CHAPTER 1

Introduction to Sun Fire Systems

Most companies want the best solution for their needs, especially when they are purchasing a computer system. However, designing a reliable system that performs well takes careful consideration and planning.

Consider a widely-used analogy—automobiles. A vast range of different types of vehicles, including sedans, sports cars, convertibles, sport-utility vehicles, trucks, and hybrids is available. Each of these types of vehicles has advantages and disadvantages. When trying to decide what type of vehicle to purchase, you must consider your needs. How many people must it be able to carry? How much power does it need? How much cargo must it hold? Is size a concern? And so forth.

You would probably be making a mistake if you purchased a two-seater sports car when you had a family of five that regularly attended soccer games. You would probably be making a similar mistake if you bought a sedan with the intent of using it to haul your trailer and dirt bikes out to the desert on the weekends. For a purchase to be worthwhile, you must ensure that it properly fits your specific requirements.

This analogy may seem trite, but many people do not put nearly as much thought into purchasing a high-end server as they do a car. Instead, a system often is purchased based on its processor speed or size alone. Little, if any, consideration is given to concerns such as how to layout the I/O, how much memory to get, and how much expansion capacity is really needed.

As with the car analogy, you must ensure that the design of your server configuration meets your company's needs. This design process requires time and effort, but it pays off in better system reliability, availability, serviceability (RAS) and performance.

This chapter summarizes the RAS and performance features of the Sun Fire system hardware. The definitions of the RAS features are:

- *Reliability*—The ability of the system to run without interruption, to continue to operate when correctable errors are detected, and to prevent data corruption.
- *Availability*—The percentage of time the customer's system is able to do productive work. The ability to always recover after a failure by testing and bypassing failed components.

- *Serviceability*—The system ensures that repair time (downtime) is minimized.

This chapter covers these topics in two sections—RAS and Performance.

RAS

The RAS goals for the Sun Fire system are to protect the integrity of the customer's data and to maximize availability. The focus is on three areas:

- Problem detection and isolation—knowing what went wrong and ensuring that the problem is not propagated
- Tolerance and recovery—absorbing abnormal system behavior and fixing or dynamically circumventing it
- Redundancy—replicating critical components

To ensure data integrity at the hardware level, all data is error correction code (ECC) protected, and address and control buses are protected by parity checks. These checks ensure the containment of errors.

For tolerance of errors, resilience capabilities are designed into the Sun Fire system to ensure that the system continues to operate, even in a degraded mode. The Sun Fire system can function with one or more processors disabled. In recovering from a problem, the system is checked quickly to determine the fault and to ensure minimum downtime. To reduce downtime, redundant hardware can be configured into the system.

Reliability

Sun Fire systems have five categories of reliability capabilities:

- Reducing the probability of errors
- Detecting and correcting errors using error correction code (ECC)
- Detecting uncorrectable errors with ECC and parity checking
- Redundant power and cooling
- Environmental sensing

Availability

Availability is the ability of a system to be continually accessible and useful to the customer. Sun Fire systems have many features that contribute to this quality, including the ability to:

- Test, identify, and de-configure failed components following a system interrupt
- Configure and boot a usable configuration with a subset of the original configuration
- Change the configuration without interrupts using dynamic reconfiguration (DR).

For higher levels of availability Sun Fire systems can be clustered.

Serviceability

To reduce repair time, the Sun Fire systems are designed with a number of maintenance capabilities and aids. These are used by the Sun Fire system administrator and by the service provider.

Failing components are listed in the failure logs in such a way that the field-replaceable unit (FRU) is clearly identified. You can remove and replace most system components in a properly configured system during system operation without scheduled downtime. If properly configured, CPU/Memory boards, I/O boards, I/O controllers, fans and power supplies can all be replaced while the system is running.

Sun Fire RAS Features

CPU and System Interconnect

The most important reliability feature in any system is the protection of data integrity.

The UltraSPARC® III processor has parity protection for all major internal caches and ECC protection for transactions to and from the external caches.

The data interconnect has ECC and parity protection throughout the system. The address interconnect has parity protection.

CPU Error Protection

The CPU corrects errors detected in the internal data and instruction cache SRAMs. When an internal error is detected, the CPU invalidates the cache and retries the data or instruction load.

Two SRAM modules per CPU reside on the CPU/Memory system board. These modules contain the external cache (Ecache). The CPU corrects all parity errors detected during data cache or instruction cache by invalidating and flushing the cache line. The hardware corrects all single bit errors detected by ECC during data transfers on the fly. ECC also detects and reports any uncorrectable (multibit) errors to the Solaris OE.

System Interconnect Error Protection

The system has three types of error protection—data interconnect, address interconnect, and error isolation.

Data Interconnect

The system protects all data interconnect pathways by using ECC and parity protection. It generates ECC and parity bits for all data blocks sourced by processors and PCI I/O controllers (system devices). All data switches in the path for each transfer check ECC and parity. The receiving system device checks and corrects ECC.

Address Interconnect

The system has parity protection for all address interconnect pathways and checks parity between all system devices and address repeaters. On the Sun Fire ™15K/12K systems there is also ECC protection on the address and address response crossbars for transactions across the centerplane.

Error Isolation

Because each of the data switches in the path for every data transfer checks ECC, the source of ECC errors can be identified in most cases. However, some types of ECC errors are difficult to isolate. For ECC errors, such as a CPU writing bad ECC into memory, finding the source is difficult because multiple system devices may read and report the bad ECC. In such cases, the ECC error usually can be isolated to a dual CPU data switch or its pair of processors. This is an improvement over the previous architecture in which it was more difficult to isolate these types of ECC errors.

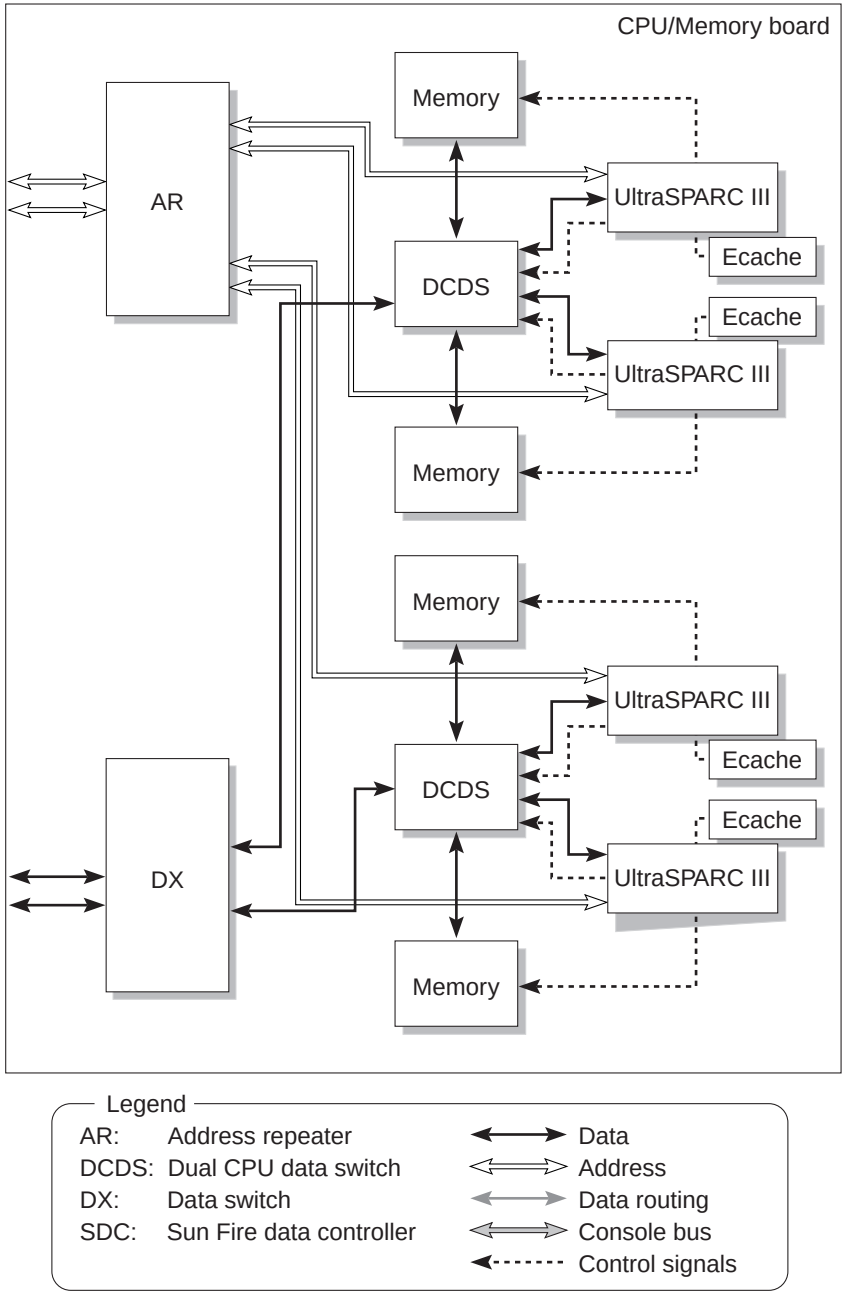


FIGURE 1-1 CPU Board Block Diagram

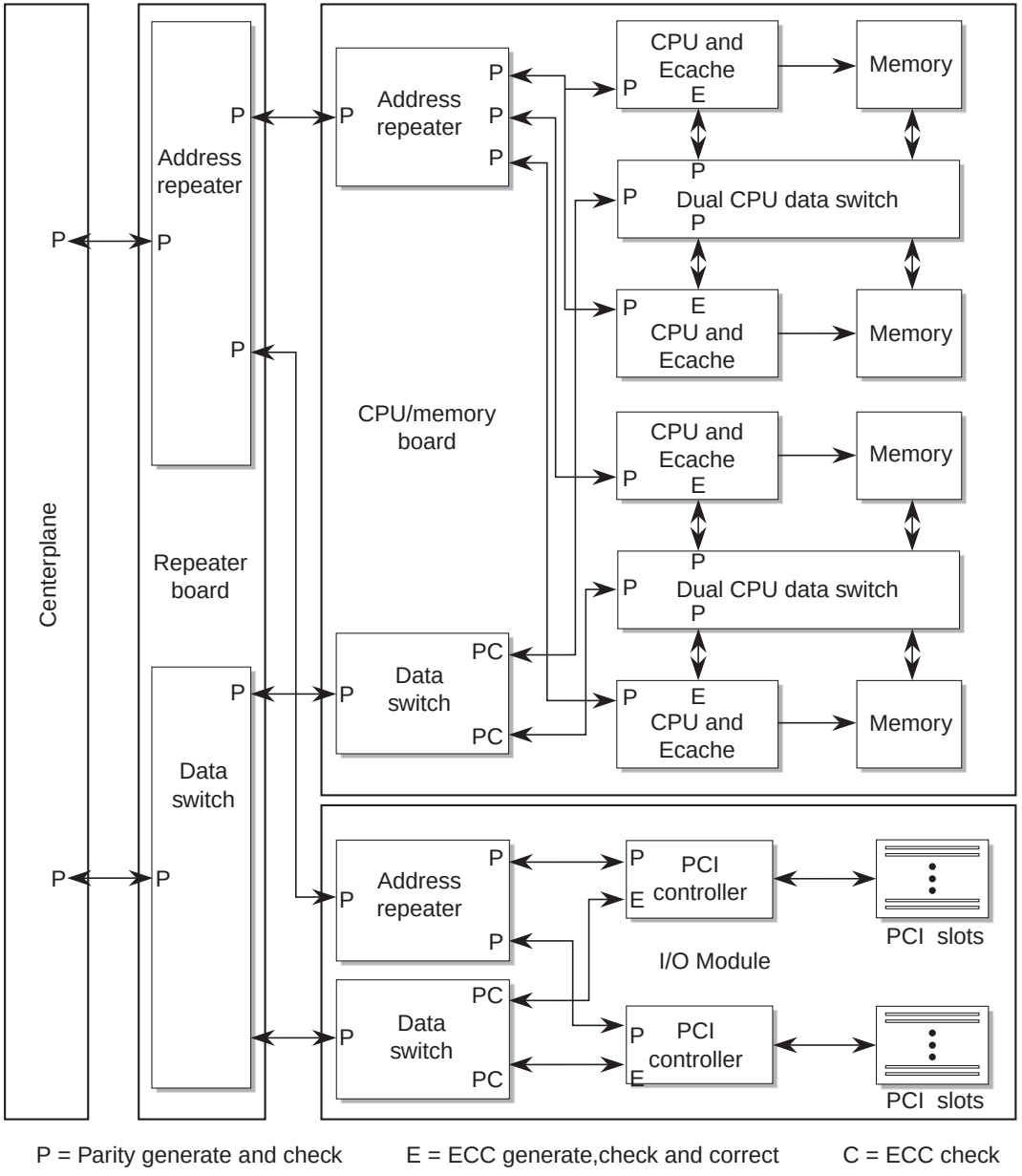
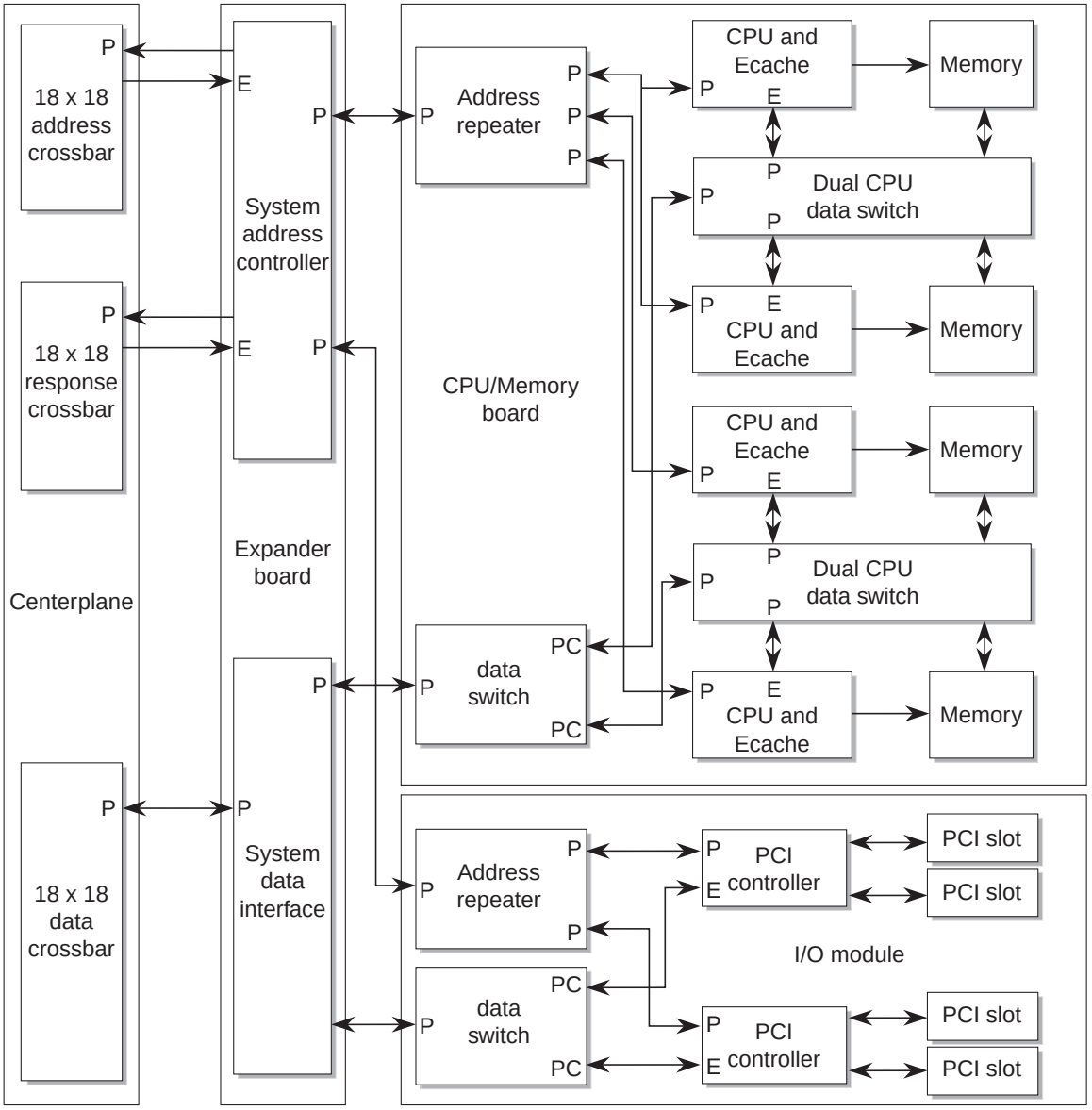


FIGURE 1-2 Sun Fire 6800/4810/4800/3800 Systems Interconnection Diagram



P = Parity generate and check E = ECC generate,check and correct C = ECC check

FIGURE 1-3 Sun Fire 15K / 12K Systems Interconnection Diagram

System Controller

A small system called the system controller (SC) manages the Sun Fire systems. The SC is responsible for all of the functions required to test, configure, and boot domains. It supplies all system clocks and virtual TOD clocks for the domains and monitors power and environmental status. It also provides domain control with virtual key switches and console connections for each domain.

Testing and Configuration

Upon power on or a reset, the SC runs a self test called SCPOST. It then starts the platform management software.

The SC configures and coordinates the initialization, testing, and boot processes. After a system failure, and when the virtual key switch of a domain is turned on, the SC powers on the system components associated with the domain and runs SPOST for the domain. SPOST controls the running of LPOST and IPOST. LPOST tests the CPU/Memory boards, I/O boards, and the system interconnect. IPOST tests the PCI controllers. The SC then configures the domain based on the components that pass the tests and starts the boot sequence.

Environmental Monitoring

The SC monitors the following conditions:

- Voltage, current, and temperatures for power supplies
- Voltage and temperatures for all system boards and processors
- Temperatures of ASICS
- Fan status and speed

If safe thresholds are exceeded, the SC shuts down components to prevent damage to the system.

System Administration and Maintenance

The SC provides access for platform and domain administration. Access to the SC is by the included RS-232 serial connection or by network connection.

Tasks performed at the platform level are:

- SC setup and configuration
- Allocation of system resources for domains
- Domain creation

- Status display of all domains
- Power control for all system components
- Logical enable and disable of components
- Individual CPU/Memory board tests
- Component and environmental status display
- Platform error message administration
- Platform password setup (Sun Fire 6800/4810/4800/*3*800 systems)
- Platform and Domain Security configuration (Sun Fire 15K/12K systems)

Tasks performed at the domain level are:

- Power and boot control
- Domain status display
- Logical domain component enable and disable
- Individual CPU/Memory board tests
- Domain error message administration
- Domain password setup (Sun Fire 6800/4810/4800/*3*800 systems)

Redundant System Components

All systems in the Sun Fire system product line can be configured with redundant components. The ability to run with a subset of configured components increases availability of the system. As long as one processor with memory and one I/O module are functional, the system can run. If the system is configured with redundant connections to storage and network, the access can be maintained using these alternate paths.

Redundant components include:

- CPU/Memory boards
- I/O modules
- PCI cards
- System Controller boards
- Repeater boards (Sun Fire 6800/4810/4800/*3*800 systems)
- Sun™ Fireplane interconnect (Sun Fire 15K/12K systems)
- Fan trays
- Power supplies

CPU/Memory Boards

Depending on the model, Sun Fire systems can support up to 18 CPU/Memory boards. Each board contains two or four CPUs and is capable of running independently or together with other boards in a larger domain.

I/O Modules

Depending on the model, Sun Fire systems can support up to 18 I/O modules:

- Sun Fire 15K system supports up to 18 I/O modules, with four PCI slots each.
- Sun Fire 12K system can have up to nine I/O modules with four PCI slots each.
- Sun Fire 6800 system can be configured with a combination of four I/O modules with eight PCI or four Compact PCI slots each.
- Sun Fire 4810/4800 can be configured with two I/O Modules with eight PCI or four Compact PCI slots each.
- Sun Fire 3800 system is configured with two I/O modules with six Compact PCI slots each.

TABLE 1-1 summarizes these configurations.

TABLE 1-1 I/O Module Configurations

System	I/O Modules	PCI Slots	Compact PCI Slots
Sun Fire 15K	18	4 each	0
Sun Fire 12K	9	4 each	0
Sun Fire 6800	4	8 each	or 4 each
Sun Fire 4810/4800	2	8 each	or 4 each
Sun Fire 3800	2	0	6 each

PCI Cards

Redundant PCI cards can be configured to provide alternate paths to all peripheral connections.

System Controller Boards

With two SC boards configured in the system, a failure of the primary SC does not cause a domain interrupt. The system clocks, virtual TOD clocks, and all other SC functions fail over to the secondary SC without causing a domain failure.

Repeater Boards

The Sun Fire 6800/4810/4800 systems contain pairs of Repeater boards. The Repeater boards contain data and address repeater switches. The Sun Fire 6800 contains two pairs of Repeater boards. If one board fails, the system can run in degraded mode on the other pair of boards. The Sun Fire 4810/4800 system is configured with one pair of Repeater boards. If one board fails, the system can run in degraded mode on the other board. The Repeater boards can be replaced while the system is running. Although the Sun Fire 3800 system has the same ability to run in degraded mode in the case of a failure, the functionality of the Repeater boards is built into the centerplane and cannot be replaced without an interrupt of the system. TABLE 1-2 summarizes these configurations.

TABLE 1-2 Repeater Board Configurations

System	Repeater boards
Sun Fire 3800	Functionality is built into the centerplane
Sun Fire 4810/4800	1 pair
Sun Fire 6800	2 pairs

Sun implements the Sun Fire 15K/12K interconnect differently than the midrange systems. The interconnect consists of three independent 18 way crossbars—one for data, one for addresses, and one for address responses. Some of the interconnect is made of multiple ASICs that reside on the centerplane that cannot be replaced while the system is running. If one of the crossbars fails, the system can run in degraded mode while the other crossbars continue with full bandwidth. Depending on the failure, the degraded crossbar may be configured to affect a single domain.

Fan Trays

All Sun Fire systems can be configured with redundant fan trays. All fan trays can be replaced while the system is running.

Power Input and Supplies

The Sun Fire 6800/4810/4800/3800 systems can be configured with up to four separate AC input connections. The systems and peripherals can be connected to two different AC input power grids using two Redundant Transfer Units (RTU), each with two redundant transfer switches (RTS).

The Sun Fire 15K/12K systems are configured with six dual AC-DC power supplies, each with two AC input connections. The power supplies convert the AC voltage to 48 VDC. The 48 VDC is supplied to all of the system modules. Each system module contains its own on-board DC-DC converter to supply the lower DC voltages needed by the logic components local to each module. A failure of a DC-DC converter only affects that board.

Domains and Partitions

The definitions of these features are:

Domain—The ability to create logically independent multiple sections within a partition, with each domain running its own operating system. The Sun Fire 6800 system can have up to four domains. The Sun Fire 4810/4800/3800 systems can each have up to two domains. Each instance of the Solaris Operating Environment (Solaris OE) runs in its own domain. Domains do not depend on each other and do not interact with each other.

- *Partition*—A partition differs from a domain in the level of isolation. The Repeater boards are logically isolated from each other so the system functions as two separate servers. On a Sun Fire 6800 system, segments can be configured to reside completely within a single internal Sun Fire 6800 power grid. The Sun Fire 15K/12K systems interconnect does not use Repeater boards and does not support multiple partitions.

A Sun Fire system can be logically divided into multiple domains. Since each domain is comprised of one or more system boards, a domain can have anywhere from two to 106 processors (on the Sun Fire 15K system). Each Sun Fire system has at least one domain to support the main functionality of the system.

Additional domains can be used for:

- Testing new applications
- Operating system updates
- Configuring several domains to support separate departments

Each domain runs its own instance of the operating system and has its own peripherals and network connections. Domains can be configured without interrupting the operation of other domains on the same system.

While production work continues on the remaining (and usually larger) domain, there is not any adverse interaction between any of the domains. You can gain confidence in the stability of applications or upgrades without disturbing production work. When the testing work is complete, the system can be rejoined logically without rebooting (there are no physical changes when you use domains). Thus, if problems occur, the rest of your system is not affected.

The Sun Fire 15K/12K systems can be configured with up to 18/9 domains. The Sun Fire 15K/12K systems do not use partitions. The Expander boards are responsible for domain separation.

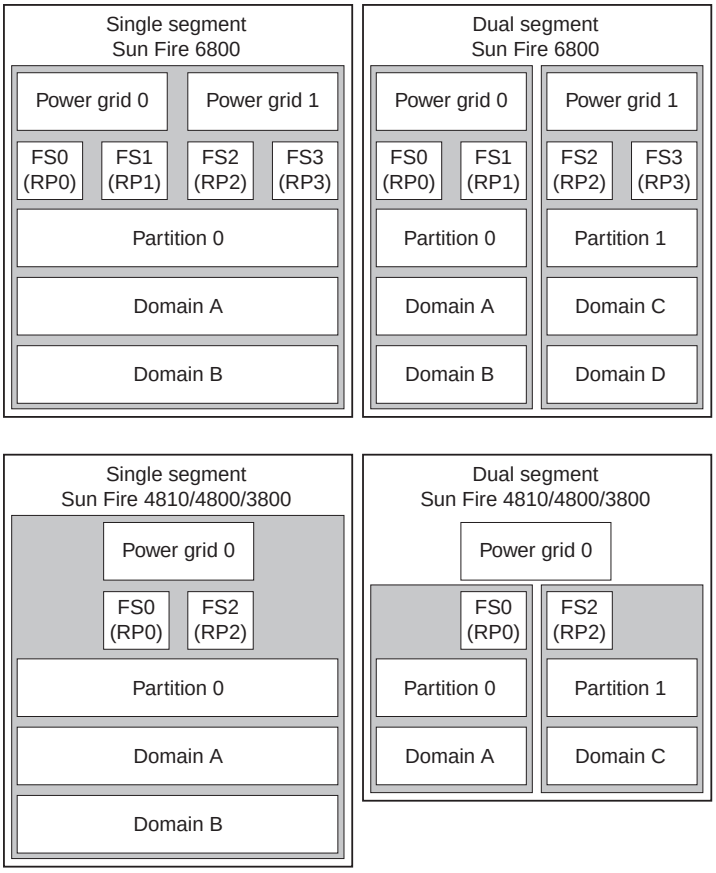


FIGURE 1-4 Sun Fire 6800/4810/4800/3800 Domain and Partition Allocations

Mechanical Serviceability

Connectors are keyed so that boards cannot be installed upside down. Special tools are *not* required to access the inside of the system. This is because all voltages within the cabinet are considered extra-low voltages (ELVs) as defined by applicable safety agencies.

No jumpers are required for configuration of the Sun Fire system. This makes it much easier to install new and/or upgraded system components. There are no slot dependencies other than the special slots required for the SC and Repeater boards.

The Sun Fire system cooling-system design consists of redundant, hot-swappable modules. Standard proven parts and components are used wherever possible. Sun designs the field-replaceable units (FRUs) and subassemblies for quick and easy replacement with minimal use of tools required.

Performance

This section describes the performance features of the Sun Fire system hardware.

UltraSPARC III Processor

The UltraSPARC III is a high-performance implementation of the 64-bit SPARC® V9 architecture.

Features of the UltraSPARC III CPU include:

- Full binary compatibility with all UltraSPARC CPU applications
- VIS instruction set for better 2D and 3D graphic processing
- 4-way superscalar
- 14-stage, non-stalling pipeline
- Integrated memory controller
- L1 caches: integrated instruction (32 kilobytes) and data (64 kilobytes)
- L2 cache: 8 megabytes
- MP scalability: Over 1000 CPUs per system
- System bus: Sun Fireplane interconnect at 150 MHz
- Error checking and correction (ECC) in external cache

TABLE 1-3 lists the UltraSPARC III CPU performance improvements over the UltraSPARC II CPU performance.

TABLE 1-3 UltraSPARC III CPU Performance Improvements

Item	Improvement
Clock frequency	Approximately 2X
Data cache size	4X
Instruction cache size	2X
Memory bandwidth	2 to 3X
External cache bandwidth	2X

CPU/Memory Board

The CPU/Memory board is common to all Sun Fire mid-range and high-end systems. The CPU/Memory board holds up to four processors. Each processor has an associated memory subsystem of eight DIMMs, so memory bandwidth and capacity both scale as processors are added. The memory capacity of the board is 32 gigabytes using 1-gigabyte DIMMs. The maximum memory bandwidth on a board is 9.6 gigabytes per second. The CPU/Memory board has a 4.8 gigabyte per second connection to the rest of the system.

I/O Modules

Each Sun Fire system I/O module contains two PCI controllers. Each controller provides one 66 MHz PCI bus, and one 33 MHz PCI bus. Each PCI bus contains two or more slots for PCI cards. A Sun Fire I/O Module has a 2.4 gigabyte per second connection to the rest of the system.

System Interconnect

All Sun Fire systems use the Sun Fireplane interconnect architecture, which is the coherent shared-memory protocol used by the UltraSPARC III/IV processor generation. Sun Microsystems uses an improved system interconnect with each new processor generation to keep system performance scaling with CPU performance.

The Sun Fireplane architecture is an improvement over the previous generation Ultra Port Architecture (UPA). The system clock rate is increased by fifty percent from 100 megahertz to 150 megahertz. The snoops-per-clock is doubled from one half to one. Taken together, these improvements triple the snooping bandwidth to 150 million addresses per second. The maximum data bandwidth for the Sun Fire 6800 and smaller systems is 9.6 gigabytes per second—triple that of the previous generation Sun Enterprise™ 6500/5500/4500/3500 systems.

The Sun Fireplane architecture also adds a new layer of point-to-point directory-coherency protocol, for use in systems that require more bandwidth than a single snooping bus can provide coherency for. This facility allows coherency to be maintained between multiple snooping buses, and is used in the Sun Fire 15K/12K systems. TABLE 1-4 lists the Sun Fire system interconnect specifications.

TABLE 1-4 Sun Fire System Interconnect Specifications

	Sun Fire 3800 system	**Sun Fire 4810 / 4800 systems**	**Sun Fire 6800 system**	**Sun Fire 15K/12K systems**
System clock	150 MHz			
Coherency protocol	Snooping			Snooping on each board set, directory across centerplane
Address interconnect	1 snooping bus			18 snooping buses, 18x18 global address crossbar, 18x18 global response crossbar
CPU/Memory board internal bisection bandwidth	4.8 Gbytes/sec			
CPU/Memory board external bandwidth	4.8 Gbytes/sec			
I/O board external bandwidth	2.4 Gbytes/sec			
Inter-board data interconnect	4 x4 crossbar	5 x 5 crossbar	10 x10 crossbar	18 3x3 crossbars, 18x18 global crossbar
Same-board bandwidth	9.6 Gbytes/sec			121 Gbytes/sec
Different-board bandwidth	4.8 Gbytes/sec	9.6 Gbytes/sec		43/21.6 Gbytes/sec

Centerplane Implementation

The centerplane implementation depends upon system size. The Sun Fire 6800 and 4810/4800 systems have passive centerplanes, with switch ASICs located on the Repeater boards. The Sun Fire 6800 system uses four Repeater boards to implement a 10x10 data crossbar that connects the CPU/Memory boards and the I/O modules together. The Sun Fire 4810/4800 systems use two Repeater boards to implement a 5x5 data crossbar. The Sun Fire 3800 system uses an active centerplane. The Repeater board functionality is included on the centerplane to implement a 4x4 data crossbar.

The Sun Fire 15K/12K systems use an Expander board to implement a 3x3 switch between a CPU/Memory board, an I/O module, and the centerplane port.

The Sun Fire 15K/12K system has three 18x18 crossbars on the active centerplane to provide connections between the Expander boards. The three crossbars are separate busses for address, response, and data transfers. This method keeps address traffic from interfering with data traffic. The peak Sun Fire 15K system centerplane bandwidth is 43 gigabits per second and 21.6 gigabits per second for the Sun Fire 12K system.

System Configurations

TABLE 1-5 lists the Sun Fire system maximum configurations.

TABLE 1-5 Sun Fire System Maximum Configurations

	Sun Fire 3800 system	Sun Fire 4800 system	Sun Fire 4810 system	Sun Fire 6800 system	Sun Fire 12K system	Sun Fire 15K system
CPU/Memory boards	2	3		6	9	18
Processors	8	12		24	38/52[1]	72/106[1]
Number of DIMMs	64	96		192	288	576
Memory capacity (with 1 Gbyte DIMMs)	64 Gbytes	96 Gbytes		192 Gbytes	288 Gbytes	576 Gbytes
Centerplane	Active	Passive			Active	
Repeater boards	0	2		4	NA	
Expander boards	NA				9	18
Domains	2	2		4	9	18
I/O/MaxCPU Modules	2	2		4	9	18

TABLE 1-5 Sun Fire System Maximum Configurations *(Continued)*

	Sun Fire 3800 system	Sun Fire 4800 system	Sun Fire 4810 system	Sun Fire 6800 system	Sun Fire 12K system	Sun Fire 15K system
PCI card types	hot-swap Compact-PCI	PCI and hot-swap CompactPCI			hot-swap PCI	
PCI slots per assembly	6	8 per PCI, 4 per cPCI			4	
Max total PCI slot	12	16		32	36	72
Bulk power supplies	2	3		6		6
Power requirements	100–120 VAC or 200–240 VAC	200–240 VAC				
System Controller boards	2					
Redundant cooling	Yes					
Redundant AC input	Yes					
Enclosure	Rackmount	Deskside or Rackmount	Rackmount	Sun Fire 6800 cabinet	Sun Fire 12K cabinet	Sun Fire 15K cabinet

1. Maximum CPU count is attained by putting MaxCPU boards in the I/O slots.

System Board Designations and Locations

TABLE 1-6 through TABLE 1-15 list the system board designations and locations.

TABLE 1-6 System Board Abbreviations

Abbreviation	System Board	Sun Fire System
IB[6-9]	I/O Module	6800/4810/4800/3800
FP	Filler Panel	All
IO[0-17]	I/O Board	15K/12K
MaxCPU[0-17]	Dual Processor Board	15K/12K
SB[0-17]	CPU/Memory Board	All
SC[0-1]	System Controller	All
SCPER[0-1]	System Controller Peripheral Board	15K/12K

TABLE 1-7 Sun Fire 6800 System Board Locations

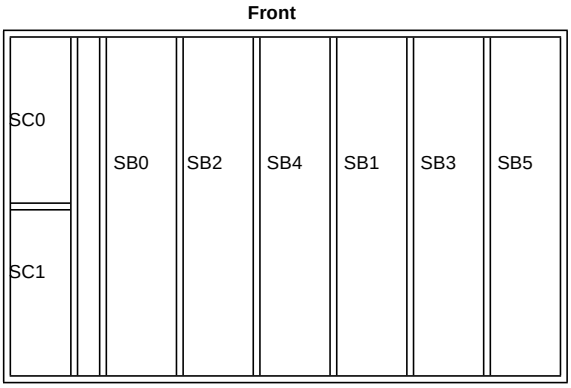


TABLE 1-8 Sun Fire 6800 System Board Locations

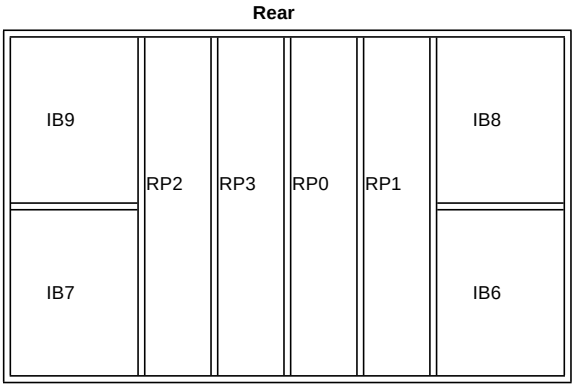


TABLE 1-9 Sun Fire 4810 System Board Locations

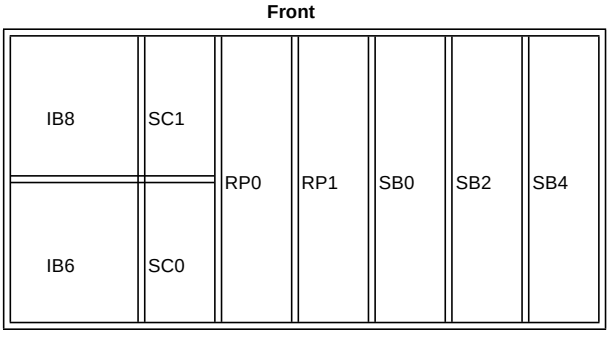


TABLE 1-10 Sun Fire 4800 System Board Locations

Rear

IB8	SC1	RP0	RP1	SB0	SB2	SB4
IB6	SC0					

TABLE 1-11 Sun Fire 3800 System Board Locations

Front

SC1	
SB2	
SC0	
SB0	
IB6	IB8

TABLE 1-12 Sun Fire 15K System Board Locations

Front

SB8	SB7	SB6	SB5	SB4	SB3	SB2	SB1	SB0	SC0
IO8 / Max CPU	IO7 / Max CPU	IO6 / Max CPU	IO5 / Max CPU	IO4 / Max CPU	IO3 / Max CPU	IO2 / Max CPU	IO1 / Max CPU	IO0 / Max CPU	SCPER0

TABLE 1-13 Sun Fire 15K System Board Locations

Rear

SB17	SB16	SB15	SB14	SB13	SB12	SB11	SB10	SB9	SC1
IO17 / Max CPU	IO16 / Max CPU	IO15 / Max CPU	IO14 / Max CPU	IO13 / Max CPU	IO12 / Max CPU	IO11 / Max CPU	IO10 / Max CPU	IO9 / Max CPU	SCPER1

TABLE 1-14 Sun Fire 12K System Board Locations

Front

SB8	SB7	SB6	SB5	SB4	SB3	SB2	SB1	SB0	SC0
IO8 / Max CPU	IO7 / Max CPU	IO6 / Max CPU	IO5 / Max CPU	IO4 / Max CPU	IO3 / Max CPU	IO2 / Max CPU	IO1 / Max CPU	IO0 / Max CPU	SCPER0

TABLE 1-15 Sun Fire 12K System Board Locations

Rear

FP	FP	FP	FP	FP	FP	FP	FP	FP	SC1
FP	FP	FP	FP	FP	FP	FP	FP	FP	SCPER1

System Dimensions and Footprints

FIGURE 1-5 shows the dimensions and footprints of the Sun Fire 4810/4800/3800 systems. FIGURE 1-6 shows the dimensions and footprints of the Sun Fire 15K/6800 systems.

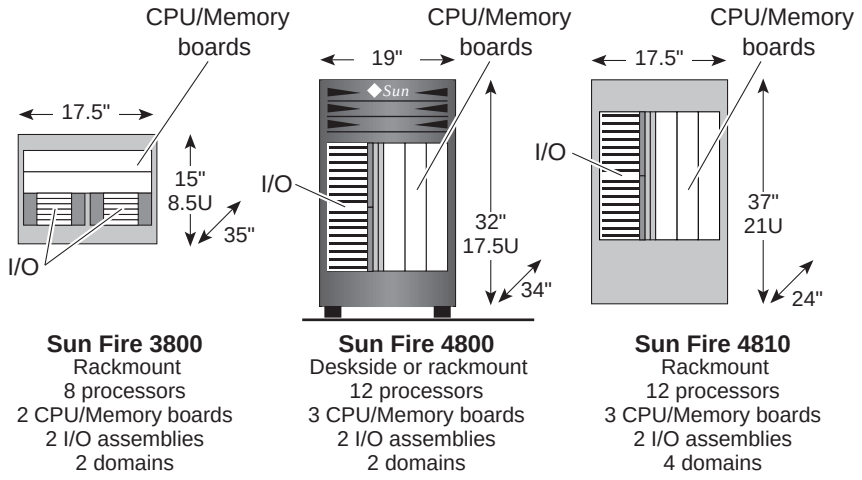


FIGURE 1-5 Sun Fire 4810/4800/3800 Systems Footprints and Dimensions

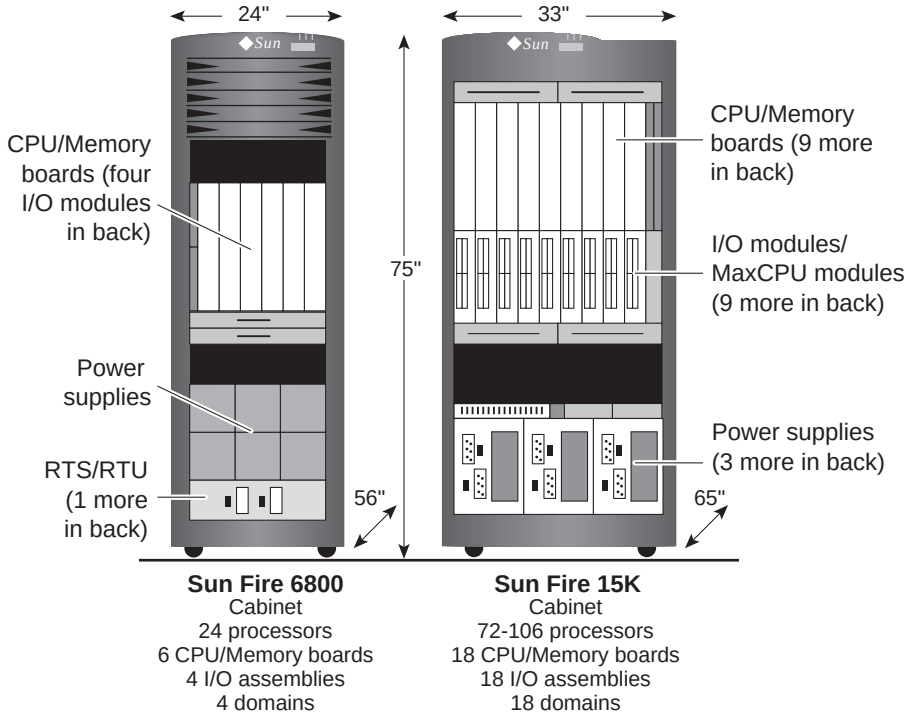


FIGURE 1-6 Sun Fire 15K/6800 Systems Footprints and Dimensions