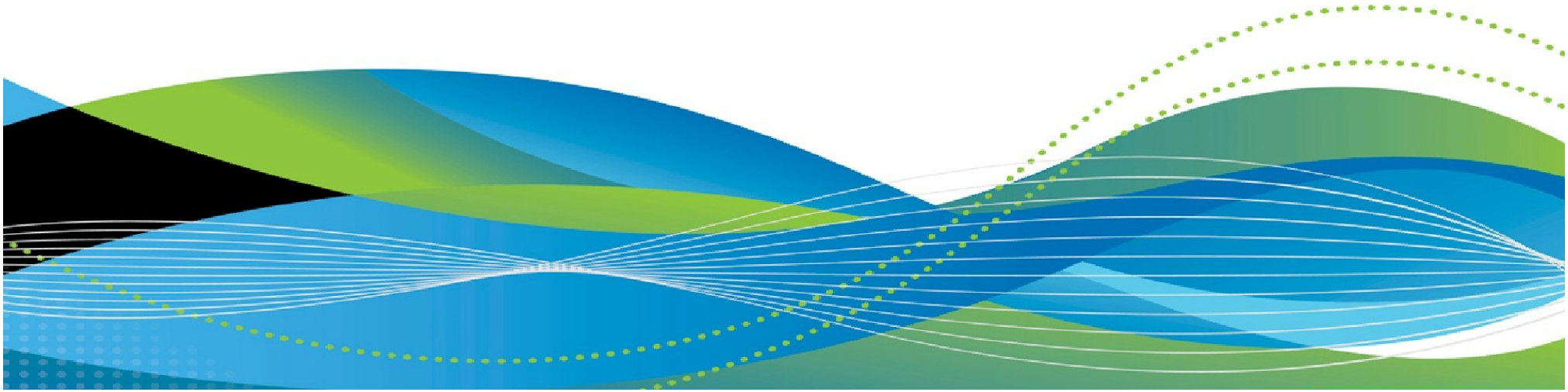
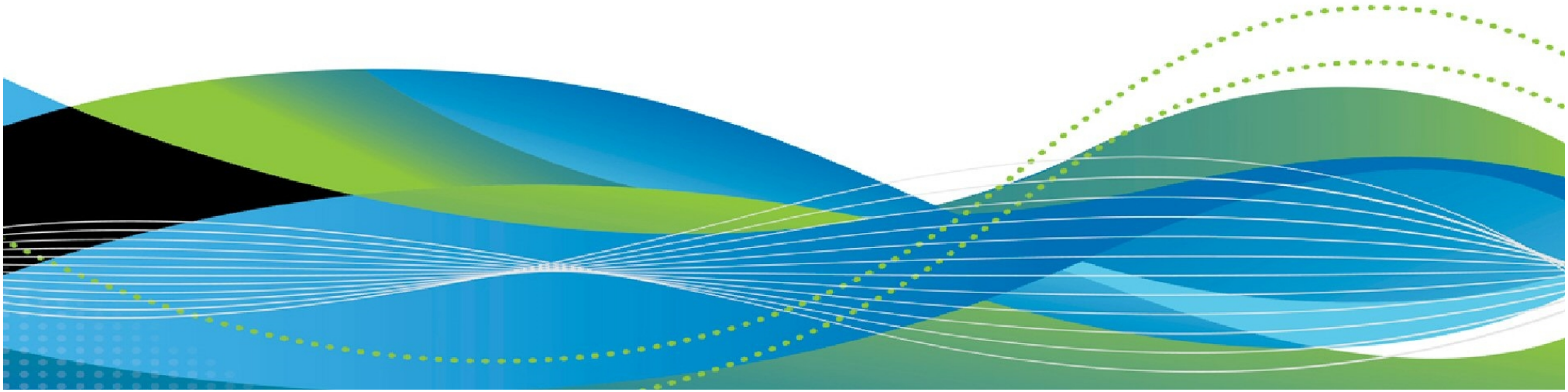
Introduction to IBM Z Hardware



What are we going to learn today?

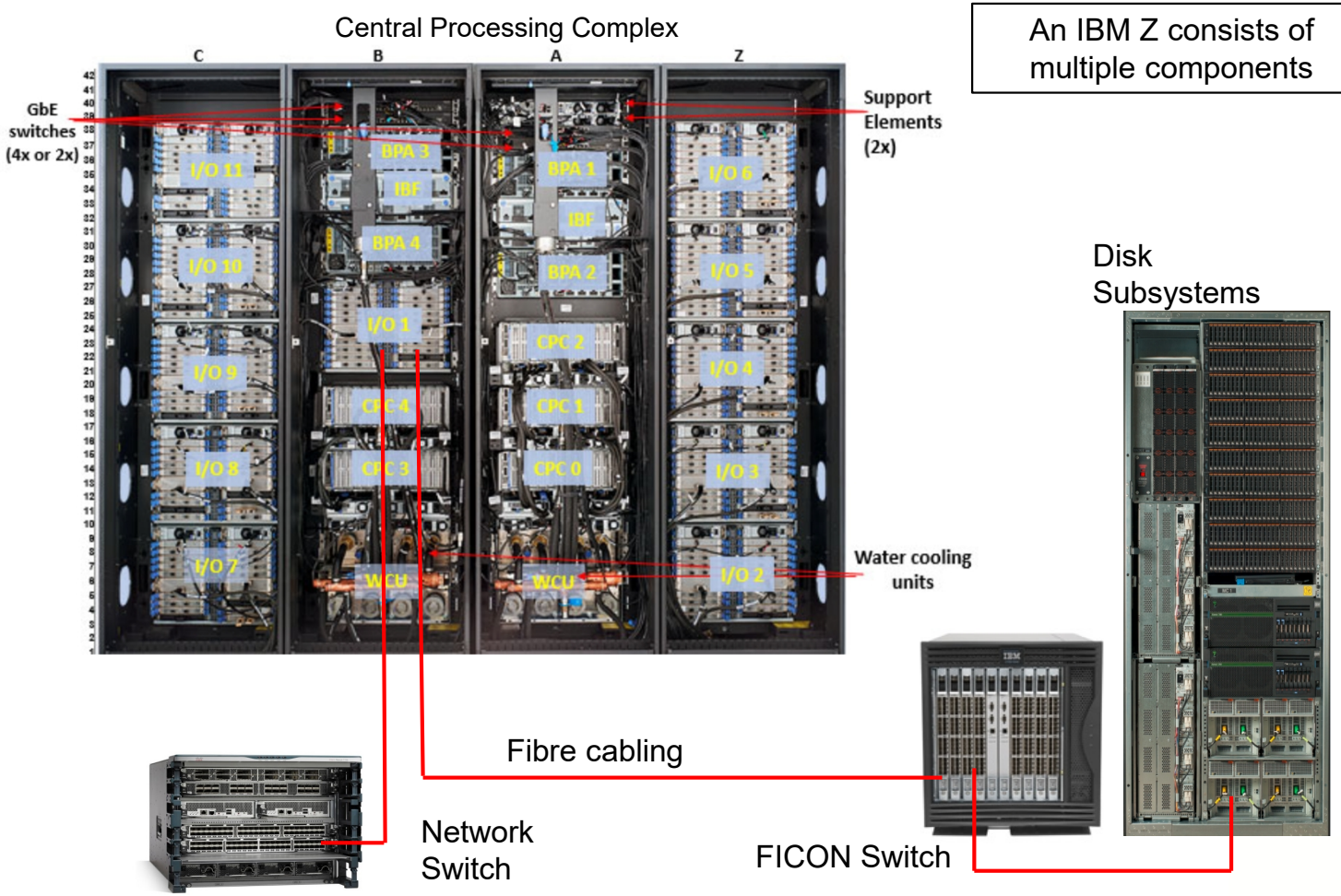
- Introduction
- Parts of the Central Processing Complex
  - Support Elements
  - Processors
  - Memory
  - I/O
- Disk Subsystems
- Hardware Management Console
- IOCP and LPARs

Introduction



What Makes IBM Z Special? Innovation Across the Stack!

An IBM Z consists of multiple components



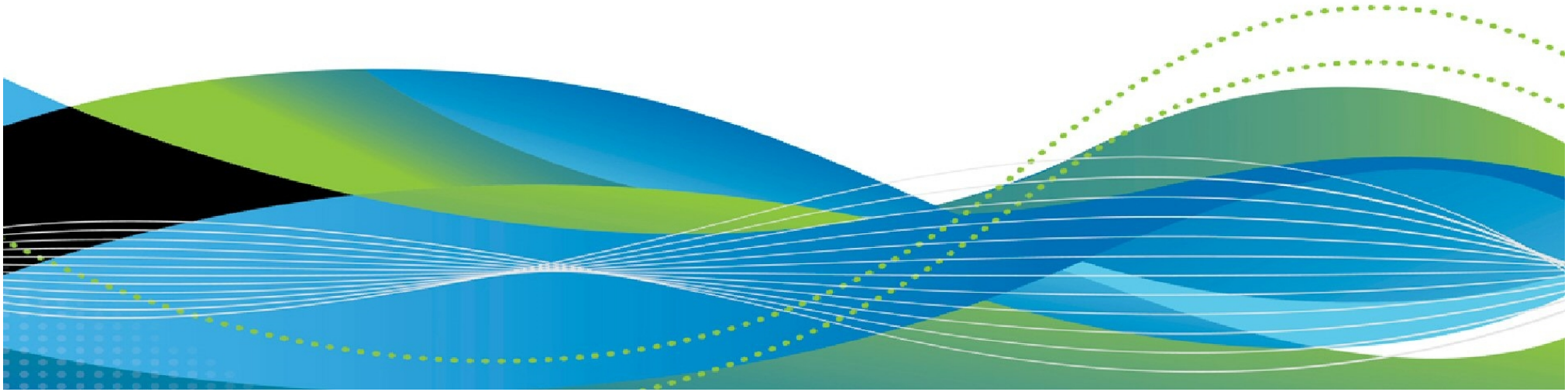
Interactive Processor Demos

- IBM Interactive Product Tour Catalog

Z15-T02

Z17-ME1

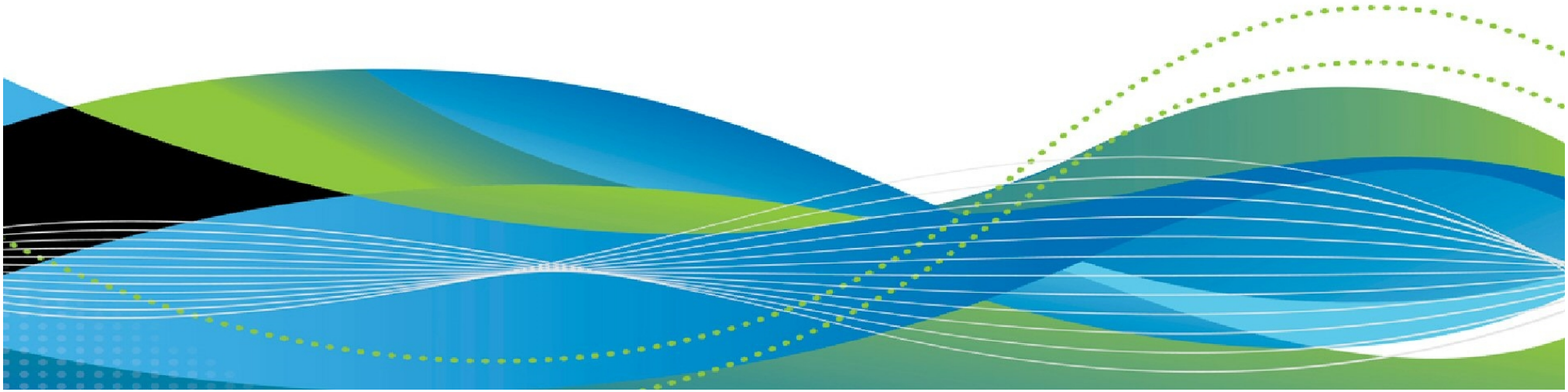
Support Elements



Support Elements, Displays, Keyboards

- Provide platform management functions
- Manage power-up and power-down
- Hold product firmware
- Monitor all system functions
- Service the system
- Two fully redundant appliances

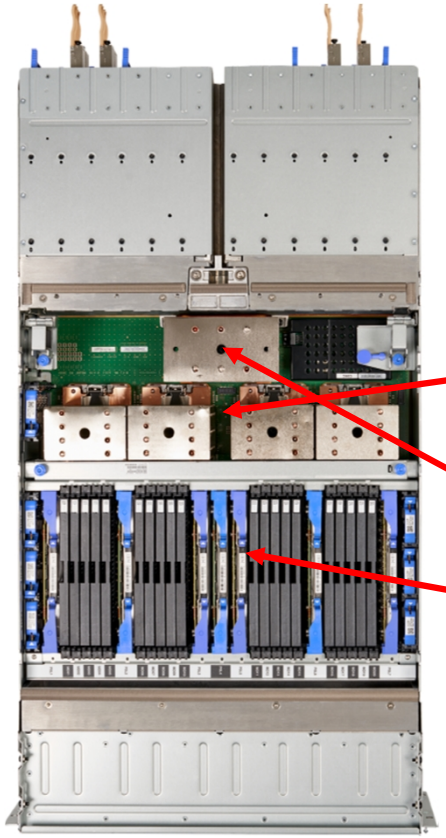
Cores



CPC Drawers

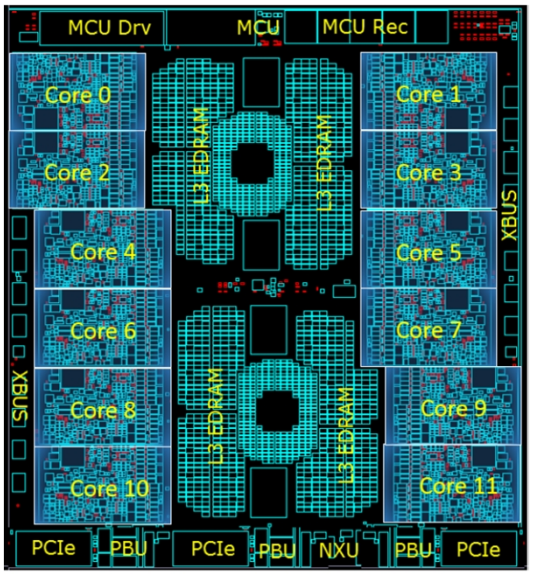
- Packaging to contain processors, memory, and connectors to:
  - PCIe I/O drawers
  - I/O drawers through InfiniBand features
  - Coupling links to other CPCs
- Single z15 can have up to 5 CPC drawers
- Facilitates non-disruptive changes

Z15 Processor Drawer (Top View)



- Each PU SCM:
  - 14nm
  - Four PU SCMs
  - One Memory Controller per PU Chip
  - Five DDR4 DIMM slots per Memory Controller
    - 20 DIMMs total per drawer
- Each drawer:
  - Two logical PU clusters (0 and 1)
  - Up to Four PU Chips per CPC Drawer
    - 12 PUs / SCMs, (7-11 active cores PNs)
    - 8,17,27,38 active PUs per drawer
    - 41 active PUs per drawer – Max4, Max13, Max21, Max31 and Max65
  - One SC Chip (960 MB L4 cache)
  - DIMM slots: 20 DIMMs to support up to 8 TB of addressable memory (10 TB RAIM)
  - Two Flexible Support Processors/ OSC Cards
  - 12 fanout slots for PCIe+ I/O drawer or PCIe coupling fanouts (ICA SR).

Processor Chips – z15

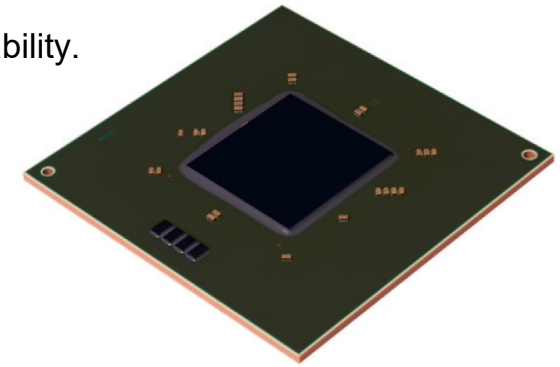


- **14nm SOI Technology**
  - 12 Cores
  - 17 layers of metal
  - 696 mm2 chip area
  - 9.2B transistors versus 6.2B on z14 ZR1
- **20% reduction area**
- **20% reduction in power**

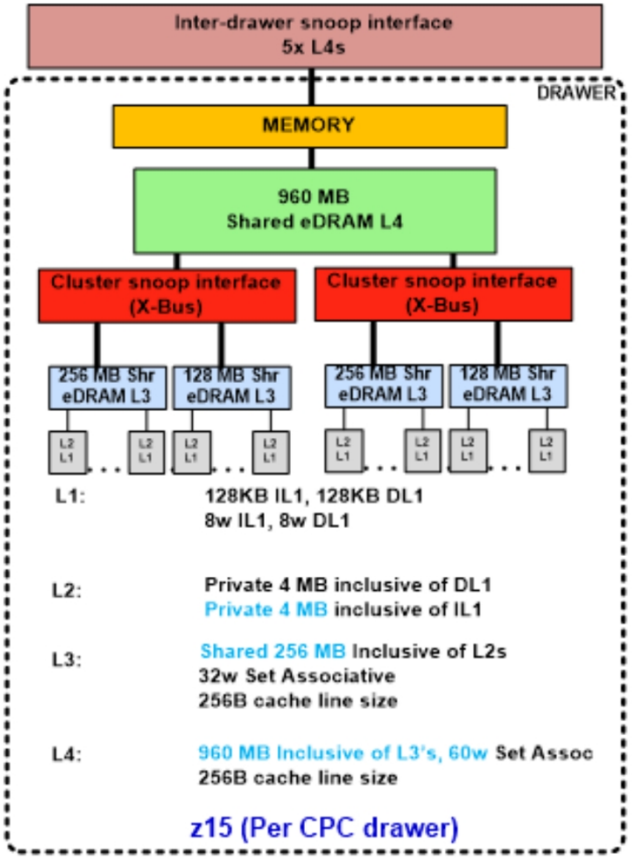
- **4.5 GHz core frequency**
- **7, 8, 9, 10 or 11 active cores per chip**
- **IBM Integrated Accelerator for z Enterprise Data Compression (zEDC)**
  - On-chip compression accelerator (NXU)
- **On Core L1/L2 Cache**
  - L2-I from 2MB to 4MB per core
- **On chip L3 Cache**
  - Shared by all on-chip cores
  - Communicates with cores, memory, I/O and system controller single chip module.
  - L3 from 128MB to 256MB per chip
- **I/O buses**
  - Each CP chip will support up to 3 PCIe buses
    - PCIe+ I/O Drawer Fanout
    - ICA SR 1.1 Coupling Links

Processor Chips

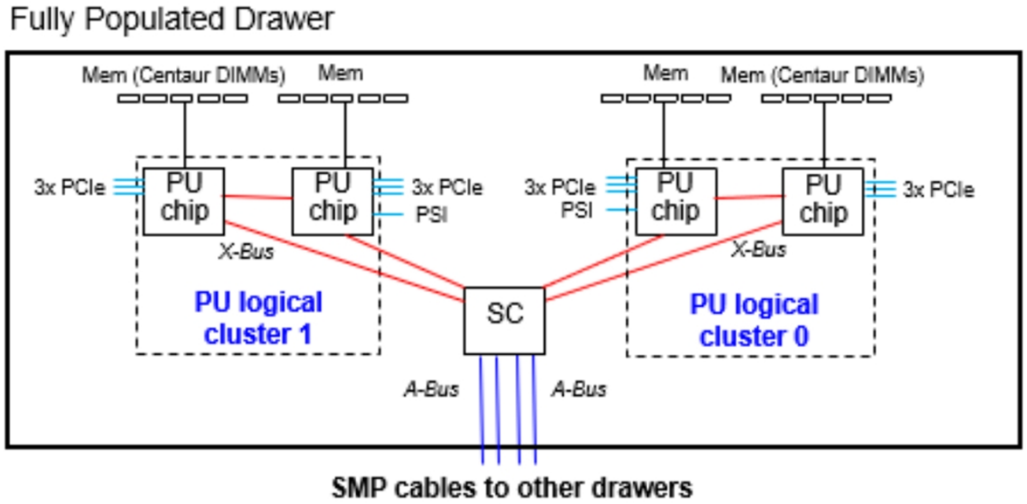
- Each core “characterized” by firmware for a particular use
  - Spare – operations moved to a spare in case of core failure
  - System assist processor (SAP) to handle I/O operations
  - Integrated firmware processor
  - zIIP – z Integrated Information Processor to run eligible Java, z/OS XML, DB2 DRDA, or z/OS Com Server IPSec workloads
  - IFL – Integrated Facility for Linux processor to run z/VM and Linux workloads
  - ICF – Internal Coupling Facility processor
  - CP – general purpose processor for all processing including sub-capacity capability.

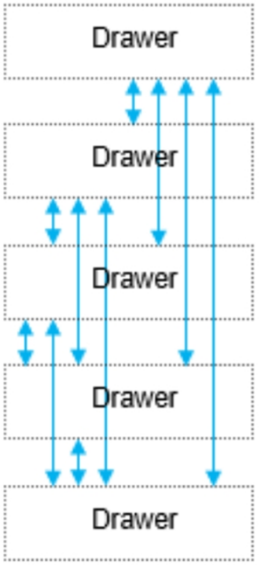


z15 CPC Drawer Cache Hierarchy Detail



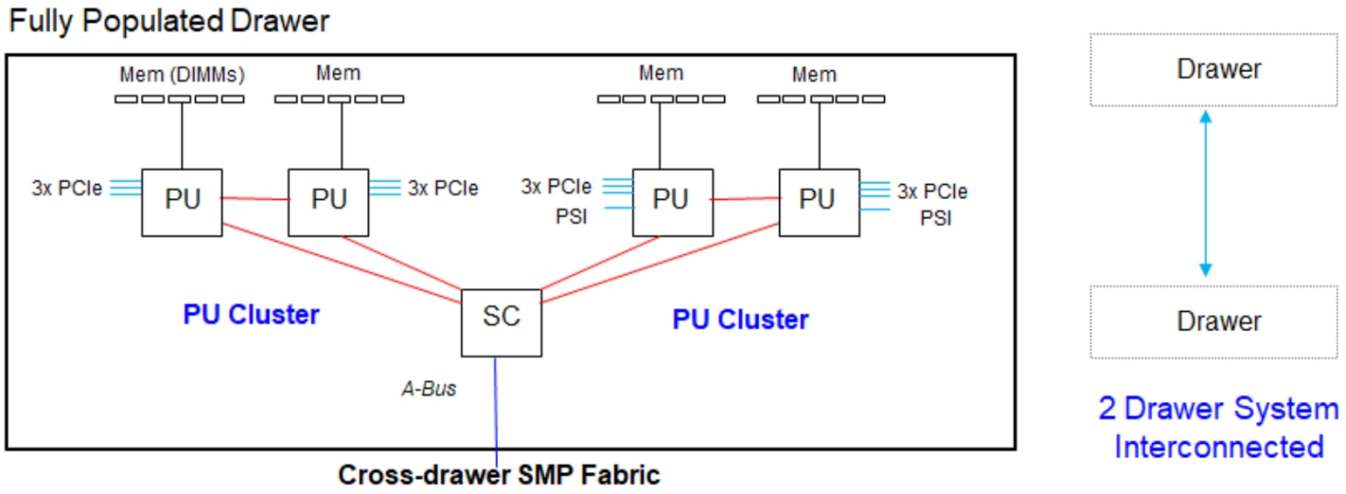
z15 Multi-frame CPC Drawer Structure and Interconnect



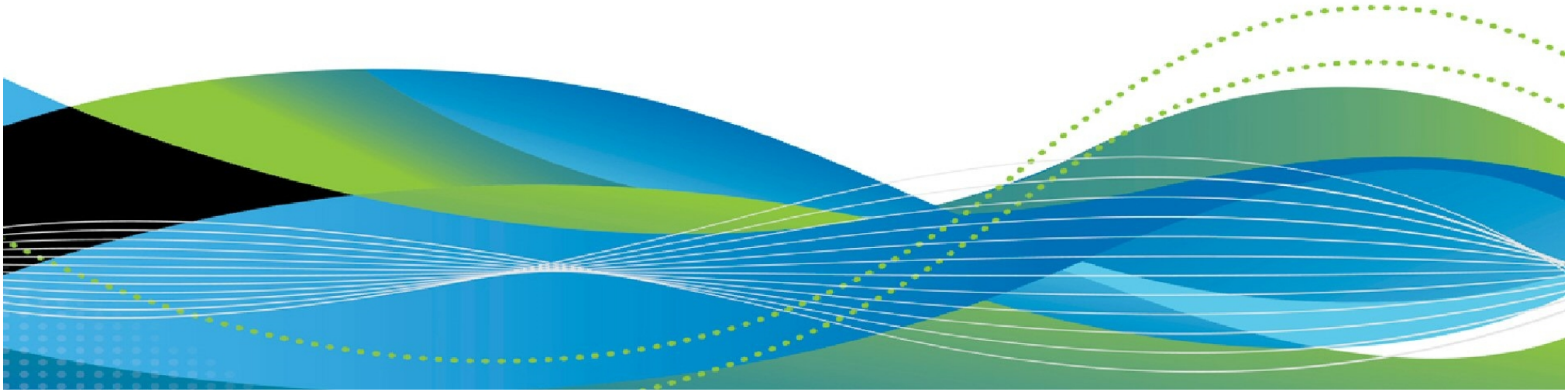


5 Drawer System
Fully Interconnected

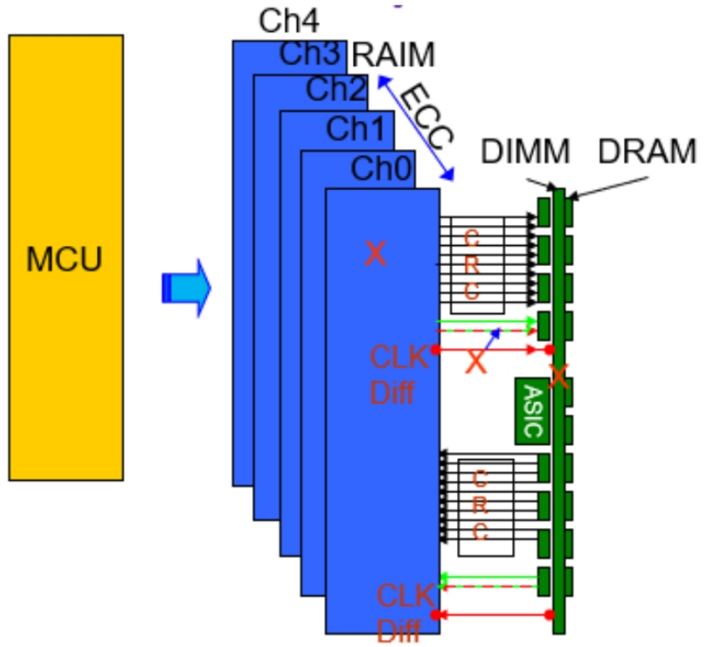
z15 Single-frame CPC Drawer Structure and Interconnect



Memory



z15 5-Channel RAIM Memory Controller Overview
RAIM = Redundant Array of Independent Memory



Layers of memory recovery

ECC

- Powerful 90B/64B Reed Solomon code

DRAM failure

- Marking technology, no half sparing is needed
- 2 DRAM can be marked
- Call for replacement on third DRAM failure

Lane Failure

- CRC with retry
- Data - lane sparing
- CLK - RAIM with lane sparing

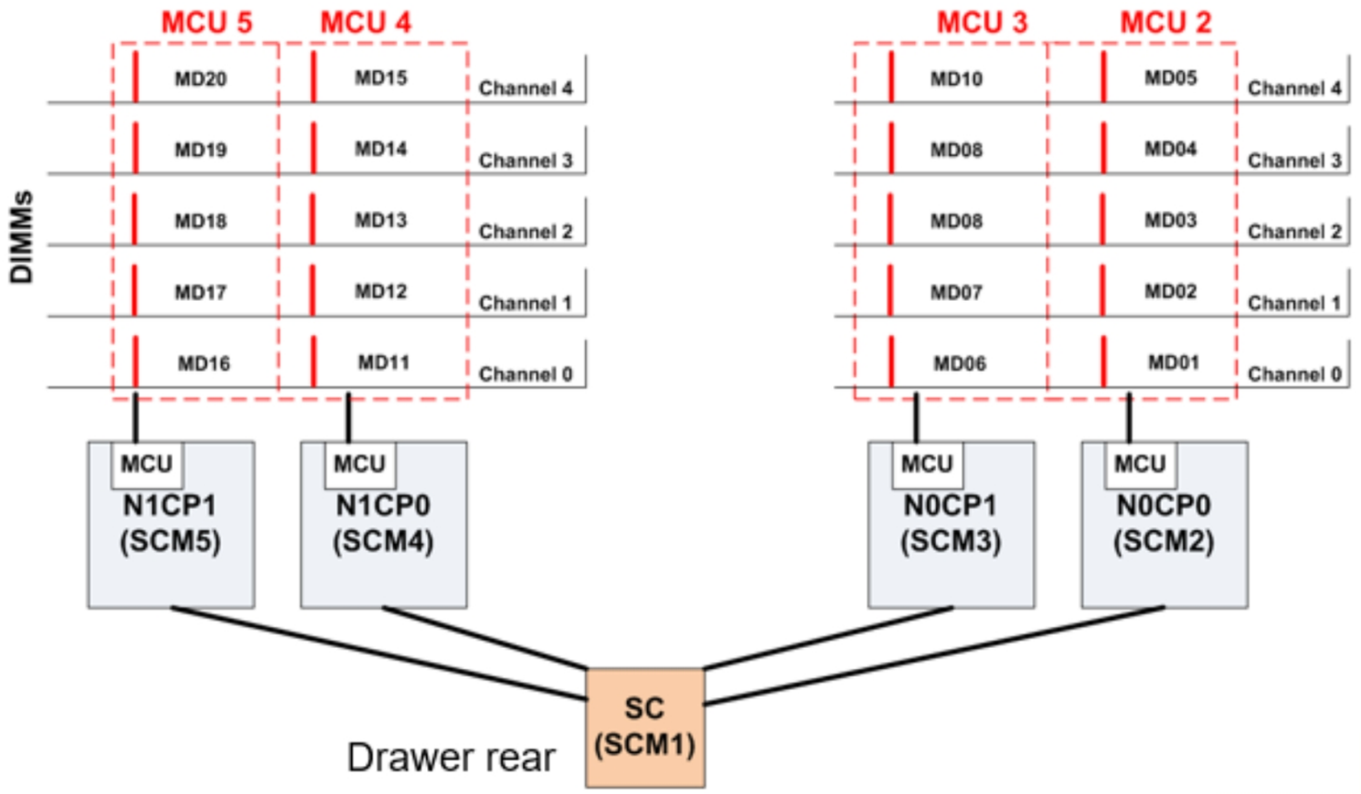
DIMM Controller ASIC failure

- RAIM Recovery

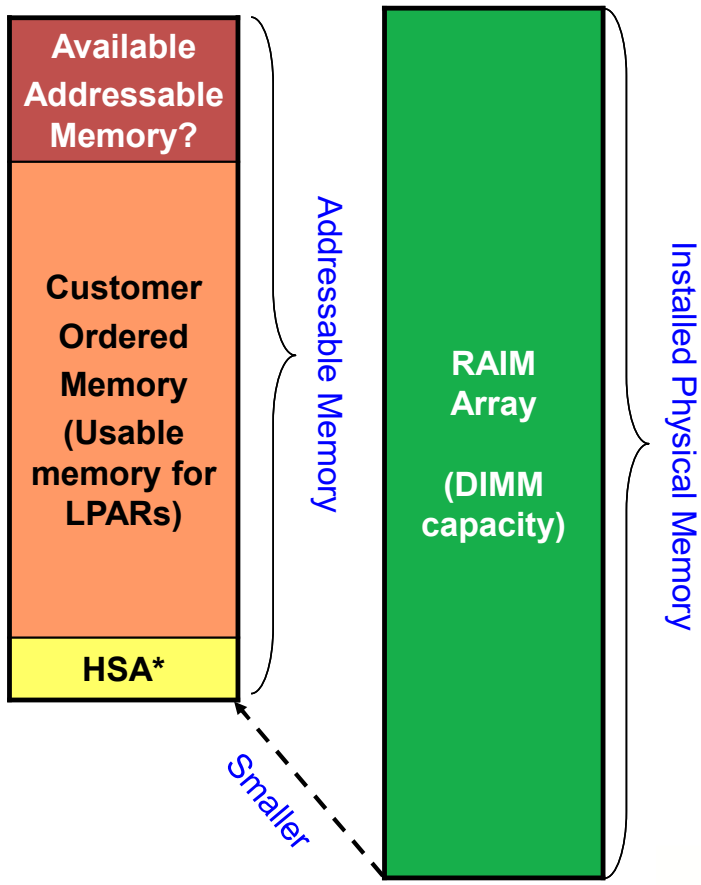
Channel Failure

- RAIM recovery

Processor Drawer Memory Topology

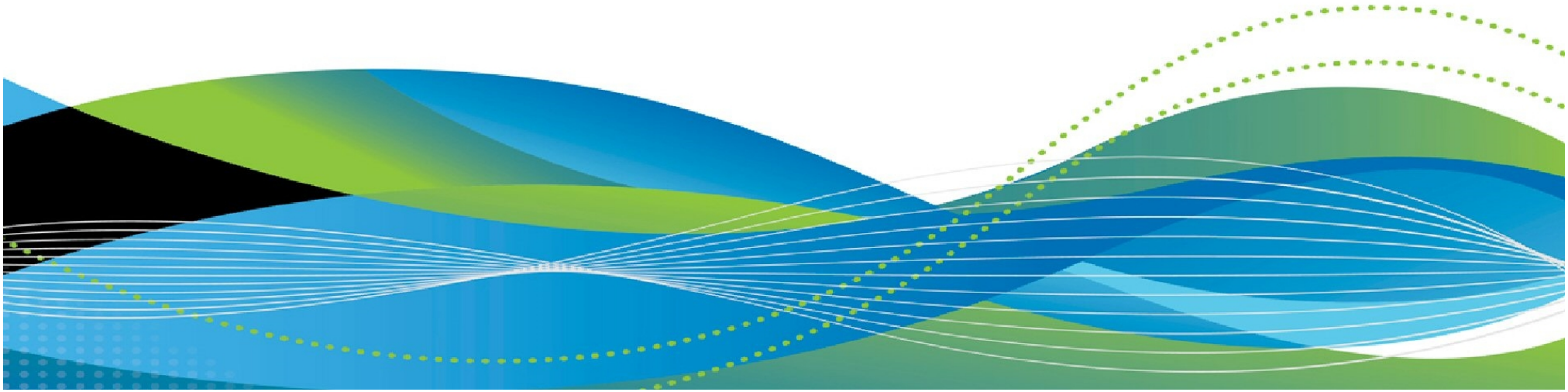


z15 Memory Usage and Allocation



*HSA size is 160 GB on z15

I / O



IBM Z I/O Subsystem

- Channel Subsystem (CSS)
  - Directs flow of information between I/O devices and main storage
  - Relieves instruction processors of communicating with I/O devices
    - System Assist Processor (SAP) interfaces with channel instead of instruction processors
    - Allows I/O operations to continue independently within instruction processors
  - Performs path management
  - Defines operating environment for correct running of all system I/O operations
  - Communicates with external devices through channels

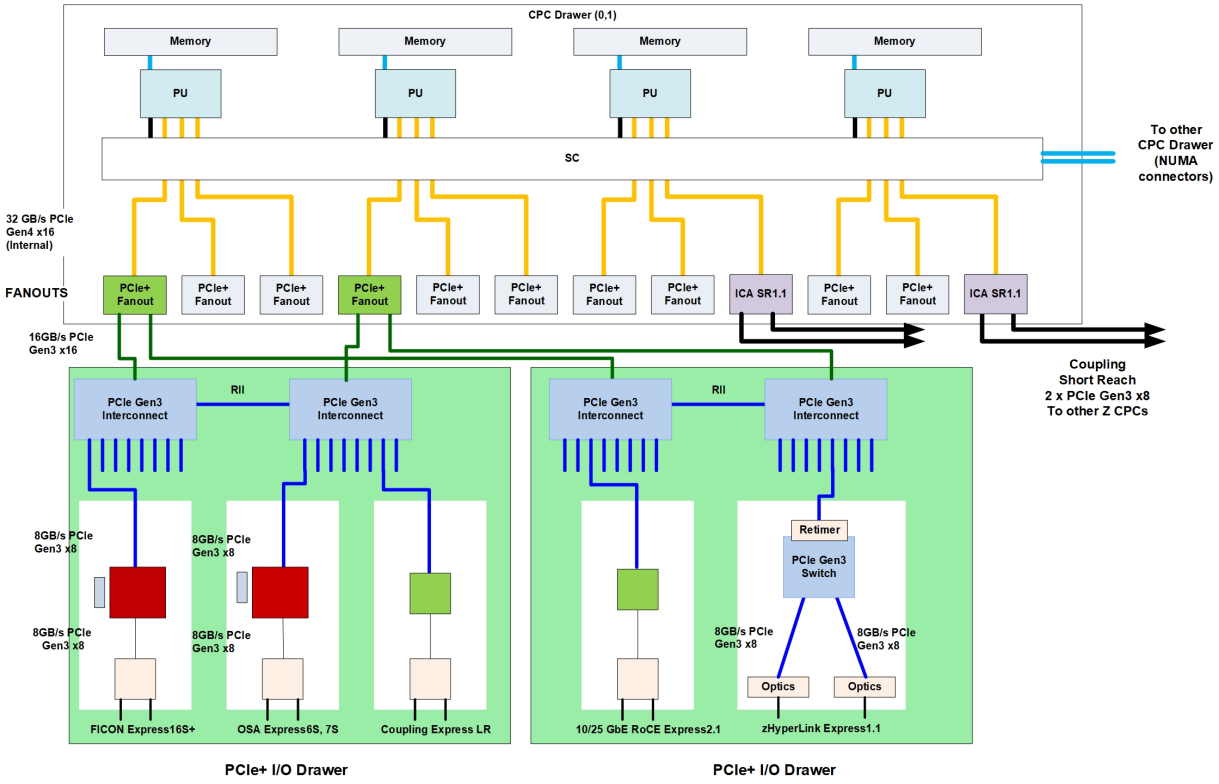
IBM Z I/O Subsystem (cont.)

- Channel or Channel Path
  - Component manages a single I/O interface between a channel subsystem and a set of control units
  - Unit transfers data concurrently with other channels and CPU
  - All devices plugged into it
    - E.g. 64 channels will transfer 64 streams of data concurrently
  - Hardware resides in PCIe I/O drawer with other I/O interconnect features
  - Hipersocket - type of channel implemented in firmware for network connectivity between logical partitions.

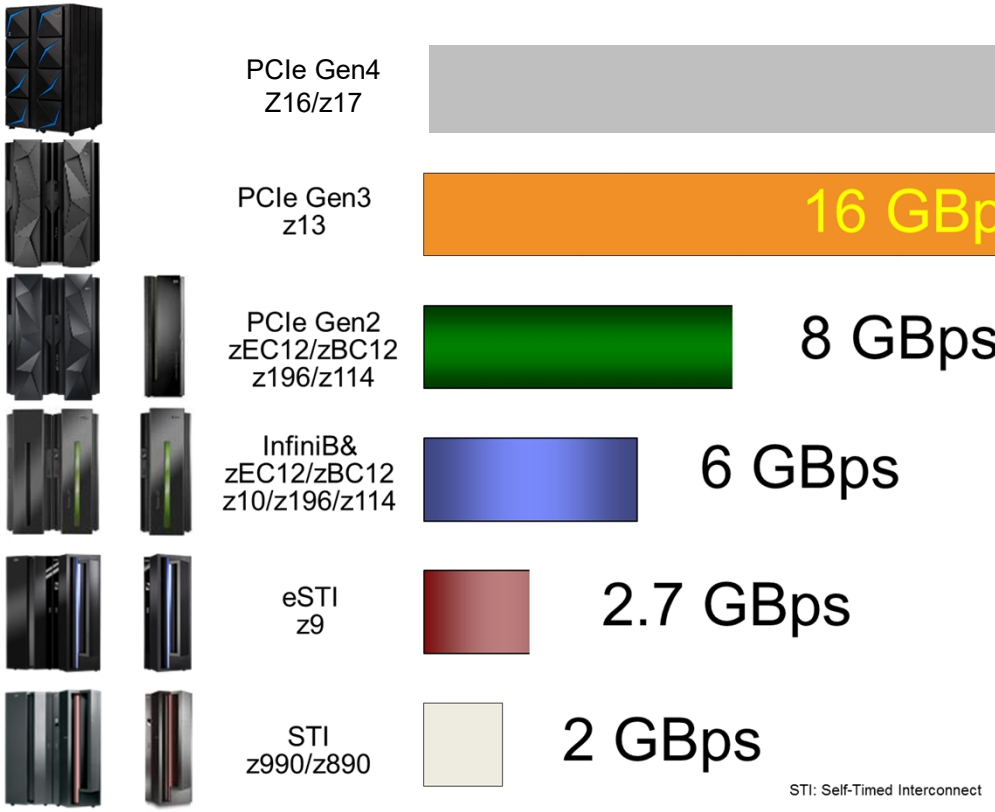
IBM Z I/O Subsystem (cont.)

- Subchannel
  - Logical representation of device to operating system
  - Contains information for sustaining a single I/O operation
  - One is assigned for each device defined in a Logical Partition
- Control Unit
  - Logical capabilities necessary to operate and control I/O device
  - Adapts characteristics of each device to standard form of control provided by Channel Subsystem
  - Housed separately or physically and logically integrated with I/O device, channel subsystem or system itself
- I/O device
  - Attached to one control unit
  - Accessible through one or more channels

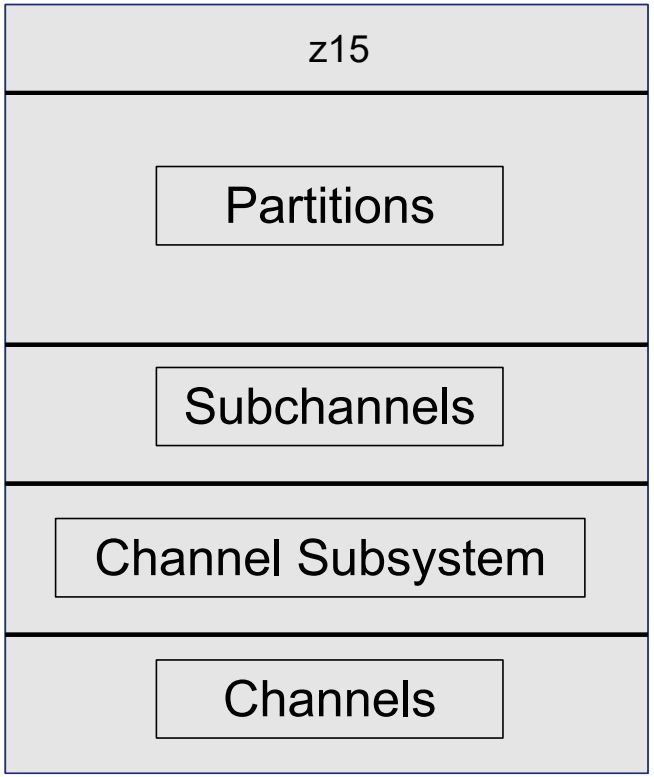
z15 I/O Infrastructure



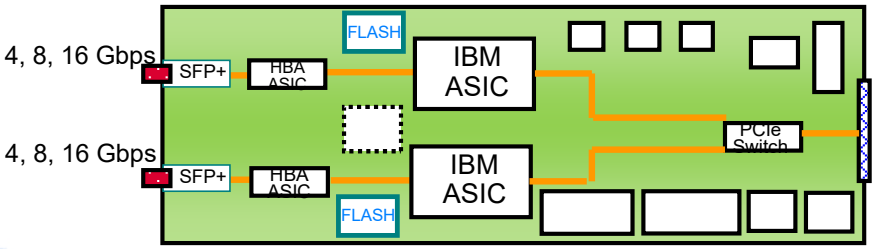
IBM Z I/O Subsystem Internal Bus Interconnect Speeds (GBps)



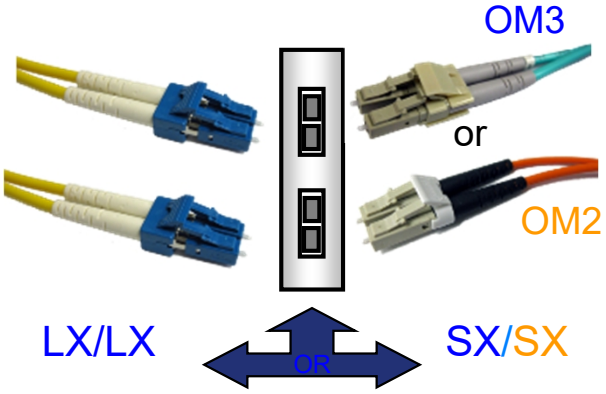
Channel Subsystem Building Blocks



FICON Express16S – SX and 10KM

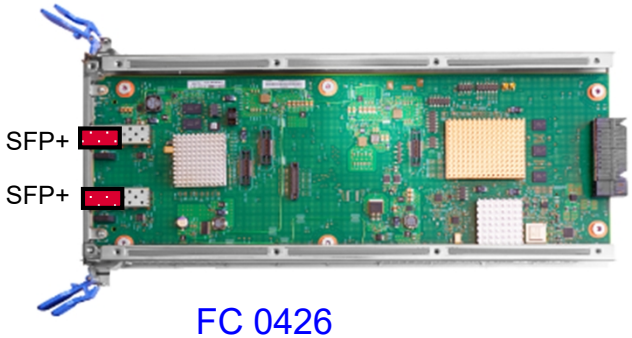


FC 0418 – 10KM LX, FC 0419 – SX



OSA-Express6S 1000BASE-T Ethernet Feature - PCIe I/O Drawer

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)

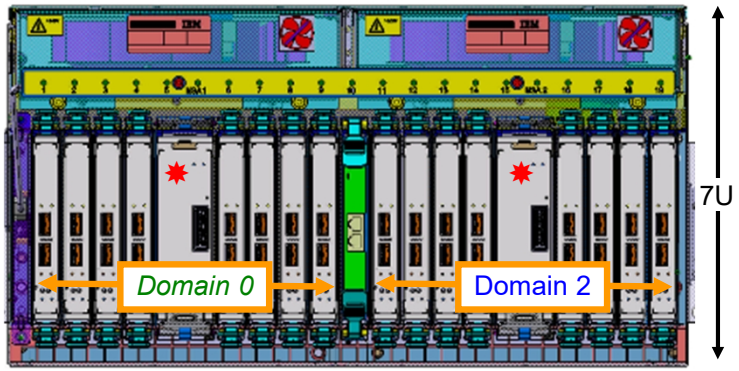


FC 0426

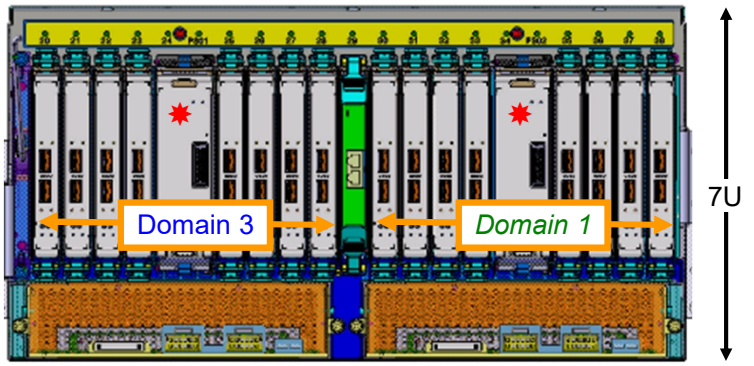
Connector = RJ-45

PCIe 32 I/O slot drawer

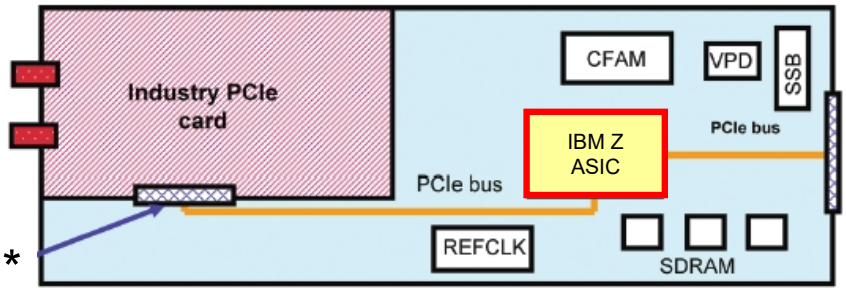
Front



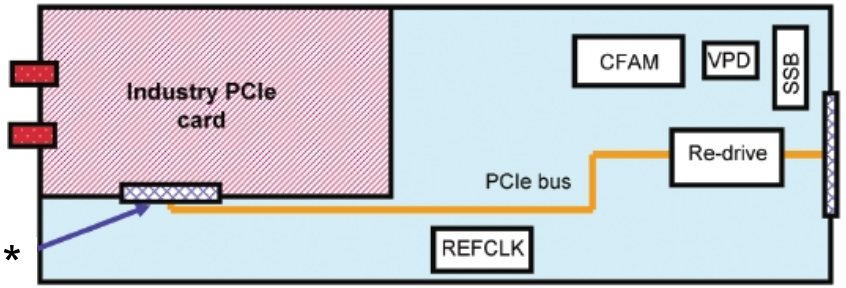
Rear



PCIe I/O Features – “Native” (AKA “Direct Attach”) PCIe Flash Express, zEDC Express and 10GbE RoCE Express



Traditional IBM Z I/O PCIe Features: FICON Express16S and 8S, OSA-Express5S and 4S, *Crypto Express4S*.



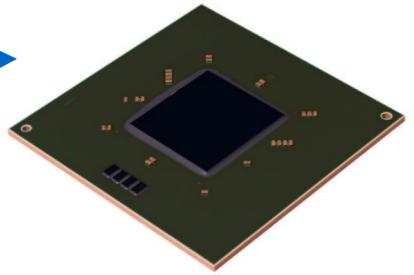
Native PCIe Feature: zEDC Express, 10GbE RoCE Express, *Flash Express, and Crypto Express5S*

Overview – HW Crypto support in IBM Z

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

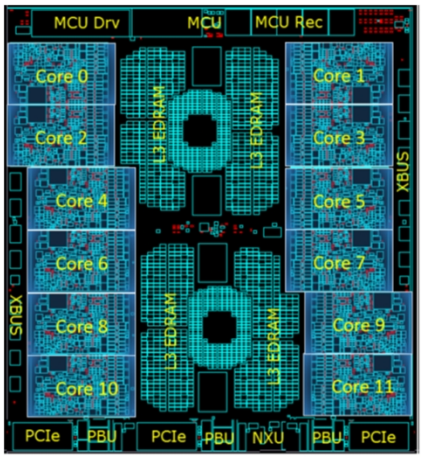
Trusted Key Entry (TKE) workstation

Smart Cards

Smart Card Readers

CPACF - CP Assist For Cryptographic Functions

- Provides a set of symmetric cryptographic and hashing functions
- Enhances encryption/decryption performance of clear-key operations
- 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 IBM Z
- Protected key support for additional security of H/W protected keys over S/W protected cryptographic keys



Crypto Express5S

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

Accelerator	
TKE	N/A
CPACF	NO
UDX	N/A
CDU	N/A

Clear Key RSA operations and SSL acceleration

CCA Coprocessor	
TKE	OPTIONAL
CPACF	REQUIRED
UDX	YES
CDU	YES(SEG3)

Secure Key crypto operations

EP11 Coprocessor	
TKE	REQUIRED
CPACF	REQUIRED
UDX	NO
CDU	NO

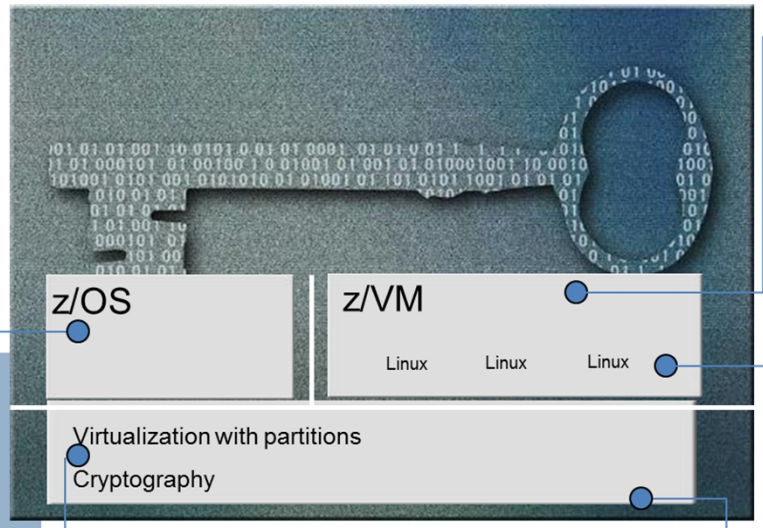
Secure Key crypto operations

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 Z Security Certifications

The IBM z13 is designed for Common Criteria Evaluation Assurance Level 5+ certification for security of logical partitions.



z/OS

- Common Criteria EAL4+
  - with CAPP & LSPP
  - z/OS V1.7 → V1.10 + RACF
  - z/OS V1.11 + RACF (OSPP)
  - z/OS V1.12, z/OS V1.13,
  - z/OS V2R1 (OSPP) Sept '14
- Common Criteria EAL5+
  - RACF V1R12 (OSPP)
  - RACF V1R13 (OSPP)
  - RACF V2R1 OSPP) in process
- z/OS 1.10 IPv6 Certification by JITC
- IdenTrust™ certification for z/OS PKI Services
- FIPS 140-2
  - System SSL z/OS V1.10 → V1.13
  - z/OS ICSF PKCS#11 Services
    - z/OS V1.11 → z/OS V1.13
- Statement of Integrity

- zEnterprise 196 & zEnterprise 114
  - Common Criteria EAL5+ with specific target of Evaluation – LPAR: Logical partitions
- System zEC12 & BC12
  - Common Criteria EAL5+ with specific target of evaluation -- LPAR: Logical partitions
- Common Criteria evaluation
- Crypto Express2 Coprocessor, Crypto Express3 & Crypto Express4s
  - FIPS 140-2 level 4 Hardware Evaluation
  - Approved by German DK (Deutsche Kreditwirtschaft)
  - Crypto Express5S is designed for FIPS140-2 Certification.
- CP Assist
  - FIPS 197 (AES)
  - FIPS 46-3 (TDES)
  - FIPS 180-3 (Secure Hash)

z/VM

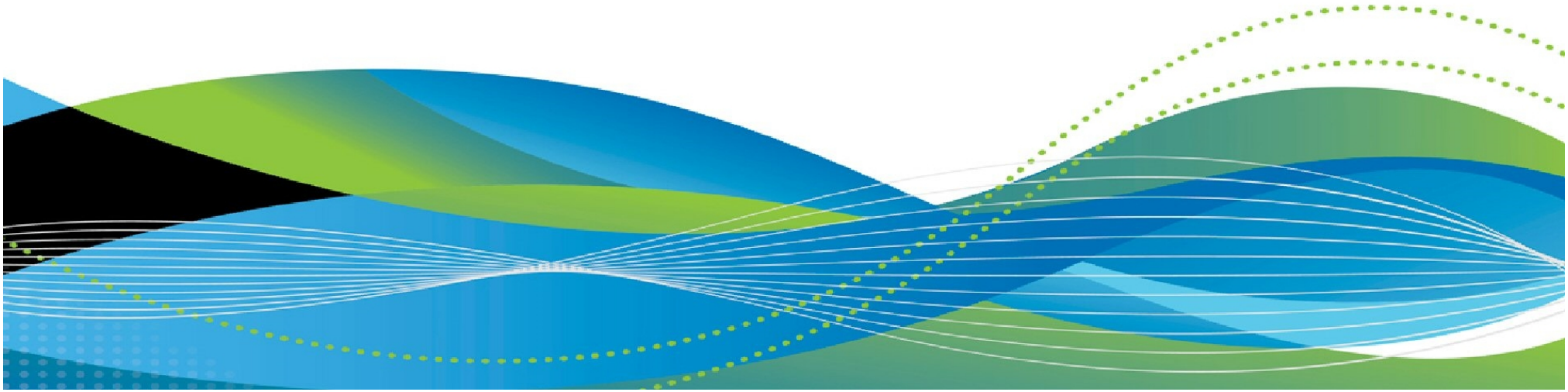
- Common Criteria
  - z/VM V6.1 is EAL 4+ for OSPP
- z/VM V6.3 System SSL FIPS 140-2 certification pending.
- Statement of Integrity

Linux on z Systems

- Common Criteria
  - SUSE SLES11 SP2 certified at EAL4+ with OSPP
  - Red Hat EL6.2 EAL4+ with CAPP & LSPP
- OpenSSL - FIPS 140-2 Level 1 Validated
- CP Assist - SHA-1 validated for FIPS 180-1 - DES & TDES validated for FIPS 46-3

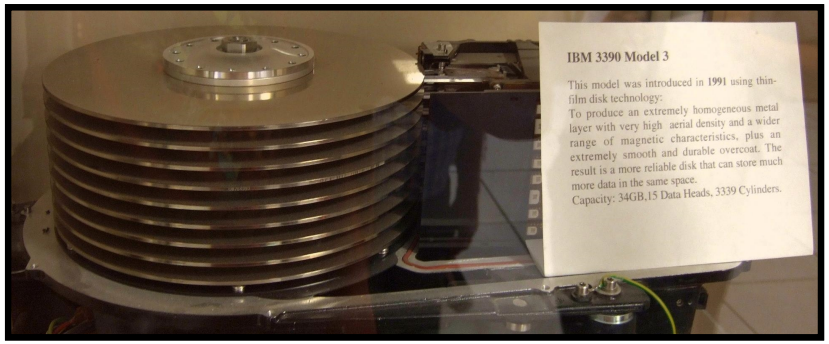
The Common Criteria program establishes an organizational & technical framework to evaluate the trustworthiness of IT Products & protection profiles

Disk Subsystems



Mainframe Disk

- Also known as CKD, ECKD, FICON disk, DASD…
  - DASD = Direct Access Storage Device (i.e. magnetic disk)
  - Format of physical devices had count field, gap, optional key field, gap and data
  - Variable size key and data portions
  - Doing I/O involved building a channel program consisting of channel command words (CCWs) that located data and would read or write the data
  - Fast search of key fields by the control unit for random access
  - Control unit would implement the channel program



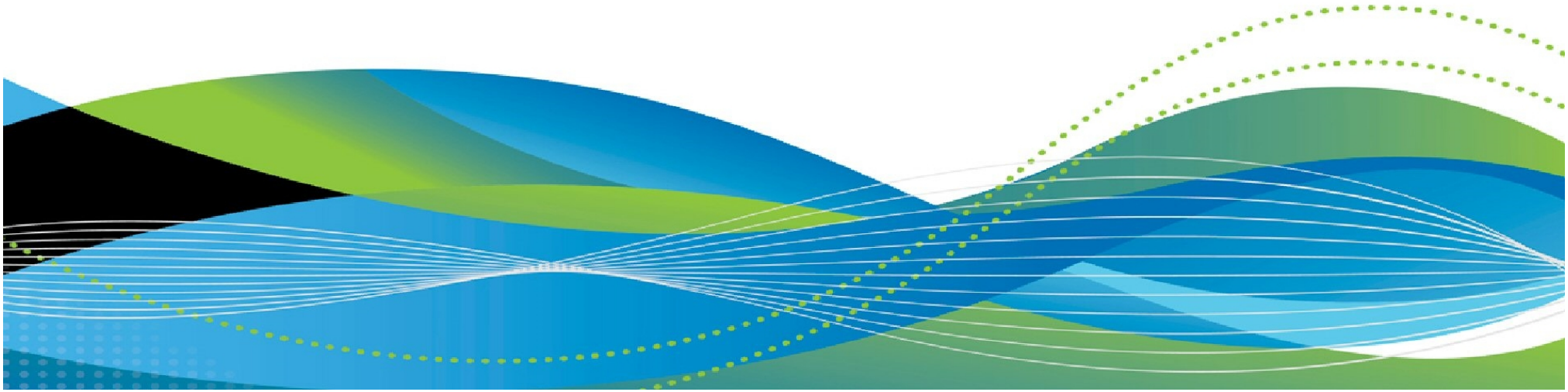
Mainframe Disk (cont.)

- Today disk subsystems are much different
  - Cabinets full of less expensive scsi SSDs or disks which operate in block mode (fixed size, addressed via index with zero origin)
  - Sophisticated processors which emulate CKD to the mainframe
  - Data spread over many disks with parity data to protect from loss and corruption
  - Dynamic rebuild to recover from disk failures
  - Hot swappable devices
  - Redundant Array of Independent Disks (RAID) architecture
- Connected by either FICON (mainframe) or FBA (open)
  - FICON
    - Operating systems still perceive individual disk devices as in the past and perform I/O operations with ECKD channel programs.
      - *Disk subsystem hides the details of physical disk I/O in the subsystem*
  - FBA
    - Same devices as an open system – Linux, Power, etc

Mainframe Disk (cont.)

- FICON
  - Architecture allows only 1 active I/O per sub channel (i.e. device)
    - Queuing occurs in operating system
    - Bottlenecks relieved by defining Hyper Parallel Access Volumes (HyperPAVs)
      - *Additional sub channels that point to the same base device*
      - *Allow more than one I/O operation to a given base device*
      - *Storage subsystems can handle more than one I/O to the same device*
  - Very well instrumented for identifying I/O performance problems and determining which component is causing the constraint (channel, control unit, device)

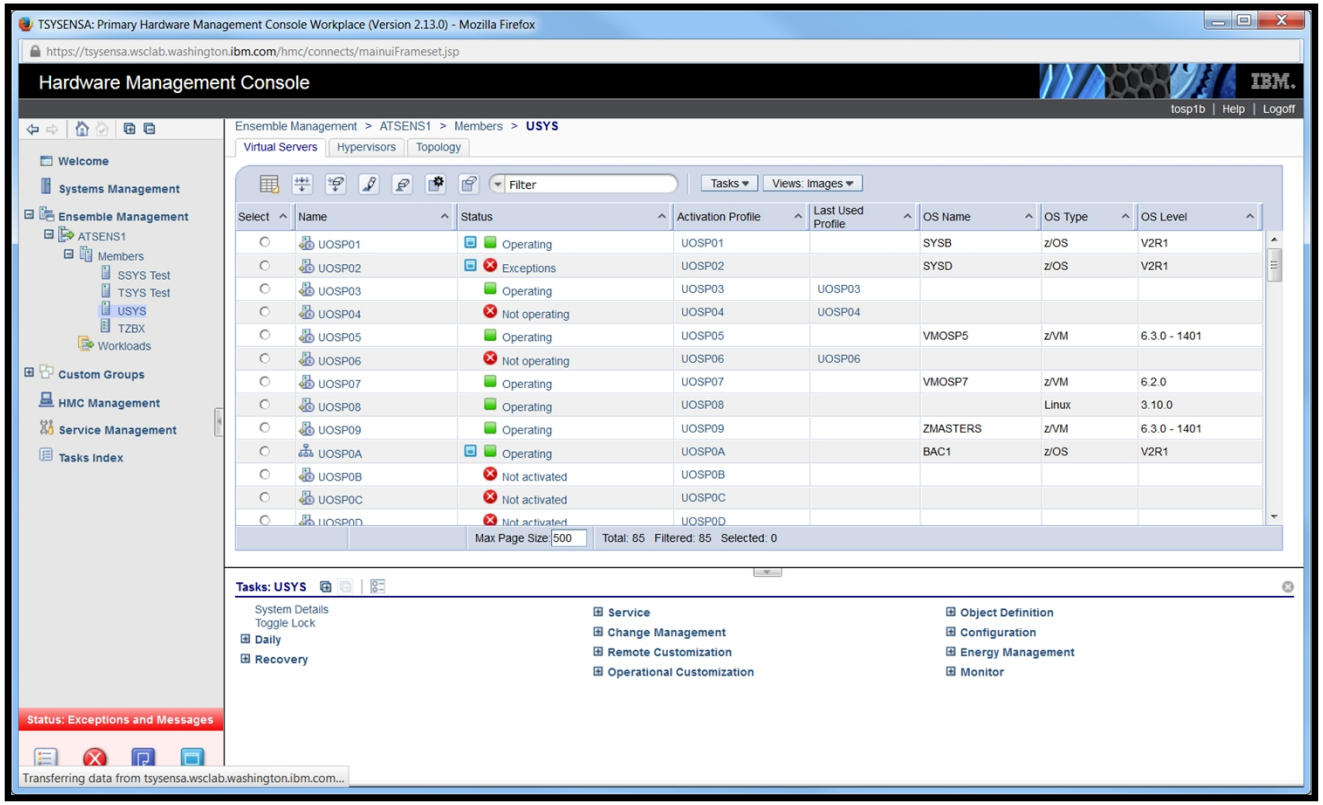
Hardware Management Console (HMC)



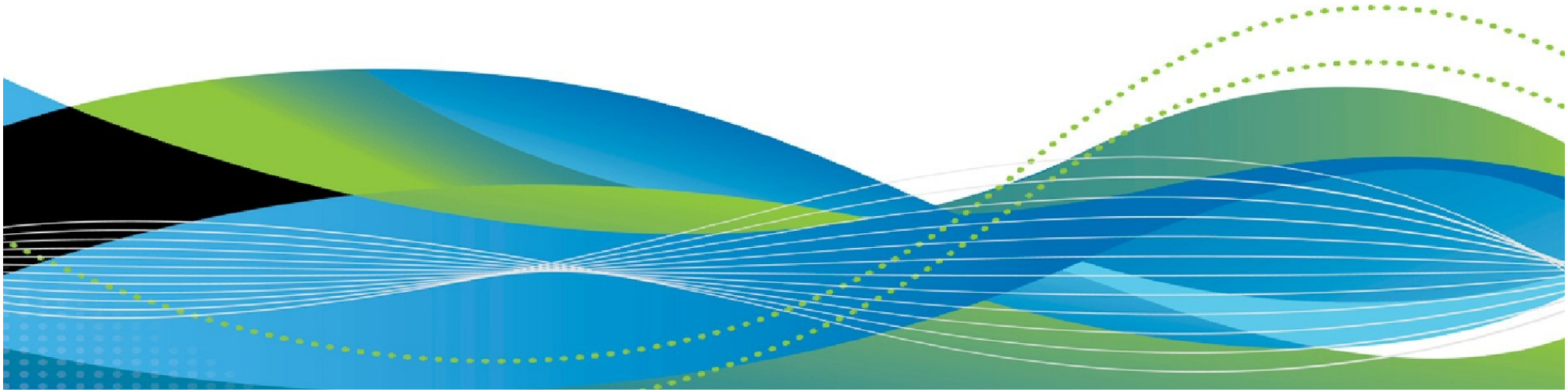
Hardware Management Console (HMC)

- Communicates with CPC through the CPC's Support Element (SE)
  - On z16 and later, HMC function runs as a task on the Hardware Management Appliance with SE function
- Tasks performed at HMC result in commands being sent to 1 or more SEs which issue commands to their CPCs
- CPCs can be grouped at the HMC so that a single command can be passed along to all of the CPCs defined to the HMC
- 1 HMC can control up to 100 SEs
- 1 SE can be controlled by up to 32 HMCs
- This is the interface a System Programmer uses to configure and manage an IBM Z

HMC Web Interface



IOCP and LPARs

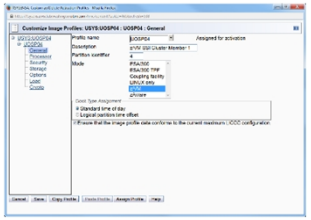


How Does Hardware Know About Configuration?

- Input/Output Configuration Program (IOCP)
  - Provides info for I/O subsystem necessary to control channel operations
    - Logical partitions
    - Channel paths and assignment to logical partitions
    - Control units attached to channel paths
    - I/O devices assigned to control units
  - Uses assembler-like macro instructions
  - Saves configuration into a hardware Input/Output Configuration Dataset (IOCDS)
- z/OS and z/VM Hardware Configuration Definition (HCD) and Hardware Configuration Management (HCM) software simplify the process
  - Especially with large configurations (i.e. 65,000 devices and lots of interconnects)

IOCP Macros

```
RESOURCE PARTITION=((CSS(0),(LINUXA,3),(LINUXB,4),(LINUXC,5),…
```



```
CHPID PATH=(CSS(0),06),SHARED,PARTITION=((LINUXA,LINUXB,… PCHID=162 …
```

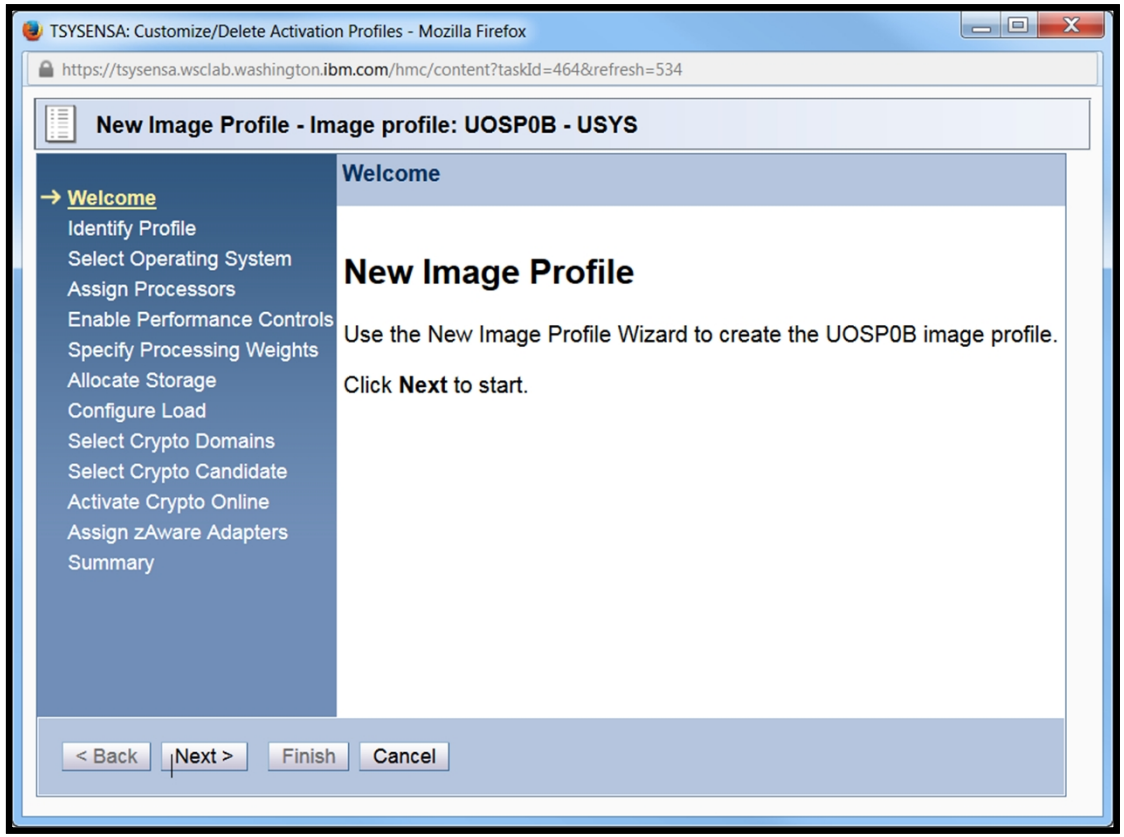
```
CNTLUNIT CUNUMBR=0060,PATH=((CSS(0),06,07,08,09)),
                UNITADD=((00,128)),CUADD=0,UNIT=3990
```

```
IODEVICE ADDRESS=(1600,128),CUNUMBR=(0060),STADET=Y,UNIT=3390
```

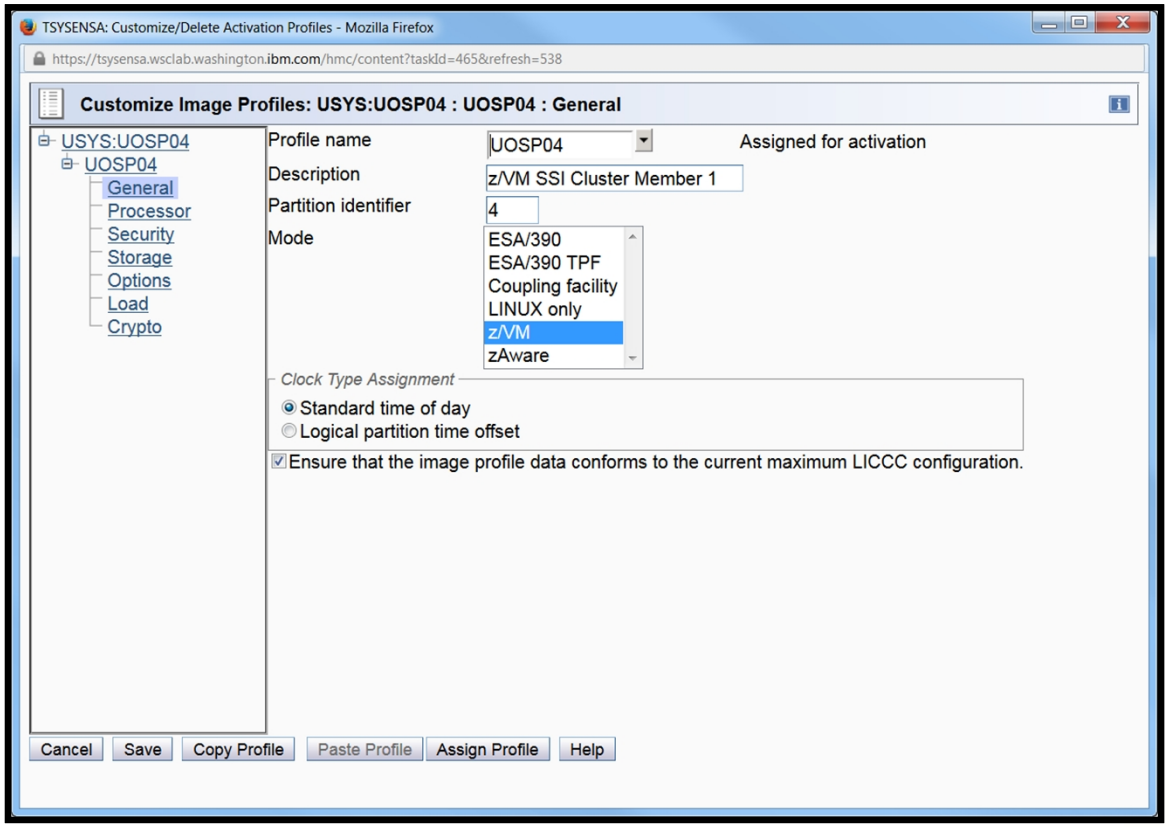
FICON Director (aka Switch)

- Provides dynamic connections between IBM Z and control units
  - Logically connect multiple control units to set of channels minimizing cabling
  - Maximize channel utilization
  - Enhances resiliency and availability
  - Makes large environments manageable and possible
- Necessary for use of FCP channels with Node Port ID Virtualization (NPIV)
  - Linux for IBM Z commonly uses this
- Enables growth
- Enables resource sharing
- IOCP identifies link addresses for connections between channels and control units

New Image Profile



Existing Image Profile



z13 LPAR Dynamic PU Reassignment

- PR/SM dynamic relocation of running processor cores to different physical core locations
- Designed to optimize physical processor location for the current LPAR's logical processor configuration
- Triggers: Partition activation/deactivation, machine upgrades/downgrades, logical processors configured on/off
- Designed to provide the most benefit for:
  - Multiple drawer machines
  - Dedicated partitions and wide partitions with HiperDispatch active