Broadcom Ships Tomahawk Ultra: Reimagining the Ethernet Switch for HPC and AI Scale-up

July 15, 2025

Ultra-low Latency, 64B Line-rate Switching, Lossless Fabric and In-Network Collectives Define a New Level of Performance for Ethernet in HPC and AI Scale-up

PALO ALTO, Calif., July 15, 2025 (GLOBE NEWSWIRE) -- Broadcom Inc. (NASDAQ:AVGO), a global leader in semiconductor and infrastructure software solutions, today announced the shipment of its breakthrough Ethernet switch — the Tomahawk Ultra. Engineered to transform the Ethernet switch for high-performance computing (HPC) and AI workloads, Tomahawk Ultra delivers industry-leading ultra-low latency, massive throughput, and lossless networking.

"Tomahawk Ultra is a testament to innovation, involving a multi-year effort by hundreds of engineers who reimagined every aspect of the Ethernet switch," said Ram Velaga, senior vice president and general manager of Broadcom's Core Switching Group. "This highlights Broadcom's commitment to invest in advancing Ethernet for high-performance networking and AI scale-up."

Shattering Myths, Redefining Performance

Built from the ground up to meet the extreme demands of HPC environments and tightly coupled AI clusters, Tomahawk Ultra redefines what an Ethernet switch can deliver. Long perceived as higher-latency and lossy, Ethernet takes on a new role:

- **Ultra-low latency**: Achieves 250ns switch latency at full 51.2 Tbps throughput.
- **High performance**: Delivers line-rate switching performance even at minimum packet sizes of 64 bytes, supporting up to 77 billion packets per second.
- **Adaptable, optimized Ethernet headers**: Reduces header overhead from 46 bytes down to as low as 10 bytes, while maintaining full Ethernet compliance —boosting network efficiency and enabling flexible, application-specific optimizations.
- **Lossless fabric**: Implements Link Layer Retry (LLR) and Credit-Based Flow Control (CBFC) to eliminate packet loss and ensure reliability.

"AI and HPC workloads are converging into tightly coupled accelerator clusters that demand supercomputer-class latency — critical for inference, reliability, and in-network intelligence from the fabric itself," said Kunjan Sobhani, lead semiconductor analyst, Bloomberg Intelligence. "Demonstrating that open-standards Ethernet can now deliver sub-microsecond switching, lossless transport, and on-chip collectives marks a pivotal step toward meeting those demands of an AI scale-up stack — projected to be double digit billions in a few years."

Built for HPC and AI Scale-Up

Tomahawk Ultra is optimized for the tightly coupled, low-latency communication patterns found in both high-performance computing systems and AI clusters. With ultra-low latency switching and adaptable optimized Ethernet headers, it provides predictable, high-efficiency performance for large-scale simulations, scientific computing, and synchronized AI model training and inference.

When deployed with Scale-Up Ethernet (SUE specification available to the public here), Tomahawk Ultra enables sub-400ns XPU-to-XPU communication latency, including the switch transit time — setting a new benchmark for tightly synchronized AI compute at scale.

By reducing Ethernet header overhead from 46 bytes to just 10 bytes, while maintaining full Ethernet compliance, Tomahawk Ultra dramatically improves network efficiency. This optimized header is adaptable per application, offering both flexibility and performance gains across diverse HPC and AI workloads.

Tomahawk Ultra incorporates lossless fabric technology that eliminates packet drops during high-volume data transfer. Incorporating LLR, the switch detects link errors using Forward Error Correction and automatically retransmits packets, avoiding drops at the wire level. Simultaneously, CBFC prevents buffer overflows that traditionally caused packet loss. Together, these mechanisms create a truly lossless Ethernet fabric, delivering the level of reliability demanded by today's most data-intensive workloads.

Tomahawk Ultra also accelerates performance through In-Network Collectives solving one of the most persistent bottlenecks in AI and machine learning workloads. Rather than burdening XPUs with collective operations like AllReduce, Broadcast, or AllGather, Tomahawk Ultra executes these directly within the switch chip. This can reduce job completion time and improve utilization of expensive compute resources. Importantly, this capability is endpoint-agnostic, enabling immediate adoption across a wide range of system architectures and vendor ecosystems.

Designed with innovations in topology-aware routing to support advanced HPC topologies including Dragonfly, Mesh and Torus, Tomahawk Ultra is also compliant with the UEC standard and embraces the openness and rich ecosystem of Ethernet networking.

Introducing SUE-Lite

As part of Broadcom's Ethernet-forward strategy for AI scale-up, the company has introduced SUE-Lite — an optimized version of the SUE specification tailored for power and area-sensitive accelerator applications. SUE-Lite retains the key low-latency and lossless characteristics of full SUE, while further reducing the silicon footprint and power consumption of Ethernet interfaces on AI XPUs and CPUs.

This lightweight variant enables easier integration of standards-compliant Ethernet fabrics in AI platforms, promoting broader adoption of Ethernet as the interconnect of choice in scale-up architectures.

Platform for AI Scale-Up and HPC Scale-Out

Together with the 102.4 Tbps Tomahawk 6, Tomahawk Ultra forms the foundation of a unified Ethernet architecture: enabling scale-up Ethernet for AI, and scale-out Ethernet for HPC and distributed workloads.

Now Shipping

Tomahawk Ultra is 100% pin-compatible with Tomahawk 5, ensuring a very fast time-to-market. It is shipping now for deployment in rack-scale AI training clusters and supercomputing environments. To learn more about the Broadcom Tomahawk Ultra family click here. Explore the full Scale-Up/Scale-Out media kit for resources and insights into Broadcom's scalable solutions here.

About Broadcom
Broadcom Inc. (NASDAQ: AVGO) is a global technology leader that designs, develops, and supplies a broad range of semiconductor, enterprise software and security solutions. Broadcom's category-leading product portfolio serves critical markets including cloud, data center, networking, broadband, wireless, storage, industrial, and enterprise software. Our solutions include service provider and enterprise networking and storage, mobile device and broadband connectivity, mainframe, cybersecurity, and private and hybrid cloud infrastructure. Broadcom is a Delaware corporation headquartered in Palo Alto, CA. For more information, go to www.broadcom.com.

Broadcom, the pulse logo, and Connecting everything are among the trademarks of Broadcom. The term "Broadcom" refers to Broadcom Inc., and/or its subsidiaries. Other trademarks are the property of their respective owners.

Press Contact:
Jon Piazza
Global Communications
press.relations@broadcom.com
Telephone: +1 310 498 5254

Industry Quotes

Michael KT Lee, Senior Vice President, Research and Development Center, Accton
"Networking needs within an XPU node are as critical as those between nodes. With 51.2 Tbps Ethernet switching, 250-ns latency, credit-based flow control, and configurable optimized header, the Tomahawk Ultra is a perfect solution for building high-bandwidth, high-reliability, high-efficiency, and low-latency lossless systems ready for scale-up AI and HPC applications. Accton is excited to embrace the launch and collaborate closely with Broadcom to bring the Tomahawk Ultra solutions to the market."

Forrest Norrod, Executive Vice President and General Manager, Data Center Solutions Group, AMD
"Low latency is essential to unleashing the full potential of AI — from reducing training times to powering real-time inference. By combining Broadcom's new Tomahawk Ultra switch with AMD Instinct ™ GPUs and EPYC ™ processors, we're enabling high-performance, standards-based Ethernet solutions for AI infrastructure. Together, we're advancing an open ecosystem that brings our vision of AI everywhere, for everyone, closer to reality."

Simon Capper, Principal Engineer for AI networking, Arista Networks
"Arista appreciates the combination of ultra-low latency and scale-Up Ethernet innovations of Tomahawk Ultra for AI networking. Once again Broadcom is setting the pace in the AI and the switch industry."

Shekar Ayyar, Chairman and CEO, Arrcus
"The launch of Broadcom's Tomahawk Ultra marks a groundbreaking advancement in Ethernet innovation, particularly for AI and HPC scale-up environments. With ultra-low latency and a lossless fabric, it significantly accelerates job completion times — critical for modern AI workloads. At Arrcus, we're proud to champion an open, standards-based networking ecosystem. Combined with our high-performance ArcOS network operating system, customers can unlock scalable infrastructure that is both flexible and future-ready."

Wangson Wang, General Manager of Data Networks Infrastructure, Delta Electronics
"Delta Electronics is constantly looking ahead, and we're thrilled to confirm that our 51.2T Ethernet switch platform is ready to harness the full power of Broadcom's Tomahawk Ultra chip. We see Tomahawk Ultra as a game-changer for AI scale-up and HPC. The collaboration between Delta and Broadcom demonstrates our dedication to pushing the boundaries of what is possible in Data Center network infrastructure. Building on the success of Delta's current 800G switches, the newly launched Tomahawk Ultra chips enable us to deliver advanced solutions that offer not only unmatched performance and efficiency, but also high reliability and scalability for the most demanding network workloads — supporting rapid AI/ML network deployments for our customers."

Praveen Jain, Senior Vice President and General Manager, AI Clusters and Cloud Ready Data Center, HPE Networking
"HPE is committed to delivering open, high-performance and easy-to-manage Ethernet-based solutions for the modern data center. We commend Broadcom on its new offering, and its ultra-low latency, high throughput and support for in-network collectives align perfectly with what today's workloads demand. It reflects our shared vision for building the most advanced and open data center infrastructure solutions with operational simplicity at its core."

Saurabh Kulkarni, Vice President, AI Technical Product Management, Intel
"Broadcom's Tomahawk Ultra Series with its high throughput and ultra-low latency enables all-to-all connectivity across up to 64 Intel ® Gaudi® 3 AI accelerators per rack with total HBM bandwidth of 76.8TB/s, capable of scaling the connectivity across multiple racks. This rack-level bandwidth unlocks new possibilities for training and real-time inference of the most complex LLMs, redefining industry SLAs. Through our collaboration with Broadcom, Intel is showcasing the open architecture and modular design advantage and full capability of our rack scale platform built for large-scale, global AI deployments."

Vincent Lin, General Manager, Inventec EBG
"Inventec congratulates Broadcom on the launch of Tomahawk Ultra Ethernet switch, which significantly enhances the efficiency and sustainability of AI solutions by delivering the industry's lowest switch latency, 250 nanoseconds, and leading power efficiency with 800W at 51.2T performance. At Inventec, our vision is to develop cutting-edge artificial intelligence products that drive sustainable change for humanity and the environment through close partnership with Broadcom to deliver high-performance, scalable solutions, supporting customers' evolving AI and high-performance computing

needs.”

Kiyo Oishi, CEO, IPI
“The Tomahawk Ultra represents a bold leap forward in AI workloads and HPC clusters, delivering an unmatched combination of bandwidth, latency, and cutting-edge features like In-Network Collectives and scale-up Ethernet. By leveraging non-proprietary Ethernet, the Tomahawk Ultra will empower customers to scale their data intensive applications with unparalleled performance, efficiency, and reliability — paving the way for groundbreaking innovations in data-intensive computing.”

Andrew Qu, CEO, Micas Networks
“Broadcom’s Tomahawk Ultra is a major step forward for scale-up Ethernet in AI and HPC. With 250ns latency, 51.2 Tbps switching, and advanced features like Link Layer Retry, In-Network Collectives, and the AI Fabric Header, it delivers the performance, reliability, and efficiency our customers need for AI at scale. Thanks to pin compatibility with Tomahawk 5, Micas can rapidly bring Tomahawk Ultra-based systems to market, enabling seamless upgrades to meet the demands of next-generation AI infrastructure.”

Anshul Sadana, Founder and CEO, Nexthop AI
“With Tomahawk Ultra, Broadcom has driven AI Networking to a new level, allowing us to enable a new generation of low latency and lossless scale-up Ethernet solutions. Along with Nexthop SONiC, we now offer some of the most efficient scale-up and UEC compatible scale-out Ethernet solutions for the world’s largest hyperscalers.”

Mike Yang, President, Quanta Cloud Technology
“At QCT, we are committed to delivering next-generation AI and HPC infrastructure that meets the demands of extreme scale, performance, and efficiency. Broadcom’s Tomahawk Ultra Ethernet switch is a game-changer for the AI era, enabling 51.2 Tbps of switching capacity with ultra-low 250ns latency to dramatically accelerate AI training and inferencing workloads. We are excited to continue collaborating with Broadcom to push the next frontier of AI with Ethernet-based infrastructure.”

Vincent Ho, CEO of UfiSpace
“The Tomahawk Ultra delivers high performance and full pin-to-pin compatibility with Tomahawk 5. This seamless upgrade path shortens our development cycle for next-generation platforms, and we’re excited to integrate it into our upcoming solutions.”

Robert CL Lin, President of Enterprise and Networking Business Group, Wistron
“Broadcom’s Tomahawk Ultra sets a new benchmark in Open Ethernet for AI and HPC. Designed for GPU scale-up, the Tomahawk Ultra achieves 250ns latency at 51.2 Tbps, supporting 64B line-rate switching and lossless fabrics. This innovation represents a significant step forward for the industry. Wistron is seamlessly aligning these scalable AI systems, and the Tomahawk Ultra solution offering.”

Johnson Hsu, Senior Vice President and General Manager, WNC
“We’re proud to partner with Broadcom on the innovative Tomahawk Ultra. Purpose-built for the demands of AI and HPC, this advanced platform combines high performance with open Ethernet flexibility — enabling our customers to deploy scalable, reliable, and future-ready networks.”



Source: Broadcom Inc.