Ethernet for AI Networking: How Does RoCEv2 Stack Up Against InfiniBand?

Eric Fairfield, Technical Solution Architect – High Performance Networking

World Wide Technology

Eric Fairfield

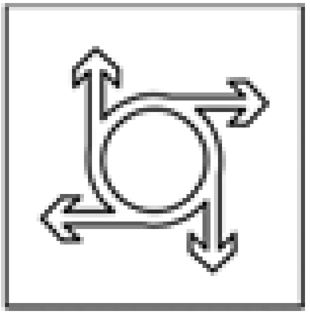
Technical Solution Architect

High-Performance Networking

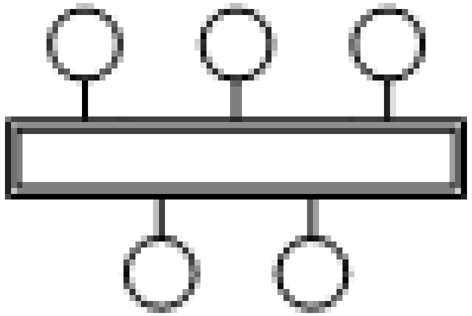
- *Contact Info:*
  - *Email: eric.fairfield@wwt.com*
  - *Linkedin: https://www.linkedin.com/in/eric-fairfield*
  - *608-628-6445*

Ethernet vs InfiniBand: The Battle Royale

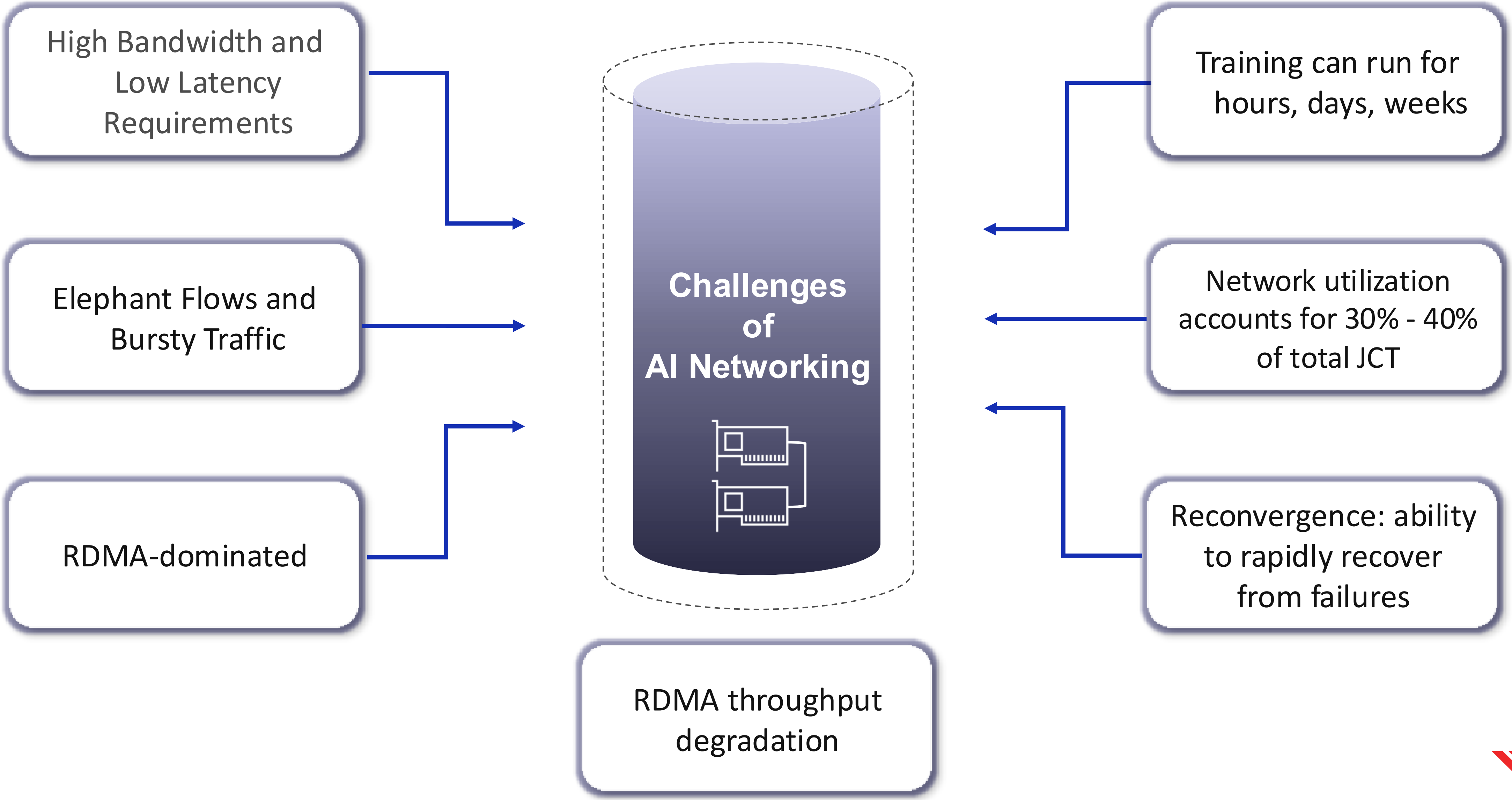
In this corner: born in 1999 and hailing from the mad labs of Silicon Valley, weighing in with an MTU of 4096 bytes, the Lord of Latency … InfiniBand!



And in this corner: coming to us from the hallowed labs of Xerox PARC, with a heavyweight MTU of 9216 bytes, the Swiss Army Knife of transport protocols, the god of "Good Enough" … Ethernet!

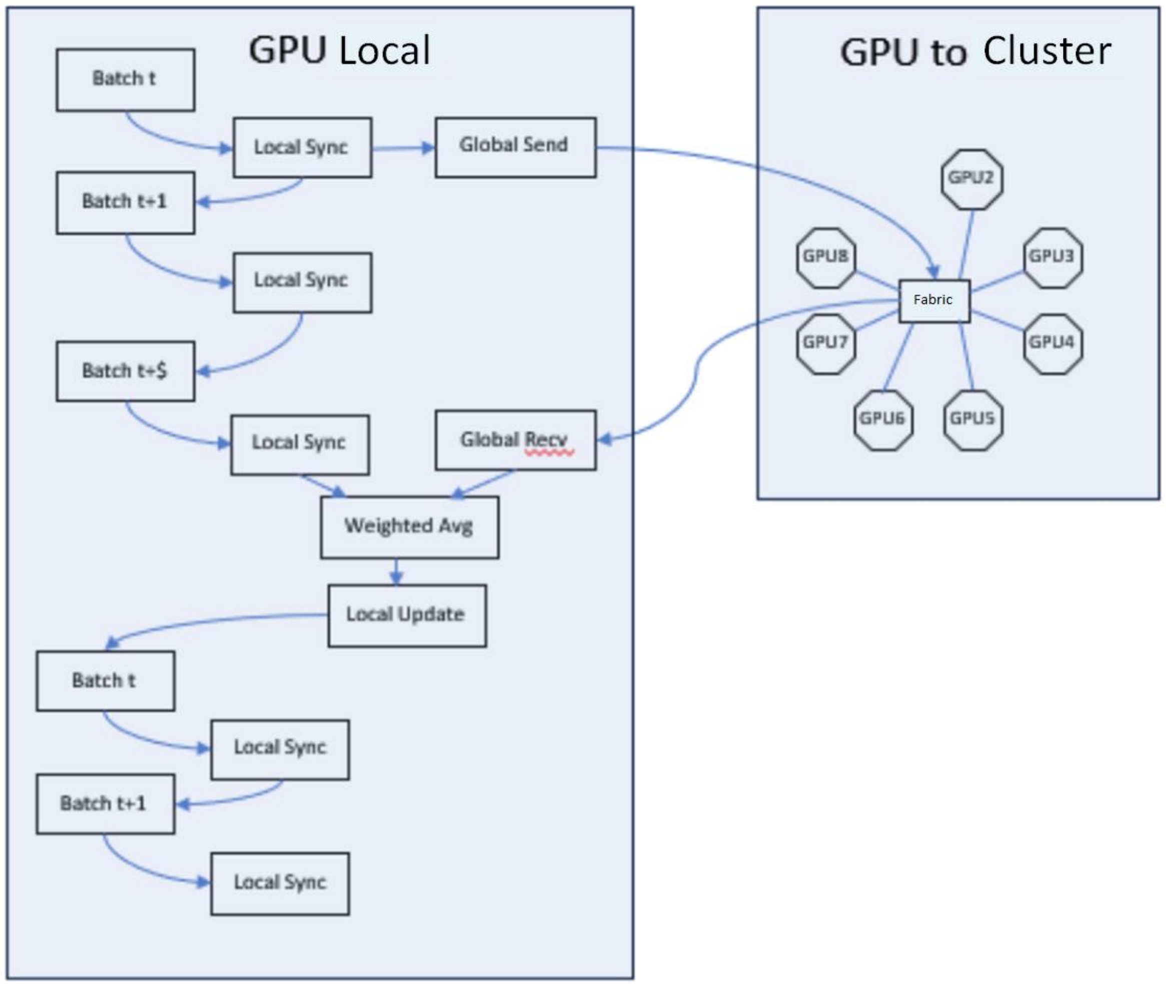


Why Does it Matter?



Compute Clustering

- GPUs communicate inside servers via specialized interconnects such as NVLink
    - NVLink is very fast but mostly non-routable
    - NVLink 5.0 can connect up to 576 GPUs
- GPUs typically communicate between servers in clusters ranging from 16 to 100,000 nodes/GPUs
- Cluster interconnects can be over Ethernet or InfiniBand
- Regardless of the transport, RDMA is leveraged to maximize throughput between GPUs



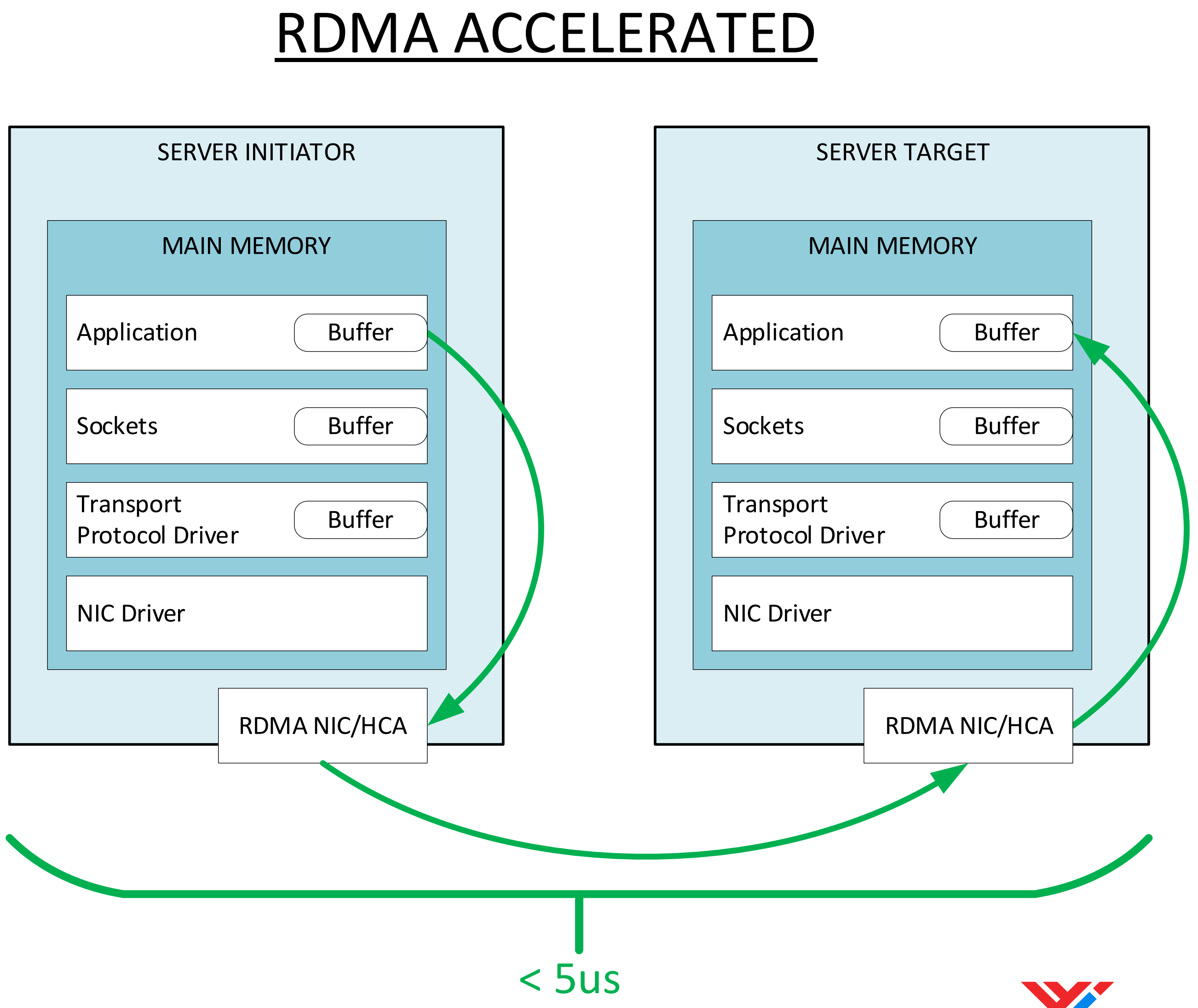
Cluster Workflow

World Wide Technology
Make a new world happen

Misconception: InfiniBand is Faster

WRONG!

- Both protocols are at parity in terms of ASIC latency and bandwidth
- In a direct kernel-to-kernel race, vanilla IP (over IB or Ethernet) takes about 50 microseconds regardless of transport
- The "go-fast juice" is RDMA, which is supported by both.
    - Drops latencies by 90-95%



Misconception: InfiniBand is More Reliable

WRONG AGAIN!

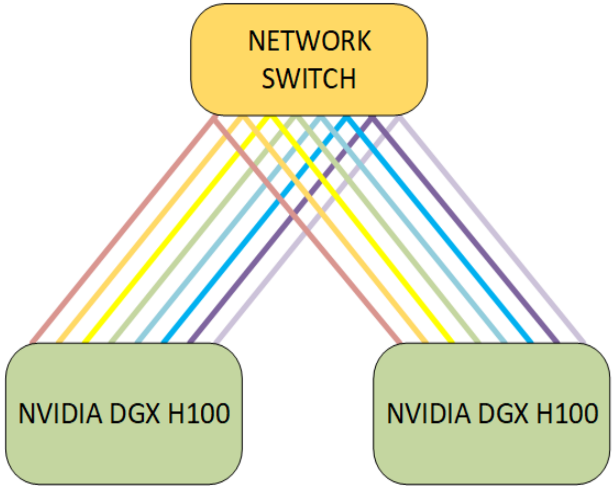
- Both protocols leverage different methods to achieve the same results
- InfiniBand uses scheduled fabric
  - Link-level flow control is used to avoid drops
- Ethernet uses a variety of tools to ensure reliable delivery
  - Current
    - PFC/ECN to manage congestion
      - **NOTE: This was sufficient for Meta to train LLAMA3 on a 24,000 GPU cluster**
    - Sender/Receiver collaboration and mid-fabric adaptive routing (eg, SpectrumX)
  - Upcoming (UET)
    - Multipath Packet Spraying
    - Receiver-based token for congestion control
    - Packet trimming as an alternative to Go_Back_N

How Do They Compare on Paper?

	ETHERNET	INFINIBAND
Max Bandwidth	800 gbps	800 gbps
MTU	9216 bytes	4096 bytes
Layer 3 Support	Yes	No
Delivery	Best Effort, enhanced to lossless	Scheduled Fabric
Load Balancing	Hash Values	Deterministic
RDMA Support	RoCEv2	Native
Enhancements	• Dynamic Load Balancing • Weighted ECMP • VOQ • Disaggregated Scheduled Fabric (DSF) • Adaptive Routing • EtherLink • Performance Isolation • DDP	• Adaptive Routing • SHARP
Pros	• Handles multi-workload fabrics (i.e., several different AI's with varying requirements) • Easily adapted skillset for existing network engineers	• Simple to install • Self-optimizing
Cons	• At present, requires a few QoS modifications to optimize performance	• Rare skillset • Operationally difficult to support when something goes wrong

How Do They Compare In Our AIPG Lab?

- Stripped down to the fundamentals
- Equipment:
    - (2) NVIDIA DGX H100 (16 GPU total)
    - (1) NVIDIA Quantum2 InfiniBand switch
    - (1) Cisco 9332D-GX2B Ethernet switch
- Eliminate every variable not related to network
    - Same GPU's, optics, cables leveraged for Ethernet and IB
    - Disable NVLink on DGX to force traffic through network
    - Enable basic PFC/ECN on Ethernet (for parity with native IB functions)
- Run MLCommons benchmarks on each topology and switch to gauge Generative and Inference performance



Atomic Test - Results

BENCHMARK	MODEL	ETHERNET	INFINIBAND
MLPerf Training	BERT-Large	3:01:06	3:02:31
MLPerf Inference	LLAMA2-70B-99.9	52.362	52.003

- **Generative**
    - Performance delta between InfiniBand and Ethernet was statistically insignificant (less than 0.03%)
    - Ethernet was faster than InfiniBand's best time in 3 out of 9 generative tests (although the margin was only by a few seconds)
    - Variance between OEM's was approximately 1.5% (5 minutes over 3 hour benchmark)
- **Inference**
    - Ethernet averaged 1.0166% slower (approximately 0.359 seconds)

World Wide Technology
Make a new world happen

How About Some More Testing?

BENCHMARK	MODEL	ETHERNET FEATURES	DURATION
MLPerf Training	BERT-Large	ECMP Only, NVLink Off	3:01:33
		ECMP/ECN/PFC On, NVLink Off	3:00:26
		ECMP/ECN/PFC On, NVLink On	3:00:31
		DLB Packet Spray/ECN/PFC/NVLink On	3:00:36

- **Spine/Leaf instead of a single switch**
- **Same exact tests of MLPerf Training**
- **Run three tests of each scenario (Average shown)**
    - **ECMP Only with NVLink Off**
    - **ECMP/ECN/PFC with Nvlink Off and On**
    - **ECN/PFC with packet spraying and NVLink On**

What Should I Be Considering For Ethernet?

- For Ethernet to be a viable GPU cluster interconnect, three main questions need to be answered:

    - To Schedule or not to Schedule?

    - Packet spray or flowlet (ie, “perfect load balancing”)?
        - Spraying means out of order delivery (roughly 25% of the time)
        - Regular Ethernet hashing is based on standard 5 tuples and can lead to major oversubscription.
        - Flowlets increase entropy by incorporating time between identical flows to mark them for separate egress interfaces

    - Reorder in network or NIC?
        - Some vendors incorporate endpoint intelligence (ie, DPUs) into switch/smartnic solutions that coordinate traffic patterns through the fabric and reorder at the receiving point on the server
        - Others will reorder at the last hop in the network prior to delivering traffic to the endpoint

What QoS Mechanism Should I Use?

- DCQCN (Data Center Quantized Congestion Notification)
  - Congestion control mechanism designed for performance-sensitive workloads (eg, RDMA)

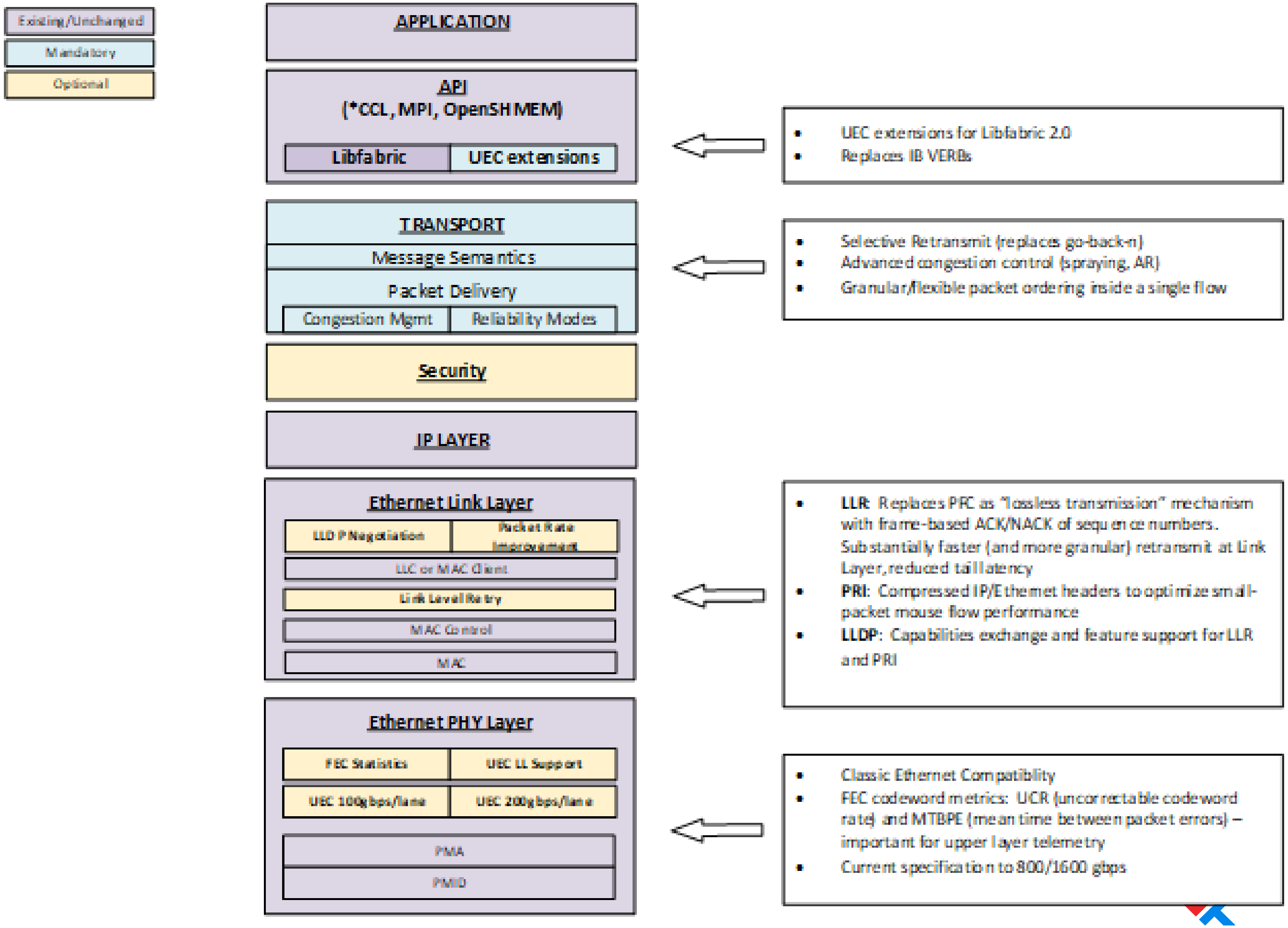
- PFC (Priority-Based Flow Control)
  - Link-level flow control mechanism that pauses traffic at the port and priority level when congestion is detected

- ECN (Explicit Congestion Notification)
  - End-to-end congestion control mechanism that marks packets to signal congestion to the sender, allowing it to reduce its transmission rate.

Where is Ethernet Going?

- Traditional Ethernet has challenges
  - Inefficient Load-Balancing
  - Engineering efforts associated with reliable delivery
  - PFC/ECN/DCQCN can be difficult to scale, tune, and adapt

- UltraEthernet (under development) addresses these
  - Major updates to Physical, DataLink, and Transport Layers of TCP stack
  - Scales to 1,000,000+ endpoints
  - Overhaul of RDMA



Why Pick Ethernet Over InfiniBand?

- **Question**
  - If Ethernet and InfiniBand deliver statistically identical results, why choose one over the other?

- **Answer**
  - *Operations*
    - InfiniBand is easy to setup and self-optimizing, but tricky to troubleshoot
  - *Support*
    - IB expertise is a rare skillset
    - In comparison, a network team can be brought up to speed on RoCE in a matter of weeks
  - *Reusability*
    - What's fast today becomes ho-hum tomorrow
    - 400gig Ethernet switches can be re-used for general datacenter use
    - 400gig InfiniBand is special-purpose