BROADCOM

Ethernet fabrics for AI Networking

Niranjan Vaidya

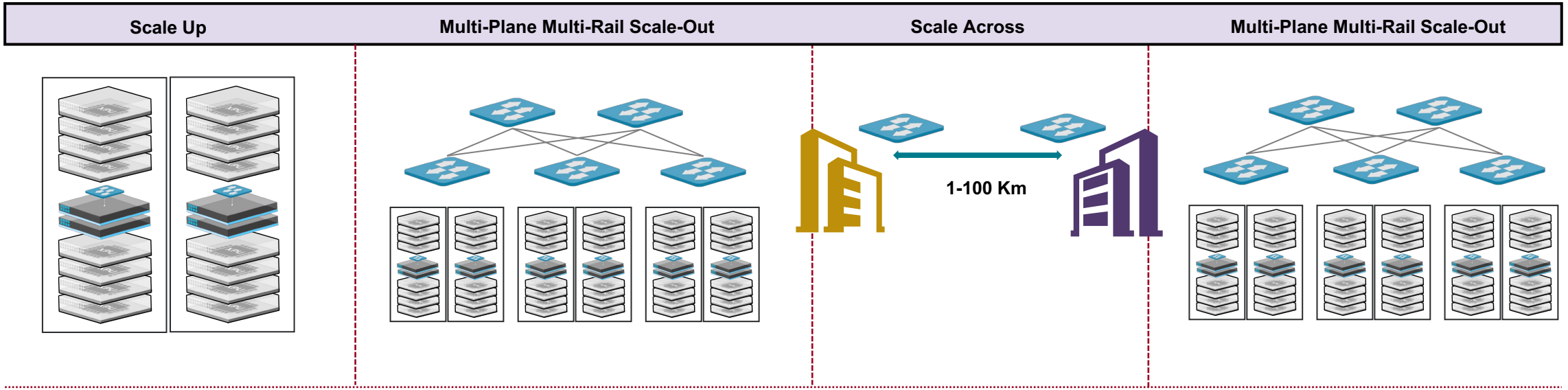
Distinguished Engineer / Core Switching Group

August 2026

- **The largest AI data centers are now based on Ethernet.**
- **Multiple GPU/XPU designers are using Ethernet for their high-performance scale-up interconnects.**
- **Ethernet continues to be the mainstay for front-end and inter-DC connectivity.**

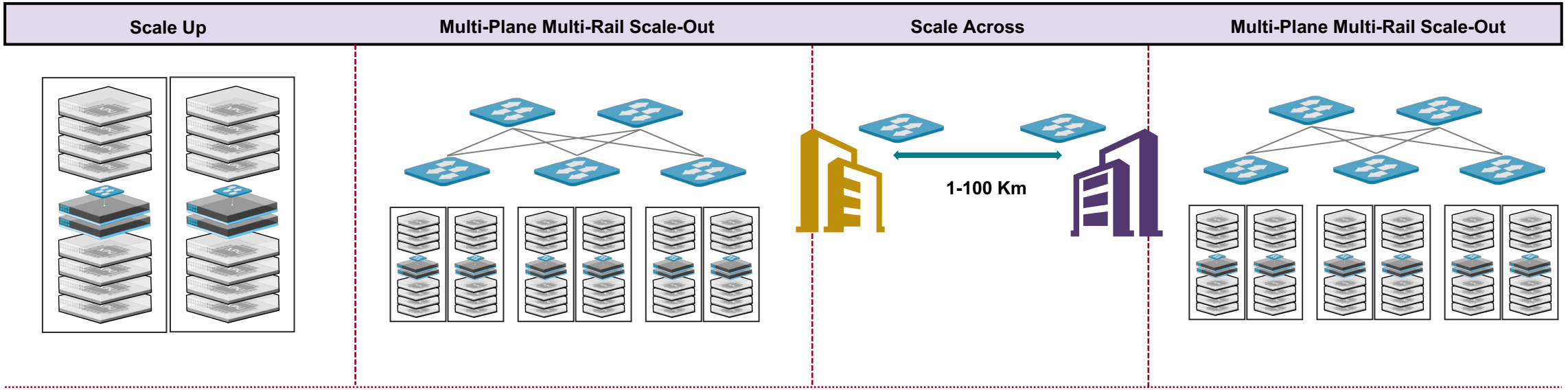
- **What attributes and capabilities does Ethernet bring to the solution space?**
- **How is Ethernet evolving to address HPC/AI needs?**

A Familiar Infrastructure View



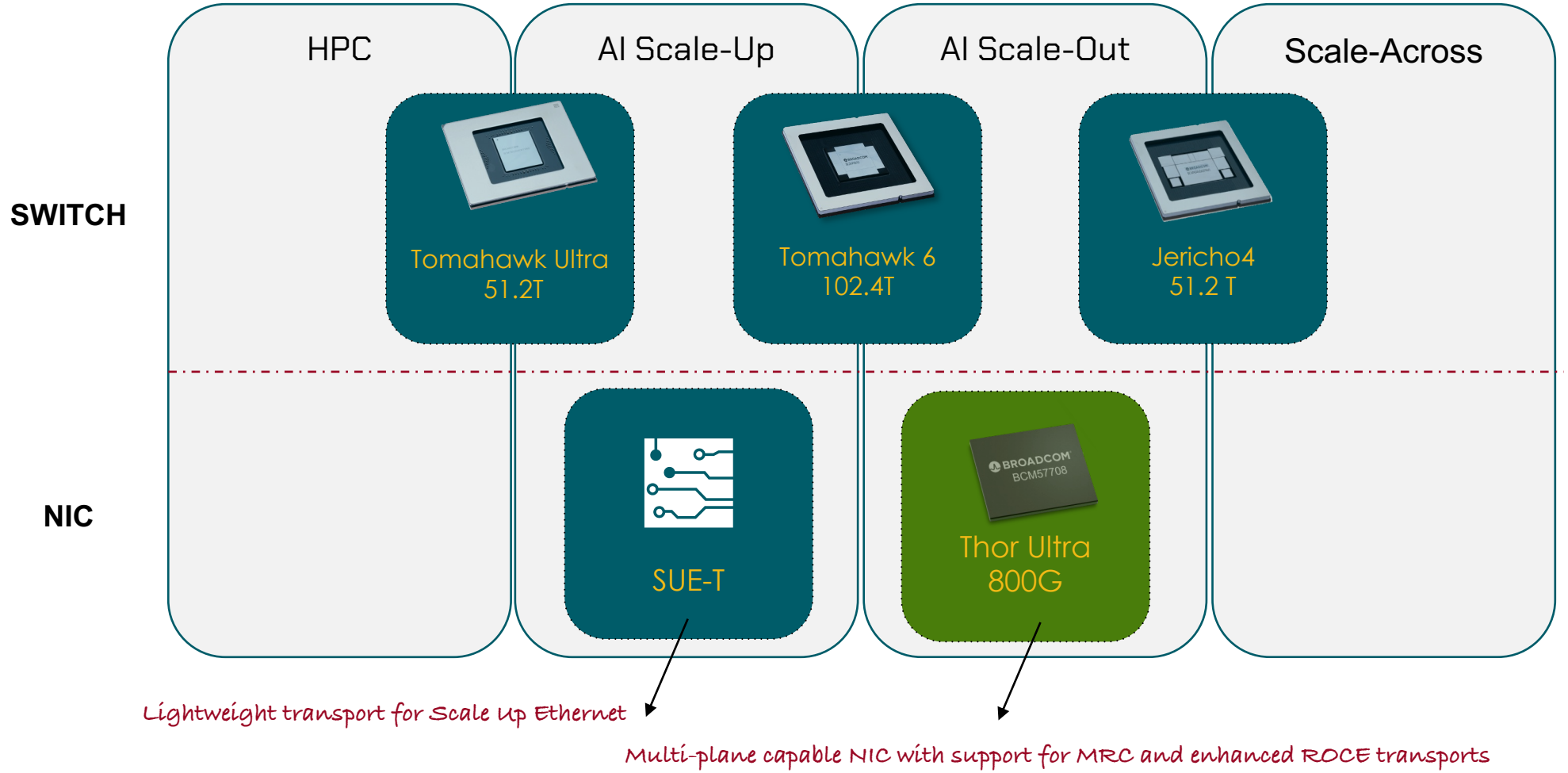
< 2K	< 512 K	1M +	**Scale**
< 5 us	< 15 us	milliseconds	**RTT**
Low	Medium	High	**Fabric Feature Richness**
Tightly coupled memory semantics	Message Passing		**Communication Paradigm**
High	Medium		**Latency Sensitivity**
High	Medium		**Jitter Sensitivity**
Point to Point	End to End		**Reliability**
Low (Very Disruptive)	Low (Disruptive)		**Fault Tolerance**

The Ethernet Stack - Reuse and Innovation



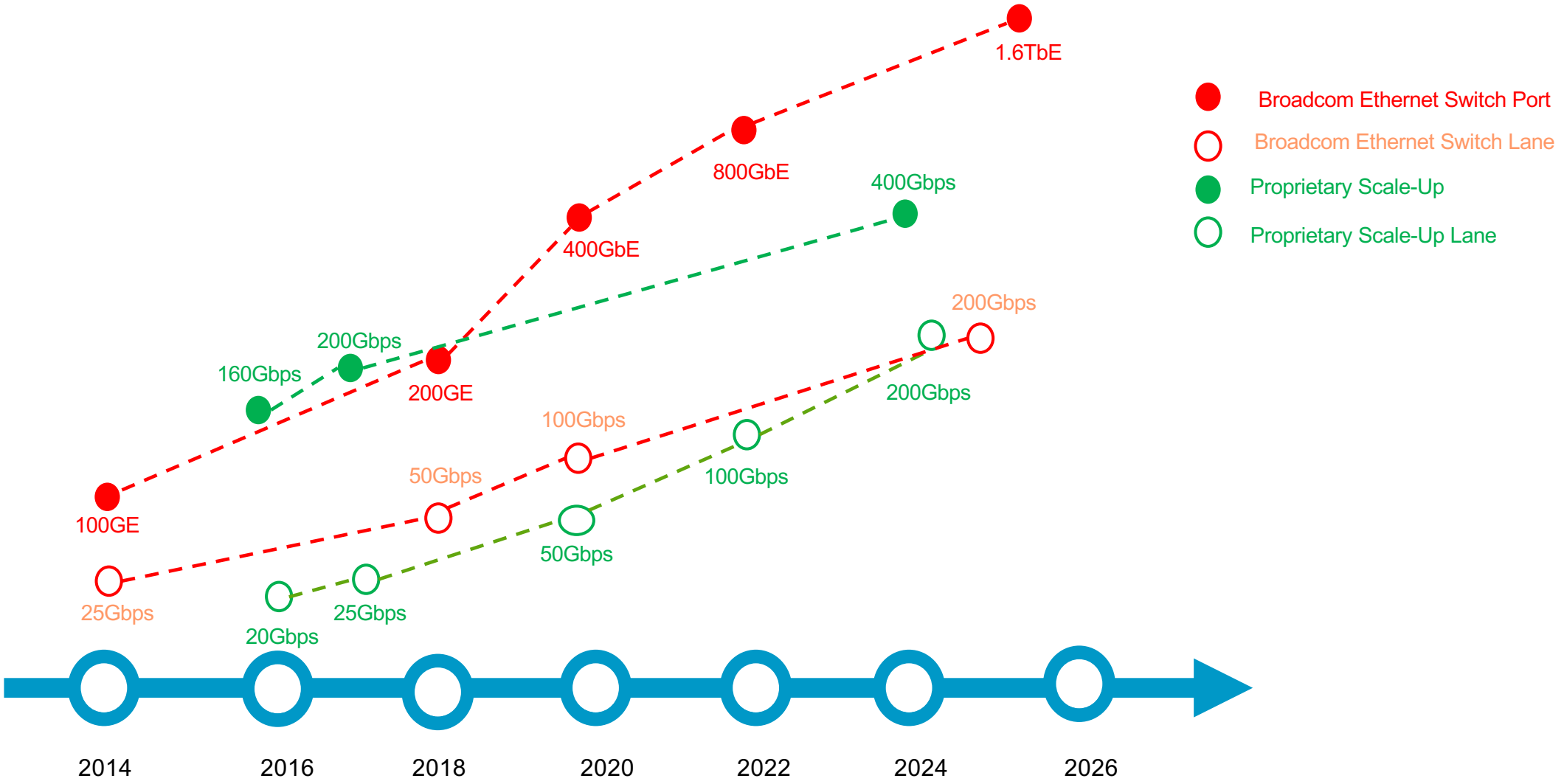
Familiar protocols, management tools and SDN frameworks		**Management**
MACSec, PSP, Proprietary	MACSeC, PSP, UET TSS Group Keys	**Security**
VLANs	VxLAN or other	**Virtualization / Multi Tenancy**
SUE-T, ULN, UALoE, Meta RoCE	UET, MRC, Various enhanced RoCE variations	**Transport**
ESUN/AFH, UFH	IP, Adaptive Routing / DLB, Fast Failover, SRv6	**Network**
LLR, CBFC, RLM	Standard Ethernet	**Link**
Low latency RS272 FEC, CPO	Standard Ethernet RS572	**Physical**

This table is illustrative, not comprehensive

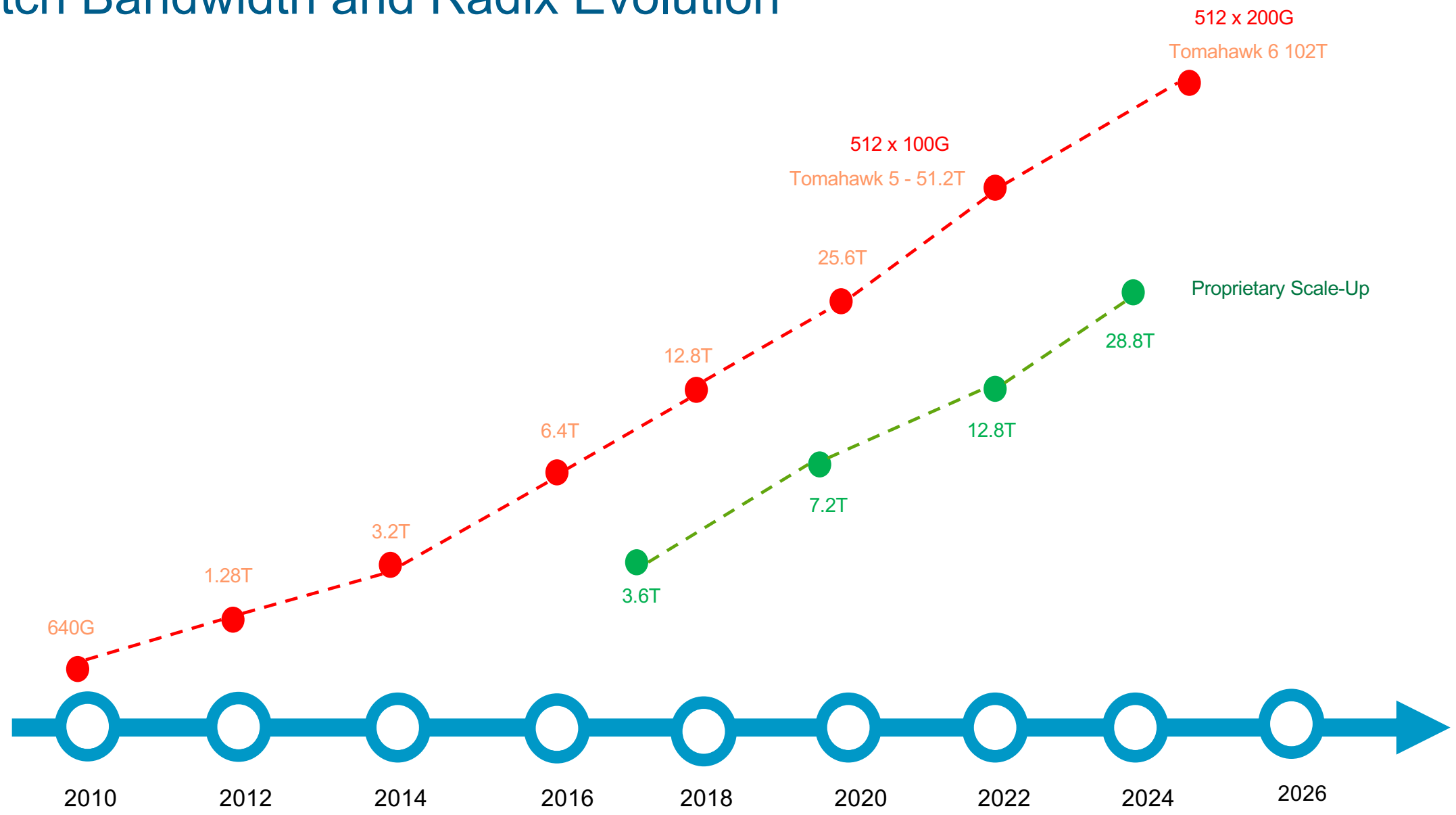


Addressing needs of Scale-Up fabrics

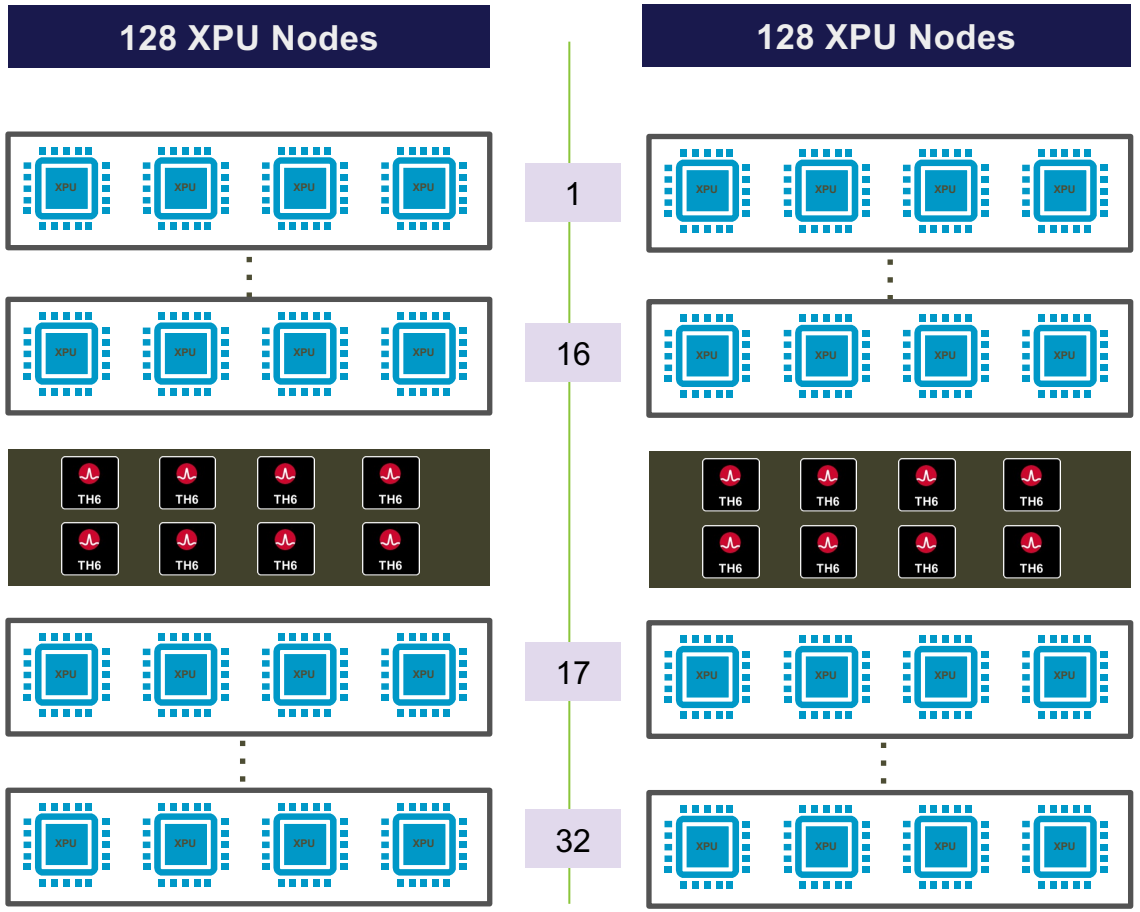
Port and Lane Speed Evolution



Switch Bandwidth and Radix Evolution



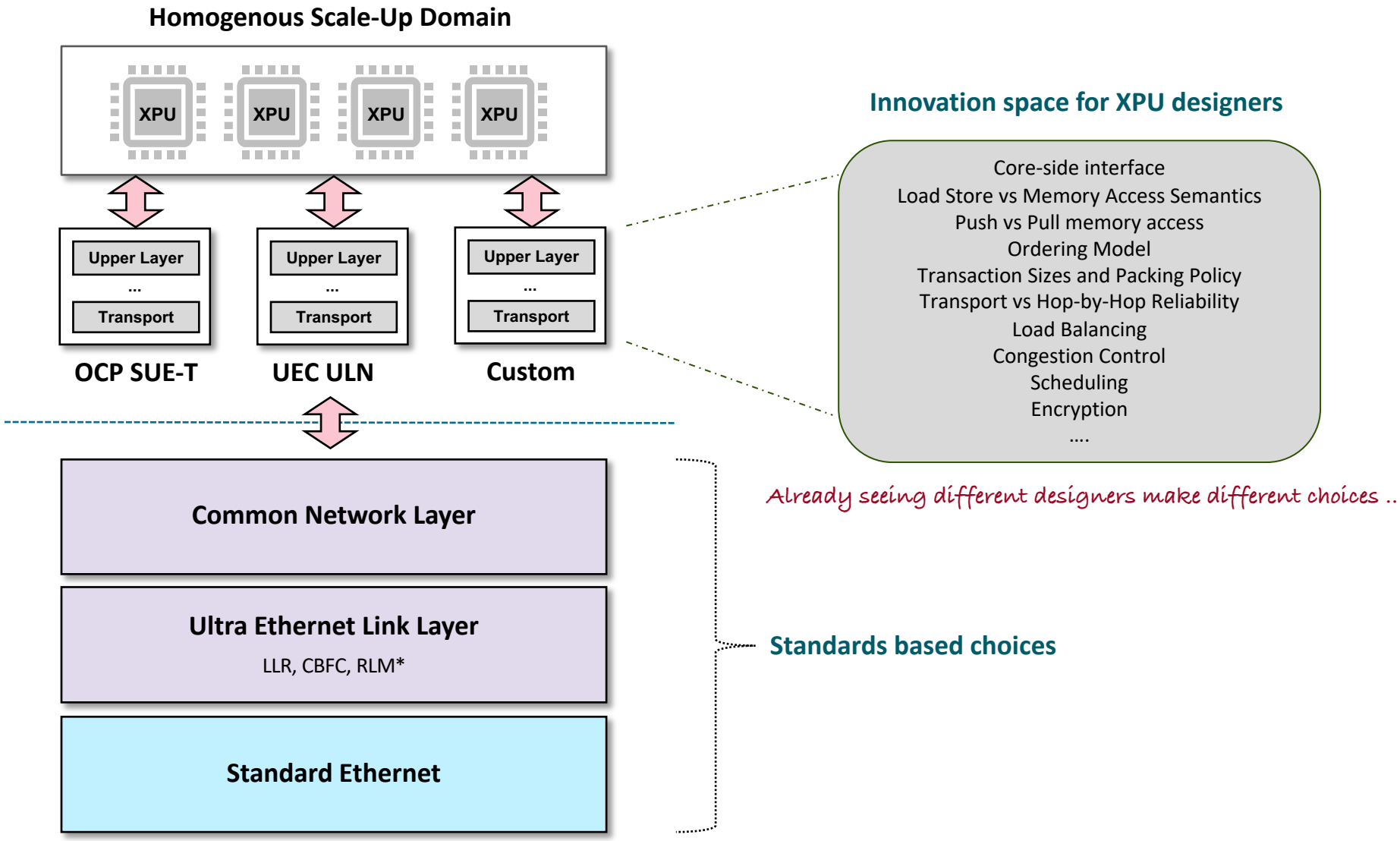
Cluster Scale Example



2026

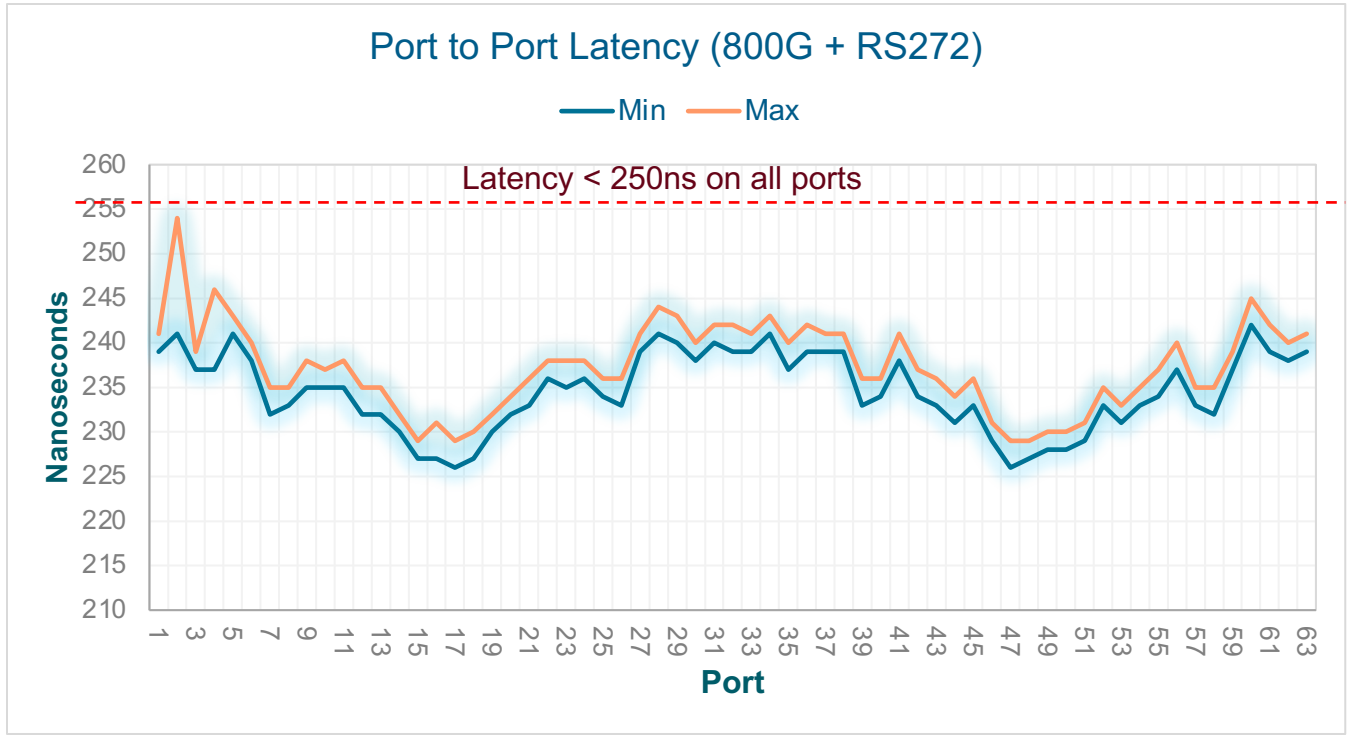
High radix Ethernet switches enable 128/256/512+ XPU Scale-Up domains in a single tier, multi-plane fabric.

Modular Scale-Up Architecture



* Reduced Lane Mode is in development

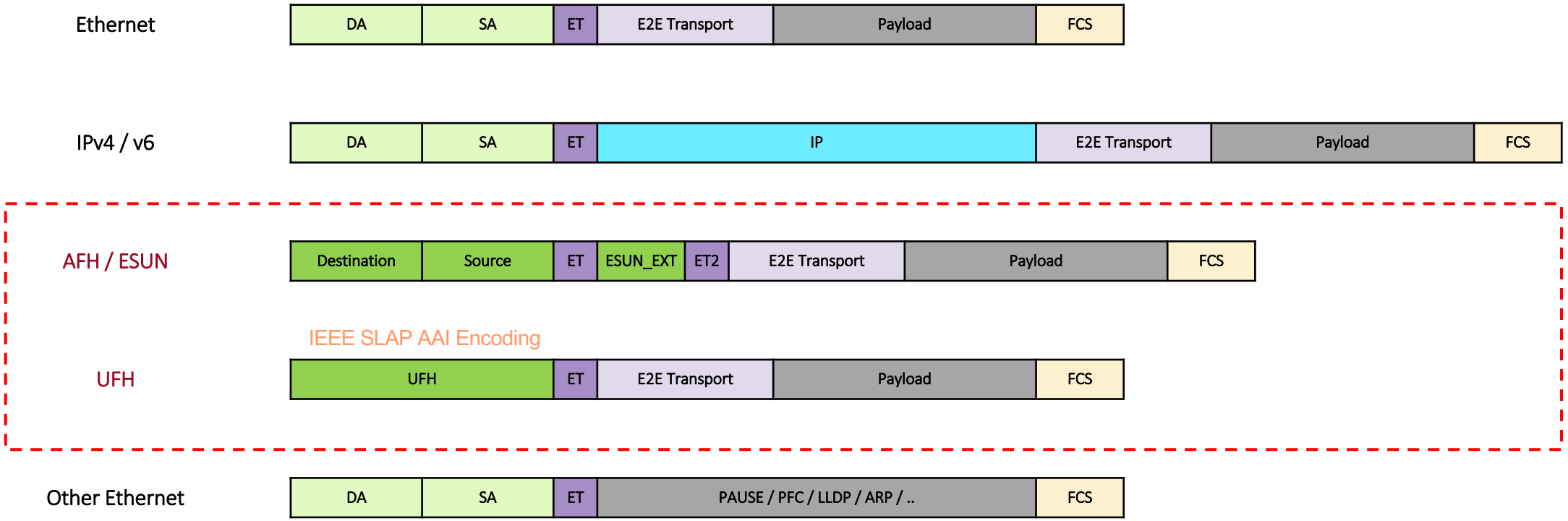
Addressing small packet throughput and latency



PktSz	Time (sec)	Total Pkts	Total Bytes	Exp.Rate (Gbps)	Obs.Rate (Gbps)
64	30	42876153178	2915578416104	800	799.99
64	30	42876922472	2915630728096	800	799.99
64	30	42879084392	2915777738656	800	799.99
64	30	42879184021	2915784513428	800	799.99
64	30	43279114690	2942979798920	800	799.97
64	30	43284217326	2943326778168	800	799.99
64	30	43289563949	2943690348532	800	799.99
64	30	43290198852	2943733521936	800	799.99
64	30	43290689501	2943766886068	800	799.99
64	30	43293097582	2943930635576	800	799.99
64	30	43294384425	2944018140900	800	799.99
64	30	43296920187	2944190572716	800	799.99
64	30	43297910995	2944257947660	800	799.99
64	30	43300131355	2944408932140	800	799.99
64	30	43301764787	2944520005516	800	799.99
64	30	43302575458	2944575131144	800	799.99
64	30	43303395516	2944630895088	800	799.99
64	30	43304192682	2944685102376	800	799.99
64	30	43304135586	2944681219848	800	799.99
64	30	43303536082	2944640453576	800	799.99
64	30	43304256696	2944689455328	800	799.99
64	30	43303355608	2944628181344	800	799.99
64	30	43304211274	2944686366632	800	799.99
64	30	43304223620	2944687206160	800	799.99
64	30	43306237255	2944824133340	800	799.99
64	30	43308006873	2944944467364	800	799.99
64	30	43305978615	2944806545820	800	799.99
64	30	43305350239	2944763816252	800	799.99
64	30	43307250806	2944893054808	800	799.99
64	30	43308722178	2944993108104	800	799.99
64	30	43308174785	2944955885380	800	799.99
64	30	43308956075	2945009013100	800	799.99
64	30	43309719314	2945060913352	800	799.99

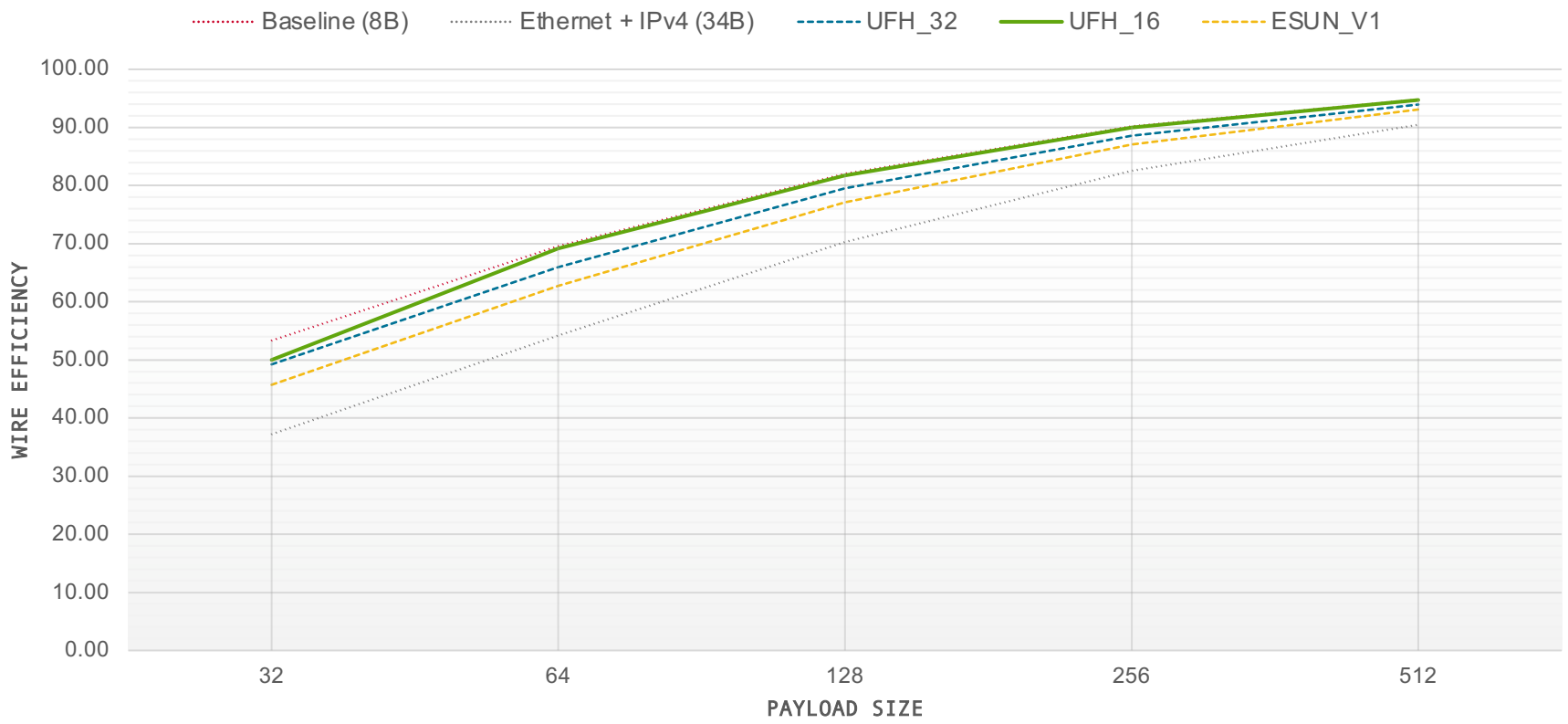
Tomahawk Ultra is proof Ethernet switches can meet Scale-Up and HPC performance needs

Wire Efficiency with Optimized Headers



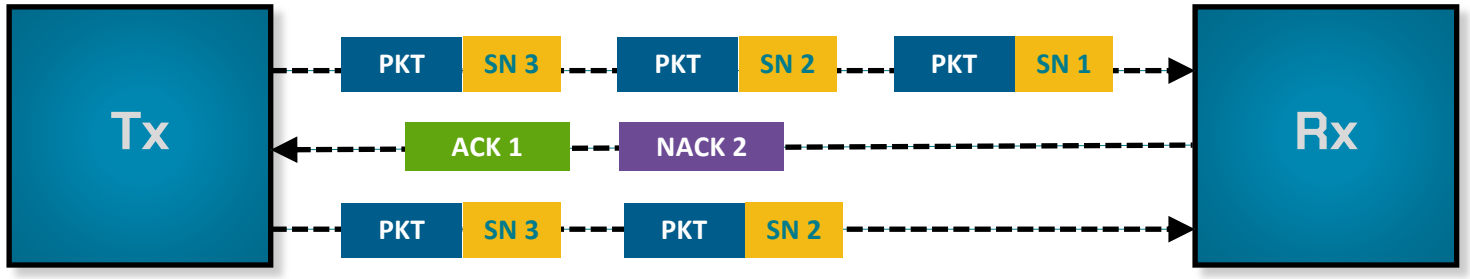
- Optimized headers coexist with standard Ethernet + IP headers on the wire.
- Familiar Ethernet + IP Management Tools.
- Ethernet silicon ecosystem reuse.

Wire Efficiency with Optimized Headers



- Retains all advantages of the Ethernet + IP ecosystem, including IP addressing.
- Single hop or multi-hop fabrics with DSCP and ECN marking.

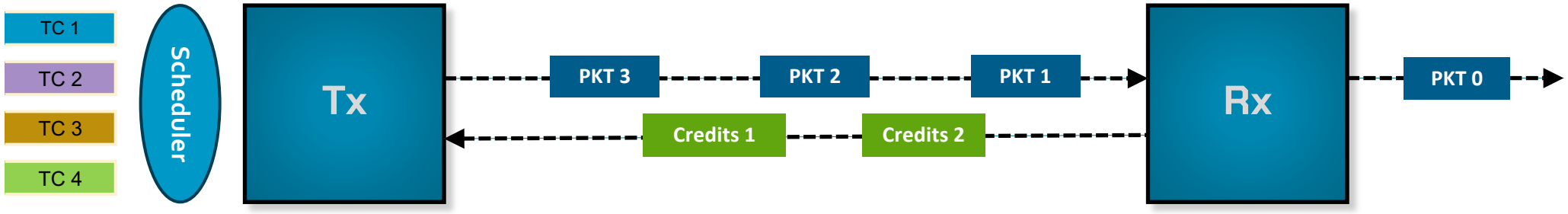
Link Level Retry



Retransmit on receiving NACK. Go back N.

- LLR removes the need for delayed end-to-end retransmissions.
- Important for memory semantic communication paradigms.

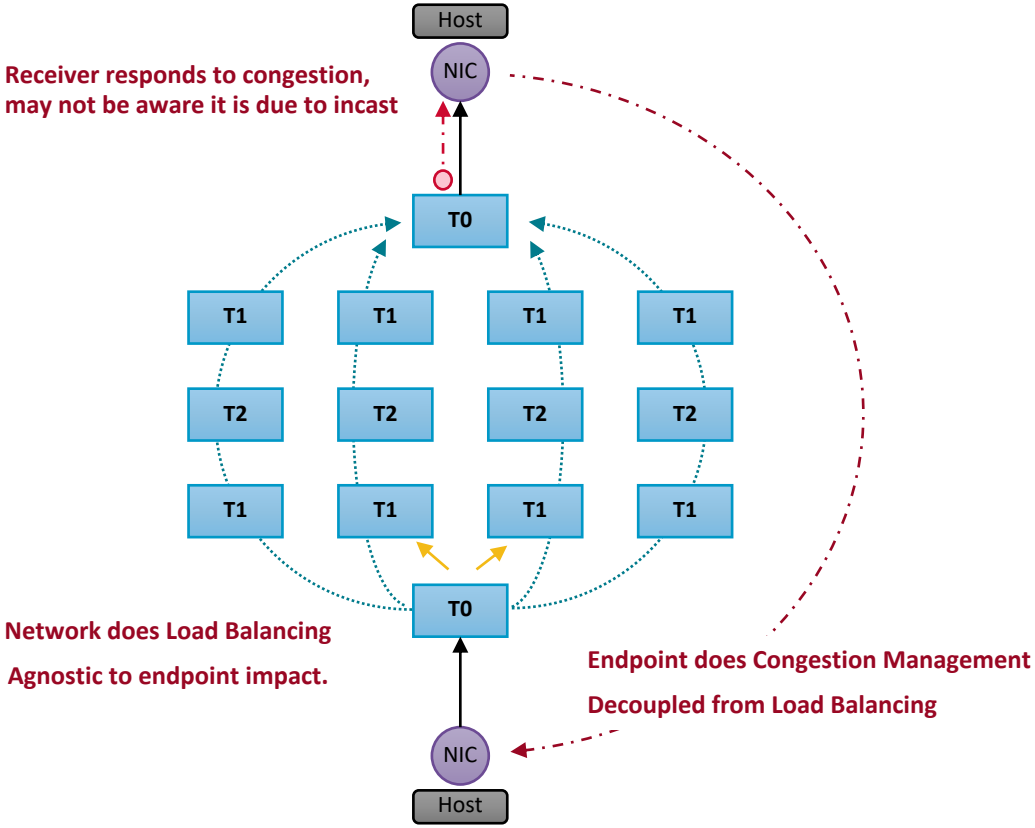
Credit Based Flow Control



- Lossless fabrics with tight control over buffer usage.
- Per traffic class or shared credits.
- Can coexist with Ethernet PFC.
- LLR and CBFC usage is negotiated via Ethernet LLDP.

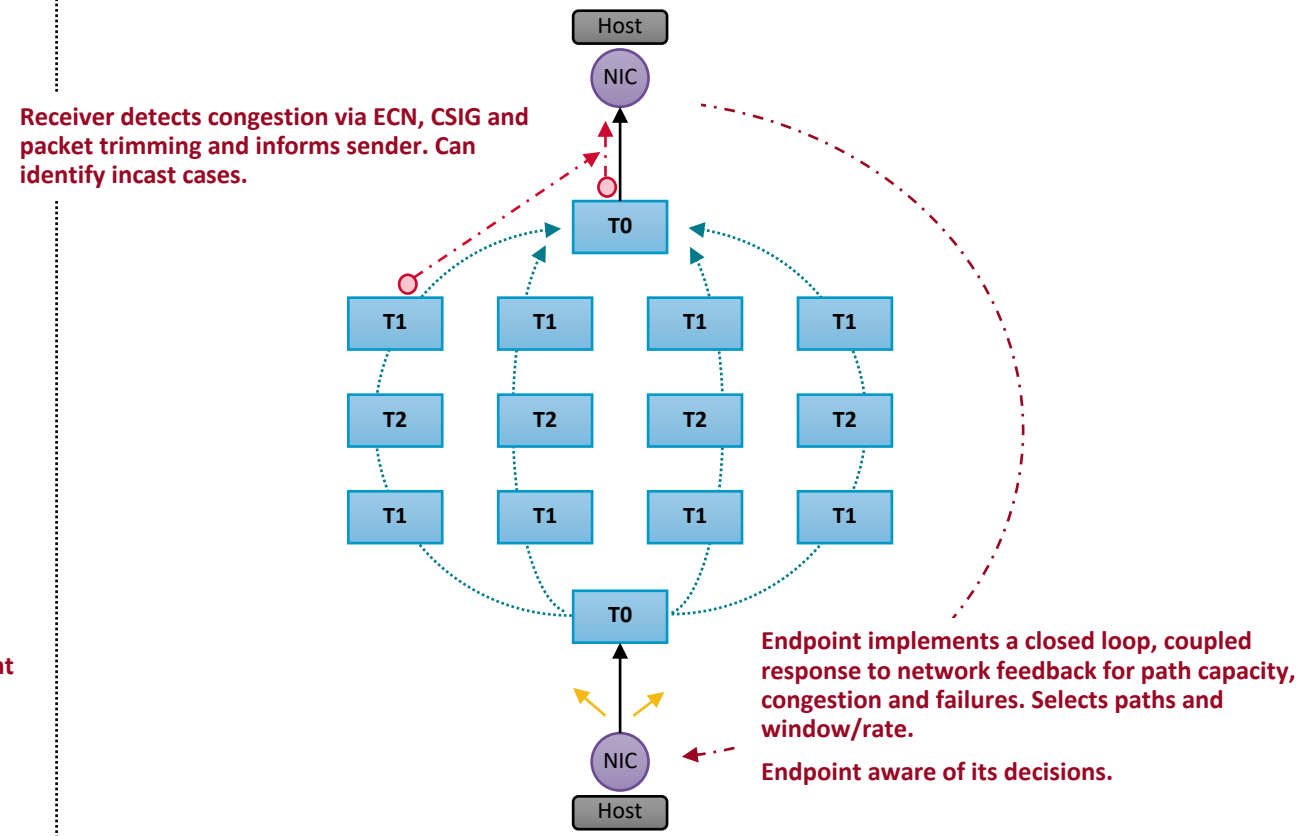
High performance Scale Out Fabrics

A Tale of Two Models



Traditional Network Model

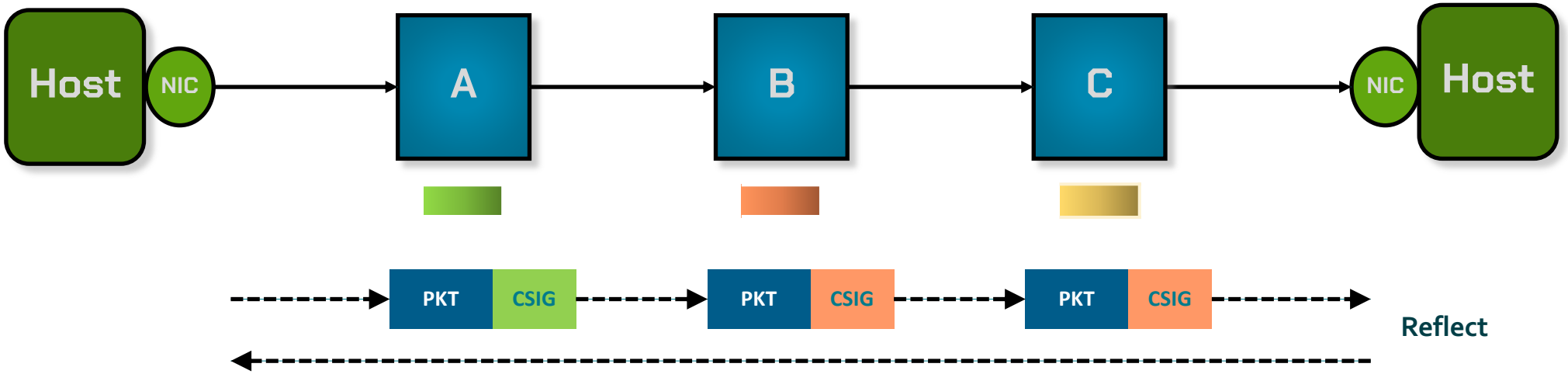
RoCE / Enhanced RoCE



Moving the intelligence to the edge

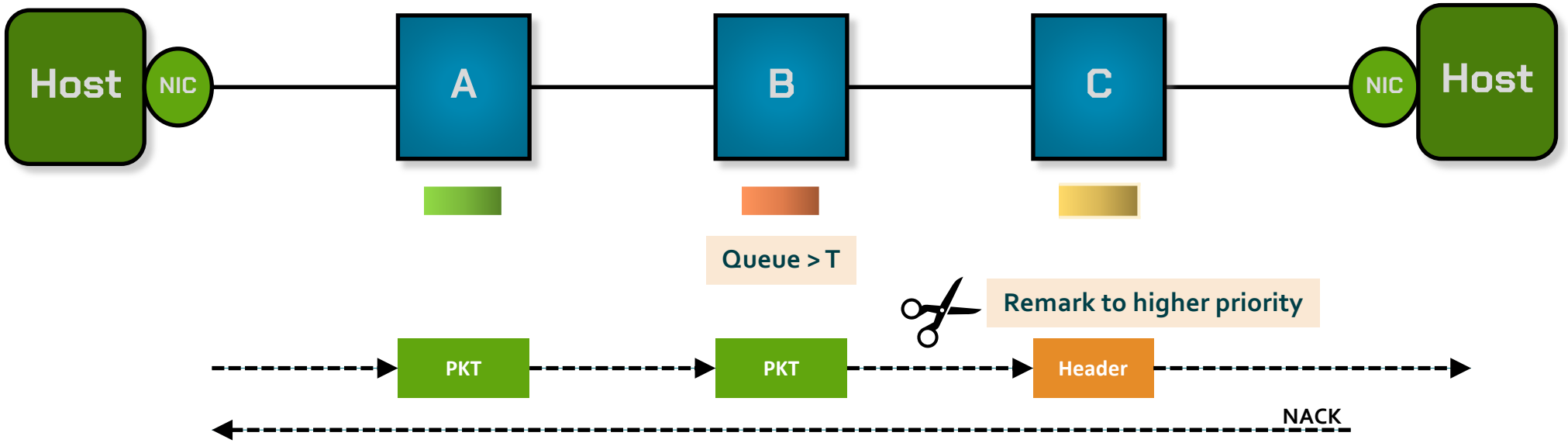
Ultra Ethernet Transport (UET), MRC, Google Poseidon

Congestion Signaling (CSIG)



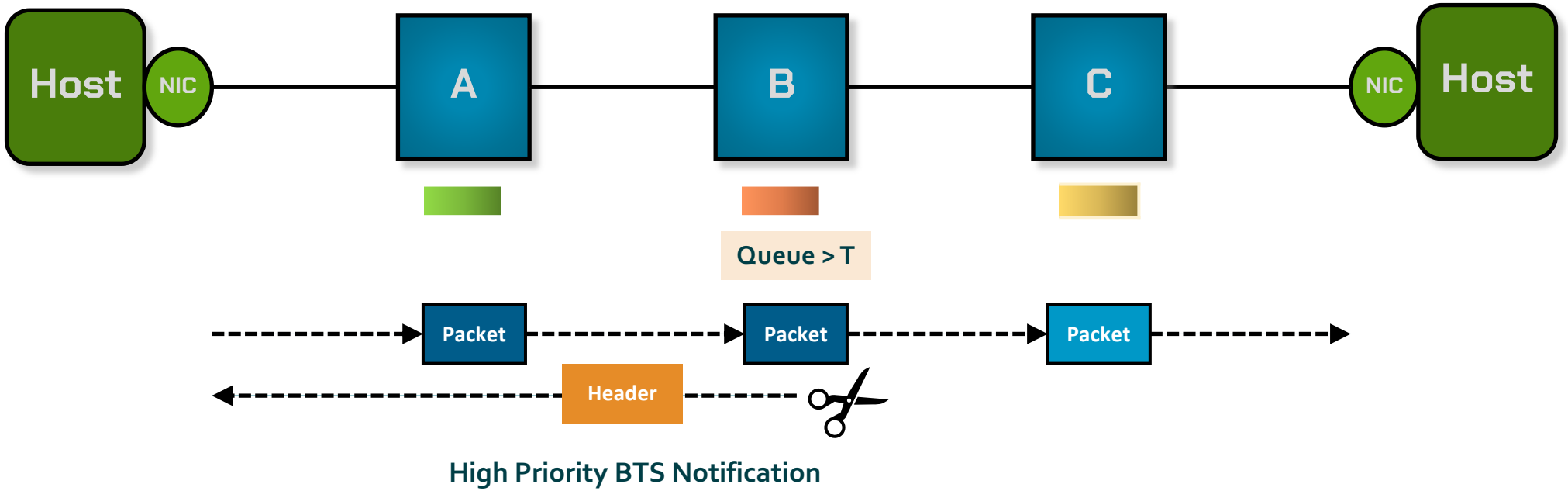
- Wire-efficient mechanism to identify path bottleneck and available capacity.
- Applications to transport design, congestion control, load balancing and network telemetry.
- CSIG is included in UEC 1.1.
- Google paper with deployment experience at SIGCOMM 2026.

Packet Drop Notifications aka “trimming”



- Classic problem in transport design – how fast should a cold sender ramp up?
  - Conservative mechanisms such as TCP slow-start leave capacity on the table.
  - Lossless transports build congestion trees.
- Drop notifications enables senders to ramp up quickly without multi-RTT delays.

Back to Sender Notifications



- High bandwidth links create sub-RTT flows. Long links create large RTTs.
- ‘Back to Sender’ notifications allow senders to control Tx without RTT delays.

Take Aways ..

Ethernet is evolving rapidly on multiple dimensions

- Many enhancements have been around for years (e.g., adaptive routing); more are coming.
- Any reference to "commodity Ethernet" in the HPC/AI space is using the wrong baseline.

Synergistic arms race

- The Ethernet ecosystem drives all layers of the ~~network~~ innovation stack.
- In return, innovation feeds the ecosystem and standardization.

Optionality for system designers and operators

- Solution characteristics, technology choices and economic.