#SHAREorg

SHARE
Technology · Connections · Results

A first look into the Inner Workings and Hidden Mechanisms of FICON Performance

- David Lytle, BCAF
- Brocade Communications Inc.

- Thursday February 7 2013 – 8:00am to 9:00am

- Session Number - 13010

QR Code

SHARE in San Francisco 2013

Legal Disclaimer

- All or some of the products detailed in this presentation may still be under development and certain specifications, including but not limited to, release dates, prices, and product features, may change. The products may not function as intended and a production version of the products may never be released. Even if a production version is released, it may be materially different from the pre-release version discussed in this presentation.
- NOTHING IN THIS PRESENTATION SHALL BE DEEMED TO CREATE A WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, STATUTORY OR OTHERWISE, INCLUDING BUT NOT LIMITED TO, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NONINFRINGEMENT OF THIRD-PARTY RIGHTS WITH RESPECT TO ANY PRODUCTS AND SERVICES REFERENCED HEREIN.
- Brocade, Fabric OS, File Lifecycle Manager, MyView, and StorageX are registered trademarks and the Brocade B-wing symbol, DCX, and SAN Health are trademarks of Brocade Communications Systems, Inc. or its subsidiaries, in the United States and/or in other countries. All other brands, products, or service names are or may be trademarks or service marks of, and are used to identify, products or services of their respective owners.
- There are slides in this presentation that use IBM graphics.

Notes as part of the online handouts

I have saved the PDF files for my presentations in such a way that all of the audience notes are available as you read the PDF file that you download.

If there is a little balloon icon in the upper left hand corner of the slide then take your cursor and put it over the balloon and you will see the notes that I have made concerning the slide that you are viewing.

This will usually give you more information than just what the slide contains.

I hope this helps in your educational efforts!

A first look into the Inner Workings and Hidden Mechanisms of FICON Performance

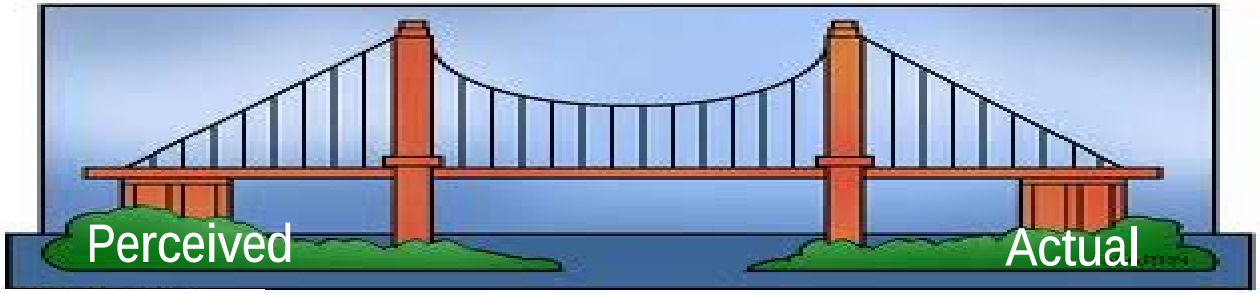
AGENDA – # 13010: 1st Look into the Inner Workings:

- Discuss some architecture and design considerations of a FICON infrastructure.

AGENDA – # 13009: A Deeper Look into the Inner Workings:

- Focused more on underlying protocol concepts:
  - FICON Link Congestion
  - How Buffer Credits are used with FICON
  - Oversubscription
  - Slow Draining devices
  - RMF reporting of Buffer Credits

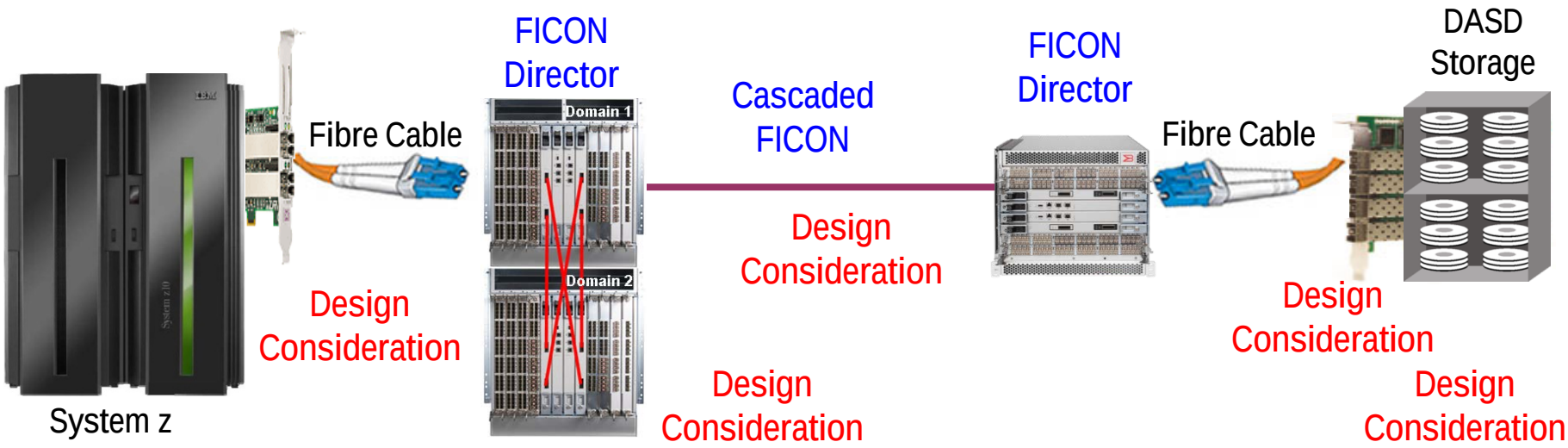
The FICON GAP



When Deploying FICON, There Is Often A Gap Between What You Expect For Its Performance And What You Actually Get!

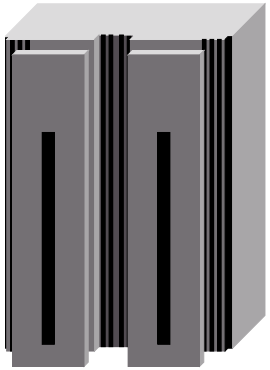
I Will Help You Bridge Some Of That Gap Here!

End-to-End FICON/FCP Connectivity



- From End-to-End in a FICON infrastructure there are a series of Design Considerations that you must understand in order to successfully satisfy your expectations with your FICON fabrics
- This short presentation is just a 50,000 foot OVERVIEW!

Mainframe Hardware and Software



System z

Evolving Interconnects and IOS Considerations

- I/O Interconnect Technologies
- Channel Path Groups

Channel Interconnect Speeds Hurt/Help Performance

EoS

PCIe
zEC12

8 GBps

PCIe
Later z196/z114

8 GBps

July 2011

InfiniBand
z10/ early z196

6 GBps

One Reason Why FICONX8 Cannot Do 8Gbps!

Jun 2010

STI
z9

2.7 GBps

Apr 2008

STI
z990/z890

2.0 GBps

Jan 2005

STI
z900/z800

1.0 GBps

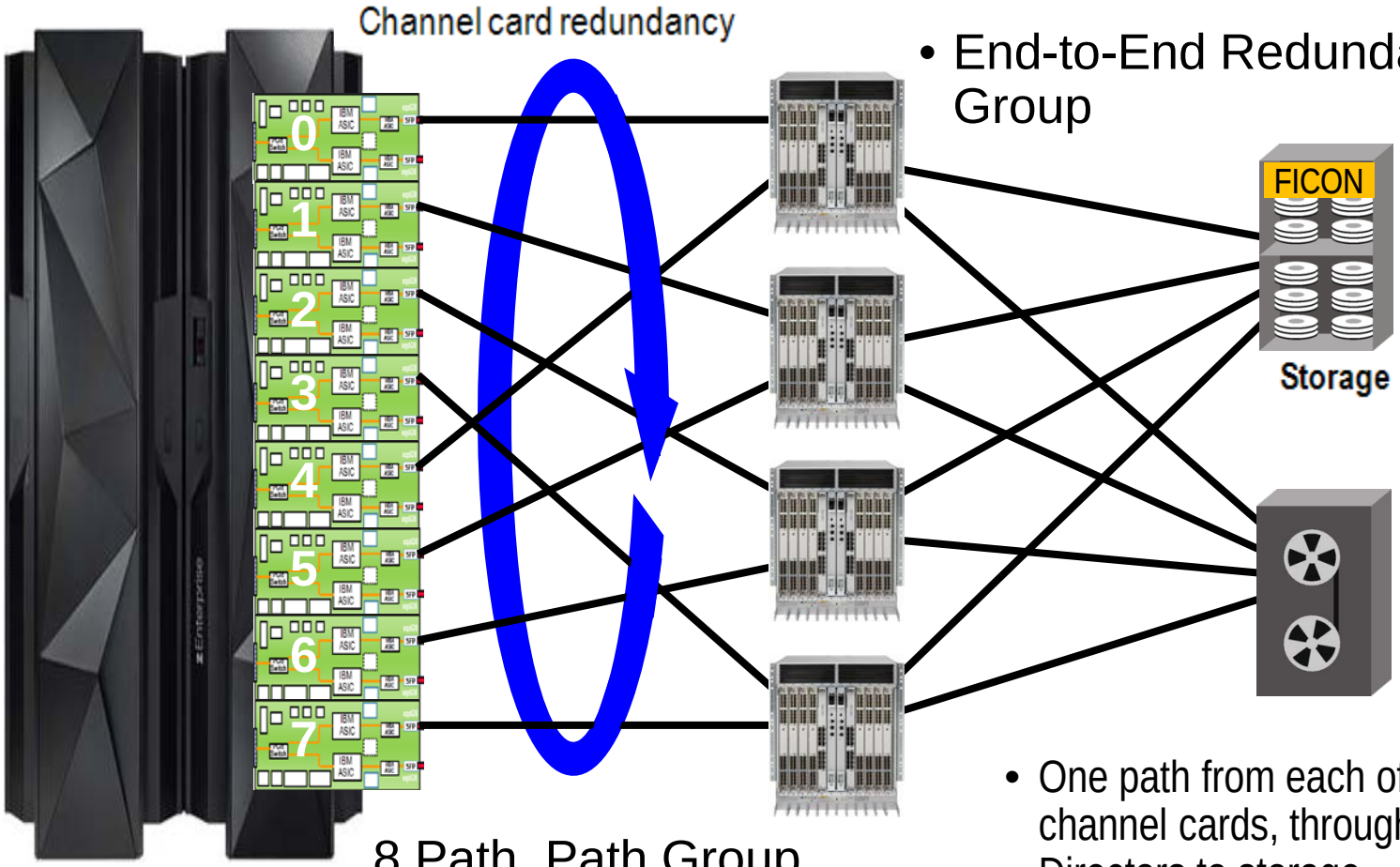
STI: Self-Timed Interconnect
PCIe: Peripheral Component Interconnect Express

I/O Channel Path Groups – since the 1980s

- The System z® operating system has a built in capability known as “Path Group” to balance and provide performance-oriented I/O.
- On the mainframe a user can group up to 8 of their physical connections between the Channel Path IDs (CHPIDs), which are the mainframe I/O ports, out to connected storage ports.
- It is the mainframe channel subsystem that decides which path in the path group will be used by deciding which path is least busy and which paths are operational, etc.
- Path Groups allow I/O to be automatically spread evenly and fairly across a number of physical channel paths without over-subscribing any given I/O path.
- Path Groups provide instantaneous fail over to operational links if a path group link fails

I/O Channel Path Group

To provide excellent I/O service time and highly available I/O activity, mainframe FICON channel Path Groups must be well architected for their role in I/O delivery



- End-to-End Redundant Path Group

- One path from each of eight Express8S channel cards, through four FICON Directors to storage.

Channel Sub-System Enhancements for zEC12 Path Group Enhancement

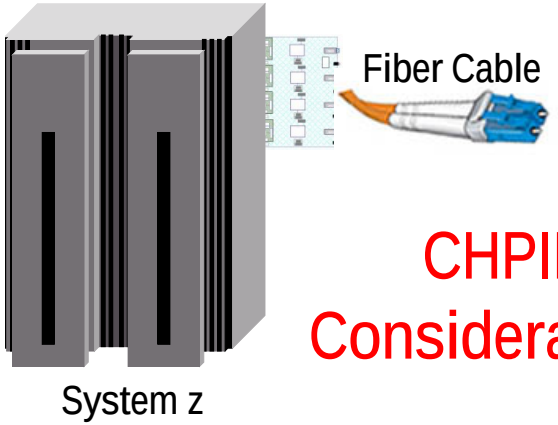
- The zEC12 channel subsystem has been enhanced with **channel path selection algorithms** designed to provide improved throughput and I/O service times when abnormal conditions occur.
- Abnormal conditions include the following:
  - Multi-system work load spikes
  - Multi-system resource contention in the I/O Fabric(s) or at the CU ports
  - I/O Fabric congestion
  - Destination port congestion
  - Firmware failures in the I/O Fabric, channel extenders, DWDMs, CUs
  - Hardware failures - link speeds did not initialize correctly
  - Mis-configuration
  - Cabling Errors
  - Dynamic changes in fabric routes

zEC12 Channel Sub-System Enhancements Path Group Enhancement

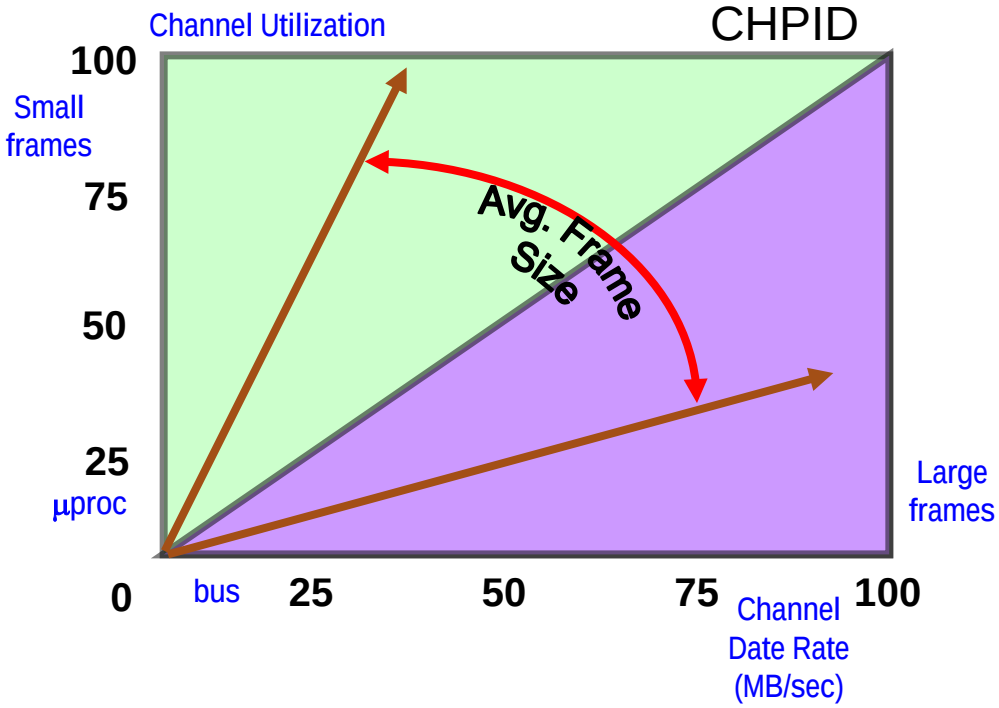
- When conditions occur that cause an imbalance in performance (e.g. I/O latency/throughput):
  - The channel subsystem will bias the path selection away from poorer performing paths toward the well performing paths.

- This is accomplished by exploiting the in-band I/O instrumentation and metrics of System z FICON and zHPF protocols and new intelligent algorithms in the channel subsystem to exploit this information.

Mainframe Channel Considerations

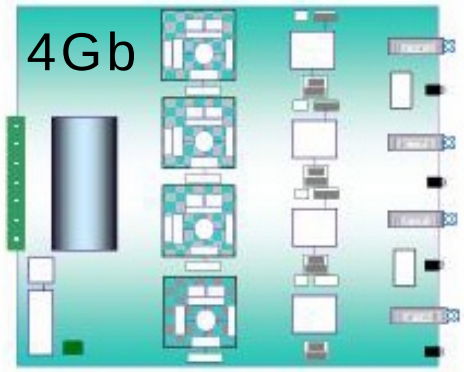


CHPID Considerations



- Channel Microprocessors and PCI Bus
- Average frame size for FICON
- Buffer Credit considerations

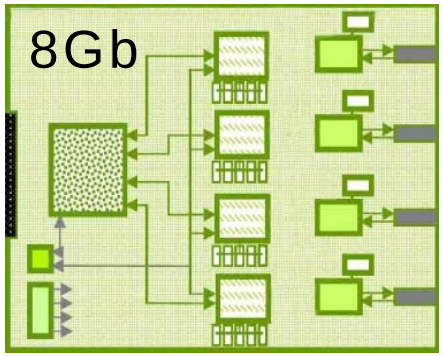
Current Mainframe Channel Cards (Features)



FICON Express4

- z196, z114, z10, z9
- 4 ports per feature
- 4km & 10km LX
- Shortwave (SX)
- 1, 2 or 4 GBps link rate
- 200 Buffer Credits/CHPID

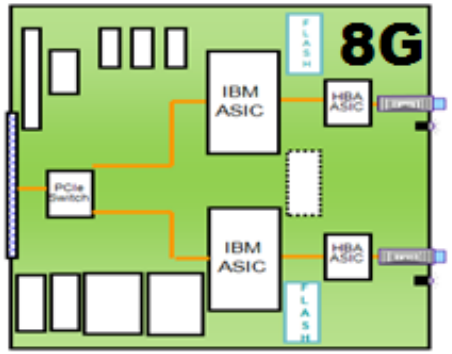
FICON Express4 provides the last native 1Gbps CHPID support



FICON Express8

- zEC12, z196, z114, z10
- 4 ports per feature
- Longwave (LX) to 10km
- Shortwave (SX)
- 2, 4 or 8 GBps link rate
- 40 Buffer Credits/CHPID

FICON buffer credits have become very limited per CHPID



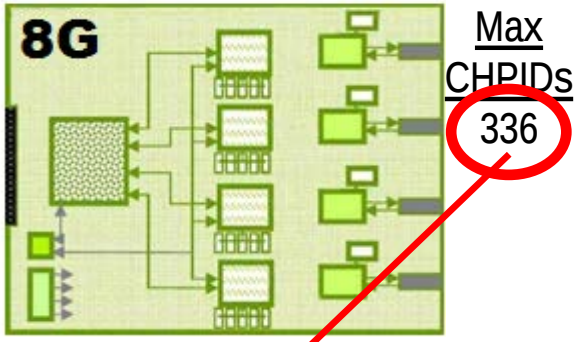
FICON Express8S

- zEC12, z196, z114
- 2 ports per feature
- Longwave (LX) to 10km
- Shortwave (SX)
- 2, 4 or 8 GBps link rate
- 40 Buffer Credits/CHPID

Reduced Ports per feature
...BUT...
Better Performance

Let's Look At This Information In More Detail.....

Mainframe Channel Cards

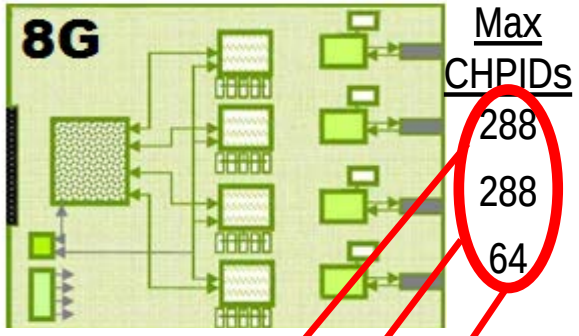


FICON Express8 – 4 ports
800MBps+800MBps=1600MBps

FICON Express8

- z10
- 2, 4 or 8 GBps link rate
- Cannot Perform at 8Gbps!
- Standard FICON Mode:
  32% <= 620 MBps Full Duplex out of 1600 MBps
- zHPF FICON Mode:
  46% <=770 MBps Full Duplex out of 1600 MBps
- 40 Buffer Credits per port
  - Out to 5km assuming 1K frames

FICON switching devices will provide BCs for long distances



FICON Express8 – 4 ports
800MBps+800MBps=1600MBps

FICON Express8

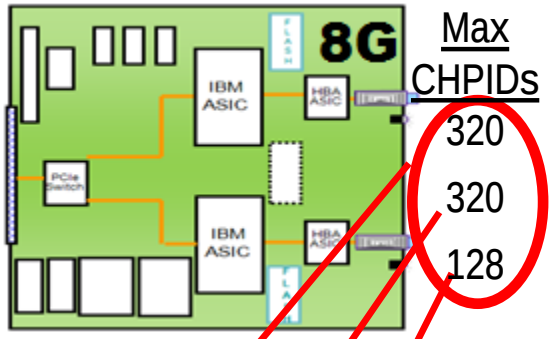
- zEC12, z196, z114
- 2, 4 or 8 GBps link rate
- Cannot Perform at 8Gbps!
- Standard FICON Mode:
  32% <= 620 MBps Full Duplex out of 1600 MBps
- zHPF FICON Mode:
  46% <=770 MBps Full Duplex out of 1600 MBps
- 40 Buffer Credits per port
  - Out to 5km assuming 1K frames

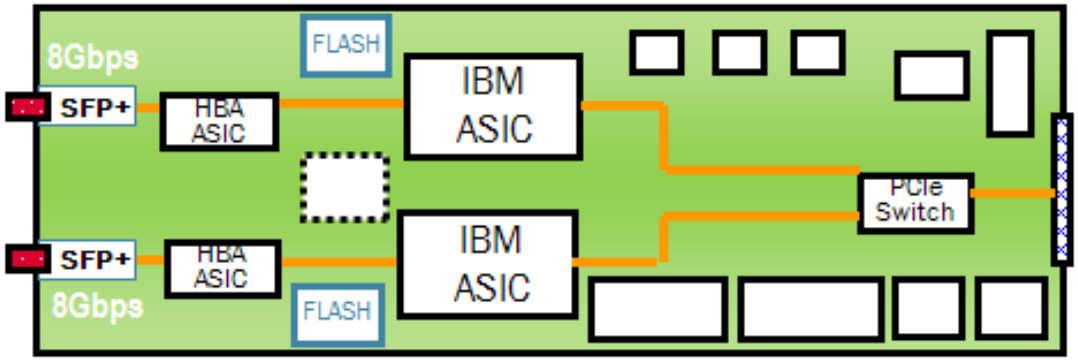
Faster Processors but fewer total CHPIDs available

One or more IBM graphics are used above

Mainframe Channel Cards

Standard, long PCIe card





FICON Express8S

- zEC12, z196, z114
- 2, 4 or 8 GBps link rate
- zHPF Performs at 8Gbps!
- Standard FICON Mode:
  32% <= 620MBps Full Duplex out of 1600 MBps
- zHPF FICON Mode:
  100% <=1600 MBps Full Duplex out of 1600 MBps
- 40 Buffer Credits per port
  - Out to 5km assuming 1K frames

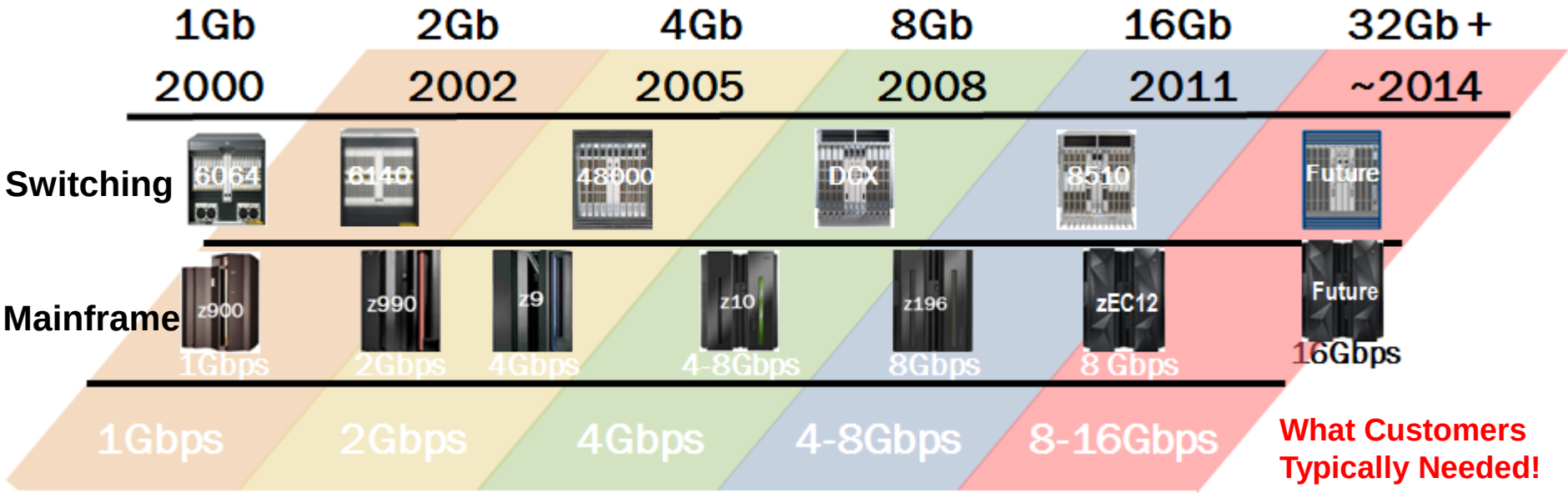
- For FICON, zHPF, and FCP environments
  - CHPID types: FC and FCP
- Auto-negotiates to 2, 4, or 8Gbps
- Increased performance versus FICON Express8
- 10KM LX - 9 micron SM fiber
  - Unrepeated distance - 10 kilometers which is 6.2 miles
  - Receiving device must also be LX
- SX - 50 or 62.5 micron multimode fiber
  - Distance variable with link data rate and fiber type
  - Receiving device must also be SX
- 2 channels of LX or SX (no mix)

Potentially – additional FICON Connectivity

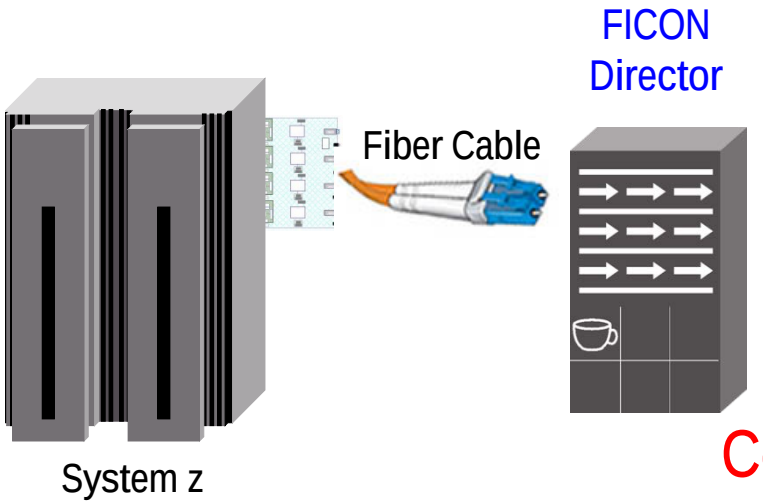
- While here at SHARE this week, one of the attendees in an earlier session that I taught, mentioned that there is a new RPQ that allows some features to be removed from the zEC12 making additional space for FICON Express8S features.
- I am trying to confirm this information right now
- If true, FICON Express8S connectivity might reach 336 connections once again.

FICON Channel Cards, Line Rate and Storage

- FICON ExpressX features typically provide less throughput capability than line rate would suggest….However,
- Customers seldom have application throughput requirements that exceed host and storage throughput rates
- Faster switching devices typically get installed before host/storage



FICON/FCP Switching Devices



FICON Director Chassis Considerations

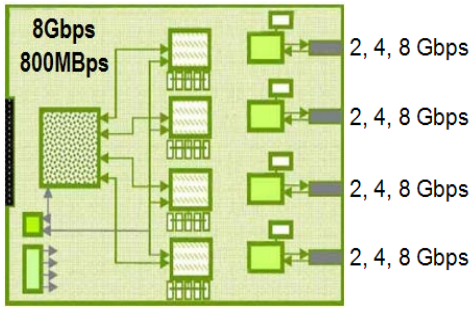
- Switched-FICON
- Direct-attached (point-to-point) versus switched FICON connectivity
- Redundant fabrics to position for five-9s of availability

Key Reasons For Using Switched-FICON

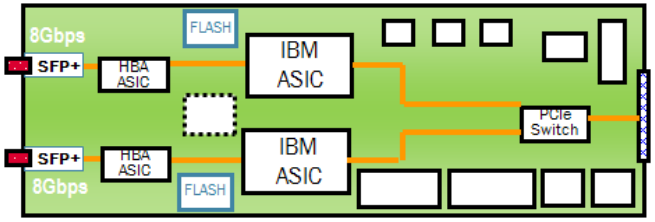
- There are 5 key technical reasons for connecting storage control units using switched-FICON fabrics:
  - Overcome buffer credit limitations on FICON 8Gbps channels.
  - Build Fan-in, Fan-out architecture designs for maximizing resource utilization.
  - Localize failures for improved availability.
  - Increase scalability and enable flexible connectivity for continued growth.
  - Leverage new FICON technologies.

FICON Connectivity from the Mainframe

FICON Express8



FICON Express8S

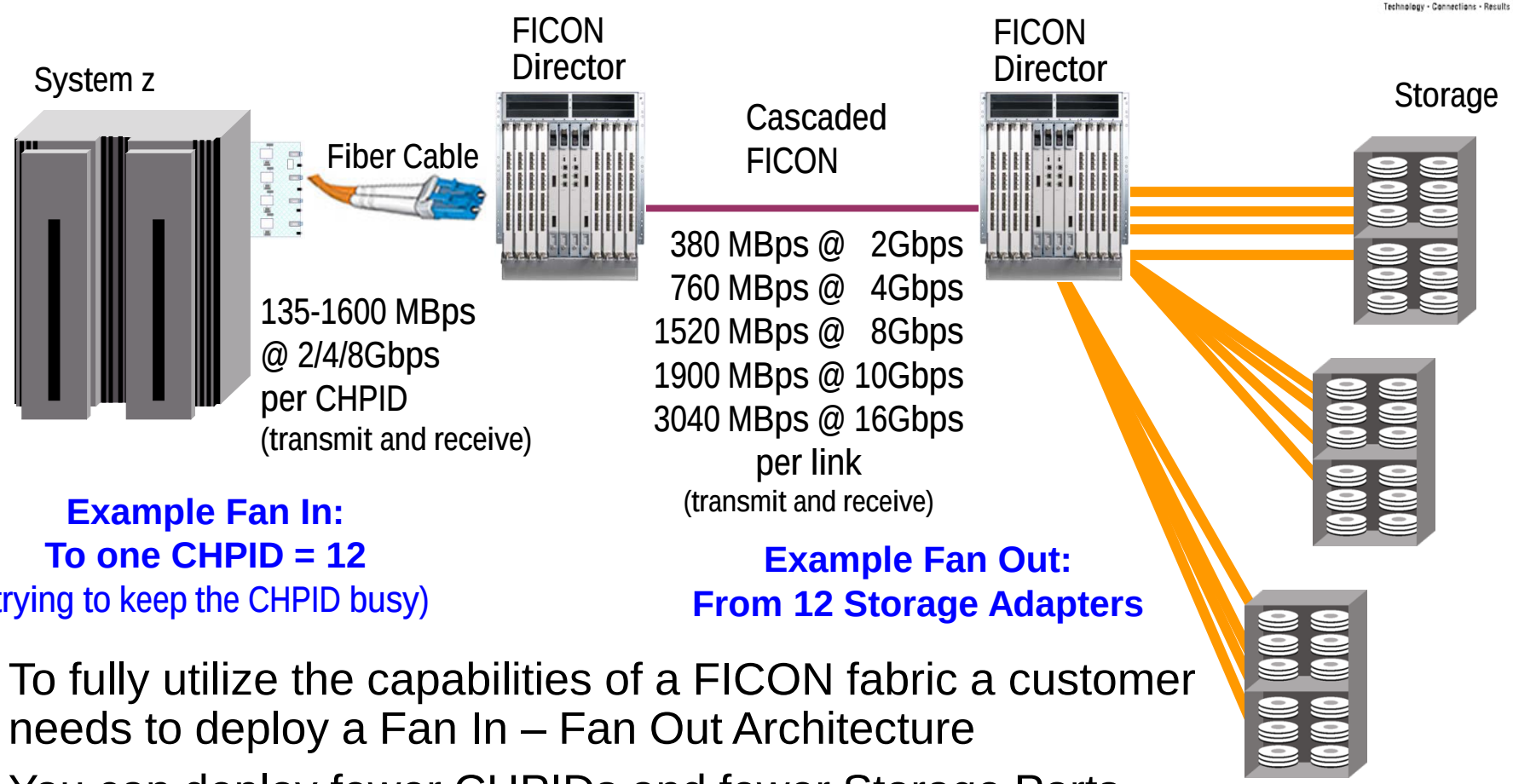


FICON 8Gbps CHPIDs

- 40 Buffer Credits per port
  - 10km @ 8G full frame / port
  - 5km @ 8G half frame / port
  - 4km @ 8G 819 byte payloads

- To the left is just an example of the limitations of buffer credits provided on mainframe CHPIDs
- FICON switching devices can provide from 2K to 4K buffer credits on a single port
- If a dark fiber connection, or any long distance connection, requires more than 5-10km of distance then switched-FICON can provide connectivity ports that can reach far further, at full path utilization, than a CHPID can provide

Fan In – Fan Out For FICON Channel Efficiency

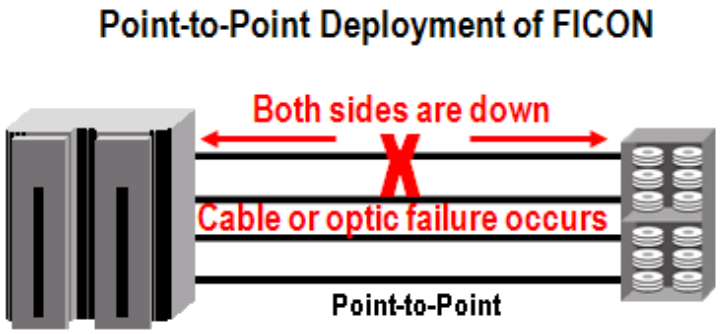


Example Fan In:
To one CHPID = 12
(trying to keep the CHPID busy)

Example Fan Out:
From 12 Storage Adapters

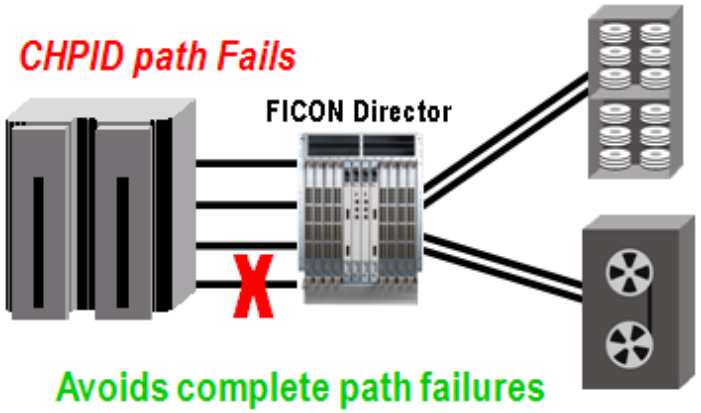
- To fully utilize the capabilities of a FICON fabric a customer needs to deploy a Fan In – Fan Out Architecture
- You can deploy fewer CHPIDs and fewer Storage Ports
- You can utilize the assets you have purchased at 80-100%
- You can scale up very easily without purchasing a lot of hardware
- You actually achieve a higher level of system availability

Availability After A Component Failure



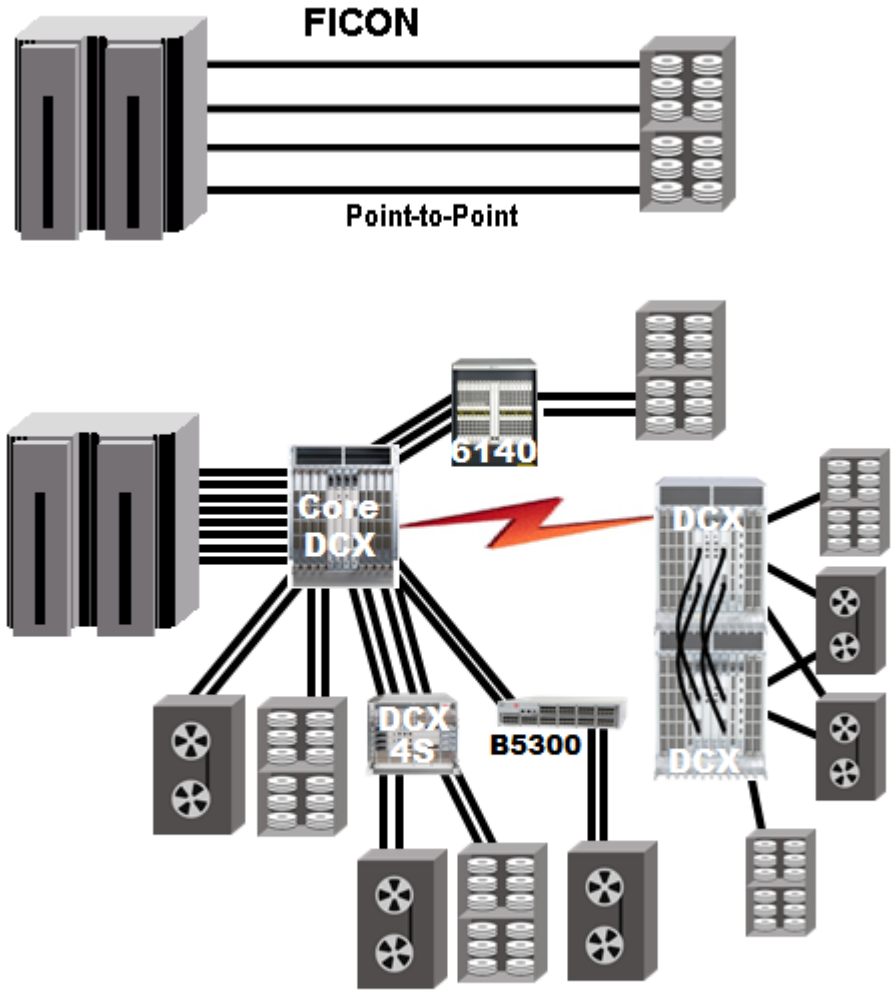
...BUT... Storage Port Remains Available!

- A failure of a FICON CHPID or cable or storage port means that you lose two valuable resources:
  - Channel port will become unavailable AND
  - Storage port becomes unavailable for everyone!
- A failure ***anywhere*** affects both the mainframe connection and the storage connection
  - The WORST possible reliability and availability is provided by a direct-attached FICON and/or SAN storage topology!



- In a switched-FICON environment, only a connection segment is rendered unavailable:
  - The non-failing side remains available
  - If the storage port has not failed, its port is still available to be used by other CHIPDs
  - If the CHPIP has not failed, its port is still available to be used by other storage ports

Robust General Scalability



FICON switching allows for dynamic connectivity in a local or remote environment

Point-to-Point does not allow for easy, dynamic growth and scalability

- 1 CHPID is tied to 1 storage port

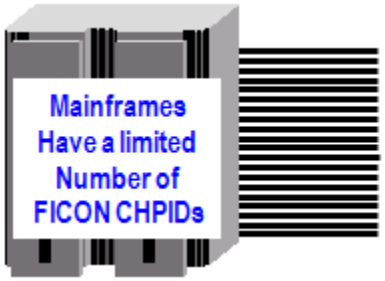
In a switched-FICON environment, you can provide dynamic connectivity

- Better use of all channel resources
- Better use of all storage resources (Fan in-Fan out)

In a switched-FICON environment, you can provide dynamic scalability if you implement FICON cascading

- Better use of all channel resources
- Better use of all storage resources
  - Fan in-Fan out
- Efficient utilization of all resources
- Quick response to new connectivity demands
- Proven Core-to-Edge connectivity
- Ability to utilize ICLs for cascading

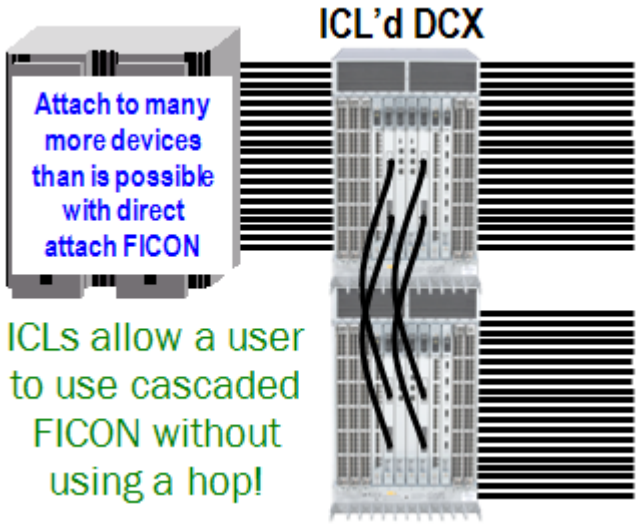
Scalability Beyond System z CHPID Limitations



Maximums

z9BC: 112 FICON Express4
z9EC: 336 FICON Express4
z10BC: 112 FICON Express8
z10EC: 336 FICON Express8
z196: 288 FICON Express8
z196: 336 FICON Express8S
z114: 64 FICON Express8
z114: 336 FICON Express8S
zEC12: 320 FICON Express8S

- Each storage port in a direct attached FICON connection will require its own, non-shared physical port connection to a CHPID
  - There is a finite limit to the number of FICON channels – depends upon the model of the mainframe that you are using
  - zEC12, using FX8S, has 5% fewer CHPIDs than z9, z10 and z196 possess
  - What happens when you need more FICON connectivity than you have CHPIDs?
- How do you continue to scale when you run out of mainframe CHPIDs?
  - When you are out of CHPIDs you have to buy a new mainframe to gain more FICON connectivity
- In a switched-FICON environment, the Fan In – Fan Out ratios solve this problem just like it solves many other connectivity and scalability problems
  - If you run out of FICON CHPIDs then simply continue to Fan-Out to more storage ports
  - Or simply use fewer FICON channel cards on your Fan-In into storage depending upon your bandwidth requirements



Besides Fan In – Fan Out, use DCX family ICL's to provide FICON Cascading to act as a CHPID multiplier for obtaining access to storage devices

Leverage New z/OS and System z Functionality

Some functionality REQUIRES customers to deploy switched–FICON:

- **FICON Dynamic Channel Management:** Ability to dynamically add and remove channel resources at Workload Manager discretion can be accomplished only in switched-FICON environments.
- **zDAC:** Simplified configuration of FICON connected disk and tape through z/OS FICON Discovery and Auto Configuration (zDAC) capability of switched-FICON fabrics.
- **NPIV:** Excellent for Linux on the Mainframe, Node_Port ID Virtualization allows many FCP I/O users to interleave their I/O across a single physical channel path

Some of my favorite photos

In Technical Sessions, Your Brain Should Be Allowed To Take A Break!

Visiting Paris

Pioneer‘s Oregon Trail

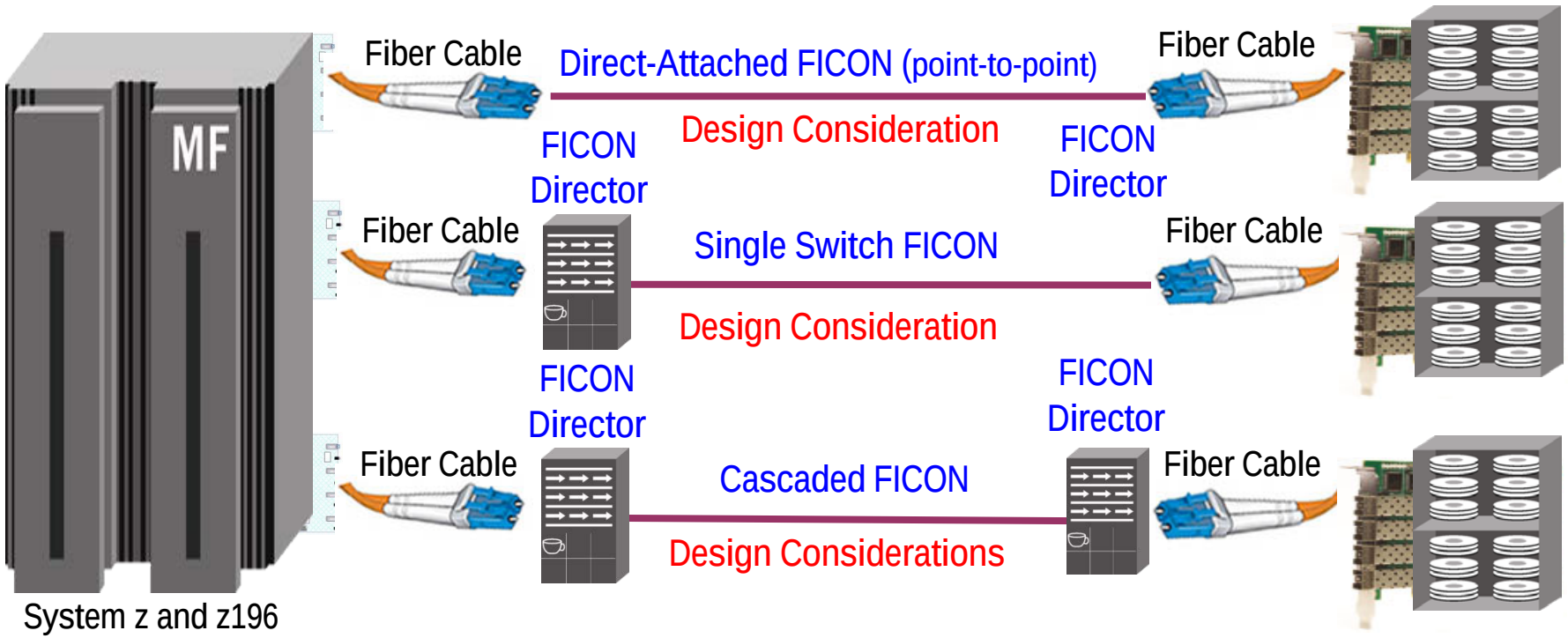
800 year old Olive Tree

Golf with Buddies

Brain Interlude Is Over….

Back to Work!

End-to-End FICON Connectivity

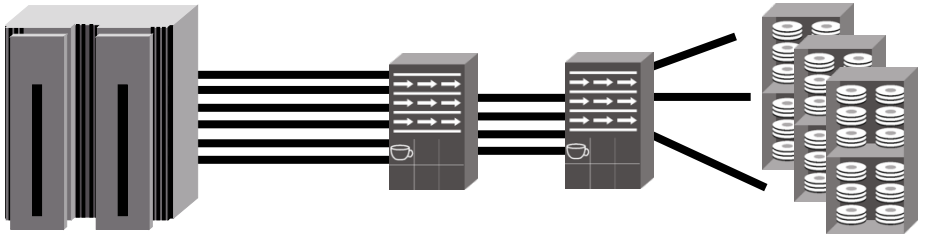


- These are the typical ways that FICON is deployed for an enterprise.
  - Switch device Long wave ports (Single Mode cables) can go to 10km or 25km (ELWL) possibly even farther
  - Switch device Short wave ports (Multimode cables) can go from 50-500 meters

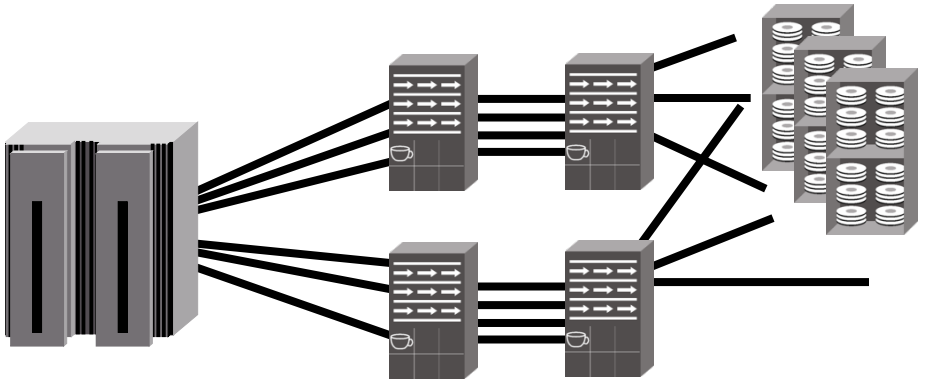
Native FICON with Simple Cascading (FC)

- Uses FICON switching devices
- Single fabrics provide no more than four-9s of availability – if a switching devices fails (a very rare occurrence) it could take down all connectivity [1]
- Redundant fabrics might provide five-9s of availability – a fabric failure would not take down all connectivity – but, loss of bandwidth is another consideration to create five-9s environments

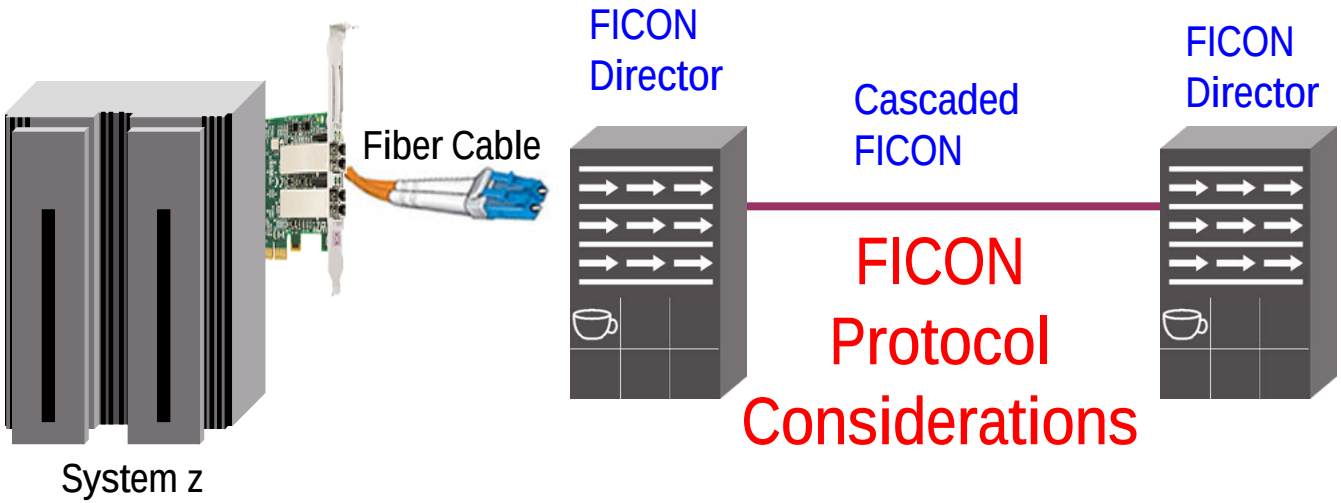
Switched-FICON and a Cascaded FICON Fabric



Redundant Switched-FICON and Cascaded FICON Fabrics



End-to-End FICON/FCP Connectivity



- **With 8b/10b, ~ 20% overhead per full frame on FICON links**
- **With 64b/66b, ~ 2% overhead per full frame on FICON links**

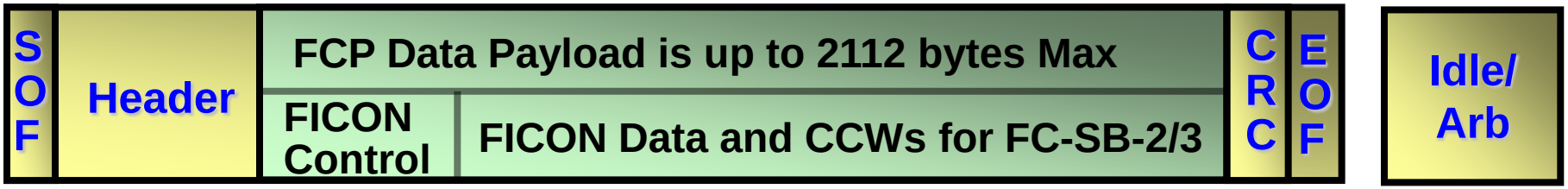
Customers can use 2/4/8/16G and/or 10G for ISL traffic today

The FICON Protocol uses 8b/10b data encoding for most link rates – but there is 20% frame payload overhead associated with it

Newer 64b/66b data encoding (10G and 16G) is also in use and is more performance oriented (only 3% data payload overhead)

MIDAW & zHPF make very good use of 8G FICON switch links

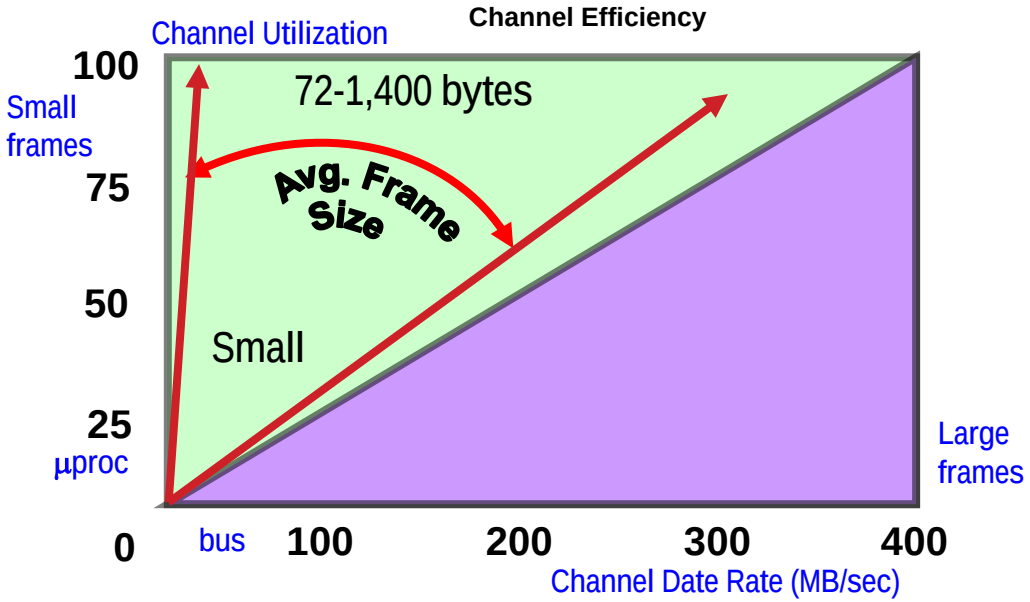
FICON FC-SB-2/3 – Channel Efficiency



64 bytes <===FICON: 2048 bytes Max =====>

<=== 2112 bytes without FICON Control ===>

<==Except for 1st frame, 2148 bytes Max out of 2148 possible==>



FC-SB-2/3

FC-SB-2/3 FICON tends to have an average frame size of between 72 and 1400 bytes

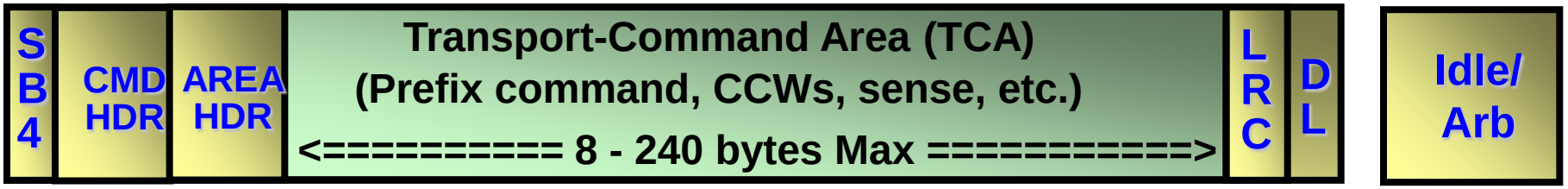
FC-SB-2/3 is used for all FICON Tape data sets and normally used for BSAM, QSAM and EXCP datasets

High Performance FICON (zHPF)

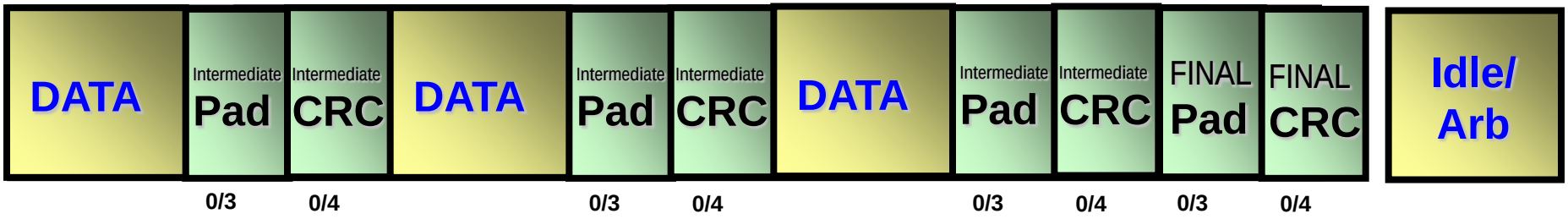
- Available since October 2008
  - zHPF is qualified on, but is not technically part of, switching
  - Partly z/OS IOS code and partly DASD control unit code
  - Available on specific IBM, HDS and EMC DASD units
- zHPF is a performance, reliability, availability and serviceability (RAS) enhancement of the z/Architecture and the FICON channel architecture
- Implemented exclusively in System z10, z196, z114 & zEC12
- Exploitation of zHPF by the FICON channel, the z/OS operating system, and the DASD control unit is designed to help reduce the FICON channel overhead
  - This is achieved through protocol simplification, CCW encapsulation within a frame, and sending fewer frames in an I/O exchange resulting in more efficient use of the channel

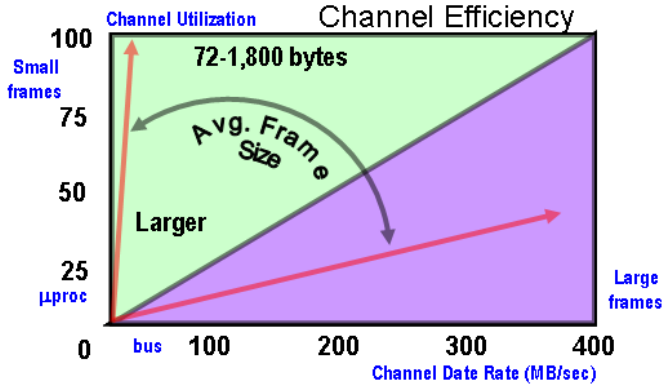
FICON FC-SB-4 zHPF – More Data, Fewer Frames

FICON Transport-Command IU for FC-SB-4



FICON Transport-Data IU for FC-SB-4 – larger average frame sizes



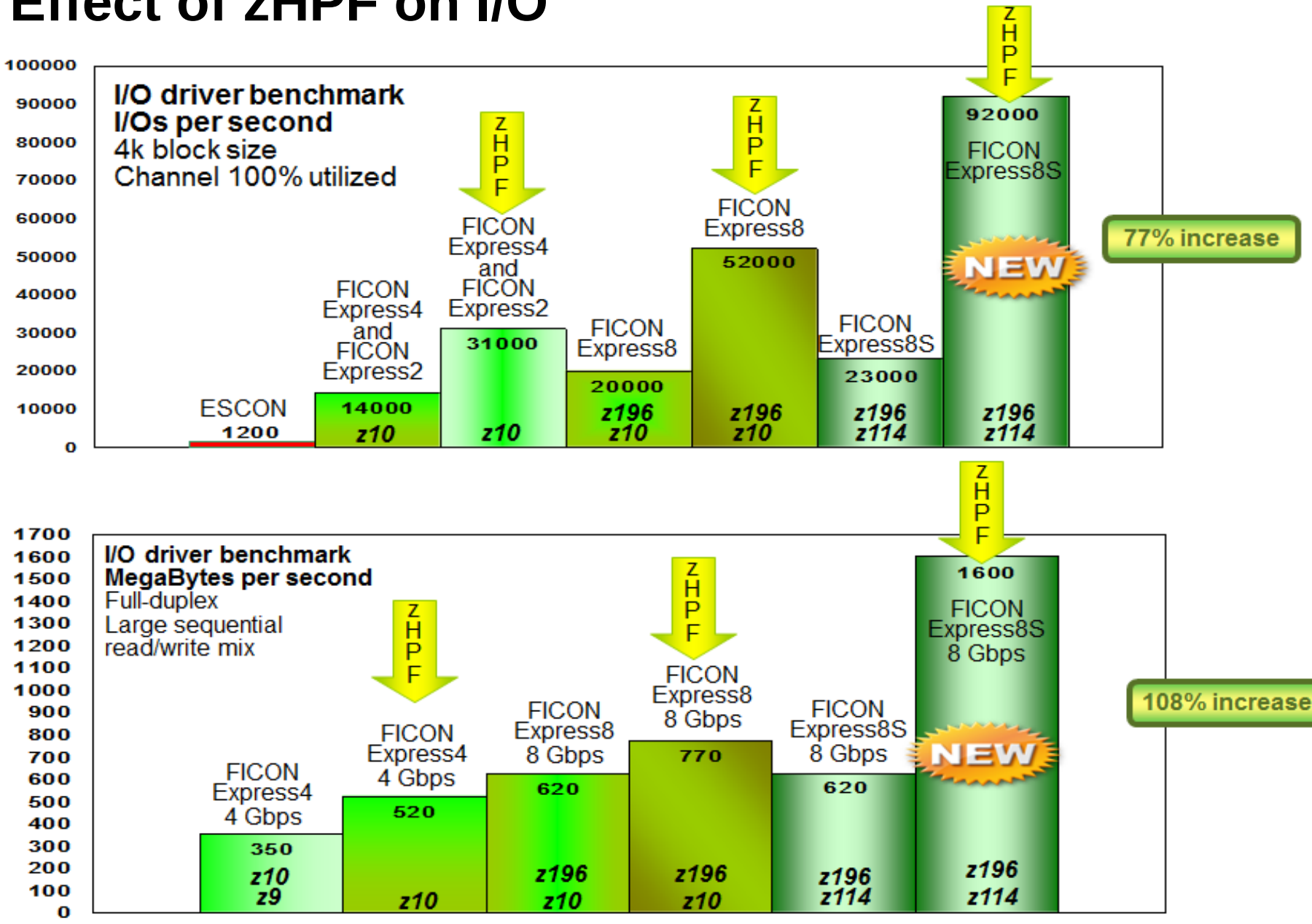


FC-SB-4

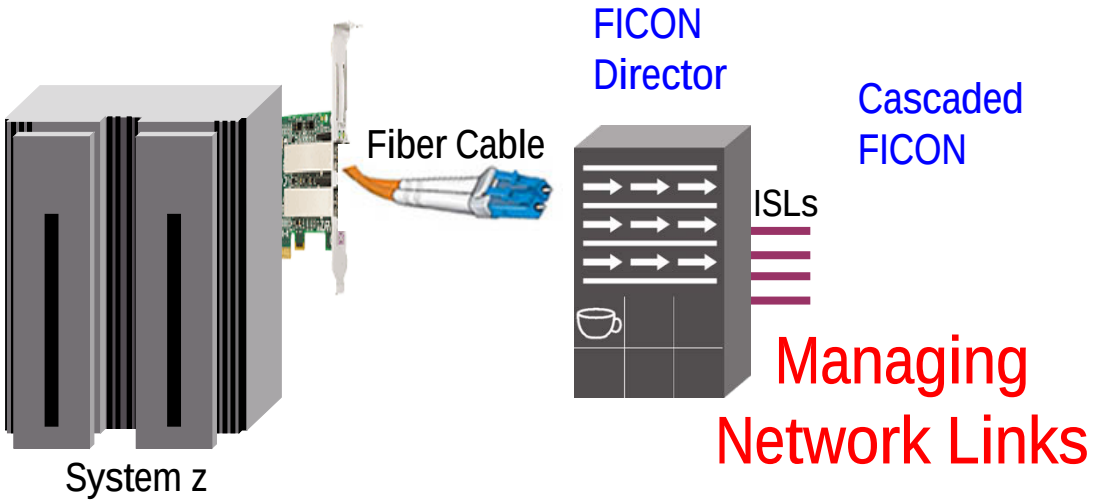
FC-SB-4 FICON tends to have an average frame size of between 72 and 1,800 bytes

FC-SB-4 can be used for FICON Media Manager Datasets like VSAM, DB2, PDSE (basically Extended Format DS) as well as BSAM, QSAM and BPAM DASD datasets

Effect of zHPF on I/O



End-to-End FICON/FCP Connectivity



Many Possible Topics

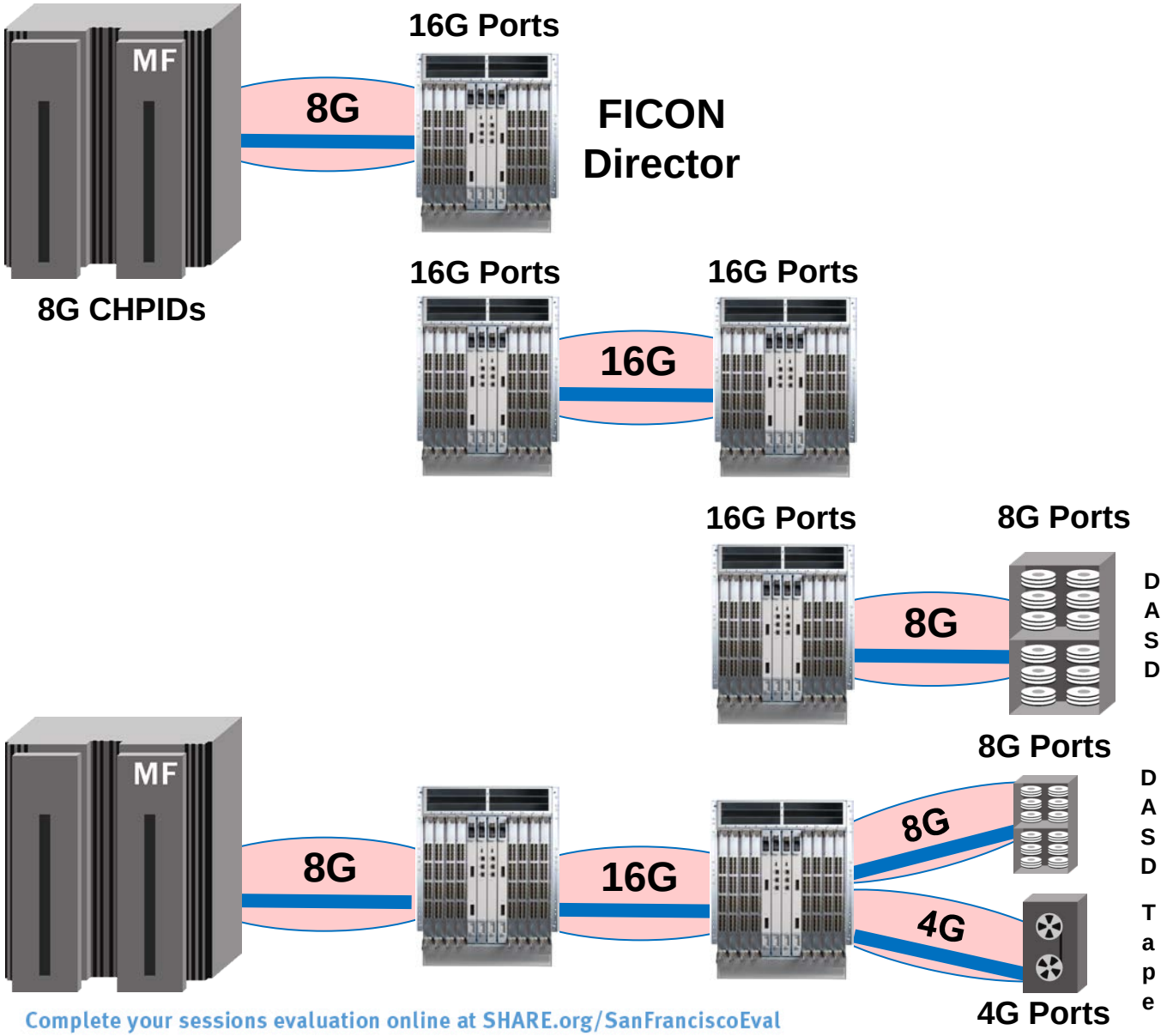
- **Fabric Link rates**
- **FICON Fabric Scalability**
- **Hops and hop issues**
- **Managing ISL Congestion**
- **Trunking**
- **Protocol Intermixed FICON Fabrics**
- **Buffer Credits**
- **Control Unit Port (CUP)**
- **Distance Extension**

Here we are at cascaded links (ISLs)

There are too many design considerations with switch-to-switch and data center-to-data center connectivity to do it all today

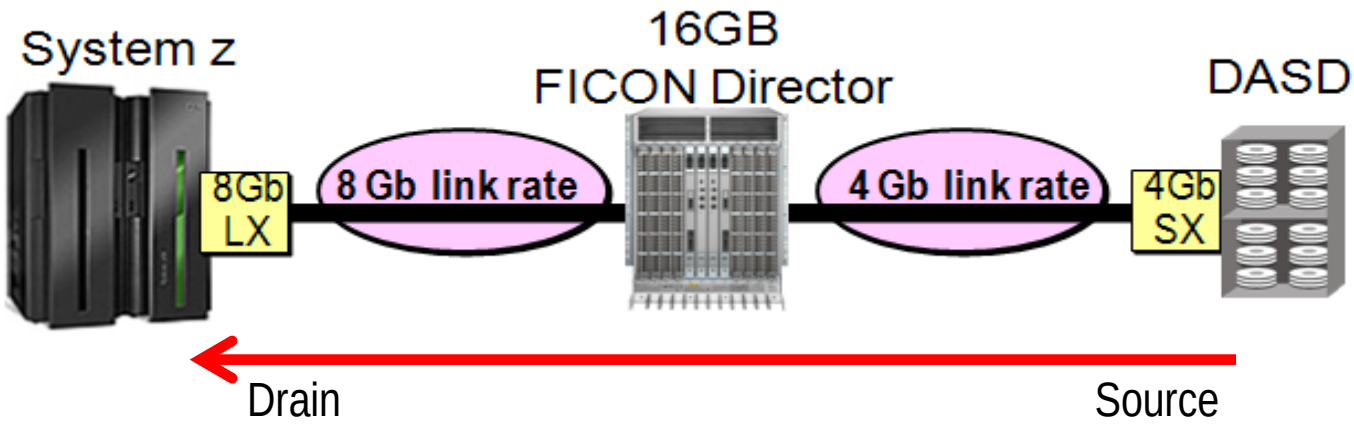
I will just spend a moment to discuss fabric link rates.

There is no such thing as End-to-End Link Rate



- Some I/O traffic will flow faster through the fabric than other I/O traffic will be capable of doing

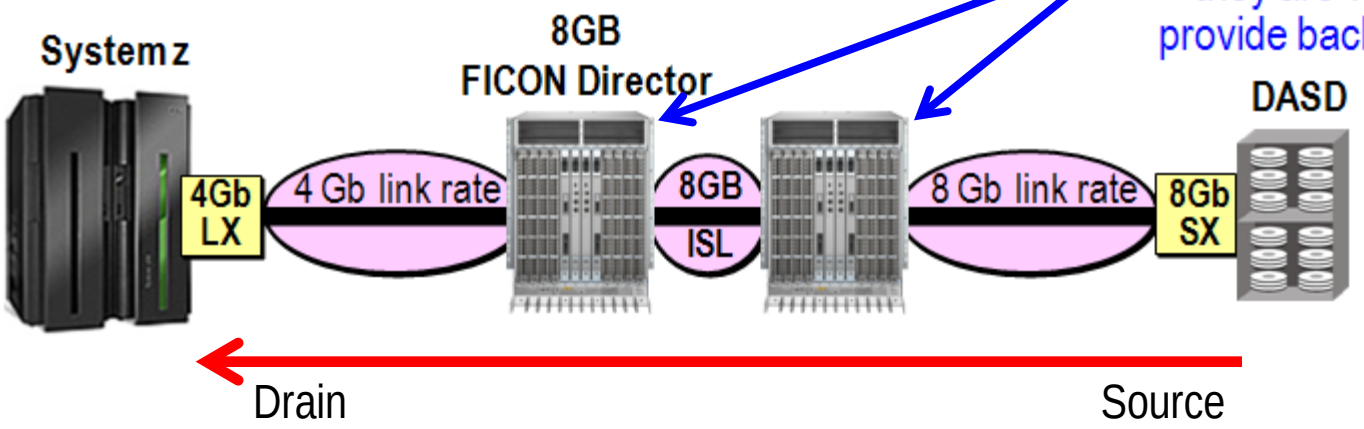
A Discussion On The Affects Of Link Rates



- Assuming no buffer credit problems, and assuming the normal and typical use of DASD, is the above a good configuration?
- If you deployed this configuration, is there a probability of performance problems and/or slow draining devices or not?

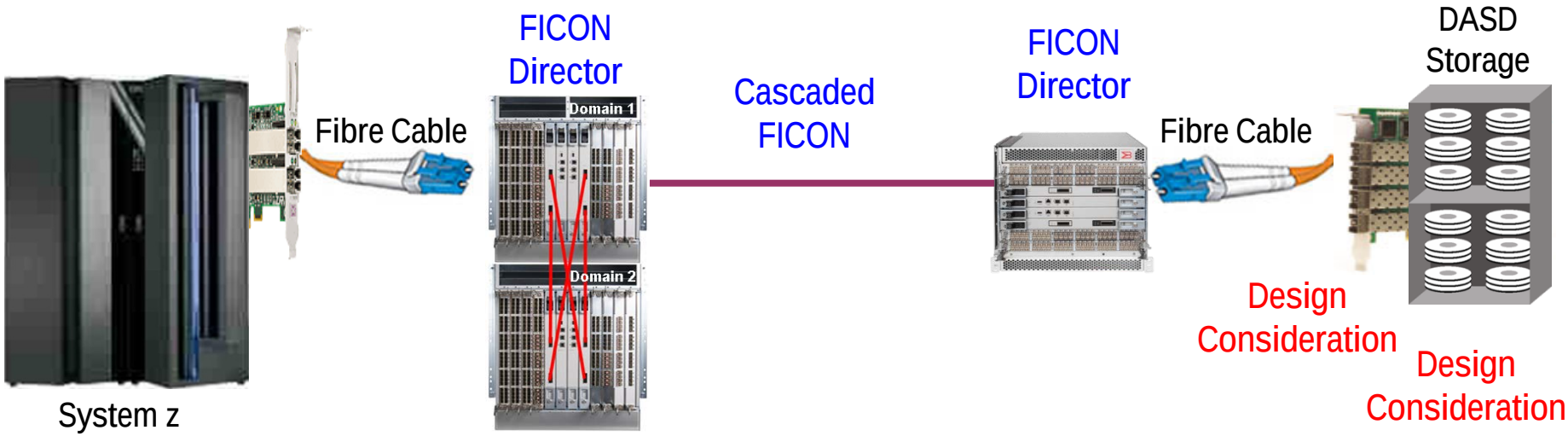
- This is actually the ideal model!
- Most DASD applications are 90% read, 10% write. So, in this case the "drain" of the pipe are the 8Gb CHPIDs and the "source" of the pipe are 4Gb storage ports.
- The 4G source (DASD in this case) cannot overrun the drain (8G CHPID)

The Affects Of Link Rates
(Adding 8G DASD with only 4G CHPIDs)



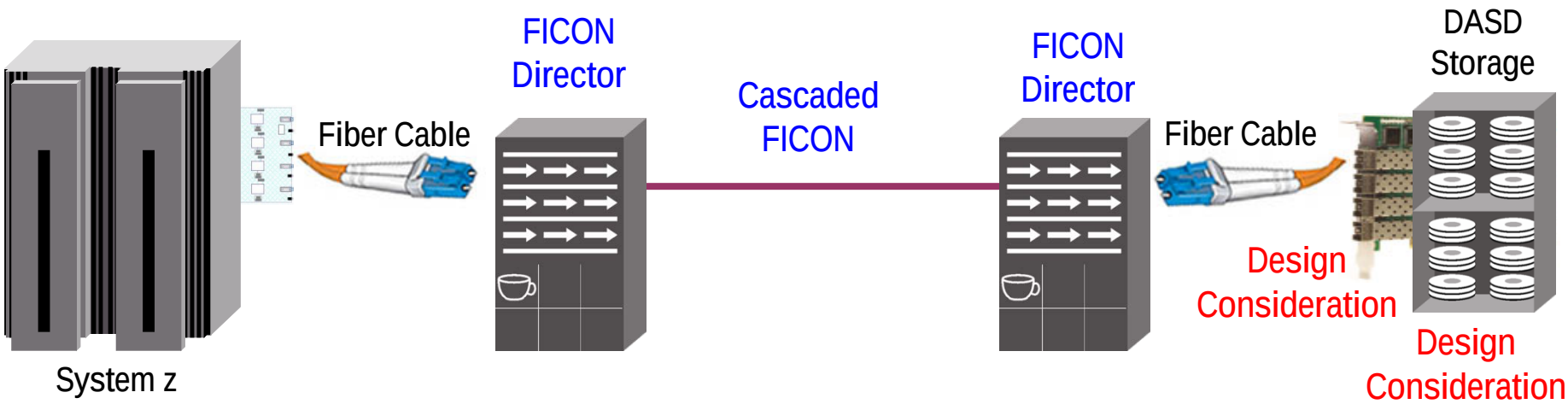
- Assuming no ISL or BC problems, and assuming the normal and typical use of DASD, is the above a good configuration?
- If you deployed this configuration, is there a probability of performance problems and/or slow draining devices or not?
- This is potentially a very poor performing, infrastructure!
- Again, DASD is about 90% read, 10% write. So, in this case the "drain" of the pipe are the 4Gb CHPIDs and the "source" of the pipe are 8Gb storage ports.
- The Source can out perform the Drain. This can cause congestion and back pressure towards the CHPID. The CHPID becomes a slow draining device.

End-to-End FICON/FCP Connectivity



- Your most challenging considerations most likely occur due to DASD storage deployment

Connectivity with storage devices



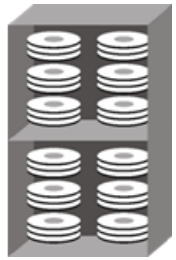
Storage adapters can be throughput constrained

- Must ask storage vendor about performance specifics
- Is zHPF supported/enabled on your DASD control units?

Busy storage arrays can equal reduced performance

- RAID used, RPMs, volume size, etc.
- Let's look a little closer at this

Connectivity with storage devices



How fast are the Storage Adapters?

- Mostly 2 / 4Gbps today – but moving to 8G – where are the internal bottlenecks?

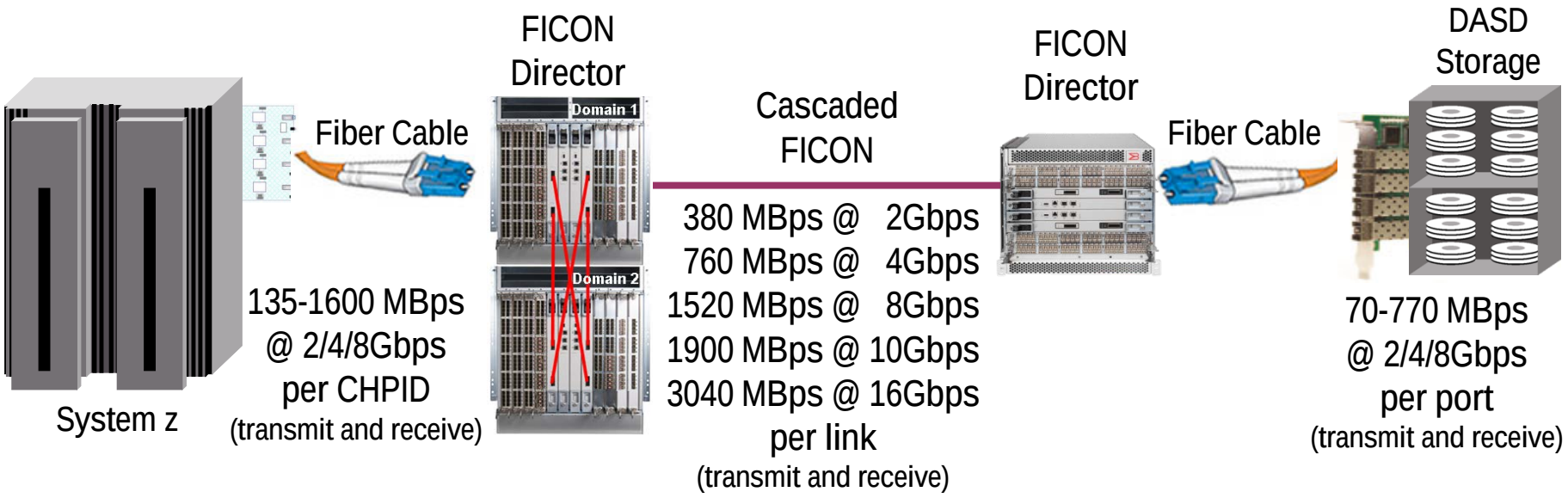
What kinds of internal bottlenecks does a DASD array have?

- 7200rpm, 10,000rpm, 15,000rpm
- What kind of volumes: 3390-3; 3390-54; EAV; XIV
- How many volumes are on a device? HiperPAV in use?
- How many HDDs in a Rank (arms to do the work)
- What Raid scheme is being used (RAID penalties)?
- Etc.

Intellimagic, Performance Associates and a host of other vendors can provide you with great tools to assist you to understand DASD performance much better

These tools perform mathematical calculations against raw RMF data to determine storage HDD utilization characteristics – use them or something like them to understand I/O metrics!

End-to-End FICON/FCP Connectivity



- In order to fully utilize the capabilities of a FICON fabric a customer needs to deploy a Fan In – Fan Out Architecture
- Fan In – Fan Out really helps to overcome many of the scalability and performance issues inherent in FICON!

Brocade Proudly Presents…
Our Industries ONLY FICON Certification

Brocade
Certified Architect
for FICON

Industry Recognized Professional Certification

We Can Schedule A Class In Your City – Just Ask!

SHARE
Technology · Connections · Results

Brocade FICON Certification

Brocade Certified Architect for FICON

Certification for Brocade Mainframe-centric Customers – Available since Sept 2008

For people who do or will work in FICON environments

Brocade provides a free on-site or in area 2-day class (Brocade Design and Implementation for FICON Environments – FCAF200), to assist customers in obtaining the knowledge to pass this certification examination – ask your local sales team about this training – also look at www.brocade.com under Education

Certification tests a person's ability to understand IBM System z I/O concepts, and demonstrate knowledge of Brocade FICON Director and switching fabric components

After the class a participant should be able to design, install, configure, maintain, manage, and troubleshoot Brocade hardware and software products for local and metro distance (100 km) environments

Check the following website for complete information:

- http://www.brocade.com/education/certification-accreditation/certified-architect-ficon/index.page

#SHAREorg

SHARE
Technology · Connections · Results

.....My Next Presentation.....

A Deeper Look into the Inner Workings and Hidden Mechanisms of FICON Performance

- David Lytle, BCAF
- Brocade Communications Inc.
- Thursday February 7, 2013 -- 9:30am to 10:30am
- Session Number - 13009

SHARE in San Francisco 2013

SAN Sessions at SHARE this week

Thursday:

Time-Session

0930 – 13009: A Deeper Look Into the Inner Workings and Hidden Mechanisms of FICON Performance
1300 – 13012: Buzz Fibrechannel - To 16G and Beyond

Mainframe Resources For You To Use

Visit Brocade's Mainframe Blog Page at:

http://community.brocade.com/community/brocadeblogs/mainframe

Also Visit Brocade's New Mainframe Communities Page a

http://community.brocade.com/community/forums/products_and_solutions/mainframe_solutions

You can also find us on Facebook at:

https://www.facebook.com/groups/330901833600458/

#SHAREorg

SHARE
Technology · Connections · Results

Please Fill Out Your Evaluation Forms!!

This was session:

13010

Thank You For Attending Today!

5 = "Aw shucks. Thanks!"
4 = "Mighty kind of you!"
3 = "Glad you enjoyed this!"
2 = "A Few Good Nuggets!"
1 = "You Got a nice nap!"

And Please Indicate On Those Forms If There Are Other Presentations You Would Like To See In This Track At SHARE.

QR Code

SHARE in San Francisco 2013