#SHAREorg

SHARE

A First Look Into The Inner Workings and Hidden Mechanisms of FICON Performance

- David Lytle, BCAF
- Brocade Communications Inc.

- Thursday August 15, 2013 – 8:00am to 9:00am

- Session Number - 14269

QR Code

SHARE in Boston

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

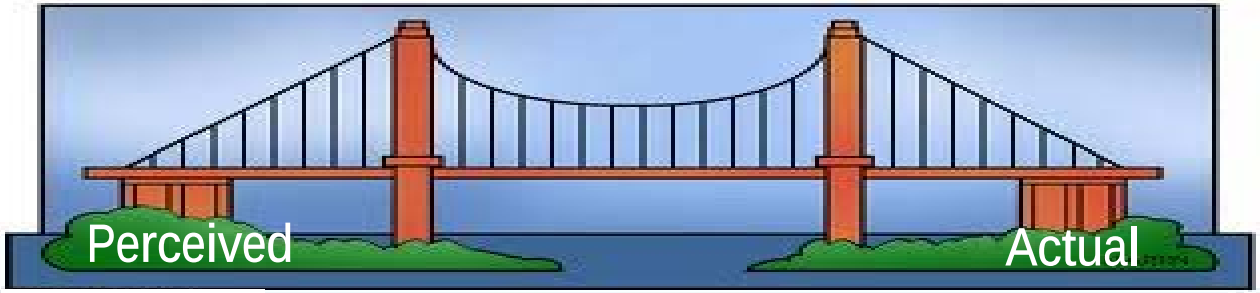
AGENDA – # 14269: 1st Look into the Inner Workings:

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

AGENDA – # 14268: 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!

Question

- In order to optimize connections and ensure performance, which of the following components must be considered when deploying your FICON infrastructure?
  - Internal mainframe wiring
  - z/OS and IOS
  - Channel Path and Components
  - Switching Devices
  - Storage Devices
  - All of the above
  - None of the above

Answer

- In order to optimize connections and ensure performance, which of the following components must be considered when deploying your FICON infrastructure?
  - ➢ **Internal mainframe wiring**
  - ➢ **z/OS and IOS**
  - ➢ **Channel Path and Components**
  - ➢ **Switching Devices**
  - ➢ **Storage Devices**
  - ✓ **ALL OF THE ABOVE**

Key Reasons For Using Switched-FICON

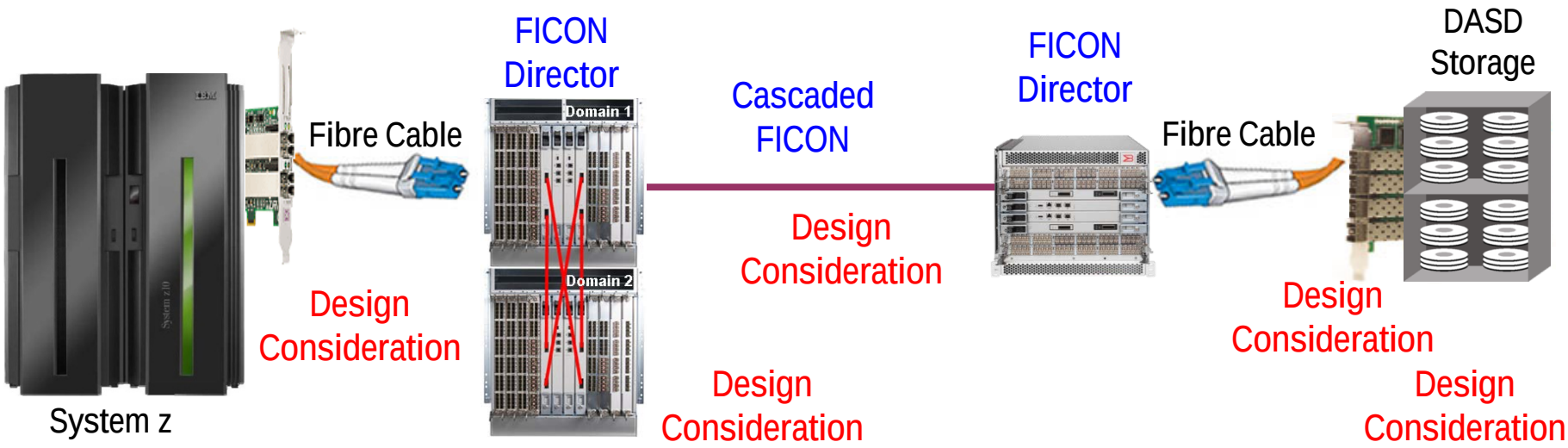
- There are 5 key technical reasons for connecting storage control units using switched-FICON fabrics:
  - To Improve reliability and availability of the deployed I/O infrastructure
  - To Improve data frame efficiency and reduce ISL link bit errors
  - For Building I/O fabrics that allow path and component consolidation while at the same time maximizing the utilization of all of those existing components
  - For Deploying long distance links that can be fully utilized beyond 10km (6.2 miles)
  - To Leverage evolving mainframe and FICON technologies.

Leverage New z/OS and System z Functionality

Some functionality REQUIRES customers to deploy switched–FICON:

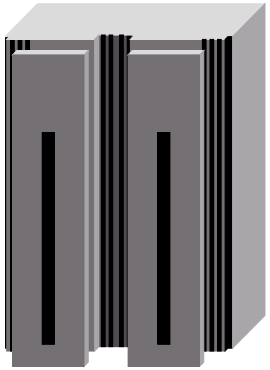
- **FICON Dynamic Channel Management:** Ability to dynamically add and remove channel resources at Workload Manager discretion can be accomplished only in switched-FICON environments.
- **zDAC:** Simplified configuration of FICON connected disk and tape through z/OS FICON Discovery and Auto Configuration (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

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 / zBC12

8 GBps

PCIe
Later z196/z114

8 GBps

InfiniBand
z10/ early z196

6 GBps

One Reason Why FICONX8 Cannot Do 8Gbps!

STI
z9

2.7 GBps

STI
z990/z890

2.0 GBps

STI
z900/z800

1.0 GBps

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

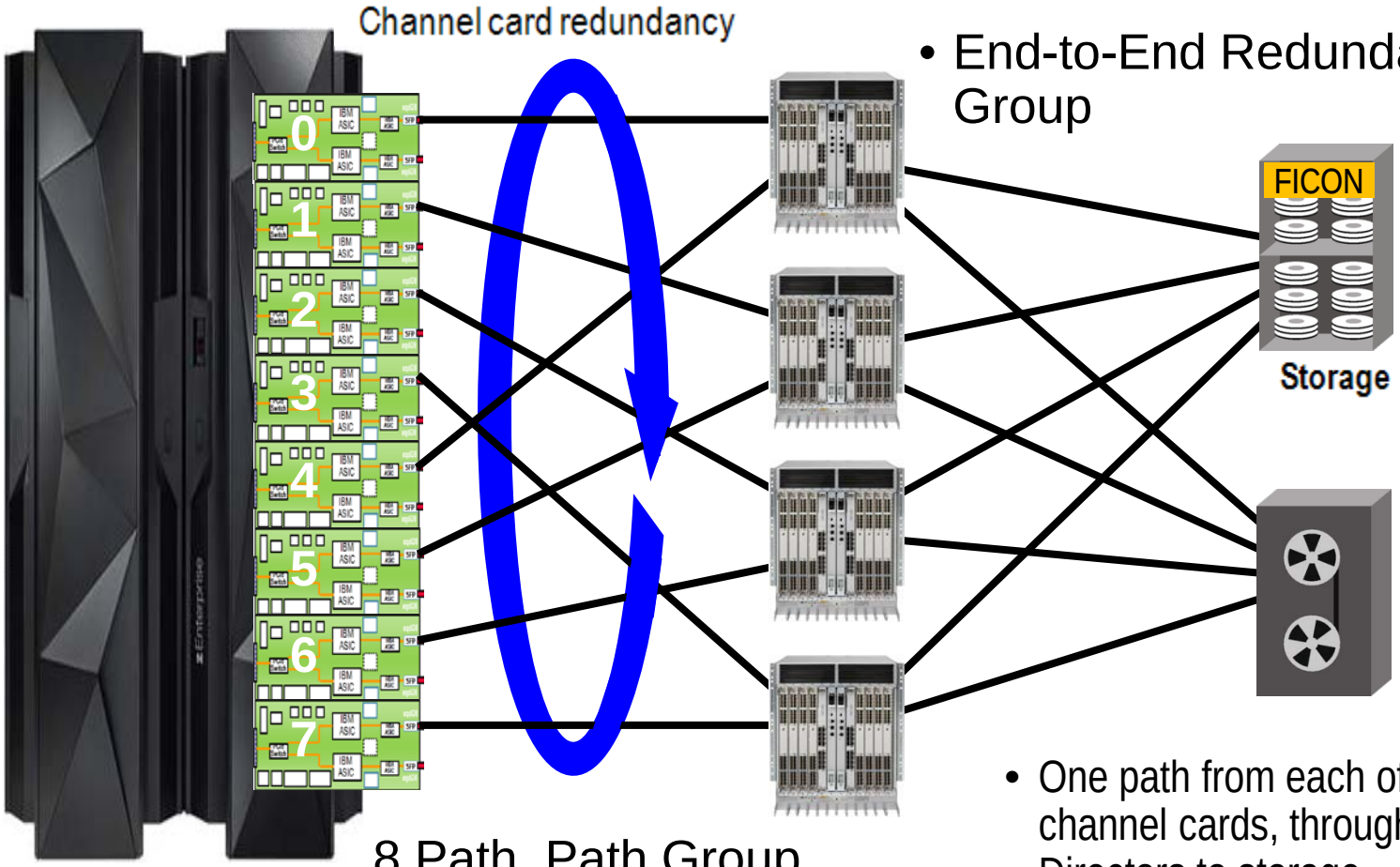
SHARE in Boston

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 / zBC12 Path Group Enhancement

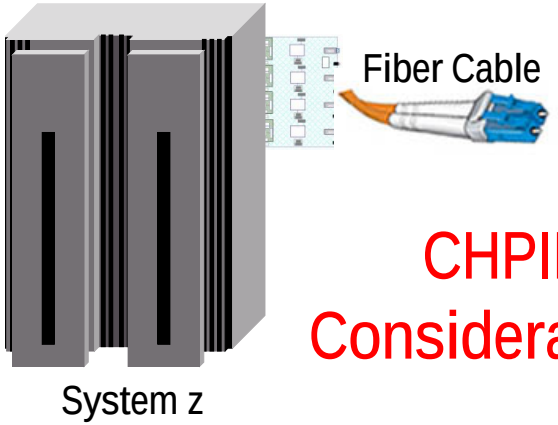
- The channel subsystem has updated **channel path selection algorithms** designed to provide improved throughput and I/O service times when abnormal conditions occur.
- Abnormal conditions include the following:
  - Multi-system work load spikes
  - Multi-system resource contention in the I/O Fabric(s) 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

Channel Sub-System Enhancements for zEC12 / zBC12 Path Group Enhancement

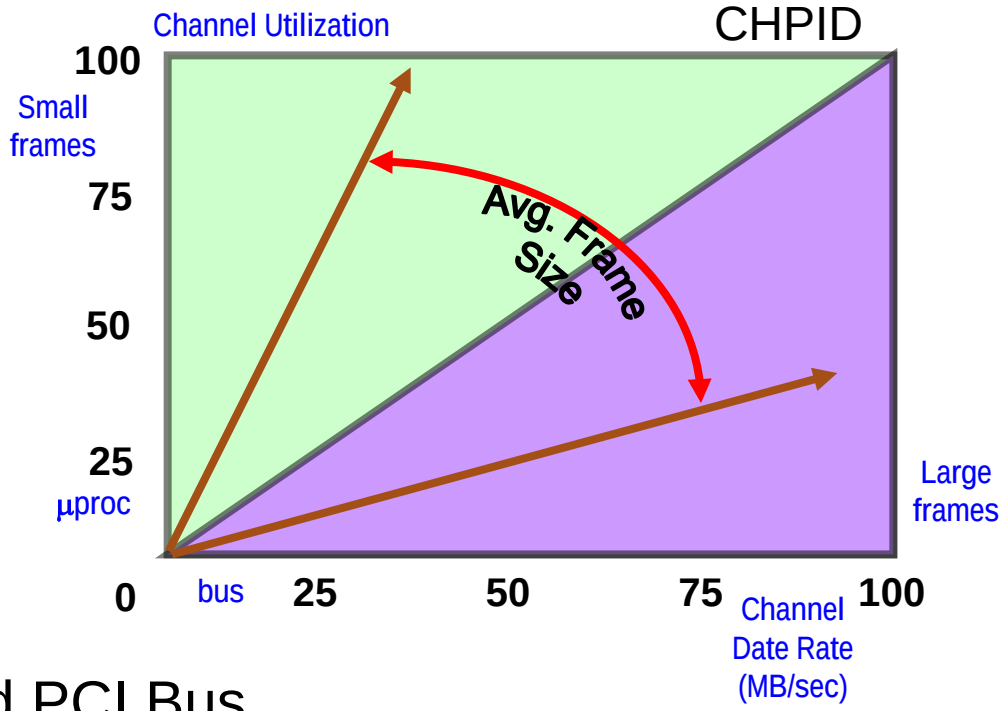
- When conditions occur that cause an imbalance in performance (e.g. I/O latency/throughput):
  - The channel subsystem will bias the path selection away from poorer performing paths toward the well performing paths.

- 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

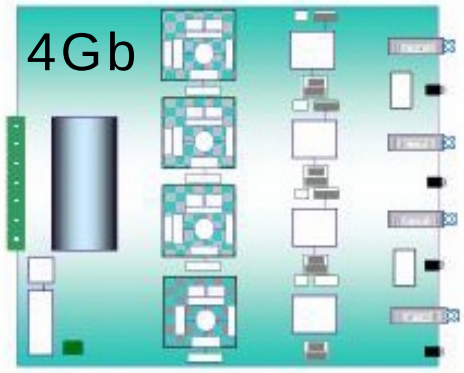


CHPID Considerations



- Channel Microprocessors and PCI Bus
- Average frame size for FICON
- Command Mode FICON
- High Performance FICON
- Buffer Credit considerations

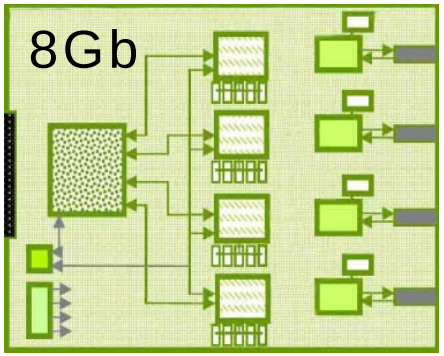
Current Mainframe Channel Cards (Features)



FICON Express4

- z196, z114, z10, z9
- 4 ports per feature
- <= 620 MBps Full Duplex out of 1600 MBps
- zHPF FICON Mode: <=770 MBps Full Duplex out of 1600 MBps
- 200 Buffer Credits/CHPID

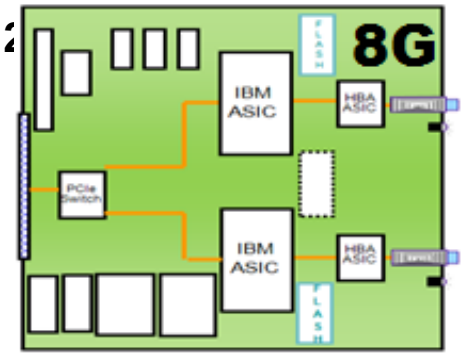
FICON Express4 provides the last native 1Gbps CHPID support



FICON Express8

- zXC12, z196, z114, z10
- 4 ports per feature
- <= 620 MBps Full Duplex out of 1600 MBps
- zHPF FICON Mode: <=770 MBps Full Duplex out of 1600 MBps
- 40 Buffer Credits/CHPID

FICON buffer credits have become very limited per CHPID

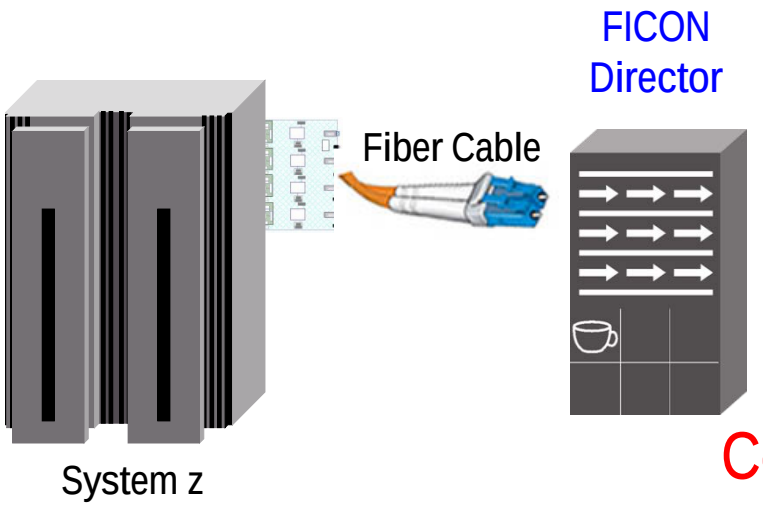


FICON Express8S

- zXC12, z196, z114
- 2 ports per feature
- <=620 MBps Full Duplex out of 1600 MBps
- zHPF FICON Mode: <=1600 MBps Full Duplex out of 1600 MBps
- 40 Buffer Credits/CHPID

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

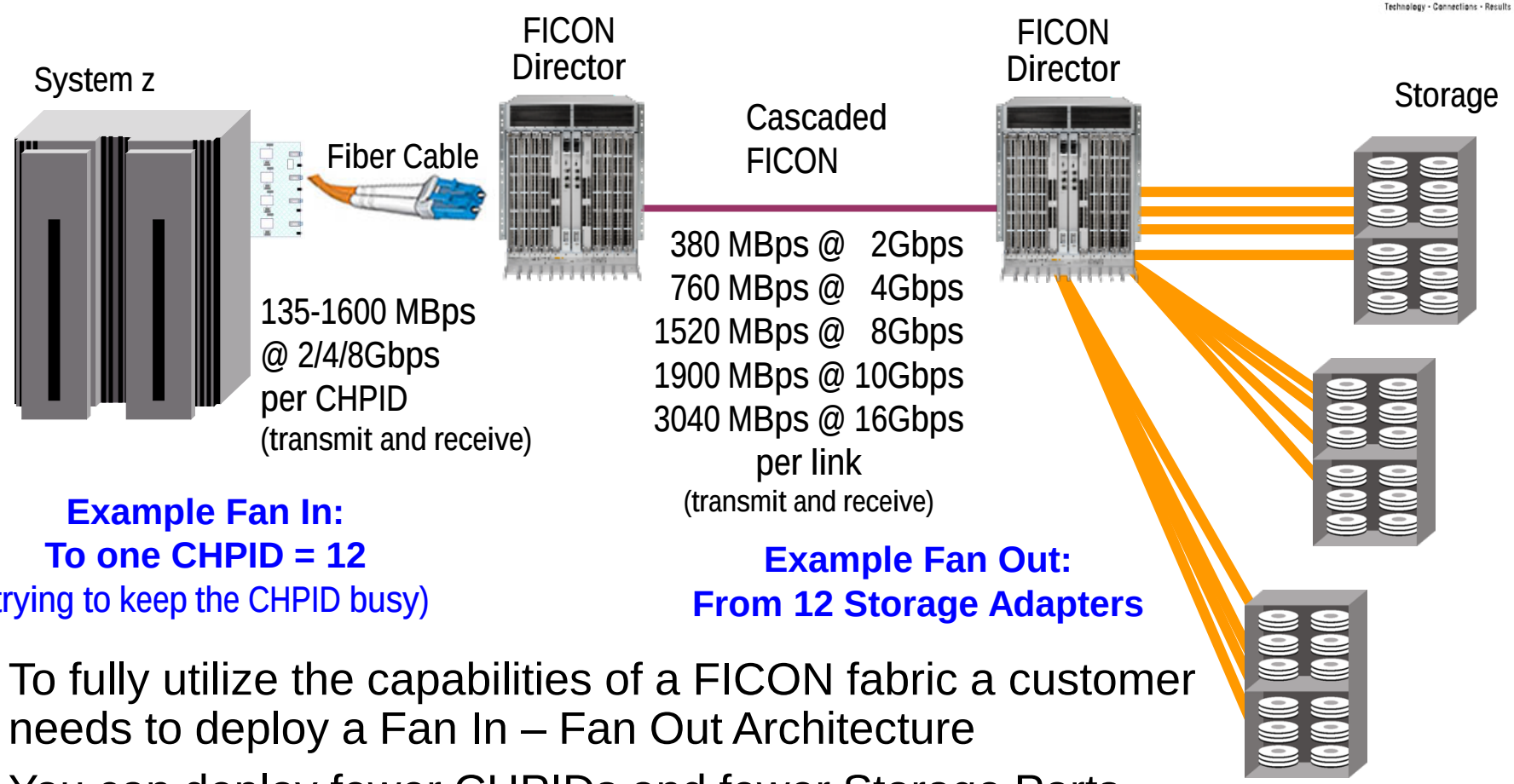
FICON/FCP Switching Devices



FICON Director Chassis Considerations

- Switched-FICON Networks are an I/O infrastructure best practice
- Provides for massive buffer credits per ISL link for long distance connectivity – as many as 4,000 or 5,000 on a link
- Architect for Fan In Fan Out consolidation and to effectively utilize resources
- Provide the best possible five-9s of availability for I/O

Fan In – Fan Out For FICON Channel Efficiency

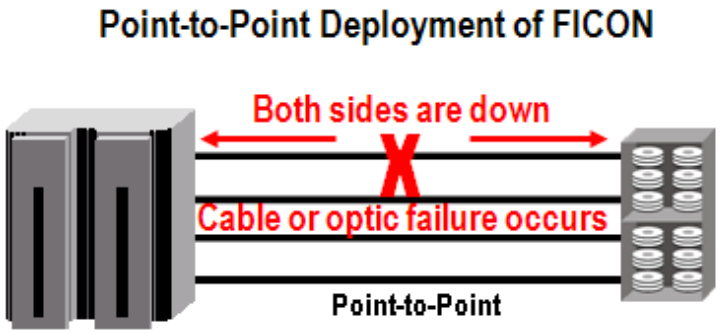


Example Fan In:
To one CHPID = 12
(trying to keep the CHPID busy)

Example Fan Out:
From 12 Storage Adapters

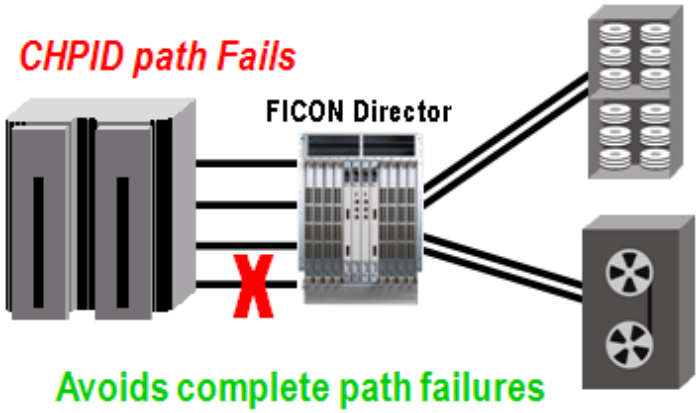
- To fully utilize the capabilities of a FICON fabric a customer needs to deploy a Fan In – Fan Out Architecture
- You can deploy fewer CHPIDs and fewer Storage Ports
- You can utilize the assets you have purchased at 80-100%
- You can scale up very easily without purchasing a lot of hardware
- You actually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

Availability After A Component Failure



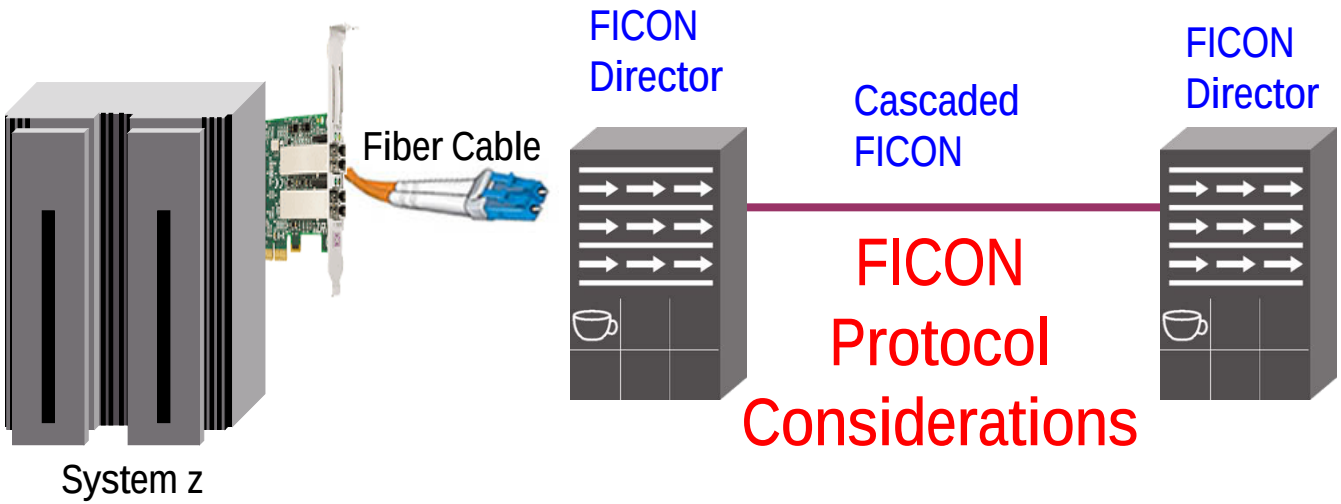
...BUT... Storage Port Remains Available!

- 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

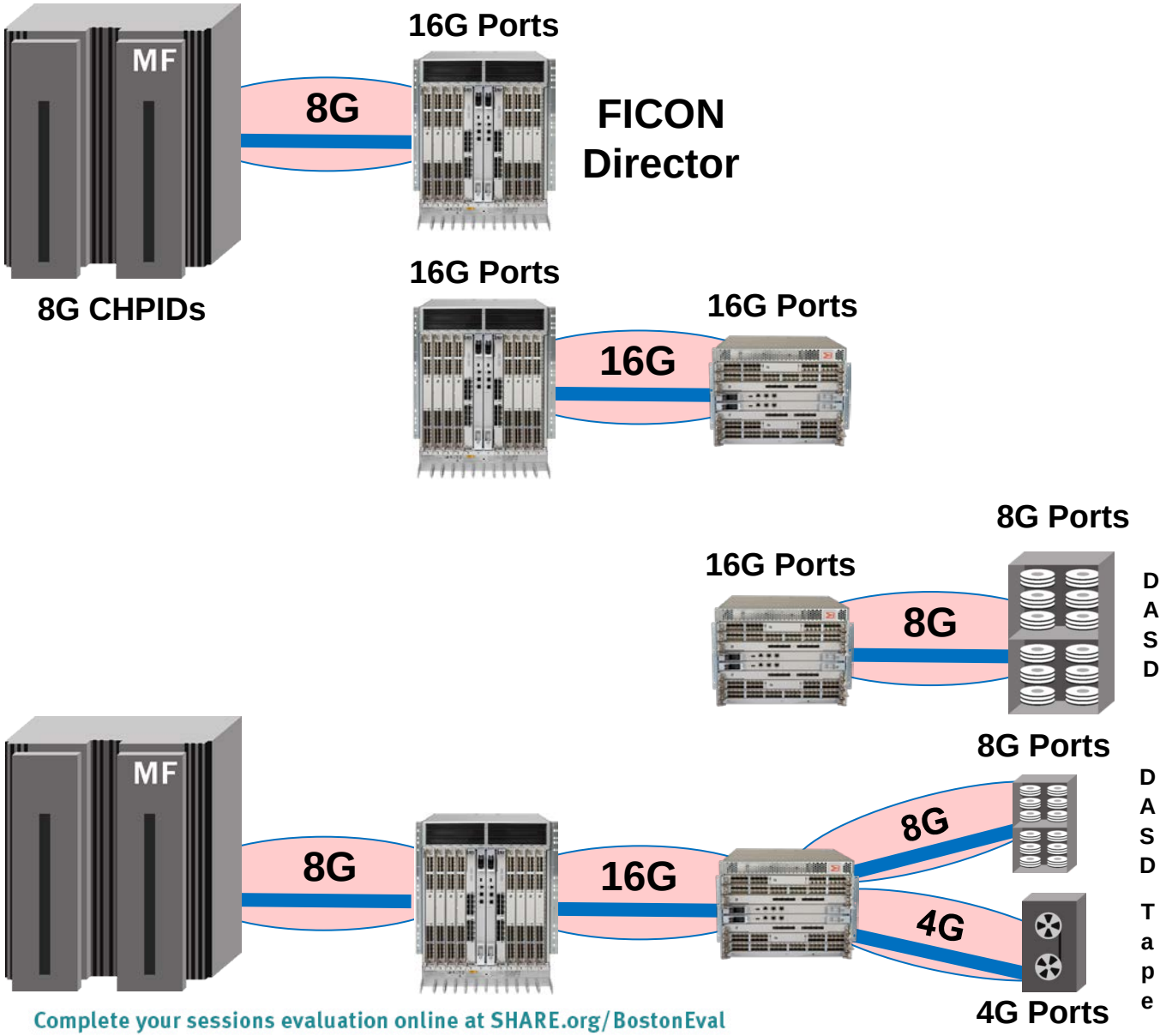
End-to-End FICON/FCP Connectivity



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

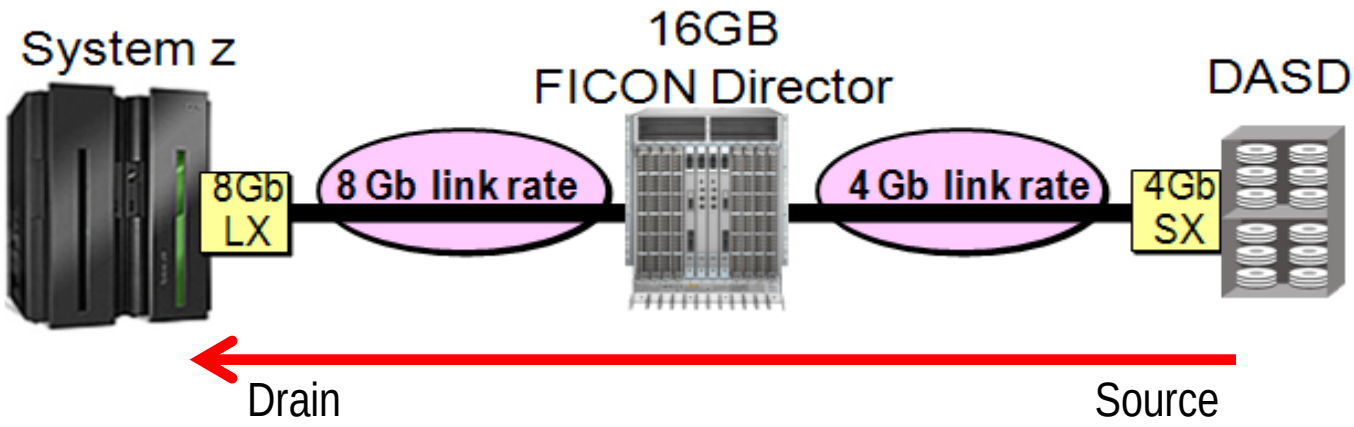
- Customers can use 2/4/8/16G and/or 10G for ISL traffic today
- The FICON Protocol uses 8b/10b data encoding for most link rates – but there is 20% frame payload overhead associated with it
- Newer 64b/66b data encoding (10G and 16G) is also in use and is more performance oriented (only 2% data payload overhead)
- 64b/66b facilitates Forward Error Correction to clean up ISL link bit errors
- MIDAW and zHPF make very good use of 8G FICON switch links

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

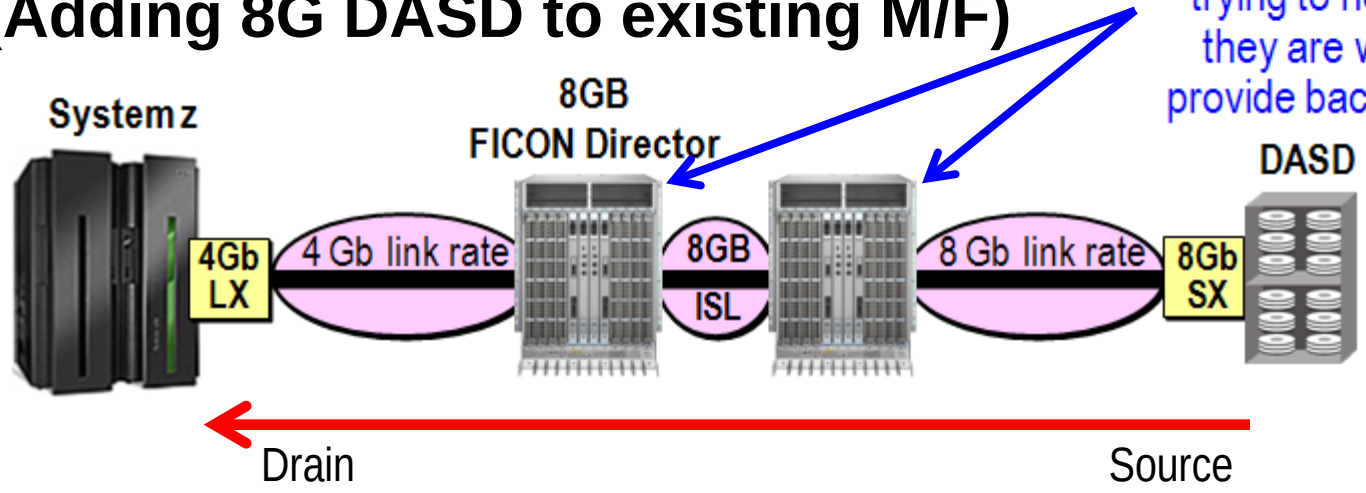
A Discussion On The Affects Of Link Rates



- 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)

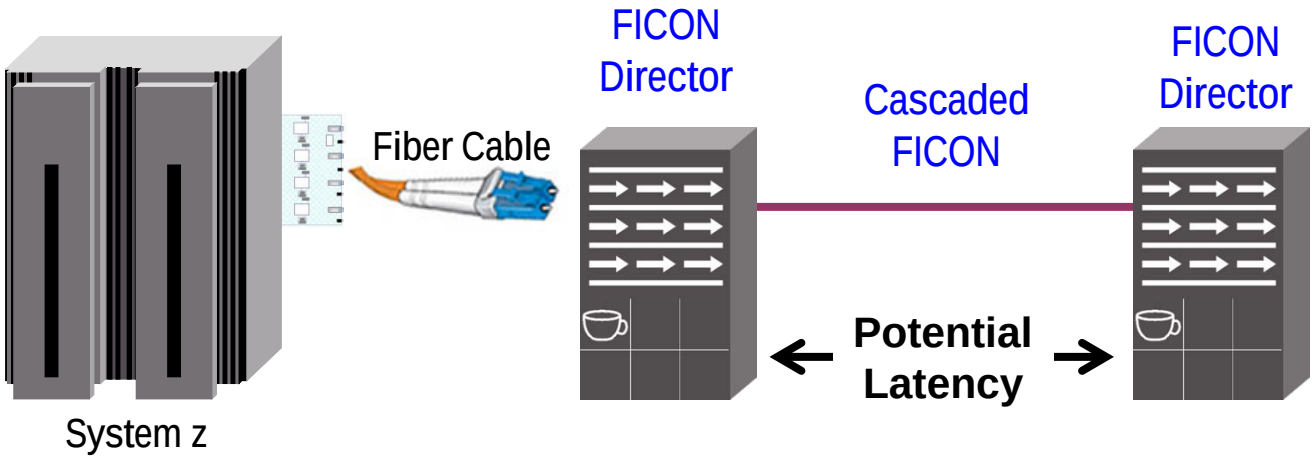
The Affects Of Link Rates
(Adding 8G DASD to existing M/F)



- 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.

HDD / SSD Storage and Storage Fabrics Must Have Synergy

User's want optimized hardware to avoid I/O bottlenecks and long latency times



Factors affecting Latency

- **Store-and-Forward**
- **Link bit errors**
- **Switch Latency**
- **ISL Congestion**
- **Fabric Backpressure**
- **Slow drain devices**
- **Long distance links**
- **Etc.**

Mainframe channels can produce 23,000 to 92,000 IOPS (FX8S)

- For zHPF, on Express8S, the minimum I/O time is 1/92,000 part of a second
- Or, from measured values, it takes a around 10.8695 microseconds to do one zHPF I/O operation when using FICON Express8S channel cards

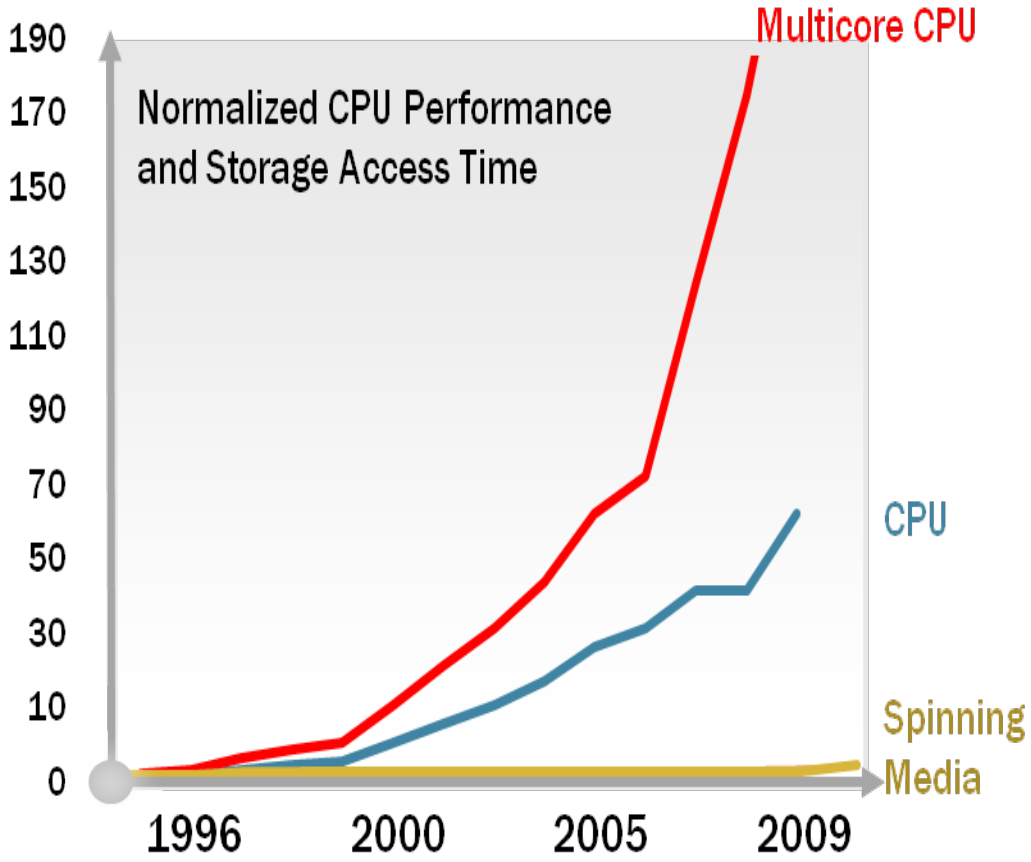
Special consideration should be paid to your deployed switch's latency

- Every µs of switch latency reduces the ability to maintain a max I/O flow
- High switch or path latency can reduce your system's maximum IOPS

HDD / SSD Storage and Storage Fabrics Must Have Synergy

User's want optimized hardware to avoid I/O bottlenecks and long latency times

User's must minimize all data path latencies or suffer reduced performance!



175x
CPU PERFORMANCE IMPROVEMENT

1.3x
SPINNING MEDIA PERFORMANCE IMPROVEMENT

- Dramatically lower access times
- Increased IOPS (100x)
- Greater reliability, lower power and cooling
- 71% CAGR (2010−2014)

Performance and Switch Latency

Brocade switching device latency DOES NOT impact maximum IOPS

When IBM was determining the number of I/O Operations that they would ultimately publish for FICON Express8S, the numbers were based on “measured values” in their lab rather than artifically calculated values.

Because IBM used Brocade switching devices and IBM storage in their labs for this testing, the “measured values” include Brocade switch latency as well as the responding device latency.

So the published FICON Express8S IOPS figures for Command Mode FICON and High Performance FICON includes all of time for the I/O and its acknowledgement from storage.

- At 23,000 IOPS (FICON Express8S using Command Mode FICON at 8Gbps), the calculation would be that each I/O operation takes 1/23,000 part of a second
- 1 second / 23,000 IOPS = 43.48 µs so it takes that long to do one CM I/O operation
- At 92,000 IOPS (FICON Express8S using zHPF at 8Gbps), the calculation would be that each I/O operation takes 1/92,000 part of a second
- 1 second / 92,000 IOPS = 10.87 µs so it takes that long to do one zHPF I/O operation
- Each measured I/O had 4.2µs of Brocade switch latency included in the measured results
  - 2.1 µs for the I/O and 2.1 µs for the acknowledgment

Performance – Some Math about Switch Latency

Charting The Affect That Switch Latency Has On IOPS

SHARE
Technology · Connections · Results

A CHART TO CAUSE BLINDNESS!

IBM tests their FICON Express Cards through Brocade Switching Devices to determine Max achieable IOPS

M/F	CHPID	Type	Rated IOPS	Each I/O µs	Max I/O with Switch Latency of 2x 0.7 µs	Max IOPS %	Max I/O with Switch Latency of 2x 2.1 µs	Brocade Max IOPS %	Max I/O with Switch Latency of 2x 5 µs	Other Max IOPS %	Max I/O with Switch Latency of 2x 10 µs	Other Max IOPS %	Max I/O with Switch Latency of 2x 100 µs	Other Max IOPS %
z10	FX2 and FX4	CM	14000	71.43	14000	100.00%	14000	100.00%	12281	87.72%	10938	78.13%	3684	26.32%
z10	FX2 and FX4	zHPF	31000	32.26	31000	100.00%	31000	100.00%	23664	76.34%	19136	61.73%	4306	13.89%
z10	FX8	CM	20000	50.00	20000	100.00%	20000	100.00%	16667	83.33%	14286	71.43%	4000	20.00%
z10	FX8	zHPF	52000	19.23	52000	100.00%	52000	100.00%	34211	65.79%	25490	49.02%	4561	8.77%
z114	FX8S	CM	23000	43.48	23000	100.00%	23000	100.00%	18699	81.30%	15753	68.49%	4107	17.86%
z114	FX8S	zHPF	92000	10.87	92000	100.00%	92000	100.00%	47917	52.08%	32394	35.21%	4742	5.15%
z196	FX8	CM	20000	50.00	20000	100.00%	20000	100.00%	16667	83.33%	14286	71.43%	4000	20.00%
z196	FX8	zHPF	52000	19.23	52000	100.00%	52000	100.00%	34211	65.79%	25490	49.02%	4561	8.77%
z196	FX8S	CM	23000	43.48	23000	100.00%	23000	100.00%	18699	81.30%	15753	68.49%	4107	17.86%
z196	FX8S	zHPF	92000	10.87	92000	100.00%	92000	100.00%	47917	52.08%	32394	35.21%	4742	5.15%
zEC12	FX8	CM	20000	50.00	20000	100.00%	20000	100.00%	17921	89.61%	15198	75.99%	4068	20.34%
zEC12	FX8	zHPF	52000	19.23	52000	100.00%	52000	100.00%	39951	76.83%	28546	54.90%	4650	8.94%
zEC12	FX8S	CM	23000	43.48	23000	100.00%	23000	100.00%	20293	88.23%	16870	73.35%	4179	18.17%
zEC12	FX8S	zHPF	92000	10.87	92000	100.00%	92000	100.00%	59990	65.21%	37496	40.76%	4839	5.26%

Performance and Switch Latency for New Channels / Disk

Brocade switching devices will provide synergy for new channels and protocols

Latency Matters! Regardless of what causes it in the I/O infrastructure.

M/F	CHPID	Type	Rated IOPS	Each I/O µs	Max I/O with Switch Latency of 2x 10 µs	Max IOPS %	Max I/O with Switch Latency of 2x 50 µs	Max IOPS %	Max I/O with Switch Latency of 2x 100 µs	Max IOPS %
zEC12	FX8S	CM	23000	43.48	16870	73.35%	7180	31.22%	4179	18.17%
zEC12	FX8S	zHPF	92000	10.87	37496	40.76%	9375	10.19%	4839	5.26%

This chart does double the latency to account for the frame acknowledgements.

Any number in red indicates more than a 50% loss in maximum IOPS due to path latency.

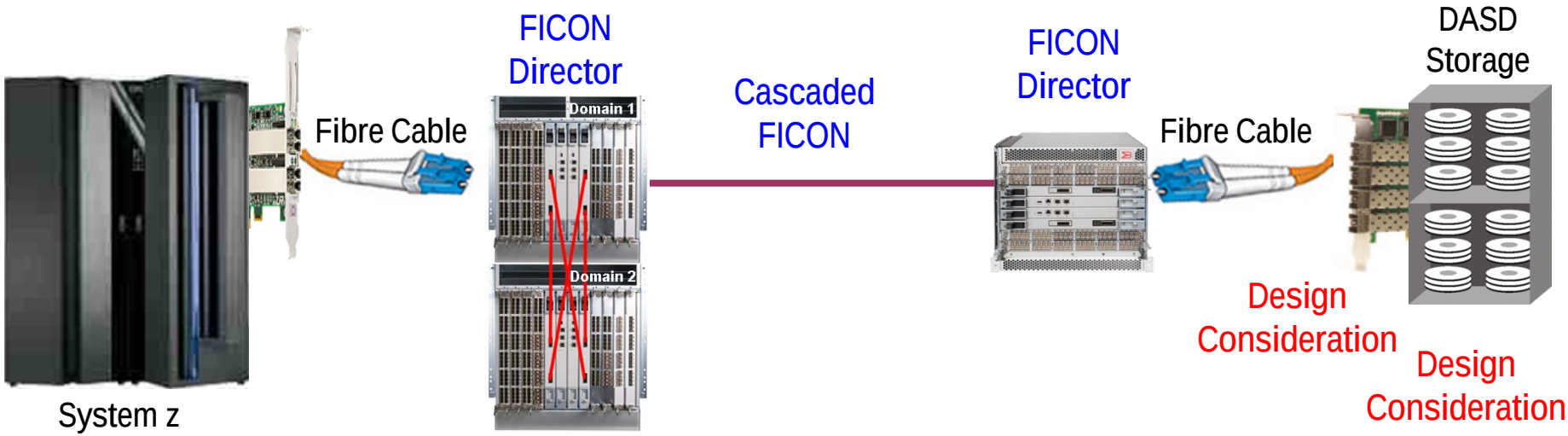
It does not take very much latency in the link to cause significant throttling down of the maximum possible IOPS.

Store-and-Forward frame routing, fabric congestion, long distance connectivity, slow draining devices and other similar issues would all throttle IOPS.

In the future, additional latency in a fabric will reduce the performance increase achieved by newer channel hardware (driver and protocol) as well as new disk.

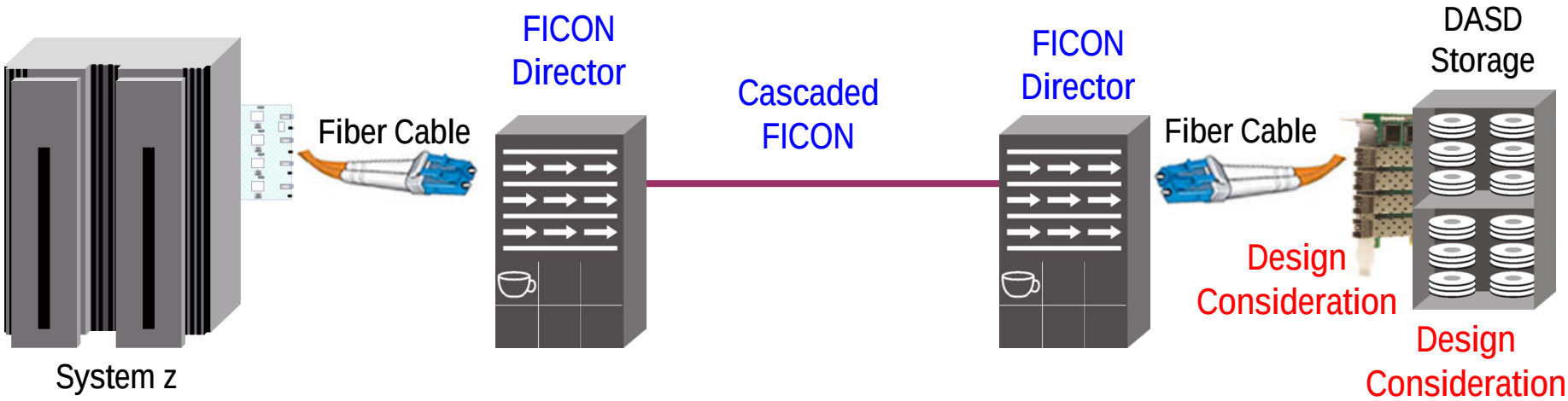
Customers who want to realize the full performance of their new channel systems should match them with switches that will meet those performance demands!

End-to-End FICON/FCP Connectivity



- A user's most challenging considerations most likely occur due to DASD storage deployment

Connectivity with storage devices



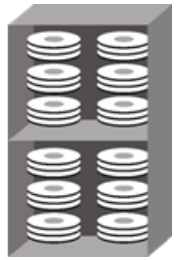
Storage adapters can be throughput constrained

- 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?

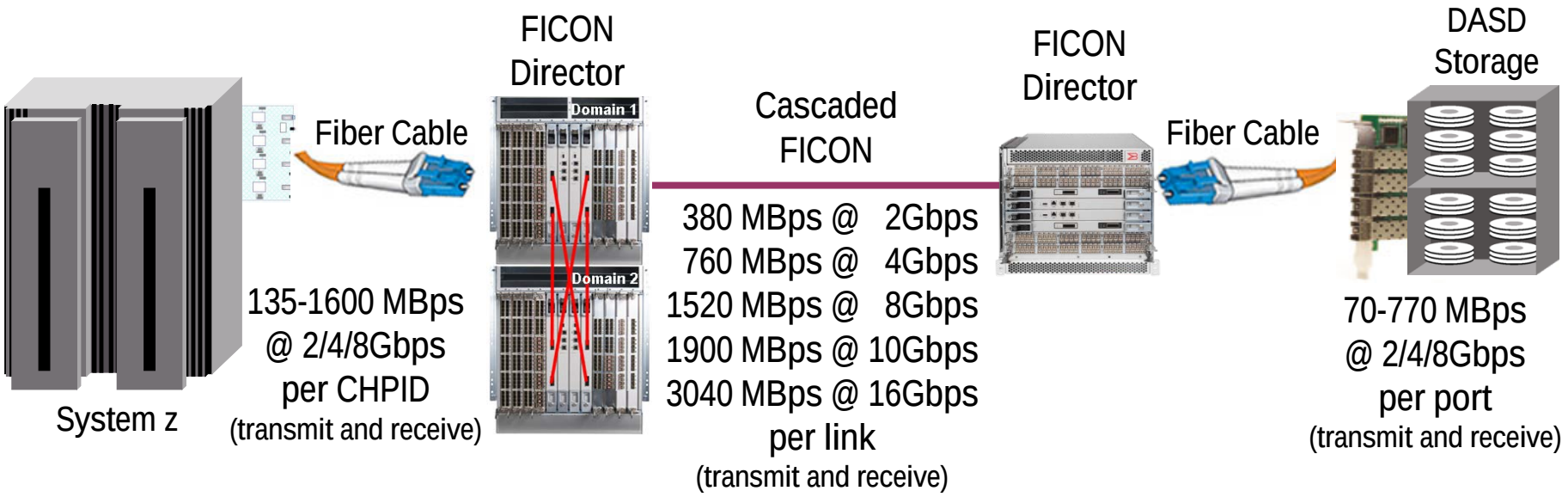
What kinds of internal bottlenecks does a DASD array have?

- 7200rpm, 10,000rpm, 15,000rpm (might never be any 20,000rpm)
- 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!

Brocade FICON Certification

Brocade Certified Architect for FICON

- Certification for Brocade Mainframe-centric Customers
- Available since September 2008
- Updated for 8Gbps in June 2010
- Updated for 16Gbps in November 2012
- This certification tests the attendee's ability to understand IBM System z I/O concepts, and demonstrate knowledge of Brocade FICON Director and switching infrastructure components
- ***Brocade would be glad to provide a free 2-day BCAF certification class for Your Company or in Your City!***
- Ask me how to make that happen for you!

SHARE in Boston

Brocade Certified Architect for FICON (BCAF)

This FICON Certification is Unique in the Industry

BCAF is a Preparatory Certification Seminar – 2 days

- These classes teach the certification material
- Certification is awarded only after successful completion of the examination
- We have been holding classes since mid-2008
- This is good for mainframers who desire to become professionally certified as FICON subject matter experts
- This uses advanced materials and is not well suited for professionals with less than 1 year of experience

Total number of attendees at these seminars since 2008: **455** (as of May 2013)

Total number of Brocade FICON Certifications awarded: **222+**

We also have a Brocade Accredited FICON Specialist credential (based on WBT training and an exam): **122** awarded

Brocade Mainframe Social Media

Visit Brocade's Mainframe Blog Page at:

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

Almost 250,000 hits

Also Visit Brocade's New Mainframe Communities Page at:

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/

- www.linkedin.com Groups

My Next Presentation:

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

Thursday August 15, 2013 -- 9:30am to 10:30am -- Session 14268

Please Fill Out Your Evaluation Forms!!

Thank You For Attending Today!

QR Code

This was session:

14269

And Please Indicate on those forms if there are other presentations you would like to see in this track at SHARE!

Your Eval ⬇

My Reaction!

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

Thursday August 15, 2013 -- 8:00am to 9:00am -- Session 14269