Unscrambling the Alphabet Soup

An Interconnect Primer

Dr Matthew Grosvenor
Principal Engineer (Exablaze)
Global STAC Live

Announcements

Exablaze is now part of Cisco.

Better Together Now

Immediate improvements to customer experience

Hire and train global Exablaze specialist sales team

Better Together Now

Immediate improvements to customer experience

- ✓ Hire and train global Exablaze specialist sales team
- ✓ Train global Exablaze specialist technical support team

Better Together Now

Immediate improvements to customer experience

- Hire and train global Exablaze specialist sales team
- Train global Exablaze specialist technical support team
- Retain Exablaze product range and roadmap

Better Together 2020

Longer term integration work

- Cisco global 24/7 technical support and RMA options
- Cisco global partner, reseller and distributor network
- Cisco manufacturing and procurement integration
- Cisco global shipping and logistics

New Product Announcements

New Smart NICs

ExaNIC X25

New Smart NICs

ExaNIC X25

ExaNIC X100

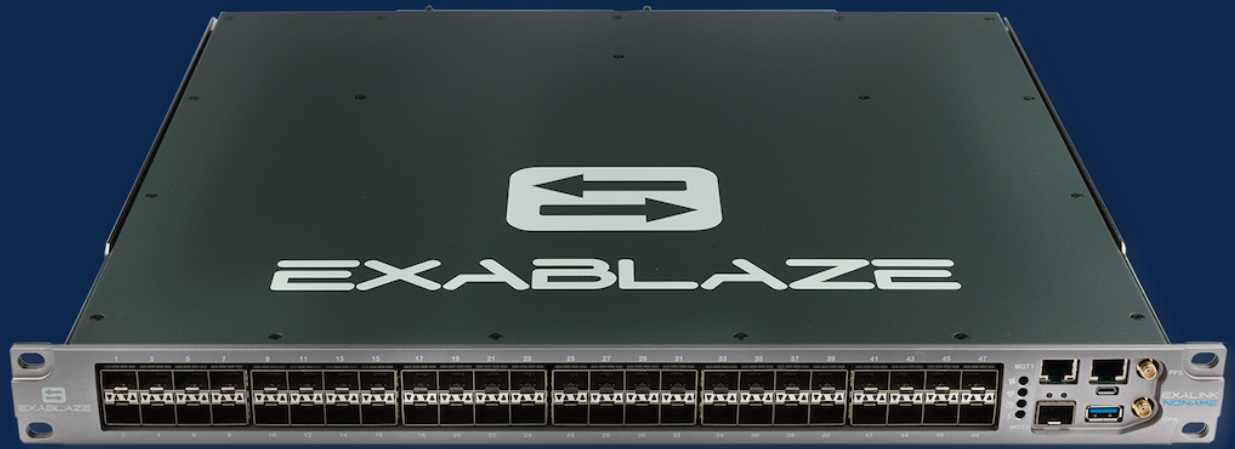
ExaNIC X25 / X100 Features

Ultra-low Latency Smart NIC cards optimized for financial trading

- New Xilinx KU3P Ultrascale+ based FPGA SmartNICs
- Ultra-low latency 10GbE NIC
- 568ns trigger to response latency*
- 2 port SFP28 / QSFP28 form factors
- Optional 4 / 9GB DDR4 onboard
- 25GbE capable hardware**
- Firmware Development Kit

* Not a STAC Benchmark (yet)
**25GbE firmware late 2020

New Switch Platform



ExaLINK Triton

ExaLINK Triton Features

Ultra-low Latency Smart Switch optimized for financial trading

- Xilinx Ultrascale+ based FPGA Smart Switch platform
- Initial L3+ switch firmware release (Aug)
  - L2 switching (MAC learning, VLAN, IGMP, STP)
  - L3 switching (IP routing, BGP, PIM)
  - L4 switching (NAT)
  - Latency 100-200ns*
- 25G capable hardware**
- Firmware Development Kit (VU35P-3)

* Not a STAC Benchmark
**25GbE firmware in later release

Unscrambling the Alphabet Soup

An Interconnect Primer

The Interconnect Alphabet

The Interconnect Alphabet

- AGP
- BlueLink
- CCIX
- CXL
- GenZ
- Ethernet
- FSB
- Infiniband
- ISA
- NVLink
- OpenCAPI
- PCIe 3.0 / 4.0 / 5.0 / 6.0
- QPI
- UPI
- VCLX

The Interconnect Alphabet Soup

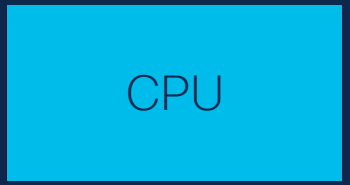
- AGP
- BlueLink
- CCIX
- CXL
- GenZ
- Ethernet
- FSB
- Infiniband
- ISA
- NVLink
- OpenCAPI
- PCIe 3.0 / 4.0 / 5.0 / 6.0
- QPI
- UPI
- VCLX

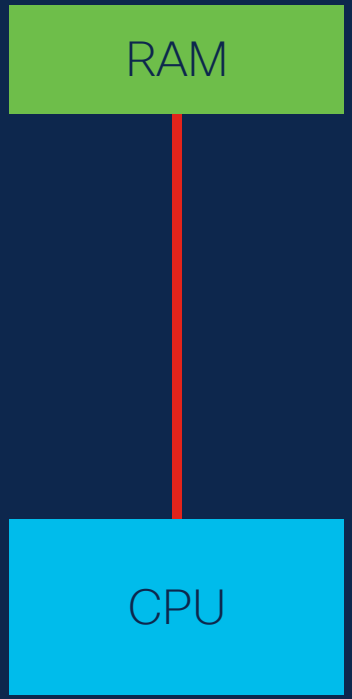
What is an interconnect anyway?

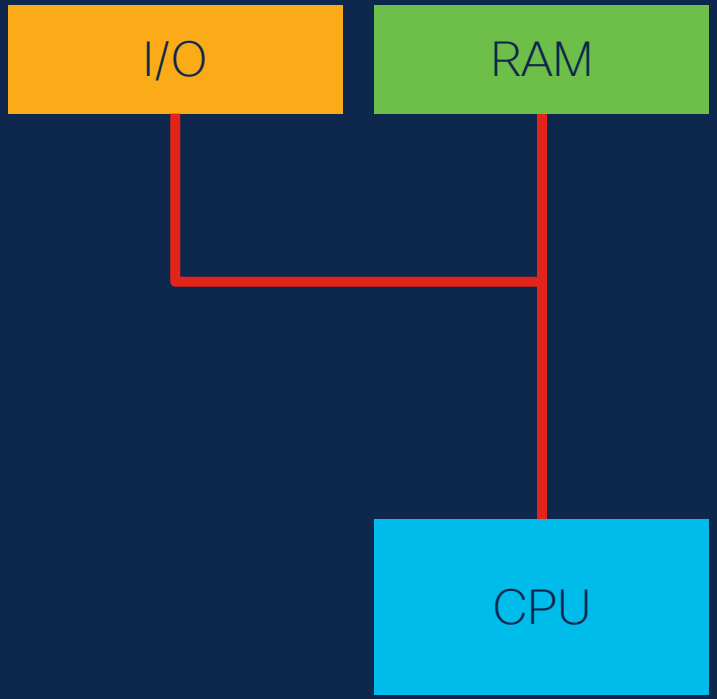
What is cache coherence?

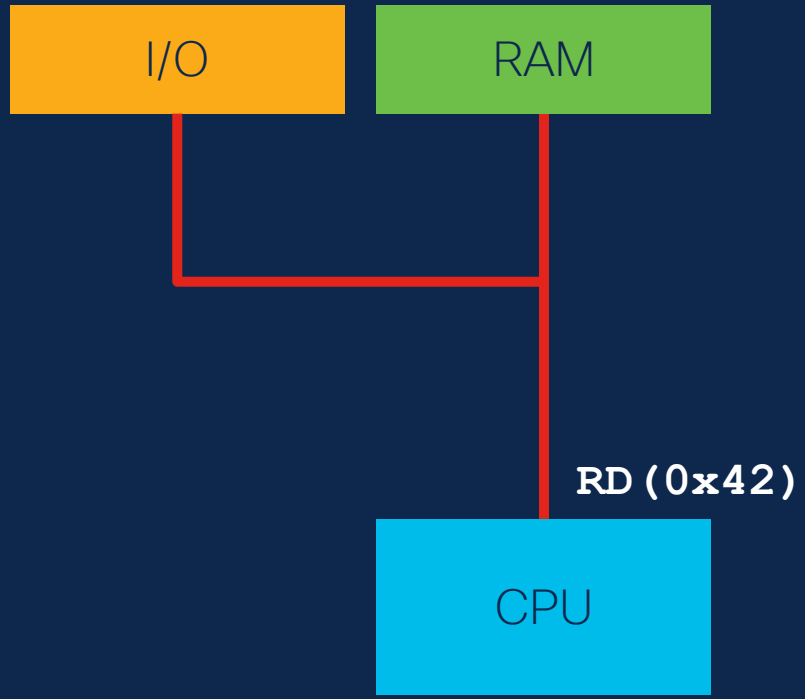
Before we go forward, we must go…

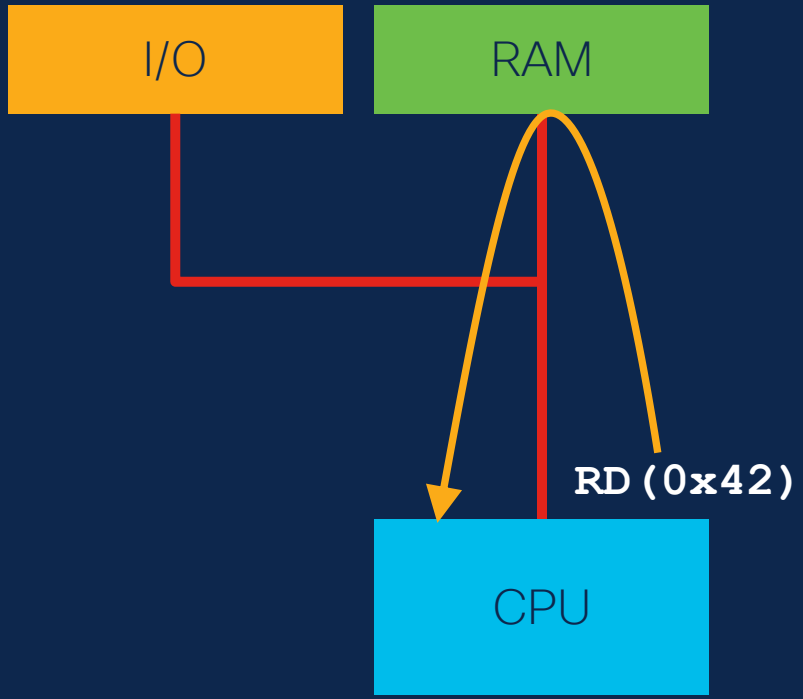

CALIFORNIA
BACKINTIME
BACK IN TIME

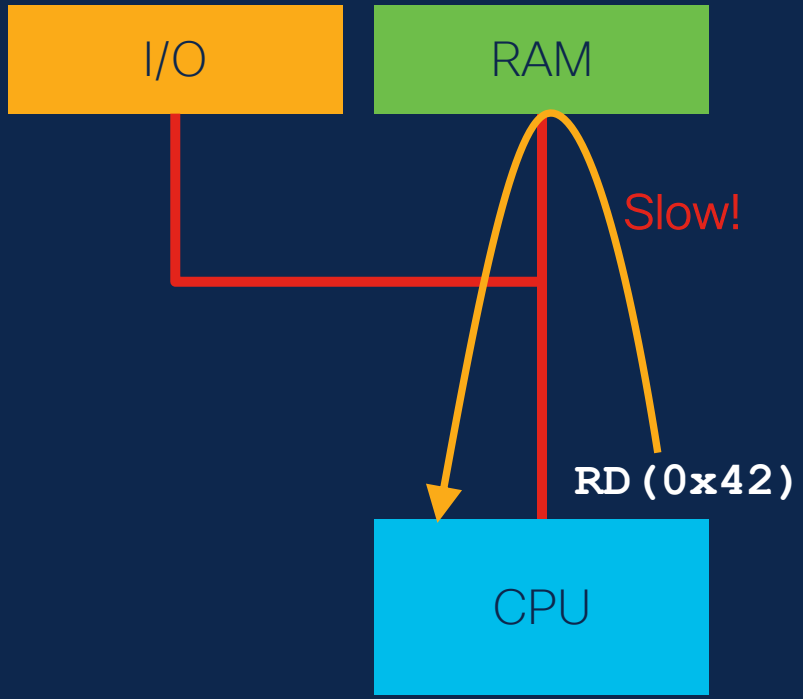


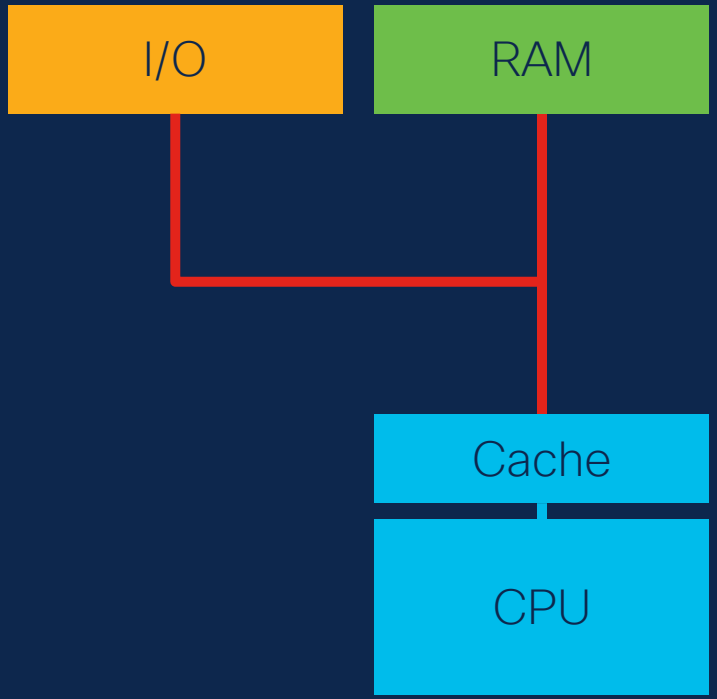


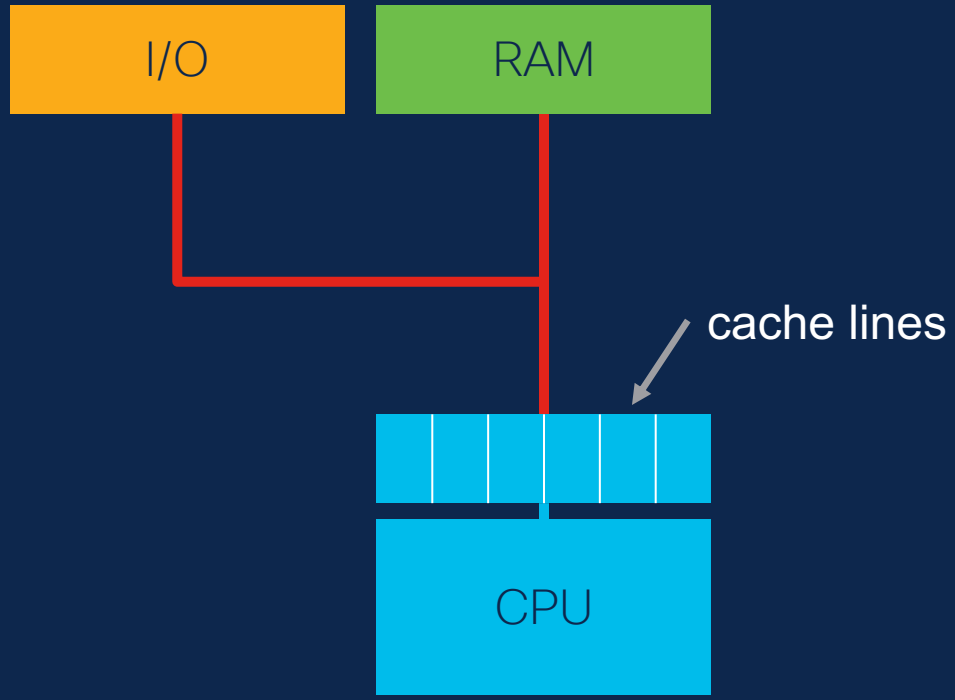


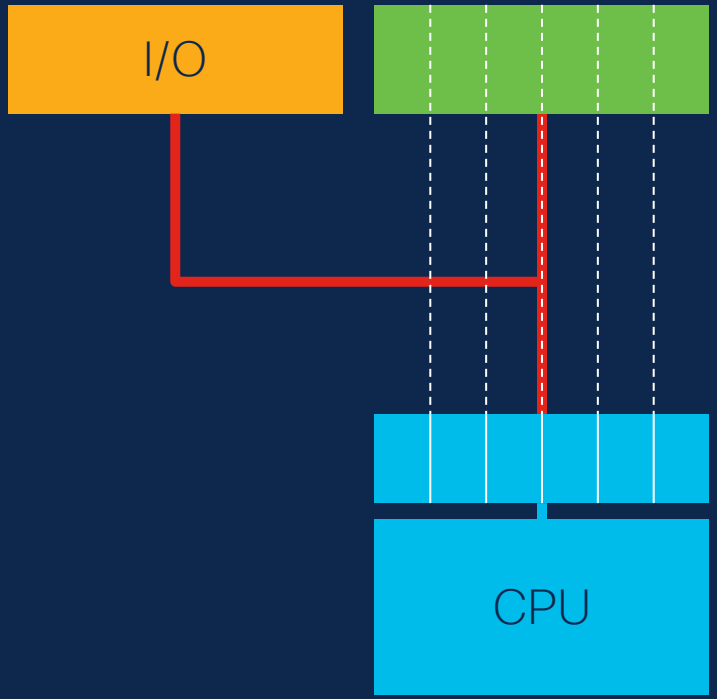


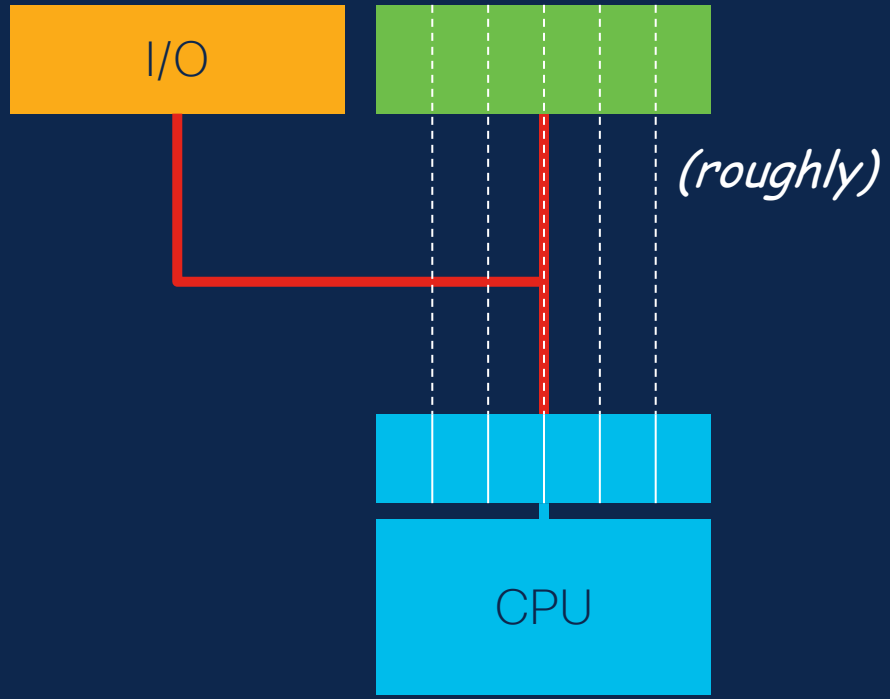


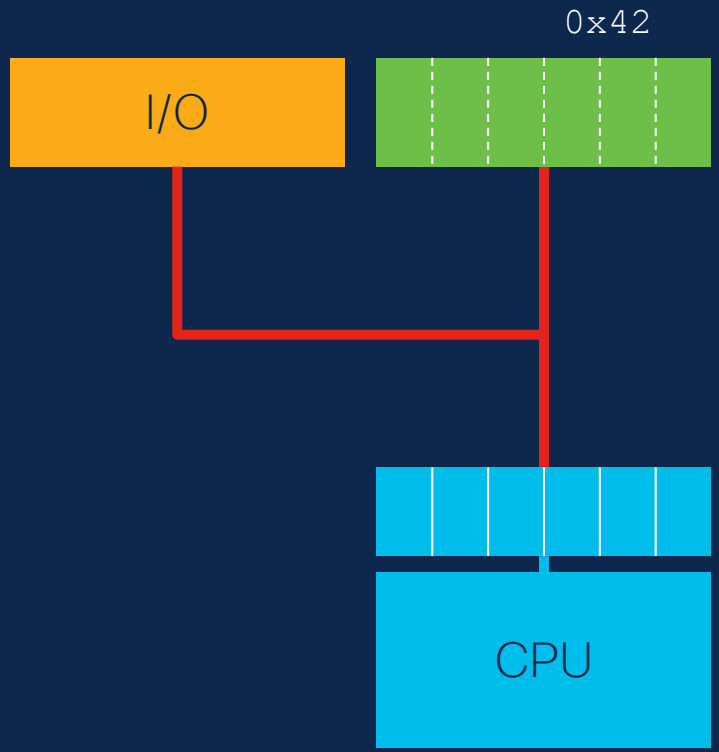


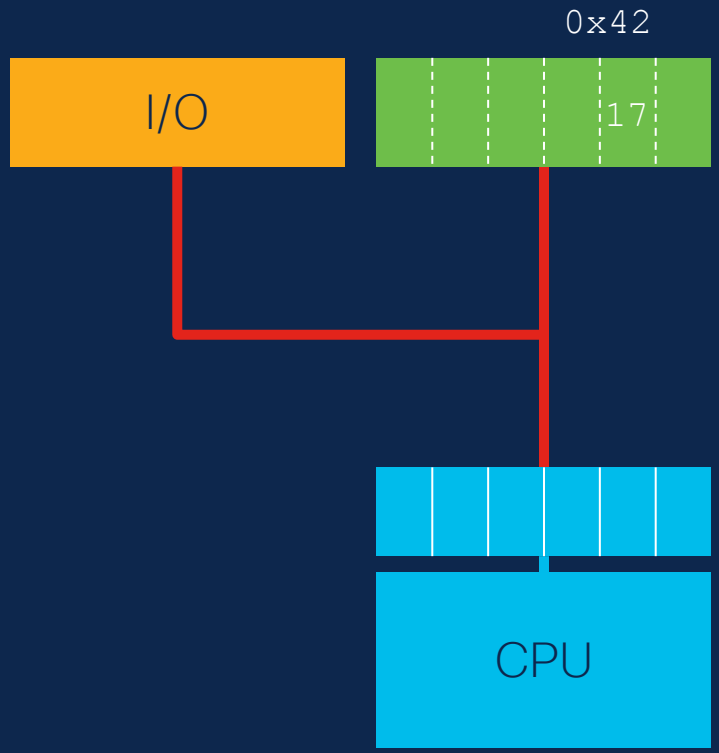


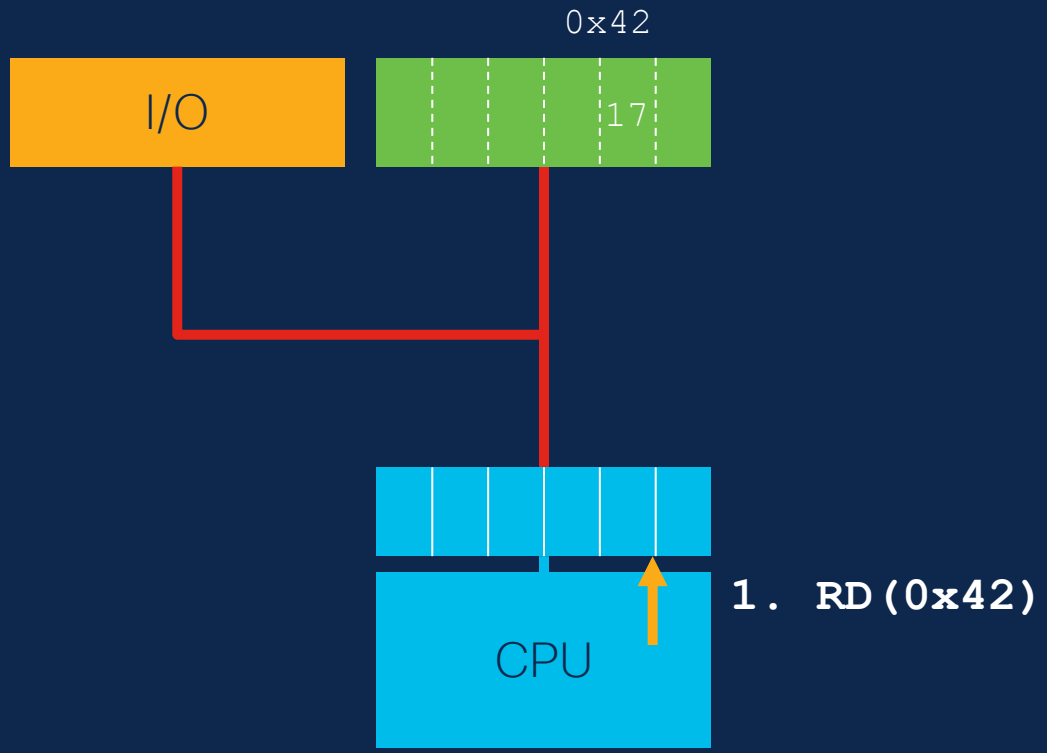


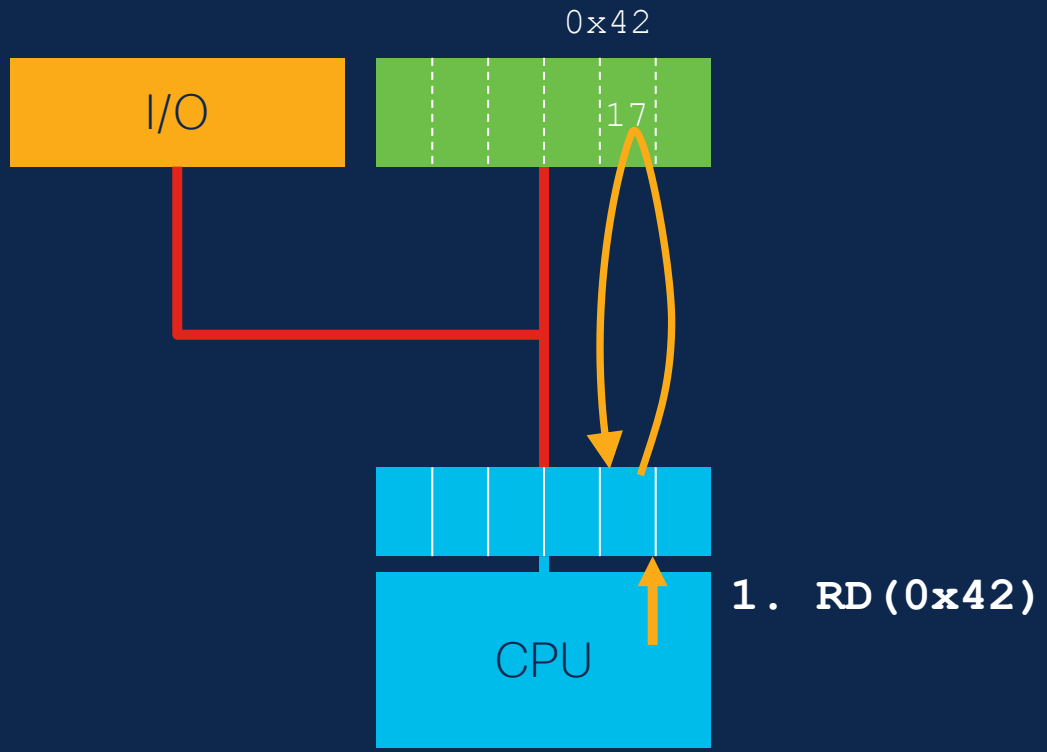


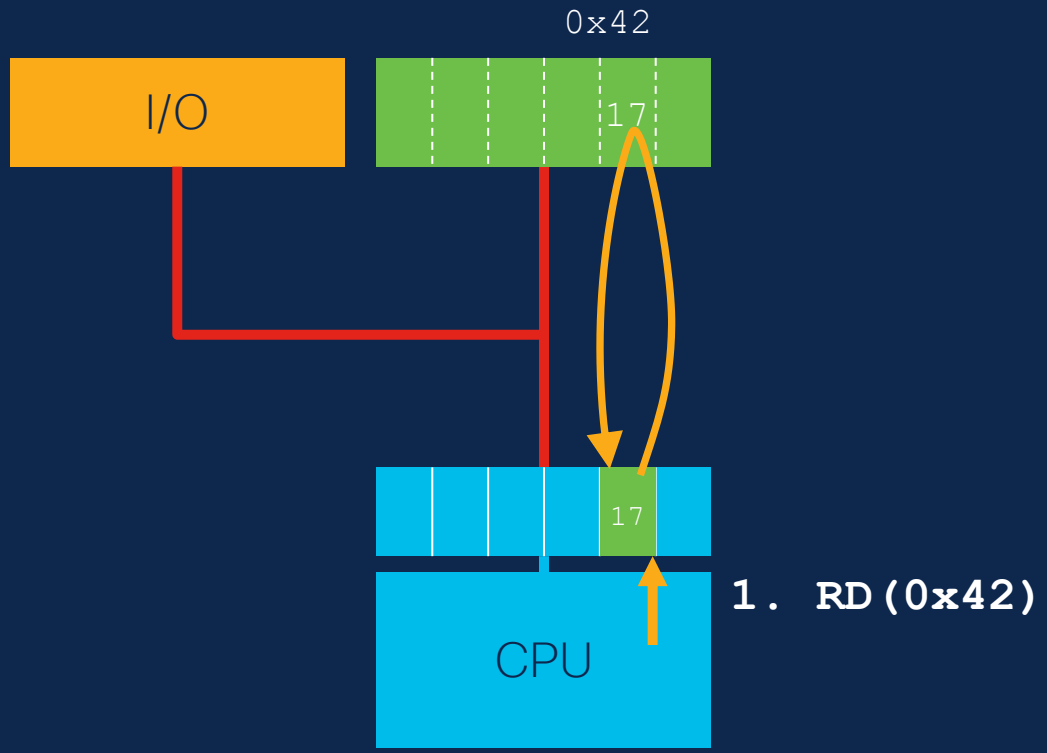


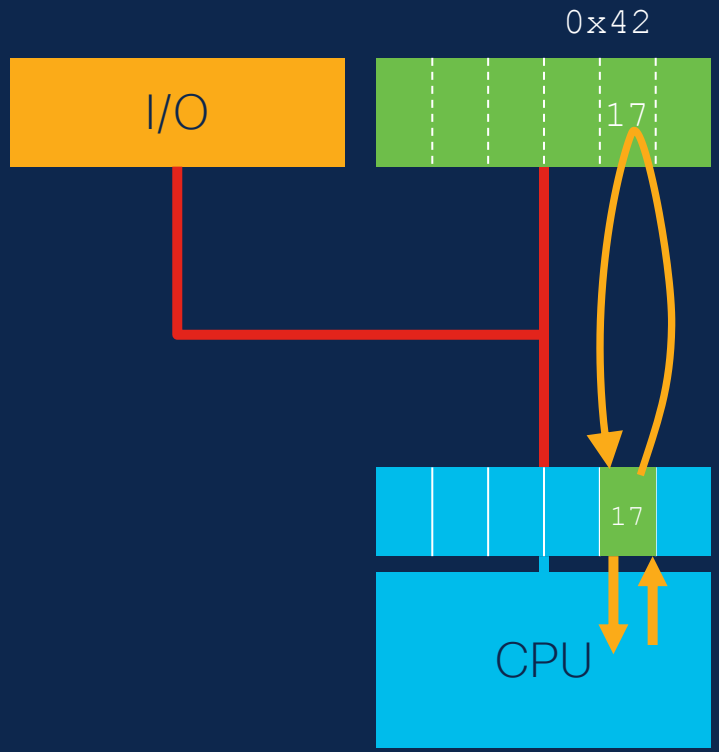




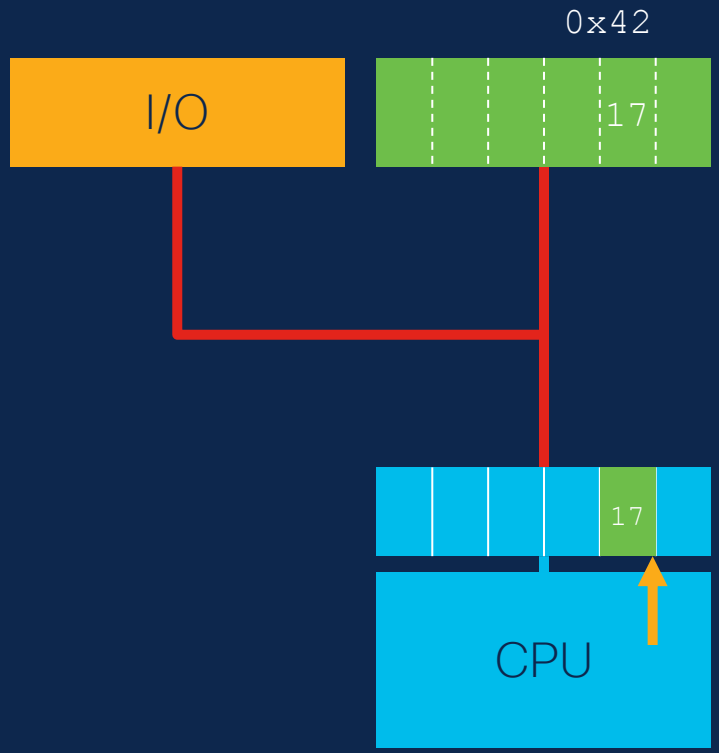




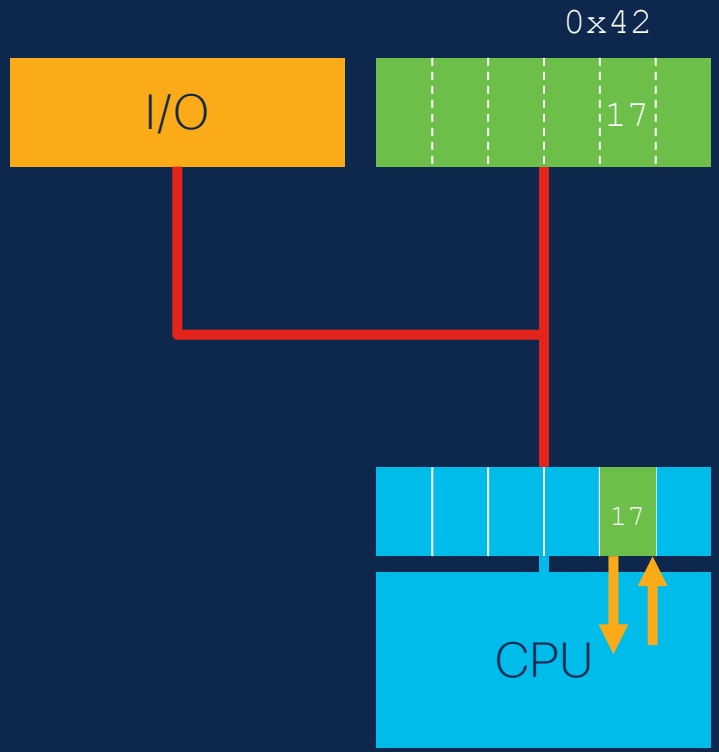




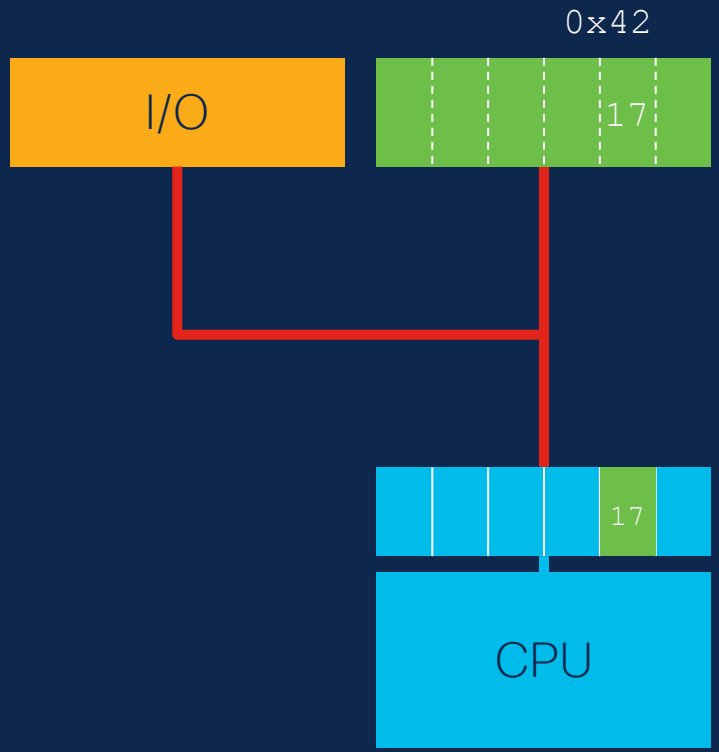
1. **RD(0x42) = 17** Slow!



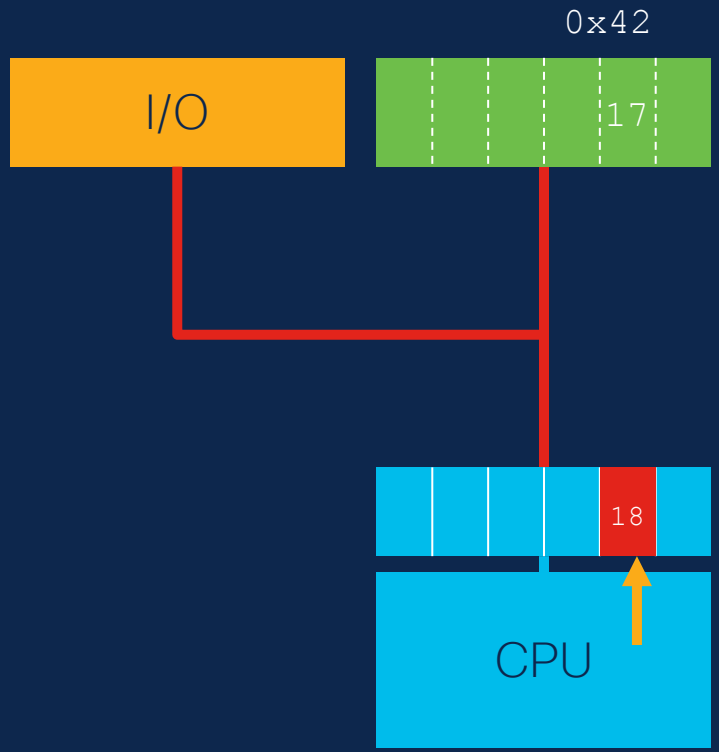
1. RD(0x42) = 17
2. RD(0x42)



1. RD(0x42) = 17
2. RD(0x42) = 17 Fast!

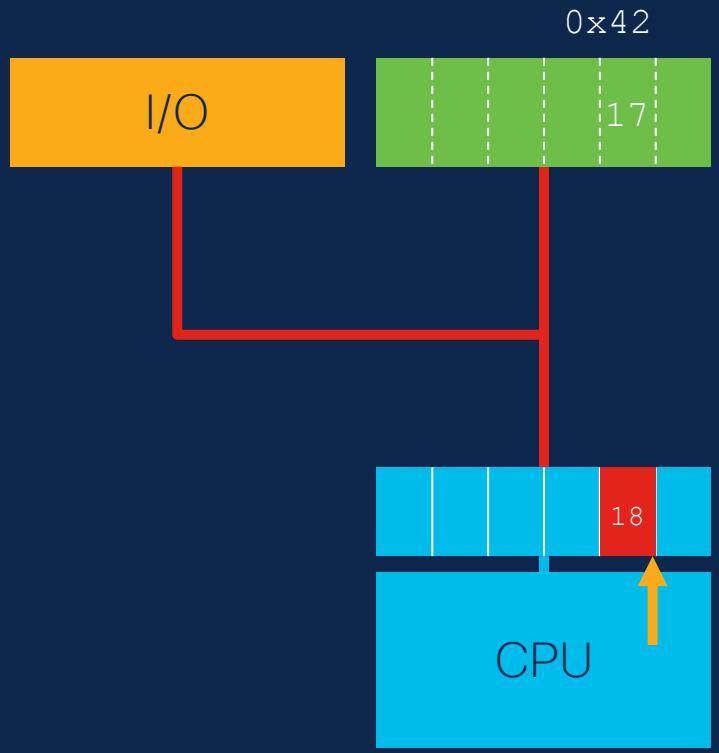


1. RD(0x42) = 17
2. RD(0x42) = 17
3. WR(0x42) = 18

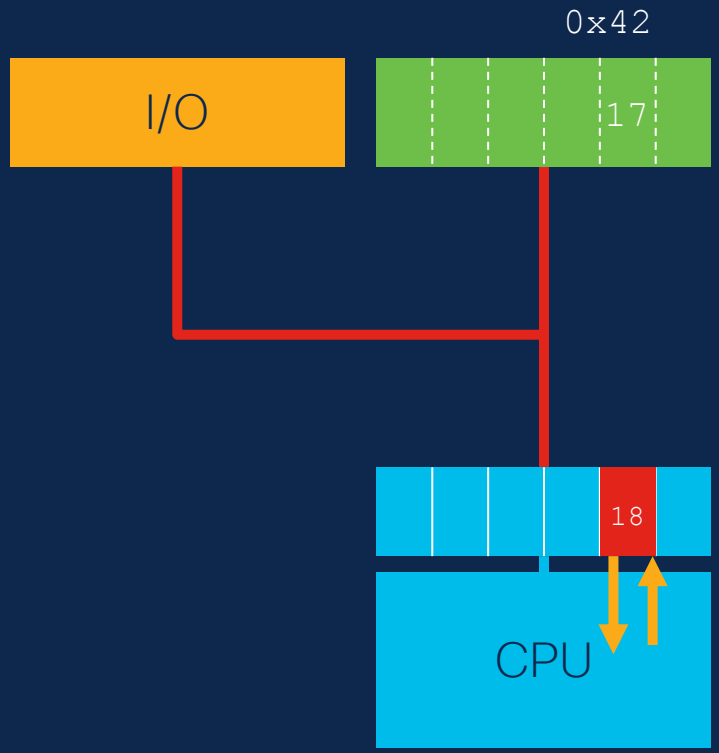


1. RD(0x42) = 17
2. RD(0x42) = 17
3. WR(0x42) = 18

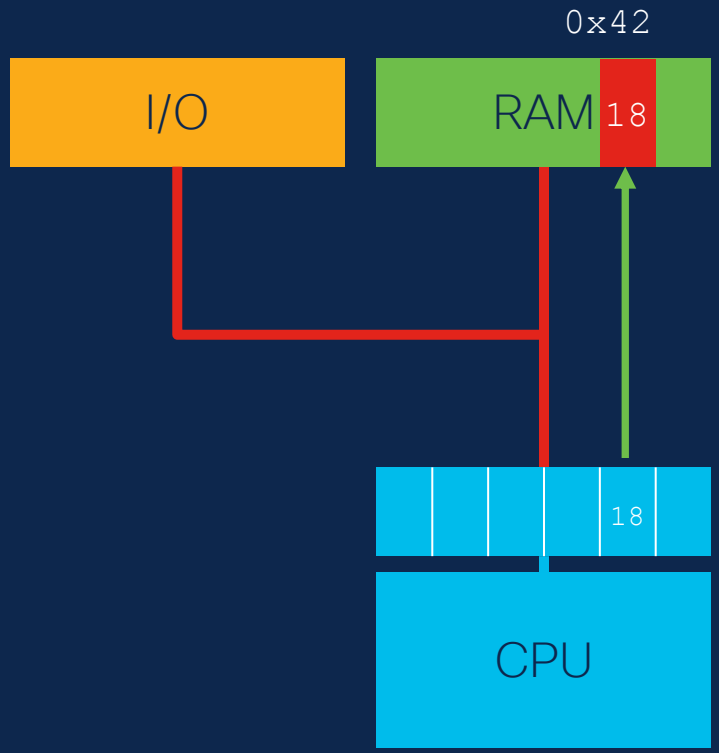
Fast!



1. RD(0x42) = 17
2. RD(0x42) = 17
3. WR(0x42) = 18
4. RD(0x42)

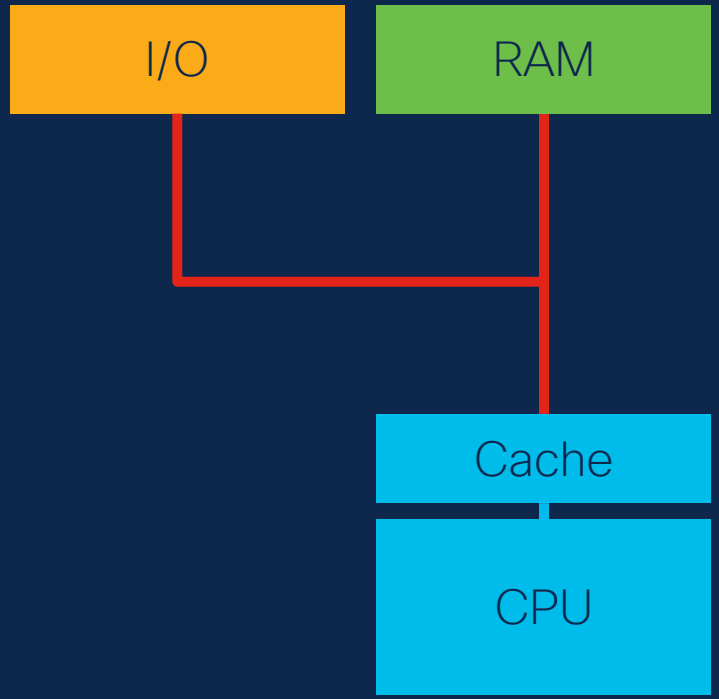


1. RD(0x42) = 17
2. RD(0x42) = 17
3. WR(0x42) = 18
4. RD(0x42) = 18 Fast!



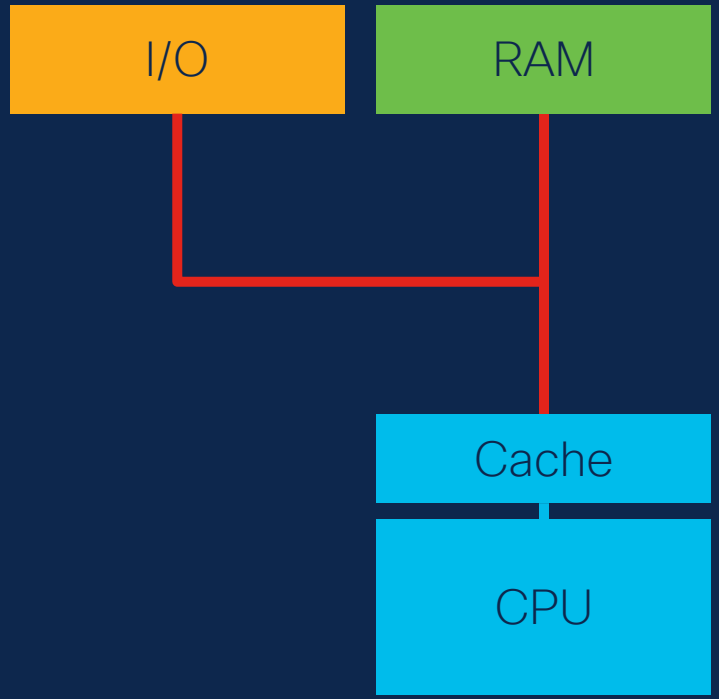
1. RD(0x42) = 17
2. RD(0x42) = 17
3. WR(0x42) = 18
4. RD(0x42) = 18

... FL



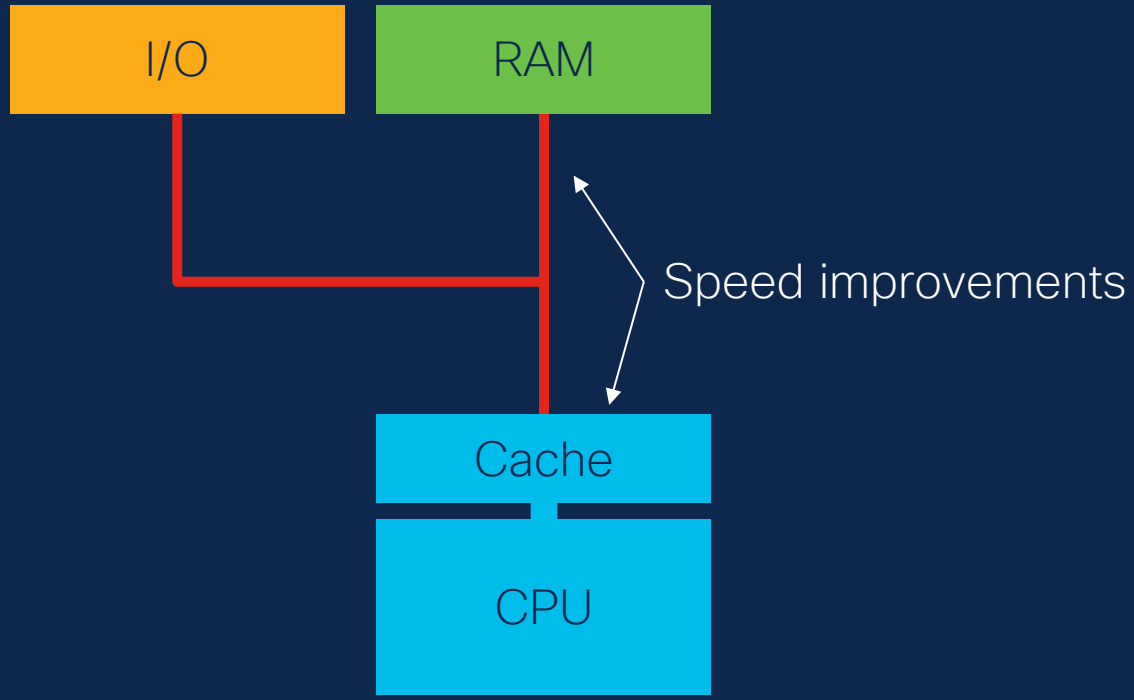
Caching:

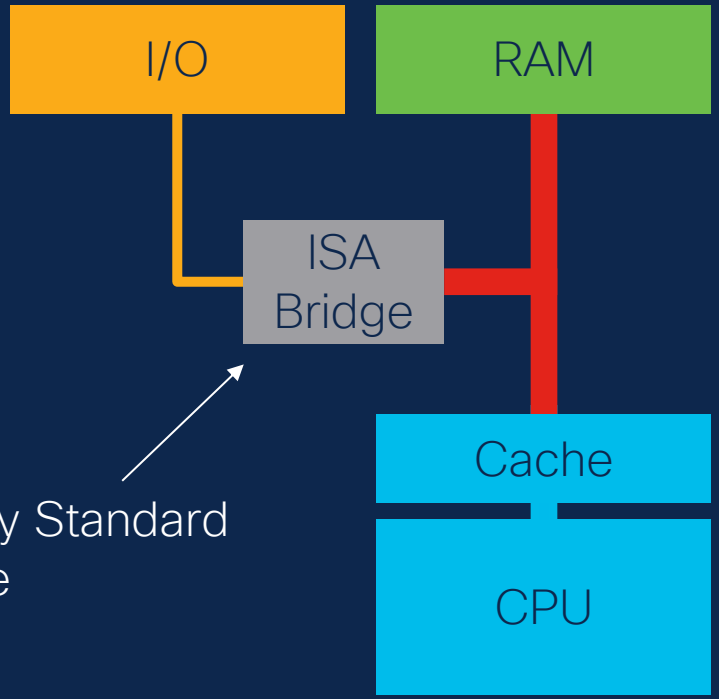
- Copy of recently used data



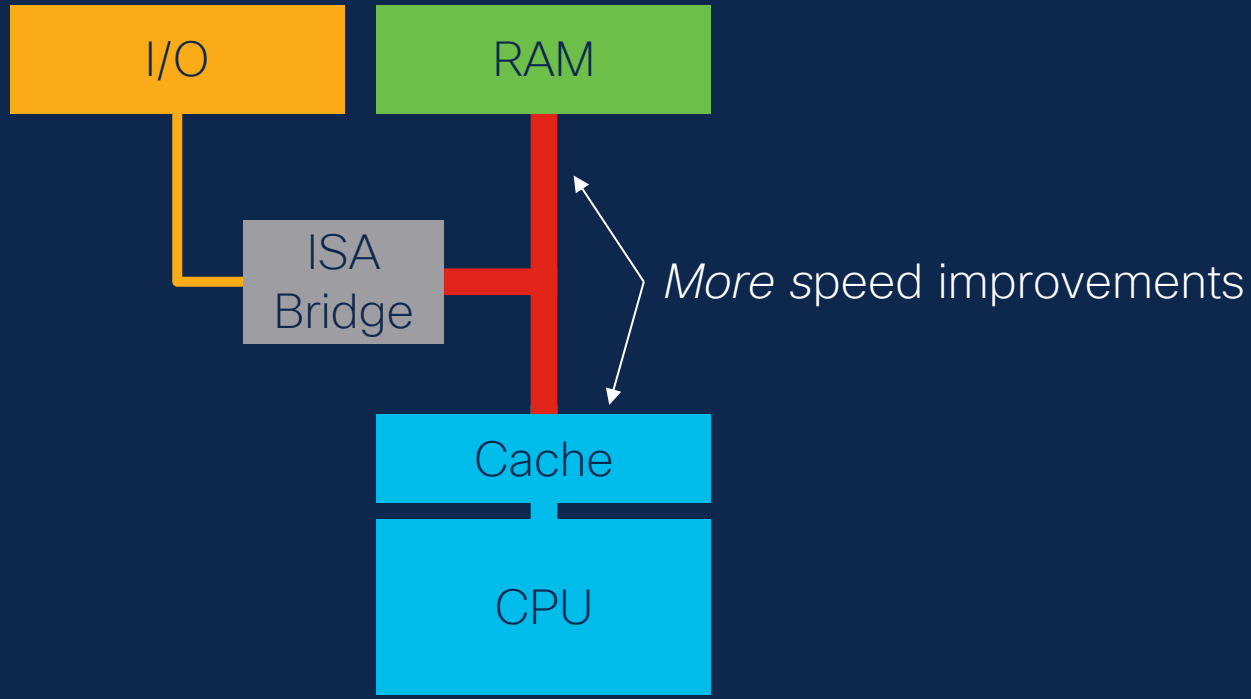
Caching:

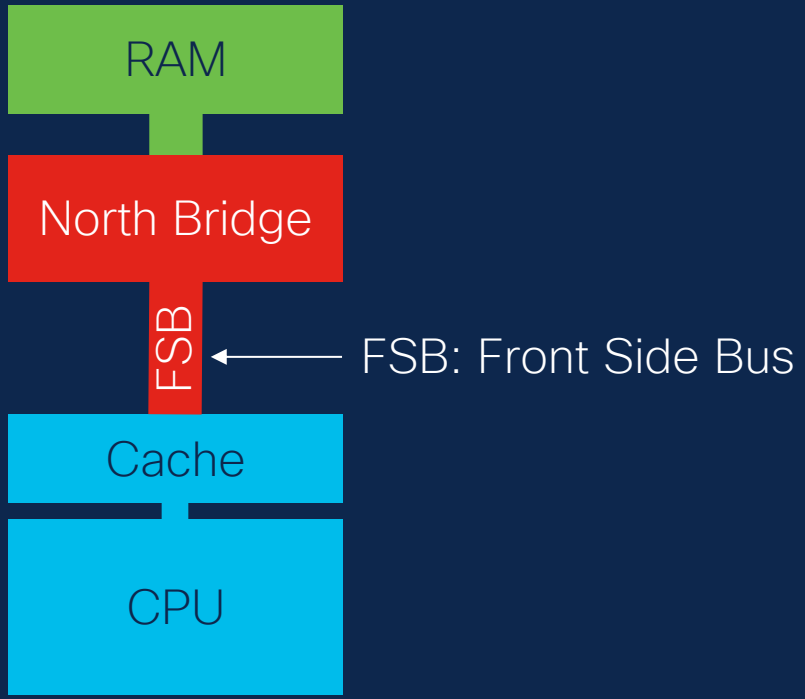
- Copy of recently used data
- Avoid RAM long delays



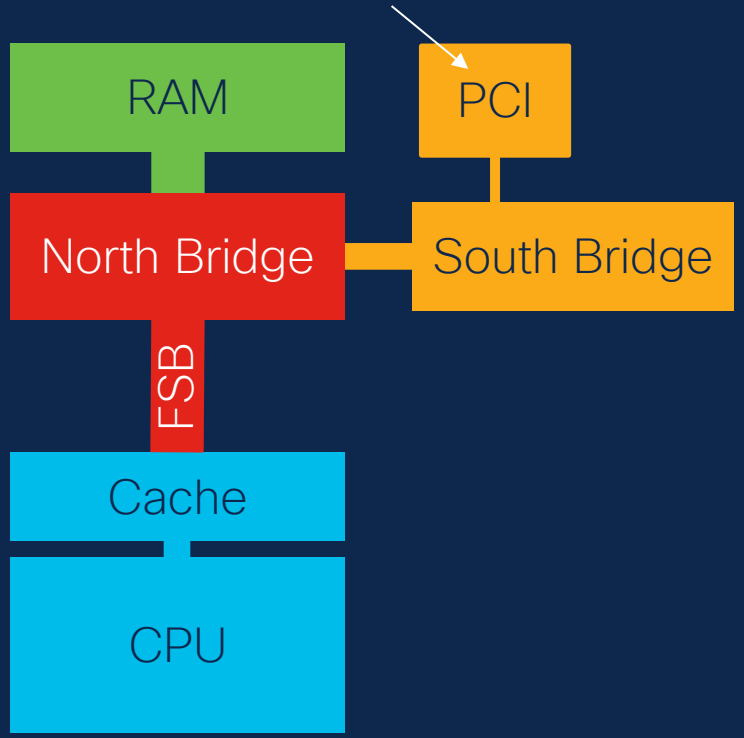


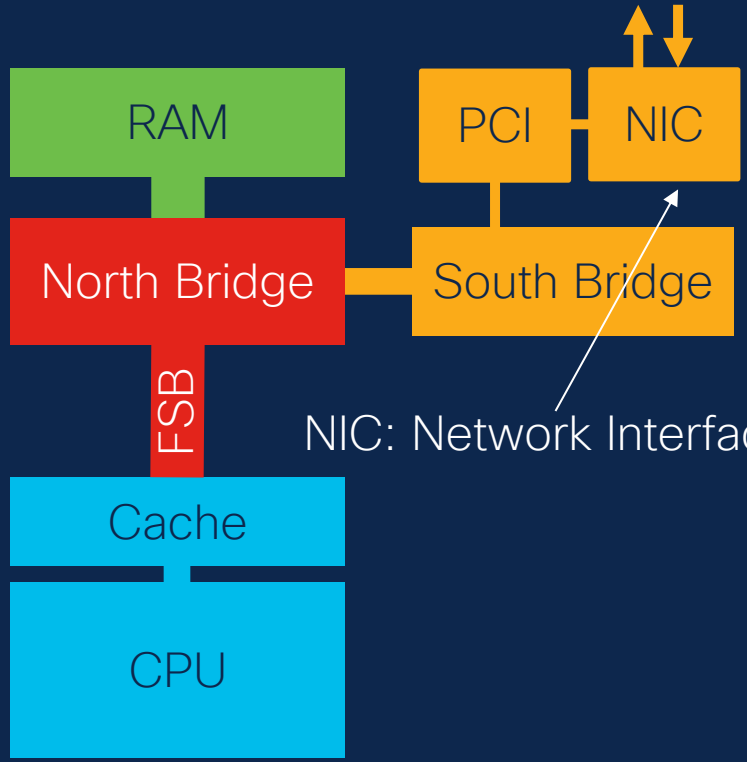
ISA: Industry Standard Architecture

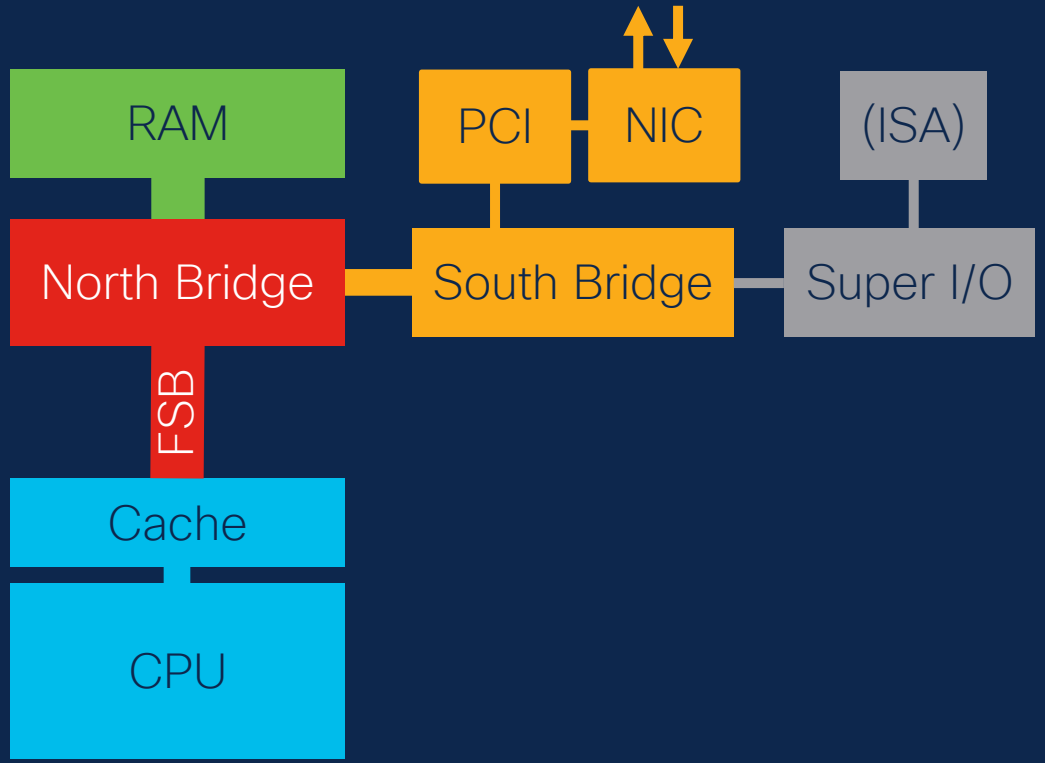


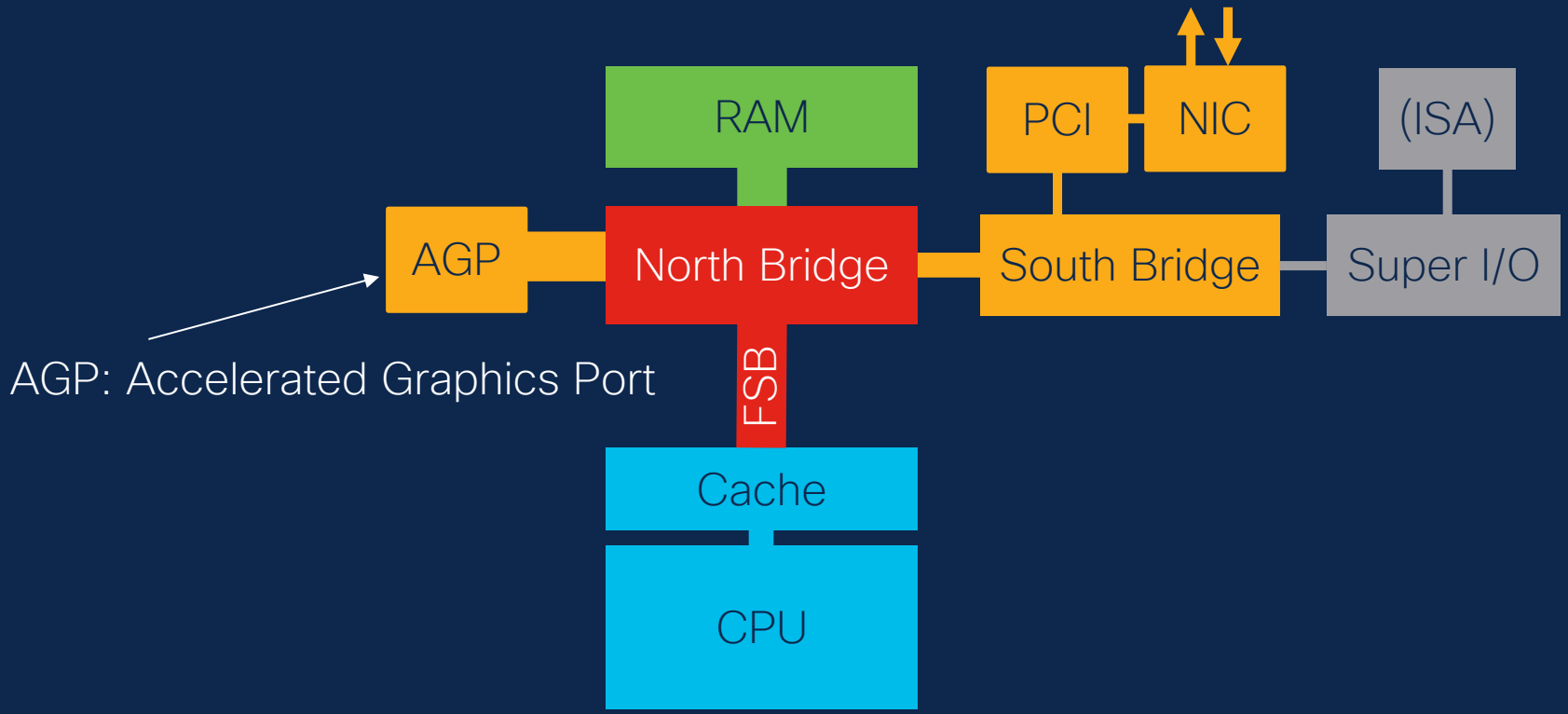


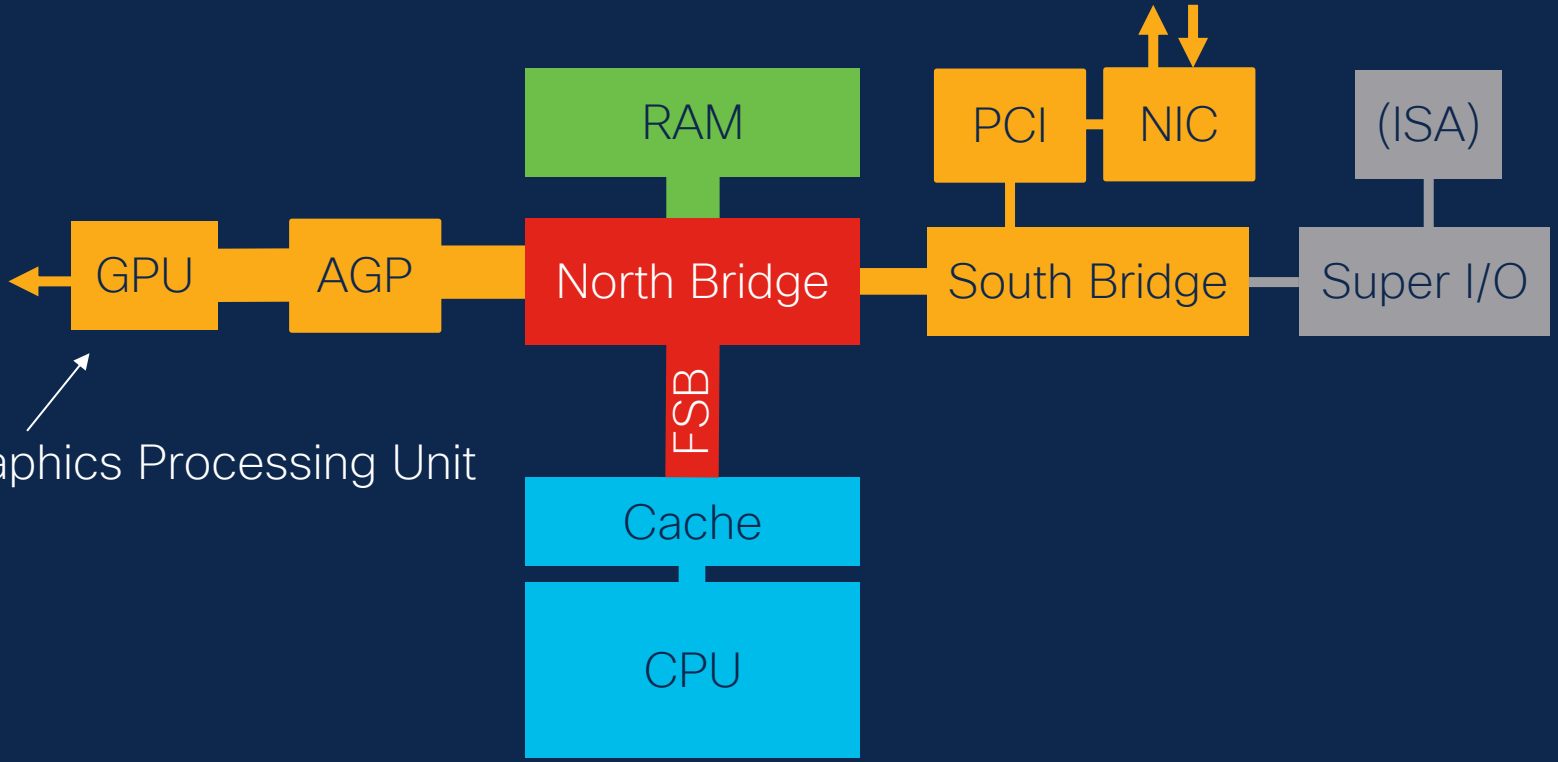
PCI: Peripheral Component Interconnect (bus)



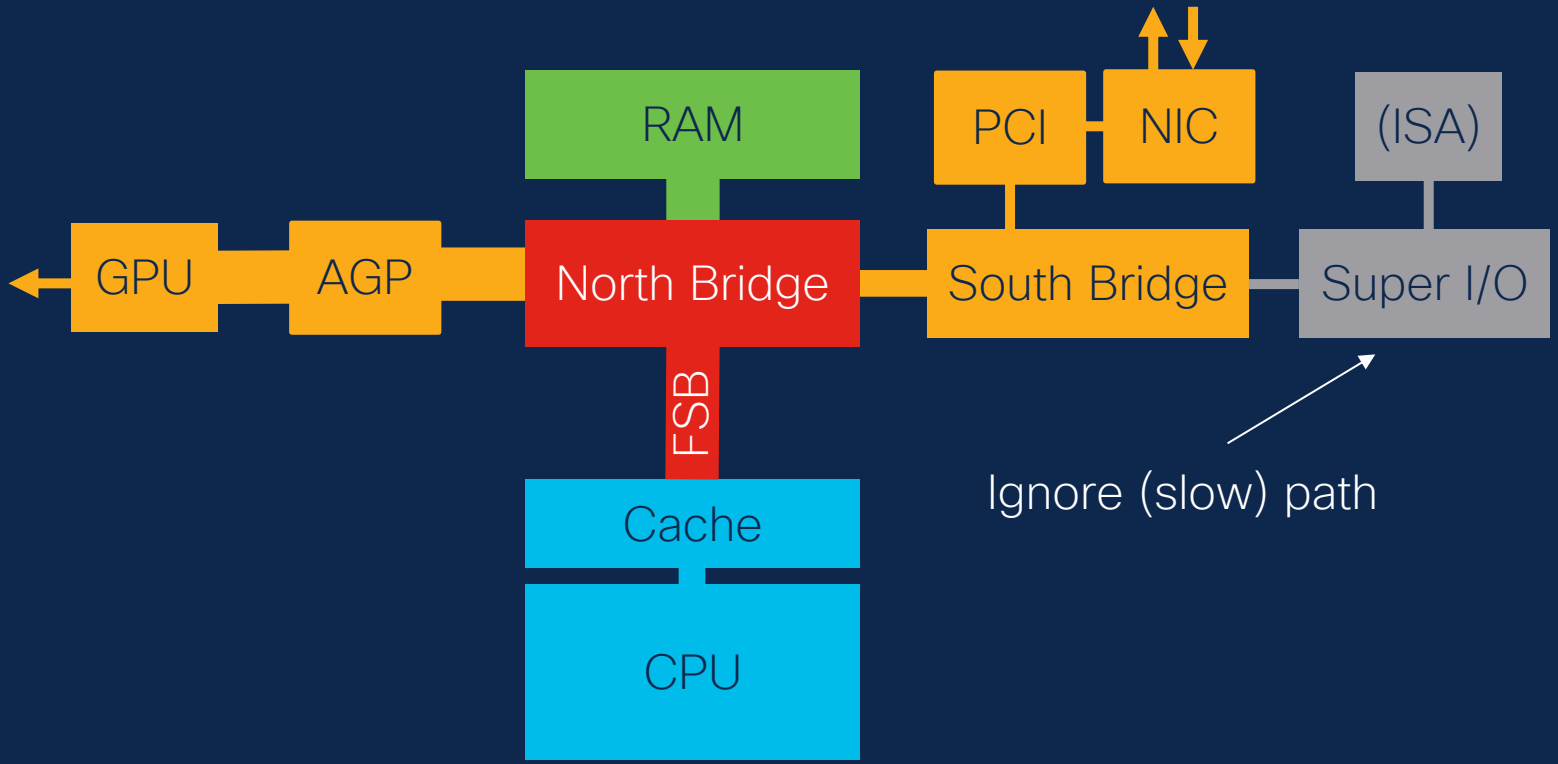


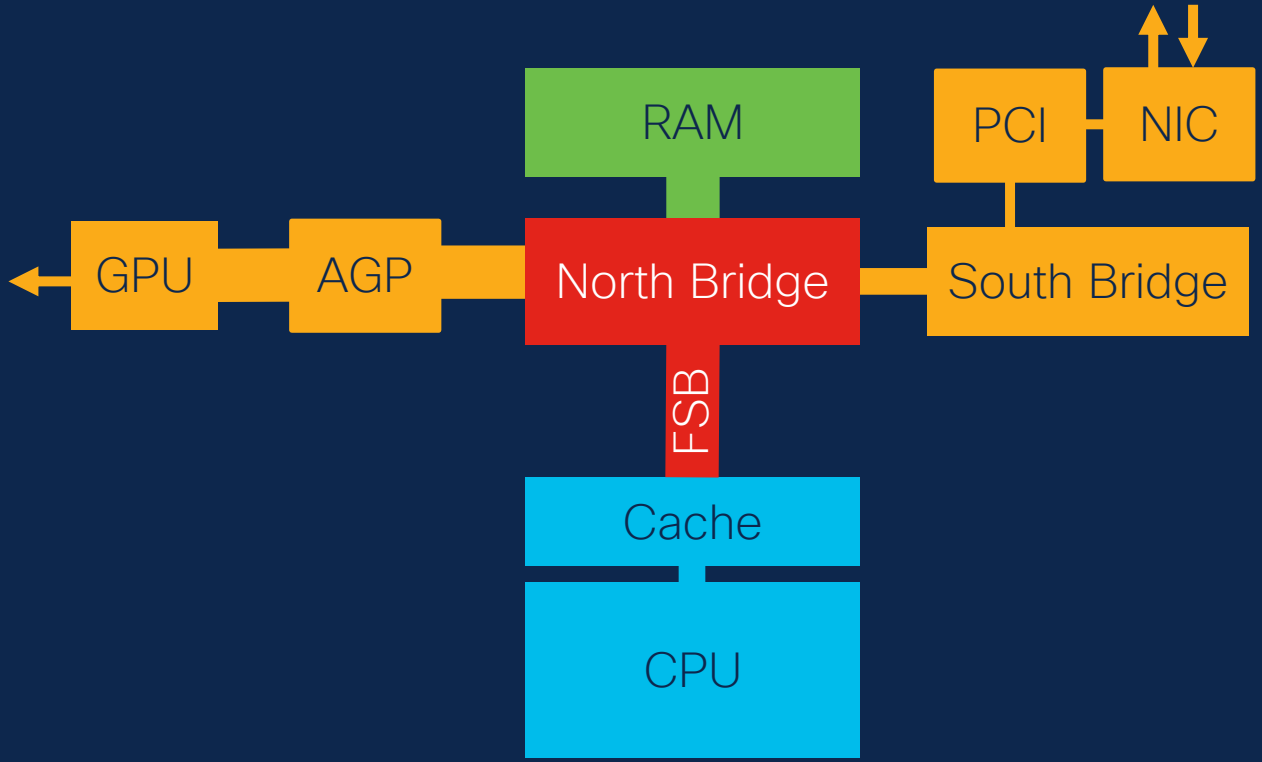


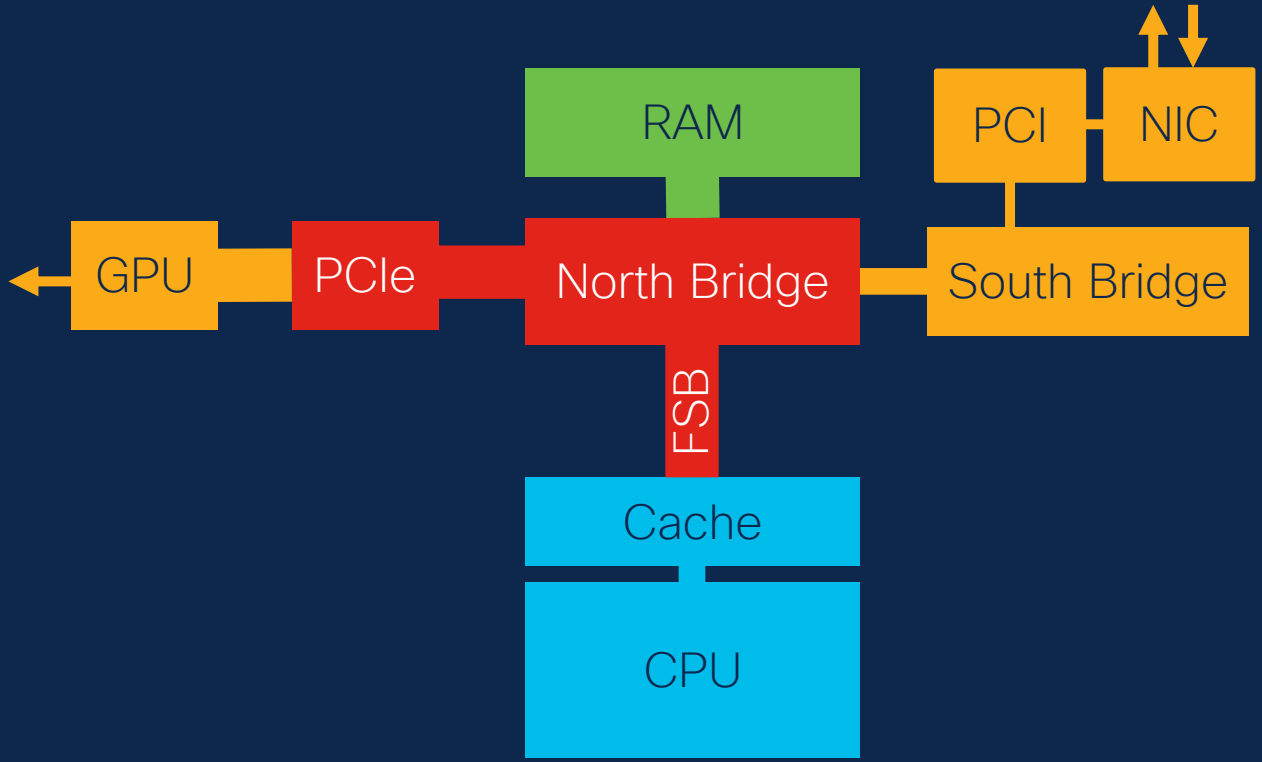


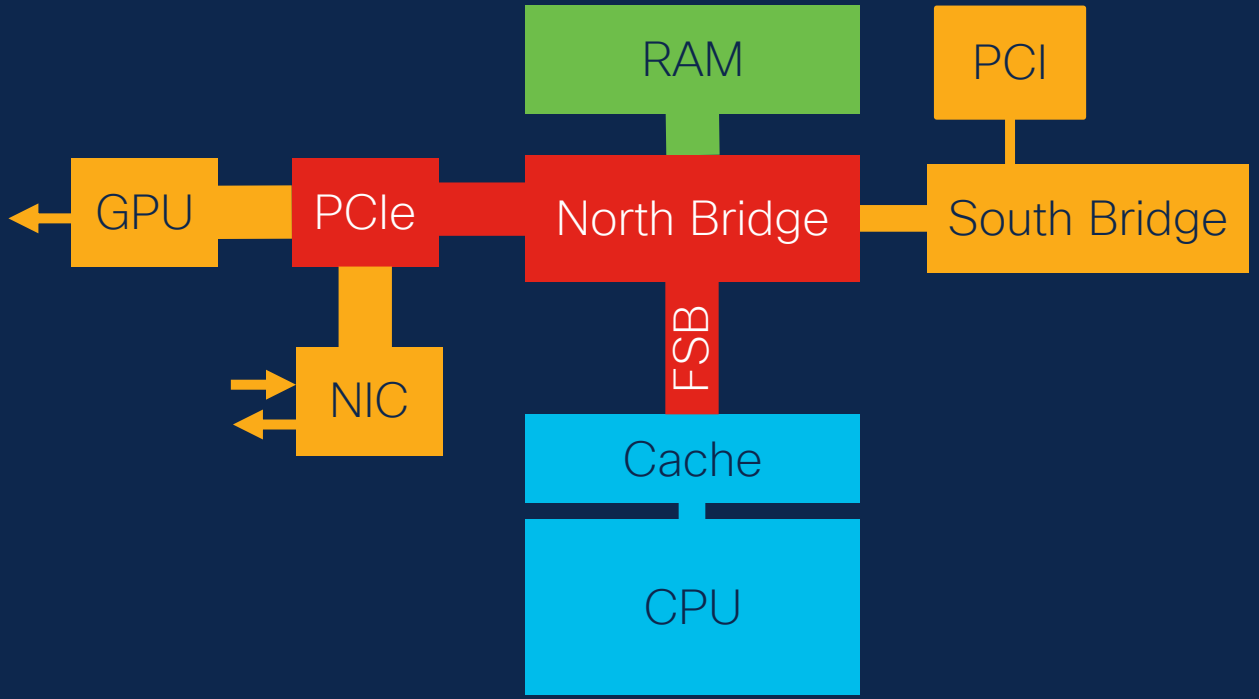


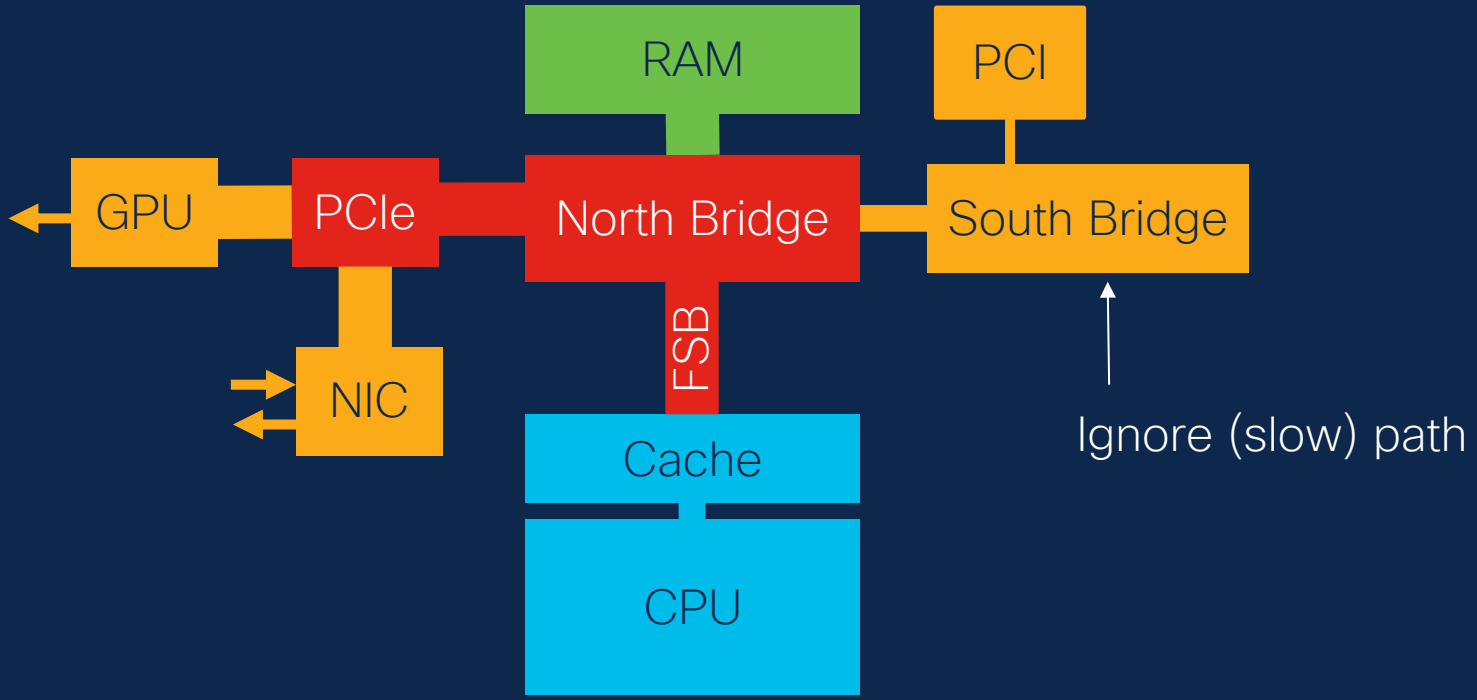
GPU: Graphics Processing Unit

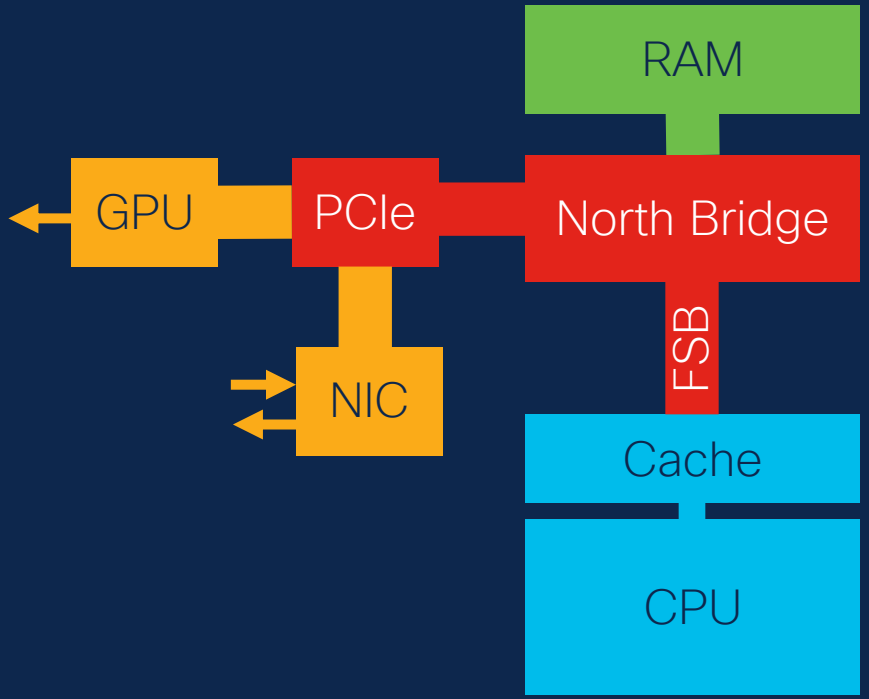


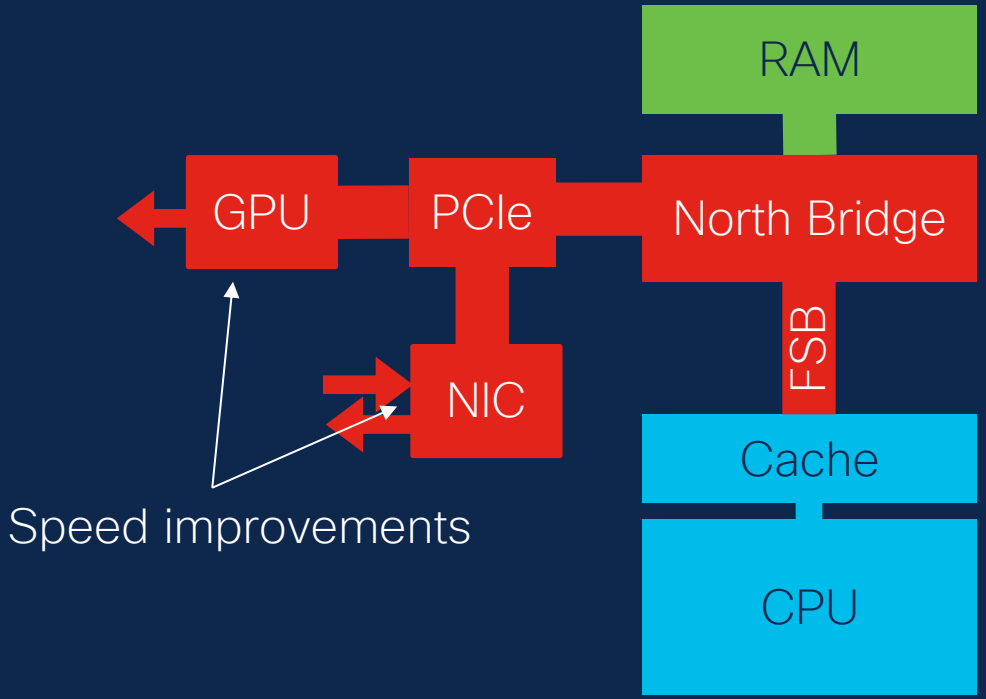


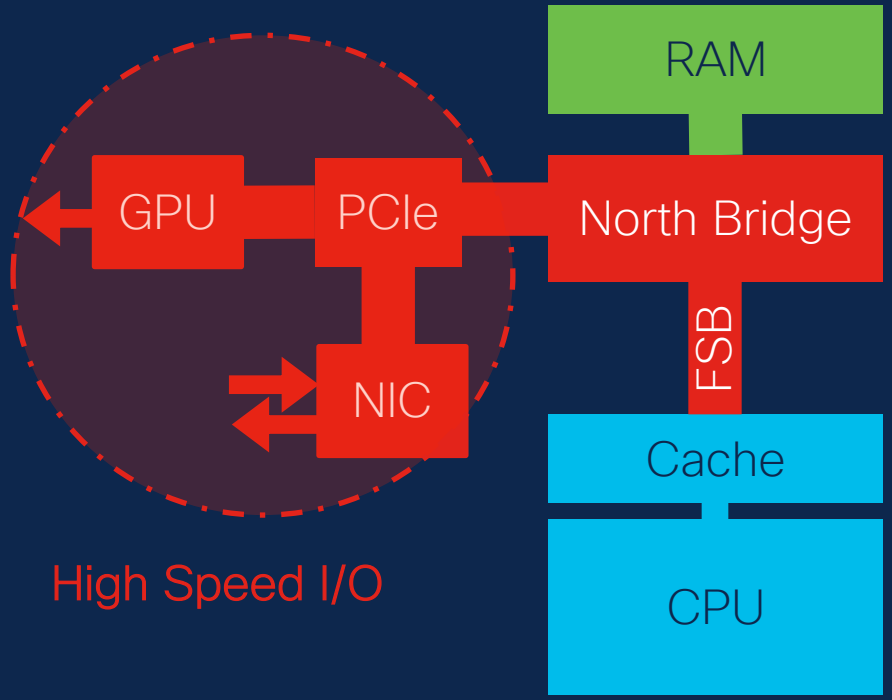


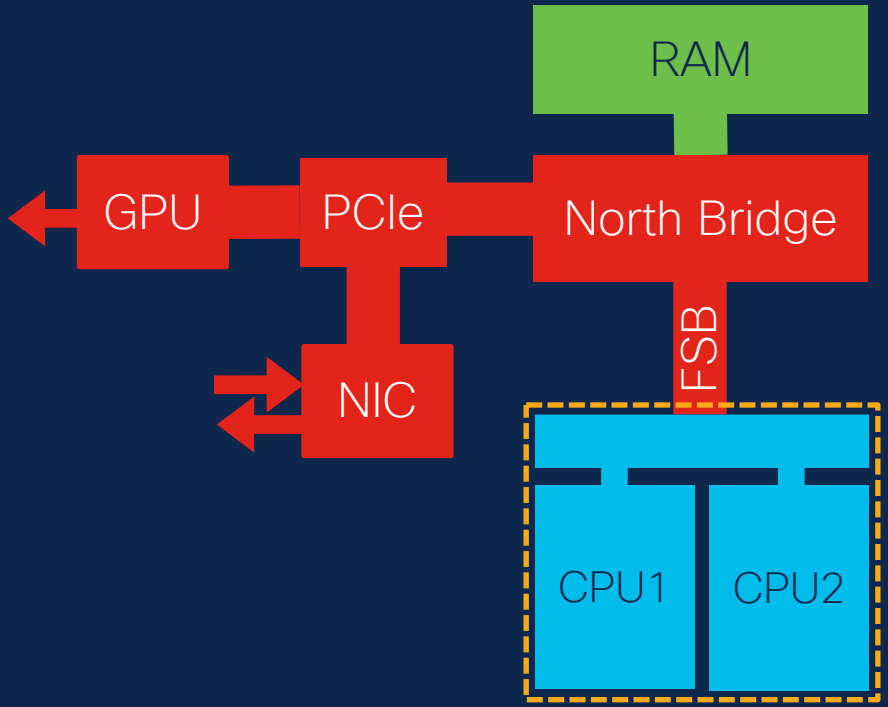


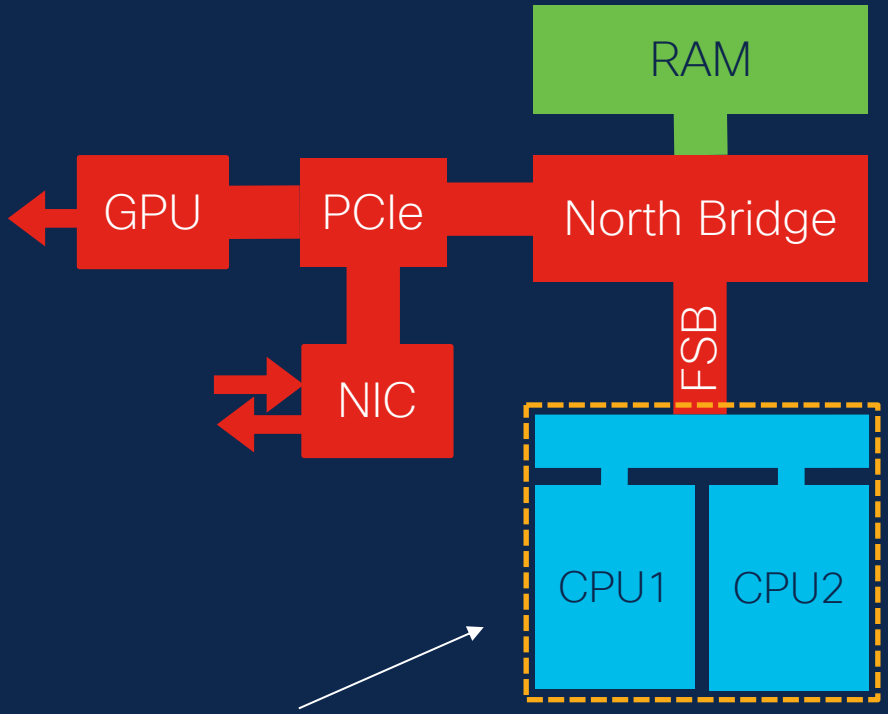




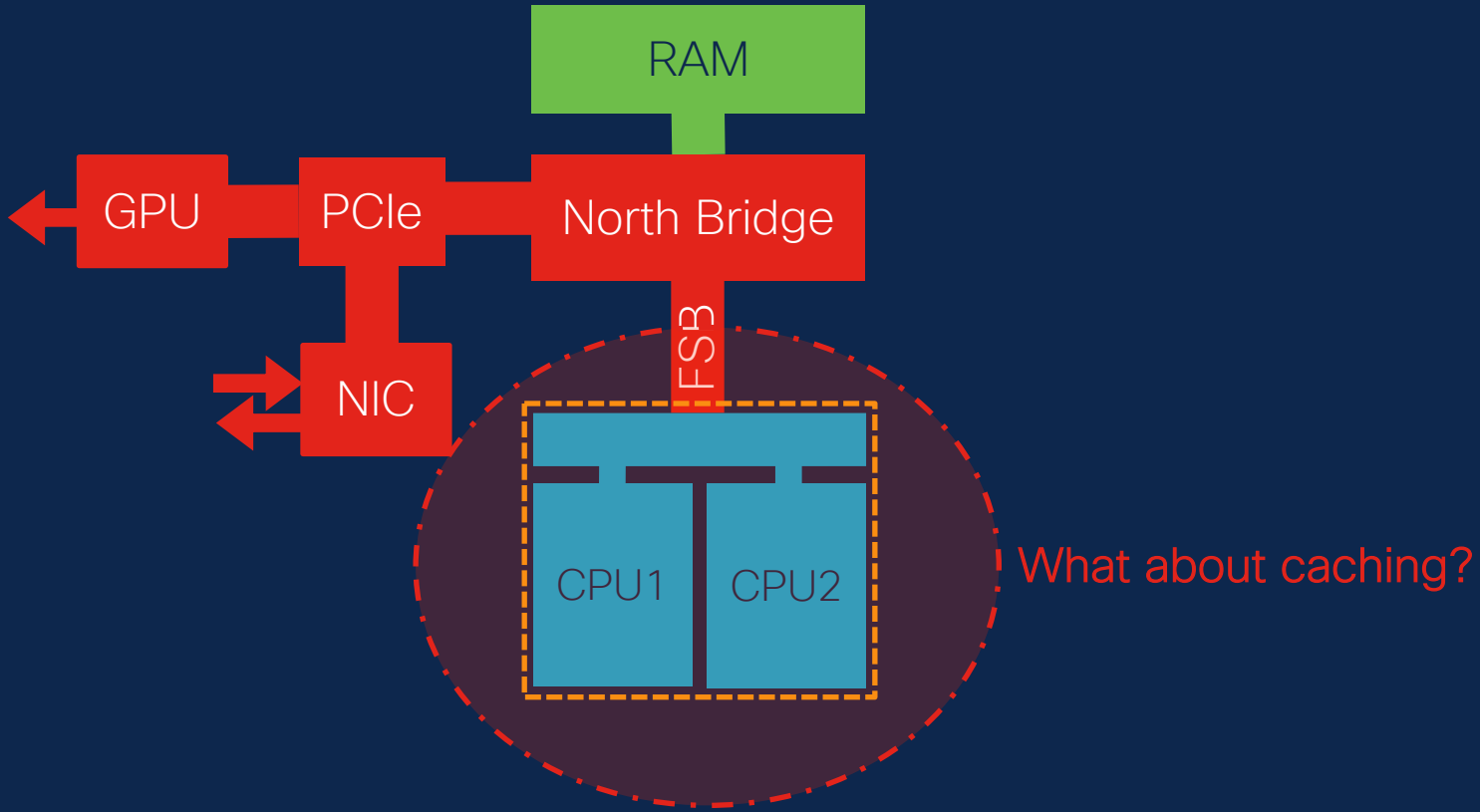


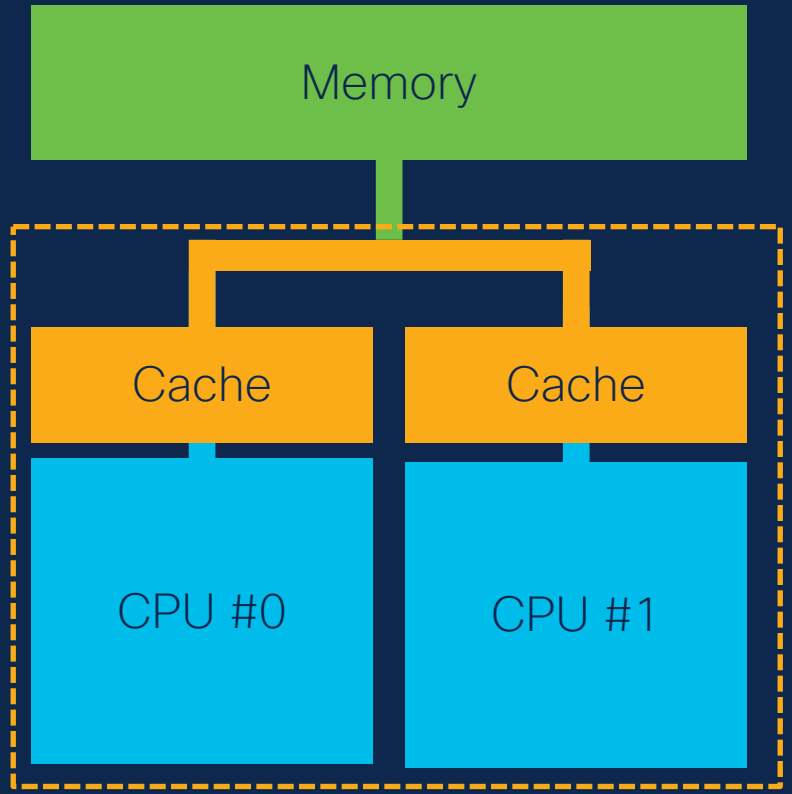


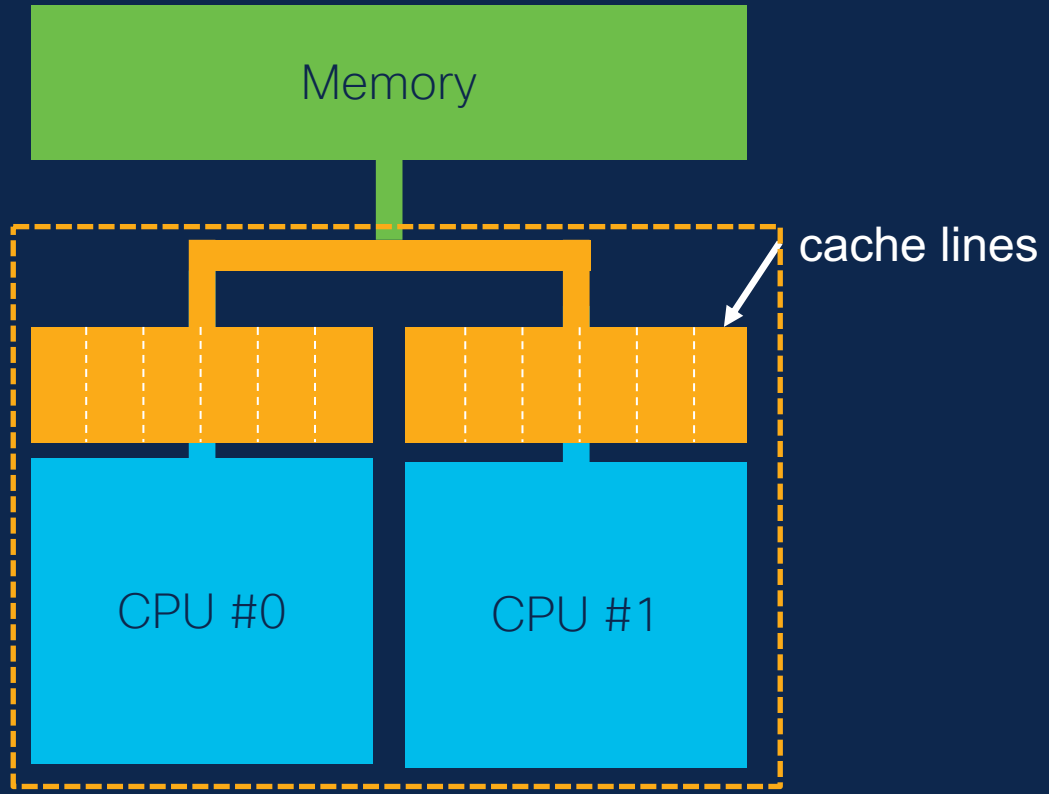


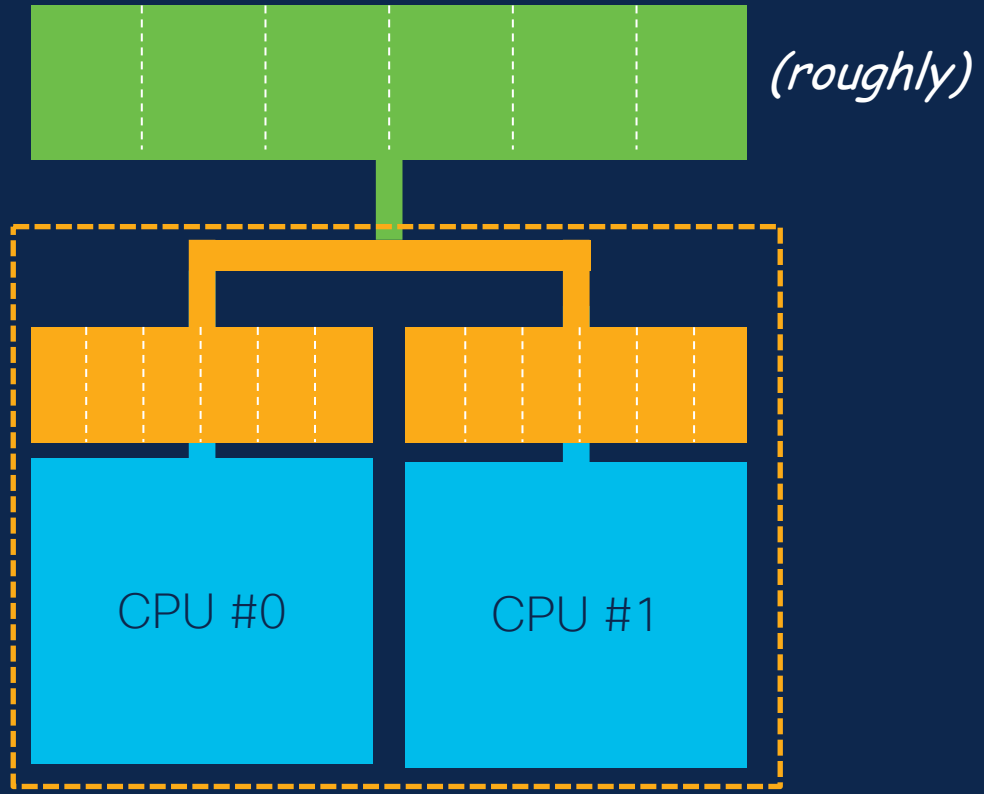


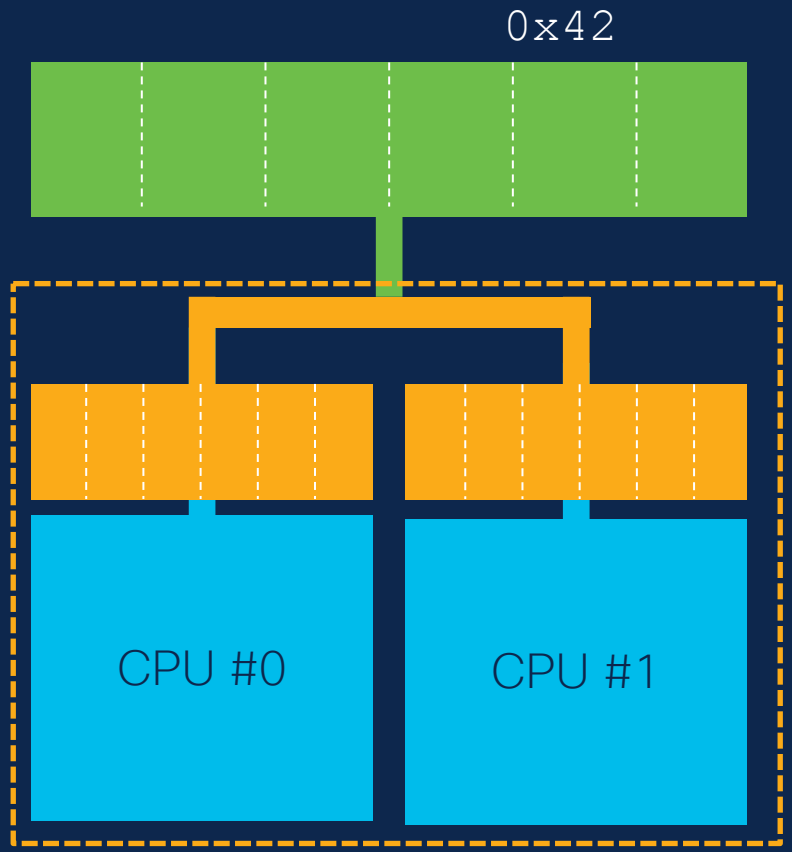
SMP: Symmetric Multiprocessing

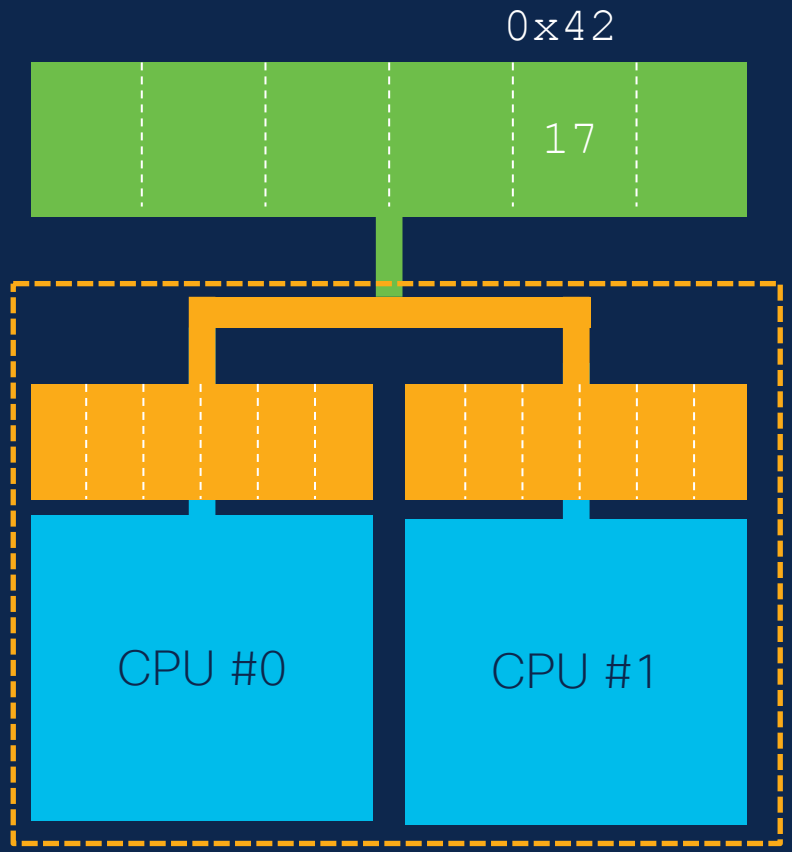


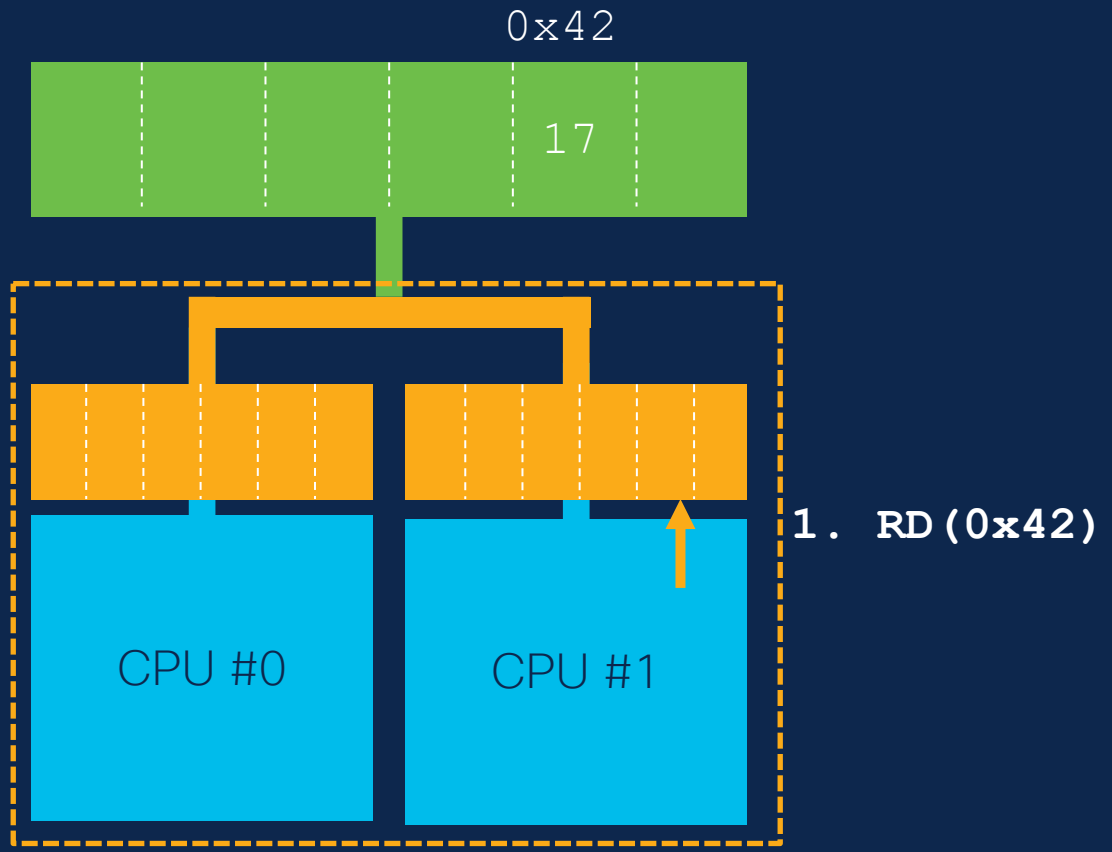


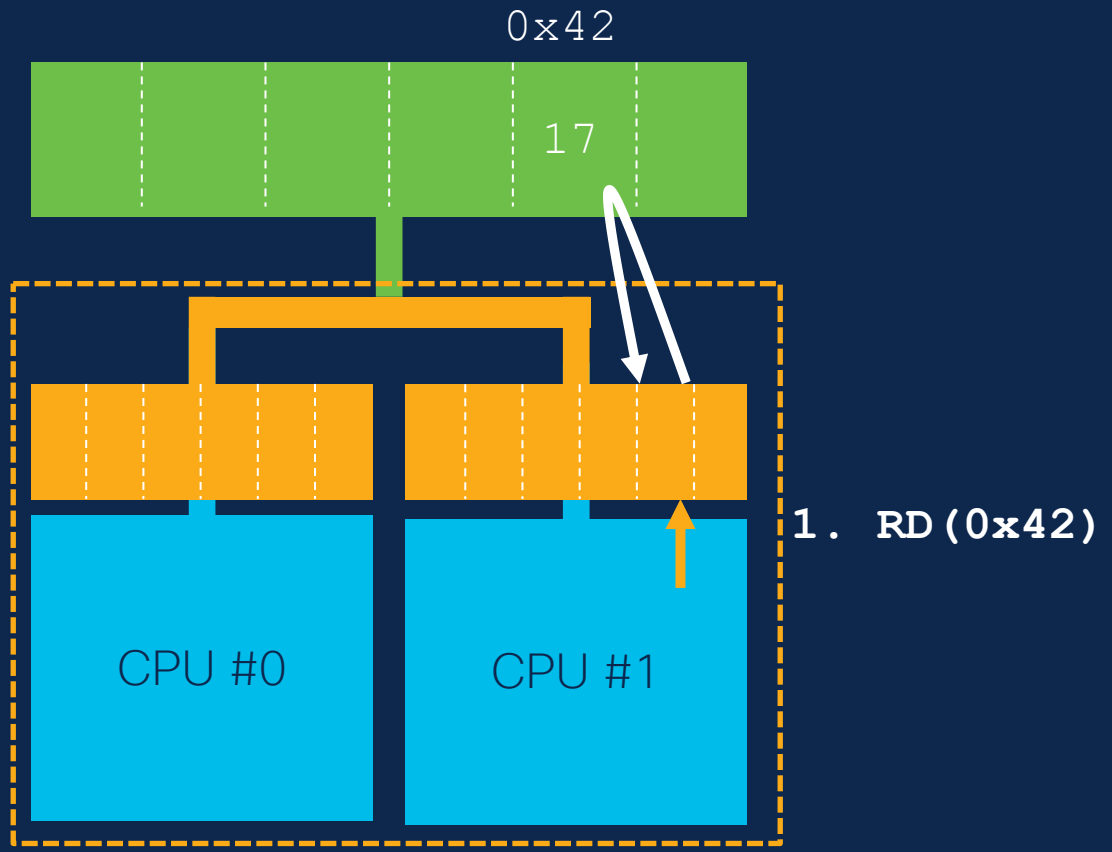


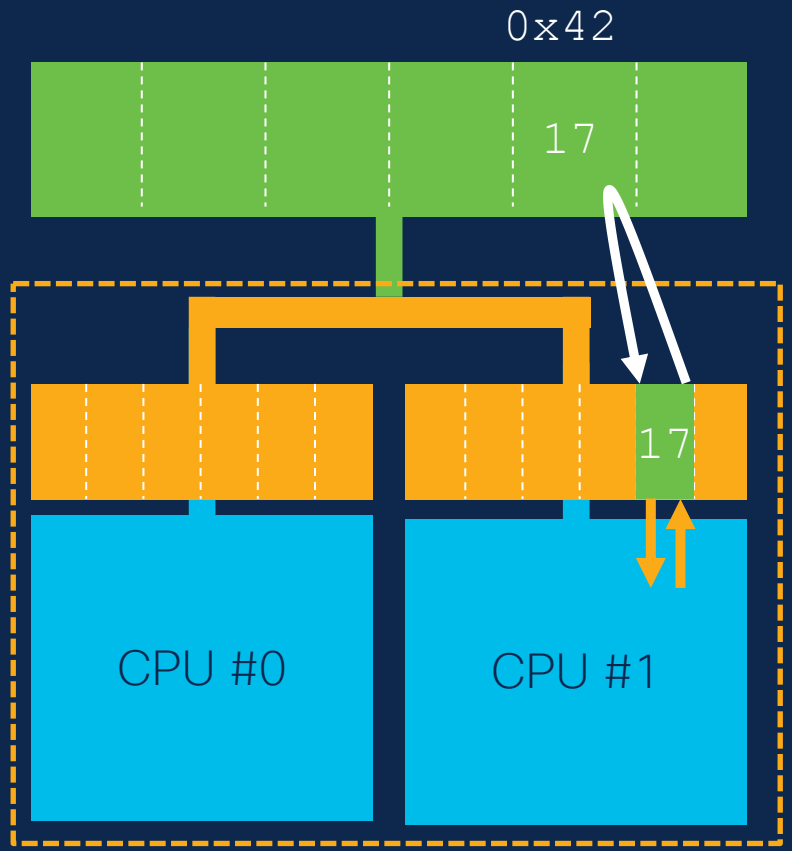






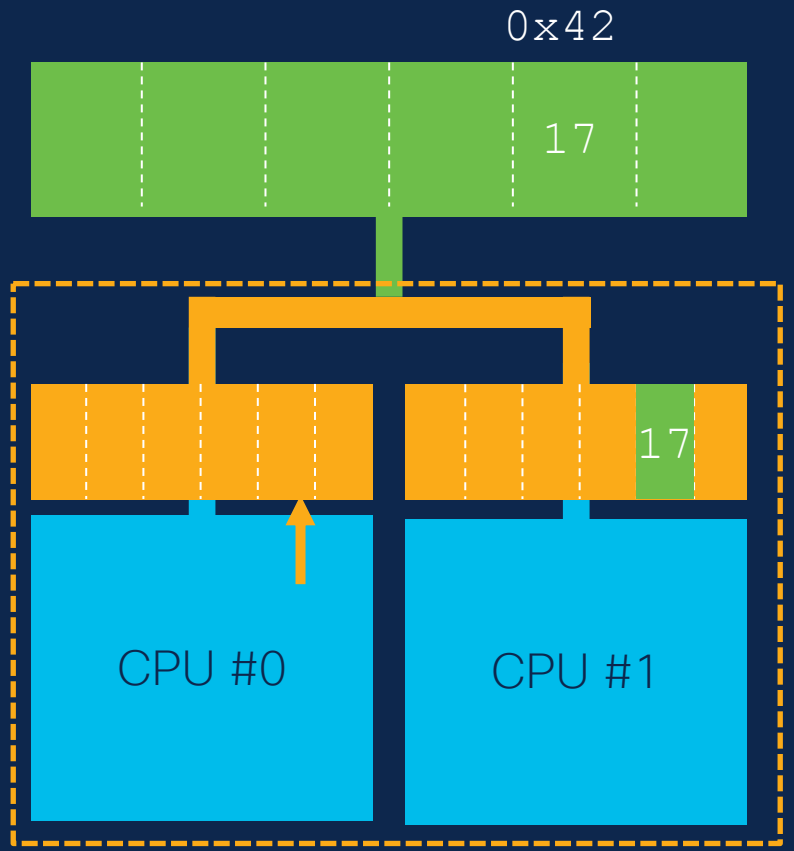






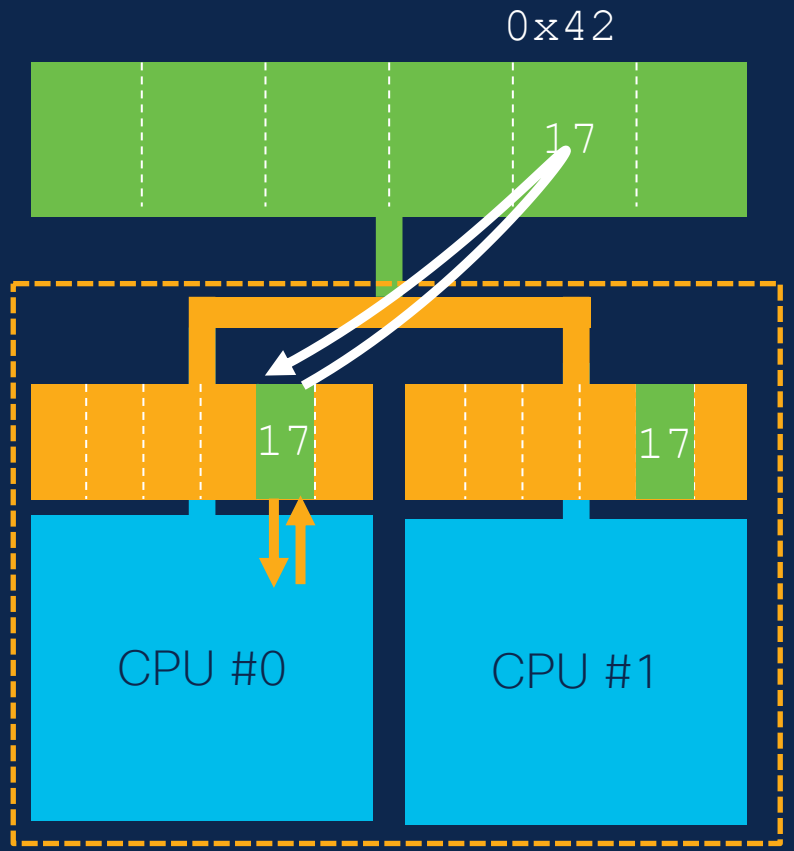
1. **RD(0x42)= 17**

1. RD(0x42) = 17



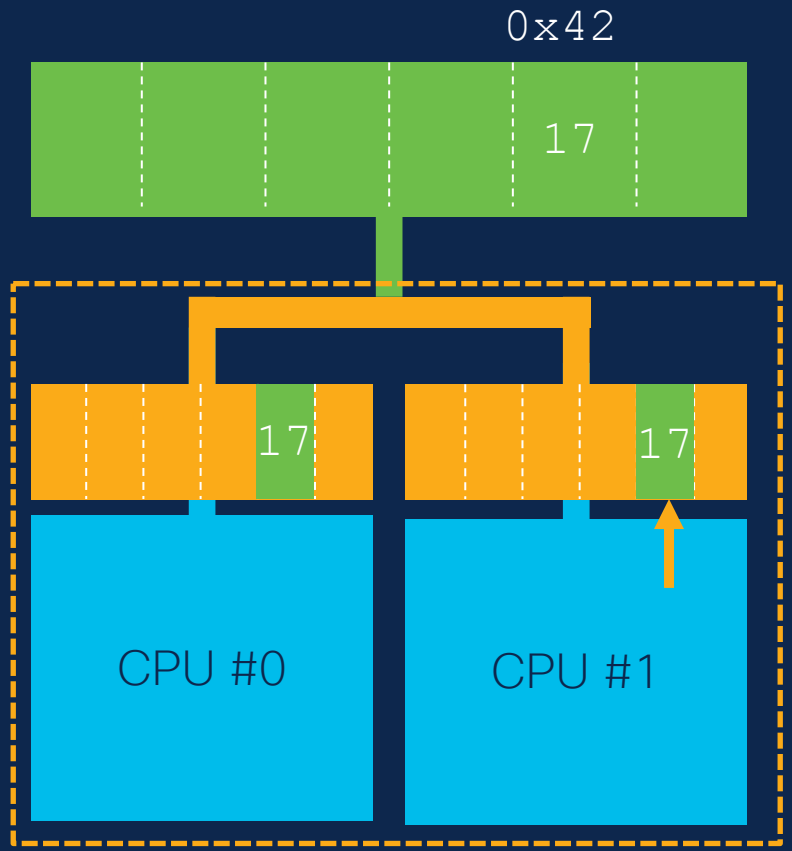
1. RD(0x42)= 17

1. RD(0x42) = 17



1. RD(0x42)= 17

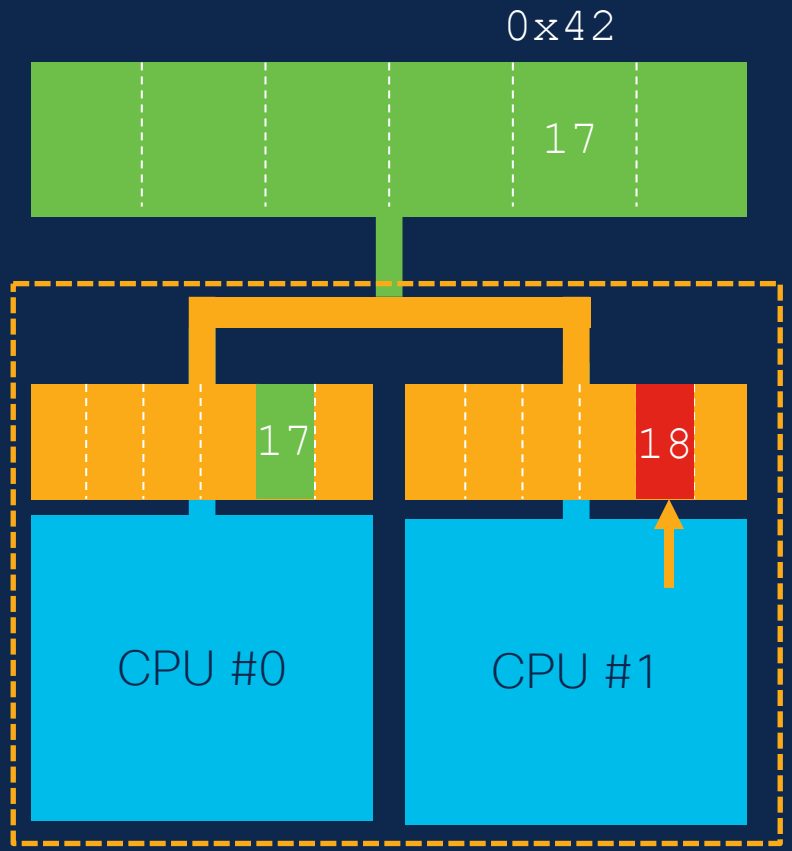
1. RD(0x42) = 17



1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

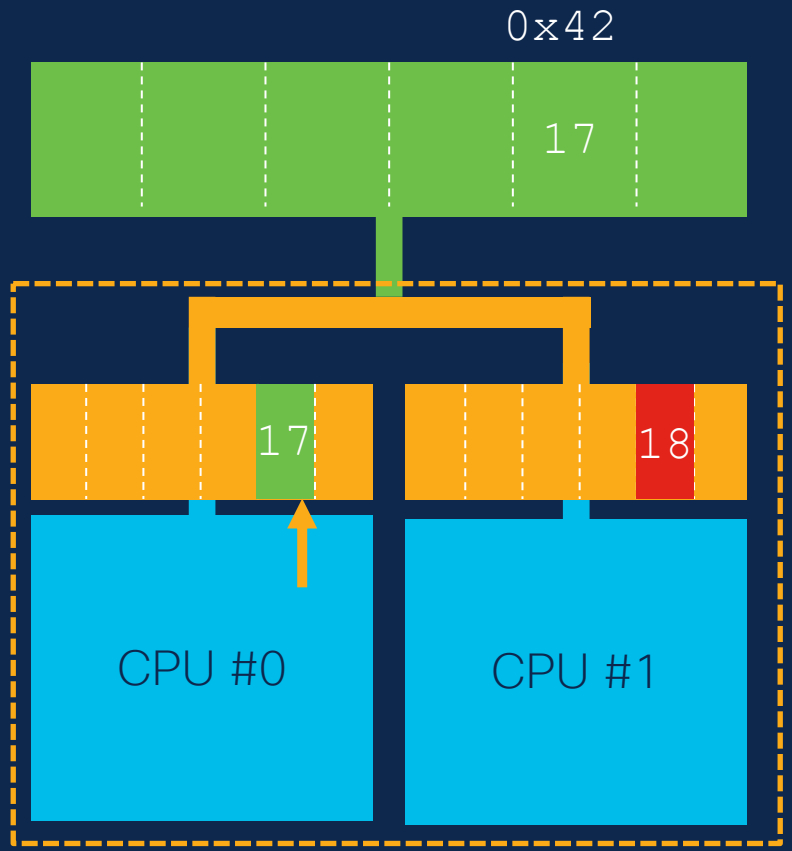


1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

2. RD(0x42)

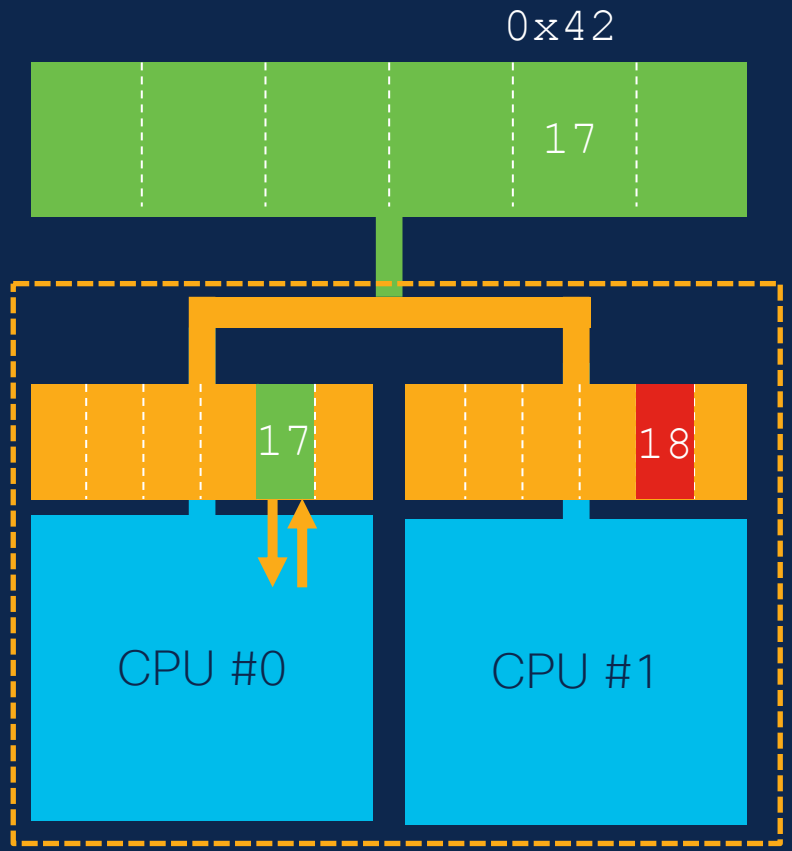


1. RD(0x42)= 17

2. WR(0x42)= 18

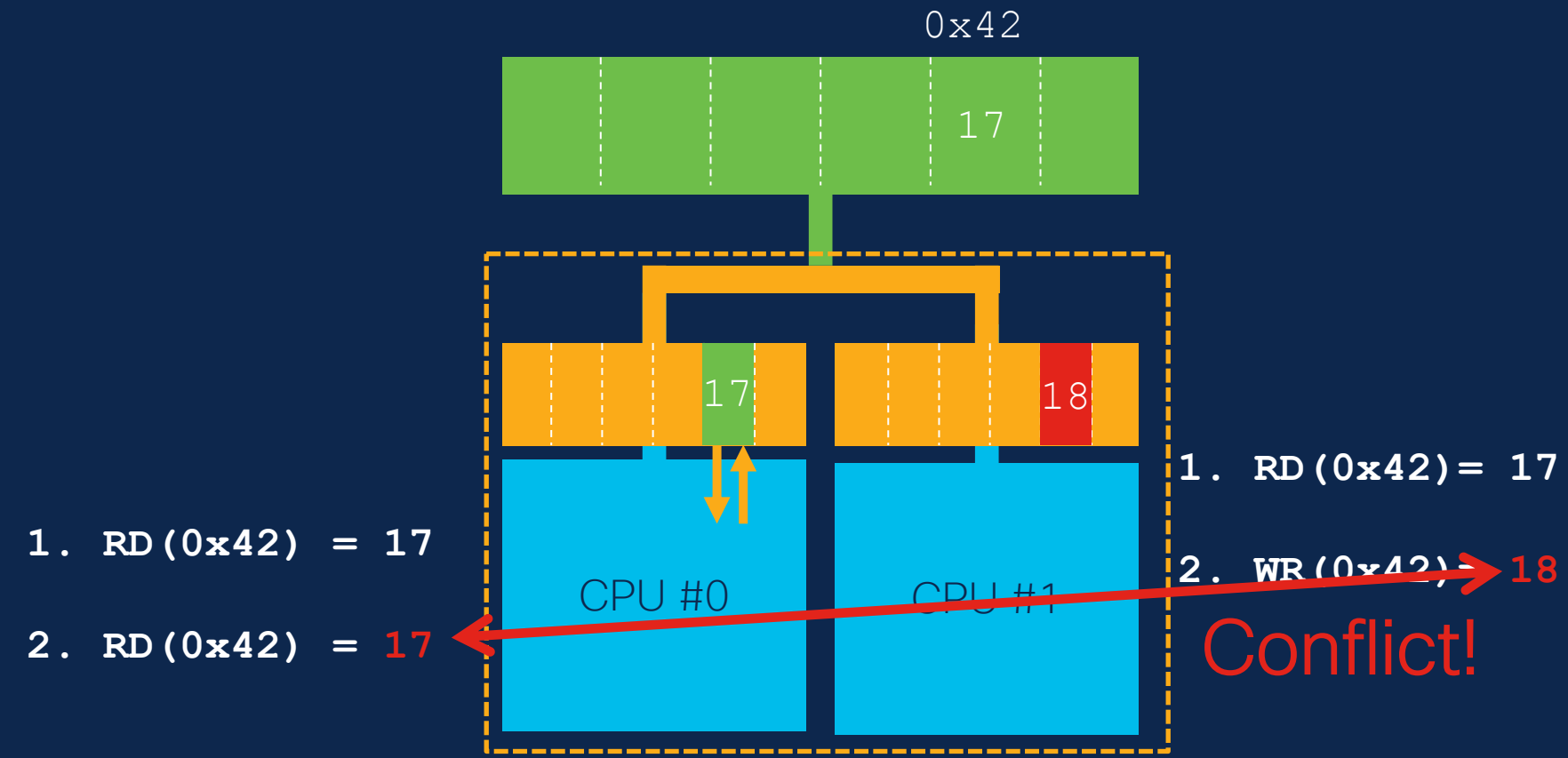
1. RD(0x42) = 17

2. RD(0x42) = 17



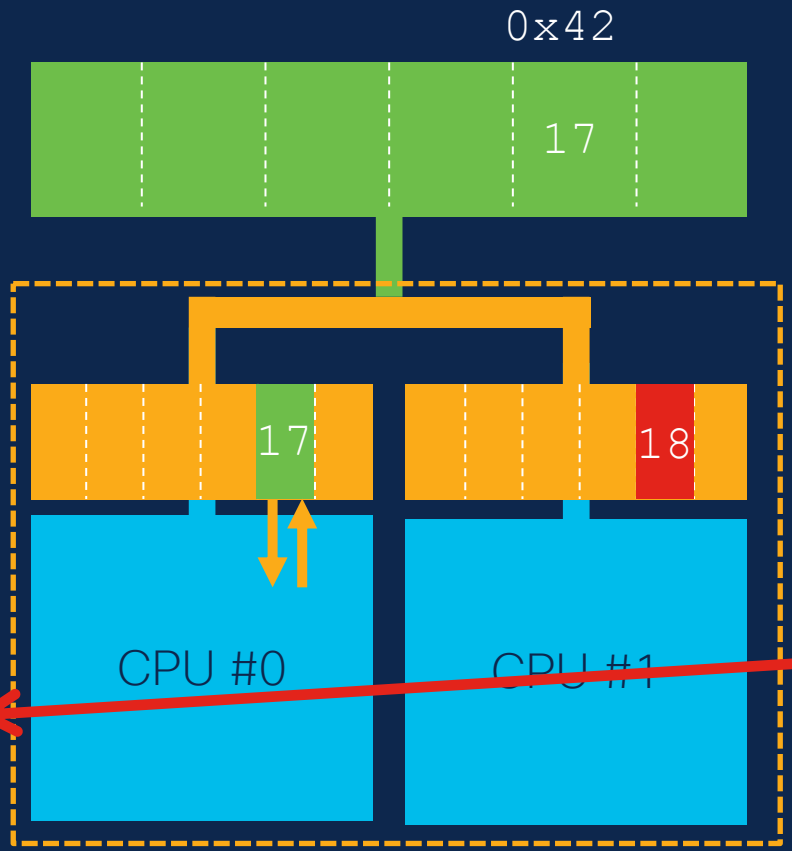
1. RD(0x42)= 17

2. WR(0x42)= 18



Cache Coherency:

Ensure that all caches have a consistent view of memory



1. RD(0x42) = 17

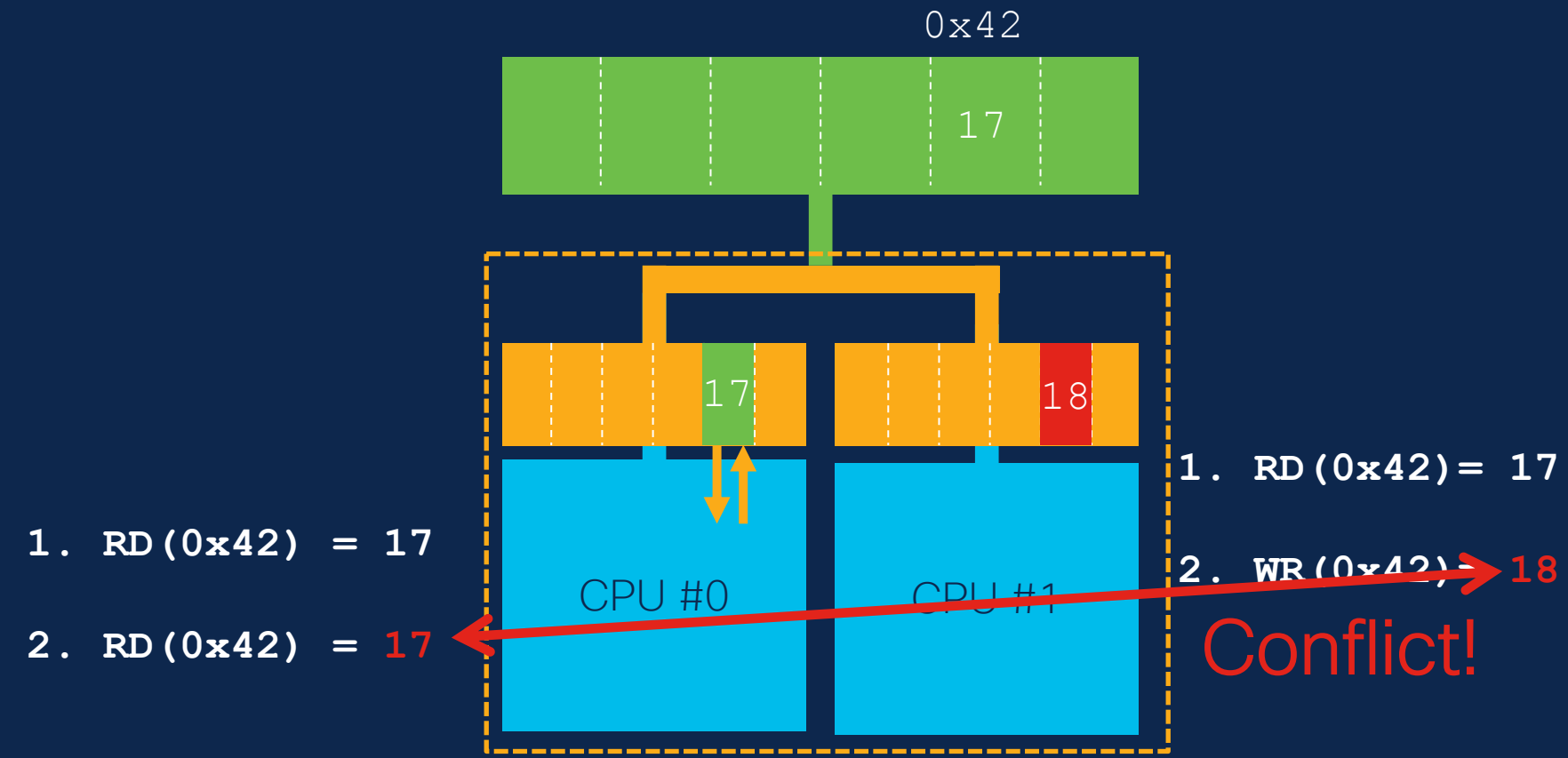
2. RD(0x42) = 17

1. RD(0x42)= 17

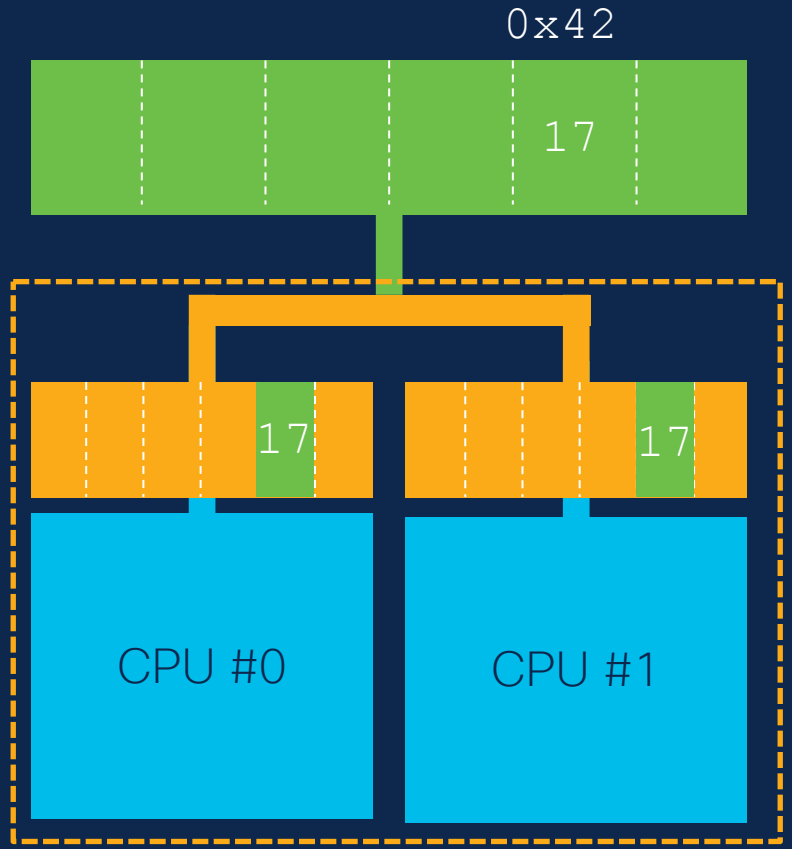
2. WR(0x42)= 18

Conflict!

A sort of half answer to the CC problem…

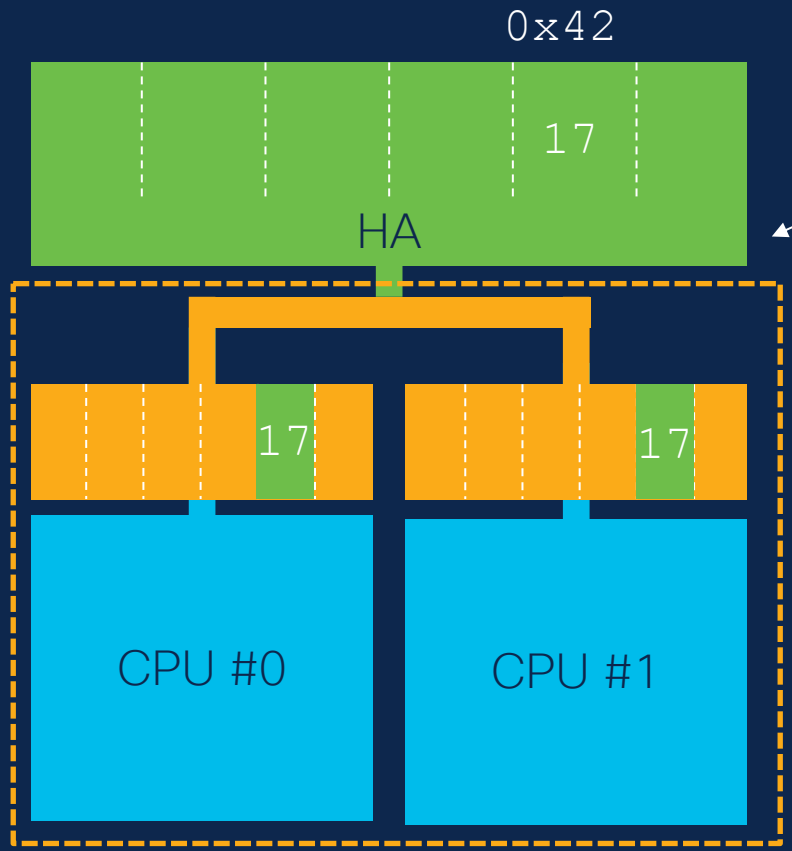


1. RD(0x42) = 17



1. RD(0x42)= 17

2. WR(0x42)= 18

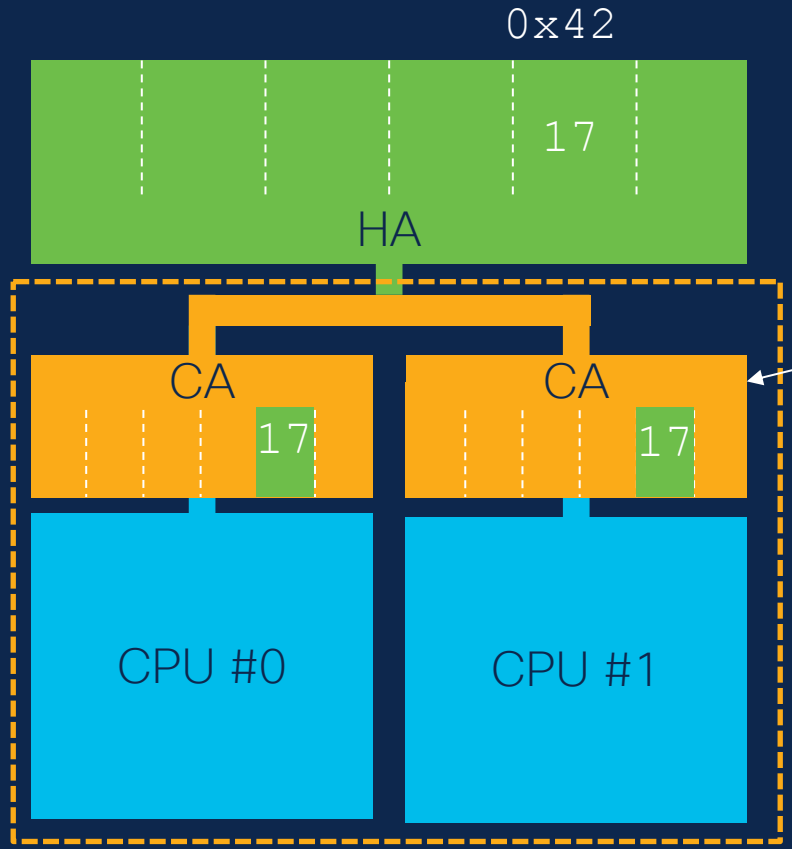


1. **RD(0x42) = 17**

1. **RD(0x42)= 17**

2. **WR(0x42)= 18**

1. RD(0x42) = 17

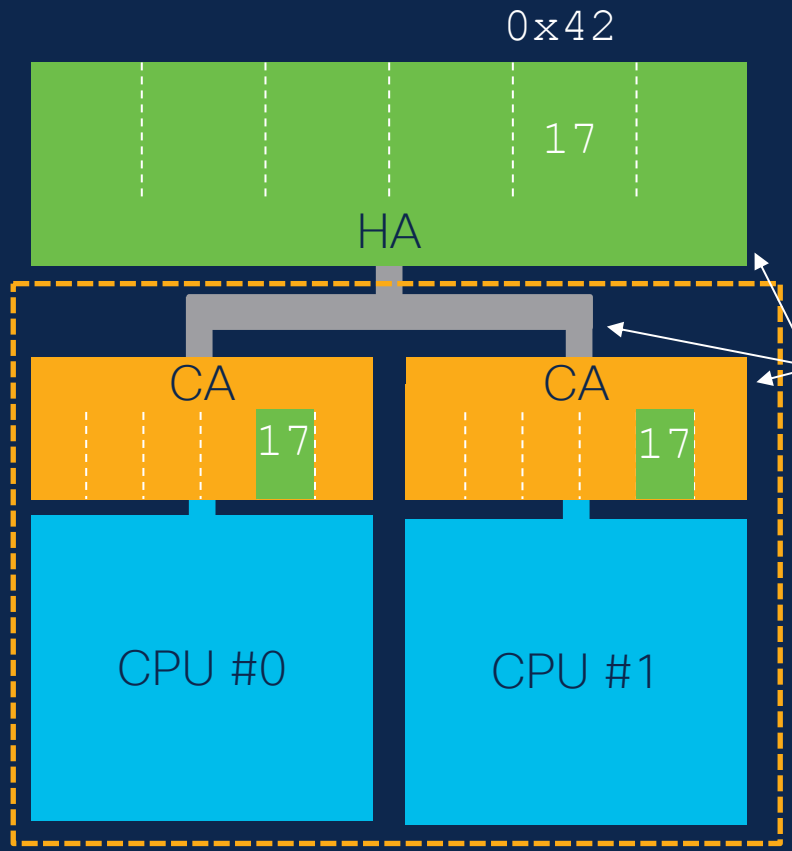


CA: Cache Agent

1. RD(0x42)= 17

2. WR(0x42)= 18

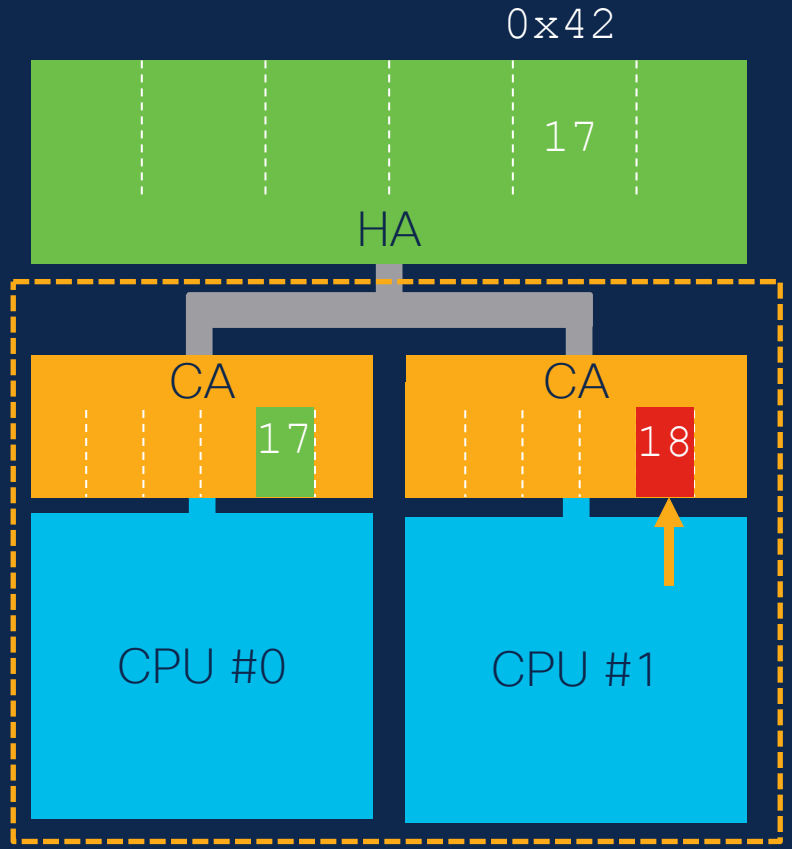
1. RD(0x42) = 17



1. RD(0x42)= 17

2. WR(0x42)= 18

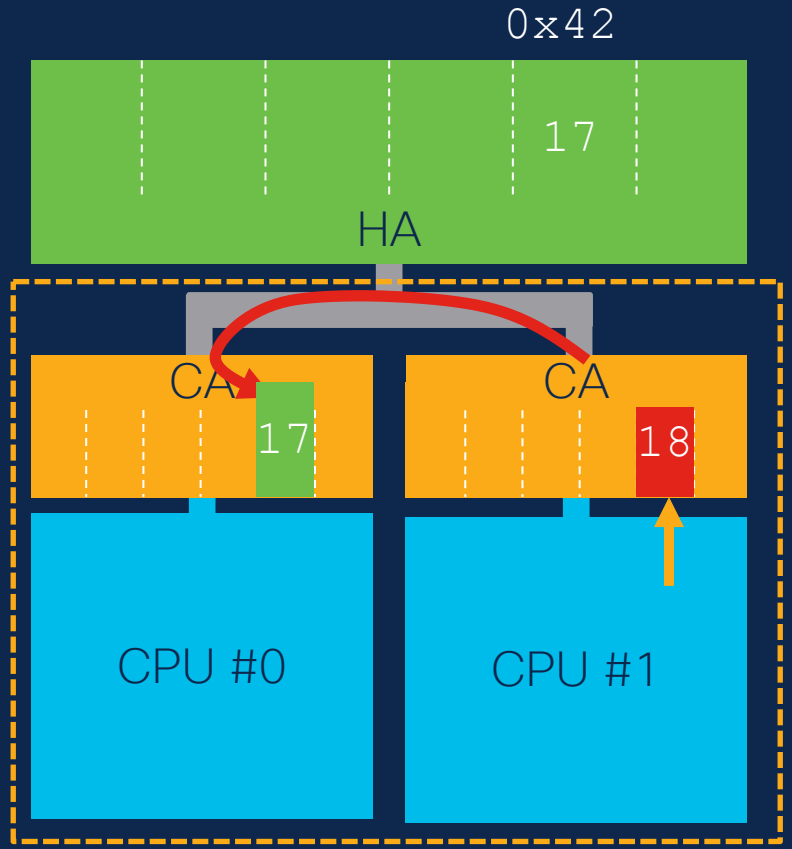
1. RD(0x42) = 17



1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

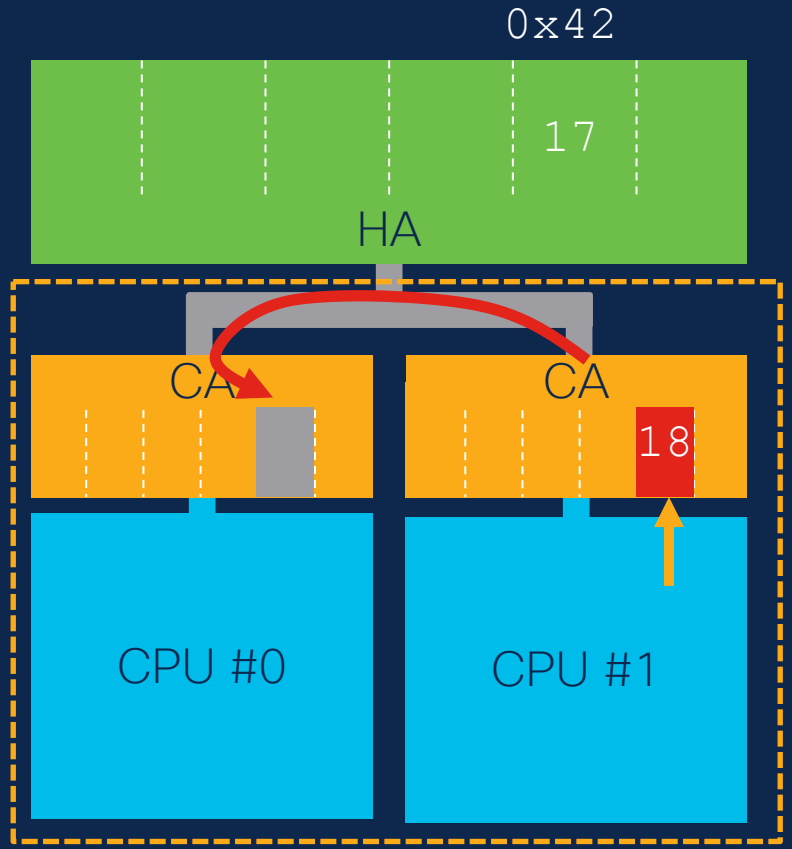


Invalidation msg

1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

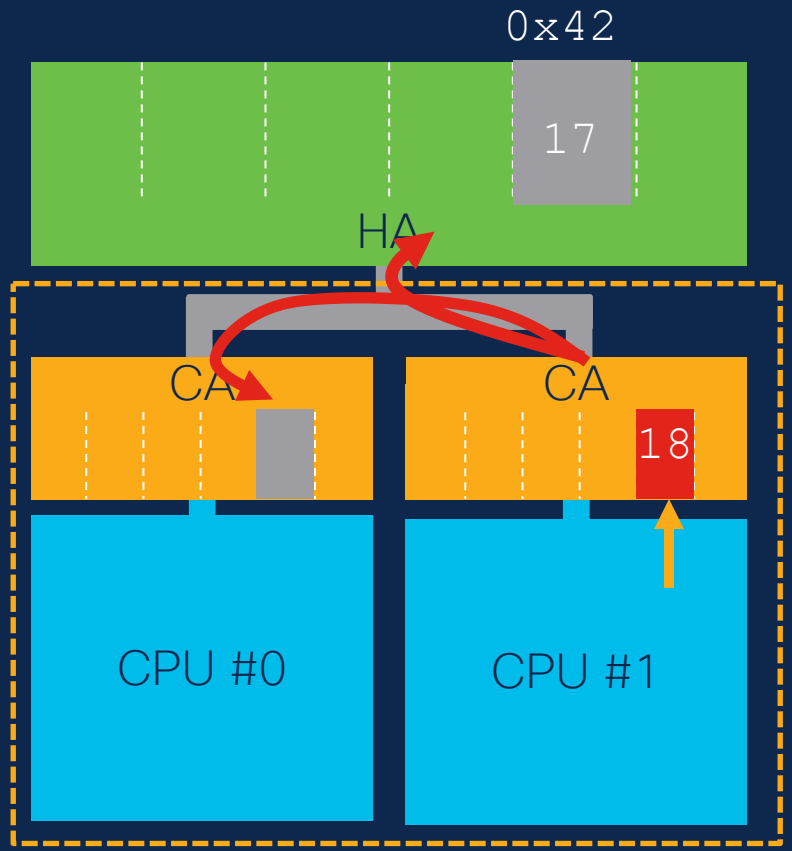


Invalidation msg

1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17



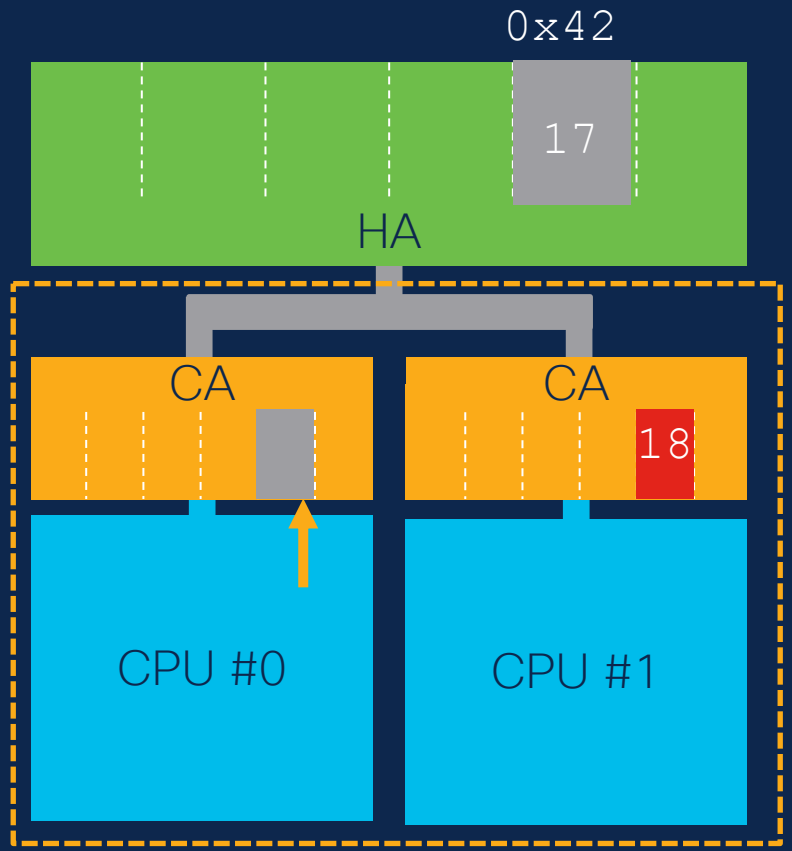
Invalidation msg

1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

2. RD(0x42)

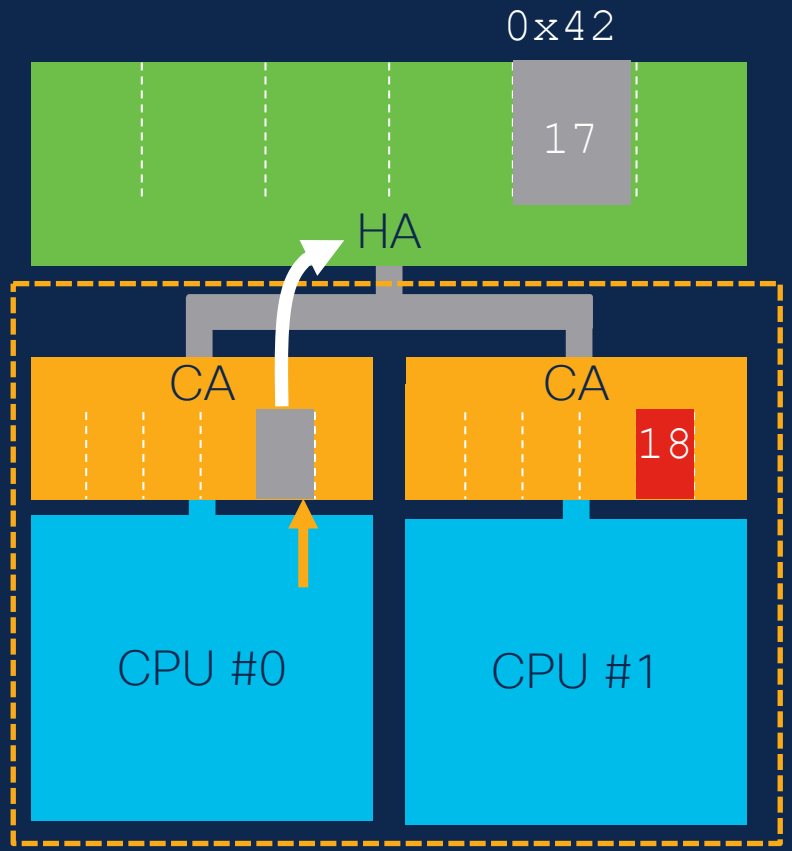


1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

2. RD(0x42)

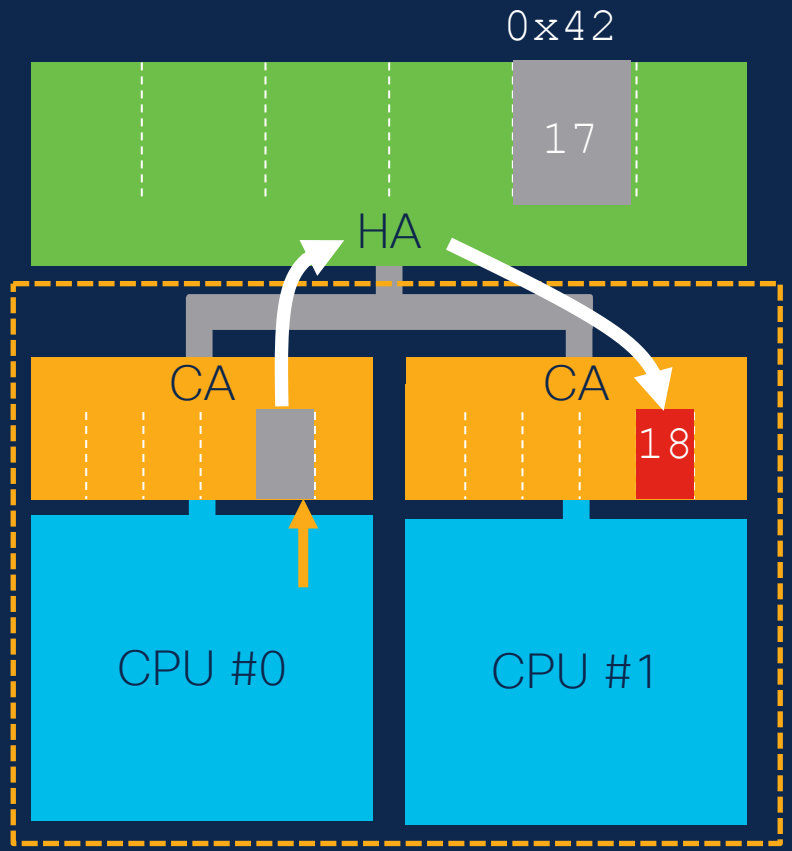


1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

2. RD(0x42)

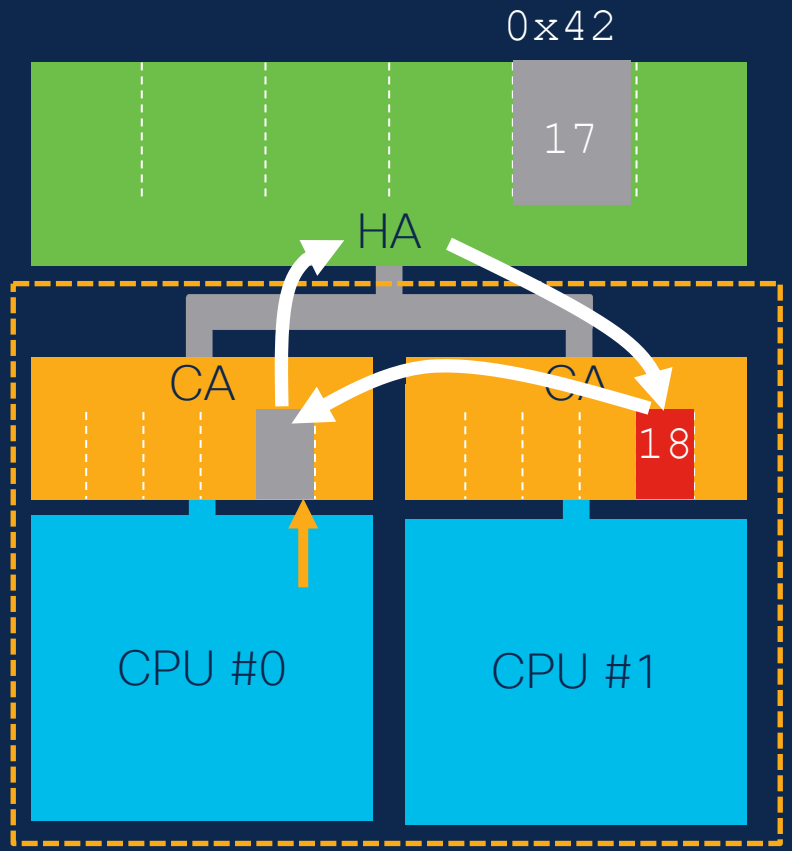


1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

2. RD(0x42)

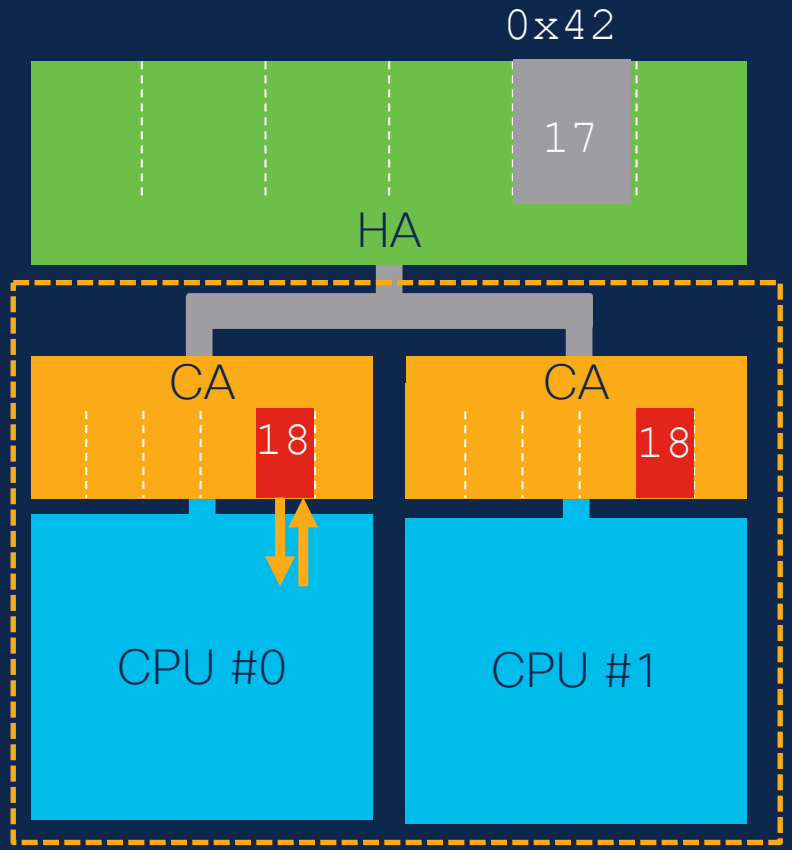


1. RD(0x42)= 17

2. WR(0x42)= 18

1. RD(0x42) = 17

2. RD(0x42)

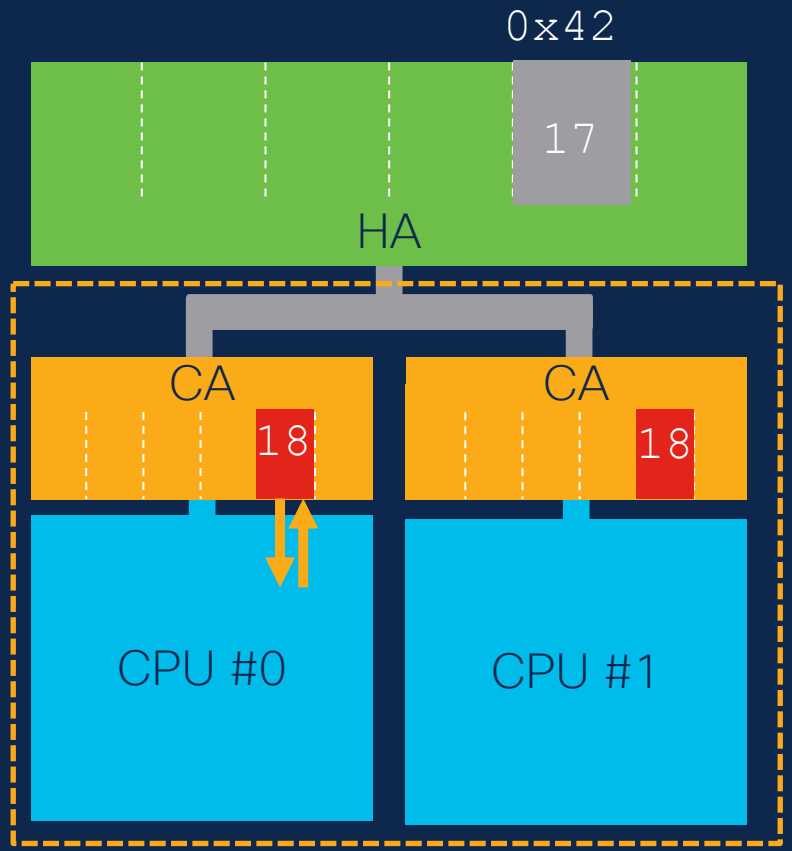


1. RD(0x42)= 17

2. WR(0x42)= 18

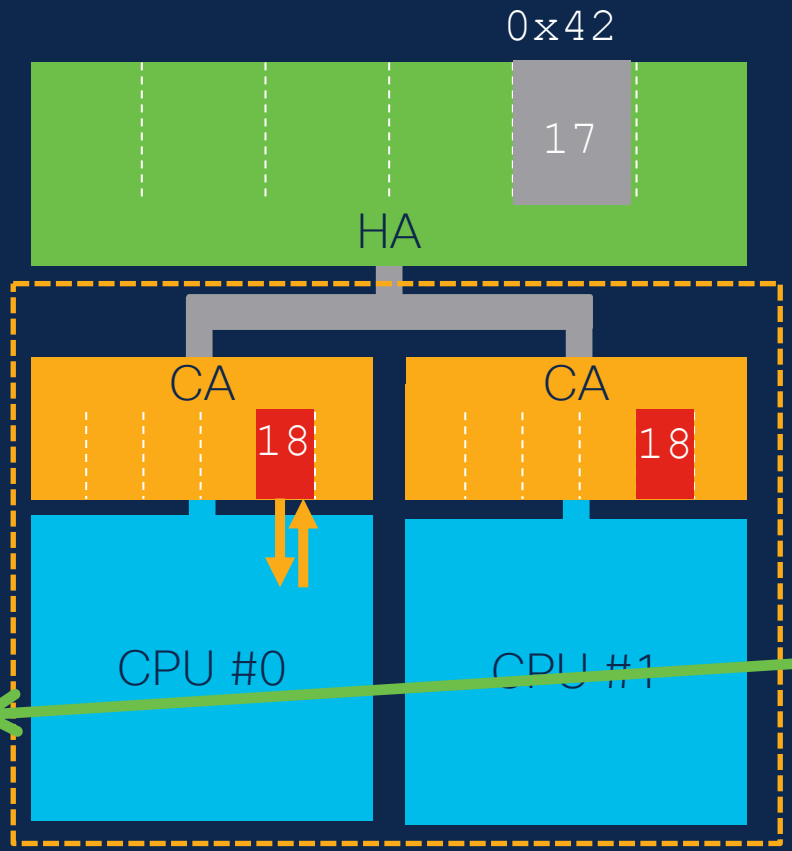
1. RD(0x42) = 17

2. RD(0x42) = 18



1. RD(0x42)= 17

2. WR(0x42)= 18



1. RD(0x42) = 17

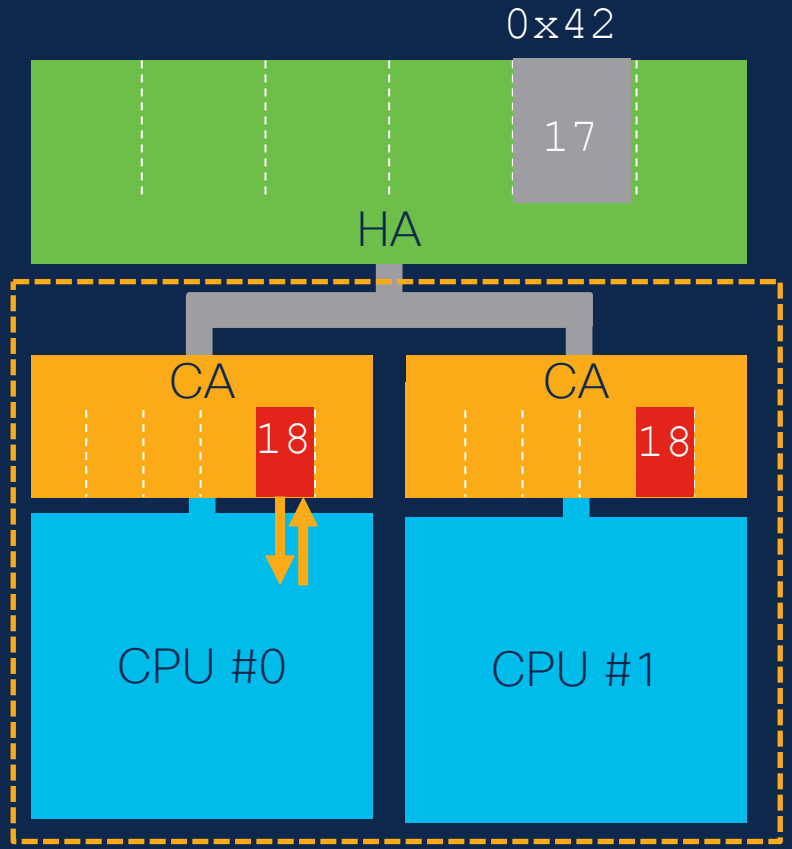
2. RD(0x42) = 18

1. RD(0x42)= 17

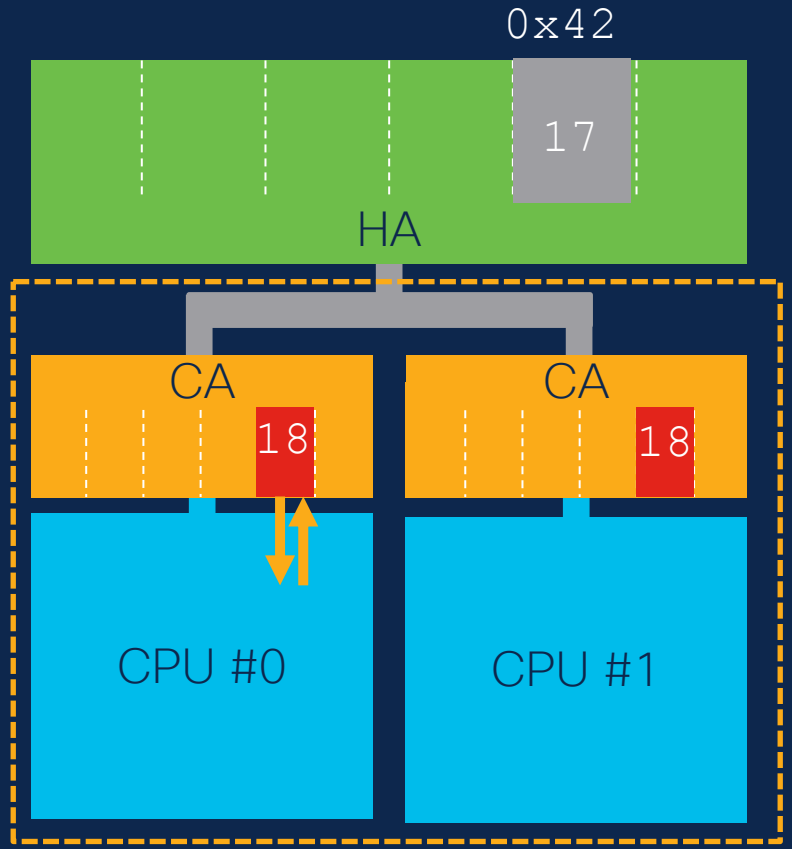
2. WR(0x42) = 18

✔ Coherent

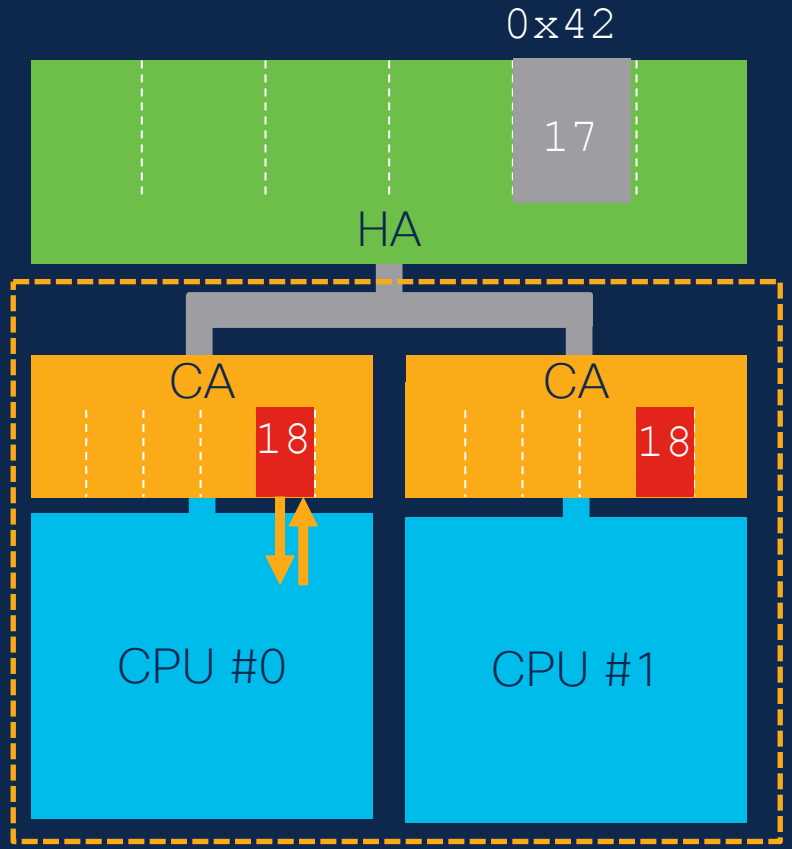
- Cache coherency implementations are *complicated*

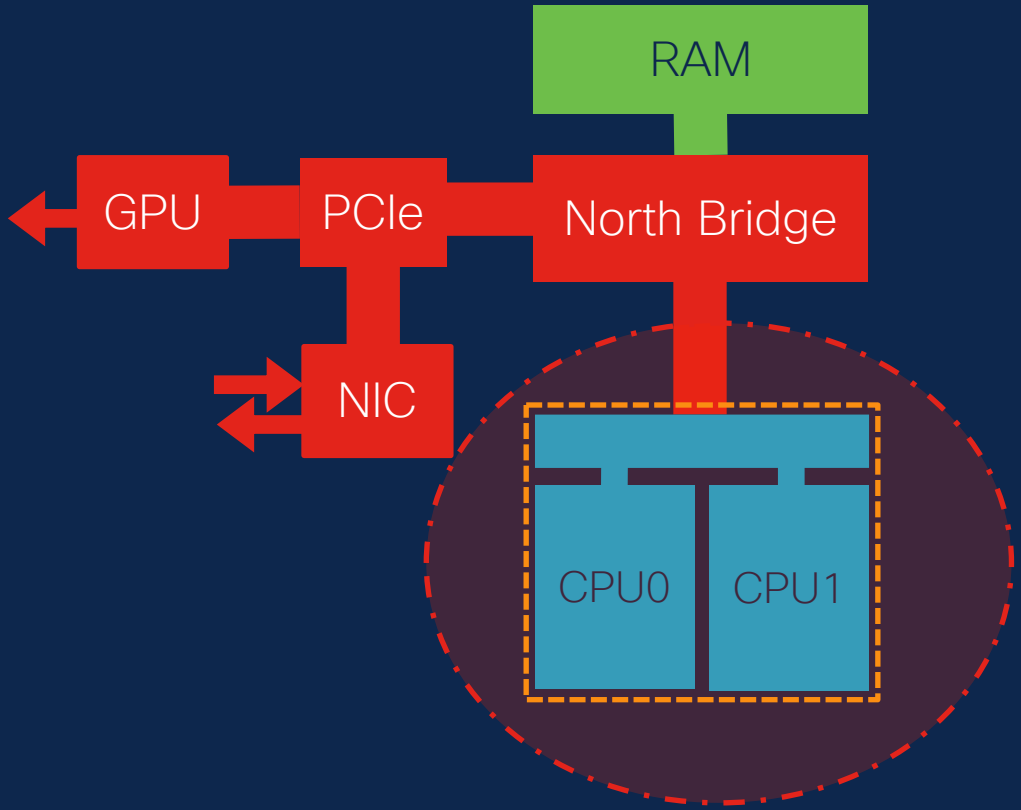


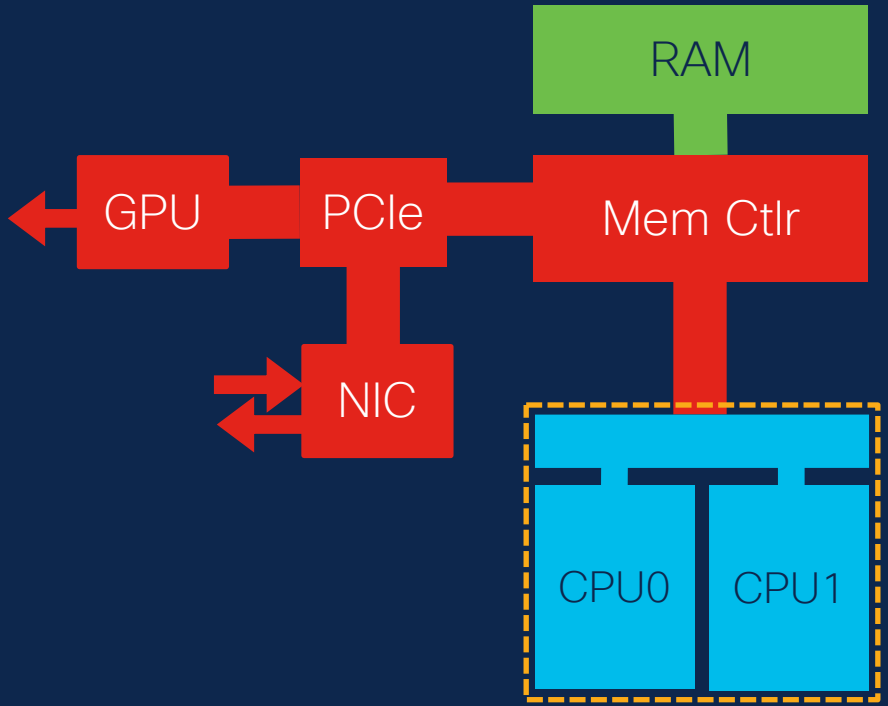
- Cache coherency implementations are *complicated*
- *T*ypically some sort of “MESI” protocol.

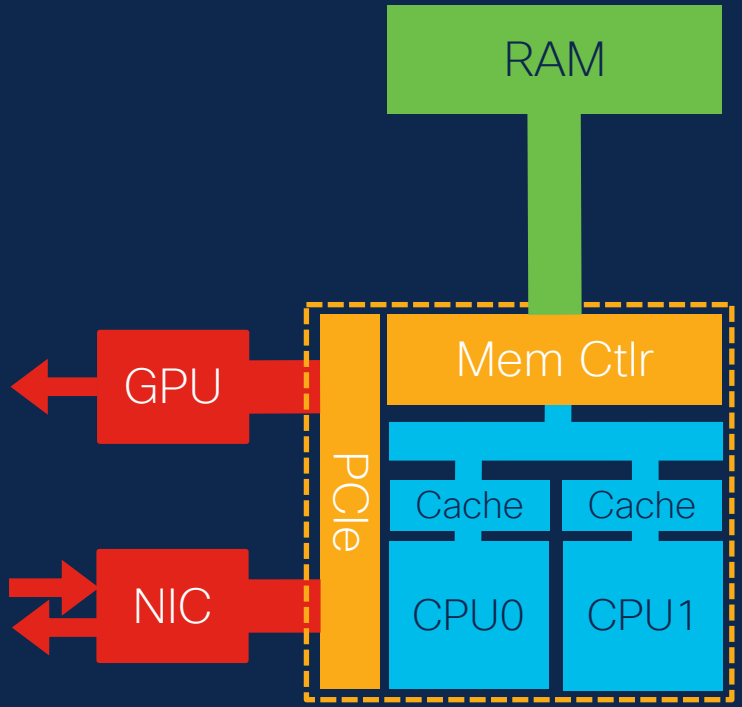


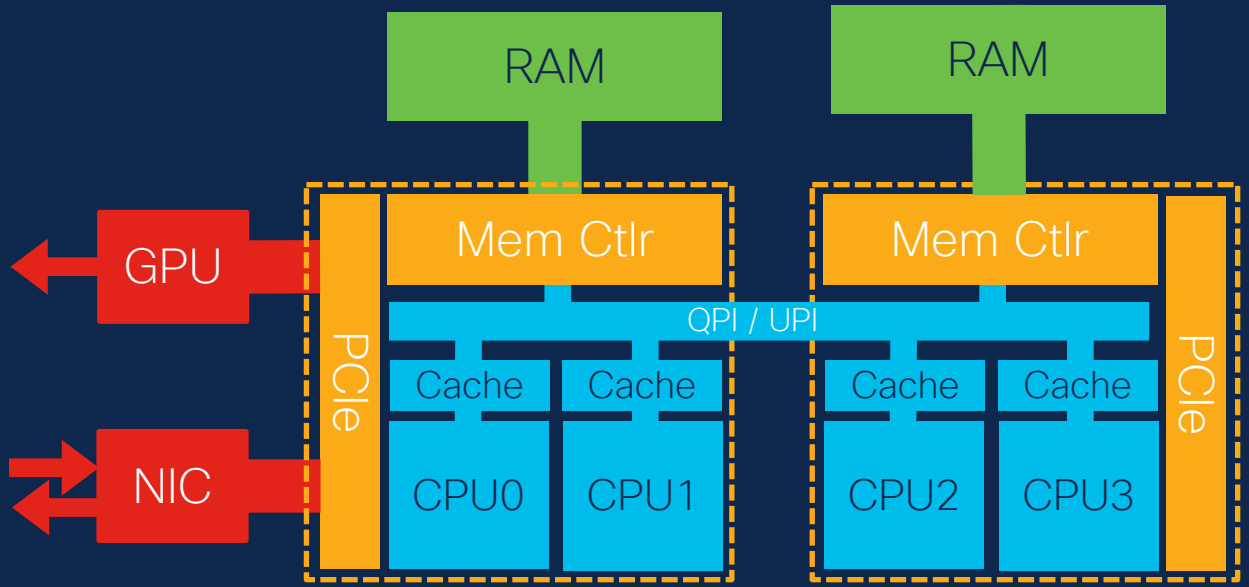
- Cache coherency implementations are *complicated*
- *T*ypically some sort of “MESI” protocol.
- Memory and caches need to be aware of the protocol (i.e. require CA/HA)





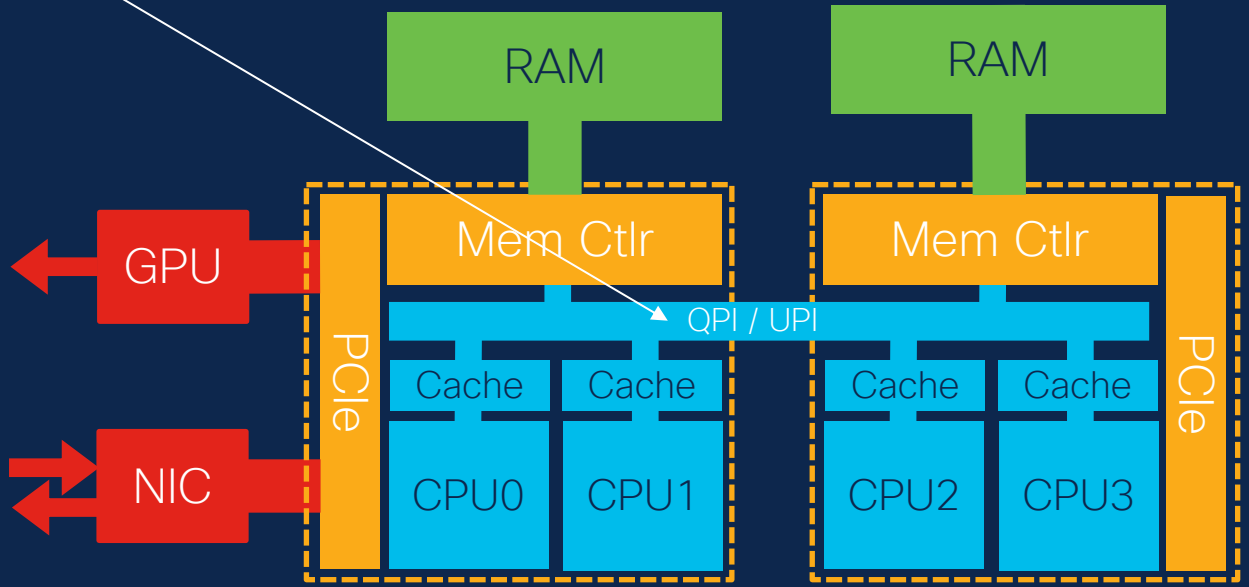




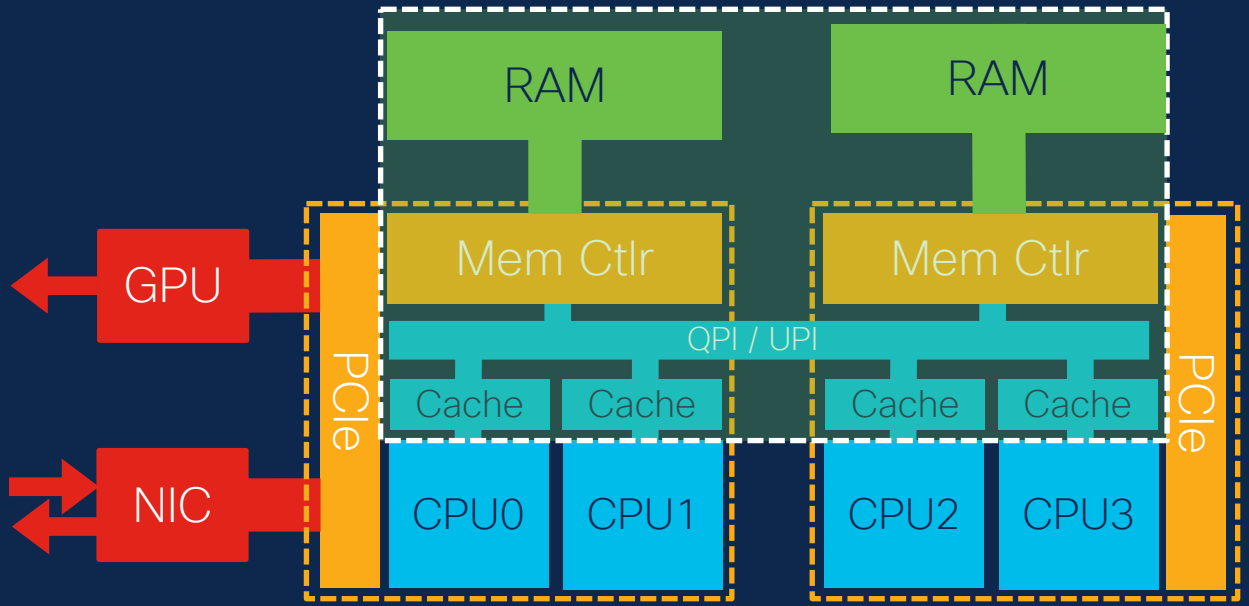


NUMA: Non-uniform memory access

QPI: Quick Path Interconnect
UPI: Ultra Path Interconnect

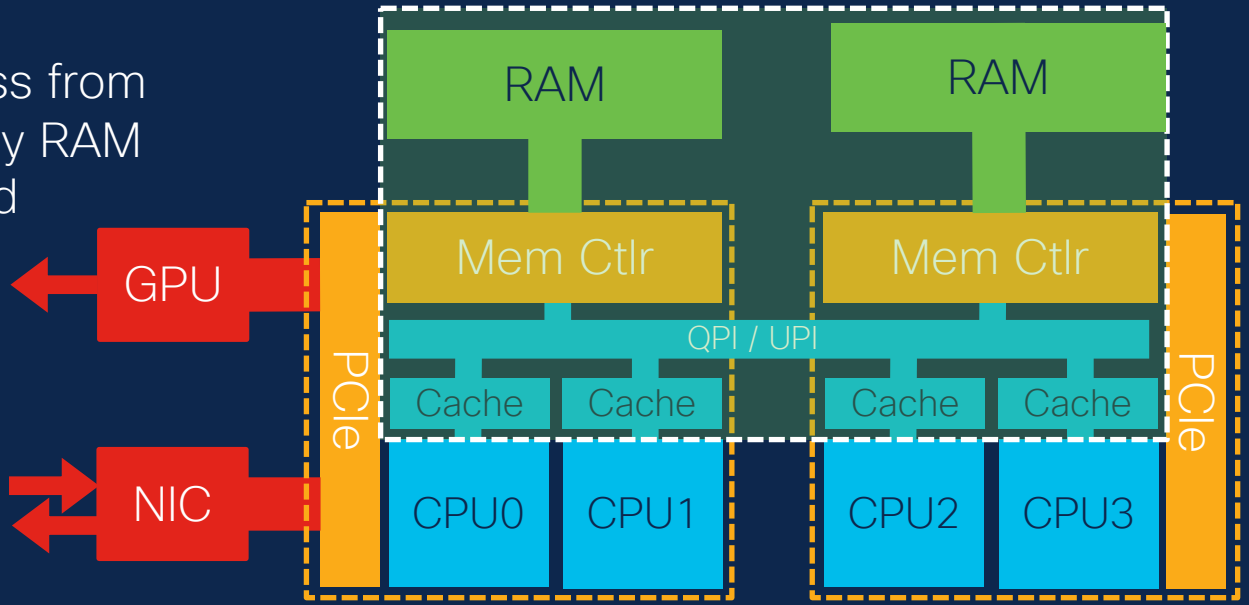


Cache Coherency Domain



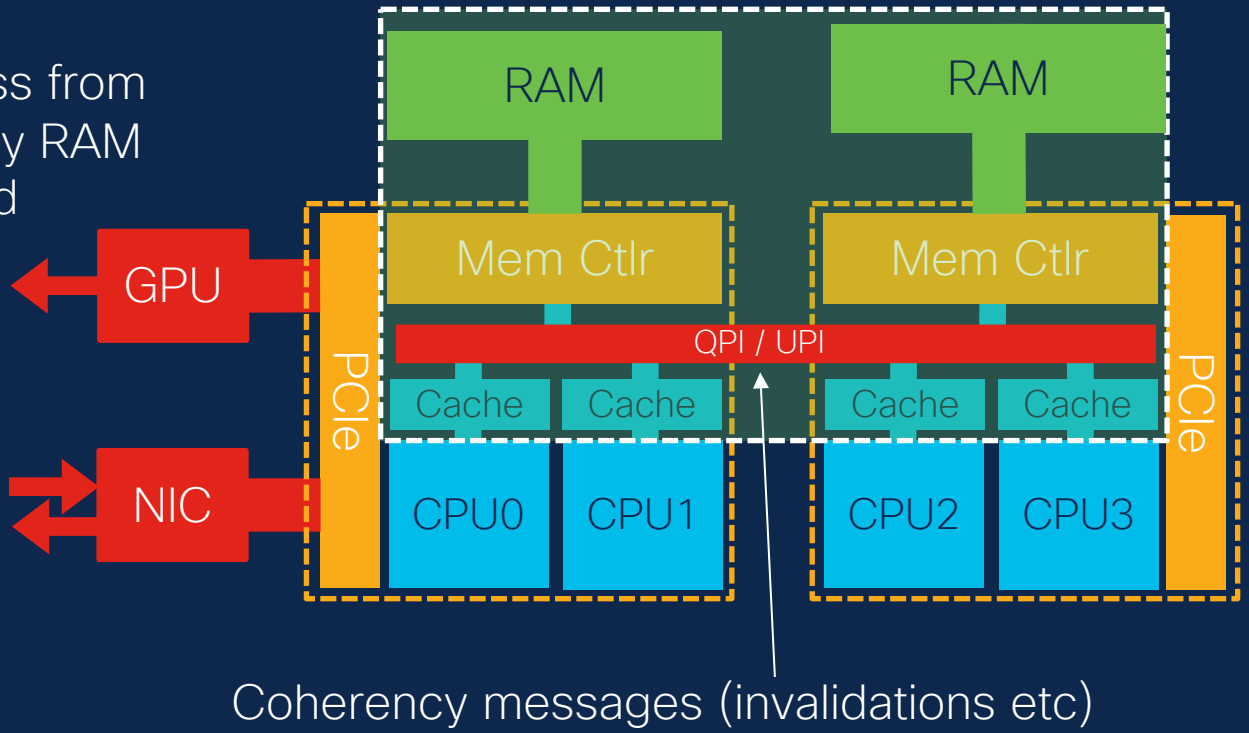
Cache Coherency Domain

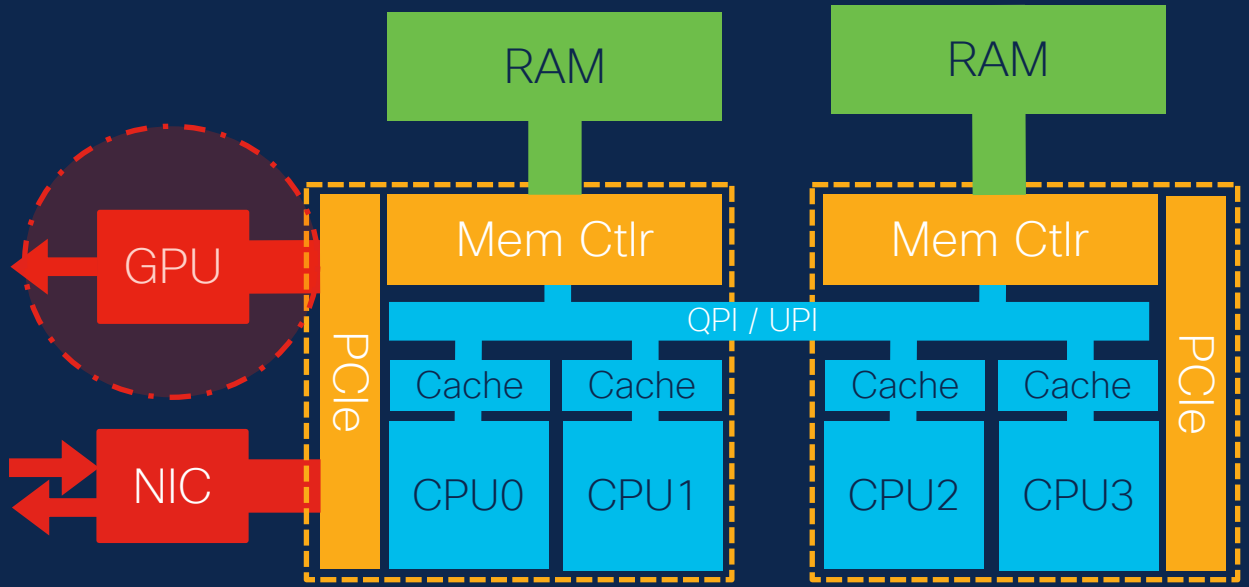
Memory access from any CPU to any RAM is implemented consistently



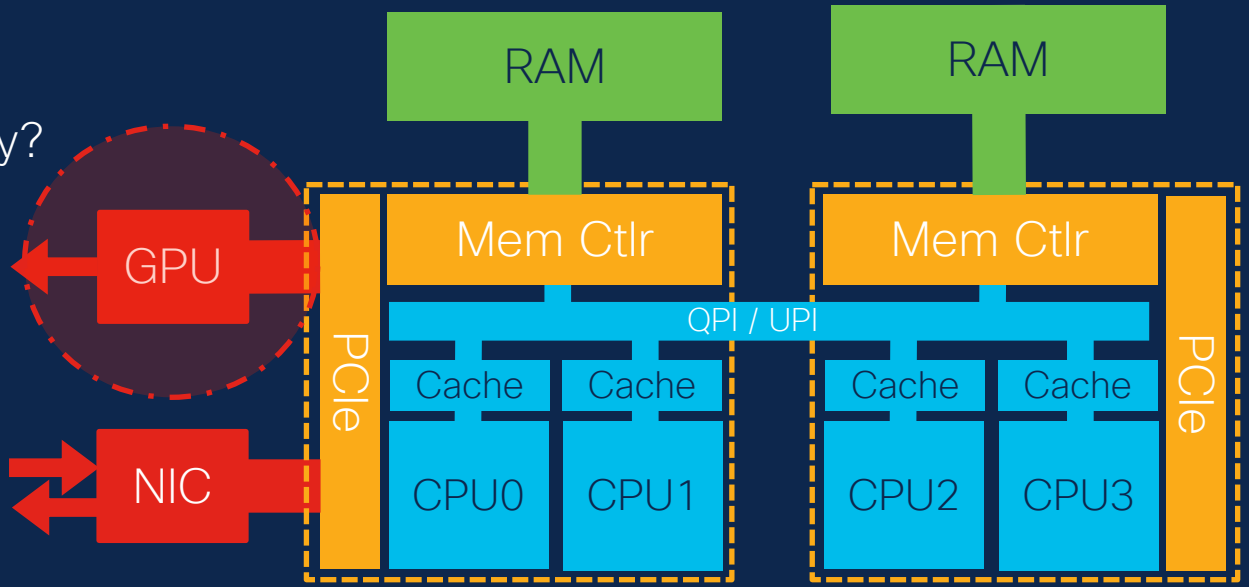
Cache Coherency Domain

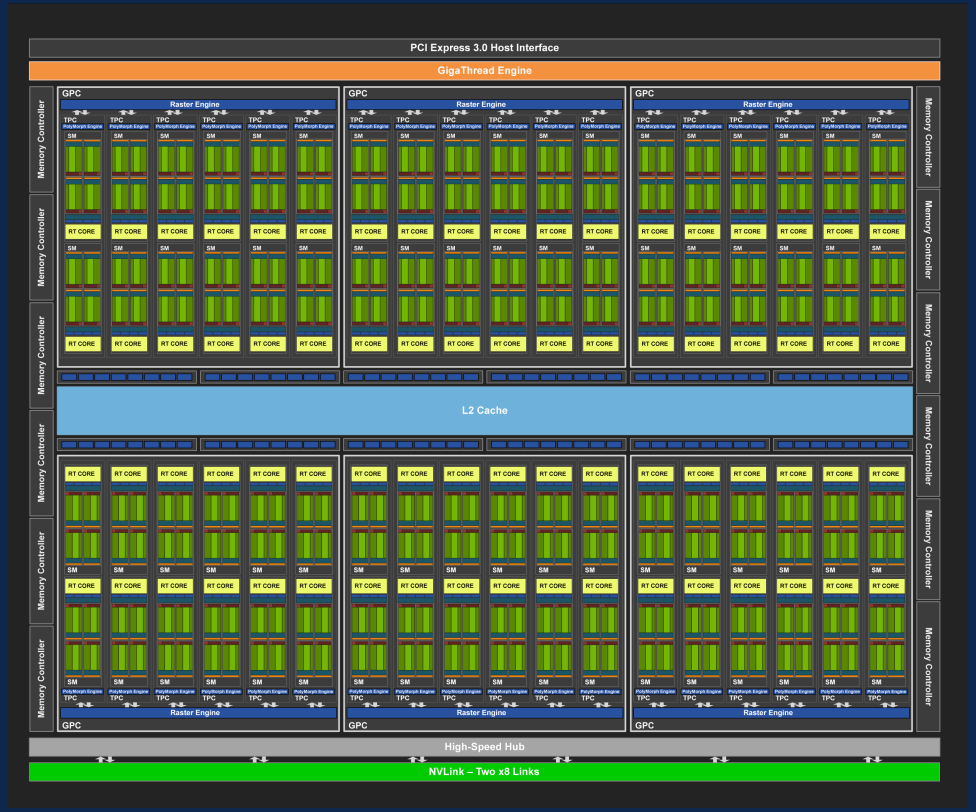
Memory access from any CPU to any RAM is implemented consistently





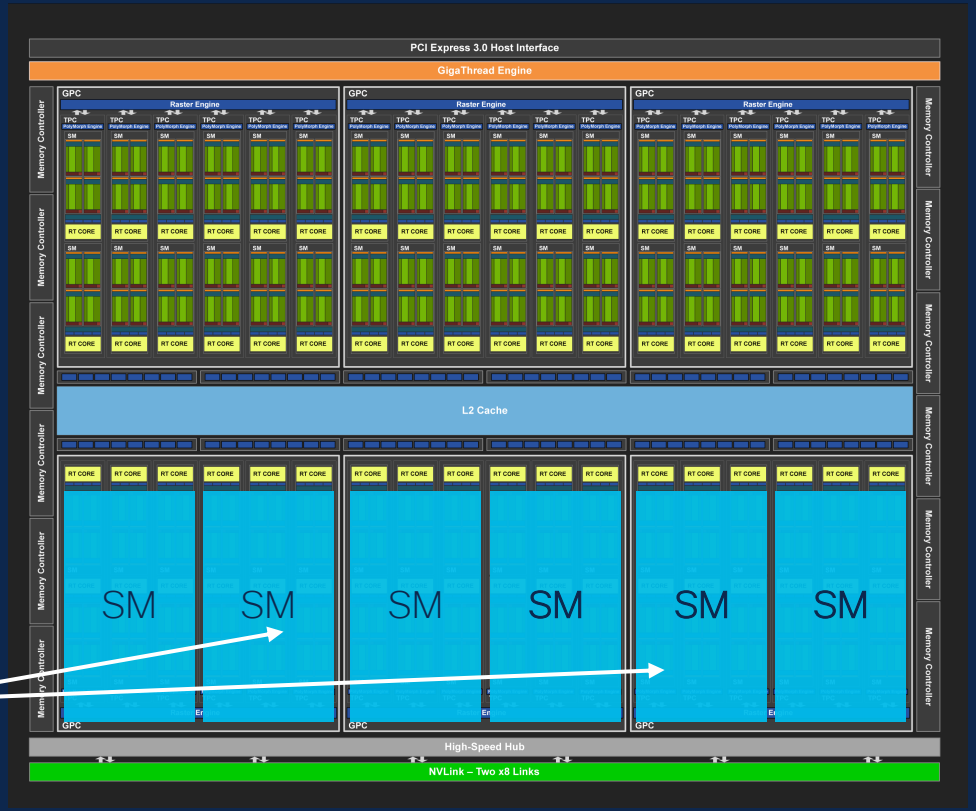
Graphics only?





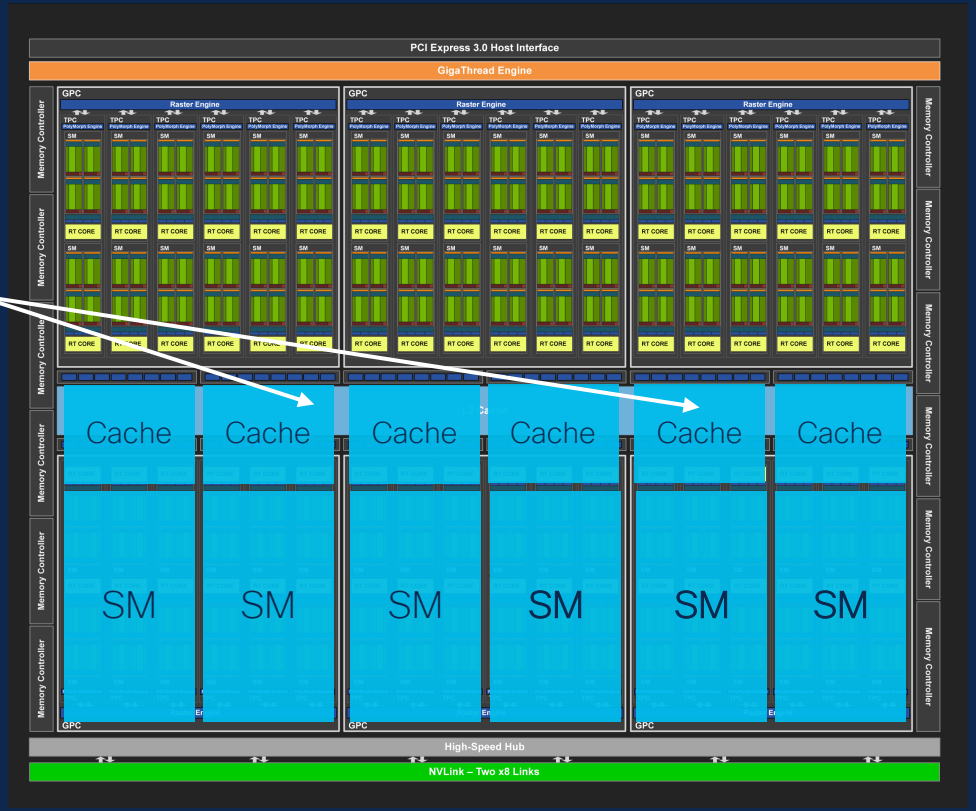
NVIDIA Turing GPU Architecture

Streaming Multiprocessors

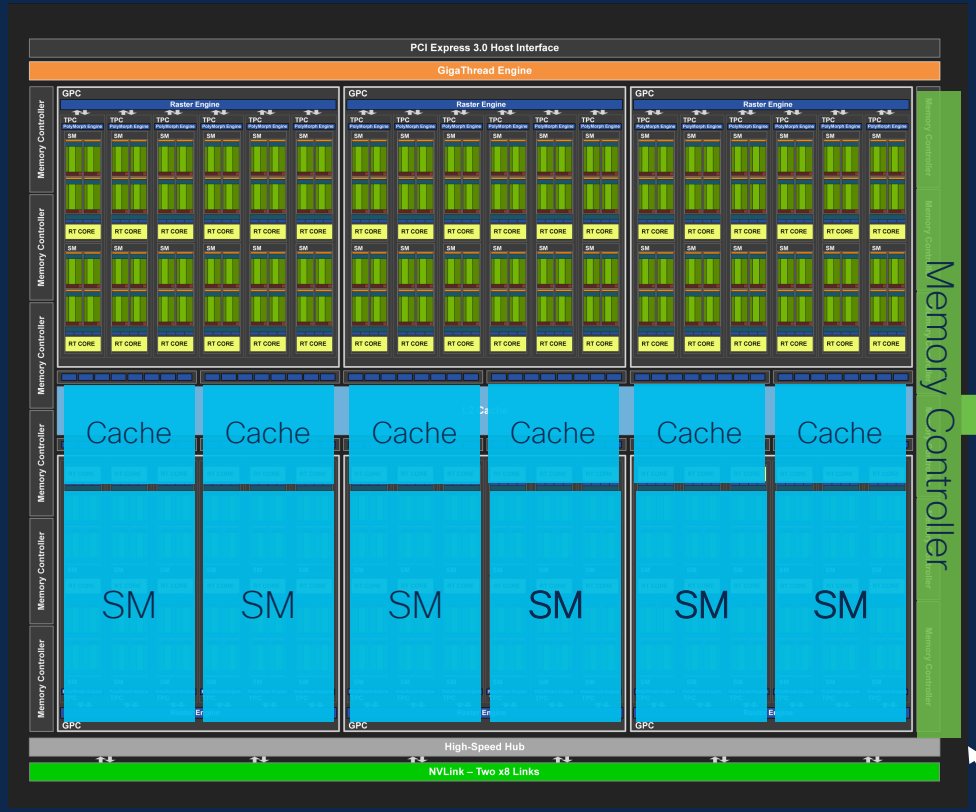


NVIDIA Turing GPU Architecture

Coherent Caching



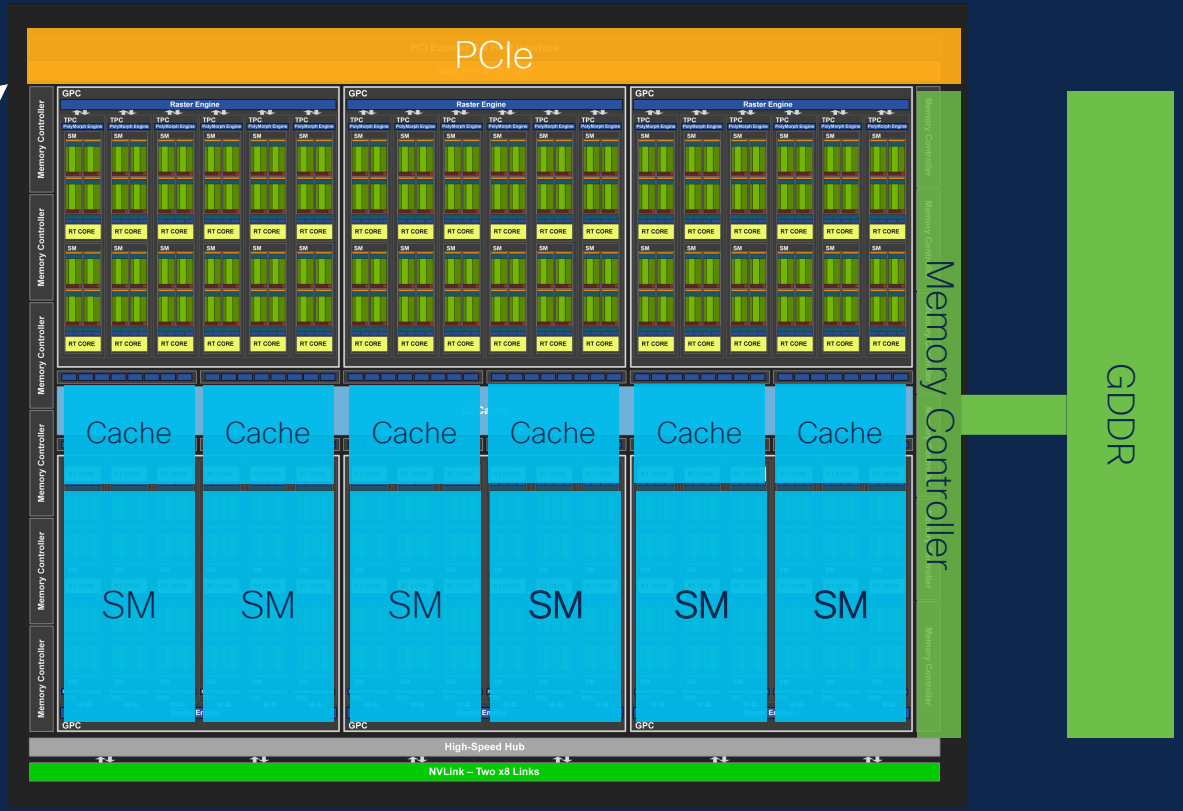
NVIDIA Turing GPU Architecture



NVIDIA Turing GPU Architecture

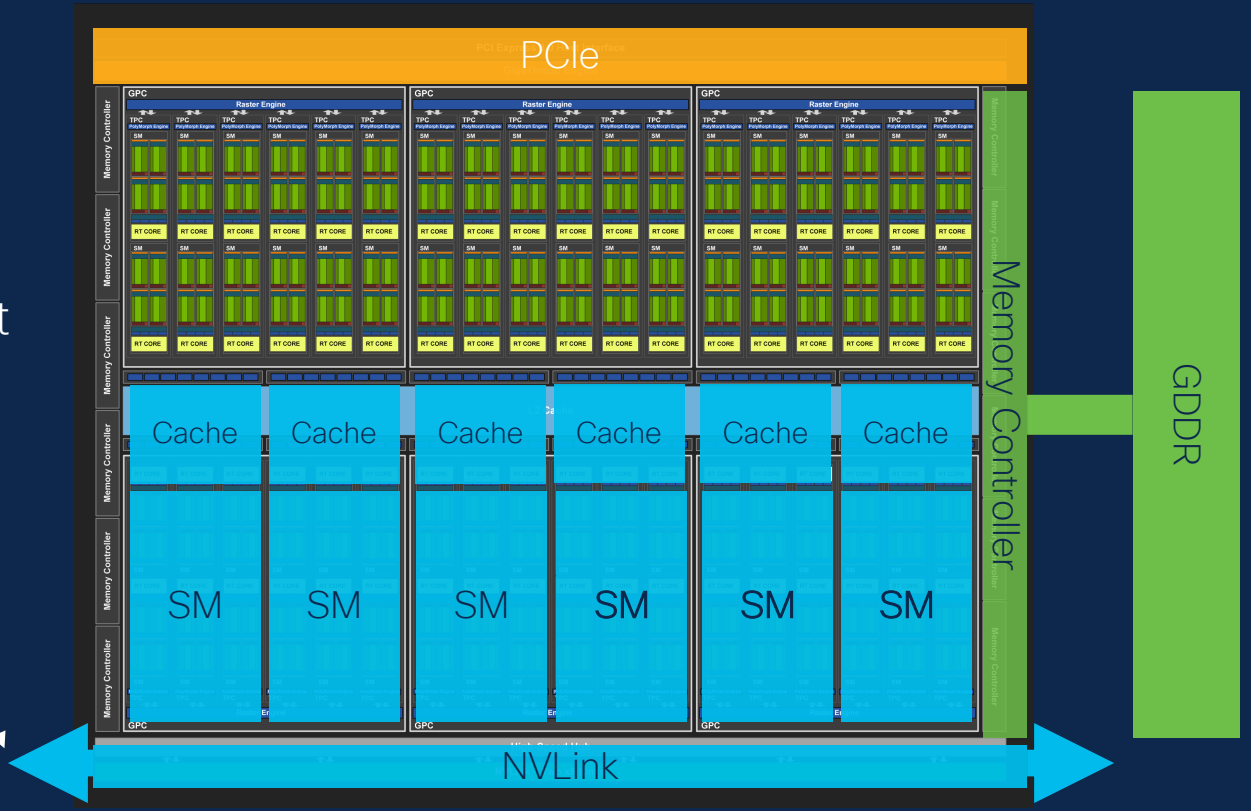
Memory controller and
High bandwidth memory

PCIe controller

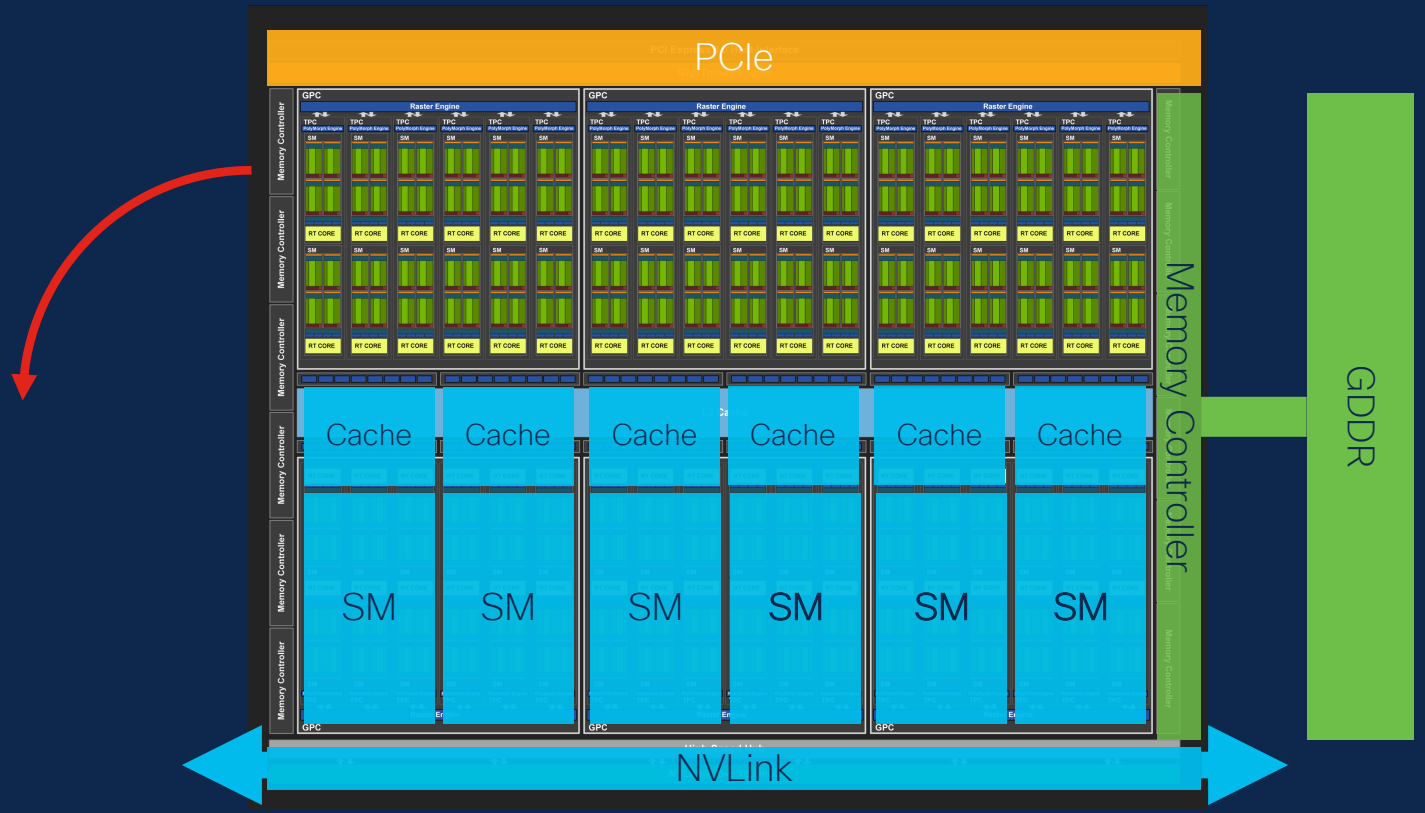


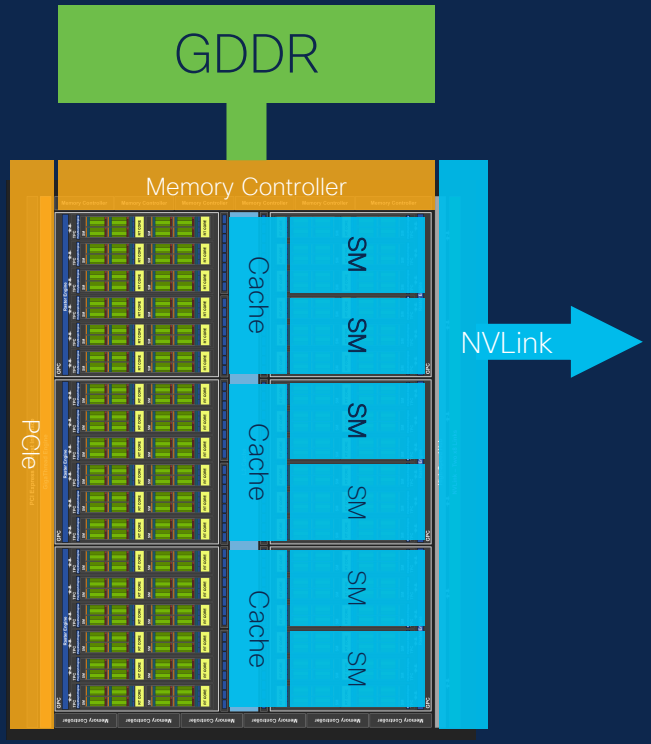
NVIDIA Turing GPU Architecture

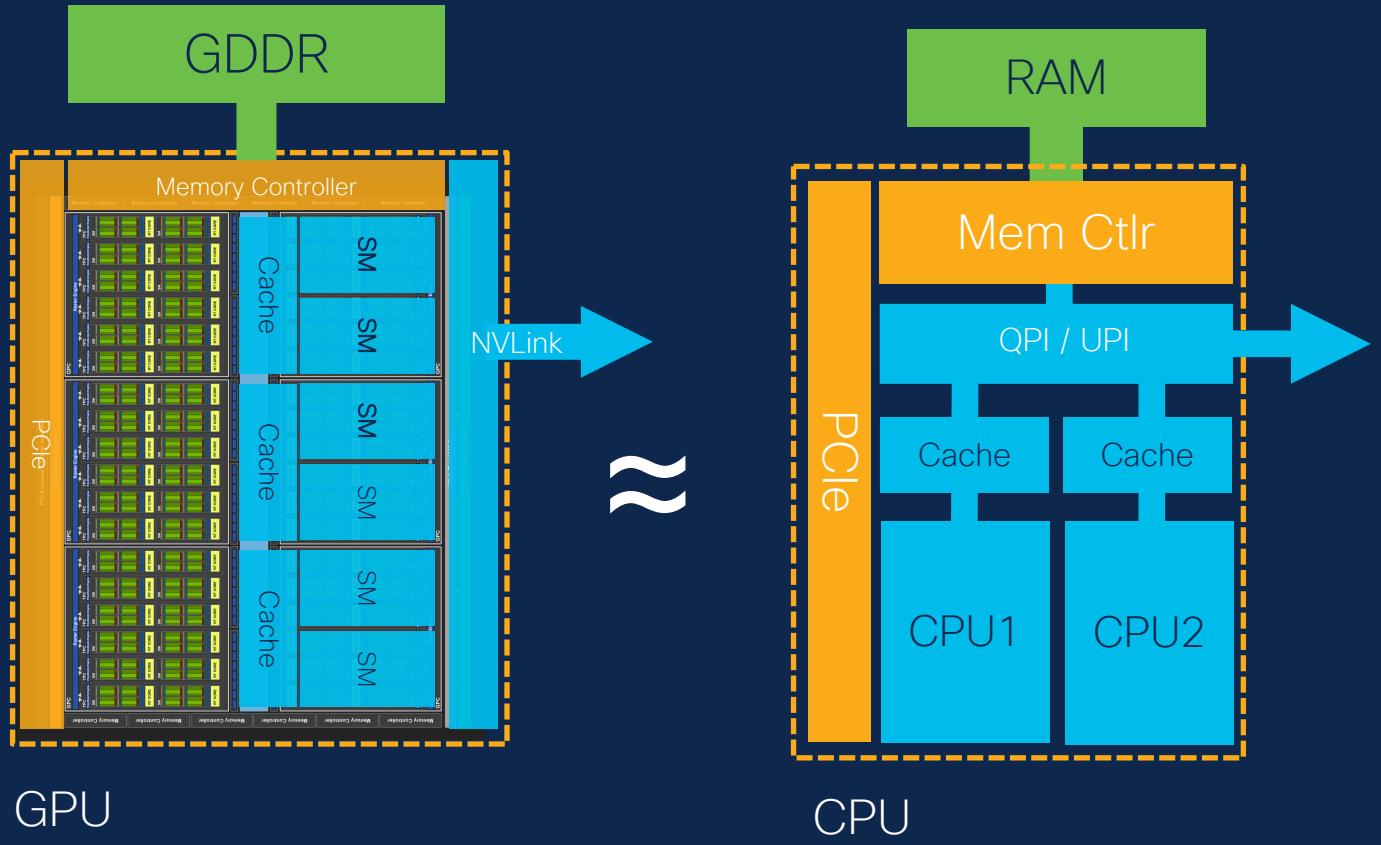
Cache Coherent
NUMA Interface



NVIDIA Turing GPU Architecture

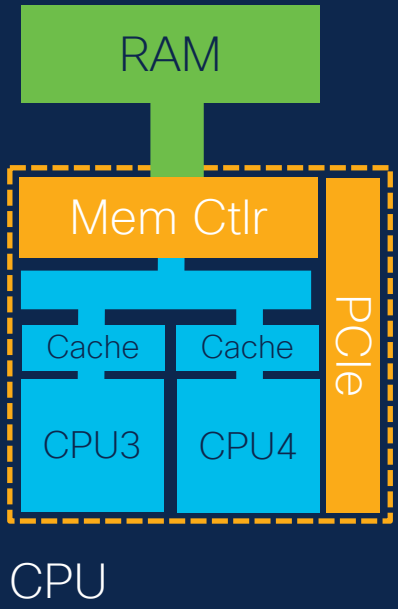




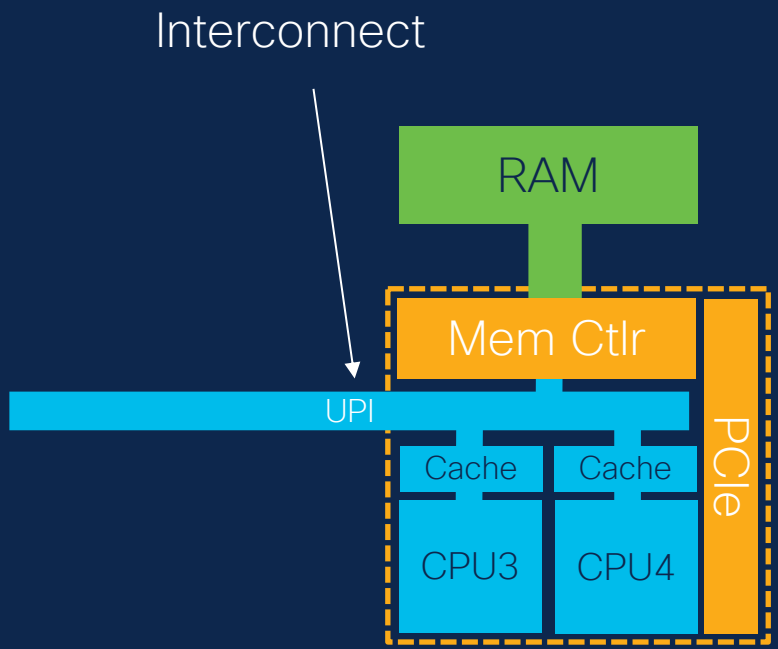


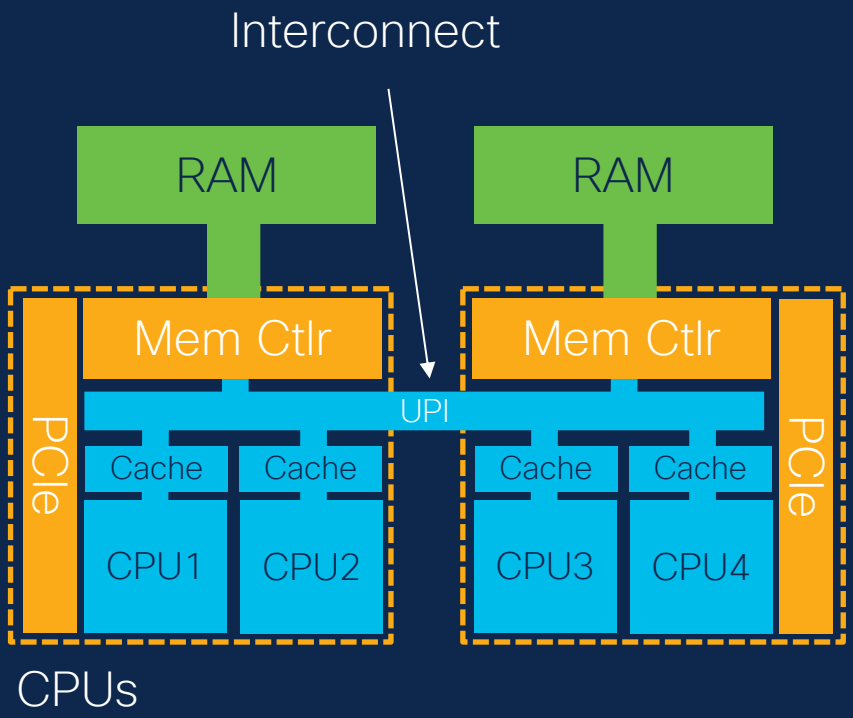
What is an interconnect?

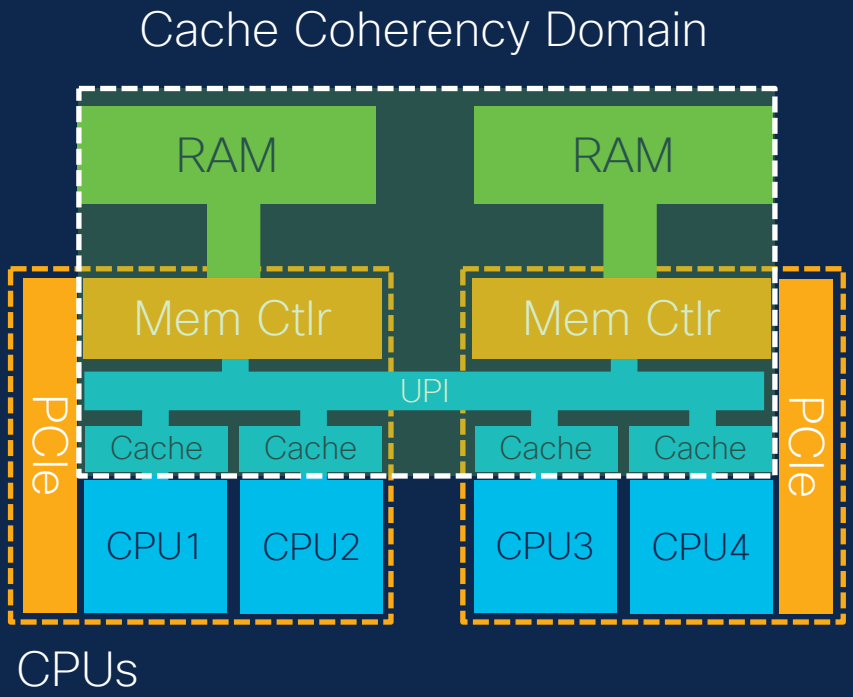
What is cache coherence?

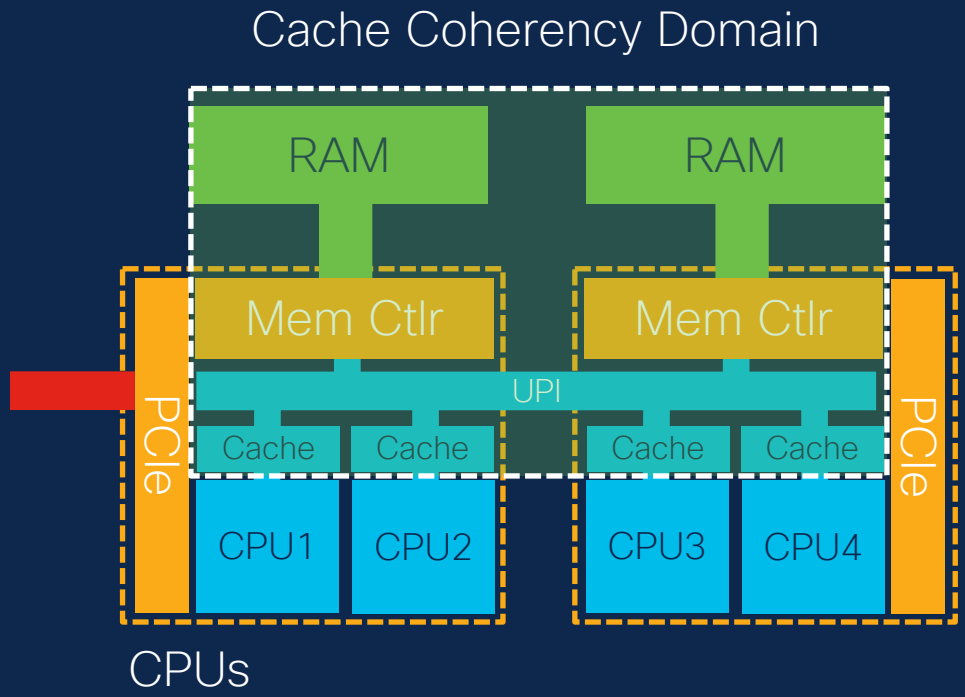


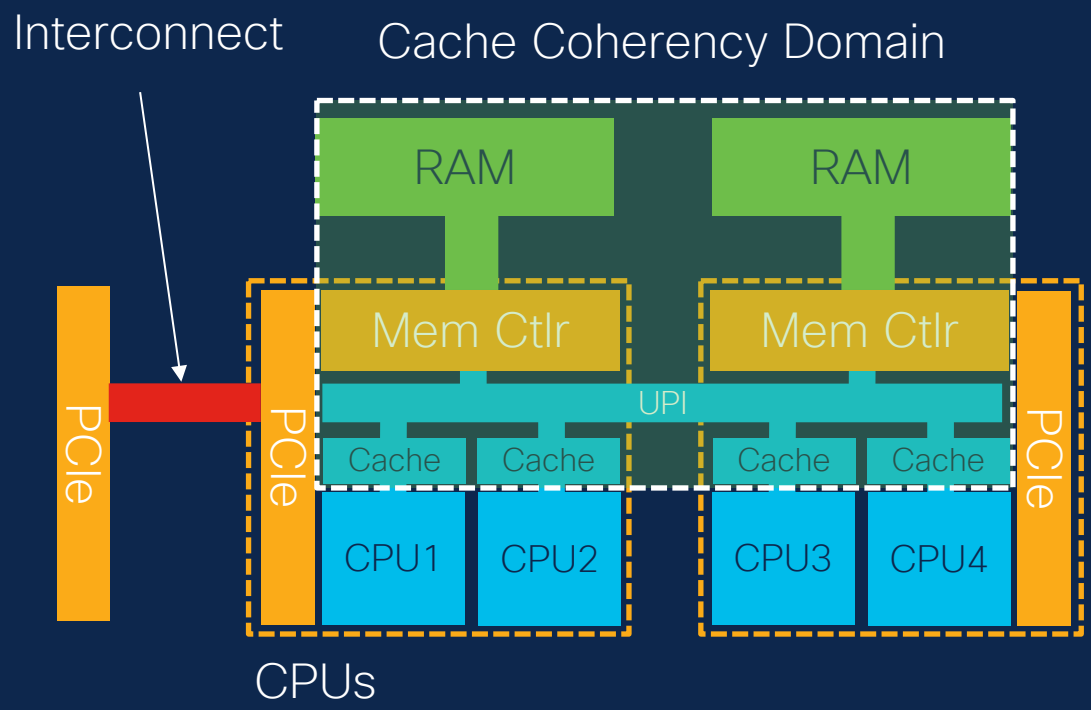
CPU

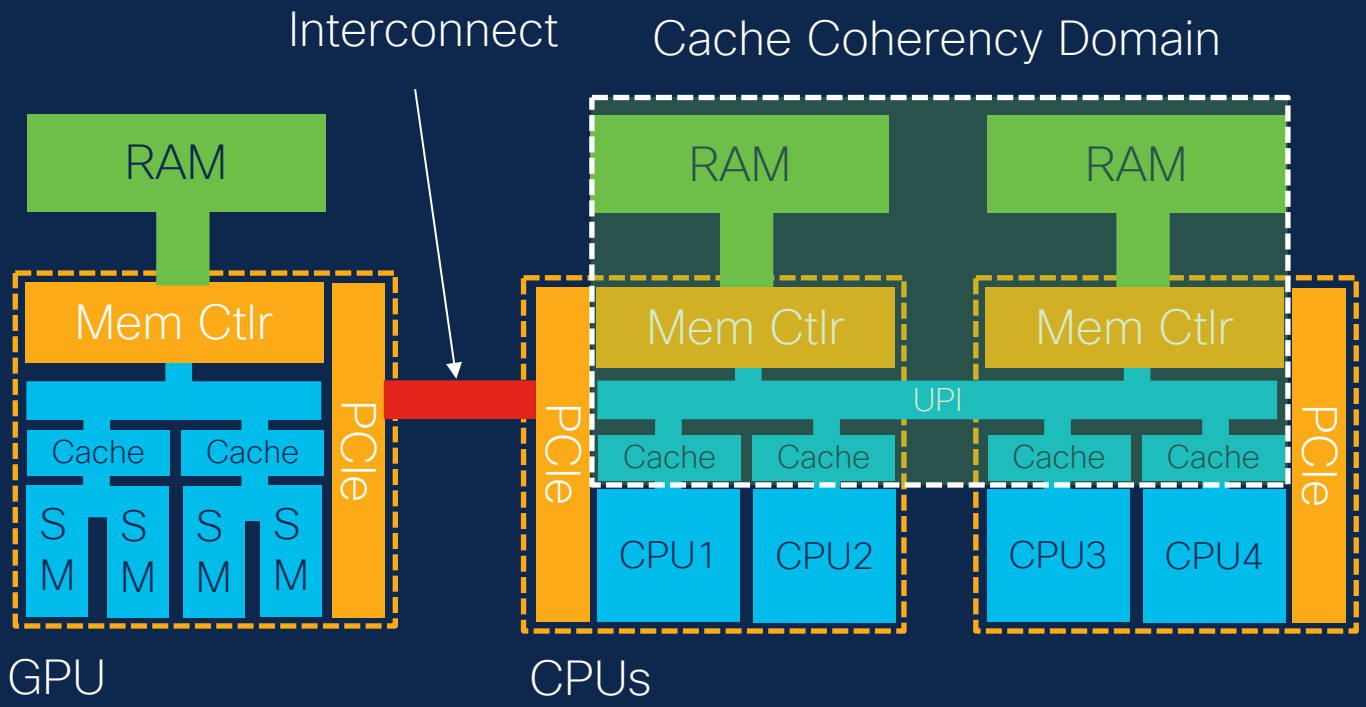


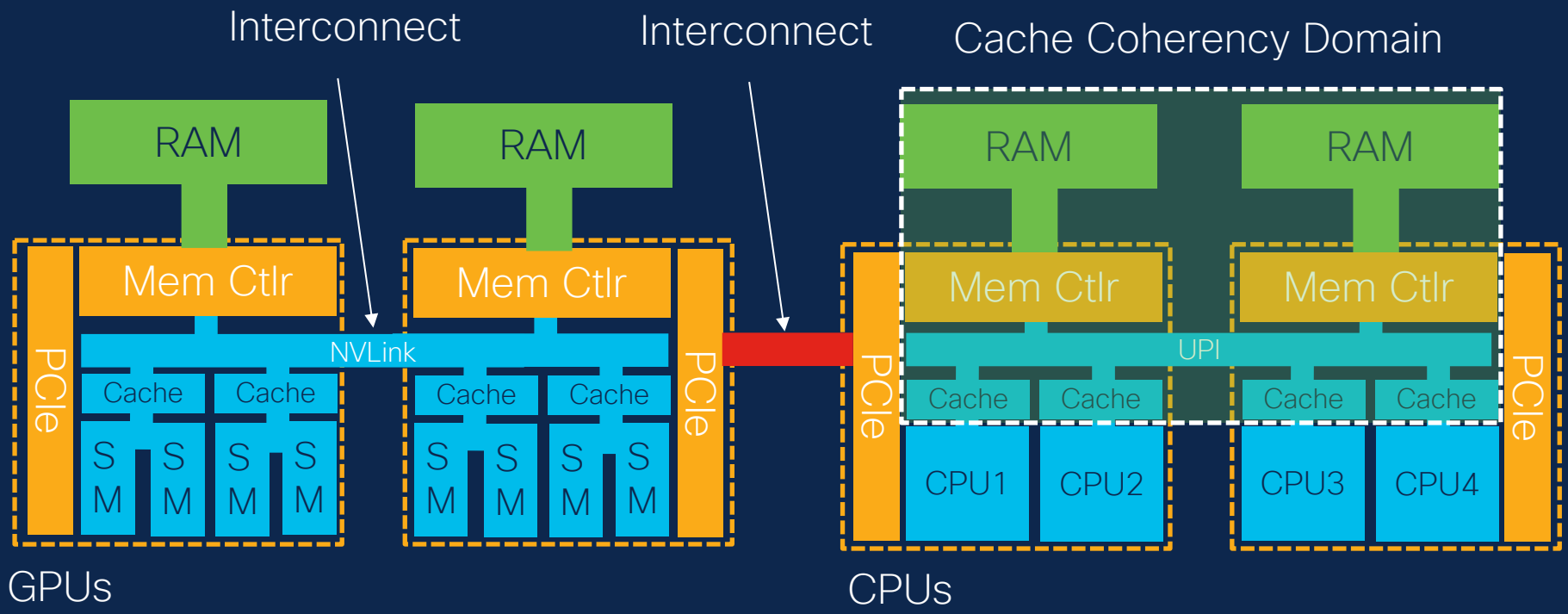


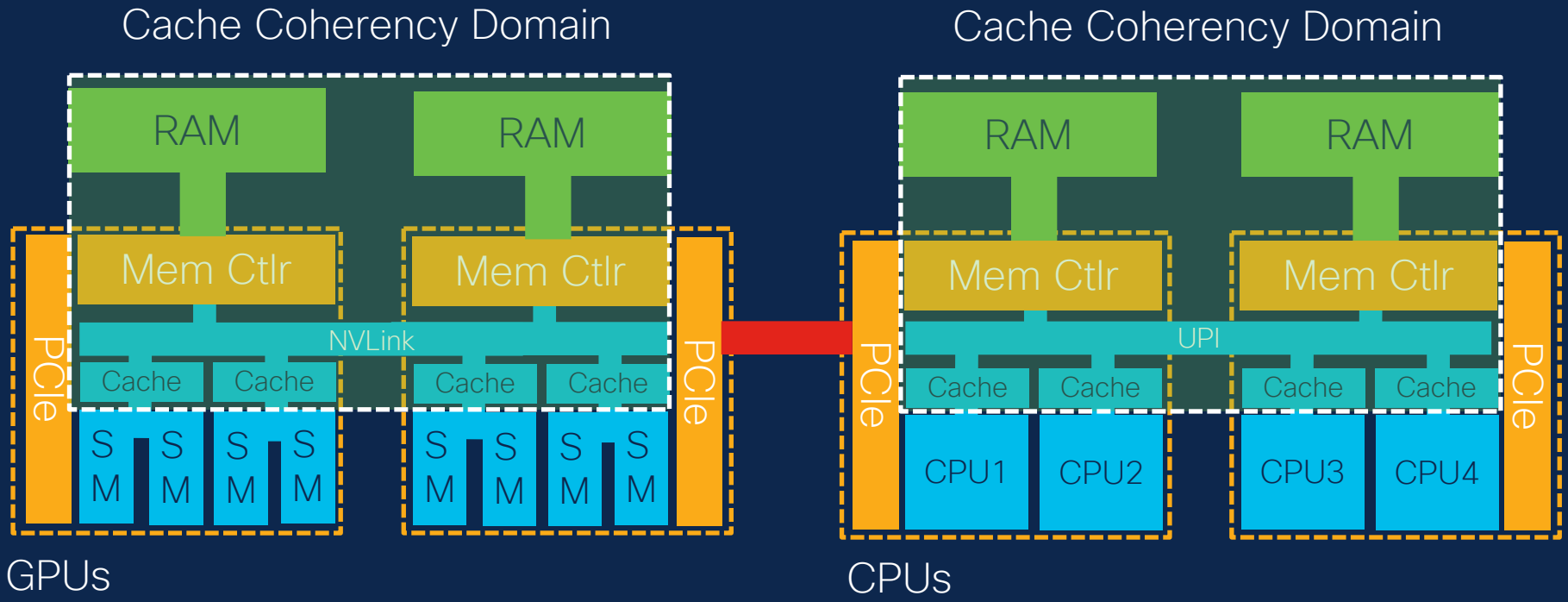




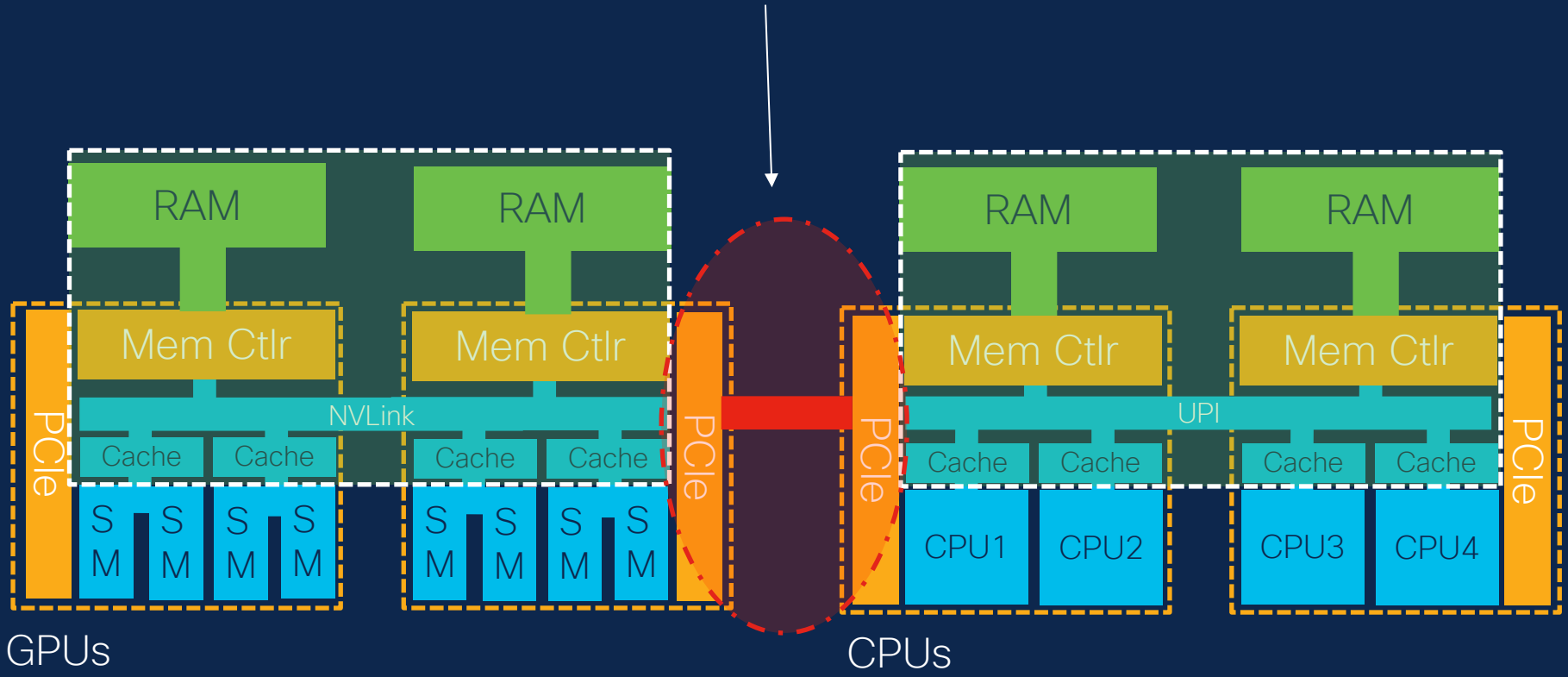




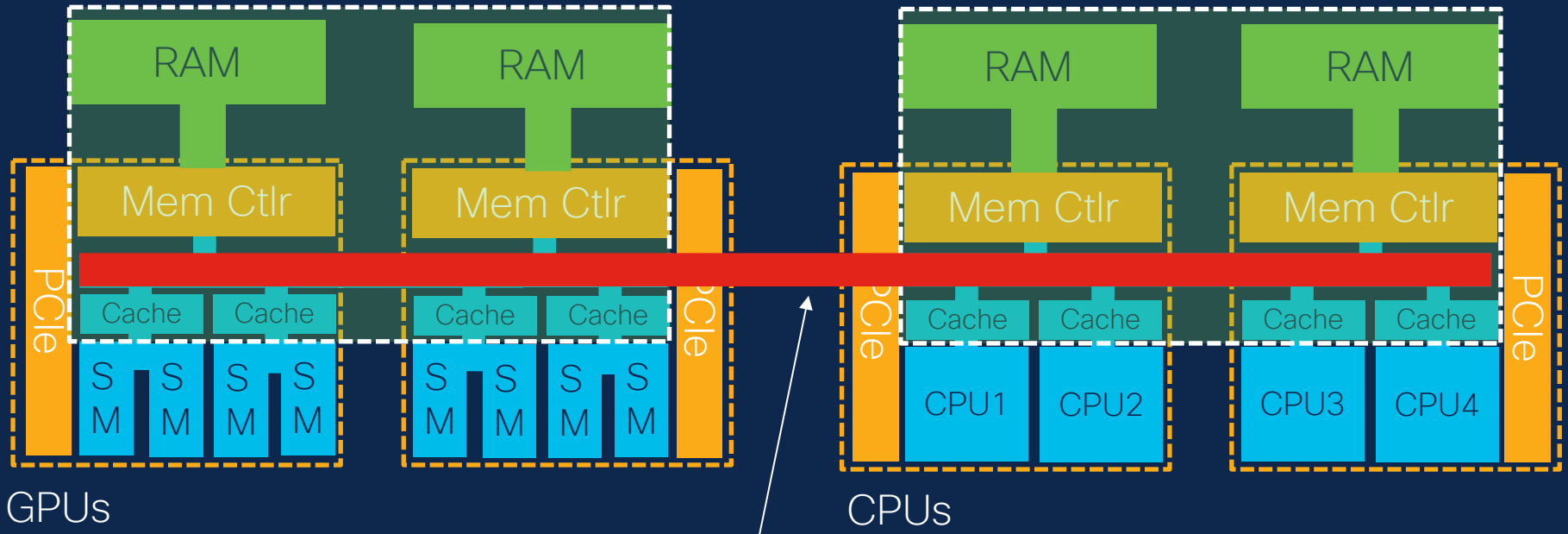




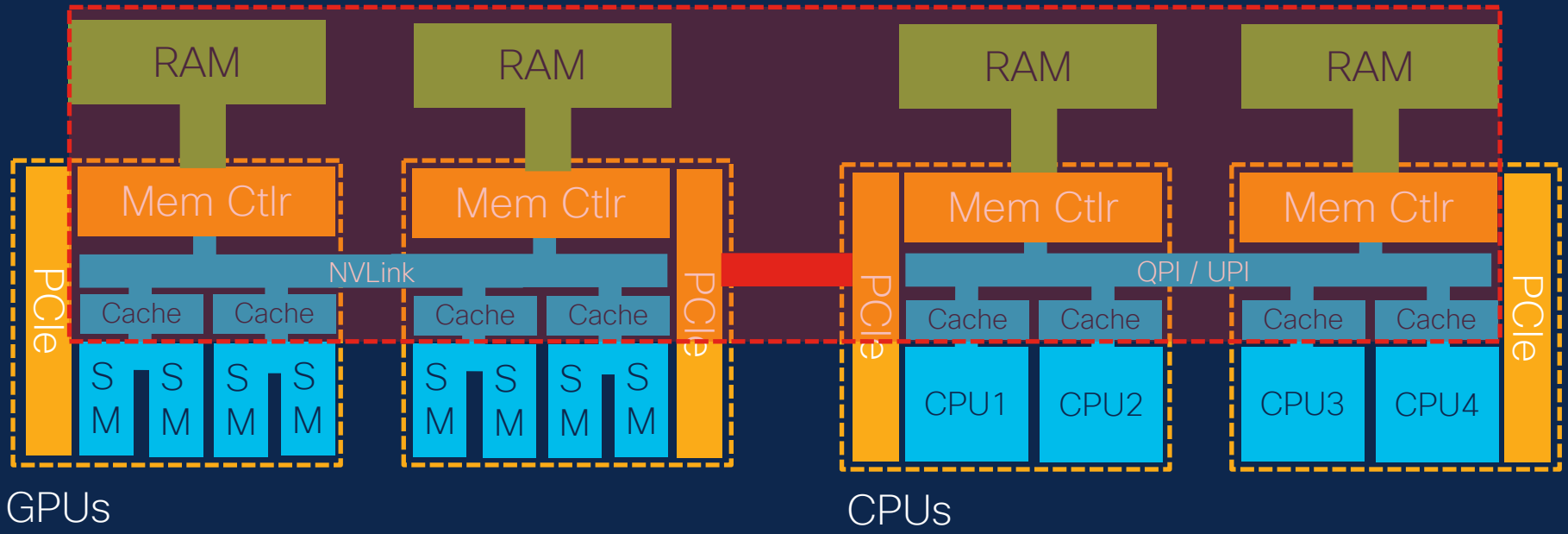
Is PCIe the right *accelerator interconnect*?



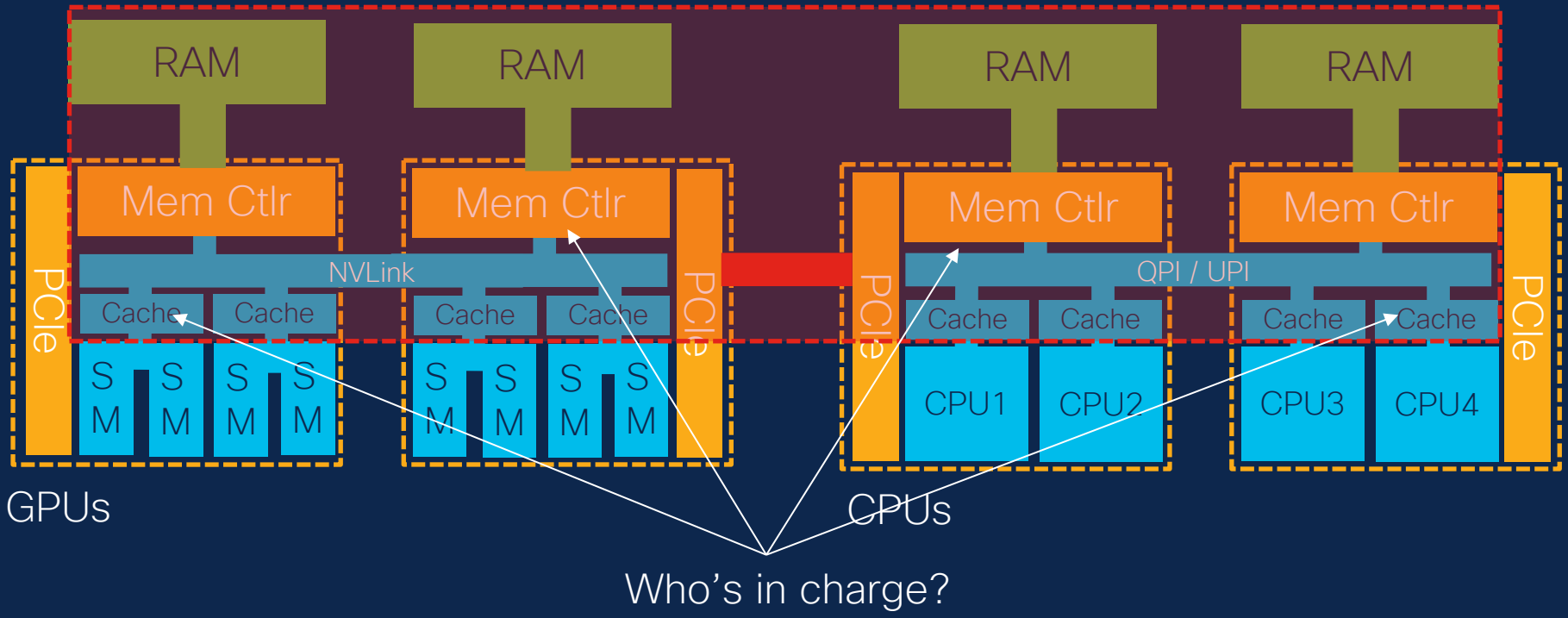
Is PCIe the right *accelerator interconnect*?



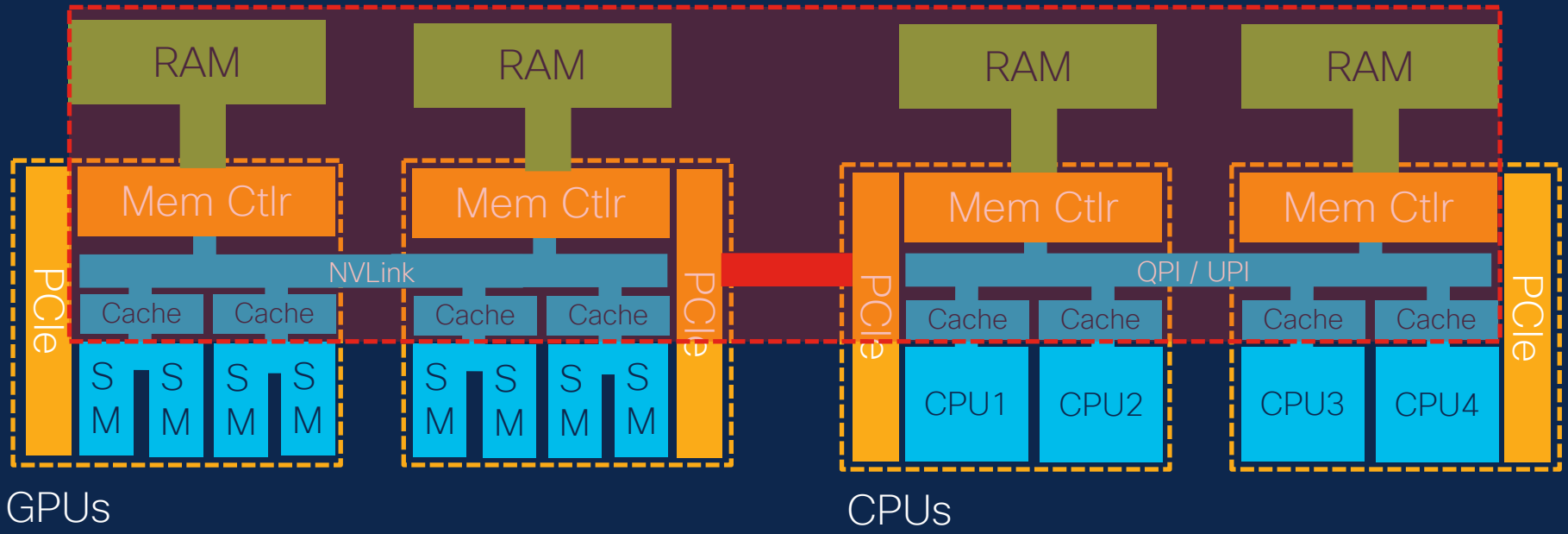
Is PCIe the right *accelerator interconnect*?
What about *cache coherency*?



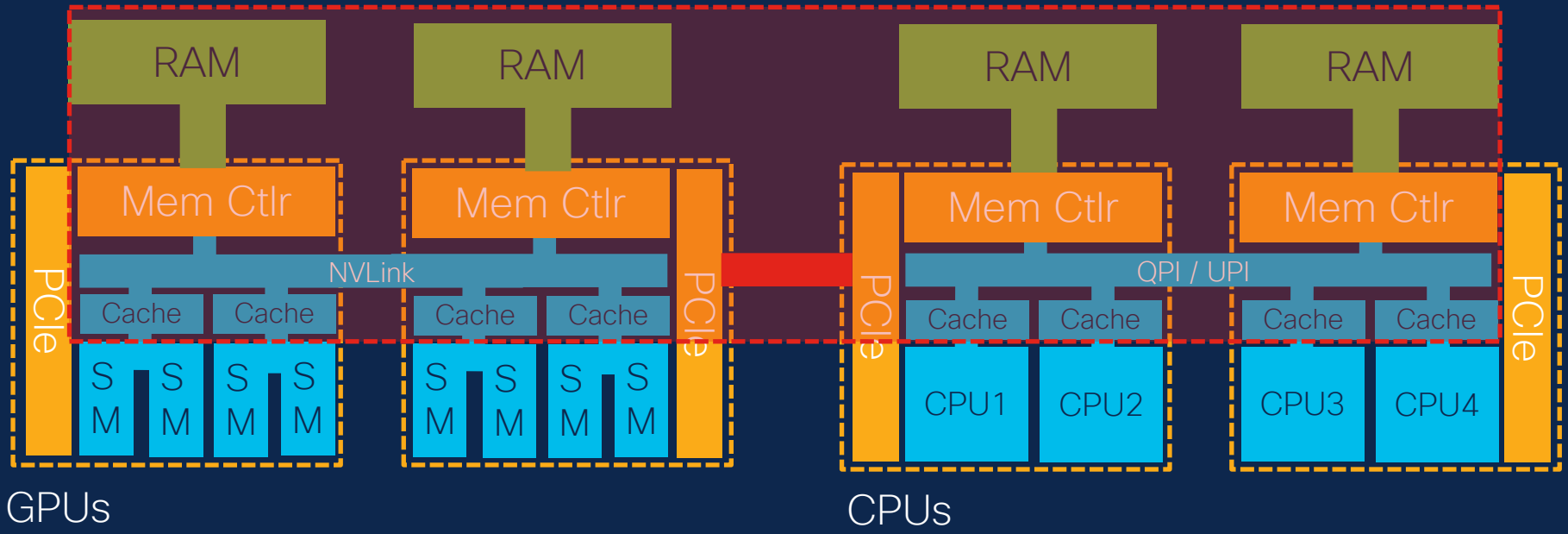
Is PCIe the right *accelerator interconnect*?
What about *cache coherency*?

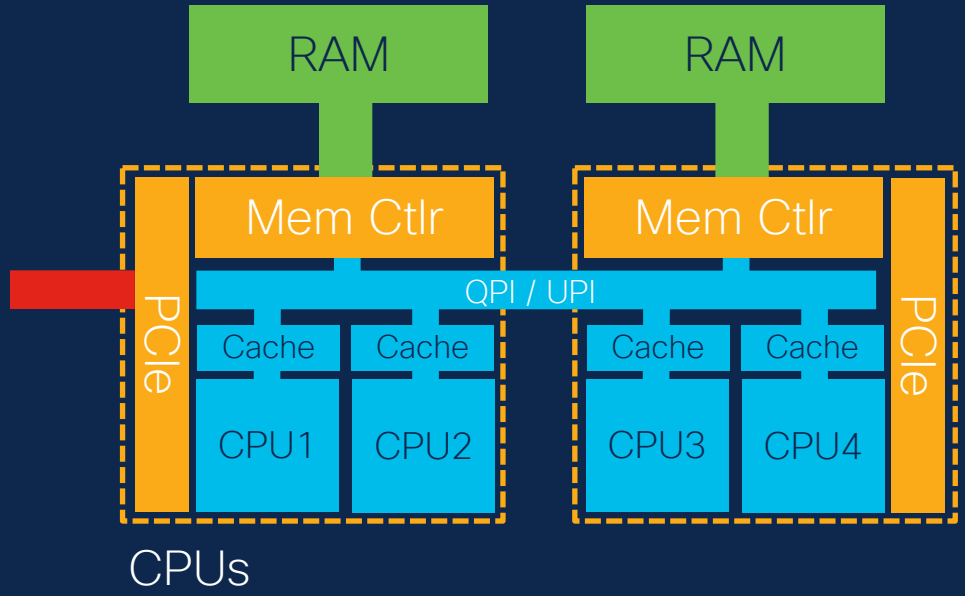


Is PCIe the right *accelerator interconnect*?
What about *cache coherency*?
Does this apply to *GPUs only*?

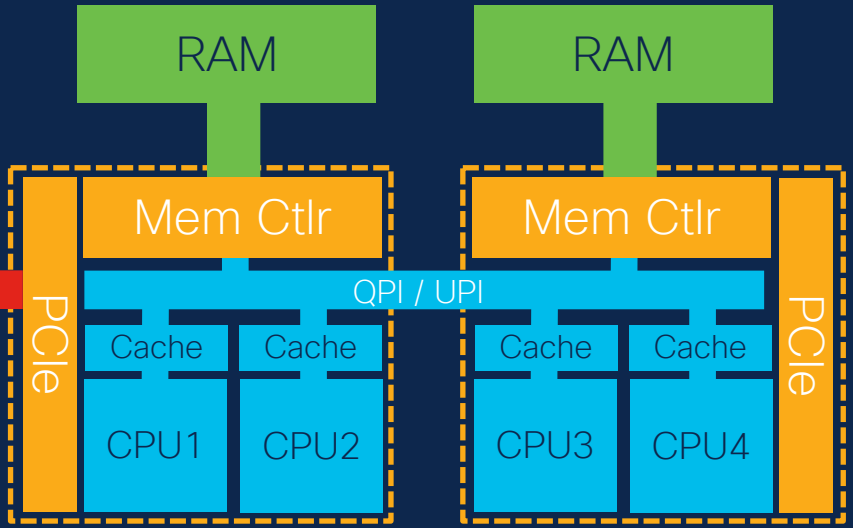


Is PCIe the right *accelerator interconnect*?
What about *cache coherency*?
Does this apply to *GPUs only*? No!

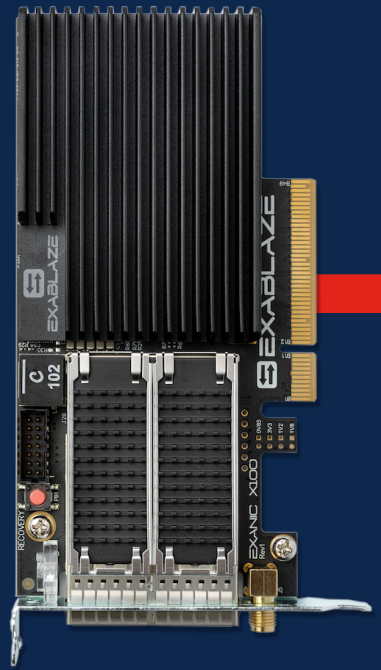




ExaNIC X100



CPUs



RAM

RAM

Mem Ctlr

Mem Ctlr

QPI / UPI

PCIe

PCIe

Cache

Cache

Cache

Cache

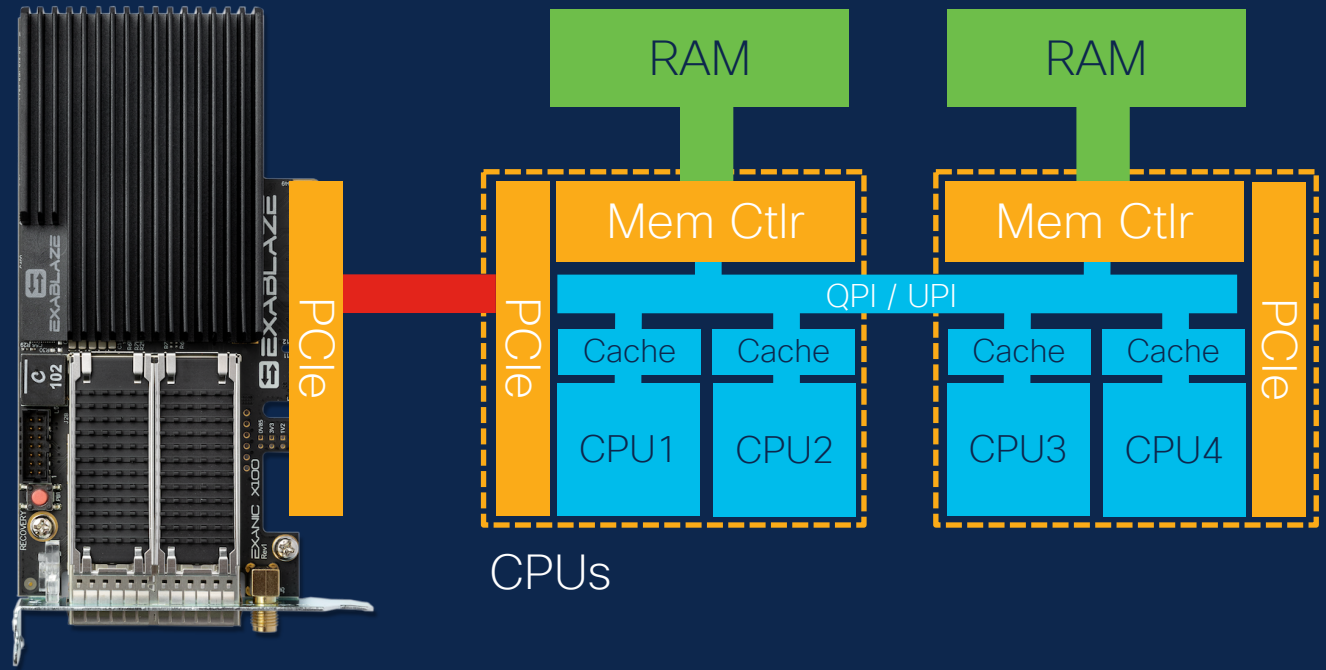
CPU1

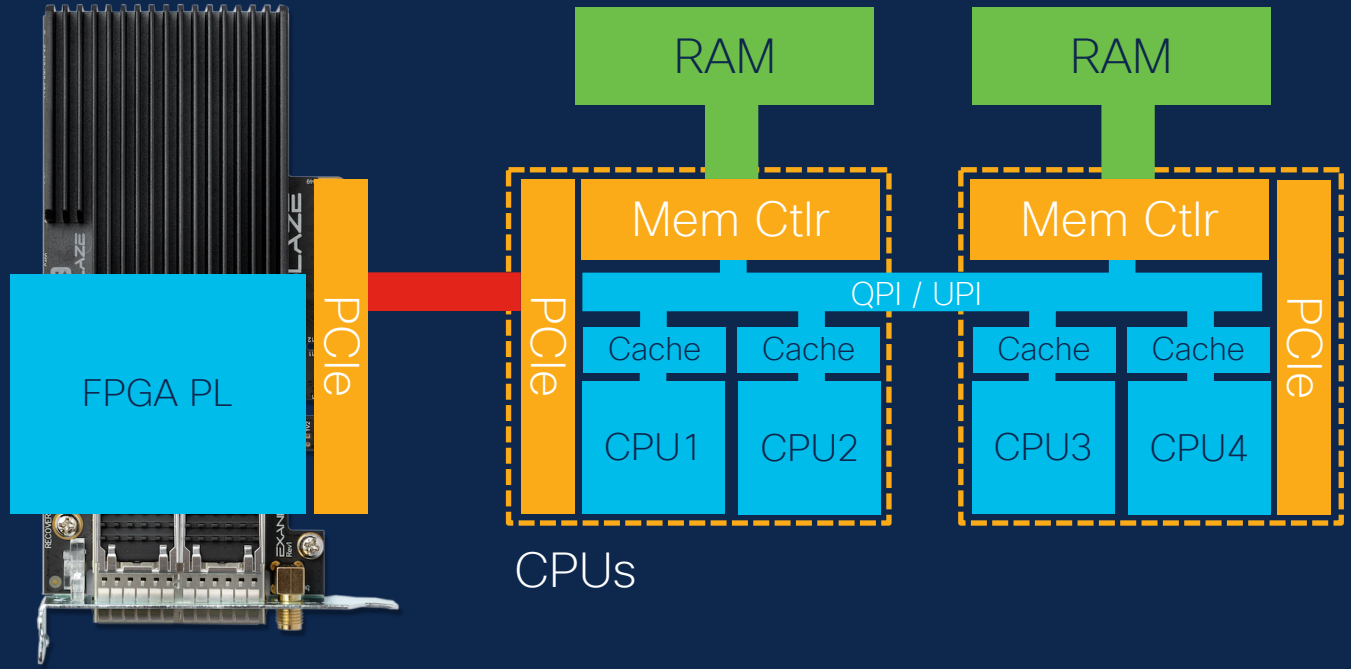
CPU2

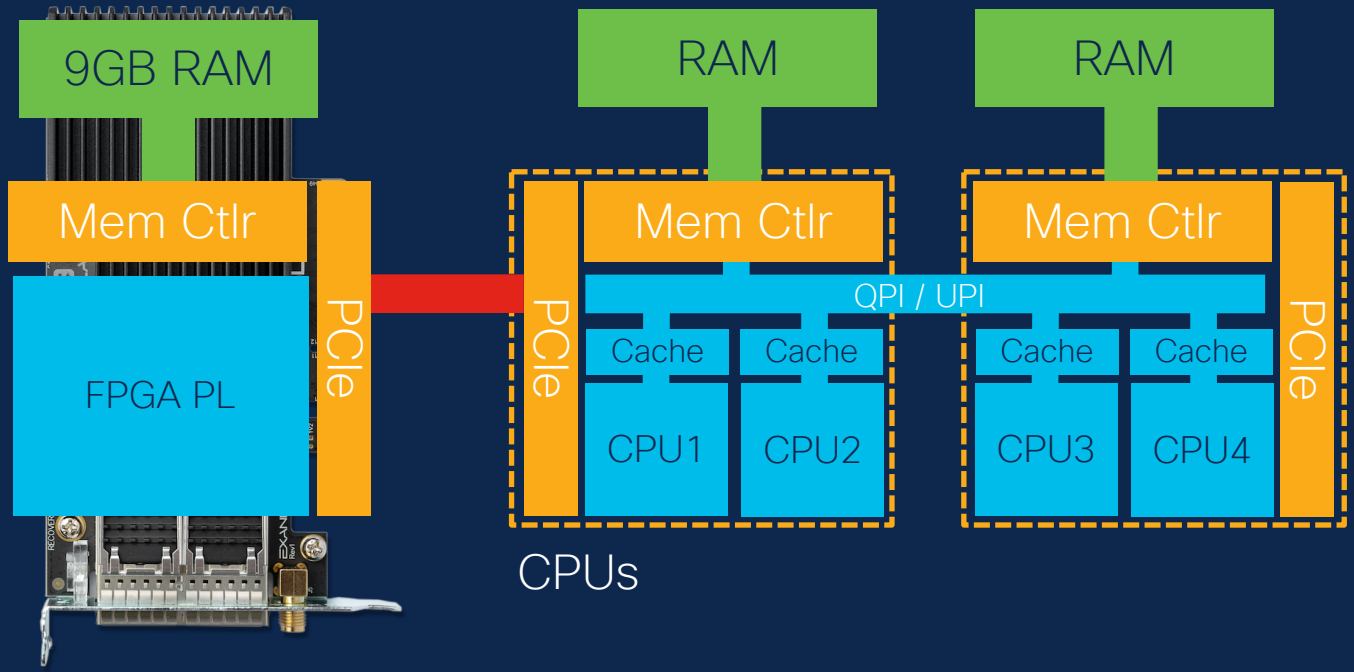
CPU3

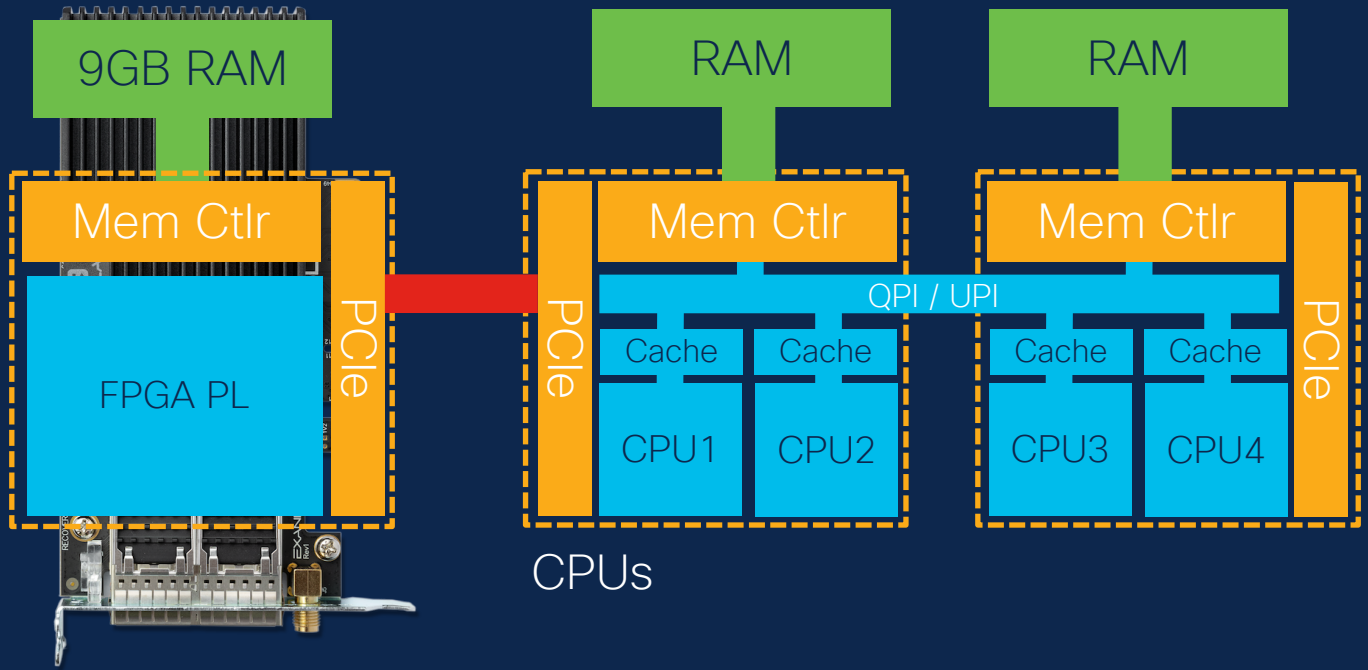
CPU4

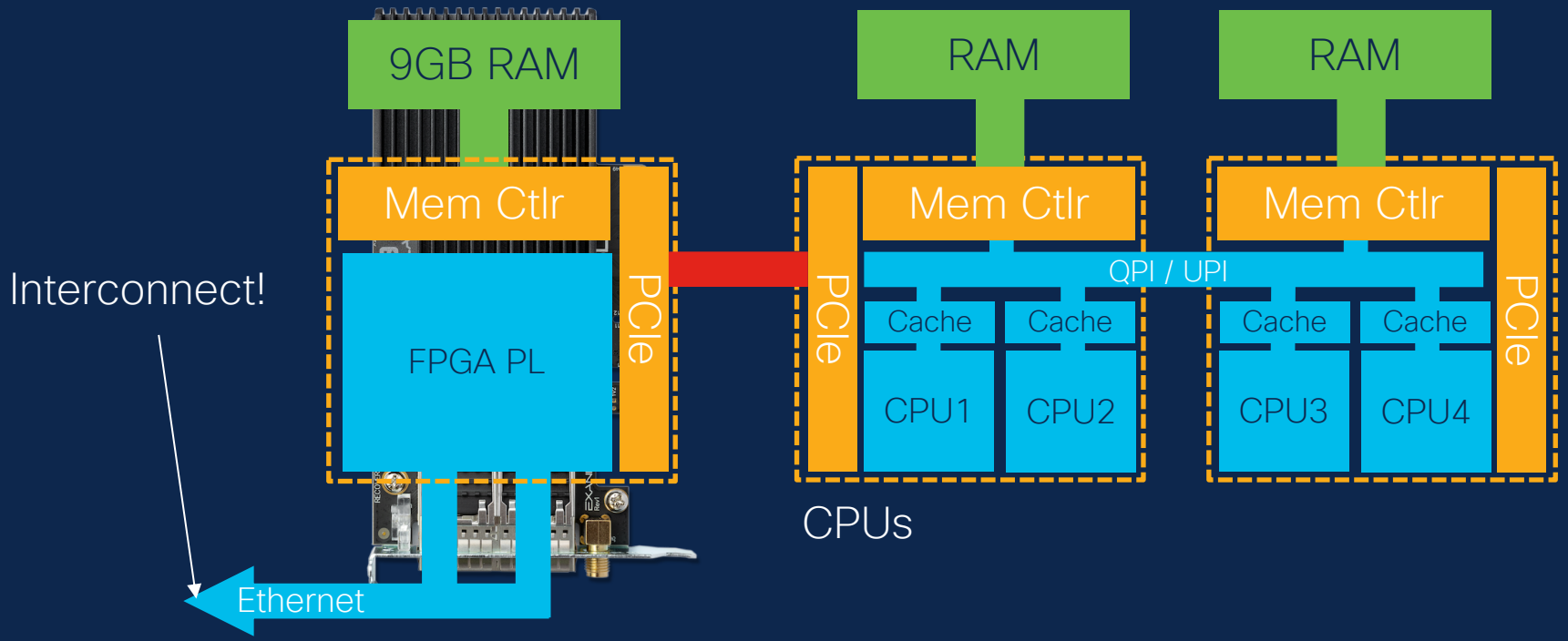
CPUs

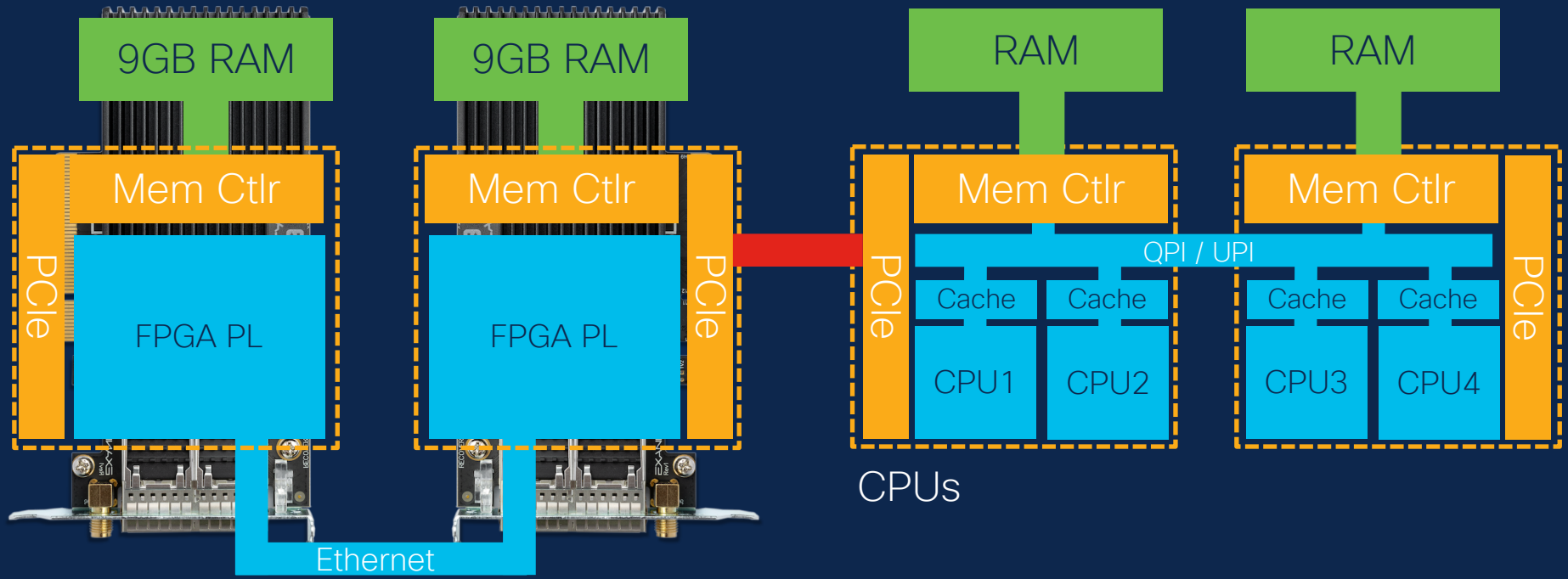


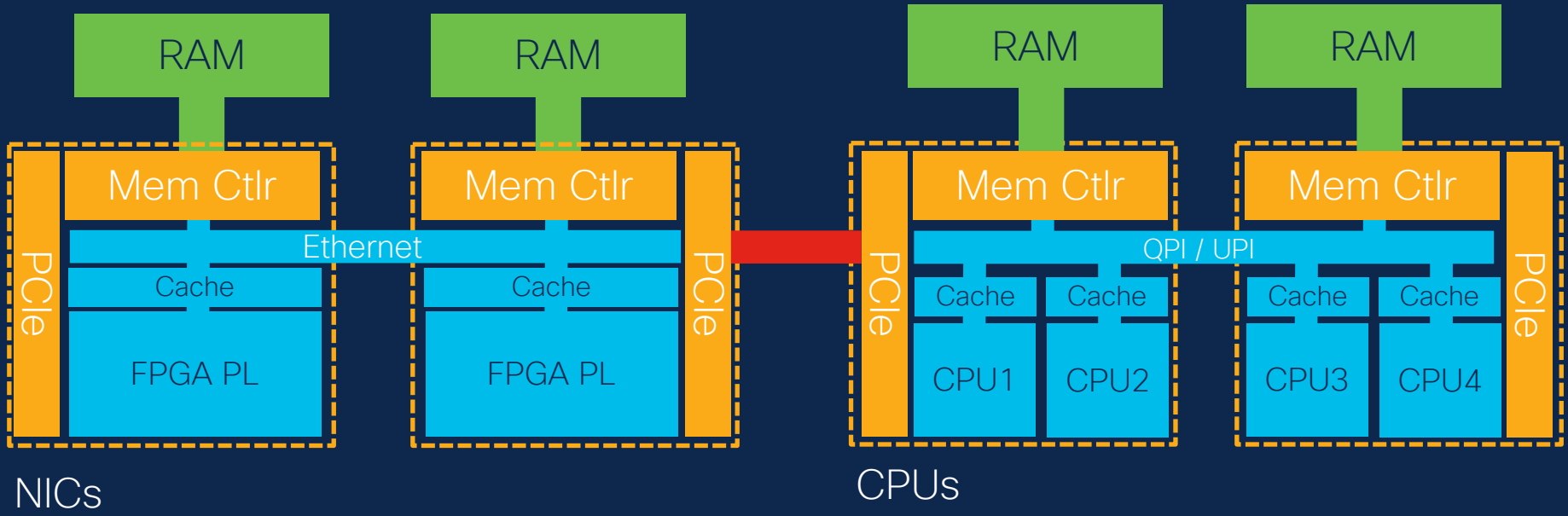


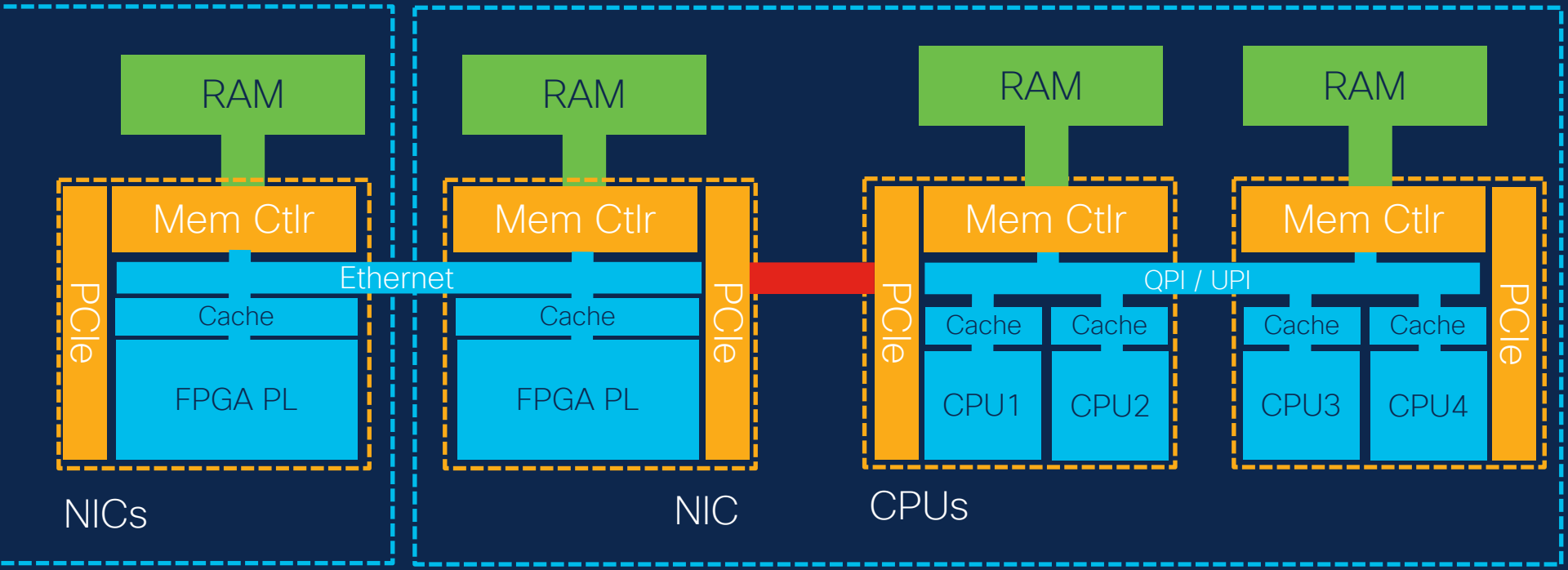


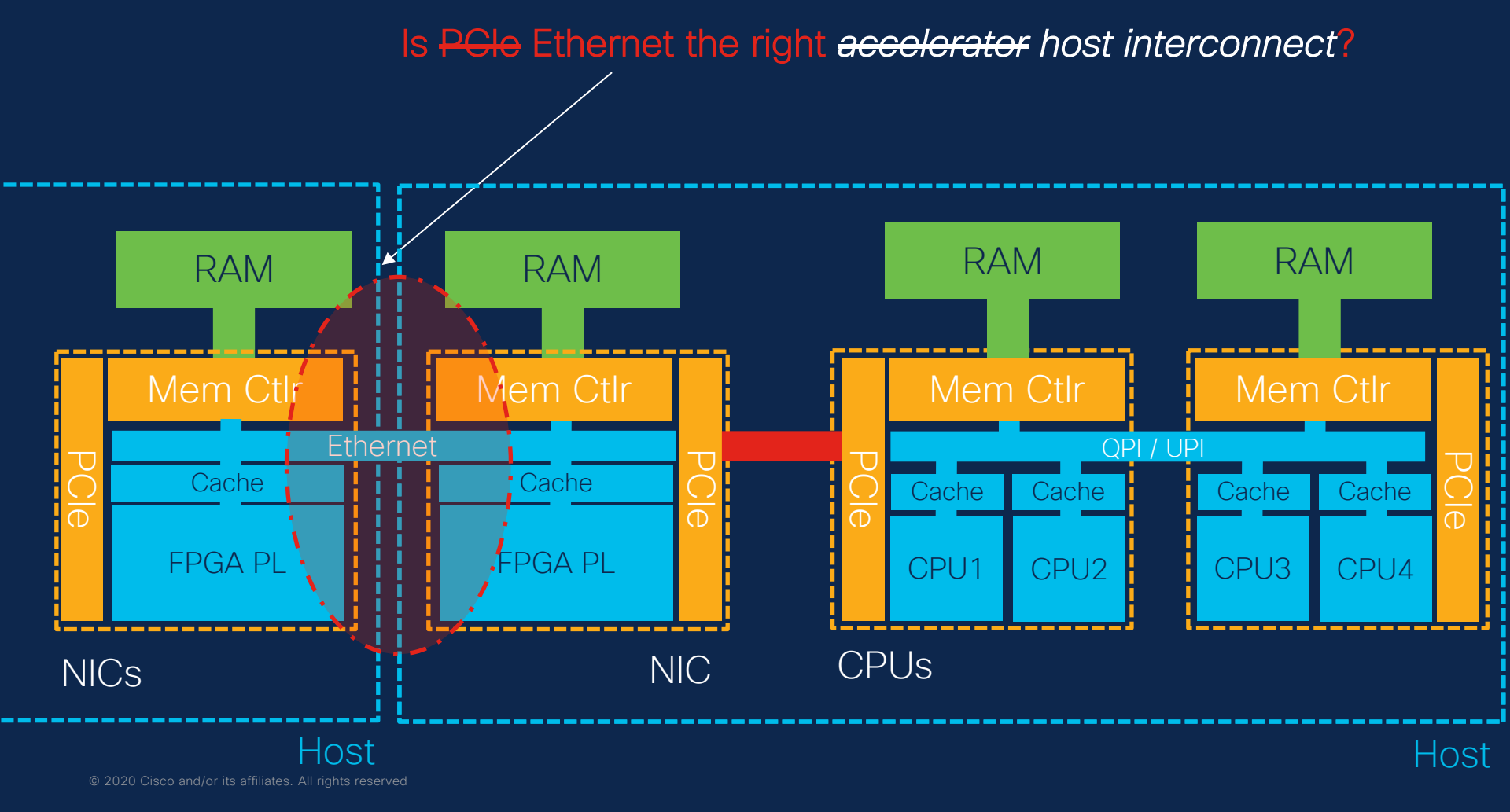




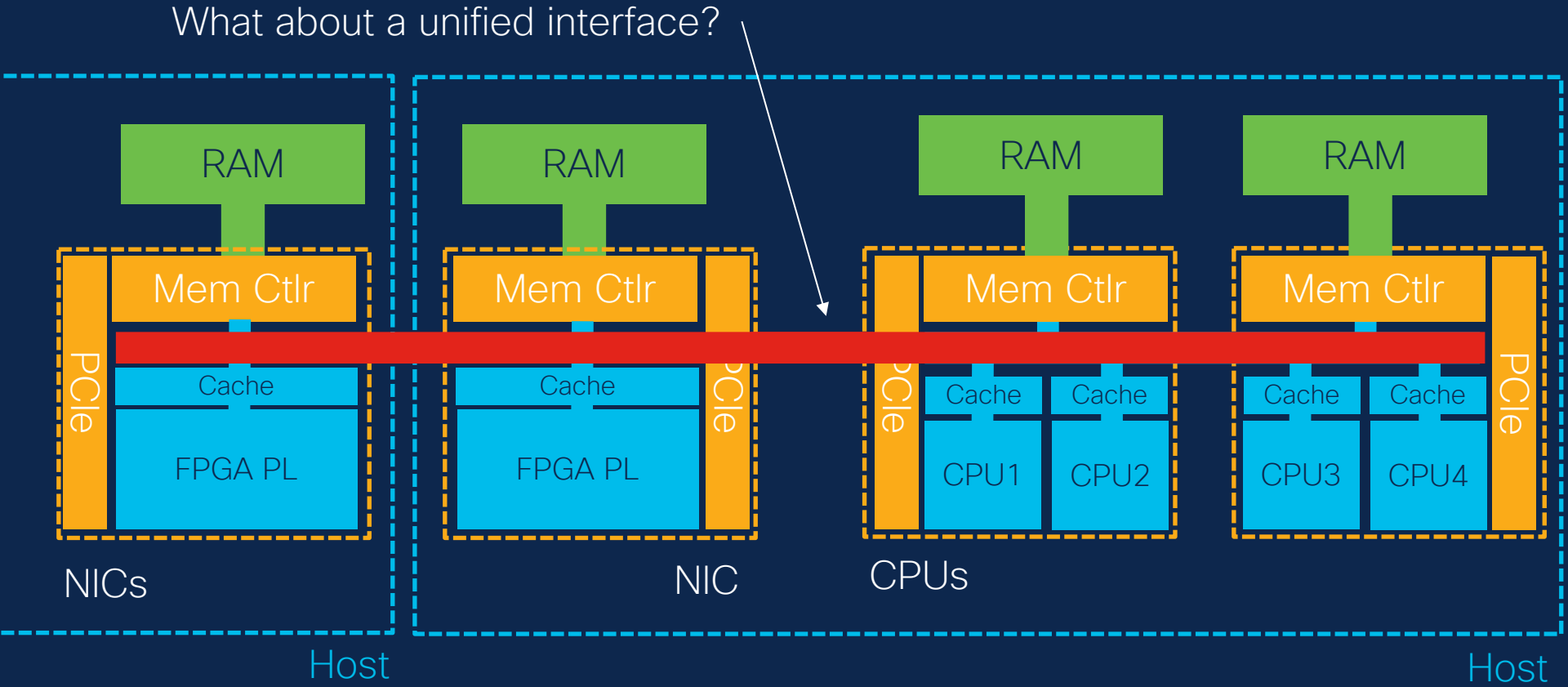




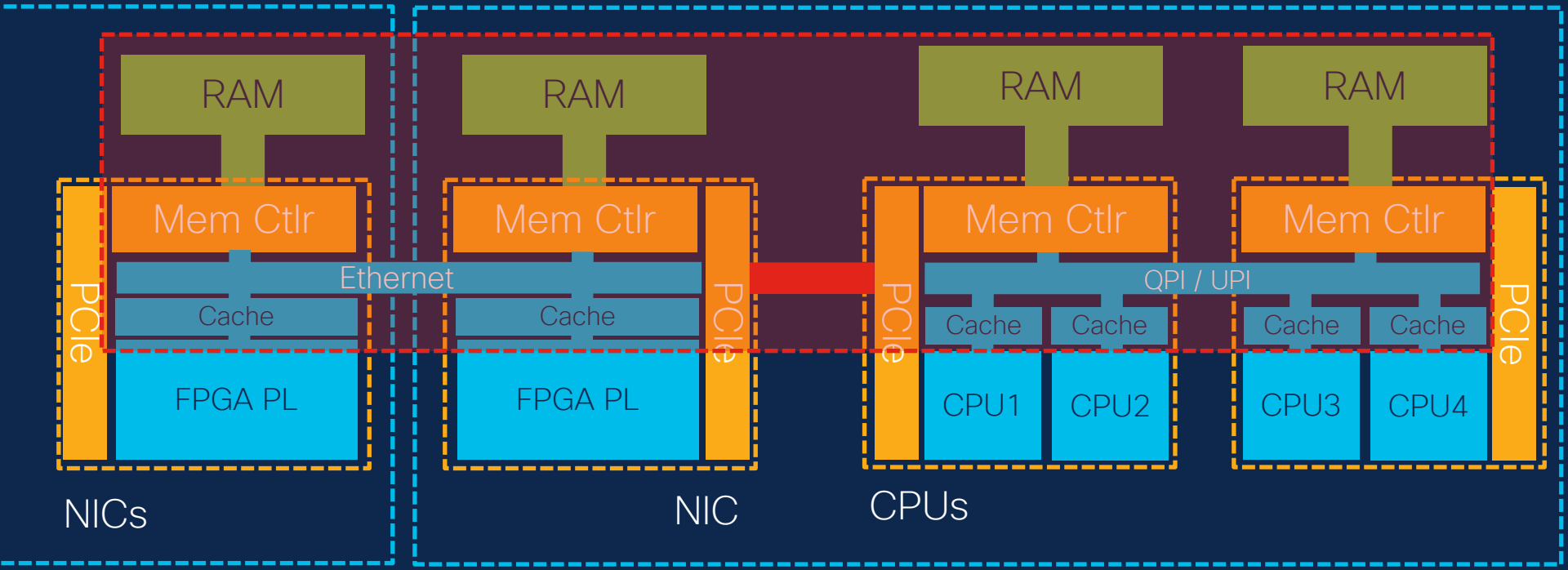




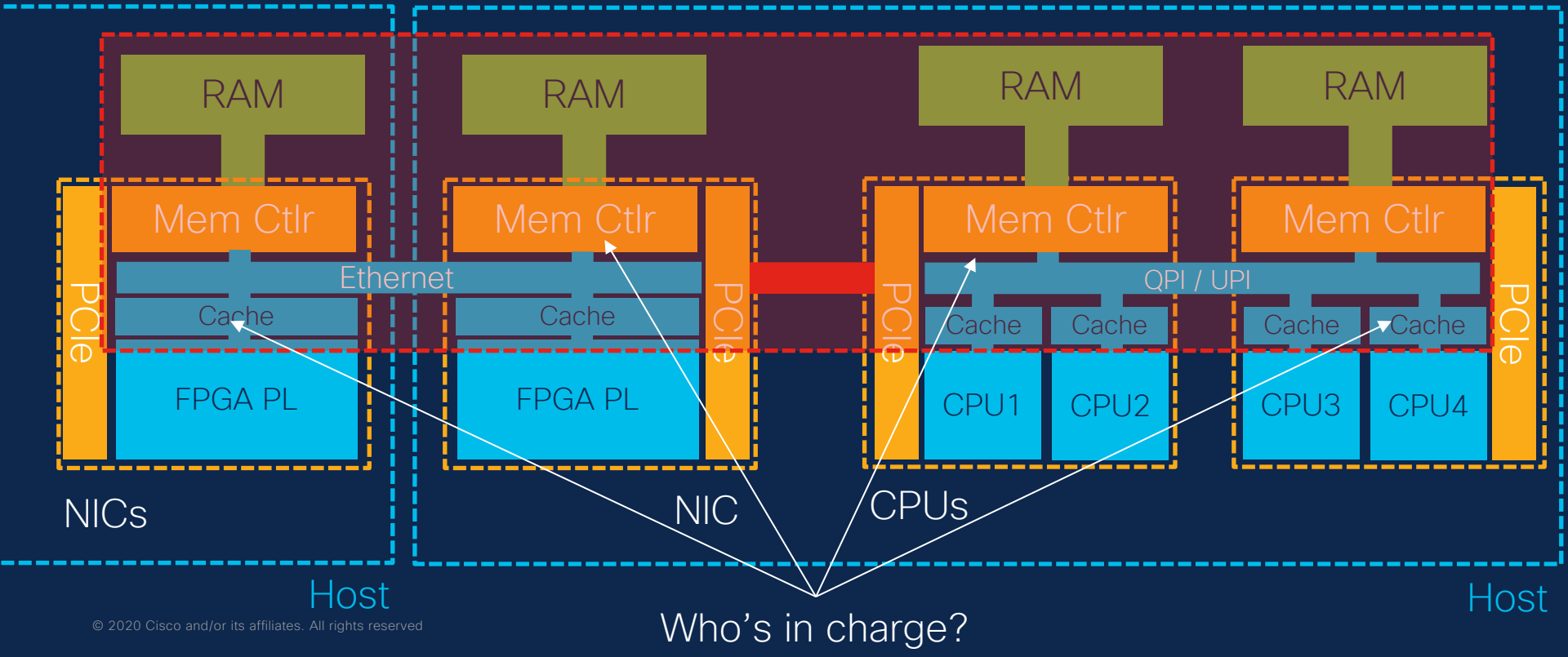
Is ~~PCIe~~ Ethernet the right *accelerator interconnect*?



Is ~~PCIe~~ Ethernet the right *accelerator interconnect*?
What about *unified cache coherency*?



Is ~~PCIe~~ Ethernet the right *accelerator interconnect*?
What about *unified cache coherency*?



Unscrambling the alphabet (soup?)…

The Interconnect Alphabet

- AGP
- BlueLink
- CCIX
- CXL
- GenZ
- Ethernet
- FSB
- ISA
- NVLink
- OpenCAPI
- PCIe 4.0 / 5.0
- QPI
- UPI
- VCLX

The Interconnect Alphabet

- AGP
- FSB
- ISA
- QPI
- BlueLink
- NVLink
- UPI
- VCLX
- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- AGP
- FSB
- ISA
- QPI

} Obsolete

- BlueLink
- NVLink
- UPI
- VCLX
- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- BlueLink
- NVLink
- UPI
- VCLX
- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- BlueLink
- NVLink
- UPI

} Proprietary

- VCLX
- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- VCLX
- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- VCLX → I made this one up
- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0
- GenZ
- Ethernet

The Interconnect Alphabet

- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0

- GenZ

- Ethernet

The Interconnect Alphabet

- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0

Mostly inside a host

- GenZ

- Ethernet

The Interconnect Alphabet

- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0

} Mostly inside a host

- GenZ
- Ethernet

} Datacenter

The Interconnect Alphabet

- CCIX
- CXL
- OpenCAPI
- PCIe 4.0 / 5.0

} Mostly inside a host

- GenZ

} Chip-to-chip up to rack-scale

- Ethernet

} Datacenter

Comparisons

Standard

Comparisons

Standard
Ethernet

Comparisons

Standard	PHY
Ethernet	802.3

Comparisons

Standard	PHY	Link Rate (Gb/s)
Ethernet	802.3	100

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links
Ethernet	802.3	100	8

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)
Ethernet	802.3	100	8	800

Comparisons

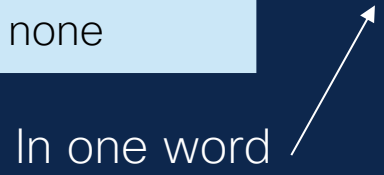
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence
Ethernet	802.3	100	8	800	none

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	



Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	

In one word

AKA: my outrageous opinions

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16				

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256		

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late

PCIe 4.0 announced in 2011
Device availability 2020

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32				

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early

PCIe 5.0 announced in 2017
Devices availability 2021?

Comparisons

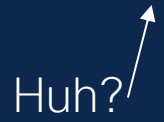
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800		

Comparisons

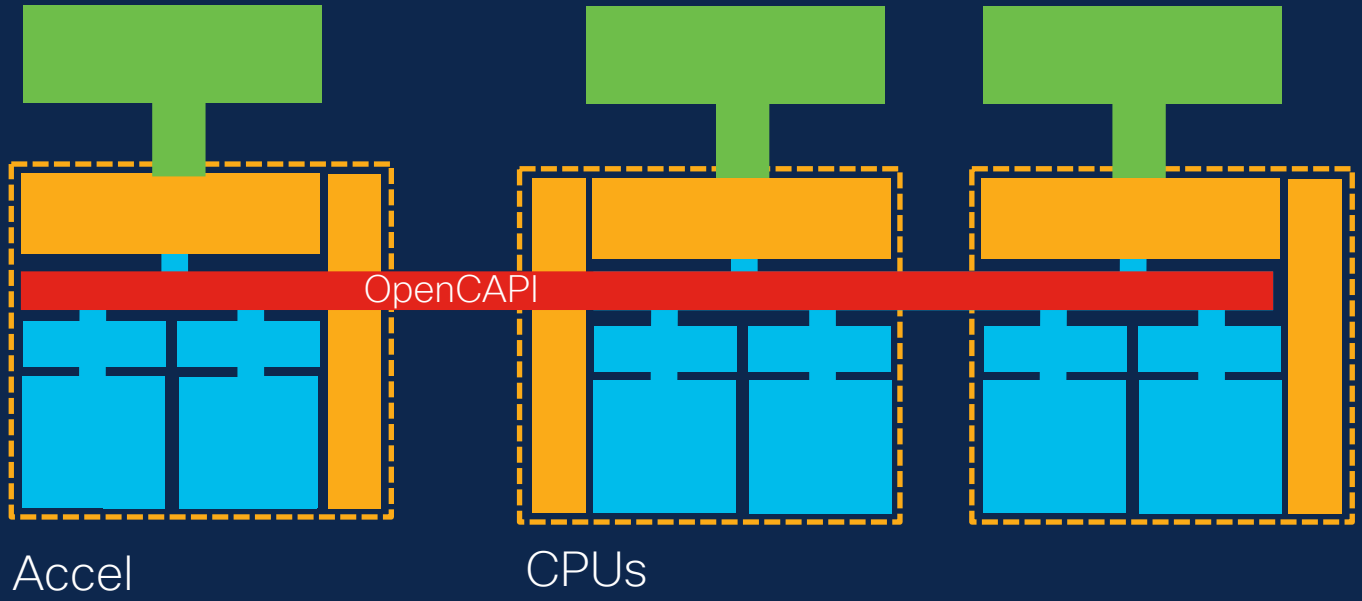
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	

Comparisons

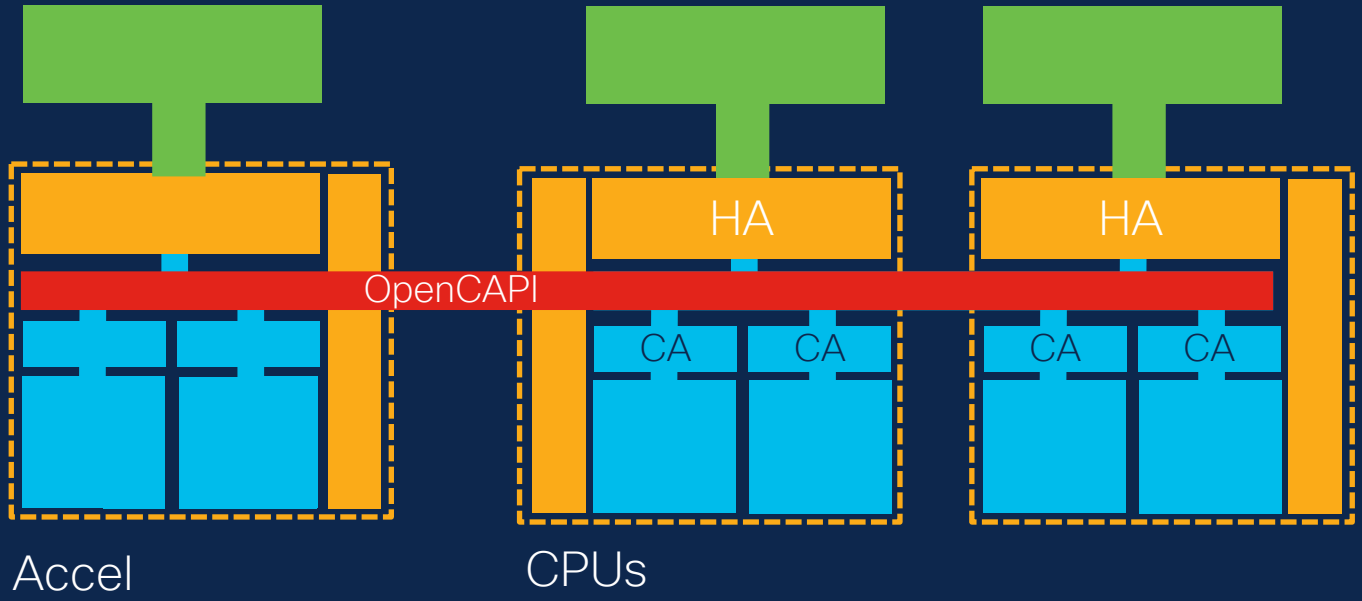
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	



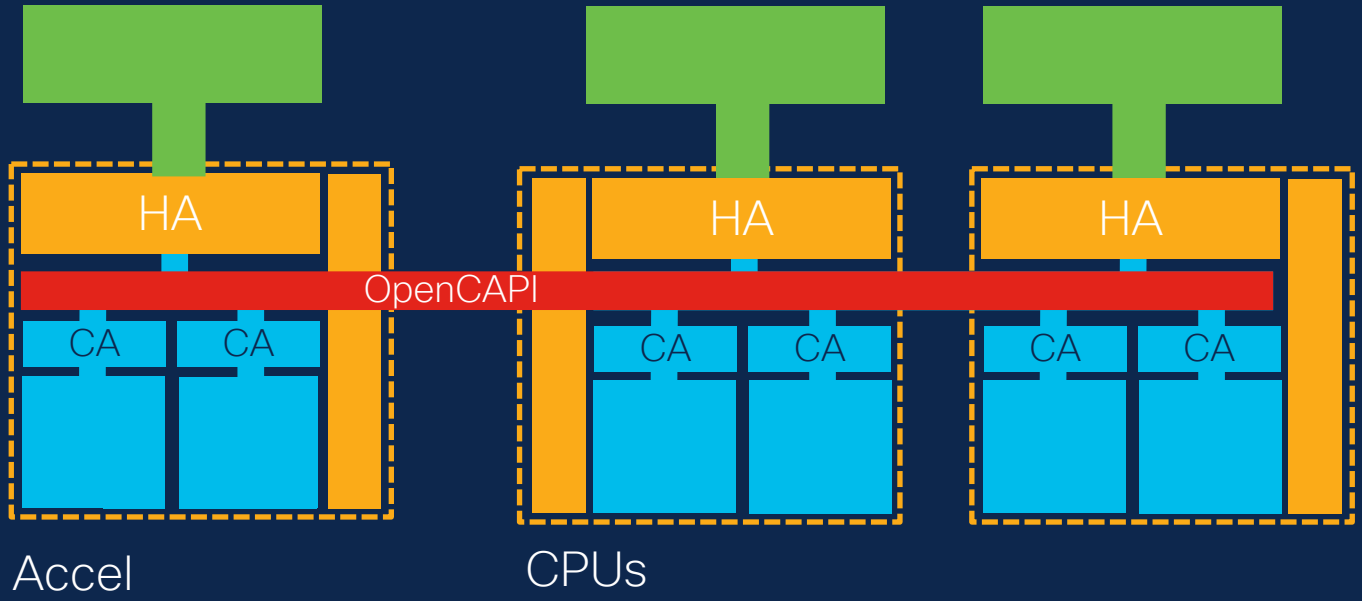
Full Cache Coherence



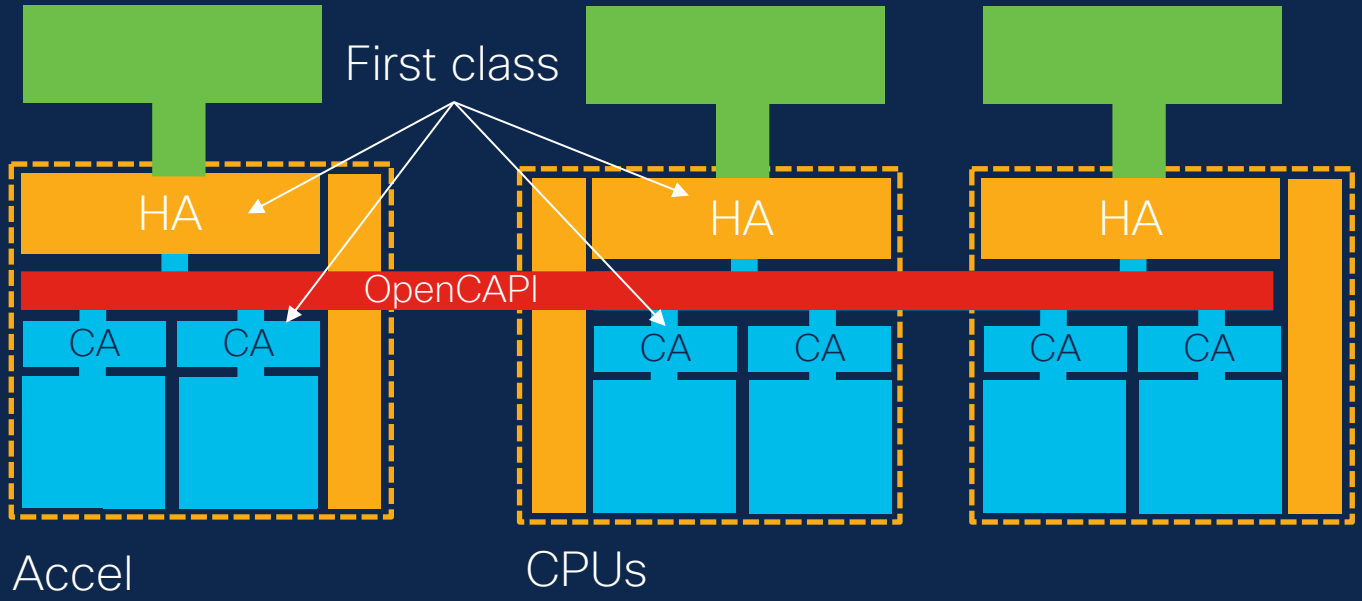
Full Cache Coherence



Full Cache Coherence



Full Cache Coherence



Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9

Who runs Power9?

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9

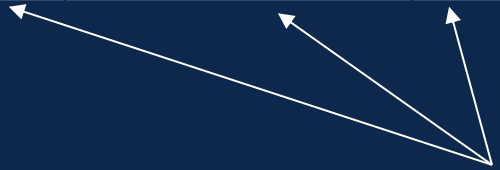
Who runs Power9?
Intel has 90%+ DC market share

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16			

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16			

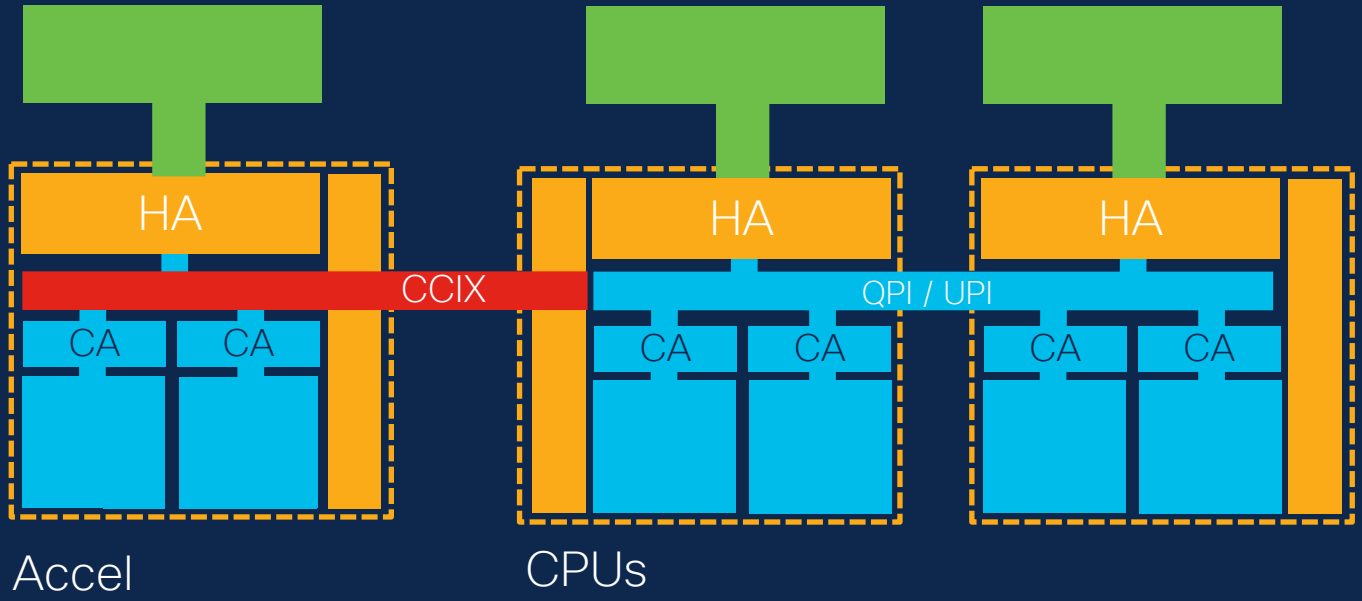


Reuse PCIe PHY chips,
retimers, connectors etc

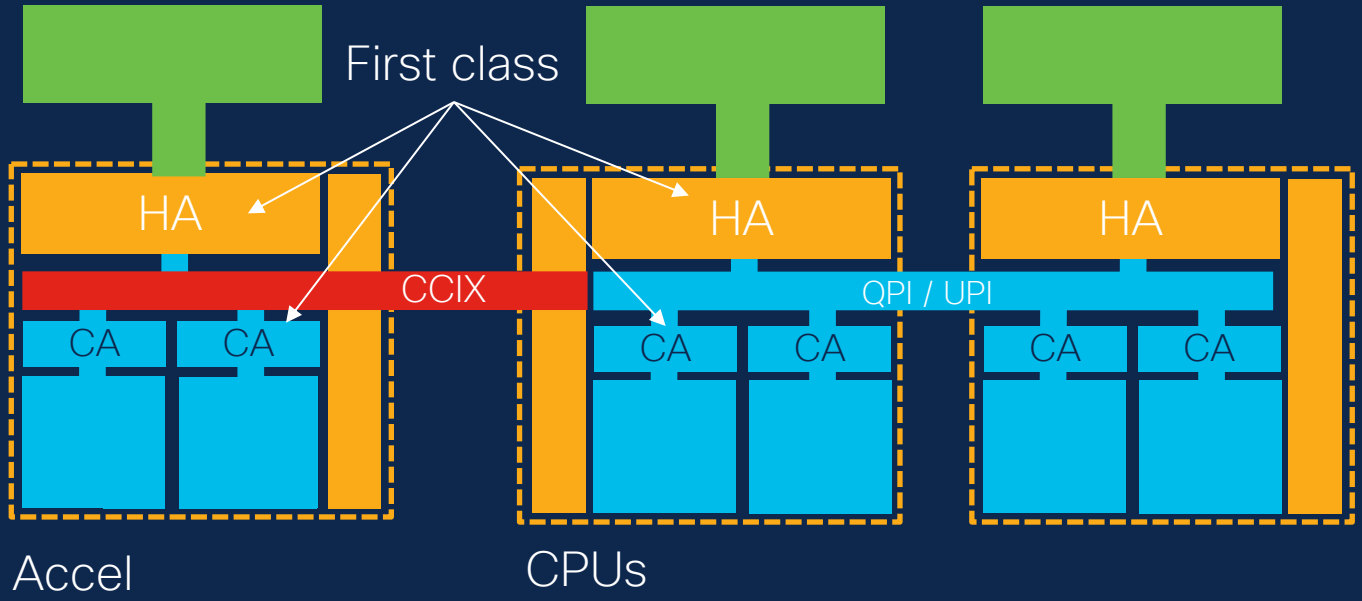
Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	

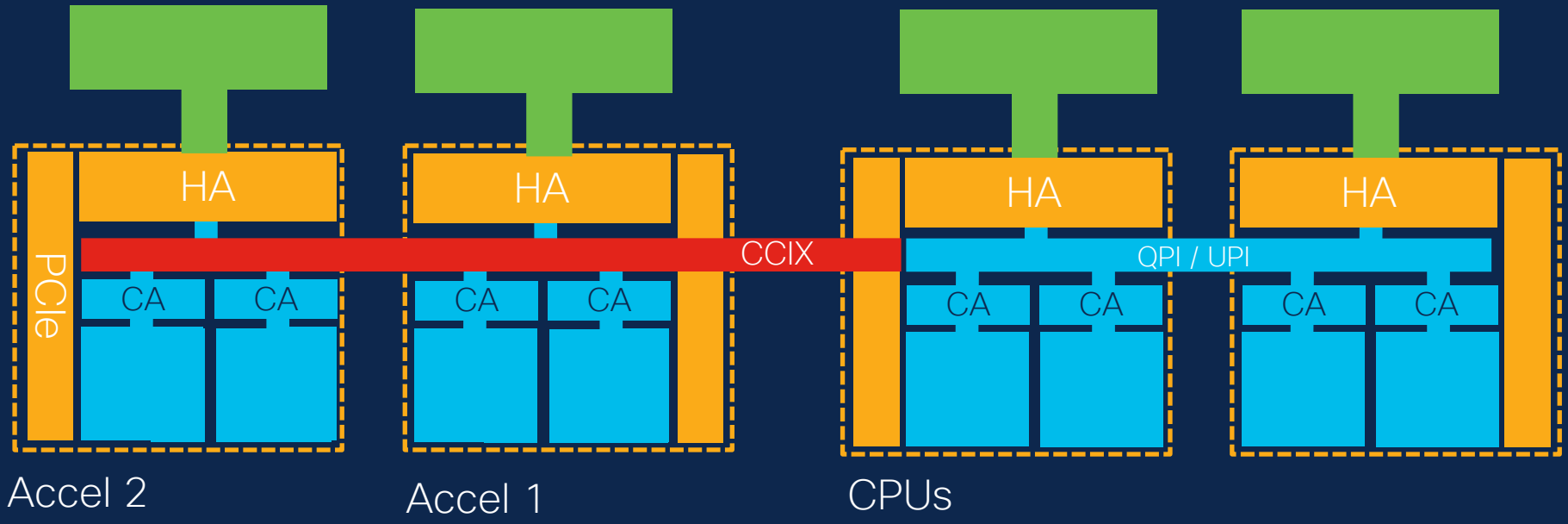
Full Cache Coherence



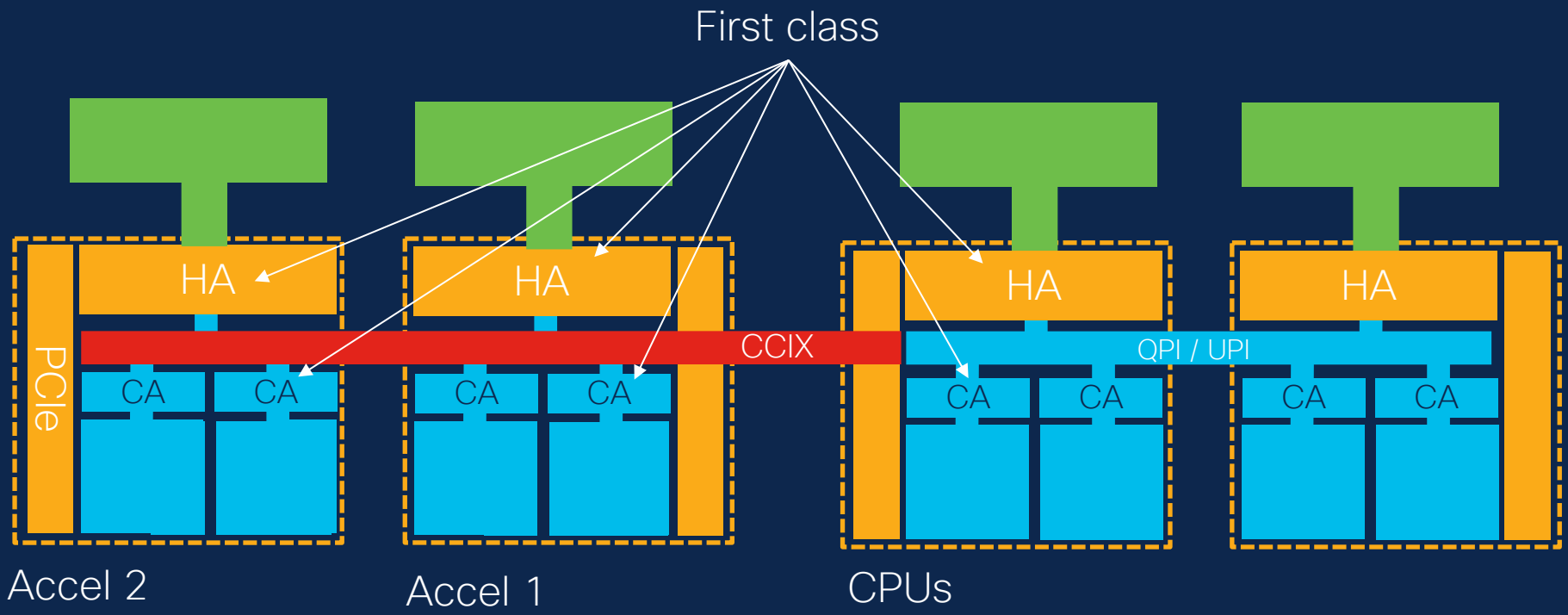
Full Cache Coherence



Full Cache Coherence



Full Cache Coherence



Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Perfect

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead

Killed by CXL
Why?

Comparisons

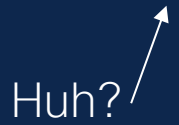
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512		

Comparisons

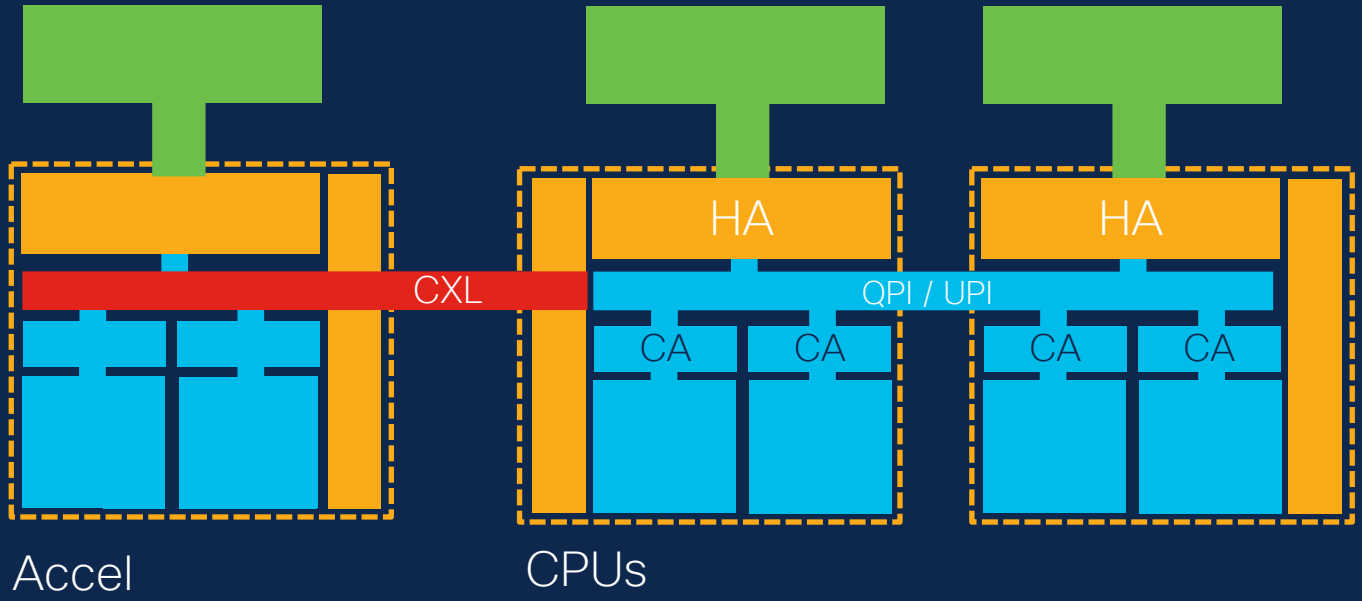
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	

Comparisons

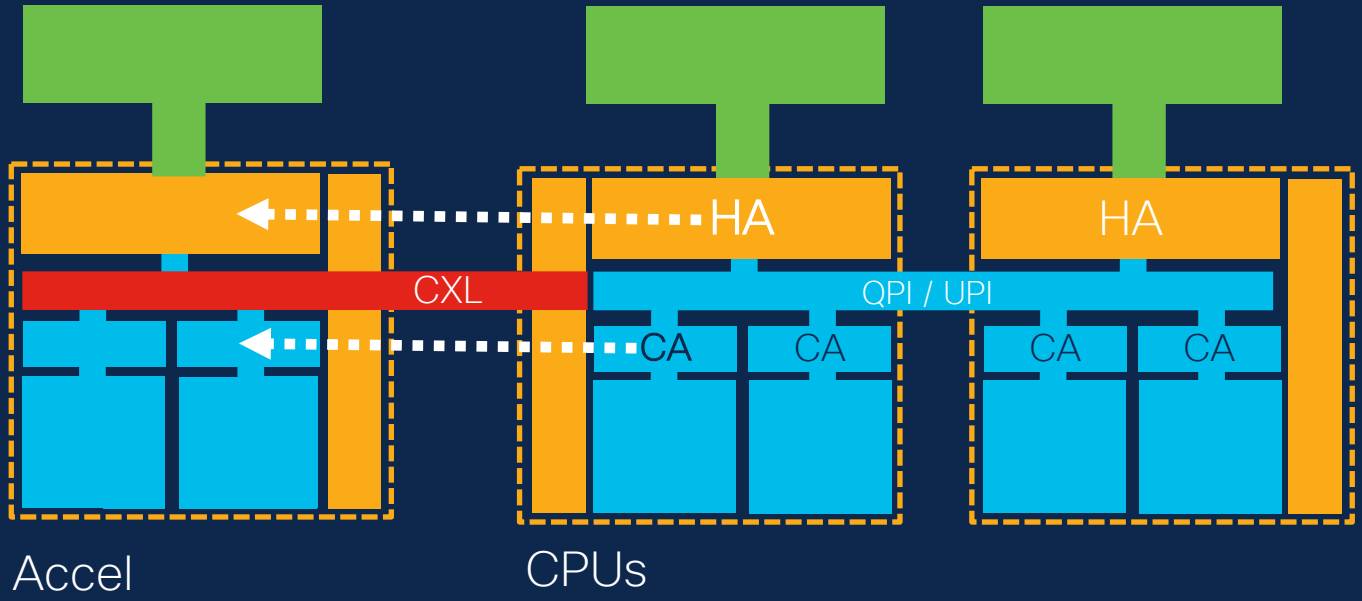
Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	



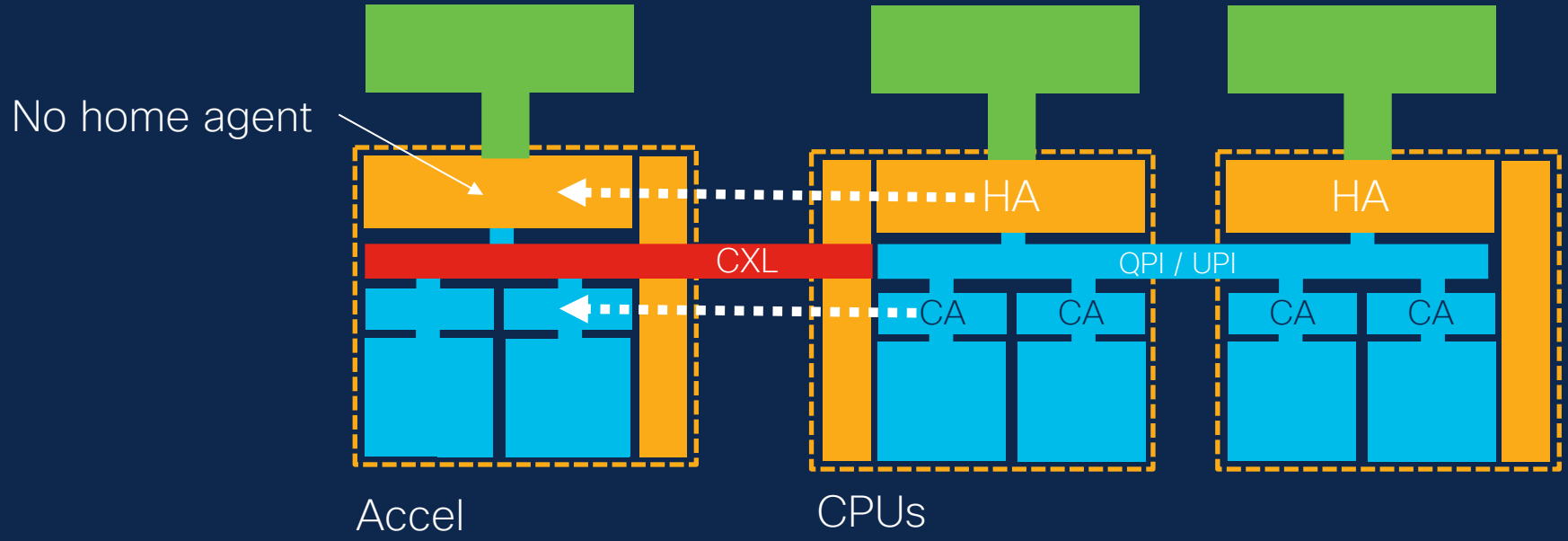
Asymmetric Cache Coherence



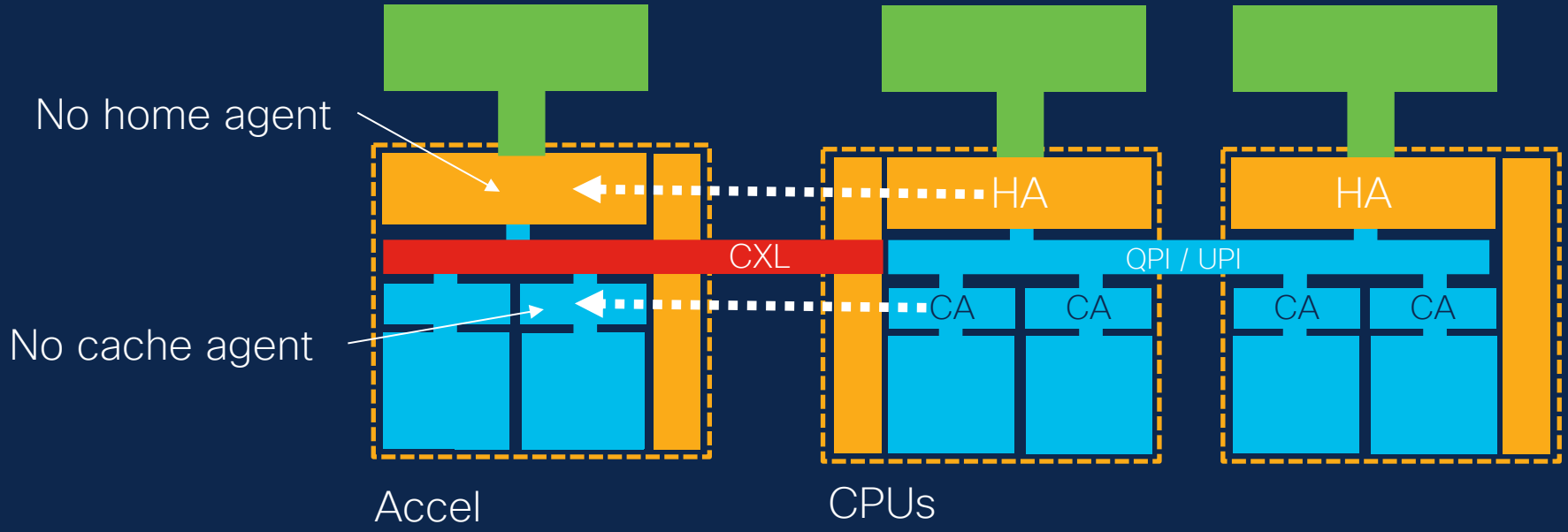
Asymmetric Cache Coherence



Asymmetric Cache Coherence



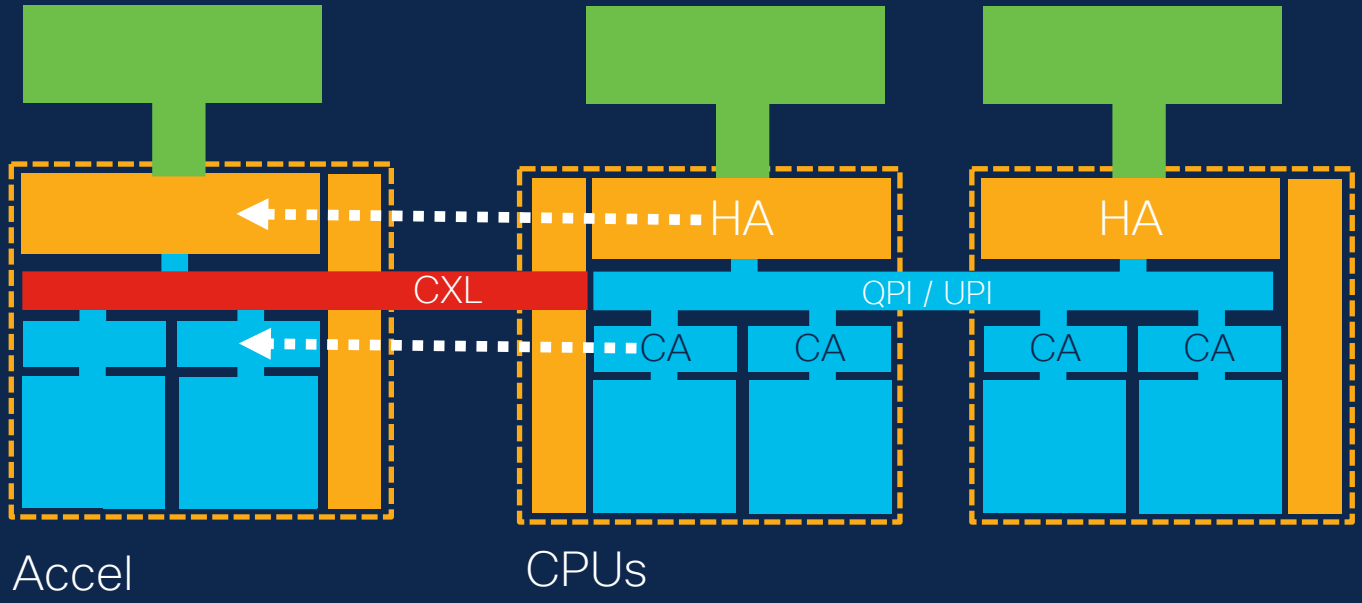
Asymmetric Cache Coherence



Asymmetric Cache Coherence

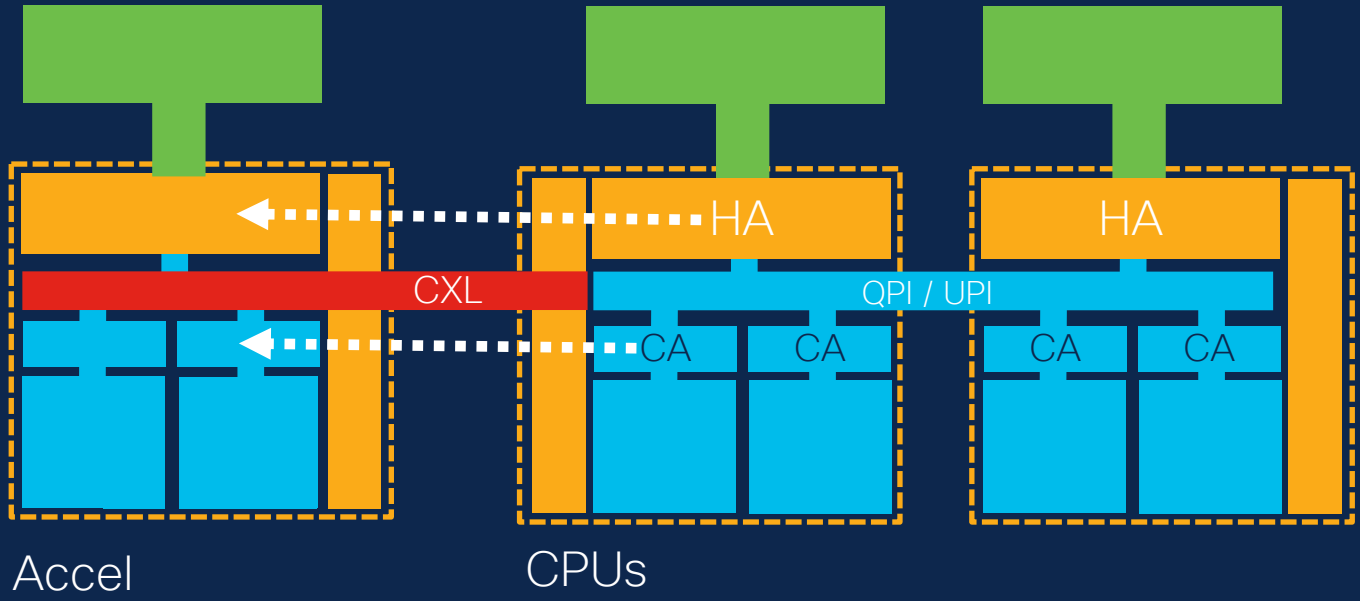
Pros:

- simplify accel



Asymmetric Cache Coherence

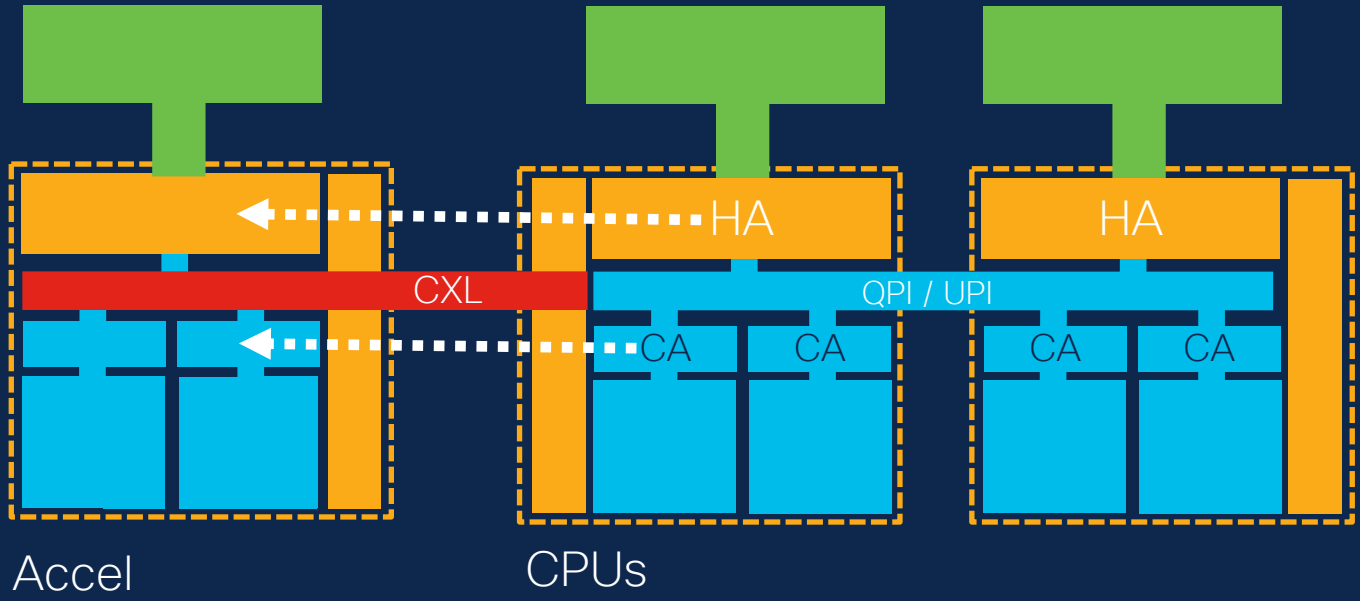
Pros:
- simplify accel
- sys perf



Asymmetric Cache Coherence

Pros:
- simplify accel
- sys perf

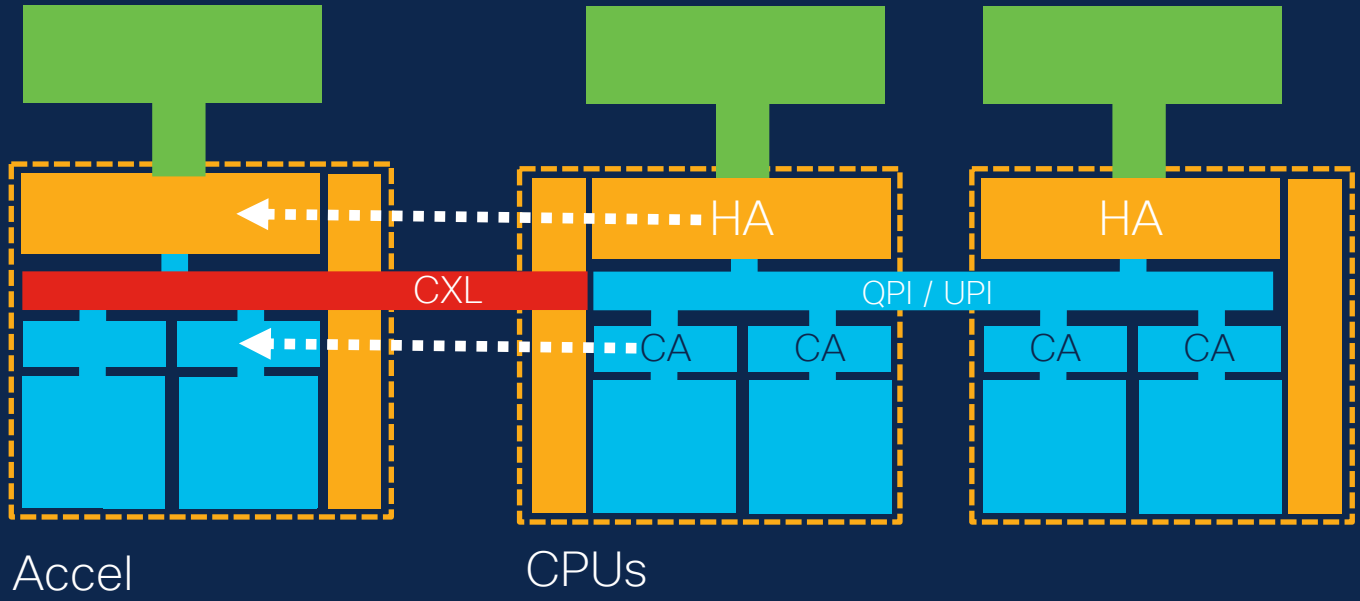
Cons:
- CPU deps



Asymmetric Cache Coherence

Pros:
- simplify accel
- sys perf

Cons:
- CPU deps
- Limited vis



Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel

AKA Winner

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel

AKA Winner

Intel has 90%+ DC market share!!

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3					

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	← Huh?				

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3					

← Short, medium and long range
(scales from chip-2-chip, to rack scale)

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	← Huh?			

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	← PAM4 50GbE			

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800	← Huh???	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800		

256x50G

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800		

256x50G = Never!

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800	Full	← Huh?

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800	Full	

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800	Full	Kitchen Sink

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800	Full	Kitchen Sink

Never going to happen

Comparisons

Standard	PHY	Link Rate (Gb/s)	Max Links	Datarate (Gb/s)	Coherence	IOW
Ethernet	802.3	100	8	800	none	Ubiquitous
PCIe 4.0	PCIe 4	16	16	256	none	Late
PCIe 5.0	PCIe 5	32	16	512	none	Early
OpenCAPI	OCAPI	25	32	800	full	Power9
CCIX	PCIe 5	32	16	512	full	Dead
CXL	PCIe 5	32	16	512	asymm	Intel
GenZ	PCIe 5.0 / 802.3	56	256	12,800	Full	Kitchen Sink

Questions?

Questions?

m0110@cisco.com