



# Processor Details: What Has Changed with z17?

Scott Chapman

Enterprise Performance Strategies, Inc.

[Scott.Chapman@EPStrategies.com](mailto:Scott.Chapman@EPStrategies.com) 😊



SHARE Webinar  
May 2026

# Contact, Copyright, and Trademarks



## Questions?

Send email to [performance.questions@EPStrategies.com](mailto:performance.questions@EPStrategies.com), or visit our website at <https://www.epstrategies.com> or <http://www.pivotor.com>.

## Copyright Notice:

© Enterprise Performance Strategies, Inc. All rights reserved. No part of this material may be reproduced, distributed, stored in a retrieval system, transmitted, displayed, published or broadcast in any form or by any means, electronic, mechanical, photocopy, recording, or otherwise, without the prior written permission of Enterprise Performance Strategies. To obtain written permission please contact Enterprise Performance Strategies, Inc. Contact information can be obtained by visiting <http://www.epstrategies.com>.

## Trademarks:

Enterprise Performance Strategies, Inc. presentation materials contain trademarks and registered trademarks of several companies.

The following are trademarks of Enterprise Performance Strategies, Inc.: **Health Check<sup>®</sup>, Reductions<sup>®</sup>, Pivotor<sup>®</sup>**

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries: IBM<sup>®</sup>, z/OS<sup>®</sup>, zSeries<sup>®</sup>, WebSphere<sup>®</sup>, CICS<sup>®</sup>, DB2<sup>®</sup>, S390<sup>®</sup>, WebSphere Application Server<sup>®</sup>, and many others.

Other trademarks and registered trademarks may exist in this presentation

# Abstract (why you're here!)



Many customers will only upgrade their mainframe machines every other generation. This can make it easier to miss changes that happened in the interim generation. In this webinar, **Scott Chapman** will discuss the differences and new capabilities for both the z16 and z17, relative to the z15. Some brief discussion of prior generations may occur as well. If you're considering a machine upgrade to the current z17 generation, then this webinar will be a good overview of some of the new features and capabilities that upgrade might afford you.

# Agenda



- Processor Design
- Logical Processors
- Measurements and Comparisons

# EPS™: We do z/OS performance...



- Pivotor - Reporting and analysis software and services
  - Not just reporting, but analysis-based reporting based on our expertise
- Education and instruction
  - We have taught our z/OS performance workshops all over the world
- Consulting
  - Performance war rooms: concentrated, highly productive group discussions and analysis
- Information
  - We present around the world and participate in online forums
    - <https://www.pivotor.com/content.html>
    - <https://www.pivotor.com/webinar.html>



# z/OS Performance workshops available



During these workshops you will be analyzing your own data!

- Essential z/OS Performance Tuning
  - March 30 – April 3, 2026 (4 days, excl Wednesday the 1<sup>st</sup>)
- WLM Performance and Re-evaluating Goals
  - July 27 – 31, 2026 (4 days, excl Wednesday the 29<sup>th</sup>)
- Parallel Sysplex and z/OS Performance Tuning
  - May 12-13, 2026
- Also... please make sure you are signed up for our free monthly z/OS educational webinars! (email [contact@epstrategies.com](mailto:contact@epstrategies.com))

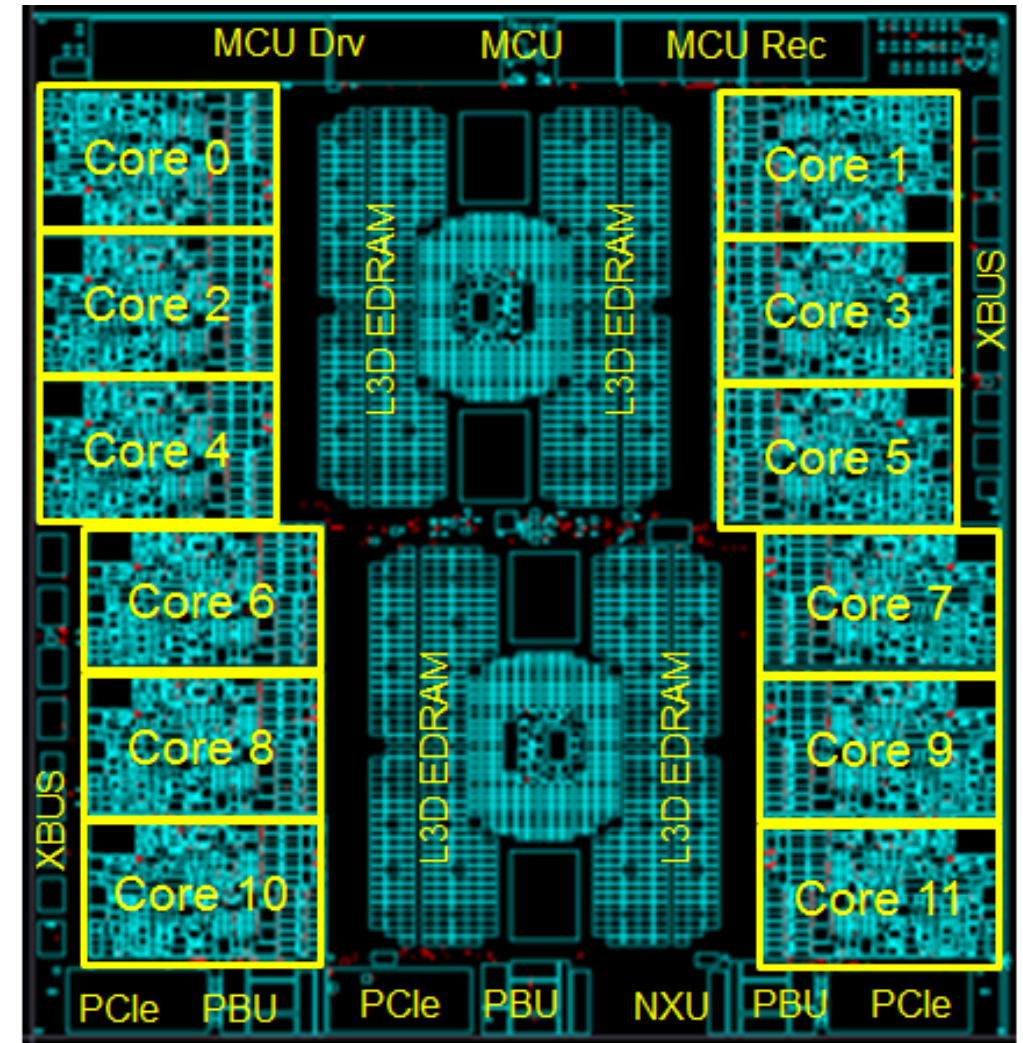


# Processor Design

# z15 PU Chip



- This is one z15 PU (Processor Unit) Chip
  - About 1" square (25.3mmx27.5mm)
  - 9.2B transistors
- 4 chips per drawer (each on SCM)
- 12 cores (9, 10, or 11 “active”) per chip
  - 41 active cores per drawer < Max190
  - 43 active cores per drawer Max190
  - Wafer yields improved by utilizing chips that have some cores disabled
- Notice amount of chip area for L3 cache
  - Note cores rotated to orient L2 near L3
  - Distance matters!
- L4 is a separate chip in the drawer

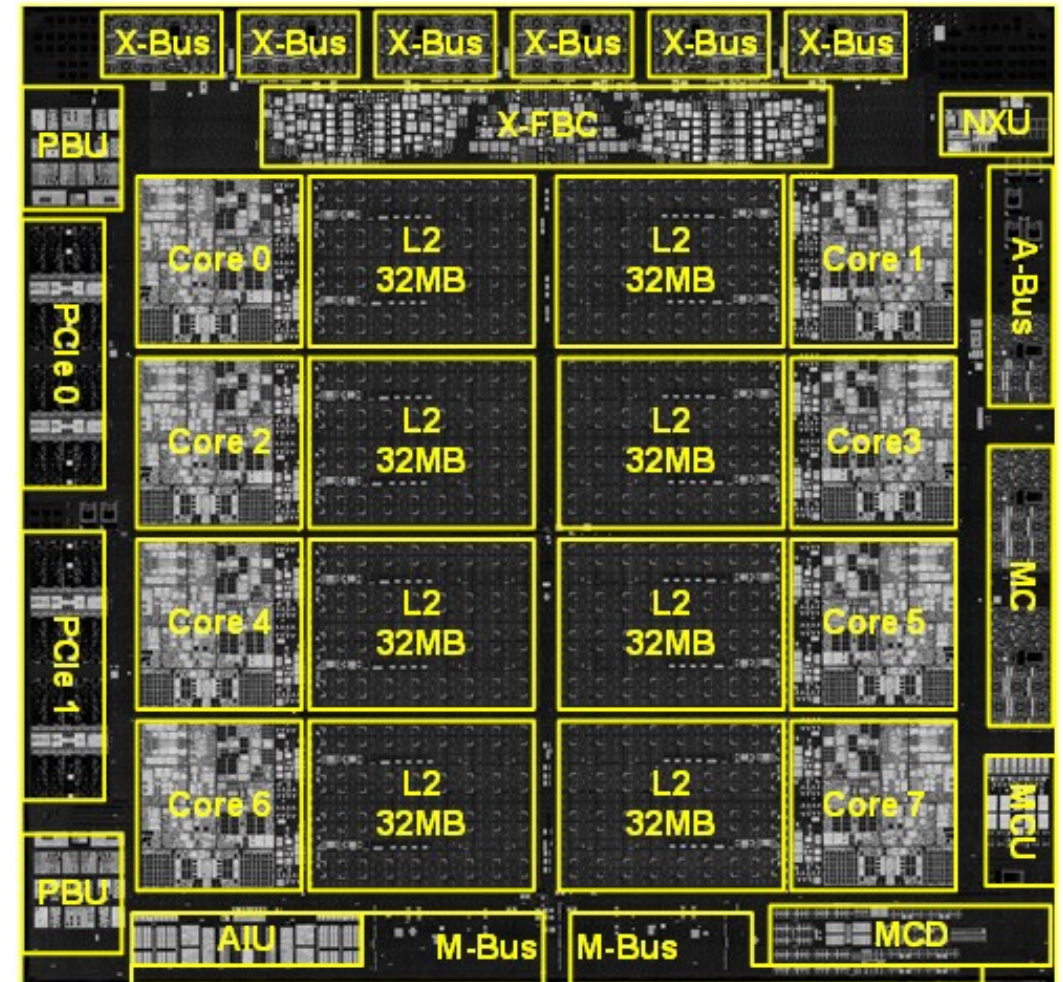


© IBM

# z16 PU Chip - Telum



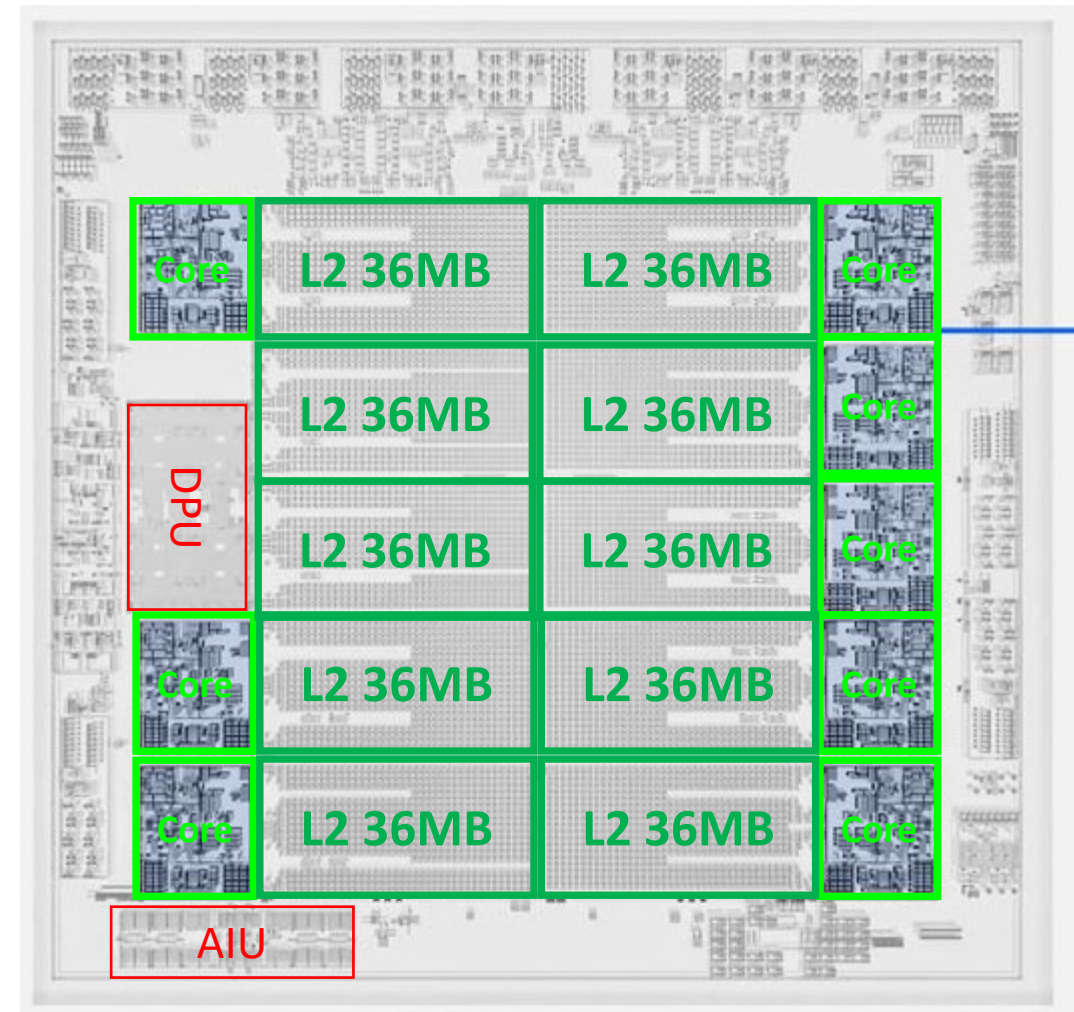
- This is one z16 PU (Processor Unit) Chip
  - A bit under 1" square (530 mm<sup>2</sup>)
  - 22.5B transistors
- 2 chips per DCM, 4 DCMs per drawer
- 8 cores per PU (not all may be active)
  - 48 active cores per drawer < Max200
  - 57 active cores per drawer Max200
  - Wafer yields improved by utilizing chips that have some cores disabled
- Note large L2 and no specific L3/L4
  - Virtual L3/L4 from sharing L2 between cores



# z17 – Telum II Chip



- This is one z17 PU (Processor Unit) Chip
  - A bit under 1" square (566 mm<sup>2</sup>)
  - 43B transistors (31B using z16 count method)
- 2 chips per DCM, 4 DCMs per drawer
- Still 8 z Cores, but 10 L2 cache areas
  - 50 active cores per drawer < Max208
  - 60 active cores per drawer Max208
  - Wafer yields improved by utilizing chips that have some cores disabled
- Similar L2/L3/L4 cache design, but more of it
- DPU (Data Processing Unit) core takes up space from two of the z cores & replaces custom ASICs on FICON cards



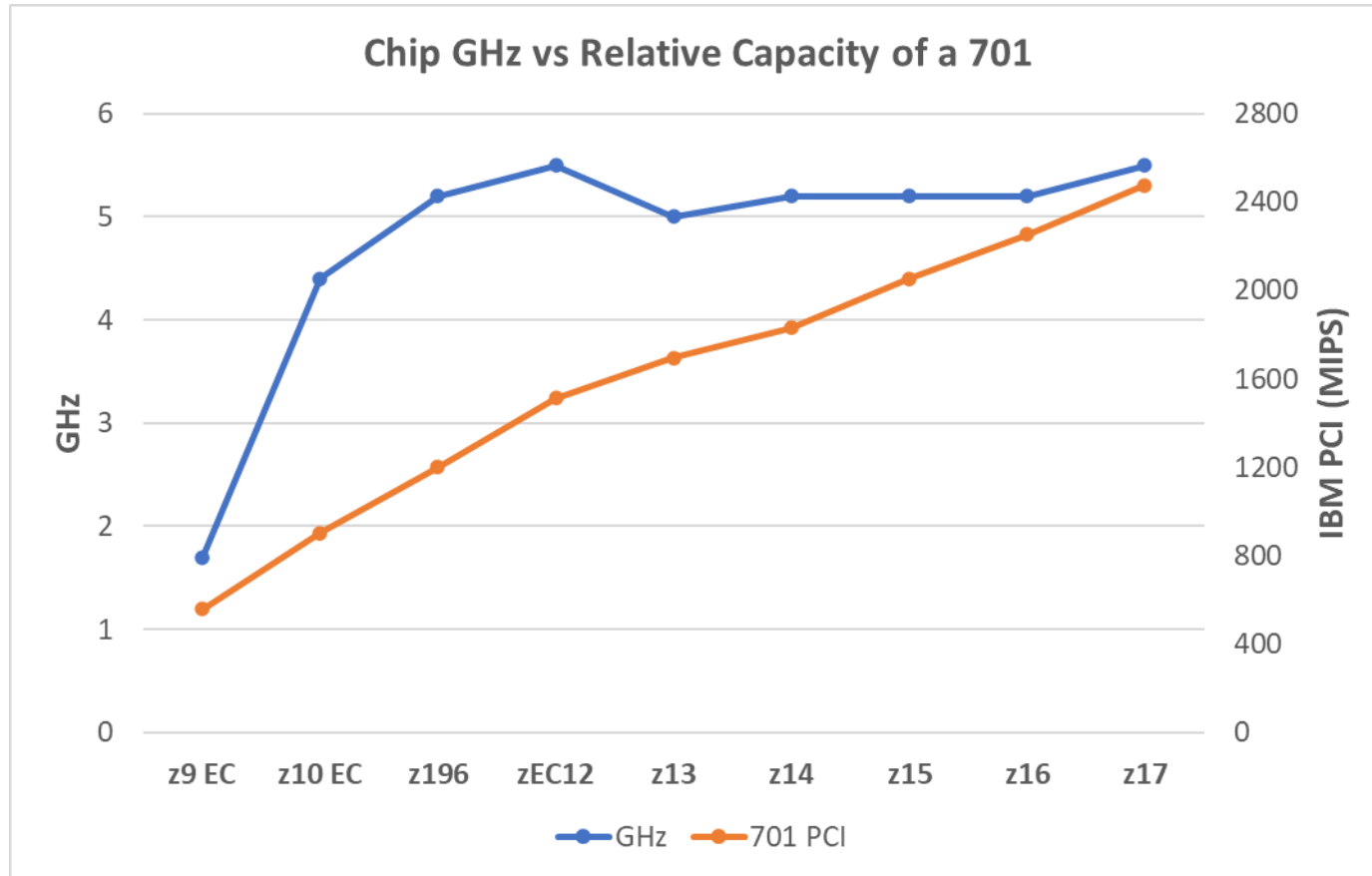
# Clock speed...



- “Clock speed” or “Cycle time” or “Clock cycle” basically refers to a quantum of time that is used to control the flow of work through the processor
  - Think of it as a metronome for the processor
  - Often referred to as a frequency, e.g. 5.5Ghz = 0.18 nanoseconds
    - Also: 54.51mm (distance light can cover in a vacuum in 180 picoseconds)
  - Represents a commit point for in-flight operations
    - Electrical signals take time to propagate around the chip so need a point in time of truth
  - Faster clock speed generally means more work done per unit of time
    - Because we have shorter quanta of time
- Note that the clock speed was mostly flat recently despite continual increases in the capacity of the processors
  - Higher clock speeds can require more power = more heat = more problems
  - Also, hard to get the necessary things done, especially considering distances

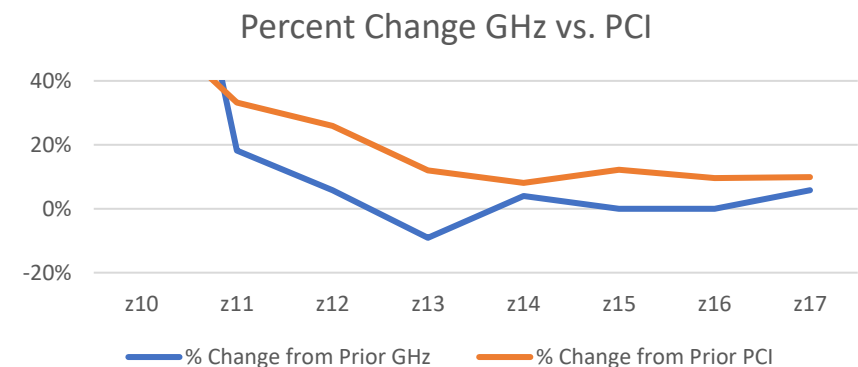
Physical distance to the data matters!

# Increased speed without clock speed



• *Architectural* improvements represent most (sometimes all) of the capacity/speed improvements over the past 10+ years

- Cache changes
- New instructions
- Micro/milli/pico-code changes



# Processor Changes Detailed



zGen	Name	Year	Mach Type	GHz	701 PCI	701 MSUs	Max per first book-drawer			Cores/chip	Processor Cache					
							Memory	CPs	PU Chips		Core-level				Chip	Book-dwr
										L1-Data	L1-Instr	L2-Data	L2-Instr	L3/chip	L4/bk-dwr	
z9	z9 EC	2005	2094	1.7	560	81	128G	8	8	2	256K	256K	n/a	n/a	n/a	40M
z10	z10 EC	2008	2097	4.4	902	115	384G	12	5	4	128K	64K	3M		n/a	48M
z11	z196	2010	2817	5.2	1202	150	704G	15	6	4	128K	64K	1.5M		24M	192M
z12	zEC12	2012	2827	5.5	1514	188	704G	20	6	6	96K	64K	1M	1M	48M	348M
z13	z13	2015	2964	5	1695	210	2464G	30	6	8	128K	96K	2M	2M	64M	960M
z14	z14	2017	3906	5.2	1832	227	8000G	33	6	10	128K	128K	4M	2M	128M	672M
z15	z15	2019	8561	5.2	2055	253	8000G	34	4	12	128K	128K	4M	4M	256M	960M
z16	z16	2022	3931	5.2	2253	278	9984G	39	4x2	8	128K	128K	up to 32M		up to 224M	up to 1.75G
z17	z17	2025	9175	5.5	2477	306	16TB	43	4x2	8	128K	128K	up to 36M		up to 324M	up to 2.88G

- Other measures go up or down, but there's always a cache size that goes up
  - Fast access to data is critical for increasing performance
  - L1 cache size limited by clock frequency
- z17 got clock speed increase and all cache levels stayed same or increased
  - At z12 and z14 clock speed bumps, cache changes were mixed

# Speed and capacity



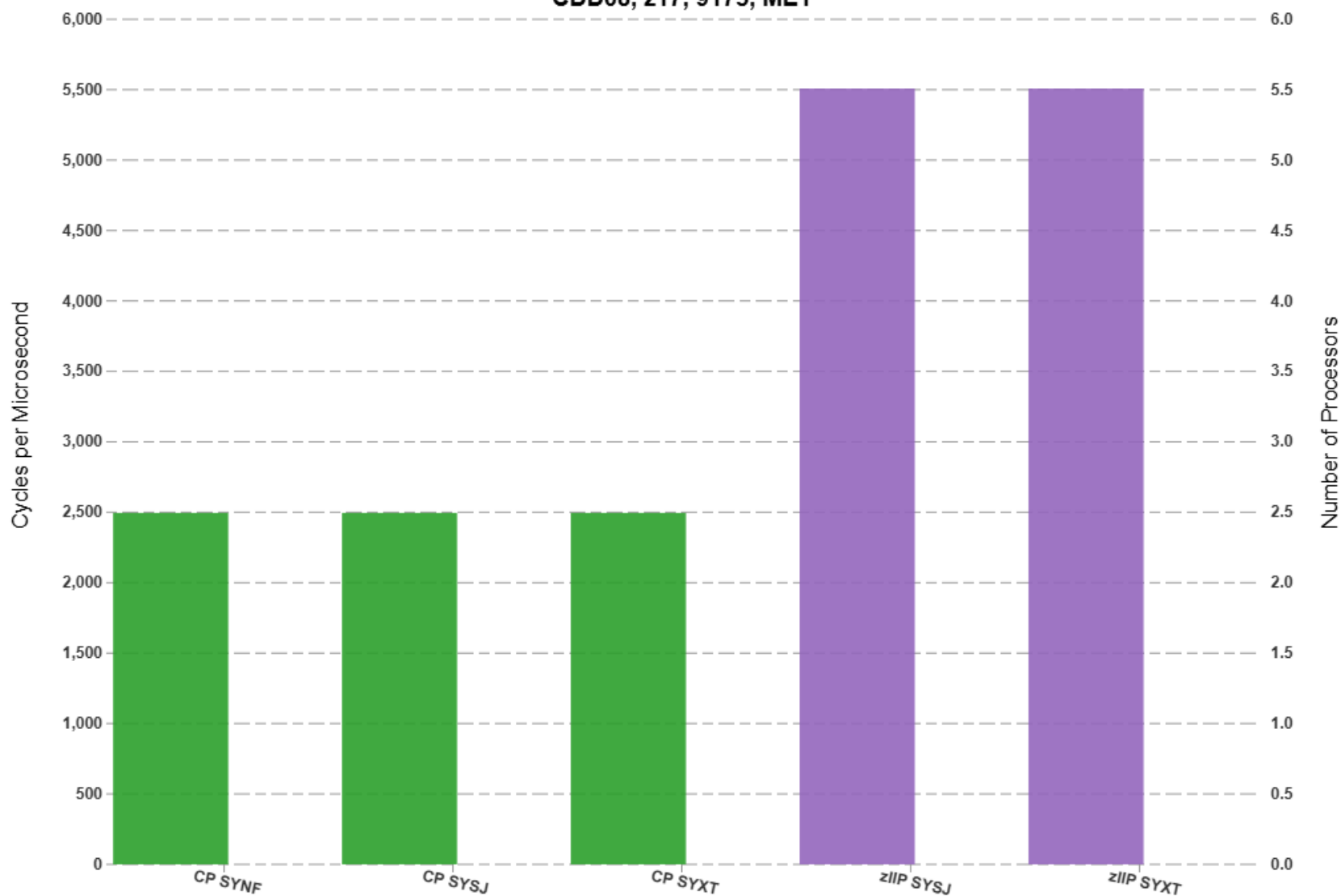
- All of the (e.g.) z17 processors run at the same clock speed
- But some GPs may be “sub-capacity” engines, which we often say are “slower” CPs
  - “slower” = “get less work done per unit of time” (make sense)
  - But the physical clock speed is not any slower
    - “Virtually”, in some measurements, it may appear to be
    - Notionally, think about the no-ops being injected into the instruction stream
- Or we say a new machine has “faster” CPs when the clock speed hasn’t changed
- I.E. we often talk about the “speed” of the CPs when we really are referring to the capacity of the individual CPs
  - I’m mostly ok with this ambiguity, but feel compelled to point it out



# Processor Speed (in Cycles per Microsecond)

SMF 113

CDD08, z17, 9175, ME1



Virtual clock speed difference from the 113s.

zIIP = 5508

CP = 2495

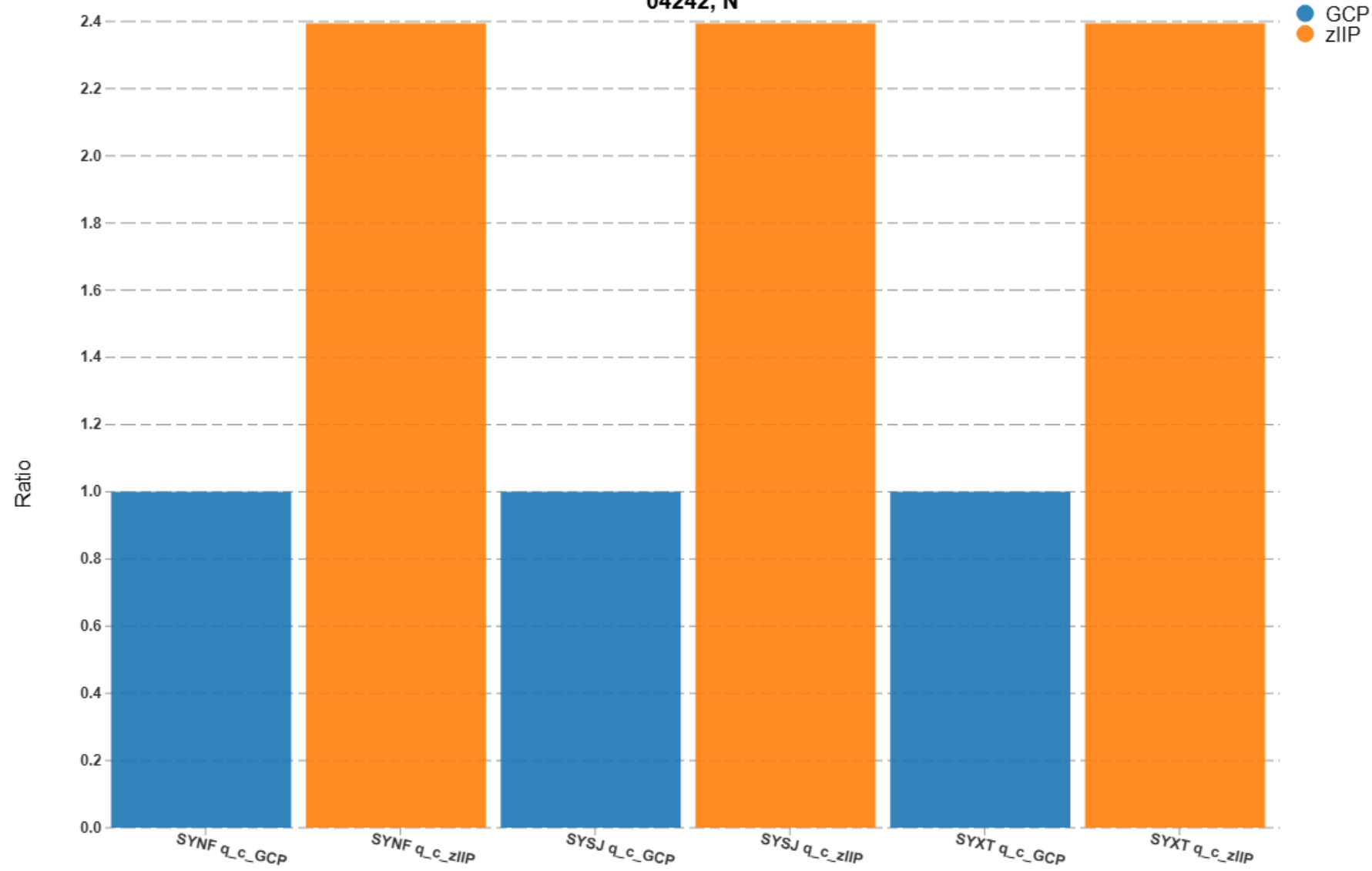
Ratio = 2.208



# zIIP to GCP Ratio

CEC, Speedboost

04242, N

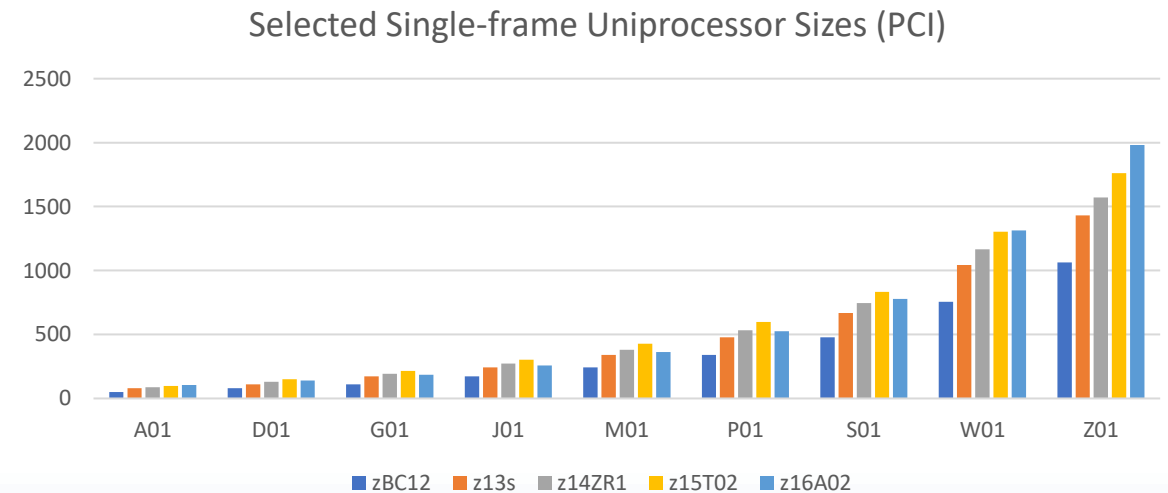
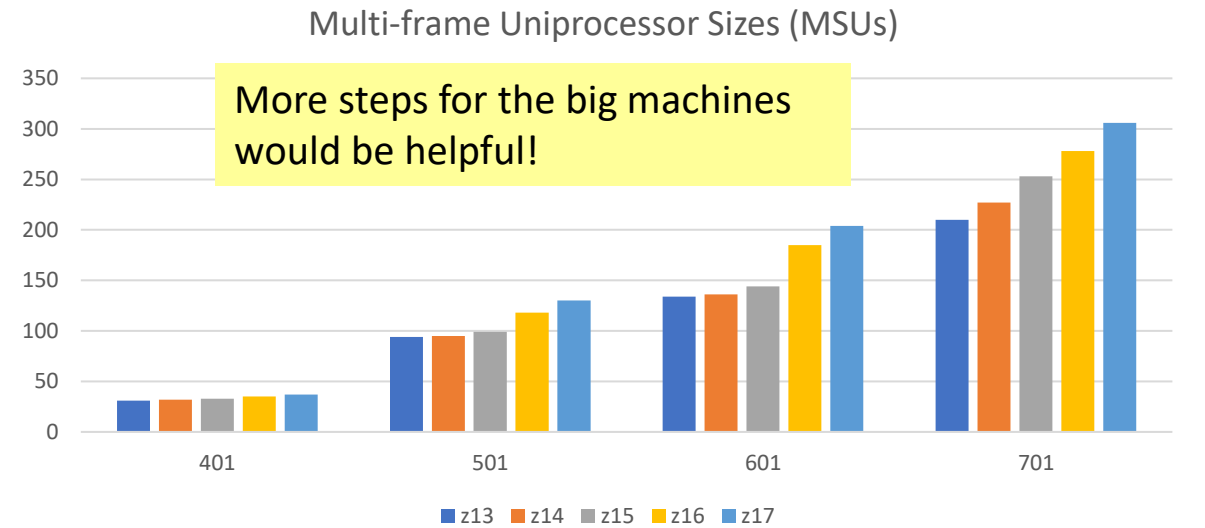


Ratio used by system in CPU calculations to normalize CPs to zIIPs.  
Ratio = 2.3945

# Sub-capacity Capacity Increases (or not)



- IBM sets the capacity of the sub-capacity models
- Sub-capacity models may not see the same per-processor capacity/performance increase that the full-speed machines see
  - z16 started adding capacity to the sub-capacity models after IBM held them mostly steady for 3 generations
  - Interesting that for some z16 A02 capacity settings, they dialed capacity *down* from the z15 T02 level for the same step
- Whether this is good or bad depends on your specific situation
  - Always use zPCR to model your proposed upgrade!



# Performance impact by workload

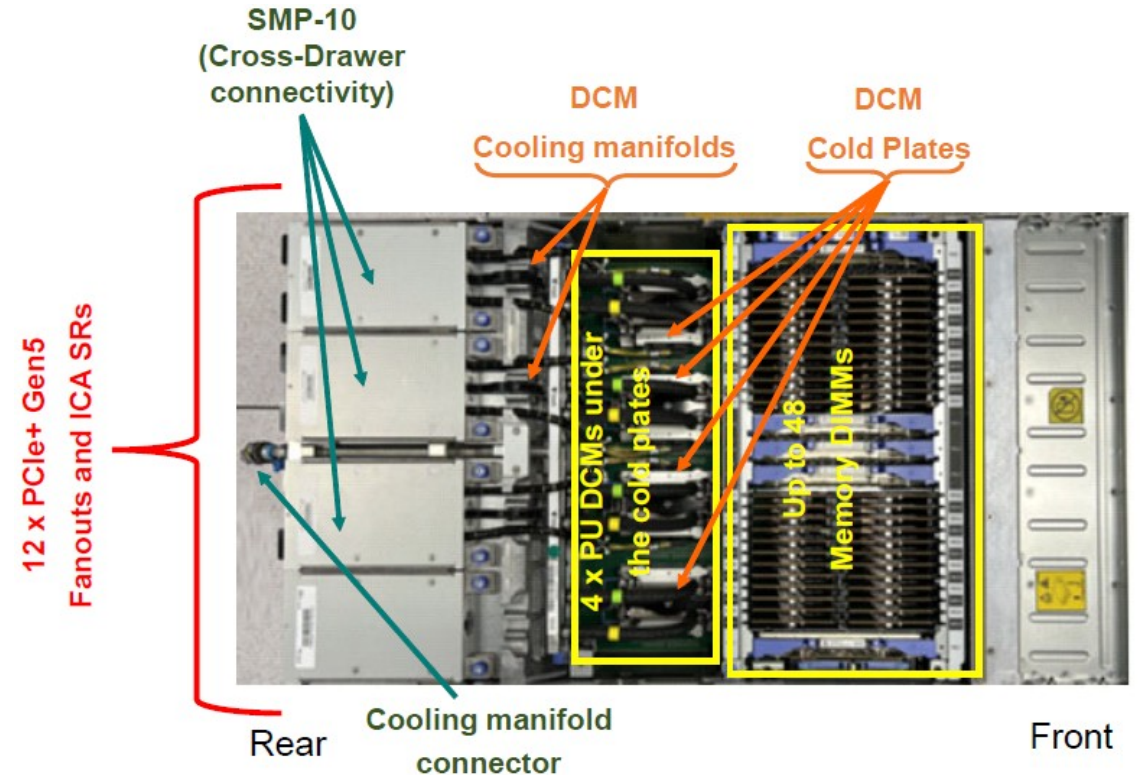


- Clock speed increase can improve all workloads
  - Note though that cache misses still take time so improvement may not be entirely uniform across all workloads
- Architectural changes will impact some workloads more than others, E.G.
  - If a workload fits all within L3 cache, increasing L3/L4 cache won't help
    - But increasing L2 probably would be helpful
  - z14 DAT improvement was a significant improvement for many systems
  - SORTL has shown limited benefits for customers I've talked to
  - Cache-unfriendly workloads may benefit more from faster memory
- Understanding your workloads can help you understand how a potential migration might affect those workloads
  - zPCR will help with this, providing better impact estimates than just using the MIPS/MSU ratings

# z17 Drawer



- PU chips are mounted on 4 dual chip modules (DCM) that are in turn mounted on the drawer “motherboard”
- 4 – 6 populated memory controller banks, each with 8 DIMMs (8 channel R-S RAIM)
  - DIMMS are 32GB-512GB each
  - All DIMMS same size per bank
  - RAIM is on the DIMMs
  - **DDR5** (or DDR4 carry-forward)
- 1 to 4 drawers per z17



© IBM

© IBM

# CPs per Drawer



Gen\Drawers	Max Active CPs/IFLs/ICFs					Max Active zIIPs					SAPs				
	1	2	3	4	4 Full	1	2	3	4	4 Full	1	2	3	4	4 Full
z14	33	69	105	141	170	22	46	70	94	112	5	10	15	20	23
z15	34	71	108	145	190	22	46	70	96	126	4	8	12	16	22
z16	39	82	125	168	200	25	54	82	110	132	5	10	15	20	24
z17	43	90	136	183	208	42	89	135	182	207	5	10	16	21	24

- Not shown:
  - 2 spares for all configs on all these generations
  - 2 IFPs on z16 and z17, 1 IFP on z14/z15
  - Optional purchased SAPs (that “nobody” purchases, **not available on z17**)
- Note change in available zIIPs based on changing policy
- From z15 on, selected via Feature MaxN where N is max active CPs
- Single drawer sets limit for max sub-capacity CPs
- **If you need a 4-drawer, consider the “4 Full”**
  - May be worth a little extra money to give PR/SM more room to avoid splitting LPARs across drawers



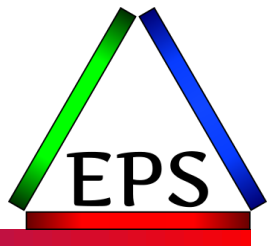
# Features

# Generation-tied Functionality



- IBM MLC price reductions for new machines
  - Get a new discount tier by moving to new machine (can help pay for the upgrade)
  - Staying on old machines forever may not be as cost-effective as upgrading!
- IBM often ties new functionality to a new machine generation
  - There may be no technical reason why some of these couldn't be available on older machines... other than that wouldn't help drive upgrades
  - Software Pricing models (e.g. WLC, TFP)
  - SRB
  - CFCC (at least partially)
- Hardware updates usually only released with new processor generation
  - E.G. FICON and CF link updates
  - Crypto Cards generally add new functionality at each release (not so much for z17)

# Notable changes by generation



Gen	Changes
z13	SMT 3-for-1 memory deal leads to more affordable, larger memory sizes CFCC Levels 20 & 21 bring larger memory support and async CF duplexing
z14	DAT changed from pico-code to multiple hardware engines (a big part of the MIPS increase) zHyperLink SMT enhancements and enabled for IOPs (SAPs) <b>Clock speed increase</b>
z15	System Recover Boost (SRB) zEDC on chip replaced zEDC Express PCIE cards (keeping data closer to the core) SORTL instruction (although of questionable value)
z16	Cache restructuring to virtual L3 and L4 and is now all faster SRAM instead of eDRAM RAIM moved to the DIMMs AI Unit SRB for Middleware Recovery
z17	DPU (Data Processing Unit) moves FICON functionality from I/O card to the processor chip DDR5 memory (increased bandwidth) Significantly enhanced AI Unit plus available AI accelerator cards (Spyre) <b>Clock speed increase</b>

# z17 AI

Different architecture  
not easily comparable  
to (eg) NVIDIA H100



- Second-generation AI Unit (AIU)

- Each AIU now capable of up to 24 TOPS (trillions of ops/sec), up 4x vs z16
- Add support for int8 to better support models using that data type
- NNPA instructions also enhanced to better support more LLMs
- Each processor core can access any of the 8 AIUs on the same drawer
  - z16 cores could only access the AIU on the same chip

- New Spyre AI Accelerator card

- 75 watts, 32 accelerator core (“similar to AIU”), 128GB of memory
- Sold in groups of 8 (total 1TB memory, 256 cores)
  - Max 6 groups per CEC
- Requires 3 LPARs to manage it
  - 1 Appliance Control Center for all CECs (2 shared IFLs, 16 GB mem, 50GB disk)
  - 2 Spyre Support Appliances per CEC (each 2 shared IFLs, 50GB memory, 50GB disk)

# z17 TFP for Workloads / Workload Classification Pricing

I'm not 100% certain of the naming here



- New z17 Workload Instrumentation Facility allows z/OS to tag executables to differentiate what they're used for
  - Will allow differentiated HW pricing based on executable's purpose
  - Pay for some always-on HW capacity on a "cloud-like" on-demand basis
- Initially, two workload classes will get differentiated pricing
  - AI Inference
  - z/OS Container Extensions virtualizing Linux on z/OS
- Additional classes being considered and could affect both HW and SW
- TFP for Workloads requires z17 and TFP for Hardware

# z17 DPU



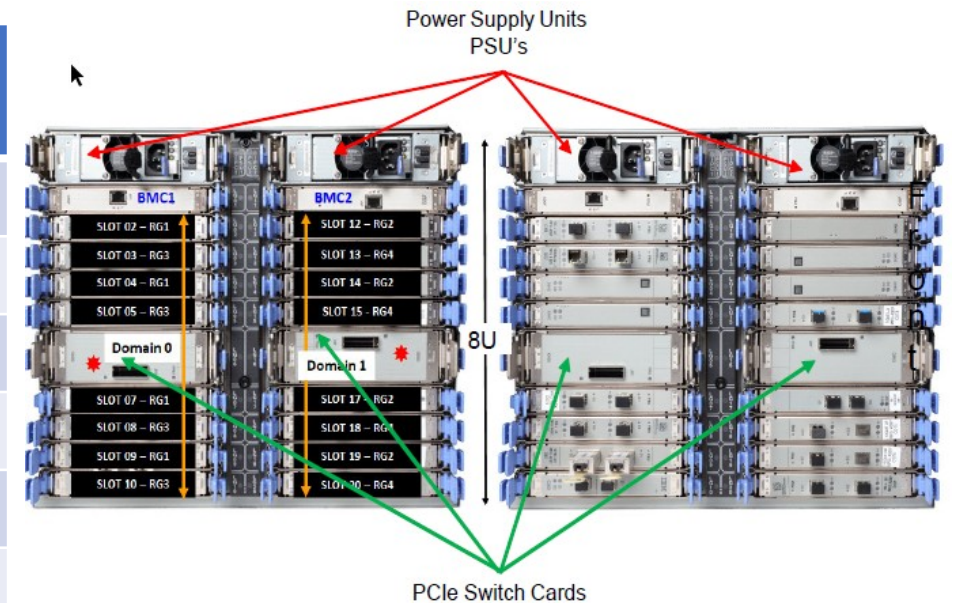
- Potentially most significant architectural change in z17
  - Yes, ignoring AI because not everybody will use that
- DPU has 32 cores organized into 4 clusters plus its own private L1 cache
- DPU avoids need for custom ASICs on FICON I/O cards to handle FICON
  - Each DPU consolidates up to 16 ASICs and supports up to 64 PCHIDs
- Expected to enable more throughput for small I/Os
  - Up to 400,000/sec/port (far above what we generally see sustained)

# Internal I/O Connectivity



- Channel Adapter cards are installed in cages or drawers inside the CEC
  - Connected to fanouts on the processor book/drawer
    - “Recently” either 6GB/s Infiniband, 8GB/s PCIe Gen2, 16GB/s PCIe Gen 3, 32 GB/s PCIe Gen 4
    - z15 & z16 processors have PCIe gen 4 that is split into PCIe Gen 3 fanouts
    - z17 processor has PCIe gen 5 that is split into PCIe Gen 4 fanouts
  - All have redundant connectivity to the books/drawers
  - Some older ones can carry forward to the next machine

Machine Generation	I/O Cage (Infiniband)	I/O Drawer (Infiniband)	PCIe I/O Drawer
z12	Yes	Yes	Gen2
z13	n/a	Yes	Gen3
z14	n/a	n/a	Gen3
Z15	n/a	n/a	Gen3
z16	n/a	n/a	Gen3
z17	n/a	n/a	Gen4



z17 I/O Drawer  
© IBM

# I/O Cards



- Prior to z17, most modern I/O Cards have 2 connections / card
  - OSA 10GbE & 25GbE are single port cards
  - Old 1x IFB CF Links were 4-port cards
- z17 doubles ports on some cards
  - FICON now 4 ports at 32 Gbps (same as z16, auto-negotiate to 8 or 16 Gbps)
  - New z17 Network Express 2-port card replaces OSA, RoCE, and LR Coupling
    - Capable of both 10GbE and 25GbE
- CF Links still 2 port cards
  - ICA SR2.0 still 2 port cards, still 8GBps signaling rate
    - Does now use PCIe Gen4 instead of PCIe Gen3
  - Coupling Express3 Long Reach also still 2 port cards, 10 or 25Gbps
  - Support for N-2 (z15) connectivity
    - 10Gbps only for LR connectivity to z15/z16
    - Also, 10 to 10 only and 25 to 25 only

# z17 Planning notes



- If you think you might want Spyre cards in the future, plan for that
  - If not planned for, adding Spyre cards might not be entirely non-disruptive
- Significantly more memory available
  - Now 16TB/drawer (up from 10TB in z15/z16)
- No field upgrade from 3 to 4 drawers
  - Same as z15 and z16
- No optional SAPs
- *Not* z17-specific, but:
  - Fibre Channel Endpoint Security required for zNext and DASD/Tape subsystems GA'ed after Dec 31 2024 (Currently none have)



# What should you buy?

# Fewer/Faster vs. More/Slower



- Although dependent on the LPAR configuration and software particulars, I'm often (almost always!) a fan of more/slower vs. fewer faster CPUs
  - E.G. a 410 vs. a 503 or 620 vs. 710
- More/slower can get you more:
  - L1/L2 cache
  - More TLB
  - More vertical high CPUs
- All of the above can result in a more efficient overall system when you have more than 1 significant LPAR on the machine
- Multiple LPARs sharing a few fast CPUs, each end up getting a small time slice, resulting in them processing much like slower CPUs, albeit with less total cache

# Logical and Physical CPUs



- Processor = CP = CPU = GCP or zIIP or any other processor type
  - All the same bit of silicon: a core on a physical chip
- You pay for a certain number of physical processors (CPs)
  - **A processor can only be processing one stream of instructions at a time**
    - Absent SMT, which doesn't apply to GCPs and which we're not going to discuss here
- You define LPARs, each with a certain number of logical, shared CPs
  - For each LPAR Logical CPs  $\leq$  physical CPs, although can have reserved CPs
  - Most machines have multiple LPARs
- z/OS dispatches work to its (logical) CPs
- PR/SM dispatches logical CPs to physical CPs
  - A logical CP can't do any work when it's not dispatched to a physical CP
  - If you only have 1 physical CP, only 1 LPAR is doing anything at any given instant

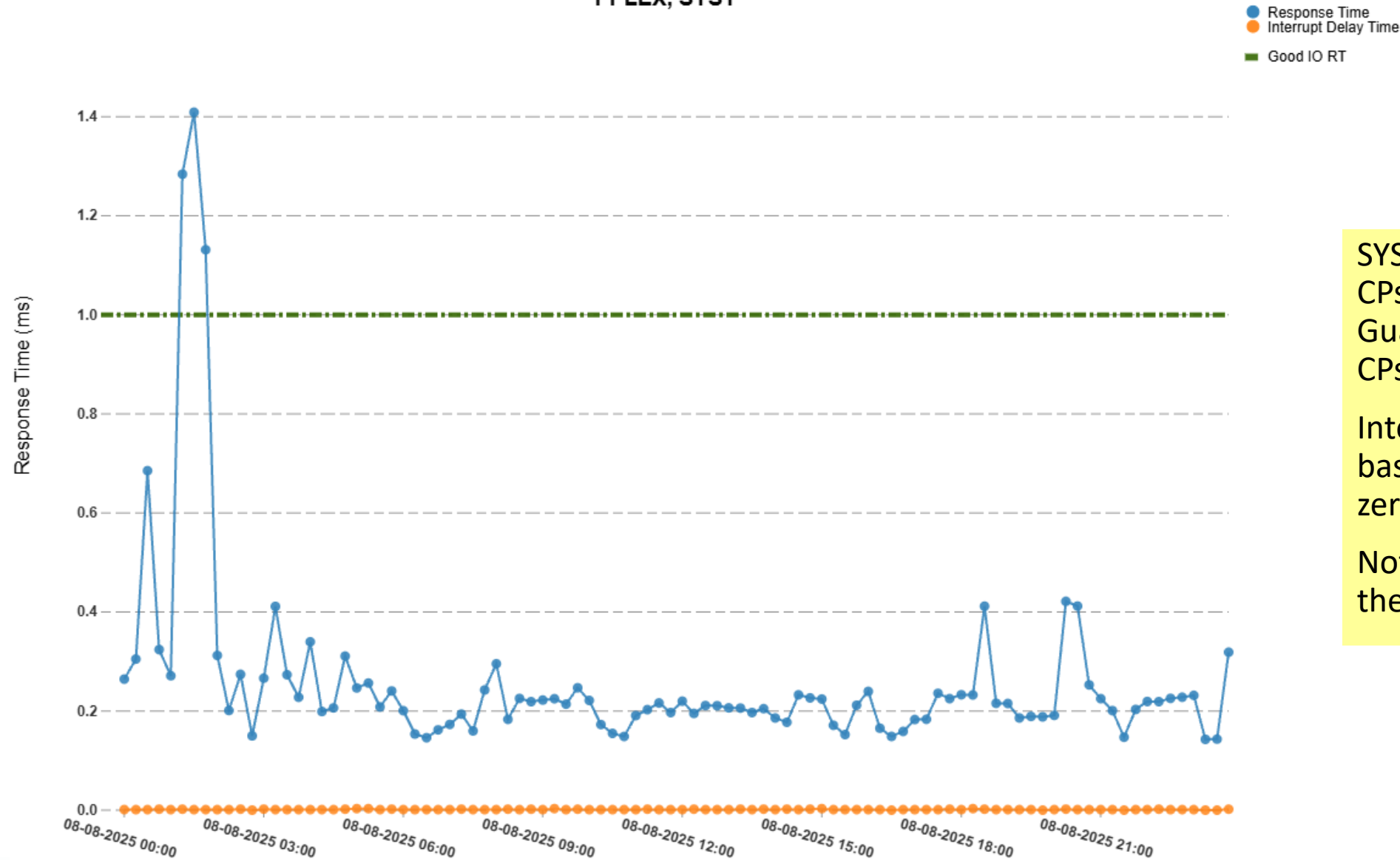
# CP Availability impact to I/O times



- DASD Response Time = IOSQ + PEND + DISCONNECT + CONNECT
- Not (usually) included in the response time value: **Interrupt delay time**
  - This is delay that it takes for the interrupt to be processed by z/OS
  - On larger LPARs generally rounds to zero
  - On small LPARs with low weights and only 1 or 2 CPs, can be quite significant
    - Remember, if a PCP isn't dispatched to the LPAR, the LPAR can't process the interrupt
    - Seen small systems with fast DASD where interrupt delay is  $\geq$  RT
    - Could be a risk point for migrating to shared outsourcer system
- When available, High Pool CPs will be the first ones enabled for I/O interrupts
  - Avoids most of the potential delay due to other LPARs using the CPs

# Average Response Time versus Interrupt Delay Time

PPLEX, SYS1

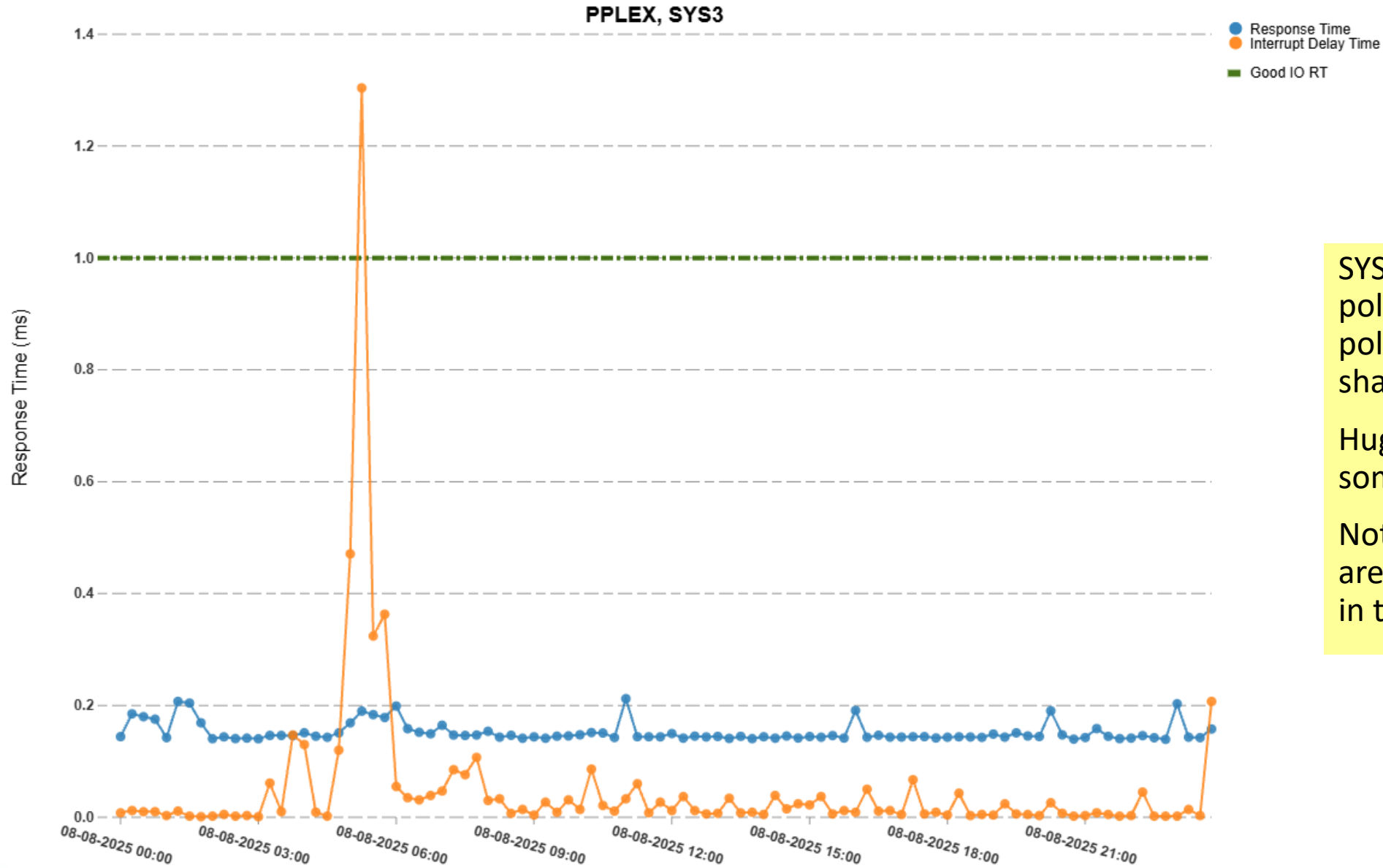


SYS1 has 2 high polarity CPs and 2 mediums. Guaranteed share is ~3.2 CPs capacity.

Interrupt delay is basically rounding to zero.

Note this comes from the SMF 74 data.

# Average Response Time versus Interrupt Delay Time

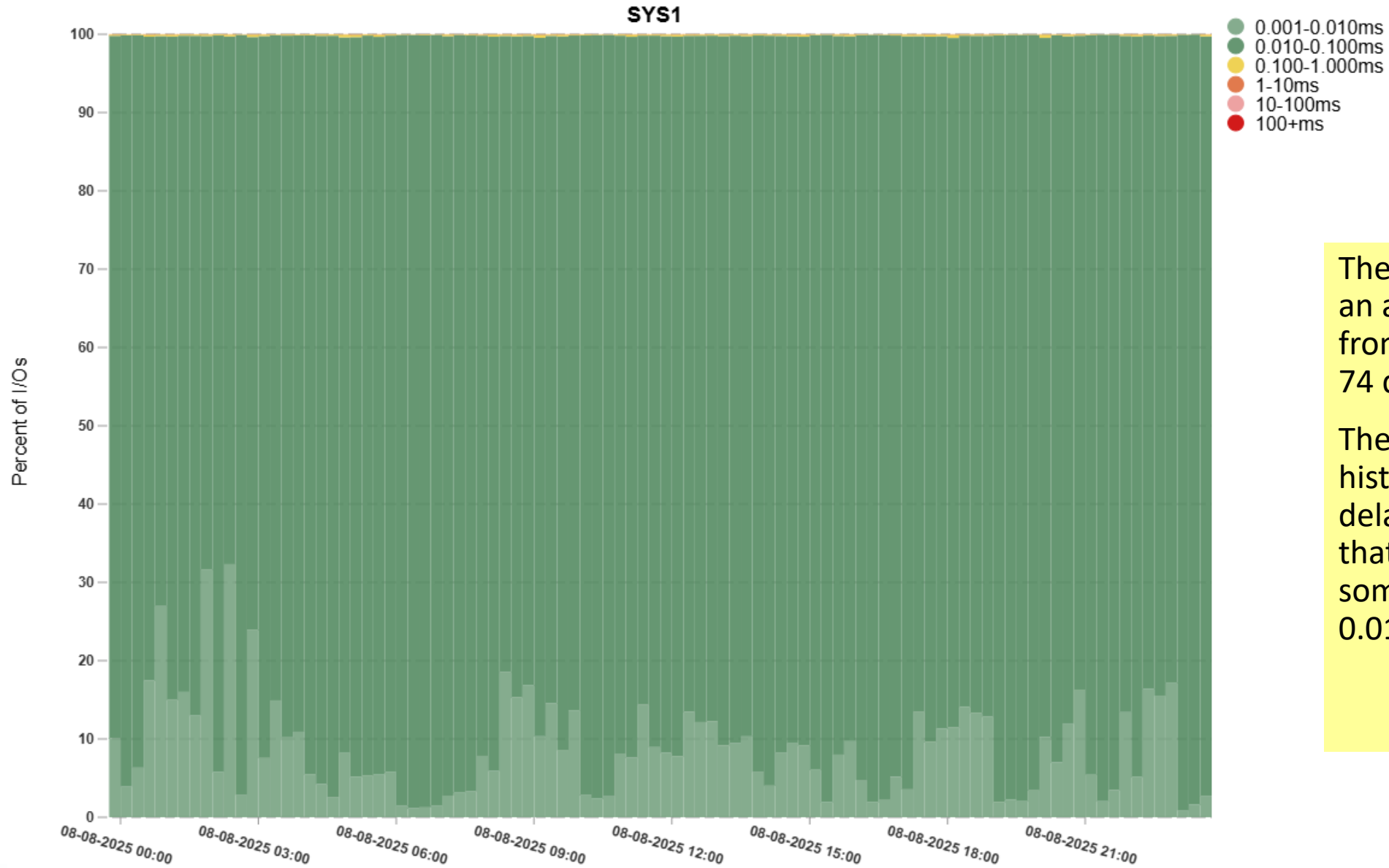


SYS3 has 1 medium polarity CPs and 3 low polarity CPs. Guaranteed share is only 0.16 CPs.

Huge interrupt delay in some intervals!

Note these two systems are on the same CEC and in the same sysplex.

# System I/O Interrupt Delay Distribution



The previous chart was an average calculated from values in the SMF 74 data.

The SMF 42 data has a histogram of interrupt delays. That data shows that most I/Os suffered some interrupt delays of 0.010 to 0.100ms.

# More CPUs = More Better



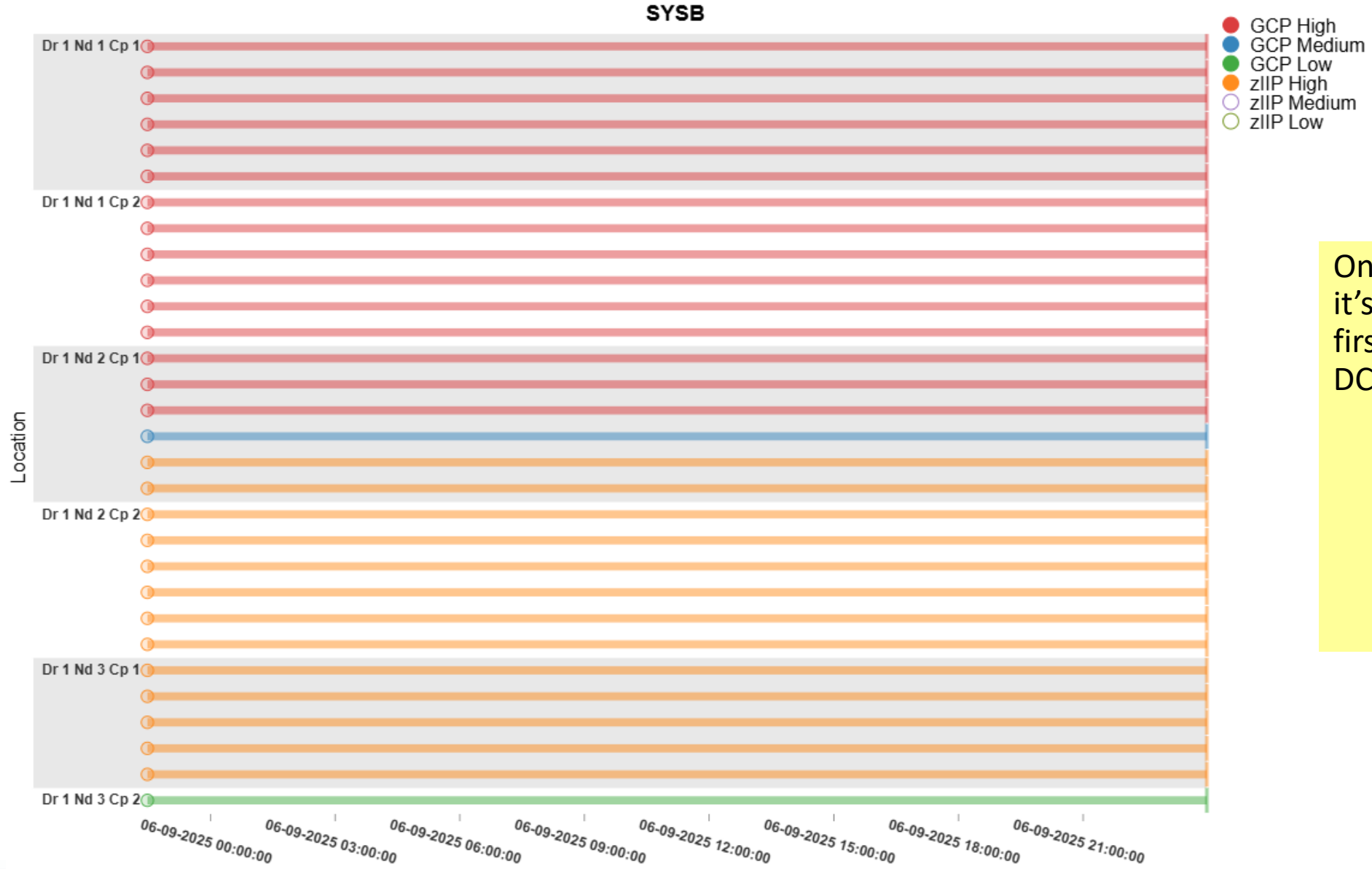
- Getting more CPUs by switching to “slower” CPUs can be quite beneficial
  - More L1/L2 cache and less cache contention = less CPU consumed / unit of work
    - I’ve seen customers make a fewer/faster decision and immediately see a bump up in their MLC software bill
  - More CPUs mean more concurrently executing tasks = more throughput
  - Avoiding execution on low pool or low weight mediums avoids possibly stranding work on CP waiting for PR/SM
  - High pool CPUs lead to less I/O interrupt delay
- Most customers should be using sub-capacity machines
- Many customers should be looking at slower CPUs than they currently have
- The more important LPARs on the machine, the more important this is

# LPAR Sizes



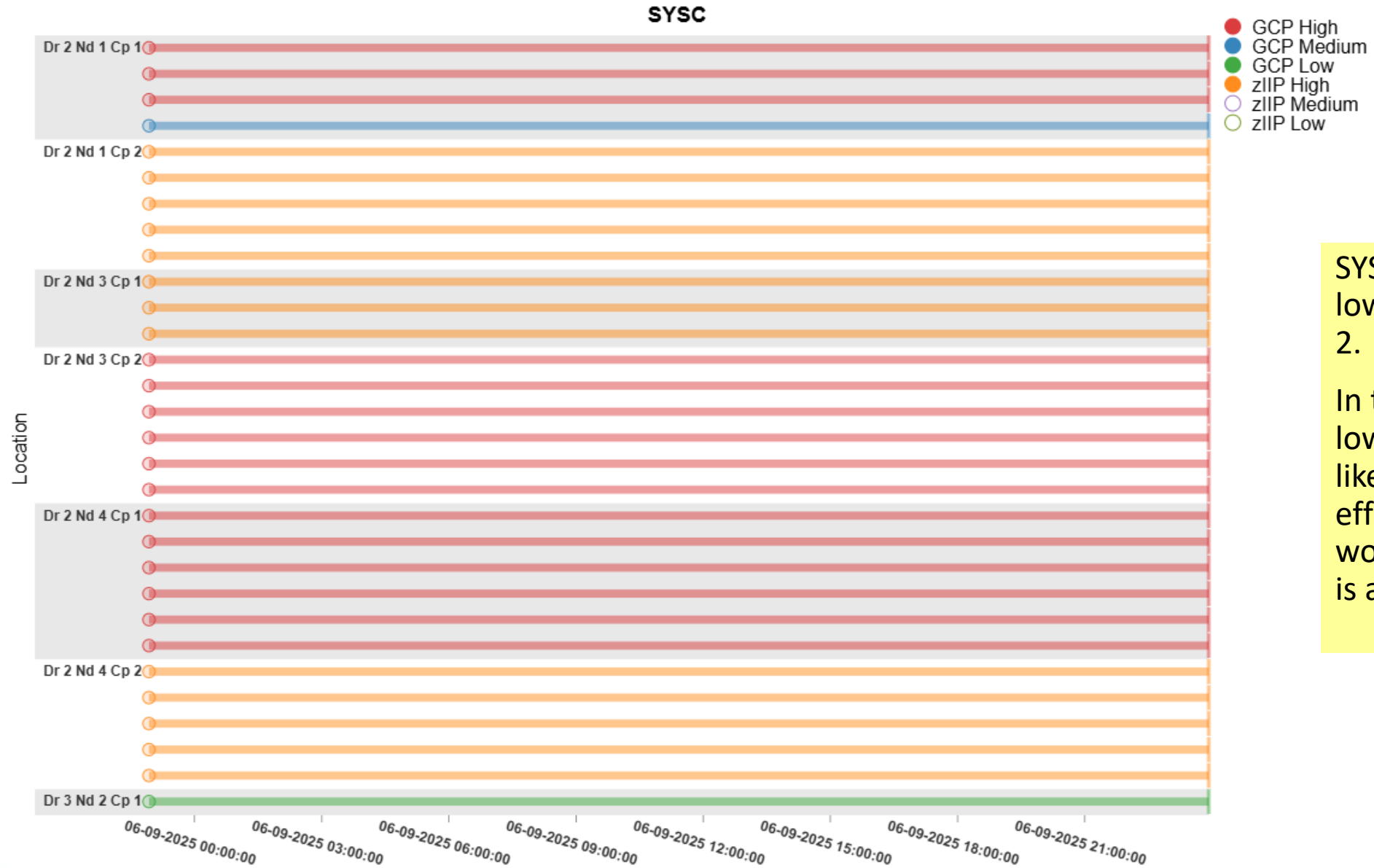
- If you have very large LPARs and a multiple drawer machine...
  - Large = dozens of CPs and zIIPs for an LPAR and/or multiple TBs of memory
- Ideally keep an individual LPAR “small” enough to fit into a single drawer
  - CPs and zIIPs total count  $\leq$  max per drawer
    - Generally easy to plan for
  - Memory  $\leq$  drawer max
    - May be harder to plan for, discuss with IBM during configuration planning
    - Probably somewhat less important than CPs/zIIPs since it only affects L4 cache misses
- Max208 may be better than a Max183 for this reason
  - IBM claims the cost difference is “not that much” (YMMV?)
- May want to consider upsizing memory so there’s enough on each drawer (check memory configurations)

# Processor Location Assignments



On this z16, SYSB has all its GCPs and zIIPs in the first drawer on 3 of the 4 DCMs.

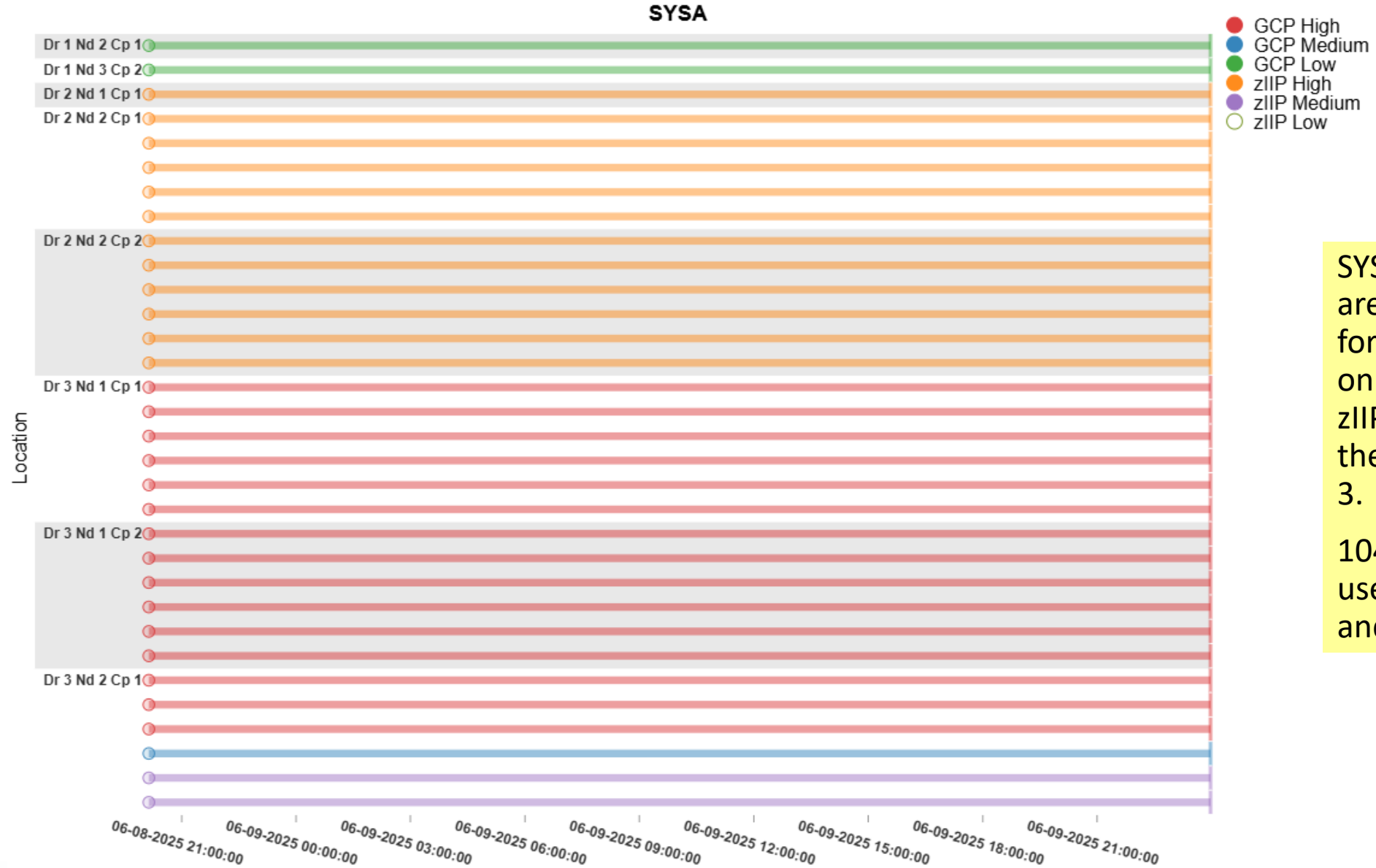
# Processor Location Assignments



SYSC has all but its one low-pool GCP on drawer 2.

In this situation, that low-pool processor will likely run noticeably less efficiently (when it runs work) because the data is all in the other drawer.

# Processor Location Assignments



SYSA is scattered: GCPs are on drawer 3, except for the lows which are on drawer 1. The high zIIPs are on drawer 2 but the medium zIIPs are on 3.

104 (of 125) CPs are in use, including 12 ICFs and IFLs.

# Speaking of memory...



- Generally: buy more!
  - Especially true for the “small” machines that start at 64GB instead of 512GB
- Use what you have!
  - Especially true for “large” machines that grew from smaller historic machines
- Use memory to avoid I/O
  - The only good I/O is no I/O is still true
  - Increasing buffer pools is easier, faster, and less CPU compared to zHyperLink

# Finally: zIIPs



- Buy more if currently using SMT to get extra capacity
  - 2 zIIPs are much better than 1 zIIP with SMT
- My usual preference: Don't use SMT on a regular basis
  - Keep it in your back pocket in case you need more capacity before the next upgrade
- There are some situations where SMT is useful on a regular basis though
  - Systems with lots of concurrent zIIP eligible tasks
  - Low-weight LPARs that have sometimes significant zIIP needs
- Also: Buy more if you will likely need them during the life of the machine
  - Probably will not get cheaper than ordering them as part of the upgrade
  - Unless you're looking to do a separate deal on something IBM is trying to incent

# Summary



- Processors get faster over time
  - “Faster” = more useful work done per unit of time
  - “Faster” ≠ faster clock speed
- Architectural changes often more important than clock speed changes
  - But z17 gets a clock speed bump too
- Early results with z17 migrations have been pretty much as expected
  - I.E. if you change your CPU counts or installed capacity you should expect some changes
  - Have seen channel measurement issues that have been resolved (apparently) with maintenance