UNIVERSITY OF CAMBRIDGE

P51: High Performance Networking

Lecture 5: High Throughput Devices

Dr Noa Zilberman
noa.zilberman@cl.cam.ac.uk

Lent 2017/18

Very High Throughput Switches

The Truth About Switch Silicon Design

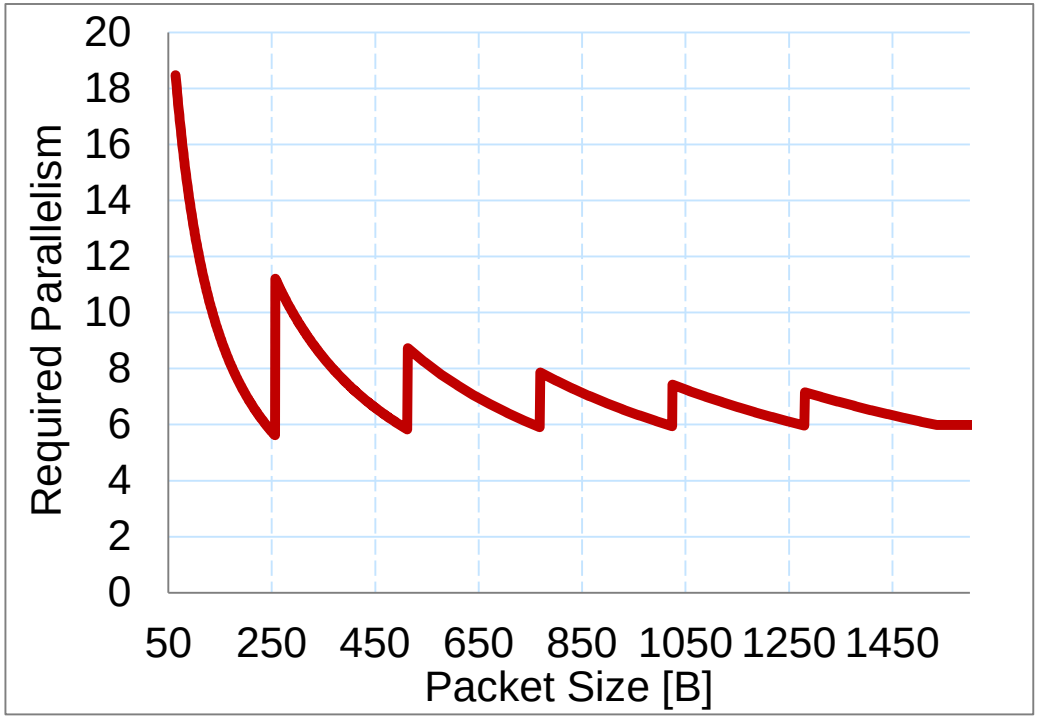
12.8Tbps Switches!

Lets convert this to packet rate requirements:

5.8 Gpps @ 256B

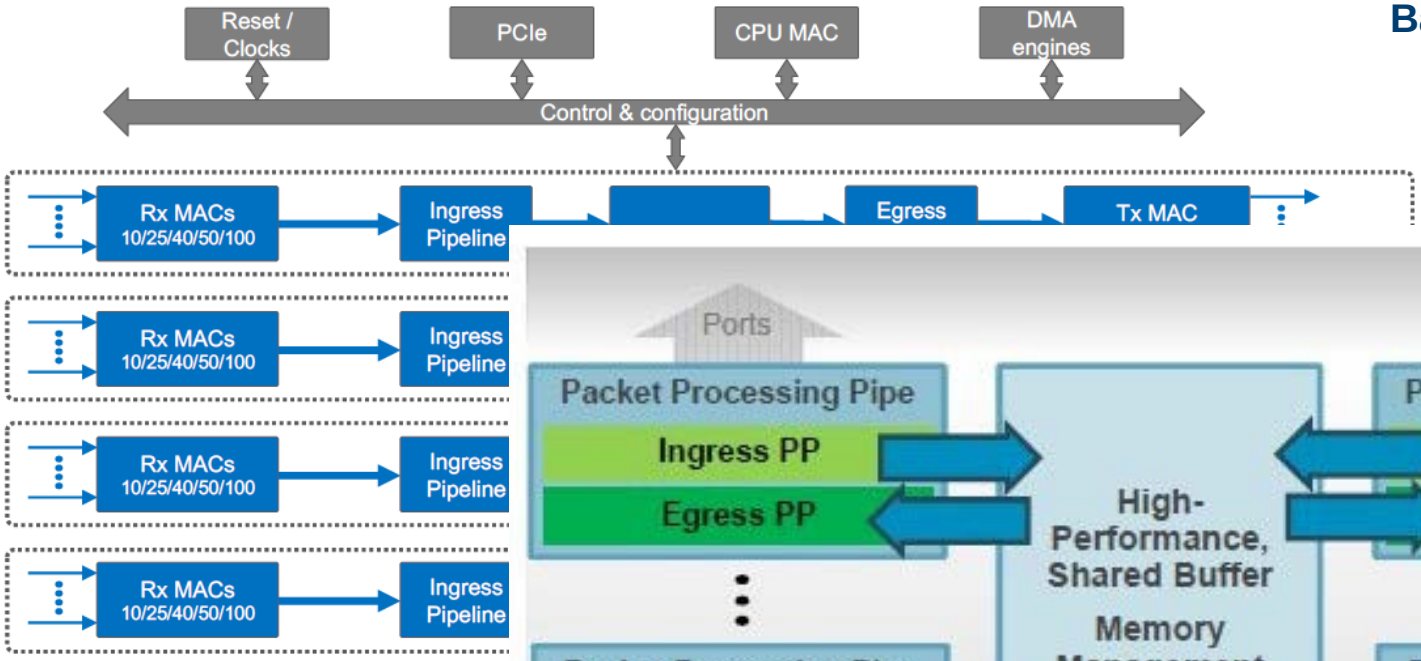
19.2 Gpps @ 64B

But clock rate is only ~1GHz….

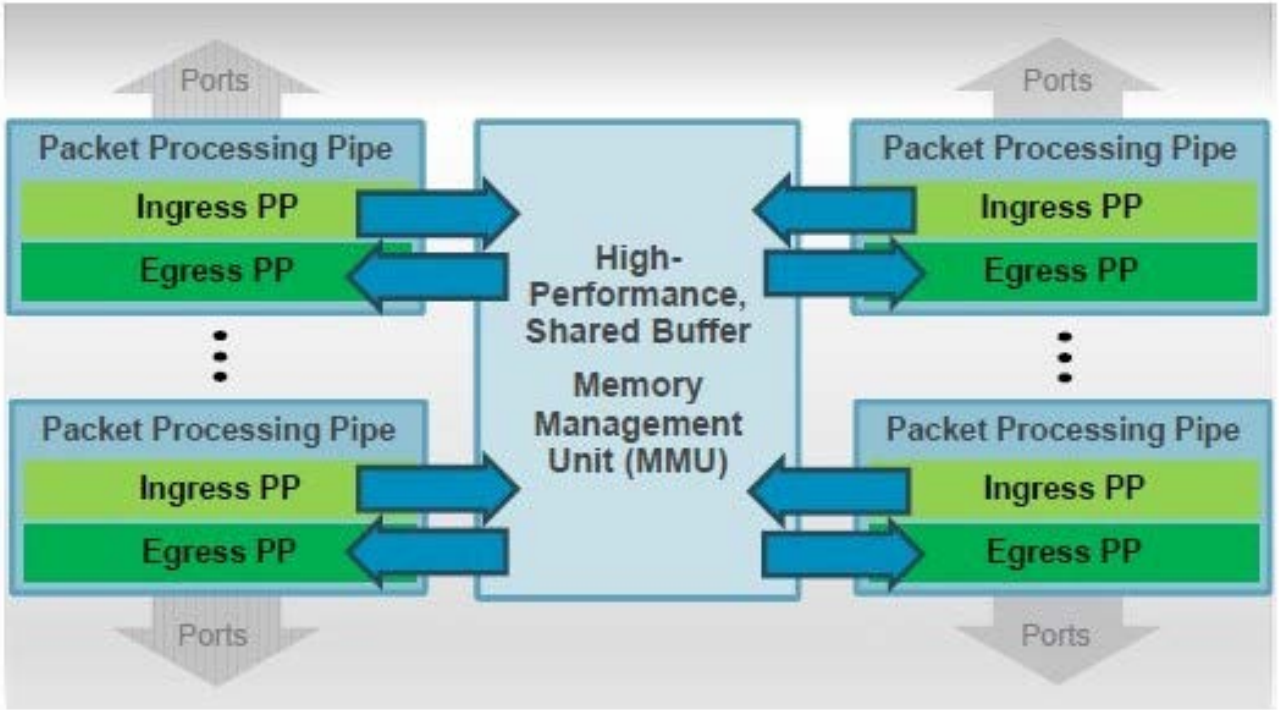


Multi-Core Switch Design

Barefoot Tofino



Each pip
Additional ports



Broadcom Tomahawk 3

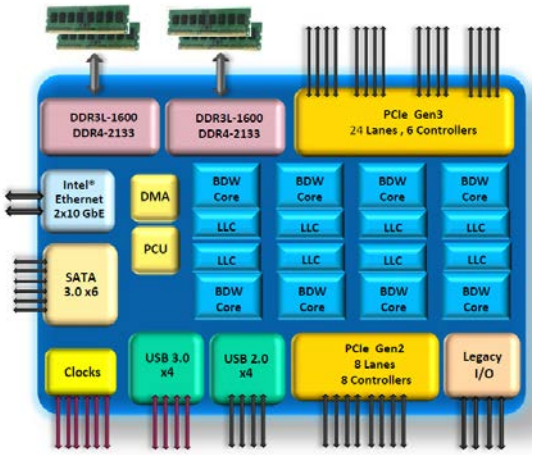
Image sources: https://p4.org/assets/p4_d2_2017_programmable_data_plane_at_terabit_speeds.pdf
https://www.nextplatform.com/2018/01/20/flattening-networks-budgets-400g-ethernet/

Multi Core Switch Design

- So what? Multi-core in CPUs for over a decade
- Network devices are not like CPUs:
  - CPU: Pipeline - instructions, memory – data
  - Switch: pipeline – data, memory – control
- Network devices have a strong notion of *time*
  - *Must* process the header on cycle X
  - Headers are split across clock cycles
  - Pipelining is the way to achieve performance

Multi Core Switch Design

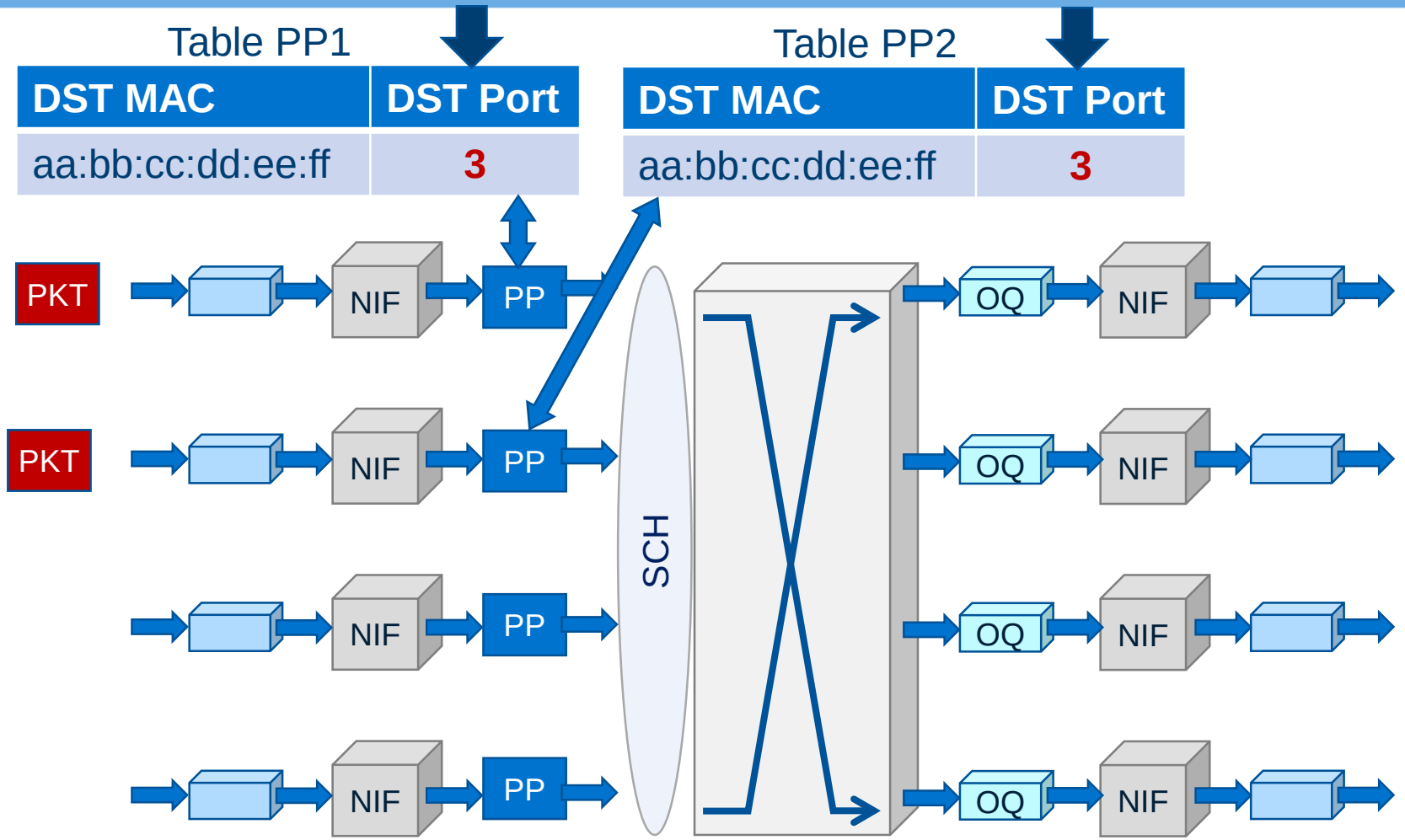
- The limitations of processing packets in the host:
- DPDK: can process a packet in 80 clock cycles
  - Lets assume 4GHz clock (0.25ns/cycle)
  - Can process $4 \times 10^9 \div 80 = 50 \times 10^6$ pkts/sec
  - 50Mpps is not sufficient for 40GE. 30% of 64B packets at 100GE.
  - Can dedicate multiple cores…
  - And this is just sending / receiving, not operating on the packet!



Multi Core Switch Design

- The problem with multi-core switch design: look up tables.
  - Shared tables:
    - need to allow access from multiple pipelines
    - need to support query rate at packet rate
  - Separate tables:
    - wastes resources
    - need to maintain consistency
      - Not everyone agree with this assumption

Multi Core Switch Design

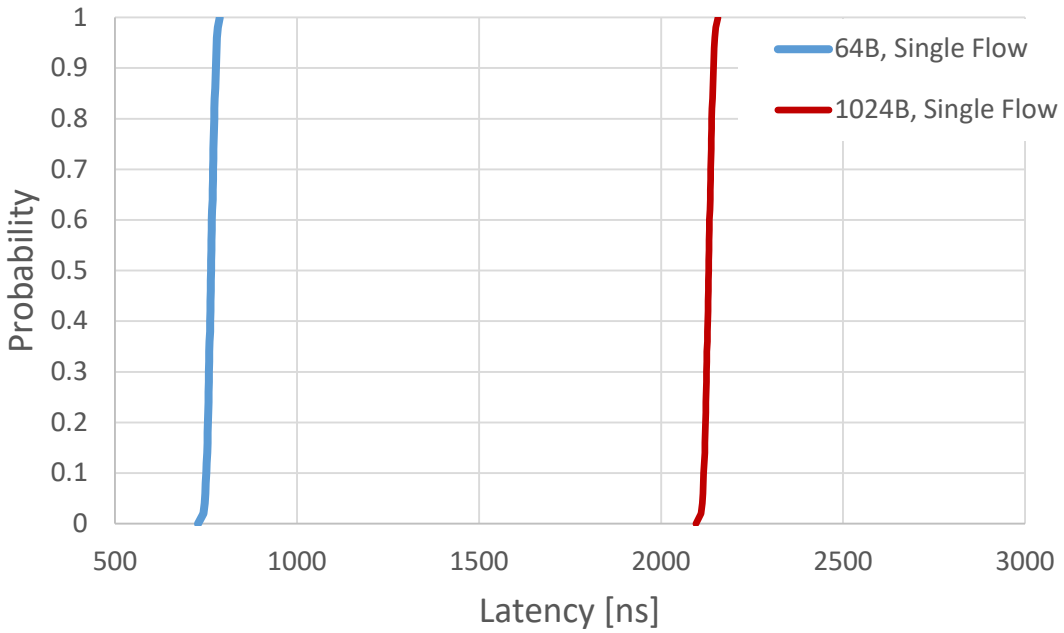


Inferring Switch Architecture

refer to:
http://www.mellanox.com/tolly/

Inferring NetFPGA Switch Architecture

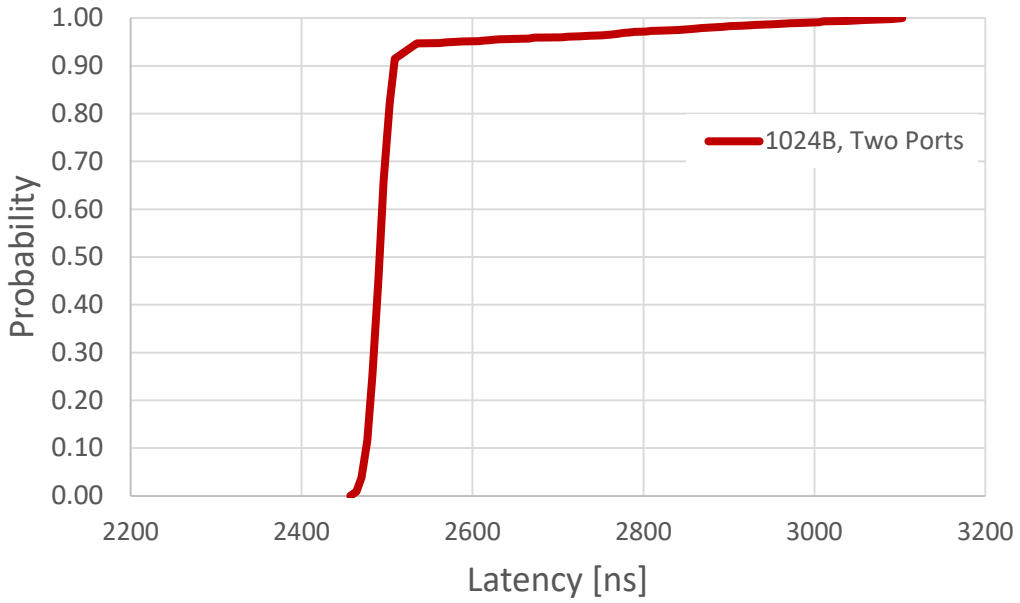
- The following graph shows the latency CDF of a single packet going through the switch:



- What does this graph tell us?

Inferring NetFPGA Switch Architecture

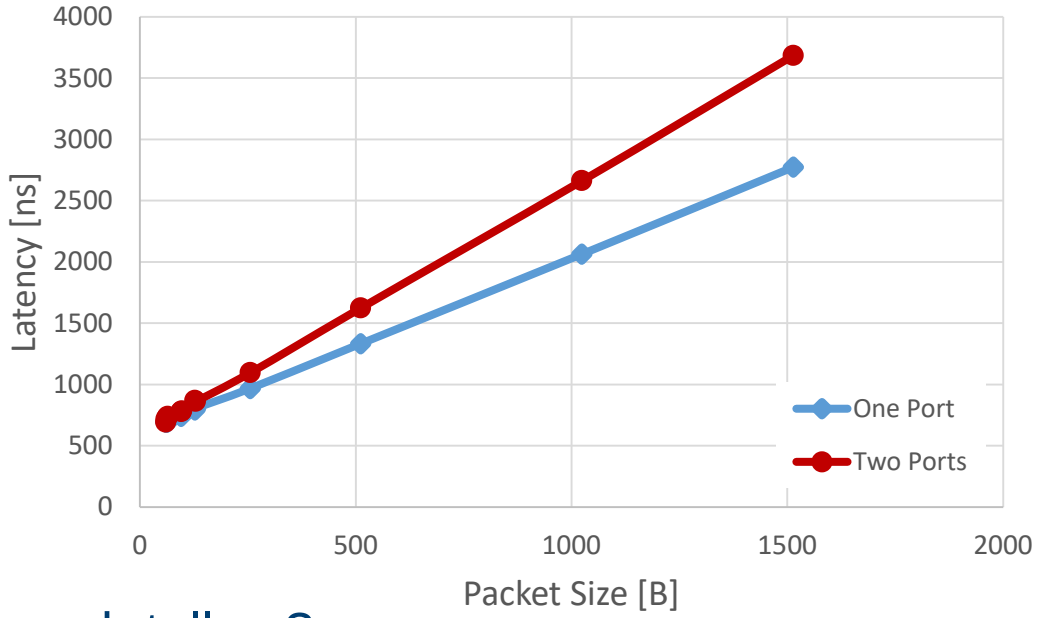
- The following graph shows the latency CDF of packets sent from Port 0 to Port 1, where there are (in parallel) packets sent from Port 1 to Port 0:



- What does this graph tell us?

Inferring NetFPGA Switch Architecture

- The following graph shows the maximum latency of packets sent from Port 0 to Port 1, (a) when there is no other traffic and (b) when there are (in parallel) packets sent from Port 1 to Port 0:



- What does this graph tell us?