Session 17926

Hints, Tips, and Best Practices For Implementing FICON fabrics (As well as how to avoid common pitfalls).

Mike Blair **Cisco Systems** **mblair@cisco.com**

#SHAREorg

SHARE is an independent volunteer-run information technology association that provides education, professional networking and industry influence.

Session Abstract

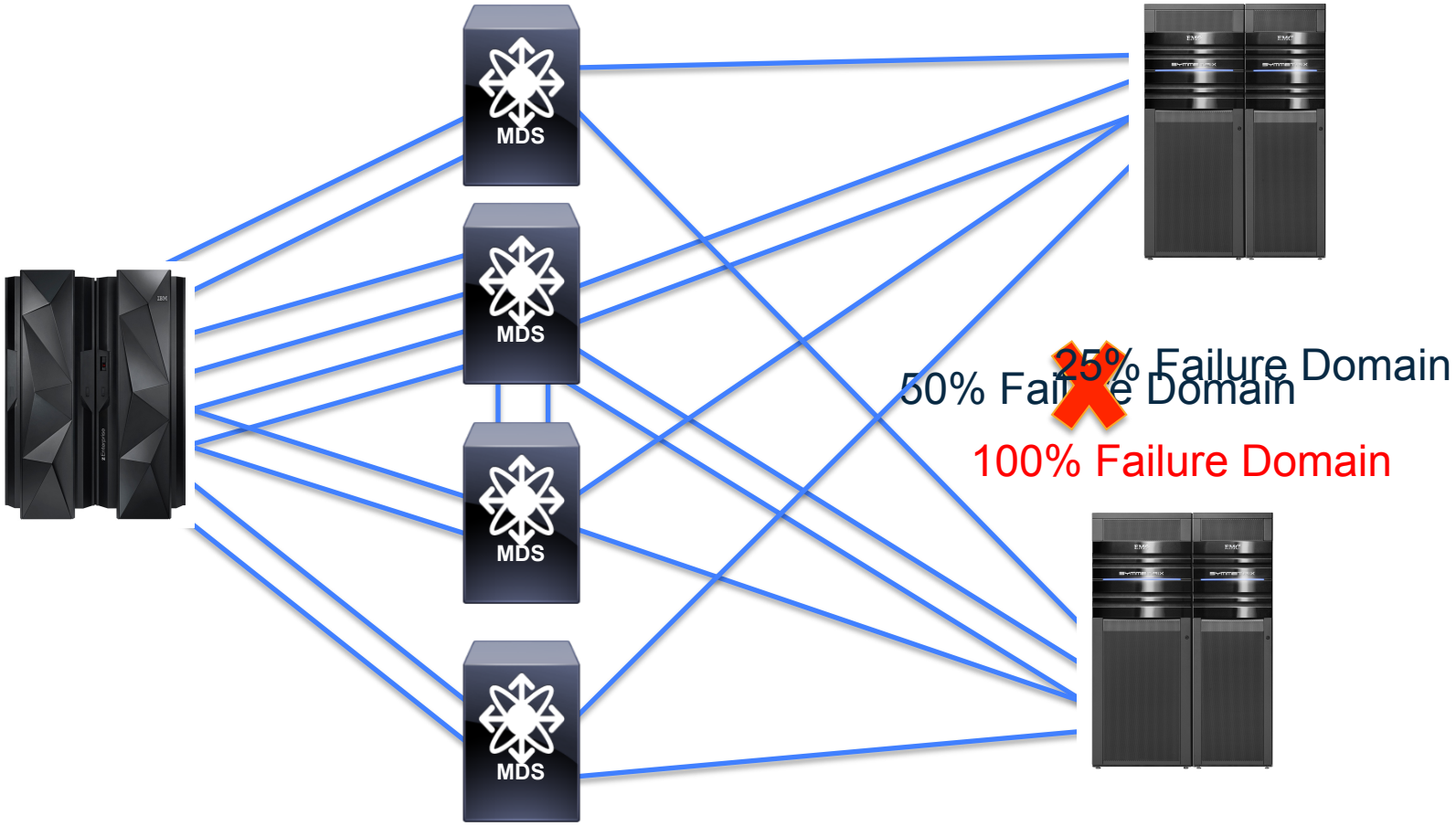
- Want to learn the best practices for FICON fabric design ? What things are significant and which are not ? How do/ should virtual fabrics play into today's designs ? How many buffer credits are enough and how can I monitor them to make sure I estimate correctly ? What technologies are available for link aggregation and when are these recommended ? What are the most common pitfalls seen during initial FCIP SAN extension deployments ?

- This session will cover real life examples of deployments and issues seen during customer deployments - and how to avoid these.

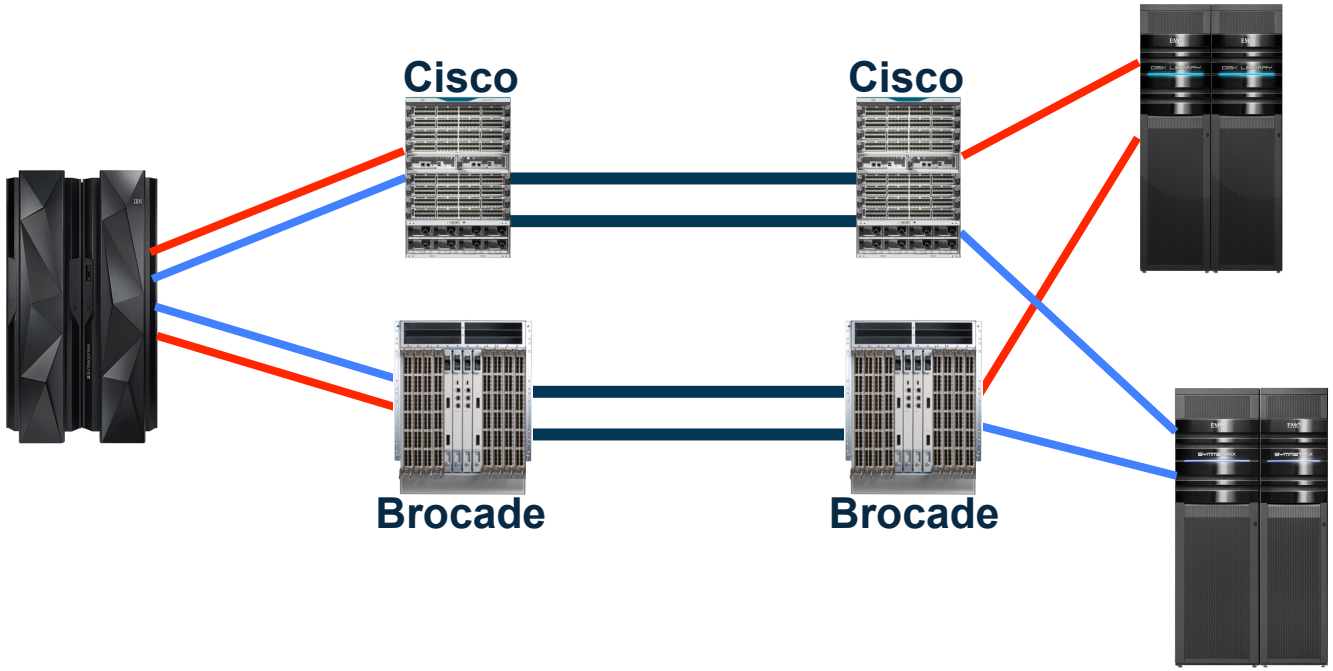
Agenda

- Fabric Design Discussion
- Virtual SANs / Virtual Fabrics
- Local Switching FICON
  - FICON Addressing
  - CUP Usage
- Cascaded FICON
  - Speeds, Optics, Fibers, and Infrastructure
  - ISL High Availability
  - ISLs Over Distance
  - Buffer Credits
- Long Distance FICON
  - FICON over IP
  - Channel Extension

Fabric Design – Redundancy / Failure Domain

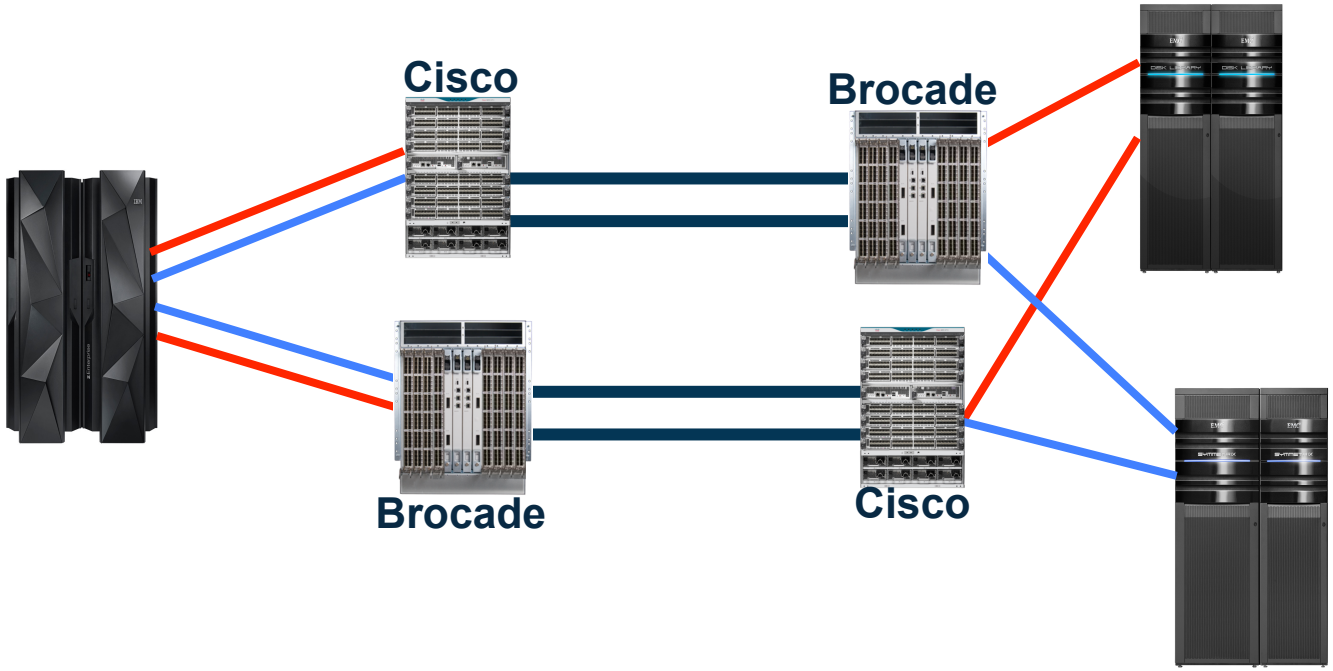


Fabric Design – Vendor Intermix



- Supported!
  - 2 Separate Fabrics with no Interconnection
  - Fully supported by IBM, EMC, and HDS
  - Common during Vendor Migration

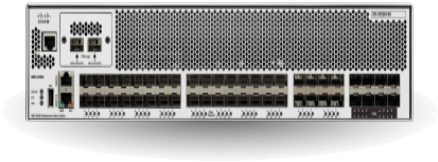
Fabric Design – Vendor Intermix



- Not Supported!
  - FICON VSANs from Different Vendors will not talk
    - Unlike Open Systems which can do this

FICON Directors Versus Switches

VS



- FICON Director
  - High Port-Density Modular Platform
  - Built for Investment Protection / Growth of Features, Functions, and Speed
  - Full Line-Rate Ports
  - N:N Power Grid Redundancy
  - N+1 Power Redundancy within Grid
  - N+1 Fabric Redundancy
  - In-Service Component Repair
  - In-Service Software Upgrade
  - In-Service Upgrade to Higher Speeds

FICON Switch

- Low Port-Density Fixed Configuration
- Often with a Pay-As-You-Grow Model
- Full Rate Ports but Less Memory
- Redundant Power (but not Grid)
- Disruptive / Replacement Repairs
- Disruptive Upgrade to Higher Speeds
- In-Service Software Upgrades
- Often Purpose Built

FICON Directors Versus Switches

- Directors Most Suited
  - Production Mainframe Environment
  - Disk Replication Environment
  - Tape Replication Environment
  - DR Facility (It could become the Production Environment)
- Switches May be Sufficient
  - Test / Staging / Burn-in Environments
  - Development Environments
  - Remote Archival of Tape
  - Tape Replication Environment

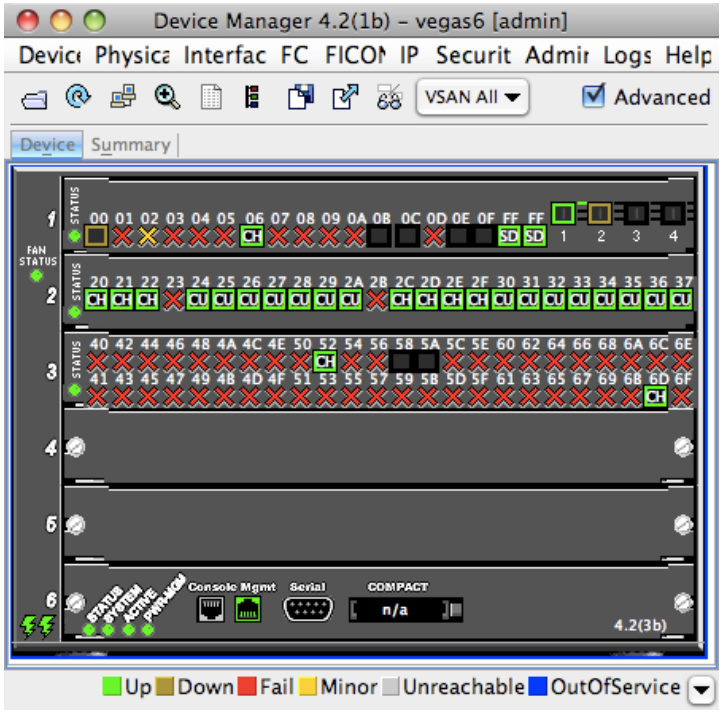
Fabric Design – Tips

- Decide on your Redundancy Based on Failure Domain
- Maintain Fabric Isolation
- No Vendor Mixing for FICON Directors
- For Most FICON Applications, Director Class is Recommended

Virtual SANs / Virtual Fabrics

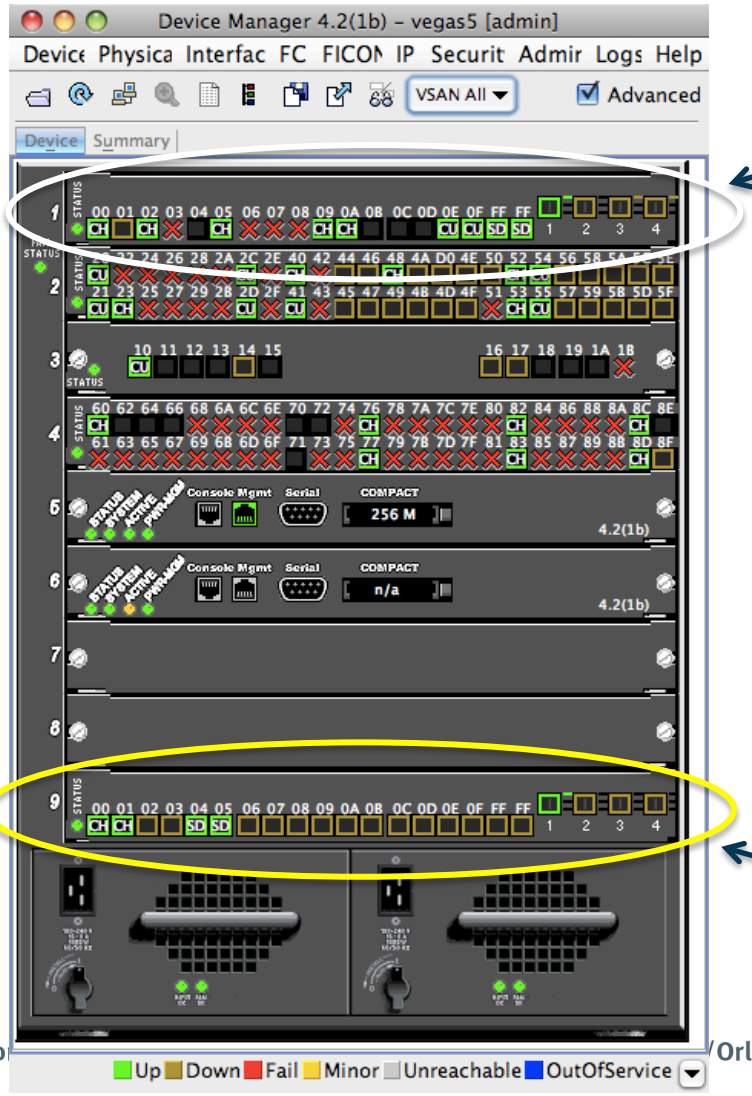
- Isolation / Partitioning Mechanism for SAN Workloads
- 1 Physical Director/Switch becomes multiple “boxes”
- Each VSAN Has
  - Its own Fabric Services (LOGIN, Name, Security)
  - Its own Serial Number
  - Its own Domain Number
  - FICON Port Addressing
  - FICON Management CUP Port
  - Traffic Routing Space (Ports cannot talk between VSANs)
- VSANs can be divided up on a port basis

FICON Addressing is Per VSAN



- FICON Port Addresses
  - Every port has a 1-byte FICON port address
  - Valid values are 0x00 – 0xFD
  - 0xFE reserved for CUP
  - Ports are not useable without a port address
  - Defaults are relatively worthless – throw away
  - Port addresses must be assigned left-> right
- Special FICON Port addresses
  - Every physical port has a port address
  - FCIP links and Port Channels use logical
  - Logical port addresses come from logical pool

FICON VSANs – Port Addressing Duplicates



VSAN 2

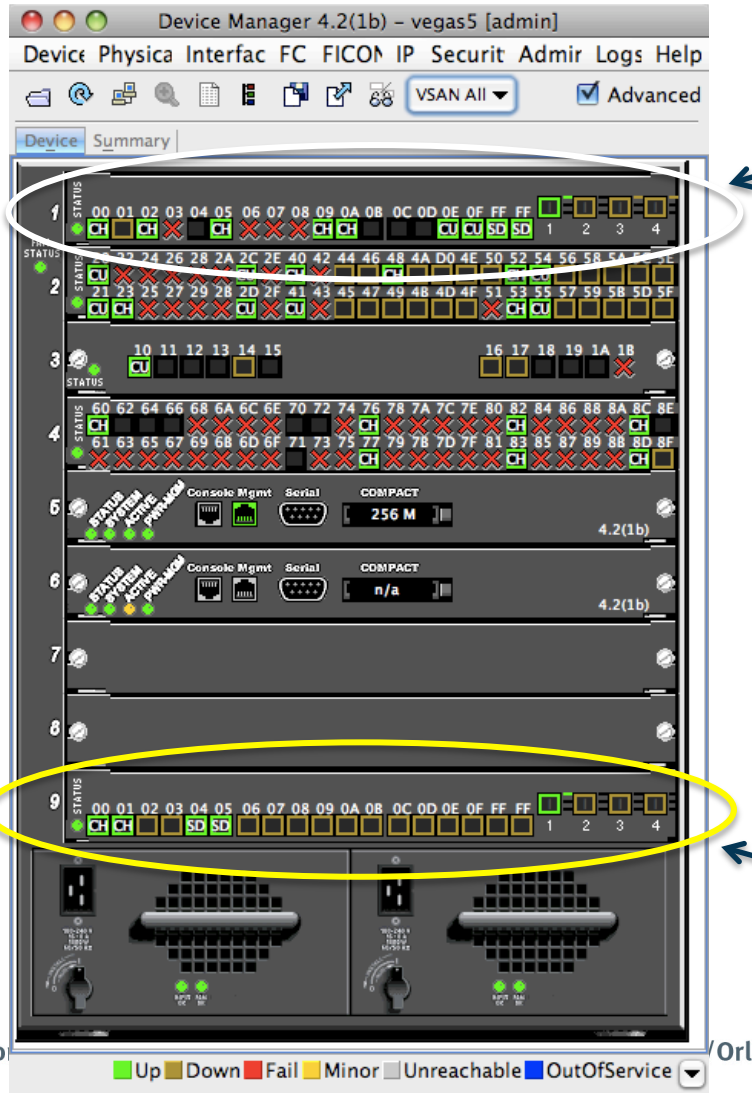
Port Addresses 0x00 – 0x0F

- Duplicate Port Numbers
  - Allowed when in different VSANs

VSAN 10

Port Addresses 0x00 – 0x0F

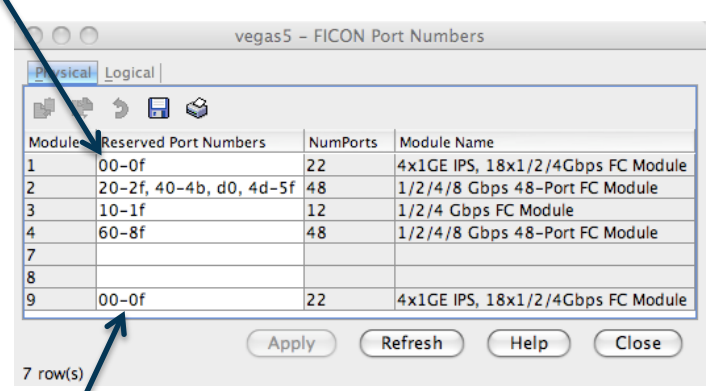
FICON VSANs – Port Addressing Duplicates



VSAN 2

Port Addresses 0x00 – 0x0F

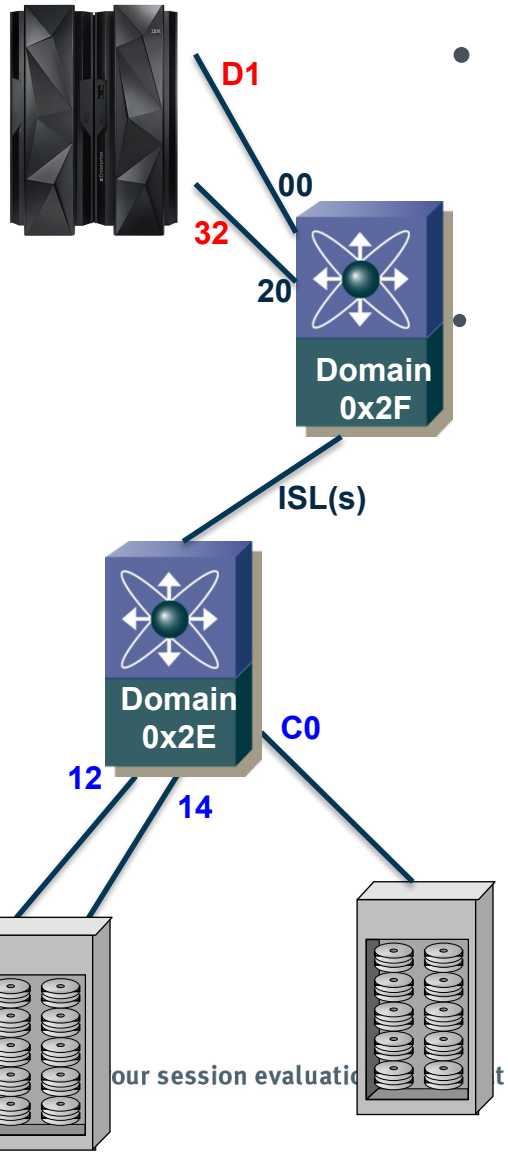
- Duplicate Port Numbers
  - Allowed when in different VSANs



VSAN 10

Port Addresses 0x00 – 0x0F

VSAN Numbers Versus Domain Numbers

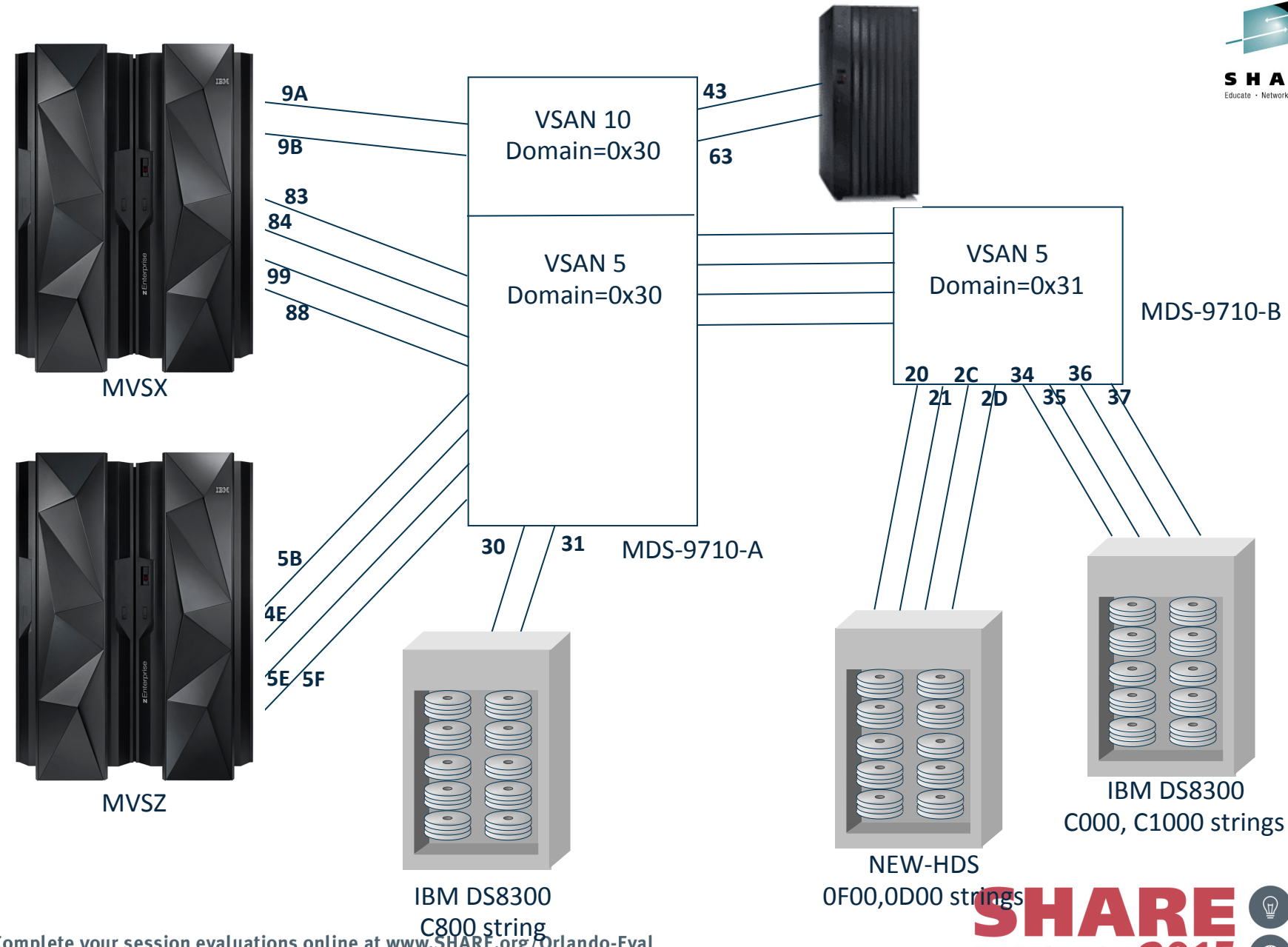


- VSAN Number
  - Arbitrary Number that defines the “Fabric”
  - Must be present in all Directors for ports to Talk
  - Mainframe has no visibility to the VSAN
- Domain Number
  - Must be different within each Director for the same VSAN
  - All Domains must be Unique per Mainframe

```
CHPID PATH=(CSS(0),D1),SHARED,                                  *
      PARTITION=((LPARMVSY),(LPARMVSX,LPARMVSA,LPARMVSB,VMLPAR*
      02)),SWITCH=2F,PCHID=1B1,TYPE=FC
CHPID PATH=(CSS(0),32),SHARED,                                  *
      PARTITION=((LPARMVSY),(LPARMVSX,LPARMVSA,LPARMVSB,VMLPAR*
      02)),SWITCH=2F,PCHID=1B3,TYPE=FC

CNTLUNIT CUNUMBR=0CC0,PATH=((CSS(0),D1,32)),                    *
      LINK=((CSS(0),2E12,2E14)),                                *
      UNIT=2105,CUADD=B,UNITADD=((00,032))

CNTLUNIT CUNUMBR=0C00,PATH=((CSS(0),D1)),                       *
      LINK=((CSS(0),2EC0)),                                     *
      UNIT=2105,CUADD=C,UNITADD=((00,032))
```


9A
9B
83
84
99
88
VSAN 10
Domain=0x30
VSAN 5
Domain=0x30
43
63
VSAN 5
Domain=0x31
MDS-9710-B
MVSX
MVSZ
5B
4E
5E
5F
30
31
MDS-9710-A
20
21
2C
2D
34
35
36
37
IBM DS8300
C800 string
NEW-HDS
0F00,0D00 strings
IBM DS8300
C000, C1000 strings

CONFUSED!

MVSX

MVSZ

9A
9B
83
84
99
88

VSAN 10
Domain=0x30
Serial Number=AA

VSAN 5
Domain=0x30
Serial Number=BB

MDS-9710-A

43
63

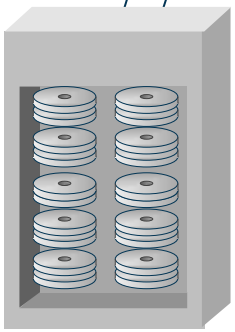
VSAN 5
Domain=0x31

MDS-9710-B

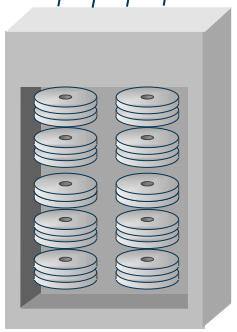
20 21 2C 2D 34 35 36 37

5B
4E
5E 5F

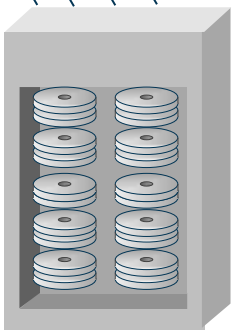
30 31



IBM DS8300
C800 string



NEW-HDS
0F00,0D00 strings



IBM DS8300
C000, C1000 strings

All Good

MVSX

MVSZ

9A
9B
83
84
99
88

VSAN 10
Domain=0xD1
Serial Number=AA

VSAN 5
Domain=0x30
Serial Number=BB

MDS-9710-A

43
63

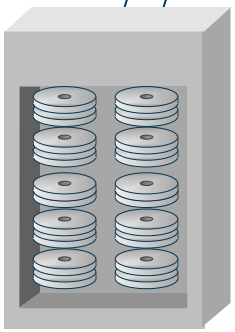
VSAN 5
Domain=0x31

MDS-9710-B

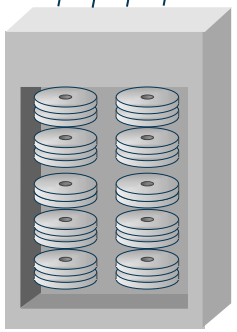
20 2C 34 36
21 2D 35 37

5B
4E
5E 5F

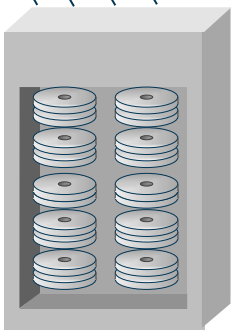
30 31



IBM DS8300
C800 string



NEW-HDS
0F00,0D00 strings



IBM DS8300
C000, C1000 strings

Virtual SANs – When To Use

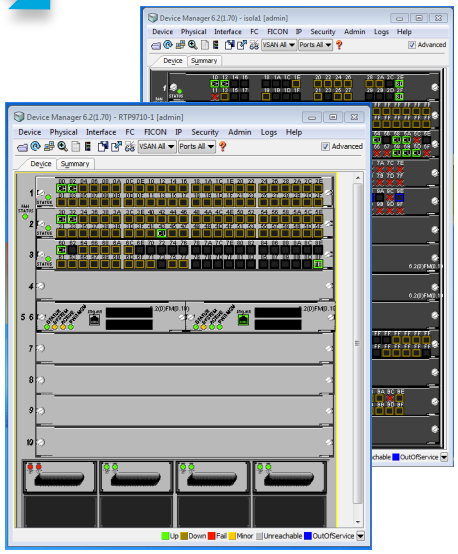
- Rule #1 – Keep it Simple
  - Just Because You can Complicate it – Resist!
- FICON Versus Open Systems
  - Separate Replication VSAN but share FICON Director
  - Separate Linux LPARs from FICON VSAN
- Divide FICON VSANs by Customer / Group
- Divide FICON VSANs by Local / Remote (Cascade)
- Divide to Isolate traffic

Virtual SANs / Virtual Fabrics – Tips

- Keep VSAN Implementation as Simple as Possible
- There is no non-VSAN mode for Cisco FICON
  - Very like IBM System Z – No Single Image Mode Anymore
- Don't use VSAN number 1 for FICON
- Don't use the same Domain number on 2 different VSANs
- Don't make the VSAN Number the Domain Number
- Keep Domains Unique Regardless of VSANs
  - Don't Allow Mainframe to see Same Domain in 2 different places
- Keep FICON Port Address Contiguous if possible

FICON CUP – What is it / When to use it ?

FICON In-Band Communications (CUP)



- Control Unit Port
  - IBM Specified in-Band Mgmt interface for Directors
- Must be defined to the IOCDS
- One CUP per VSAN (automatically)
- Allows Control / Time Sync
- Allows Gathering of Statistics
  - Stats are gathered by IO Subsystem
  - Gathered based on RMF Interval
  - SA/IO-Ops not required
  - FICON Director Records are logged as an SMF Record
  - Everyone Should Run This!
- No Extra License Fee

RMF Report Based on CUP Data

```
SWITCH DEVICE: EF20   SWITCH ID: 20     TYPE: 0MDS9K   MODEL: 513   MAN: CSC  PLANT: 00   SERIAL: 000DEC3B7DC2
PORT     -CONNECTION-   AVG FRAME     AVG FRAME SIZE     PORT BANDWIDTH (MB/SEC)     ERROR
ADDR     UNIT      ID      PACING       READ   WRITE     -- READ --  -- WRITE --     COUNT
 0E      CHP-H     51           0         93     887          0.00         0.0         0
 0F      CHP-H     9C           0       1190    1154         28.86        24.1         0
 7A      ------  ----       P O R T   O F F L I N E
 E1      SWITCH  ----           0       1545    1432         24.73        29.5         0
```

- Director Report Interval same as other RMF Reports
  - Allows for Correlation of events
- Visibility to Average Port Bandwidth (Write and Read)
- Visibility to Average Frame Size (Important Later)
- Provides Hints of Port Congestion (Frame Pacing)

CUP Data Collection Tips

- Limit FICON Stats Gathering to one LPAR per VSAN
  - FICON Stats Gathering is Enabled on a Per LPAR Basis
  - CUP Processing is CPU Intensive on the Director
- Use CUP to Gather Stats on Replication / Open VSANs
  - Costs one CHPID to be placed into the VSAN
  - VSAN must be defined as a FICON VSAN
- FICON Stats are gathered for Logical Links
  - Port Channels are monitored
  - FCIP Tunnels are also monitored
- Vary CUP Devices Offline during Code Upgrade
  - Avoids one IFCC per VSAN

Speeds, Optics, Fibers, and Infrastructure

Speed Name	Throughput (MBps)	Line Rate (GBaud)	Encoding	Retimers in the module	Transmitter Training
1GFC	100	1.0625	8b/10b	No	No
2GFC	200	2.125	8b/10b	No	No
4GFC	400	4.25	8b/10b	No	No
8GFC	800	8.5	8b/10b	No	No
10GFC	1200	10.53	64b/66b	Yes	No
16GFC	1600	14.025	64b/66b	Yes	Yes

- Open Systems moving to 16G FC - mostly 8G now
- 16G Switches shipping for multiple years – large 8G base
- FICON is mostly 4G/8G – 16G is finally here!!
- Long Distance is 4G / 8G – 16G on horizon

Speeds, Optics, Fibers, and Infrastructure

Speed Name	Multimode OM1 Link Distance 62.5 um core and 200 MHZ*km	Multimode OM2 Link Distance 50 um core and 500 MHz*km	Multimode OM3 Link Distance 50um core and 2000 MHz*km	Multimode OM4 Link Distance 50um core and 4700 MHz*km	Single-mode OS1 Link Distance 9um core and ~infinite MHz*km
1GFC	300	500	860	*	10,000
2GFC	150	300	500	*	10,000
4GFC	50	150	380	400	10,000
8GFC	21	50	150	190	10,000
10GFC	33	82	300	*	10,000
16GFC	15	35	100	125	10,000

* The link distance on OM4 fiber has not been defined for these speeds.

Small Form-Factor Pluggable (SFP)

1G, 2G, 4G – Round Handle
Black – MM (all speeds), Dark Blue SM (2G)
Light Blue SM (4G), Grey SM (4G)

8G, 10G, 16G – Metal Handle
Blue – SM
Black – MM

Spectrum Colors
DWDM, CWDM Optics
Various Speeds

Small Form-Factor Pluggable (SFP)

- 4G SFPs
  - 1G / 2G / 4G Auto
- 8G SFPs
  - 2G / 4G / 8G Auto
- 16G SFPs
  - 4G / 8G / 16G Auto
- 10G SFPs
  - 10G Only (ISLs Only)

- 8G Cards Generally Support 4G and 8G SFPs (Some Add 10G)
- 16G Cards Support and 8G, 16G, and 10G SFPs (NOT 4G)
- SFPs can be different per port

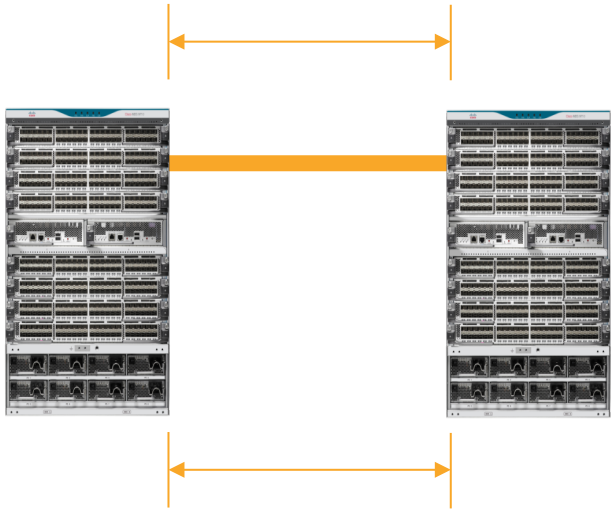
Optical Domain + 16G FICON Tips

- Use Optical Scopes to Verify Fiber/SFP Condition
  - Cleanliness is More Important as Speed Increases (16G)
  - “Blind” Cleaning is not enough
- For any Greenfield Installations consider SM Fiber
- Benchmark Optical power for Ports
  - Verify no/low Power Variance After Changes
- With Director Purchase insure SFP Speed support
- Optics are Vendor Specific – Unsupported even if Working
- For new FICON Director Purchases – 16G Optics
  - 16G FICON is (finally) here
  - Price Delta for 16G Optics is reducing

16G Forward Error Correction

Additional protection against application performance impacts

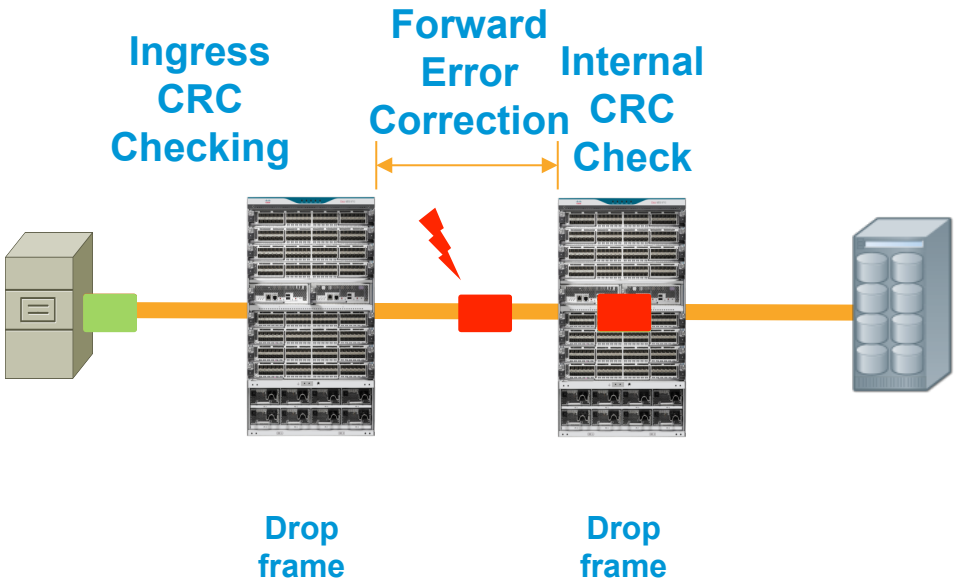
- Lossy Media such as loose SFPs, dirty cable can result in packets getting corrupted on ISLs
- FEC improves reliability of links
  - Corrects errors to some degree on the media between switches to reduce the number of corrupted frames that have to be dropped
- Even if bad packets do get through, CRC validation and packet drops will ensure that corrupted frames do not enter the network



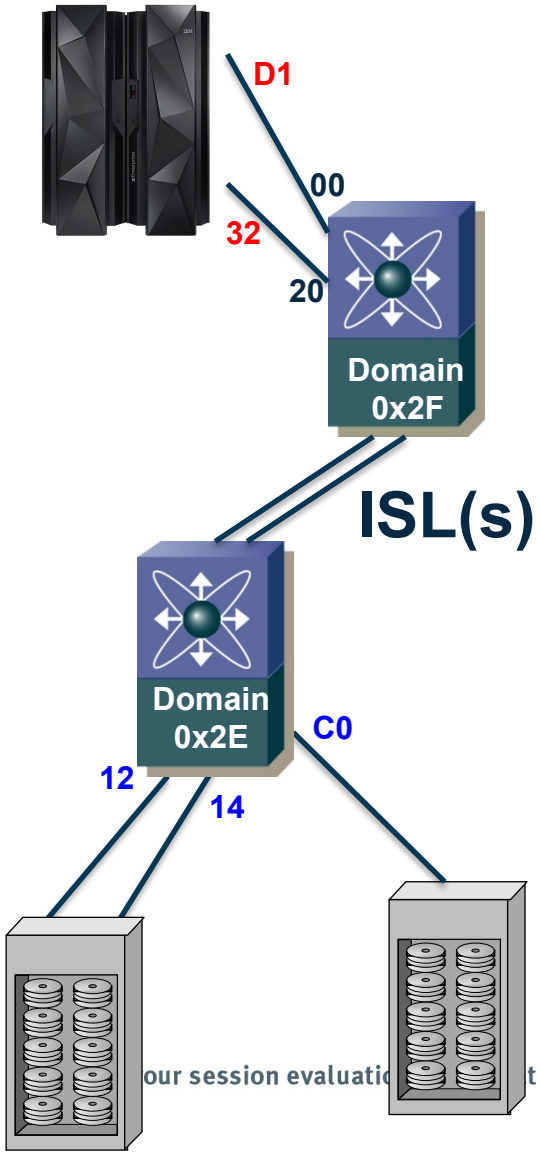
Available today for ISLs

Fabric Resiliency
CRC Detection and 16 Gbps FC Forward Error Correction (FEC)

- Faulty equipment, loose SFPs, dirty or damaged cables, can result in packets getting corrupted
- CRC Checking drops corrupted frames from end devices or internally corrupted frames
- FEC corrects frames corrupted in-flight to preserve frames
- FEC is hop-by-hop
  - Today it is between switches only
  - Tomorrow could be CH or CU links

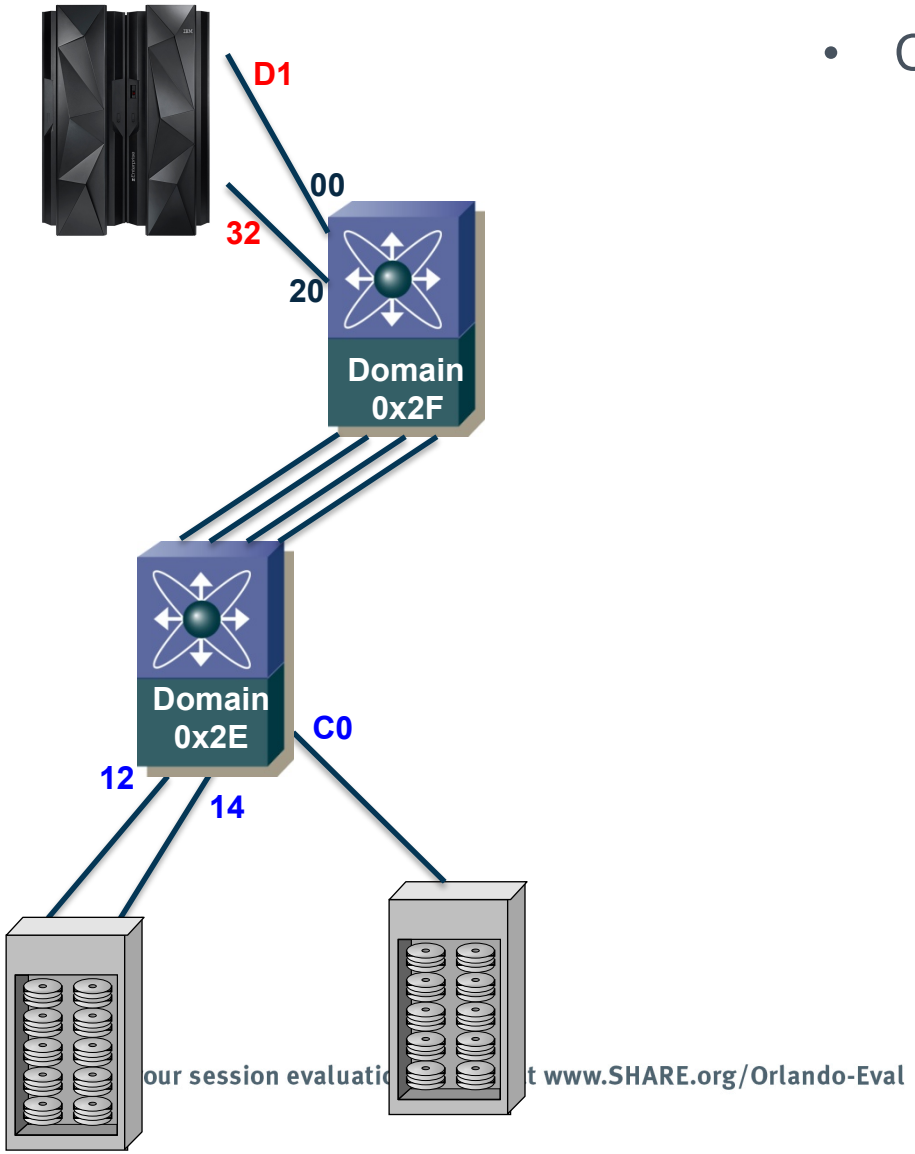


InterSwitch Links (ISLs)



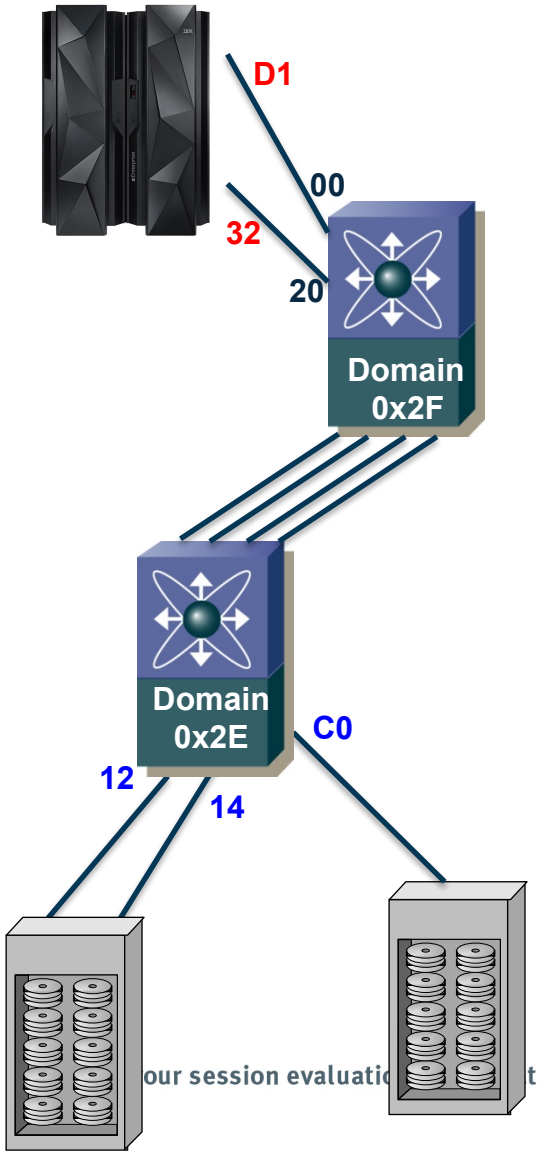
- InterSwitch Links
  - Also Called Expansion(E) Ports
- Connects 2 switches to create a fabric
- ISLs can carry one or more VSANs
  - E port carries a single VSAN
  - TE (Trunking E) port carries multiple VSANs
- Can be Fibre Channel or FCIP

InterSwitch Links (ISLs) - Routing



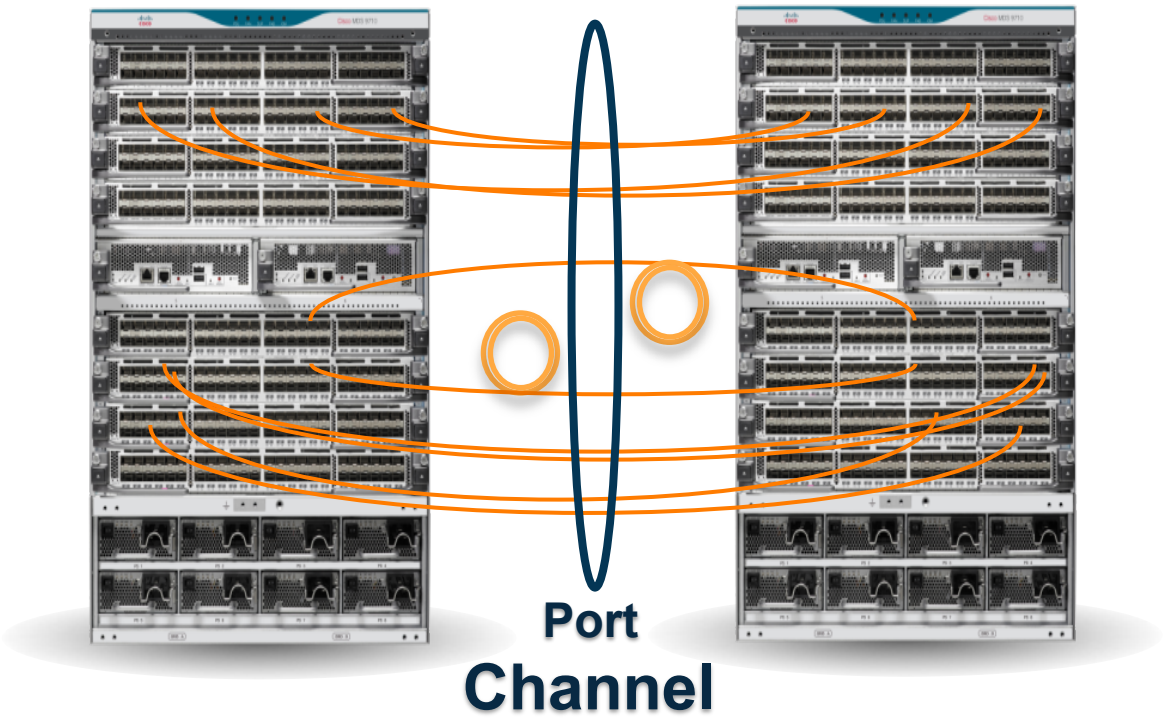
- Current FICON ISL Load-Balancing
  - For In-Order Delivery uses Source port and Destination Port Addresses to make ISL Selection
  - Link Selection holds until a Topology Change
  - This method was Selected for FICON because Exchange-based ISL Balancing Surfaced Issues in Testing
  - Because the Selection is Static, Mapping of a Path between Source and Destination is Possible
  - Can Exhibit Less than Optimal Balancing if not enough CH – CU Pairs compared to number of ISLs

InterSwitch Links (ISLs) - Routing



- New Exchange-Based Routing
  - Work in Progress to Enable the System z to Correctly Interoperate with Exchange-Based Load Balancing
  - On Cisco FICON Directors this is called SRC/DST/OXID Balancing
  - Load Balancing is Configured on a per VSAN Basis
  - Previous Code Levels Restricted FICON VSANs from SRC/DST/ OXID Balancing
  - With EBR, Different Flows between Source and Destination Ports will use all available ISLs
  - Load Balancing of ISLs is very uniform with EBR

Building High Availability for InterSwitch Links



>1 PORT GROUPS

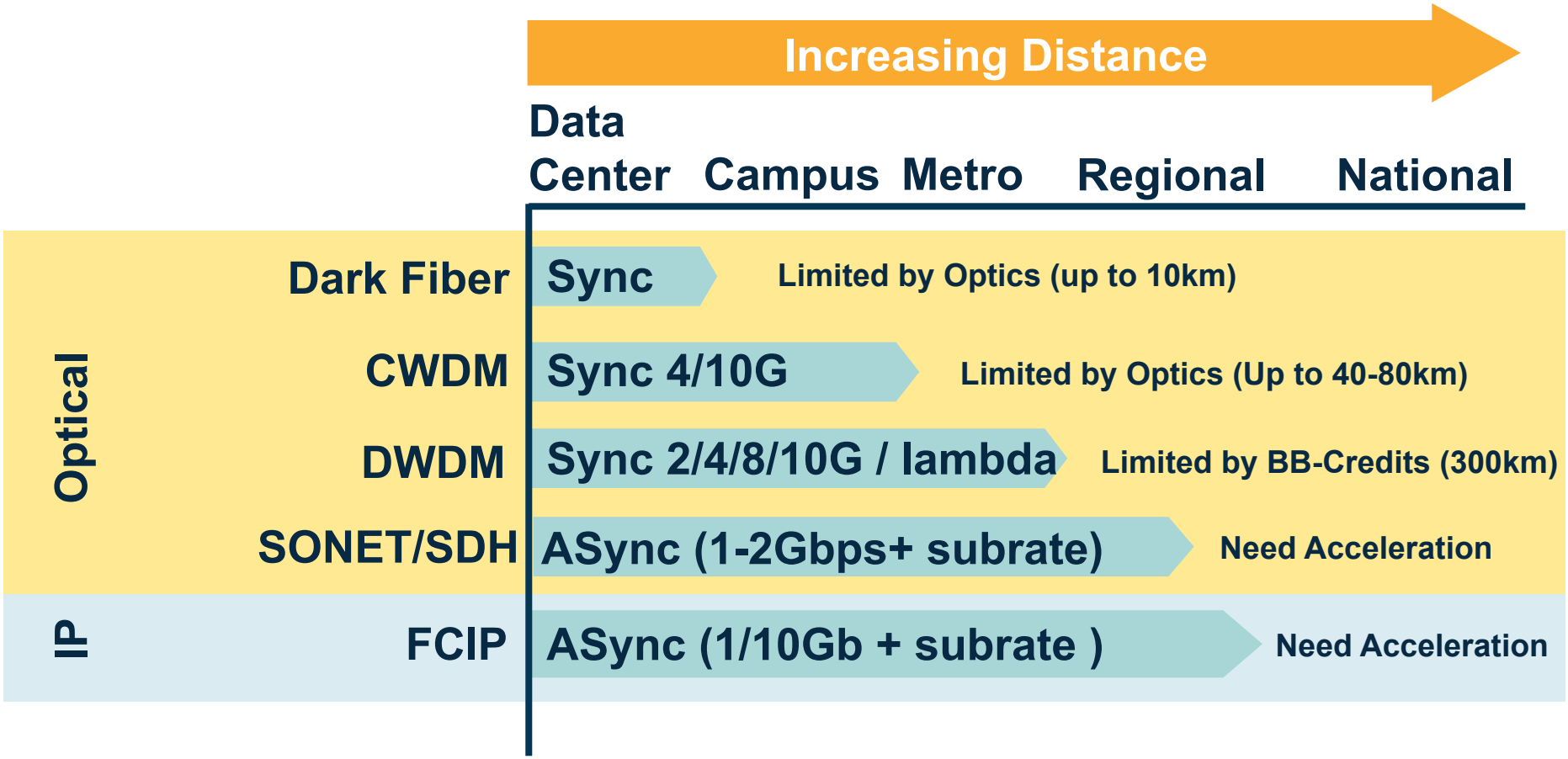
>1 ASICS SPANNED

MULTIPLE LINE CARDS

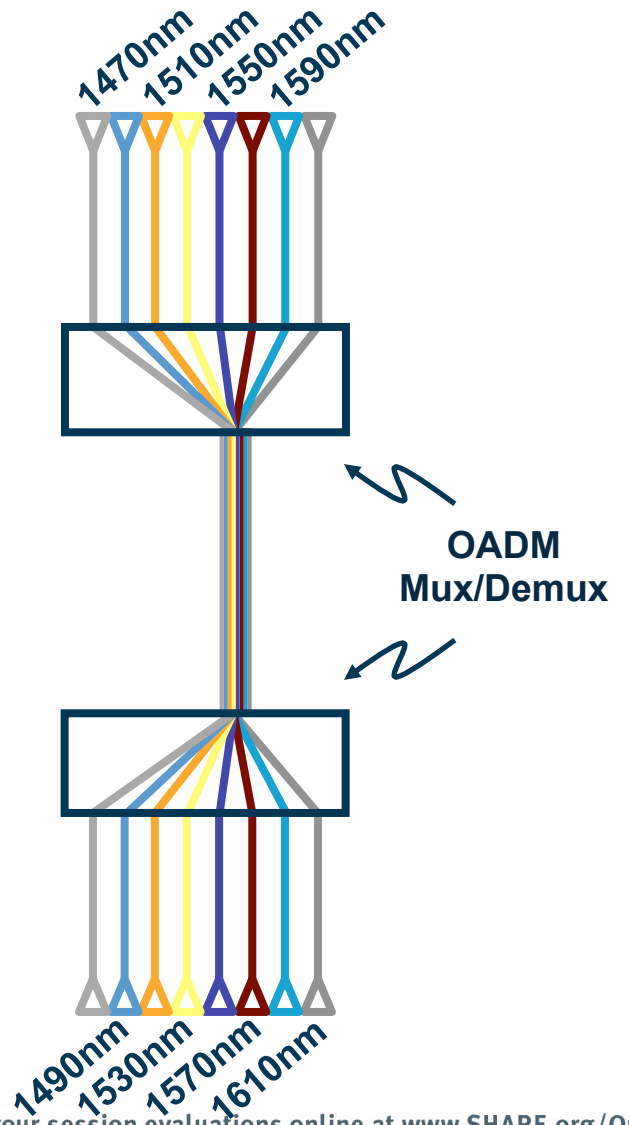
DISPARATE ROUTES

- Up to 16 Links of the Same Speed (Unequal Lengths – Not a Problem)
- Non-disruptive Activation / Deactivation of member links
- Maintains In-Order Delivery (Required for FICON)

Long Distance ISLs – Distance vs Technology



CWDM (Course Wavelength Division Multiplexing)



- 8 Channel WDM at 20nm Spacing
- “Colored” Optics in FICON Director
- Optical Multiplexing accomplished with a CWDM OADM (Optical Add/Drop Multiplexer)
  - Passive (unpowered) Device
- Power Budget (distance) depends on Speed
  - 2G optics support approx 100km
  - 4G optics support approx 40km
  - 10G optics support approx 50km
- Not Able to be Amplified
- Very Cost effective for some distance/scale

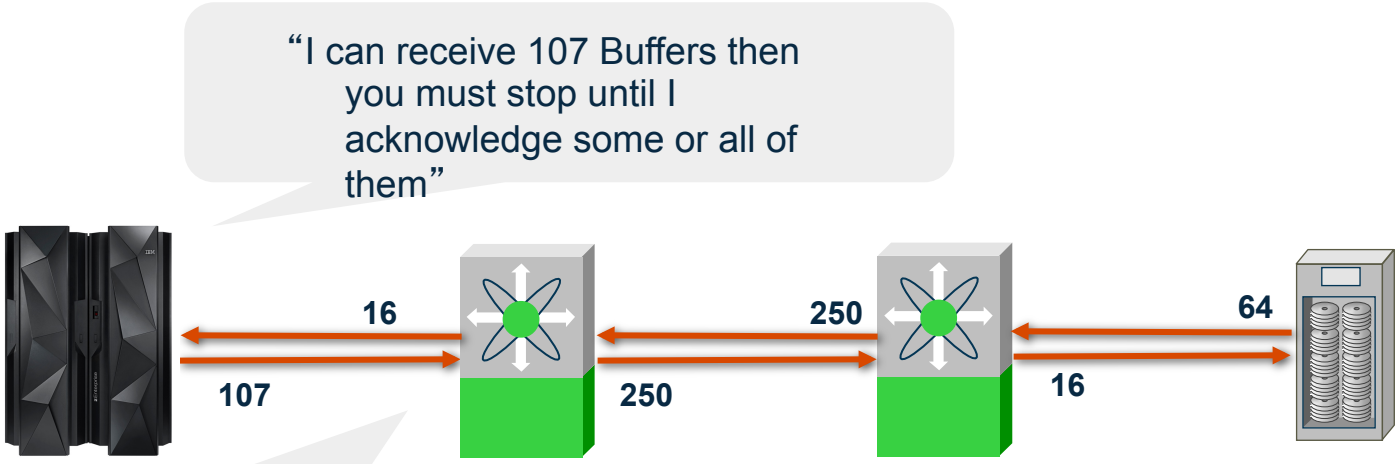
DWDM High Level Overview

- Dense Wavelength Division Multiplexing
  - 80+ Wavelength in Narrow Industry Defined Range
- Optical Transponders / Muxponders accept “Grey” optics
  - Regular SM or MM FC optics plug into optical cards
  - Optical Cards often bundle up number of lower speeds
  - Signal is usually Regenerated / Retimed
- DWDM Muxes also accept Foreign Wavelengths
  - DWDM Colored Optics directly in FICON Director
- Various Optical Amplifiers used to Increase Distance
- Support for 1/2/4/8/10/16G FC today
- Various Optical Domain Protections Available

FC Buffer Credit / Flow Control Concept

- Flow Control Mechanism used for FC / FICON Netwoks
  - Based on the Idea of no Packet Loss / Drops
- Each Port has a Configurable (sometimes) number of Buffer Credits
  - May be based on Port Type or Speed
- At Fabric Login, Number of Buffer Credits are Exchanged Between Peers
  - This is not a Negotiation – Each Side sets their Number of Credits
- Hosts can never have more than BC frames outstanding to peer
  - When a Host/CU Transmits a Frame, it Decrements the Number of Available Buffer Credits
  - Once the Number of Buffer Credits hits 0, Not frames can be sent
- Frames are Acknowledged at the Receiver with a R_RDY
- Each Buffer Credit is 2K

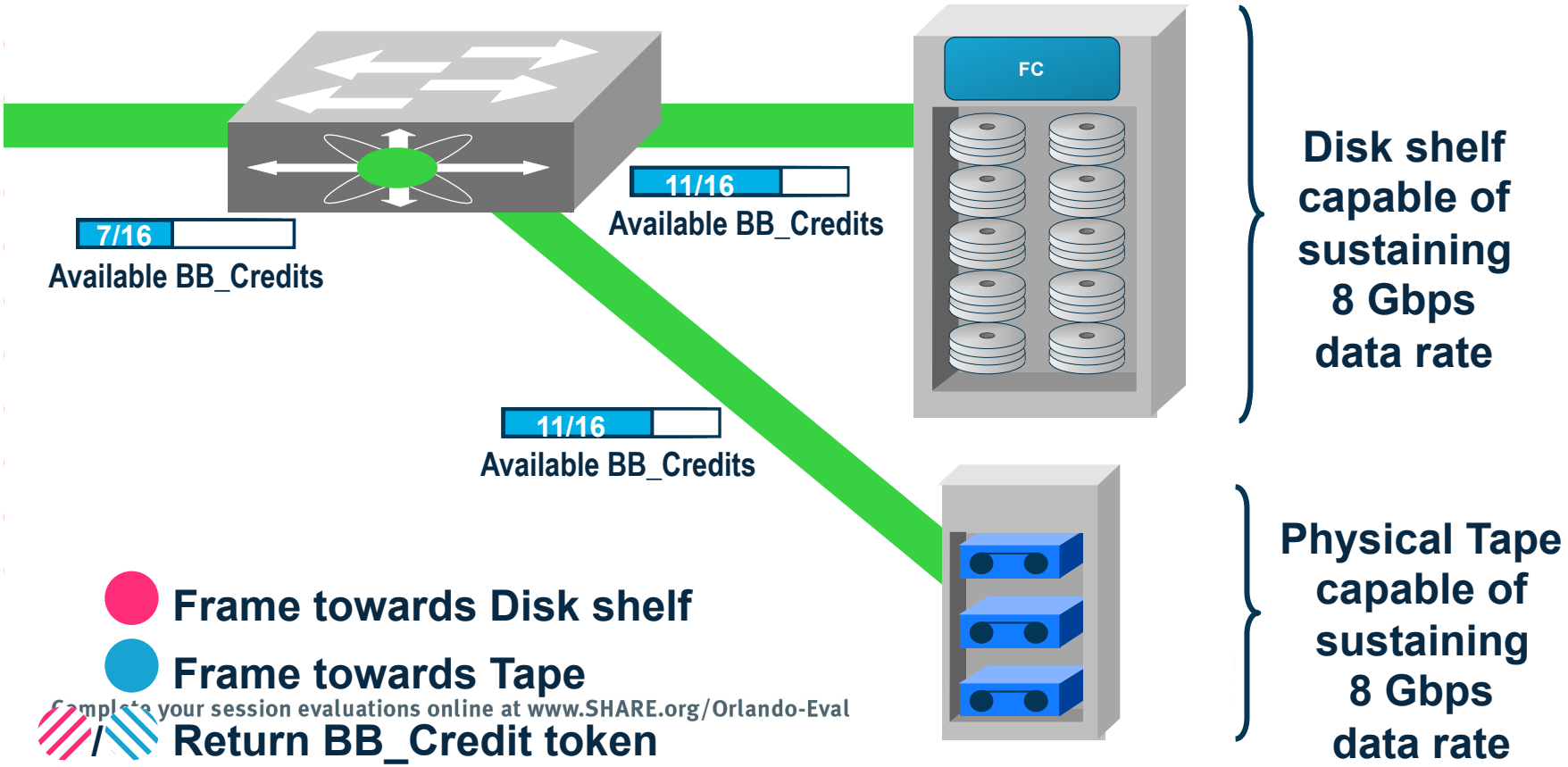
FC Buffer to Buffer Credits / FC Flow Control



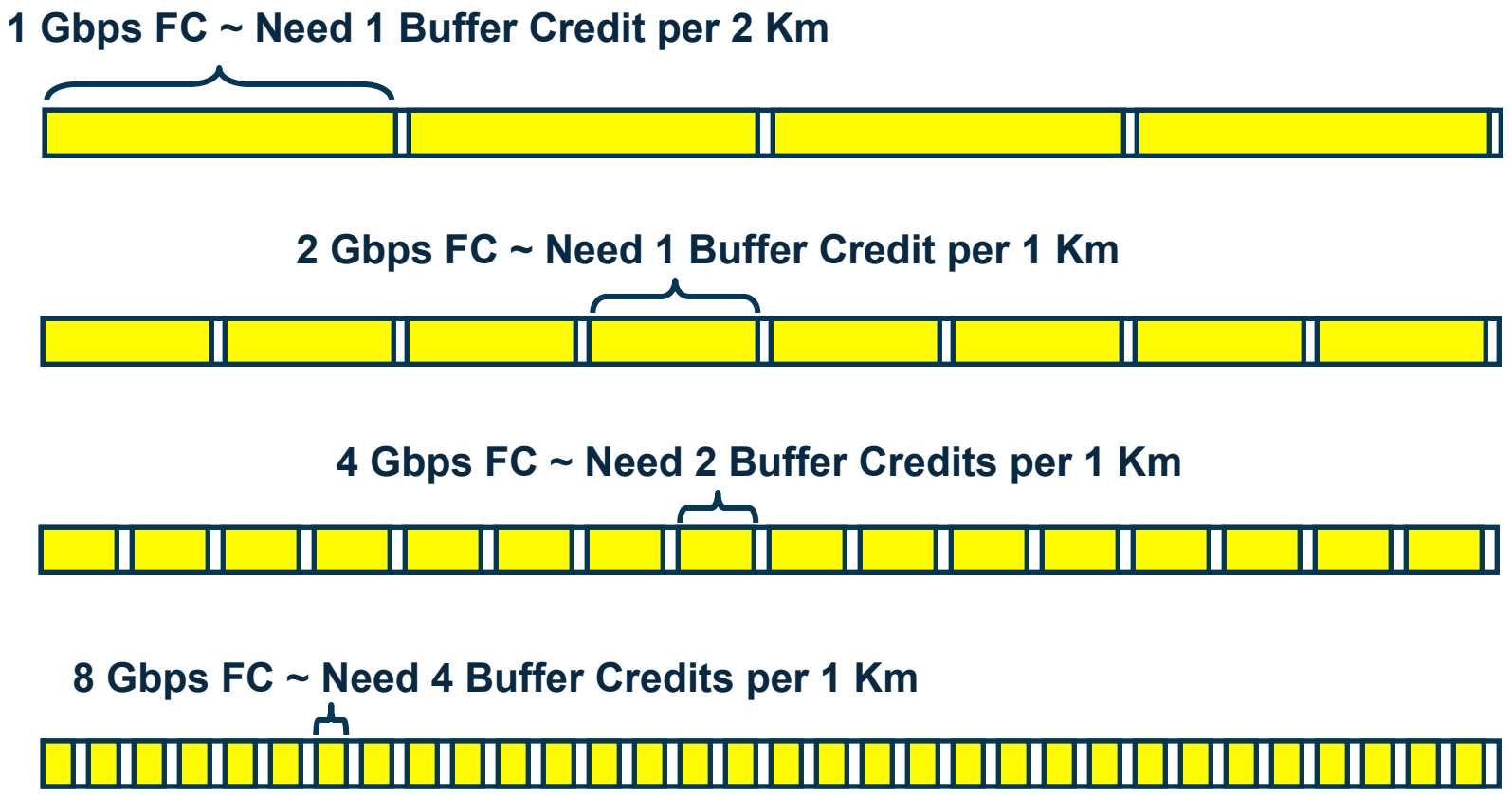
Buffer Credits (BB_Credits): Working Clean

Buffer Credits are a “Flow Control” mechanism to assure that frames are sent correctly

In an ideal FC network, there is no blocking in any device connected to the fabric. (All devices can process frames at the same rate and negotiate equal levels of BB_Credits)



Relationship of Speed to Buffer Credits



Based on Full 2K Frames

FC Buffer Credit / Flow Control - Further

- Average Frame Size for most Networks is not 2K
  - More Realistic Number Given FICON is between 800 and 1000
- So for Average Frame Size of 1000, the Theoretical Number of Buffer Credits needs to be Approximately Doubled
- Recall that the Average Frame Size for Ports is Reported in the FICON Director Activity Report within RMF
  - Based on data from the CUP
- For Instance, 8G ISL at 100Km needs approx. 800 Buffer Credits
- Depending on the Number of Buffer Credits Needed, the Cisco MDS can have up to 4K per port (with added license feature)

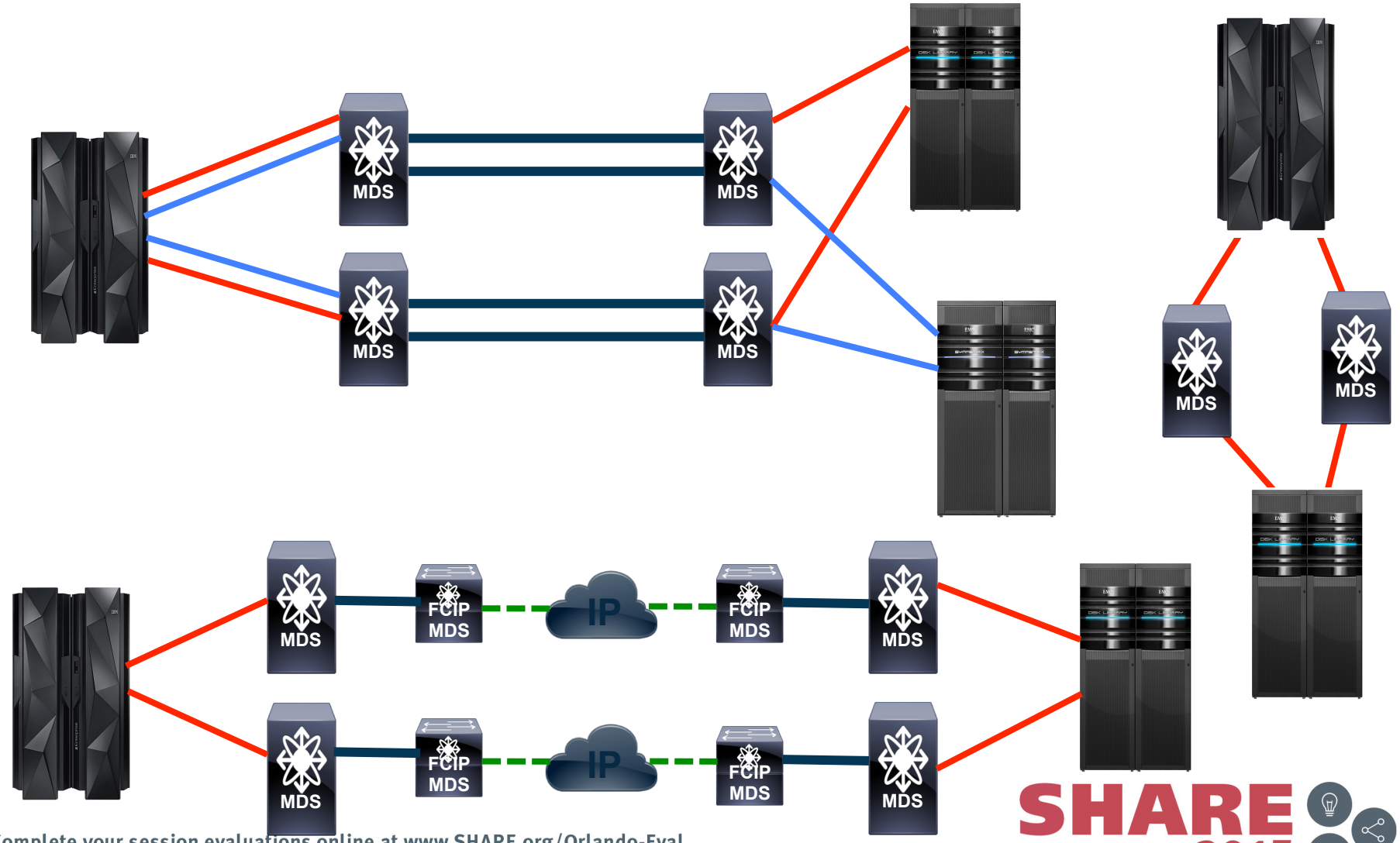
ISL Tips

- Use Port Channels for highly redundant ISLs
  - Port Channels are a base feature and require no extra license
- Insure that FICON VSANs use only SRC/DST balancing
  - Until EBR Feature is released, do not use
  - If used without System z fixes, IFCCs can occur
- For DWDM Systems, Check Interoperability with System z
  - GDPS Qualification on IBM site or from Vendor
- Check for Interoperability of DWDM features with FICON Director Features
- For System z, Long Distance ISLs should always used a FICON Director
  - Recent System z Cards have Stopped Raising Buffer Credits

ISL Tips

- Take Care in Estimating Needed Buffer Credits for ISLs
  - Validate Average Frame Size if Possible
- Monitor Buffer to Buffer transitions to 0
  - Some Number is ok but Excessive Number can be an Issue
- For ISLs connecting 2 Datacenters
  - What is the Expectation during a Trunk Cut
  - Is 2x Provisioning Required based on this ?

IBM FICON Supported Topologies



IBM Qualification Letters / Information - Tips

- IBM Publishes FICON Qualification Letters
  - Verify that FICON Qualified Code is being Implemented
  - Carefully read about any IBM Notes / Caveats from Testing
  - Qualification Letters are available on IBM web / Vendors
- IBM GDPS Qualification Letters for DWDM equipment
  - Verify IBM Testing and Support of Cards

Questions ?

Comments ?

Tips You Want to SHARE ?

Thank You For Attending!

Session 17926

Hints, Tips, and Best Practices For Implementing FICON fabrics
(As well as how to avoid common pitfalls).

Please Fill Out You Session Evaluation

#SHAREorg

60 YEARS OF SHARE
Influencing IT Since 1955