Architect of an Open World™

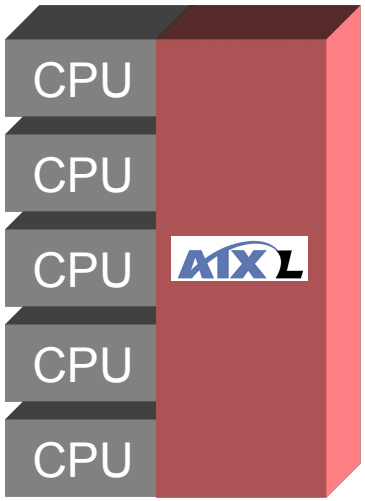
Virtualization for Power6

Dietrich Ziegler

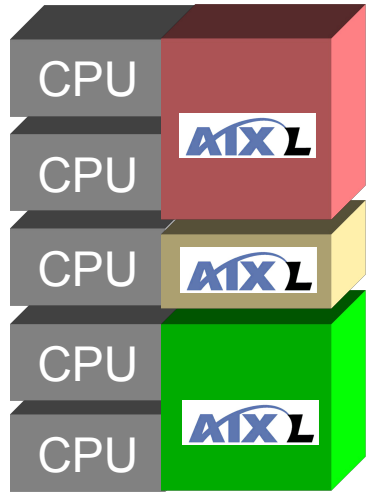
Support Specialist Open Systems

LIBERATE IT

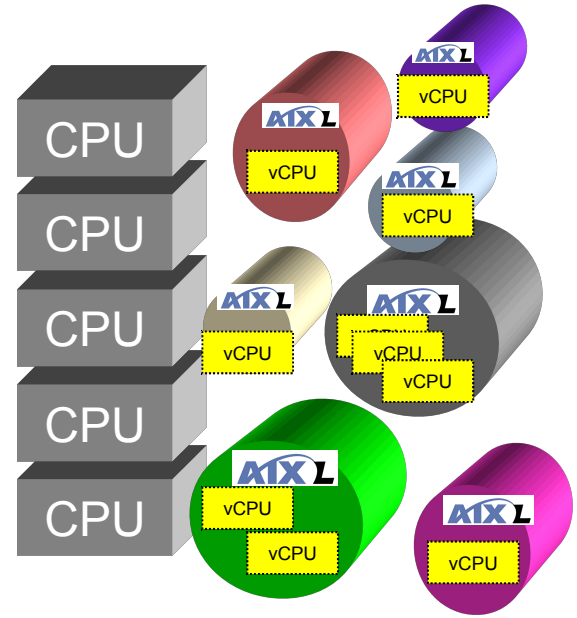
Partitioning evolution



Traditional Server:
- one Server
- one OS



Logical Partitioning:
- multiple OS's
- granularity based on physical CPU's



Micro-partitioning:
- virtual CPU's
- more OS's than physical CPU's
- resource allocation to OS's more granular

PowerVM Features

- The main functions of PowerVM are:
  - **Shared Processor**
  - Virtual I/O server places the

following virtual resources

to the Client Partitions for the order:

  - **Virtual Network**
  - **Virtual SCSI**

These functions can be used together or separately.

Virtualization Features - Licence

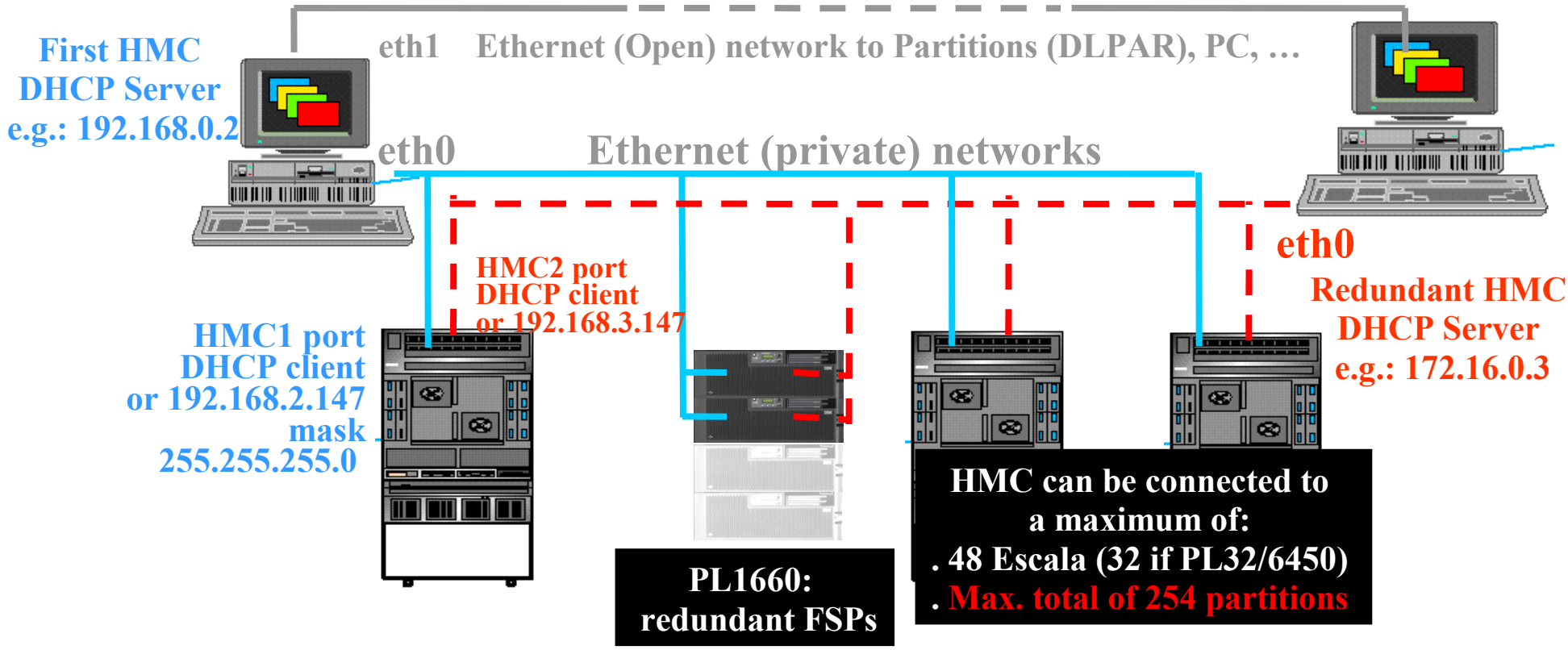
- **PowerVM** introduces three new level of virtualization

- **PowerVM Express**
  - Only available for PL160,260,460,860,E5-700
  - Max. 3 LPARs managed through IVM
  - No HMC attachment, only IVM

- **PowerVM Standard**
  - Available for all Escala models
  - Does not include LPM and AMS

- **PowerVM Enterprise**
  - For all Escala Power7 and PL Power6
  - Includes all virtualization options

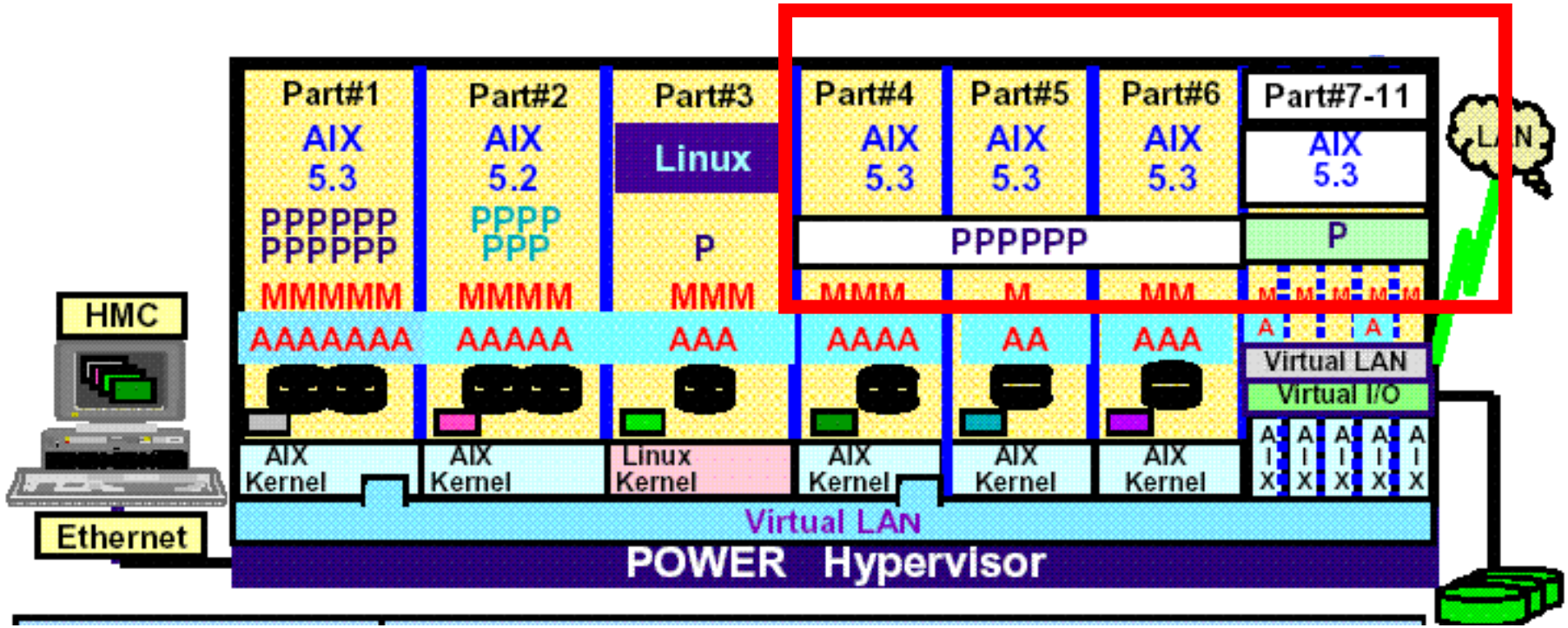
HMC(s) connection to Escala Server(s)



Port eth0 is either **additional Ethernet** or **bottom-right port of an HMC integrated in a rack.**
- **First HMC**, **port eth0** is connected to **HMC1** Ethernet port of each LPAR sever(s) through a private network. HMC port eth0 is configured as **DHCP server (e.g.: range 192.168.0.2)**
- **Redundant (second) HMC**, **port eth0** is connected to **HMC2** Ethernet port of each LPAR server(s) through a second physically SEPARATED private network. Port eth0 on redundant HMC is configured as **DHCP server (e.g.: range 172.16.0.3).**

VIRTUAL VIRTUAL Shared Processors:

micro-partitions Shared Processors:
micro-partitions



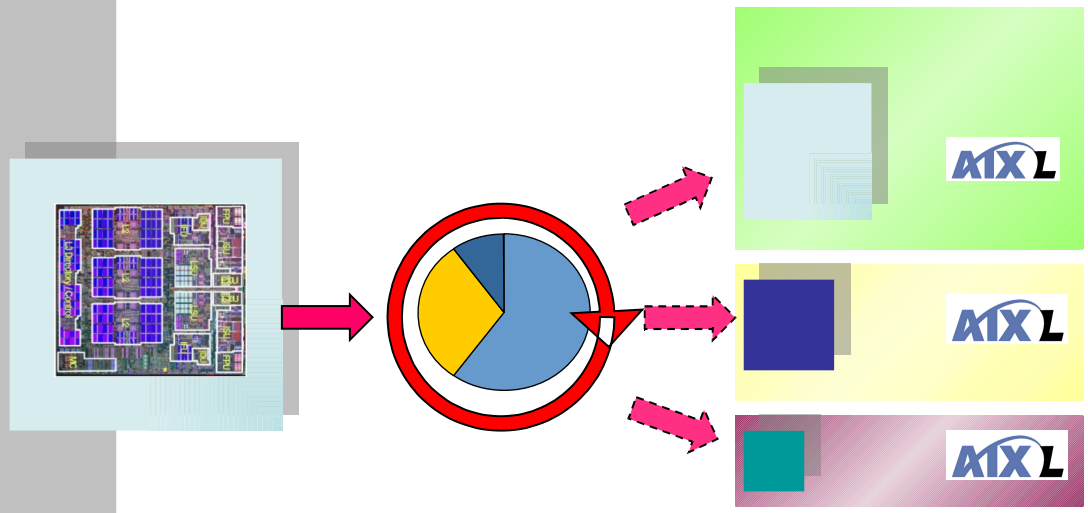
Shared-Processors and Micro-Partitions

- **Partitions** can be defined either with **dedicated** or **shared processors**
  - Traditional **dedicated** processor assignments for **optimal performance** (no hypervisor overhead)
  - By default, all processors **not dedicated** to currently active partitions belong to the **shared processors** pool (except if disabled at partition profile level).

- **Fractional allocation of Shared Processors to Micro-Partitions:**
  - **Each partition gets a percentage of the execution dispatch time on the processors in the pool, based on its entitled processing capacity.**
  - **Minimum** assignable capacity of **1/10th** of a processor (0.10):
    each processor can be shared by up to 10 micro-partitions.
  - Additional **increments** of **1/100th** (0.01)of a processor.

  - The **sum of all entitled processing capacities assigned to active micro-partitions** must be **less than (or equal to) the number of physical processors inside the shared pool.**

Fine grained 1% share granularity



Physical CPU

Hypervisor timeslicing

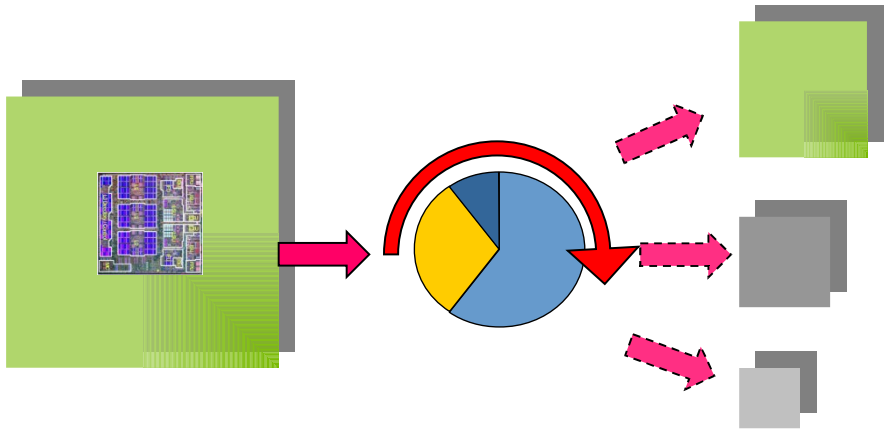
Virtual CPU's

Micro partitions

- Micro-partitions are based on Virtual processors:
  - Smallest partition size: 10% of a physical CPU
  - Smallest increment: 1%
- Micro-partitions can contain multiple Virtual processors

Allows partition sizes smaller than 1 physical CPU

How does it work ?

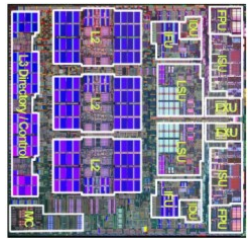


Physical CPU | Hypervisor timeslicing | Virtual CPU's

- Virtual Processors are 'Time sliced' on the physical processors
- Scheduling controlled by the Hypervisor
- 'Size' of virtual processor depends of the number of time-slices given be the administrator (shares)
  - Min.: 10 shares
  - Max.: 100 shares

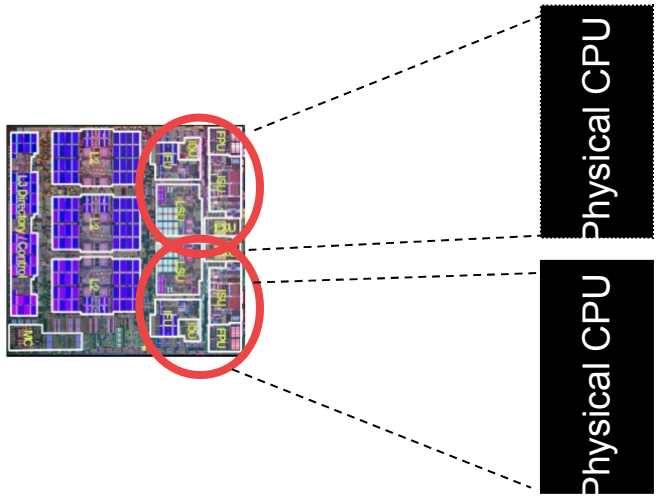
Split a physical CPU into multiple virtual CPU's

Micro-partitioning principles: CPU types and terminology



Power
Dual-core processor

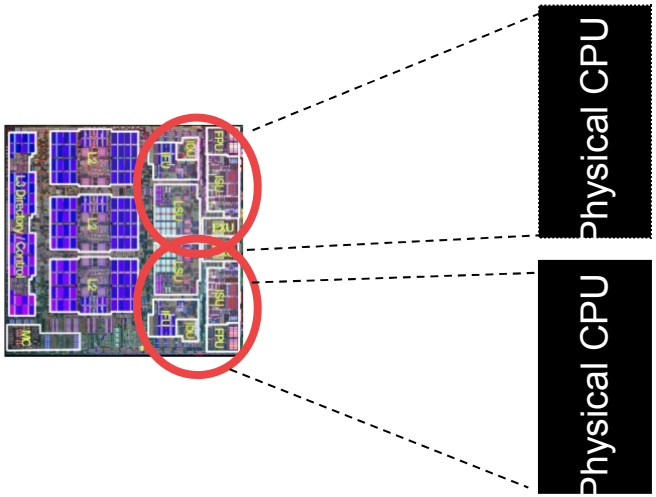
Micro-partitioning principles: CPU types and terminology



Power
Dual-core processor

Physical CPU
- physical resource
- one CPU of a Power 5 processor

Micro-partitioning principles: CPU types and terminology



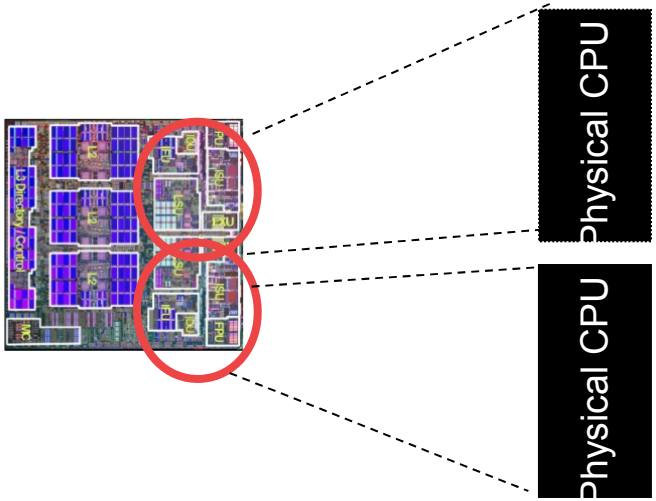
100 processing units

100 processing units

Power
Dual-core processor

Physical CPU
- physical resource
- one CPU of a Power 5 processor

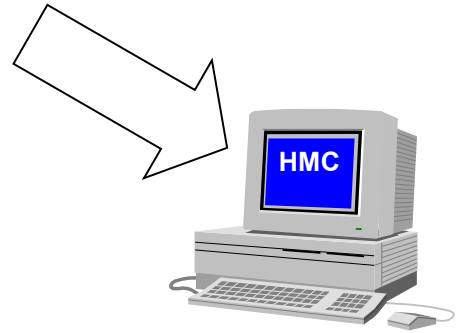
Micro-partitioning principles: CPU types and terminology



Power
Dual-core processor

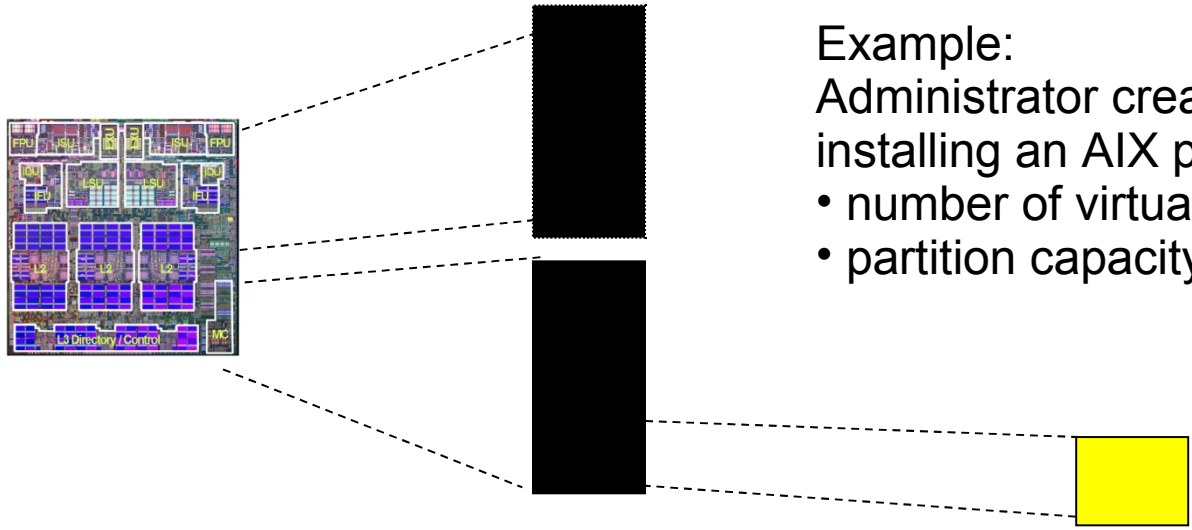
Physical CPU
- physical resource
- one CPU of a Power 5 processor

200 processing units



CPU shares
- are the 'HMC dispatchable' resources
- what the Administrator 'sees'

Micro-partitioning principles: CPU types and terminology



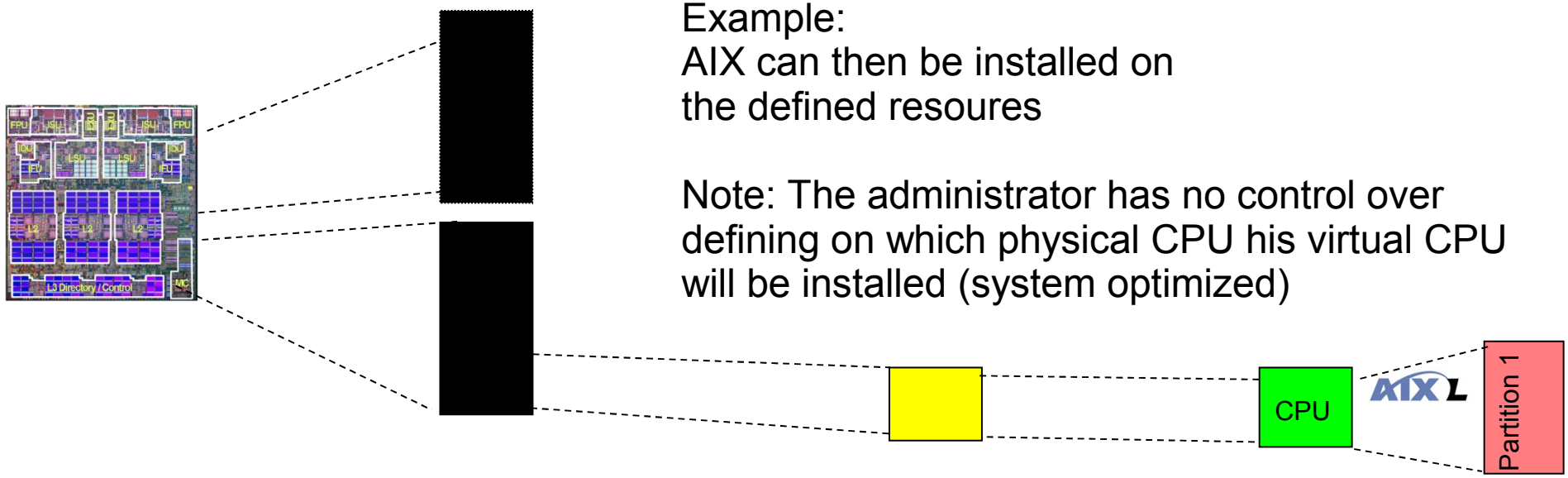
Example:
Administrator creates a virtual CPU for installing an AIX partition:
- number of virtual CPU's: one
- partition capacity: 30 processing units

Power processor
- two CPU's per Power 5 chip

Physical CPU
- physical resource
- one CPU of a Power 5 processor

Virtual CPU
- % of physical processor shares
- Administrator defined
- behaves like a Power 5 CPU

Micro-partitioning principles: CPU types and terminology



Example:
AIX can then be installed on the defined resoures

Note: The administrator has no control over defining on which physical CPU his virtual CPU will be installed (system optimized)

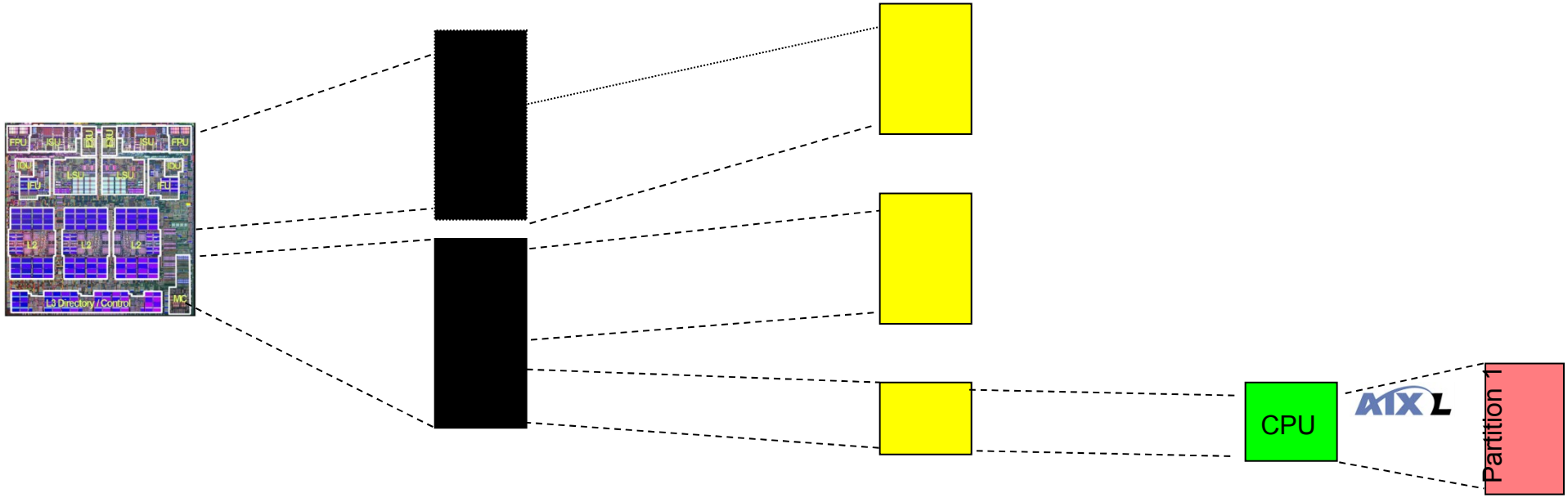
Power processor
- two CPU's per Power 5 chip

Physical CPU
- physical resource
- one CPU of a Power 5 processor

Virtual CPU
- % of physical processor shares
- Administrator defined
- behaves like a Power 5 CPU

Logical CPU
- HW thread
- influenced by SMT on/off
- OS view of a dispachable unit

Micro-partitioning principles: CPU types and terminology

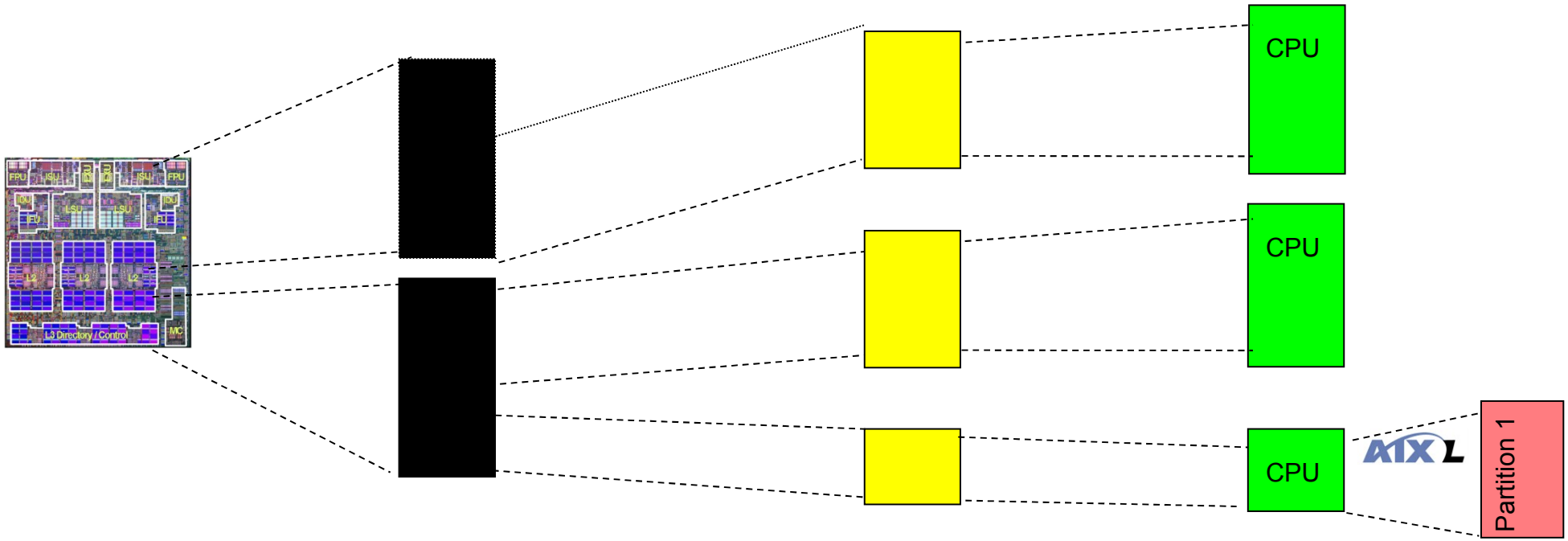


Note: when configuring partitions which require more resources than one physical CPU (>100 shares) - the HMC will force automatically the creation of more than one virtual CPU

Example: Creation of a larger partition:

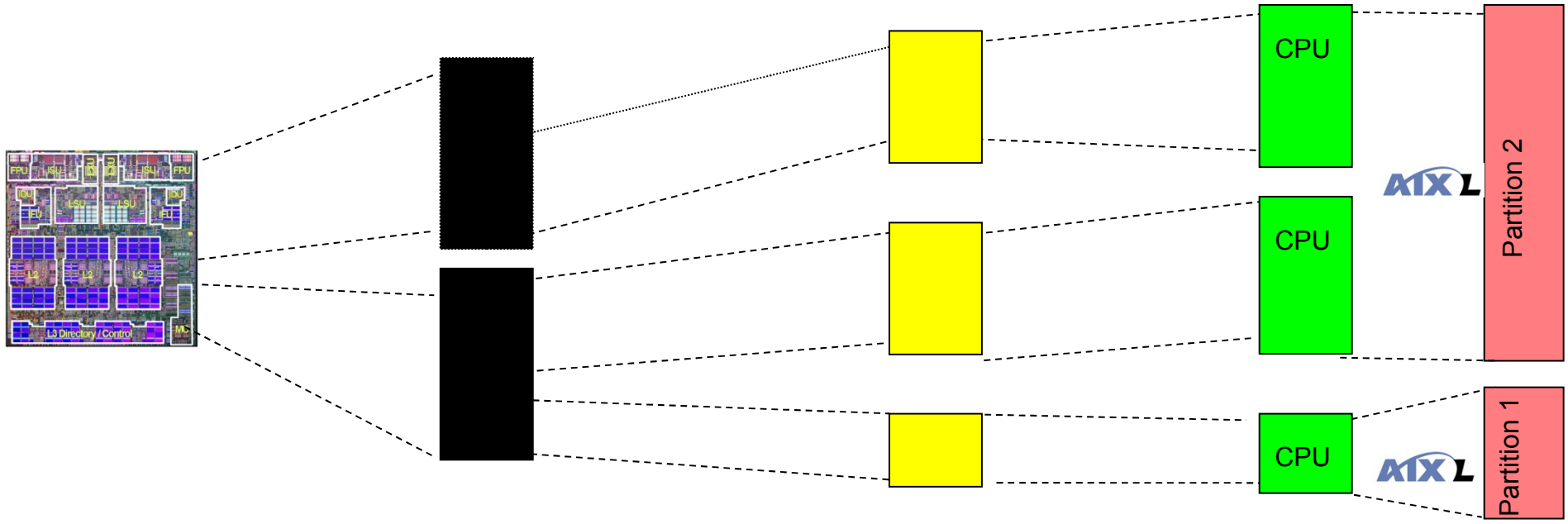
- partition capacity:
  120 processing units
- number of virtual CPU's:
  2 each with 60 processing units

Micro-partitioning principles: CPU types and terminology



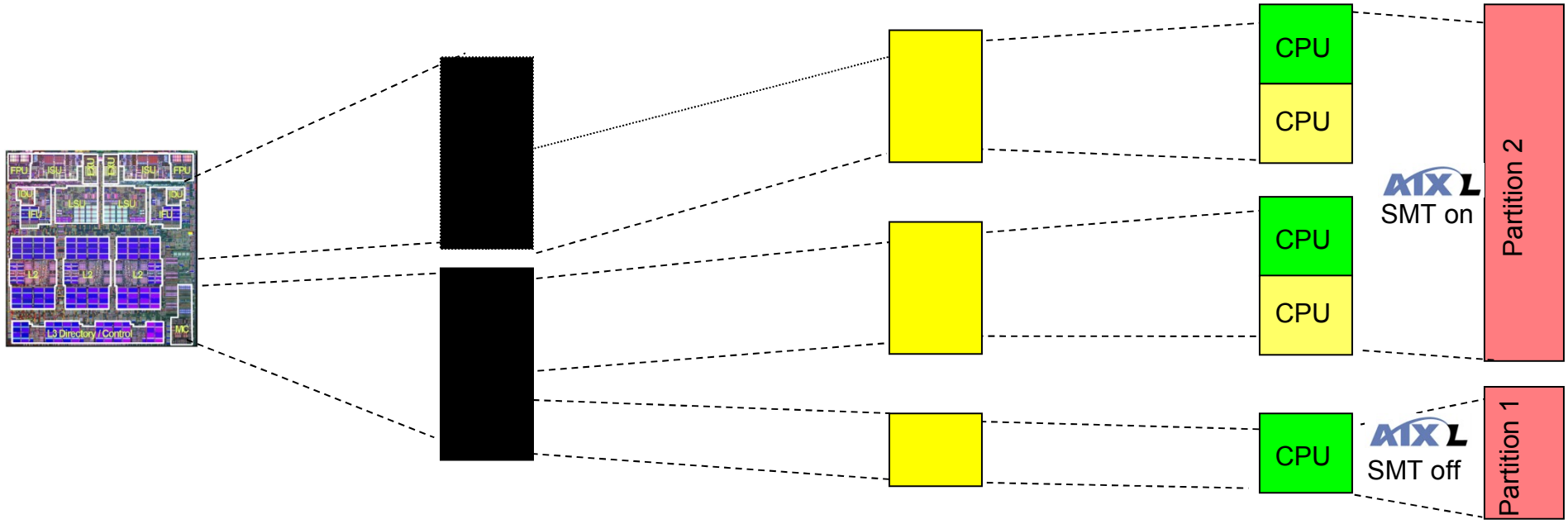
A ‘virtual SMP’ systems with two CPU’s has been created

Micro-partitioning principles: CPU types and terminology



…. where AIX can now be installed

Micro-partitioning principles: CPU types and terminology

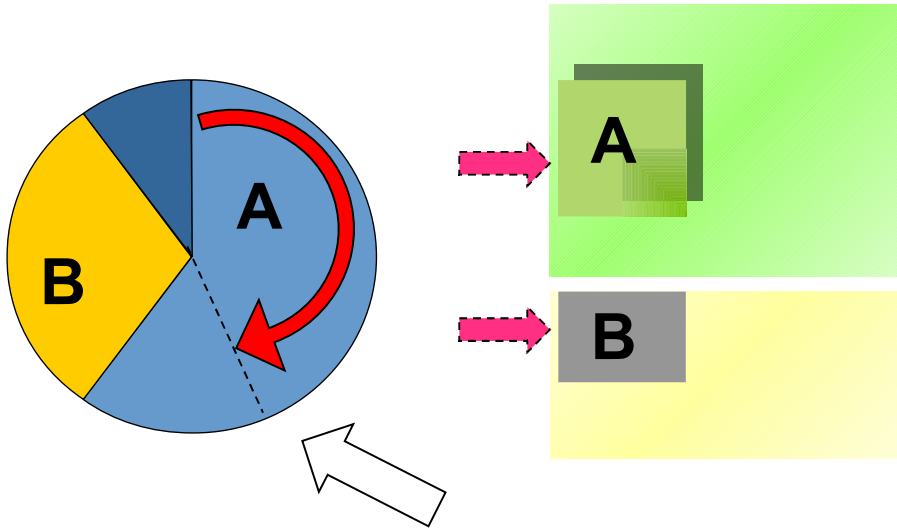


Note: SMT on/off decission is on partition level – no influence to settings on other partitions – even if they share the same physical CPU

SMT 'on' in Partition 2

- Partition appears to AIX as a 4-way system (4 logical CPU's)
- SMT can be turned on/off by the administrator during runtime

Sharing of resources – uncapped partitions

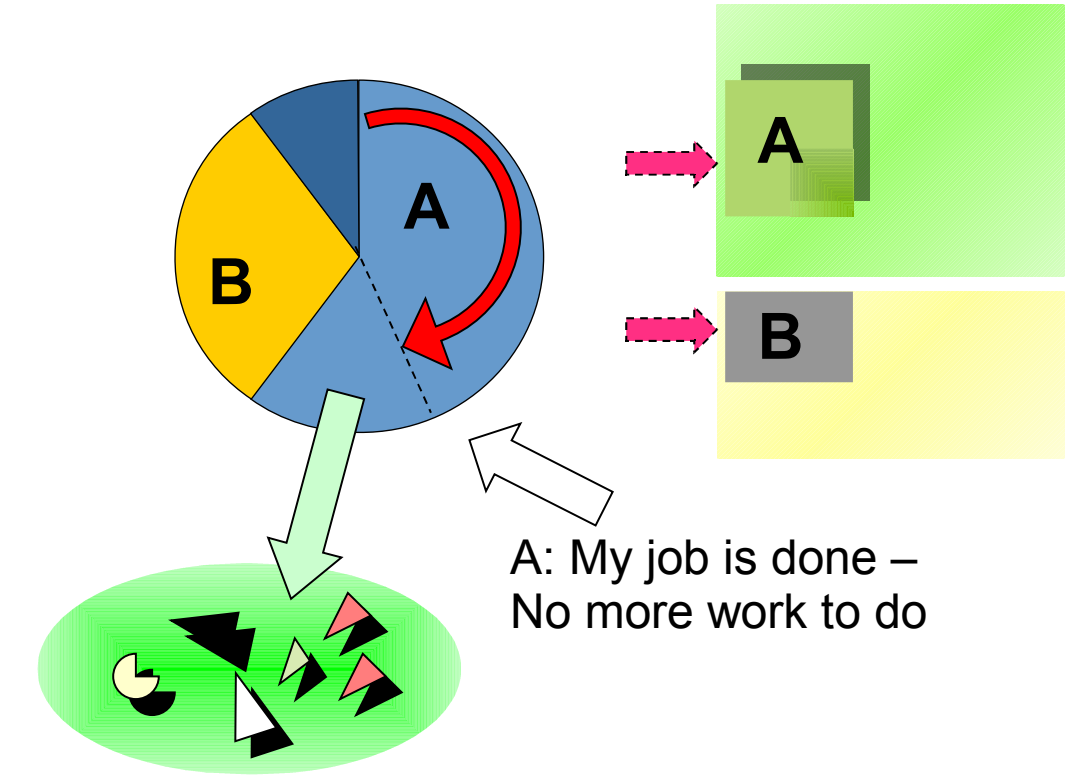


A: My job is done –
No more work to do

- Partitions won't (can't) use constantly their full CPU entitlement
  - Waiting for lock's, I/O access, user interaction, ..

Automatic cross-partition workload balancing

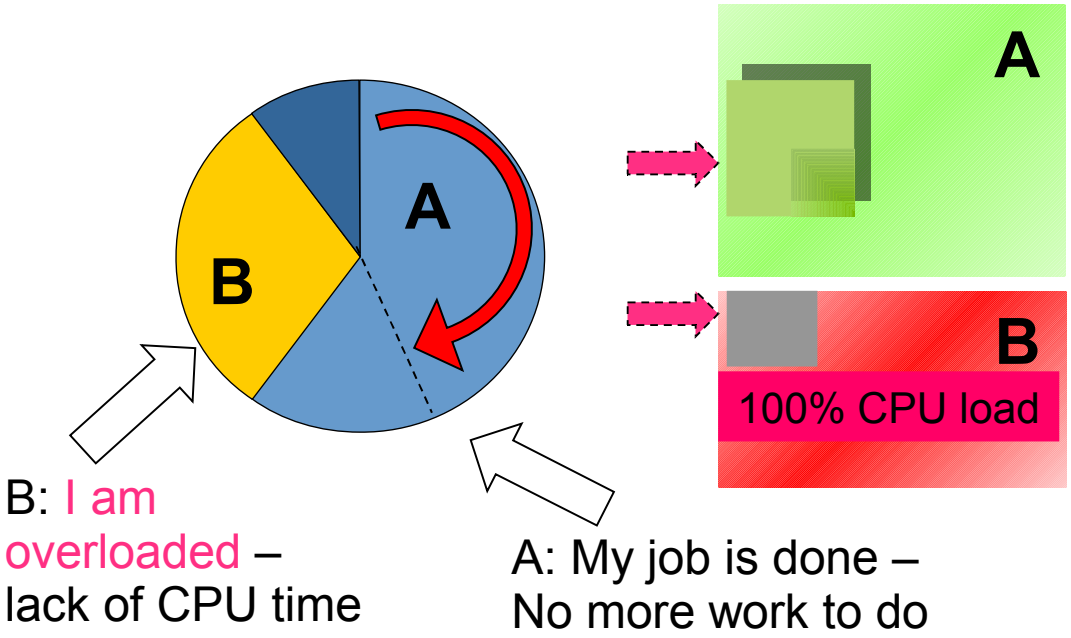
Sharing of resources – uncapped partitions



- Unused time slices are given back automatically to the 'shared processor pool'
  - Cede system call

Automatic cross-partition workload balancing

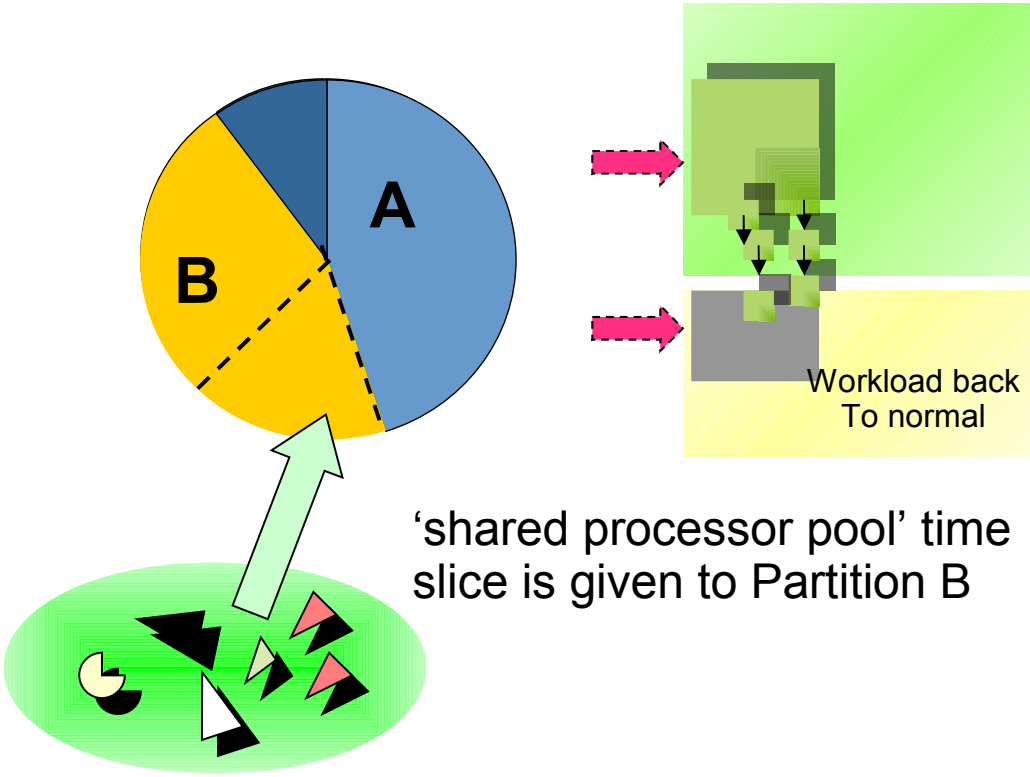
How an uncapped partition works



- Partition B is defined as ‘uncapped’ partition
  - CPU load approaches 100%

Automatic cross-partition workload balancing

How does it work ?

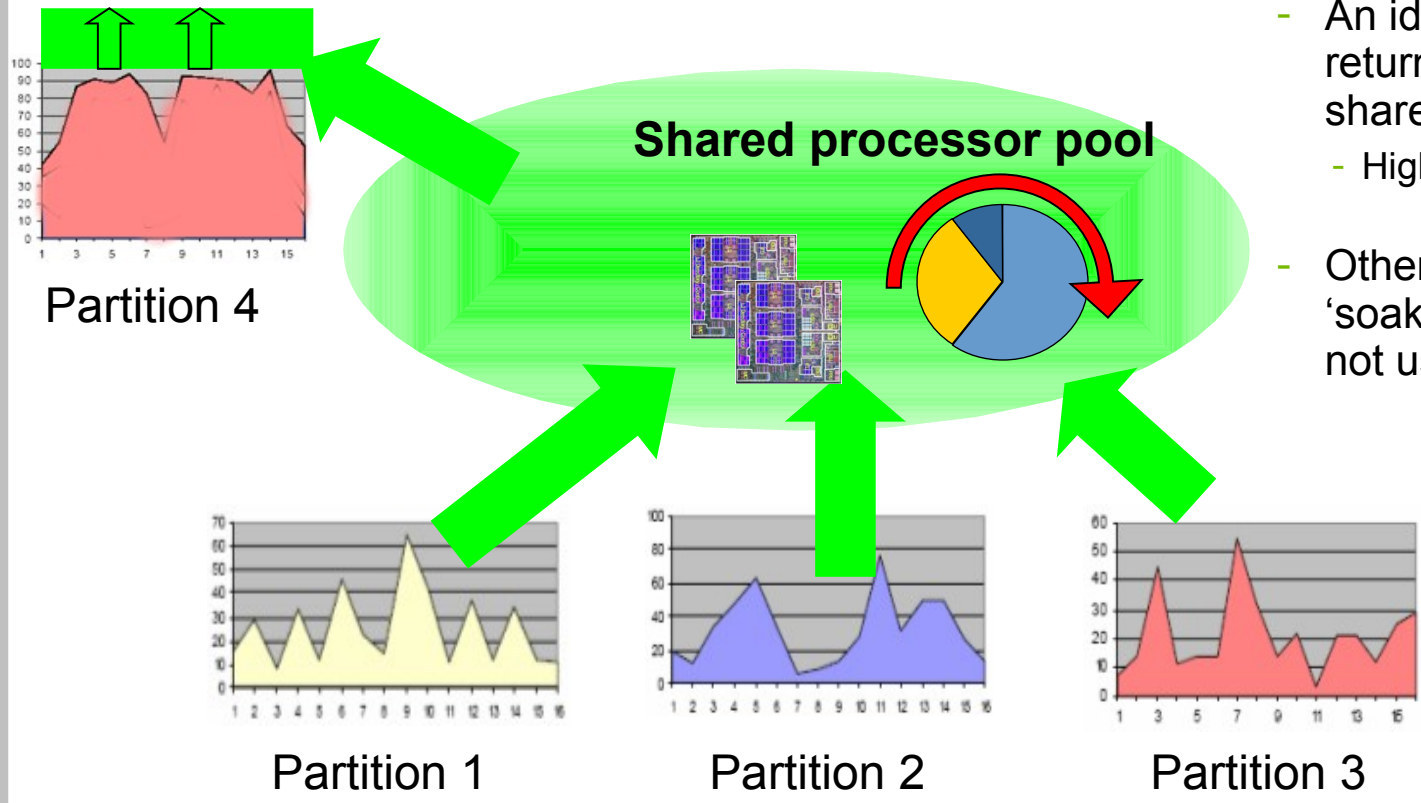


'shared processor pool' time slice is given to Partition B

- Partition B is defined as 'uncapped' partition
  - CPU load approaches 100%
- Partition B can benefit dynamically from additional CPU shares out of the shared processor pool

Automatic cross-partition workload balancing

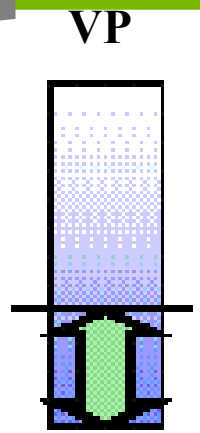
Highly efficient used or resources



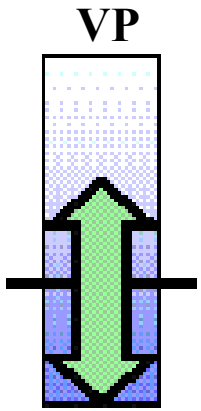
- An idle virtual processor can return CPU time back to the shared pool
  - Highly efficient use of resources
- Other (uncapped) partitions can ‘soak-up’ CPU power which is not used

Unused CPU-cycles can be re-distributed

Capped vs Uncapped Processor Sharing



- **Capped Processing Mode for Micro-Partition (default** mode**):**
  - **The partition is limited to its (desired) "entitled capacity" (ent)**
    - Actual utilization depends on the partition's workload
    - The partition is always guaranteed its full entitled capacity, whenever it needs it
    - The partition **never uses more** than its entitled capacity, unless the capacity changes
  - **Unused capacity is ceded to other (uncapped) partitions**
  - Capped should be used only in case of CPU capacity billed to users



- **Uncapped Processing Mode for Micro-Partition (recommended** mode**):**
  - The partition is still **always guaranteed its full « entitled capacity » (ent)**
  - but it may utilize **more than its entitled desired capacity,** up to its current (dedired) number of **virtual processors** (see Notes below)
    - **If it has more work to do** (needs more CPU cycles) than its entitled capacity
    - **and if there is « unused » capacity** in other partitions (capped or uncapped)
    - **or if all shared-processor capacity is not assigned** to currently active partitions
  - **All uncapped partitions share this "unused" capacity, according to their weight**
    - Extra cycles are dispatched between multiple partitions needing extra cycles according to **need**, and **user-defined partition weight** (between 0 and 255, default value is 128) **Variable Weight=** *% share (priority) of surplus capacity* **(see Notes)**
  - When ALL shared-pool capacity is used, ***Reserve Capacity on Demand*** **processors, if present, may be used**
  - **Uncapped mode is RECOMMENDED for optimal CPU utilization**
- **Power6 allows Dedicated CPUs idle cycles Donation to Pool**

Metering CPU usage by Partitions

- **How to know which partitions are using all the processor cycles?**

- **AIX 5.3 Partition commands:**
  - **enhancements to** *sar –P ALL, topas –L, topas -C*
    to show physical processor and percentage of entitled processing capacity consumption.
  - **new commands:** *lparstat –I lparstat [-h] 2 2 mpstat –s 2 2*
    See examples next page
  - **Redbooks: IBM @server p5 Virtualization Performance Considerations SG24-5768**
    AIX 5L Differences Guide Version 5.3 Edition SG24-7463-00

- **Processor utilization metrics, from any AIX partition:**
  - ***topas –C*** **show full-screen Cross-LPAR (CEC) panel (dynamic view) related to CPU activity**
  - ***topas –R*** **to record cross-LPAR CPU activity,**
    topasout –R **to display recorded activity**
    For implementation and usage **see Notes** in bottom comment part below
  - **BPR-SEBull Performance Report Standard Edition since July 2007 BullEnh 530_09 CD**

  - at HMC level, select a **partition -> *Properties* -> *Hardware***
    Tick ***Allow performance information collection.***
    Stop (*shutdown -F*) then reactivate partition:
    lparstat 2 2 app **(available processor pool) column will show global usage of shared processors pool capacity by ALL micro-partitions**

AIX 5.3: Example of *lparstat* & *mpstat* outputs

- **To see uncapped behaviour of a micro-partition: *# lparstat* or *# topas -L***

  Ex.: Micro-Partition Uncapped with 0.30 of CPU Entitled Capacity (*%entc* or *ent*), **one Virtual Processor, SMT is enabled:** Virtual processor appears to AIX as **TWO Logical CPUs**

  # lparstat [-h] 2 3 **(** -h shows time spent in hypervisor but NOT a measure of hypervisor/partitioning overhead)

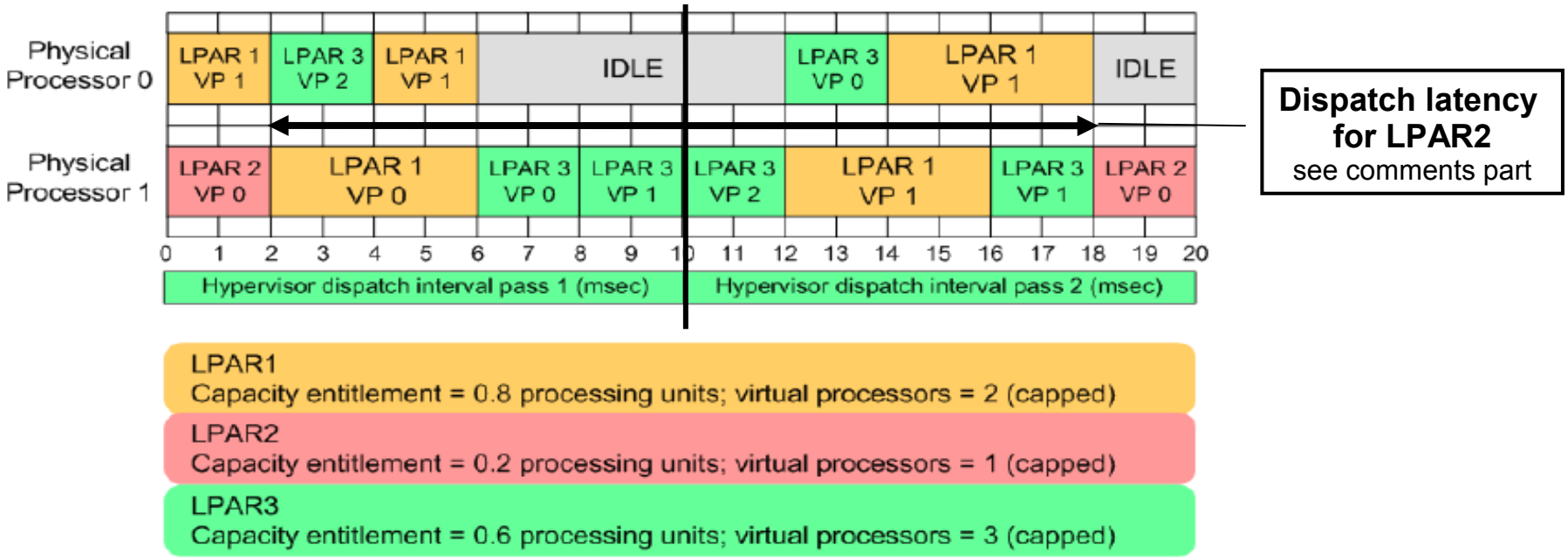
  System configuration: type=**Shared** mode=**Uncapped** **smt=On lcpu=2** mem=512 **ent=0.30**

%user	**%sys**	**%wait**	**%idle**	**physc**	**%entc**	lbusy	**app**	vcsw	phint	**[%hypv**	**hcalls]**
3.0	11.5	1.8	83.6	0.05	16.7	5.0	-	586	8	1.4	240
15.6	76.9	0.6	**7.0**	**0.99**	**329.1**	89.2	-	707	296	0.9	150
0.0	0.3	0.0	99.7	0.00	0.8	0.0	-	574	0		

  When system is busy (0% idle), because free CPU capacity is available,
  uncapped partition uses **329.1% of 0.30 CPU = 1** (virtual and physical) **CPU = 0.99 physc**
  where a **capped** partition would stay limited to **100% of 0.30** (virtual and physical)**CPU**

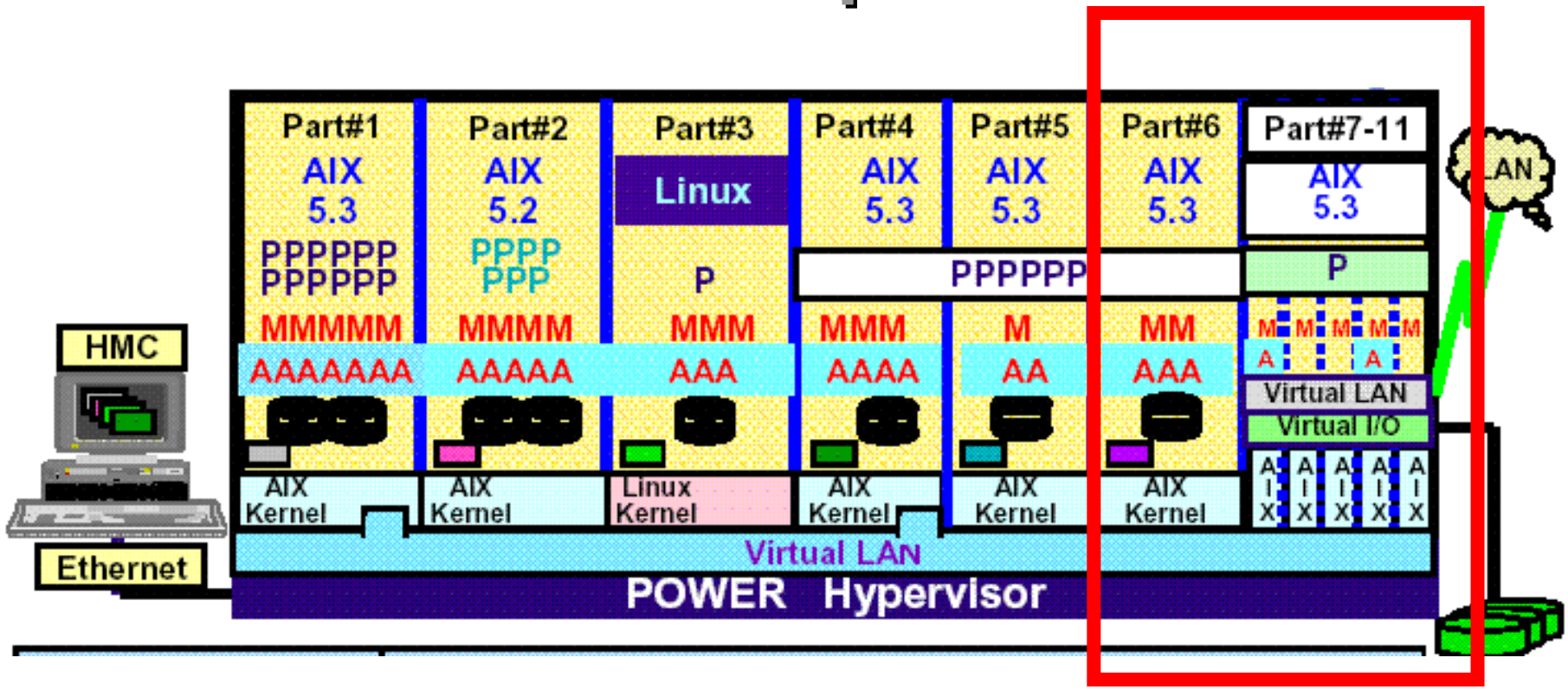
  ***app* column shows global usage of shared processors pool capacity by ALL micro-partitions**
  if « Allow shared processor pool utilization authority this partition » has been set in this partition Properties:
  Properties -> Hardware -> Processors and memory

Dispatch of Processing Capacity cross Virtual and Physical Processors



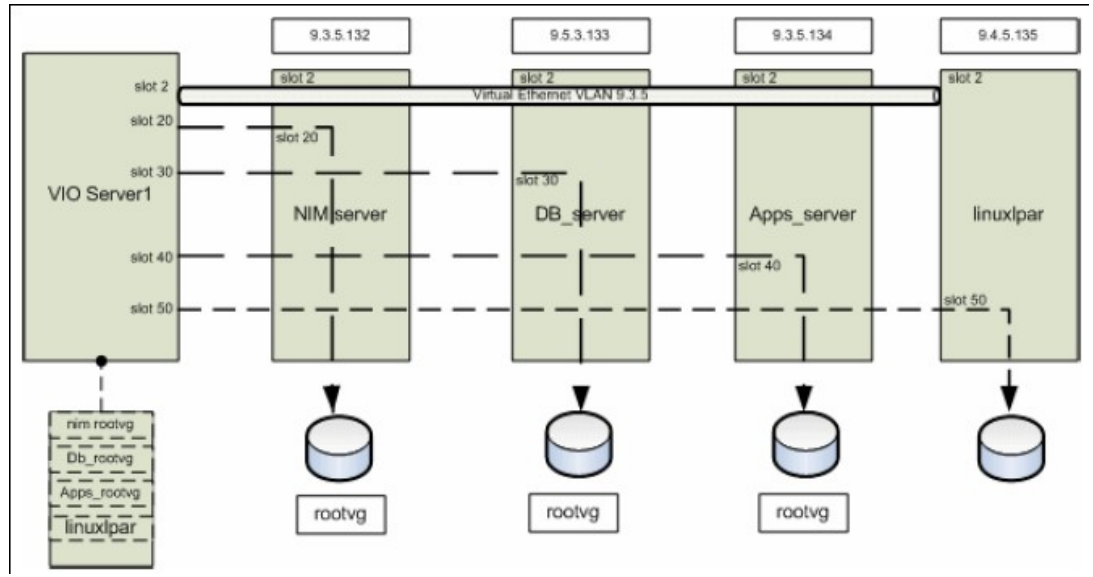
- **LPAR3 can be defined with 3 Virtual Processors even though there are only 2 Physical Processors**
- **POWER Hypervisor attempts to maintain physical processor affinity when dispatching virtual processors.** It will always first try to dispatch the virtual processor on the same physical processor as it last ran on, and depending on resource utilization will broaden its search out to the other processor on the POWER5 chip, then to another chip on the same MCM, then to a chip on another MCM

VIRTUAL I/Os:
Virtual Ethernet Adapters and Disks



LIBERATE IT

Typical Configuration: one VIO server and client partitions



Configuration:

One VIO server with:

- **One physical Ethernet adapter** and a virtual Ethernet adapter on a VLAN.
  SEA allows client partitions communication with external network
- **One virtual SCSI server adapter *(vhost<x>)* per client partition, to virtualize :**
  - Logical Volumes in Volume Groups
  - physical disks (LUNs from SAN)
  which will be used as virtual disks for boot and data by client partitions
- **DVD drive** virtualized to ALL client partitions via **a particular virtual SCSI server adapter**

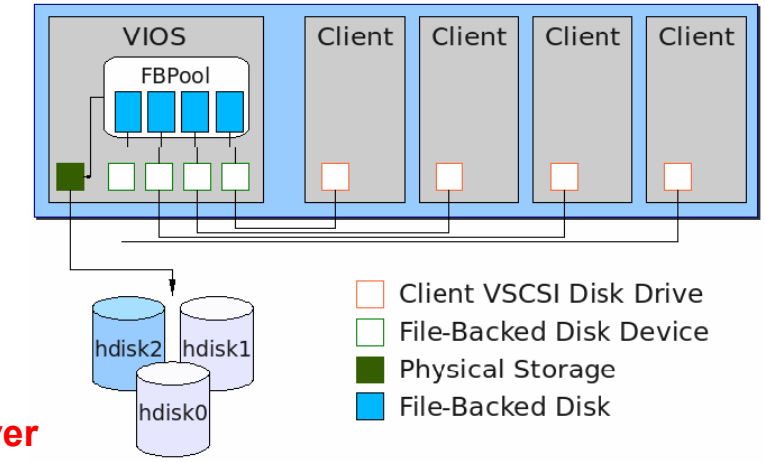
Several Virtual I/O client partitions, each one with:

- **A virtual Ethernet** adapter for inter-partition communication and access to external network via Shared Ethernet Adapter **(SEA)** and physical Ethernet adapter in VIO server.
- **A virtual SCSI client adapter *(vscsi<x>)*linked to one *vhost<x>* VIO virtual SCSI server adapters**
  AIX installed in a virtual SCSI disk located in a logical volume of the VIO server *rootvg*
- **A particular virtual SCSI client adapter to access DVD in VIO server**

File-backend Virtual Disk and Virtual Optical Drive (VIO 1.5)

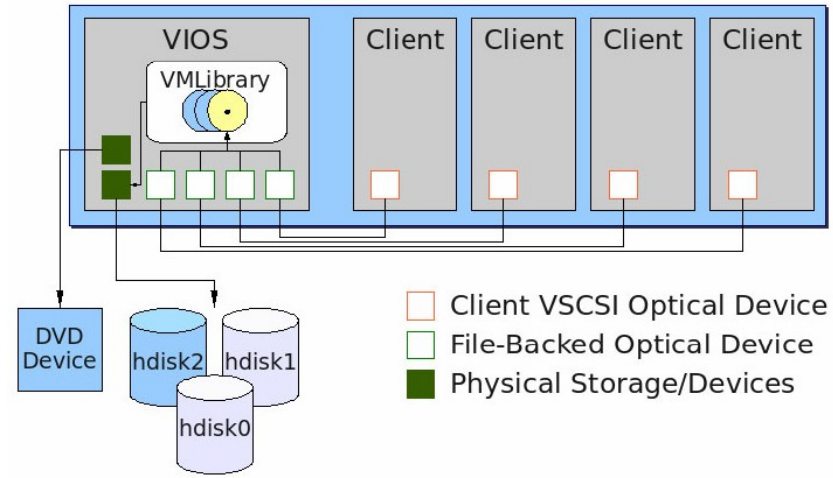
- **File Backed Virtual Disk:**
  A file in a VIO server filesystem (Storage Pool **SP**)
  can be mapped as client partition virtual disk.
  (in the same way as LVs or physical hdisk).
  mkbdsp –sp <SPname> -bd <file> -vadapter vhost<x>

  Better than LVs, when using VGs in single VIO internal disks, for virtual disks:
  - make system backup easier
  - allows easy transfer (ftp) of an AIX clone to another Escala server

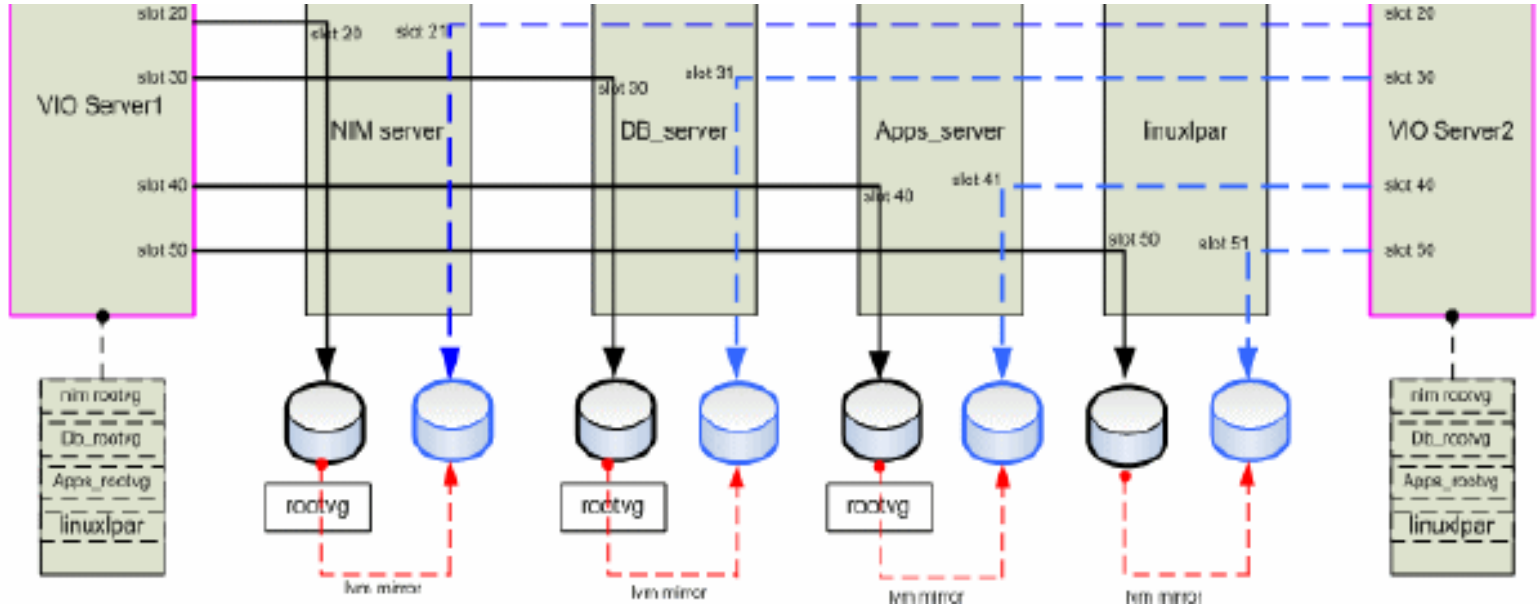


- **File Backed Virtual Optical Disk and Drive.**
  2 components:
  - Virtual Optical Drive = virtual DVD device for client partitions
  - Virtual Optical Disk = a file
    with the content of a physical CD / DVD
    or system backup made with *smit mkdvd*
    loaded into a Virtual Optical Drive

Virtual Media (*padmin*) Commands:
mkvdev –fbo… / rmrep / mkrep / lsrep,
mkvopt / rmvopt / lsvopt / loadopt / unloadopt …



Dual VIO Servers, Virtualizing LVs or Files (VIO 1.5)

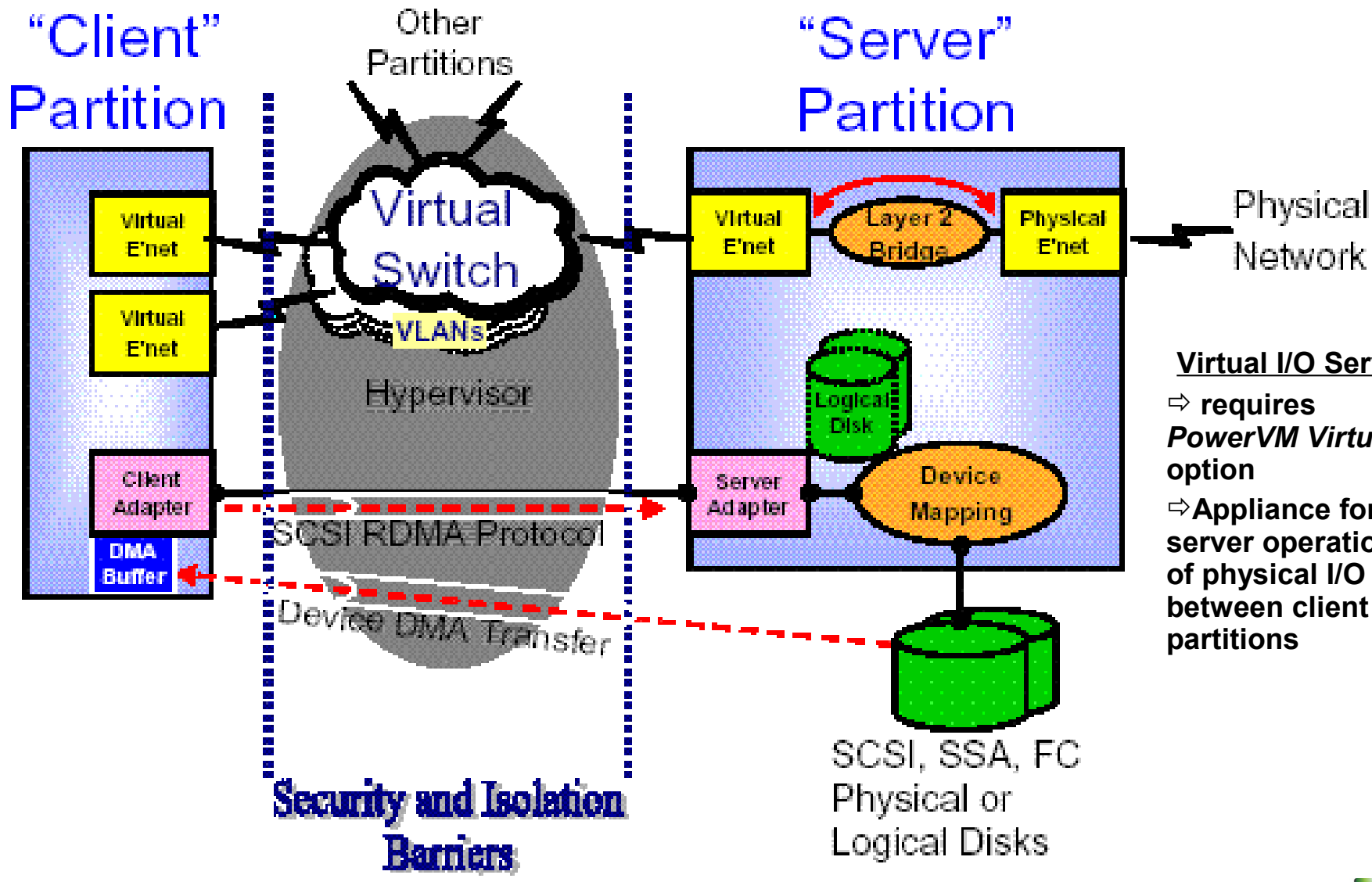


Two VIO Servers, each one with:

- **Physical disk adapter(s) and disk(s)**
- **Only one Virtual SCSI Server adapter (*vhost*) PER CLIENT partition and Logical Volumes** (not necessarily mirrored)**,** used as mirrored boot disks (***rootvg***) by client partitions.

- **Several Client partitions,** each one with**:**
  - **Two Virtual SCSI Client adapters (*vscsi*), each one connected to a different VIO Server**
  - ***rootvg* AIX-mirrored** on two virtual disks, each one mapped from a different VIO server.

Virtual I/O Operations



Virtual I/O Server:

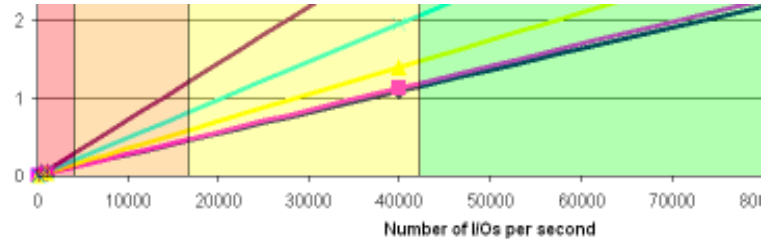
⇨ **requires *PowerVM Virtualization* option**

⇨**Appliance for virtual I/O server operations:sharing of physical I/O resources between client logical partitions**

Virtual I/O Server Sizing for VSCSI

- **With Shared or Dedicated Processors:**
  1) number of CPU instructions per IO operation of various sizes :

	4K	8K	32K	64K	128K
Phys. Disk	45000	47000	58000	81000	120000
LVM	49000	51000	59000	74000	105000



2) 15 Krpm disk drive is capable of approximately, **per second: 200 I/O's of 8KB** (1.5 MB/s) or 400 I/O's of 128KB (50 MB/s) **.**

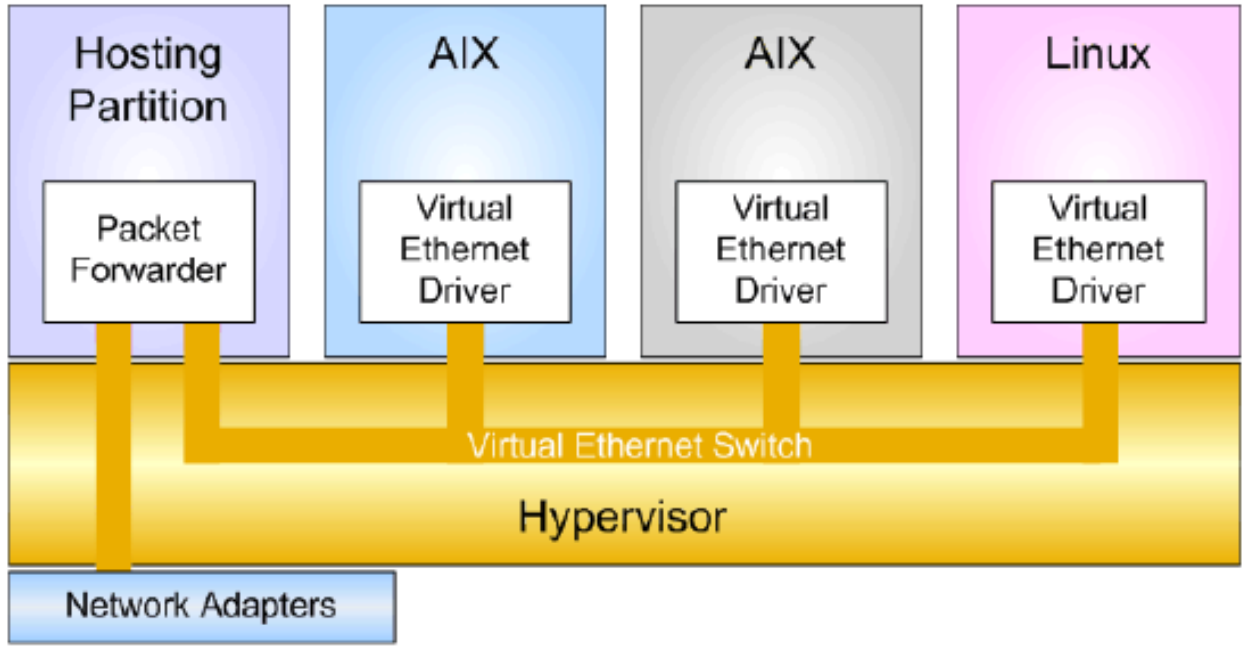
3) Ultra-320 SCSI adapter throughput: 320 MB/sec., Fibre Channel adapter 4Gbit/sec = 500 MB/sec.

➢ **8KB IO operations** on one disk need **200 IO/s x 47 000 inst**. = **9 400 000 CPU inst.**
With **1.65 GHz CPU** providing 1 650 000 000 inst./s.:
9 400 000 / 1 650 000 000 = **0.60 % CPU per disk or 160 disks driven by one CPU.**
Ultra-320 SCSI: **320 000 / 8KB** = 40 000 IO/s. **40 000 / 200 IO/s** corresponds to **200 disks**
FC adapter: **500 000 / 8KB** = 62 000 IO/s. 62 000 / 200 corresponds to **300 disks,**
600 disks with 2 adapters

➢ **128KB IO operations** on one disk need 400 IO/s x 120 000 inst. = **48 000 000 CPU inst.**
48 000 000 / 1 650 000 000 = **3 % CPU per disk or 30 disks driven by one CPU.**
Ultra-320 SCSI: **320 000 / 128KB** = 2 500 IO/s. 2 500 / 400 IO/s corresponds to **6 disks**
FC adapter: **500 000 / 128KB** = 4000 IO/s. 4000 / 400 IO/s corresponds to **10 disks,**
20 disks with 2 adapters

- **Memory allocation for the VSCSI server :**
  - AIX 5.3 can't start without a minimum of 512 MB of memory
  - **1GB memory is recommended with PowerPath** and is sufficient for large I/O configurations and very high data rates,. .

Virtual Ethernet and Shared Ethernet Adapter (SEA)



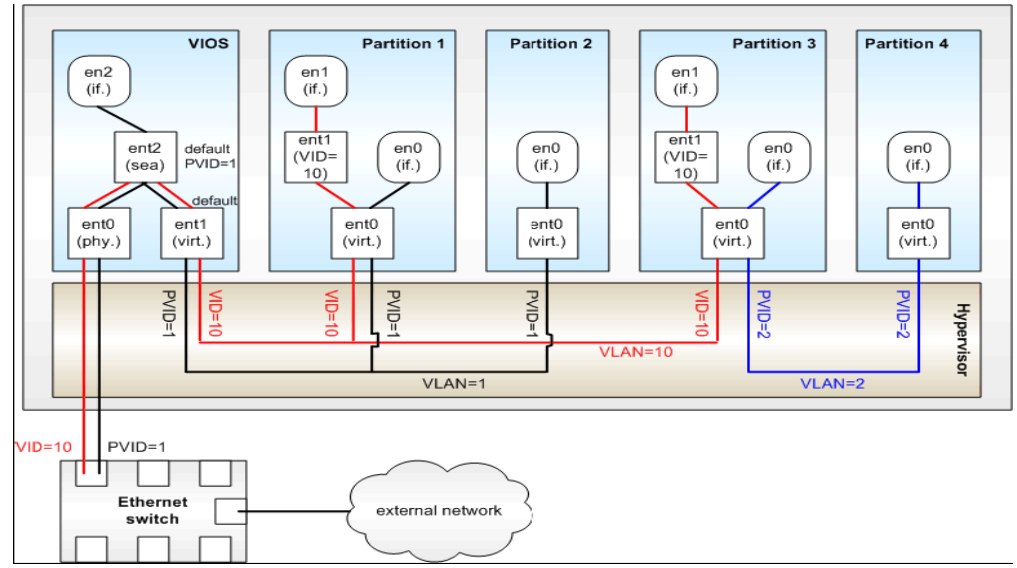
Virtual Ethernet :

- **Created at HMC level**
- **Fast interconnect between partitions:** Gbit/s and more
- Configured like a **standard Ethernet:** *smit mktcpip*
- "Virtual MAC" address based on partition ID and slot nb.
  xxxx6000**p**00**s**
- **Multiple virtual connections per partition :**
  - up to 256 Virtual Ethernet per Partition
  - each Virtual Ethernet can be attached to up to 21 VLANs (20 VLANs + Port VLAN ID)

Shared Ethernet Adapter (SEA):

- created in **I/O server partition** with *mkvdev* command:
- provides **IP forwarding** to external network **bridging** between **virtual and physical Ethernet**
- **Up to 16 virtual Ethernet** adapters to the same **physical via one Shared Ethernet Adapter** (SEA)
- **requires I/O server partition and *PowerVM Virtualization* option**

Virtual Ethernet and Virtual LANs (VLAN)

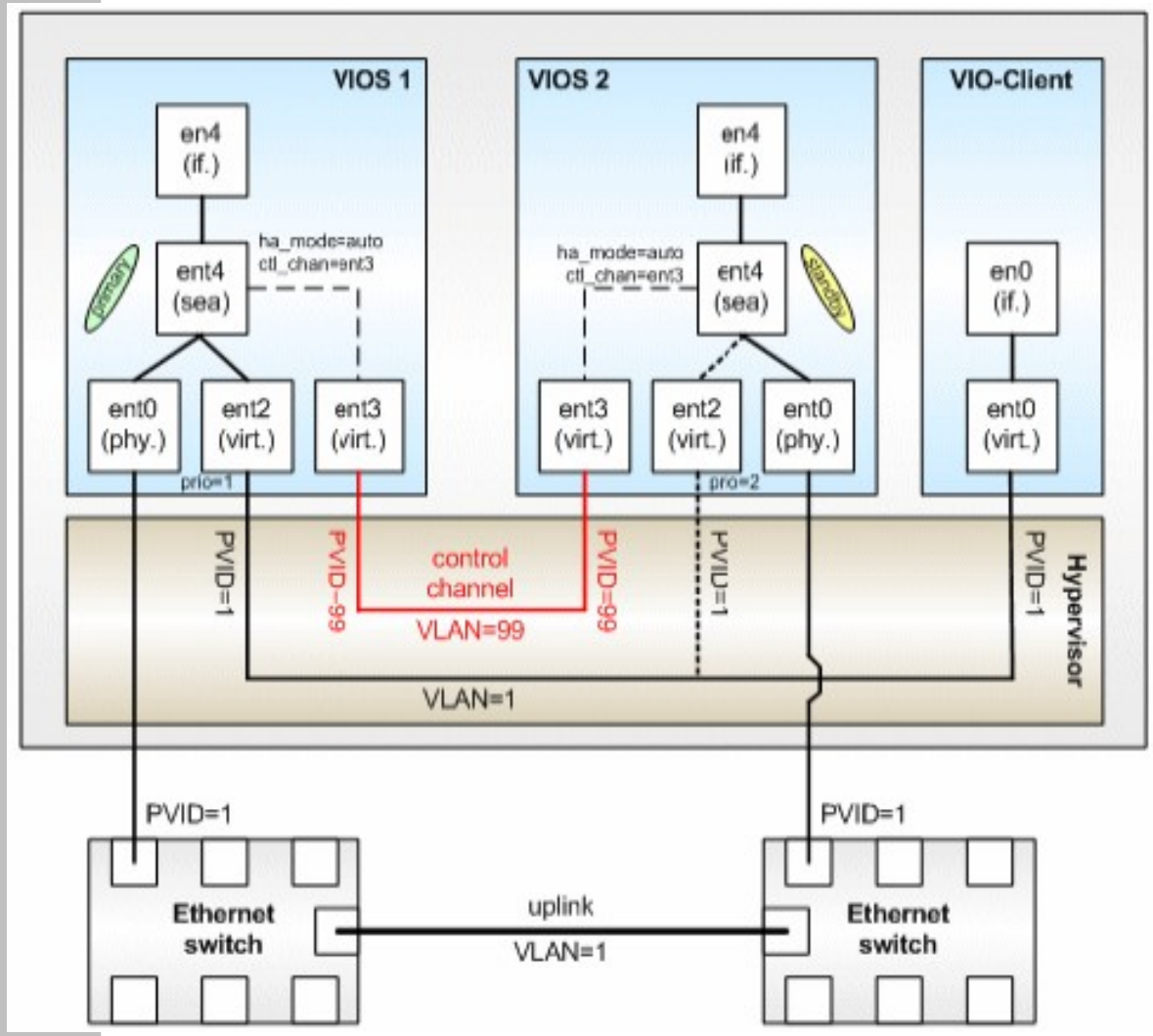


- **VID – VLAN ID: identifies a VLAN**
- **PVID – Port VLAN ID: is default VID of a switch port (or virtual adapter)**
- **If host sends untagged packets via base interface ent0, they are tagged with PVID by switch port or Hypervisor (virtual Ethernet port)**
- **Packets are tagged with VID by (AIX) host when sent by en<x> corresponding VLAN interface**

(created with *smit vlan*)

- **An adapter only accepts untagged packets or packets with a tag corresponding to its PVID or VIDs.**
- **Recommendations:**
  - If adapter connects to a **single network** (ex. **Partition 2** or **Partition 4**): use base interface ***en0*** and **PVID (=1 or =2)**
  - If **multiple networks** per adapter (ex. **Partition 1 or Partition 3**), add **VLAN IDs** and use ***smit vlan*** to create **additional *en1* interface** corresponding to **VLAN ID (ex.: VID = 10)**
- **A router that belongs to both VLAN segments** and forwards packets between them is required to communicate between hosts on different VLAN segments like HMC, NIM master or backup server

SEA Failover: High Availability for VLAN-aware client partitions



- **If client partitions are VLANs aware (using VLAN tagging), ALL virtual adapters must be on these same VLANS.**

  SEA failover mode between I/O servers must be used, instead of Etherchannel (which requires different VLAN IDs) :
  - **no load-balancing between VIOs**

On VIOS1, with HMC:

. Virtual Ethernet for SEA:
- Port Virtual LAN ID=**1**
- **Access external network** button checked
- Trunk priority=**1**

. Virtual Ethernet for control channel:
- Port Virtual LAN ID=**99**

On VIOS2, with HMC:

. Virtual Ethernet for SEA:
- Port Virtual LAN ID=**1**
- **Access external network** button checked
- Trunk priority=**2**

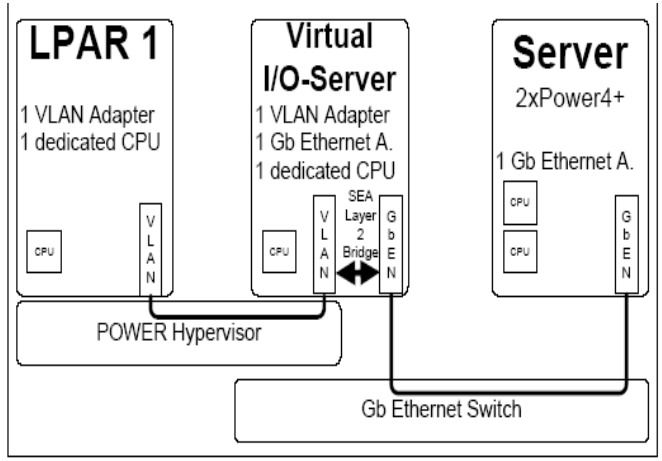
. Virtual Ethernet for control channel:
- Port Virtual LAN ID=**99**

When heartbeat mechanism of control Channels detects failure of VIOS1 with priority=1, SEA in VIOS2 priority=2 will take over.

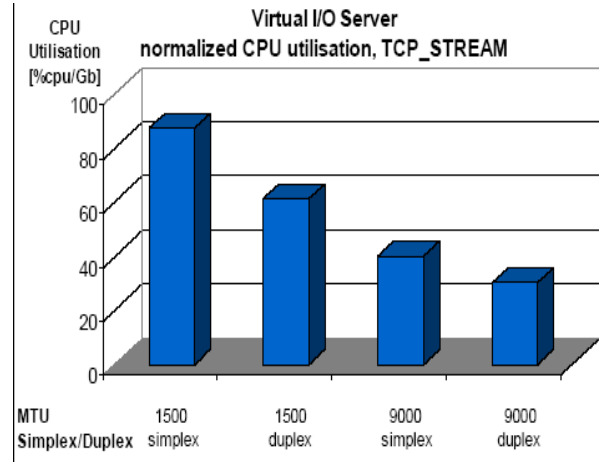
Virtual I/O Server Sizing for Ethernet SEA

Planning for the Virtual I/O Server -> Capacity Planning

- **Virtual Ethernet I/O Server Sizing**



CPU Utilization in the Virtual I/O Server for 1Gbit/s SEA data throughput.



- **1 Gigabit/sec Ethernet throughput** (with a MTU size of 1500) uses about **80% of a 1.65GHz processor.**

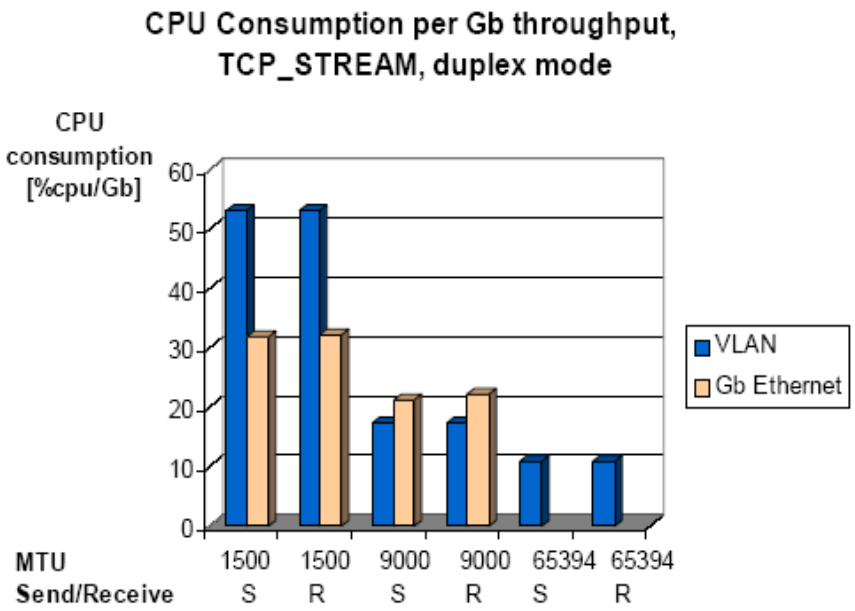
- **Shared Ethernet Adapter on Virtual I/O Server usage not recommended:**
  - if heavy network traffic between Virtual LANs and local networks
  - for latency critical applications

  use a dedicated Ethernet adapter instead.

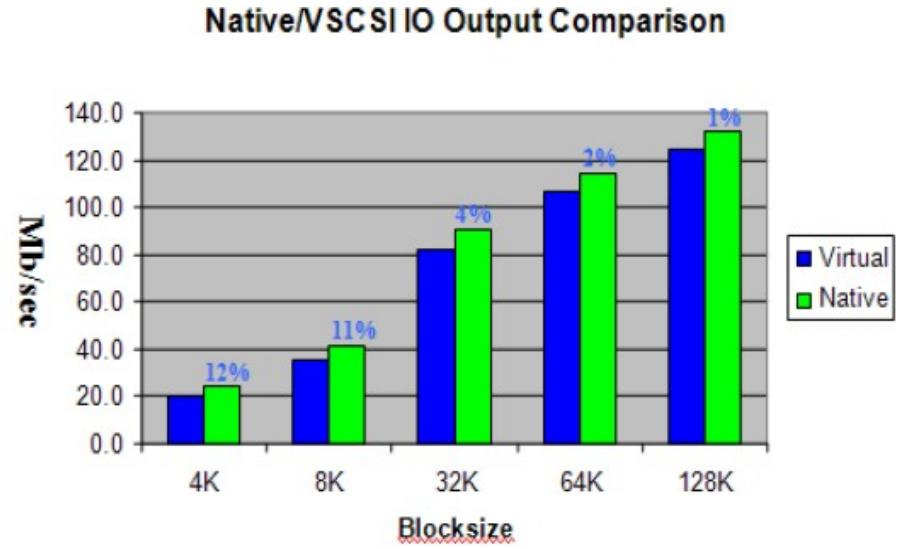
Virtualization and Performances

Shared Processors, Micro-partitioning Overhead:
NFS test, throughput measured on 4 partitions, each one with **1 physical dedicated processor**, then on 4 micro-partitions with shared processors and **1.00 entitlement capacity**.
The throughput in each partition was the same in both cases. Processor usage is about 2% higher in the case of shared processor partitions. Latency to dispatch Virtual Processor, when several micro-partitions share the same physical processor, increase response time,

Virtual Ethernet versus physical Gbit Ethernet Processor (1.65GHz) Consumption for 1Gbit/s in CLIENT partition



Native to Virtual Disk I/O Output comparison

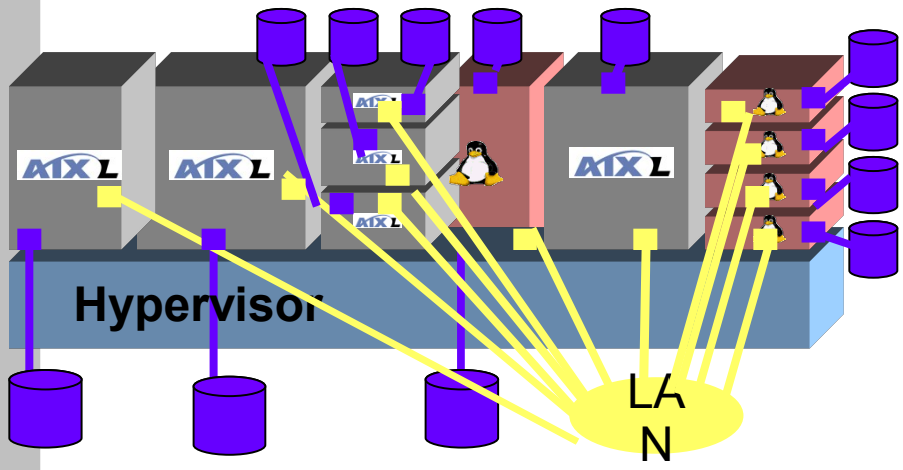


Benefits of Virtualization

Virtualization doesn't improve performances **but optimizes resources utilization:**

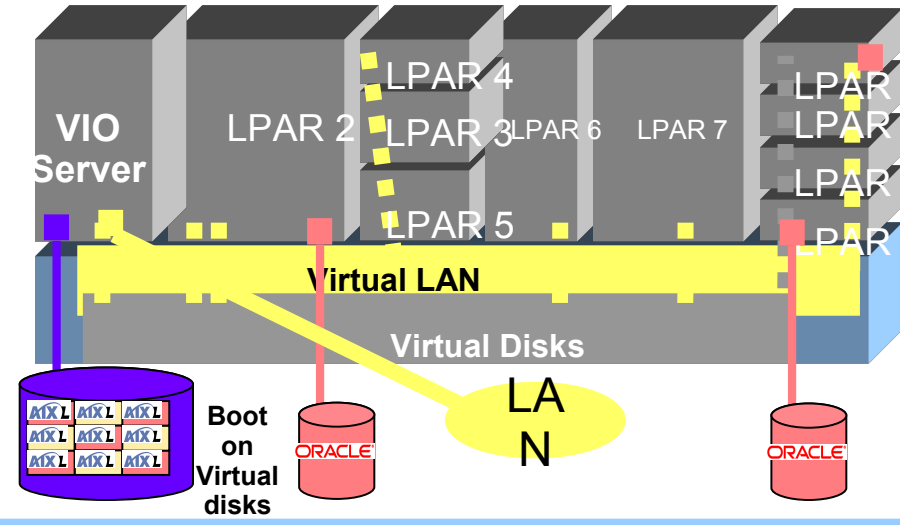
- ☐ **CPU usage:** ex. PL250, 2 CPUs, one micro-partition with 80%=1.60 and another one with 20%=0.40 resources. With dedicated CPUs each partition would have one CPU = 50% of resources
- ☐ **Disk and Network bandwidth and connectivity**
- ☐ **Easy to dynamically create / remove a partition**

Servers or DLPAR Without Virtualization



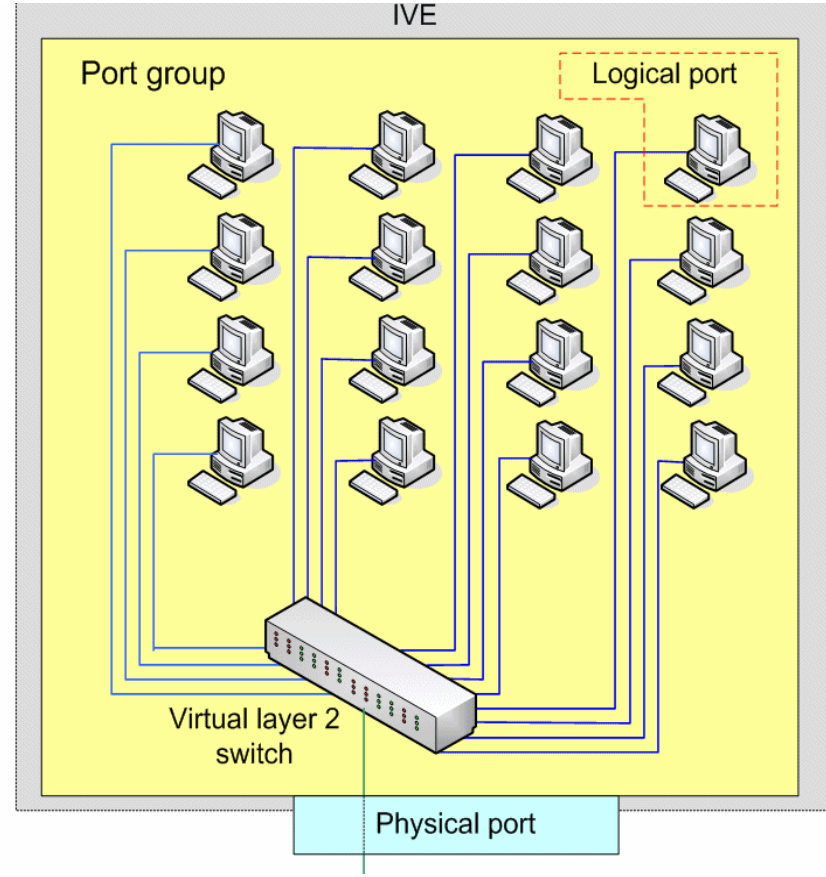
Minimum PER PARTITION (or server) :
1 CPU
1 SCSI adapter + 1 disk
1 LAN adapter

Escala PLx50 and Virtualization



-254 partitions maxi. -> only 160 I/O slots
- Less CPUs, adapters and disk drawers, network and SAN switch ports, LUN masking, …
- **Virtual SCSI is GOOD for BOOT disk (SAN)**
- **Virtual Ethernet, e.g.: fast internal NIM network**

Other new features related with Power6: SPPool, HEA, AMS, NPIV and LPM



LIBERATE IT

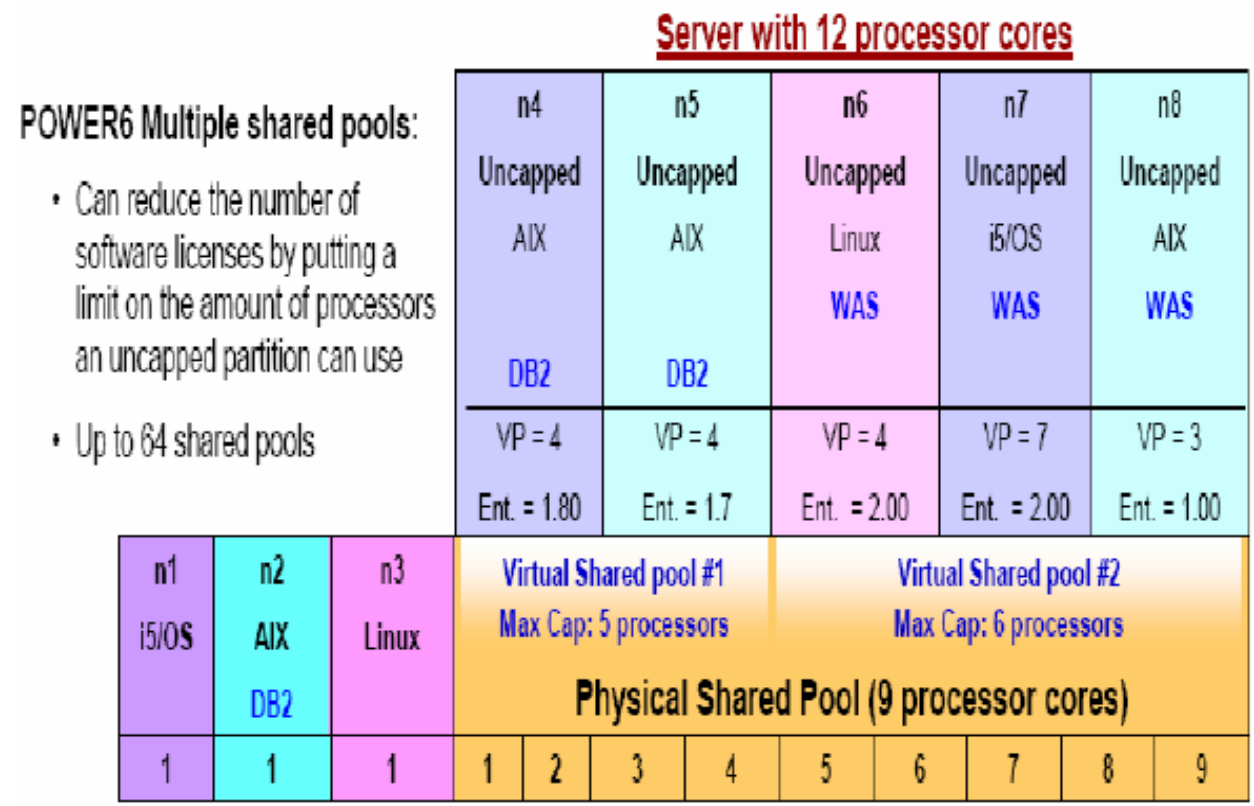
Processor Pools

LIBERATE IT

Power6: **Multiple Shared Virtual Processor Pools**

- Primary motivation: reduce cost of licenses for micro-partitions when based on the number of CPU in the virtual pool (HACMP, Oracle, …)
  - Uncapped partition limited to **Virtual Pool Maximum number of CPUs** not Physical Shared Pool size

Up to 64 pools,
HMC V7 R320, not IVM
Firmware EM320,
AIX 5.3 TL07 (mini).



- **Ex.: Pool #1 Partitions: n4, n5 Σ Entc = 3.50 VP = 4 < Max. Cap.= 5 CPU < 9 CPU**
  Pool #2 Partitions: n6, n7, n8 Σ Entc = 5.00 VP = 7 < Max. Cap.= 6 CPU < 9 CPU
  Phys. Pool Pool Ent. Capacity Σ = 8.50 < 9 CPU

HEA – Host Ethernet Adapter

LIBERATE IT

How to use Host Ethernet Adapter

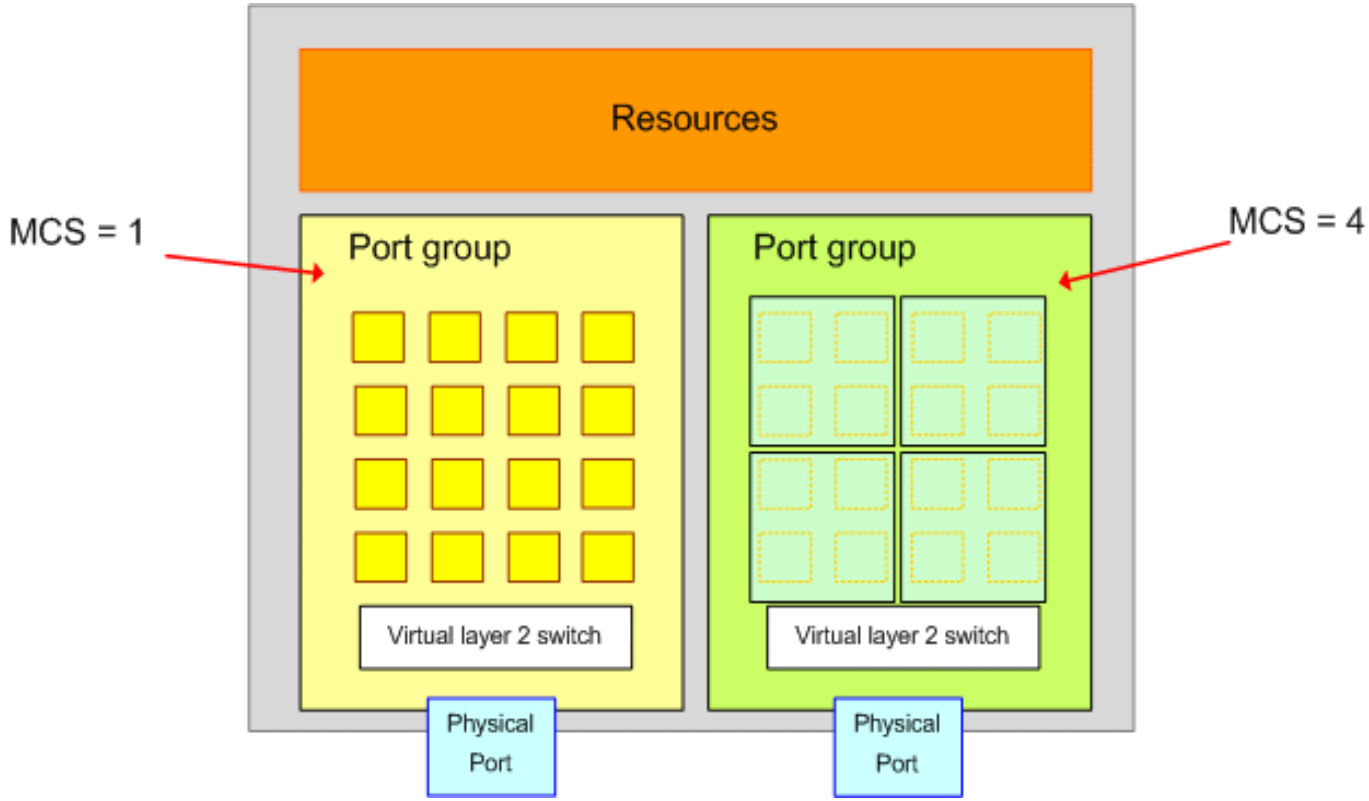
- **3 possibilities:**
  - Escala Server is **not partitioned** and not management by an HMC:
    - Use physical HEA ports, configured within the operating system (SMIT).
  - **Logical HEA** (LHEA) allows up to 16-32 partitions in the same server to directly use and share a same HEA port group, instead of SEA in VIO-Server.
  - SEA with port „**promiscuous mode**“ in VIO-Server:
    - To connect more than 16-32 partitions through ONE HEA port or with Partition Mobility or AMS where all adapters must be virtual.

Physical HEA to Logical LHEA

- **Each physical port has it's own location code.**
- **Physical ports are associated with a Port Group**
- **Each port Group can have up to 16 LHEA (16/MCS Multi-Core Scaling factor)**
  - **Dual port** 1Gbps HEA provides one port group: up to *16 LHEAs*
  - **Quad port** 1Gbps and Dual port 10Gbps HEA provide two port groups up to *32 LHEAs*
  - Each Logical Port (LHEA) can be owned by a separate partition
  - A partition can have **only one LHEA** logical port **per Physical HEA port**
  - A partition can have multiple LHEA logical ports, each one **on a different HEA port**

Multiple Core Scaling factor

- **Each HEA port group has 16 receive (RX) and transmit (TX) queue pairs sp (QP)**
  - To break network traffic into multiple streams dispatched to multiple Power6 processors and take advantage of parallel processing.
- **Multi-Core Scaling (MCS) value**:
  - Each LHEA port of the sane HEA port group uses an „MCS“ number of queue pairs (Qps)
  - Thus the number of LHEA ports in an HEA port group is 16/MCS
  - With the default MCS=4, the number of LHEA ports per port group is 4 (16 LHEA ports if MCS=1)



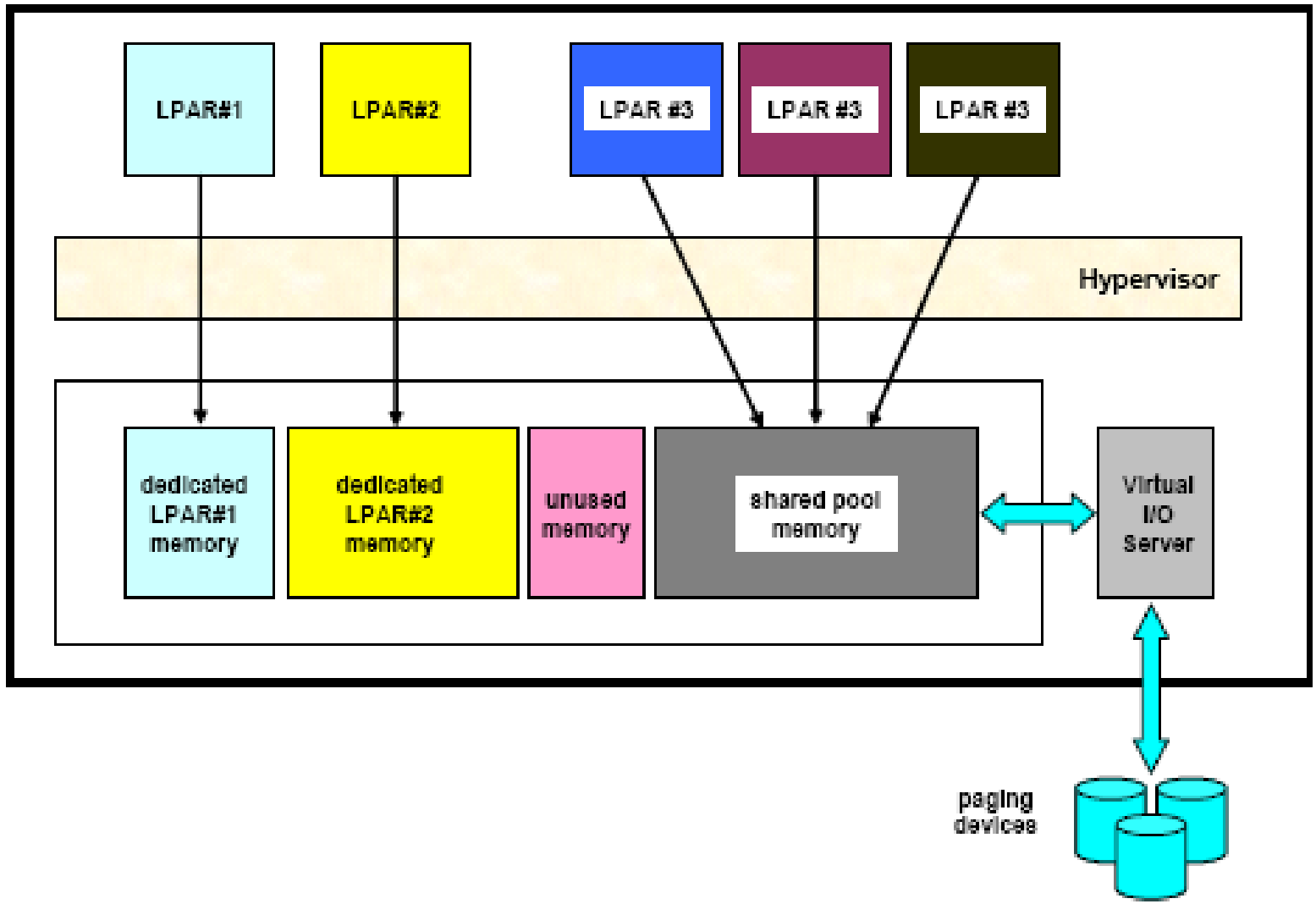
AMS – Active Memory Sharing

LIBERATE IT

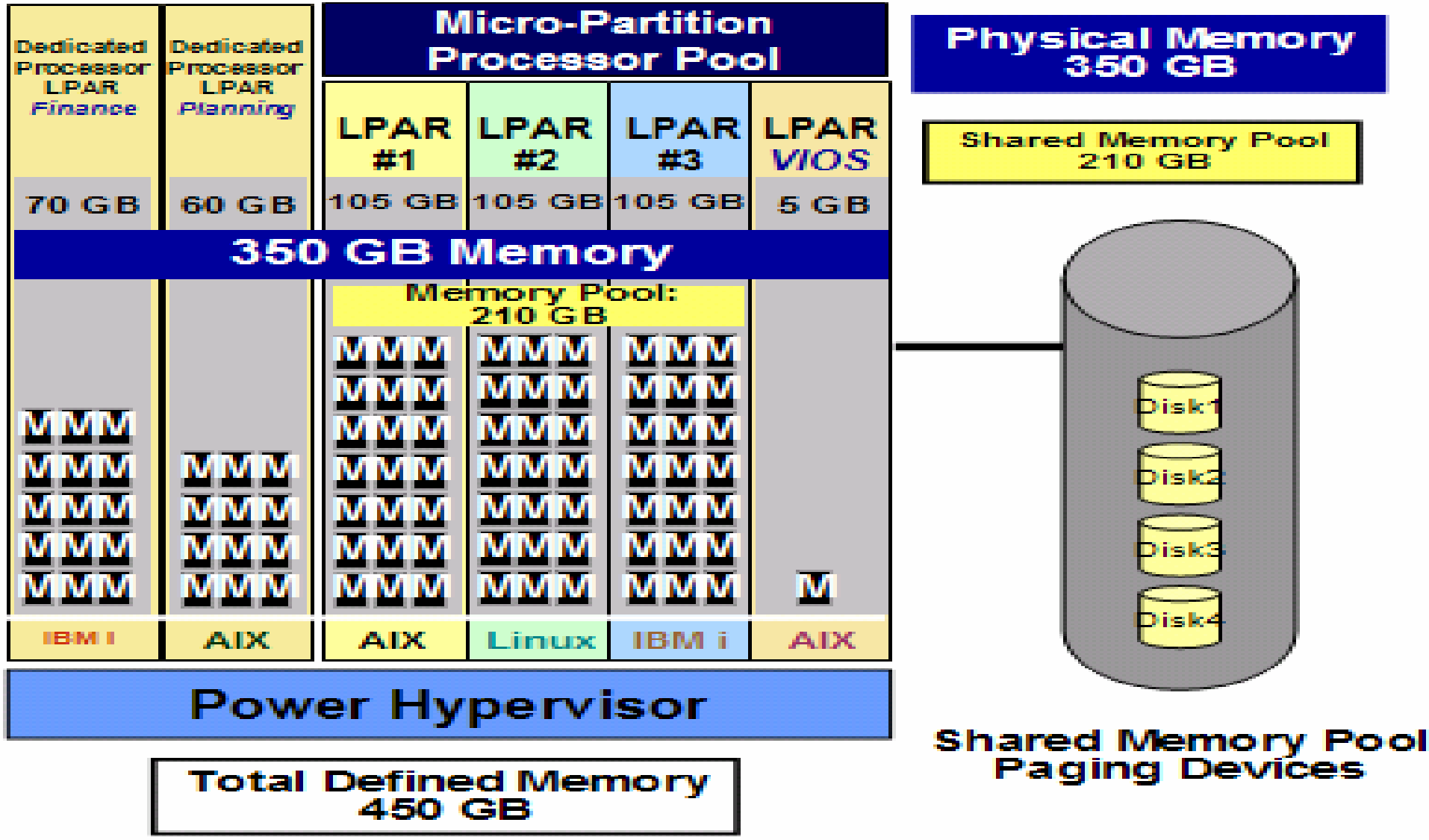
AMS – Active Memory Sharing – Why ?

- **CPU and I/O resources are already consolidated for better exploitation.**
- **DLPAR was the only way to move memory between partitions:**
  - **Can be slow**, depending on the amount of memory
  - **May fail,** if application does not release the memory
  - Application needs **to be aware** of memory changes
  - **Administrator have to initiate** the movement of memory

AMS – Active Memory Sharing



AMS – Example



NPIV – FC N_Port ID Virtualization

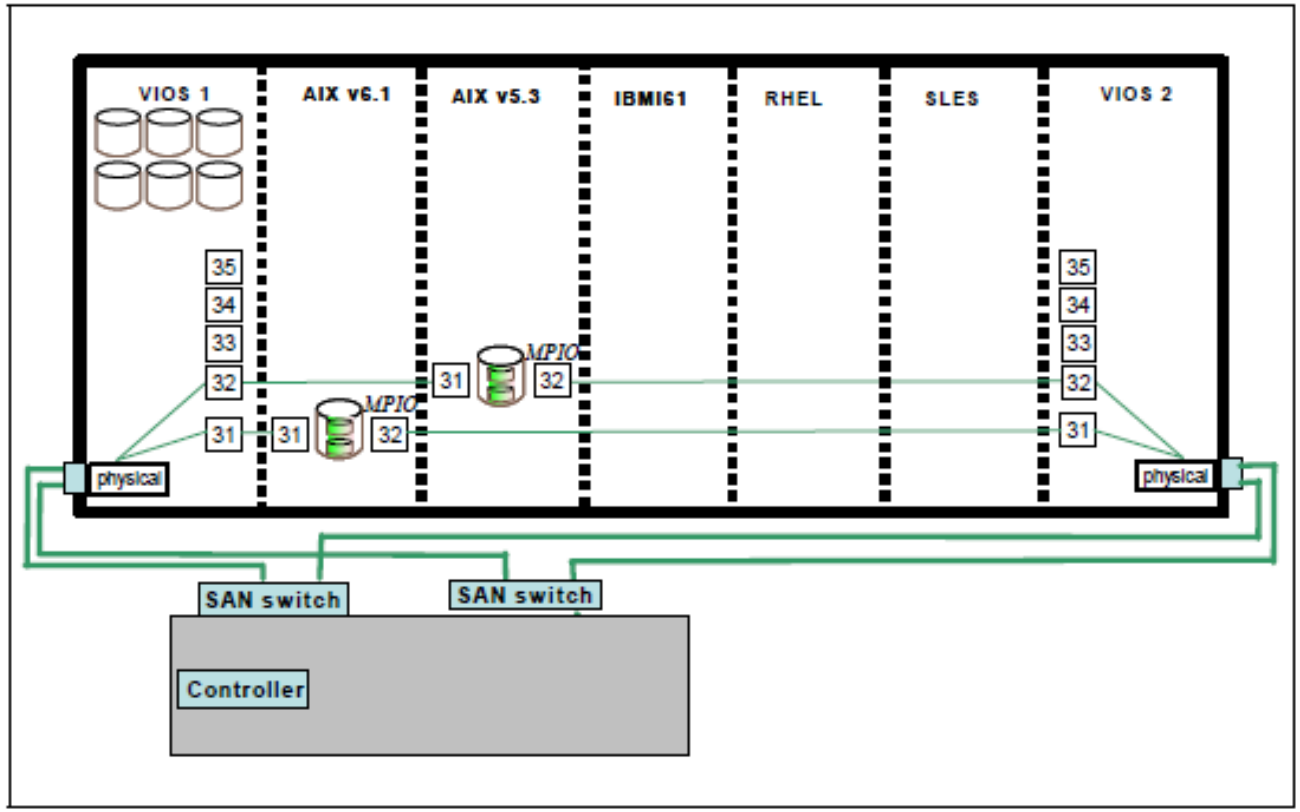
LIBERATE IT

NPIV – FC N_Port ID Virtualization

- **VIOS 2.1.0.10 (FP20.1) supports N_Port ID Virtualization for new 8 Gb FC HBA**

  - NPIV is a FC standard wich allows a physical HBA to be logically partitioned into multiple logical ports (WWNs) so that it can support multiple initiators.
  - **N_Port Virtulization means:**
    - **Have one physical port which can be divided in many up to 64 logical (virtual) ports.**
  - VIOS provision FC logical ports to client LPARS, rather than individual LUNs.
  - Client partitions with this type of logical port operate as though they have dedicated FC adapter.
  - Performances for disk access with NPIV similar to dedicated FC adapter (virtual SCSI = -5%)

NPIV – FC N_Port ID Virtualization



LPM – Live Partition Mobility

LIBERATE IT

LPM - Benefits

- **Allows to move a partition, between two physical Power6 Escala servers**
  - inactive (unbooted) partition: move LPAR definition to another server
  - **active partition, without taking the application off-line**

- Benefits: increase data center optimization
  - Reduce impact of planned outages for hardware maintenance
  - Relocate workloads to optimize existing resources utilization

- **Doesn't replace ARF or HA for automatic failover:**
  because both Escala servers must be **up and running** when moving.
  - ARF reactivates application from a DEAD partition to a LIVE partition
  - LPM moves a partition from a LIVE server to a server with available "DEAD" resources

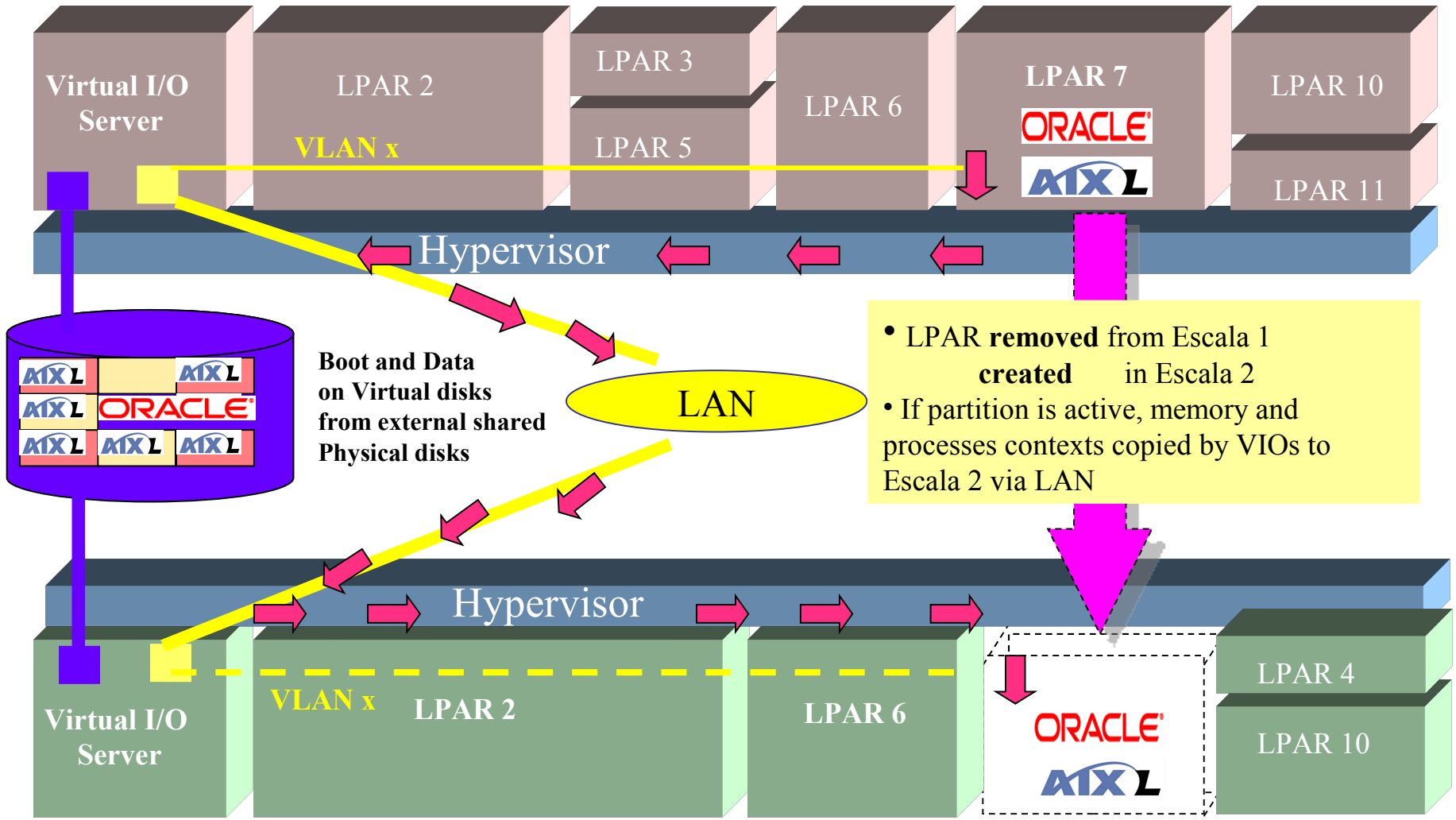
- **Machines must be close together:**
  1 Gbps network inter-connection (dedicated network recommended)
  Direct access to Virtual disks in External Shared (SAN) Disk Subsystems

LPM - Configuration

Configuration Requirements:

- **VIOS** with "**Mover Service Partition**" attribute and VASI adapters
- VIOS inter-connected via (1Gbps) Ethernet network
- Virtual disks of mobile partitions already configured in VIOS of BOTH Escala systems
- Partition **without dedicated physical adapters**:
  - **boot and data disks virtualized from external disks (SAN LUNs)**
    visible by **VIOS of both Power6** systems
    disks **without reservation** attribute at VIO level
  - **virtual Ethernet via VIO SEA (no LHEA)**
  - **FC Adaptaters 8Gb NPIV virtualized**
- **Enough free CPU / Memory resources** in destination machine to create partition

LPM - Cross-PLx60 Partition Migration



Architect of an Open World™

LIBERATE IT