19 września 2014

POWER8

POWER8 Processor

Technology

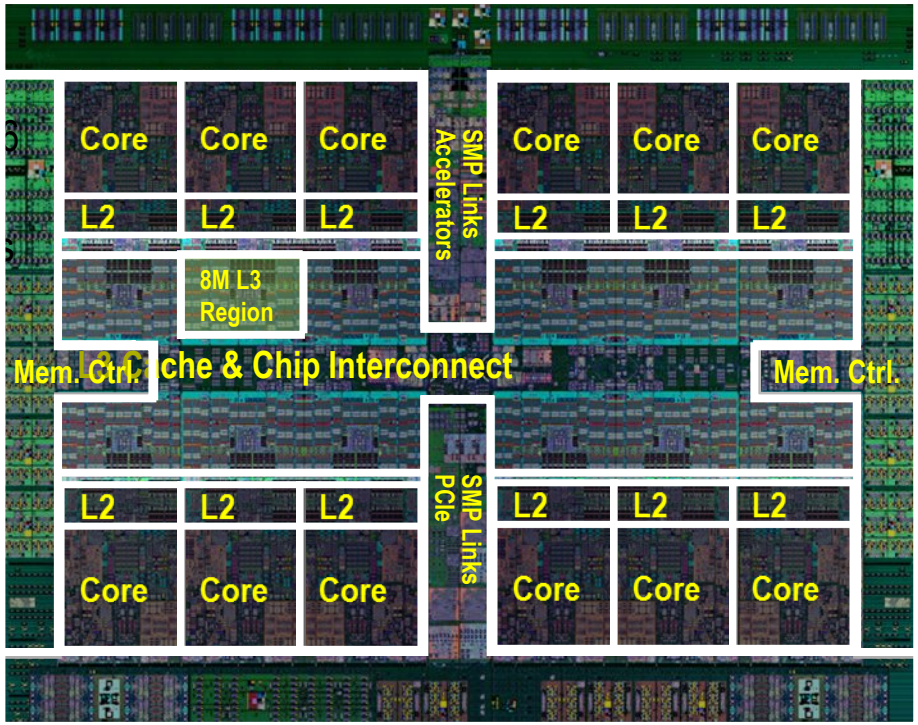
22nm SOI, eDRAM, 15 ML 650mm2

Cores

- 12 cores (SMT8)
- 8 dispatch, 10 issue, 16 exec pipe
- 2X internal data flows/queues
- Enhanced prefetching
- 64K data cache, 32K instruction cache

Accelerators

- Crypto & memory expansion
- Transactional Memory
- VMM assist
- Data Move / VM Mobility



Caches

- 512 KB SRAM L2 / core
- 96 MB eDRAM shared L3
- Up to 128 MB eDRAM L4 (off-chip)

Memory

- Up to 230 GB/s sustained bandwidth

Bus Interfaces

- Durable open memory attach interface
- Integrated PCIe Gen3
- SMP Interconnect
- CAPI (Coherent Accelerator Processor Interface)

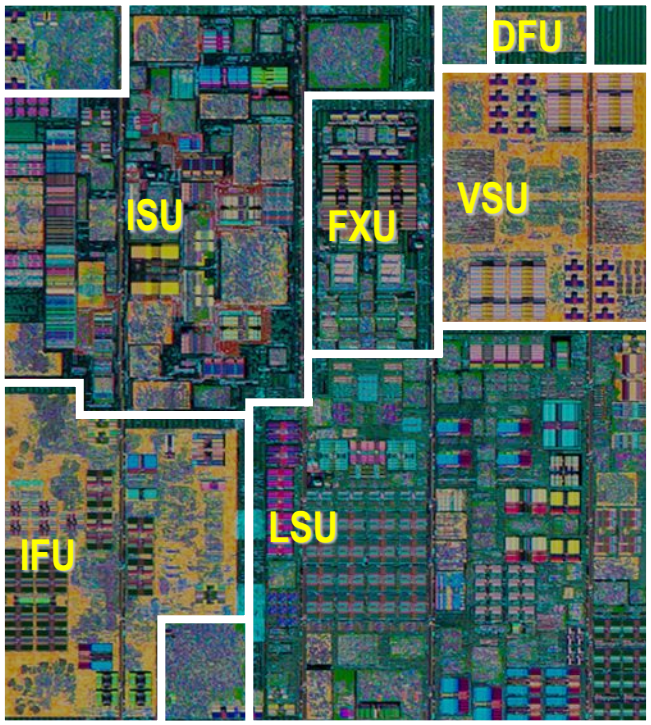
Energy Management

- On-chip Power Management Micro-controller
- Integrated Per-core VRM
- Critical Path Monitors

POWER8 Core

Execution Improvement vs. POWER7

- SMT4 → SMT8
- 8 dispatch
- 10 issue
- 16 execution pipes:
  - 2 FXU, 2 LSU, 2 LU, 4 FPU, 2 VMX, 1 Crypto, 1 DFU, 1 CR, 1 BR
- Larger Issue queues (4 x 16-entry)
- Larger global completion, Load/Store reorder
- Improved branch prediction
- Improved unaligned storage access



Core Performance vs . POWER7
~1.6x Thread
~2x Max SMT

Larger Caching Structures vs. POWER7

- 2x L1 data cache (64 KB)
- 2x outstanding data cache misses
- 4x translation Cache

Wider Load/Store

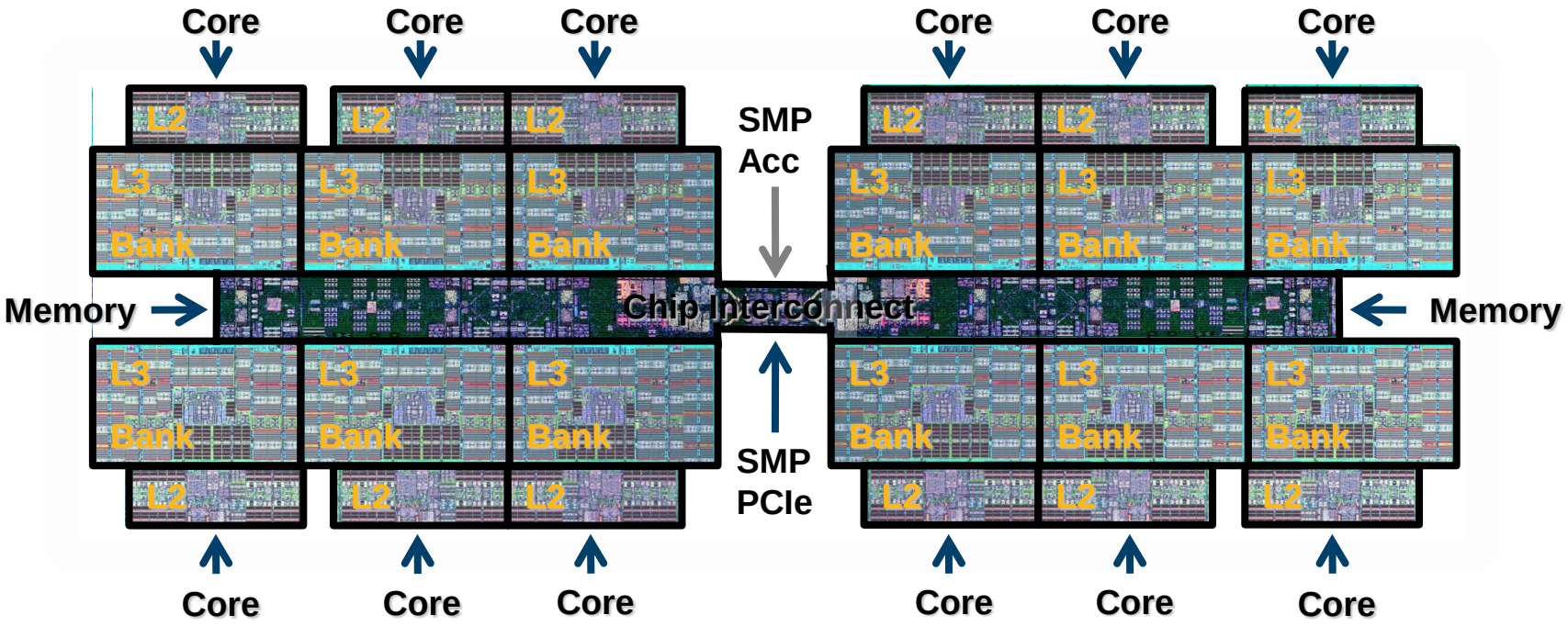
- 32B → 64B L2 to L1 data bus
- 2x data cache to execution dataflow

Enhanced Prefetch

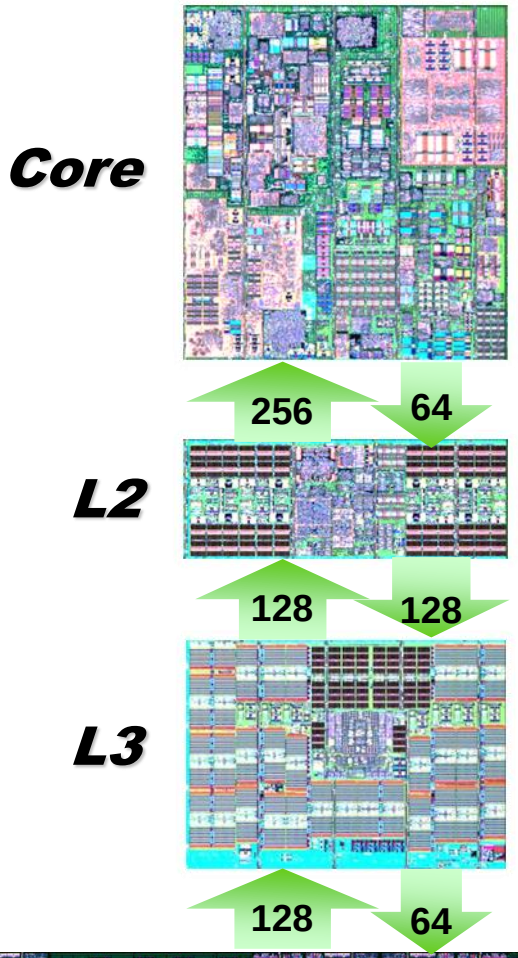
- Instruction speculation awareness
- Data prefetch depth awareness
- Adaptive bandwidth awareness
- Topology awareness

POWER8 On-chip Caches

- **L2: 512 KB 8 way per core**
- **L3: 96 MB (12 x 8 MB 8 way Bank)**
- **"NUCA" Cache policy (**Non-Uniform Cache Architecture)
  - Scalable bandwidth and latency
  - Migrate "hot" lines to local L2, then local L3 (replicate L2 contained footprint)
- **Chip Interconnect: 150 GB/sec x 12 segments per direction = 3.6 TB/sec**

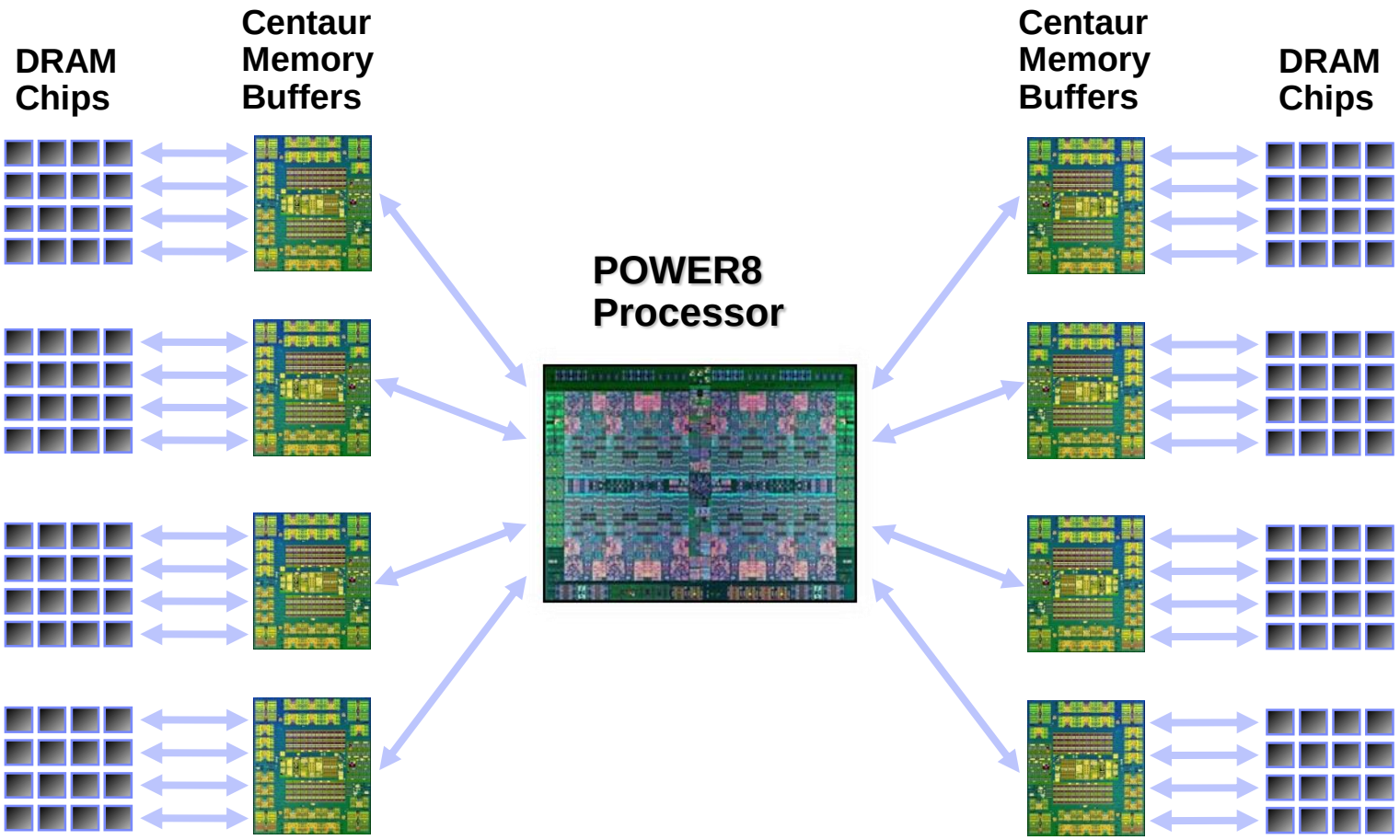


Cache Bandwidth



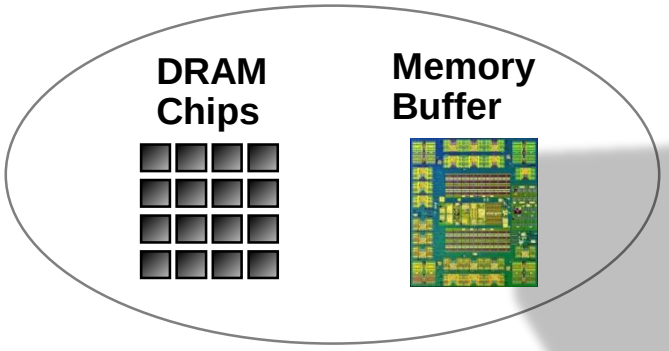
- **GB/sec shown assuming 4 GHz**
  - Product frequency will vary based on model type
- **Across 12 core chip**
  - 4 TB/sec L2 BW
  - 3 TB/sec L3 BW

Memory Organization



- ***Up to 8 high speed channels, each running up to 9.6 Gb/s for up to 230 GB/s sustained***
- ***Up to 32 total DDR ports yielding 410 GB/s peak at the DRAM***
- ***Up to 1 TB memory capacity per fully configured processor socket***

Memory Buffer Chip …*with 16MB Cache…*



Intelligence Moved into Memory

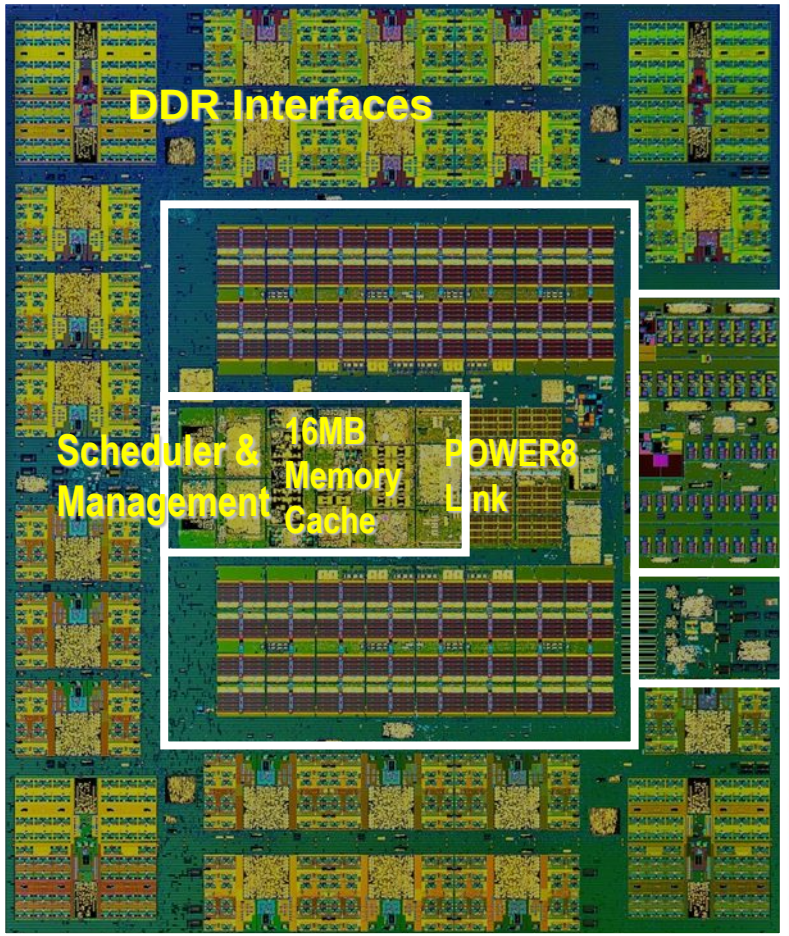
- Scheduling logic, caching structures
- Energy Mgmt, RAS decision point
  - Formerly on Processor
  - Moved to Memory Buffer

Processor Interface

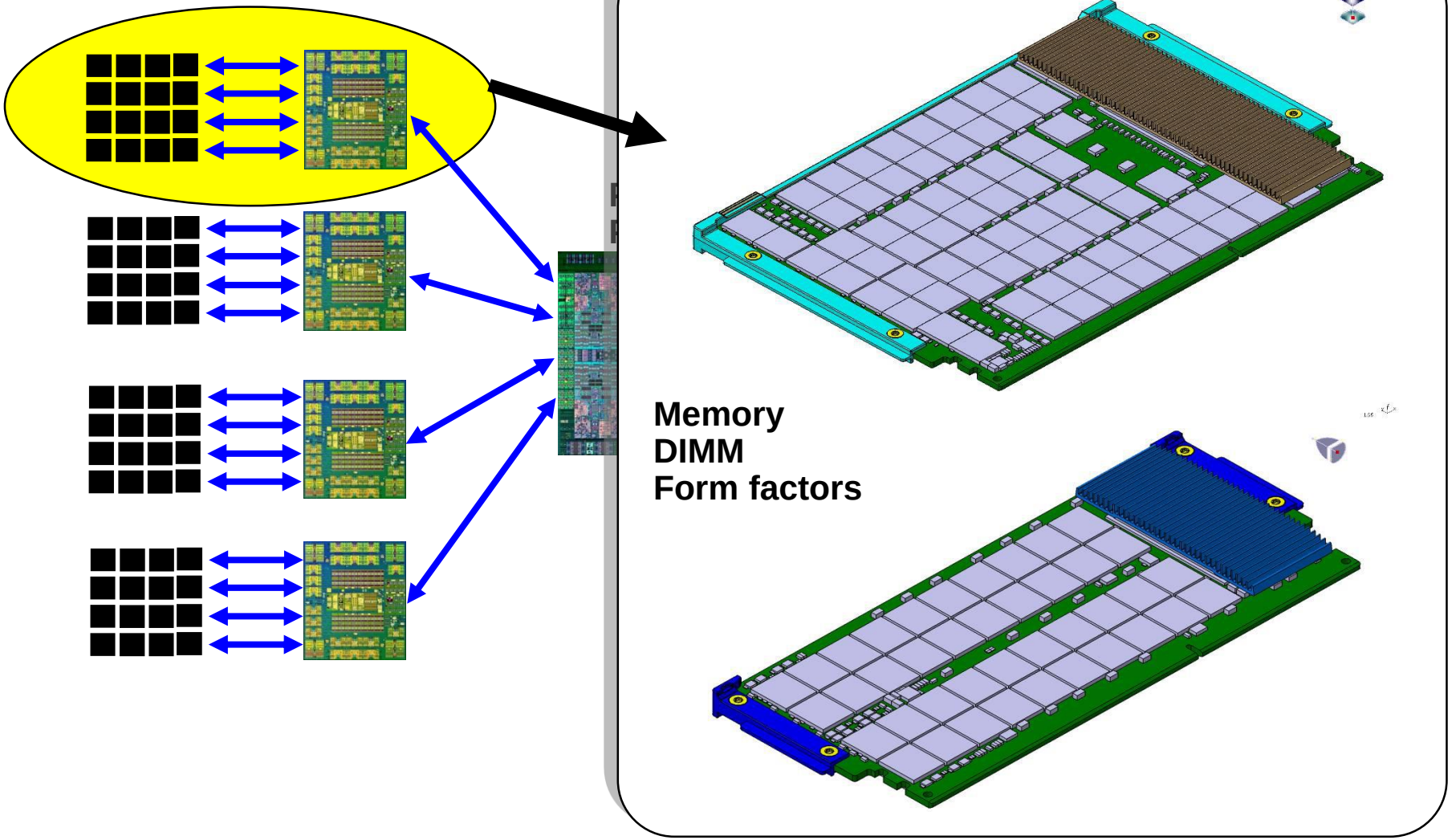
- 9.6 GB/s high speed interface
- More robust RAS
- "On-the-fly" lane isolation/repair
- Extensible for innovation build-out

Performance Value

- End-to-end fastpath and data retry (latency)
- Cache → latency/bandwidth, partial updates
- Cache → write scheduling, prefetch, energy
- 22nm SOI for optimal performance / energy
- 15 metal levels (latency, bandwidth)

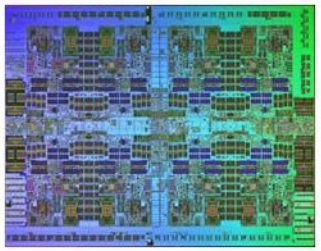


Centaur Memory DIMM



Integrated PCIe Gen3

POWER7



GX Bus

I/O Bridge

PCIe G2

PCI Device

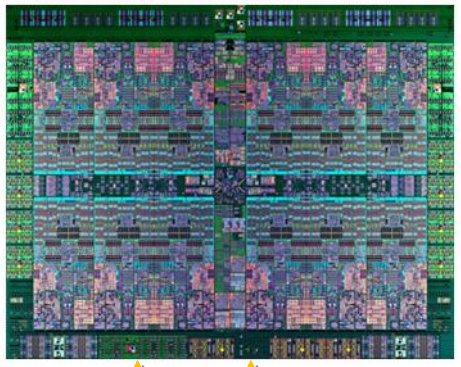
Native PCIe Gen 3 Support

- Direct processor integration
- Replaces proprietary GX/Bridge
- Low latency
- Gen3 x16 bandwidth (16 Gb/s)

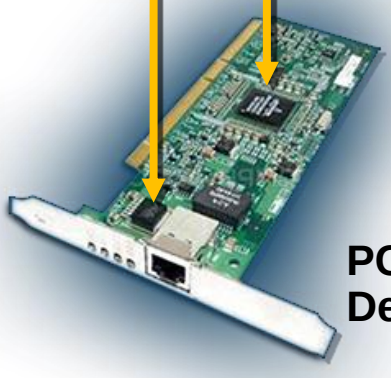
Transport Layer for CAPI Protocol

- Coherently Attach Devices connect to processor via PCIe
- Protocol encapsulated in PCIe

POWER8



PCIe G3

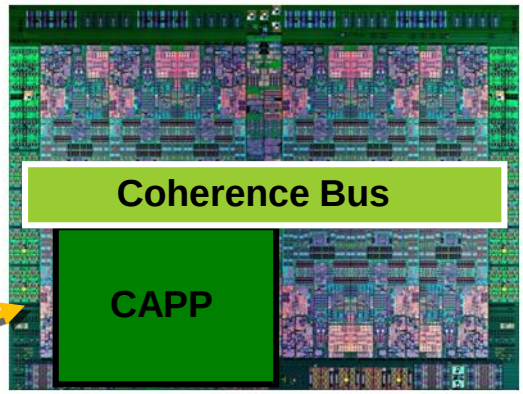


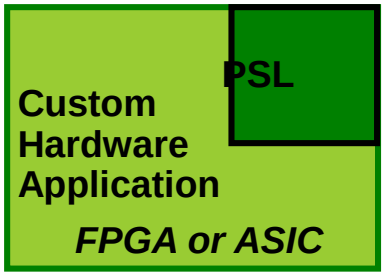
PCI Device

CAPI *(Coherent Accelerator Processor Interface)*

Virtual Addressing
•Accelerator can work with same memory addresses that the processors use
•Pointers de-referenced same as the host application
•Removes OS & device driver overhead

Hardware Managed Cache Coherence
•Enables the accelerator to participate in “Locks” as a normal thread Lowers Latency over IO communication model





PCIe Gen 3
Transport for encapsulated messages

Customizable Hardware Application Accelerator
•Specific system SW, middleware, or user application
•Written to durable interface provided by PSL

Processor Service Layer (PSL)
•Present robust, durable interfaces to applications
•Offload complexity / content from CAPP

Processor DCM Feature Codes

	S814	S824	S822	S812L	S822L
	8286-41A	8286-42A	8284-22A	8247-21L	8247-22L
Number Sockets	1S	2S (Opt 1S)	2S (Opt 1S)	1S	2S
Processor DCM	#EPX0 6-core 3.02 GHz	One or Two #EPXE 6-core 3.89 GHz	One or Two #EPX1 6-core 3.89 GHz	One #ELPD 10-core 3.42 GHz	Two #ELPD 10-core 3.42 GHz
Processor DCM	#EPX6 8-core 3.72 GHz	One or Two #EPXF 8-core 4.15 GHz	One or Two #EPXD 10-core 3.42 GHz	One #ELP3 12-core 3.02 GHz	Two #ELP4 12-core 3.02 GHz
Processor DCM		Two #EPXH 12-core 3.52 GHz			

Power System Roadmap

2013 **2014**

Enhanced HMC Management

KVM

POWER8

PowerVC

PowerVP

4U Systems: S814 & S824

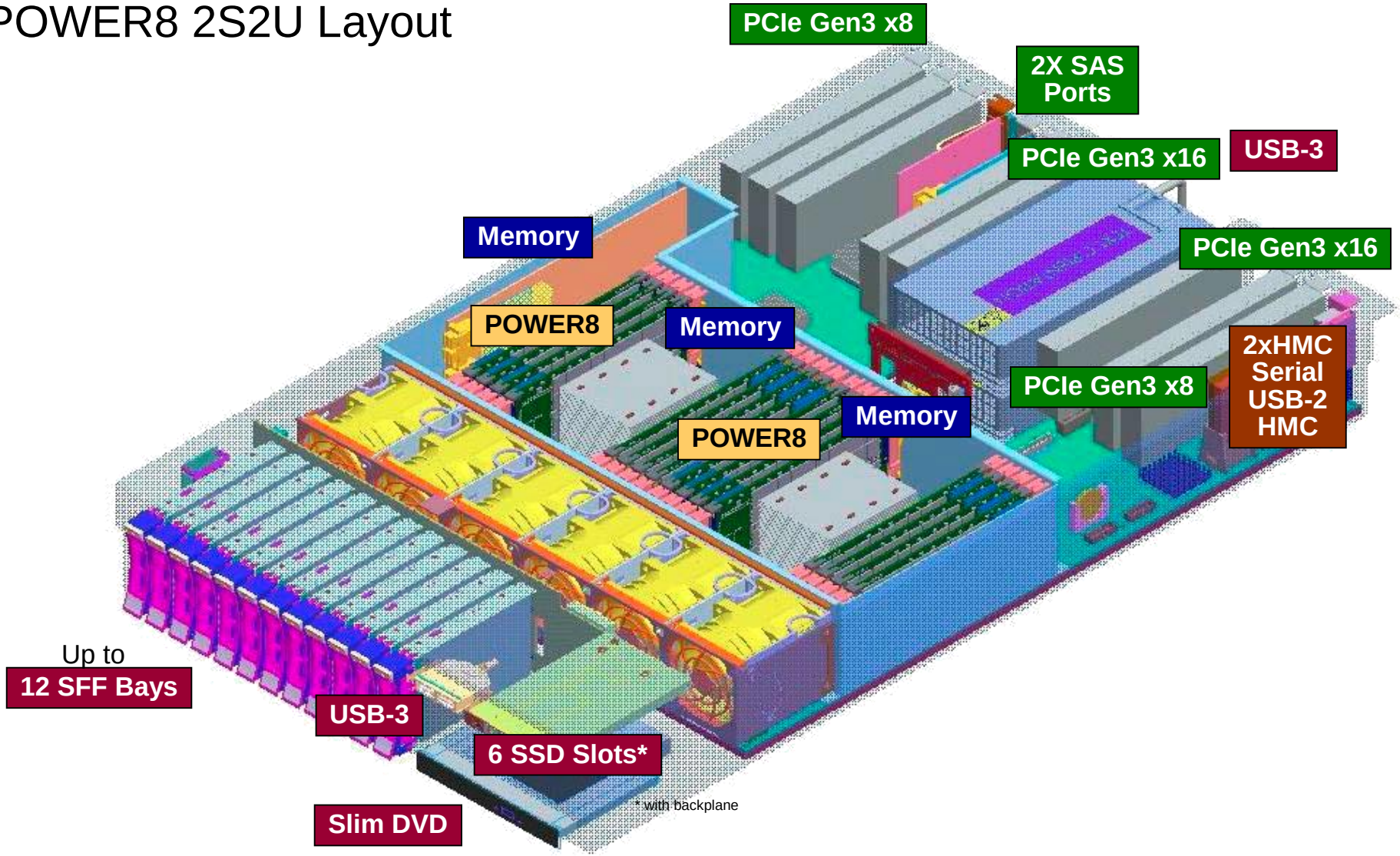
Power Linux Engines

2U System: S822

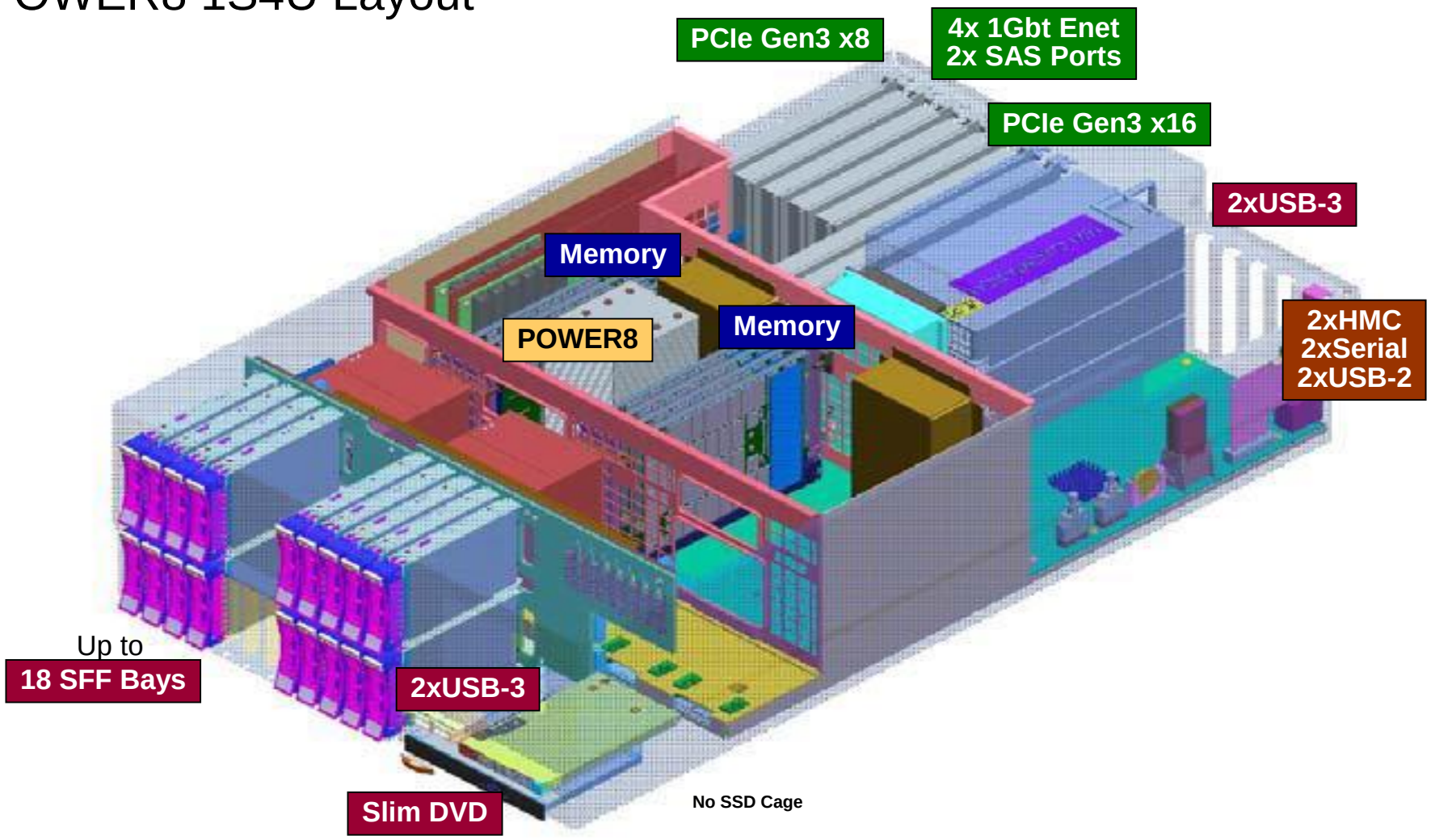
Enterprise Pools

2U Linux: S822L

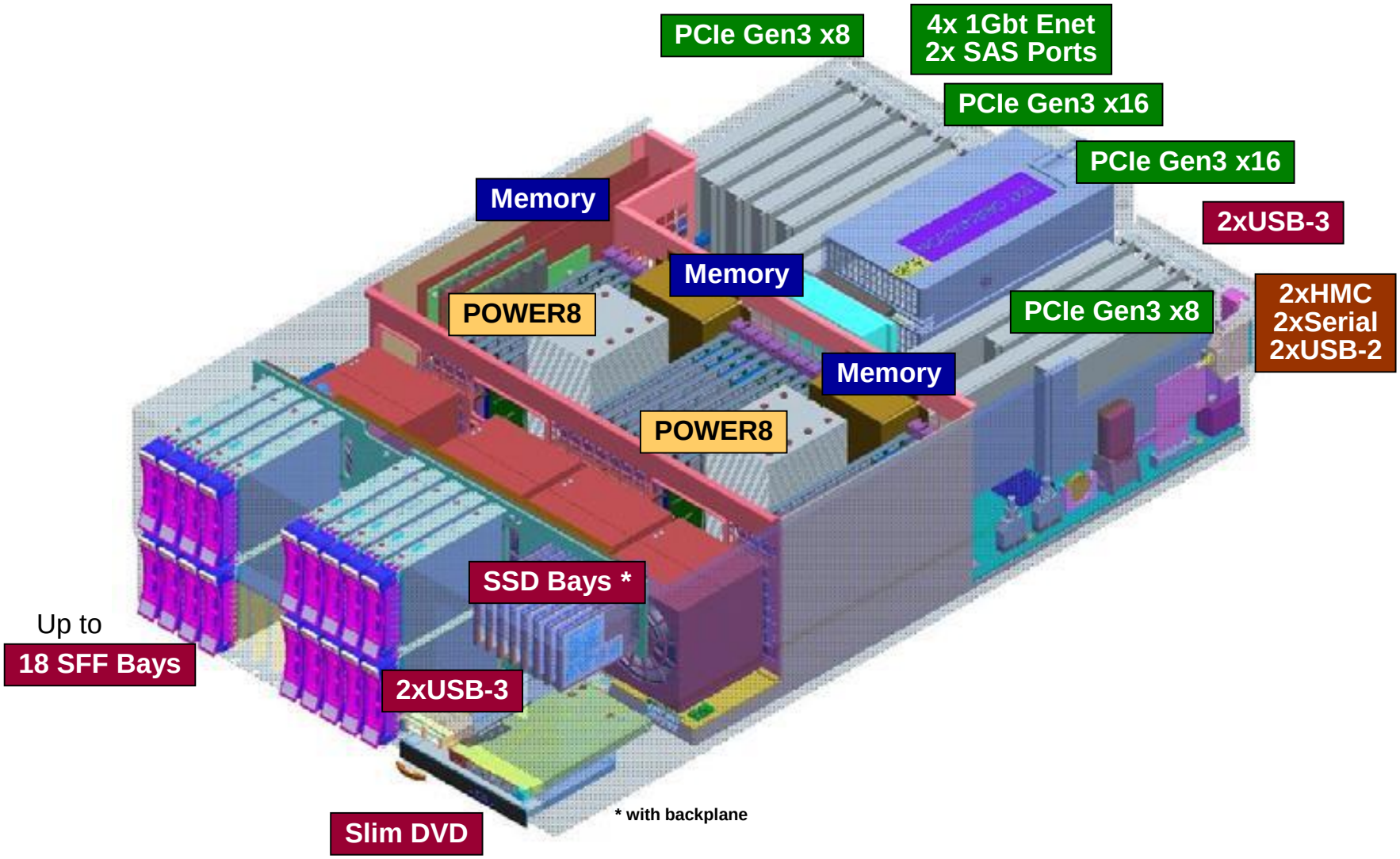
POWER8 2S2U Layout



POWER8 1S4U Layout



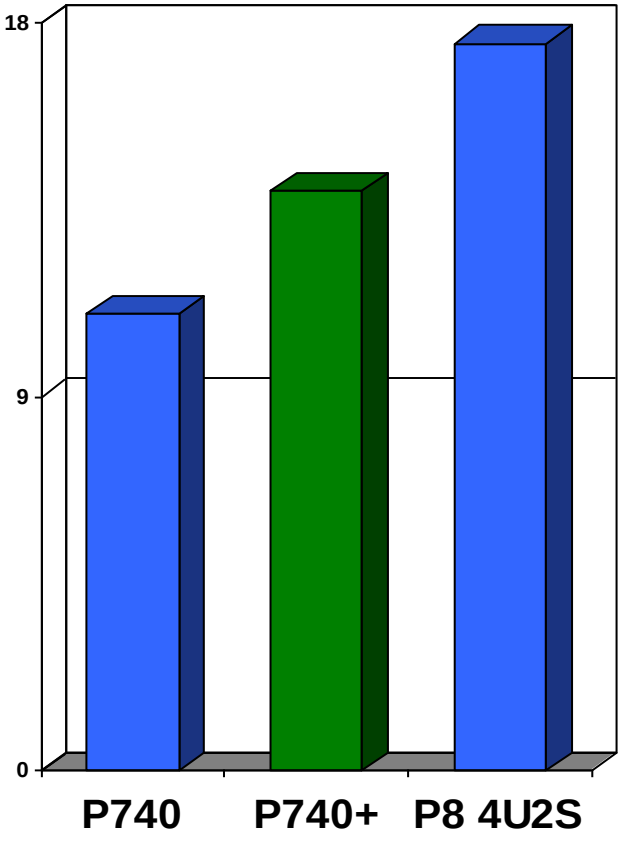
POWER8 2S4U Layout



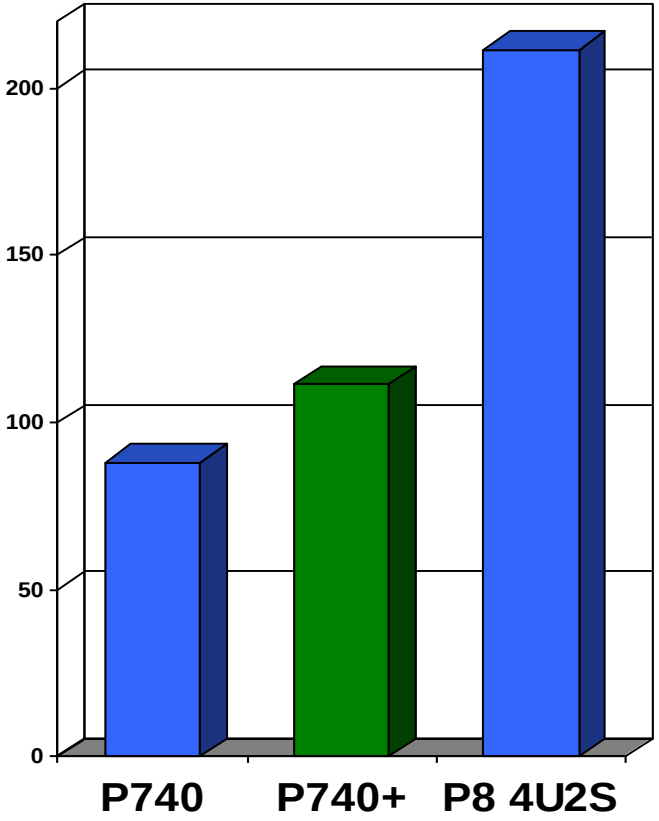
Wydajność

Power 740+ / POWER8 S824 rPerf Comparisons

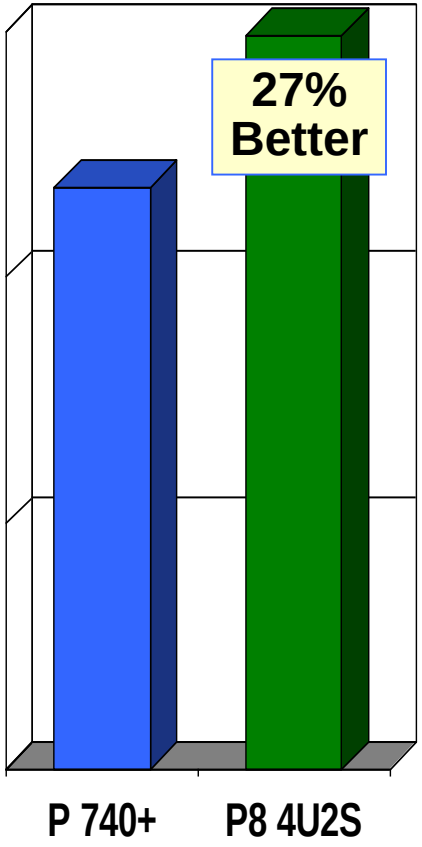
Performance per Core



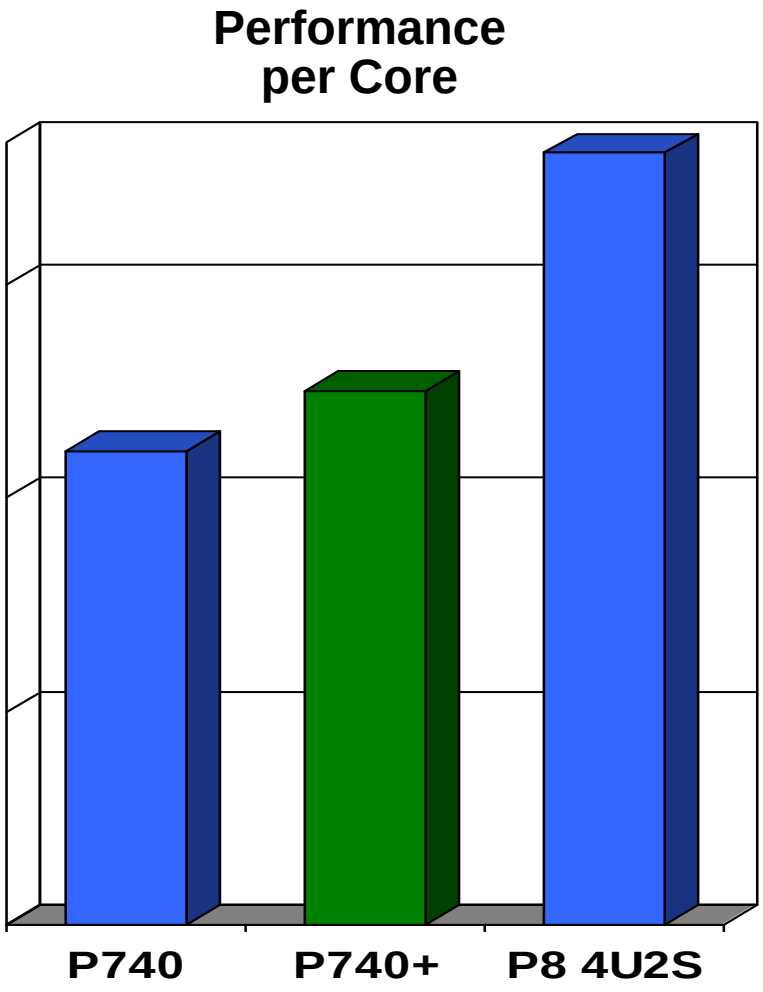
Performance per Socket

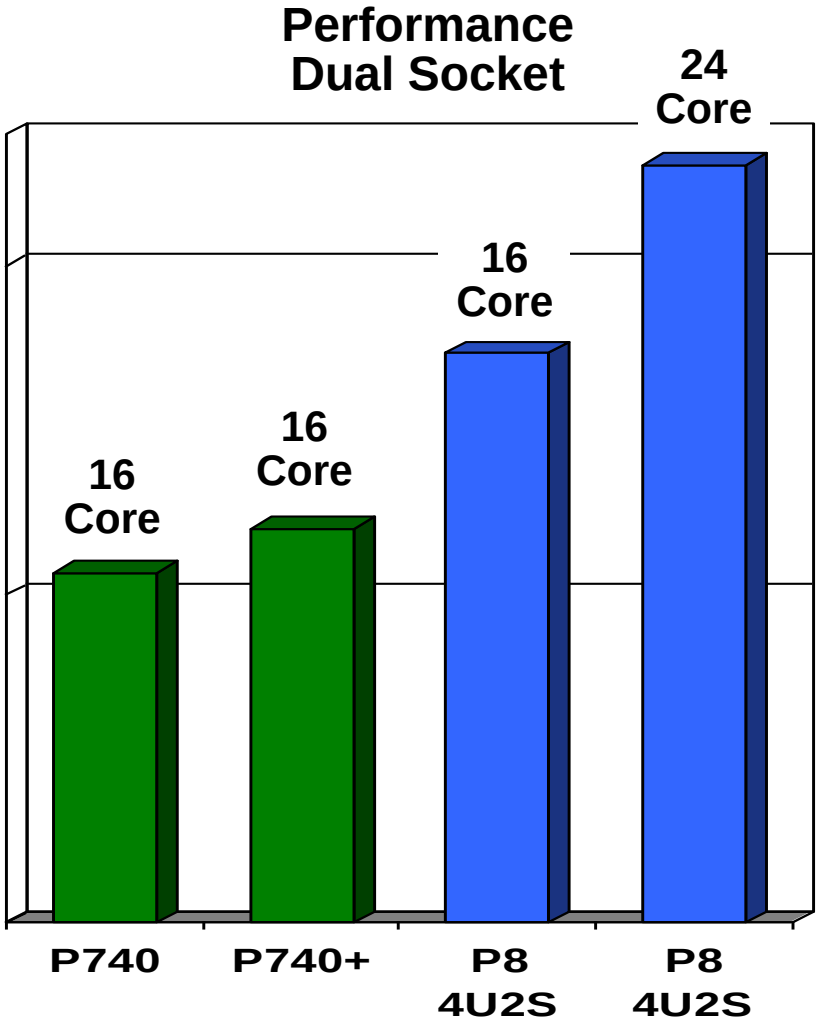


Performance per KW



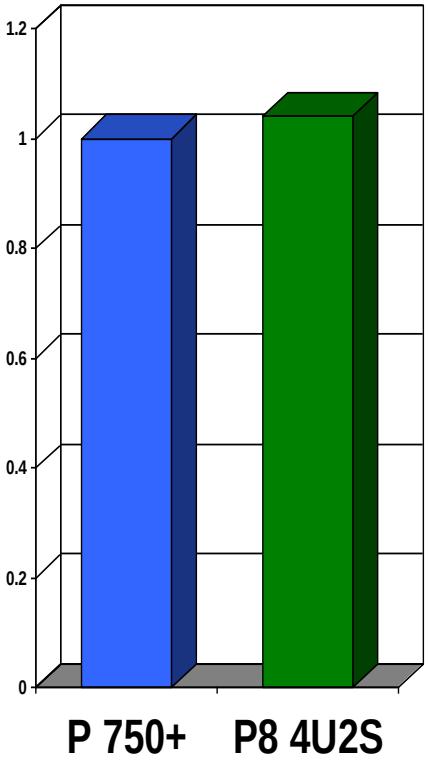
Power 740 / POWER8 S824 CPW Comparisons





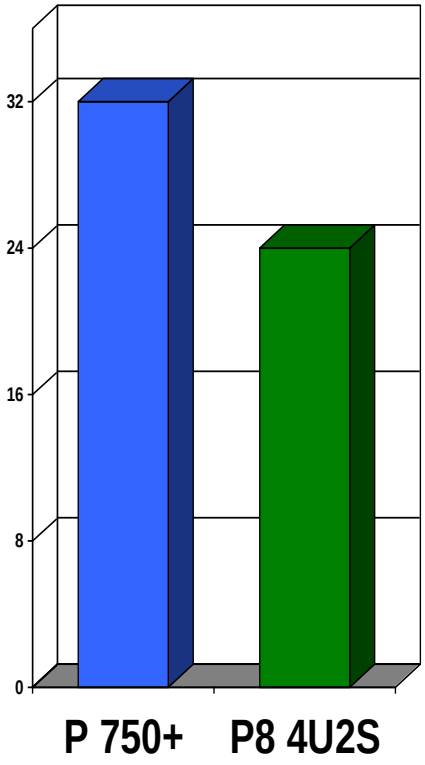
Power 750+ vs POWER8 4U2S

Performance



POWER8 ≥ Performance

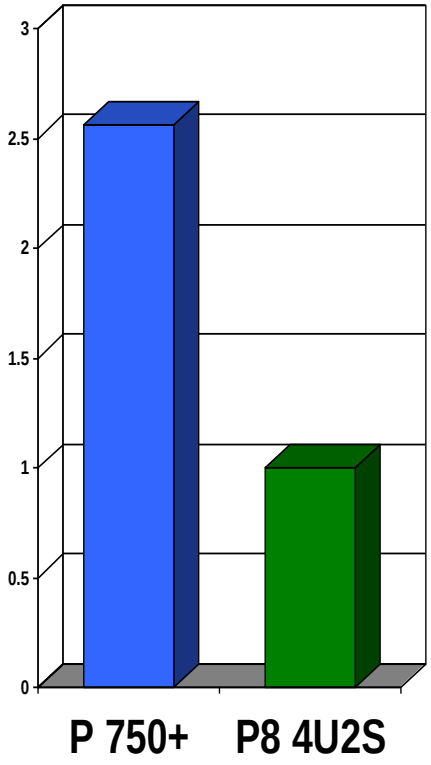
Total # Cores



Fewer cores means Lower Software Costs

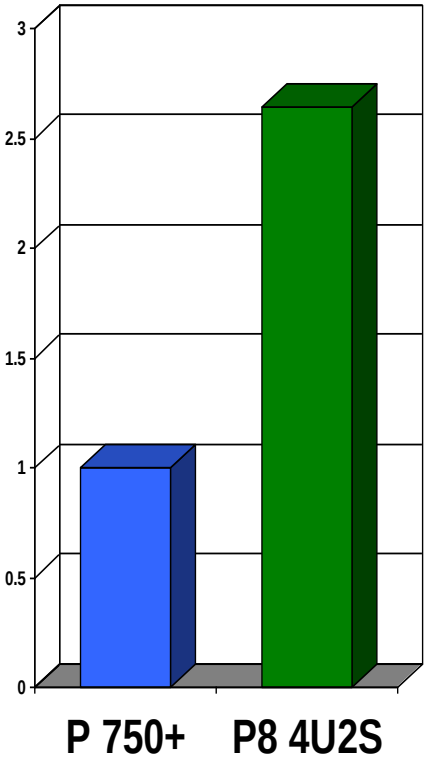
Does not reflect lower PVU savings

TCA



Lower investment costs

Performance per $



Better performance per dollar spent

POWER8 Processor Is Purpose Built Which Results in Superior Performance

	Sandy Bridge EP	Ivy Bridge EP E5-26xx v2	Ivy Bridge EX E7-88xx v2	POWER 7+ Systems	POWER8
Clock rates	1.8–3.6GHz	1.7-3.7GHz	1.9-3.4 GHz	3.1–4.4 GHz	3.0-4.1 GHz
SMT options	1,2*	1, 2*	1, 2*	1, 2, 4	1, 2, 4, 8
Max Threads / sock	16	24	30	32	96
Max L1 Data Cache	32KB	32KB*	32KB*	32KB	64KB
Max L2 Cache	256 KB	256 KB	256 KB	256 KB	512 KB
Max L3 Cache	20 MB	30 MB	37.5 MB	80 MB	96 MB
Max L4 Cache	0	0	0	0	128 MB
Memory Bandwidth	31.4-51.2 GB/s	42.6-59.7 GB/s	68-85** GB/s	100 – 180 GB/sec	230 - 410 GB/sec

* Intel calls this Hyper-Threading Technology (No HT and with HT)

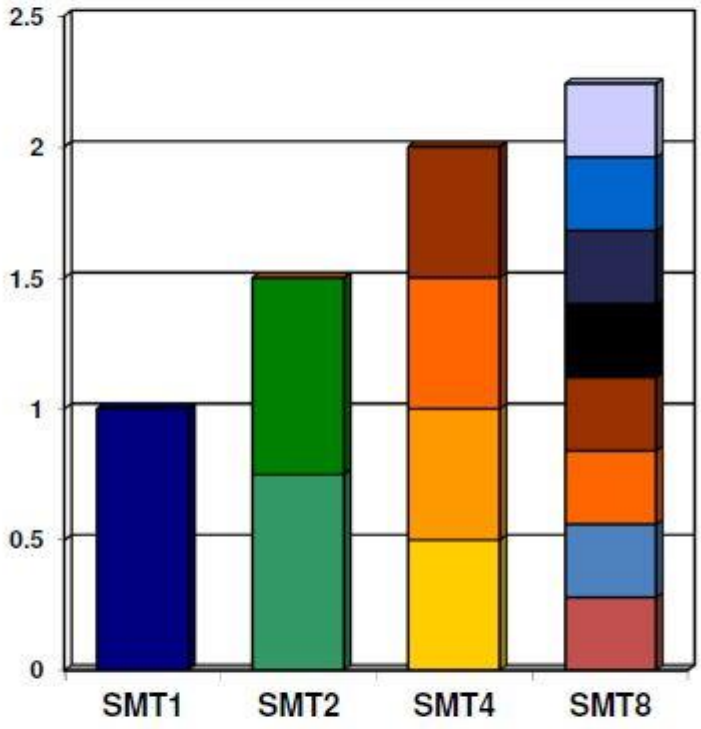
*32KB running in "Non-RAS mode" Only 16KB in RAS mode

**85GB running in "Non-RAS mode" = dual-device error NOT supported

Simultaneous Multi-Threading Can Be a Major Performance Enhancement

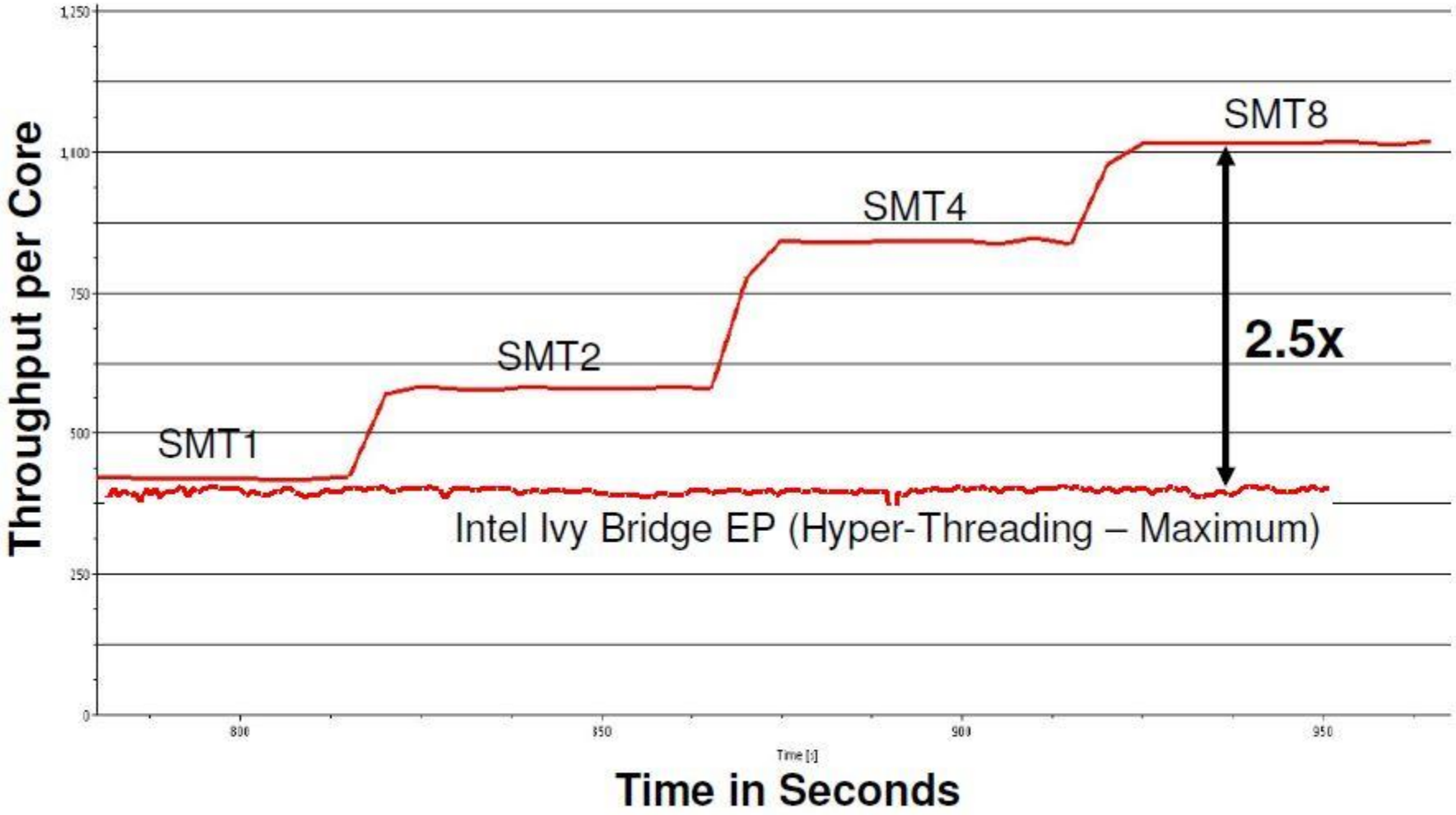
SMT allows separate instruction streams, or threads, to run concurrently on the same physical processor, or core

- **SMT1:** Largest unit of execution work
- **SMT2 / SMT4:** Smaller units of work, but provides greater amount of execution work per cycle
- **SMT8:** Smallest unit of work, but provides the maximum amount of execution work per cycle
- Can dynamical shift between modes as required: SMT1 / SMT2 / SMT4 / SMT8
- Mixed SMT modes supported within same LPAR



SMT = Simultaneous Multi-Threading

DEMO: POWER8 SMT Delivers More Performance Per Core Than Intel Ivy Bridge



Power's Basic RAS Features are Better Than x86

RAS Feature	POWER7/7+	POWER8 (PowerVM)	x86 Ivy Bridge-EX
Application/Partition			
Move running workload partitions to another server	Yes	Yes	Yes
Partition Availability priority	Yes	Yes	No
System			
First Failure Data Capture	Yes	Yes	No
Memory Keys ensure secure hardware separation of memory	Yes	Yes	No
Processor			
Dedicated Service Element processor	Yes	Yes	No
Processor Instruction Retry	Yes	Yes	No
Alternate Processor Recovery	Yes	Yes	No
Predictive Processor Deallocation	Yes	Yes	No (1)
Dynamic Processor Sparing with Capacity on Demand (CoD)	Yes	No(3)	No (1)
Memory			
Chipkill - Redundant Memory Capabilities	Yes	Yes	Yes
Chipkill - Survives Double Memory Failures with x8 DRAMS	Yes(2)	Yes(2)	Yes (4)
I/O			
VIOS - Dynamic adapter failover	Yes	Yes	No

1 = Capability built in but not supported
2 = Chipkill on Systems with DRAM sparing
3 = CoD supported on Enterprise Servers (not the 1 & 2 Socket)
4= Performance degradation and only x4 DRAM

POWER8 2U Storage Backplane Options

Base: 12 SFF HDD/SSD Bays
- **RAID 0,5,6,10 / JBoD**

Optional: Split Disk Backplane = 6 + 6

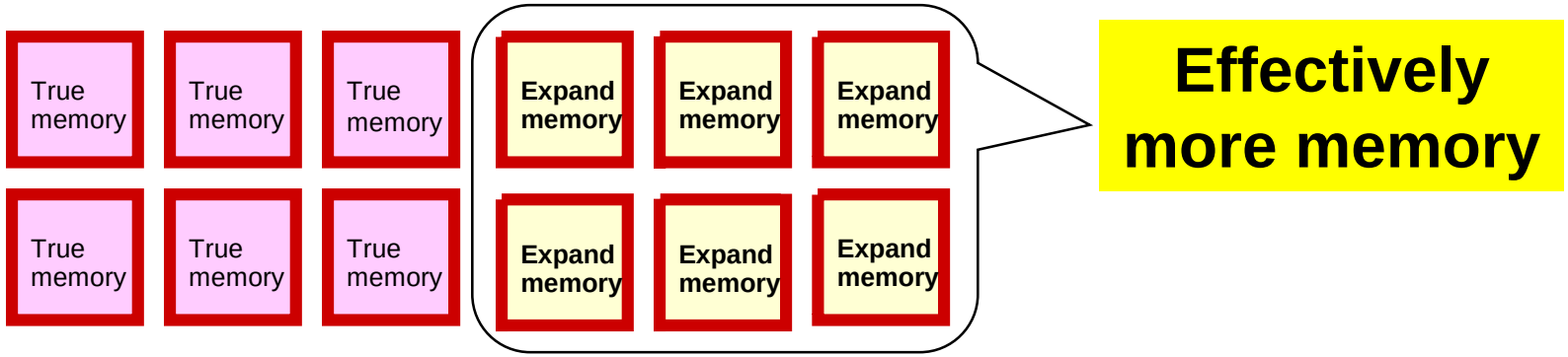
Optional: High Performance RAID 0,5,6,10,5T2,6T2,10T2
- **8 SFF Bays (HDD / SSD) and 6 SSD**

POWER8 4U Front View

Base RAID 0,5,6,10 Feature: 12 SFF HDD/SSD disks
Split disk 6+6 feature (optional)

High Performance RAID 0,5,6,10,5T2,6T2,10T2 Feature (optional)

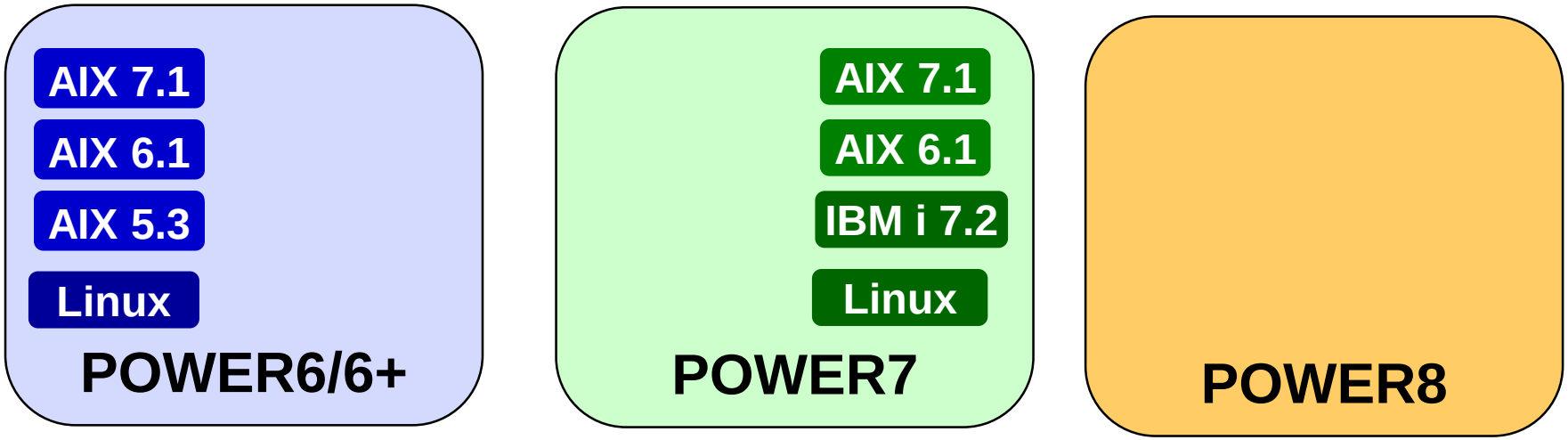
POWER8 Active Memory Expansion



- Like POWER7, provides POWER8 advantage
- Expand memory beyond physical limits
- More effective server consolidation
  - Run more application workload / users per partition
  - Run more partitions and more workload per server
- 60-day trial like Power 7xx
- AIX only

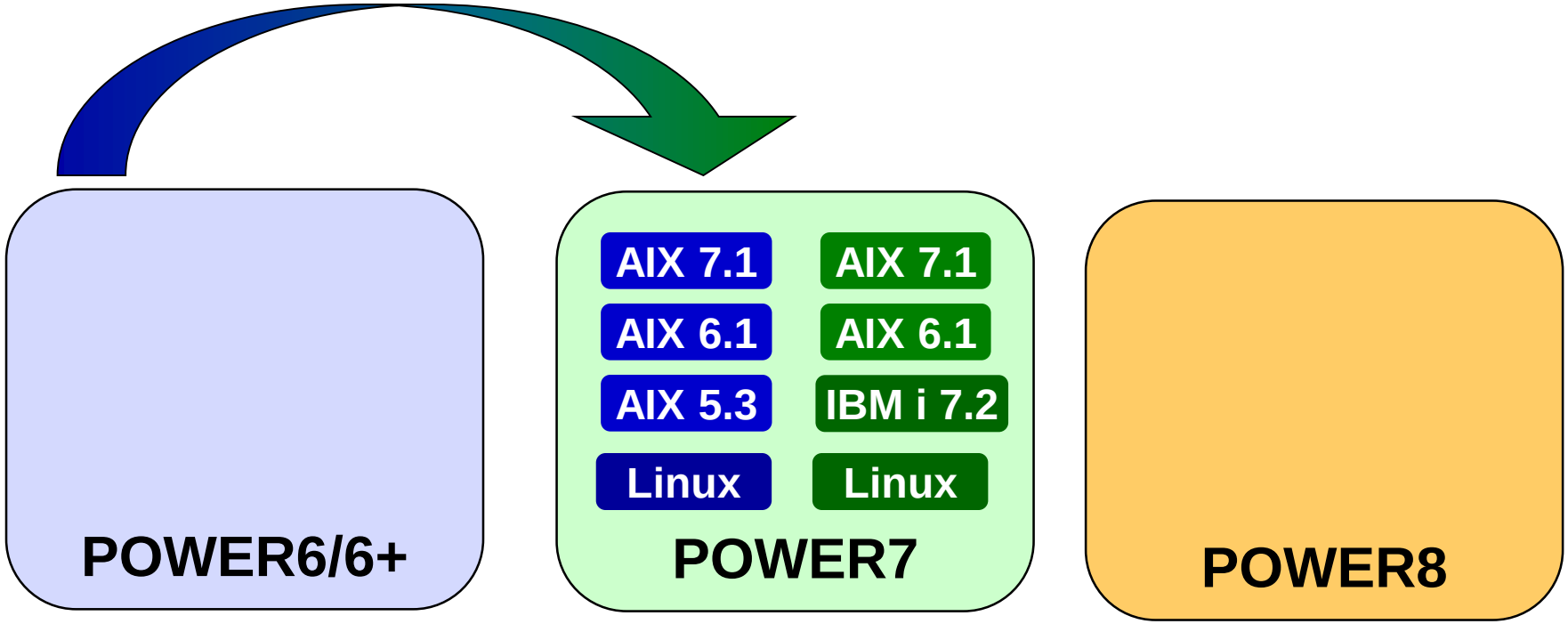
Note expansion percentage is very workload dependent

POWER6 / POWER7 / POWER8 Partition Mobility



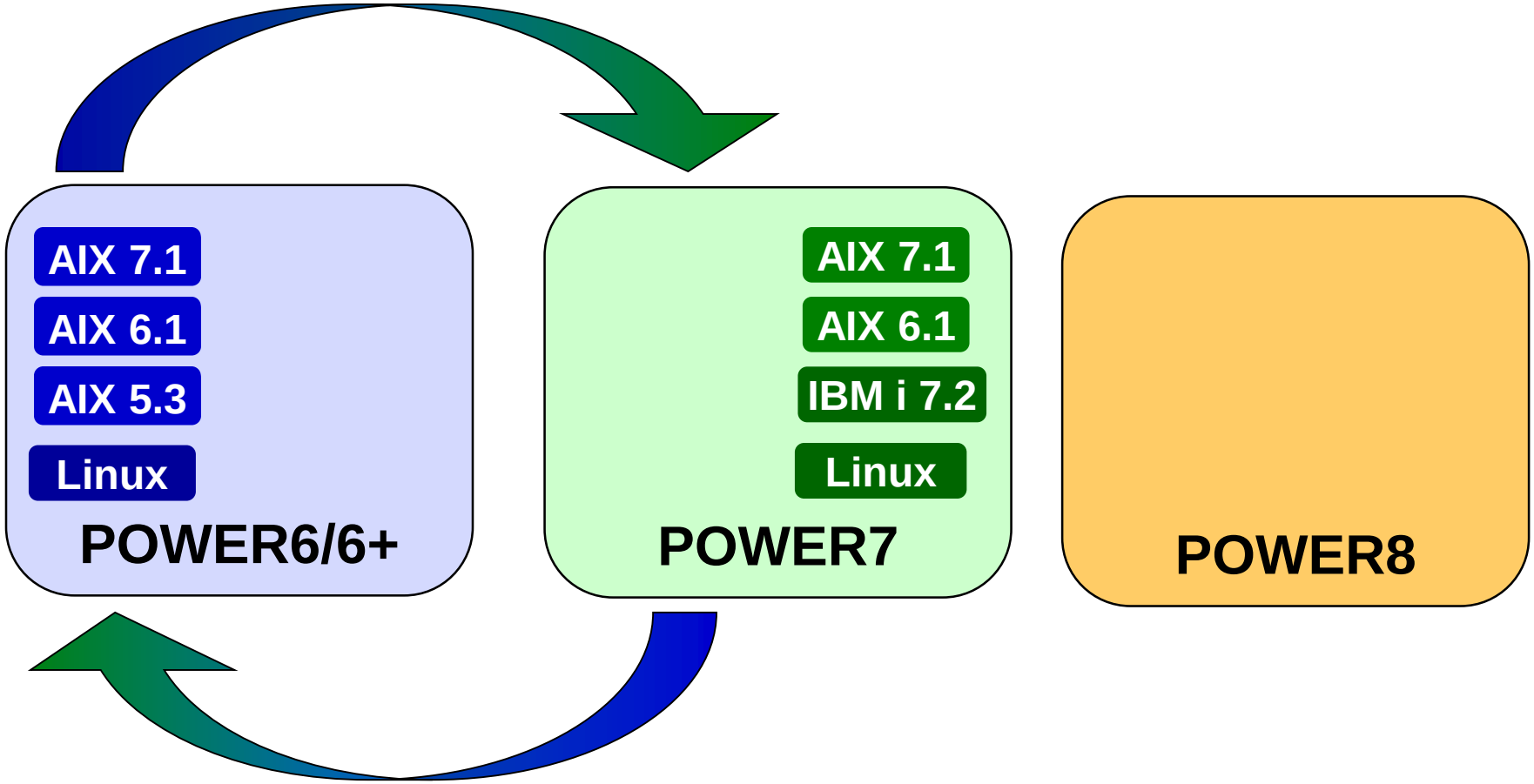
Leverage POWER6 / POWER7 Compatibility Modes
LPAR Migrate between POWER6 / POWER7 / POWER8 Servers
Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER6 / POWER7 / POWER8 Partition Mobility



Leverage POWER6 / POWER7 Compatibility Modes
LPAR Migrate between POWER6 / POWER7 / POWER8 Servers
Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER6 / POWER7 / POWER8 Partition Mobility

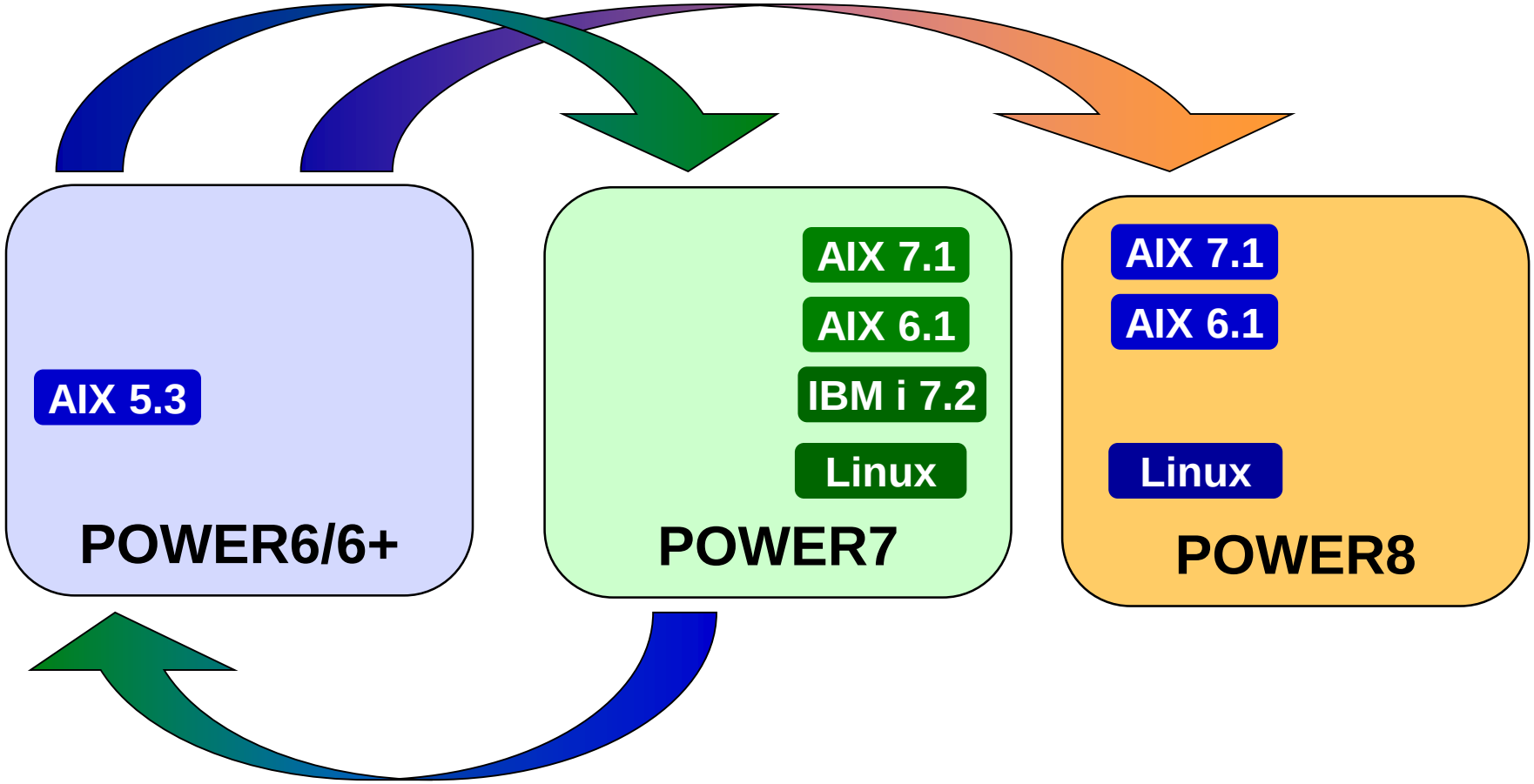


Leverage POWER6 / POWER7 Compatibility Modes

LPAR Migrate between POWER6 / POWER7 / POWER8 Servers

Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER6 / POWER7 / POWER8 Partition Mobility

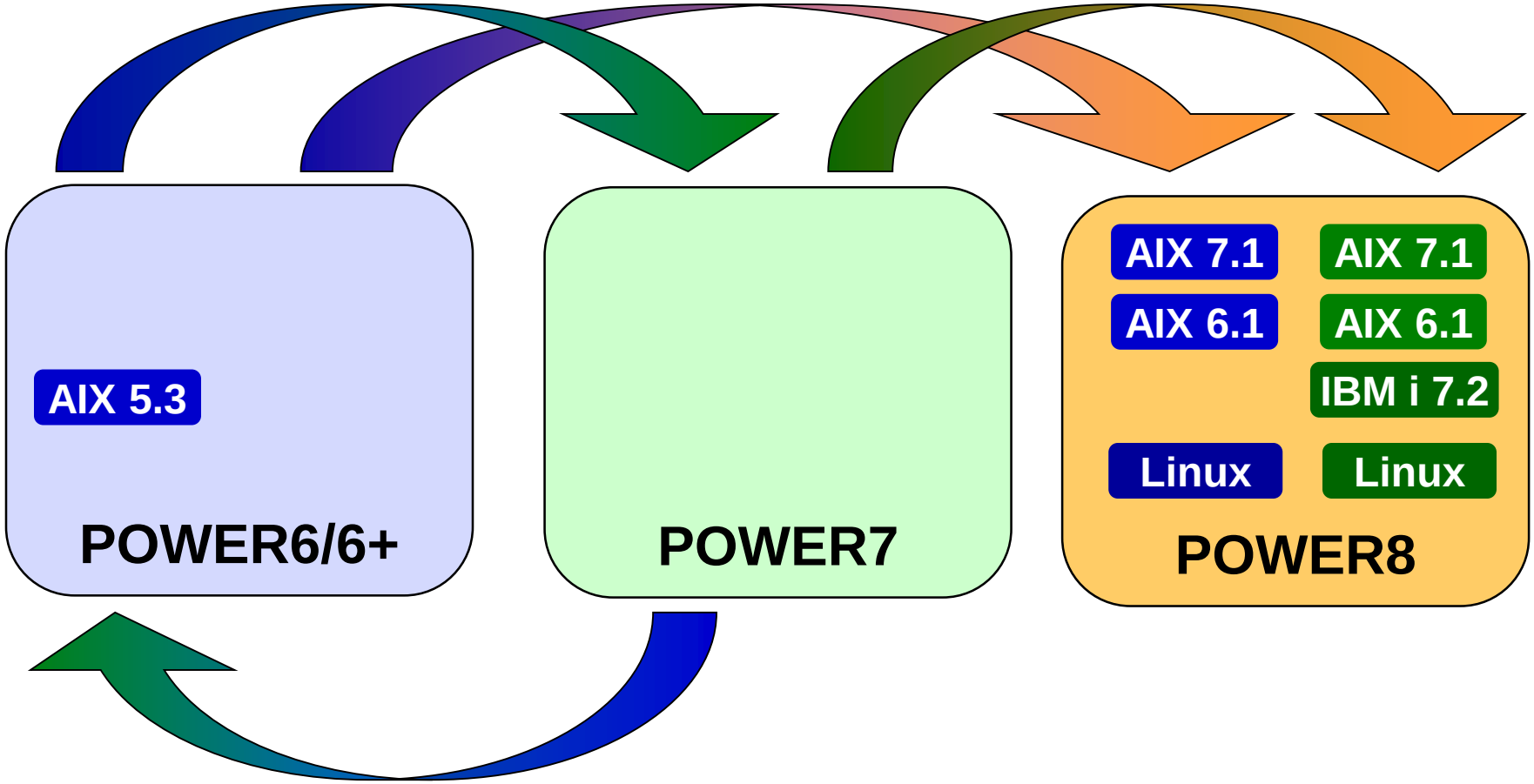


Leverage POWER6 / POWER7 Compatibility Modes

LPAR Migrate between POWER6 / POWER7 / POWER8 Servers

Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER6 / POWER7 / POWER8 Partition Mobility

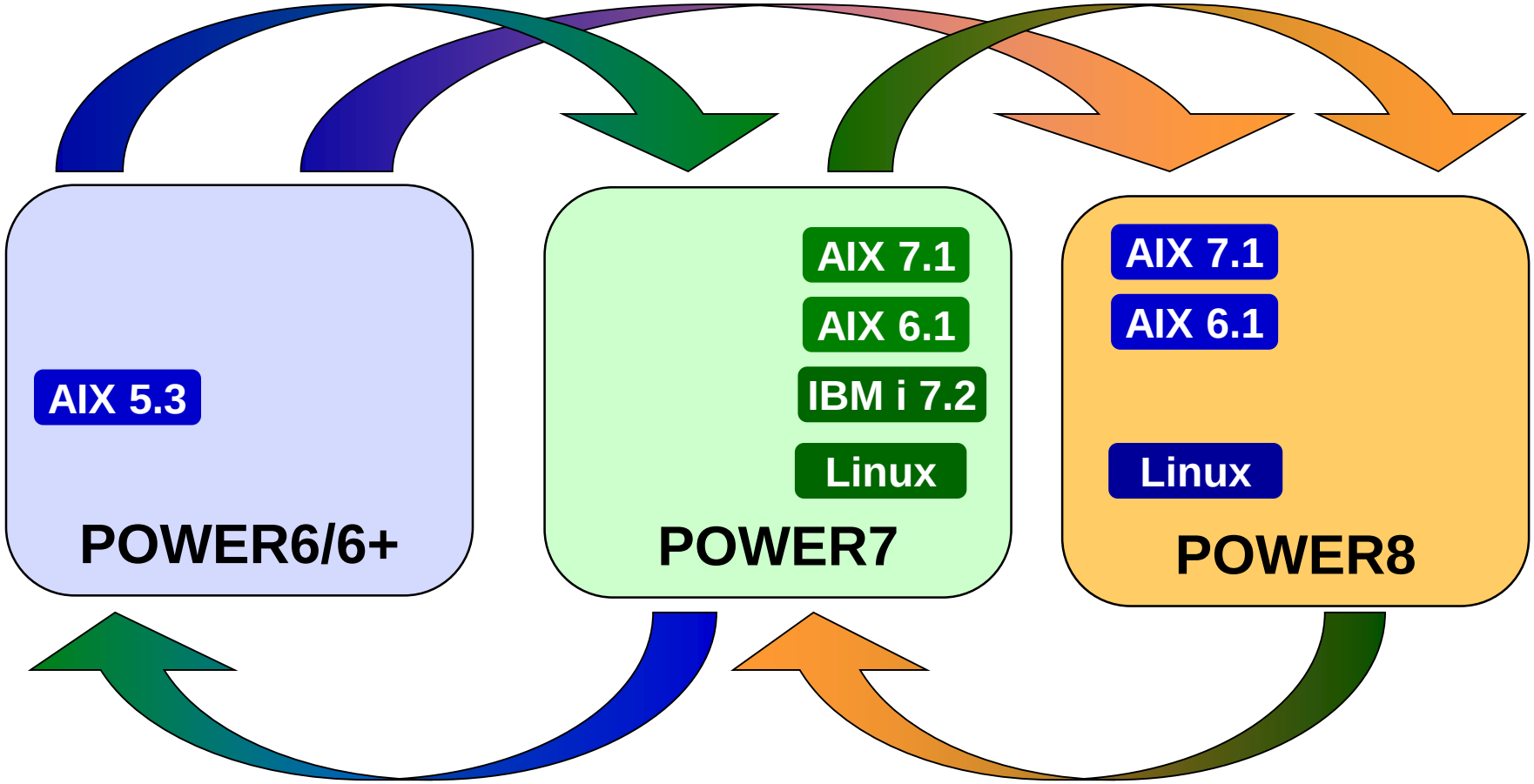


Leverage POWER6 / POWER7 Compatibility Modes

LPAR Migrate between POWER6 / POWER7 / POWER8 Servers

Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER6 / POWER7 / POWER8 Partition Mobility

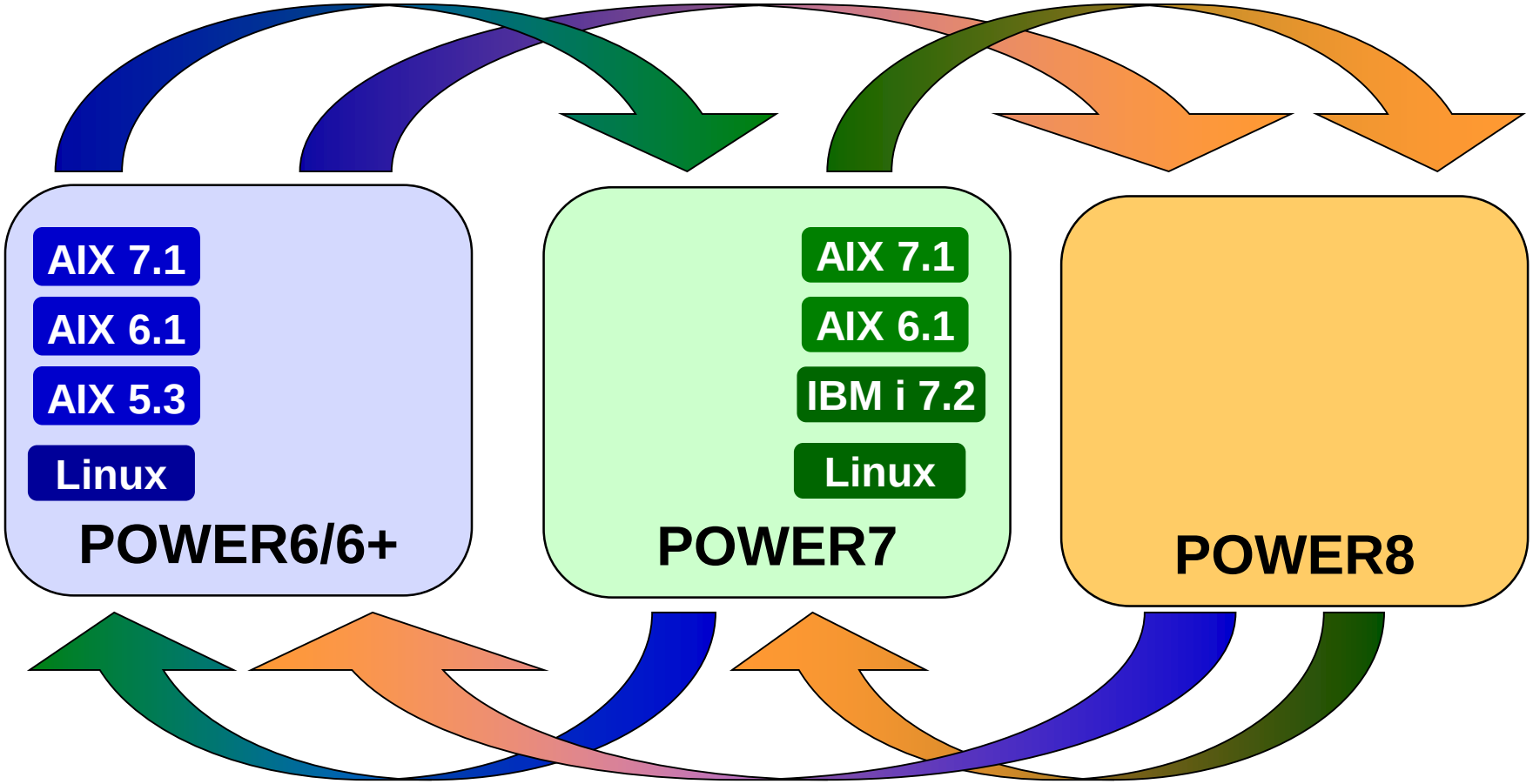


Leverage POWER6 / POWER7 Compatibility Modes

LPAR Migrate between POWER6 / POWER7 / POWER8 Servers

Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER6 / POWER7 / POWER8 Partition Mobility



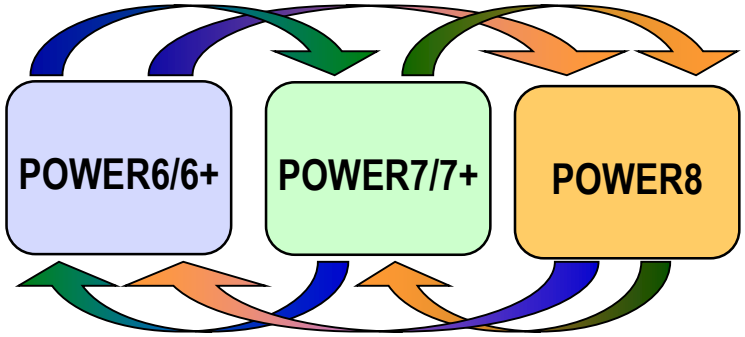
Leverage POWER6 / POWER7 Compatibility Modes

LPAR Migrate between POWER6 / POWER7 / POWER8 Servers

Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

POWER8 LPM

- Can Utilize 10 Gbt Ethernet
  - Minimize movement time
  - Move more LPARs concurrently
- Used for Migrations
  - POWER6 to POWER8
  - POWER7 to POWER8
- Used for Maintenance Activities…
  - Primary resource for concurrent maintenance functions for POWER8
  - No Application outages
  - Mode status of LPAR dictates LPM movement
- Used for Workload balancing
  - Provide better distribution server resources



धन्यवाद
Hindi

谢 谢
Simplified Chinese

תודה רבה
Hebrew

Спасибо
Russian

Gracias
Spanish

Thank You
English

شكرا
Arabic

Obrigado
Brazilian Portuguese

Danke
German

Grazie
Italian

감사합니다
Korean

Merci
French

நன்றி
Tamil

謝 謝
Traditional Chinese

ขอบคุณ
Thai