To MIPS or Not to MIPS

That is the CP Question!

CMG 2016
La Jolla, CA

Gary King
IBM

November 7, 2016

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

AlphaBlox*
APPN*
CICS*
CICS/VSE*
Cool Blue
DB2*
DFSMS
DFSMShsm
DFSMSrmm
DirMaint
DRDA*
DS6000
DS8000
ECKD
ESCON*
FICON*
FlashCopy*
GDPS*
HiperSockets
HyperSwap
IBM*
IBM eServer
IBM logo*
IMS
Language Environment*
Lotus*
Large System Performance Reference™ (LSPR™)
Multiprise*
MVS
OMEGAMON*
Parallel Sysplex*
Performance Toolkit for VM
PowerPC*
PR/SM
Processor Resource/Systems Manager
RACF*
Redbooks*
Resource Link
RETAIN*
REXX
RMF
S/390*
Scalable Architecture for Financial Reporting
Sysplex Timer*
Systems Director Active Energy Manager
System/370
System p*
System Storage
System x*
System z
System z9*
System z10
Tivoli*
Tivoli Storage Manager
TotalStorage*
VSE/ESA
VTAM*
WebSphere*
zEnterprise
xSeries*
z9*
z10
z10 BC
z10 EC
z/Architecture*
z/OS*
z/VM*
z/VSE
zSeries*

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* 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.

MIPS - can one number fit all?

- It's commonplace to assign IBM z Systems processors a capacity rating called MIPS
- The MIPS rating is very often used
  - to track and set workload capacity requirements
  - to select the proper size processor for the workload
- Today, we will discuss
  - just what are MIPS and where do they come from?
  - for a given processor, do all workloads run at the same MIPS?
  - how much trouble can using MIPS get us into?
    - and what to do about it

Just what are MIPS?

- Once upon a time, MIPS really meant Millions of Instructions Per Second
- As commonly used today, MIPS has become a RELATIVE indicator of AVERAGE processor CAPACITY
- MIPS are based on capacity RATIOS between processors
- MIPS are still in the ballpark of real Mi/sec
- Generally speaking,

  MIPS of new processor =

  MIPS of old processor x the AVERAGE CAPACITY RATIO new:old

Average Capacity Ratio

- IBM z Systems sets average capacity ratios among processors based on a variety of measured workloads which are published in the Large System Performance Reference (LSPR)
  - https://www.ibm.com/servers/resourcelink/lib03060.nsf/pages/lsprindex
- Old and new processors are measured in the same environment with the same workloads at high utilizations (>=90%)
- Over time, workloads and environment are updated to stay current with customer profiles
  - old processors measured with new workloads/environment may have different average capacity ratios compared to when they were originally measured

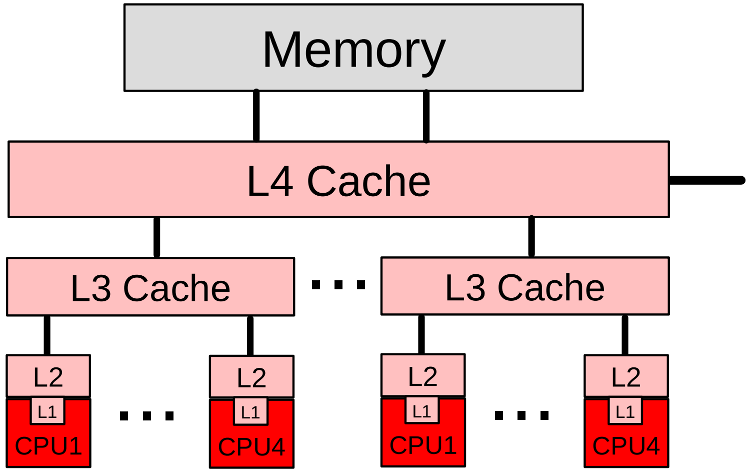
So, can one number (MIPS) fit all?

- To find out we have to ask …
  - When is it okay to use an average and when is it not?

- Sources of variation from average capacity ratio
  - System design
  - Workload characteristics
  - Workload scaling
  - CPU utilization
  - LPAR configurations
  - Coupling technology

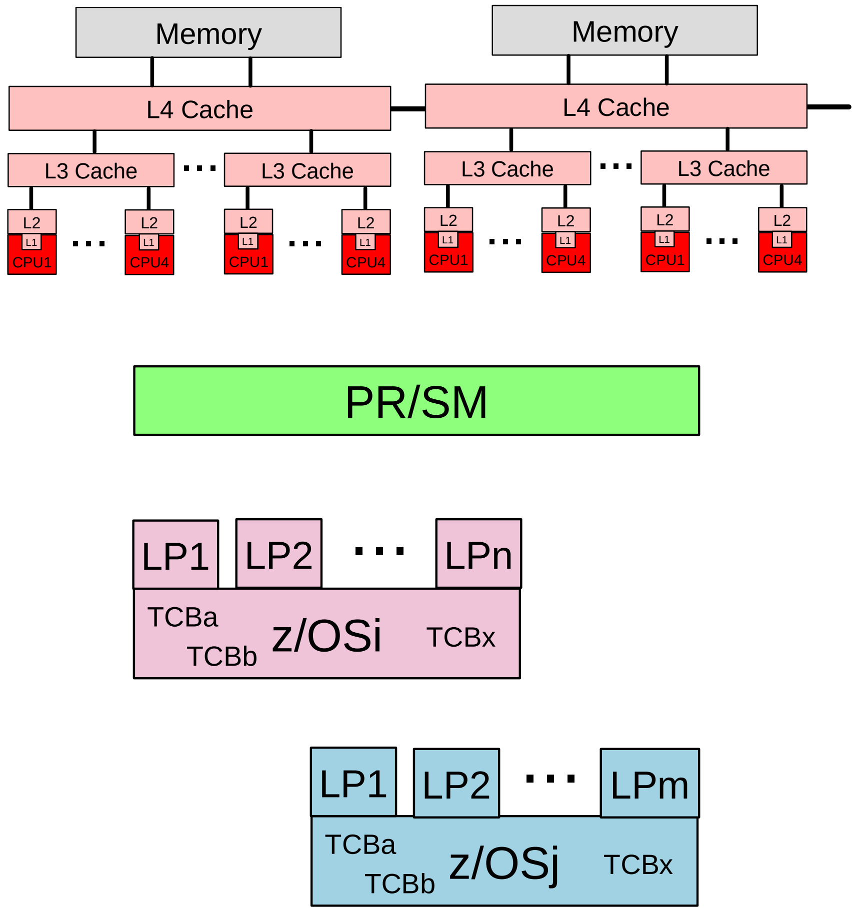
System Design: Processor

- Processor Design
  - CPU (core)
    - cycle time
    - pipeline
    - branch prediction
    - hardware vs. millicode
  - memory hierarchy (nest)
    - high speed buffers (caches)
      - on chip, on book/node
      - private, shared
    - buses
      - number, bandwidth
    - latency
      - distance
      - speed of light



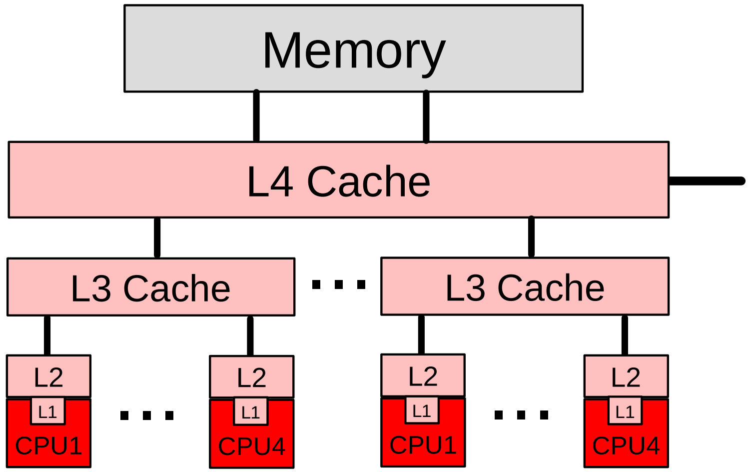
System Design: Hipervisor and OS

- Hipervisor (PR/SM)
  - virtualization layer at OS level
  - distributes physical resources
    - memory
    - processors
      - logicals dispatched on physicals
      - dedicated
      - shared
      - affinities
- OS
  - virtualization layer at addrspc level
  - distributes logical resources
    - memory
    - processors
      - tasks dispatched on logcials
- Enhanced cooperation
  - HiperDispatch
    - z/OS + PR/SM
    - z/VM + PR/SM



Workload Characteristics

- Workload Characteristics
  - CPU
    - instructions
      - mix
      - sequence
      - branch characteristics
    - task dispatch profile
      - locked in or chatty
  - memory
    - size
    - locality of reference
    - multiprogramming level
  - I/O rate



LSPR z/OS workload primitives

- CB-L commercial batch long job steps
- WASDB WebSphere-focus application server and data base
- OLTP-T traditional online transaction processing
- OLTP-W webenabled access to legacy data

CHARACTERISTICS MORE IMPORTANT THAN NAME

	CPU use profile	I/O	Memory Hierarchy
CB-L	heavy appl, light OS	light	light
WASDB	medium appl and OS	light	light/moderate
OLTP-T	medium appl and OS	heavy	moderate
OLTP-W	medium appl and OS	moderate	stress

NOW RUN IN VARIOUS MIXES TO PRODUCE WORKLOADS MATCHING CUSTOMER PROFILE OF MEMORY HIERARCHY STRESS OR RELATIVE "NEST" INTENSITY (RNI)

Relative Nest Intensity (RNI)

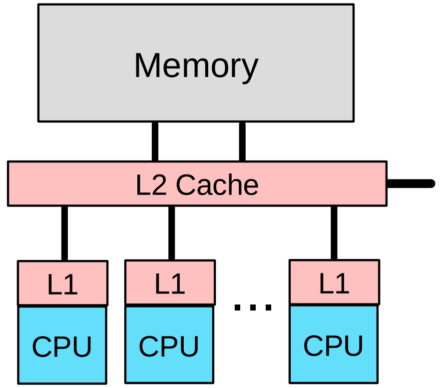
- Activity beyond private cache(s) is the most sensitive area
- Reflects distribution and latency of sourcing from shared caches and memory
- Data for calculation available from CPU MF (SMF 113) starting with z10

LSPR Workload Categories

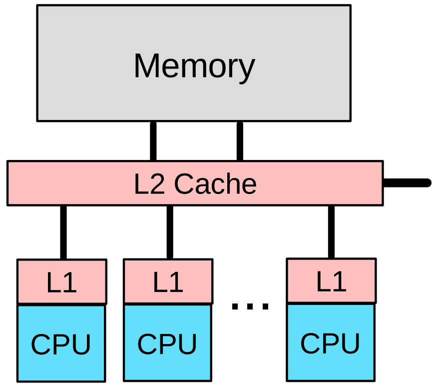
- Categories developed to match the profile of data gathered on customer systems
  - over 100 data points (LPARs) used in the profiling
- Various combinations of prior workload primitives are measured to reflect the new workload categories
  - Applications include CICS, DB2, IMS, OSAM, VSAM, WebSphere, COBOL, utilities
- **LOW** (relative nest intensity)
  - Workload curve representing light use of the memory hierarchy
  - Similar to past high Nway scaling workload primitives
- **AVERAGE** (relative nest intensity)
  - Workload curve expected to represent the majority of customer workloads
  - Similar to the past LoIO-mix curve
- **HIGH** (relative nest intensity)
  - Workload curve representing heavy use of the memory hierarchy
  - Similar to the past DI-mix curve
- zPCR extends these published categories
  - Low-Avg: 50% LOW and 50% AVERAGE
  - Avg-High: 50% AVERAGE and 50% HIGH

System Design + Workload Characteristics
Variation from Average: sometimes small
Example: z990 to z9 EC

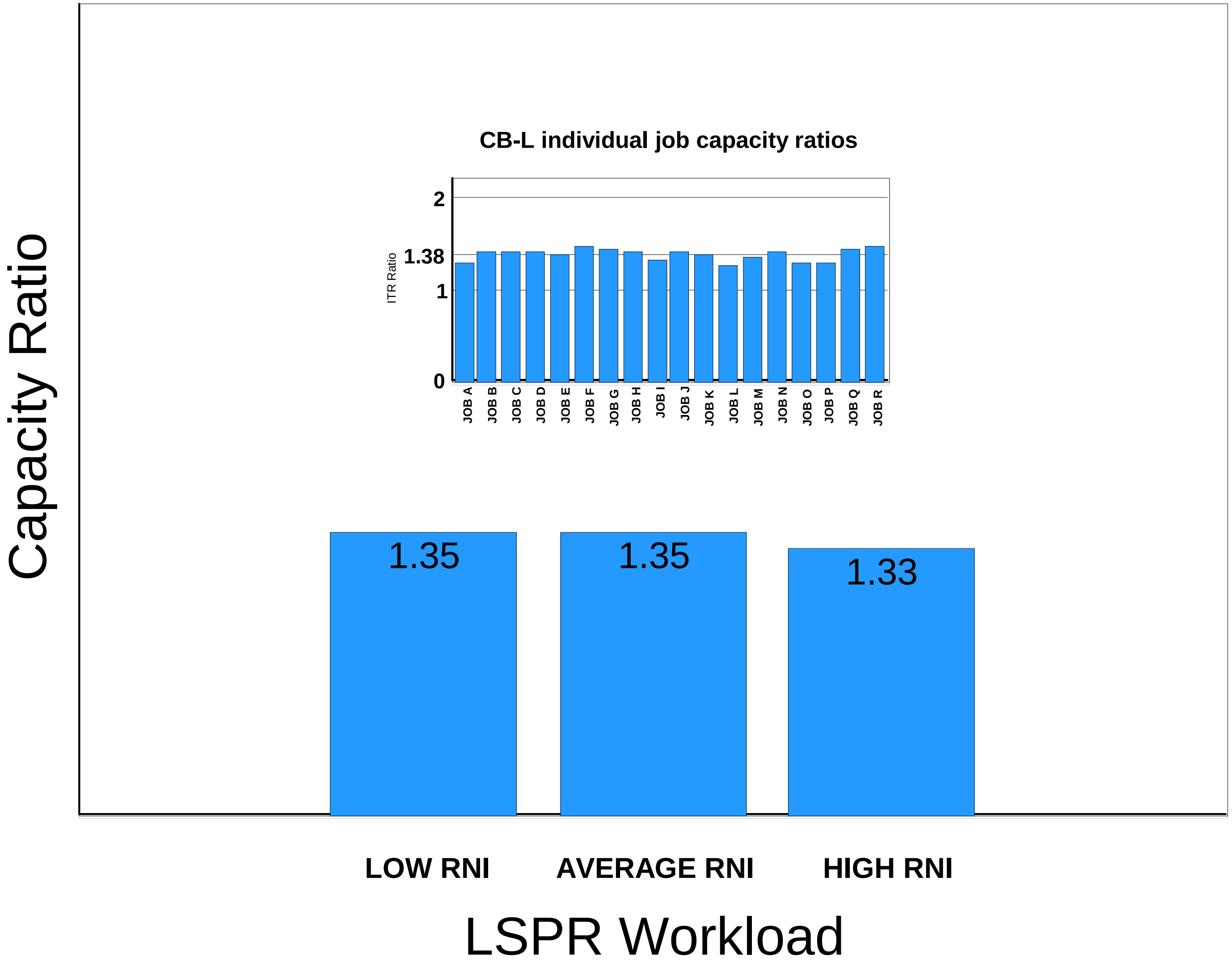
- z990
  - CPU
    - 1.2 GHz
    - superscalar
  - Caches
    - L1 private 256k i, 256k d
    - L2 shared 32 MB / book
    - book interconnect: ring



- z9 EC
  - CPU
    - 1.7 GHz
    - superscalar
  - Caches
    - L1 private 256k i, 256k d
    - L2 shared 40 MB / book
    - book interconnect: ring

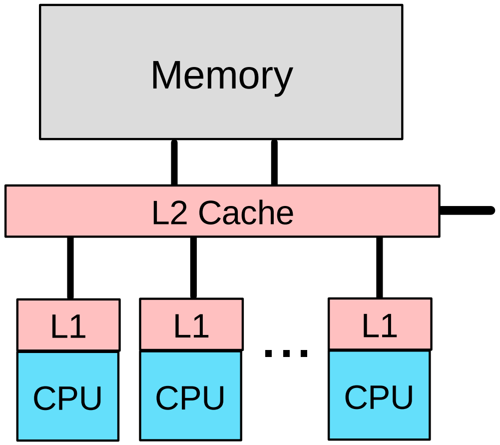


LSPR Single Image Capacity Ratios
10way: z9 EC versus z990

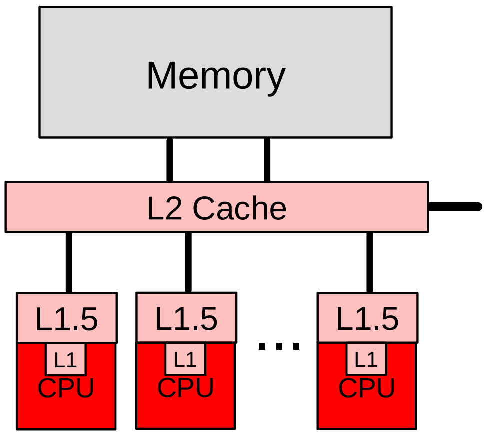


System Design + Workload Characteristics
Variation from Average: sometimes large
Example: z9 EC to z10 EC

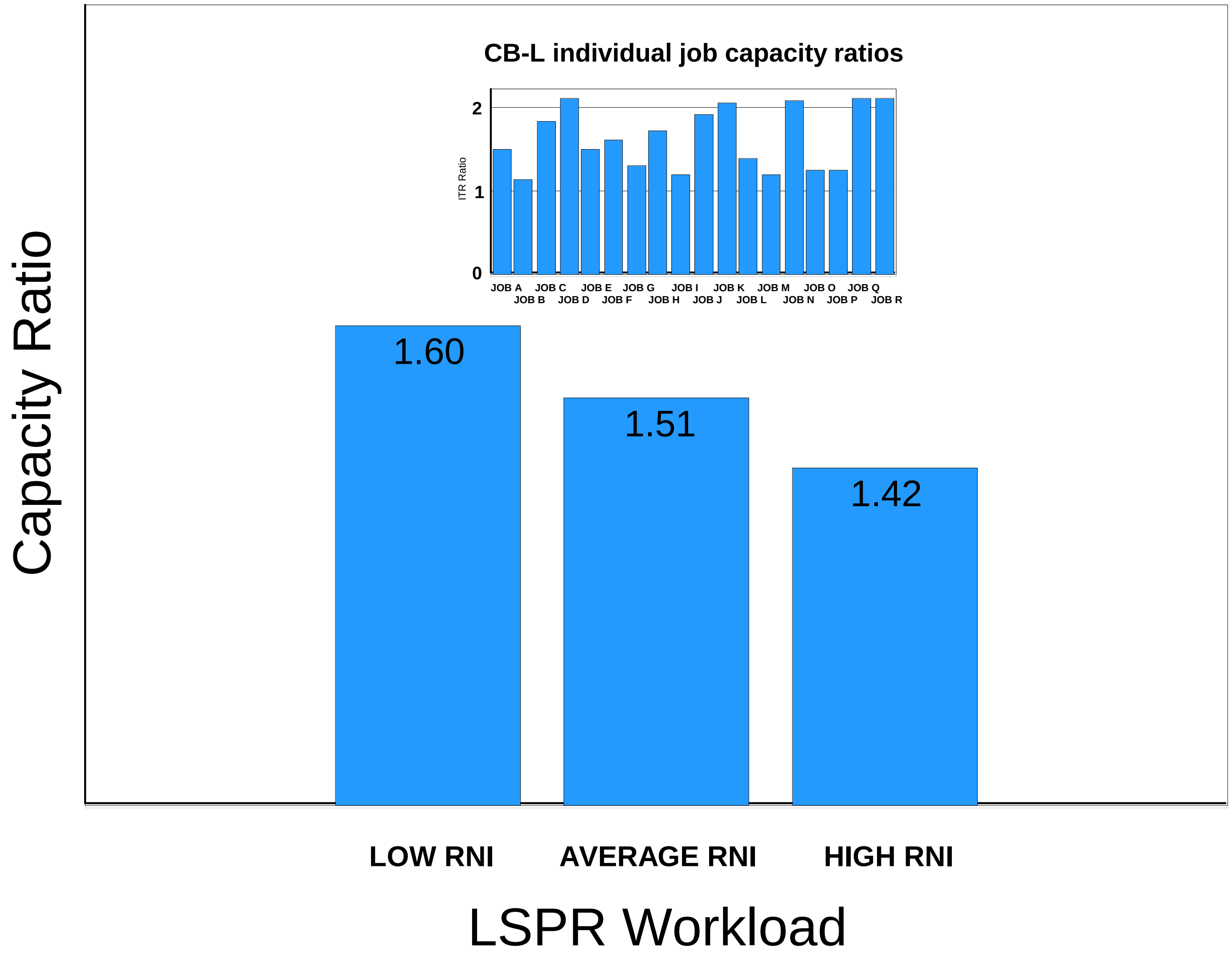
- z9 EC
  - CPU
    - 1.7 GHz
    - superscalar
  - Caches
    - L1 private 256k i, 256k d
    - L2 shared 40 MB / book
    - book interconnect: ring



- z10 EC
  - CPU
    - 4.4 GHz
    - redesigned pipeline
    - superscalar
  - Caches
    - L1 private 64k i, 128k d
    - L1.5 private 3 MB
    - L2 shared 48 MB / book
    - book interconnect: star

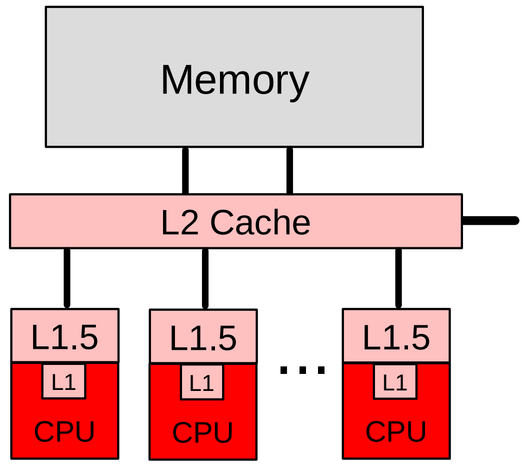


LSPR Single Image Capacity Ratios
10way: z10 EC versus z9 EC

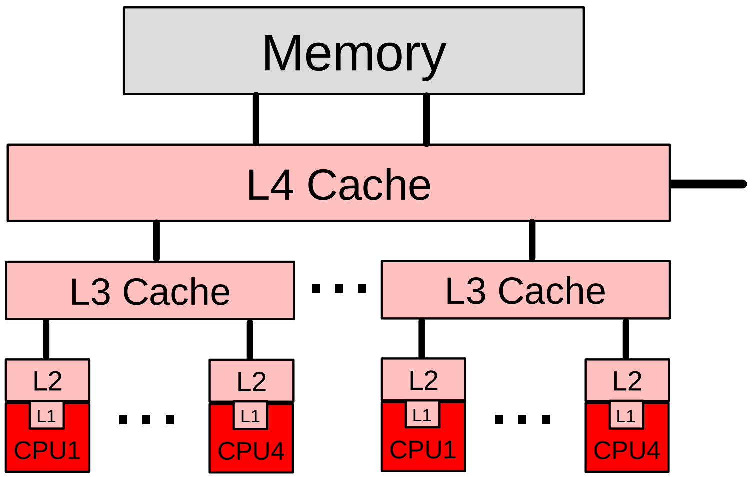


System Design + Workload Characteristics
Variation from Average: sometimes inbetween
Example: z10 EC to z196

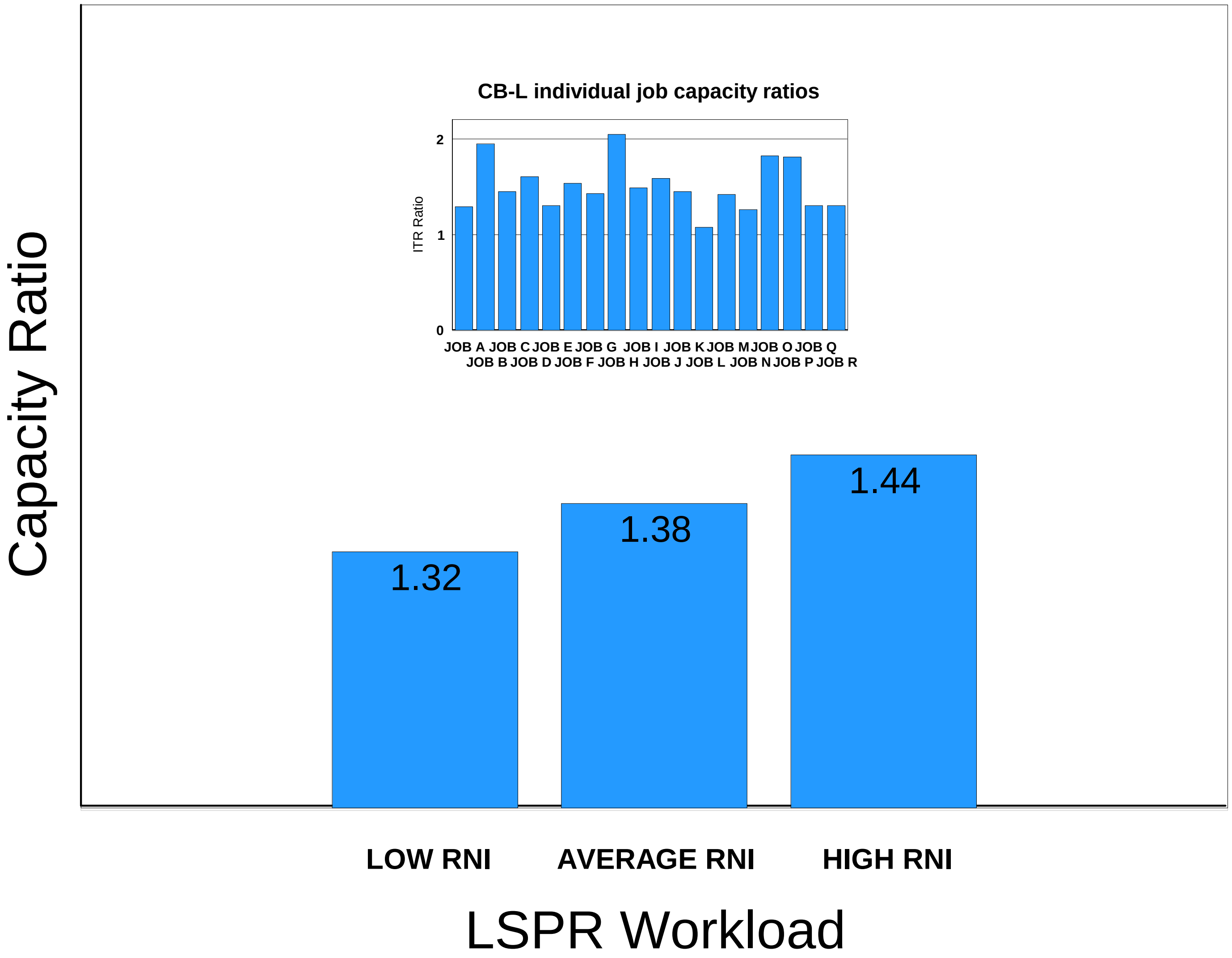
- z10 EC
  - CPU
    - 4.4 GHz
  - Caches
    - L1 private 64k i, 128k d
    - L1.5 private 3 MB
    - L2 shared 48 MB / book
    - book interconnect: star



- z196
  - CPU
    - 5.2 GHz
    - Out-Of-Order execution
  - Caches
    - L1 private 64k i, 128k d
    - L2 private 1.5 MB
    - L3 shared 24 MB / chip
    - L4 shared 192 MB / book
    - book interconnect: star

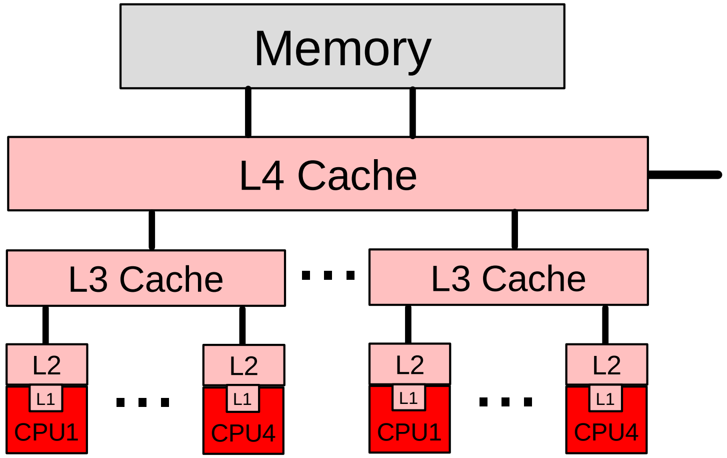


LSPR Single Image Capacity Ratios
10way: z196 versus z10 EC

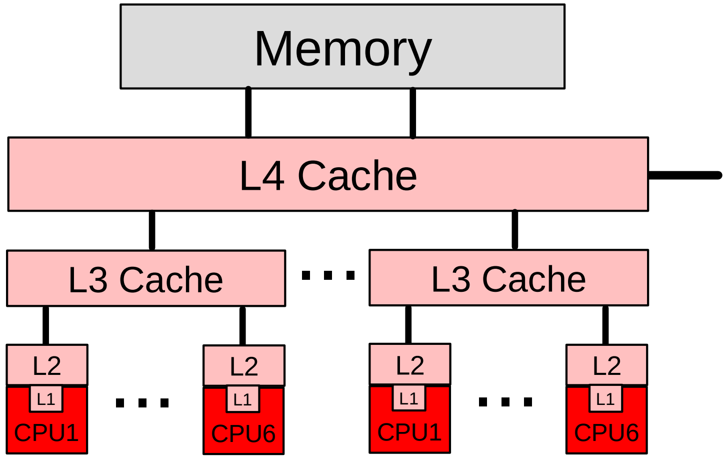


System Design + Workload Characteristics
Variation from Average: sometimes fairly small
Example: z196 to zEC12

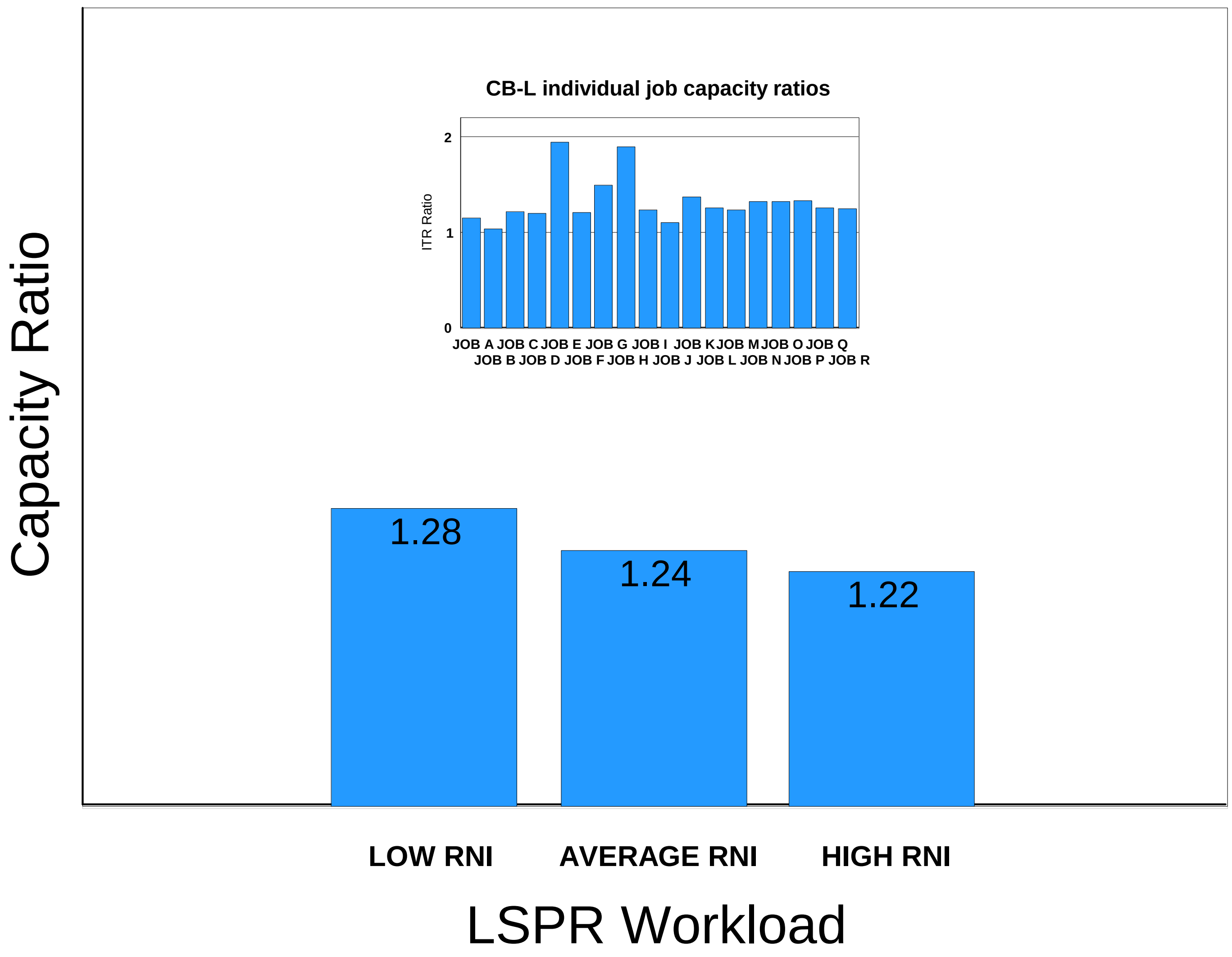
- z196
  - CPU
    - 5.2 GHz
    - Out-Of-Order execution
  - Caches
    - L1 private 64k i, 128k d
    - L2 private 1.5 MB
    - L3 shared 24 MB / chip
    - L4 shared 192 MB / book



- zEC12
  - CPU
    - 5.5 GHz
    - Enhanced Out-Of-Order
  - Caches
    - L1 private 64k i, 96k d
    - L2 private 1 MB i + 1 MB d
    - L3 shared 48 MB / chip
    - L4 shared 384 MB / book



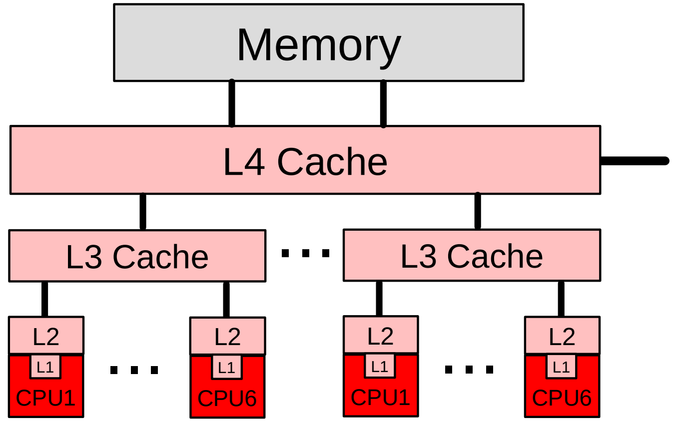
LSPR Single Image Capacity Ratios
10way: zEC12 versus z196



z13 versus zEC12 hardware comparison

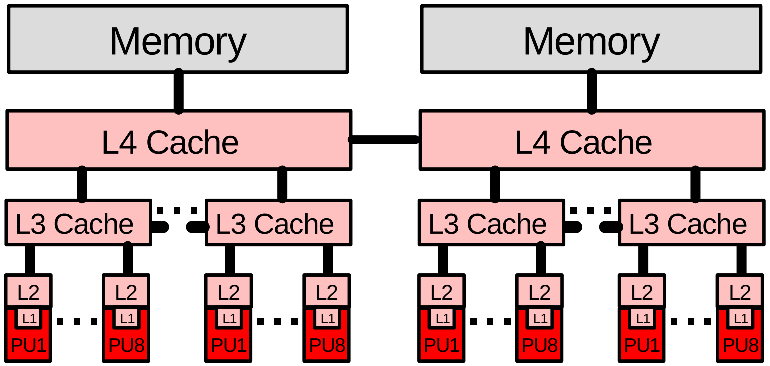
- zEC12
  - CPU
    - 5.5 GHz
    - Enhanced Out-Of-Order
  - Caches
    - L1 private 64k i, 96k d
    - L2 private 1 MB i + 1 MB d
    - L3 shared 48 MB / chip
    - L4 shared 384 MB / book

Single Book View

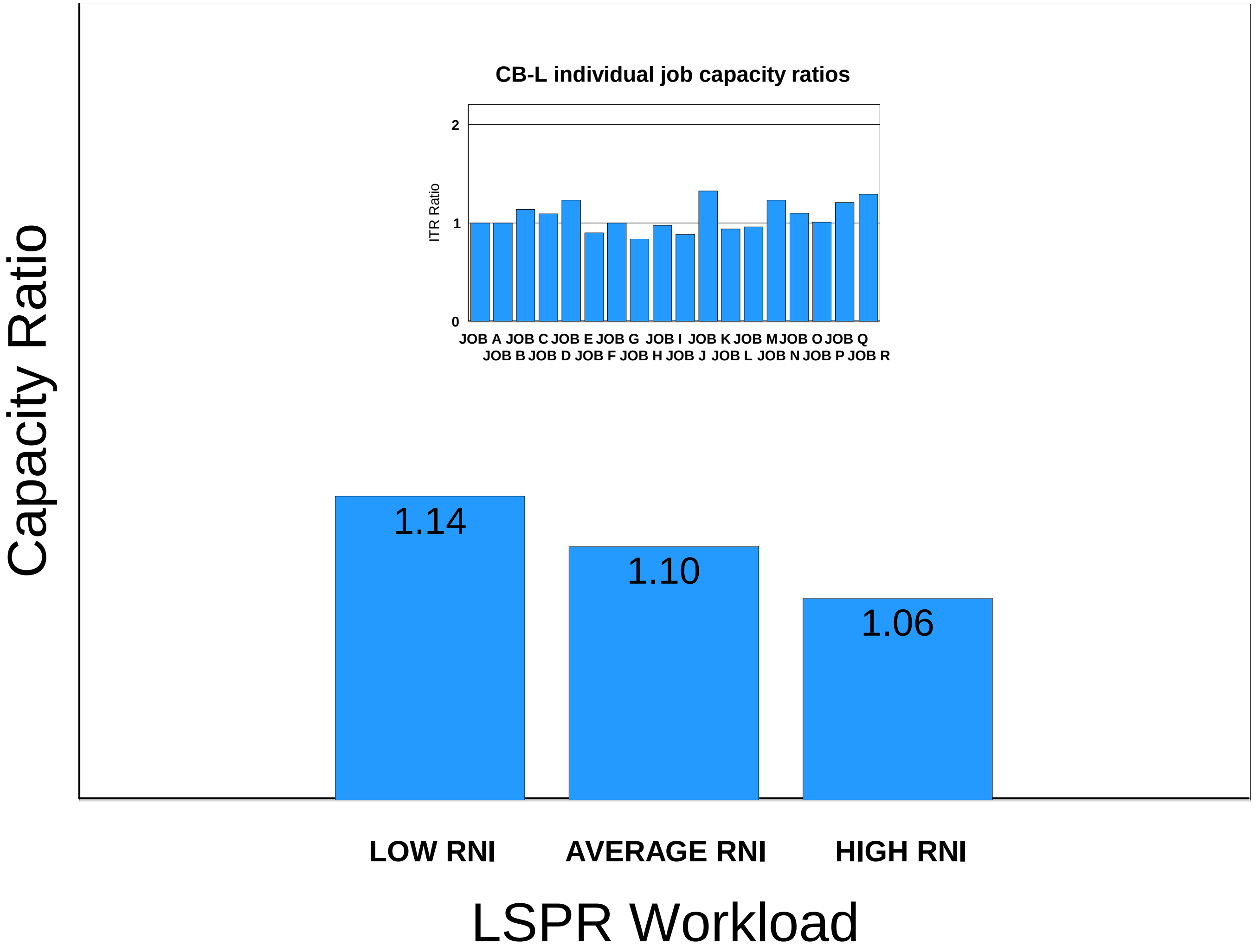


- z13
  - CPU
    - 5.0 GHz
    - Major pipeline enhancements
  - Caches
    - L1 private 96k i, 128k d
    - L2 private 2 MB i + 2 MB d
    - L3 shared 64 MB / chip
    - L4 shared 480 MB / node
      - plus 224 MB NIC

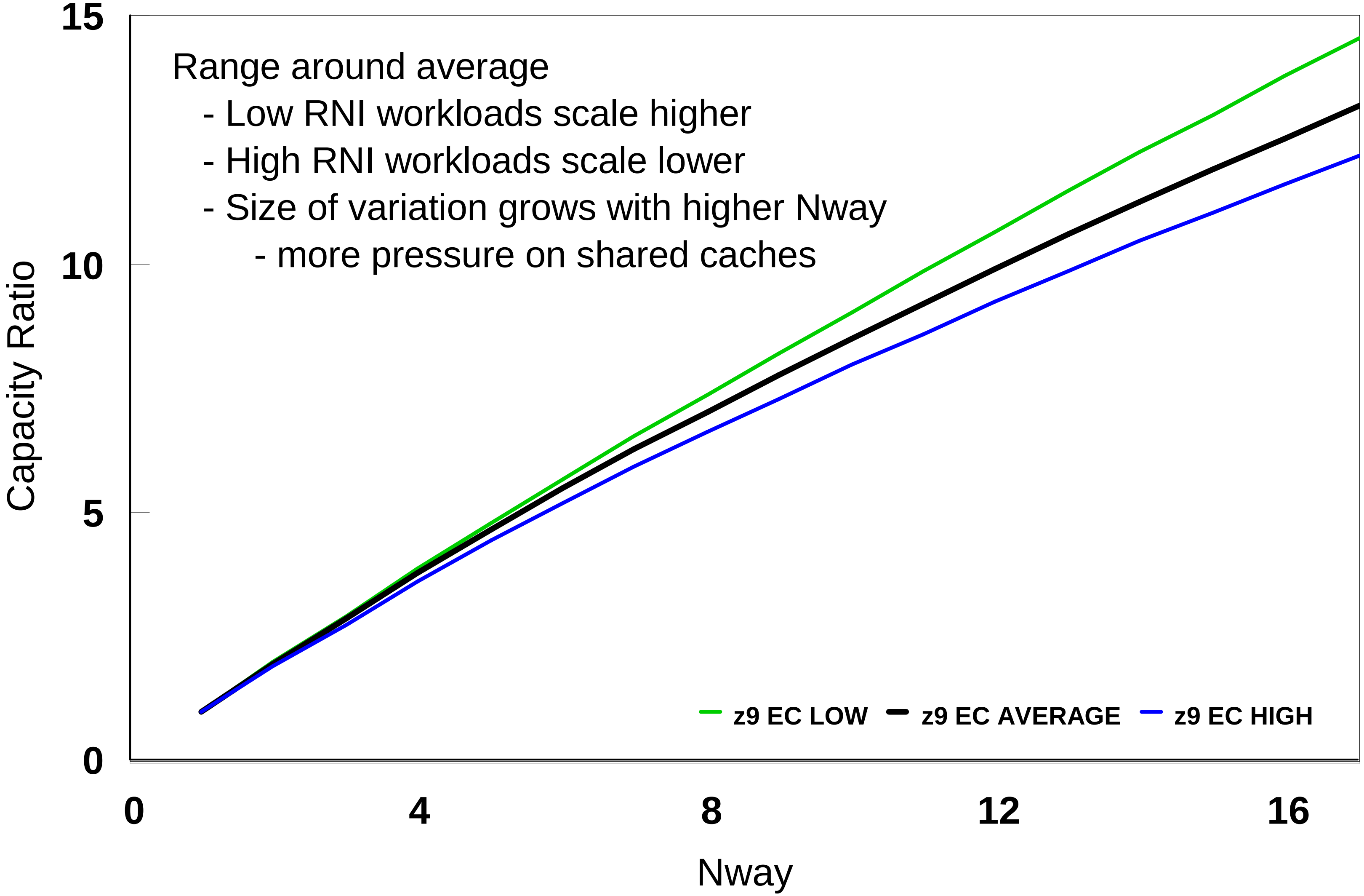
Single Drawer View



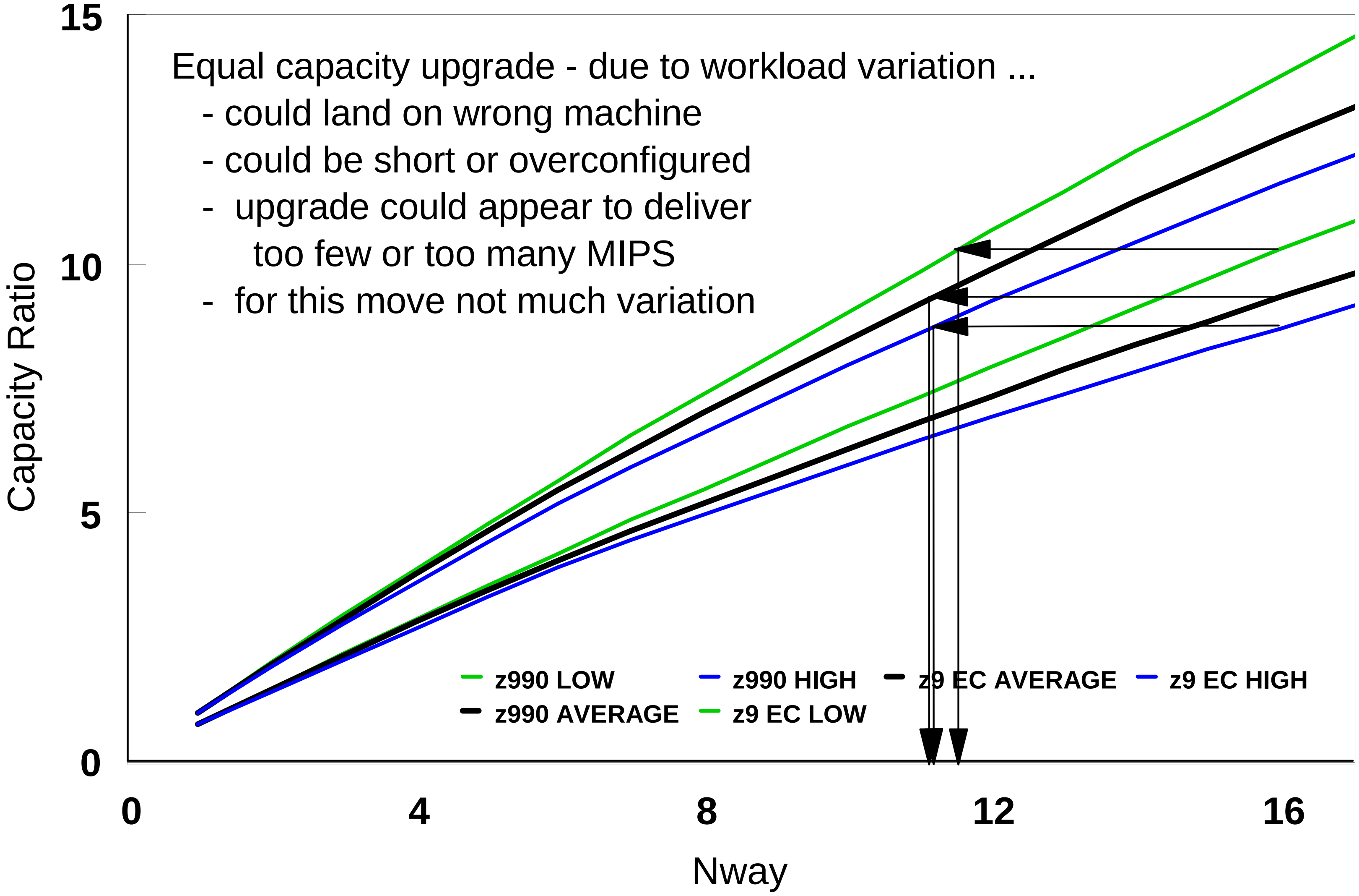
LSPR Single Image Capacity Ratios
16way: z13 versus zEC12
Example of Workload Variability



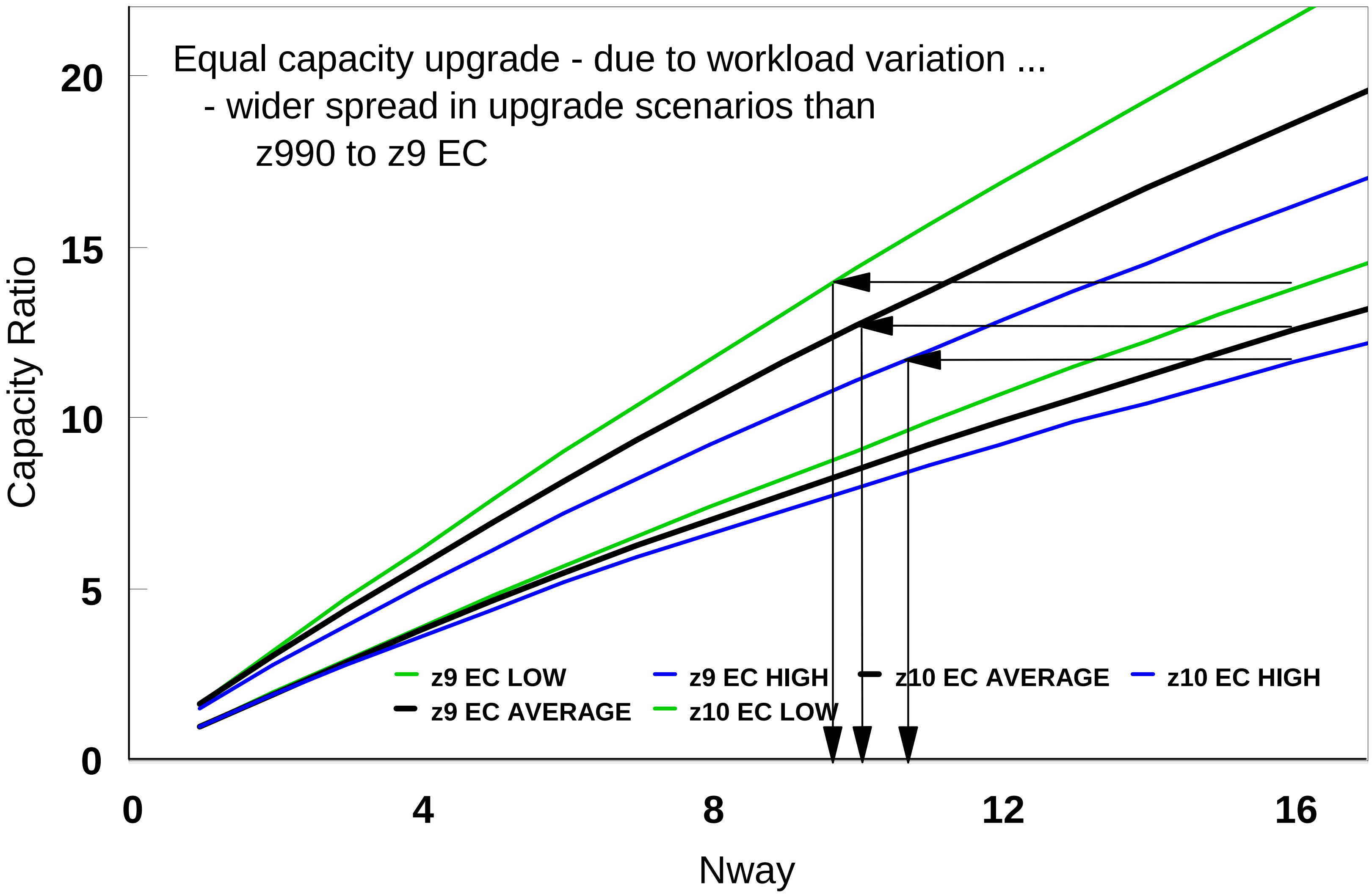
Workload Scalability
Variation from Average
Example: z9 EC



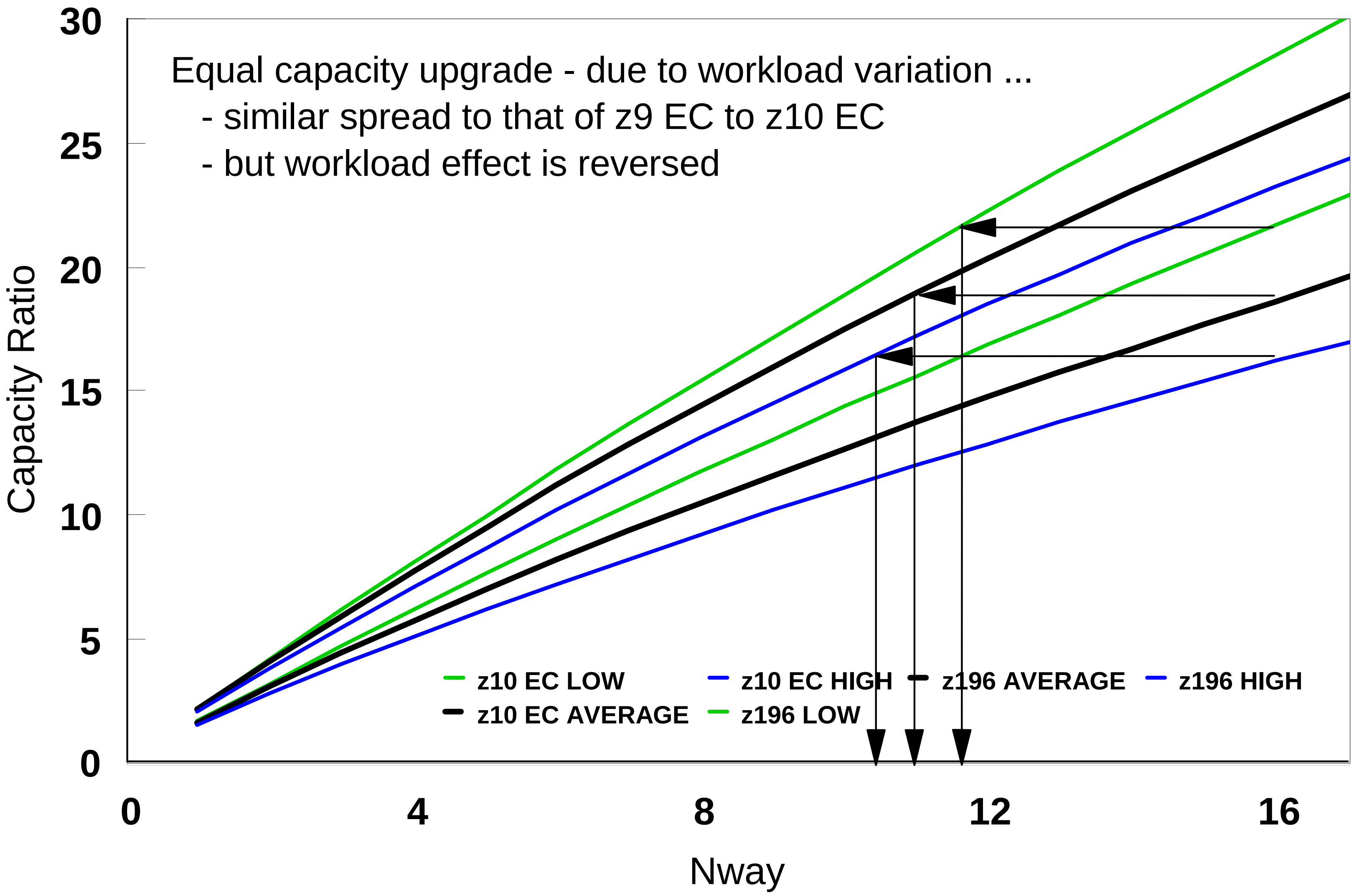
Workload Scalability
Variation from Average
Example: moving from z990 to z9 EC



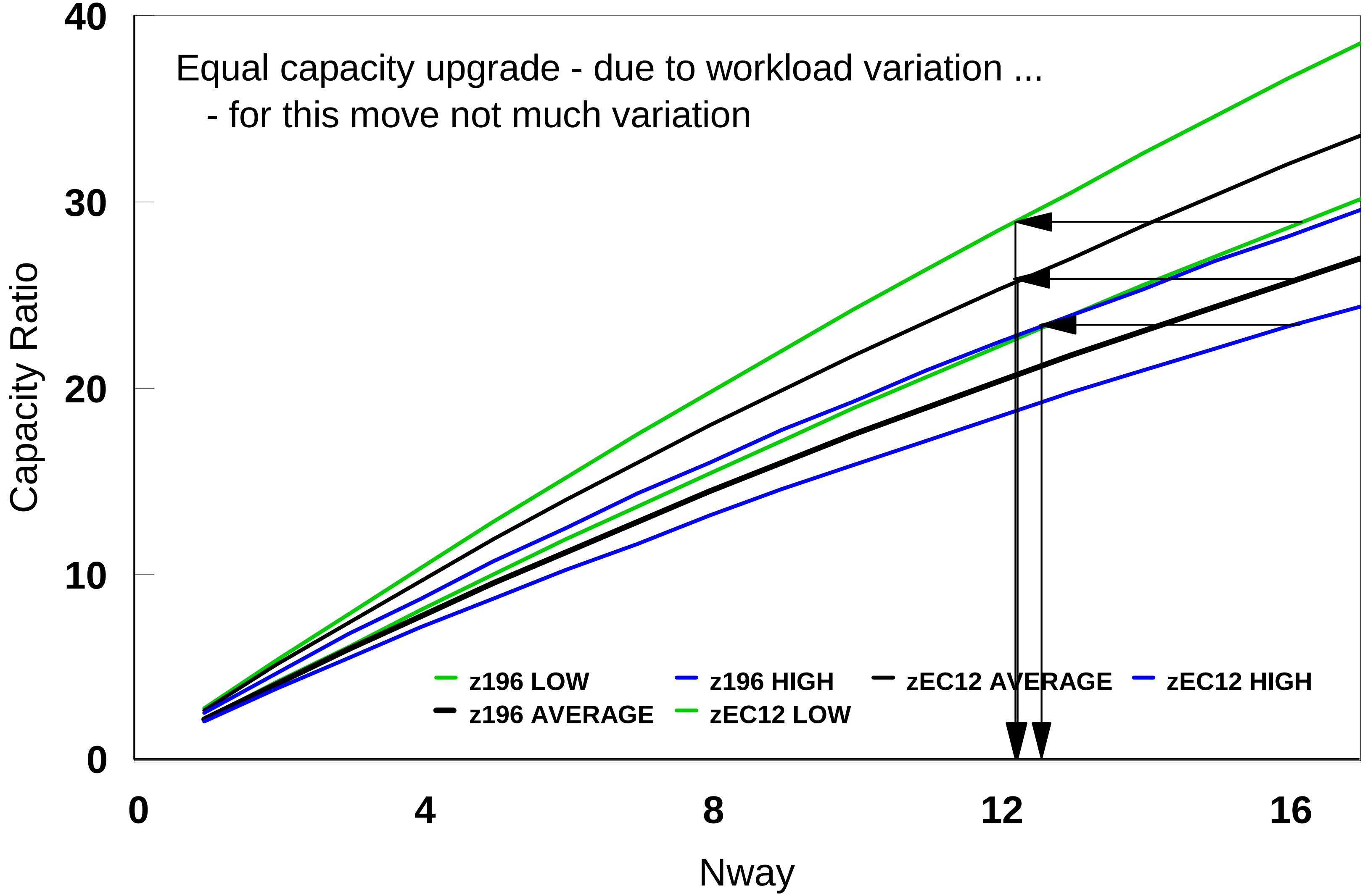
Workload Scalability
Variation from Average
Example: moving from z9 EC to z10 EC



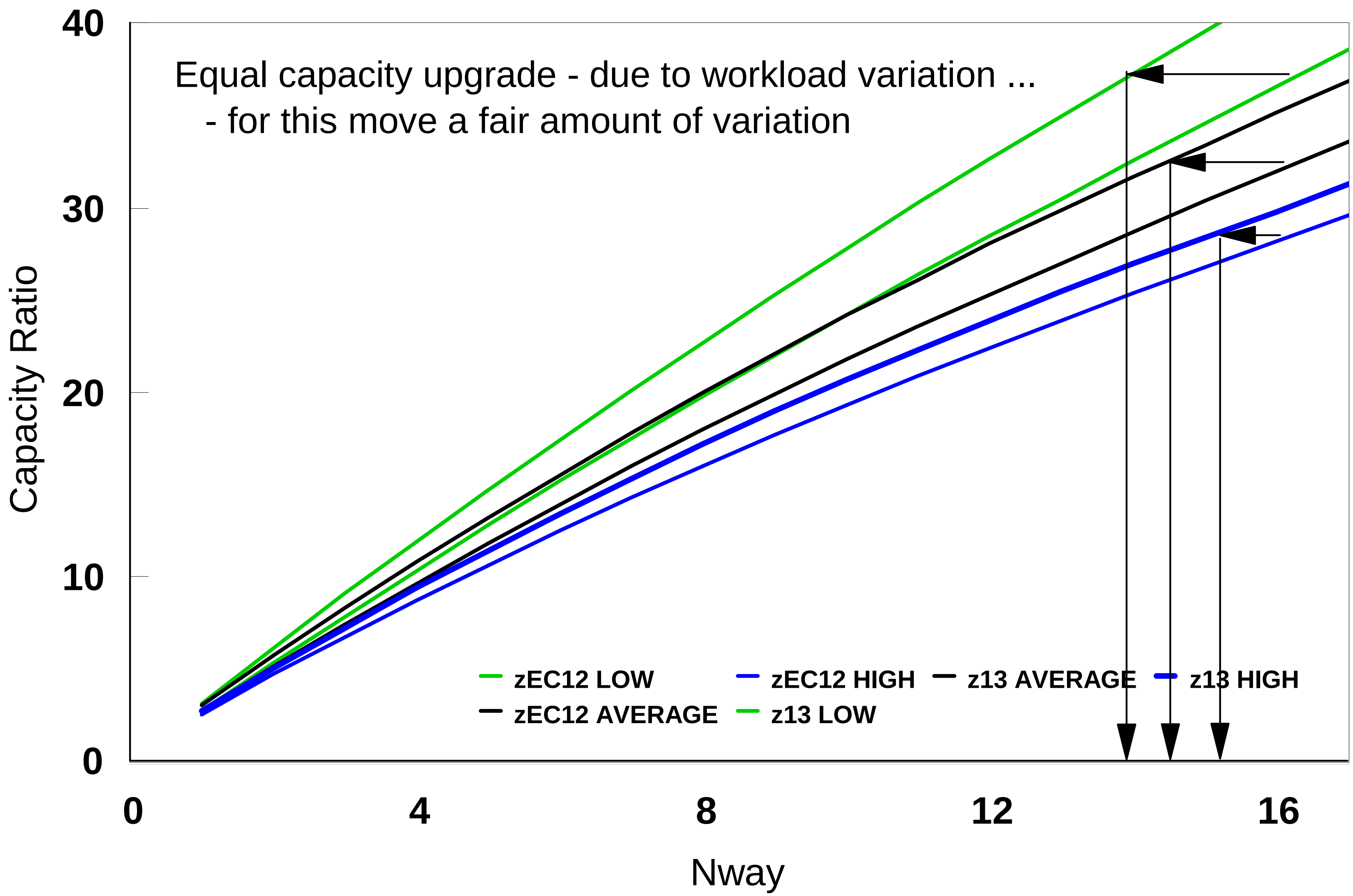
Workload Scalability
Variation from Average
Example: moving from z10 EC to z196



Workload Scalability
Variation from Average
Example: moving from z196 to zEC12



Workload Scalability
Variation from Average
Example: moving from zEC12 to z13



CPU Utilization
Source of Variation

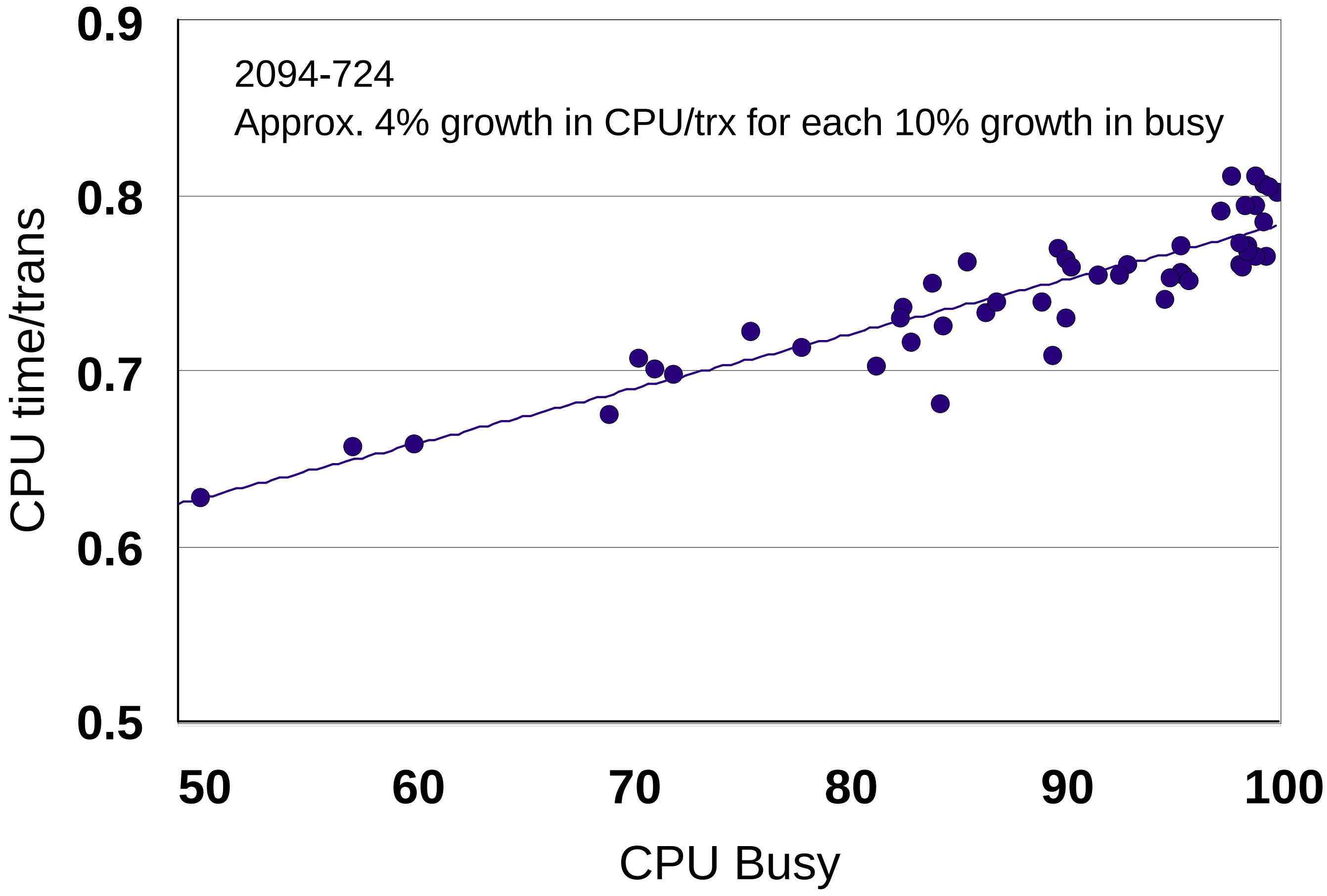
- CPU utilization generally reflects the amount of work flowing through a fixed HW/SW configuration
  - the higher the workload rate, the higher the utilization

- As more work flows through a fixed HW/SW configuration, the efficiency of the HW and SW is reduced
  - less shared HW resources (caches, buses) available to each work unit
  - SW manages more work units - longer queues, more contention
  - CPU time per transaction or job will grow

- Magnitude of the effect is related to
  - workload characteristics
    - higher RNI workloads (as measured at higher utilizations) see higher impact
  - size of the processor
    - smallest Nways (say 1-4way) are somewhat less sensitive

OLTP Client Workload Example

Growth in CPU time/trans as CPU busy increases



CPU Utilization
Impact to Capacity Planning When Using MIPS

- Impact to capacity planning comes in two flavors
  - may have less headroom on the box than you think
  - when moving a workload, it may not fit in the new container

- Example
  - assume a workload is running at 50% busy on a 2000 MIPS box
    - without factoring in utilization effect, it will be called a 1000 MIPS workload
    - in fact, it may be an 1200 MIPS workload when running at the efficiency of a 90% busy box
  - caution #1: there is NOT room to double this workload on the current box
  - caution #2: if moved to a new box or LPAR, it will likely need a 1200 MIPS container (not 1000 MIPS) to fit

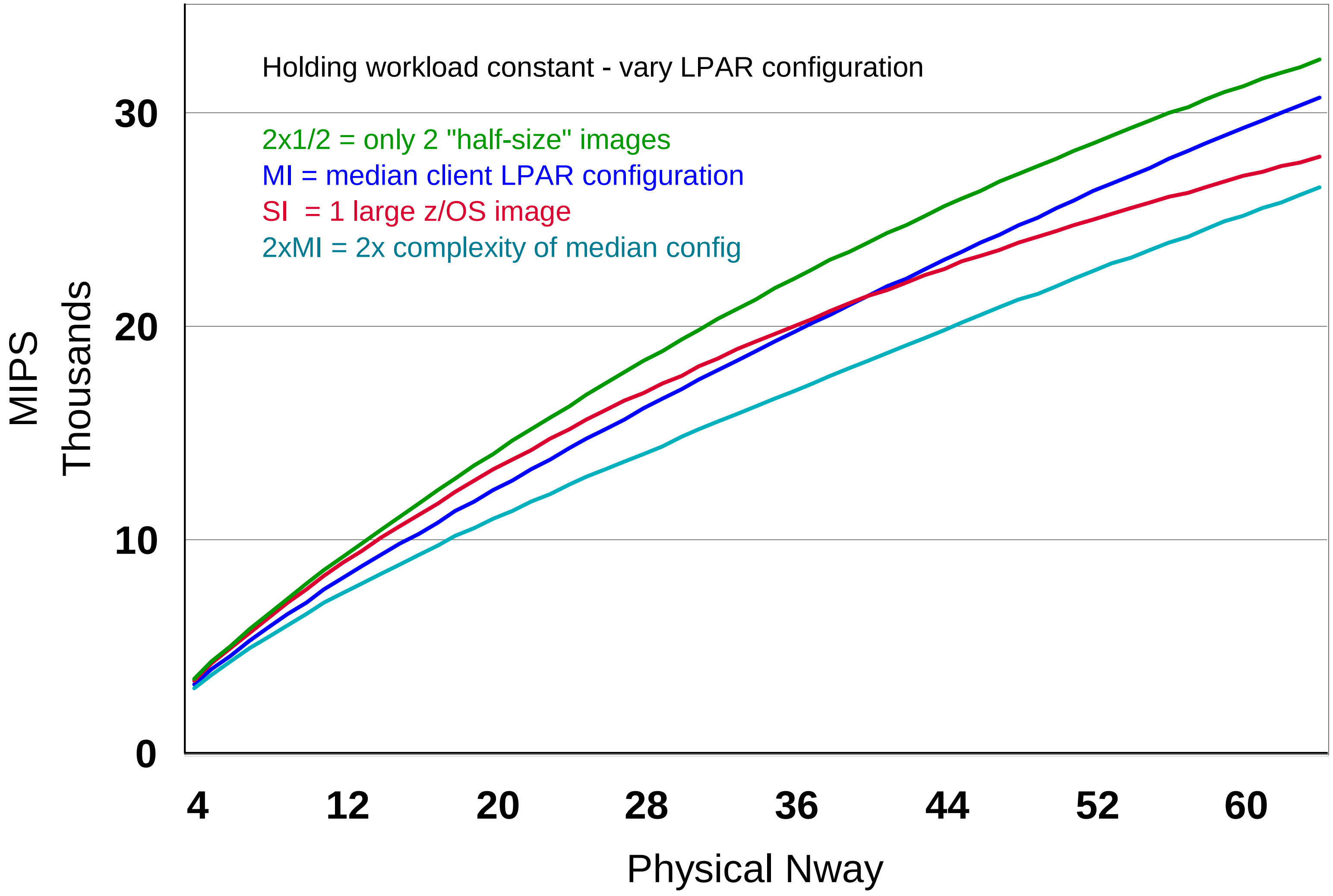
- Estimating the impact - conservative approach
  - For a change in utilization of 10%, plan for the capacity effect to be
    - 3% for LOW RNI workloads
    - 4% for AVERAGE RNI workloads
    - 5% for HIGH RNI workloads

LPAR Configurations
Variation from Average MIPS

- LPAR configurations affect the efficiency of the HW and SW
  - key factors
    - workload characteristics
    - number of LPARs
    - number of logical processors and weight of each LPAR
    - overall ratio of logical to physical processors

- MIPS ratings are based on AVERAGE wkld and median client LPAR config
  - median client LPAR configuration varies by Nway
    - number of LPARs
      - 5 at low-end, 9 at high-end
      - generally 2 are major (>20% of weight), rest are minor
    - size of major LPARs
      - close to Nway of box for low/mid-range Nways, well less than Nway at high-end
    - logical:physical ratio
      - 5:1 at low end, 2:1 for most, 1.3:1 at high end

Example LPAR Configurations
Effect on MIPS



Coupling Technology
Impact on MIPS

- Sysplex configurations affect the efficiency of the HW and SW
  - key factors
    - workload characteristics - rate of operations to the coupling facility
    - speed of coupling technology (CPU and links) versus speed of host technology
  - example host effects
    - 2% for light coupling workload
    - 5-7% for medium coupling workload with speed-matched CF technology
    - 9% for medium coupling workload with "slow" CF technology
    - 10-14% for heavy coupling workload with speed-matched CF technology
    - 18% for heavy coupling workload with "slow" CF technology
- When upgrading the host, must consider impact of CF technology on MIPS requirement

So, what have we learned about MIPS?

- When there is a big change in a sensitive factor - be careful
  - move to new processor technology
  - change in workload characteristics
  - change in CPU utilization
  - change in LPAR configuration
  - change in coupling technology
- But, most of the time, the items above are stable or change only a little
  - for example, adding an engine to an existing processor
- And over the long run many variations tend to "even out"
  - for example, when moving to a new technology, a below average workload this time is often an above average workload the next time

Conclusions about MIPS

- MIPS are fine for long term workload trending
- MIPS are okay for short term planning where there are only minor changes in any of the sensitive factors
- But whenever there is to be a major change, there is a risk of significant variation from average (MIPS) and additional analysis should be done
- Useful tool to help with "additional analysis" - zPCR

zPCR

- Capacity sizing tool available for download from
  - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381
  - or just search for "zPCR download"
- Through customization, zPCR can provide insight on many of the sensitive factors discussed in this presentation
  - system design
  - workload characteristics
  - workload scaling
  - LPAR configurations