Session 9857

Introduction to Storage Technologies
SAN (**S**torage **A**rea **N**etworking)
FICON (**FI**ber **CON**nection)

David Lytle - Brocade
dlytle@brocade.com

Mike Blair   - Cisco
mblair@cisco.com

SAN 101

S H A R E
Technology · Connections · Results

Agenda for Session's 9857 and 10079

1st Session 9857

- **History**

- **Terminology, Connectors, Cables, and Wavelengths**

- **ESCON Status, zHPF and NPIV**

2nd Session 10079

- **Types and Components of Storage**

- **Let's talk Fibre Channel**

- **FC Buffer Credits**

- **Fabric Routing / Virtual Fabrics / Partitioning**

- **Security / Zoning**

Brocade, Cisco and IBM Graphics are used throughout this presentation.
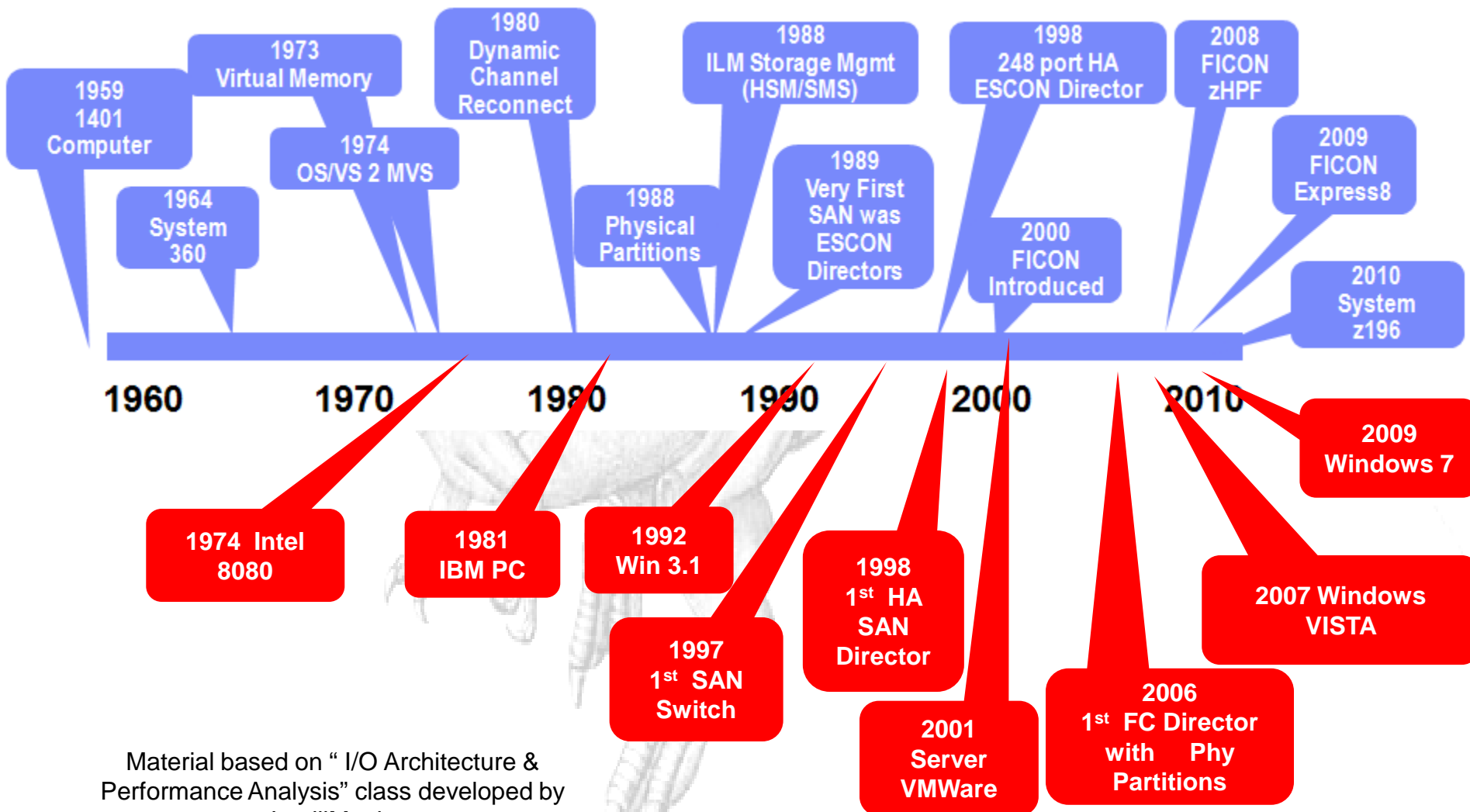
Agenda for Session 9857 and 10079

Session 9857

- **History**

- Terminology, Connectors, Cables, and Wavelengths

- ESCON Status, zHPF and NPIV

Mainframe And Open System Time Lines

1959
1401
Computer

1964
System
360

1973
Virtual Memory

1974
OS/VS 2 MVS

1980
Dynamic
Channel
Reconnect

1988
Physical
Partitions

1988
ILM Storage Mgmt
(HSM/SMS)

1989
Very First
SAN was
ESCON
Directors

1998
248 port HA
ESCON Director

2000
FICON
Introduced

2008
FICON
zHPF

2009
FICON
Express8

2010
System
z196

1960   1970   1980   1990   2000   2010

1974 Intel
8080

1981
IBM PC

1992
Win 3.1

1997
1st SAN
Switch

1998
1st HA
SAN
Director

2001
Server
VMWare

2006
1st FC Director
with   Phy
Partitions

2007 Windows
VISTA

2009
Windows 7

Material based on " I/O Architecture &
Performance Analysis" class developed by
IntelliMagic

4

Agenda for Session 9857 and 10079

Session 9857

- History
- **Terminology, Connectors, Cables, and Wavelengths**
- ESCON Status, zHPF and NPIV

SAN Terminology

- "Fibre" is the protocol/architecture used to transport frames
  - As in "the fibre channel protocol"

- "Fiber" is the glass cable used to attach connections
  - As in "the fiber cable"

- Allows up to 16 million nodes of connectivity

- Historically has used 8-bit/10-bit encoding/decoding to translate an 8-bit byte of information to 10-bit format for serial transmission within the payload of a fibre channel frame

- Variable frame size with a maximum user payload of 2112 bytes. Supports block transfers up to 128MB in size.

- Supports full duplex transmission (simultaneous send and receive)

SAN Terminology

- Light wavelengths in fiber are expressed in nanometers

- The speed of light in fiber cable is 2/3 the speed of light in a vacuum
  - Light travels at ~5nsec per meter in glass

- **Multimode fiber** is used for numerous frequencies which are all short-wave frequencies (62.5, 50 micron) of laser light.
  - Always used with short wave optics (transceivers)
  - This is used for local distance connectivity (~50-500 m)

- **Single-mode fiber** has a smaller core that allows only one frequency of light (9 micron) which is long-wave laser light.
  - Always used with long wave optics (transceivers)
  - This is used for longer distance connectivity (up to 10km)

- Optical power budgets, or link loss budgets, measured in decibels (dBs), are used to manage optical signal loss.

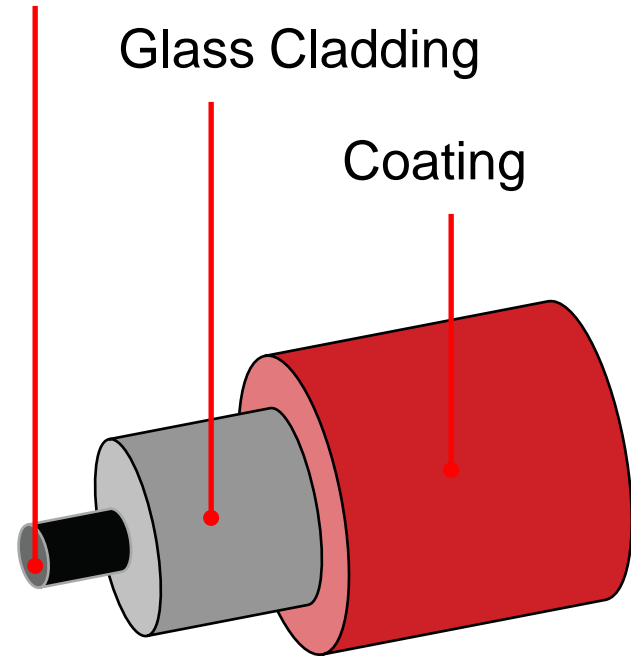# SAN Terminology

Glass Core

Glass Cladding
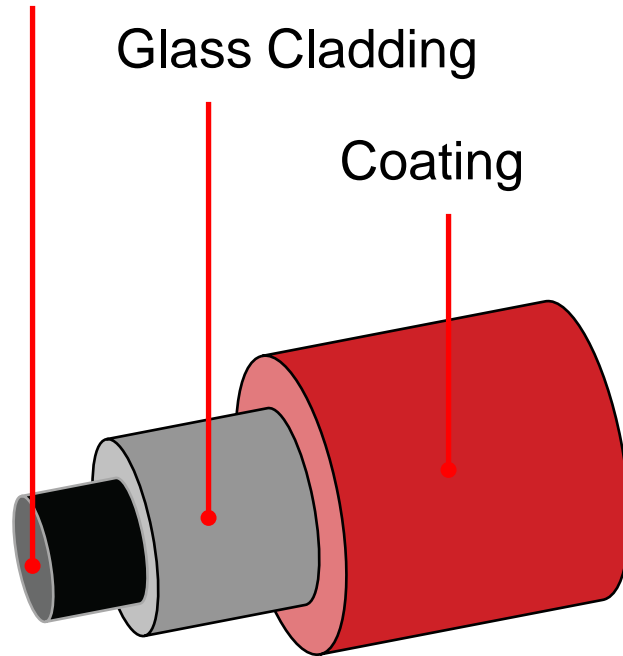
Coating

Glass Core

Glass Cladding

Coating

Multimode Fiber
For local distances and/or
1/2/4/8Gbps
MM – use with short wave optics
62.5 micron, 50 micron
(62.5m has little value today)

Single-mode Fiber
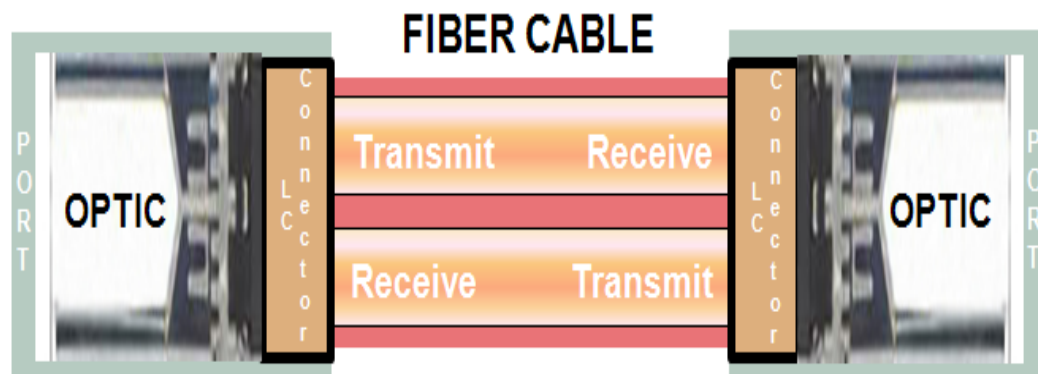For long distances and/or
8Gbps+ link rate
SM – use with long wave optics
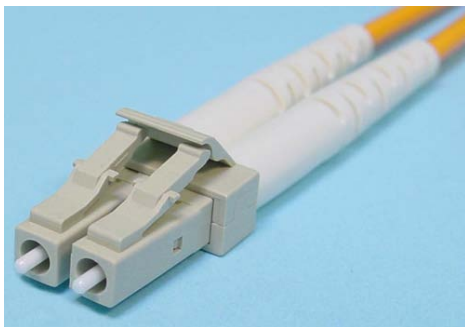9 micron

SAN Terminology
Fibre Channel Link

- **A link consists of**
  - 2 unidirectional "fibers" transmitting in opposite directions
  - May be either:
    - Optical fiber or Copper

- **Transmitters may be:**
  - Long wave laser
    - There can be multiple distances for these – ie. 4km/10km
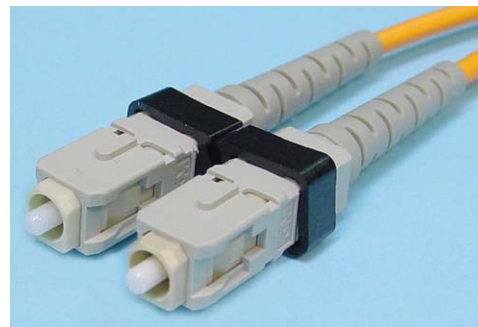 - Short wave laser
  - LED
  - Electrical

SAN Terminology

OPTICAL

- **LC Optical Connector**
  - **Standard on 2-8Gbps Switches**
  - **Most widely used connector**

  Cable Shown as Bonded Duplex

- **SC Optical Connector**
  - **Standard on 1 Gbps Switches**
  - **Little used any longer**

  Cable Shown as Bonded Duplex

COPPER

- **HSSDC Copper Connector**
  - **Smaller than older connectors**
  - **Easier to insert/remove**

- HSSDC2 Copper Connector
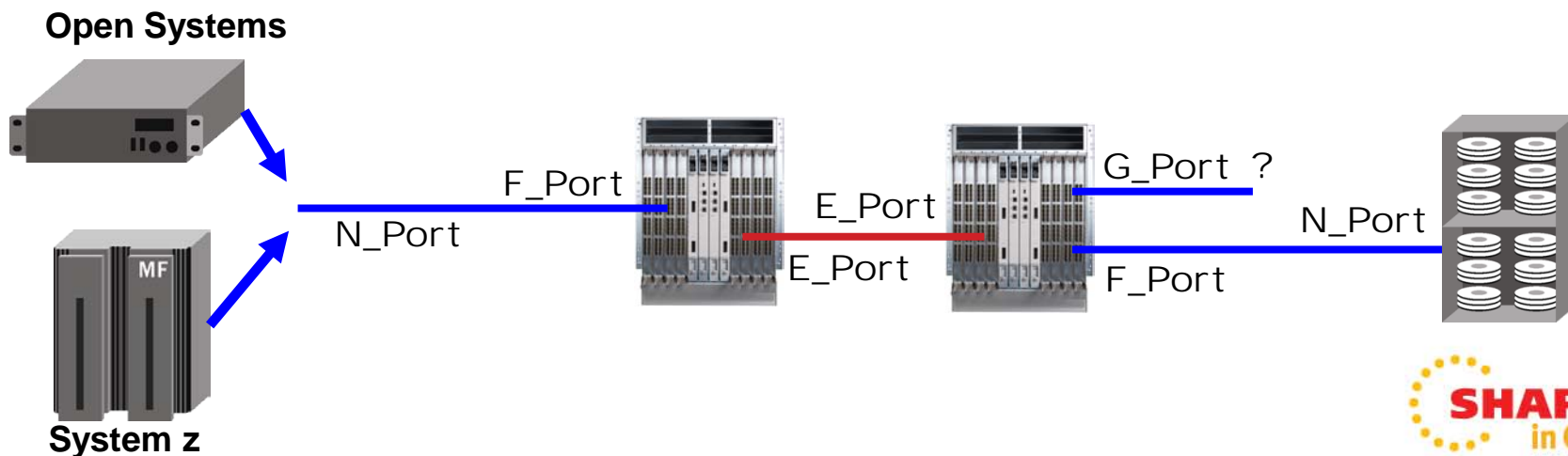 - Fits in SFP Media Slots
  - Smaller than HSSDC

SAN Terminology

Device Ports

- N_Port – Node Port, a Fabric device directly attached

Switch Ports

- G_Port – Generic Port, a port waiting to become an E or F_Port
- F_Port – Fabric Port, a port to which an N_Port attaches
- E_Port – Expansion port used for inter-switch links

Open Systems

F_Port

N_Port

G_Port ?

E_Port

E_Port

N_Port

F_Port

System z

SAN Terminology

- The interconnection between switches is called the Inter-Switch Link (ISL) or in FICON a Cascaded Link
  - E_Port to E_Port (so it is also known as an E_Port)
  - For FICON, a 10Gbps link can ONLY BE a cascaded link (ISL)

- Allows switches to be connected together to create a multi-switch Fabric

- Supports all classes of service
  - Class 1, 2, 3, and a special Class F

- The FC Standard permits consecutive frames of a sequence to be routed over different, parallel ISL links for maximum throughput
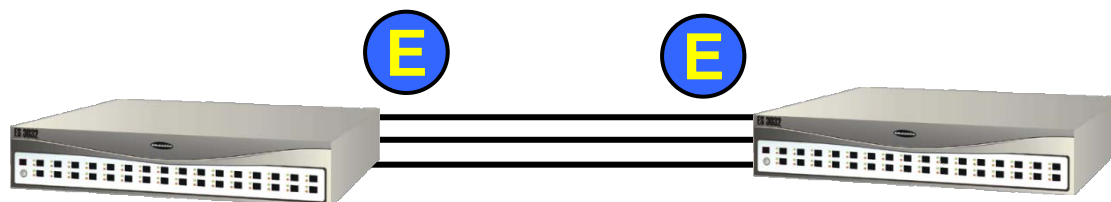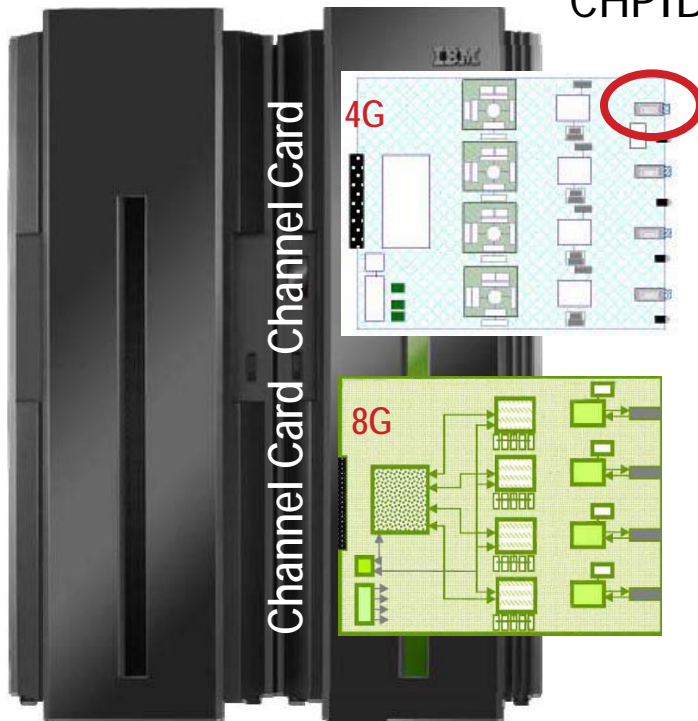
Why Customers Use Mainframes

- Organizations that run an IBM System z platform have unique business requirements.

- They need a **very powerful, highly available, robust and scalable environment** that can support high-volume transaction processing with demanding batch windows and large, critical application portfolios.

- There are tremendous opportunities (like private cloud computing) for leveraging the vast computing power of System z to help greatly simplify the computing environment while substantially reducing overall costs and power consumption.

- All of this makes System z a critically important platform for the future of large enterprise-scale computing!

Mainframe Terminology

System z

This is one CHPID

4G

8G

Channel Card   Channel Card

Up to 336 FICON CHPIDs on z9 / z10
Up to 288 FICON CHPIDs on z196
Up to 128 FICON CHPIDs on z114

All CHPIDs on a channel card must
be the same -- LX or SX – no mixture

Channel or Channel Path Identifier (CHPID)

A physical connectivity port which is embedded in the mainframe processor's channel system

A CHPID will be either long wave or short wave and depending upon the installed channel card it could be 1Gbps, 2Gbps, 4Gbps or 8Gbps referred to as FICON Express or FICON Express$\underline{X}$ where X is 2, 4 or 8

Mainframe Terminology

System z Mainframe (M/F)

LPARs

From 1 to 60 LPARS on a System z
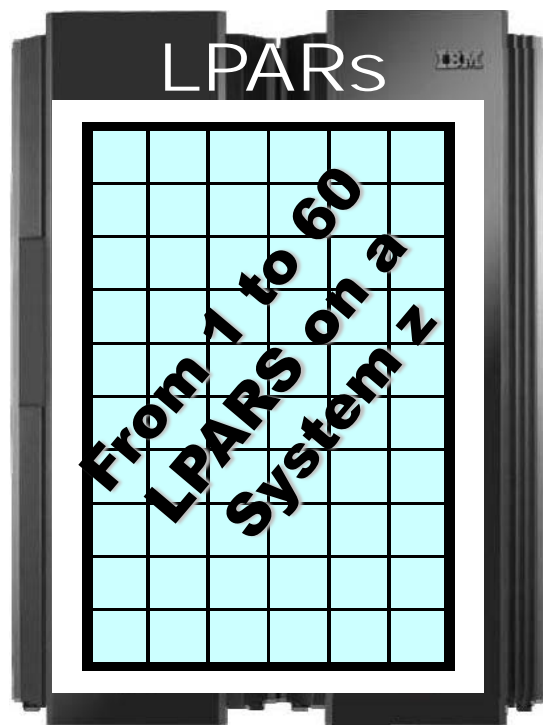
Each LPAR runs its own operating system and has some number of processors assigned to it as CPs and specialty engines

Logical Partitions (LPARs)

- Available since 1988 this is a way of dividing up a mainframe's capacity into **L**ogical **PAR**titions and isolating each LPAR from every other LPAR to ensure RAS.

- System 390 and zSeries could run up to 15 partitions per physical mainframe.

- System z can run up to 60 partitions per physical mainframe.

Mainframe Terminology

Channel Path

The fiber between the channel and the storage subsystem, as well as the interface adapter on the subsystem and any interleaving directors

A channel path can be Point-to-Point path or it can be a switched path
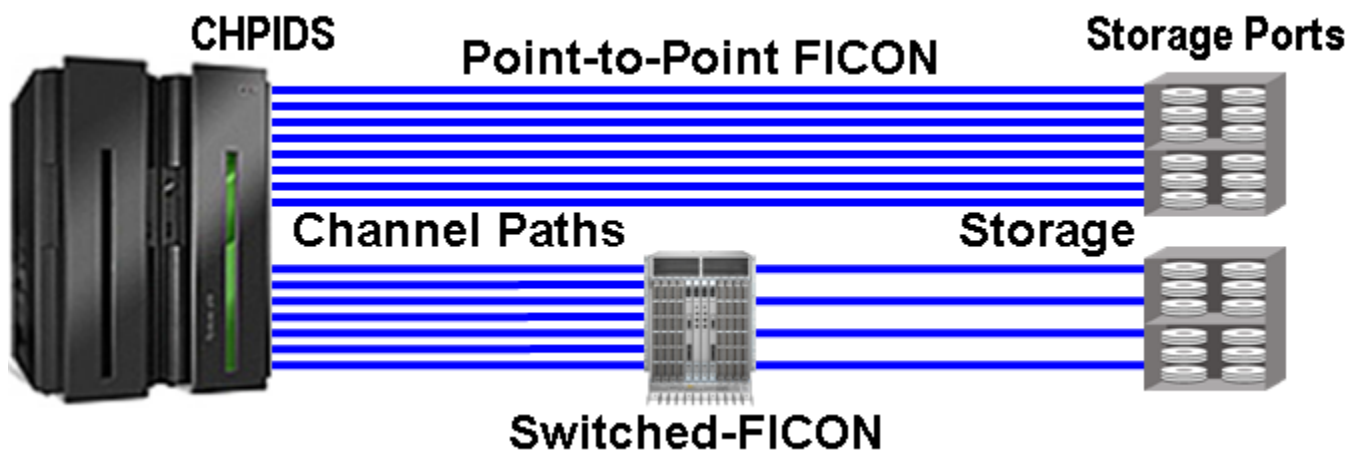
Mainframe Terminology

ESCON Channel Paths – 1990

- This old protocol provides a circuit switched, unidirectional data transfer mechanism.

- Once a data transfer for an I/O from channel to subsystem or subsystem to channel has begun, no other I/O operations can employ the channel until that transfer has been completed.
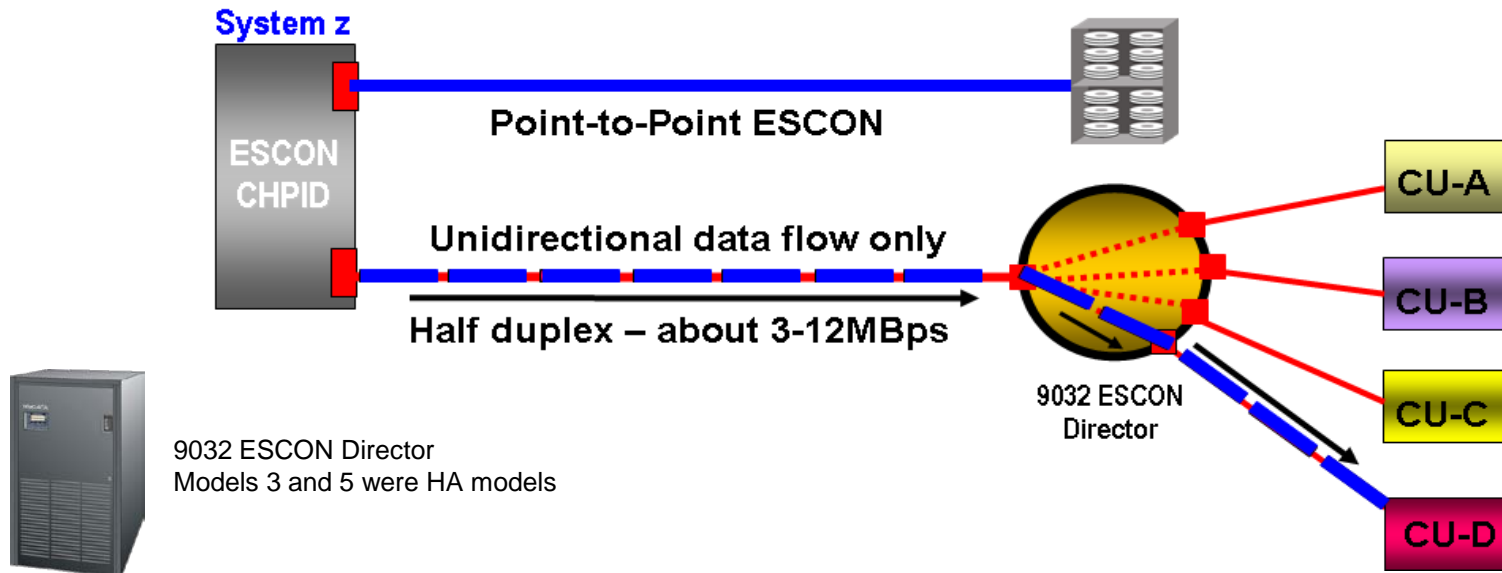
9032 ESCON Director
Models 3 and 5 were HA models

Mainframe Terminology

FICON Channel Paths - 2000

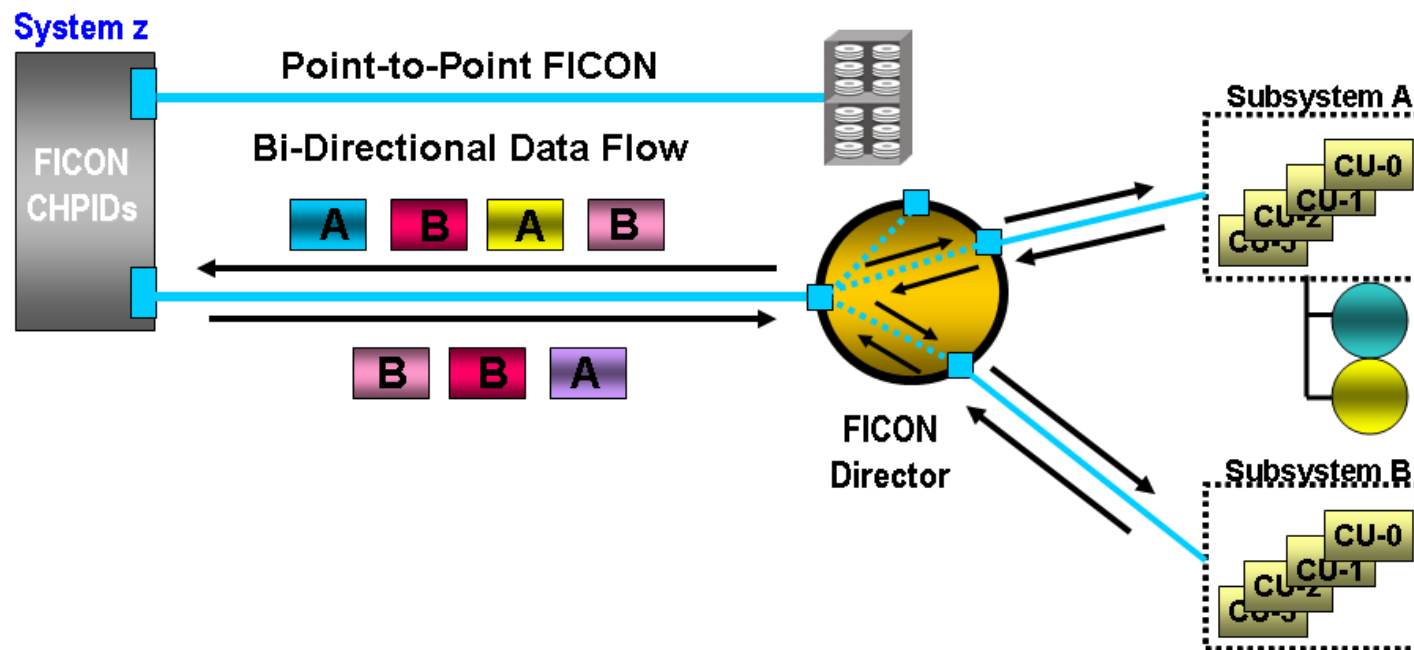- This newer but mature protocol provides a very robust packet switched, bi-directional data transfer mechanism.

- System z10, z196 and z114 can do up to 64 open exchanges unless zHPF is driving the I/O in which case ~600 OEs can be active.

ESCON versus FICON I/O Protocol

- **ESCON** among other things is:

  - Very old (September 1989)
  - Proprietary protocol
  - Half-Duplex at ~10-14 MBps
  - Maximum of about 1,200 IOps
  - Short distances of ~3-9 km
  - Restricted number of control units per channel - 15
  - Only 1,024 devices per channel allowed
  - No channel consolidation going from bus-tag to ESCON
  - Lots of multi-mode optical cables used for connectivity needed to be managed

- **FICON** among other things is:

  - Relatively new (December 2000)
  - FC Standards-based protocol
  - Full-Duplex at 130-1600 MBps
  - 23,000 (CCW) or 92,000 (TCW) IOps
  - Long distances of Local-to-100 km
  - Better use of the 15 control units than ESCON
  - 16,384 devices per channel are now allowed
  - Channel consolidation from 2:1 up to 16:1 when going to FICON
  - Fewer single-mode and/or multi-mode cables are needed makes management easier

So What Is FICON (FIber CONnection)?

- FICON has always been non-proprietary – based upon Fibre Channel Standards

- FICON was a major technical improvement over ESCON and continues to mature and get better and faster all the time

- Worldwide, there are still about ~20 to 30% of ESCON users that still need to completely migrate to FICON
  - The z114 mainframe will be the last mainframe to host ESCON CHPIDs
  - ESCON Directors will soon go to End of Support

- Many factors are pushing end users towards FICON:
  - Mainframes are getting faster and more flexible
  - DASD and Tape storage is getting faster and faster
  - Customer requirements are much wider and deeper
  - IBM support for ESCON is waning

Mainframe Terminology

1971 / 1974
S/390 Mainframe (M/F)

SVS-MVS were early M/F operating systems that managed virtual resources
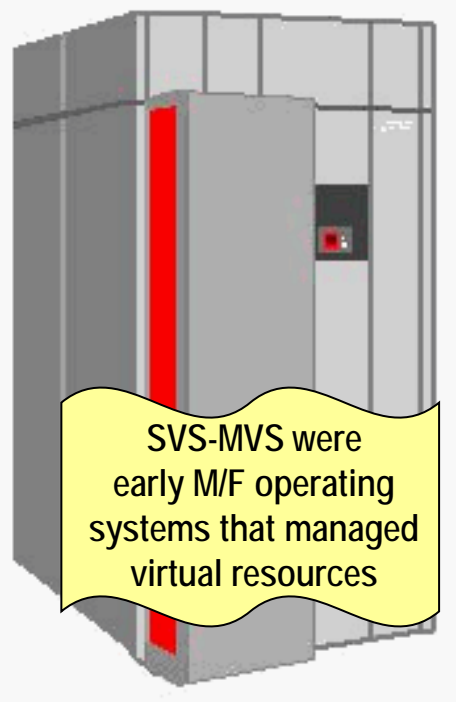
MVS could be run in up to 15 partitions on a M/F or could be intermixed with other operating systems running in other LPARs

Single Virtual Storage (SVS): Multiple Virtual Storage (MVS)

MVS was introduced in 1974, and it is often mentioned, even today, by old mainframer's as their primary operating system

MVS has become the de facto name for the mainframe operating system

This z/OS ancestor was a 24-bit, virtualized, batch processing-oriented operating system that managed lots of memory and DASD space for its time.

Mainframe Terminology

System z Mainframe (M/F)

z/OS is the most current M/F operating system that manages virtual resources

z/OS can run in up to 60 partitions on a M/F or it can be intermixed with other operating systems running in other LPARs

Z Operating System (z/OS)

- Was introduced in 2000 when the zSeries mainframes became available.

- z/OS is a 64-bit server operating system, the latest IBM mainframe operating system, combining MVS and UNIX System Services (formerly known as MVS Open Edition, or OpenMVS).

- As of 2010, on z196, z/OS will also directly connect and manage an IBM BladeCenter with Power 7 blades

Mainframe Terminology

System z Mainframe (M/F)
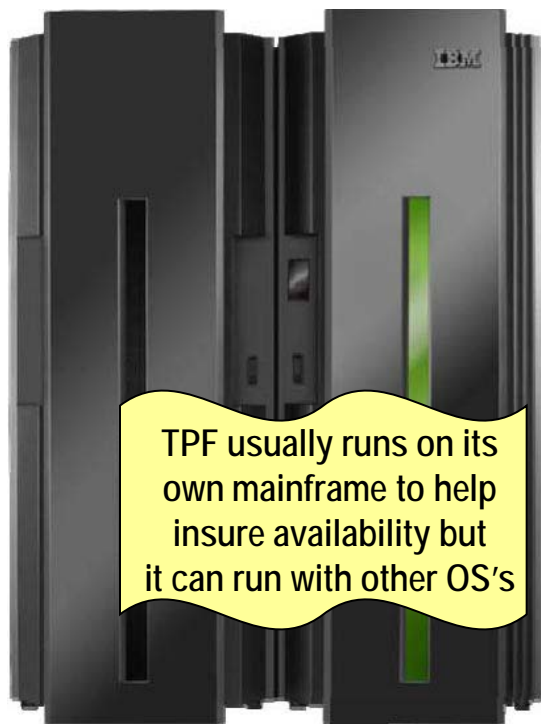
TPF usually runs on its own mainframe to help insure availability but it can run with other OS's

TPF is a risk adverse system that demands the highest performance and availability levels!

Transaction Processing Facility (TPF)

This is a mature real-time operating system that processes many requests very quickly and runs on mainframes like *zSeries* and *System z.*

Any company that needs to process very high volumes of transactions (hotels, airlines, cruise lines, etc.) often utilizes this operating system which requires very high availability.

23

Mainframe Terminology

System z Mainframe (M/F)

VM runs other operating systems, such as Linux, that are "guests" of the VM system
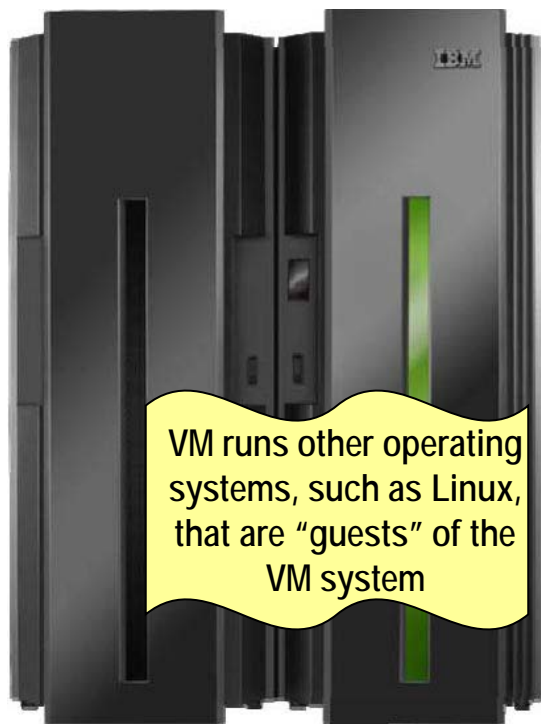
Linux on the System z is most often run as a "guest" under VM and VM actually does the I/O on behalf of Linux

Virtual Machine (VM)

Officially called VM/ESA (Enterprise Systems Architecture), it is an operating system for mainframes that can host other operating systems, including z/OS and Linux.

Each of the guest OS's seems to have its own complete system of software and hardware resources (data storage, processor,…) but are actually sharing resources via VM services.

Mainframe Terminology

System z Mainframe (M/F)

Linux on System z usually runs under VM but it can also run in its own partition(s)

FICON channels in FCP mode for use with zLinux use the Queued Direct Input/Output (QDIO) I/O for communication with the operating system.
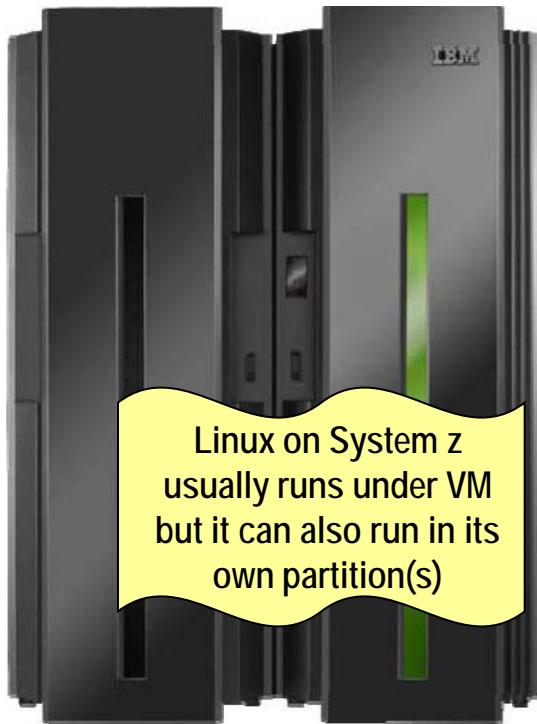
Linux On System z (zLinux)

- Linux on System z allows a customer to leverage their highly available, reliable and scalable mainframe along with their highly available and powerful FCP and FICON infrastructure capabilities

- A Linux administrator now simply administers Linux on a "Big Server"

- Linux has been running on System z since 1999 – twelve years ago!

Mainframe Terminology

Hardware Configuration Definition (HCD)
Define the mainframe's computing and I/O environment

Example FICON IOCP

```
ID

RESOURCE PARTITION=(LP1,1),(LP2,2),(LP3,3)
CHIPID PATH=(AD),SHARED, SWITCH=01,TYPE=FC
CHIPID PATH=(B2),SHARED, SWITCH=01,TYPE=FC
CHIPID PATH=(B6),SHARED, SWITCH=01,TYPE=FC
CHIPID PATH=(BA),SHARED, SWITCH=01,TYPE=FC

CNTLUNIT CUNUMBER=5000, PATH (AD,B2,B6,BA),
    UNITADD=((00,256)),LINK=(0114,0116,011A,011C),
    CUADD=0,UNIT=2105

    *

CNTLUNIT CUNUMBER=5700, PATH (AD,B2,B6,BA),
    UNITADD=((00,256)),LINK=(0104,0106,010A,010C),
    CUADD=7,UNIT=2105

IODEVICE
```

LP1-1 | LP2-2 | LP3-3

System z

AD B2 B6 BA

FICON
Switch.ID 01
Switch.@01

14 16 1A 1C

CU
LCU-7
LCU-n
LCU-0
CU 5000

Devices
Devices
Devices 5000-50FF

HCD is an element of z/OS that provides the interactive tool which is used to define the hardware configuration for both a processor's channel subsystem and the operating system running on the processor.
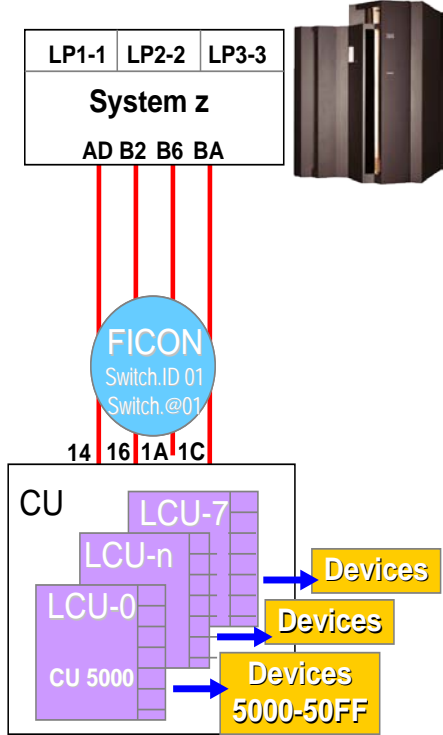
Hardware Configuration Definitions (HCD)

- HCD provides an interactive interface that allows customers to define the hardware configuration for both a processor's channel subsystem and the OS running on the processor.

- There really isn't a comparable facility in the distributed world.

Mainframe Terminology

Sysplex and Parallel Sysplex
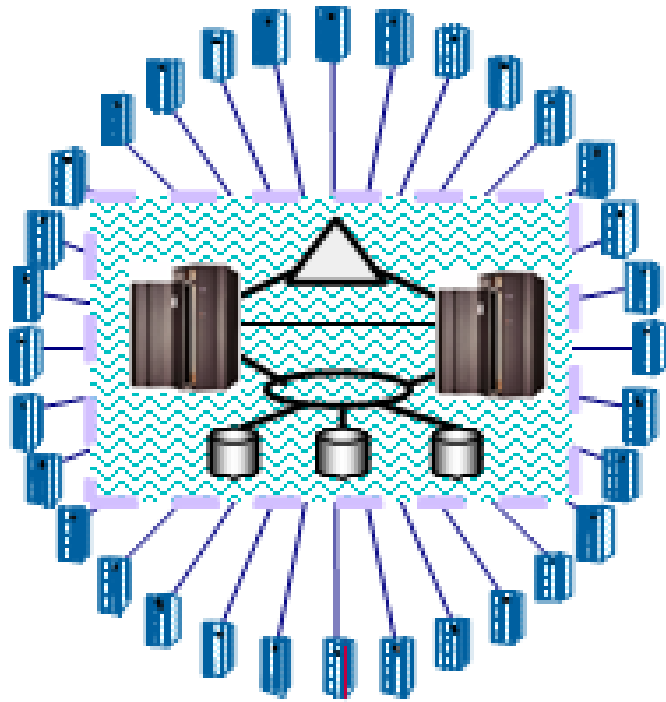
1 to 32 Mainframes

A Parallel Sysplex provides a clustered environment to provide both parallelism for application processing as well as better reliability/availability

Sysplex

- IBM's **SYS**tems Com**PLEX** was introduced in 1990 as a platform for the MVS/ESA operating system for IBM mainframe servers. The sysplex consists of the computer or multiple computers that make up the Computer Electronics Complex (*CEC*).

- *Parallel Sysplex* is a Sysplex evolution providing a clustering architecture that has improved communication capabilities and supports more connected Central Processing Complexes (*CPCs*) and more copies of the operating system and is often deployed with a *Workload Manager capability*.

Mainframe Terminology

Geographically Dispersed Parallel Sysplex (GDPS)

Parallel Sysplex

1 to 32 Systems

GDPS

Might Be A Parallel Sysplex

Might Be A Parallel Sysplex
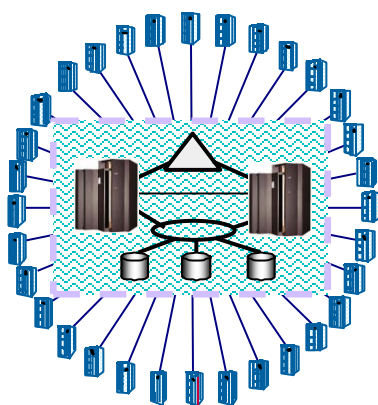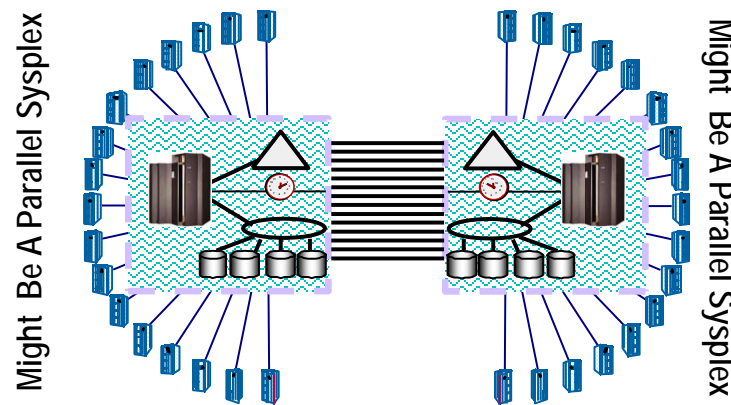
Site 1   **Site 2**

- **Addresses Planned / Unplanned Hardware / Software Outages**

- **Flexible, Non-disruptive Growth**
  - **Capacity beyond largest CEC**
  - **Scales better than SMPs**

- **Dynamic Workload/Resource Management**

- **Addresses Planned / Unplanned Hardware / Software Outages**

- **Addresses Site Failure and Site Maintenance**

- **Robust DASD/Tape Data Replication**
  - **Eliminates SPOF**
  - **Zero / Very Little Data Loss**

- **Application Independent**

What is a GDPS Solution?

- **GDPS is:**
  - **Services offering**
  - **Software**
  - **Automation code**

- **Are FICON Switching Devices GDPS qualified?**
  - **Not automatically!**
    - There is a GDPS qualification beyond FICON qualification
  - **Most, if not all, FICON Directors available today are GDPS qualified**

- **A GDPS Solution includes:**
  - **Two (or more) data center infrastructures**
  - **Plus data replication technology (PPRC, Global Mirror z/OS XRC)**
  - **Plus application replication technology (Parallel Sysplex)**
  - **Plus network infrastructure (for data replication & any-to-any connectivity)**

- **Note about Sysplex Protocols**
  - **ISC3/STP do not run across FICON Directors**
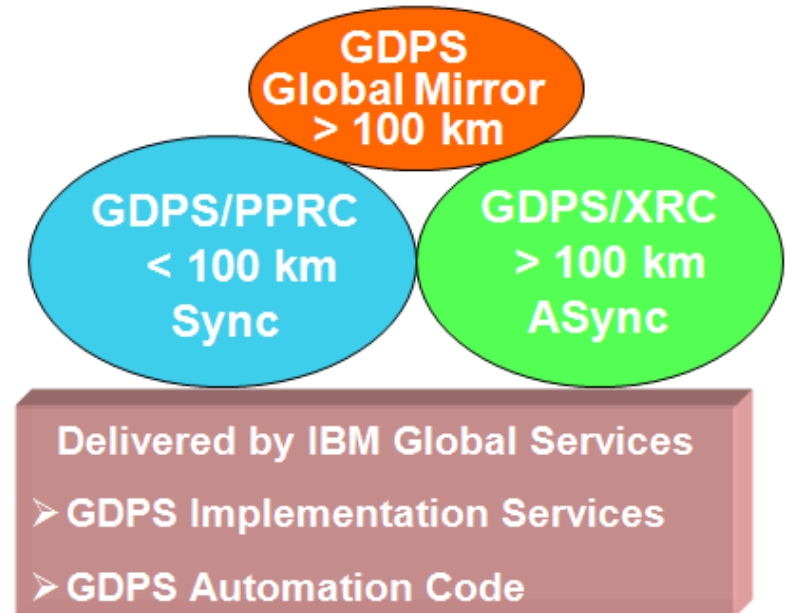
Components of a SAN Connection

Server **Channel** **Transceiver** **Connector** **Fiber** **Storage Network** **Disk** **Tape**

FC HBA or
FICON Channel
Card

LED (no longer used)
Laser-SX
Laser-LX

LC Duplex MM
LC Duplex SM
SC Duplex MM
SC Duplex SM

50 or 62.5 micron
Multimode (MM)

9 micron SM
 - Jumpers
 - Patch Cables
 - Conversion Kits

Switches
Directors

Storage **Storage**
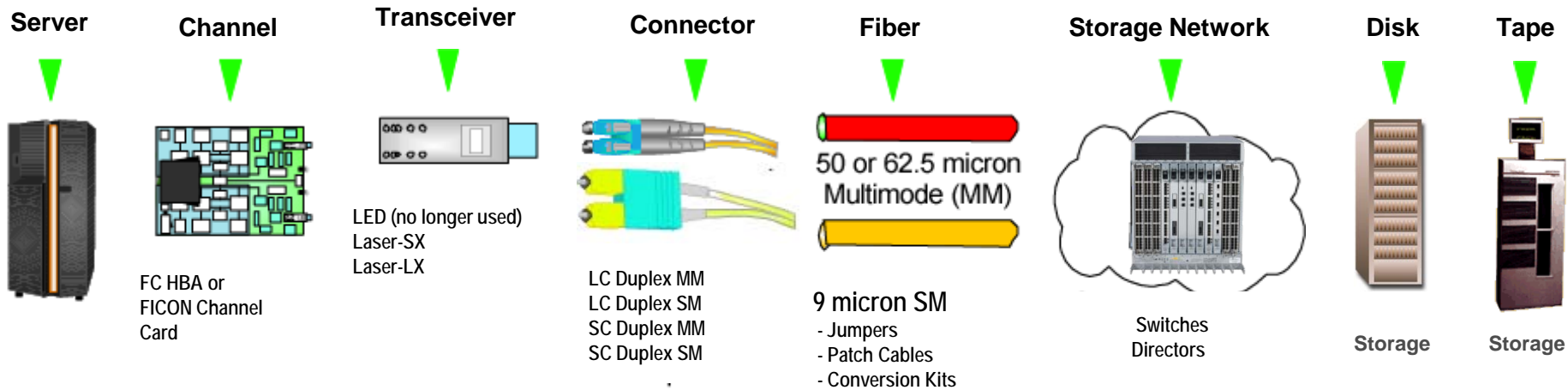
This is a component summary of how everything works together in order to effectively deploy a fibre channel I/O infrastructure.

Open Systems compared to Mainframe

Disk

HOSTS Open Systems

Server HBA

Uses Zoning, PDCM, Trunking and Virtual Fabrics

Switch / Director

Storage Adapter

Optic
2/4/8Gbps
SX / LX

Fibre Cable
MM / SM

Fibre Cable
SM / MM

Optic
2/4/8Gbps
LX / SX

LUNs

Path

2/4/8/10Gbps

FICON Express Channel Card

Uses HCD, Zoning, PDCM, Trunking and Virtual Fabrics

Director / Switch

DASD

Storage Adapter

Optic
2/4/8Gbps
LX / SX

Fibre Cable
SM / MM

Fibre Cable
MM / SM

Optic
2/4/8Gbps
SX / LX

Volumes, Datasets, Unit Control Blocks (UCBs)

CHPID/Channel

2/4/8/10Gbps

Mainframe

Multi-mode OM2
50 micron 500mHz
2/4 Gbps

Multi-mode OM3
50 micron 2000mHz
2/4/8/10Gbps

Multi-mode OM4
50 micron 4700mHz
2/4/8/10Gbps

Single-mode OS1
9 micron
2/4/8/10Gbps ~ 10k

SHARE Technology · Connections · Results

SHARE in Orlando 2011

Agenda for Session 9857 and 10079

Session 9857

- History
- Terminology, Connectors, Cables, and Wavelengths
- **ESCON Status, zHPF and NPIV**

ESCON Statement of Direction

ESCON channels will be phased out

* It is IBM's intent for ESCON channels to be phased out. System z10 EC and System z10 BC will be the last server to support greater than 240 ESCON channels.

* Released April 28, 2009

* Currently, 1024 channels are supported on z10 EC and 480 channels are supported in the z10 BC

* Only 240 ESCON channels are supported on z196 …AND… z196 and z114 will be the last mainframes to natively support ESCON channels

FICON to ESCON Converter

Promotes FICON infrastructure modernization

Preserves ESCON infrastructure investments

Replaces ESCON directors, FICON bridge, and ESCON extenders

ISL

FICON

Prizm 2:8

Native FICON

FICON Tape & DASD

ISL

FICON

Prizm 2:8

ESCON Tape
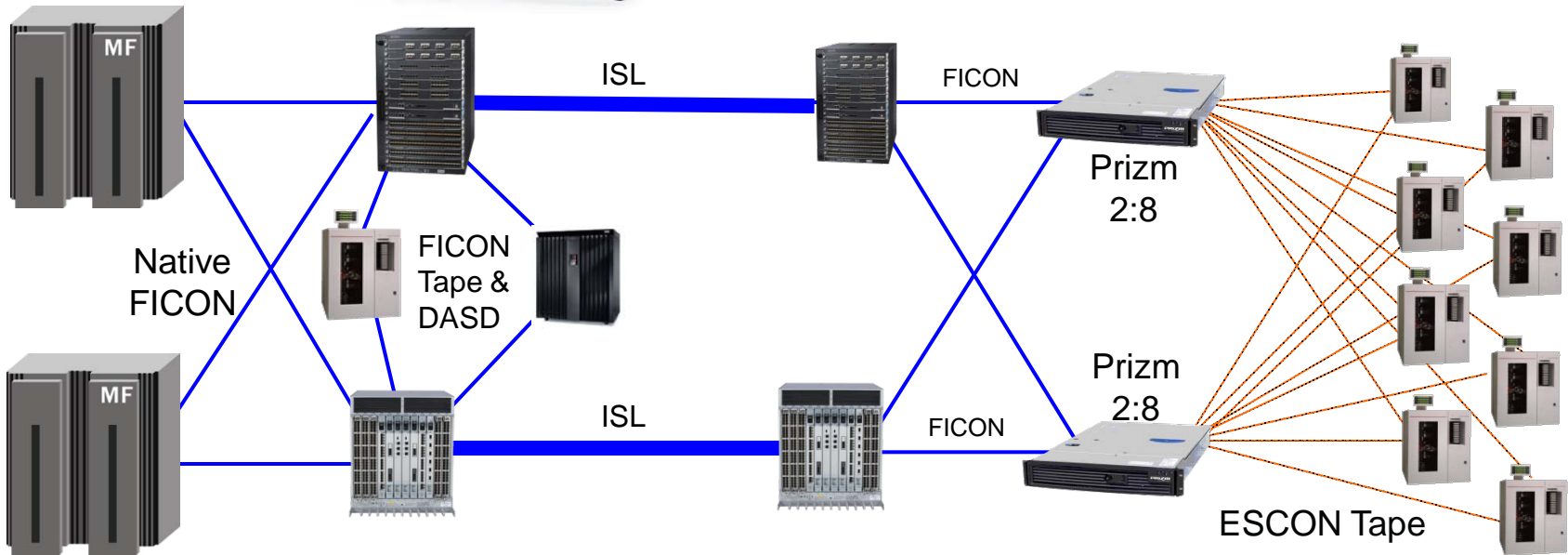
zHPF – High Performance FICON

- It was first introduced in 2008

- It continues to be enhanced

- Currently used with Media Manager, extended format data sets

- Encapsulates multiple FICON CCWs into a single frame

- z10, z196 or z114 must be the host

- Supported by the major storage vendors

- Enhanced Performance

FICON CCWs into FCP-like Frames

FICON

FICON
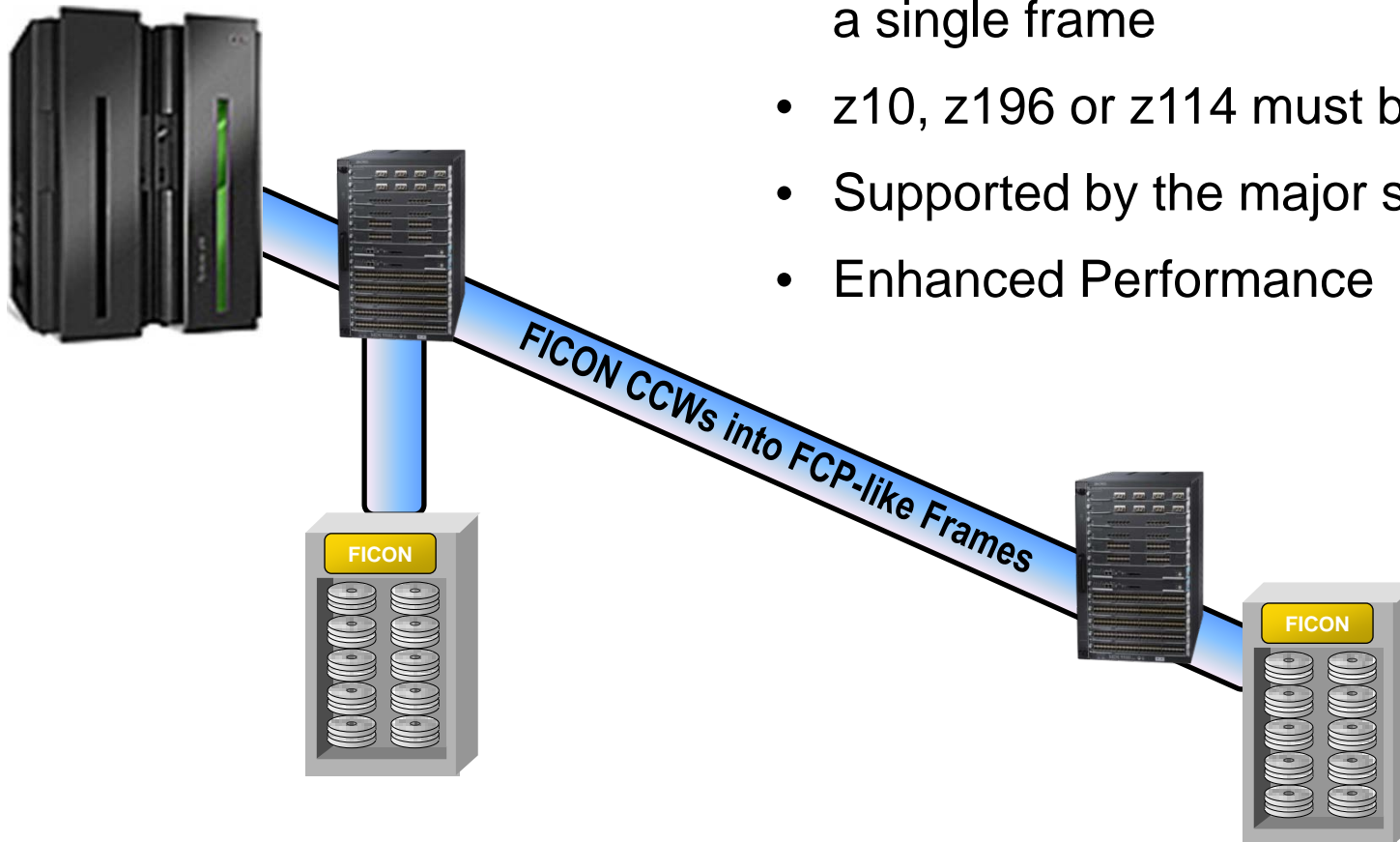
NPIV – Node_Port ID Virtualization

- NPIV is standards based

- NPIV is used on the mainframe when Linux utilizes FCP Ports

System z using NPIV

Linux Guests

Linux A

Linux B

Linux C

Linux n

V M

I/O

A

B

C

n

IOS

NPIV works only when using switched-FICON

One FCP channel for many Linux guests

n   C   B   A

Lots of Parallelism

Fewer switch ports required!

NPIV enabled Switches and Directors

Much better I/O bandwidth utilization per path
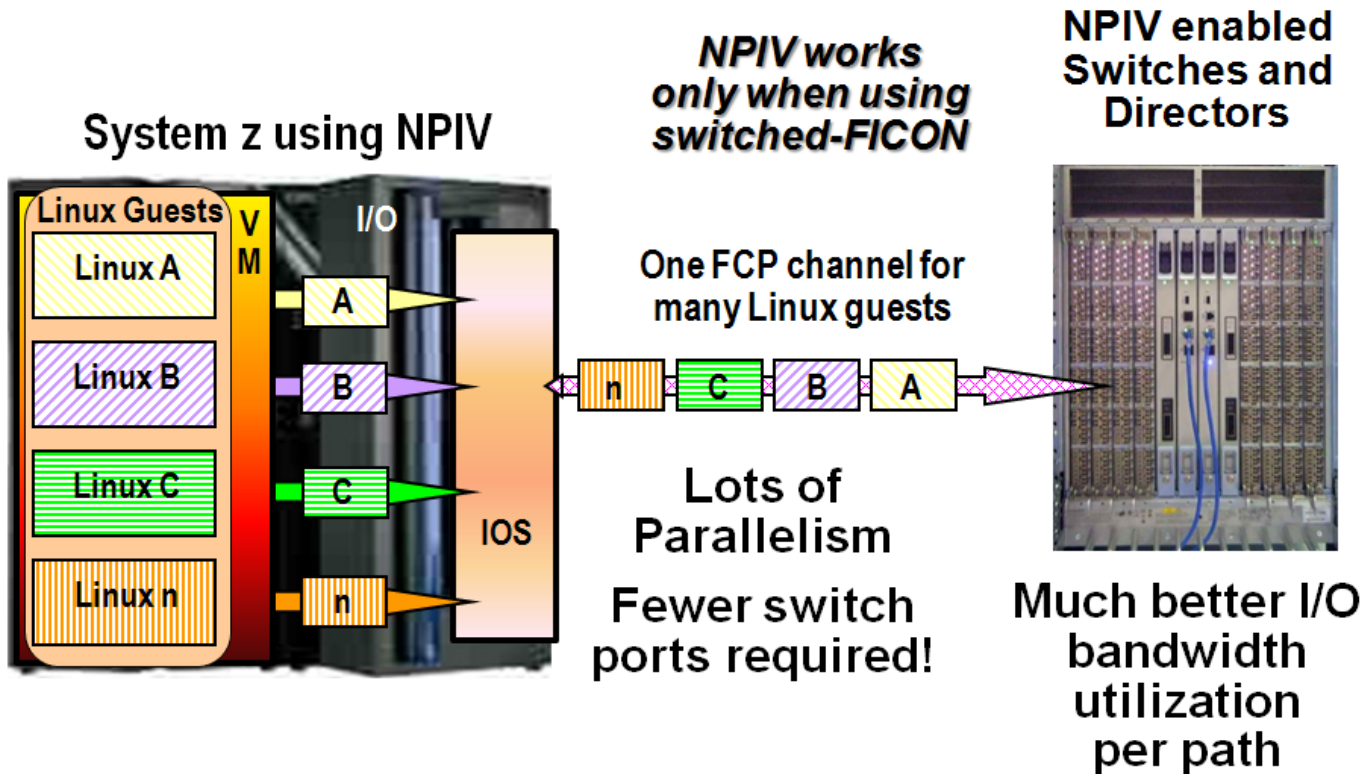
8 Gbps Is Great For NPIV!

This Is The End Of Part 1

The 2nd of this 2-part session is

Session **10079**

Monday: 11:00AM – 12:00PM

SAN Sessions at SHARE this week

Monday:

Time-Session

0930 - 9857: <u>SAN (Storage Area Networking) Overview – Part 1</u>

1100 - 10079: <u>SAN (Storage Area Networking) Overview - Part 2</u>

1330 - 9671: <u>Managing SAN for Linux on z/VM - a Nationwide Perspective</u>

1500 - 9931: <u>Buffer-to-Buffer Credits, Exchanges, and Urban Legends</u>

1630 - 9934: <u>Evolution of the System z Channel</u>

Tuesday:

Time-Session

1100 - 9368: <u>A First Look at the Inner Workings and Hidden Mechanisms of FICON Performance</u>

1500 - 10079: <u>A deeper look into the Inner Workings and Hidden Mechanisms of FICON Performance</u>

Wednesday:

Time-Session

0800 - 9479: <u>Planning and Implementing NPIV for System Z</u>

0930 - 9864: <u>zSeries FICON and FCP Fabrics - Intermixing Best Practices</u>

Thursday:

Time-Session

0800 - 9853: <u>FICON Over IP - Technology and Customer Use</u>

0800 - 9899: <u>Planning for ESCON Elimination</u>

0930 - 9933: <u>Customer Deployment Examples for FICON Technologies</u>

1500 - 9316: <u>SAN Security Overview</u>

1630 - 10088: <u>FICON Director and Channel Free-for-all</u>

THANK YOU FOR ATTENDING!

Session 9857

Please fill out your evaluation forms

Session 10079

Introduction to Storage Technologies
SAN (Storage Area Networking)
FICON (FIber CONnection)

Mike Blair   - Cisco
mblair@cisco.com

David Lytle - Brocade
dlytle@brocade.com

SAN 201

SHARE
Technology · Connections · Results

Agenda for Session 9857 and 10079

Session 9857

- History

- Terminology, Connectors, Cables, and Wavelengths

- ESCON Status, zHPF and NPIV

Session 10079

- Types and Components of Storage

- Let's talk Fibre Channel

- FC Buffer Credits

- Fabric Routing / Virtual Fabrics / Partitioning

- Security / Zoning

Agenda for Session 9857 and 10079

Session 10079

- **Types and Components of Storage**
- Let's talk Fibre Channel
- FC Buffer Credits
- Fabric Routing / Virtual Fabrics / Partitioning
- Security / Zoning

Storage Network Components

Multiple Infrastructure Vendors (eg. Brocade and Cisco)
Several components required to build a SAN

- Servers with Host Bus Adapters (HBAs)
- Mainframes with FICON Express Channels
- Storage systems
  - RAID (Redundant Array of Independent/Inexpensive Disks)
  - JBOD (Just A bunch of Disks)
  - Tape
  - VTS/VSM (Virtual Tape)
- Fibre Channel / FICON Switches or Directors
- Ethernet Switches (iSCSI / FCoE)
- SAN management software

Direct Attached Storage

- Direct Attached Storage (DAS)

- Storage is captive 'behind' the server, limited mobility

- Limited scalability due to limited devices

- No storage sharing possible

- Costly to scale

- Management can be complex

- Often cannot take full advantage of the technology
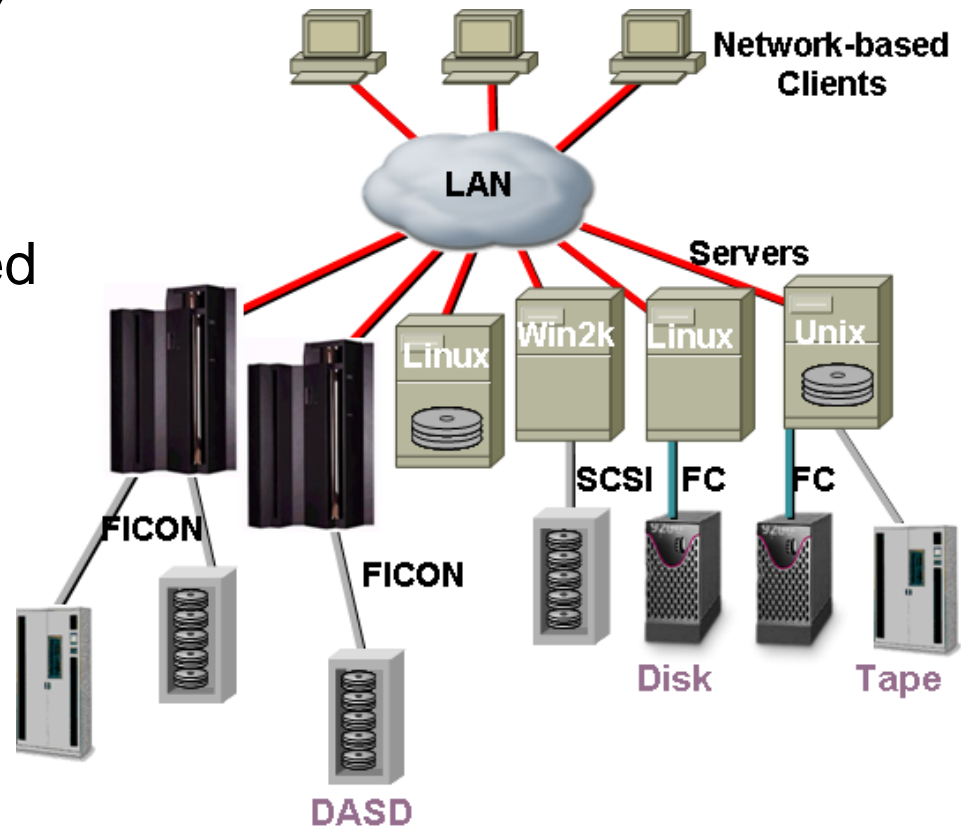
Network Attached Storage (NAS)

- Dedicated file server

- Optimized for file-based access to shared storage over an IP network

- Suitable for applications involving file serving/sharing

- High-performance access, data protection, and disaster recovery

- Capable of storage partitioning

- Uses network file system protocols such as NFS or CIFS

IP Network

NAS Appliance

45
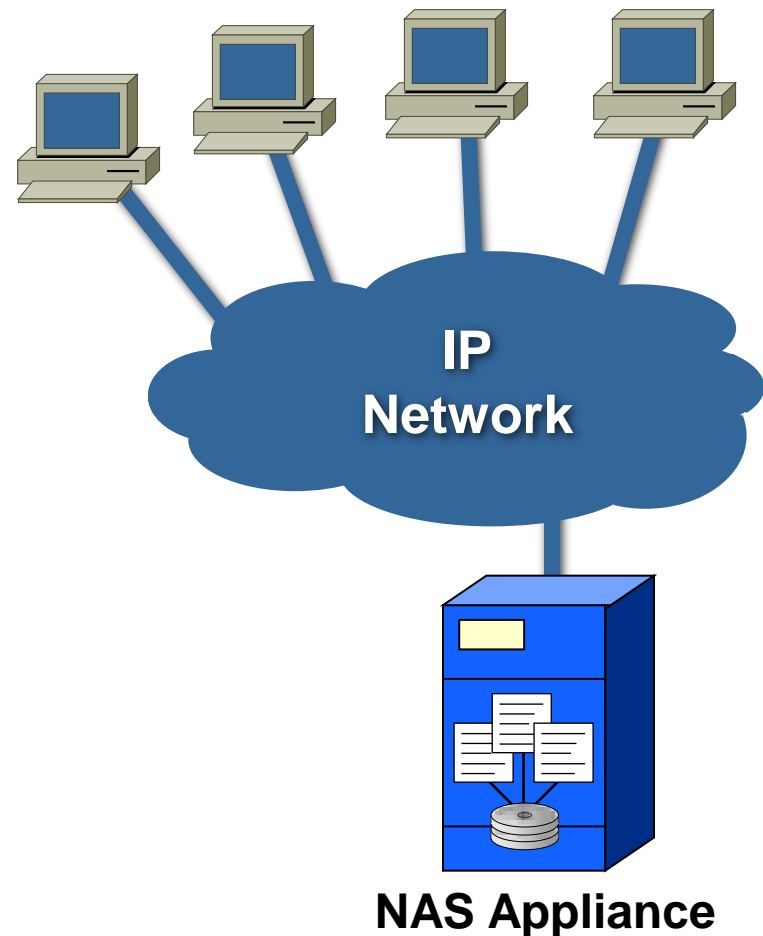
Storage Area Network (SAN)

Separation of Storage from the Server

- Storage is accessed Block-level via SCSI/FICON and can be in a switched environment

- High performance interconnect providing high I/O throughput

- Lower TCO relative to direct attached storage, storage can be shared

- Have to consider Vendor Interoperability / Qualifications

- More Complex management due to size/scale

Clients

LAN

Database/App Servers

FC/FICON SAN

Block Storage Devices

46
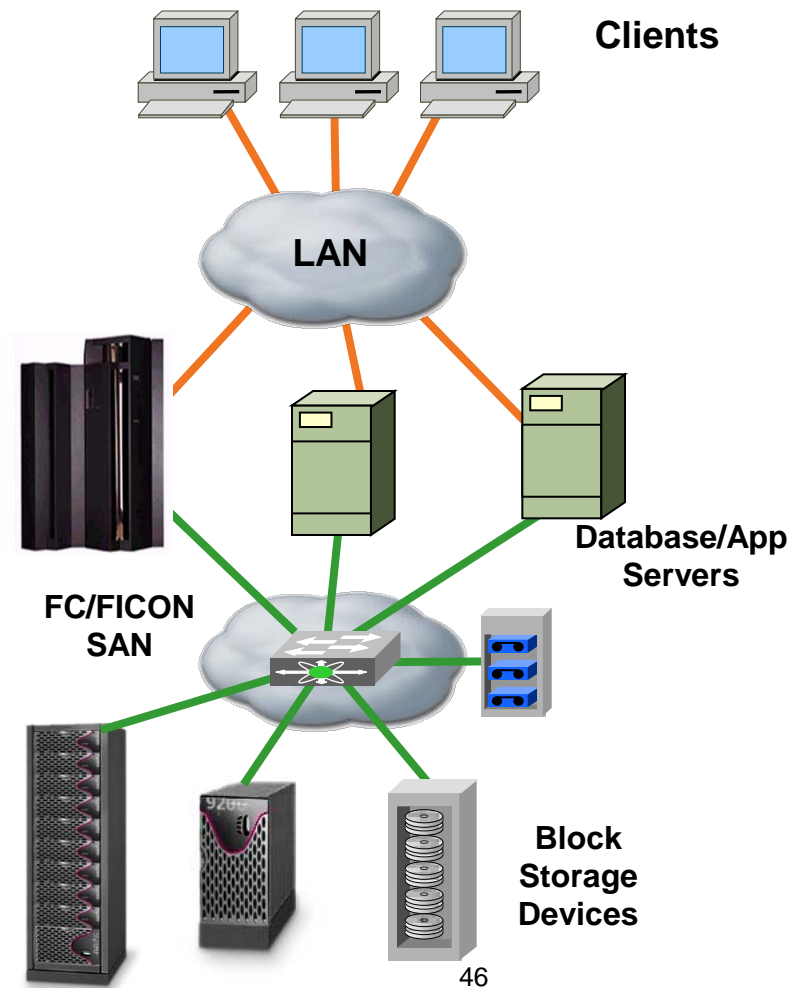
FICON or Storage Area Network (SAN)

Agenda for Session 9857 and 10079

Session 10079

- Types and Components of Storage
- **Let's talk Fibre Channel**
- FC Buffer Credits
- Fabric Routing / Virtual Fabrics / Partitioning
- Security / Zoning

Fibre Channel Protocol

Fibre Channel (FC) provides high speed transport for Upper level (ie. FICON or SCSI) payloads

- FC is the "protocol" for a Storage Network – attributes are:
    - Highly Scale - Addressing for up to 16 million nodes
    - Various Switched Topologies
    - High Speeds - 100, 200, 400 or 800 MBps …and… 10Gb ISLs
    - Segments of up to 100 Km between switches
    - Support for multiple protocols like FICON and OPEN (SAN)
    - Support for Security via Zoning and Prohibit/Allow Matrix

The Fibre Channel Protocol

- **FCP and FICON** are just a part of the upper layer (FC-4) protocol

- **They** are compatible with existing lower layers in the protocol stack

FC-4	**Protocol Mapping Layer** Upper Level Protocol (ULP)	**FCP/FICON/HIPPI/Multi-media, etc.**
FC-3	**Common Services**	**Login Server, Name Server, Alias Server**
FC-2	**Framing Protocol / Flow Control**	**Data packaging, Class of service, Port Login / logout, Flow control...**
FC-1	**Transmission Protocol - Encode / Decode**	**Serial Interface (one bit after another) Frame Transfer (up to 2048 byte payload) 8b/10b or 64b/66b data encode / decode**
FC-0	**Interface/Media – The Physical Characteristics**	**Cables, Connectors, Transmitters & Receivers...**

Fibre Channel Overview

- Fibre Channel is primarily utilized as a networked form of SCSI (open systems - SAN) or CCWs (System z - FICON)

  - **Actually, the lower layers of Fibre Channel are generic**

    - Able to transport multiple data types such as video on demand and Internet Protocol

  - **But, most common deployment is Fibre Channel Protocol (FCP)**

    - FCP is an upper layer protocol that provides for the transmission of SCSI commands and data over the Fibre Channel transport layers

  - **Next most common deployment is Fiber Connection (FICON)**

    - FICON is an upper layer protocol that provides for the transmission of mainframe CCW commands and data over the Fibre Channel transport layers

World Wide Names

- Each switch element is assigned a 64 bit WWN at time of manufacture

- Each switch port is assigned a 64 bit WWPN at the time manufacture

- During Fabric Logon (FLOGI) the switch identifies the WWN in the service parameters of the accept frame

These Address Assignments Can then Correlate
Each Fabric Port with Switch Routing and the Fiber
Channel ID (FCID)

Fabric Addressing

- The 24 bit FCID address is partitioned into 3 fields
  - Device or NPIV
  - Area
  - Domain
- This partitioning helps speed up routing
- Switch element assigns the address to N_Ports
- Address portioning is transparent to N_Ports

	8 bits	8 bits	8 bits
Switch/Open Topology	**Switch Domain**	**Area**	**Device**
FICON Topology	**Switch Domain**	**Port Address**	**0 ….. or NPIV virtual addr.**
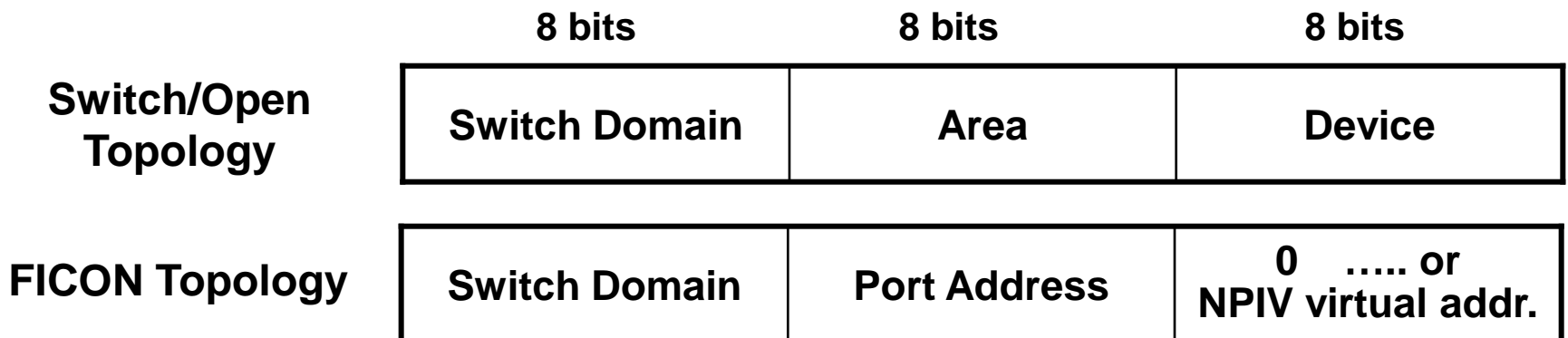
Fibre Channel Frame Format

- All FC-2 frames follow the general frame format as shown below

- Idles are 'Ordered Sets' used for synchronization and basic signaling

- SOF – Start-of-Frame, EOF – End-of-Frame

- 8b/10b frame encoding for 1, 2, 4 and 8Gbps

- 64b/66b frame encoding for 10 and 16Gbps

General FC-2 Frame Format

Frame Content

Idles*	SOF	Frame Header	Data Field	CRC	EOF	Idles*
	(4)	(24)	(0-2112)	(4)	(4)	

0-528 Transmission Word
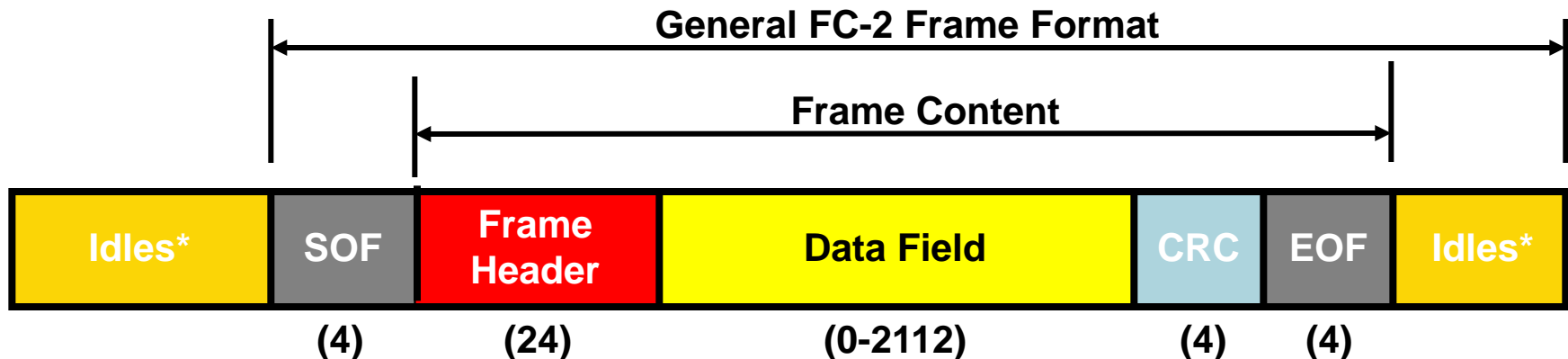
* 6 Idle words (24 bytes) requires by TX
2 Idle words (8 bytes) guaranteed to RX

53

Storage Networking Topology
Dual Star (non-cascaded for FICON)

SAN and/or FICON Fabrics

Application **A**

Application **B**

Application **C**

Shared Storage Resources
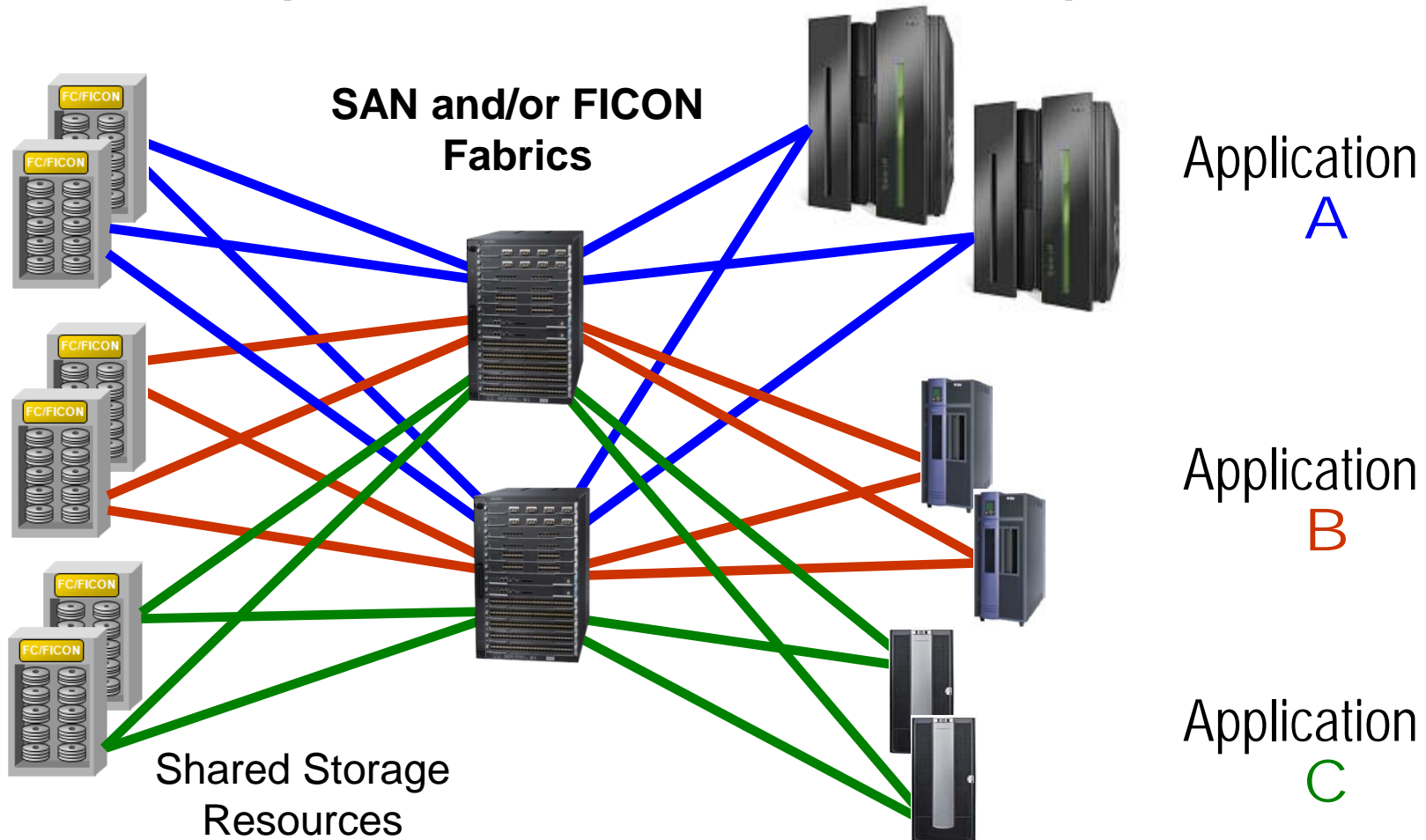
- Provides an opportunity to deploy fabrics to meet five-9′s of availability
- Still must insure <=50 min/annual fabric downtime

Typical SAN Topology
Core-to-Edge (distributed systems only)

**SAN
Fabrics**

Application
B

Shared Storage
Resources

Storage
Resources

Application
C
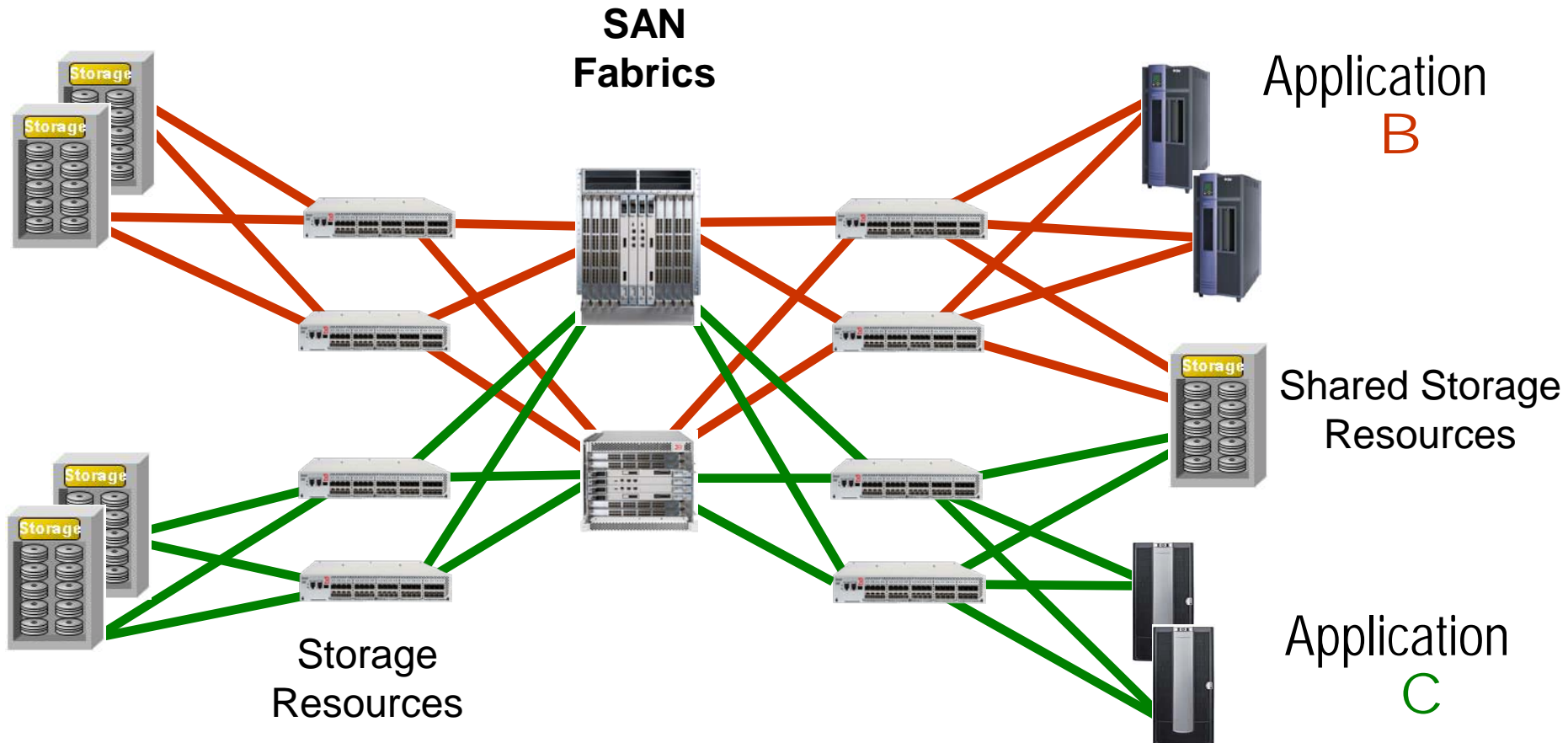
- Requires multi-hop so it is not currently supported for FICON
  due to IBM Qualification

Storage Networking Topology Cascaded FICON and FCP
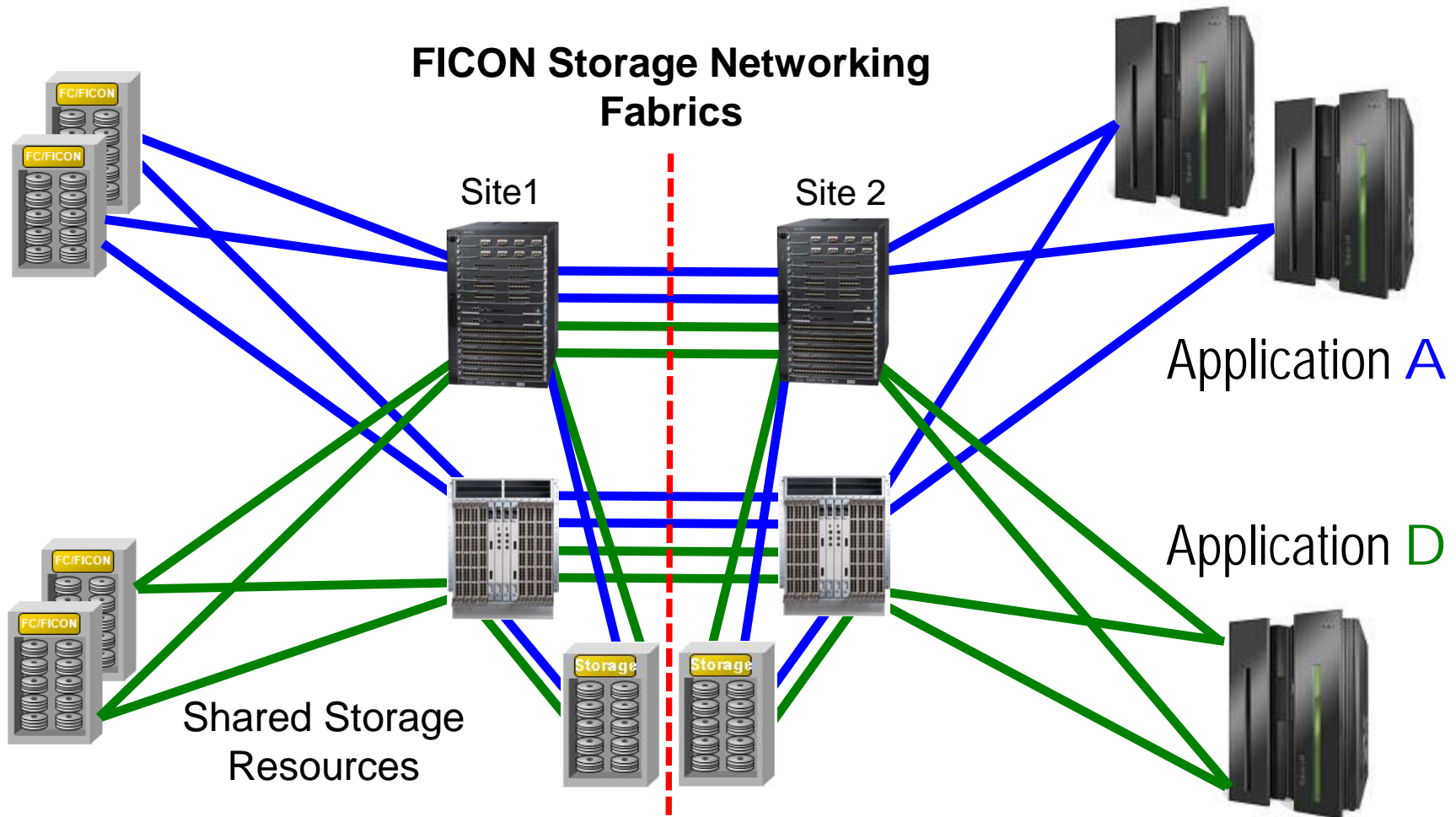
FICON Storage Networking Fabrics

Site1

Site 2

Application **A**

Application **D**

Shared Storage Resources

Storage

Storage

- Provides an opportunity to deploy fabrics to meet five-9's of availability
- Only one vendor per fabric and Only one hop per fabric

Storage Networking Topology Cascaded FICON and FCP
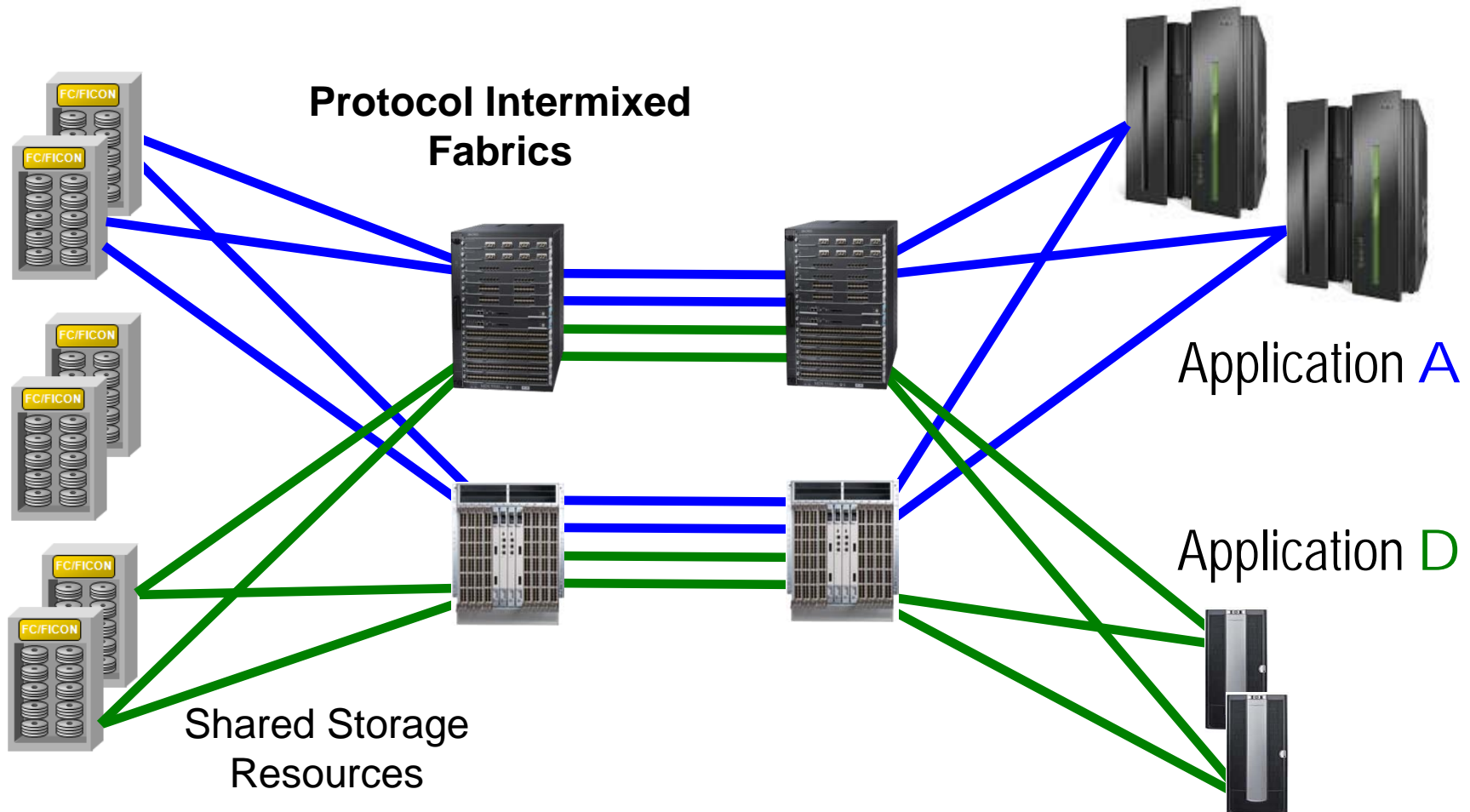
Protocol Intermixed Fabrics

Application **A**

Application **D**

Shared Storage Resources

- Provides an opportunity to deploy fabrics to meet five-9's of availability
- Only one vendor per fabric and Only one hop

ISL Aggregation

Port Aggregation Is Used to Create a Single Logical ISL from multiple Physical ISLs

Different names depending upon the vendor
    Brocade = Trunking
    Cisco = Port Channel

- Increases bandwidth and availability

- Simplifies Topology

- Usually some load balancing

- Interfaces can both be added and removed in a non-disruptive manner in production environments

- Preserves FC guarantee of in-order delivery (IOD)
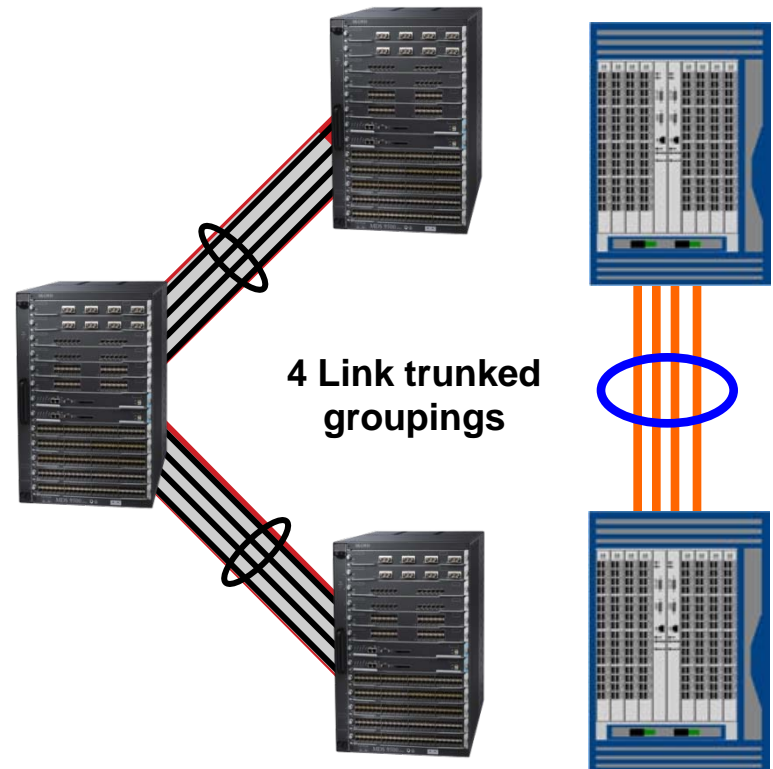
4 Link trunked groupings

Agenda for Session 9857 and 10079

Session 10079

- Types and Components of Storage
- Let's talk Fibre Channel
- **FC Buffer Credits**
- Fabric Routing / Virtual Fabrics / Partitioning
- Security / Zoning

59

FC Buffer Credits and Flow Control

I can receive 40 frames (buffer credits) but after that you will have to stop sending frames until I acknowledge some or all of them.

We are long distance ISL links and therefore we need extra buffer credits in order to keep the link fully utilized.

Director or Switch

Storage

←16
40 →

←202
202 →

←200
16 →
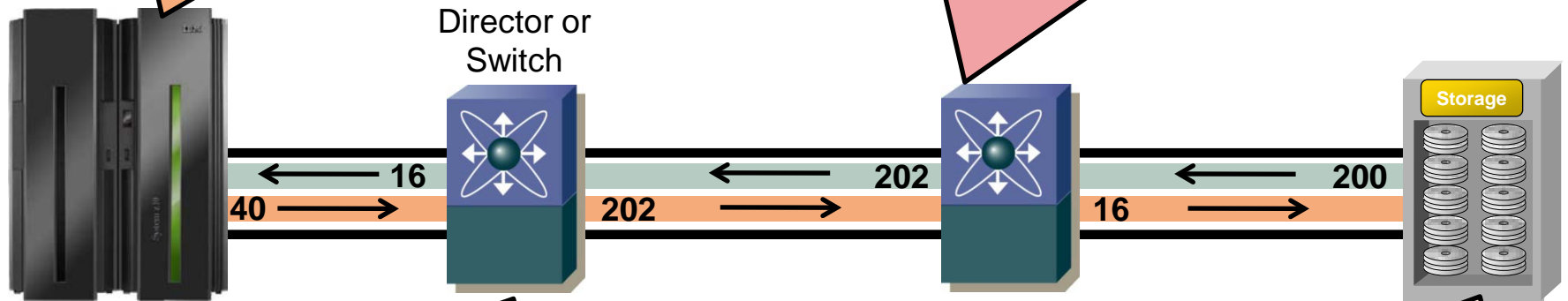
That is fine – I can only receive 16 frames myself. This is OK since we are only a few feet/meters apart anyway.

Like most modern DASD and Tape storage, I have a set number of frames that I can handle and I will let you know what the maximum BC limit will be.

Fiber Cable

Strand 1
Strand 2
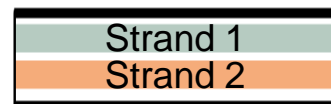
FC Buffer Credits and Flow Control

- BB_Credits are the "admission control" mechanism in FC to ensure that FC switches don't run out of buffers (FC Switches cannot drop frames)

- For Devices operating at FC Class 3 (most devices), Buffer Credits are determined at login.

- BB_Credits are the only flow-control mechanism for FC Class 3.

11 available BB_Credits

15 available BB_Credits

15 available BB_Credits

12 available BB_Credits

16 available BB_Credits

🔴 **Frame towards Disk shelf**

▨ **Return BB_Credit token**
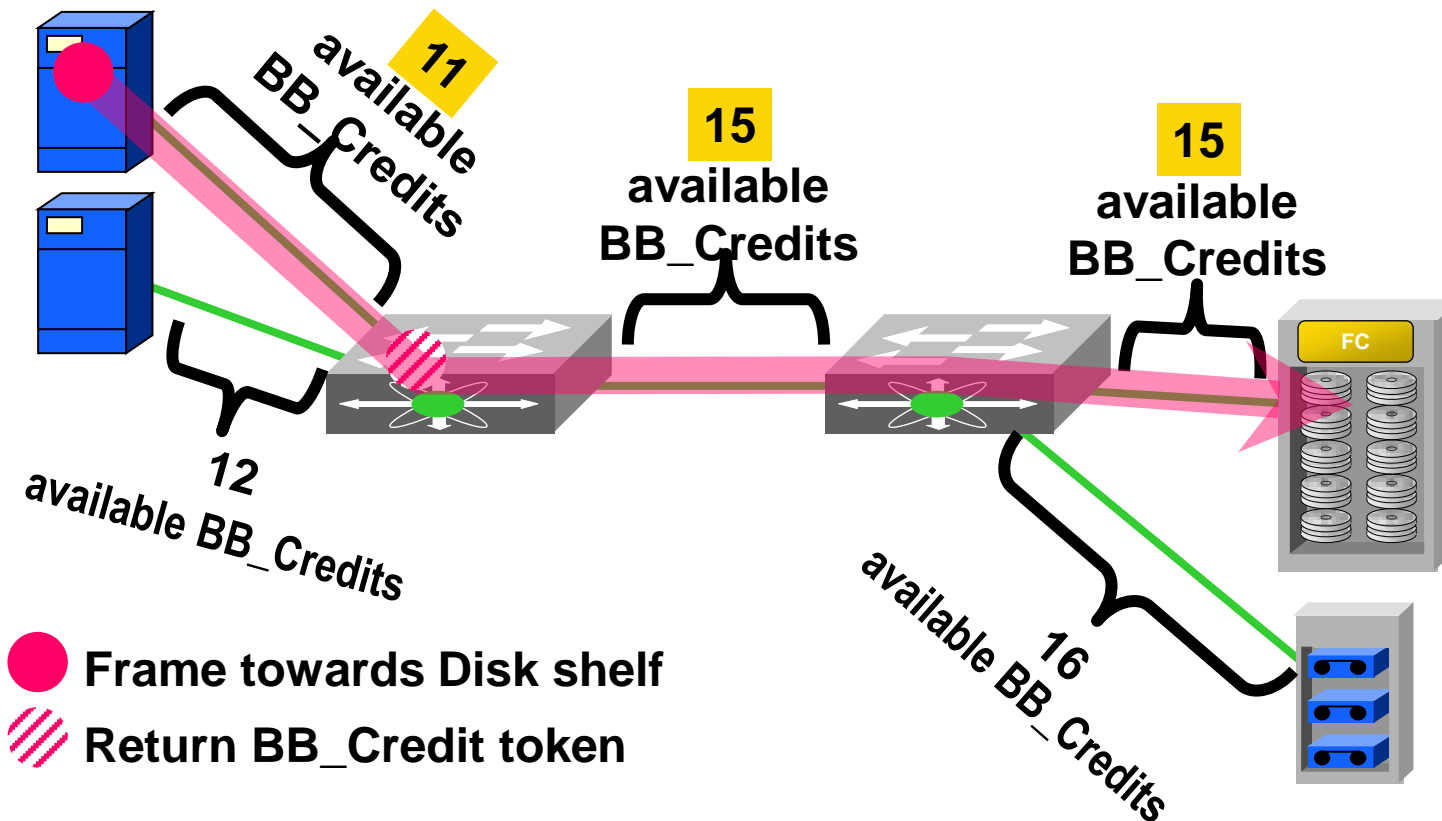
Buffer Credits (BB_Credits): Working Clean

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)

11/16
Available BB_Credits

7/16
Available BB_Credits

FC

Disk shelf capable of sustaining 200MB/sec

11/16
Available BB_Credits

Frame towards Disk shelf

Frame towards Tape

/ Return BB_Credit token

Tape capable of sustaining 15MB/sec
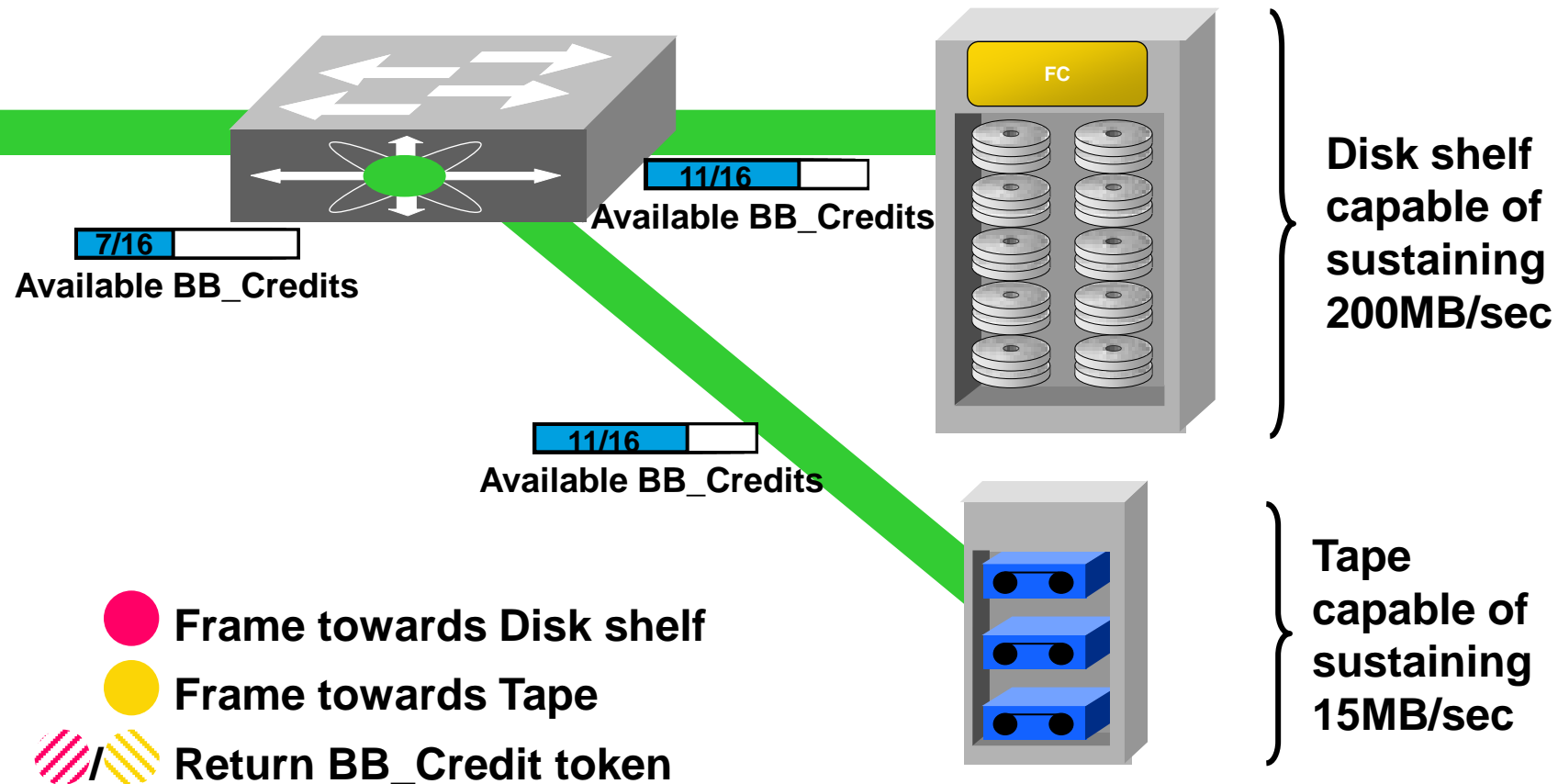
FC BB_Credits and Distance

~2 km per Frame

1 Gbps FC

32 Km

~1 km per Frame

2 Gbps FC

32 Km
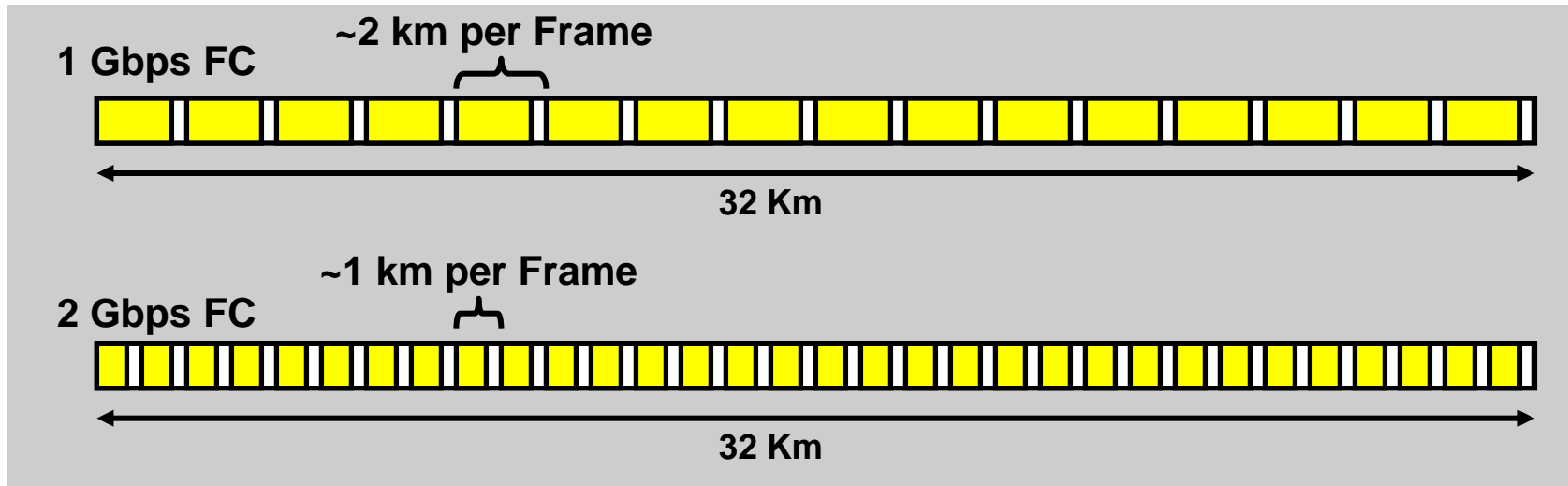
- **BB_Credits are used to ensure enough FC frames in flight**

- **A full (2112 byte) FC frame is approximately 2 km long @ 1 Gbps and approximately 1 km long @ 2 Gbps and approximately .5 km long @ 4 Gbps**

- **As distance increases, the BB_Credits need to increase as well**

- **Shorter frames will require additional BB_Credits to ensure a "full pipe" – and *USUALLY* it is smaller frames that are sent!**

- **Insufficient BB_Credits will throttle performance—no data will be transmitted until R_RDY is returned**

Number of Buffer Credits - Reality

Control `76`

Write `2048` `2048` `72`

Status Accept `68`

- Simple 4K write
- Will not fit into 2 buffers because of headers for FC as well as SB3

Average = (76+2048+2048+72+68) / 5 = 862 Bytes

Buffer Credits Required
By Size of Frame and Link Speed

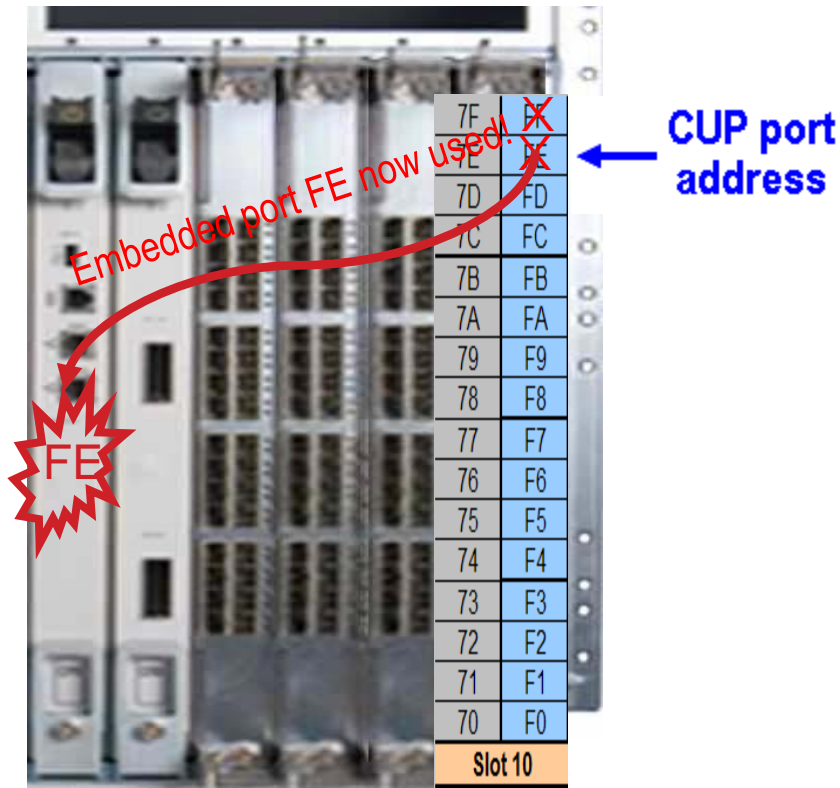| SOF, Header, CRC, EOF	Payload	Total Frame Bytes	Smaller than full frame by x%	2Gbps Buffer Credits Required 8b10b	4Gbps Buffer Credits Required 8b10b	8Gbps Buffer Credits Required 8b10b	10Gbps Buffer Credits Required 64b66b
A distance of 20km with the link 100% utilized							
36	2112	2148	0.000%	20	40	80	117
36	2002	2038	5.138%	21	42	84	124
36	1902	1938	9.809%	22	44	88	130
36	1802	1838	14.481%	24	47	93	137
36	1702	1738	19.152%	25	49	98	145
36	1602	1638	23.823%	26	52	104	154
36	1502	1538	28.494%	28	56	111	164
36	1402	1438	33.165%	30	60	119	175
36	1302	1338	37.836%	32	64	128	188
36	1202	1238	42.507%	35	69	138	203
36	1102	1138	47.179%	38	75	150	221
36	1002	1038	51.850%	41	82	164	243
36	902	938	56.521%	46	91	182	268
36	819	855	60.398%	50	100	199	294
36	700	736	65.957%	58	116	232	342
36	600	636	70.628%	67	134	268	396
36	500	536	75.299%	80	159	318	469
36	400	436	79.970%	98	195	390	577
36	300	336	84.641%	127	254	507	748
36	200	236	89.312%	181	361	721	1065
36	100	136	93.984%	313	626	1251	1848
36	75	111	95.151%	383	766	1532	2264
36	50	86	96.319%	495	989	1978	2922

As distance link speed grows, so does the need for buffer credits!

Mainframe Terminology

CUP should be deployed on every FICON Director! It allows you to provide the M/F environment with port performance information and buffer credit flow control information.
2 ports could be lost when CUP is used!

Control Unit Port (CUP)

- Customers install FICON Management Server (FMS) as a license on their FICON director.

- FMS enables CUP to be utilized for in-band management and FICON Director performance reporting up to the mainframe environment.

- Port address 0xFE is always the port address exclusively defined for the CUP – but it can also be a physical port which cannot be used for connectivity when CUP is enabled.

FICON Director Activity Report (RMF 74-7)

(requires CUP Code)

```
                                    F I C O N    D I R E C T O R    A C T I V I T Y

        z/OS V1R4                 SYSTEM ID KS01              START 10/03/2005-13.55.00  INTERVAL 000.05.00
                                  RPT VERSION V1R2 RMF        END   10/03/2005-14.00.00  CYCLE 1.000 SECONDS
   IODF = 0C    CR-DATE: 07/07/2005   CR-TIME: 18.47.41   ACT: POR
                                                                MAN: MCD    PLANT: 01    SERIAL: 0000013A2:
                            AVG FRAME                           DTH (MB/SEC)        ERROR
             UNIT      ID    PACING       READ                  -- WRITE --         COUNT
             CHP-H     47       0          433                        1.52            0
   05        CU       ----      0          830        1199            5.81            0
   06        CU       ----      0          140         69             0.00            0
   07        CU      C052       0          591         85             0.00            0
             CU      C050                                             1.55            0
   08        CHP-H     45       0          400        577             1.00            0
             CU       ----      0         1355        32              0.00            0
             ------   ----      0            0                        0.00            0
                                0                                     1.56            0
                                                                      1.41            0
                                                                      0.00            0
   0F        CU      C053       0         1773         78             0.08     0.00    0
             CU      C051
   10        CHP-H     46       0          366        716             1.06     2.32    0
   11        CU       ----      0         1099        393             4.62     0.97    0
   12        ------   ----      0            0          0             0.00     0.00    0
   13        ------   ----      0            0          0             0.00     0.00    0
   14        CHP-H     50       0          533        832             0.32     0.58    0
   15        CU       ----      0          868       1223             2.25     5.55    0
   16        CU       ----      0          158         72             0.00     0.00    0
   17        CU      C053       0         1761         77             0.09     0.00    0
             CU      C051
   18        CHP-H     49       0          378        745             1.04     2.33    0
   19        CU       ----      0         1118        399             4.83     0.99    0
   1A        ------   ----      0            0          0             0.00     0.00    0
   1B        ------   ----      0            0          0             0.00     0.00    0
   1C        CHP-H     51       0          737        535             0.34     0.17    0
   1D        CU       ----      0          877       1230             3.22     7.91    0
   1E        CU       ----      0            0          0             0.00     0.00    0
   1F        CU      C052       0          590         82             0.00     0.00    0
             CU      C050
   20        CHP-H     4A       0          374        756             1.04     2.40    0
   21        CU       ----      0         1472        413             3.51     0.36    0
   22        ------   ----      0            0          0             0.00     0.00    0
   23        ------   ----      0            0          0             0.00     0.00    0
```
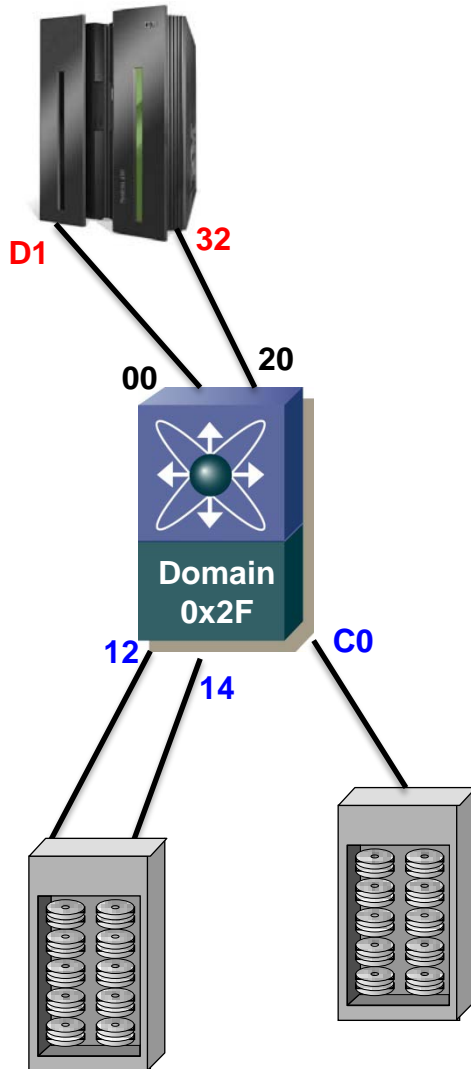
Agenda for Session 9857 and 10079

Session 10079

- Types and Components of Storage
- Let's talk Fibre Channel
- FC Buffer Credits
- **Fabric Routing / Virtual Fabrics / Partitioning**
- Security / Zoning
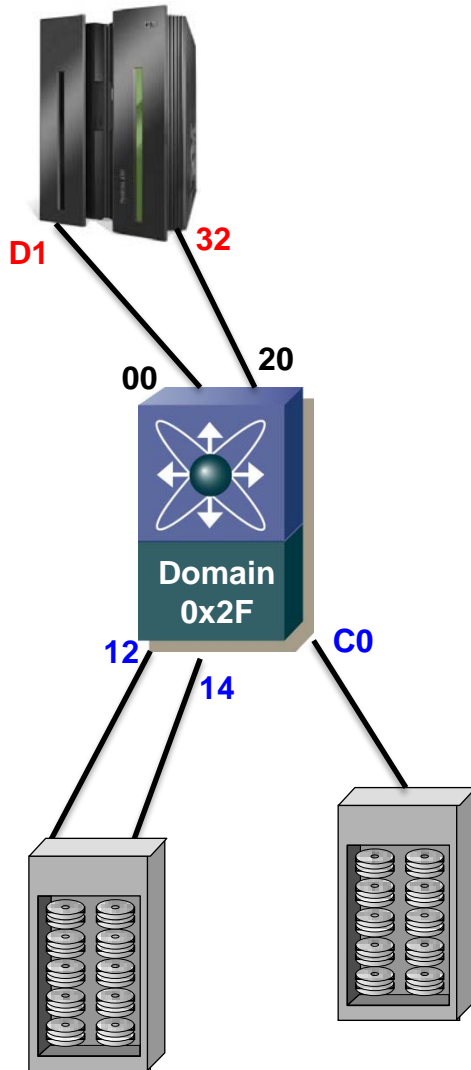
Fabric Routing Across ISLs: FSPF

- For FSPF a domain ID identifies a single switch
  - This limits the max number of switches that can be supported in the Fabric to 239 when FSPF is supported
    - Each Cisco VSAN is treated as a separate Fabric
    - Each Brocade Virtual Fabric is treated as a separate fabric
    - Each Brocade Physical Partition is treated as a separate fabric

- FSPF performs hop-by-hop routing
  - Each Cisco VSAN runs it's own FSPF process
    - Routing between VSAN's is done with Inter VSAN Routing (IVR)
  - Brocade will use FSPF routing if TRUNKING is not used
    - Old CNT/McDATA devices use FSPF for ISL routing

- FSPF supports hierarchical path selection
  - Provides the scalable routing tables in large topologies
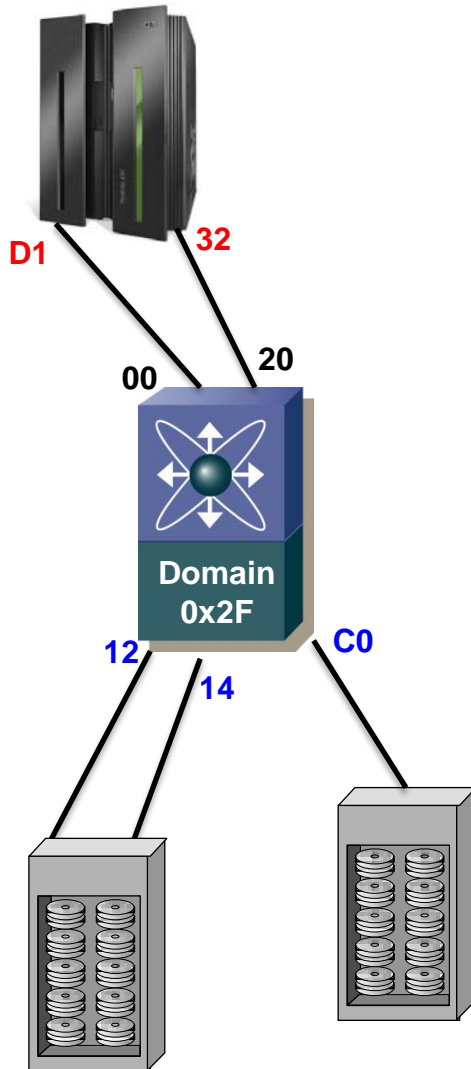
FICON Traffic Routing – Single Switch

- Mainframe uses static FCID-based routing for devices
  - CHPIDs 32 and D1 are connected to a single switch
  - Switch has statically defined domain of 0x2F
  - Host IOCDS or HCD defines route to devices
  - Specifying the CHPID(s) and corresponding output port

D1   32

00   20

Domain
0x2F

12   14   C0

FICON Traffic Routing – Single Switch

- Mainframe uses static FCID-based routing for devices
  - CHPIDs 32 and D1 are connected to a single switch
  - Switch has statically defined domain of 0x2F
  - Host IOCDS or HCD defines route to devices
  - Specifying the CHPID(s) and corresponding output port

```
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
```

D1   32

00   20

Domain
0x2F

12   C0
  14

FICON Traffic Routing – Single Switch

- Mainframe uses static FCID-based routing for devices
    - CHPIDs 32 and D1 are connected to a single switch
    - Switch has statically defined domain of 0x2F
    - Host IOCDS or HCD defines route to devices
        - Specifying the CHPID(s) and corresponding output port

```
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),12,14)),                                         *
      UNIT=2105,CUADD=B,UNITADD=((00,032))

CNTLUNIT CUNUMBR=0C00,PATH=((CSS(0),D1)),                            *
      LINK=((CSS(0),C0)),                                            *
      UNIT=2105,CUADD=C,UNITADD=((00,032))
```
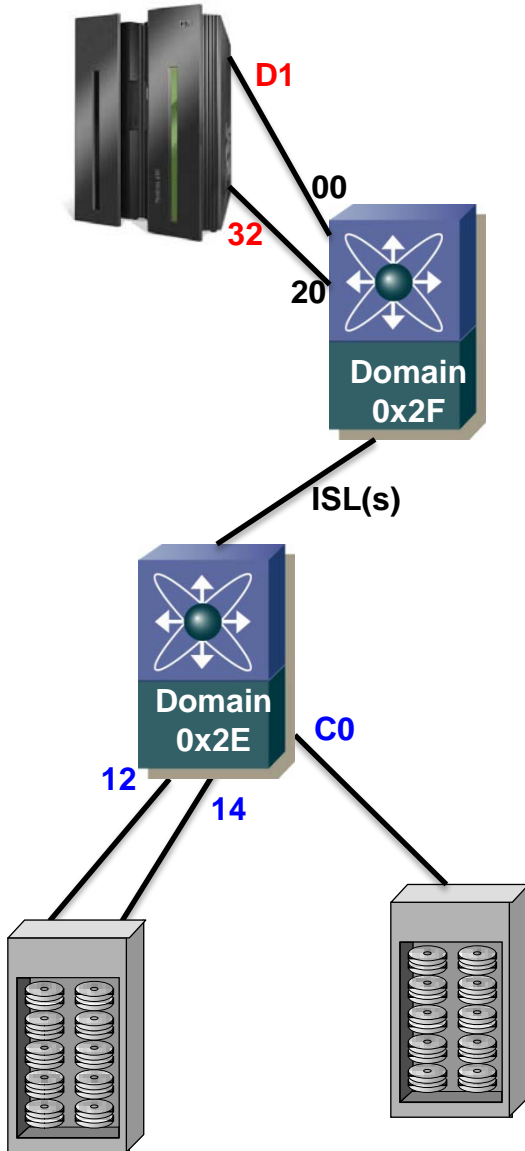
D1 32

00 20

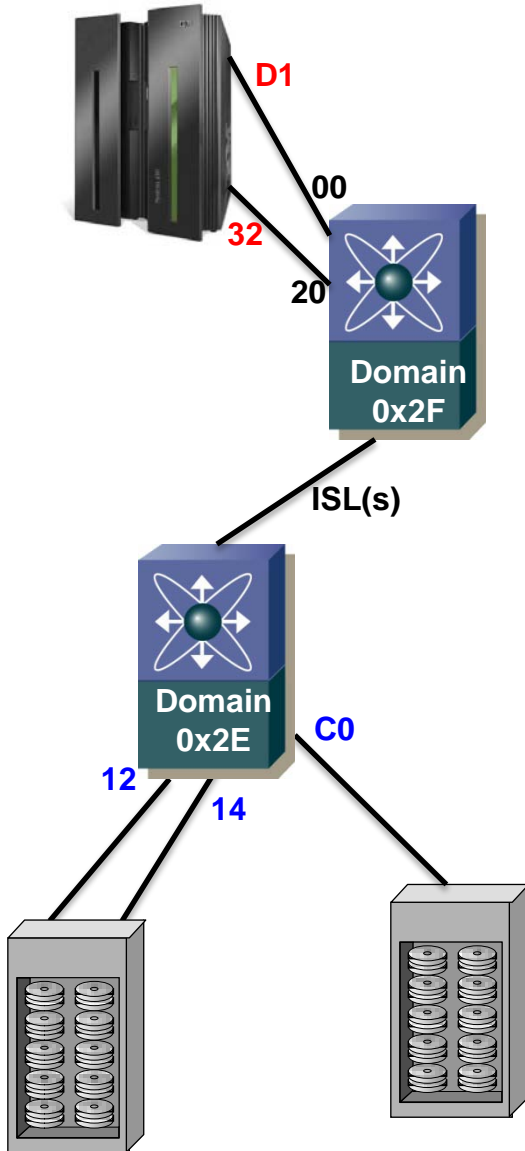**Domain 0x2F**

12 14 C0

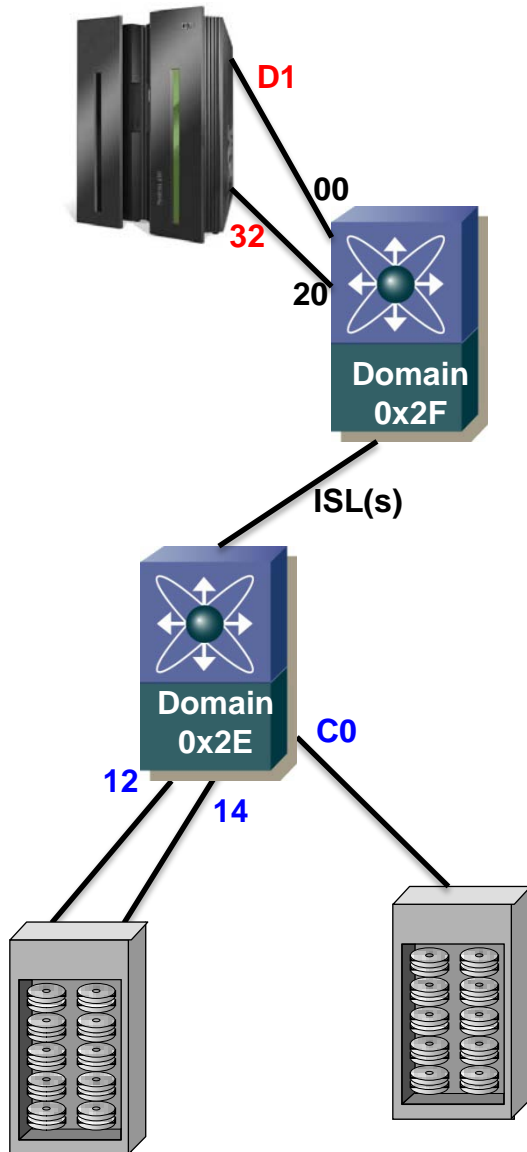# FICON Traffic Routing – Cascade

- Mainframe two switch routing
  - CHPIDs 32 and D1 are connected to a host-side switch (0x2F)
  - DASD is attached to a different switch (0x2E)
  - Host IOCDS or HCD defines route to devices
  - - Specifying the CHPID(s) and corresponding output domain/port
  - Mainframe is "blind" to the the ISL

FICON Traffic Routing – Cascade

D1

00

32

20

**Domain
0x2F**

ISL(s)

**Domain
0x2E**

C0

12

14

- Mainframe two switch routing
  - CHPIDs 32 and D1 are connected to a host-side switch (0x2F)
  - DASD is attached to a different switch (0x2E)
  - Host IOCDS or HCD defines route to devices
  - - Specifying the CHPID(s) and corresponding output domain/port
  - Mainframe is "blind" to the the ISL

```
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
```

FICON Traffic Routing – Cascade

D1

00

32

20

**Domain
0x2F**

ISL(s)

**Domain
0x2E**

C0

12

14

- Mainframe two switch routing
  - CHPIDs 32 and D1 are connected to a host-side switch (0x2F)
  - DASD is attached to a different switch (0x2E)
  - Host IOCDS or HCD defines route to devices
  - - Specifying the CHPID(s) and corresponding output domain/port
  - Mainframe is "blind" to the the ISL

```
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))
```

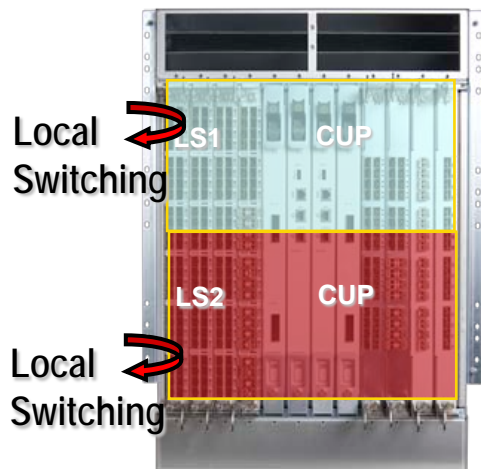# Virtual Fabrics / Virtual SANs

- Multiple Virtual SANs in a single hardware entity
  - Used to Converge SAN Islands
  - IETF RFC 4747
  - Although it is a Standard – Implementations are different per Vendor

Virtual Fabrics (VF) for DCX Family
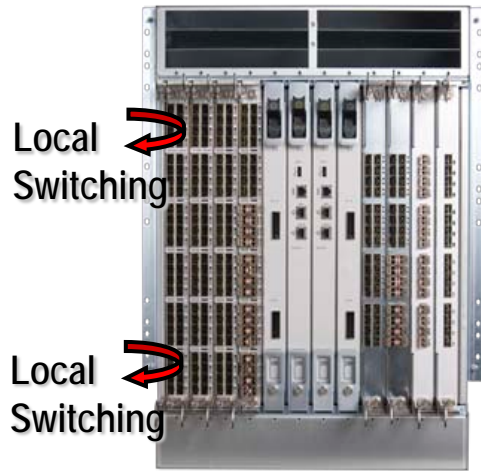
We expect most customers to deploy FICON on a non-virtualized chassis

Local Switching

Local Switching

An infrastructure that is supporting intermixed protocol use (FICON/FCP) or hosts different customers on the same infrastructure might desire virtualization

Local Switching

LS1   CUP

LS2   CUP

Local Switching

CUP   CUP

LS1   LS2   LS3   LS4

- VF is an OPTIONAL feature on the Brocade DCX family of FICON/SAN Directors

- But DCX 48-port blades for FICON can be used only when Virtual Fabrics are enabled on a DCX
  - DCX-4S never requires VF

- Creating a Logical Switch, which enables Virtual Fabric, requires FOS 6.2+

- VFs create new services within each Virtual Fabric
  - Domain ID
  - Name Server, etc

- Provides additional isolation of ports within a chassis

Cisco's VSANs (Virtual SAN)

A way to Partition a Switch or SAN into a Virtual/Logical environment

- Virtual SANs created from larger cost-effective redundant physical fabric

- Reduces wasted ports of the older "island" approach

- Hardware-based isolation

- Statistics can be gathered per VSAN

- Management per VSAN

- Unique Serial Number / CUP per FICON VSAN
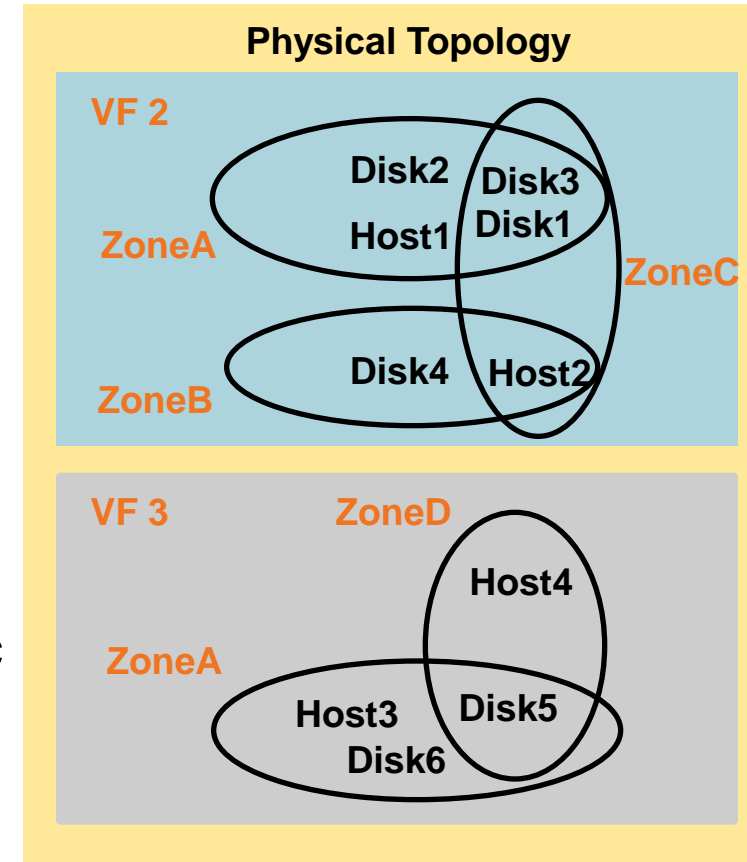
- Service and process level Isolation between VSANs
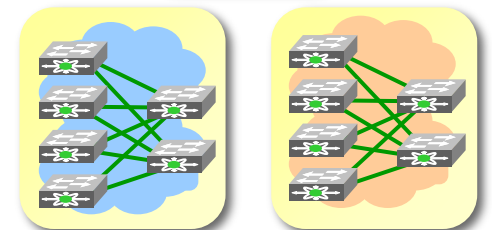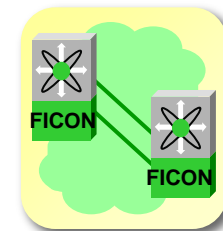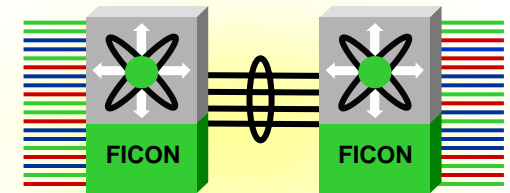
Agenda for Session 9857 and 10079

Session 10079

- Types and Components of Storage

- Let's talk Fibre Channel

- FC Buffer Credits

- Fabric Routing / Virtual Fabrics / Partitioning

- **Security / Zoning**

Zoning

- A logical grouping of fabric connected devices within a SAN (or virtual fabric)

- Zoning establishes access control
  - Devices within a zone can access each other

- Zoning increases security
  - Limiting access prevents unauthorized access

- Zone membership might be configured by:
  - Port World Wide Name (pWWN)—device
  - Fabric World Wide Name (fWWN)—fabric
  - Fibre Channel Identifier (FCID)
  - Fibre Channel Alias (FC_Alias)
  - IP address
  - Domain ID/port number
  - Interface

Physical Topology

VF 2

ZoneA

Disk2 Disk3
       Disk1
Host1

ZoneC

ZoneB

Disk4 Host2

VF 3   ZoneD

Host4

ZoneA

Host3   Disk5
Disk6

SHARE
in Orlando
2011

FICON Port Security

- Zoning
  - Able to be used with FICON (some vendors require it, others don't)

- Single Domain
  - Prohibit / Allow Matrix (ala. ESCON like port blocking/unblocking)

- Cascaded Configurations
  - Access to Cascaded resources controlled from IOCDS
  - Can also be done with Zoning but most choose not to

You can block or prohibit ports, eg:
- 05 is blocked and prohibited on all
- 04 can reach 07 and 0C
- 07 is prohibited from 0C

SAN Sessions at SHARE this week

Monday:

Time-Session

0930 - 9857: <u>SAN (Storage Area Networking) Overview – Part 1</u>

1100 - 10079: <u>SAN (Storage Area Networking) Overview - Part 2</u>

1330 - 9671: <u>Managing SAN for Linux on z/VM - a Nationwide Perspective</u>

1500 - 9931: <u>Buffer-to-Buffer Credits, Exchanges, and Urban Legends</u>

1630 - 9934: <u>Evolution of the System z Channel</u>

Tuesday:

Time-Session

1100 - 9368: <u>A First Look at the Inner Workings and Hidden Mechanisms of FICON Performance</u>

1500 - 10079: <u>A deeper look into the Inner Workings and Hidden Mechanisms of FICON Performance</u>

Wednesday:

Time-Session

0800 - 9479: <u>Planning and Implementing NPIV for System Z</u>

0930 - 9864: <u>zSeries FICON and FCP Fabrics - Intermixing Best Practices</u>

Thursday:

Time-Session

0800 - 9853: <u>FICON Over IP - Technology and Customer Use</u>

0800 - 9899: <u>Planning for ESCON Elimination</u>

0930 - 9933: <u>Customer Deployment Examples for FICON Technologies</u>

1500 - 9316: <u>SAN Security Overview</u>

1630 - 10088: <u>FICON Director and Channel Free-for-all</u>

Questions and Discussion

THANK YOU FOR ATTENDING!

Session 10079

Please fill out your evaluation forms