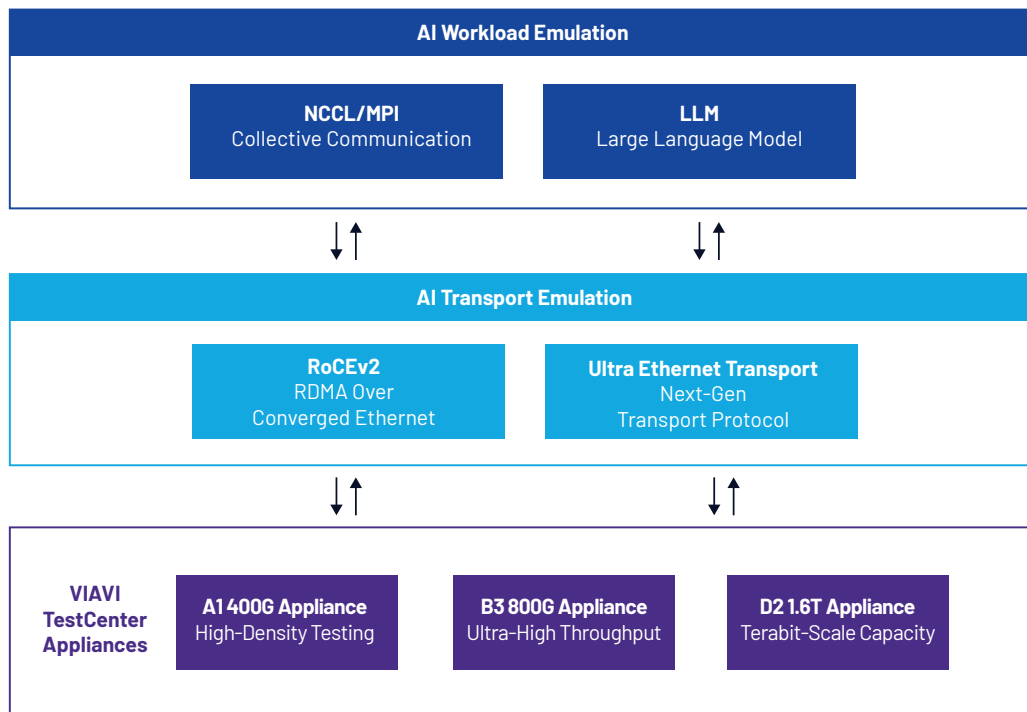


# Understanding the Dynamics of AI Fabric Testing

## The Evolution of AI Back-End Fabric Validation

AI infrastructure scalability has hit a wall. As clusters scale to tens of thousands of GPUs, the backend fabric has become the primary bottleneck. Traditional testing methods that rely on simple bandwidth checks are no longer enough to guarantee performance and resiliency of AI fabrics.

To achieve production-grade AI at scale, organizations must pivot to AI workload aware, GPU-free validation. By leveraging high-fidelity AI workload traffic emulation over technologies like **RoCEv2** and **Ultra Ethernet Transport (UET)**, teams can now validate their infrastructures under real-world conditions without tying up millions of dollars in expensive GPU resources. This delivers efficient, reliable, and cost-effective validation of next generation AI data center fabrics.



2371.900.0626

AI Workload Emulation for Next-Gen AI Fabrics

## Performance Metrics for AI Fabric Validation

### Throughput Considerations and Tail Latency

- **Business Value:** Distributed AI training lives and dies by synchronization. Even a 10% drop in RDMA throughput can inflate training costs by 15-20%, representing millions in wasted computational spend.
- **Technical Focus:** Benchmark high-flow environments to ensure load balancing via ECMP, Adaptive Scheduling, or Flowlet switching/Packet spray which prevents bottlenecks.
- **Primary KPIs:** JCT, Tail Latency (P95/P99/P99.9), Queue Pair (QP) scaling, Flow establishment rates, out-of-sequence packets, and packet latency.

### Congestion Control in Low-Latency, Lossless Fabrics

- **Business Value:** AI back-end networks need to handle training orchestration and data pipelines but also face unique congestion patterns such as microburst traffic and massive parameter synchronization (all-reduce operations) with bulk gradient transfers. Any packet loss or high packet latency can considerably slow down the training process, thus increasing operational costs by many folds.
- **Technical Focus:** Use in-network telemetry for efficient congestion avoidance and adaptive per flow congestion control. Optimize ECN/PFC threshold ratios, such that congestion is managed primarily at pre flow level using congestion control algorithms and not at port level using PFC. This reduces flow victimization and is a step towards low latency lossless fabric.
- **Primary KPIs:** ECN percentage, Sender window size, Receiver Credit Utilization, Packet Latency, packet drop count, PFC count.

### Link Layer Performance and Workload Metrics

- **Business Value:** In AI, every microsecond is a dollar. Sub-microsecond retransmission at the link layer is 100x faster than transport-layer retries, directly impacting your Job Completion Time (JCT).
- **Technical Focus:** Transition from standard PFC to Credit-Based Flow Control (CBFC) to minimize the frequency of head-of-line blocking.
- **Primary KPIs:** Analyze Bus Bandwidth (BusBW) and JCT decomposition to separate network-induced delays from computational bottlenecks.

## Scalable, GPU-Free Validation Approaches

Using GPUs for testing AI fabrics is not efficient or economically viable. VIAVI provides the industry's most comprehensive suite for validating next-gen AI infrastructures:

- **GPU-free Stress Testing:** Emulate massive CCL and LLM traffic patterns without tying up expensive accelerator resources.
- **High-Fidelity Emulation:** Replicate production incast and microburst scenarios over RoCEv2 and UET.
- **Network Benchmarking:** Gain real-time insights of AI fabric behavior using real time stats like Tail Latency (P95/P99/P99.9) JCT, BusBW, out-of-order packets, packet latency and more.

## Operational Implications for AI Fabric Validation

Organizations leveraging automated, high-fidelity fabric testing realize immediate operational advantages:

- **De-Risk Upgrades:** Validate infrastructure changes in the lab, not production, to eliminate downtime and costly deployment failures.
- **Proactive Defense:** Automated continuous validation alerts operators to performance degradation before it impacts active training jobs.
- **Cost Containment:** Prevent multi-million-dollar training resets by ensuring consistent, high-performance infrastructure – from day one.

## The Future of AI Infrastructures

### Ultra Ethernet Transport: The Next Frontier

The development of UET represents the next phase in network infrastructure development. UET addresses the growing complexity of managing multiple protocols and traffic types within modern data centers.

### The Convergence Vision

UET emerged from the recognition that even RoCEv2 deployments often required multiple network fabrics for different purposes. Storage traffic, management traffic, and AI training traffic each had distinct requirements that traditional approaches addressed through separate physical networks. The vision behind UET involves creating a single, unified fabric capable of handling all traffic types with appropriate service levels.

The era of “simple connectivity” is over. As models grow in complexity, the network is no longer just a pipe, it is the backbone of AI success. The convergence of RoCEv2, UET, and sophisticated testing methodologies is now a competitive necessity.

The future belongs to organizations that treat their AI back-end fabric as a strategic asset. By mastering GPU-free AI workload-aware validation today, you aren't just testing a network, you're building the foundation for the next wave of AI breakthroughs.

**Ready to harden your AI infrastructure? Explore our [AI Data Center Networking Test solutions](#) today.**

### Sources Cited

- RDMA over Ethernet for Distributed AI Training at Meta Scale 2024, <https://cs.stanford.edu/~keithw/sigcomm2024/sigcomm24-final246-acmpaginated.pdf>
- Microsoft Research: “RDMA for AI Training at Scale” presentation detailing Azure’s ROCEv2 deployment and performance metrics. [Microsoft Research](#)
- NVIDIA GTC 2023 talk: “Networking Innovations for Large Scale AI Training,” covering latency and CPU utilization improvements with ROCEv2. [NVIDIA GTC](#)
- OpenAI blog post on infrastructure for GPT-4 training mentioning Azure’s network performance enabling large-scale model training. [OpenAI](#)



Contact Us

**+1844 GO VIAVI**  
(+1 844 468 4284)

To reach the VIAVI office nearest you, visit [viasolutions.com/contact](https://viasolutions.com/contact)

© 2026 VIAVI Solutions Inc.

Product specifications and descriptions in this document are subject to change without notice. Patented as described at [viasolutions.com/patents/aifabrictesting-fly-hse-nse-ae](https://viasolutions.com/patents/aifabrictesting-fly-hse-nse-ae) 30195101 900 0626

[viasolutions.com](https://viasolutions.com)