

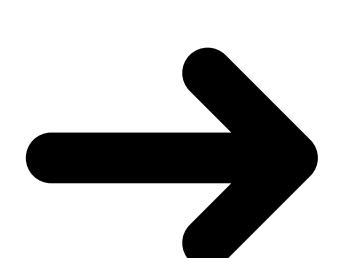
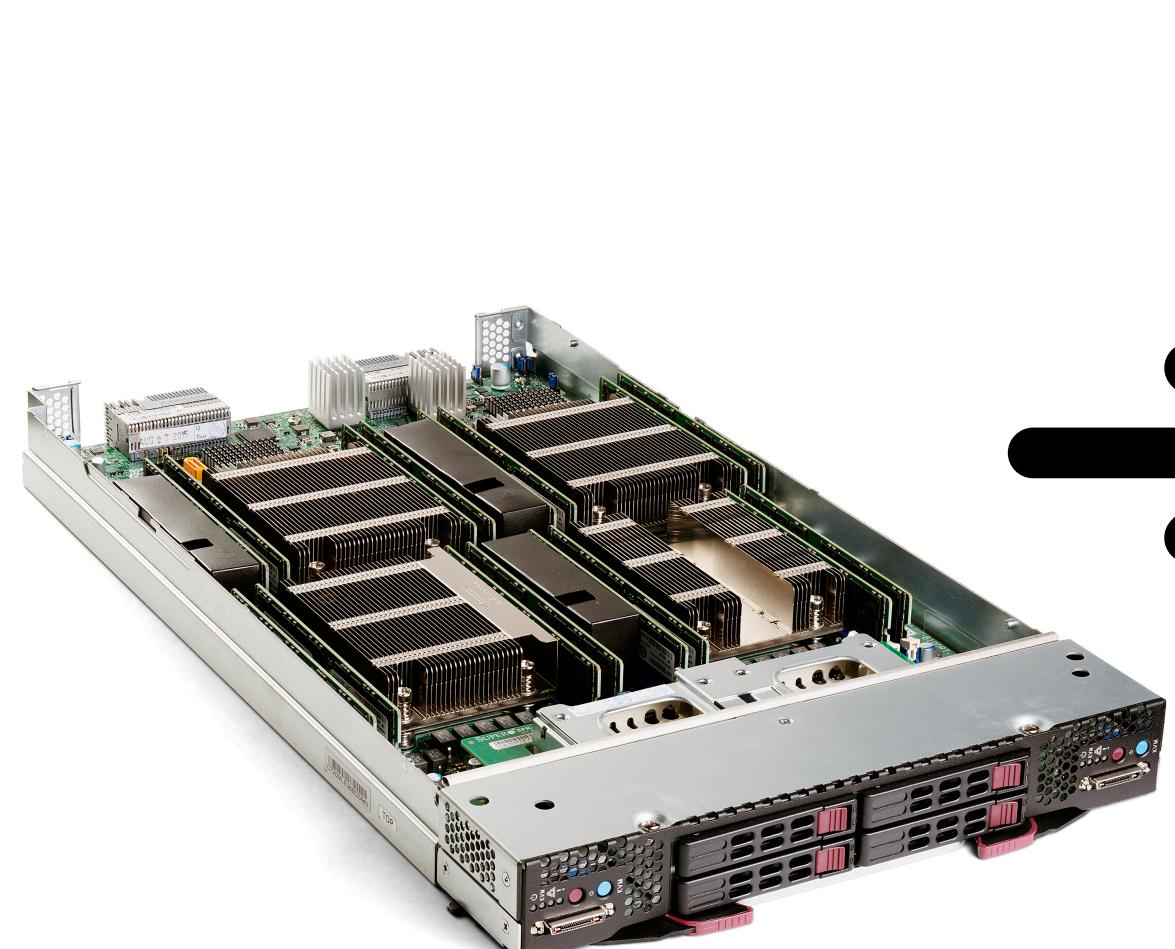
EDM: An Ultra-low Latency Ethernet Fabric for Memory Disaggregation

Weigao Su, Vishal Shrivastav

PURDUE UNIVERSITY

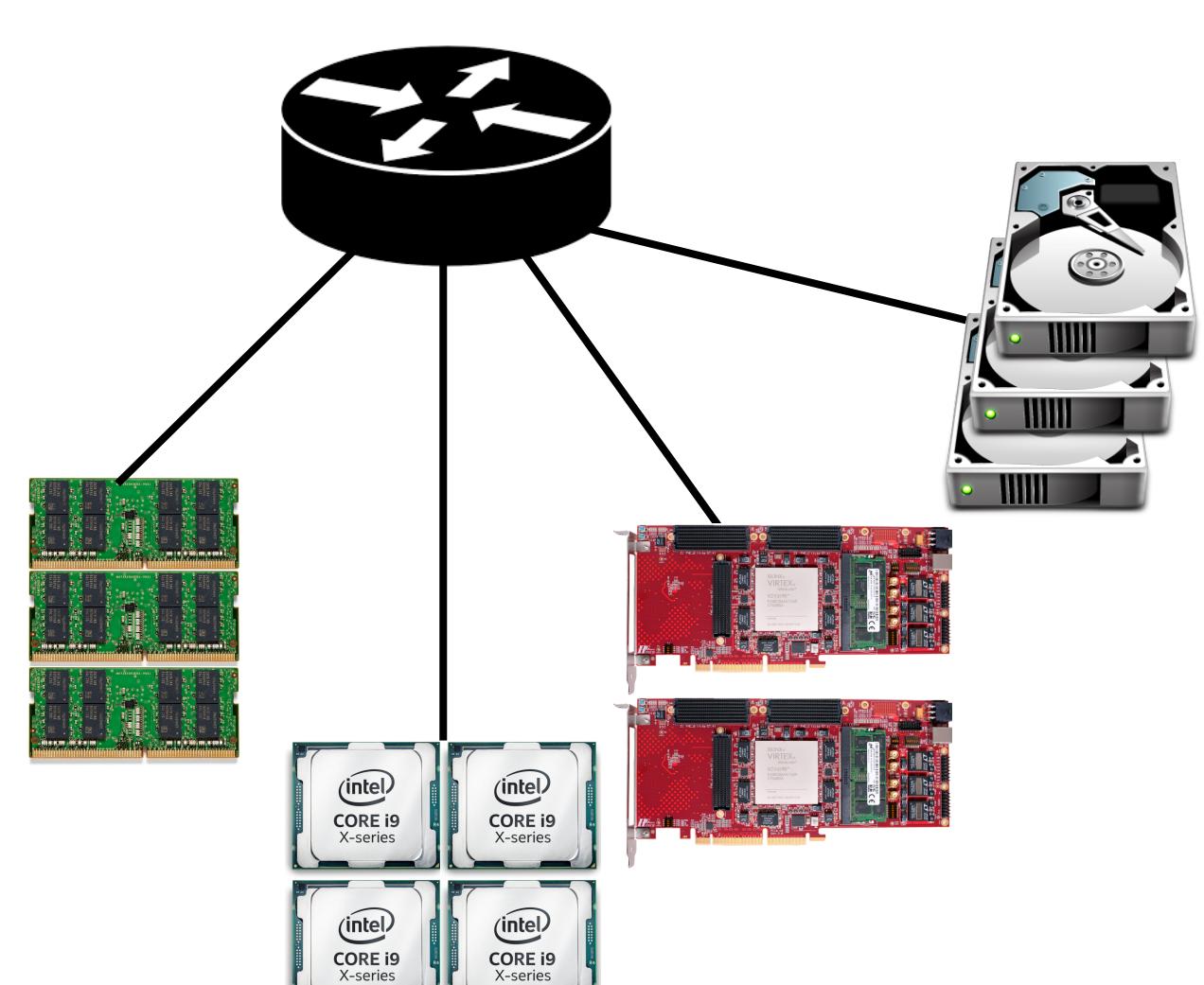
Why disaggregation?

- Surging need for memory
- Constrained monolithic blade



Why Ethernet?

- High bandwidth (Tbps);
- Easy management, etc...

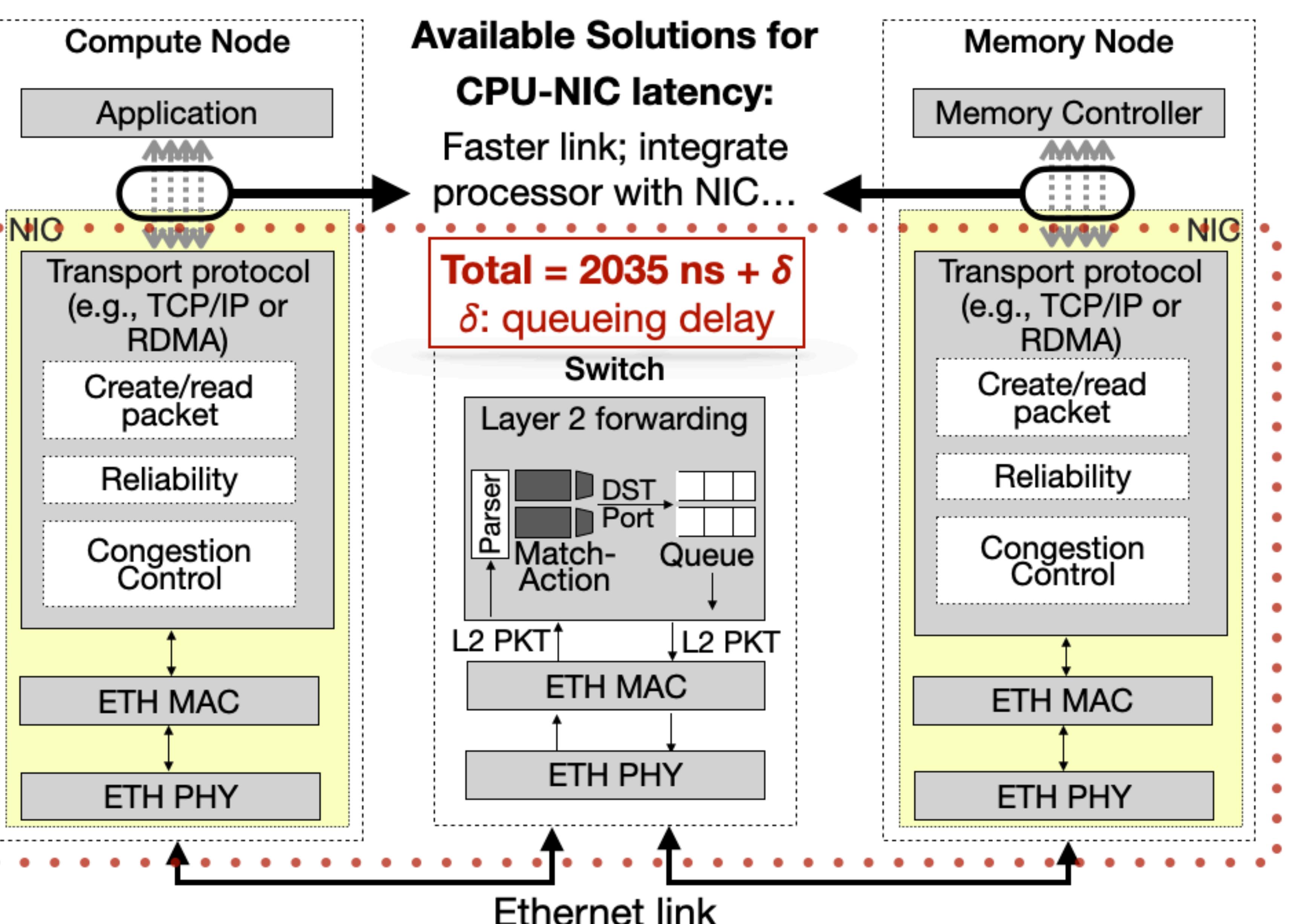


Existing Memory Disaggregation over Ethernet

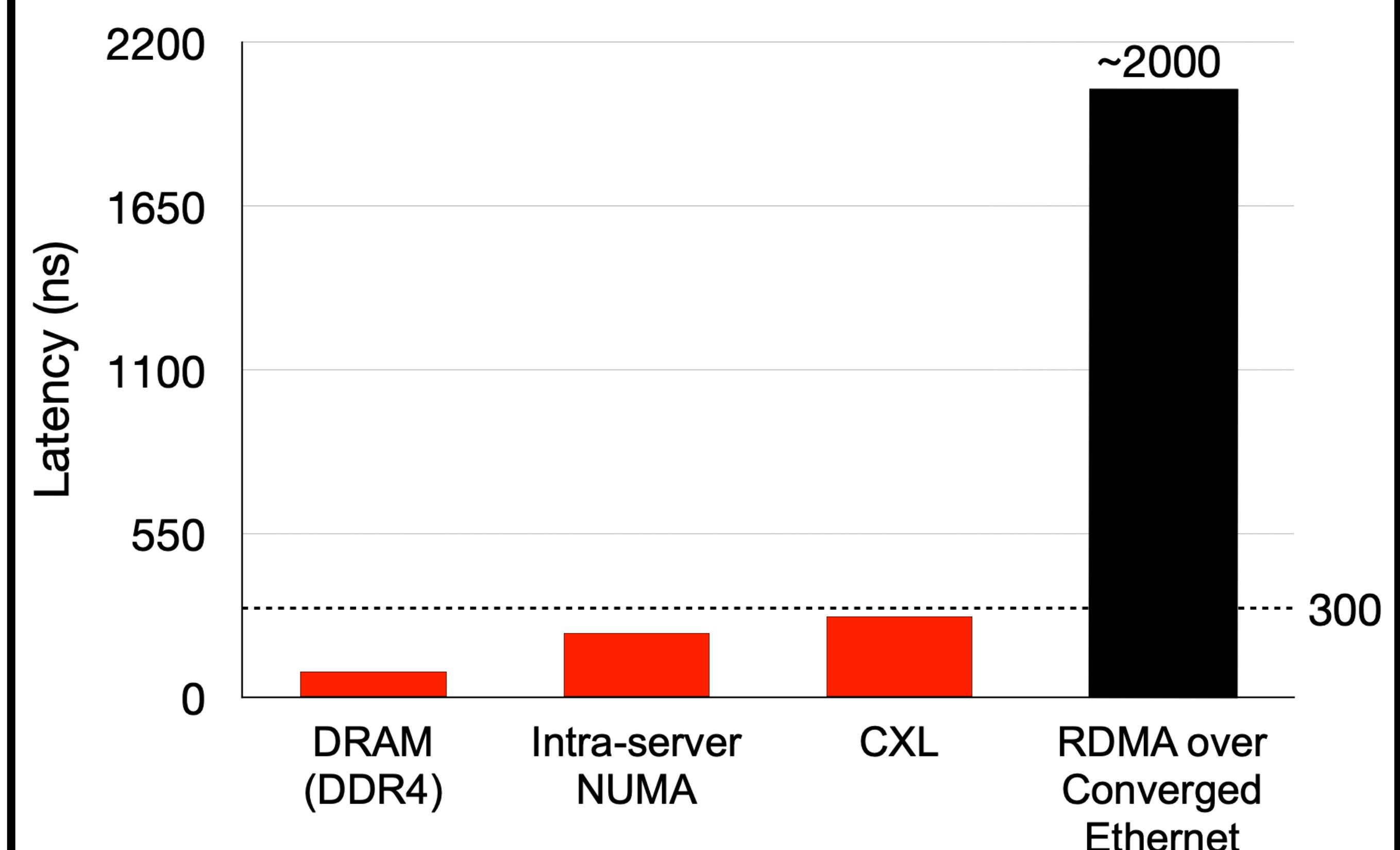
Available Solutions for CPU-NIC latency:

Faster link; integrate processor with NIC...

Total = 2035 ns + δ
 δ : queueing delay

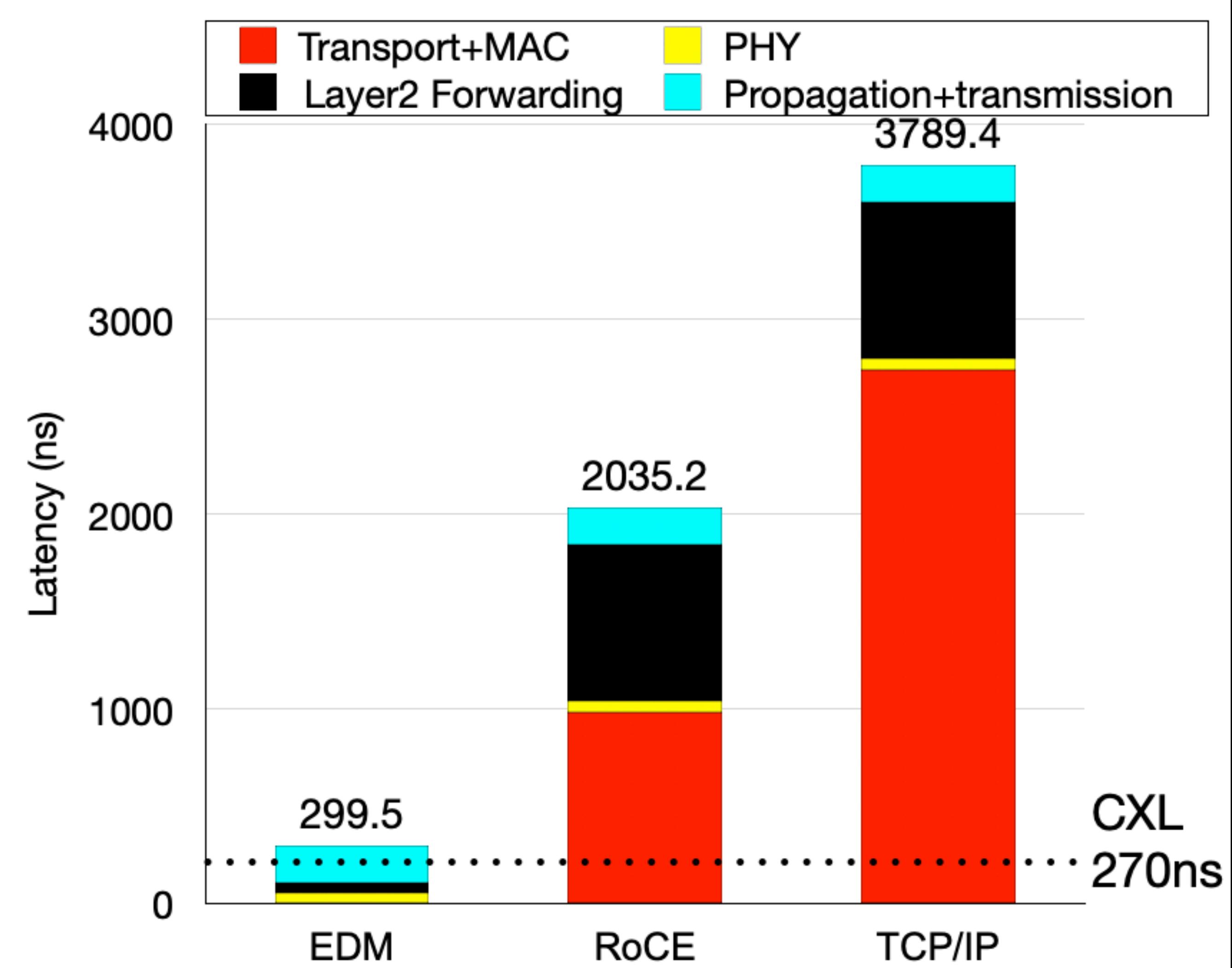


Prohibitive Ethernet Latency



Ideal state: an *unified fabric with low latency and high bandwidth*.

Evaluation: End-to-end Latency



Research Goal: Achieve near **intra-server memory latency** over rack-scale **Ethernet** (while maintaining high bandwidth utilization)

EDM (Ethernet Disaggregated Memory)

- Implement **protocol for remote memory access** within Ethernet **PHY**
- Implement a **centralized memory traffic scheduler** in the **PHY** of the switch

Implementation

