Optical Interconnection Networks in Data Centers: Recent Trends and Future Challenges

Speaker: Lin Wang

Research Advisor: Biswanath Mukherjee

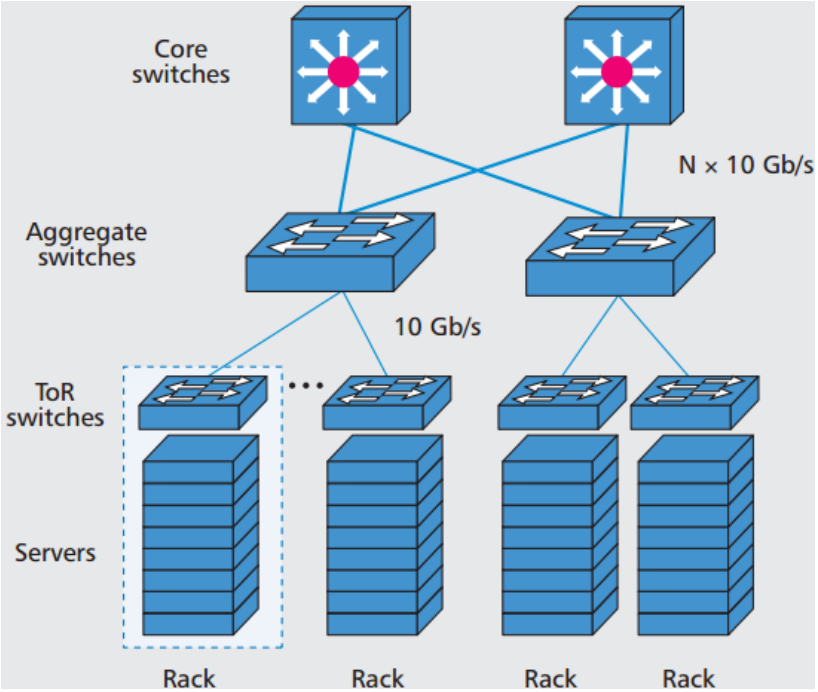
Kachris C, Kanonakis K, Tomkos I. Optical interconnection networks in data centers: Recent trends and future challenges[J]. IEEE Communications Magazine, 2013, 51(9): 39-45.

UCDAVIS

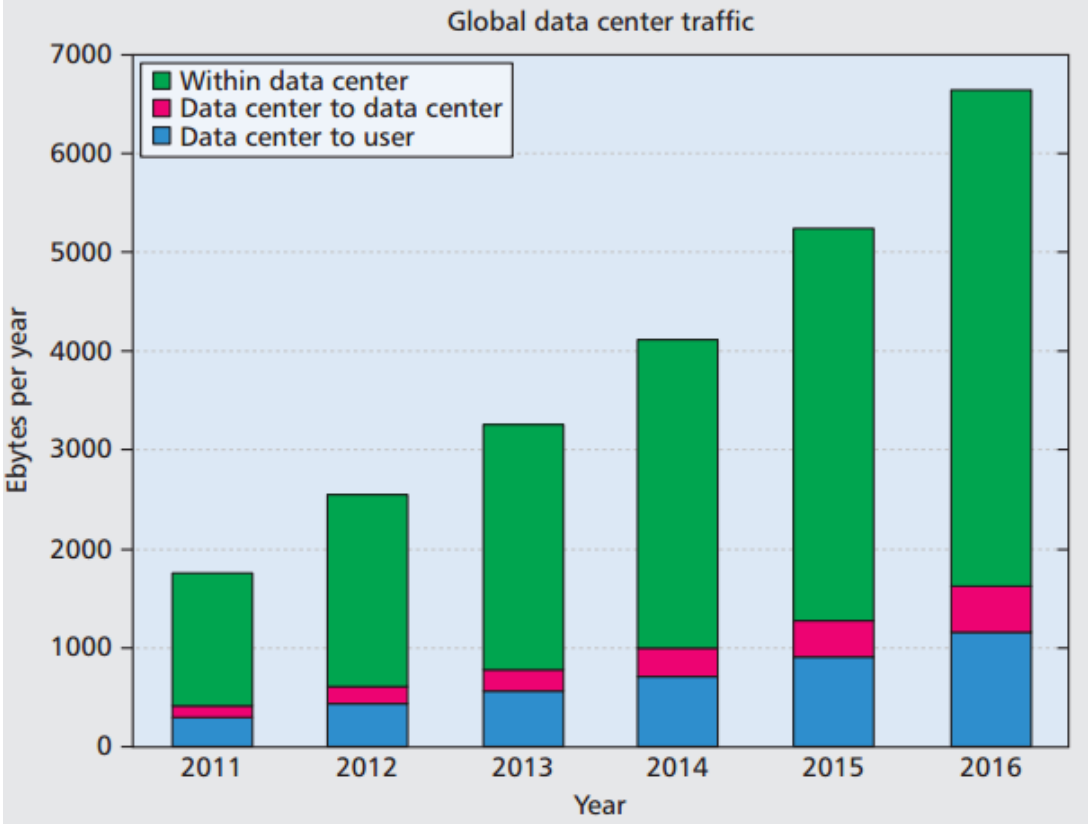
Motivation

- Intra-data center networks (DCNs) are currently based on commodity Ethernet switches connected in a fat-tree topology.

- Optical interconnection networks could be a solution to address limited throughput, high latencies, and high power consumption.



Data Center Traffic Requirements



The majority of network traffic is within data centers: because most applications hosted in data centers are based on parallel programming frameworks such as MapReduce.
More and more processing cores are integrated into a single processor, higher-bandwidth interfaces will be required for the communication of these cores with other cores residing on separate racks.

Power Consumption Requirements

Feature	2012	2016	2020
(Bidi) bandwidth	1 Pbytes/s	20 Pbytes/s	400 Pbytes/s
Overall power consumption	5 MW	10 MW	20 MW
Network power consumption	0.5 MW	1 MW	2 MW optical interconnection networks

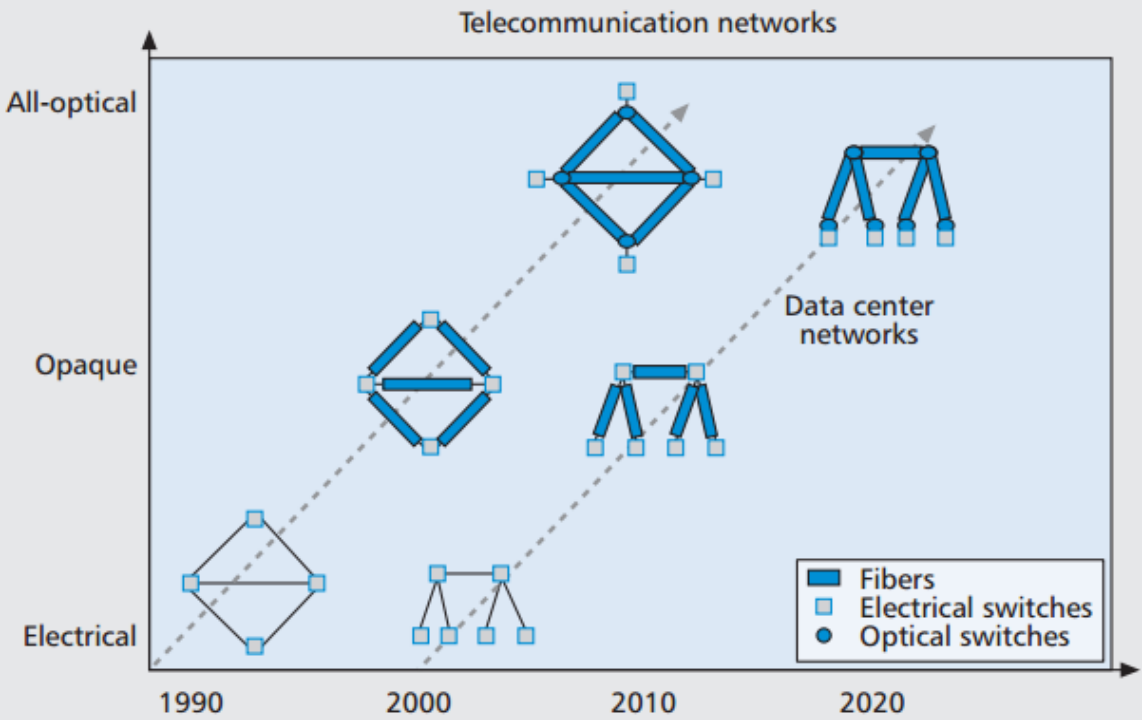
While traffic is expected to increase from 1 to 400 Pbytes/s, the projected allowable power consumption will only increase from 5 to 20 MW.

The affordable DCN power consumption in 2020 could be as low as around 2 MW for 400 Pbytes/s.

State-of-art: optical technology in DCNs has been utilized mainly in the form of point-to point communication between the switches using optical fibers and optical transceivers.

Drawbacks:
1. power is wasted in the electrical to optical (E/O) and optical to electrical (O/E) conversions;
2. Latency is further aggravated due to electrical buffering to address contention.

Migration from electrical to all-optical networks



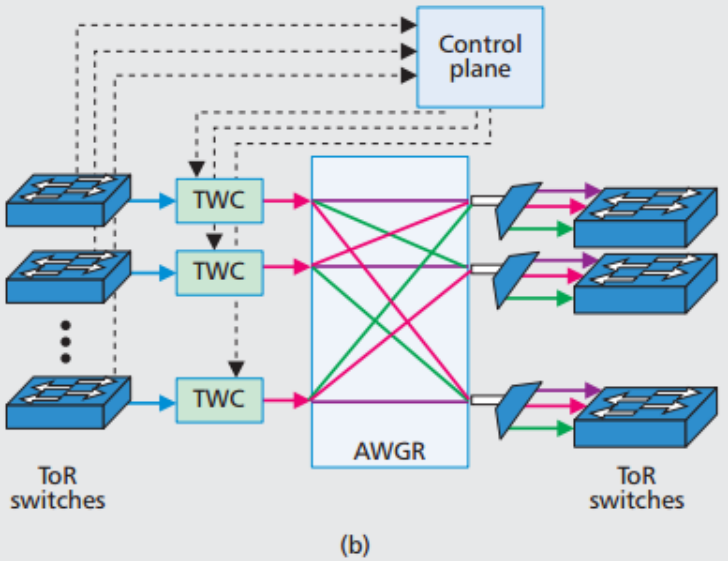
At the telecommunications networks side, opaque networks based on point-to-point optical fibers have recently been replaced by all-optical (transparent) networks.

Similar paradigm shift should be expected to take place in the case of future intra-DCNs

Circuit VS. Packet-switching for DCN

Circuit-based schemes: long-term bulky data transfers are required between racks.
Advantages: they are usually based on micro electro-mechanical switches (MEMS), which have increased reconfiguration time (in the orders of few milliseconds).
Weakness: in order to guarantee all-to-all communication, significant over-provisioning might be required.

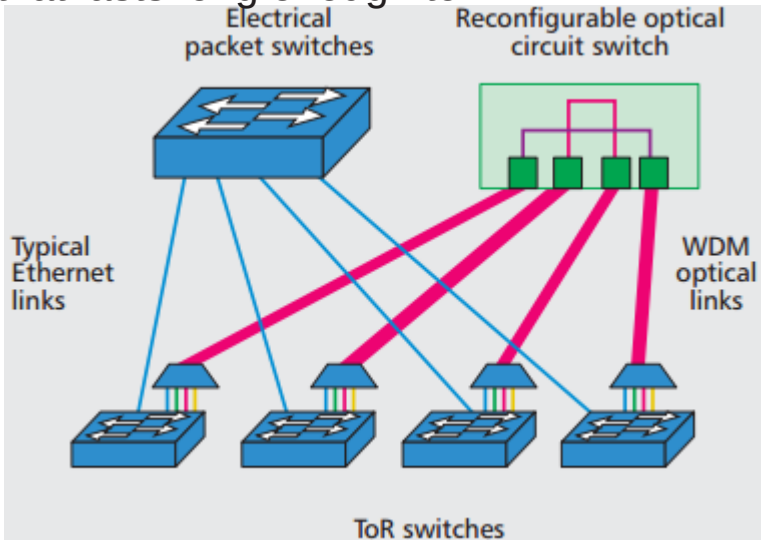
Packet-based schemes: burstier traffic and all-to-all connectivity.
Advantages: Packet-based switching assumes either an array of fixed lasers or fast tunable transmitters for addressing a specific destination port by selecting the appropriate wavelength.
Tunable wavelength converters (TWCs) and transceivers can be configured much faster than optical MEMS



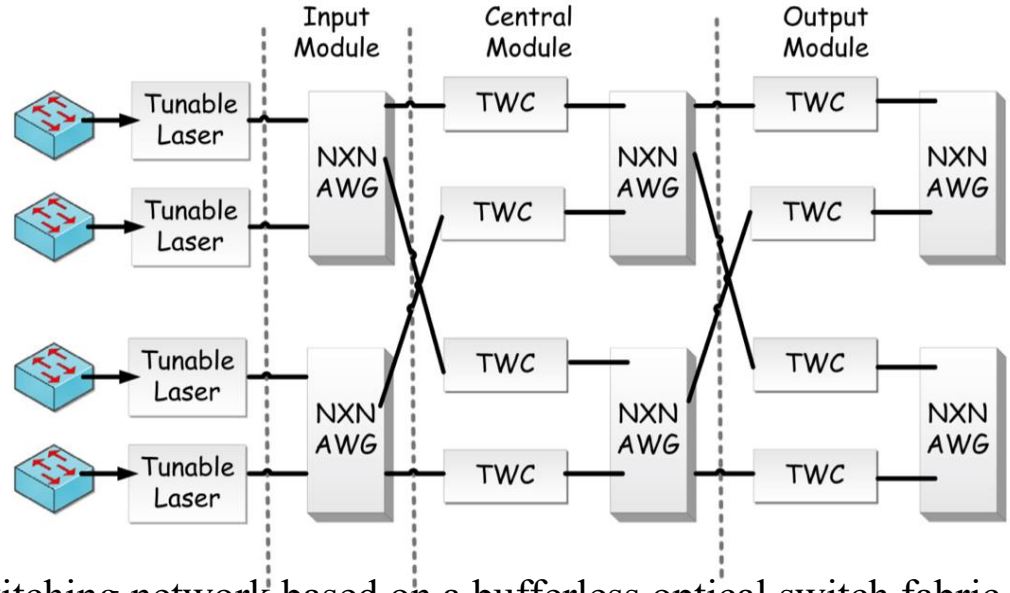
Hybrid VS. All-optical architectures in DCN

Complete replacement of commodity electrical switches means high capital expenditure (CAPEX).
Hybrid schemes offer the advantage of an incremental upgrade of an operating data center with commodity switches, thus reducing the associated CAPEX.

For example:
ToR switches can be enhanced by adding optical modules, which will increase bandwidth and reduce latency.
A portion of the traffic demands consists of traffic that lasts long enough to compensate for the reconfiguration overhead



Recent Optical Interconnection Architectures



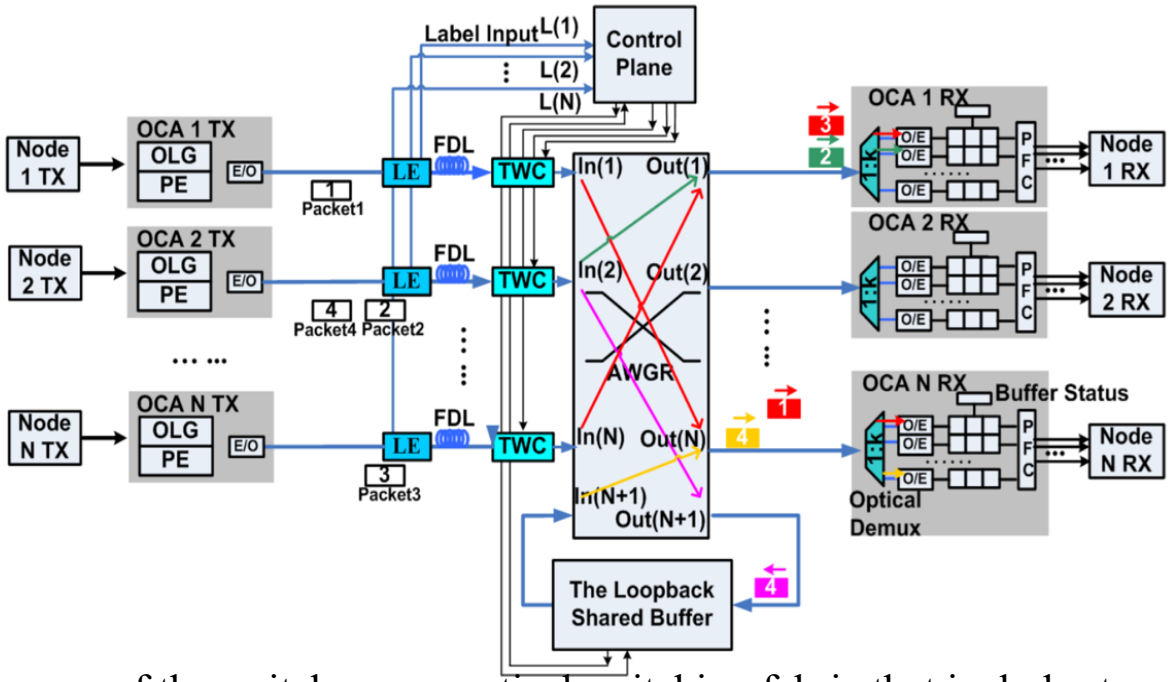
Petabit: full optical switching network based on a bufferless optical switch fabric.

Advantages:

1. Overcome the issues with oversubscription, bottlenecks, latency, wiring complexity and high power consumption.
2. Flattened the network by designing one switch that is capable of connecting all racks within the data center.

Architecture: three-stage network fabric with Input Modules (IMs), Central Modules (CMs) and Output Modules (OMs), where each module has an AWGR.

Recent Optical Interconnection Architectures

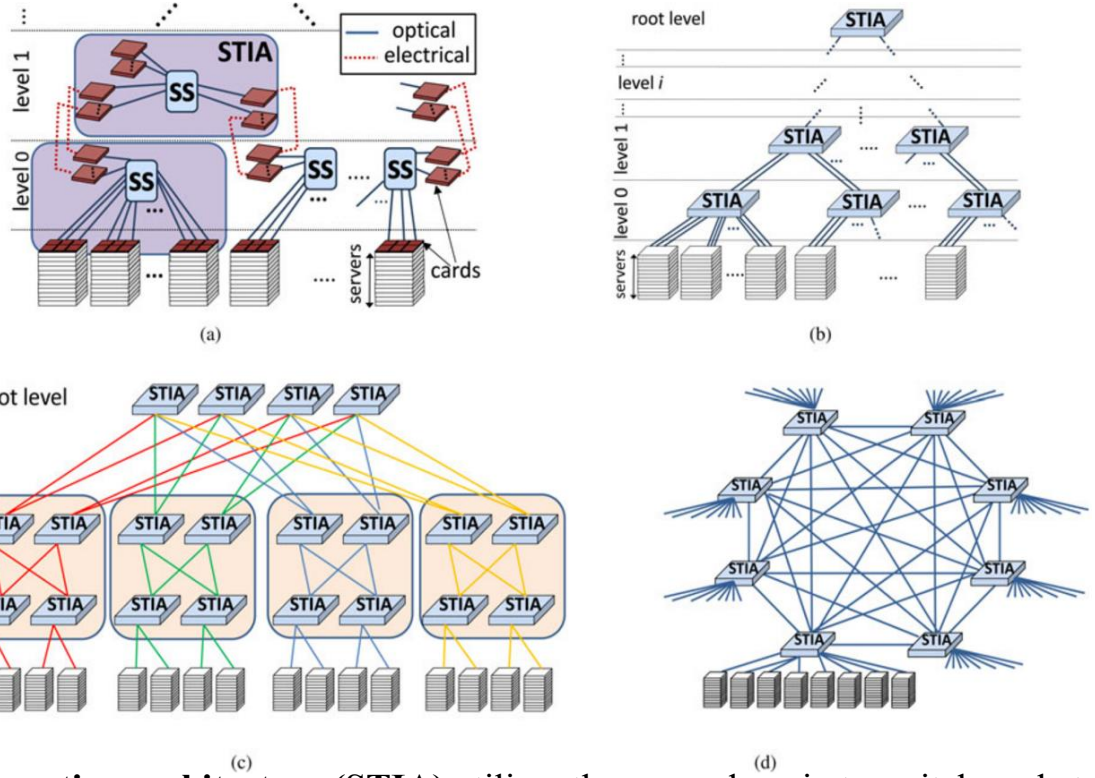


DOS contains a core of the switch are an optical switching fabric that includes tunable wavelength converters (TWC), a uniform loss and cyclic frequency (ULCF) AWGR, and a loopback shared buffer system. a control block that processes the label of the packet and then arbitrates each packet by checking resource availability on the output port side.

Advantages:

Latency is almost independent of the number of input ports and remains low even at high input loads. Because packets have to traverse only one optical switch and thus avoid the delay of the electrical switch's buffers

Recent Optical Interconnection Architectures



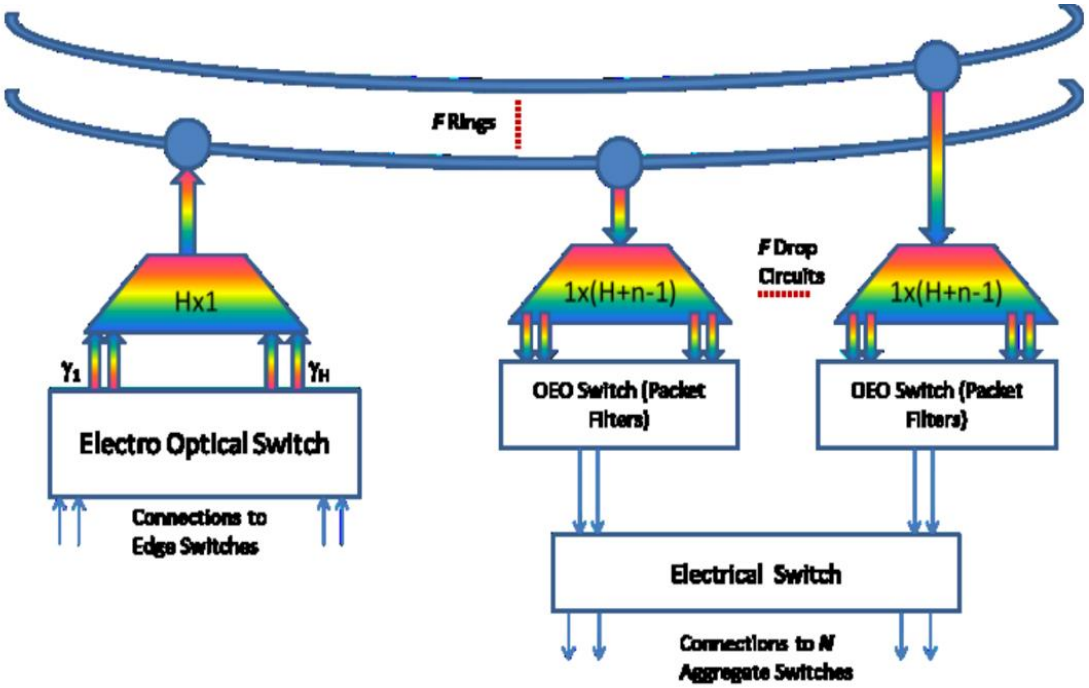
Space-time interconnection architecture (STIA) utilizes the space domain to switch packets and the time domain to switch packets to different nodes.

A space switch is used as a central node for the switching of packets.

A wave- length domain is used in order to increase the throughput by encoding packets on multiple wavelengths.

Advantages: efficient scalability using folded Clos or flattened butterfly topologies and can provide more than an order of a magnitude lower power consumption compared to Ethernet-based networks.

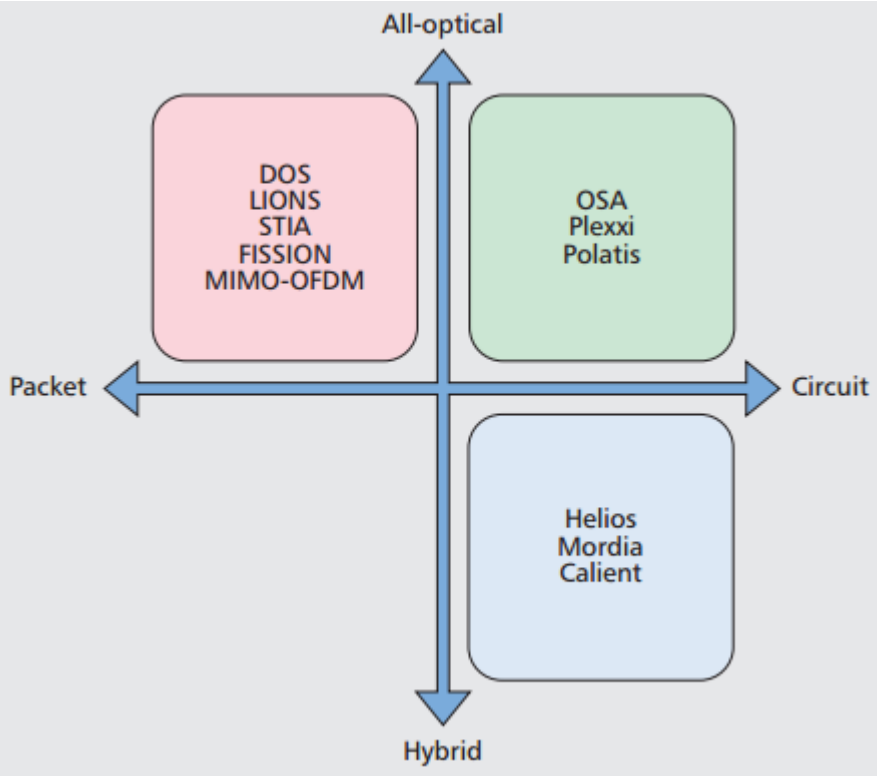
Recent Optical Interconnection Architectures



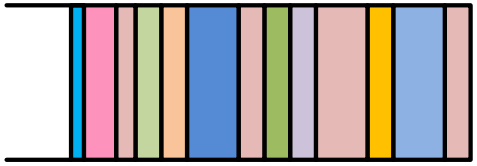
Fission DCN architecture comprises fiber rings that are used to interconnect the core data center switches by deploying ultra-dense WDM (UDWDM) technology.
Each of the nodes (called electrical-optical switches, EOSs) connected to the fiber ring utilizes a WSS that is used to add and/or drop wavelengths into and from the fiber ring, respectively.
Advantages: it can be scaled efficiently by sup- porting several rings depending on the number of nodes and the bandwidth requirements.

Classification: qualitative comparison

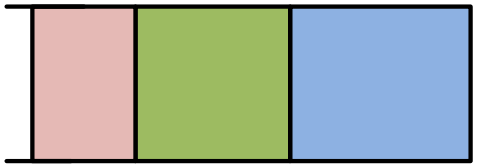
- Circuit-based schemes: the hosted applications usually require long lived traffic flows that transfer big chunks of data to other nodes.
- Packet-based circuits schemes: provide all-to-all communication, and due to the fast switching time, they can support both long-lived and short-lived traffic flow.

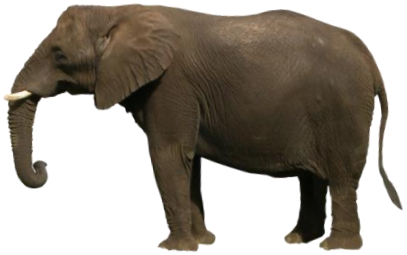


Mice flow VS elephant flow



- Small size packet
- Short flow
- Large number
- Short-lived

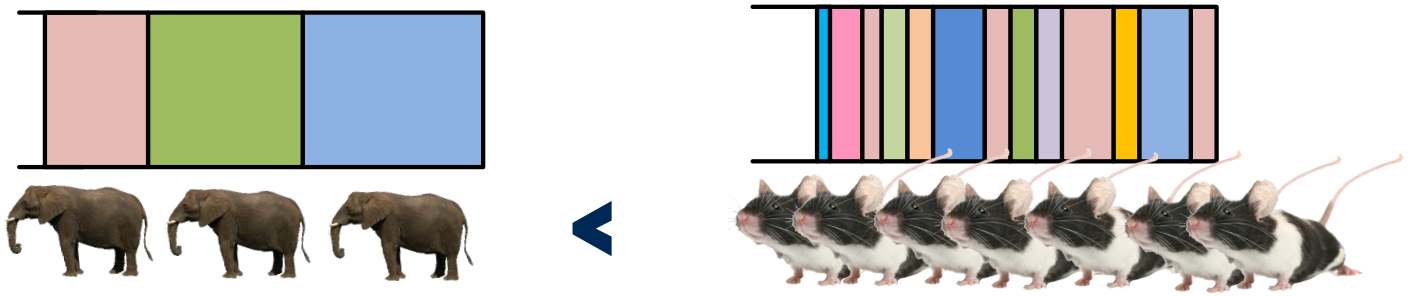




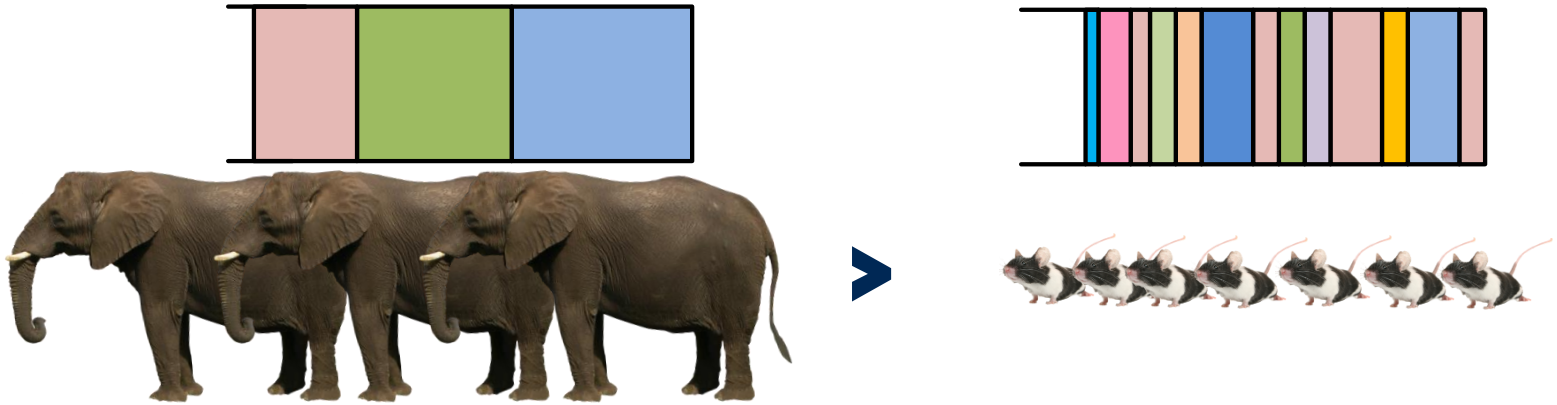
- Large size packet
- Large volume flow
- Small number
- Long lasting

Mice flow VS elephant flow

- If we only care about the number of packets in the queue, elephant flow transmission is easy to be degraded.



- If we only care about the total size of packets in the queue, mice flow transmission is easy to be degraded.

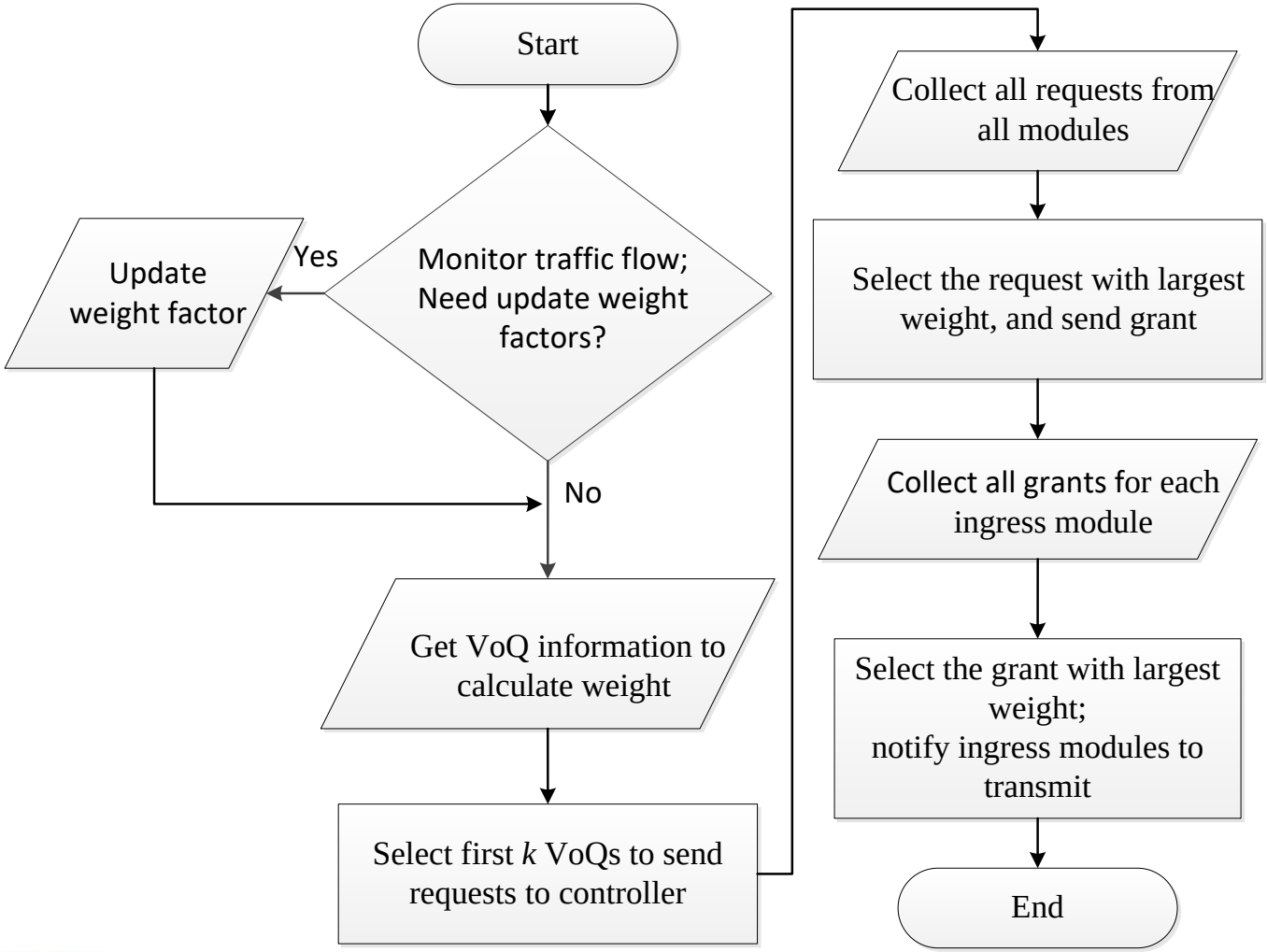


Our work

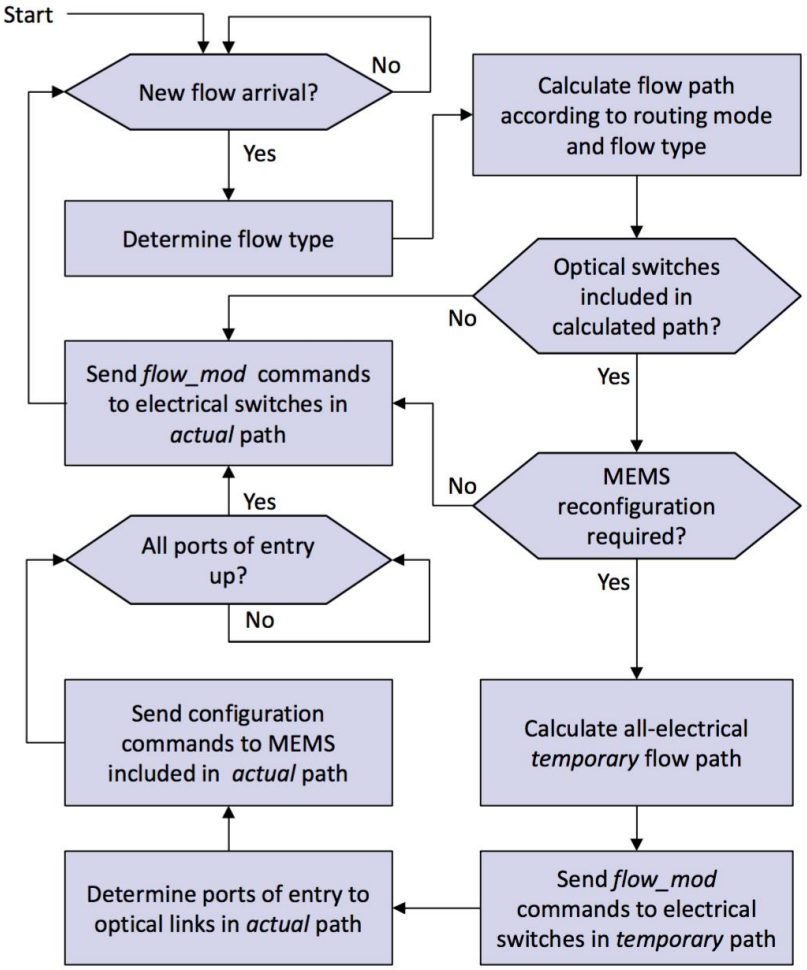
- **Priority-aware scheduling algorithm.**
  - We can update the weight factors based on the classification of traffic flow.

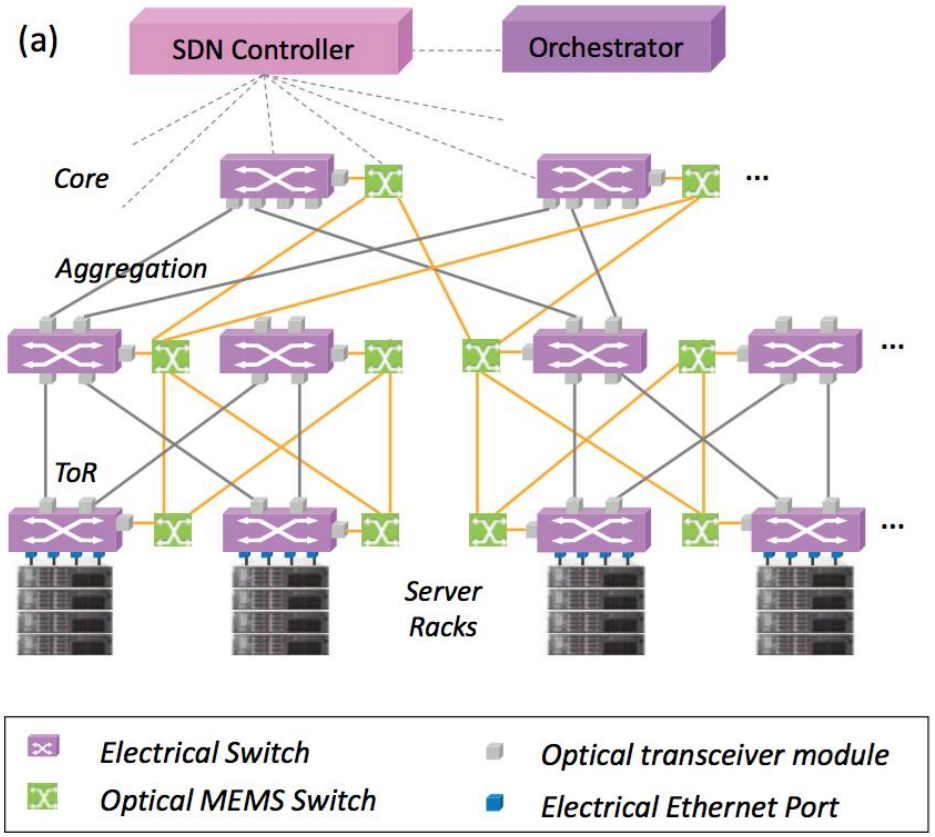
- **Hybrid Optical/Electrical DCN**
  - We can study the migration from traditional electrical DCN to all optical DCN. This is highly related with the demands of traffic flow.
  - Note: only flows which last long enough and can consume the optical capacity (e.g. “elephant” flows) should be offloaded to the optical domain, while bursty, low-bitrate “mouse” flows should use the electrical part of the topology.

Priority-aware Scheduling Algorithm for PSON



Hybrid Optical/Electrical DCN





Kanonakis K, Yin Y, Ji P N, et al. SDN-controlled routing of elephants and mice over a hybrid optical/electrical DCN testbed[C]//Optical Fiber Communications Conference and Exhibition (OFC), 2015. IEEE, 2015: 1-3.

UCDAVIS
UNIVERSITY OF CALIFORNIA

amlwang@ucdavis.edu