



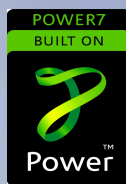
# PowerVM Processors

Gareth Coates

[gaz@uk.ibm.com](mailto:gaz@uk.ibm.com)

IBM Power Systems

Advanced Technology Support, EMEA



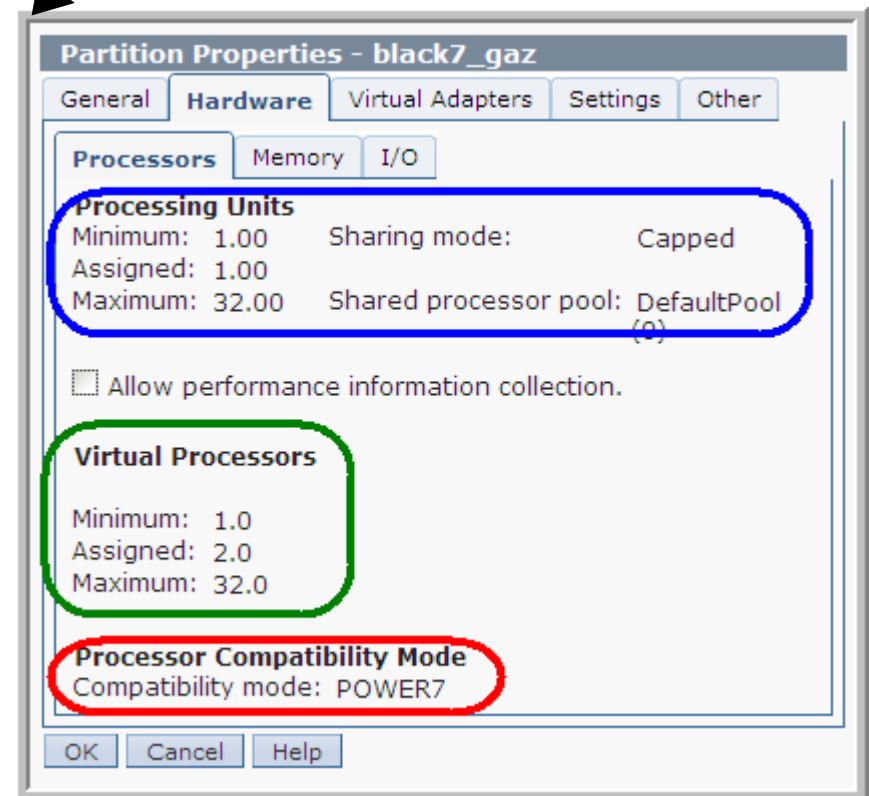
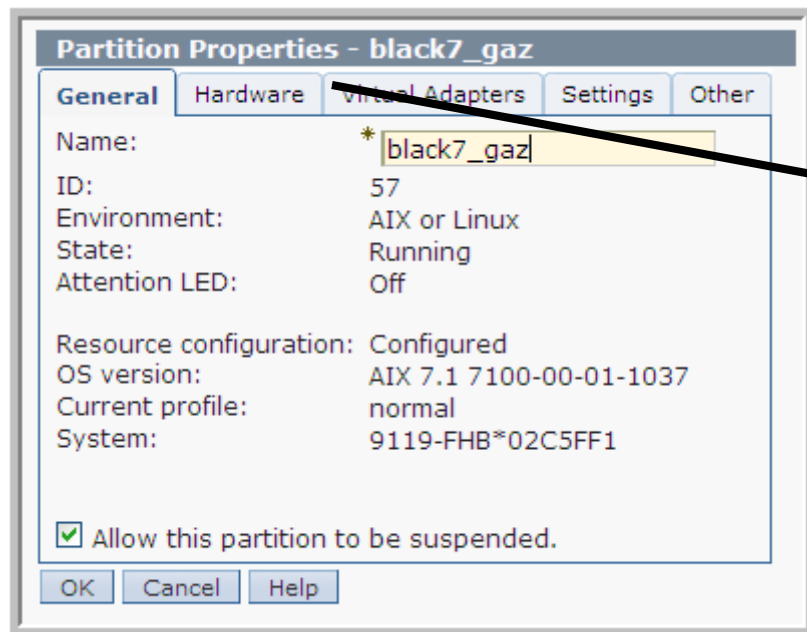
© 2011 IBM Corporation

## Message to presenter

- Remember to start recording



# Properties of the VM or LPAR



## Properties of the profile

ibm.com https://hmc11.aixncc.uk.ibm.com/hmc/wd/T866

**Logical Partition Profile Properties: normal @ black7\_gaz @ black-9119-FHB-02C5FF1 - black7\_gaz**

General **Processors** Memory I/O Virtual Adapters Power Controlling Settings HCA

Detailed below are the current processing settings for this partition profile.

**Processing mode**

Dedicated  
 Shared

**Processing units**

Total managed system processing units : 64.00  
Minimum processing units :   
Desired processing units :   
Maximum processing units :   
Shared processor pool:

**Virtual processors**

Minimum processing units required for each virtual processor : 0.10  
Minimum virtual processors :   
Desired virtual processors :   
Maximum virtual processors :

**Sharing mode**

Uncapped      Weight :

Processor compatibility mode:

OK Cancel Help

## Processor

- During Virtual Machine Start up
  - Minimum - If less than this available = do not start
  - Desired - What you really would like
  - Maximum – Ignored
- Hopefully, you get Desired but might get less
- Minimum used to stop VM starting with so little CPU time that it would cause problems
- Once running what you got is called Entitlement = how much CPU the LPAR is entitled to use.

## Processor

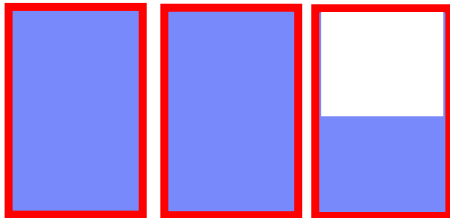
- Virtual Machine + Dynamic Online Changes
  - Minimum - Entitlement can't go lower than this
  - Desired - Ignored
  - Maximum – Entitlement can't go higher than this
- Sanity checks  
so operators can't do too much damage!!
- Often a “pain in the #&\$%”  
as you can't change it while running

## Processor Pool

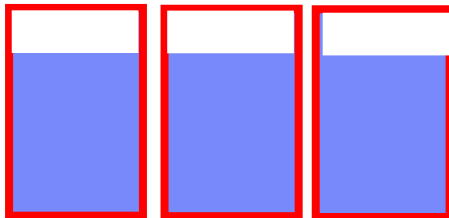
- Used to sub divide all the CPUs
  - ROT: less than 8 CPUs not used much – use Default Pool
  
- Why?
  - License control – only pay for pool's CPU for software
    - 4 CPUs for VIOS
    - 30 CPU for 4 DB2 VMs      – only pay for 30 licenses
    - 30 CPU for 2 Oracle VMs      – only pay for 30 licenses
    - 64 CPU for 16 WAS VMs      – only pay for 64 licenses
  
  - Prioritise CPUs for important production VMs
    - 48 CPUs for production
    - 8 CPUs for dev/test

## Virtual Processor

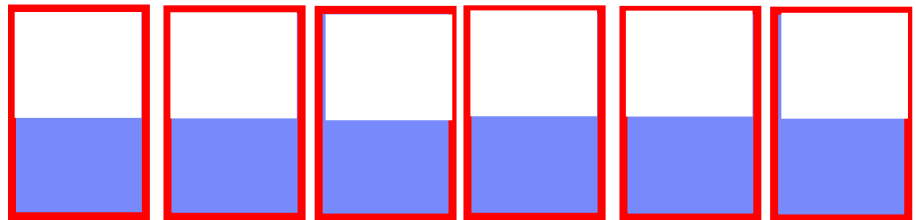
- A major confusion to many
  - Physical, logical and now virtual CPUs!!
- If it was called “spreading factor” it would help
  - The Entitlement CPU cycles are spread across the VP



E=2.5 spread across 3 VP



E=2.5 spread across 6 VP







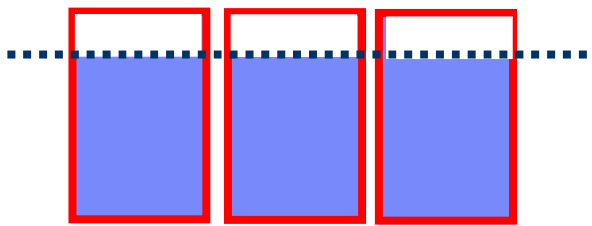
## Virtual Processor

- VP must be same or higher than Entitlement
  - Entitlement=2.5 CPU
  - Obviously can't run on just 2 CPUs
  - So VP must be 3 or more
- Must have 0.1 of a CPU or more for every VP
  - Mandatory and good for efficiency
- Have a physical CPU for every VP in the pool/box
  - More is dumb!

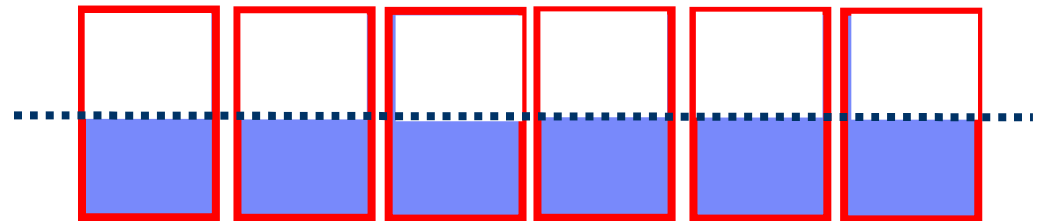
## To cap or not to cap

### ■ Capped

- The VM can't use more than Entitlement, unused cycles go in the pool for other VM's or are wasted
- Can be used for SW license control



E=2.5 on 3 VP



E=2.5 on 6 VP

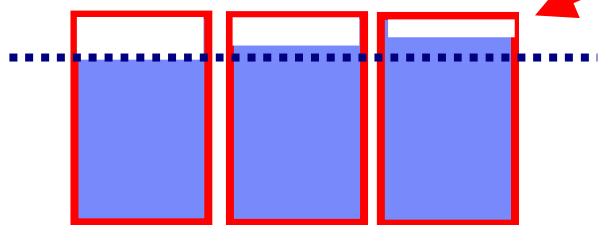
## Cap or uncap

### ■ Capped

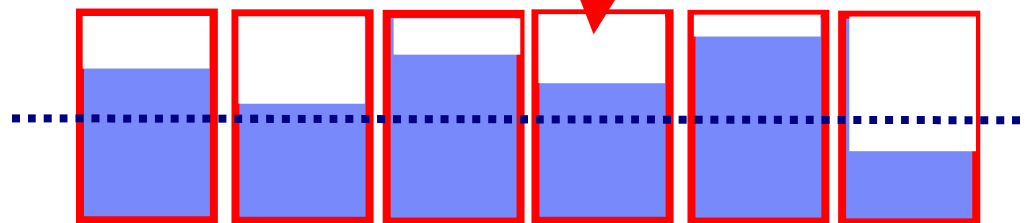
- The VM can't use more than Entitlement, unused cycles go in the pool for other VM's or are wasted
- Can be used for SW license control

### ■ UnCapped

- This VM competes for spare pool CPU cycles and can get more work done
- Entitlement = 2.5
- VP=3 up to 3 CPU
- VP=6 up to 6 CPU



E=2.5 on 3 VP



E=2.5 on 6 VP

## Weight factor

- 0 to 255 (0  $\equiv$  capped)
  - but you can tweak it dynamically
  - ~~– Toggling Capped/Uncapped requires a VM restart~~
- How much of the "spare" can each VM get?
  - The weights of all VM's who have had their entitlement are compared and unused resource is shared accordingly
- Let's do the arithmetic



# Weight factor

$$\text{AdditionalCapacityShare} = \frac{UCk \times \frac{WPn}{rP}}{\sum We}$$

Don't panic :-)

## Where:

### AdditionalCapacityShare

Share of unused processing capacity to be allocated to a particular partition (in *processor units x 100*)

UCk Unused processor capacity available in their Shared-Processor Pool for the dispatch window (in processor units)

WPn Uncapped weight of the particular uncapped micro-partition

rP The number of runnable (eligible) micro-partitions for this dispatch window

$\sum We$  Sum of the uncapped weights of all runnable uncapped micro-partitions

## Weight factor

- Consider 3 VMs



8 physical CPUs installed and activated	CE	VP	WEIGHT
PROD	6.0	8	255
TEST	1.0	8	1
DEV	1.0	8	1

- If PROD is idle, DEV and TEST can use all of the 8 CPUs between them and have equal shares
- If PROD becomes busy, it will take its CE up to 6.0 and if eg: TEST is idle and DEV is busy, it will get nearly all of TEST's entitlement, DEV getting very little.

# Properties of the VM or LPAR

**Partition Properties - black7\_gaz**

General Hardware **Virtual Adapters** Settings Other

Name: \*black7\_gaz

ID: 57

Environment: AIX or Linux

State: Running

Attention LED: Off

Resource configuration: Configured

OS version: AIX 7.1 7100-00-01-1037

Current profile: normal

System: 9119-FHB\*02C5FF1

Allow this partition to be suspended.

OK Cancel Help

**Partition Properties - black7\_gaz**

General **Hardware** Virtual Adapters Settings Other

Processors Memory I/O

**Processing Units**

Minimum: 1.00	Sharing mode: Capped
Assigned: 1.00	
Maximum: 32.00	Shared processor pool: DefaultPool (0)

Allow performance information collection.

**Virtual Processors**

Minimum: 1.0
Assigned: 2.0
Maximum: 32.0

**Processor Compatibility Mode**

Compatibility mode: POWER7

OK Cancel Help

# Properties of the profile

ibm.com https://hmc11.aixncc.uk.ibm.com/hmc/wd/T866

**Logical Partition Profile Properties: normal @ black7\_gaz @ black-9119-FHB-02C5FF1 - black7\_gaz**

General **Processors** Memory I/O Virtual Adapters Power Controlling Settings HCA

Detailed below are the current processing settings for this partition profile.

**Processing mode**

Dedicated  
 Shared

**Processing units**

Total managed system processing units : 64.00  
Minimum processing units :   
Desired processing units :   
Maximum processing units :   
Shared processor pool:

**Virtual processors**

Minimum processing units required for each virtual processor : 0.10  
Minimum virtual processors :   
Desired virtual processors :   
Maximum virtual processors :

**Sharing mode**

Uncapped Weight :

Processor compatibility mode:

OK Cancel Help



## What can be changed Dynamically

- Lots of things – to make your life easy
- Capacity Entitlement
  - +/- 0.01 between MIN and MAX
- Virtual Processors
  - integers between MIN and MAX
- Uncapped Weight
  - integers between 0 and 255
- Memory
  - LMBs between MIN and MAX

**LMB=Logical Memory Block**

The granularity that the hardware uses for allocating memory.

Between 16 and 256MB.

# Iparstat -i dedicated

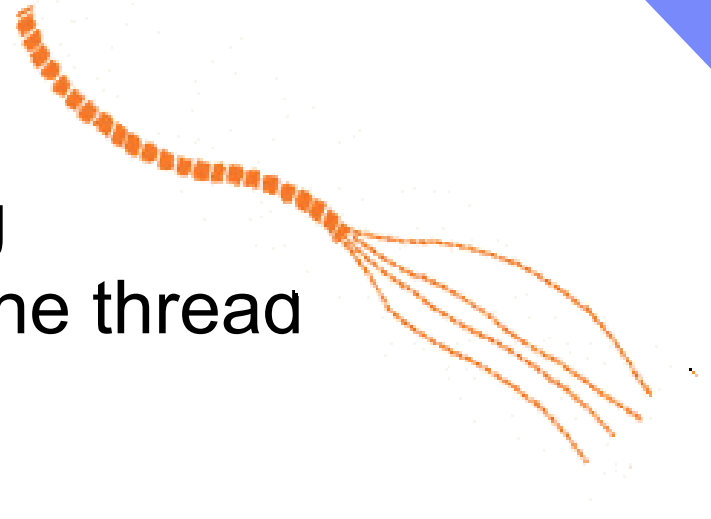
```
# lparstat -i
Node Name : mantova
Partition Name : purple3-hpc
Partition Number : 13
Type : Dedicated-SMT-4
Mode : Capped
Entitled Capacity : 17.00
Partition Group-ID : 32781
Shared Pool ID : -
Online Virtual CPUs : 17
Maximum Virtual CPUs : 32
Minimum Virtual CPUs : 1
Online Memory : 32768 MB
Maximum Memory : 131072 MB
Minimum Memory : 2048 MB
Variable Capacity Weight : -
Minimum Capacity : 1.00
Maximum Capacity : 32.00
Capacity Increment : 1.00
Maximum Physical CPUs in system : 64
Active Physical CPUs in system : 32
Active CPUs in Pool : -
Shared Physical CPUs in system : 0
Maximum Capacity of Pool : 0
Entitled Capacity of Pool : 0
Unallocated Capacity : -
Physical CPU Percentage : 100.00%
Unallocated Weight : -
Memory Mode : Dedicated
Total I/O Memory Entitlement : -
Variable Memory Capacity Weight : -
Memory Pool ID : -
Physical Memory in the Pool : -
Hypervisor Page Size : -
Unallocated Variable Memory Capacity Weight : -
Unallocated I/O Memory entitlement : -
Memory Group ID of LPAR : -
Desired Virtual CPUs : 17
Desired Memory : 32768 MB
Desired Variable Capacity Weight : -
Desired Capacity : 17.00
Target Memory Expansion Factor : -
Target Memory Expansion Size : -
Power Saving Mode : Disabled
```

# lparstat -i shared

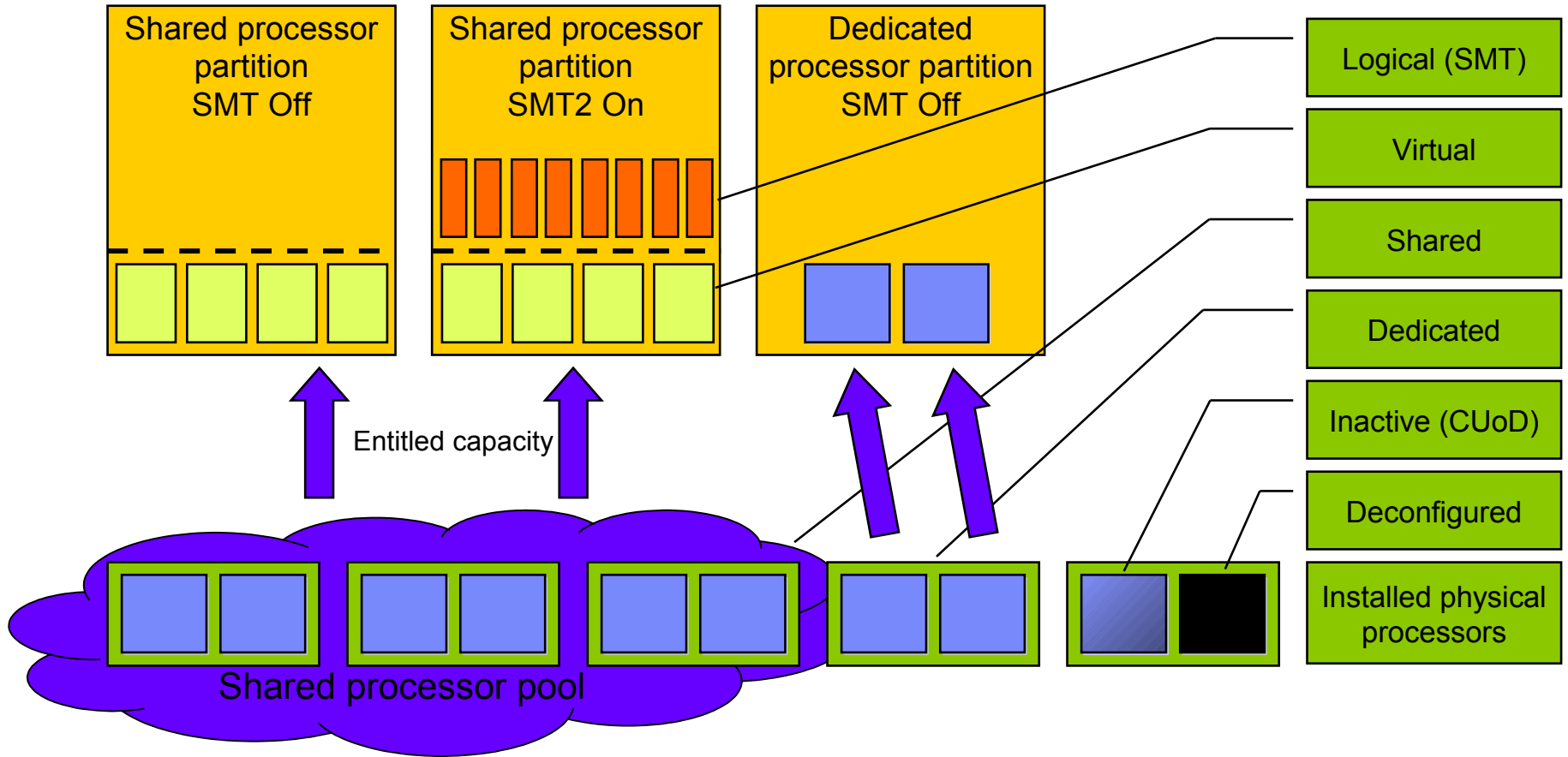
```
# lparstat -i
Node Name : mantova
Partition Name : purple3-hpc
Partition Number : 13
Type : Shared-SMT-4
Mode : Uncapped
Entitled Capacity : 6.00
Partition Group-ID : 32781
Shared Pool ID : 0
Online Virtual CPUs : 8
Maximum Virtual CPUs : 30
Minimum Virtual CPUs : 2
Online Memory : 32768 MB
Maximum Memory : 131072 MB
Minimum Memory : 2048 MB
Variable Capacity Weight : 128
Minimum Capacity : 2.00
Maximum Capacity : 24.00
Capacity Increment : 0.01
Maximum Physical CPUs in system : 64
Active Physical CPUs in system : 32
Active CPUs in Pool : 32
Shared Physical CPUs in system : 32
Maximum Capacity of Pool : 3200
Entitled Capacity of Pool : 1763
Unallocated Capacity : 0.00
Physical CPU Percentage : 75.00%
Unallocated Weight : 0
Memory Mode : Dedicated
Total I/O Memory Entitlement : -
Variable Memory Capacity Weight : -
Memory Pool ID : -
Physical Memory in the Pool : -
Hypervisor Page Size : -
Unallocated Variable Memory Capacity Weight : -
Unallocated I/O Memory entitlement : -
Memory Group ID of LPAR : -
Desired Virtual CPUs : 8
Desired Memory : 32768 MB
Desired Variable Capacity Weight : 128
Desired Capacity : 6.00
Target Memory Expansion Factor : -
Target Memory Expansion Size : -
Power Saving Mode : Disabled
```

## SMT & Logical CPUs

- Simultaneous Multi-Threading
- The ability to run more than one thread
  - on the same core
  - at the same time
  - really, the same time, ie: in the same clock cycle!



































































# Processor terminology



# Multi-threading Evolution

## Single thread Out of Order

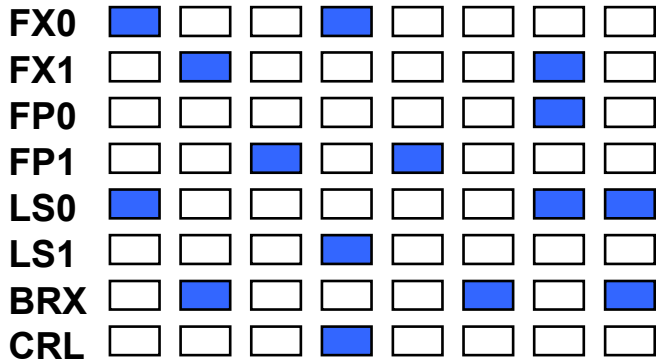
FX0								
FX1								
FP0								
FP1								
LS0								
LS1								
BRX								
CRL								

 No Thread Executing

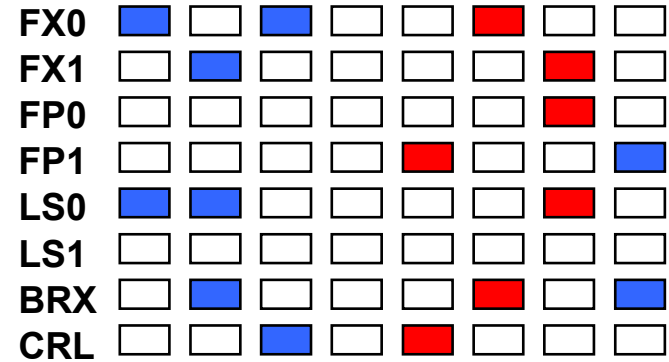
 Thread 0 Executing

# Multi-threading Evolution

## Single thread Out of Order



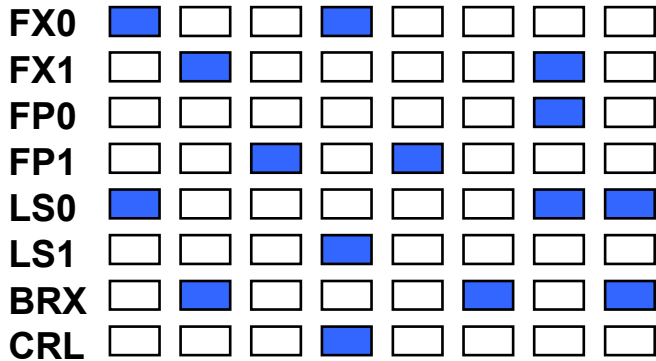
## S80 HW Multi-thread



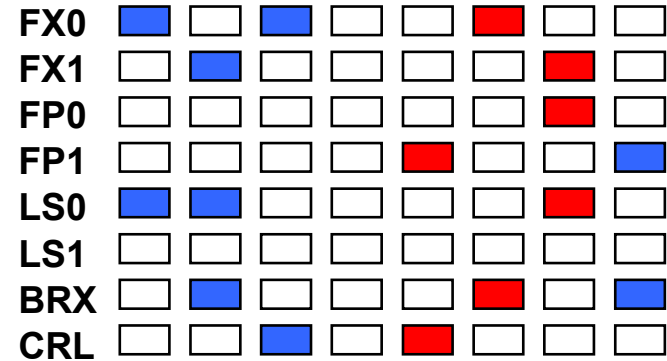
□ No Thread Executing    █ Thread 0 Executing    █ Thread 1 Executing

# Multi-threading Evolution

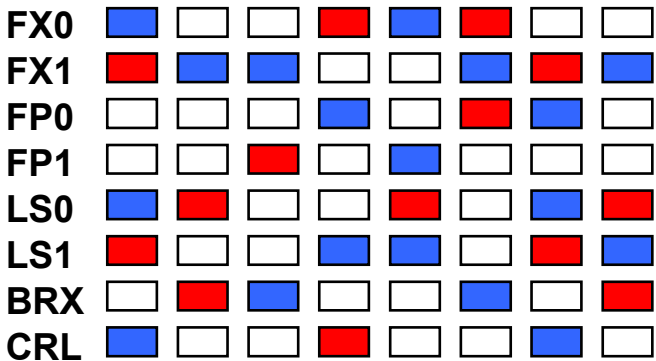
## Single thread Out of Order



## S80 HW Multi-thread



## POWER5 2 Way SMT

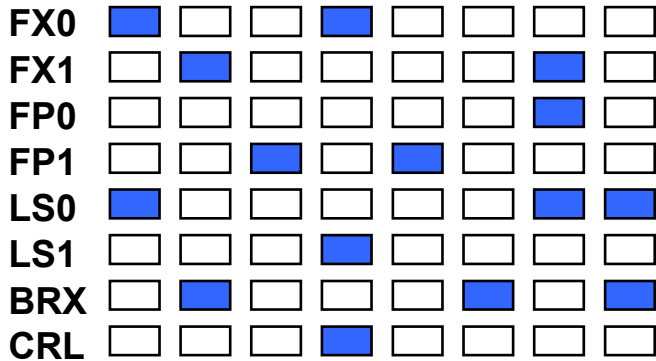


□ No Thread Executing    █ Thread 0 Executing    █ Thread 1 Executing

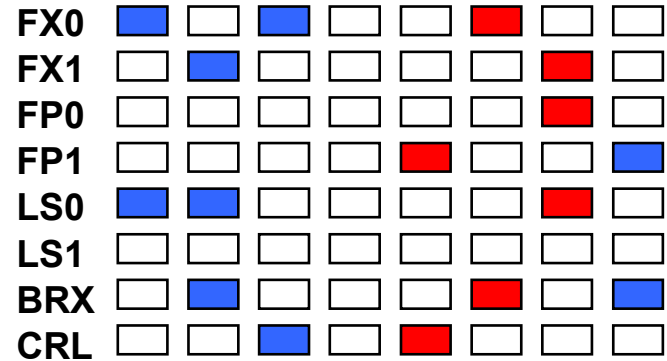


# Multi-threading Evolution

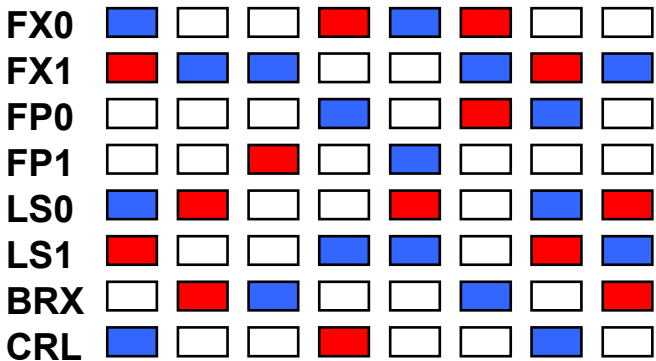
## Single thread Out of Order



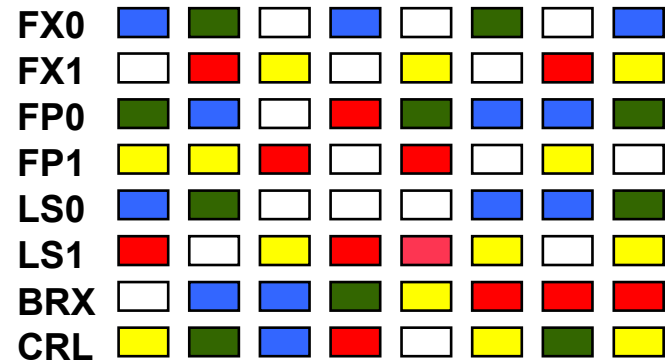
## S80 HW Multi-thread



## POWER5 2 Way SMT



## POWER7 4 Way SMT



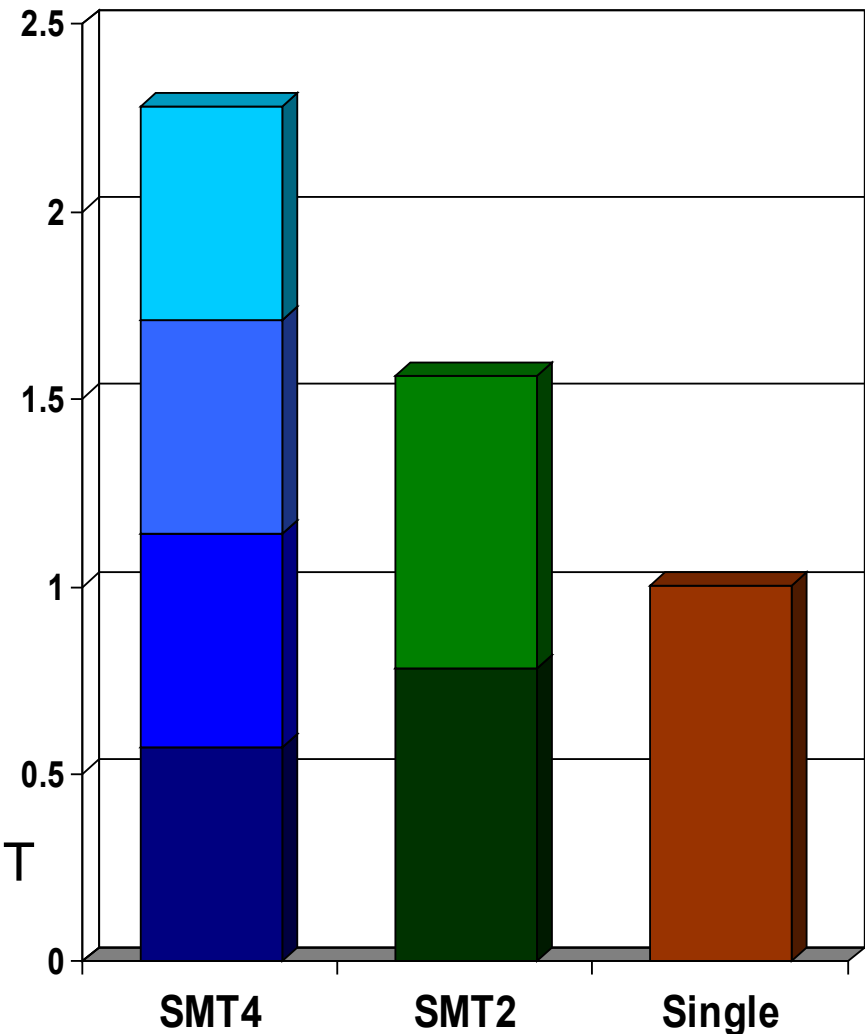
□ No Thread Executing    █ Thread 0 Executing    █ Thread 1 Executing  
 █ Thread 2 Executing    █ Thread 3 Executing

## SMT & Logical CPUs

- Historically, applications have used homogeneous systems
- In reality, different pieces of code have different needs of performance
- POWER7 offers different SMT modes to provide varying level of throughput and thread performance

SMT Modes are dynamically changeable and may be set by partition.

- OS may also dynamically change SMT levels based on thread availability



# Demonstration