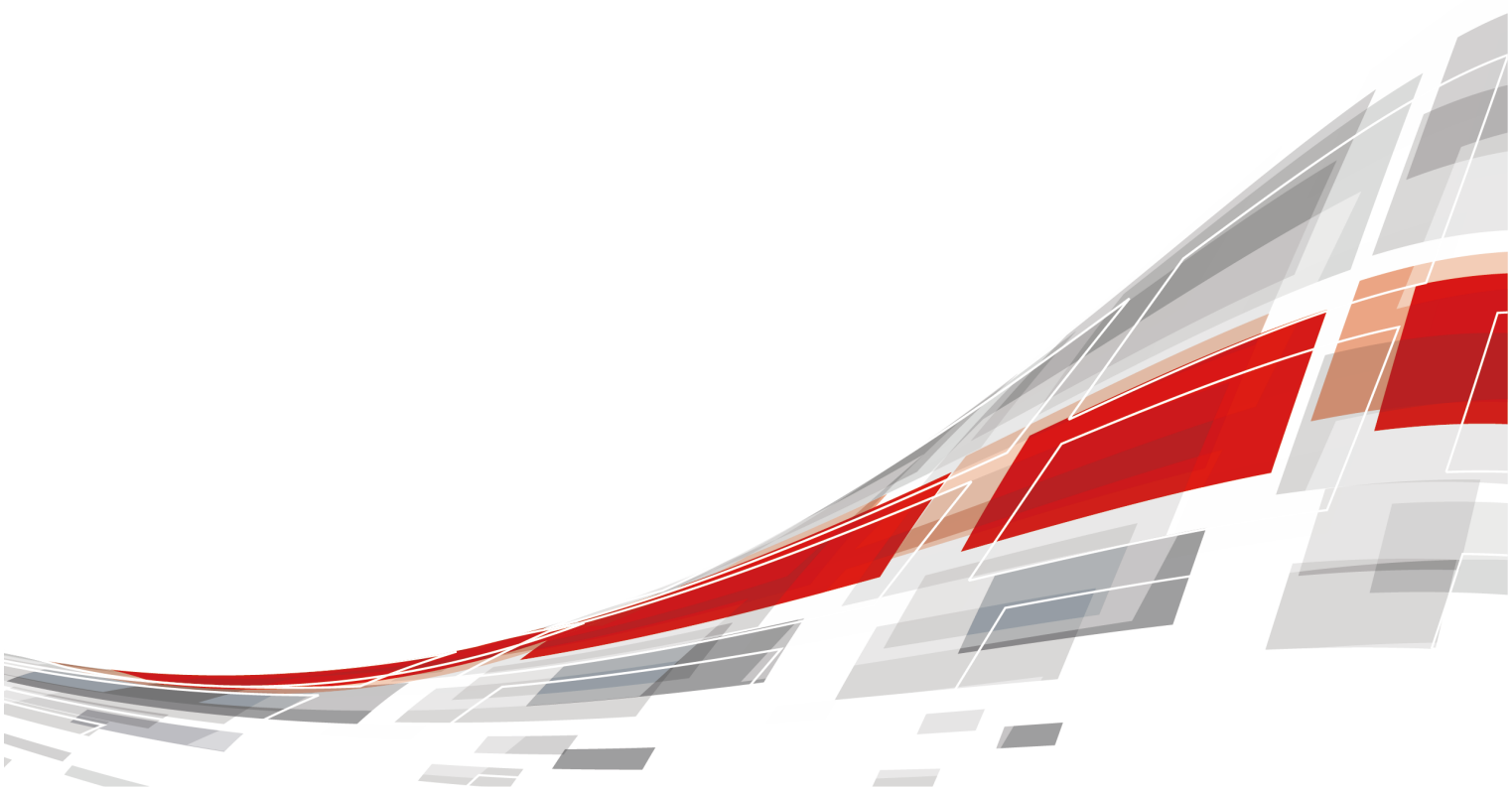
FusionServer G5500 Server

Technical White Paper

Issue 04

Date 2023-11-30



xFusion Digital Technologies Co., Ltd.

Trademarks and Permissions

XFUSION and other xFusion trademarks are trademarks of xFusion Digital Technologies Co., Ltd. All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

In this document, "xFusion" is used to refer to "xFusion Digital Technologies Co., Ltd." for concise description and easy understanding, which does not mean that "xFusion" may have any other meaning. Any "xFusion" mentioned or described hereof may not be understood as any meaning other than "xFusion Digital Technologies Co., Ltd.", and xFusion Digital Technologies Co., Ltd. shall not bear any liability resulting from the use of "xFusion".

The purchased products, services and features are stipulated by the contract made between xFusion and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

xFusion Digital Technologies Co., Ltd.

Address: 9th Floor, Building 1, Zensun Boya Square, Longzihu Wisdom Island
Zhengdong New District 450046
Zhengzhou, Henan Province
People's Republic of China

Website: https://www.xfusion.com

About This Document

Purpose

This document describes the overall chassis architecture, modules, hardware configuration, power supply, and heat dissipation design of the FusionServer G5500 server. For details about the compute node, see the corresponding technical white paper.

Intended Audience

This document is intended for pre-sales engineers.

Symbol Conventions

The symbols that may be found in this document are defined as follows.

Symbol	Description
DANGER	Indicates a hazard with a high level of risk which, if not avoided, will result in death or serious injury.
WARNING	Indicates a hazard with a medium level of risk which, if not avoided, could result in death or serious injury.
CAUTION	Indicates a hazard with a low level of risk which, if not avoided, could result in minor or moderate injury.
NOTICE	Indicates a potentially hazardous situation which, if not avoided, could result in equipment damage, data loss, performance deterioration, or unanticipated results. NOTICE is used to address practices not related to personal injury.

Symbol	Description
NOTE	Supplements the important information in the main text. NOTE is used to address information not related to personal injury, equipment damage, and environment deterioration.

Change History

Issue	Date	Description
04	2023-11-30	• Updated **3.1 Module List**. • Added **9 Waste Product Recycling**.
03	2022-11-11	Updated **3.1 Module List**.
02	2022-06-25	Updated **Certifications**.
01	2021-10-22	This issue is the first official release.

Contents

1 Product Overview

1.1 Overview

Rapid advancements in applied computing, such as artificial intelligence (AI), high-performance computing (HPC), and big data analytics, mean growing demand for computing power. As we reach the limit of Moore's law, memory and power consumption become the limiting factor to facing this challenge. A viable solution is heterogeneous computing, which allows for flexible combinations of CPU, GPU, FPGA, and ASIC processors and multiplies computing performance and energy efficiency exponentially.

Nowadays, deep learning algorithms are developing rapidly in the AI field. Typical examples are that AlphaGo defeats the world's No.1 Go player and the face recognition accuracy offered by visual computing outperforms human eyes. With the help of heterogeneous computing in data training and reasoning, outstanding achievements are also recorded in the fields such as financial, auxiliary clinical treatment, automatic driving, and public security.

FusionServer G5500 (G5500 for short) is a heterogeneous computing server dedicated for AI, HPC, cloud computing, and big data processing. It has been optimized to adapt to service scenarios such as HPC as well as data training and inference for deep learning. It supports heterogeneous processors including CPU, GPU, and FPGA and applies both to enterprise and public cloud deployments.

The G5500 is a 4U modular server that features excellent computing performance, flexible orchestration, high-performance large-capacity built-in storage, and easy maintenance.

Market positioning of the G5500 is as follows:

- A flexibly orchestratable, high-performance, and easy-to-maintain basic hardware platform for AI training and cloud services in Internet and data center scenarios
- A high-density high-bandwidth AI inference hardware platform for Smart City video analytics

- A high-performance hardware platform for HPC applications

Figure 1-1 shows the G5500 appearance.

Figure 1-1 G5500 appearance

1.2 Product Features

Excellent heterogeneous computing capabilities

- Supports NVIDIA GPUs and Intel CPUs of all thermal design power (TDP) specifications.
- Supports configuration of two CPUs and 24 DIMMs to address applications requiring high memory capacity, for example, large-scale neural network model training.

Flexibly orchestratable for performance optimization in HPC or AI scenarios

- Supports GPU topology orchestration with one-click configuration.
- Supports configuration of single or dual nodes and different ratios between GPUs and CPUs.

Industry-leading storage performance and capacity

- Supports RAID 1 configuration for the disk on which the OS is installed.
- Provides built-in NVMe disks that support a maximum bandwidth of 24 GB/s for AI training and HPC high-performance storage.
- Provides built-in 80 TB large-capacity RAID 6 permanent storage, significantly reducing the cost of external NAS storage devices.

Modular architecture, easy to maintain

- Supports GPU and CPU decoupling and independent evolution to protect customer investment.
- Adopts a modular design that helps improve the maintenance efficiency while reducing the maintenance cost and service interruption time.

Optimized heat dissipation and power supply design, ensuring system stability, security, and reliability

- Independent air channel for heat dissipation; chassis support CPUs and GPUs running with full loads.
- Uses a passive busbar backplane design that supports high current, more secure and stable.

- Supports redundant fan modules and PSUs as well as RAID configuration for storage, preventing data loss and service interruption.

Manageable and secure

- The following requirements in NIST SP 800-147B are met:
  - The BIOS firmware digital signature update mechanism is supported. During the upgrade, the digital signature is verified to prevent unauthorized BIOS firmware upgrade.
  - The flash security protection mechanism is supported to prevent unauthorized modification of the flash memory in the OS.

2 System Architecture

2.1 Logical Architecture

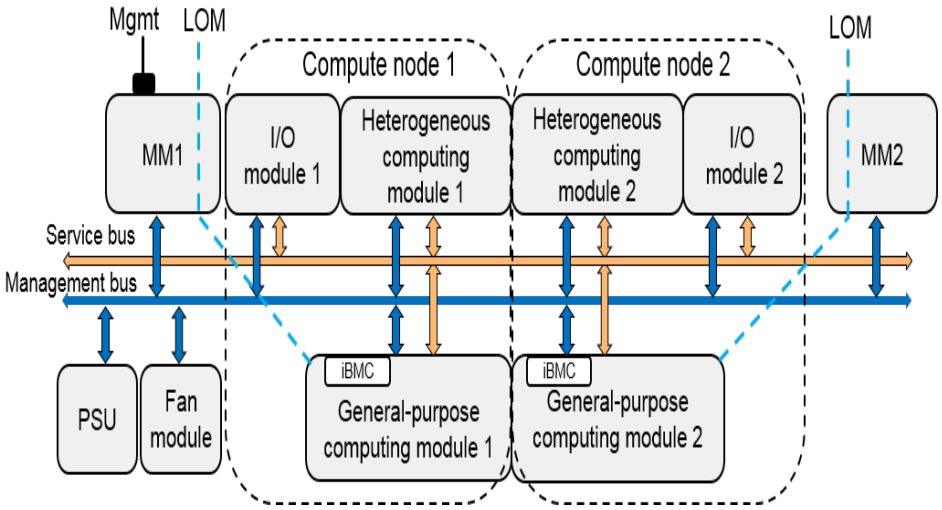
Logically, the G5500 contains six modules: general-purpose compute module (x86), heterogeneous compute module (GPU/FPGA), I/O module, chassis management module, fan module, and PSU.

The general-purpose compute module, heterogeneous compute module, and I/O module logically comprise a unified compute node through the backplane high-speed service bus (PCIe). The compute node supports two form factors, full-width and half-width. In the full-width configuration, a chassis can house a single node; in the half-width configuration, a chassis can house two nodes. The I/O module provides external ports through a standard PCIe card, and the compute node provides external ports through the management module (MM) by way of LAN on motherboard (LOM).

All modules communicate through the backplane management bus. In addition to modules, the management module also manages nodes, fans, and PSUs of the entire chassis.

- **Figure 2-1** shows the logical system architecture.

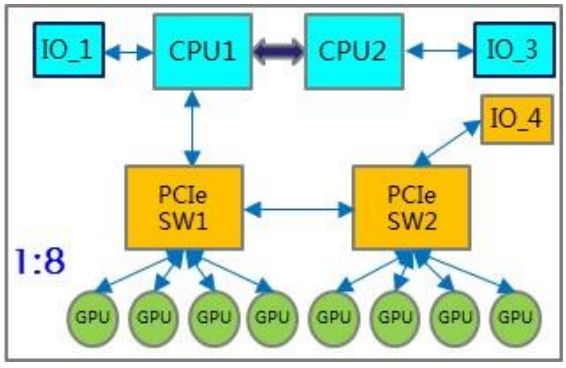
Figure 2-1 Logical architecture of the G5500



The G5500 supports multiple logical topologies. Full-width nodes support three types of topologies: cascaded, balanced, and NVLink™ topologies. Topologies can be configured using the management software to quickly adapt to different service scenarios to achieve the optimal performance. Half-width nodes support high-density inference and balanced topologies. The details are as follows:

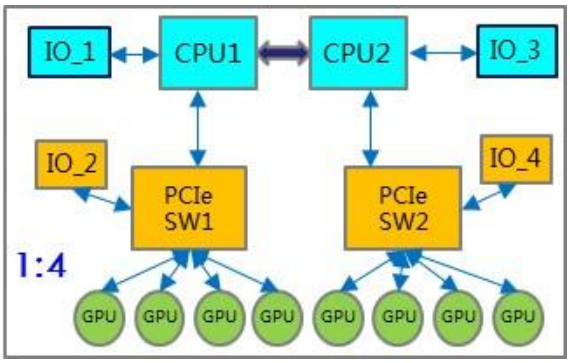
- **Figure 2-2** shows the full-width node cascaded topology. The ratio of CPUs to GPUs is 1:8. This topology supports P2P between GPUDirect™ RDMA and eight GPUs for large-scale deep learning.

Figure 2-2 Full-width node cascaded topology



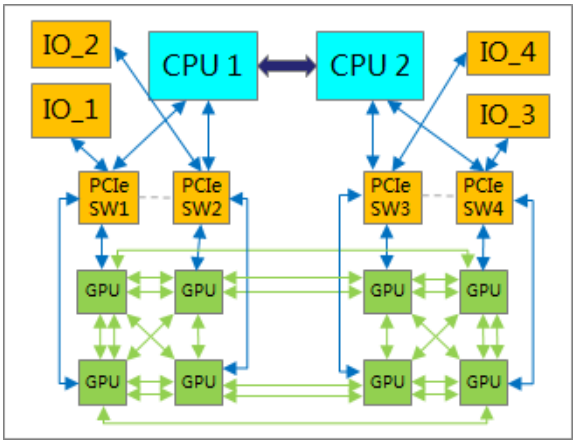
- **Figure 2-3** shows the full-width node balanced topology. The ratio of CPUs to GPUs is 1:4. This topology supports P2P between GPUDirect™ RDMA and four GPUs, and is applicable to small- and medium-scale deep learning training and inference, public cloud, and HPC scenarios.

Figure 2-3 Full-width node balanced topology



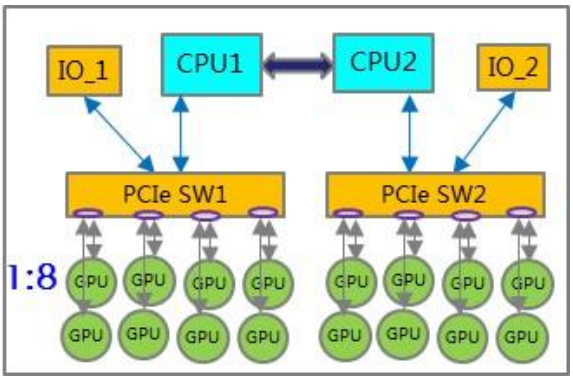
- **Figure 2-4** shows the full-width node NVLink™ topology. The ratio of CPUs to GPUs is 1:4. This topology supports P2P between GPUDirect™ RDMA and eight GPUs for large-scale deep learning.

Figure 2-4 Full-width node NVLink™ topology



- **Figure 2-5** shows the half-width node high-density inference topology. A maximum of 16 half-height half-length P4 GPU cards are supported. The ratio of CPUs to GPUs is 1:8. This topology is applicable to high-density inference scenarios in large-scale deployment.

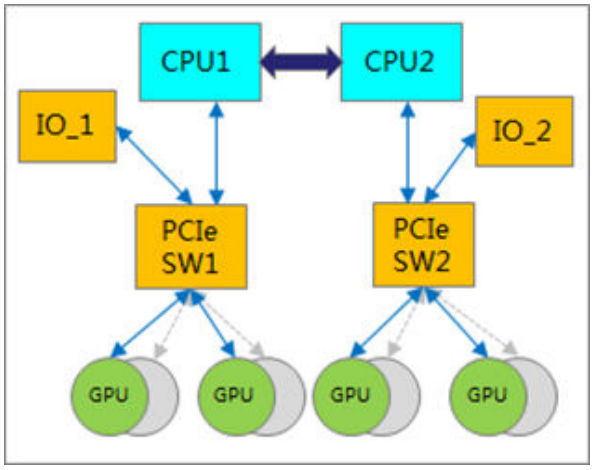
Figure 2-5 Half-width node high-density inference topology



- **Figure 2-6** shows the half-width node balanced topology. This topology supports four full-height full-length dual-slot GPU cards or eight full-height half-length

single-slot GPU cards, which are applicable to HPC and small-and medium-scale deep learning training and inference scenarios.

Figure 2-6 Half-width node balanced topology



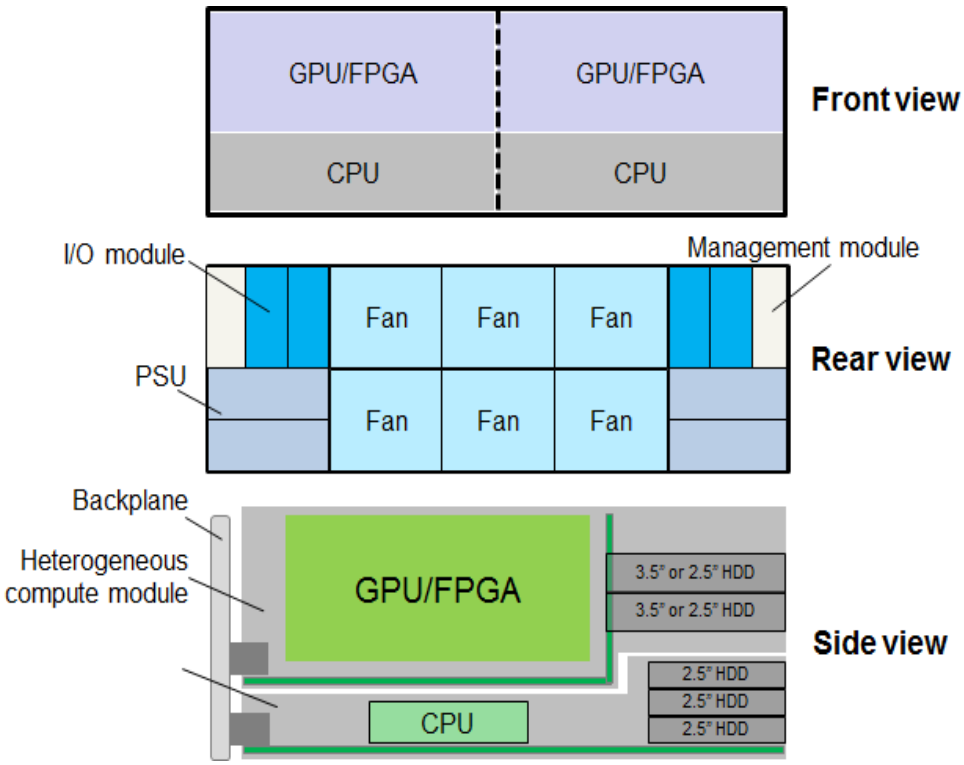
2.2 Server Design

The design used by the G5500 has the following features:

- The server is 4U high and can be installed in a standard 19-inch cabinet. Two types of compute nodes, full-width and half-width, are supported. A chassis can house one full-width compute node or two half-width compute nodes.
- The server adopts the full modular design. Components including fans, PSUs, I/O modules, chassis management modules, general-purpose compute modules, and heterogeneous compute modules are hot-swappable and can be maintained without being removed.
- PSUs share a 12 V DC bus and support N+N redundancy.
- Fans are configured as fan trays in N+1 redundancy mode to improve the heat dissipation efficiency and reliability and tolerate single failures of fans.

Figure 2-7 shows the system architecture of the G5500.

Figure 2-7 System architecture of the G5500



NOTE

- To make full use of space and prevent heterogeneous modules from being violently removed, the heterogeneous compute modules in the upper layer and the general-purpose compute modules in the lower layer are staggered. To maintain these modules, you need to follow the specified sequence.
- For installation, install the heterogeneous compute modules in the upper layer and then the general-purpose compute modules in the lower layer.
- For removal, remove the general-purpose compute modules in the lower layer and then the heterogeneous compute modules in the upper layer.

2.3 Heat Dissipation Design

- Front-to-rear ventilation: The system draws in air from the front of the chassis, delivers the cool air to the hard disks, processors, and dual in-line memory modules (DIMMs), and then discharges warm air through the rear of the chassis.
- Independent air channel for CPUs and GPUs: CPUs and GPUs are separately located in the upper and lower slots. An independent air channel is equipped for each channel, helping to achieve the optimal heat dissipation effect.
- Heat dissipation for PSUs: PSUs are located on both sides at the rear of the chassis. An independent air channel is reserved on either chassis side to enable cooling of PSUs by using built-in fans.
- Heat dissipation for the management module and I/O module: The two modules are located in the upper part at the rear of the chassis. Cooling of the two modules is implemented through the returned air of built-in fans.
- Counter-rotating fans with high wind pressure have 10% higher air volume than common fans. The refined ventilation channel design concentrates heat

dissipation capability of the system for heat-sensitive components. Proper layout of temperature sensors cover all hot spots, and support precise speed adjustment. The honeycomb hole design on the panel has a porosity rate of 66%, which is 10% higher than the porosity rate of square holes. The fan speed is adjusted using the PID algorithm, ensuring that the system can run stably at 35°C (95°F) for a long time.

Figure 2-8 and **Figure 2-9** show the ventilation channels of the G5500.

Figure 2-8 Ventilation channels (front view)

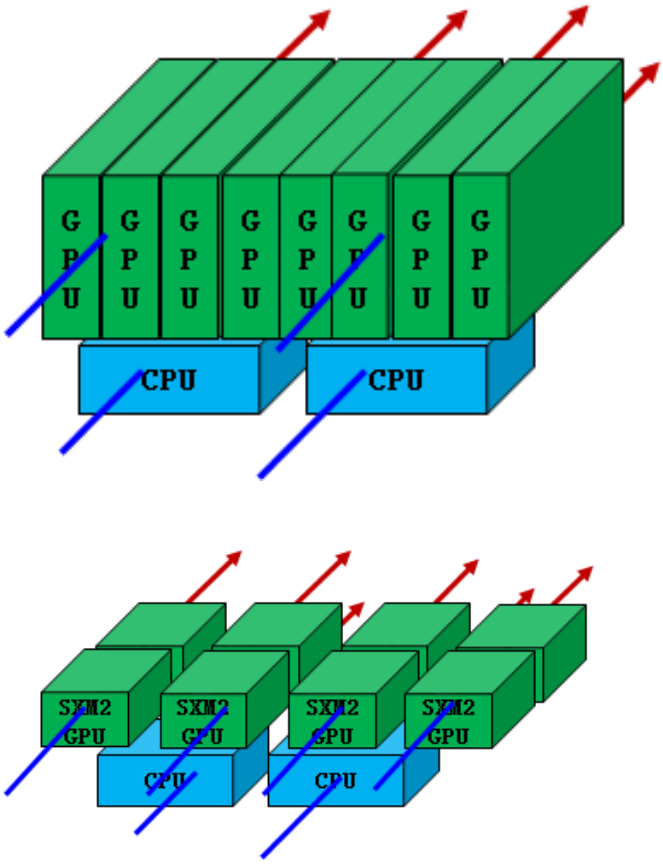
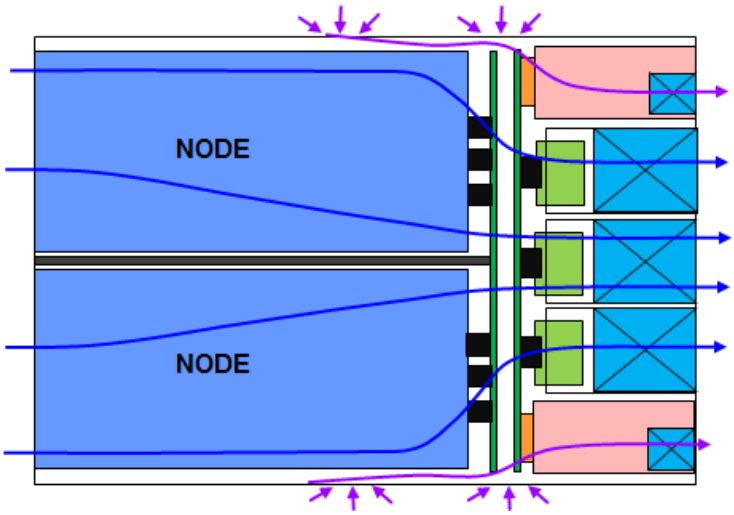


Figure 2-9 Ventilation channels (top view)

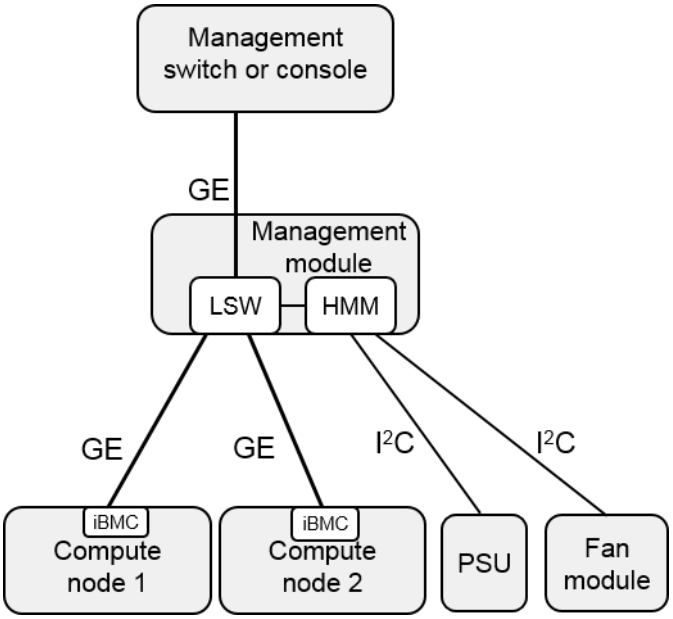


2.4 Management System

- The system uses a dual-layer architecture that comprises the Integrated Baseboard Management Controller (iBMC) and the Hyper Management Module (HMM). The iBMC on each compute node manages the node through the Intelligent Platform Management Interface (IPMI), KVM, or virtual DVD-ROM drive. The HMM manages fan modules, PSUs, and chassis assets.
- Ports of the iBMC and HMM are converged through the built-in switch of the chassis management module. Users can access the HMM and iBMC over a GE port to carry out chassis and node management.
- The HMM adjusts speeds and manages alarms of all fan modules through the I2C bus centrally. It sends management commands to fan modules, which then deliver pulse-width modulation (PWM) signals to fans.
- The HMM determines proper fan speeds based on the temperature of temperature-sensitive components on each board and the ambient temperature, and then sends the speed adjustment command to the fan modules.
- The HMM manages each two PSUs through one I2C bus. It detects the operating status and PWROK status of the PSUs by using GPIO pins and reports alarms when detecting an exception. It supports query of the output power and service status of PSUs as well as reporting of power alarms.

  Figure 2-10 shows the management system of the G5500.

Figure 2-10 Management system of the G5500



3 Hardware Description

3.1 Module List

The G5500 adopts a modular design. **Table 3-1** describes the modules of the G5500.

Table 3-1 List of modules

Model	Type	Description
G5500	4U chassis	G5500 4U chassis.
	Chassis management module	Module used by the G5500 for chassis management.
	I/O module	G5500 I/O module. Each module supports two PCIe cards.
	General-purpose compute module	• The G560 V5 and G530 V5 support Intel® Xeon® Scalable (Skylake and Cascade Lake) processors. • The G560 supports Intel®Xeon® E5-2600 V4 processors

Model	Type	Description
	Heterogeneous compute module	• A maximum of 16 NVIDIA Tesla GPU cards (P4, T4, P40, P100, or V100, support a display memory of 12, 16, or 32 GB) • A maximum of 8 double-slot NVIDIA GPU cards (such as A800, A100, and A30 GPU cards)

Table 3-2 shows the compatible G5500 compute nodes and how they can be combined.

Table 3-2 Combinations of G5500 compute nodes

General-Purpose Compute Module	Supported Heterogeneous Compute Module
G560 V5	GP608, GS608
G530 V5	GP316, GP308

NOTE

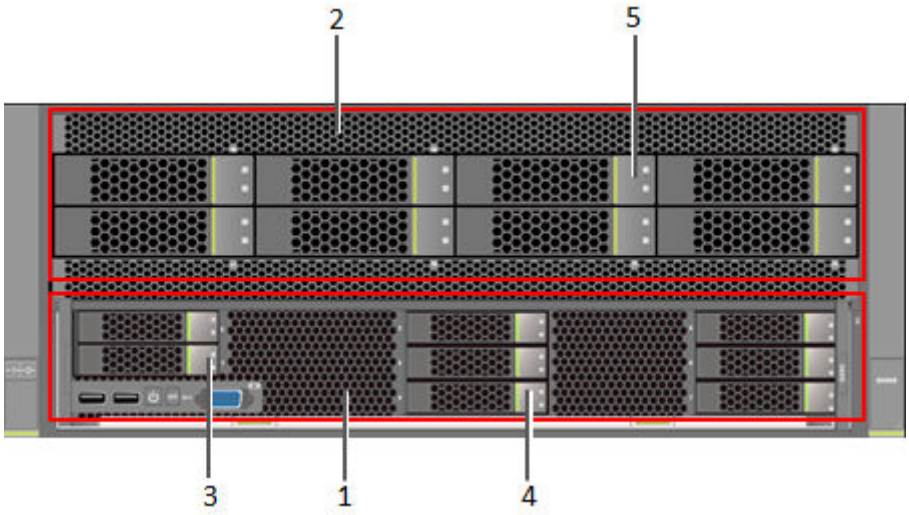
Both the general-purpose compute modules and heterogeneous compute modules do not support mixed configuration of full-width and half-width nodes.

3.2 Appearance

Front Panel

- Full-width configuration 1: Figure 3-1 shows the front panel of a server with full-width nodes (G560 V5 + GP608).

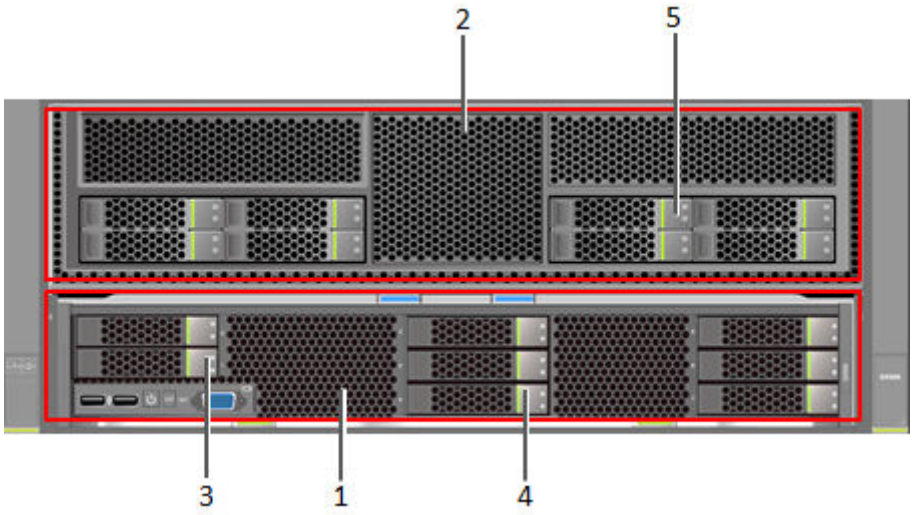
Figure 3-1 G560 V5 + GP608 front panel



1	General-purpose compute module 1 (G560 V5)	2	Heterogeneous compute module 1 (GP608)
3	2.5" SAS/SATA drives or M.2 SSDs	4	2.5" SAS/SATA drives or NVMe SSDs
5	3.5" SAS/SATA drives	-	-

- Full-width configuration 2: **Figure 3-2** shows the front panel of a server with full-width nodes (G560 V5 + GS608).

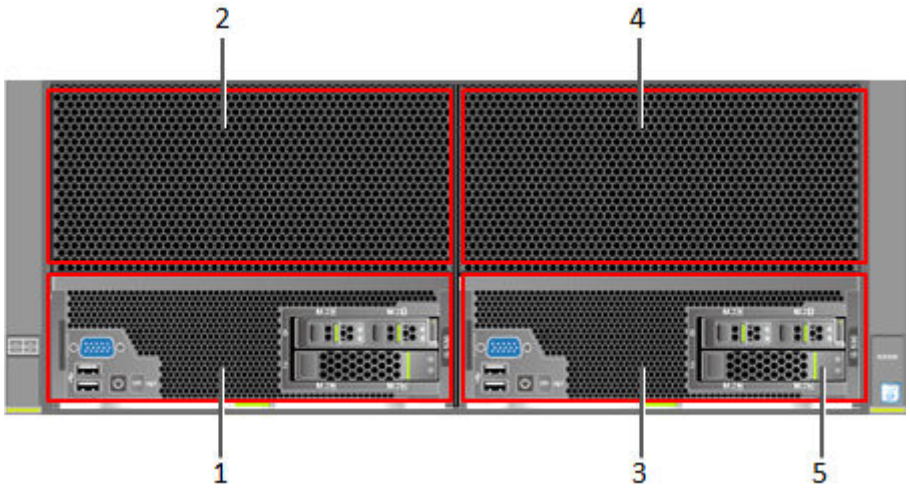
  Figure 3-2 G560 V5 + GS608 front panel



1	General-purpose compute module 1 (G560 V5)	2	Heterogeneous compute module 1 (GS608)
3	2.5" SAS/SATA drives or M.2 SSDs	4	2.5" SAS/SATA drives or NVMe SSDs
5	2.5" SAS or SATA drives	-	-

- Half-width configuration 1: **Figure 3-3** shows the front panel of a server with half-width nodes (G530 V5 + GP316).

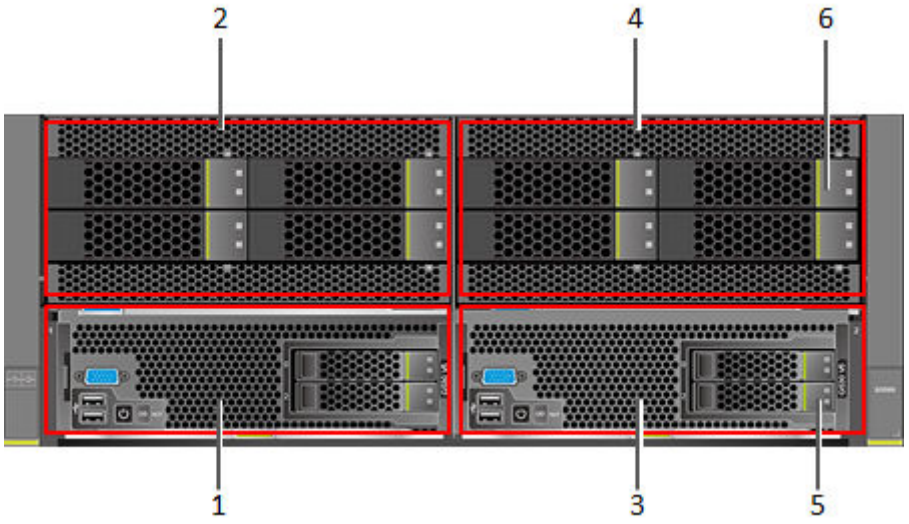
Figure 3-3 G530 V5 + GP316 front panel



1	General-purpose compute module 1 (G530 V5)	2	Heterogeneous compute module 1 (GP316)
3	General-purpose compute module 2 (G530 V5)	4	Heterogeneous compute module 2 (GP316)
5	2.5" SAS/SATA/NVMe drives or M.2 SSDs	-	-

- Half-width configuration 2: **Figure 3-4** shows the front panel of a server with half-width nodes (G530 V5 + GP308).

Figure 3-4 G530 V5 + GP308 front panel



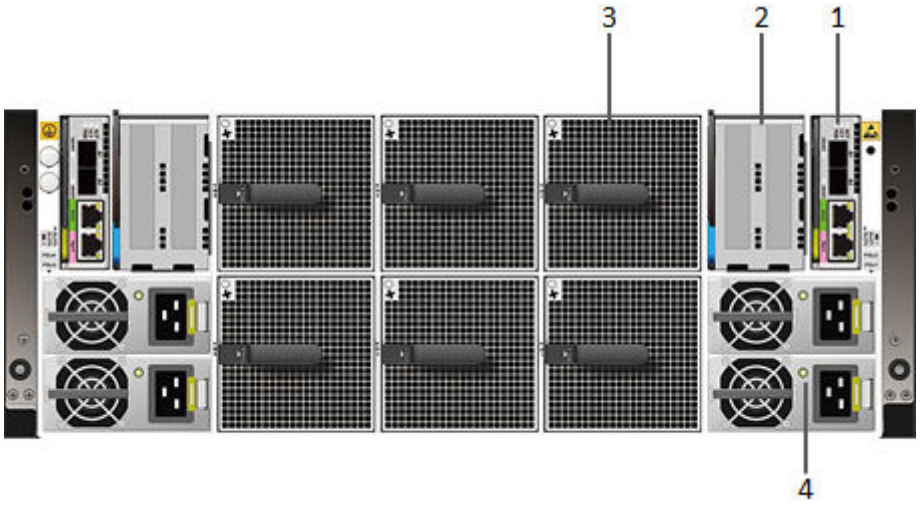
1	General-purpose compute module 1 (G530 V5)	2	Heterogeneous compute module 1 (GP308)

3	General-purpose compute module 2 (G530 V5)	4	Heterogeneous compute module 2 (GP308)
5	2.5" SAS/SATA/NVMe drives or M.2 SSDs	6	3.5" SAS/SATA drives

Rear Panel

- The rear panel of the G5500 chassis consists of chassis management modules, I/O modules, fan modules, and PSUs, as shown in **Figure 3-5**.

Figure 3-5 Rear panel of the G5500



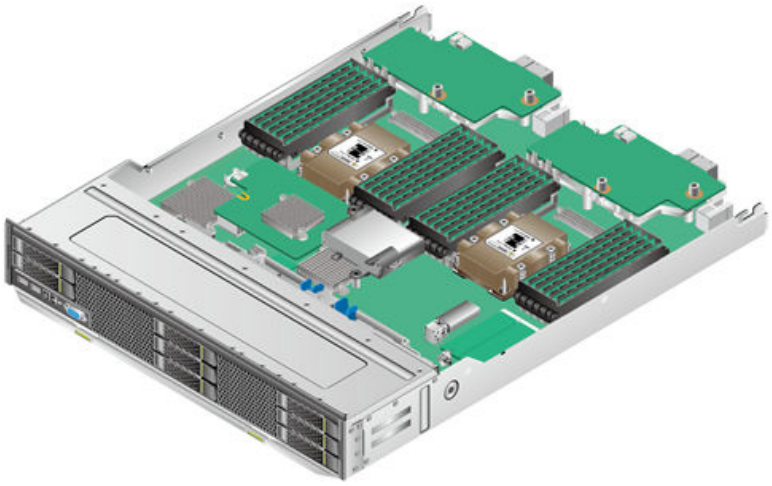
1	Chassis management module	2	I/O module
3	Fan module	4	PSU

3.3 Compute Nodes

The G5500 contains a general-purpose compute module (x86) and a heterogeneous compute module.

- **Figure 3-6** shows a full-width general-purpose compute module G560 V5.

Figure 3-6 G560 V5 appearance

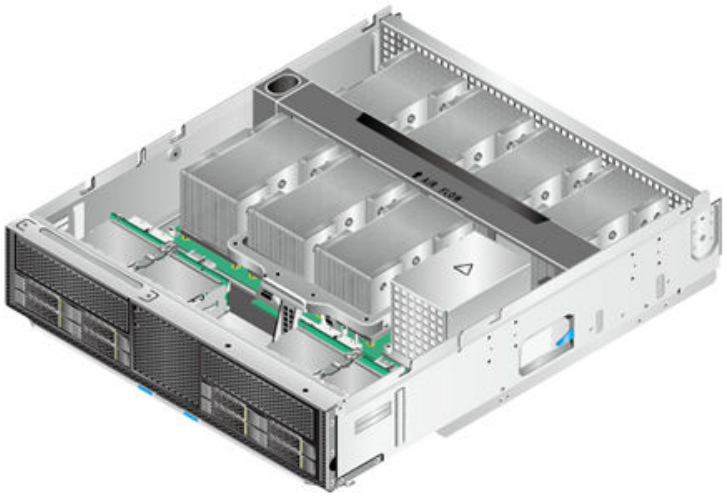


- **Figure 3-7** shows a full-width heterogeneous compute module GP608.

Figure 3-7 GP608 appearance

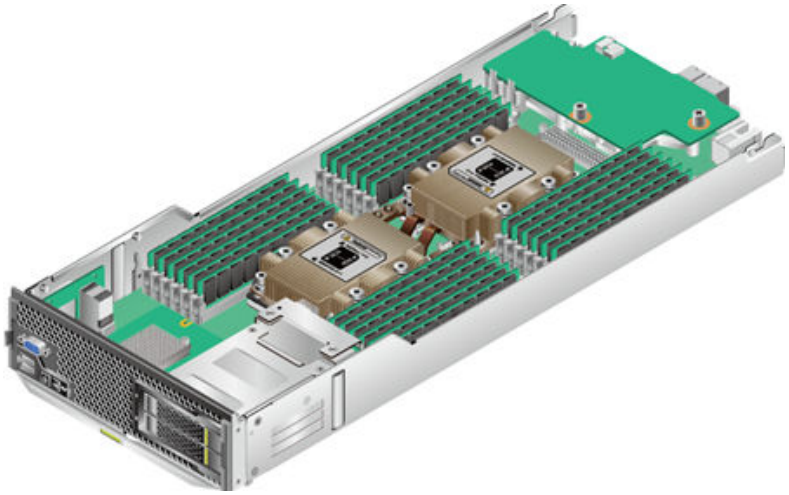
- **Figure 3-8** shows a full-width heterogeneous compute module GS608.

Figure 3-8 GS608 appearance



- **Figure 3-9** shows a half-width general-purpose compute module G530 V5.

Figure 3-9 G530 V5 appearance

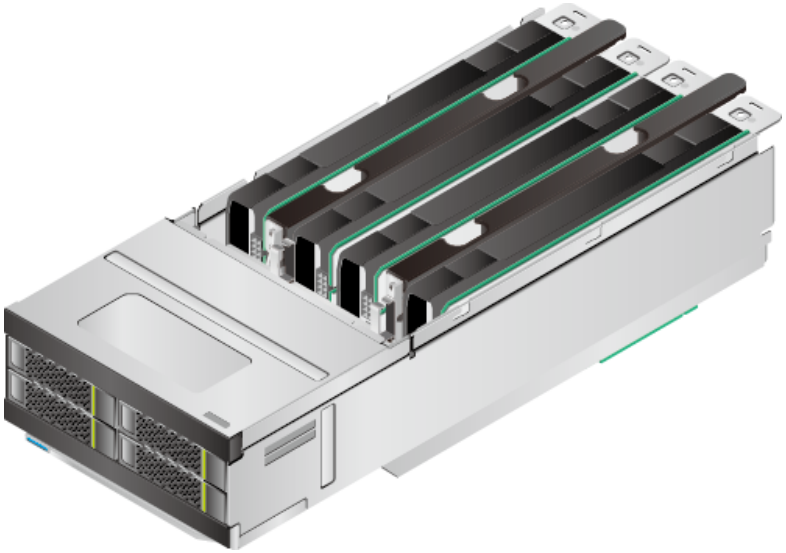


- **Figure 3-10** shows a half-width heterogeneous compute module GP316.

Figure 3-10 GP316 appearance

- **Figure 3-11** shows a half-width heterogeneous compute module GP308.

Figure 3-11 GP308 appearance



NOTE

- For details about full-width compute nodes, see the *G5500 Server G560 V5 Compute Node Technical White Paper*.
- For details about half-width compute nodes, see the *G5500 Server G530 V5 Compute Node Technical White Paper*.

3.4 I/O Modules

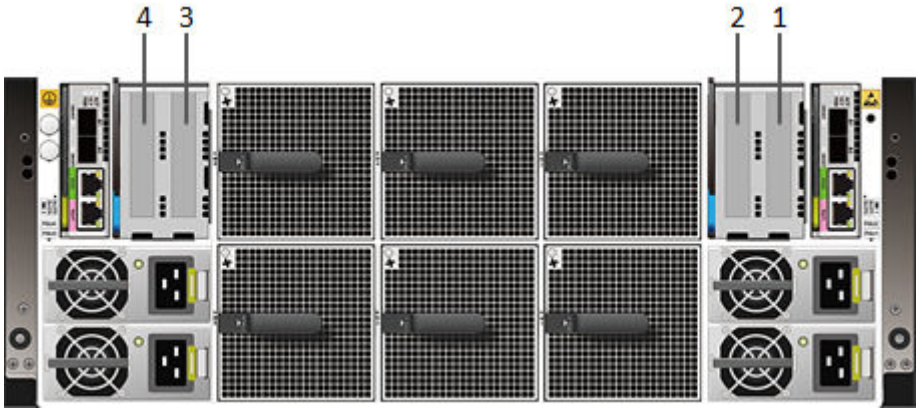
The G5500 provides two I/O modules at the rear of the chassis. Each I/O module provides two HHHL PCIe x16 slots. The entire chassis supports four PCIe slots for installing Ethernet, InfiniBand, and Fibre Channel NICs. For details, see the **Compatibility Checker**.

- **Figure 3-12** shows an I/O module.

Figure 3-12 Appearance of an I/O module

- Each I/O module has two PCIe slots, as shown in **Figure 3-13**.

Figure 3-13 Positions of PCIe slots



1	SLOT 1	2	SLOT 2
3	SLOT 3	4	SLOT 4

- **Table 3-3** lists the mapping between PCIe slots and compute nodes.

Table 3-3 Mapping between PCIe slots and compute nodes

PCIe Slot	Full-Width Node Configuration	Half-Width Node Configuration
SLOT 1	Compute node 1	Compute node 1
SLOT 2	Compute node 1	Compute node 1
SLOT 3	Compute node 1	Compute node 2

PCIe Slot	Full-Width Node Configuration	Half-Width Node Configuration
SLOT 4	Compute node 1	Compute node 2

- Table 3-4 lists the technical specifications of PCIe slots.

Table 3-4 Technical specifications of PCIe slots

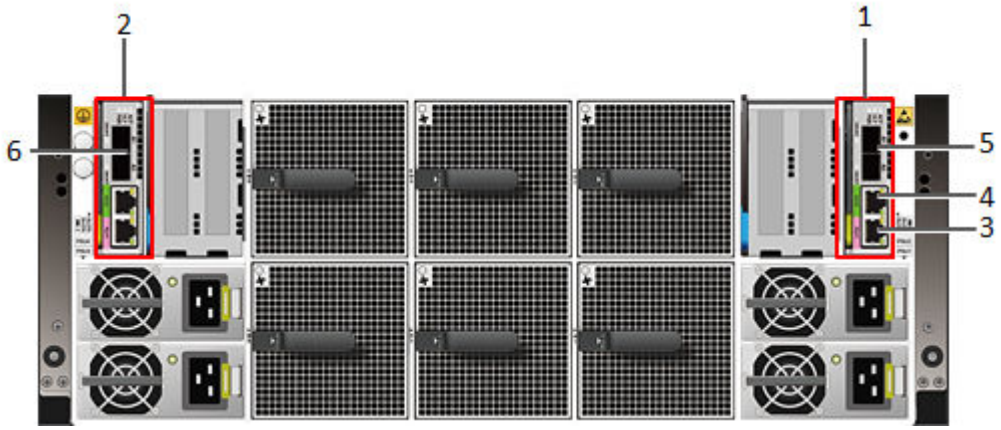
Item	Specifications	Description
Dimensions (H x L)	68.90 mm × 167.65 mm (2.71 in. x 6.6 in.)	Half-height half-length
Maximum power consumption	50 W	Single slot

3.5 Chassis Management Modules

The MM510, chassis management module of the G5500, manages the fans, PSUs, and assets of the chassis. In addition, it converges the management ports of compute nodes and provides unified management ports externally.

Figure 3-14 shows the positions of chassis management modules.

Figure 3-14 Positions of chassis management modules



1	Management module 1 (MM1), mandatory	2	Management module 2 (MM2), optional
3	1000BASE-T management port (Mgmt)	4	Management serial port

5	LOM network port, connecting to compute node 1 in both full-width and half-width configurations	6	LOM network port, which is not connected in full-width configuration and is connected to compute node 2 in half-width configuration.

NOTE

For details about the supported NIC types, see the white paper of each compute node.

The main functions and configuration principles of the chassis management module are as follows:

- MM1: Mandatory. It supports the management port convergence function. It also provides two LOM ports for compute node 1.
- MM2: Optional. It does not support the management port convergence function. It also provides two LOM ports for compute node 2.
- It is recommended that a chassis management module be configured in the MM2 slot only when half-width compute node 2 is configured and LOM ports need to be provided.

Figure 3-15 shows a chassis management module.

Figure 3-15 Chassis management module

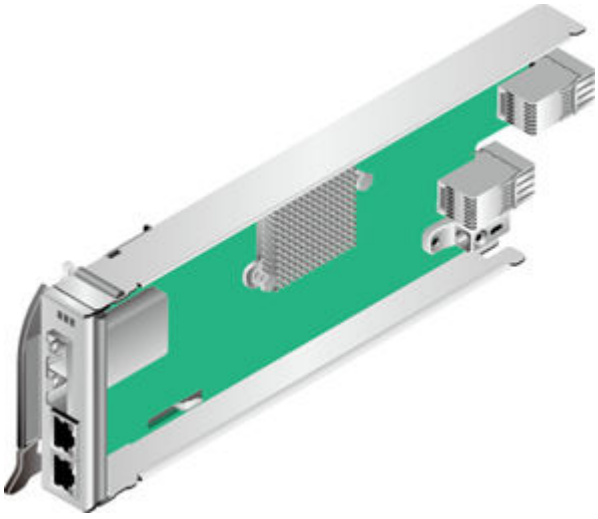
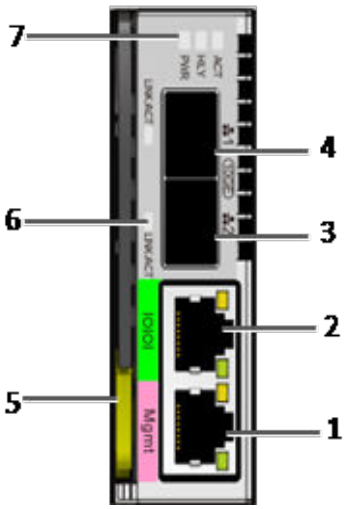


Figure 3-16 shows the panel of the chassis management module.

Figure 3-16 Panel of the chassis management module



1	1000BASE-T management port	2	Management serial port
3	10GE port 2 (LOM)	4	10GE port 1 (LOM)
5	Ejector lever	6	10GE port indicator
7	Management module indicator	-	-

NOTE

- The LOM network interface card (NIC) of the compute node provides external ports through the management module. Replacing the management module affects the LOM function.
- From G560 V5 and G530 V5 series, the LOM function will be supported.

3.6 Fan Modules

The G5500 is equipped with six hot-swappable fan modules. It adopts a design that provides front-to-back ventilation channels to accelerate air flows.

Figure 3-17 shows a fan module.

Figure 3-17 Appearance of a fan module

Table 3-5 describes the indicators on a fan module.

Table 3-5 Indicators on a fan module

Name	Color	State	Description
Operating status indicator	Red or green	Off	There is no power supply.
		Blinking red at 0.5 Hz	The fan module has an alarm.
		Blinking green at 0.5 Hz	The fan module is communicating with the management module properly.
		Blinking green at 4 Hz	The fan module is not communicating with the management module properly.

3.7 PSUs

The G5500 is equipped with four PSUs, which support both AC input and high-voltage DC input. The G5500 currently supports two types of PSUs, 2000 W and 2200 W PSUs. The 2000 W PSU is a newly developed PSU that will gradually replace the 2200 W PSU.

NOTE

Only one type of PSUs can be configured in a chassis.

3.7.1 2000 W PSU

The input and output voltage of a G5500 PSU are as follows:

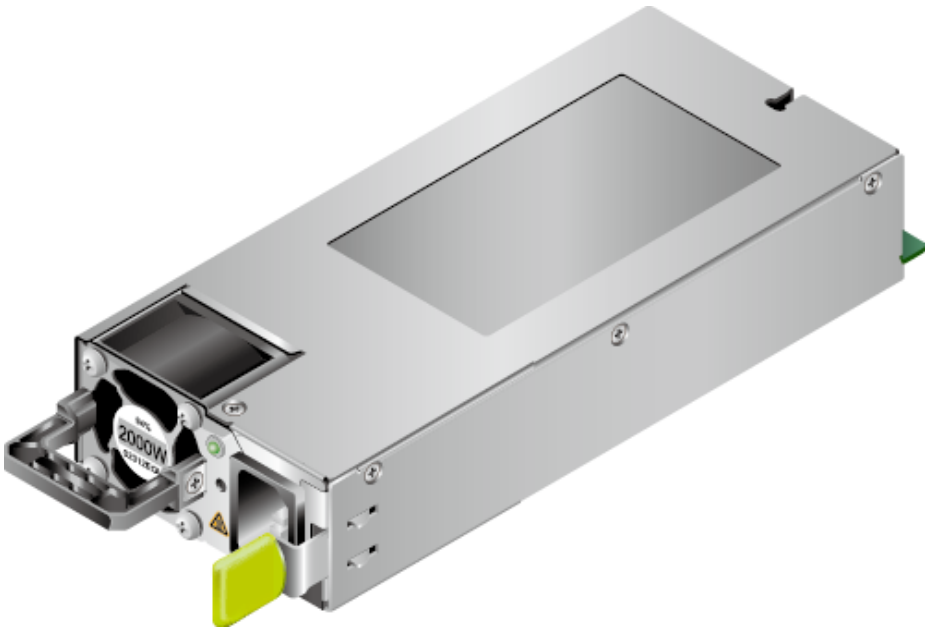
- AC input: 200 V AC to 240 V AC, 2000 W

- AC input: 100 V AC to 120 V AC, 1000 W
- High-voltage DC input: 240 V DC, 2000 W

The output voltage of the PSUs is 12 V DC.

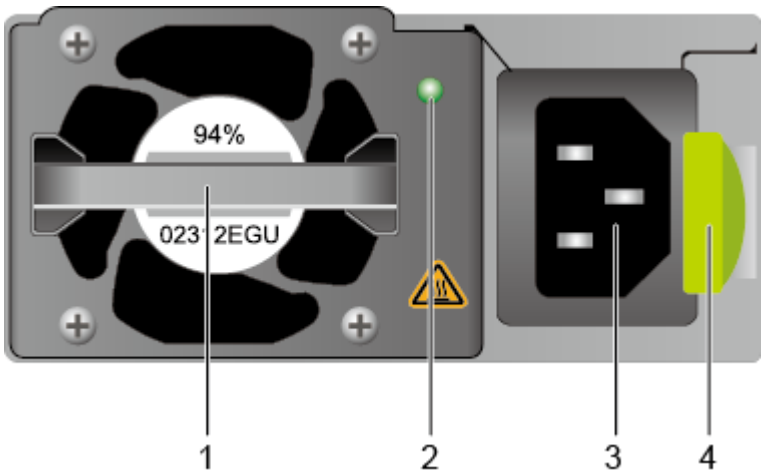
Figure 3-18 shows a PSU.

Figure 3-18 Appearance of a 2000 W PSU



The PSU has a handle, an AC power input socket, an operating status indicator, and a latch, as shown in **Figure 3-19**.

Figure 3-19 PSU panel



1	Handle	2	Indicator
3	C14 power socket	4	Latch

Table 3-6 describes the indicator on an AC PSU.

Table 3-6 Indicator on an AC PSU

Name	Color	State	Description
PSU status indicator	Green/Orange	Off	There is no power supply, or the PSU is faulty.
		Green	The PSU is operating properly.
		Orange	Power protection is enabled and no output power is supplied.

NOTE

If the input current of the 2000 W power supply exceeds 10 A, a 10 A power cable and terminal are required. The PSU figures in this document are for reference only.

3.7.2 2200 W PSU

The input and output voltage of a G5500 PSU are as follows:

- AC input: 200 V AC to 240 V AC, 2200 W
- AC input: 100 V AC to 120 V AC, 1200 W
- High-voltage DC input: 240 V DC, 2200 W

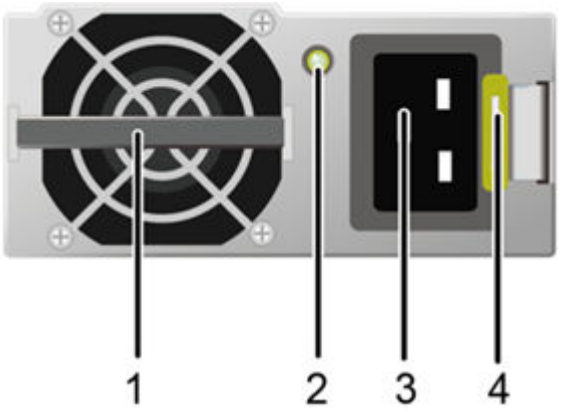
The output voltage of the PSUs is 12 V DC.

Figure 3-20 shows a PSU.

Figure 3-20 Appearance of a 2200 W PSU

The PSU has a handle, an AC power input socket, an operating status indicator, and a latch, as shown in Figure 3-21.

Figure 3-21 PSU panel



1	Handle	2	Indicator
3	C20 power socket	4	Latch

Table 3-7 describes the indicator on an AC PSU.

Table 3-7 Indicator on an AC PSU

Name	Color	State	Description
PSU status indicator	Green/ Orange	Off	There is no power supply, or the PSU is faulty.
		Green	The PSU is operating properly.
		Orange	Power protection is enabled and no output power is supplied.

NOTE

If the input current of the 2200 W power supply exceeds 10 A, a 16 A power cable and terminal are required. The PSU figures in this document are for reference only.

4 Security Management

4.1 TPM

Compute nodes support security solutions based on the Trusted Platform Module (TPM).

TPM is a security chip that complies with the Trusted Computing Group (TCG) TPM specification. The TPM is a hardware-based system security function module. TPM 2.0 provides functions such as data encryption, password protection, authentication, and internal resource protection.

A TPM is installed on a compute node and connected to the compute node over a port. To ensure data security at the hardware level, users are allowed to access a compute node with the TPM only after authorization and authentication and in the specified way.

4.2 Management Module

Security management of the management module consists of hyper management module (HMM) command-line based hierarchical protection, remote Secure Shell (SSH) logins, and Simple Network Management Protocol (SNMP) encrypted authentication.

Command Line-based Hierarchical Protection

When a user logs in to an HMM through an Ethernet port, the HMM authenticates the user to ensure security. Only the user that passes the authentication can log in to the HMM to perform configuration and maintenance.

The HMM uses a hierarchical protection mode for commands. Commands are classified into the following three levels in ascending order: monitoring level, configuration level, and management level. Similarly, login users are also classified into three levels: common user, operator, and administrator. Users can use only the

commands of the levels that are equal to or lower than their own levels. This mechanism effectively controls the authority of login users.

Remote SSH Login

The HMM supports Secure Shell (SSH). SSH provides authentication for user logins and helps defend against various attacks on a non-secure network. The HMM also supports Secure File Transfer Protocol (SFTP), which provides encryption protection for file transfer.

SNMP Encrypted Authentication

The HMM supports SNMPv3 and SNMPv3 Trap encrypted authentications. If the HMM is managed by a network management system through the SNMP protocol, the HMM uses the encrypted authentication mode offered by the user-based security model (USM) to ensure security.

The SNMP version can be SNMPv3 or SNMPv3 Trap. Both versions support the MD5 or SHA authentication modes and the DES or AES encryption modes. The MD5 authentication mode is used by default, which is not secure and easy to be cracked. Therefore, the SHA algorithm is recommended.

NOTE

The HMM also supports the SNMPv1, SNMPv2c, FTP, and Telnet protocols, which are disabled by default for security purposes. If you need to use these functions, contact technical support.

5 Management Features

The G5500 provides the HMM+iBMC dual-layer management architecture:

- Independent node management

  The node integrates the new-generation iBMC intelligent management system to provide highly reliable hardware monitoring and management functions.
- Out-of-band aggregation management

  The G5500 uses the HMM to manage fans, PSUs, and assets in each chassis. The HMM is not directly accessible by users. Its GUI is incorporated on the iBMC. Users can access the iBMC over the GE port on MM1.

5.1 iBMC Features

The iBMC provides the following features:

- KVM and text console redirection
- Remote virtual media
- IPMI V2.0
- SNMPv3
- Redfish 1.0.
- Common information model (CIM)
- Login using a web browser
- Black box

Table 5-1 describes the iBMC features.

Table 5-1 iBMC features

Feature	Description
Support for multiple management ports	Supports integration with any standard management system using the following interfaces: ● IPMI 2.0 ● CLI ● HTTPS ● SNMPv3 ● Web ● Redfish 1.0
Node fault detection	Detects hardware faults and accurately locates faults to the component level.
Node alarm management	Supports alarm management and reports alarms using the SNMP Trap, Simple Mail Transfer Protocol (SMTP), Event, SMTP or syslog service to ensure 24/7 operating.
Integrated virtual KVM	Provides remote maintenance measures to improve the troubleshooting efficiency. A maximum resolution of 1280 x 1024 is supported.
Integrated virtual media	Virtualizes local media devices or images to media devices on a remote server, which simplifies OS installation. A maximum transmission rate of 8 MB/s is supported.
WebUI	Provides a user-friendly GUI that helps simplify users' configuration and query operations. The iBMC WebUI supports the following browsers: ● Internet Explorer 9.0/10.0/11.0 ● Mozilla Firefox 26.0/39.0 ● Chrome 21.0/44.0 ● Safari 8.0 ● JRE 1.7.0 U40 ● JRE 1.8.0 U45
Fault reproduction	Reproduces faults to facilitate fault diagnosis.
Screen snapshots and screen videos	Allows you to view screenshots and videos without login, which facilitates routine preventive maintenance inspection (PMI).
Domain Name Service (DNS)/Active Directory (AD)	Provides the DNS and AD services, which significantly simplify network and configuration management.
Dual-image backup	Starts software from a backup image if the software fails.
IPv6	Supports IPv6 to ensure sufficiency of IP addresses.

5.2 HMM Features

The HMM features are available on the iBMC WebUI. The HMM provides the following features:

- IPMI V2.0
- Login over CLI
- Out-of-band aggregation management
- Heat dissipation and power management

Table 5-2 describes the HMM features.

Table 5-2 HMM features

Feature	Description
Server management port	Supports integration with any standard management system using the following interfaces: ● IPMI 2.0 ● CLI
HMM hot swap	Supports replacement of a failed HMM without powering off the system to quickly restore the HMM back to normal operation.
Chassis asset management	Supports intelligent asset management.
Out-of-band aggregation management	Integrates the out-of-band management of all server nodes in the chassis, which helps reduce cabling in the management network and shorten the maintenance time.

6 Technical Specifications

6.1 Chassis Specifications

Chassis Parameters

Table 6-1 provides the G5500 technical specifications.

Table 6-1 G5500 technical specifications

Item	Description
Form factor	4U modular server
Compute node	Supports one full-width node or two half-width nodes.
PCIe module	• Provides two I/O modules, which support four PCIe x16 slots. • Provides an PCIe smart card module, which supports four PCIe x16 slots.
Chassis management module	Provides a 1000BASE-T management port for external connection and two 10GE LOM ports for compute nodes through MM1.
PSU	Provides four PSUs, which can work in N+N redundancy mode. The PSU specifications are 2000 W or 2200 W. A chassis support only PSUs of the same model.
Fan module	Provides six fans, which can work in N+1 redundancy mode.

Item	Description
Dimensions (W x D x H)	447 mm x 790 mm x 175 mm (17.6 in. x 31.1 in. x 6.89 in.)
Weight	• Net weight: 20 kg (44.1 lb) • Fully-configured: 70 kg (154.35 lb) • Packaging materials: 9.1 kg (20.07 lb)

6.2 Node Specifications

Table 6-2 provides the models and specifications of G5500 server nodes.

Table 6-2 Hardware list

Type	Model	Specifications
Full-width	G560 V5	Two Intel® Xeon® Scalable CPUs, a maximum of 24 DDR4 DIMMs, 6 x 2.5" SAS drives/SATA drives/NVMe SSDs, and 2 x 2.5" SAS/SATA drives.
	GP608	• A maximum of eight full-height dual-slot GPU cards • Eight 3.5-inch SAS/SATA hard disks
	GS608	A maximum of eight SXM2 GPUs and 8 x 2.5" SAS or SATA drives can be configured.
Half-width configuration	G530 V5	• Two Intel® Xeon® Scalable processors • A maximum of 24 DDR4 DIMMs • Two 2.5-inch SAS/SATA hard disks or NVMe SSDs
	GP316	A maximum of 16 HHHL PCIe cards
	GP308	• Four FHFL dual-slot GPUs and 4 x 3.5" SAS or SATA drives • Eight FHHL single-slot GPUs and 4 x 3.5" SAS or SATA drives

6.3 Power Supply and Power Specifications

The G5500 provides four PSU slots. Table 6-3 provides the power supply and power specifications.

Table 6-3 Power supply and power specifications

Power Type	Power Input	Maximum Output Power	Power Cable
2000 W PSU	• 200–240 V AC at 50 or 60 Hz • 240 V DC	Single module: 2200 W 2+2 redundancy: 4400 W 1+1 redundancy: 2000 W	10 A power cables complying with IEC320 C13 and IEC320 C14
	100–120 V AC at 50 or 60 Hz	Single module: 1000 W 2+2 redundancy: 2000 W 1+1 redundancy: 1000 W	
2200 W PSU	• 200–240 V AC at 50 or 60 Hz • 240 V DC	Single module: 2200 W 2+2 redundancy: 4400 W 1+1 redundancy: 2200 W	16 A power cables complying with IEC320 C19 and IEC320 C20
	100–120 V AC at 50 or 60 Hz	Single module: 1200 W 2+2 redundancy: 2400 W 1+1 redundancy: 1200 W	

NOTE

- To ensure the normal running of the system and to support PSU backup, you need to configure sufficient PSUs according to the power calculator to provide sufficient power supply for the system.
- The model referenced by the power calculator is the nominal maximum power consumption of a component or module. The instantaneous overshoot power consumption caused by the EDP feature unique to NVIDIA GPU is not considered in the power calculator. For example, the nominal maximum power consumption of the NVIDIA V100 16 GB dual-slot PCIe card is 250 W, but its 200 μs EDP pulse current is 58 A at 12 V. If the EDP feature is enabled for multiple modules at the same time, the system may not work properly.

6.4 Environmental Specifications

Table 6-4 provides the G5500 environmental specifications.

Table 6-4 Environmental specifications

Item	Description
Temperature	● Operating temperature: 5°C to 35°C (41°F to 95°F) ● Storage temperature (within three months): –30°C to +60°C (–22°F to +140°F) ● Storage temperature (within six months): –15°C to +45°C (5°F to 113°F) ● Storage temperature (within one year): –10°C to +35°C (14°F to 95°F) ● Maximum change rate: 20°C/hour (36°F/hour), 5°C/15 minutes (9°F/15 minutes)
Humidity (RH, non-condensing)	● Operating humidity: 8% to 90% ● Storage humidity (within three months): 8% to 85% ● Storage humidity (within six months): 8% to 80% ● Storage humidity (within one year): 20% to 75% ● Change rate: < 20% RH/hour
Altitude	● ≤ 3048 m (9999.88 ft). For altitudes above 900 m (2952.72 ft), the operating temperature decreases by 1°C (1.8°F) every 300 m (984.24 ft). ● HDDs are not supported when the altitude is higher than 3000 m (9842.4 ft).
Acoustic noise	Noise emissions are measured in accordance with ISO 7999 (ECMA 74) and declared in accordance with ISO 9296 (ECMA 109). The data listed in the following is the declared A-weighted sound power levels (LWAd) and declared average bystander position A-weighted sound pressure levels (LpAm) when the server is operating in a 23°C (73.4°F) ambient environment: Idle: ● LWAd: 6.7 Bels ● LpAm: 50 dBA **NOTE** The actual sound levels generated when the server is operating vary depending on the server configuration, workload, and ambient temperature.
Vibration	● One cyclical sweep in each axial direction at the rate of 0.1 oct/min, a total of three axial directions ● 5 Hz to 10 Hz: 5 mm (0.2 in., peak-to-peak value) ● 10 Hz to 100 Hz: 1 m/s2
Shock	Half sine wave, peak acceleration of 2 g (g-force), 11 ms, 3 times for each surface, a total of three axial directions

Item	Description
Corrosive gaseous contaminant	• Copper corrosion rate test requirements: The corrosion product thickness growth rate is lower than 300 Å/month (meeting G1 requirements of the ANSI/ISA-71.04-2013 standard on gaseous corrosion). • Silver corrosion rate test requirements: The corrosion product thickness growth rate is lower than 200 Å/month.
Particle contaminant	• In compliance with ISO 14664-1 Class 8 requirements A professional organization is required to monitor particle contaminants in the equipment room. • No explosive, conductive, magnetic conductive, and corrosive dust
Power consumption	The power consumption parameters vary with server configurations, including the configurations complying with EU's energy-related products (ErP) requirements. For details, see the **Power Calculator**.

7 Maintenance and Warranty

For details about the maintenance policy, visit **Customer Support Service**.

For details about the warranty policy, visit **Warranty**.

8 Certifications

Region	Country	Certification	Certification Mark
China	China	CCC	
		RoHS	
Europe	European Union	CE	
		REACH	NA
		WEEE	
	Russia	EAC&GOST	
	UK	UKCA	NA
North America	US	FCC	This device complies with Part 15 of the FCC Rules.Operation is subject to the following two conditions: (1) this device may not cause harmful interference,and (2) this device must accept any interference received,including interference that may cause undesired operation.
	Canada	IC	CAN ICES-3(A)/NMB-3(A)
Asia Pacific	Japan	VCCI	
Global	IECEE members	CB	NA

9 Waste Product Recycling

If product users need product recycling service provided by xFusion after products are scrapped, contact technical support for services.

A Appendix

A.1 Acronyms and Abbreviations

A	
AC	Alternating Current
AES NI	Advanced Encryption Standard New Instruction Set
AI	Artificial Intelligence
ARP	Address Resolution Protocol
ASIC	Application-Specific Integrated Circuits
AVX	Advanced Vector Extensions
B	
BBU	Backup Battery Unit
BMC	Baseboard Management Controller
C	
CD	Calendar Day
CIM	Common Information Model
CLI	Command-line Interface
D	
DC	Direct Current
DDR4	Double Data Rate 4
DEMT	Dynamic Energy Management Technology
DIMM	Dual In-line Memory Module
DVD	Digital Video Disc

E	
ECC	Error Checking and Correcting
ECMA	European Computer Manufacturers Association
EDB	Execute Disable Bit
EN	European Efficiency
ETS	European Telecommunications Standards
F	
FC	Fiber Channel
FPGA	Field-Programmable Gate Array
FTP	File Transfer Protocol
G	
GE	Gigabit Ethernet
GPIO	General Purpose Input/Output
GPU	Graphics Processing Unit
H	
HDD	Hard Disk Drive
HMM	Hyper Management Module
HPC	High-performance Computing
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
HVDC	High Voltage Direct Current
I	
ICMP	Internet Control Message Protocol
IDC	Internet Data Center
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
iBMC	Integrated Baseboard Management Controller
IO	Input/Output
IOPS	Input/Output Operations per Second
IP	Internet Protocol
IPC	Intelligent Power Capability
IPMB	Intelligent Platform Management Bus

IPMI	Intelligent Platform Management Interface
K	
KVM	Keyboard Video and Mouse
L	
LC	Lucent Connector
LDIMM	Load Reduced DIMM
LED	Light Emitting Diode
LOM	LAN on motherboard
M	
MAC	Media Access Control
N	
NBD	Next Business Day
NC-SI	Network Controller Sideband Interface
NVMe	Non-Volatile Memory Express
O	
OS	Operating system
P	
P2P	Peer to Peer
PCIe	Peripheral Component Interconnect Express
PHY	Physical Layer
PID	Proportional–Integral–Derivative
PMBUS	Power Management Bus
POK	Power OK
PWM	Pulse-width Modulation
Q	
QPI	QuickPath Interconnect
R	
RAID	Redundant Array of Independent Disks
RDIMM	Registered Dual In-line Memory Module
RDMA	Remote Direct Memory Access
RJ45	Registered Jack 45
S	

SAS	Serial Attached Small Computer System Interface
SATA	Serial Advanced Technology Attachment
SGMII	Serial Gigabit Media Independent Interface
SMTP	Simple Mail Transfer Protocol
SM_CLP	Server Management Command Line Protocol
SNMP	Simple Network Management Protocol
SSD	Solid-state Drive
T	
TACH	Tachometer signal
TBT	Turbo Boost Technology
TCG	Trusted Computing Group
TDP	Thermal Design Power
Telnet	Telecommunication Network Protocol
TXT	Trusted Execution Technology
TFTP	Trivial File Transfer Protocol
TPM	Trusted Platform Module
U	
UEFI	Unified Extensible Firmware Interface
UID	Unit Identification Light
UL	Underwriter Laboratories Inc.
UPI	UltraPath Interconnect
USB	Universal Serial Bus
V	
VGA	Video Graphics Array
VRD	Voltage Regulator-Down

