IBM Inside Sales

International Technical Support Organization Global Content Services

ITSO – z System Hardware Workshop

www.ibm.com/redbooks

Part 3 – I/O Subsystem

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml

BladeCenter®, DB2®, e business(logo)®, DataPower®, ESCON, eServer, FICON, IBM®, IBM (logo)®, MVS, OS/390®, POWER6®, POWER6+, POWER7®, Power Architecture®, PowerVM®, S/390®, System p®, System p5, System x®, z Systems®, System z9®, System z10®, WebSphere®, X-Architecture®, zEnterprise®, z9®, z10®, z196®, z114®, zEnterprise System z196®, zEnterprise System z114®, zEnterprise System zEC12®, zEnterprise System zBC12®, z13®, z/Architecture®, z/OS®, z/VM®, z/VSE®, zSeries®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Important information about today's workshop

- The ITSO z hardware team created 7 IBM z13 presentations to be delivered today
  - Part 1 – IBM z13 – Positioning / introduction
  - Part 2 – z13 CPC Details Capacity and Performance
  - Part 3 – z13 I/O Subsystem
  - Part 4 – Native PCIe Adpters – zEDC and RoCE (what's new with z13)
  - Part 5 – HMC, CoD and RAS and zAware
  - Part 6 – Installation Planning
  - Part 7 – Software Support

- The main references for the presentations today are:.
  - IBM z13 Technical Guide – Redbook – SG24-8251
  - IBM z13 Technical Introduction – Redbook - SG24-8250

- **Part of the available material may not be presented..** ☹
  - Even if we don't cover the presentations entirely,
    - The material can be download from:
      - http://www.redbooks.ibm.com/Redbooks.nsf/pages/addmats

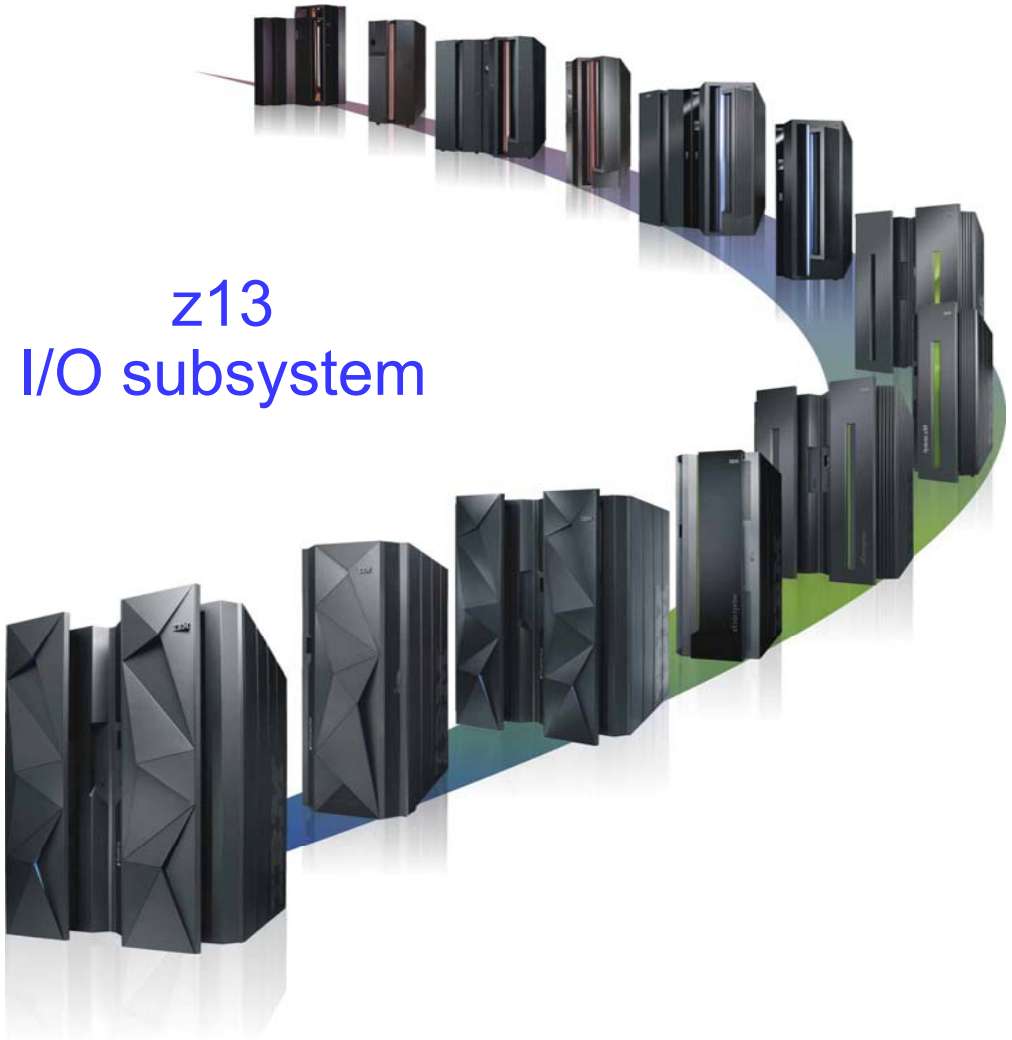
- The material being presented may not fully match the copied version you have

- **You can always get the latest version .. If you want it, just ask !** ☺

- **Please ask questions, make comments and share your own experiences at any time**

- **Thank You !**

z13
I/O subsystem



Glossary for z13 I/O

Acronym	Full Name	Description / Comments
N/A	I/O drawer	I/O drawer introduced with z10 BC and also supported on zEC12, z196, z114 and z13; has 8 I/O card slots
N/A	PCIe switch	Industry standard PCIe switch Application-specific integrated circuit (ASIC) used to fanout (or multiplex) the PCI bus to the I/O cards within the PCIe I/O drawer
N/A	PCIe I/O drawer	I/O drawer that supports PCIe bus I/O infrastructure; has 32 I/O card slots
PCI-IN	PCIe interconnect	Card in the I/O Drawer or the PCIe I/O drawer that contains the PCIe switch ASIC. Connects to the PCIe fanout in the CPC drawer
N/A	PCIe Gen3 fanout	Card on front of CPC drawer that supports PCIe Gen3 bus; used exclusively to connect to the PCIe I/O drawer; PCIe fanout supports FICON Express8S, FICON Express16S, Crypto Express4S, Crypto Express5S, OSA-Express4S, OSA-Express5S, Flash Express, 10 GbE RoCE and zEDC Express
HCA2-C	HCA2-C fanout	HCA2 Copper fanout used to connect CPC Drawer to the I/O Drawer to support FICON Express8 I/O features
HCA3 or HCA3-O LR	HCA3-O LR fanout for 1x IFB	For 1x InfiniBand at unrepeated distances up to 10 km; 5 Gbps link data rate; *4 ports per fanout*; may operate at 2.5 Gbps or 5 Gbps. Based upon capability of DWDM. Can communicate with an HCA2-O LR fanout; third generation Host Channel Adapter
HCA3 or HCA3-O	HCA3-O fanout for 12x IFB	For 12x InfiniBand at 150 meters; supports 12x IFB and 12x IFB3 protocols; *increased service times when using 12x IFB3 protocol;* 6 GBps link data rate; two ports per fanout; can communicate with an HCA2-O fanout; third generation Host Channel Adapter
RoCE	RDMA over CE	High speed inter communication fabric facilitating data movement between z System and other platforms
ICA SR	PCIe-O SR fanout	A new generation of short reach (SR) PCIe-based Coupling link, connects to ICA SRs (up to 150m) in other z13 systems. 8 GBps link date rate; two ports per fanout.

Glossary for z Systems Coupling

Acronym	Full name	Comments
AID	Adapter identification	Fanout has AID instead of a PCHID
CIB	Coupling using InfiniBand	CHPID type used by z13, zEC12, zBC12, z196, z114, z10, System z9
CS5	Coupling using ICA SR	CHPID type used by z13
HCA	Host Channel Adapter	Path for communication
PSIFB	Parallel Sysplex using InfiniBand	InfiniBand Coupling Links
12x IFB	12x InfiniBand	12 lanes of fiber in each direction
1x IFB	1x InfiniBand	Long Reach - one pair of fiber
12x IFB3	12x InfiniBand3	Improved service times of 12x IFB on HCA3-O

Type	z13	zEC12/zBC12/z196/z114
HCA2-C fanout	Copper – Connects to I/O Drawer	Copper - Connects to I/O Cage (zEC12 and z196 only) or I/O Drawer
HCA2-O fanout	N/A	Optical - Coupling 12x InfiniBand
HCA2-O LR fanout	N/A	Optical - Coupling 1x InfiniBand
PCIe Gen3 fanout	Copper – Connects to PCIe I/O drawer	Copper - Connects to PCIe I/O drawer
HCA3-O fanout	Optical - Coupling 12x InfiniBand	Optical - Coupling 12x InfiniBand
HCA3-O LR fanout	Optical - Coupling 1x InfiniBand	Optical - Coupling 1x InfiniBand
ICA SR	Optical – Coupling PCIe Gen3	N/A

z System I/O Interface Evolution

	z9	z196 z10	zEC12 zBC12 z196 z114	z13
Extended Link (processor to I/O Cage/Drawer)	eSTI (IBM)	InfiniBand	PCI Express Gen2	PCI Express Gen3
I/O Cage/drawer Backplane	mSTI (IBM)	InfiniBand	PCI Express Gen2	PCI Express Gen3
Coupling Link	ISC (IBM)	InfiniBand	InfiniBand	PCI Express Gen3, InfiniBand

z System I/O Subsystem Internal Bus Interconnect Speeds

PCIe Gen3
z13

16 GBps

PCIe Gen2
zEC12/zBC12
z196/z114

8 GBps

InfiniBand
zEC12/zBC12
z10/z196/z114

6 GBps

eSTI
z9

2.7 GBps

STI
z990/z890

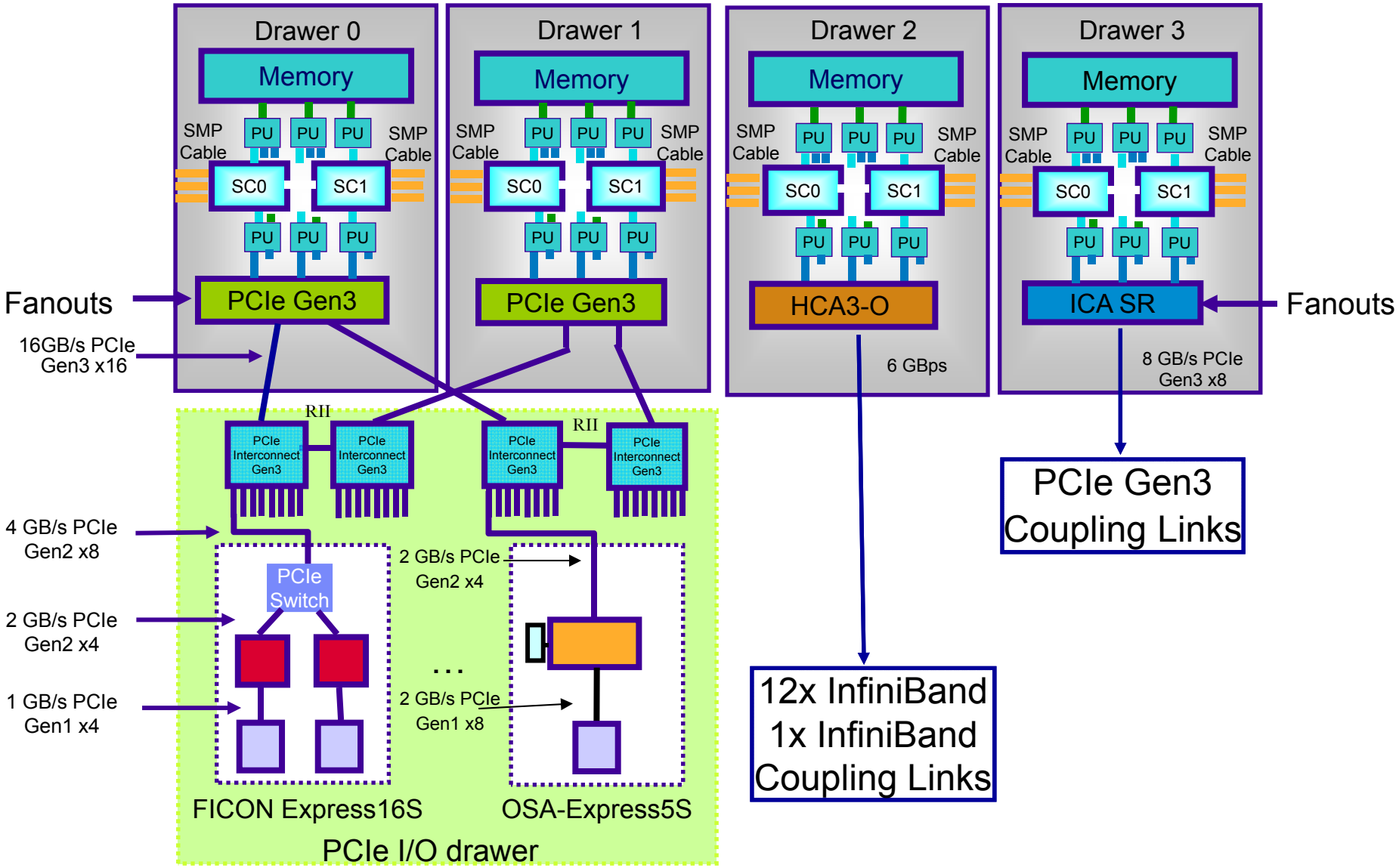
2 GBps

STI: Self-Timed Interconnect

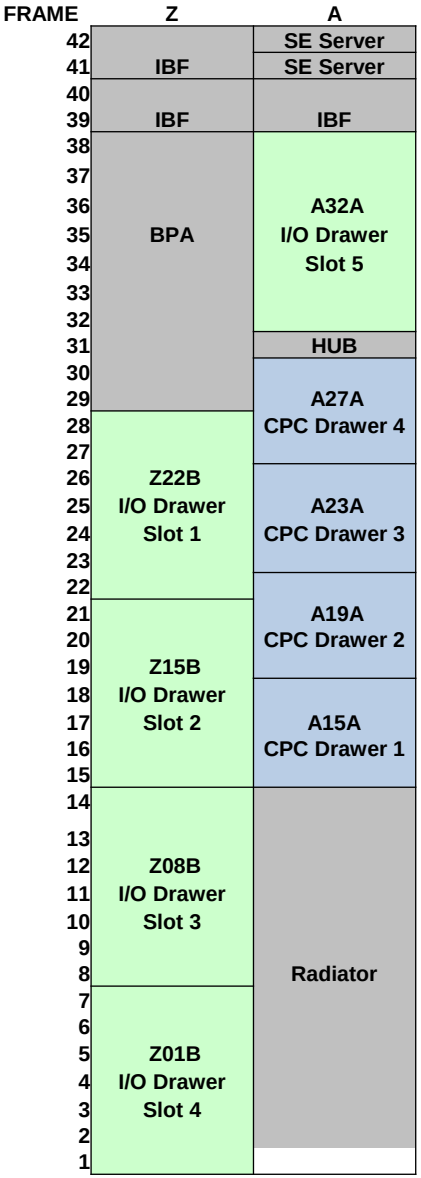
z13 supports two different internal I/O bus infrastructures

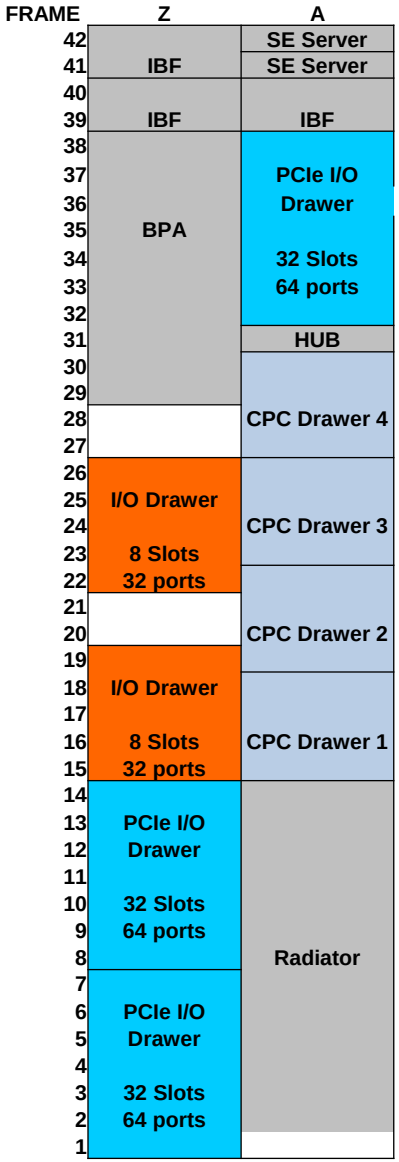
- **The InfiniBand I/O infrastructure first introduced on the z10:**
  - InfiniBand fanouts supporting the 6 GBps InfiniBand I/O interconnect
  - InfiniBand I/O card domain multiplexers with Redundant I/O interconnect in:
    - The 5U, 8-slot, 2-domain I/O drawer (carry forward only)
  - Selected non-PCIe I/O features
    - FICON Express8 LX (FC 3325) and FICON Express8 SX (FC 3326) carry forward for Frame Roll MES only
- **PCI Express Generation 3 (PCIe Gen3) I/O infrastructure introduced with z13**
  - PCI Express Generation 2 (PCIe Gen2) I/O infrastructure introduced with z196/z114
  - PCIe Gen3 fanouts and PCIe Interconnect Gen3 supporting the 16 GBps PCIe I/O interconnect will be shipped in **ALL** new builds, Migration Offerings, z System Exchange Program, and Frame Roll MES
    - Frame Roll MES cannot carry forward the existing Gen 2 PCIe fanout or Gen 2 PCIe Interconnect
  - PCIe Interconnect Gen3 (i.e. PCIe Switch) with Redundant I/O interconnect for I/O domains in a 7U, 32-slot, 4-domain PCIe I/O drawer
  - PCIe I/O drawers for New Builds, Migration Offerings and System z Exchange Programs
- **New Build z13 supports PCIe I/O Drawers ONLY**
  - Maximum 5 drawers
- **Upgrades from zEC12 and z196 will allow carry forward of:**
  - I/O Drawer (Maximum 2)

z13 GA I/O Infrastructure



z13 Frame Layout





- An I/O frame slot is a physical location in the A or Z frame for an I/O drawer or PCIe I/O drawer to be inserted = 7u
- PCIe I/O drawer uses 1 I/O frame slot = 7u
  - 32 two port I/O slots = 64 ports each
  - 5 drawers maximum = 160 slots, 320 ports total
- I/O drawer uses 0.7 frame slot = 5u
  - 8 four port I/O slots = 32 ports total
  - 2 drawers carry forward **ONLY** maximum in I/O frame slots 1 and 2 only

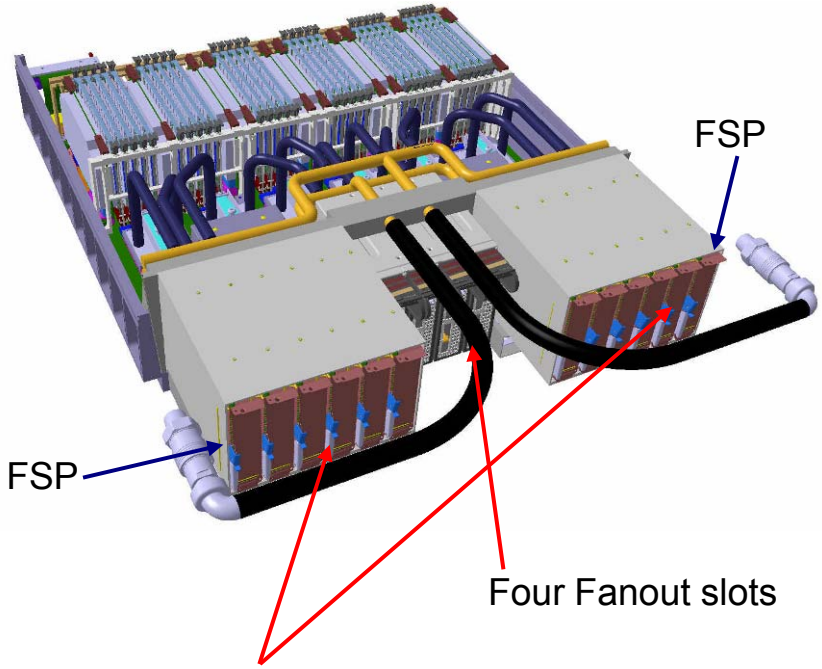
z13 Carry Forward (Field Upgrade) Rules for I/O Features (All PCIe I/O Features Can be Carried Forward)

FICON Express8 Features Carried Forward	8-slot I/O Drawers Required	Maximum PCIe Drawers/Slots available for other 'I/O' features
0	0	5/160
1 to 8	1	4/128
9 to 16	2	3/96
17 or more	Not Supported!	

Empty slots in a carried forward drawer can NOT be filled by MES.

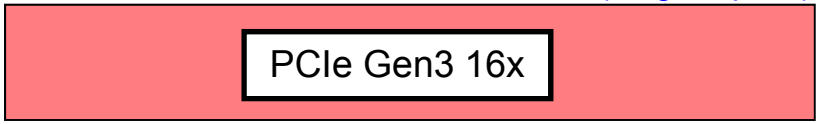
Note: Large I/O configurations may require two or more CPC drawers.

z13 Processor Drawer Connectivity for I/O and Coupling

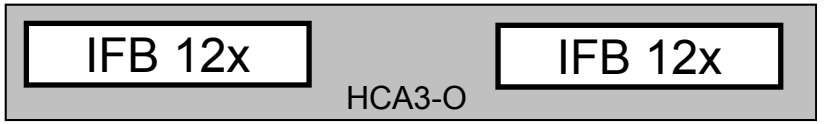


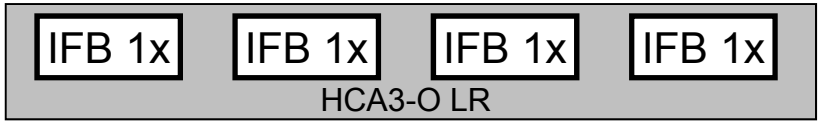
- Ten PCIe fanout slots per drawer (40 maximum)
  - ICA (ICA SR) two-port 8 GBps PCIe Gen3 fanout 150 meter fiber optic coupling link

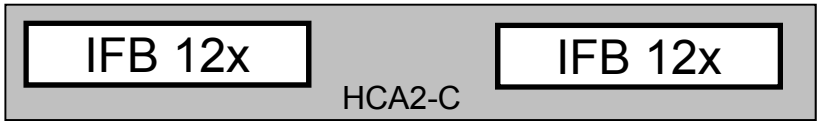
  - PCIe Gen3 one-port 16 GBps PCIe fanout connects to a switch card for an 8-slot PCIe I/O domain (Plugs in pairs)



- Four IFB HCA fanouts fanout slots per drawer (16 maximum on a four drawer system)
  - HCA2-C 2-port 6 GBps I/O drawer fanout (plugs in pairs)
  - HCA3-O 2-port 12x IFB Coupling Link fanout
  - HCA3-O LR 4-port 1x IFB Coupling Link fanout



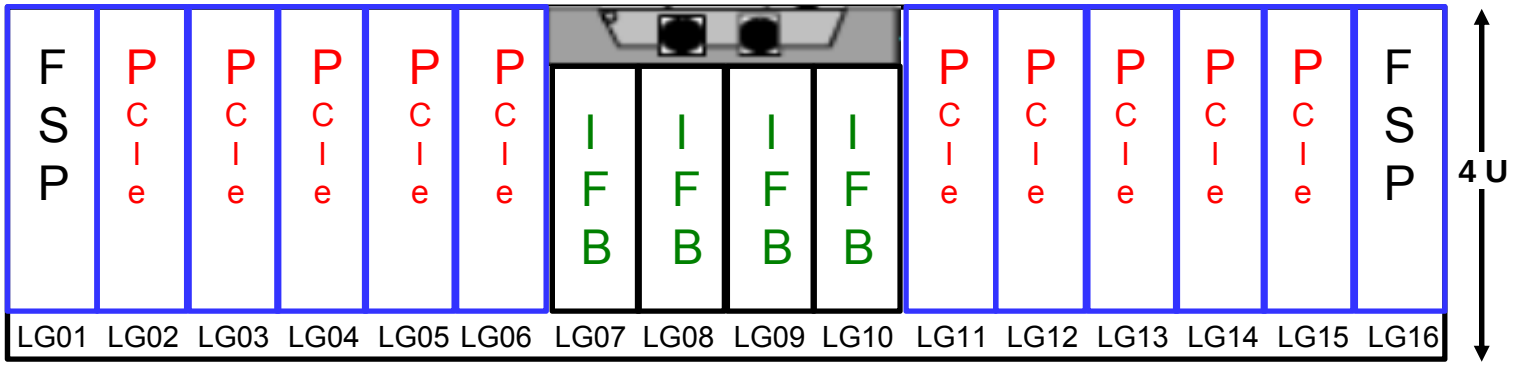




Carry forward
(One pair only)

Carry forward or
New Build

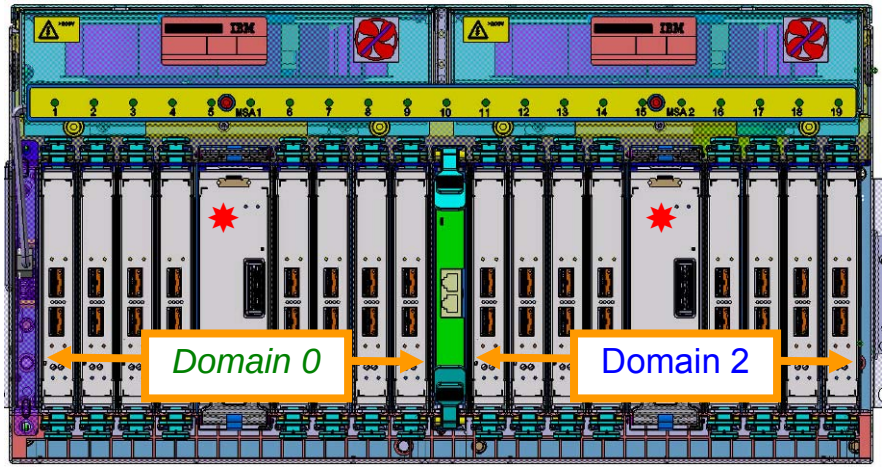
CPC Drawer I/O Fanout and Flexible Support Processor Locations



- PCIe Fanout Slots (Ten), slots LG02 – LG06 and LG11 – LG15, can support:
  - Up to 10 one-port PCIe 16 GBps I/O fanouts to support up to 10 domains in 32-slot PCIe I/O drawers
    Note: A zEC12 book with eight two-port 8 GBps PCIe fanouts supports up to 16 domains in 32-slot PCIe I/O drawers but with 50% less bandwidth per domain
  - Up to 10 ICA (PCIe-SR) two-port coupling fanouts to support up to 20 8 GBps coupling links
- IFB Fanout Slots (Four), LG07 – LG10, can support:
  - Up to four HCA3-O 12x InfiniBand coupling fanouts, 8 12x 6 GBps links – Two per fanout
  - Up to four HCA3-O LR 1x InfiniBand coupling fanouts 16 1x 5 Gbps links – Four per fanout
    Note: A zEC12 book with 8 two-port HCA3-O 12x InfiniBand coupling fanouts can support 16 12x links
    A zEC12 book with 8 four-port HCA3-O LR 1x InfiniBand coupling fanouts can support 32 1x links
  - Up to two two-port HCA2-C 6GBps I/O fanouts (2 8-slot I/O drawers) with two slots left
- Slots LG01 and LG16 always have Flexible Support Processors (FSPs)

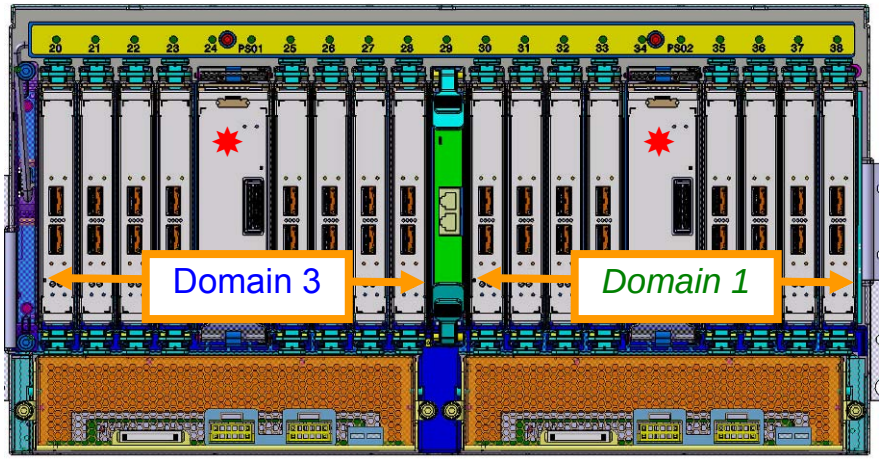
PCIe 32 I/O slot drawer

Front



7U

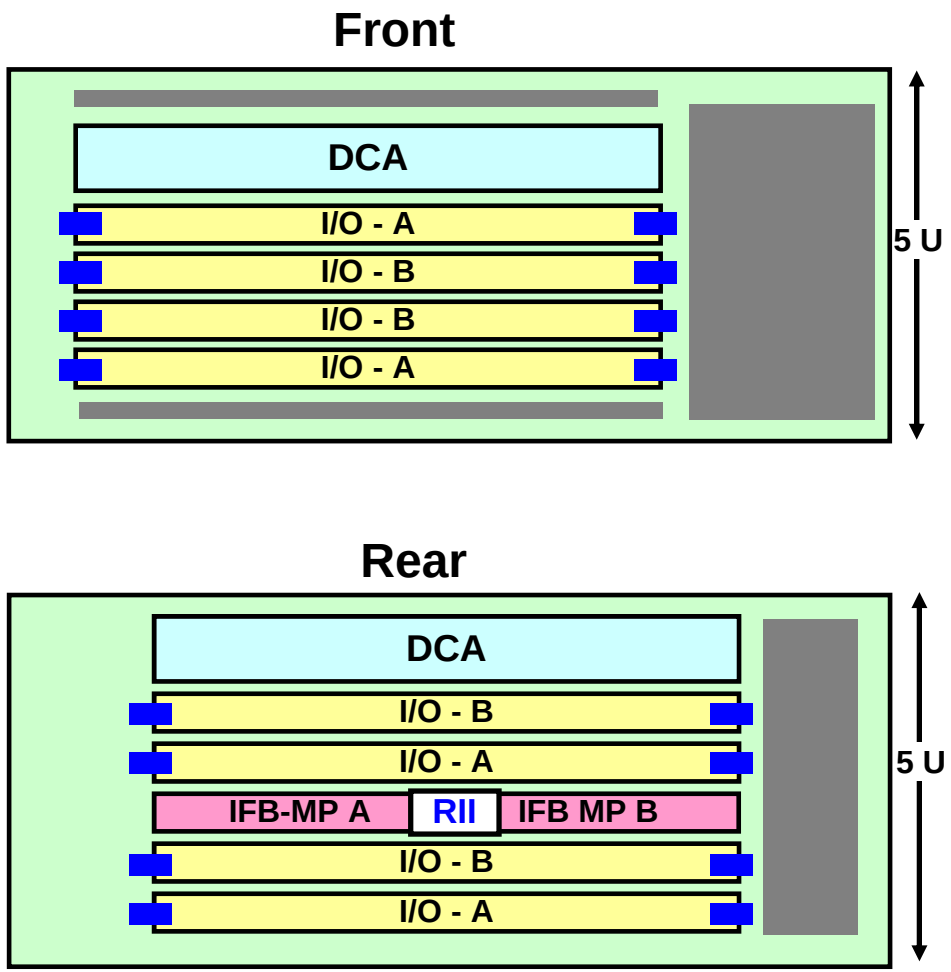
Rear



7U

- Supports only PCIe I/O cards
  - z13: Up to five drawers
  - zEC12: Up to five drawers
- Supports 32 PCIe I/O cards, 16 front and 16 rear, vertical orientation, in four 8-card domains (shown as 0 to 3).
- Requires four 16 GBps PCIe switch cards (✸), each connected to a 16 GBps PCIe I/O interconnect to activate all four domains.
- To support Redundant I/O Interconnect (RII) between front to back domain pairs *0-1* and 2-3 the two interconnects to each pair will be from 2 different PCIe fanouts. (All four domains in one of these cages can be activated with two fanouts.)
- Concurrent field install and repair.
- Requires 7 EIA Units of space (12.25 inches ≈ 311 mm)

8-slot I/O Drawer (Carry Forward from zEC12 and z196 only)

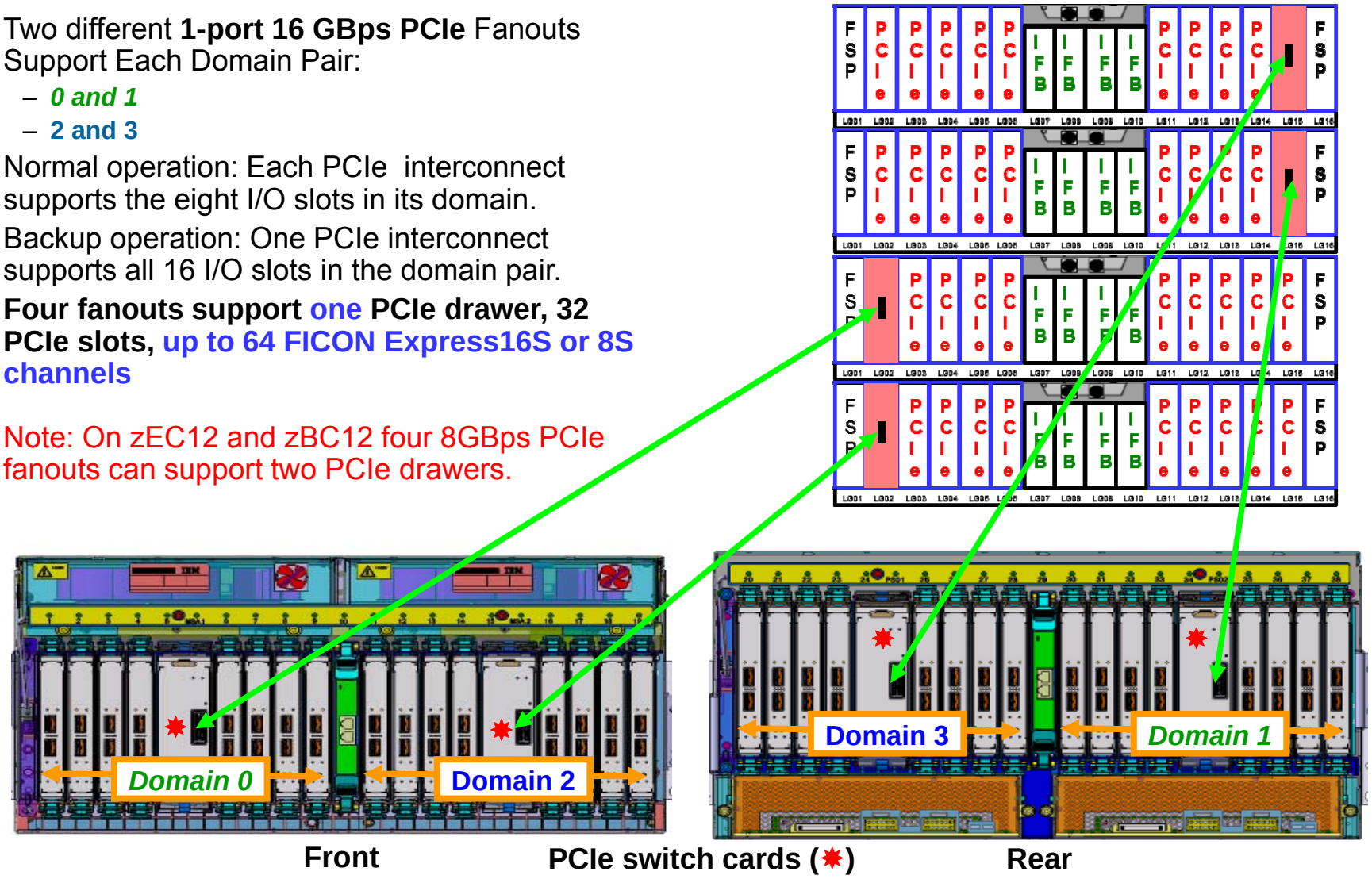


- Supports carry forward of FICON Express8 4-port features
- Supports 8 I/O slots, 4 front and 4 back, horizontal orientation, in two 4-slot domains (shown as A and B)
- Requires two IFB-MP daughter cards, each connected to a 6 GBps InfiniBand interconnect to activate both domains.
- To support Redundant I/O Interconnect (RII) between the two domains, the two interconnects must be from two different InfiniBand fanouts. (Two fanouts can support two of these drawers.)
- Concurrent add, repair.
- Limits LPAR to 1 TB
- Requires 5 EIA Units of space (8.75 inches ≈ 222 mm)

z13 NE1 or NC9 Redundant I/O Interconnect – PCIe I/O Drawer

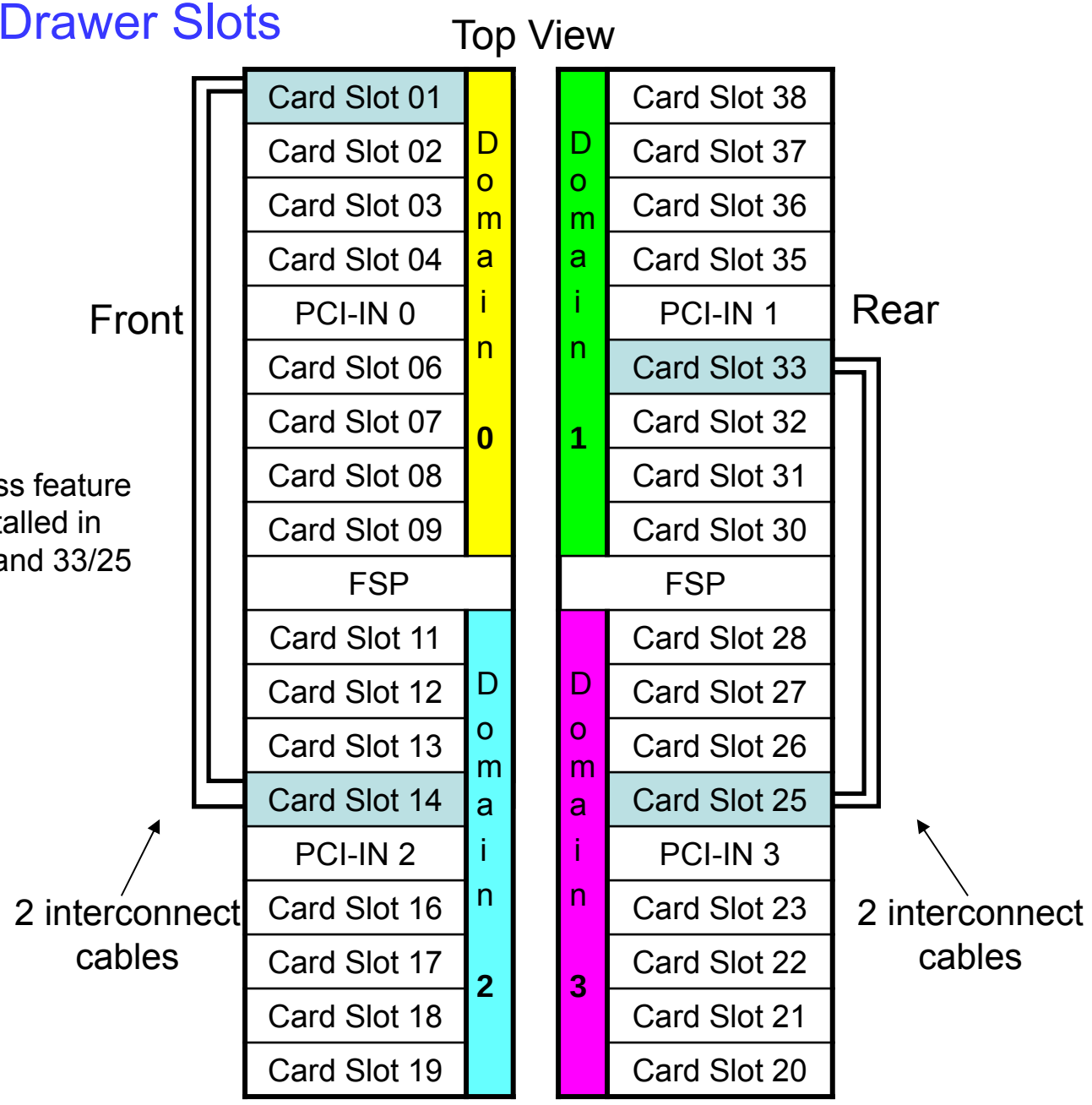
- Two different **1-port 16 GBps PCIe** Fanouts Support Each Domain Pair:
  - ***0 and 1***
  - **2 and 3**
- Normal operation: Each PCIe interconnect supports the eight I/O slots in its domain.
- Backup operation: One PCIe interconnect supports all 16 I/O slots in the domain pair.
- **Four fanouts support one PCIe drawer, 32 PCIe slots, up to 64 FICON Express16S or 8S channels**

Note: On zEC12 and zBC12 four 8GBps PCIe fanouts can support two PCIe drawers.



z13 NE1 or NC9 Redundant I/O Interconnect – PCIe I/O Drawer

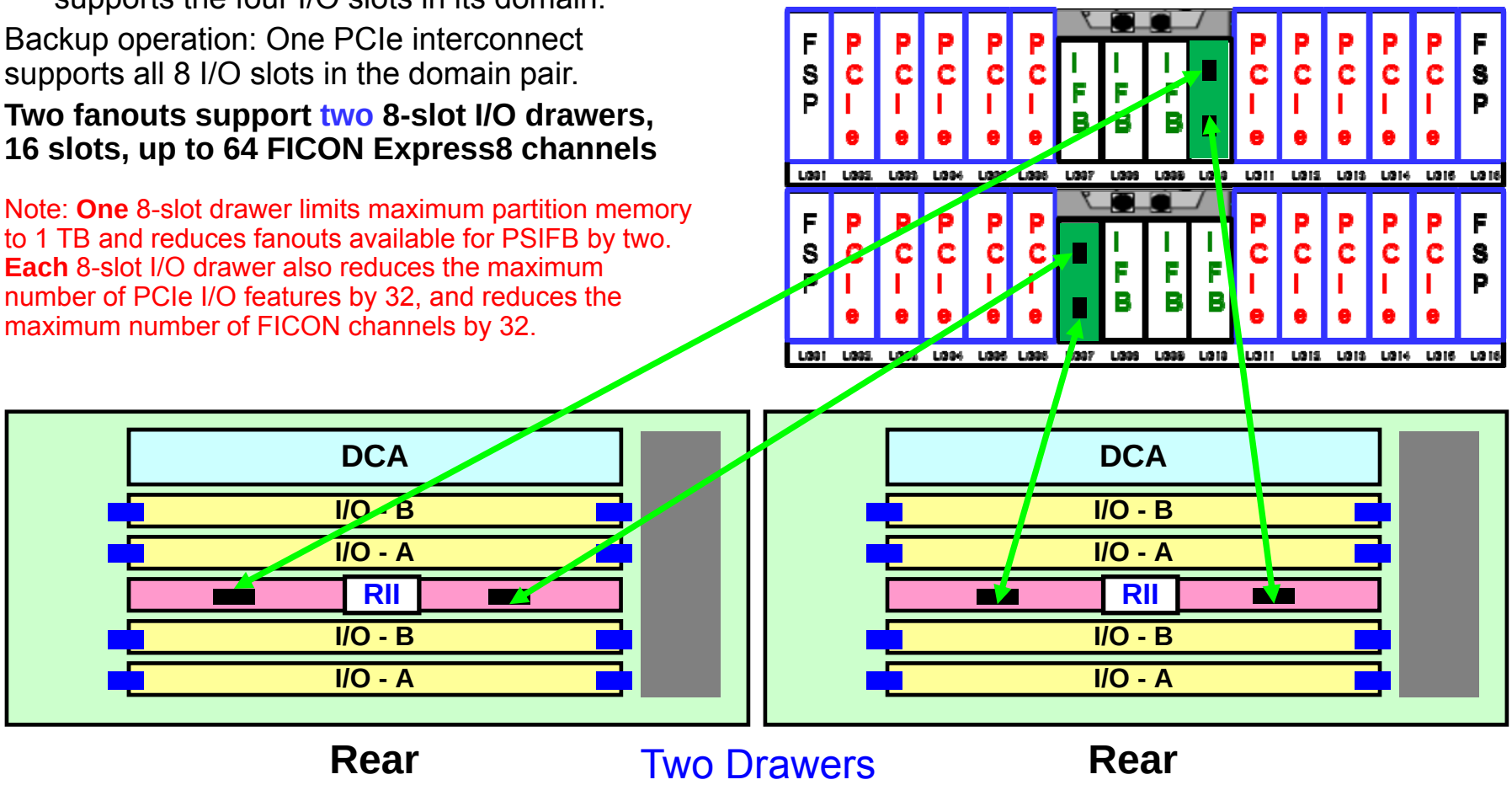
PCIe I/O Drawer Slots



z13 N63 Redundant I/O Interconnect – Two 8-slot I/O Drawers

- Two different **2-port 6 GBps HCA2-C** Fanouts Support Each Domain Pair:
  - **A and B**
  - Normal operation: Each PCIe interconnect supports the four I/O slots in its domain.
- Backup operation: One PCIe interconnect supports all 8 I/O slots in the domain pair.
- **Two fanouts support two 8-slot I/O drawers, 16 slots, up to 64 FICON Express8 channels**

Note: **One** 8-slot drawer limits maximum partition memory to 1 TB and reduces fanouts available for PSIFB by two. **Each** 8-slot I/O drawer also reduces the maximum number of PCIe I/O features by 32, and reduces the maximum number of FICON channels by 32.



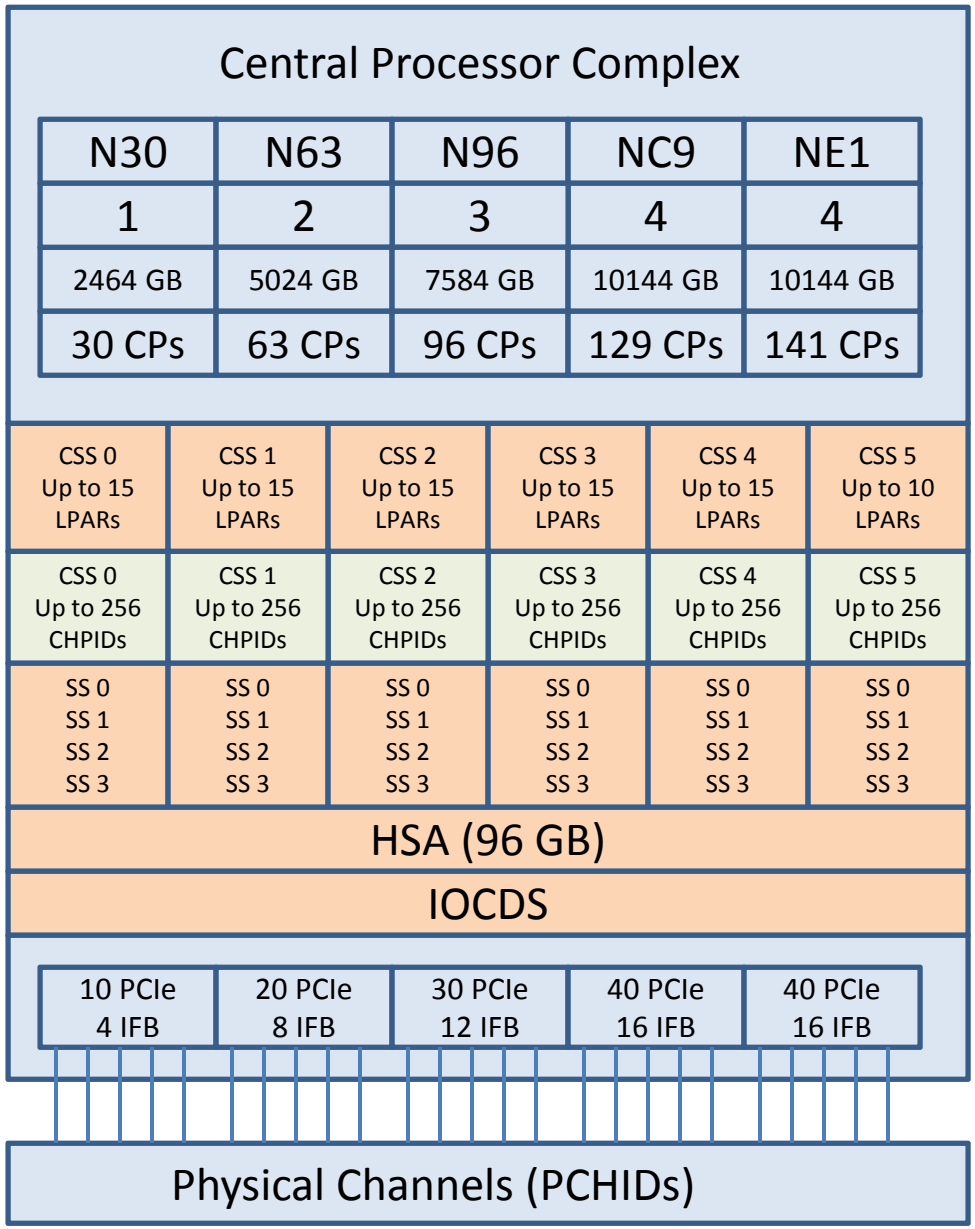
Logical channel subsystems (CSS)

Subchannel sets (SS), Function Definitions, and Logical Partitions on z13

- Six Logical Channel Subsystems (CSS) each with four subchannel sets (SS) and up to 256 channels
  - Maximum channel count includes channels spanned to more than one CSS
  - Total physical channels depend on I/O features configured
  - Up to 63.75k base IODEVICEs in SS 0 and 64 k alias IODEVICEs each in SS 1 to SS 3 per CSS
- FUNCTION definition support for virtualized RoCE and zEDC independent of CSS
- Up to 85 Logical Partitions: 15 each in CSS 0 – 4, 10 in CSS 5 (Partitions B – F Reserved)
  - Only channels and IODEVICEs defined in its CSS can be assigned to an LPAR
  - Any defined FUNCTION can be assigned to any LPAR

z13					
Function Definitions for up to16 RoCE features (31 LPARs each) and 8 zEDC features (15 LPARs each)					
CSS 0	CSS 1	CSS 2	CSS 3	CSS 4	CSS 5
Up to 15 Logical Partitions	**Up to 15 Logical Partitions**	**Up to 15 Logical Partitions**	**Up to 15 Logical Partitions**	**Up to 15 Logical Partitions**	**Up to 10 Logical Partitions**
Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k	**Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k**	**Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k**	**Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k**	**Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k**	**Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k**
Up to 256 Channels	Up to 256 Channels	Up to 256 Channels	Up to 256 Channels	Up to 256 Channels	Up to 256 Channels

CPC Models and channel subsystems (CSS) relationship



- Up to four CPC Drawer
- Maximum memory size
- Maximum Processing Units
- Up to 15 LPARs in CSS 0 - 4
  Up to 10 LPARs in CSS 5
- Up to 6 Logical Channel Subsystems
- Up to four Subchannel Sets per CSS
- Single / fixed HSA
- One active IOCDS
- PCIe and IFB Adapter per Drawer
- PCIe and IFB copper cables
- PCIe Drawer and I/O Drawer

z13
Supported I/O Features

z13 “New Build” I/O and MES Features Supported

New Build Features

- **Features – PCIe I/O drawer**
  - **FICON Express16S** (SX and LX, 2 SFPs, 2 CHPIDs)
  - **FICON Express8S** (SX and LX, 2 SFPs, 2 CHPIDs)
  - **OSA-Express5S**
    - 10 GbE LR and SR (1 SFP, 1 CHPID)
    - GbE SX, LX, and 1000BASE-T (2 SFPs, 1 CHPID)
  - **10 GbE RoCE Express** (2 supported SR ports)
  - **zEDC Express**
  - **Crypto Express5S**
  - **Flash Express**

PCIe I/O drawer

32 I/O slots

- **Integrated Coupling Adaptor (ICA) Fanout**
  - PCIe-O SR two 8 Gbps PCIe Gen3 Coupling Link
- **InfiniBand Coupling Feature Fanouts**
  - HCA3-O two 12x 6GBps InfiniBand DDR Coupling Links
  - HCA3-O LR four 1x 5Gbps InfiniBand DDR or SDR Coupling Links

z13 "Carry Forward" I/O Features Supported

Carry Forward Features

- **Features – PCIe I/O drawer**
  - FICON Express8S (SX and LX, 2 SFPs, 2 CHPIDs)
  - OSA-Express5S (All)
  - OSA-Express4S (All)
  - 10 GbE RoCE Express (Both ports supported on z13)
  - zEDC Express
  - Flash Express
  - Not Supported – Crypto Express4S

PCIe I/O drawer 32 I/O slots

- **Features – I/O drawer (No MES adds)**
  - FICON Express8 (SX and LX, 4 SFPs, 4 CHPIDs)
  - Not Supported: ESCON, FICON Express4, OSA-Express3, ISC-3, and Crypto Express3

I/O drawer 8 I/O slots

- **InfiniBand Coupling Features (Fanouts)**
  - HCA3-O two 12x 6GBps InfiniBand DDR Coupling Links
  - HCA3-O LR four 1x 5Gbps InfiniBand DDR or SDR Coupling Links
  - NOT Supported: HCA2-O 12x, HCA2-O LR 1x InfiniBand Coupling Links

Legacy I/O Drawers - FC 4008

To keep or not to keep, that is the question ?

Can limit future I/O growth - Should not be used for configurations that require high I/O connectivity or high coupling using HCA3-O (PSIFB) links

No RPQ's, cannot add or fill drawers once they are carried forward

Disadvantages:

- Reduced I/O counts
  - With two legacy I/O drawers (16 slots) installed, limits PCIe drawer plugging or lost space for 2 PCIe drawers (lost 64 slots). Carrying forward only a couple of cards is even worse as you are using up valuable fanout space !
- Reduced PSIFB links
  - Legacy I/O drawers requires two HCA2-C fanouts which only leaves 2 free adapter slots for use with PSIFB links.
- Carry forward for FICON Express8 cards only, other I/O card types not supported
- Limits partition memory to 1TB
- If contending for adapter space with IFB links, could drive the need to add additional CPC drawer resulting in more cost then what is saved by carrying forward older FICON Express8 cards

FICON Convertor Support

- IBM Maintenance and Support
  - North America: IBM is contracted for onsite support
  - EMEA/AP/LA: IBM and Optica partners are contracted for onsite support
  - Japan: Optica partners are contracted for onsite support

Sean Seitz
Optica Technical Services
sean.seitz@opticatech.com
+1-720-214-2800 ext 811

Partnership

Old → **IBM 9034 (ESCON to Parallel)** → **Optica 34600 FXBT (ESCON to Parallel)** → **Optica PRISM (FICON to ESCON)** → **Current**

PRIZM is available from IBM GTS Site & Facilities as part of the EFM Service (ESCON to FICON Migration - offering # 6948-97D)

- http://www.ibm.com/services/us/index.wss/itservice/igs/a1026000?cm_re=masthead-_-itservices-_-site
- The order process for PRIZM is the same as it is for IBM cabling systems

PRIZM Hardware Versions

- All PRIZM hardware platforms are 2U in height, 19 inch rack mount servers
- All PRIZM platforms are functionally equivalent and backward compatible (i.e. Gen-3 can replace a Gen-2 or 1; Gen-2 can replace a Gen-1)

FICON interface specifications and Serial Number Schemes:

Generation-1:

- 2Gb optics / supports 8, 4, 2 and 1Gb channels
- SN: 12 character format beginning with “BZDR” (i.e. BZDR12345678)

Generation-2:

- 8Gb optics / strapped to 4Gb / supports 16, 8, 4 and 2Gb channels
- SN: “R6:” followed by 12 numbers w/ 7 leading 0’s (i.e. R6:000000012345)

Generation-3:

- 8Gb optics / strapped to 4Gb / supports 16, 8, 4 and 2Gb channels
- SN: 13 character scheme with dashes (i.e. AC1-000-KQ2BKK7). First 6 characters are often dropped, leaving the last 7 digits such as “KQ2BKK7”
  - Note: Optica will work with PRIZM customers who have maintenance coverage to help them manage the transition to the16S channel - techsupport@opticatech.com

3rd Generation PRIZM

Last 7 digits of S/N

Solid State Drives in RAID1 Configuration

Power

ID

Front View

2U

Pull-out Status Panel

FICON LC Duplex Ports

ESCON Harness Ports

Rear View

B0

B1

0

1

1

2

550W

550W

Power Supplies

NIC Ports

IMM Module / KVM Port

ID Light

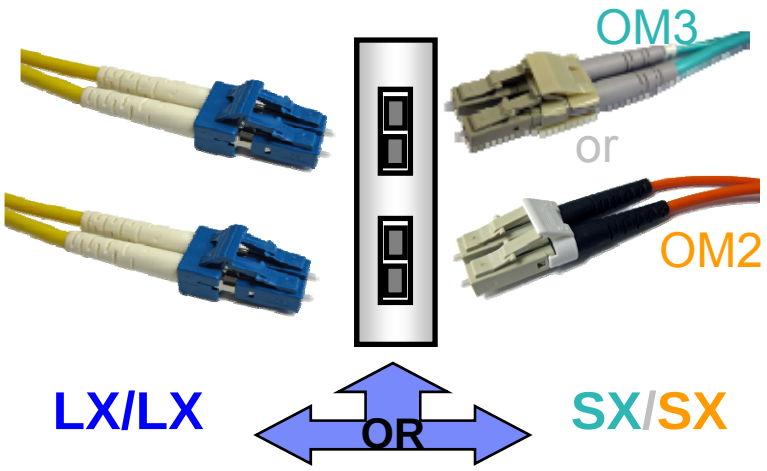
Note: Previous PRIZM generations slides can be found within the additional presentation materials

z13
Connectivity for Storage

FICON Express16S – SX and 10KM

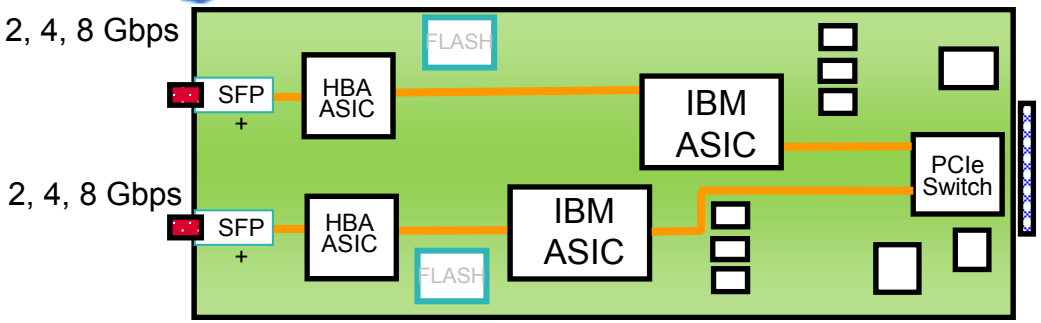
- For FICON, zHPF, and FCP environments
  - CHPID types: FC and FCP
    - 2 PCHIDs/CHPIDs
- Auto-negotiates to 4, 8, or 16 Gbps
  - 2Gbps connectivity NOT supported
  - FICON Express8S will be available to order for 2Gbps connectivity
- Increased I/O Devices (subchannels) per channel for all FICON features:
  - TYPE=FC: Increased from 24k to 32k to support more base and alias devices
- Increased bandwidth compared to FICON Express8S
- 10KM LX - 9 micron single mode fiber
  - Unrepeated distance - 10 kilometers (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)
- Small form factor pluggable (SFP) optics
  - Concurrent repair/replace action for each SFP

FC 0409 – 10KM LX, FC 0410 – SX

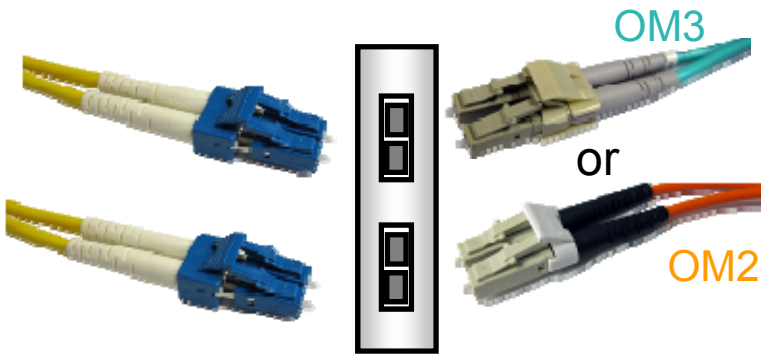


FICON Express8S – SX and 10KM LX in the PCIe I/O drawer

- For FICON, zHPF, and FCP environments
  - CHPID types: FC and FCP
    - 2 PCHIDs/CHPIDs
- Auto-negotiates to 2, 4, or 8 Gbps
- Increased I/O Devices (subchannels) per channel for all FICON features:
  - TYPE=FC: Increased from 24k to 32k to support more base and alias devices
- Increased performance compared to FICON Express8
- 10KM LX - 9 micron single mode fiber
  - Unrepeated distance
    - 10 kilometers (6.2 miles)
  - Receiving device must also be LX
  - Note: Only LX has FQC support
- 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)
- Small form factor pluggable (SFP) optics
  - Concurrent repair/replace action for each SFP



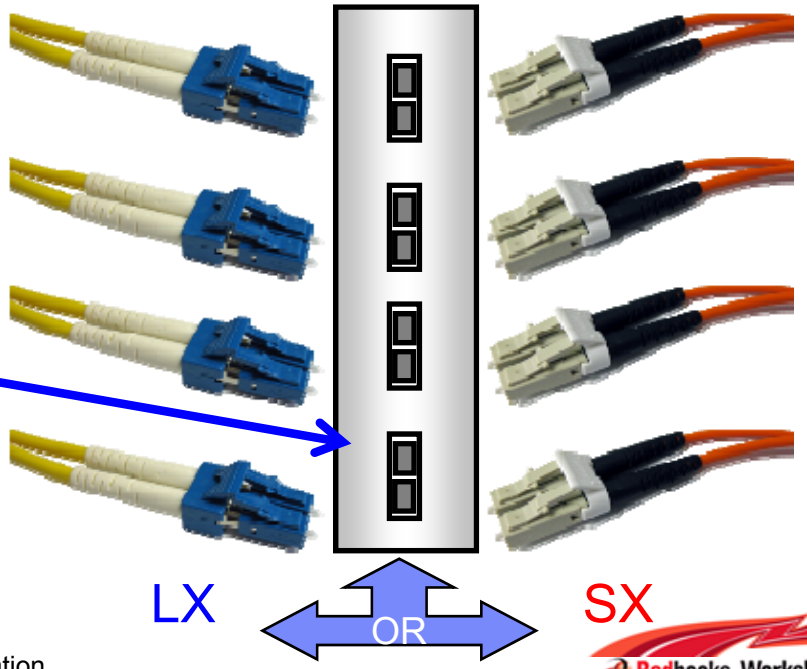
FC 0409 – 10KM LX, FC 0410 – SX



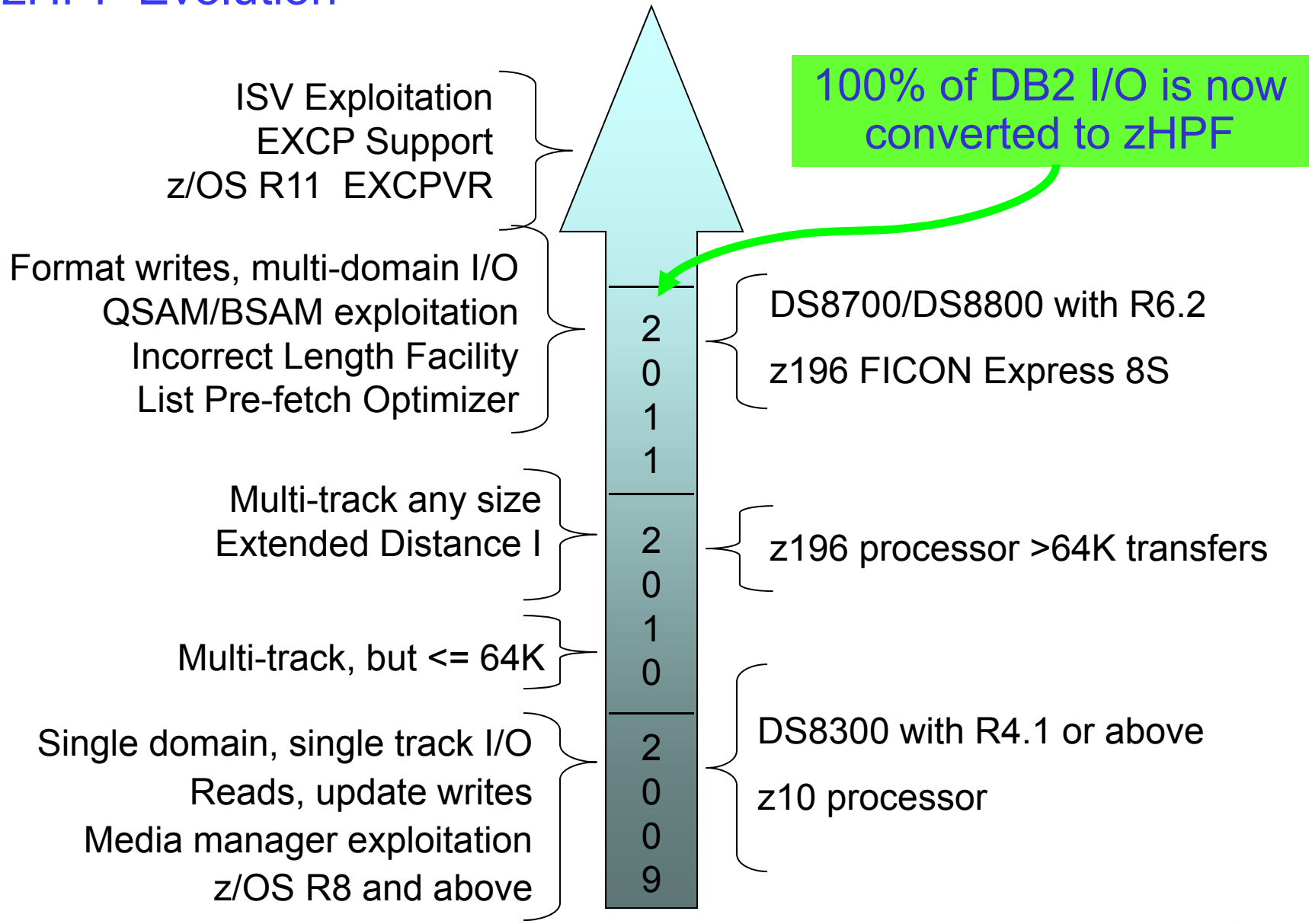
FICON Express8 (carry forward only)

- Auto-negotiate to 2, 4, or 8 Gbps
  1 Gbps devices not supported point-to-point
- Increased I/O Devices (subchannels) per channel for all FICON features:
  - TYPE=FC: Increased from 24k to 32k to support more base and alias devices
- Connector - LC Duplex
- Four LX ports (FC 3325)
  - 9 micron single mode fiber
  - Unrepeated distance - 10 km (6.2 miles)
  - Receiving device must also be LX
- Four SX ports (FC 3326)
  - 50 or 62.5 micron multimode fiber
    (50 micron fiber is preferred)
  - Unrepeated distance varies fiber type and link data rate
  - Receiving device must also be SX
- Small Form Factor Pluggable (SFP) optics
  Concurrent repair/replace action for each SFP
- LX and SX performance is identical
- Additional buffer credits supplied by a director or DWDM are required to sustain performance beyond 10 km

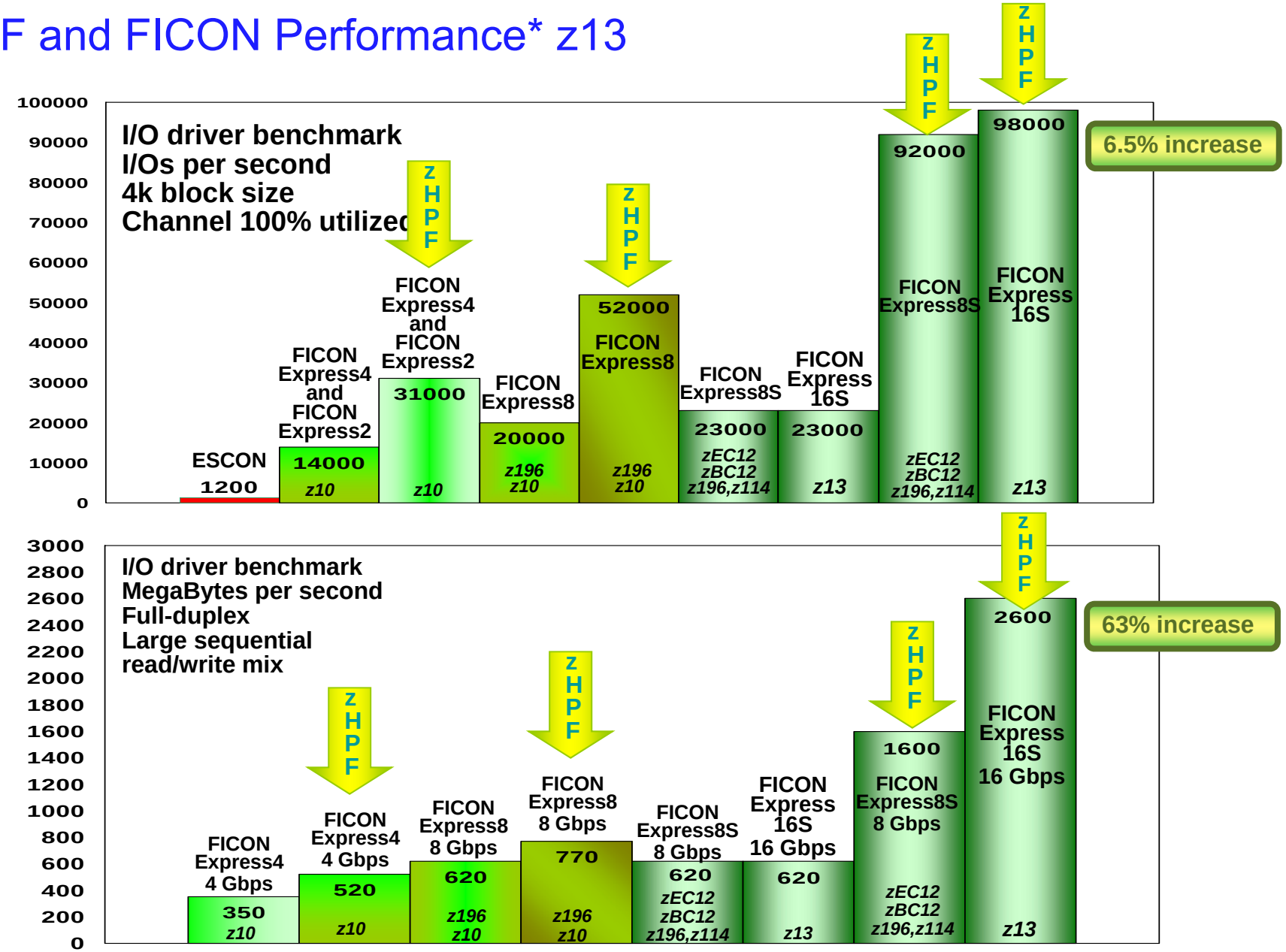
FC 3325 – 10KM LX, FC 3326 – SX



zHPF Evolution

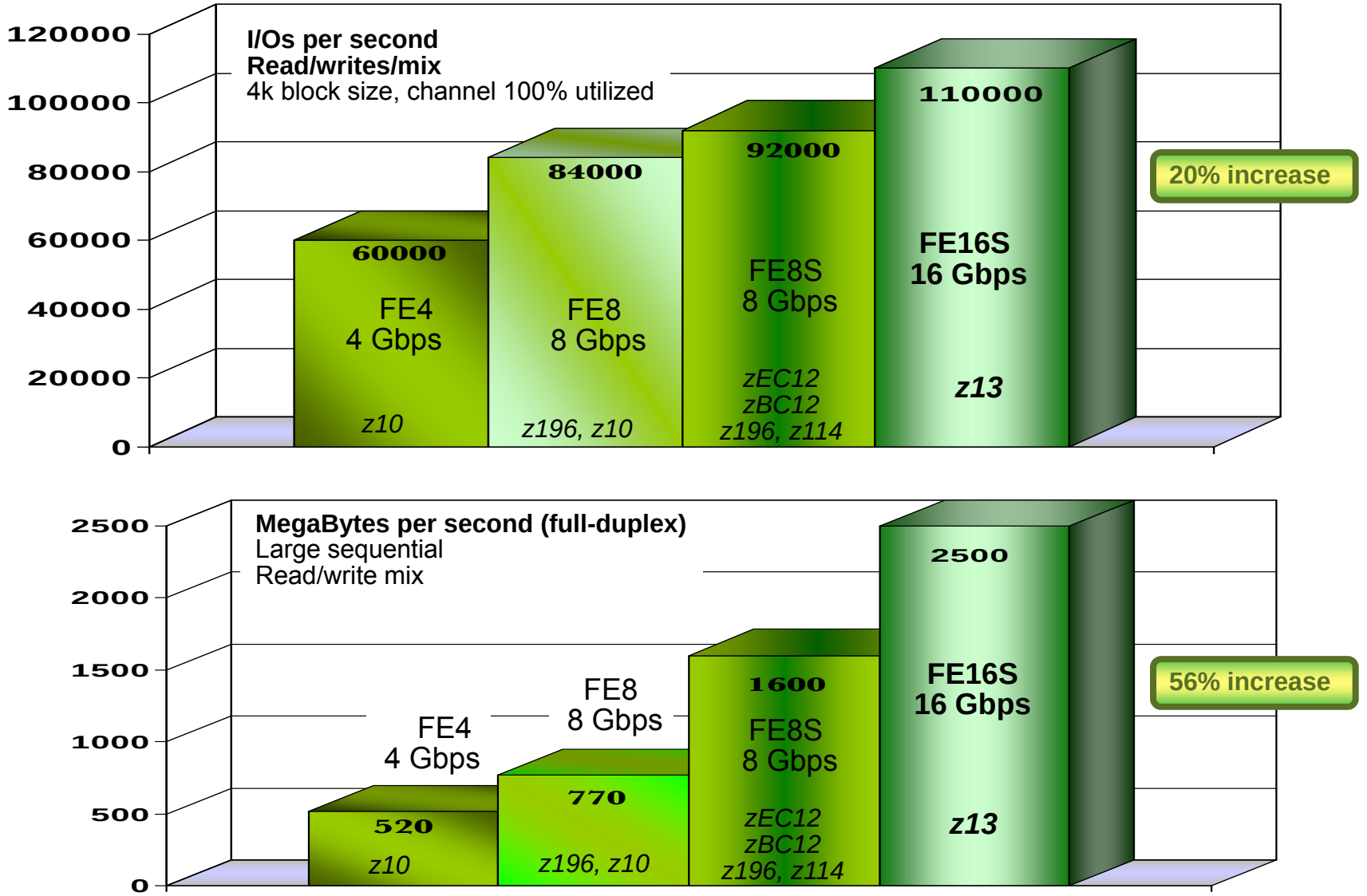


zHPF and FICON Performance* z13



*This performance data was measured in a controlled environment running an I/O driver program under z/OS. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed.

FCP Performance* for z13



*This performance data was measured in a controlled environment running an I/O driver program under z/OS. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed.

FICON Express8S vs FICON Express16S

Common Features Supported

- For FICON, zHPF, and FCP environments
  - CHPID types: FC and FCP
    - 2 PCHIDs/CHPIDs
- 10KM LX - 9 micron single mode fiber
  - Unrepeated distance - 10 kilometers (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)
- Small form factor pluggable (SFP) optics
  - Concurrent repair/replace action for each SFP

Differences

- FICON Express16S auto-negotiates to 4, 8, or 16 Gbps
  - 1 & 2Gbps connectivity NOT supported
  - Increased bandwidth compared to FICON Express8S
- FICON Express8S auto-negotiates to 2, 4, or 8 Gbps
  - 1Gbps & 16Gpbs connectivity NOT supported

Recommended use of FICON Express8S cards should be to support attachment to 2Gpbs devices, and connections to older I/O optic's that are not certified for 16Gpbs.

z13 Storage Connectivity Options

Description	F/C	Ports	Available	Comments
FICON Express16S 10KM LX	0418	2	New	z13 Only
FICON Express16S SX	0419	2	New	z13 only
FICON Express8S 10KM LX	0409	2	New and carry forward	Carry forward from zEC12 / z196
FICON Express8S SX	0410	2	New and carry forward	Carry forward from zEC12 / z196
FICON Express8 10KM LX	3325	4	Carry Forward only	Carry forward from zEC12 / z196
FICON Express8 SX	3326	4	Carry Forward only	Carry forward from zEC12 / z196

Maximum FICON features varies with mix of Drawers types and Model of the System

All use LC Duplex connectors

Fibre Channel Physical Interface standard

- Applies to FICON Express16S (4, 8, 16 Gps) FICON Express8S and FICON Express8 (2, 4, 8 Gbps), FICON Express4 (1, 2, 4 Gbps) FICON Express2 (1, 2 Gbps), and FICON Express (1, 2 Gbps) features
- CHPID types FC (FICON, zHPF, CTC) and FCP (Fibre Channel Protocol)
- Unrepeated distances in kilometers (km), meters (m), and feet (ft)

	1 Gbps		2 Gbps		4 Gbps		8 Gbps		16 Gbps		10 Gbps ISLs	
Fiber Core (µ) Light source	**Distance meters feet**	*** Link loss budget**	**Distance meters feet**	*** Link loss budget**	**Distance meters feet**	*** Link loss budget**	**Distance meters feet**	*** Link loss budget**	**Distance meters feet**	*** Link loss budget**	**Distance meters feet**	*** Link loss budget**
9µ SM LX laser	10 km 6.2 miles	7.8 dB	10 km 6.2 miles	7.8 dB	10 km 6.2 miles	7.8 dB	10 km 6.2 miles	6.4 dB	10 km 6.2 miles	6.4 dB	10 km 6.2 miles	6.0 dB
9µ SM LX laser	4 km # 2.5 miles	4.8 dB #	4 km # 2.5 miles	4.8 dB #	4 km # 2.5 miles	4.8 dB #	N/A	N/A	N/A	N/A	N/A	N/A
50µ MM OM3 2000 MHz-km SX laser	860 m 2822 ft	4.62 dB	500 m 1640 ft	3.31 dB	380 m 1247 ft	2.88 dB	150 m 492 ft	2.04 dB	100 m 328 feet	1.86 dB	300 m 984 ft	2.6 dB
50µ MM OM2 500 MHz-km SX laser	500 m 1640 ft	3.85 dB	300 m 984 ft	2.62 dB	150 m 492 ft	2.06dB	50 m 164 ft	1.68 dB	35 m 115 feet	1.63 dB	82 m 269 ft	2.3 dB
62.5µ MM OM1 200 MHz-km SX laser	300 m 984 ft	3.00 dB	150 m 492 ft	2.10 dB	70 m 230 ft	1.78 dB	21 m 69 ft	1.58 dB	15 m 49 feet	1.56 dB	33 m 108 ft	2.4 dB

Inter-Switch Links (ISLs) is the link between two FICON directors; FICON features do not operate at 10 Gbps

* The link loss budget is the channel insertion loss as defined by the standard.

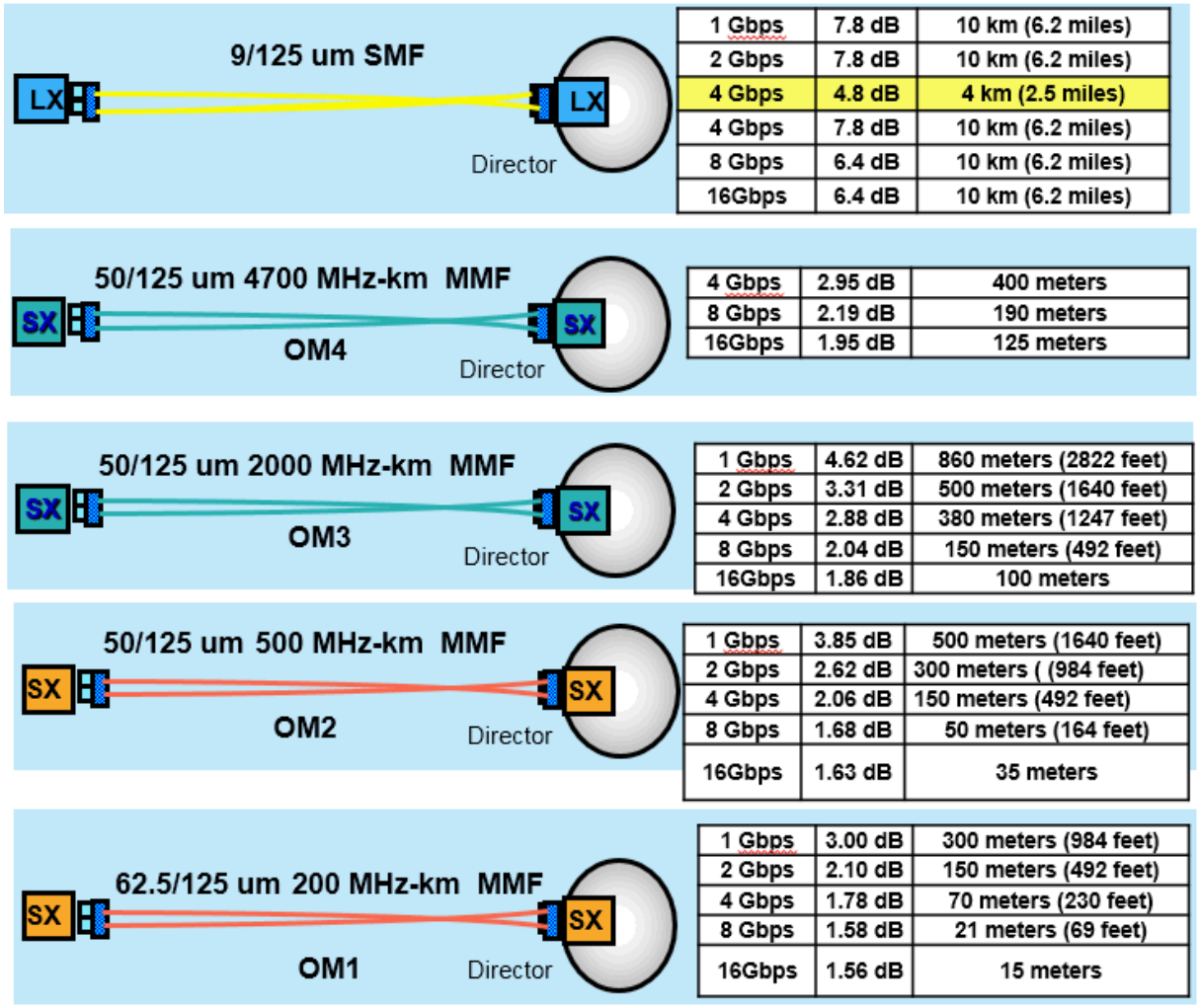
\# This distance and dB budget applies to FICON Express4 4KM LX features

FICON Optical Lengths

There are several factors that limit distance for a FICON connection, DB loss, number of connections, cable type and speed all impact distance. Customers that plan on ordering and using FICON Express16S SX connections should plan out their installations carefully !

Table list various speeds, FICON Express16S only supports 4, 8 and 16 Gbps

LX = Long wavelength 1310 nm transceiver SX - Short wavelength 850 nm



Link Speed	Link Budget	Unrepeated Distance
1 Gbps	7.8 dB	10 km (6.2 miles)
2 Gbps	7.8 dB	10 km (6.2 miles)
4 Gbps	4.8 dB	4 km (2.5 miles)
4 Gbps	7.8 dB	10 km (6.2 miles)
8 Gbps	6.4 dB	10 km (6.2 miles)
16Gbps	6.4 dB	10 km (6.2 miles)
4 Gbps	2.95 dB	400 meters
8 Gbps	2.19 dB	190 meters
16Gbps	1.95 dB	125 meters
1 Gbps	4.62 dB	860 meters (2822 feet)
2 Gbps	3.31 dB	500 meters (1640 feet)
4 Gbps	2.88 dB	380 meters (1247 feet)
8 Gbps	2.04 dB	150 meters (492 feet)
16Gbps	1.86 dB	100 meters
1 Gbps	3.85 dB	500 meters (1640 feet)
2 Gbps	2.62 dB	300 meters ((984 feet)
4 Gbps	2.06 dB	150 meters (492 feet)
8 Gbps	1.68 dB	50 meters (164 feet)
16Gbps	1.63 dB	35 meters
1 Gbps	3.00 dB	300 meters (984 feet)
2 Gbps	2.10 dB	150 meters (492 feet)
4 Gbps	1.78 dB	70 meters (230 feet)
8 Gbps	1.58 dB	21 meters (69 feet)
16Gbps	1.56 dB	15 meters

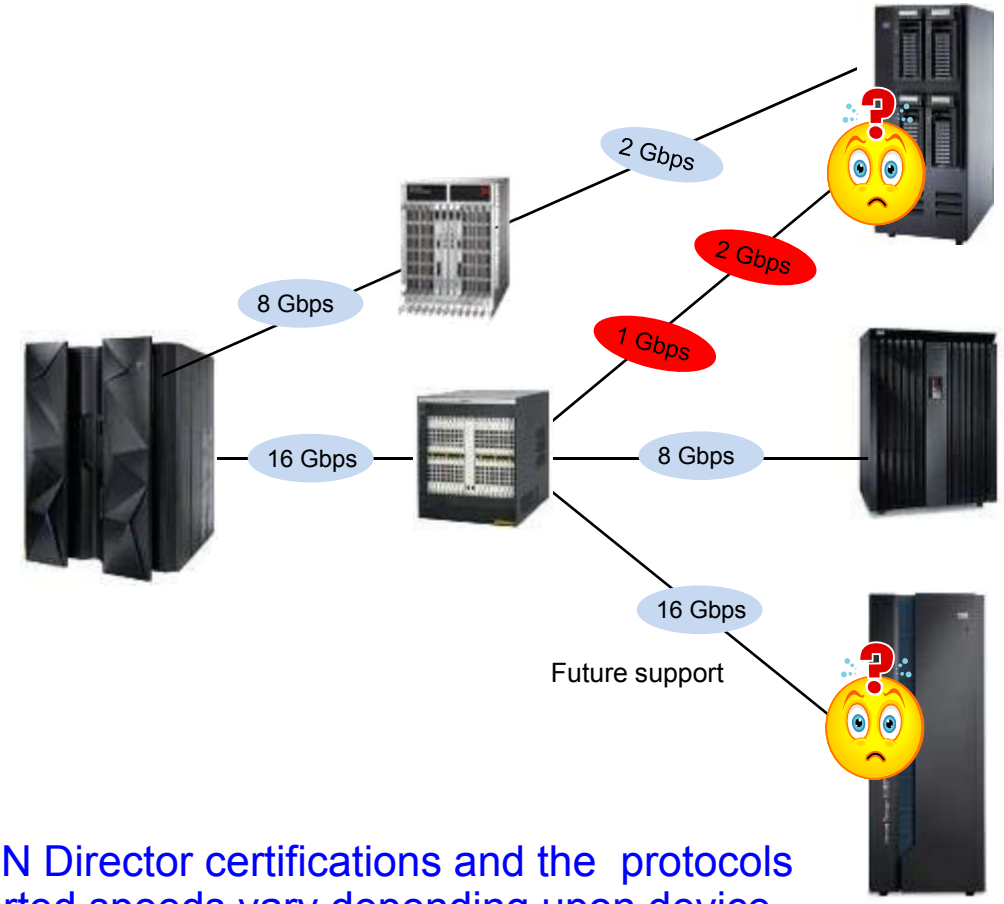
FICON Director Considerations

It is extremely important to understand the planned use of FICON Directors and to review I/O connectivity requirements to determine the correct mix of FICON Cards for your configuration.

Check certification list on ResourceLink

- Will older directors be certified for 16 Gpbs ?
- What speeds will newer directors support ?

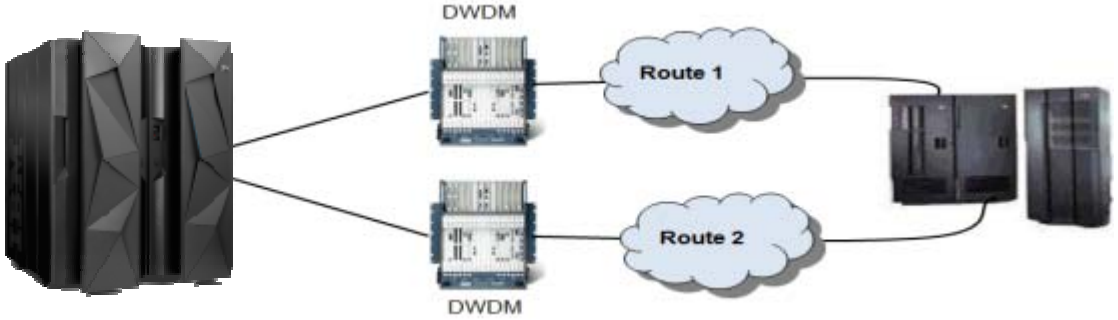
Customers may need to continue to use FICON Express8 or Express 8S cards and/or also maintain different set of directors that support lower speeds. This is true especially when transitioning from older equipment to new equipment where the older devices will need to be maintained for a period of time.



Check IBM ResourceLink for current FICON Director certifications and the protocols and line speeds that are supported. Supported speeds vary depending upon device model and microcode installed.

DWDM Considerations

IBM will only support WDM products qualified by IBM z Systems for usage in GDPS solutions. To obtain this qualification, WDM vendors obtain licensed IBM patents, knowledge, and intellectual property related to the GDPS architecture.



No Support for ICA SR

For GDPS environments with DWDM always check the vendor certification list that is available under IBM Resourcelink when adding new link or channel types.

Advanced planning for updates to DWDM equipment can involve:

1. Current device compatible with features being ordered, FICON, OSA, Links, doesn't support ICA SR ?
2. Use of links and DWDM on z13 requires Infiniband 1x Coupling Links, validate certifications?
3. DWDM management Vendors can / will charge a fee to upgrade microcode
   - Upgrade maybe concurrent but may/will need to be scheduled during maint window?
   - Consider time for ordering and vendor resource availability for installation
4. DWDM capacity planning and feature ordering, need to add shelves, or supporting cards?

IBM Resourcelink ➔ Library ➔ z Systems Qualified Wavelength Division Multiplexer (WDM) products for GDPS solutions
https://www.ibm.com/servers/resourcelink/lib03020.nsf/pages/systemzQualifiedWdmProductsForGdpsSolutions?OpenDocument&pathID=

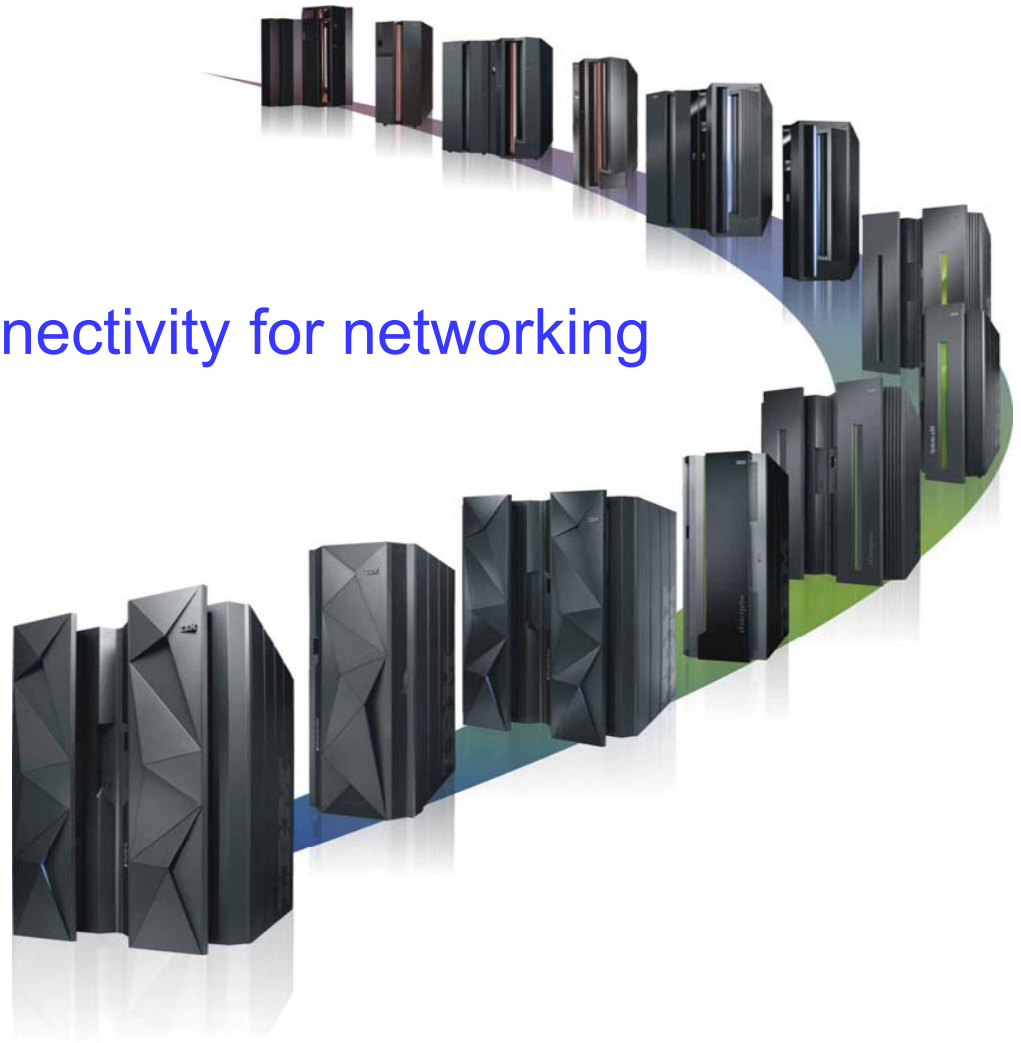
New FICON Function for z13

- 16 Gbps Link Speeds (March 9, 2015)
  - Designed to reduce I/O latency to improve response time for performance-critical middleware and to shrink the batch window required to accommodate I/O bound batch work
- 6th Logical Channel Subsystem (March 9, 2015)
  - Up to 85 Logical Partitions: More flexibility for server consolidation
- 4th Subchannel Set (March 9, 2015)
  - Simplifies I/O configurations for a 2nd synchronous copy of data
  - With multi-target PPRC, can do HyperSwap and still maintain synchronous copy for 2nd HyperSwap
- Preserve Virtual WWPNs for NPIV configured FCP channels
  - Designed to simplify migration to a new-build z13 (March 9, 2015)
- 32K devices per FICON channel (March 9, 2015)
  - Up to 85 Logical Partitions: More flexibility for server consolidation
- zHPF Extended I/O execution at Distance (June 26, 2015)
  - Up to 50% I/O service time improvement for remote write
  - Designed to help GDPS HyperSwap configurations with secondary DASD in remote site
- FICON Dynamic Routing (September 25, 2015)
  - Designed to allow ISL sharing by FC and FCP traffic to optimize use of ISL bandwidth in the SAN fabric for both types of traffic
- Forward Error Correction Codes (September 25, 2015)
  - Designed to addresses high bit-error rate on high frequency (>= 8Gb/s) links
  - Estimated equivalence to doubling optical signal power
- SAN Fabric I/O Priority (September 25, 2015)
  - Extends z/OS WLM policy into the SAN fabric
  - Gives important work priority to get through SAN traffic congestion (e.g. after SAN hardware failures)

IBM z Systems I/O Exerciser Tool

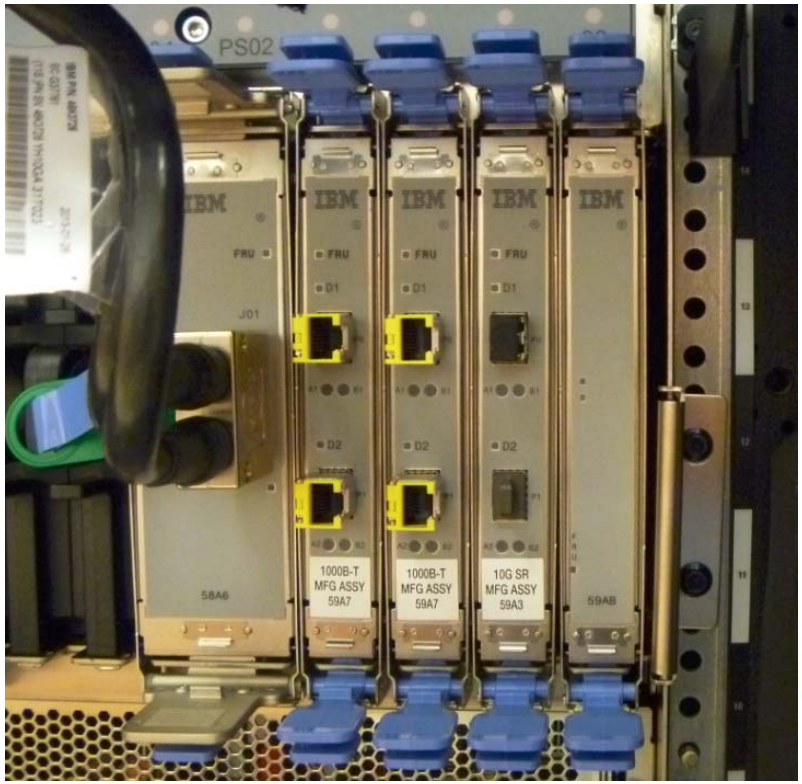
- **Problem**
  - After upgrades (processor or storage) FICON connections are not found to be faulty until the production z/OS work load is run
- **Solution**
  - Provide a way to verify quality of the cable connections before running z/OS production work
- **IBM z Systems I/O Exerciser**
  - *https://www.ibm.com/services/forms/preLogin.do?source=swg-beta-ibmioexzos*
    New tool made available March 4, 2014
  - Runs in a stand-alone LPAR or z/VM Guest Machine
  - Tests all the FICON devices available to that partition via the IOCDS

Connectivity for networking



OSA-Express5S, 1000BASE-T, GbE and 10 GbE SFPs

- Common OSA-Express5S feature using the following SFPs:
  - 1000BASE-T, two ports
  - GbE, two ports, both either SX or LX
  - 10 GbE, one port, either SX or LX. Port D2 not used
- Same performance as OSA-Express4S



OSA-Express5S 1000BASE-T Ethernet Feature - PCIe I/O Drawer

- PCI-e form factor feature supported by PCIe I/O drawer
  - One two-port CHPID per feature
    Half the density of the OSA-Express3 version
- Small form factor pluggable (SFP+) transceivers
  - Concurrent repair/replace action for each SFP
- Exclusively Supports: Auto-negotiation to 100 or 1000 Mbps and full duplex only on Category 5 or better copper
- RJ-45 connector
- Operates at “line speed”
- CHPID TYPE Support:

FC 0417

Connector = RJ-45

Mode	TYPE	Description
OSA-ICC	OSC	TN3270E, non-SNA DFT, OS system console operations
QDIO	OSD	TCP/IP traffic when Layer 3, Protocol-independent when Layer 2
Non-QDIO	OSE	TCP/IP and/or SNA/APPN/HPR traffic
Unified Resource Manager	OSM	Connectivity to intranode management network (INMN)
OSA for NCP (LP-to-LP)	OSN	NCPs running under IBM Communication Controller for Linux (CCL)

Note: OSA-Express5S feature are designed to have the same performance and to require the same software support as equivalent OSA-Express4S features.

OSA-Express5S Fiber Optic Features – PCIe Drawer

- **10 Gigabit Ethernet (10 GbE)**
  - CHPID types: OSD, OSX
  - Single mode (LR) or multimode (SR) fiber
  - One port of LR or one port of SR
    - 1 PCHID/CHPID
  - Small form factor pluggable (SFP+) optics
    - Concurrent repair/replace action for each SFP
  - LC duplex
- **Gigabit Ethernet (1 GbE)**
  - CHPID types: OSD (OSN not supported)
  - Single mode (LX) or multimode (SX) fiber
  - Two ports of LX or two ports of SX
    - 1 PCHID/CHPID
  - Small form factor pluggable (SFP+) optics
    - Concurrent repair/replace action for each SFP
  - LC Duplex

Note: OSA-Express5S features are designed to have the same performance and to require the same software support as equivalent OSA-Express4S features.

FC 0415 – 10 GbE LR, FC 0416 – 10 GbE SR

FC 0413 – GbE LX, FC 0414 – GbE SX

Open Systems Adapter in the PCIe I/O drawer

Description	Feature Code	Ports	Available	CHPID
OSA-Express4S GbE LX	0404	2[1]	Carry Forward	OSD
OSA-Express4S GbE SX	0405	2[1]	Carry Forward	OSD
OSA-Express4S 10 GbE LR	0406	1	Carry Forward	OSD, OSX
OSA-Express4S 10 GbE SR	0407	1	Carry Forward	OSD, OSX
OSA-Express4S 1000BASE-T	0408	2[1]	Carry Forward	OSC, OSD, OSE, OSM, OSN

Description	Feature Code	Ports	Available	CHPID
OSA-Express5S GbE LX	0413	2[1]	New and Carry Forward	OSD
OSA-Express5S GbE SX	0414	2[1]	New and Carry Forward	OSD
OSA-Express5S 10 GbE LR	0415	1	New and Carry Forward	OSD, OSX
OSA-Express5S 10 GbE SR	0416	1	New and Carry Forward	OSD, OSX
OSA-Express5S 1000BASE-T	0417	2[1]	New and Carry Forward	OSC, OSD, OSE, OSM, OSN

[1] Two ports per CHPID

Summary: OSA-Express CHPID types to control operation

CHPID type	Purpose / Traffic	Operating Systems
OSC **1000BASE-T** **z13, zEC12, zBC12, z196, z114, z10, z9**	OSA-Integrated Console Controller (OSA-ICC) Supports TN3270E, non-SNA DFT to IPL CPCs & LPs	z/OS, z/VM z/VSE
OSD **All OSA features** **z13, zEC12, zBC12, z196, z114, z10, z9**	Supports Queue Direct Input/Output (QDIO) architecture TCP/IP traffic when Layer 3 (uses IP address) Protocol-independent when Layer 2 (uses MAC address)	z/OS, z/VM z/VSE, z/TPF Linux on z Systems
OSE **1000BASE-T** **z13, zEC12, zBC12, z196, z114, z10, z9**	Non-QDIO; for SNA/APPN/HPR traffic and TCP/IP “passthru” traffic	z/OS, z/VM z/VSE
OSM **1000BASE-T** **z13, zEC12, zBC12, z196, z114**	OSA-Express for Unified Resource Manager Connectivity to intranode management network (INMN) from z13, zEC12, z196, or z114 to Unified Resource Manager functions	z/OS, z/VM* Linux on z Systems
OSN **GbE, 1000BASE-T** **z13, zEC12, zBC12, z196, z114, z10, z9** **No OSN support for OSA-Express4S and 5S GbE**	OSA-Express for NCP Appears to OS as a device supporting CDLC protocol Enables Network Control Program (NCP) channel-related functions Provides LP-to-LP connectivity OS to IBM Communication Controller for Linux (CCL)	z/OS, z/VM z/VSE, z/TPF Linux on z Systems
OSX **10 GbE** **z13, zEC12, zBC12, z196, z114**	OSA-Express for zBX Connectivity and access control to intraensemble data network (IEDN) from z13, zEC12, z196, or z114 to zBX	z/OS, z/VM*, z/VSE 5.1, Linux on z Systems

*CHPIDs OSX and OSM supported by z/VM V6.2 and z/VM V6.3 to define, modify, and delete OSX CHPID types when z/VM is the controlling LPAR for dynamic I/O

10 Gigabit Ethernet cabling options

Channel insertion loss + additional insertion loss allowed as defined by the IEEE 802.3 standard

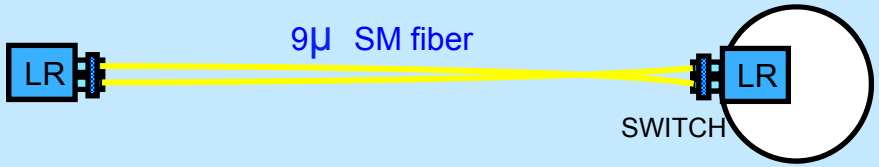
Single mode (SM)

(OM3)

(OM2, OM1)

LR = Long reach 1310 nm transceiver

SR - Short reach 850 nm transceiver



- LR transceiver/feature on each end

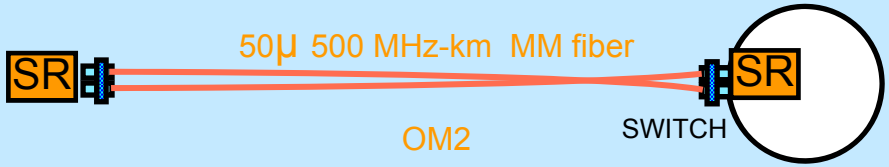
10 Gbps	6.0 dB	10 km (6.2 miles)
Link Speed	Link Budget	Unrepeated Distance

Single mode (SM) fiber

- SR transceiver/feature on each end

10 Gbps	2.6 dB	300 meters (984 feet)
Link Speed	Link Budget	Unrepeated Distance

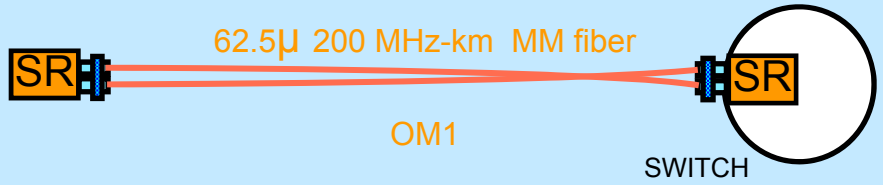
Multimode (MM) fiber



- SR transceiver/feature on each end

10 Gbps	1.8 dB	82 meters (269 feet)
Link Speed	Link Budget	Unrepeated Distance

Multimode (MM) fiber



- SR transceiver/feature on each end

10 Gbps	1.6 dB	33 meters (108 feet)
Link Speed	Link Budget	Unrepeated Distance

Multimode (MM) fiber

All connectors = LC Duplex

micron = µ

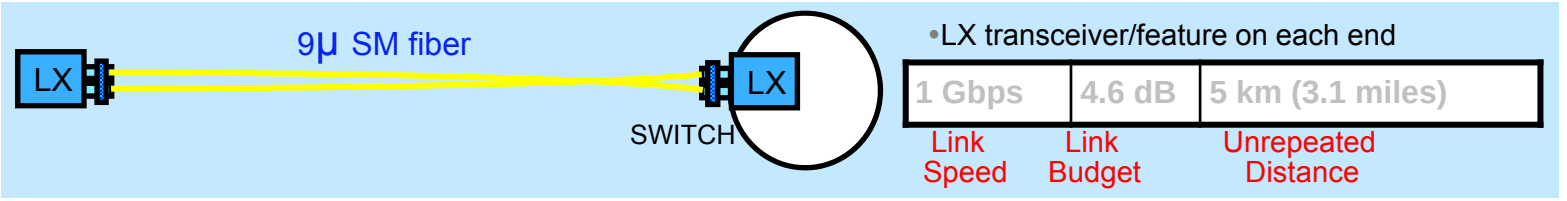
Gigabit Ethernet cabling options

Single mode (SM)

LX = Long wavelength 1310 nm transceiver

(OM2, OM1)

SX - Short wavelength 850 nm transceiver

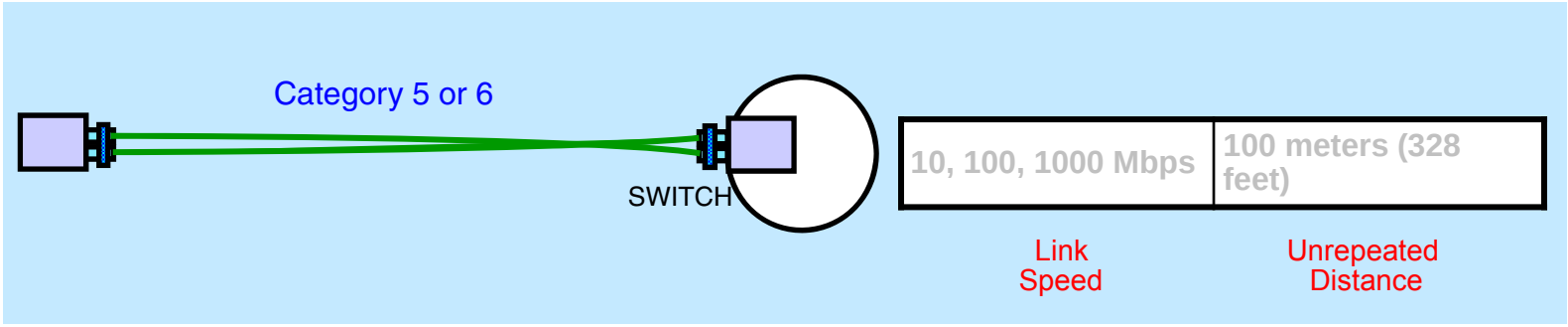


OM3 (2000 MHz-km): No changes have been made to the standard (1000BASE-SX) and no new variants. The distance remains at 550 meters for 50 um fiber; independent of whether it is OM2 or OM3.

All connectors = LC Duplex micron = µ

1000BASE-T Ethernet cabling

Cable type: EIA/TIA Category 5 or 6 Unshielded Twisted Pair (UTP) cable



Connector = RJ-45

I/O Connectivity Summary

z13 I/O Features, Channels, Ports, Domains, and Functions

Features	Offered As	Maximum # of features	Channels, Ports, Domains, Functions	Increments per Feature	Purchase increments
FICON – (Maximum of 160 features (320 channels) total only if all are FICON Express16S or 8S features.)					
FICON Express16S[1]	NB[2]	160	320[4] channels maximum	2 channels/feature	2 channels
FICON Express8S[3]	NB/CF	160	320[4] channels maximum	2 channels/feature	2 channels
FICON Express8[4]	CF	16	64 channels maximum	4 channels/feature	**CF Only**
Networking – (No more than 48 networking features total counting features of all types. One channel per feature)					
OSA-Express5S	NB	48	96[5]ports maximum	Ports: 2, 10 GbE 1	1 feature
OSA-Express4S	CF	48	96[5]ports maximum	Ports: 2, 10 GbE 1	**CF Only**
Crypto – (No more that 16 crypto features total counting features of both types.)					
Crypto Express5S[1]	NB	16	85 Domains/Adapter	1 PCIe Adapter	2, 3 - 8
Special purpose – These features provide Native PCI FUNCTIONs or Storage Class Memory (SCM)					
10GbE RoCE Express	NB	16	31 FUNCTIONs/Adapter	2 ports/Adapter	1 feature
Flash Express[1] (FC#0403)	NB	8 (4 Pairs)	1.4 TB SCM per pair	1 PCIe Adapter	2 (1 Pair)
Flash Express (FC#0402)	CF	8 (4 Pairs)	1.4 TB SCM per pair	1 PCIe Adapter	**CF Pairs Only**
zEDC Express	NB	8	15 FUNCTIONs/Adapter	1 PCIe Adapter	1 feature

Notes:
1. Bold blue text indicates new features for z13
2. NB = New Build, and if previously offered Carry Forward, CF- Carry Forward ONLY
3. FICON Express8S is offered on New Build to support point to point 2 Gbps attachment
4. Any 8-slot drawer limits maximum memory in any LPAR to 1 TB; One 8-slot drawer limits maximum FICON channels to 288, two 8-slot drawers limit maximum FICON channels to 256. (These numbers are REDUCED by 4 for each empty slot in an 8-slot drawer.)
5. 48 ports for 10 GBE

z13 and zEC12 – I/O and Coupling Migration Comparisons

Type - Model	Books or Drawers	PCIe 16 GB Fanouts	PSIFB or PCIe 8 GB Fanouts	Maximum PCIe Drawers	Maximum PSIFB Fanouts	Drawers with Maximum PSIFB	Maximum PCIe Coupling Fanouts
Notes:		1	2, 3	4, 5	3	4,5	6
2827-H20	1	N/A	8	4	8	0	N/A
2964-N30	1	10	4	2.5	4	2.5	10
2827-H43	2	N/A	16	5	16	0	N/A
2964-N63	2	20	8	5	8	5	16
2827-H66	3	N/A	20	5	16	2	N/A
2964-N96	3	30	12	5	12	5	16
2827-H89 or HA1	4	N/A	24	5	16	4	N/A
2964 NC9 or NE1	4	40	16	5	16	5	16

- Notes: These comparisons assume no 8-slot I/O drawers, no FICON Express8 carried forward
  1. PCIe 16 GBps z13 only: One port to a PCIe I/O drawer domain or two ports of PCIe Coupling
  2. PCIe 8 GBps fanout zEC12 only: Two ports to PCIe I/O drawer domains
  3. PSIFB fanout: Two ports 6 GBps 12x IFB short distance coupling or four ports 5 Gbps 1x IFB long distance coupling
  4. A fully populated PCIe I/O drawer in z13 needs four 16 GBps PCIe fanouts
  5. A fully populated PCIe I/O drawer in zEC12 needs two 8 GBps PCIe fanouts
  6. PCIe Coupling Fanout z13 only: Two ports 8 GBps PCIe short distance coupling

z13 Coupling Link Connectivity Summary

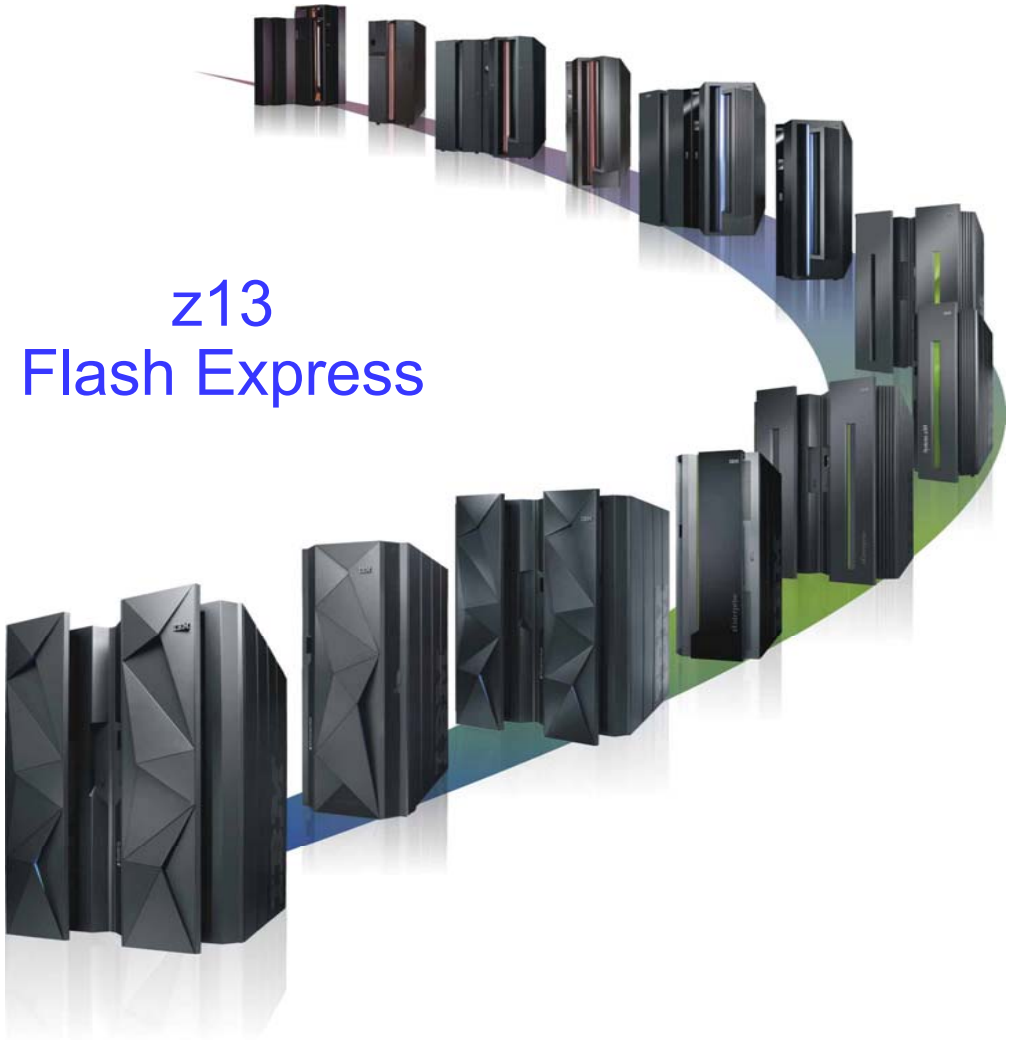
Features	Offered As	Maximum # of Features	Maximum Connections	Increments per Feature	Purchase Increments
ICA SR	NB	16	32 links[1]	2 links	2 links
HCA3-O LR (1x) [2]	NB/CF	16	64 links	4 links	4 links
HCA3-O (12x) [2]	NB/CF	16	32 links	2 links	2 links

Notes: [1] Same physical number of links as 12X PSIFB on zEC12
[2] Same physical number of links as 1X PSIFB on zEC12

NB = New build, Migration Offering, System z Exchange Program
CF = Carry Forward

Link Type	Port Qty	CHPID	Protocol	Link Data Rate	Fiber Core	Fiber Bandwidth	Fiber Type	Light Source	Cable	Connector	Maximum Distance	Repeated Distance
Short Distance												
HCA3-O fanout (12X IFB)	2	CIB	IFB	6 GBps	50 micron	2 GHz-km @ 850 nm	OM3 Multimode	SW	Duplex 12x 24-fiber cable assembly	MTP (split) TX & RX	150 meters	N/A
Integrated Coupling Adapter (ICA SR) CPC Fanout	2	CS5	PCIe Gen3	8 GBps	50 micron	4.7 GHz-km @ 850 nm	**OM4** Multimode	SW	Single 24-fiber cable assembly	MTP (new)	**150** meters	N/A
						2 GHz-km @ 850 nm	**OM3** Multimode	SW	Single 24-fiber cable assembly	MTP (new)	**100** meters	N/A
Long Distance												
HCA3-O LR fanout (1X IFB)	4	CIB	IFB	5 Gbps	9 micron	50 GHz-km @ 1310 nm	Single Mode	LW	1 fiber pair	LC Duplex	10 km	100 km

z13
Flash Express



Why Flash Express on z13?

- Provides Storage Class Memory
  - Implemented via NAND Flash SSDs (Solid State Drives) mounted in PCIe Flash Express features
  - Protected by strong AES Encryption done on the features
  - Not defined as I/O devices or with PCIe FUNCTIONs
  - Assigned to partitions similarly to Main Memory; but, not in the partition Image Profile. Reconfigurable.
  - Accessed using the new z System architected EADM (Extended Asynchronous Data Mover) Facility
  - Designed to enable extremely responsive paging of 4k pages to improve z/OS availability
  - Enables pageable large (1 MB) pages
- Flash Express Exploitation
  - z/OS V2.1, V1.13 + PTFs and RSM Enablement Offering
    - With z/OS Java SDK 7 SR3: CICS TS V5.1, WAS Liberty Profile V8.5, DB2 V11, IMS 12 and higher, SOD: Traditional WAS 8.0.0x*
    - CFCC Level 19 with WebSphere MQ for z/OS Version 7 MQ Shared Queue overflow support (March 31, 2014)
  - Linux on z Systems
    - SLES 11 SP3 and RHEL 6.4

10x

Faster response time and 37% increase in throughput compared to disk for morning transition

28%

Improvement in DB2 throughput leveraging Flash Express with Pageable Large Pages (PLP)

19%

Reduction in total dump time for a 36 GB standalone dump

~25%

Reduction in SVC dump elapsed time

*Note: All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

Flash Express Implementation Requirements - Review

Flash Express:

Prereqs:

- zEC12, zBC12 or z13, with Flash Express feature
- Flash card pair memory size is 1.4 TB
  - Min: 1 Card Pair - Max: 4 Card Pairs
  - Mirroring across pairs for resiliency
  - Typical configuration: 1 card pair

Configuration/Setup:

- Prior to Enablement; cards will be on line, unformatted
- Firmware updates delivery will be handled through z13 Bundle process
- Firmware update via concurrent patch Config off/on for enablement

+

z/OS

Prereqs:

- V1.13 + Web deliverable or V2.1 + zEC12, zBC12 and z13
- Web deliverable also contains:
  - Dynamic reconfiguration
  - 2GB Pages (not dependent on Flash)
- 4GB Real Storage needed for large pages (not specific to Flash)

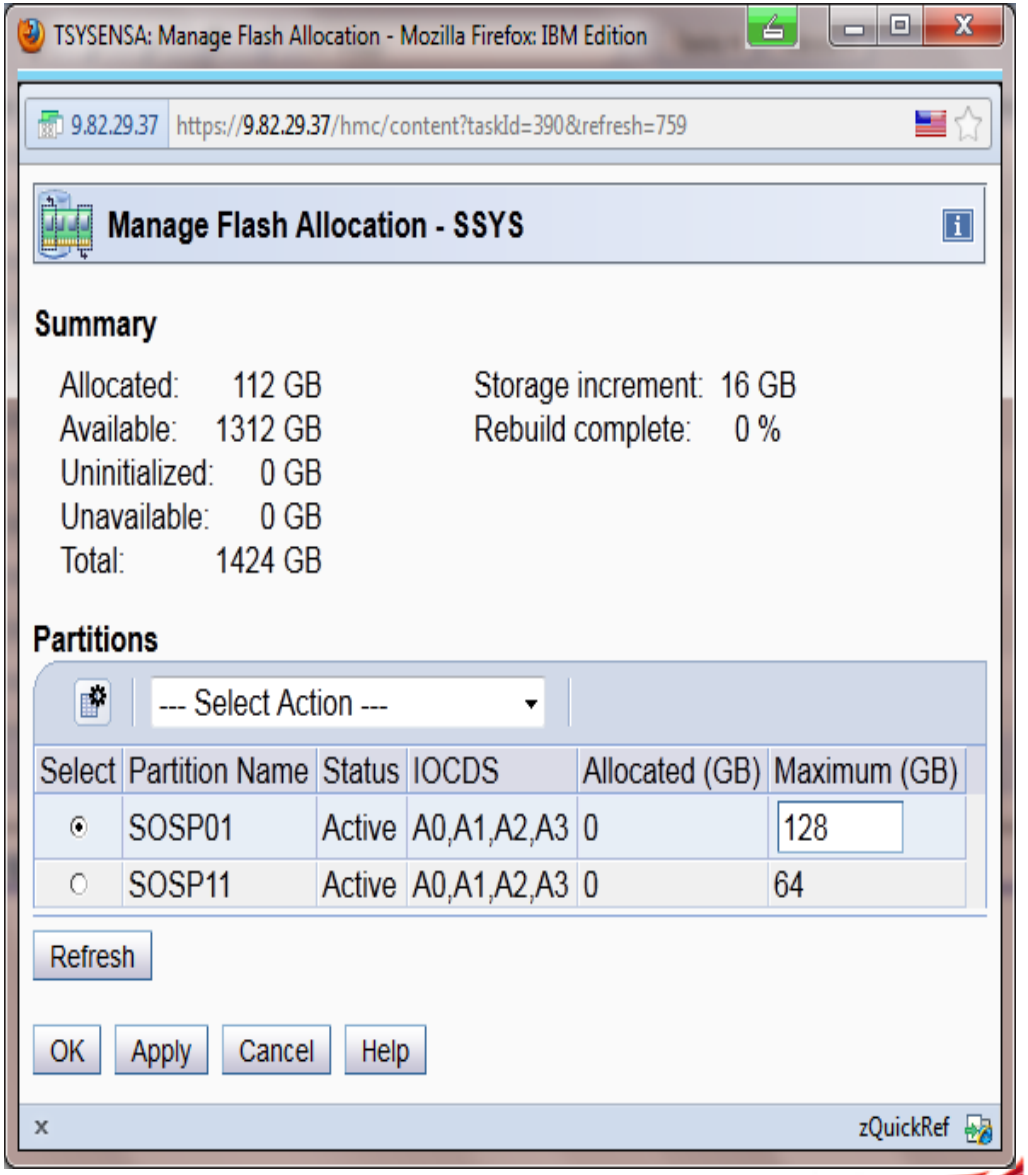
Configuration/Setup:

- At IPL, z/OS detects if flash is assigned to the partition. z/OS automatically uses Flash for paging unless specified otherwise
- All paging data can easily reside on Flash
- No need for fine grained capacity planning
- No definition required for placement of data on flash

Flash Allocation

Allocating Flash to a partition

- The *initial* and *maximum* amount of Flash Memory available to a particular logical partition is specified at the SE or HMC via a new Flash Memory Allocation panel
- Can dynamically change maximum amount of Flash Memory available to a logical partition
- Additional Flash Memory (up to the maximum allowed) can be configured online to a logical partition dynamically at the SE or HMC
  - For z/OS this can also be done via an operator command
- Can dynamically configure Flash Memory offline to a logical partition at the SE or HMC
  - For z/OS this can also be done via an operator command
  - Predefined subchannels, no IOCDS



Nothing is Changing for Flash Express for z13

- **z/OS Clients Migrating to z13 with Flash Express can improve availability and CPU performance if they are coming from systems that do not have Flash Express**
  - Customers with Flash Express are generating Minutes of Additional Up Time each week
  - Customers with Flash Express are seeing improved CPU performance that translates to Software License Fee savings

- **Clarifications..**
  - **Database:** z System uses Flash Memory in the DISK CONTROL UNIT not Flash Express to hold database data. This solution performs very well and works with all existing DR solutions. Flash Express is not used for database solutions on z/OS
  - **Client Paging Rate:** Clients with no paging see the Availability and CPU performance savings from Flash Express. No z/OS Clients should have a significant steady state paging rate. The availability value of Flash Express is on transitions. The performance value of Flash Express comes from using 1MB Pages, not from CPU savings during paging. Clients are not currently paging at any significant rate, and should not be encouraged to run at higher paging rates

Flash Express Extensions

- **CF support for Flash Express**
  - Requires z/OS V1.13 (with PTF) or above running on zEC12, zBC12 and later servers with CFLEVEL 19 or later
  - Support Flash Express for certain Coupling Facility list structures
  - Can allow keyed list structure data to be migrated to Flash Express memory
    - For example, when data consumers do not keep up with creators
    - Designed to migrate it back to real memory to be processed
  - With WebSphere MQSeries® for z/OS Version 7 (5655-R36):
    - Can buffer enterprise messaging workload spikes
    - Provide support for storing very large amounts of data in shared queue structures
    - Potentially allow several hours' worth of data to be stored without causing interruptions in processing
  - z/OS V2.1 RMF™ provides measurement data and reporting capabilities for Flash Express on Coupling Facilities
- **Available with the PTF for APAR OA40747**
  - CFSIZER also updated for Flash Express:
    - http://www.ibm.com/systems/support/z/cfsizer/

Note: New CF (coupling facility) structure attributes: SCMMAXSIZE and SCMALGORITHM can be specified in the CFRM policy to allow CF SCM (storage-class memory) to be associated with the structure.

Flash Express PCIe Adapter Card

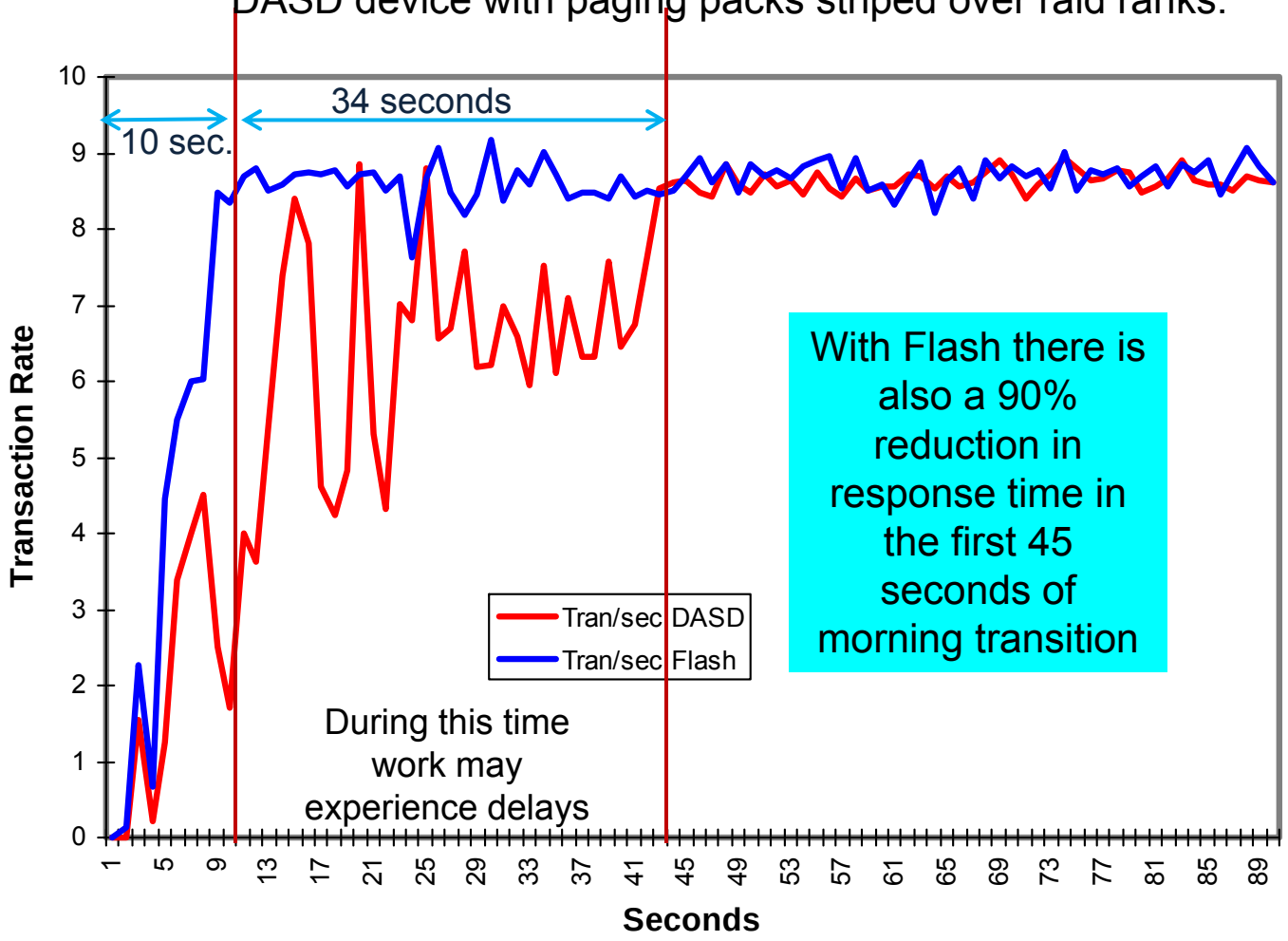
Four 400 GByte ($G=10^{9}$) SSDs support 1.4 TBytes ($T=2^{40}$) of Storage Class Memory (AES encrypted)

Cable connections to form a RAID 10 Array across a pair of Flash Express Cards.

Note: For z13, the Flash Express feature (FC 0403) is just a technology refresh of the SSDs. There are no performance or usage differences between FC 0403 and the prior FC 0402. FC 0402 will still be used during 1H2015 for certain configurations

Results with Flash Express - Morning Transition

Workload transition test was run using a dedicated DS8800 Model 2107-951 DASD device with paging packs striped over raid ranks.

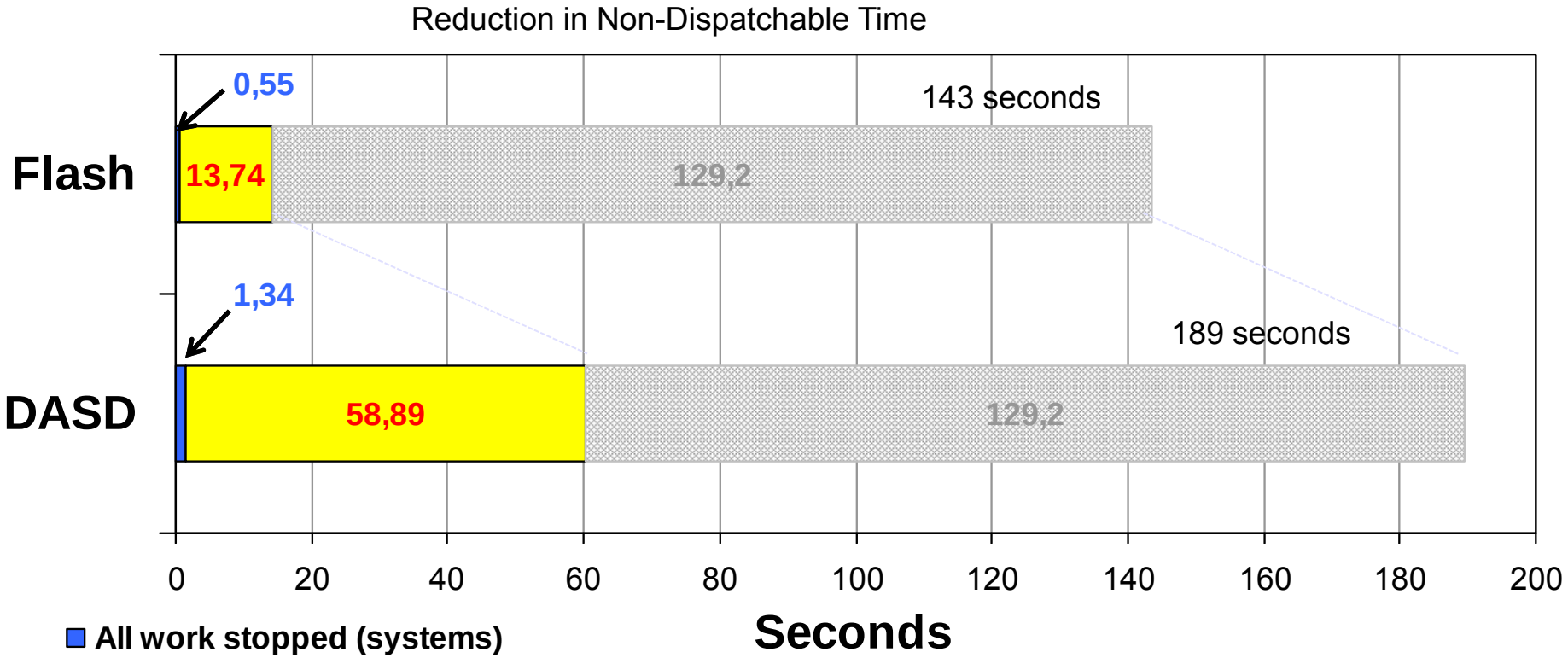


Paging to DASD required about 44 seconds for the workload to reach steady state

Paging to Flash required only 10 seconds for the workload to reach steady state

SVC Dump – Address Space and System Availability Improvements

In SVC dumps, availability was up to 4 times higher for workloads and up to twice as high for systems



Note: Gray bar indicates there is no significant time difference in writing the dump

System non-dispatchable time & address space non-dispatchable time were dramatically reduced, enabling processing that would otherwise have been stopped.
VERBX IEAVTSFS Shows total dump capture time, system/task non-dispatch time, page operations required to dump address space

Flash Express - Stand-Alone Dump

- **Improvements in Stand-Alone Dump time when dumping data that are paged out**
- **Overall 37 sec reduction in dump time due to faster page-in of data from aux when using Flash Express**

Tests	Total dump time	Paging I/O wait time	Batch read rate MB/sec	Total GB dumped	GB of data from aux
Dasd Page data sets	00:03:12.92	00:00:41.30	438.06	36.2	17.7
Flash for paging	00:02:35.03	00:00:10.38	1612.30	36.3	16.3

WAS benchmark: z/OS Performance for Pageable Large Pages

The WAS Day Trader benchmarks showed up to an 8% performance improvement using Flash Express.

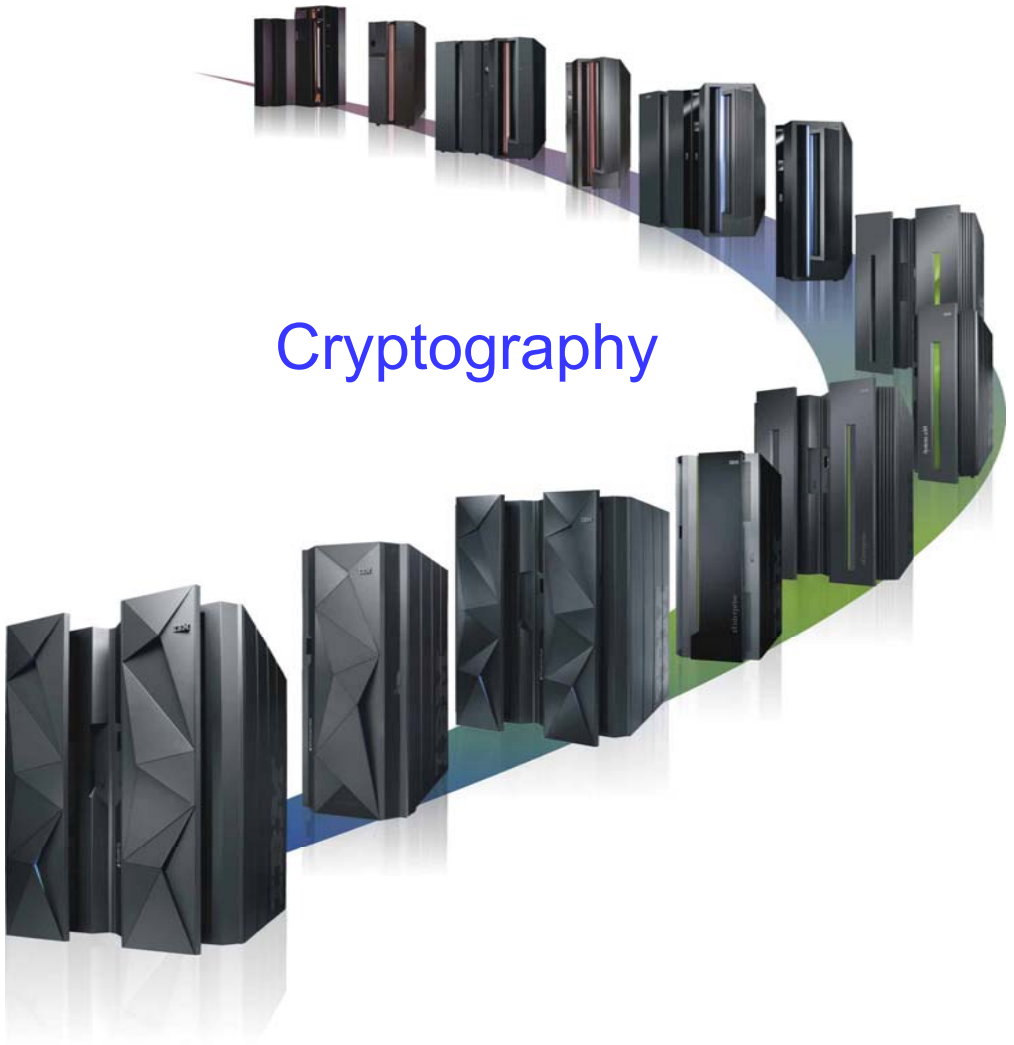
Java 7 SR3	JIT	Java Heap	Multi Threaded	WAS Day Trader 2.0
31 bit	yes	yes	4%	
64 bit	yes		1%	3%
64 bit		yes	4%	5%

* WAS Day Trader 64-bit Java 7 SR3 with JIT code cache & Java Heap

DETAILS

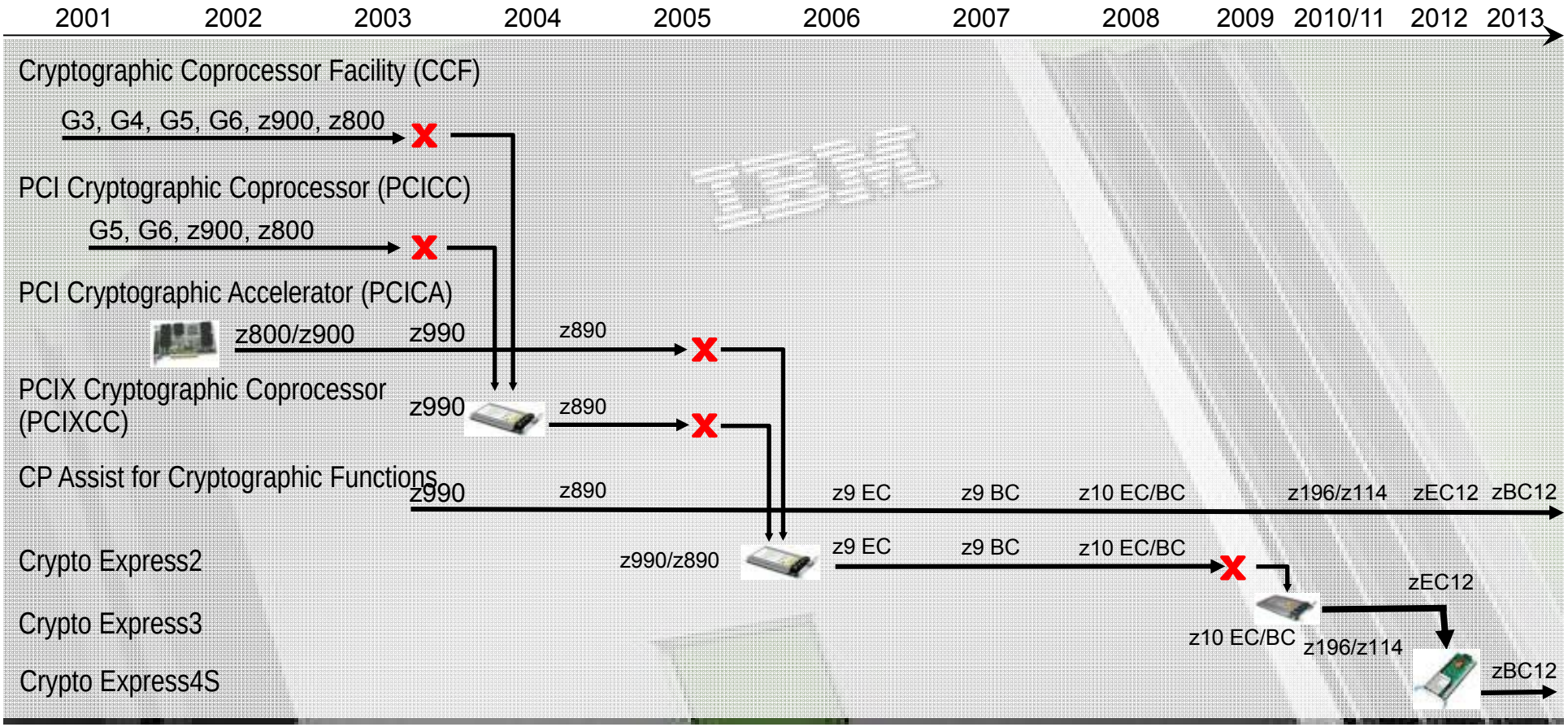
- **64-bit Java heap (**1M fixed large pages (FLPs) or 1M Pageable (PLPs)) versus 4k pages
  - Java heap 1M PLPs improve performance by about
  - **4%** for Multi-Threaded workload
  - **5 %** for WAS Day Trader 2.0
- **64-bit** Java 7 SR3 with **JIT code cache** 1M PLPs vs without Flash
  - **3** % improvement for traditional WAS Day Trader 2.0*
  - **1** % improvement for Java Multi-Threaded workload
- **31-bit** Java 7 SR3 with **JIT code cache and Java heap** 1M PLPs vs without Flash
  - **4 %** improvement for Java Multi-Threaded workload

* Note: This test used 64-bit Java 7 SR3 with JIT code cache & Java Heap leveraging Flash and pageable large pages. Also , tests used WAS Day Trader app that supports PLP; earlier version of 31-bit Java did not allocate 1M large pages



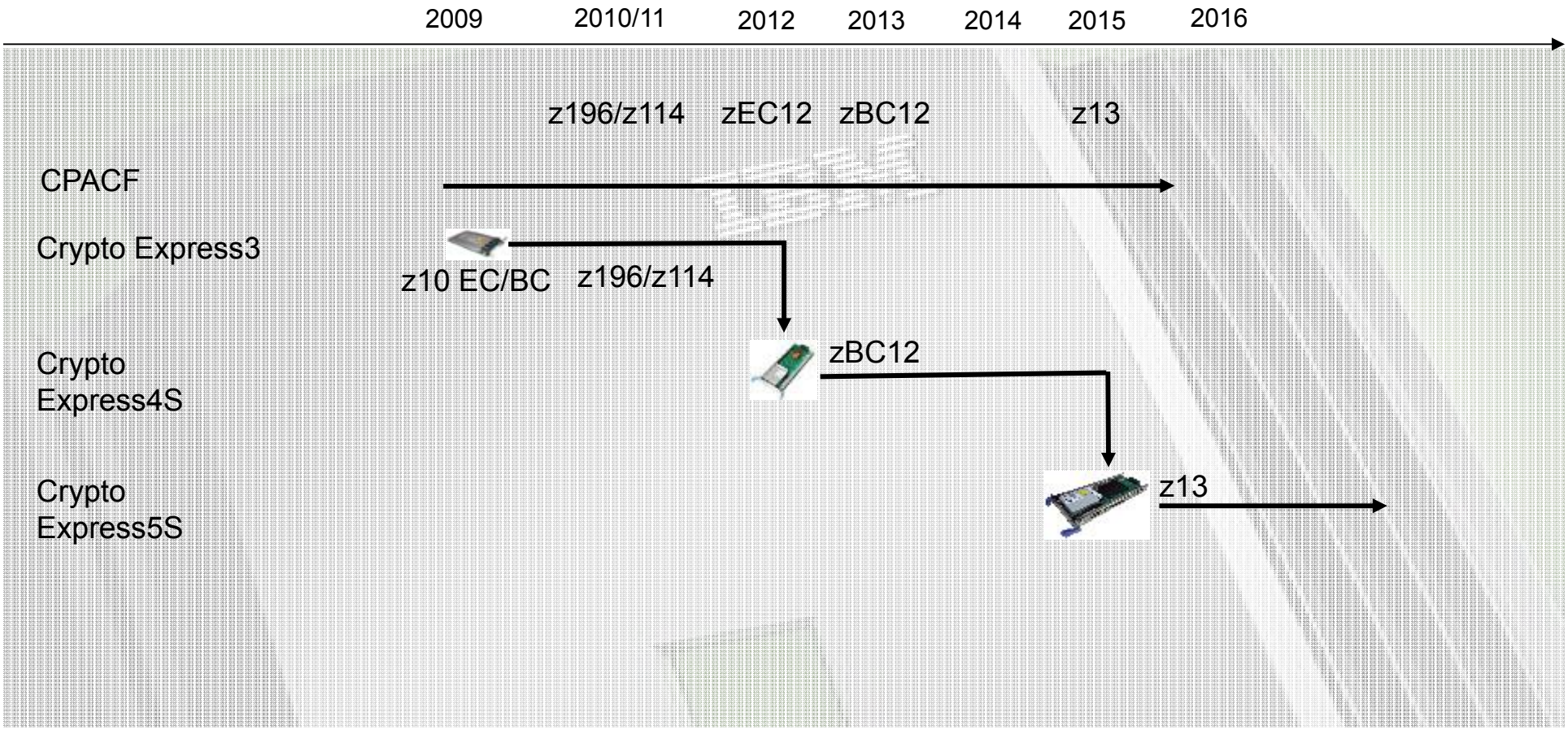
Cryptography

z System Crypto History



- Cryptographic Coprocessor Facility – Supports “Secure key” cryptographic processing
- PCICC Feature – Supports “Secure key” cryptographic processing
- PCICA Feature – Supports “Clear key” SSL acceleration
- PCIXCC Feature – Supports “Secure key” cryptographic processing
- CP Assist for Cryptographic Function allows limited “Clear key” crypto functions from any CP/IFL
  - NOT equivalent to CCF on older machines in function or Crypto Express2 capability
- Crypto Express2 – Combines function and performance of PCICA and PCICC
- Crypto Express3 – PCIe Interface, additional processing capacity with improved RAS
- Crypto Express4S - IBM Standard PKCS #EP11

z System Crypto History

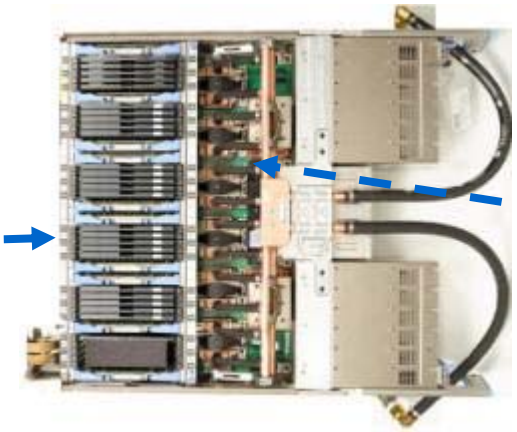


- CP Assist for Cryptographic Function allows limited “Clear key” crypto functions from any CP/IFL
  - NOT equivalent to CCF on older machines in function or Crypto Express2 capability
- Crypto Express3 – PCIe Interface, additional processing capacity with improved RAS
- Crypto Express4S - IBM Standard PKCS #EP11
- Crypto Express5S - ECC HW acceleration and more RSA Engines

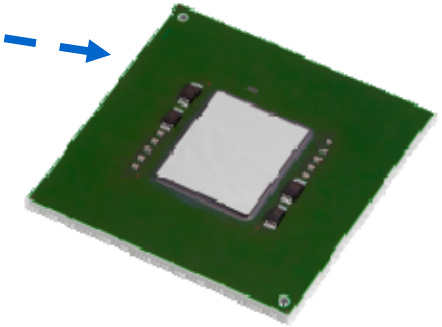
Overview – HW Crypto support in z System

PCIe I/O drawers

CPC Drawer



PU SCM:
Each PU is capable of having the CPACF function



Crypto Express5S

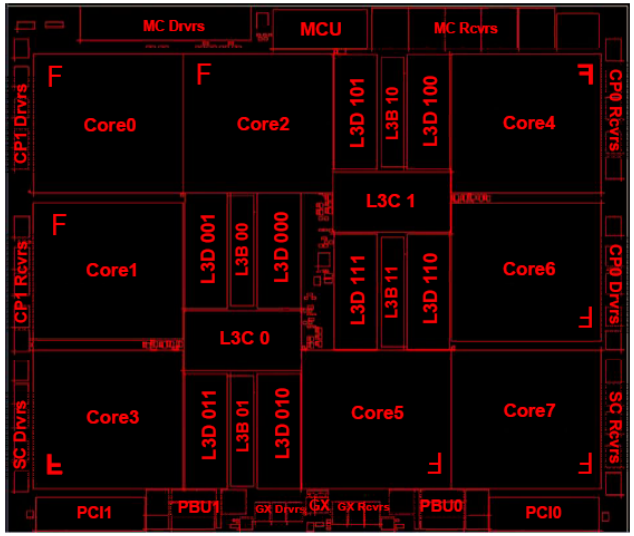
TKE:
recommended for management of Crypto Express5S and required for EP11 mode

Trusted Key Entry (TKE) workstation

Smart Cards

Smart Card Readers

CPACF - CP Assist For Cryptographic Functions

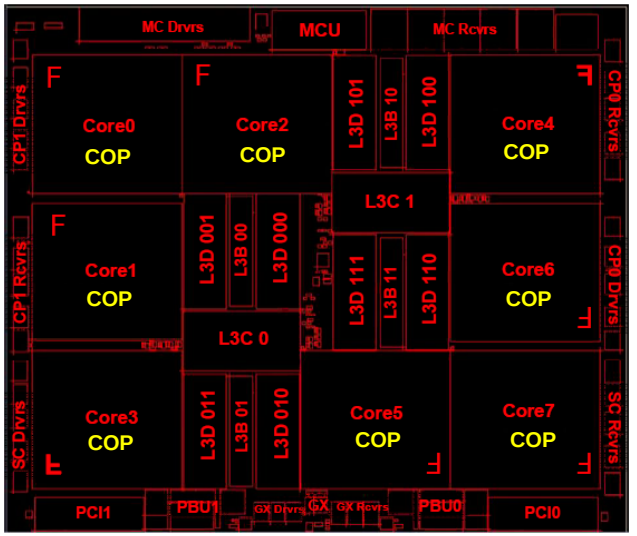


Supported Algorithms	Clear Key	Protected Key
DES, T-DES	Y	Y
AES128	Y	Y
AES192	Y	Y
AES256	Y	Y
SHA-1	Y	N/A
SHA-256	Y	N/A
SHA-384	Y	N/A
SHA-512	Y	N/A
PRNG	Y	N/A
DRNG	Y	N/A

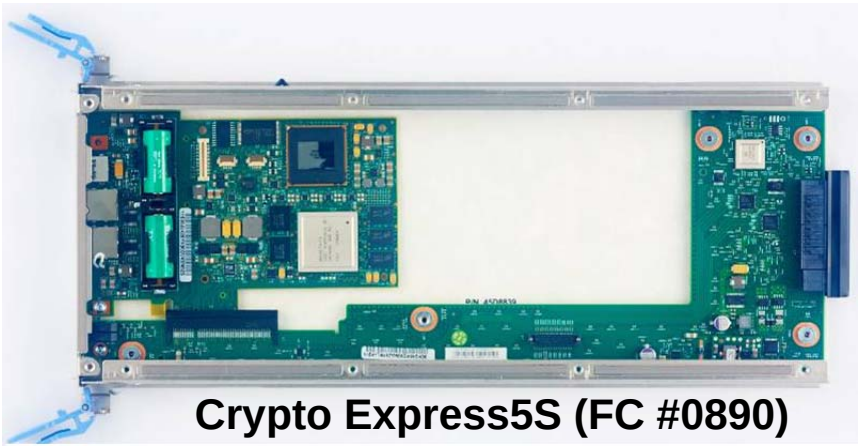
- Provides a set of symmetric cryptographic functions and hashing functions for:
  - Data privacy and confidentiality
  - Data integrity
  - Random Number generation
  - Message Authentication
- Enhances the encryption/decryption performance of clear-key operations for
  - SSL
  - VPN
  - Data storing applications
- Available on every Processor Unit defined as a CP, IFL and zIIP
- Supported by z/OS, z/VM, z/VSE, z/TPF and Linux on z System
- Must be explicitly enabled, using a no-charge enablement feature (#3863),
  - SHA algorithms enabled with each server
- Protected key support for additional security of cryptographic keys
  - Crypto Express5S required in CCA mode

z13 CPACF

- **CP Assist for Cryptographic Function Co-processor redesigned from "ground up"**
- **Design estimates are “steady state”**
  - Estimates do not include overhead for COP start/end and cache effects
  - Design estimates for large blocks of data
    - AES: 2x throughput vs. zEC12 (estimate)
    - TDES: 2x throughput vs. zEC12 (estimate)
    - SHA: 4x throughput vs. zEC12 (estimate)
- **Exploiters of the CPACF benefit from the throughput improvements of z13's CPACF such as:**
  - DB2/IMS encryption tool
  - DB2® built in encryption
  - z/OS Communication Server: IPsec/IKE/AT-TLS
  - z/OS System SSL
  - z/OS Network Authentication Service (Kerberos)
  - DFDSS Volume encryption
  - z/OS Java SDK
  - z/OS Encryption Facility
  - Linux on z Systems; kernel, openssl, openCryptoki, GSKIT



Crypto Express5S Details


Crypto Express5S (FC #0890)

- Native PCIe card (FC #0890)
  - Resides in the PCIe I/O drawer
  - Requires CPACF Enablement (FC #3863)
- New Crypto Module
  - Designed to more than double Crypto Express4S performance (Added L2 Cache, New Crypto ASIC and processor upgrade)
  - Designed to support up to 85 domains for logical partitions or z/VM guests
- Designed to Meet Physical Security Standards
  - FIPS 140-2 level 4
  - ANSI 9.97
  - Payment Card Industry (PCI) HSM
  - Deutsche Kreditwirtschaft (DK)
- New Functions, Standard and Compliance
  - Drivers: NIST via FIPS standards and implementation guidance requirements; emerging banking standards: and strengthening of cryptographic standards for attack resistance
  - VISA Format Preserving Encryption (VFPE) for credit card numbers
  - Enhanced public key Elliptic Curve Cryptography (ECC) for users such a Chrome, Firefox, and Apple's iMessage
- New Trusted Key Entry Workstation
  - Workstation and LIC – FC #0847 with new crypto module and TKE LIC 8.0 is required
  - Required: EP11 (PKCS #11) Mode, Recommended: Common Cryptographic Architecture (CCA) Mode
  - Additional Smart Cards (FC #0892) – Support for stronger encryption than previous cards

Crypto Express5S

- One PCIe adapter per feature
  - Initial order – two features
- Designed for 2X performance increase over CEX4

Three configuration options for the PCIe adapter

- Only one configuration option can be chosen at any given time
- Switching between configuration modes will erase all card secrets
  - Exception: Switching from CCA to accelerator or vice versa

Accelerator		CCA Coprocessor		EP11 Coprocessor	
TKE	N/A	TKE	OPTIONAL	TKE	REQUIRED
CPACF	NO	CPACF	REQUIRED	CPACF	REQUIRED
UDX	N/A	UDX	YES	UDX	NO
CDU	N/A	CDU	YES(SEG3)	CDU	NO
Clear Key RSA operations and SSL acceleration		*Secure Key crypto operations*		*Secure Key crypto operations*	

Business Value

- High speed advanced cryptography; intelligent encryption of sensitive data that executes off processor saving costs
- PIN transactions, EMV transactions for integrated circuit based credit cards(chip and pin), and general-purpose cryptographic applications using symmetric key, hashing, and public key algorithms, VISA format preserving encryption(VFPE), and simplification of cryptographic key management.
- Designed to be FIPS 140-2 Level certification to meet regulations and compliance for PCI standards

Crypto Express5S Standards supported

- DES/TDES w DES/TDES MAC/CMAC
- AES, AESKW, AES GMAC, AES GCM, AES XTS mode, CMAC
- MD5, SHA-1, SHA-2 (224,256,384,512), HMAC
- VISA Format Preserving Encryption (VFPE)
- RSA (512, 1024, 2048, 4096) -> Performance improvement
- ECDSA (192, 224, 256, 384, 521 Prime/NIST)
- ECDSA (160, 192, 224, 256, 320, 384, 512 BrainPool)
- ECDH (192, 224, 256, 384, 521 Prime/NIST)
- ECDH (160, 192, 224, 256, 320, 384, 512 BrainPool)
- Montgomery Modular Math Engine
- RNG (Random Number Generator)
- PNG (Prime Number Generator) -> NEW
- Clear Key Fast Path (Symmetric and Asymmetric)

IBM CCA LIC

Architected set of cryptographic functions and application programming interfaces (APIs) that provide both general-purpose functions and a broad set of functions designed specifically to secure financial transactions and keys

Enhancements

- AES PIN support for German Banking Industry
- PKA Translate UDX functions into CCA
- VISA Format Preserving Encryption (FPE)
- >16 Domain Support

CCA Enhancement: >16 Domain support

Description & Value	z System has support for up to 85 logical partitions (LPARs). The z Systems crypto architecture was designed to support 16 domains (which matched the LPAR maximum at the time). In current customer environments where the number of LPARs is larger than 16, crypto workload separation can be complex. These customers have to map a large set of LPARs to a small set of crypto domains.
Details	With the adjunct processor (AP) extended addressing (APXA) facility installed, the z Systems crypto architecture can support up to 256 domains in an AP. So, the Crypto Express cards will be enhanced to handle 256 domains and the z System firmware will expose up to 85 domains to customers (to match the current LPAR maximum). Customers will have the flexibility of mapping individual LPARs to unique crypto domains or continuing to share crypto domains across LPARs.
Customers interested in this function	System administrators exploiting Crypto Express cards on z System CPCs with multiple LPARs.
HW/SW requirements	▪ Hardware Requirements: z13 with the following CCA Support on : − Crypto Express 5S with CCA 5.0 firmware ▪ Software Requirements: − z/OS V2.1 with the Cryptographic Support for z/OS V1R13-z/OS V2R1 web deliverable (FMID HCR77B0) − z/OS V1.13 with the Cryptographic Support for z/OS V1R13-z/OS V2R1 web deliverable (FMID HCR77B0) − Also available with HCR7780, HCR7790, HCR77A0, HCR77A1 (previous WDs with PTFs) − z/VM 6.2 and 6.3 with PTFs for guest exploitation

CCA Enhancement: VISA Format Preserving Encryption (VFPE)

Description & Value	Format Preserving Encryption (VFPE) refers to a method of encryption where the resulting cipher text has the same form as the input clear text. The form of the text can vary according to use and application. One of the classic examples is a 16 digit credit card number. After using VFPE to encrypt a credit card number, the resulting cipher text will be another 16 digit number. This helps allow legacy databases to contain encrypted data of sensitive fields without having to restructure the database or applications.
Details	VFPE allows customers to add encryption to their applications in such a way that the encrypted data can flow through their systems without requiring massive redesign of their application. In the above example – if the credit card number is VFPE encrypted at the point of entry – the cipher text still behaves like a credit card number and can flow through business logic until it meets that back end transaction server which can VFPE decrypt it to get the original credit card number to process the transaction.
Customers interested in this function	Financial institutions that want to exploit the VISA format preserving encryption algorithms to protect cardholder data. Application programmers that want to provide field/cell level encryption to their database.
HW/SW requirements	▪ CCA 5.0 − Available on Crypto Express5S ▪ Software Requirements: −z/OS V2.1 with the Cryptographic Support for z/OS V1R13-z/OS V2R1 web deliverable (FMID HCR77B0) −z/OS V1.13 with the Cryptographic Support for z/OS V1R13-z/OS V2R1 web deliverable (FMID HCR77B0) −z/VM 6.2 and 6.3 with PTFs for guest exploitation

TKE 8.0 – New Features

- **TKE 8.0 required to manage host Crypto Express5S**
- **Only TKE 8.0 can be used to manage domains higher than the 16th**
- **Full Function Migration Wizard requirement when data is applied to a host Crypto Express5S**
  - If data will be applied to a host Crypto Express5S, the collect must be done using Key Part Holder Certificates from KPH smart cards created on TKE 8.0
    - Recommendation: If data will be applied to a Crypto Express5S, do the collect from the source module from TKE 8.0

TKE Hardware Support and Migration Information

On z13, if TKE workstation is required, TKE 8.0 must be used

TKE Workstation					Server Supported					Manage Host Crypto Module				
TKE Release (LIC)	HW Feature Code	LIC	Smart Reader	Smart Card	z196	z114	zEC12	zBC12	z13	Crypto Express3 (CCA)	Crypto Express4S (CCA)	Crypto Express4S (EP11)	Crypto Express5S (CCA)	Crypto Express5S (EP11)
TKE 7.2	0841	0850	0885	0884	yes	yes	yes			yes	yes	yes [1]		
TKE 7.3	0842	0872	0885	0884 [2]	yes	yes	yes	yes		yes	yes	yes [1]		
TKE 8.0	0847	0877	0891	0892			yes	yes	yes		yes	yes [1]	yes	yes [1]

Table Notes:

1. A Crypto Express4S running in IBM Enterprise PKCS #11 (EP11) mode, as a CEX4P, requires smart cards to hold administrator certificates and master key material. The smart card must be part #74Y0551.
2. Older smart cards 45D3398 (FC 0884) and 74Y0551 (FC 0884) may be used on TKE 8.0 (Available from System z10)

The TKE is unaware of the type of CPC where the host crypto module is installed. That is, the TKE does not care whether a Crypto Express3 is running on a z10, z196/z114 or zEC12/zBC12. This means that TKE LIC can support any CPC where the coprocessor is supported, but the TKE LIC must support the specific crypto module.

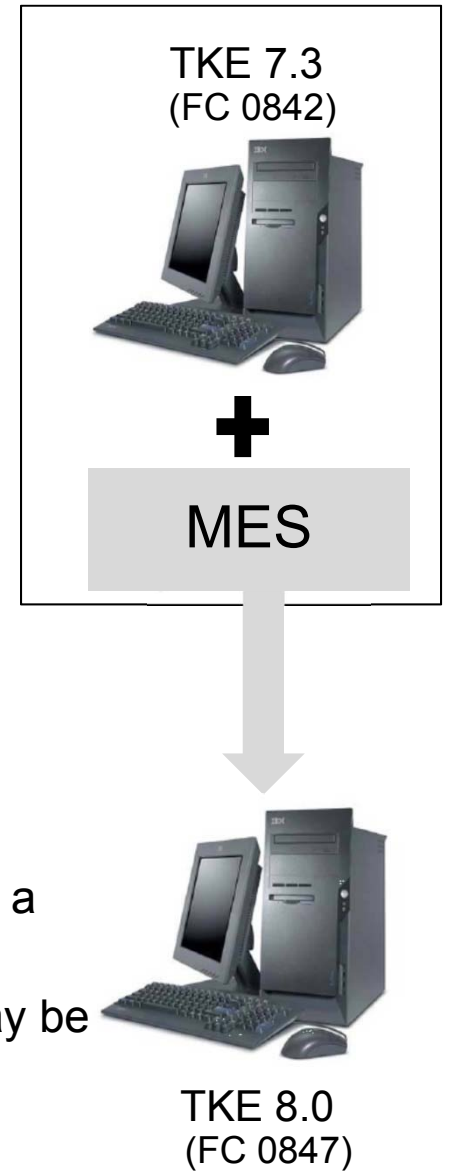
TKE 8.0 Workstation

TKE 8.0 Workstation, FC 0847

- Available for ordering on z13, zEC12 and zBC12
  - MES availability dates for zBC12 & zEC12 on February 16, 2015
- Same workstation used in older FC 0842 but requires:
  - New crypto adapter 4767
- TKE 7.3 workstation FC 0842 can be upgraded TKE 8.0 workstation FC 0847
  - MES generates FC 0894 to add 4767 adapter

Things that can ONLY be managed by TKE 8.0

- A host Crypto Express5
- Greater than 16 Domain support
- Full Function Migration Wizard requirement when data is applied to a host Crypto Express5
- Older smart cards 45D3398 (FC 0884) and 74Y0551 (FC 0884) may be used on TKE 8.0
- Smart card 00JA710 to be released (74Y0551 EOL)



TKE 8.0 – Installation:

- **Support upgrade of FC 0842 to FC 0847**

– Process if MES path from FC 0842 to FC 0847

1. Save Upgrade Data on TKE 7.x to USB memory to save client data
   - Recommended step if running TKE 7.3: Prior to Save Upgrade Data, run the TKE Workstation Setup wizard and perform *Save User Roles and Profiles* task.
2. Replace the 4765 crypto adapter with the 4767 crypto adapter
3. Upgrade firmware to TKE 8.0
   - Note: The order of steps 2 and 3 can be switched
4. Perform Frame Roll install to apply Save Upgrade Data (customer data) to TKE 8.0 system
5. Run the TKE Workstation Setup wizard
   - Note: Wizard will install proper code on 4767 crypto adapter

Note: see a complete step by step TKE migration sequence within the extras at the end of this presentation

Summary of Crypto Features

Description	F/C	zEC12	zBC12	z13
CPACF enablement	3863	New/CF	New/CF	New/CF
Crypto Express5S	0890	NA	NA	New
Crypto Express4S	0865	New	New	NA
Crypto Express3	0864	CF	CF	NA
Crypto Express3-1P	0871	NA	CF	NA
TKE 8.0 LIC	0877	NA	NA	New
TKE 7.3 LIC	0872	New	New	CF*
TKE 7.2 LIC	0850	CF	CF	NA
TKE Workstation	0847	NA	NA	New
TKE Workstation	0842	New	New	CF*
Smart Card Reader	0885	New/CF	New/CF	New/CF
Add'l Smart Cards	0884	New/CF	New/CF	New/CF

* Hw upgrade EC required

Software Requirements

Description	z/OS	z/VM	z/VSE	z/TPF	Linux	TKE LIC	Smart card
CPACF enablement	**z/OS v1.12, z/OS v1.13, z/OS v2.1**	**z/VM V6.2, z/VM V6.3**	**z/VSE 5.1 z/VSE 5.2**	**z/TPF v1.1**	**Yes**	**N/A**	**N/A**
CCA 4.5/5.0	**z/OS v1.13, z/OS v2.1 WD#14 HCR77B0**	**z/VM V6.2, z/VM V6.3**	**z/VSE 5.1, z/VSE 5.2 (with PTF)**	**z/TPF v1.1 with APAR supports Accelerator mode only**	**SLES 11 & 12(tol) SLES 13 (Exp) RHEL 6 & 7(Tol) RHEL 8(exp)**	**8.0 LIC (optional)**	**Optional**
EP 11	**z/OS v1.13, z/OS v2.1 WD#14 HCR77B0**	**z/VM V6.2, z/VM V6.3**	**N/A**	**N/A**	**Future Objective**	**8.0 LIC required**	**Required**

Questions ?

– Ewerson Palacio

bird@br.ibm.com

– Frank Packheiser

F.Packheiser@de.ibm.com

– Parwez Hamid

pnh@us.ibm.com

IBM z13

Reinventing enterprise IT
for digital business

Additional Slides

(available as additional material – Backup)

z13 FICON Enhancements Details

Additional Slides

(available as additional material – not presented)

New FICON Functions for z13

- 16 Gbps Link Speeds
  - Designed to reduce I/O latency to improve response time for performance-critical middleware and to shrink the batch window required to accommodate I/O bound batch work
- FICON Dynamic Routing
  - Designed to allow ISL sharing by FC and FCP traffic to optimize use of ISL bandwidth in the SAN fabric for both types of traffic
- SAN Fabric Priority
  - Extends z/OS WLM policy into the SAN fabric
  - Gives important work priority to get through SAN traffic congestion (e.g. after SAN hardware failures)
- zHPF Extended Distance II
  - Up to 50% I/O service time improvement for remote write
  - Designed to help GDPS HyperSwap configurations with secondary DASD in remote site
- 32K devices per FICON channel
  - Up to 85 Logical Partitions: More flexibility for server consolidation
- Preserve Virtual WWPNs for NPIV configured FCP channels
  - Designed to simplify migration to a new-build z13
- Forward Error Correction Codes
  - Designed to addresses high bit-error rate on high frequency (>= 8Gb/s) links
  - Estimated equivalence to doubling optical signal power
- 6th Logical Channel Subsystem
  - Up to 85 Logical Partitions: More flexibility for server consolidation
- 4th Subchannel Set
  - Simplifies I/O configurations for a 2nd synchronous copy of data
  - With multi-target PPRC, can do HyperSwap and still maintain synchronous copy for 2nd HyperSwap

Leadership with FICON Express16S

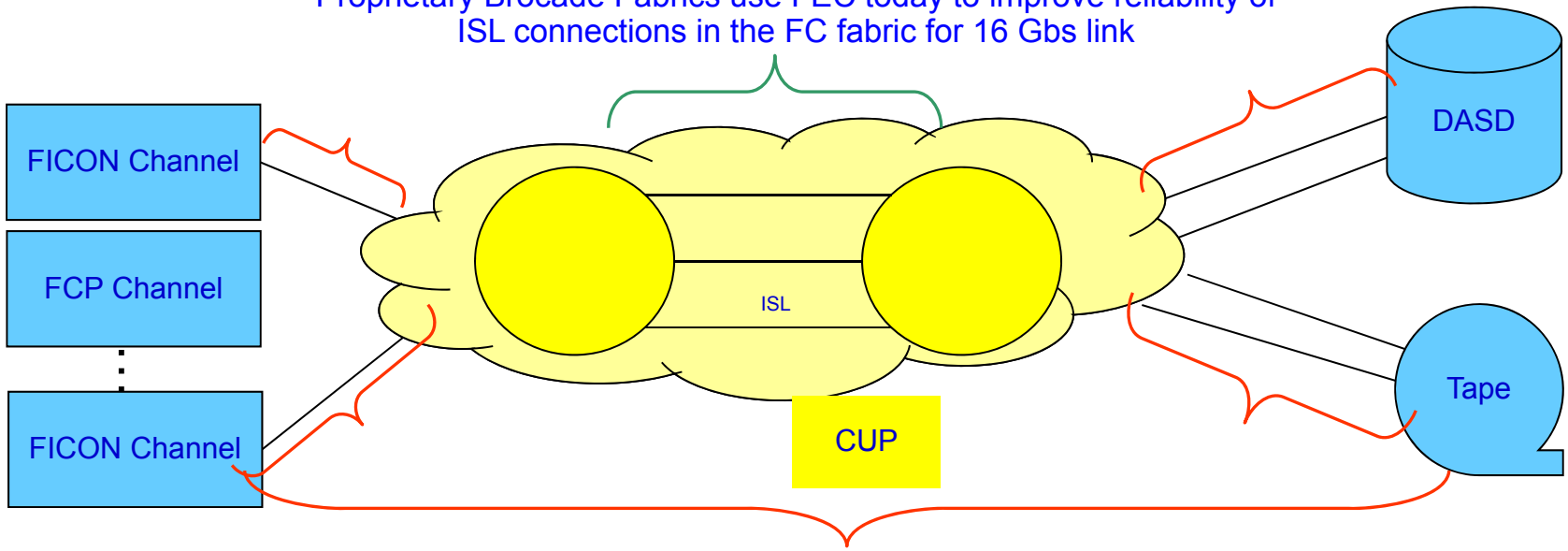
Faster Link Speed yields latency improvements for large block transfers

- **z/OS DB2 Writes to the LOG**
  - 12-14% latency reduction in 128K log writes
  - Yields 5-6% improvement in Transaction Latency, depending on workload
- **z/OS Managed File Transfer**
  - DS8000 zDDB feature for exchanging data through the SAN
  - New z/OS IOS FBA Access Method
  - New Connect:Direct exploitation - 16 MB reads/writes
  - Lower CPU cost with zDDB adds flexibility for when managed file transfer (MFT) is done
- **Reduce the Batch Window**
  - ISV's (e.g. FASTVSAM 1 MB/SSCH)
  - DSS backup/restore

Prevent I/O Errors with Forward Error Correction (FEC) Codes

- New standard for transmission of data on 16 Gbs Links
  - T11.org FC-FS-3 standard defines use of 64b/66b encoding
    - Designed to improve efficiency to 97% vs. 80% with 8b/10b encoding
  - FEC codes provide error correction on top of 64b/66b encoding
    - Designed to improve reliability by reducing bit errors (adds equivalent of 2.5 db signal strength)
    - Up to 11 bit errors per 2112 bits can be corrected
    - IBM is leading new standards required to enable FEC for optical links

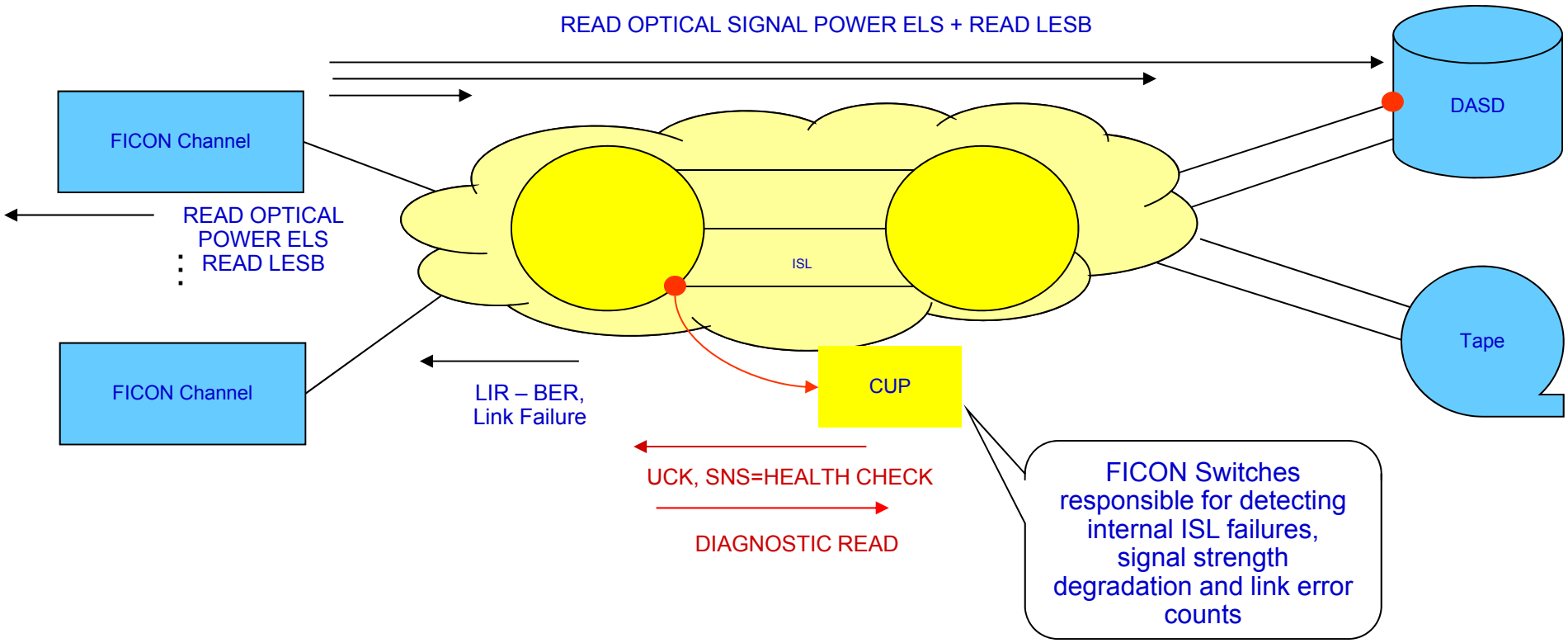
Proprietary Brocade Fabrics use FEC today to improve reliability of ISL connections in the FC fabric for 16 Gbs link



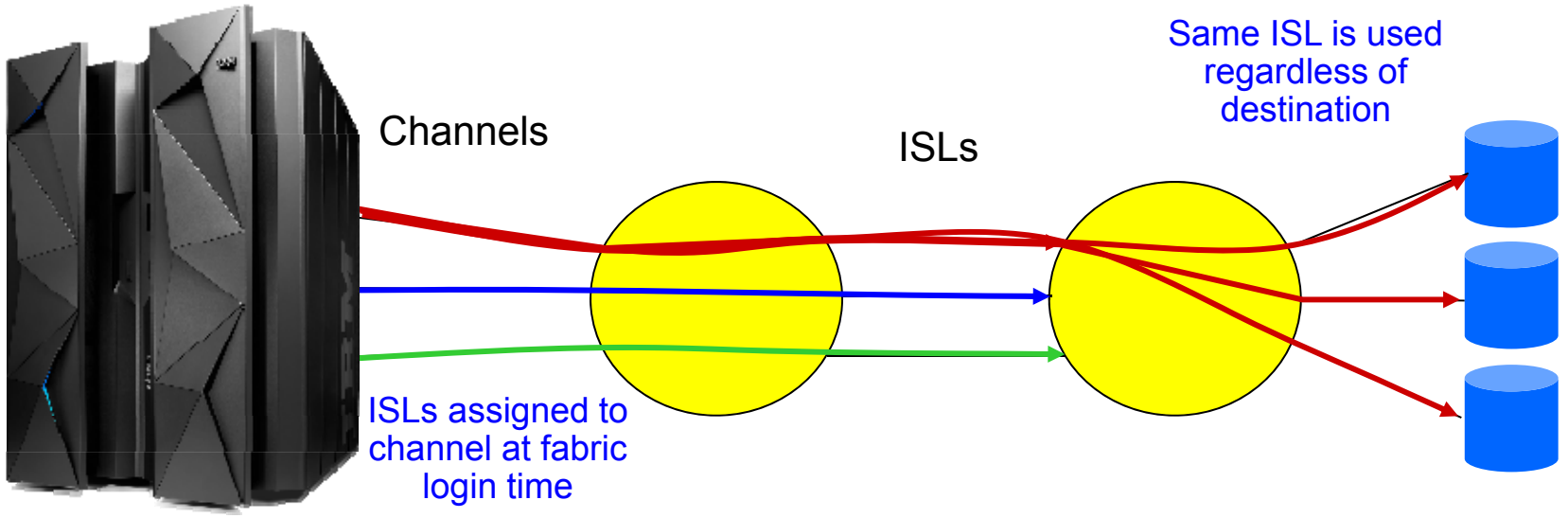
z13 and DS8870 will extend the use of FEC to the fabric N_Ports for complete end-to-end coverage for new 16 Gbs FC links

Policy Based SAN Alerts

- **Extend SAN Diagnostics to trigger on policy based alerts**

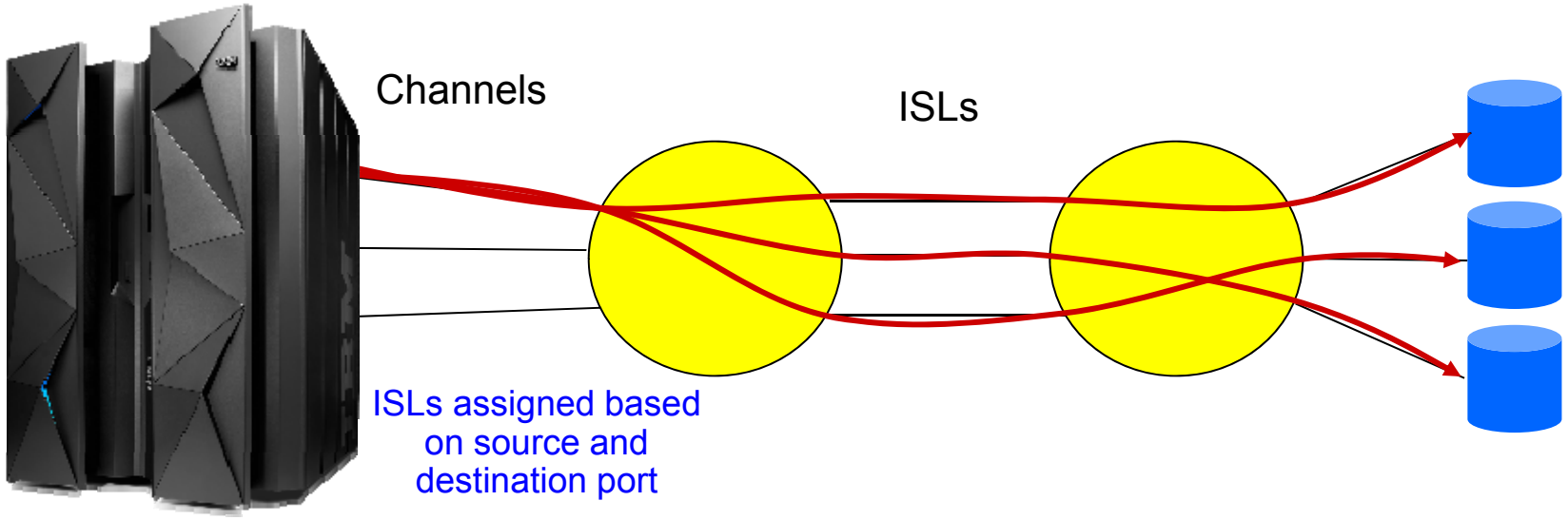


Static Routing (Port Based Routing, PBR)



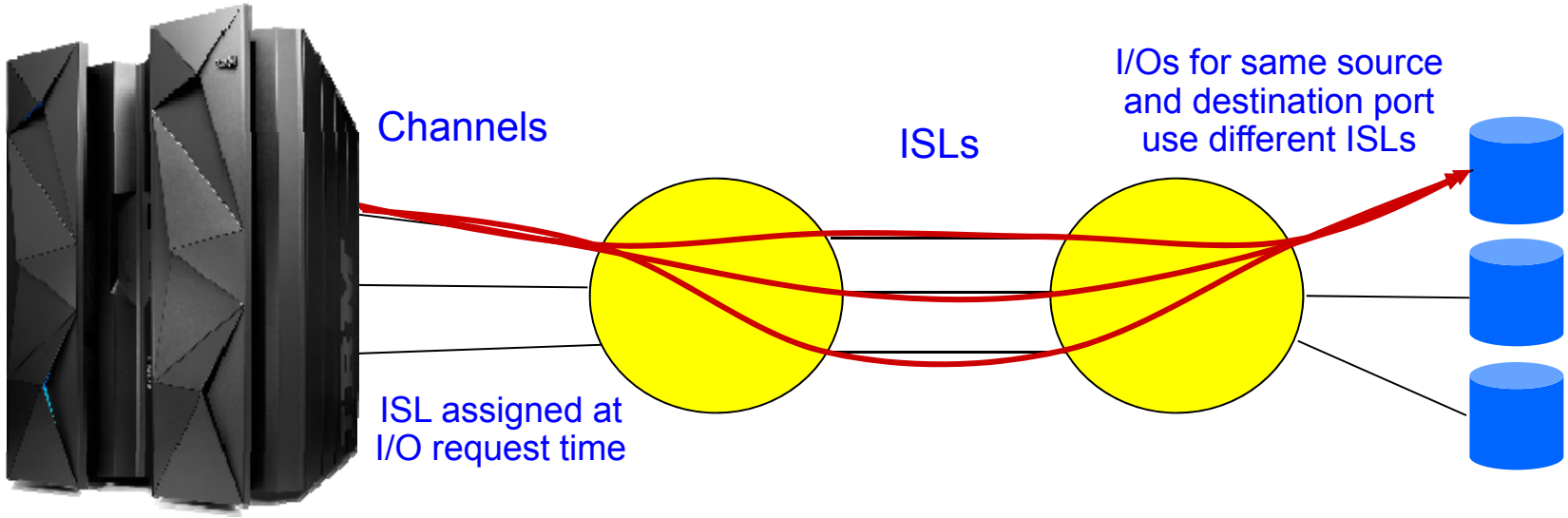
- Port Based Routing (PBR) assigns the ISL (route) statically based on “First Come, First Served” at fabric login (FLOGI) time
- The ISL is assigned “Round Robin” as ports log in
- Switch has no idea how the port that is logging-in will use the ISL and whether it will cause bottlenecks
- This can result in some ISLs overloaded, some under utilized
- Allow/Prohibit manual controls are too complicated to manage
- Routing tables change after every POR
- z Systems channel selection algorithm (zEC12) will move work away from congested ISLs

Static Routing (Device Based Routing, DBR)



- Device Based Routing (DBR) assigns the ISL (route) statically based on a hash of the source and destination port. Same as CISCO default routing.
- Spreads load across ISLs much better than PBR, but no guarantee that the same ISL won't be assigned
- z Systems channel selection algorithm (zEC12) will move work away from congested ISLs
- In some configurations, ISLs may go unused – e.g., four PPRC ports in a fabric with 8 ISLs

Dynamic Routing (Exchange Based Routing, OxID)



- Dynamic Routing (Brocade EBR or CISCO OxID) dynamically changes the routing between the channel and control unit based on the “Fibre Channel Exchange ID”
- Each I/O operation has a unique exchange id
- Client Value:
  - Reduces cost by allowing sharing of ISLs between FICON, FCP (PPRC or distributed)
  - I/O traffic is better balanced between all available ISLs
  - Improves utilization of switch and ISL hardware – ~37.5% bandwidth increase
  - Easier to manage
  - Easier to do capacity planning for ISL bandwidth requirements
  - Predictable, repeatable I/O performance
  - Positions FICON for future technology improvements, such as work load based routing

FICON Dynamic Routing Enablement

- **HCD/HCM will prevent a user from accidentally defining a dynamic routing device to a fabric, which is not capable of dynamic routing**
- **A new dynamic routing attribute will be added to switches and control units, which can be considered in the HCD/HCM dialogs**
  - Based on this flag HCD will perform a check at build production IODF, whether dynamic routing is supported on the complete path
  - A warning message will be shown, if a mixture of dynamic routing capable and incapable control units and switches is found
- **To enable FICON Dynamic Routing:**
  - Install the firmware on the switch that supports the Control Unit Port (CUP) function,
  - Define the switches as z/OS devices in the IODF
  - Bring these online

Health Check for FICON Dynamic Routing

- **z13 is changing the channel microcode to support dynamic routing**
  - No changes are needed in z/OS to support dynamic routing
  - However, if a customer configures their switches for dynamic routing and either the processor and/or storage controllers do not support it, then this can lead to I/O errors
    - Therefore, a health check is being provided that interrogates the switch to determine if dynamic routing is enabled in the switch fabric.

- **No action is required to enable the health check**
  - It will automatically be enabled at IPL and react to changes that might cause problems. For example:
    - Varying on the first device for a control unit that does not support dynamic routing, when the control unit is connected to a switch fabric enabled for dynamic routing
    - Configuring a new channel online that is connected to a switch fabric enabled for dynamic routing
  - The health check can be disabled via modify commands, PARMLIB or via SDSF
    - If so, then the dynamic routing capability of the processor and control units is checked.
    - If either do not support dynamic routing, then a health check exception is surfaced.
    - Running the health check in VERBOSE(YES) mode will cause the health check to check whether all devices support dynamic routing, regardless of whether the devices are connected to a switch fabric that is enabled for dynamic routing.

Rollout of Routing Policy Enhancements

FICON Dynamic Routing (FIDR) is the z13 feature that supports the use of dynamic routing policies in the switch, such as EBR for Brocade and OxID for CISCO

- June 2013: Brocade Device Based Routing (DBR)
  - Uses source and destination id as a hash value to map to the ISL to be used
  - Improved ISL Performance (I/O service time)
  - Improved ISL Bandwidth (workload spikes)
  - Deployed at large financial institutions
- z13: FICON Dynamic Routing (EBR/OxID)
  - New function in the channel subsystem to handle out of order arrival
  - Device qualification required as timing changes may expose implementation errors
    - Some older devices may never support it
    - DS8870 7.5 release will be qualified in June 2015
  - Migration Plan
    - Clients define two virtual fabrics, EBR capable and non-EBR capable
      - z/OS HCD will police which devices are allowed to be defined on EBR/OxID switches
      - Health Check Infrastructure checks switches, processors and devices
      - ISLs need to be divided between virtual fabrics
    - Potential pitfalls
      - Adds complication
      - Requires partitioning ISLs, which defeats the purpose of EBR

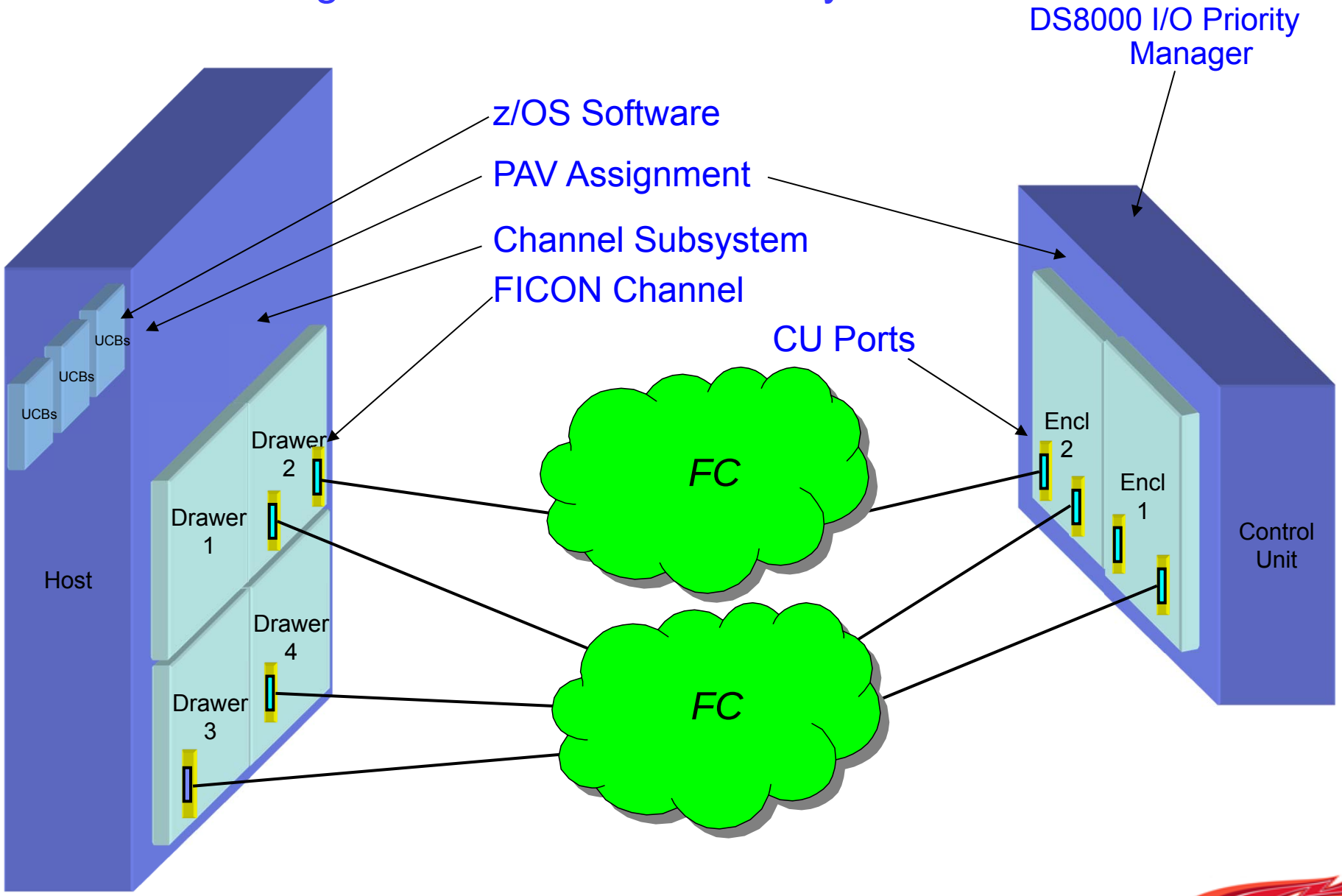
Fabric Priority for an I/O request

- **Provide the ability for z/OS to pass a fabric priority for an I/O request.**
  - The fabric priority is used when the processor is connected to the devices via one or more FICON switches or directors
  - The fabric priority is used by the switch when congestion occurs to determine which I/O requests should be given preference
- **The fabric priority range to be used is obtained by z/OS from the control unit port (CUP) device associated with the switch**
  - The information returned contains the maximum priority used by the switch (the minimum is assumed to be 1).
  - This global maximum priority is used by WLM to assign a fabric priority to an address space or enclave based on the performance goals of the unit of work.
  - IOS extracts the priority assigned by WLM and passes it to the channel subsystem when the I/O is started if in WLM's Service Definition the function for I/O Priority Management is set active
  - The channel that is selected adds the fabric priority to each of the frames associated with the I/O request, which can then be used by the switch to prioritize the frames.

Fabric Priority for an I/O request

- **To enable the function. the following is required:**
  - LPAR Image profile must have Fabric I/O priority on
  - Install the firmware on the switch that supports the Control Unit Port (CUP) function
  - Define the switches as z/OS devices, and enable I/O priority queueing in the IODF
    - This will enable all three priorities (CSS, Fabric and CU)
  - Bring them online.
    - This allows z/OS to obtain the maximum priority for the switch
  - DS8000 and TS7700 microcode should be installed that echoes the priority in the frames for read I/O requests
  - Set I/O Priority Management active in your WLM Service Definition
  - IECIOSxx or SETIOS FICON,FABRICPRTY=YES|NO can be used to enable or disable the function.
    - The default is **YES** (Enable)

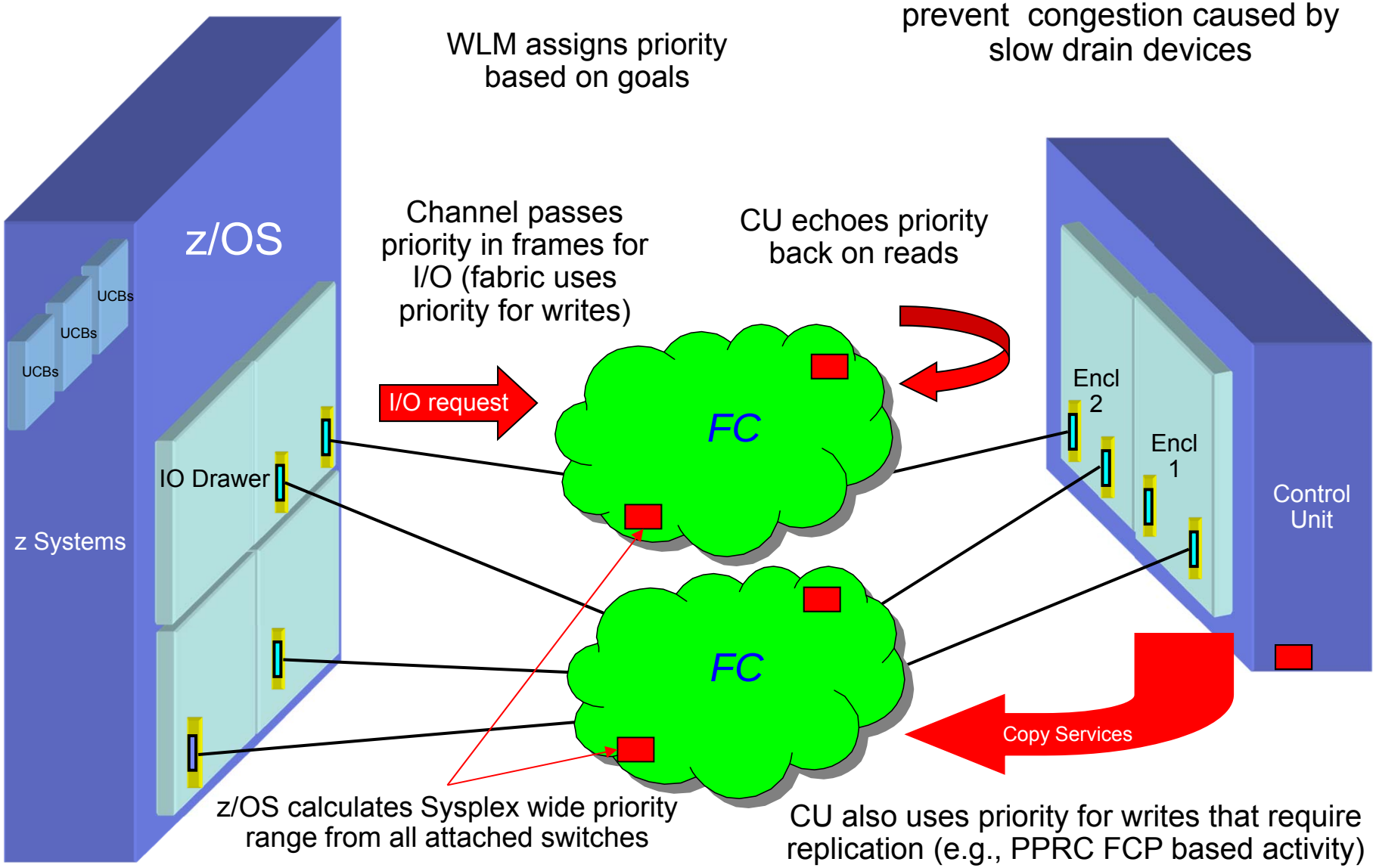
Work Load Manager and I/O Priorities Today



z Systems Fabric I/O Priority

Will be used to alleviate or prevent congestion caused by slow drain devices

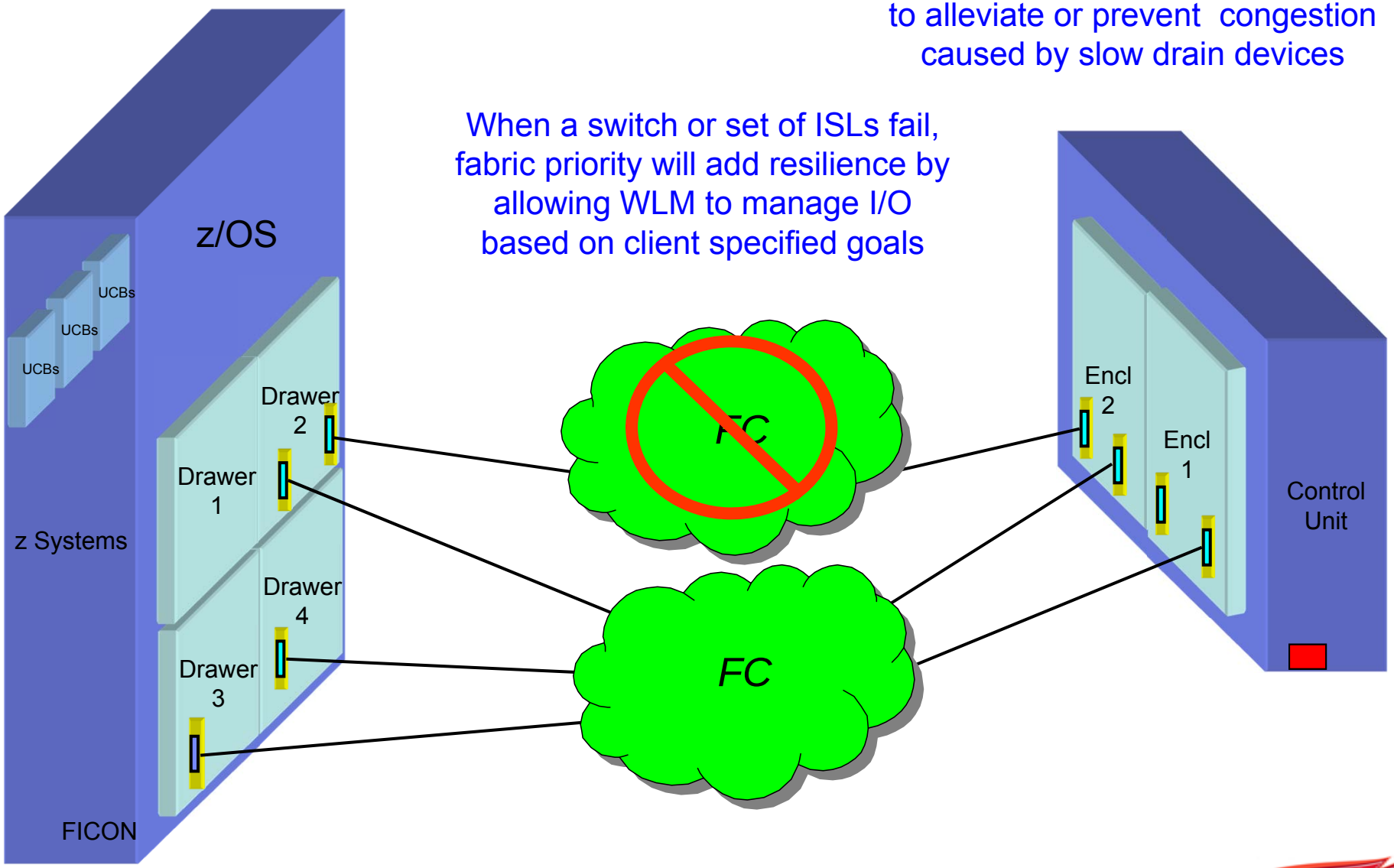
WLM assigns priority based on goals



z Systems Fabric I/O Priority

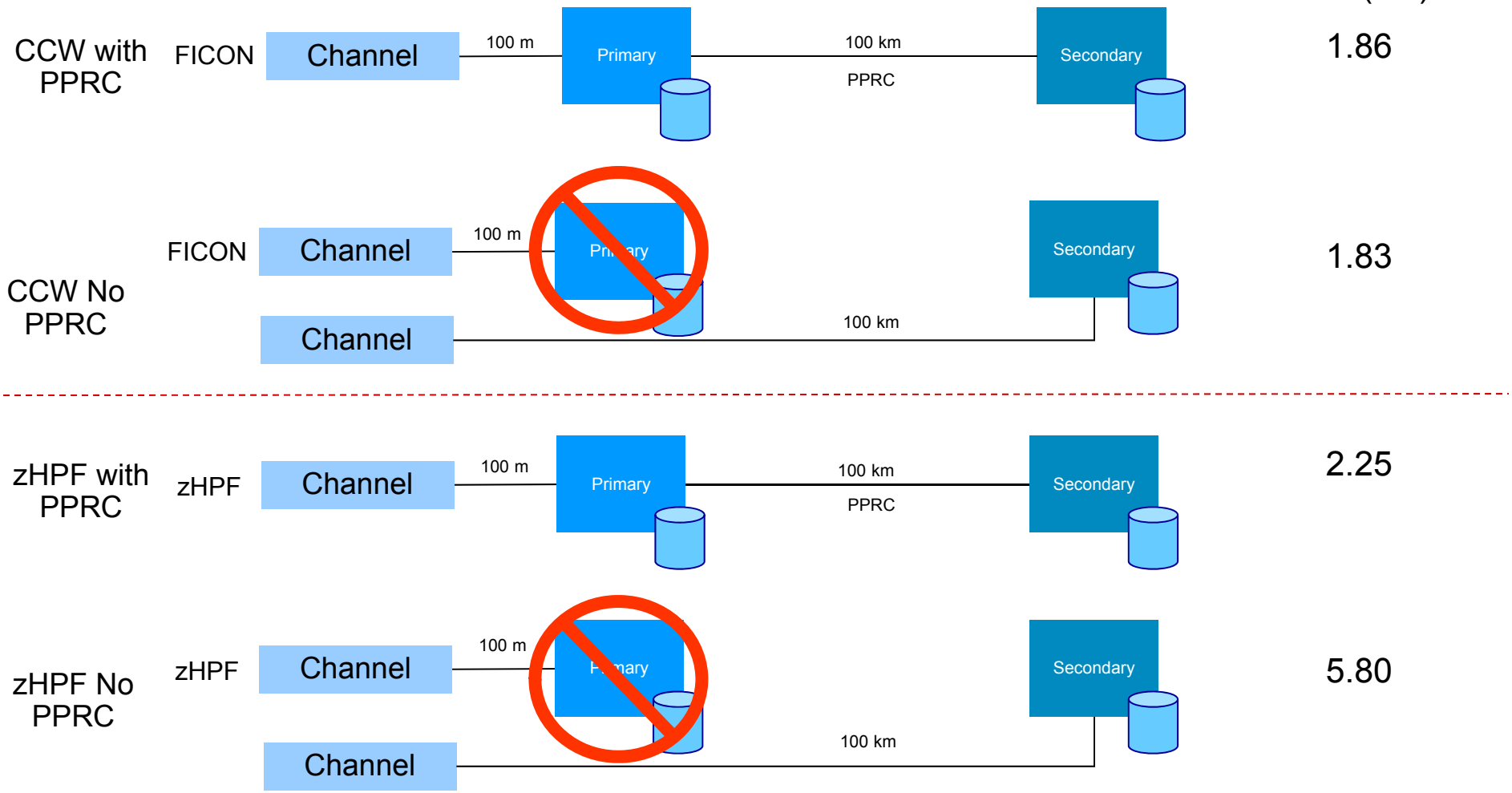
Will be used to operating system to alleviate or prevent congestion caused by slow drain devices

When a switch or set of ISLs fail, fabric priority will add resilience by allowing WLM to manage I/O based on client specified goals

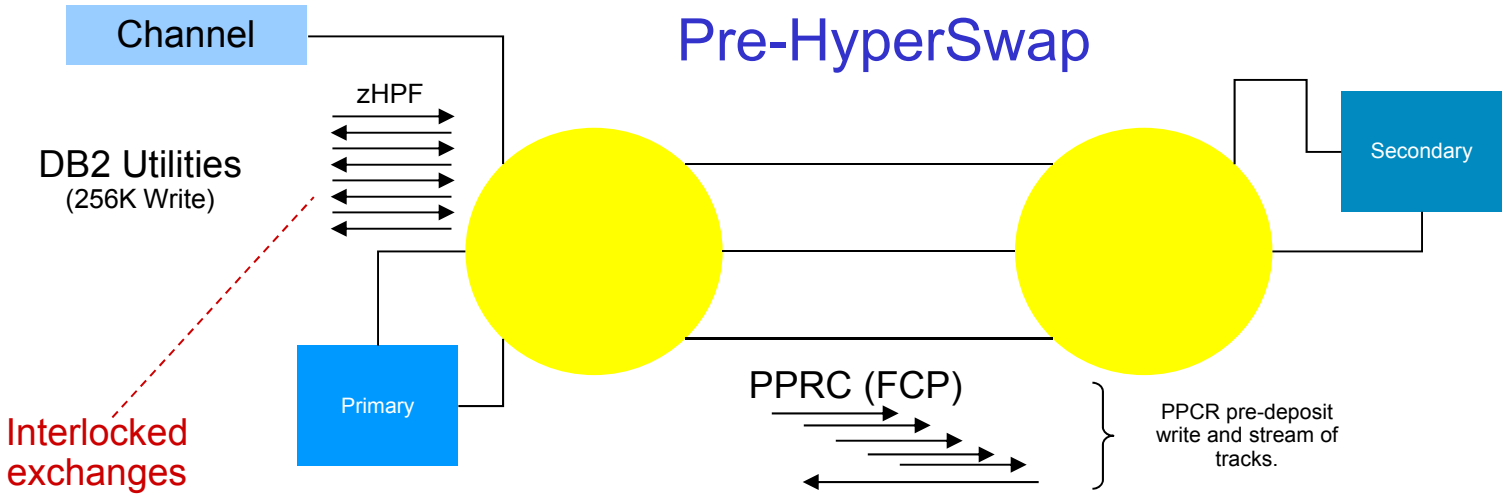


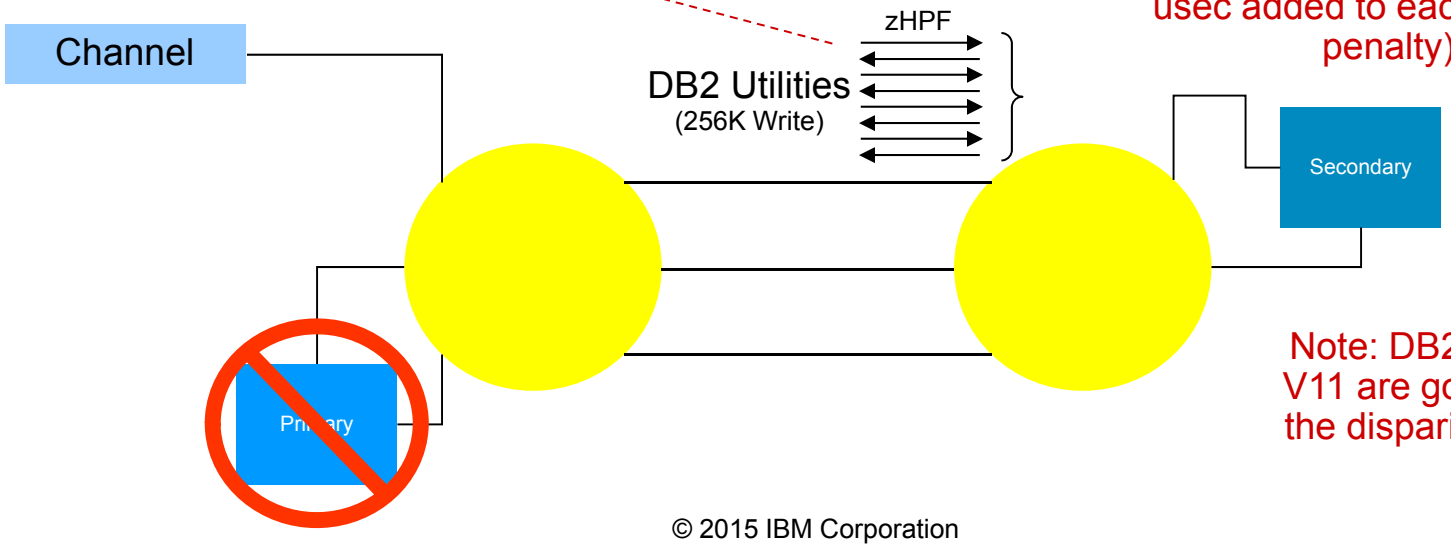
Pre-zHPF Extended Distance II – Experimental Results

256K Writes – Typical of DB2 Utilities, ISV products, etc.



Pre-zHPF Extended Distance II

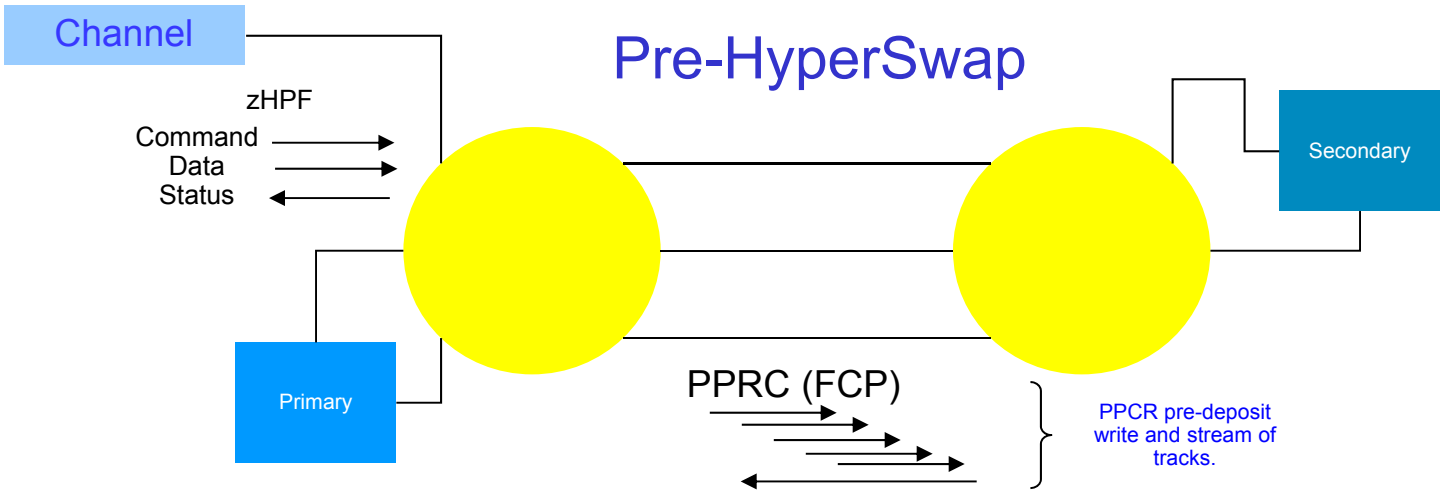


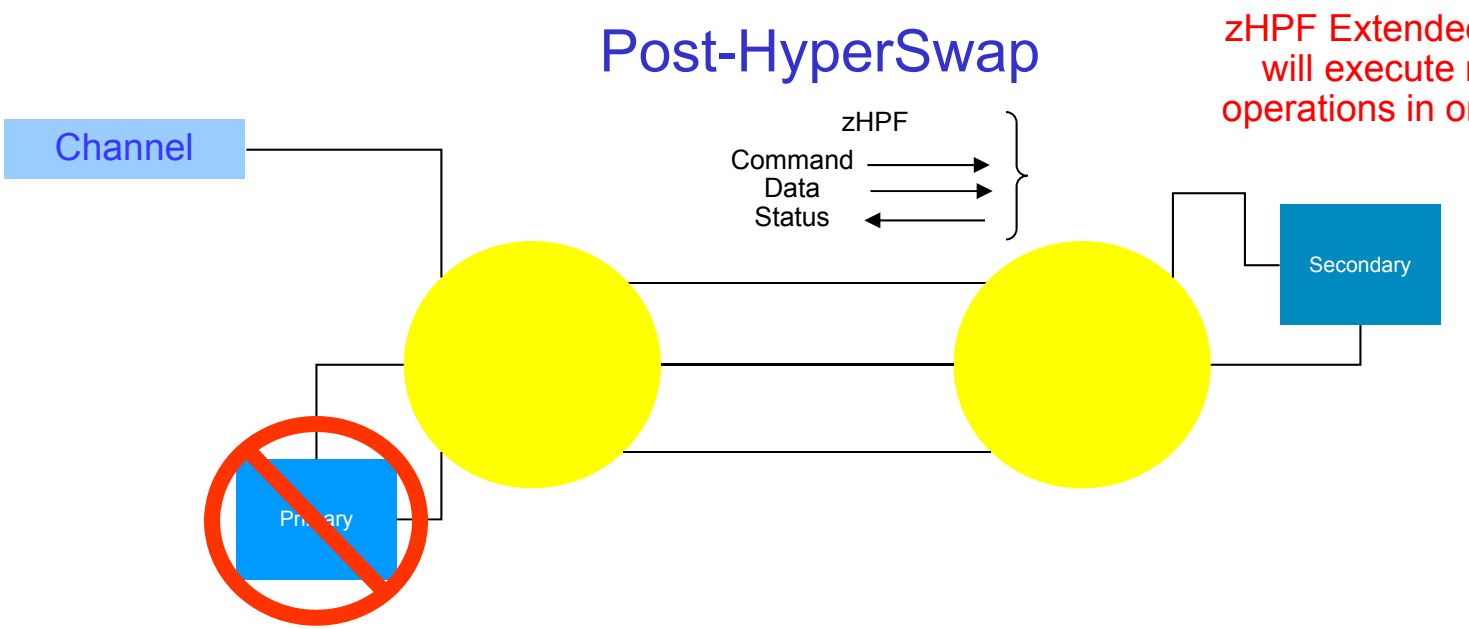


At least four interlocked exchanges at long distance. At 10 km, ~400 usec added to each I/O (50% penalty).

Note: DB2 utility writes in V11 are going to 512K, so the disparity will be worse

zHPF Extended Distance II





zHPF Extended Distance II

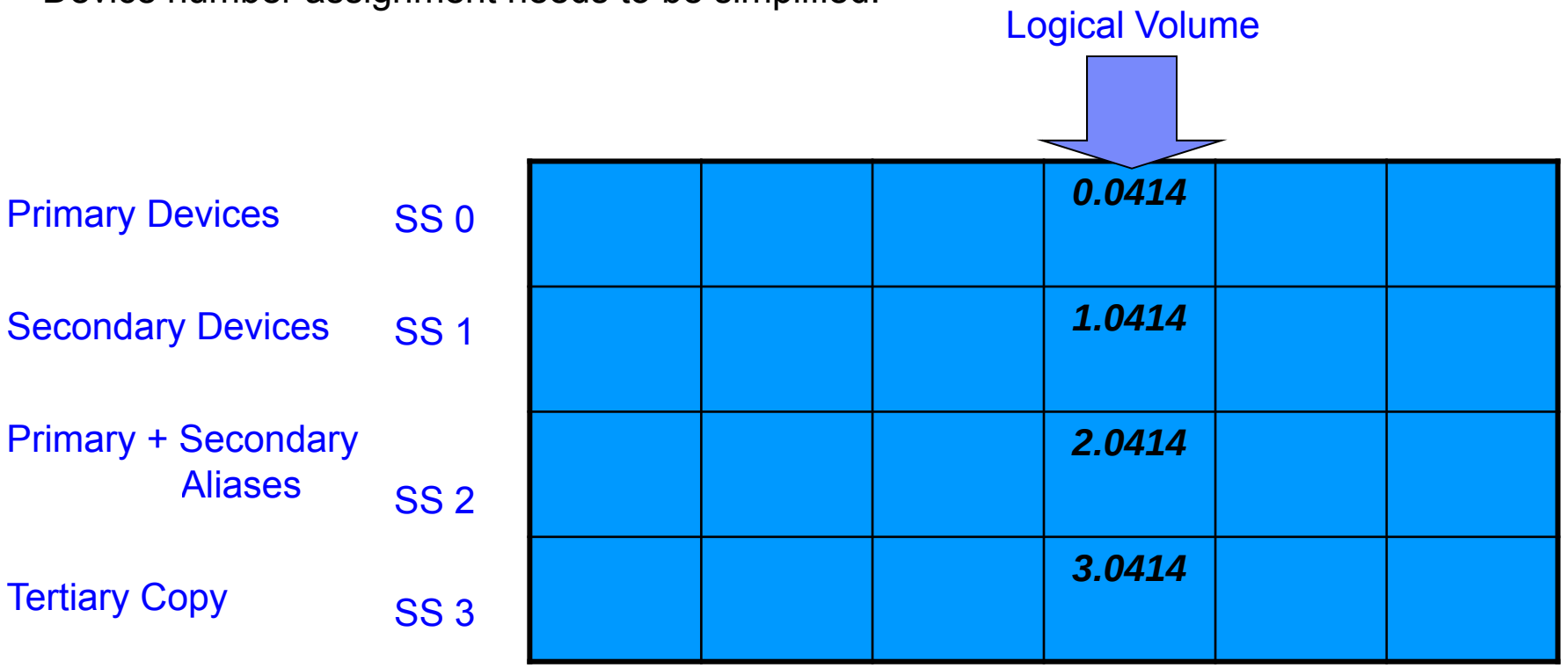
- RMF Monitor I & II data gatherers are enhanced to request new Format-2 Secondary-Queue Measurement blocks whenever the Store-Channel-Subsystem Characteristics response block indicates that the extended I/O-measurement-block-2 facility is installed

- The I/O Queuing configuration data sections of SMF record 78 subtype 3 and SMF record 79 subtype 14 are extended to provide new transport-mode related performance counters for each configured CHPID
  - Transport-Mode-Write-Count
  - First-Transfer-Ready-Disabled write count

- **First-transfer-ready-disabled inhibited ratio is reported through new RMF Postprocessor Overview condition**

Fourth Subchannel Set

Maintain HyperSwap readiness after the primary or a secondary fails.
Device number assignment needs to be simplified:



Compliments multi-target PPRC by simplifying the configuration changes needed to define 3rd copy of data in large configurations.

Compliments Global Mirror by facilitating return to the primary site after the after disaster is over

I/O Serial Number (WWPN) Migration Forward Order Process

- **Problem: FCP World Wide Port Names are determined based on:**
  1. the IO serial number of the CPC,
  2. the IOCDS configuration details (for NPIV WWPNs), and
  3. PCHID values (for physical WWPNs).

  When a customer purchases a new ‘replacement ‘ CPC and does a ‘push-pull’ (maintains current IO Configuration definitions), the IO serial number changes as most likely do the PCHID values.

  This means that the FCP WWPNs will change, and the customer will need to reconfigure zoning in their Storage Area Network (SAN) switches and LUN masking in their storage controllers. This is tedious, time-consuming, and error-prone.

- **Solutions:**
  1. Provide for a feature in the order and fulfillment processes that enables a client to have the option to keep the same IO serial number on their replacement CPC to facilitate WWPN persistence.
  2. Provide a means to export PCHID values of new system into .csv file, assign ‘alias’ definitions to them (to match old system PCHID values), and import edited file

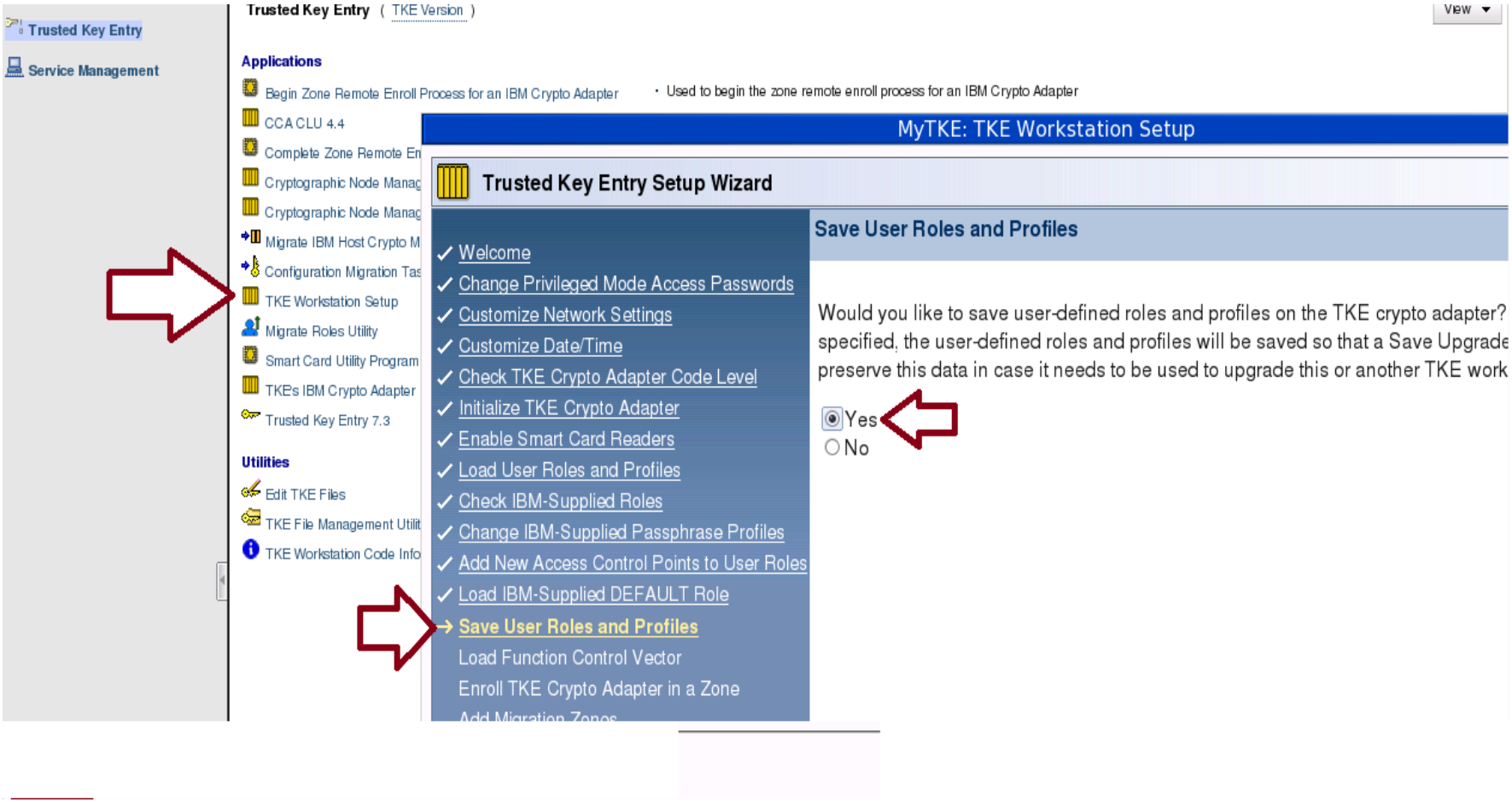
z13 TKE 8.0 migration steps

Additional Slides

(available as additional material – not presented)

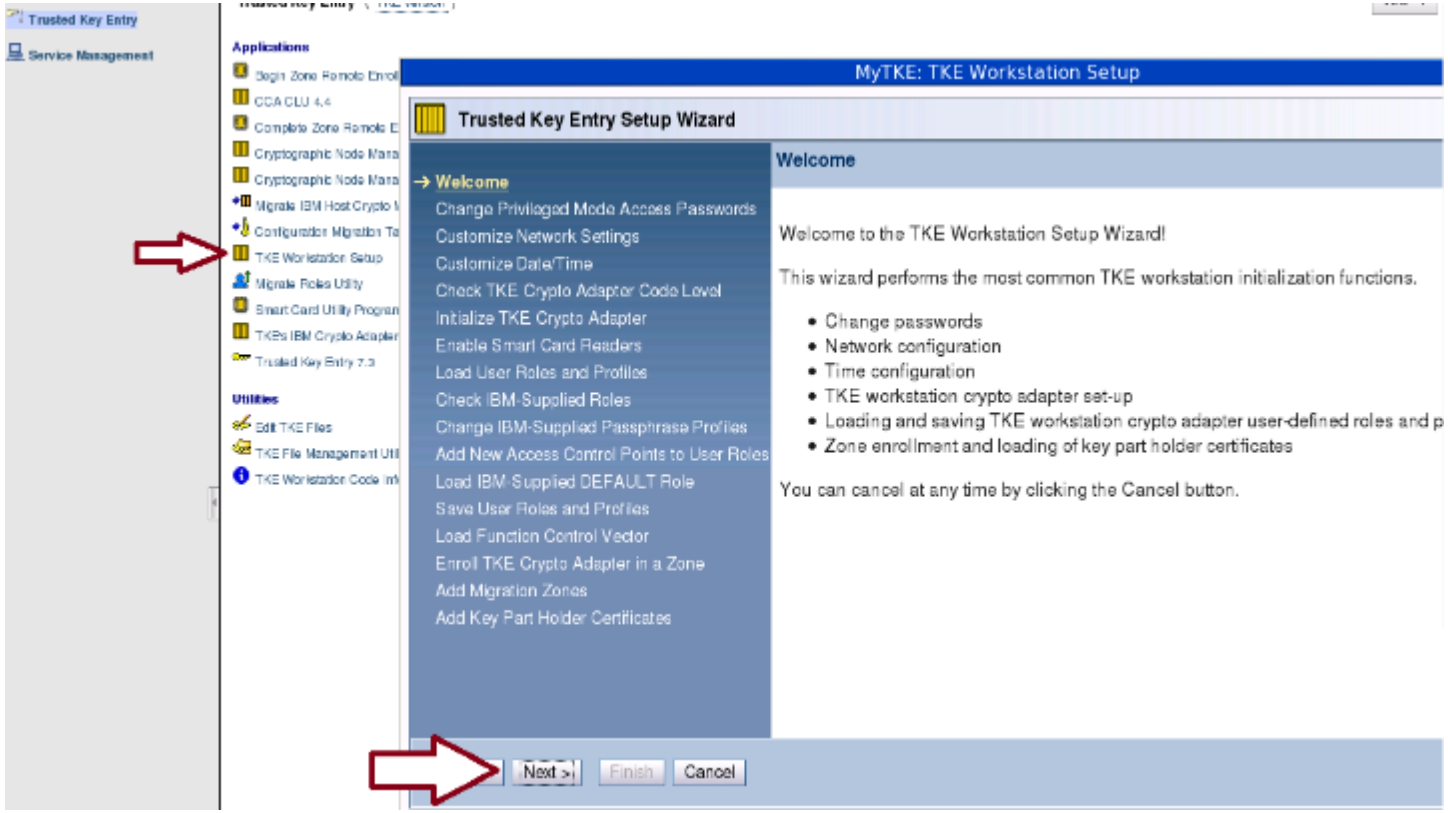
TKE Migration Step 1

- **On TKE 7.3, use the workstation setup wizard to *Save User Roles and Profiles* before the Save Upgrade Date is done**



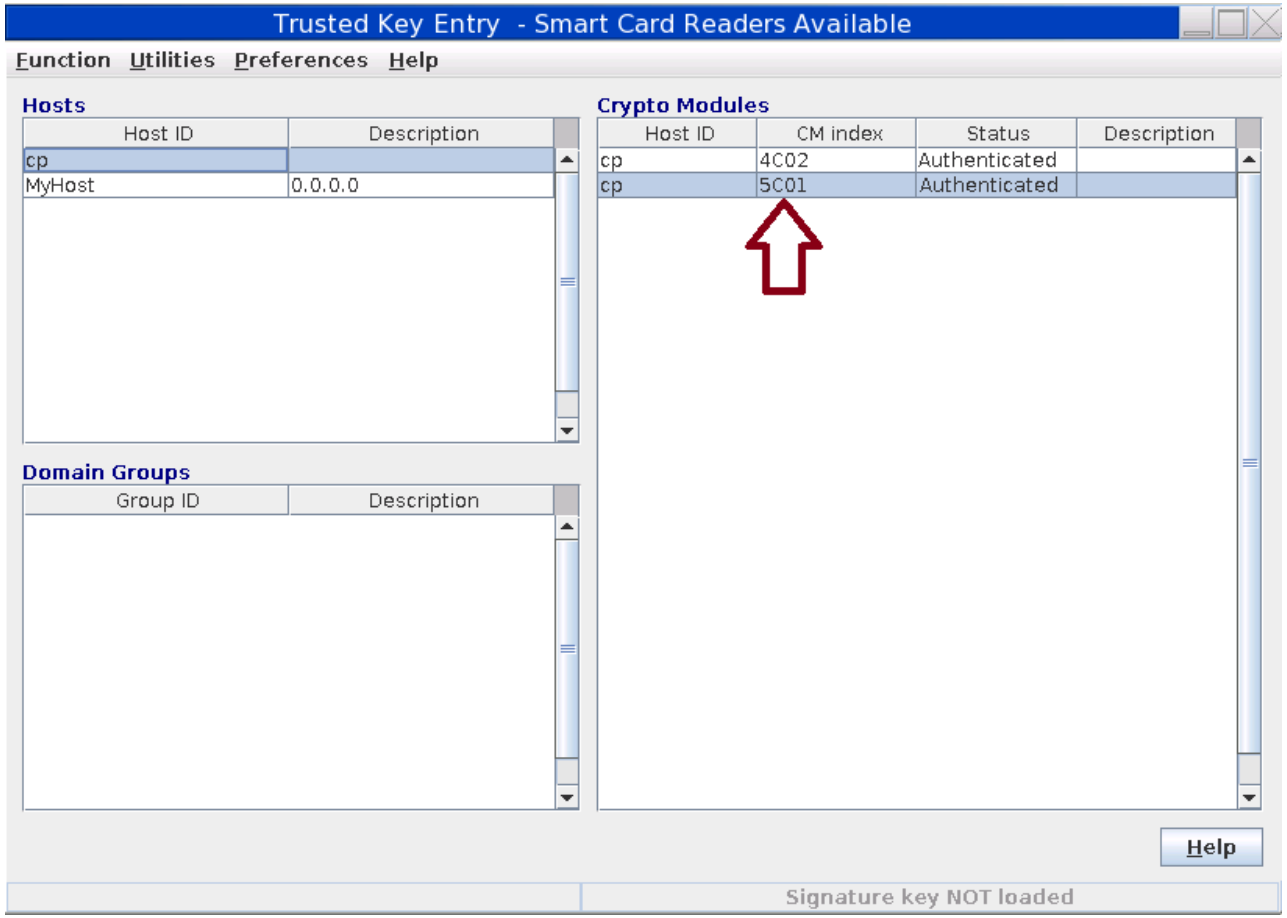
TKE Migration Step 2

- **When a client receives a new TKE 8.0 system, they should run the workstation setup wizard.**
  - Shows the launch point of TKE Workstation Setup wizard and the initial screen.



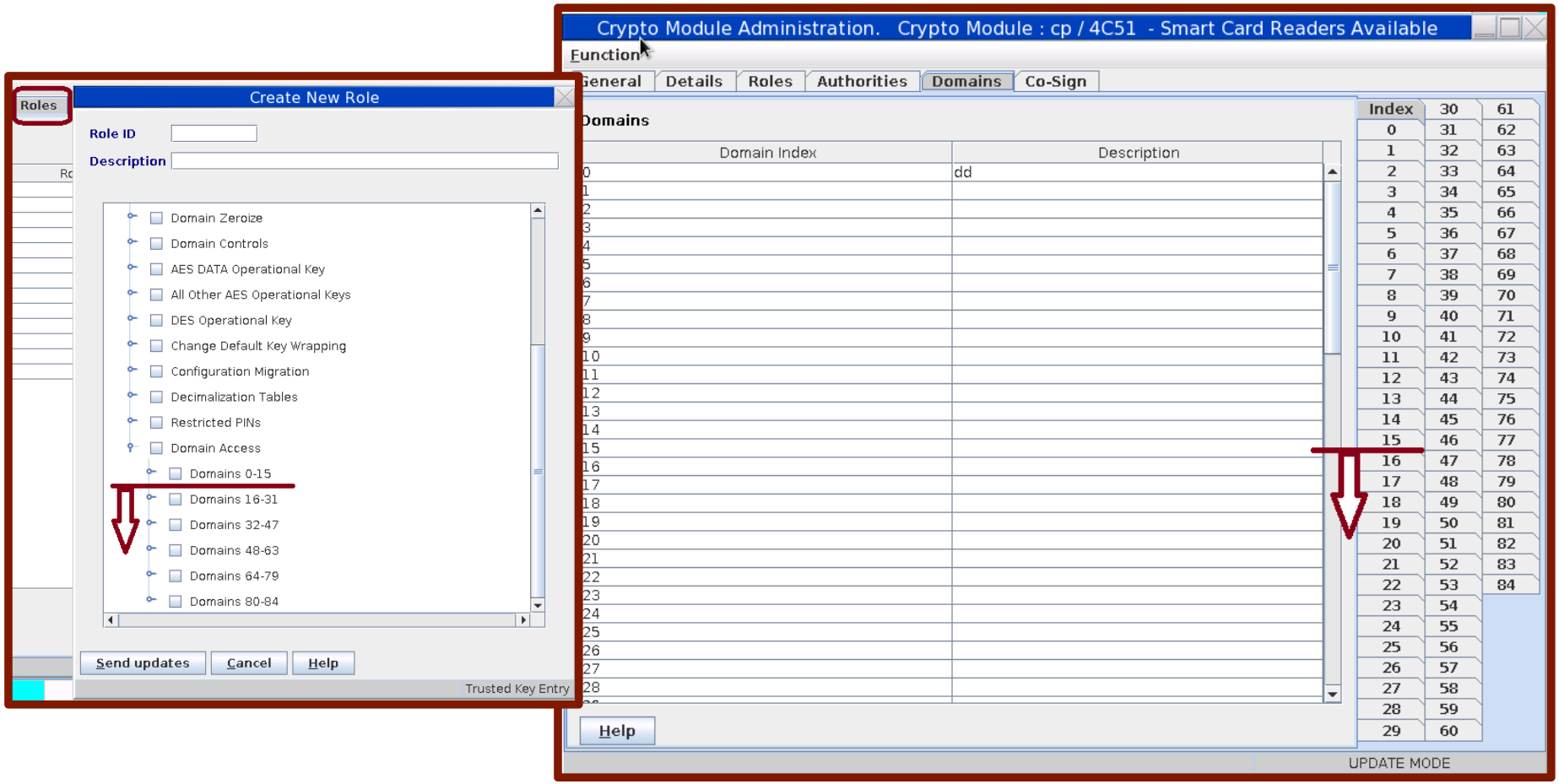
TKE Migration Step 3

- **Only TKE 8.0 can be used to manage a Crypto Express5**
  - Shows a host with a Crypto Express5



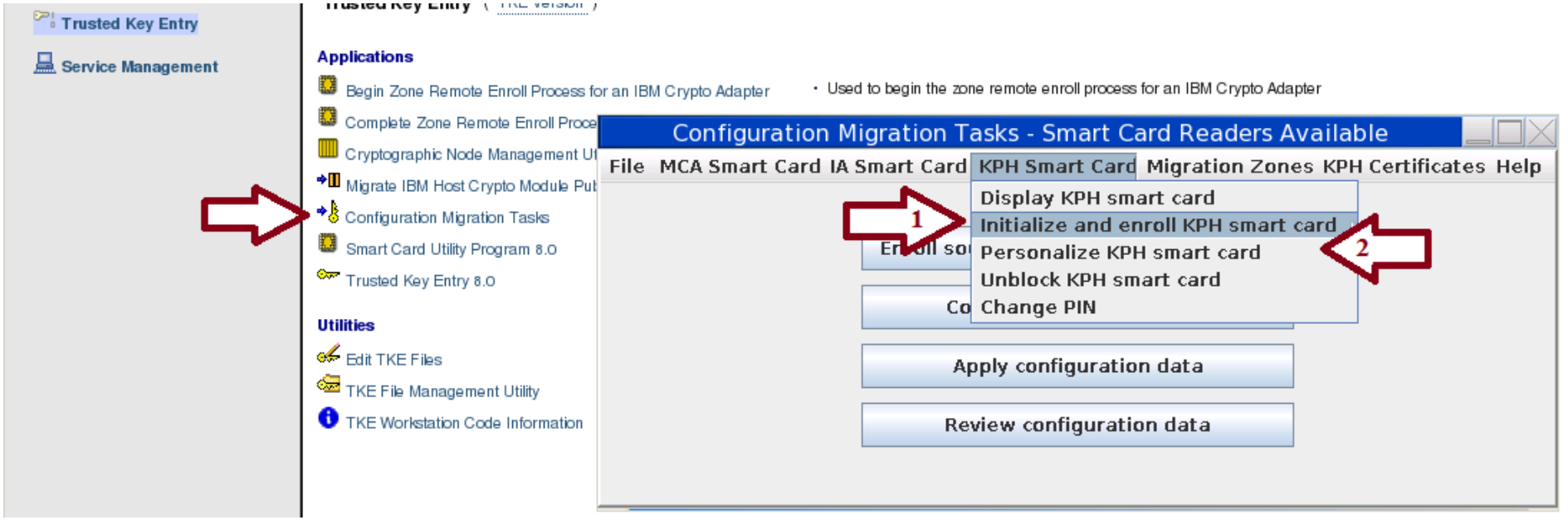
TKE Migration Step 4

- **Only TKE 8.0 can be used to manage a domain greater than 15:**
  - Example impacts: *Domain Access* on roles, and managing *Domains*



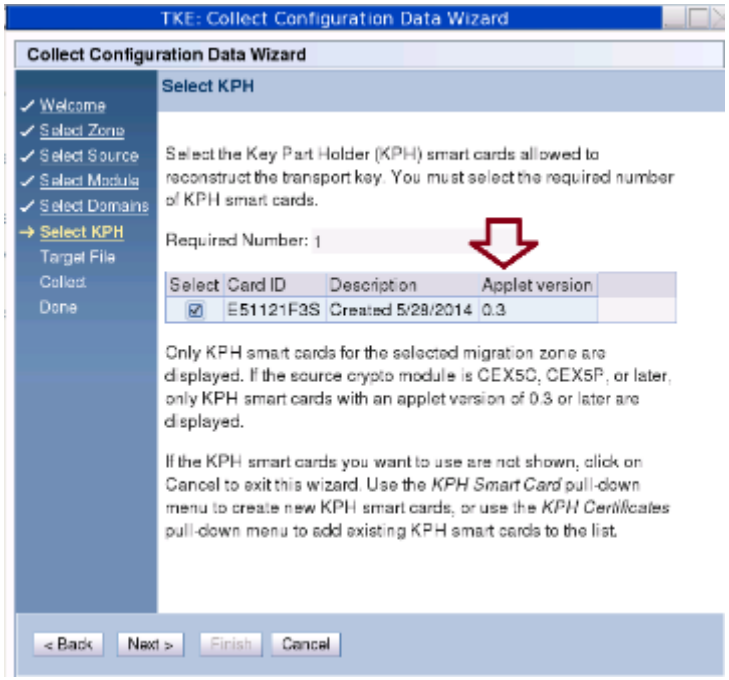
TKE Migration Step 5

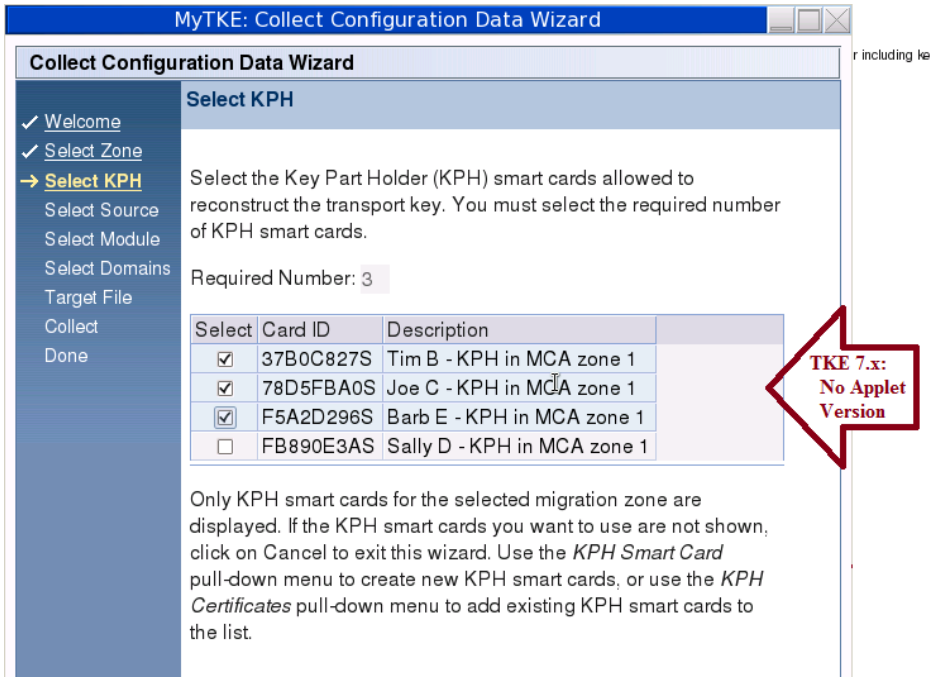
- **Creating Key Part Holder (KPH) smart cards on a TKE 8.0 system**
  - Start the *Configuration Migration Tasks* application
    1) Initialize and enroll KPH smart card
    2) Personalize KPH smart card
       - Note: MCA smart card is needed. MCA created on Pre-TKE 8.0 can be used.



TKE Migration Step 6

- **If collecting data that will be applied to a Crypto Express 5, KPH version 0.3 certificates must be used:**
  - Notes:
    - You can put version 0.3 KPH certificates on Pre-TKE 8.0 systems
    - Only TKE 8.0 exposes the KPH Certificate applet version during the collect
    - Recommended: If data will be applied to a Crypto Express5, do the collect from a TKE 8.0 systems.



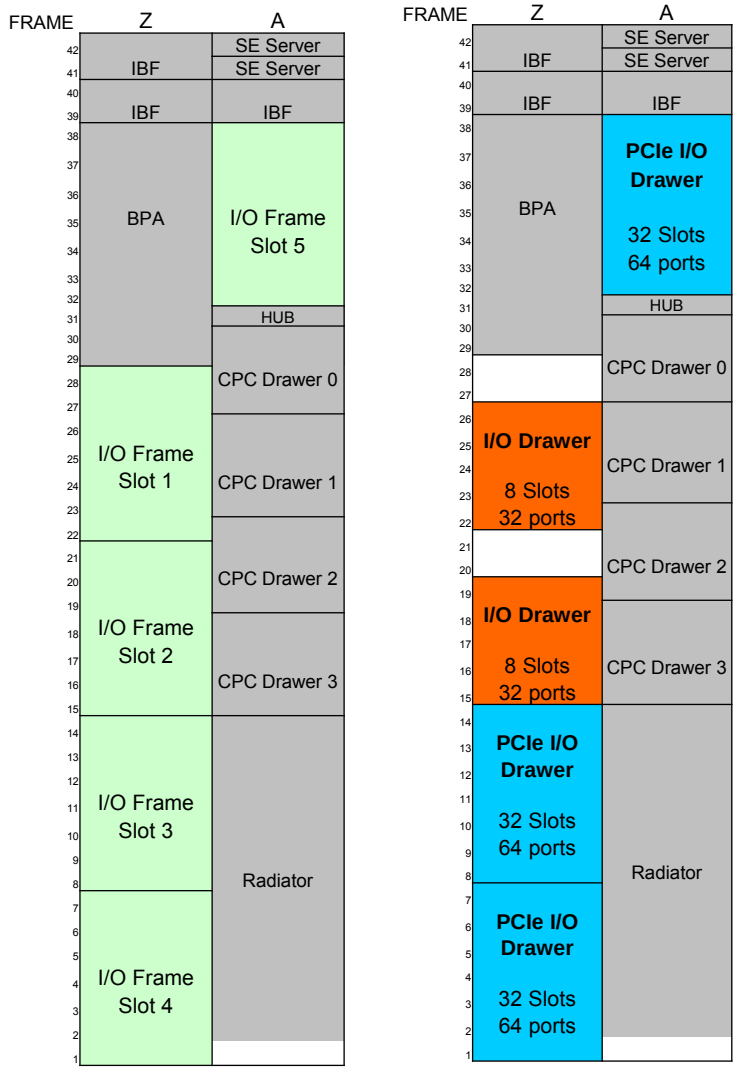


Additional Slides

(available as additional material – not presented)

z13
I/O Layout

z13 Frame Layout for Carry Forward I/O – Air Cooled* System



- An I/O frame slot is a physical location in the A or Z frame for an I/O drawer or PCIe I/O drawer to be inserted = 7u
- PCIe I/O drawer uses 1 I/O frame slot = 7u
  - 32 two port I/O slots = 64 ports
  - 5 drawers maximum = 160 slots
- I/O drawer uses 0.7 frame slot = 5u
  - 8 four port I/O slots = 32 ports
  - Requires 2u of free space for future upgrade to the PCIe I/O drawer
  - 2 drawers carry forward only maximum in I/O frame slots 1 and 2 only
- I/O Cage not supported in z13

* Frame locations differ if water cooled; but the number of I/O frame slots is identical.

Carry Forward (Field Upgrade) Rules for I/O Drawers

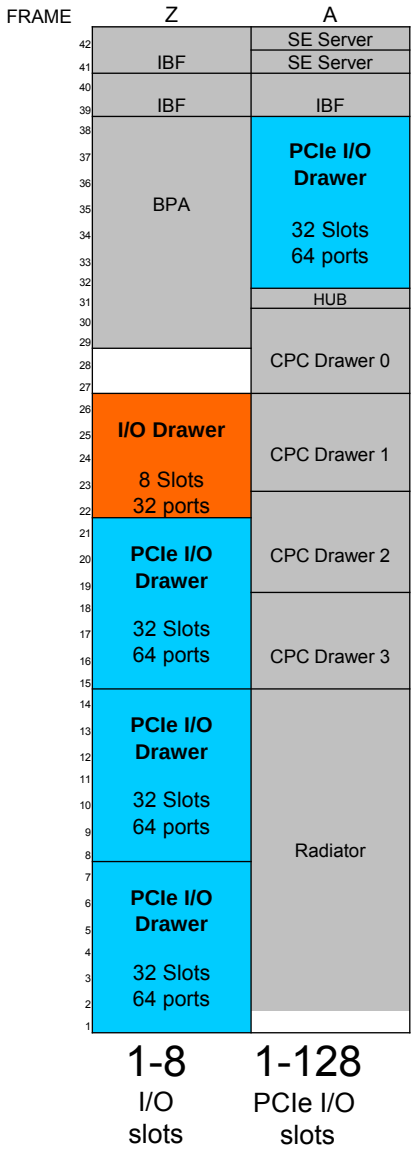
Non-PCIe Features Carried Forward*	8-slot I/O Drawers	Maximum PCIe Drawers/Slots (CF or Add)
0	0	**5/160**
1 to 8	1	**4/128**
9 to 16	2	**3/96**
17 or more	Not Supported!	

Note: The only Non-PCIe Feature that can be carried forward is FICON Express8 LX (FC 3325) and FICON Express8 SX (FC 3326)

Empty slots in a carried forward I/O drawer CANNOT be filled by MES.

z13 I/O Cage/Drawer Offerings with Box MES

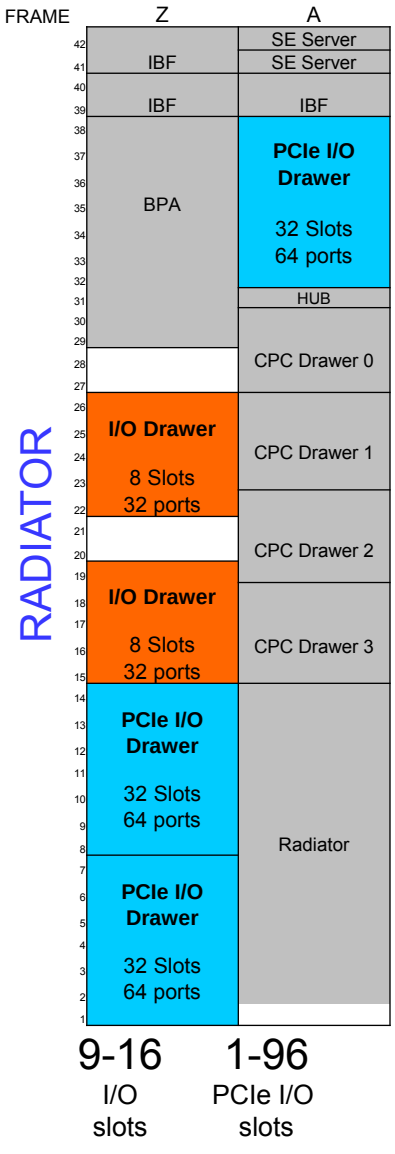
Max I/O Offering w/ 1 I/O Drawer



1-8 I/O slots

1-128 PCIe I/O slots

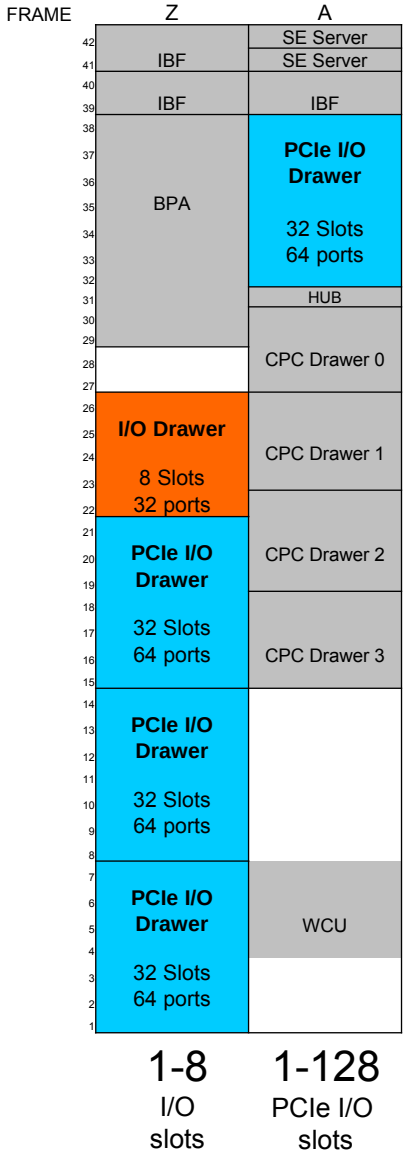
Max I/O Offering w/ 2 I/O Drawers



9-16 I/O slots

1-96 PCIe I/O slots

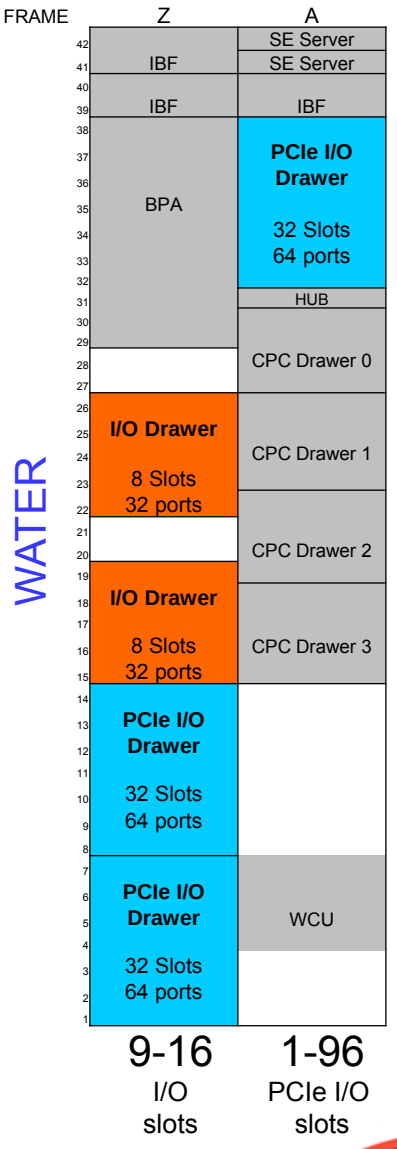
Max I/O Offering w/ 1 I/O Drawer



1-8 I/O slots

1-128 PCIe I/O slots

Max I/O Offering w/ 2 I/O Drawers



9-16 I/O slots

1-96 PCIe I/O slots

Note: Drawings Not to Scale

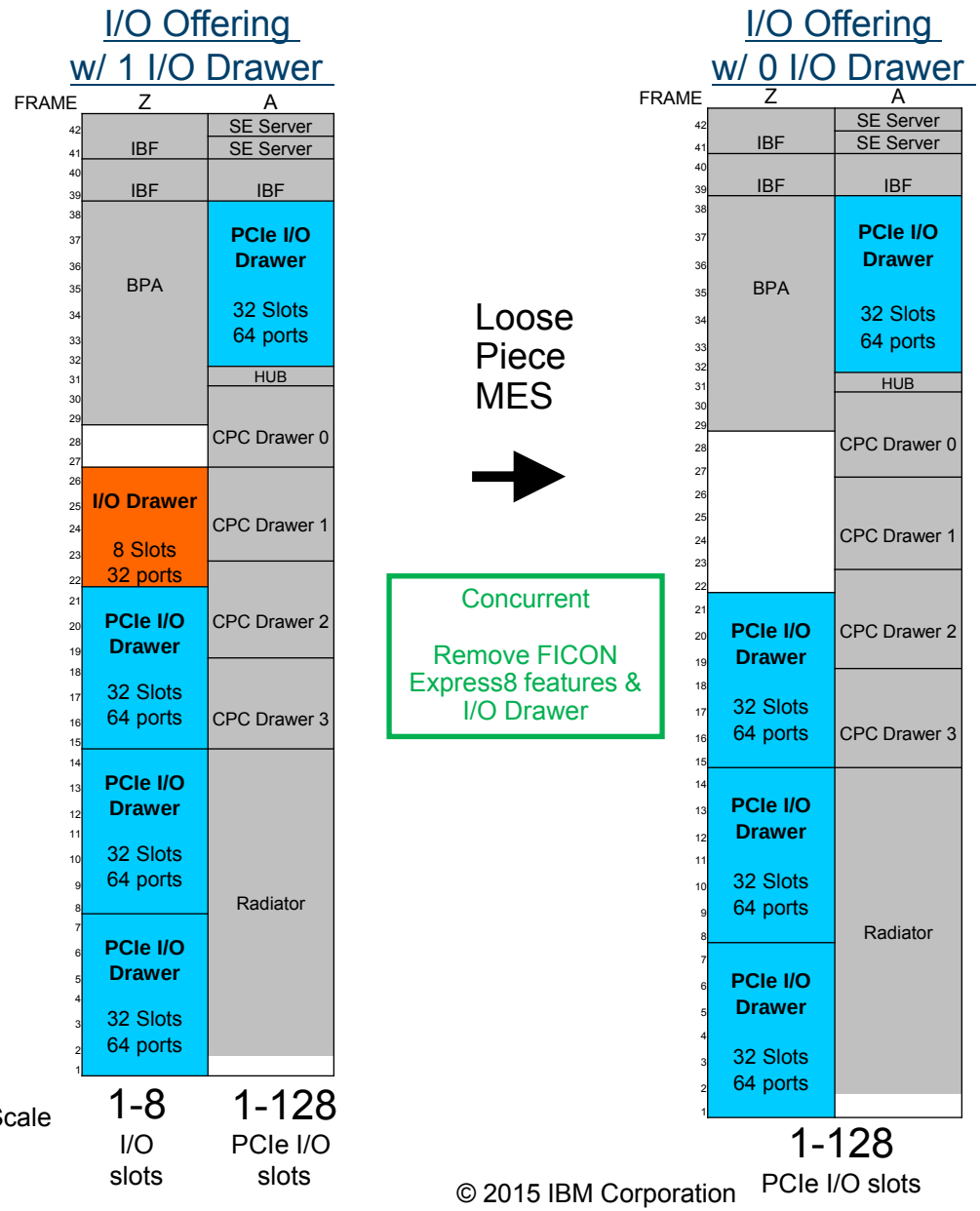
z13 Frame Roll MES

- The only legacy I/O drawer based features that can be carried forward to z13 are (FC 3325) FICON Express8 LX and (FC 3326) FICON Express8 SX
  - Maximum number of carry forward I/O Drawers is 2 which equals 16 features
  - Any empty slots in the I/O drawer(s) after the Frame Roll MES cannot be filled
  - NOTE: >1TB Memory in an LPAR is NOT supported with an I/O drawer present

- If I/O drawers are present in the zEC12 or z196, then the I/O drawer(s) will be physically moved by the SSR to the z13

- If the System to be upgraded (z196/zEC12) configuration does not contain any I/O drawers (may have I/O Cages) or not enough I/O drawers, then the appropriate number will be shipped from manufacturing in the z13
  - See the next chart regarding RoHS exemptions and z13 Frame Roll MES

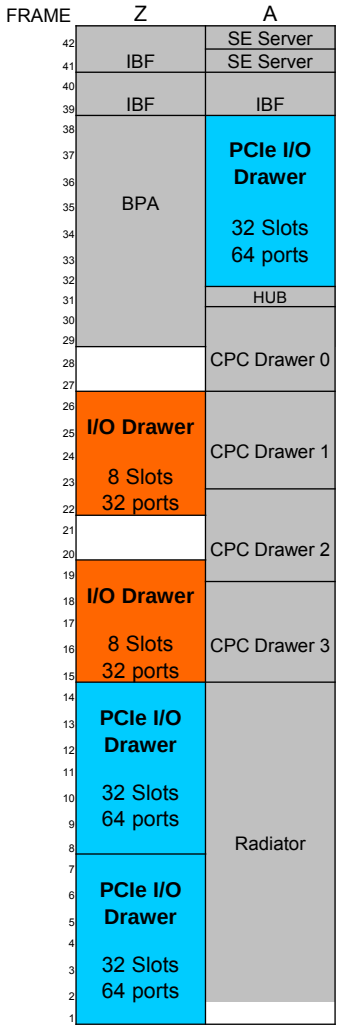
z13 I/O Drawer Offerings with Loose Piece MES



Note: Drawings Not to Scale

z13 I/O Drawer Offerings with Loose Piece MES– Air Cooled

I/O Offering w/ 2 I/O Drawers



Note: Drawings Not to Scale

Loose Piece MES

DISRUPTIVE
Move FICON Express8 features to the remaining I/O Drawer (PCHID changes)

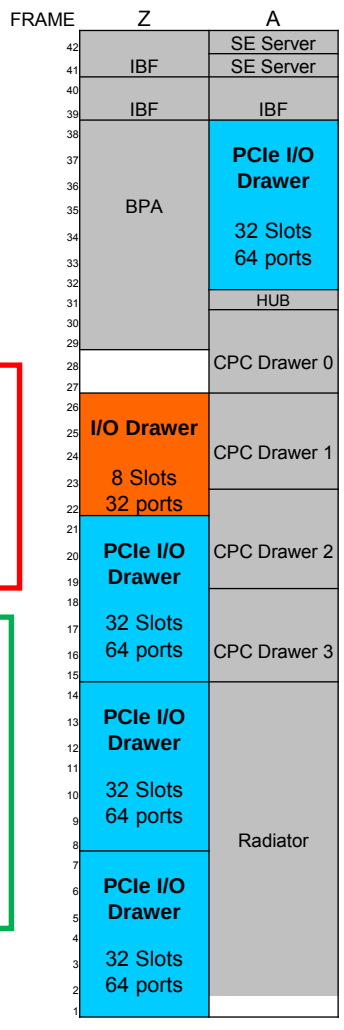
Concurrent

Remove empty I/O drawer

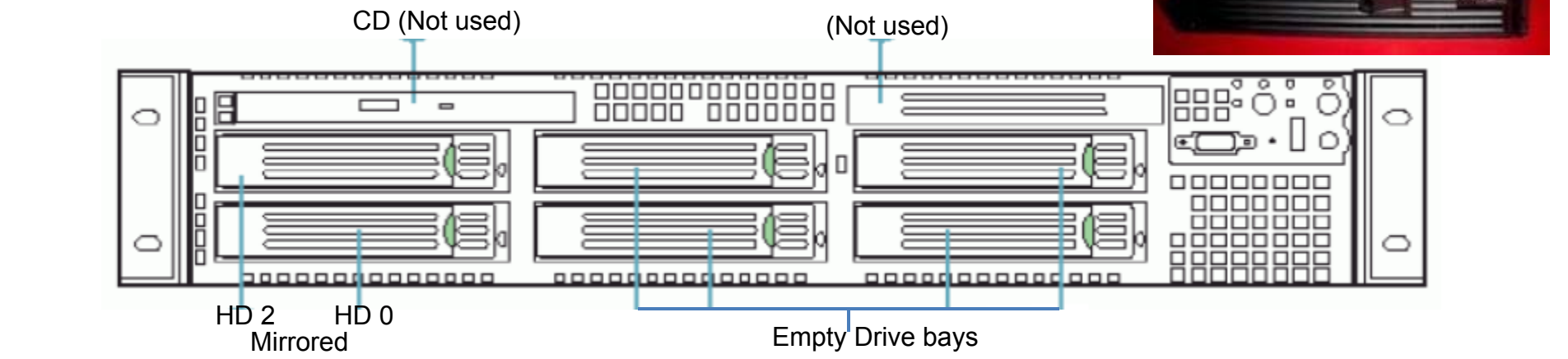
Add 1 PCIe I/O drawer

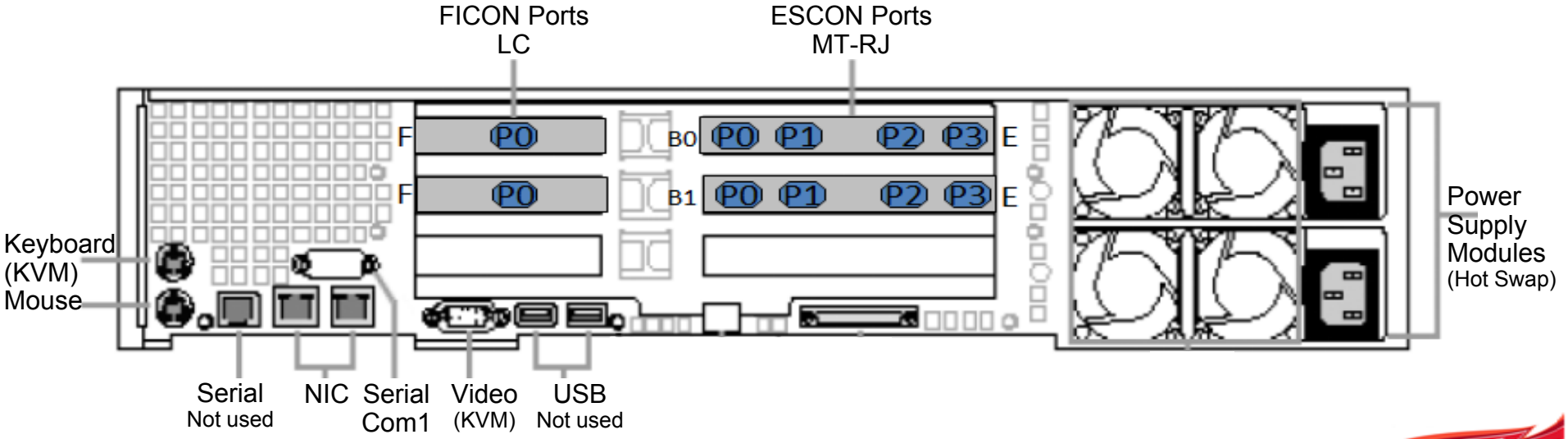
Add PCIe I/O features

I/O Offering w/ 1 I/O Drawer

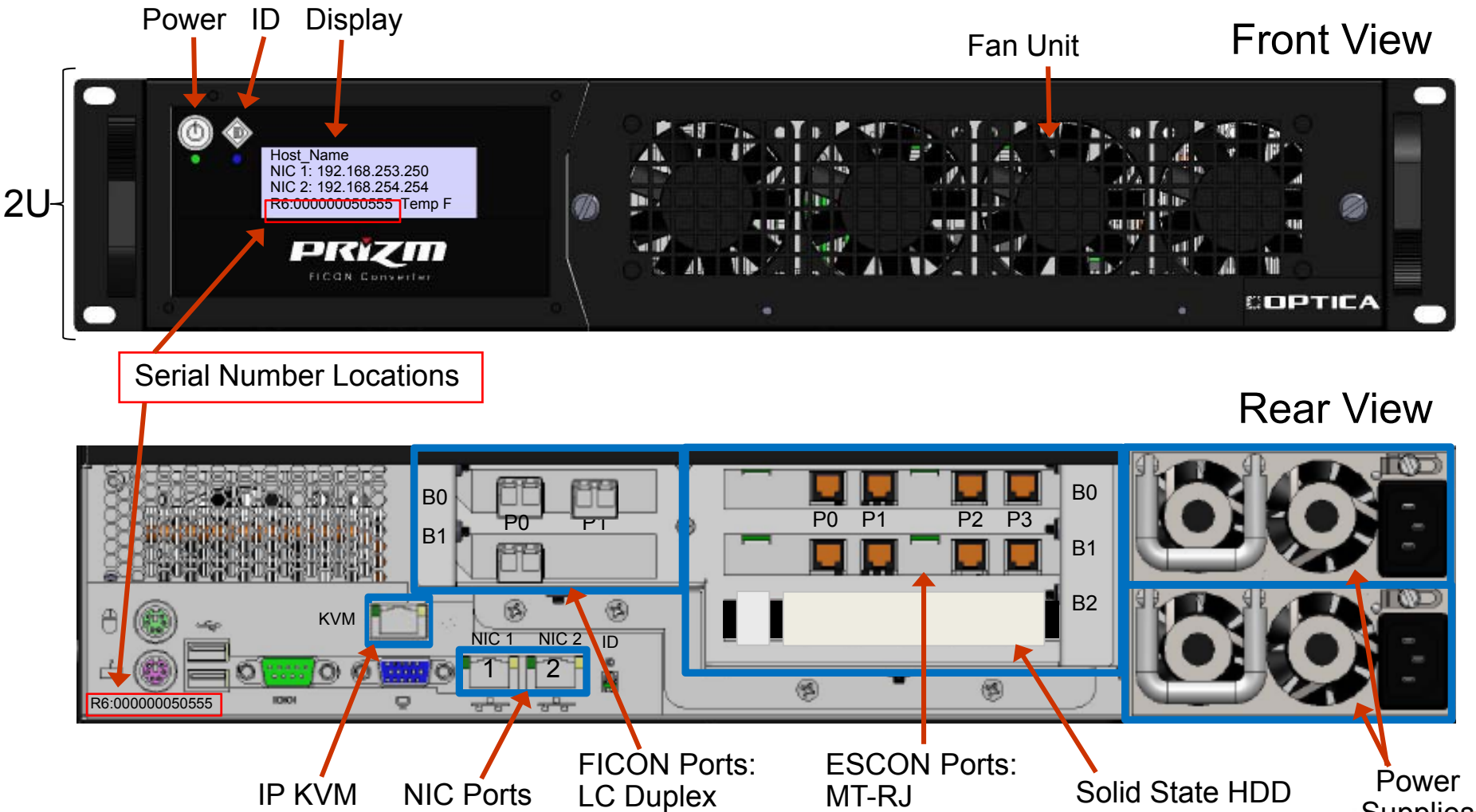


1st Generation PRIZM





2nd Generation PRIZM



z13 Frame Roll MES and RoHS Exemption 7b

- In July 2016, the Restrictions of Hazardous Substances (RoHS) exemption 7b will expire which will impact some z13 Frame Roll MES scenarios
  - 7b) Lead in solders for servers, storage and storage array systems, network infrastructure equipment for switching, signaling, transmission as well as network management for telecommunications
- The RoHS affected components are all associated with the I/O Drawer (the I/O drawer itself, the DCA, MDA, STI-A8 mother card, the IFB-MP daughter card, the HCA2-C fanout) and the FICON Express8 features
- Starting in June of 2016, customers in RoHS countries will find new eConfig rules regarding the carry forward of the FICON Express8 features (FC 3325 & 3326) and the I/O Drawer (FC 4008)
  - If the existing machine does not have any I/O drawers in the zEC12 or z196, then eConfig will not allow the carry forward of any FICON Express8 features
  - If the existing machine has 1 I/O drawer in the zEC12 or z196, then eConfig will allow the carry forward of up to 8 FICON Express8 features
  - If the existing machine has 2 I/O drawers in the zEC12 or z196, then eConfig will allow the carry forward of up to 16 FICON Express8 features
- For non-RoHS countries, manufacturing will ship any missing I/O Drawers for the remainder of the z13 platform

z13 “Loose Piece” MES Scenarios Post GA

- Any loose piece MES additions after z13 GA will be for PCIe I/O Drawer features ONLY

- The only support for z13 loose piece MES for legacy I/O will be to decrement the FICON Express8 features (FC 3325 & 3326) which can be done concurrently
  - If a machine config has >8 FICON Express8 features (i.e. 2 I/O drawers) and the customer decrements the FICON Express8 features to <8, eConfig will remove cards across both the I/O drawers and the 4 total domains.
  - If the new config requires moving FICON Express8 cards into a single drawer from both I/O drawers (in order to replace an I/O drawer with a PCIe I/O drawer) then the MES will be DISRUPTIVE because the moving of any I/O card from one PCHID location to any other PCHID location is a disruptive move
  - If the new config is completely removing all the FICON Express8 cards, then the MES can be done concurrently
    - Remember that in order to support >1TB memory in a single LPAR, the I/O drawers must be completely removed from the system not just empty with no FICON Express8 cards in them

z13 I/O Subsystem Enhancements with IBM Storage

- **GOALS**
  - **Performance**
    - Measureable latency Reduction for DB2 transactions
    - Substantial throughput improvement for database logs including DB2 and IMS
  - **Batch Window Reduction**
    - Add client value differentiation from prior generations with higher I/O throughput with the same HW footprint, cabling infrastructure and architectural addressing limits
  - **Scale**
    - More devices per channel, larger devices
    - More logical channel subsystems and LPARs
  - **Resilience**
    - Lead in Software Defined Environment (SDE) by extending WLM policy based management for I/O into the SAN fabric across all z Systems servers, increasing scale and enhanced resilience for mainframe data
    - Reduce impact to production work when I/O components fail
    - Reduction in false repair actions
    - Improved resilience by allowing automatic re-routing of channel traffic through SAN after switch failures
    - Simplify migration to new machines for FCP users

Supporting Technologies

- *Managed File Transfer Acceleration*
- *zHyperWrite for DB2*
- *New Tooling for I/O Resilience*
- *FICON Expres16S*
- *Forward Error Correction Codes*
- *FICON Dynamic Routing*
- *Fabric I/O Priority*
- *zHPF Extended Distance II*
- *FICON IU Pacing Enhancement*
- *SPoF Elimination w/Storage*
- *32K UA/Channel, 6 LCSS, 4 SS*
- *WWPN Preservation II*

Crypto support in z/OS