SUE-T – An Ethernet transport for Scale-Up Networking

Tom Emmons *Software Engineer, Arista Networks Inc.*

Karen Schramm *Vice President, Broadcom Inc.*

OCP PLATINUM Member

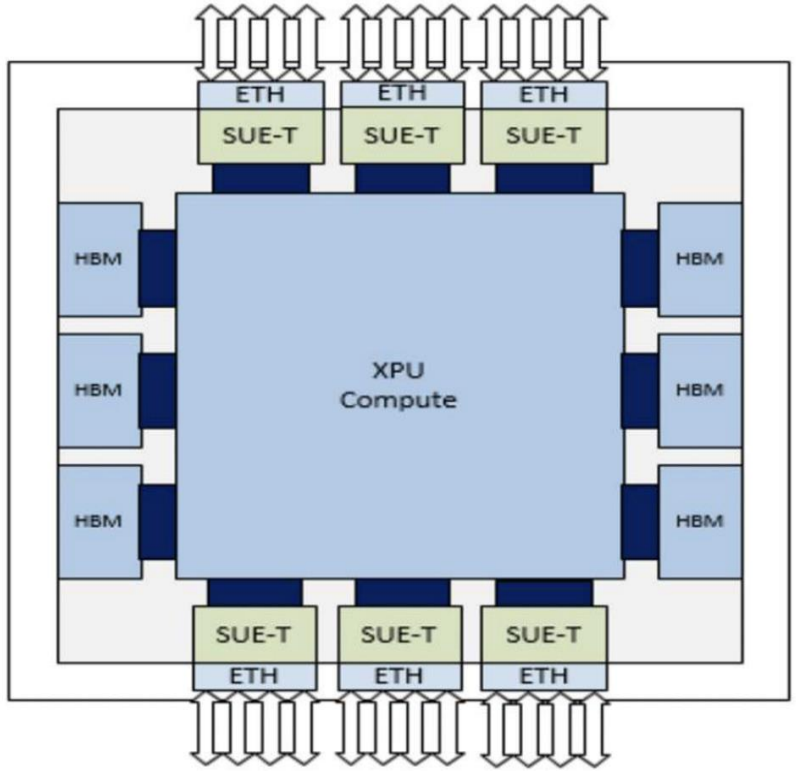
Why SUE-T?

- Scale up transport on XPUs have used proprietary solutions
  - Open solutions did not meet the efficiency and latency requirements
- Broad industry investment in advancing transports and ETH link optimizations
  - OCP ESUN, Ultra Ethernet Transport (UET), UEC LLR and CBFC, etc.
- Value of using standardized protocols is increasing
  - Scale-up and scale-out lines are converging, storage is directly attached
  - In-network compute (INC) available in switches

SUE-T specifies a network interface with guidance on trade-offs optimized for wire efficiency and low latency area efficient solution based on open standards

SUE-T Scope

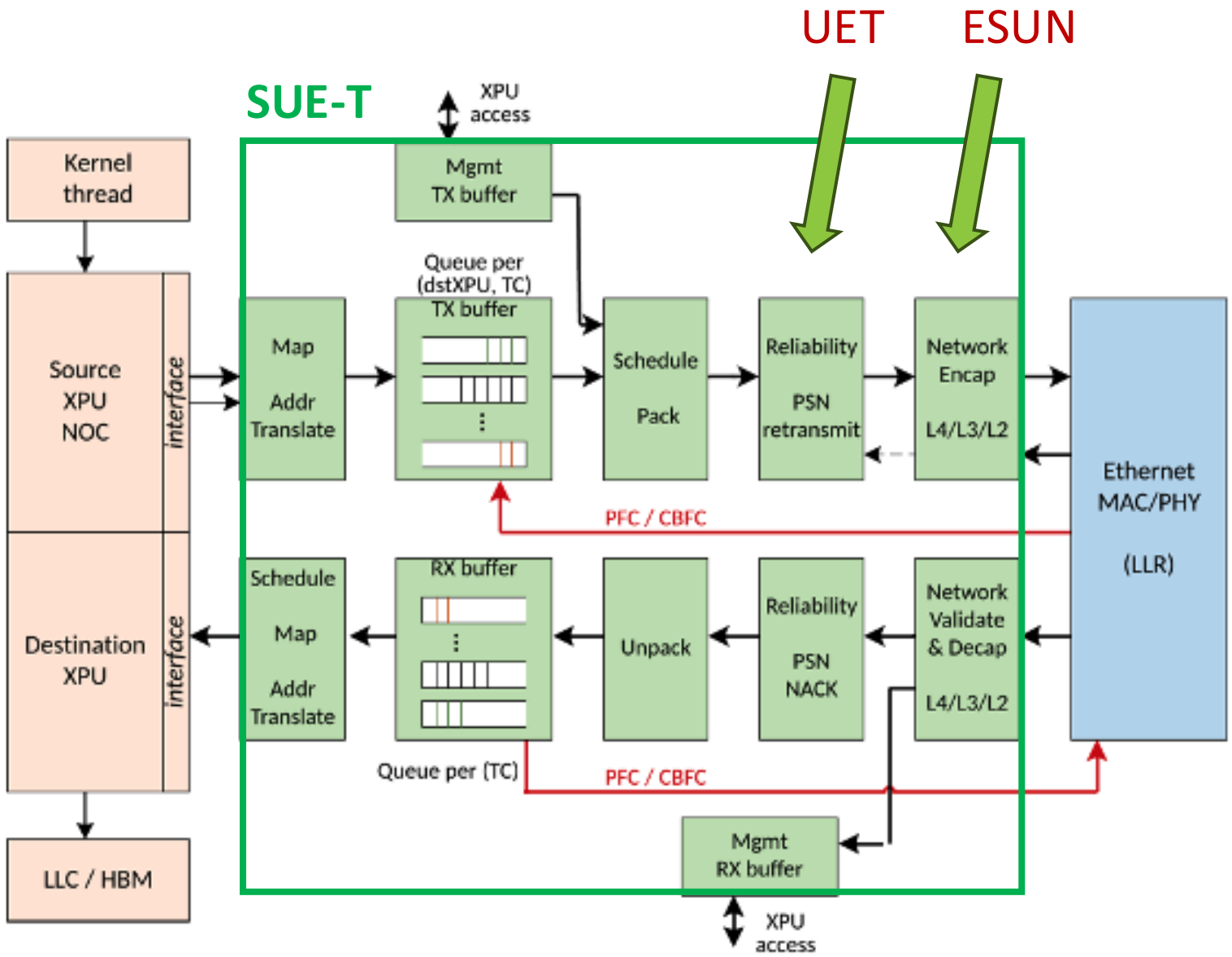
Feature	Primary Target	Notes
Network scale	Single hop	Extensible to multi-hop
Number of XPUs	Up to 4k	Ethernet addressing, no practical limit
Transaction types	Load/Store/Atomics	Optimized for small transactions
Memory architecture	Shared memory	Direct address translation
E2E latency	< 2us RTT	No hard limit Affects buffer size
Traffic Classes	2+	Separate request and response



Area Optimized

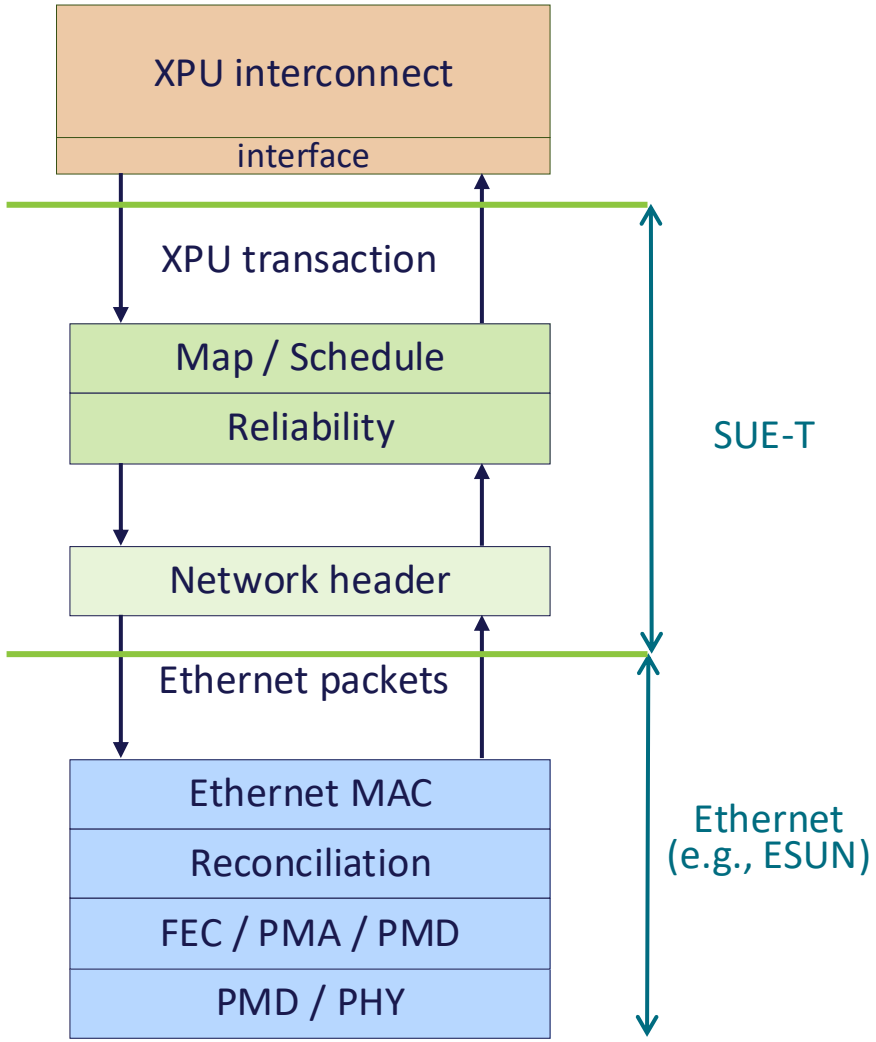
SUE-T Overview

- Accept XPU transactions
  - Buffer per destination
- Schedule packets to the network
- UET RODL transport
  - Reliable, ordered delivery
- OCP ESUN
  - Compact L2/L3 header
- Symmetric receive path



SUE-T Structure

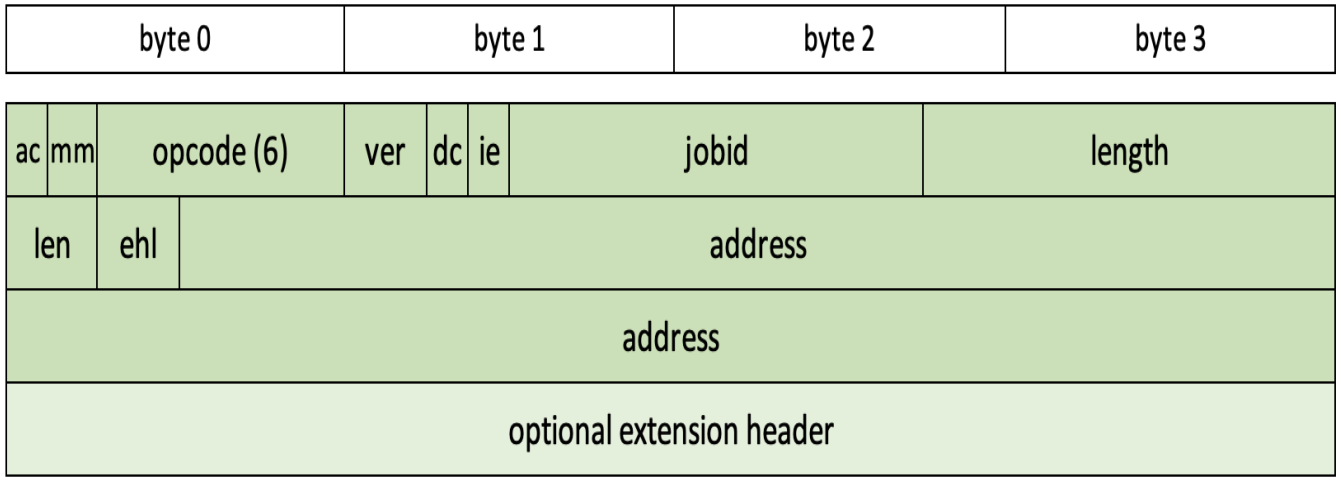
- XPU to SUE-T connection
  - AXI or simple signaled interface
- Packing, buffering and scheduling
  - Transactions are pipelined with minimal buffering
  - Buffering enables packing multiple transactions in a single packet for increased wire efficiency
- Transport and network headers are optimized
  - UET Semantic header for opcode and control
  - UET RODL transport header for reliable delivery
  - ESUN header for L2/L3 forwarding



UET Semantics

- SES header carries opcode and control
- Compact 12B header
  - use ehl option to carry XPU specific fields
  - single transactions up to 16K-1 bytes

SUE-T Semantic Header



- Enables:
  - opcode write, read, atomic
  - jobid user isolation
  - mm packing multiple transactions in one packet (more messages)
  - ac address compression, reduce header to 8 bytes

UET Reliability - RODL

- Reliable ordered protocol
  - Avoid bitmap tracking
- Implementation efficient
  - Keep endpoint simple
  - Choose retransmission or LLR
- Uses:
  - cumulative ack — ACK received packets, minimize need for ACK packets
  - clear — ACK read/atomic responses, optional
  - flags — static or dynamic connections

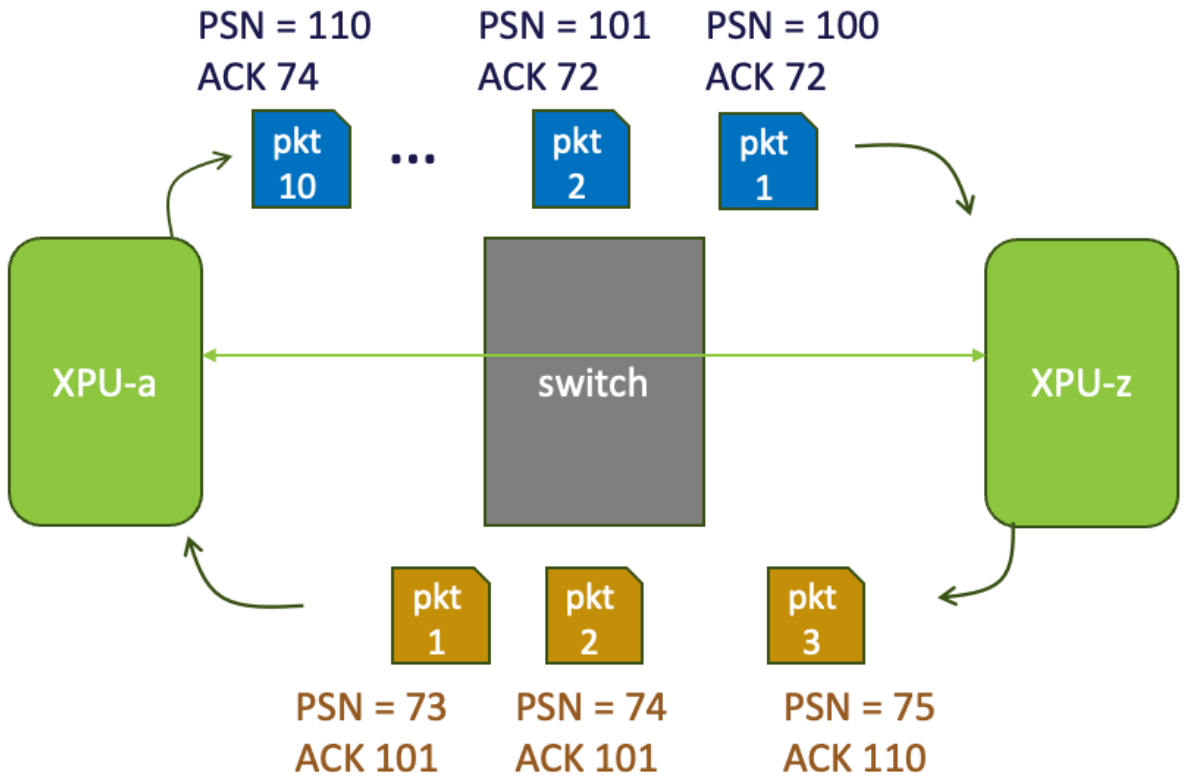
SUE-T Reliability Header

byte 0	byte 1	byte 2	byte 3
type = RODL	next_hdr	flags	cumulative ACK / clear
psn			
source connection ID		destination connection ID	

Wire efficiency (per payload size)	256B – 85% 1KB – 96%
Die size (5nm estimate)	LLR only < 2mm2 Retx (6us) < 3mm2

Acknowledgements and Error Handling

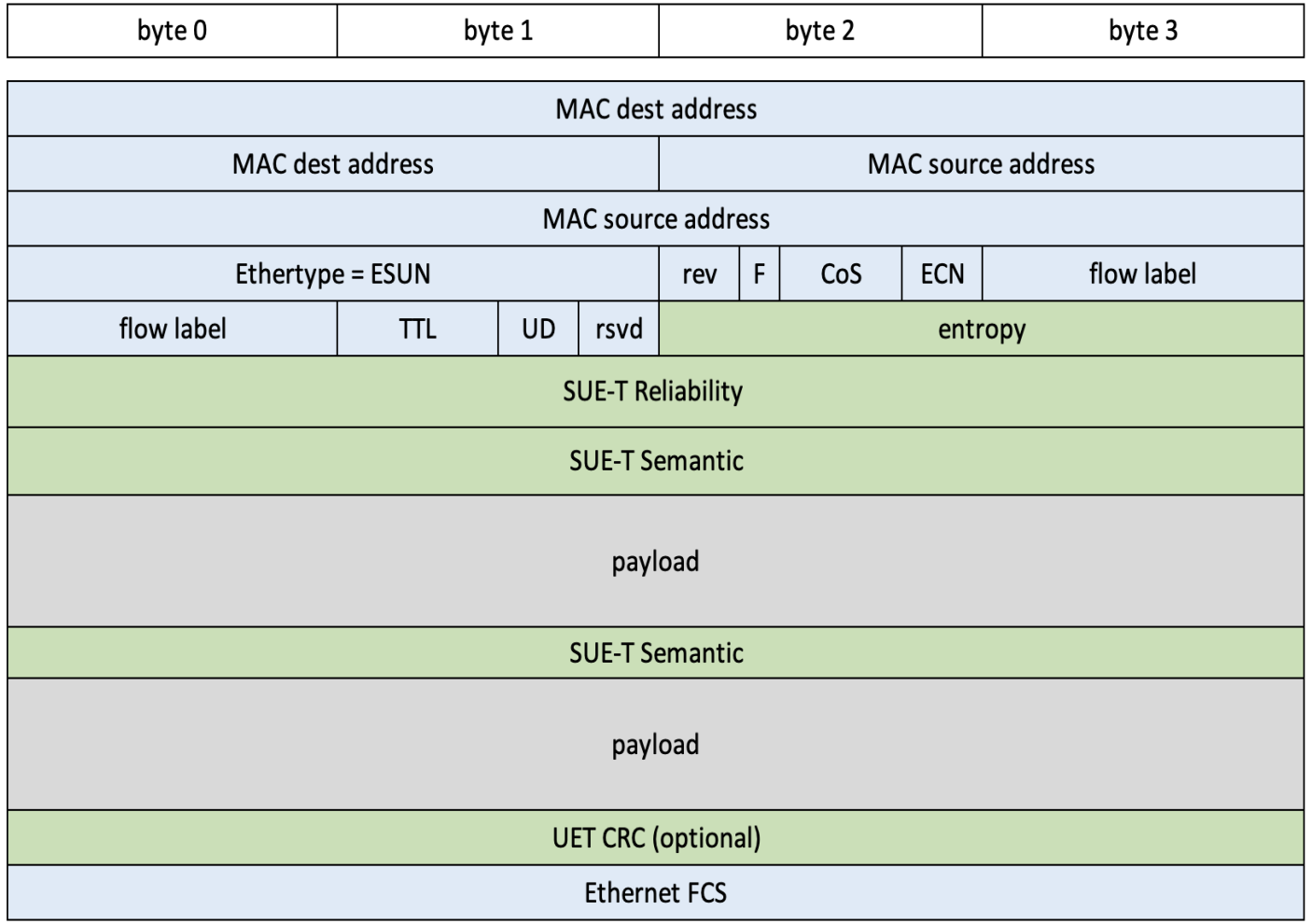
- UET Mode – enhanced error handling
  - ACK indicates destination successfully processed the transaction
  - Unsuccessful transactions return error – this error state is saved until cleared
  - ‘Operation ACK mode’
- Delivery Mode
  - ACK only indicates the transaction was delivered successfully to destination
  - Separate message used to carry response to transaction
  - ‘Delivery ACK mode’



OCP ESUN

- ESUN Header
  - Mac header addressing
  - IP-like Forwarding Semantics
- Lossless Link Technologies
  - CBFC, LLR
- UDP encapsulation supported
  - UET destination port 4793

SUE-T over ESUN



Come work with us

- Join the SUE-T working group
  - https://ocp-all.groups.io/g/ocp-net-sue-t
- Join UEC to contribute to ULN and other adjacent technologies
  - https://ultraethernet.org/