Unified Computing Technical Background

Even Solberg

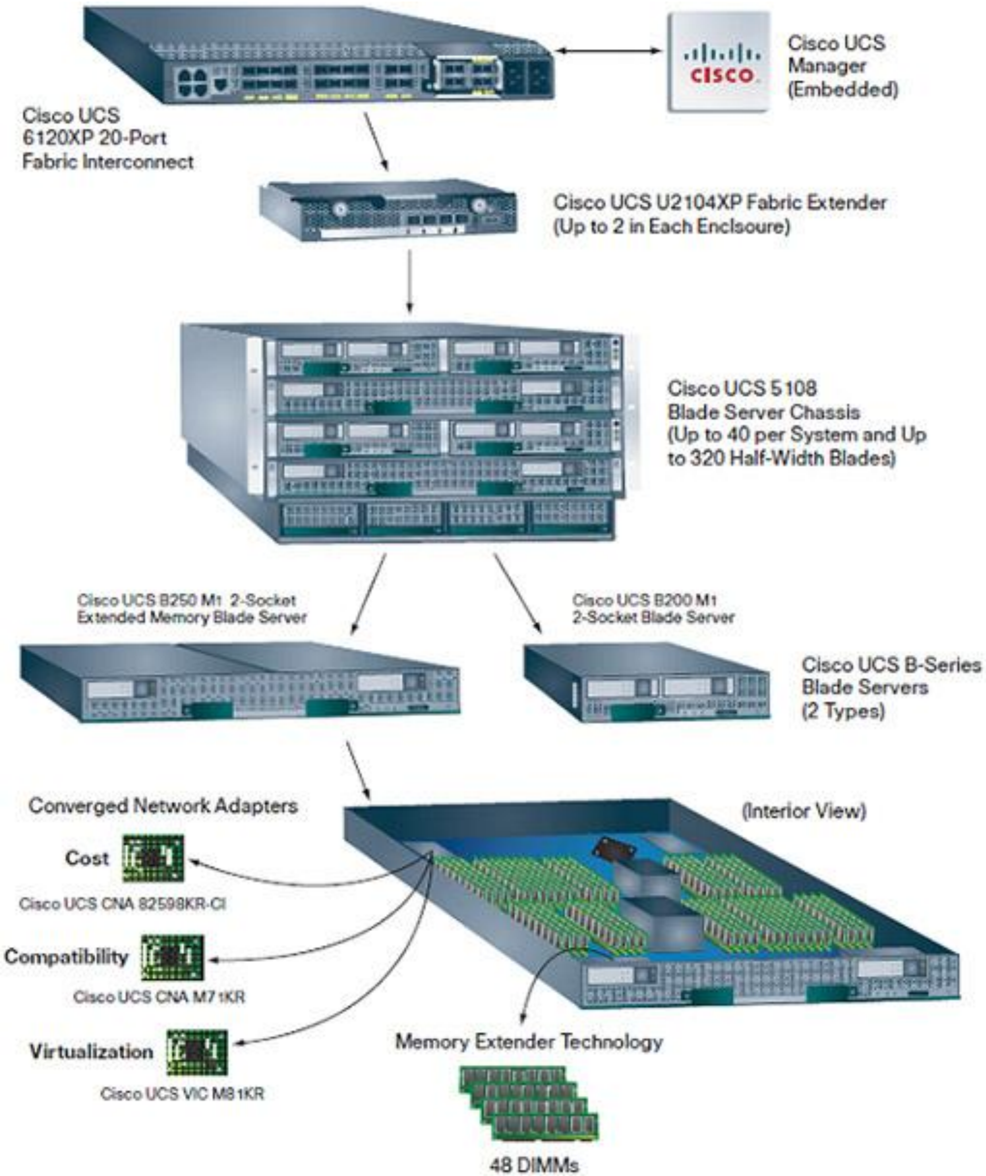
Agenda

- Background / Intro
- Memory Expansion
- I/O
- Fabric Extender Connections
- More Information

Unified Computing System

- A single system that encompasses:
  - –Network: Unified fabric
  - –Compute: Industry standard x86
  - –Storage: Access options
  - –Virtualization: optimized
- Unified management
  - –Dynamic resource provisioning
- Efficient Scale
  - –Same effort to manage 8 blades as 320 blades
- Lower cost
  - –Fewer servers, switches, adapters, cables
  - –Lower power consumption

Unified Computing System


Cisco UCS Manager (Embedded)
Cisco UCS 6120XP 20-Port Fabric Interconnect
Cisco UCS U2104XP Fabric Extender (Up to 2 in Each Enclsoure)
Cisco UCS 5108 Blade Server Chassis (Up to 40 per System and Up to 320 Half-Width Blades)
Cisco UCS B250 M1 2-Socket Extended Memory Blade Server
Cisco UCS B200 M1 2-Socket Blade Server
Cisco UCS B-Series Blade Servers (2 Types)
Converged Network Adapters
Cost
Cisco UCS CNA 82598KR-CI
Compatibility
Cisco UCS CNA M71KR
Virtualization
Cisco UCS VIC M81KR
(Interior View)
Memory Extender Technology
48 DIMMs

Memory Expansion

Memory Expansion

- With modern CPUs, memory becomes a bottleneck
  - Limits server performance
- Each server socket has a limited number of sockets and speeds it can connect to
  - Intel Xeon 5500 (Nehalem EP) supports up to 48 GB of memory per socket on 2 x DDR3 RDIMMS.
  - Current max RDIMM is 8 GB
- Nehalem can address larger memories
  - Electrical issues, memory density and details of processor pinout are the real limiting factors.

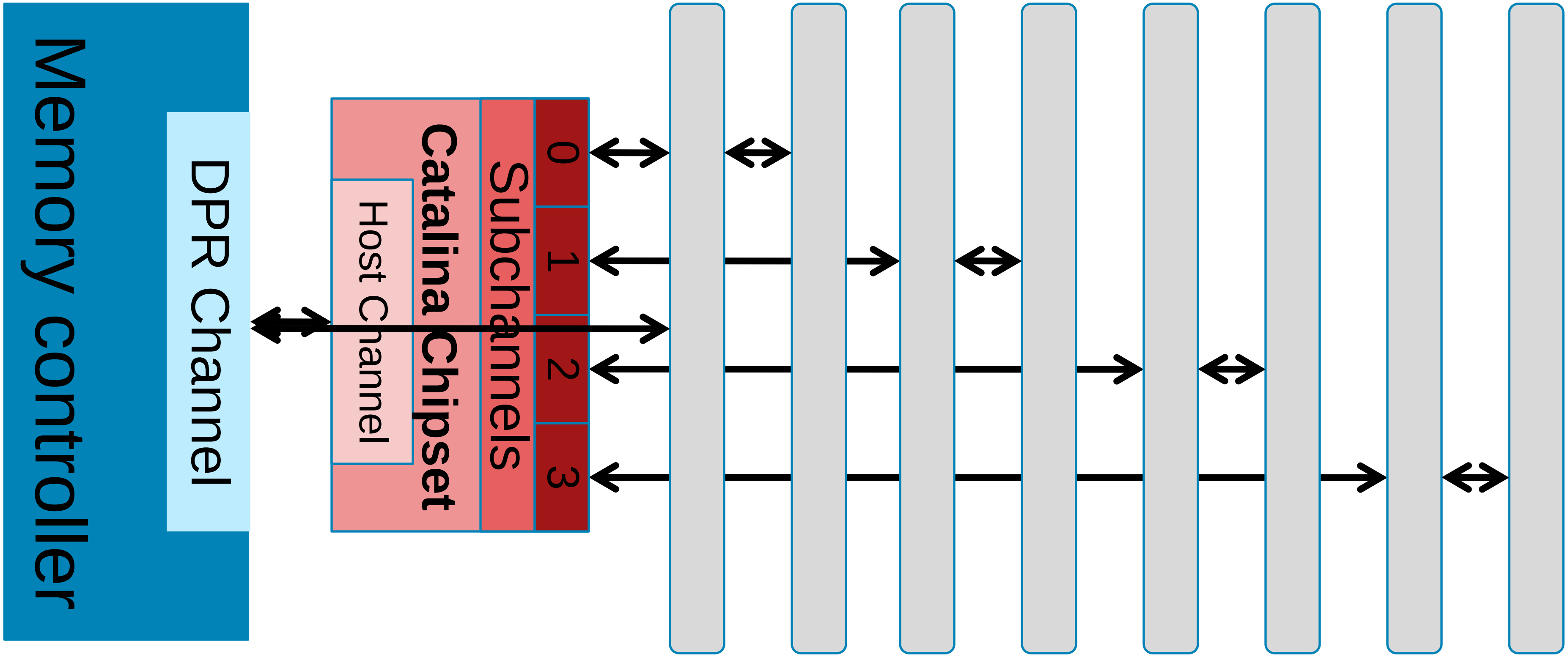
The UCS Approach

- Uses Cisco ASICs called the ”Catalina chipset”
  - Expands the number of memory sockets that can be connected to each single memory bus
  - ASICs are inserted between the processor and the DIMMs on the memory bus, minimizing electrical load
  - Bypasses the control signal limitations of the Nehalem CPU design
- Achieved memory expansion factor: 4x
  - Using 8 GB DIMMS, a Nehalem socket can connect up to 192 GB of RAM, and a dual CPU socket blade can host up to 384 GB of RAM.
  - DIMMs are unmodified JEDEC standard DDR3 RDIMMs.
  - Expansion is done at electrical level and is completely transparent to operating system and applications
  - BIOS is extended to initialize and mointor ASICs and perform error reporting.

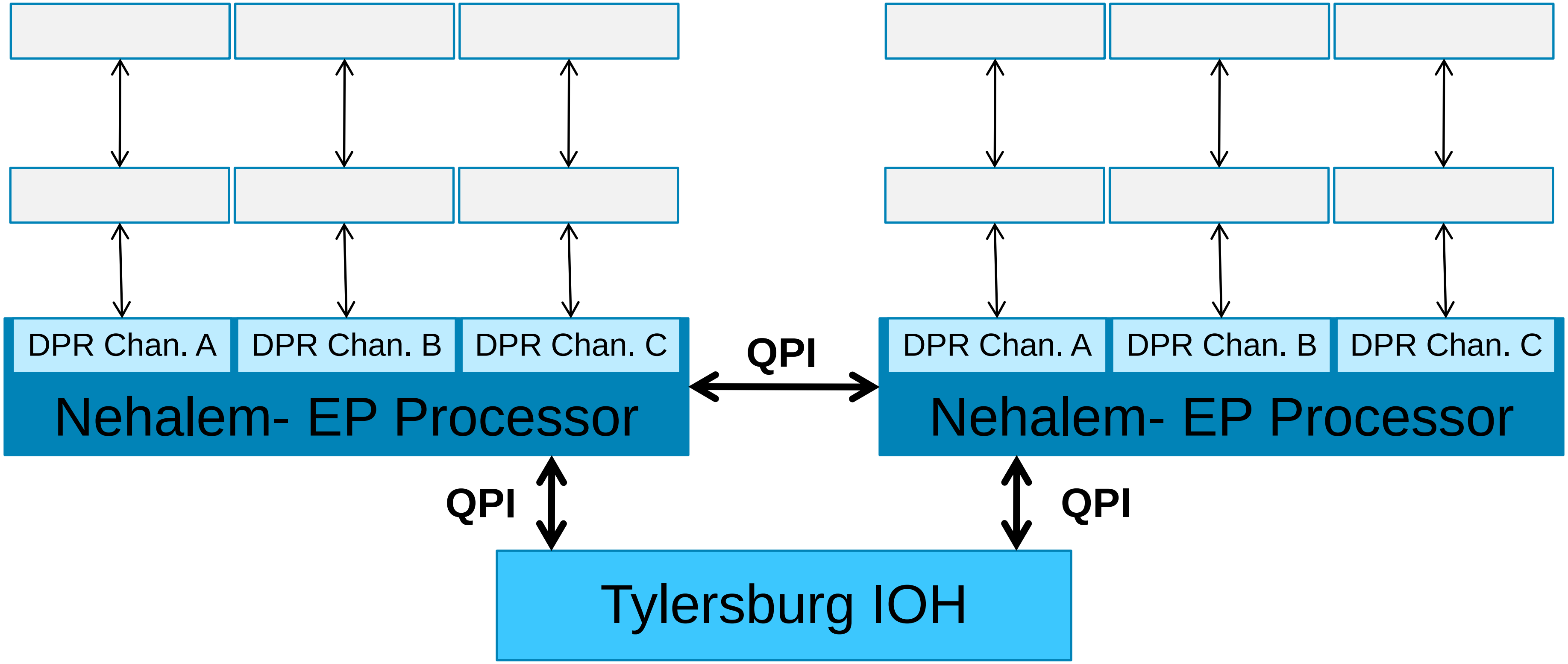
Catalina Usage

- The Catalina chipset advantage can be exploited in two ways:
  - Building the largest memories possible using highest density DIMMs. Used for memory-hungry apps such as database, engineering tools like circuity synthesis and simulation, as well as virtualization.
  - Building a medium to large memory configuration using inexpensive DIMMs

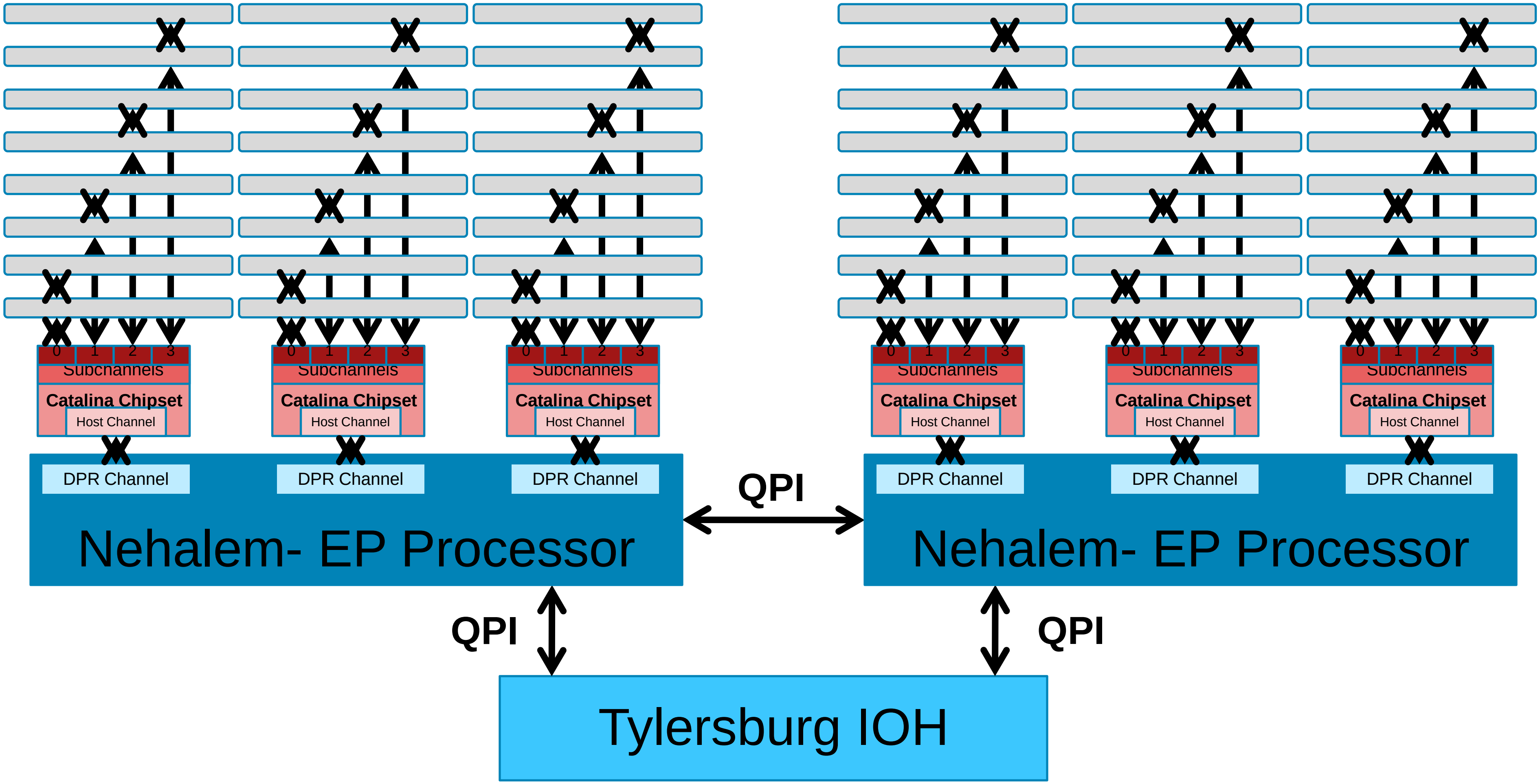
The Catalina ASIC chipset



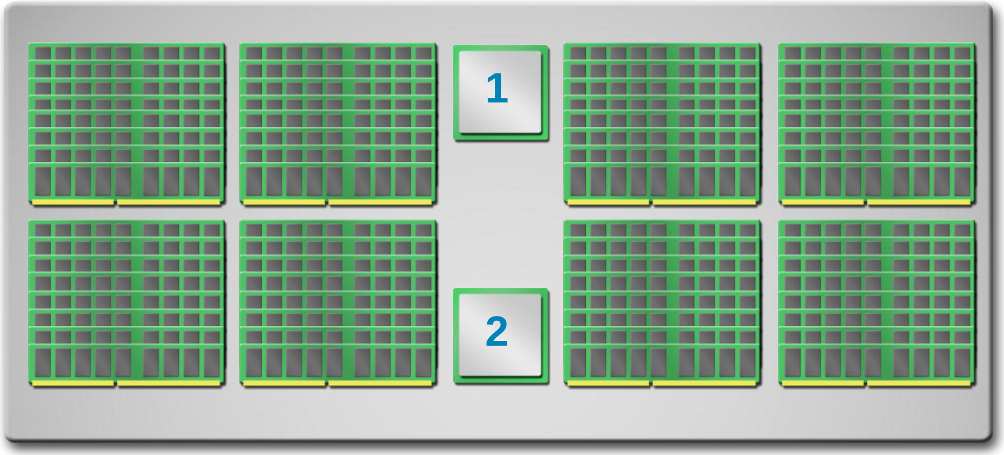
Nehalem system without memory expansion



Nehalem system with memory expansion



Savings with Expanded Memory



Memory Capacity & Pricing

Capacity	Speed	DIMMs	Cost*	DIMMs	Cost*	Savings
64 GB	1033Mhz	4x 8GB & 8x 4GB	$3,544	32x 2GB	$2544	28%
96 GB	1033Mhz	12x 8GB	$6000	48x 2GB	$3,816	36%
192 GB	1033Mhz	12x 16GB**	$N/A	48x 4GB	$9,264	N/A
384 GB	1033Mhz	12x 32GB**	N/A	48x 8GB	$24,000	N/A

*** DDR2 pricing as of 3/09**

**** Nonexistent or Nonstandard DIMM (MetaRAM or other). Est. MetaRAM pricing.**

I/O

Cisco Palo

- CNA developed by Cisco
- 2 x 10 GE + Up to 128 vNICs
- Currently only exists in the UCS solution
- Only available as a mezzanine card
- Palo ASIC uses 8 W power, mezzanine card 18 W
  - Max established by PCIe standard: 25 W

Palo System Level Features

- 32 Gb/s raw bandwidth, more than enough for 2 x 10 Gb/s ports
- I/O consolidation and Unified Fabric
- DCE compliant Ethernet interface
- Supports Active/Active & Active/Standby
- Create I/O devices on demand for virtualization
- Supports kernel and hypervisor bypass
- Low latency and high bandwidth
- SRIOV compliant hardware
- Native support for VNTag
- Native support for FCoE

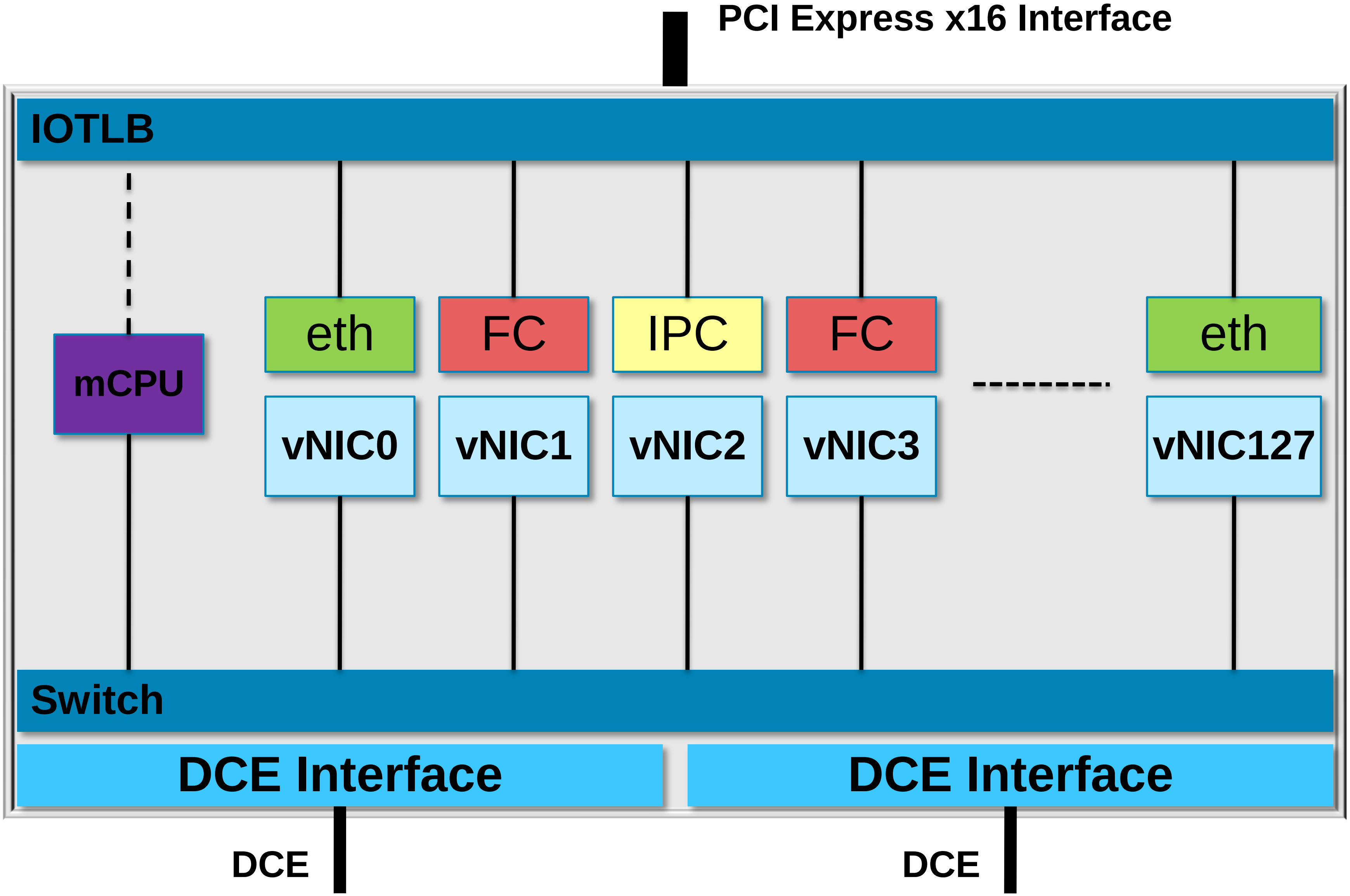
Palo System Level Features

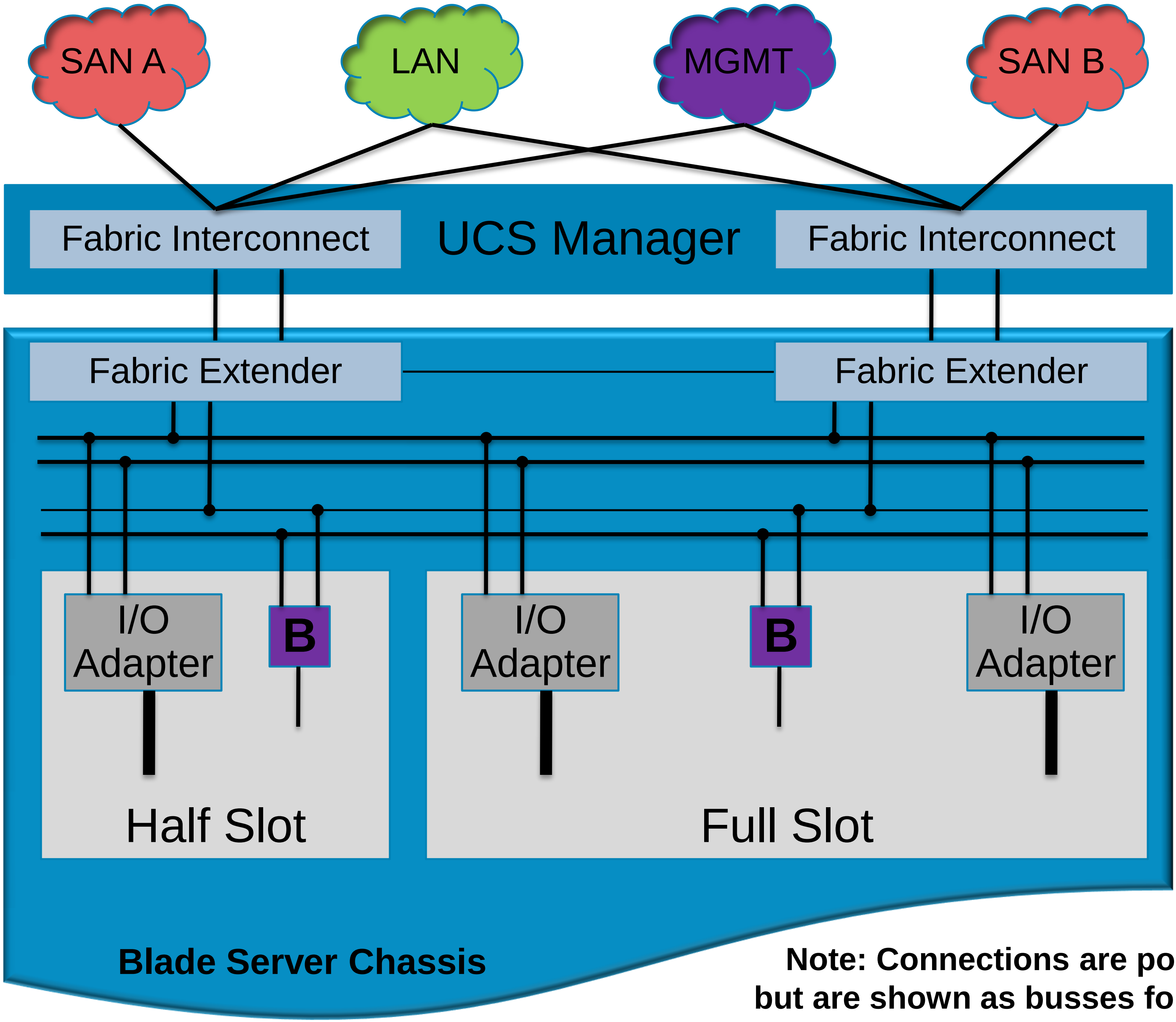
- Has an internal switch that allows it to work as a Virtual Ethernet Bridge (VEB)
  - In the adapter
  - In the switch through the use of VNTag
- Management CPU creates vNICs that are seen by the OS and Hypervisor as regular PCI devices
- Supports up to 128 vNICs
  - Can be either Ethernet NIC or FC HBA
  - Other types are supported in HW, but not used

Palo Preferred Mode of Operation

- Palo assigns a VNTag to each vNIC
- Uses an external switch to switch between them
- VNTag also provides traffic separation between VMs that share the same Palo adapter

Palo Logical Diagram





SAN A

LAN

MGMT

SAN B

Fabric Interconnect

UCS Manager

Fabric Interconnect

Fabric Extender

Fabric Extender

I/O Adapter

B

I/O Adapter

B

I/O Adapter

Half Slot

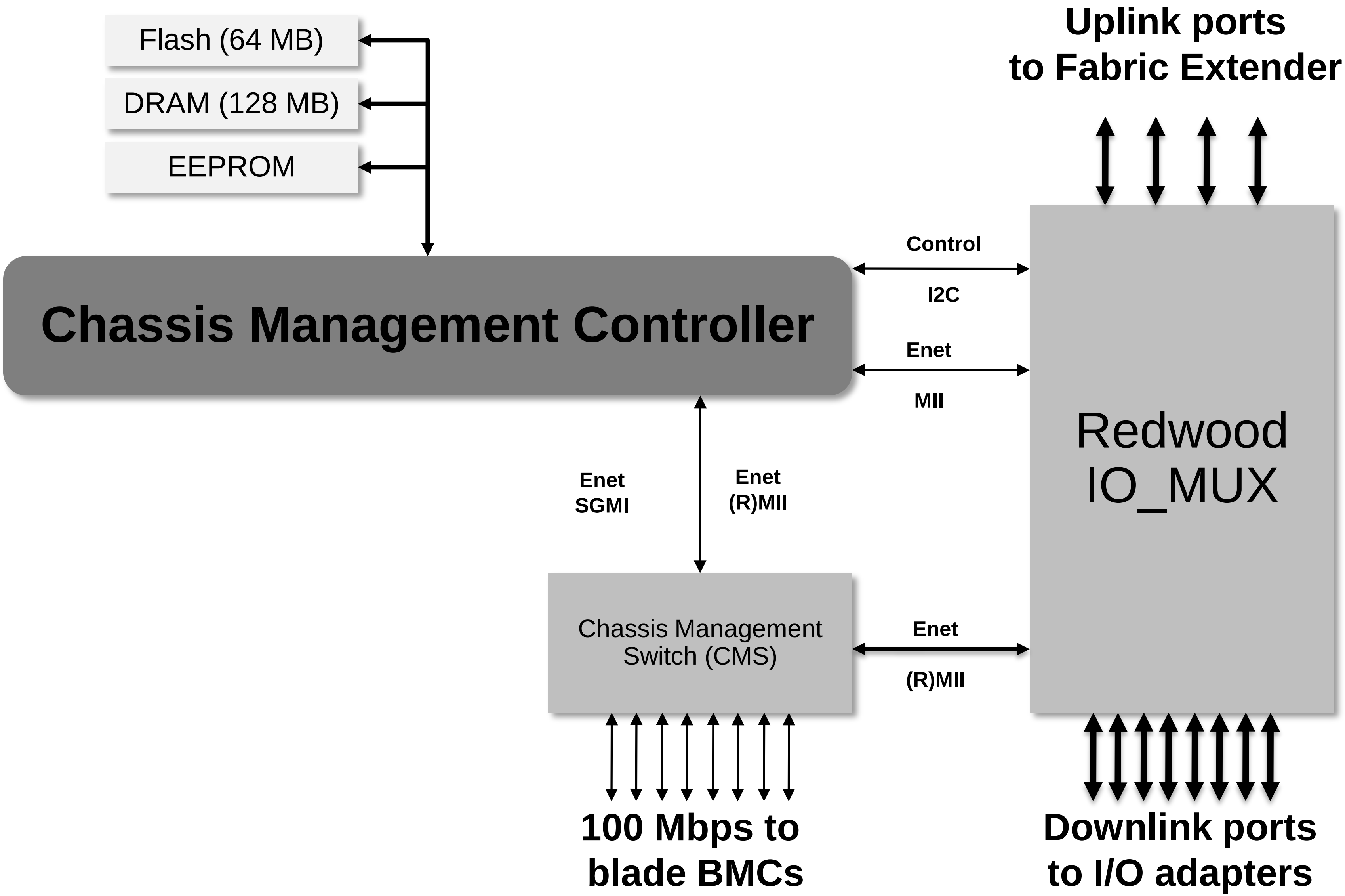
Full Slot

Blade Server Chassis

Note: Connections are point to point, but are shown as busses for simplicity.

Fabric Extender

Fabric Extender Block Diagram



Redwood IO_MUX

- Bridge between server blades and Fabric Interconnect
- The ASIC that implements the dataplane of the Fabric Extender
- Provides:
  - 8 x 10GE external downlink ports to server blades
  - 4 x 10GE external uplink ports to the fabric interconnect
  - 1 x 1GE internal port to connect to the CMS
  - 1 x 100Mbps internal port towards the CMC
- By default mezzanine adapters installed on server blades get pinned to uplinks in a pre-determined fashion.

Chassis Management Controller – CMC

- Processor embedded in the Fabric Extender
- Interacts with the UCS manager and Baseboard Management Controllers (BMCs) on the server blades.
- Admin does not interact directly with CMC – only through UCSM
- CMC main function is to provide overall chassis discovery and management and to report the result to the UCSM.

Chassis Management Controller – CMC

- Implements 7 main functions:
  - Controls the chassis fan
  - Monitors and logs fan speed
  - Monitors and logs ingress and egress temperatures
  - Powers up & down power supplies, monitoring & logging voltages, currents and temperatures inside the chassis
  - Detects presence, insertion and removal of UCS blades
  - Reads the IDs of the chassis, UCS blades, and Fabric Extenders
- CMC does *not* manage UCS blades.
- Will cluster if 2 are present in a chassis, in active/passive configuration.

Chassis Management Switch – CMS

- Provides connectivity to the BMC on each server blade
- 8 x 100 Mbps + 1 x 1GE connection available to CMS
- Each slot has a dedicated 100 Mbps connection to the blade BMC
- 1 x 1GE is used to connect CMS to Redwood IO_MUX
- CMS is an unmanaged switch that requires not config

More Info

Recommended Reading

Project California:
a Data Center Virtualization Server
UCS (Unified Computing System)
Silvano Gai
Tommi Salli
Roger Andersson

- Silvano Gai, Tommi Salli, Roger Andersson, *“Project California: a Data Center Virtualization Server - UCS (Unified Computing System)”*
- ISBN 978-0-557-05739-9
- Cisco Systems, April 2009
- http://www.lulu.com/content/6579826

CISCO