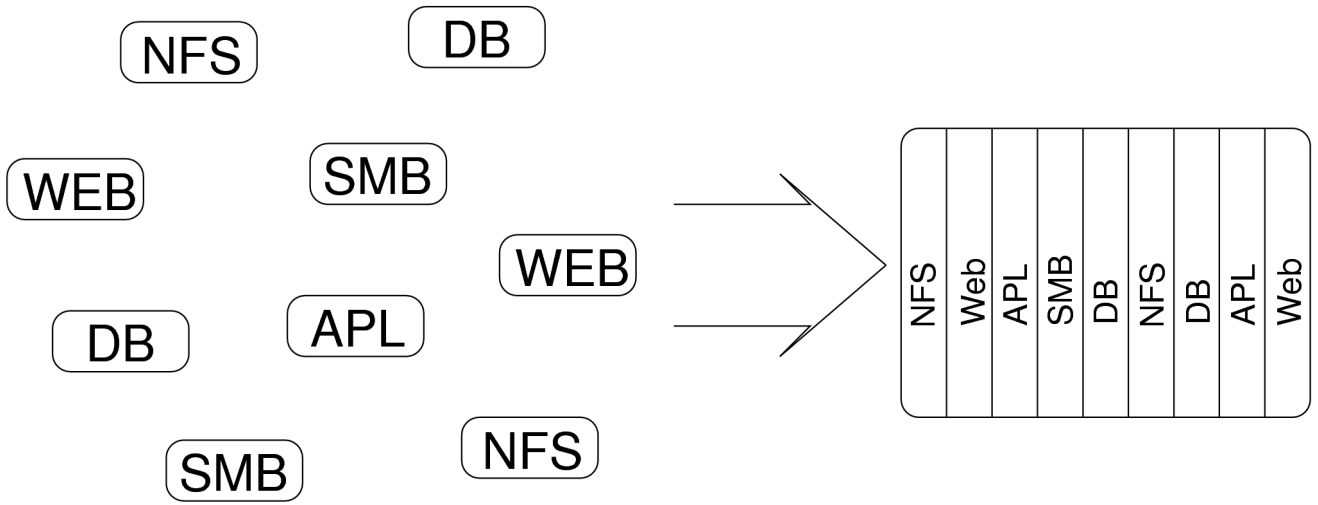
Partitioning Computers

Rolf M Dietze

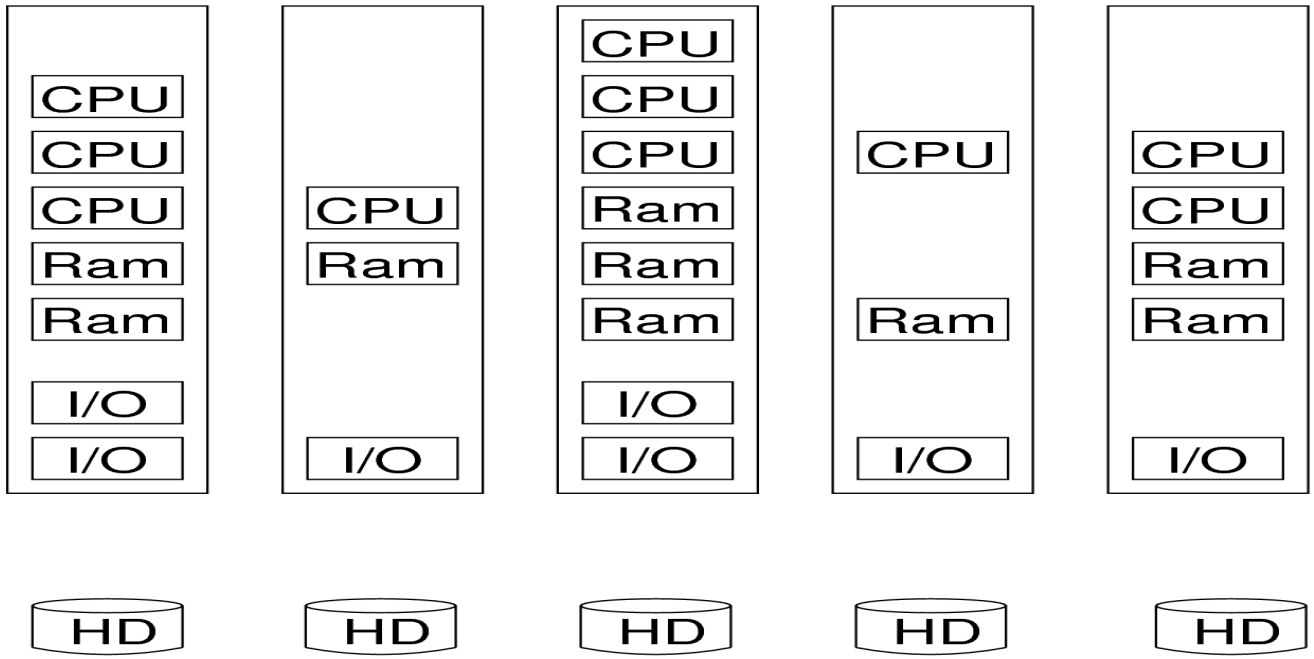
rolf.dietze@dietze-consulting.de

Consolidation/Partitioning



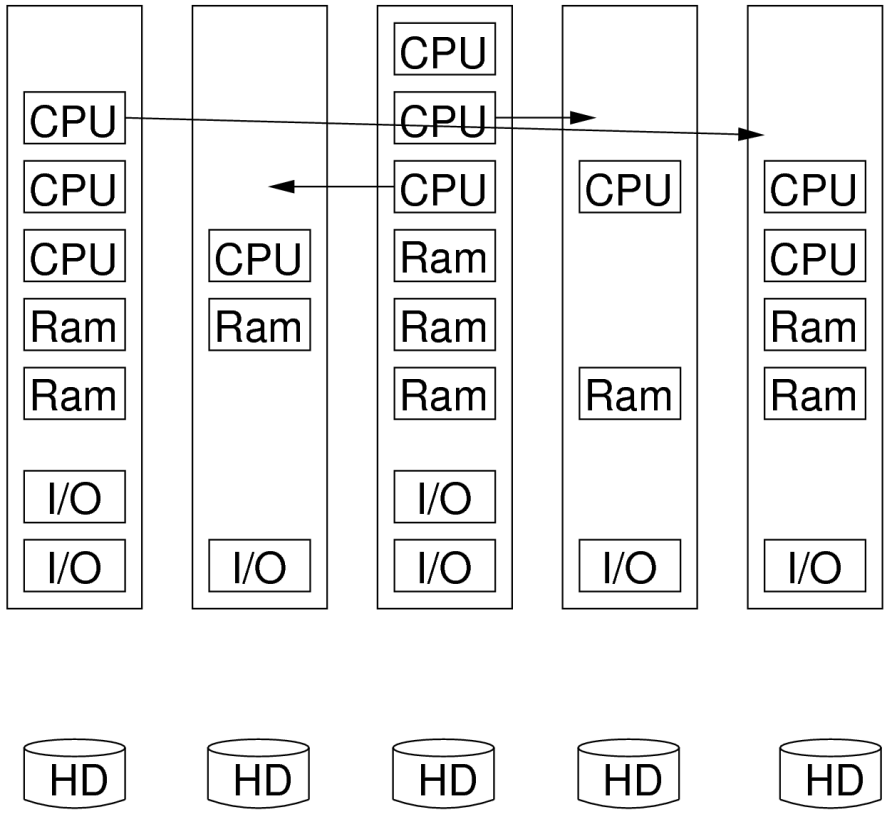
Rolf M Dietze, rmd@dietze-consulting.de

Hardware Partitioning



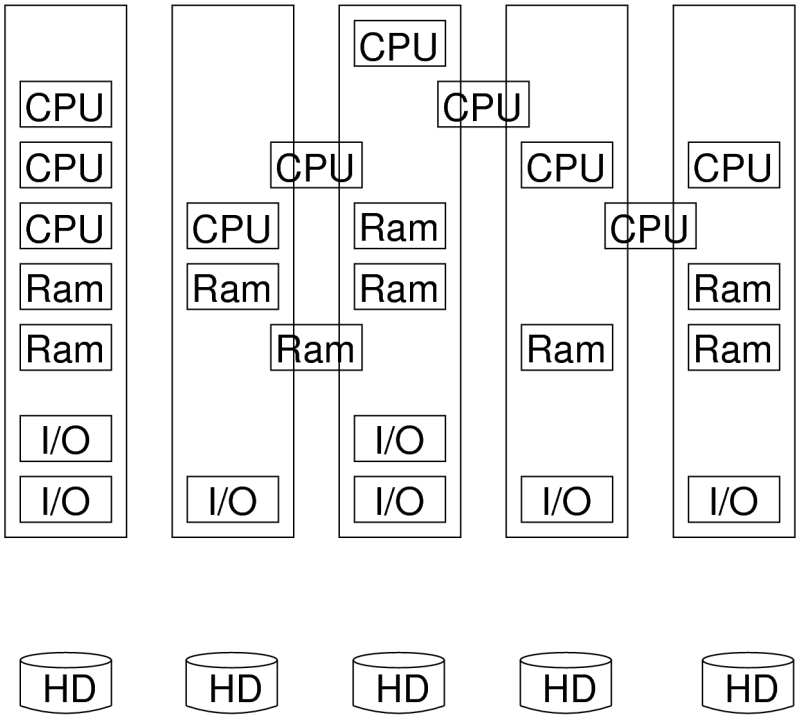
Rolf M Dietze, rmd@dietze-consulting.de

Resource Direction



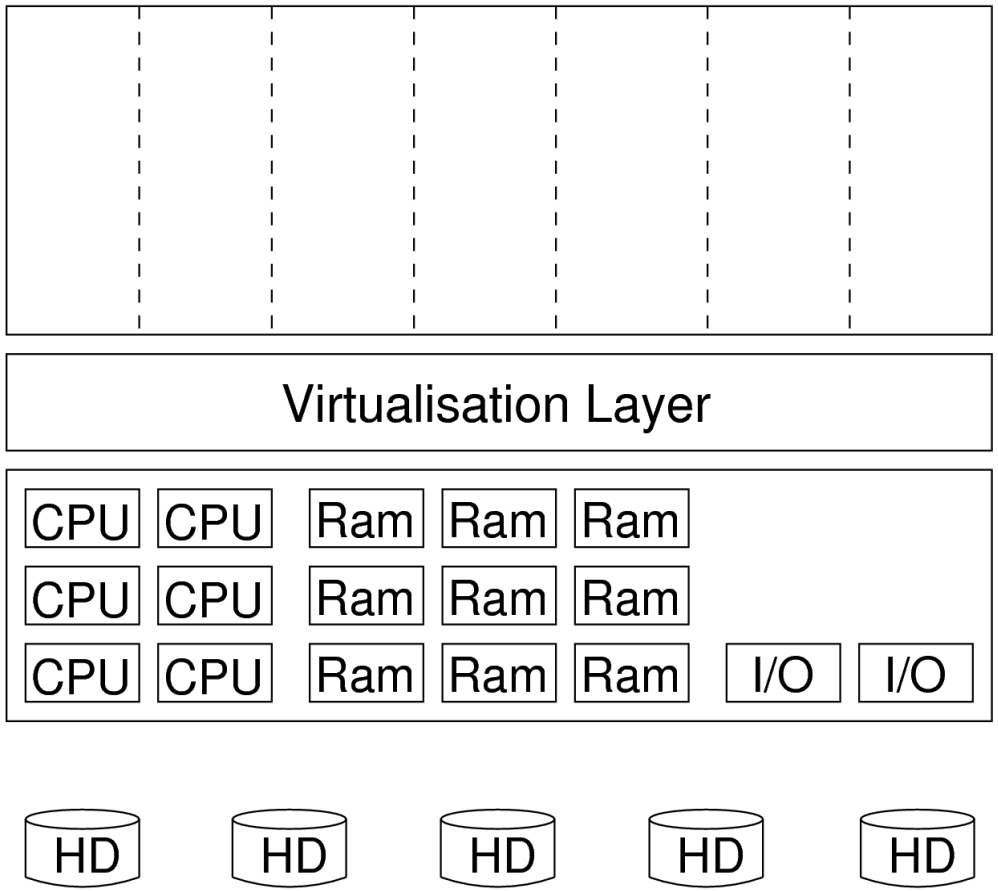
Rolf M Dietze, rmd@dietze-consulting.de

Resource Shareing



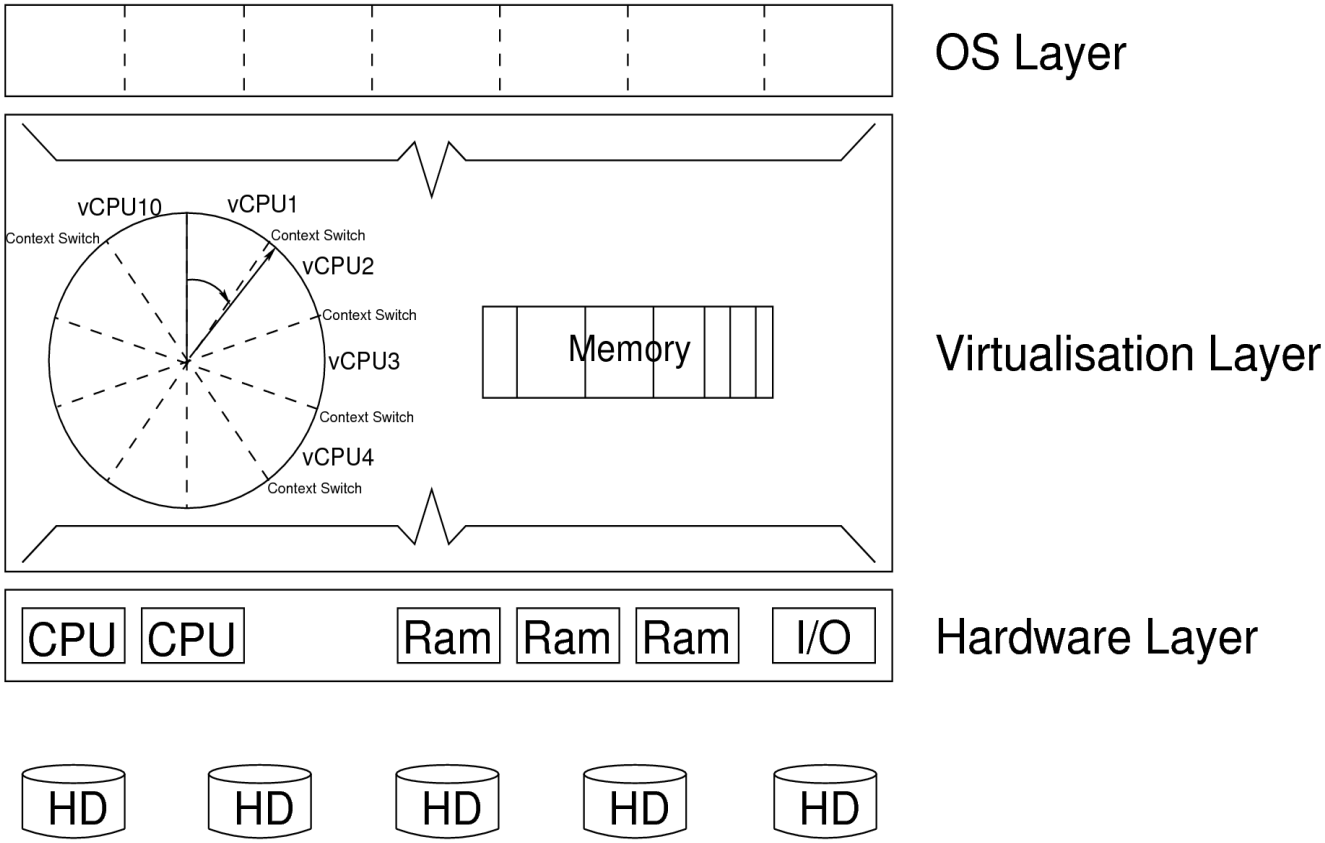
Rolf M Dietze, rmd@dietze-consulting.de

Soft Partitioning



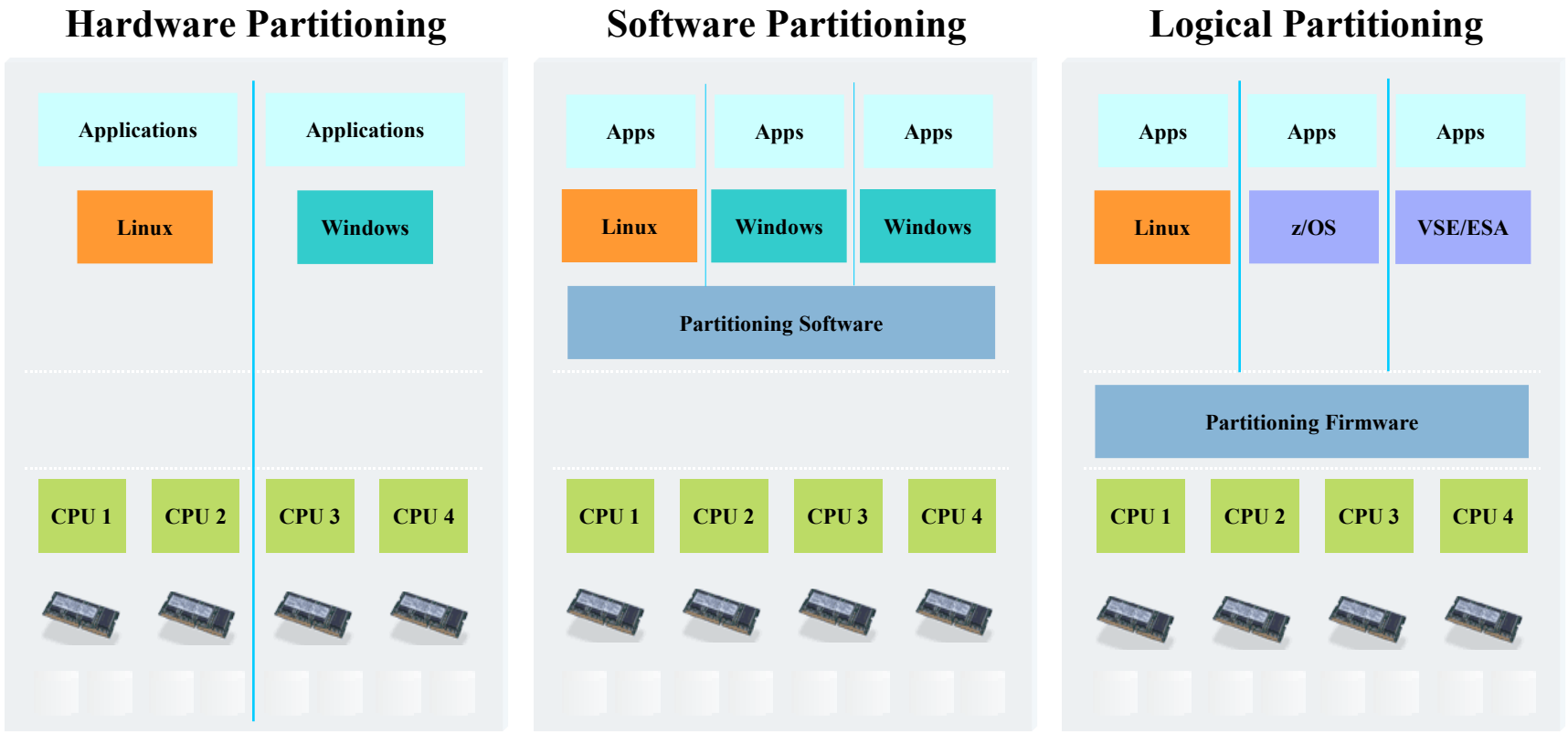
Rolf M Dietze, rmd@dietze-consulting.de

Hypervisor



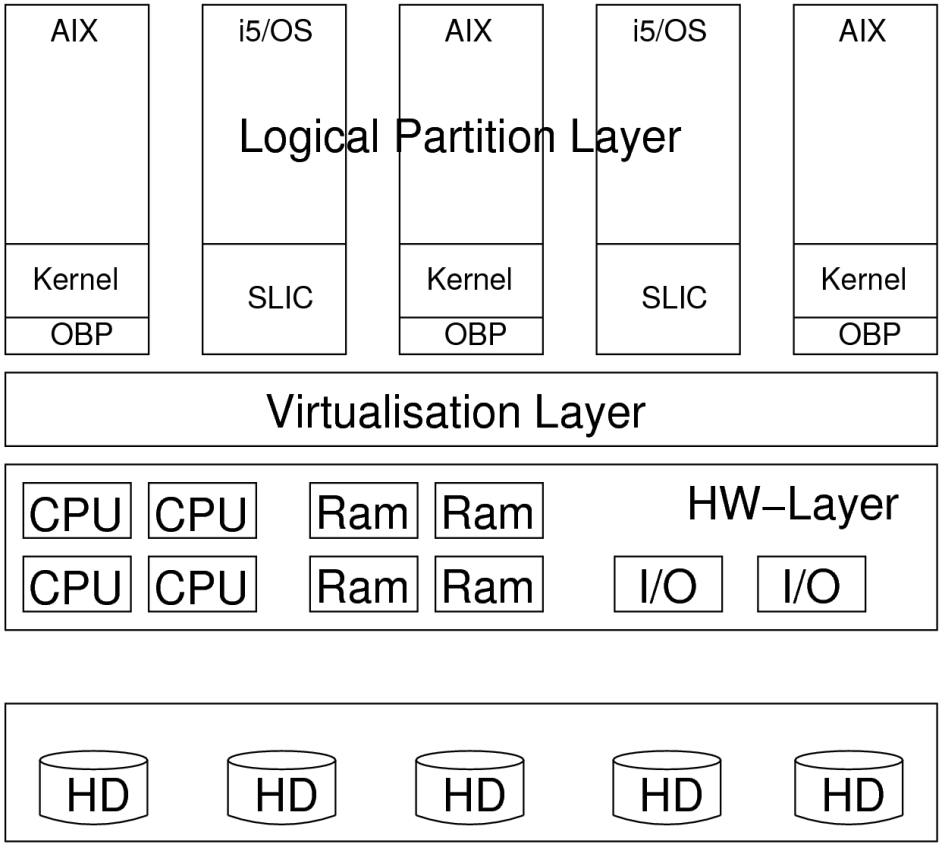
Rolf M Dietze, rmd@dietze-consulting.de

Classifying Server Partitioning



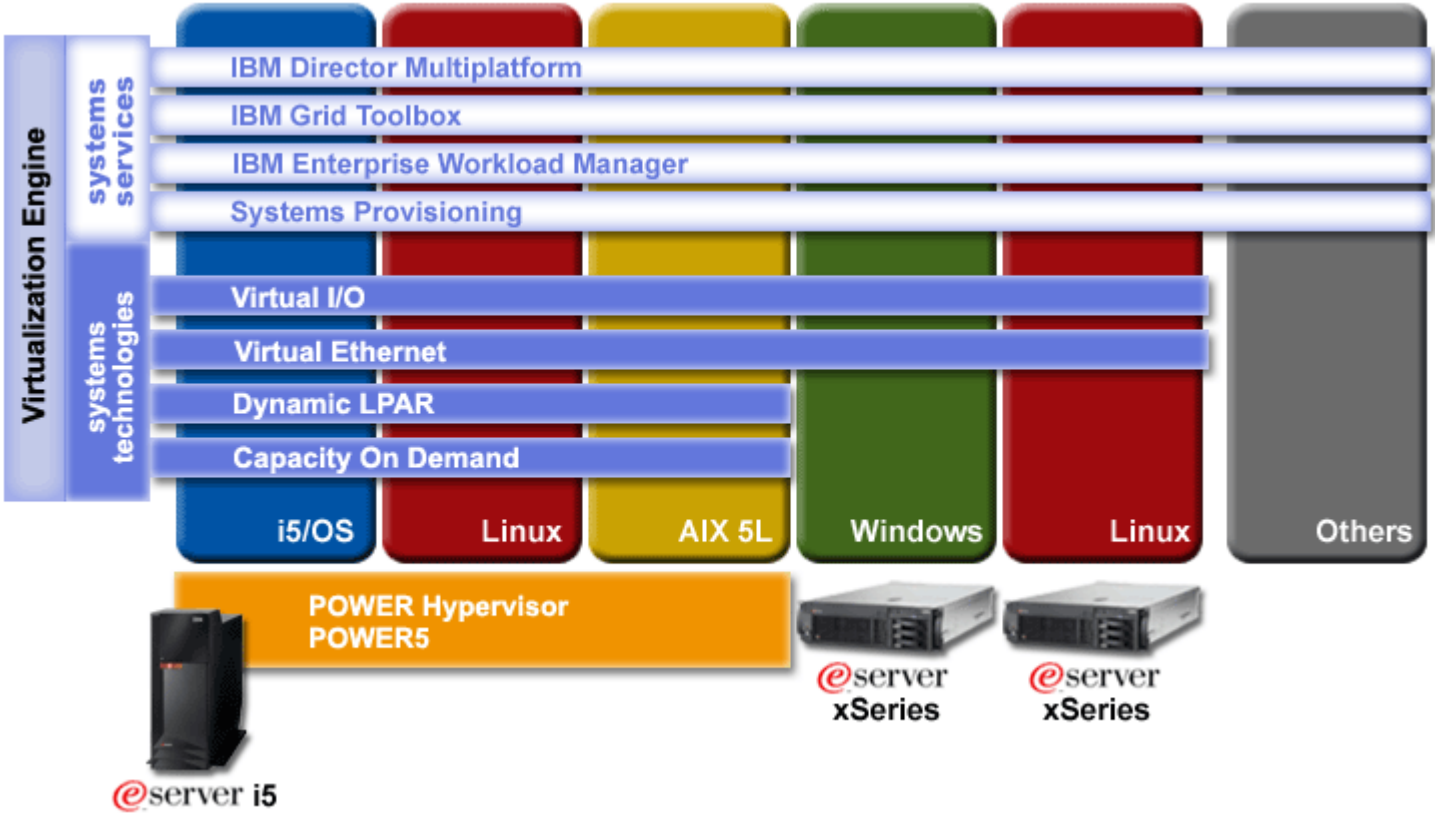
Rolf M Dietze, rmd@dietze-consulting.de

Virtualisation for Partitioning



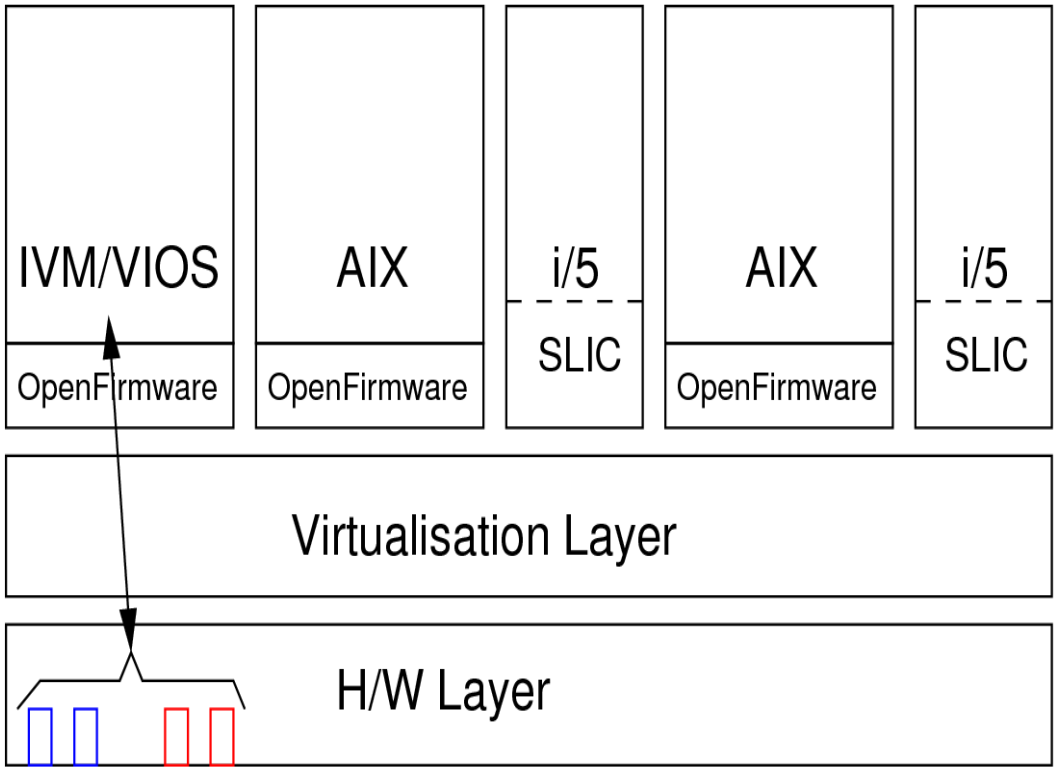
Rolf M Dietze, rmd@dietze-consulting.de

Advanced Virtualization Technologies



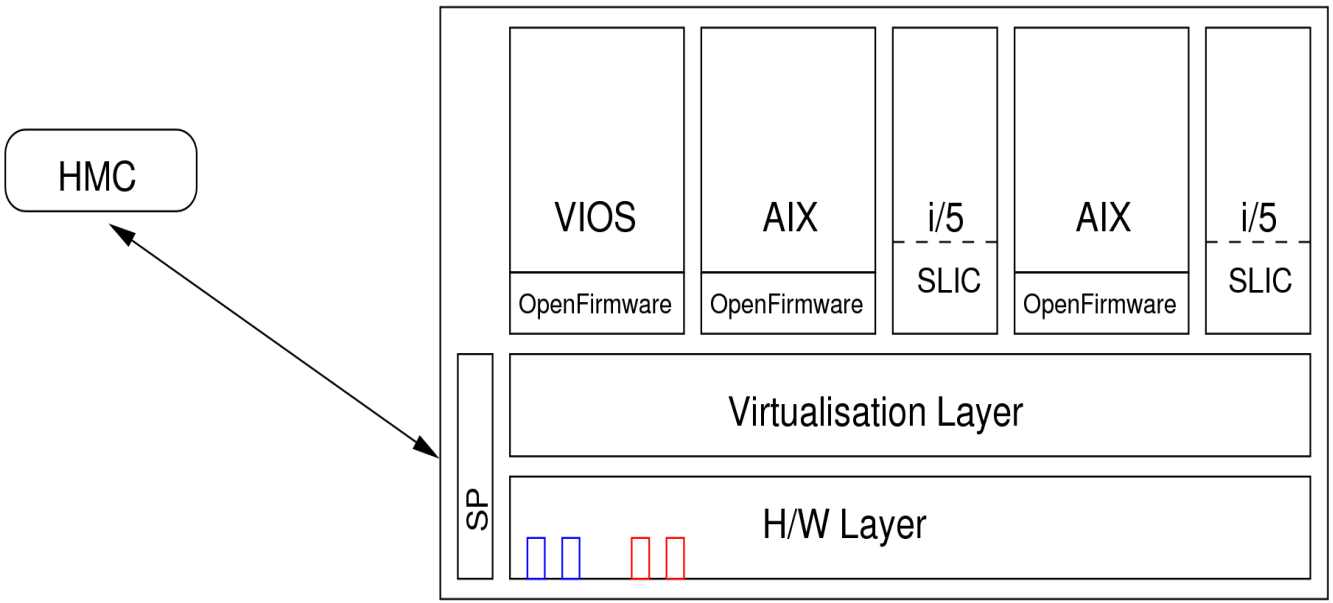
Rolf M Dietze, rmd@dietze-consulting.de

IVM/VIOS



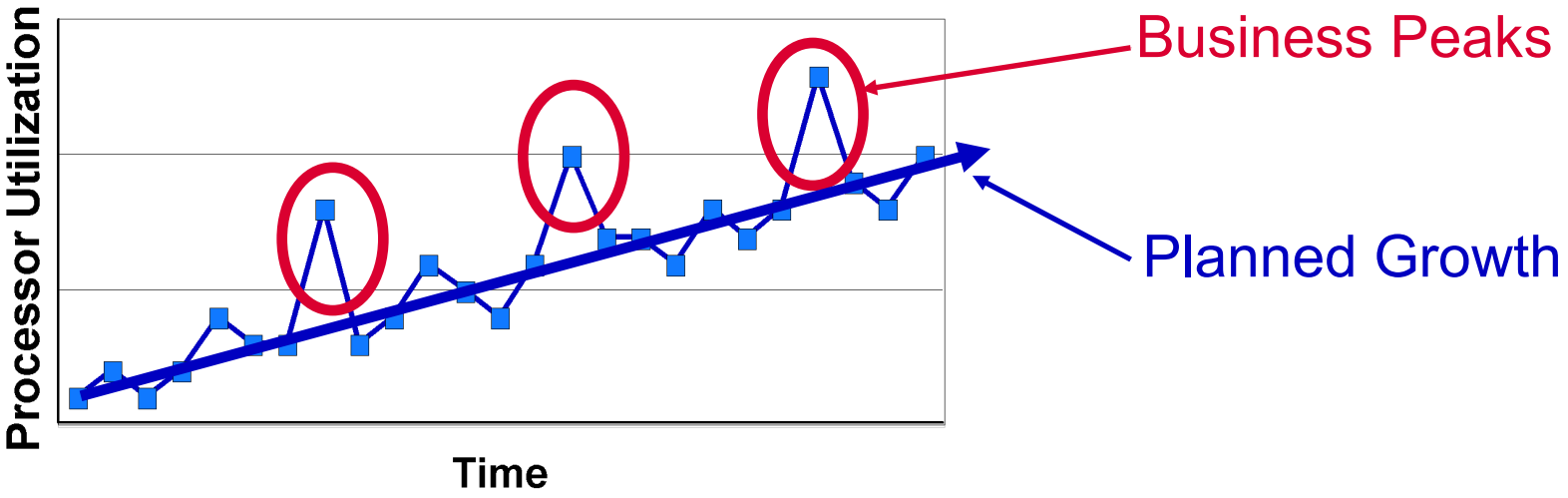
Rolf M Dietze, rmd@dietze-consulting.de

HMC + VIOS



Rolf M Dietze, rmd@dietze-consulting.de

Capacity on Demand

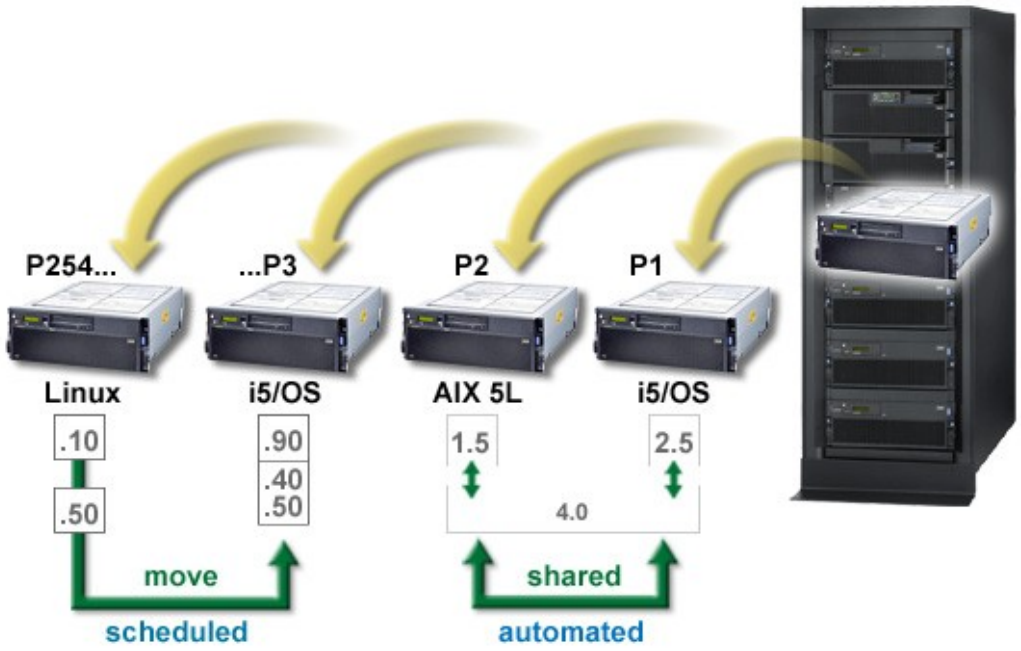


- **Permanent Capacity:** CUoD … pay when purchased (processors & memory)
- **Temporary Capacity:** On/Off CoD … pay after use (processors & memory)
- **Reserve Capacity:** CoD … pay before use (processors)
- **Trial Capacity:** CoD … no-charge for use (processors & memory)

Rolf M Dietze, rmd@dietze-consulting.de

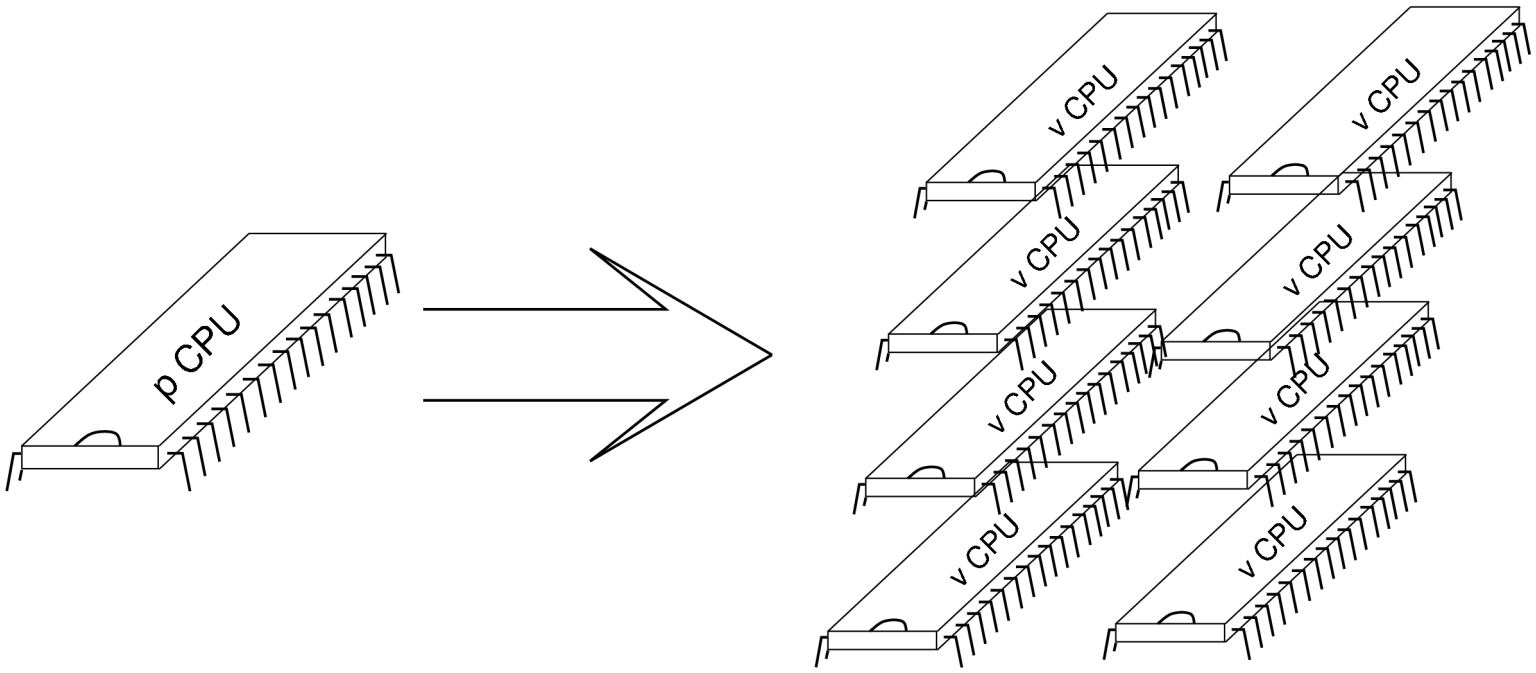
i5/OS Dynamic Logical Partitioning

- New POWER Hypervisor™ for ~ i5 supports i5/OS, AIX 5L and Linux and up to 254 partitions
  - Up to 10 Partition per processor
- Increase server utilization rates across multiple workloads
  - Dynamic resource movement
  - Automatic processor balancing with uncapped partitions



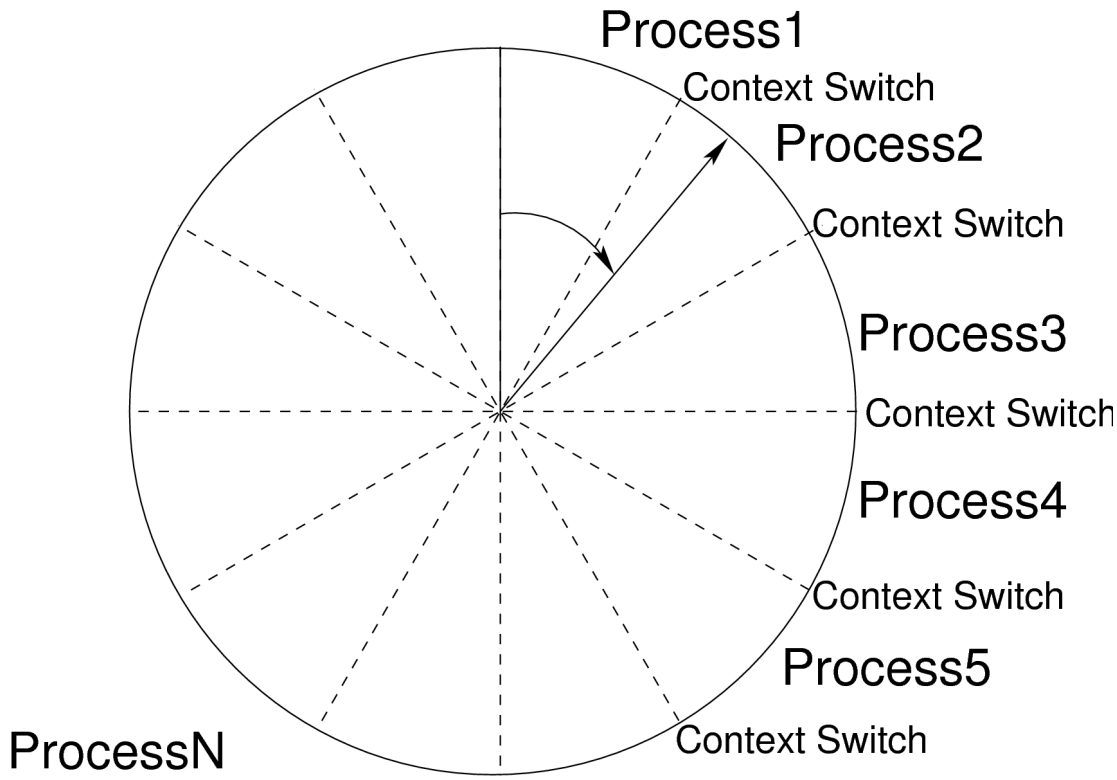
Rolf M Dietze, rmd@dietze-consulting.de

Der Weg zur logischen CPU



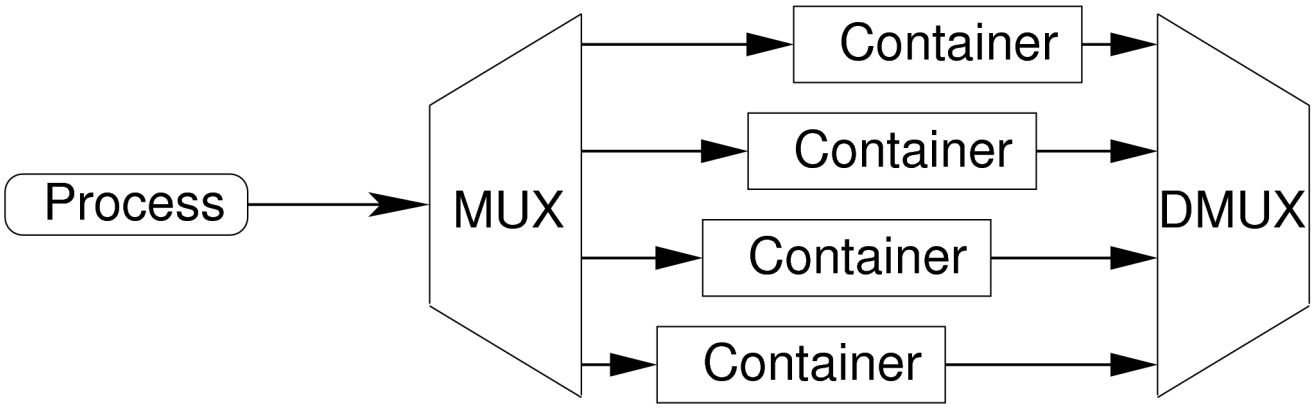
Rolf M Dietze, rmd@dietze-consulting.de

Scheduling Processes



Rolf M Dietze, rmd@dietze-consulting.de

Spreading to Containers



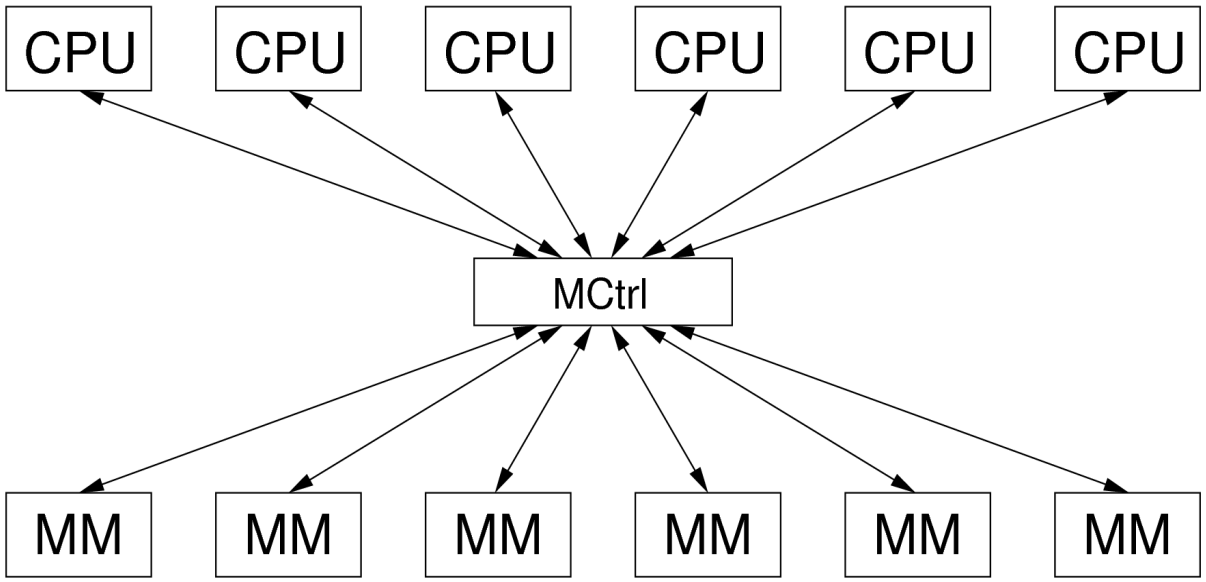
Rolf M Dietze, rmd@dietze-consulting.de

MP Konzepte

- SMP: Symmetric MultiProcessing
  - uniformer Zugriff aller CPUs auf RAM+I/O
- NUMA: Non-Uniform Memory Access
  - variable RAM-Zugriffslatenz
- CC-Numa: Cache Coherent NUMA
  - Hardwaregesteuerte Cachecoherence
- COMA: Cache Only Memory Architecture
  - Hardwaregesteuerte Replikation und Coherence
- S-COMA: Simple-COMA
  - Software-Replikation, Hardware-Coherence

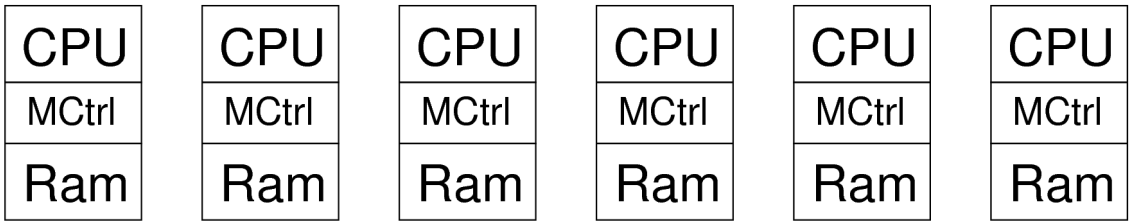
Rolf M Dietze, rmd@dietze-consulting.de

SMP Topology Restrictions

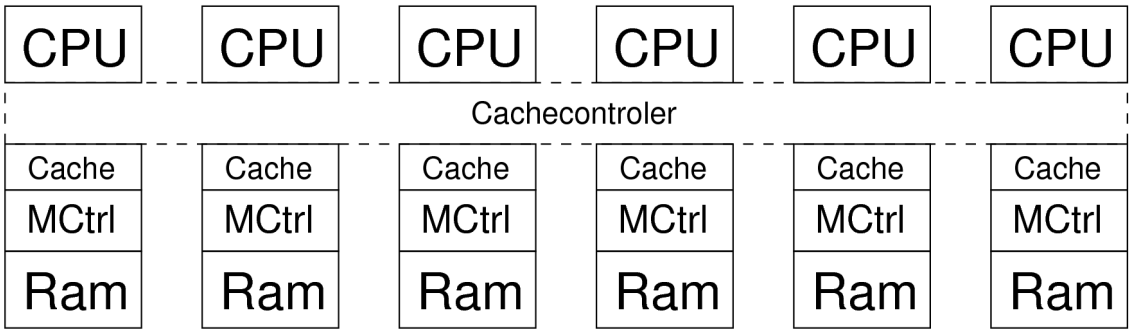


The bottleneck of the Von-Neumann Architecture

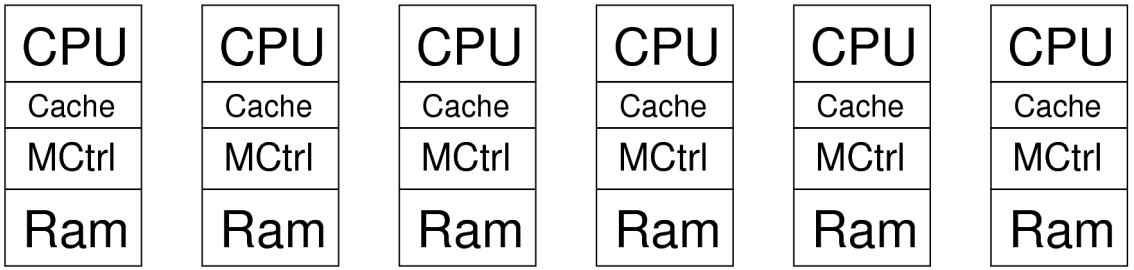
Rolf M Dietze, rmd@dietze-consulting.de



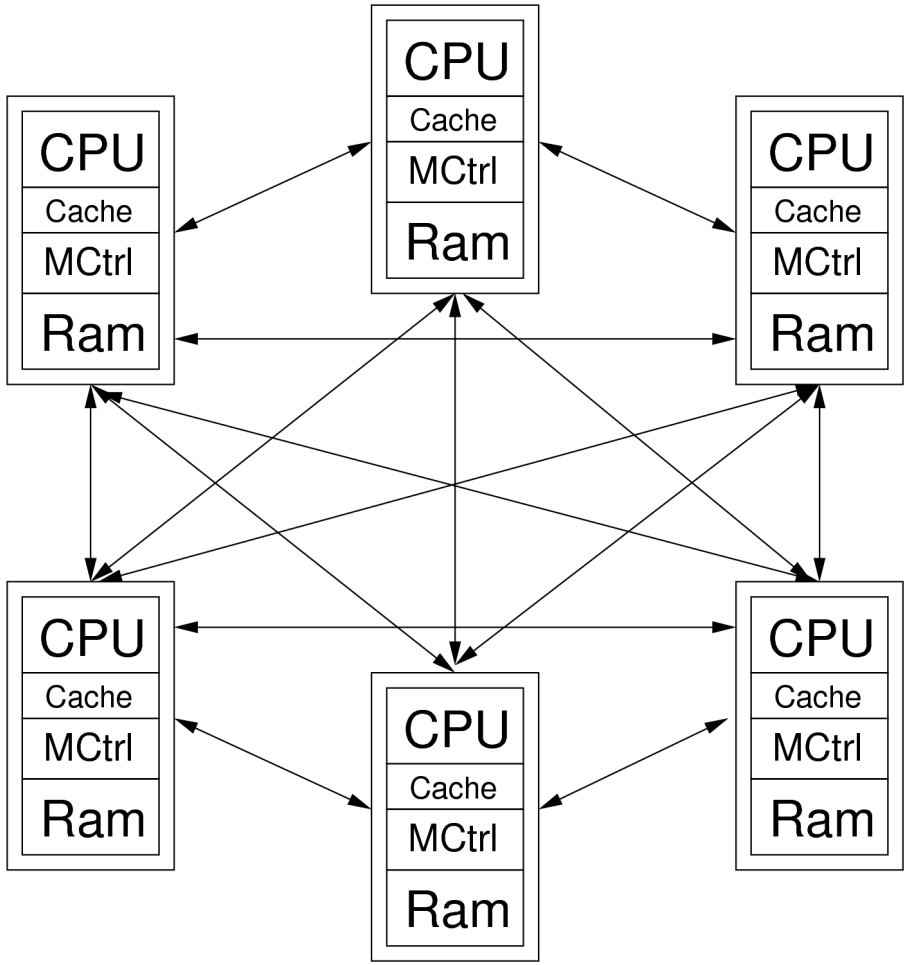
Rolf M Dietze, rmd@dietze-consulting.de



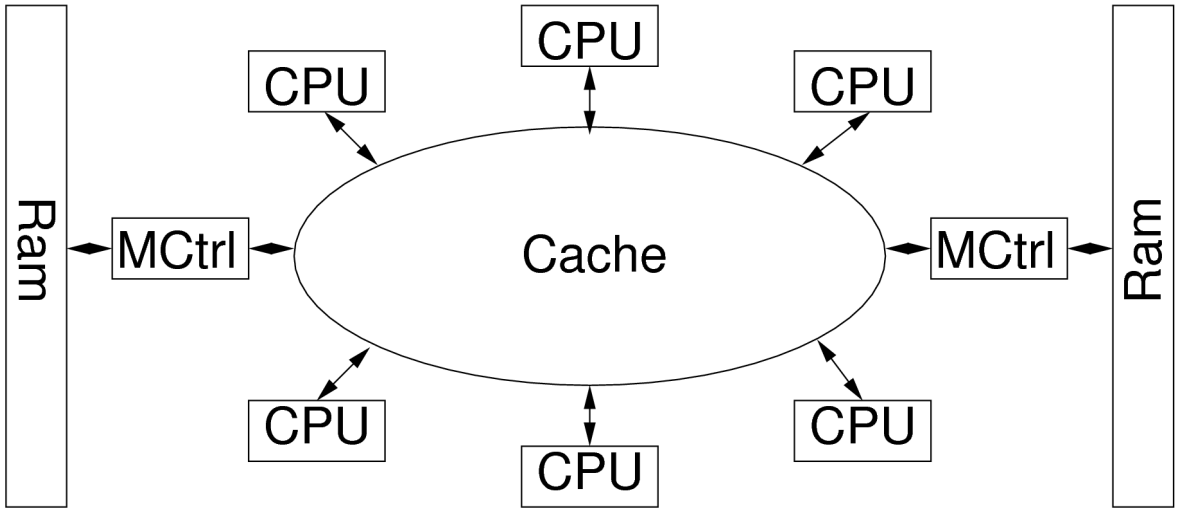
Rolf M Dietze, rmd@dietze-consulting.de



Rolf M Dietze, rmd@dietze-consulting.de

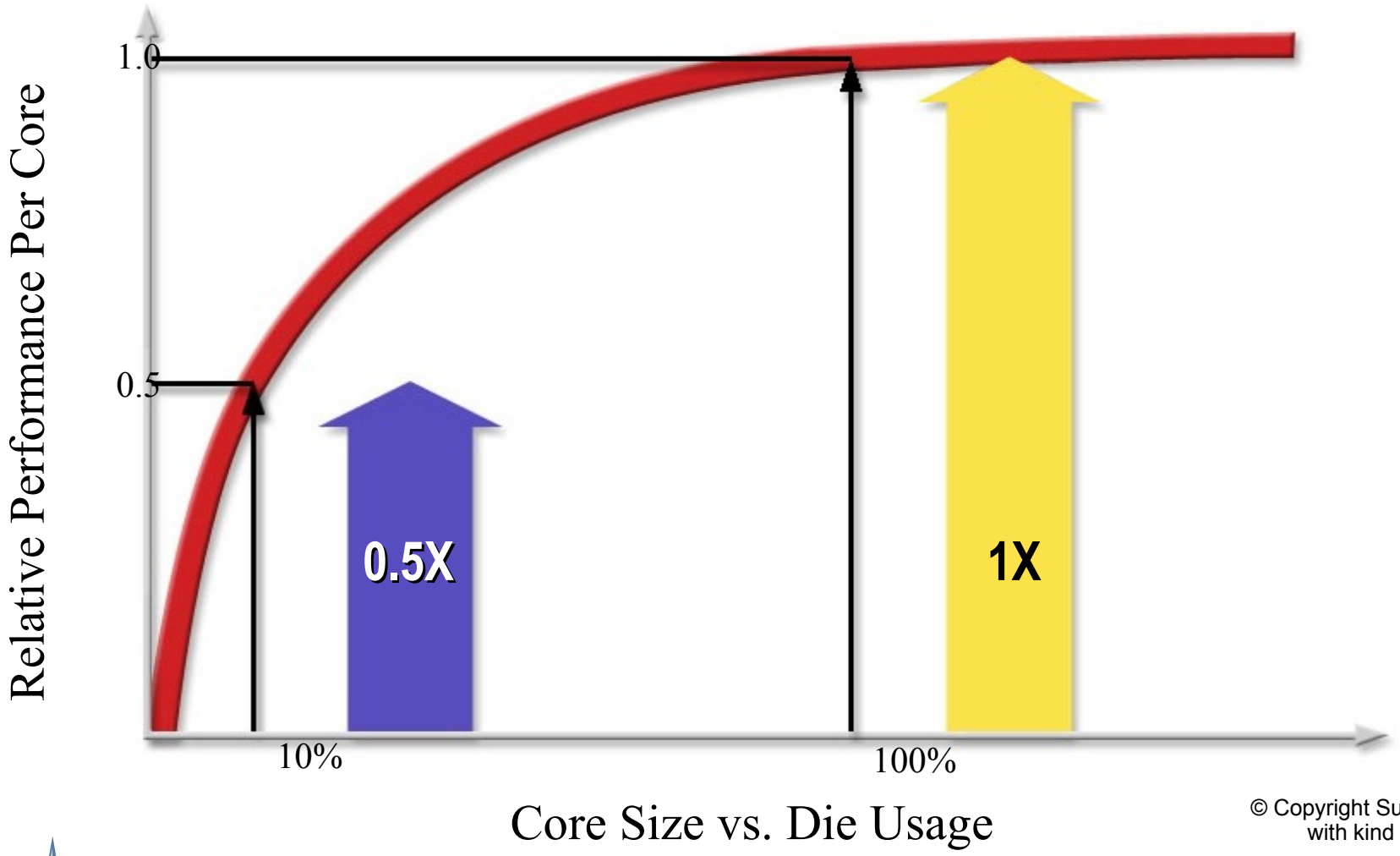


Rolf M Dietze, rmd@dietze-consulting.de

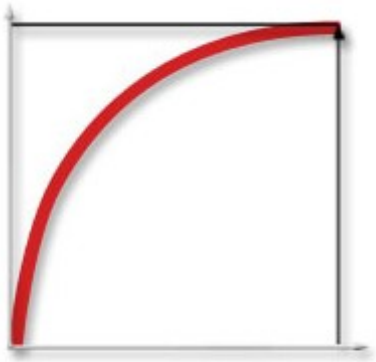


Rolf M Dietze, rmd@dietze-consulting.de

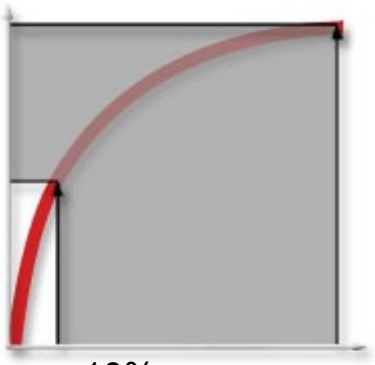
Die 80-20 Regel gilt auch für CPUs



Die Idee des Chip Multithreading



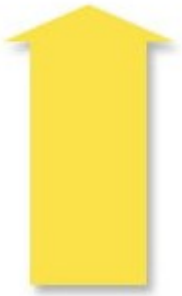
100%



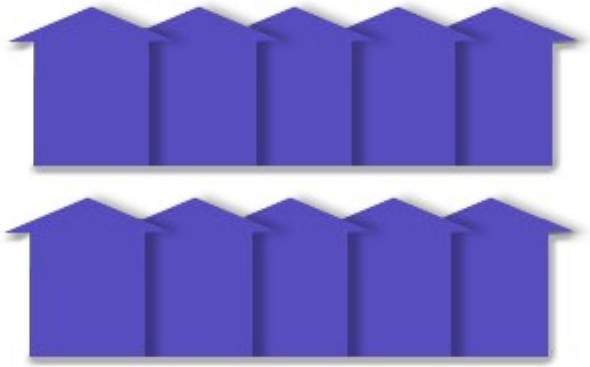
10% 100%

100% x 1 =

1x

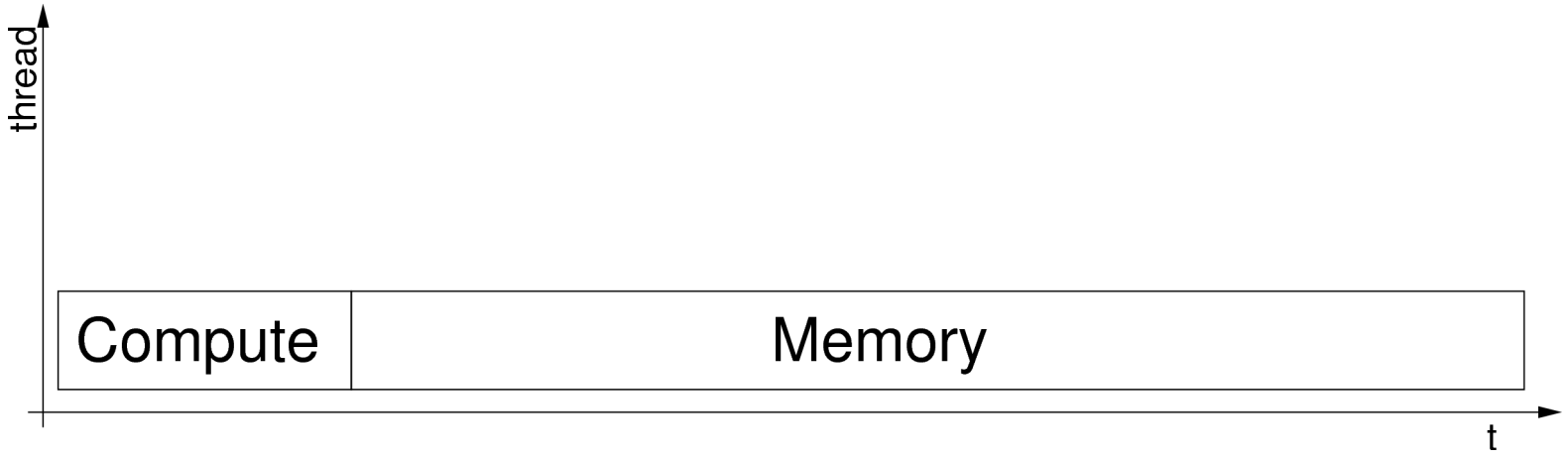


50% x 10 = 5x



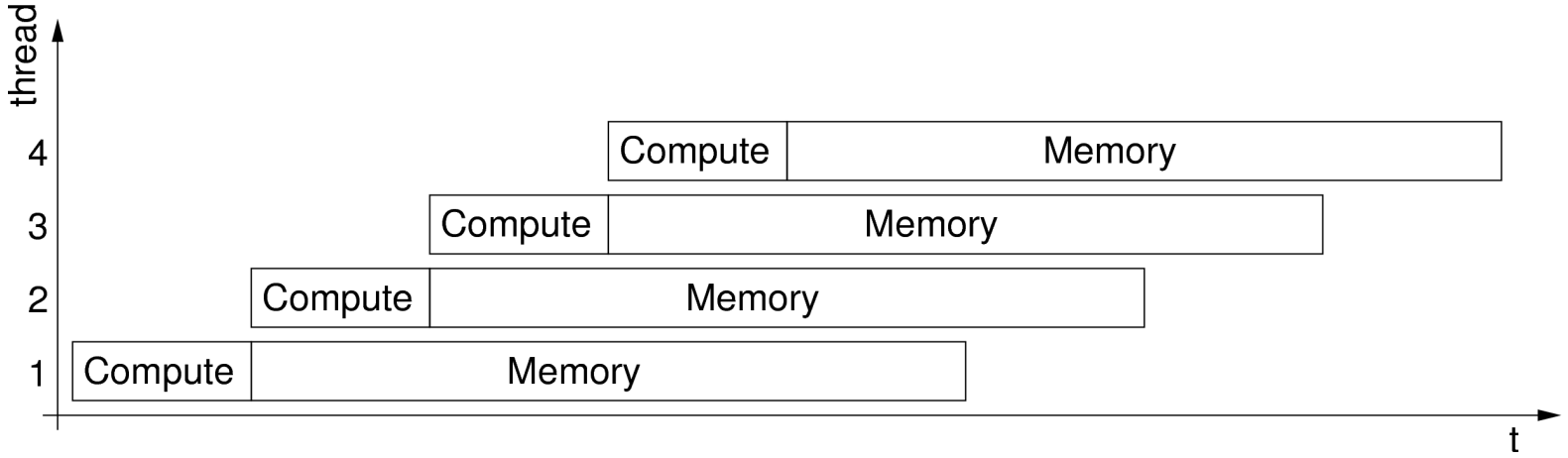
Rolf M Dietze, rmd@dietze-consulting.de

Compute- vs. Memorycycle



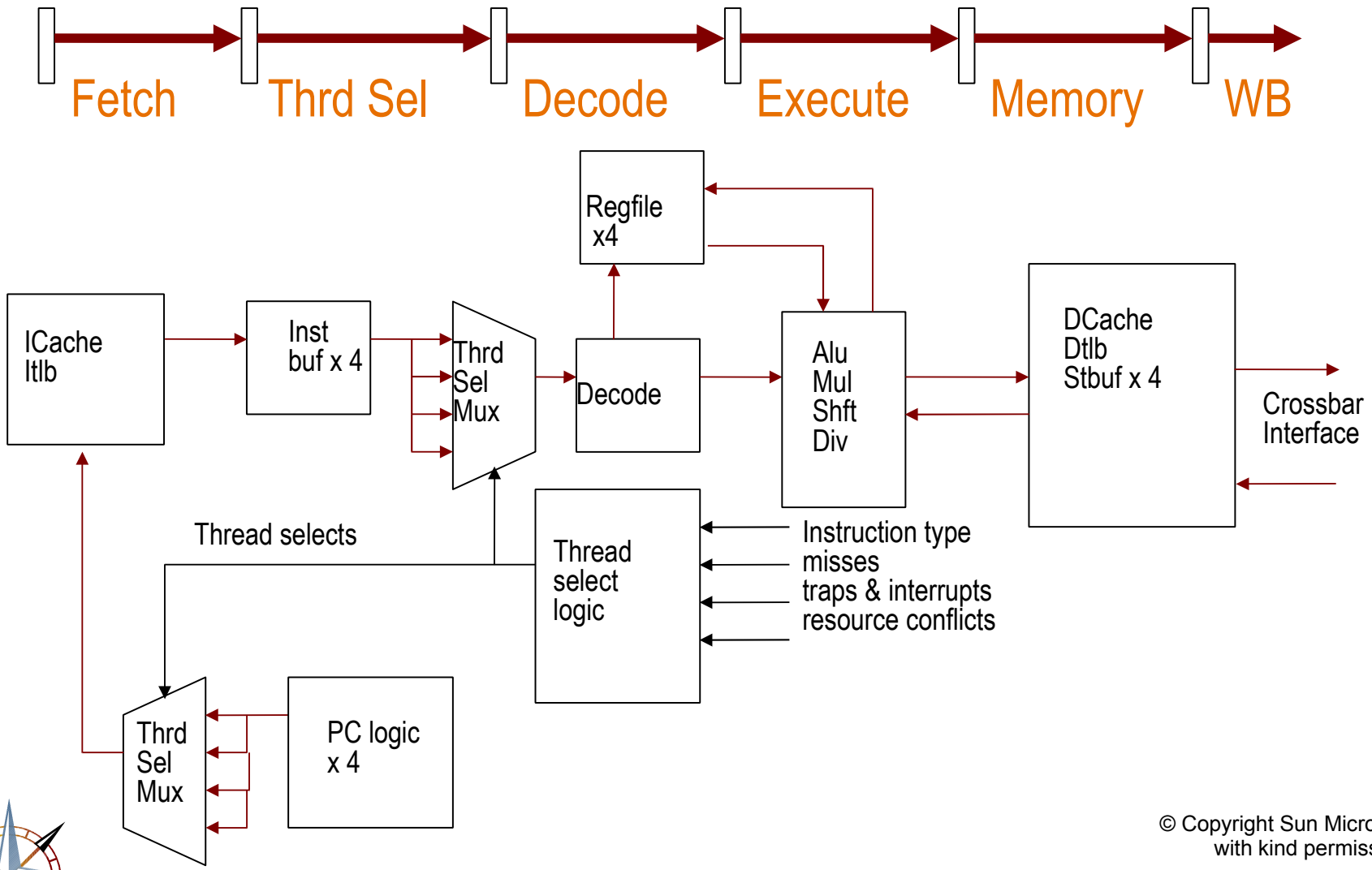
Rolf M Dietze, rmd@dietze-consulting.de

Thread CPU Model



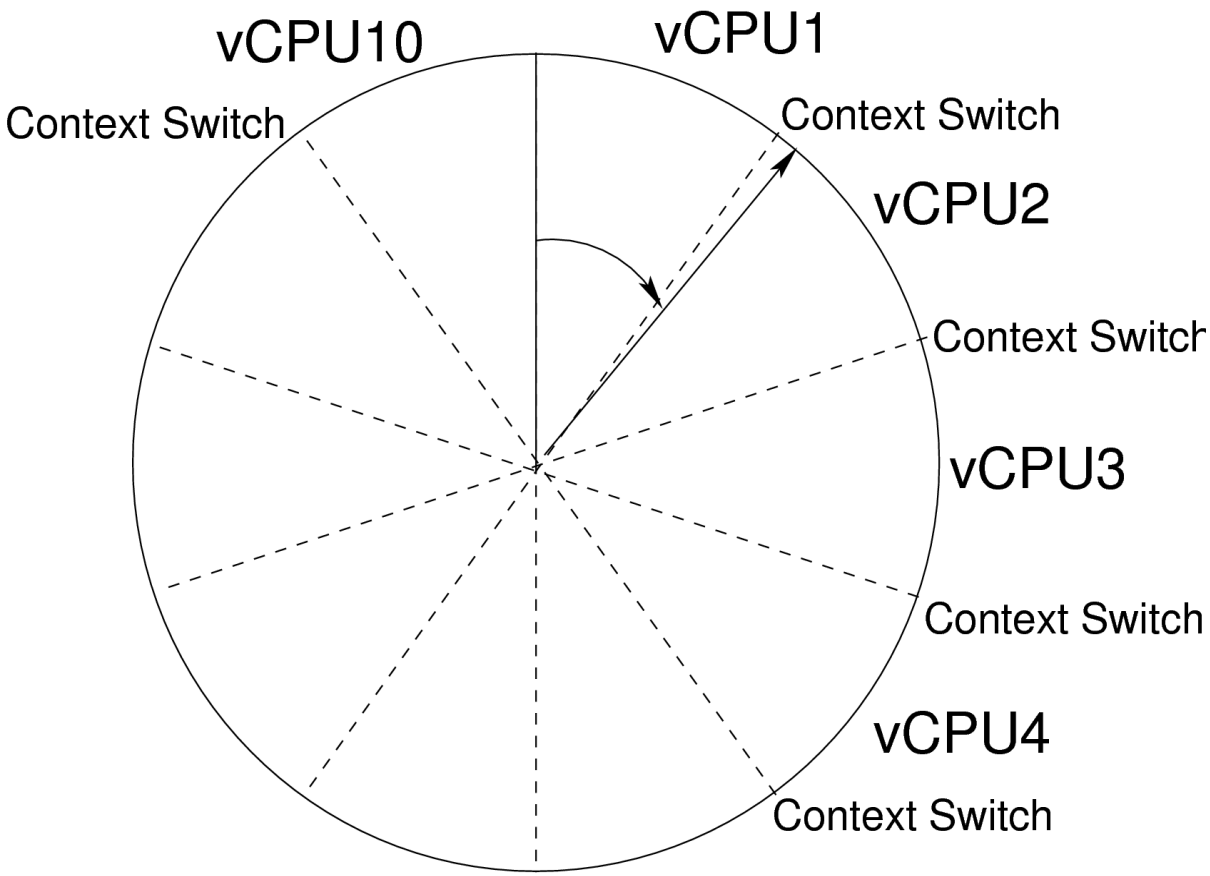
Rolf M Dietze, rmd@dietze-consulting.de

UltraSPARC T1 Pipeline Diagramm



Rolf M Dietze, rmd@dietze-consulting.de

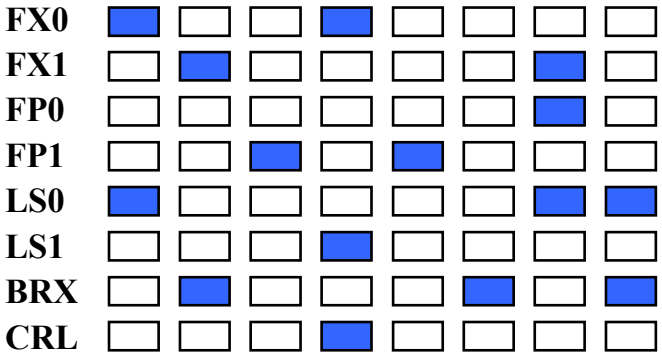
Timeslice CPU Model



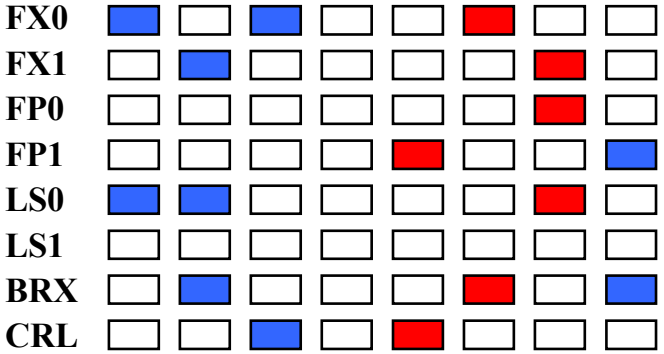
Rolf M Dietze, rmd@dietze-consulting.de

Multithreading Evolution

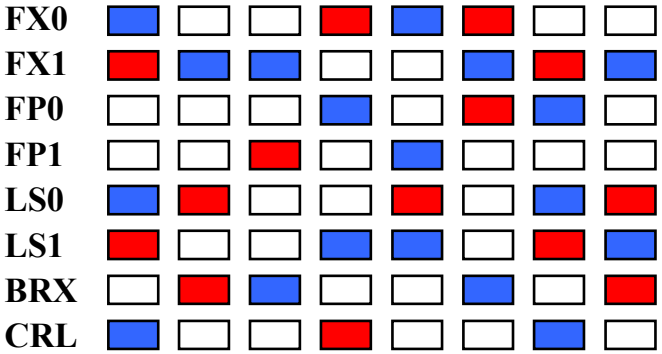
Single thread Out of Order



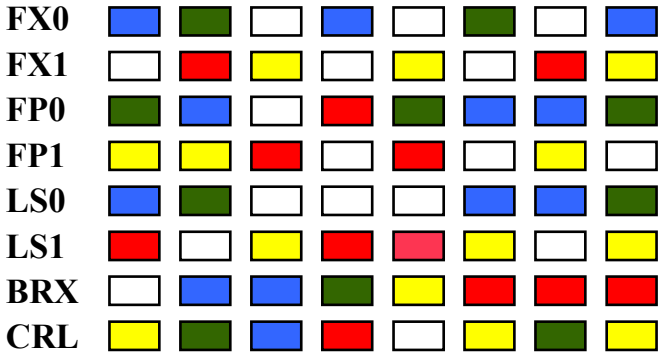
S80 Hardware Multi-thread



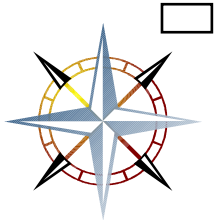
POWER5 2 Way SMT



POWER7 4 Way SMT



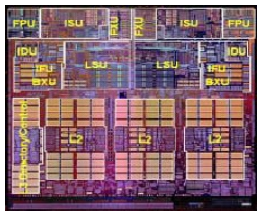
No Thread Executing Thread 0 Executing Thread 1 Executing Thread 2 Executing Thread 3 Executing



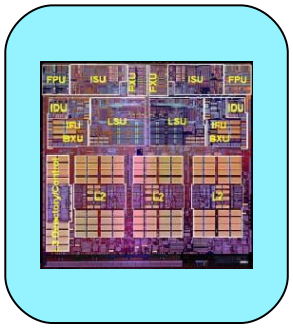
Rolf M Dietze, rmd@dietze-consulting.de

Understanding Shared Processors

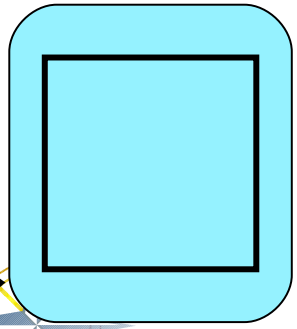
To understand Processing Units – there are four main concepts



1. One single processor is equivalent to 1.00 Processing Units, 2 Processors = 2.00 Processing Units, etc.
   0.5 processing units is NOT same as half a processor.



2. Shared processor pool. A processor must live in the shared processor pool (now the default) to become Processing units.

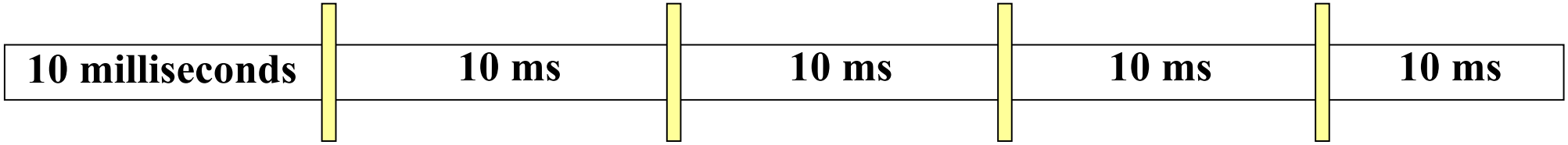


3. Virtual Processor – how many processors do you want the partition to be able to use (run jobs/threads on) simultaneously. It's also the number of processors that the operating system thinks it has to use.

Rolf M Dietze, rmd@dietze-consulting.de

10 Milliseconds Time Slice

4. The iSeries processors run on 10 ms time slices

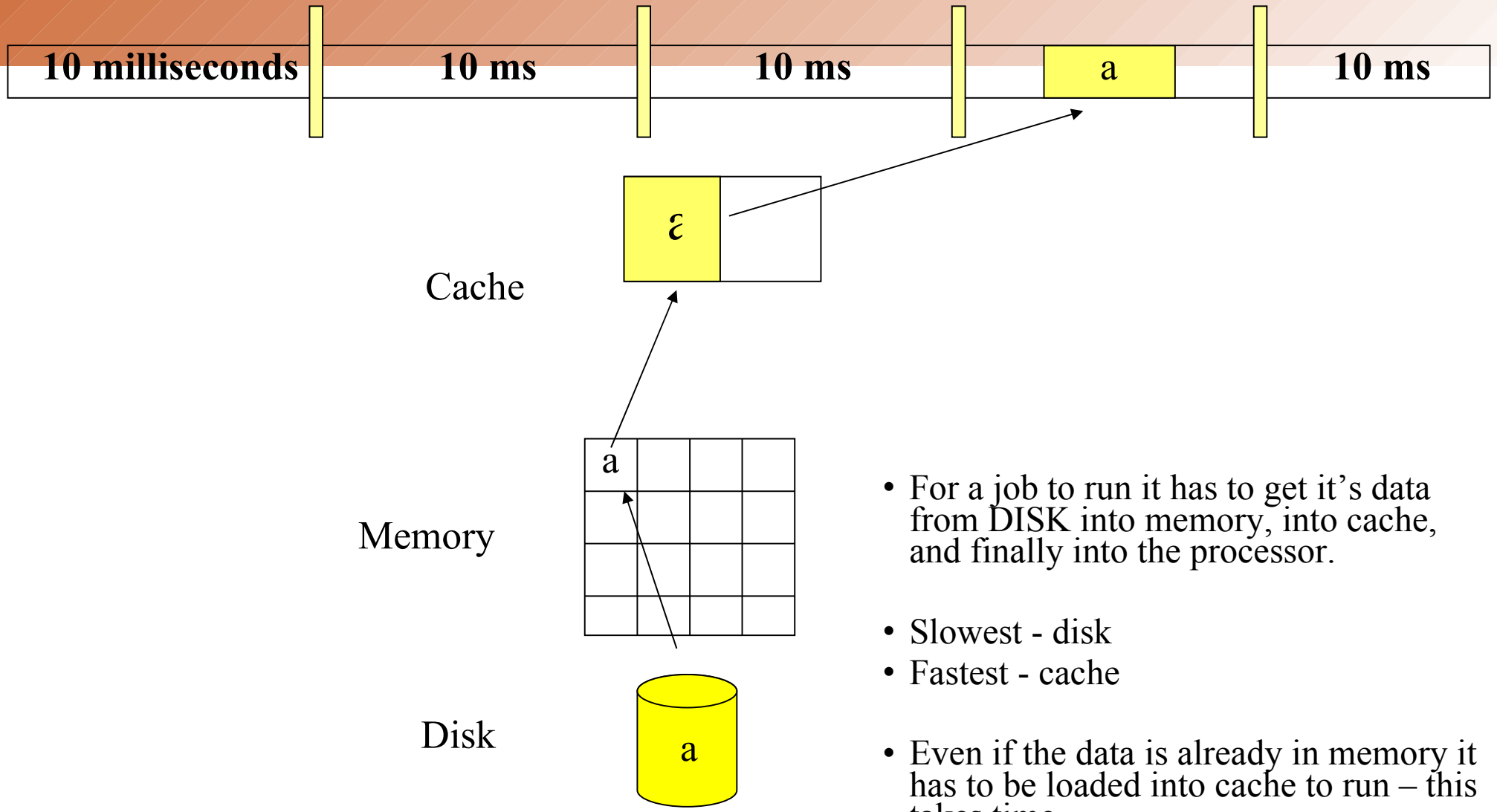


Each Processor use is allocated within a 10 ms Cycle

- A partition's use of a processor is limited to its allocation during each 10 ms cycle.
- For example, 80% of a processor (.80 Processing Units) yields up to 8 ms of processing time out of this 10 ms time slice. It also yields .8 X CPW rating of the processor.
- Every 10 ms this cycle repeats itself.

Rolf M Dietze, rmd@dietze-consulting.de

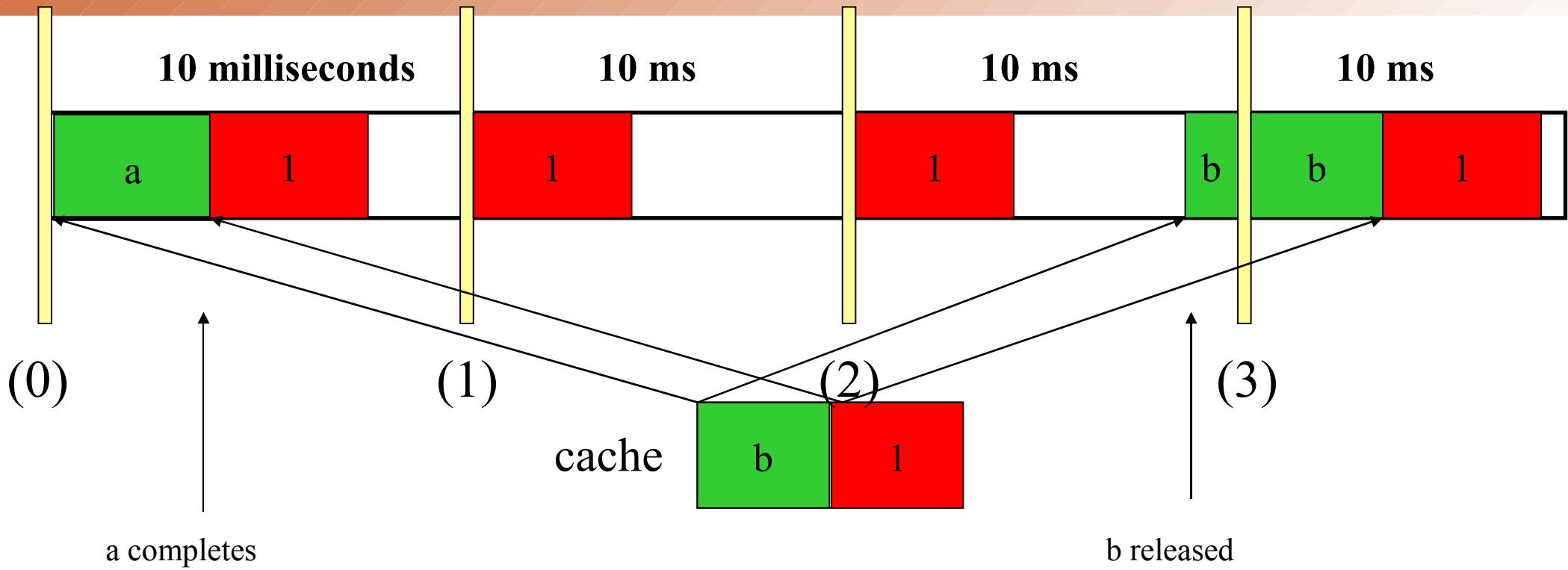
How Does a Job Get Into the Processor?



- For a job to run it has to get it's data from DISK into memory, into cache, and finally into the processor.
- Slowest - disk
- Fastest - cache
- Even if the data is already in memory it has to be loaded into cache to run – this takes time.

Rolf M Dietze, rmd@dietze-consulting.de

Example of Two Partitions Sharing a Processor ("capped")



Partition dog jobs a,b,c allocated .6 Processing Units

Partition cat jobs 1,2,3 allocated .4 Processing Units

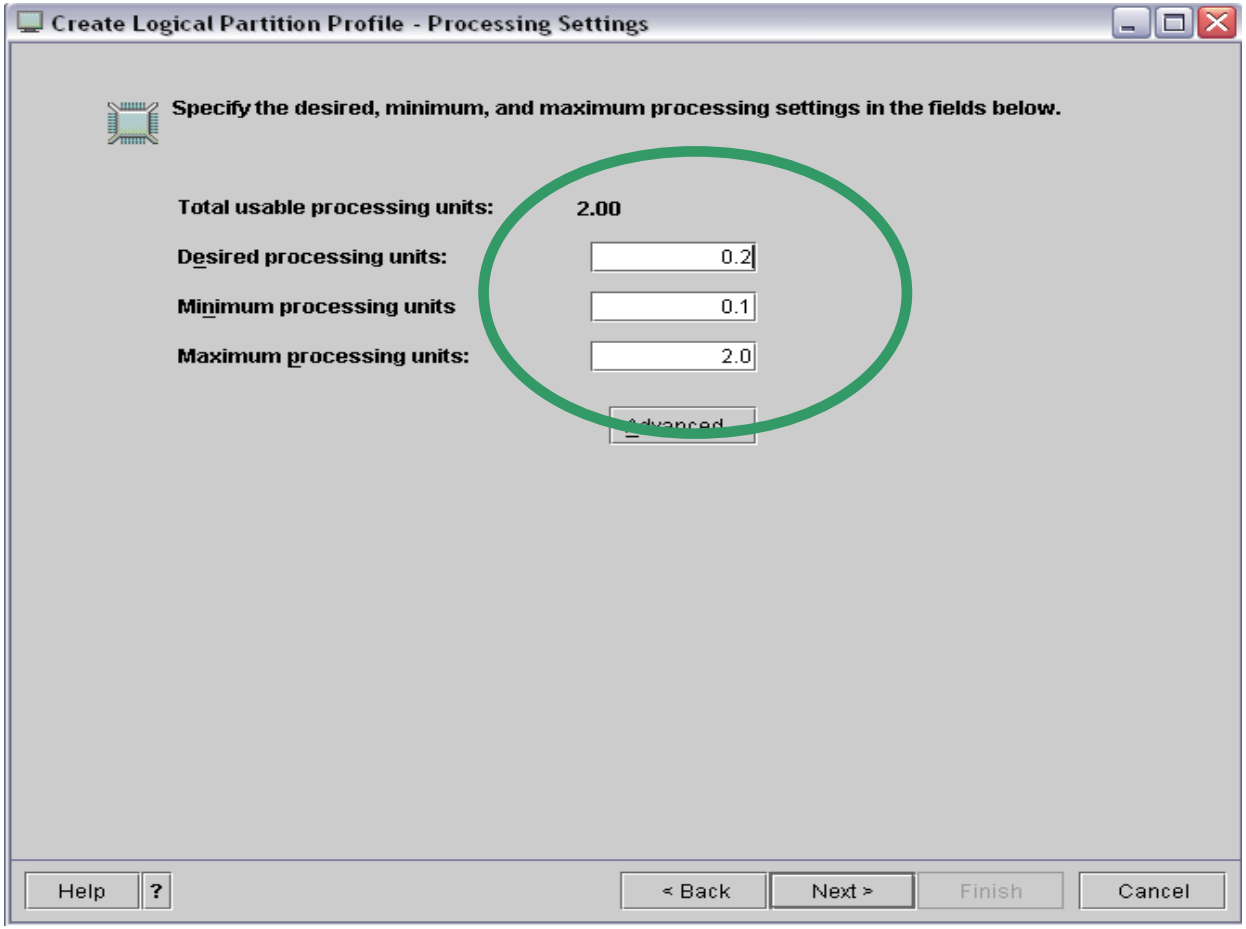
Rolf M Dietze, rmd@dietze-consulting.de

Potential Shared Processor Penalty

- There is a potential for a performance penalty (from 0 to 10%) when using shared processors, due to:
  - Increasing the possibility that jobs won't complete, and
  - Having to be redispatch and potentially reload cache, and
  - Increasing the chance of a cache miss
- Reduce the chance for processor and memory affinity
- The POWER Hypervisor overhead of:
  - Managing multiple processors
  - Tracking each partitions use of its allotted milliseconds
  - Managing time slices from multiple partitions
- All of the above are affected by how you allocate your virtual processors – next couple of foils

Rolf M Dietze, rmd@dietze-consulting.de

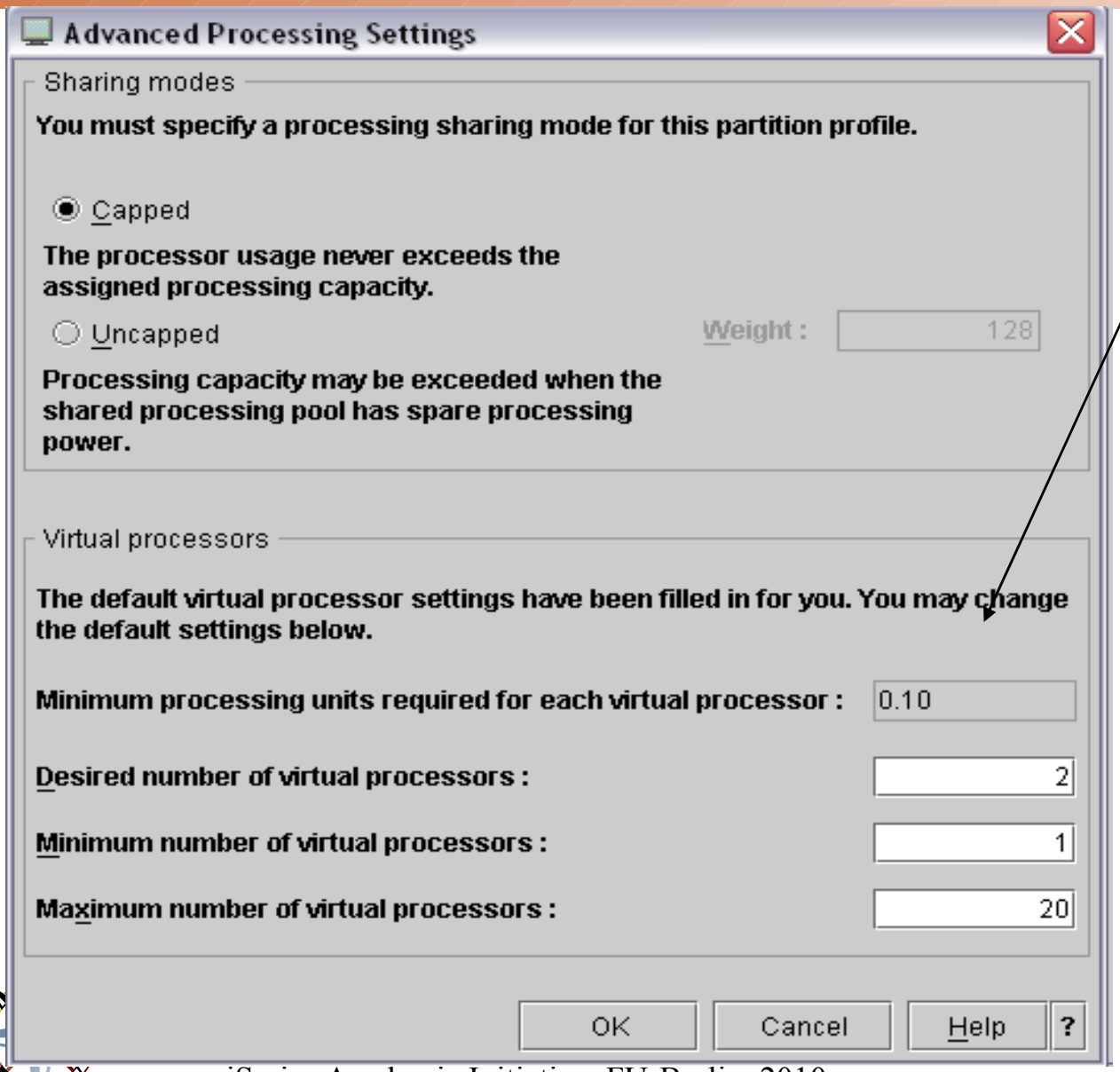
Desired Minimum/Maximum Processing Units



- How about 0.2 Processing Units
- Minimum of .1
- Maximum of 2.00
- Select Advanced

Rolf M Dietze, rmd@dietze-consulting.de

Capped Partitions

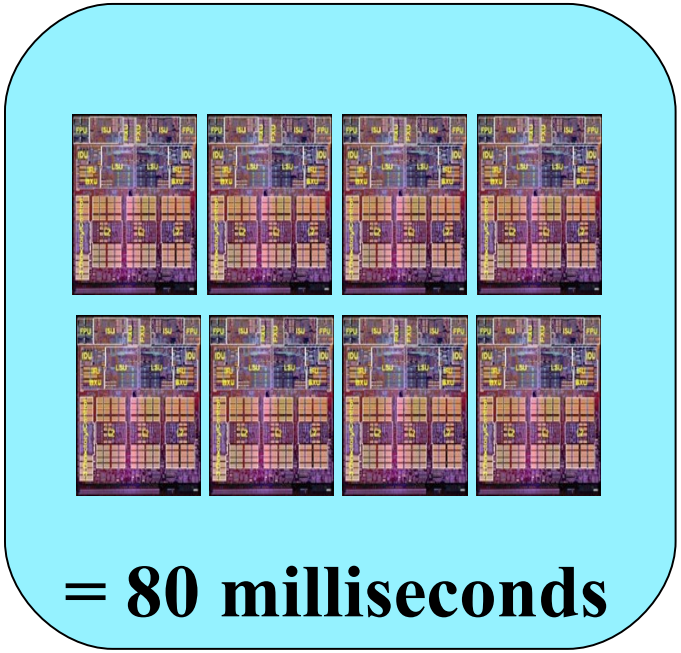


- First time through we select Capped
- You can't have less than .10 processing units per virtual processor
- I'll allocate two virtuals for my .2 PUs
- What's a virtual processor?

Rolf M Dietze, rmd@dietze-consulting.de

Introduction to Virtual Processors

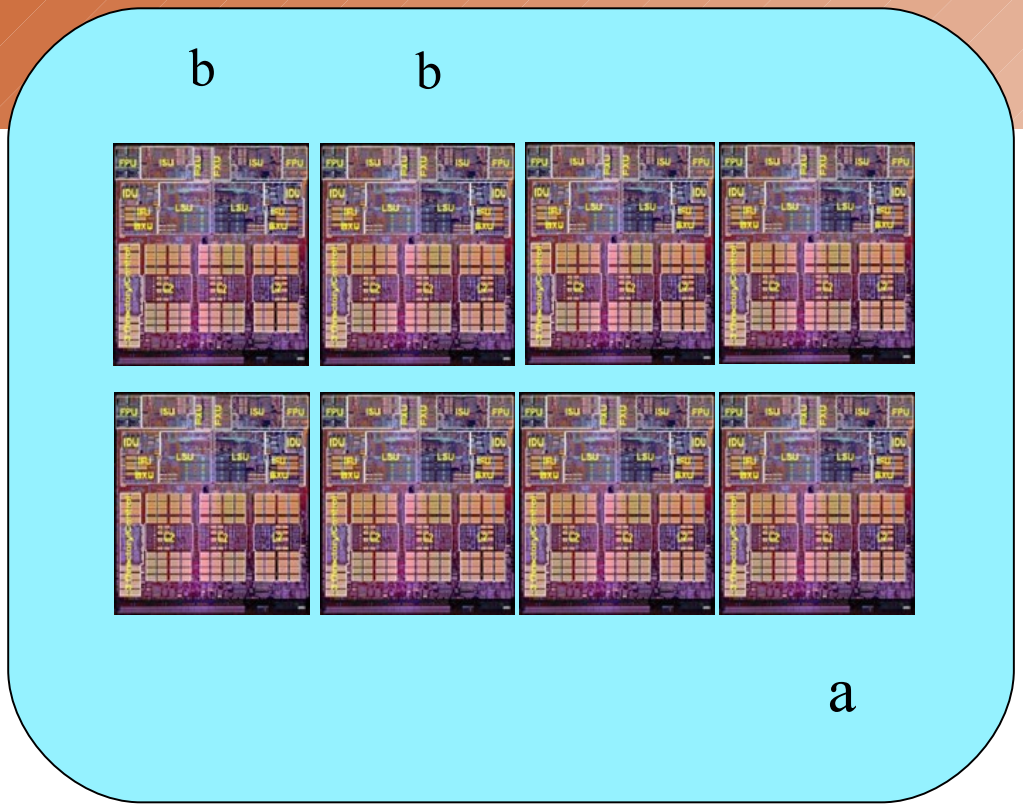
For every 10 milliseconds of wall clock time each processor in the shared pool is capable of 10 milliseconds of processing time

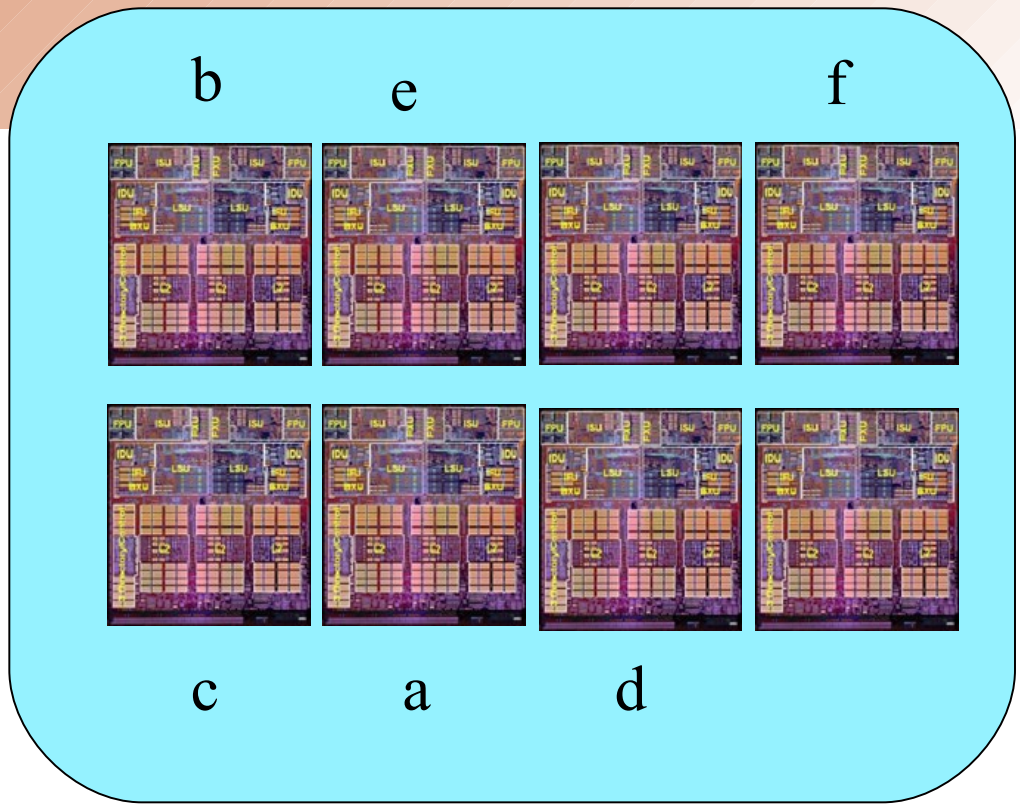


- If you give Partition x .5 processing units it could use (up to) 5 milliseconds of processing time – capped. (more on capped soon)
- But you have ABSOLUTELY no control over which processors your jobs/threads run on
- All you CAN control is how many of the processors in the pool, your jobs/threads do run on (potentially) simultaneously, via Virtual Processors

Rolf M Dietze, rmd@dietze-consulting.de

Virtual Processors - Capped





P1 1.5 processing unit default of **2** virtual processors – max of **15 milliseconds** – capped. Each job potentially could get 7.5 milliseconds (15/2 = 7.5)

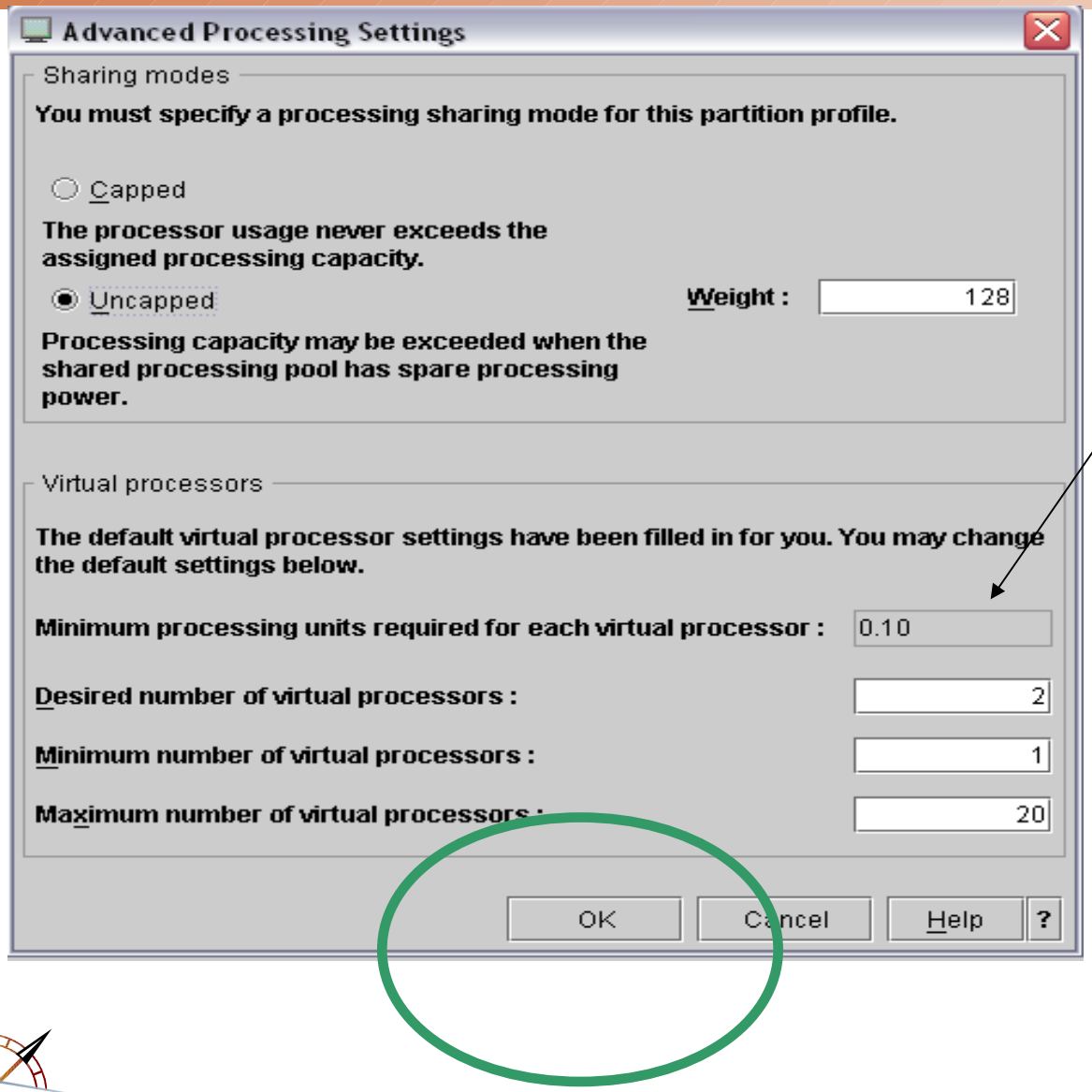
P2 1.5 processing unit but using **6** virtual processors – max of **15 milliseconds** – capped. But if all 6 Jobs ran at same time each may get no more than 2.5 milliseconds per job. (15/6 = 2.5)

Uncapped - Introduction

- As of IBM eServer i5/OS, and POWER5-based servers, it is now possible, by using the uncapped mode, to use more milliseconds than are allocated to a partition.
- An uncapped partition can use excess pool Processing Units.
- But even an uncapped partition could still be limited by setting the number of virtual processors too low. The number of processors it can use simultaneously is still limited by the number of virtual processors assigned.
- As of Power5 it's also possible to allocate more virtual processors than there are processors in the pool, since the actual number of processors in the pool is a 'floating' number. However, you still cannot allocate less than 1 ms (.10 PUs) per processor per job (virtual processor). For example, .5 PUs and 6 virtuals is a dog that doesn't hunt. 5 (milliseconds)/6 (jobs) < 1 milliseconds per job.

Rolf M Dietze, rmd@dietze-consulting.de

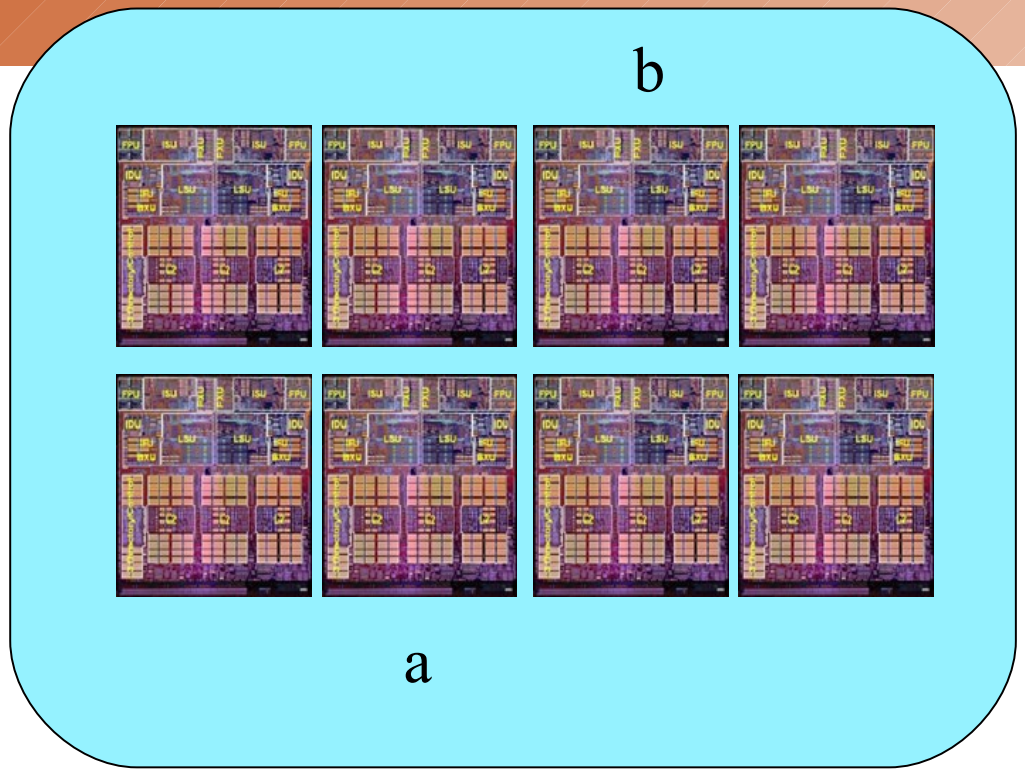
Uncapped - Configuring

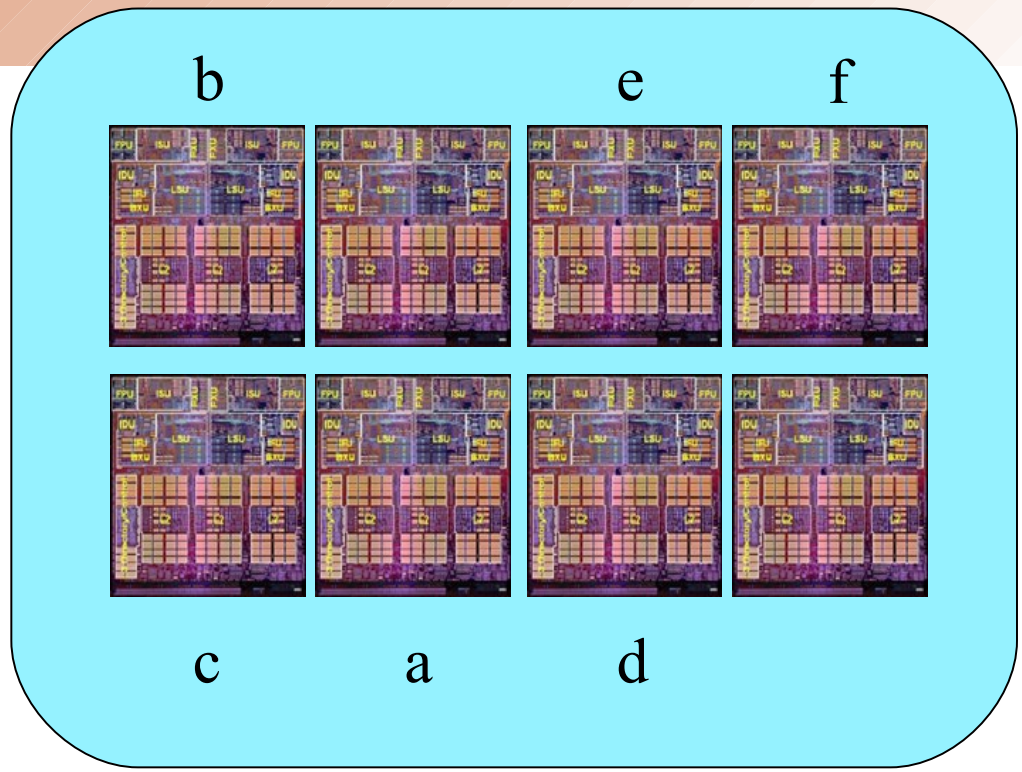


- This time we deal with uncapped
- You can't have any less than .10 processing units per virtual processor
- Allocate two virtuals for my .2 PUs
- Select OK

Rolf M Dietze, rmd@dietze-consulting.de

Virtual Processors (Limited) – Uncapped

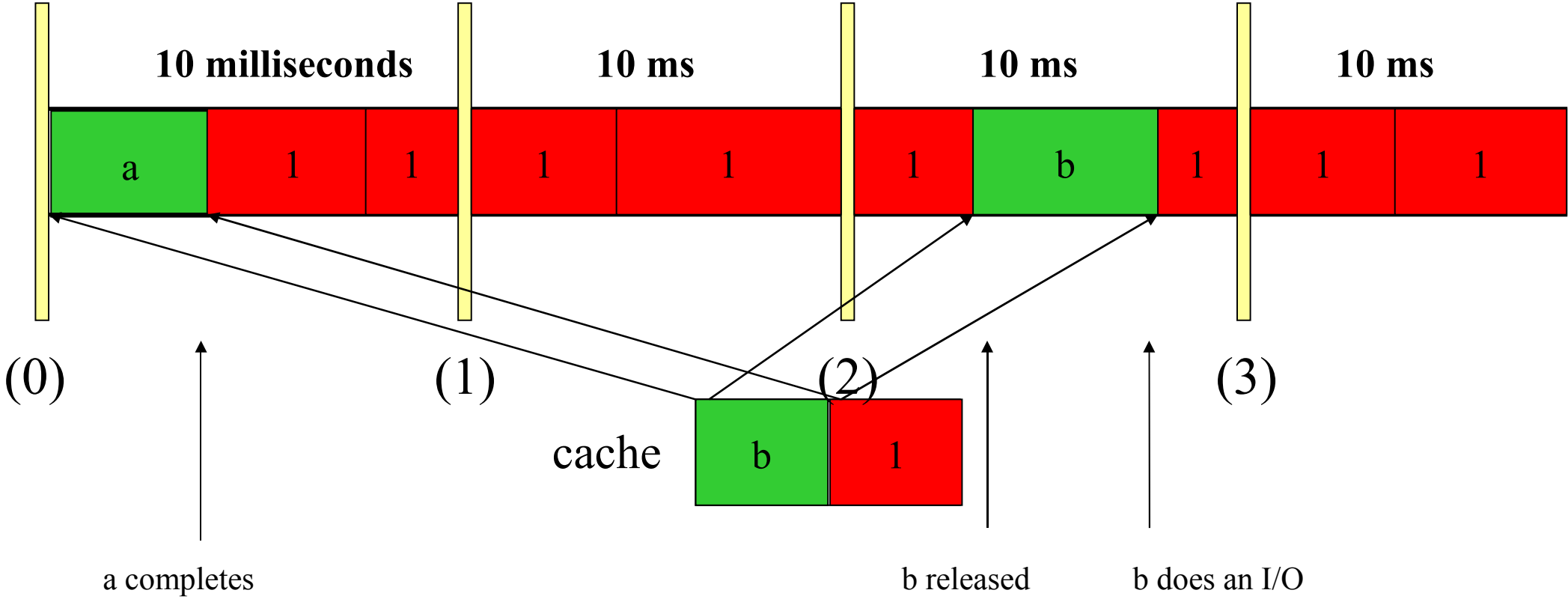




P1 1.5 processing unit default of **2** virtual processors – max of **20 milliseconds** – uncapped because you are limited to only use 2 Processors simultaneously

P2 1.5 processing unit **6** virtual processors – max of **60 milliseconds** – uncapped

Example of Two Partitions Sharing a Processor ("uncapped")



Partition dog jobs a,b,c allocated .6 Processing Units

Partition cat jobs 1,2,3 allocated .4 Processing Units

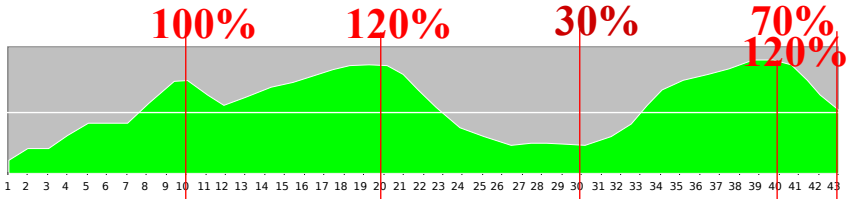
Rolf M Dietze, rmd@dietze-consulting.de

iSeries “uncapped shared pool” with CoD

Shared Processor Pool mit 3 Prozessoren

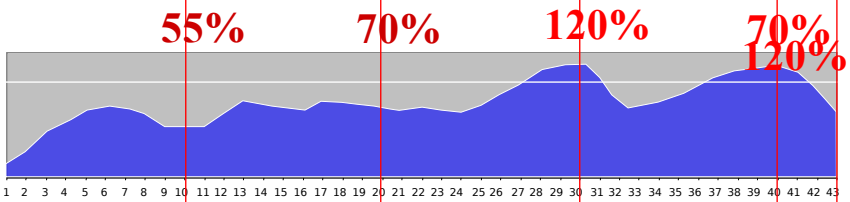
LPAR 1 (uncapped)

Optimized capacity of 0.4 processors for LPAR1



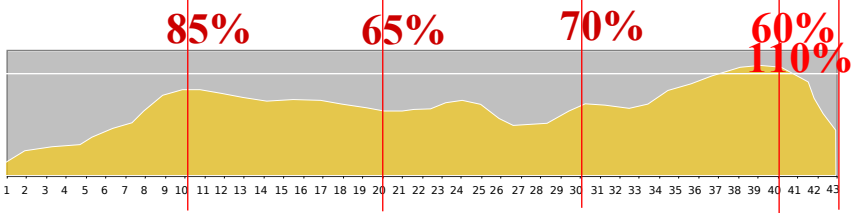
LPAR 2 (uncapped)

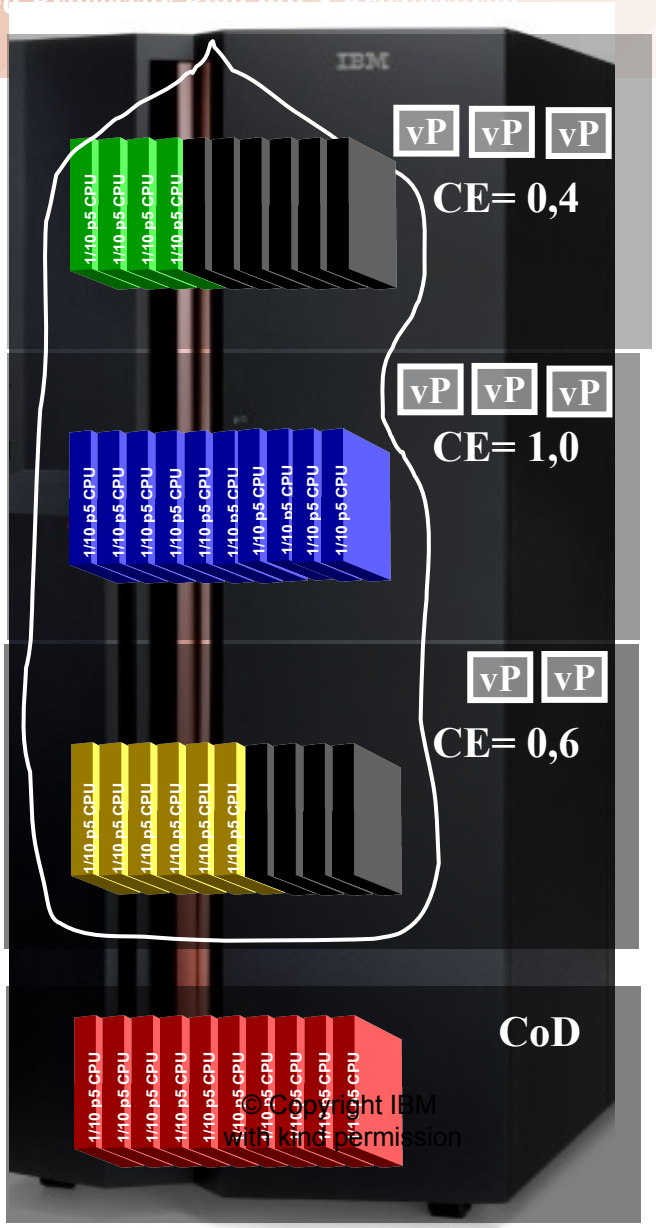
Optimized capacity of an entire processor for LPAR 2



LPAR 3 (uncapped)

Optimized capacity of 0.6 processors for LPAR3

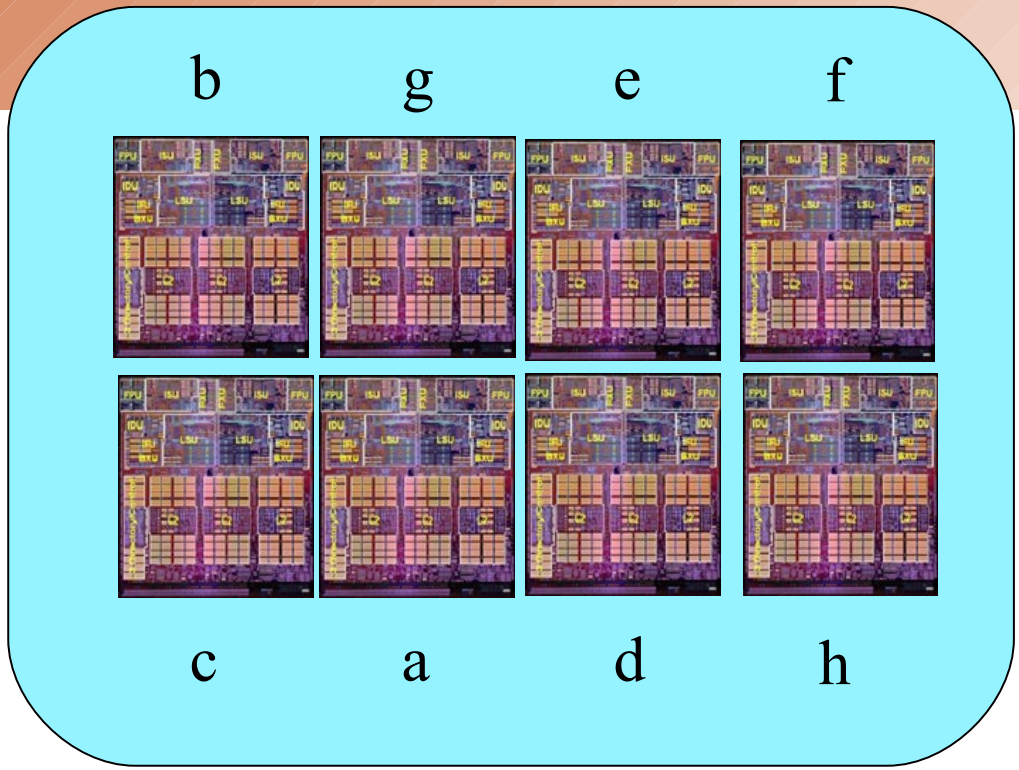




- **Ressources can be requested by any partition**
- **Unused resources can be released**
- **Priorities can be assigned**
- **Unused resources CPUs/MEM will automatically be used to solve failures in a running operating environment**

Rolf M Dietze, rmd@dietze-consulting.de

Virtual Processors (Unlimited) - Uncapped

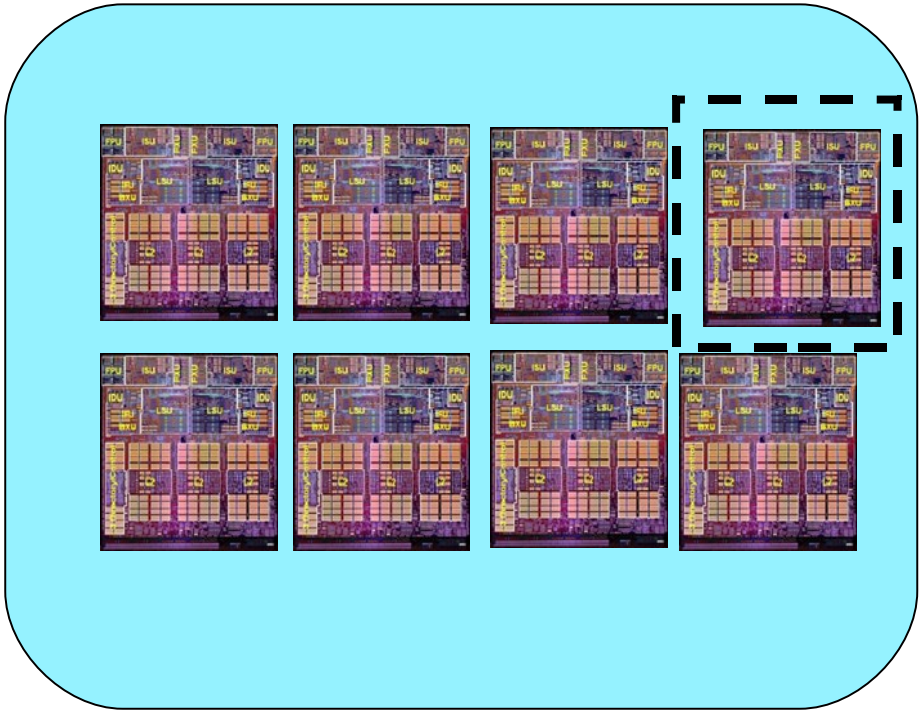


P2 1.5 processing unit with 15 Virtual Processors (maximum allowed) gives Partition – potentially – ***ALL* 80 milliseconds** of Processing time – for every 10 physical milliseconds BUT only as long as the other shared processor partitions DON'T have jobs ready to run!

Rolf M Dietze, rmd@dietze-consulting.de

Floating Processors

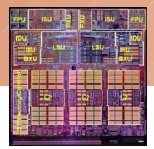
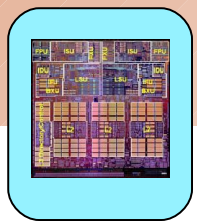
- You have eight processors on your system. Seven are in the pool and one partition uses a dedicated processor.



- Dedicated partitions can allow its processors to be used for uncapped capacity (returned to the shared pool) when the partition is powered off, or a processor is removed from the partition. This is the default.

Rolf M Dietze, rmd@dietze-consulting.de

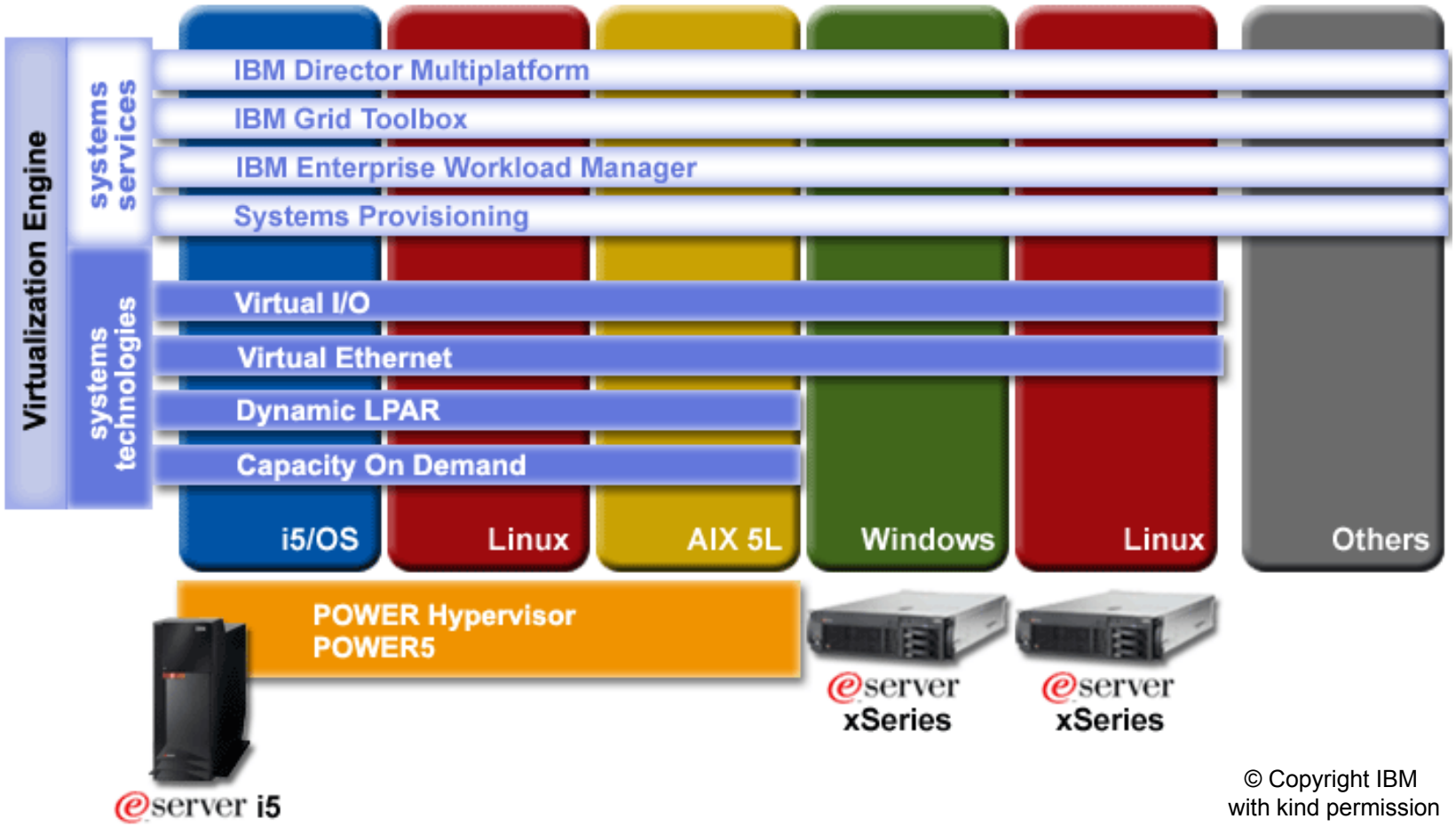
Dedicated or Shared / Capped or Uncapped?


?


- The best performance may well be achieved by using dedicated processors. However, dedicated processors cannot utilize excess capacity.
- For both capped and uncapped partitions, setting the virtual processor number too high may degrade performance.
- Shared uncapped allows use of excess capacity of the processors in the shared pool. Setting virtual processors too low limits the amount of excess you can use. Setting too high may negatively impact performance.
- Also be aware for uncapped partitions the operating system sees the number of desired virtual processors as equal to the number of physical processors, you need an OS license (i5/OS, Linux and AIX 5L) for the lesser of the number of virtual processors or the number of processors in the shared pool.
- So what could be recommended? The right answer depends on workload.

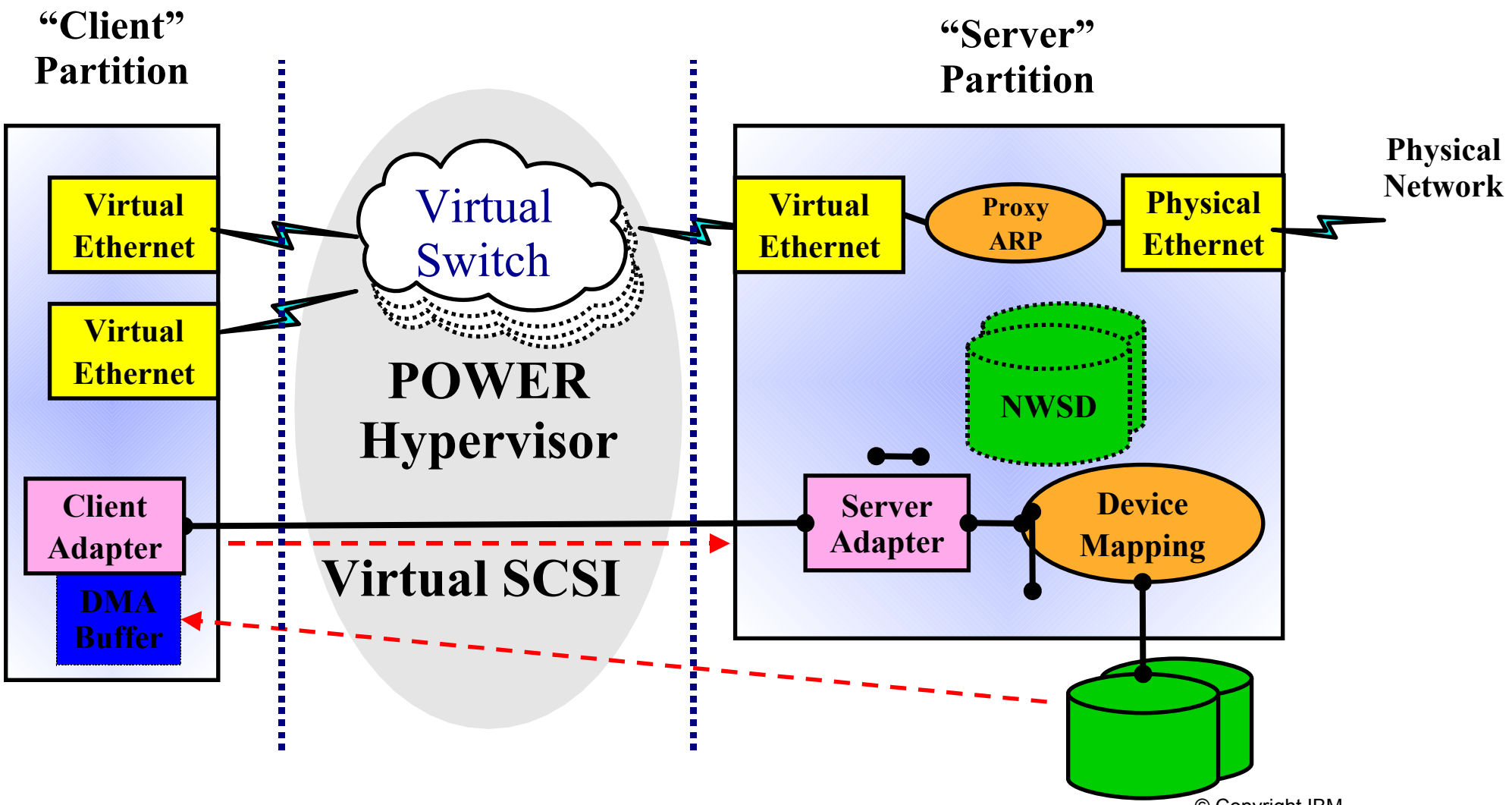
Rolf M Dietze, rmd@dietze-consulting.de

Partitioning Computers

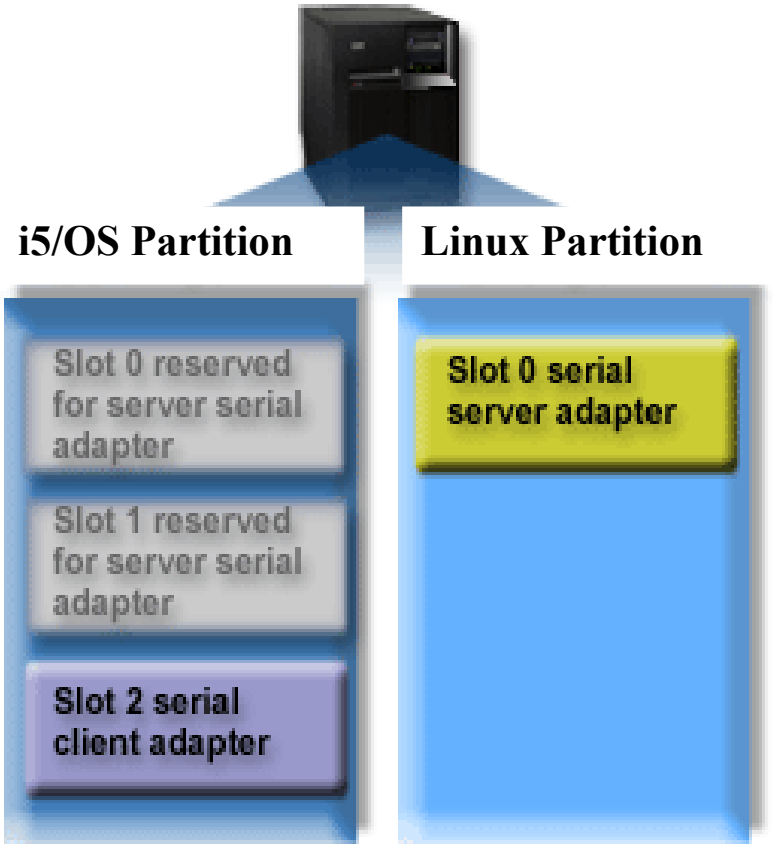


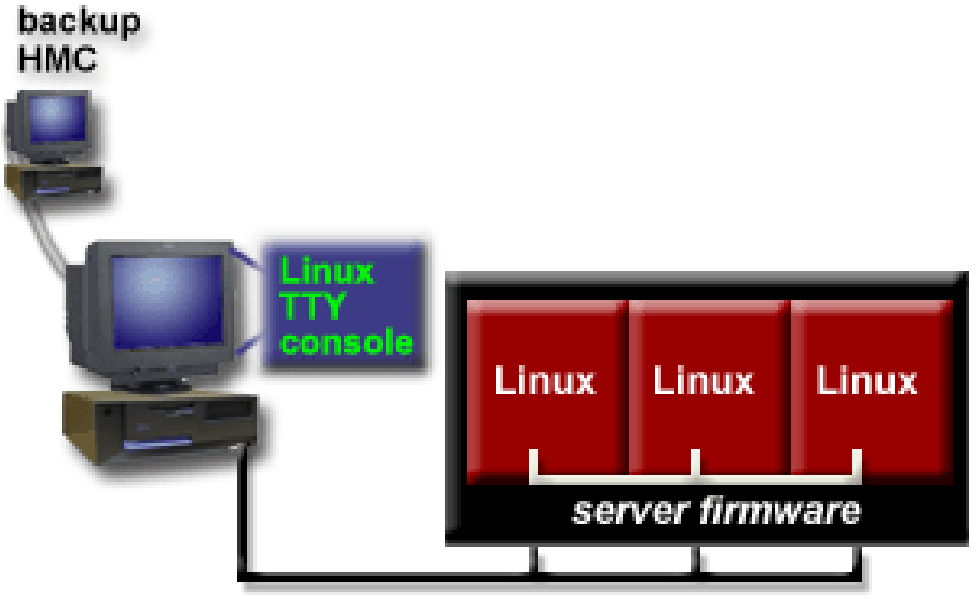
Rolf M Dietze, rmd@dietze-consulting.de

Virtual I/O Example



Virtual Serial

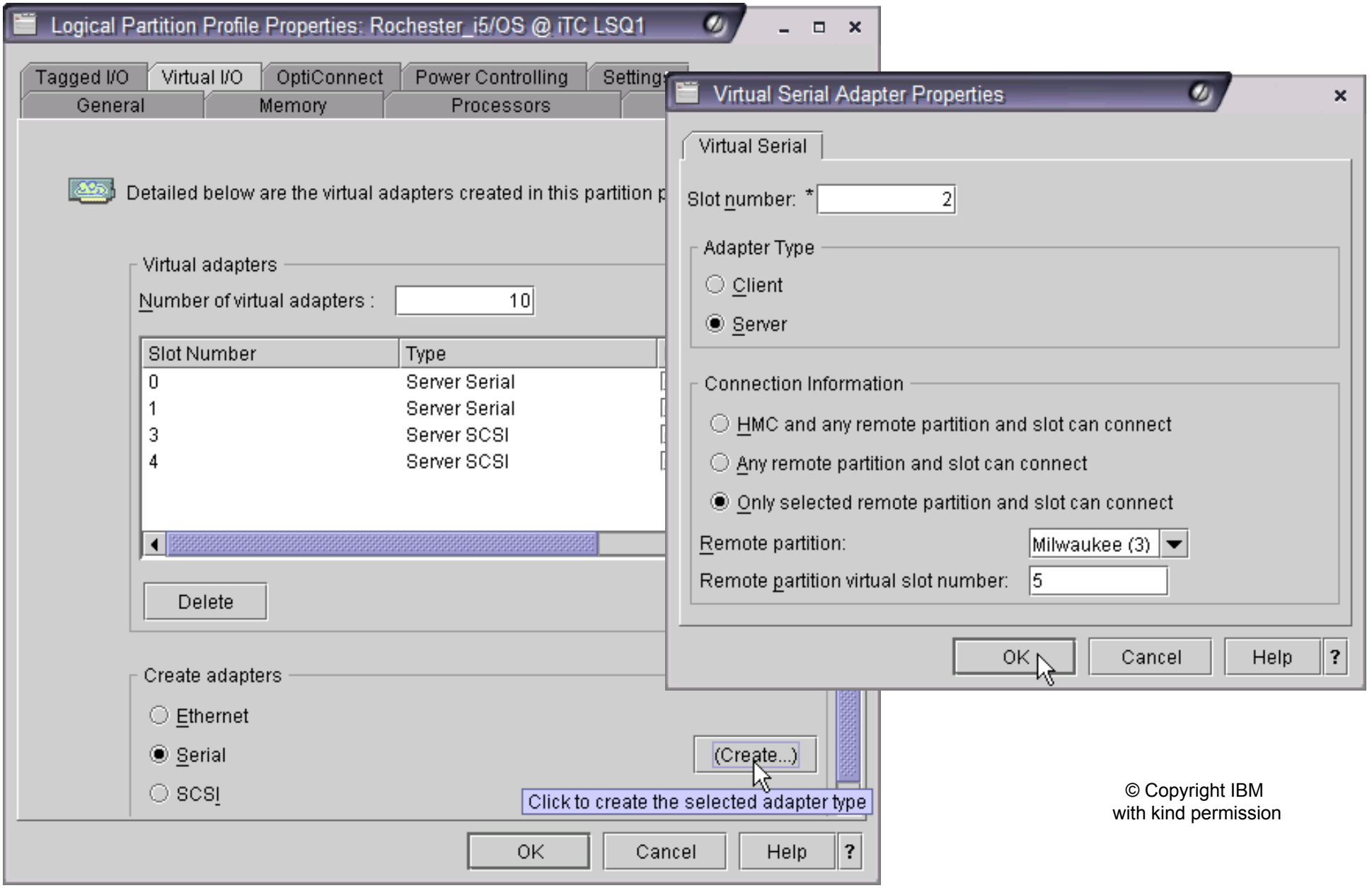




- First 2 virtual slots in every partition reserved for virtual serial server adapters for system console in HMC

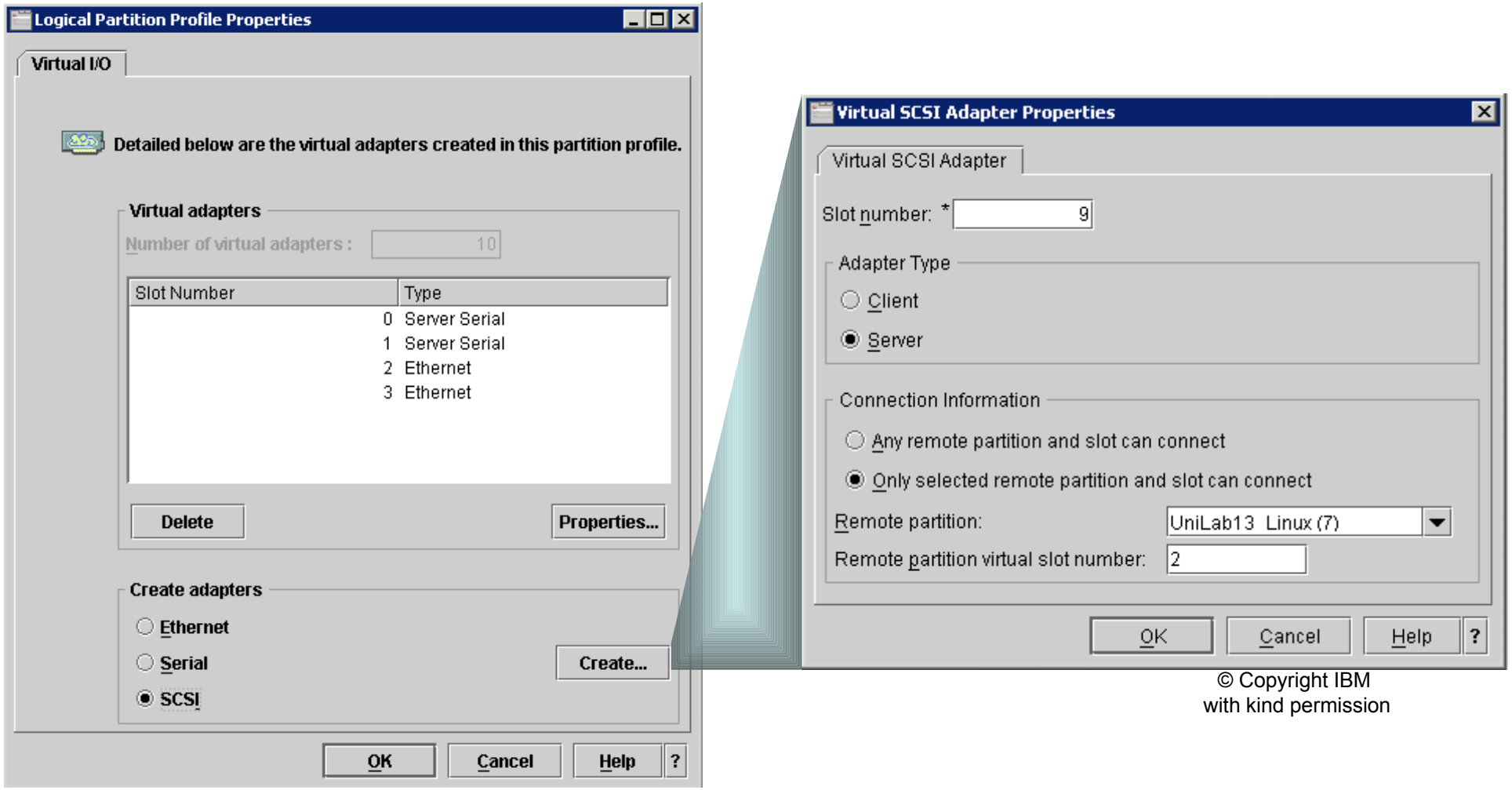
- For i5/OS, virtual serial adapters provide 5250 console
- For Linux and AIX 5L, they provide character console

Virtual Serial



Adding SCSI Adapter via DLPAR

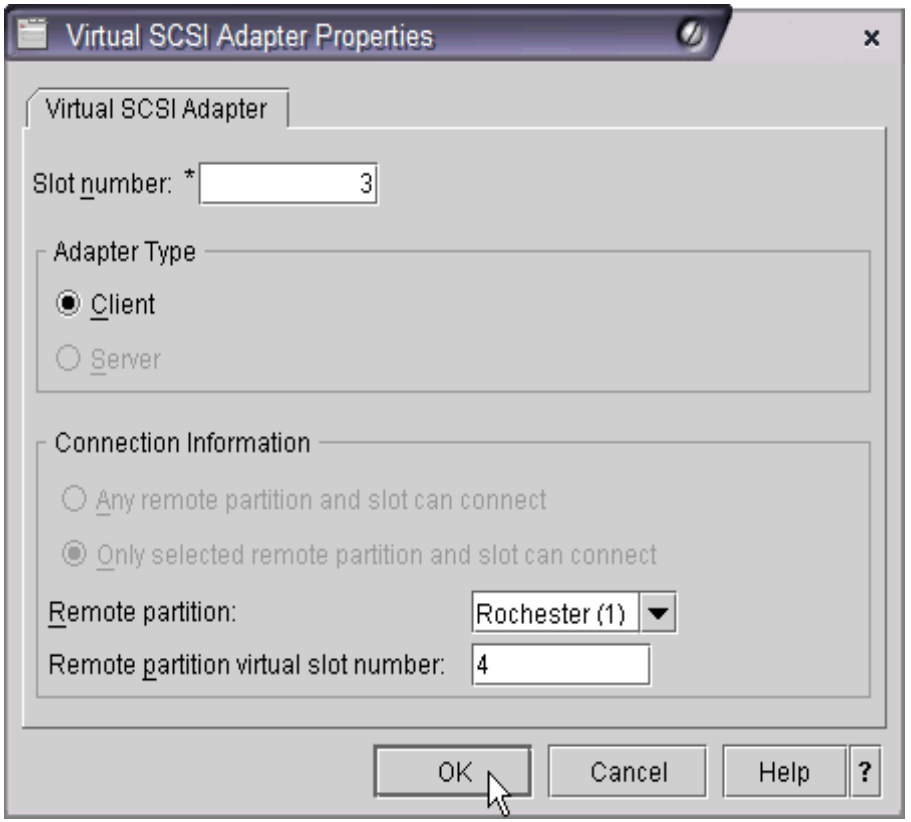
- Create a virtual SCSI client adapter via the partition creation wizard
- Create a virtual SCSI server adapter via Dynamic LPAR on the i5/OS partition – Does not require any restart



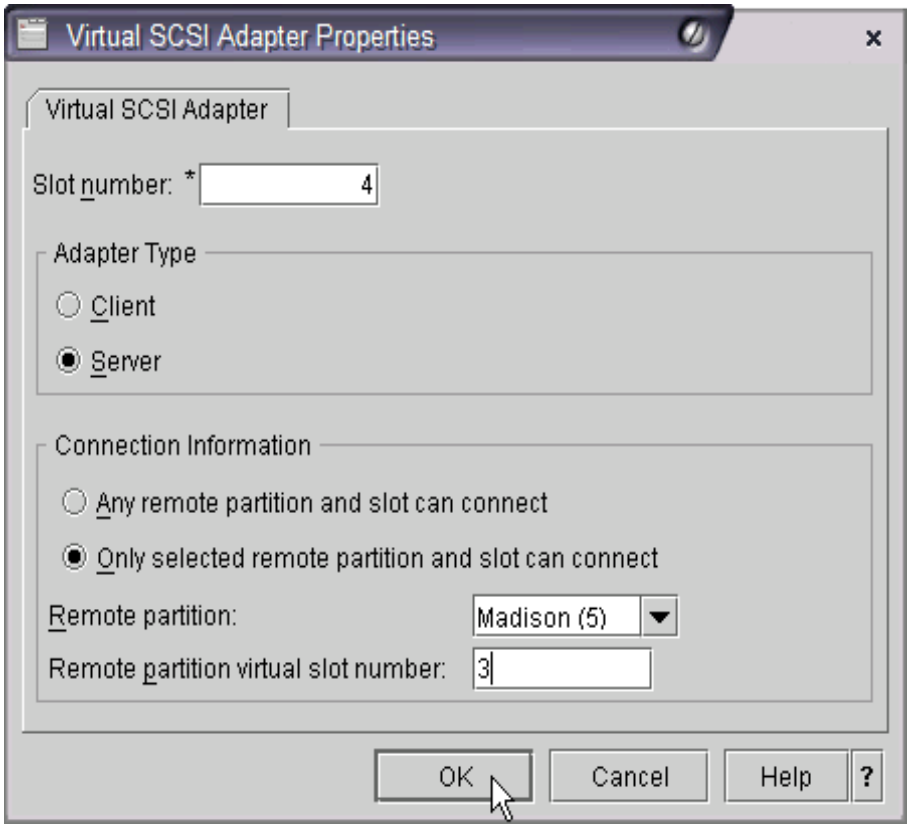
© Copyright IBM
with kind permission

Virtual SCSI

Linux/AIX 5L



i5/OS



Individual slot numbers do no matter, as long as they are configured in pairs