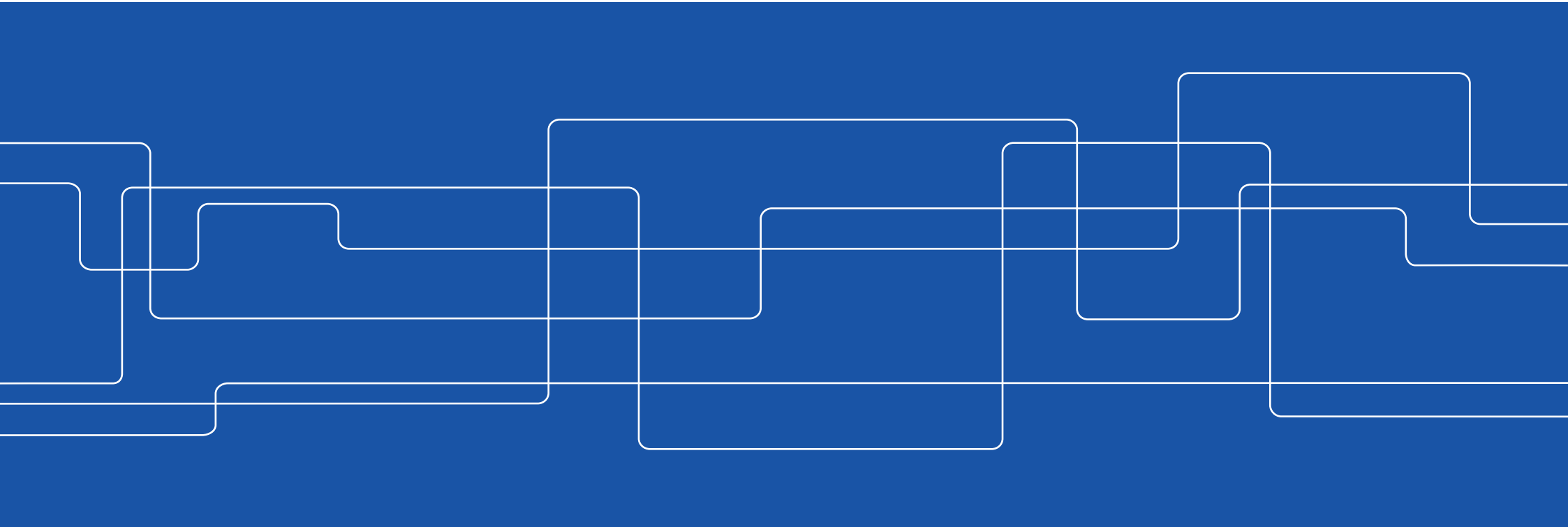
Optical Interconnects for Data Centers

Matteo Fiorani

Optical Networks Lab (ONLab)
Communication Systems Department (COS)
KTH Royal Institute of Technology
Stockholm (Sweden)



ROYAL INSTITUTE OF TECHNOLOGY

ONLab

About myself

- **2010** Master in Telecommunications Engineering

University of Modena and Reggio Emilia (UNIMORE), Italy

- **2014** PhD in Information and Communications Technologies

University of Modena and Reggio Emilia (UNIMORE), Italy

Vienna University of Technology (TUW), Austria

- **2014-Now** Postdoc in Optical Networks

Royal Institute of Technology (KTH), Sweden

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

My research

- **5G Transport**
- **Optical interconnects for data centers**
- **Multi-domain SDN orchestration**

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Optical Interconnects for Data Centers

- Optical interconnects at Top-of-Rack for energy efficient data centers
  1. J. Chen, Y. Gong, M. Fiorani, S. Aleksic, "Optical Interconnects at Top of the Rack for Energy-Efficient Datacenters", IEEE Communications Magazine, Vol. 53, Issue 8, pp.140-148, August 2015.
  2. M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen, "Energy-Efficient Elastic Optical Interconnect Architecture for Data Centers", IEEE Communications Letters, Vol. 18, No. 9, pp. 1531-1534, Sept. 2014.

- Spatial division multiplexing (SDM) for ultra-high capacity modular data centers
  1. Now in progress…

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Optical Interconnects for Data Centers

- Optical interconnects at Top-of-Rack for energy efficient data centers
  1. J. Chen, Y. Gong, M. Fiorani, S. Aleksic, "Optical Interconnects at Top of the Rack for Energy-Efficient Datacenters", IEEE Communications Magazine, Vol. 53, Issue 8, pp.140-148, August 2015.
  2. M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen, "Energy-Efficient Elastic Optical Interconnect Architecture for Data Centers", IEEE Communications Letters, Vol. 18, No. 9, pp. 1531-1534, Sept. 2014.

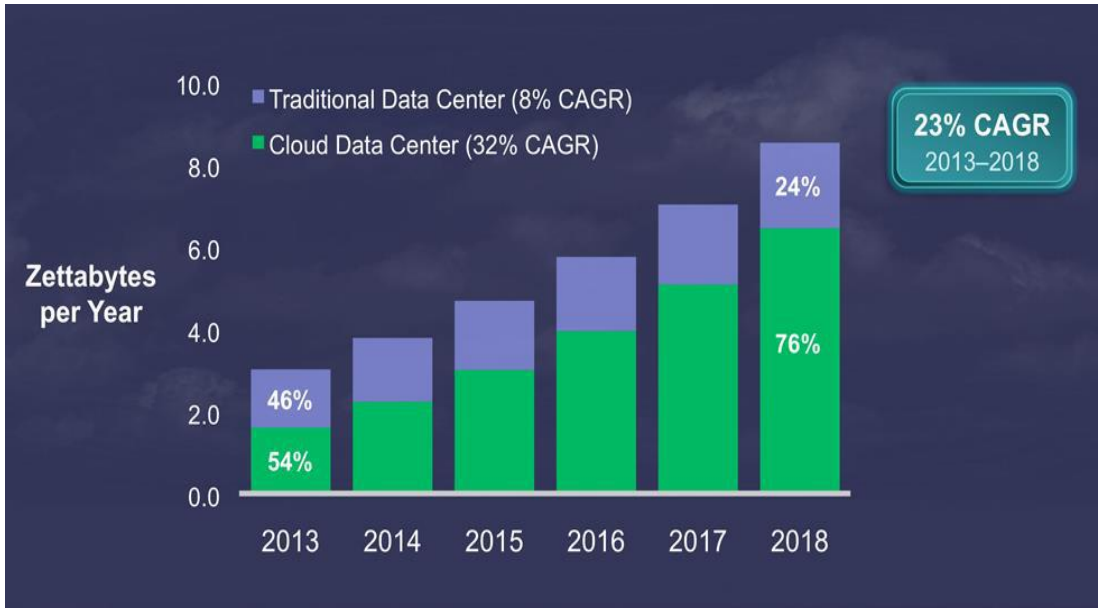
- Spatial division multiplexing (SDM) for ultra-high capacity modular data centers
  1. Now in progress…

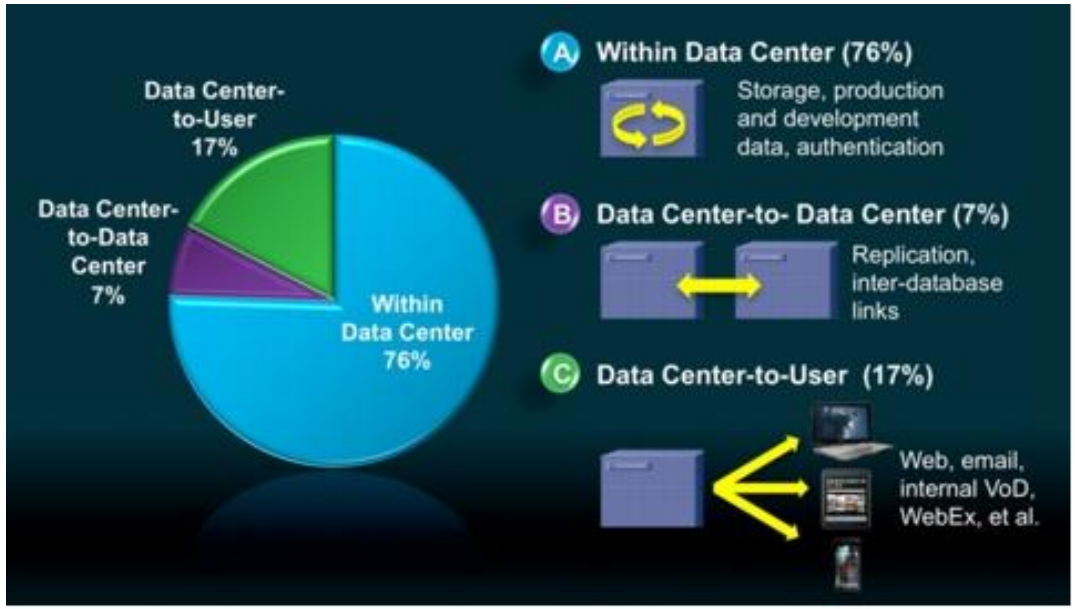
ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Motivation (1/2)

- Data center traffic[1] :
  1. Increases with a compound annual growth rate (CAGR) of 23%
  2. Main contributor is cloud computing traffic
  3. Majority of traffic is exchanged among serves inside same data center





*[1] *Cisco white paper: "Cisco Global Cloud Index: Forecast and Methodology, 2013-2018," 2014.*

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Motivation (2/2)

Source: Cisco (www.cisco.com)

Collaboration

Empowered User

SLA Metrics

Global Availability

Reg. Compliance

New Business Pressures

Challenges

Scalability

Power & Cooling

Provisioning

Security Threats

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

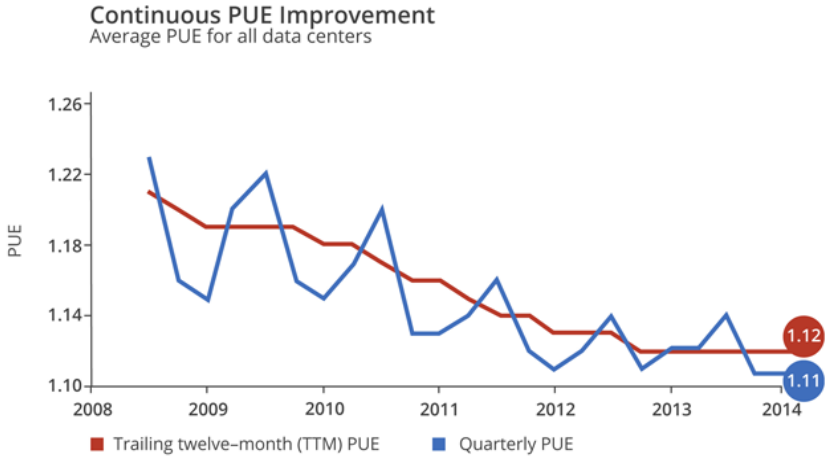
Power consumption (2/3)

- Datacenter power consumption:
  1. IT equipment (P_{IT})
  2. Cooling system (P_{CS})
  3. Power supply chain (P_{PSC})

- Power Usage Effectiveness (PUE)

$$PUE = \frac{P_{IT} + P_{CS} + P_{PSC}}{P_{IT}}$$

- Exploiting efficient power control systems modern data centers reach PUE ≈1
- Further PUE improvement can be achieved by a "wise" data center location



Average PUE for Google® data centers
www.google.com/about/datacenters/efficiency/internal/

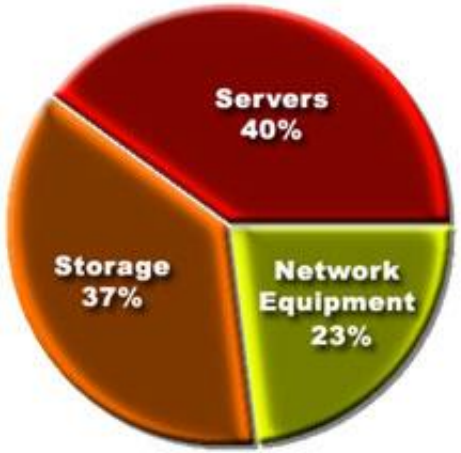
The Facebook® Artic data center in Luleå (Sweden)

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Power consumption (3/3)

- IT power consumption (P_{IT}):
  1. Servers
  2. Storage
  3. Network (≈ 23% [1])



- Datacenter traffic increase in a large cloud computing data center [2]:

Features	2012	2016	2020
Bandwidth	1 Pbytes/s	20 Pbytes/s	400 Pbytes/s
Network Power Consumption	0.5 MW	1 MW	2 MW

  - Traffic can increase up to 400 times
  - Affordable power consumption increase is only 4 times
  - **Objective**: decrease the data center network power consumption (W/bit/s) by **100 times**

[1] "Where does power go?", GreenDataProject, available online at http://www.greendataproject.org/
[2] C. Kachris, et al., "Optical interconnection networks in data centers: recent trends and future challenges," IEEE Comm. Mag. September 2013

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

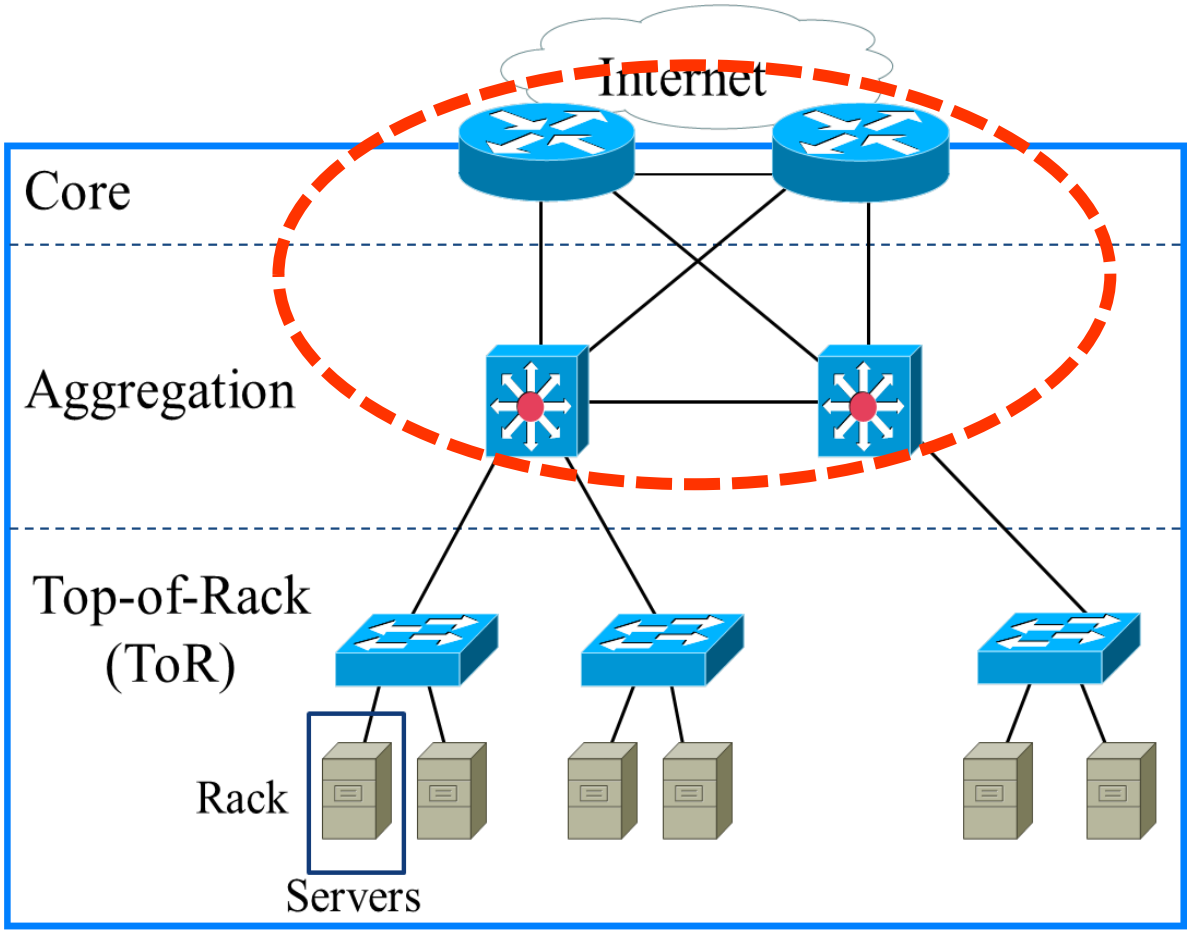
Limitations of current solutions (1/5)

- Current large scale data center architecture:
  - Fat-tree 3-tiers topology
  - Electronic switching
    - High power consumption
    - Poor scalability

- Major energy savings are not possible in the approaches based on electronic switches [3][4]

- **Solution**:

Optical switching in data center networks



[3] S. Aleksic, "Analysis of power consumption in future high-capacity network nodes", IEEE/OSA J. Opt. Comm. And Net., 2009
[4] M. Fiorani, et al., "Performance and power consumption analysis of a hybrid optical core node", IEEE/OSA J. Opt. Comm. And Net., 2011

ROYAL INSTITUTE OF TECHNOLOGY

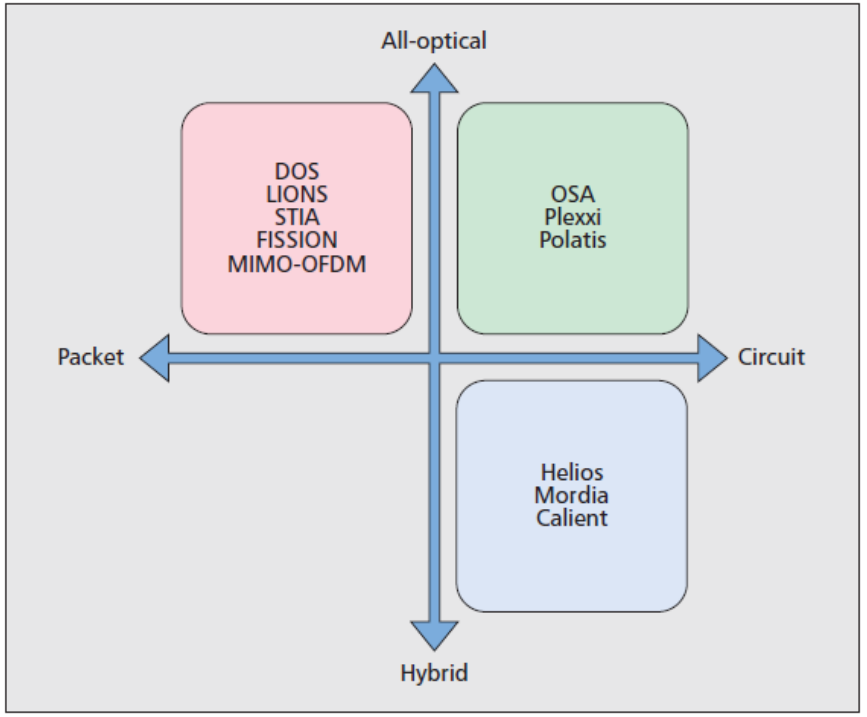
ONLab

Limitations of current solutions (2/5)

- Current optical switching solutions for data centers:
  - Replace the **aggregation** and **core** tires with a single large optical core switch
  - Can be categorized basing on the adopted switching technology

1. Hybrid Optical/Electronic
   - Circuit switching in the optical domain
   - Packet switching in the electronic domain
2. Optical Circuit Switching
3. Optical Packet Switching

C. Kachris, et al., IEEE Communications Magazine, September 2013



ROYAL INSTITUTE OF TECHNOLOGY

ONLab

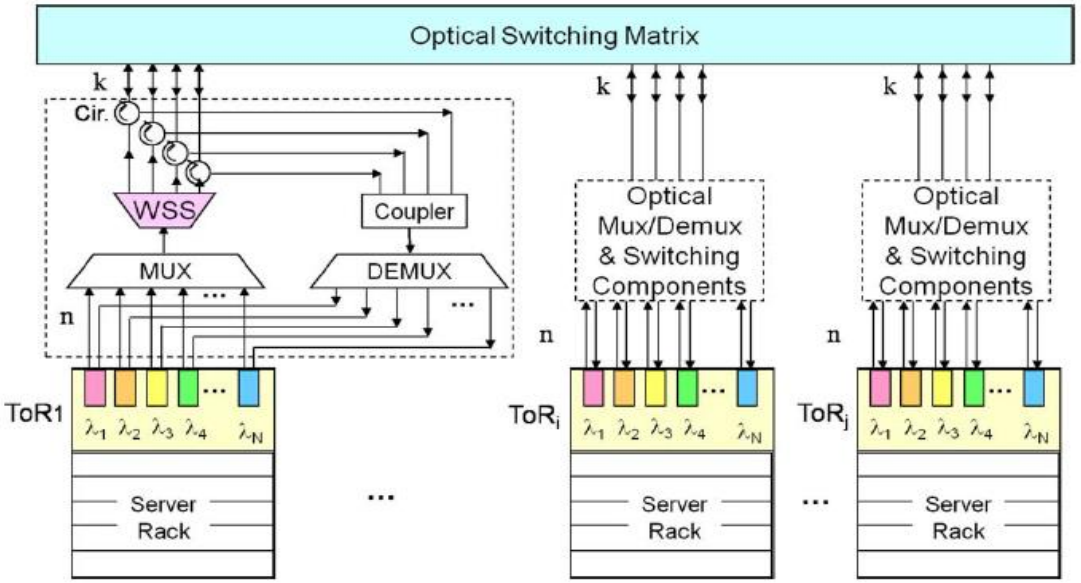
Limitations of current solutions (3/5)

- Current optical switching solutions for data centers:
  - Replace the **aggregation** and **core** tires with a single large optical core switch
  - Can be categorized basing on the adopted switching technology

1. Hybrid Optical/Electronic
2. Optical Circuit Switching

   OSA: An Optical Switching Architecture for Data Center Networks With Unprecedented Flexibility
3. Optical Packet Switching

K. Chen, et al., IEEE/ACM Transactions on Networking, 2014



ROYAL INSTITUTE OF TECHNOLOGY

ONLab

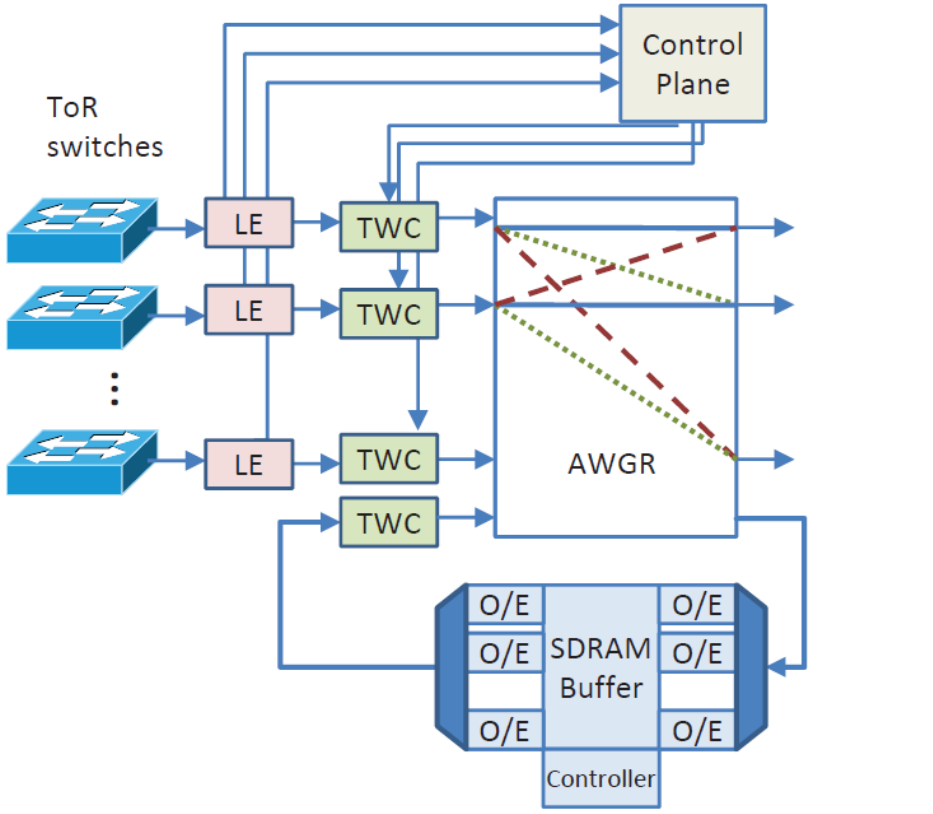
Limitations of current solutions (4/5)

- Current optical switching solutions for data centers:
  - Replace the **aggregation** and **core** tires with a single large optical core switch
  - Can be categorized basing on the adopted switching technology

1. Hybrid Optical/Electronic
2. Optical Circuit Switching
3. Optical Packet Switching

   LIONS: An AWGR-Based Low-Latency Optical Switch for High-Performance Computing and Data Centers

Y. Yawei, et al, IEEE J. Selected Topics in Quantum Electronics, 2013

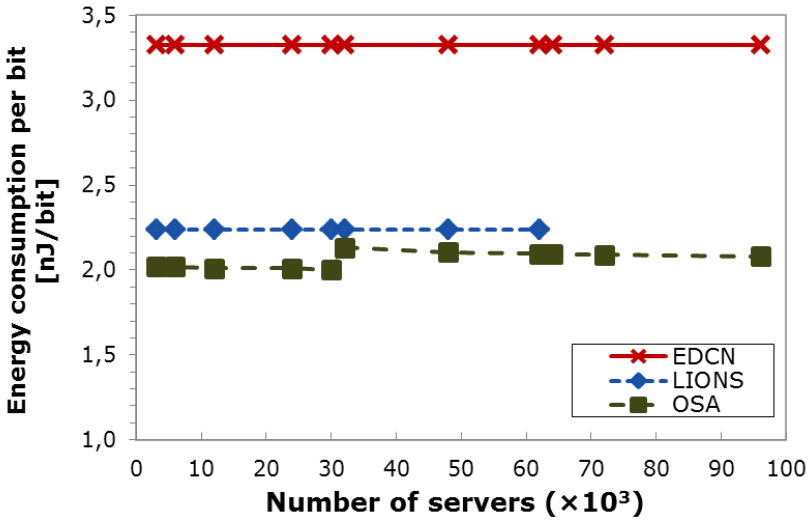


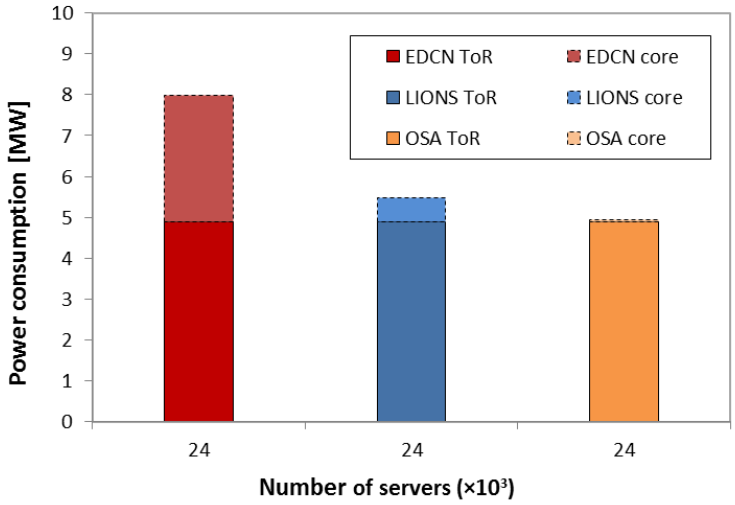
Limitations of current solutions (5/5)

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

- We assess the power consumption of OSA and LIONS and compare against a traditional data center network based on electronic switching (EDCN)





- Reduce the energy consumption per bit by a factor 2

… Not enough to deal with the expected traffic increase

- **Limitations:** high energy consumption of conventional electronic ToR switches
- **Idea:** Optical interconnects at the ToR

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Active vs. passive switching at ToR

- Active optical switching:
  - **OCS**: has very coarse granularity → not suitable for the very busty traffic at the edge tier of the data center network
  - **OPS**: has fundamental technical problems:
    - components are very expensive
    - to achieve good performance often requires electronic buffering
- Passive optical switching:
  - Only uses passive components (e.g., splitters/combiners, arrayed waveguide gratings — AWGs) to interconnect servers
    - Fine switching granularity
    - Low cost and energy consumption
    - High reliability

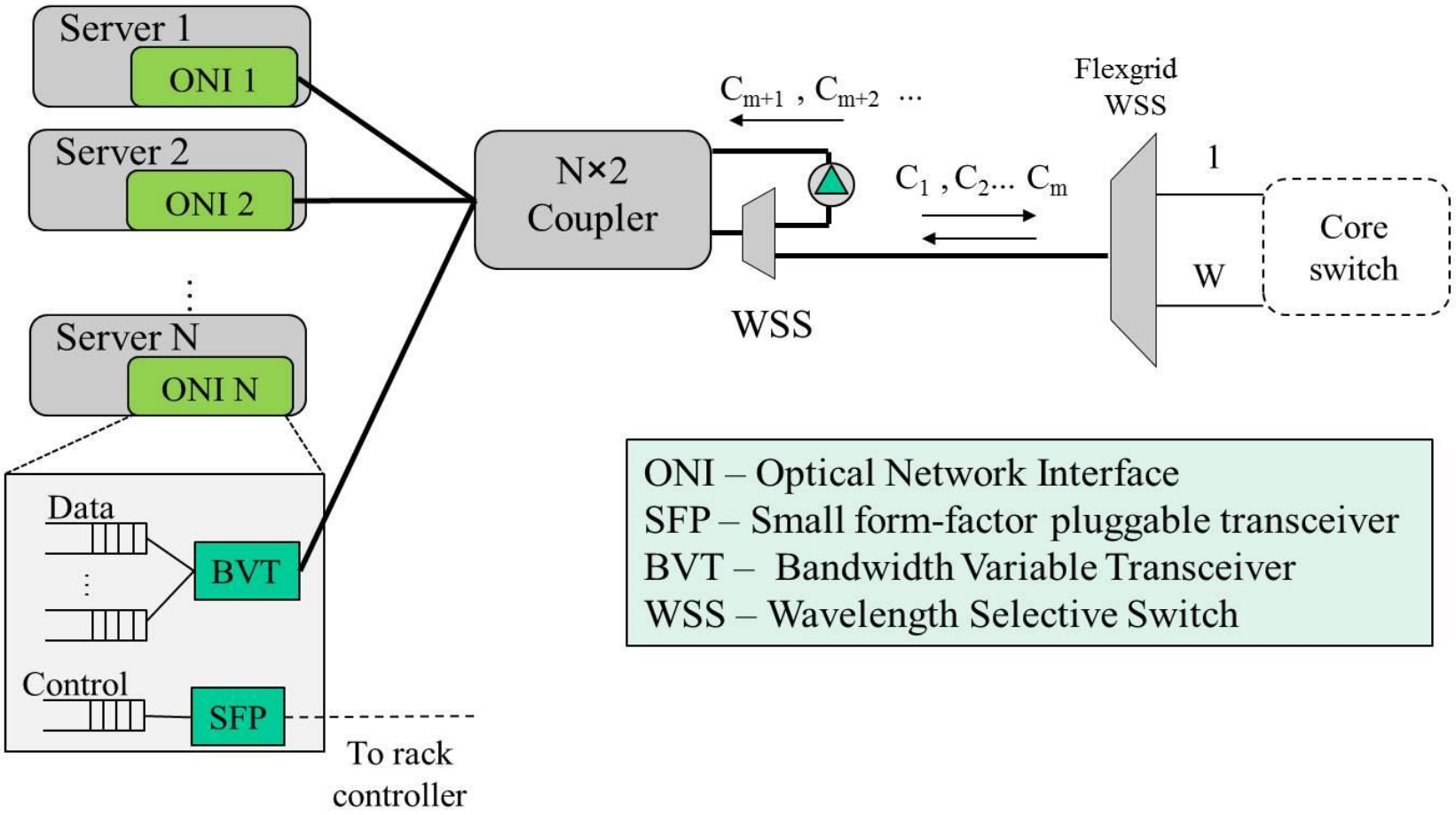
J. Chen, Y. Gong, M. Fiorani, S. Aleksic, "Optical Interconnects at Top of the Rack for Energy-Efficient Datacenters", IEEE Communications Magazine, Vol. 53, Issue 8, pp.140-148, August 2015.

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Passive optical interconnect (1/5)

- Coupler-based passive optical interconnect architecture:



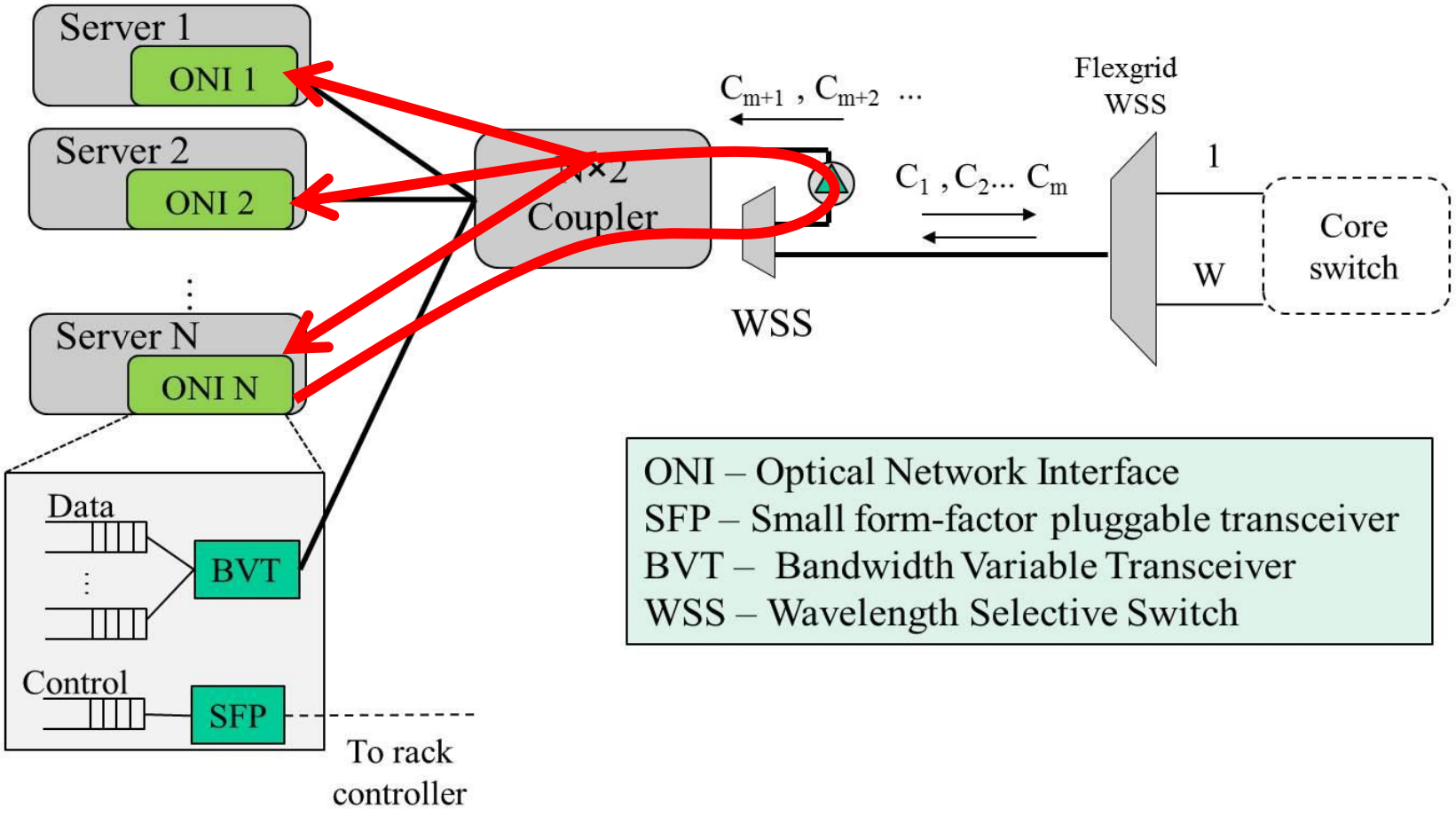
M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen, "Energy-Efficient Elastic Optical Interconnect Architecture for Data Centers", IEEE Communications Letters, Vol. 18, No. 9, pp. 1531-1534, Sept. 2014.

Passive optical interconnect (2/5)

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

- Multipoint-to-multipoint connection for intra-rack communication:
  - time and wavelength allocation performed by the rack controller
  - broadcast and select (native support to multicast and broadcast)

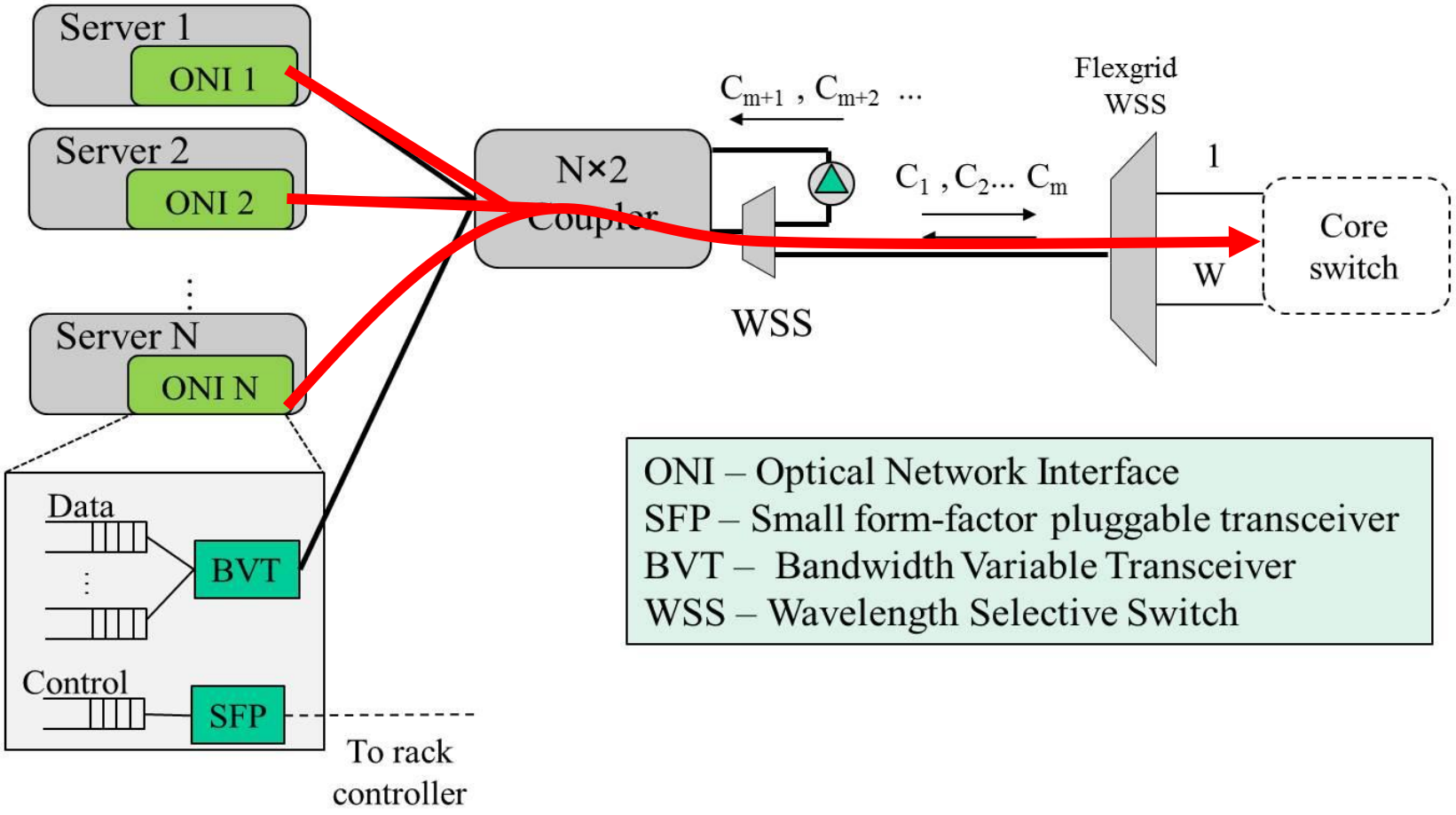


Passive optical interconnect (3/5)

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

- Multipoint-to-point connection for inter-rack communication:
  - time and wavelength allocation performed by the rack controller
  - coordination with other racks required if core switch is all-optical



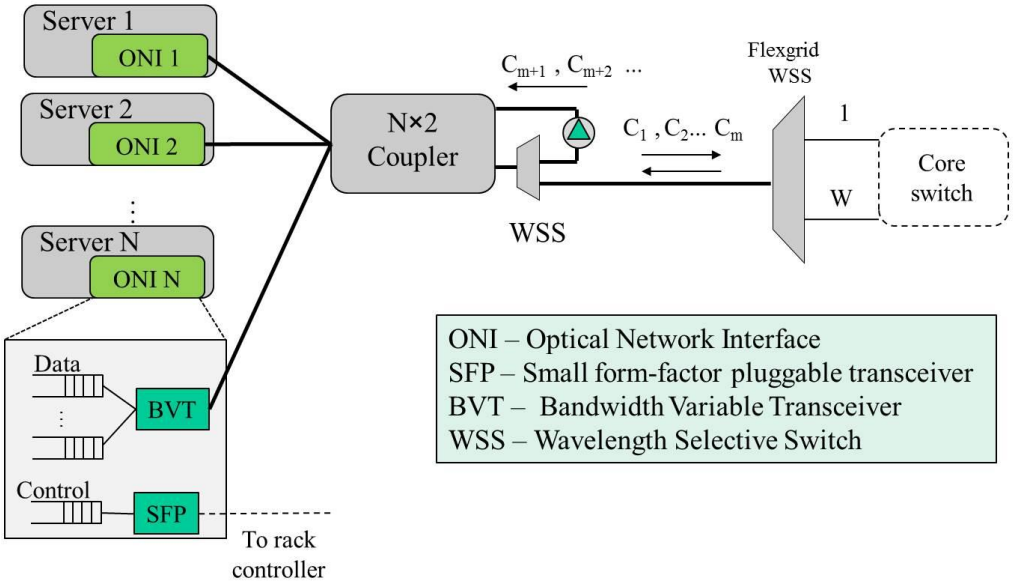
ROYAL INSTITUTE OF TECHNOLOGY

ONLab

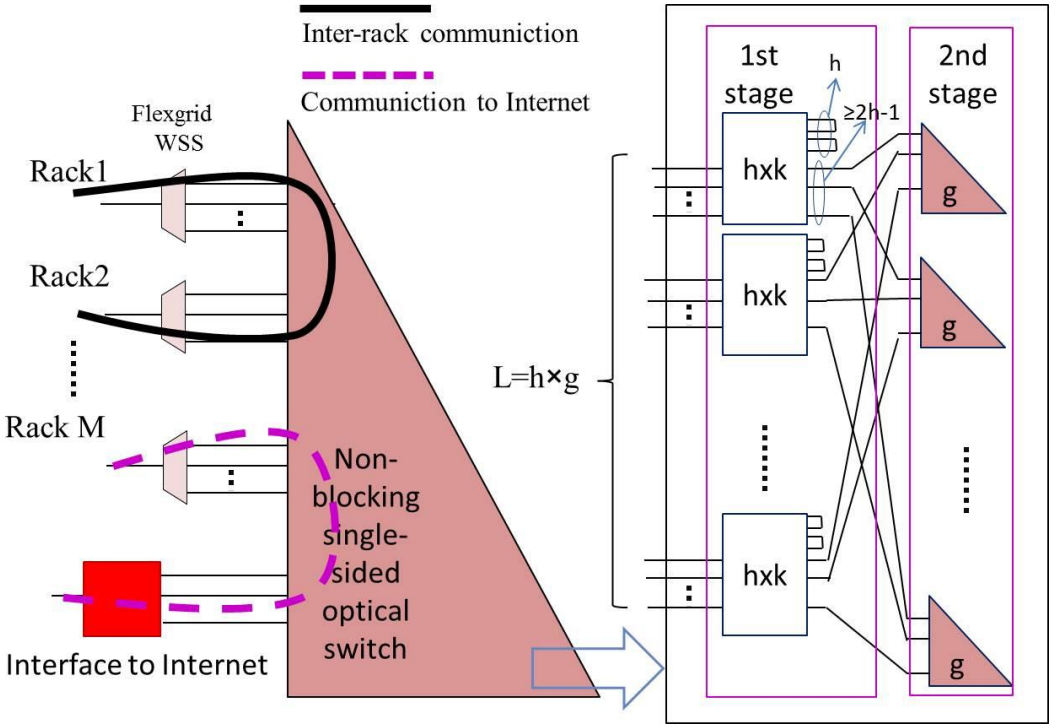
Passive optical interconnect (4/5)

- Elastic optical data center network (EODCN): combining optical interconnect at ToR and elastic optical core switch:

Intra-rack architecture



Inter-rack architecture



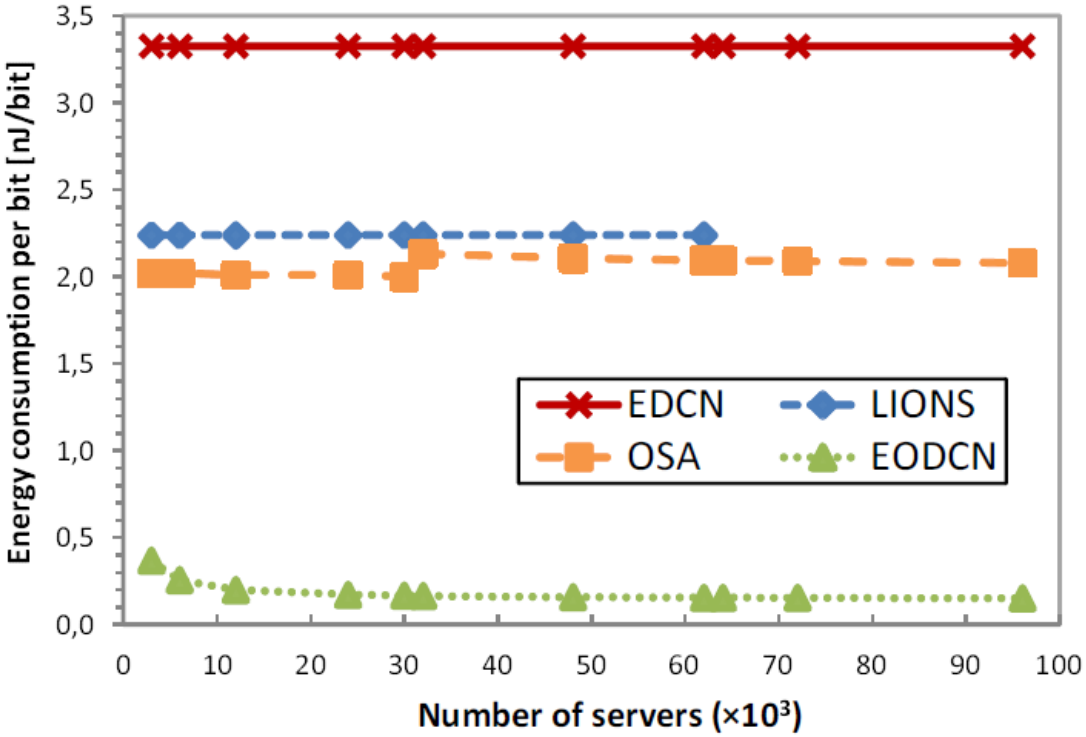
M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen, "Energy-Efficient Elastic Optical Interconnect Architecture for Data Centers", IEEE Communications Letters, Vol. 18, No. 9, pp. 1531-1534, Sept. 2014.

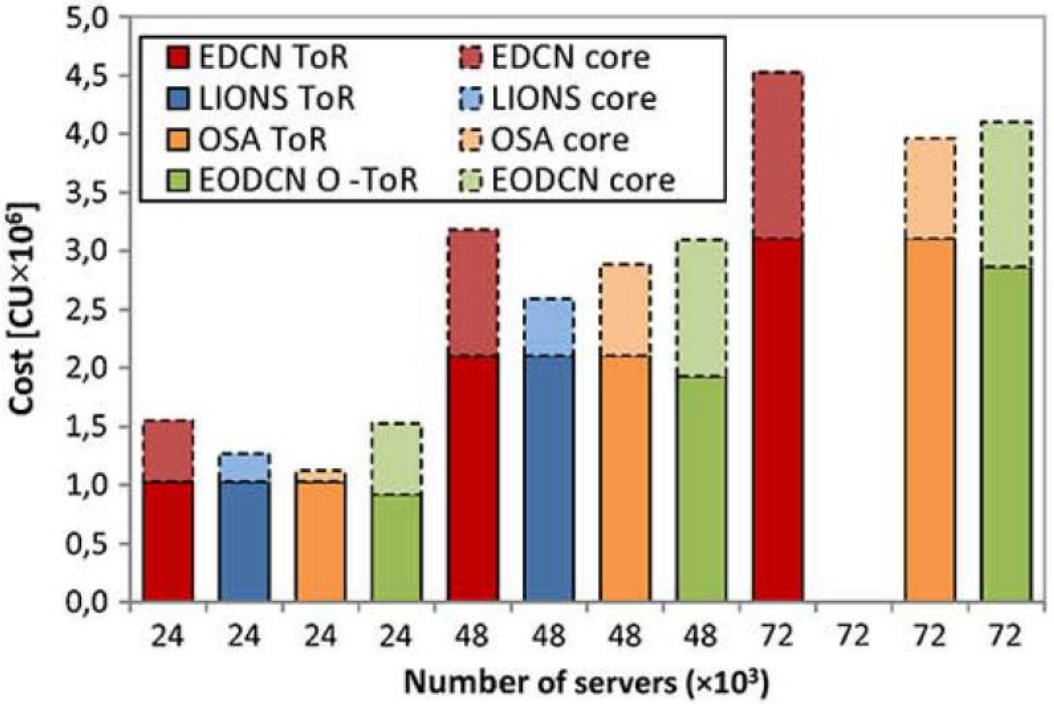
Passive optical interconnect (5/5)

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

- Energy efficiency of EODCN:
  - Reduces energy consumption per bit by a factor 10 with respect to EDCN (factor 5 with respect to OSA and LIONS) - at a similar cost





M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen, "Energy-Efficient Elastic Optical Interconnect Architecture for Data Centers", IEEE Communications Letters, Vol. 18, No. 9, pp. 1531-1534, Sept. 2014.

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Optical Interconnects for Data Centers

- Optical interconnects at Top-of-Rack for energy efficient data centers
  1. J. Chen, Y. Gong, M. Fiorani, S. Aleksic, "Optical Interconnects at Top of the Rack for Energy-Efficient Datacenters", IEEE Communications Magazine, Vol. 53, Issue 8, pp.140-148, August 2015.
  2. M. Fiorani, S. Aleksic, M. Casoni, L. Wosinska, J. Chen, "Energy-Efficient Elastic Optical Interconnect Architecture for Data Centers", IEEE Communications Letters, Vol. 18, No. 9, pp. 1531-1534, Sept. 2014.

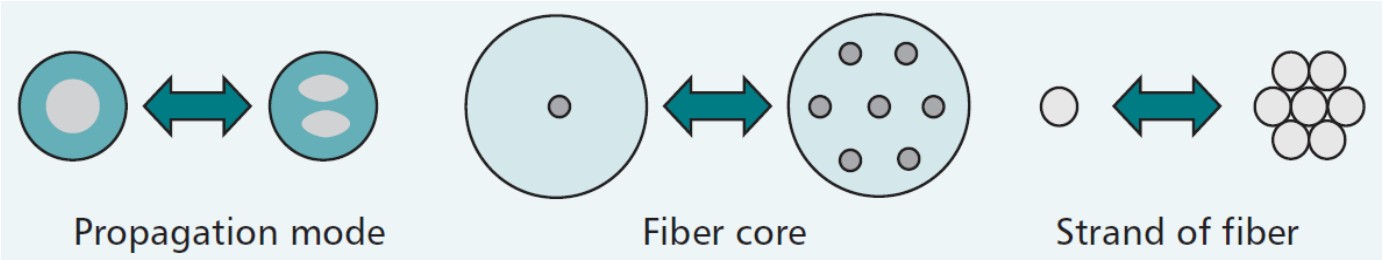
- Spatial division multiplexing (SDM) for ultra-high capacity modular data centers
  1. Now in progress…

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Spatial division multiplexing (1/3)

- Spatial flexibility: controllable arrangement of optical signals in the spatial domain
- SDM options:
  1. Multi-mode (FMF)
  2. Multi-core (MCF)
  3. Multi-element (fiber bundle) (MEF)
- SDM can be broadly categorized into two categories:
  - Uncoupled SDM: the parallel optical channels do not coupled to each other; hence, existing transponders for single-mode fibers can be reused
  - Coupled SDM: the parallel optical channels coupled to each other; thus, multiple-input multiple-output (MIMO) signal processing is required to untangle crosstalk



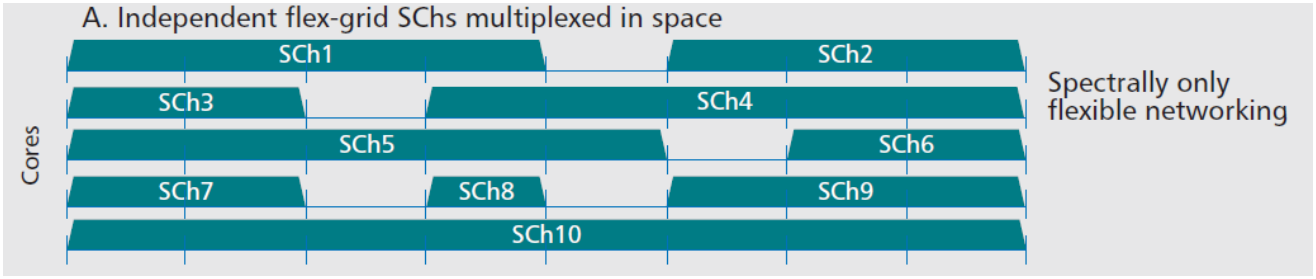
Xia, T.J.; Fevrier, H.; Ting Wang; Morioka, T., "Introduction of spectrally and spatially flexible optical networks," *Communications Magazine, IEEE*, vol.53, no.2, pp.24,33, Feb. 2015

ROYAL INSTITUTE OF TECHNOLOGY

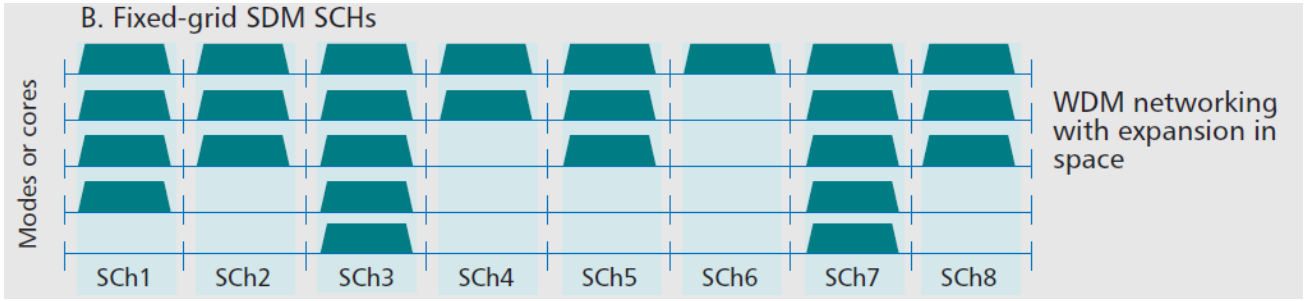
ONLab

Spatial division multiplexing (2/3)

- Combining SDM and WDM – different options:



- Uncoupled SDM / flex-grid WDM
- Elasticity only in spectral domain



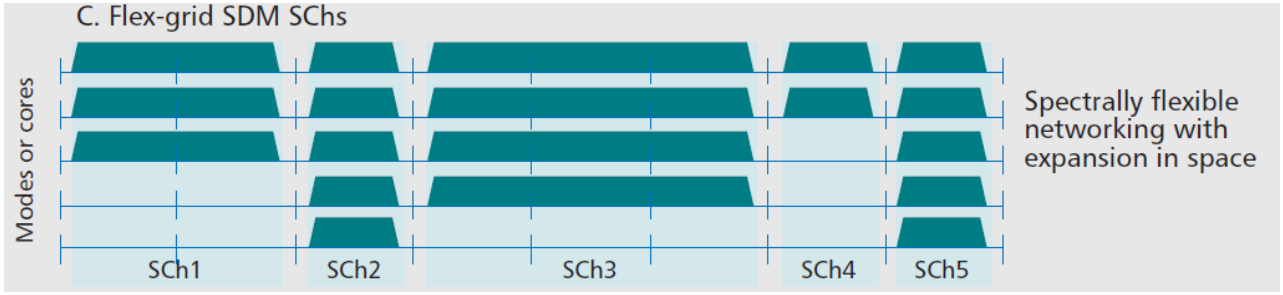
- Coupled SDM / fixed-grid WDM
- Expansion in spatial domain

Klonidis, D.; Cugini, F.; Gerstel, O.; Jinno, M.; Lopez, V.; Palkopoulou, E.; Sekiya, M.; Siracusa, D.; Thouenon, G.; Betoule, C., "Spectrally and spatially flexible optical network planning and operations," Communications Magazine, IEEE , vol.53, no.2, pp.69,78, Feb. 2015

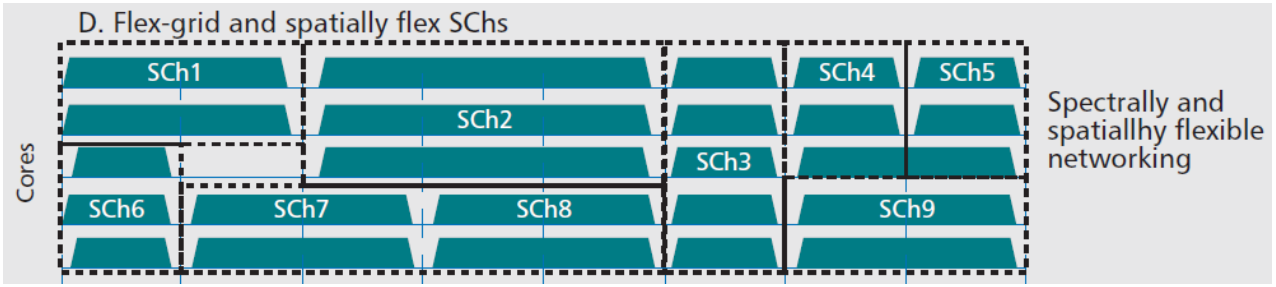
ROYAL INSTITUTE OF TECHNOLOGY

ONLab

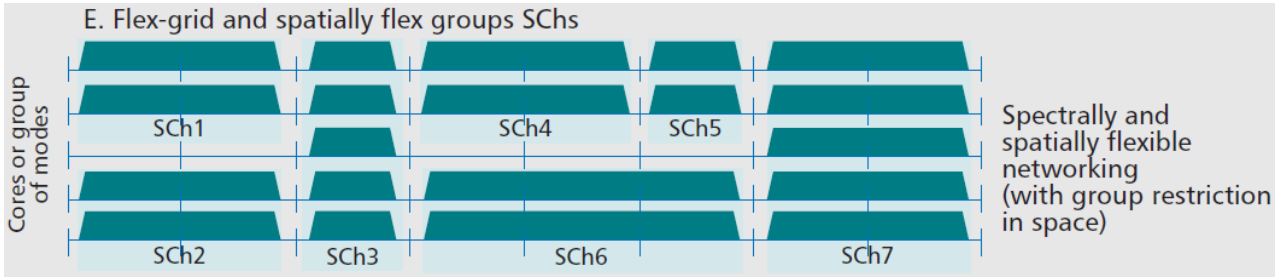
Spatial division multiplexing (3/3)



- Coupled SDM / flex-grid WDM
- Expansion in spatial domain



- Coupled SDM / flex-grid WDM
- Elasticity in spectral and spatial domains



- Coupled SDM / flex-grid WDM
- Elasticity in spectral and spatial domains with group restriction

Klonidis, D.; Cugini, F.; Gerstel, O.; Jinno, M.; Lopez, V.; Palkopoulou, E.; Sekiya, M.; Siracusa, D.; Thouenon, G.; Betoule, C., "Spectrally and spatially flexible optical network planning and operations," Communications Magazine, IEEE , vol.53, no.2, pp.69,78, Feb. 2015

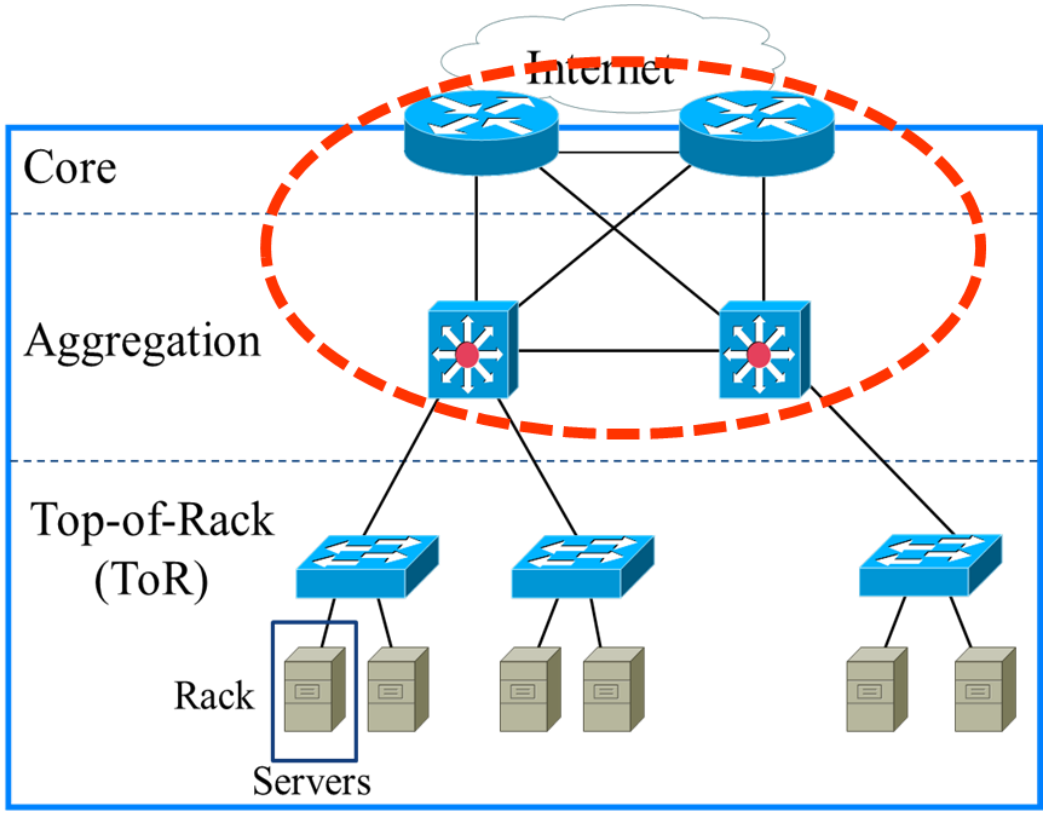
SDM in data center networks

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

- SDM has high potential to be used for core switch in data centers*:
  1. Reduces cabling complexity
  2. Offers high scalability to support:
     1. higher capacity per server
     2. higher # of servers
  3. Short-reach: reduced effect of physical impairments

- The potential of SDM in data center networks has never been explored
- **Idea:**
  - Assess the performance of different SDM options in data centers;
  - Identify best SDM option → best trade-off between cost and performance



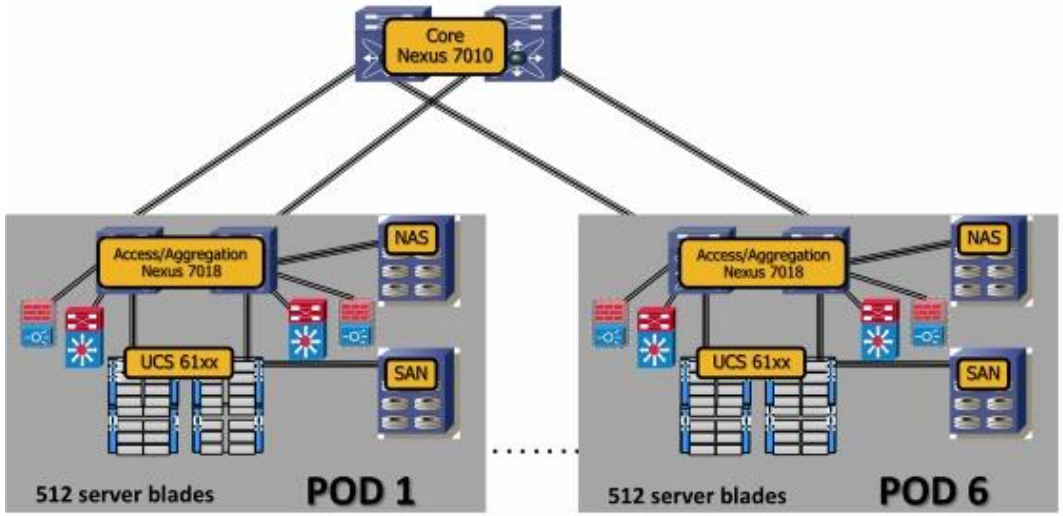
*Shuangyi Yan; Hugues-Salas, E.; Rancano, V.J.F.; Yi Shu; Saridis, G.M.; Rofoee, B.R.; Yan Yan; Peters, A.; Jain, S.; May-Smith, T.; Petropoulos, P.; Richardson, D.J.; Zervas, G.; Simeonidou, D., "Archon: A Function Programmable Optical Interconnect Architecture for Transparent Intra and Inter Data Center SDM/TDM/WDM Networking," Lightwave Technology, Journal of , vol.33, no.8, pp.1586,1595, April, 2015

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Reference topology (1/2)

- The capacity generated by a single ToR is not enough to justify the use of a SDM fiber:
  - ToR: 96 servers & 100 Gb/s = 9.6 Tb/s (max)
  - SDM fiber: 320 slots & 10 cores = 200 Tb/s (max) [16 slots → 1 Tb/s]*
- SDM fits better to modular data centers based on the POD concept:
  - POD: is a set of defined compute, network and storage resources. It includes several ToRs connected using aggregation switches**



*O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" IEEE Commun. Mag., vol. 50, no. 2, pp. s12–s20, Feb. 2012.

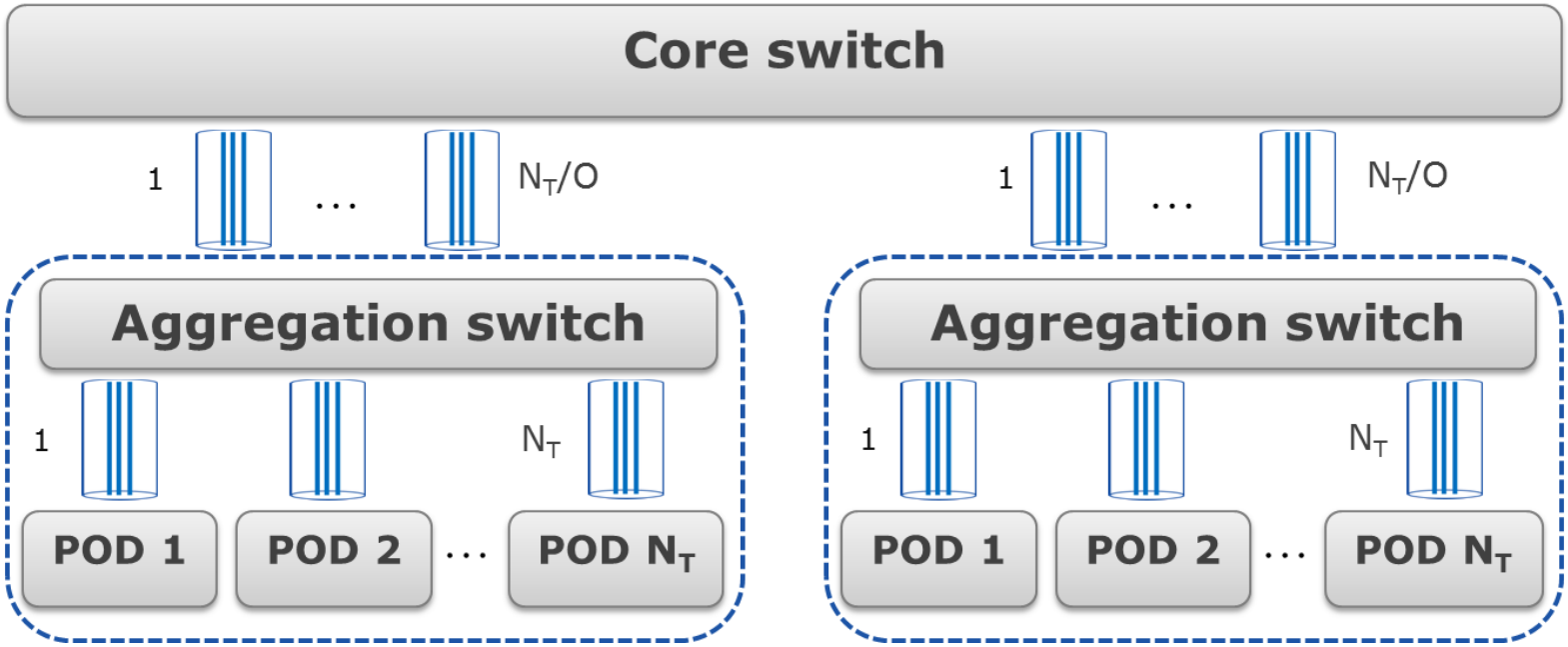
**Data Center Top-of-Rack Architecture Design, Cisco, white paper.

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Reference topology (2/2)

- Example of modular data center:
  - POD: 20 racks & 96 servers & 100 Gb/s = 192 Tb/s (max)
  - SDM fiber: 320 slots & 10 cores = 200 Tb/s (max) [16 slots → 1 Tb/s]*
  - 250 POD = 480,000 servers**
- Two levels switch topology:



O = overprovisioning factor. Depends on the ratio between intra- and inter-cluster traffic

*O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" IEEE Commun. Mag., vol. 50, no. 2, pp. s12–s20, Feb. 2012.

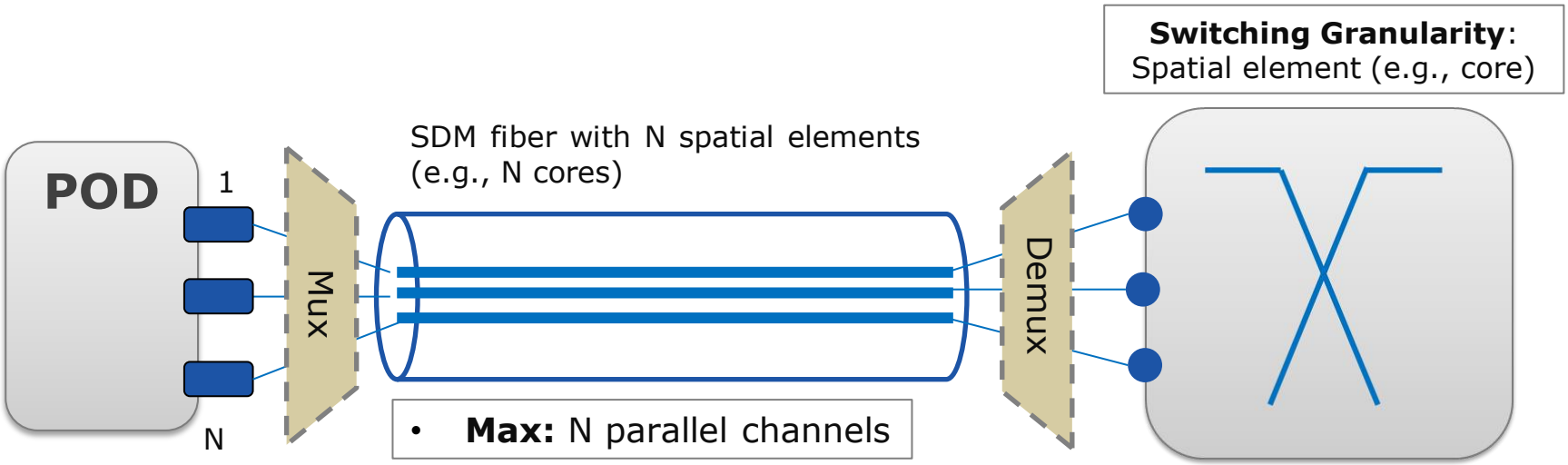
**http://www.datacenterknowledge.com/archives/2013/05/24/microsoft-will-back-xbox-one-300000-servers/

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Architecture 1

- **Uncoupled SDM & no WDM*:**



Transceivers:
- No tunable
- No flexgrid
- No MIMO
- N transceivers per POD

Optical switch:
- 1 port per spatial element
- N ports per fiber
- N ports per POD

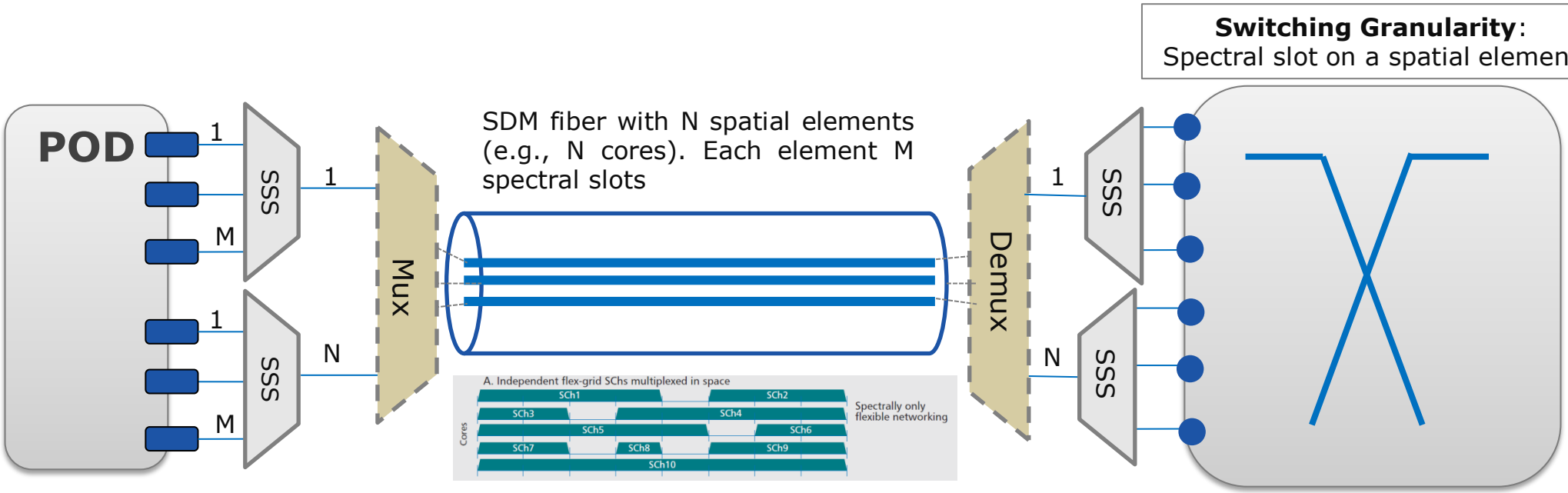
Shuangyi Yan; Hugues-Salas, E.; Rancano, V.J.F.; Yi Shu; Saridis, G.M.; Rofoee, B.R.; Yan Yan; Peters, A.; Jain, S.; May-Smith, T.; Petropoulos, P.; Richardson, D.J.; Zervas, G.; Simeonidou, D., "Archon: A Function Programmable Optical Interconnect Architecture for Transparent Intra and Inter Data Center SDM/TDM/WDM Networking," Lightwave Technology, Journal of , vol.33, no.8, pp.1586,1595, April, 2015

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Architecture 2

- **Uncoupled SDM & flex-grid WDM:**



- **Max:** N x M parallel channels

Transceivers:
- Tunable
- Flexgrid
- No MIMO
- N x M transceivers per POD

Mux and Demux:
- 2 x N SSS per ToR

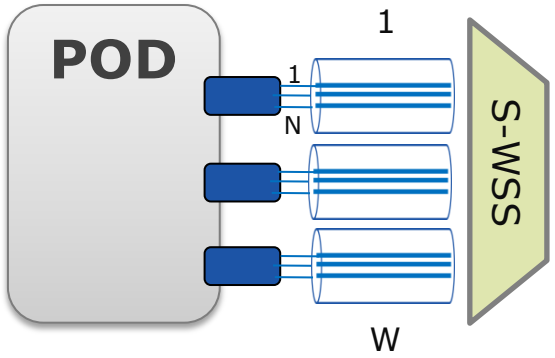
Optical switch:
- 1 port per spatial element per spectral slot
- N x M ports per fiber
- N x M ports per ToR

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

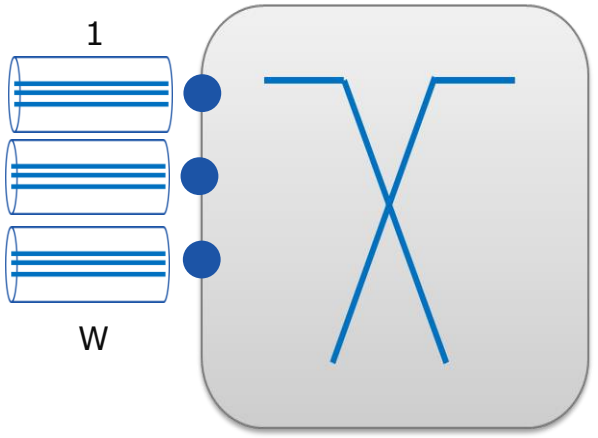
Architecture 3

- **Coupled SDM & fixed-grid WDM** – expansion in spatial domain:



SDM fiber with N spatial elements (e.g., N cores). Each element W (<M) wavelengths

Switching Granularity: wavelength



Transceivers:
- Tunable
- No flexgrid
- MIMO
- W transceivers per POD

Mux and Demux:
- 2 S-WSS per POD

- **Max:** W (<M) parallel channels

Optical switch:
- 1 port per wavelength
- W ports per fiber
- W ports per POD

L. E. Nelson, M. D. Feuer, K. Abedin, X. Zhou, T. F. Taunay, J. M. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, and D. M. Marom, "Spatial Superchannel Routing in a Two-Span ROADM System for Space Division Multiplexing," J. Lightwave Technol. 32, 783-789 (2014)

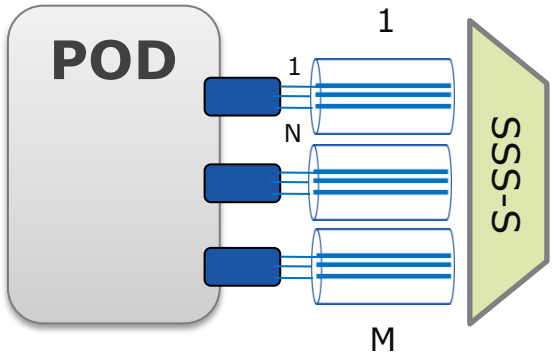
Ryf, R.; Chandrasekhar, S.; Randel, S.; Neilson, D.T.; Fontaine, N.K.; Feuer, M., "Physical layer transmission and switching solutions in support of spectrally and spatially flexible optical networks," Communications Magazine, IEEE , vol.53, no.2, pp.52,59, Feb. 2015

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Architecture 4

- **Coupled SDM & flex-grid WDM** – expansion in spatial domain:



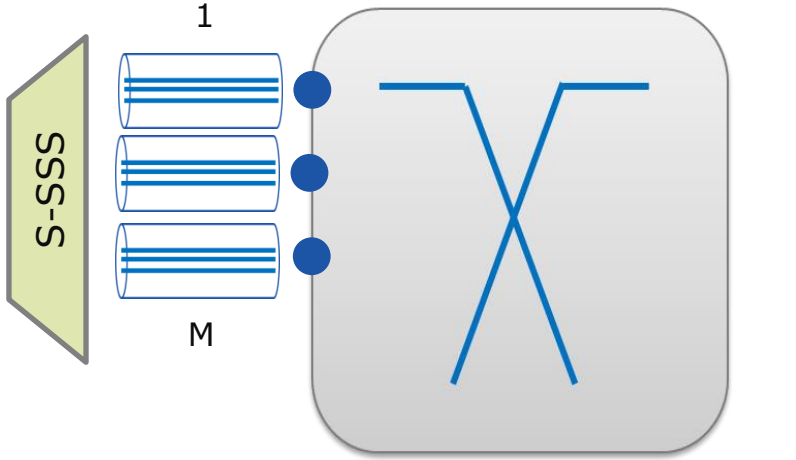
Transceivers:
- Tunable
- Flexgrid
- MIMO
- M transceivers per POD

Mux and Demux:
- 2 S-SSS per POD

SDM fiber with N spatial elements (e.g., N cores). Each element M spectral slots

- **Max:** W (<M) parallel channels

Switching Granularity: spectral slot



Optical switch:
- 1 port per spectral slot
- M ports per fiber
- M ports per POD

L. E. Nelson, M. D. Feuer, K. Abedin, X. Zhou, T. F. Taunay, J. M. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, and D. M. Marom, "Spatial Superchannel Routing in a Two-Span ROADM System for Space Division Multiplexing," J. Lightwave Technol. 32, 783-789 (2014)

Ryf, R.; Chandrasekhar, S.; Randel, S.; Neilson, D.T.; Fontaine, N.K.; Feuer, M., "Physical layer transmission and switching solutions in support of spectrally and spatially flexible optical networks," Communications Magazine, IEEE , vol.53, no.2, pp.52,59, Feb. 2015

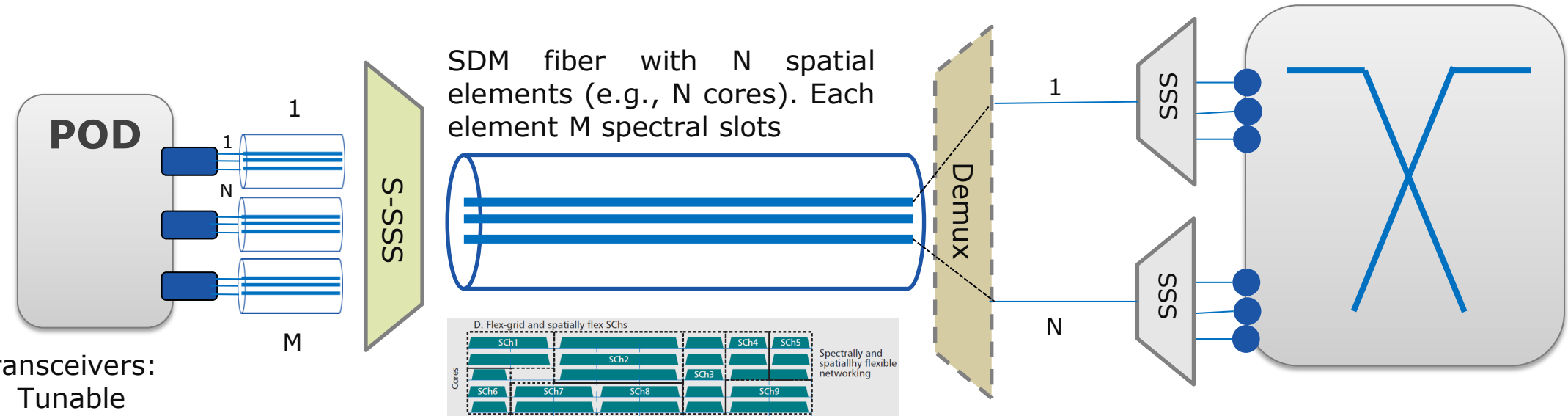
ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Architecture 5

- **Coupled SDM & flex-grid WDM** – spectral and spatial flexibility:

Switching Granularity: Spectral slot on a spatial element



Transceivers:
- Tunable
- Flexgrid
- MIMO
- M transceivers per POD

Mux and Demux:
- 1 S-SSS per POD
- N SSS per POD

- **Max:** N x M parallel channels

Optical switch:
- 1 port per spatial element and per spectral slot
- M x N ports per fiber
- M x N ports per POD

L. E. Nelson, M. D. Feuer, K. Abedin, X. Zhou, T. F. Taunay, J. M. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, and D. M. Marom, "Spatial Superchannel Routing in a Two-Span ROADM System for Space Division Multiplexing," J. Lightwave Technol. 32, 783-789 (2014)

Ryf, R.; Chandrasekhar, S.; Randel, S.; Neilson, D.T.; Fontaine, N.K.; Feuer, M., "Physical layer transmission and switching solutions in support of spectrally and spatially flexible optical networks," Communications Magazine, IEEE , vol.53, no.2, pp.52,59, Feb. 2015

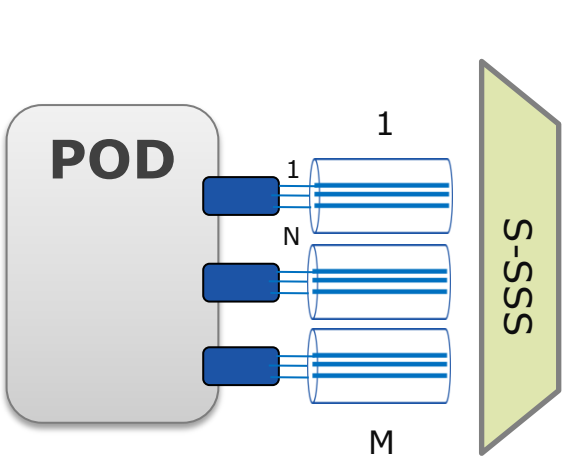
ROYAL INSTITUTE OF TECHNOLOGY

ONLab

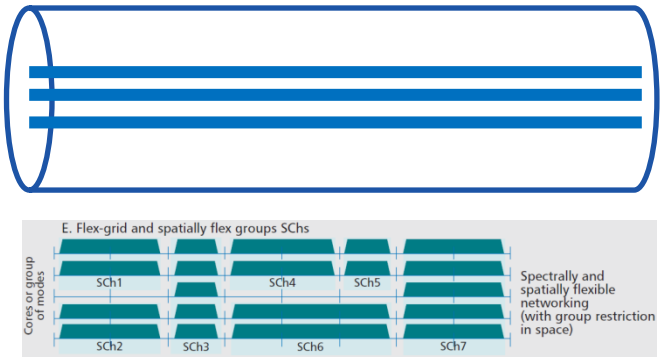
Architecture 6

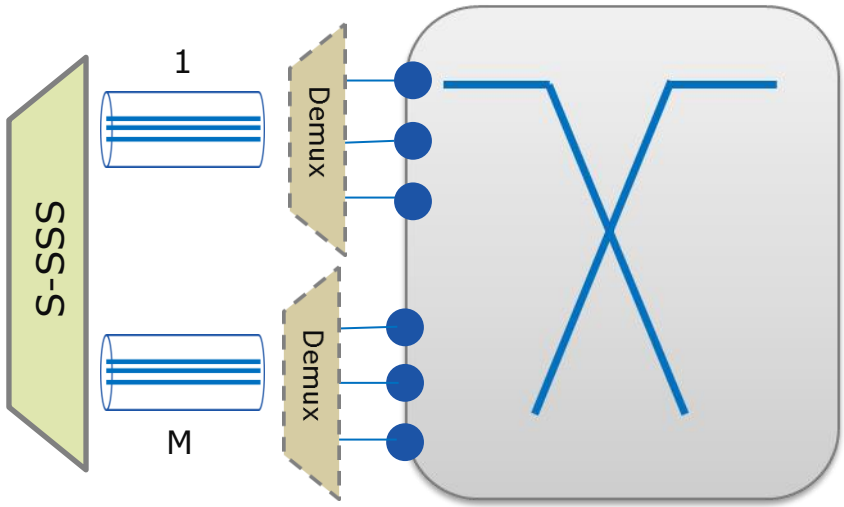
- **Coupled SDM & flex-grid WDM** – spectral and spatial flexibility w group restriction:

Switching Granularity: Spectral slot on a spatial element



SDM fiber with N spatial elements (e.g., N cores). Each element M spectral slots





Transceivers:
- Tunable
- Flexgrid
- MIMO
- M transceivers per POD

Mux and Demux:
- 2 S-SSS per POD

- **Max:** N x M parallel channels

Optical switch:
- 1 port per spatial element and per spectral slot
- M x N ports per fiber
- M x N ports per POD

L. E. Nelson, M. D. Feuer, K. Abedin, X. Zhou, T. F. Taunay, J. M. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, and D. M. Marom, "Spatial Superchannel Routing in a Two-Span ROADM System for Space Division Multiplexing," J. Lightwave Technol. 32, 783-789 (2014)

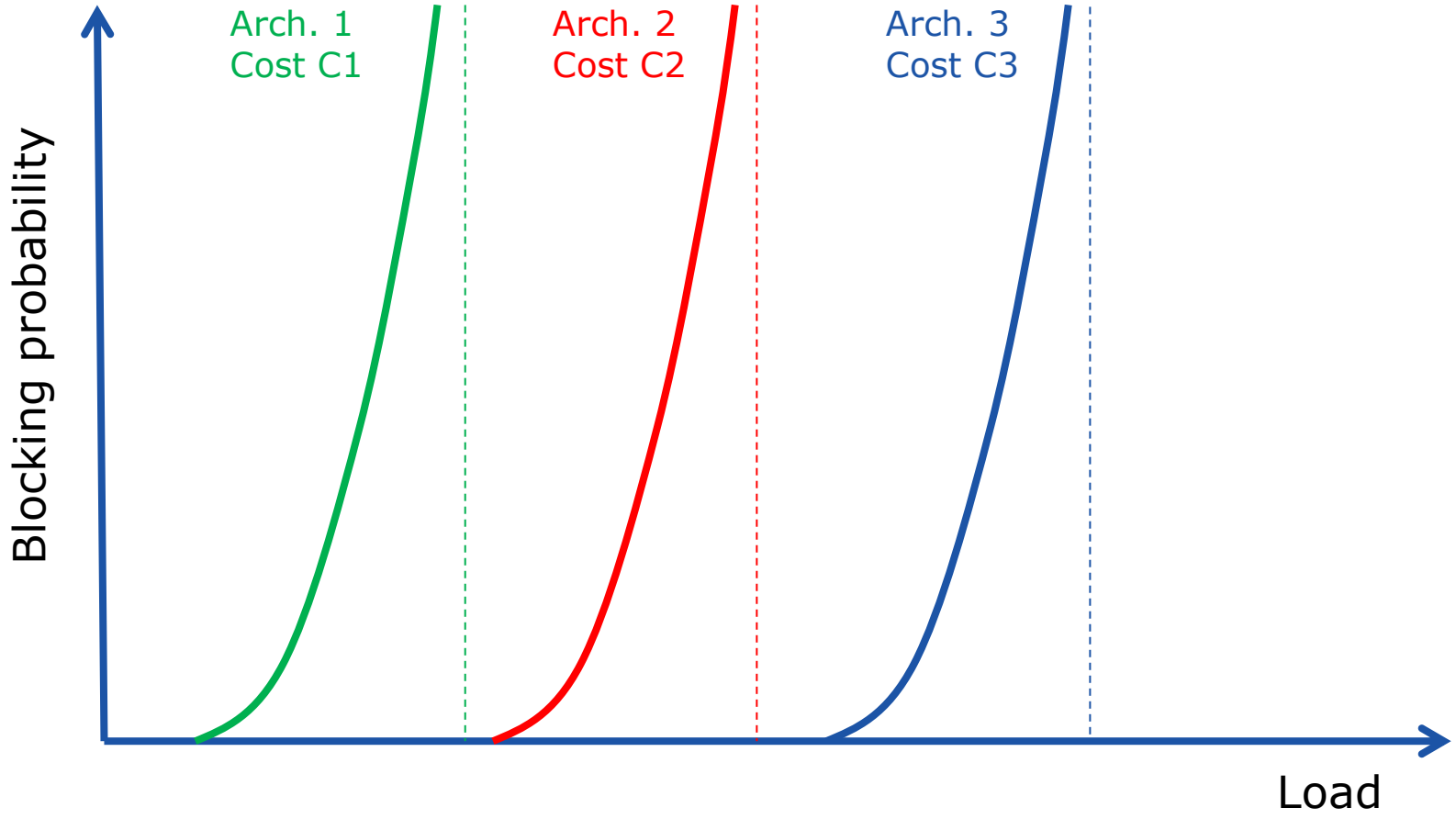
Ryf, R.; Chandrasekhar, S.; Randel, S.; Neilson, D.T.; Fontaine, N.K.; Feuer, M., "Physical layer transmission and switching solutions in support of spectrally and spatially flexible optical networks," Communications Magazine, IEEE , vol.53, no.2, pp.52,59, Feb. 2015

ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Expected results (example)

- Cost vs. performance of the different SDM architectures:

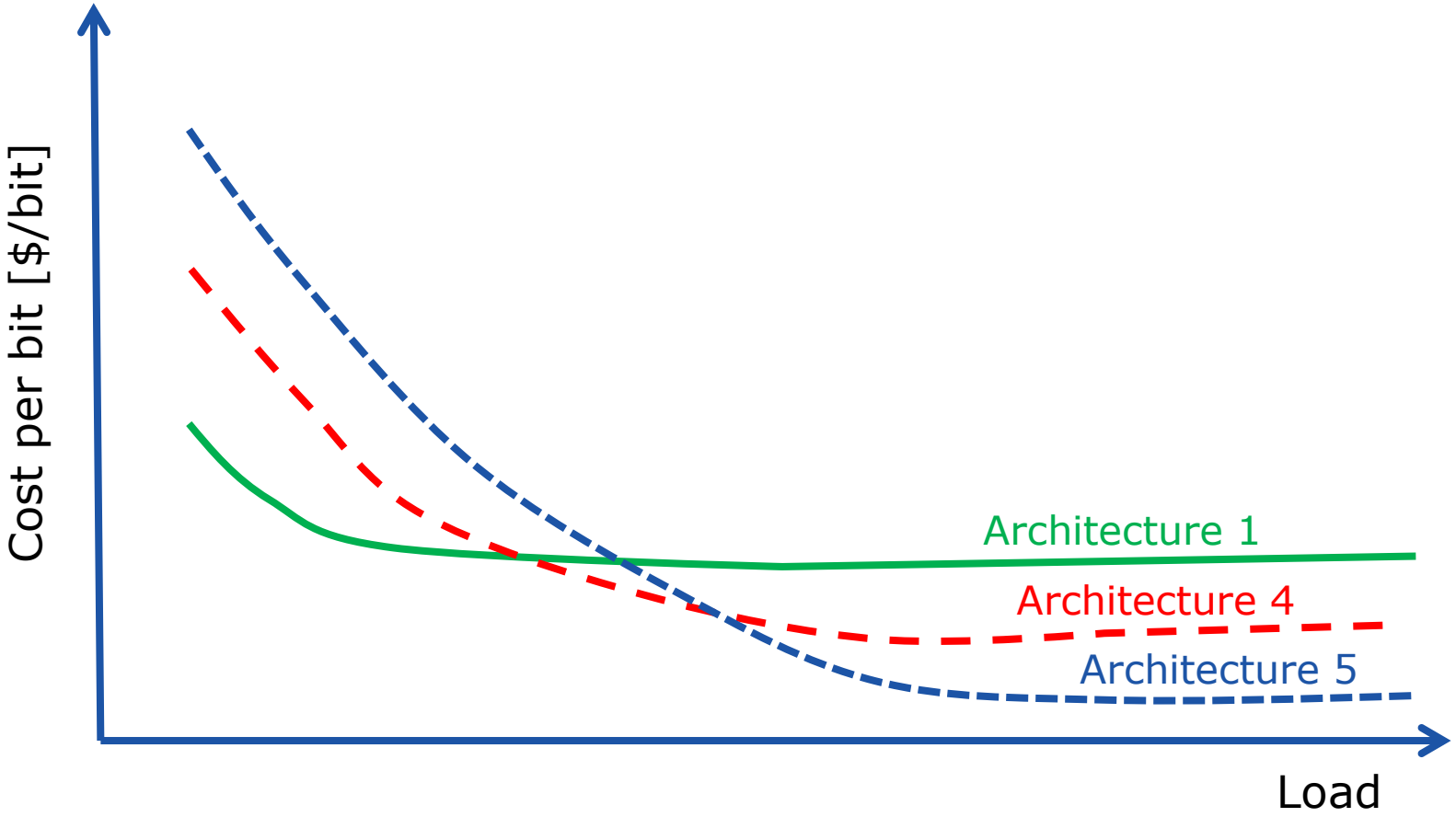


ROYAL INSTITUTE OF TECHNOLOGY

ONLab

Expected results (example)

- Cost vs. performance of the different SDM architectures:



ROYAL INSTITUTE
OF TECHNOLOGY

ONLab

Thank you for your attention!

Matteo Fiorani

Email: fiorani@kth.se

Optical Networks Lab (ONLab)
Communication Systems Department (COS)
KTH Royal Institute of Technology
Stockholm (Sweden)