zIIP stealing GCP MSUs time for Capacity Management

Donald Zeunert

Tuesday 7th November, 2017 (10:45 – 11:45)
Session LI in Woodcote

Donald_Zeunert@BMC.com

Topics

- **Elimination of zAAP good or bad?**
- **Why should we care ? –**
- **Who are zIIP exploiters? –**
- **How to get more zIIP capacity -**
- **How to control overflow to GCP -**
- **How to reduce zIIP usage -**

Eliminate zAAP Good?

Elimination of zAAP

Pros

- CP utilization - Improved
- MP effect - Reduced
- Queueing time – Reduced w/ more eligible CPs
- MLC save w/ less hardware
  - Small zIIP & zAAP workload could justify an engine

Cons

- Application JAVA code contending with system background tasks
- Eliminated separate overflow controls
  - Workloads impacted by delays were not controllable

zAAP Demise

- zAAP on zIIP
  - z/OS V1.11
    - ZAAPZIIP=YES
  - z/OS V1.12
    - APAR to allow zAAP to run on zIIP when zAAP present

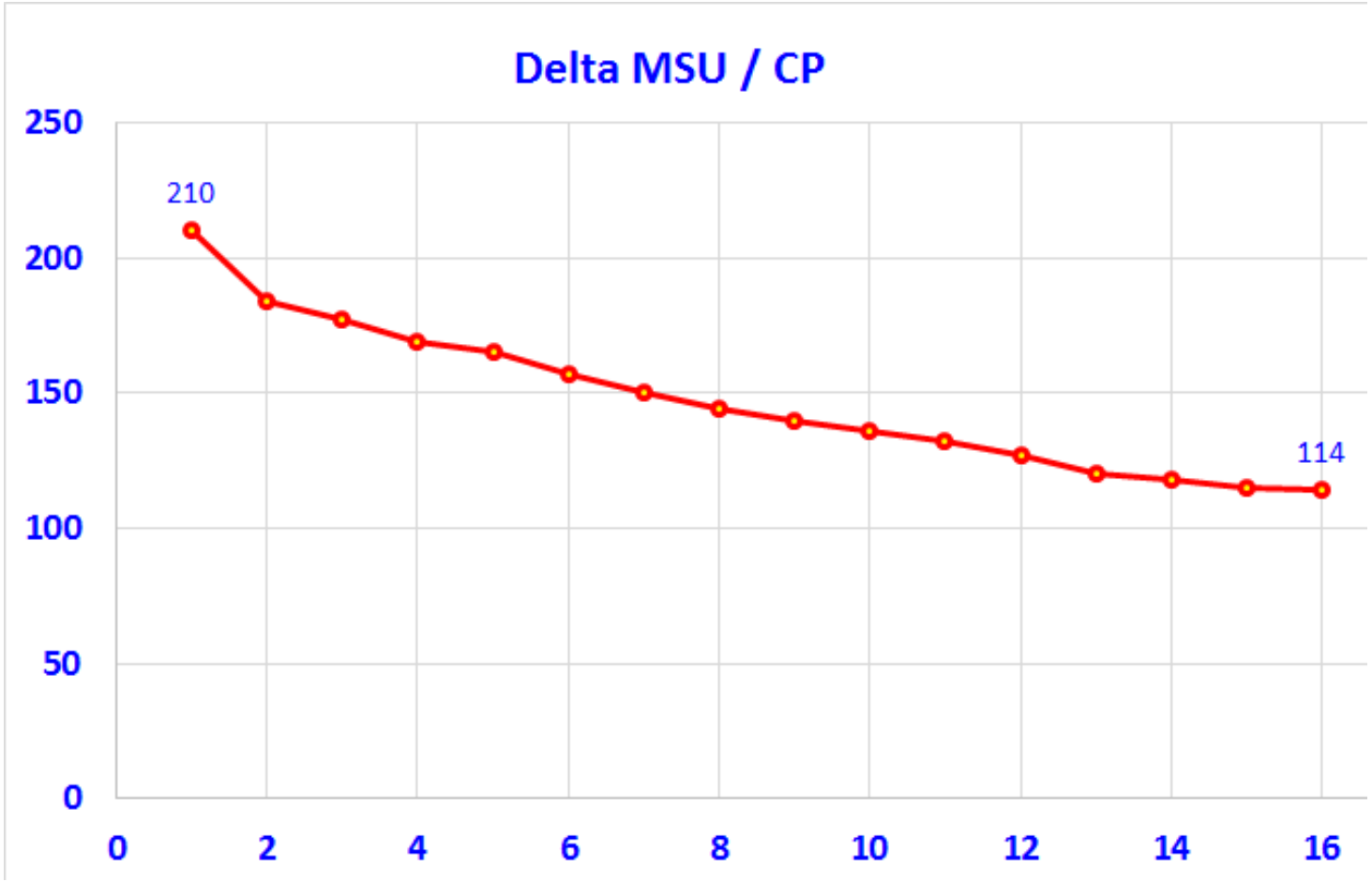
zEnterprise 196 & 114

- zIIP to CP ratio <= 1:1

zEC12 and zBC12

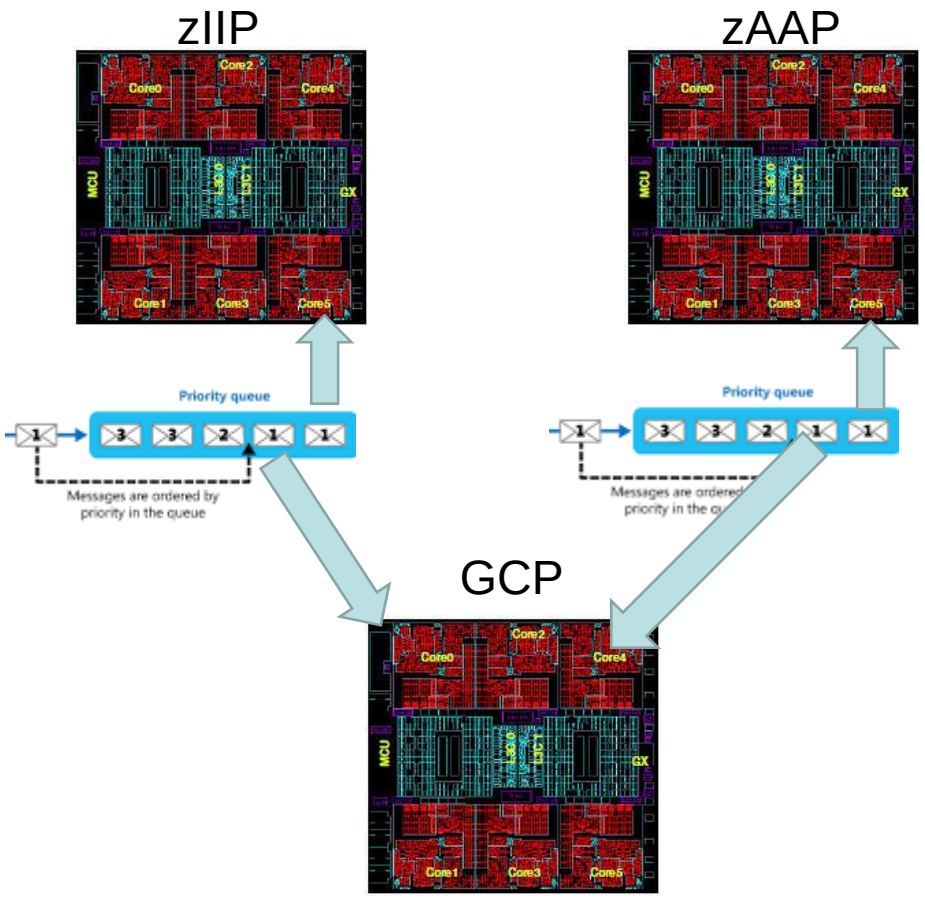
- zIIP to CP ratio <= 2:1
- Last CPCs w/ zAAPs

Multiprocessor (MP) effect



- Original underutilized
  - 8 zIIP and
  - 8 zAAP
- If fewer CPs then lower MP effect and better utilization / CP

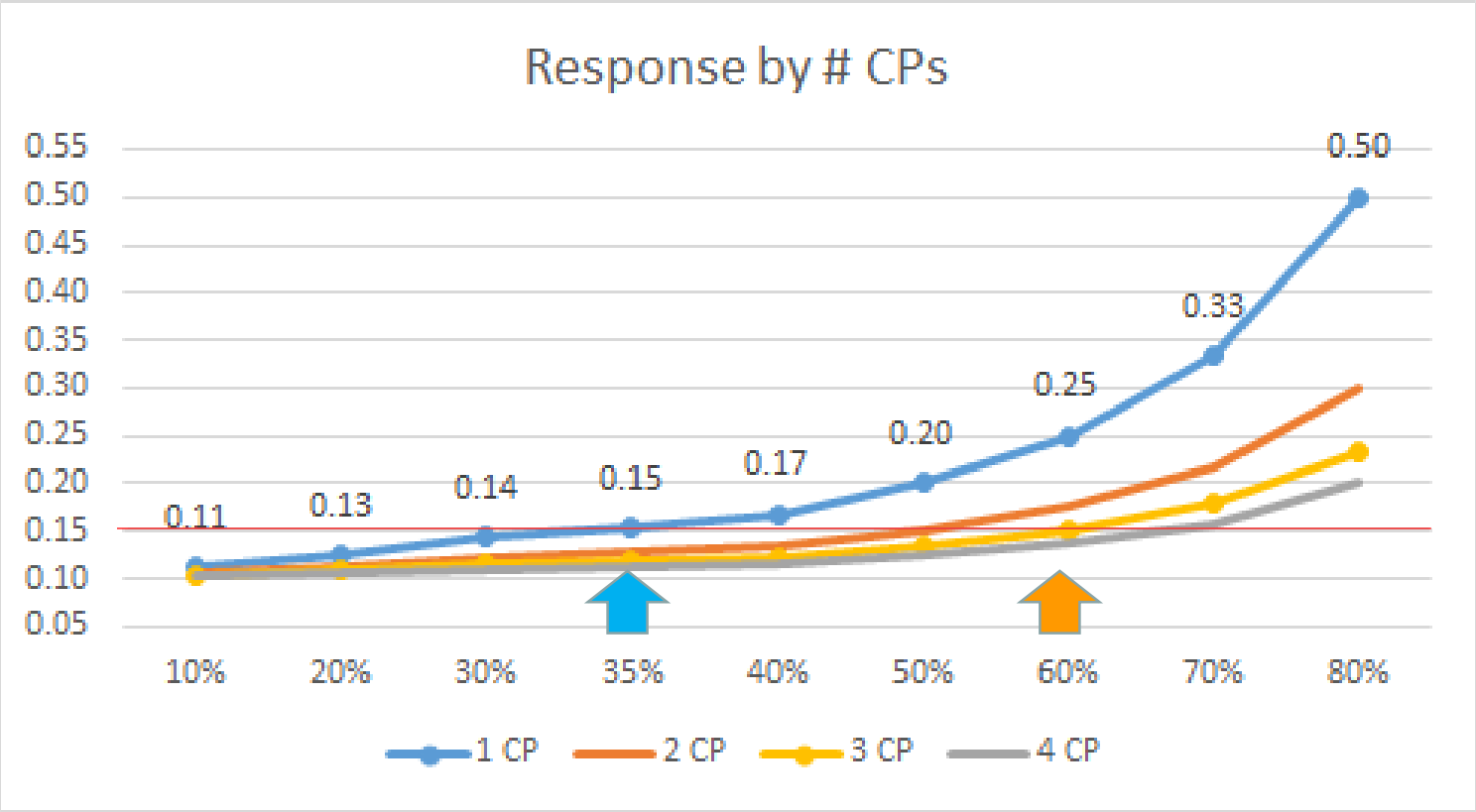
Why / how overflow



Defaults allow specialty engine to overflow to GCP

- Spare specialty of other type wasted
- Controls were LPAR wide
- zAAP had more controls
- Help signaled if all busy of 1 specialty type

zIIP % Busy and Crossover



- 1 Specialty engine @ 35%
- 2 Specialty CP @ 50-60%
- Only 1 type = less likely all busy to ask for Help

Why Should I Care?

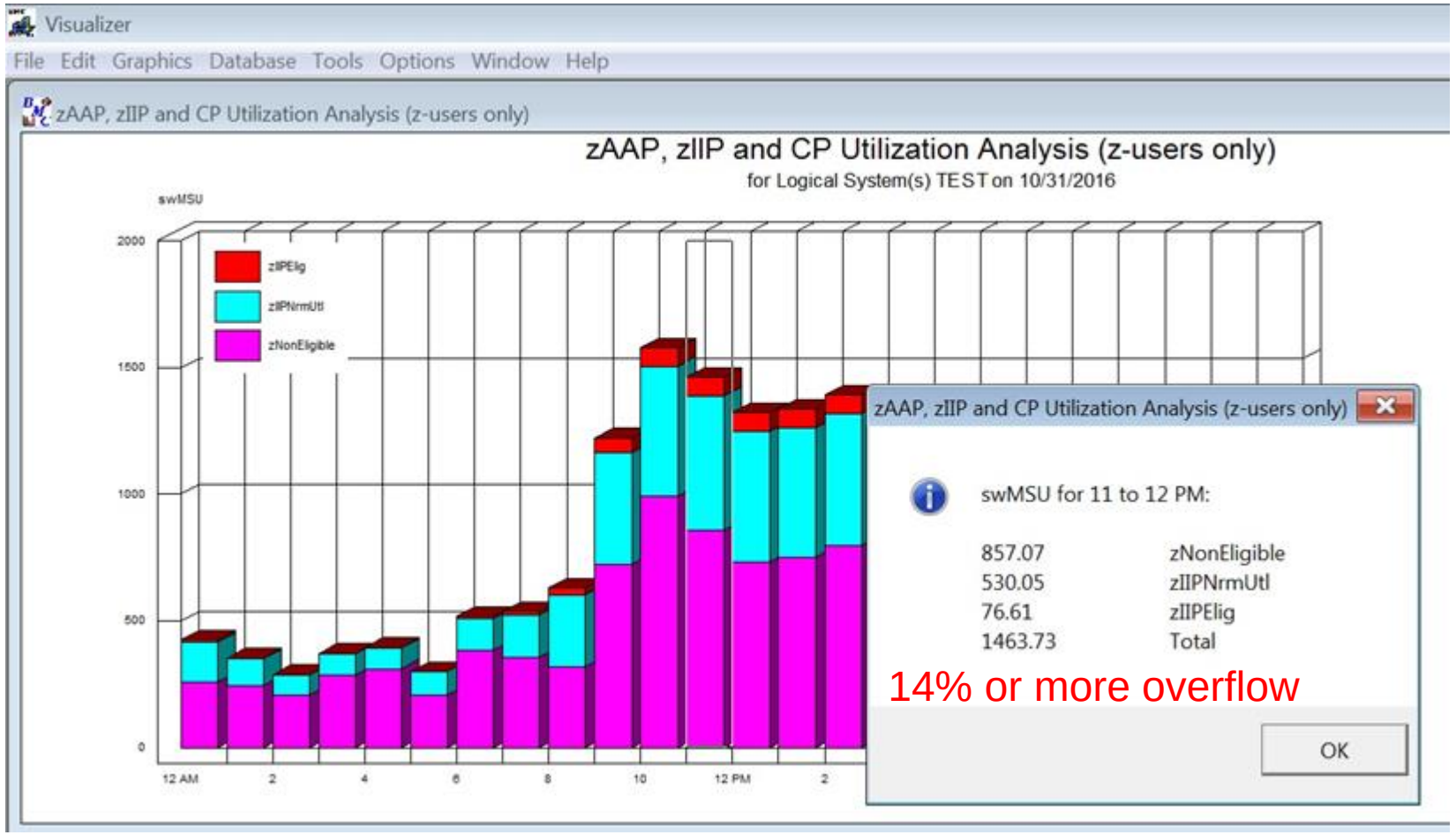
Overloaded zIIPs cost GCP MSUs

- zIIPs overloaded w/o zAAP work
- Large MSUs overflowing to GCPs
  - Slowing down workloads
  - Costing hardware and software $$$
- More zIIP exploiters on horizon

IMPACTING SLAs & Costs

Overflow MSU to GCPs = MLC $

MLC Impact : $250K – $750K
per year …per engine wasted



- 77 MSUs @ $400 / MSU = $372K / yr
- Smaller sites pay > $/ MSU larger sites waste more MSUs

Who are zIIP Exploiters?

Who is using my zIIP

Traditional zIIP Users

- DB2 v8+ & IMS v12+ (more to come)
- Batch (SyncSort)
- XML Sys Service, ComSrvr IPSec
- Global Mirror DFSMS Data Mover
- Performance monitors

zAAP on zIIP Users

- Batch JAVA (JZOS)
- Transactional JAVA
  - WAS, CICS Liberty, IMS JMP, Spark, etc.

Rock 'Em Sock 'Em Robots ©Mattel

DB2 zIIP Eligible

- **DB2 buffer pool processing** - up to 100% for
  - prefetch, Deferred write, Page set castout (CF Read GBP),
- **DB2 Utilities** – up to 100% of
  - LOAD, REORG, and REBUILD INDEX
- **Log processing** - read, and write
- **DRDA over IP workloads** – SQL request up to 60%
- **Parallel query** - child processes up to 80%
- **XML Schema processing** - up to 100% of
  - schema validation and non-validation parsing

Note: IIPHONORPRIORITY = NO, DB2 prevents system tasks becoming zIIP eligible

IMS zIIP Eligible

IMS V12 – CQS access ShrMsgQs

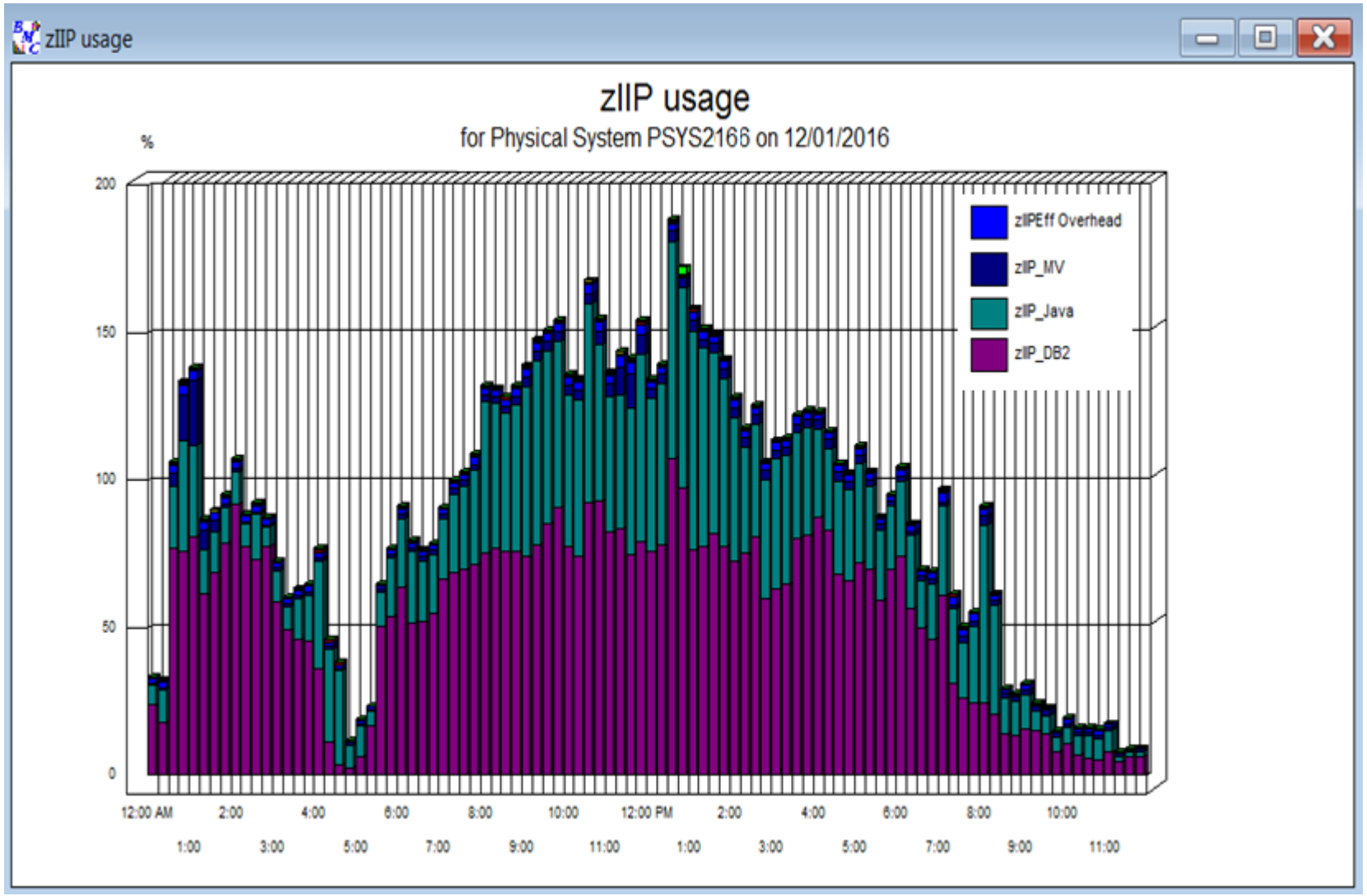
IMS V13

- **IMS Connect** thread requests via TCP/IP for;
  - SOAP messages
  - DRDA®
  - MSC
  - ISC
- **ODBM address space**
  - DRDA® threads requests via TCP/IP
  - Threads requests via CSLDMI API.

zAAP on zIIP

- Java Message Processing (JMP) regions
- Java Batch Processing (JBP) regions
- Message Processing Programs (MPP) w/ COBOL to JAVA

Harder to tell what is using zIIP?



- 55% to DB2
  - Prefetch
  - Utilities
- 35% to Java
  - Application
  - Garbage Collection
- 2% to Monitors
- 3% to Overhead
- 5% spare?

How to get more zIIP capacity?

Preventing Overflow – Increase Capacity

- Buy more zIIP or can't due to
  - GCP Ratio restrictions
  - Hardware $s
- Other Options
  - Sub-capacity CPCs zIIP MSUs to GCP ratio better
  - Simultaneous Multithreading (*SMT*) on z13+ family

Subcap – zIIP Relative Speed

z13	Model	MSUs	2 zIIP	Eff Ratio
2964	401	31	420	13.55
2964	501	94	420	4.47
2964	601	134	420	3.13
2964	701	210	420	2

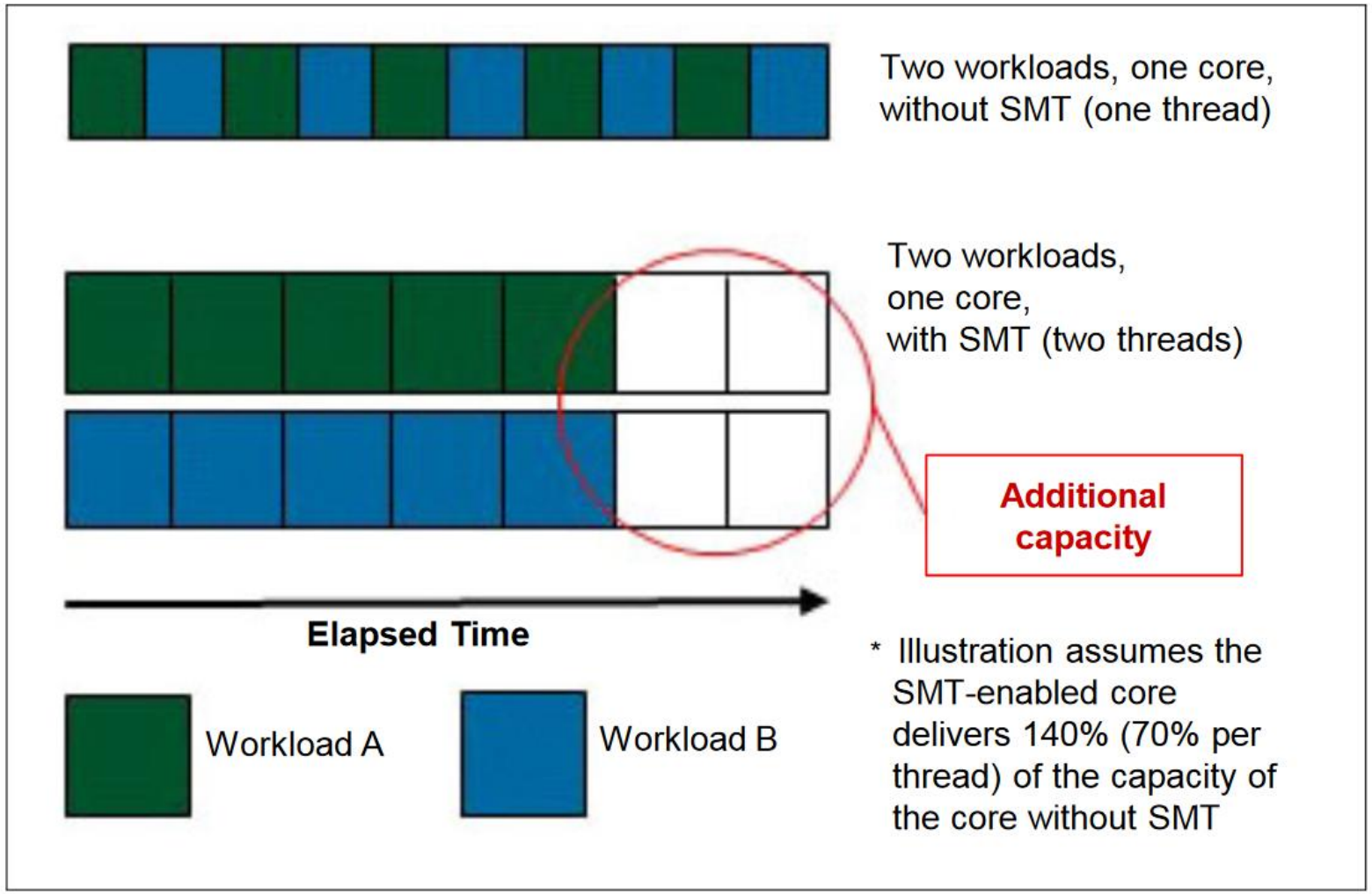
- zIIP, zAAP run at full 701 speed on all models
- 2 zIIPs / GCP can be 13.5x more MSUs instead of 2x on 7xx models

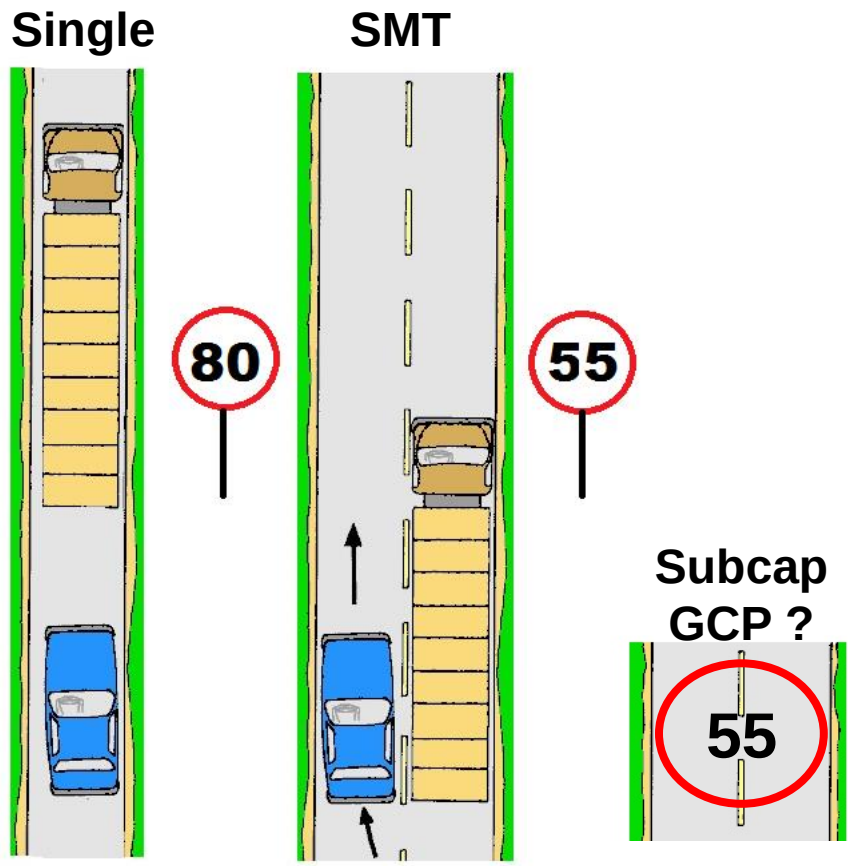
Cost of Subcapacity GCP crossover

- Is cross-over occurring in 4HRA CPC peak?
- Higher for Sub-capacity model w/ GCPs slower than zIIPs
  - 10% overflow of one zIIP = 21 MSUs or 68% of 1w 401 GCP
    - 87% of 1 CP on 2964-416
  - Normalized zIIP Seconds in Monitors & SMF Recs
    - Address space(30), Processor(70), SrvCls(72)

z13	Model	MSUs	1 zIIP	Eff Ratio
2964	401	31	210	6.77

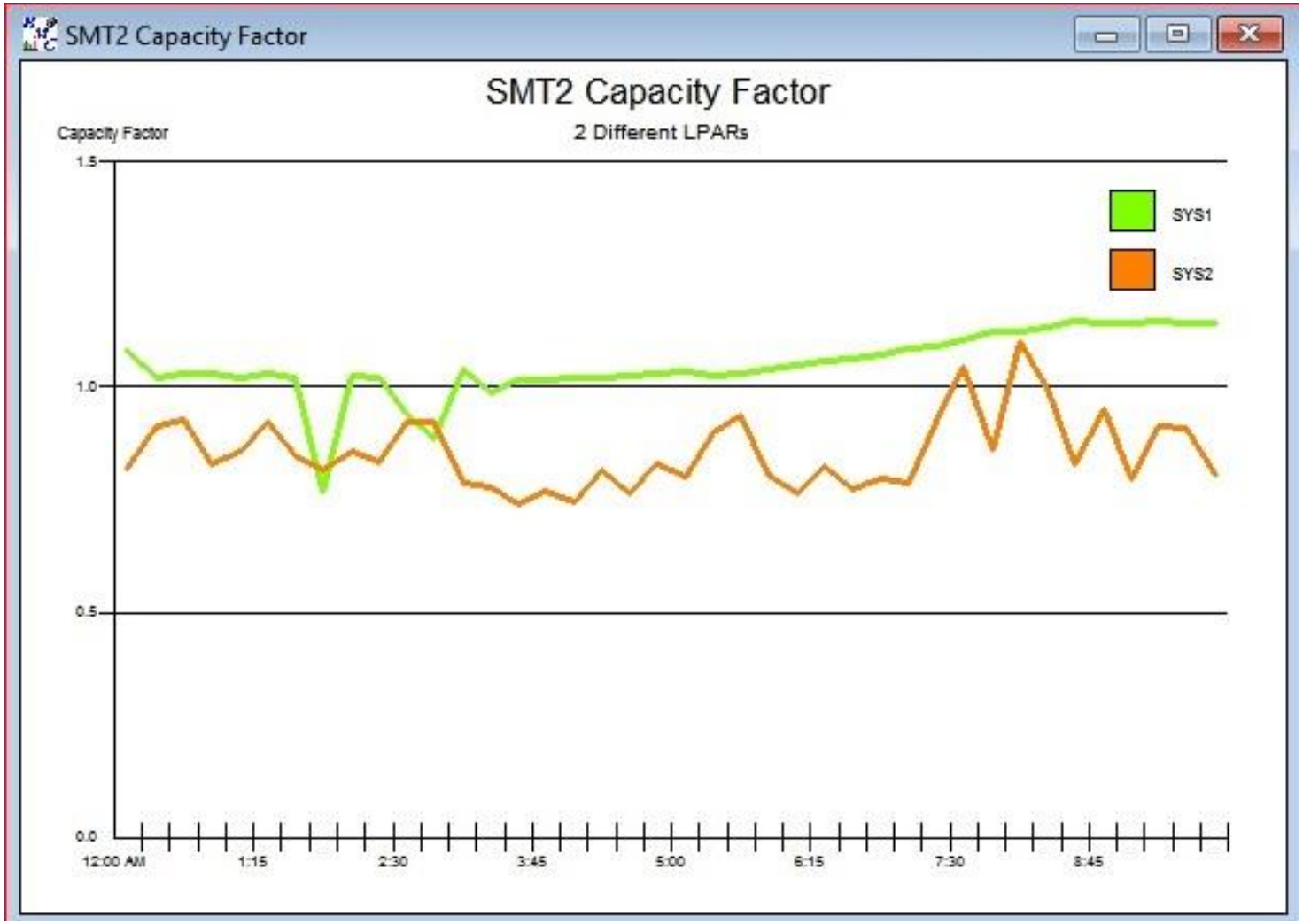
z13 Simultaneous MultiThreading (*SMT*)





SMT - Slow to pass

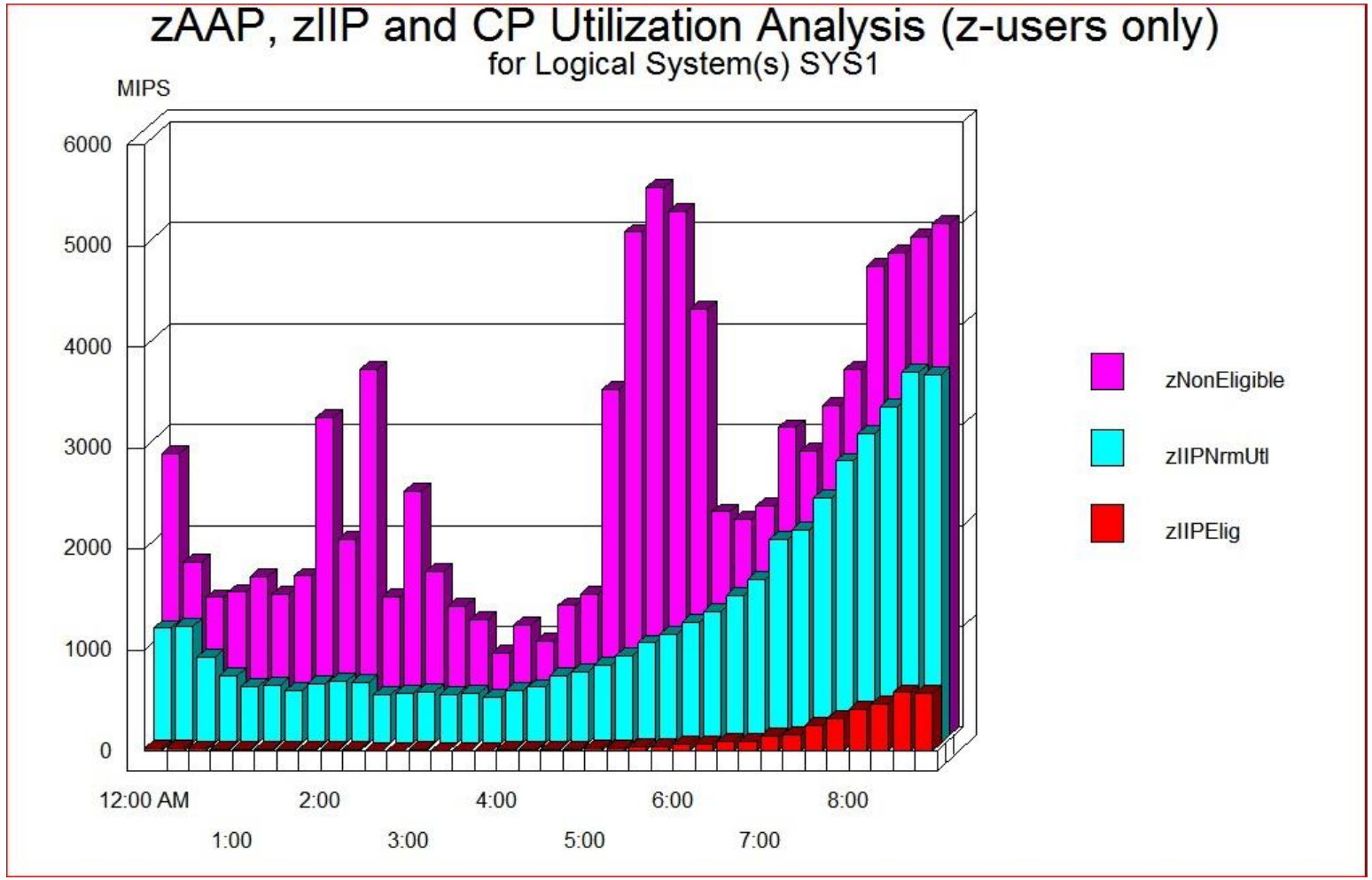
SMT Effectiveness



Capacity Factor

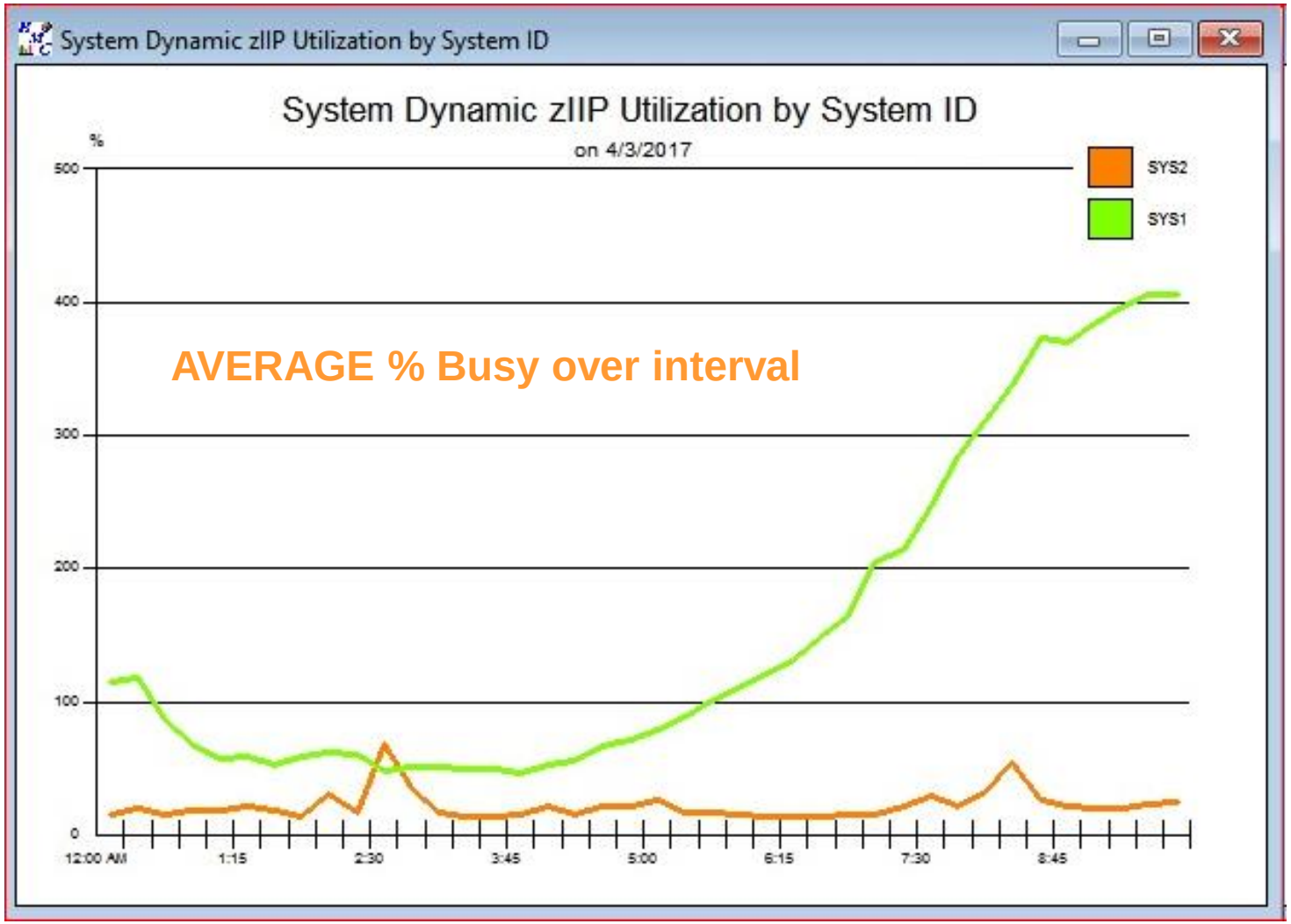
- Any CF # > 1 means SMT is delivering additional zIIP capacity

zIIP Overflow



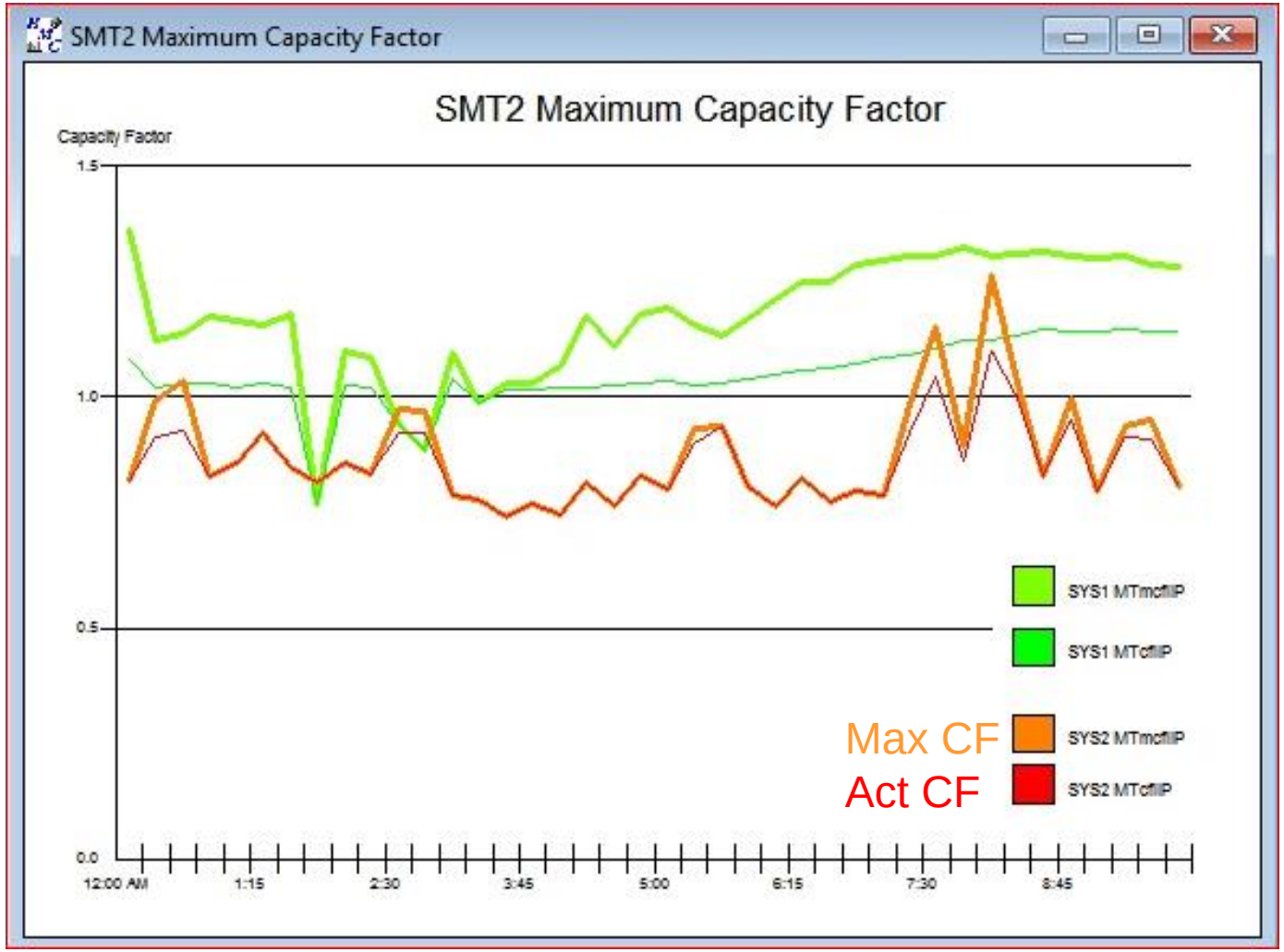
- Overflow still occurring, but would have been more MSUs

Overflow w/o > 65+%



- 44% Busy
  - 4 of 9 zIIP busy
- Queueing theory
  - predicts elapsed time elongation based on # of servers / zIIPs
  - > 35% for 1 is begin of CP queueing
  - Does not predict probability of all busy for 1 microsec

SMT2 – not at Max Capacity



- Just like SMT1 mode most work does not wait for zIIP if all zIIPs busy at the moment in time

SMT – Other benefits

- zIIP initiated CF SYNC calls w/ SMT do not spin both threads
  - Hiperlinks
  - DB2 Utility page locks

How to control overflow to GCP?

Specialty CP on GCP Controls

IEAOPTxx Options

Crossover	Honor Priority	General Purpose Processors Behavior
NO	YES	zIIP / zAAP work on GCPs only when help needed
YES	NO	zAAP work on GCPs only when no non-zAAP work is ready (1)
YES	YES	zAAP work on GCPs when help needed and no non-zAAP work ready
NO	NO	No zIIP / zAAP work on GCPs (1)

1) HonorPrio=NO; Specialty engine eligible only runs on GCPs to resolve resource conflicts w/ GCP work

Help needed is signaled when all specialty CPs of a type are busy. Less likely if all 1 type.

zAAP

- IFACROSSOVER
- IFAHONORPRIORITY
  - Parm z/OS 1.6 (2005)

zIIP

- zIIP Crossover = NO
  - GCPs help if zIIP asks
- IIPHONORPRIORITY =YES
  - Parm since z/OS 1.8 (2007)

ZAAPAWMT & ZIIPAWMT

- **Need for help** - checking frequency
  - HIPERDISPATCH=NO
    - 12 ms default
  - HIPERDISPATCH=YES
    - 3.2 ms default
    - Automatically increased likelihood to request Help
- **Increasing the value(s)**
  - Help requested after all zAAPs busy for a longer time
  - Delays GCPs helping
    - Reduces 4HRA impact
  - Reduces overhead of waking up to check for work when low utilization
    - Specialty work waits longer

Workload level GCP overflow controls

NEW 2017

```
Create a Service Class

Service Class Name . . . . . JZOSLOW
Description  . . . . . ____________________________
Workload Name  . . . . . CB (name or ?)
Base Resource Group  . . . . . ________ (name or ?)
Cpu Critical . . . . . NO (YES or NO)
I/O Priority Group . . . . . NORMAL (NORMAL or HIGH)
Honor Priority . . . . . NO (DEFAULT or NO)

Specify BASE GOAL information. Action Codes:
I=Insert new period, E=Edit period, D=Delete period.
```

Prevent specialty overflow to GCPs by

- WLM Service class
  - Periods not supported

APAR **OA50760** – March 29th, 2017

Component	PTF
R7AJ PSY	UA91371
R7A0 PSY	UA91369
R79J PSY	UA91372
R790 PSY	UA91370

Hiperdispatch – zIIP Parking

C P U A C T I V I T Y

z/OS V1R11 SYSTEM ID Z2 START 09/11/2009-02.30.00
RPT VERSION V1R11 RMF END 09/11/2009-03.00.00

CPU 2097 MODEL 737 H/W MODEL E56 SEQUENCE CODE 0000000000699FF HIPERDISPATCH=YES

---CPU---		TIME %				LOG PROC		--I/O INTERRUPTS--	
NUM	TYPE	ONLINE	LPAR BUSY	MVS BUSY	PARKED	SHARE %		RATE	% VIA TPI
0	CP	100.00	96.60	96.74	0.00	100.0	HIGH	1593	2.64
1	CP	100.00	97.51	97.69	0.00	100.0	HIGH	1607	2.73
2	CP	100.00	96.02	96.23	0.00	96.0	MED	5.12	29.30
3	CP	100.00	39.26	80.81	51.23	0.0	LOW	0.00	0.00
4	CP	100.00	48.71	79.90	38.77	0.0	LOW	0.00	0.00
5	CP	100.00	41.06	79.34	48.01	0.0	LOW	0.00	0.00
6	CP	100.00	12.42	78.35	84.11	0.0	LOW	0.00	0.00
7	CP	100.00	0.00	-----	100.00	0.0	LOW	0.00	0.00
8	CP	100.00	0.00	-----	100.00	0.0	LOW	0.00	0.00
9	CP	100.00	33.05	80.34	58.68	0.0	LOW	199.6	1.01
TOTAL/AVERAGE			46.46	89.73		296.0		3405	2.62
A	IIP	100.00	57.35	88.68	0.00	32.0	MED		
B	IIP	100.00	46.71	92.85	17.56	0.0	LOW		
C	IIP	100.00	45.27	90.82	17.79	0.0	LOW		
D	IIP	100.00	52.81	85.00	0.00	0.0	LOW		
TOTAL/AVERAGE			50.78	89.09		32.0			

- Overflow when online zIIPs are busy
- Unpark > 80%
- Park < 66%
- Focus on VH/VM avoid wasted VLs

How to reduce zIIP usage?

How to tune to reduce zIIP usage

- Overflow Reduction
  - Disable those that can wait
  - Make others wait longer

- Tune consumers to reduce overflow
  - JAVA (GC, Application tuning, etc.),
  - DB2 sequential pre-fetch waste

Prevent Overflow ?

Consider

- JAVA batch = wait
- Inefficient zIIP users –
  - Disable in product?
- DB2 Test ? = No zIIP

Avoid

- DB2 Production as never use
- Monitors –
  - if can't disable in product
  - May cause delay / inaccuracies in sampling

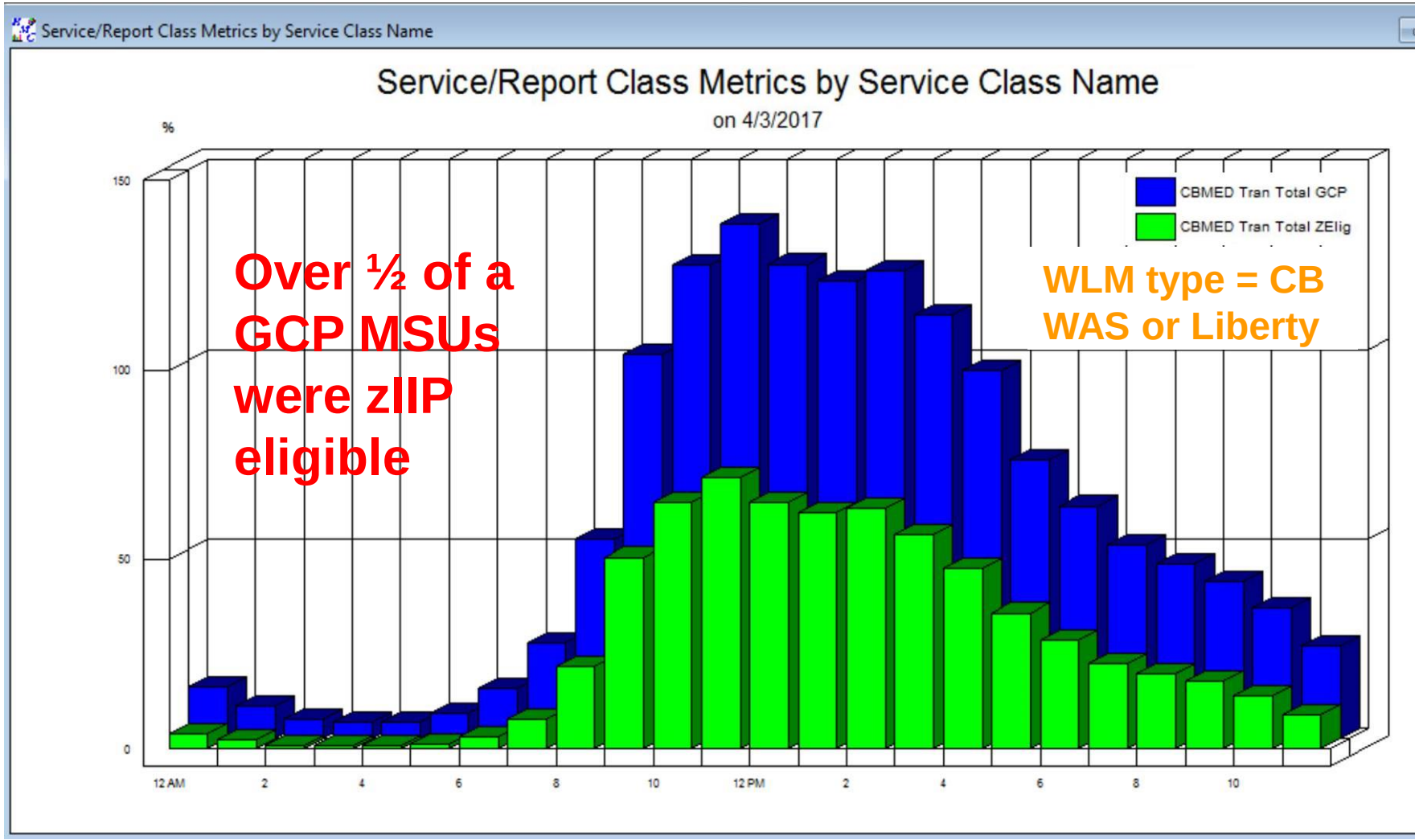
DB2 Sequential Prefetch tuning

- Sequential prefetch - Table scans and table space scans
  - Change DB IDX or SQL to reduce # scans
  - Prefetch in DB2 STC so can't control w/ overflow w/ WLM
    - w/o impacting everyone (none zIIP eligible)
  - Reduce prefetch duration & reduce zIIP busy when # < #DB2+1
    - Move table w/ large scan / prefetch to smaller buffer pool
      - Reduces prefetch size and associated zIIP or overflow CPU
    - Reduce % of buffer pool allocated to sequential reads
      - Default 80%
      - VPSEQT = 0 disables pre-fetch

DB2 Utilities

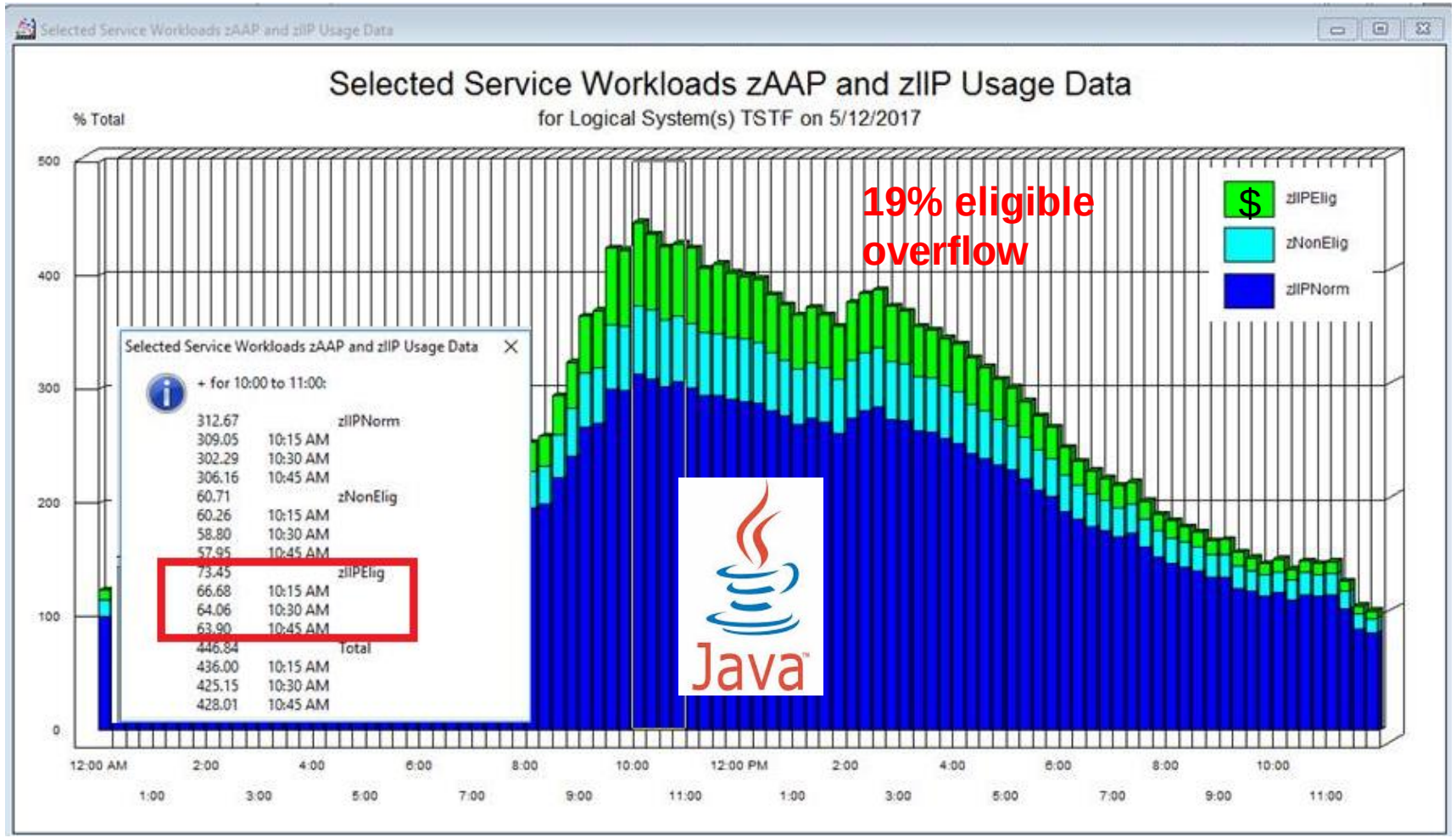
- Run outside of 4HRA or demand peaks
- Reorg fewer databases based on need, not schedule
- Use more efficient DB2 Utilities
- Use reorg while open outside of 4HRA peak

not Free if eligible on GCP



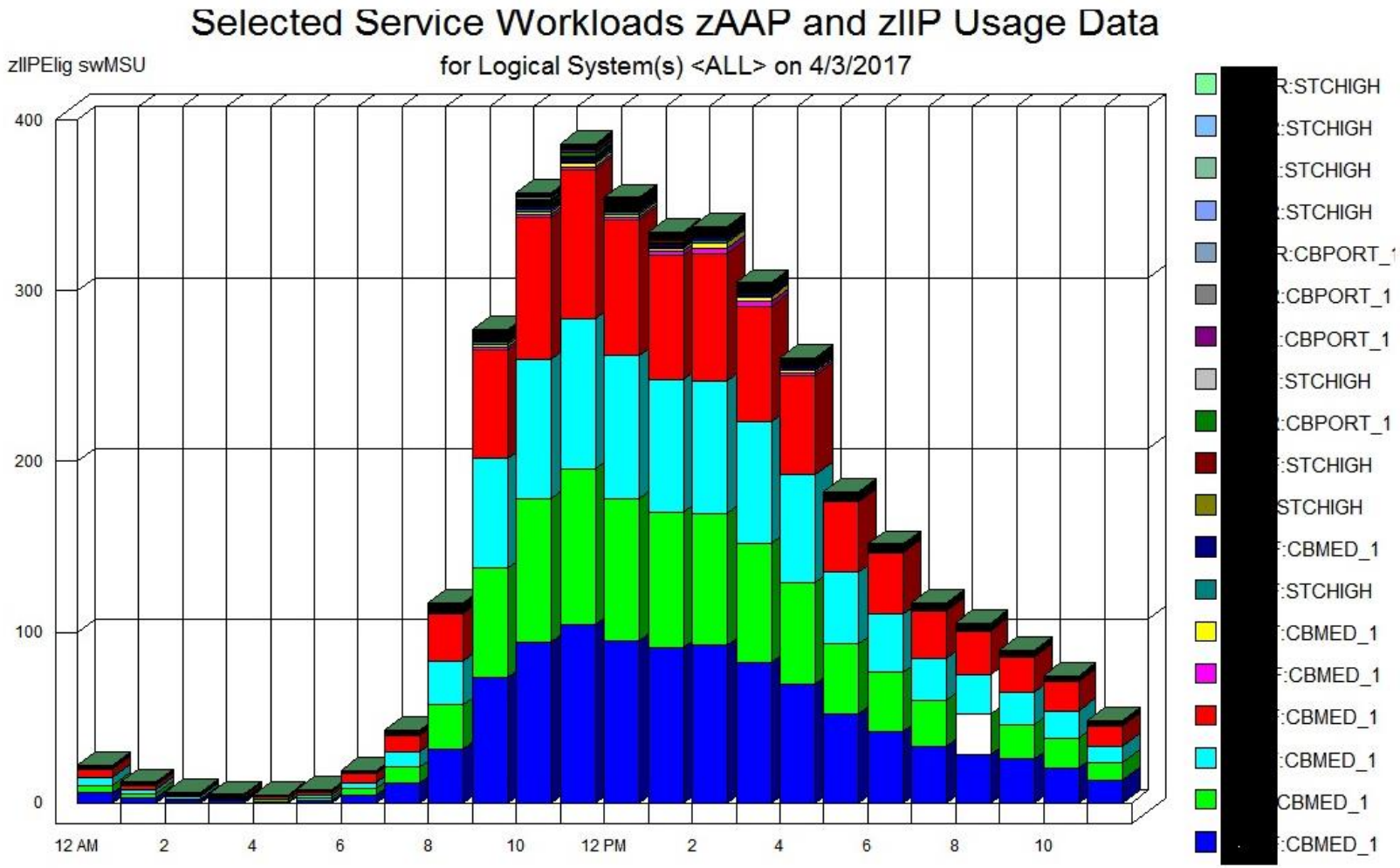
- 125% of 2 zIIP or 62% each
- Overflow to GCP > 50% of a GCP

JAVA – zIIP overflow

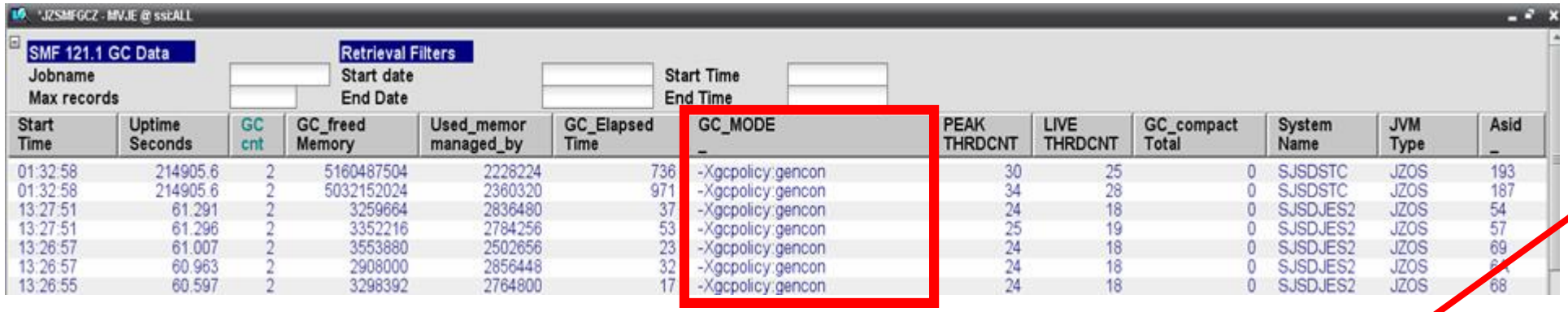


- Normalized
  - Using 3 zIIPs
- zIIP Eligible
  - Overflow 73% of a GCP
- Non Eligible
  - 61% of a GCP

WebSphere = ~ 400 MSUs overflow



Garbage Collection Modes



Start Time	Uptime Seconds	GC cnt	GC_freed Memory	Used_memor managed_by	GC_Elapsed Time	GC_MODE	PEAK THRDCNT	LIVE THRDCNT	GC_compact Total	System Name	JVM Type	Asid
01:32:58	214905.6	2	5160487504	2228224	736	-Xgcpolicy:gencon	30	25	0	SJSDSTC	JZOS	193
01:32:58	214905.6	2	5032152024	2360320	971	-Xgcpolicy:gencon	34	28	0	SJSDSTC	JZOS	187
13:27:51	61.291	2	3259664	2836480	37	-Xgcpolicy:gencon	24	18	0	SJSDJES2	JZOS	54
13:27:51	61.296	2	3352216	2784256	53	-Xgcpolicy:gencon	25	19	0	SJSDJES2	JZOS	57
13:26:57	61.007	2	3553880	2502656	23	-Xgcpolicy:gencon	24	18	0	SJSDJES2	JZOS	69
13:26:57	60.963	2	2908000	2856448	32	-Xgcpolicy:gencon	24	18	0	SJSDJES2	JZOS	6[illegible]
13:26:55	60.597	2	3298392	2764800	17	-Xgcpolicy:gencon	24	18	0	SJSDJES2	JZOS	68

≠ Lower CPU

Policy Mode	Optimized for	Default for
concurrentScavenge (AKA Pause-less)	Large heaps and reduced pause times NEW	Optional Java 8 SR5 on z14 only
gencon	Transactional workloads	Java 6.1,7,8
optthruput	Throughput	Java 5 & 6.0
balanced	Large heaps	
optavgpause	Reduced pause times	Obsolete

GSE UK Conference 2018

Better, stronger, faster; The Mainframe….. the Machine!

GSE GUIDE SHARE EUROPE UK REGION

Tuesday 6th November

Start	End	Stream	Room	Title	Speaker
11:45	12:45	IMS	Wellington B	The No Cost Way to Manage the IMS Catalog	David Schipper
15:00	16:00	IMS	Wellington B	Current Trends in IMS Analytics	David Schipper
16:30	17:30	zCMPA	Woodcote	zIIP stealing GCP MSUs for Capacity Management	Donald Zeunert

Wednesday 7th November

Start	End	Stream	Room	Title	Speaker
09:30	10:30	Db2	Nurburgring	Know your onions when it comes to Db2 indexes	Randy Bright
09:30	10:30	IMS	Wellington B	IMS Checkpoint Pacing	David Schipper
10:45	11:45	zCMPA	Woodcote	How many GCP MSU is my CF stealing?	Don Zeunert

Stop by the BMC booth for more information about sessions or other Q/A with speakers

bmc

Session feedback

- Please submit your feedback at http://conferences.gse.org.uk/2018/feedback/lLI
- Paper feedback forms are also available from the Chair person
- Session is LI

Contact: Donald_Zeunert@bmc.com