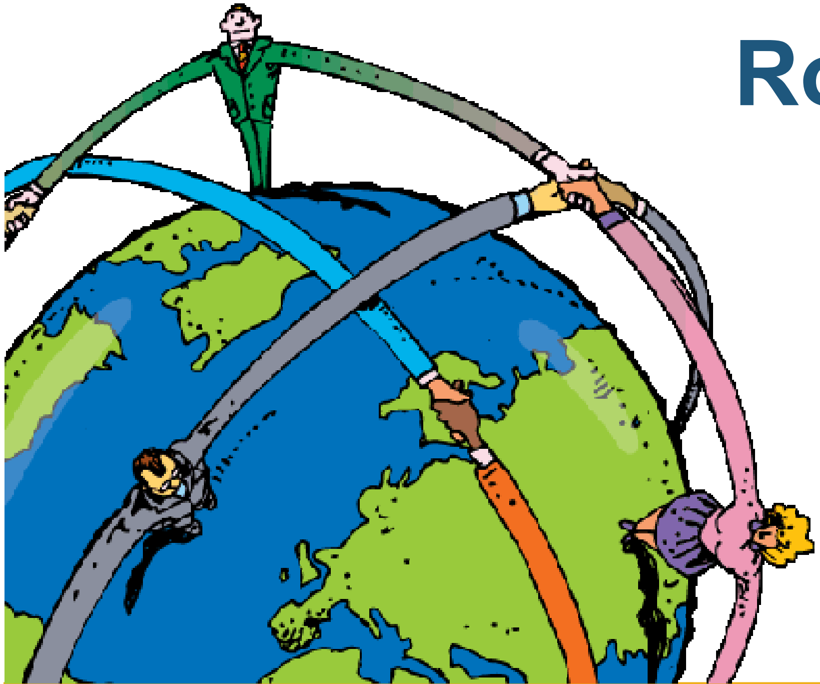


Router Transport beyond 10 Gbps

Julian Lucek

jlucek@juniper.net

Agenda

- **Historical background**
  - Router speed versus transmission speed
- **Router evolution and advent of 40 Gbps interfaces**
- **100 Gbps Ethernet**
- **Router and Optical Network integration**
  - Tunable router interfaces
  - GMPLS

Historical background: router interface speed versus transport interface speed

- **Originally, router interface speeds were much less than maximum transmission link speed**
  - E.g. mid-1990s: E3/STM1 on routers versus STM4/STM16/STM64 on transmission networks
- **In year 2000, router interface speed caught up with maximum deployed transmission speed**
  - OC192/STM64 interfaces available on router for first time
- **Currently, maximum router interface speed still at parity with maximum transmission speed**
  - OC768/STM256
- **Sometimes, situation is reversed with desired router-to-router bit rate greater than available transmission capacity**

Core router evolution: 30 times increase in <10 years!

1998: M40

Size = ½ rack

8 slots

Slot speed = 3.2 Gbps

Interfaces up to STM16

2007: T1600

Size = ½ rack

8 slots

Slot speed = 100 Gbps

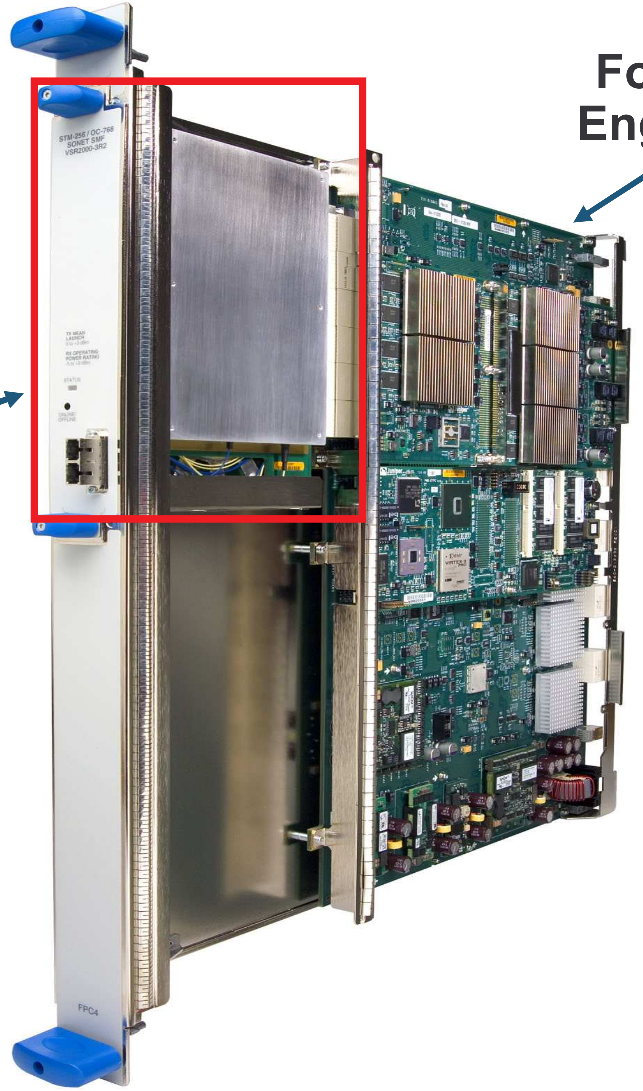
Interfaces up to STM256

(up to 100 GE in future)

Use of 40 Gbps (OC768/STM256) today

- **OC768/STM256 router interfaces have been available for a couple of years now**
- **Deployed by several service providers**
- **Today mainly used for intra-PoP connections**
  - Also used for ultra-high speed private peering
- **Also for inter-site connections within the same city, over dark fibre**
- **Less deployed so far for inter-city connections**
  - Not everyone has infrastructure capable of carrying 40 Gbps wavelengths..
    - The fibre may not be capable of carrying 40 Gbps wavelengths due to PMD etc
    - Or the optical transport equipment may not support 40 Gbps
  - In some cases, spectral efficiency may be worse for 40 Gbps wavelengths than 10 Gbps wavelengths
    - Some trans-oceanic systems

OC768/STM256 interface

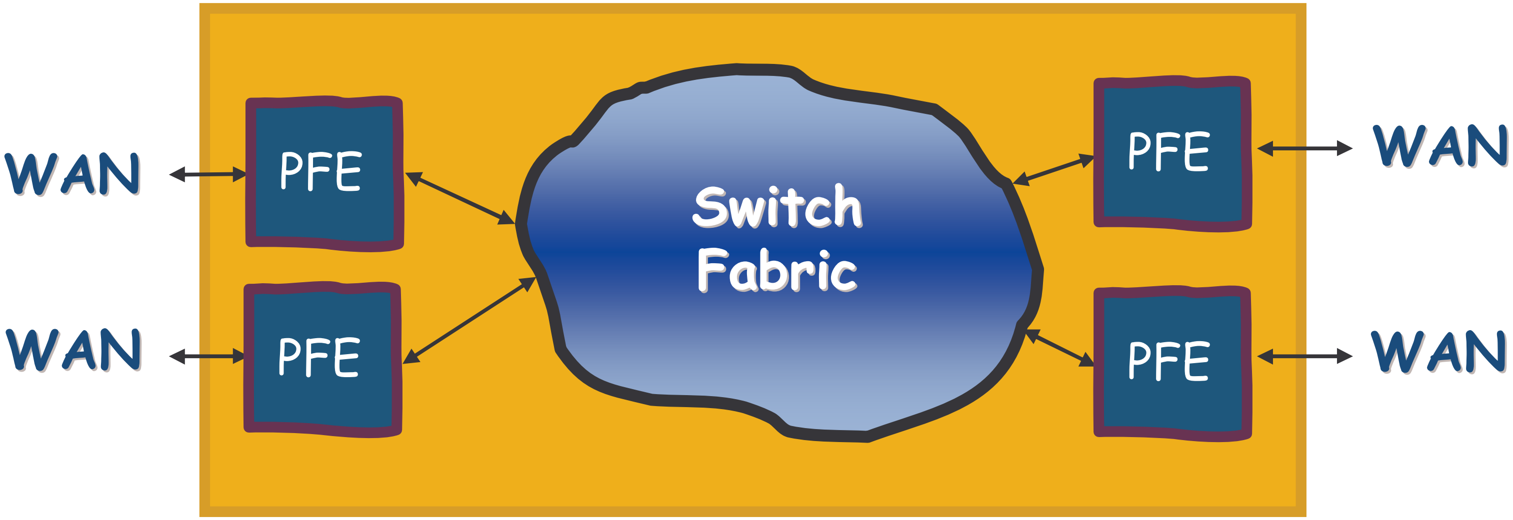


Packet Forwarding Engine (PFE)

Interface card ("PIC")

Contains SONET/SDH framer ASIC and 40 Gbps optics operating in 1550 nm region

Support of 40 Gbps interfaces

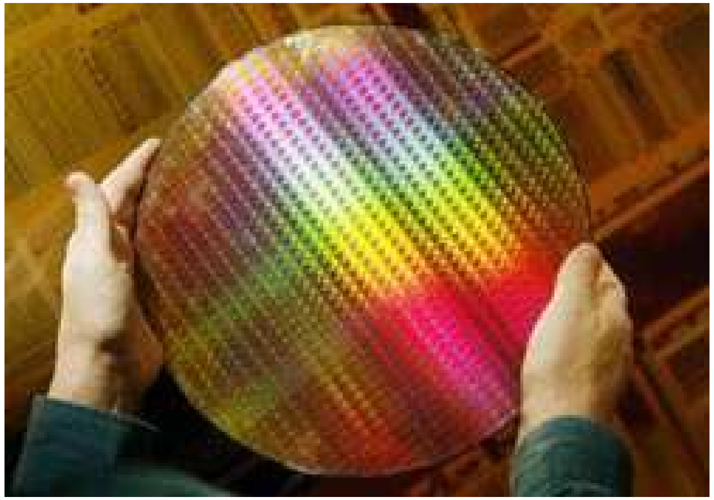


- Each Packet Forwarding Engine (PFE) needs to support packet processing functions (encapsulation/de-encapsulation, route lookup, queueing, firewall filtering) at rate of at least 40 Gbps in each direction
- Switch fabric needs to have sufficient capacity to interconnect PFEs at full line rate
- ASICs used in PFE and switch fabric to achieve these requirements

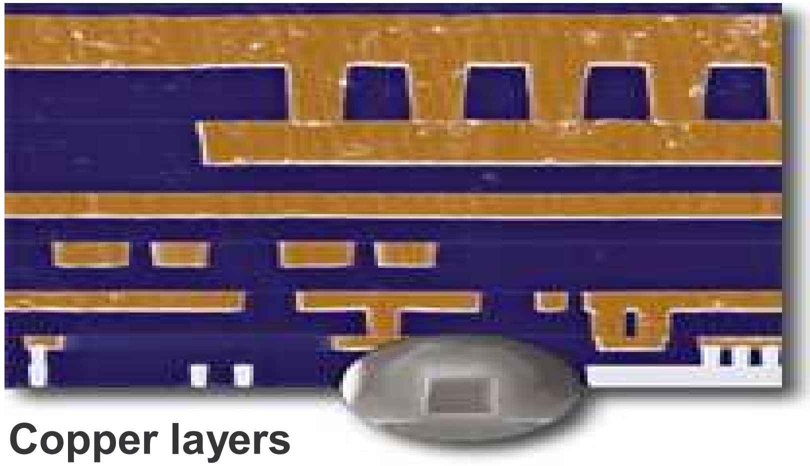
Why ASICs?

- **A high degree of integration is desirable, because it**
  - Reduces complexity of logic design
  - Reduces system cost
  - Reduces power
  - Increases system reliability
- **FPGAs**
  - ✓ Faster development, field upgrades
  - ✓ Smaller Non-Recurring Engineering (NRE) cost
  - × Lower performance & density
  - × Higher per part price
- **Custom ICs**
  - ✓ Higher performance & density
  - ✓ Lower per part price
  - × More complexity, longer development time
  - × Larger NRE
- **Network processors**
  - ✓ More flexibility
  - × Lower performance

ASIC Manufacturing



300mm wafer



Copper layers

300mm wafer fab

Packaging

ASIC Verification

- **Goal: First-time-right silicon**
  - Avoid expensive ASIC respins
  - Simulations are far easier to debug than real chips
  - No feature or function makes it into the chip without a rigorous test for it
- **At least as many verification engineers as design engineers per chip**
- **More verification code than design code!**
- **Performed at multiple levels**
  - Block level
  - Chip level
  - Sub-system level
  - System level

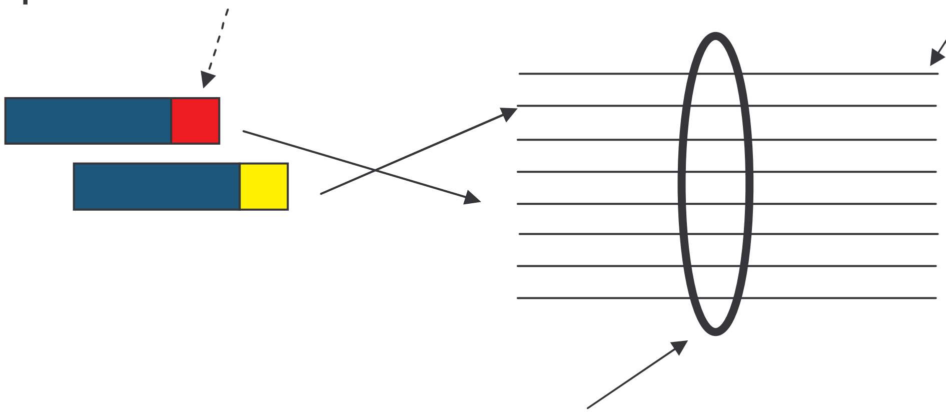
Router-to-router transport over multiple links

- **Sometimes the required router-to-router capacity is greater than the speed of an individual transport link**
- **For example, the WAN transport infrastructure may be based on 10 Gbps but need (for example) 40 Gbps router-to-router capacity**
- **Three main solutions exist:**
  - Packet load balancing across multiple links without link aggregation
  - Packet load balancing across multiple links with link aggregation
  - Inverse multiplexing across multiple links

Link Aggregation

Hash algorithm based on MPLS labels, source/destination IP address, source/destination port, protocol

Up to 16 physical interfaces (SONET/SDH or Ethernet)



Single logical interface as far as Layer 3 is concerned

- one ip address
- single IGP adjacency

Hash function to ensure that packets belonging to same flow always use same physical link

Inverse multiplexing

•Operates at L1: each packet is striped across the set of links

•Single logical interface as far as router is concerned

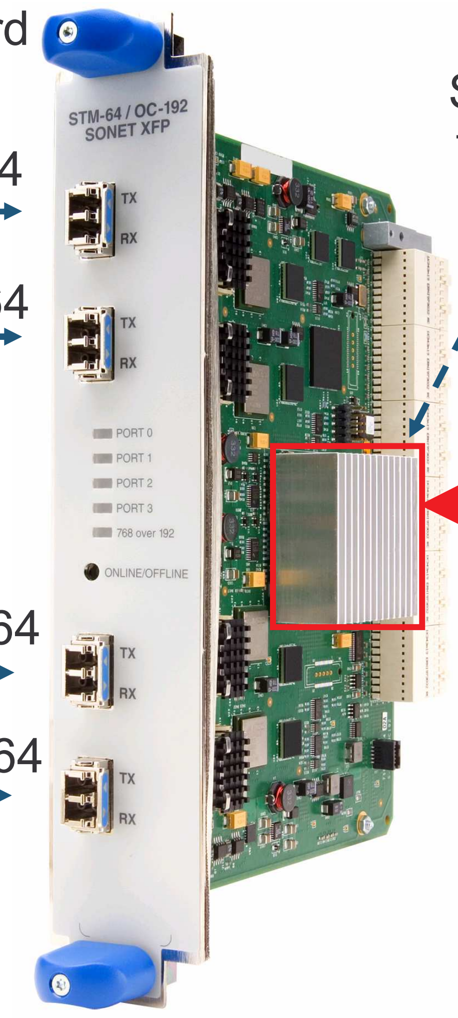
Four port OC192/STM64 card

10 Gbps OC192/STM64

10 Gbps OC192/STM64

10 Gbps OC192/STM64

10 Gbps OC192/STM64



SONET/SDH framer ASIC

Single 40 Gbps packet stream

Packet Forwarding Engine (PFE)

100 Gbps Ethernet

100 Gbps Ethernet

- The industry has recently begun discussions about 100 Gbps Ethernet
- Within the IEEE, the High Speed Study Group (HSSG) started meeting in September 2006 to discuss 100 Gbps Ethernet
- Has had seven meetings so far (during each 802.3 Plenary and also at interim meetings)
- HSSG home-page is on the IEEE website at:
  - http://grouper.ieee.org/groups/802/3/hssg/index.html

HSSG deliverables

- **Statement of Objectives**
- **Response to 5 Criteria**
  - Technical Feasibility
  - Economic Feasibility
  - Unique Identity
  - Compatibility
  - Broad Market Potential
- **Project Authorization Request (PAR)**
  - The formal request to 802.3 to create a task force
- **Tutorial**

Reaches to be specified for 100 Gbps

- **At least 40km on single-mode fibre**
- **At least 10km on single-mode fibre**
- **At least 100m on multi-mode fibre**
- **At least 10m over a copper cable assembly**

- **Physical format to be determined –likely to use parallel approach rather than serial approach**
  - 4 x 25 Gbps?
  - 10 x 10 Gbps?
  - 1 x 100 Gbps considered not achievable in required timescale at reasonable cost

Technology leverage

- **Unlike previous Ethernet generations, there is less scope for 100 Gbps to leverage already existing technology**
  - 10 Gbps Ethernet was able to leverage OC192/STM64 SONET/SDH
  - 1 Gbps Ethernet was able to leverage Fibre Channel
- **However, the expected use of parallelism (e.g. 10 X 10 Gbps) in 100 Gbps Ethernet could allow some leverage..**

40 Gbps Ethernet?

- **There was some debate in the HSSG over whether the IEEE should also define a specification for 40 Gbps Ethernet**
  - Strong interest from the server and storage area network industry
- **In the end, the Project Authorisation Request (PAR) submitted in July 2007 proposed to work on *both* 40 Gbps Ethernet and 100 Gbps Ethernet**
  - http://grouper.ieee.org/groups/802/3/hssg/PAR/par_0707.pdf

Core router evolution: what about the next 10 years?

1998: M40

Size = ½ rack

8 slots

Slot speed = 3.2 Gbps

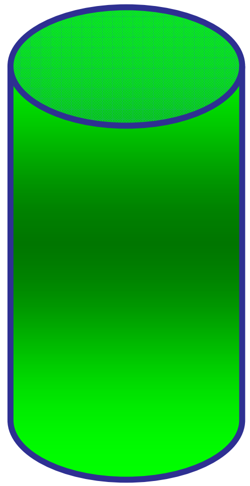
2H 2007: T1600

Size = ½ rack

8 slots

Slot speed = 100 Gbps

2H 2016/17: ??



Size = ½ rack??

8 slots??

Slot speed = 3000 Gbps??

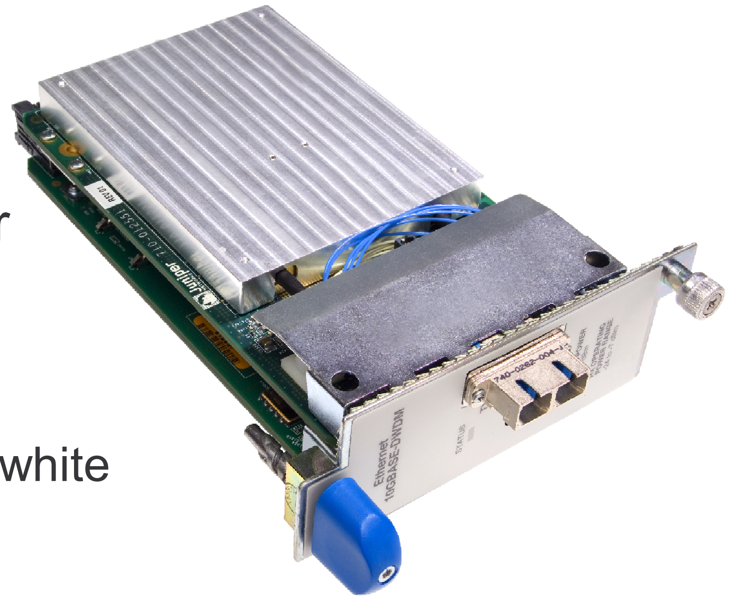
Router-Optical Network Integration

Tunable lasers on router interfaces

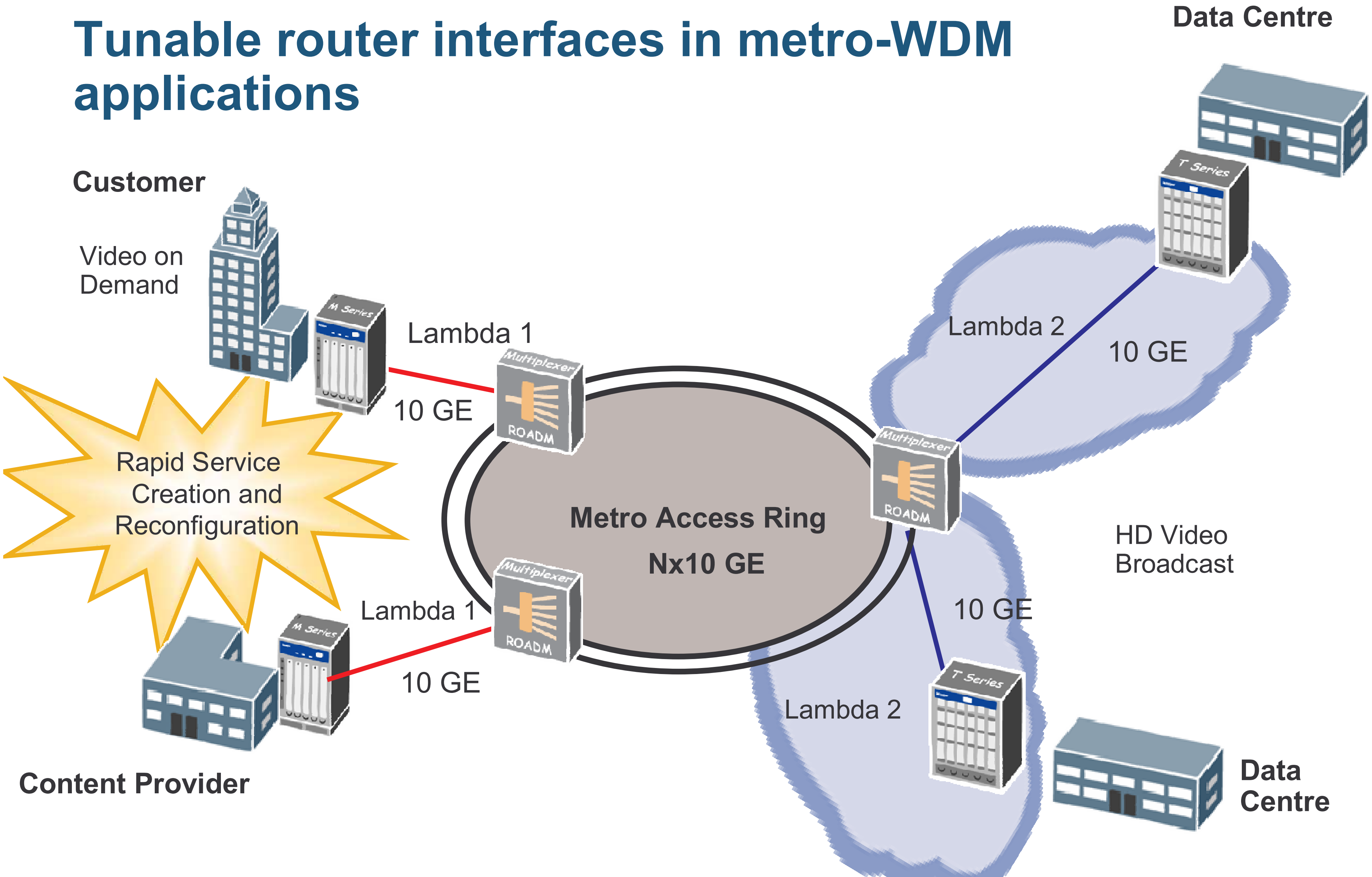
- **Traditionally, routers have “grey” transmitter on interface card**
  - Non-specific wavelength in 1300nm or 1550nm band
  - Signal is carried between routers over dark fibre
  - Or plugs into transponder on WDM system (OEO conversion to a specific wavelength)
- **Alternative approach: tunable laser on router interface**
  - allows elimination of transponder on the WDM equipment..

Tunable 10GE Interface

- **First optically integrated router interface (Q1 2006)**
- 80 km reach
- Tunability within C-Band via CLI
  - Choice of 45 wavelengths
- Tunability gives more flexibility and simpler sparing than fixed-wavelength modules
- Diagnostic/monitoring Capabilities
- Further reading:
  - http://www.juniper.net/solutions/literature/white_papers/200202.pdf



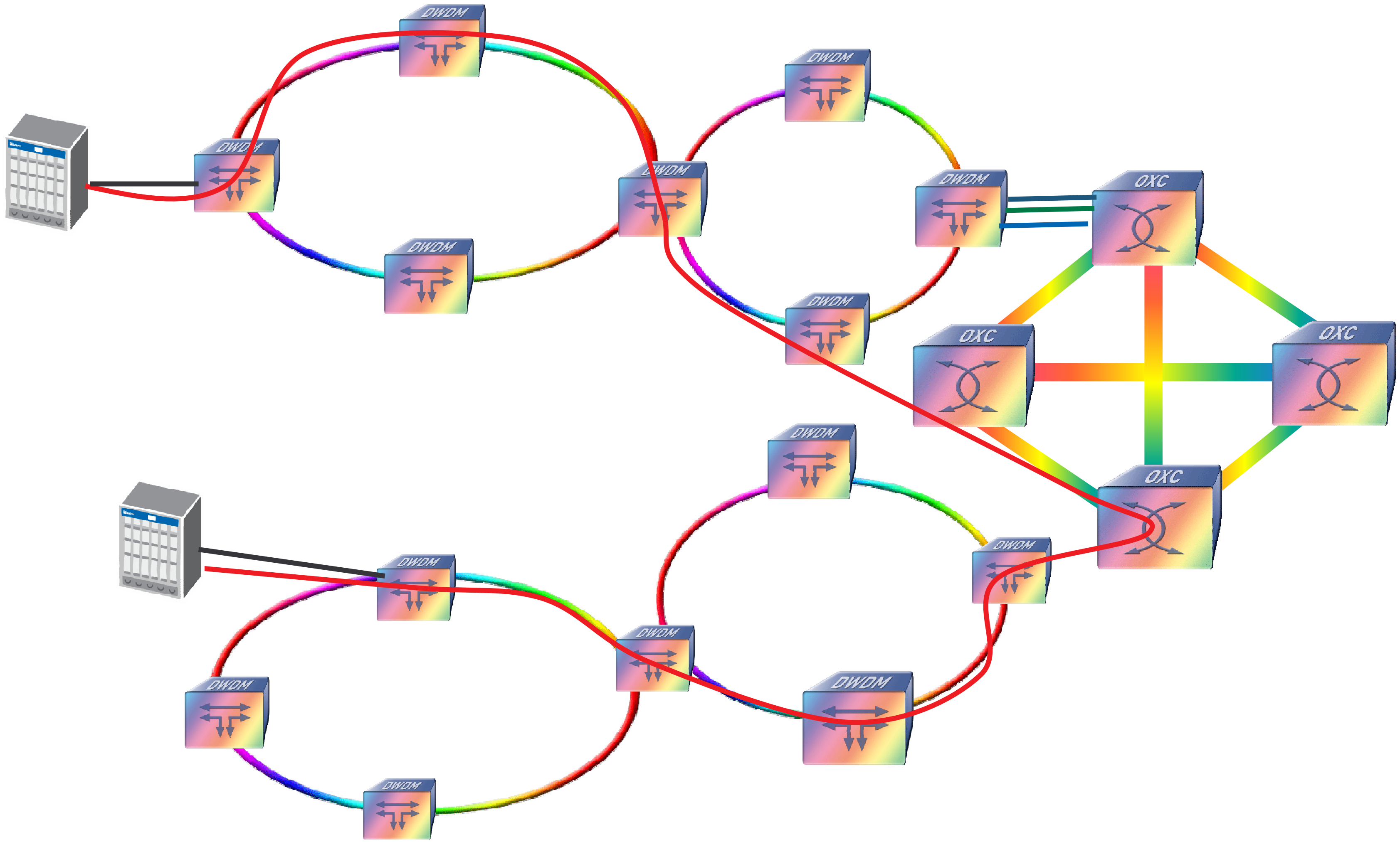
Tunable router interfaces in metro-WDM applications



Trends in long-haul optical transport

- Traditionally, long haul optical transport has been in the form of point-to-point DWDM line systems, with manual patching between line systems
- More recently, Reconfigurable Optical Add-Drop Multiplexers (ROADM) have become available, which allow an end-to-end path to be created more easily across the network
- Internally, a ROADM either performs switching in the optical domain, or in electronic domain with OEO conversion

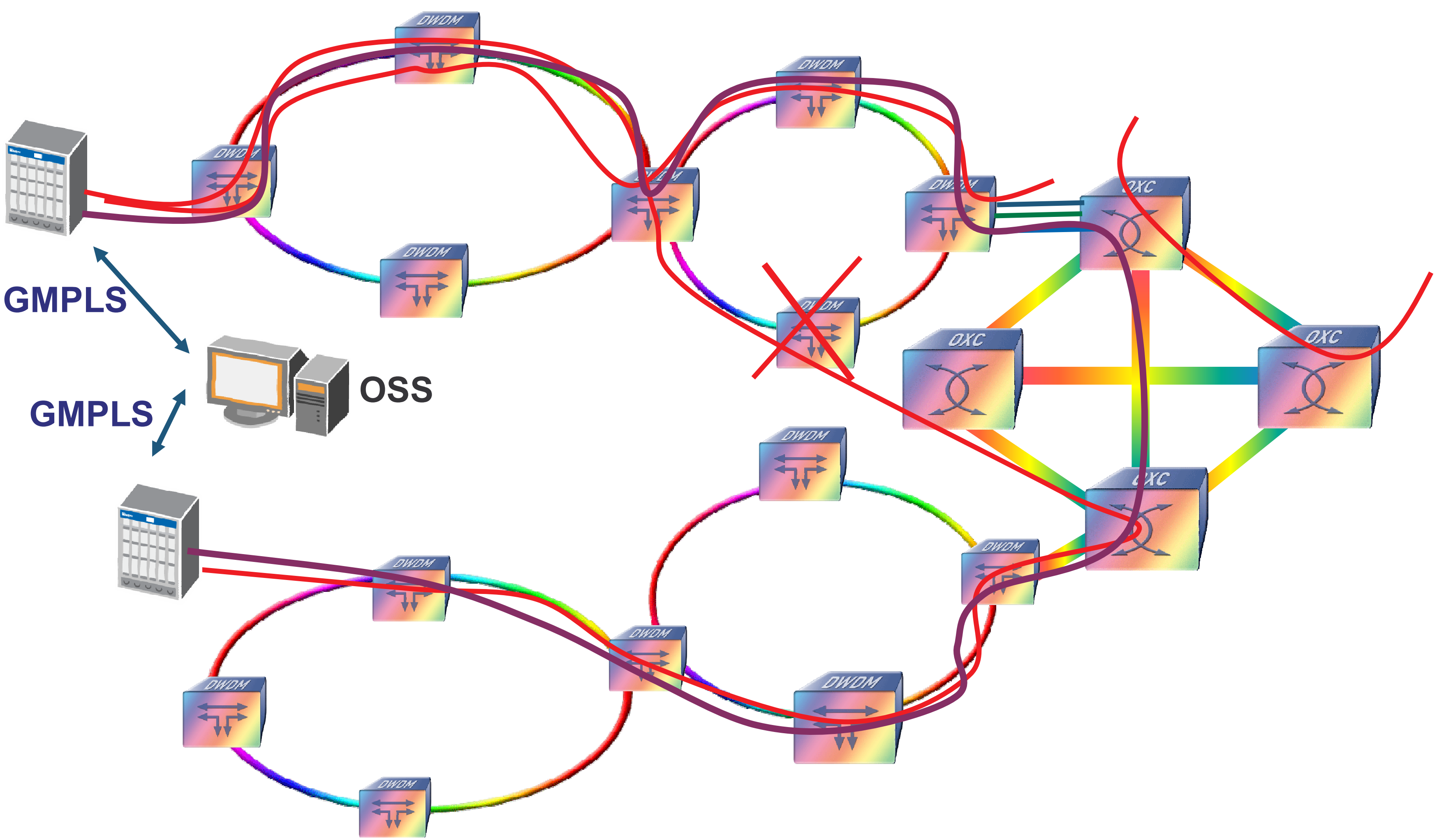
The New ROADM Network



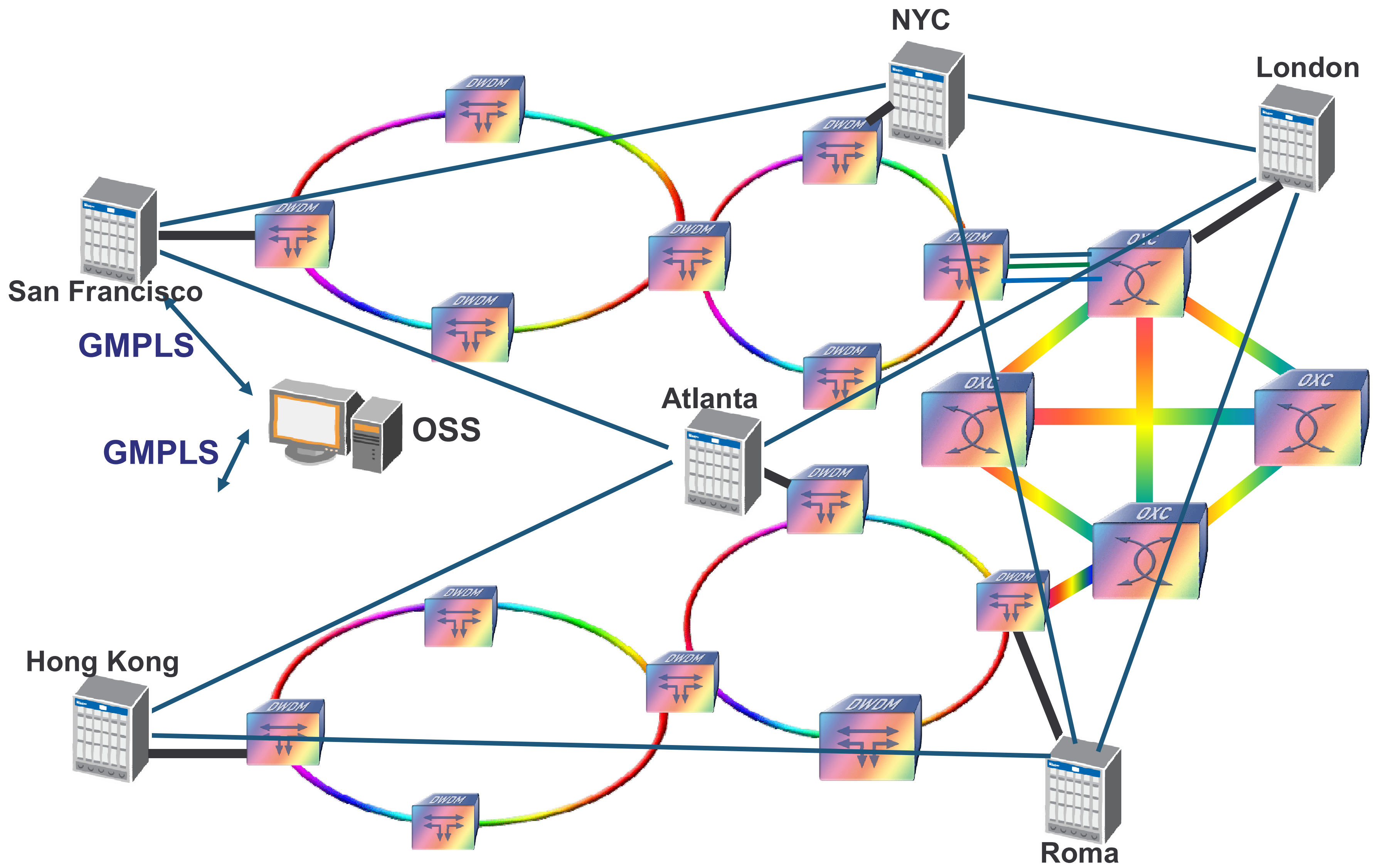
Control plane for Router-Optical network: GMPLS

- MPLS has the ability to compute/signal paths through a network of Label Switch Routers
- Generalised MPLS (GMPLS) extends this capability to other forms of network equipment, for example Sonet/SDH cross-connects, optical cross-connects etc
- A GMPLS control plane therefore brings closer integration between routers and the optical network..

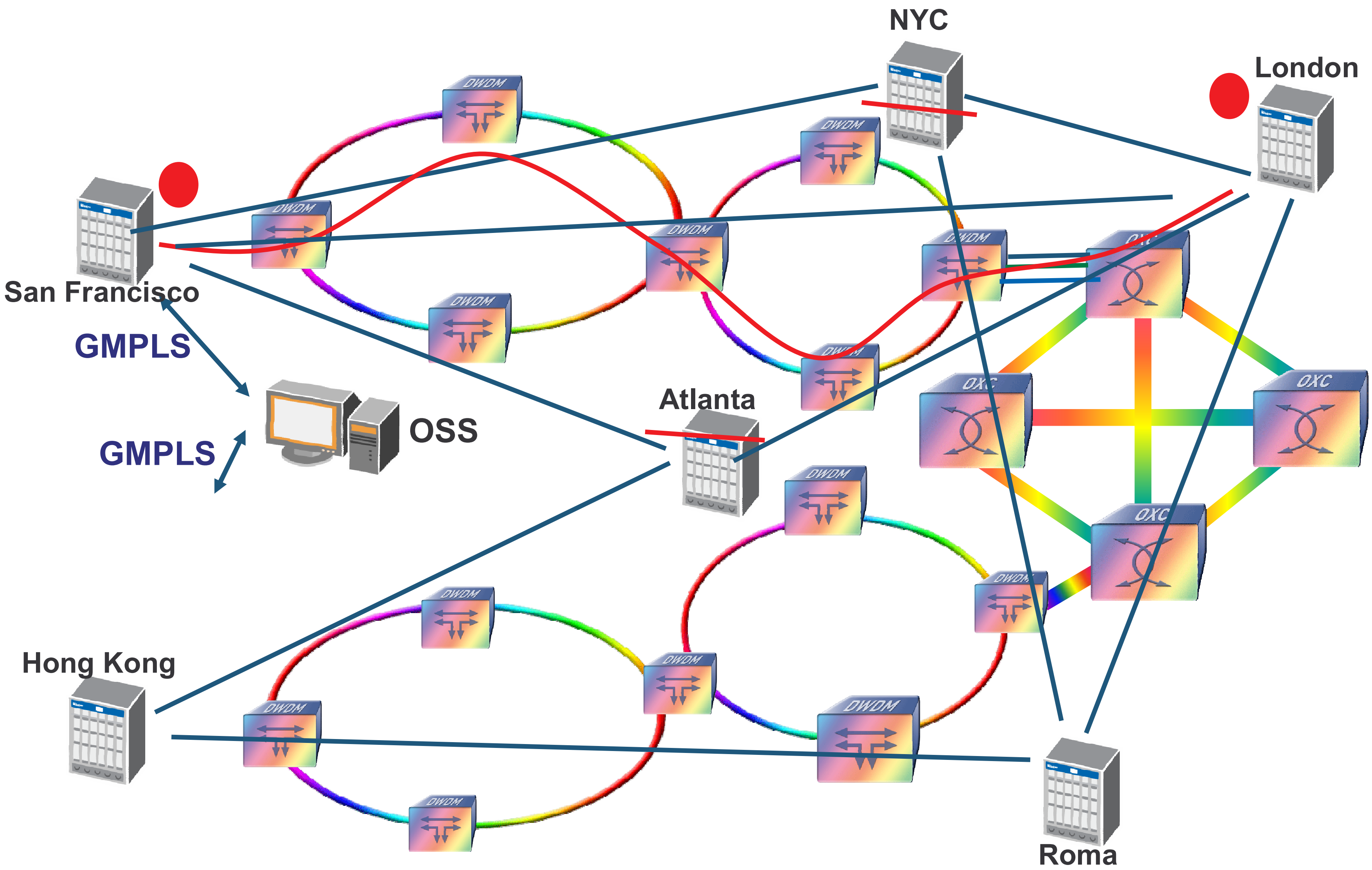
Restoration Capabilities



Capacity Management



Capacity Management



Summary

- Modern core routers support multiple 40 Gbps interfaces today, and 100 Gbps Ethernet interfaces once the standards are created and the transceivers are available
- Novel inverse multiplexing techniques enable router-to-router connectivity at higher speeds than available from an individual transmission link
- Tunable lasers and the use of a GMPLS control plane give closer integration of routers with the optical network

Juniper your Net™