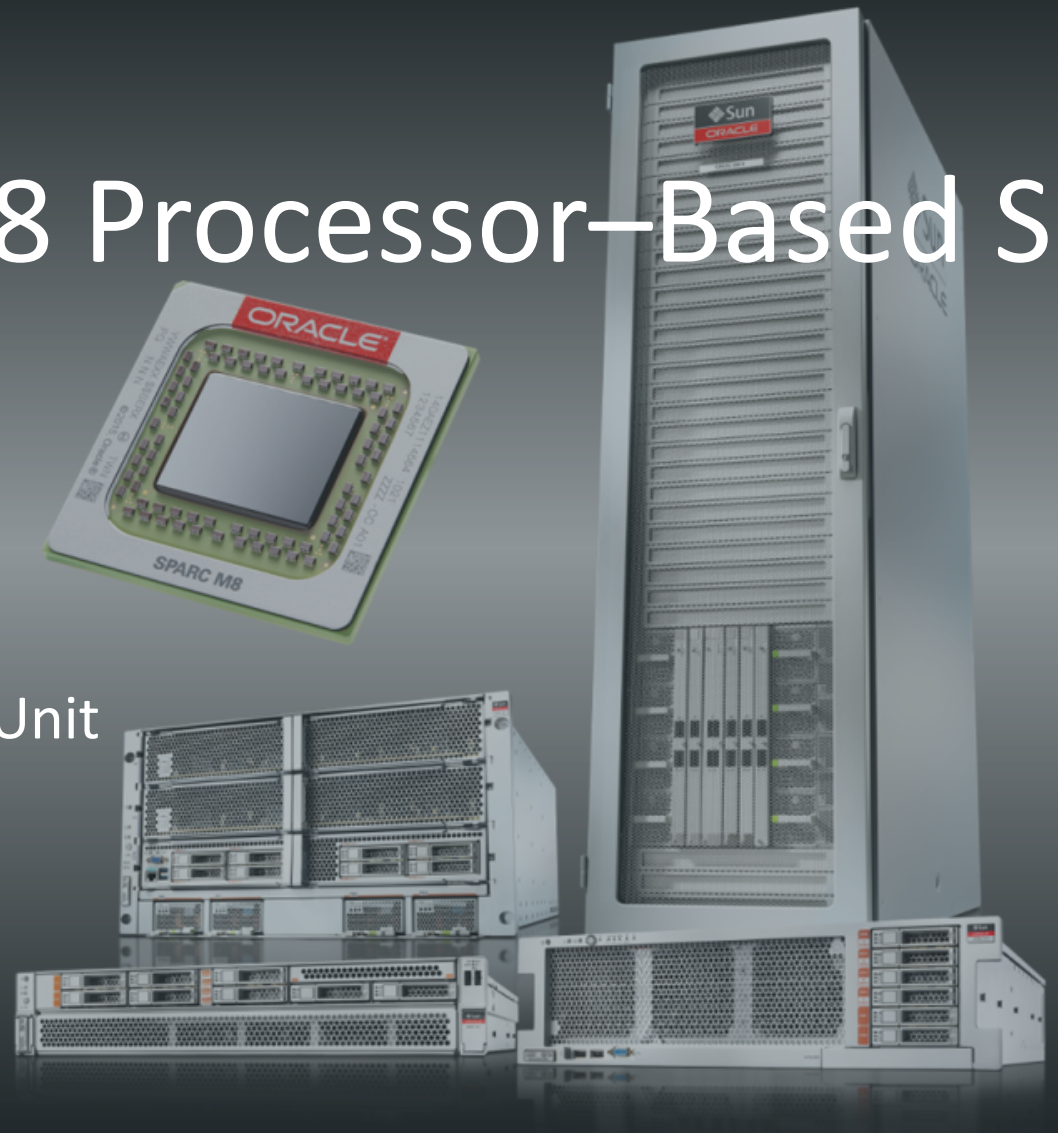

ORACLE®

Oracle’s SPARC M8 Processor–Based Servers

Technical Deep Dive



Presenter’s Name
Presenter’s Title
Organization, Division, or Business Unit
Month 00, 2018

ORACLE

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Agenda

1. Overview
2. System Details
3. New Technologies
4. RAS
5. Platform Management
6. Virtualization
7. Summary

SPARC M8 Processor–Based Servers

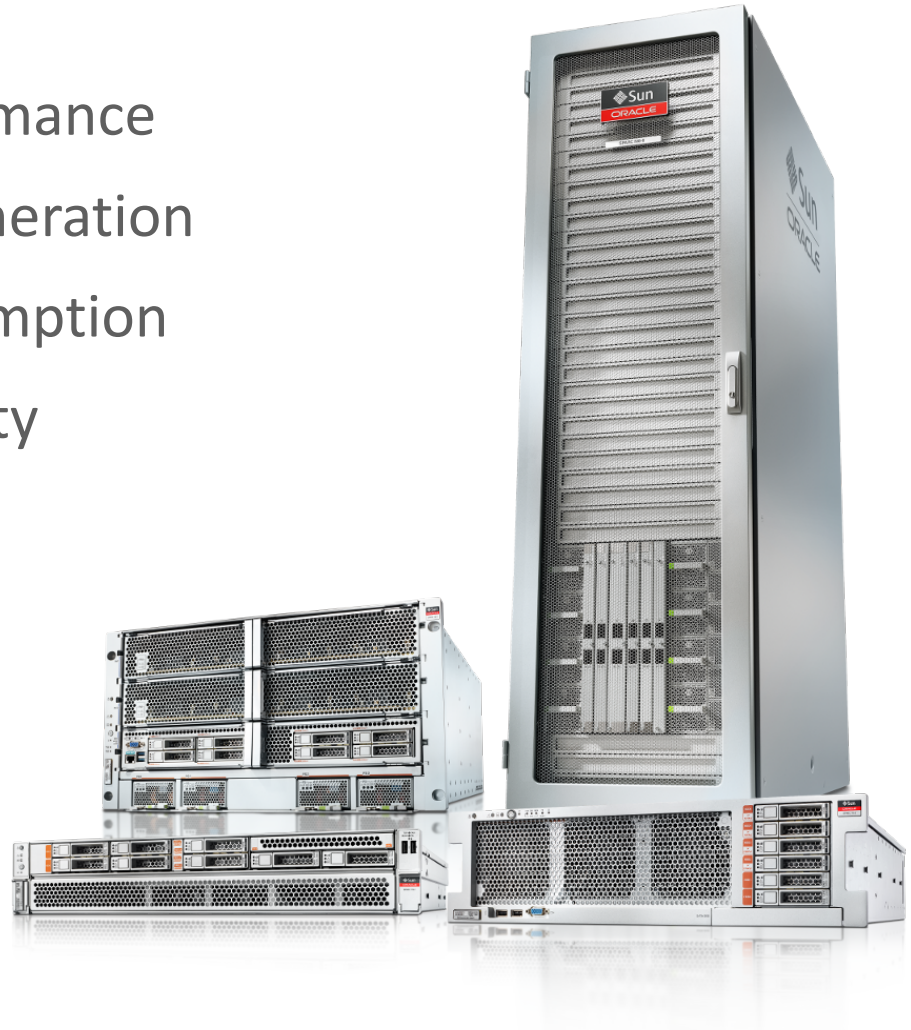
	SPARC T8-1	SPARC T8-2	SPARC T8-4	SPARC M8-8
Processors	1	2	2 or 4	Up to 8 [1]
Max Cores	32	64	128	256
Max Threads	256	512	1,024	2,048
Max Memory [2]	1 TB	2 TB	4 TB	8 TB
Form Factor	2U	3U	5U	Rack / 10U
Domaining	Logical domains (LDoms)	LDoms	LDoms	LDoms, PDoms [1]

(1) Factory configured with one (up to 8 processors) or two (up to 4 processors each) static physical domains (PDoms)
(2) Maximum memory capacity is based on 64 GB DIMMs

SPARC M8 Processor–Based Servers

Common Characteristics

- New 5.0 GHz SPARC M8 processor for higher performance
- Revolutionary Software in Silicon, now in second generation
- DDR4-2400 memory: Faster and lower power consumption
- DIMM sparing: Standard feature for higher availability
- New larger NVMe flash drives for high performance
- x16 capable PCIe 3.0 slots: More I/O throughput
- On-board SAS3 (for SPARC T8-1, SPARC T8-2, and SPARC T8-4)
- Oracle Solaris 11.3 SRU 23 or later required
  - Oracle Solaris 10 1/13 and later in guest domains



SPARC T8 and M8 Systems: Software

Take full advantage of second generation Software in Silicon

- Oracle Solaris 11.3 SRU 23 or later required
  - Oracle Solaris 10 9/10 and later in guest domains
- Oracle Database 12c (12.2.0.1 + patches)[1]
  - In-Memory Query Acceleration and In-Line Decompression available with Oracle Database In-Memory option
  - Encryption acceleration available with the Oracle Database Transparent Data Encryption (TDE) option
  - Supported through Database Bundle Patch
  - In-Line Decompression is applicable to OZIP compression
  - Oracle Numbers acceleration
- Oracle Solaris Studio 12.4
- Oracle Enterprise Manager 13c
- Oracle Enterprise Manager Ops Center 12c (12.3.3)
- Oracle Solaris Cluster
  - Version 4.3 SRU 7 (w/ Oracle Solaris 11)
  - Version 3.3U2 + KU Patch 150400-49 (w/ Oracle Solaris 10)

ORACLE SOLARIS

ORACLE DATABASE

ORACLE SOLARIS STUDIO

ORACLE ENTERPRISE MANAGER 13c

ORACLE ENTERPRISE MANAGER OPS CENTER 12c

1) See notes

Java Support for Oracle Solaris for SPARC Servers

Java SE Development Kit and JVM*

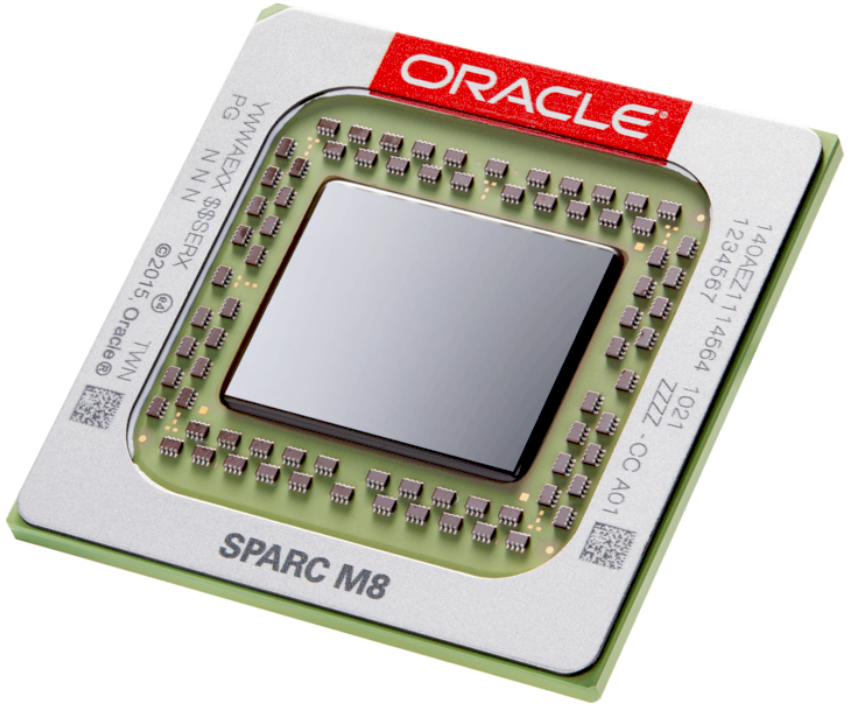
	Java 8	Java 7	Java 6	Java 5	Java 4
Solaris 11	8u60 b27	7u85 b33	1.6.0_141	1.5.0_85 [1]	1.4.2_42 [1]
Solaris 10	8u60 b27 [2]	7u85 b33 [2]	1.6.0_141	1.5.0_85 [1]	1.4.2_42 [1]
Solaris 9 [1]	Not supported	Not supported	1.6.0_141	1.5.0_85 [1]	1.4.2_42 [1]
Solaris 8 [1]	Not supported	Not supported	1.6.0_141	1.5.0_85 [1]	1.4.2_42 [1]

- Java version listed is minimum version required
- You must install the minimum version of Solaris required for the platform as well as specified for the Java build
- [1] The versions of Java and Solaris listed which are past the EOSL are listed for completeness, this in no way constitutes a change to that policy or those dates
- [2] Requires Solaris 10 Update 9 or later

* This applies to SPARC T8, M8, T7, M7, and S7 servers

SPARC M8 Processor Overview

- Second Generation Software in Silicon features NEW
  - Silicon Secured Memory
  - Enhanced Encryption Acceleration
  - In-Memory Query Acceleration
  - Enhanced In-Line Decompression
  - Oracle Numbers Acceleration
- 5th Generation SPARC Core NEW
  - 32 cores at 5.0 GHz, 256 hardware threads
  - Multiple performance enhancements
- Dynamic threading supporting critical threads
- Up to 16 DDR4 DIMMs per processor
- Up to 8-way fully connected glueless SMP
- Technology: TSMC 20nm



Advancing the State of the Art

Key Microprocessor Capabilities

	SPARC M8	SPARC M7	SPARC T5
Enhanced Cores	32 (5th Gen)	32 (4th Gen)	16 (3rd Gen)
Larger Cache per Core	2 MB	2 MB	0.5 MB
More Memory Bandwidth	185 GB/sec	160 GB/sec	79 GB/sec
Faster Memory Access	127 ns	131 ns	163 ns
Leading I/O Bandwidth	145 GB/sec	145 GB/sec	32 GB/sec
Faster Processor Frequency	5.0 GHz	4.13 GHz	3.60 GHz

2nd Generation Software in Silicon

Continuing the Revolution

	SPARC M8: 2nd Generation	M7, S7: 1st Generation
Data Analytics Accelerators (DAX)	Native OZIP format added	Native NZIP-Huffman format, conversion from OZIP
	Translate table expanded to 64K entries	Translate table holds 32K entries
	16 byte alignment on Translate and OZIP tables	64 byte alignment on Translate and OZIP tables
	Fixed width bit packed formats extended up to 23 bits	Fixed width bit packed formats up to 15 bits
	Pipelined operations	-
	Runs at 2.2 GHz	Runs at 1.8 GHz
Core	Oracle Number acceleration units	-
	24 New HPK data instructions run on two VIS pipes	-
Cryptography	SHA-3 added	15 cyphers and hashes accelerated

SPARC M8 Processor: Software in Silicon Features

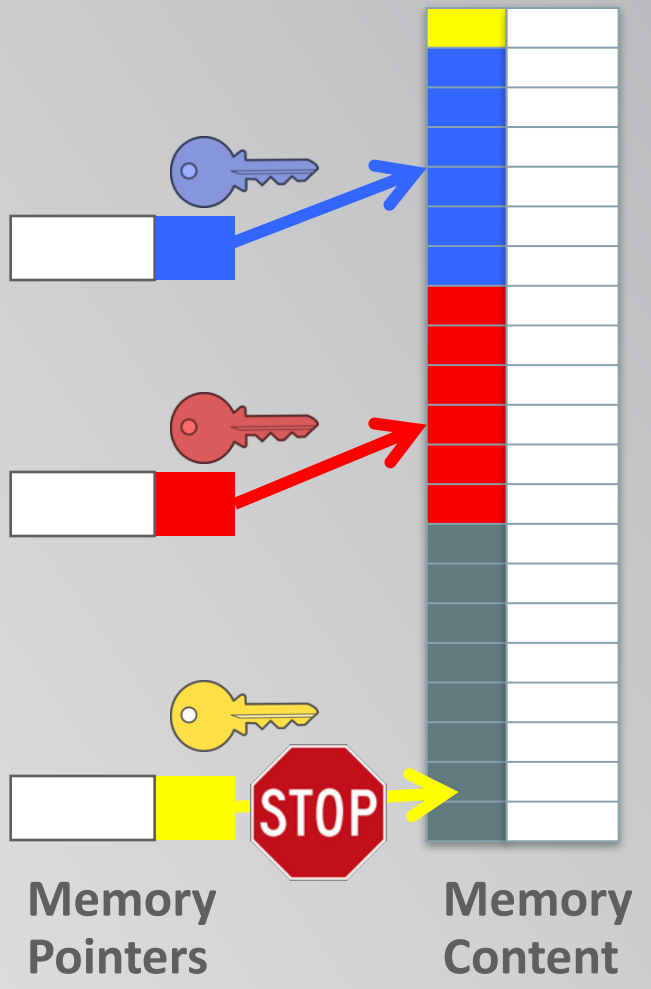
Get the Benefits, Without the Performance Overhead

Feature	Benefits
Silicon Secured Memory	• Prevention of malicious memory access attacks • Protection against silent data corruption in memory • Faster and higher-quality code development
Cryptographic Acceleration	• End-to-end security at full speeds, everything always encrypted • Secured network and database transactions
In-Memory Query Acceleration	• Analytics performed on live OLTP data • Reduced number of systems needed for analytics and reporting
In-Line Decompression	• Reduced memory footprint and lower TCO • Increased in-memory databases and performance
Java Streams Acceleration	• Significant performance gain for Java application (Java Stream operations)
Oracle Numbers Acceleration	• Increased analytics and database performance with faster arithmetic operations on the Oracle Number, a primitive data type in Oracle Database

Security in Silicon: Silicon Secured Memory

Revolutionary Hardwired Protection Against Data Corruption

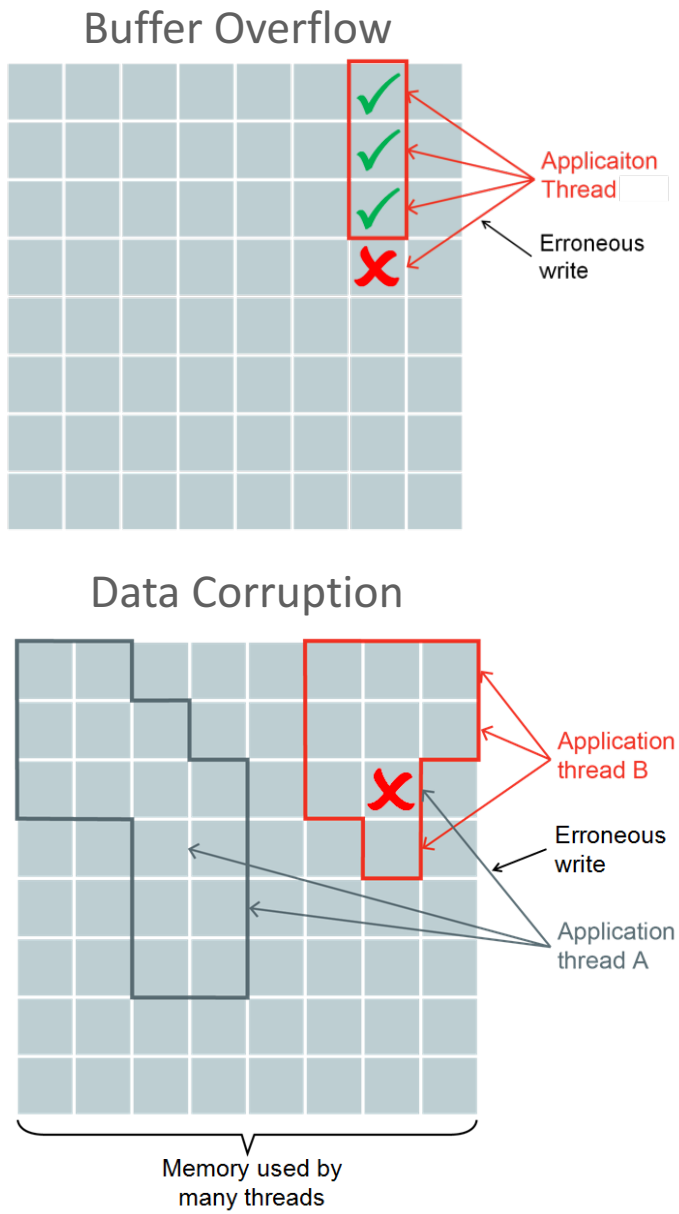
- SPARC M7 and M8 processors stop memory corruption
  - Helps stop malicious code from accessing secure data
  - When access is denied, the problem is reported
- Hidden "color" bits added to pointers and to memory
- Pointer color must match content color or access is aborted
- Can be used in production code with near-zero impact on performance



Protected Memory

Unauthorized and Erroneous Access Prevented

- Oracle Database In-Memory places terabytes of data in memory
  - More vulnerable to corruption by bugs/attacks than storage
- Silicon Secured Memory prevents malicious attacks, invalid/stale references, and buffer overflows
  - Buffer overflow
  - Freed or stale pointers (silent data corruption)
- Enables applications to inspect faulty references, diagnose, and take appropriate recovery actions
- Can be used in optimized production code and by using Silicon Secured Memory–enabled libraries
- Oracle Solaris Studio 12.4 supports Silicon Secured Memory
  - Higher-quality code and shorter development time



A Couple Famous Examples: Heartbleed and Venom

Preventable in Production and Detectable in Development with Silicon Secured Memory

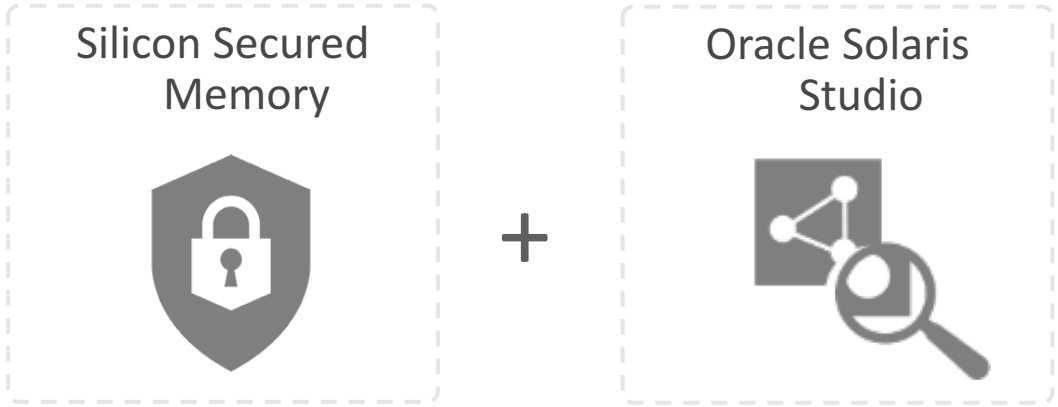
Buffer Over-Read Attack

Buffer Over-Write Attack

Secure Software Made Simple: A Case Study

With Silicon Secured Memory and Oracle Solaris Studio

- Large enterprise app with heavy use of memory-intensive processing
- Time to value
  - **4 cross-platform** bugs tagged in 2 days
  - **180x faster** bug identification
    - Other memory validation tool: 3 hours
    - Silicon Secured Memory and Oracle Solaris Studio's Discover tool: 1 minute



Integrated. Simple. Fast.

Silicon Secured Memory

Support for Both Development and Deployment

DEVELOPMENT: Oracle Solaris Studio provides detailed diagnostics for developers to find and fix memory corruptions

PRODUCTION: Oracle Solaris enables applications to take appropriate recovery actions in real time *

Oracle Solaris Studio 12.4 Discover tool

`libdiscoveradi`

`libadimalloc`

Oracle Solaris kernel
(Provides system calls for user-level applications)

SPARC M7 or M8 hardware
(Enables software stack for Silicon Secured Memory checking)

* App must be coded to use Application Data Integrity (ADI) APIs

How To Use Silicon Secured Memory?

- Run applications that are enabled with Silicon Secured Memory:
  - Oracle Database 12*c* (12.1.0.2 and later) uses Silicon Secured Memory in the system global area (SGA)
  - Some ISV software has been developed with Silicon Secured Memory
- Enable your existing software; No need to recompile!
  - Check application binaries with Oracle Solaris Studio 12.4
  - Link with correct Oracle Solaris libraries, for example, `malloc()` enhanced with ADI: `libadimalloc`
  - Certify on your test environment
- Develop your applications with Silicon Secured Memory
  - C/C++ 64-bit code, Oracle Solaris API for ADI (Silicon Secured Memory)
  - Comprehensive tools available with Oracle Solaris Studio 12.4

Security in Silicon: Cryptographic Acceleration

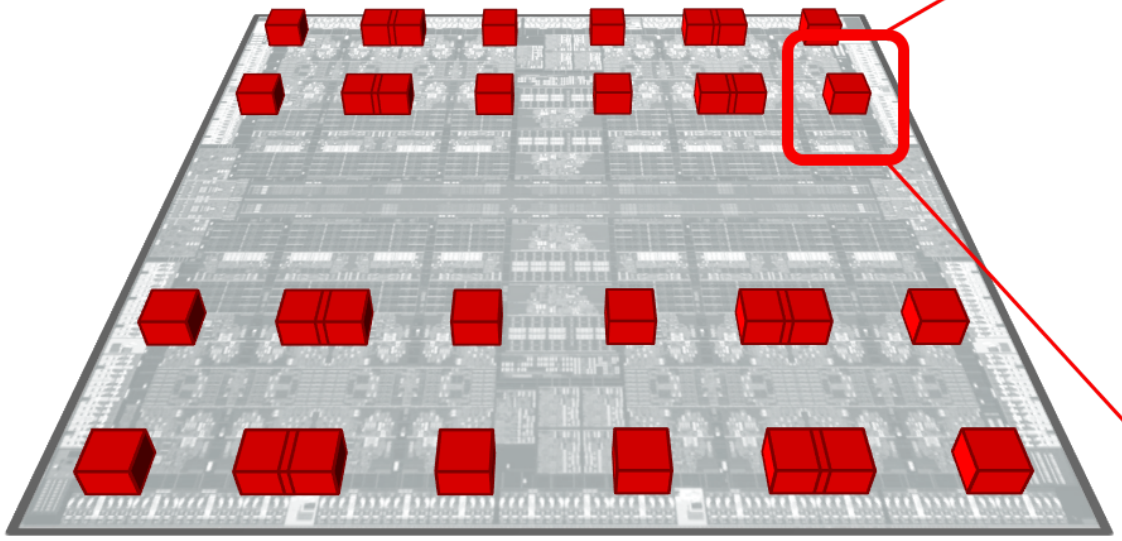
- 32 cryptographic accelerators per SPARC M8 processor, one in each core
- Wire-speed encryption capabilities for secure data center operation without a performance penalty
- Support for 16 industry-standard cryptographic algorithms plus random number generation, now including SHA-3
- Easy to deploy via Oracle Solaris Cryptographic Framework
- Oracle Database 11.2.0.3 or later: Acceleration is automatic with Oracle Database Enterprise Edition when Transparent Data Encryption is enabled

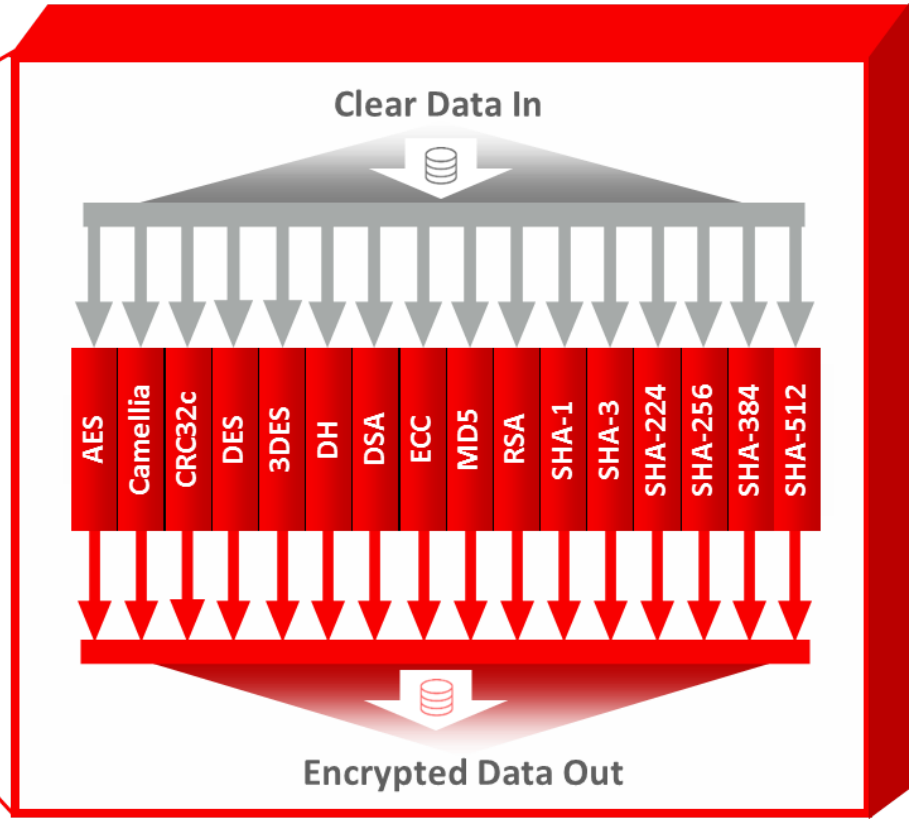
SPARC M8 Processor: Cryptographic Acceleration

Broadest Set of Ciphers for End-to-End Encryption

NEW: SHA-3

32 Crypto Accelerators per Processor

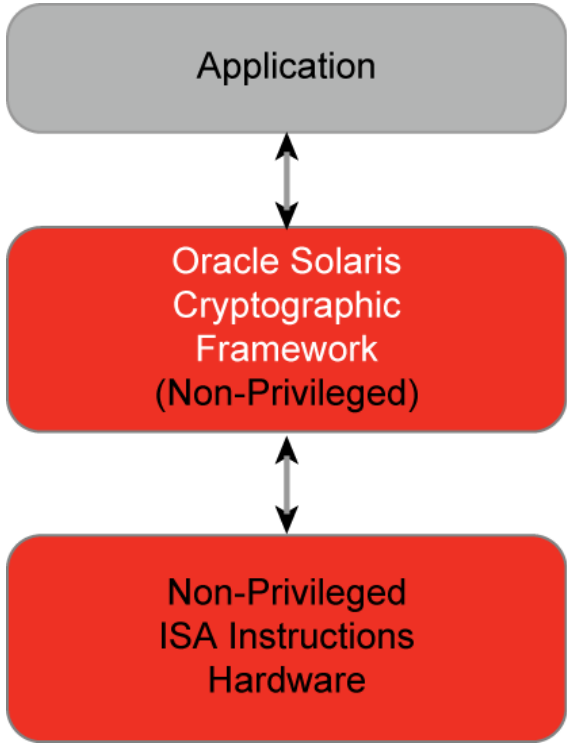




Supported Algorithms and Operational Model

Algorithm Type	Algorithm
Accelerator Driver	Userland (no drivers required)
Public-Key Encryption	RSA, DSA, DH, ECC
Bulk Encryption	AES, DES, 3DES, R4, Camelia
Message Digest	CRC32c, MD5, SHA-1, SHA-3, SHA-224, SHA-256, SHA-384, SHA-512
APIs	PKCS#11 Standard, Ucrypto APIs, Java Cryptography Extensions, OpenSSL

Operational Model



How To Use Encryption Acceleration?

Leverage the Oracle Solaris Cryptographic Framework

- Encryption acceleration automatically enabled with the Transparent Data Encryption (TDE) feature of Oracle Database Advanced Security Option
  - Starting with Oracle DB 11.2.0.3, requires ASO
- Encryption will be automatically accelerated as long as these are using the Oracle Solaris cryptographic framework
  - See documentation: Managing Encryption and Certificates in Oracle Solaris 11.3
  - Example: Encryption in WebSphere MQ will be automatically accelerated through KSSL
  - Oracle WebLogic Server automatically leverages the hardware cryptoaccelerators when enabling secure client transactions via SSL/TLS

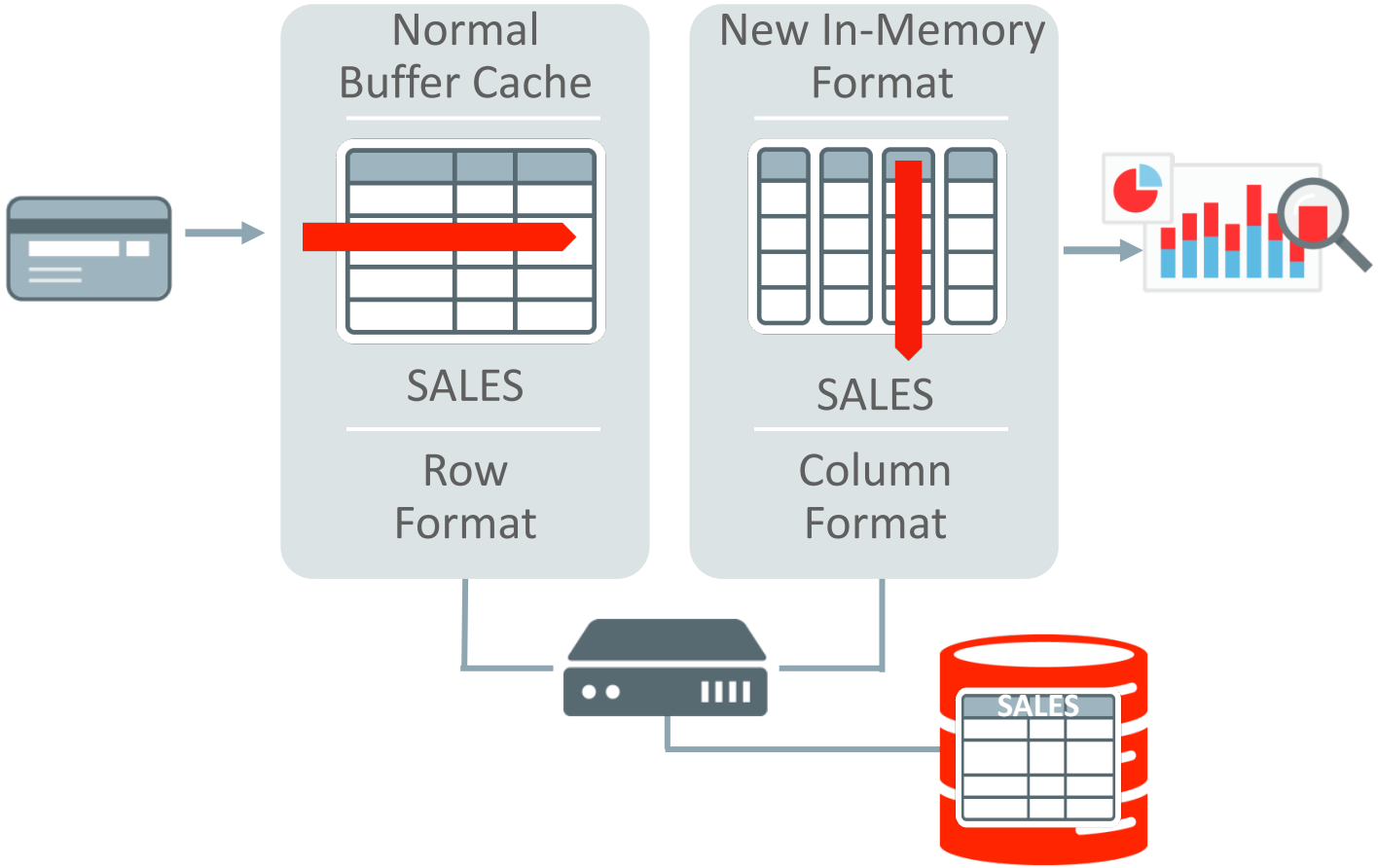
Encryption Use Cases

Protect your data. Encrypt it!

Cryptography Types	Common Use Examples	Mainly Used	Defining Characteristic
Asymmetric key / public key encryption	Web browsers, phone calls, VPN, secure FTP	To protect data or files in transit	2 keys, one public & one private/ sender and recipient
Symmetric key / bulk encryption	Databases, credit card and social security numbers, plus other private info	To protect data or files at rest or stored: disks, backup tapes, and so on	One key to both encrypt and decrypt
Message digests / hash functions	Data lookup and authentication, digital signatures, message authentication codes (MAC)	To compress/create a short summary from a data chunk and not expose it; To detect duplicate or corrupt data	One-way operation: Data in creates a unique value, Value in yields the original data
Random number generation	Ubiquitous	To generate keys	Required in many cryptography aspects

Oracle Database 12*c* Breakthrough: Dual-Format Database

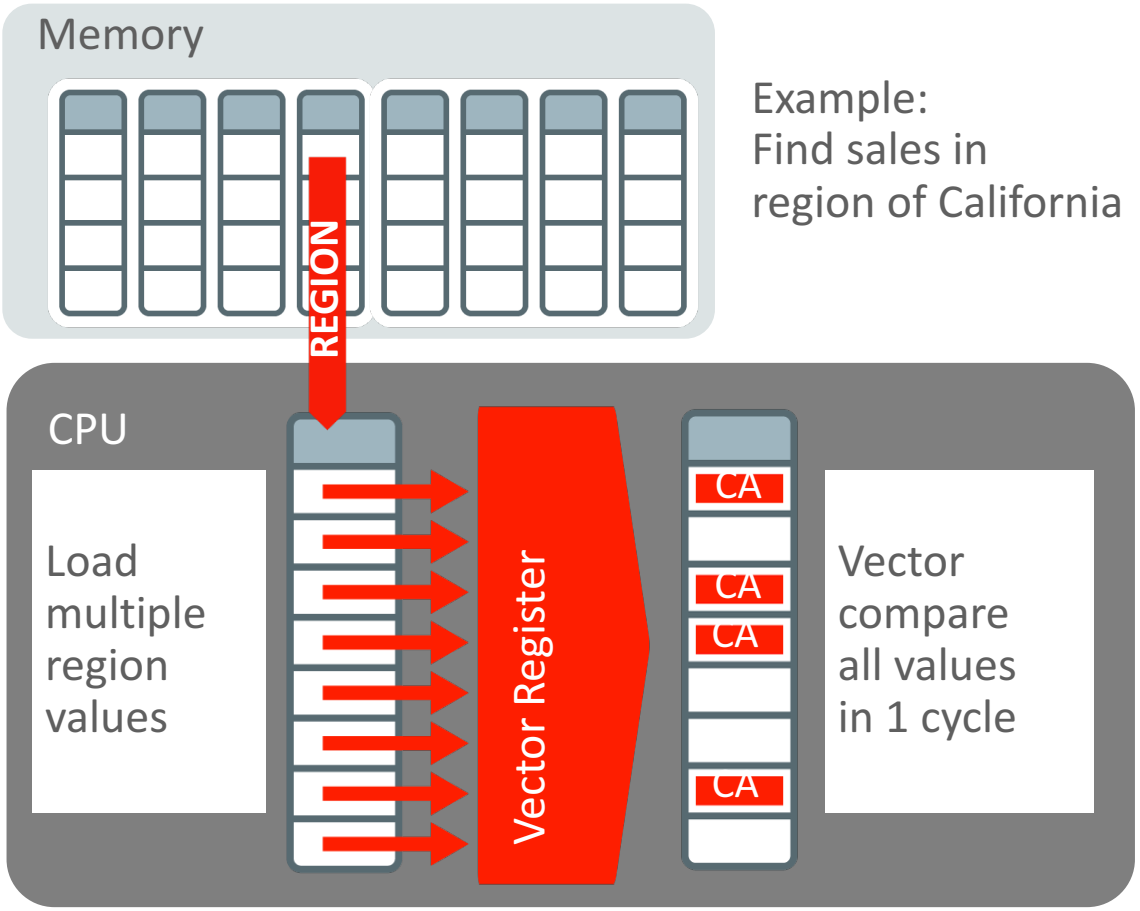
Dual-Format Database—Enables In-Memory Query Acceleration



- **BOTH** row and column formats for the same table
- Simultaneously active and transactionally consistent
- Analytics and reporting use new in-memory column format
- OLTP uses proven row format

Orders-of-Magnitude Faster Analytic Data Scans

Oracle Database 12*c* In-Memory Option

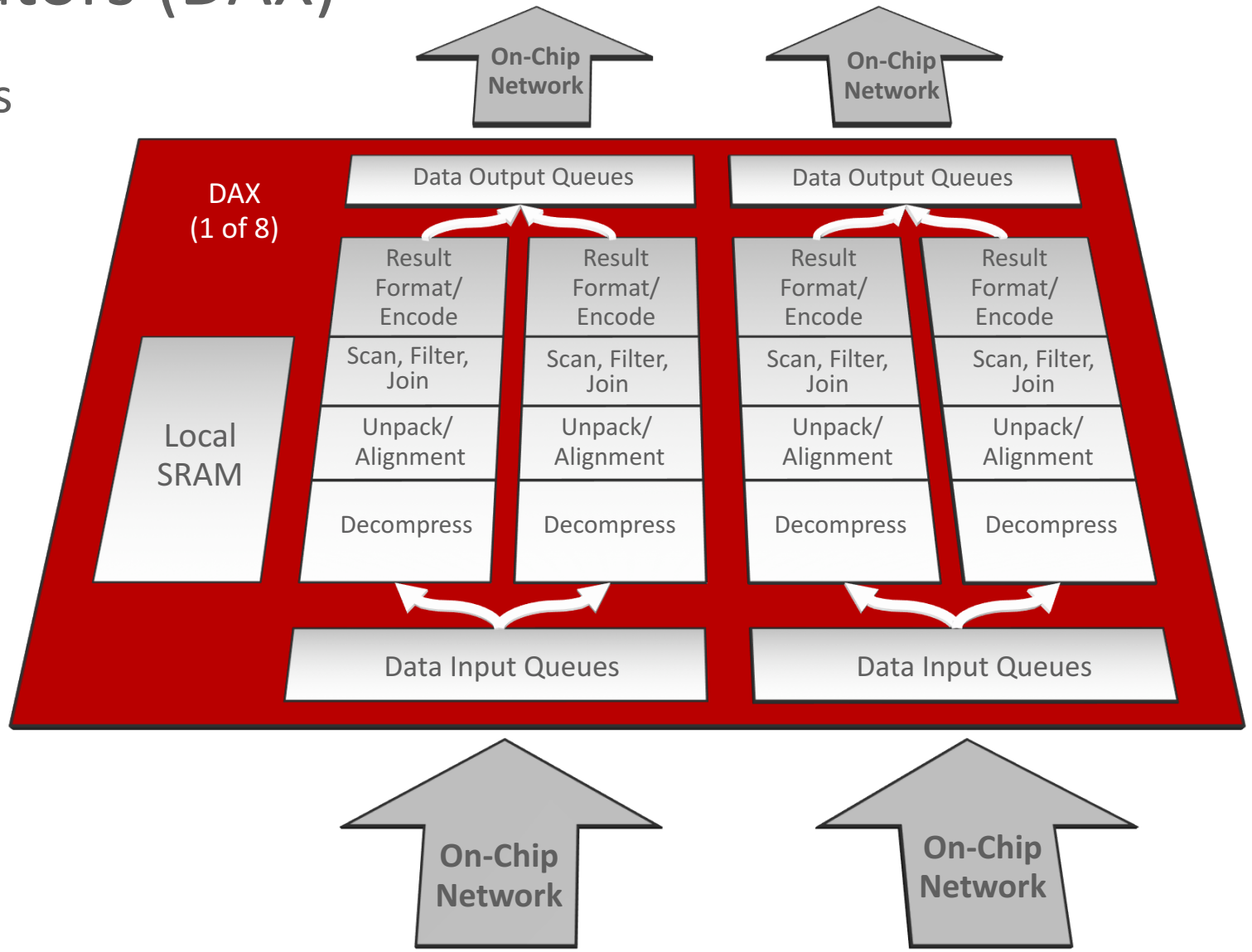


> 100x Faster

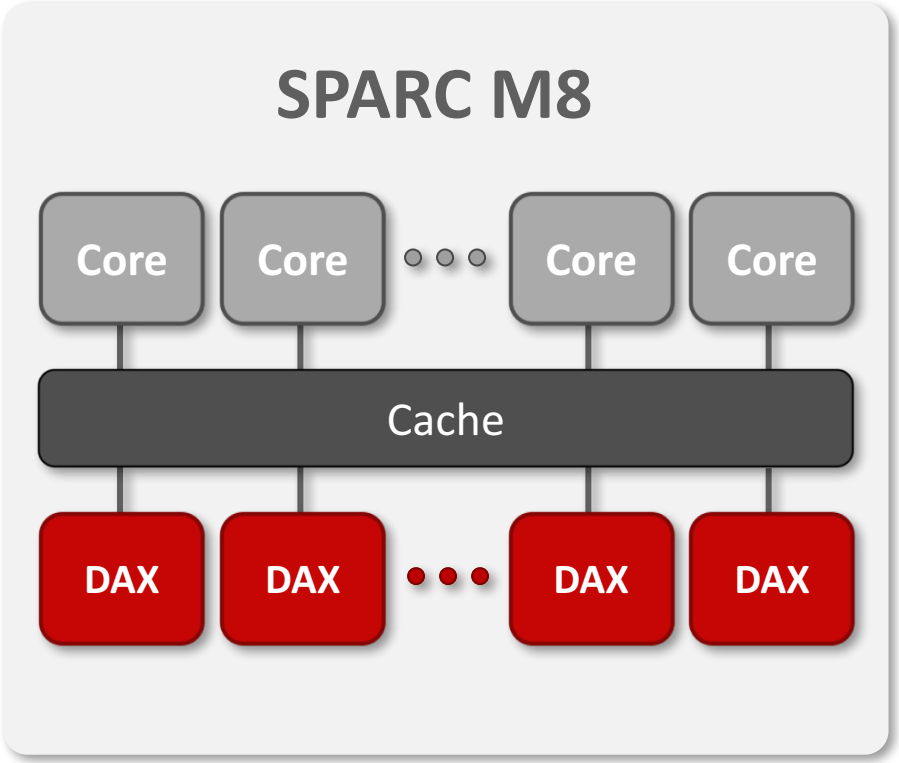
- Each CPU core scans local in-memory columns
- Scans use super-fast SIMD vector instructions
  - Originally designed for graphics and science
- Billions of rows/sec scan rate per CPU core
  - Row format is millions/sec

Data Analytics Accelerators (DAX)

- 32 in-silicon accelerator engines
  - 8 DAXs, 4 pipelines per DAX
  - Using less than 1% of chip space
- Independently process streams of columns
  - Up to 170 billion rows per second!
- Cores/threads operate synchronously or asynchronously to accelerator engines
- User-level synchronization through shared memory

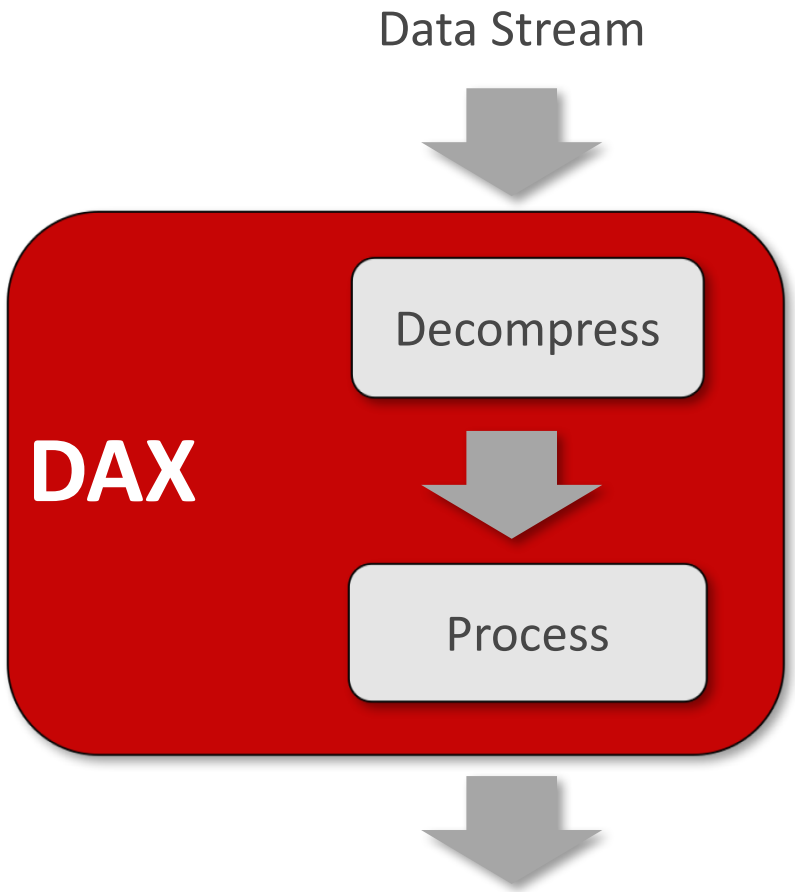


SQL in Silicon: In-Memory Query Acceleration



- SPARC M8 processor includes on-chip Data Analytics Accelerators (DAX)
  - Independently process streams of database column elements placed in system memory
  - Example: Find all values that match “California”
  - 8 accelerators per chip, with 4 pipelines each
- Frees processor cores to run higher-level SQL functions, or other applications
- Reads data directly from memory, processes it, and places results in cache for core usage

SQL in Silicon: In-Line Decompression



- Key to placing more data in memory
- The performance of decompression on today's processors is a huge bottleneck
- Solution: acceleration unit runs decompression at memory speeds
  - Equivalent to 16 decompression PCI cards, 60 CPU cores
- Increases useable memory

SQL in Silicon with Oracle Database

In-Memory Query Acceleration and In-Line Decompression

- In-Memory Query Acceleration and In-Line Decompression available only with the Oracle Database In-Memory option
- Oracle Database 12.1.0.2 + latest patches required
- Oracle Database 12.2.0.1 + BP170718 recommended
- Oracle Applications certified for Oracle Database 12*c* support DAX
  - Performance will be application-specific
  - Internal workloads have been heavily optimized with indexes over time
  - More opportunities with customer database

Accelerate Your Analytics with Open APIs for DAX

Software in Silicon Developer Program

- Leverage hardware acceleration for big data analytics, machine learning, and more
- Developer resources
  - Oracle Solaris 11 open APIs for DAX
  - Example integration for Apache Spark
  - Documented use cases and code examples
- Expands existing Silicon Secured Memory developer resources

Accelerating Analytic Query Operations with DAX

Example Algorithms Accelerated

- Key-value pairs: both simple and complex
- Finding top <N> items from an ordered list
- Building analytic cubes
- JSON processing
- Outlier detection

Example Use Cases

- Fraud and intrusion detection
- Use patterns
- Buying recommendations
- Trend detection
- Market segmentation
- Classification and regression

Agenda

1. Overview
2. System Details
3. New Technologies
4. RAS
5. Platform Management
6. Virtualization
7. Summary

SPARC T8-1 Server

System Details

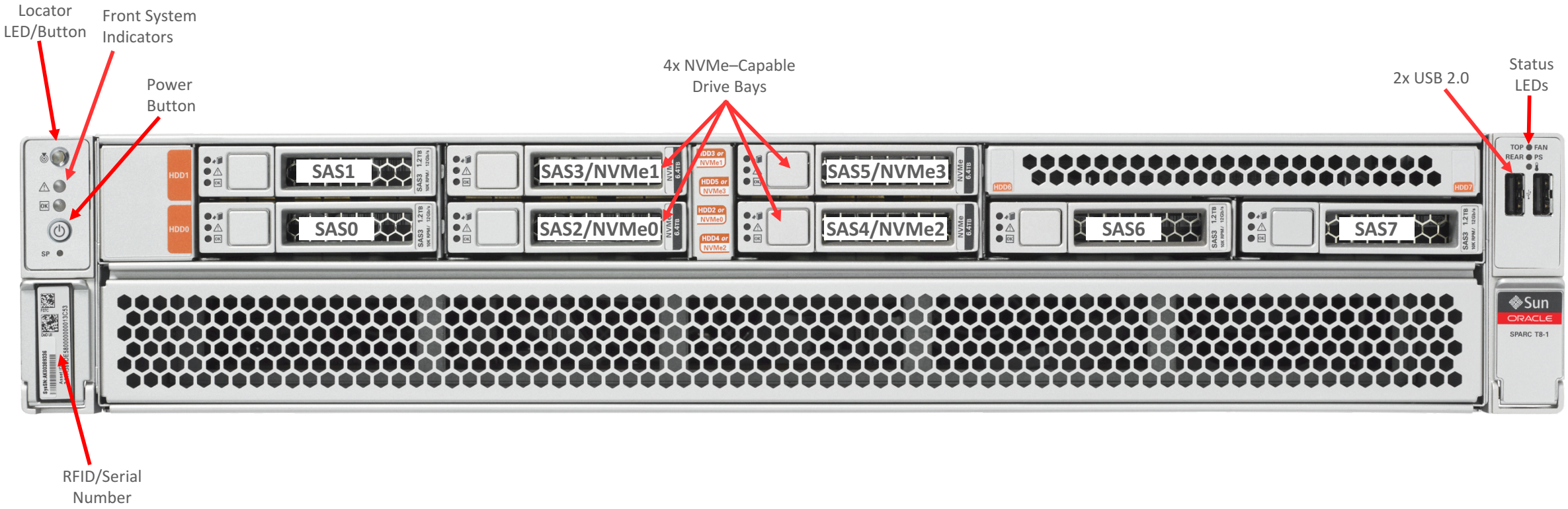
SPARC T8-1 Server: Overview

- 2U chassis, fits into 900 mm deep rack
- One SPARC M8 processor (32 cores/256 threads)
- Up to 1 TB of memory
  - 8 or 16 DIMMs, 16 GB, 32 GB or 64 GB each
- Eight 2.5” SFF hot-pluggable disk drive bays
  - Up to 8 SAS HDDs/SSDs
  - Up to 4 NVMe SSDs
  - Mixing SAS and NVMe drives is supported
- One onboard SAS3 HBA (hardware RAID 0/1/10/1E)
- One optional NVMe PCIe switch card (factory-installed)
- Four 10GBASE-T ports
- 6 low-profile PCIe 3.0 x8 slots, or 2 x16 + 2 x8 slots (4 PCIe buses)
- Local (front and rear USB, rear video) and remote KVMS
- Four hot-swappable redundant fan modules (top loading)
- Two 1,200 W (output) hot-pluggable power supplies (1+1)

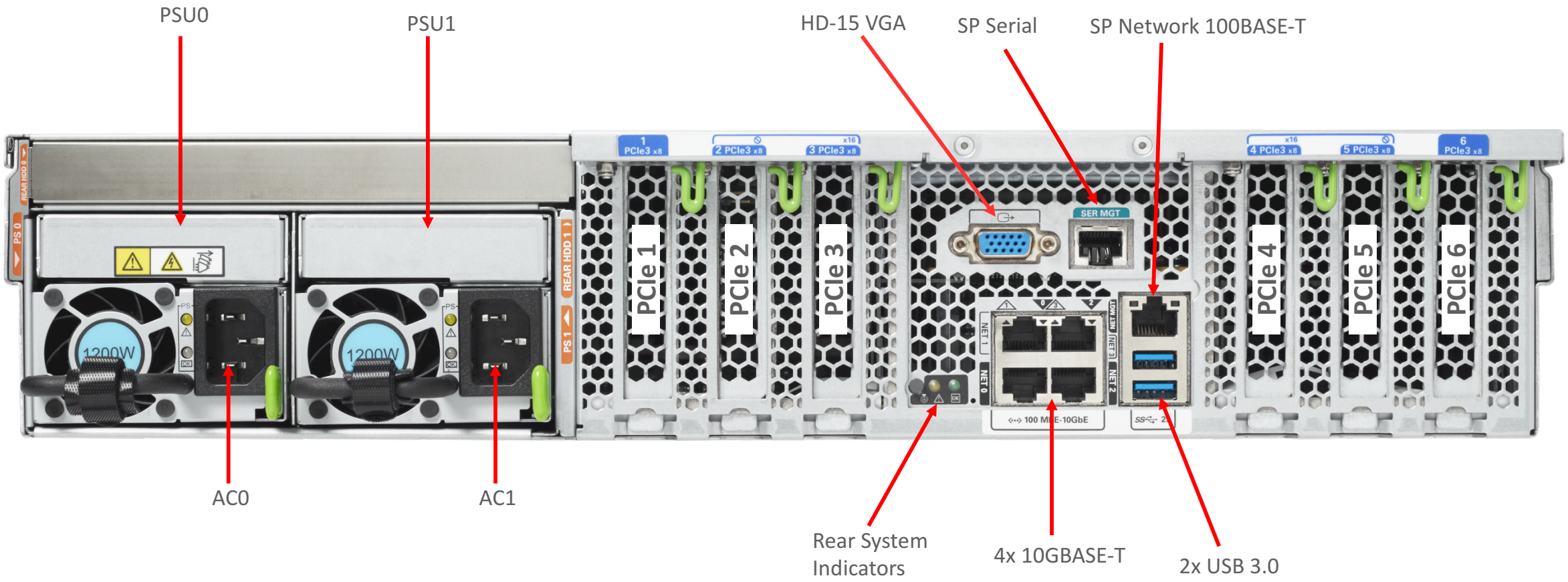
SPARC T8-1 Versus SPARC T7-1

Feature	SPARC T8-1	SPARC T7-1
Form Factor	2U, 737 mm (29”) deep	2U, 737 mm (29”) deep
Processor	1x 5.0 GHz SPARC M8, 32 cores, (32 cores, 256 threads)	1x 4.13 GHz SPARC M7, 32 cores, (32 cores, 256 threads)
Memory	DDR4-2400, 16x slots, Max. 1 TB w/ 64 GB DIMMs	DDR4-2133, 16x slots, Max. 1 TB w/ 64 GB DIMMs
Integrated Network Ports	4x 10GBASE-T	4x 10GBASE-T
Internal Storage	8x 2.5” hot-pluggable SFF bays 8x SAS 3.0 HDD and/or SSD, 4x NVMe SSDs	8x 2.5” hot-pluggable SFF bays 8x SAS 3.0 HDD and/or SSD, 4x NVMe SSDs
Removable Media	-	DVD drive
Management Ports	1x Serial (RJ-45), 1x 100BASE-T, 1x VGA (HD-15)	1x Serial (RJ-45), 1x 100BASE-T, 1x VGA (HD-15)
USB Ports	2xUSB 2.0, 2x USB 3.0	2xUSB 2.0, 2x USB 3.0
PCI Express Slots	6x PCIe 3.0 (x8) low-profile slots, or 4x PCIe 3.0 slots, 2 (x16) and 2 (x8)	6x PCIe 3.0 (x8) low-profile slots, or 4x PCIe 3.0 slots, 2 (x16) and 2 (x8)
Hot-swappable Fans	4x redundant dual fan modules	4x redundant dual fan modules
Hot-swappable Power Supplies (Nominal Output)	2x 1200 watt AC, 1+1	2x 1000 watt AC, 1+1

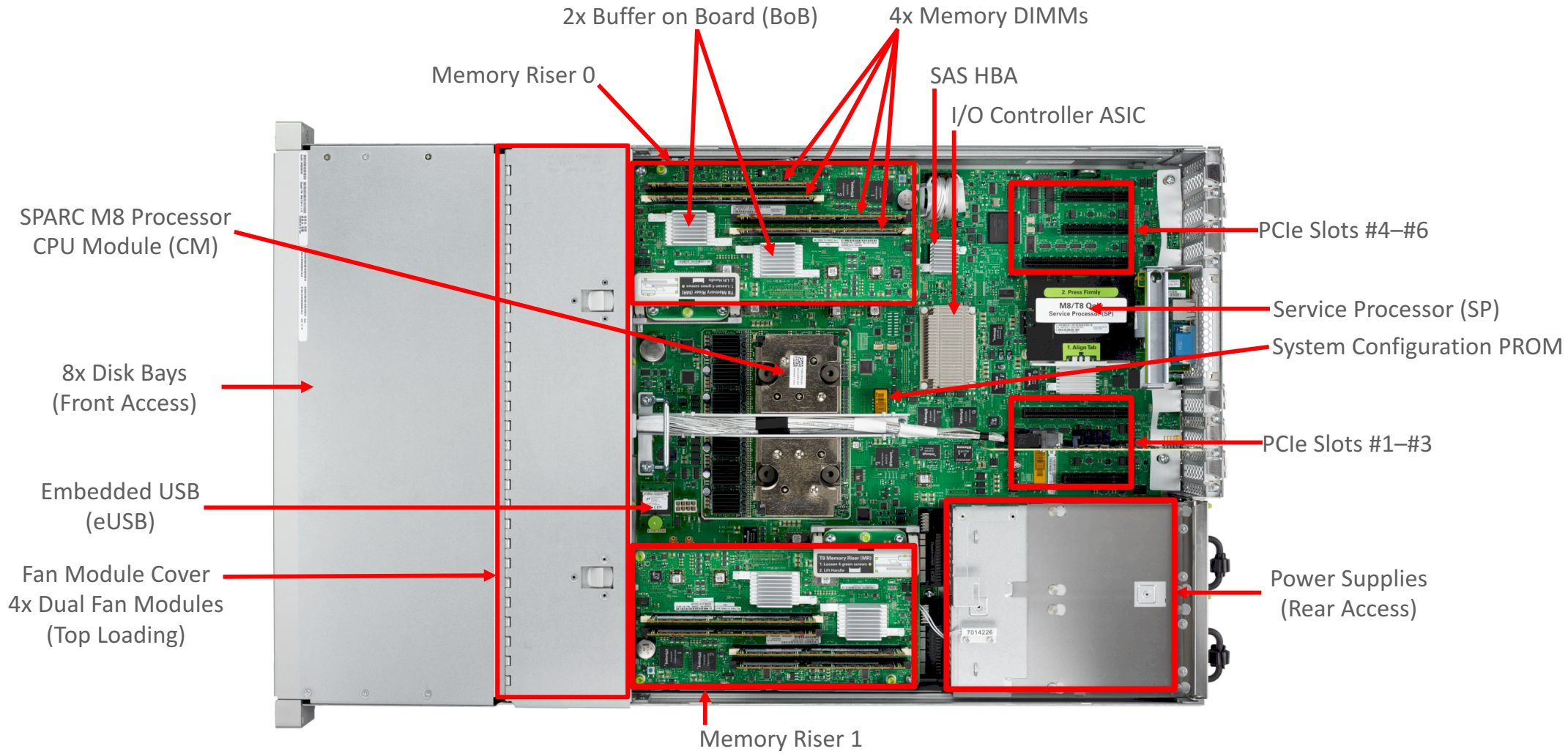
SPARC T8-1 Server: Front View



SPARC T8-1 Server: Rear View



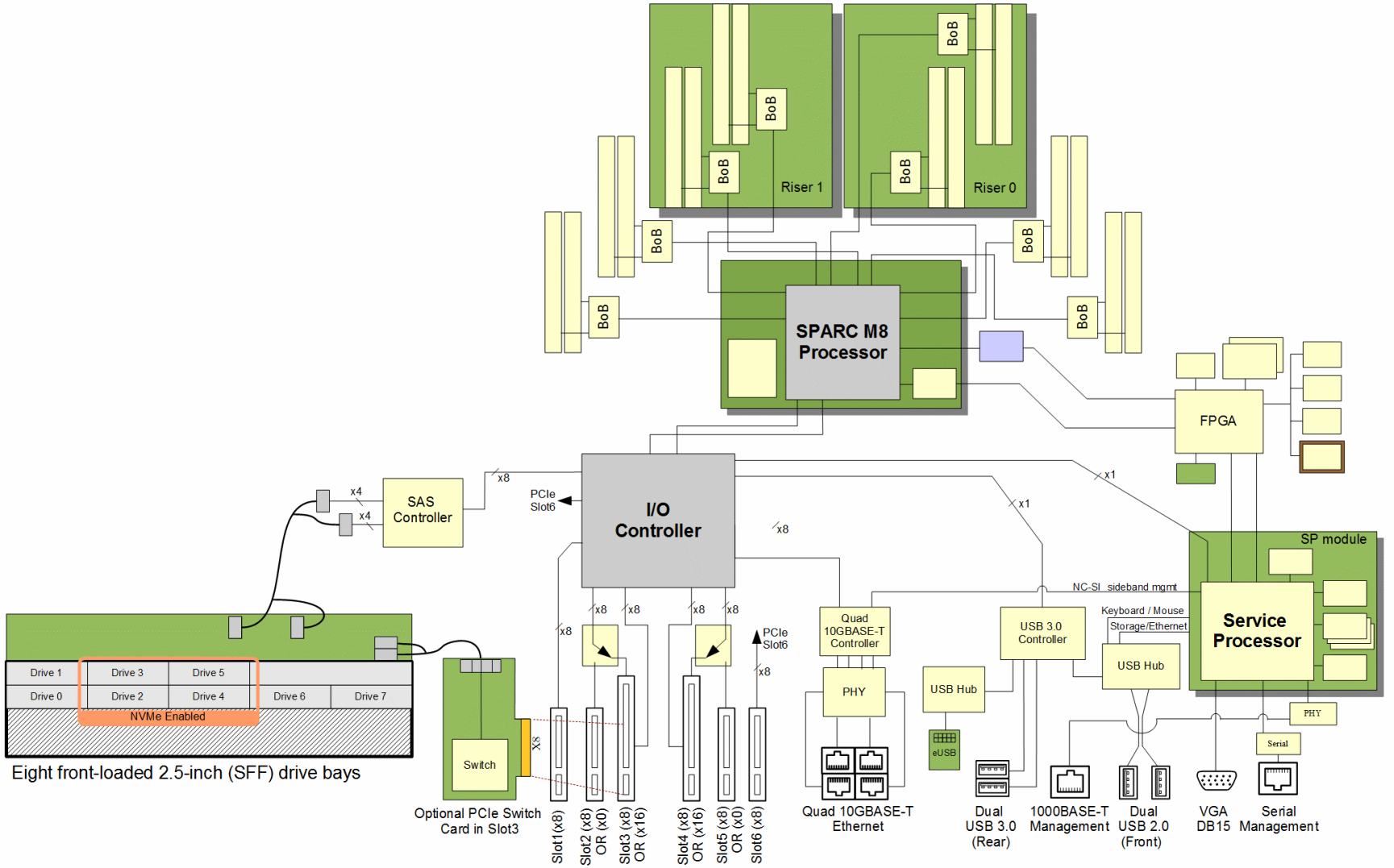
SPARC T8-1 Server: Top View



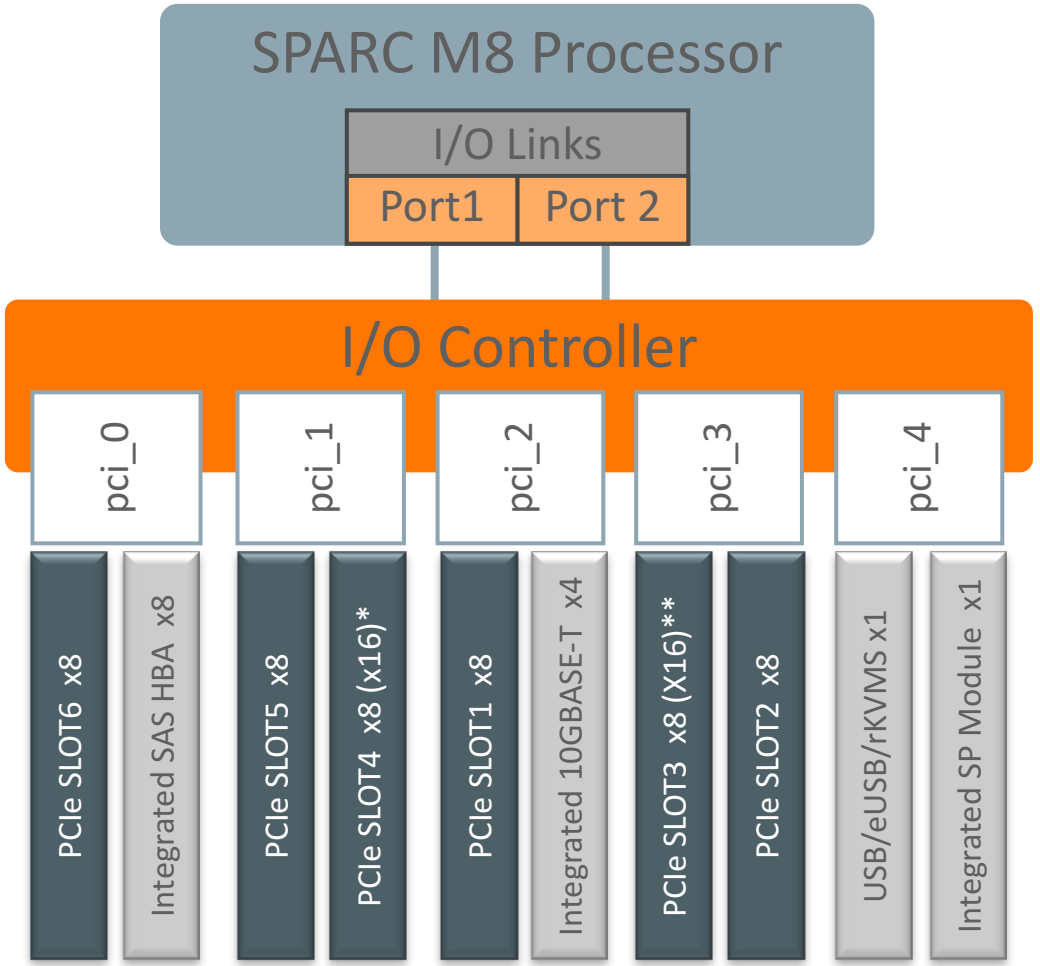
SPARC T8-1 Server: Configuration Policy

- Memory
  - Half-populated system memory is supported (8x DIMMs on motherboard [MB])
  - Fully populated system memory is supported (8x DIMMs on MB and 8x DIMMS on memory risers)
  - Both X- and ATO options require all DIMM sizes to be the same: either all 16 GB, 32 GB or 64 GB
- Fans
  - System will continue to operate with fan failure(s) unless an over-temperature scenario occurs
  - System will power-on with one **failed** fan in any of the fan modules
  - System will *not* power on with a **missing** fan module
- Power Supply Units (PSUs)
  - Base package includes 1+1 redundant hot-swappable power supplies
  - System continues to operate and will power-on with one failed or missing power supply
  - System warns if an AC cord is not attached
  - Requires 200–240 VAC, 50/60 Hz

SPARC T8-1 Server: Block Diagram



SPARC T8-1 Server: PCIe Root Complex Map



* SLOT4 is x16 when SLOT5 is empty
** SLOT3 is x16 when SLOT2 is empty
*** Optional NVMe PCIe switch card must be in SLOT3

Legend:

SPARC T8-1 Server: PCIe I/O Mapping

Name	RC	Port	Width	Path	Notes
NET0 NET1 NET2 NET3	2	0	x8	/pci@300/pci@1/network@0 /pci@300/pci@1/network@0,1 /pci@300/pci@1/network@0,2 /pci@300/pci@1/network@0,3	Integrated quad 10GBASE-T
SAS0	0	1	x8	/pci@301/pci@2/scsi@0	Integrated SAS HBA, drive bays HDD #0–#7
Slot1	2	1	x8	/pci@300/pci@2	Dedicated root complex (RC)
Slot2	3	1	x8 or NC**	/pci@303/pci@2	Slot2 is x8–capable, see Slot3 notes
Slot3	3	0	x8 or x16	/pci@303/pci@1	Slot3 is x16–capable when Slot2 is empty. This slot supports the optional NVMe PCIe switch.
Slot4	1	0	x8 or x16	/pci@302/pci@1	Slot4 is x16–capable when Slot5 is empty
Slot5	1	1	x8 or NC**	/pci@302/pci@2	Slot5 is x8–capable; see Slot4 notes
Slot6	0	0	x8	/pci@301/pci@1	Root complex shared with integrated SAS HBA
USB	4	0	x1	/pci@304/pci@1/usb@0	See Installation Guide for device specific paths
SPM	4	0	x1	/pci@304/pci@1/usb@0	

* Refer to the device map
** NC = not connected

SPARC T8-1 Server: CRUs and FRUs

- Hot-service customer replaceable units (CRUs)
  - Fan modules
  - Power supplies
  - 2.5” SAS HDDs/SSDs
  - 2.5” NVMe SSDs
- Cold-service CRUs
  - Memory riser
  - Memory DIMMs
  - PCIe cards
  - Service processor module (SPM)
  - System configuration PROM
  - System battery
  - Embedded USB flash memory (eUSB)
- Cold-service field replaceable units (FRUs)
  - Motherboard
  - Disk backplane cage
  - LED indicator modules

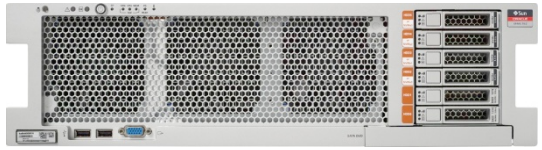
SPARC T8-2 Server

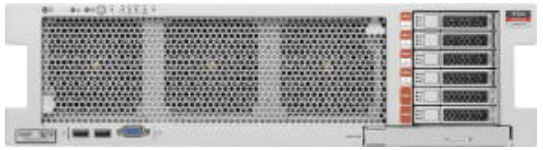
System Details

SPARC T8-2 Server: Overview

- 3U chassis, fits into 900 mm deep rack
- Two SPARC M8 processors (64 cores/512 threads)
- Up to 2 TB GB of memory
  - 16 or 32 DIMMs, 16 GB, 32 GB or 64 GB each
- Six 2.5” SFF hot-pluggable disk drive bays
  - Up to 6 SAS HDDs/SSDs
  - Up to 4 NVMe SSDs
  - Mixing SAS and NVMe drives is supported
- Dual onboard SAS3 HBAs (HW RAID 0/1/10/1E)
- One or two optional NVMe PCIe switch cards (factory-installed)
- Four 10GBASE-T ports
- 8 low-profile PCIe 3.0 slots, 4 x16 + 4 x8 (8 PCIe buses)
- Local (front and rear USB, rear video) and remote KVMS
- Six hot-swappable redundant fan modules (top loading)
- Two 2,000 W (output) hot-swappable power supplies (1+1

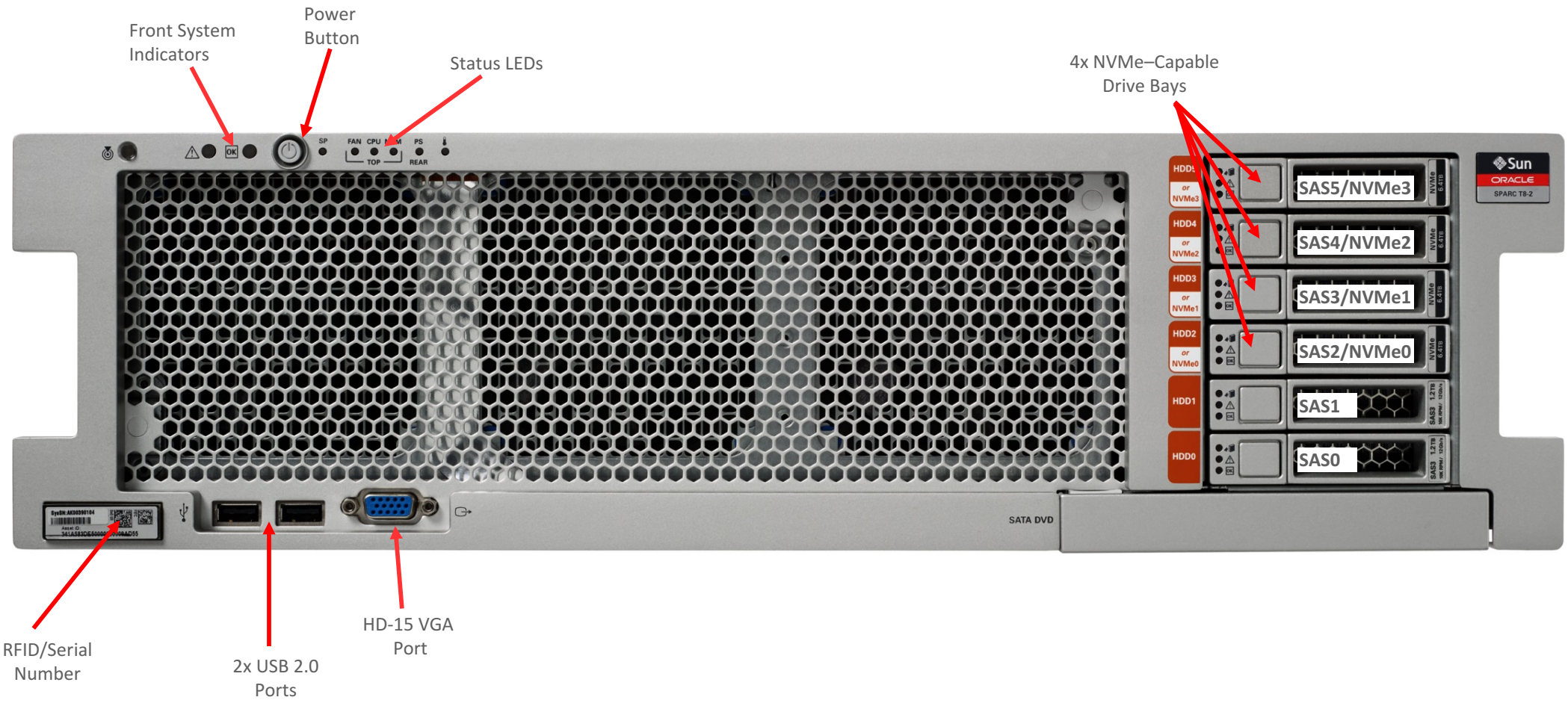
SPARC T8-2 Versus SPARC T7-2



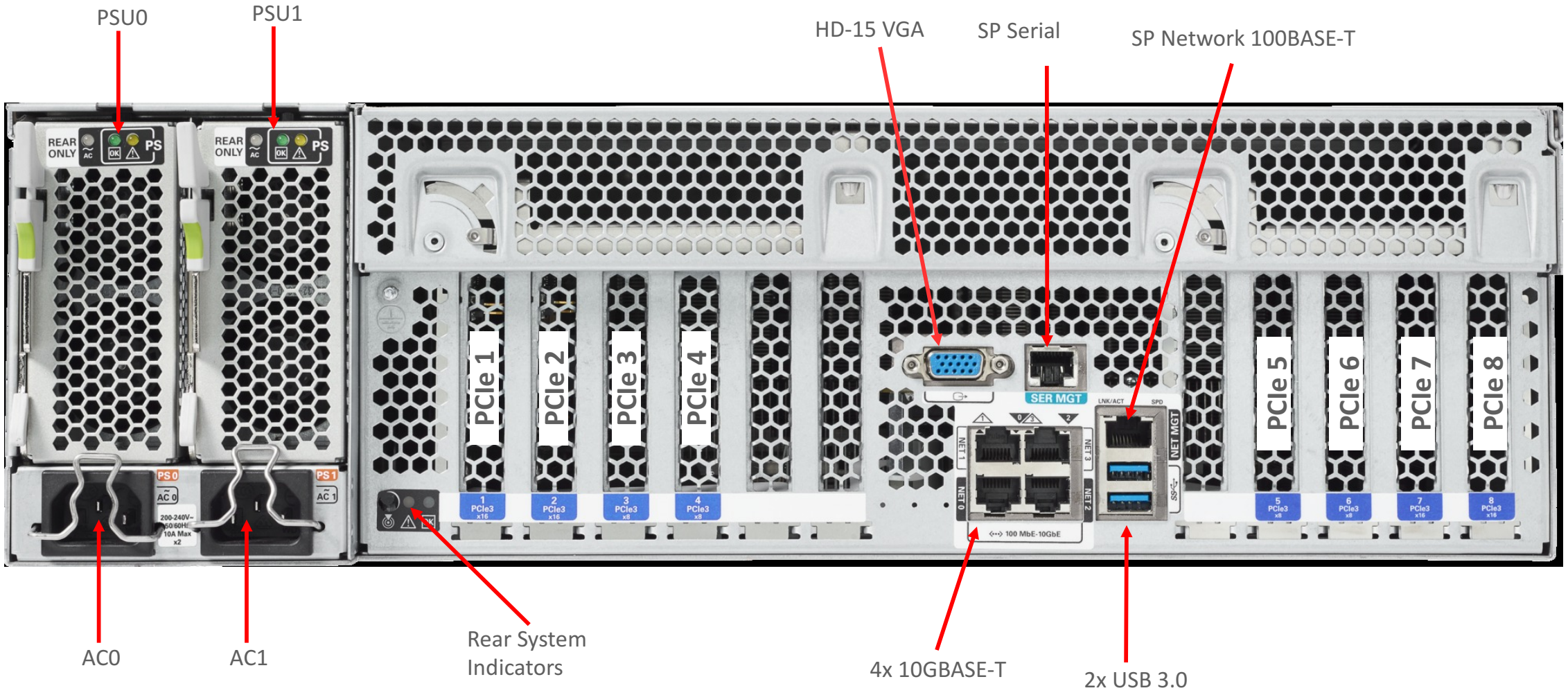


Feature	SPARC T8-2	SPARC T7-2
Form Factor	3U, 753 mm (29.6”) deep	3U, 753 mm (29.6”) deep
Processor	2x 5.0 GHz SPARC M8, 32-cores, (64 cores, 512 threads)	2x 4.13 GHz SPARC M7, 32-cores, (64 cores, 512 threads)
Memory	DDR4-2400, 32x slots, Max. 2 TB w/ 64 GB DIMMs	DDR4-2133, 32x slots, Max. 2 TB w/ 64 GB DIMMs
Integrated Network Ports	4x 10GBASE-T	4x 10GBASE-T
Internal Storage	6x 2.5” hot-pluggable SFF bays 6x SAS 3.0 HDD or SSD, 4x NVMe SSDs	6x 2.5” hot-pluggable SFF bays 6x SAS 3.0 HDD or SSD, 4x NVMe SSDs
Removable Media	-	DVD drive
Management Ports	1x serial (RJ-45), 1x 100BASE-T, 2x VGA (HD-15)	1x serial (RJ-45), 1x 100BASE-T, 2x VGA (HD-15)
USB Ports	2xUSB 2.0, 2x USB 3.0	2xUSB 2.0, 2x USB 3.0
PCI Express Slots	8x PCIe 3.0 low-profile slots, 4 (x16) and 4 (x8)	8x PCIe 3.0 low-profile slots, 4 (x16) and 4 (x8)
Hot-swappable Fans	6x redundant fans	6x redundant fans
Hot-swappable Power Supplies (Nominal output)	2x 2000 watt AC, 1+1	2x 2000 watt AC, 1+1

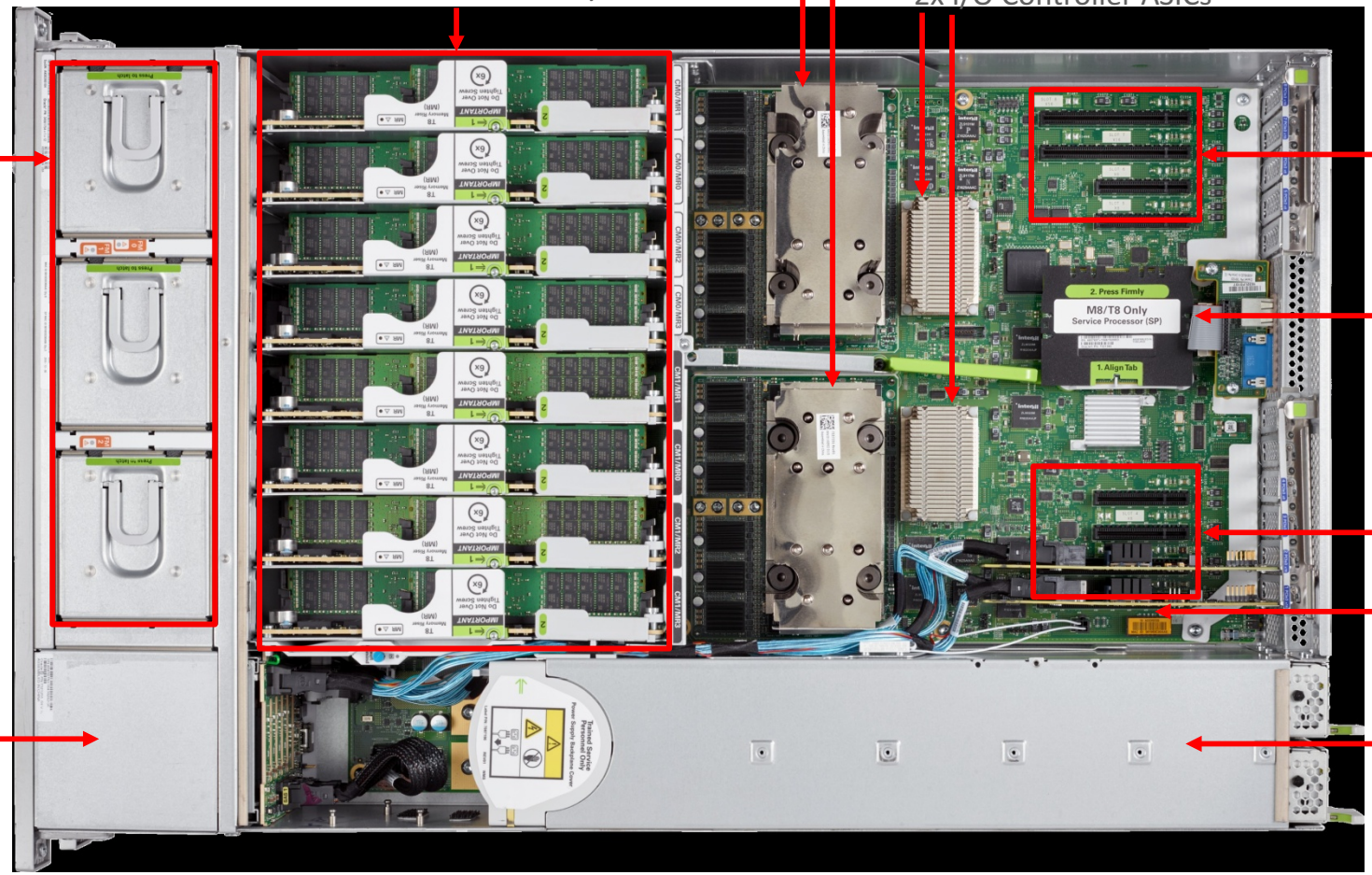
SPARC T8-2 Server: Front View



SPARC T8-2 Server: Rear View



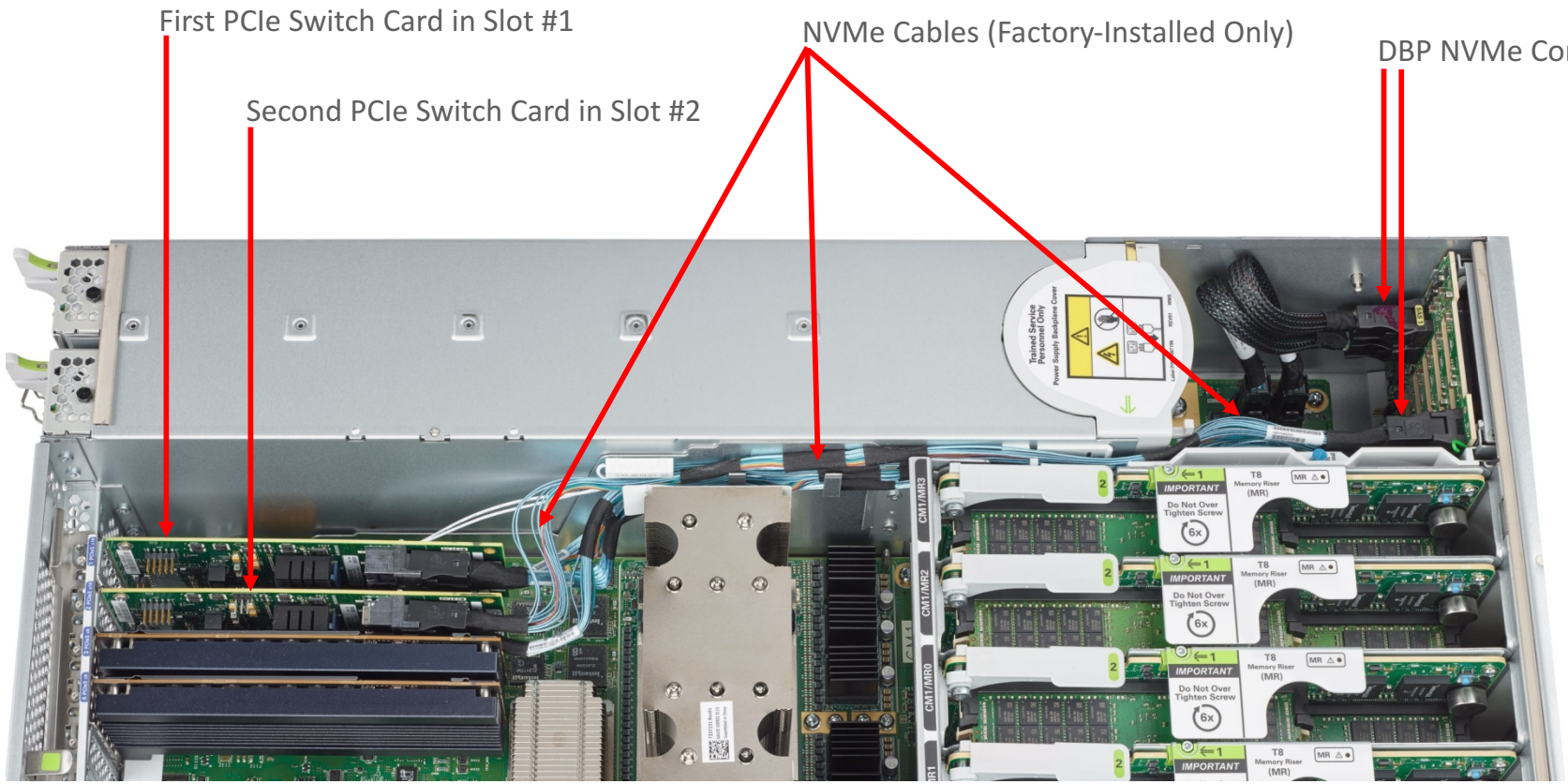
SPARC T8-2 Server: Top View



SPARC T8-2 Server: Configuration Policy

- Memory
  - Half-populated system memory is supported (2 DIMMs per memory riser)
  - ATO options : Total of 16 or 32 DIMMs, all DIMMs of the same size
  - Mixing of X-option DIMM sizes is supported; all DIMMs associated with a CPU must be same size
    - Example 1: 4x 32G DIMMs on memory risers 0-0, 0-1, 0-2, 0-3 + 4x 16G DIMMs on memory risers 1-0, 1-1, 1-2, 1-3
    - Example 2: 2x 32G DIMMs on memory risers 0-0, 0-1, 0-2, 0-3 + 4x 16G DIMMs on memory risers 1-0, 1-1, 1-2, 1-3
- Fans
  - System will continue to operate with a single fan failure (any 1 of the 6 failed) unless an over-temperature scenario occurs
  - System will power-on if 5 of the 6 fan modules are present; any single fan can be missing
- Power Supply Units (PSUs)
  - Base package includes 1+1 redundant hot-swappable power supplies
  - System continues to operate and will power-on with one failed PSU
  - Requires 200–240 VAC , 50/60 Hz

SPARC T8-2 Server: PCIe Switch Card and NVMe Cables



One PCIe Switch Card

- PCIe Switch Card in Slot #1
- Supports NVMe Bays 0, 1, 2, 3*

Two PCIe Switch Cards

- PCIe Switch Card in Slot #1
- Supports NVMe Bays 0, 1*

- PCIe Switch Card in Slot #2
- Supports NVMe Bays 2, 3*

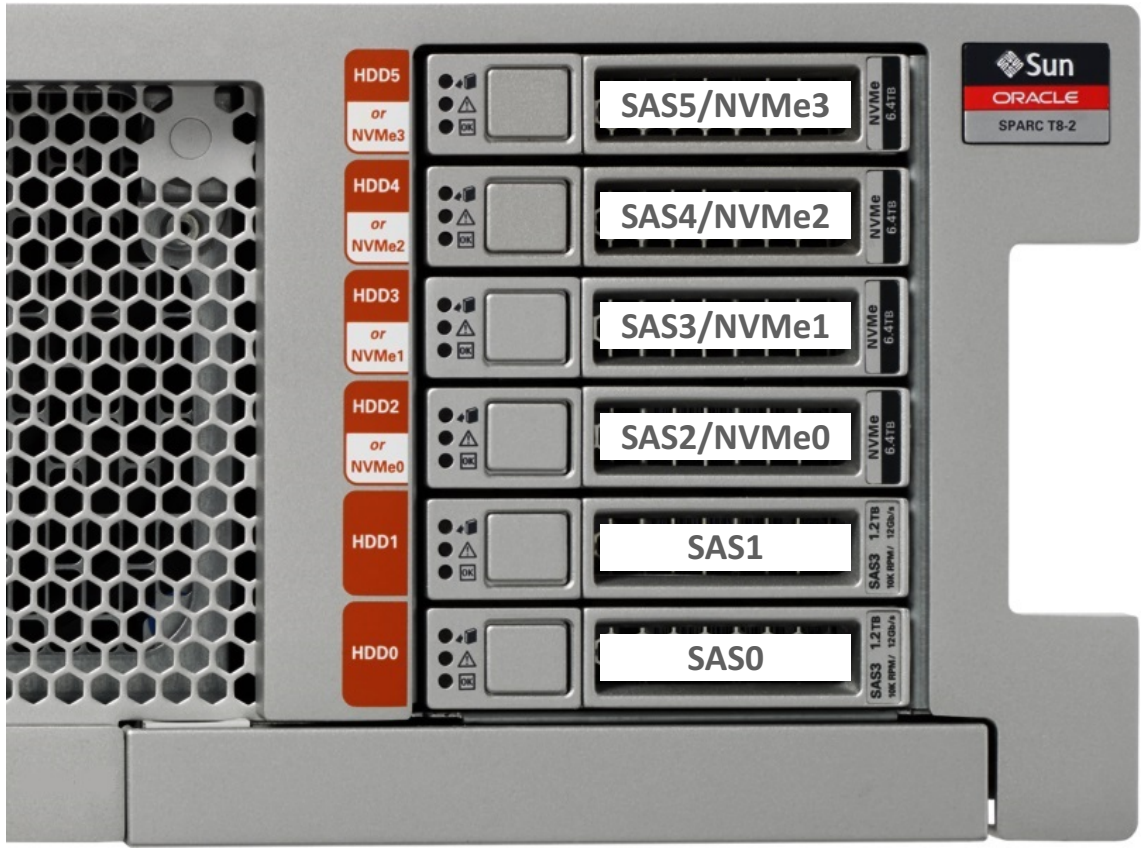
* Note that NVMe and SAS drive bay numbering are different

SPARC T8-2 Server: NVMe 2.5” SFF Drives

Device Nomenclature and Path Names

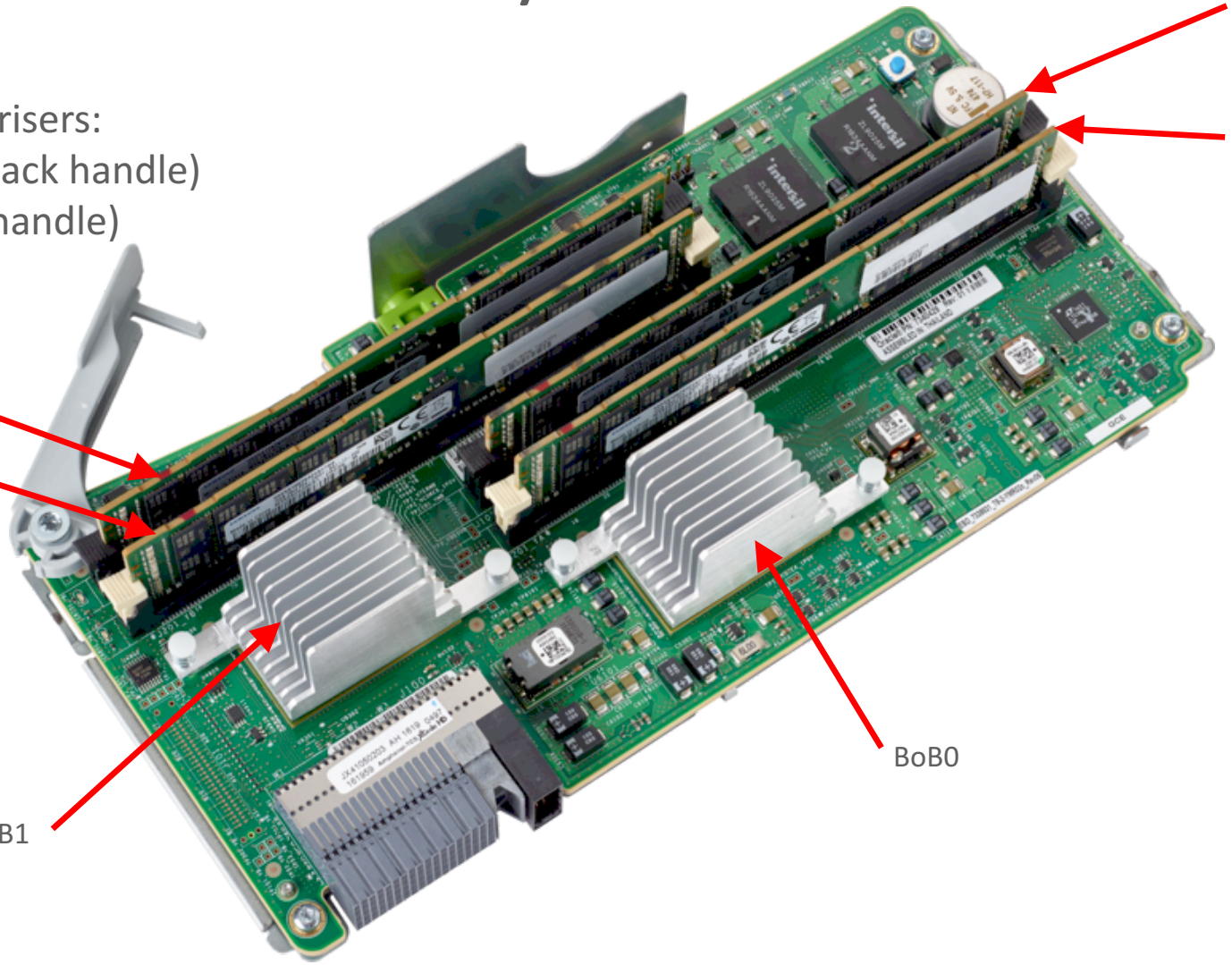
NVMe Bay	Shared SAS Bay	One PCIe Switch Card (PCIe Slot 1)
NVMe0	HDD2	/pci@306/pci@1/pci@0/pci@4/nvme@0/disk@1
NVMe1	HDD3	/pci@306/pci@1/pci@0/pci@5/nvme@0/disk@1
NVMe2	HDD4	/pci@306/pci@1/pci@0/pci@6/nvme@0/disk@1
NVMe3	HDD5	/pci@306/pci@1/pci@0/pci@7/nvme@0/disk@1

NVMe Bay	Shared SAS Bay	Two PCIe Switch Cards (PCIe Slots 1,2)
NVMe0	HDD2	/pci@306/pci@1/pci@0/pci@6/nvme@0/disk@1
NVMe1	HDD3	/pci@306/pci@1/pci@0/pci@7/nvme@0/disk@1
NVMe2	HDD4	/pci@307/pci@1/pci@0/pci@6/nvme@0/disk@1
NVMe3	HDD5	/pci@307/pci@1/pci@0/pci@7/nvme@0/disk@1



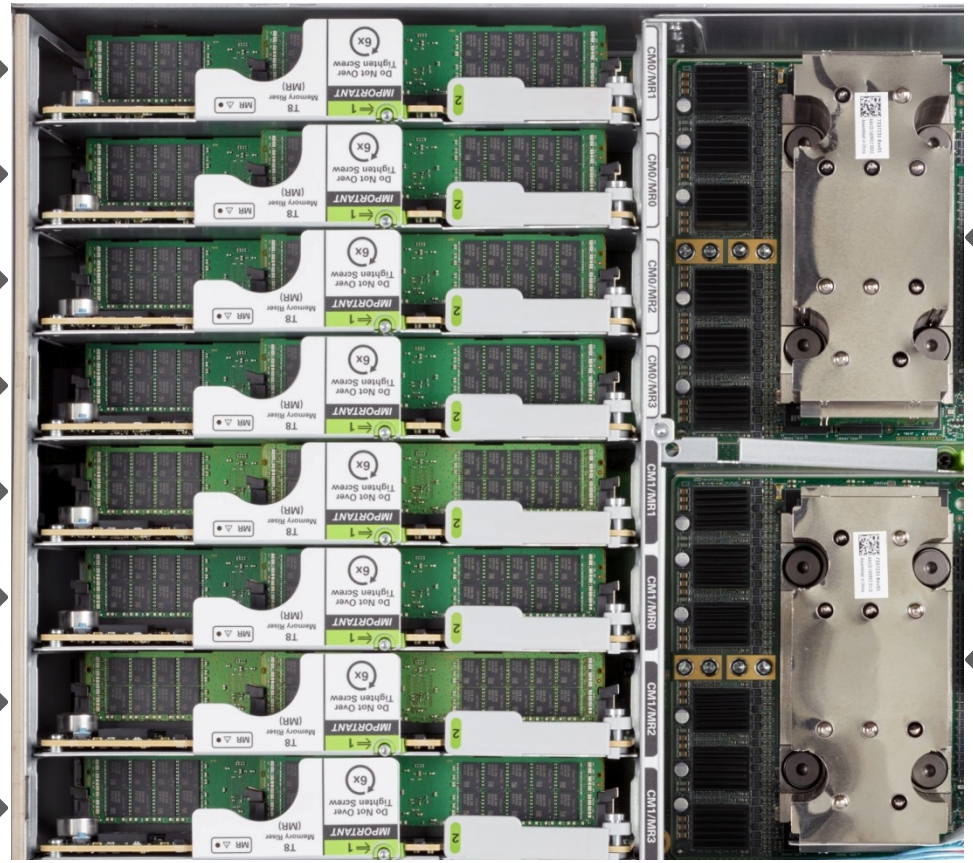
SPARC T8-2 Server: Memory Riser

Half-populated memory risers:
- Channel0 populated (black handle)
- Channel1 empty (light handle)



SPARC T8-2 Server: Memory Risers

CPU0, Riser1 →
CPU0, Riser0 →
CPU0, Riser2 →
CPU0, Riser3 →
CPU1, Riser1 →
CPU1, Riser0 →
CPU1, Riser2 →
CPU1, Riser3 →

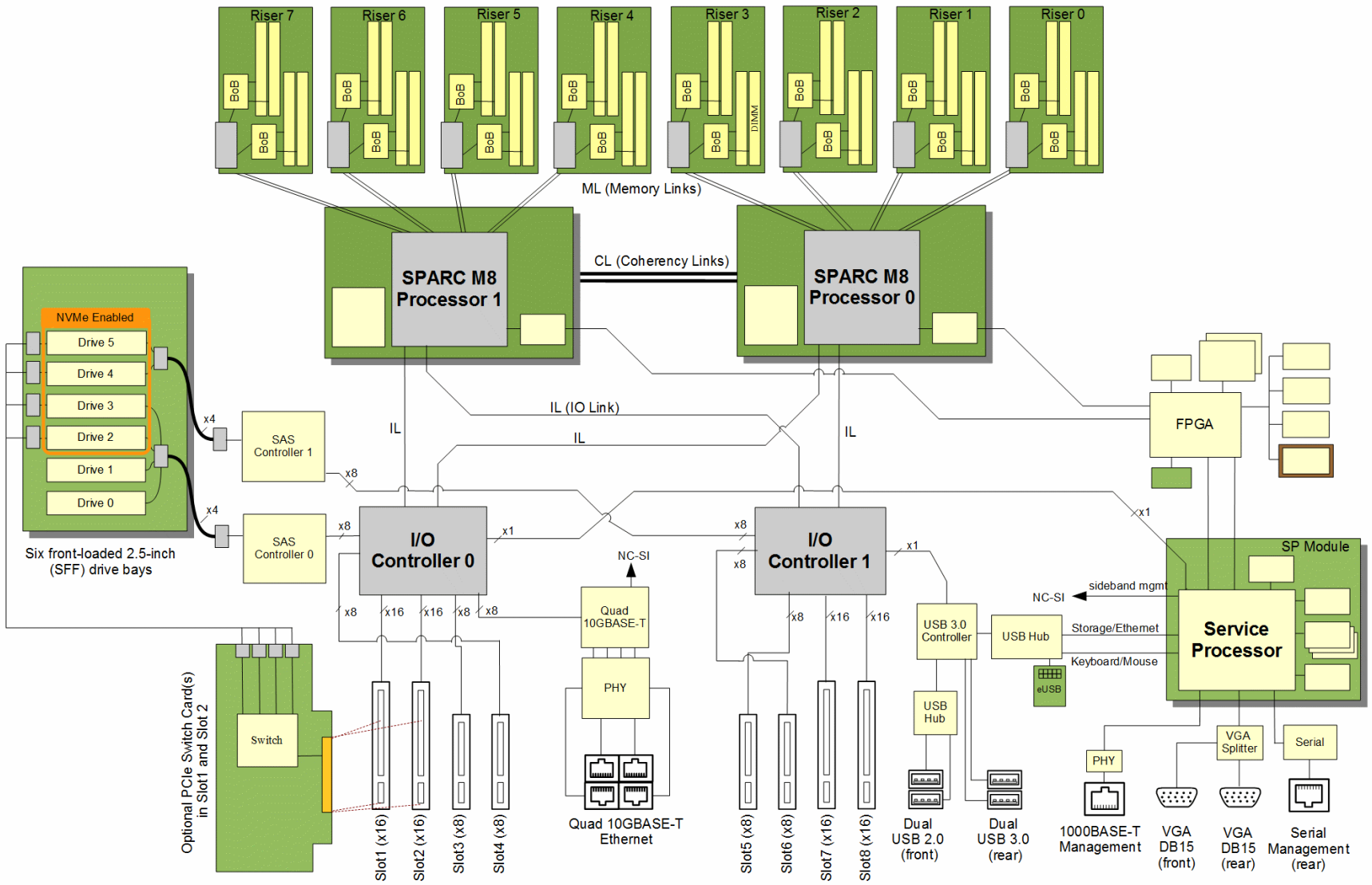


← CPU0

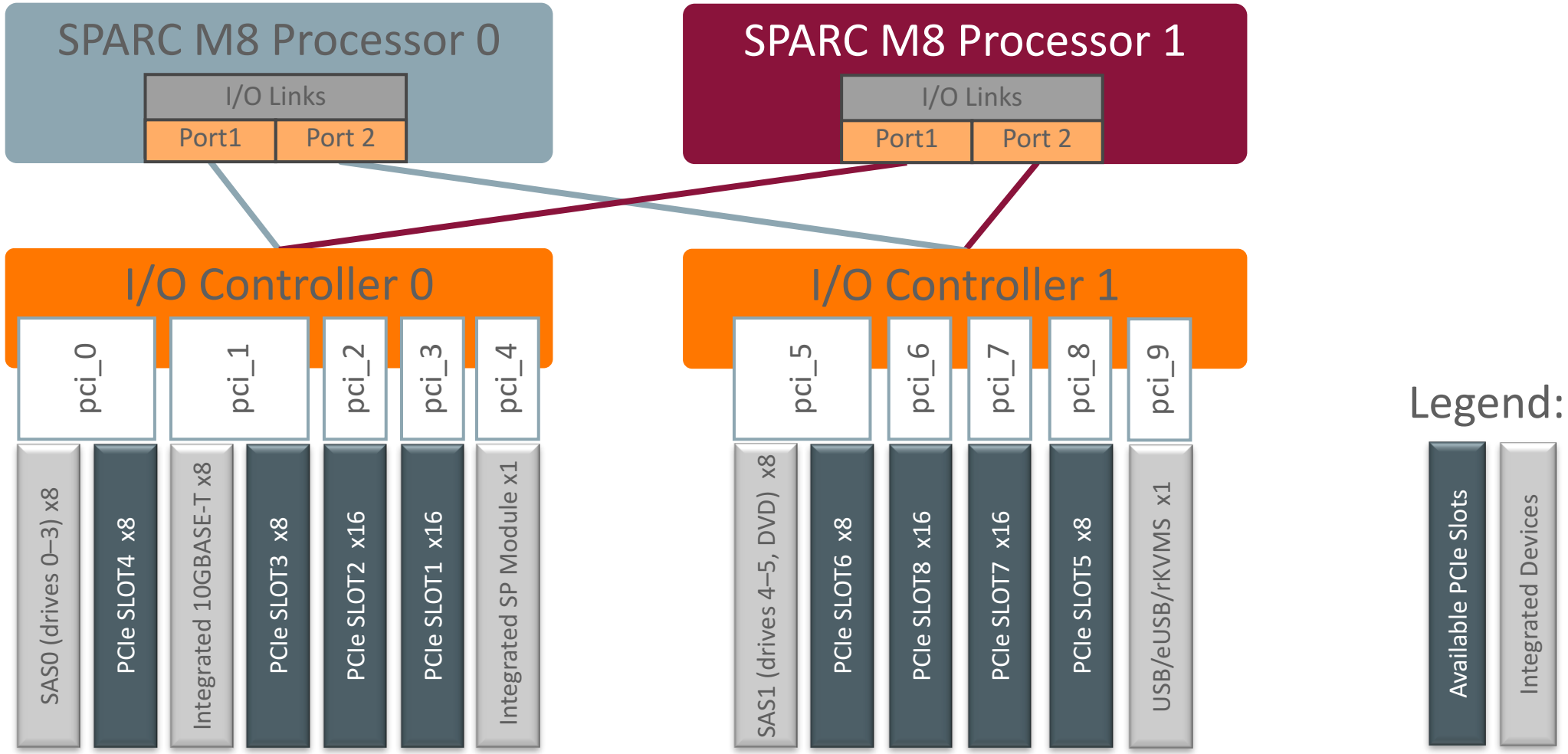
Rear of the Enclosure →

← CPU1

SPARC T8-2 Server: Block Diagram



SPARC T8-2 Server: Device Map



* First optional NVMe PCIe switch card (x8) for NVMe drive support goes into SLOT1
** Second optional NVMe PCIe switch card (x8) for NVMe drive support goes into SLOT2

SPARC T8-2 Server: PCIe I/O Mapping

Name	I/O ASIC	RC	Port	Width	pci_x*	Path	Notes
NET0-NET4	0	1	0	x8	pci_1	/pci@300/pci@1	Integrated 10GBASE-T (Net0, Net1, Net2, Net3)
SAS0	0	0	0	x8	pci_0	/pci@301/pci@1/scsi@0	Integrated SAS HBA #0, drive bays HDD #0–#3
SAS1	1	0	0	x8	pci_5	/pci@303/pci@1/scsi@0	Integrated SAS HBA #1, drive bays HDD #4–#5
SLOT1	0	3	0	x16	pci_3	/pci@306/pci@1	Dedicated root complex. Option: First NVMe PCIe switch, bays NVMe #0–#3 or #0–#1
SLOT2	0	2	0	x16	pci_2	/pci@307/pci@1	Dedicated root complex. Option: Second NVMe PCIe switch #1, bays NVMe #2–#3
SLOT3	0	1	1	x8	pci_1	/pci@300/pci@2	Root complex shared with Net0, Net1, Net2, and Net3
SLOT4	0	0	1	x8	pci_0	/pci@301/pci@2	Root complex shared with SAS0
SLOT5	1	3	1	x8	pci_8	/pci@302/pci@2	Dedicated root complex
SLOT6	1	0	1	x8	pci_5	/pci@303/pci@2	Root complex shared with SAS1
SLOT7	1	2	0	x16	pci_7	/pci@304/pci@1	Dedicated root complex
SLOT8	1	1	0	x16	pci_6	/pci@305/pci@1	Dedicated root complex
USB	1	4	0	x1	pci_9	/pci@308/pci@1/usb@0	See Installation Guide for device specific paths
SPM	0	4	0	x1	pci_4	/pci@309/pci@1/usb@0	

* Refer to the device map

SPARC T8-2 Server: CRUs and FRUs

- Hot-service CRUs
  - Fan modules
  - Power supplies
  - 2.5” SAS HDDs/SSDs
  - 2.5” NVMe SSDs
- Cold-service CRUs
  - Memory riser
  - Memory DIMMs
  - PCIe cards
  - System configuration PROM
  - System battery
  - Embedded USB flash memory (eUSB)

- Cold-service FRUs
  - Motherboard
  - Service processor (SP)
  - Disk backplane cage
  - Power supply backplane
  - Fan board

SPARC T8-4 Server

System Details

SPARC T8-4 Server: Overview

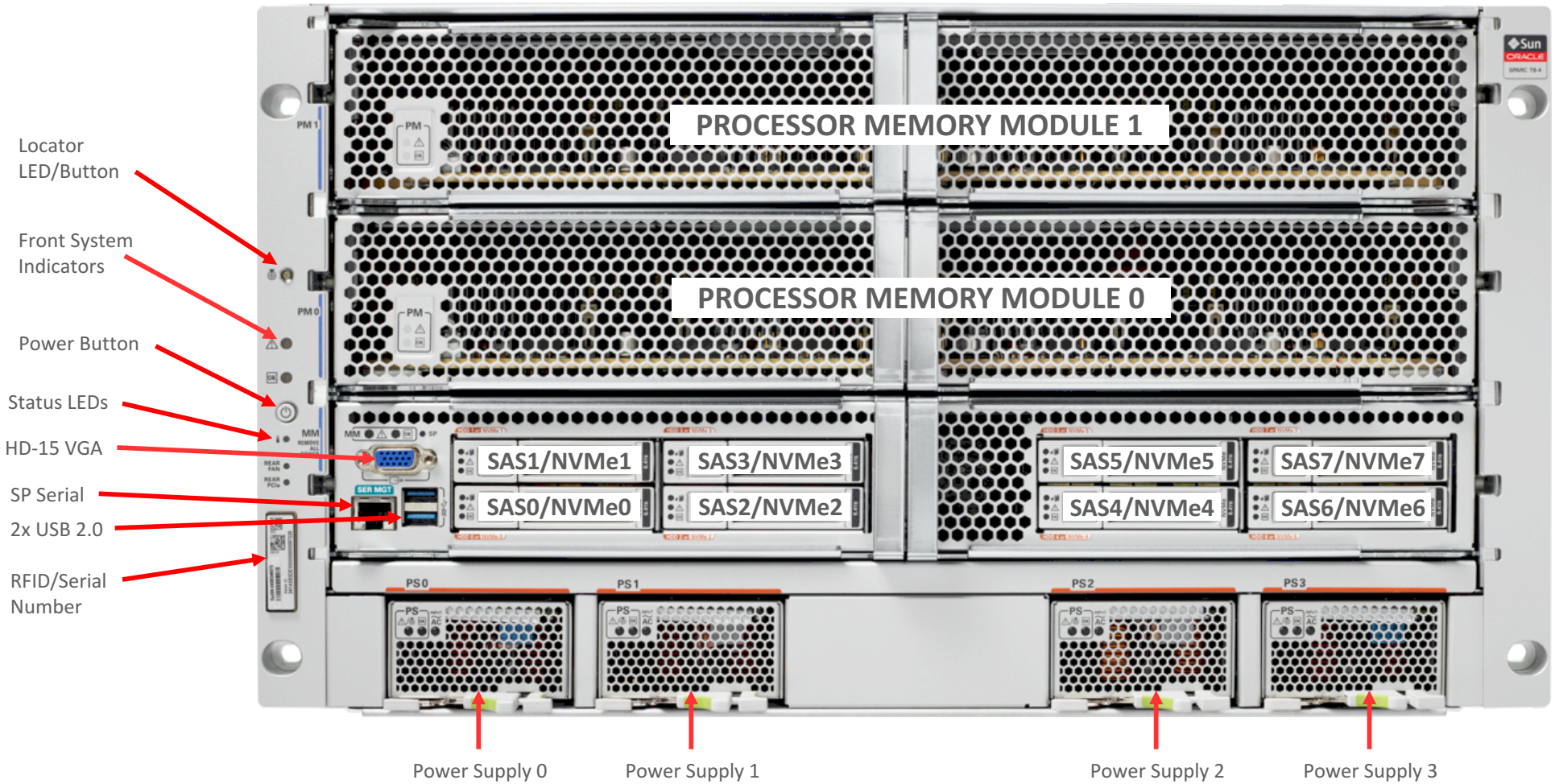
- 6U chassis, fits into 1,000 mm deep rack
- Two or four SPARC M8 processors (up to 128 cores/1024 threads)
- Up to 4 TB of memory
  - Half or full populated DIMM slots, up to 64 DIMMs
  - 32 GB and 64 GB DIMMs available
- Eight 2.5” SFF hot-pluggable disk drive bays
  - Up to 8 SAS HDDs/SSDs and/or NVMe SSDs
  - Mixing SAS and NVMe drives is supported
- Dual onboard SAS3 HBA (hardware RAID 0/1/10/1E)
- Two optional NVMe PCIe switch cards*
- Four 10GBASE-T ports
- 16 low-profile PCIe 3.0 slots, 8 x16 + 8 x8 (twelve PCIe buses)**
- Local (front and rear USB, rear video) and remote KVMS
- Five hot-swappable redundant fan modules (rear loading)
- Four 3,000 W (output) hot-swappable power supplies (N+N)

* Factory-configured or field add-on, installed in a pair
** Plus two internal slots for NVMe PCIe switch cards

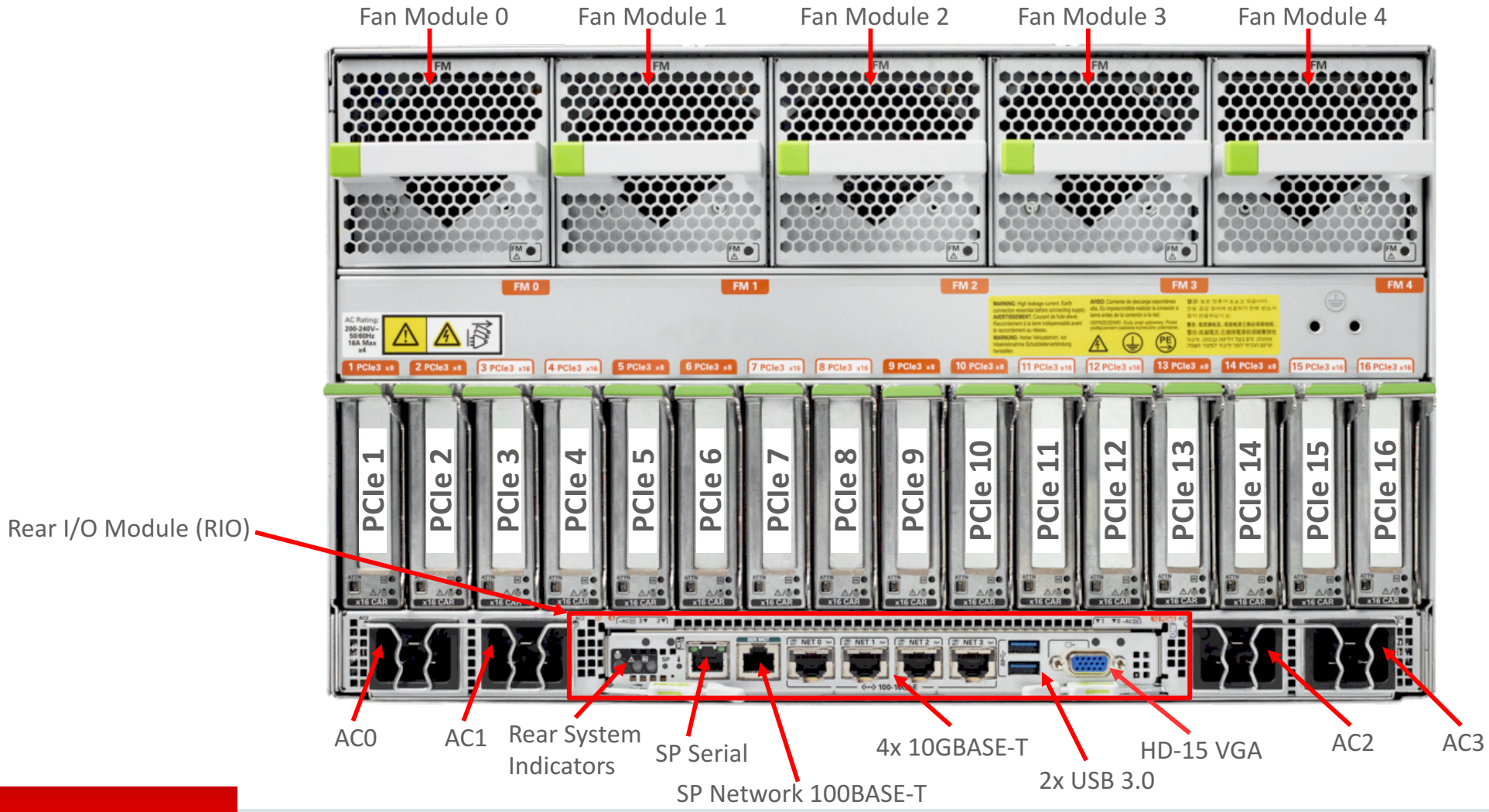
SPARC T8-4 Versus SPARC T7-4

Feature	SPARC T8-4	SPARC T7-4
Form Factor	6U, 31.9” deep	5U, 31.5” deep
CPU	4x SPARC M8 5.0 GHz (1,024 threads)	4x SPARC M7 4.13 GHz (1,024 threads)
Memory	DDR4-2400, 64x slots, Max. 4 TB w/ 64 GB DIMMs	DDR4-2133, 64x slots, Max. 4 TB w/ 64 GB DIMMs
Integrated Network Ports	4x 10GBASE-T	4x 10GBASE-T
Internal Storage	8x 2.5” hot-pluggable SFF bays 8x SAS 3.0 HDD or SSD, 8x NVMe SSDs	8x 2.5” hot-pluggable SFF bays 8x SAS 3.0 HDD or SSD, 8x NVMe SSDs
Removable Media	via rKVMS	via rKVMS
Management Ports	1x serial (RJ-45), 1x 100BASE-T, 2x VGA (HD-15)	1x serial (RJ-45), 1x 100BASE-T, 2x VGA (HD-15)
USB Ports	4x USB 3.0	4x USB 3.0
PCI Express Slots	16x PCIe 3.0 slots, 8 (x16) and 8 (x8) Hot-pluggable low-profile slots with card carrier	16x PCIe 3.0 slots, 8 (x16) and 8 (x8) Hot-pluggable low-profile slots with card carrier
Hot-swappable Fans	5 x redundant dual fan modules	5 x redundant dual fan modules
Hot-swappable Power Supplies (Nominal Output)	4 x 3000 watt AC, N+N	4 x 3000 watt AC, N+N

SPARC T8-4 Server: Front View



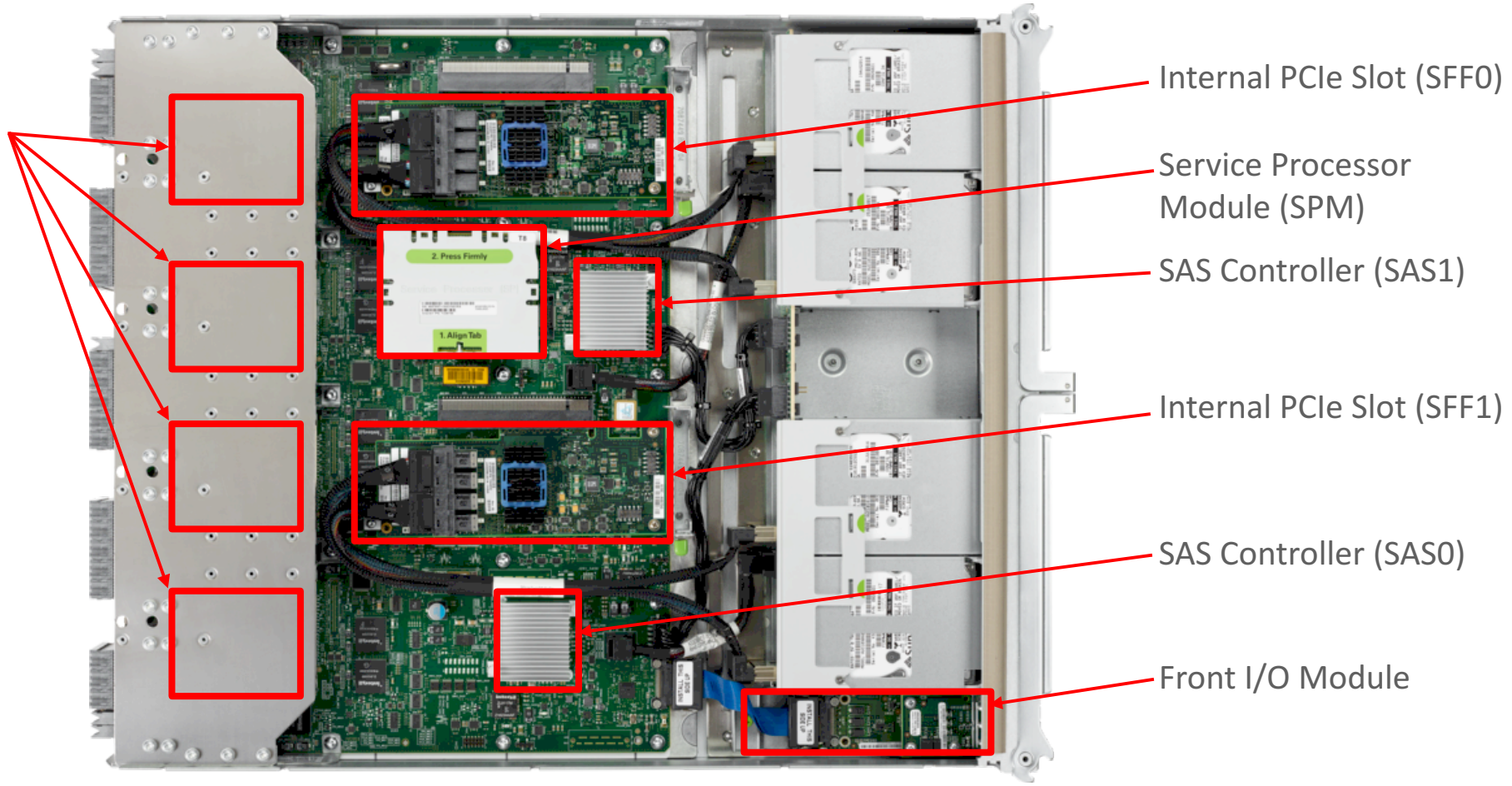
SPARC T8-4 Server: Rear View



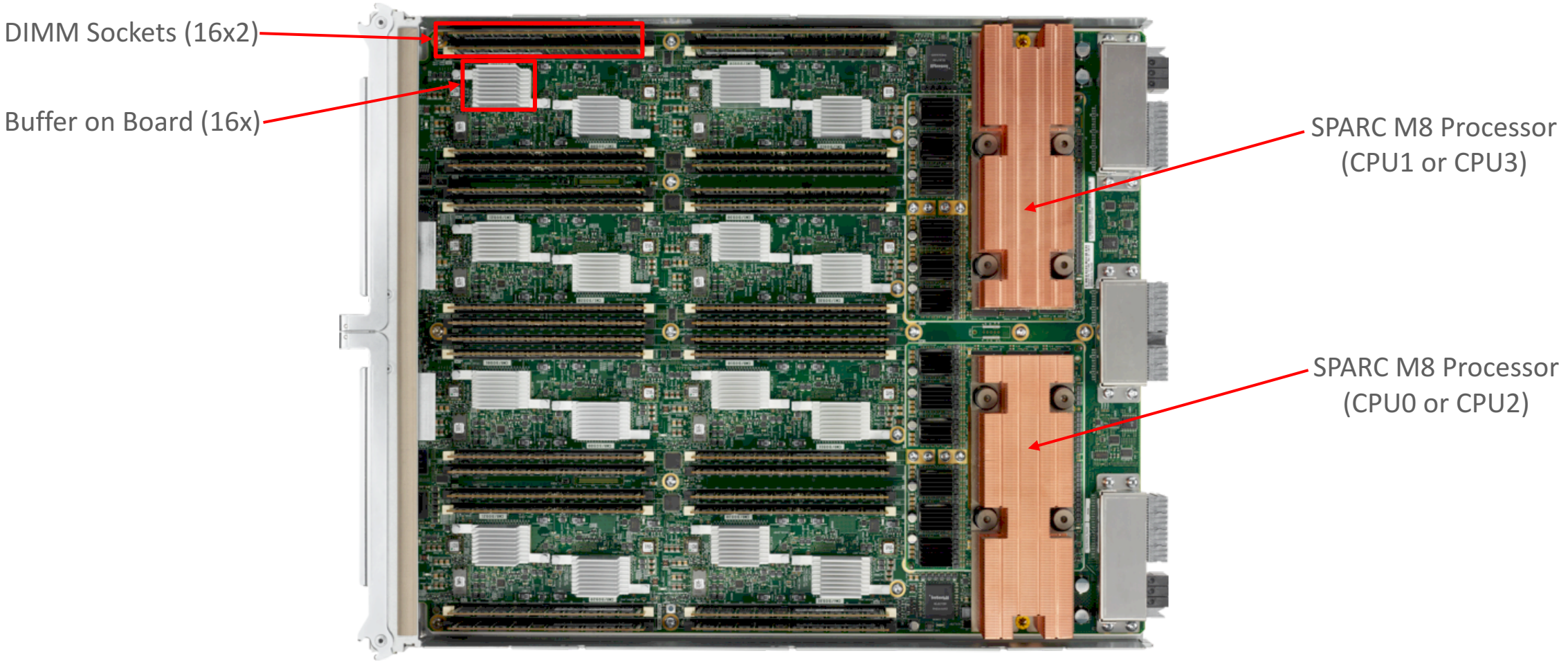
SPARC T8-4 Server: Configuration Policy

- Memory
  - Initial server order must have only one type of DIMMs, either half- populated or fully populated
    - Half-populated: 16 DIMMs per processor module (PM), 1 DIMM per BoB
    - Fully populated: 32 DIMMs per PM
  - Memory configuration on one PM0 can be different from PM1 via X-options
  - Both ATO and X-options: All DIMMs on a PM must be of same type and capacity
- Fans
  - Base package includes five N+1 redundant hot-swappable dual fan modules
  - System will continue to operate with fan failure(s) unless an over-temperature scenario occurs
  - System will power-on with one failed fan
  - System will *not* power-on with a missing fan module
- Power Supply Units (PSUs)
  - Base package includes 2+2 redundant hot-swappable power supplies
  - Two PSUs must be present and powered; requires 200–240 VAC , 50/60 Hz
  - System continues to operate with two failed PSUs or a failed grid
  - System will power-on with two failed PSUs or a failed grid

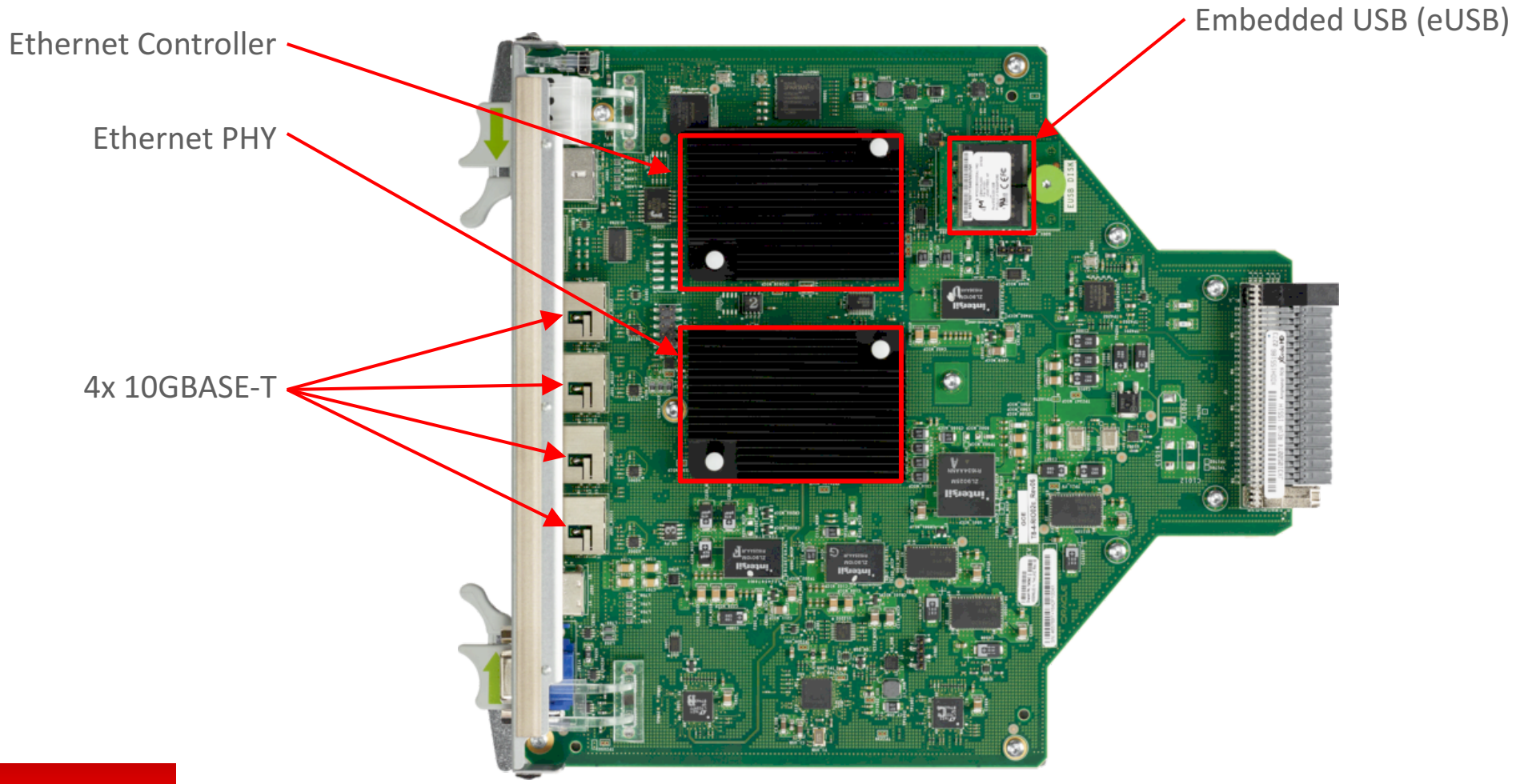
SPARC T8-4 Server: Main Module Top View



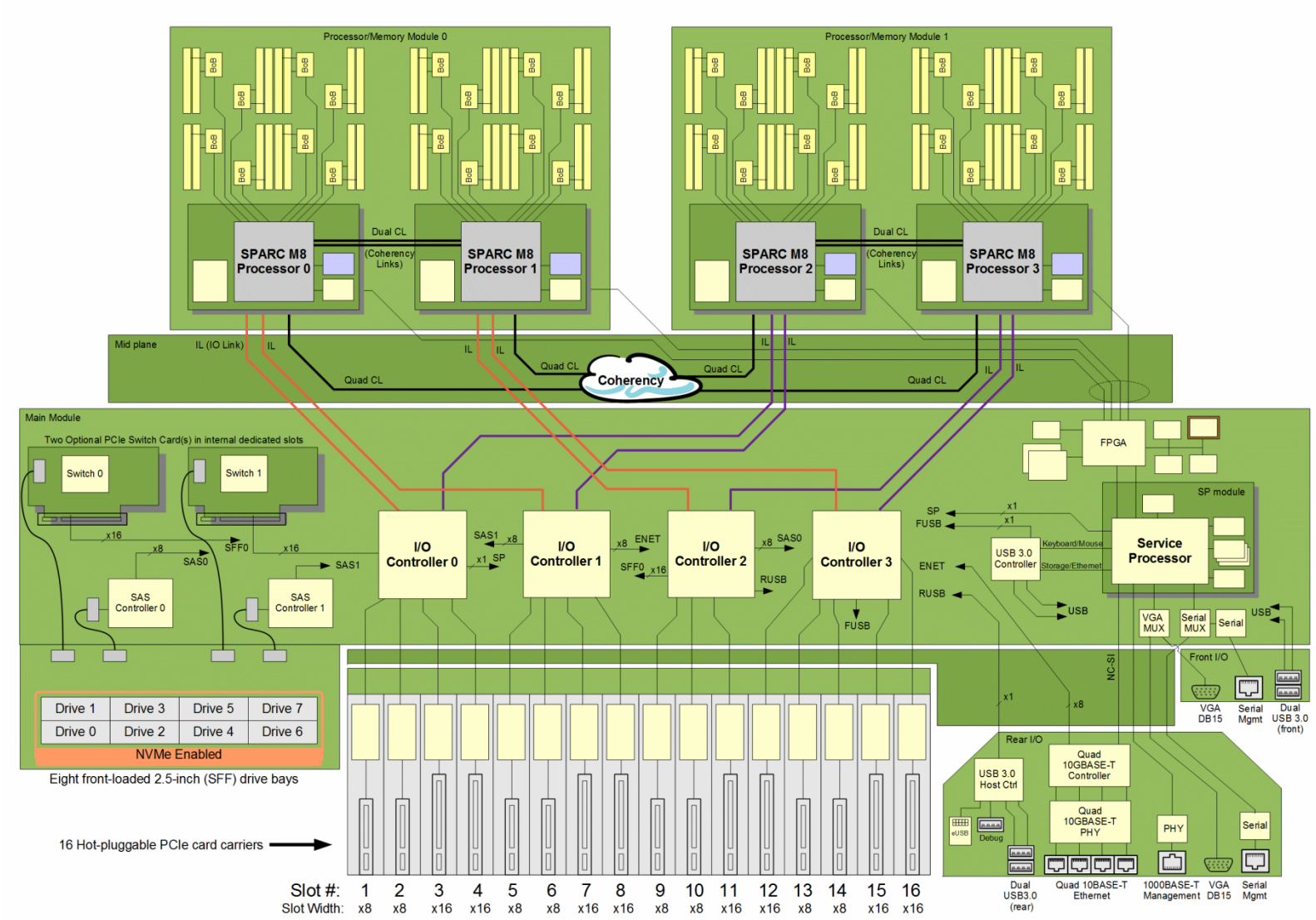
SPARC T8-4 Server: Processor Module Top View



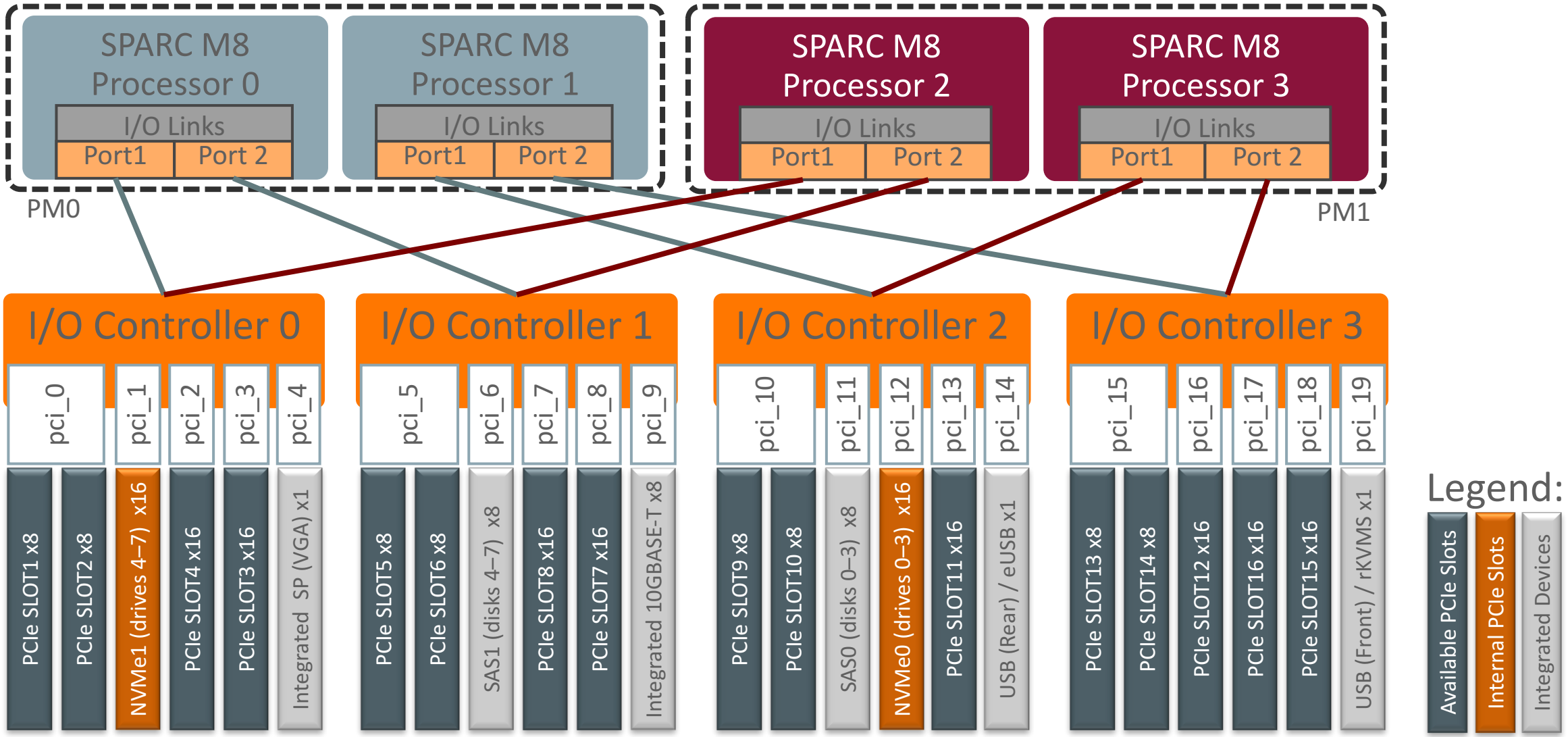
SPARC T8-4 Server: Rear I/O Module Top View



SPARC T8-4 Server: Block Diagram



SPARC T8-4 Server: Device Map



SPARC T8-4 Server: PCIe I/O Mapping (1 of 2)

Name	I/O ASIC	RC	Port	Width	pci_x*	Path	Notes
SLOT1	0	0	1	x8	pci_0	/pci@305/pci@2	Root complex shared with Slot2
SLOT2	0	0	0	x8	pci_0	/pci@305/pci@1	Root complex shared with Slot1
SFF 1	0	1	0	x16	pci_1	/pci@304/pci@1	Internal NVMe PCIe switch #1, bays NVMe #4–#7
SLOT4	0	2	0	x16	pci_2	/pci@307/pci@1	Dedicated root complex
SLOT3	0	3	0	x16	pci_3	/pci@306/pci@1	Dedicated root complex
SP	0	4	1	x4	pci_4	/pci@313/pci@1	SP (incl. VGA graphics), wired x1
SLOT5	1	0	1	x8	pci_5	/pci@308/pci@2	Root complex shared with Slot6
SLOT6	1	0	0	x8	pci_5	/pci@308/pci@1	Root complex shared with Slot5
SAS1	1	1	0	x8	pci_6	/pci@301/pci@1	Integrated SAS HBA #1, drive bays HDD #4–#7
SLOT8	1	2	0	x16	pci_7	/pci@30a/pci@1	Dedicated root complex
SLOT7	1	3	0	x16	pci_8	/pci@309/pci@1	Dedicated root complex
NET	1	4	0	x8	pci_9	/pci@301/pci@1	Integrated quad 10GBASE-T

* Refer to the device map

SPARC T8-4 Server: PCIe I/O Mapping (2 of 2)

Name	I/O ASIC	RC	Port	Width	pci_x*	Path	Notes
SLOT10	2	0	0	x8	pci_10	/pci@30b/pci@1	Root complex shared with Slot9
SLOT9	2	0	1	x8	pci_10	/pci@30b/pci@2	Root complex shared with Slot10
SAS0	2	1	1	x8	pci_11	/pci@300/pci@1	Integrated SAS HBA #0, drive bays HDD #0–#3
SFF0	2	2	0	x16	pci_12	/pci@303/pci@1	Internal NVMe PCIe switch #0, bays NVMe #0–#3
SLOT11	2	3	0	x16	pci_13	/pci@30c/pci@1	Dedicated root complex
RUSB	2	4	0	x4	pci_14	/pci@312/pci@2	Rear USB ports and eUSB, wired x1
SLOT14	3	0	0	x8	pci_15	/pci@30e/pci@1	Root complex shared with Slot13
SLOT13	3	0	1	x8	pci_15	/pci@30e/pci@2	Root complex shared with Slot14
SLOT12	3	1	0	x16	pci_16	/pci@30d/pci@1	Dedicated root complex
SLOT16	3	2	0	x16	pci_17	/pci@310/pci@1	Dedicated root complex
SLOT15	3	3	0	x16	pci_18	/pci@30f/pci@1	Dedicated root complex
FUSB	3	4	0	x4	pci_19	/pci@311/pci@1	Front USB ports and SP (rKMS), wired x1

* Refer to the device map

SPARC T8-4 Server: CRUs and FRUs

- Hot-service CRUs
  - Fan modules
  - Power supplies
  - 2.5” SAS HDDs/SSDs
  - 2.5” NVMe SSDs
  - PCIe cards in the carrier
- Cold-service CRUs
  - Main module (MM)
  - Processor module (PM)
  - Memory DIMMs
  - Front I/O assembly
  - Internal PCIe cards (no carrier)
- Cold-service FRUs
  - Service processor module (SPM)
  - System configuration PROM
  - System battery
  - Rear I/O module (RIO)
  - Embedded USB flash memory (eUSB)
  - Rear chassis subassembly

SPARC M8-8 Server

System Details

SPARC M8-8 Server: Overview

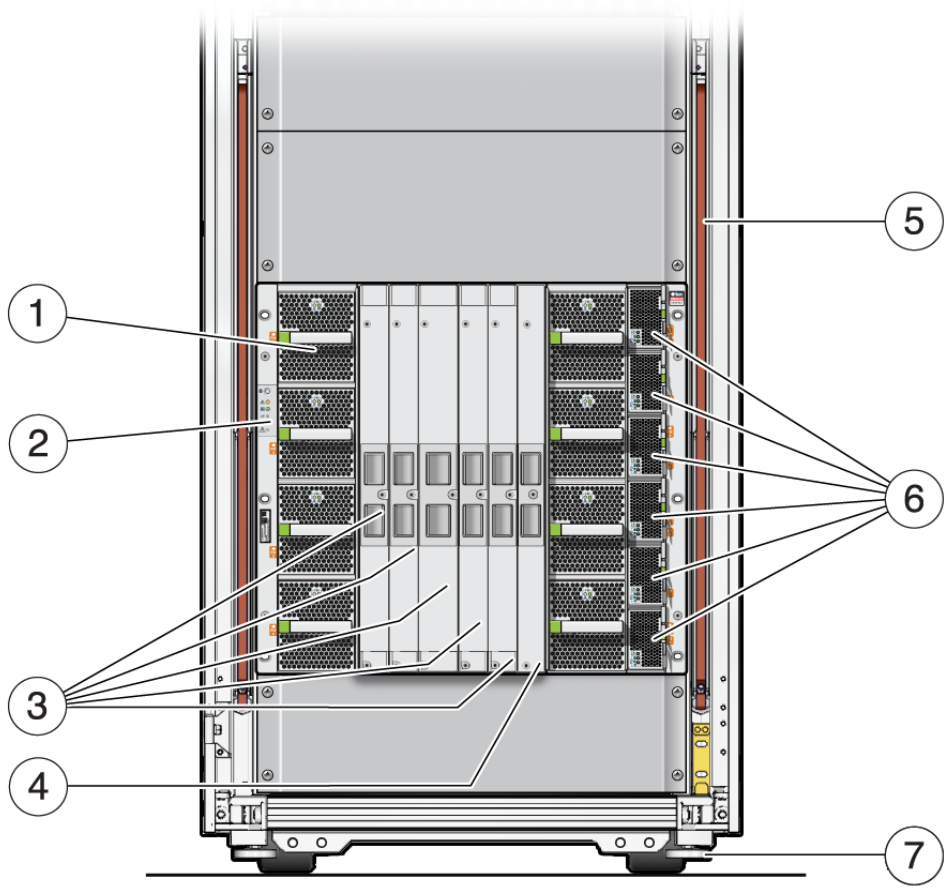
- Two factory-configured options:
  - One static physical domain with up to 8 processors
  - Two static physical domains, each up to 4 processors
- 2 to 8 SPARC M8 processors (up to 256 cores and 2,048 threads)
- Up to 8 TB of memory
  - 8 or 16 DIMMs per processor (32 GB or 64 GB DIMMs)
- Up to 24 PCIe 3.0 (x16) slots
  - 3 PCIe slots per processor, dedicated PCIe buses per slot
- Redundant hot-swappable SPs with automatic failover
- Redundant system clocks
- Eight hot-swappable redundant fan modules
- Six 3,000 W (output) hot-swappable power supplies (N+N)
- Power input: 3-phase PDUs, or 1-phase direct
- Factory mounted in a system rack (Oracle Rack Cabinet 1242)
- Optional: Standalone 10U chassis (requires 1,200 mm deep rack)

SPARC M8-8 Versus SPARC M7-8

Feature	SPARC M8-8	SPARC M7-8
Form Factor	System rack: 600 mm wide, 1200 mm deep, 2000 mm (78.74”) high Standalone: 10U, 813 mm (32”) deep chassis; Min. 1,200 mm deep rack	System rack: 600 mm wide, 1200 mm deep, 1998 mm (78 .66”) high Standalone: 10U, 813 mm (32”) deep chassis; Min. 1,200 mm deep rack
Physical Domains	1 or 2	1 or 2
Processor	SPARC M8, 32 cores, 5.0 GHz	SPARC M7, 32 cores, 4.13 GHz
Max. Processors/Cores/Threads	8/256/2048	8/256/2048
Memory	DDR4-2400, 128x slots; Max. 8 TB w/ 64 GB DIMMs	DDR4-2133, 128x slots; Max. 8 TB w/ 64 GB DIMMs
Integrated Network Ports	PCIe NIC cards	PCIe NIC cards
Internal Storage	Up to 16 x Oracle Flash Accelerator F640 PCIe Card, hot-pluggable	Up to 16 x Oracle Flash Accelerator F320 PCIe Card, hot-pluggable
Removable Media	via rKVMS	via rKVMS
Management Ports	Dual SPs, ports per SP: 1x or 2x serial (RJ-45), 1x 1000BASE-T	Dual SPs, ports per SP: 1x or 2x serial (RJ-45), 1x 1000BASE-T
PCI Express Slots	Up to 24x PCIe 3.0 (x16) slots Hot-pluggable low-profile slots with card carrier	Up to 24x PCIe 3.0 (x16) slots Hot-pluggable low-profile slots with card carrier
Fans	8 x redundant hot-swappable dual fan modules	8 x redundant hot-swappable dual fan modules
Power Supplies (Nominal Output)	6 x 3000 watt AC, N+N	6 x 3000 watt AC, N+N

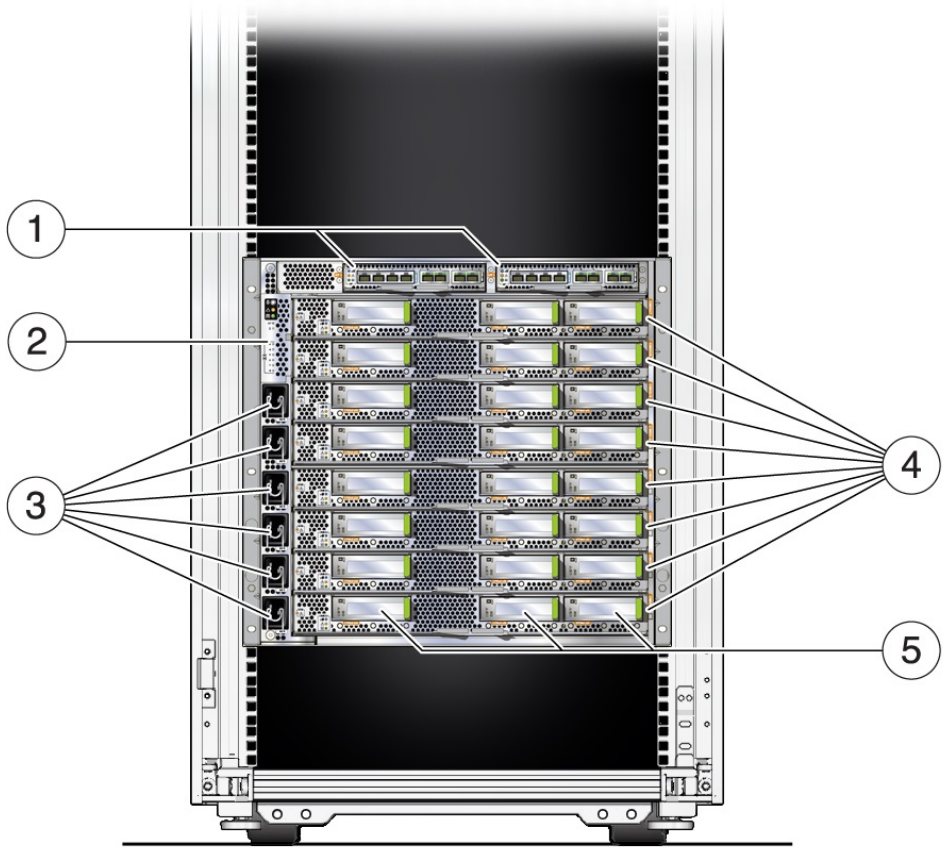
SPARC M8-8 Server: Front View

#	Description
1	Fans
2	Front indicator panel
3	Coherence link (CL) interconnects
4	SP interconnect
5	Anti-tilt bar (one of two)
6	Power supplies
7	Leveling feet



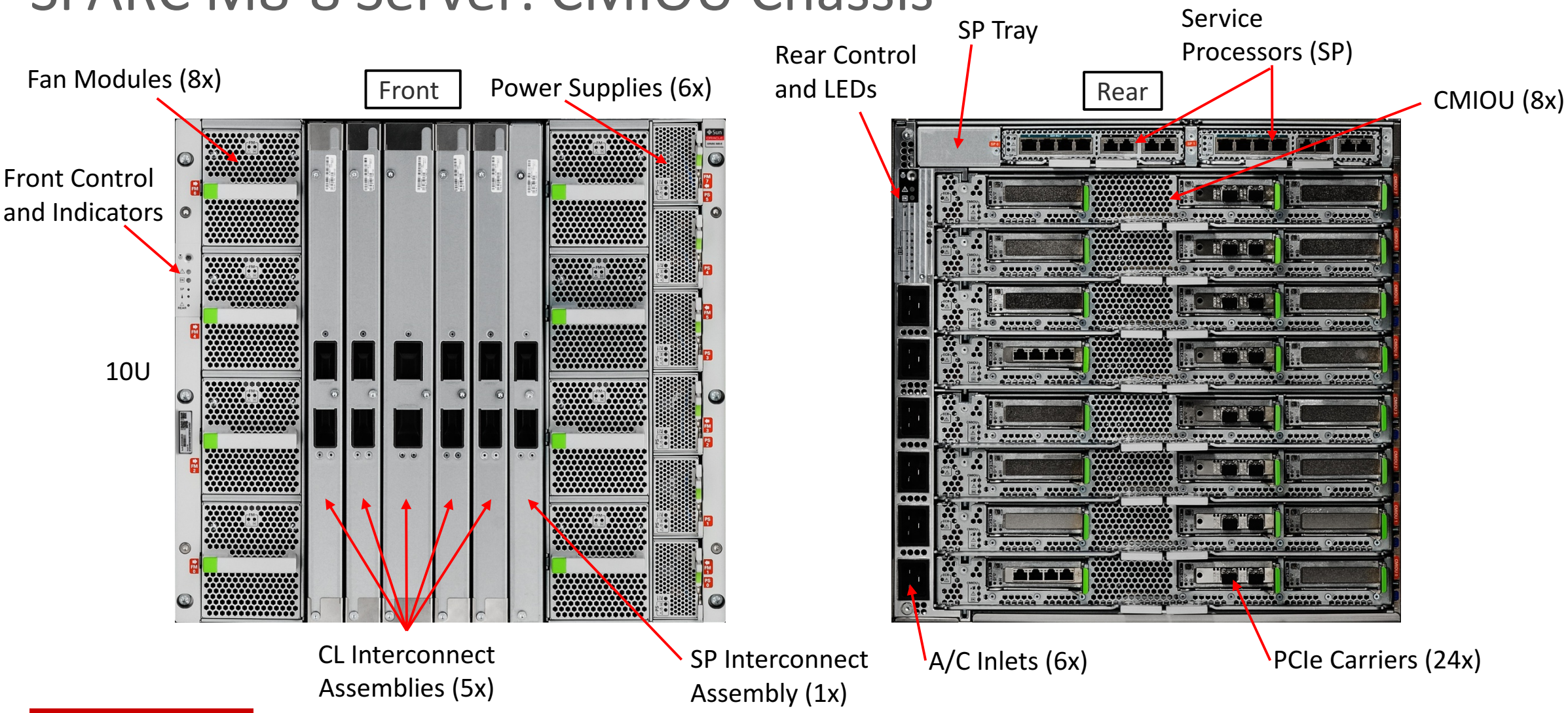
SPARC M8-8 Server: Rear View

#	Description
1	Service processors
2	Rear indicator panel
3	6x AC inputs (1-phase, C20)
4	8x CMIOU[1] boards
5	24x PCIe slots with carriers



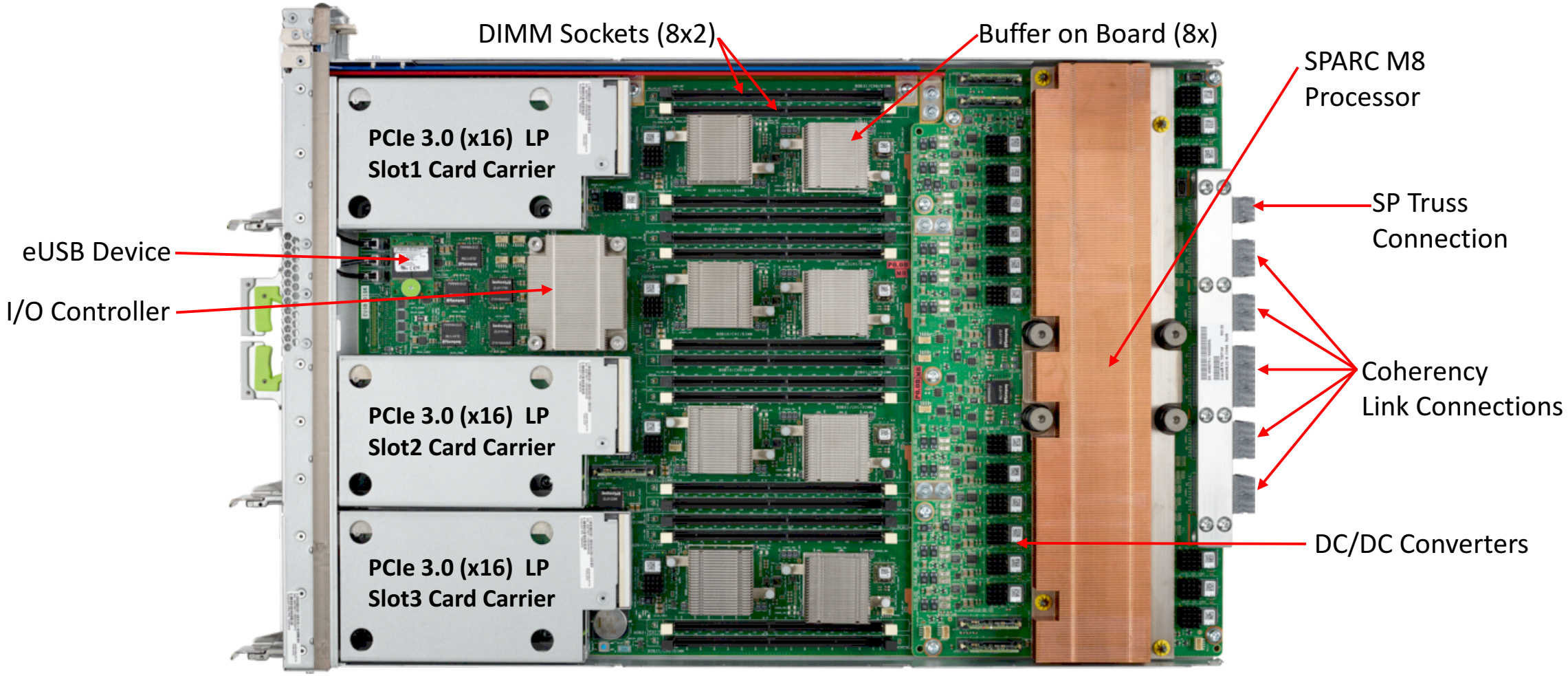
(1) CPU, memory and I/O unit (CMIOU)

SPARC M8-8 Server: CMIOU Chassis



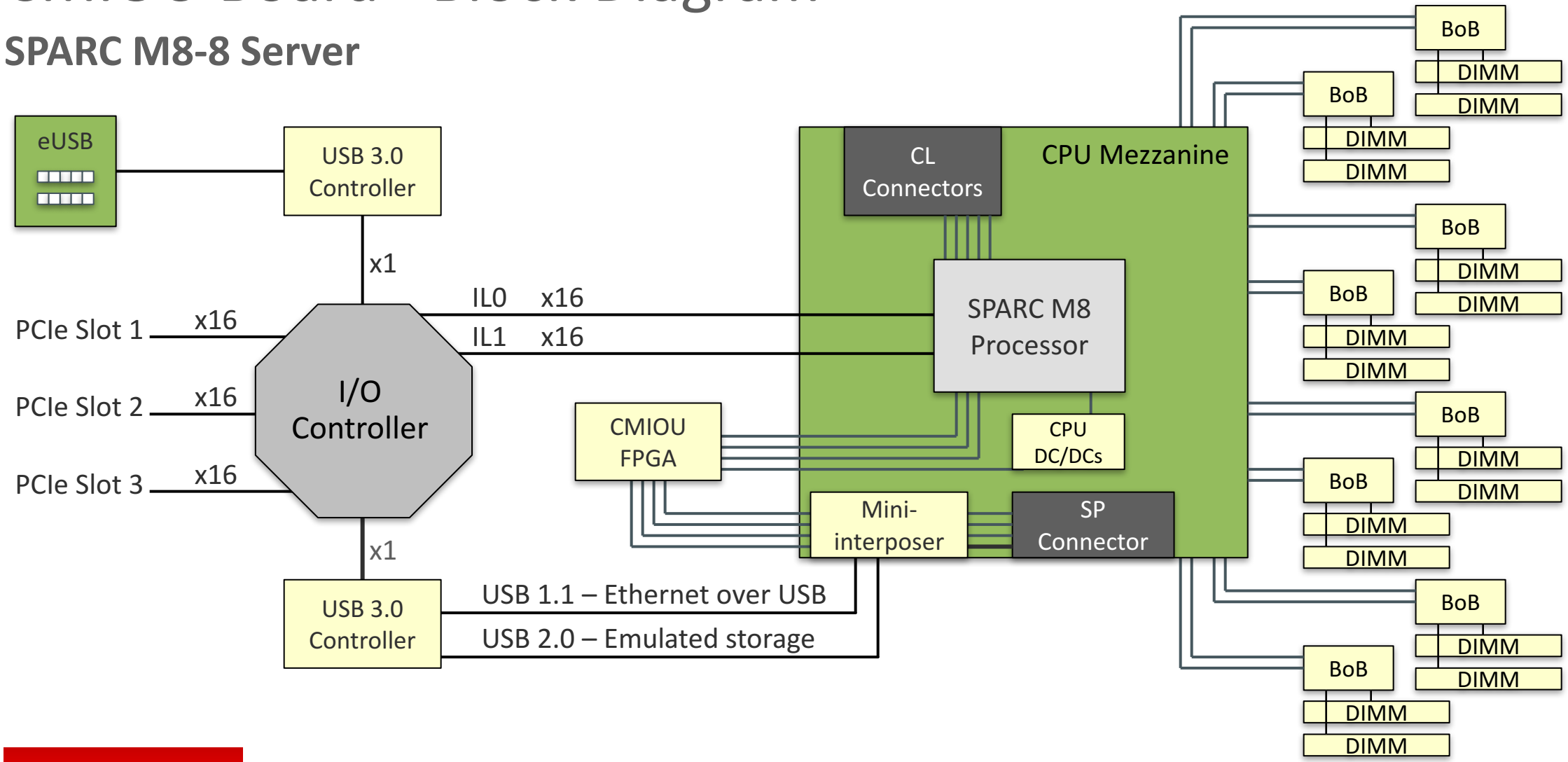
CMIOU Board

SPARC M8-8 Server

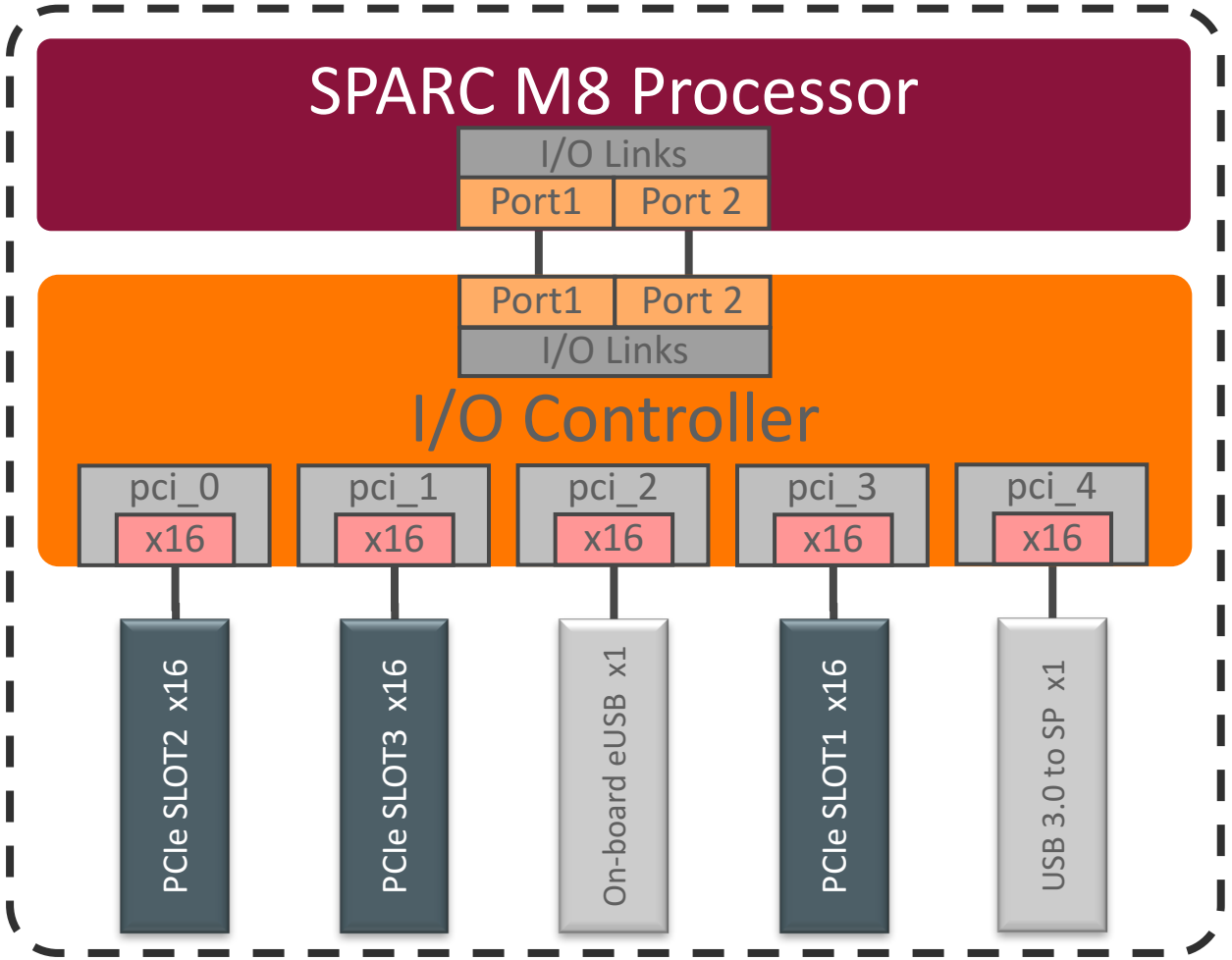


CMIOU Board - Block Diagram

SPARC M8-8 Server



SPARC M8-8 CMIOU Device Map



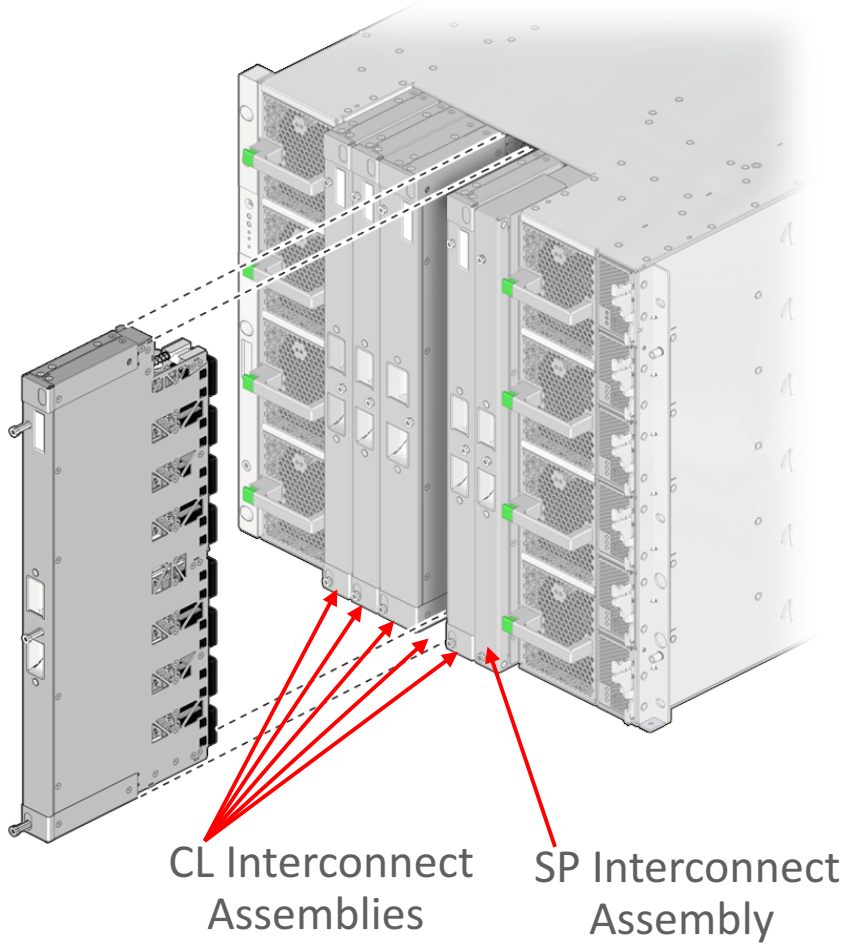
- A SPARC M8-8 server include two or more CMIOU boards.
- A simplified internal architecture of a CMIOU board is shown (e.g. processor coherency links to other CMIOUs are not shown).
- SP connections are used only in the first two DCU slots.
- "pci_0" root complex enumeration is shown for CMIOU 0 and varies depending on the location of the CMIOU board.

Legend:

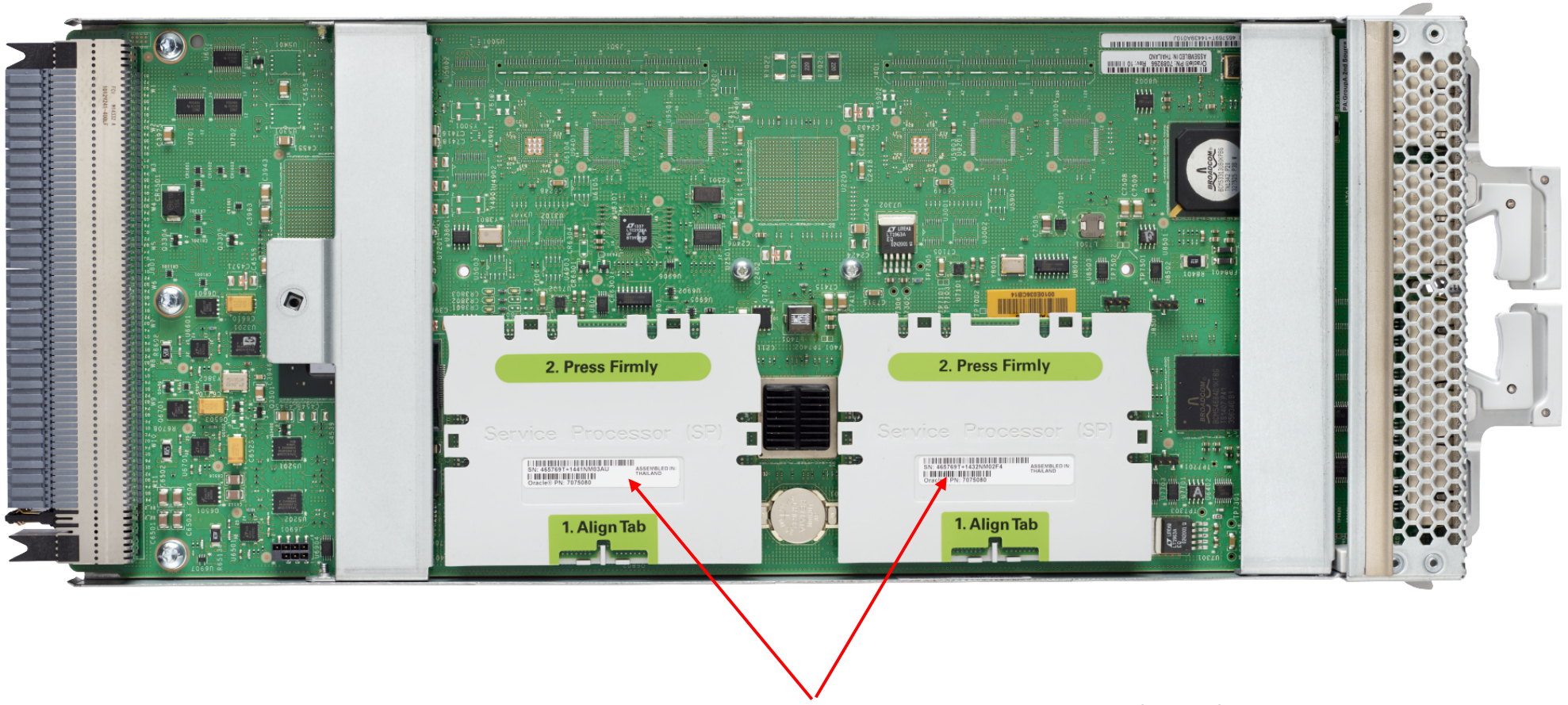
SPARC M8-8 8 Server

System Interconnect Assemblies

- Connect devices internally within the chassis
- CL interconnect assemblies
  - Direct coherence link connectivity
  - Two form factors: Wide and narrow
- SP interconnect assemblies
  - CMOIU-to-SP connectivity
- Wiring is different for the two server variants
  - Single or dual PDoms



SPARC M8-8 Server: Service Processor Top View



Dual Service Processor Modules (SPM)

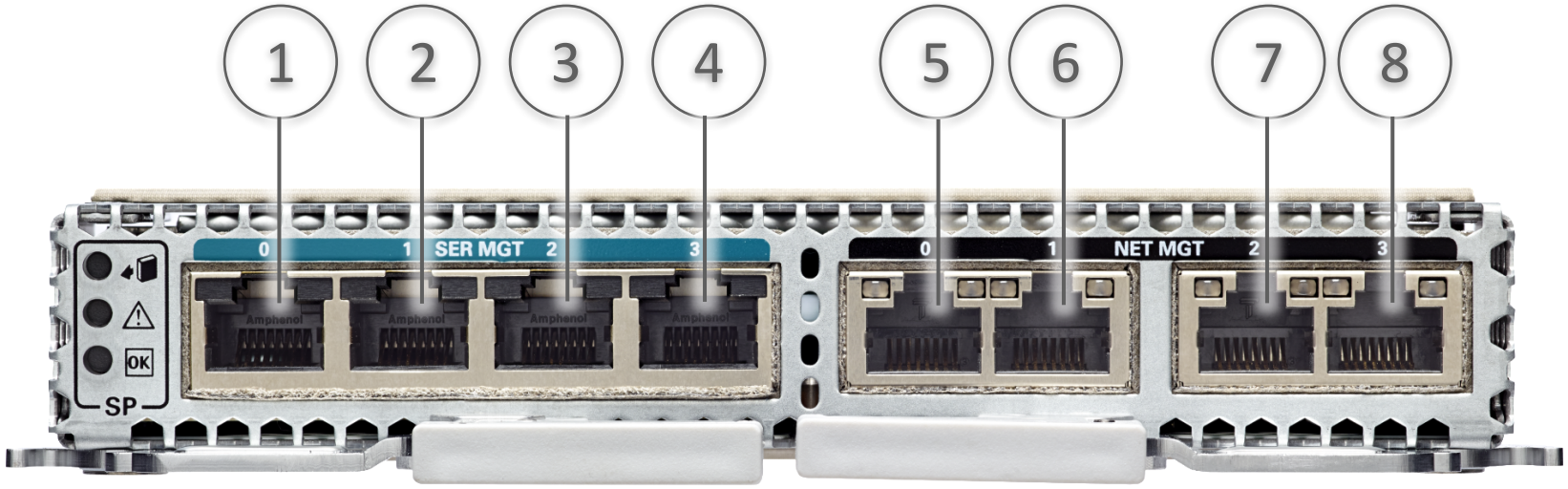
SPARC M8-8 Server: Service Processor (SP)

- All models have dual SPs with an active/standby configuration
  - Automatic failover in the event of a failure
  - One or two serial (RJ-45) ports (one per SPM)
  - One 1000BASE-T port per SP
- SP is responsible for:
  - Initialization and managing the CPU, memory controller, and DIMMs
  - Environmental monitoring, such as tracking sensors and adjusting fans/cooling
  - Fault management
  - Hosting remote console (VNC) and emulated storage for the PDom
- All PDoms are supported with redundant SP functionality
  - SPARC M8-8 with 1 PDom: Dual SPs with 1 SPM in each supporting 1 PDom
  - SPARC M8-8 with 2 PDoms: Dual SPs with 2 SPMs in each supporting 2 PDoms

SPARC M8-8 Server: Service Processor Module

- The service processor module (SPM) contains the processor that runs the SP function
- Each SP includes one or two SPMs depending on model
  - SPARC M8-8 with one PDom: Each SP has only one SPM
  - SPARC M8-8 with two PDoms: Each SP has two SPMs
- Each PDom is connected to two SPMs in separate SPs providing an active/standby configuration with failover capability
- In dual PDom server, one SPM manages hardware items that are not associated with PDoms (PSUs, fans, and so on)

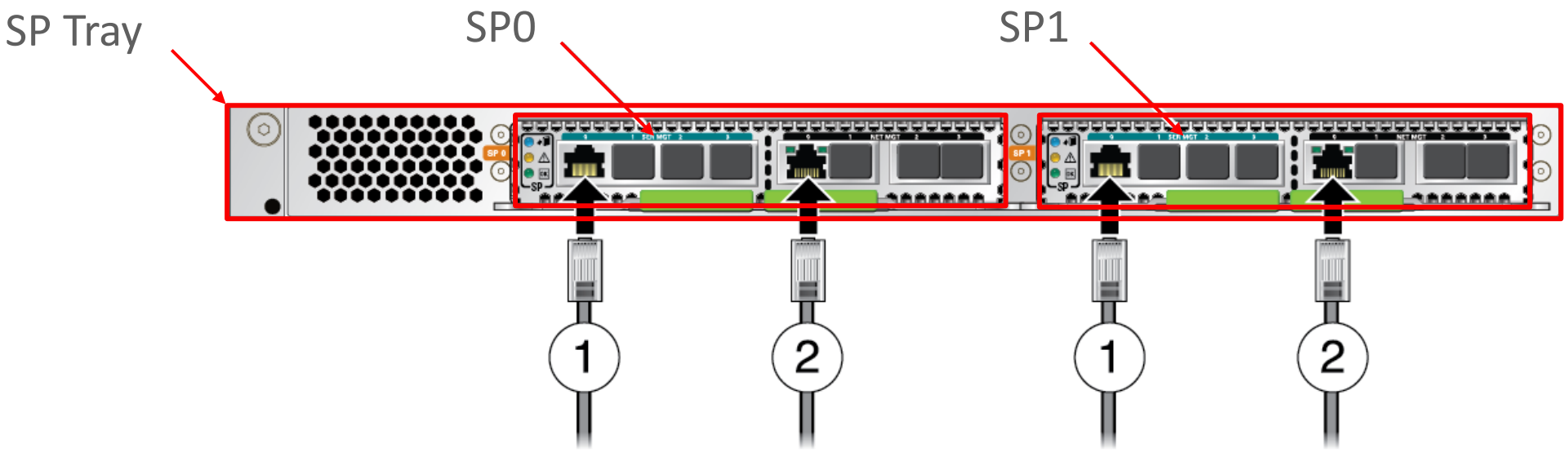
SPARC M8-8 Server: SP Rear Panel View



#	Description
1–4	Serial management ports (RJ-45 connector)
5–8	1000BASE-T network management ports

Note: Ports 3, 4, 6, 7, and 8 are not used in SPARC M8-8 server

SPARC M8-8 Servers: SP Management Ports



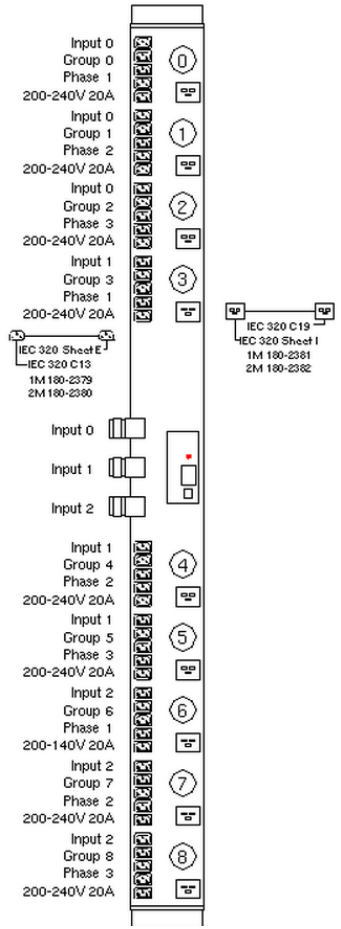
#	Description
1	Serial management ports (one or two SER MGT 0 ports per SP)
2	1000BASE-T network management ports (one NET MGT 0 port per SP)

SPARC M8-8 Server: Oracle Rack Cabinet 1242

- Recommended: Rack-mounting at factory
- Dual 3-phase PDUs must be included
- 29Us available after factory-installed SPARC M8-8
  - SPARC M8-8 chassis: 10U
  - 3-phase power cables: 3U at bottom or top
- Dimensions and doors
  - Height: 2,000 mm, 78.74 inches (42U)
  - Width: 600 mm, 23.62 inches
  - Depth: 1,200 mm, 47.24 inches
  - Single front door and split rear door

SPARC M8-8 Server: Power Distribution Unit (PDU)

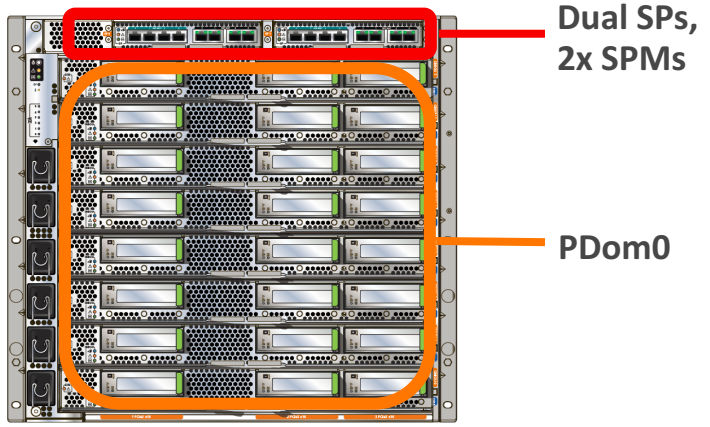
- Inputs (3-phase)
  - Low voltage (total of 26 kVA)
    - Three inputs: 24 A @ 208 VAC
    - Wiring: 4 W + ground
    - Connector : NEMA L21 5-pin 30 A
  - High voltage (total of 33 kVA)
    - Three inputs: 16 A @ 230/400 VAC
    - Wiring: 4 W + ground
    - Connector : Walther 210 5-pin 16 A
- Additional features
  - Power monitoring module



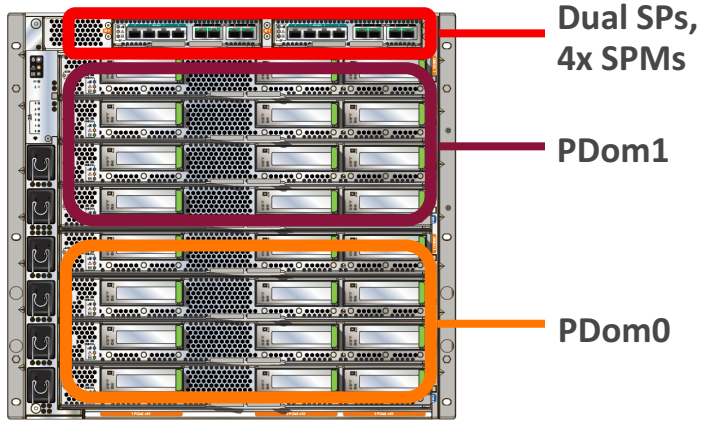
- Outputs (1-phase)
  - Low voltage
    - Nine 20 A 250 V 2-pole breaker
    - Nine C19 plugs ; forty-two C13 plugs
  - High voltage
    - Nine 20 A 25 0V 1-pole breaker
    - Nine C19 plugs ; forty-two C13 plugs
- Mechanical
  - Zero RU vertical PDU mounted in the rack side
  - Two PDUs per rack

SPARC M8-8 Server: PDom Population Rules

- SPARC M8-8 with one PDom: includes up to 8 processors
- SPARC M8-8 with two PDoms: each up to 4 processors
- Minimum of two CMIOU boards per system
- Minimum of two CMIOUs per PDom (unless empty), additional CMIOUs in increments of one
- Either half-populated or fully populated memory on CMIOU boards
  - All the memory on the CMIOU must be the same density and type
  - Mixing memory in PDoms is supported
- SPARC M8-8 CMIOUs are supported in the PDom1 of SPARC M7-8 server with dual PDoms (restrictions apply)



SPARC M8-8 with one PDom



SPARC M8-8 with two PDoms

SPARC M8-8 Server with One PDom

Glueless System Interconnect

Coherency Links (CL)

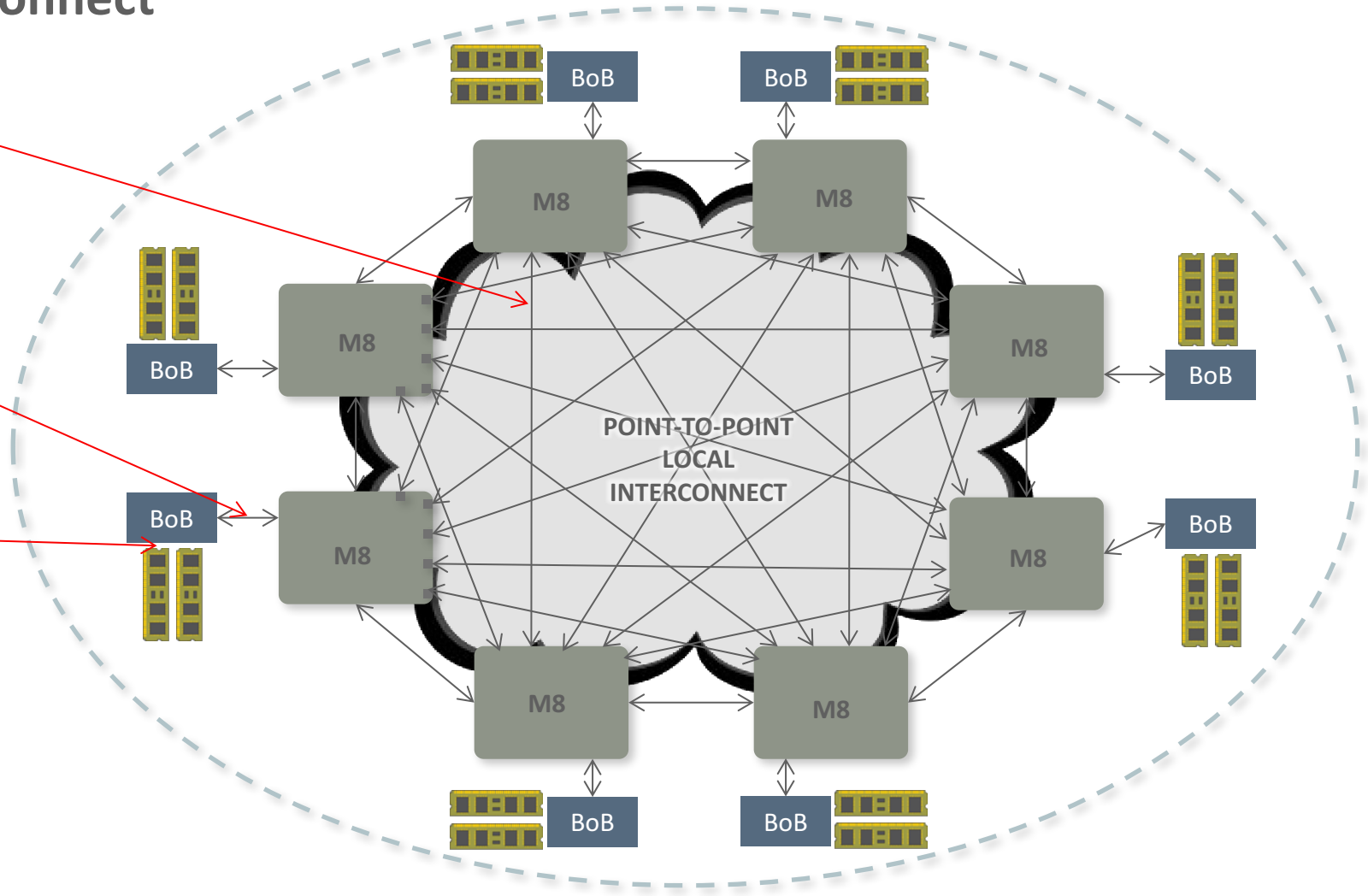
All-to-All local interconnect. The 8-way CPU-to-CPU fabric is seven Coherency Links from each CPU to the other 7 CPUs.

Memory Links (ML)

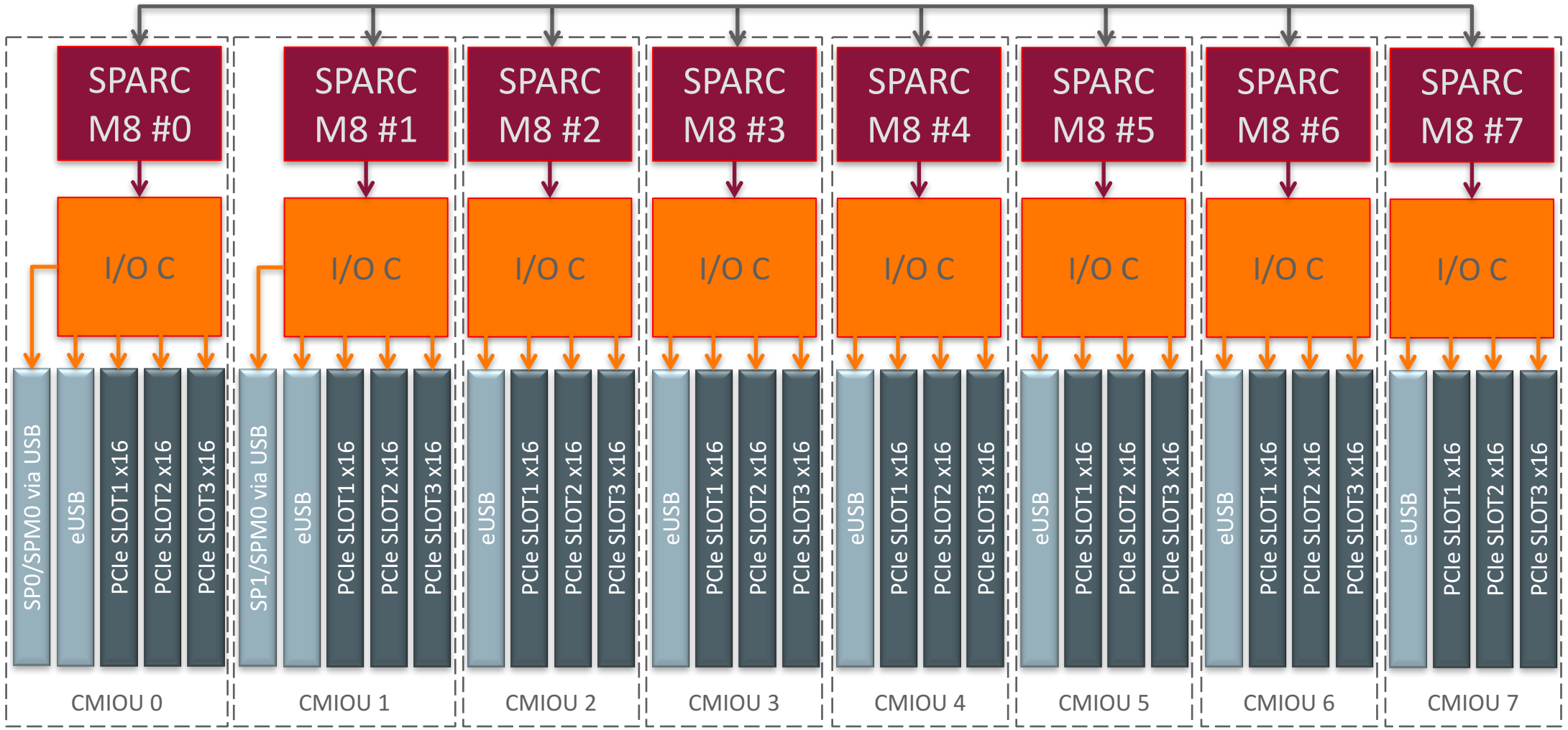
CPU-to- BoB connectivity. Each CPU is connected to eight Buffer-on-Board (BoB) chips.

DDR4 Channel

BoB-to-DIMM connectivity. Each BoB is connected to two DDR4 DIMMs. There are total of 16 DIMMs/CPU



SPARC M8-8 Server with One PDom: Device Map

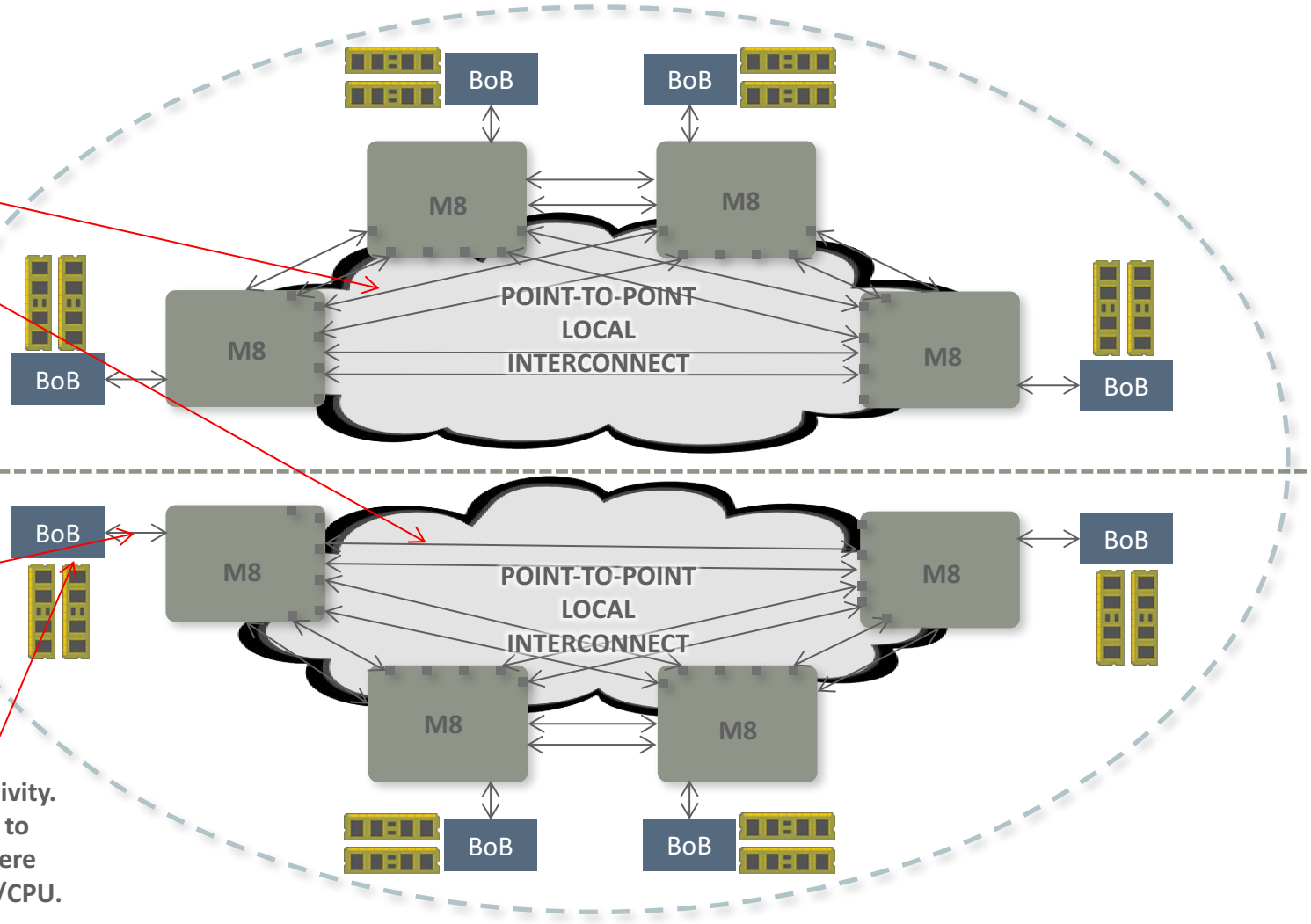


SPARC M8-8 Server with Two PDoms

Glueless System Interconnect

Coherency Links (CL)

Two separate All-to-All local interconnects. The 4-way CPU-to-CPU fabric is six Coherency Links from each CPU to the other 3 CPUs.



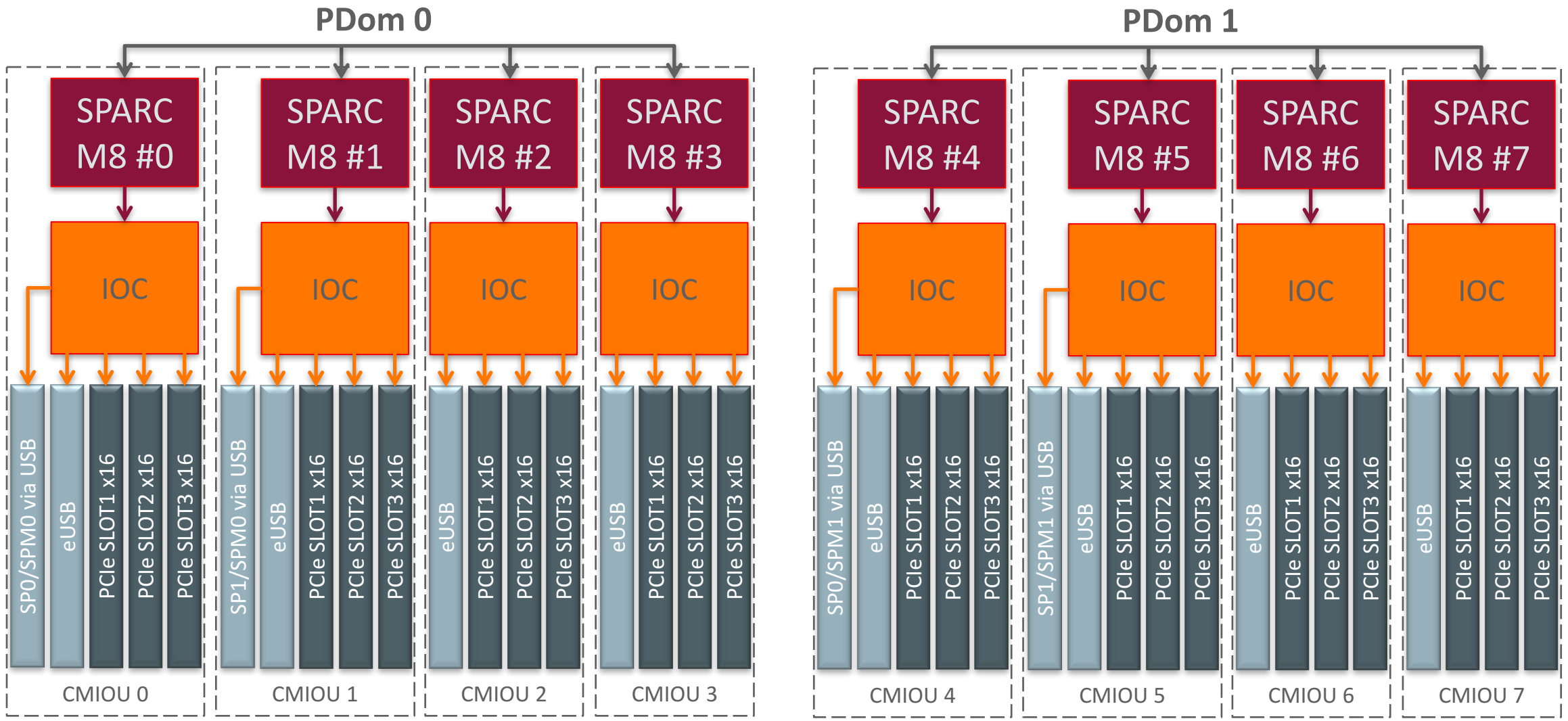
Memory Links (ML)

CPU-to- BoB connectivity. Each CPU is connected to 8 Buffer-on-Board (BoB) chips.

DDR4 Channel

BoB-to-DIMM connectivity. Each BoB is connected to two DDR4 DIMMs . There are total of 16 DIMMs/CPU.

SPARC M8-8 Server with Two PDoms: Device Map



Select CRUs and FRUs

SPARC M8-8 Server

- Hot-serviceable CRUs
  - Fan modules
  - Power supplies
  - PCIe cards in the carrier
- Cold-serviceable FRUs
  - Service processor tray
  - Interconnect assemblies
  - Front indicator panel
  - Power module

- Hot-serviceable FRUs
  - CMIOU board[1]
  - Memory DIMMs[1]
  - Service processor (SP)[2]
  - Service processor module on SP[2]
  - System battery on SP[2]
  - Embedded USB flash memory (eUSB)[1]
  - Power distribution units (PDU)

1) Can be hot-serviced if the CMIOU is not active
2) Can be hot-serviced if the SP is not active

SPARC M8-8 Server: Terminology

Term	Definition/Description
CPU Memory and I/O Unit (CMIOU)	Board equipped with one SPARC M8 processor, 16 DIMM slots, and three PCIe slots
CMIOU Chassis	The enclosure that houses up to 8 CMIOU boards and the service processors (SPs)
Buffer on Board (BoB)	An ASIC that interfaces between the memory DIMMs and the MCU on the SPARC M8 processor
Memory Controller Unit (MCU)	Each SPARC M8 processor has four MCUs in order to communicate with total of eight BoBs
Physical Domain (PDom)	An electrically fault-isolated hard partition
Service Processor (SP)	Connects to CMIOUs/SWUs/SPPs and communicates externally for monitoring and management
Service Processor Module (SPM)	Includes the SP processor chip that runs Oracle ILOM and connects to the CMIOUs. Resides on SP.
Interconnect Assemblies (aka Cable Trusses)	Wired connections between CMIOUs and/or SPMs
Coherency Link (aka CL or Clink)	Direct connection between two or more SPARC M8 processors forming the glueless system interconnect
I/O Link (aka IL or ILink)	Connection between SPARC M8 processor and the I/O controller that provides the PCIe root complexes

Agenda

1. Overview
2. System Details
3. New Technologies
4. RAS
5. Platform Management
6. Virtualization
7. Summary

Common Features and Technologies

SPARC M8 Processor–Based Servers

- DDR4 Memory Subsystem
  - DIMM sparing
- SPARC M8 Processor–Based Systems Interconnect
  - High speed direct coherency links, processor to processor
  - Plesiochronous system clock
- I/O Controller
  - X16–capable PCIe 3.0 with up to 5 buses per SPARC M8 processor
- NVM Express (NVMe) Interface for PCIe SSDs
- Embedded USB (eUSB) and Boot over InfiniBand

DIMM Sparing

Increased Availability and Reliability

- Automatically retires a DIMM that is potentially failing
- Used in presence of correctible errors; protects the system against potential future hard failures
- Retirement is done dynamically while the system is running
- Memory capacity remains unchanged and error protection is intact
- On a per-CPU basis, 1/16 of the memory capacity is held in reserve, to be used for isolating one entire DIMM from the configuration
- The content of one of the 16 DIMMs can be remapped into the reserved memory that resides in each of the other 15 DIMMs

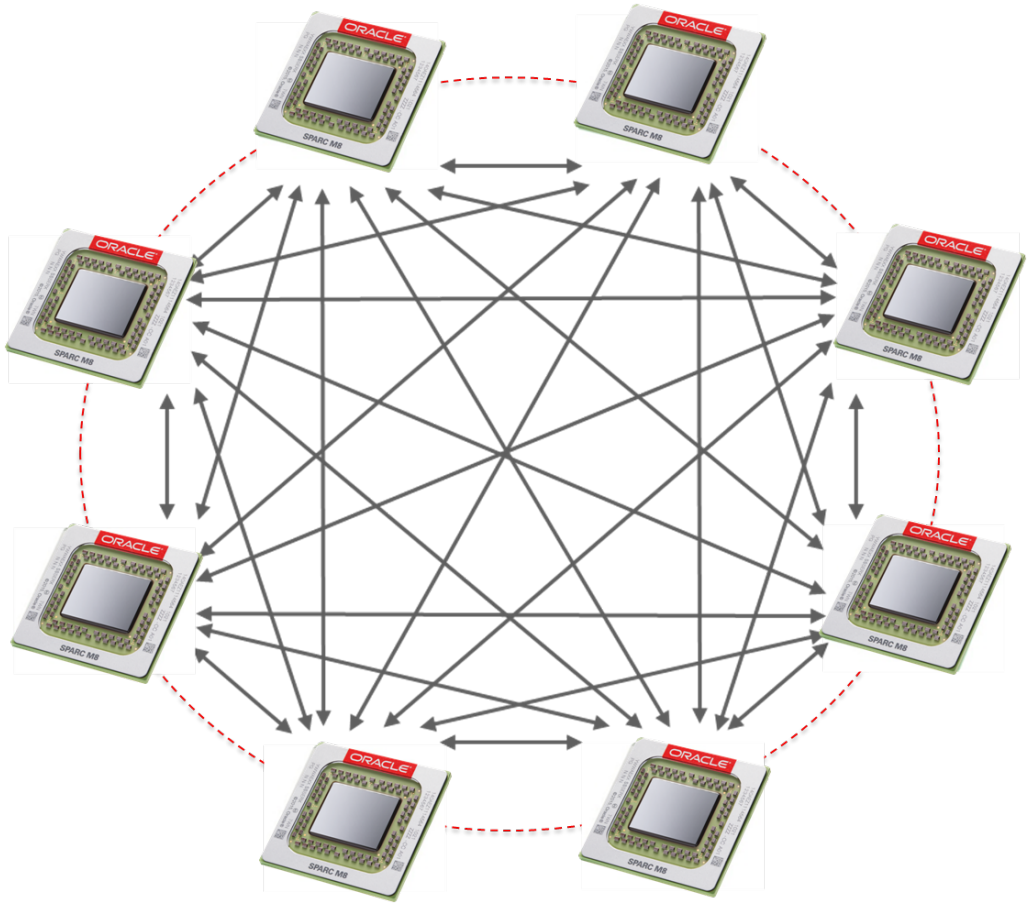
DIMM Sparing (cont.)

SPARC M8 Processor–Based Servers

- Applies to every CPU with all 16 DIMM slots populated
  - CPUs with only 8 DIMM slots populated cannot support sparing
  - No memory is held in reserve on CPUs with only 8 DIMM slots populated
- When a DIMM is spared, no service notification is generated
  - No change in capacity, no change in error protection capabilities
  - Increased availability due to avoiding scheduled maintenance outage
- Every CPU can spare independently
- If a second DIMM attached to the same CPU is determined to be faulty, a service notification will be generated
  - Service both faulty DIMMs at same time

SPARC M8 Processor–Based Servers: SMP Scalability

- Up to 8 processors directly connected (glueless)
- Dynamic congestion avoidance for data
- Coherence links
- Link-level reliability, availability, and serviceability (RAS) in hardware
  - CRC check and automatic message retry
  - Automatic lane retire (per direction)
  - Automatic link retrain and reinitialization
  - Built in PRBS testing during link training

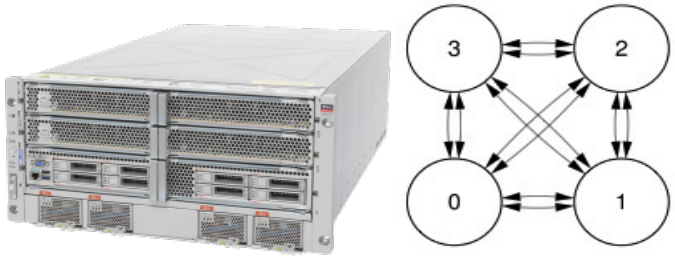


SPARC T8-2, T8-4, and M8-8 Servers: Systems Interconnect

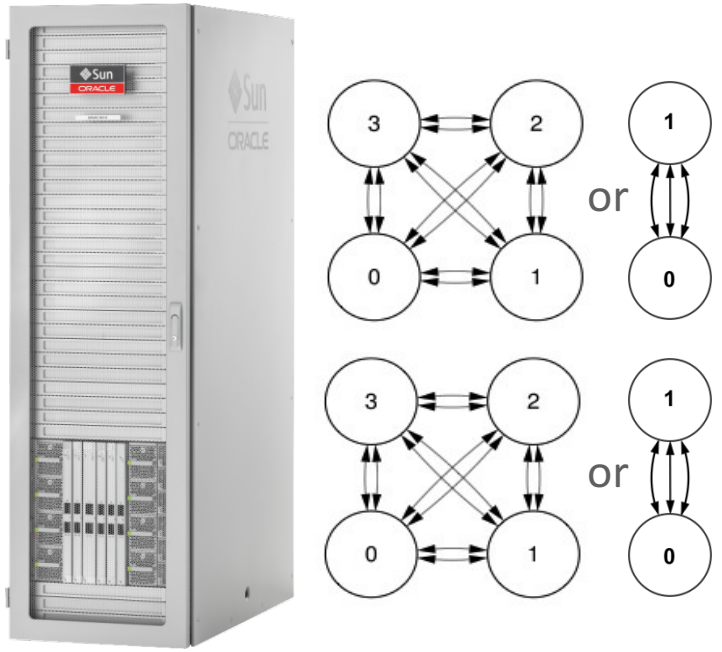
Glueless Connectivity with Coherency Links

SPARC T8-2

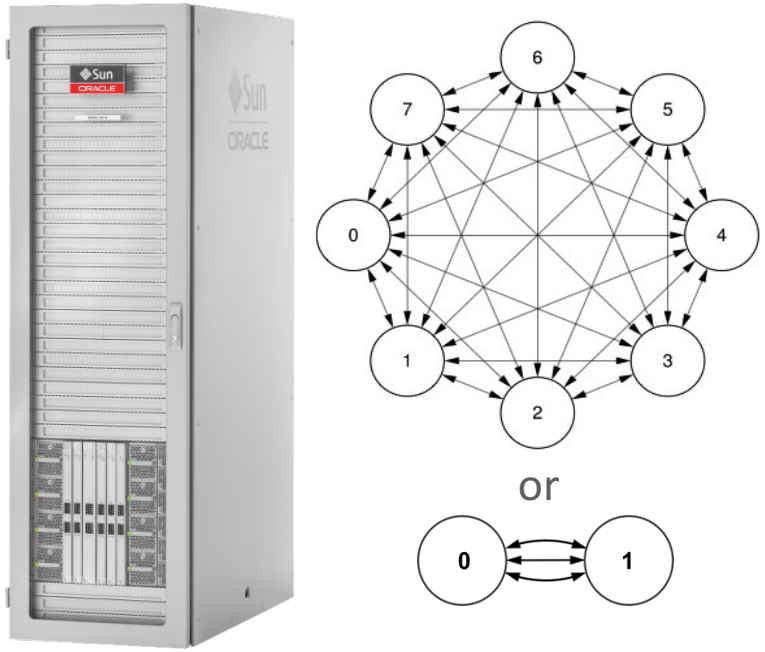
SPARC T8-4



SPARC M8-8 w/ 2 PDOMs



SPARC M8-8 w/ 1 PDOM



SPARC M8-8 Server

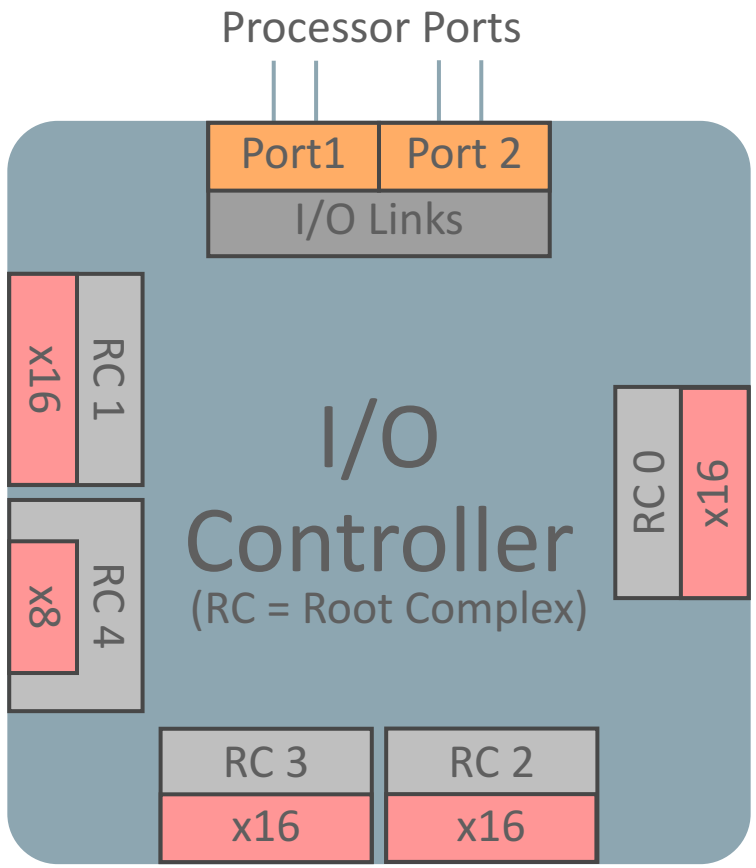
Plesiochronous System Clock

- Plesiochronous (pronounced plee-see-AH-krun-us)
  - Means “almost synchronous”
  - From Greek plesos (close), and chronos, (time)
- In general, plesiochronous systems behave similarly to synchronous systems
- Support for plesiochronous clocking makes a reliable system that can operate regardless of occasional “sync slips”
- Each CMIOU and switch board has dual, redundant clock synthesizers

SPARC M8: I/O Controller

I/O Controller ASIC (aka I/O Hub or PCIe Controller)

- Dual-host processor failover
- Changes in CPUs have no impact on PCIe paths
- 5 PCIe 3.0 root complexes
  - 4 PCIe 3.0 (x16) quadfurcatable ports (1 x16, 4 x4, or 2 x8)
  - 1 PCIe 3.0 (x8) bifurcatable port (1x 8 or 2 x4)
- 2 processor ports, 4 I/O links to the SPARC M8 processor
  - x8 lanes per link, 18.1 Gb/sec link rates per lane
  - 72.5 GB/sec I/O bandwidth, per direction
- Over 4x I/O bandwidth vs. SPARC T5 and M6 processors
- SR-IOV–compliant
- Address translation (HV/Oracle Solaris) per DMA stream
- Relaxed packet ordering per DMA stream



NVM Express (NVMe) Technology

- NVMe = Non-volatile memory host controller interface for PCI Express (PCIe) bus
- NVMe is a high-performance interface for solid-state drives (SSDs)
  - Eliminates the SAS host bus adapter (HBA) by connecting directly into PCIe bus
  - Improves both random and sequential performance
  - Reduces latency, increases parallelism, and provides streamlined command set
- NVMe uses PCIe signaling
  - 8 GB/sec, x4 interface per drive (four times the wire count vs. SAS)
- Same connector as the 2.5” SFF SAS HDDs/SSDs (more pin-out used)
- NVMe drives also exist in other form factors (for example, PCIe and M.2 cards)

NVMe Technology (cont.)

Offered with SPARC M8 processor–based servers

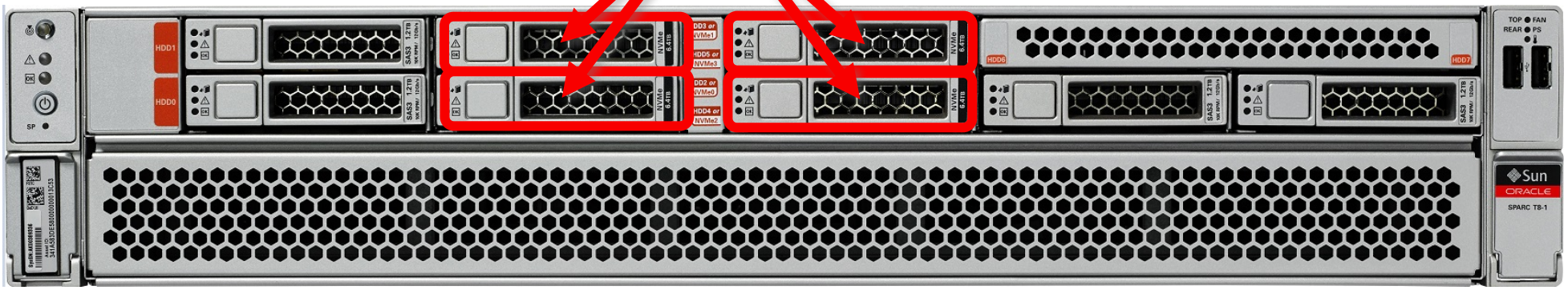
- Oracle Flash Accelerator F640 PCIe Card with 6.4 TB NVMe memory
  - Bootable device; supported in all SPARC M8 processor–based servers
- Internal 2.5” SFF NVMe drives in SPARC T8-1, T8-2, and T8-4 servers
  - SPARC T8-1: Up to 4 NVMe SFF drives (requires one PCIe switch card)
  - SPARC T8-2: Up to 4 NVMe SFF drives (requires one or two PCIe switch cards)
  - SPARC T8-4: Up to 8 NVMe SFF drives (requires two PCIe switch cards)
- PCIe switch card supports up to 4 NVMe drives
  - SPARC T8-1 and T8-2: Factory-configured option only
- Mixing of SAS and NVMe SFF drives is supported

Hot-Swappable, High-Bandwidth NVMe Flash

2.5-Inch SSD Drives in SPARC T8-1, T8-2, and T8-4 Servers

Uses same 2.5-inch SFF drive bays as the SAS HDDs and SSDs

For example: Four drive bays in SPARC T8-1 are designated as NVMe-capable



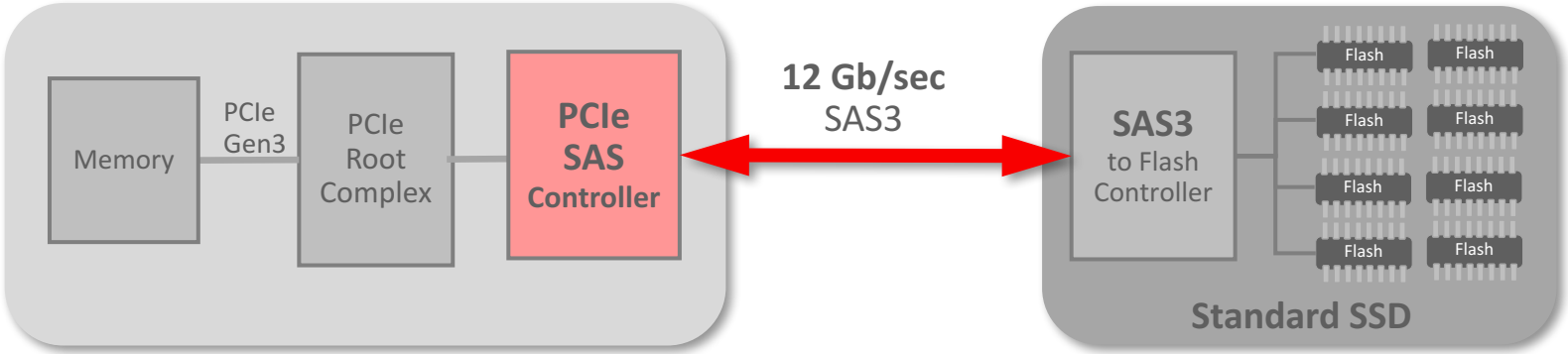
Unique Oracle design connects the disk backplane to an optional PCIe switch that bypasses the SAS controller

NVMe Bandwidth Breakthrough

Bandwidth to Each SSD Increased by 2.6x

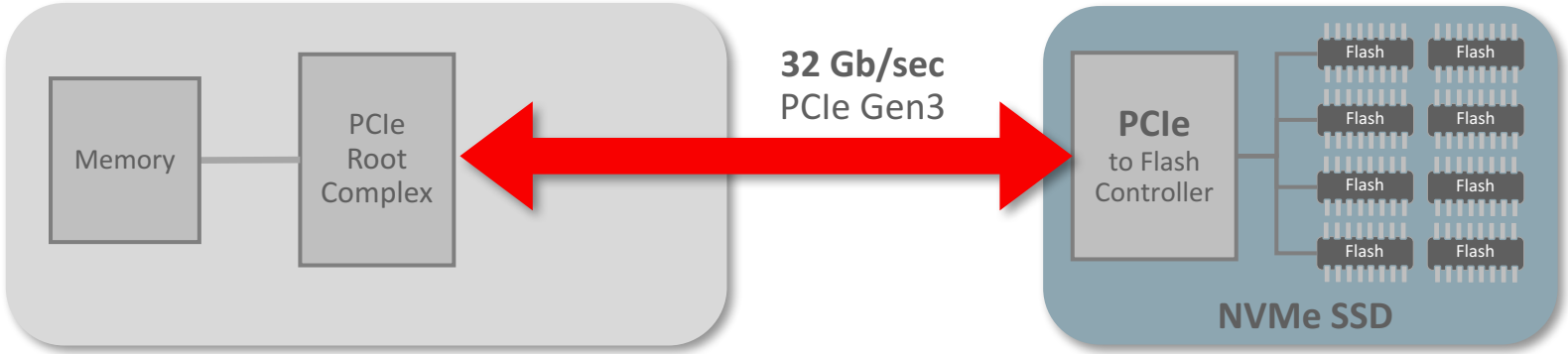
Conventional SSDs

- Flash data is first transformed to SAS protocol
- Bandwidth to each SSD limited by SAS3 bandwidth



NVMe SSDs

- Eliminate the protocol transformation to SAS
- Interface with root complex over a 4-lane PCIe Gen3 interface



Oracle Flash Accelerator F640 PCIe Card

- 6.4 TB NVMe flash memory
  - High-performance, low-latency storage
  - Internal boot device
  - Hot-pluggable (with the carrier)
- Occupies one PCIe LP slot (PCIe 3.0 x8)

Performance	Sun Flash Accelerator F80 PCIe Card	Oracle Flash Accelerator F640 PCIe Card
Interface	SAS 3.0	NVMe
8 k random read (IOPS)	155 k	356 k
Read/write throughput (GB/sec)	2.1 / 1.2	3.2 / 2.2
Write latency (µs)	84 (8k)	18 (4k)

SPARC M8-8 Server: Boot Options

Each PDom must have a boot source, which can be one of the following:

Boot Option	PCIe Adapter/Option	Connects to
Internal NVMe SSD	• Oracle Flash Accelerator F640 PCIe Card	N/A
Fibre Channel (SAN attached)	• Oracle Storage Dual-Port 16 Gb or 32 Gb Fibre Channel PCIe HBA • Sun Storage Dual 16 Gb FC HBA	• Oracle ZFS Storage ZS5-2 and ZS5-4 • Oacle ZFS Storage ZS3 and ZS4[1] • Oracle FS1 Flash Storage[1] • Sun Storage 6180, 6580, and 6780 Array[1] • Pillar Axiom 300, 500, and 600[1] • Sun ZFS Storage[1]
iSCSI (Ethernet)	• Oracle Quad Port 10GBase-T Adapter • Sun Quad Port GbE PCIe 2.0, UTP • Sun Dual 10 GbE SFP+ PCIe Adapter • Oracle Quad 10Gb or Dual 40Gb Adapter	• Oracle ZFS Storage ZS5-2 and ZS5-4 • Oracle ZFS Storage ZS3 and ZS4[1] • Pillar Axiom 600[1] • Sun ZFS Storage[1]
iSCSI (InfiniBand)[2]	• Oracle Dual Port QDR InfiniBand Adapter M3	• Oracle ZFS Storage ZS5-2 and ZS5-4 • Oracle ZFS Storage ZS3 and ZS4[1]

1) The product is EOL'd but supported.
2) Requires a specific boot process

Embedded USB (eUSB) and Boot over InfiniBand (IB)

Oracle Solaris Boot Process

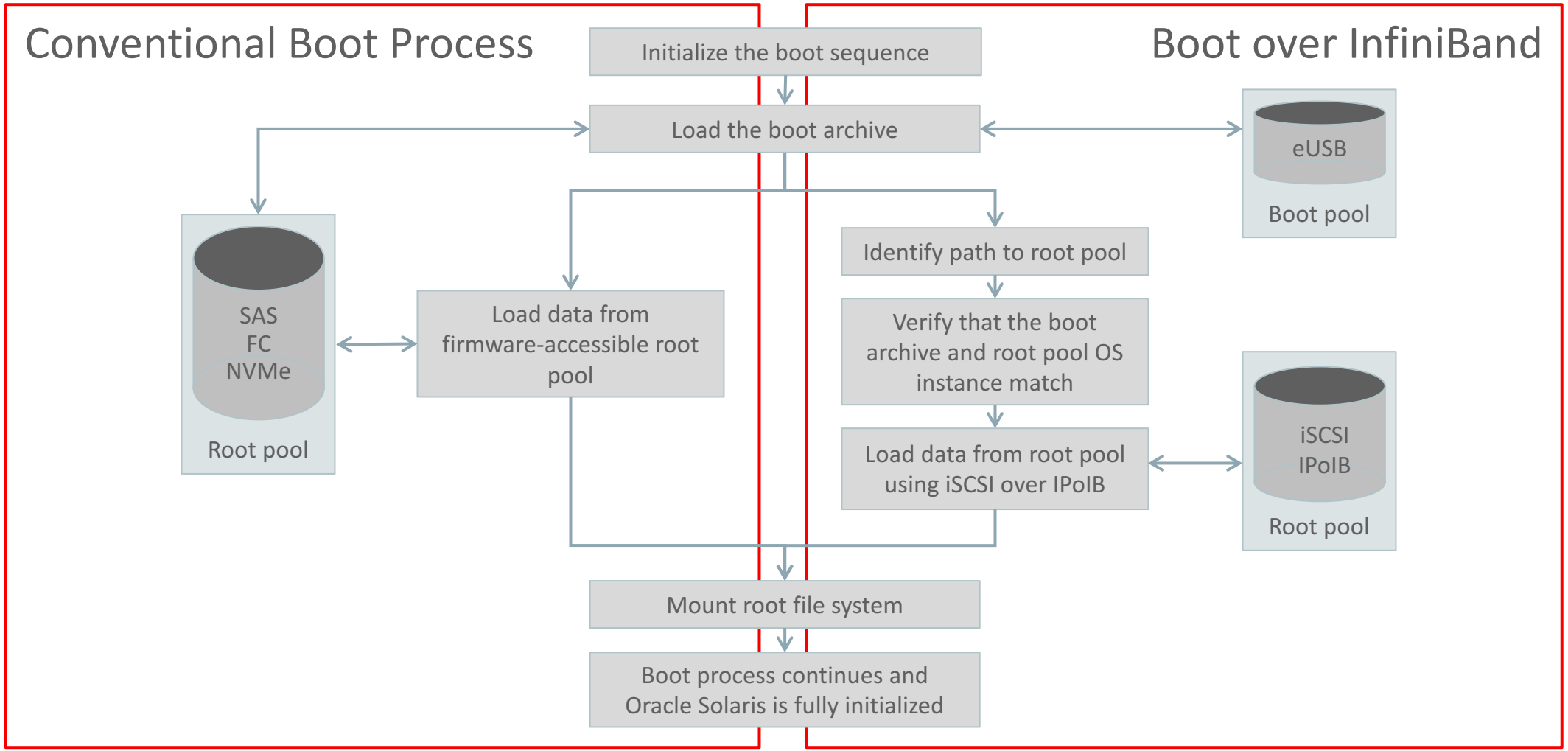
- Conventional network boot is not possible over IB
  - IB storage devices are not accessible by the OpenBoot PROM firmware
- Boot program and data loaded from local flash memory
  - New servers include internal, embedded USB (eUSB) storage device
  - One or more eUSB devices form the boot pool
  - Boot pool stores the firmware-accessible boot archives

- Boot archive allows root file system mount using iSCSI over IPoIB
- Fall-back: A boot archive exists on the system service processor

Oracle Solaris Boot Process Terminology

Term	Definition/Description
Boot environment (BE)	A bootable instance of the Oracle Solaris operation system plus any other application software packages installed into that instance.
Boot archive	A subset of a root file system. It contains the kernel modules and configuration files. The files in the boot archive are read by the kernel before the root file system is mounted. After the root file system is mounted, the boot archive is discarded from memory by the kernel.
Boot pool	A distinct pool that is used to store boot archives. This pool also includes boot loader data files, as well as recovery data. Each data set in the boot pool is linked to a boot environment. The boot pool will exist only when an installation is done off a root pool that is not accessible by the system firmware.
Root pool	The device or pool of devices containing the root file system.
InfiniBand (IB)	A computer-networking communications standard featuring high throughput and low latency.
IPoIB	Internet Protocol over InfiniBand is a protocol stack that enables TCP/IP over an IB network.
SCSI	The Small Computer System Interface is a set of parallel interface standards for attaching disk drives, printers, scanners, and other peripherals to computers.
iSCSI	Internet SCSI; works on top of the Transport Control Protocol (TCP) and allows the SCSI command to be sent end-to-end over local-area networks (LANs), wide-area networks (WANs), or the internet.

Oracle Solaris Boot Process



Agenda

1. Overview
2. System Details
3. New Technologies
4. RAS
5. Platform Management
6. Virtualization
7. Summary

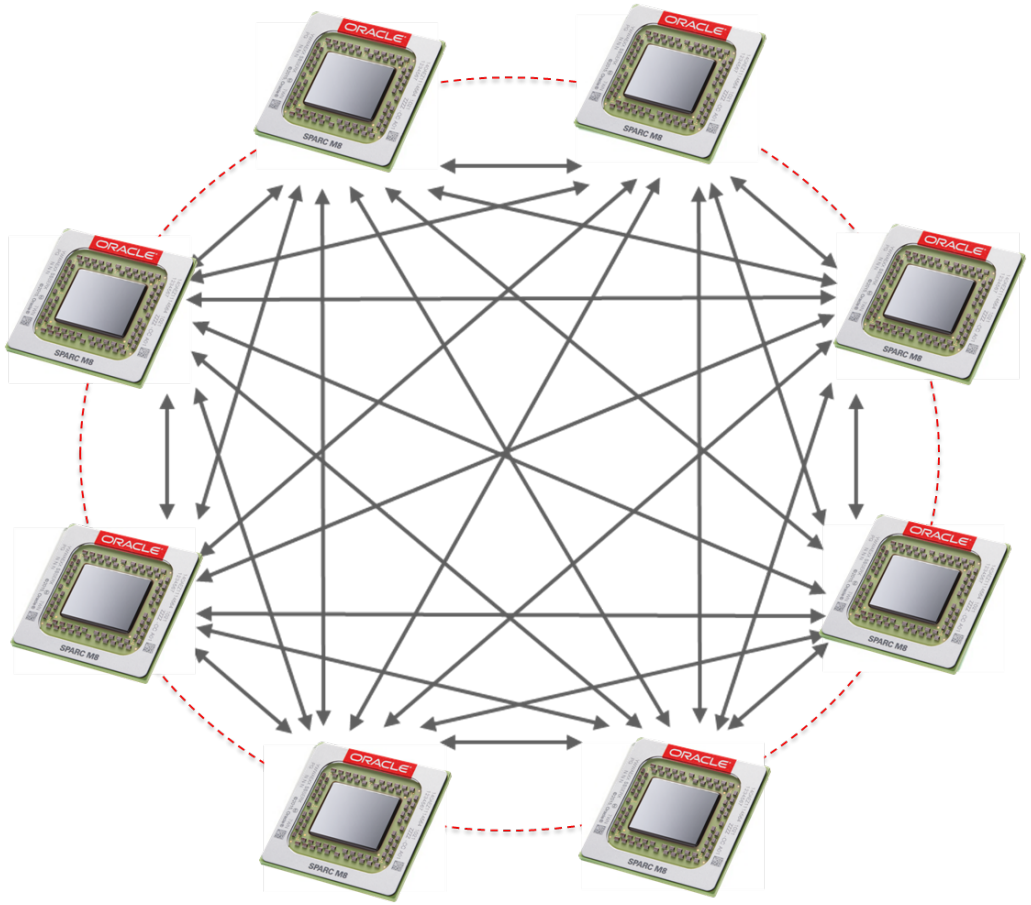
Reliability, Availability, and Serviceability

SPARC M8 Processor–Based Servers

- Designed to minimize part count and operating temperature to enhance reliability
- Fault Management Architecture (FMA) support in Oracle Solaris and Oracle ILOM
- Processor and memory protection
  - CPU core and thread off-lining
  - SDRAM error protection, DIMM sparing, page retirement, and memory scrubbing
- End-to-end data protection detecting and correcting errors
  - Includes cyclic redundancy check (CRC)
  - In hardware: Message retry and lane retire
- Major components are redundant and hot-serviceable
  - Fans, power supplies with dual grid support
  - SPARC T8-1, T8-2, and T8-4 servers: Internal disks with RAID capability

SPARC M8 Processor–Based Servers: SMP Scalability

- Up to 8 processors directly connected (glueless)
- Dynamic congestion avoidance for data
- Coherence links
- Link-level reliability, availability, and serviceability (RAS) in hardware
  - CRC check and automatic message retry
  - Automatic lane retire (per direction)
  - Automatic link retrain and reinitialization
  - Built in PRBS testing during link training



SPARC M8 Processor–Based System RAS

System and I/O

- Redundant hot serviceable SP, with automatic failover*
- Diagnosis to the FRU level on first fault
- Independent clock sources*
- Redundant clocking*
- Redundant DC/DC power*
- Intelligent fan control
- Redundant hot-swappable PSUs and fans
- Dual grid power
- Up to 5 PCIe buses per processor*
- PCIe end-to-end CRC and PCIe link retry
- Hot-pluggable PCIe cards*

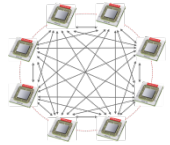
Fault Management Architecture

- Diagnosis engine on SP and Oracle Solaris
- Auto-reconfigure on failure
- Soft error rate discrimination (SERD)
- Bad page retirement
- OS and SP watchdogs

M8 Processor

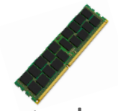
- L1$ tag, status, and data
  - Parity protection, retry on error
- L2$/L3$ data
  - SEC/DED protection, inline correction
  - Cache-line sparing
- L2$/L3$ status and directory
  - SEC/DED protection, inline correction
- Architectural registers
  - SEC/DED protection
  - Precise trap and hypervisor correction/retry

Systems Interconnect



- Cyclic redundancy check (CRC)
- Message retry in hardware
- Lane retire in hardware
- PDom isolation*
- Easy-to-service cable trusses*
- Independent and distributed clock sources*

Memory



- SDRAM error protection
  - Error correction within a single SDRAM device
  - Triple-bit error detection across SDRAM devices
- DIMM sparing
- Memory channel interconnect
  - Message retry in hardware
  - Lane retire in hardware
  - Cyclic redundancy check (CRC)

Hypervisor

ORACLE VM for SPARC

- Dynamic assignment of CPU, memory, I/O
- Dynamic PCIe bus assignment
- Logical domains virtualization and failure containment
- Processor support for error clearing, correction, and collection

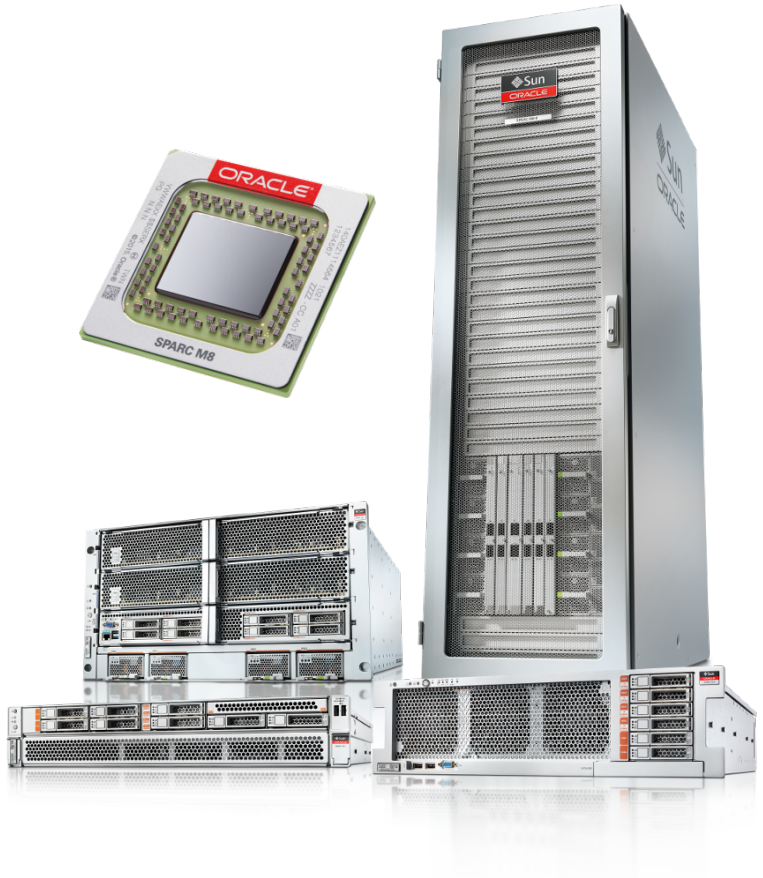
* Varies by model

Definition of Terms

- Hot-serviceable
  - Components can be removed and replaced while the server is running.
  - Hot-swappable components do not require any preparation prior to servicing. Typical examples: power supply, fan module, or a single RAID disk drive
  - Hot-pluggable components do require preparation prior to servicing (for example, invocation of a CLI command or actuating a hot-service button on the component to be removed). The system will notify the user when it is safe to remove the component. Typical examples: a PCIe card in a hot-pluggable carrier or non-RAID disk drive.
- Cold-serviceable
  - Components require that the server be shut down. In addition, some service procedures require that the power cables be disconnected between the power supplies and the power source.

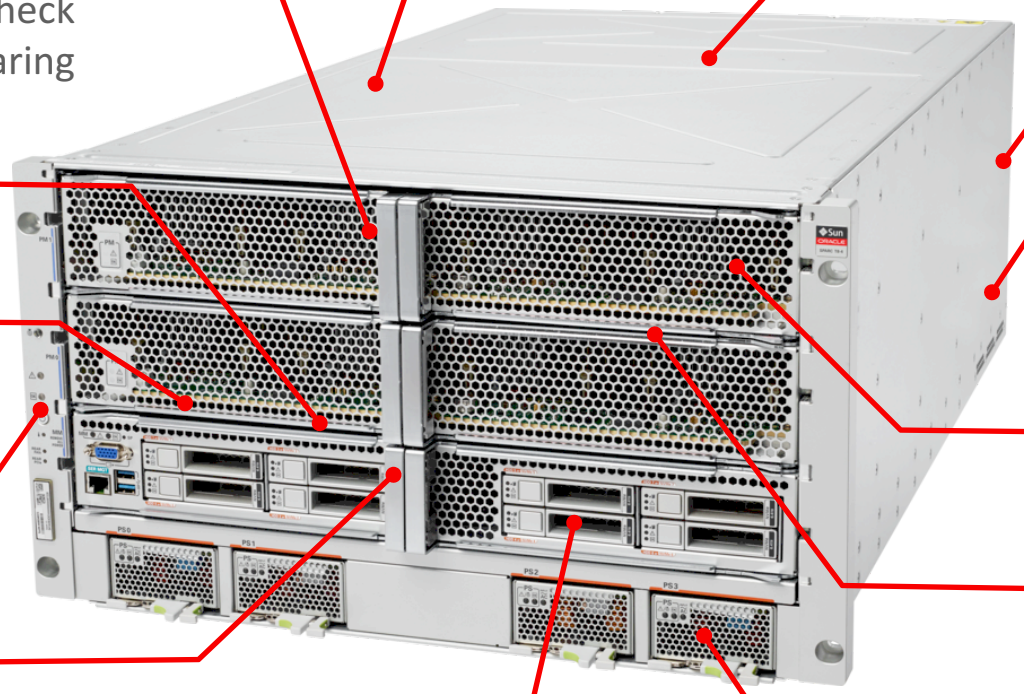
Common RAS Features

SPARC M8 Processor–Based Servers



Item	Feature
System	Fault Management Architecture incl. Predictive Self Healing in Oracle Solaris and Oracle ILOM
	Redundant hot-swappable cooling fans and power supplies with dual grids
	Multiple root complexes and multipath storage (MPxIO)/networking (IPMP)
	Power-on tests and Oracle Auto Service Request
Software	Oracle Solaris Zones, Oracle VM Server for SPARC, ZFS file system
	Live Oracle Solaris operating system upgrades
	Firmware updates during system operation
	Oracle Solaris Cluster, Oracle Real Application Clusters (Oracle RAC)
Processor	Instruction retry, core isolation/deconfiguration
	Cache parity protection
Memory	SDRAM error protection and DIMM Sparing
	Memory channel interconnect: Message retry, lane retire, CRC protection
	Soft error rate discrimination (SERD) and bad-page retirement

SPARC T8-1, T8-2, and T8-4 Servers: Availability Features



SDRAM error protection, message retry in hardware, lane retire in hardware, cyclic redundancy check (CRC), and DIMM sparing

Power-On Self-Test

Redundant, hot-swappable cooling fans

Hot-pluggable I/O adapters (SPARC T8-4)

Hardware RAID 0, 1, 10, and 1E integrated in disk controller

Redundant I/O controllers with MPxIO for storage devices and IPMP for network devices

ZFS with software RAID 5, 6, and 10 for file system resilience

Multiple PCIe root complexes for isolated I/O virtualization

Front and rear system indicator LEDs

Independent power domains on processor modules (SPARC T8-4)

Dual integrated SAS HBAs or NVMe PCIe switches for drive isolation (SPARC T8-2 and T8-4)

Redundant, hot-swappable power supplies

Hot-pluggable drives

SPARC M8-8 Server: Availability Features

Isolated PDoms

SDRAM error protection, message retry in hardware, lane retire in hardware, cyclic redundancy check (CRC), and DIMM sparing

Redundant, hot-serviceable SPs with automatic failover

Distributed, independent and redundant clocking

CRC protected systems interconnect with message retry and lane sparing in hardware

Hot-pluggable I/O adapters

Redundant I/O controllers with MPxIO for storage devices and IPMP for network devices

Dedicated PCIe buses for every slot for isolated I/O virtualization

ZFS with software RAID 5, 6, and 10 for file system resilience

Redundant, hot-swappable cooling fans

Redundant DC power on CMIOU boards and CPU module

Redundant, hot-swappable power supplies, dual grid power input

Predictive Self Healing in Oracle SPARC Servers

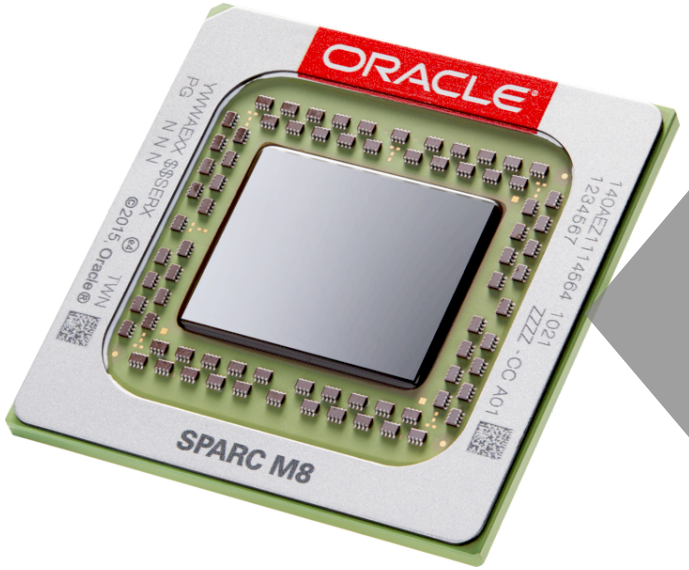
Hardware faults that otherwise might cause a system restart are isolated to the affected services. Services are automatically restarted after hardware and software faults.

- Fault Manager Architecture (FMA)
  - Automated diagnosis and isolation of hardware faults
  - Structured logs and tools for telemetry data
  - Live diagnosis updates without system reboots
  - Standardized fault messaging
- Service Manager Facility (SMF)
  - Integrated, automatic restart of failed software services
  - Automated, guided troubleshooting for failed services
  - Faster boot, improved disaster recovery, security

SPARC M8 Processor: RAS Features

- Instruction retry
- Processor bus protection for address/control, caches, and data networks
- Dynamic processor core deallocation
- Node-to-node bus dynamic repair/lane retire
- ECC for architectural states, that is, caches and data networks
- CRC for serial links; can detect multiple-bit errors

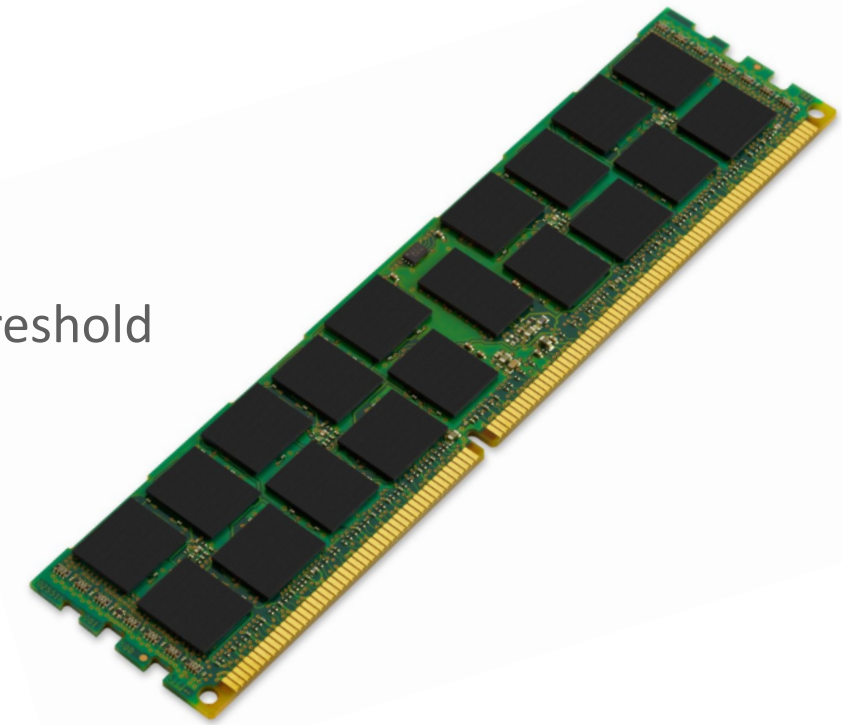
SPARC M8 Processor: Dynamic Voltage and Frequency Scaling (DVFS)



- Continued server operation even during unexpected increased ambient temperature
- If needed, processor throttles processing via the clock signal
- Slower clock speed reduces power consumption and heat dissipation
- DVFS is also used for efficient power management

SPARC M8 Processor–Based Servers: Memory

- SDRAM error protection
  - Error correction (up to 32 bits) within a single SDRAM device
  - Triple-bit error detection across SDRAM devices
- DIMM sparing
  - Retire a DIMM w/o loss of capacity or error protection
  - Dynamic, no loss of capacity, and error protection intact
- Memory page retirement upon exceeding soft error rate threshold
- Memory channel interconnect
  - Cyclic redundancy check (CRC)
  - Message retry in hardware
  - Lane retire in hardware
- Memory “scrubber”
  - Periodical event to prevent multiple soft errors in one line



DIMM Sparing

Increased Availability and Reliability

- Automatic and dynamic retirement of a failing memory DIMM
- Protection against potential future hard failures
- Error protection and memory capacity remain intact
- How it works
  - FMA software detects high error rate, decides to not use one potentially failing DIMM anymore
  - Content from failing DIMM is migrated to the other 15 DIMMs
  - DIMM sparing happens dynamically in live system; no noticeable impact on performance
  - Software managed migration w/ hardware assist; region in migration kept live and coherent

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	DIMM Number
94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	94%	% Used

DIMM #7 retired

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	DIMM Number
100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	% Used

Distributed, Independent and Redundant Clocking

SPARC M8-8 Server

- Independent clock sources on every CMIOU and switch board
- Dual clock sources on every board
  - A board remains operational after one of the clock sources fails
- Plesiochronous system that deals better with "sync slips"
  - Sync slips are a natural characteristic of the system
  - No need to pause and synchronize the system clock
  - Reduced risk of data corruption due to timing errors

Power-On Self-Test (POST)

- Tests processors, caches, memory, and for the presence of adapters
- Coherency, scalability, and I/O links
- PCI topology links
- Message to system processor (SP) for any components not passing diagnostics
- Tests devices on service processor and its Ethernet port
- Component failure during operation
  - The boot process will automatically deconfigure the affected component(s) upon the next boot

A Sample of Availability Features within Oracle Solaris

- Enhanced drivers verify data integrity
  - Support for PCIe card hot-pluggability and dynamic reconfiguration
  - Enhanced SR-IOV support with I/O resiliency in Oracle VM Server for SPARC
- Robust file system: ZFS
  - Metadata not stored with the data
  - Advanced RAID levels independent of controllers
- Predictive Self Healing
  - Fault Manager Architecture (FMA)
  - Service Management Facility (SMF)
- Layered virtualization

Availability Features Within ZFS

- Transactional file system
  - Never in an inconsistent state through journaling
- Shadow migration
  - Migrate data from an old to a new file system while allowing access and modification of the new file system
- File snapshots freeze data at a point in time
  - Initially consumes no additional space
- RAID-Z, variations on RAID 5 or RAID 6 and beyond
- Ability to split a mirrored pool
- Checksums and self-healing data
  - Scrubbing, similar to memory

RAS Features with Oracle VM Server for SPARC

- Independent and isolated virtual machines, operating system images
- Dynamic resource management
  - Add/remove resources as needed, without interruption to LDoms
  - Dynamic PCIe bus assignment
- Options for I/O resources
  - LDom owns the root complex and all of its devices
  - Single Root I/O Virtualization (SR-IOV)
  - Virtualized I/O for sharing and failover
- Live migration
- Physical-to-virtual conversion for earlier Oracle Solaris versions

Oracle Solaris Cluster

- Built for mission-critical cloud environments
- Fast failure detection via kernel integration
- Faster, automatic failover for business applications
  - Safe, automated recovery from site failure for data center/campus environment or remote/contingency center
- Pretested, out-of-the-box support for Oracle Database, Oracle E-Business Suite, Oracle WebLogic Server, MySQL, Oracle's PeopleSoft applications, SAP, and more
- Easy migration from single zone or server to a multinode cluster

Agenda

1. Overview
2. System Details
3. New Technologies
4. RAS
5. Platform Management
6. Virtualization
7. Summary

Oracle ILOM

Oracle Integrated Lights Out Manager

Oracle ILOM on SPARC M8 Processor–Based Servers' SP

- Manages, monitors, and responds to faults
  - Interacts with hypervisor and OS
  - Manages IP/MAC addresses
  - Monitors FRUs and sensors for power/temperature control
  - Provides remote power-on/power-off
  - Facilitates firmware updates

- New Oracle ILOM V4.0.1 software available on all servers
  - Black-box recorder to monitor Oracle ILOM processes and resource usage
  - Oracle ILOM is part of system FMA: analyzes errors and manages recovery
- Oracle ILOM integrates with third-party management tools from HP, IBM/Tivoli, Microsoft, and CA

Oracle ILOM on SPARC M8 Processor–Based Servers

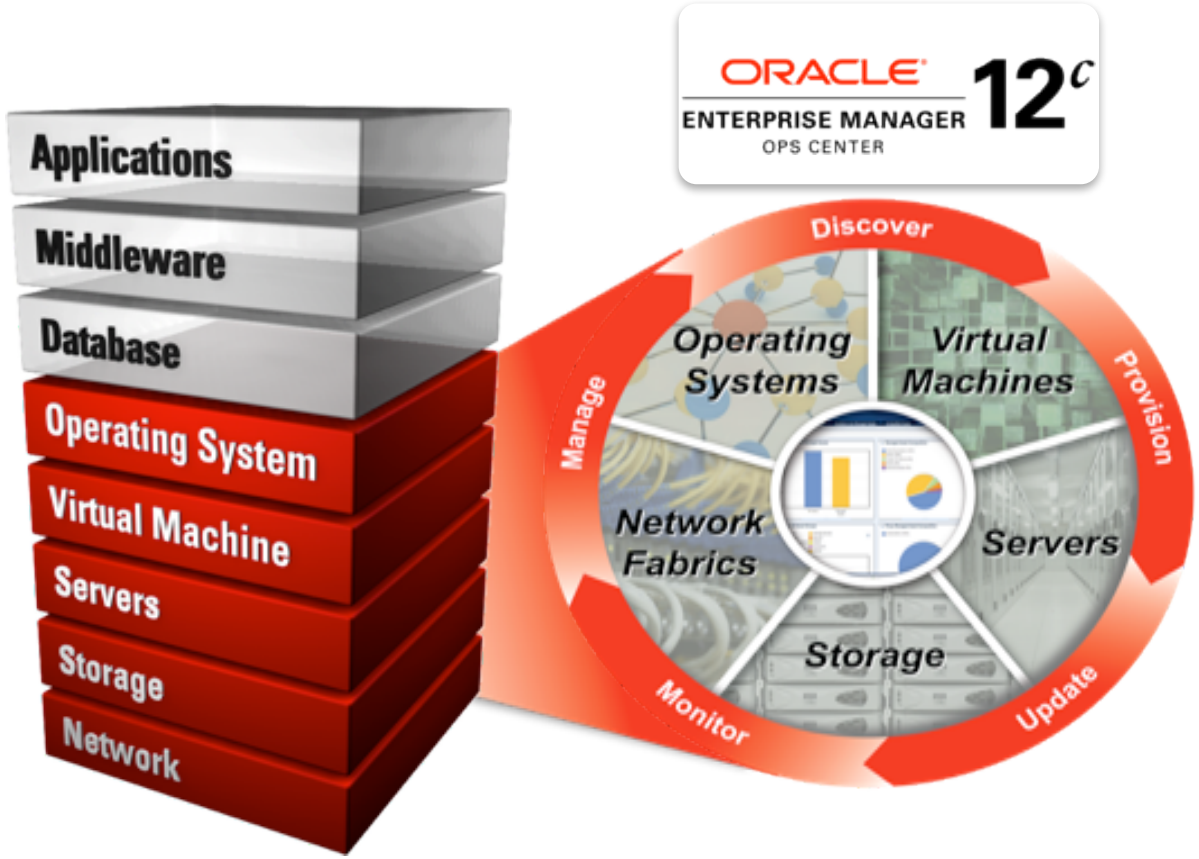
- SPARC M8 processor–based servers use Oracle ILOM Version 4.0.1 or later
- Oracle ILOM looks/behaves like Oracle ILOM on other platforms
  - Management interfaces: CLI, BUI, IPMI, SNMP
  - Remote host management
  - Inventory and component management
  - System monitoring and alert/fault management
  - User account management
  - Power consumption management
- Supports the SP/SPP/SPM structure in SPARC M8-8, M7-8 and M7-16 servers

Oracle Enterprise Manager Ops Center

Hardware and VM Management

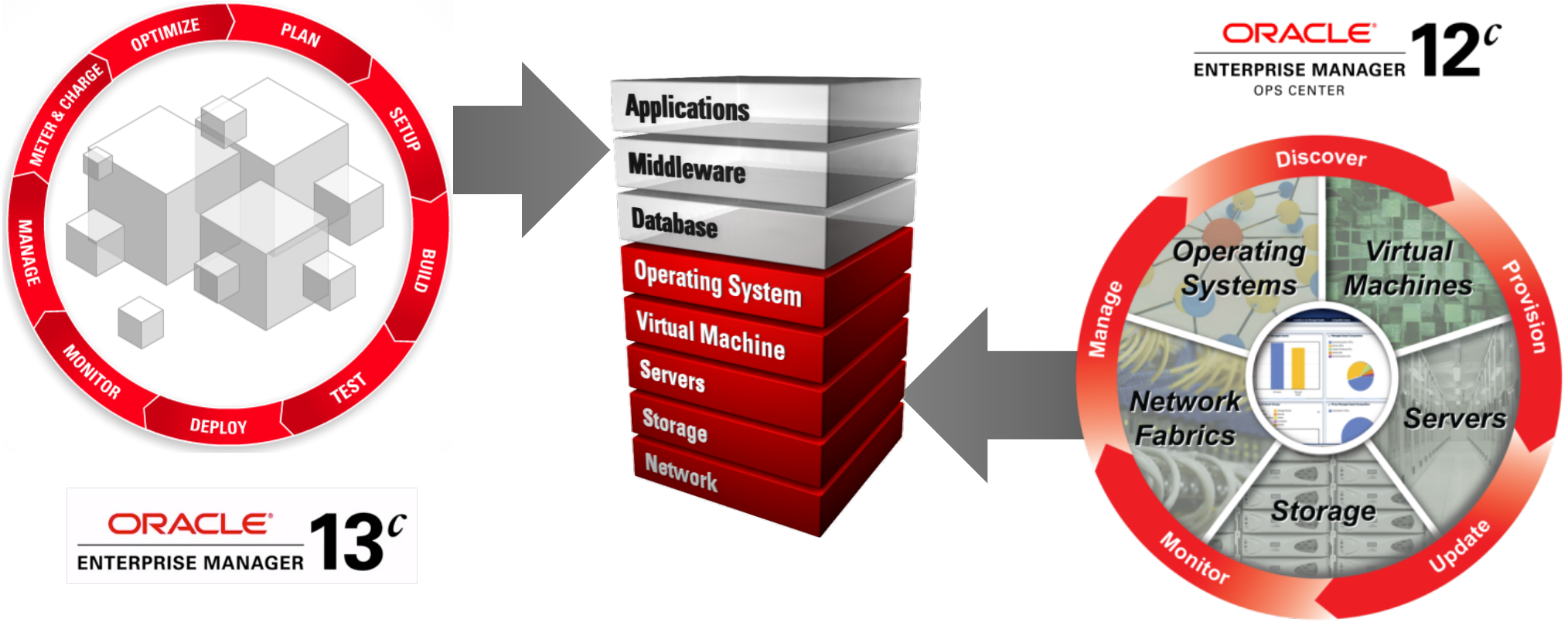
Oracle Enterprise Manager Ops Center 12*c*

Industry's First Converged Hardware Management Solution

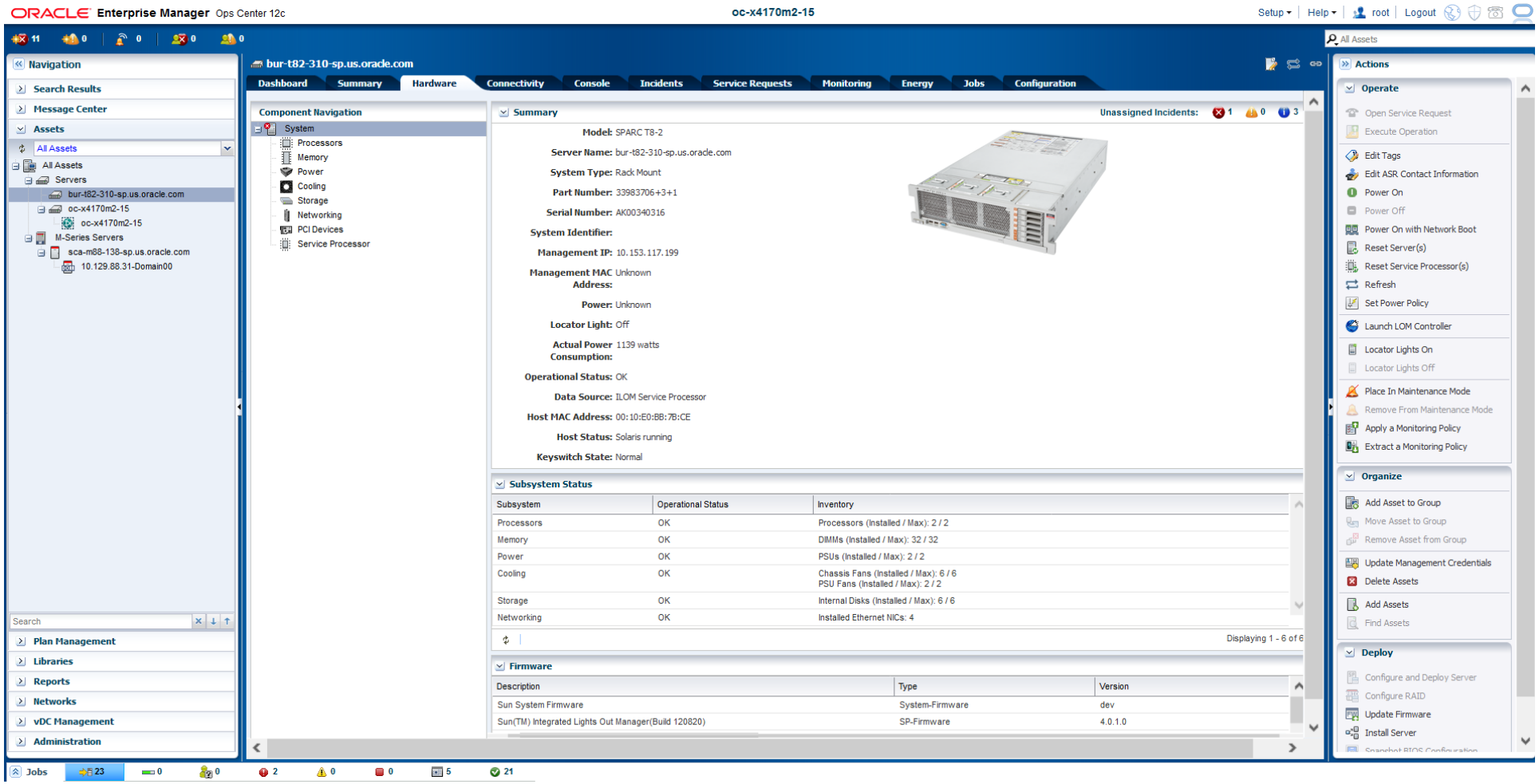


Integrated Infrastructure Management
+
Integrated Applications-to-Disk Management
+
Integrated Lifecycle Management
+
Integrated Systems Management and Support

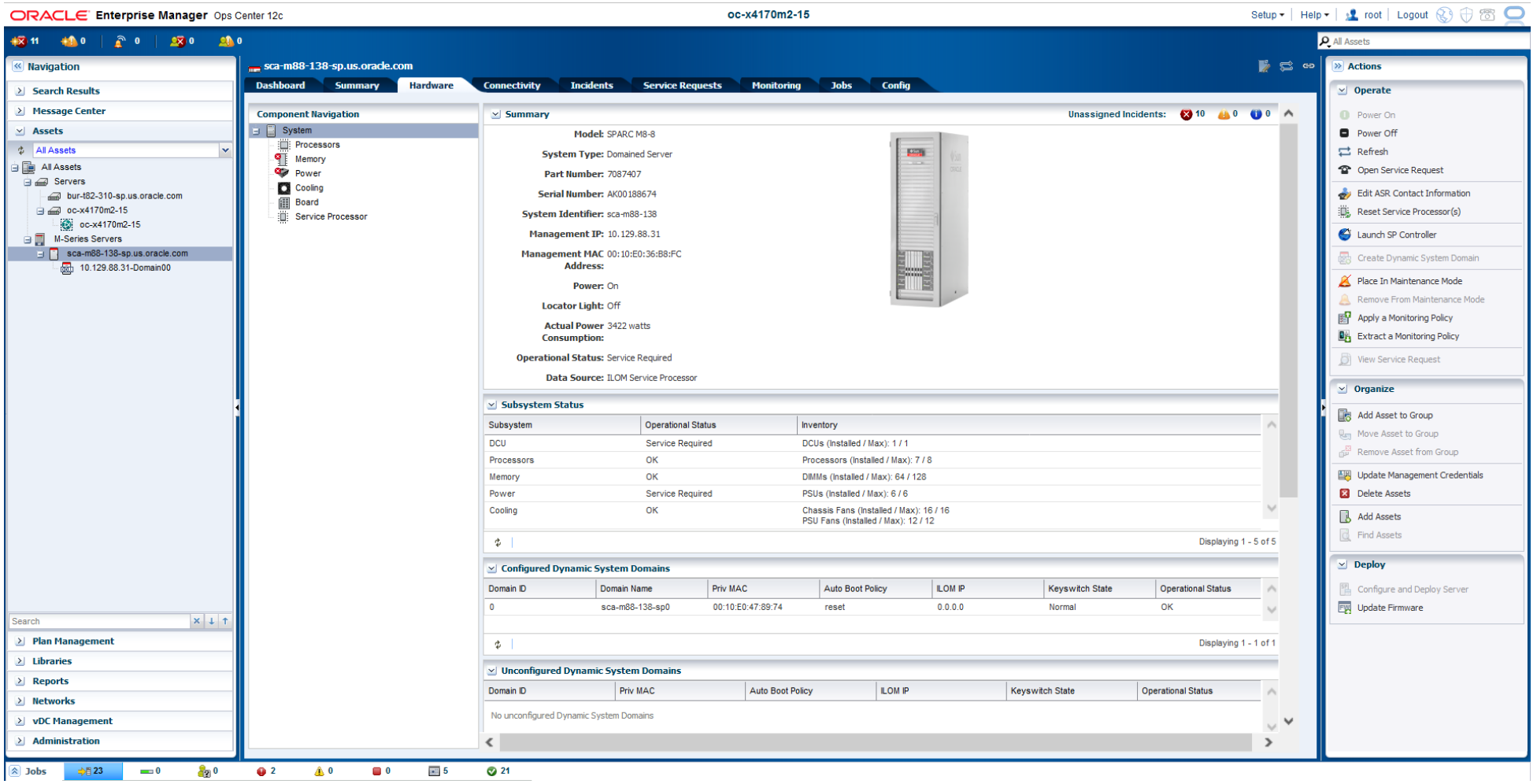
Management of Applications to OS to Virtualization to Disk



Oracle Enterprise Manager Ops Center: Screen Shot



Oracle Enterprise Manager Ops Center: Screen Shot (cont.)



Power Management

Power Management Interfaces

Rich Choice of Management Options

ORACLE ENTERPRISE MANAGER 12c
OPS CENTER

Oracle Solaris 11.3

- `poweradm`
- `pwconfig` and `/etc/power.conf`

Oracle ILOM

- `ssh` CLI
- HTTP

Power Management Features

SPARC M8 Processor–Based Servers

Feature	Comments
Dynamic Voltage and Frequency Scaling (DVFS)	SPARC M8 processor adjusts real-time voltage and/or frequency within each quadrant based on software-defined policies.
Cycle Skipping	SPARC M8 processor can be set to not work during a cycle in order to consume less power.
Power Supplies	SPARC T8-1: (A258): Platinum SPARC T8-2: (A263): Platinum SPARC T8-4: (A261): Platinum SPARC M8-8: (A265): Titanium
Intelligent Fan Control	Fans are automatically adjusted to increase air flow if temperatures exceed thresholds.

Power Management Policies

SPARC M8 Processor–Based Servers

- Disabled
  - All components run at full speed (legacy performance policy)
- ILOM power management policy: Performance (default)
  - Unallocated (unused) components are power-managed
  - Power savings features with insignificant performance impact are enabled
- ILOM power management policy: Elastic
  - Unallocated and allocated but idle components are power-managed
- Oracle ILOM provides the interface to set and manage power capping

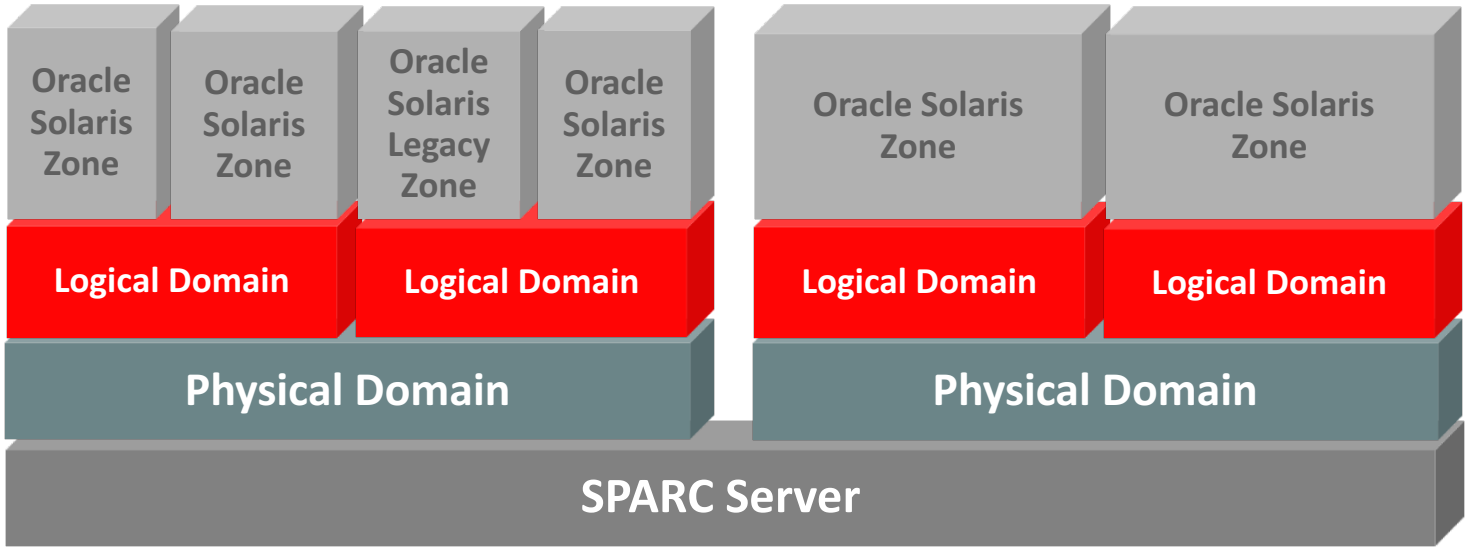
Agenda

1. Overview
2. System Details
3. New Technologies
4. RAS
5. Platform Management
6. Virtualization
7. Summary

Most-Extensive Virtualization Infrastructure

Layered Virtualization

- Zero-performance overhead
- Maximizes utilization, optimizes for availability
- Built-in, no-cost virtualization



Layered Virtualization

OS Virtualization
Oracle Solaris **Zones**

- Thousands of lightweight VMs
- Dynamic, zero overhead
- Oracle Solaris 11, 10, 9, and 8 zones
- Zone clusters
- Finest granularity

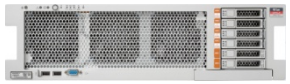
Hardware Virtualization
Oracle VM Server for SPARC
Logical Domains

- Up to 128 domains per PDom
- Dynamic, very low overhead
- Isolation of separate OSs
- Domain clusters
- Live migration
- Support of rolling upgrades

Logical Domains and Oracle Solaris Zones

Hardware and OS virtualization in a single PDom

SPARC T8-1



SPARC T8-2

SPARC T8-4

SPARC M8-8

ORACLE DATABASE 12c

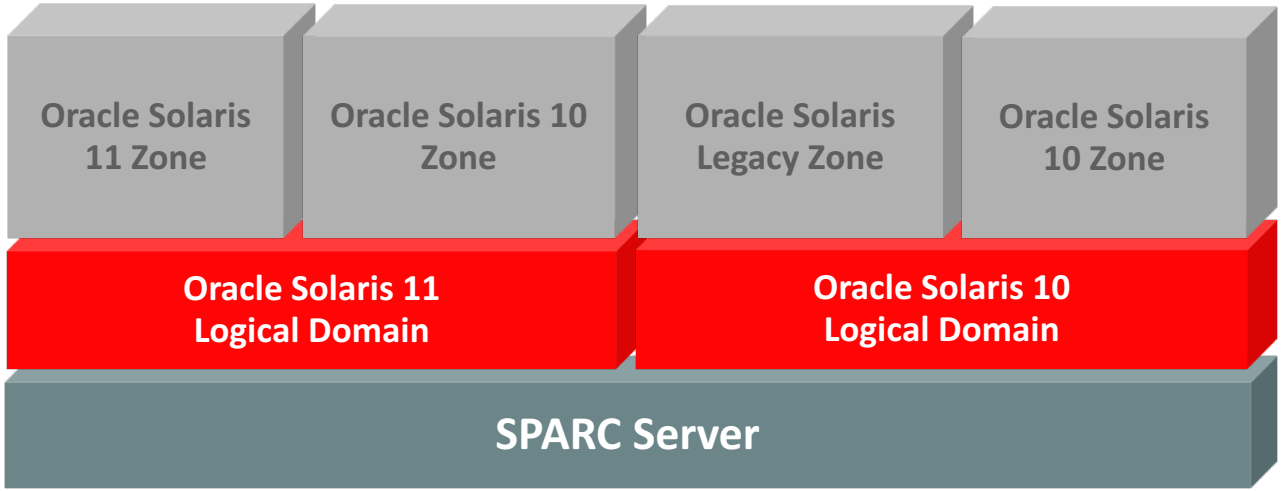
ORACLE FUSION MIDDLEWARE 11g

ORACLE E-BUSINESS SUITE

ORACLE ENTERPRISE MANAGER 13c

ORACLE PEOPLESOFT ENTERPRISE

ORACLE SIEBEL



Hardware Isolation with Physical Domains

SPARC M8-8 servers with dual PDoms

ORACLE DATABASE 12c

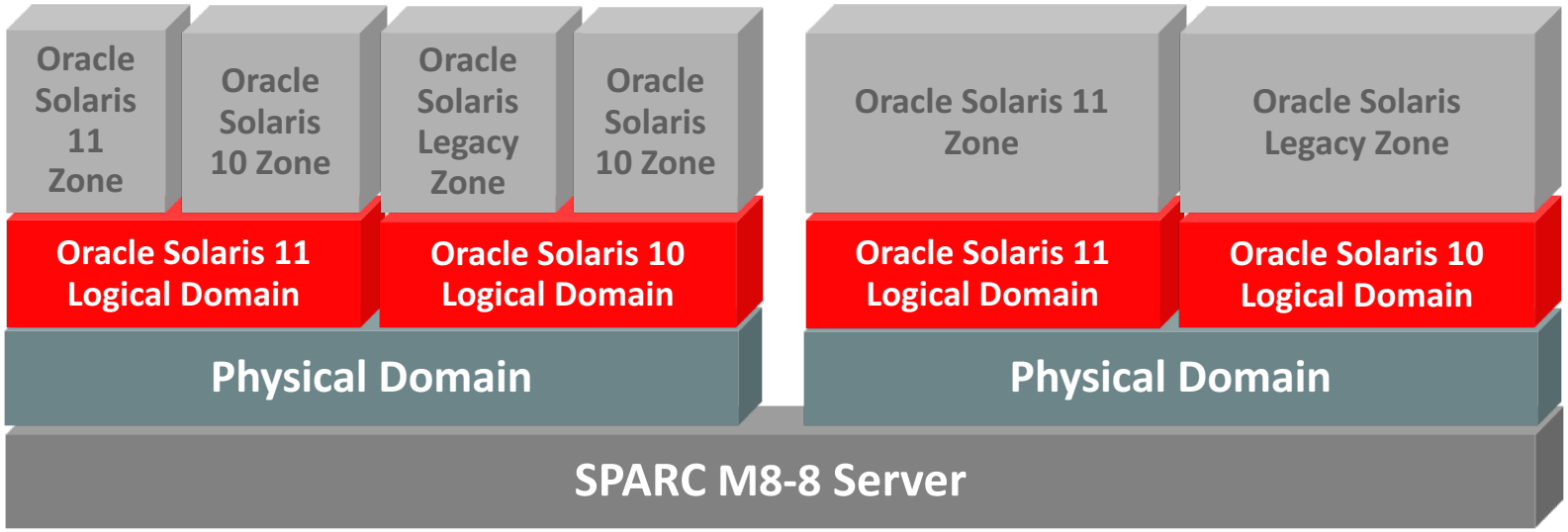
ORACLE FUSION MIDDLEWARE 11g

ORACLE E-BUSINESS SUITE

ORACLE ENTERPRISE MANAGER 13c

ORACLE PEOPLESOFT ENTERPRISE

ORACLE SIEBEL

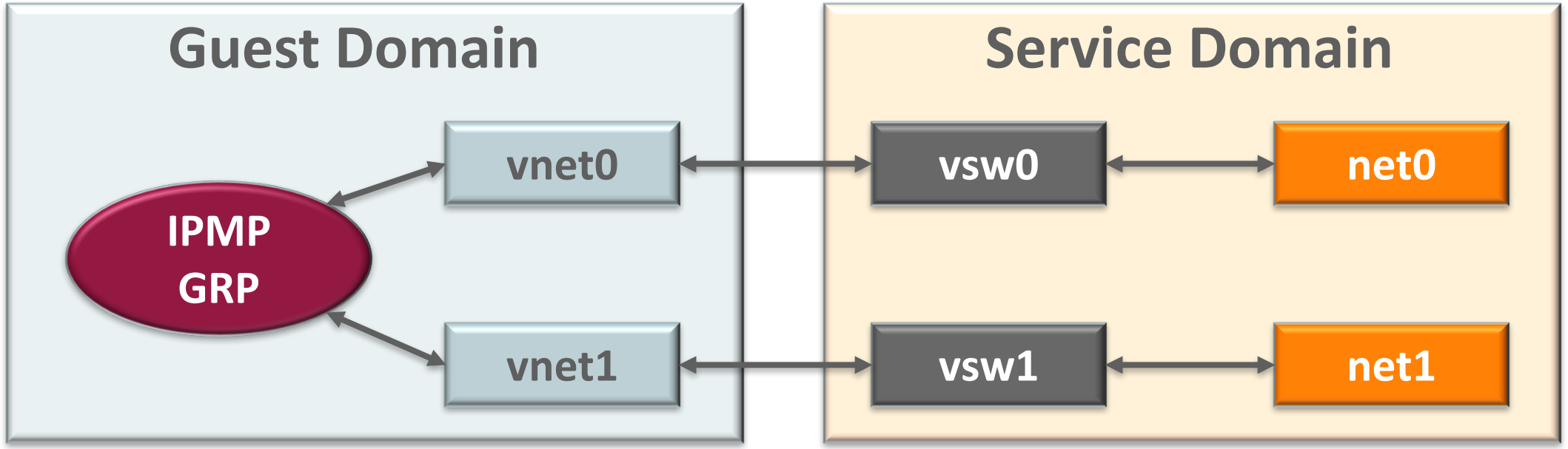


Characteristics of Logical Domains

- Guest domains—for applications; consume virtual device services provided by one or more service domains
- Service domains—provide virtual network and disk devices to guest domains
- I/O domains—have direct access to a physical I/O device or own a PCIe SR-IOV virtual function
- Root domain—I/O domain that owns one or more whole root complexes
- Control domain (aka primary)—management control point for configuring domains and managing resources; one per PDom

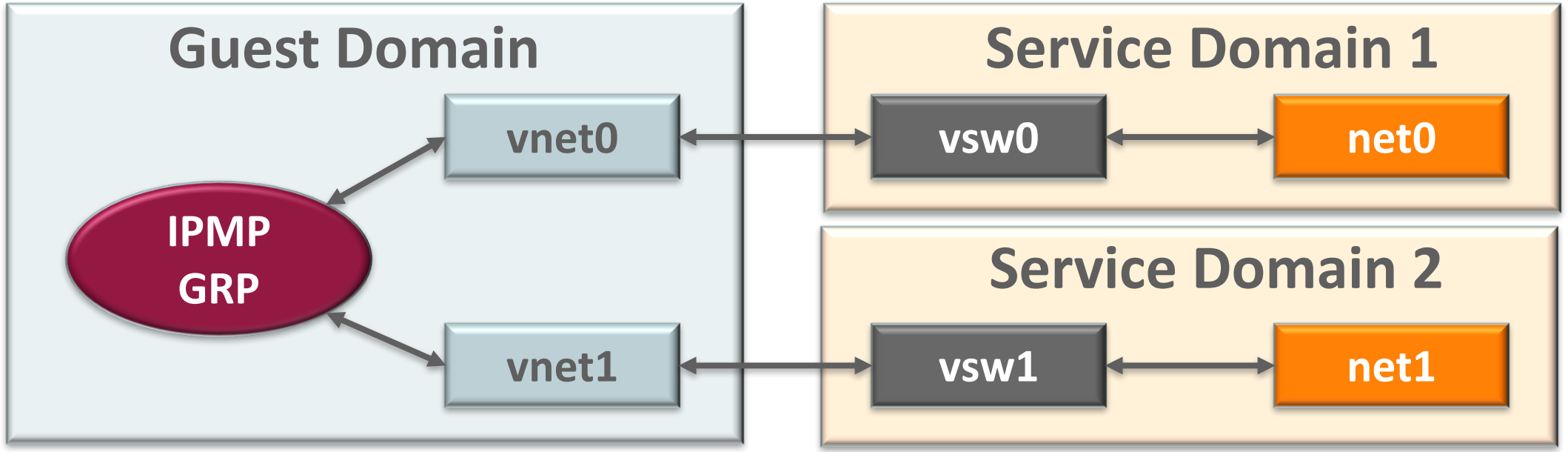
Redundant Virtual I/O Services

- Configure redundant connections for best availability
- Each go to the same network or storage
- Apply MPxIO for storage or IPMP for networks



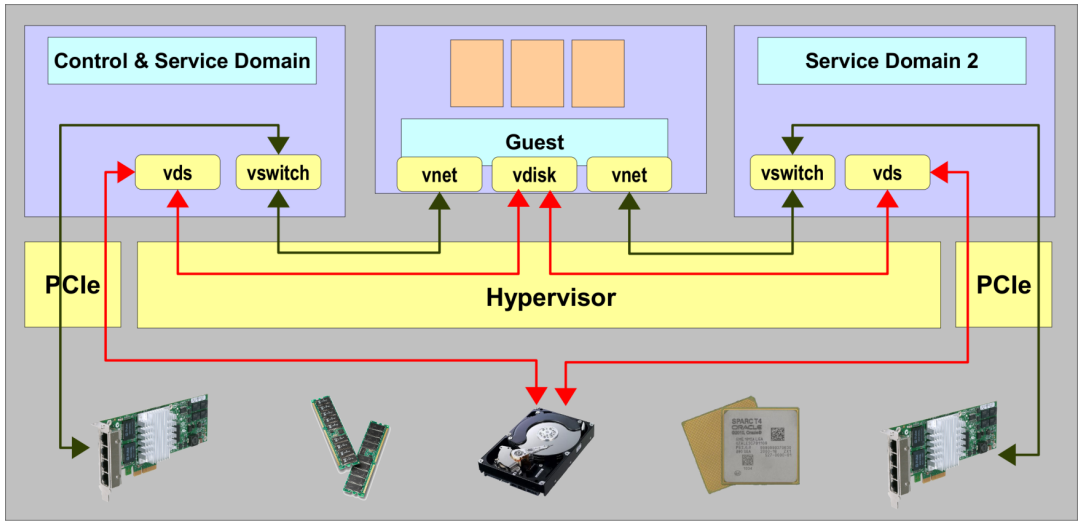
Redundant Service Domains

- Define redundant service domains
  - Each go to the same network or storage
  - Apply MPxIO for storage or IPMP for networks
  - Use separate root complexes to maximize redundancy



Enhanced Serviceability as a Bonus—Rolling Upgrades

- SPARC M8 processor–based servers offer large number of root complexes providing exceptional support for I/O virtualization
- With redundant service domains, each to the same entity, you can accomplish “rolling upgrades”
  - Take down one of the two service domains; I/O activity continues transparently in the other
  - Upgrade the service domain just taken down with a revision to the OS, add new drivers, add another adapter, and so on
  - Restore the upgraded service domain
  - Upgrade the second service domain



Number of Root Complexes per Server

SPARC M8 Processor–Based Servers

	I/O Controller(s)	Root Complexes for PCIe Slots	Total Number of Root Complexes
SPARC T8-1	1 on motherboard	2 + 2[1]	5
SPARC T8-2	2 on motherboard	2 + 6[1]	10
SPARC T8-4	4 on main module	12	20
SPARC M8-8	1 per CMIOU	24	Up to 40

1) N + M, N = dedicated to PCIe slots; M = shared with other devices

SPARC M8 Processor–Based Systems

The world's most advanced systems for enterprise workloads, with unique capabilities for information security, database acceleration, and Java acceleration

Safe Harbor Statement

The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Integrated Cloud

Applications & Platform Services

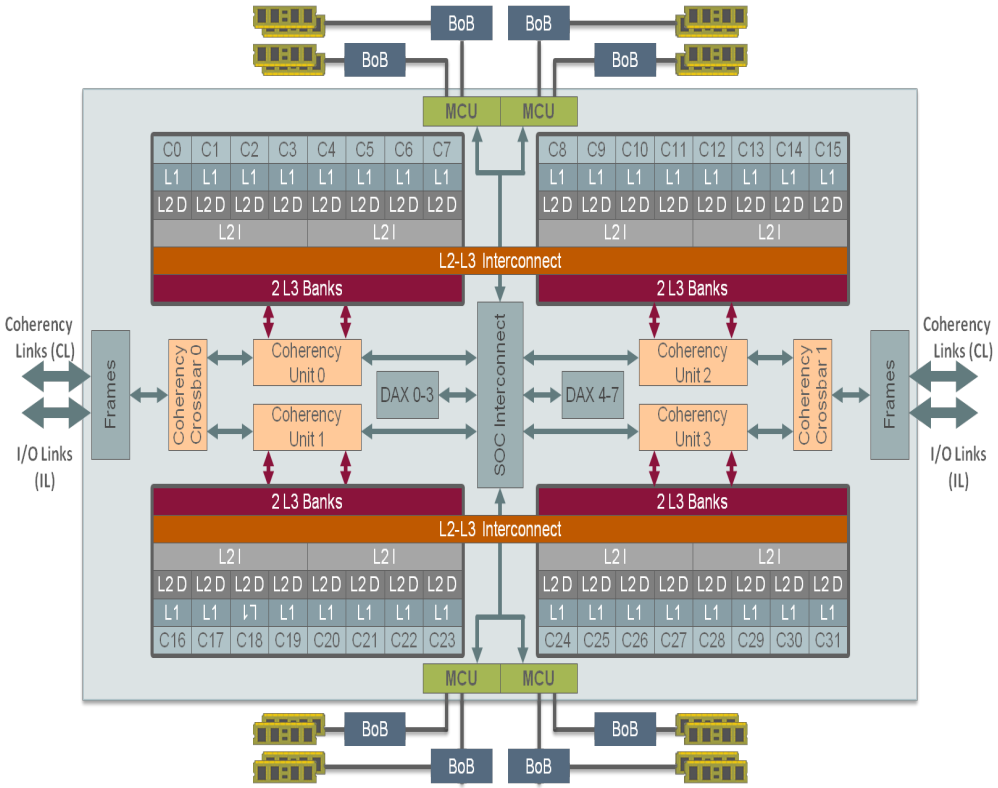

ORACLE®

BACKUP SLIDES

SPARC M8 Processor

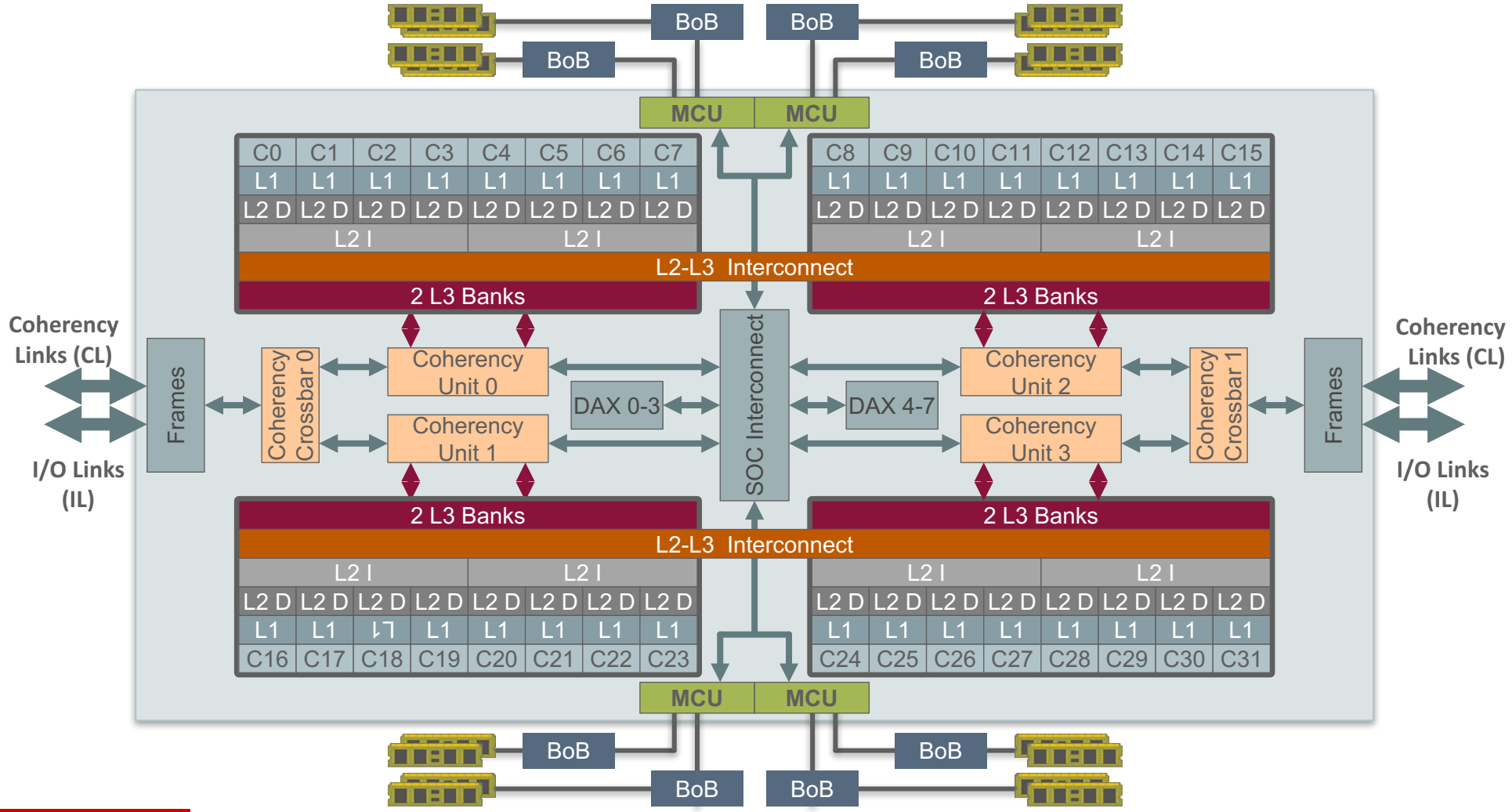
Details

SPARC M8 Processor Overview



- 32 SPARC cores
  - Dynamically threaded, 1 to 8 threads per core
  - Organized as two partitions, each containing 16 cores
- 2nd Generation Software in Silicon NEW
  - Enhanced DAX
  - Enhanced cryptographic acceleration
  - Oracle Numbers unit
- 5th generation SPARC cores NEW
  - 4-wide instruction issue
  - Enhanced cache design
  - Misaligned access support
- Glueless SMP scalability 8 processors
- PCIe 3.0 support via I/O controller ASICs

SPARC M8 Processor Block Diagram



2nd Gen Software in Silicon Performance Features

Feature	HV	OS	DB	Java	Remarks
New Misaligned Ld/St instructions	✓	✓	✓	✓	Improves LZ4, OZIP compression/decompression performance, ZLIB, HPK, and Java.
High b/w Random Number Generator		✓	✓	✓	Security acceleration by using a non-deterministic entropy generator, used in CryptoLibs and OpenSSL libraries.
SHA-3		✓	✓	✓	Acceleration for Keccak algorithm, improves throughput (MB/sec). Used in CryptoLibs and OpenSSL libraries.
REVBYTES, REVBITS		✓	✓	✓	Improves Star Schema Benchmark performance in DB, word and double-word bit and byte reversals in Java
ONadd, ONsub, ONdiv, Onmul			✓	✓	HW acceleration for ON operations in DB – substantial query speedups
Dictionary Unpack		✓	✓	✓	Eliminates need for bmask/bshuffle for dictionary scans in DB queries. Improves HPK and OZIP performance.
RLE Burst, RLE Length		✓	✓	✓	Decompress run length encoded data and calculate length of the input bit stream. Improves HPK and OZIP performance.
New VIS instructions for Partitioned Compare & Shift, Range Compare & Shift		✓	✓	✓	Eliminates the need for separate offsetting/shifting results Merges two independent comparisons plus an AND into a single operation. HPK and OZIP performance improvements,
Context Masking, Multiple Shared Contexts		✓	✓	✓	Enables low-overhead sharing of memory objects
✓ - Denotes changes required to take advantage of feature or enable higher level SW to access feature					

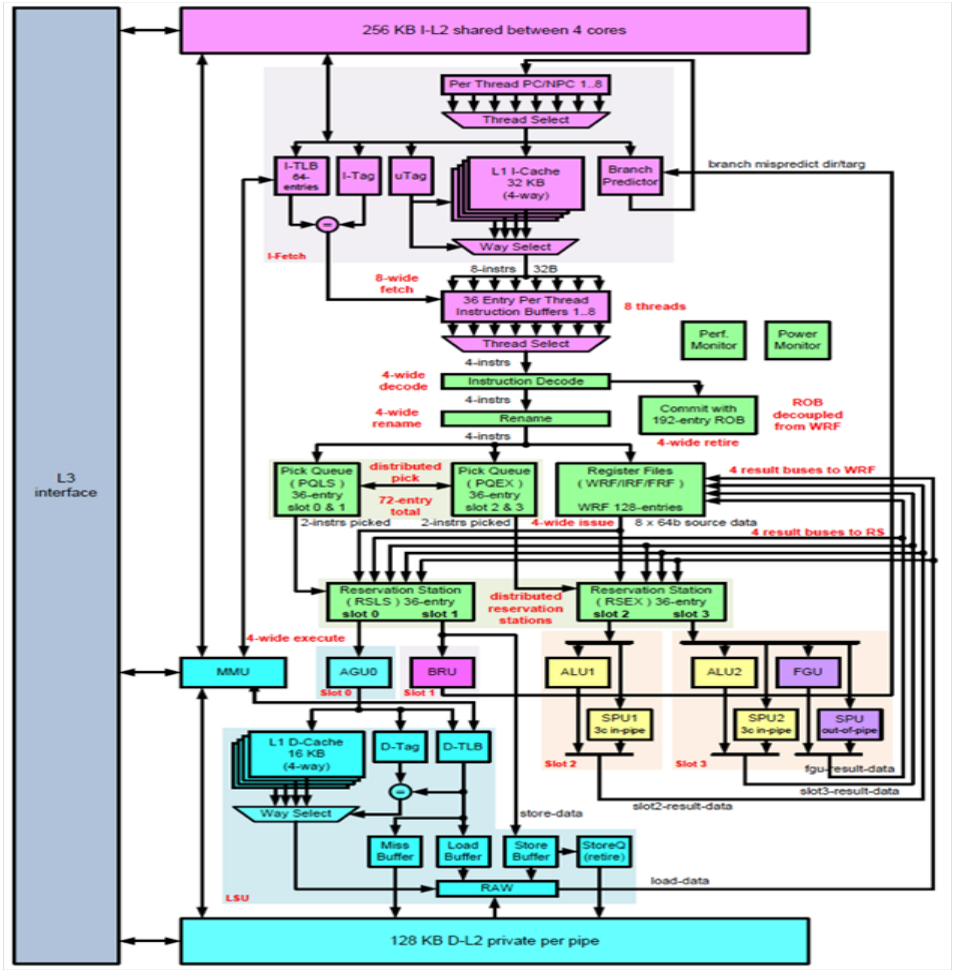
SPARC M8 Processor

Improvements in Microprocessor Capability

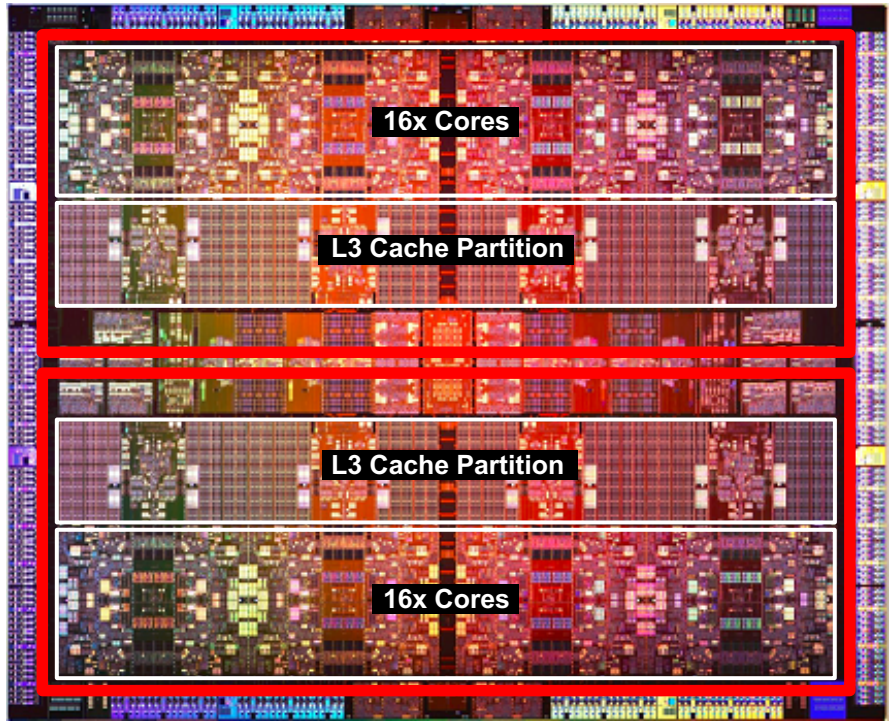
	SPARC M8	SPARC M7
Frequency	**5.0**	4.13
Instruction Issue Width	**4**	2
L1 Cache	**32 KB** four-way I$ 16 KB four-way D$	16 KB four-way I$ 16 KB four-way D$
L2 Instruction Cache	Shared 256 KB four-way I$ per 4 cores	
L2 Data Cache	128 KB eight-way per core	Shared 256 KB eight-way per core pair
L3 Cache	Shared 64 MB, 2 partitions	Shared 64 MB, 4 partitions

SPARC M8 Processor Core

- Dynamically threaded, 1 to 8 threads
- 8 Wide Fetch, 4 Wide Decode, 4 Wide Issue
- Enhanced Branch Prediction NEW
  - Can predict up to 4 branches every cycle
- Oracle Numbers unit NEW
  - 4 new instructions: ONadd, ONsub, ONmul, ONdiv
  - Native support for all ON lengths (up to 22 Bytes)
- Enhanced cryptographic acceleration NEW
  - Added support for SHA-3
- Misaligned access support NEW
  - Micro-architectural handling of a misaligned load/store trap for existing binaries
  - New SPARC instructions for native execution



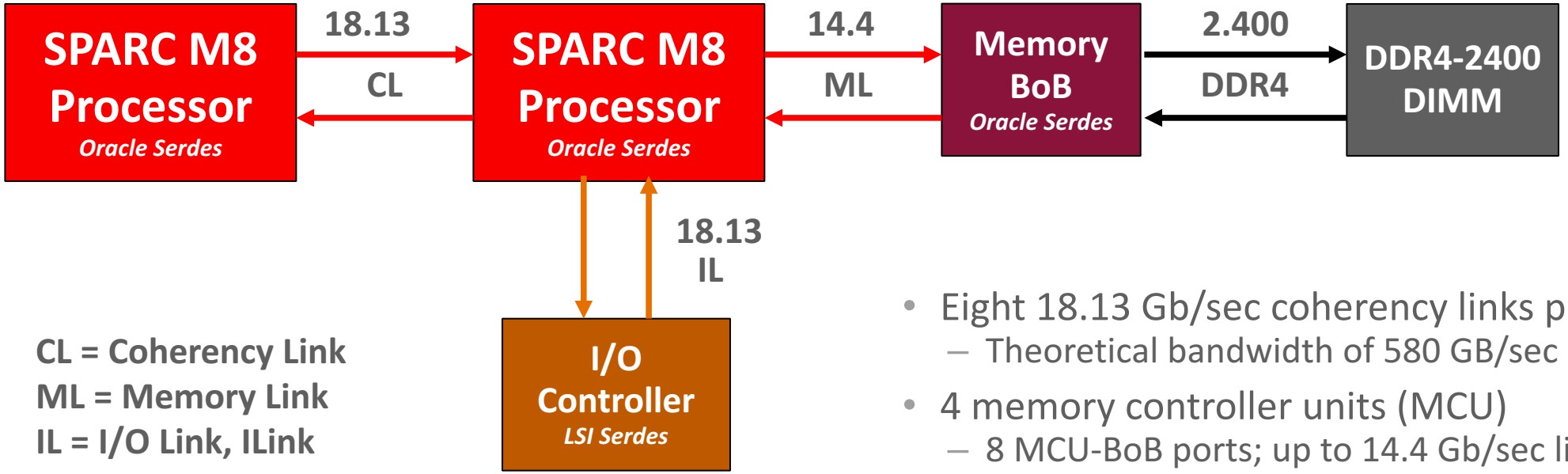
SPARC M8 Fine-Grained Power Management



- On-die power estimator per core
  - Generates dynamic power estimates by tracking internal core activities
  - Estimates updated at 250 nanosecond intervals
- On-die power controller
  - Estimates total power of cores and caches on a partition basis (16 cores + 4 L3$ banks)
  - Accurate to within a few percent of measured power
  - Individually adjusts voltage and/or frequency within each partition based on software-defined policies
- Performance at power optimizations
  - Highly responsive to workload temporal dynamics
  - Can account for workload non-uniformity between partitions
- Partitions may be individually power gated

SPARC M8 Processor Communication Links

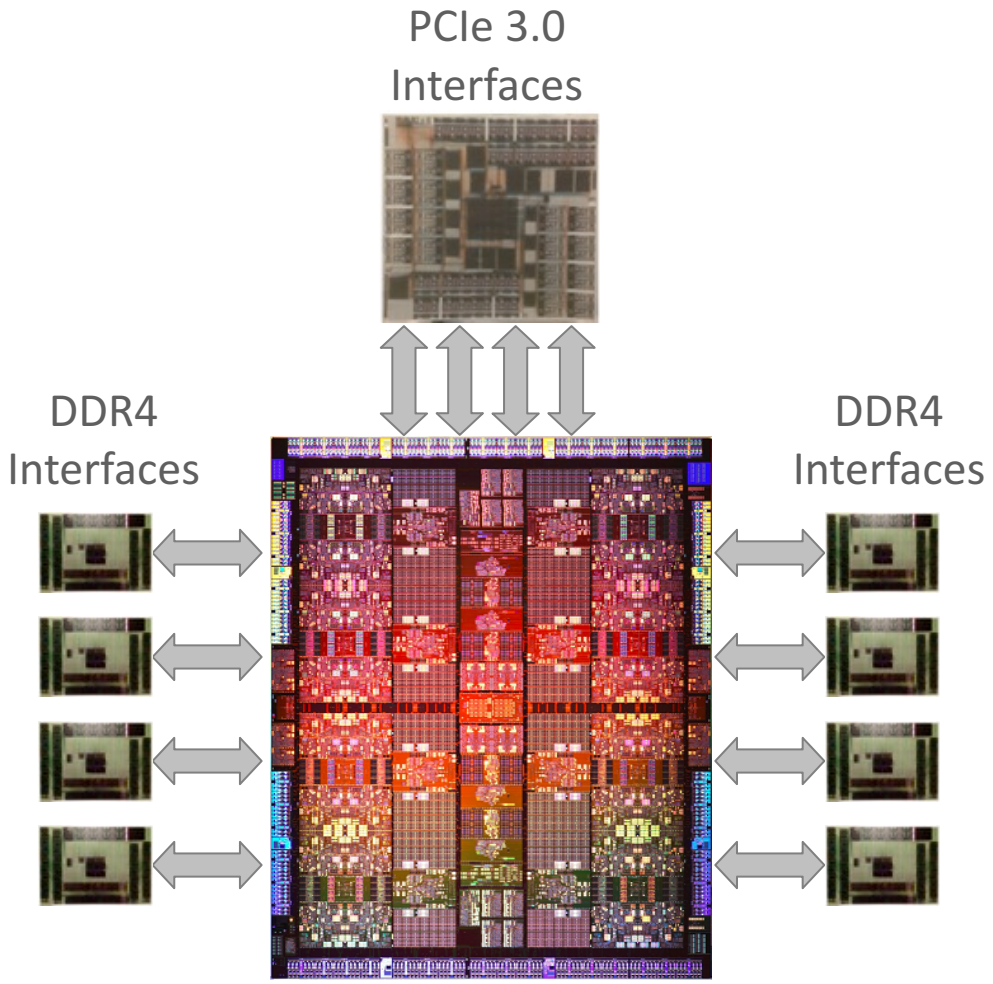
SERDES Link Speeds (Gb/sec per lane or MT/sec per DDR4 channel)



CL = Coherency Link
ML = Memory Link
IL = I/O Link, ILink

- Eight 18.13 Gb/sec coherency links per processor
  - Theoretical bandwidth of 580 GB/sec
- 4 memory controller units (MCU)
  - 8 MCU-BoB ports; up to 14.4 Gb/sec link rates
  - 374 GB/sec raw memory bandwidth per processor
- PCIe 3.0 support via I/O controller ASICs
  - 4 (x8) I/O links; 18.1 Gb/sec/lane link rates
  - 145 GB/sec raw I/O bandwidth per processor

SPARC M8 Memory and I/O



- 4 DDR4 memory controllers
  - 16 DDR4-2400 channels
  - 307 GB/sec raw bandwidth for 16 DDR4 channels
  - 180 GB/sec measured memory bandwidth (preliminary)
  - DIMM retirement without system stoppage
- Memory links to buffer chips
  - 14.4 Gb/sec link rate with DDR4-2400, yielding a 374 GB/sec raw memory bandwidth
  - Lane failover with full CRC protection
- Speculative memory read
  - Reduces local memory latency by prefetching on local L3$ partition miss
  - Dynamic per request, based on history (data, instruction) and threshold settings
- Links to I/O controller ASIC
  - 4 internal links supporting 145 GB/sec raw bandwidth

SPARC M8 Buffer on Board (BoB)

Memory Buffer ASIC

- M8 memory link to the DDR4 interface
  - 12 Tx + 14 Rx at 14.4 Gb/sec/lane
  - 41.6 GB/sec per memory link, bidirectional
- Dual DDR4 channels per BoB
- One DIMM per BoB DDR4 channel
- One or two DIMMs present per BoB
- Two BoBs per SPARC M8 MCU
- 8 BoBs per SPARC M8 processor
- 8 or 16 DIMMs present per SPARC M8 CPU