# DNX (Jericho) Buffering

# Buffer Sizing Meeting @ Stanford – Mar 2019

**Golan Schzukin**

March 2019

BROADCOM®

# Agenda

- High Level Overview of Jericho Buffering
- Configurations
- Measurements and Visibility

- What can impact buffer size
- What are the Metrics for success or failure



BROADCOM®

# Overview – Jericho

- On-Chip-Buffer 16MB
- External buffer: 6GB (8xGDDR5/DDR4), 2KB buffers
- Buffering resources are allocated to core by API
- VOQ can be  designated as either
  - OCB-Only
  - DRAM-Mix: Use OCB buffers up to a threshold, above which it uses DRAM resources.
- VOQ admission attributes
  - Tail-Drop
  - WRED curve per DP
  - ECN marking curve
  - FADT



BROADCOM

# Congestion Management

- 3 Resources maintained
  - SRAM-Buffers
  - SRAM-Packet-Descriptors
  - DRAM-Buffers

- VOQ & VSQ metrics
  - SRAM-Buffers
  - SRAM-Packet-Descriptors
  - Total-Words; Sum of SRAM and DRAM bytes(16B) resolutions
- Global Metrics
  - Free PDB and BDBs



BROADCOM

# Buffering Allocation Model

- Guaranteed Space
  - Per VOQ, Source-Port, PG
- Shared Space
  - Can be partitioned into 2 pools – Port & TC based
    - E.g. Lossy and Lossless
  - Limits per VOQ, PG, Source-Port, VSQ …
    - Static & dynamic (FADT)
  - When over limit
    - Lossy – Drop
    - Lossless – Generate PFC and use Headroom
- Headroom Space
  - Dedicated space - Absorb lossless after PFC
  - Limits per PG, Source Port, Total (sharing option)



BROADCOM

# Buffering controls

- Allocation of Guaranteed, Shared and Headroom Spaces
  - DRAM & OCB-Only
- Allocation of more specific limits per Space
  - Per VOQ, Source Port, PG, VSQ
- Setting of VOQ, VSQ parameters – Tail drop, WRED, ECN

- Allocation of DRAM bandwidth
  - For what % of traffic deep buffers are needed
  - #of DRAM devices
  - Shaping DRAM access BW -?



BROADCOM

# Visibility

- Dropped packets/octets per VOQ/globally
- Max VOQ size – Per VOQ, 8 Largest
- Max VSQ size – Selecting virtual group of packet/VOQs
- Max packet latency - Per Latency-Flow, 8 largest
- Low water marks of free resources
  - DRAM total
  - OCB total
  - Shared
- Rates – Counters that can be activated for a time period

- Time stamping INT and other instrumentation capabilities in recent devices



BROADCOM

# General

# What can impact buffer size

- Lossless & PFC - How much of the memory is assigned to headroom? and how to prevent/minimize PFC and PFC blocking nature
- Network utilization  - Higher unitization requires larger buffers.
- End to End congestion control – E2E Scheduling, fast congestion indication from network (ECN, TCP, RDMA, SCH at NICs) ...
- End to End latency - Reaction of the end points to network congestion
- Load balancing - Non ideal load balancing has some adverse effects on utilization
- Applications – What is the nature of applications. Do we control them them (Burst, Max rate, Synced In-cast)
- Robustness – Should it work for Current load or be ready for future app/loads
- Resiliency - Behavior under failure scenarios – Impact vector
- Rate mismatch between network layers
- QOS, SLA – Shaping



BROADCOM

# Metrics for results

- Throughput / Good-put
- Drops – What is a passing threshold?
- FCT
- SLA contract
- User experience – Like Video quality, Reaction time



BROADCOM