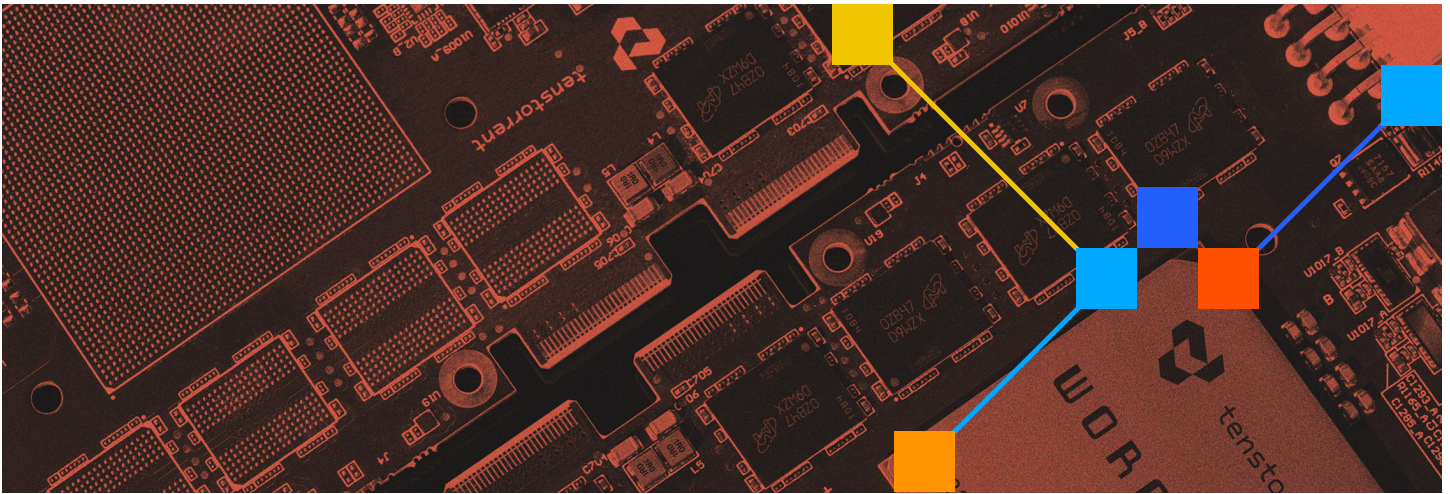


Case Study

Cache Coherent and Non-Coherent NoCs Connect AI and HPC SoCs and Chiplets



Customer Overview

Tenstorrent's high-performance RISC-V CPUs, modular chiplets, and scalable compute systems give developers full control at every layer of the stack, at any scale from a single-node experimentation to data center-scale deployment. All of their hardware is supported by open source full stack software supporting 20+ customer models running at top speeds and 700+ models running out of the box enabled by the compiler TT-Forge.

Tenstorrent believes in an open future. Their architecture and software are designed to be edited, forked, and owned. Tenstorrent's products are gaining traction and momentum in US, European, Asian, and Middle East markets with those building next-gen and sovereign AI solutions.

"We build high-performance compute AI using silicon-proven solutions that are configurable and customizable to meet the requirements of advanced AI within the chip. Arteris' NoC solutions are mature, silicon proven, and can be configured and customized based on our requirements enabling rapid SoC design."

Aniket Saha
VP of Product Management
Tenstorrent

arteris.com

Business Challenges

- To build "computers for AI" aimed at delivering unprecedented performance, scalability, and flexibility across applications including HPC/AI, automotive, and robotics.

Design Challenges

- To develop high performance intellectual property (IP) cores, including RISC-V CPU cores and Tensix AI Cores, and chiplets (building blocks of SoCs).
- To connect multiple different types of IP within its chiplets to handle the immense data traffic generated by modern AI workloads.

Arteris Solution

- FlexNoC® non-coherent NoC IP
- Ncore™ cache coherent NoC IP

Results

- FlexNoc fully addressed Tenstorrent's non-coherent NoC requirements in its current generation of chiplets.
- Ncore addressed Tenstorrent's cache coherent NoC requirements in next generation of chiplets in planning.

Tenstorrent designs chiplets: smaller, specialized, and reusable silicon building blocks. This modularity offers advantages in terms of cost, time-to-market, and the ability to mix and match different technologies. Chiplets are optimal for plug-and-play components across technologies and vendors unlocking real innovation. This vision of a "composable" hardware future is being driven by Open Chiplet Architecture (OCA), a standard that ensures interoperability between chiplets from different vendors.

The fundamental building blocks of Tenstorrent's architecture are Tensix AI Cores and high-performance Ascalon RISC-V CPU cores. Tensix Cores are highly programmable and are designed to efficiently execute the complex mathematical operations at the heart of AI models. Ascalon CPUs provide the high-performance, general-purpose compute capabilities necessary to run operating systems and manage workloads. Tenstorrent implements these cores as its foundational IP. It then uses multiple instances of its foundational IP, plus other IP blocks to construct its own chiplets for compute, memory, and I/O.

The Challenge

One of the challenges Tenstorrent faces lies in managing the data traffic generated by AI workloads within and across chiplets. As the company looks to incorporate next-generation memory standards like GDDR7, which promise improved bandwidth, the demands on the internal fabric of the chiplets will be high. The GDDR7 specification, with its high-speed PAM-3 signaling, requires pristine integrity and a meticulously designed physical interface (PHY).

When the Tenstorrent team originally set out to build the highest-performing AI solutions available, they decided to base their designs on open-source solutions, develop differentiating IP in-house, and employ trusted third-party IP. Key to this approach is highly configurable cache coherent and non-coherent Network-on-Chips (NoCs) that scale up and down as required.

The Solution

Tenstorrent required mature, high-bandwidth low-latency fabrics that were stable and silicon-proven. Arteris offers Ncore cache coherent and FlexNoC non-coherent NoC IPs, both of which can be easily configured to address the company's requirements. These NoCs support the customizability needed for various use cases, from automotive AI to data center deployments, ensuring large volumes of data are moved quickly with minimal latency.

"In the high-end HPC and AI space, especially for data centers, you need something that can guarantee quality-of-service (QoS) between cores. You also need error detection and correction technologies, and future applications may require functional safety (FuSa) technologies. Arteris NoCs have all these capabilities."

Aniket Saha
VP of Product Management
Tenstorrent

Results and Future Plans

The Arteris FlexNoC non-coherent NoC IP fully addressed the non-coherent requirements for Tenstorrent's compute, memory, and I/O chiplets. In addition, Tenstorrent is planning to deploy Arteris Ncore coherent NoC in its next generation chiplets.

Using Arteris FlexNoC, Tenstorrent's memory chiplet can meet streaming read / write bandwidth expectation at 144GB/s as specified in the JEDEC GDDR7 memory specifications.

Based on the Tenstorrent team's experience with Arteris technologies and support from Arteris, the company intends to use Arteris technology in future generations of products on its roadmap.

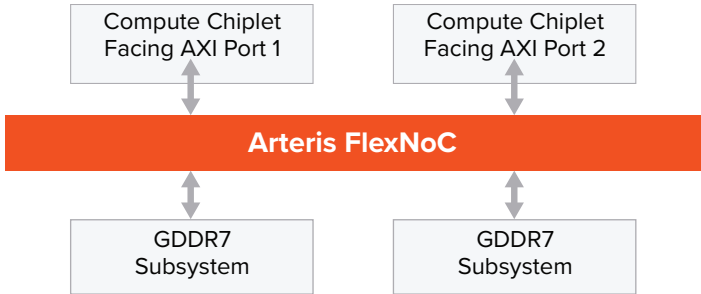


Figure 1: FlexNoC for Tenstorrent's chiplet

"A big aspect of IP is the level of support provided by the vendor. Many people don't realize how difficult it is to support customers who are building their own IP. Customers often customize the product, use it in different configurations, and run into various issues. Our team has been very happy with working with the Arteris team, especially the support group, which has been extremely responsive."

Aniket Saha
VP of Product Management
Tenstorrent

Scenario	Source	Target	BW Expectation	Utilization
Streaming Read	Compute Chiplet Facing AXI Port 0	DRAM	144GB/s read data	98%~100%
Streaming Write	Compute Chiplet Facing AXI Port 0	DRAM	144GB/s write data	
Streaming Read	Compute Chiplet Facing AXI Port 1	DRAM	144GB/s read data	
Streaming Write	Compute Chiplet Facing AXI Port 1	DRAM	144GB/s write data	

Figure 2. Streaming read / write bandwidth expectations

Tenstorrent Product Summary

IP (Ascalon / Tensix-Neo)



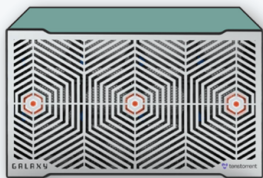
- Scales from mW to MW for efficiency and performance
- IP available for licensing
- Industry-leading performance
- Modular design available in varied configurations

Chips & Chiplets



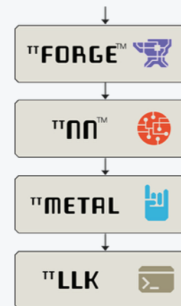
- Portfolio of cards powered by scalable Tensix AI cores
- Inference and Training, CNN and NLP, Recommendation Engines, all on the same silicon
- Hardware available for purchase, as well as IP available for licensing
- Multi-component modular chiplets

Servers (Galaxy)



- Galaxy Server – 32 high performance cards in a custom chassis – shipping now in 2025
- Servers are easily combined into a Galaxy Rack for high bandwidth chip-to-chip connectivity

Software



- ML compilers that scale from one chip to thousands
- TT-Forge | Automated AI/ML Compiler
- TT-NN | Operator Library
- TT-Metalium | Bare Metal Software Stack
- TT-LLK | Low Level Kernels

About Arteris

Arteris is a global leader in system IP used in semiconductors to accelerate the creation of high-performance, power-efficient silicon. Arteris network-on-chip (NoC) interconnect IP and system-on-chip (SoC) integration automation software are used by the world's top semiconductor and technology companies to improve overall performance, engineering productivity, reduce risk, lower costs, and bring complex designs to market faster.

[Learn more at arteris.com](https://www.arteris.com).

arteris-support@arteris.com +1 408 470 7300

Copyright © 2004-2025 Arteris, Inc. All rights reserved worldwide. Arteris, Arteris IP, the Arteris IP logo and the other Arteris marks found at <https://www.arteris.com/trademarks> are trademarks or registered trademarks of Arteris, Inc. or its subsidiaries. All other trademarks are the property of their respective owners.