

P51: High Performance Networking

Lecture 1: Introduction

Introduction to the course

Administrivia

Scope:

- High performance networking design and usage.

Course structure:

- Lectures – 6 hours – FS07
- Supervised Labs – 10 hours - SW02 (ACS lab)

Assessment:

- Practical Assignment (100%) – 24/04/2019 12:00

Schedule

Week	Lecture	Lab
1	General architecture of high performance network devices	Introduction to NetFPGA (SW01)
2	Programmable devices	Introduction to NetFPGA (Cont.) Project selection
3	High throughput devices – Part I	Project architecture
4	High throughput devices – Part II	Performance profile
5	Low latency devices - Part I	Evaluation
6	Low latency devices - Part II	

Project

- Starting point: a reference design of a network device
- Goal: Design a high performance application
- Examples:
 - Line-rate network monitoring
 - Line-rate KVS
 - More examples on the website
- Projects done in pairs
- More information in Lab 1

Some logistics for 2018-19

Web page: <http://www.cl.cam.ac.uk/teaching/current/P51/>

Mailing list: *cl-ac-s-p51-announce@cam.ac.uk*

Grades:

Mphil (ACS) – Pass / Fail - based on a mark out of 100

All others (DTC) – Mark out of 100

Next steps

- Explore the web page

<http://www.cl.cam.ac.uk/teaching/current/P51/>

- Decide if you still want to take the class – promptly

- Project:

- Pair with a classmate
- Register to NetFPGA repository

http://netfpga.org/site/#/SUME_reg_form/

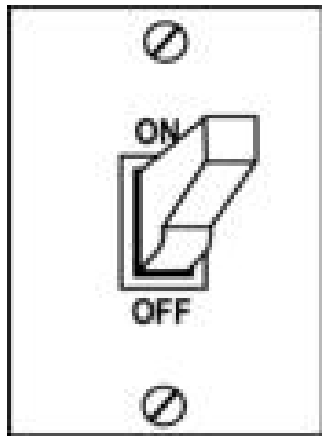
Register to the P4-NetFPGA repository

<https://goo.gl/forms/h7RbYmKZL7H4EaUf1>

General architecture of high performance network devices

What Is a Switch?

We use switches all the time!



ON / OFF



Left / Right

What Is a Network Switch?

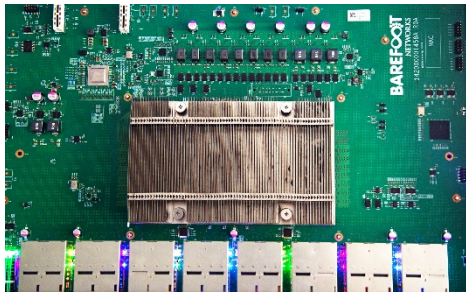
Conceptually, a left / right switch...

- Receives a packet through port $\langle N \rangle$
 - Decides through which port to send it
 - *A forwarding decision*
- + Some “real world” considerations



Real World Switches

- High Throughput Switch Silicon: 6.4Tbps (64x100G) – 12.8Tbps (32x400G)
Top of Rack Switches
 - E.g. Broadcom Tomahawk III, Barefoot Tofino, Mellanox spectrum II
- High Throughput Core Switch System: >100Tbps
 - E.g. Arista 7500R series, Huawei NE5000E, Cisco CRS Multishelf

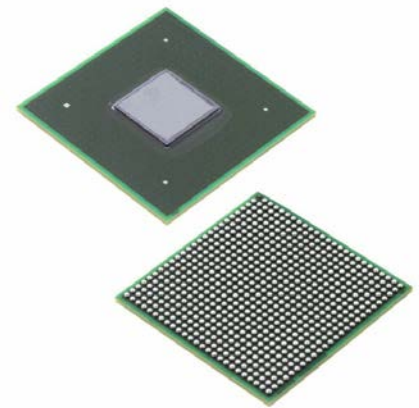


Real World Switches


- Low latency switch (Layer 1): ~5ns fan-out, ~55ns aggregation
- Low latency switch (Layer 2): 95ns - 300ns
 - Examples: g. Mellanox spectrum II, Exablaze Fusion
- Low latency NIC: <1us (loopback)
 - E.g. Mellanox Connect-X, Solarflare 8000, Chelsio T6, Exablaze ExaNIC
- Low latency switches don't always support full line rate!

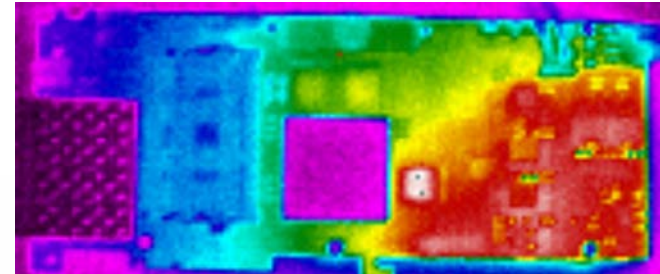
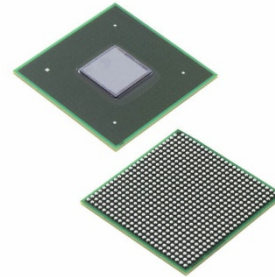
Real World Switch Silicon in Numbers

- Up to 20 Billion Transistors
- Manufacturing process of down to 7nm
- Silicon size: 400 to 600 square mm
- Clock Rate: ~1GHz (typical)
- Packet Rate: ~10 Billion packets per second
- Buffer Memory: ~16MB-30MB on-chip
- Ports: Up to 256
- Power: ~100W-300W
- 2018 Numbers



What Drives The Architecture of a Switch?

- Cost 
- Manufacturing limitations (e.g. maximum silicon size)
- Power consumption
- General purpose or user specific?
- I/O on the package
- Number of ports:
 - Front panel size (24,32,48 ports in 19inch rack)
 - MAC area



Packet Rate as a Performance Metric

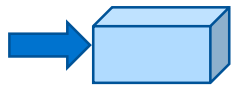
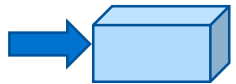
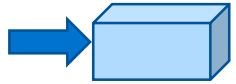
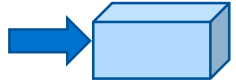
- Bandwidth is misleading
 - For example: full line rate for 1024B packets
but not for 64B packets...
- Packet Rate: how many packets can be processed every second?
- Unit: packets per second (PPS)

- An easy way to calculate the packet rate:
$$(\text{Clock Frequency}) / (\text{Number of Clock Cycles per Packet})$$

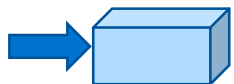
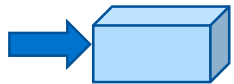
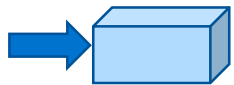
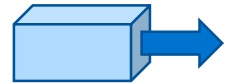
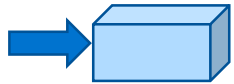
Switch Internals 101

What defines the architecture of a switch?

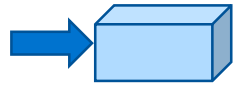
Input Ports



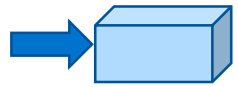
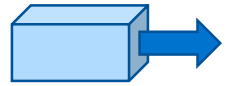
Output Ports



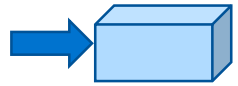
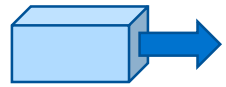
Header Processing



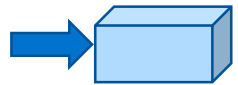
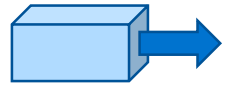
HP



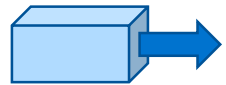
HP



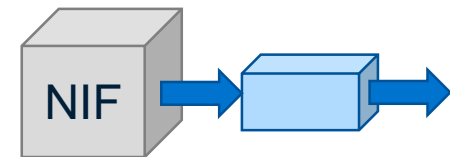
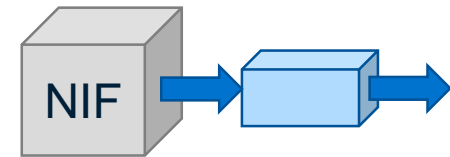
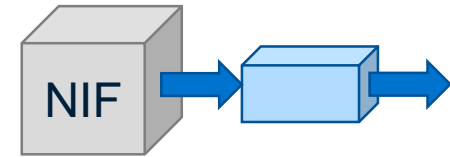
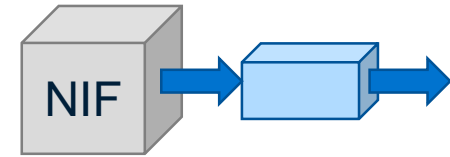
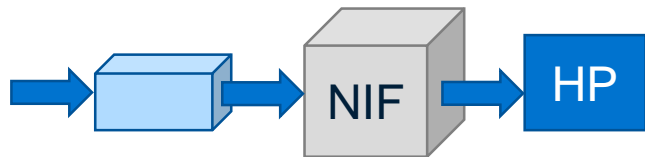
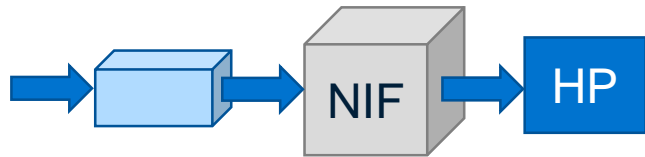
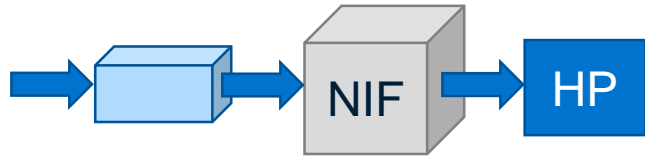
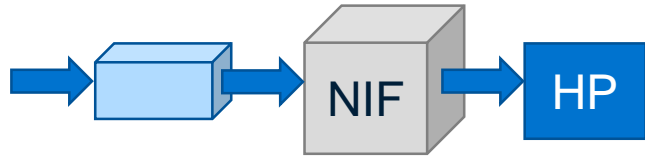
HP



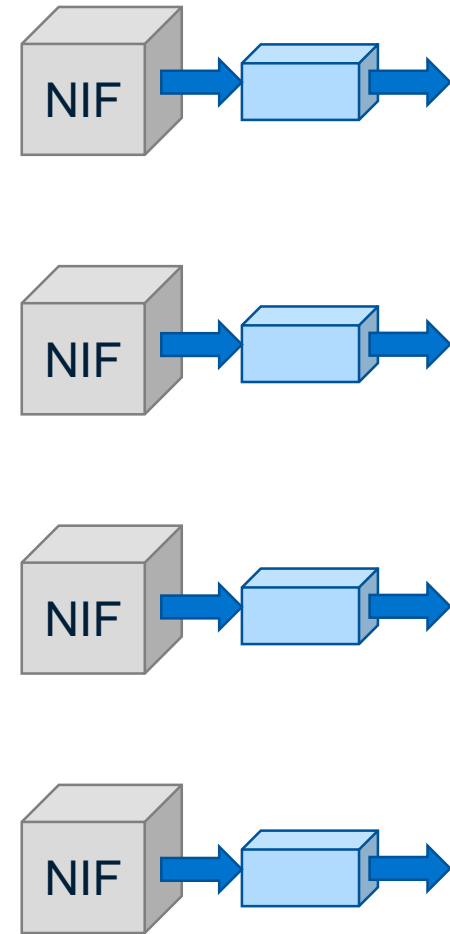
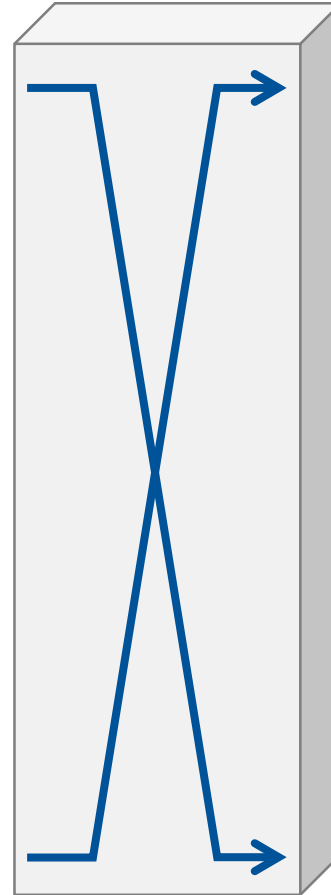
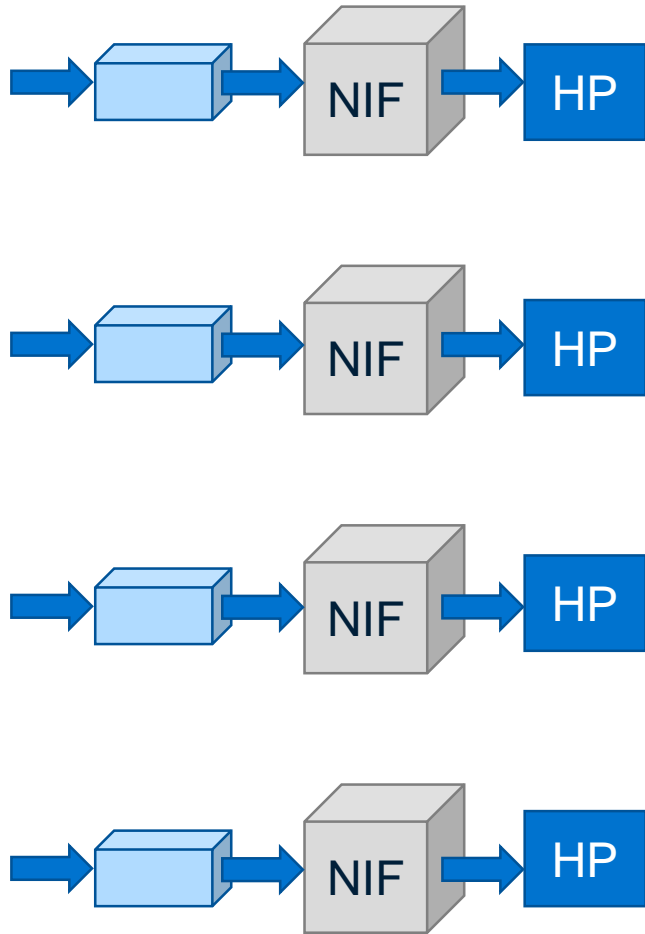
HP



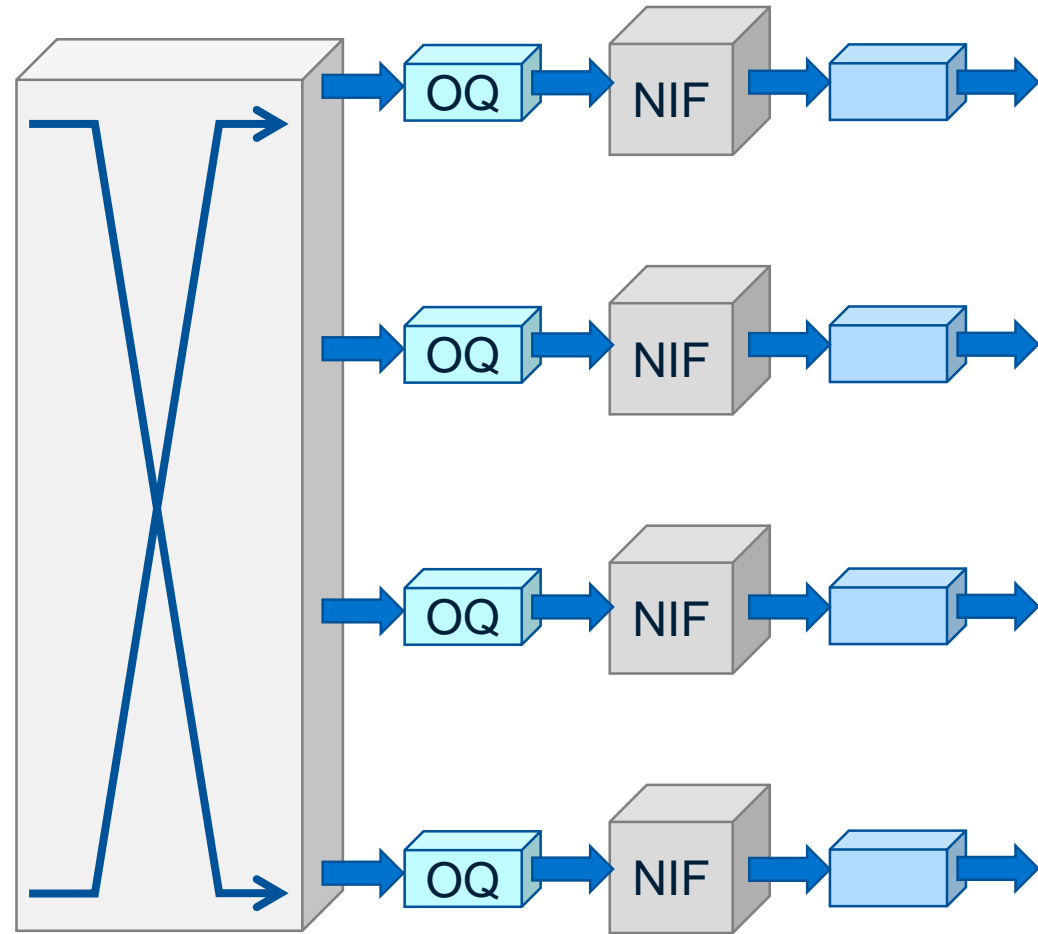
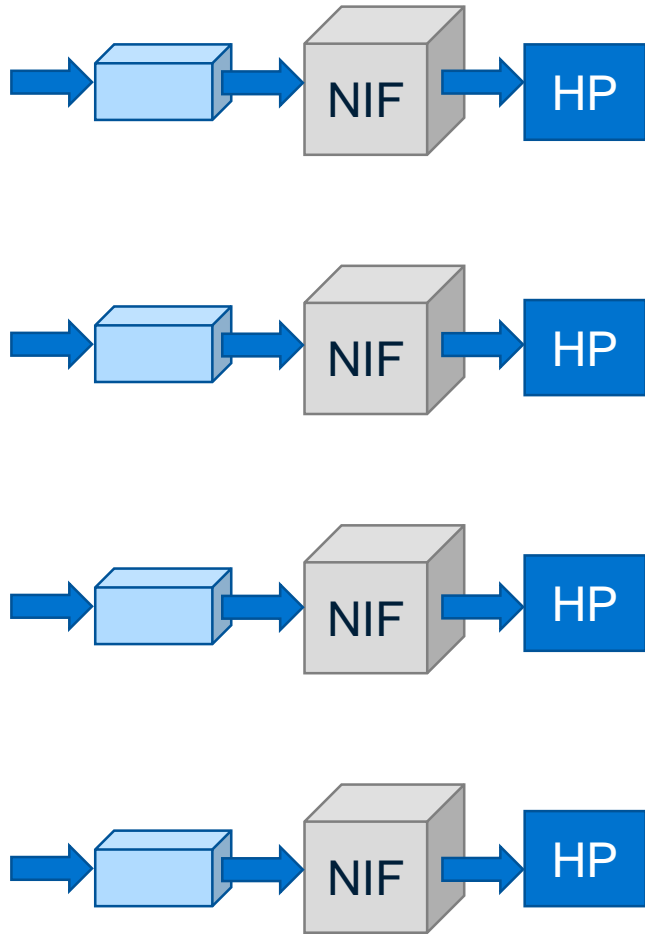
Network Interfaces



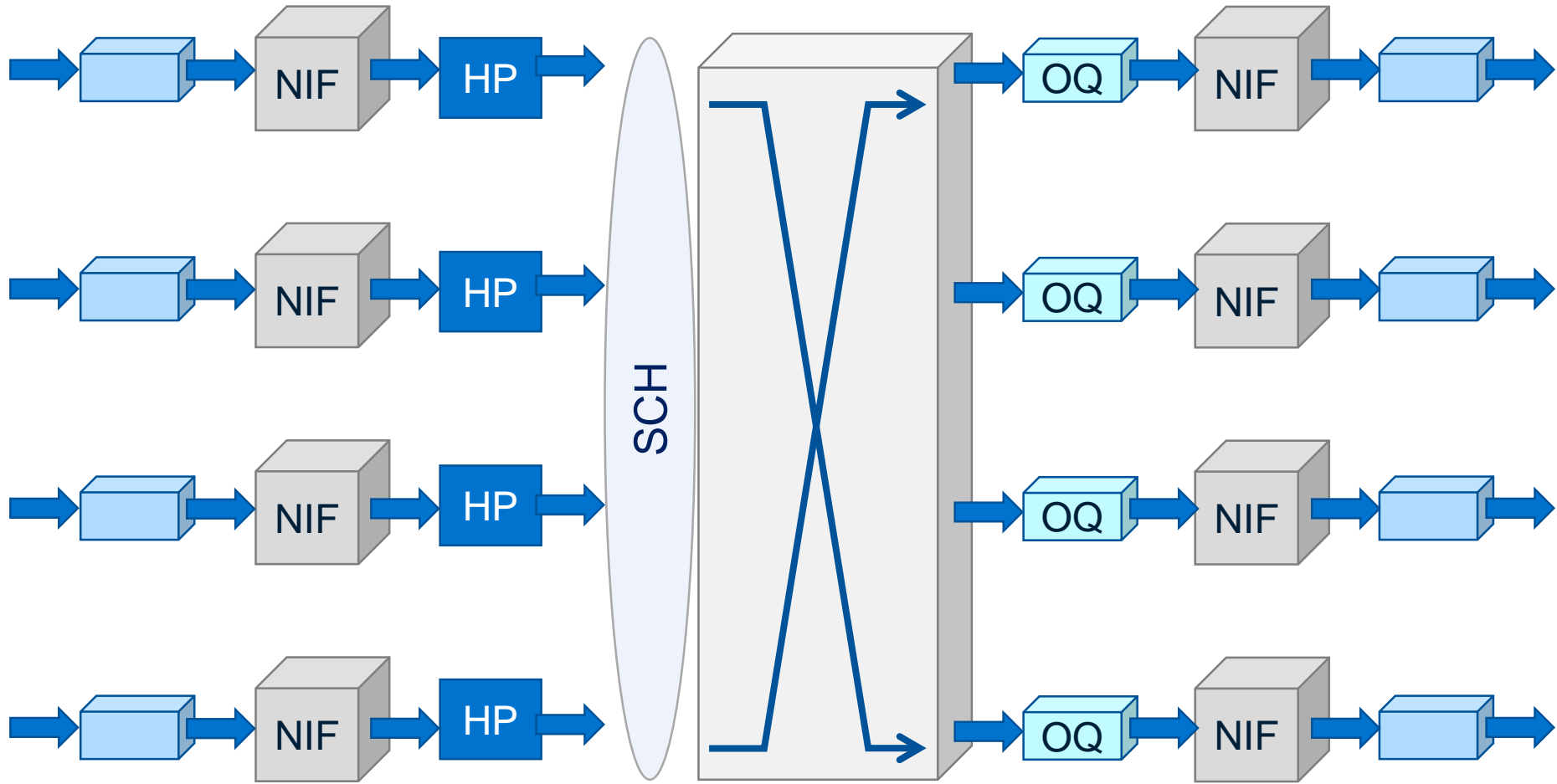
Switching



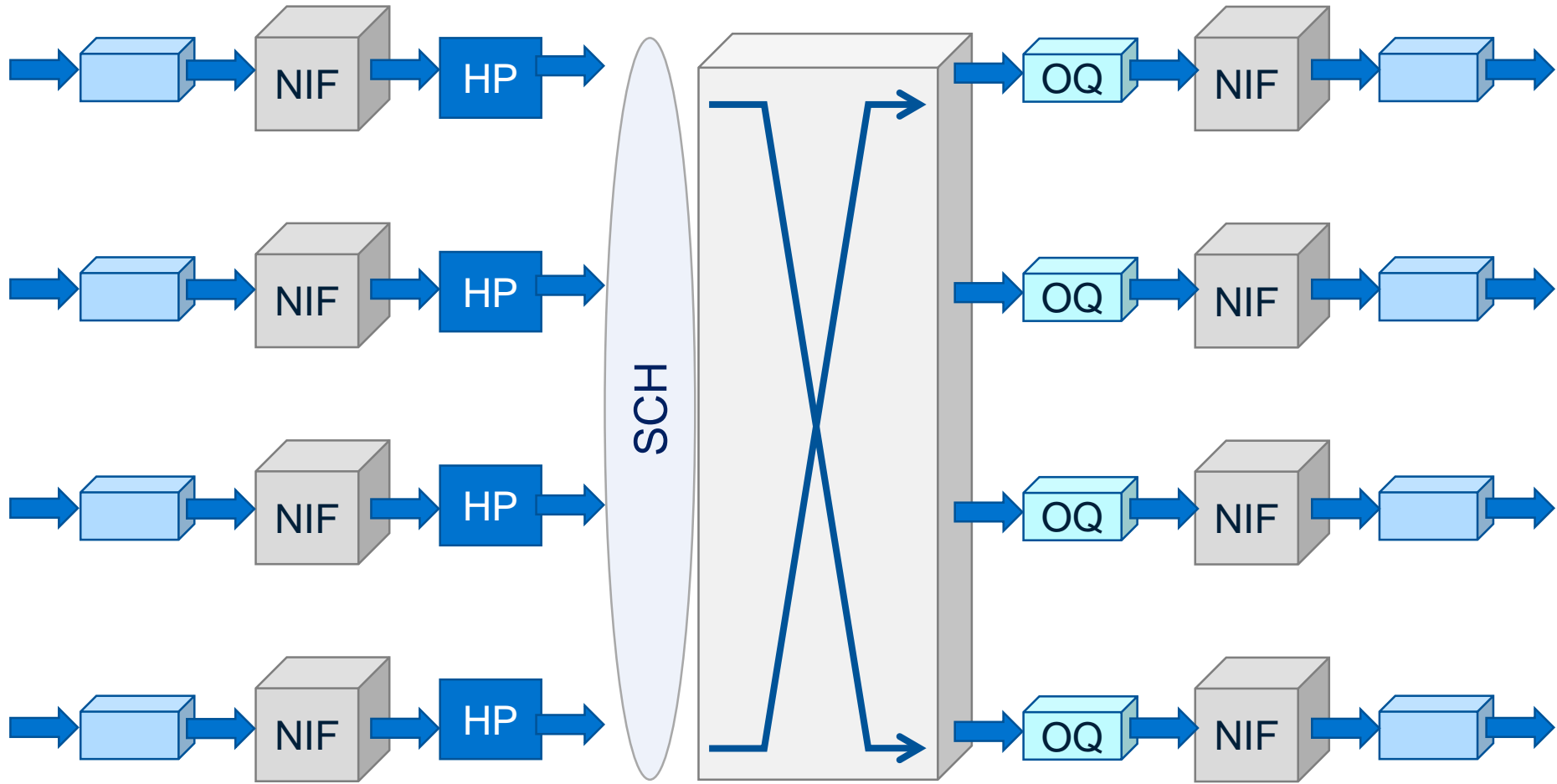
Output Queues



Scheduling



Is This A Real Switch?



Recall What Drives Real World Switches

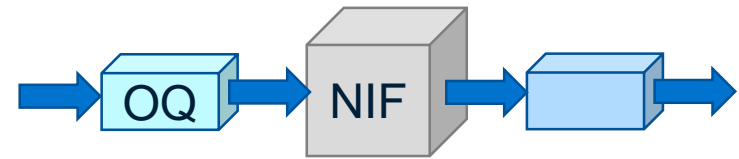
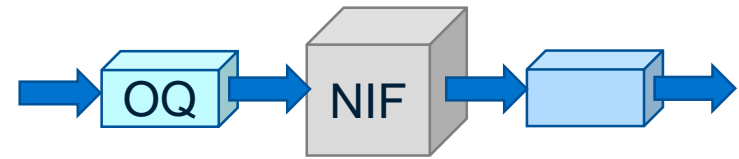
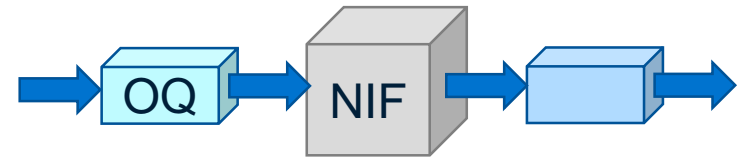
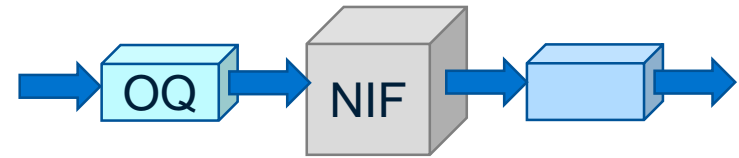
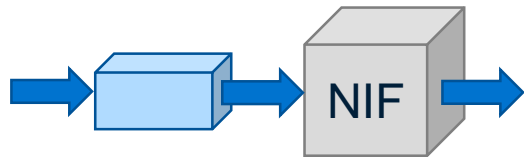
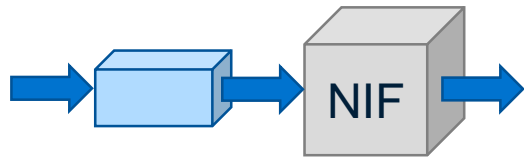
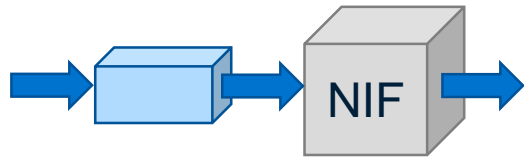
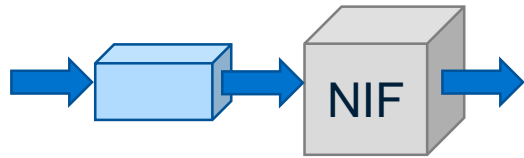
- Cost
- Power
- Area



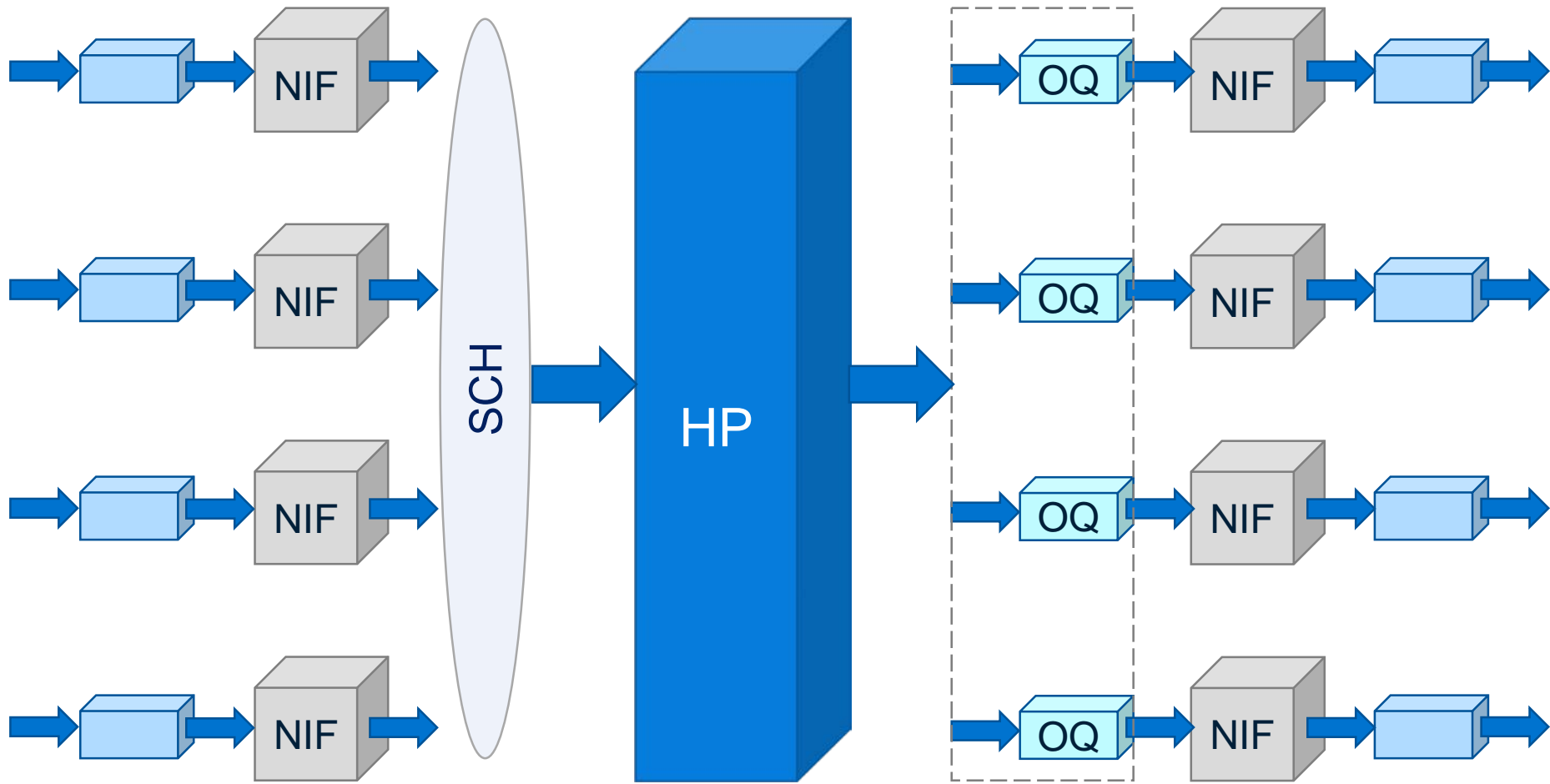
Sharing Resources Is Good!

- Single header processor (if possible)
- Shared memories
- No concurrency problems
 - Also no need to synchronise tables, no need to send updates,

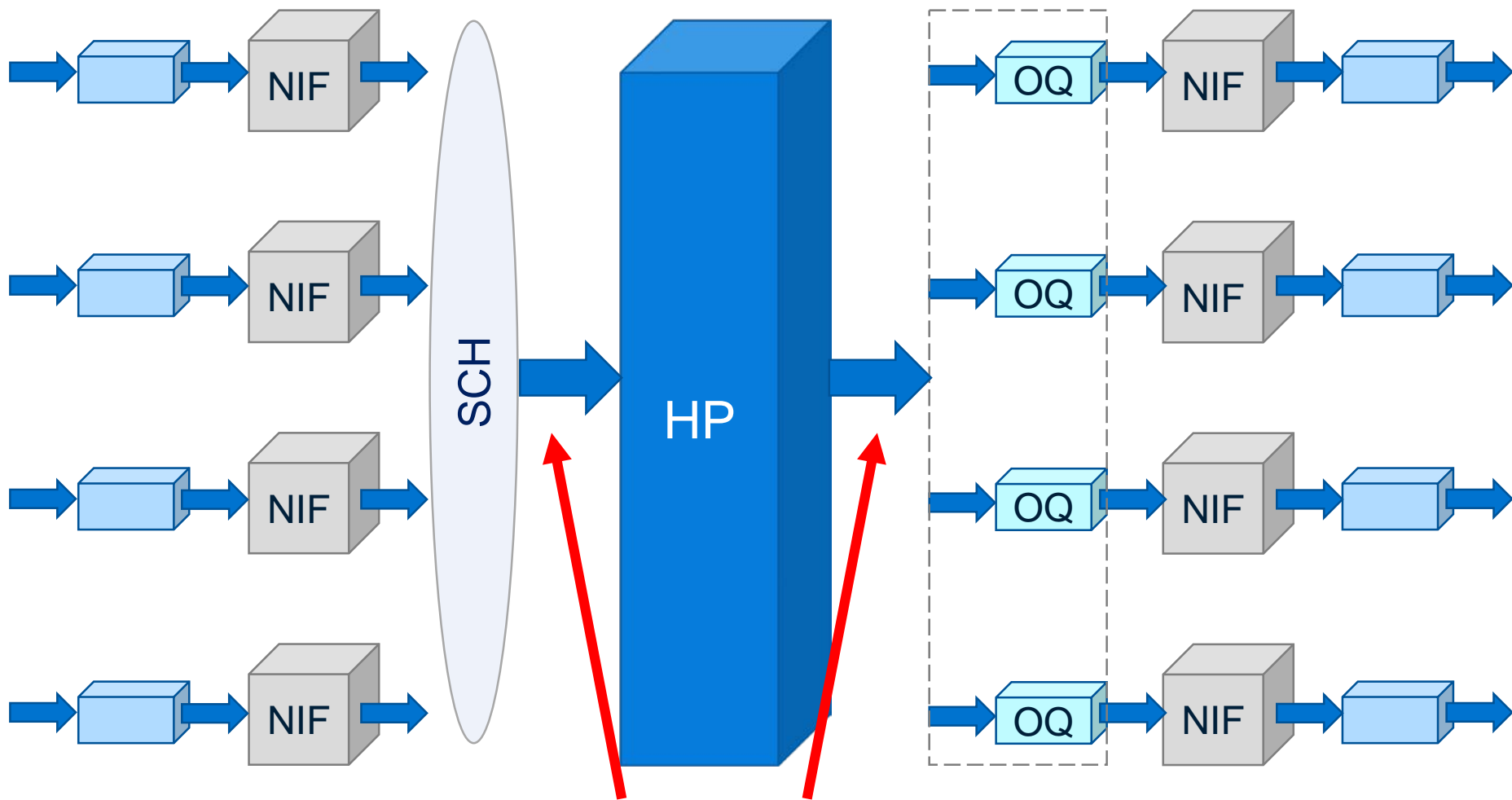
Rethinking The Switch Architecture



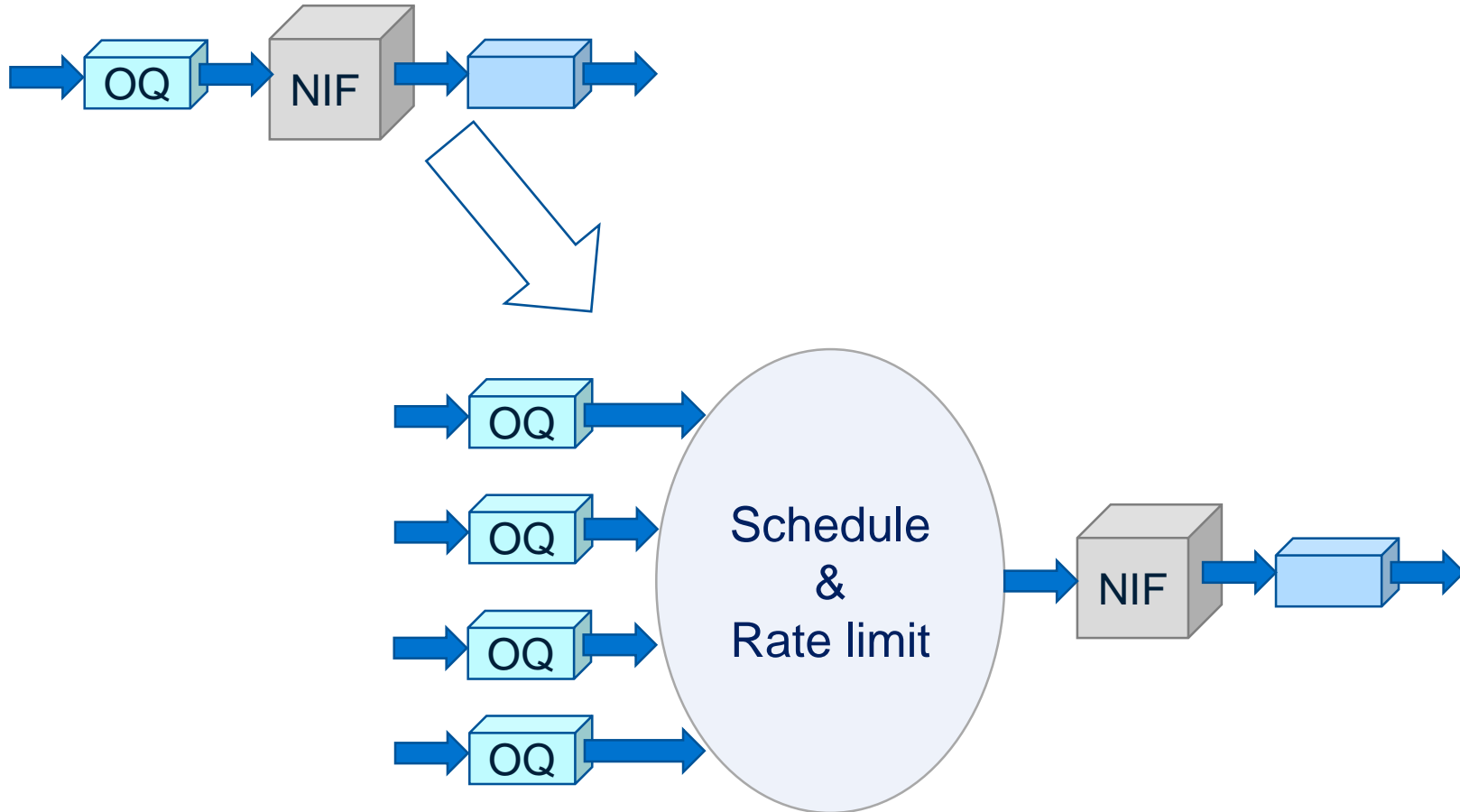
Rethinking The Switch Architecture



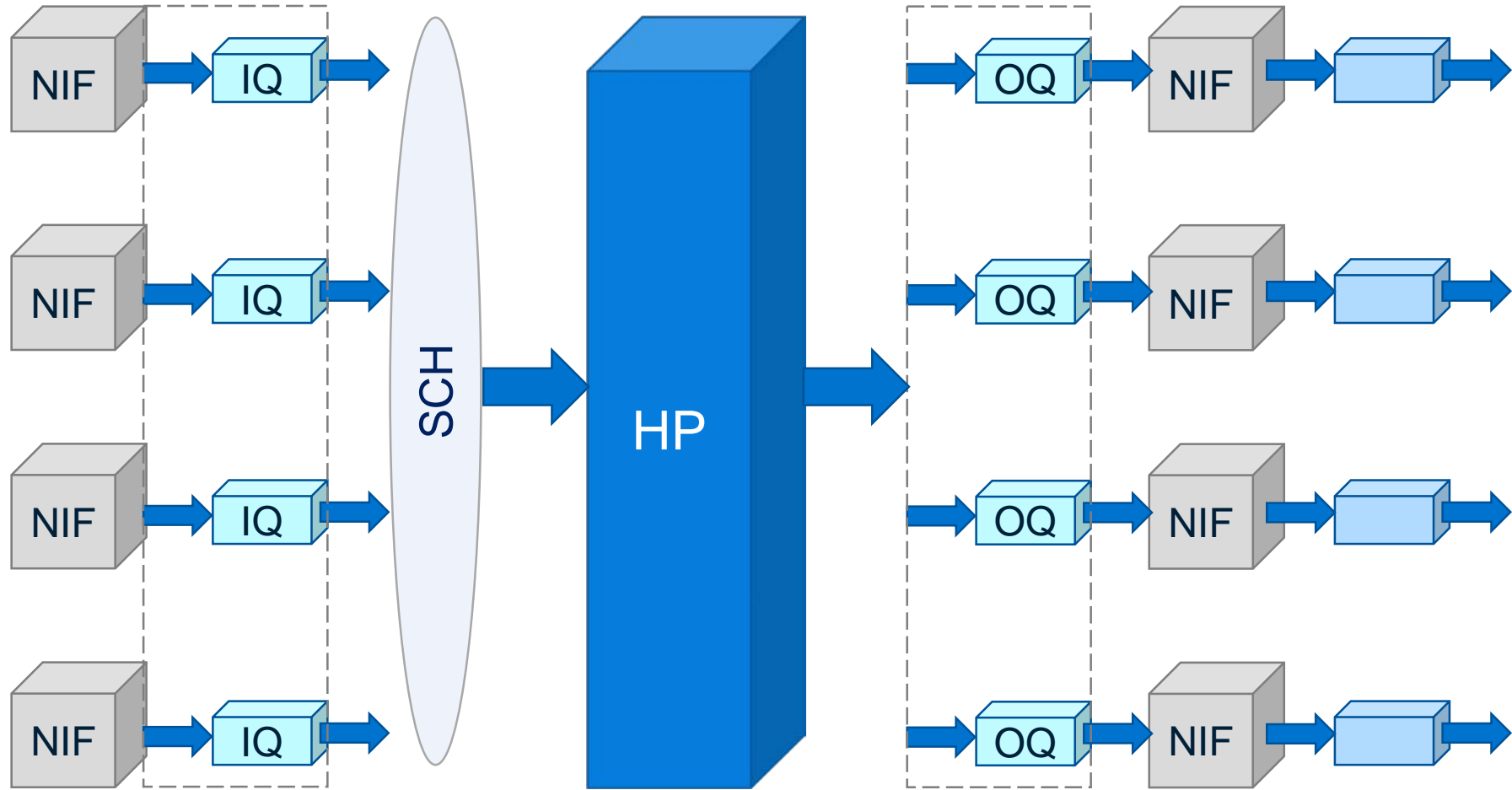
Where Is The Switching?



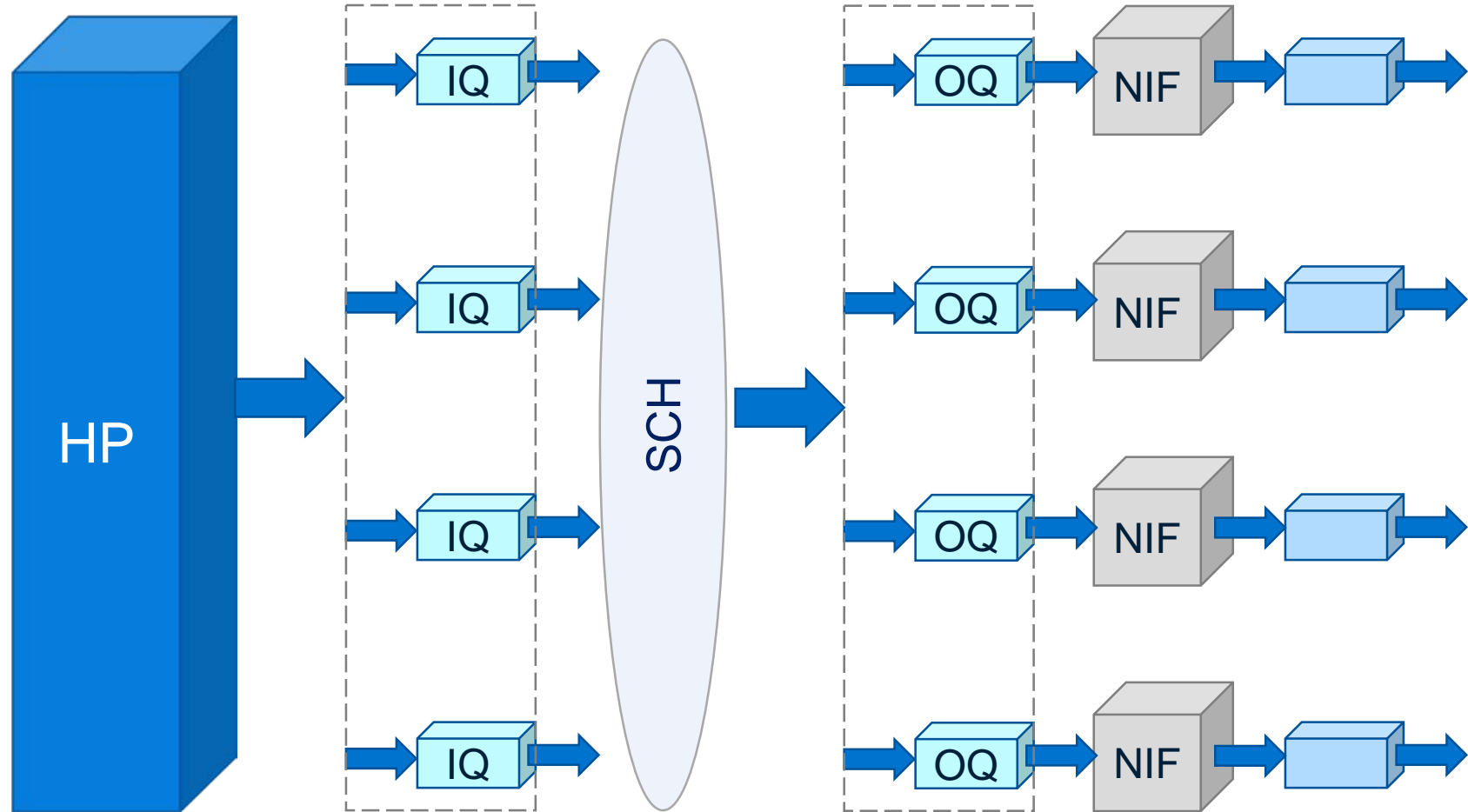
Output Queueing



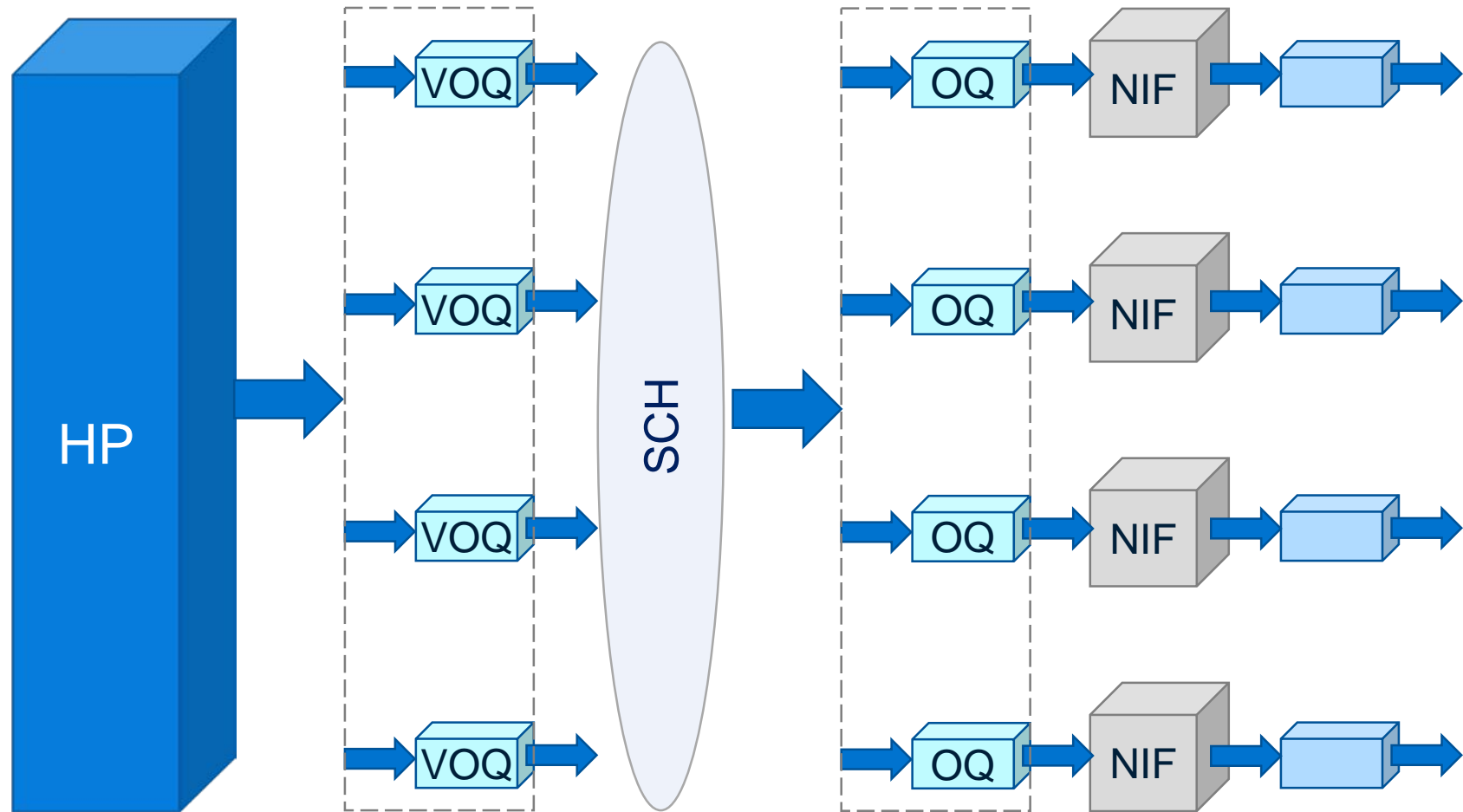
Input Queueing



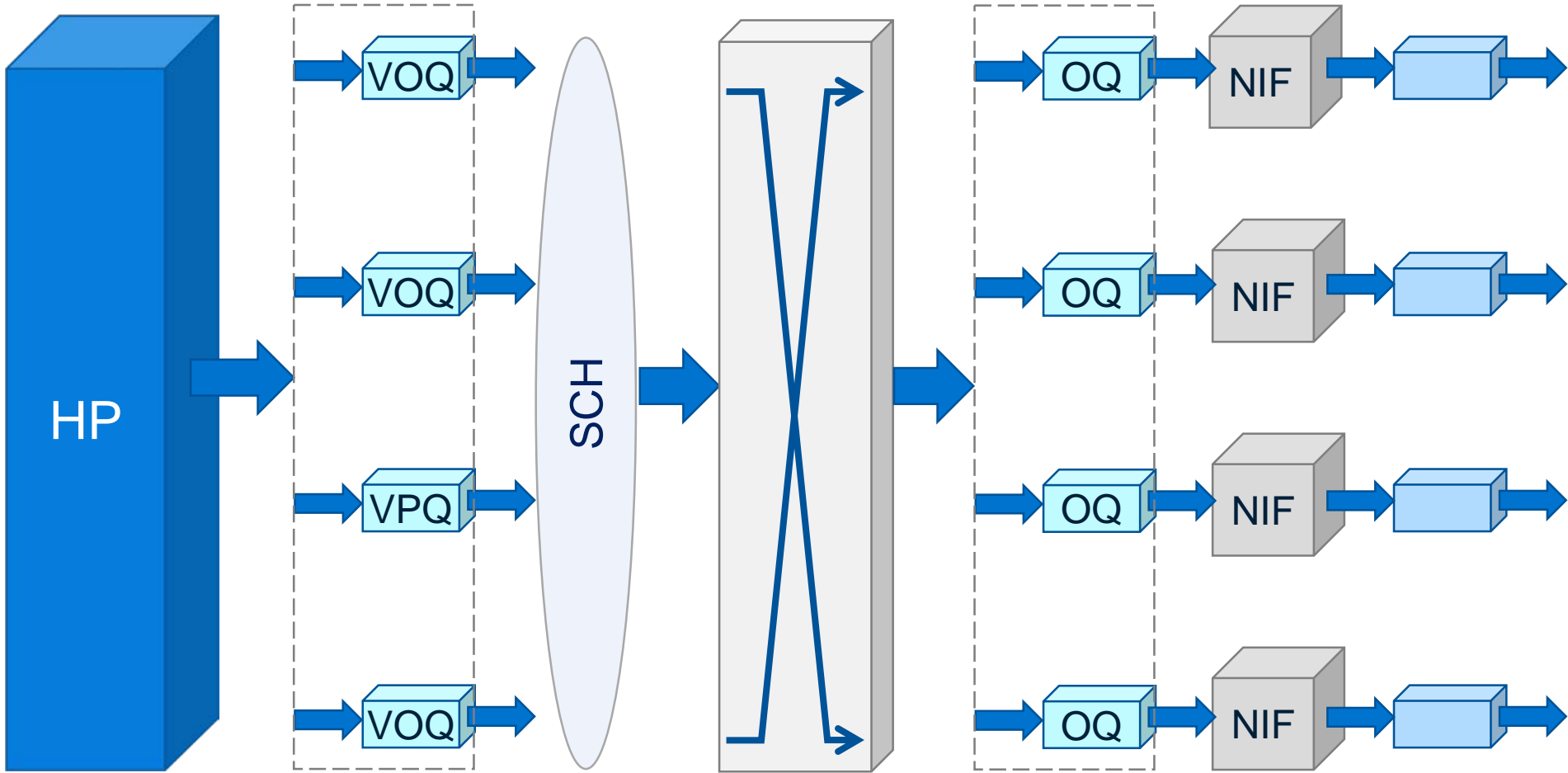
Virtual Output Queueing



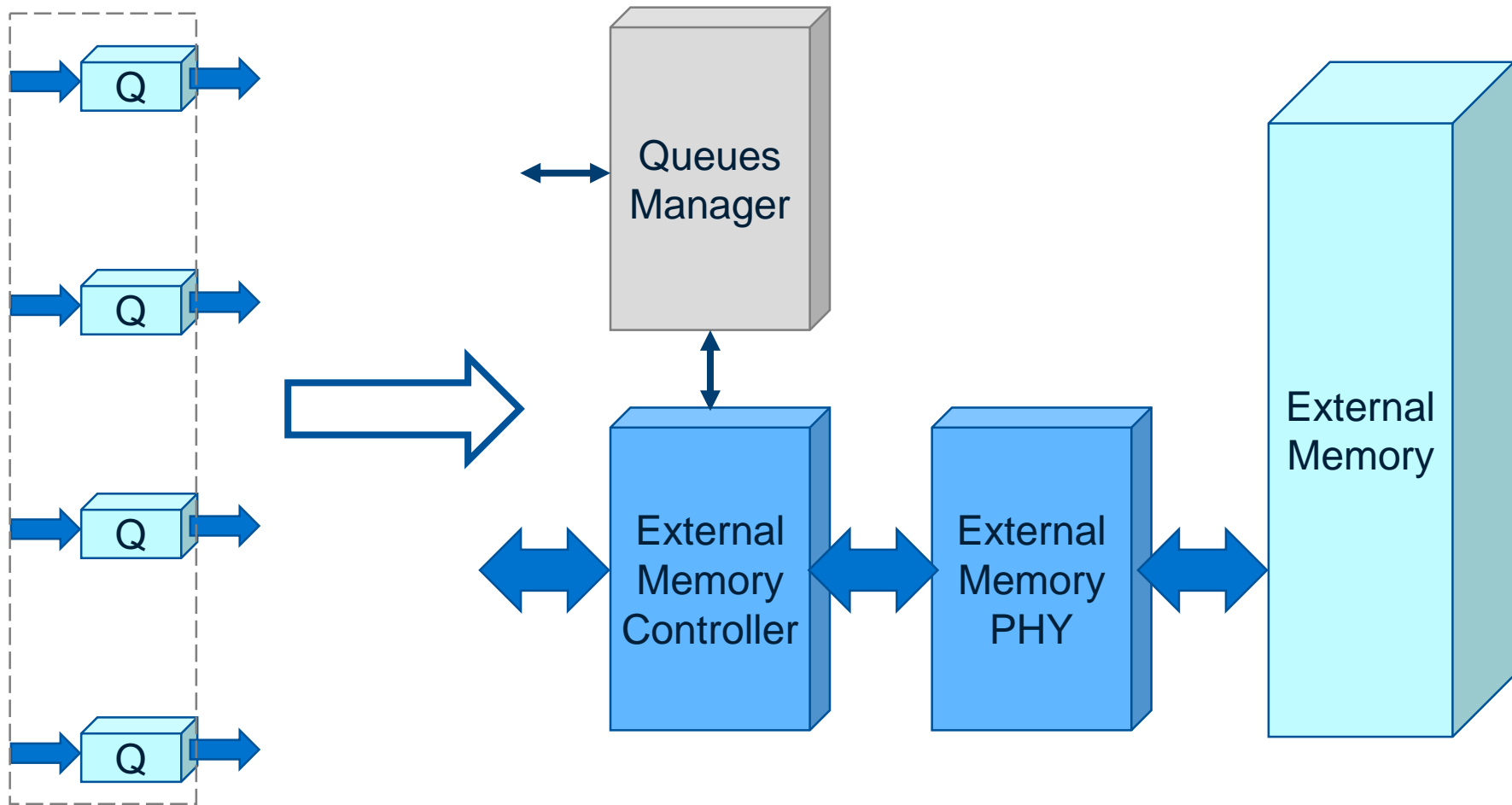
Virtual Output Queueing



Virtual Output Queueing



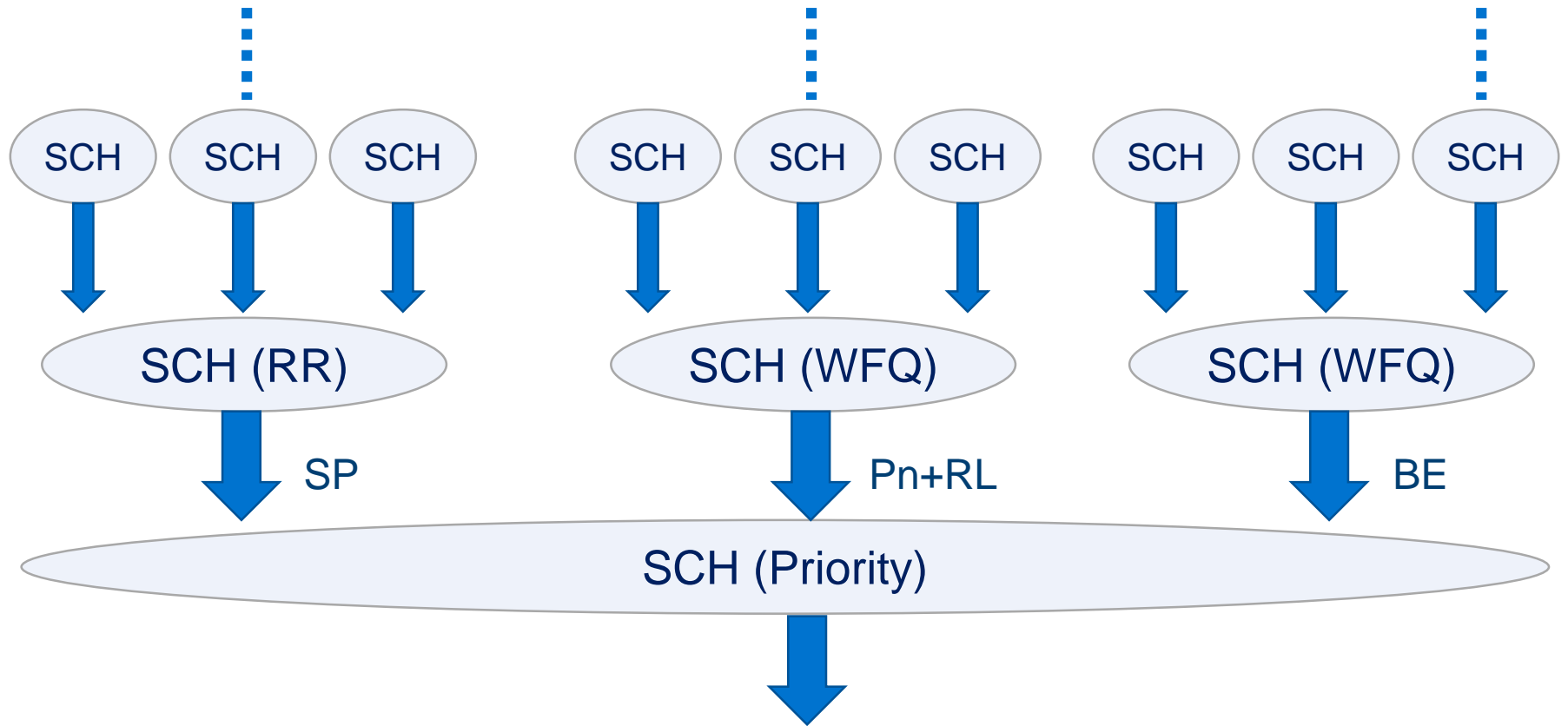
Deep Buffers



Scheduling

- Different operations within the switch:
 - Arbitration
 - Scheduling
 - Rate limiting
 - Shaping
 - Policing
- Many different scheduling algorithms
 - Strict priority, Round robin, weighted round robin, deficit round robin, weighted fair queueing...

Scheduling Hierarchies



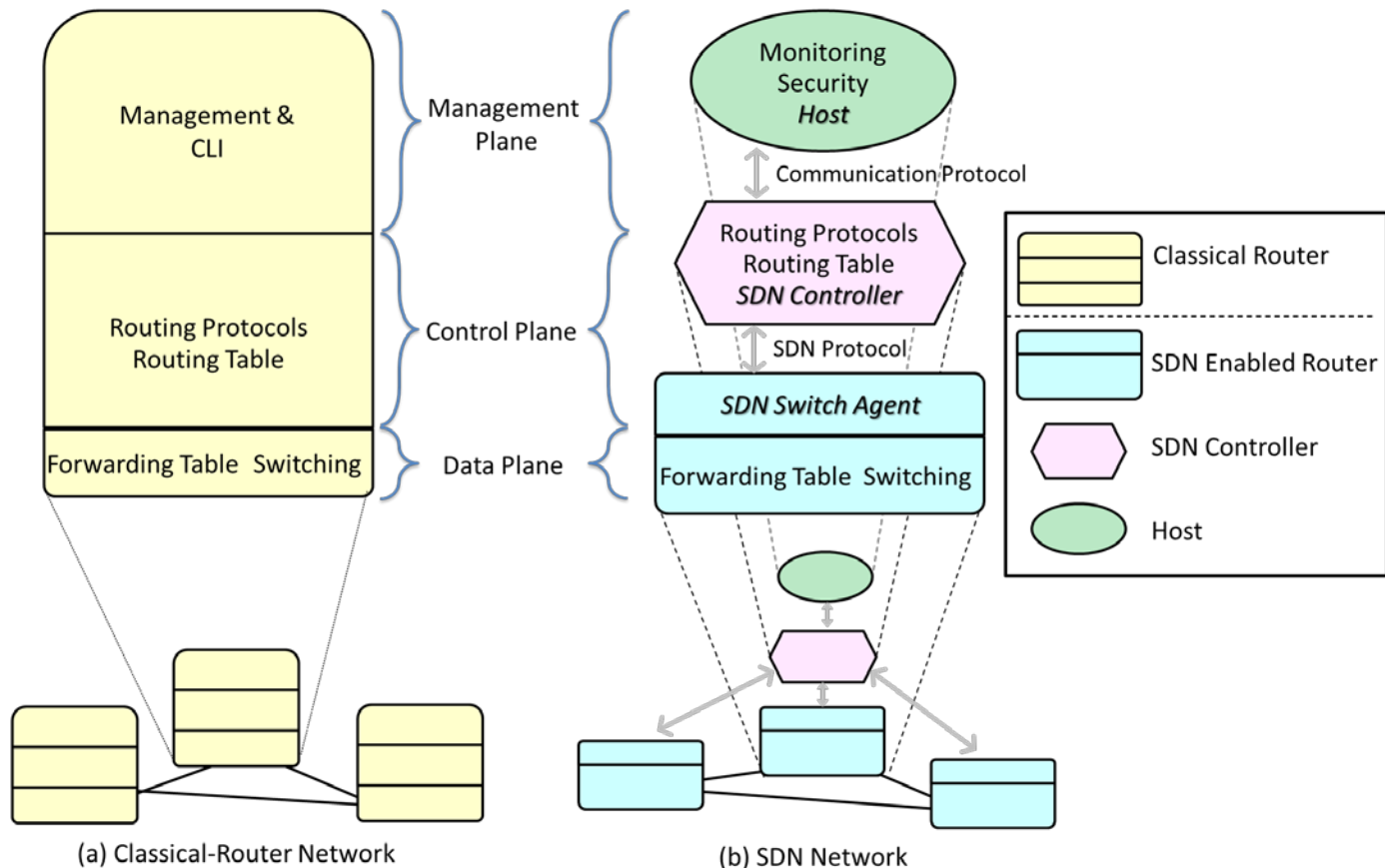
SP – Strict Priority
Pn – Priority <n>

BE – Best Effort
RL – Rate Limiting

WFQ – Weighted Fair Queueing
RR – Round Robin

Software Defined Networking (SDN)

Key Idea: Separation of Data and Control Planes



Switch Architecture and SDN

