



Enabling Topological Flexibility for Data Centers Using OmniSwitch

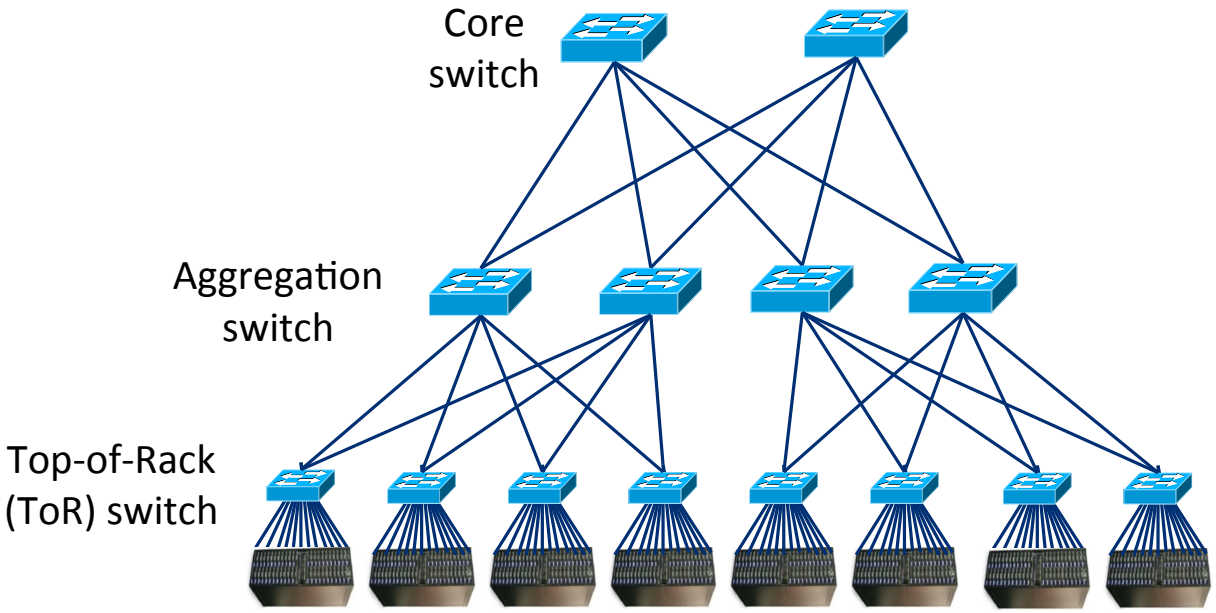
Yiting Xia[§], Mike Schlansker[*], T. S. Eugene Ng[§], Jean Tourrilhes[*]

Rice University[§], HP Labs[*]

HotCloud 2015

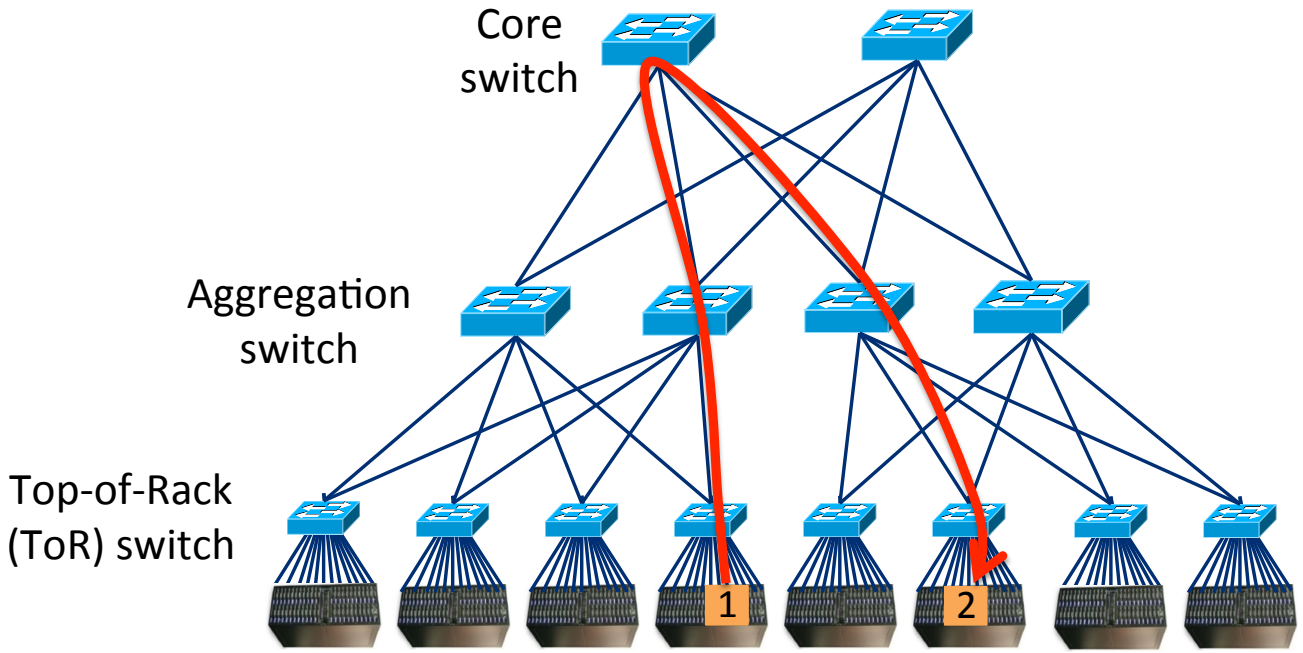
Traditional data centers

- Fixed network topology
- Multi-rooted tree structure of Ethernet switches
- Bandwidth oversubscription
- Optical fibers and transceivers for fast transmission



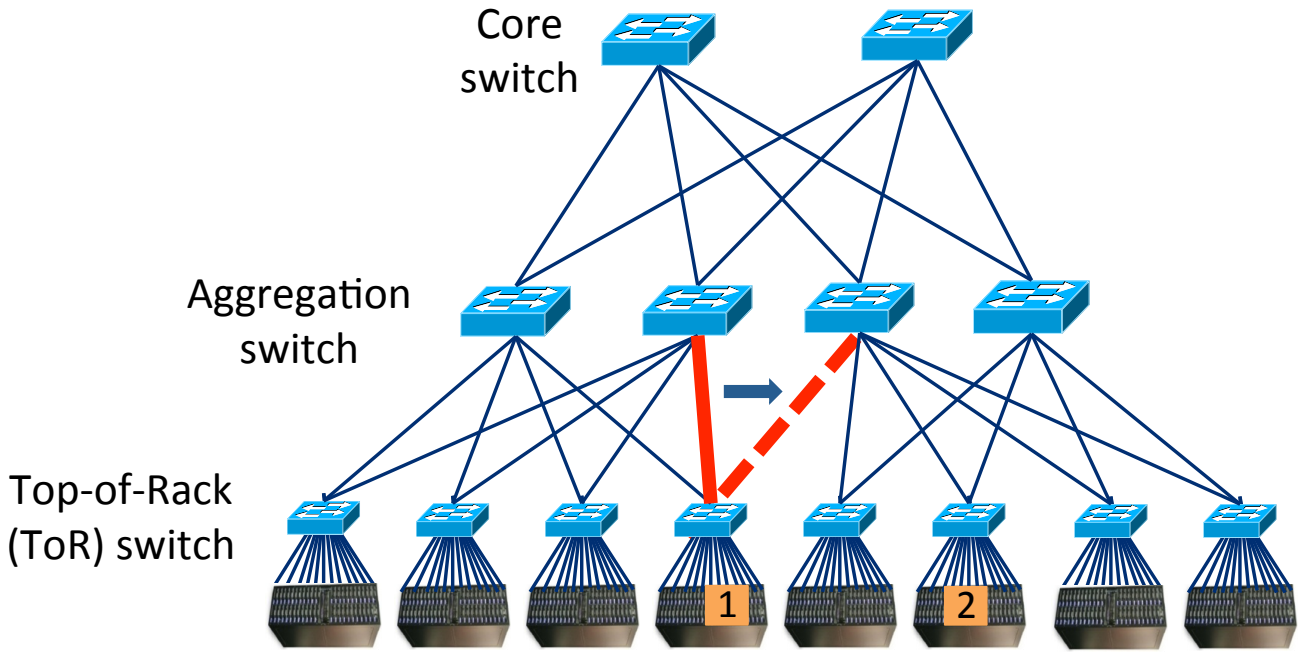
Problems of fixed topology

- Limited traffic optimization



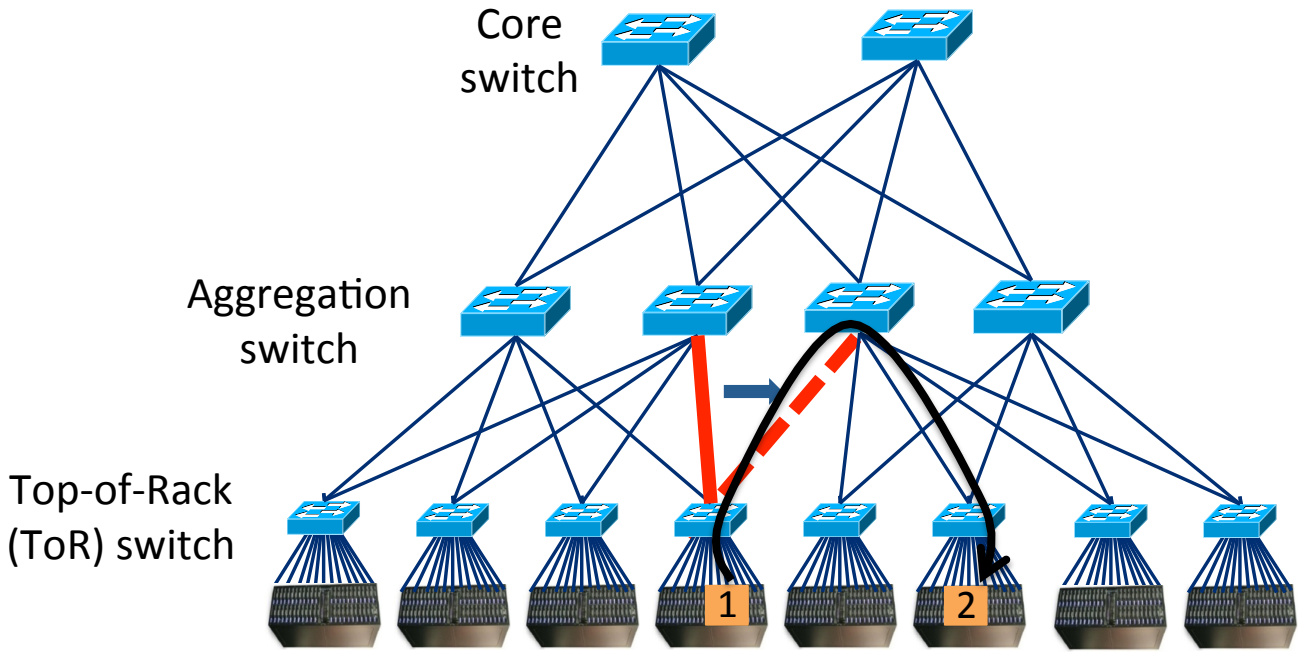
Problems of fixed topology

- Limited traffic optimization



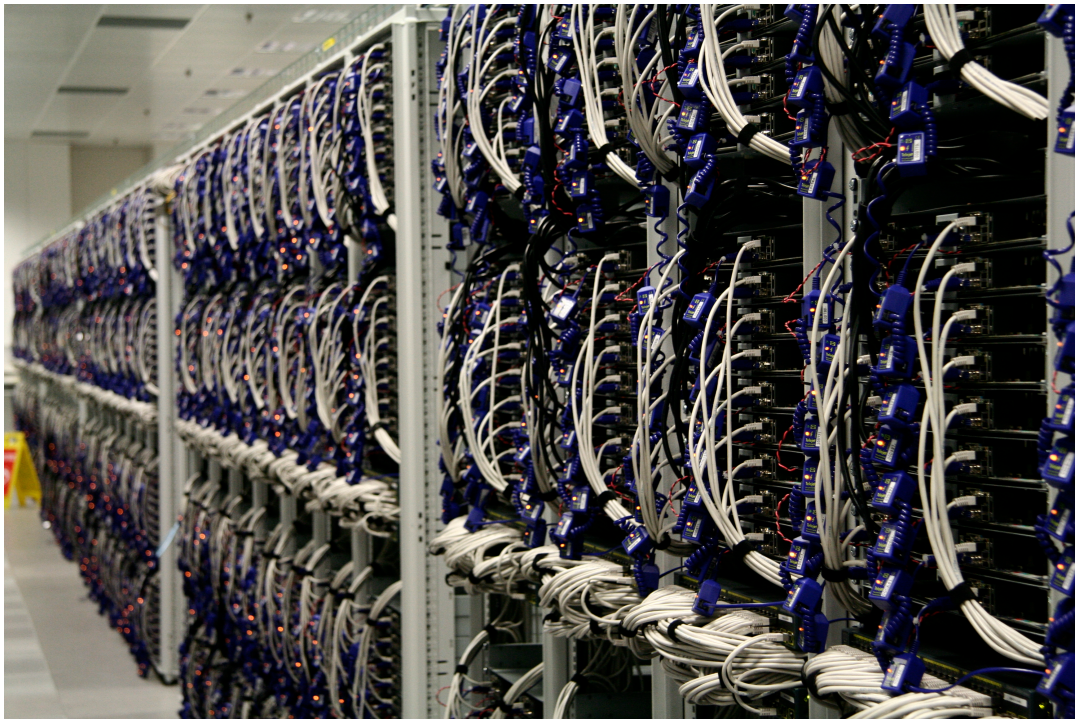
Problems of fixed topology

- Limited traffic optimization



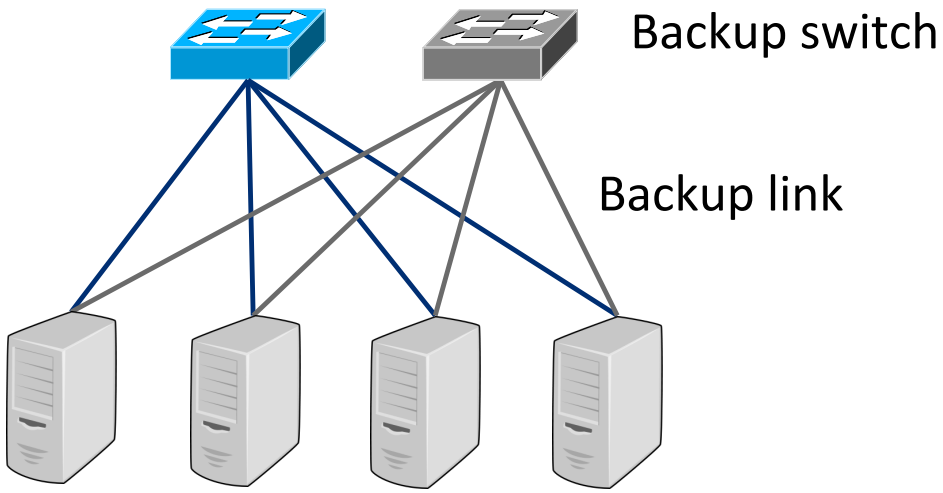
Problems of fixed topology

- Limited traffic optimization
- Complex wiring
- Poor expansion



Problems of fixed topology

- Limited traffic optimization
- Complex wiring
- Poor expansion
- High redundancy cost



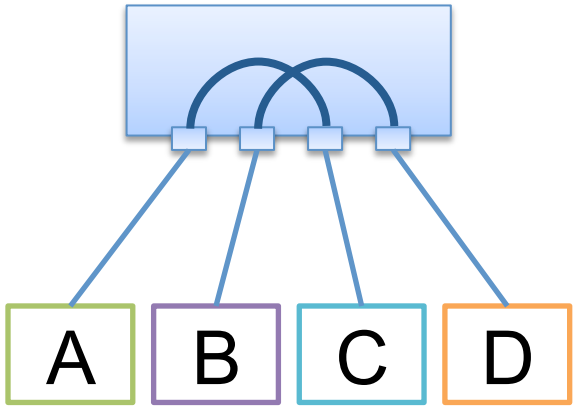
Solution: topological flexibility

- Dynamic cable rewiring
- Can be achieved using circuit switch

Solution: topological flexibility

- Dynamic cable rewiring
- Can be achieved using circuit switch

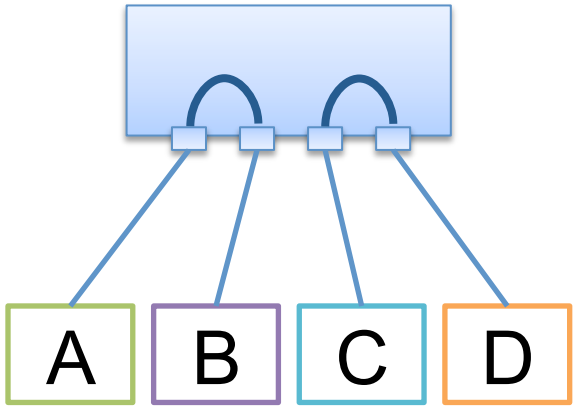
- Circuit switch
  - *dedicated links*
  - *many possible topologies*
  - *configured by software*



Solution: topological flexibility

- Dynamic cable rewiring
- Can be achieved using circuit switch

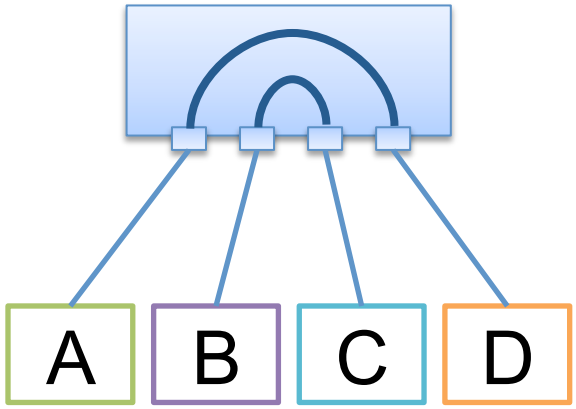
- Circuit switch
  - *dedicated links*
  - *many possible topologies*
  - *configured by software*



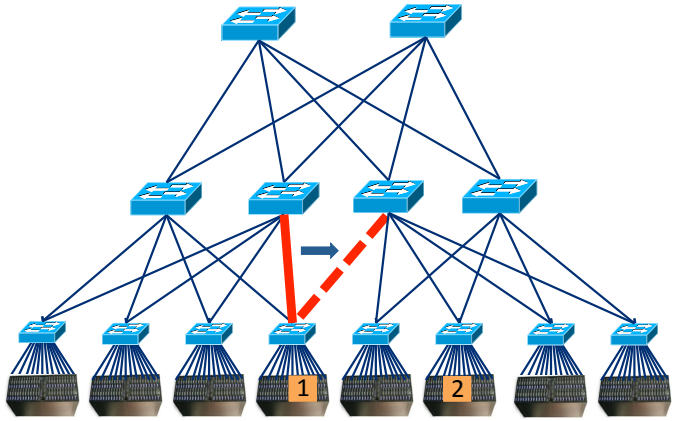
Solution: topological flexibility

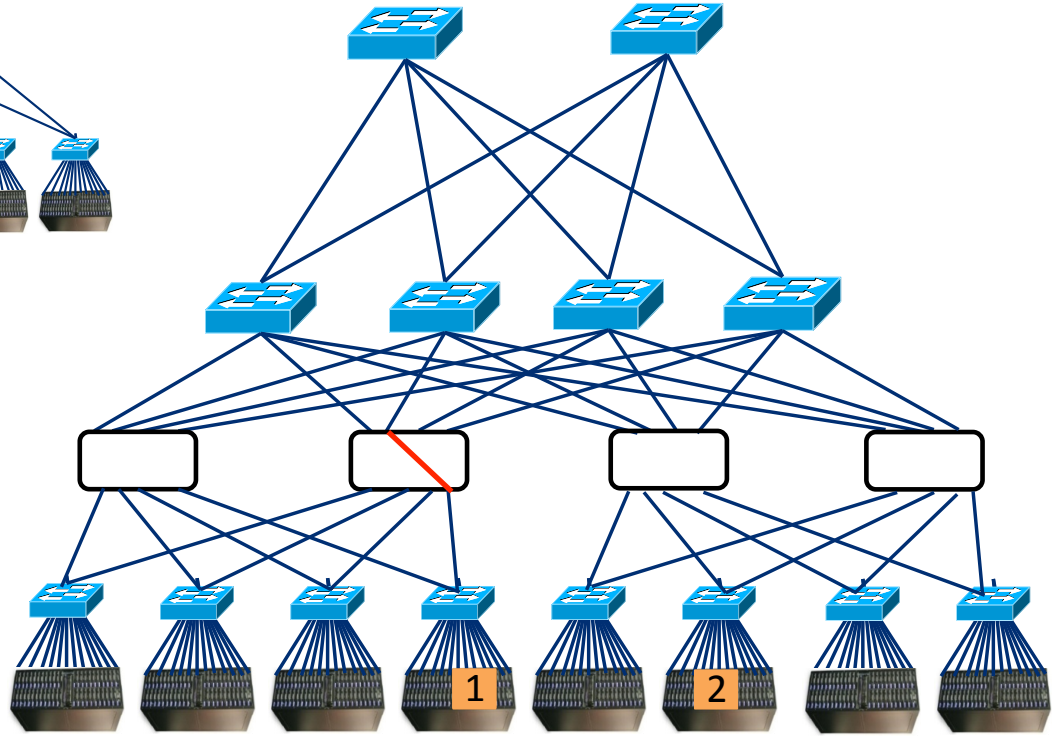
- Dynamic cable rewiring
- Can be achieved using circuit switch

- Circuit switch
  - *dedicated links*
  - *many possible topologies*
  - *configured by software*

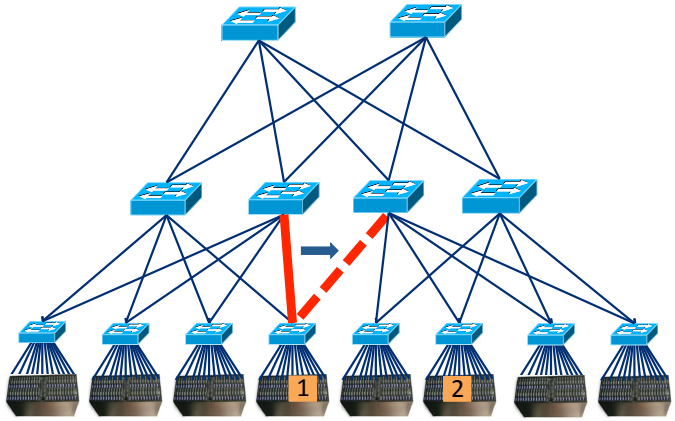


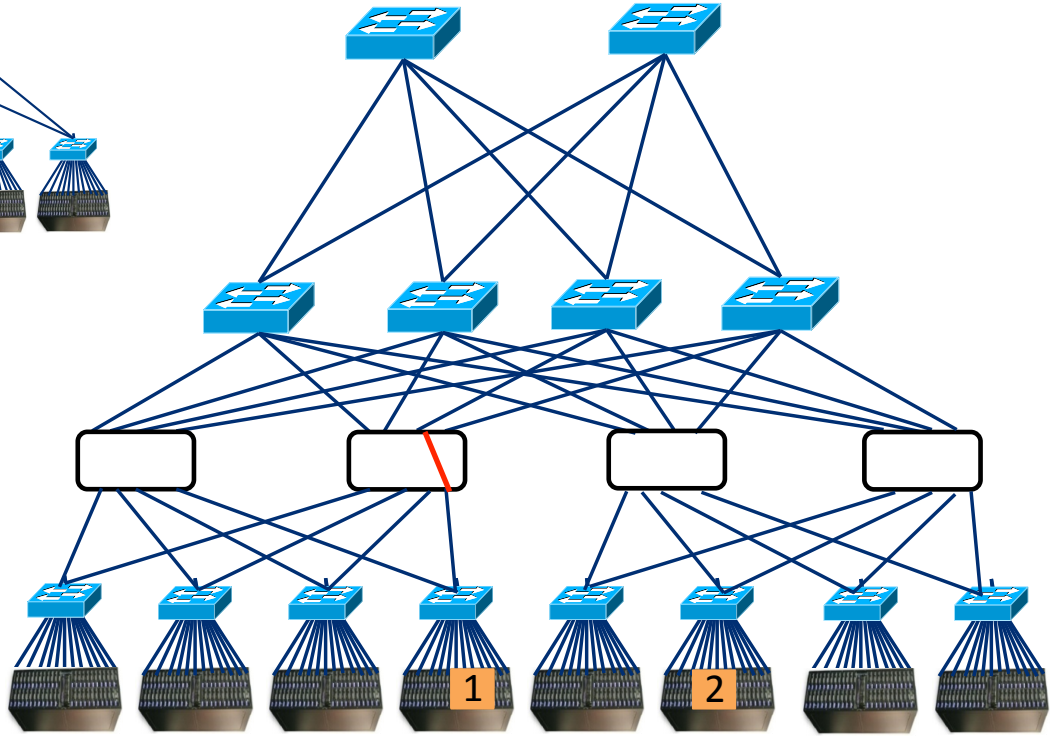
Topological flexibility example



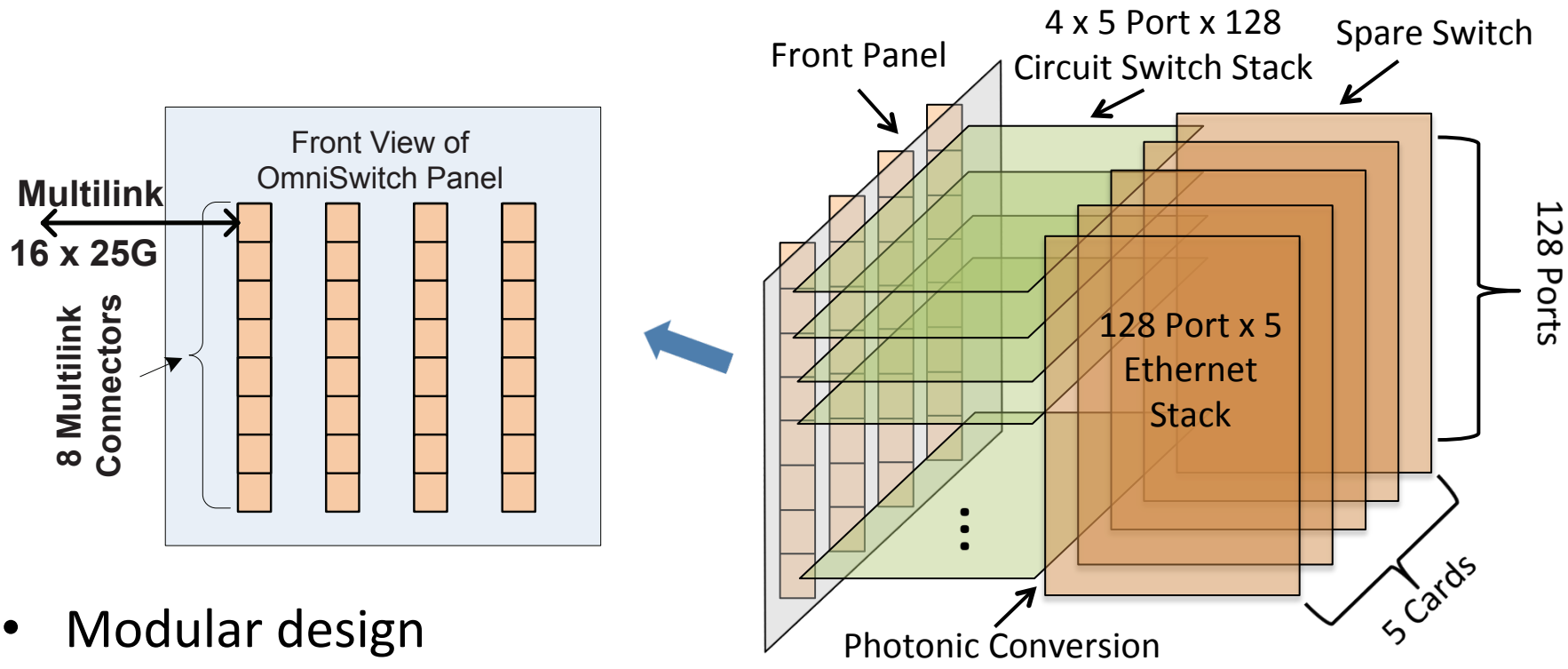


Topological flexibility example



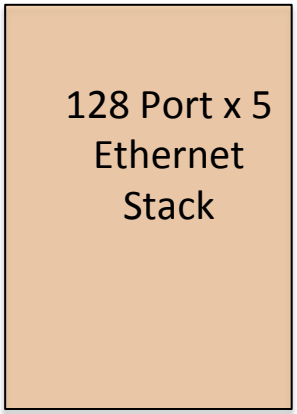


OmniSwitch architecture



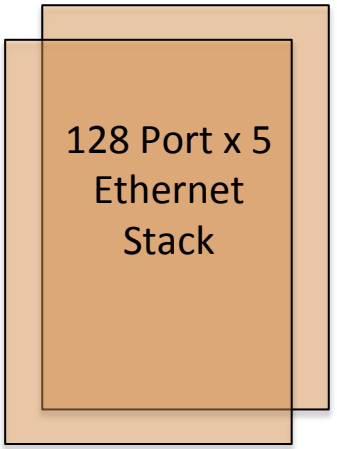
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



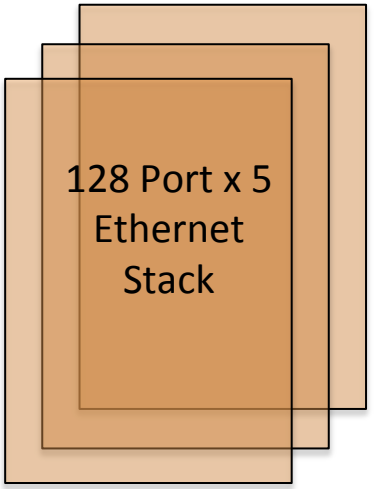
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



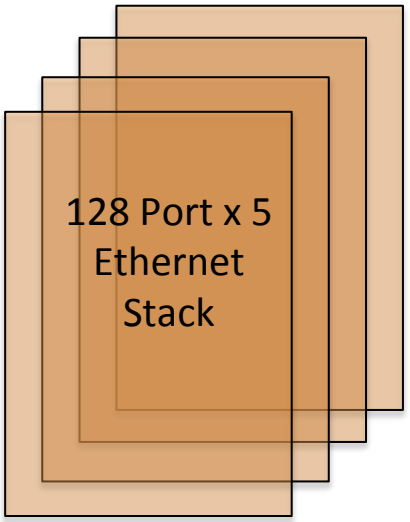
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



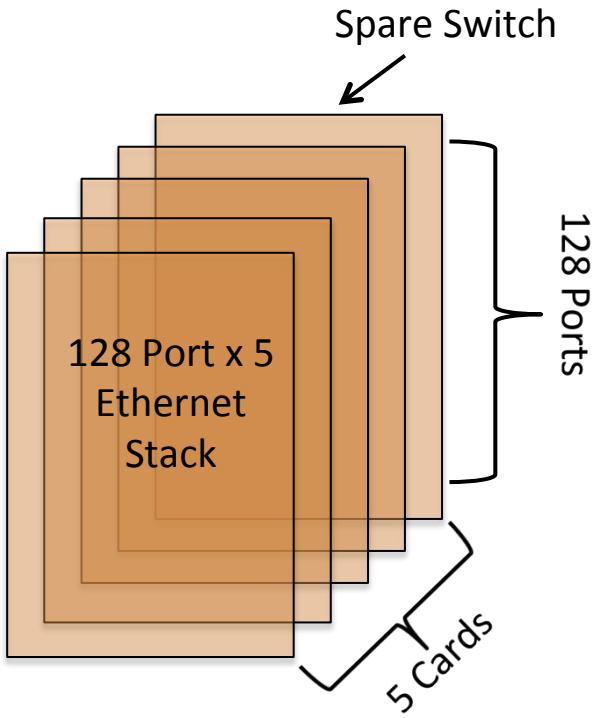
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



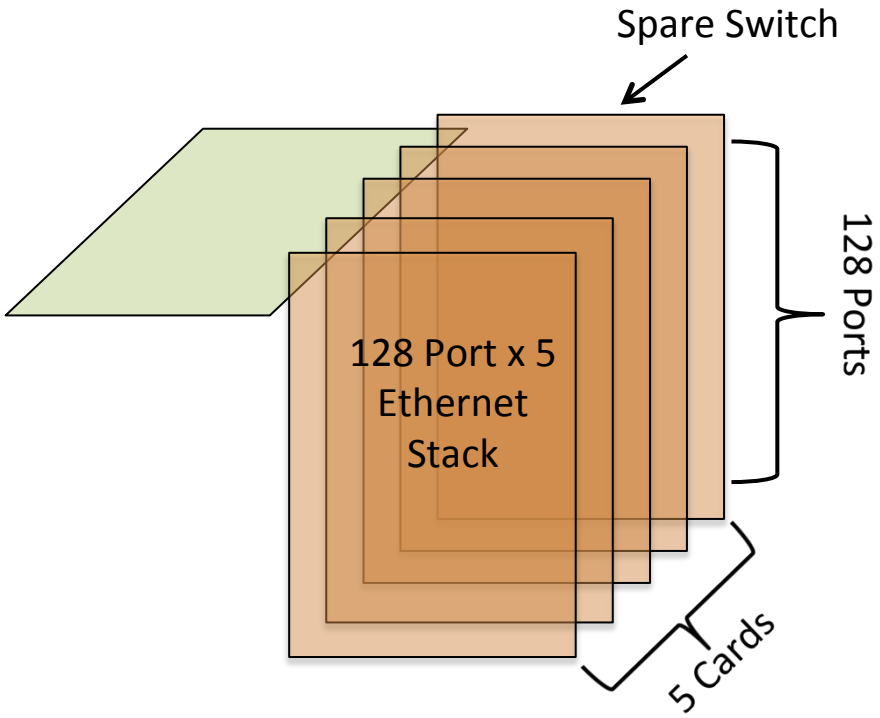
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



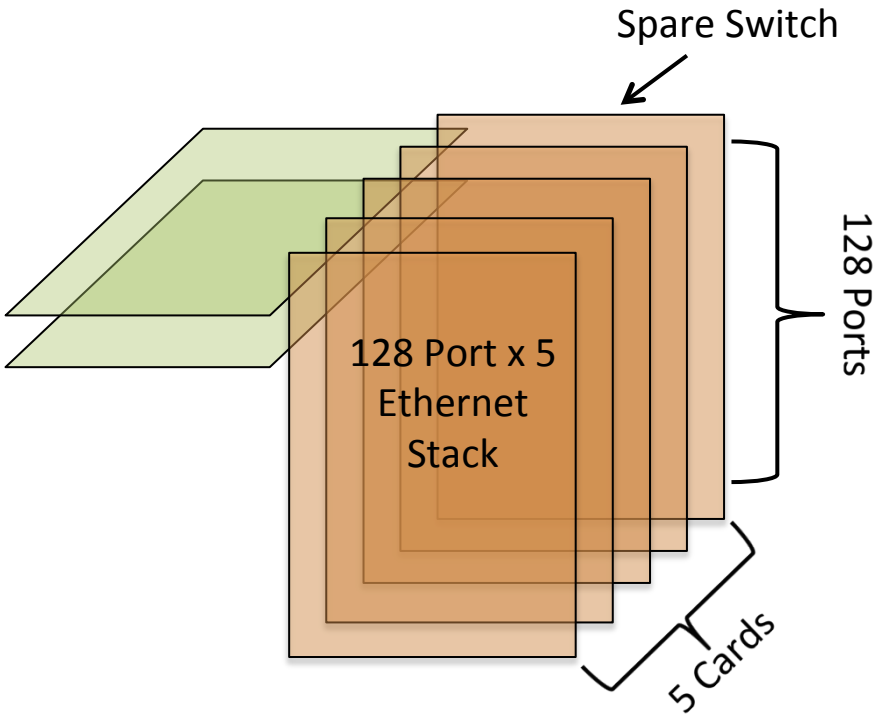
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



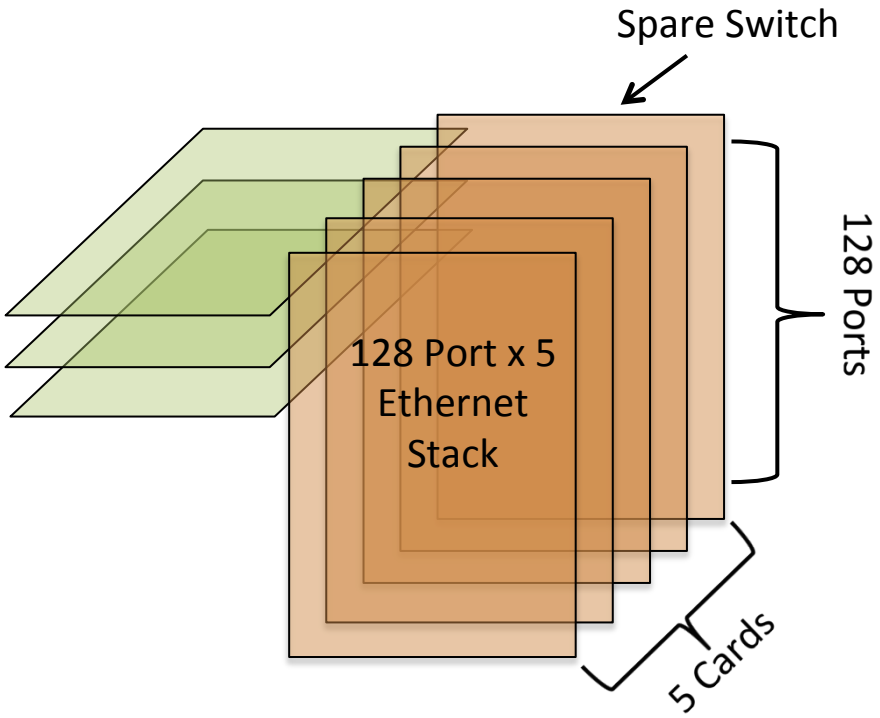
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



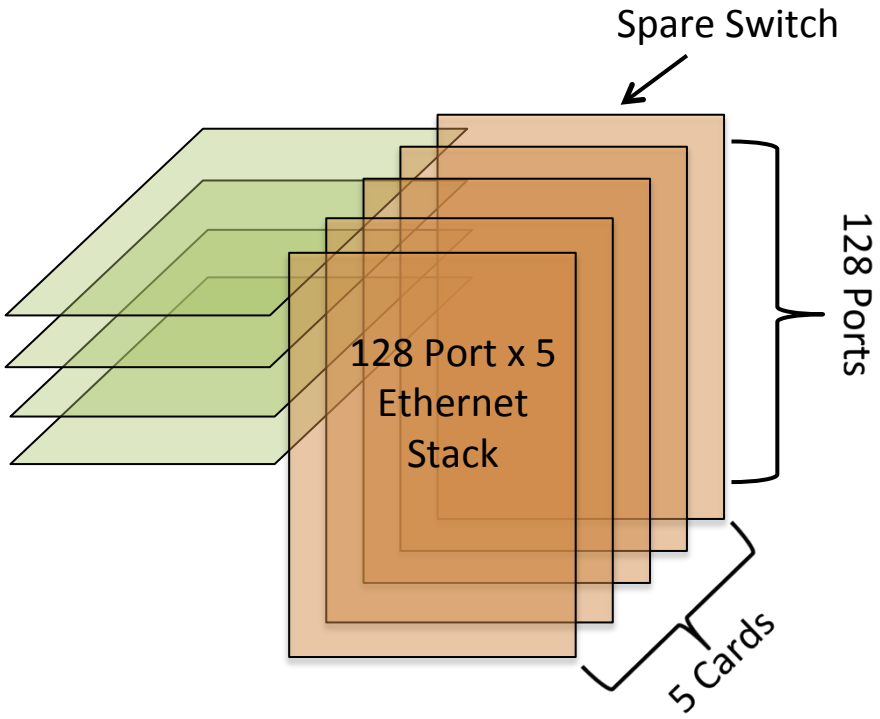
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



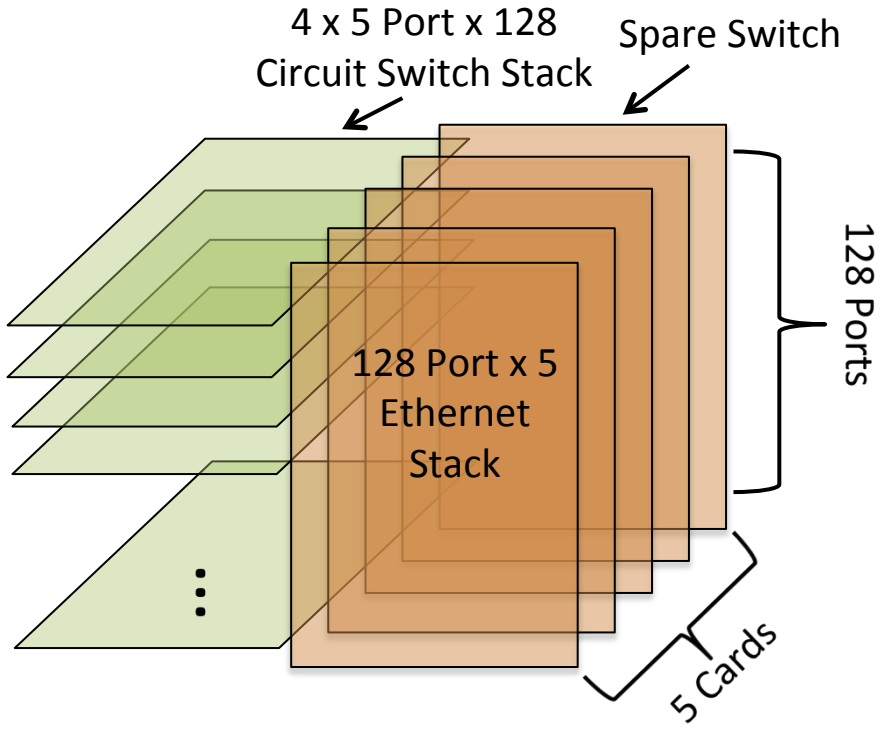
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



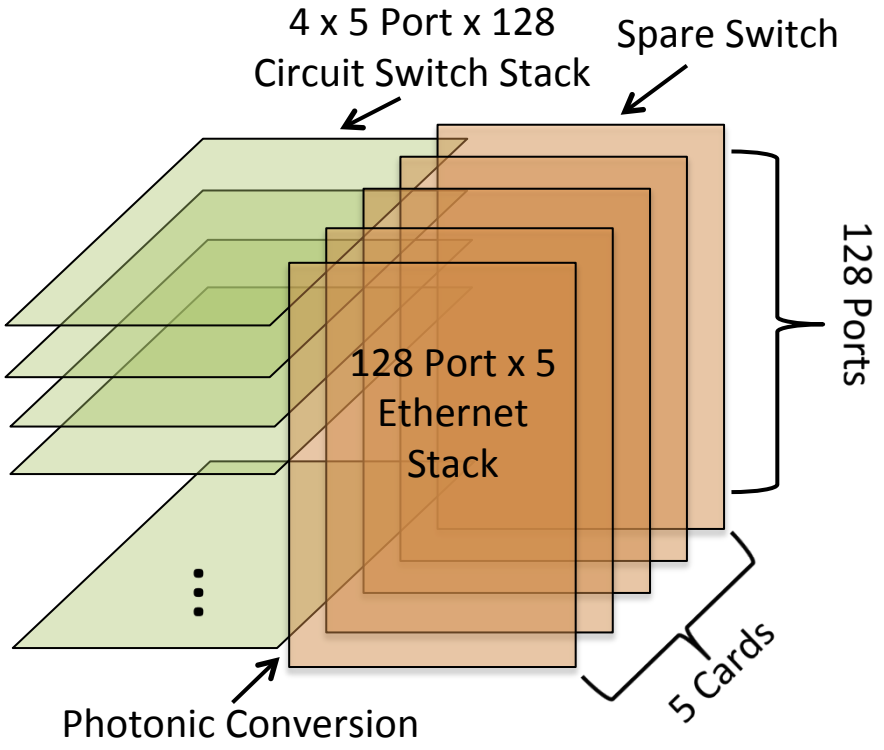
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



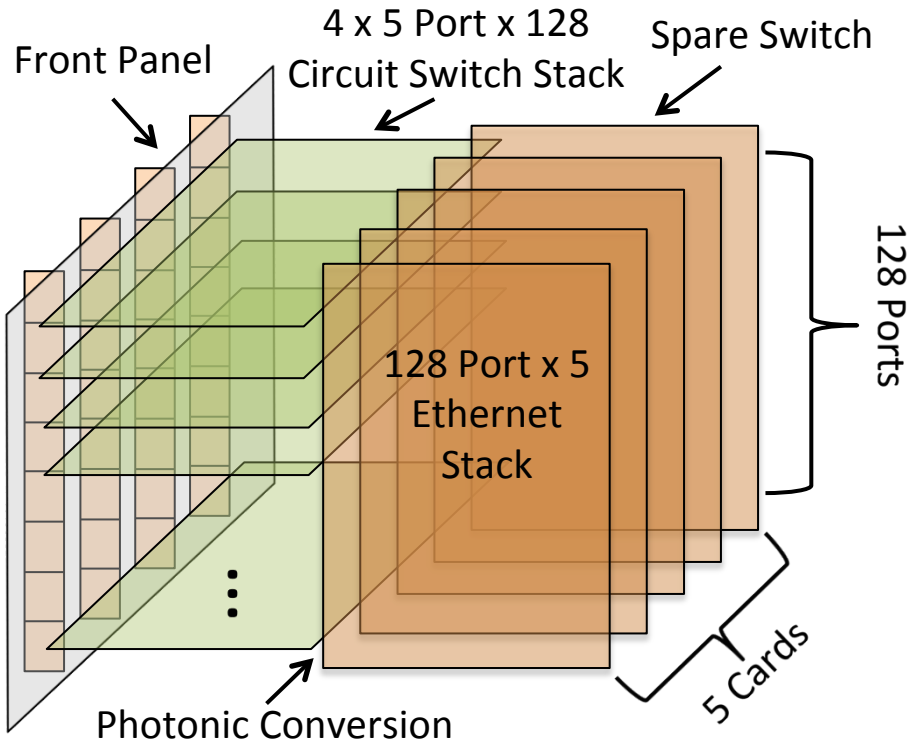
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



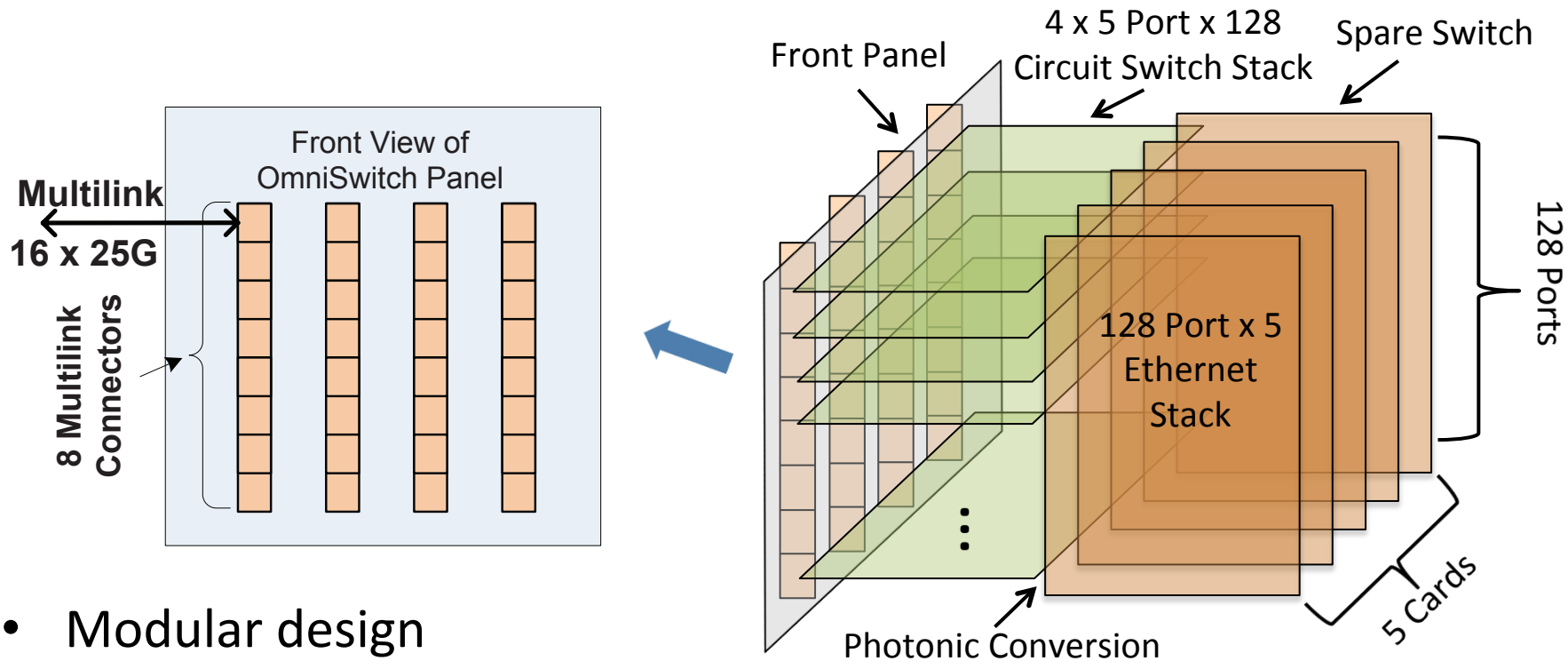
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



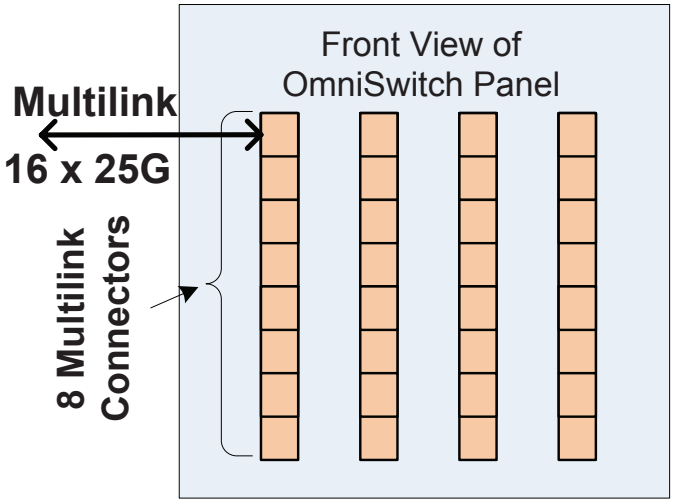
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



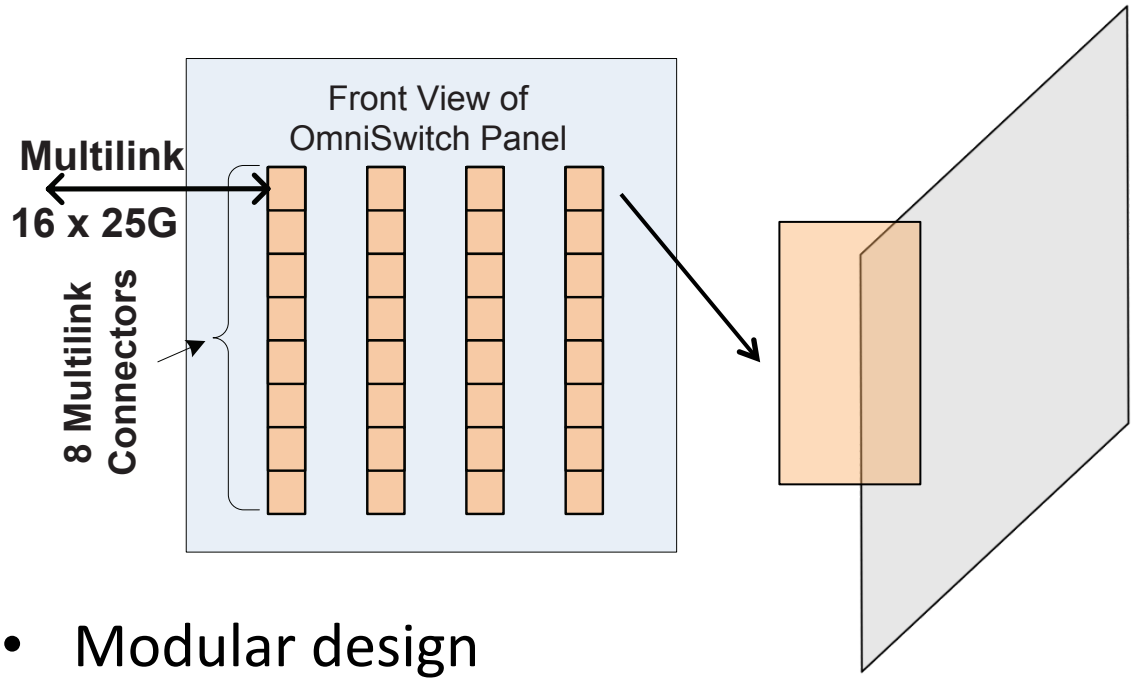
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



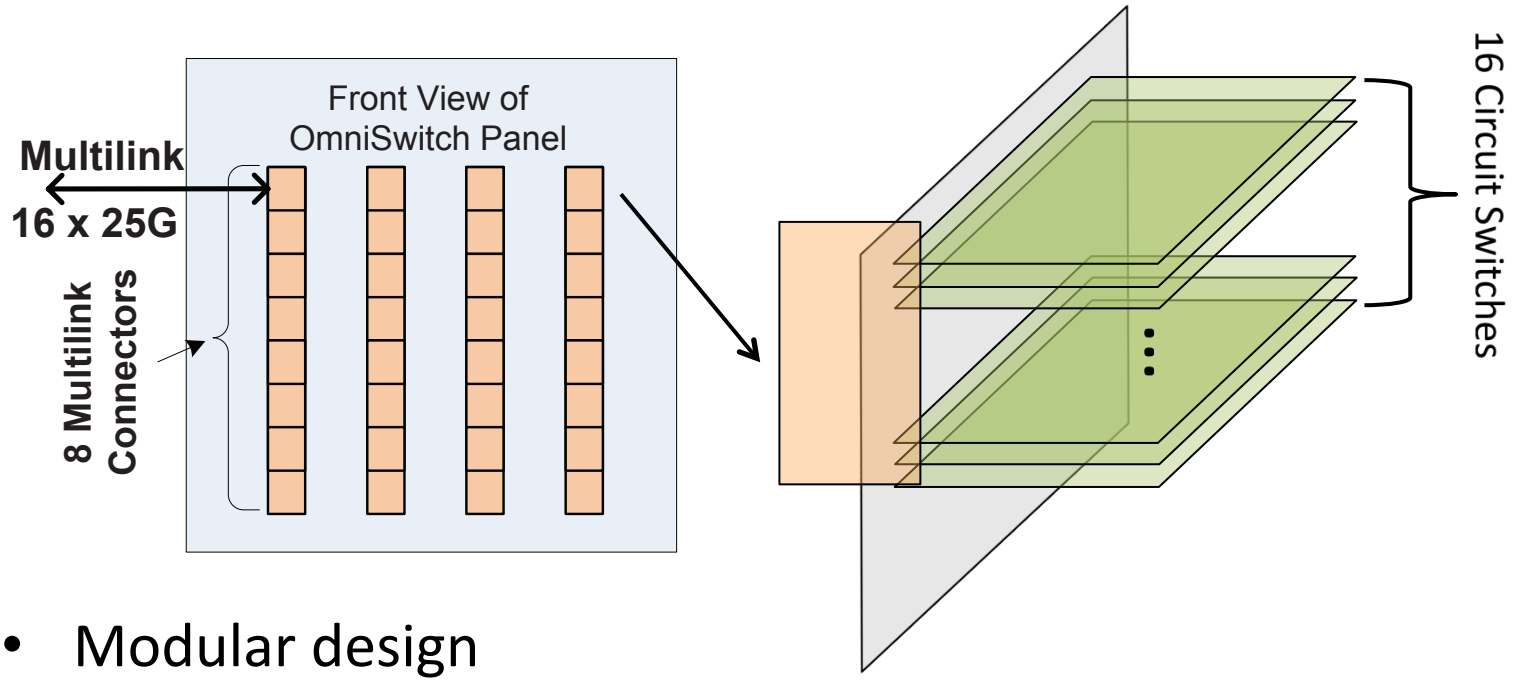
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



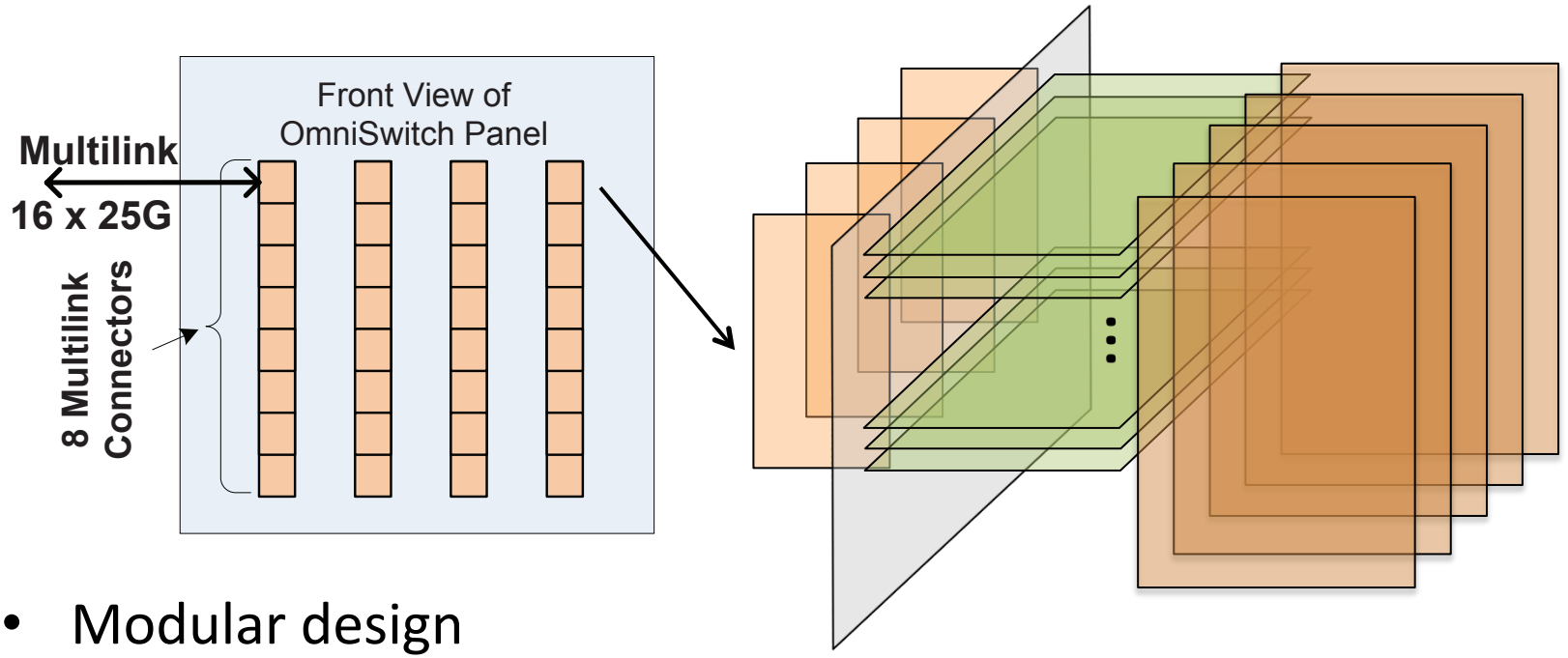
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



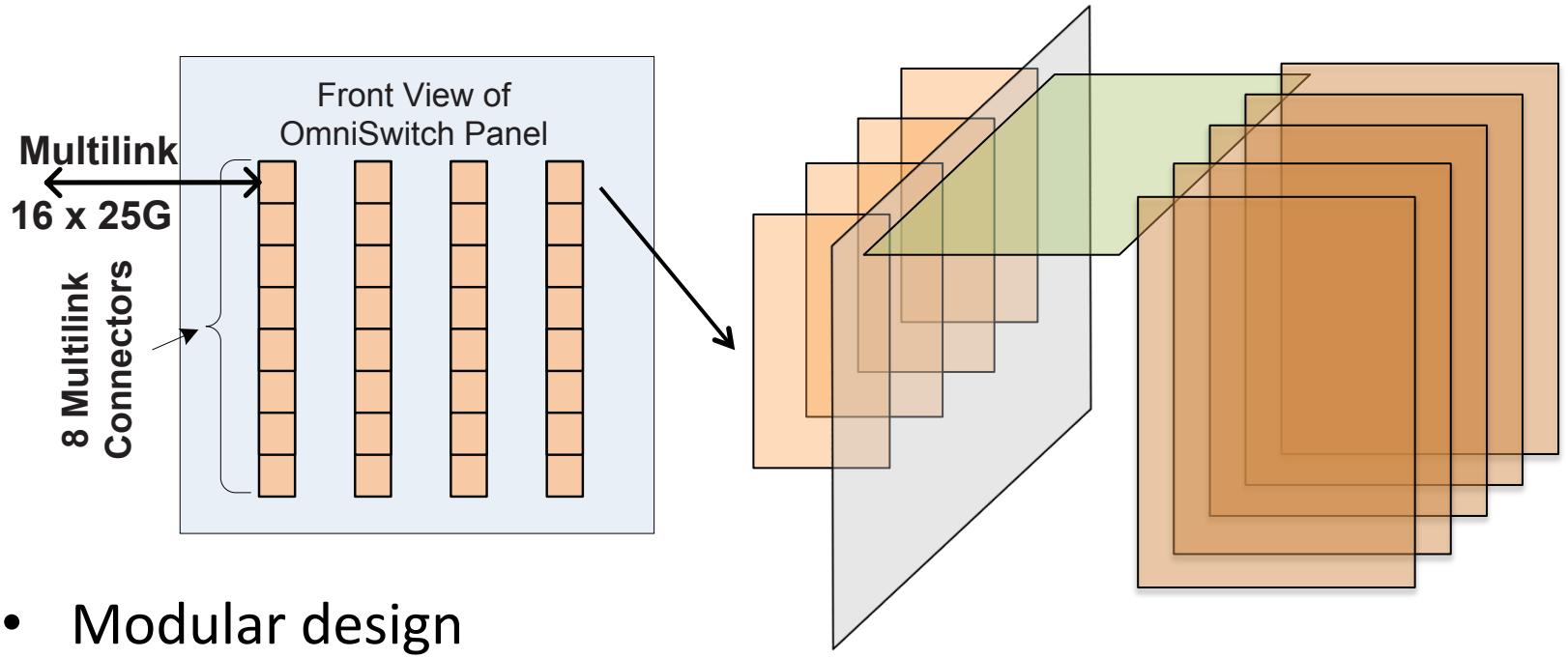
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



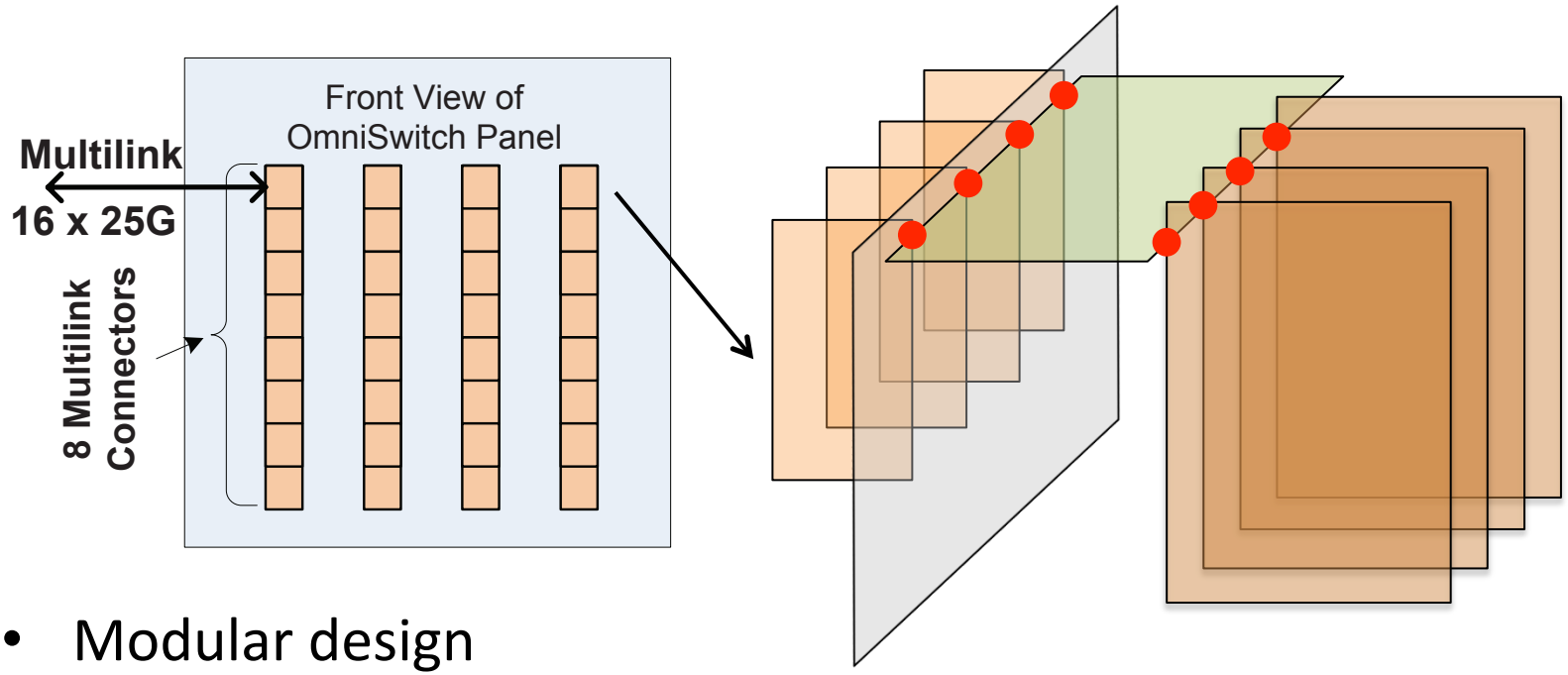
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



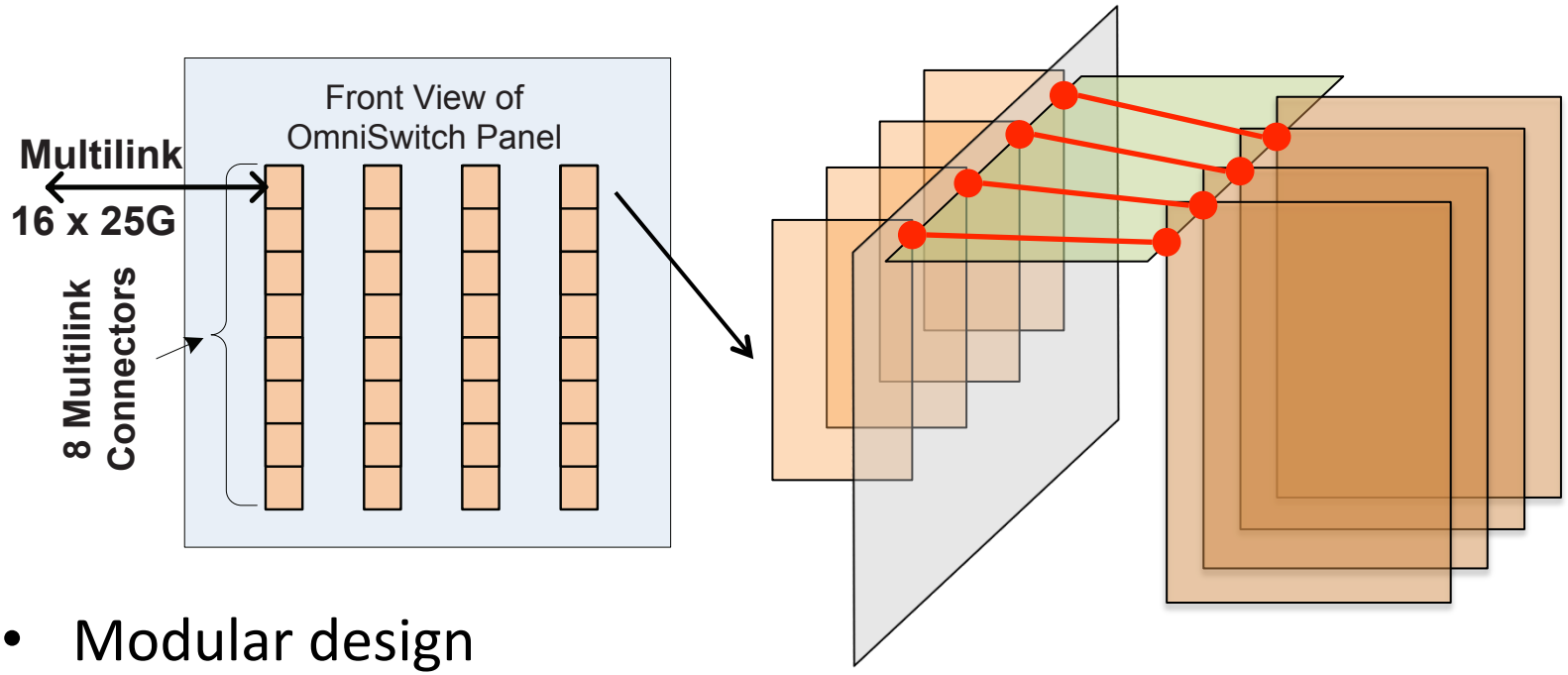
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



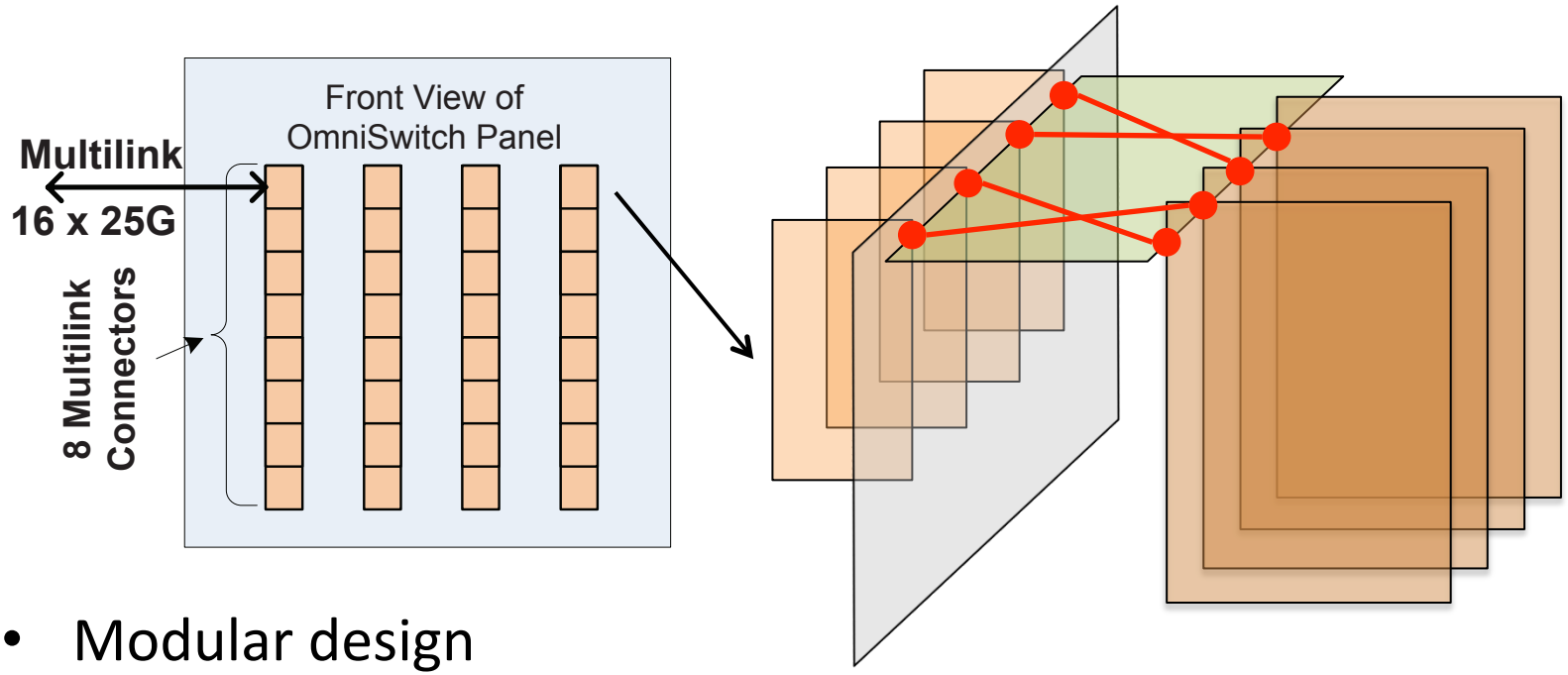
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



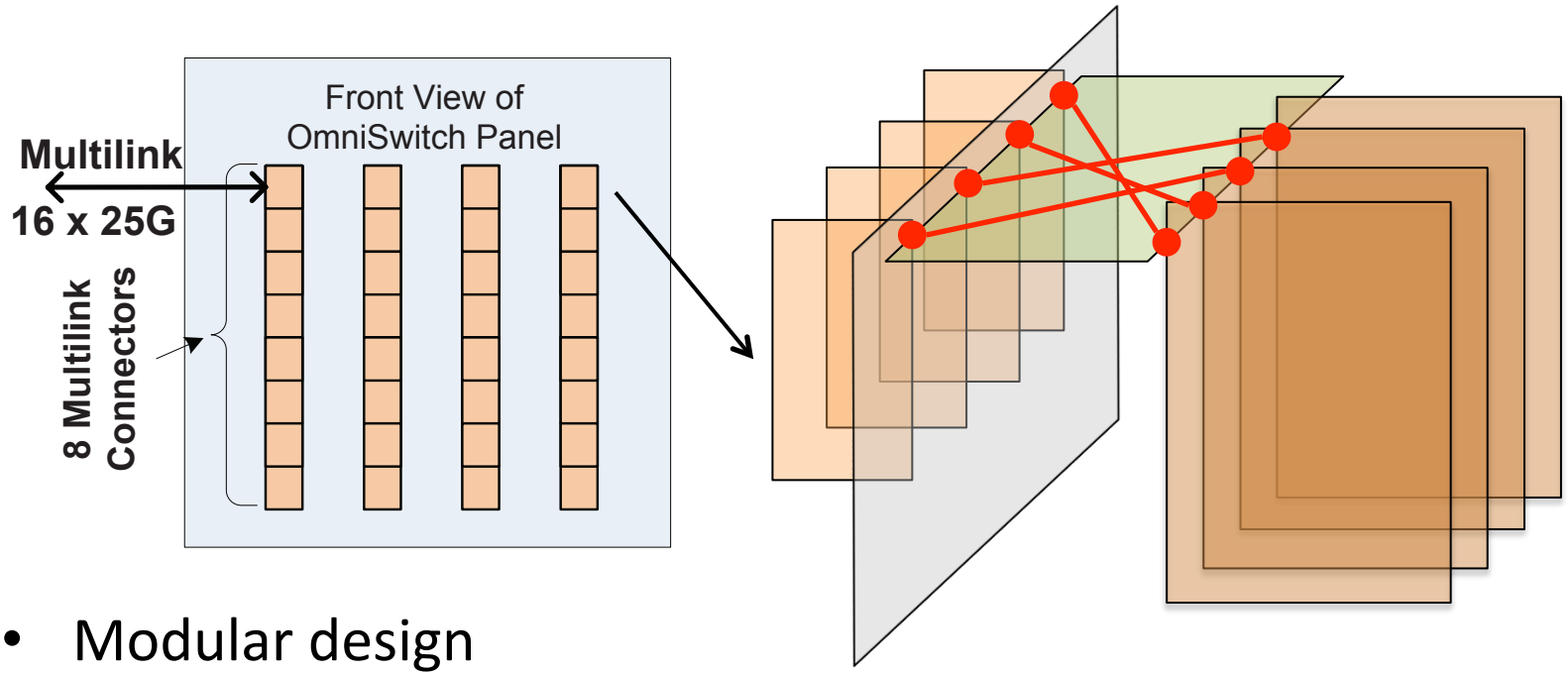
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



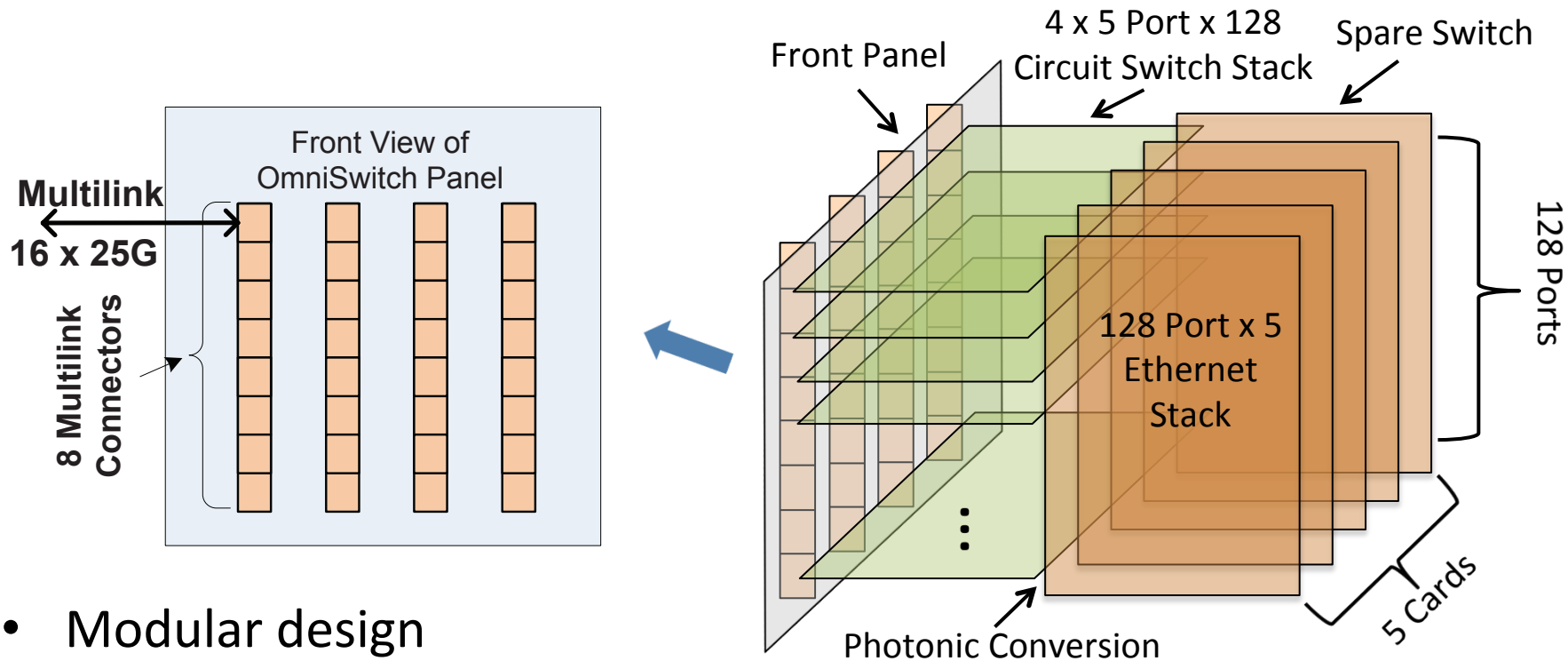
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture



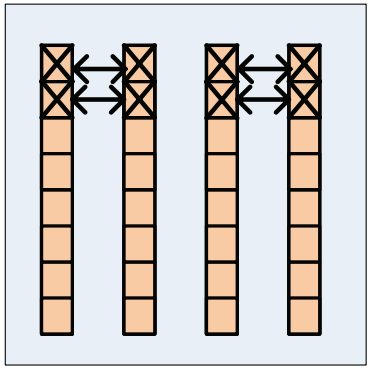
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

OmniSwitch architecture

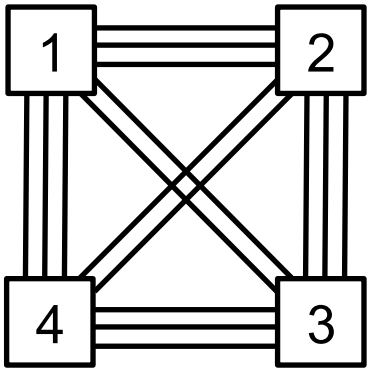


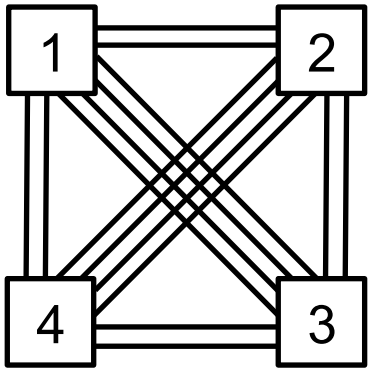
- Modular design
- Small optical circuit switches
- Interleaving optical and Ethernet switches
- 1 out of 5 Ethernet switch as backup
- 16 fibers bundled into 1 multilink

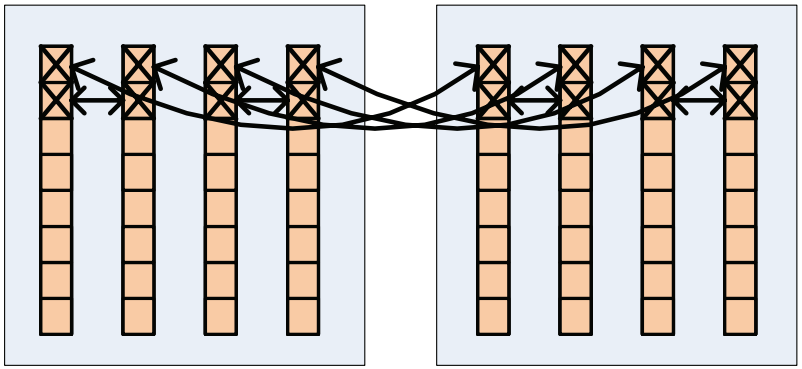
OmniSwitch mesh network



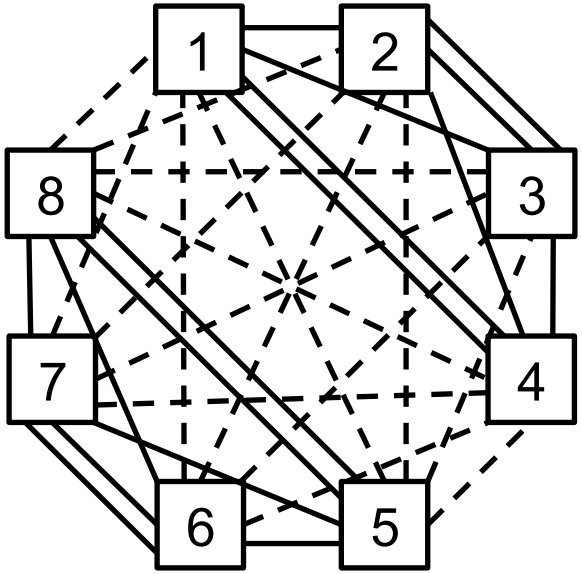
384 Port



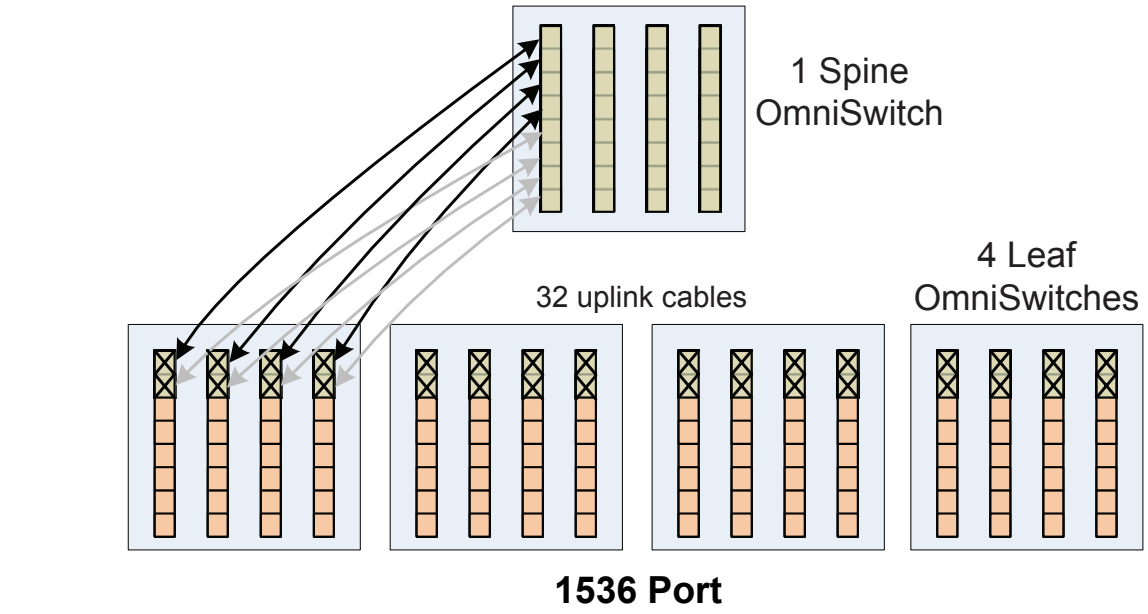


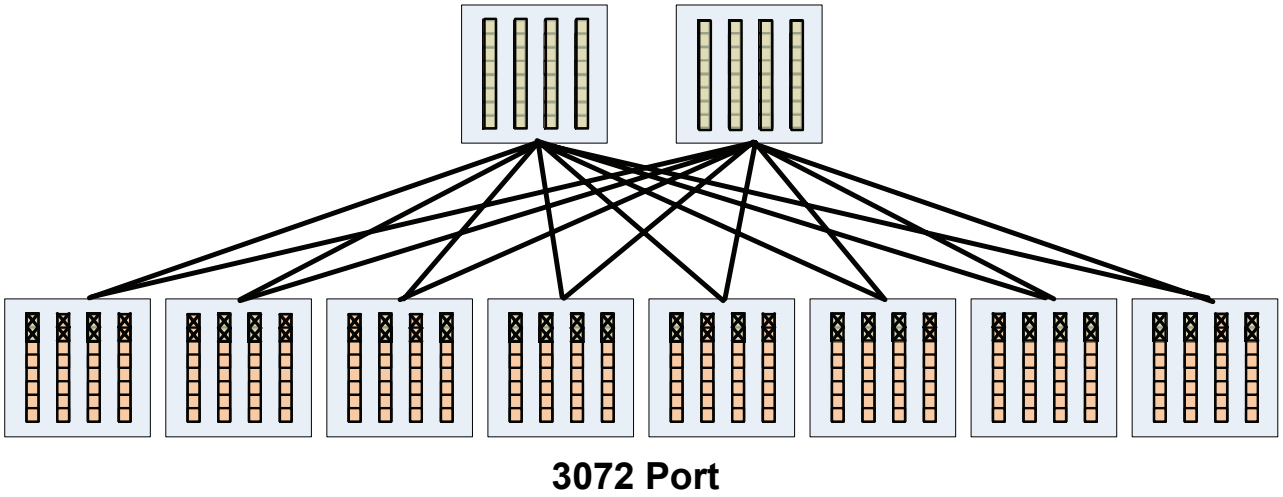


768 Port



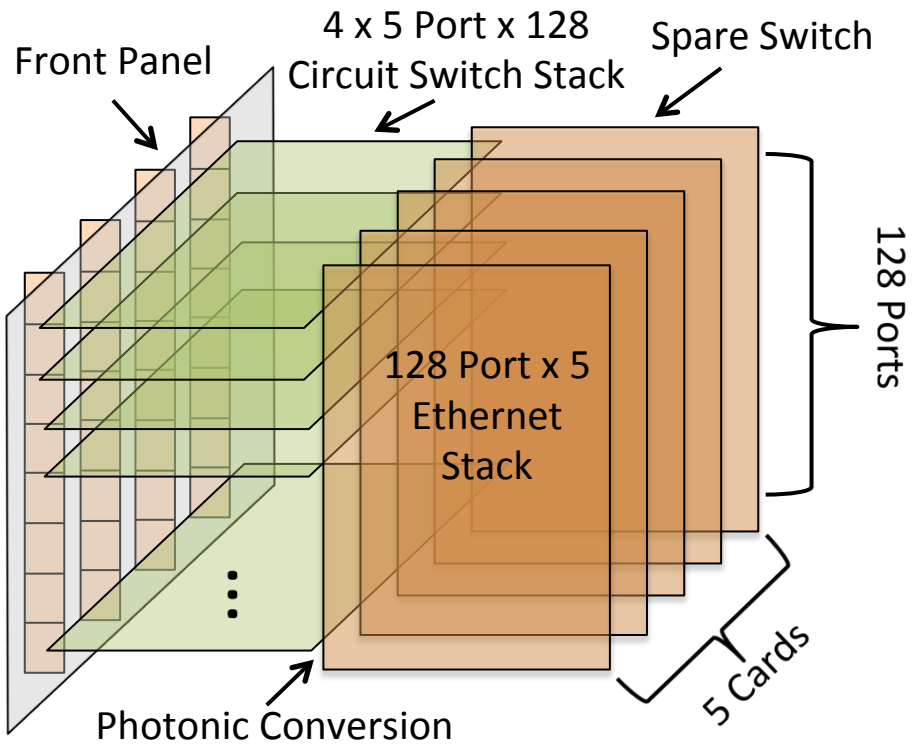
OmniSwitch tree network





Benefits of OmniSwitch

- Easy wiring
  - *multilinks*
  - *automatic internal wiring*
- Incremental expansion
  - *add equipments in cabinet*
  - *purchase new cabinets*
- Efficient backup
  - *1:5 vs. 1:1 redundancy*
- Traffic optimization
  - *traffic locality*
  - *load balancing*
- Cost effectiveness
  - *cheap small optical circuit switches*
  - *constrained flexibility*

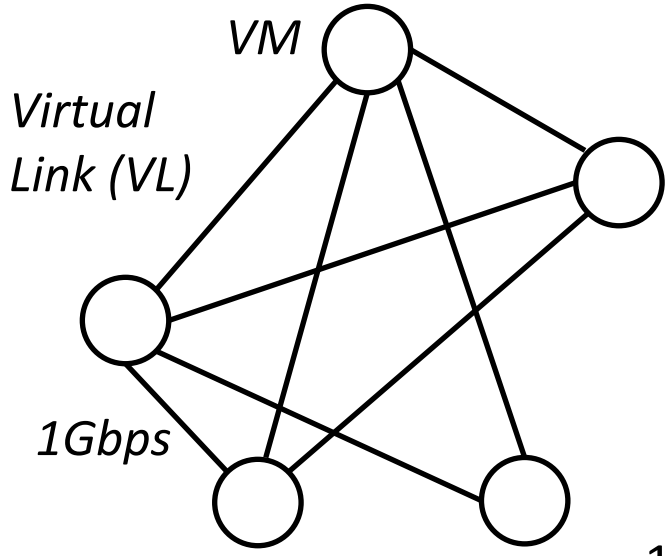


Control plane

- Existing routing schemes do not work
  - *need to configure topology & routing*
- Data center administered by single entity
  - *centralized network controller*
- Optical switches configured by software
- Routing enforced by many possible mechanisms
  - *OpenFlow protocol*
  - *VLAN*
  - *source routing*
- Control logic customized to use cases

VM clustering: a case study

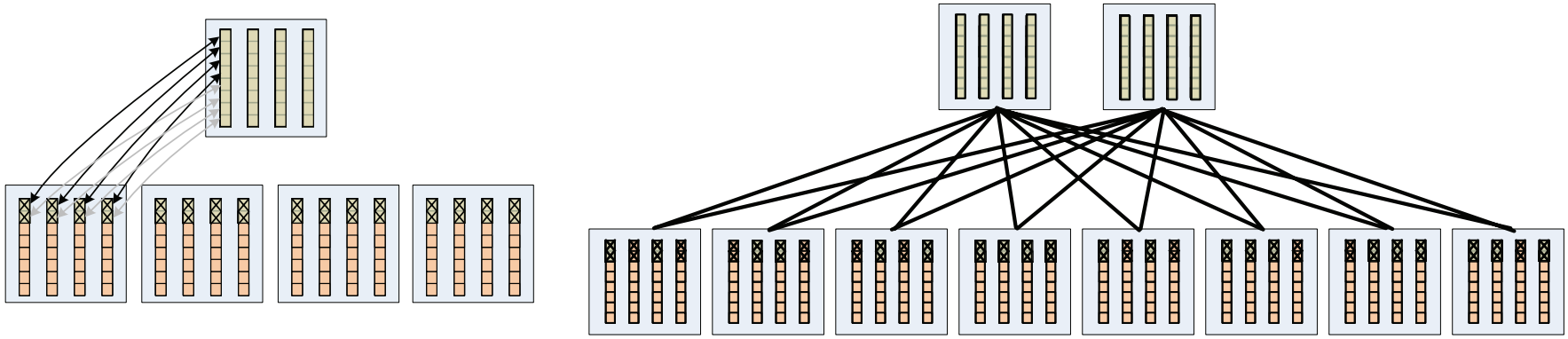
- Tenant refers to a cluster of VMs
- VM only talks to other VMs in the same tenant
- Localize traffic within tenant
- Tenant management, not flow management
- Input: pipe model
- Output:
  1) optical switch settings
  2) placement of VLs
     on physical links



Control algorithm

- Hard problem
  - *NP-complete for fixed topology*
  - *search space larger for OmniSwitch*

Control algorithm



- Tree-structured OmniSwitch cabinets
- Iterative from bottom up
- VMs already placed → better topology & routing
- Search space small

  - uplinks on switches

  - permutations on small optical switches
- Configure topology and routing at the same time

Evaluations

- Tenant provisioning scenario
- Smart VM placement vs. flexible topology
- Systems compared

✓ dumb

- contiguous VM placement + ECMP routing + fixed Clos network

✓ SecondNet

- SecondNet VM placement and routing + fixed Clos network

✓ OmniSwitch

- contiguous VM placement + VM clustering algorithm + OmniSwitch network

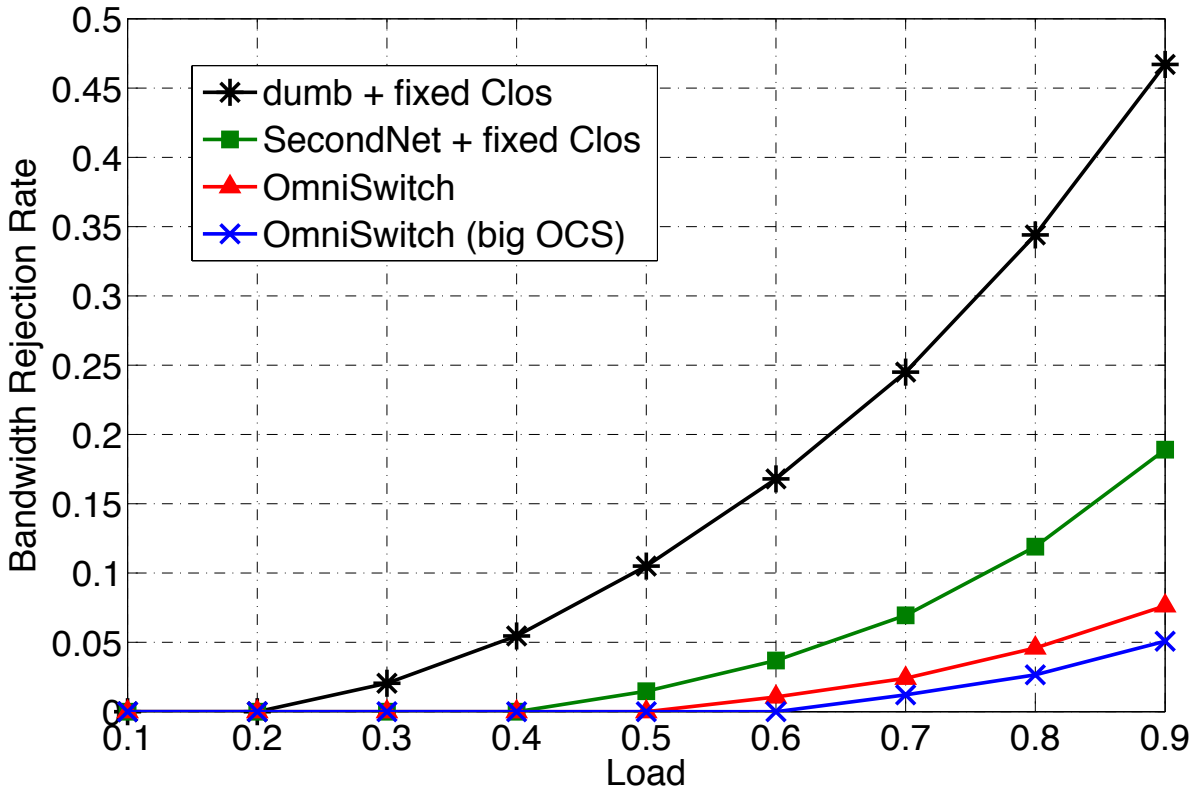
✓ OmniSwitch (big OCS)

- same as above, but 1 big optical switch in each OmniSwitch cabinet

Bandwidth rejection rate

bw rejection rate = rejected bw / requested bw

load = tenant size * arrival rate * duration / #VM

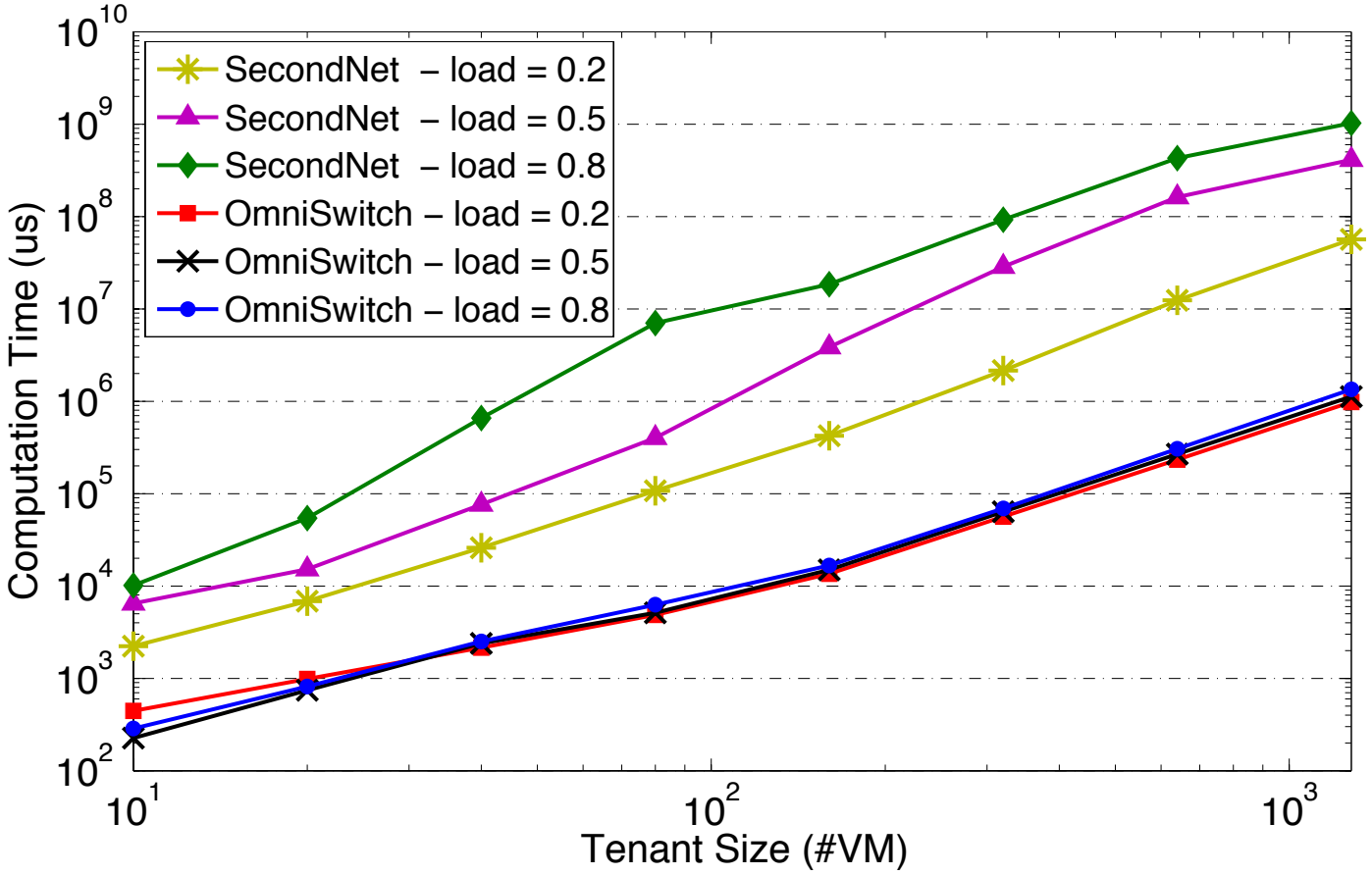


Transmission hop count

Average hop count when load = 0.8

dumb	SecondNet	OmniSwitch	OmniSwitch (big OCS)
4.422	4.164	3.217	3.048

Computation time



17min

1s

Conclusion

- OmniSwitch is a modular data center architecture
- Integrated optical switches and Ethernet switches for topological flexibility and large-scale connectivity
- Topological flexibility improves traffic optimization and simplifies network management
- VM clustering as a case study

  - *reduce transmission path length*

  - *host more bandwidth*

  - *low computation time*