

# Fewer/Faster vs More/Slower: Practical Considerations

**Scott Chapman**

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



z/OS Performance  
Education, Software, and  
Managed Service Providers



Creators of Pivotor®

Enterprise Performance Strategies, Inc.  
3457-53rd Avenue North, #145  
Bradenton, FL 34210

<http://www.epstrategies.com>

<http://www.pivotor.com>



# Contact, Copyright, and Trademark Notices

---



## Questions?

Send email to Scott at [scott.chapman@EPStrategies.com](mailto:scott.chapman@EPStrategies.com), or visit our website at <http://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®**, **Reductions®**, **Pivotor®**

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries: IBM®, z/OS®, zSeries®, WebSphere®, CICS®, DB2®, Db2®, S390®, WebSphere Application Server®, and many, many others.

Other trademarks and registered trademarks may exist in this presentation.

# 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

# z/OS Performance workshops available

---



**During these workshops you will be analyzing your own data!**

- Essential z/OS Performance Tuning
  - Milwaukee WI, June 10-14, 2019
- Parallel Sysplex and z/OS Performance Tuning
  - Via the internet, November 12-14, 2019
- WLM Performance and Re-evaluating Goals
  - Virginia Beach VA, October 21-25, 2019

# Like what you see?

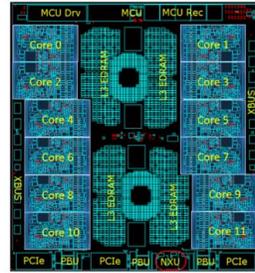
---



- The z/OS Performance Graphs you see here come from Pivotor™ but should be in most of the major reporting products
- If not, or you just want a free cursory review of your environment, let us know!
  - We're always happy to process a day's worth of data and show you the results
  - See also: <http://pivotor.com/cursoryReview.html>



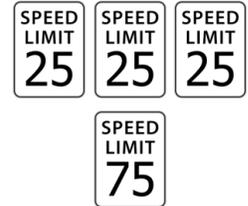
## Processor Performance Details



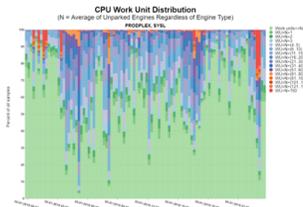
## What should you choose?



## What should you expect from more/slower vs. fewer/faster?



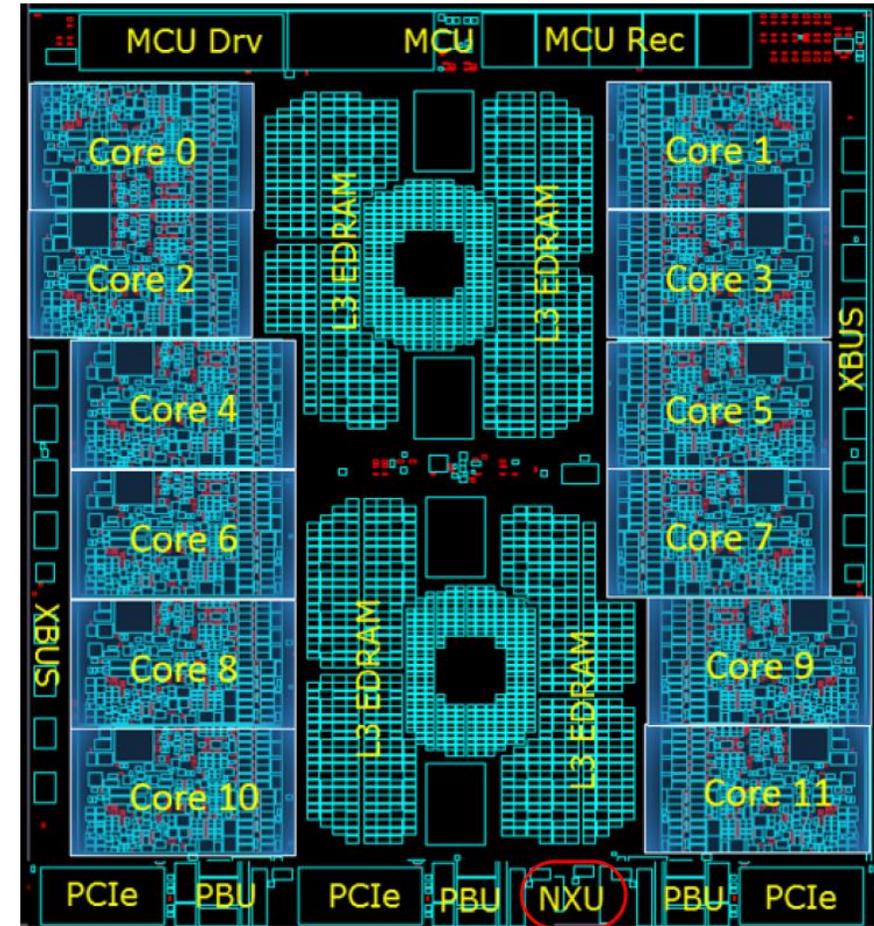
## Interesting measurements



## Practical Considerations - Summary



# Processor Performance Details

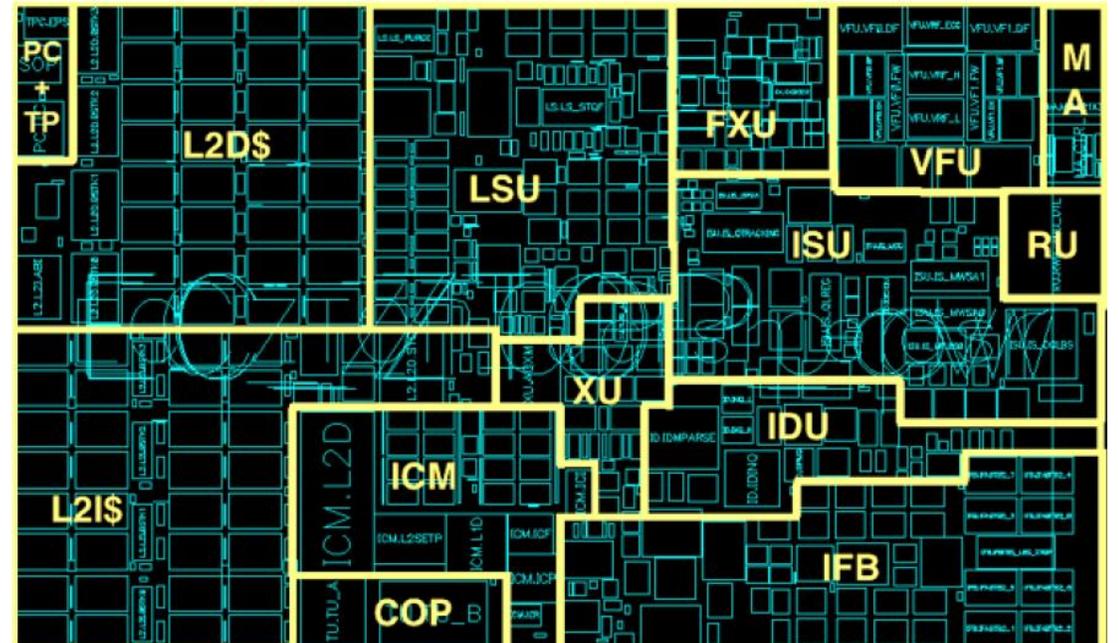


# Clock cycles and effective capacity



- Ideally, you'd like to get real work done each clock cycle
- z Processor speeds are really fast
  - z10 – 4.4 Ghz
  - z196 – 5.2Ghz
  - zEC12 – 5.5Ghz
  - z13 – 5.0 Ghz
  - z14 – 5.2 Ghz
  - z15 – 5.2 Ghz
- So 1ms to wait for an I/O = millions of clock cycles

Billions of cycles per second  
1 Clock cycle = fraction of a **nanosecond** (0.192ns for z14/z15)

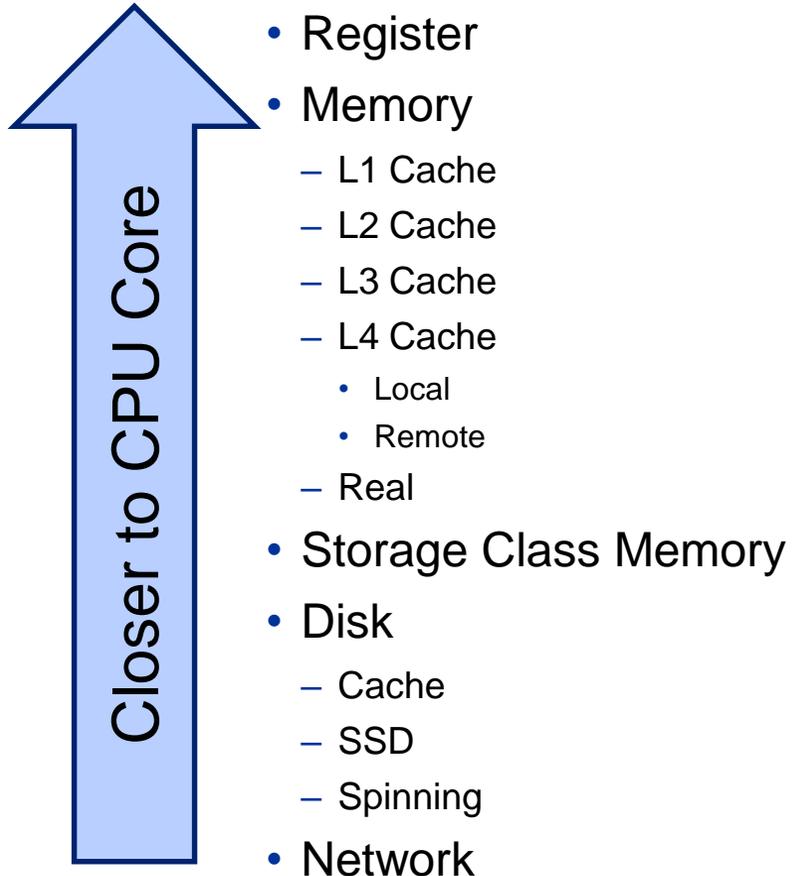


# How long is a cycle again?

---

- Just over 2 inches
  - Light, in a vacuum
  - Electrical signal in a circuit is much slower (40-70% of  $c$ )
  - 1 meter in fiber  $\sim 5$  ns ( $>25$  clock cycles!)
- Need to make a round trip
- Signal paths aren't as the mosquito flies
  - 7.7 Miles of wire in a zEC12 chip
  - Over 13 miles in z13, 14 in z14, 15.6 in z15
- Physical distance matters!

# Data access hierarchy



The farther the data is away from the processor, the more clock cycles will be spent accessing it.

Optimal performance & capacity utilization = keeping data as close to processor as possible!

# Cache utilization & performance

---



- Memory is far away from the processor core and relatively slow
- Effective use of processor cache is important to keeping the processor “fed”
- Cache effectiveness measurements are in the Hardware Instrumentation Services SMF 113 records
  - Requires z/OS 1.8 +PTFs & z10 GA2
- **Enable HIS and record the 113 records**
  - Required for effective capacity planning on upgrade

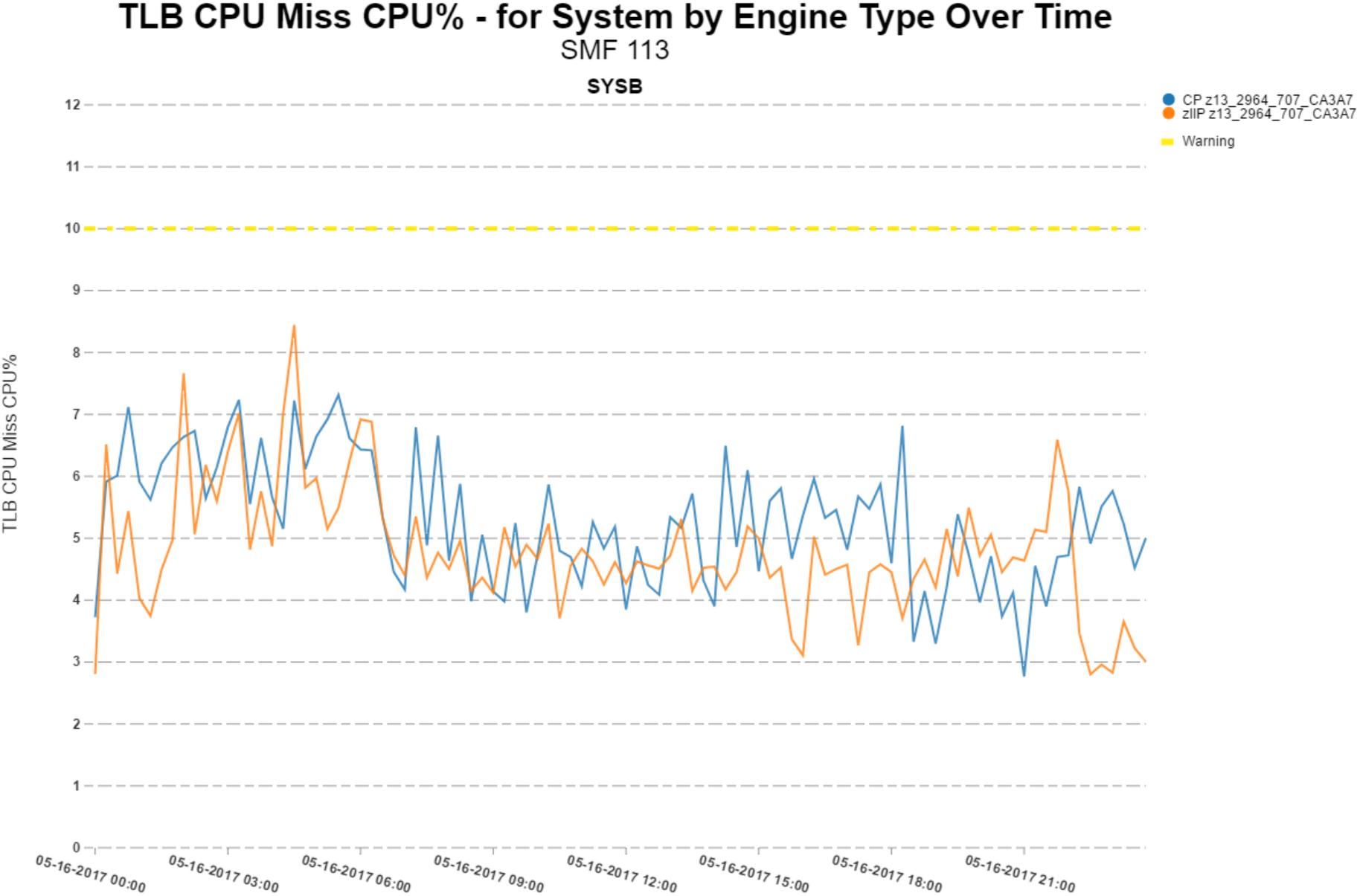
# Dynamic Address Translation

---

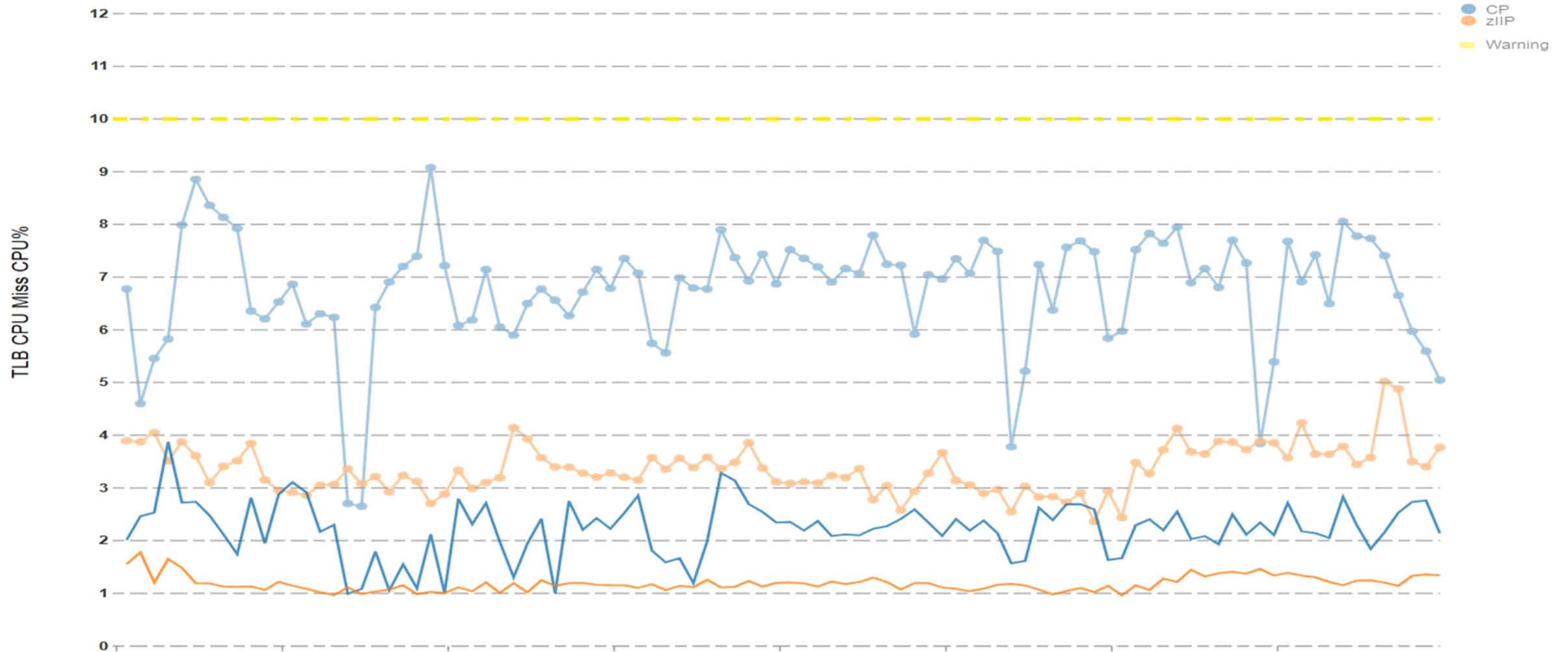


- DAT performed using multiple tables that point to different ranges of storage
- DAT is not free!
- Result of DAT cached in Translation Look-aside Buffers (TLB)
- TLBs are in L1 cache and managed by the hardware
  - Relatively small
- **1MB & 2GB pages make TLBs more effective**
  - **Larger pages = Fewer pages = Fewer TLB entries required**
    - 100GB = 50 2GB pages = 102,400 1MB pages = 26,214,400 4K pages

# Estimated impact of TLB Misses



# zEC12 to z14 TLB CPU Miss % improvement



zEC12: pale lines with markers, z14: darker lines without markers

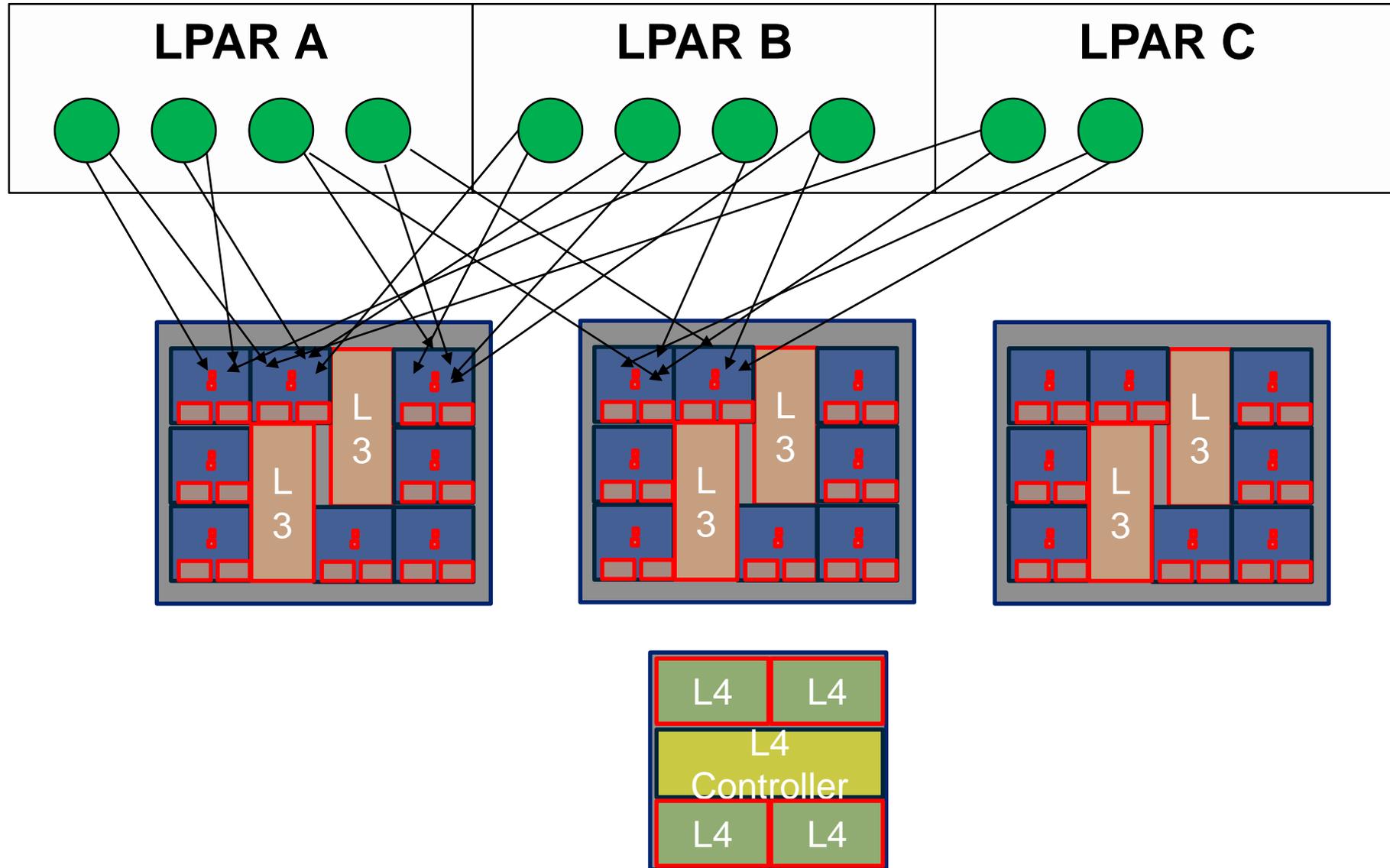
# HiperDispatch Terms

---

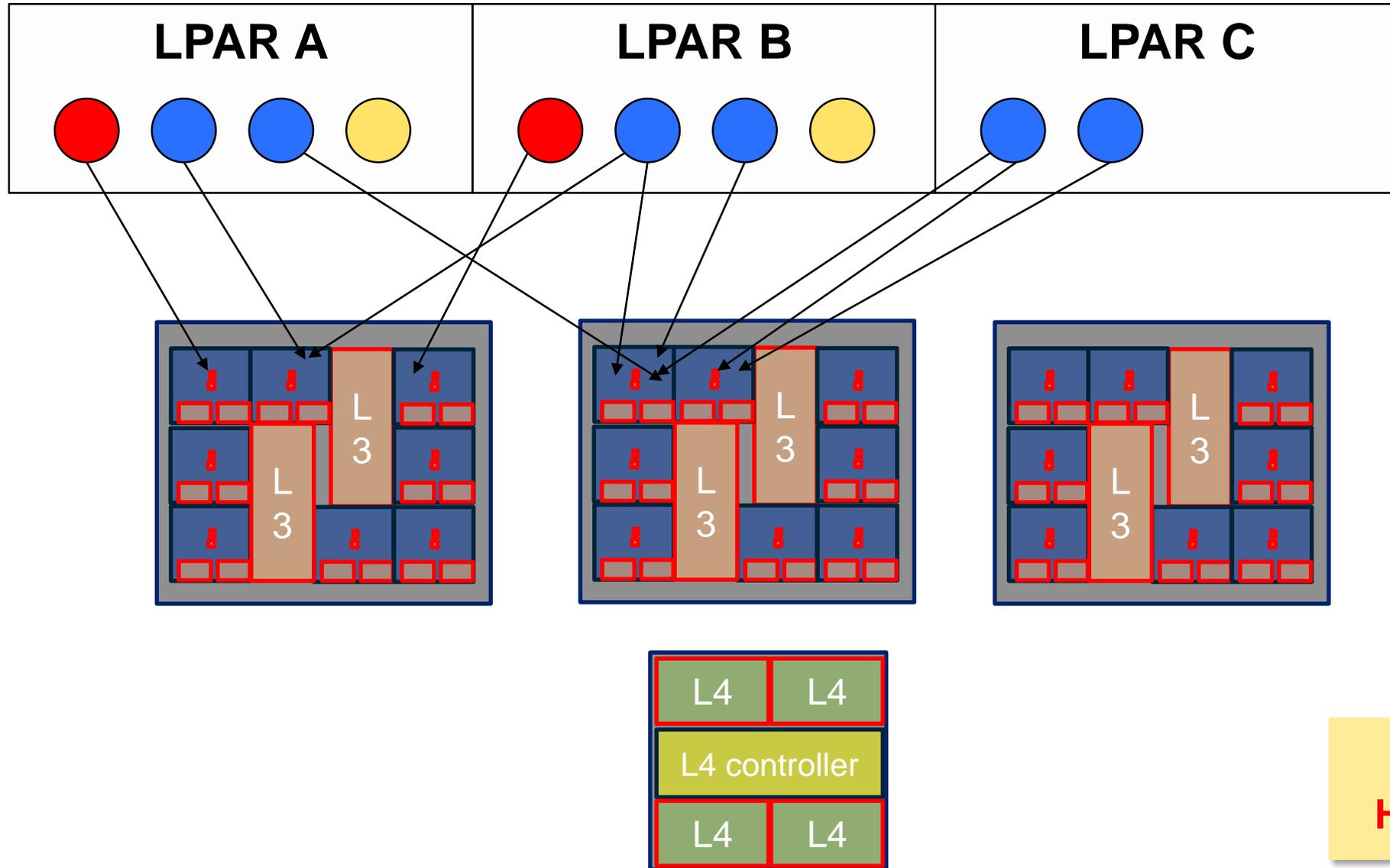


- Logical processors classified as:
  - High – The processor is essentially dedicated to the LPAR (100% share)
  - Medium – Share between 0% and 100%
  - Low – Unneeded to satisfy LPAR's weight
- This processor classification is sometimes referred to as “vertical” or “polarity” or “pool”
  - E.G. Vertical High = VH = High Polarity = High Pool = HP
- Parked / Unparked
  - Initially, VL processors are “parked”: work is not dispatched to them
  - VL processors may become unparked (eligible for work) if there is demand and available capacity

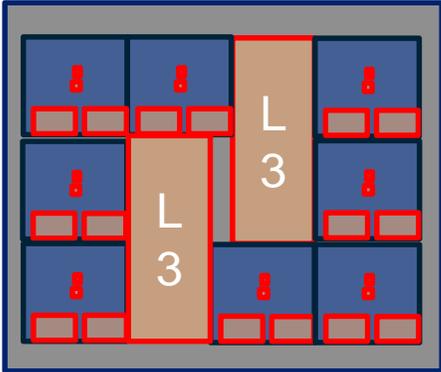
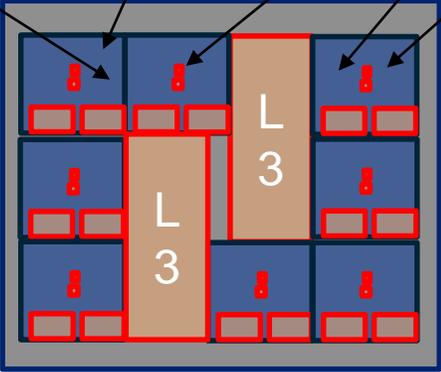
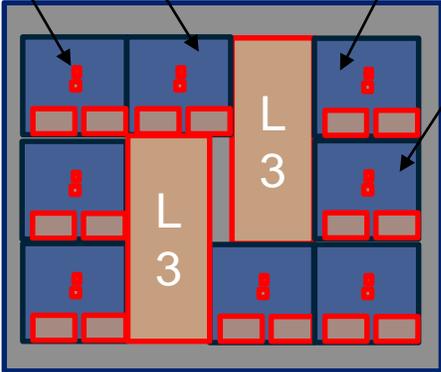
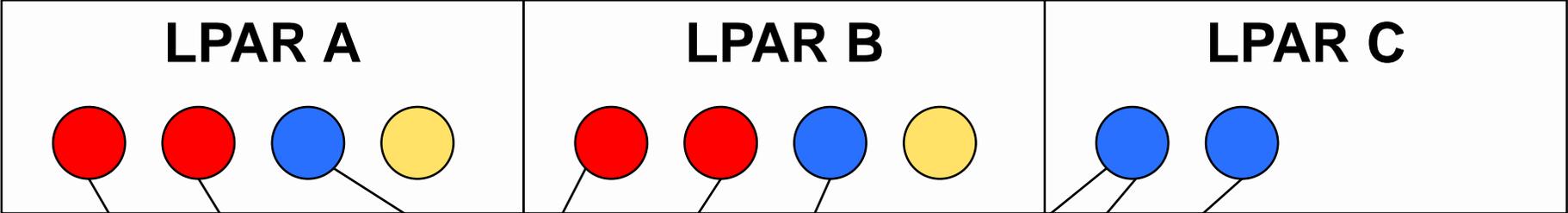
# HiperDispatch off (5 CPUs)



# HiperDispatch on (5 CPUs)

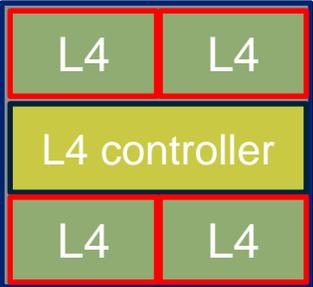
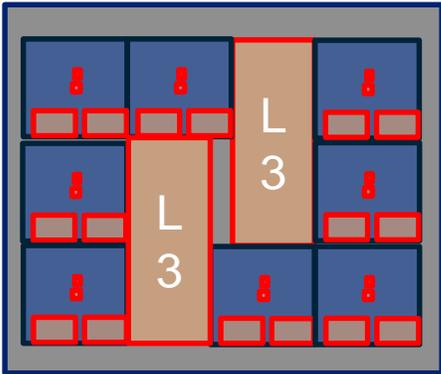
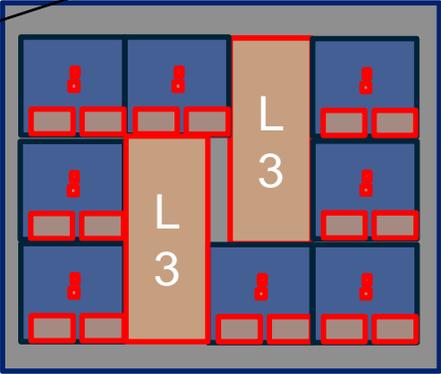
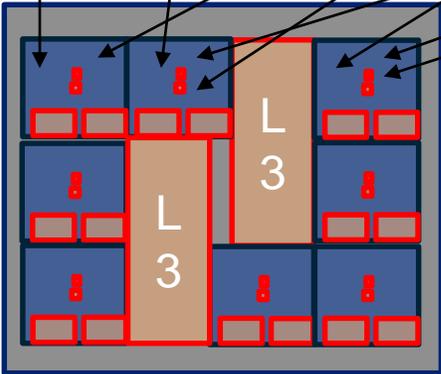
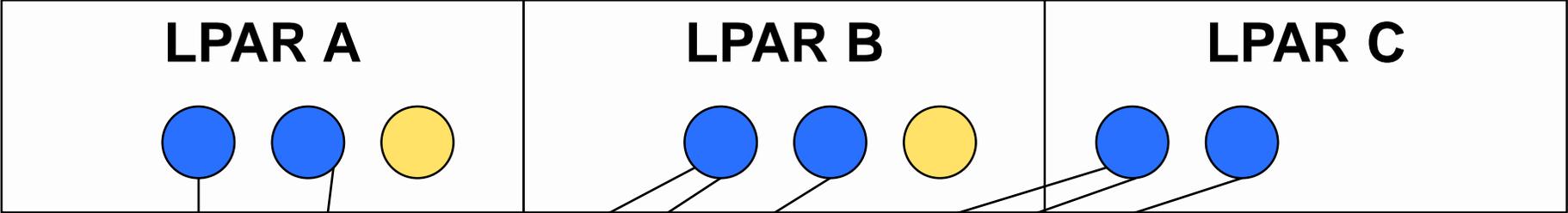


# More/Slower (7 CPUs)



More L1/L2 cache for the work

# Fewer/faster (3 CPUs)



Less L1/L2 cache for the work

# How can you improve cache effectiveness?

- Enable HiperDispatch
- Make good use of large pages
- Upgrade to newer machine

zGen	Name	Year	GHz	701 PCI	701 MSUs	Processor Cache					
						Core-level				Chip	Book-dwr
						L1-Data	L1-Instr	L2-Data	L2-Instr	L3/chip	L4/bk-dwr
z9	z9 EC	2005	1.7	560	81	256K	256K	n/a	n/a	n/a	40M
z10	z10 EC	2008	4.4	902	115	128K	64K	3M		n/a	48M
z11	z196	2010	5.2	1202	150	128K	64K	1.5M		24M	192M
z12	zEC12	2012	5.5	1514	188	96K	64K	1M	1M	48M	348M
z13	z13	2015	5	1695	210	128K	96K	2M	2M	64M	960M
z14	z14	2017	5.2	1832	227	128K	128K	4M	2M	128M	672M
z15	z15	2019	5.2	2055	253	128K	128K	4M	4M	256M	960M

- Consider more/slower CPUs instead of fewer/faster
  - More CPUs = More L1/L2/TLB

# What should you choose?



# What's in a name (or machine type)?



- General form: mmmm-snn
  - mmmm = machine type
  - s = relative engine speed
    - “EC” machines: 4 (slowest) to 7 (fastest)
    - “BC” machines: A (slowest) to Z (fastest)
  - nn = number of general purpose engines
- Examples:
  - 8561-705 = z15 with 5 full speed engines
  - 3906-604 = z14 with 4 2<sup>nd</sup> fastest speed engines
  - 2964-410 = z13 with 10 slowest-speed engines
  - 2827-507 = zEC12 with 7 2<sup>nd</sup> slowest speed engines

Common name	Machine Type
z15	8561
z14	3906
z14 ZR1	3907
z13	2964
z13s	2965
zEC12	2827
zBC12	2828
z196	2817
z114	2818
z10EC	2097
z10BC	2098
z9EC	2094
z9BC	2096

# Scott's ROTs for CPU counts

---



- Don't configure z/OS with less than 2 logical CPs
  - Possible exception: LPAR is a largely unused sandbox & not in a Sysplex
- Less than 3 physical CPs on a machine troublesome
  - 2 possibly ok if single primary LPAR
- Even “minor” changes may make a difference, depending on your LPAR configuration
  - Going from 3 faster to 5 slower not as dramatic as 5 to 10 or 3 to 8
  - “not as dramatic” doesn't mean ignore it
  - If a couple of extra CPs means more high-polarity processors, that might be a very good thing
- **Upgrading to new generation may require changing engine speeds**

# Upgrade scenario examples



- Upgrade z196/z114 to z13
- Upgrade z13 to z15
- Keep MSUs about the same
  - Easily done for MLC sub-capacity, but consider the ISVs
- Explore the options with zPCR
  - Completely fictional configs, but hopefully somewhat representative

**z196 2817-704**  
**531 MSUs**

**A**

6 LPARs

**z196 2817-504**  
**265 MSUs**

**B**

4 LPARs

**z114 2818-V02**  
**92 MSUs**

**C**

3 LPARs

**z13 2964-705**  
**905 MSUs**

**D**

5 LPARs

# Scenario D, Total Capacity



Host Capacity Comparison Summary zPCR V9.3

**LPAR Host Capacity Comparison Report**  
 Capacity basis: 2094-701 @ 1.000 ITRR for a shared single-partition configuration  
 Capacity for z/OS on z10 and later processors is represented with HiperDispatch turned ON

LPAR Configuration				Full Capacity (based on usable RCP count)					
Identity	Hardware	SMT	GP*	zAAP	zIIP	IFL	ICF	Total	
#1	z13-705, 905 MSUs, 7392 PCI	2964-N30/700: GP=5 zIIP=2 ICF=1	12.715	n/s	5.017			17.732	
#2	z15-704, 914 MSUs, 7467 PCI	8561-T01(Max34)/700: GP=4 zIIP=2 ICF=1	12.743	n/s	6.307			19.050	
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		+0.2%		+25.7%			+7.4%	
#3	z15-608, 966 MSUs, 7890 PCI	8561-T01(Max34)/600: GP=8 zIIP=2 ICF=1	13.909	n/s	6.114			20.023	
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		+9.4%		+21.9%			+12.9%	
#4	z15-511, 875 MSUs, 7138 PCI	8561-T01(Max34)/500: GP=11 zIIP=2 ICF=1	12.685	n/s	5.942			18.627	
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		-0.2%		+18.4%			+5.0%	
#5	z15-512, 941 MSUs, 7690 PCI	8561-T01(Max34)/500: GP=12 zIIP=2 ICF=1	13.783	n/s	5.941			19.724	
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		+8.4%		+18.4%			+11.2%	
#6	z15-705, 1117 MSUs, 9170 PCI	8561-T01(Max34)/700: GP=5 zIIP=2 ICF=1	15.866	n/s	6.272			22.137	
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		+24.8%		+25.0%			+24.8%	

Content Control

Show Capacity Deltas

Based on "z13-705, 905 MSUs, 7392 PCI "

Incremental

Show capacity as

Full CPC

Single-CP

For significant configuration changes such as upgrading the processor family, consider capacity comparisons to have a +/-5% margin-of-error.

IBM does not guarantee the results from this tool. This information is provided "as is", without warranty, expressed or implied. You are responsible for the results obtained from your use of this tool.

# Scenario D, Single CP



Host Capacity Comparison Summary zPCR V9.3

**LPAR Host Capacity Comparison Report**  
 Capacity basis: 2094-701 @ 1.000 ITRR for a shared single-partition configuration  
 Capacity for z/OS on z10 and later processors is represented with HiperDispatch turned ON

LPAR Configuration				Single-CP Capacity (based on usable RCP count)				
Identity	Hardware	SMT	GP*	zAAP	zIIP	IFL	ICF	Total
#1  z13-705, 905 MSUs, 7392 PCI	2964-N30/700: GP=5 zIIP=2 ICF=1		2.543	n/s	2.509			2.533
#2  z15-704, 914 MSUs, 7467 PCI	8561-T01(Max34)/700: GP=4 zIIP=2 ICF=1		3.186	n/s	3.154			3.175
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		+25.3%		+25.7%			+25.3%
#3  z15-608, 966 MSUs, 7890 PCI	8561-T01(Max34)/600: GP=8 zIIP=2 ICF=1		1.739	n/s	3.057			2.002
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		-31.6%		+21.9%			-21.0%
#4  z15-511, 875 MSUs, 7138 PCI	8561-T01(Max34)/500: GP=11 zIIP=2 ICF=1		1.153	n/s	2.971			1.433
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		-54.7%		+18.4%			-43.4%
#5  z15-512, 941 MSUs, 7690 PCI	8561-T01(Max34)/500: GP=12 zIIP=2 ICF=1		1.149	n/s	2.971			1.409
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		-54.8%		+18.4%			-44.4%
#6  z15-705, 1117 MSUs, 9170 PCI	8561-T01(Max34)/700: GP=5 zIIP=2 ICF=1		3.173	n/s	3.136			3.162
	Percent Delta from "z13-705, 905 MSUs, 7392 PCI "		+24.8%		+25.0%			+24.8%

Content Control

Show Capacity Deltas

Based on "z13-705, 905 MSUs, 7392 PCI "

Incremental

Show capacity as

Full CPC

Single-CP

For significant configuration changes such as upgrading the processor family, consider capacity comparisons to have a +/-5% margin-of-error.

IBM does not guarantee the results from this tool. This information is provided "as is", without warranty, expressed or implied. You are responsible for the results obtained from your use of this tool.

# Scenario A, Total Capacity



## LPAR Host Capacity Comparison Report

Study ID: Scenario A

Capacity basis: 2094-701 @ 1.000 for a shared single-partition configuration  
 Capacity for z/OS on z10 and later processors is represented with HiperDispatch turned ON

LPAR Configuration			Full CPC Capacity (based on usable RCP count)					
Identity	Hardware	SMT	GP	zAAP	zIIP	IFL	ICF	Total
#1  z196 704, 531 MSUs	2817-M15/700: GP=4 zIIP=2 ICF=2		7.124		3.656		3.778	14.557
#2  z13 703, 571 MSUs	2964-N30/700: GP=3 zIIP=2 ICF=2		7.551	n/s	5.091		4.413	17.055
	Percent Delta from "z196 704, 531 MSUs"		+6.0%		+39.3%		+16.8%	+17.2%
#3  z13 605, 577 MSUs	2964-N30/600: GP=5 zIIP=2 ICF=2		7.863	n/s	5.094		4.372	17.329
	Percent Delta from "z196 704, 531 MSUs"		+10.4%		+39.3%		+15.7%	+19.0%
#4  z13 507, 552 MSUs	2964-N30/500: GP=7 zIIP=2 ICF=2		7.695	n/s	5.061		4.333	17.089
	Percent Delta from "z196 704, 531 MSUs"		+8.0%		+38.4%		+14.7%	+17.4%
#5  z13 423, 527 MSUs	2964-N30/400: GP=23 zIIP=2 ICF=2		7.842	n/s	4.622		4.025	16.489
	Percent Delta from "z196 704, 531 MSUs"		+10.1%		+26.4%		+6.5%	+13.3%

### Content Control

Show Capacity Deltas

Based on "z196 704, 531 MSUs"  
 Incremental

### Show capacity as

Full CPC  
 Single-CP

For significant configuration changes such as upgrading the processor family, consider capacity comparisons to have a +/-5% margin-of-error.

IBM does not guarantee the results from this tool. This information is provided "as is", without warranty, expressed or implied. You are responsible for the results obtained from your use of this tool.

# Scenario B, Total Capacity



## LPAR Host Capacity Comparison Report

Capacity basis: 2094-701 @ 1.000 for a shared single-partition configuration  
 Capacity for z/OS on z10 and later processors is represented with HiperDispatch turned ON

LPAR Configuration			Full CPC Capacity (based on usable RCP count)					
Identity	Hardware	SMT	GP	zAAP	zIIP	IFL	ICF	Total
#1  z196 504, 265 MSUs	2817-M15/500: GP=4 zIIP=1 ICF=2		3.679		1.910		3.792	9.381
#2  z13 602, 249 MSUs	2964-N30/600: GP=2 zIIP=1 ICF=2		3.376	n/s	2.695		4.449	10.521
	Percent Delta from "z196 504, 265 MSUs"		-8.2%		+41.1%		+17.3%	+12.1%
#3  z13 503, 255 MSUs	2964-N30/500: GP=3 zIIP=1 ICF=2		3.483	n/s	2.697		4.428	10.608
	Percent Delta from "z196 504, 265 MSUs"		-5.3%		+41.2%		+16.8%	+13.1%
#4  z13 504, 333 MSUs	2964-N30/500: GP=4 zIIP=1 ICF=2		4.592	n/s	2.670		4.408	11.671
	Percent Delta from "z196 504, 265 MSUs"		+24.8%		+39.8%		+16.3%	+24.4%
#5  z13 410, 258 MSUs	2964-N30/400: GP=10 zIIP=1 ICF=2		3.732	n/s	2.519		4.292	10.543
	Percent Delta from "z196 504, 265 MSUs"		+1.4%		+31.9%		+13.2%	+12.4%
#6  z13 411, 281 MSUs	2964-N63/400: GP=11 zIIP=1 ICF=2		4.032	n/s	2.483		4.377	10.892
	Percent Delta from "z196 504, 265 MSUs"		+9.6%		+30.0%		+15.4%	+16.1%
#7  z13 603, 363 MSUs	2964-N30/600: GP=3 zIIP=1 ICF=2		4.971	n/s	2.697		4.428	12.096
	Percent Delta from "z196 504, 265 MSUs"		+35.1%		+41.2%		+16.8%	+28.9%

### Content Control

Show Capacity Deltas

- Based on "z196 504, 265 MSUs"  
 Incremental

### Show capacity as

- Full CPC  
 Single-CP

For significant configuration changes such as upgrading the processor family, consider capacity comparisons to have a +/-5% margin-of-error.

# Scenario C, Total Capacity



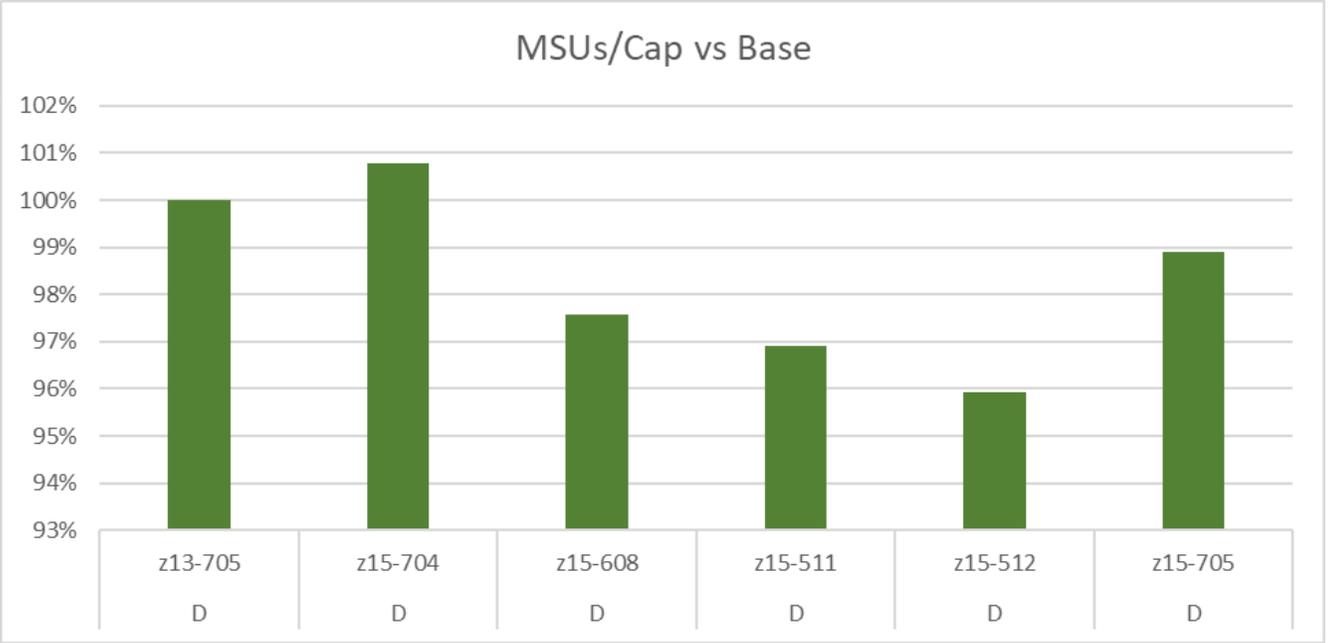
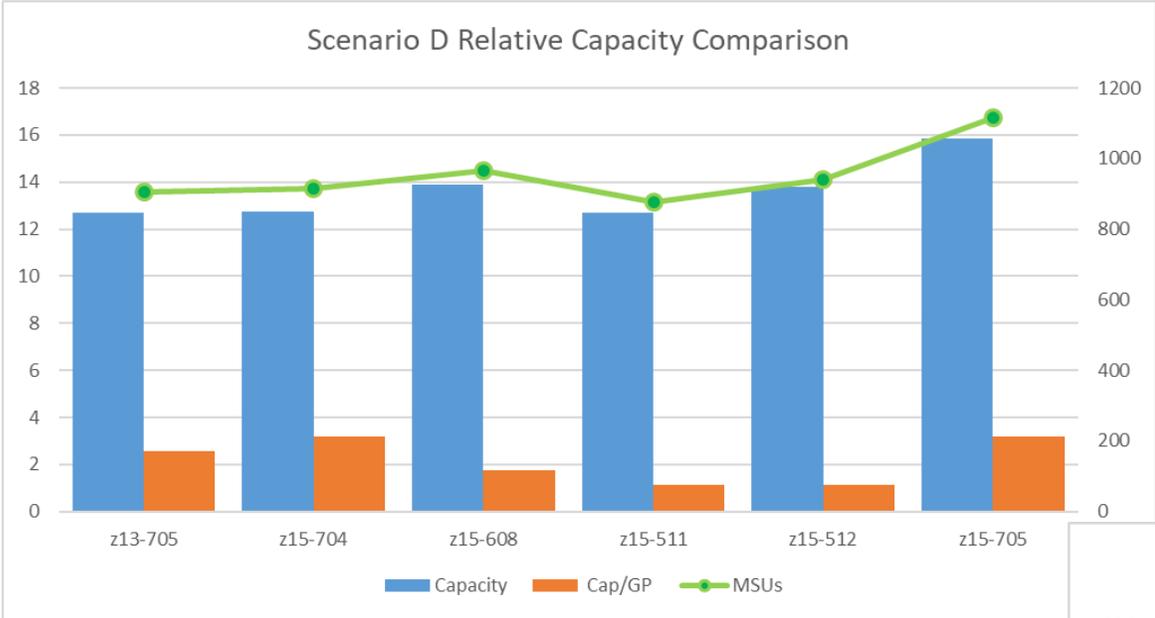
zPCR V8.8

## LPAR Host Capacity Comparison Report

Capacity basis: 2094-701 @ 1.000 for a shared single-partition configuration  
 Capacity for z/OS on z10 and later processors is represented with HiperDispatch turned ON

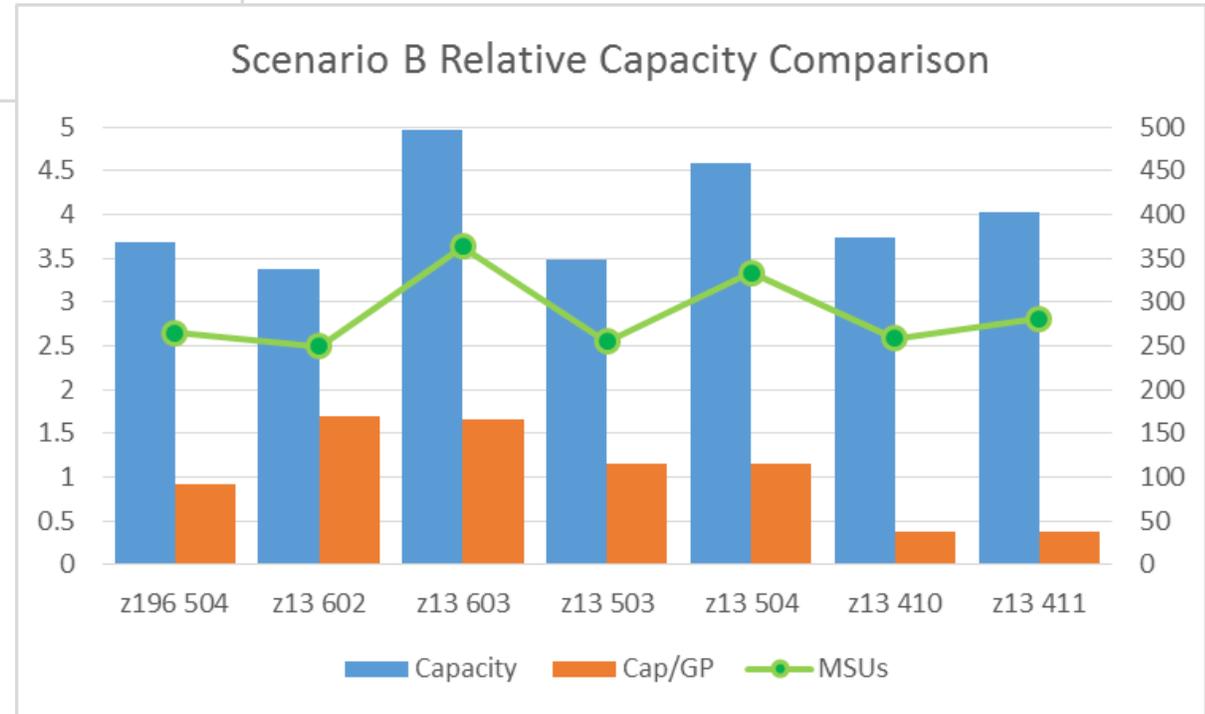
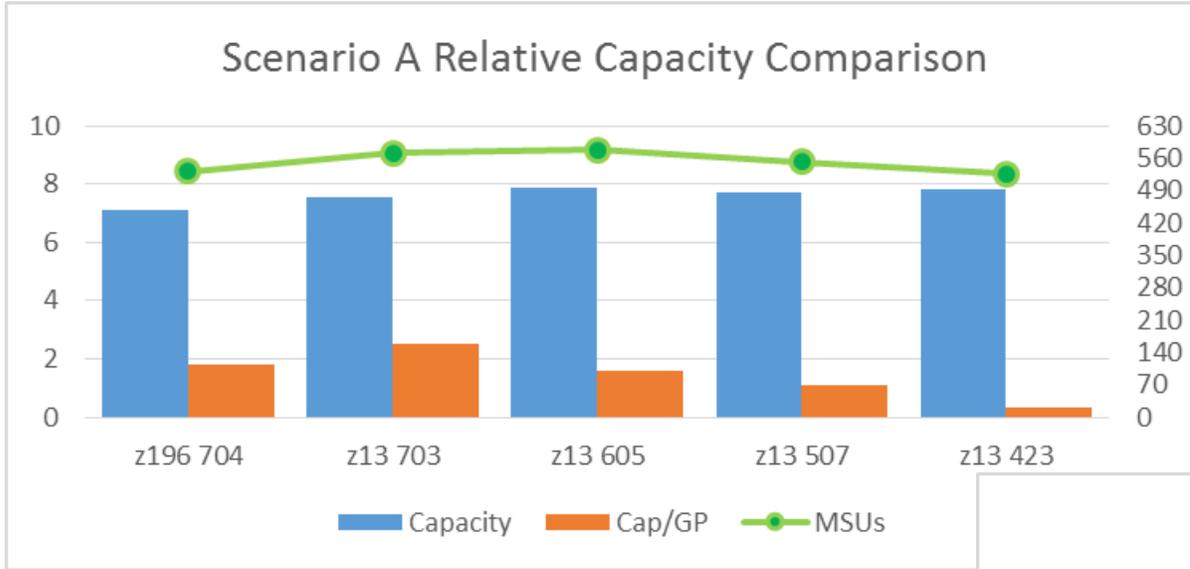
LPAR Configuration				Full CPC Capacity (based on usable RCP count)				
Identity	Hardware	SMT	GP	zAAP	zIIP	IFL	ICF	Total
#1	z114 V02, 98 MSUs	2818-V02: GP=2 zIIP=1		1.307		1.199		2.506
#2	z13s N10 O02, 100 MSUs	2965-O02: GP=2 zIIP=1		1.360		2.305		3.665
		Percent Delta from "z114 V02, 98 MSUs"		+4.0%		+92.2%		+46.2%
#3	z13s N10 K03, 92 MSUs	2965-K03: GP=3 zIIP=1		1.268		2.252		3.521
		Percent Delta from "z114 V02, 98 MSUs"		-3.0%		+87.8%		+40.5%
#4	z13s N10 I04, 95 MSUs	2965-I04: GP=4 zIIP=1		1.311		2.203		3.514
		Percent Delta from "z114 V02, 98 MSUs"		+0.3%		+83.7%		+40.2%
#5	z13s N10 G05, 92 MSUs	2965-G05: GP=5 zIIP=1		1.278		2.156		3.433
		Percent Delta from "z114 V02, 98 MSUs"		-2.3%		+79.8%		+37.0%
#6	z13s N10 F06, 93 MSUs	2965-F06: GP=6 zIIP=1		1.310		2.111		3.420
		Percent Delta from "z114 V02, 98 MSUs"		+0.2%		+76.0%		+36.5%

# Scenario D Comparisons

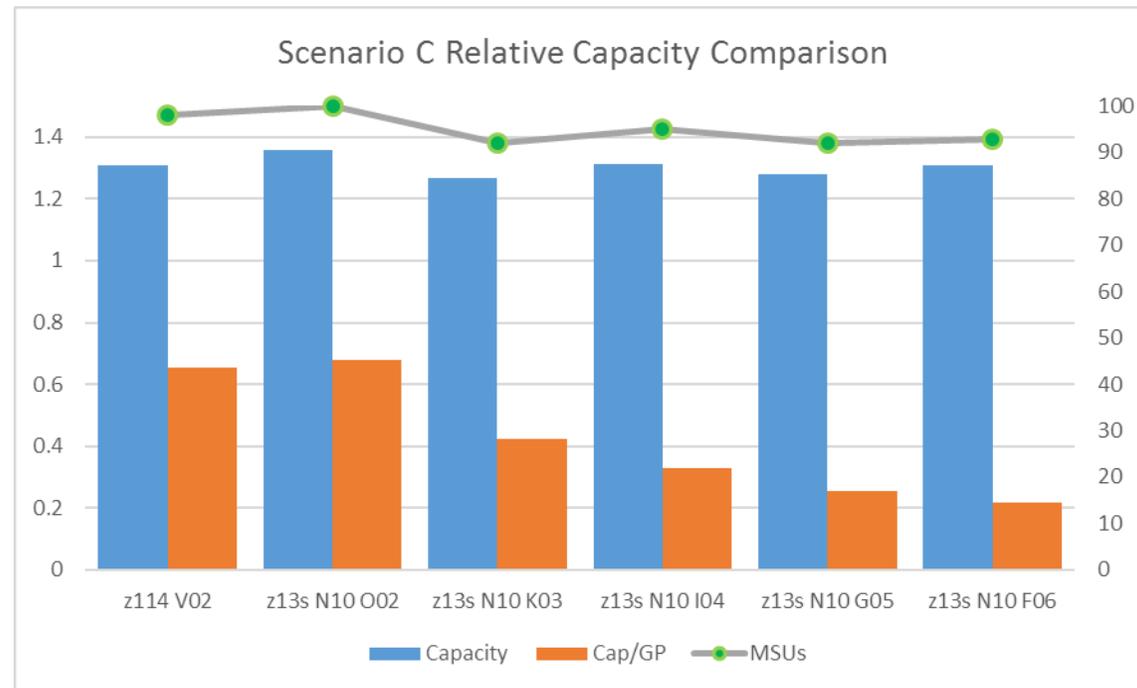


The slower speed engines deliver more relative capacity per MSU!

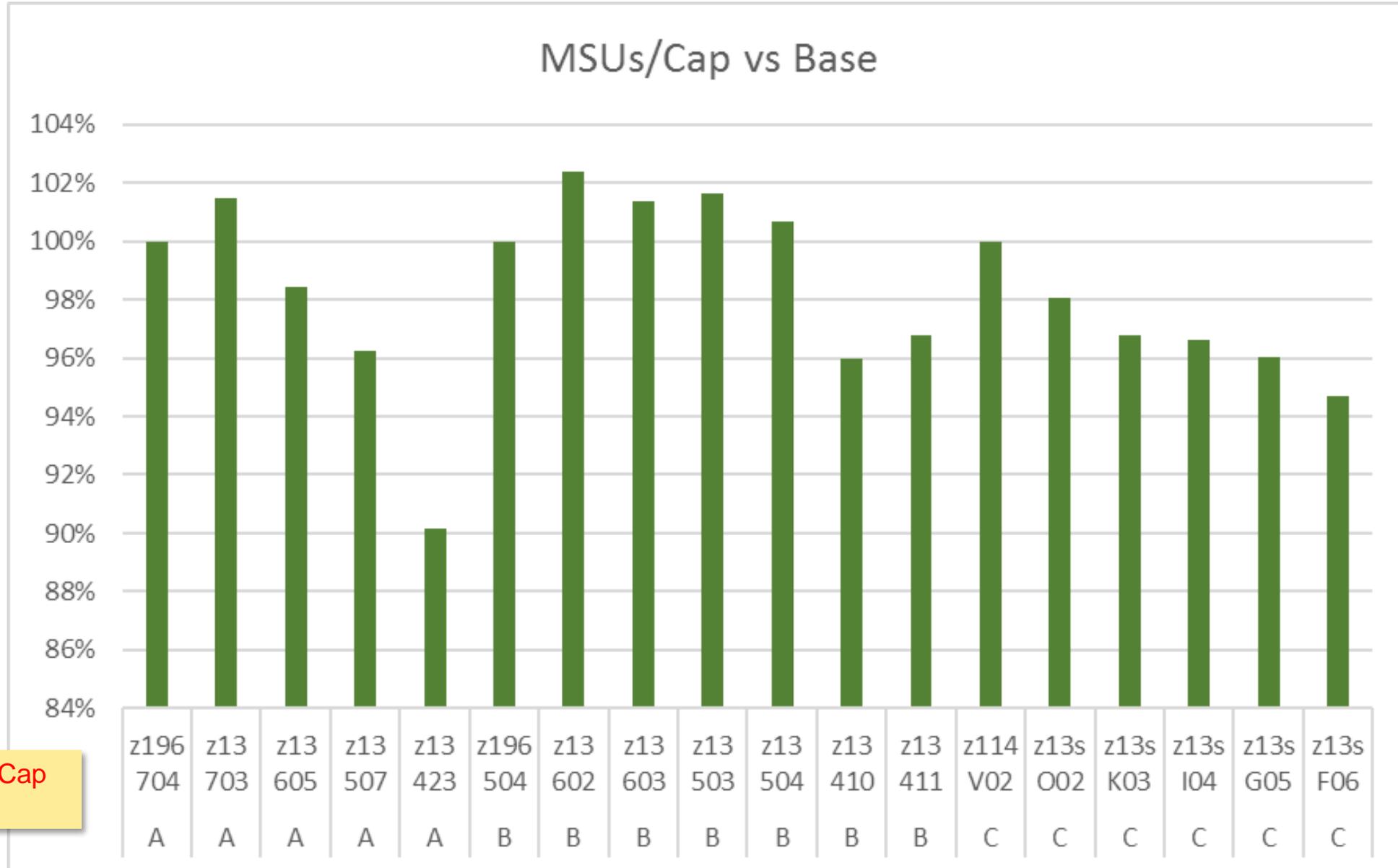
# Scenario Comparisons



# Scenario C comparisons



# Impact on software cost



Fewer MSUs/Cap is better!

# What does this mean?

---



- When considering an upgrade you should consider all engine speed options
- The choice might impact your software bill, at least slightly
- New z15 benefit of slow speed engines: more System Recovery Boost capacity!
- The big question is: what happens if you change the engine speed?

**What should you expect from more/slower vs. fewer/faster?**





# Components of elapsed times

---

- CPU time – time spent using the CPU
  - Directly impacted by CPU speed
- CPU wait/delay time – time spent waiting to get on a CPU
  - Related to number & speed of CPUs
- I/O time – time spent doing I/O
  - Likely won't change substantially with CPU changes
- Database or other subsystem request time
  - This will likely include some CPU time that's not charged back to the job
  - Also likely includes some I/O time
- Serialization (locking, enqueues, latches, etc.)
  - Can be related to how fast other work is running

# What to expect (fewer/faster)

---



- Lower CPU times
  - If you double the processor speed, then CPU time should drop by about half
  - Of course, things will not be so simple
- Possibly higher CPU delays
  - Fewer processors = more queueing
- Possibly more variable throughput
  - High priority, CPU-intensive tasks can monopolize more of the total capacity
  - Fewer high-polarity CPUs could mean more waiting for other LPARs
  - Sync CF requests consume more of your capacity, unless they get faster too
    - Always best to keep the CF technology on the same level as your CPU

# What to expect (more/slower)

---



- Higher CPU times
  - If you halve the processor speed, then CPU time should double
  - Of course, things will likely not be so easy
- Possibly lower CPU delays
  - More processors = less queueing
- Possibly more consistent throughput
  - High priority, CPU-intensive tasks can monopolize less of the total capacity
  - More high-polarity processors mean less waiting for other LPARs
  - Sync CF requests consume less of your capacity
    - Note CF engines always run at full speed
  - More z15 System Recovery Boost capacity

# Professionals and Jailbirds



## Fewer/Faster

- + Lower CPU time
- + Better for single-threaded CPU-intensive workloads
- + Specialty engine capacity constraints cause less impact
- + Scalable to multiple books/drawers
- CPU spinning/waiting is more of total capacity (Higher Parallel Sysplex overhead)
- Less L1/L2 cache
- Potentially fewer high polarity processors (More inter-LPAR impacts)
- Fewer concurrently executing tasks

## More/Slower

- Higher CPU times
- Worse for single-threaded CPU-intensive workloads
- Specialty engine capacity constraints cause more impact
- Not scalable past one book/drawer
- + CPU spinning/waiting is less of total capacity (Lower Parallel Sysplex overhead)
- + More L1/L2 cache
- + Potentially more high polarity processors (Less inter-LPAR impact)
- + More concurrently executing tasks

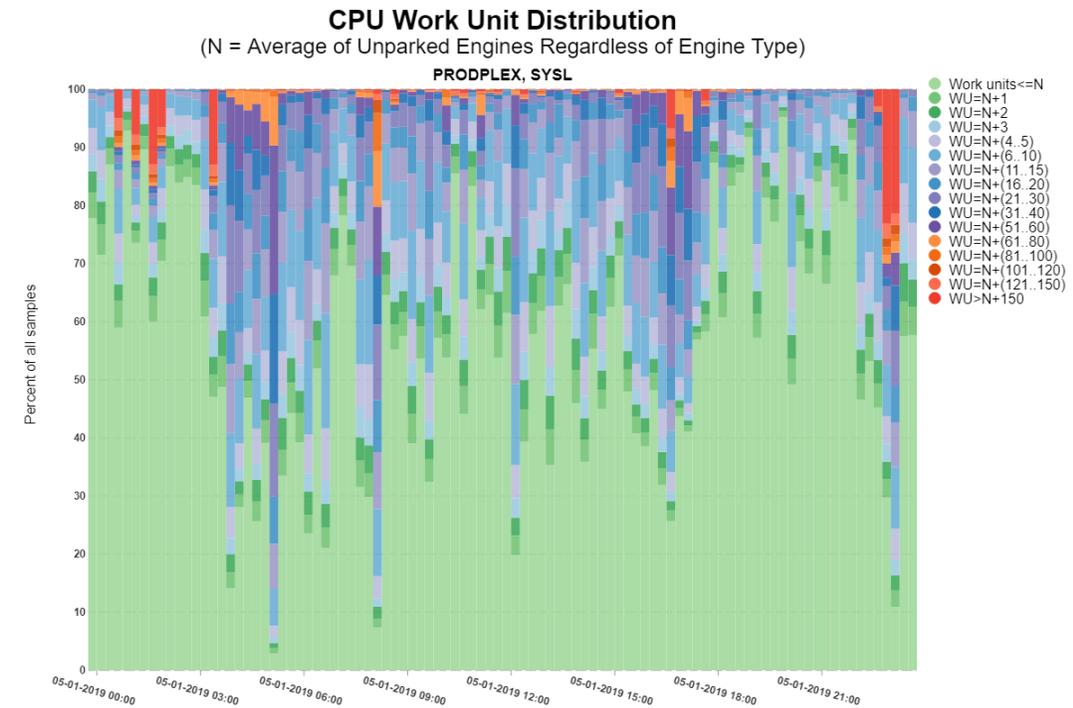


Best for  
single-task  
performance



Best for  
system efficiency

# Interesting measurements



# Work units

---

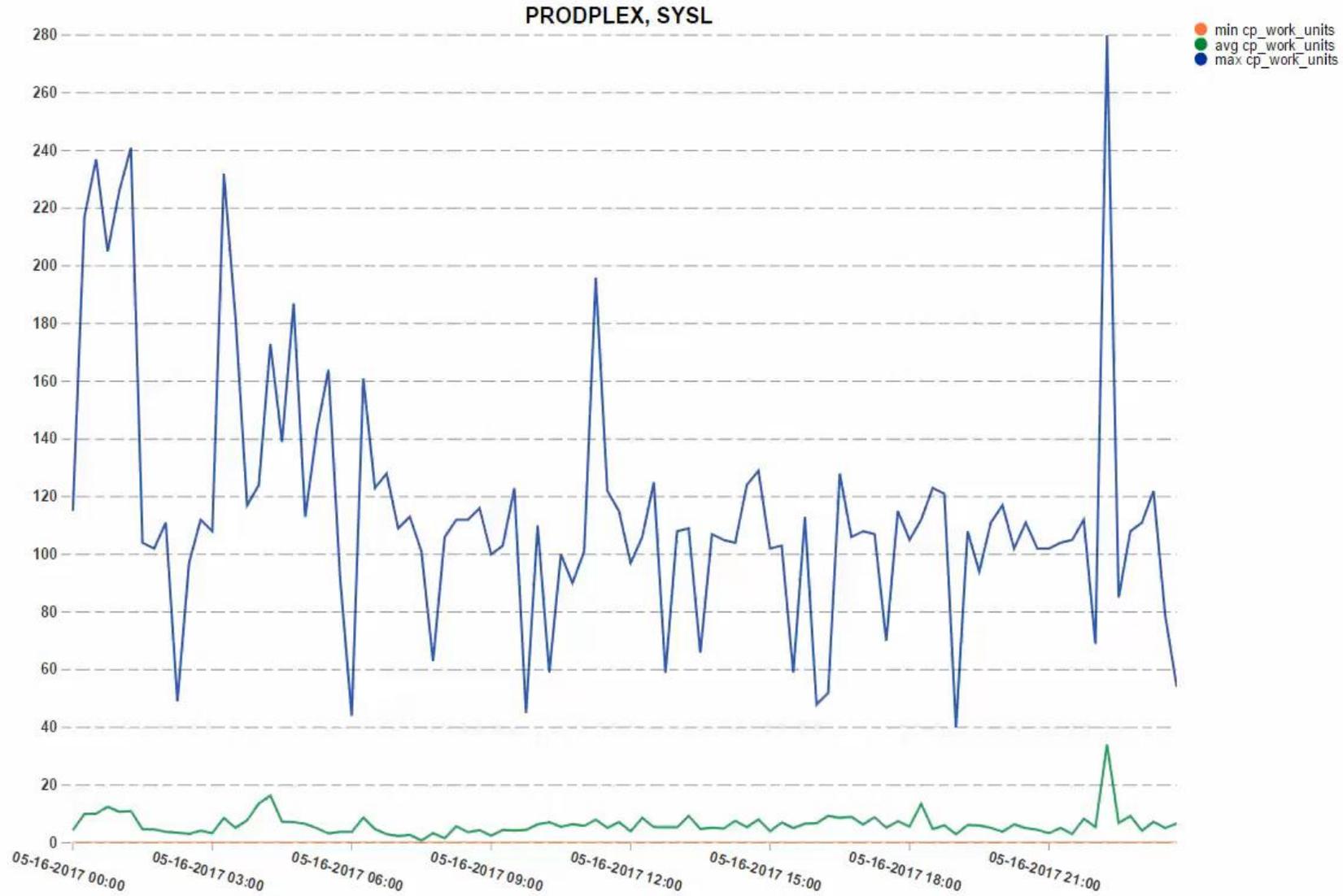


- Work units is the replacement for the “in-ready” address space counts
  - Counts running or waiting work units
- More accurate representation of work because we have an increasing number of multi-threaded address spaces
- Values by processor type (GP, zIIP, zAAP)
- Plot min, average, max over time
  - Max is often far larger than average
- Distribution of observations
  - Based on the number of online and not parked processors (N)
  - Counts in buckets: N, N+1, N+2, N+3, N+5, N+10, N+15, N+20, N+30...

# Work units over time



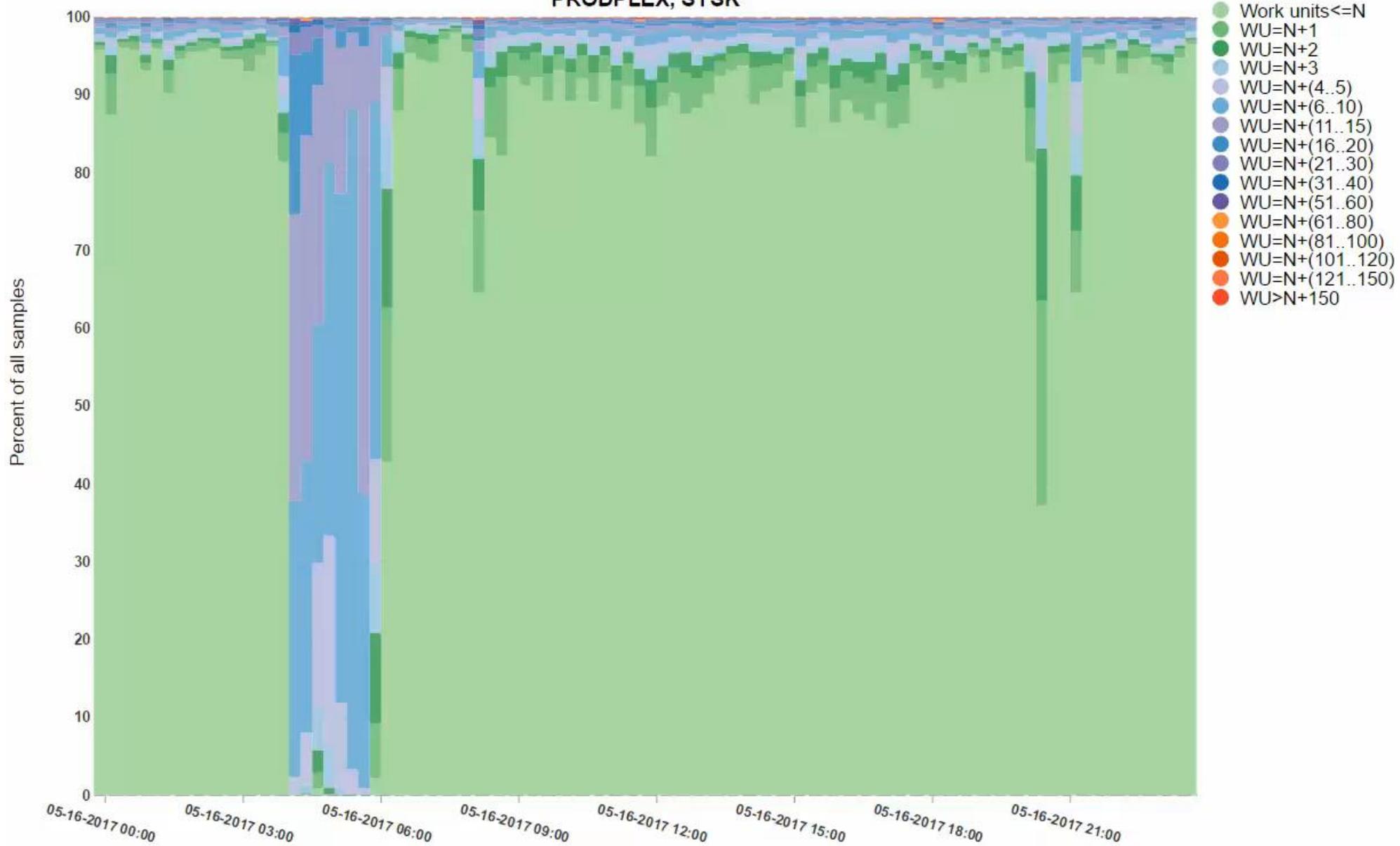
## CP Work Units - Min, Avg, Max



# CPU Work Unit Distribution

(N = Average of Unparked Engines Regardless of Engine Type)

PRODPLEX, SYSK

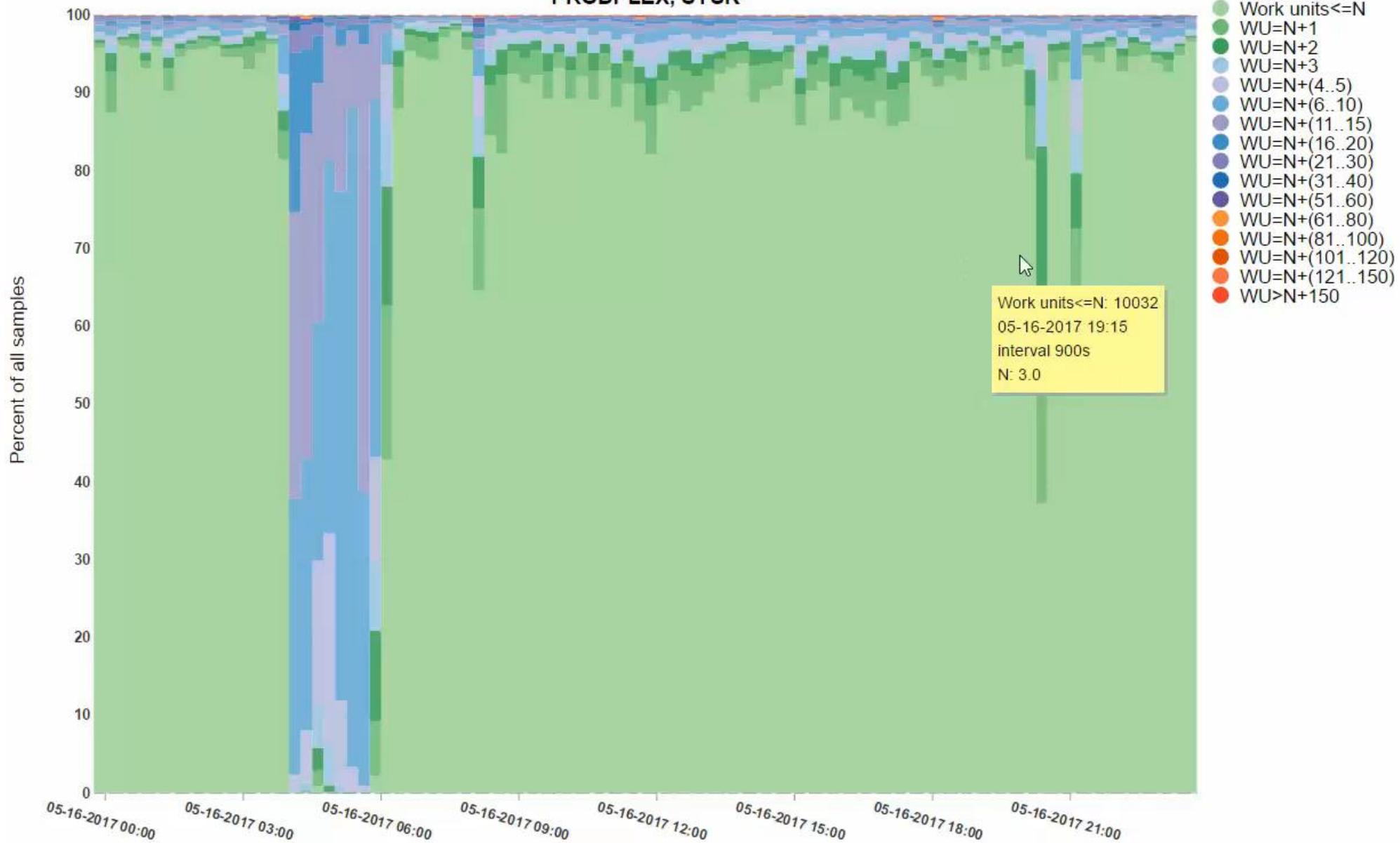


# CPU Work Unit Distribution

(N = Average of Unparked Engines Regardless of Engine Type)



PRODPLEX, SYSK



# Highest task CPU percent

---

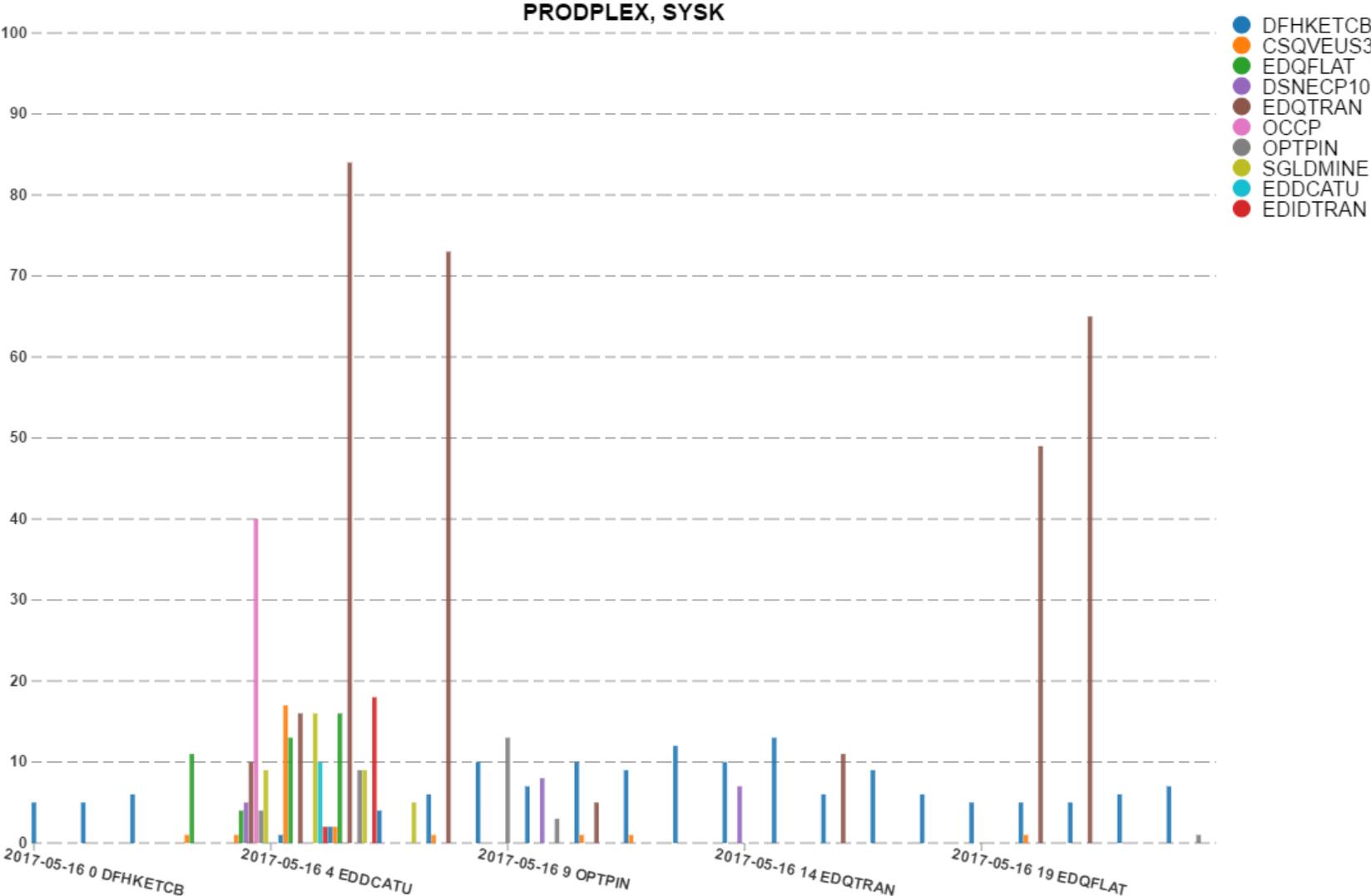


- “New” field in SMF30: SMF30\_Highest\_Task\_CPU\_Percent
  - For interval records: largest percentage of CPU time used by any task in the address space =  $\text{round}(\text{TCB} / \text{interval} * 100)$
  - For step/job end records: largest reported interval value
- Related field: SMF30\_Highest\_Task\_CPU\_Program
  - Name of the program loaded by the task that had the largest percentage
- As value approaches 100, per-CPU speed becomes more important
  - Threshold for worry: “it depends”
- We don’t know anything about the second-most busy task
- “Spikey” TCBs might be under-represented
  - Larger intervals hide larger spikes

# Highest task CPU – high level overview



## Highest Task CPU Percent Programs by Hour



# CICS: QR TCB time

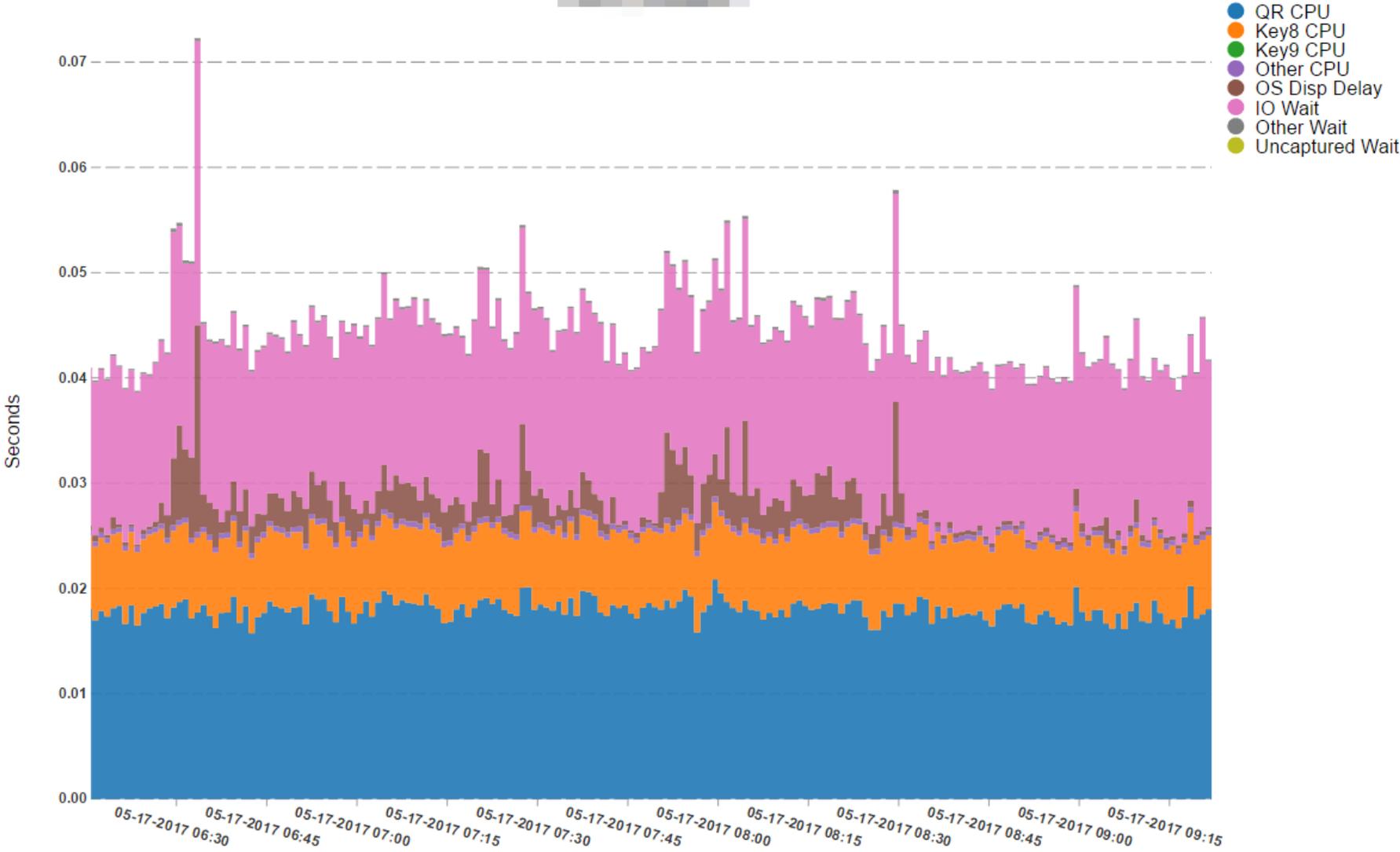


- Each CICS region has multiple TCBs, but only one QR TCB
  - Single TCB = scalability bottleneck
- Historically all application code ran on the QR TCB
- Today: many more TCBs
  - Prime example: SQL and MQ commands run on L8 TCB
  - L9: User Key OpenAPI programs
    - But use CICS Key for programs doing DB2/MQ
  - L8/L9 TCB pool limit = ~~(2 \* max task) + 32~~ something more reasonable
- CICS SMF data will break CPU consumption down by TCB
  - Will also show things like the amount of dispatch delay and number of dispatches

# CICS Response Time Breakdown



## CICS Average Response Time Components (Excluding system transactions)



# Specialty engine crossover

---



- SMF 30 and SMF 72 record the amount of CPU time consumed on GPs that could have been done on zIIPs (or zAAPs)
- Mostly an indication of queueing for specialty engines
- Can be mostly stopped by setting HONORPRIORITY=NO for the service class
  - See OA50845
  - May not want to do that for very important work (e.g. production DB2 system tasks)
- In certain specific situations ZIIPAWMT (in IEAOPTxx) may be helpful
- Slower GPs mean the execution time will be longer
  - Maybe should just wait for a zIIP to become available?
    - ZIIPAWMT can tune that, but probably rarely needed
- If you have significant failover happening, ideally just buy another zIIP
  - Or turn on SMT on z13 and above
  - Note SMT or not is also a more/slower vs. fewer/faster question



# Sysplex overhead

---



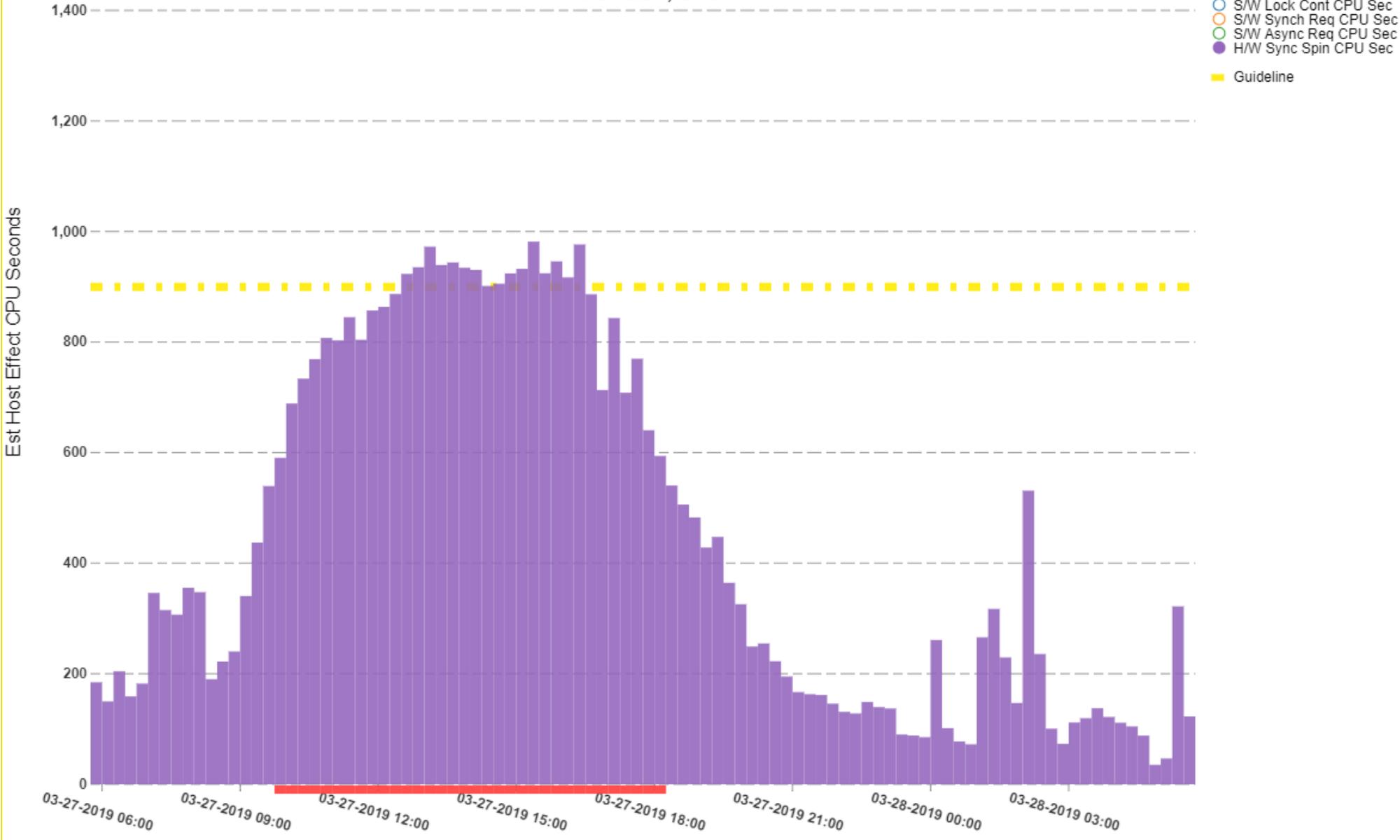
- While sync requests are executing, the requesting CPU spins waiting for a response
- Faster engine with same CF response time = more lost capacity
- More / slower CPs means a smaller portion of your capacity
- Keep CF technology matched as best as possible



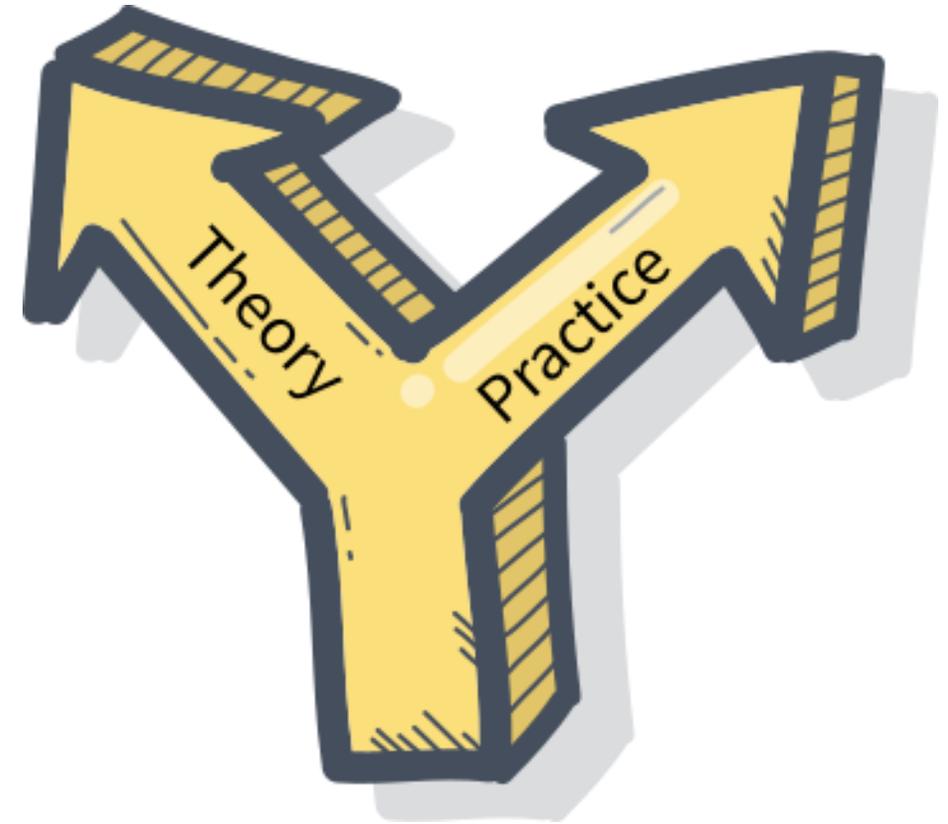
# CF Host Effect - Estimated Host Seconds Breakdown for System

Software times scaled from 2094-701 estimates

PRODPLEX, SYSA



# Practical Considerations - Summary

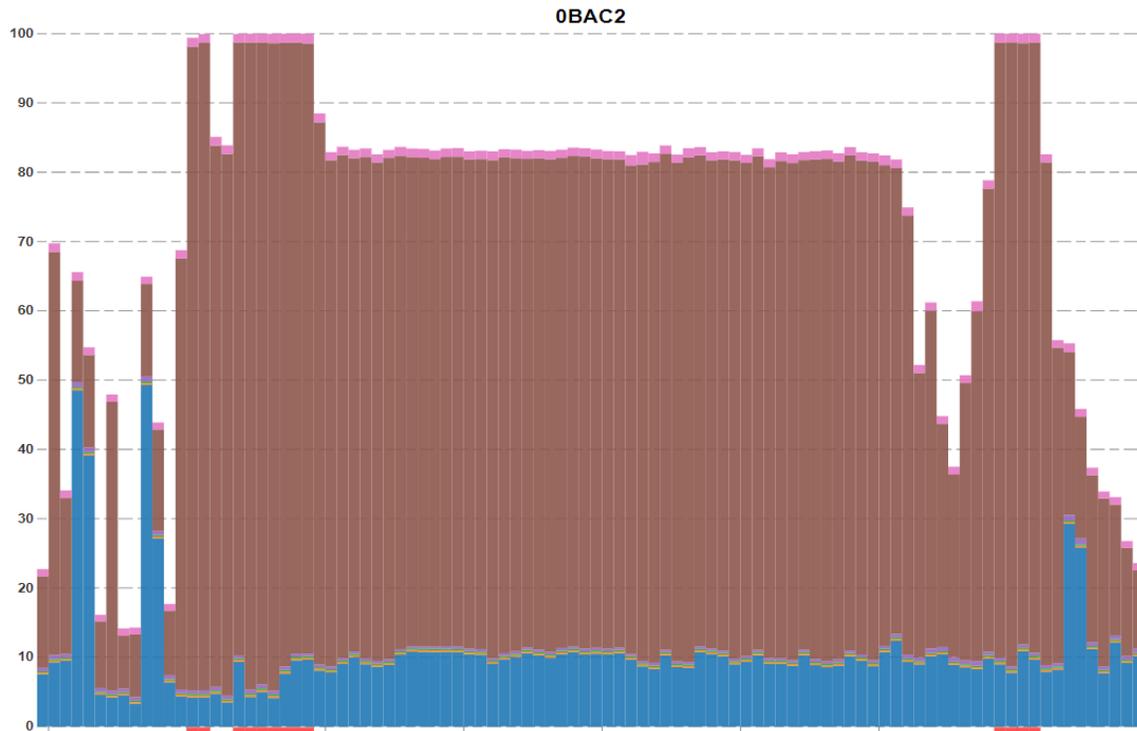


# LPAR Count

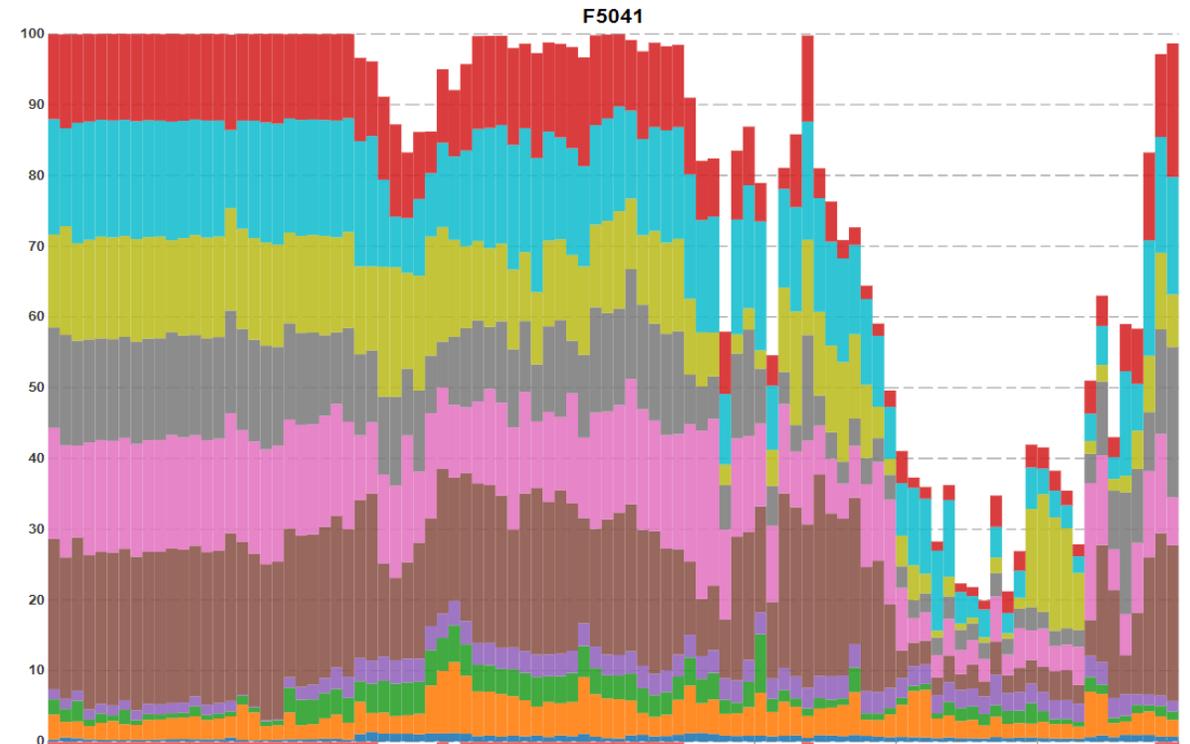


- As number of LPARs per CEC increases, be more wary of having a small number of CPs to dispatch across
- Particularly true when there are more “significant” LPAR

CEC Physical Machine CP Busy% by CEC Serial Number



CEC Physical Machine CP Busy% by CEC Serial Number



- Find candidates that will be impacted
  - Consider filtering by batch jobs that consume more than trivial CPU and/or run more than a trivial amount of time
  - Consider filtering by CPU intensity
  - Consider batch jobs that are CPU-bound when the system is not CPU-constrained
- Understand how much headroom you have in your batch SLAs
  - E.G. if you're meeting your window by 3 hours today, some delay may be fine
  - **Make sure the SLAs are representative of real business need!**
- Do you care about low-importance batch?
  - Maybe: dev/test work needs to get done too!
- **Watch out for S322s, possibly adjust limits before changing**

# Started tasks

---



- Find candidates that will be impacted
  - SMF30\_Highest\_Task\_CPU\_Percent is a good place to start
  - But don't forget to look for other started tasks with a high CPU intensity
- If moving to fewer/faster beware high-priority workloads' ability to monopolize CPUs

- Often the CPU time & CPU delay per transaction is so small as to be unnoticeable to the end user
  - Most transactions are using milliseconds of CPU—milliseconds + milliseconds of wait may not be enough to matter
  - But beware applications which trigger multiple transactions per user interaction
- How busy are the QR TCBs in each region?
  - Note that this is a single server queueing model:  $\sim 70\%$  busy = wait = 2x service time, ie. response = 3x service time
  - As QR TCB busy exceeds 70% and approaches 100%, increased caution warranted
- If at all possible, make applications threadsafe



# zIIP-eligible workloads

---

- Slower engines differ more from full-speed zIIPs
- Consider whether you want to allow failover based on
  - How often failover occurs
  - Whether it occurs during your R4HA peaks
  - How important the work that fails over is
  - Discrepancy between GPs and zIIPs

# Can you use OOCoD to help?

---



- On/Off Capacity on Demand is great for trying different combinations
  - “Delivered” capacity can be less than “purchased” capacity
  - Changes within purchased capacity can be done for no additional hardware charge
- But there are rules that you have to be aware of:
  - There are various agreements to sign
  - You can not decrease the number of physical CPs to less than what was delivered
  - You can not decrease the capacity to less than what was delivered
  - You can not go beyond what’s physically installed or 2x purchased capacity
  - Pre-paid maintenance may be problematic
    - Pre-pay based on purchased vs. delivered capacity?
    - Try during warranty, lock in before maintenance kicks in?

# OOCoD Matrices (Upgrade Scenario B)



Which of these options will work best?

Speed	Number of CPUs										
	1	2	3	4	5	6	7	8	9	10	11
4xx	250	478	697	910	1,118	1,321	1,520	1,716	1,907	2,093	2,277
5xx	746	1,417	2,067	2,694	3,306	3,903	4,485	5,052	5,606	6,145	6,671
6xx	1,068	2,019	2,938	3,827	4,690	5,528	6,342	7,132	7,899	8,644	9,368
7xx	1,695	3,196	4,644	6,041	7,392	8,700	9,964	11,188	12,371	13,515	14,622

Bringing in the machine as a 410 precludes using OOCoD to try the other options

Speed	Number of CPUs										
	1	2	3	4	5	6	7	8	9	10	11
4xx	250	478	697	910	1,118	1,321	1,520	1,716	1,907	2093	2277
5xx	746	1417	2067	2694	3,306	3,903	4,485	5,052	5,606	6,145	6,671
6xx	1,068	2019	2938	3827	4,690	5,528	6,342	7,132	7,899	8,644	9,368
7xx	1,695	3,196	4,644	6,041	7,392	8,700	9,964	11,188	12,371	13,515	14,622

Blue = delivered capacity

Green = OOCoD possibilities

Starting at a 503 is better

Speed	Number of CPUs										
	1	2	3	4	5	6	7	8	9	10	11
4xx	250	478	697	910	1,118	1,321	1,520	1,716	1,907	2093	2277
5xx	746	1417	2067	2694	3,306	3,903	4,485	5,052	5,606	6,145	6,671
6xx	1,068	2019	2938	3827	4,690	5,528	6,342	7,132	7,899	8,644	9,368
7xx	1,695	3,196	4,644	6,041	7,392	8,700	9,964	11,188	12,371	13,515	14,622

But we really want to start as a 602 (or 502) even though we don't expect to use that

Speed	Number of CPUs										
	1	2	3	4	5	6	7	8	9	10	11
4xx	250	478	697	910	1,118	1,321	1,520	1,716	1,907	2093	2277
5xx	746	1417	2067	2694	3,306	3,903	4,485	5,052	5,606	6,145	6,671
6xx	1,068	2019	2938	3827	4,690	5,528	6,342	7,132	7,899	8,644	9,368
7xx	1,695	3,196	4,644	6,041	7,392	8,700	9,964	11,188	12,371	13,515	14,622

# Summary

---



- Choice of engine speed can affect system efficiency
- Engine speed choice can possibly affect real capacity / MSU
- Slower engines may be a better choice
- Many measurements to review to help you decide
- Examine your workloads