# Innovations in IP market driven by AI demands



The interface intellectual property (IP) market is experiencing significant innovations primarily driven by the escalating demands of artificial intelligence (AI) workloads. As documented by Semiconductor Engineering, AI model parameters are doubling approximately every four to six months, which creates an urgent requirement for advancements in hardware to bridge the gap between the rapid expansion of AI capabilities and the slower advancements traditionally dictated by Moore’s Law, which recommends a cycle of 18 months.

This growing demand places immense pressure on hardware design, resulting in a necessity for more robust computational capacities, enhanced resource availability, and higher bandwidth interconnects. Both Central Processing Units (CPUs) and Graphics Processing Units (GPUs) are now facing challenges as their performance has exceeded the standard reticle limits. Consequently, a new ultra-efficient networking infrastructure tailored to support AI workloads and high-performance computing (HPC) is required. This new infrastructure must ensure low latency, high-density connections, and optimal energy efficiency for effective chip-to-chip communication.

In response to these challenges, new standards are emerging to facilitate efficient hardware development. A key focus is on scaling up and out chip-to-chip architectures. The industry is transitioning from monolithic dies to systems incorporating multiple dies, facilitated by parallel interfaces such as High Bandwidth Memory (HBM) and Universal Chiplet Interconnect Express (UCIe). These solutions allow for more versatile compute architectures, operating in conjunction with established connections through PCI Express (PCIe) and Compute Express Link (CXL) for memory expansion, as well as Ethernet for extensive network architectures.

As the demands of AI and HPC continue to amplify, two notable standards have been developed: Ultra Ethernet and Ultra Accelerator Link (UALink). Ultra Ethernet has been crafted as an open, high-performance architecture that supports the scaling of AI workloads, with backing from major industry players, including providers of switches, networking devices, semiconductors, and system providers. Speaking to Semiconductor Engineering, Jon Ames, a senior staff product marketing manager for Ethernet IP at Synopsys, elaborated on how the Ultra Ethernet Consortium (UEC) technology addresses limitations found in traditional networking systems. These include constraints created by in-order packet delivery and inefficient load balancing structures, which can lead to significant performance costs for AI operations.

Ultra Ethernet is structured into clusters comprising nodes and fabric infrastructure. As detailed in the article, nodes connect via advanced Fabric Interfaces and function in two operational modes—Parallel Job Mode, which facilitates simultaneous communication for high-performance computing tasks, and Client/Server Mode, which manages storage tasks through focused communication between nodes.

In parallel, UALink has emerged as a scale-up fabric that provides a network of high-bandwidth connections designed specifically for large networks of AI accelerators. UALink enables rapid and efficient memory access among multiple interconnected accelerators, allowing the network to function as a cohesive, single unit. This interconnectivity is crucial as accelerators can share memory resources effectively, thus reducing overall latency and enhancing performance.

Both Ultra Ethernet and UALink offer substantial technical advancements. Ultra Ethernet supports compatibility with IEEE 802.3 and introduces innovative features such as link-level retry for lossless transmissions. On the other hand, UALink boasts capabilities such as high bandwidth of up to 200 Gbps per lane and specific memory-sharing functionalities that cater to high-performance AI applications.

Synopsys has positioned itself at the forefront of these advancements with its offerings of UALink and Ultra Ethernet IP solutions, providing significant bandwidth and low-latency communication essential for developing expansive AI accelerator clusters.

As the AI landscape grows, the adoption of these standardised interfaces will be pivotal for driving innovation and complexity reduction in business practices. The establishment of collaborative, open-standard solutions marks a critical phase in enhancing both the efficiency and performance of AI infrastructure, ensuring that enterprises can leverage these technologies effectively in their operations.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

* <https://www.gsdvs.com/post/will-moore-s-law-be-impacted-by-ai> - This article explains how AI model parameters are increasing rapidly, creating a need for advancements in hardware to keep up with the demands of AI workloads, and how this impacts Moore’s Law.
* <https://ourworldindata.org/grapher/exponential-growth-of-parameters-in-notable-ai-systems> - This source details the exponential growth of parameters in AI models, which is relevant to the increasing computational demands of AI workloads.
* <https://www.evalueserve.com/blog/why-ai-is-transforming-intellectual-property/> - This article discusses the impact of AI on intellectual property processes, highlighting the need for advanced computational capacities to handle AI-driven tasks.
* <https://www.gsdvs.com/post/will-moore-s-law-be-impacted-by-ai> - This article elaborates on how AI is driving innovations in hardware, including the development of AI-specific chips and the optimization of semiconductor manufacturing processes.
* <https://www.gsdvs.com/post/will-moore-s-law-be-impacted-by-ai> - It discusses the role of AI in extending Moore’s Law by optimizing chip design and exploring new computing paradigms.
* <https://www.evalueserve.com/blog/why-ai-is-transforming-intellectual-property/> - This source highlights the need for robust computational capacities and efficient data processing to support AI workloads in intellectual property services.
* <https://ourworldindata.org/grapher/exponential-growth-of-parameters-in-notable-ai-systems> - It explains the challenges posed by the increasing number of parameters in AI models, such as the need for more advanced and efficient hardware.
* <https://www.gsdvs.com/post/will-moore-s-law-be-impacted-by-ai> - This article mentions the transition from monolithic dies to systems incorporating multiple dies, facilitated by parallel interfaces like HBM and UCIe, to address the computational demands of AI.
* <https://www.gsdvs.com/post/will-moore-s-law-be-impacted-by-ai> - It discusses the emergence of new standards and technologies, such as Ultra Ethernet and UALink, to support high-performance computing and AI workloads.
* <https://iplawusa.com/revolutionizing-the-future-the-role-of-intellectual-property-in-ai-innovation/> - This source emphasizes the importance of efficient hardware development and standardized interfaces for supporting AI innovation and high-performance computing.
* <https://www.gsdvs.com/post/will-moore-s-law-be-impacted-by-ai> - It highlights the role of companies like Synopsys in providing IP solutions for Ultra Ethernet and UALink, ensuring significant bandwidth and low-latency communication for AI accelerator clusters.
* <https://semiengineering.com/how-ultra-ethernet-and-ualink-enable-high-performance-scalable-ai-networks/> - Please view link - unable to able to access data