CS-Storm™ 500GT 3U Server Hardware Guide

(Rev C)

H-6150

Contents

About the CS-Storm 500GT 3U Server Hardware Guide

The *Cray® CS-Storm 500GT™ 3U Server Hardware Guide H-6150* describes the 3U server (Model 7201) components and features. This guide does not include information about peripheral switches or network fabric components. Refer to the manufacturer's documentation for peripheral equipment.

Document Versions

Table 1. Record of Revision

Publication Title	Date	Updates
CS-Storm™ 500GT 3U Server Hardware Guide H-6150 Rev C	Feb 2018	Volta 100 GPU.
CS-Storm™ 500GT Hardware Guide H-6150 Rev B	Oct 2017	Technical updates.
CS-Storm™ 500GT Hardware Guide H-6150 Rev A	Sept 2017	Original publication.

Scope and Audience

This document provides information about the CS-Storm 500GT 3U server. Installation and service information is provided for users who have experience maintaining high performance computing (HPC) equipment. Installation and maintenance tasks should be performed by experienced technicians in accordance with the service agreement.

Related Publications

- *CS-Storm 500GT Hardware Replacement Procedures H-6159*

Acronyms and Terms

The following table lists the acronyms and their definitions used in this guide.

Acronym	Definition
Accelerator	Specialized hardware that performs some functions more efficiently than is possible with software running on a more general-purpose CPU. GPU-accelerated computing is the use of a GPU together with a CPU to accelerate scientific, analytics, engineering, consumer, and enterprise applications. In use, GPU accelerator is often shortened to GPU.
ASHRAE	American Society of Heating Refrigeration and Air Conditioning Engineers.
BIOS	Basic Input/Output System. Non-volatile firmware used to perform hardware initialization during the booting process, and to provide runtime services for the operating system.

Acronym	Definition
Bridge board	Bridge board. A PCI board/card that provides front panel control signals from the motherboard to the power backplane and SATA signals from the motherboard to the disk backplane.
FPGA	Field Programmable Gate Array. An integrated circuit designed to be configured by a customer after it is manufactured.
GPU	Graphics Processing Unit (GPU). A processor chip that performs rapid mathematical calculations, primarily for the purpose of rendering images. GPUs perform parallel operations on multiple sets of data.
KVM	Keyboard Video Mouse (KVM). A rackmounted drawer unit with display screen, keyboard, and mouse or touch pad used to control multiple computers in a data centers.
I²C	Inter-Integrated Circuit. A multi-master, multi-slave, packet switched, single-ended, serial computer bus. It is typically used for attaching lower-speed peripheral ICs to processors and microcontrollers in short-distance, intra-board communication. I²C is often spelled I2C and pronounced I-two-C.
IFB	Interface board. A printed circuit board (PCB) assembly used for the transmission of signals between different components/systems within the server.
MDC	Management daughter card. A printed circuit board (PCB) assembly with IO interface used to configure, monitor, and manage server subsystems and components.
NVMe	Non-Volatile Memory Express (NVMe). A logical device interface specification for accessing non-volatile storage media attached through a PCI Express (PCIe) bus. NVMe is commonly flash memory that comes in the form of solid-state drives (SSDs).
PCIe 3.0	Peripheral Component Interconnect Express, 3rd generation I/O.
PCIe switch board	A PCIe expansion backplane with 10 PCIe x16 Gen3 slots that expand the motherboard PCIe lanes and computing resources.
PLX	PLX Technology, Inc. is the manufacturer of the PEX8796 PCIe 3.0 multiple-host switching integrated circuit (IC) chips used on the PCIe switch board.
RU	Rack unit. Abbreviated RU or U, is a height measurement defined as 44.5 mm (1.75 in). Most frequently refers to the overall height of 19-inch and 23-inch rack frames, as well as the height of servers/equipment that mounts in these frames.
SATA	Serial AT Attachment (SATA). A computer bus interface that connects host bus adapters to mass storage devices such as hard disk drives and solid-state drives.
SMBus	System Management Bus. A single-ended, simple, two-wire bus used for lightweight communication. It is typically used in computer motherboards for on/off communication with the power source.
SSD	Solid-state storage device (SSD). SSDs use integrated circuit assemblies as memory to store data persistently so the data can continue to be accessed. SSDs have no moving mechanical components as do traditional electromechanical magnetic disks such as hard disk drives (HDDs).

Acronym	Definition
U.2	U.2 formerly known as SFF-8639, is a computer interface for connecting SSDs to a computer. It uses up to four PCI Express lanes.
UPI	Intel® UltraPath® Interconnect. UPI is a point-to-point processor interconnect capable of up to 10.4 GT/s. With the Intel Xeon Scalable processor family (formerly code-named Skylake-SP), UPI replaces the Intel QuickPath Interconnect (QPI).

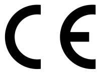
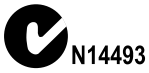
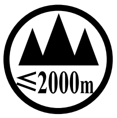
Product EMC Compliance

- FCC Part 15 (USA)
- EN55022 (Europe)
- ICES-003 Emissions (Canada)
- VCCI Emissions (Japan)
- KC Certification (Korea)

Product Regulatory Compliance Markings

The CS-Storm 500GT model 7201 chassis and system components are marked with the following regulatory and certification markings.

Regulatory Compliance	Country	Marking
FCC Marking (Class A)	USA	INFORMATION TO THE USER This equipment has been tested and found to comply with the limits for a Class A digital device, pursuant to part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference in which case the user will be required to correct the interference at his own expense. WARNING Changes or modifications not expressly approved by the manufacturer could void the user's authority to operate the equipment.
NRTL (National Recognized Test Laboratory)	USA/Canada	Safety tested Production monitored C TÜV SÜD US

Regulatory Compliance	Country	Marking
CE Mark	Europe	 WARNING This is a class A product. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.
EMC Marking (Class A)	Canada	This Class [A] digital apparatus complies with Canadian ICES-003. Cet appareil numerique de la classe [A] est conforme a la norme NMB-003 du Canada.
VCCI Marking (Class A)	Japan	この装置は, 情報処理装置等電波障害自主規制協議會 (VCCI) の基準に基づくクラス A 情報技術装置です. この装置を家庭環境で使用すると電波妨害を引き起こすことがあります. この場合には使用者が適切な對策を講ずるよう要求されることがあります.
C-Tick Marking (Class A)	Australia	
Replaceable Lithium battery Warning Information	UL Safety	**CAUTION** RISK OF EXPLOSION IF BATTERY IS REPLACED BY AN INCORRECT TYPE. DISPOSE OF USED BATTERIES ACCORDING TO THE INSTRUCTIONS
Low Altitude Use	China	Only use at altitude not exceeding 2000m. 
AC Symbol	All	IEC 60417-5032 Alternating current

Regulatory Compliance	Country	Marking
Stand-by Symbol	All	IEC 60417-5009 Stand-by ⏻

Trademarks

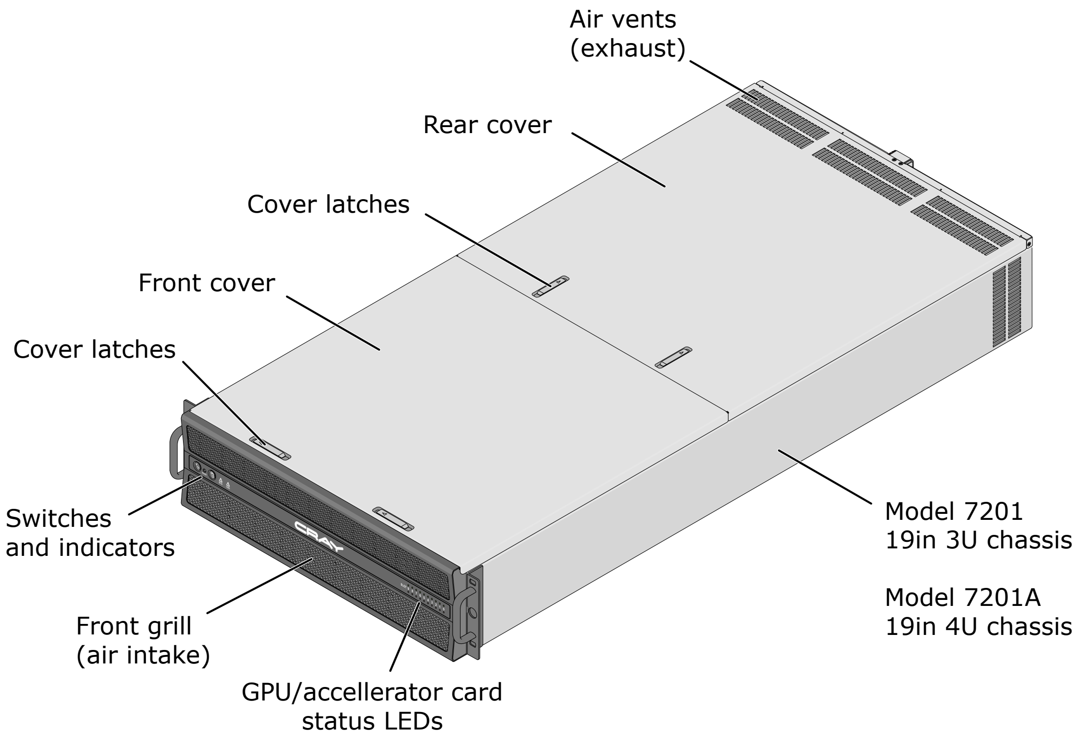
System Description

The CS-Storm™ 500GT system is a dense 3U or 4U 19-inch wide rackmount server that is optimized to support today's highest power GPU or FPGA accelerator cards.

Each 500GT server contains two Intel® Xeon® Scalable processors, up to 1536GB of memory, eight 2.5-in drive bays, and up to 16 DIMMs. However, for optimal memory performance, 12 DIMMs are recommended to achieve maximum performance.

Each CS-Storm 500GT server supports up to 10 PCIe GPU or FPGA accelerator cards.

Figure 1. CS-Storm 500GT Server



Server Configuration Options:

- **Balanced PCIe Configuration**
  - GPU host-to-peer optimized server.
  - Balanced PCIe CPU-to-GPU bandwidth. The balanced PCIe architecture offers balanced performance for codes that have high data parallelism and use both the CPUs and GPUs in workload processing.
- **Custom Accelerator Card Configuration**
  - A 4U chassis balanced PCIe server implements the same system PCIe architecture and hardware components but supports extended height custom-sized FPGA accelerator cards.

Table 2. CS-Storm 500GT Server Specifications

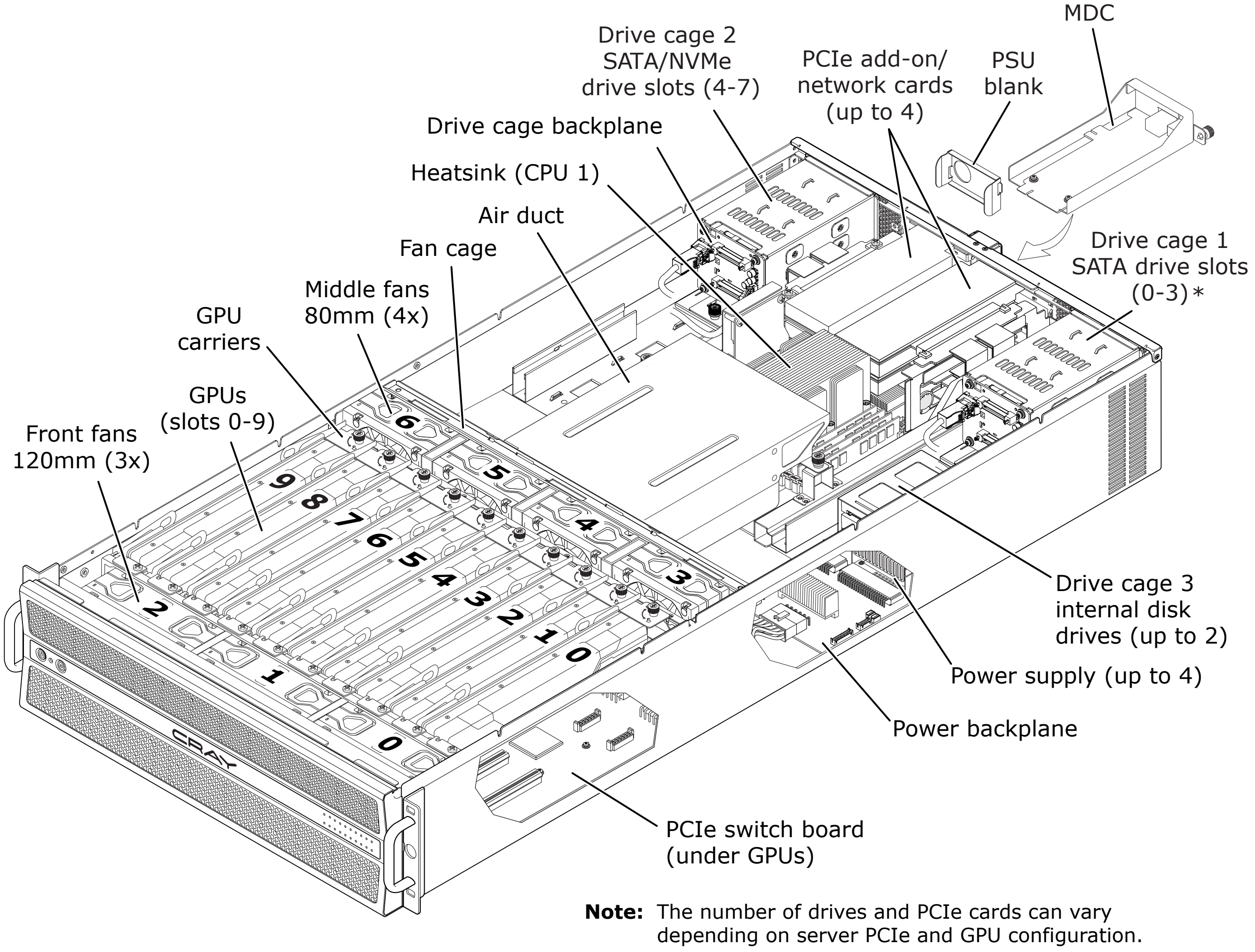
Feature	Description
Rack options	19in rack, 42RU and 48RU options
Chassis	● 19-inch wide, 3U or 4U rackmounted chassis ● Up to 15 server chassis in a 48RU rack ● Chassis weight: ○ Up to 76 lb (34kg) without PCIe cards ○ Up to 135 lb (62kg) fully loaded ● 3U Dimensions: (HxWxD) 5.1 x 17.7 x 36.4in (130 x 449 x 925mm) ● 4U Dimensions: (HxWxD) 6.8 x 17.7 x 36.4in (173 x 449 x 925mm)
Accelerators	Up to 10 PCIe accelerators (up to 400W continuous each): ● NVIDIA® Tesla® P40 or P100 ● NVIDIA® Tesla® V100 ● Custom extended height full-size FPGA accelerators (4U chassis only)
Custom 4U Chassis	● Up to eight 425W custom cards ● N+1 power supply redundancy ● Not ASHRAE compliant ● Balanced PCIe configuration
Motherboard	Intel® S2600BP
Processors	Two Intel Xeon Scalable family processors (up to 165W TDP)
Memory Capacity	Up to 12 of 16 available DIMM slots Up to 1536GB DDR4 (12 x 128GB DIMMs) For optimal memory performance, 12 DIMMs (1 DIMM per channel, 6 DIMMs per CPU) are highly recommended.
Storage	2.5in drive bays NVMe U.2 drive configuration depends on PCIe topology. Spinning disks are not supported (all SATA disks must be SSDs). ● Up to 8 SATA SSDs in external drive bays (hot swap) ● 4 NVMe SSDs (external bays 4-7) ● 1 or 2 fixed internal SATA SSDs Some configurations require an additional add-in storage controller Total number and type of drives vary with configuration and PCIe topology.
Expansion slots	● 2 PCIe 3.0 x16 slots

Feature	Description
	● 2 additional PCIe 3.0 x16 slots can be added with 8 GPUs
Network adapter cards	● Omni-Path (100Gb/s) ● InfiniBand™ EDR (100Gb/s) or HDR (200Gb/s) ● Ethernet (100Gb/s)
Cooling	Air cooled (front to rear air flow) ● Seven fans ○ Three 120mm fans (front) ○ Four 80mm fans (middle) ○ Active/manual fan speed control through MDC or `hydrad` daemon ● Built-in air duct ● Passive processor heatsinks ● Passive GPU/FPGA heatsinks ● Two in-line fans in each power supply unit
Power Supplies	Support for both N+1 and N+N power configurations. Up to four 2200W AC power supplies, 200-277VAC (gold level efficiency) ● 2+2 redundancy with 10 (300W) accelerators ● 3+1 redundancy with 10 (400W) accelerators (3+1 PSUs required) Server supports multiple PCIe topologies and configuration options
Node management	● Integrated Baseboard Management Controller (BMC) (IPMI 2.0) ● Management daughter card (MDC) ● MDC supports `hydrad` daemon to manage fans, GPUs, and PSUs ● Intel remote management module 4 (RMM4) ● RMM4 supports remote KVM and Intel Dedicated Server Management NIC ● On-board RJ45 management port ● Support for Intel System Management Software
IO Ports	● 2 RJ45 10GBase-T LAN ports ● 1 RJ45 dedicated management LAN port ● 2 USB 3.0 ports ● Optional: VGA or serial port

Server Components

The major components in the CS-Storm 500GT server are shown in the following figure.

Figure 2. CS-Storm 500GT 3U Chassis Components



Note: The number of drives and PCIe cards can vary depending on server PCIe and GPU configuration.

Fans and fan cage

There are 7 pluggable fans, each with a 4-pin connector on the bottom that plugs into a fan interface board (front and middle). The fan distribution boards provide an interface between the fans and the power backplane. The middle fan cage can be removed to provide access to other chassis components. Air flow runs from front to back.

PCIe switch board

The balanced four PCIe switch board (4 PLX) supports 10 PCIe slots for GPU or FPGA cards with the option of using slots 4 and 5 for network add-in cards. See *PCIe Architecture* on page 20. The switch board provides an interface between the motherboard and GPUs through the PLX device. The switch board also provides a direct power connection to each GPU slot. GPU status signals are routed to the front panel through the PCIe switches and over the SMBus.

Power backplane

All power supplies (PSUs) plug into the power backplane. The power backplane distributes power along with monitoring features to all printed circuit assemblies (PCAs) in the chassis including the motherboard, front panel controls and indicators, fans, and accelerators. The power backplane provides a control signal interface between the motherboard and front control panel. See *Power Distribution* on page 23.

Drive Cages and Disk Backplane

The drive cages support multiple local storage configuration options. See *Drive Support and Configuration* on page 17. Depending on the configuration, one or more of the drive cages may not be included in the chassis. If a rear-accessible drive cage is not included, a cover is used to fill the chassis opening.

The rear-accessible drive cages have a disk backplane that extends the motherboard SATA ports to the drives. The backplanes provide power for the drives and include separate cable connectors for SATA and NVMe drives. The disk backplane provides PCIe 3.0 x4 for each drive slot. Drive cage 1 can support NVMe drives with additional PCIe cables, if PCIe lanes are available.

The internal 2.5in drive cage does not have a backplane. Direct cable connections for power and SATA signals are used.

Management Daughter Card (MDC)

The MDC is used to configure, monitor, and manage server subsystems and components. Primary maintenance functions include fan and thermal monitoring and power consumption monitoring. See *Management Daughter Card (MDC)* on page 35.

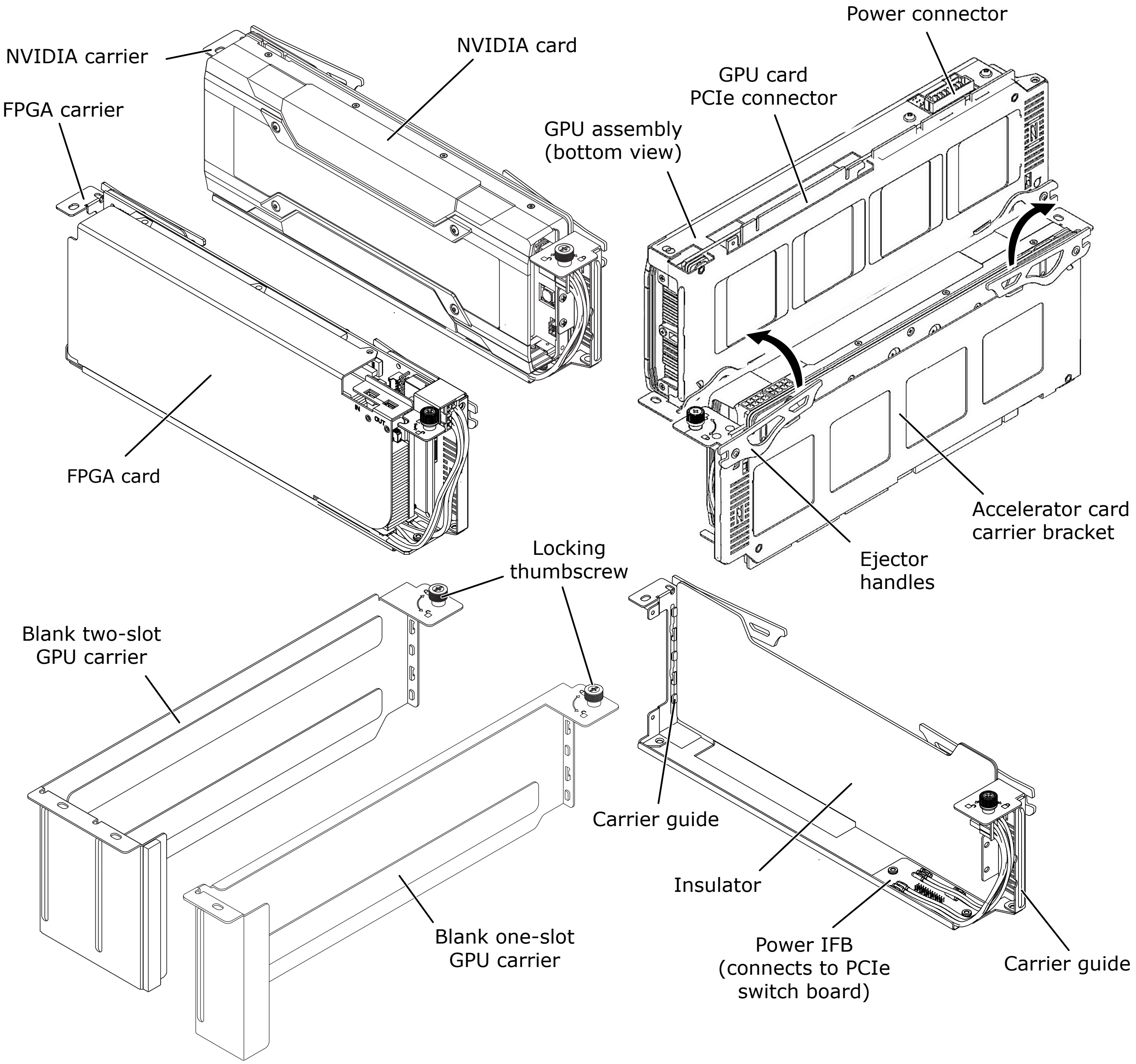
Air duct

The air duct provides proper air flow for the motherboard, DIMMs, and CPU heatsinks. The air duct is mounted behind the middle fan assembly. Always operate the CS-Storm 500GT with the air duct in place. The air duct is required for proper airflow within the server chassis.

Accelerator Carriers

Accelerator cards are mounted/screwed to a carrier frame. The accelerator card and carrier assembly is lowered down into the chassis and seated to the PCIe switch board. Guides on each end of the carrier fit into slots in the front and middle fan trays. Ejector handles are used to seat/unseat the assembly. A locking thumbscrew secures the assembly in place.

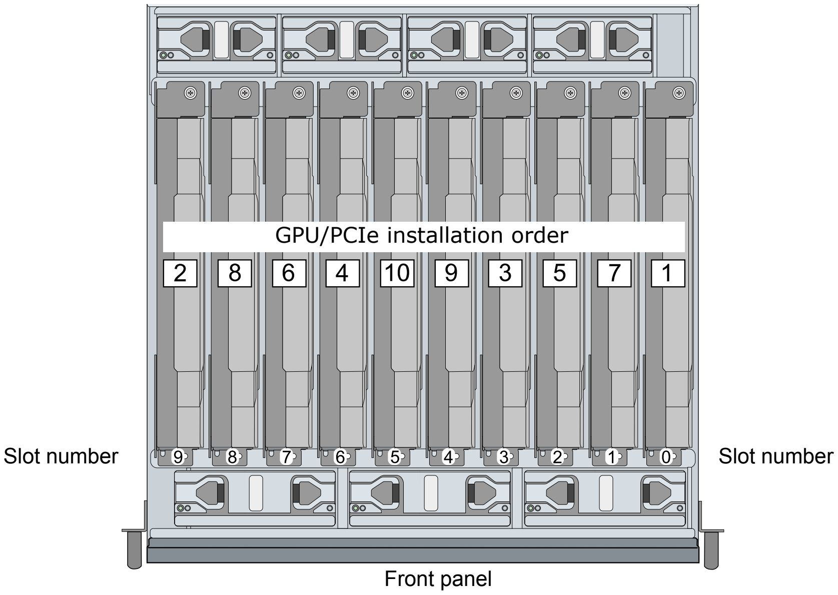
Figure 3. GPU Carriers, Multiple Views, CS-Storm 500GT



GPU Installation Order. The figure shows the order for installing GPUs if the chassis is not fully populated. This installation sequence must be followed to maintain a proper thermal environment. For example, for a 4 GPU configuration, slots 0, 9, 3, 6, are populated. Slots 2, 7, 1, 8, 4, and 5 would be empty.

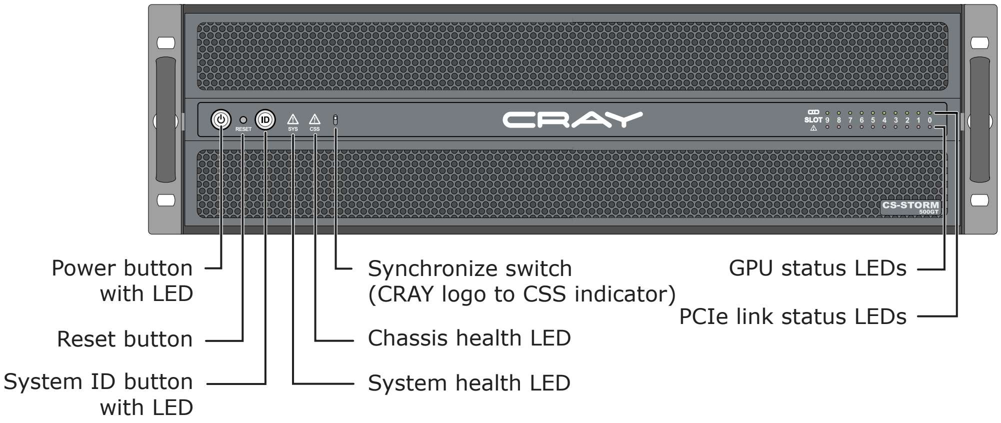
If all GPU slots are not fully populated, empty two-slot carriers are installed in the empty slots to maintain proper cooling.

Figure 4. CS-Storm 500GT GPU Installation Order



Controls and Indicators

Figure 5. Front Controls - CS-Storm 500GT



⏻ **Power button [blue]**. The power button LED lights blue to indicate system power is on. The power button is used to apply power to server components. Pressing the power button initiates a request to the Baseboard Management Controller (BMC) integrated into the motherboard, which forwards the request to the ACPI power states in the motherboard chip set. The power button is monitored by the BMC and does not directly control power on the power supplies.

○ RESET

Reset button. Press the Reset button to shut down, clear memory, and reset devices to their initialized state.

System ID button [white]. This LED lights white to visually identify a specific server within a rack/cabinet. The System ID button toggles the state of the LED. If the LED is off, pushing the System ID button lights the ID LED. It remains lit until the button is pushed again or until a chassis identify command is received to change the state of the LED.

Chassis health (CSS) LED [amber]. The chassis health LED indicates:

Off — Normal operation

Blinking — Fan, PSU, or SSD failure

Solid On — PCI, PLX, GPU, PSU, fan failures, SMBus errors

System health (SYS) LED [amber]. The system status LED indicates a fatal or non-fatal error in the system as reported through the BMC or by the management daughter card (MDC). The System Status LED is set to a steady amber color for all fatal errors that are detected during processor initialization. A steady amber color indicates that an unrecoverable system failure condition has occurred:

Off — Normal operation

Solid On — Fatal error

Blinking — Non-fatal error

Synchronize switch. This switch sets/synchronizes the CRAY logo to display the same conditions as the CSS LED. Synchronize is on (default) when the switch is in the up position, as shown.

CRAY logo [blue/amber].

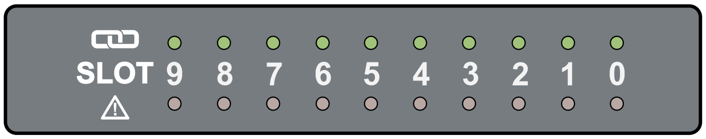
On (blue) — Server is powered on or server is powered off but power cable is connected to power. Unplug the server to remove all power.

On (amber) — Optional. Can be synchronized to have the same indications as the Chassis Health LED.

GPU and PCIe Status Indicators

The front panel has a series of LED status indicators that are used to quickly identify issues. The 20 LEDs on the right side provide GPU status from the PCIe switch board.

Figure 6. GPU and PCIe LEDs - CS-Storm 500GT



= **GPU Status LEDs**

Off — Normal operation.

Solid (red) — Fatal alarm. Indicates over temperature, over current, or communication error.

⚠ = **PCIe Status LEDs**

These LEDs indicate the status and transfer speed for the PCIe connection through the PCIe (PLX) switch to the GPU.

Off — No link or GPU is not detected.

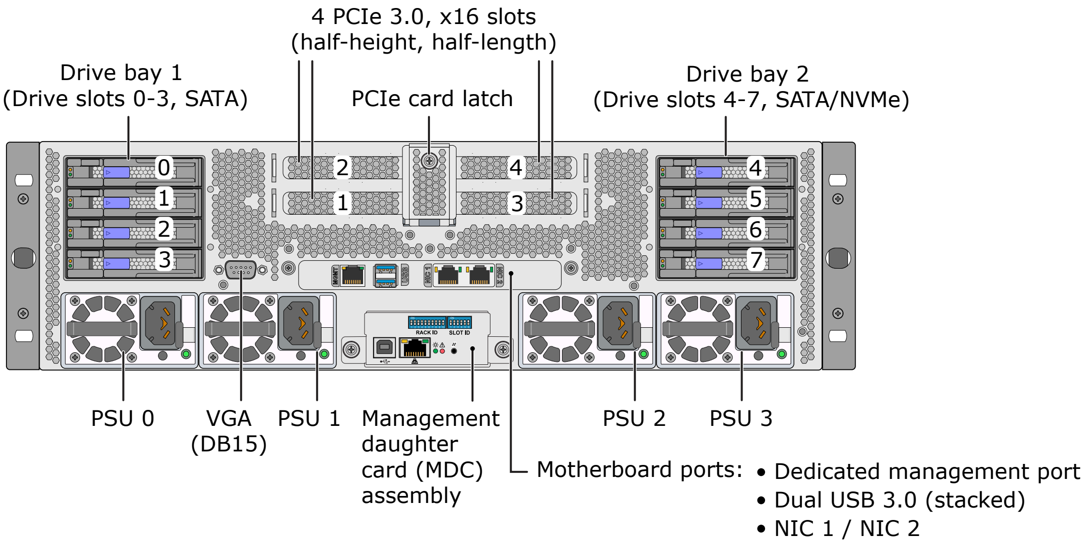
On (red) — PCIe link is up (8.0 GT/s - Gen3).

Blinking (2 Hz) — PCIe link is up (5.0 GT/s - Gen2).

Blinking (1 Hz) — PCIe link is up (2.5 GT/s - Gen1).

Rear PCIe Slots, I/O Connectors, and LEDs

Figure 7. Rear Controls and Connectors - CS-Storm 500GT



Note: The number of drives and PCIe cards can vary depending on server PCIe and GPU configuration.

Four PCIe 3.0 slots

The low-profile card slots support use of add-in cards with standard low-profile brackets. No customized brackets are required. The PCIe card release closes to secure the cards in place and opens to release tension so cards can be added/replaced. A thumbscrew secures the PCIe card release in place.

- Dual rail configuration (default) – PCIe 3.0 slots 1 and 2.
- Quad rail configuration (optional) – PCIe 3.0 slots 1-4. Additional slots 3 and 4 come from the center slots of the PCIe switch board through twin-axial (twin-ax) ribbon cable assemblies.

VGA (DB15) optional

The optional VGA port is implemented through a 12-pin ribbon cable connection to the motherboard (default).

Management Daughter Card (MDC)

The MDC is used to configure, monitor, and manage server subsystems and components. Primary maintenance functions include fan and thermal monitoring and power consumption monitoring. Refer to *Management Daughter Card (MDC)* on page 35 for details.

Drive Support and Configuration

The flexible design of the CS-Storm 500GT server enables numerous storage configuration options. The following two drive configurations are offered as standard options. Other drive configurations will be considered upon request:

- 8 SATA drives
  - Bay 1 - 4 SATA
  - Bay 2 - 4 SATA
- 4 SATA and 4 NVMe drives
  - Bay 1 - 4 SATA
  - Bay 2 - 4 NVMe (requires NVMe cables from slots 2 and 3)

NVMe Support

- Drive bay 2 supports up to 4 NVMe drives

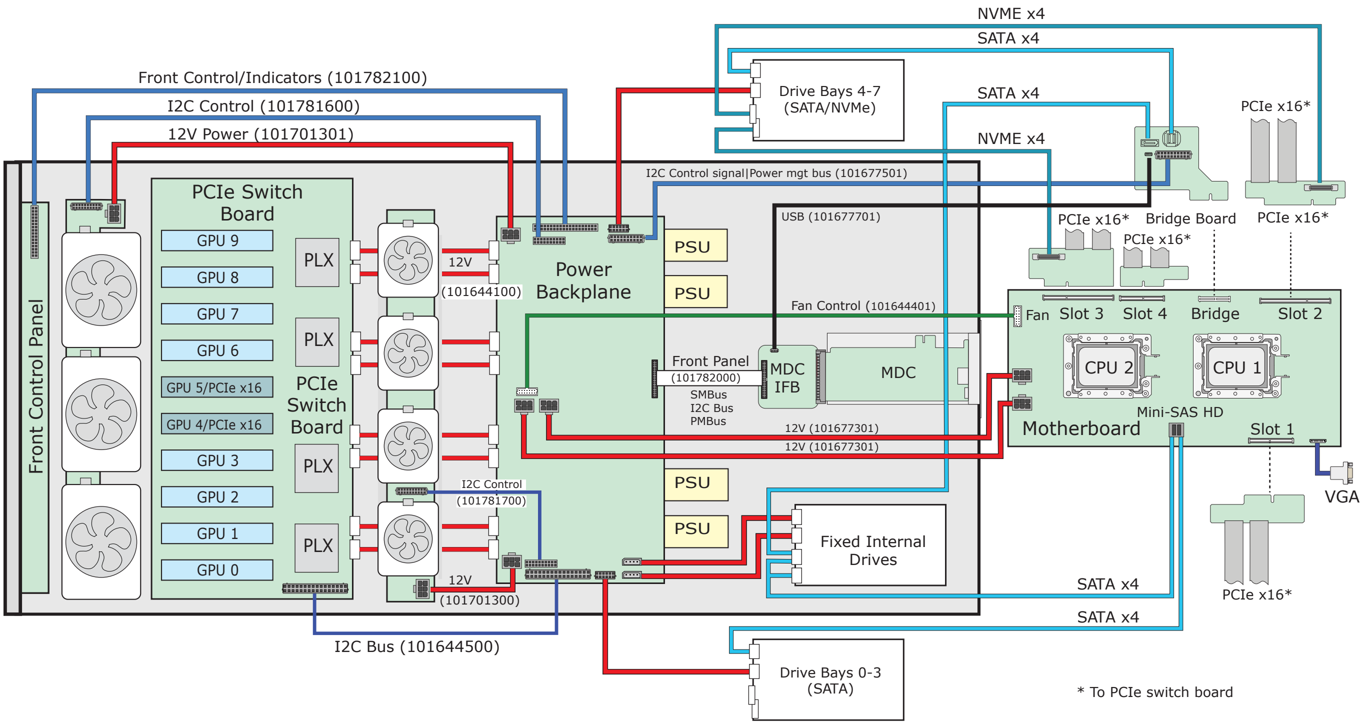
Internal SATA SSD Drives

- Up to 2 internal fixed SATA SSDs are supported, with different cabling and/or changes to drive support in bay 2.
- This drive bay does not have a disk backplane. The drives are cabled directly to power and SATA cables in the chassis.

System Interconnect Diagram

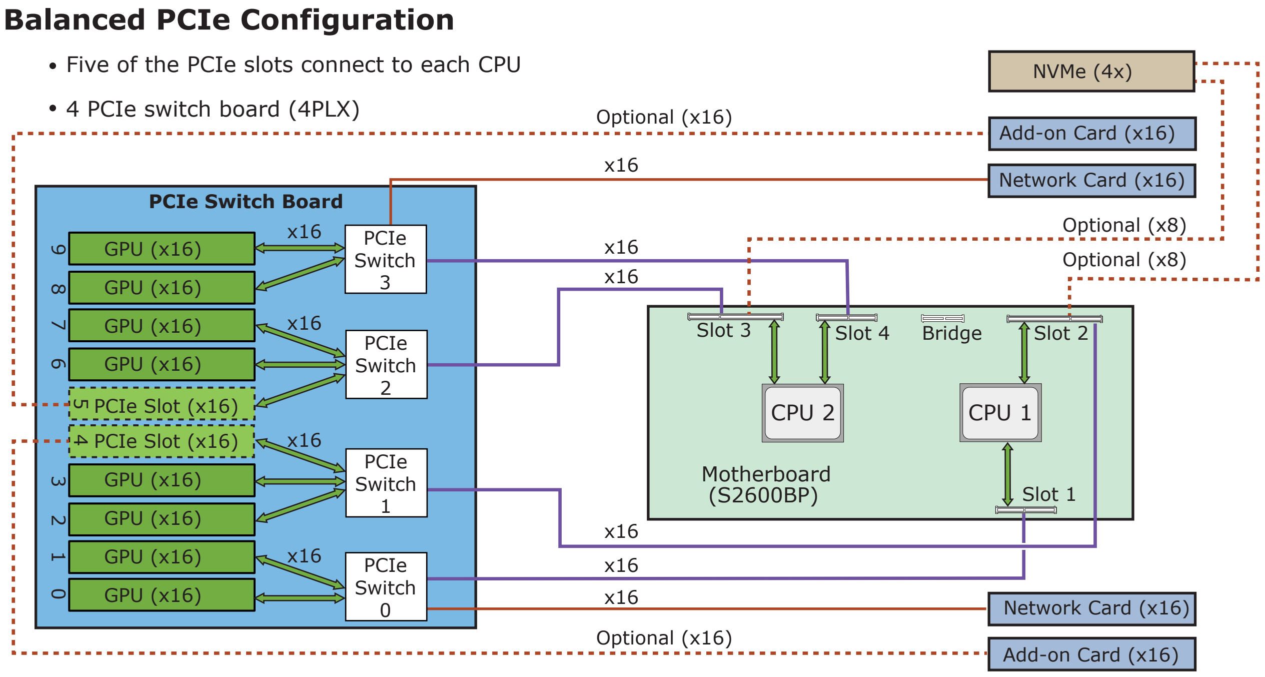
The figure shows the CS-Storm 500GT server interconnect and cable connections between each of the major subsystem components.

Figure 8. Balanced PCIe 4PLX Interconnect Diagram - CS-Storm 500GT



PCIe Architecture

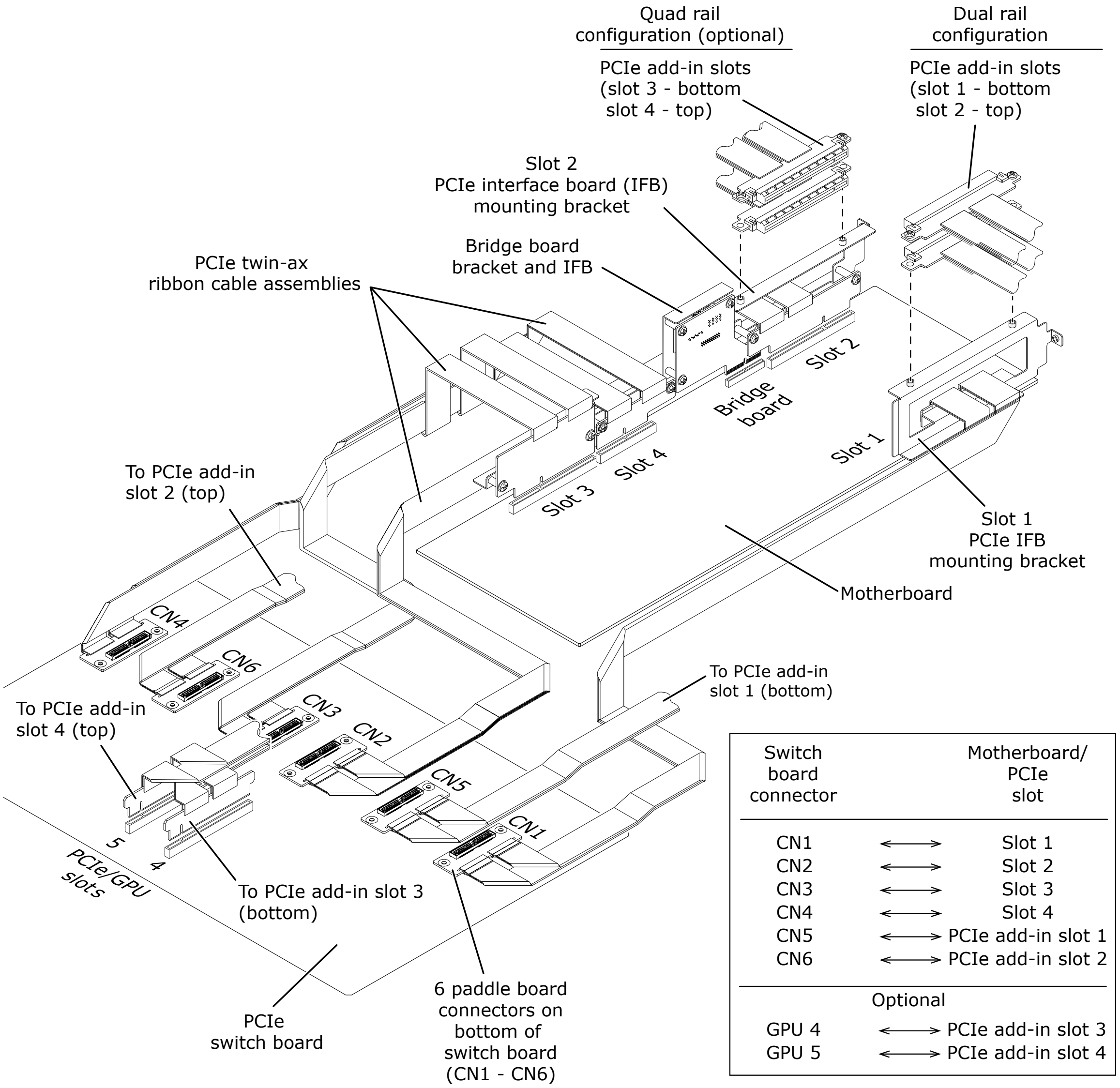
Figure 9. Balanced 4PLX PCIe Block Diagram - CS-Storm 500GT



PCIe Connections and Cabling

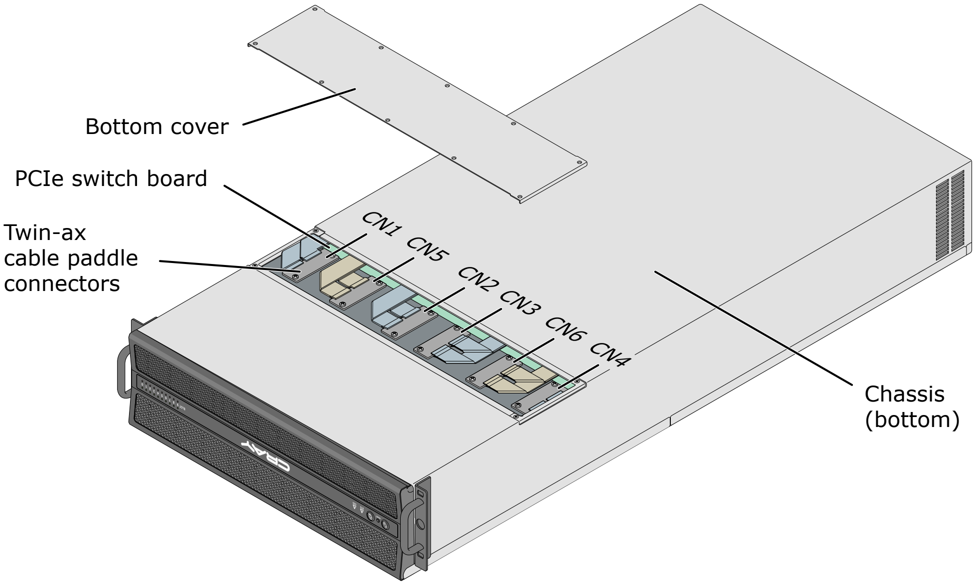
The CS-Storm 500GT supports up to four PCIe 3.0 x16 slots for high-speed network adapter cards. The PCIe lanes for HSN adapter cards come from the motherboard through the PCIe switch board and on to the add-in slot connectors through twin-axial ribbon cable assemblies. Two of the available 10 GPU slots are used to support the two additional PCIe slots in the quad-rail configuration.

Figure 10. PCIe Connections for Balanced Configuration - CS-Storm 500GT


Quad rail configuration (optional)
PCIe add-in slots (slot 3 - bottom slot 4 - top)
Dual rail configuration
PCIe add-in slots (slot 1 - bottom slot 2 - top)
Slot 2 PCIe interface board (IFB) mounting bracket
Bridge board bracket and IFB
PCIe twin-ax ribbon cable assemblies
Slot 2
Bridge board
Slot 4
Slot 3
Slot 1
Slot 1 PCIe IFB mounting bracket
Motherboard
To PCIe add-in slot 2 (top)
CN4
CN6
CN3
CN2
CN5
CN1
To PCIe add-in slot 4 (top)
To PCIe add-in slot 1 (bottom)
5
4
PCIe/GPU slots
To PCIe add-in slot 3 (bottom)
PCIe switch board
6 paddle board connectors on bottom of switch board (CN1 - CN6)

Switch board connector		Motherboard/ PCIe slot
CN1	⟷	Slot 1
CN2	⟷	Slot 2
CN3	⟷	Slot 3
CN4	⟷	Slot 4
CN5	⟷	PCIe add-in slot 1
CN6	⟷	PCIe add-in slot 2
	Optional	
GPU 4	⟷	PCIe add-in slot 3
GPU 5	⟷	PCIe add-in slot 4

Figure 11. PCIe Connectors Bottom View - CS-Storm 500GT



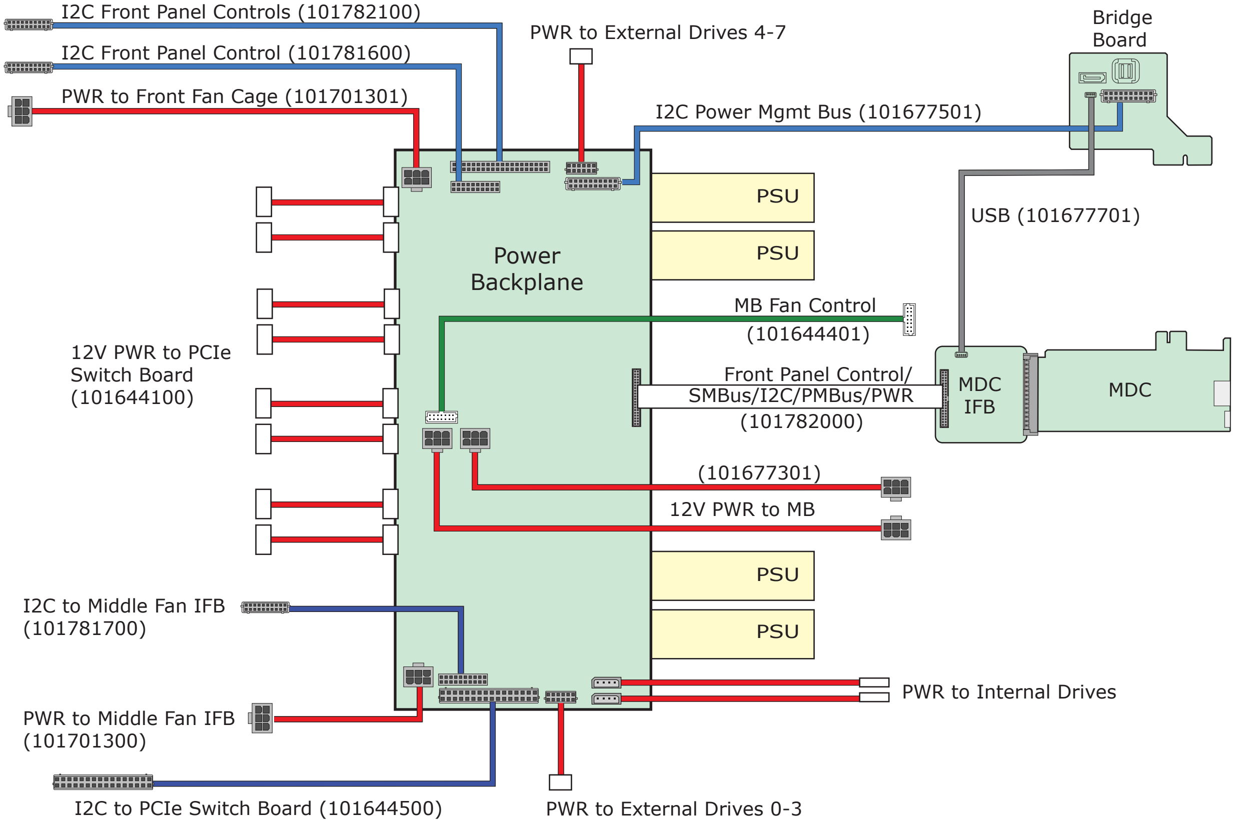
Power Distribution

CS-Storm 500GT system PDU choices may be based on data center facilities/requirements, customer preferences, and system/rack equipment configurations. Each 2200W power supply (PSU) in the server connects to a PDU outlet through a 1.5 m power cord.

Chassis Power Distribution

Up to 4 (N+1) power supplies in the chassis receive power from the rack PDU. The power supplies are installed in the rear of the chassis and distribute power to all the components in the chassis through the power backplane. The power backplane is located at the bottom of the chassis, below the motherboard plate assembly and air guide plate assembly. Up to four 2200W PSUs plug into the power backplane. The power backplane distributes power and monitors all printed circuit assemblies (PCAs) in the chassis including the motherboard, front panel controls and indicators, fans, and accelerators. The power backplane provides a control signal interface between the motherboard and front control panel.

Figure 12. Power Backplane Interconnect Diagram - CS-Storm 500GT


I2C Front Panel Controls (101782100)
I2C Front Panel Control (101781600)
PWR to Front Fan Cage (101701301)
PWR to External Drives 4-7
Bridge Board
I2C Power Mgmt Bus (101677501)
PSU
PSU
USB (101677701)
Power Backplane
MB Fan Control (101644401)
12V PWR to PCIe Switch Board (101644100)
Front Panel Control/ SMBus/I2C/PMBus/PWR (101782000)
MDC IFB
MDC
(101677301)
12V PWR to MB
PSU
PSU
I2C to Middle Fan IFB (101781700)
PWR to Internal Drives
PWR to Middle Fan IFB (101701300)
I2C to PCIe Switch Board (101644500)
PWR to External Drives 0-3

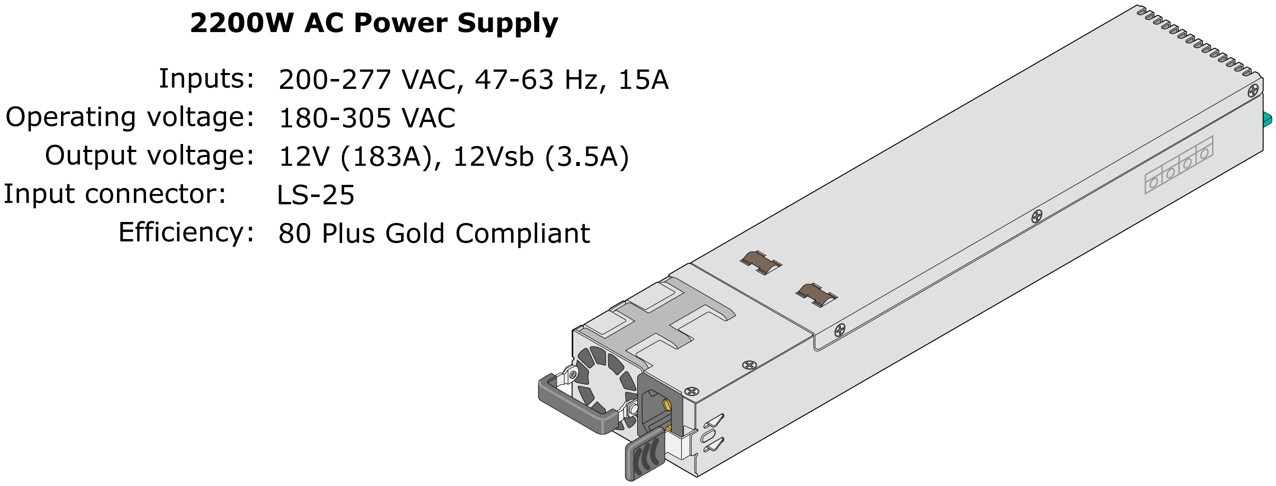
CS-Storm 500GT Power Supplies

The CS-Storm 500GT uses up to four 2200W high-efficiency power supplies to support different PCIe/GPU configurations. Each power supply receives power from a rack PDU. The PSUs support 2+1 or 2+2 redundancy with hot-swap capability and provide up to 4.4kW of power.

The power supplies support Power Management Bus (PMBus™) technology and are managed over this bus. The power supplies can receive 277VAC or 208VAC (200-277VAC input).

The 500GT balances the load on all available PSUs. If a PSU fails, the load is rebalanced across the remaining PSUs. If using only two PSUs, without redundancy, the PSUs should be placed in the first two slots with two PSU blank assemblies in slots 3 and 4.

Figure 13. 2200W Power Supply



2200W AC Power Supply

Inputs:	200-277 VAC, 47-63 Hz, 15A
Operating voltage:	180-305 VAC
Output voltage:	12V (183A), 12Vsb (3.5A)
Input connector:	LS-25
Efficiency:	80 Plus Gold Compliant

Hydra Fan Control Utility

The hydra fan control utility monitors and controls GPUs and fans in CS-Storm 500GT servers. This utility controls Cray designed PCIe expansion and fan control logic through the motherboard BMC. The utility runs as a Linux service daemon (`hydrad`) and is distributed as an RPM package.

Fan control utility (`hydrad`):

- Supports GPUs or custom accelerators
- Supports Intel® motherboards
- Provides active/manual fan control for GPUs with fan localization (left or right)
- Supports Red Hat Enterprise Linux (RHEL) 6 and 7
- Enables individual GPU power on/off
- Enables user-programmable fan control parameters
- Provides power data monitoring with energy counter for PSU, motherboard, and GPUs
- Provides direct access the GPU/MIC/FPGA devices via SMBus

The fan control utility RPM package includes the following:

`/usr/sbin/hydrad`

The `hydrad` daemon is the main part of the hydra utility and runs as a service daemon on Linux OS. It starts and stops by the init script at runlevel 3, 4, and 5. When the service starts, `hydrad` parses the `/etc/hydra.conf` file for runtime environment information, then identifies/discovers the motherboard BMC, GPU, and fan control hardware logic on the system. The service then monitors the GPU status and fan speed every second. The fan speed varies according to GPU temperature, or what is defined in `hydra.conf`.The hydrad service updates the data file `/var/tmp/hydra_self` whenever the GPU or fan status has changed.

`/usr/sbin/hydrad.sh`

This script is called by `/etc/rc.d/init.d/hydra` and invokes the `hydrad` service. It generates a `/tmp/hydrad.log` file.

`/usr/sbin/hydra`

The hydra fan utility provides following command line interface (CLI) to users.

- Show GPU status
- Control GPU power on/off
- Show fan status
- Set active/manual fan control mode
- Set fan speed under manual mode

`/etc/hydra.conf`

This file contains the running environment for the `hydrad` service. The running parameters for fan speed and GPU temperature can be adjusted on the system. Restart the `hydrad` service to apply changes made to the `hydra.conf` file.

RPM Package

After installing the hydra RPM package, the hydra utility automatically registers and starts up the `hydrad` daemon. To change parameters, modify the `/etc/hydra.conf` file, then stop and start `hydrad`.

Install:

```
# rpm -ihv ./hydra-1.4-0.x86_64.rpm
```

The `hydrad` service starts automatically during install and continues running as a service daemon unless the package is removed.

Remove:

```
# rpm -e hydra
```

The `/etc/hydra.conf` file is moved to `/etc/hydra.conf.rpmsave` for the next installation.

Data File

A data file, `/var/tmp/hydra_self`, is created when `hydrad` starts. It contains all GPU and fan information that `hydrad` collects. Both `hydrad` and `hydra` use this data file to monitor and control the system. This file can be used as a snapshot image of the latest system status.

Configuration Parameters

The `hydrad` runtime environment is modified using the `/etc/hydra.conf` configuration file. Use the `hydra config` command to display/verify the current GPU environment settings. Modify the `/etc/hydrad.conf` then restart the `hydrad` service. The `/etc/hydra.conf` file contains following parameters:

- `activefan` (`on`, `off`, default is `on`). Selects active or manual fan control mode
- `debug` (`on`, `off`, default is `off`). When this option is set to on, hydrad outputs debug messages to `/tmp/hydrad.log`.
- `discover` (`on`, `off`, default is `on`). `hydrad` responds if there is a broadcast packet issued from `hscan.py` on the network UDP 38067 port.
- `fanhigh` (`fannormal` − 100%, default is 85%). The PWM duty value of high speed. If the GPU maximum temperature is higher than `hightemp`, the fan speed is set by this high duty value. The default setting is full speed.
- `fanlow` (5% - `fannormal`, default is 10%). The pulse-width modulation (PWM) duty value for low speed. If the GPUs maximum temperature is lower than `normaltemp`, the fan speed is set according to this low duty value. The default value 10%, set for the idled state of the GPU, which reduces fan power consumption.
- `fannormal` (`fanlow` − `fanhigh`, default is 65%). The PWM duty value of normal fan speed. If the GPU maximum temperature is higher than normaltemp, and lower than `hightemp`, the fan speed is set run by this normal duty value.
- `fanspeed` (5 - 100%, default is 85%). The default fan speed after you set manual fan control mode.

- ⚠ **CAUTION:**
  - **GPU Overheating**
  - Manually setting the default fan speed to low can overheat the GPUs. Monitor GPU temperature after manually setting the fan speed to avoid damage to the GPU or accelerator card.
- gpuhealth (on, off, default is on). Set gpuhealth to off to disable the GPU monitoring function if GPUs are not installed in the system
- gpumax (0°C - 127°C, default is 90°C). The maximum GPU temperature allowed. If a GPU exceeds the gpumax value, hydrad issues an event in the event log. Set the proper gpumax temperature for the type of GPU installed in the system.
- gpu_type (auto, K10, K20, K40, K80, MIC, default is auto). You can define the type of your GPU/MIC. If you set auto, hydrad will automatically detect the type of GPU (requires additional time).
- hightemp (normaltemp - 127°C, default is 75°C). The minimum temperature where the fan runs at high speed. If a GPU exceeds this high temperature value, the fan runs at high speed.
- login_node (on, off, default is off). When this option is set to on, hydrad operates for a login or I/O node.
- loglevel (info, warning, critical, default is info). Controls what events hydrad logs to the /tmp/hydrad.log file
- nodepower (on, off, default is off). hydrad monitors motherboard power consumption.
- normaltemp (0°C - hightemp, default is 60°C). The minimum temperature where the fan runs at normal speed. If a GPU temperature exceeds the normal temperature value, the fan runs at normal speed.
- polling (1 - 100 seconds, default is 2 seconds). Controls how often hydrad service accesses the GPU and fan controller
- psu_health (on, off, default is off). hydrad monitors GPU power consumption.
- psupower (on, off, default is on). hydrad checks and monitors power status and consumption of the three PSUs.
- sysloglevel (info, warning, critical, default is warning). The hydrad service also supports the syslog facility using this log level. hydrad event logs are written to /var/log/messages.

⚠ **CAUTION:**

- **GPU Overheating**
- Manually setting the default fan speed to low can overheat the GPU. Monitor GPU temperature after manually setting the fan speed to avoid damage to the GPU or accelerator.

hydra Commands

To start the fan control service:

```
# service hydra start
```

To stop the fan control service:

```
# service hydra stop
```

Fan control utility settings are controlled from the /etc/hydra.conf configuration file when hydrad is started.

To disable or enable active fan control:

```
# hydra fan [on|off]
on:  Active Fan Control by GPU temperature
off: Manual Fan Control
```

To set manual fan control to a specific PWM duty value (% = 10 to 100):

```
# hydra fan off
# hydra fan [%]
```

Command line options (examples shown below):

```
# hydra
Usage: hydra [options] <command>
Options:
   - D            :display debug message
   - f <file>     :use specific hydrad data file. default: /var/tmp/hydra_self
   - v            :display hydra version
Commands:
   config         :display running hydrad settings
   gpu [on|off]   :display or control GPU power
   node           :display node status
   sensor         :display GPU temperatures
   fan [%|on|off] :display fan status, set duty cycle, active control, manual
control
   power [node|gpu|clear] :display PSU, motherboard and GPU power status or reset
the energy counter
```

hydra config: Display Configuration Parameters

The `hydra config` command displays parameter values that the hydra service is currently using. Values can be changed in the `/etc/hydrad.conf` file and implemented by stopping and starting the hydra service.

```
# hydra config
uid=0
cid=0
id=0
gpu_map=00000000
gpu_type=auto
normaltemp=60
hightemp=75
gpumax=90
fanspeed=100
low=50
normal=80
high=100
polling=2
loglevel=info
sysloglevel=warning
activefan=on
gpu_health=on
psu_health=on
nodepower=off
gpupower=off
login_node=off
debug=off
ok
[root@hydra3]#
```

hydra gpu: GPU Power Control

The CS-Storm has power control logic for the all GPUs that can be controlled using a hydrad CLI command. GPU power can be disabled to reduce power consumption. The default initial power state for GPUs is power on. If the GPU power is off, the GPU is not powered on when powered on, unless GPU power is enabled using the CLI command.

The following limitations exist for GPU power control:

- The OS may crash if GPU power is set to off while the operating system is active due to the disabled PCI link.
- Reboot the operating system after enabling power to a GPU so that the GPU is recognized.

Show the GPU status or on/off the GPU power. The power operation is performed for all installed GPUs. Individual GPU control is not allowed. Status information includes Bus number, PCI slot, Mux, power status, GPU Type, Product ID, firmware version, GPU slave address, temperature, and status. Use the following commands to enable or disable power to the GPUs.

```
Args:
  <non>: Display GPU status: Bus(PCI#,Mux), Power, Type, Product ID, FWVer for
MIC, GPU slave address, Temperature and Status.
  on   : Turn on the all GPU power.
  off  : Turn off the all GPU power.

[root@hydra3]# hydra gpu
#  Slot:Mux Loc   Power Type   PID  FWVer Addr Temp  Min  Max  Status
      0  PCI1:1   R/R/B on    Pascal 15f8 -     9eH    29   29   29  ok
      1  PCI1:2   R/R/T on    Pascal 15f8 -     9eH    31   31   31  ok
      2  PCI2:1   R/L/B on    Pascal 15f8 -     9eH    32   32   32  ok
      3  PCI2:2   R/L/T on    Pascal 15f8 -     9eH    30   30   30  ok
      4  PCI3:1   F/L/B on    Pascal 15f8 -     9eH    31   31   31  ok
      5  PCI3:2   F/L/T on    Pascal 15f8 -     9eH    30   30   30  ok
      6  PCI4:1   F/R/B on    Pascal 15f8 -     9eH    32   32   32  ok
      7  PCI4:2   F/R/T on    Pascal 15f8 -     9eH    32   32   32  ok
ok
# hydra gpu off
ok
# hydra gpu on
ok
#
```

hydra node: Motherboard BMC Status

The `hydra node` command displays motherboard BMC status, Product ID, BMC firmware version and IP settings.

```
# hydra node
Prod-ID: 004e
BMC Ver: 1.20
BMC CH1: 00:1e:67:76:4e:91
 ipaddr: 192.168.1.57
netmask: 255.255.255.0
gateway: 192.168.1.254
BMC CH2: 00:1e:67:76:4e:92
 ipaddr: 0.0.0.0
netmask: 0.0.0.0
gateway: 0.0.0.0
Sensors: 4
       p1_margin: ok ( -49.0 'C)
       p2_margin: ok ( -55.0 'C)
           inlet: ok (  31.0 'C)
          outlet: ok (  45.0 'C)
ok
#
```

hydra fan: Display Fan Status and Set Control Mode

The `hydrad fan` command displays fan status and changes fan control mode and speed. When active fan control is disabled, the fan speed is automatically set to the default manual fan speed. Use the `hydrad fan` command to display controller chip revision, slave address, control mode and fan status.

```
Args:
  <none>: Display FAN status: Chip Rev, slave addr, control mode and FAN status.
  on    : Set Active Fan control mode.
  off   : Set Manual Fan control mode.
  %     : Set FAN speed duty. 5-100(%)

[root@hydra3]# hydra fan
ADT7462 Rev : 04h
ADT7462 Addr: b0h
Active Fan  : on
Fan Stat RPM   Duty
FAN1 ok  9591  50
FAN2 ok  9574  50
FAN3 ok  9574  50
FAN4 ok  9574  50
Ok
```

Set fan control mode to manual:

```
# hydra fan off
ok
# hydra fan
ADT7462 Rev : 04h
ADT7462 Addr: b0h
Active Fan  : off
Fan Stat  RPM   Duty
FAN1 ok   13300 100
FAN2 ok   12980 100
FAN3 ok   13106 100
FAN4 ok   13466 100
Ok
```

Set fan duty cycle to 70%:

```
# hydra fan 70
ok
# hydra fan
ADT7462 Rev : 04h
ADT7462 Addr: b0h
Active Fan  : off
Fan Stat  RPM    Duty
FAN1 ok   12356  70
FAN2 ok   12300  70
FAN3 ok   12300  70
FAN4 ok   12244  70
Ok
```

Set fan control mode to active.

```
# hydra fan on
Ok
```

hydra sensor: Display GPU Temperatures

The `hydra sensor` command displays GPU temperatures

```
# hydra sensor
PCI1-A    PCI1-B    PCI2-A    PCI2-B    PCI3-A    PCI3-B    PCI4-A    PCI4-B
31        33        32        33        31         32
ok
[root@hydra3 ~]#
```

hydra power: Display Power Values

The `hydra power` command displays PSU, motherboard and GPU power status and can be used to reset the peak/average and energy counters.

```

Args:
  <none>: Display PSU power status
  node  : Display Motherboard power status
  gpu   : Display GPU power status
  clear : Reset all Peak/Average and Energy Counters

# hydra power
No Pwr  Stat Temp  Fan1  Fan2  +12V Curr ACIn Watt Model
00 on   ok     27  7776  6656  11.9   42  207  572 PSSH16220 H
01 on   ok     27  6144  5248  11.9   43  207  572 PSSH16220 H
02 -
Power : 84.0 A 1122 W  (Peak 1226 W, Average 1129 W)
Energy: 3457.5 Wh in last 11013secs(3h 3m 33s)
ok

# hydra power node
PMDev : ADM1276-3 0 (ok) p: 368.0 a: 187.6
Power : 12.2 V 10.5 A 193 W  (Peak 228 W, Average 188 W)
Energy: 576.1 Wh in last 11011secs(3h 3m 31s)
ok

# hydra power gpu
No Slot Stat  +12V  Curr  Watt  Peak  Avrg Model
1  PCI1 ok    12.2  20.0 366.5 495.0 367.8 ADM1276-3 0
2  PCI2 ok    12.2  20.8 386.9 485.2 387.7 ADM1276-3 0
3  PCI3 ok    12.1  20.0 365.5 480.6 364.3 ADM1276-3 0
4  PCI4 ok    12.2  18.4 339.3 483.2 340.6 ADM1276-3 0
Power : 78.9 A 1450 W  (Peak 1534 W, Average 1407 W)
Energy: 4310.9 Wh in last 11019secs(3h 3m 39s)
ok

# hydra power clear
ok

# hydra power
No Pwr  Stat Temp  Fan1  Fan2  +12V Curr ACIn Watt Model
00 on   ok     27  7776  6656  11.9   42  207  560 PSSH16220 H
01 on   ok     27  6144  5248  11.9   42  207  558 PSSH16220 H
02 -
```

```
Power : 84.0 A 1118 W  (Peak 1118 W, Average 1129 W)
Energy: 1.9 Wh in last 1secs(0h 0m 1s)
ok
#
```

Fan Speeds by GPU Temperature

As described above, fan speeds increase and decrease based on GPU termperatures. If one of GPU gets hot and exceeds the next temperature region, `hydrad` immediately changes the fan speed to reach target speed. As the GPU gets back to a low temperature below the region, `hydrad` will decrease the fan speed step by step.

```
         Duty %
   fanhigh |----------------------------
           |                # ^
           |               /  |
           |              L   |
 fannormal |---------+=======>+---------
           |       / ^
           |      /  |
           |     /   |
           |    /    |
           |   L     |
    fanlow |========>+------------------
           |
           +---------------------------- Temperature 'C
                 normaltemp  hightemp
```

GPU and Fan Localization

Each group of fans is controlled independently. The GPU temperature for a group does not affect the other group's fan speed. The fan speeds are determined by the GPUs within the same group.

No Power or Unknown GPU States

If there is no power or the GPU state is unknown, `hydrad` sets the fans speeds to either:

- Idle Speed (10%), if all of GPUs are powered off
- Full Speed (100%), if one GPU is unidentified or in an abnormal state (no thermal status reported for example)

Fan Control Watchdog Timeout Condition

The system includes hardware watchdog timeout logic to protect the GPUs from overheating in the event `hydrad` malfunctions. The fan speed is set to full speed after 5-10 seconds if any of the following conditions occur:

- System crash
- BMC crash
- `hydrad` crash
- `hydrad` service is stopped
- If the hydra fan utility package is removed

Discover Utility

A discovery utility (hscan.py) identifies all systems/nodes that are running hydrad. The hscan.py utility provides the following information from hydrad. (hydrad contains the internal identification/discovery service and provides information through UDP port 38067.) You can turn off the discover capability using the discover=off option in the hydra.conf file for each system.

- system: IP address, MAC of eth0, hostname, node type, hydrad version
- gpu: GPU temperature and type
- fan: fan status, pwm (L/R) and running speed (RPM)
- power: PSU, node, GPU power status

If hydra is not running on the system, hscan.py will not display any information even though the system is running and online.

> **NOTE:** The fan, power and temperature information can not be displayed together. So the -T, -S, -P, -N, -G and -F options can not be combined.

```
Usage:  ./hscan.py [options] <command>
Options:
      -h        : display this message
      -w <time> : waiting time for response packet
      -i <nic>  : specific ethernet IF port (eth0, eth1,...)
      -m        : display Mac address
      -c        : display Current IP address
      -l        : display system Location (cid-uid)
      -n        : display host Name
      -t        : display Node type
      -v        : display hydrad Version
      -d        : display hydrad Date
      -F        : display Fan status
      -T        : display gpu Temperature
      -S        : display pci Slot devices
      -P        : display PSU Power status
      -N        : display Node Power status
      -G        : display GPU Power status
```

Getting hscan.py

The hscan.py binary file is located /usr/sbin/hscan.py after installing the RPM package.

```
# rpm -ihv hydra-1.4-0.x86_64.rpm
Preparing...                ##################################### [100%]
1:   hydra                  ##################################### [100%]
     chkconfig --level 345 hydra on
     Starting hydra: [  OK  ]

# which hscan.py
/usr/sbin/hscan.py#
```

Copy the hscan.py binary file to a directory in the default path, or the ccshome directory.

```
$ scp root@s074:/usr/sbin/hscan.py .
hscan.py 100% 7764     7.6KB/s  00:00
$
```

Option Handling

Options for the discovery utility are displayed in the order they are entered:

```
$ ./hscan.py -mcn
00:1e:67:56:11:fd 192.168.100.74 sona

$ ./hscan.py -ncm
sona 192.168.100.74 00:1e:67:56:11:fd
$
```

System Information

When you run `./hscan.py` without option, each hydrad displays basic system information to your command window.

```
$ ./hscan.py
0-0 cn 00:1e:67:56:11:fd 192.168.100.74 v1.0rc3(Sep/23/20140) sona
$
```

GPU Information

GPU information cannot be displayed with fan information on the same command line. The `-G`, and `-P` options display GPU information.

```
$ ./hscan.py -lcG
00-0 192.168.100.74 1A:0 1B:0 2A:29 2B:28 3A:25 3B:26 4A:0 4B:0 01-0

$ ./hscan.py -lcP
00000-00 192.168.100.74 1A:- 1B:- 2A:K40 2B:K40 3A:K20 3B:K20 4A:- 4B:-
```

Fan Information

Fan information cannot be displayed with GPU information on the same command line. The `-F` option, displays fan information.

```
$ ./hscan.py -lcF
00000-00 192.168.1 A5h 10% 10% 0(bad) 5115(ok) 4631(ok) 0(bad)
```

Power Information

If you enter any one of the `-P`, `-N`, or `-T` options, `./hscan.py` displays PSU, Node and GPU power information.

```
$ ./hscan.py -lcP
00000-00 192.168.100.106 PSU 74A 1026W (1086/1014) 251Wh-14m
00000-00 192.168.100.19 PSU 42A 620W (1250/667) 160Wh-14m

$ ./hscan.py -lcN
00000-00 192.168.100.106 Node 16A 319W (332/311) 77Wh-14m
00000-00 192.168.100.19 Node 9A 185W (204/186) 45Wh-14m

$ ./hscan.py -lcG
00000-00 192.168.100.106 GPU 59A 1130W (1228/1125) 282Wh-15m
00000-00 192.168.100.19 GPU 33A 634W (1564/696) 171Wh-14m
$
```

Management Daughter Card (MDC)

The CS-Storm 500GT MDC configures, monitors, and manages server subsystems and components. The primary functions of the card are:

- Automatic detection of GPUs, PSUs, and HDD/SSDs
- Component management
- Temperature monitoring
- Power/energy consumption
- Power supply voltage/current monitoring
- Fan status/speed/control
- GPU/PCIe status
- Command line interface (CLI)
- IPMI support

The MDC reports complete system health. From a remote client, an administrator can easily use the MDC to monitor the state of the server. The MDC provides a CLI that can be used for server management. No CDs or additional installation steps are required to use the CLI.

The MDC can be inserted and removed during operation (hot-swapped).

MDC Control Panel

Figure 14. MDC Control Panel - CS-Storm 500GT

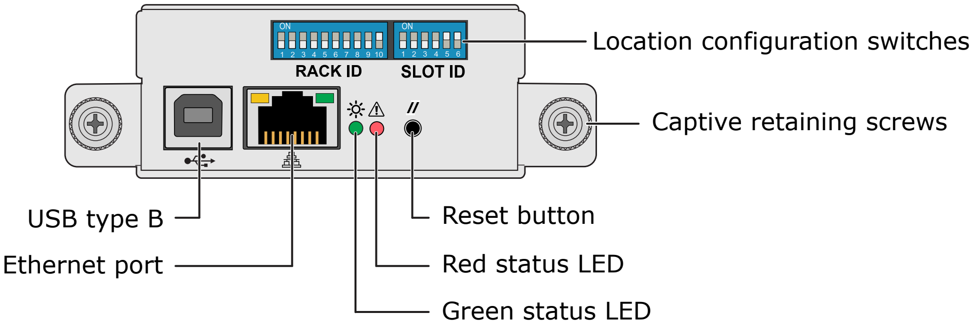


Table 3. CS-Storm 500GT MDC Control Panel Status LEDs

Green Status LED	Steady	Preparing MDC service
	Blinking† (0.5 sec)	MDC service is running
Red Status	Steady	Error during MDC boot-up sequence
	Blinking (0.2 sec)	Internal bus error

	Blinking (0.5 sec)	Node/fan control error
Green and Red Status LEDs	Steady	Cold start then could not boot system
	Blinking (0.5 sec)	Health of the server is abnormal
	Winking† (0.2 sec)	Firmware flashing in progress
	Winking (0.5 sec)	Initializing resources
Reset Button	Pressing the Reset button discharges power from the MDC assembly and initiates a start-up process. Pressing the Reset button does not affect components in the server.	
USB type-B Port	Provided for remotely flashing MDC firmware.	
Ethernet Port	Enables remote access to the MDC through an IP address (default address or configured by IDSW2-1). Refer to the "MDC DIP Switch Configuration" section. This is a standard full/half-duplex, 10/100 base-T Ethernet port with link and activity LEDs.	

† **Blinking**: LED is On and Off for equal amount of time.

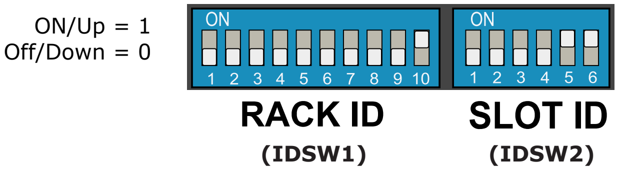
Winking: LED is On for a short period of time and Off for a longer period.

MDC DIP Switch Configuration

There are two DIP switches on the front panel of the CS-Storm 500GT MDC assembly. These two switches are used to configure the location of the server within the computer system (rack and slot) and to define the source of the IP address used by the MDC.

There are three DIP switches on the main board of the MDC assembly. These switches are used for defining the boot mode source and device, flashing firmware (volume production), and printing log messages (for debugging).

Figure 15. MDC DIP Switches - CS-Storm 500GT



IDSW1

1	2	3	4	5	6	7	8	9	10	Rack ID
0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	0	1	0	2
0	0	0	0	0	0	0	0	1	1	3
0	0	0	0	0	0	0	1	0	0	4
0	0	0	0	0	0	0	1	0	1	5
0	0	0	0	0	0	0	1	1	0	6
0	0	0	0	0	0	0	1	1	1	7
0	0	0	0	0	0	1	0	0	0	8
0	0	0	0	0	0	1	0	0	1	9
0	0	0	0	0	0	1	0	1	0	10
…	…	…	…	…	…	…	…	…	…	…
1	1	1	1	1	1	1	1	1	1	1023

IDSW2

1	IP Address Type
0	DHCP Default
1	Static

2	3	4	5	6	Slot ID
0	0	0	0	1	1
0	0	0	1	0	2
0	0	0	1	1	3
0	0	1	0	0	4
0	0	1	0	1	5
0	0	1	1	0	6
0	0	1	1	1	7
0	1	0	0	0	8
0	1	0	0	1	9
0	1	0	1	0	10
0	1	0	1	1	11
0	1	1	0	0	12
…	…	…	…	…	…

Figure 16. MDC Internal DIP Switches - CS-Storm 500GT

SW2
Boot Mode Settings

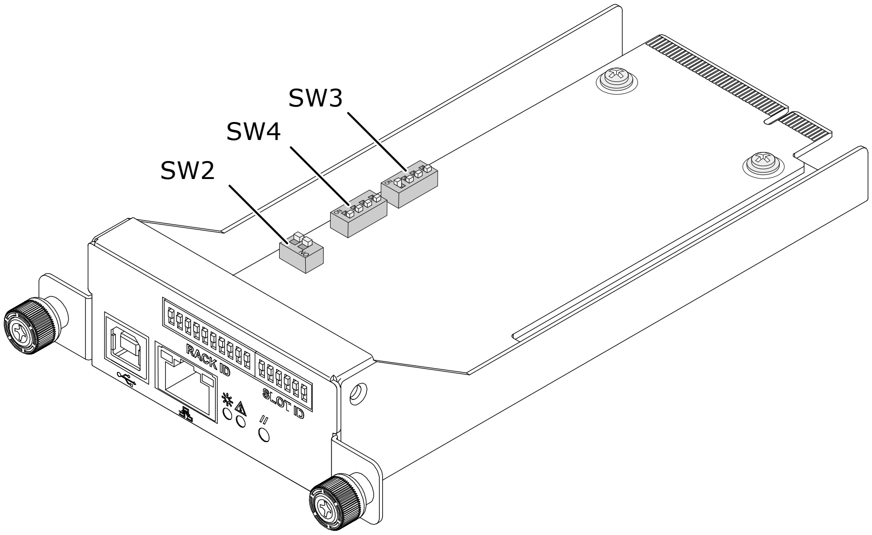
2-1	2-2	Boot Mode
Off	Off	Default. Internal boot
Off	On	Serial boot over USB on-the-go (OTG) [USB connected host]

SW3
Internal Boot Device

3-1	3-2	3-3	3-4	Internal Boot Device
On	Off	Off	Off	Default. NAND flash (256 MB)
Off	Off	On	On	SPI EEPROM (16 Mb)
*	*	*	*	N/A

SW4
Option Switches

4-1	4-2	4-3	4-4	Option
Off	Off	Off	Off	Default. Normal operation.
On	Off	Off	Off	Debug mode. MDC will log debug messages into /tmp/mdcd.log. No password validation.
Off	**On**	Off	Off	Diagnostic Mode. Not yet implemented.
Off	Off	**On**	Off	Clear NVRAM on the power backplane.
Off	Off	Off	**On**	N/A.

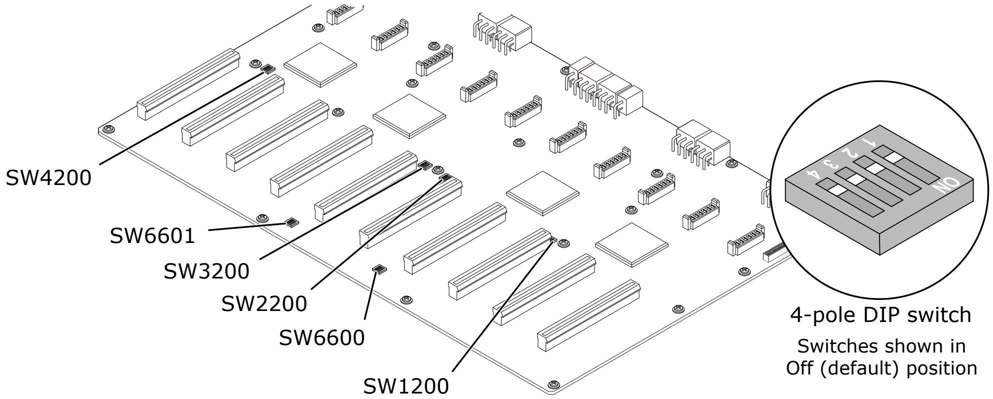


PCIe Bifurcation of the 4 PCIe Switch Board

This feature is reserved for custom FPGA configurations.

These DIP switches are all set to Off (default) for all other configurations. The CS-Storm 500GT PCIe PLX switch board has four DIP switches that are used to configure bifurcation of PCIe lanes out of each PCIe switch chip (PEX8796). The figure shows the location and function of each 4-pole DIP switch and its function.

Figure 17. 4PLX Switch Board Bifurcating Switches - CS-Storm 500GT



DIP Switch	Switch Positions				Function			
	1	2	3	4				
SW1200	Off	Off	Off	Off	Default	Slot 0: x16	Slot 1: x16	
	Off	Off	**On**	**On**	Bifurcated Slots 0:1	Slot 0: x8x8	Slot 1: x8x8	
SW2200	Off	Off	Off	Off	Default	Slot 2: x16	Slot 3: x16	Slot 4: x16
	On	Off	**On**	**On**	Bifurcated Slots 2:3:4	Slot 2: x8x8	Slot 3: x8x8	Slot 4: x8x8
SW3200	Off	Off	Off	Off	Default	Slot 5: x16	Slot 6: x16	Slot 7: x16
	Off	**On**	**On**	**On**	Bifurcated Slots 5:6:7	Slot 5: x8x8	Slot 6: x8x8	Slot 7: x8x8
SW4200	Off	Off	Off	Off	Default	Slot 8: x16	Slot 9: x16	
	Off	Off	**On**	**On**	Bifurcated Slots 8:9	Slot 8: x8x8	Slot 9: x8x8	
SW6600	1	2	3	4	Switches for JTAG0 debug port GPU slots 0-4			
	Off	Off	Off	Off	Default.			
SW6601	1	2	3	4	Switches for JTAG1 debug port GPU slots 5-9			
	Off	Off	Off	Off	Default.			

Environmental Specifications

The table lists shipping, operating, and storage environment specifications for CS-Storm 500GT servers.

CS-Storm 500GT servers comply with ASHRAE Class A2 specifications when configured with 250W accelerator cards and ASHRAE Class A1 specifications with 300W/400W accelerators.

Table 4. CS-Storm 500GT Environmental Requirements

Environmental Factor	Requirement
Operating	
Operating temperature	Up to 250W accelerators: 41° to 95°F (5° to 35°C), up to 5,000ft (1,500m) Up to 400W accelerators: 41° to 90°F (5° to 32°C), up to 5,000ft (1,500m) Derate maximum temperature of 95°F (35°C) by 1.8°F (1°C) 1°C per 1,000ft (305m) of altitude above 5,000ft (1525m) Temperature rate of change must not exceed 18°F (10°C) per hour
Operating humidity	8% to 80% non-condensing Humidity rate of change must not exceed 10% relative humidity per hour
Operating altitude	Up to 10,000ft. (up to 3,050m)
Shipping	
Shipping temperature	-40° to 140°F (-40° to 60°C) Temperature rate of change must not exceed 36° F (20° C) per hour
Shipping humidity	10% to 95% non-condensing
Shipping altitude	Up to 40,000ft (up to 12,200m)
Storage	
Storage temperature	41° to 113°F (5° to 45°C) Temperature rate of change must not exceed 36°F (20°C) per hour
Storage humidity	8% to 80% non-condensing
Storage altitude:	Up to 40,000ft (12,200m)

S2600BP Motherboard Description

The Intel® S2600BP (Buchanan Pass) motherboard is designed to support the Intel Xeon® Scalable processor family, previously codenamed “Skylake". Previous generation Xeon processors are not supported.

Figure 18. Intel® S2600BP Motherboard

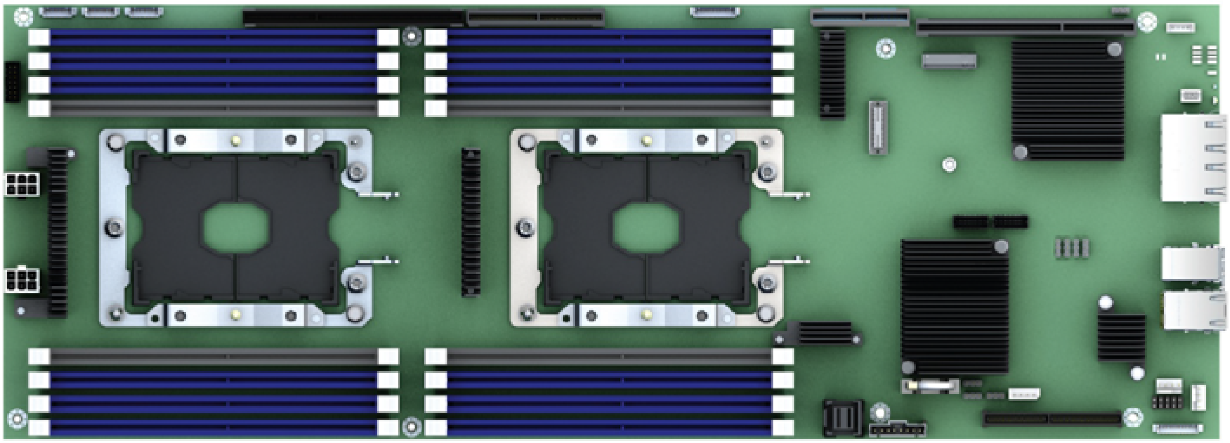


Table 5. Intel® S2600BP Motherboard Specifications

Feature	Description
Processor Support	Support for two Intel Xeon Scalable processors: ● Two LGA 3647, (Socket-P0) processor sockets ● Maximum thermal design power (TDP) of 165W ● 40 lanes of Integrated PCIe 3.0 low-latency I/O
Memory	● 16 DIMM slots in total across six memory channels ● Support for DDR4 DIMMs only. DDR3 DIMMs are not supported ● Registered DDR4 (RDIMM), Load Reduced DDR4 (LRDIMM) ● Memory DDR4 data transfer rates of 1600/1866/2133/2400/2666 MT/s ● Up to two DIMM slots per channel ● 1,536GB memory (maximum)
Chipset	Intel 621 Platform Controller Hub (PCH)
External I/O Connections	● One dedicated management port. RJ45, 100Mb/1Gb/10Gb, for remote server management (Embedded dedicated NIC module from BMC) ● Stacked dual port USB 3.0 (port 0/1) connector ● Two RJ-45 10GbE network interface controller (NIC) ports

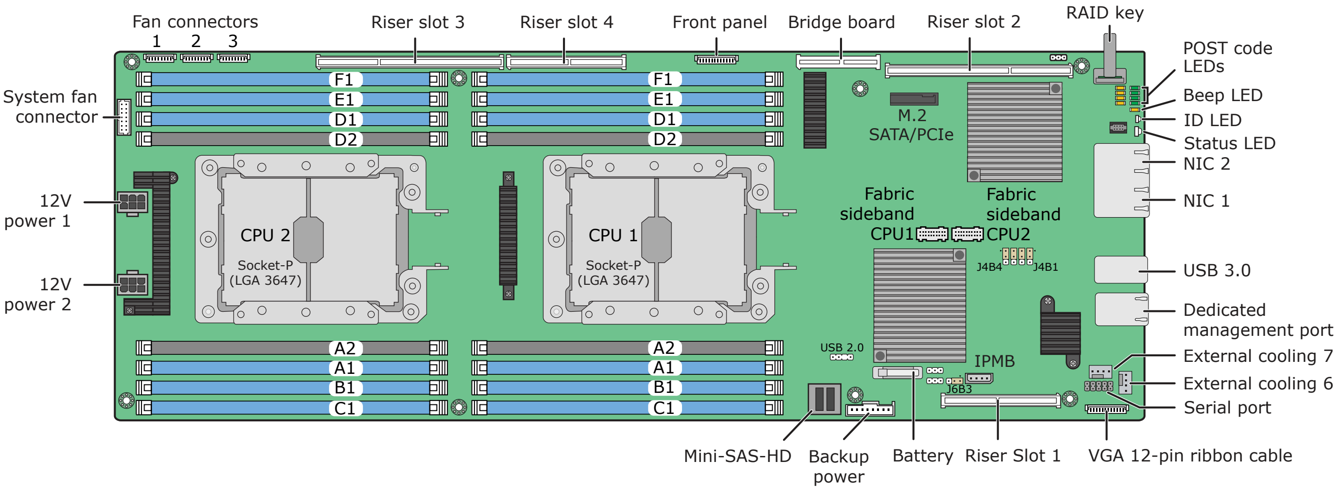
Feature	Description
Internal I/O Connectors	● Bridge slot to extend board I/O ● One 1x12 internal Video header ● One 1x4 IPMB header ● One internal USB 2.0 connector ● One 1x12 pin control panel header ● One DH-10 serial Port connector ● One 2x4 pin header for Intel® RMM4 Lite ● One 1x4 pin header for Storage Upgrade Key ● Two 2x12 pin header for Fabric Sideband CPU1/CPU2
PCIe Support	PCIe 3.0 (2.5, 5, 8 GT/s)
Power Connections	● Two sets of 2x3 pin connectors (main power 1/2) ● One backup power 1x8 connector
System Fan Support	● One 2x7 pin fan control connector ● Three 1x8 pin system fan connectors
Riser Card Support	● One bridge board slot for board I/O expansion ● Riser Slot 1 (Rear Right Side) ● VGA Bracket is installed on Riser slot 1 as a standard ● Riser Slot 2 (Rear Left Side) providing a x24 PCIe 3.0 lanes: CPU1 ● Riser Slot 3 (Front Left Side) providing a x24 PCIe 3.0 lanes: CPU2 ● Riser Slot 4 (Middle Left Side) providing a x16 PCIe 3.0 lanes: CPU2
Video	● Integrated on ASPEED AST2500 BMC ● 16MB of DDR4 video memory (512 total for BMC)
Onboard Storage Controllers and Options	● One M.2 SATA/PCIe connector (42 mm drive support only) ● Four SATA 6Gbps ports via Mini-SAS HD (SFF-8643) connector
Fabric	Dual port Intel Omni-Path fabric via processor or Single port Omni-Path fabric via x16 Gen 3 PCIe adapter
Server Management	● Onboard ASPEED AST2500 BMC controller ● Support for optional Intel Remote Management Module 4 Lite (RMM4) ● Intel Light-Guided Diagnostics on field replaceable units ● Support for Intel System Management Software

Feature	Description
	● Support for Intel Intelligent Power Node Manager (Require PMBus compliant power supply)
RAID Support	● Intel Rapid Storage RAID Technology (RSTe) 5.0 ● Intel Embedded Server RAID Technology 2 (ESRT2) with optional Intel RAID C600 Upgrade Key to enable SATA RAID 5

S2600BP Component Locations

Intel® S2600BP component locations and connector types are shown in the following figure. The motherboard includes a status and ID LED for identifying system status. These LEDs and the rear ports are described below.

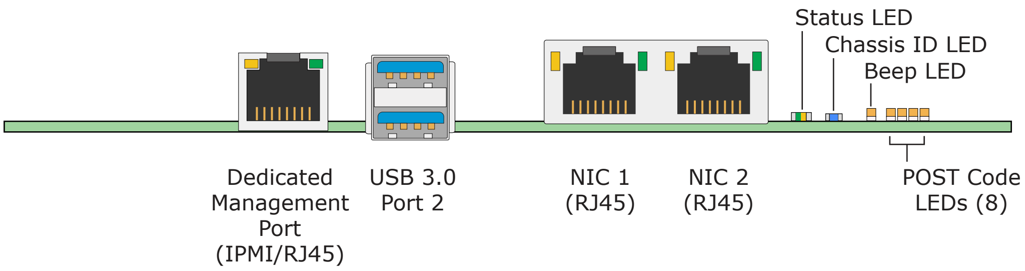
Figure 19. Intel® S2600BP Component Locations



Jumpers. The motherboard includes several jumper blocks that can be used to configure, protect, or recover specific features of the motherboard. These jumper blocks are shown in the default position in the above figure. Refer to *S2600BP Configuration and Recovery Jumpers* on page 53 for details.

POST code LEDs. There are several diagnostic (POST code and beep) LEDs to assist in troubleshooting motherboard level issues.

Figure 20. Intel® S2600BP Rear Connectors



Dedicated management port. This port with a separate IP address to access the BMC. It provides a port for monitoring, logging, recovery, and other maintenance functions independent of the main CPU, BIOS, and OS. The management port is active with or without the RMM4 Lite key installed. The dedicated management port and the two onboard NICs support a BMC embedded web server and GUI.

Dedicated management port/NIC LEDs. The link/activity LED (at the right of the connector) indicates network connection when on, and transmit/receive activity when blinking. The speed LED (at the left of the connector) indicates 10-Gbps operation when green, 1-Gbps operation when amber, and 100-Mbps when off. Figure 58 provides an overview of the LEDs.

Table 6. Intel® S2600BP NIC LEDs

LED	Color	LED State	NIC State
Left	Green	Off	LAN link not established
		On	LAN link is established
		Blinking	LAN transmit and receive activity
Right	--	Off	100 Mbit/sec data rate is selected
	Amber	On	1 Gbit/sec data rate is selected.
	Green	On	10 Gbit/sec data rate is selected

Status LED. This bicolor LED lights green (status) or amber (fault) to indicate the current health of the server. Green indicates normal or degraded operation. Amber indicates the hardware state and overrides the green status. The state detected by the BMC and other controllers are included in the Status LED state. *TRUE? The Status LED on the chassis front panel and this motherboard Status LED are tied together and show the same state.* When the server is powered down (transitions to the DC-off state or S5), the Integrated BMC is still on standby power and retains the sensor and front panel status LED state established prior to the power-down event.

The Status LED displays a steady Amber color for all Fatal Errors that are detected during processor initialization. A steady Amber LED indicates that an unrecoverable system failure condition has occurred.

A description of the Status LED states follows.

Table 7. Intel® S2600BP Status LEDs

Color	State	Criticality	Description
Off	System is not operating	Not ready	● System is powered off (AC and/or DC) ● System is in Energy-using Product (EuP) Lot6 Off mode/regulation[1] ● System is in S5 Soft-off state.
Green	Solid on	OK	Indicates the system is running (in S0 state) and status is healthy. There are no system errors. AC power is present, the BMC has booted, and management is up and running. After a BMC reset with a chassis ID solid on, the BMC is booting Linux. Control has been passed from BMC uboot to BMC Linux. Remains in this state for approximately 10-20 seconds.
	Blinking (~1 Hz)	Degraded: System is operating in a degraded state	System degraded: ● Power supply/fan redundancy loss

Color	State	Criticality	Description
		although still functional, or system is operating in a redundant state but with an impending failure warning.	● Fan warning or failure when the number of fully operational fans is less than minimum number needed to cool the system ● Non-critical threshold crossed (temperature, voltage, power) ● Power supply failure ● Unable to use all installed memory ● Correctable memory errors beyond threshold ● Battery failure ● Error during BMC operation ● BMC watchdog has reset the BMC ● Power Unit sensor offset for configuration error is asserted ● HDD HSC is off-line or degraded
Amber	Solid on	Critical, non-recoverable - system is halted	Fatal alarm: System has failed or shutdown
	Blinking (~1 Hz)	Non-critical: System is operating in a degraded state with an impending failure warning, although still functioning.	Non-fatal alarm: System failure likely ● Critical threshold crossed (temperature, voltage, power) ● VRD Hot asserted ● Minimum number of fans to cool the system not present or failed ● Hard drive fault ● Insufficient power from PSUs
[1] The overall power consumption of the system is referred to as System Power States. There are a total of six different power states ranging from: S0 (the system is completely powered ON and fully operational), to S5 (the system is completely powered OFF), and the states (S1, S2, S3, and S4) referred to as sleeping states.			

Chassis ID LED. This blue LED is used to visually identify a specific motherboard/server installed in the rack or among several racks of servers. The ID button on front of the server/node toggles the state of the chassis ID LED. There is no precedence or lock-out mechanism for the control sources. When a new request arrives, all previous requests are terminated. For example, if the chassis ID LED is blinking and the ID button is pressed, then the ID LED changes to solid on. If the button is pressed again with no intervening commands, the ID LED turns off.

Table 8. Intel® S2600BP ID LED

LED State	State
On (steady)	The LED has a solid On state when it is activated through the ID button. It remains lit until the button is pushed again or until an `ipmitool chassis identify` command is received to change the state of the LED.
Blink (~1 Hz)	The LED blinks after it is activated through a command.

LED State	State
Off	Off. Pushing the ID button lights the ID LED.

BMC Boot/Reset Status LED Indicators. During the BMC boot or BMC reset process, the System Status and Chassis ID LEDs are used to indicate BMC boot process transitions and states. A BMC boot occurs when AC power is first applied to the system. A BMC reset occurs after a BMC firmware update, after receiving a BMC cold reset command, and upon a BMC watchdog initiated reset. These two LEDs define states during the BMC boot/ reset process.

Table 9. Intel® S2600BP Boot/Reset LEDs

BMC Boot/Reset State	Chassis ID LED	Status LED	Condition
BMC/Video memory test failed	Solid blue	Solid amber	Non-recoverable condition. Contact Cray service for information on replacing the motherboard.
Both universal bootloader (u-Boot) images bad	Blink blue (6 Hz)	Solid amber	Non-recoverable condition. Contact Cray service for information on replacing the motherboard.
BMC in u-Boot	Blink blue (3 Hz)	Blink green (1 Hz)	Blinking green indicates degraded state (no manageability), blinking blue indicates u-Boot is running but has not transferred control to BMC Linux. System remains in this state 6-8 seconds after BMC reset while it pulls the Linux image into Flash.
BMC booting Linux	Solid blue	Solid green	Solid green with solid blue after an AC cycle/BMC reset, indicates control passed from u-Boot to BMC Linux. Remains in this state for ~10-20 seconds.
End of BMC boot/reset process. Normal system operation	Off	Solid green	Indicates BMC Linux has booted and manageability functionality is up and running. Fault/Status LEDs operate normally.

Beep LED. The S2600BP does not have an audible beep code component. Instead, it uses a beep code LED that translates audible beep codes into visual light sequences. Prior to system video initialization, the BIOS uses these Beep_LED codes to inform users on error conditions. A user-visible beep code is followed by the POST Progress LEDs.

Table 10. Intel® S2600BP Beep LEDs

Beep_LED Sequence	Error Message	POST Progress Code	Description
1 blink	USB device action	N/A	Short LED blink whenever USB device is discovered in POST, or inserted or removed during runtime.

Beep_LED Sequence	Error Message	POST Progress Code	Description
1 long blink	Intel® TXT security violation	0xAE, 0xAF	System halted because Intel® Trusted Execution Technology detected a potential violation of system security.
3 blinks	Memory error	Multiple	System halted because a fatal error related to the memory was detected.
3 long blinks followed by 1	CPU mismatch error	0xE5, 0xE6	System halted because a fatal error related to the CPU family/core/cache mismatch was detected.
The following “Beep_LED” Codes are lighted during BIOS Recovery.			
2 blinks	Recovery started	N/A	Recovery boot has been initiated.
4 blinks	Recovery failed	N/A	Recovery has failed. This typically happens so quickly after recovery is initiated that it lights like a 2-4 LED code.

The Integrated BMC may generate beep codes upon detection of failure conditions. Beep codes are translated into visual LED sequences each time the problem is discovered, such as on each power-up attempt, but are not lit continuously. Codes that are common across all Intel server boards and systems that use the same generation of chipset are listed in the following table. Each digit in the code is represented by a LED lit/off sequence of whose count is equal to the digit.

Table 11. Intel® S2600BP Beep LEDs

Code	Associated Sensors	Reason for Beep LED lit
1-5-2-1	No CPUs installed or first CPU socket is empty.	CPU1 socket is empty, or sockets are populated incorrectly. CPU1 must be populated before CPU2.
1-5-2-4	MSID Mismatch	MSID mismatch occurs if a processor is installed into a system board that has incompatible power capabilities.
1-5-4-2	Power fault	DC power unexpectedly lost (power good dropout) – Power unit sensors report power unit failure offset
1-5-4-4	Power control fault (power good assertion timeout).	Power good assertion timeout – Power unit sensors report soft power control failure offset
1-5-1-2	VR Watchdog Timer sensor assertion	VR controller DC power on sequence was not completed in time.
1-5-1-4	Power Supply Status	The system does not power on or unexpectedly powers off and a Power Supply Unit (PSU) is present that is an incompatible model with one or more other PSUs in the system.

POST Code Diagnostic LEDs

There are two rows of four POST code diagnostic LEDs (eight total) on the back edge of the motherboard. These LEDs are difficult to view through the back of the server/node chassis. During the system boot process, the BIOS executes a number of platform configuration processes, each of which is assigned a specific hex POST code number. As each configuration routine is started, the BIOS displays the given POST code to the POST code LEDs. To assist in troubleshooting a system hang during the POST process, the LEDs display the last POST event run before the hang.

During early POST, before system memory is available, serious errors that would prevent a system boot with data integrity cause a System Halt with a beep code and a memory error code to be displayed through the POST Code LEDs. Less fatal errors cause a POST Error Code to be generated as a major error. POST Error Codes are displayed in the BIOS Setup error manager screen and are logged in the system event log (SEL), which can be viewed with the `selview` utility. The BMC deactivates POST Code LEDs after POST is completed.

S2600BP Processor Socket Assembly

Unlike previous Intel® Xeon® E5-2600 (v3/v4) processors that use an integrated loading mechanism (ILM), the S2600BP motherboard includes two Socket-P0 (LGA 3647) processor sockets that supports Intel Xeon Scalable E5-2600 v5 processors. The socket features a rectangular shape which is different from the square shape of previous sockets. The LGA 3647 socket contains 3647 pins and provides I/O, power, and ground connections from the motherboard board to the processor package. Because of the large size of the socket, it is made of two C-shaped halves. The two halves are not interchangeable and are distinguishable from one another by the keying and pin1 insert colors: The left half keying insert is the same color as the body, the right half inserts are white.

The LGA 3647 socket also supports both the E5-2600 v5 processors with embedded Intel Omni-Path Host Fabric Interconnect (HFI).

The parts of the socket assembly is described below and shown in the following figure.

Important: The pins inside the processor socket are extremely sensitive. No object except the processor package should make contact with the pins inside the processor socket. A damaged socket pin may render the socket inoperable, and will produce erroneous CPU or other system errors.

Processor Heat Sink Module (PHM)

The PHM refers to the sub-assembly where the heatsink and processor are fixed together by the processor package carrier prior to installation on the motherboard.

Processor Package Carrier (Clip)

The carrier is an integral part of the PHM. The processor is inserted into the carrier, then the heatsink with thermal interface material (TIM) are attached. The carrier has keying/ alignment features to align to cutouts on the processor package. These keying features ensure the processor package snaps into the carrier in only one direction, and the carrier can only be attached to the heatsink in one orientation.

The processor package snaps into the clips located on the inside ends of the package carrier. The package carrier with attached processor package is then clipped to the heatsink. Hook like features on the four corners of the carrier grab onto the heatsink. All three pieces are secured to the bolster plate with four captive nuts that are part of the heatsink.

Important: Fabric supported processor models require the use of a Fabric Carrier Clip which has a different design than the standard clip shown in the figure below. Attempting to use a standard processor carrier clip with a Fabric supported processor may result in component damage and result in improper assembly of the PHM.

Bolster Plate

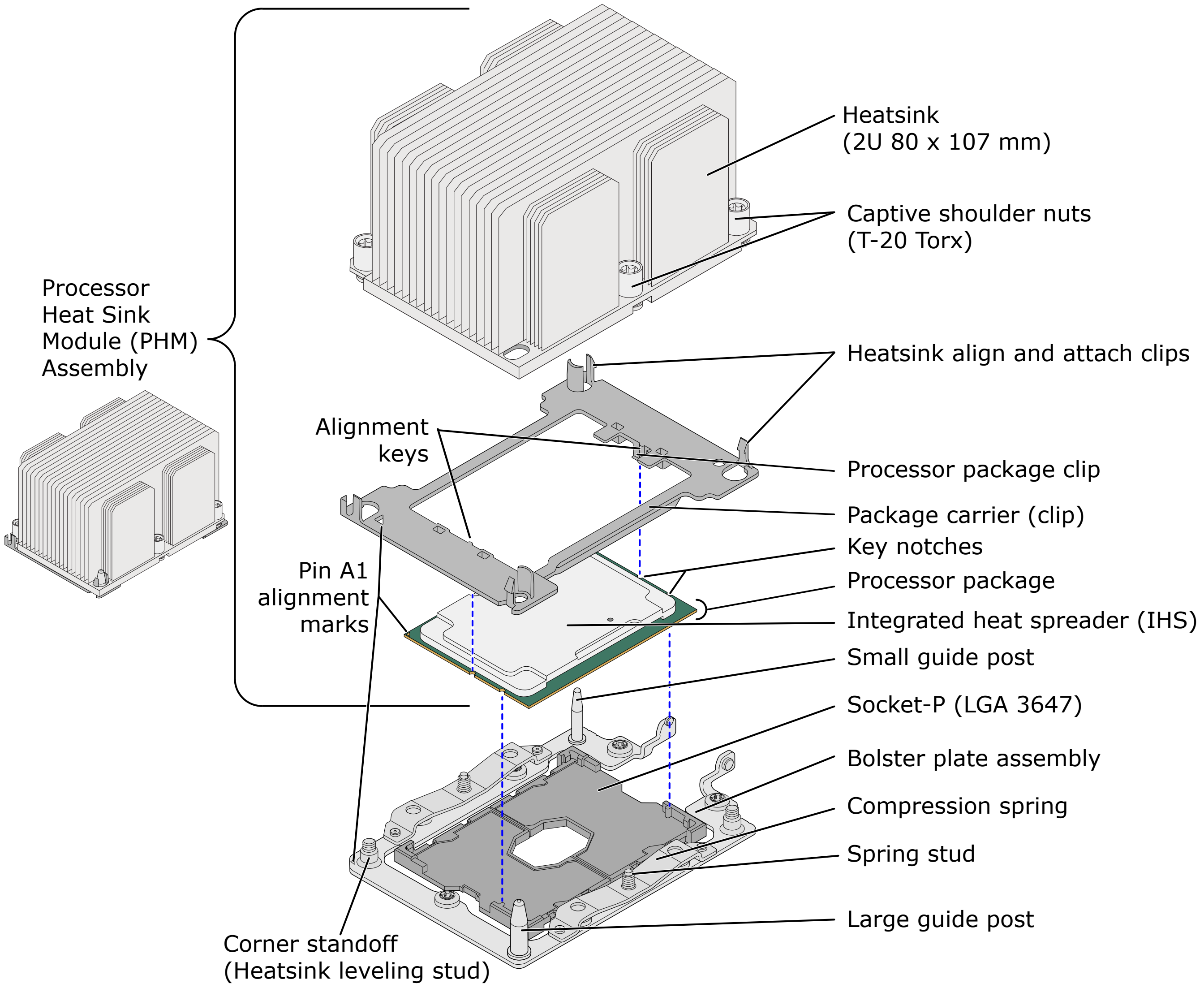
The bolster plate is an integrated subassembly that includes two corner guide posts placed at opposite corners and two springs that attach to the heatsink via captive screws. The springs are pulled upward as the heatsink is lowered and tightened in place, creating a compressive force between socket and heatsink. The bolster plate provides extra rigidity, helps maintain flatness to the motherboard, and provides a uniform load distribution across all contact pins in the socket.

Heatsink

The heatsink is integrated into the PHM which is attached to the bolster plate springs by two captive nuts on either side of the heatsink. The bolster plate is held in place around the socket by the backplate. The heatsink's captive shoulder nuts screw onto the corner standoffs and bolster plate studs. Depending on the manufacturer/model, some heatsinks may have a label on the top showing the sequence for tightening and loosening the four nuts.

There are two types of heatsinks specially for each of the processors. These heatsinks are NOT interchangeable and must be installed on the correct processor/socket, front versus rear.

Figure 21. Intel® S2600BP Processor Socket Assembly

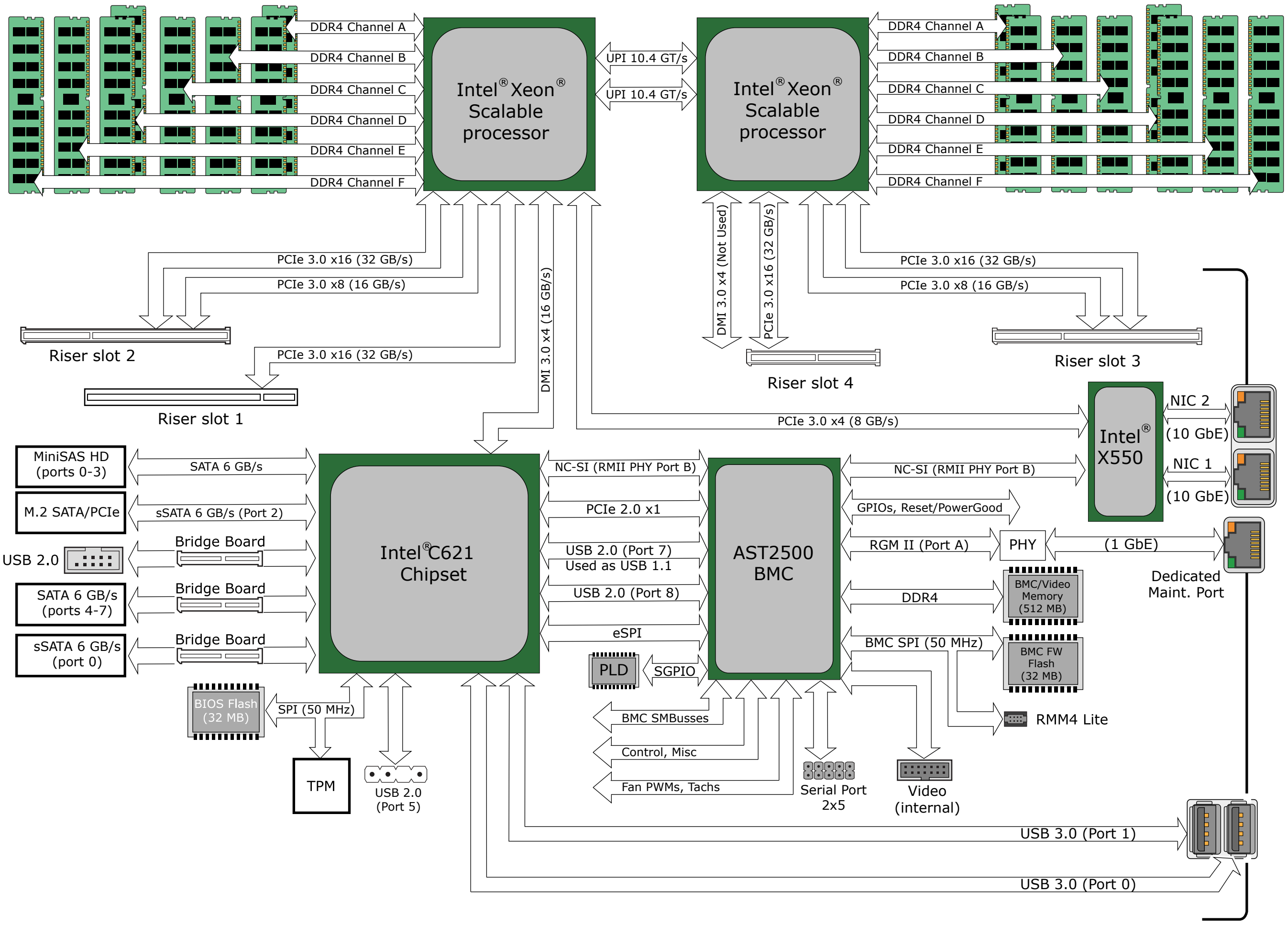


S2600BP Architecture

The architecture of Intel® Server Board S2600BP is developed around the integrated features and functions of the Intel Xeon Scalable processor family, the Intel C621 Series Chipset family, Intel® Ethernet Controller X550, and the ASPEED* AST2500* Server Board Management Controller.

The figure provides an overview of the S2600BP architecture, showing the features and interconnects of each of the major subsystem components.

Figure 22. Intel® S2600BP Block Diagram



S2600BP Processor Population Rules

Although the Intel® S2600BP motherboard supports using different processors on each socket, Cray performs platform validation only on systems that are configured with identical processors. For optimal system performance and reliability, install identical processors.

If needed, the S2600BP may operate with one processor in the CPU 1 socket. However, some board features may not be functional if a second processor is not installed. Riser slots 3 and 4 can be used only in dual processor configurations.

When two processors are installed, both processors must:

- be of the same processor family,
- have the same number of cores,
- and have the same cache sizes for all levels of processor cache memory.

Processors with different core frequencies can be mixed in a system, given the prior rules are met. If this condition is detected, all processor core frequencies are set to the lowest common denominator (highest common speed) and an error is reported.

Processors that have different Intel UltraPath (UPI) Link Frequencies may operate together if they are otherwise compatible and if a common link frequency can be selected. The common link frequency would be the highest link frequency that all installed processors can achieve.

Processor stepping within a common processor family can be mixed as long as it is listed in the processor specification updates published by Intel Corporation.

S2600BP Memory Support and Population Rules

Each Intel® S2600BP motherboard processor includes two integrated memory controllers (IMC), each capable of supporting three DDR4 memory channels. Each memory channel is capable of supporting two DIMM slots. Channel A and channel D support two memory slots and B, C, E, and F supports one memory slot, for a total possible of 16 DIMMs.

The processor IMC supports the following:

- For maximum memory performance, 12 DIMMs (one DIMM per channel) are recommended
- Registered DIMMs (RDIMMs), Load Reduced DIMMs (LRDIMMs) and LRDIMM 3DS are supported
- DIMMs of different types (RDIMM, LRDIMM) may not be mixed – this results in a Fatal Error during memory initialization at the beginning of POST
- DIMMs using x4 or x8 DRAM technology DIMMs organized as Single Rank (SR), Dual Rank (DR), or Quad Rank (QR)
- Maximum of 8 logical ranks per channel
- Maximum of 10 physical ranks loaded on a channel
- DIMM sizes of 4 GB, 8 GB, 16 GB, 32 GB, 64 GB and 128 GB depending on ranks and technology
- DIMM speeds of 1600/1866/2133/2400/2666 MT/s
- Only Error Correction Code (ECC) enabled RDIMMs or LRDIMMs are supported
- Only RDIMMs and LRDIMMs with integrated Thermal Sensor On Die (TSOD) are supported

Memory Population Rules

Although mixed DIMM configurations are supported on the Intel® S2600BP motherboard, Cray performs platform validation only on systems that are configured with identical DIMMs installed. Each memory slot should be populated with identical DDR4 DIMMs.

- The memory channels from processor socket 1 and processor socket 2 are identified as “CPU# plus A, B, C, D, E or F” respective channel.
- The memory slots associated with a given processor are unavailable if the corresponding processor socket is not populated.
- The silk screened DIMM slot identifiers on the board provide information about the channel, and therefore the processor to which they belong. For example, CPU1_DIMM_A1 is the first slot on Channel A on processor 1; CPU2_DIMM_A1 is the first DIMM socket on Channel A on processor 2.
- A processor may be installed without populating the associated memory slots, if a second processor is installed along with its associated memory. In this case, the memory is shared by the processors. However, the platform suffers performance degradation and latency due to the remote memory.
- The S2600BP uses a “2-1-1” configuration--populate first the slot closest to processor in the channel with 2 slots.

- Processor sockets are self-contained and autonomous. However, all memory subsystem support (such as Memory RAS and Error Management) in the BIOS setup is applied commonly across processor sockets.
- Mixing DIMMs of different frequencies and latencies is not supported within or across processor sockets.
- A maximum of 8 logical ranks can be used on any one channel, as well as a maximum of 10 physical ranks loaded on a channel.
- DIMM slot 1 closest to the processor socket must be populated first in the channel with 2 slots. Only remove factory installed DIMM blanks when populating the slot with memory. Intel MRC will check for correct DIMM placement

S2600BP Configuration and Recovery Jumpers

The Intel® S2600BP motherboard has several 3-pin jumper blocks that can be used to configure, protect, or recover specific features of the motherboard. Refer to *Intel S2600BP Component Locations* on page 43 to locate each jumper block on the motherboard. Pin 1 of each jumper block can be identified by the arrowhead (▼) silk screened next to the pin. The default position for each jumper block is pins 1 and 2.

BMC Force Update (J6B3)

When performing a standard BMC firmware update procedure, the update utility places the BMC into an update mode, allowing the firmware to load safely onto the flash device. In the unlikely event the BMC firmware update process fails due to the BMC not being in the proper update state, the server board provides a BMC Force Update jumper (J6B3) which will force the BMC into the proper update state. The following procedure should be followed in the event the standard BMC firmware update process fails.

Normal BMC functionality is disabled with the Force BMC Update jumper set to the enabled position. You should never run the server with the BMC Force Update jumper set in this position. You should use this jumper setting only when the standard firmware update process fails. This jumper should remain in the default/disabled position when the server is running normally

To perform a Force BMC Update, follow these steps:

1. Move the jumper (J6B3) from the default operating position (covering pins 1 and 2) to the enabled position (covering pins 2 and 3).
2. Power on the server by pressing the power button on the front panel.
3. Perform the BMC firmware update procedure as documented in the Release Notes included in the given BMC firmware update package. After successful completion of the firmware update process, the firmware update utility may generate an error stating the BMC is still in update mode.
4. Power down the server.
5. Move the jumper from the enabled position (covering pins 2 and 3) to the disabled position (covering pins 1 and 2).
6. Power up the server.

ME Force Update (J4B1)

When this 3-pin jumper is set, it manually puts the ME firmware in update mode, which enables the user to update ME firmware code when necessary.

Normal ME functionality is disabled with the Force ME Update jumper set to the enabled position. You should never run the server with the ME Force Update jumper set in this position. You should only use this jumper setting when the standard firmware update process fails. This jumper should remain in the default/disabled position when the server is running normally.

To perform a Force ME Update, follow these steps:

1. Move the jumper (J4B1) from the default operating position (covering pins 1 and 2) to the enabled position (covering pins 2 and 3).
2. Power on the server by pressing the power button on the front panel.
3. Perform the ME firmware update procedure as documented in the Release Notes file that is included in the given system update package.
4. Power down the server.
5. Move the jumper from the enabled position (covering pins 2 and 3) to the disabled position (covering pins 1 and 2).
6. Power up the server.

Password Clear (J4B2)

The user sets this 3-pin jumper to clear the password. This jumper causes both the User password and the Administrator password to be cleared if they were set. The operator should be aware that this creates a security gap until passwords have been installed again.

No method of resetting BIOS configuration settings to the default values will affect either the Administrator or User passwords.

This is the only method by which the Administrator and User passwords can be cleared unconditionally. Other than this jumper, passwords can only be set or cleared by changing them explicitly in BIOS Setup or by similar means

The recommended steps for clearing the User and Administrator passwords are:

1. Move the jumper (J4B2) from the default operating position (covering pins 1 and 2) to the enabled position (covering pins 2 and 3).
2. Power on the server by pressing the power button on the front panel.
3. Boot into the BIOS Setup. Check the Error Manager tab for POST Error Codes:
   - 5221 Passwords cleared by jumper
   - 5224 Password clear jumper is set
4. Power down the server.
5. Move the jumper from the enabled position (covering pins 2 and 3) to the disabled position (covering pins 1 and 2).
6. Power up the server.
7. **Strongly recommended:** Boot into the BIOS Setup immediately, go to the Security tab and set the Administrator and User passwords if you intend to use BIOS password protection.

BIOS Recovery Mode (J4B3)

If a system is completely unable to boot successfully to an OS, hangs during POST, or even hangs and fails to start executing POST, it may be necessary to perform a BIOS Recovery procedure, which can replace a defective copy of the Primary BIOS.

The BIOS introduces three mechanisms to start the BIOS recovery process, which is called Recovery Mode:

- The Recovery Mode Jumper causes the BIOS to boot in Recovery Mode.
- The Boot Block detects partial BIOS update and automatically boots in Recovery Mode.
- The BMC asserts Recovery Mode GPIO in case of partial BIOS update and FRB2 time-out.

The BIOS Recovery takes place without any external media or Mass Storage device as it utilizes the Backup BIOS inside the BIOS flash in Recovery Mode. The Recovery procedure is included here for general reference. However, if in conflict, the instructions in the BIOS Release Notes are the definitive version

When Recovery Mode Jumper is set, the BIOS begins with a “Recovery Start” event logged to the SEL, loads and boots with the Backup BIOS image inside the BIOS flash itself. This process takes place before any video or console is available. The system boots up into the Shell directly while a “Recovery Complete” SEL logged. An external media is required to store the BIOS update package and steps are the same as the normal BIOS update procedures. After the update is complete, there will be a message displayed stating that the “BIOS has been updated successfully" indicating the BIOS update process is finished. The User should then switch the recovery jumper back to normal operation and restart the system by performing a power cycle.

If the BIOS detects partial BIOS update or the BMC asserts Recovery Mode GPIO, the BIOS will boot up with Recovery Mode. The difference is that the BIOS boots up to the Error Manager Page in the BIOS Setup utility. In the BIOS Setup utility, boot device, Shell or Linux for example, could be selected to perform the BIOS update procedure under Shell or OS environment.

Again, before starting to perform a Recovery Boot, be sure to check the BIOS Release Notes and verify the Recovery procedure shown in the Release Notes.

The following steps demonstrate this recovery process:

1. Move the jumper (J4B3) from the default operating position (covering pins 1 and 2) to the BIOS Recovery position (covering pins 2 and 3).
2. Power on the server.
3. The BIOS will load and boot with the backup BIOS image without any video or display.
4. When the compute module boots into the EFI shell directly, the BIOS recovery is successful.
5. Power off the server.
6. Move the jumper (J4B3) back to the normal position (covering pins 1 and 2).
7. Put the server back into the rack. A normal BIOS update can be performed if needed.

BIOS Default (J4B4)

This jumper causes the BIOS Setup settings to be reset to their default values. On previous generations of server boards, this jumper has been referred to as “Clear CMOS”, or “Clear NVRAM”. Setting this jumper according to the procedure below will clear all current contents of NVRAM variable storage, and then load the BIOS default settings.

This jumper does not reset Administrator or User passwords. In order to reset passwords, the Password Clear jumper must be used.

The recommended steps to reset to the BIOS defaults are:

1. Move the jumper from pins 1-2 to pins 2-3 momentarily. It is not necessary to leave the jumper in place while rebooting.
2. Restore the jumper from pins 2-3 to the normal setting of pins 1-2.

3. Boot the system into Setup. Check the Error Manager tab, and you should see POST Error Codes:
   - 0012 System RTC date/time not set
   - 5220 BIOS Settings reset to default settings
4. Go to the Setup Main tab, and set the System Date and System Time to the correct current settings. Make any other changes that are required in Setup – for example, Boot Order.

S2600BP BIOS Features

Hot Keys Supported During POST

The Intel® S2600BP BIOS-supported Hot Keys are recognized by the BIOS during the system boot-time POST process only. A Hot Key is recognized as an unprompted command input, where the operator is not prompted to press the Hot Key.

After the POST process has completed and handed off the system boot process to the OS, BIOS-supported Hot Keys are no longer recognized.

Table 12. Intel® S2600BP Boot-time POST Hot Keys

Hot Key Combination	Function
<F2>	Enter the BIOS Setup Utility
<F6>	Pop-up BIOS Boot Menu
<F12>	Network boot
<Esc>	Switch from Logo Screen to Diagnostic Screen
<Pause>	Stop POST temporarily

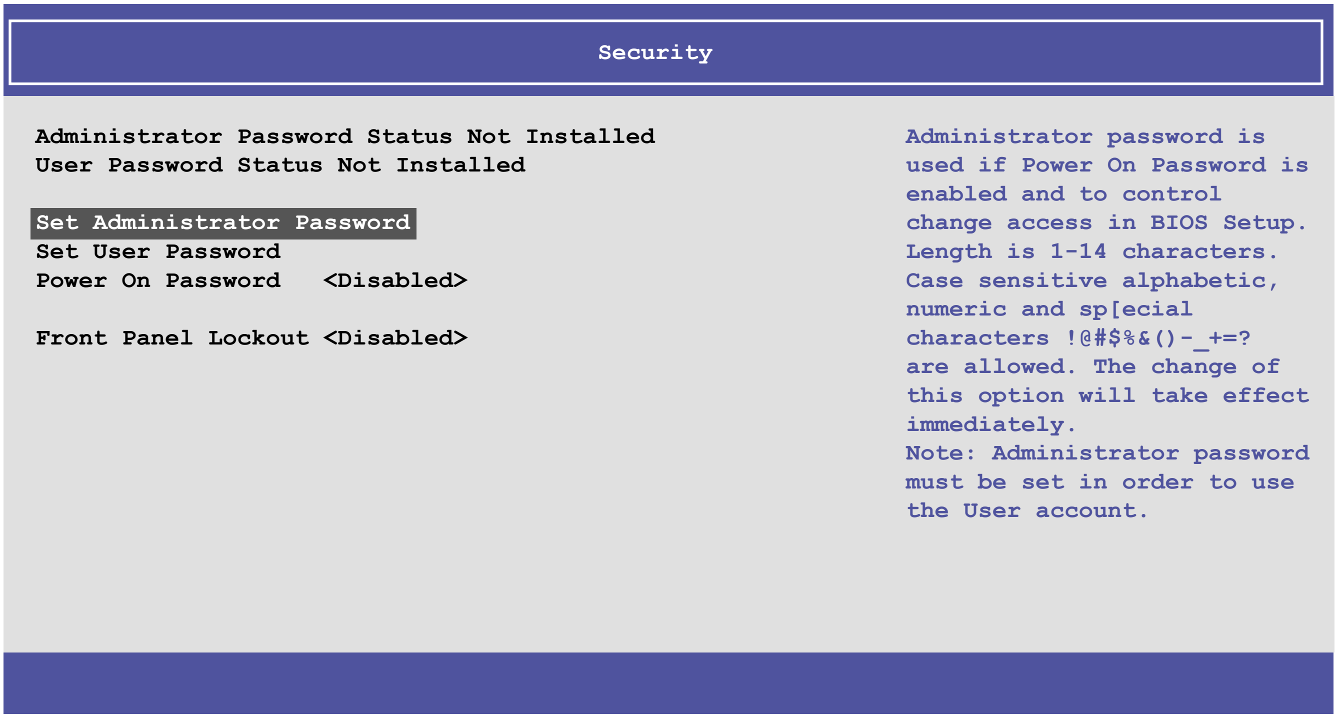
BIOS Security Features

The motherboard BIOS supports the following system security options designed to prevent unauthorized system access or tampering of server/node settings:

- Password protection
- Front panel lockout

The <F2> BIOS Setup Utility, accessed during POST, includes a Security tab with options to configure passwords and front panel lockout.

Figure 23. BIOS Setup Security Tab



```
Security

Administrator Password Status Not Installed
User Password Status Not Installed

Set Administrator Password
Set User Password
Power On Password   <Disabled>

Front Panel Lockout <Disabled>
```

```
Administrator password is
used if Power On Password is
enabled and to control
change access in BIOS Setup.
Length is 1-14 characters.
Case sensitive alphabetic,
numeric and sp[ecial
characters !@#$%&()-_+=?
are allowed. The change of
this option will take effect
immediately.
Note: Administrator password
must be set in order to use
the User account.
```

Entering BIOS Setup

To enter the BIOS Setup Utility using a keyboard (or emulated keyboard), press the <F2> function key during boot time when the POST Diagnostic Screen is displayed. If using a USB keyboard, it is important to wait until the BIOS “discovers” the keyboard and beeps. Until the USB Controller has been initialized and the USB keyboard activated, key presses will not be read by the system.

When the Setup Utility is entered, the Main screen is displayed initially. However, in the event that a serious error occurs during POST, the system will enter the BIOS Setup Utility and display the Error Manager screen instead of the Main screen.

Typically, changing BIOS settings has been done primarily through the BIOS Setup utility. After navigating through the menu screen and making desired changes, <F10> is pressed to "Save and Exit" the utility. BIOS changes are saved and the system is rebooted to make all changes take effect.

Password Setup

The BIOS uses passwords to prevent unauthorized access to the server. Passwords can restrict entry to the BIOS Setup utility, restrict use of the Boot Device pop-up menu during POST, suppress automatic USB device reordering, and prevent unauthorized system power on. It is strongly recommended that an Administrator Password be set. A system with no Administrator password set allows anyone who has access to the server to change BIOS settings.

An Administrator password must be set in order to set the User password.

The maximum length of a password is 14 characters and can be made up of a combination of alphanumeric (a-z, A-Z, 0-9) characters and any of the following special characters:

! @ # $ % ^ & * ( ) – _ + = ?

Passwords are case sensitive.

The Administrator and User passwords must be different from each other. An error message will be displayed and a different password must be entered if there is an attempt to enter the same password for both. The use of “Strong Passwords” is encouraged, but not required. In order to meet the criteria for a strong password, the password entered must be at least 8 characters in length, and must include at least one each of alphabetic, numeric, and special characters. If a weak password is entered, a warning message will be displayed, and the weak password will be accepted.

Once set, a password can be cleared by changing it to a null string. This requires the Administrator password, and must be done through BIOS Setup or other explicit means of changing the passwords. Clearing the Administrator password will also clear the User password. Passwords can also be cleared by using the Password Clear jumper on the motherboard. Resetting the BIOS configuration settings to default values (by any method) has no effect on the Administrator and User passwords.

As a security measure, if a User or Administrator enters an incorrect password three times in a row during the boot sequence, the system is placed into a halt state. A system reset is required to exit out of the halt state. This feature makes it more difficult to guess or break a password.

System Administrator Password Rights

When the correct Administrator password is entered when prompted, the user has the ability to perform the following actions:

- Access the <F2> BIOS Setup Utility
- Configure all BIOS setup options in the <F2> BIOS Setup Utility
- Clear both the Administrator and User passwords
- Access the <F6> Boot Menu during POST

If the Power On Password function is enabled in BIOS Setup, the BIOS will halt early in POST to request a password (Administrator or User) before continuing POST.

Authorized System User Password Rights and Restrictions

When the correct User password is entered, the user has the ability to perform the following:

- Access the <F2> BIOS Setup Utility
- View, but not change any BIOS Setup options in the <F2> BIOS Setup Utility
- Modify System Time and Date in the BIOS Setup Utility
- If the Power On Password function is enabled in BIOS Setup, the BIOS will halt early in POST to request a password (Administrator or User) before continuing Post

In addition to restricting access to most Setup fields to viewing only when a User password is entered, defining a User password imposes restrictions on booting the system. In order to simply boot in the defined boot order, no password is required. However, the F6 Boot pop-up menu prompts for a password, and can only be used with the Administrator password. Also, when a User password is defined, it suppresses the USB Reordering that occurs, if enabled, when a new USB boot device is attached to the system. A User is restricted from booting in anything other than the Boot Order defined in the Setup by an Administrator.

Front Panel Lockout

If enabled in BIOS setup, this option disables the following front panel features:

- The OFF function of the Power button
- System Reset button

If [Enabled], the power and reset buttons on the server front panel are locked, and they must be controlled via a system management interface.