CWDM/DWDM designs for DC interconnect

Jaromír Pilař (jpilar@cisco.com)
Consulting Systems Engineer, CCIE 2910

Cisco Expo 2008

Enable Your Network
Empower Your Business

Agenda

- DR solutions - motivation, parameters, components
- Options for DC interconnection
- WDM system anatomy
- Design scenarios and examples
- Protocol and applications interactions with transport systems
- Certifications

Disaster recovery and business continuance solutions

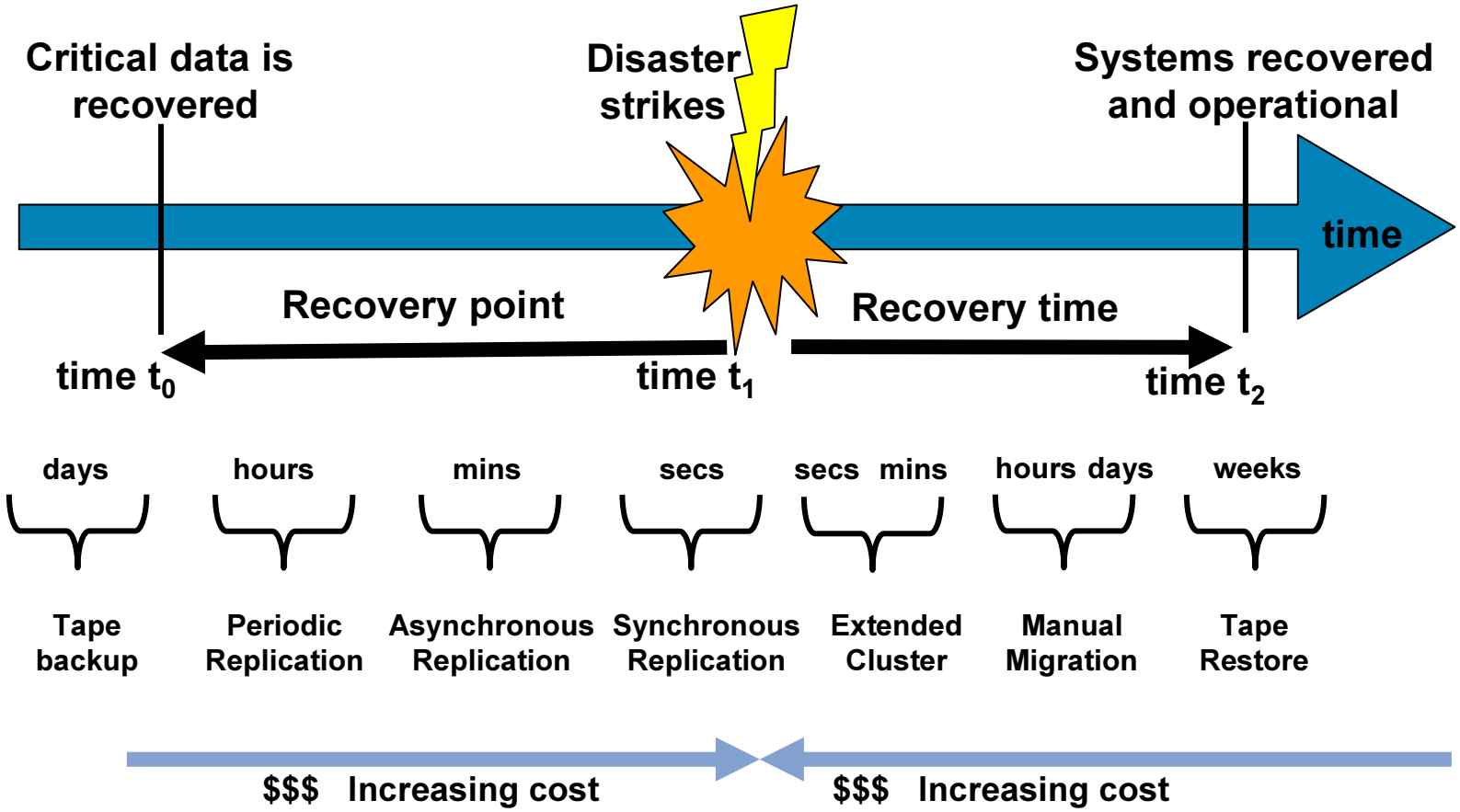
Motivation, components and parameters

Business continuance solutions

- Motivation for BC/DR solutions
  - Protect operation of enterprise or organization during unexpected event (‘disaster’)
  - Sometimes enforced by law or other regulation (Basel II etc.)
- What have to be done to create BC/DR policy
  - Identification of Critical Applications
  - Distance between disaster and recovery zone
  - Mode of operation (active-active, active-stanby)
  - Tolerable Application down time
- What parameters should BC/DR policy have
  - Must be measurable
  - RTO, RPO, RAO

Disaster recovery parameters

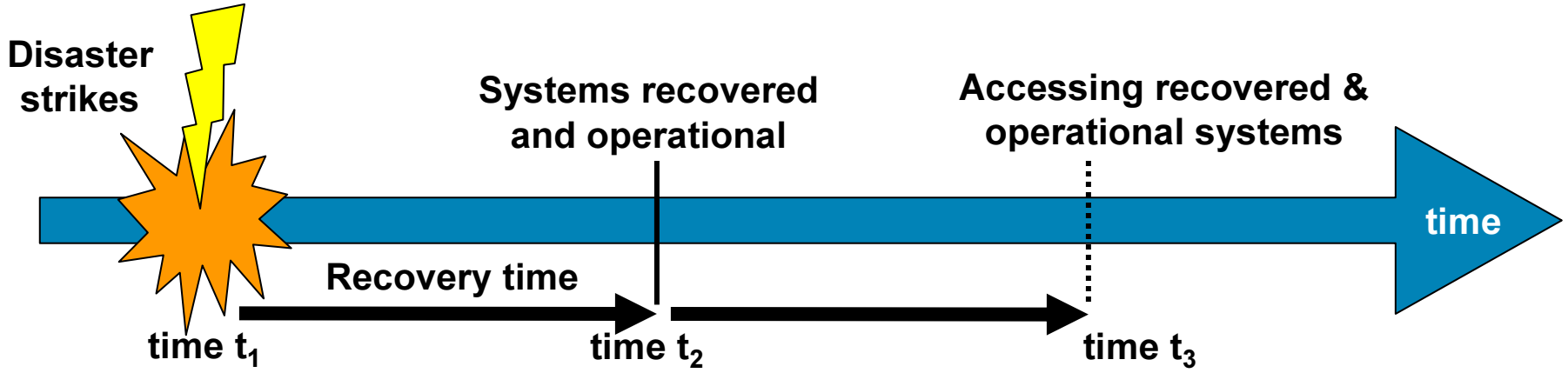
Recovery Time Objective and Recovery Point Objective



- How current or fresh is the data after recovery?
- How quickly can systems and data be recovered?

Disaster recovery parameters

Recovery Access Objective



(t_2)→ Recovery Time Objective
($t_3 – t_2$)→ Recovery Access Objective

Time taken by network to converge and provide a path for clients to access the applications and data
Note: RAO can be +ve or –ve w.r.t. Recovery time (t2)

Business Continuance / Disaster Recovery

Logical solution components

Front end: Site selection

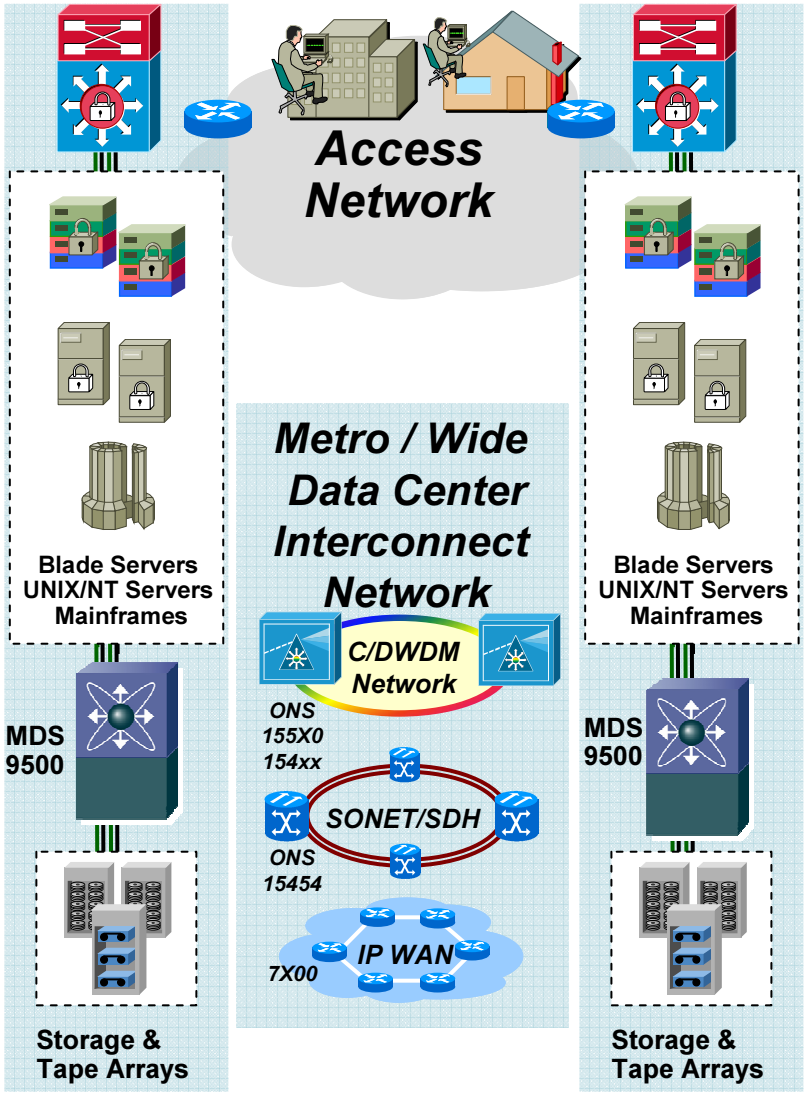
- pointing users to operational site
  - DNS based solutions
  - solutions based on routing protocols (RHI)
  - HTTP redirection

Application: Content switching

- selecting the appropriate server to perform requested operation
  - load balancing
  - load and health monitoring

Back end: Data replication and inter datacenter transport

- ensuring data availability in case of disaster or failure
  - solutions for storage array based mirroring
  - CDP solutions
  - optical solutions (DWDM, CWDM, SDH)
  - FCIP



Disaster recovery and business continuance solutions

Options for datacenter interconnect

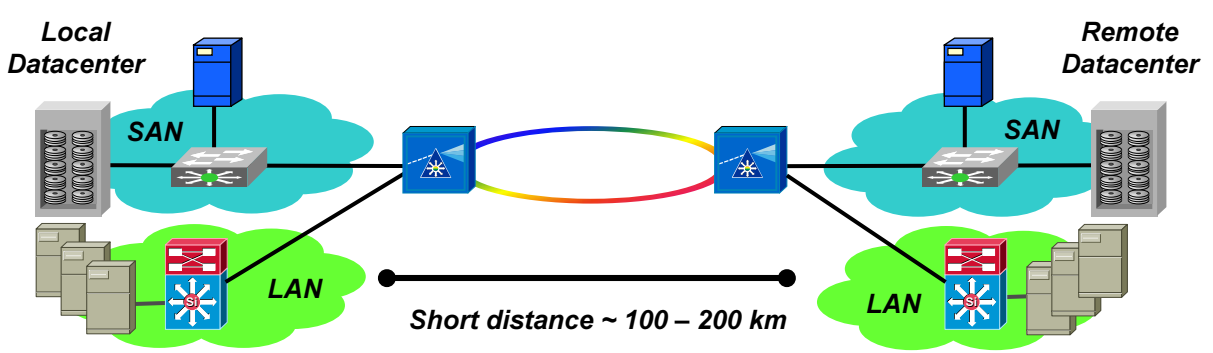
Inter datacenter channels

- **Data network (LAN/MAN)**
  - ✓Gigabit Ethernet
  - ✓10 Gigabit Ethernet
- **Storage Area Network (SAN)**
  - ✓Fiber Channel (1, 2, 4 and 10 Gbps)
  - ✓FICON (1, 2, 4 and 10 Gbps)
  - ✓ESCON
- **Channels for cluster applications**
  - ✓Heartbeat signals (GE, FE)
  - ✓GDPS/GDPS 2 (Sysplex ETR/CLO, ISC-1, ISC-2, ISC-3, STP)
- **Others**
  - ✓E1 or E3 for voice
  - ✓STM-x for ATM
  - ✓FDDI

Datacenter interconnect options

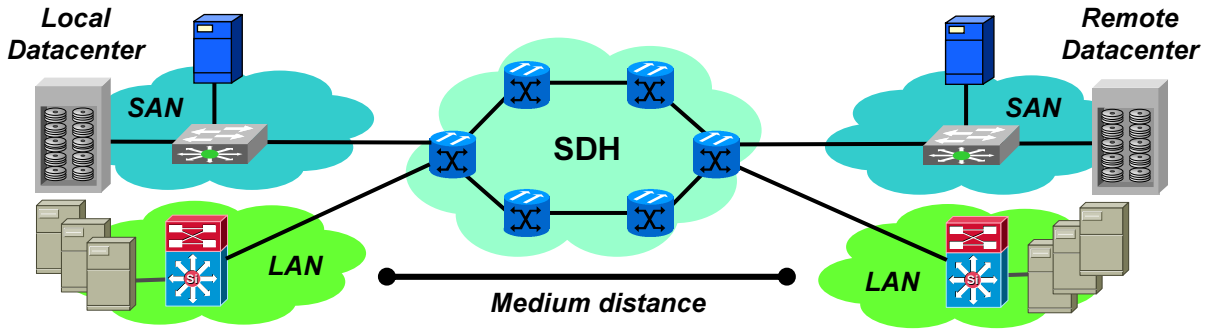
DWDM/CWDM

- most often short distance
- dark fiber must be available
- dedicated channels for LAN, SAN and other signals



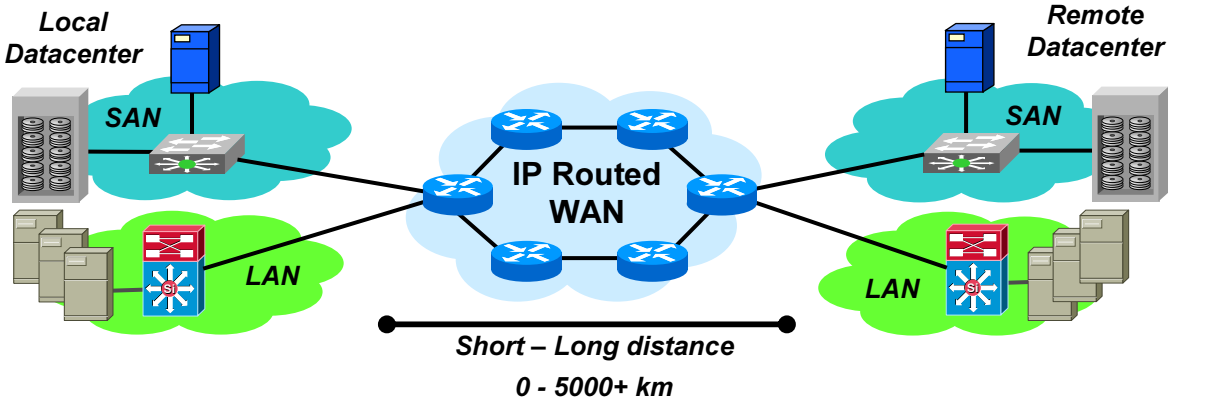
SONET/SDH

- most often short – intermediate distance
- dark fiber not avail. – distance, cost, exhaust
- links may be shared
- EoSDH and FCoSDH



IP, IP/MPLS, Metro Ethernet

- short – long distance
- dark fiber not available
- links may be shared
- FCIP for FC and/or FICON



Single lambda vs. multiple lambdas

- Single lambda
  - One channel only (e.g. 1000BaseZX, 10GBaseZR etc.)
  - More 'channels' – using TDM (SDH, EoSDH, FCoSDH) or IP (FCIP, CEoIP)
- Multiple lambdas (grids defined by ITU standard)
  - CWDM – 20 nm grid (usually 8 or 16 channels)
  - DWDM – 200 GHz, 100 GHz or 50 Ghz grid
  - WWDM
- In ITU terminology
  - DWDM: channel spacing is less than 1000 GHz (8 nm at 1550 nm)
  - CWDM: channel spacing is greater than 1000 GHz (8 nm at 1550 nm) but less than 50 nm
  - WWDM: channel spacing is greater than 50 nm (for example joint 1310-1550 dual bands)

Transmission Bands

- Optical transmission is conducted in wavelength regions, called “bands”.
- Commercial DWDM systems typically transmit at the C-band

  Mainly because of the Erbium-Doped Fiber Amplifiers (EDFA).
- Commercial CWDM systems typically transmit at the S, C and L bands.
- ITU-T has defined the wavelength grid for xWDM transmission

  G.694.1 recommendation for DWDM transmission, covering S, C and L bands.

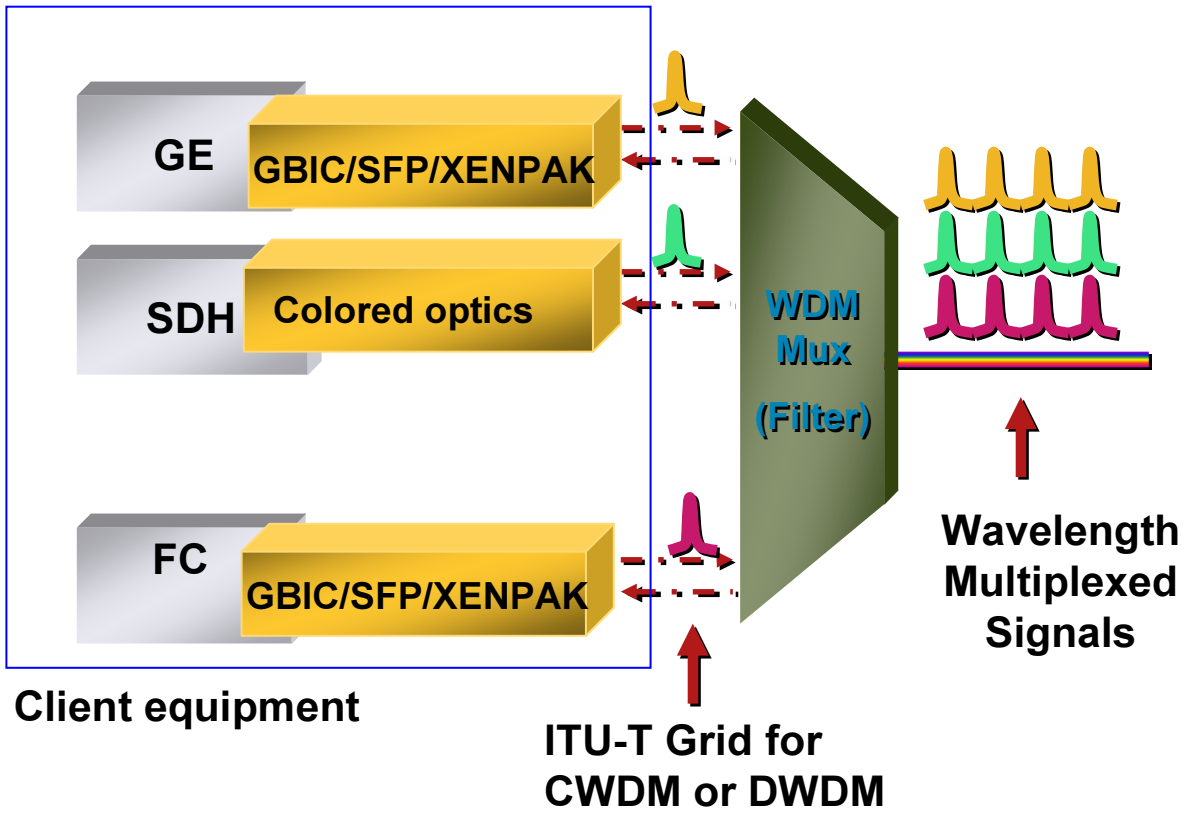
  G.694.2 recommendation for CWDM transmission, covering O, E, S, C and L bands.

Band	*Wavelength (nm)*
O	1260 – 1360
E	1360 – 1460
S	1460 – 1530
C	1530 – 1565
L	1565 – 1625
U	1625 – 1675

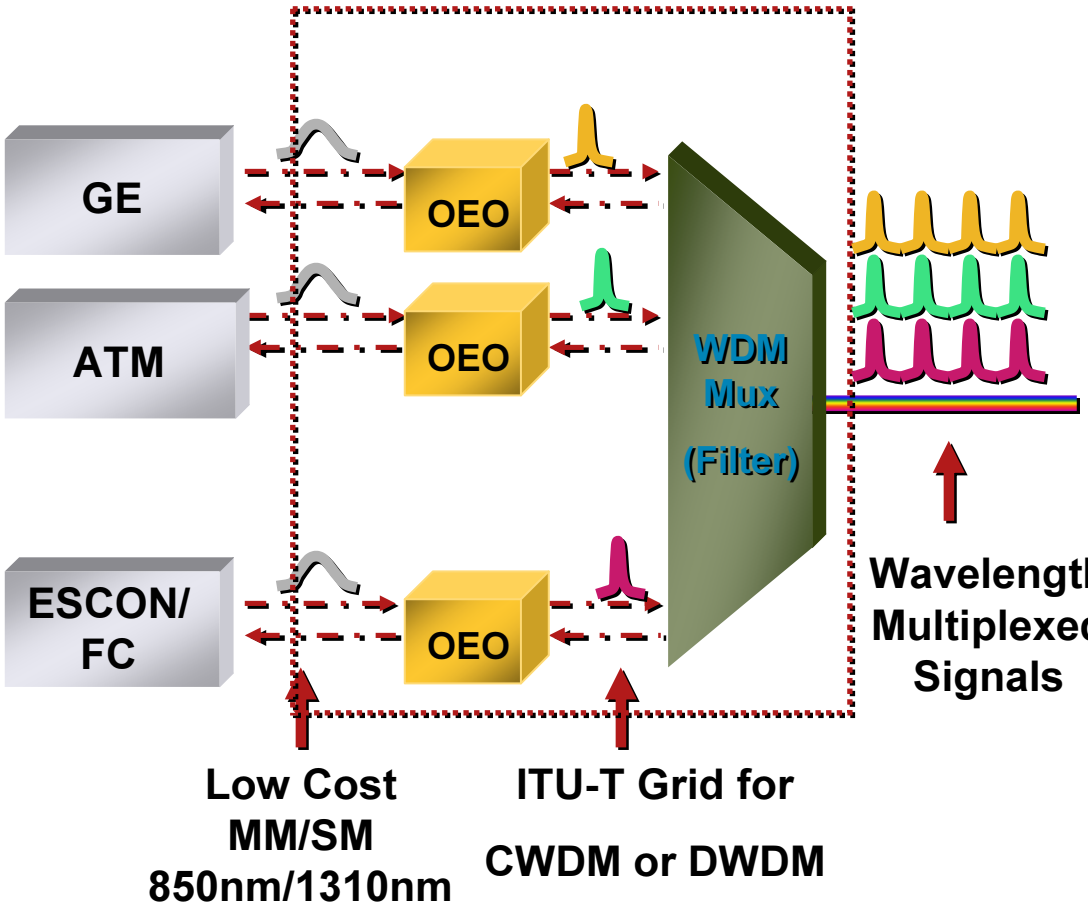
CWDM/DWDM designs for DC interconnect

WDM system anatomy

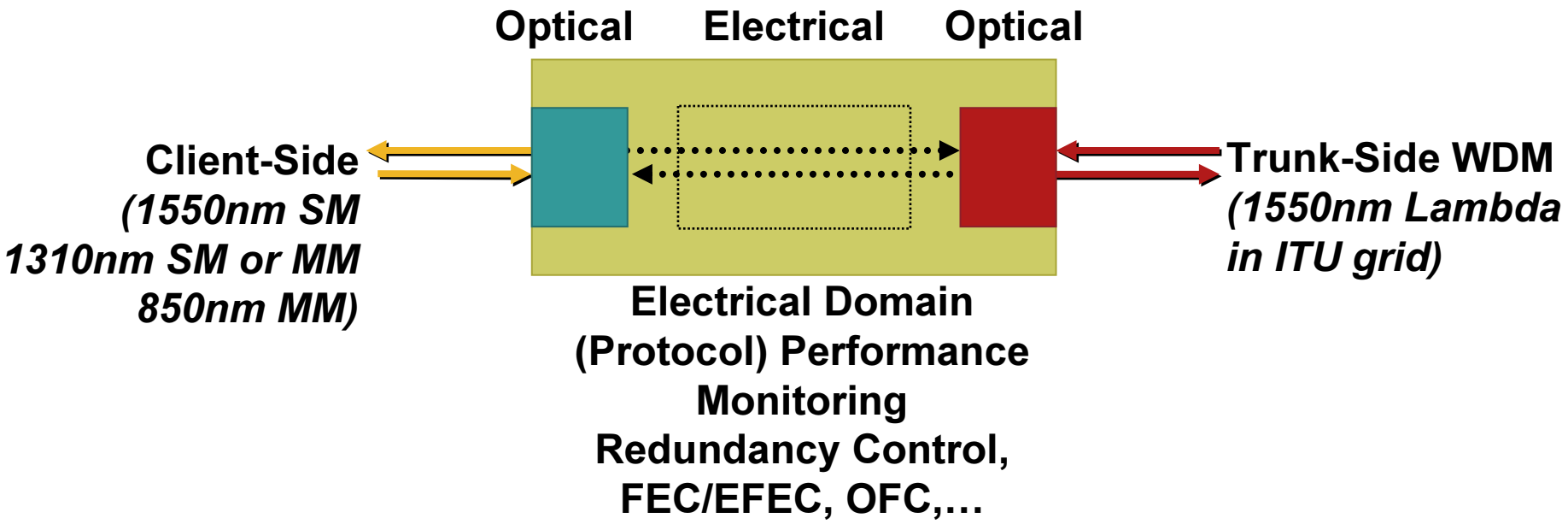
WDM system anatomy - colored clients



WDM system anatomy - transponders



WDM system components - Transponder



- 2R—Regenerate and Reshape
- 3R—Regenerate, Reshape and Retime

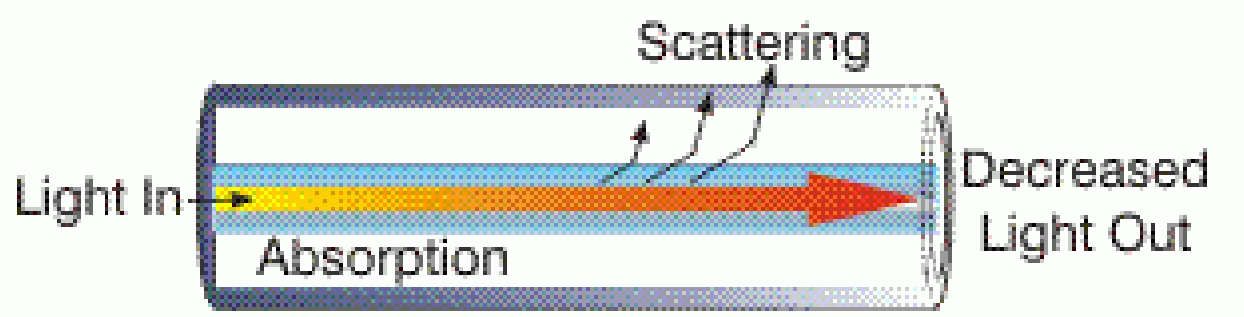
Fiber attenuation

- The decrease in optical level along a fiber optic waveguide
- Two main components:

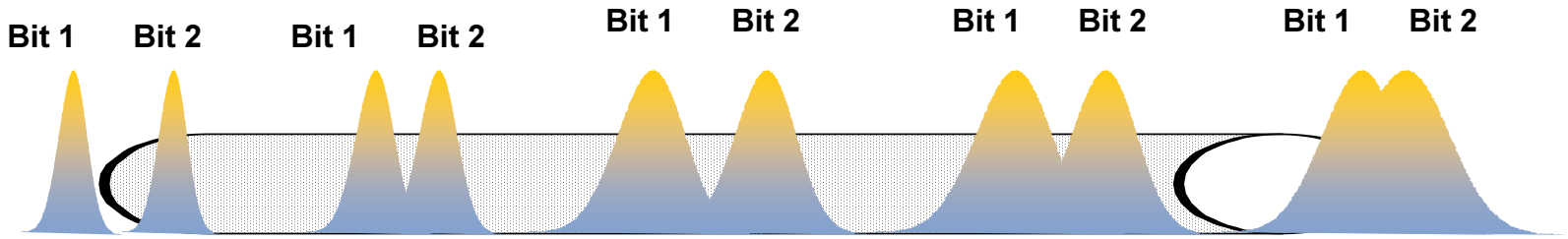
  Absorption: portion of optical attenuation in optical fiber resulting from the conversion of optical power to heat. Caused by impurities in the fiber such as hydroxyl ions.

  Scattering: change of direction of light rays or photons after striking small particles. It may also be regarded as the diffusion of a light beam caused by the non-homogeneity of the transmitting material.

- Different for different wavelength (lowest in C-band)



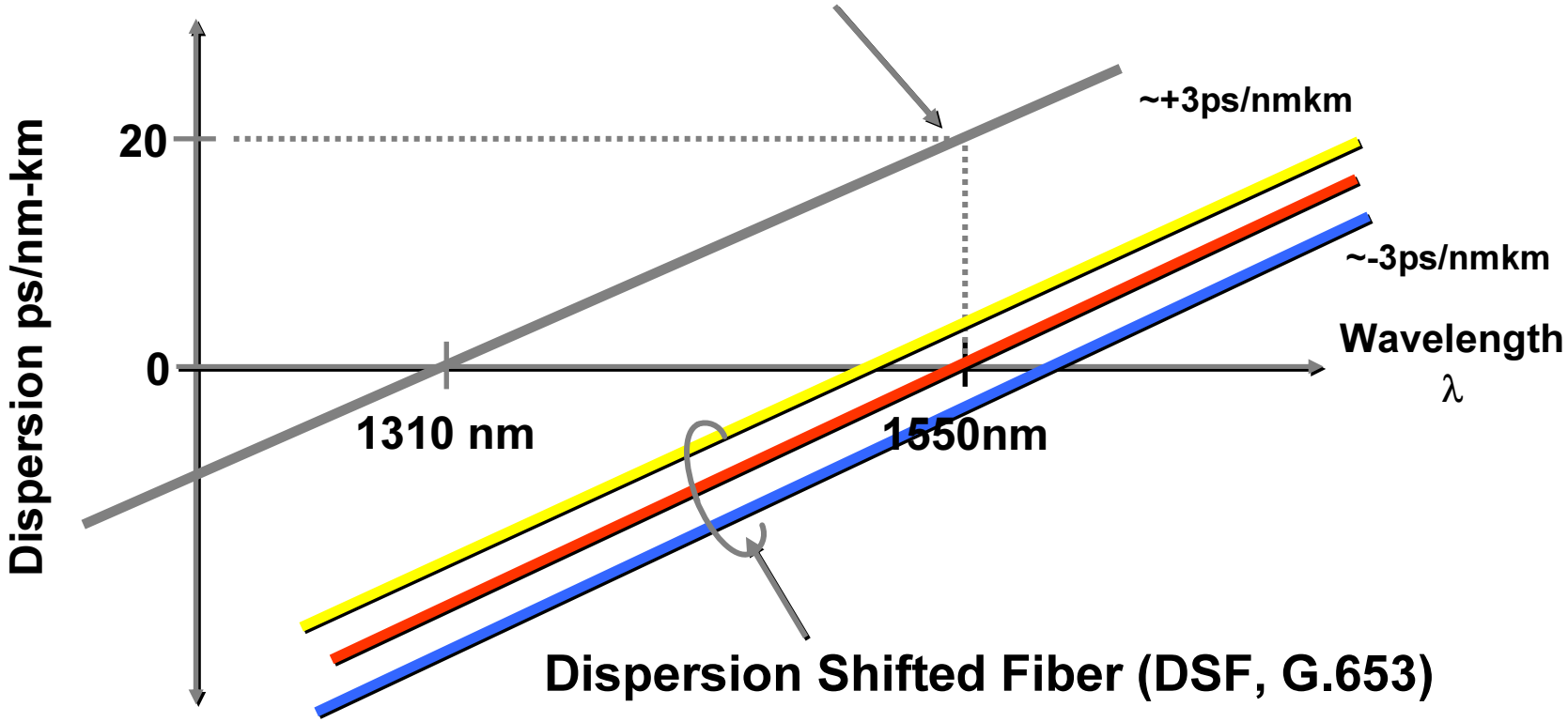
Chromatic Dispersion (CD)



- The optical pulse tend to spread as it propagates down the fiber, limiting either the bit rate or the maximum achievable distance at a specific bit rate
- Chromatic Dispersion = Material Dispersion + Waveguide Dispersion
- Chromatic Dispersion is a fiber characteristic (D=17 ps/nm.km), but also depends on the light source
- Physics background: The refractive index has a wavelength dependent factor

Fiber Classification

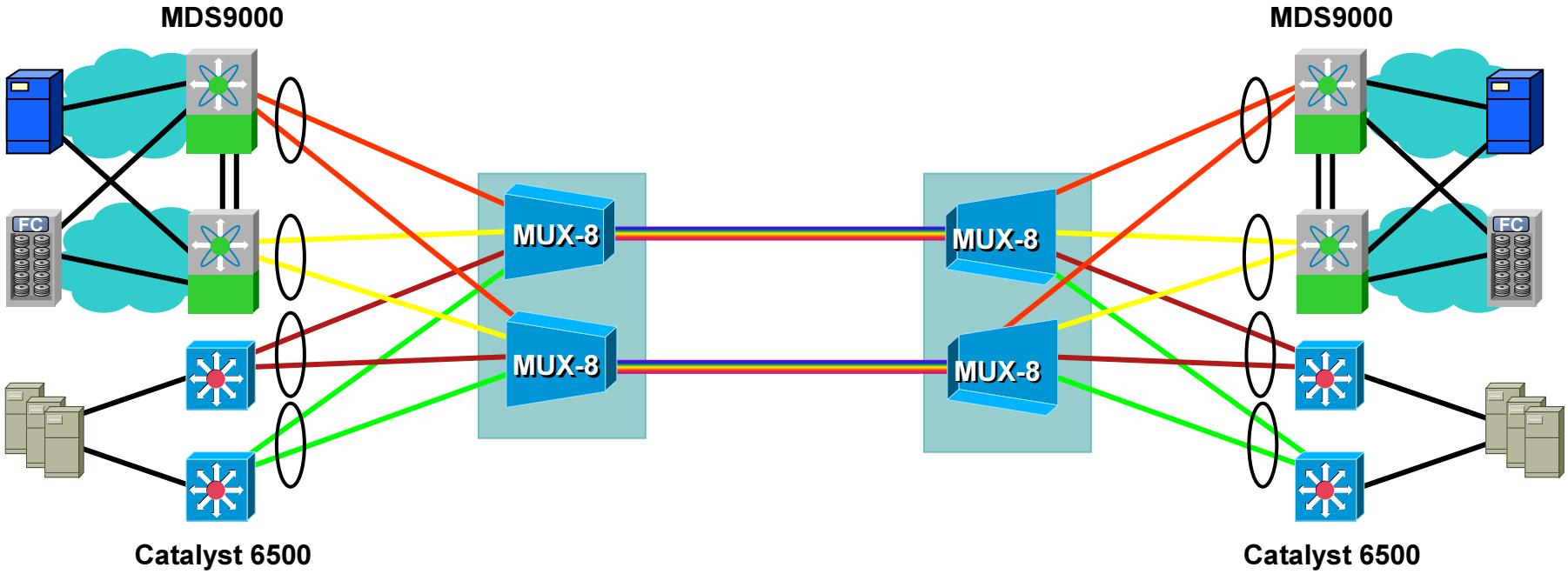
Normal Single Mode Fiber
(SMF-28, G.652) >95% of Deployed Plant
(new SMF-28e. G.652.C, ZWPF)



CWDM/DWDM designs for DC interconnect

Design scenarios

DR solution with integrated WDM optics



- Uses colored interfaces (GBIC, SFP, XENPAK, X2, XFP etc.) in CWDM or DWDM wavelength grid plugged directly in communication devices (ethernet or FC switches) and passive DWDM or CWDM filters
- Lower cost than transponder based system but less functionality
- Can be combined with transponder based solution

Cisco CWDM filters and pluggable modules

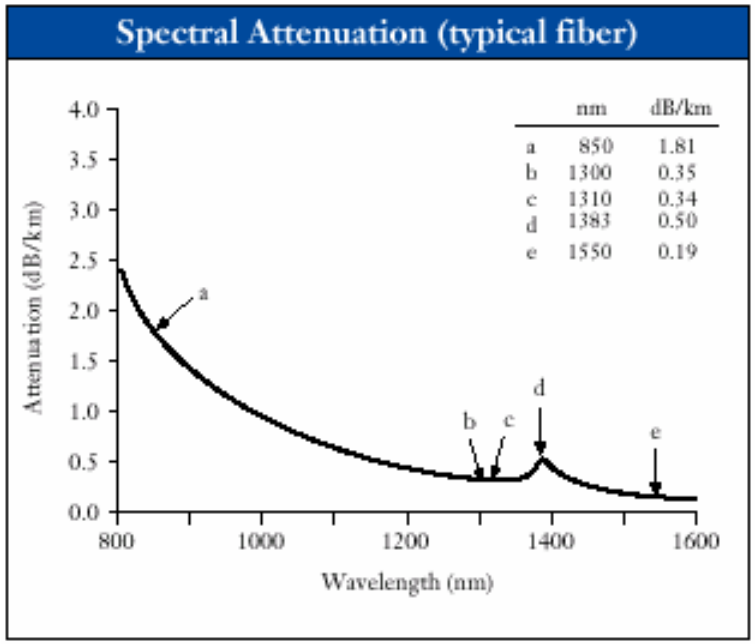
Passive CWDM solution

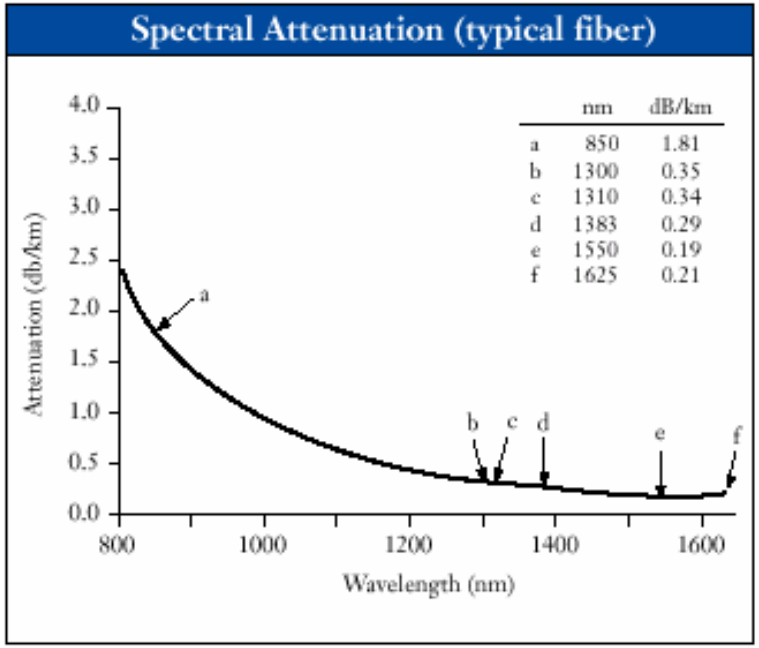
- ✓8 channels (20nm spacing, 1470nm-1610nm)
- ✓low insertion loss for new generation modules
- ✓ring, bus, p-t-p
- ✓GBICs, SFPs and passive filters
- ✓GE, FC (1G, 2G and 4G)
- ✓Optical budget
  - min. 30 dB for CWDM GBICs (GE)
  - min. 28 dB for CWDM SFPs (2G FC)
  - min. 16 dB for CWDM SFPs) (4G FC)

- ✓Single fibre solution - 4 channels
- ✓Standalone 2 channel transponder for cases where colored interface is not directly supported in device

ITU-T G.652.C Zero Water Peak Fiber (ZWPF)

- Designed to support CWDM application with more than 8 channels, the ZWPF removes the absorption peak around 1383 nm
- From a Chromatic Dispersion standpoint it behaves as a standard G.652 fiber





Source: Corning data sheets for SMF28™ and SMF28e™

Cisco DWDM filters and pluggable modules

Passive DWDM solution

- ✓32 channels (100GHz spacing)
- ✓Cisco ONS 15216 Flexlayer family
- ✓ring, bus, p-t-p
- ✓GBICs, SFPs, XENPAKs and passive filters
- ✓GE, FC (1g and 2g) and 10GE

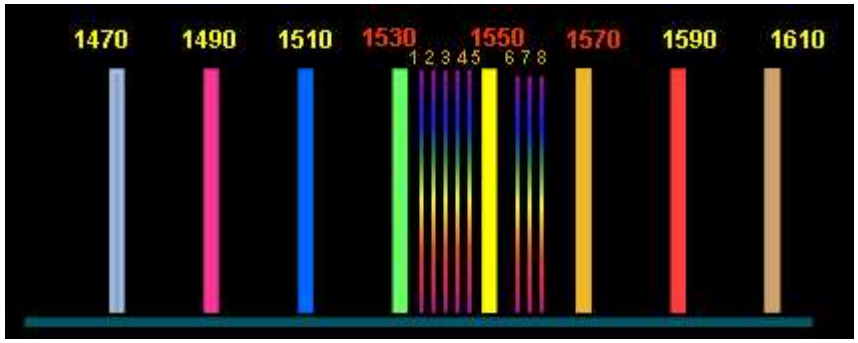
- ✓other colored interfaces in DWDM grid
- ✓Optical budget
  - min. 28 dB for DWDM GBICs
  - min. 23 dB for DWDM XENPAKs

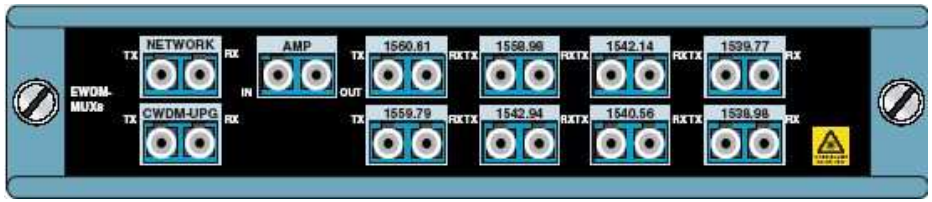
- ✓Single fibre solution - 16 channels
- ✓Standalone 2 channel transponder for cases where colored interface is not directly supported in device

Cisco eWDM filters

Combination of CWDM and DWDM

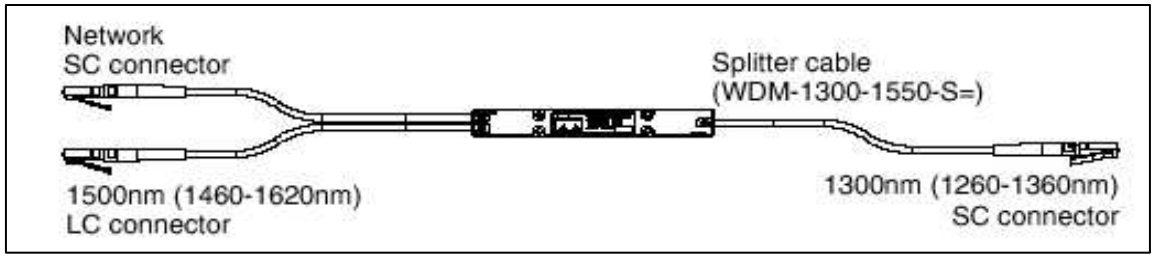
- ✓16 channels (8 CWDM + 8 DWDM)
- ✓upgrade of existing CWDM networks to 16 channels
- ✓upgrade to 10Gbps
- ✓ mux/demux, OADM (2 and 4 channels)
- ✓EDFA for DWDM channels
- ✓CWDM and DWDM pluggable modules
- ✓Standalone 2 channel transponder for cases where colored interface is not directly supported in device



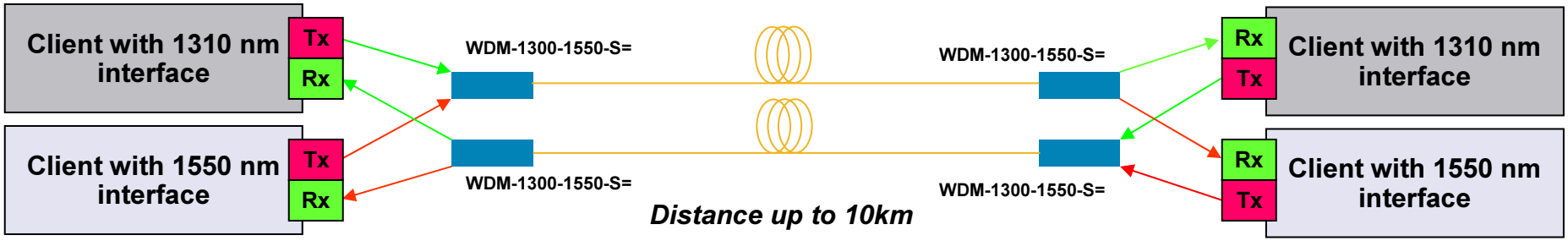


Design with 1310/1550 nm splitter cable

- Two unprotected clients, one using 1310nm and second 1550nm
- Simplest WDM 'device' - falls into WWDM category
- Low insertion loss, Simplex cable - two needed per site
- Can be combined with CWDM/DWDM filters connected to 1550nm port

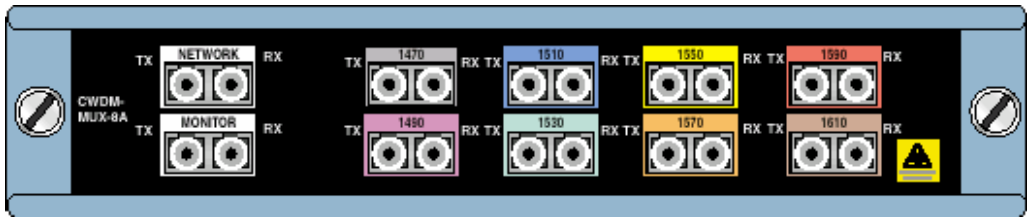


Parameter	Path	Minimum	Maximum	Unit
1310 Insertion Loss	ADD/DROP	-	1.2	dB
1550 Insertion Loss	ADD/DROP	-	0.8	dB

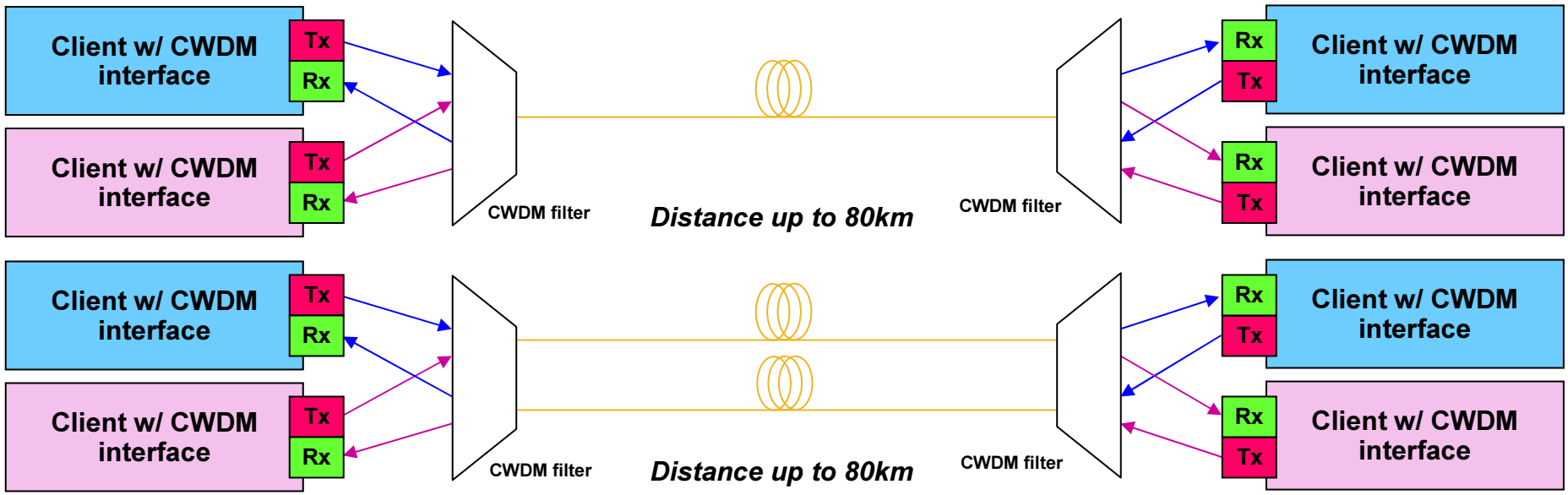


Design with passive CWDM filters

- Support for 8 unprotected channels over fiber pair or 4 unprotected channels over single fiber strand

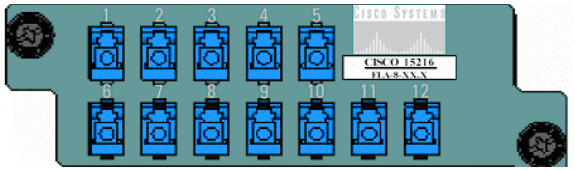


Insertion loss	Add	Drop	Pass
MUX 8	2.2	2.2	-
OADM 4	1.8	1.8	2.1
OADM 1	1.5	1.5	1.5
Single fiber	3.0	3.0	-



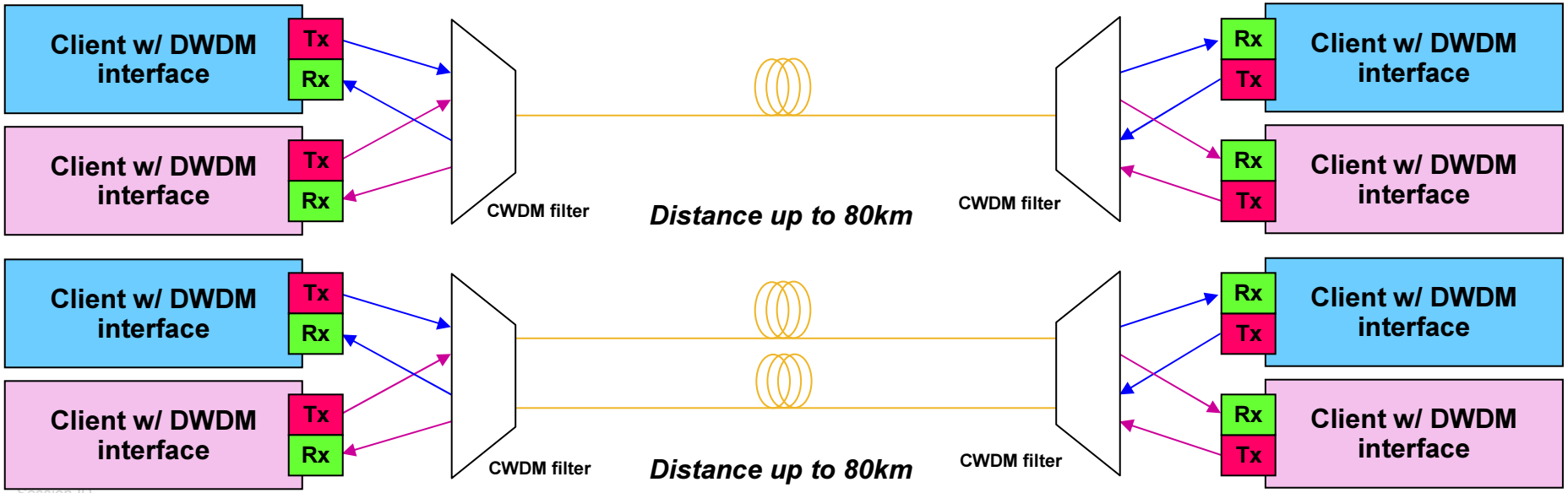
Design with passive DWDM filter

- Support for up to 32 unprotected channels over fiber pair or up to 16 unprotected channels over single fiber strand



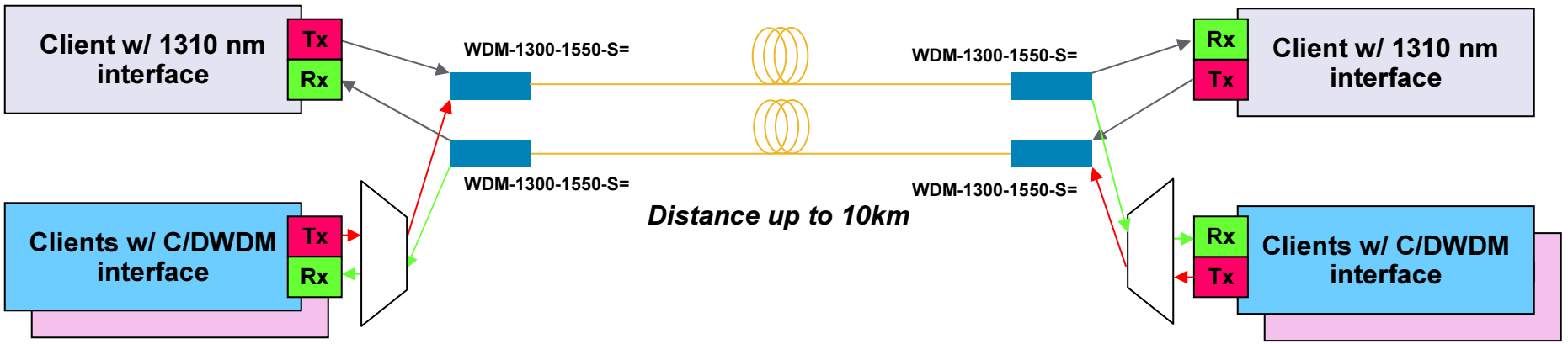
Available modules

- 8 and 2 channel OADM (two required per site)
- 4 band splitter/combiner (two required per site)
- 2,3,4 way splitter/combiner
- VOA module



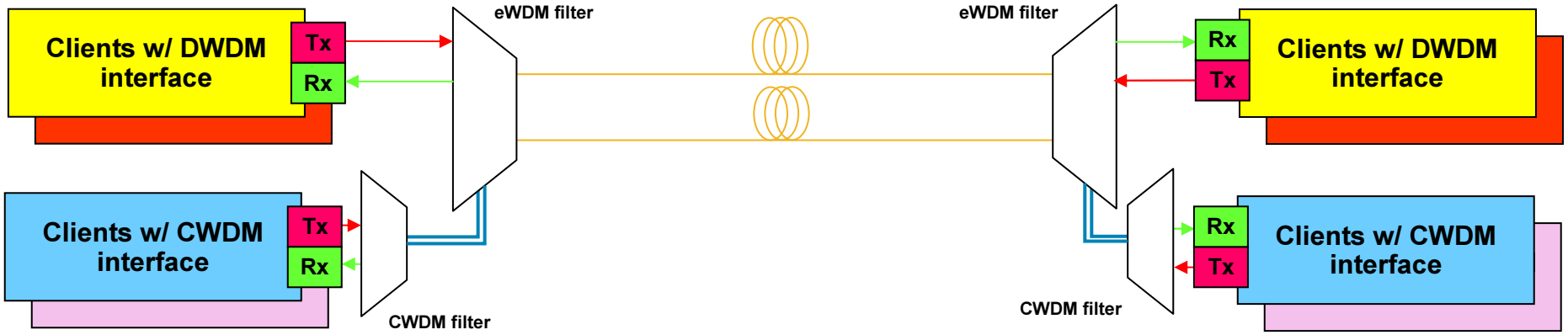
Design with 1310/1550 nm splitter cable in combination with C/DWDM filter

- Combination of one client using 1310nm and others using CWDM or DWDM wavelengths (up to 8 channels with CWDM, up to 32 channels with DWDM)
- Allows support of 'legacy' 1310nm channel, combine 10GE with CWDM, etc.
- Fiber pair only, no single strand



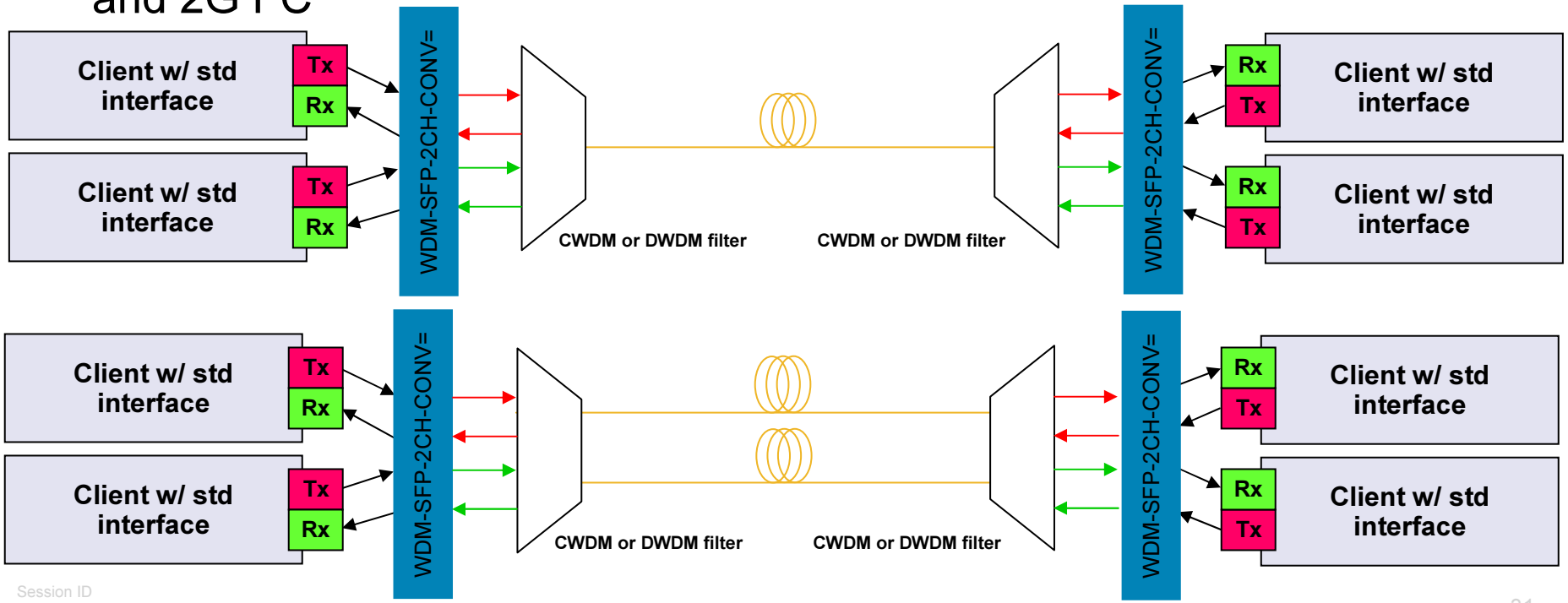
Design using eWDM solution

- Combination of 8 channels in CWDM grid and 8 channels in DWDM grid
- Cost effective upgrade of CWDM networks
- Cost effective option for small (approx. 40%less for 8 channel DWDM network comparing to 15216 Flexlayer)

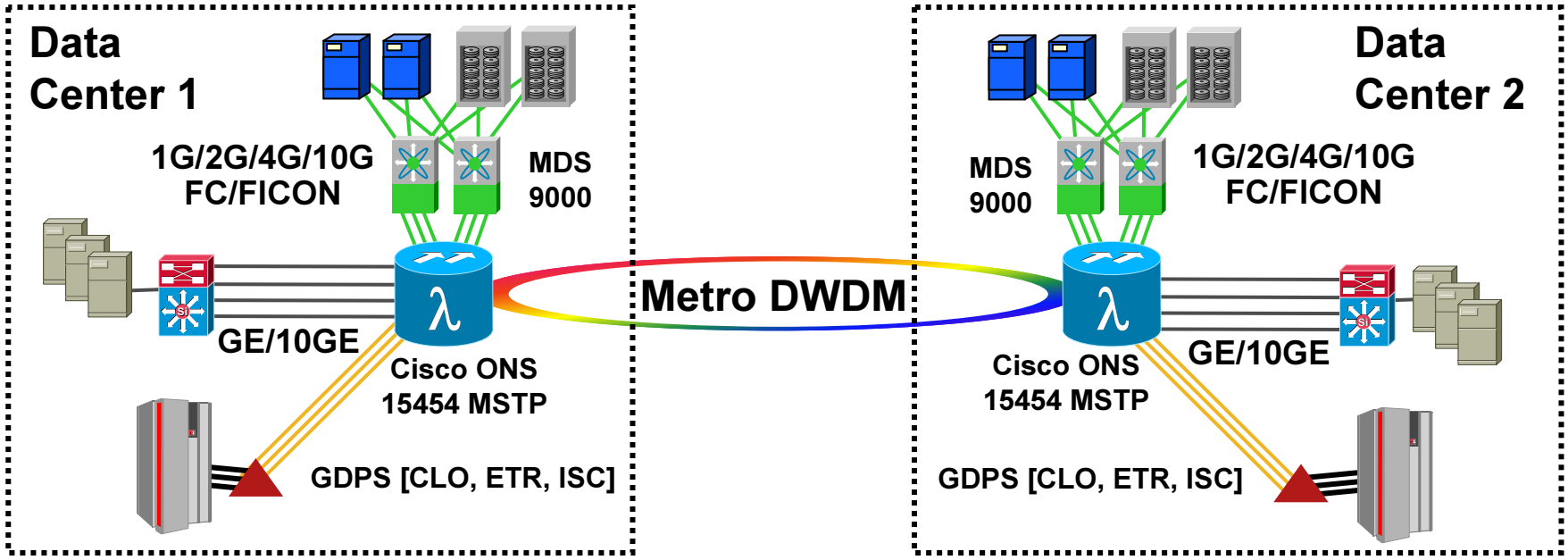


Complementing designs with standalone transponder unit

- When CWDM or DWDM interface is not directly available at client than WDM-SFP-2CH-CONV= can be used to convert signal from 850 nm or 1310 nm to CWDM or DWDM grid on both sides. One WDM-SFP-2CH-CONV= can handle two channels
- WDM-SFP-2CH-CONV= requires 'grey' SFP in client port and C/DWDM SFP in network port and supports STM-1, STM-4, STM-16, GE, 1G FC and 2G FC



DR solution with transponder based DWDM

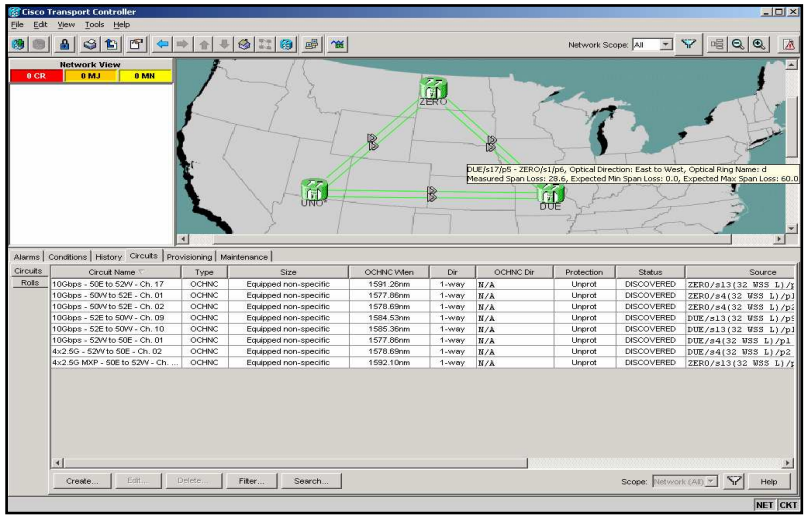


- Support of many different channel types: GE, 10GE, FC/FICON (1/2/4/10G), SDH (STM-1/4/16/64/256), ESCON, IBM solution specific interfaces (CLO, ETR, ISC), video interfaces, 2R transparent signal etc.
- Cost-effectively aggregates data and storage services into 2.5 or 10 Gbps lambda
- Buffer-to-buffer credits for distance extension
- Optical performance and comprehensive protocol (payload) monitoring

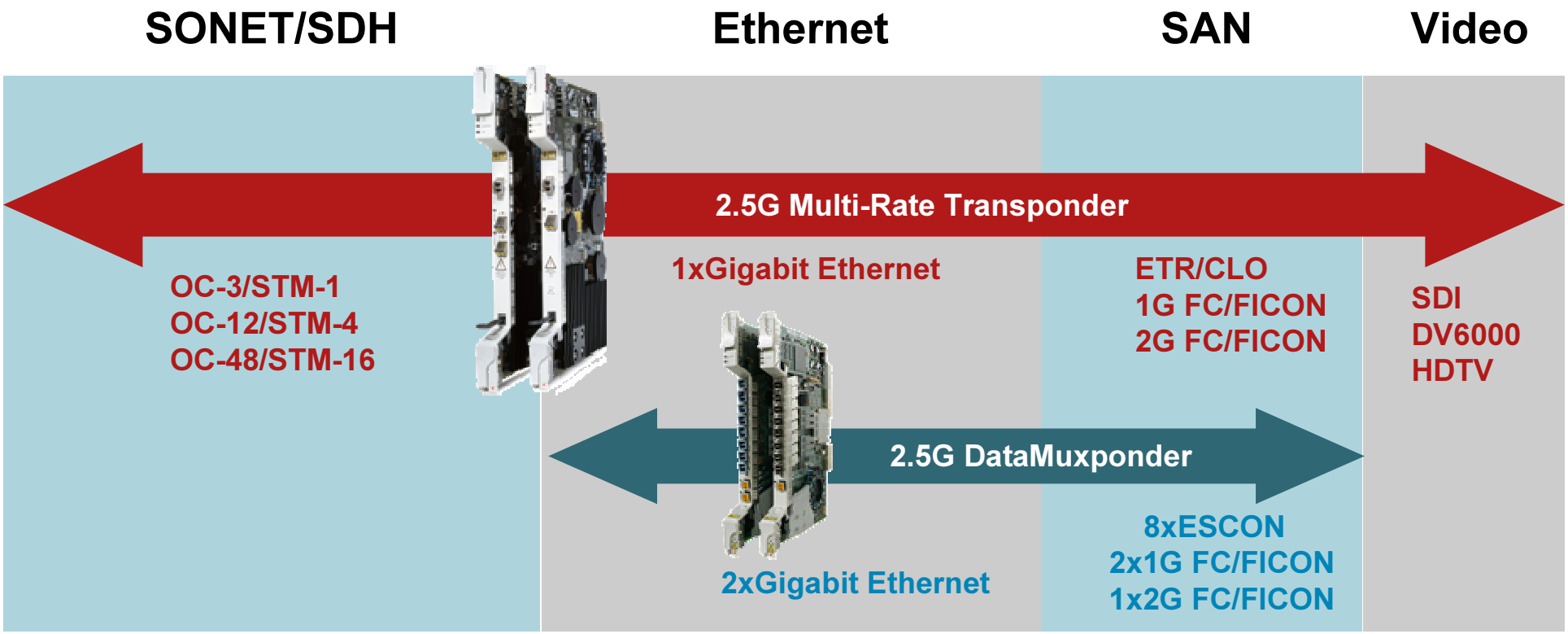
Cisco universal optical platform

Cisco ONS 15454E MSTP

- ✓*32/40/80 wavelengths (expandable to 112)*
- ✓*ring, bus, p-t-p, mesh*
- ✓*50/100GHz spacing, C-band/L-band*
- ✓*fixed OADMs and ROADMs*
- ✓*various traffic protection schemes*
- ✓*Transponders:*
  - *8Mbps-2.5Gbps*
  - *10GE,STM64,10GFC*
  - *STM-256 support*
  - *CLO/ETR*
- ✓*Service aggregation (muxponding):*
  - *1/2/4G FC/FICON, ISC1/ISC3;*
  - *GE, ESCON, 1G FC/FICON;*
  - *STM-16*
- ✓*Integrated EDFAs with DCU support*
- ✓*support for FEC and EFEC*
- ✓*more than 2000 km without regeneration*
- ✓*enterprise/regional/SP*
- ✓*certified by major storage vendors*

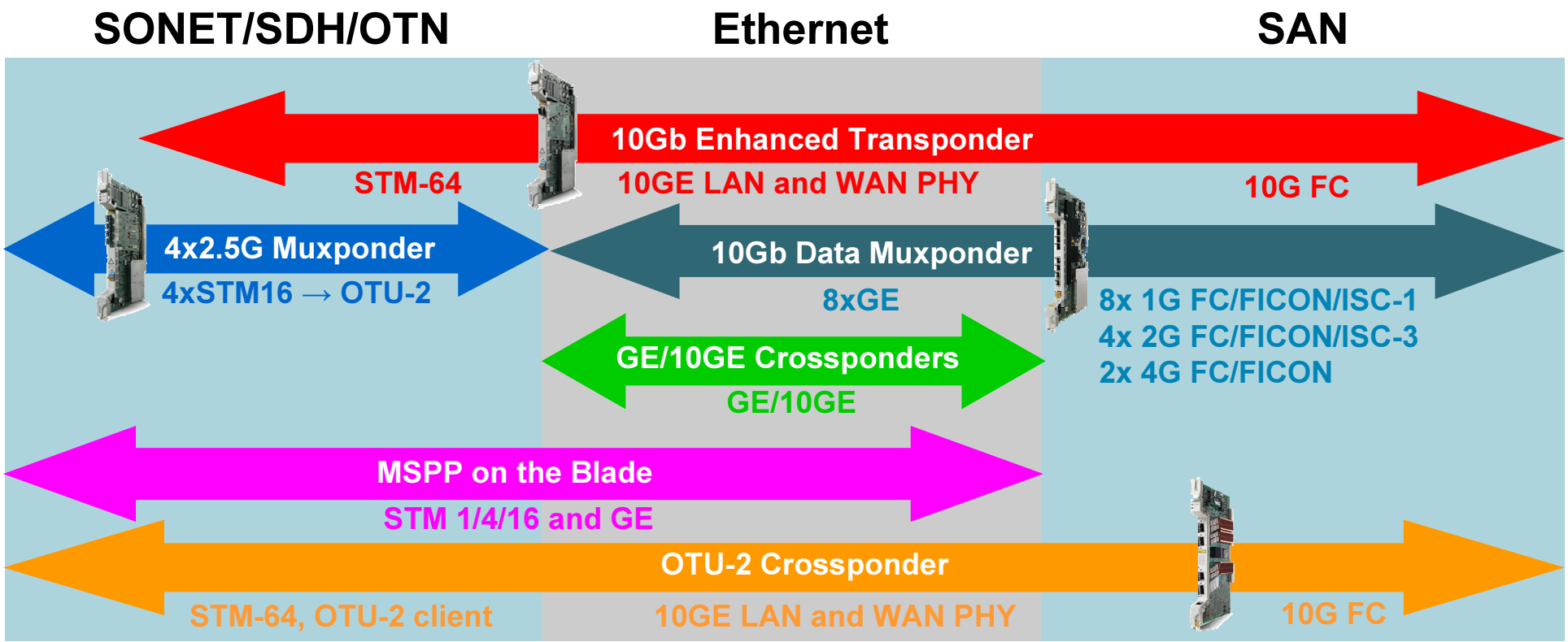


Cisco ONS 15454 MSTP
2.5Gbps Service Cards



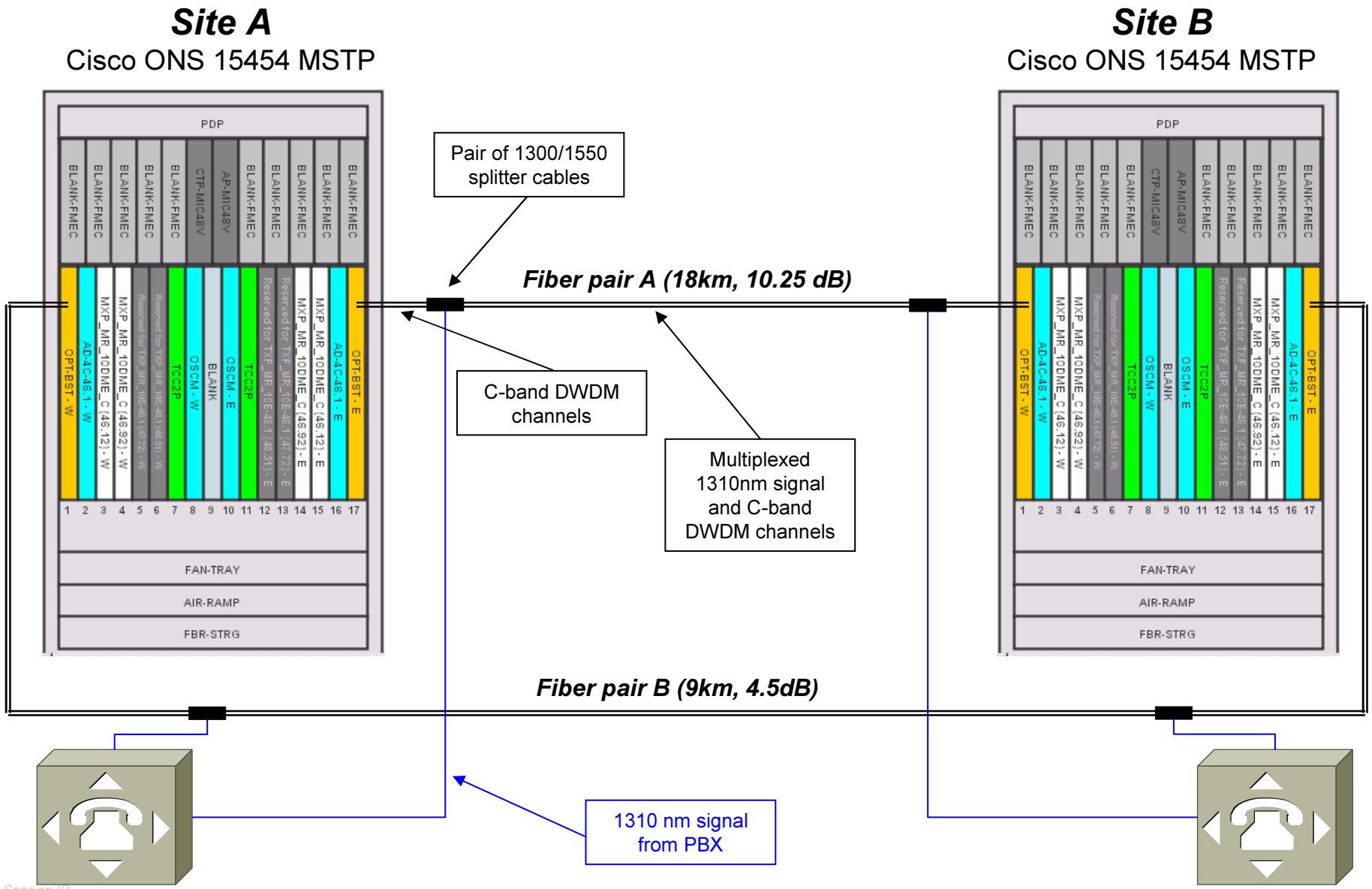
- Simple planning, sparing, and ordering with multi-rate, multiprotocol and pluggable optics
- Optical and payload monitoring
- FEC support at 2.5Gbps transponder
- G.709 support
- Client 1+1, Y-cable and splitter protection

Cisco ONS 15454 MSTP
10Gbps Service Cards



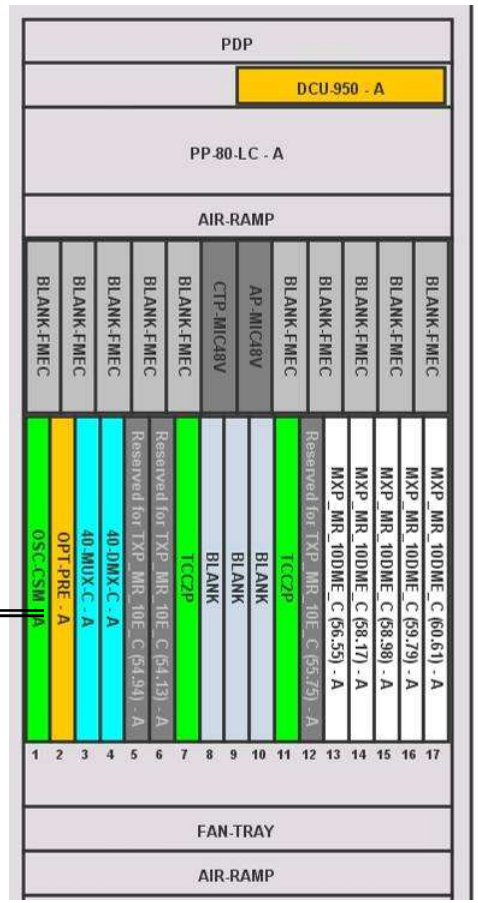
- All 10G applications covered by 1 transponder,
- Aggregation cards reduce the cost of service delivery
- Full C-band or L-band tunability - 80 channels @ 50GHz spacing (*except crossponders)*
- FEC and EFEC support (G.975, G.975.1), G.709 support
- Optical and payload monitoring, Client 1+1, Y-cable protection and 'splitter' (XP)

Cisco ONS 15454 - example 1 (bank)



Cisco ONS 15454 - example 2 (insurance)

- Two identical systems



Site A
Cisco ONS 15454 MSTP

59km @ 0.25dB/km

Site B
Cisco ONS 15454 MSTP

Cisco ONS 15454 interfaces

TDM	Data	Storage	Video	2R
•STM-1 •STM-4 •STM-16 •STM-64 •STM-256 •E1, E3 (MSPP integration)	•FE •GE •10GE LAN PHY •10GE WAN PHY	•1G FC/FICON •2G FC/FICON •4G FC/FICON •10G FC •ESCON •ISC 1 •ISC 3 •Sysplex CLO •Sysplex ETR	•DV-6000 •HDTV •SDI •D1 video •DVB ASI	•Any rate from 100 Mbps to 2.5 Gbps

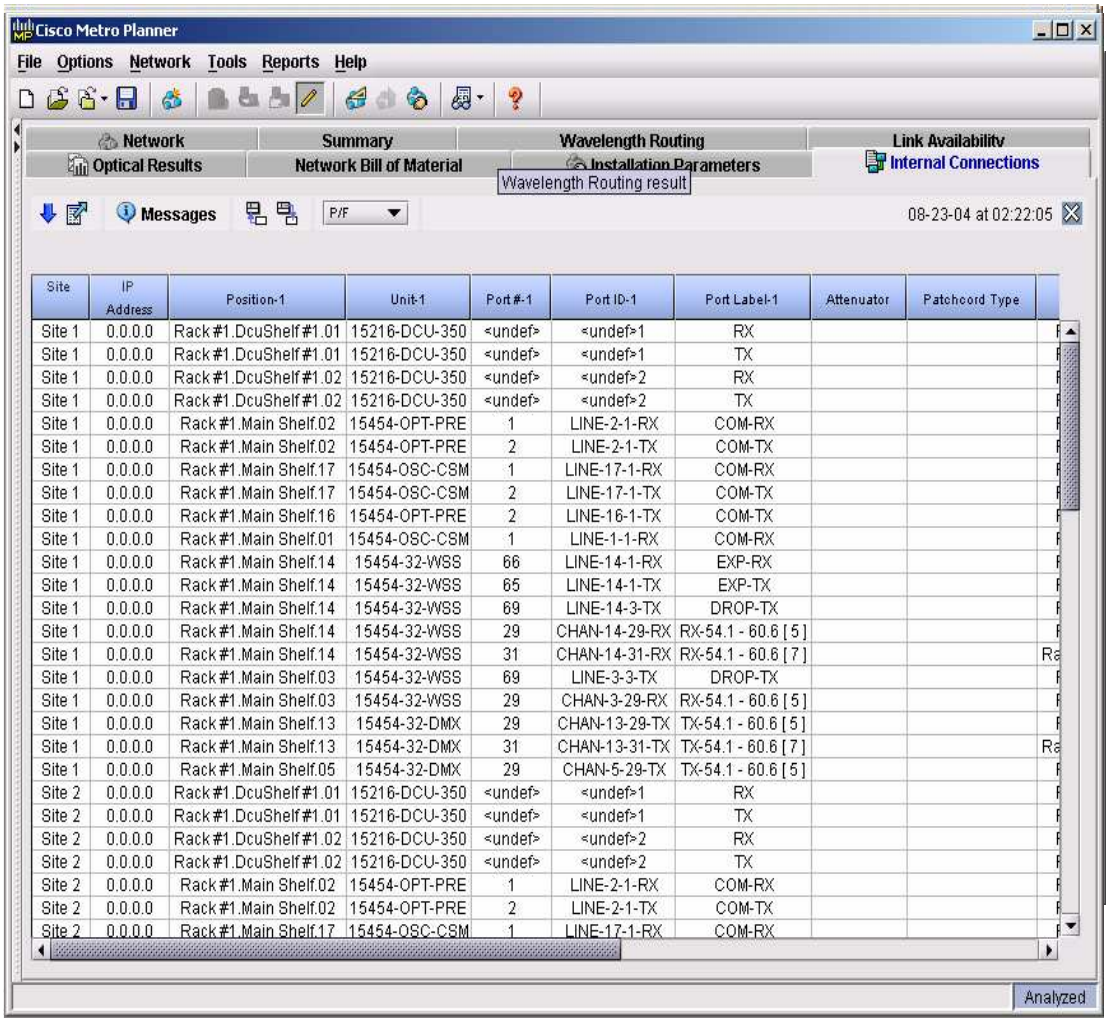
- High flexibility in system deployment, most of applications covered
- Broad range of potential service offerings
- 40Gbps support allows for further bandwidth scaling

Pluggable client interfaces

Type/category	Example
Grey optics 850 nm SFP	1000BaseSX, MMF FC clients
Grey optics 1310 nm SFP	1000BaseLX, SDH SR/IR clients, FC SMF clients
Grey optics 1550 nm SFP	1000BaseZX, SDH LR clients
Grey optics 1310 nm XFP	10GBaseLR, 10G FC, STM-64
Grey optics 1550 nm XFP	10GBaseER/EW, STM-64 LR
CWDM client optics	GE, 1/2G FC, STM-16 *(release 8.5)*
DWDM client optics	GE, 1/2G FC, STM-16 *(release 8.5)*
Metalic client SFP	10/100/1000BaseT for GE Xponder

- **Lower opex through common sparing with other Cisco products**
- **Per port reach and rate selection**
- **Tight integration of CWDM and DWDM from network perimeter**
- **High transponder reusability for different services**

Cisco Transport Planner

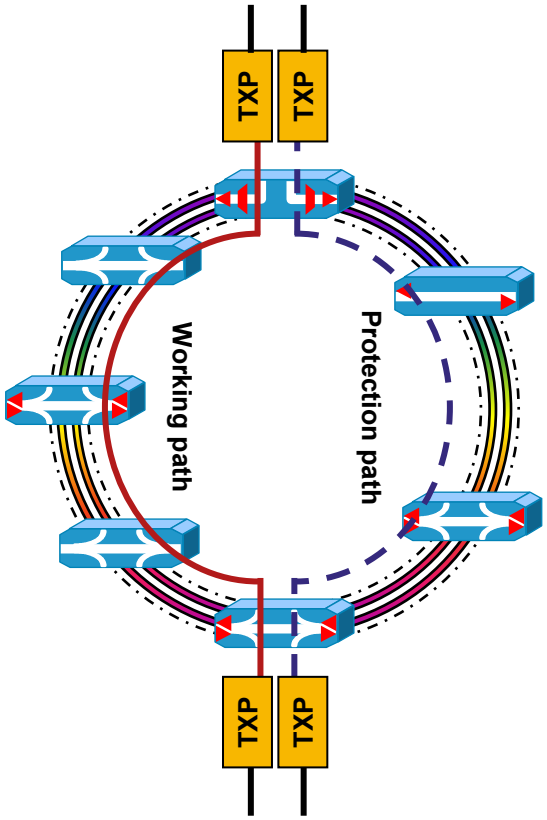


- **GUI-based Network Design Entry**
- **Traffic requirements:**
  - **Any-to-Any Demand provided by ROADM**
  - **Point-to-point demands**
- **Comprehensive Analysis checks for:**
  - **wavelength routing and selection**
  - **optical budget and OSNR**
  - **CD, PMD, amplifier tilt etc.**
- **Smooth Transition from Design to Implementation**
  - **Bill of Materials**
  - **Rack Diagrams**
  - **Step-by-Step Interconnect**

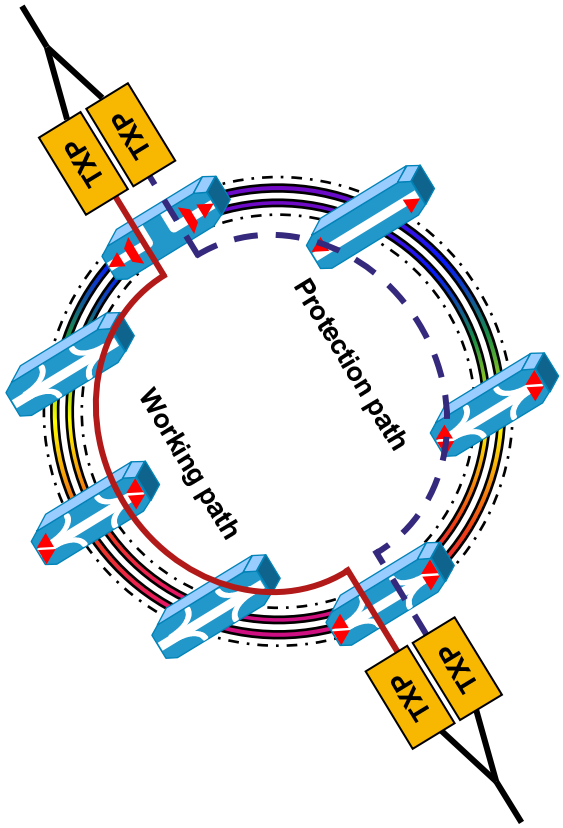
Channel protection mechanisms

- Protection mechanisms
  - Splitter based protection
  - Y-cable based protection
  - Client based protection
- Protection signalling
  - Optical/analogue parameters—LOS/LOL
  - Digital performance monitoring parameters— application-specific
  - Client-side/trunk-side
- Path switching
  - Uni-directional
  - Bi-directional

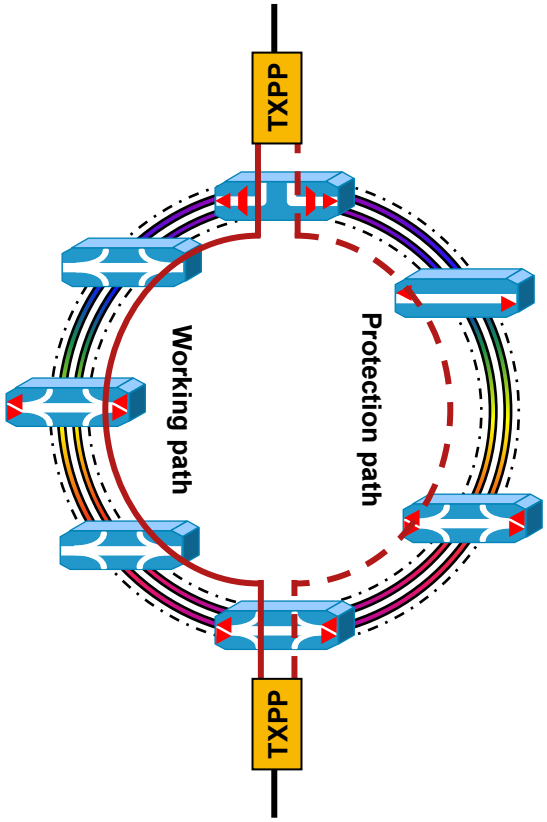
Cisco ONS 15454MSTP Protection Schemes



Path & Equipment Protection (Client protection)



Path & Equipment Protection (Y-Cable protection)

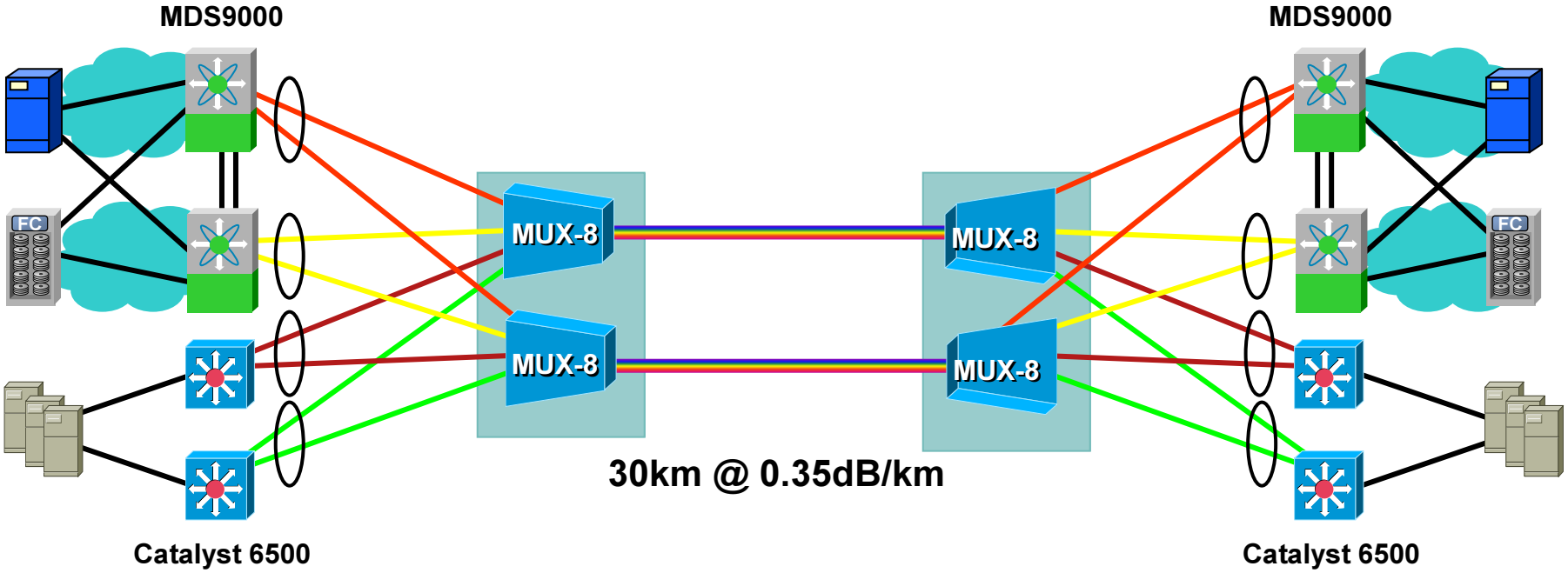


Path Protection (Splitter protection)

CWDM/DWDM designs for DC interconnect

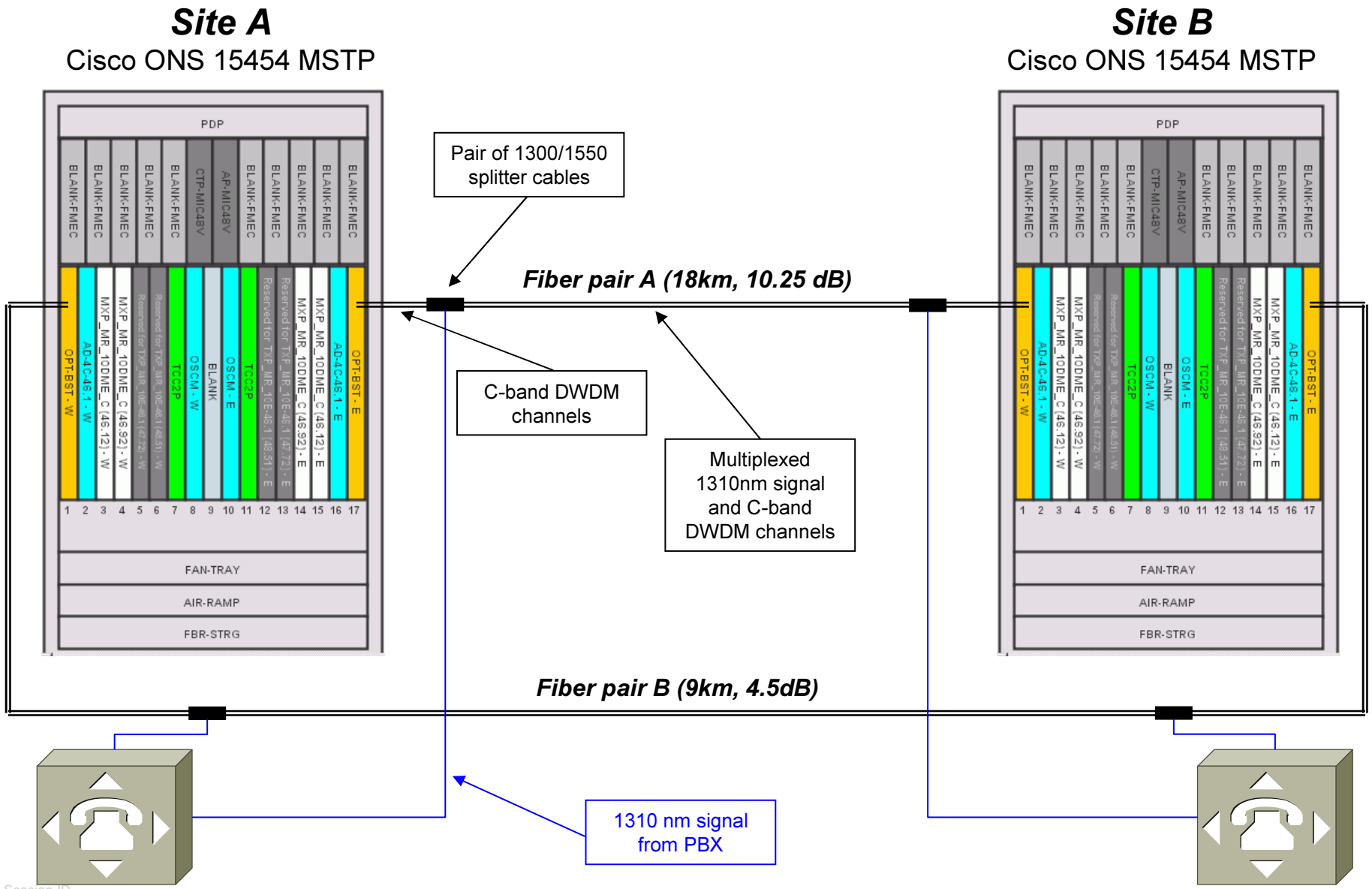
Solution examples

CWDM design example



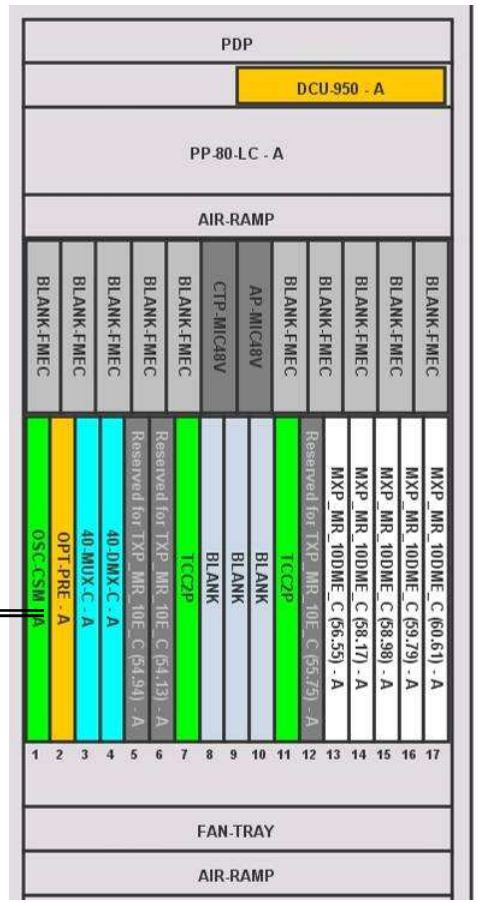
- GE and 2G FC ports, 8 channel mux/demux - insertion loss 2.2 dB
- TX (min): 0 dBm, TX (max): +5 dBm
- RX (min): -29dBm@1Gbps, -28dBm@2Gbps
- Dispersion penalty: 2dB@1Gbps, 3dB@2Gbps
- Connector margin: 0.5 dB, Aging margin: 1dB
- Worst case budget: 0 - (-28) -2*2.2 - 3 - 2*0.5 - 1 = 18.6 dB
- Span attenuation: 30*0.35 = 10.5 dB => design is OK

Cisco ONS 15454 - example 1 (bank)



Cisco ONS 15454 - example 2 (insurance)

- Two identical systems



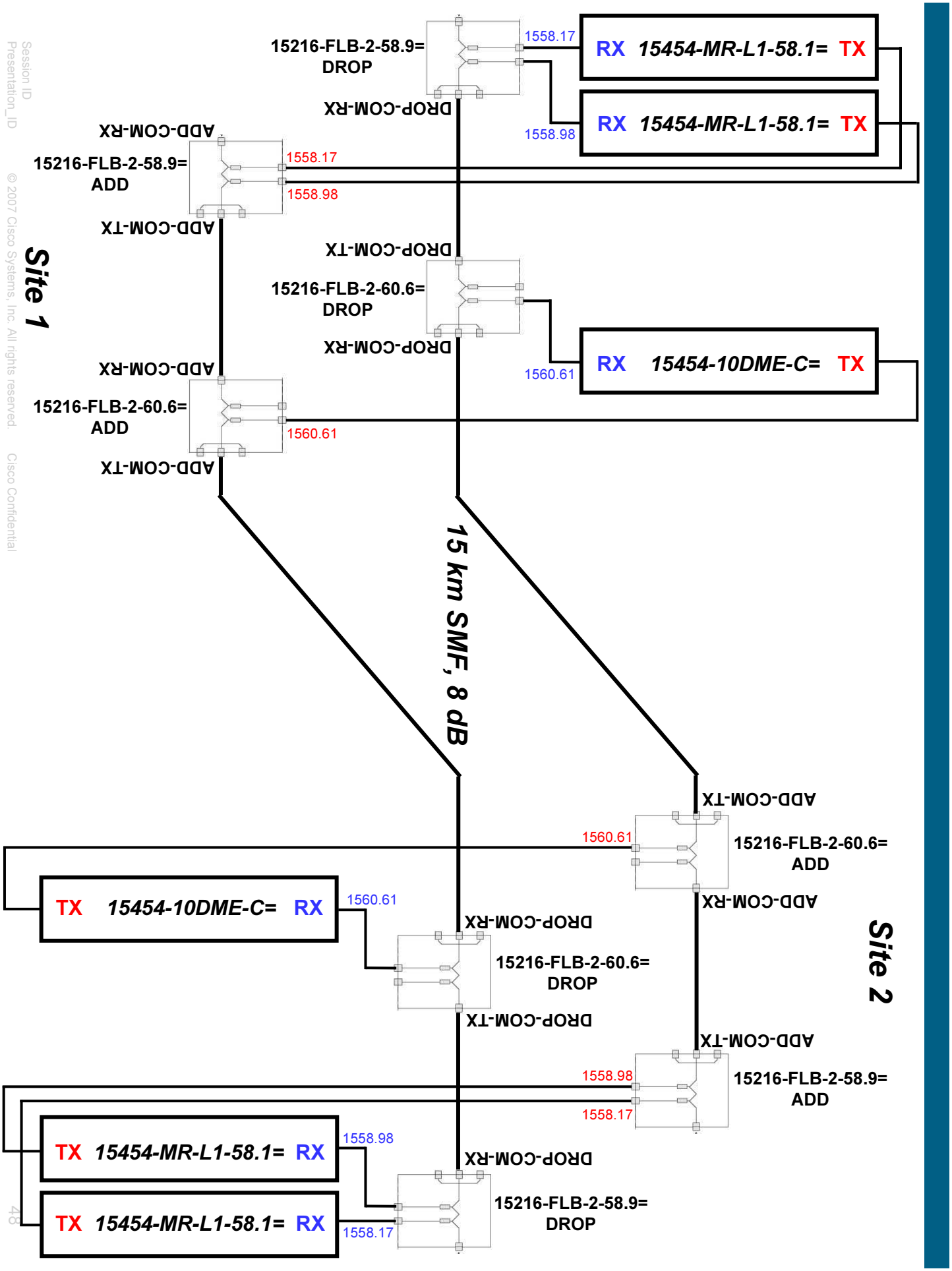
Site A
Cisco ONS 15454 MSTP

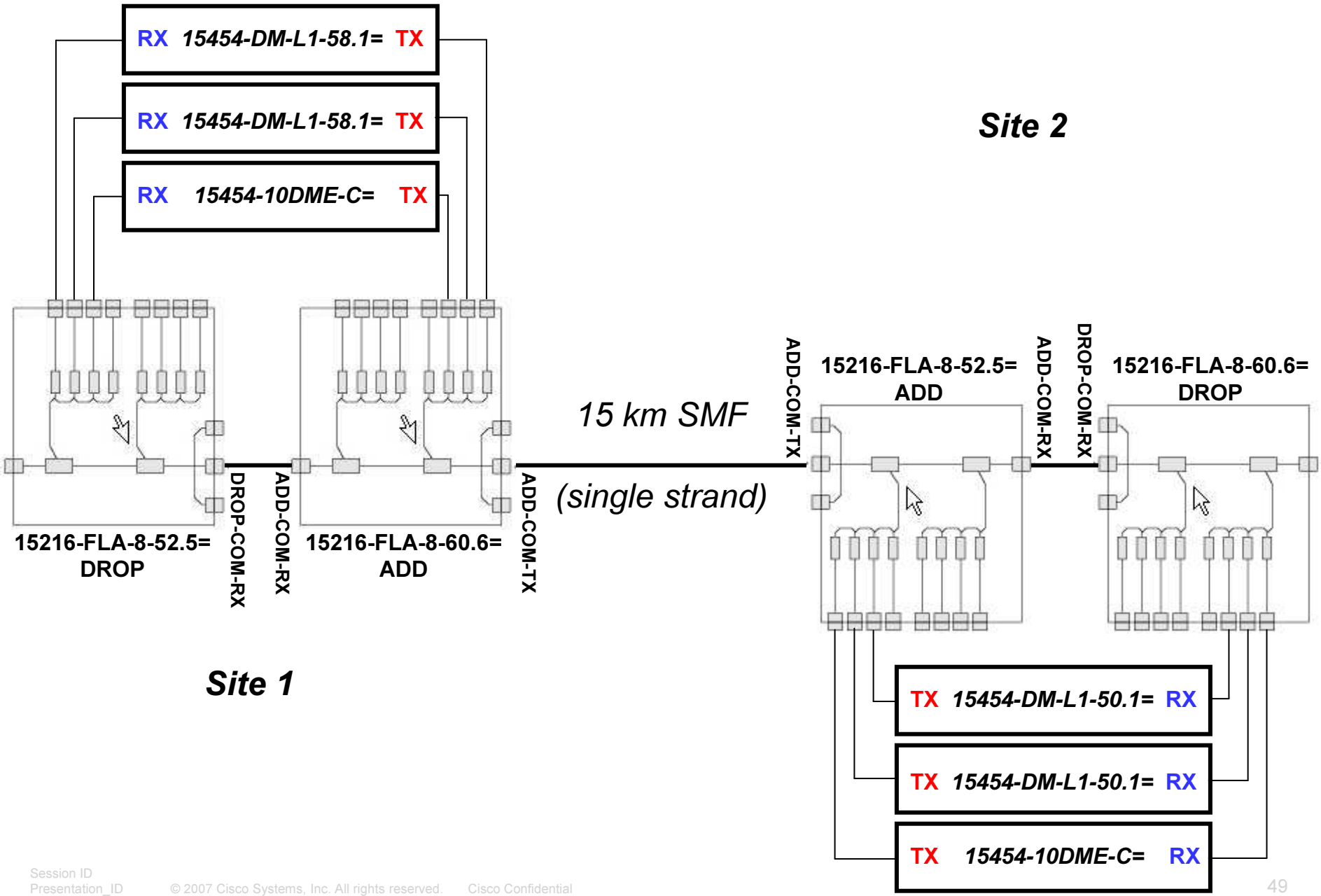
59km @ 0.25dB/km

Site B
Cisco ONS 15454 MSTP

Cisco ONS 15454/15216 Hybrid examples

- Unidirectional
- Bidirectional





CWDM/DWDM designs for DC interconnect

Protocol and application interaction with transport systems

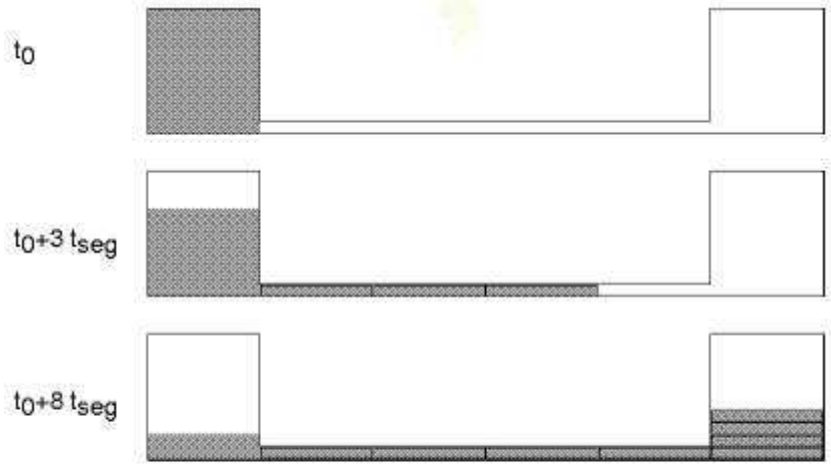
Why the distance is important and where it matters

- ✓ ***Optical transport network***
  - ✓ usually not limit – up to 2000 km without regeneration for 10GE or 10G FC
- ✓ ***Data network***
  - ✓ latency itself is usually not serious problem
  - ✓ too small default TCP window will cause performance degradation – must be increased
- ✓ ***Application***
  - ✓ for 'chatty' application protocols (like CIFS) cumulative latency is usually serious issue
  - ✓ WAAS technology can be used to optimize even traffic between datacenters
- ✓ ***SAN***
  - ✓ latency is serious problem when synchronous replication is used
  - ✓ skew can be an issue for some load balancing schemes (frame based)
  - ✓ buffer-to-buffer flow control can limit performance for long distance (use switches with large BB count or DWDM with BB spoofing)
  - ✓ FEC/EFEC cause additional latency - switch it off if not needed

TCP/IP in Long Fat Networks

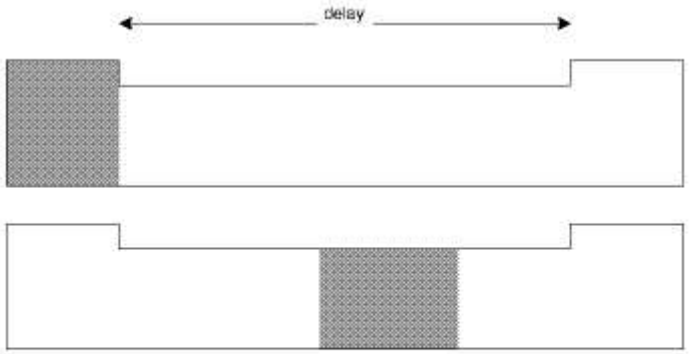
*Bandwidth*Delay product*

Slow Network



- window size is 10 KB
- every segment is 1 KB,
- capacity of the channel is 8 KB

Long Fat Network



- window size 10 KB
- 1 KB segments
- capacity of the channel is 32 KB

	5 km LAN	*100 km MAN*	*2000 km WAN*
Bytes in the Pipe:			
10 Mbit/s Ethernet	52	1040	20800
155 MBit/s ATM	703	14062	281250
1000 MBit/s GE	5208	104167	2083330

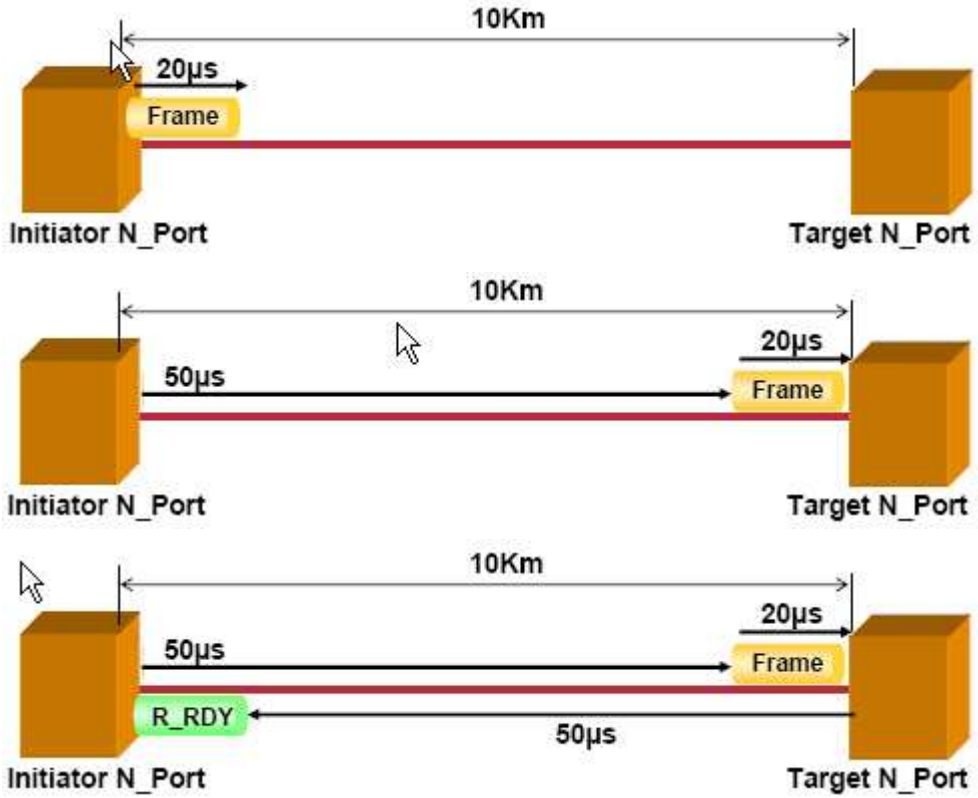
Typical Windows TCP/IP window size:

- **bellow 1 Mbps = 8kB, 1 Mbps-100 Mbps = 17kB, greater than 100Mbps = 64kB**
- **if larger is needed use window scaling option**

Flow control in FC SAN

How many BB credits do I need?

```
Credits = (Round_Trip_Time + Processing_Time) / Serialization_Time
```



Step 1: Serialization
Link rate is 1.0625 Gbps => 9.41ns/byte
1byte = 10 bits because of 8b/10b coding
Max. frame size = 2148 bytes
Serialization Time ≈ 20 µs

Step 2: Transmission and deserialization
Propagation delay ≈ 5 µs/km
Caused by speed of light in optical fibre
Time to transmit frame over 10 km ≈ 50 µs
Deserialization Time ≈ 20 µs

Step 3: Processing and confirmation
Processing is neglected in this example
R_RDY generated and sent back
Time to transmit frame over 10 km ≈ 50 µs
Deserialization Time ≈ 20 µs

Step 4: Calculation of required BB credits
Total RTT = 50+20+50=120 µs
Credits (10km) ≈ 120 / 20 = 6 (for 1 Gbps FC)
Credits (10km) ≈ 110 / 10 = 11 (for 2 Gbps FC)
Credits (10km) ≈ 105 / 5 = 21 (for 4 Gbps FC)

Fiber channel flow control
How many BB credits do we need?

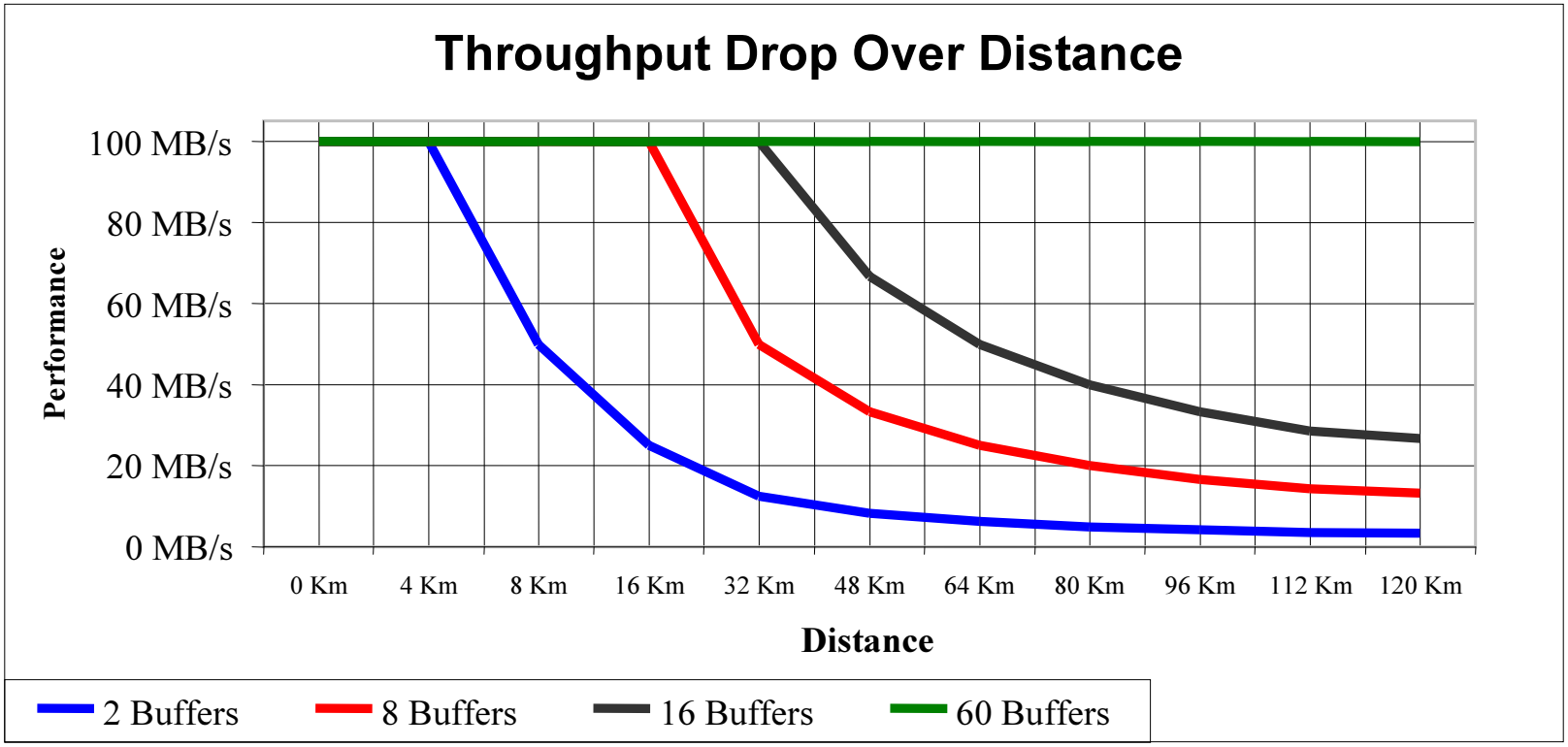
- 1G FC: 1 BB for 2 km at max frame size (2148 Bytes)
- 2G FC: 1 BB for 1 km at max frame size (2148 Bytes)
- 4G FC: 2 BB for 1 km at max frame size (2148 Bytes)
- 10G FC: 6 BB for 1 km at max frame size (2148 Bytes)

Example: 255 B2B credits available per port:

Maximum distance of 127 Km at 4 Gbps with maximum frame size packets (2,148B)

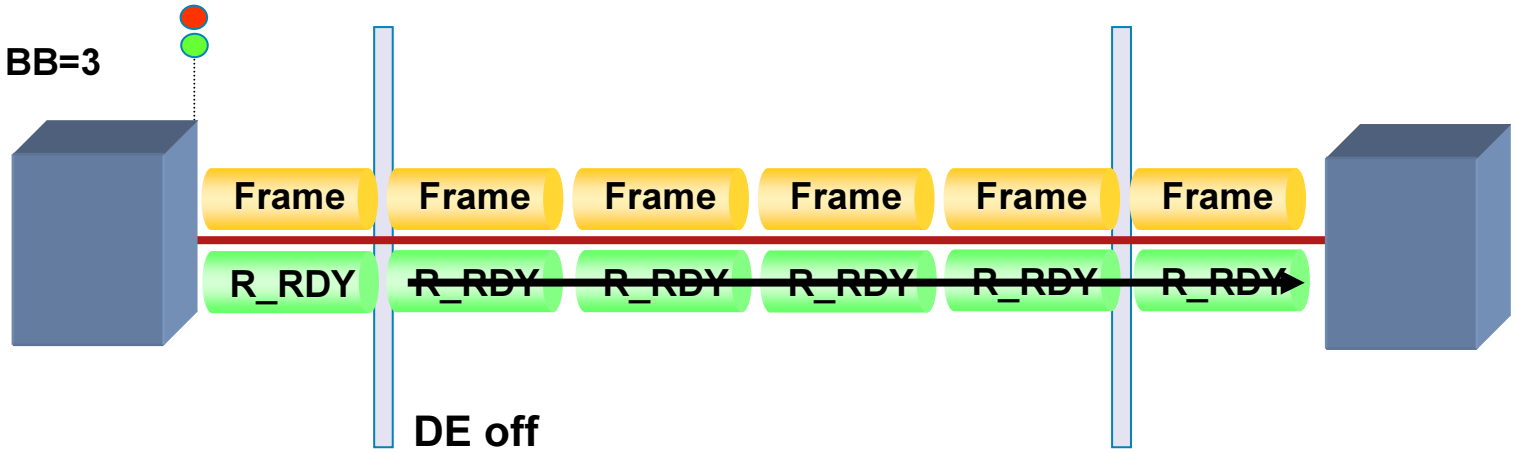
$$\frac{255}{2}\ \frac{credits}{\frac{credits}{km}\ @\ 4\ \ Gbps} \approx 127\ Km$$

Fiber channel flow control
What if I do not have enough credits?



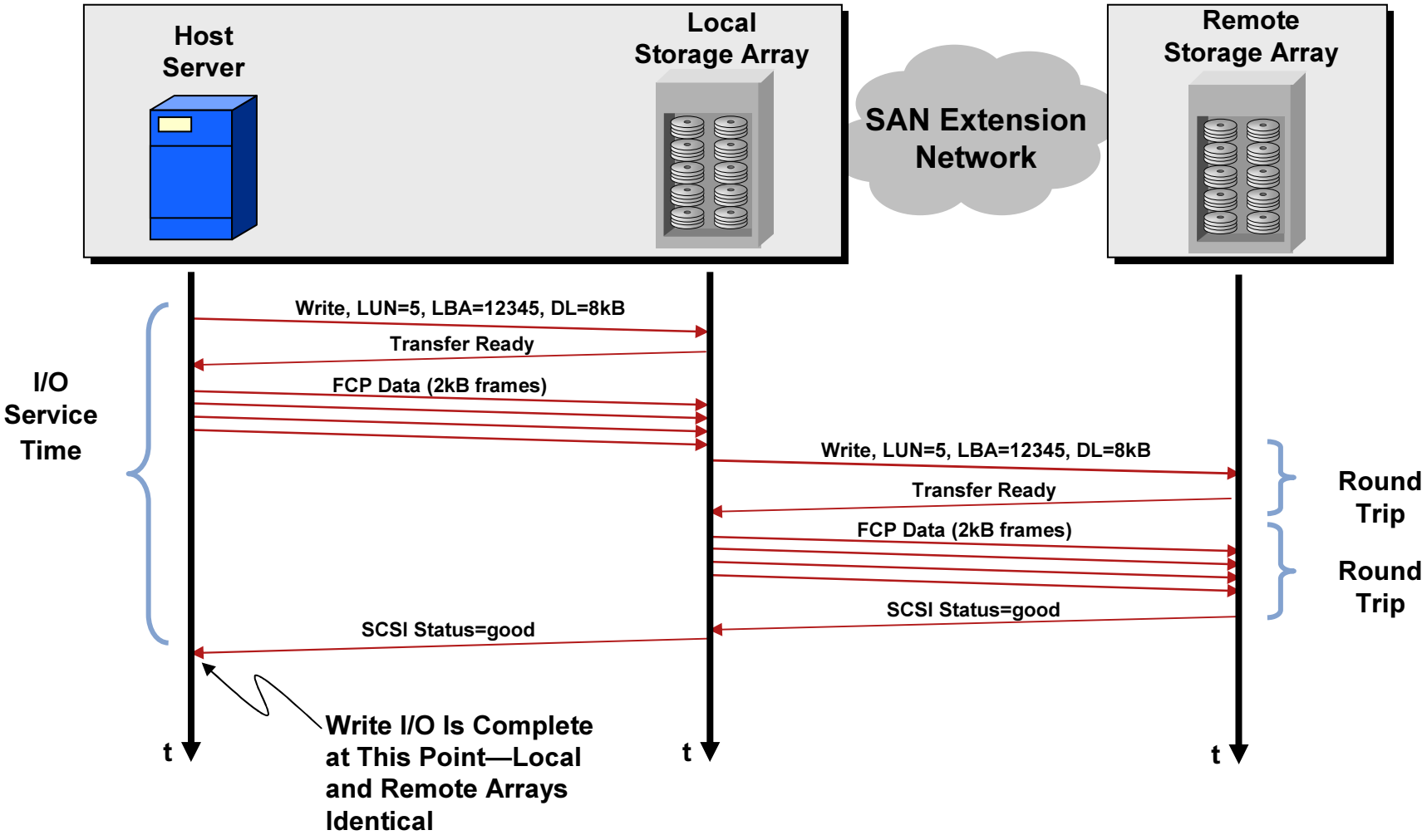
- What to do if I need more BB credits
  - *Buy switch with more credits (Cisco MDS 9500)*
  - *Use Distance extension (aka. BB credit spoofing)*

Distance extension solution

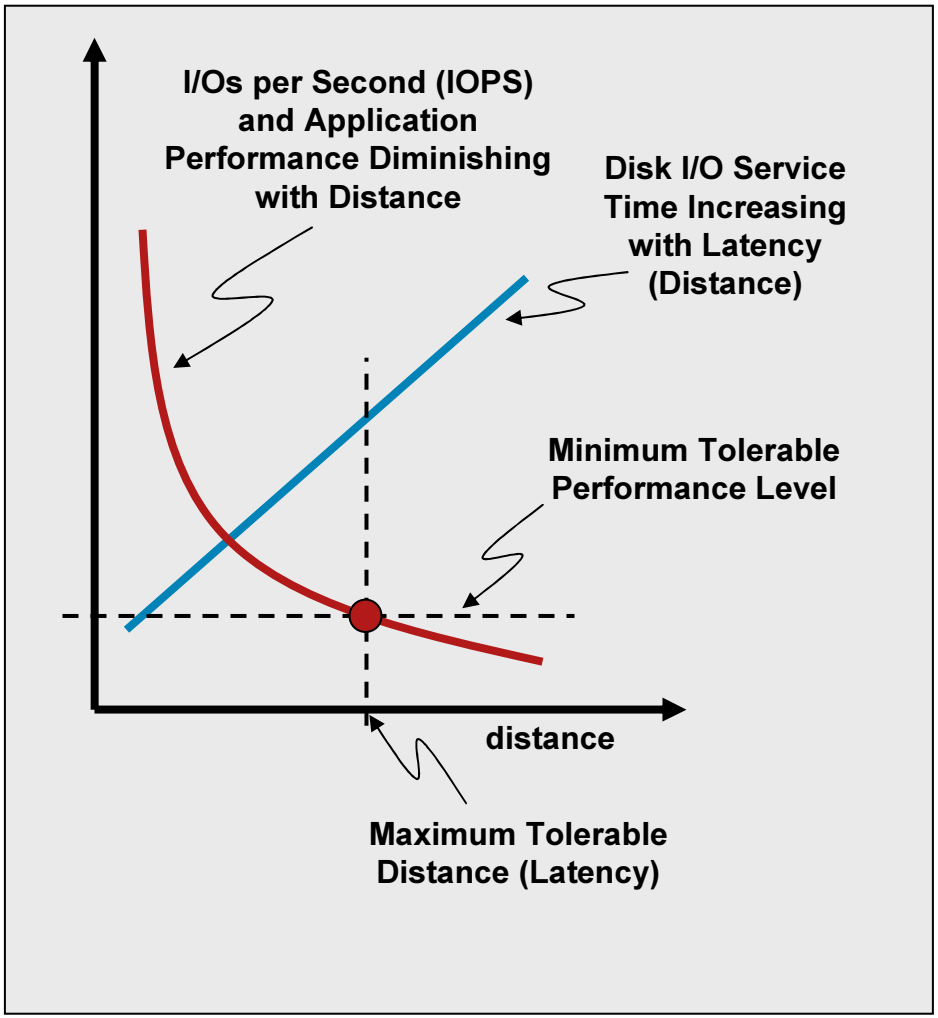


- Useful when switch does not have enough BB credits

Synchronous Replication
I/O Detail and latency consequences

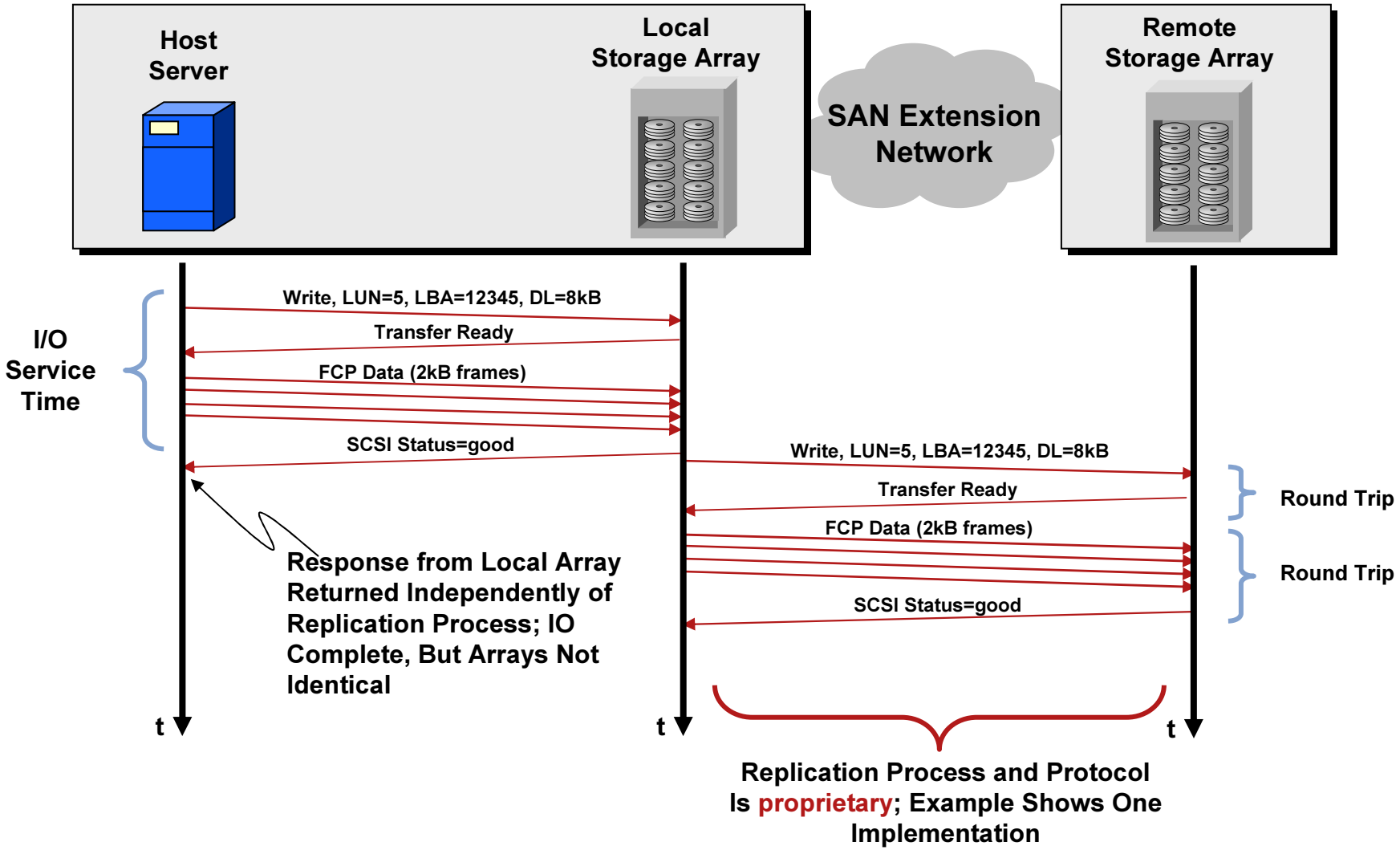


Synchronous Replication: How Far?



- Tolerable latency is up to the application (and Enterprise)
  - **Case-by-case basis**
  - **Databases are very sensitive to latency**
  - **Speed of light in fiber introduces delay of 5µs/km**
- Only Write I/Os are affected
  - **Increased "service (Response) time"**
- Maximum tolerable distance ascertained by assessing each application

Asynchronous Replication
I/O Detail and latency consequences



CWDM/DWDM designs for DC interconnect

Certifications

ONS 15454 SAN Qualification Summary June, 2007

Certification	15454-DM-L1-xx.x= 15454-DMP-L1-xx.x=	15454-MR-L1-xx.x= 15454-MRP-L1-xx.x=	15454-10DME-C= 15454-10DME-L=	15454-10E-L1-C= 15454-10E-L1-xx.x=
EMC SRDF / Mirrorview				
1G/2G/4G FC	✓ (1/2)	✓ (1/2)	✓	-
10G FC	-	-	-	✓
IBM GDPS/PPRC/XRC				
1G/2G/4G FC/FICON	✓ (1/2)	✓ (1/2)	✓	-
10G FC/FICON	-	-	-	✓
ESCON	✓	✓	-	-
ISC	-	✓	✓	-
STP (ISC)	-	✓ (MR only)	TBD	-

ONS 15454 SAN Qualification Summary June, 2007

Certification	15454-DM-L1-xx.x= 15454-DMP-L1-xx.x=	15454-MR-L1-xx.x= 15454-MRP-L1-xx.x=	15454-10DME-C= 15454-10DME-L=	15454-10E-L1-C= 15454-10E-L1-xx.x=
HP CA-EVA / DRM				
1G/2G/4G FC	✓ (1/2)	✓ (1/2)	✓	-
10G FC	-	-	-	✓
HDS TrueCopy				
1G/2G/4G FC	✓ (1/2)	✓ (1/2)	✓	-
10G FC	-	-	-	✓

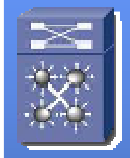
CWDM/DWDM designs for DC interconnect

Summary

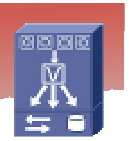
CWDM/DWDM - ideal solution for DC interconnect

- High bandwidth
- High scalability
- Traffic separation at Layer 1
- Optimise RPO
- Several options for different requirements/environments
- Cost savings by sharing one fiber pair (or even single fibre strand)

Cisco Product Portfolio for Data Center

Unified Fabric Networking	Ethernet Networking	Storage Networking	Optical Networking	Application Network Services	Infiniband Clustering	Data Center Security
						
						
Nexus 7000 Modular Switch Nexus 5000 Switch	Catalyst 6500 Series Catalyst 4900M Top-of-Rack Catalyst Blade Server Switches	MDS 9500 Storage Directors SSM MDS Fabric Switches Blade Switches	ONS 15454 MSTP ONS 15216 CWDM filters CWDM/DWDM pluggables	ACE Application Delivery – Module and Appliance Wide-Area Application Services ACE XML Gateway	SFS 7000 Infiniband Switch SFS 3000 Infiniband Gateway	Firewall Services Module

Data Center Provisioning — VFrame Server/Service Provisioning System



Data Center Management — Data Center Network Manager– Topology Visualization and Provisioning; ANM– Advanced L4-7 Services Module Management

Q and A