



IBM Systems & Technology Group

z/VM Performance Update Session 9106

Bill Bitner
VM Performance Evaluation
bitnerb@us.ibm.com

8/10/2008

© 2007 IBM Corporation

Trademarks

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: AS/400, DBE, e-business logo, ESCO, eServer, FICON, IBM, IBM Logo, iSeries, MVS, OS/390, pSeries, RS/6000, S/30, VM/ESA, VSE/ESA, Websphere, xSeries, z/OS, zSeries, z/VM

The following are trademarks or registered trademarks of other companies

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation
Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries
LINUX is a registered trademark of Linus Torvalds
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
Intel is a registered trademark of Intel Corporation
* All other products may be trademarks or registered trademarks of their respective companies.

NOTES:

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

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

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

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

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

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

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

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use.

The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

Introduction

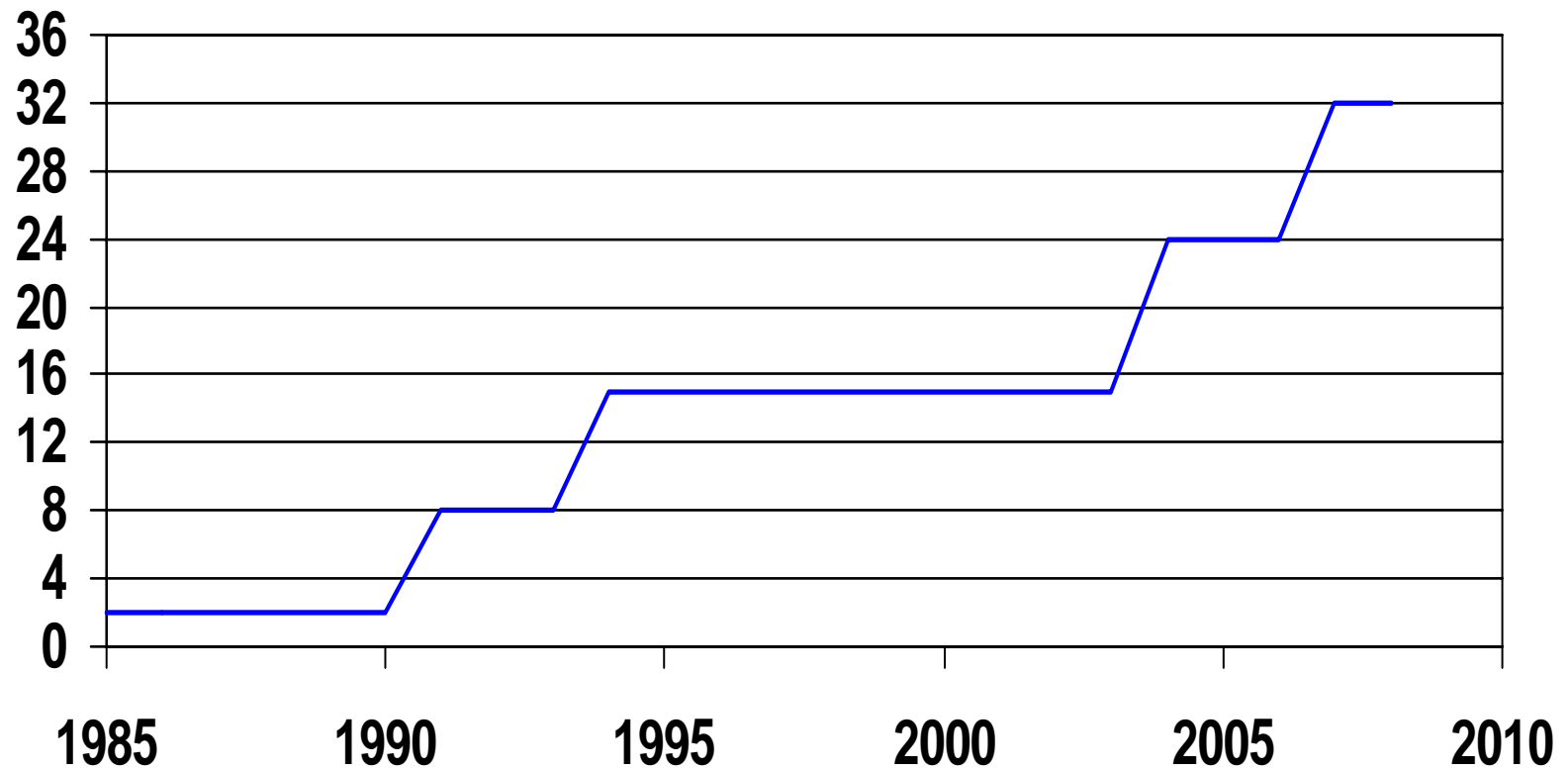
- **z/VM 5.3.0 Performance**
 - Line Items that have an impact
- **z10 Performance**
- **z/VM 5.4.0 Performance Preview**
- **See Performance Report on web for details**
 - <http://www.vm.ibm.com/perf/reports/>

z/VM 5.3.0

- **GA June 29, 2007**
 - Many more details available in the z/VM Performance Report:
 - <http://www.vm.ibm.com/perf/reports/>
- **Scalability and capability extended in several directions**
 - Processors, Memory, I/O, Network
 - What were the old limits?
 - What are the new limits?
- **Other Performance Enhancements**

Processor Scaling

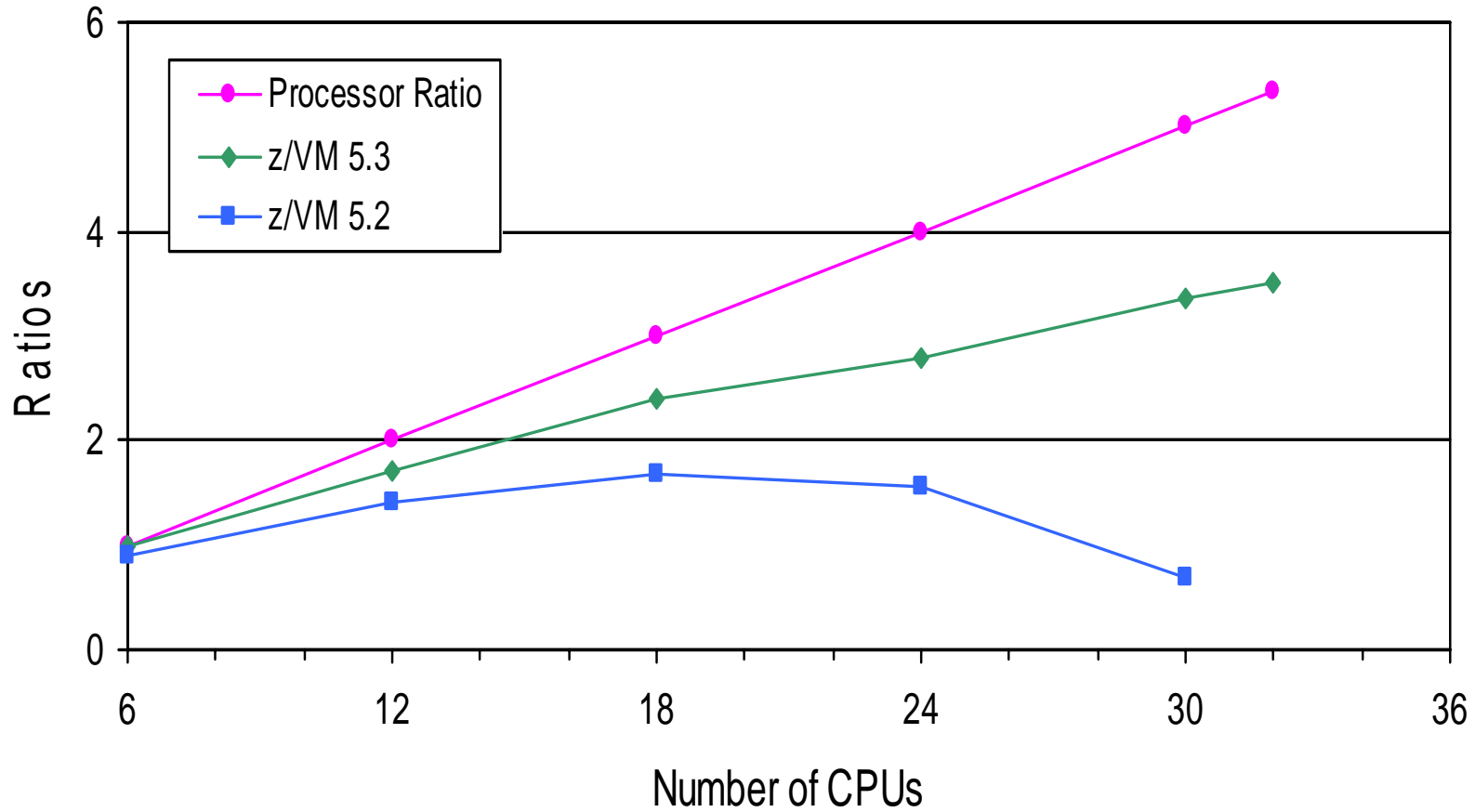
Number of Supported Processors



Greater than 24 CPU Support

- **While z/VM 5.2.0 would run on up to 31 processors, it only supported 24 due to performance limitations**
- **z/VM 5.3.0 supports 32 processors**
- **Serialization Changes**
 - General support for exclusive and shared formal spin locks
 - First to exploit is the Scheduler Lock (SRMSLOCK)
 - New lock associated with each Processor Local Dispatch Vector (PLDV) for dispatching (DSVLOCK)
- **Performance is Workload Dependent**
 - Watch for Master Processor Limitations
 - Tend to be more traditional workloads, not Linux environments
 - Single non-MP virtual machine limits
 - Example: DB2 for z/VM & VSE can only use 1 processor

Large N-Way Effects on ITR Ratio



Linux Guests with Apache Webserving

Metrics for Formal Spin Locks

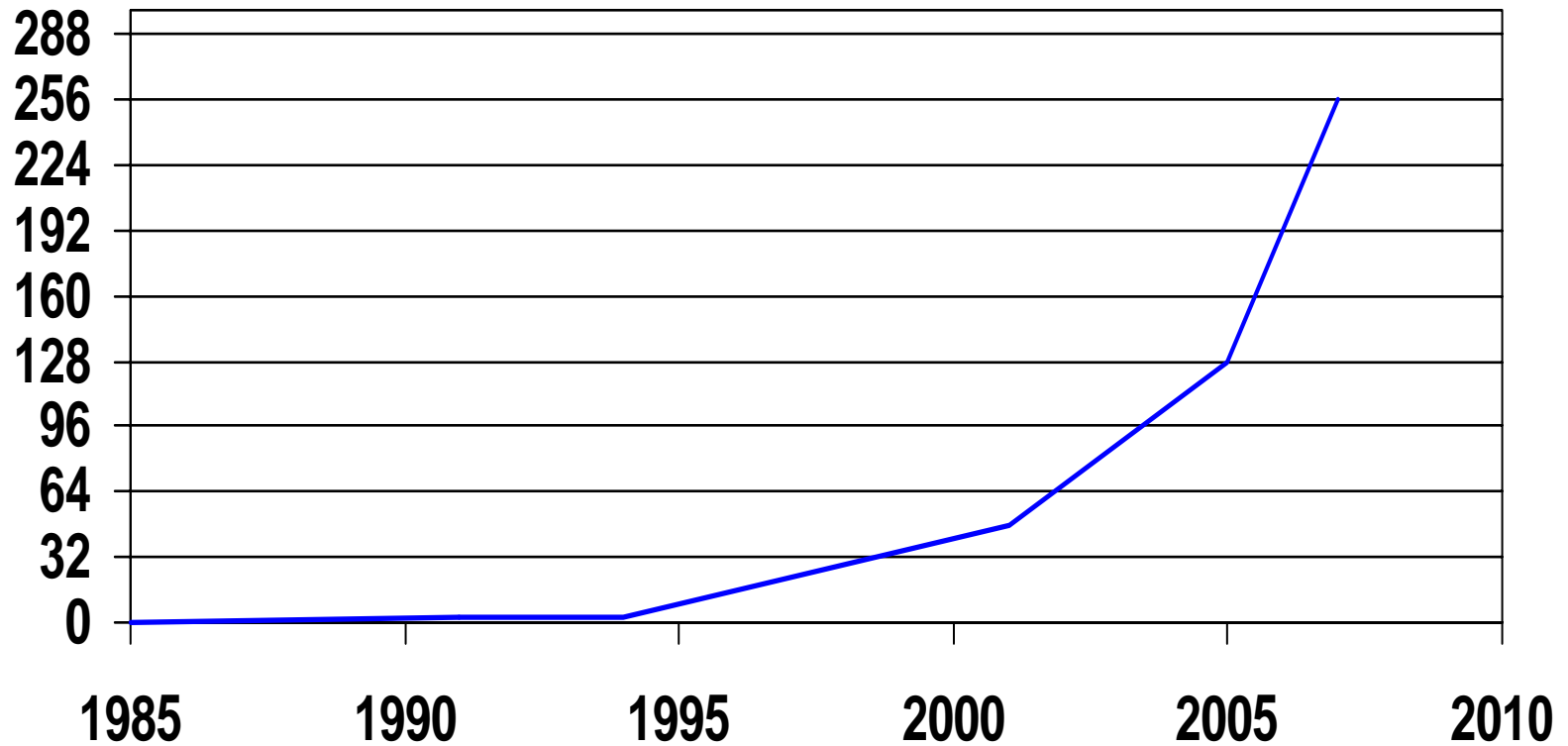
FCX265 CPU 2094 SER 19B9E Interval 02:31:51 - 12:34:01 GDLVM7

```

<----- Spin Lock Activity ----->
<----- Total -----> <--- Exclusive ---> <----- Shared ----->
Interval                Locks Average   Pct   Locks Average   Pct   Locks Average   Pct
End Time LockName      /sec   usec   Spin  /sec   usec   Spin  /sec   usec   Spin
>>Mean>> SRMATDLK      1.9    .539   .000   1.9    .539   .000   .0    .000   .000
>>Mean>> RSAAVCLK       .0    2.015   .000   .0    2.015   .000   .0    .000   .000
>>Mean>> FSDVMLK       .0   24.97   .000   .0   24.97   .000   .0    .000   .000
>>Mean>> SRMALOCK       .0    .000   .000   .0    .000   .000   .0    .000   .000
>>Mean>> HCPTRQLK      4.1    .195   .000   4.1    .195   .000   .0    .000   .000
>>Mean>> SRMSLOCK     34.0   1.096   .001  32.7   1.037   .001   1.3   .001   .000
    
```

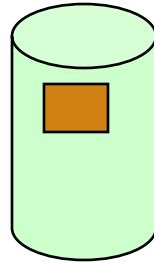
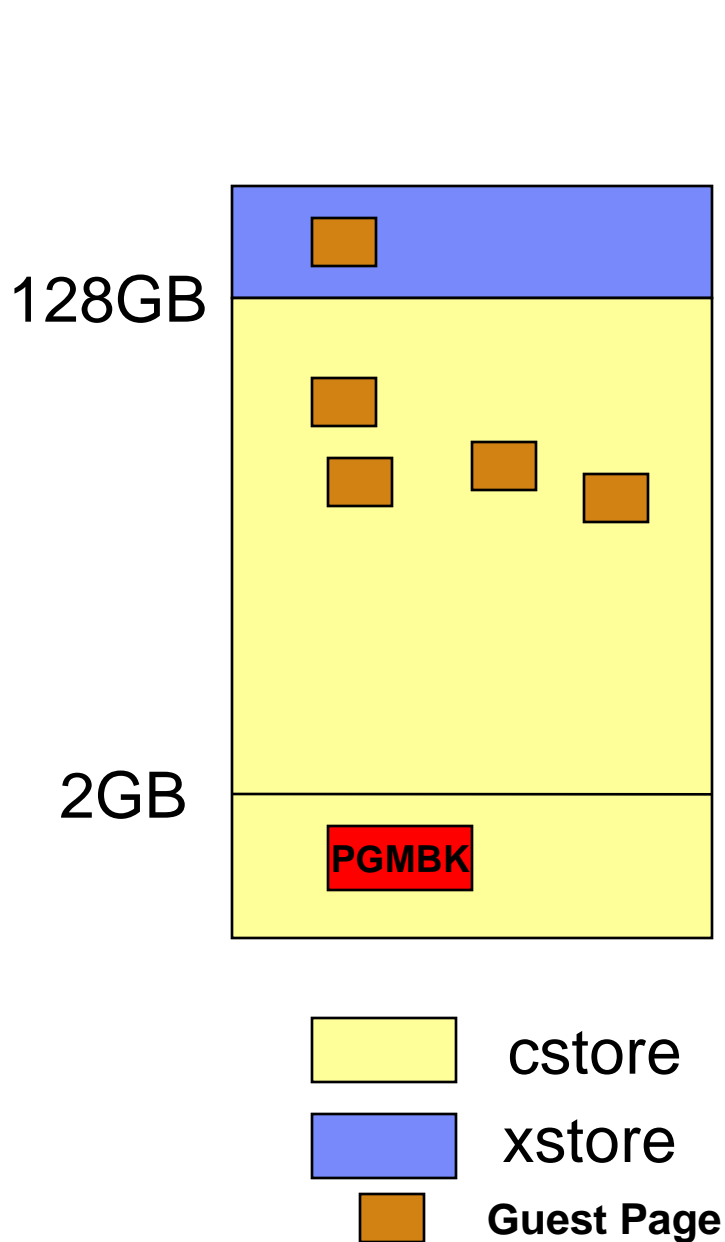

Memory Scaling

Effective Real Memory Use Limits



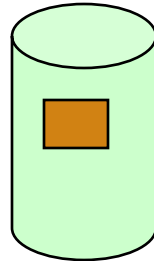
Greater than 128GB Memory Support (256GB)

- **z/VM 5.3 Improvements:**
 - PGMBKs allowed to be allocated above 2GB
 - Each PGMBK is 8KB (2 contiguous frames)
 - Enhanced contiguous frame management
- **Also seeing improvements to smaller configurations that are memory constrained**
- **Be careful with memory terminology**
 - Try to define various terms when you use them or hear them
 - Examples:
 - Defined
 - Resident
 - Backed
 - Active
 - Actively Referenced
 - Addressable

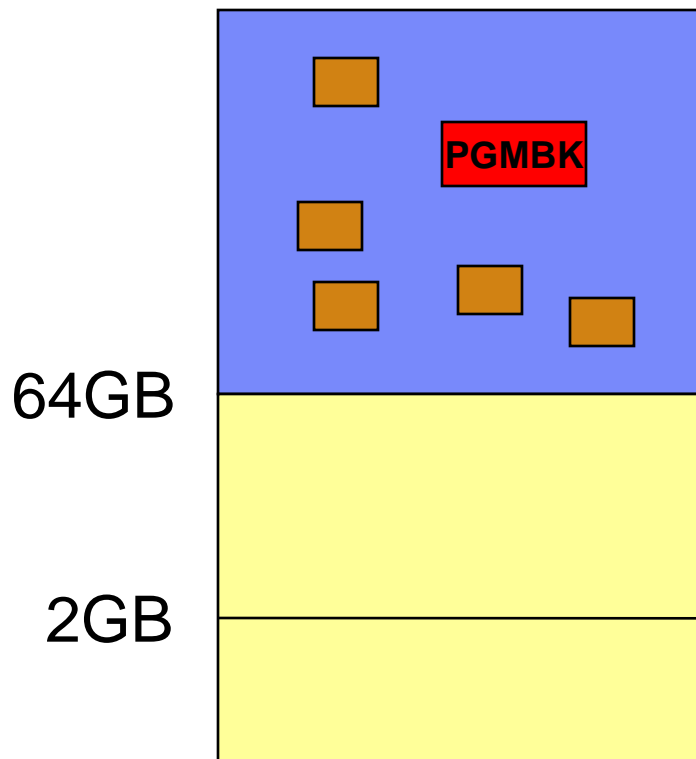


z/VM 5.2.0 Traditional Xstore:Cstore Ratio

- PGMBK is required for any guest page that is acted on
- PGMBK is 8KB (2 contiguous frames)
- Resident PGMBK located below 2GB
- PGMBKs are pageable
- PGMBK resident if any guest pages it represents are resident
- Perfect world: ~256GB of virtual memory actively being referenced
- Realistic world: ~128GB of virtual memory actively being referenced

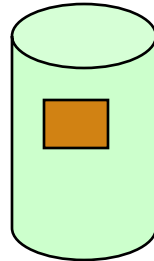
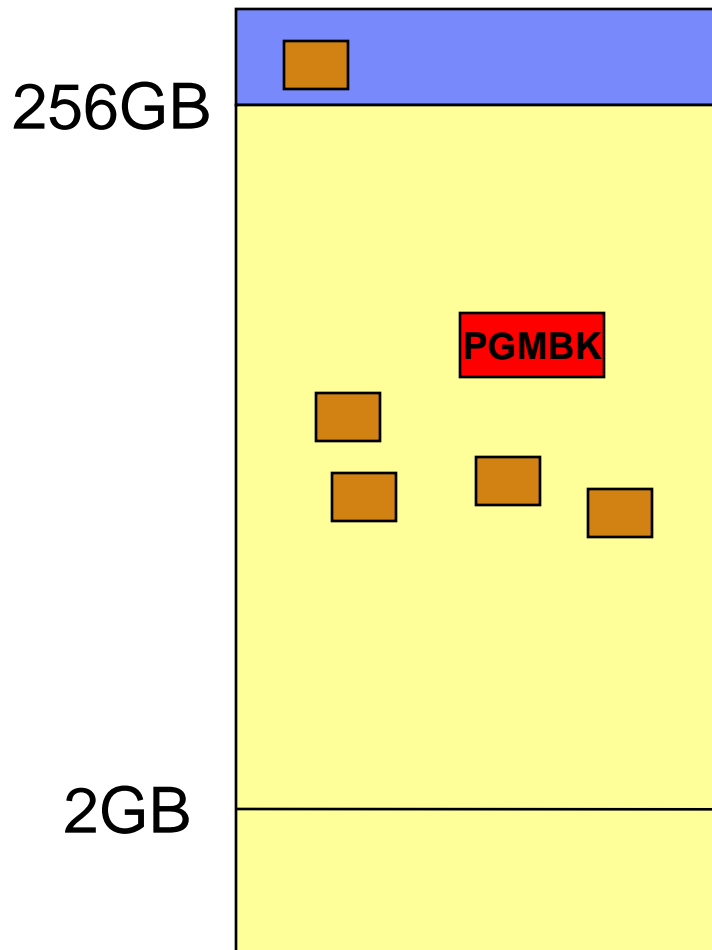


z/VM 5.2.0 Larger Xstore:Cstore Ratio



- More expanded storage can increase the likely hood that guest pages are moved out of central storage.
- This allows PGMBKs to also be moved.





z/VM 5.3.0 PGMBKs above 2GB Bar

- PGMBKs can reside anywhere but still must be better in the hierarchy than the guest pages they represent
- Next Limit: amount of PTRM space
 - Each space 4GB in size mapping 500GB of memory
 - Limit of 16 Spaces
 - Totals: 8TB of virtual machine memory



Performance Toolkit DSPACESH Report

```

FCX134      CPU 2094  SER 19B9E  Interval 13:04:01 - 13:09:01      GDLVM7
                <-----Number of Pages----->
Owning                <--Resid--> <-Locked--> <-Aliases->
Userid  Data Space Name  Total Resid R<2GB  Lock L<2GB  Count  Lockd  XSTOR  DASD
SYSTEM  PTRM0000          1049k 35602  1104      0      0      0      0      980 7502

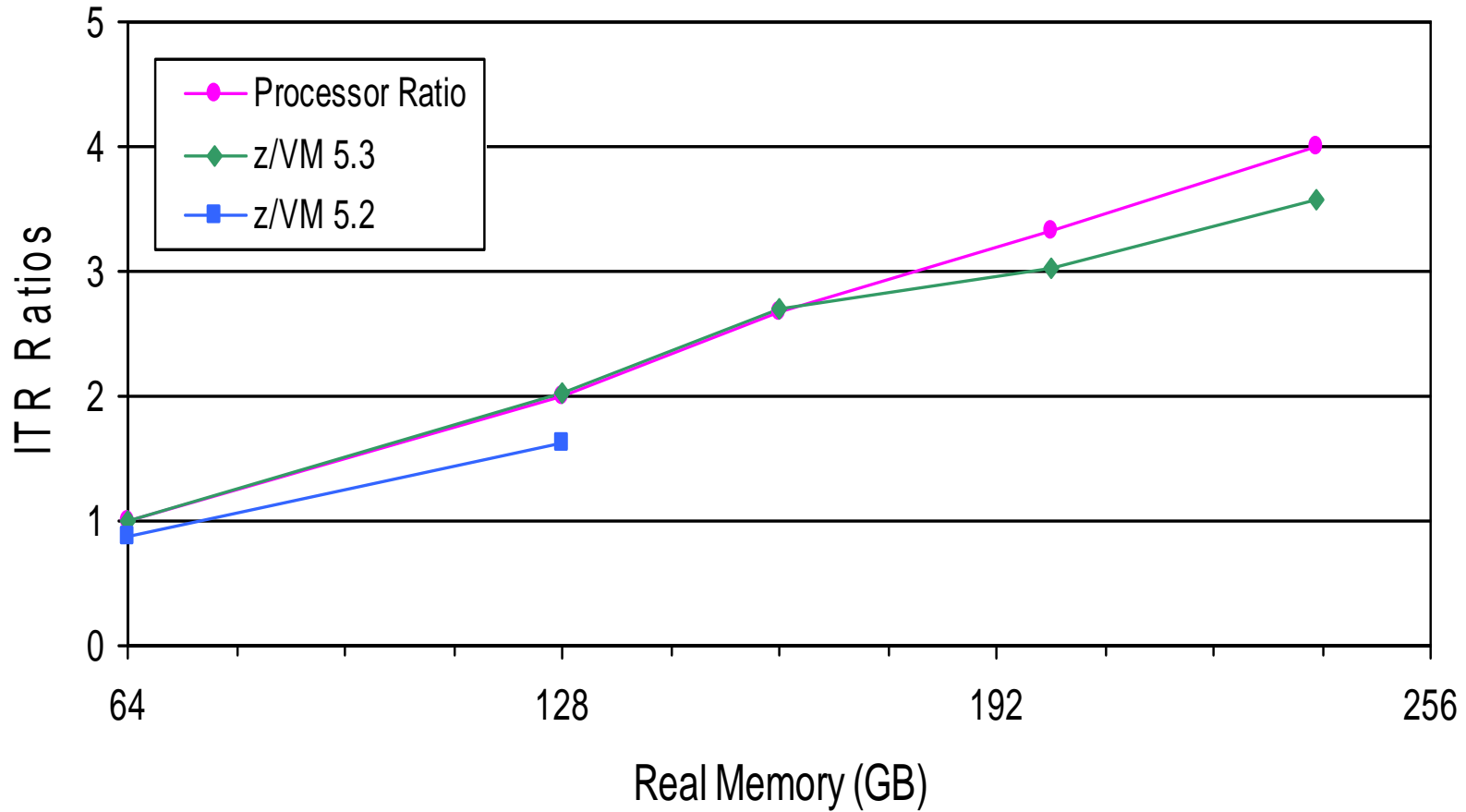
```

- Slightly edited FCX134 report
- PGMBKs live in PTRM0000, PTRM0001, ... PTRM000F
- Most systems will just have a PTRM0000

Limitations for Memory

- **Memory limitations dependent on workload & configuration**
 - 256GB Real memory
 - 8TB of 'addressable' virtual machine memory – Limit of Page Tables
 - Paging Space (optimal when <50% full)
 - 11.2TB for ECKD
 - 15.9TB for Emulated FBA on FCP SCSI
 - Virtual Machine Size (HW Dependent)
 - 1TB on z9

Scaling Memory Results - Apache Webserving



Improvement in Memory Constrained Environments

Scenario	Contention (Page Reqs per CPU-second)	Delta Thruput	Delta Total CPU/Tx
3G/4G 2084 3-way	2159	+10.3%	-9.5%
64G/2G 2094 3-way	0	+1.0%	-0.9%
64G/2G 2094 3-way	352	+15.5%	-12.7%
128G/2G 2094 6-way	291	+21.6%	-19.4%

Results of various Linux Apache measurements comparing z/VM 5.3 to z/VM 5.2

z/VM & CMM

- **z/VM Memory Management Concepts**
- **Approaches to improve the situation**
 - Asynchronous Page Fault
 - VMRM use of Cooperative Memory Management
 - Aka: CMM 1, VMRM-CMM
 - Collaborative Memory Management Assist
 - Aka: CMM 2, CMMA, MEMASSIST
- **Uses for each**

z/VM Concepts – Available Lists

- **Series of list of frames available to be used**
- **Avoids having to take time to search for a usable frame when a request from other subsystems or pagefault for guest occurs**
- **Depending on release, lists for: <2GB vs. >2GB, single frames vs. contiguous frames.**
- **Dynamic Thresholds**
 - Low threshold – kick off processing to get more available frames
 - High threshold – number of frames we want to have to stop process kicked off by low threshold
- **Most severe impact is when Available List goes empty**

Performance Toolkit FCX254 AVAILLOG

```

<----- Thresholds -----> <----- Page Frames ----->
<---Low---> <---High---> <Available> <Obtains/s> <Returns/s>
<2GB >2GB <2GB >2GB <2GB >2GB <2GB >2GB <2GB >2GB
    0   356   920  1271   175k   904   16.0  9070   1.4  9070

<-Times-> <----- Replenishment -----> Perct
<-Empty-> <---Scan1---> <---Scan2---> <-Em-Scan-> Scan Emerg
<2GB >2GB Compl Pages Compl Pages Compl Pages Fail Scan
    0    5   6472 5431k    30 33864  6101 3985k  443   50
    
```

z/VM Concepts – Demand Scan

- **Process where we try to replenish available list**
- **Series of passes through various lists of ‘owned’ frames.**
 - Pass 1, Pass 2, Emergency Pass
 - Passes oriented to ‘steal’ pages that will hurt the users and system the least
 - Example of information considered:
 - Long term dormant, short term dormant, active virtual machine
 - Referenced or not referenced
 - Shared or private
 - ...

Performance Toolkit FCX259 DEMNDLOG

```

<----- Demand Scan Pass 1 -----> <----- Demand Scan Pass 2 -----> <----- Emergency Scan ----->
<-- Ended After --> <--- Page Frames ---> <-Ended After-> <--- Page Frames ---> <-Ended After-> <-Page Frames-->
Interval  Lng Drm NSS Eli Dsp Long Dor- NSS Eli Dsp Lng Drm Eli Dsp Long Dor- NSS Eli Dsp Drm NSS Eli Dsp Dor- NSS Eli Dsp  Scan
End Time  Drm ant Shr Lst Lst Dorm mant Shr Lst Lst Drm ant Lst Lst Dorm mant Shr Lst Lst ant Shr Lst Lst mant Shr Lst Lst Failed
>>Mean>>  0  3  0  0 181 1886 9129  0  0 1M  0 13  0  0  0 21k  0  0  0  3  9  0 1k 3331 7k  0 2M 3346
09:58:36  0  0  0  0  17  0  543  0  0 81k  0  0  0  0  0  960  0  0  0  0  1  0  46  29 261  0 99k  155
09:59:06  0  0  0  0  3  0  22  0  0 42k  0  0  0  0  0 1162  0  0  0  0  0  0  49 161 229  0 77k  88
09:59:36  0  0  0  0  3  1  401  0  0 38k  0  0  0  0  0  709  0  0  0  1  0  0  47  39 117  0 62k  87
10:00:06  0  1  0  0  1  2  517  0  0 20k  0  0  0  0  0  538  0  0  0  0  0  0  27  84 197  0 37k  46
10:00:36  0  0  0  0  4  0  859  0  0 54k  0  2  0  0  0  417  0  0  0  0  1  0  35  17 789  0 0M  140
    
```

z/VM Concepts – Reorder & Page Release

■ Reorder

- Processing of reordering frame 'owned' lists
- Reference bit processing

■ Page Release

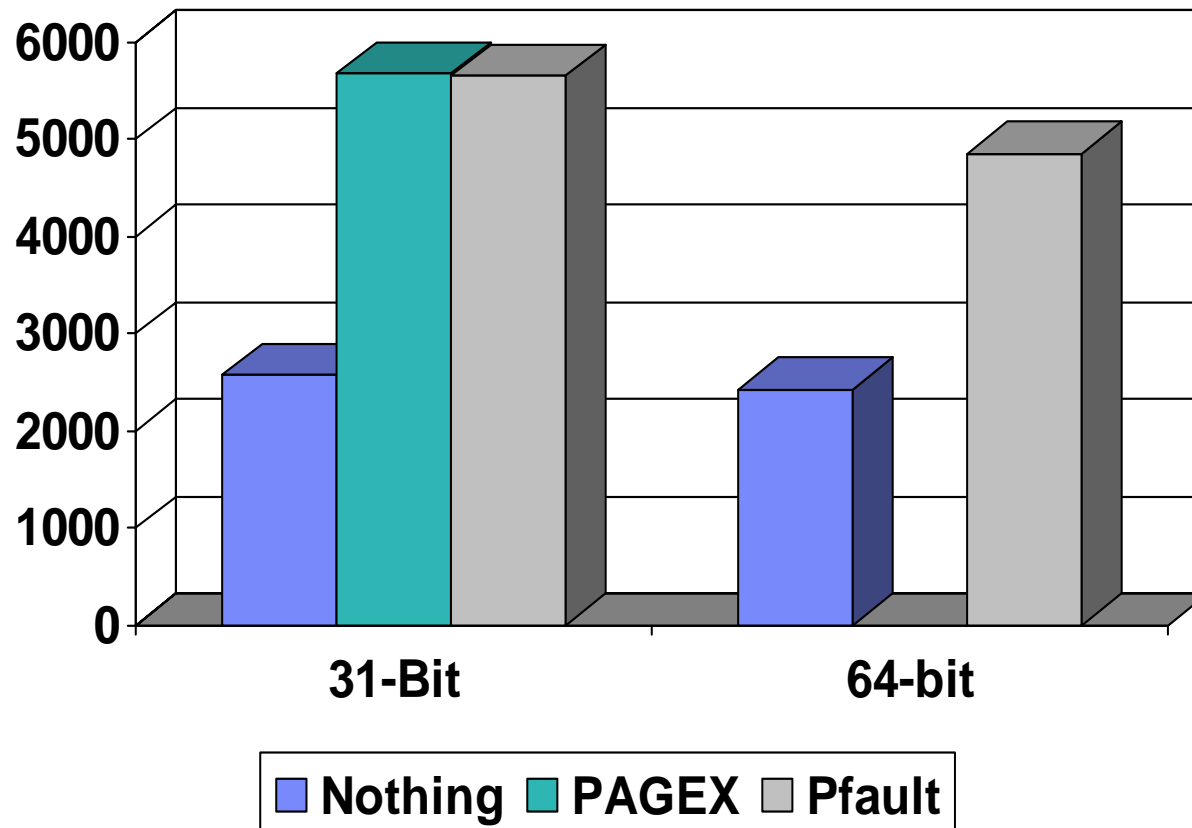
- Ability for a guest to tell CP that it no longer needs a guest real page.
- CP no longer needs to back the page
- Two methods:
 - Diagnose x'10' – Page Release – Used by Linux
 - Diagnose x'214' – Pending Page Release – Used by CMS

Asynchronous Page Fault Processing

- **Ordinarily, page faults serialize the virtual machine. This can be a throughput and response time problem for guest systems**
- **PAGEX was implemented for VSE**
 - Limited to 31-bit
- **Enhancements designed for Linux**
- **PFAULT macro**
 - Accepts 64-bit inputs
 - Provides 64-bit PSW masks
- **Diagnose x'258'**

Page Fault Tests

NonFaulting Throughput



CMM 1: Overview

- **Customer identifies guests to participate**
- **VM Resource Manager (VMRM) tracks system memory utilization / demand, computes target “resident footprint” for each guest**
- **VMRM sends messages to guest to adjust footprint**
- **Guest device driver receives messages, uses “ballooning” to comply**
 - Driver requests specified number of pages from kernel (*kmalloc()*)
 - *kmalloc()* will trigger reclaim of guest real memory
 - Capitalizes on existing guest logic to cast out the least valuable pages
 - Driver instructs VM to release page contents (Diagnose x'10')

CMM 1 Verbs

- **Sent to Linux guests through VM “SMSG”**
- **Message formats:**
 - **CMM SHRINK *npages*** – resize “permanent” balloon to *npages* (absolute count; may be higher or lower than previous SHRINK)
 - **CMM RELEASE *npages*** – inflate “temporary” balloon by an additional *npages* pages
 - **CMM REUSE *npages nseconds*** – sets deflation rate for temporary balloon
 - Pages redeployed gradually: total of *npages* pages every *nseconds* seconds, until temporary balloon deflated
 - Setting persists until next REUSE message (through multiple RELEASE cycles)

Enabling guests for VMRM-CMM

- **In Linux guests, indicate that VM RMSVM is eligible to send CMM1 commands.**
- **In VM RM, supply list of virtual-machine names to VM RM**
 - New VM RM CONFIG statement
 - NOTIFY MEMORY *userid1* [*userid2... useridx*]
 - Not related to other statements (WORKLOAD, GOAL, ..)

VMRM-CMM Suggested Levels

- **Suggested levels, please follow normal service research**
- **z/VM**
 - z/VM 5.2.0 + VM64085
 - z/VM 5.3.0
 - New APAR VM64439 for z/VM 5.2.0 & z/VM 5.3.0
 - z/VM 5.2.0 PTF UM32427
 - z/VM 5.3.0 PTF UM32428
- **Linux**
 - SLES9 SP3
 - See <http://www.vm.ibm.com/sysman/vmr/vmrmmcm.html> for additional patches.
 - SLES10
 - RHEL4 U7 2.6.9-73
 - RHEL5.1

CMM 2 - Objectives

- **Pass memory usage information from pageable guest to host**
 - Attributes per 4K-byte block of guest absolute storage
- **Make guest aware of page state changes by host**

Benefits

- **Host memory management efficiency**
 - More intelligent selection of page frames to be reclaimed
 - Reduced reclaim overhead: avoid page writes where possible
- **Guest memory management efficiency**
 - Avoid double-clearing of page on reuse
 - Option to favor host-resident pages on allocation requests
- **Reduce guest memory footprint**
- **Support greater memory overcommit ratios**

CMM 2 states

- **Cross-product of guest-specified and host states**
- **Guest-specified states (“block-usage states”)**
 - Stable (S)
 - Unused (U)
 - Volatile (V)
 - Potentially Volatile (P)
- **Host states (“block-content states”)**
 - Resident (r) – contained in host main memory
 - Preserved (p) – backed on host auxiliary storage
 - Logically zero (z) – not backed; will appear as zeros on reference

Enabling for CMMA

- **z/VM – enabled by default**
 - To disable, issue before guests boot
`CP SET MEMASSIST OFF`

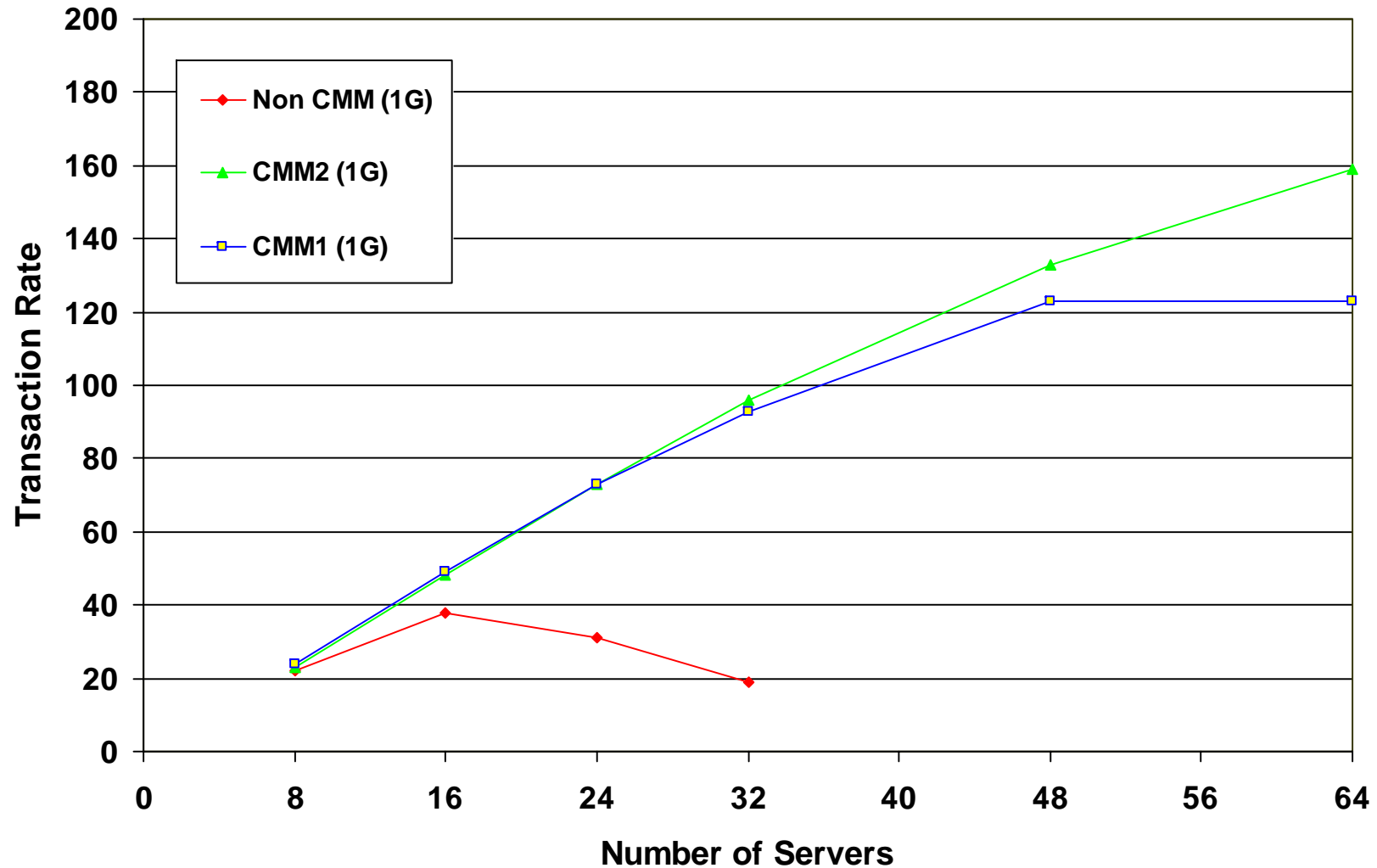
- **Linux – disabled by default**
 - To enable, add following in Linux parm file
`cmma=on`

CMMA Suggested Levels

- **Suggested levels, please follow normal service research**
- **z/VM**
 - z/VM 5.3.0 + APAR VM64265 + APAR VM64297
- **Linux**
 - SLES10 SP1 update kernel 2.6.16.53-0.18
- **Hardware**
 - z9 or z10 for ESSA instruction

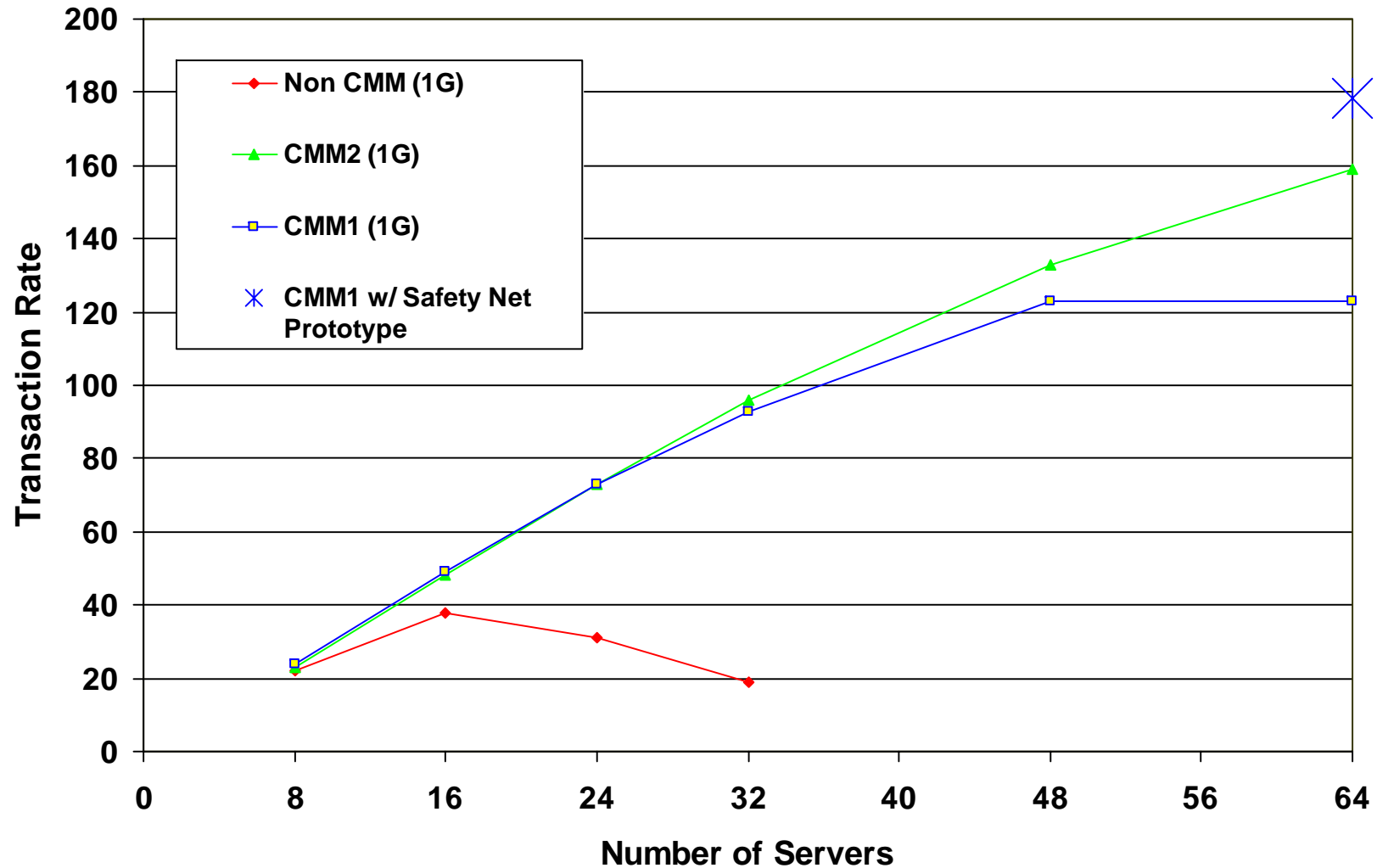
Transaction Rate vs. Number of Servers

for various Storage Management Products using Apache servers with a virtual storage size as shown in parenthesis in the legend; z9 6GB / 2GB



Transaction Rate vs. Number of Servers

for various Storage Management Products using Apache servers with a virtual storage size as shown in parenthesis in the legend



Other Uses and Approaches

- **Use VMRM-CMM for test virtual machines and leave production virtual machines off the Notify statement.**
- **Use of CP SET RESERVE for mission critical guests**
- **CMMA on older machines (pre z9)**
 - Simulation of ESSA instruction
 - Our workloads showed 26% increase in CP Processor time per transaction

More Information on CMM

- **z/VM Performance Report**

- <http://www.vm.ibm.com/perf/reports/zvm/html/530cmm.html>

- **VMRM & CMM**

- <http://www.vm.ibm.com/sysman/vmr/vmrvcmm.html>

- **z/VM Large Memory – Linux on System z**

- <http://www-03.ibm.com/support/techdocs/atmastr.nsf/WebIndex/WP101151>

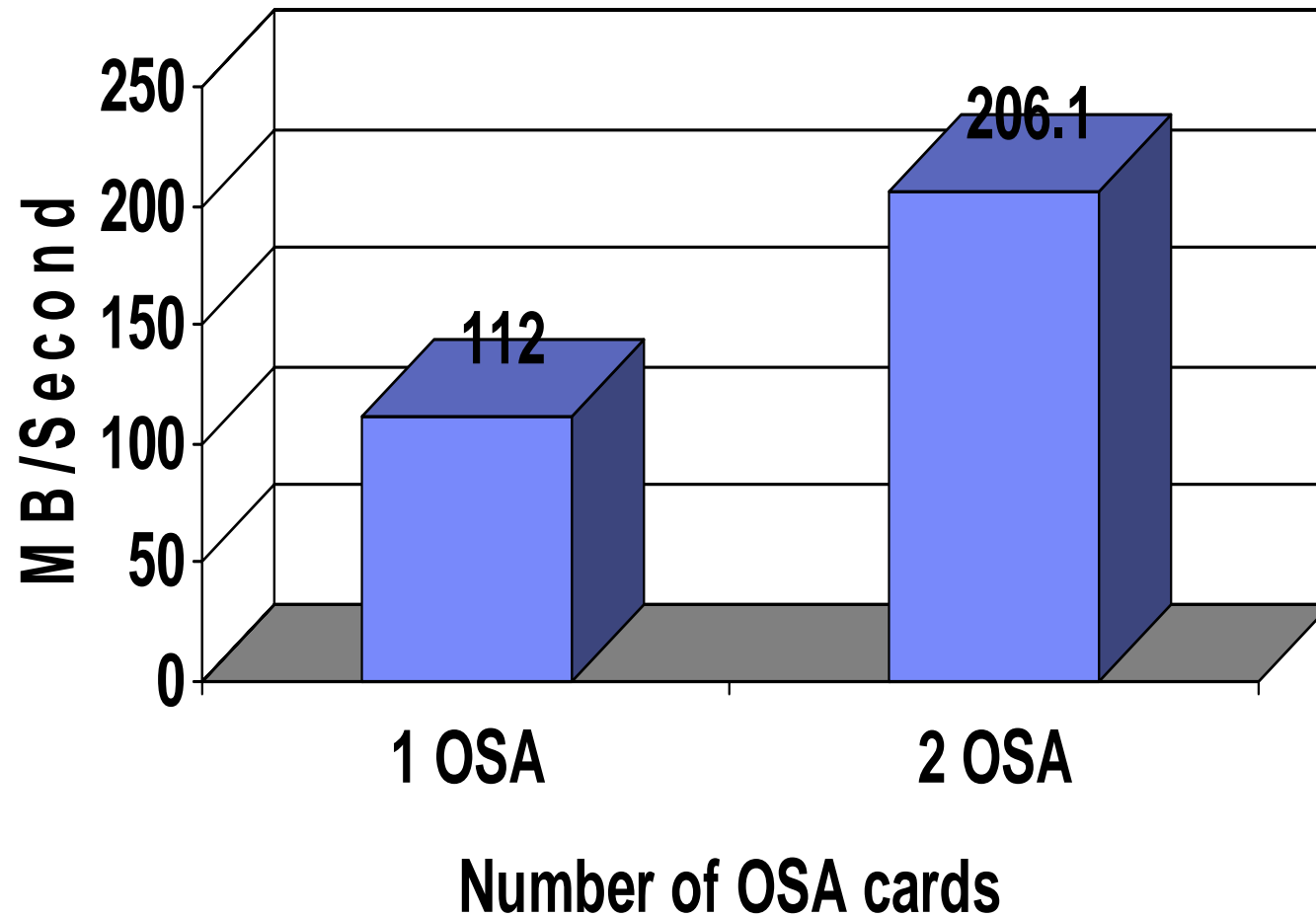
- **Linux information on Out of Memory notifier patch**

- <http://www.ibm.com/developerworks/linux/linux390/linux-2.6.16-s390-09-october2005.html>

Virtual Switch Link Aggregation

- **Ability to attach multiple OSAs to a single virtual Switch**
 - Aggregate bandwidth
 - Failover
- **Requires:**
 - z9 OSA-Express2 Support
 - Running in Layer 2 Mode
- **Dynamic Load Balancing**
 - Influenced by distribution of MAC addresses
 - Influenced by Physical Switch for inbound traffic
 - Cannot balance a single connection.
 - Example: a single data streaming connection will not get split across OSAs.

Streaming Throughput Results



I/O Improvements

- **PAV support for minidisks was provided in z/VM 5.2 via APAR VM63855 (May 2006)**
 - Should also apply VM64199 if using Minidisk Cache
- **z/VM 5.3 adds support for the HyperPAV feature of IBM System Storage DS8000**
 - Requires VM64248 (UM32072)
 - Allows for the creation of pools of Alias devices which the control unit will associate with different Bases as needed.
 - Performance characteristics similar to previous PAV support.

Improved SCSI Disk Performance

- **Exploitation of SCSI write-same function of 2105 & 2107 improves CMS FORMAT of minidisks on SCSI volumes**
- **Additional pathlength reductions**
- **CP Paging to SCSI volumes now bypasses the FBA emulation, reducing processor resource requirements**

Monitoring Enhancements

- **Lots of new fields in Monitor for new function:**
 - specialty engines
 - Scheduler changes
 - HyperPAV support
 - Memory management
- **New monitor Domain for Virtual Network Devices**
- **Additional flexibility in MONWRITE utility for starting/stopping**
- **Various changes in Performance Toolkit for VM**

z10 Performance

- **IT DEPENDS!!!**
- **Processor cycle time greatly improved over z9**
 - ~2.6 times faster (4.4 GHz)
 - Comparable to other platforms
- **Laws of Physics Must be Obeyed**
- **Tradeoffs made in order to achieve above**
 - Memory Differences
 - Key Ops
- **ITR Ratios (examples see LSPR for most current numbers)**
 - z/OS: z10 EC 701 up to **1.62 times** that of the z9 EC 701
 - LSPR z/VM Measurements: 1.30 to 1.60
 - z/VM Endicott Lab measurements: 1.23 to 2.05

It Depends On....

- **Number of Processors**
 - Fewer processors, better ITRR
- **Storage References**
 - Smaller memory footprints, better ITRR
- **Data Movement**
 - Less data movement, better ITRR
- **Virtual I/O to Real Devices**
 - Less virtual I/O, better ITRR
- **Storage Overcommitment**
 - Less over commitment, better ITRR
- **Amount of memory involved in long searches**
 - Shorter & less frequent searches, better ITRR
- **Exploitation of New Features**
 - More exploitation of features, better ITRR

z/VM 5.2.0 & z10: LPAR Monitor Records

- **z/VM 5.2.0 Customers running on z10 will not get LPAR monitor records for other partitions on the CEC.**
- **z/VM 5.3.0 is fine.**
- **Other processors are fine.**
- **A hardware problem ODT S5619 is open to correct this.**
- **Corrected with:**
 - EC Level: F85901
 - MCL No: 006
 - Driver: D73G

Setting the proper expectations

- **z10 is a great machine, with a number of excellent attributes.**
- **Care must be taken when sizing migrations from z9 to z10.**
- **Additional Information:**
 - LSPR Q & A (complete)
 - Discuss range and factors affecting
 - Pointer to z/VM Web Page
 - z/VM Web Page
 - <http://www.vm.ibm.com/perf/z10.html>
 - “To MIPS or Not to MIPS, That is the Question!” by Gary King
 - http://sharew.prod.web.sba.com/proceedingmod/abstract.cfm?abstract_id=17583

z/VM 5.4.0 Performance Preview

- **Planned General Availability September 12, 2008**
- **Areas of Improvements**
 - Memory Management
 - Dynamic Memory Upgrade
 - SET SHARE options
 - Additional Specialty Engine Support
- **Full details on z/VM Performance Web Page Sept 19, 2008**
 - <http://www.vm.ibm.com/perf/reports/>

z/VM 5.4.0: Upper DAT Structures above 2GB

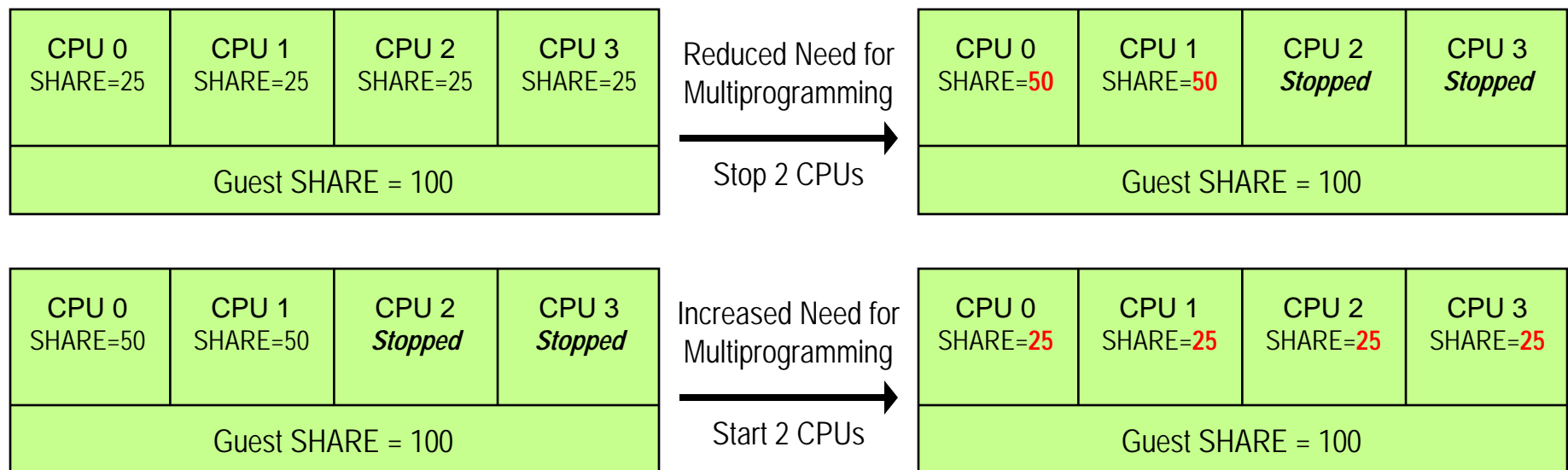
- **z/VM 5.3.0 allowed Page Tables (PGMBKs)**
- **z/VM 5.4.0 allows Segment Tables and Region Tables to be above 2GB**
 - These structures tend to require multiple contiguous frames.
 - Minimizes fragmentation and long searches
 - Resulting in better performance for some configurations
 - Prereq for increasing supported real memory (which we did not do this release)

z/VM 5.4.0: Dynamic Memory Upgrade

- **Aka Dynamic Storage Reconfiguration**
- **Dynamically add real central storage**
 - No removal; No expanded storage
- **SET STORAGE command response may be delayed**
 - Request for small increase in comparison to memory available to partition
- **Memory is not all initialized instantaneously**
- **Also virtualizes for guests that support it**
- **Some monitor event record anomalies exist**
 - Query Commands are correct
 - Will be fixed in service stream with VM64483

z/VM 5.4.0: Virtual CPU SHARE Redistribution

- **Allows z/VM guests to expand or contract the number of virtual processors it uses without affecting the overall CPU capacity it is allowed to consume**
 - Guests can dynamically optimize their multiprogramming capacity based on workload demand
 - Starting and stopping virtual CPUs does not affect the total amount of CPU capacity the guest is authorized to use
 - Linux CPU hotplug daemon starts and stops virtual CPUs based on Linux Load Average value
- **Helps enhance the overall efficiency of a Linux-on-z/VM environment**
- **Previously, Stopped virtual processors were given a portion of the guest share.**



Note: Overall CPU capacity for a guest system can be dynamically adjusted using the SHARE setting

z/VM 5.4.0: Specialty Engines Tuning

- **Now allow environment with real IFLs and real CPs in same LPAR (as well as other Specialty Engines)**
 - Called “z/VM-Mode LPAR” – be careful this phrase can be ambiguous at times
- **New options to SET SHARE allow a value for each processor type in a mixed virtual machine.**
- **CPUAFFINITY OFF will result in virtual specialty engines being dispatched on CPs**
- **Merging CP LPAR with IFL LPAR requires thought:**
 - First step, make virtual machines on IFL LPAR have virtual IFLs
 - For duplicated work (RACF, TCP/IP, etc.), need to determine which to use or in some cases which to duplicate
 - Remember that in some environments, the IFLs may be faster than the CPs.
 - Charge back model?

z/VM 5.4.0: DCSSs loaded above 2GB

- **DCSS can be loaded above 2GB**
 - DCSSs still limited to 2047MB
 - Continue to allow multiples, thus aggregate > 2GB
- **Requires additional Linux support:**
 - To use a segment above 2GB
 - To allow ‘stacked’ use of contiguous DCSSs to appear effectively as one large DCSS
- **Benefits of sharing much larger amounts of memory**
 - Varies significantly based on configuration and workload.
 - Example: Throughput improvements of ~35% for static web serving observed with prototype Linux code.

z/VM 5.4.0: DCSS Loading Time & Trade-offs

- **Support for DCSSs above 2GB means great amount of memory can be used for DCSSs.**
- **Performs well once DCSS is built and data saved.**
- **Placing at higher than necessary addresses results in more memory required for data structures in guest and in z/VM.**
- **Time required to save or fill a DCSS is non-trivial.**
- **Two primary methods of defining segments for use by Linux with unique pros and cons**
 - SR: Shared read-only access.
 - SN: Shared read/write access, no data saved.

z/VM 5.4.0: Trade Offs SN vs. SR

DCSS Attribute	SN: Shared R/W; not saved²	SR: Shared R/O
Non-Volatile (Saved to spool)	NO	YES
Initial elapsed time to 'make ready' the file system¹	~1 minute per GB	~5 minutes per GB
Spool processing for DCSS can delay other spool activity	NO	YES

¹ making ready the file system involves various steps of defining, copying data into the DCSS, and saving as necessary.

² technically the DCSS is read-write. After data is populated into the DCSSs, one would want to mount the files system read-only

VM64082 MDC Arbiter Changes

- **VM64082 (z/VM 5.2.0 & 5.3.0). Not on any RSU**
- **Rolled into base of z/VM 5.4.0**
- **This change improves MDC Arbiter choices in some environments (doesn't over bias for MDC).**
- **Also, makes other environments worse by biasing too much against MDC.**
- **Use SET MDC command to bound the range of memory to avoid problems.**
- **Work continues to make improvements here; look for future APAR.**

Summary

- **z/VM 5.2 Improvements via Service**
- **z/VM 5.3 significantly extends the capacity of:**
 - Processor
 - Memory
 - I/O
- **z/VM 5.4 incremental performance improvements**
- **See z/VM Performance Report for more details**
 - <http://www.vm.ibm.com/perf/reports/>