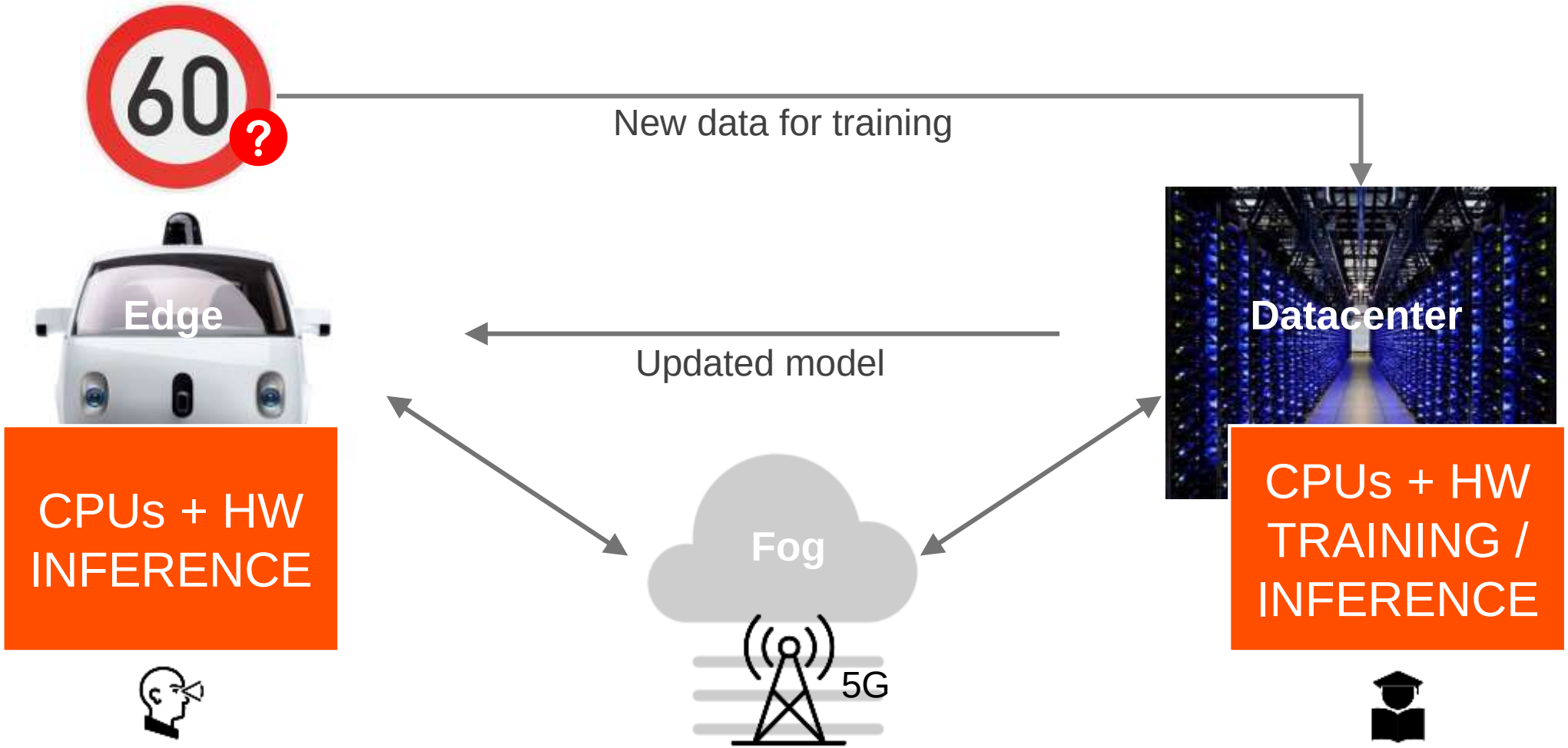
Arteris IP Ncore® cache coherent interconnect IP

Arteris IP FlexNoC® non-coherent interconnect IP

Arteris IP CodaCache® last level cache

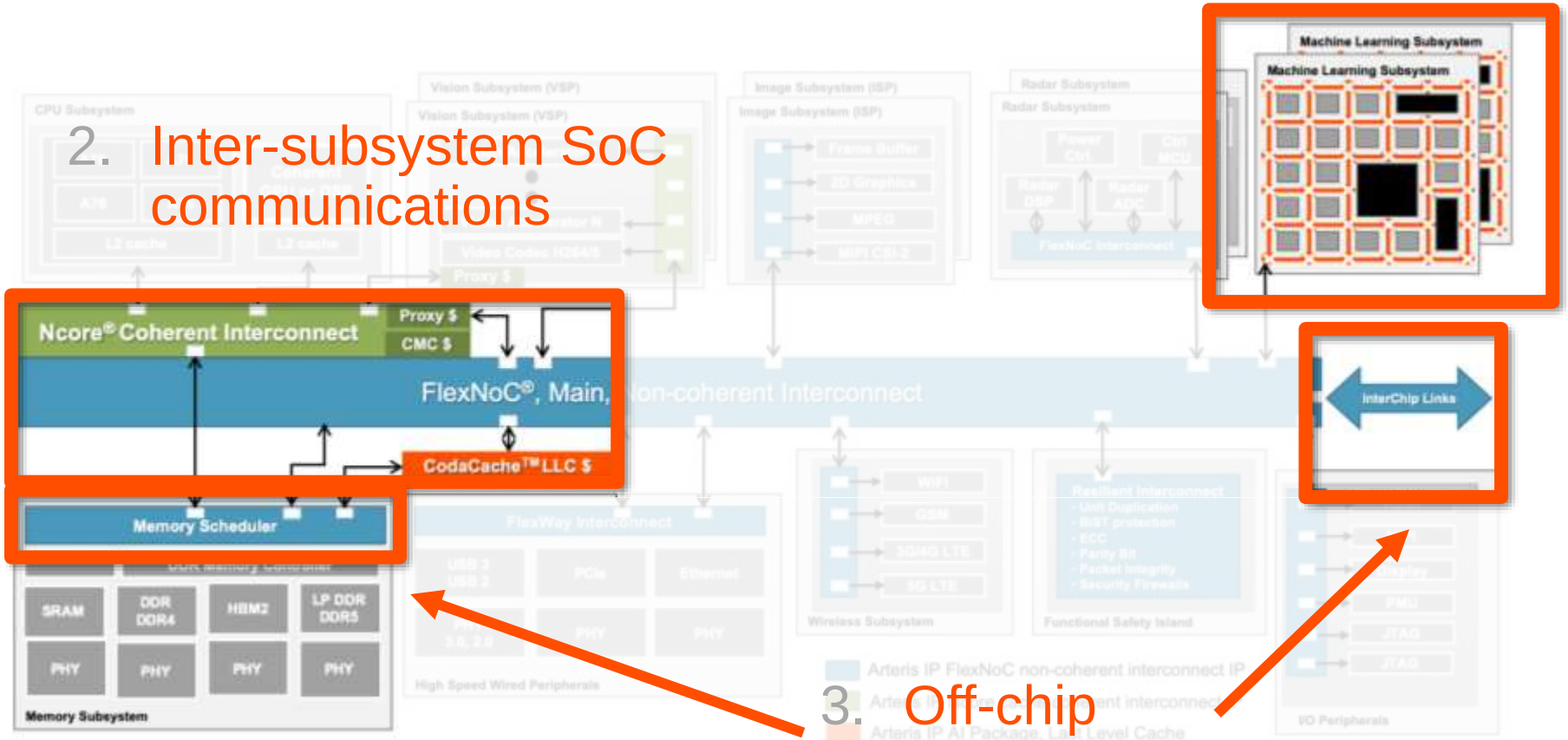
Automotive is driving high-value AI/ML systems

AT EDGE (VEHICLE), DATACENTER, AND FOG (INFRASTRUCTURE AND COMMUNICATIONS)



AI/ML systems implement HW acceleration at 3 levels

OPTIMIZE SOC FOR BEST COMBINATION OF ON-CHIP POWER CONSUMPTION & PERFORMANCE



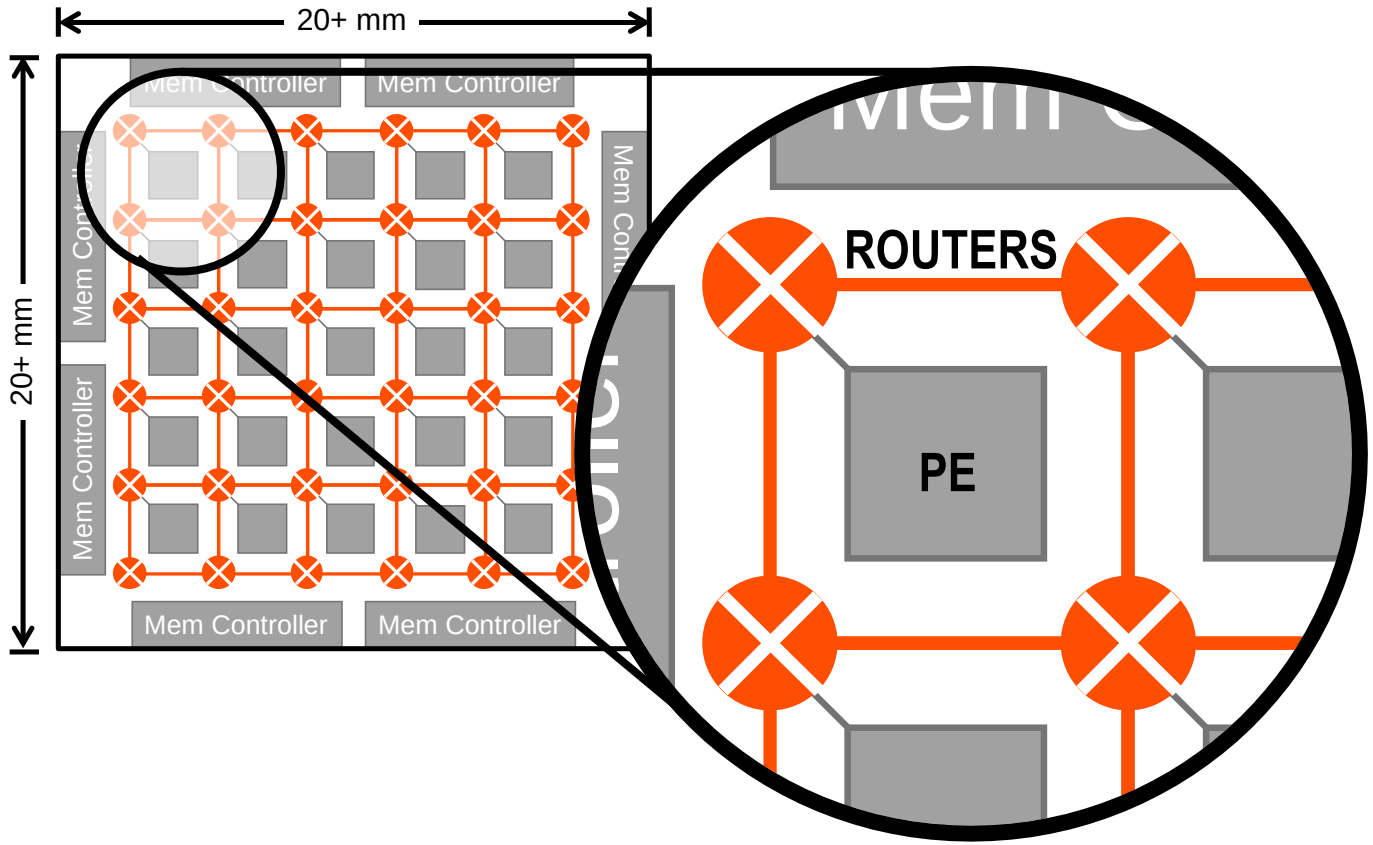
1. Processing elements / Intra-subsystem communications

2. Inter-subsystem SoC communications

3. Off-chip communications

Trends in AI HW accelerators

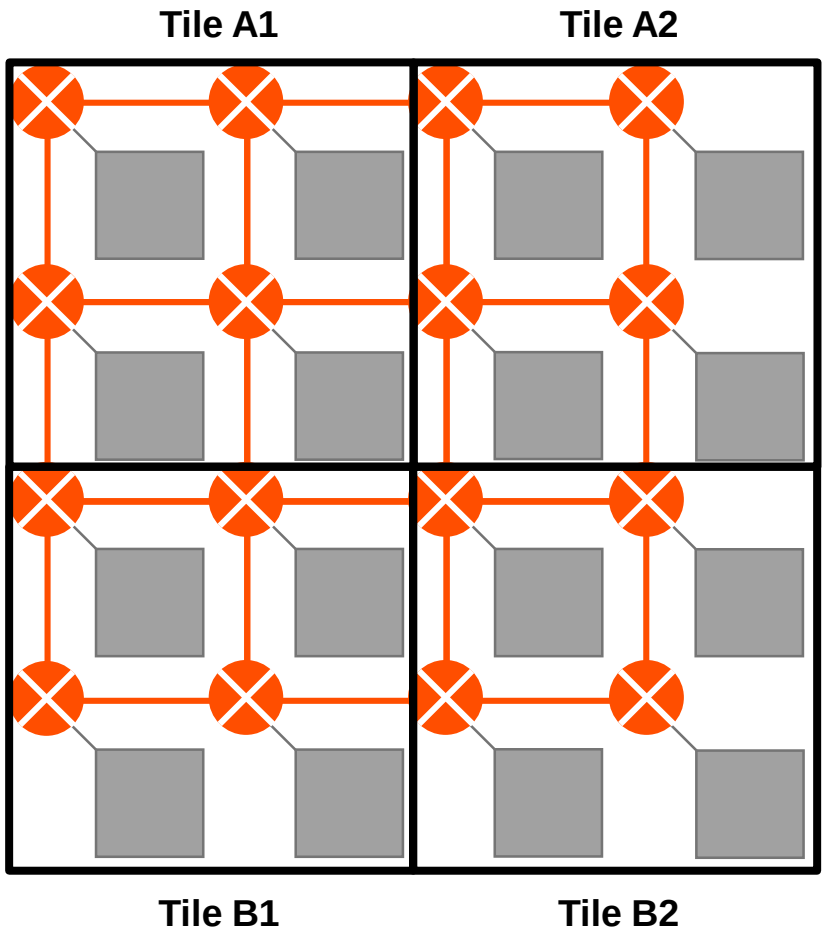
MESH AND OTHER “REGULAR” TOPOLOGIES FOR “GENERAL PURPOSE” TRAINING OR INFERENCE



- Huge scale for datacenter SoCs
- One or more processing elements (PE), memories (caches) or I/O controllers per corner router
- Multicast / Broadcast writes for BW efficiency
- Sophisticated interleaving for optimal off-chip HBM2, GDDR6 or DRAM access

Tiling of subsystems allows massive scalability

BUT REQUIRES SYSTEM-LEVEL PERFORMANCE AND POWER TRADEOFFS



WHY?

- Can be easier for place & route team to massively scale (hard macros)
- Can simplify top-level interconnect, or allow scalable reuse of existing NoC architecture

LESSONS LEARNED

- Requires additional logic and memory in tile
- Adds complexity to NoC addressing
- Reduces flexibility for system-level optimizations (QoS, memory interleaving, power management, etc.)

Cache coherence is becoming common in-vehicle

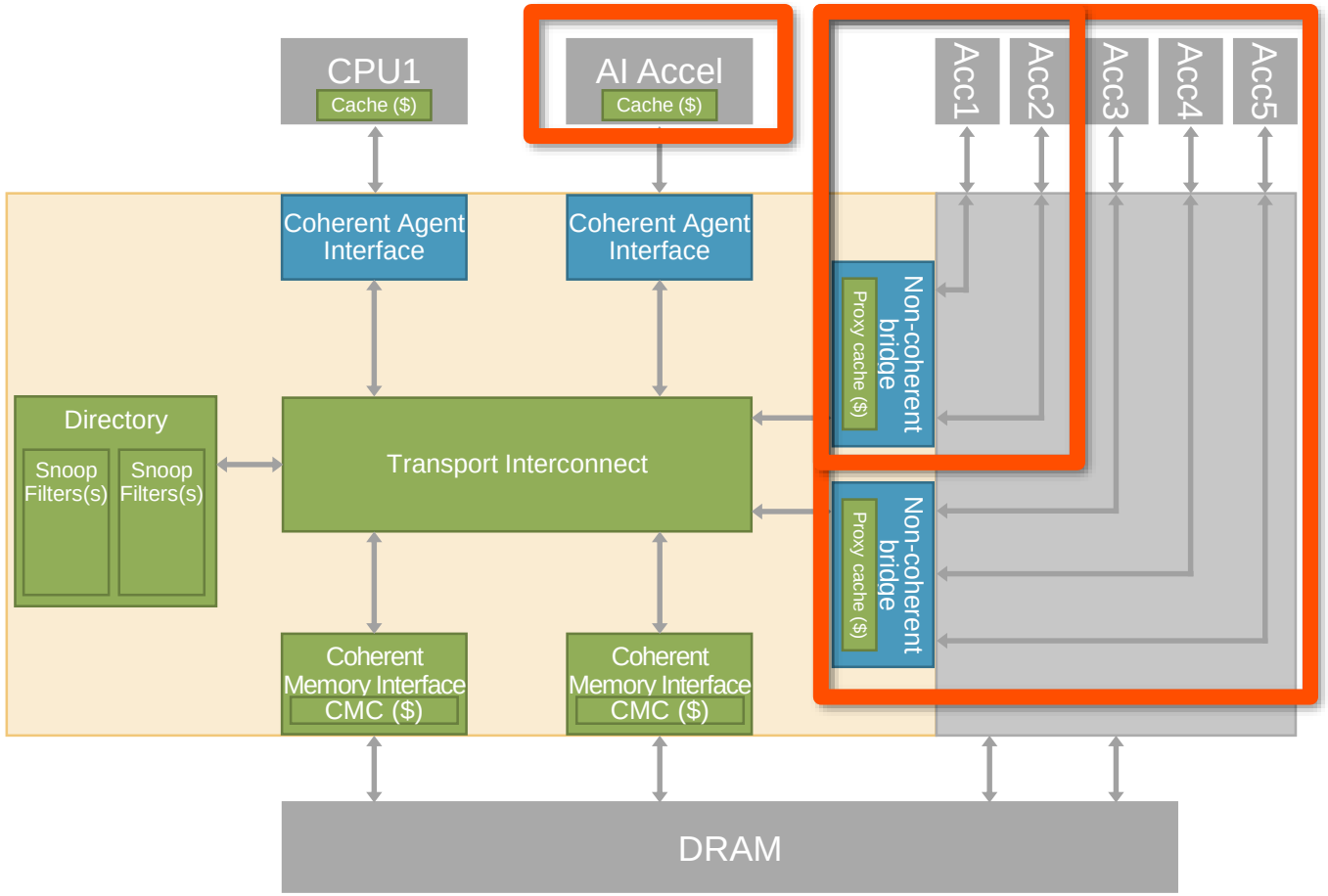
DRIVEN BY NEED FOR “FINER-SLICE” ACCELERATION OF ALGORITHMS

Trend: Integrating non-coherent accelerators into coherent systems

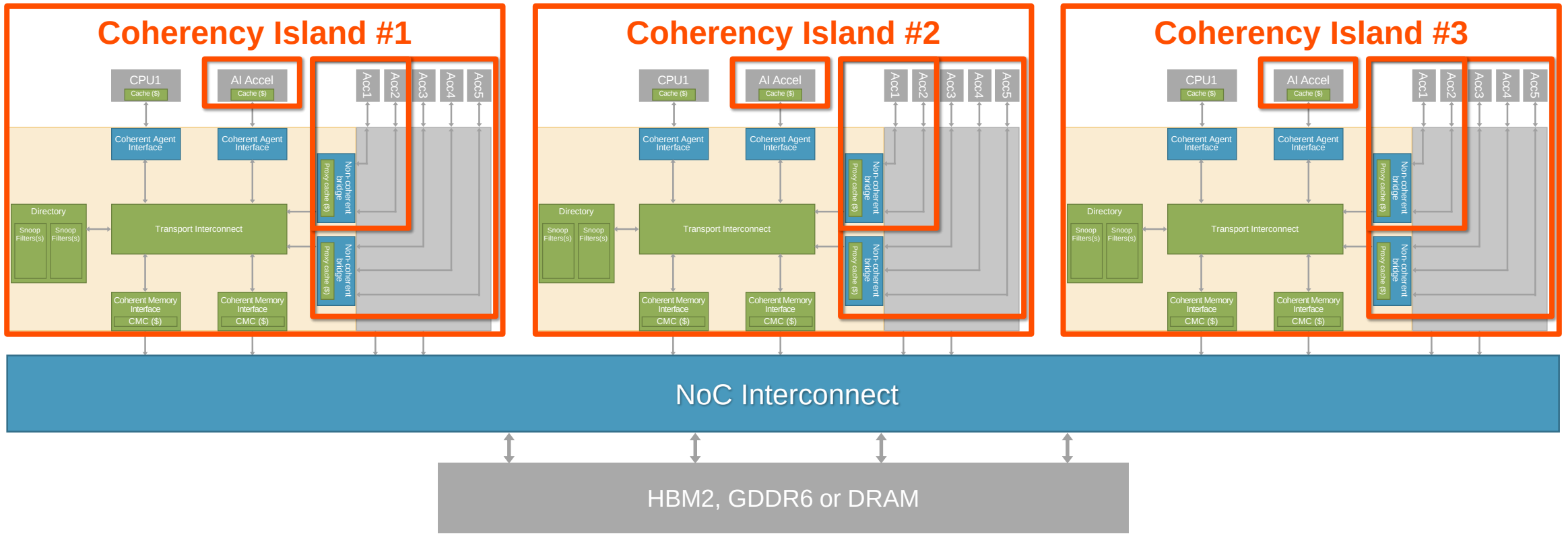
- Seen with multiple heterogeneous processing elements
- Less sensitive to memory sizing design errors (DMA for spillover)
- More freedom to adapt post-Si (Unified Memory Architecture for software)

Enhanced dataflow

- Faster, lower-power sharing between coherent and non-coherent IP
- Tailoring of dataflow to meet near real-time deadlines



Coming soon: “Islands” of cache coherency



- Multiple subsystems that are internally cache-coherent
- Each subsystem usually different, optimized for set(s) of tasks
- Can provide scalability for edge inference processing while meeting latency and power requirements

AI/ML SoCs and ISO 26262 Compliance

Challenges for ISO 26262-compliant AI SoCs

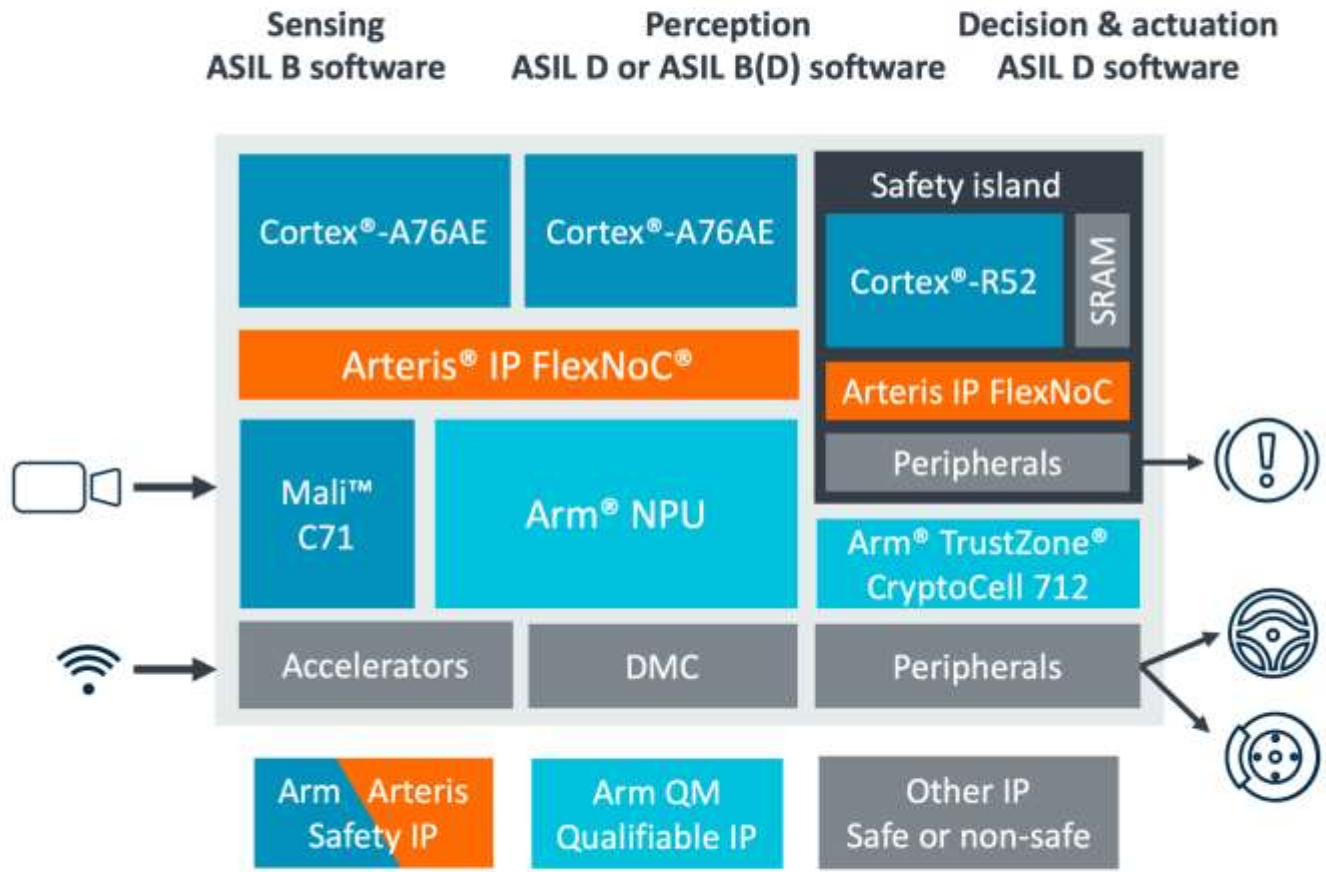
WHETHER USING OWN AI IP SUBSYSTEM OR 3RD PARTY AI IP

- **Hardware integration is only the first part**
  - Transaction protocols, performance modeling, verification
- **ISO 26262 processes, analyses, deliverables and work products are key!**
  - IP, whether internally developed or commercial, must be developed using quality processes
  - Hardware requires safety mechanisms and diagnostic capability (especially for ASIL B, C and D)
  - Analyses (DFA, FMEA, FMEDA, etc.) and validation measures (including Fault Injection)

Realize your Autonomous Future with Arm and Arteris IP

Autonomous vehicles and robots are a huge global opportunity

- Autonomous requirements
  - Advanced high-performance processing
  - Functional safety
- Arm works with the automotive market to develop next generation compute
  - High performance and functional safety
  - Machine learning (NPU)
- Arm and Arteris IP are IP leaders for functional safety systems
  - Simplifies SoC functional safety
  - Delivers necessary high performance
  - Trusted by OEMs for automotive quality



Presented jointly with Arm at ICCAD China 2018, Zhuhai

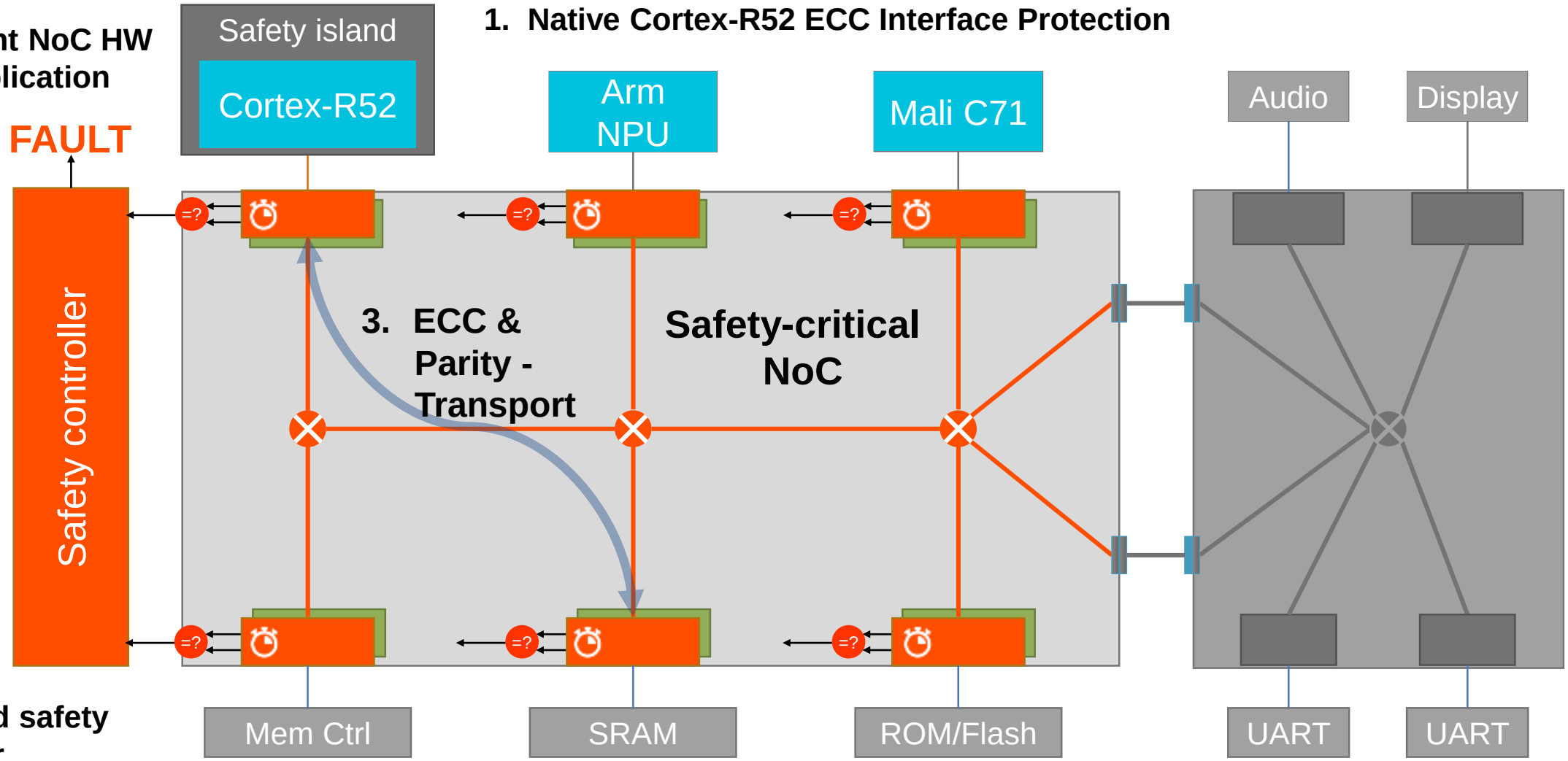
ARTERIS IP + arm

Example System:

Arteris IP Interconnect, Arm NPU, Cortex-R52,

1. **Native Cortex-R52 ECC Interface Protection**

2. **Intelligent NoC HW Unit Duplication**



4. **Integrated safety controller**

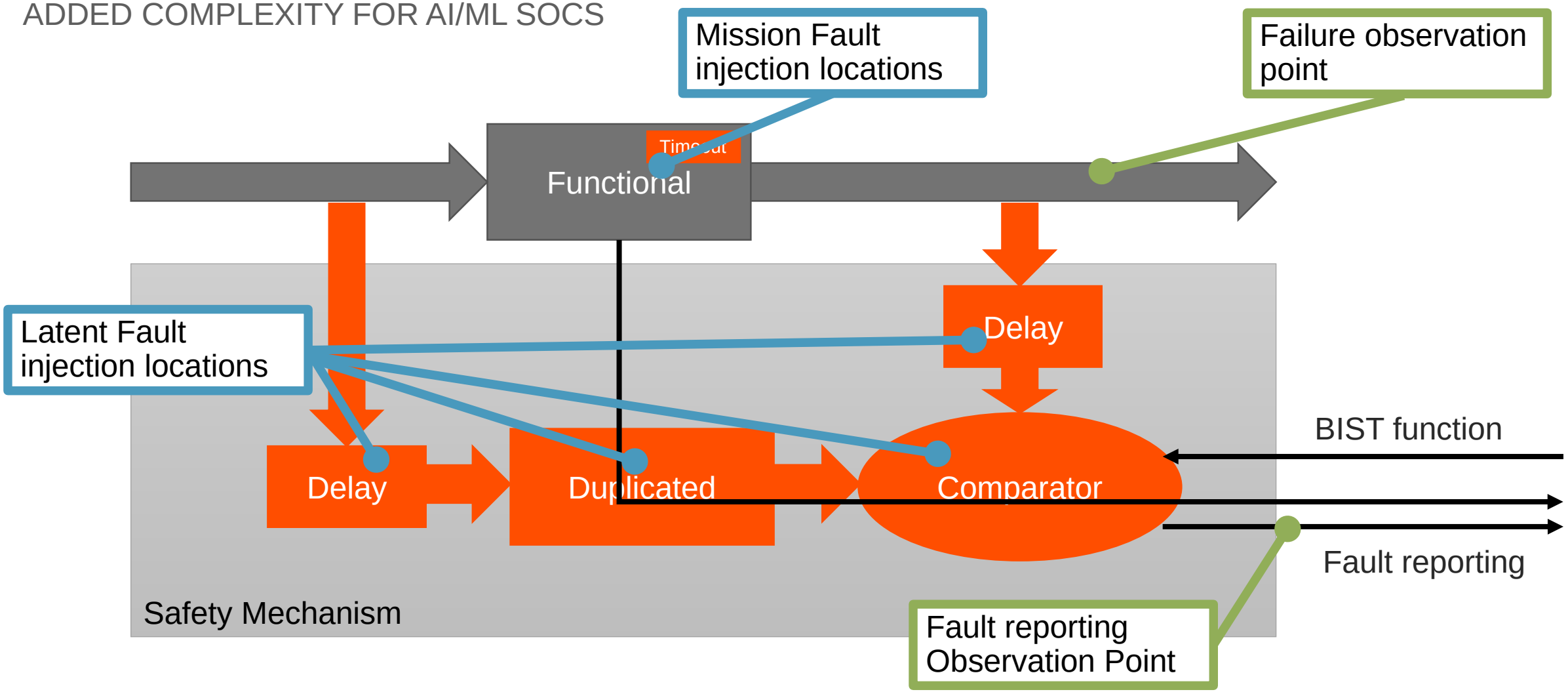
Functions, Failure Modes & Safety Mechanisms

Function	Failure Modes	Safety Mechanisms
Packetization	External interface corruption; External protocol violation; Packet corruption	External placeholder (ECC/Parity); Packet validity checker; Duplication; Initiator timeout
Transport	Packet corruption	ECC/Parity + checker; Packet validity checker; Initiator timeout
Clocking and reset	Clock / reset glitch; Frequency error;	External Timeout Assumption of Use (AoU)
	Wrong clock gating	Initiator timeout; Packet validity checker; Percentage safe AoU
Safety reporting	Missed / incorrect reporting; unexpected reporting of Fault	Register parity; Regular check AoU
Safety mechanism	Missed / incorrect reporting; unexpected reporting of Fault	BIST; Regular check AoU

Safety mechanisms within AI/ML accelerators and throughout the SoC

Validation: Fault injection of duplicated logic

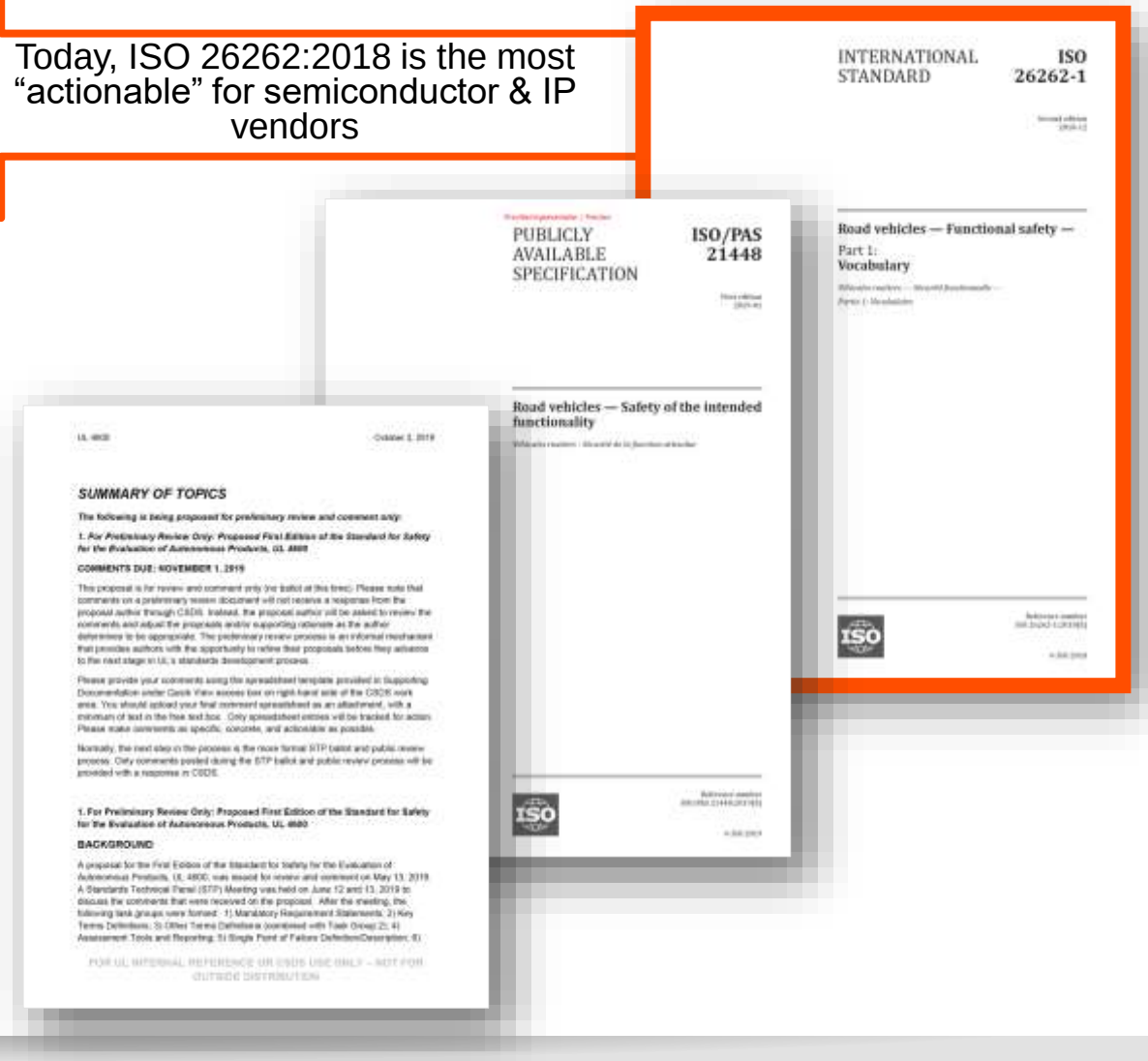
ADDED COMPLEXITY FOR AI/ML SOCS



Automotive AI/ML safety standards are evolving

- **ISO 26262:2018, "Road Vehicles – Functional Safety"**
  - Systematic & random errors leading to safety-related failures
- **ISO/PAS 21448:2019, "Safety of the Intended Functionality (SOTIF)"**
  - "The absence of unreasonable risk due to hazards resulting from **functional insufficiencies** of the intended functionality or by **reasonably foreseeable misuse** by persons…"
  - From same working groups as ISO 26262
- **UL 4600, "Standard for Safety for the Evaluation of Autonomous Products"**
  - Led by Phil Koopman from Carnegie Mellon University
  - Focuses on goal based and technology-agnostic safety case creation and validation

Today, ISO 26262:2018 is the most "actionable" for semiconductor & IP vendors

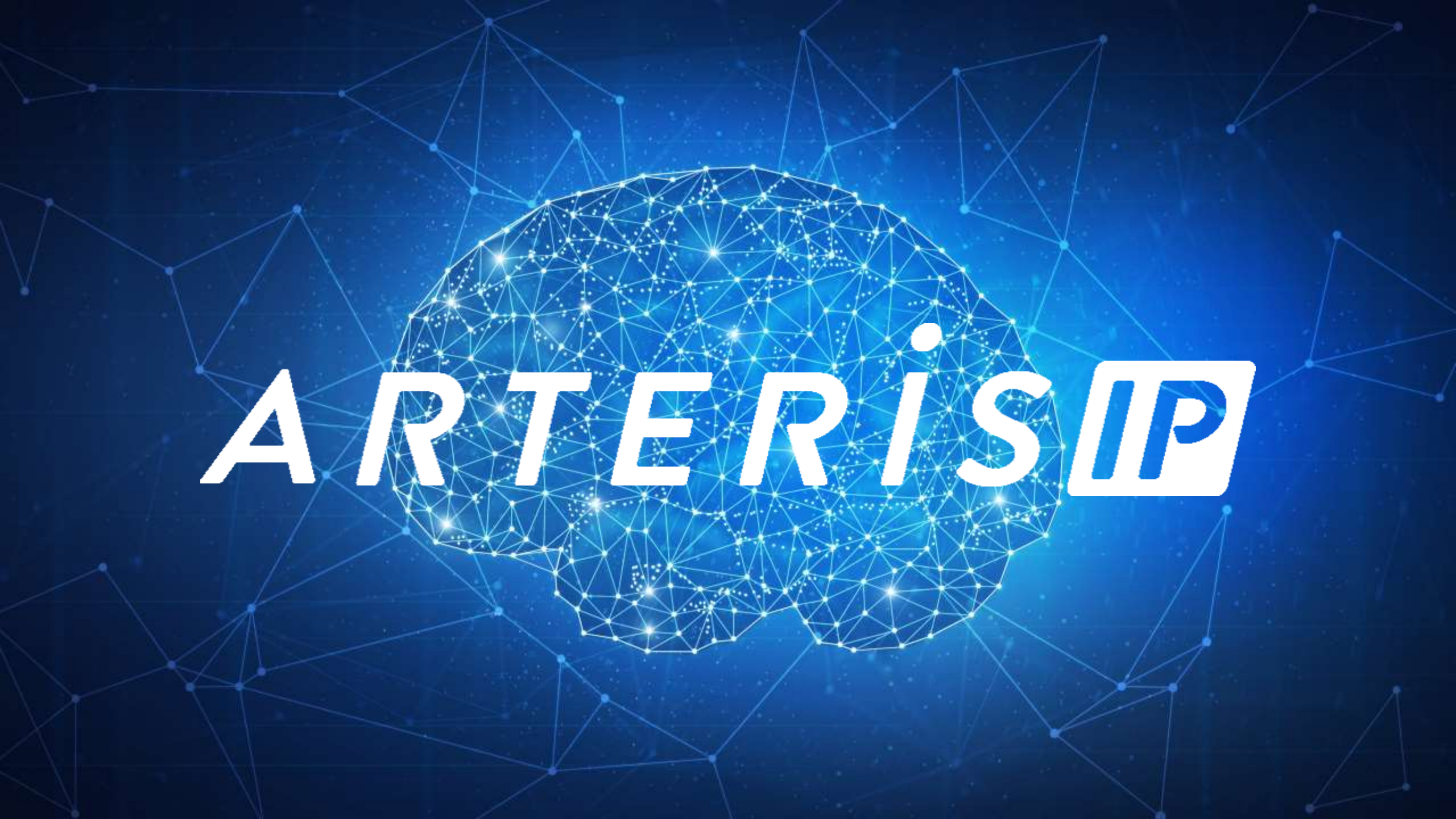


Conclusion:

Lessons learned integrating AI and ML into functionally safe automotive SoCs

- On-chip NoC interconnect is key enabler for dataflow within AI HW accelerators & at SoC level
  - Allows optimization for desired power/performance tradeoffs
  - Provides in-hardware data protection to meet functional safety requirements
- Massively scalable systems are using meshes and netlist-level tiling to connect accelerator subsystems
  - Primarily in datacenter because less power efficient
- Edge systems with more stringent power and real-time processing requirements are implementing complex cache coherent architectures
- All these systems require in-SoC functional safety mechanisms to meet automotive standards

NoC interconnect is key enabler for autonomous driving SoCs


ARTERIS IP