Science Elastic Optical Inter-Network (SEOIN)

Malathi Veeraraghavan and Xiaoyu (Sherry) Wang

University of Virginia (UVA)

mvee@Virginia.edu

July 19, 2018

Mini-Symposium on Data over Distance:
Convergence of Networking, Storage, Transport, and Software Frameworks

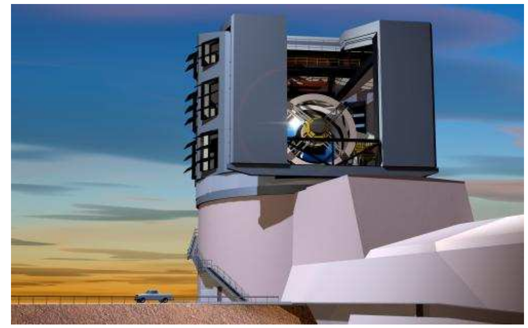
Thanks to NSF 1405171 and 1624676

Outline

- Application drivers
- Computing landscape
- Advances in optical networking
- Research challenges
- Summary: Recommendations

Application drivers for moving data over distance

- End devices used in scientific research:
  - Computers, instruments (telescopes, light sources)
- Transfers between data centers: GridFTP, Aspera's FASP
- Stream processing with online steering: reduced time when compared with sequential data acq. + processing + re-runs
  - Synchrotron Light Source Experiments reported by Argonne National Lab (APS to ALCF analysis cluster - 126 nodes) [1]
  - Center for Advanced Mathematics for Energy Research Applications (CAMERA) - Berkeley Lab



Large Synoptic Survey Telescope (LSST)
https://www.lsst.org/about

Advanced Photon Source (APS)
at Argonne Natl. Lab

[1] T. Bicer, D. Gursoy, R. Kettimuthu, Ian Foster, B. Ren, V.De Andrede, and F. De Carlo, " Real-Time Data Analysis and Autonomous Steering of Synchrotron Light Source Experiments," The 13th IEEE Intl. Conf. on eScience, Oct. 2017

Computing landscape

- Supercomputing centers: Leadership computing facilities and commercial clouds
- Edge clouds
  - To reduce latency by a factor of 10, need to increase number of datacenters by a factor of 100 [1]
  - Is latency important for streaming with control feedback loops?
  - Edge clouds at sites with experimental facilities
- Need for wide-area networking?

[1] Rick McGeer, US-Ignite

Two models of computing: HPC vs. HTC

High Performance Computing (HPC), e.g., ORNL's Summit

	Tightly coupled	Loosely coupled
***Typical* science domains**	Physical Sciences Physics, Weather, Astrophysics Computational Fluid Dynamics (CFD), Break up the problem into cells that must communicate at each time-step	Life Sciences & Statistics Genomics, , Random sampling, Some "Big-Data" Analytics Break problems into computationally-independent 'chunks' and communicate to aggregate results.
Architecture (nodes)	Homogeneous Older (slower) nodes slow down the entire calculation	Heterogeneous older (slower) nodes still contribute
Network Fabric	Expensive Need low-latency, specialized network (i.e. infiniband, 56Gbps)	Inexpensive Don't need low-latency. ommodity network (Ethernet, 0GGbps/40Gbps) is good enough
Refresh	Impulsive Must refresh the entire machine every 4-5 years. A tightly coupled calculation is only as fast as its slowest node	Continuous Incremental refresh is allowed. Add new nodes and remove old nodes as needed. An old node still speeds up a calculation.
Data Center: Space & Power	**2N** Need enough space and power for two complete machines. Old machine can't be decommissioned until its replacement is in full production.	**N + δ** Only need enough space and power for a single machine (plus a modest delta). Refresh is performed "in place".

High Throughput Computing (HTC), e.g., OSG

F. Pineda, Business Model and Governance for Life Science HPC, Presentation at UVA, July 14, 2016

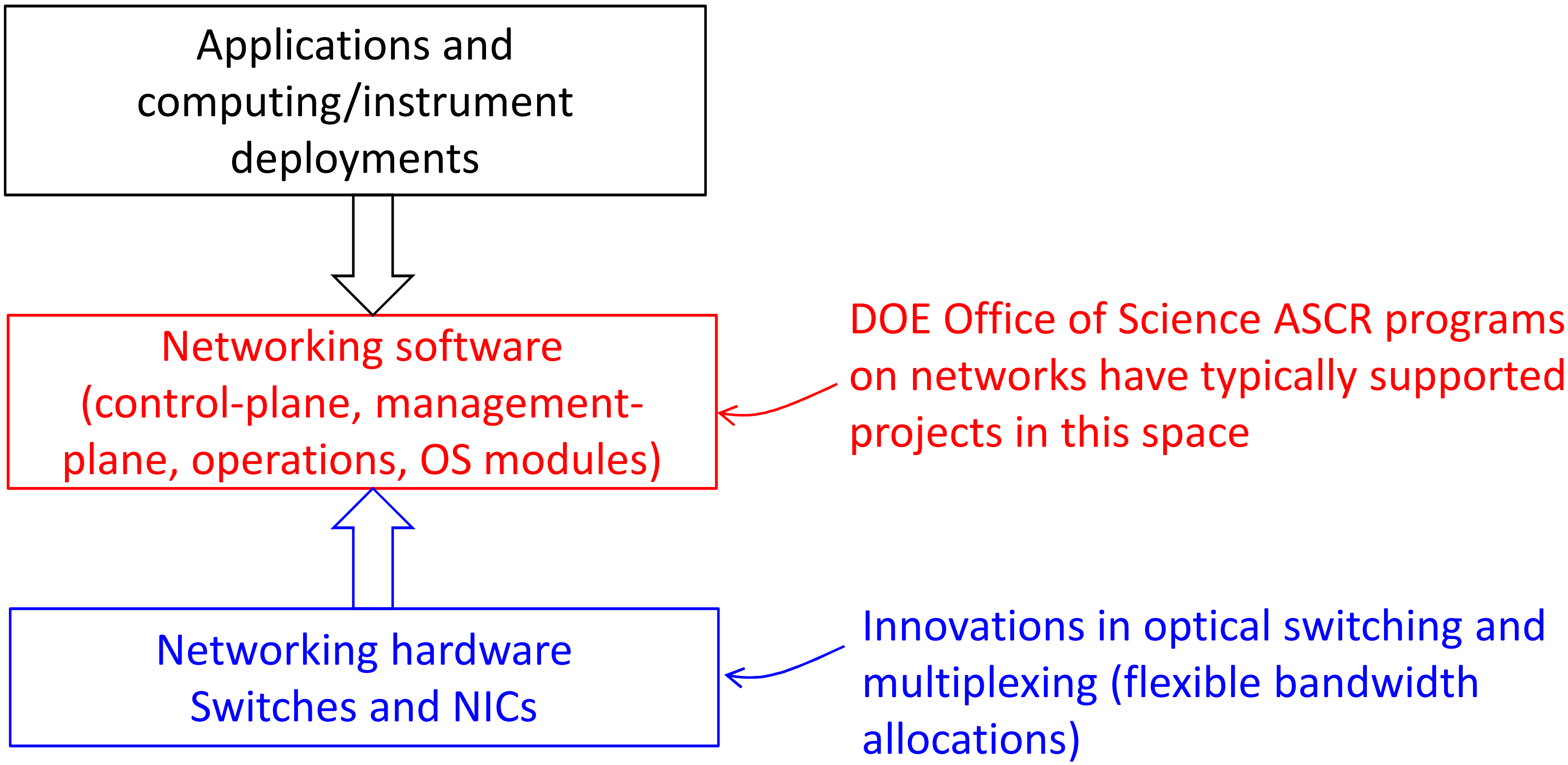
Funding: HPC vs. HTC

	Tightly coupled	Loosely coupled
Cash-flow for capital equipment	Big pulses Large capital outlay every 4-5 years	Slow drip Small continual capital outlay directly driven by individual PI grants & needs
Grant Funding	Need 'Home runs' from deep pocket sponsors, e.g. NSF MRI, NIH S10, the State or the Dean. Small number of big grants. PIs experience "double jeopardy". One competition for the grant and another competition for the computing allocation.	Only need 'base hits' Each PI writes into their grants the computing resources needed to perform their research.

F. Pineda, Business Model and Governance for Life Science HPC, Presentation at UVA, July 14, 2016

- Which of these models require WAN communications (data over distance)?

Networking innovation enablers and constraints



Outline

- Application drivers
- Computing landscape
- ➢ Advances in optical networking
- Research challenges
- Summary: Recommendations

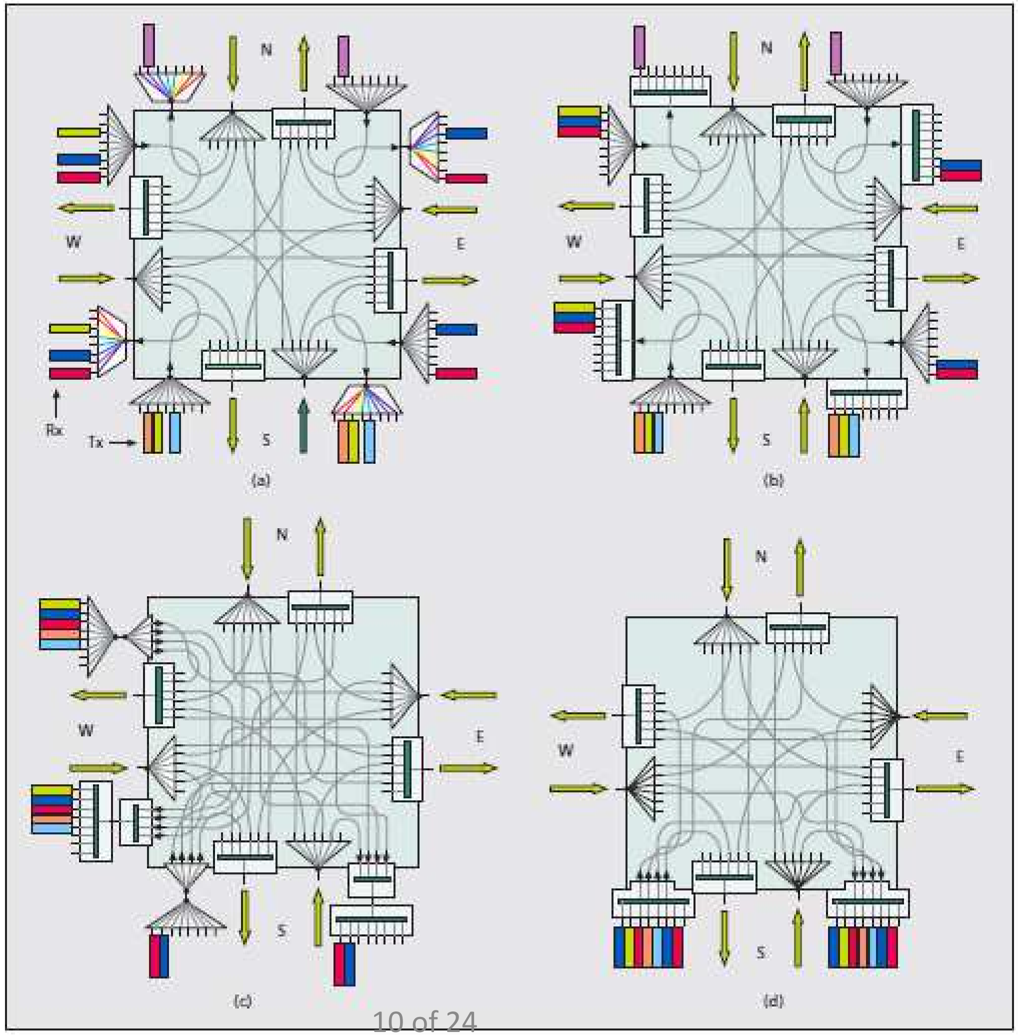
Innovations in optical multiplexing and switching

- Colorless, Directionless, Contentionless (CDC) ROADMs (Wavelength splitters (AWG), WSS, photonic switch
- FlexiGrid/Elastic Optical Networks
  - FlexiGrid CDC-ROADMs
  - Elastic-rate (bandwidth variable) transponders/transceivers
  - Coherent detection
  - Optical Frequency Combs (OFCs)
- Ethernet-to-EON Interface (E-EON)

ROADM: Reconfigurable Optical Add/Drop Multiplexer
AWG: Arrayed Wave Guide

Flexible ROADMs

Basic ROADM



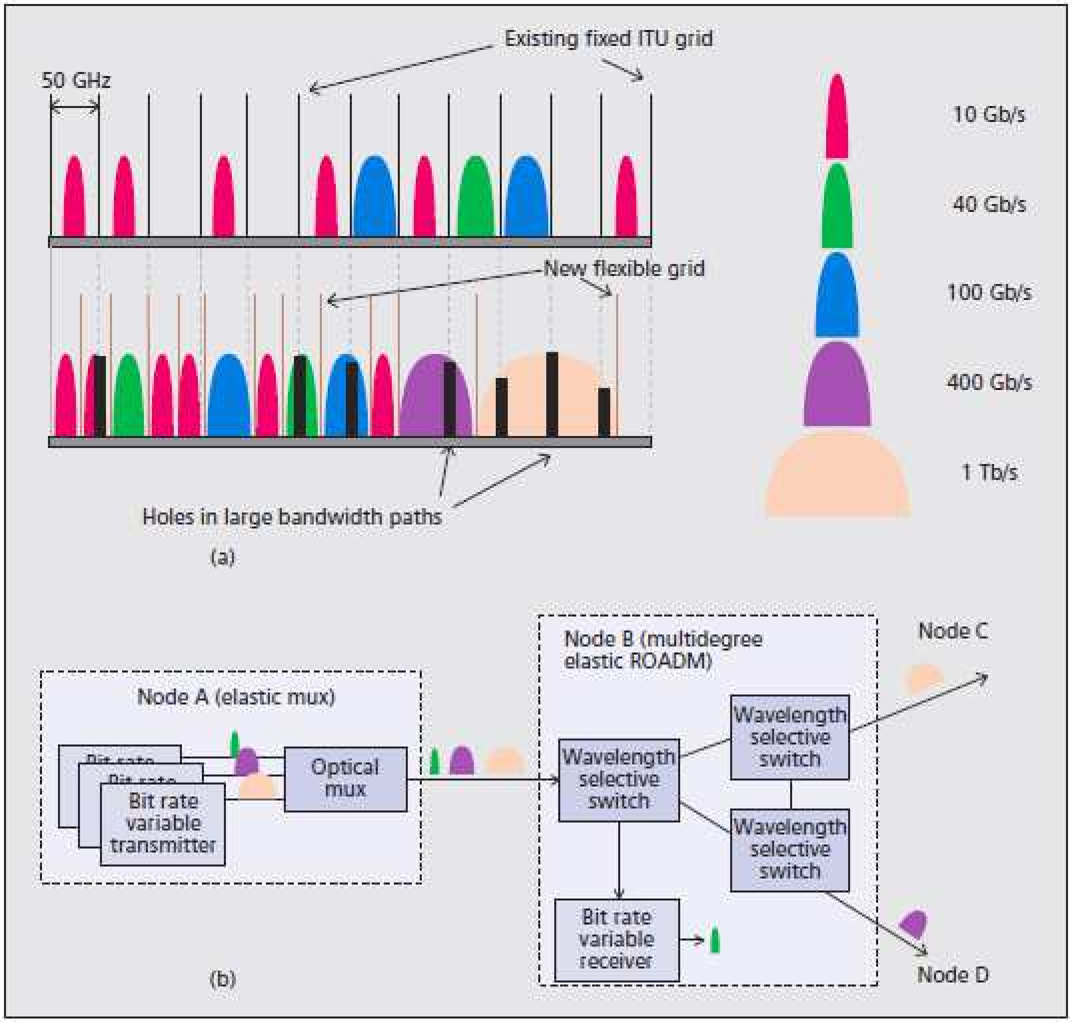
Colorless ROADM

S. Gringeri, B. Basch, V. Shukla, R. Egorov, and T. Xia, "Flexible architectures for optical transport nodes and networks," IEEE Communications Magazine, 2010.

Colorless Directionless ROADM

Colorless Directionless Contentionless ROADM

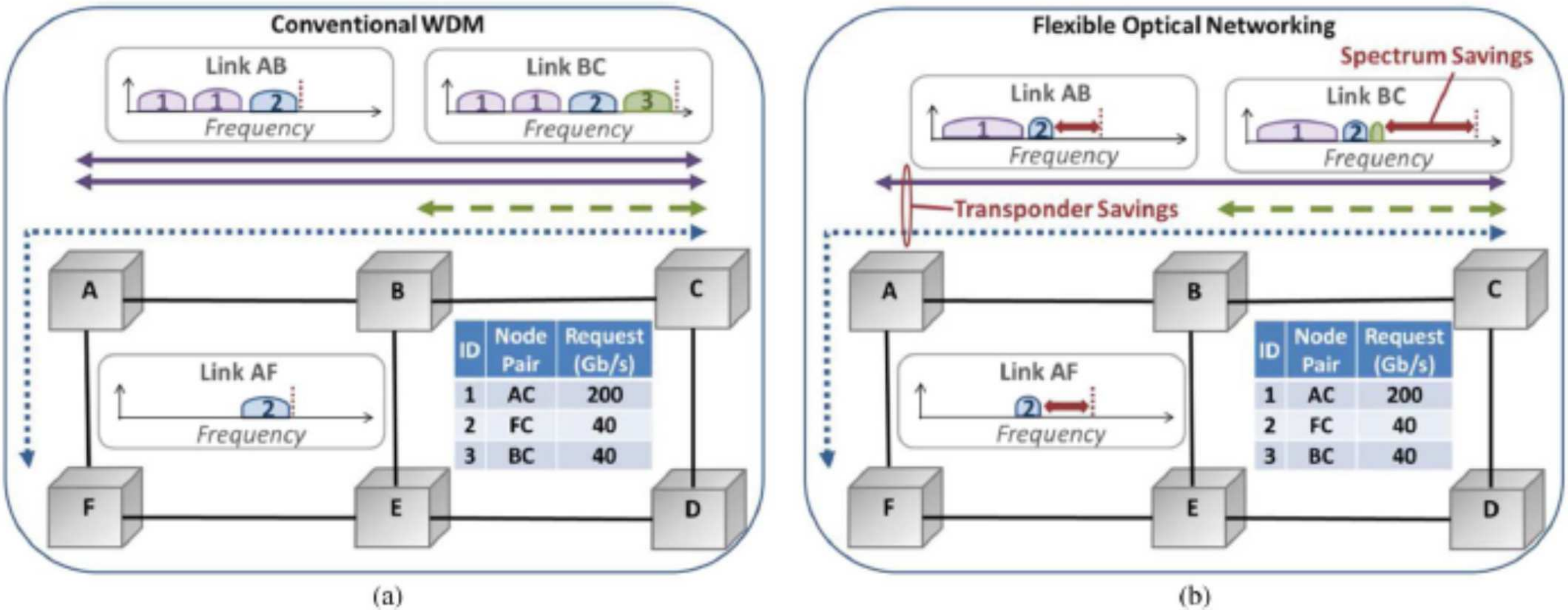
FlexiGrid and Elastic Optical Networks (EONs)



- FlexiGrid ROADM
- Bandwidth Variable Transceivers (BVTs)
- Advantages of EONs
  - support for 400 Gbps, 1Tb/s and higher bit rates
  - ability to support circuits with different bit rates
  - smaller channel spacing (coherent detection is used)
  - ability to tradeoff reach vs. spectral efficiency
  - dynamic networking

O. Gerstel, M. Jinno, A. Lord, and S. Yoo, “Elastic optical networking: a new dawn for the optical layer?” Communications Magazine, IEEE, vol. 50, no. 2, pp. s12–s20, February 2012.

Spectrum and transponder savings with EONs



I. Tomkos, S. Azodolmolky, J. Solé-Pareta, D. Careglio and E. Palkopoulou, "A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges," in Proceedings of the IEEE, vol. 102, no. 9, pp. 1317-1337, Sept. 2014.

Increased flexibility offered by EON

Potential Degrees of Flexibility	Connection Level Metrics		
	Effective Capacity	Transparent Reach	Spectral Efficiency
Modulation Format	⇑	⇓	⇑
Symbol Rate	⇑	⇓	⇑
Ratio of FEC and Payload	⇓	⇑	⇓
Number of Subcarriers	⇑	⇓	⇑
Inter-Subcarrier Spacing within a Superchannel	—	⇑	⇓
Inter-Superchannel Spacing (Guard-band)	—	⇑	⇓

I. Tomkos, S. Azodolmolky, J. Solé-Pareta, D. Careglio and E. Palkopoulou, "A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges," in Proceedings of the IEEE, vol. 102, no. 9, pp. 1317-1337, Sept. 2014.

Hybrid packet-EONs

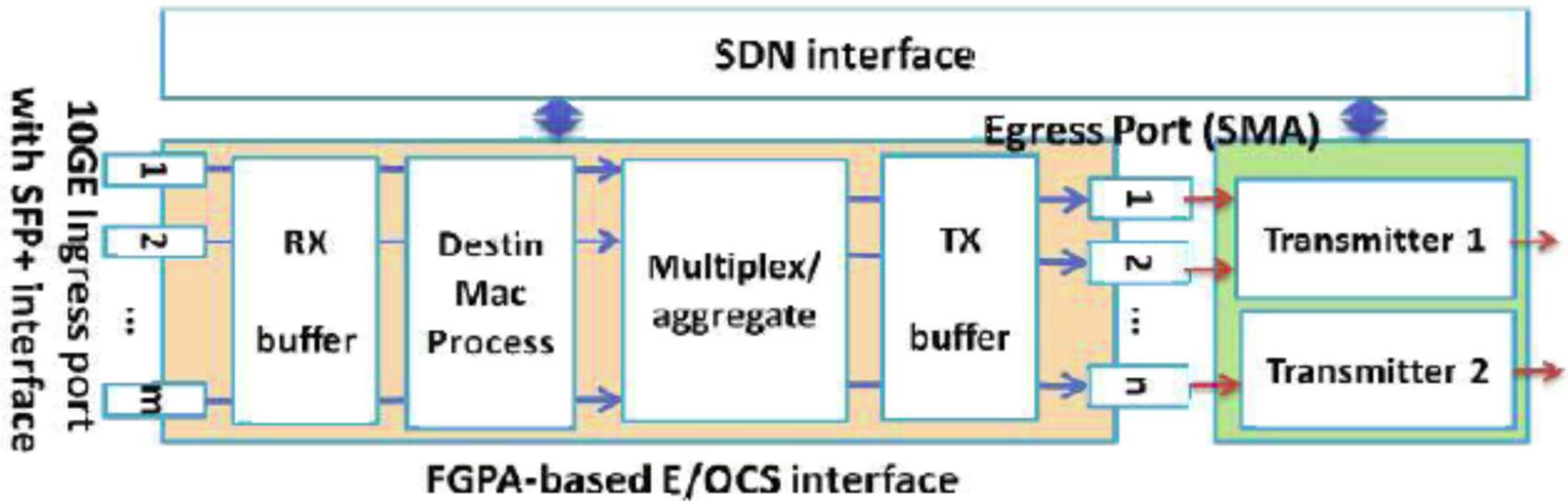


Fig.1: Design of E-EON interface based on FPGA

S. Yan et al., "Demonstration of real-time Ethernet to reconfigurable superchannel data transport over elastic optical network," 2014 The European Conference on Optical Communication (ECOC), Cannes, 2014, pp. 1-3.

Recent publications on EON

- EON SDN controllers
  - Dynamic flexi-grid wavelength circuit creation in optical access/aggregation networks [1]
  - A multi-domain testbed consisting of OFDM-based 400G OPS and flexi-grid WDM OCS domains [2]
- Optical components
  - Sliceable bandwidth variable transponders (S-BVT) [3]
    - Enable the generation of multiple optical flows to be routed into different media channels and directed toward different destinations
  - Flexi-grid WSSs [4]
    - Ultrahigh-port-count WSS (e.g., 1x95 WSS)
    - Multi-input multi-output contentionless WSS for add/drop port aggregation
    - Wavelength-selective cross-connect
  - Optical frequency combs (OFC) [5]
    - Reduced cost and inventory, improved spectral efficiency, and flexibility compared to conventional lasers

[1] N. Cvijetic *et al.*, “SDN and OpenFlow for dynamic flex-grid optical access and aggregation networks,” in J. Lightw. Technol., 2014

[2] J. M. Fàbrega *et al.*, "Demonstration of Adaptive SDN Orchestration: A Real-Time Congestion-Aware Services Provisioning Over OFDM-Based 400G OPS and Flexi-WDM OCS,” in J. Lightw. Technol., 2017

[3] N. Sambo *et al.*, "Next generation sliceable bandwidth variable transponders,” in IEEE Communications Magazine, 2015

[4] D. M. Marom *et al.*, "Survey of photonic switching architectures and technologies in support of spatially and spectrally flexible optical networking [invited]," in IEEE/OSA JOCN, 2017

[5] M. Imran *et al*, "A Survey of Optical Carrier Generation Techniques for Terabit Capacity Elastic Optical Networks," in IEEE Communications Surveys & Tutorials, 2018

Recent publications on EON

- Optical systems and networks
  - SDM-WDM ROADM [1-4]
    - Provide flexibility in both spectral and spatial domains
- Applications
  - Bulk data transfers [5,6]
  - VM migrations [6]

[1] L. E. Nelson *et al.*, “Spatial superchannel routing in a two-span ROADM system for space division multiplexing,” in *J. Lightw. Technol.*, 2014

[2] D. M. Marom and M. Blau, “Switching solutions for WDM-SDM optical networks,” in *IEEE Commun. Mag.*, 2015

[3] A. Muhammad *et al.*, “Flexible and synthetic SDM networks with multi- core-fibers implemented by programmable ROADMs,” in *Proc. ECOC*, 2014

4 Bhaumik, *et al., "* Software-defined optical networks (SDONs): a survey," Photon Netw Commun (2014).

[5] Xin Jin *et al.*, “Optimizing Bulk Transfers with Software-Defined Optical WAN,” In Proc. of SIGCOMM, 2016

[6] Payman Samadi *et al.*, "Software-defined optical network for metro-scale geographically distributed data centers," Opt. Express, 2016

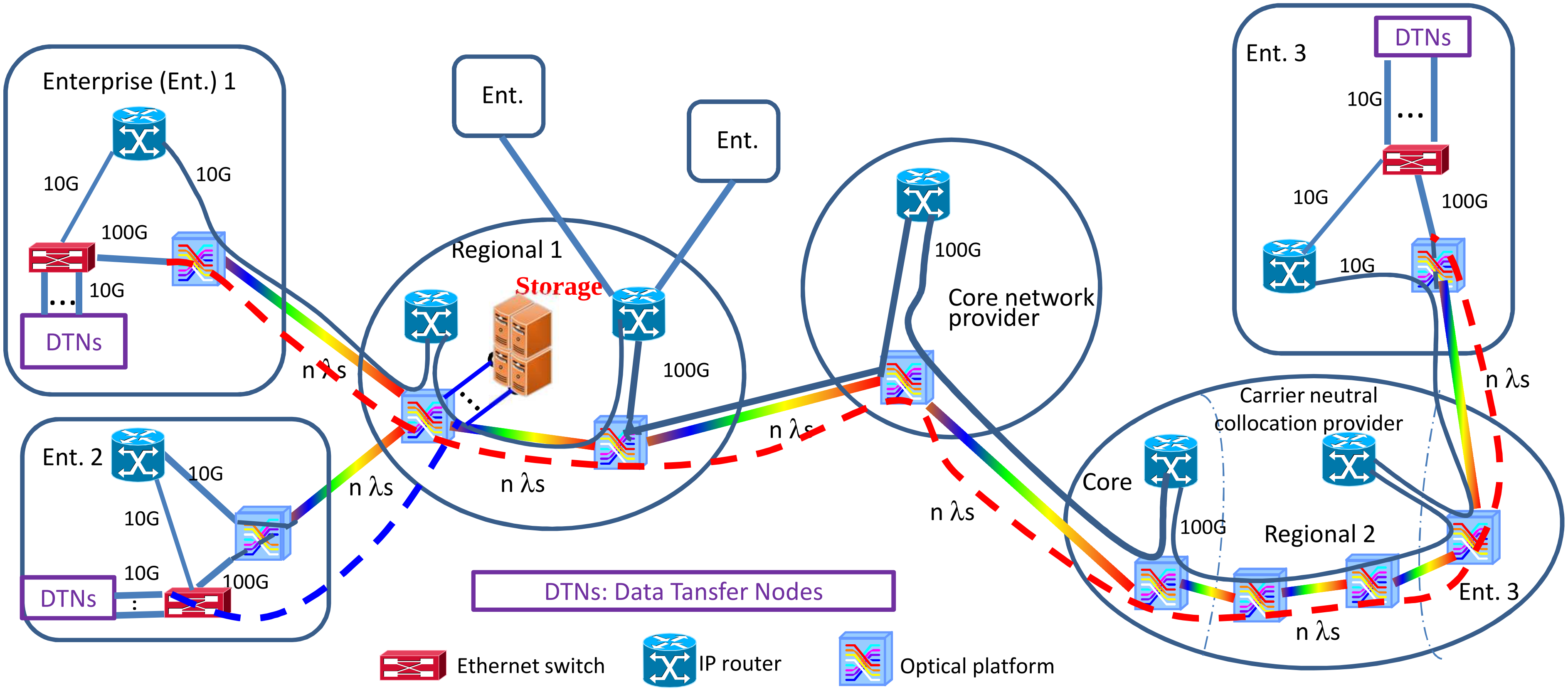
Outline

- Application drivers
- Computing landscape
- Advances in optical networking
- Research challenges
  - Integration challenges
  - Control and management plane challenges
  - Adoption challenges
- Summary: Recommendations

Integration challenges: single domain

- Science Elastic Optical Inter-Network (SEOIN)
  - FlexiGrid/EON hardware
  - SDN controllers (optical SDN)
    - Take QoT metrics into account in path computation
  - Linux features
    - Ability to plug in new congestion-control modules in TCP
    - Traffic control (match packet sending rate at NIC to the elastic rate assigned to optical circuit)
    - Tickless kernels (fine rate control at high speeds)
- Large dataset transfers on L2 virtual circuits:
  - Circuit TCP and TBF (tc): no sending rate drops when losses occur, and independence of throughput on RTT

Integration challenges: multiple domains



- Use network storage to terminate optical circuits if hybrid paths are required end2end
- Advance-reservation schedulers can improve response time for data transfers

Control-plane and management-plane challenges

- Inter-domain routing solutions
  - Need scalable routing-information distribution: path vector (e.g., BGP) or hierarchical link state
  - OSCARS topology service pushes topology to pS topology service
    - providers will not want to share their network topologies with competitors
  - Path Computation Engine Protocol with advance reservation
    - IETF draft (expires April 2018): PCEP Extensions for LSP scheduling with stateful PCE: draft-ietf-pce-stateful-pce-lsp-scheduling-01
- EON SDN controllers with support for inter-domain circuit setup
- Management-plane problems: fault management (troubleshooting), configuration management (admin errors), inter-domain performance monitoring
- Operational challenges: Certificates, topology modifications

Adoption challenges

- Need real deployment if we want our research to lead to a truly new multi-domain Internetwork
  - Otherwise scientific researchers will just make do with the current IP-based R&E networks
- What is required to achieve this transition?
  - Think NSFnet: 1985-1995 - non-commercial
  - It was during this period that we developed applications for the 1974-proposed TCP/IP based networking solution

Proposal

- Need DOE-led SEOIN backbone with colo sites to create EOIN PoPs
  - Leverages ESnet/Internet2 fiber
- At each PoP, have
  - L1 switch (CDC-Flex ROADMs, BVTs, E-EON interfaces)
  - L2 switch/L3 router (whitebox)
  - Rack of computers and storage with Incommon shared access
  - SDN controllers with Incommon shared access
- Collaborate with other US agencies (e.g., NSF) and other countries (GEFI, Fed4FIRE, Fibre)

Why a DOE-led SEOIN?

- Current state of Research-and-Education Network (REN) organizations: no root access if IT maintains your computers!
  - ESnet is a production network. Consistent, highly reliable L3 service is priority; serves many national labs
  - Campus networks: CIOs prioritize enterprise network services over research networking; same for regional RENs
  - ESnet has an SDN testbed but its user base is limited: lack of HR funding for user support?
- Need a virtual organization headed by an experimental research group with sufficient funding to "own and operate" DOE SEOIN
- Need to slowly grow the infrastructure needed to create multi-domain EONs

Summary: Recommendations

- Organize multi-agency, multi-country effort to create a DOE-led SEOIN
- Seed to enable researchers to test networking software
  - Addresses challenges in control- and management-planes
  - Integrates networking with OS, middleware, workflow managers
  - Creates applications tailored for SEOIN
  - Leverage computing solutions that fit better with WANs
    - HTC systems rather than HPC systems
    - Edge clouds rather than supercomputing centers
- For our networking software research to have impact, we need a wide-area, multi-domain research SEOIN