Why SoC Interconnects Have Outgrown the Bus

The NoC-Centric, CHI-Native Path Forward

SignatureIP Engineering Team
Center of Excellence for Compute and I/O Subsystem IP

Abstract. ***The shared-bus and crossbar interconnects that defined SoC architecture for two decades have run out of room. Workloads have changed faster than the buses meant to serve them. Processor counts have grown, coherency requirements have spread beyond CPU clusters into accelerators, and integration density has pushed past what arbitrated buses or large crossbars can support without prohibitive area, power, and timing penalties. Network-on-Chip fabrics are now the default on-die transport for AI accelerator, server-class, and high-end automotive SoCs. Four forces drive the migration: heterogeneous-coherency AI workloads, automotive ISO 26262 safety requirements, chiplet disaggregation through UCIe 2.0, and advanced-node wire-delay constraints. This paper examines each force, explains why packet-switched NoC fabrics with directory-based coherency address them, and presents the SignatureIP portfolio built for this transition.***

Index Terms. ***Network-on-Chip; AMBA CHI Issue E.b; AXI4; SoC Interconnect; Cache Coherency; Directory-Based Snoop; ISO 26262; Automotive ADAS; Chiplet Disaggregation; UCIe 2.0; Inoculator RTL Generation***

I. EXECUTIVE SUMMARY

The shared-bus and crossbar interconnects that defined SoC architecture for two decades have run out of room. Workloads have changed faster than the buses meant to serve them. Processor counts have grown, coherency requirements have spread beyond CPU clusters into accelerators, and integration density has pushed past what arbitrated buses or large crossbars can support without prohibitive area, power, and timing penalties. Network-on-Chip fabrics are now the default on-die transport for AI accelerator, server-class, and high-end automotive SoCs, and the migration is well past the early-adopter phase.

Four forces are driving the migration. AI and ML accelerators demand coherent bandwidth across heterogeneous compute domains, with snoop traffic and bandwidth profiles that broadcast-based coherency cannot serve. Automotive ADAS and EV platforms require ISO 26262 ASIL-B through ASIL-D certification, deterministic latency, and on-chip safety mechanisms that legacy interconnects were never designed to carry. Chiplet disaggregation has crossed from research into production, and on-die fabrics now must extend across die boundaries through UCIe 2.0 transport with coherency intact. At advanced nodes, wire delay dominates path delay, and routing topologies that minimize wire length per bit become the limiting factor in fabric power and timing.

Network-on-Chip (NoC) architectures address all four. Packet-based transport isolates protocol semantics inside Network Interface Units at the IP boundary, leaving the fabric free of protocol-specific logic. Directory-based coherency replaces broadcast snoop, allowing dozens of agents to share state without bandwidth collapse. The protocol layering in AMBA CHI Rev E.b separates transport from semantics, and that separation is what scales.

SignatureIP's Compute and I/O Subsystem IP division delivers a vertically integrated NoC portfolio. C-NOC carries cache-coherent traffic for CPU clusters and accelerators. NC-NOC carries non-coherent traffic for safety-critical I/O subsystems. AXI2CHI and CHI2AXI bridges connect AXI4 initiators and subordinates to CHI fabrics. ATC handles address translation, Proxy Cache buffers at the AXI/CHI boundary, and Ethernet IP provides system-level connectivity. Inoculator, the cloud-based architectural exploration and pre-configured RTL generation tool, handles configuration complexity and emits matched RTL for every block in the composition.

II. THE BUS ERA: WHAT WORKED AND WHY IT STOPPED WORKING

A. The Shared-Bus Model (AHB/ASB)

Early ARM AMBA used a single shared bus with a centralized arbiter. AHB and its predecessor ASB worked well when SoCs contained a CPU, a small handful of peripherals, and a memory controller, all running at modest clock rates. Arbitration was simple: one master at a time, fixed priority or round-robin. The model fell apart as soon as SoC designers added a second high-bandwidth master. Every additional bus master meant more contention. AHB's pipelined two-stage protocol could not be

split across clock domains without bridging logic that defeated the protocol's bandwidth ceiling.

B. The Crossbar Transition (AXI/AXI4)

AXI replaced the shared bus with point-to-point channels: separate read address, read data, write address, write data, and write response paths. Each channel could pipeline independently. Outstanding transactions, multiple IDs, and out-of-order completion gave designers a path to high sustained bandwidth.

Scaling AXI to many masters and many slaves drove the industry to crossbar fabrics. A full AXI crossbar between N masters and M slaves is an N-by-M switch with arbiters and routing for five independent channel groups. The wire count grows quadratically. At 512-bit data widths, an 8x8 or 16x16 crossbar becomes a routing problem that dominates the floorplan.

C. ACE: Coherency Bolted Onto a Non-Coherent Protocol

ARM's AXI Coherency Extensions (ACE) added snoop channels and coherent transaction types to AXI. ACE worked in 2-cluster and small 4-cluster CPU configurations, assuming broadcast snoop. Aggregate snoop bandwidth scales as N-squared with N coherent caches. By the time SoCs began integrating CPU clusters with GPU caches, AI accelerator scratchpads, and DSP local memory, the broadcast model had become a bandwidth-hungry liability. ACE was a transitional protocol; it could not carry the industry to many-core, many-accelerator coherency.

D. Where AXI and Crossbar Break

Timing closure on long crossbar arbitration paths fails first, often around the 16x16 boundary at advanced nodes. Routing congestion forces additional metal layers, raising mask cost and power per bit. Width matching between 64-bit, 128-bit, and 256-bit IP requires adapters that consume area and contribute to deadlock risk. A 30-IP AXI subsystem at 7nm or 5nm is near the upper limit of what can be closed at competitive area and power without protocol-level redesign.

III. THE NOC IMPERATIVE: WHY PACKET-SWITCHED FABRICS WIN

A. Network Interface Units: Protocol Isolation at the Boundary

A NoC fabric carries packets, not protocol transactions. Each IP block connects through a Network Interface Unit (NIU) that translates the IP's native protocol (AXI, ACE, CHI, AHB, OCP, or custom) into packets. Adding or replacing an IP block requires changes to its NIU only. The fabric topology, routing, and arbitration logic remain unchanged. The verification surface for the fabric is bounded by the packet protocol, not by the union of every connected IP's protocol.

B. CHI's Layered Architecture

AMBA CHI Rev E.b was designed for packet-based transport from the start. The protocol separates into four layers: protocol (transaction semantics), packetization, link, and node. Each layer can be modified independently—a wider data flit can be supported without changing protocol semantics, and a different arbitration scheme can be applied at the link layer without affecting transaction ordering. This is the structural reason CHI scales where ACE did not.

C. Directory-Based Coherency at Scale

Broadcast snoop sends every coherent transaction to every cache. Directory-based coherency, used in CHI through the Home Node Fully Coherent (HN-F) function, maintains a directory of cache line ownership. For a 16-agent coherent cluster, broadcast snoop generates 16 snoop messages per transaction. Directory-based snoop generates one or two. The net result is a coherent fabric that supports 32, 64, or more coherent agents without bandwidth collapse—the regime AI accelerator SoCs and server-class chiplets actually require.

D. Adoption

Adoption tracks the technical pressure. Major commercial NoC IP vendors have moved from research engagements to volume production deployments across AI accelerator, server-class, and automotive SoC programs. The acceleration is steepest at the high end, where bandwidth and coherency requirements are not satisfiable with crossbar architectures at competitive area and power.

> *Note for marketing: if SignatureIP licenses market-research data from Gartner, IDC, IBS, Semico Research, or similar, specific market size, CAGR, and adoption percentages can be reintroduced here with full source attribution before publication.*

IV. FOUR FORCES ACCELERATING THE BUS-TO-NOC MIGRATION

A. AI and ML Workloads

AI accelerator SoCs combine CPU clusters, tensor cores, vector engines, and HBM controllers in a single die. CHI distinguishes Request Node Fully Coherent (RN-F) for full coherency, Request Node I/O coherent (RN-I) for I/O coherency, and various levels of cacheability and shareability per transaction.

A NoC fabric implementing CHI can route each agent's traffic at the coherency level that agent needs, without forcing the whole fabric into the highest common denominator. Modern AI accelerators saturate aggregate fabric bandwidth in the multi-terabyte-per-second range—crossbar fabrics at this class are not buildable at competitive area and power.

B. Automotive ADAS and EV

ISO 26262 ASIL-B through ASIL-D certification requires that hardware safety mechanisms be designed into the IP, not retrofitted. The interconnect carries every transaction in the SoC, making it a single point of safety exposure. Safety mechanisms required at the interconnect layer include SECDED ECC on all data buffers, parity on control signals, safety status registers exposed to safety supervisors, and fault injection hooks for FMEDA validation. Deterministic latency is a parallel requirement. NoC fabrics with QoS arbitration and traffic class isolation can guarantee this; bus-based fabrics cannot, because head-of-line blocking on a shared resource is unbounded under adversarial traffic.

C. Chiplet Disaggregation

Reticle limits cap monolithic die area at approximately 858 mm². UCIe 2.0, ratified in August 2024 [2], defines a die-to-die transport with raw bandwidth in the multi-TB/s range. Coherency must extend across die boundaries: a coherent NoC on die A must hand off transactions to a coherent NoC on die B, with directory state, ordering, and snoop traffic preserved across the UCIe link. This maps naturally onto packet-based NoC architectures and poorly onto crossbar fabrics.

D. Advanced Nodes

At 3 nm and 2 nm, wire delay dominates total path delay. Long crossbar arbitration paths cannot close without aggressive pipelining that defeats the latency advantage of point-to-point connections. NoC mesh and torus topologies localize traffic—each hop is short, registered, and timing-clean. Power per bit at 3 nm in a well-architected NoC is significantly lower than in an equivalent crossbar at the same bandwidth. As nodes shrink further, this gap widens.

V. SIGNATUREIP'S ANSWER: A VERTICALLY INTEGRATED NOC IP PORTFOLIO

The IP family is built around a common protocol foundation (CHI Issue E.b) and a shared verification methodology. The blocks compose into complete coherent and non-coherent subsystems with consistent QoS, safety, and integration models.

A. C-NOC: Cache-Coherent Fabric for CPU and Accelerator Clusters

C-NOC implements AMBA CHI Rev E.b as a packet-switched coherent NoC with HN-F nodes performing directory lookup, snoop filter management, and ordering enforcement. The active engineering focus is direct-transfer optimization targeting the CHI Issue E.b mechanisms: Direct Memory Transfer (DMT), Direct Cache Transfer (DCT), and Direct Write Transfer (DWT)—targeting 85 to 95 percent peak throughput in cold-miss and post-eviction regimes [1]. These optimizations allow data to flow directly between RN-F nodes or between an RN-F and memory, bypassing intermediate buffering at the HN-F.

B. NC-NOC: Non-Coherent Fabric for I/O and Peripheral Subsystems

NC-NOC is an AXI4-based NoC with configurable data widths (typically 128, 256, or 512 bits). Functional safety is built in at the IP level: internal data buffers carry SECDED ECC, control paths carry parity, safety status registers expose fault state to a safety supervisor, and fault injection hooks enable FMEDA validation. The combination supports ISO 26262 ASIL-B through ASIL-D certification. QoS arbitration with traffic class isolation provides bounded worst-case latency for ADAS sensor fusion, EV powertrain control, and other deterministic-latency workloads.

C. AXI2CHI and CHI2AXI: Protocol Bridges at the Fabric Boundary

AXI2CHI translates AXI4 transactions into CHI Issue E.b transactions, presenting the AXI initiator to C-NOC as a CHI request node. Same-ID transactions must observe AXI's ordering rule, and the bridge tracks per-ID state across the protocol boundary to avoid the deadlocks naive bridges encounter under concurrent traffic. CHI2AXI accepts CHI Issue E.b requests from C-NOC and emits AXI4 transactions to AXI subordinates. Individual AXI blocks on either side can be replaced with CHI-native equivalents over time, and migration cost is bounded by the bridges' verification rather than by full-system requalification.

D. ATC: Address Translation for DMA Safety and Virtualization

ATC provides IOMMU-style translation at the NoC interface for DMA masters, caching translations to avoid repeated page-table walks. ATC supports memory isolation requirements that automotive (FFI between ASIL partitions) and data center (multi-tenant virtualization) workloads impose.

E. Proxy Cache: Local Buffering at the AXI/CHI Boundary

When the I/O subsystem reaches into the coherent compute domain through the CHI2AXI bridge, every access is a coherent transaction that must be serviced by the HN-F. Proxy Cache provides local buffering at the boundary, absorbing read and write traffic that would otherwise round-trip through the HN-F on every transaction.

F. Inoculator: Architectural Exploration and Pre-Configured RTL Generation

Inoculator is a cloud-based architectural exploration canvas that composes SignatureIP blocks into a candidate SoC architecture. The customer assembles the architecture, sets parameters, evaluates the resulting topology against bandwidth, latency, area, and safety targets, and iterates. When the architecture is finalized, Inoculator emits pre-configured RTL for every IP in the composition. Errors that previously appeared at integration time—parameter mismatches, illegal address map regions, missing safety annotations—are caught at the canvas.

G. Ethernet IP: System Connectivity for Automotive and Data Center

Automotive applications require TSN with bounded latency, IEEE 1588 time synchronization, and credit-based shaper support. Data center and edge compute applications require high-throughput Ethernet (10G, 25G, 100G and beyond) with offload capabilities. The SignatureIP Ethernet IP family covers both regimes, attaching to NC-NOC over AXI4 and reaching the coherent domain through the CHI2AXI bridge where coherent host access is required.

VI. INTEGRATION STORY: HOW THE IP FAMILY WORKS TOGETHER

A representative integration is an automotive ADAS SoC combining a coherent compute domain (CPU clusters and AI accelerators), a non-coherent I/O domain (sensor interfaces, network controllers, storage), and TSN Ethernet connectivity.

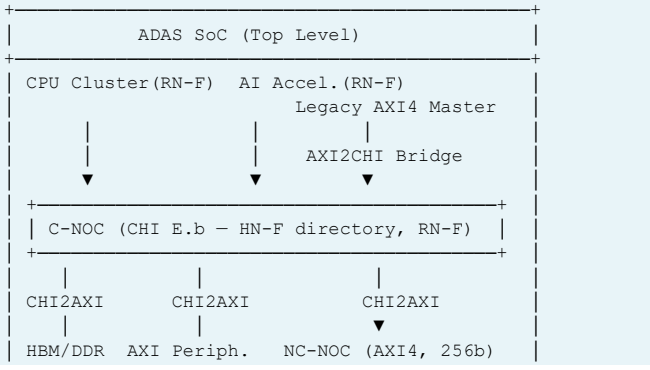

Fig. 1. Representative ADAS SoC integration. Assembled and parameterized in Inoculator; pre-configured RTL emitted for all blocks.

The CPU clusters and AI accelerators are RN-F endpoints on C-NOC, sharing the coherent address space through the HN-F directory. Legacy AXI4 master IP attaches through AXI2CHI with full ordering compliance. On the subordinate side, AXI4 memory controllers and AXI peripherals attach through CHI2AXI. The CHI fabric stays CHI-native; the AXI boundaries live at the bridges.

The non-coherent I/O subsystem is built on NC-NOC. ATC provides address translation for DMA masters, supporting virtualization and FFI between safety partitions. Proxy Cache buffers high-bandwidth I/O traffic at the AXI/CHI boundary. Ethernet IP attaches with TSN support for in-vehicle networking. The full composition is assembled and parameterized in Inoculator before any RTL is written.

The same IP family supports two other integration profiles without architectural change. An AI inference SoC drops automotive safety annotations and scales C-NOC for tens of RN-F agents. A data center accelerator extends C-NOC across UCIe die boundaries, with directory and ordering state preserved across the die-to-die link.

Customer case study. A Tier-1 automotive supplier integrated C-NOC, NC-NOC, AXI2CHI, CHI2AXI, ATC, and Proxy Cache for an ADAS SoC at an advanced foundry node using Inoculator. The fabric subsystem carried ASIL-D safety mechanisms end to end, with FMEDA validation against the configured fault injection hooks. [Customer-specific proof points to be inserted; SignatureIP marketing to confirm phrasing and disclosure level before publication.]

VII. WHAT THIS MEANS FOR SOC DESIGN TEAMS

Verification cost is the first practical implication. NoC fabrics with NIU-isolated protocol logic bound the verification surface at the NIU level, and the fabric protocol itself can be verified once and reused across SoC programs. Verification effort moves from the system level to the IP boundary.

Floorplanning flexibility follows from the same isolation. A NoC mesh routes traffic locally; IP blocks can move on the floorplan to optimize for thermal and power without forcing fabric redesign. Late-cycle floorplan changes, which are routine on real SoC programs, become tractable.

Protocol migration strategy matters for teams with substantial AXI IP investment. AXI initiator IP reaches the CHI domain through AXI2CHI. AXI subordinate IP is reached from the CHI

domain through CHI2AXI. The compute domain can go CHI-native today while the AXI ecosystem on either boundary stays in place. Replacement decisions on AXI blocks can be made per-block rather than per-system.

For design teams currently evaluating their next-generation interconnect choice, the practical path is to start with an architectural exploration in Inoculator, then engage with our team on the specific bandwidth, safety, and migration constraints of the target SoC.

VIII. CONCLUSION AND NEXT STEPS

Bus-based and crossbar interconnects served the SoC industry well during the era of single-cluster CPUs and modest accelerator counts. They cannot serve the SoCs being designed now. AI workloads require heterogeneous coherency at agent counts where broadcast snoop fails. Automotive certification requires safety mechanisms at the interconnect layer that bus protocols never carried. Chiplet integration requires fabric architectures that partition cleanly across die boundaries. Advanced nodes require routing topologies that minimize wire length per bit transferred.

SignatureIP's portfolio is built for this regime and is shipping today. C-NOC carries cache-coherent traffic at scale through CHI Issue E.b, with active direct-transfer optimization targeting 85 to 95 percent peak throughput. NC-NOC carries non-coherent and safety-critical I/O subsystem traffic with ASIL-B through ASIL-D safety mechanisms built in. AXI2CHI and CHI2AXI provide bidirectional bridging on both sides of the fabric. ATC, Proxy Cache, and Ethernet complete the subsystem. Inoculator ties the portfolio together.

Next steps for design teams evaluating interconnect IP:

- **Request access to Inoculator** and run an architectural exploration against your target SoC profile—CPU agent count, accelerator coherency model, I/O bandwidth, safety target. [signatureip.com/inoculator]
- **Schedule a technical review** with our Compute and I/O Subsystem IP team to walk through the configuration, the resulting RTL, and integration constraints specific to your design. [signatureip.com/contact]
- **Download the per-IP datasheets** for C-NOC, NC-NOC, AXI2CHI, CHI2AXI, ATC, Proxy Cache, and Ethernet to support internal evaluation. [signatureip.com/portfolio]

IX. REFERENCES

[1] Arm Limited. AMBA CHI Architecture Specification, Issue E (IHI 0050E). Reference for CHI protocol layering, RN-F/HN-F nomenclature, and the DMT/DCT/DWT direct-transfer mechanisms. https://developer.arm.com/documentation/ihi0050/latest/

[2] UCIe Consortium. Universal Chiplet Interconnect Express Specification, Version 2.0. Ratified August 2024. https://www.uciexpress.org/2-0-spec-download

[3] ISO. ISO 26262-1:2018, Road Vehicles, Functional Safety, Part 1: Vocabulary. Second edition, 2018. https://www.iso.org/standard/68383.html

About SignatureIP

SignatureIP is a provider of semiconductor IP solutions for chiplet and multi-die design across today's advanced SoC architectures. SignatureIP's network-on-chip (NoC) interconnect IP and high-speed interface IP — spanning PCIe Gen 6, CXL 3.0, and coherent and non-coherent NoC — enable system architects to build higher-performance, lower-power designs with faster time to silicon. SignatureIP's IP portfolio allows its customers to focus on product differentiation while SignatureIP handles the complexity of on-chip and die-to-die communication. Learn more at www.signatureip.ai.