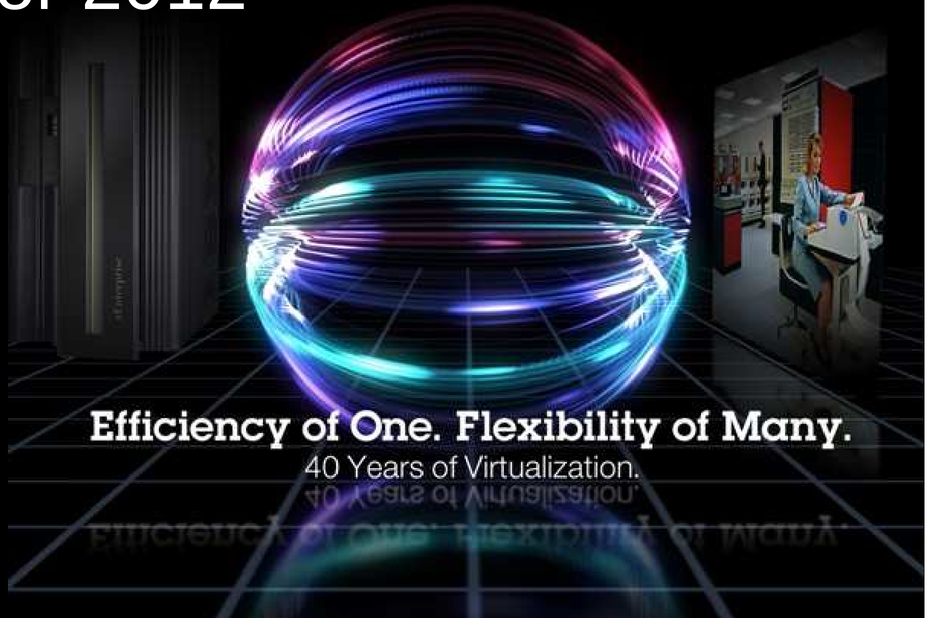
IBM Systems & Technology Group

z/VM Performance Update for 2012

SHARE Session 11796



Bill Bitner, bitnerb@us.ibm.com
Brian Wade, bkw@us.ibm.com

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.

Permission is hereby granted to SHARE to publish an exact copy of this paper in the SHARE proceedings. IBM retains the title to the copyright in this paper, as well as the copyright in all underlying works. IBM retains the right to make derivative works and to republish and distribute this paper to whomever it chooses in any way it chooses.

Acknowledgements – Your z/VM Performance Team

- **Dean DiTommaso**
- **Bill Guzior**
- **Steve Jones**
- **Virg Meredith**
- **Patty Rando**
- **Dave Spencer**
- **Susan Timashenka – Dept Manager**
- **Xenia Tkatschow**
- **Brian Wade**

Agenda

- **z/VM 6.2 thoughts**
  - LGR and SSI
    - Performance notes
    - Management and monitoring thoughts
  - Various other line items
  - Monitor record changes
  - Performance-related service
  - Other thoughts

z/VM 6.2 Highlights – A Performance View

- **Regression performance**
- **SSI and LGR considerations**
- **Memory management improvements**
- **MONDCSS and SAMPLE CONFIG increases**
- **STORBUF changes**
- **z/CMS and implications**
- **CPU Measurement Facility exploitation**
- **Monitor records**
- **z/VM Performance Toolkit changes**

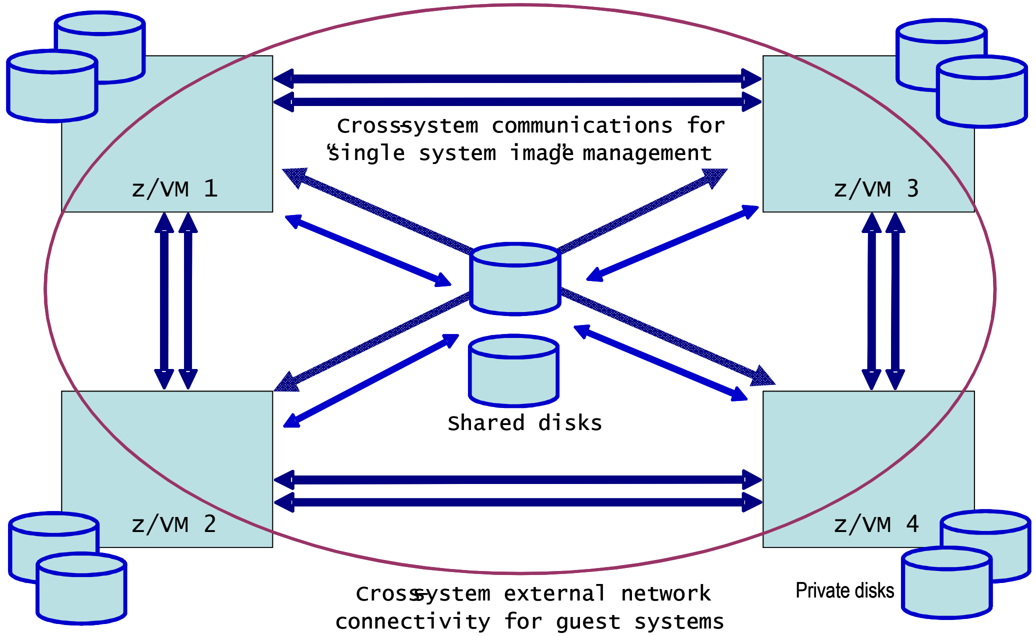
z/VM 6.2 Regression Performance

- **Ran our standard library of workloads**
  - CMS interactive, various Apache configurations
- **Results are within usual 5% regression criteria**
- **Some workloads will see improvements:**
  - Overprovisioned for logical PUs compared to utilization
  - Storage-constrained with heavy contention for <2 GB real storage
  - High virtual CPU to logical CPU overcommit with virtual CPUs often in a ready-to-run state

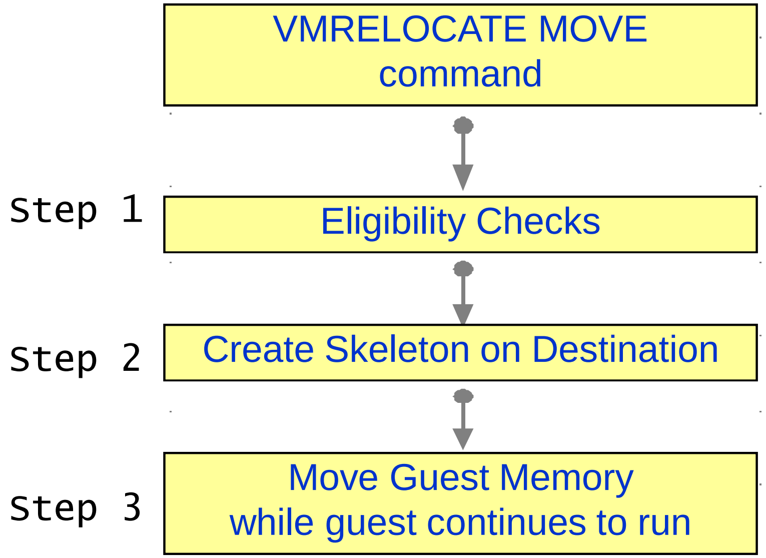
SSI and LGR Thoughts

Single System Image Feature
Clustered Hypervisor with Live Guest Relocation

- **Provided as an optional priced feature.**
- **Connect up to four z/VM systems as members of a Single System Image (SSI) cluster**
- **Provides a set of shared resources for member systems and their hosted virtual machines**
- **Cluster members can be run on the same or different System z servers**
- **Simplifies systems management of a multi-z/VM environment**
  - Single user directory
  - Cluster management from any member
    - Apply maintenance to all members in the cluster from one location
    - Issue commands from one member to operate on another
  - Built-in cross-member capabilities
  - Resource coordination and protection of network and disks

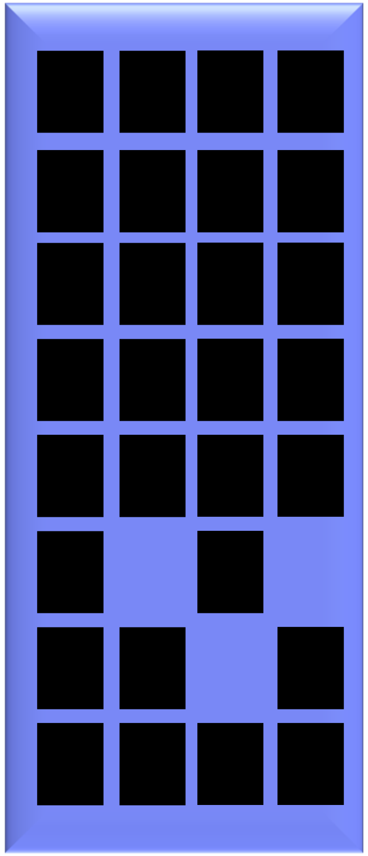


Stages of a Live Guest Relocation



LGR, High-Level View of Memory Move

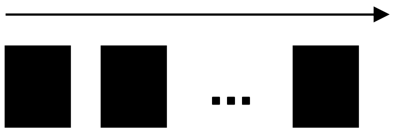
Source



Guest Address Space

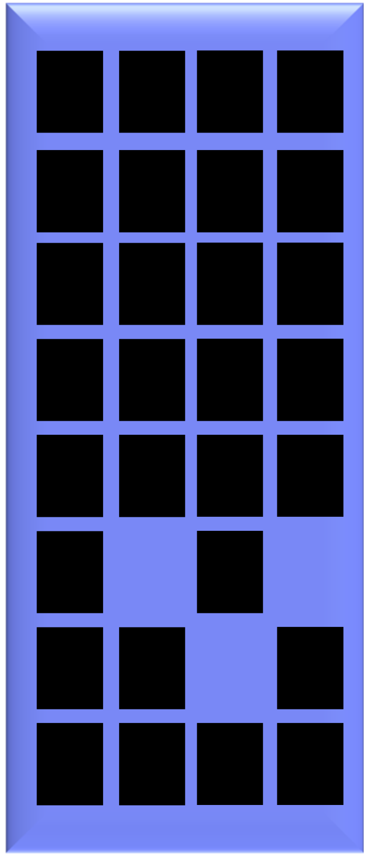
PUSH with resend

Pass 1



Walk through guest memory moving all non-zero pages

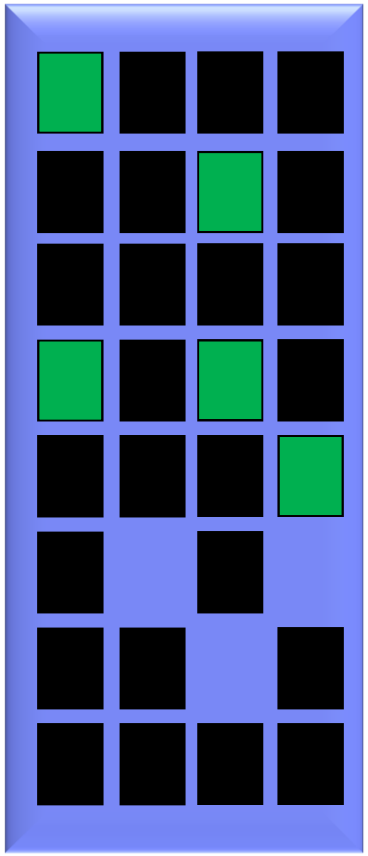
Destination



Guest Address Space

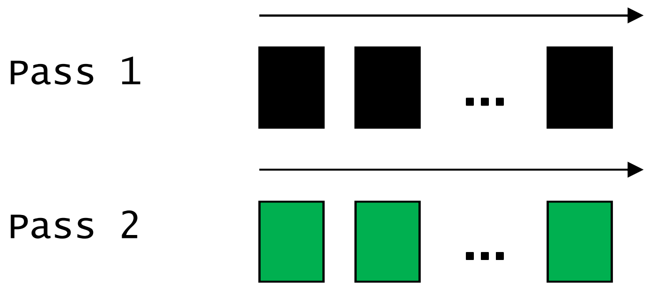
LGR, High-Level View of Memory Move

Source



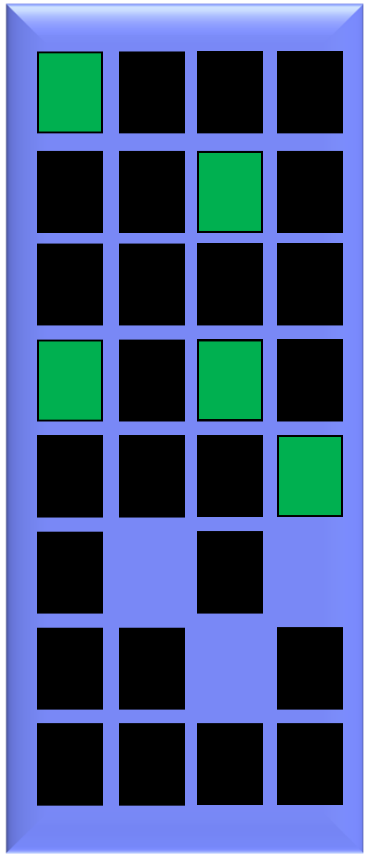
Guest Address Space

PUSH with resend



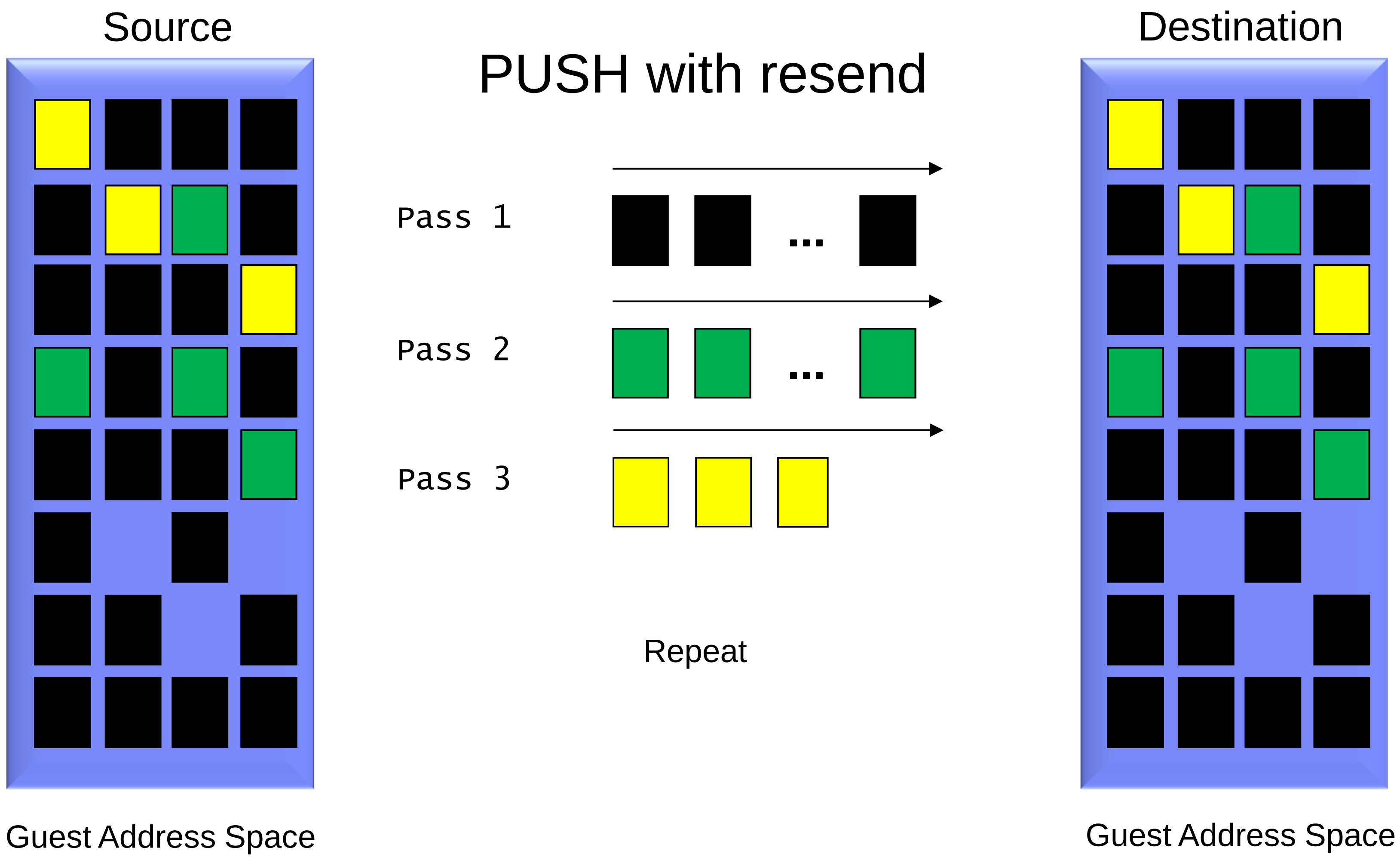
Walk through memory and resend any changed pages.

Destination

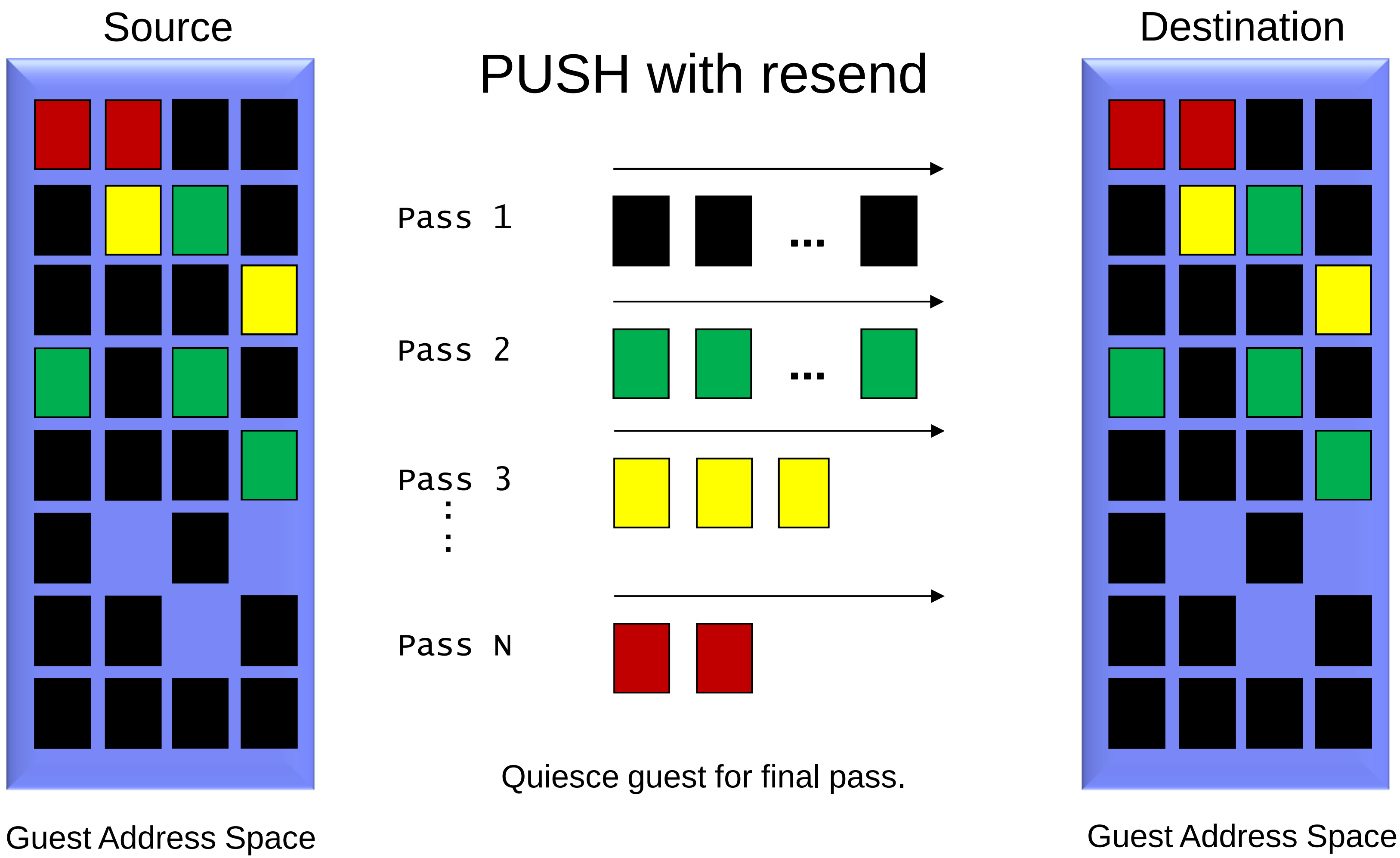


Guest Address Space

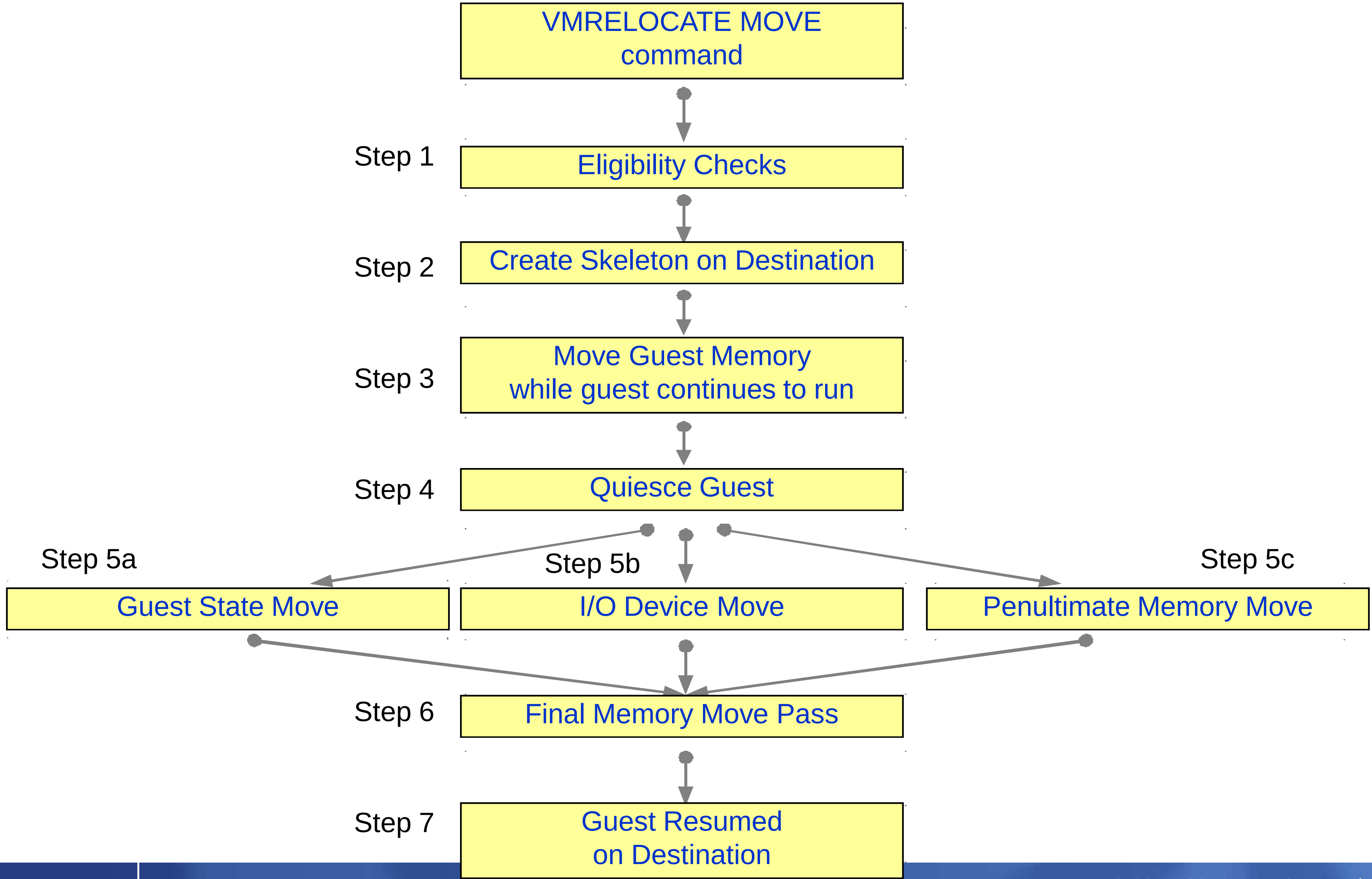
LGR, High-Level View of Memory Move



LGR, High-Level View of Memory Move



Stages of a Live Guest Relocation



Live Guest Relocation – Key Performance Metrics

- **Quiesce Time (QT)**
  - **Elapsed time that the guest is stopped (stunned) so z/VM can move the guest's last set of storage pages – probably the frequently-changed ones**
  - **To tolerate relocation, the guest and its applications must tolerate the quiesce time**
  - **VMRELOCATE can be invoked with a specified maximum quiesce time**
    - **If the quiesce would run past the maximum, z/VM cancels the relocation**

- **Relocation Time (RT)**
  - **Elapsed time from when the VMRELOCATE command is issued to when the guest is successfully restarted on the destination system.**
  - **Elapsed time must fit within the customer's window of time for planned outages for system maintenance, etc.**

Bottom line: there are some scenarios where LGR is not feasible as a result of the requirements for relocation time and quiesce time

LGR: Factors Affecting QT and RT

- **Size of the guest**
  - Amount of memory to move, time required to walk its DAT tables
- **How broadly or frequently the guest changes its pages**
  - It's an iterative memory push from source to destination
- **Time needed to relocate the guest's I/O configuration**
  - I/O device count, I/Os to quiesce, OSA recovery on target side
- **Capacity of the ISFC logical link**
  - Number of chpids, their speeds, number of RDEVs
- **Storage constraints on source and target systems**
- **Performance of paging subsystem**
- **Other work the systems are doing**
- **Other relocations happening concurrently with the one of interest**
- **Delays injected when LGR throttles itself back to prevent abends and other problems.**
  - End-to-end LGR throttling – triggered by paging intensities
  - Memory-move endpoint throttling – triggered by memory consumption
  - ISFC logical link throttling – triggered by ISFC running out of queued traffic buffers

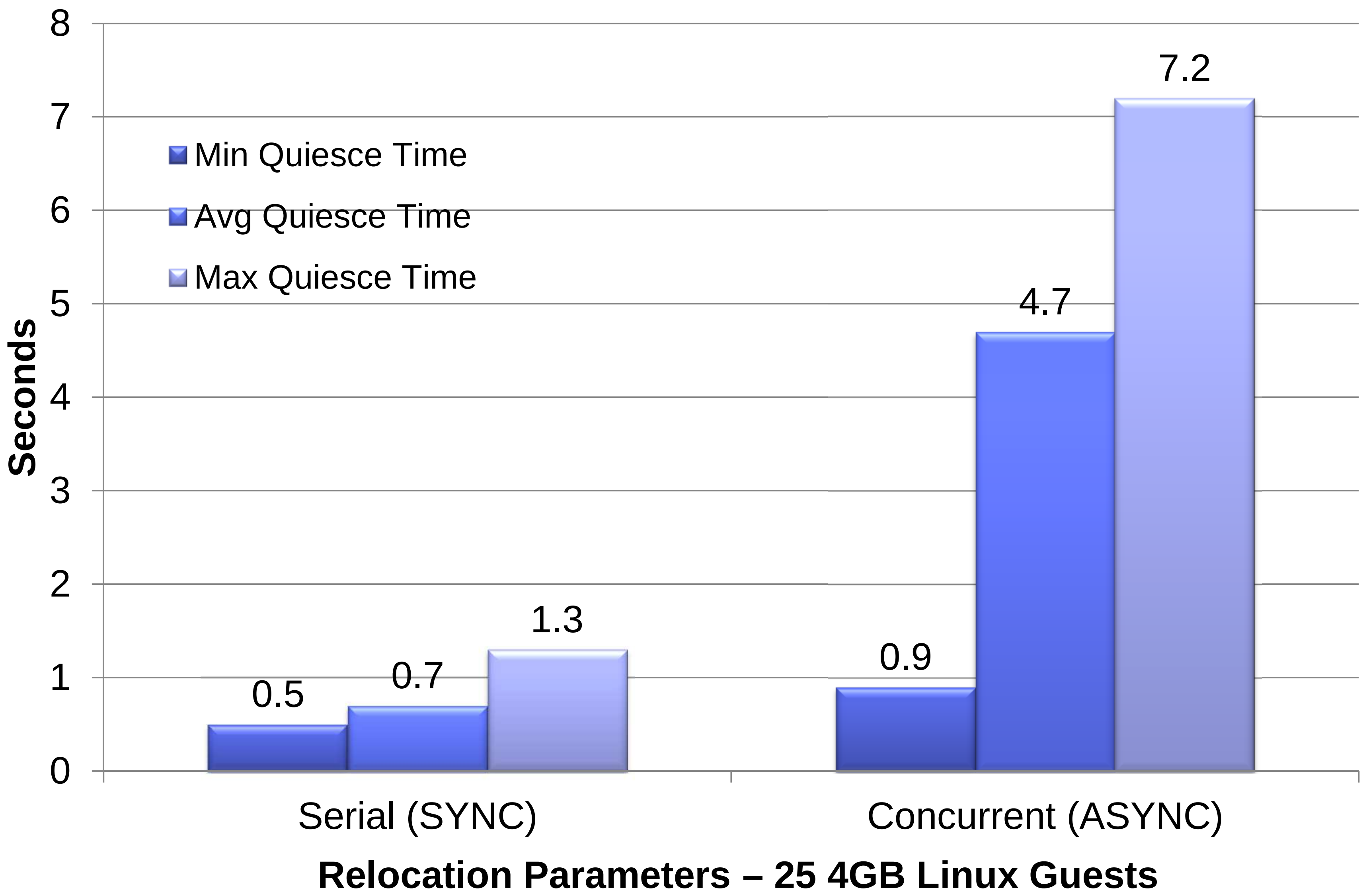
LGR: Serial vs. Concurrent Relocations

- **By default, the VMRELOCATE command operates synchronously.**
- **There is a command option (ASYNCH) to run it asynchronously (a la SPXTAPE)**
- **You could also achieve concurrent relocations by:**
  - Use the asynchronous version of VMRELOCATE multiple times.
  - Run VMRELOCATE commands in multiple users concurrently.

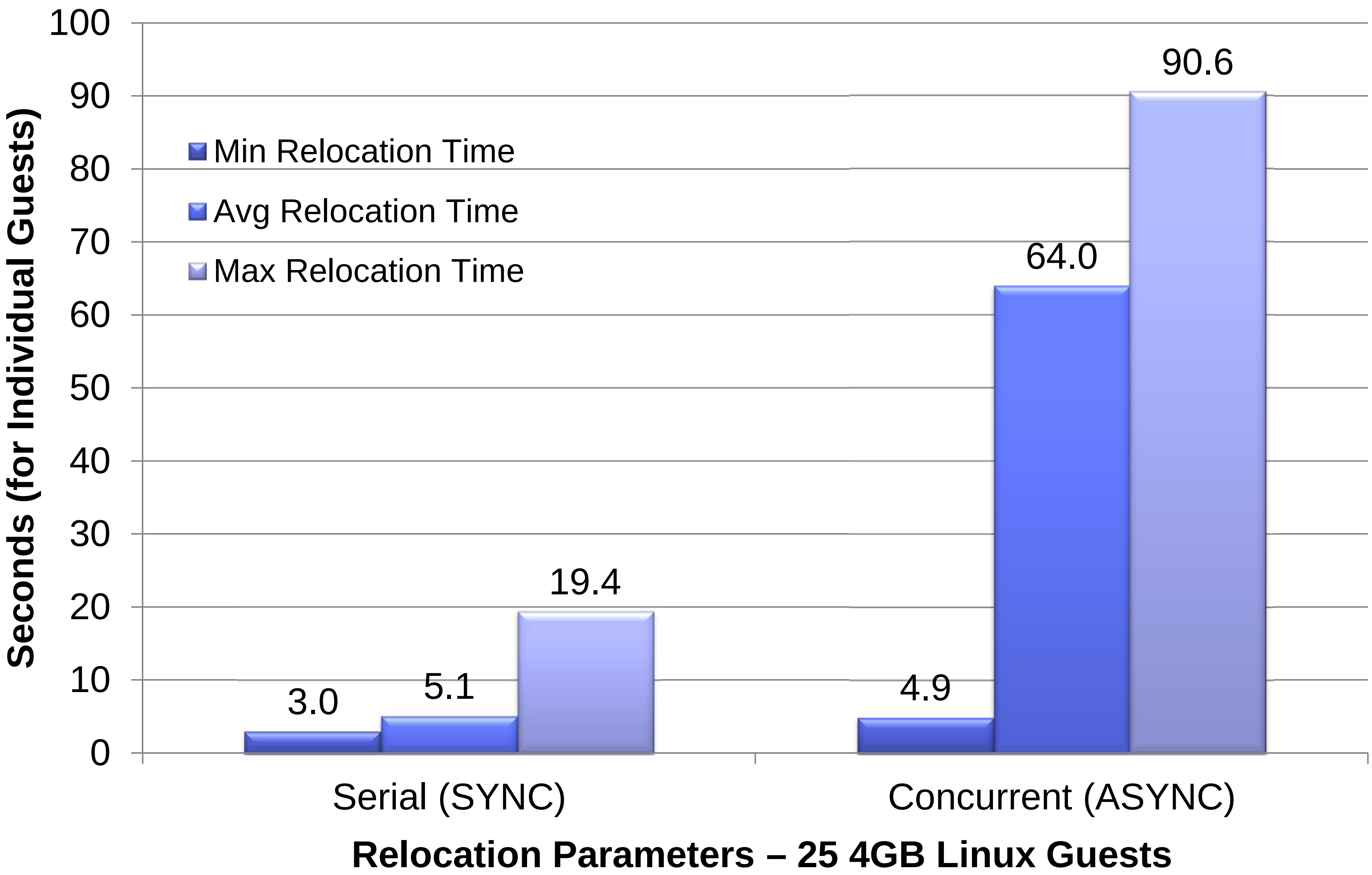
The best practice, though, is to run only one relocation at a time.

- **QT and individual RT improves substantially when relocations are done serially**
  - … and total RT elongates only slightly

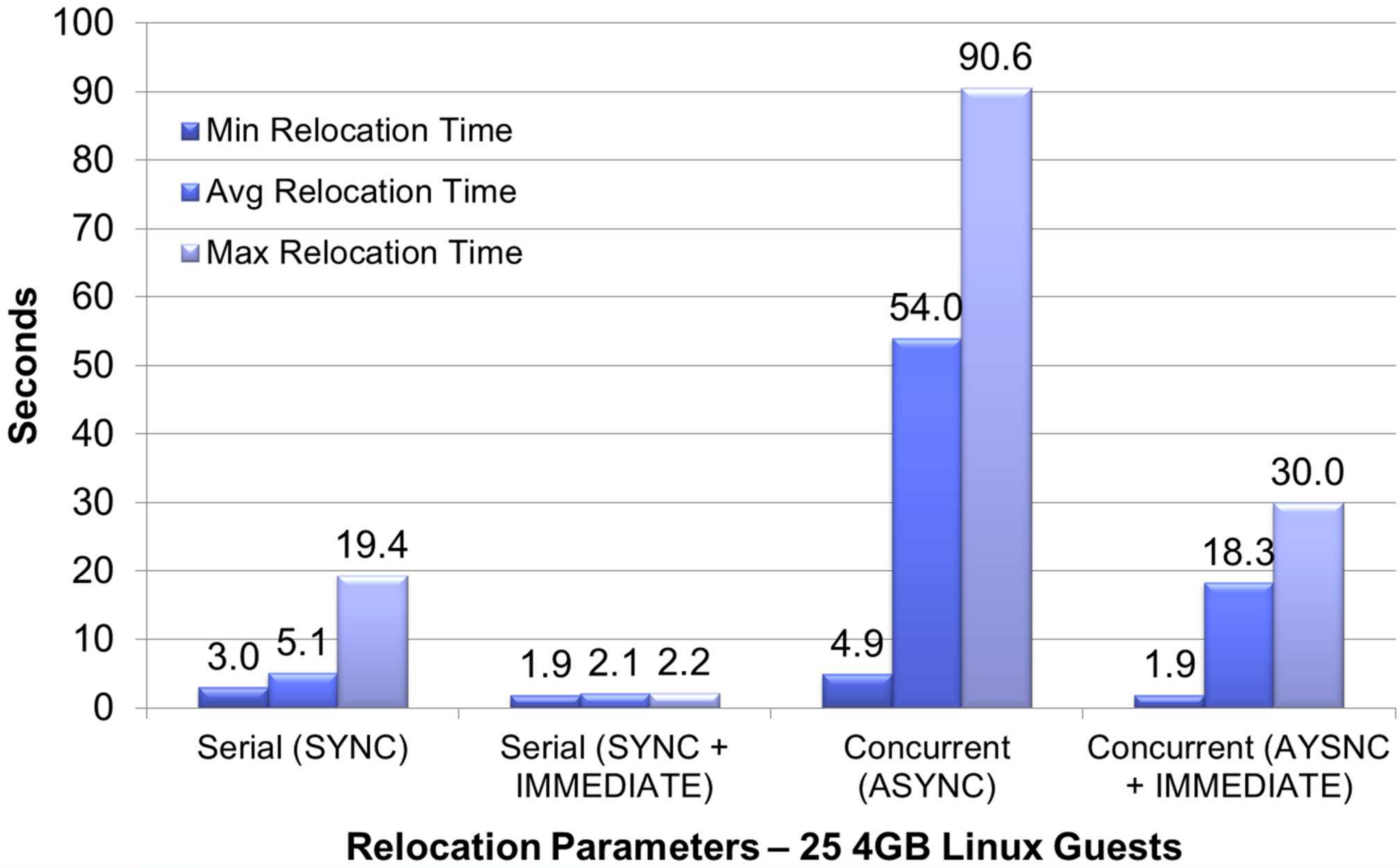
Effect of Serial vs. Concurrent on Quiesce Time



Effect of Serial vs. Concurrent on Relocation Time



Effect of IMMEDIATE option on Relocation Time



VMRELOCATE Options Summary

- **Best total relocation time for all virtual machines**
  - Concurrent (ASYNCH) + IMMEDIATE
- **Best individual relocation time**
  - Serial (SYNCH) + IMMEDIATE
- **Best quiesce times**
  - Serial (SYNCH)
- **Worst quiesce times**
  - Concurrent (ASYNC) + IMMEDIATE

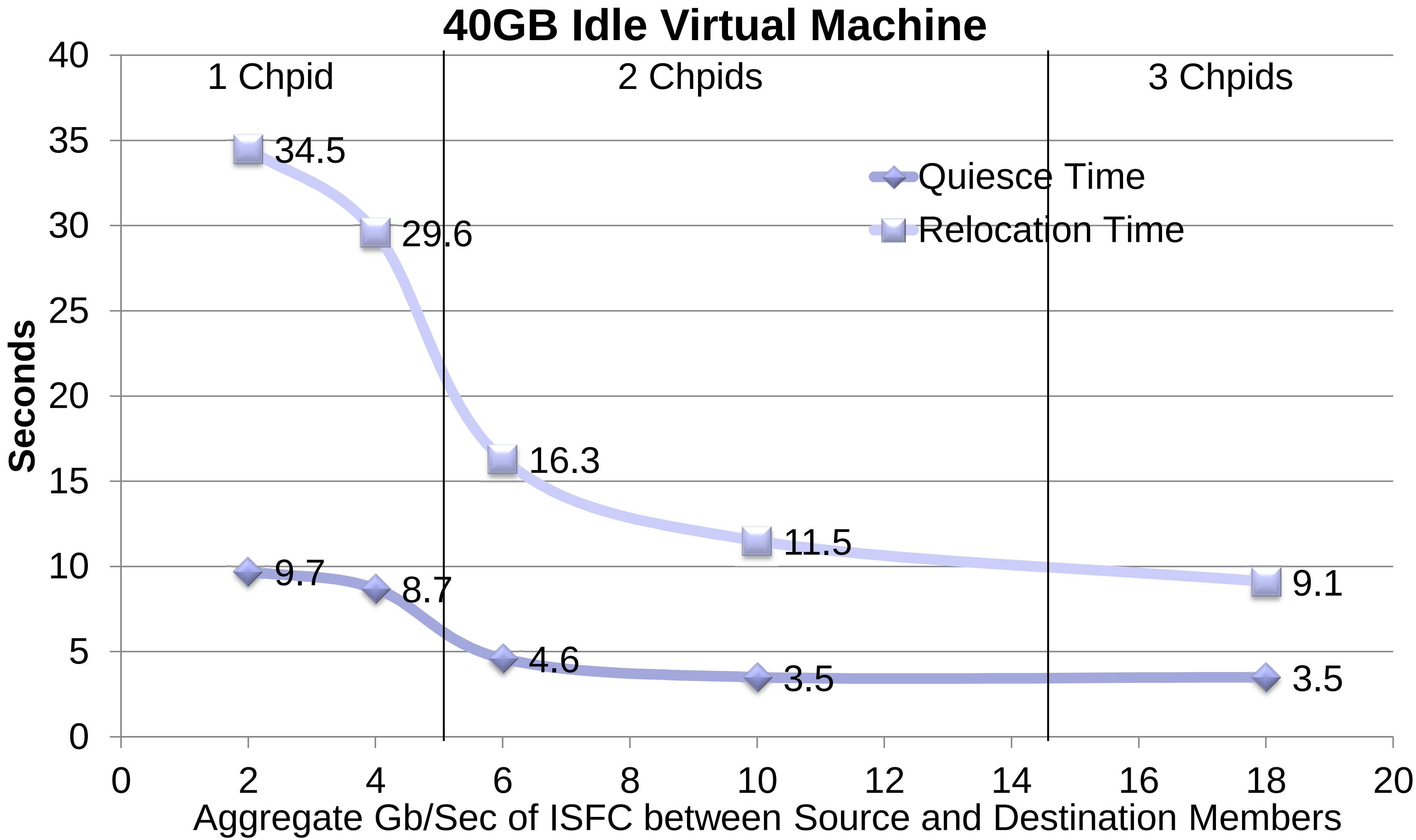
Background on ISFC Capacity Test

Table 3. Evaluated ISFC Logical Link Configurations.

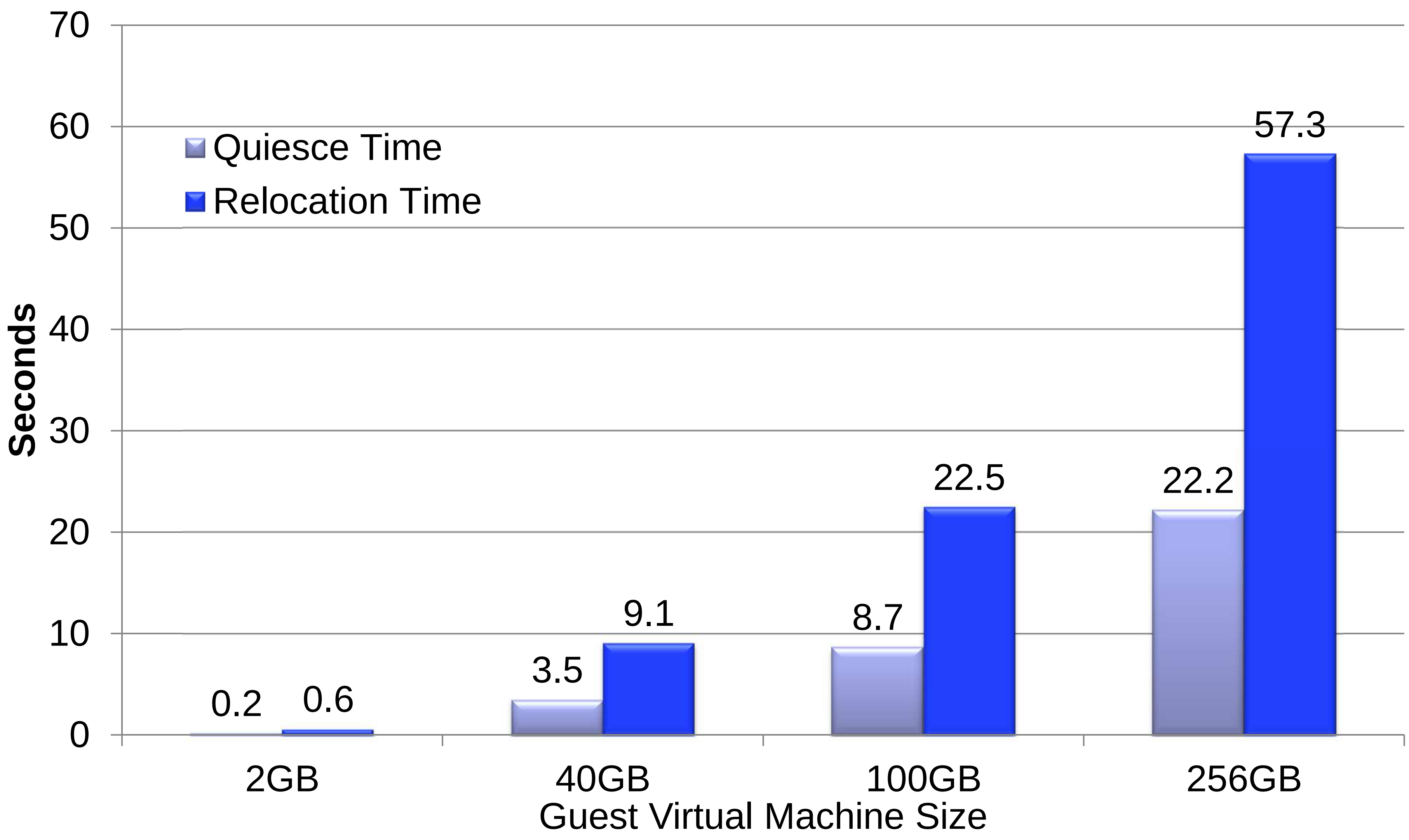
ISFC Logical Link CHPIDs	ISFC Capacity Factor *	CTCs/FICON CHPID	Total CTCs
1-2Gb, 2-4Gb, 1-8Gb	18	4	16
1-2Gb, 2-4Gb	10	4	12
1-2Gb, 1-4Gb	6	4	8
1-4Gb	4	4	4
1-2Gb	2	4	4

Note: * ISFC capacity factor is the sum of speeds of the FICON CTCs between the SSI member systems.

Effect of CTC Bandwidth on LGR

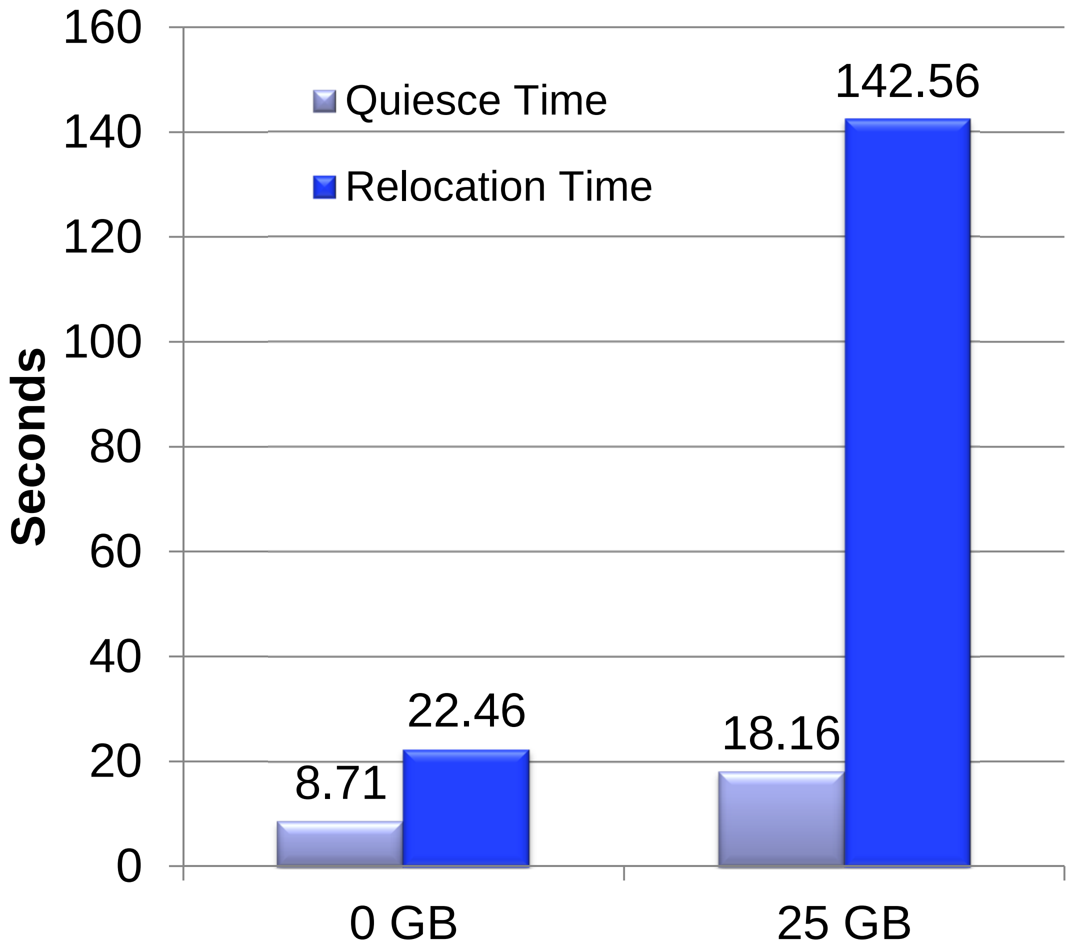


Effect of Virtual Machine Size on LGR



Impact of Virtual Machine Changing Memory on LGR

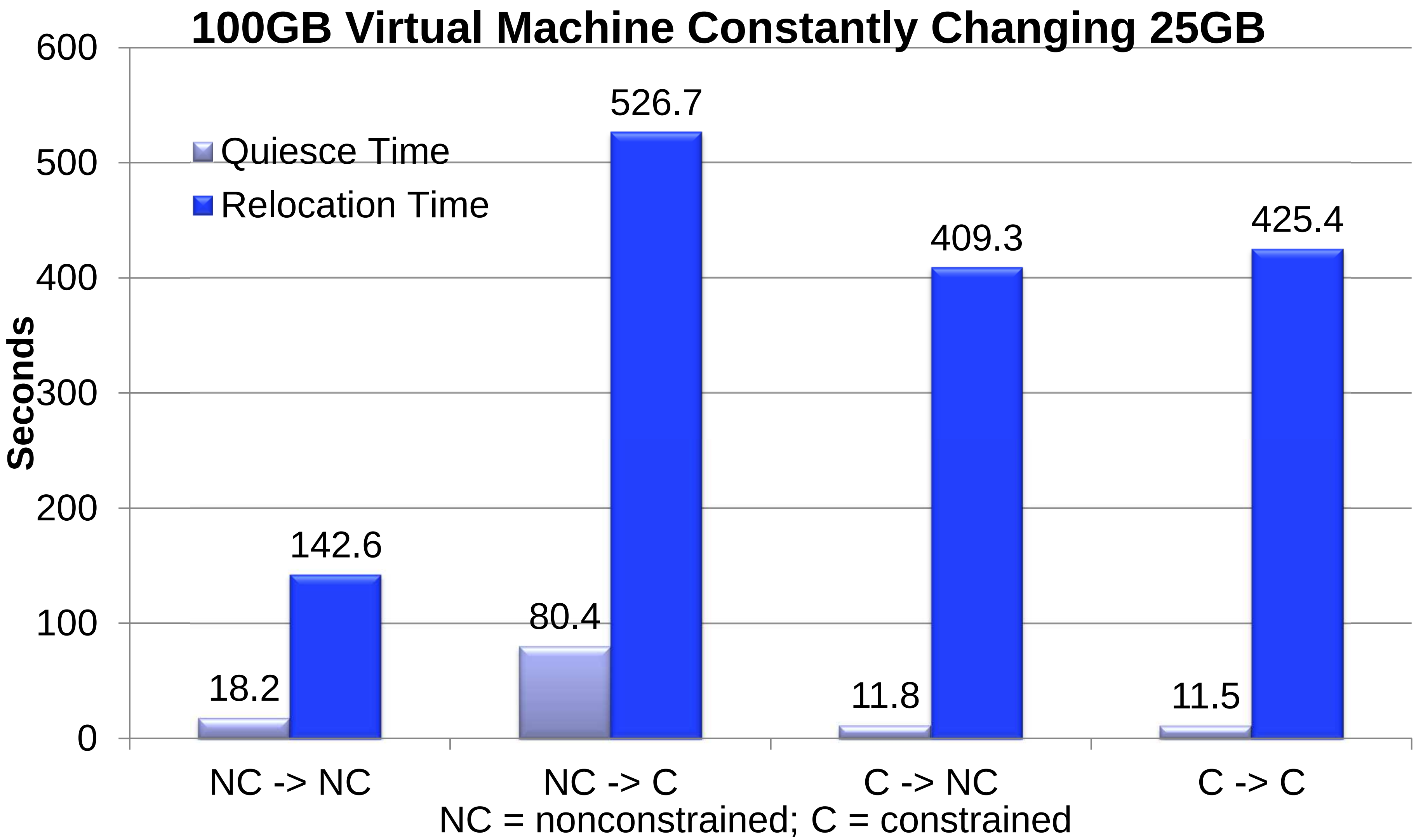
- **Idle case (0GB changing) there is less memory to move and fewer Memory Move Passes**
- **Number of Passes**
  - 0GB: 4
  - 25GB: 8
- **Total Memory Moved**
  - 0GB: 4.9GB
  - 25GB: 160GB



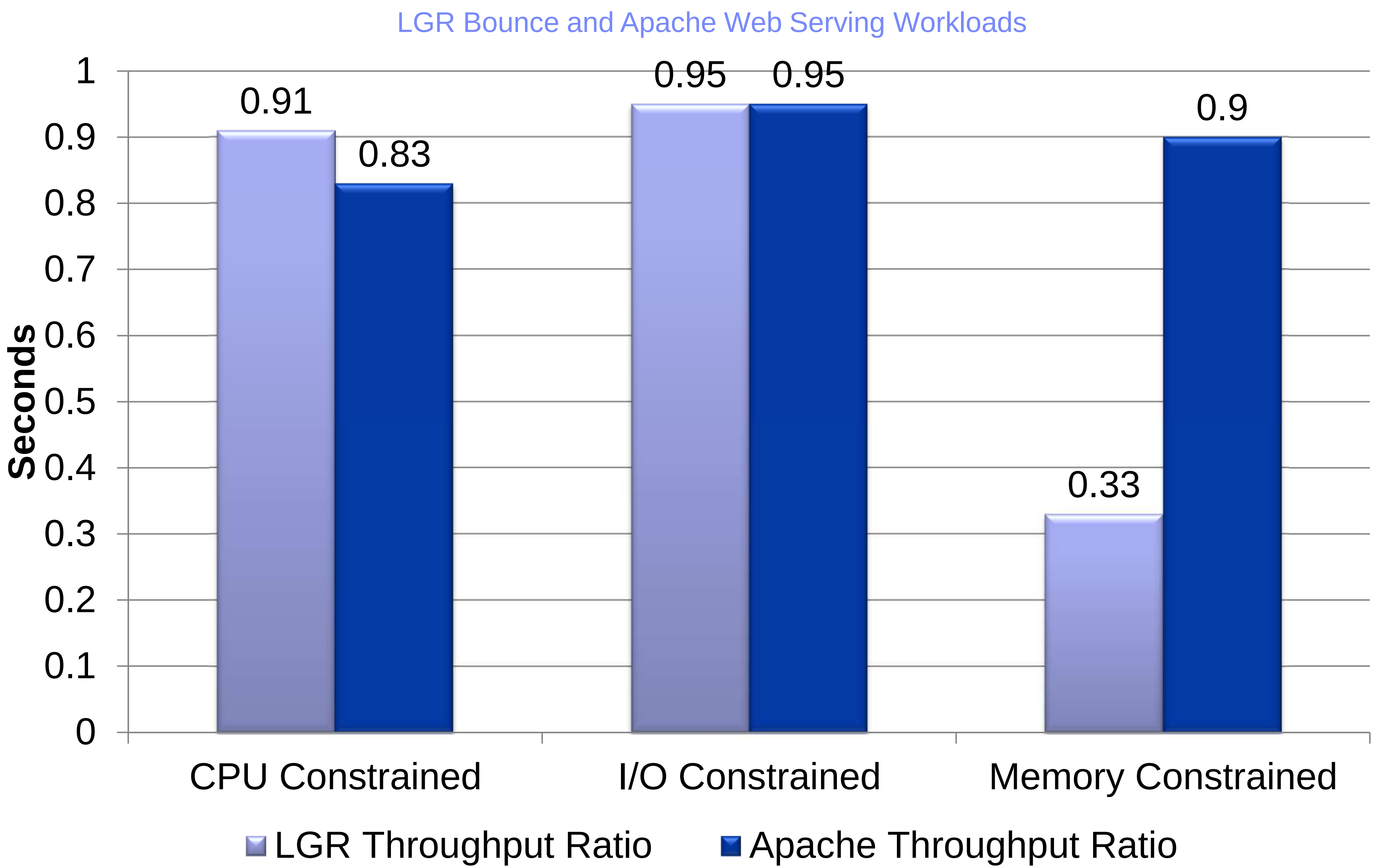
LGR: CPU and Memory Use Habits

- **CPU: generally LGR gets what it needs**
  - Taken "off the top" compared to your workload
- **Memory: CP tries really hard not to interfere**
  - End-to-end throttling, ISFC buffer limits, …
  - Socket memory-move throttling – triggered by memory consumption
  - ISFC logical link throttling – triggered by ISFC running out of queued traffic buffers
  - Considers effect on paging, memory use for specific relocations, …

Effect of System Memory Constraint on LGR



Effect of LGR on Existing Workloads



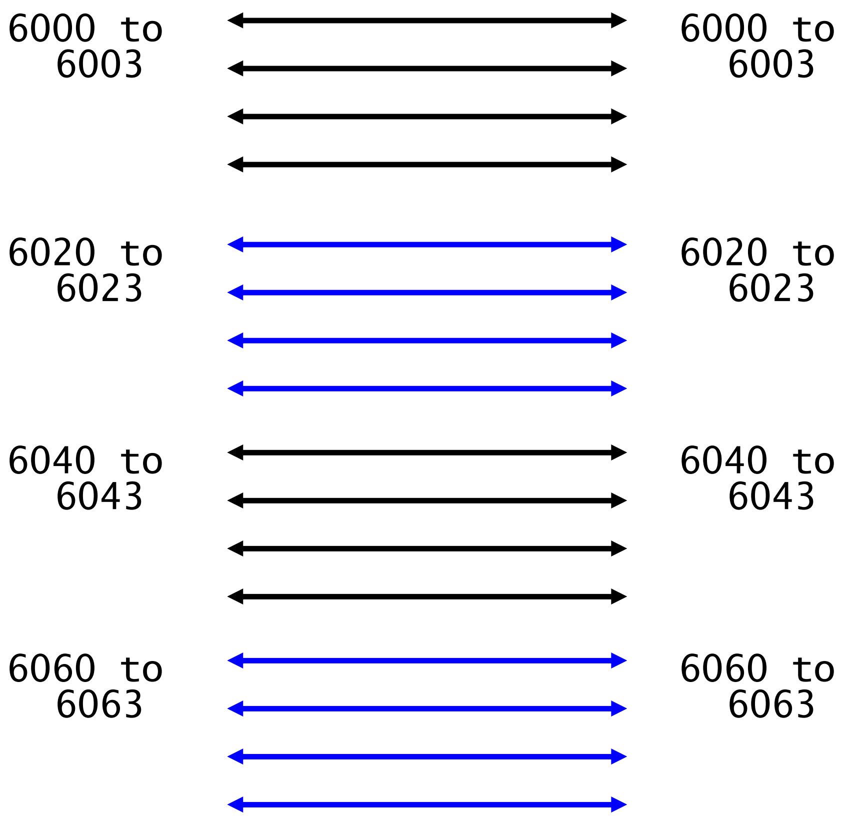
LGR: Keep These in Mind…

- **Charge back: can your procedures handle guests that suddenly disappear and then reappear somewhere else?**
- **Second-level schedulers: do you have them? Can they handle guest motion?**
- **VMRM: if VMRM-A tweaks the guest and then the guest moves to system B, what happens? And then what happens when the guest comes back?**

Best practice is not to include relocating guests in VMRM-managed groups.

SSI: ISFC Logical Link Configuration Best Practices

- **Use multiple FICON chpids of all the same speed. Up to 4 chpids.**
- **Use four CTC devices per chpid**
- **Use same RDEV numbers on both ends**
- **More esoteric configurations are certainly possible**
- **Can share the chpids but requires capacity planning**



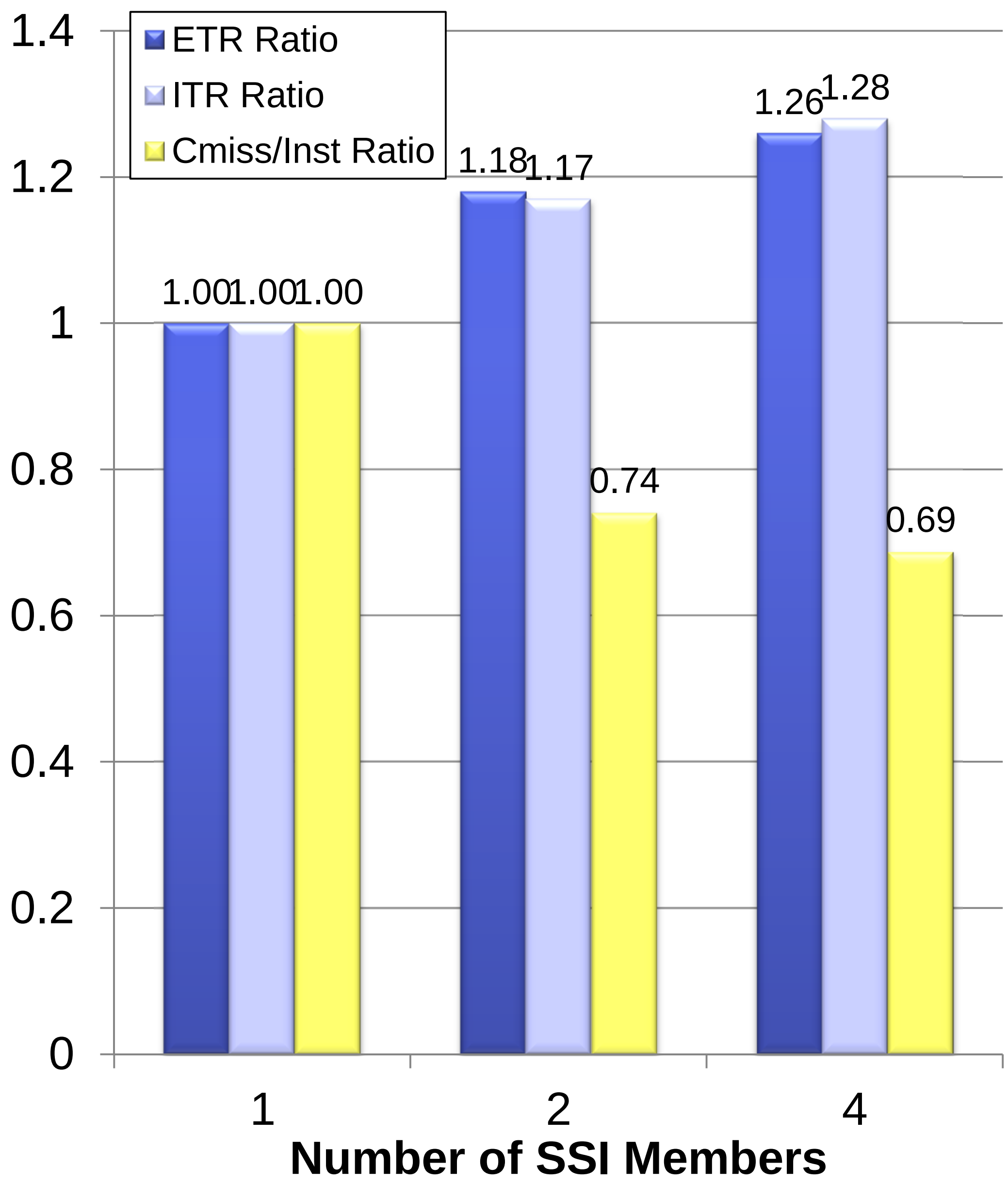
SSI Workload Distribution Measurements

Parameters	1 Member	2 Member	4 Member
Central Storage	43 GB	22 GB	11 GB
Expanded Storage	8 GB	4 GB	2 GB
Processors	12	6	3

- Series of measurements to see how a workload spread across a number of members would run compared to one larger systems of just one member.
- Resources kept the same, as shown above.
- Apache workload where clients and servers were all virtual machines was used.
  - Varied number of client and servers and use of MDC to create different stress points.

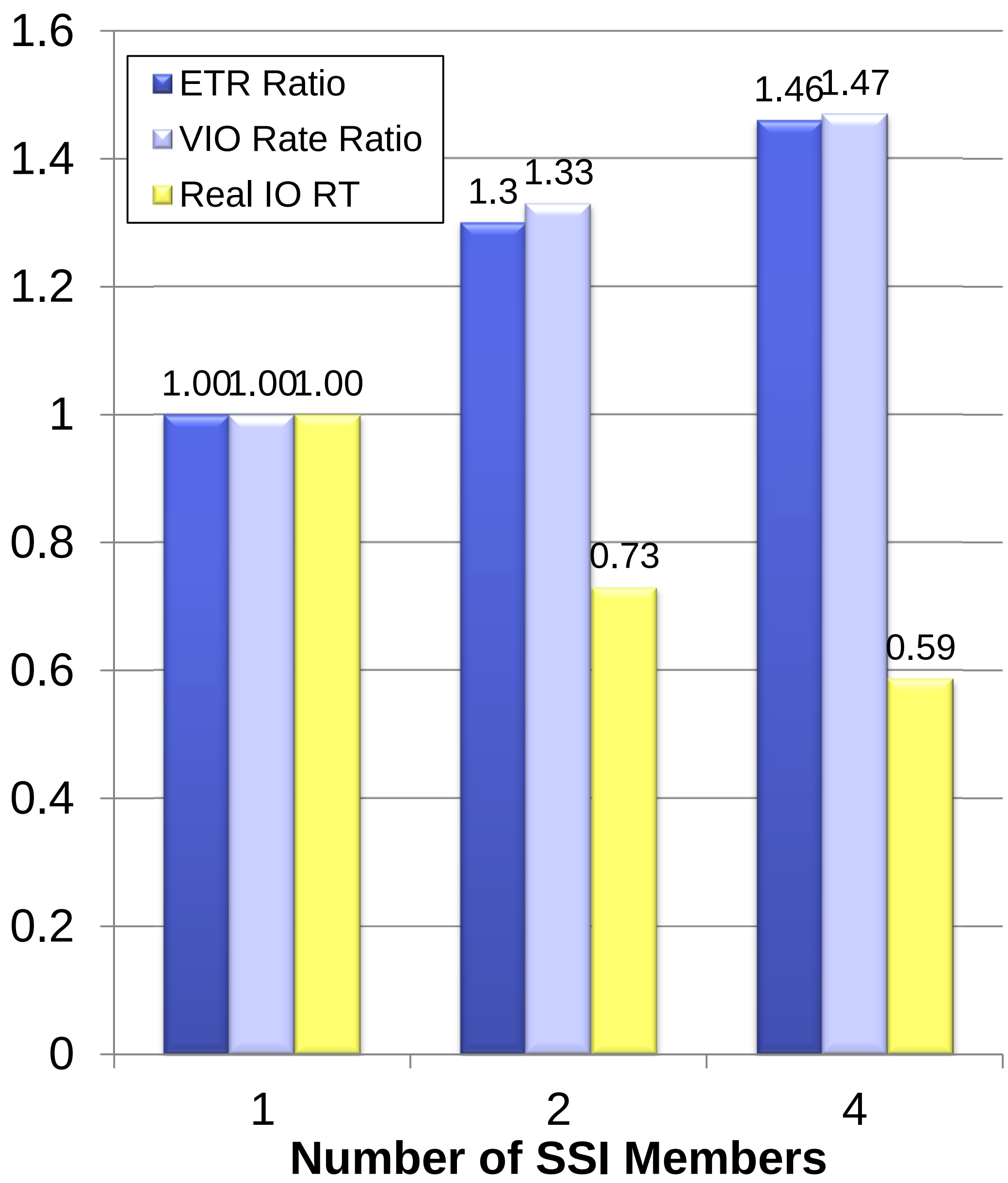
SSI Distribution: CPU Constrained Measurement

- **Keep the physical resources the same, but distribute over 1, 2, or 4 members.**
- **Apache Web Serving with the configuration being CPU bound.**
- **Benefits from running smaller n-way partitions**



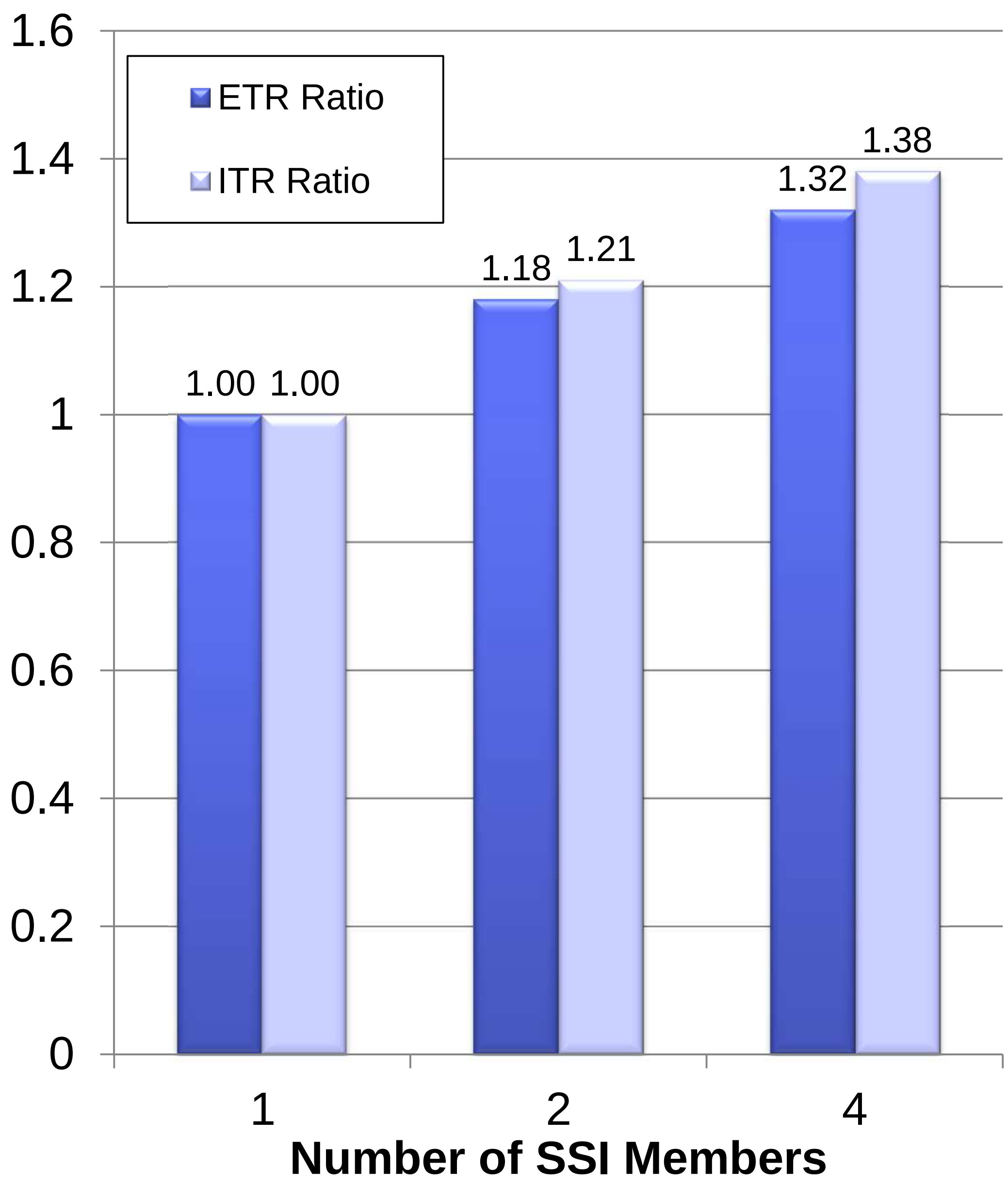
SSI Distribution: Virtual I/O Constrained Measurement

- **Keep the physical resources the same, but distribute over 1, 2, or 4 members.**
- **Apache Web Serving with the configuration being I/O bound due to virtual read I/O.**
- **PAV not used in base case, so SSI essentially gives PAV like benefits.**
- **Real I/O RT shown is for one of the shared Linux volumes containing files being served.**



SSI Distribution: Memory Constrained Measurement

- **Keep the physical resources the same, but distribute over 1, 2, or 4 members.**
- **Apache Web Serving with the configuration with there being memory constraint.**
- **Similar savings as in CPU bound measurement.**
- **Additional efficiencies in memory management.**



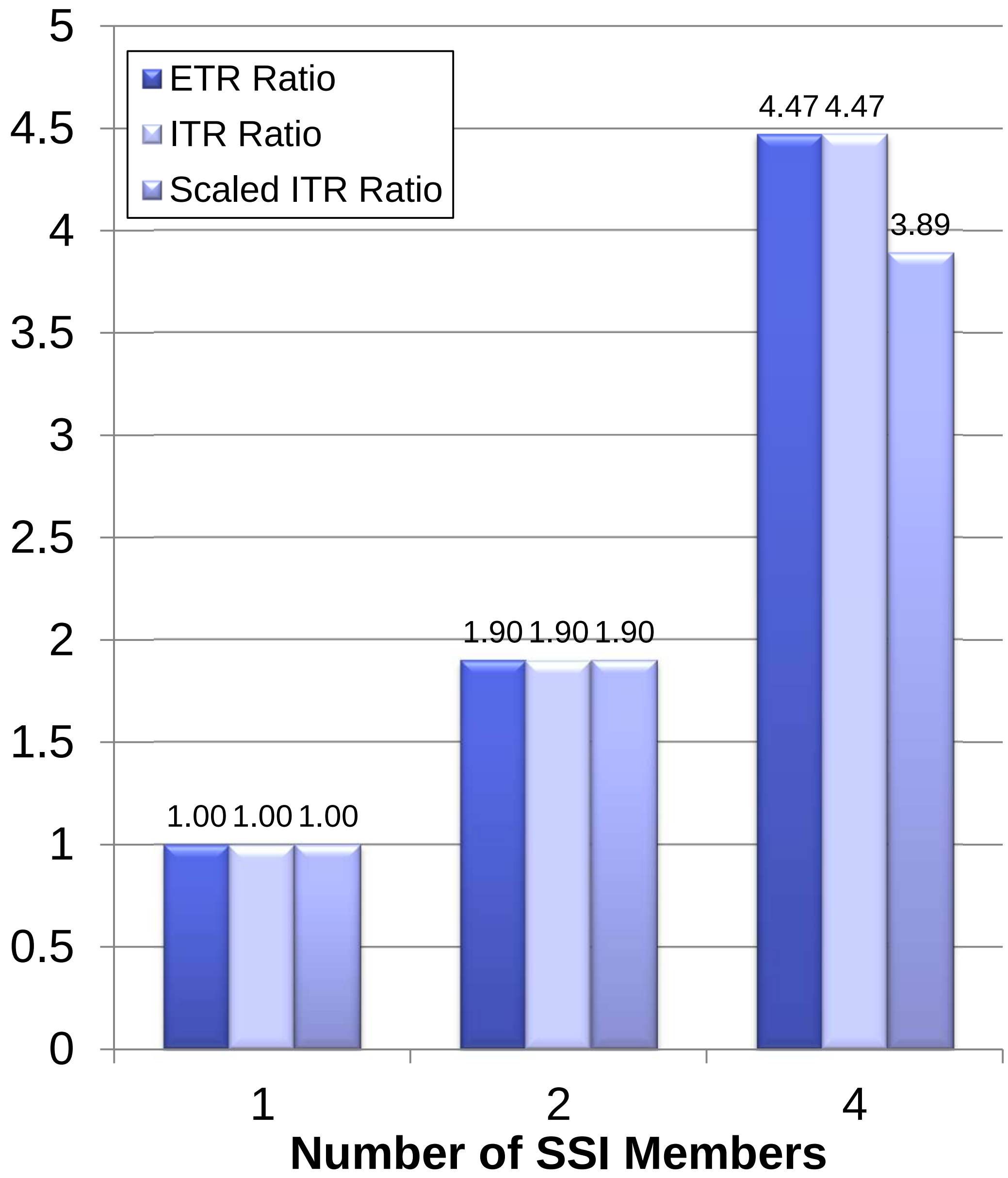
SSI Workload Scaling Measurements

z/VM Limits	1 Member	2 Member	4 Member
Central Storage	256 GB	512 GB	1 TB
IFLs	32	64	128

- Measurements were made to see how well z/VM scales within an SSI cluster.
- Resources increased with each new member added to configuration.
- Apache workload where clients and servers were all virtual machines was used.
  - Apache clients and servers scaled accordingly.
- Needed to mix processor types to get 128 IFLs, so 1 & 2 Member runs are z10, 4 member adds in z196.
- Scaled down memory to make runs more feasible.

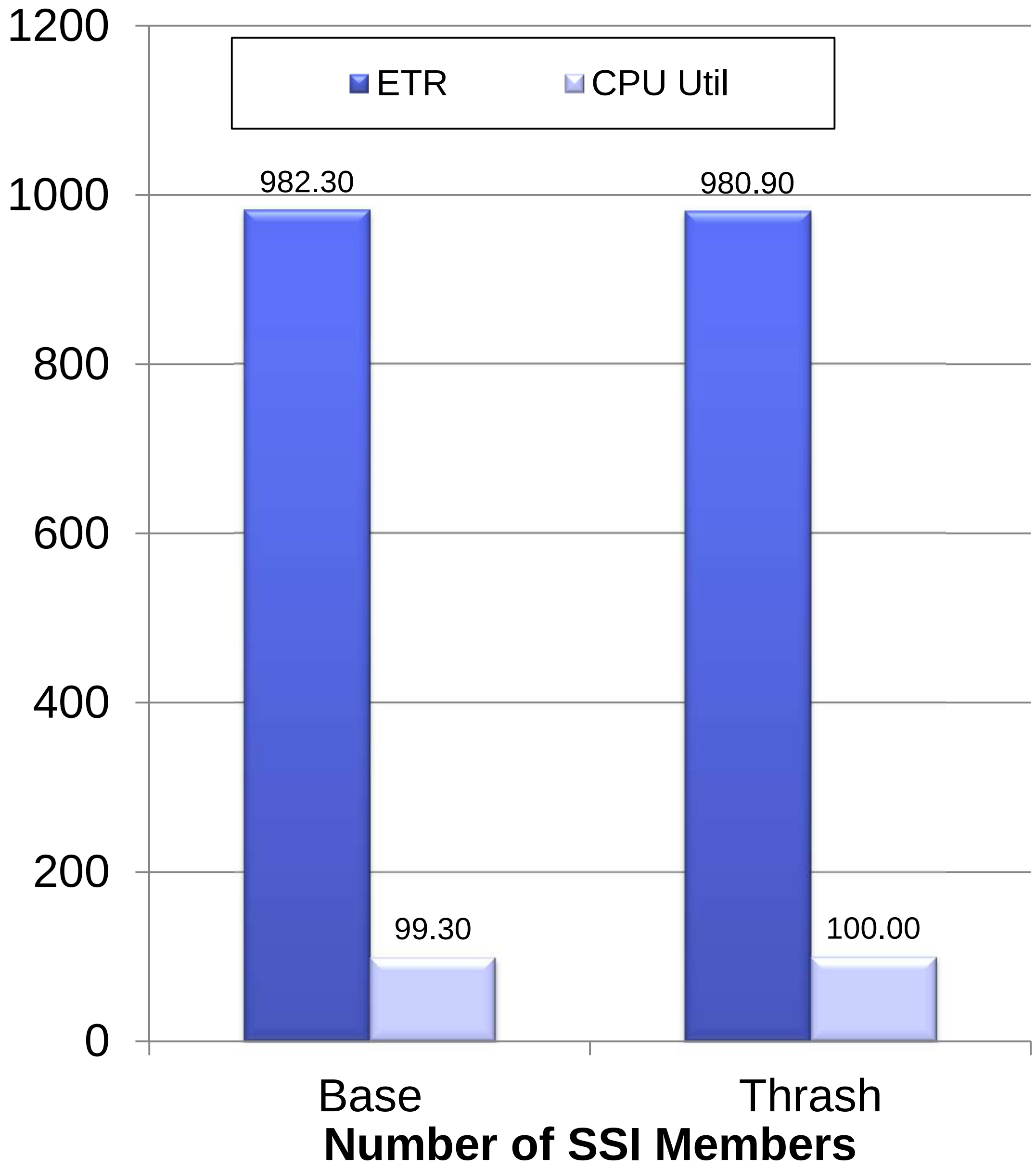
SSI Scaling Measurements

- **The SSI Cluster overhead for a running environment is very low.**
- **Note: z196s were added to get the 3rd and 4th Member.**
- **"Scaled ITR Ratio is an estimate of the Ratio if the entire cluster were on z10 processors.**

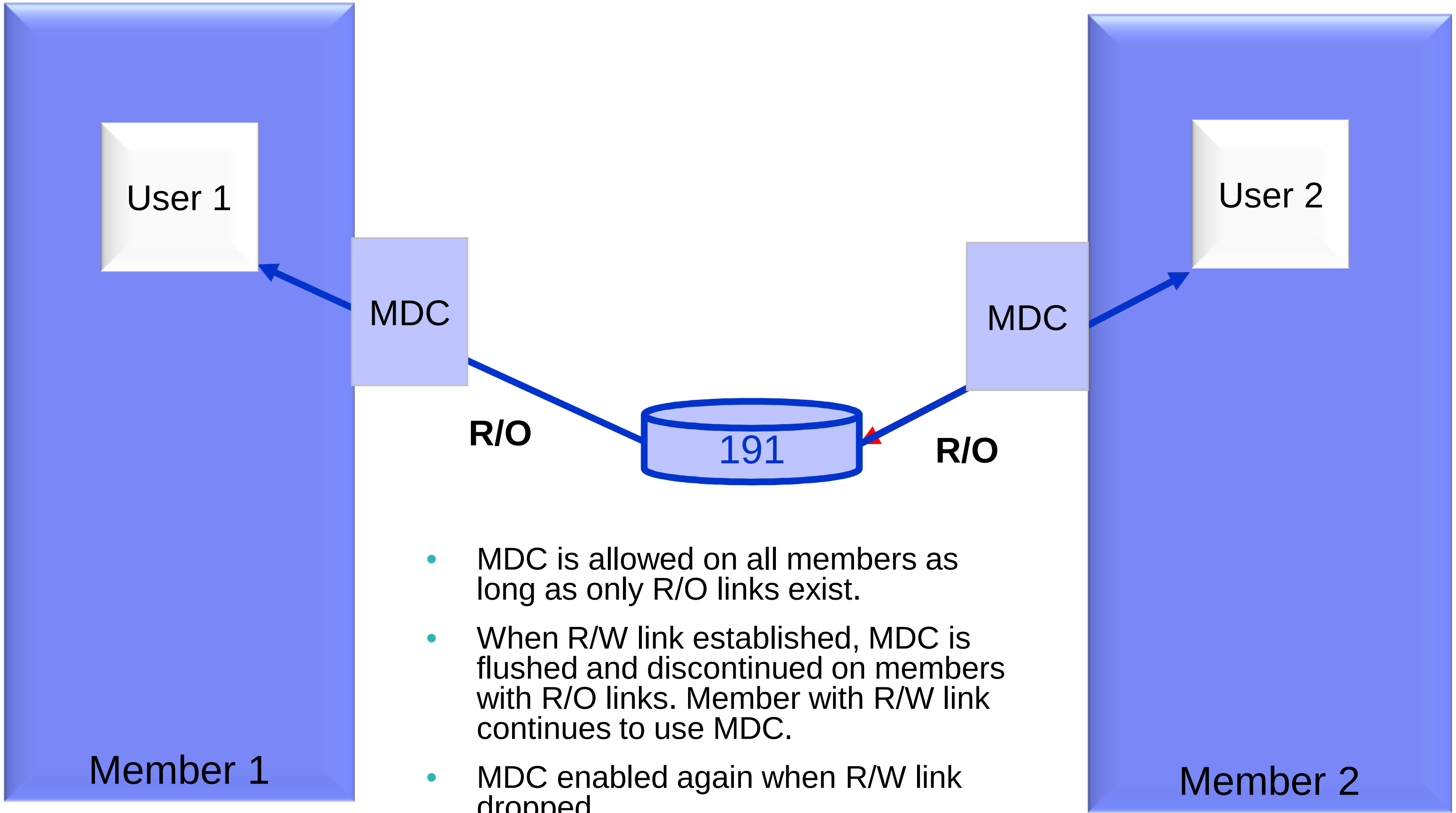


SSI Transition Measurement

- **Measurement to determine if activity or Cluster management would influence performance.**
- **Four Member environment where 3 of the members are constantly transitioning through states:**
  - Joined
  - Leaving
  - Down
  - Joining
  - *repeat*

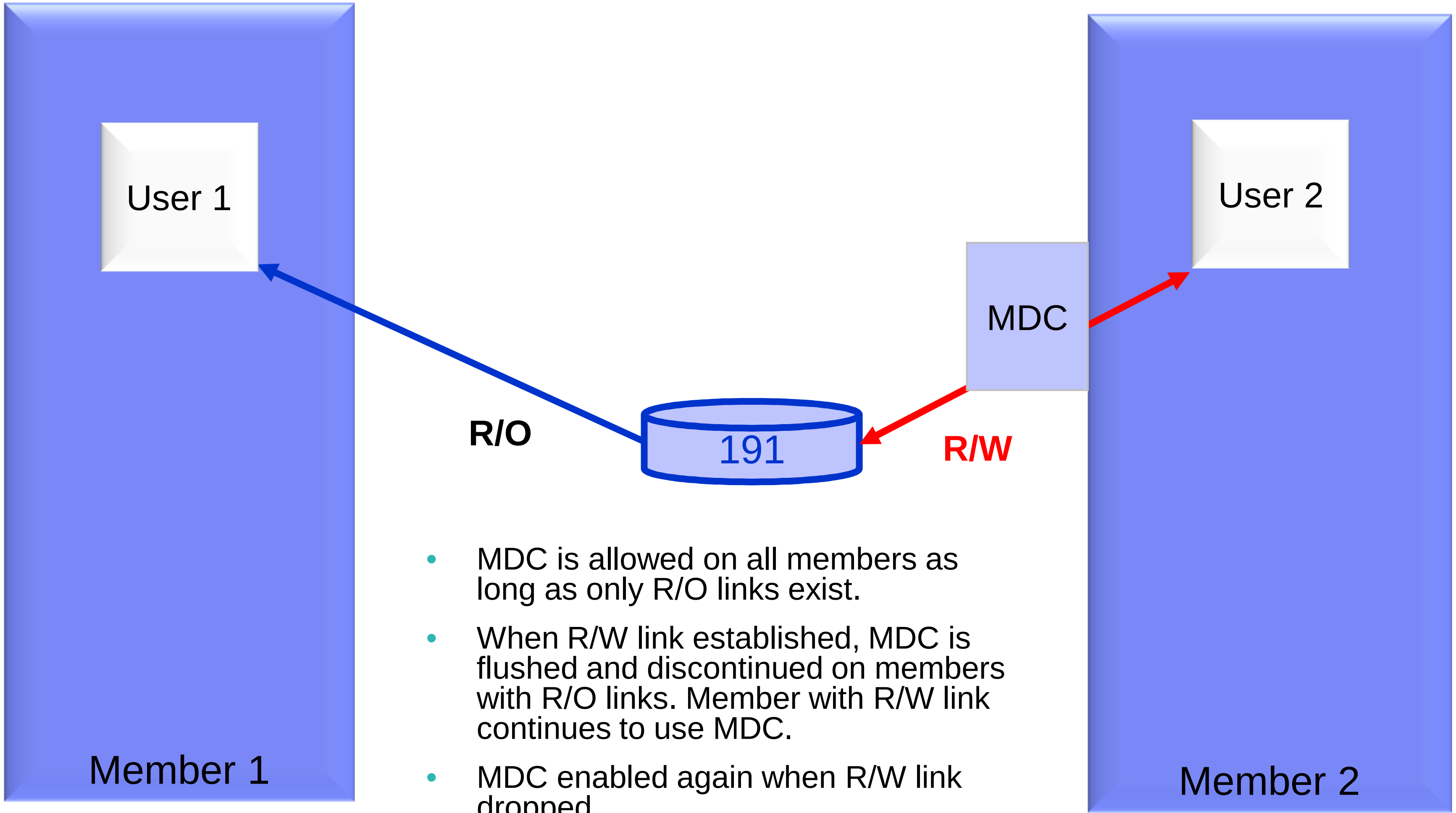


SSI: Automatic MDC Management



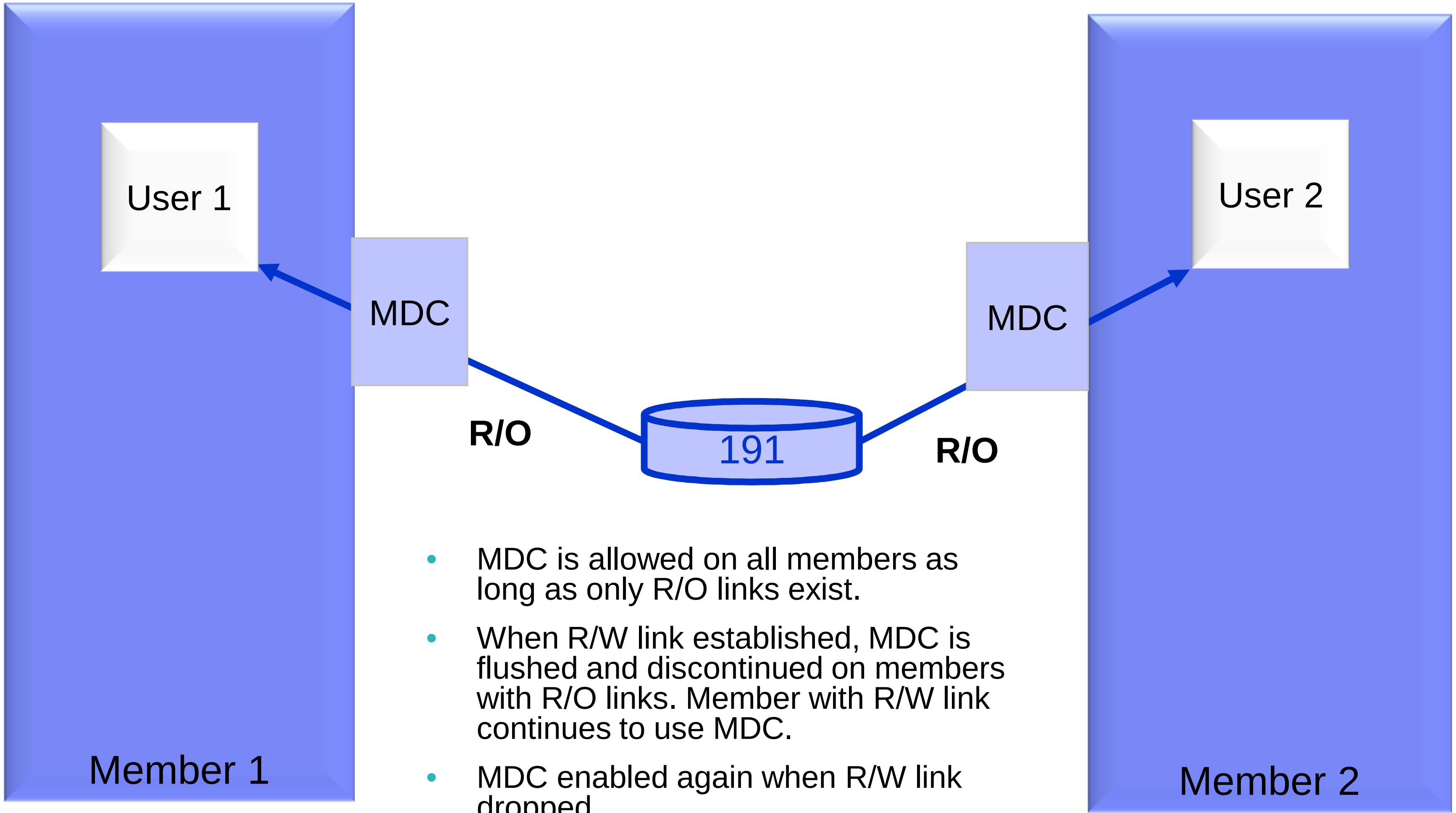
- MDC is allowed on all members as long as only R/O links exist.
- When R/W link established, MDC is flushed and discontinued on members with R/O links. Member with R/W link continues to use MDC.
- MDC enabled again when R/W link dropped.

SSI: Automatic MDC Management



- MDC is allowed on all members as long as only R/O links exist.
- When R/W link established, MDC is flushed and discontinued on members with R/O links. Member with R/W link continues to use MDC.
- MDC enabled again when R/W link dropped.

SSI: Automatic MDC Management



- MDC is allowed on all members as long as only R/O links exist.
- When R/W link established, MDC is flushed and discontinued on members with R/O links. Member with R/W link continues to use MDC.
- MDC enabled again when R/W link dropped.

SSI: Performance Toolkit, Considerations

- **Performance Toolkit continues to run separately on each member of the cluster**
  - There continues to be a unique z/VM monitor data stream for each member.
  - There will be a PERFSVM virtual machine on each member
- **Configuration and usage**
  - Configure so that you will log onto or connect to a different PERFSVM on each system.
  - Configure Performance Toolkit to use the Remote Performance Monitoring Facility, which allows local and remote performance monitoring from a single screen.
- **In general, Performance Toolkit does not produce "cluster view" reports**
  - DASD device-busy view, for example

SSI: Performance Toolkit, New Reports

- **New Reports for SSI**
  - SSICONF: SSI configuration
  - SSISCHLG: SSI state change synchronization activity log
  - SSISMILG: SSI state/mode information log

- **New ISFC reports related to SSI**
  - ISFECONF: ISFC end point configuration
  - ISFEACT: ISFC end point activity
  - ISFLCONF: ISFC logical link configuration
  - ISFLACT: ISFC logical link activity
  - ISFLALOG: ISFC logical link activity log

SSI: MONWRITE Considerations

- **IBM often asks you to run MONWRITE**
  – PMR diagnosis, for example

- **You should be running MONWRITE anyway**

- **You should now be running MONWRITE on every member of the cluster**

- **Make sure it's easy to go find the MONWRITE data for all members for a specified time interval**

z/VM 6.2

More Than Just SSI and LGR

Memory Management: Needle-in-Haystack Searches

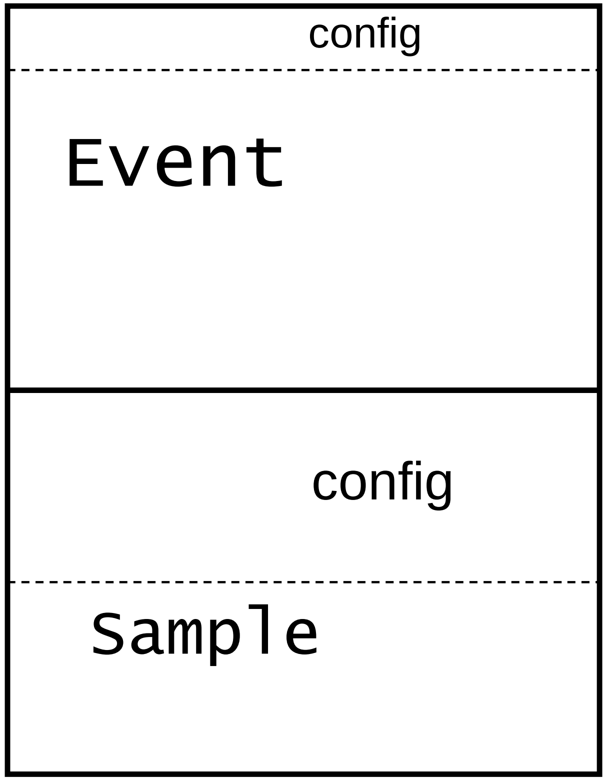
- **Searching for a below-2-GB frame in lists dominated by above-2-GB frames**
  - In months of study we identified about 10 of these searches
  - Development prototype that shut off all unnecessary use of <2GB storage gave us tremendous results
- **z/VM now does not allocate pageable buffers <2GB if:**
  - Dynamically, usable >2GB to usable <2GB is beyond a certain threshold
  - Statically, if the partition is beyond a certain size, for the life of the IPL
- **Result: no more needle searches**
- **Practically speaking, systems with 128 GB or more of real memory use below-2-GB memory only when it is architecturally required.**

MONDCSS and SAMPLE CONFIG Changes

- **The old defaults are too small for most systems nowadays**
- **So we have changed the default layout**
- **MONDCSS is 64 MB now (16384 pages)**
  - Half (32 MB) for EVENT
  - Half (32 MB) for SAMPLE
    - Half (16 MB) for SAMPLE CONFIG
- **As before, empty pages are not instantiated**
- **Remember, config pages evaporate after a short time**
- **MONWRITE 191 disk also increased to 300 cylinders.**

If you use your own MONDCSS, the new default SAMPLE CONFIG size may be too large, requiring you to set it manually or to change your MONDCSS.

MONDCSS – 16384 pages



Default STORBUF Changes

- **Many parties were noticing that the old defaults of 125 105 95 were not appropriate for Linux workloads**
- **We considered several different proposals**
  - From IBM ATS
  - From vendors
  - From Redbooks
  - From customer data
- **After careful consideration by "top people" we came to 300 250 200 as new defaults**

- If you already override defaults, the only impact would be if you also use SET SRM STORBUF INITIAL at some point.
- For CMS-intensive workloads, the old defaults might be more appropriate, and you should validate the settings for these workloads when you migrate to z/VM 6.2

z/CMS

- **Prior to z/VM 6.2, z/CMS was supplied as a sample.**
- **z/VM 6.2 supports z/CMS as an optional alternative to the standard CMS that runs in ESA and XC mode virtual machines and 31-bit addressing.**
- **z/CMS can run in a z/Architecture guest**
  - Allows programs to use z/Architecture instructions, including 64-bit addressing
- **Standard CMS function does not exploit memory above 2GB**
- **Remember that z/Architecture is not XC**
  - No VM Data Spaces
  - No SFS DIRCONTROL-in-data-space
  - No DB/2-for-VM data space use
- **The standard, usual, XC-mode CMS is still there**

CPU Measurement Facility Counters

- **CPU MF counters are a System z hardware facility that characterizes the performance of the CPU and nest**
  - Instructions, cycles, cache misses, and other processor related information
- **Available on z10 EC/BC, z196, and z114**
- **The CPU MF counter values:**
  - Help IBM to understand how your workload stresses a CEC for future design
  - Help IBM to map your workload into the LSPR curves for better sizing results
  - Help IBM better understand your system when there is a processor performance related problem.
- **z/VM 6.2, 6.1, and 5.4 can all collect the CPU MF counters from the hardware**
  - z/VM 5.4 and 6.1: VM64961, UM33440 (5.4), UM33442 (6.1)
  - Counters are put in new z/VM monitor record
- **We want volunteers to send us MONWRITE data!**
  - Your contributions will help us to understand customer workloads!

CPU MF Counters and CP Monitor, Details

- **Counter sample record is in the Processor domain**

- **MONITOR SAMPLE command manipulates counter collection**

- **QUERY MONITOR reveals whether counter collection is on**

- **z/VM writes the collected counters into the Monitor data stream**
  - Domain 5 Record 13: MRPRCMFC, Processor domain, sample record

- **The D5 R13 records land in your MONWRITE data**

IBM Wants Your CPU MF Counter Data

- **Your data will help IBM to build a library of customer workloads**
- **Collect an hour's worth of MONWRITE data…**
  - From a peak period,
  - With CPU MF counters enabled,
  - With one-minute sample intervals
- **Contact Richard Lewis at rflewis at us.ibm.com**
- **Richard will send you instructions on how to transmit the data to IBM**
- **No deliverable will be returned to you**
- **We will be ever grateful for your contribution**

Monitor Records – Highlights – New and Almost-New

- **In domain 1 (monitor), ISFC and SSI config records**
- **In domain 1, system topology record (PU-book-chip)**
- **In domain 4 (user), LGR start and LGR end**
- **In domain 5 (processor), CPU MF and system topology**
- **In domain 6 (I/O), minidisk MDC setting change event**
- **New domain 9 – ISFC performance records**
- **New domain 11 – SSI performance records**
  - On by default if running in an SSI cluster.

- **Other changes to report on LGR, mostly in user domain**

z/VM 6.2 Monitor Changes

- **Virtual Machine High Frequency State Sampling**
  - Corrected scenario being marked as “Other” state in a virtual MP configuration when the base VMDBK (virtual CPU) is actually idle but held in the dispatch list due to another virtual CPU in configuration is in dispatch or eligible list. Now more appropriately marked as in an idle state.

z196 and z114 Support for Energy Savings

- **Processor performance (capability) can change due to over heating condition or static energy savings mode.**
- **Reflected in monitor data and QUERY CAPABILITY command.**

```
Response (may only get first line on system with no changes):
CAPABILITY: PRIMARY 696      SECONDARY 696      NOMINAL 696
CAPACITY-ADJUSTMENT INDICATION 100   CAPACITY-CHANGE REASON 0
RUNNING AT NOMINAL CAPACITY.
```

```
Response for static power savings mode:
RUNNING WITH REDUCED CAPACITY DUE TO A MANUAL CONTROL SETTING.
```

```
Response possible for ambient temperature exceeded specified maximum:
RUNNING WITH REDUCED CAPACITY DUE TO AN EXTERNAL EXCEPTION CONDITION.
```

z/VM 6.2: Service Integrated in Base of z/VM 6.2

- **VM64774 SET/QUERY REORDER command**
- **All of the SSL scaling fixes**
- **VM64721 LIMITHARD now works**
  - SET SRM LIMITHARD CONSUMPTION is default now
- **VM64767/64876 VARY PROCESSOR causes hangs**
- **VM64850 VSWITCH failover buffer mixup**
- **VM64795 Enhanced Contiguous Frame Handling**
- **VM64927 Spin Lock Manager Improvement**
- **VM64887 Erratic System Performance (PLDV overflow)**
- **VM64756 Long CPEBK Chains, Master-only work, and SYSTEMMP**

Service to z/VM 6.2 – Performance Sensitive

- **VM65011 – corrects VM64943 which in combination with this avoids abends and other problems when the *Monitor system service is used on a System z Server where Global Performance Data has been disabled.**
  - R540 PTF UM33450 – future RSU candidate
  - R610 PTF UM33480 – future RSU candidate
  - R620 PTF UM33512 – future RSU candidate

z/VM Performance: April 2012 SPEs

High Performance FICON: Outline

- **VM65041 lets z/VM guests use transport-mode I/O if channels and control units are so capable**
- **Transport-mode I/O uses a simpler command word structure that is easier for the channel subsystem and FICON adapter to handle, compared to conventional command-mode I/O**
- **On comparably configured workloads, transport-mode I/O gave us:**
  - About 35% increase in I/O rate
  - About 18% decrease in I/O service time
  - About 45% to 75% in CP CPU time per I/O
- **Workloads doing large I/Os tended to benefit most**
- **www.vm.ibm.com/perf/reports/zvm/html/620jb.html**

High Performance FICON: Some Numbers

Transport-mode I/O vs. command-mode I/O, 67% reads, 1 record per I/O

Guests/vol	Run Name	I/Os/vol/sec	Serv/I/O (msec)	%Busy/vol	%CP-CPU/I/O
1	JB001238 (c) JB001239 (t) Delta %Delta	3605.1 4914.6 1309.5 36.32	0.2326 0.1855 -0.0471 -20.25	83.8508 91.1570 7.3062 8.71	0.00156 0.00090 -0.00066 -42.33

Notes: (c) denotes command-mode I/O. (t) denotes transport-mode I/O.

Transport-mode I/O vs. command-mode I/O, 67% reads, 64 records per I/O

Guests/vol	Run Name	I/Os/vol/sec	Serv/I/O (msec)	%Busy/vol	%CP-CPU/I/O
1	JB001246 (c) JB001247 (t) Delta %Delta	359.4 554.4 195.0 54.26	2.6454 1.7660 -0.8794 -33.24	95.0844 97.9029 2.8185 2.96	0.00976 0.00242 -0.00734 -75.17

Notes: (c) denotes command-mode I/O. (t) denotes transport-mode I/O.

z10, 4 ded, 30G/2G, 4 FICON Express8, switched, DS8800/6GB, z/VM 6.2 plus VM65041, IO3390.

High Performance FICON: CP Commands

Command	Notes
QUERY PATHS	Shows whether chpids are enabled for transport mode
QUERY CHPID TYPE	Shows whether chpids are enabled for transport mode
QUERY MDC	Tells whether use of transport-mode I/O caused MDC to be disabled for a minidisk
VARY ONLINE PATH	Won’t allow incompatible paths to be varied online
TRACE and TRSOURCE	Appropriate updates to outputs

High Performance FICON: Monitor Record Changes

Monitor Record	What Was Changed
D0 R19 MRSYTSYG System Data	Added total zHPF channel program translations and number that were writes
D0 R20 MRSYTEPM Extended Channel Measurement Data	Added enhanced measurement fields for zHPF chpids
D1 R4 MRMTRSYS System Configuration	Added bit to say that enhanced channel measurement data is available
D6 R3 MRIODDEV Device Activity	Added counts of zHPF I/Os to device
D6 R32 MRIODHPF HPF Feature Change Event (new)	Records change in zHPF features in control unit

Performance Toolkit: VM65044 provides basic support.

High Performance FICON: Interaction with MDC

- **Transport-mode I/O directed at a minidisk will shut off MDC for said minidisk**
- **QUERY MDC command output will reveal that it happened**

```
Off for FRED 1234
Disabled by Transport Mode I/O
```

z/VM Performance: Other Thoughts

SLES 11 and MDC

- **In SLES 11 the DASD (SSCH) driver changed the CCWs Linux uses in its ECKD channel programs**
- **The new CCWs Linux uses are not eligible for MDC**
  – Fast-trans aborts or MDC aborts
- **You can see this in Performance Toolkit:**
  – FCX108 “Avoid” column will be zero
  – FCX179 “Abort” column
- **You can work around this by switching to the Diag x’250’ discipline instead**

Summary

z/VM Performance Update: Summary

- **z/VM 6.2: SSI and LGR, plus more**
  - Loose clustering for guest mobility
  - Recognition of systems becoming larger
    - Memory management improvements
    - Better defaults: MONDCSS, SAMPLE CONFIG, STORBUF
  - CPU MF counters: help us, help you
  - Lots of good service rolled into the base
  - See http:/www.vm.ibm.com/perf/ for more details
- **The adventure continues**

Contact Info:

Bill Bitner
z/VM Customer Focus and Care
z/VM Development Lab – Endicott, NY
bitnerb@us.ibm.com
+1 607 -429 -3286

Session 11796

Scan me for session evaluation