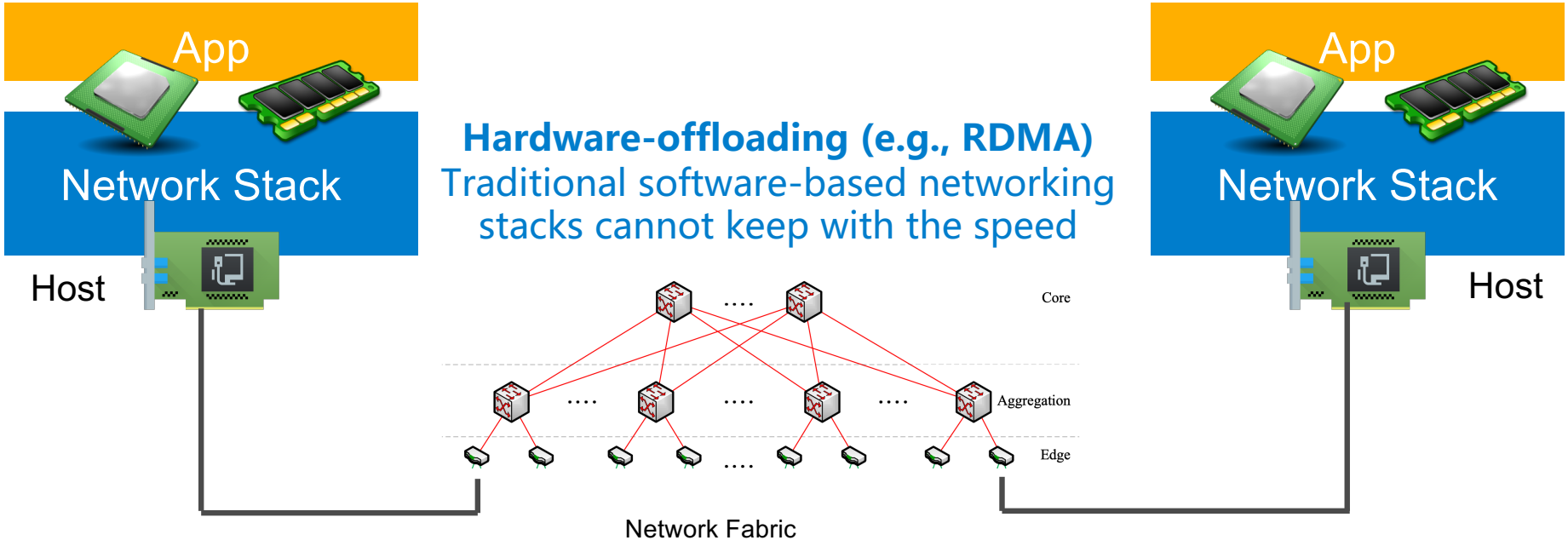
NDP with SONIC-PINS: A low latency and high performance datacenter transport integrated into SONIC

NDP Team: Costin Raiciu, Mark Handley et al

Intel Team: Reshma Sudarshan, Rong Pan, JK Lee

Long Network Stack Latency



Hardware-offloading (e.g., RDMA)
Traditional software-based networking stacks cannot keep with the speed

network is more vulnerable to congestion & packet loss
Congestion control (CC)

NDP Goals:
Latency, Latency & Latency

- **Low latency,** predictable request/response behavior
  - Zero-RTT setup
  - Fast start up
  - Very short switch queues
  - Graceful Incast behavior

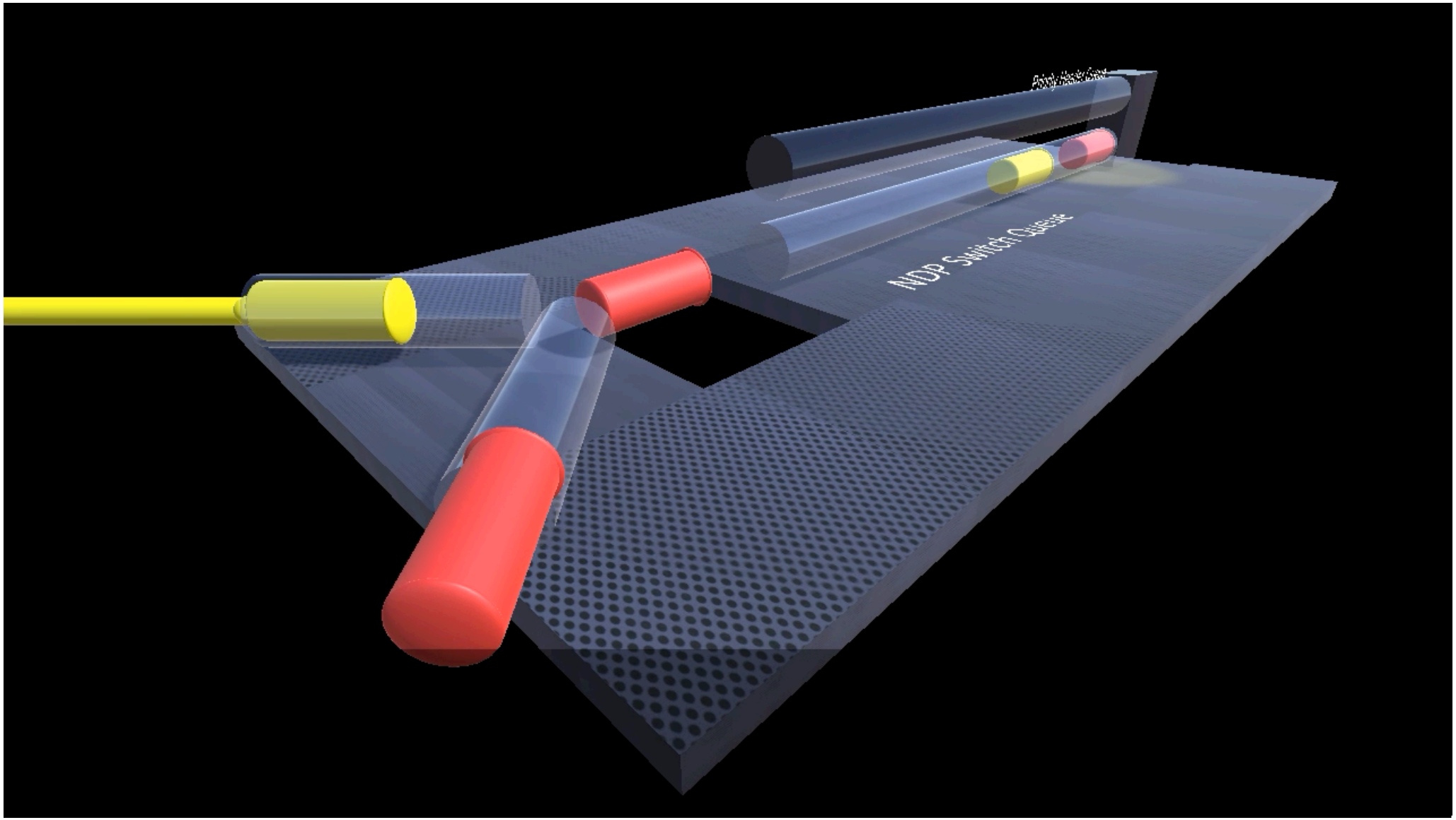
- Receiver **prioritization**
  - Receiver knows which flows are most important at this instant

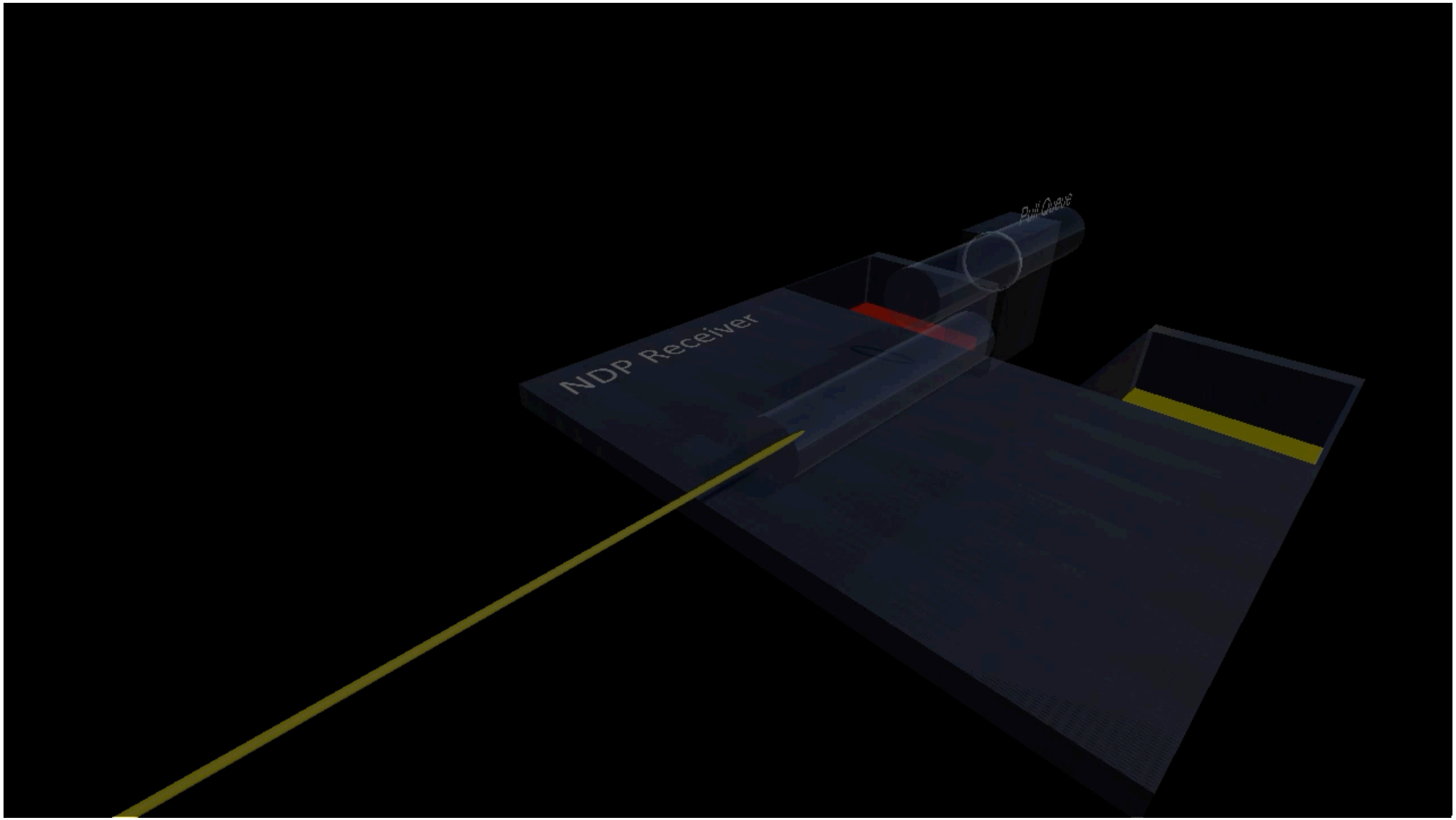
- Predictable **high throughput**
  - Per packet load balancing

Next Few Slides from NDP SIGCOMM Presentation: https://dlnext.acm.org/doi/abs/10.1145/3341302.3359768

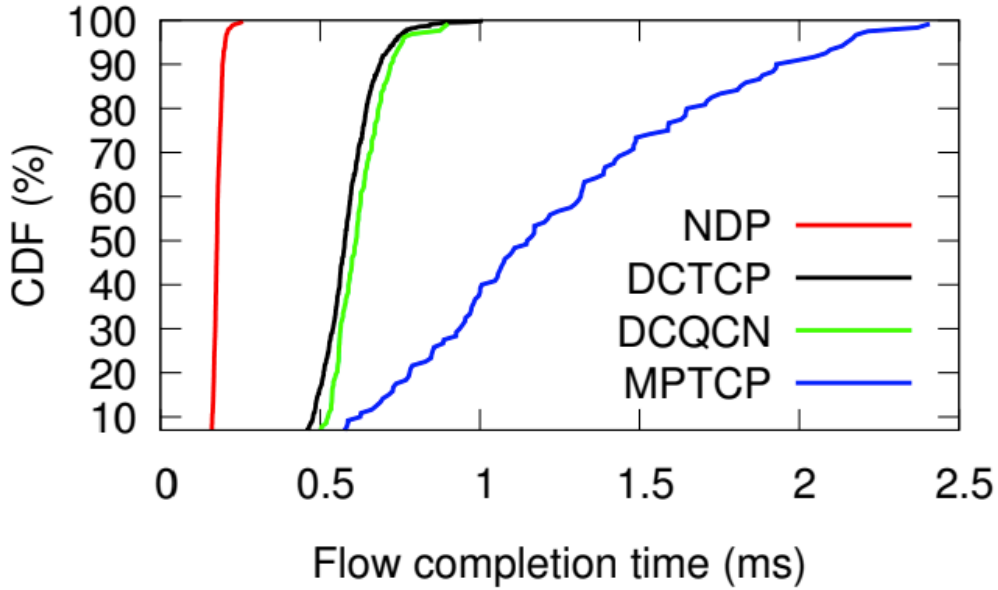
NDP: Just Start, Spray, Trim, Pull

- Sender:
  - No prior handshake (zero RTT setup)
  - Send 1st RTT of data at line rate
  - Per-packet multipath load balancing
  - fast RTX
- Network:
  - Packet Tramming + prioritization
- Receiver:
  - After 1st RTT, receiver pulls at its own rate
  - Receiver can choose which senders to pull first


NDP Switch Queue


NDP Receiver

Performance: Achieve Low Latency



- **FCT for 90KB flows with random background load, 432 node FatTree**
  - NDP and MPTCP achieve 80% network utilization, whereas DCTCP and DCQCN achieve ~75%

NDP in SONiC

Implementation

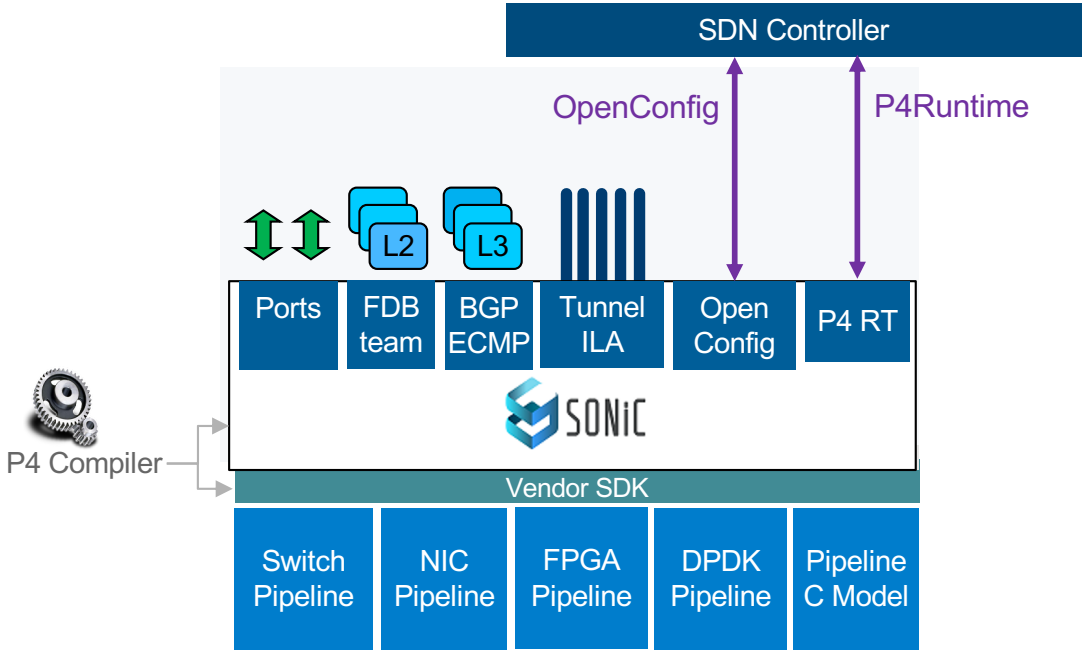
- New mirror types for DoD, trimmed, notification packets
- New metadata headers to carry NDP-specific variables between ingress and egress
- Adapted ECMP hash for new NDP hash field
- 3-level meter schema

Challenges

- To implement NDP in SONiC we require addition of new tables in SAI pipeline
- The new HW tables cannot be manipulated by regular SONiC
- Implemented NDP in proprietary method to program the P4 pipeline using BFRT
- Need to transition into standard methods to configure these tables

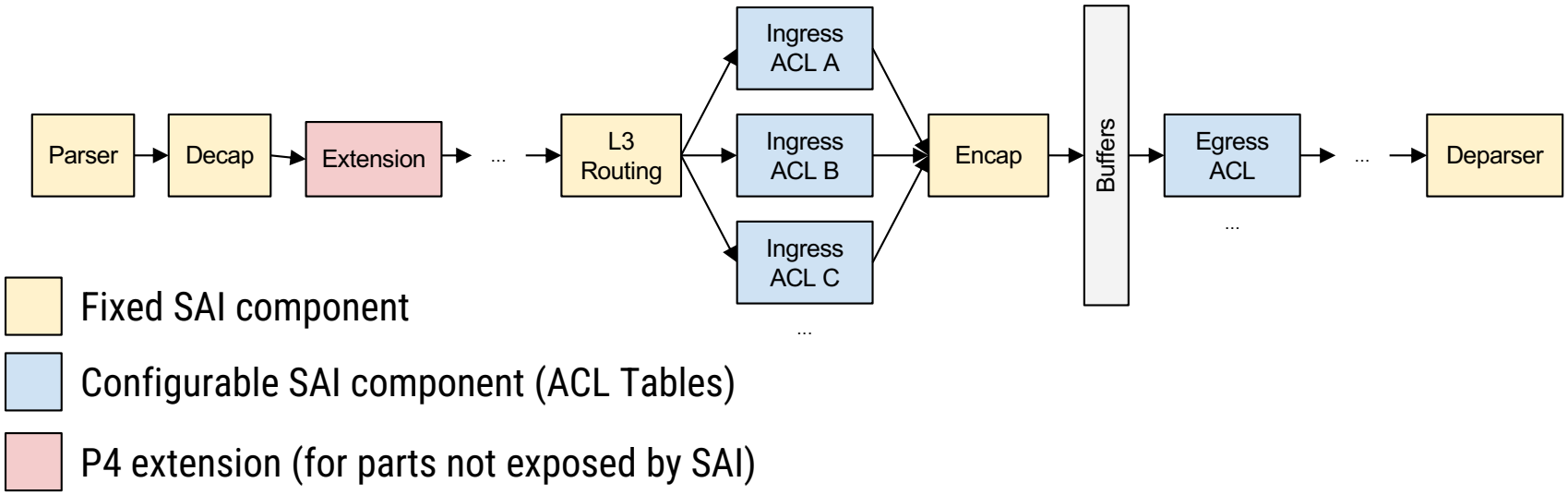
P4 Integrated Networking Stack

- SDN feature in SONiC
- Usage of SONiC for local Control Plane augmenting the Data Plane with SDN
- Uniform P4Runtime abstraction
- Solution enables
  - SDN managed SAI tables
  - Non-SAI Extension tables
- Common environment across Host, NIC & Switch
- Field upgradeable and extensible
- Brings agility and differentiation with specialized use cases
  Customer-specific network headers and data plane functions

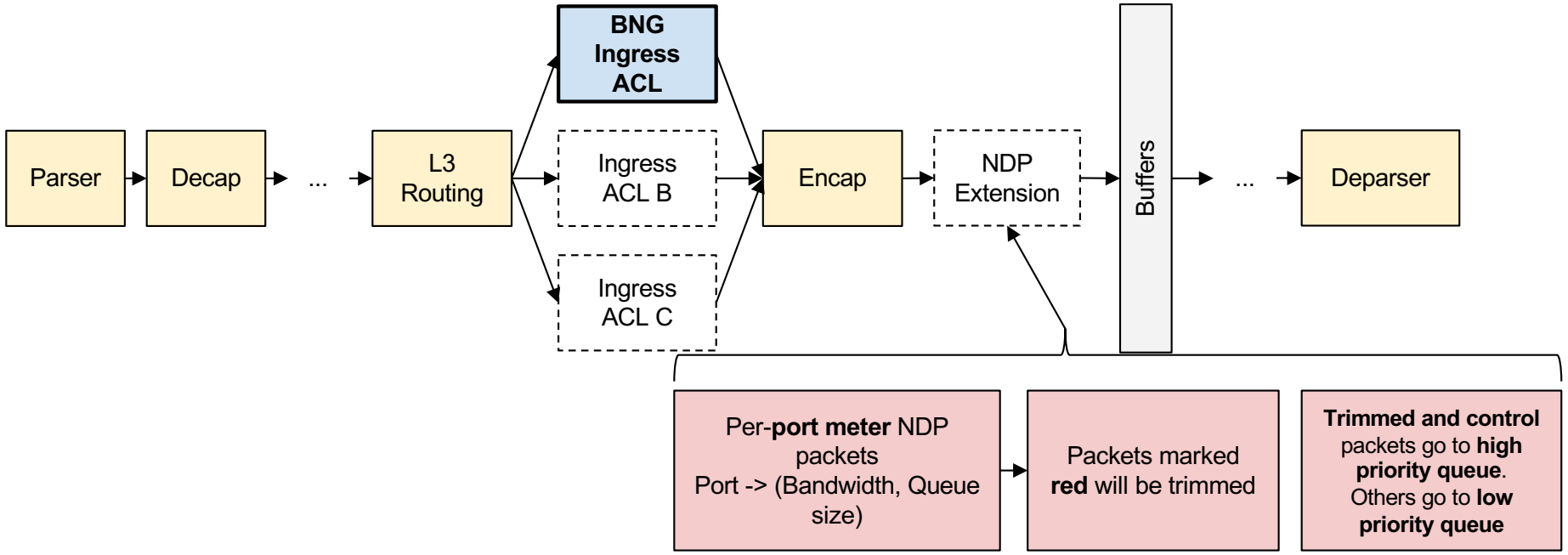


SAI Pipeline

Remote SAI with a P4Runtime based SDN interface. The interface is derived from a canonical family of P4 programs that describe the packet forwarding pipeline of SAI.

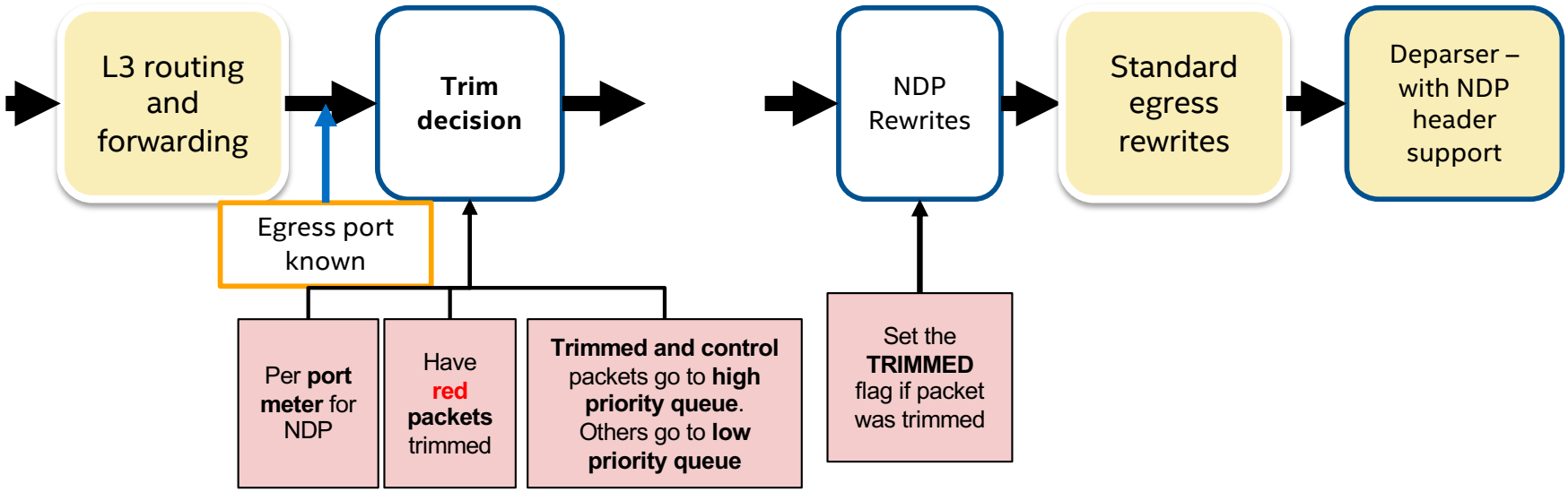


NDP Extension to SAI.p4 Pipeline



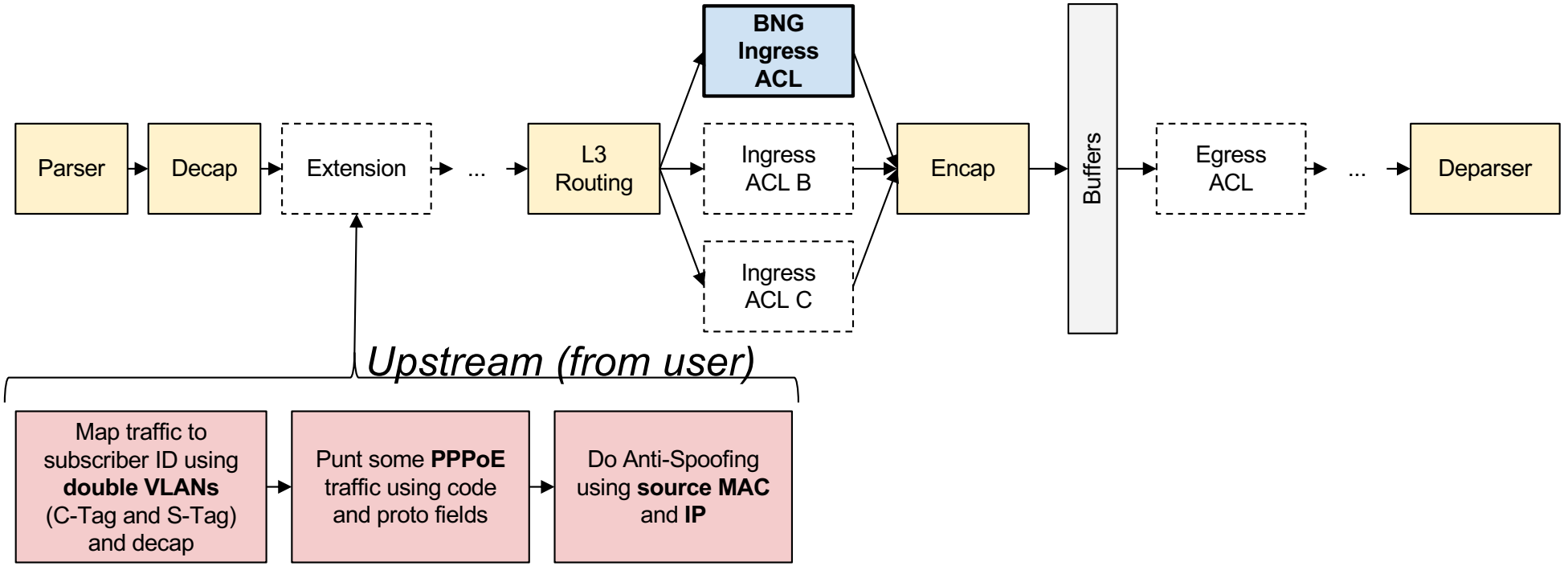
- Ingress: extract NDP headers and apply metering logic
- If packet is marked **red**, then mark the packet for trimming in egress
- Egress: if packet was marked for trimming, set the NDP TRIMMED flag

NDP SONIC P4/SAI integration



- Ingress: extract NDP headers and apply metering logic
- If packet is marked **red**, then mark the packet for trimming in egress
- Egress: if packet was marked for trimming, set the NDP TRIMMED flag

BNG Extension to SAI.p4 Pipeline



- Upstream: Decapsulate double-VLAN and PPPoE headers and validate packet
- Not pictured: MyStation Table, two counters (before and after traffic manager) for accounting, downstream

Use Cases Requiring Extensions to SAI Pipeline

SONiC as an Access Router

- BNG *for DSL/Cable/FTTP access networks*
- 5G UPF (4G SPGW) *for mobile access networks*

SONiC native support for Virtual Networking

- OVN: Virtual Overlay / Physical Underlay Network
- VNET Peering

Other Use Cases

- RTP over multicast for streaming video
- Server Load Balancer
- Custom telemetry collection

P4 is already being used to describe these pipelines in available solutions.

PINS Advantages over direct ASIC configuration

Allow SDN controllers to participate in the SONiC ecosystem with a standard way to control a running P4 forwarding switch

P4RT API is standard, open and silicon-independent and enables runtime-control of both programmable and fixed-function P4 forwarding planes

One P4 program to configure all the way to the ASIC for both SONiC/SAI objects as well as SAI-extensions P4 to SAI conversion in SONiC framework

PINS allows sharing of objects between SONiC and the Controller applications

Work in SAI compliant environment and have standard framework to extend pipeline to various proprietary applications

Thank You

Contact Information
Additional Helpful Links