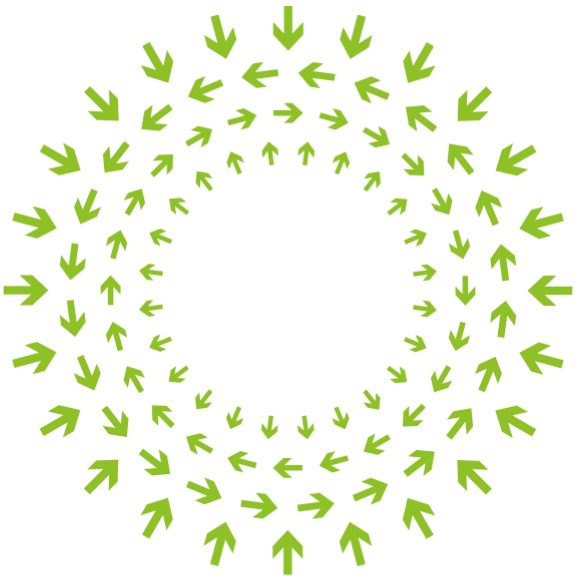


OPEN Compute Project

Telco enhanced Open Rack Server variant

Revision 1.0

Author: Samuli Toivola, Nokia

License (OCP CLA Option)

Contributions to this Specification are made under the terms and conditions set forth in **Open Compute Project Contribution License Agreement ("OCP CLA") ("**Contribution License") by:
Nokia Solutions and Networks OY

You can review the signed copies of the applicable Contributor License(s) for this Specification on the OCP website at http://www.opencompute.org/products/specsanddesign

Usage of this Specification is governed by the terms and conditions set forth in **Open Compute Project Hardware License – Permissive ("OCPHL Permissive"), ("Specification License").**

You can review the applicable Specification License(s) executed by the above referenced contributors to this Specification on the OCP website at http://www.opencompute.org/participate/legal-documents/

Note: The following clarifications, which distinguish technology licensed in the Contribution License and/or Specification License from those technologies merely referenced (but not licensed), were accepted by the Incubation Committee of the OCP: None.

Revision history

Revision	Editor	Description
V0.1 (Sep 18, 2018)	S. Toivola	First draft
V1.0 (Dec 11, 2018)	S. Toivola	Corrected document title

Contents

1 Introduction

Nokia has designed OCP Open Rack V2 compatible server variant optimized to telco NFV use cases.

New server design solves challenges to meet various customers CO site specific requirements like EMI shielding and ETSI environmental requirements.

Server design enables flexible and effective configurations for various networking and storage focused use cases.

Enhancements to NUMA performance issue have implemented by PCIe routing to CPU1 from riser slot and to storage performance with redundant NVMe SSD’s.

Open Rack Server variant is a design based on the Intel® Purley platform, including dual Skylake processors and a Lewisburg PCH chipset. The motherboard is designed to be fitted into an OCP Open Rack V2 compatible server sled. The server sled is compatible with the Cubby shelf and is compatible with the OCP rack design, Open Rack V2 (ORv2).

This document is the hardware specification of server sled, and is used to describe, at a high level, the hardware implementation details. Some features other than hardware, such as the mechanical, software, and system management are also briefly covered.

2 Hardware Structure

The following table lists in tabular format the major hardware components:

Form Factor	Board dimensions 6.5”x20”
Processor	Dual Intel® Skylake Server processors
Memory	Total 12 DIMMs DDR4 6 channels per CPU, 1 DIMM per channel Support RDIMM (SR and DR)
Chipset	Intel Lewisburg PCH (C621)
Gigabit Ethernet	1 x Shared-GbE RJ45 from Intel I210, driven by BMC through NC SI and SMBus
Serial ATA	Four SATA3 ports. Supported storage interfaces, M.2 and U.2. Two ports for each interface Advanced Host Controller Interface (AHCI) support
LPC interface	Low Pin Count (LPC) interface support
SPI	Serial Peripheral Interface, dual BIOS support. BMC is capable to perform BIOS remote upgrade and recovery
TPM	Infineon SLB9670 TPM compliant with rev 1.2/2.0 standard
BMC	ASPEED AST2500
PCIe slot	One PCIe Gen3 x32 slot, which holds x32 PCIe Gen 3 signals, x16 from CPU0 and x16 from CPU1. The slot supports two type riser cards: Type A Riser: One slot type. Slot1 is PCIe x16 connector with x16 PCIe signal from CPU0 and support FHHL PCIE card. Type B Riser: Two slots type. Bottom slot (Slot2) is x16 PCIe connector with x16 PCIe signal from CPU0 and support FHHL PCIe card. TOP slot (Slot3) is x16 PCIe connector with x16 PCIe signal from CPU1 and also support FHHL PCIe card.
OCP Mezzanine	Two PCIe x8 OCP Mezzanine Card specification 2.0 compliant connectors with dual x8 PCIe Gen 3 signal from CPU0.
USB	One USB 3.0/2.1 compliant port on the front panel
mini-B USB	One mini-B USB port on the front panel for BMC debug and in-band firmware update
M.2	Two M.2 connectors. Each connector supports one PCIe x1 and one optional PCIe x1 / SATA. Support form factor 2280 single-side and double-side M.2 module.

U.2	Two U.2 connectors. Each connector supports one SATA and one PCIe x4.

Table 1 Server Sled Hardware Component Summary

2.1 Major Hardware Components

Open Rack Server is a dual processor server board. It consists of the Intel® Skylake server processor product family, used in conjunction with the Lewisburg PCH. The reference design is the Intel® Crescent City Customer Reference board, optimized for the telco NFV use cases.

Each Skylake processor supports 6 RDIMMs. The master processor (CPU0) supports one PCIe Gen3 x16 interface for the riser slot and two OCP mezzanine interfaces. The slave processor (CPU1) supports one PCIe Gen3 x16 interface for the riser slot and two PCIe Gen3 x8 interface for the middle plane.

The PCH supports two M.2, two U.2, SATA3, and one USB 3.0 ports. One LPC interface works with BMC controlled multiplexer to support dual BIOS.

One ASPEED AST2500 BMC for system management. It covers the power sequence control, hardware monitoring, and thermal control.

One 1000 Base-T Gigabit Ethernet LOM port with RJ-45.

For the security support, the server sled hardware design supports the BIOS secure boot and the TPM for security options.

The figure below shows the Server system block diagram.

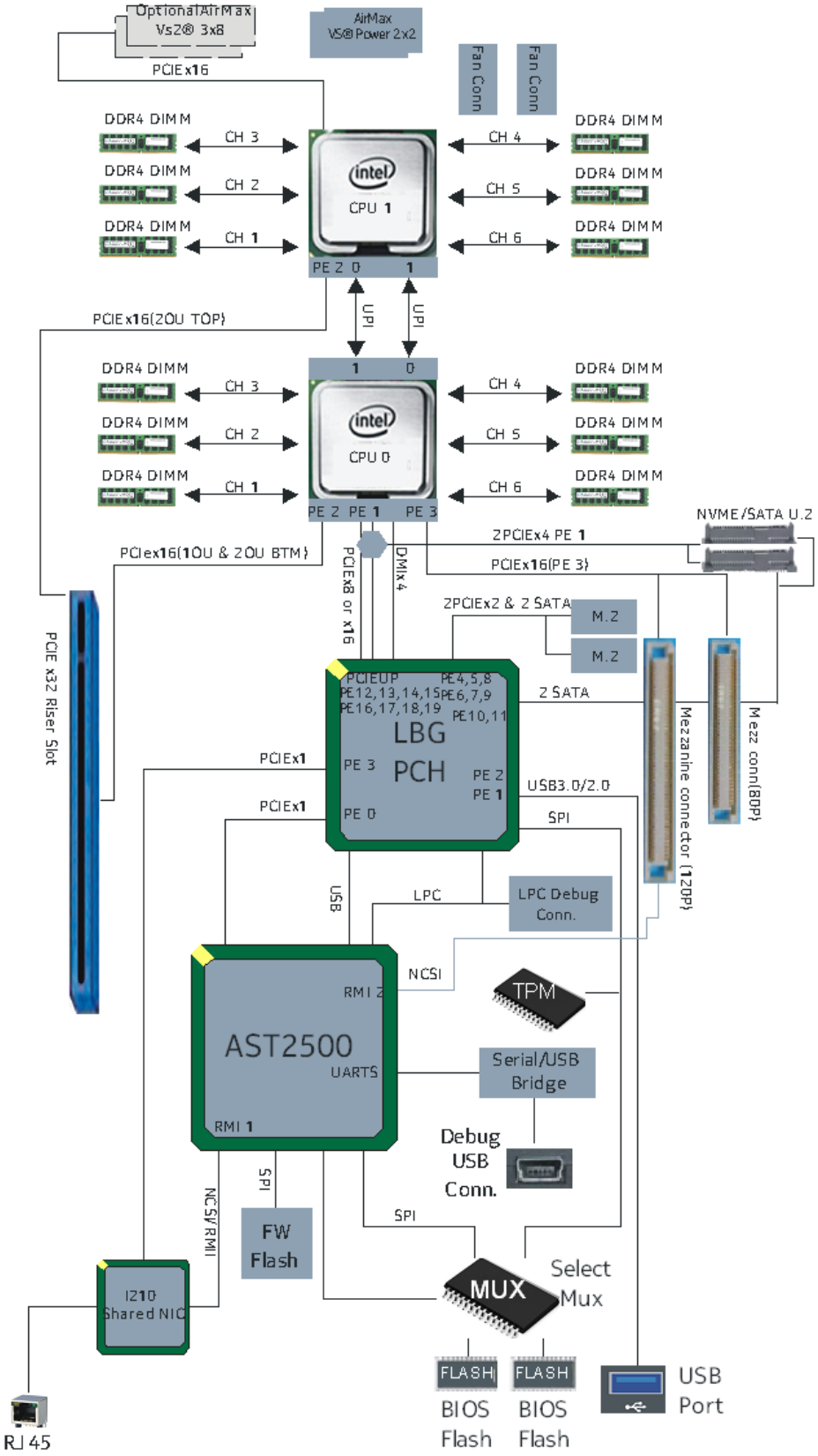

OptionalAirMax
Vs2® 3x8
AirMax
VS® Power 2x2
PCIEx16
Fan Conn
Fan Conn
DDR4 DIMM
CH 3
CH 2
CH 1
CH 4
CH 5
CH 6
CPU 1
PE 2 0 1
UPI
UPI
PCIEx16(2OU TOP)
CPU 0
1 0
PE 2 PE 1 PE 3
NVME/SATA U.2
2PCIEx4 PE 1
PCIEx16(PE 3)
PCIex16(1OU & 2OU BTM)
PCIEx8 or x16
DMIx4
2PCIEx2 & 2 SATA
M.2
M.2
PCIE x32 Riser Slot
PCIEUP
PE4,5,8
PE12,13,14,15
PE6,7,9
PE16,17,18,19
PE10,11
LBG
PCH
2 SATA
PCIEx1
PE 3
PE 2
PE 1
PCIEx1
PE 0
USB3.0/2.0
SPI
Mezzanine connector (12OP)
Mezz conn(8OP)
USB
LPC
LPC Debug
Conn.
NCSI
RMII 2
AST2500
UARTS
RMII 1
TPM
Serial/USB
Bridge
Debug
USB
Conn.
SPI
SPI
NC5/RMII
FW
Flash
I210
Shared NIC
MUX
Select
Mux
FLASH
FLASH
BIOS
Flash
BIOS
Flash
USB
Port
RJ 45

Figure 1 Server Sled Block Diagram

2.1.1 Processor

The Skylake processor is the next generation of 64-bit, multi-core server processor built on 14-nm process technology. The processor socket type is the LGA3647-0 SMT socket. The package is a 43.18 x 50.24 mm Flip-Chip Land Grid Array (FC-LGA14).

The processor TDP can be up to 165W. Concerning the thermal design capability, the following SKU's processors are available with Open Rack Server:

No.	Vendor	Model	Name	Core	Frequency	Watt
1	Intel	4116	Skylake	12 Core	2.1GHz	85W
2	Intel	5118	Skylake	12 Core	2.3GHz	115W
3	Intel	6130	Skylake	16 Core	2.1GHz	125W
4	Intel	6138	Skylake	20 Core	2.0GHz	125W
5	Intel	6140	Skylake	18 Core	2.3GHz	140W
6	Intel	8160	Skylake	24 Core	2.1GHz	150W
7	Intel	4108	Skylake	8 Core	1.8GHz	85W
8	Intel	4114	Skylake	10 Core	2.2GHz	85W
9	Intel	5115	Skylake	10 Core	2.4GHz	85W
10	Intel	6148	Skylake	20 Core	2.4GHz	150W
11	Intel	8176	Skylake	28 Core	2.1GHz	165W
12	Intel	6130T	Skylake	16 Core	2.1GHz	125W
13	Intel	6142	Skylake	16 Core	2.6GHz	150W
14	Intel	3104	Skylake	6 Core	1.7GHz	85W
15	Intel	6150	Skylake	18 Core	2.7GHz	165W

Table 2 Intel Skylake Server Processor SKU's

Single processor mode is supported.

The following table summarizes the Skylake Server processor features.

Features	Description
Memory: Technology Support	DDR4 ECC RDIMM DDR4 ECC LRDIMM DDR4 ECC LRDIMM 3DS

Memory: Speeds	DDR4: 2666, 2400, 2133 and 1866 MT/s
Memory: Max DIMMs per Socket	6
Memory: Max Capacity (w/256 GB DIMM) per Socket	1536 GB
Intel® UPI Interfaces per Socket	2
Intel® UPI Speeds	10.4 GT/s
PCI Express Lanes per Socket	48 Gen3 lanes (plus 4 DMI3 lanes)
Peripheral Controller Hub (PCH)	Intel® LBG Series Chipset (Lewisburg)
PECI	V3.1 compliant
TPM	V1.2 and V2.0 compliant

Table 3 Intel Skylake Server Processor Features

Below subsections describe the thermal reporting and protection schemes.

2.1.1.1 Platform Environmental Control Interface for thermal reporting

The Skylake CPU has an on-die Digital Thermal Sensor (DTS) for monitoring the processor die temperature. The DTS can be retrieved through S/W interface and H/W interface. The S/W interface is the processor Model Specific Register (MSR). The H/W interface is the Platform Environmental Control Interface (PECI). On server sled the PECI is implemented for accessing the DTS.

The PECI is a one-wire interface that provides a communication channel between a PECI client and a PECI master.

For a platform with BMC present, there are two different PECI connectivity options offering the PECI access to the BMC.

- PECI Proxy service of the Management Engine: PECI doesn't route to the BMC, but the BMC requires routing to the PCH SMLink0.
- PECI direct access from the BMC : PECI routes to the BMC instead of the PCH.

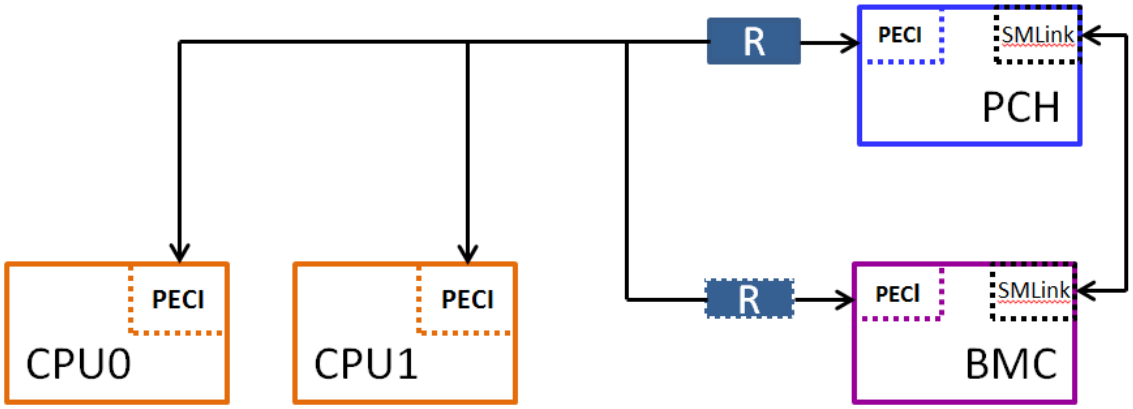


Figure 2 PECI topology on server sled

The default PECI connection is routed to the PCH. The BMC accesses the PCH SMLink interface to get the thermal sensor data from CPU, DIMM, and PCH.

2.1.1.2 Processor Thermal Protection

Basically, the processor thermal protection scheme is implemented with pure hardware level since Skylake. For previous generation processors, BIOS needs to configure the thermal monitor register such as the TM1 and TM2. Now it isn't required. The Skylake processor will select an appropriate scheme to use on a dynamic basis.

The processor has two internal thermal control signals for thermal protection. One signal is the PROCHOT#, and the other signal is THERMTRIP#.

PROCHOT#

The PROCHOT# is a bi-directional signal. It's connected between CPU and BMC. The following cases assert the PROCHOT#.

- When one or more cores of the processor reach the maximum safe operating temperature, the PROCHOT# is activated by CPU. An interrupt is generated. The interrupt is connected internally to the processor integrated local APIC controller. In this case, the PROCHOT# is sent to BMC just for status information and to make design flexible in case PROCHOT# is needed to generate NMI.
- BMC thermal sensors alarm. The BMC asserts the PROCHOT# to inform CPU speed down.
- CPU voltage regulator overheat. The regulator asserts the PROCHOT# to inform CPU speed down.

THERMTRIP#

When CPU's temperature goes to a catastrophically high point, its internal thermal protection mechanism activates the THERMTRIP# signal. The THERMTRIP# indicates one of two

possible critical over temperature conditions. One is the processor junction temperature reaches a level that the permanent silicon damage may occur, the other is the system memory interface exceeds a critical temperature limit set by BIOS.

The THERMTRIP# is routed to the CPLD. If the THERMTRIP# is asserted, the CPLD will follow the power off sequence to turn all the power rails off to prevent CPU from eternal damage.

The CPLD also re-directs the THERMTRIP# to the BMC. Once the THERMTRIP# is activated, the BMC will know the status and keep the event logged.

2.1.1.3 Use of Power Management States

The Purley platform supports the power management states such as P-States, C-States, and S-States. They are all defined by the ACPI. Brief introduction as below.

P-States

The purpose of P-state is to optimize the CPU performance and the power efficiency. In practice, it means the Intel® SpeedStep technology.

Both the CPU running frequency and the operating voltage can be changed to vary the performance and the power consumption. Each P-state assigns a CPU frequency and voltage operating point. Configure the P-states transition properly to get the optimal performance and power efficiency.

The software, mainly the OS and BIOS, has the permission to make P-state change. The software monitors the loading status and change the CPU frequency by writing to the CPU MSRs. Based on the selected frequency and the number of active processor cores, the voltage is regulated accordingly.

P-states transition rule is summarized as below.

- If the target frequency is higher than the current frequency, the voltage is ramped up in steps to an optimized voltage. This voltage is signaled by the SVID bus to the voltage regulator. Once the voltage is established, the PLL locks on to the target frequency.
- If the target frequency is lower than the current frequency, then the PLL locks to the target frequency, then transitions to a lower voltage by signaling the target voltage on the SVID bus.
- Software-requested transitions are accepted at any time. If a previous transition is in progress, then the new transition is deferred until the previous transition is completed.

The number of P-states will vary depending on the CPU SKU. P1 means the highest speed. The other P-states are the descending of P0 with a defined frequency stepping. For Skylake, the stepping is 100MHz.

NOTE : P0 is also a P-state, but it represents the turbo frequency. Only few SKUs support the Turbo mode.

C-States

ACPI defines the low-power idle states as C-state when the CPU is going to be idle. The C-states are the CPU SKU specific as well. The normal operating mode is C0. More power saving actions are taken for numerically higher C-states. In lower C-states the processor stops

executing and power is saved by turning different parts of the processor off. The lower the state, the higher latency comes back to C0.

S-States

ACPI defines the system level power states as the S-states. There are seven S-states : S0 ~ S6. The S-states definition are quite common for PC industry. S0 means system on. S1 means power on suspend. S3 means Suspend to RAM. S5 means soft off.

All the S-states transitions are initiated by the CPH

P-States, C-States, S-States combinations

The following table shows the combinations of different P-states, C-states, and S-States.

Sleep (S) State	Processor Core (C) State	System Clocks	Description
S0	C0	On	CPU On, P-states change per application
S0	C1 or C1E	On	CPU Auto-Halt
S0	C3	On	CPU Deep Sleep
S0	C6 or C7	On	CPU Deep Power Down
S1	C6 or C7	On	Power on Suspend
S3	Power off	Off, except RTC	Suspend to RAM
S4	Power off	Off, except RTC	Suspend to Disk
S5	Power off	Off, except RTC	Soft Off

Table 4 Different Power States Combinations

2.1.2 System Memory Interface

The Intel® Skylake processor product family memory interface supports the DDR4 DIMM technology. The supported features are listed below.

Type	Ranks Per DIMM and Data Width	DIMM Capacity (GB)	
		DRAM Density	
		4Gb	8Gb
RDIMM	SRx4	8GB	16GB
RDIMM	SRx8	4GB	8GB
RDIMM	DRx8	8GB	16GB
RDIMM	DRx4	16GB	32GB
RDIMM 3DS	QRx4	N/A	2H-64GB
	QRx8	N/A	4H-128GB
LRDIMM	QRx4	32GB	64GB
LRDIMM 3DS	QRx4	N/A	4H-164GB
	8Rx4	N/A	64GB

Table 5 Supported DDR4 DIMM Type List

- Maximum DIMMs per socket: six
- Each channel consists of 64 data and 8 ECC bits
- Supported data rate : 1866, 2133, 2400, 2666MT/s
- Unbuffered DIMM is not supported

NOTE : 3DS means 3D stack. 2H or 4H means the stack counts by using the through-silicon vias(TSV).

2.1.3 Lewisburg PCH

Open Rack Server incorporates the Lewisburg PCH SKU LBG-1G(C621) for extensive I/O support.

Functions and capabilities include:

Supported I/O and features	Description
DMI3	PCIe Gen3 lane x4, interface between CPU and PCH
PECI	Sideband signals between CPU and PCH
PCIe Gen 3	Root port : 20 lanes Downsteam port : 24 lanes
SATA3	AHCI mode support, RAID mode support Support 6 ports on server sled
USB 3.0	One Standard-A USB 3.0 port on the front panel
SPI	Support 3 devices at most, including two flash and one TPM
LPC v1.1	24 MHz, No DMA
SMBus v2.0	Be a SMBus host and provide up to 6 SMLink interface
ACPI v4.0a	Advanced power management support
Real Time Clock	Operate with 32.768KHz crystal to count date and time. 256 bytes of battery-backed RAM to store system data
GPIO	

Table 6 Lewisburg PCH I/O Features

Below subsections describe the LPC, SPI, and SMBus topology. They are concerning the system management scheme.

2.1.3.1 LPC Topology

. BMC and a BIOS postcode indicator are located on the LPC bus. The indicator is for debug purpose only.

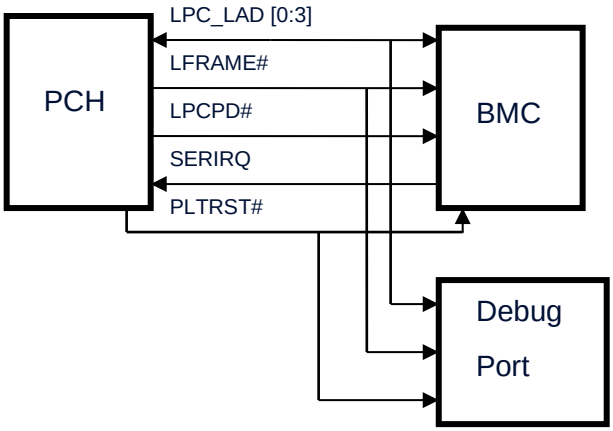


Figure 3 LPC bus topology on server sled

2.1.3.2 SPI Topology

The PCH has a SPI bus and supports up to two SPI compatible flash devices.

Open Rack Server is equipped with two boot flashes located on the SPI interface, where BIOS image and ME firmware are stored. The two flash devices keep duplicated BIOS image and ME firmware for redundancy.

ME means management engine. It's a hardware-based management scheme initiated by Intel. ME can monitor and maintain the system through out-of-band connection. This means ME can work without the presence of OS or locally installed management agent. ME needs to load firmware to work and the firmware saves in the BIOS flash.

Below a figure shows the connection between the PCH, BMC, and two BIOS flashes. There is a multiplexer to switch the BIOS flashes to the PCH or the BMC. The multiplexer is controlled by the BMC.

For normal use, the BMC controls the multiplexer to switch the BIOS flashes to the PCH SPI bus. The PCH can load BIOS code for boot up, or access the flash content for application.

For BIOS remote upgrade, the BMC controls the multiplexer to switch the BIOS flashes to the BMC SPI bus. The BMC can control the relative chip select (CS#) to choose the flash for upgrade.

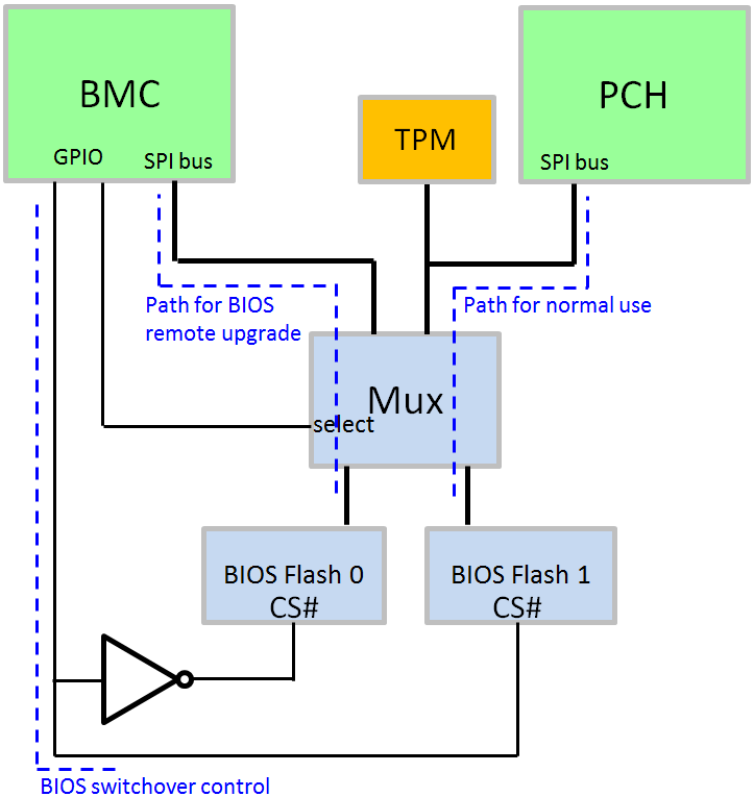


Figure 4 SPI Topology for Dual BIOS Support

Redundancy

Open Rack Server saves the same copies of BIOS image and ME firmware on two separate SPI flashes. Each copy of the BIOS portion consists of boot block and main BIOS block. In addition, the flash also contains vendor specific data and setup data. BIOS vendor shall specify and document the exact organization (memory map) of the SPI flashes.

It is controllable to set either flash 0 or flash 1 as the active BIOS flash while the other flash becoming backup. The flash selection is controlled by BMC. The SW spec details how the active flash is selected. Once selected, the active flash keeps active over any kind of resets but is defaulted to flash 0 after power down.

BMC includes two watchdog timers to monitor the BIOS boot up status. One is the FRB2 WDT, the other is the POST WDT. The FRB2 WDT monitors the BIOS code loading and decompression status. The POST WDT monitors the BIOS code running status.

When BIOS code execution starts, the FRB2 WDT is enabled and started automatically by the BMC. BIOS boot block verifies code integrity by checksum calculation. If the active BIOS boot block is found data missing or corrupted, the FRB2 WDT timeout. BMC regards the timeout as a BIOS code loading failure and switchover to the backup flash. BMC makes BIOS switchover by triggering CS# pin to set the relative flash active. After that, BMC sends a reset to reboot system.

If BIOS code loads and starts to run successfully, BIOS must send commands to BMC to disable the FRB2 WDT and enable the POST WDT. If BIOS finishes the code execution, BIOS sends commands to BMC to stop the POST WDT before transferring to OS. Otherwise, the POST WDT timeout will cause the BMC reset the system.

2.1.3.3 SMBus Topology

PCH accesses the SMBus for reading SPD data from DIMMs and configure clock synthesizer work mode.

The SMBus configuration is as below.

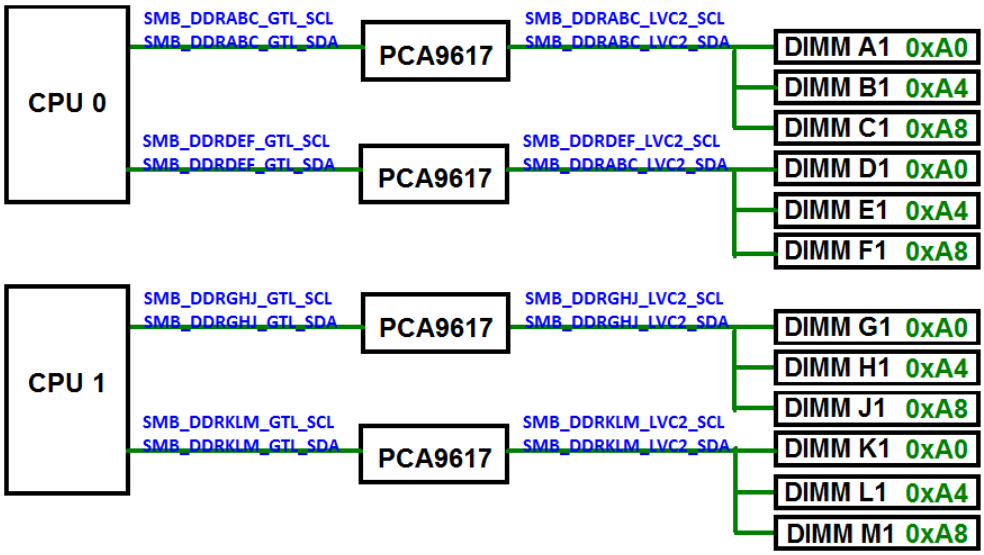


Figure 5 SMBus configuration on server sled

2.1.3.4 Real Time Clock (RTC)

The PCH contains a Motorola MC146818B-compatible real-time clock with 256 bytes of battery-backed RAM. There is two key functions for the RTC. One is to keep track of the time of day and the other is to store the system data.

To count the time of day, the RTC operates with a 32.768KHz crystal. To store the system data, a 3V battery supplies the power for the RAM to keep system data.

The RTC power circuit is implemented as below. The 3.3V standby power is ORed with the battery power through a diode. When system power on, the 3.3V standby power supplies the power for the RTC circuit to preserve the battery energy. When mechanical off or power loss, the battery discharged to supply the RTC circuit.

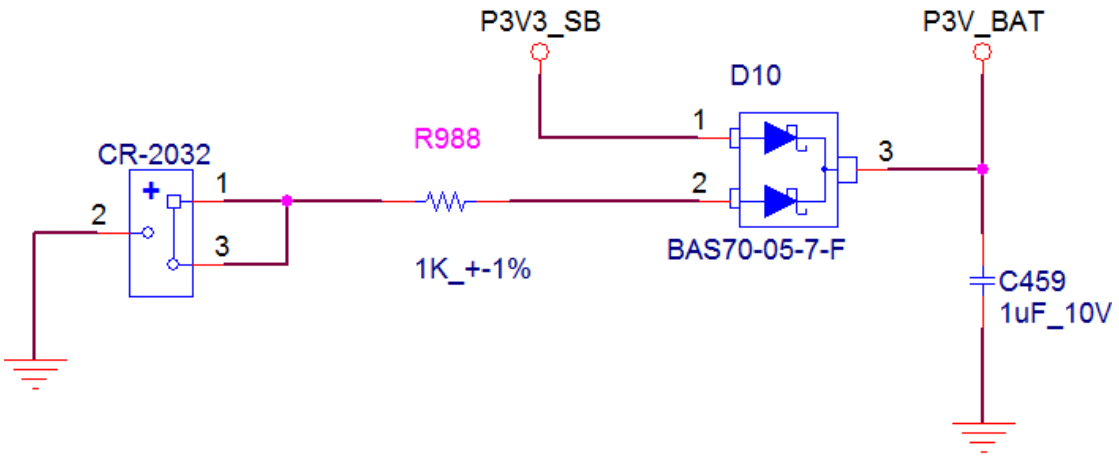


Figure 6 RTC Power Circuit on server sled

The PCH RTC circuit consumes 6uA at most. And the minimum voltage for holding the RTC data is 2V. The nominal capacity of the CR-2032 battery is 225mAh (to 2V).

225mAh / 6uA = 37500 hr ≒ 4.28 yr.

At the worst case, the RTC data can be sustained for 4.28 years.

2.1.4 PCI Express

Both Skylake CPUs and the PCH provide PCI Express resources for I/O applications.

CPU PCI Express resource

The processor PCIe lane assignment is shown as below.

CPU0

PCIe port	PCIe lane width	Lane assignment
Port 1	Lane [7:0]	These 8 lanes can optionally route to the PCH PCIe Upstream port or two U.2 connectors with BOM option. Each U.2 connector can get 4 lanes for application. Default : Route the 8 lanes to two U.2 connectors
	Lane [15:8]	Straight to PCH PCIe Upstream port
Port 2	Lane [15:0]	PCIe x32 slot
Port 3	Lane [7:0]	OCP mezzanine connector 1 (80 pin)
	Lane [15:8]	OCP mezzanine connector 2 (120 pin)
CPU1		
PCIe port	**PCIe lane width**	**Lane assignment**
Port 1	Lane [7:0]	Middle plane connector 1 (non-stuff)
	Lane [15:8]	Middle plane connector 2 (non-stuff)
Port 2	Lane [15:0]	PCIe x32 slot
Port 3	Lane [15:0]	Not used

Table 7 CPU PCIe Resource Assignment

PCH PCI Express resource - root port

The PCH implements a flexible I/O architecture. It allows several high speed interfaces such as the PCIe, USB 3.0, and SATA to be multiplexed with limited high speed I/O(HSIO here after) lanes.

There are 26 HSIO lanes on the PCH. Some of them can be configured to be PCIe Gen3 ports, and others for USB 3.0 or SATA3 depending on the I/O requirement of a platform.

The PCH HSIO lane muxing is summarized as the figure below.

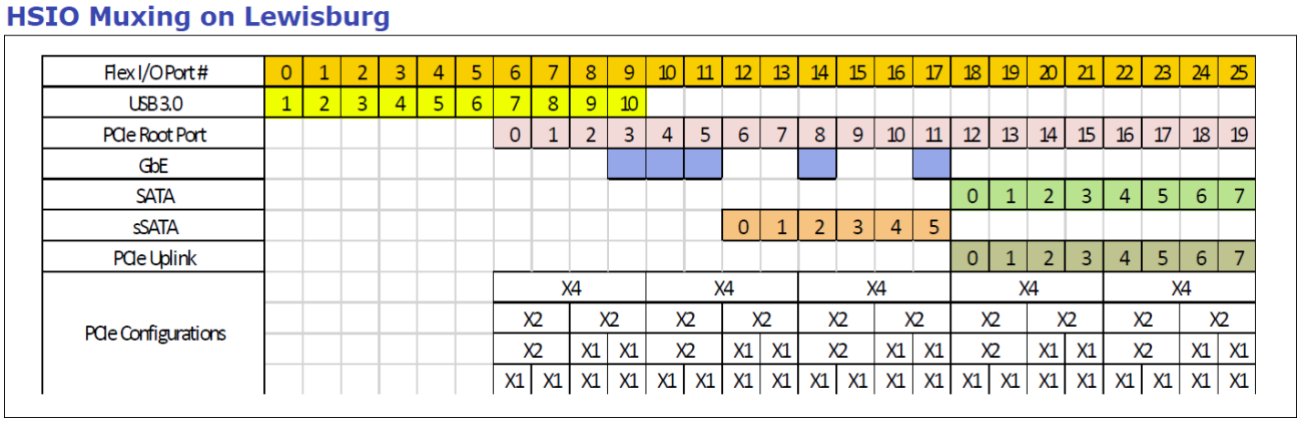


Figure 7 PCH HSIO Lane Muxing

HSIO lane width	Lane assignment
Lane [0]	Configure to be PCIe Gen3 root port for BMC
Lane [1]	Configure to be PCIe Gen3 root port for LAN controller i210-IT
Lane [3:2]	Not used
Lane [5:4]	Configure to be PCIe Gen3 root port for M.2 connector 1
Lane [7:6]	Configure to be PCIe Gen3 root port for M.2 connector 2
Lane [8]	Configure to be SATA3 port for M.2 connector 1
Lane [9]	Configure to be SATA3 port for M.2 connector 2
Lane [10]	Configure to be SATA3 port for U.2 connector 1
Lane [11]	Configure to be SATA3 port for U.2 connector 2
Lane [15:12]	Not used
Lane [19:16]	Not used

Table 8 PCH HSIO Lane Assignment

PCH PCI Express resource - uplink port

The PCH has two uplink ports to handle the PCIe traffic towards the CPU. One is x16 port, the other is x8 port. The uplink ports configuration depends on the data transfer bandwidth requirement and the special functionality support such as the Intel® QAT or 10GbE port.

The default SKU of server sled is implemented with the PCH SKU C621. It doesn't support the Intel ® QAT. Only the x16 uplink port is supported. The PCIe lane assignment is shown as below.

PCIe port	PCIe lane width	Lane assignment
x16	Lane [7:0]	CPU0 PCIe port 1 for PCIe traffic or U.2 connector (BOM option)
	Lane [15:8]	Straight to CPU0 PCIe port 1

Table 9 PCH PCIe Uplink Port Assignment

2.1.5 Ethernet

Open Rack Server has one GbE LAN port for network access. The GbE port is implemented via an Intel i210-IT single-port GbE controller.

The i210-IT is located on the PCIe x1 port of the Intel PCH. It offers a fully-integrated GbE Media Access Control (MAC), Physical Layer (PHY) port and a SGMII/NC-SI port. The highest supported transmission rate is 1000Base-T.

There is a SPI serial EEPROM device attached with the i210-IT. The EEPROM capacity is 4Mb. After power up, the BIOS assigns I/O resources to initialize the i210-IT. Then the i210-IT loads the EEPROM content for LAN port configuration.

For system management purpose, the i210-IT provides both SMBus and NC-SI interface. SMBus and NC-SI are sideband interfaces for pass-through and configuration traffic between the BMC and the GbE. Server sledi mplements both interfaces for the BMC.

For pass-through traffic, the usable bandwidth can be up to 250Mb/s. For sideband interfaces, the usable bandwidth for either direction is up to 1Mb/s when using SMBus and 100Mb/s for the NC-SI interface.

The BMC works in shared-NIC mode with separated MAC.

The concept diagram is illustrated as below.

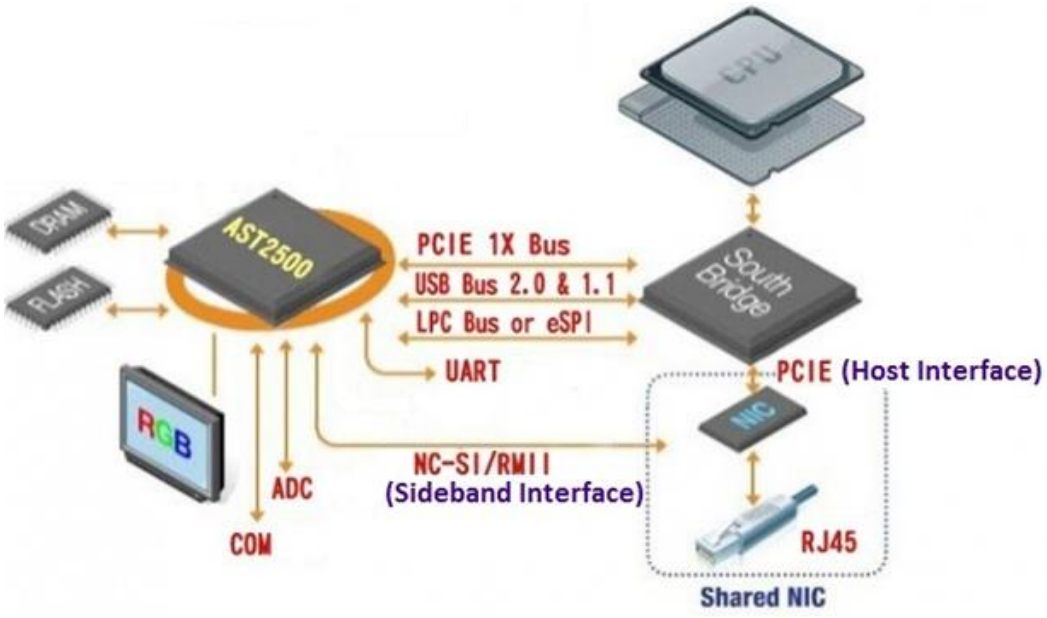


Figure 8 Shared-NIC Concept Diagram

Shared-NIC mode means a common Ethernet connection is shared between the operating system and the BMC. The system management traffic multiplexes with the regular Ethernet traffic. Shared-NIC mode has fewer cabling requirement between the server and the network. This simplifies the hardware design.

Separated MAC means the BMC has a dedicated MAC address. All kinds of traffic (HTTP, NFS, IPMI) on Ethernet can be filtered by MAC address. The management traffic is extracted and sent to the BMC via NC-SI interface.

2.1.6 BMC

Open Rack Server uses the ASPEED® AST2500 BMC for system management such as board healthy monitoring, sensor monitoring, fan speed control, remote management.

The AST2500 is an ARM1176JZF-S core based MCU. It has a local SPI interface to support two SPI flashes. Where duplicated firmware image are stored for redundancy. Both I^2C and NC-SI ports for Out-Of-Band access. The NC-SI port is from the Intel i210 GbE controller. The firmware remote upgrade is supported as well.

The BMC application block diagram is shown as the figure below.

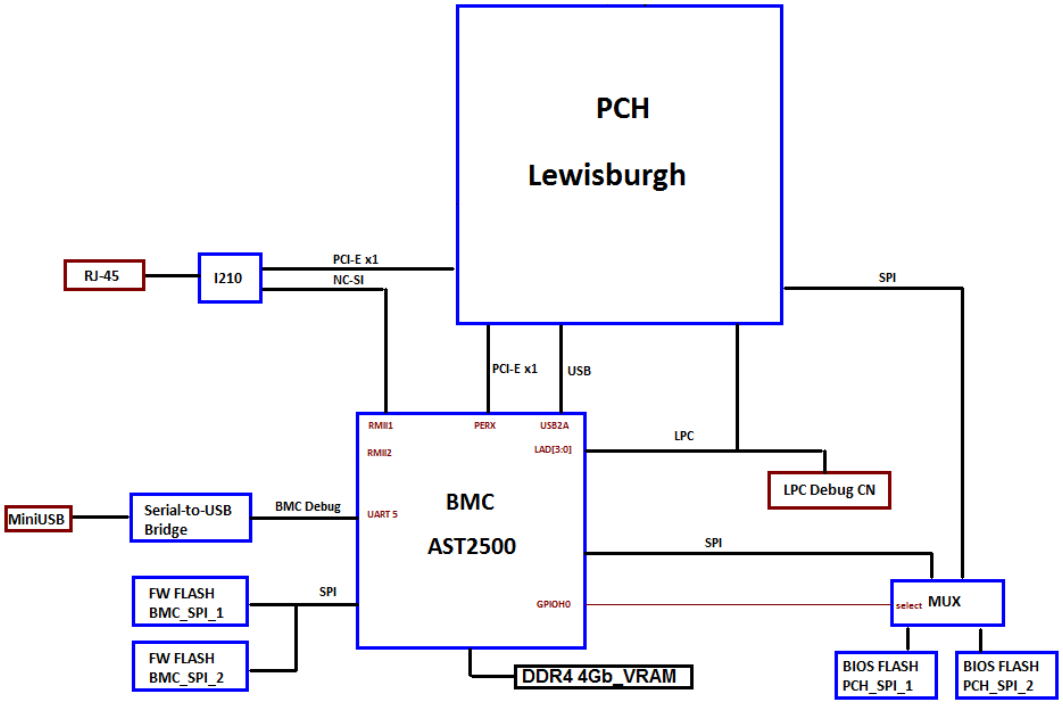


Figure 9 Server Sled BMC Application Block Diagram

AST2500 main features :

- Bus support : PCIe v2.0, USB 2.0, LPC, SPI, eSPI, PECI, I^2C, SMBus, UART
- Various analog/digital sensors for voltage and temperature monitoring
- Fan tachometer and speed control
- GPIO
- Programmable timer for alarm with interrupt generation
- Hash & Crypto Engine
- Support storage redirection
- Support KVM redirection
- JTAG
- Max power consumption (include DRAM) < 2.1W

The BMC has several I^2C compatible buses for different operation purposes.

The figure below shows the block diagram.

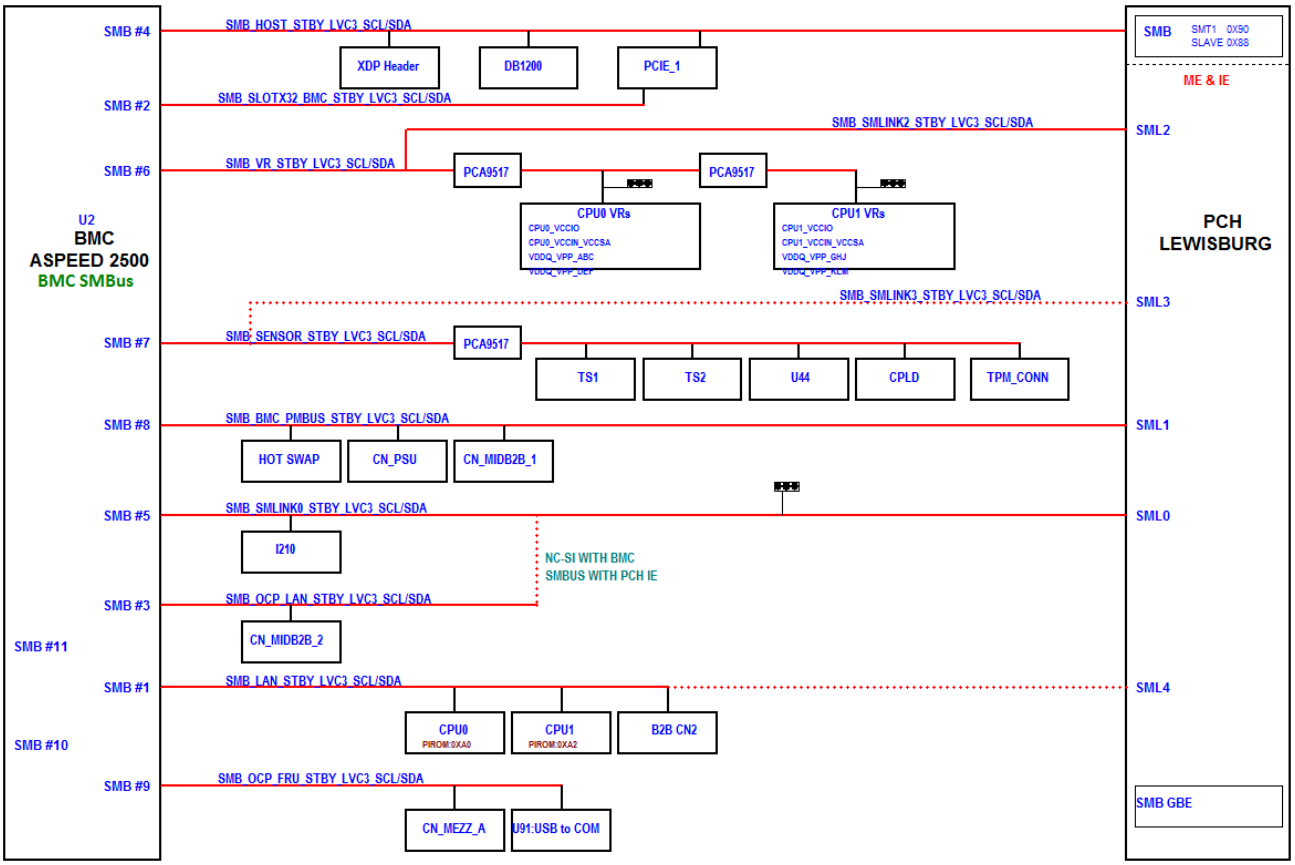


Figure 10 BMC SMBus Block Diagram

The AST2500 implements hardware monitoring to check server board status and raise an alarm if malfunction. Hardware monitor shall support the following features:

- Voltage monitoring for 14 voltage planes
- Temperature sensors: the monitoring of CPUs and chipset utilizing built in on-die sensors + additional temperature sensor located on the coolest area of PCB
- Emergency power shut down in case THERMTRIP# asserted
- PROCHOT# for CPU thermal throttling
- MEMHOT# for memory thermal throttling

Voltage monitoring

Voltage Plane	Signal Name
Vcore for CPU0	PVCCIN_CPU0
Vcore for CPU1	PVCCIN_CPU1

1.2V for DIMM ABC	PVDDQ_ABC
1.2V for DIMM DEF	PVDDQ_DEF
1.2V for DIMM GHJ	PVDDQ_GHJ
1.2V for DIMM KLM	PVDDQ_KLM
5V main power	P5V
3.3V main power	P3V3
1.0V standby for PCH	PVNN_PCH_SB
1.05V standby for PCH	P1V05_PCH_SB
1.8V standby for PCH	P1V8_PCH_SB
1.0V for CPU0 I/O	PVCCIO_CPU0
1.0V for CPU1 I/O	PVCCIO_CPU 1
Battery	VBAT_HM

Table 10 BMC Voltage Monitoring List

Thermal Sensor

Device Name	Description	Address
TS1	This sensor is located at the end far away from the fan tray. It's nearby the front panel and the mini USB connector.	0x90
TS2	This sensor is located at the fan inlet side. It' nearby the midplane connector 2 to monitor the air temperature in front of the fan tray.	0x94

Table 11 BMC Thermal Sensor List

***NOTE** : The LM73 includes three fixed I2C I/O address inside the device. The LM73 provides one ADDR pin with status float, ground, and Vdd to configure the I/O address. For the LM73 type "-0", the status float, ground, and Vdd means the I/O address is configured with 0x90h, 0x92h, and 0x94h. For the LM73 type "-1", the status float, ground, and Vdd means the I/O address is configured with 0x98h, 0x9Ah, and 0x9Ch. Server sled has two LM73-0 devices to configure a unique I2C I/O address.*

GPIO assignment for fan control

Signal name	Description	GPIO

BMC_FAN01_PWM	Fan 0 and fan 1 duty cycle control	GPION0
BMC_FAN23_PWM	Fan 2 and fan 3 duty cycle control	GPION1
BMC_FAN_TACH0	Fan speed detection for fan 0	GPIOO0
BMC_FAN_TACH1	Fan speed detection for fan 1	GPIOO1
BMC_FAN_TACH2	Fan speed detection for fan 2	GPIOO2
BMC_FAN_TACH3	Fan speed detection for fan 3	GPIOO3
FAN_WDT_OUT	This is the BMC fail-safe function. In case BMC is dead, this pin asserts to make system fan run at largest speed.	GPIOP6_TACH14

Table 12 BMC GPIO Assignment for Fan Speed Control

GPIO assignment for miscellaneous functions

Signal name	Description	GPIO
FM_PWR_BTN_N	Power button from the front panel	GPIOE2_NDSR3
FM_BMC_PWRBTN_OUT_N	Power button output to PCH	GPIOZ2_VPOG4_NORA2_SIOPBO
RST_SYSTEM_BTN_N	Reset button from the front panel	GPIOE0_NCTS3
RST_BMC_SYSRST_BTN_OUT_N	Reset button output to PCH	GPIOE1_NDCD3
FM_SLPS4_N	SLPS4_N status check	GPIOY0
FM_CPU0_SKTOCC_LVT3_N	CPU0 present	GPIOS2
FM_CPU1_SKTOCC_LVT3_N	CPU1 present	GPIOS3
CPLD_PWRGD_SYS_PWROK	System power good healthy detection	GPIOZ1
CPLD_ALL_PGD	Power good status of all the power planes	GPIOC4
FM_BIOS_POST_CMPLT_N	BIOS POST complete	GPIOG3
FM_BACKUP_BIOS_SEL_N	BIOS flash selection. H: primary BIOS flash L: secondary BIOS flash	GPIOQ4

FM_PCH_SPI_BMC_CTRL_N	BIOS SPI path control H: SPI BIOS ROM from PCH (Default) L: SPI BIOS ROM from BMC	GPIOH0
FM_I210_SPI_BMC_CTRL_N	LAN SPI flash control H: SPI LAN image ROM connected to i210. (Default) L: SPI LAN image upgrade from BMC.	GPIOH4
IRQ_BMC_PCH_SMI_LPC_N	LPC SMI	GPIOB5_LPCPD_LPCSMI#
IRQ_SML1_PMBUS_ALERT_N	Alert from the HotSwap controller	GPIOAA1_VPOR3_NORD1_SALT8
FM_PMBUS_ALERT_BUF_EN_N	PMBUS alert output	GPIOAA2_VPOR4_NORD2_SALT9
FM_FAST_PROCHOT_EN_N	PROCHOT# event output	GPIOR3_SPI2CK
FM_CPU0_PROCHOT_LVT3_N	CPU0 processor hot event input	GPIOP0
FM_MEM_THERM_EVENT_LVT3_N	Memory thermal event from the DIMMs	GPIOB3
FM_PCH_BMC_THERMTRIP_N	THERMTRIP# event input	GPIOG2_SGPS1I0
SMB_SLOTX32_SBLVC3_ALERT_N	Riser card alert input	GPIOAA0_VPOR2_NORD0_SALT7
IRQ_MEZZ_LAN_ALERT_N	Mezzanine card alert input	GPIOE5_NRTS3
IRQ_FORCE_NM_THROTTLE_N	THROTTLE# event output	GPIOR2_SPI2CS0
FM_CPU_ERR2_LVT3_N	ERROR_N[2] input. IIO or DMI3 found fatal errors.	GPIOG0_SGPS1CK
FM_CPU_MSMI_CATERR_LVT3_N	CATERR_N and MSMI_N combined input. CPU or DIMM found errors.	GPIOG1_SGPS1LD
FM_BOARD_SKU_ID0	Board SKU ID 0	GPIOG4
FM_BOARD_SKU_ID1	Board SKU ID 1	GPIOG5
FM_BOARD_SKU_ID2	Board SKU ID 2	GPIOG6

FM_BOARD_SKU_ID 3	Board SKU ID 3	GPIOG7
FM_BOARD_SKU_ID 4	Board SKU ID 4	GPIOGJ3
PCB_REV0	PCB REV0	GPIOJ0
PCB_REV1	PCB REV1	GPIOJ1
PCB_REV2	PCB REV2	GPIOJ2
PWR_LINK_B_LED	Power LED, color: Blue	GPIOO5
ALERT_R_LED	Alert LED, color: Red	GPIOO6
BMC_RDY_N	BMC ready. Output to PCH	GPIOP7
PU_BMC_GPIOS7	VGA class code	GPIOS7_VPOB9
PD_BMC_GPIOZ7	Strap for fast reset mode	GPIOZ7_VPOG9_NORA7

Table 13 BMC GPIO Assignment for Miscellaneous Functions

Ethernet traffic filtering

The BMC works in shared-NIC mode with separated MAC. The BMC has a dedicated MAC address. All kinds of traffic (HTTP, NFS, IPMI) on Ethernet can be filtered by MAC address. The normal traffic passes through PCIe when payload operating. The management traffic is extracted and sent to the BMC via NC-SI interface no matter payload on and off.

Dual boot flash for redundancy

The BMC has two SPI flashes for redundancy. There is a watchdog timer for boot time counting. When the main flash boots up successfully, the watchdog timer should be disabled before timeout. If the firmware running on the main flash doesn't disable the watchdog timer within 22 seconds, the watchdog bites and reset the BMC. After reset, the BMC boots with the 2nd flash.

2.1.7 CPLD

Server sled implements an Altera MAX 10 series CPLD device. The model number is 10M04DAU324. It's a 324-pin, 15mm x 15mm, UBGA packaged chip. There are 4K logic elements can be configured to fulfill applications. The advantage of CPLD is to make the hardware design flexible, reduce components to simplify design, and save cost.

The major functions of the CPLD are :

- Power on/off sequencing control
- Reset control
- Logic control such as logic calculation, signal multiplex, delay, buffered, level shift
- Misc signals re-route for flexibility
  - GPIO
  - I/O port expansion
  - Sensor event
  - Device present, package ID
  - Interrupt routing
  - Error reporting

2.1.8 TPM

Server sled implements an Infineon SLB9670VQ2.0 FW7.61 TPM to support secure key and file storage.

The SLB9670VQ2.0 FW7.61 is a TCG rev1.2/2.0 compliant TPM device. TCG means Trusted Computing Group. It's an industry group specialized in trusted computing.

The SLB9670VQ2.0 FW7.61 TPM connects with PCH via SPI interface. The features are summarized below.

• Compliant to TPM main specification, Family "2.0", Level 00, Revision 01.16

• Support SPI interface

• Meeting Intel TXT, Microsoft Windows and Google Chromebook certification criteria for successful platform qualification

• True Random Number Generator

• Full personalization with Endorsement Key (EK) and EK certificate

• 7206 Byte free NV memory

• Low standby power consumption. Typical 110uA

• Hardware hash accelerator for SHA-1 and SHA-256 algorithm

• Built-in support by Linux Kernel

2.2 Power Feed

Server motherboard is supplied by the 12V source from the datacenter rack. The nominal input voltage is 12.5V DC at light loading with a range of 11V ~ 13V.

At any loading, the server shall accept and operate normally with input voltage tolerance range between 10.8V ~ 13.2V.

There is a hot swap controller(HSC) in series with the power feed path. Besides the hot swap, the HSC provides input voltage monitoring and protection. The figure below shows the block diagram of the circuitry.

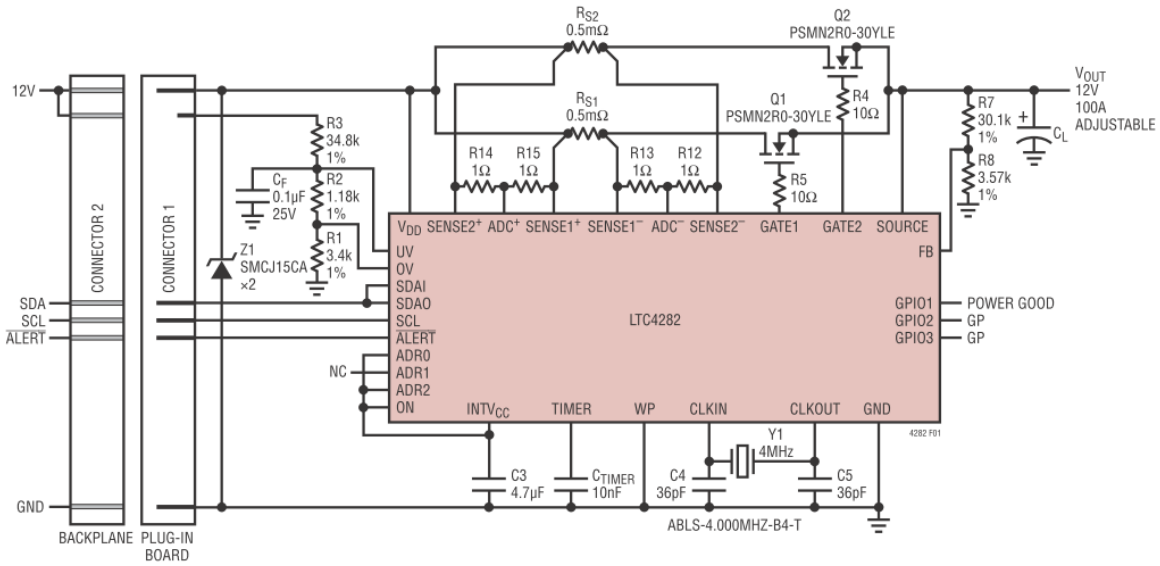


Figure 11 Hot Swap Controller Block Diagram

This HSC circuit implements several functions including:

1. In-rush current control when power on
2. Monitors Current, Voltage, Power and Energy
3. Current limiting protection for over current and short circuit conditions. Trip point: 48.8A ~ 41.4A
4. Under-voltage protection level: 10V ~ 10.1V with hysteresis
5. Send alerts when alarm thresholds exceeded
6. Fault protection for OVP and UVP events, the default scheme is auto-retry for every 50ms
7. Fault protection for OCP events, the default scheme is latch-off
8. All the trip point settings and the fault protection schemes can be reconfigured by the HSC I^2C interface

2.2.1 System Power Distribution

System Power Flow

The figure below is a System Power Flow Map. It shows where the voltage rails are regulated from. The power budget statistics and the power solutions are listed in the map as well. The power rails are derived from 12V. The total maximum power for devices in server will be up to 800W theoretically.

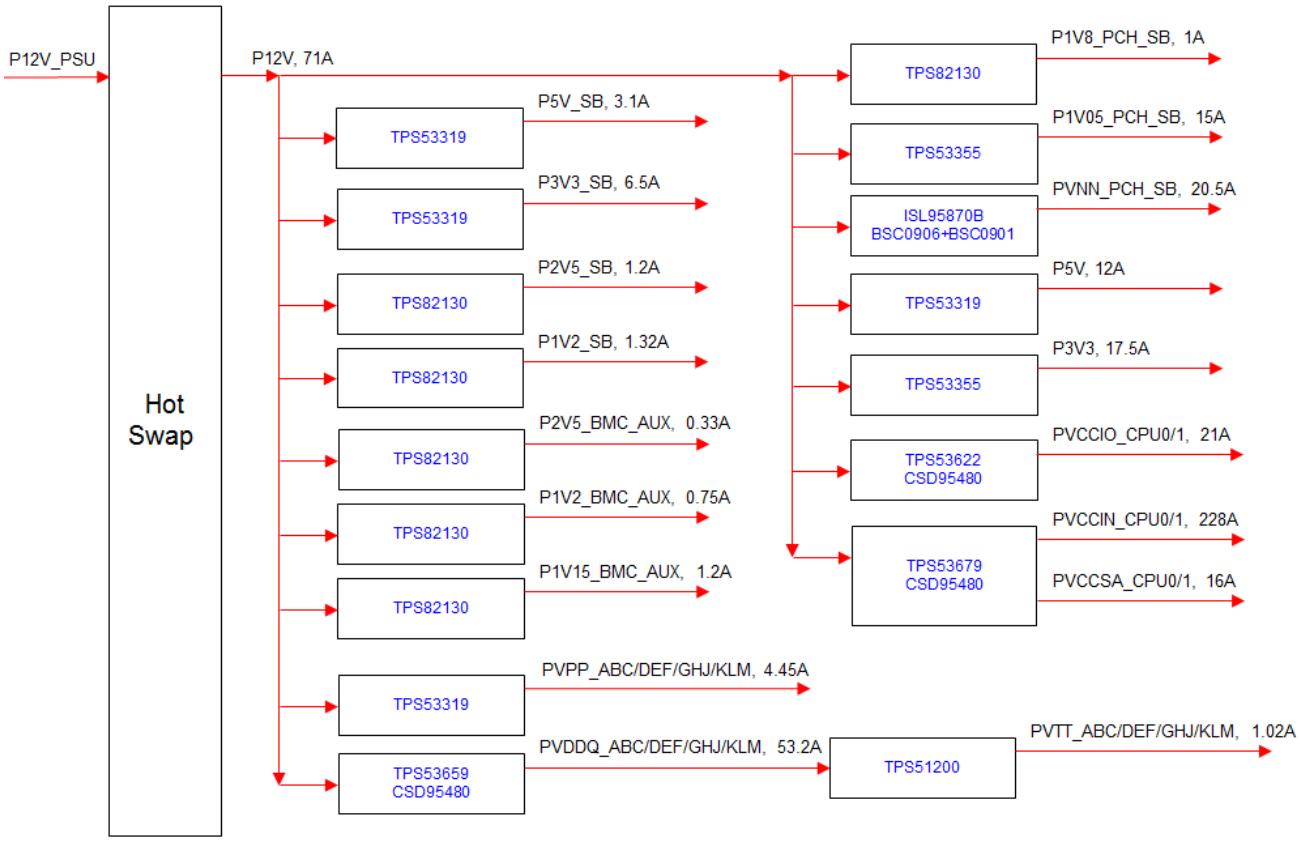


Figure 12 System Power Flow Map

System Power Budget Analysis

The table below shows the power budget analysis of all the power rails.

Page	Output Net Name	Solution	Input Net Name	Vout	Iout	Vin	η	Iin	Remark
CPU VRD									
	PVCCIN_CPU0	TPS53679_1a	P12V	1.80 V	228.00 A	12 V	94%	36.6 A	
	PVCCIN_CPU1	TPS53679_2a	P12V	1.80 V	228.00 A	12 V	94%	36.6 A	
	PVCCSA_CPU0	TPS53679_1b	P12V	0.85 V	16.00 A	12 V	92%	1.2 A	
	PVCCSA_CPU1	TPS53679_2b	P12V	0.85 V	16.00 A	12 V	92%	1.2 A	
	PVCCIO_CPU0	TPS53622	P12V	1.00 V	21.00 A	12 V	90%	1.9 A	
	PVCCIO_CPU1	TPS53622	P12V	1.00 V	21.00 A	12 V	90%	1.9 A	
Switching									
	P_VDDQ_ABC	TPS53659	P12V	1.215 V	45.22 A	12 V	91%	5.1 A	
	P_VDDQ_DEF	TPS53659	P12V	1.215 V	53.22 A	12 V	92%	5.9 A	
	P_VDDQ_GHJ	TPS53659	P12V	1.215 V	53.22 A	12 V	92%	5.9 A	
	P_VDDQ_KLM	TPS53659	P12V	1.215 V	53.22 A	12 V	92%	5.9 A	
	P_VPP_ABC	TPS53319	P12V	2.58 V	4.45 A	12 V	94%	1.0 A	
	P_VPP_DEF	TPS53319	P12V	2.58 V	4.45 A	12 V	94%	1.0 A	
	P_VPP_GHJ	TPS53319	P12V	2.58 V	4.45 A	12 V	94%	1.0 A	
	P_VPP_KLM	TPS53319	P12V	2.58 V	4.45 A	12 V	94%	1.0 A	
	PVNN_PCH_SB	ISL95870BIRZ	P12V	1.00 V	20.50 A	12 V	88%	1.9 A	
	P1V05_PCH_SB	TPS53355A	P12V	1.05 V	15.00 A	12 V	86%	1.5 A	
	P5V_SB	TPS53319	P12V	5.00 V	3.10 A	12 V	95%	1.4 A	
	P3V3_SB	TPS53319	P12V	3.30 V	6.44 A	12 V	94%	1.9 A	
	P2V5_SB	TPS82130	P12V	2.50 V	1.19 A	12 V	87%	0.3 A	
	P1V8_PCH_SB	TPS82130	P12V	1.80 V	1.00 A	12 V	83%	0.2 A	
	P1V2_SB	TPS82130	P12V	1.20 V	1.32 A	12 V	79%	0.2 A	
	P1V2_BMC_AUX	TPS82130	P12V	1.20 V	0.75 A	12 V	79%	0.1 A	
	P1V15_BMC_AUX	TPS82130	P12V	1.15 V	1.20 A	12 V	79%	0.1 A	
	P5V	TPS53319	P12V	5.00 V	12.00 A	12 V	94%	5.3 A	
	P3V3	TPS53355	P12V	3.30 V	17.50 A	12 V	94%	5.1 A	
Hot-Swap									
	P12V	LTC4282	P12V_PSU	12.000 V	70.933 A	12 V			
LDO								**PD**	**Remark**
	PVTT_ABC	TPS51200DRCT	P_VDDQ_ABC	0.608 V	1.02 A	1.2 V	50%	0.62 W	
	PVTT_DEF	TPS51200DRCT	P_VDDQ_DEF	0.608 V	1.02 A	1.2 V	50%	0.62 W	
	PVTT_GHJ	TPS51200DRCT	P_VDDQ_GHJ	0.608 V	1.02 A	1.2 V	50%	0.62 W	
	PVTT_KLM	TPS51200DRCT	P_VDDQ_KLM	0.608 V	1.02 A	1.2 V	50%	0.62 W	
	P2V5_BMC_AUX	TPS74801	P3V3_SB	2.50 V	0.33 A	3.3 V	76%	0.26 W	

Table 14 System Power Budget

2.2.2 Power Sequence

It is guaranteed that the IC devices will not be damaged during supply voltage ramp up by strictly following IC vendors' power sequence recommendations.

The 12V input is gated by the BMC. In the start-up of server, the BMC is initialized. After that, the BMC enables the DC/DC converter.

The server has a CPLD to control the power sequence. The timing is implemented according to Intel's Cresecnt CRB Revision 1.04 for CPU and chipset portion. Resistor options are also reserved as Intel's recommendation in the reference schematics. In case the CPLD design error causes boot up failure, the resistor options can bypass the CPLD control to make the board boot up successfully. But this only happens in develop phase, suppose the CPLD should design well once production.

Power-Up Sequence

After the 12V in, the power-up sequence is described in the following figure.

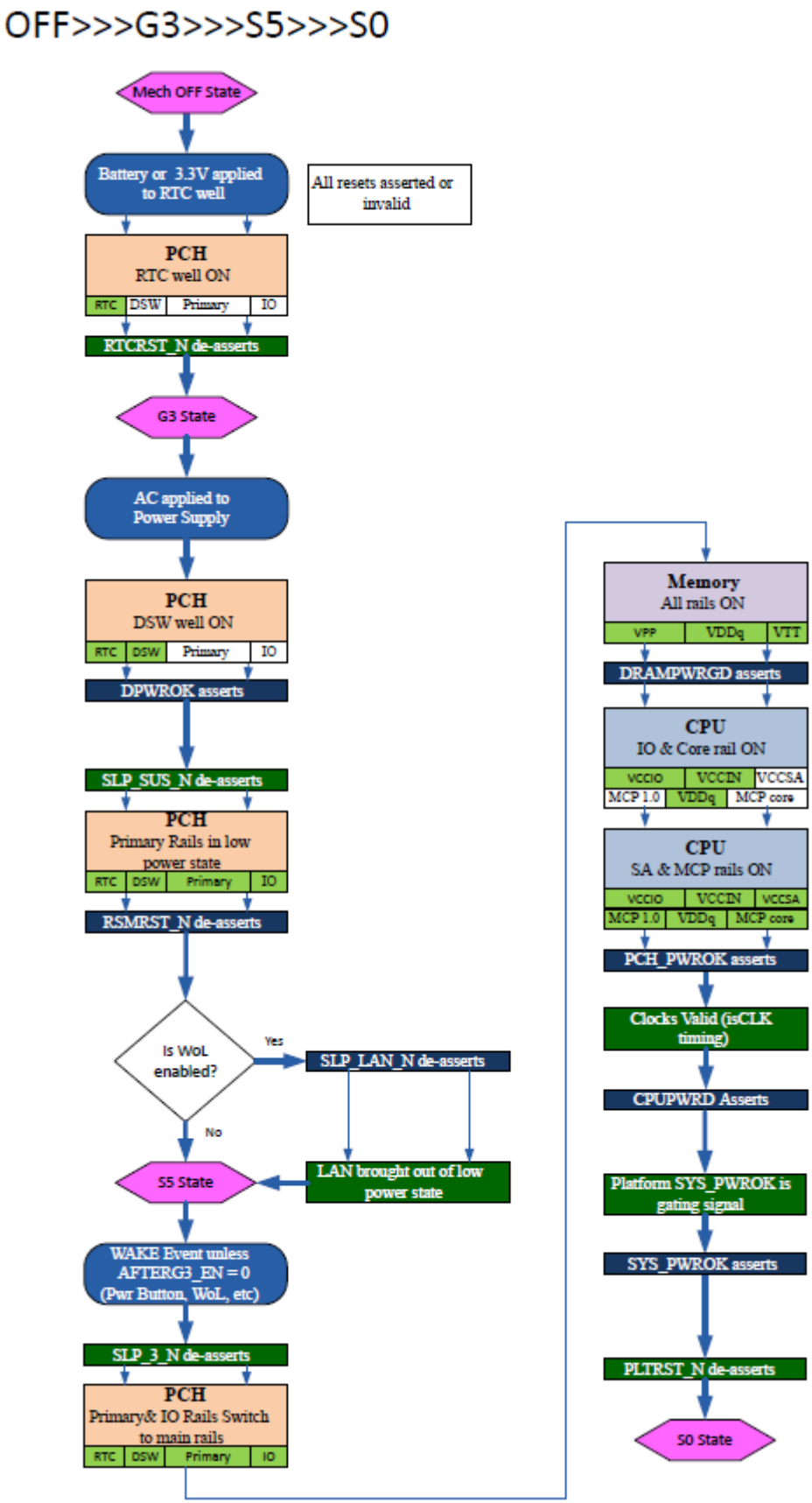


Figure 13 Server Power-Up Sequence

Power-Down Sequence

The power-Down sequence is described in the following figure.

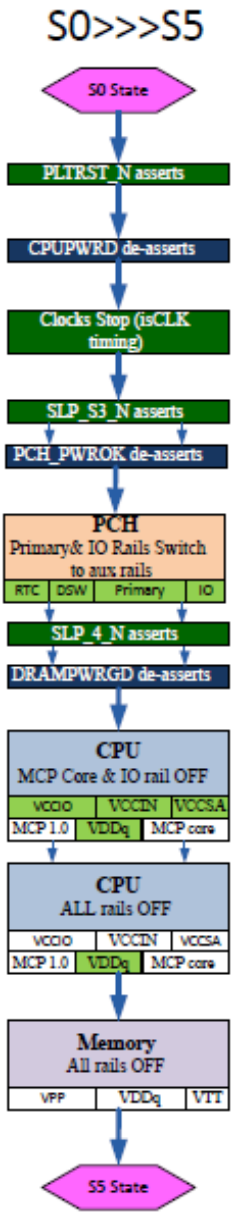


Figure 14 Server Power-down Sequence

2.2.3 Voltage Monitoring and Alarm

All the internal DC voltages of server are monitored by the BMC.

From the ramp-up of 12V, 5V and 3.3V to the all sequential system powers, the power sequence mechanism actually receives all their power-good signals and ANDs them together aggregately into the final power-good signal of SYS_PWRGD. Any voltage is out of specification will cause a hardware based reset.

All the local system power's outputs are monitored by their own voltage regulators respectively. However, some of them do not have native over voltage protection ability. For these cases, the server has external comparators for the over voltage protection.

When under voltage, over voltage or over current event happens, their controllers drive their own power good signals low. Power good signals are effectively aggregated to the last one, SYS_PWRGD. The SYS_PWRGD low will make the PCH to reset the whole system (excluding BMC).

Below table lists the set points of under voltage, over voltage and over current protections.

Power Well	VR P/N	Power Name	Voltage Vtyp (V)	UVP (V)	OVP (V)	Current Imax(A)	OCP (A)
STBY	ISL95870BIRZ	PVNN_PCH_SB	1.00 V	VOUT*84%	VOUT*116%	20.50 A	26.7 A
	TPS53355A	P1V05_PCH_SB	1.05 V	0.74 V	1.26 V	15.00 A	19.5 A
	TPS53319	P5V_SB	5.00 V	3.50 V	6.00 V	3.10 A	4.0 A
	TPS53319	P3V3_SB	3.30 V	2.31 V	3.96 V	6.44 A	8.4 A
	TPS82130	P2V5_SB	2.50 V	N/A	N/A	1.19 A	3.6 A
	TPS82130	P1V8_PCH_SB	1.80 V	N/A	N/A	1.00 A	3.6 A
	TPS82130	P1V2_SB	1.20 V	N/A	N/A	1.32 A	3.6 A
	TPS82130	P1V2_BMC_AUX	1.20 V	N/A	N/A	0.75 A	3.6 A
	TPS82130	P1V15_BMC_AUX	1.15 V	N/A	N/A	1.20 A	3.6 A
Main	TPS53679	PVCCIN_CPU0	1.80 V	VID-VDROOP-0.37	VID+0.33	228.00 A	273.6 A
	TPS53679	PVCCIN_CPU1	1.80 V	VID-VDROOP-0.37	VID+0.33	228.00 A	273.6 A
	TPS53679	PVCCSA_CPU0	0.85 V	VID-VDROOP-0.37	VID+0.33	16.00 A	20.8 A
	TPS53679	PVCCSA_CPU1	0.85 V	VID-VDROOP-0.37	VID+0.33	16.00 A	20.8 A
	TPS53622	PVCCIO_CPU0	1.00 V	VID-VDROOP-0.37	VID+0.33	21.00 A	27.3 A
	TPS53622	PVCCIO_CPU1	1.00 V	VID-VDROOP-0.37	VID+0.33	21.00 A	27.3 A
	TPS53659	P_VDDQ_ABC	1.215 V	VID-VDROOP-0.37	VID+0.33	45.22 A	58.8 A
	TPS53659	P_VDDQ_DEF	1.215 V	VID-VDROOP-0.37	VID+0.33	53.22 A	69.2 A

	TPS53659	P_VDDQ_GHJ	1.215 V	VID-VDROOP-0.37	VID+0.33	53.22 A	69.2 A
	TPS53659	P_VDDQ_KLM	1.215 V	VID-VDROOP-0.37	VID+0.33	53.22 A	69.2 A
	TPS53319	P_VPP_ABC	2.58 V	1.80 V	3.09 V	4.45 A	5.8 A
	TPS53319	P_VPP_DEF	2.58 V	1.80 V	3.09 V	4.45 A	5.8 A
	TPS53319	P_VPP_GHJ	2.58 V	1.80 V	3.09 V	4.45 A	5.8 A
	TPS53319	P_VPP_KLM	2.58 V	1.80 V	3.09 V	4.45 A	5.8 A
	TPS53319	P5V	5.00 V	3.50 V	6.00 V	12.00 A	15.6 A
	TPS53355	P3V3	3.30 V	2.31 V	3.96 V	17.50 A	22.8 A
	TPS51200DRCT	PVTT_ABC	0.608 V	N/A	N/A	1.02 A	3.00 A
	TPS51200DRCT	PVTT_DEF	0.608 V	N/A	N/A	1.02 A	3.00 A
	TPS51200DRCT	PVTT_GHJ	0.608 V	N/A	N/A	1.02 A	3.00 A
	TPS51200DRCT	PVTT_KLM	0.608 V	N/A	N/A	1.02 A	3.00 A
	TPS74801	P2V5_BMC_AUX	2.50 V	N/A	N/A	0.33 A	2.00 A

Table 15 Over Voltage and Under Voltage Threshold of Local DC/DC SMPS

2.3 Printed Circuit Board

The PCB material for server motherboard is the TU-862HF.

TU-862HF is a Hi-Tg halogen free material and made of epoxy resin and E-glass fabric. TU-862HF achieves flammability class of UL94V-0 by incorporating nitrogen compounds in the materials. This series of green materials are designed to eliminate the use of halogenated resins due to the potential hazardous effects from the environmental concerns. These products are suitable for boards that need to survive severe thermal cycles, or to experience excessive assembly work. TU-862 HF laminates also exhibit superior chemical resistance, thermal stability for lead free soldering assembly.

Industry approvals for the TU-862HF :

- IPC-4101 Type Designation : /127, /128, /130
- UL Designation - ANSI Grade : FR-4.1
- UL File Number: E189572
- Flammability Rating: 94V-0
- Maximum Operating Temperature: 130°C

In order to be compatible with the ORv2 sled mechanical design, the nominal PCB thickness should be 89.44mil with tolerance ±8 mil.

The following figure shows the PCB stack-up and the impedance control of the server motherboard.

stack-up

Project name: NDC S18-OCP Stackup

material: [illegible]

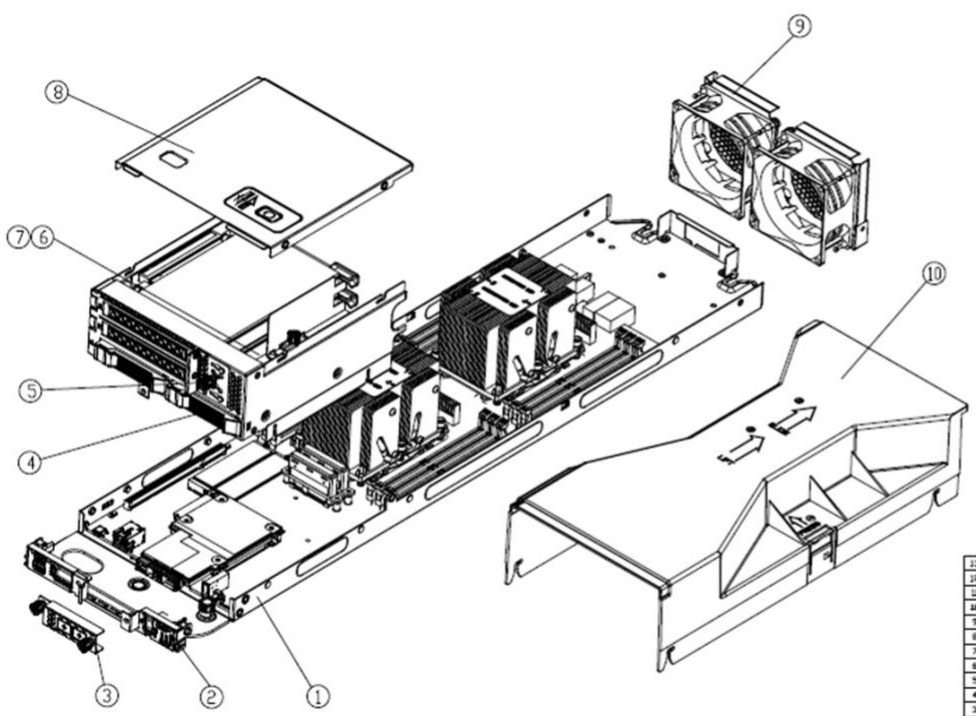
Layer Name	suggestion Build - up		Dielectric thickness (unit mil)	DK @4GHz	DF @4GHz	DK @8GHz	DF @8GHz	Estimateted residual copper	DDR 2 SPC DQ/DQS/Misc I/O; SE clocks 40 +/-4.0 ohm TraceWidth(mil)		DDR 1 SPC DQ/DQS/Misc I/O; SE clocks 50 +/-5.0 ohm TraceWidth(mil)		1) PCIe, UPI, DMI 2)Diff CLKs 3)USB2/3, SATA3 85 +/- 8.5 ohm TraceWidth(mil)		SFI 10G-KR (option 1) 93 +/- 9.3 ohm TraceWidth(mil)		SFI 10G-KR (option 2) 100 +/- 10.0 ohm TraceWidth(mil)	
	sm		0.7	4	0.022	4	0.022											
L1TOP	0.5oz+plating		1.6						8.1	40+/-4	5.1	50+/-5	5/4.5	85+/-8.5	4.5/6	93+/-9.3	4/7	100+/-10
	PP	1080-67%	3.20	3.4	0.0171	3.4	0.0181		40.8	146.21	51.4	145.58	84.8 4G:0.581dB/inch 8G:1.103dB/inch	142.99	93.9 4G:0.581dB/inch 8G:1.103dB/inch	142.77	101.1 4G:0.581dB/inch 8G:1.103dB/inch	142.77
L2GND	1oz		1.2															
	Core	1080*1	3.00	3.4	0.0166	3.4	0.0176											
L3SIG	1oz		1.2						6.5	40+/-4	4	50+/-5	4.5/5.5	85+/-8.5	4.1/8	93+/-9.3	3.5/9	100+/-10
	PP	2116-55%*2	9.10	4.2	0.0151	4.1	0.0161		40.1	156.23	50.8	156.23	85.6 4G:0.579dB/inch 8G:1.048dB/inch	156.23	94.1 4G:0.579dB/inch 8G:1.048dB/inch	156.23	101.8 4G:0.579dB/inch 8G:1.048dB/inch	156.23
L4SIG	1oz		1.2						6.5	40+/-4	4	50+/-5	4.5/5.5	85+/-8.5	4.1/8	93+/-9.3	3.5/9	100+/-10
	Core	1080*1	3.00	3.4	0.0166	3.4	0.0176		40.1	156.23	50.8	156.23	85.6 4G:0.579dB/inch 8G:1.048dB/inch	156.23	94.1 4G:0.579dB/inch 8G:1.048dB/inch	156.23	101.8 4G:0.579dB/inch 8G:1.048dB/inch	156.23
L5GND	1oz		1.2															
	PP	1080-67%+2116-55%	6.80	4.3	0.0148	4.28	0.0158											
L6SIG	1oz		1.2						5.6	40+/-4	3.4	50+/-5	4.5/7.5	85+/-8.5	3.9/10	93+/-9.3	3.4/12	100+/-10
	Core	1080*1	3.00	3.4	0.0166	3.4	0.0176		40.2	156.23	51.1	156.23	86.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	94.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	101.5 4G:0.579dB/inch 8G:1.048dB/inch	156.23
L7GND	1oz		1.2															
	PP	106-74%*2	2.70	3.8	0.0177	3.7	0.0187											
L8PWR	2oz		2.4															
	Core		12.00	4.5	0.0135	4.5	0.0145											
L9PWR	2oz		2.4															
	PP	106-74%*2	2.70	3.8	0.0177	3.7	0.0187											
L10GND	1oz		1.2															
	Core	1080*1	3.00	3.4	0.0166	3.4	0.0176											
L11SIG	1oz		1.2						5.6	40+/-4	3.4	50+/-5	4.5/7.5	85+/-8.5	3.9/10	93+/-9.3	3.4/12	100+/-10
	PP	1080-67%+2116-55%	6.80	4.3	0.0148	4.28	0.0158		40.2	156.23	51.1	156.23	86.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	94.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	101.5 4G:0.579dB/inch 8G:1.048dB/inch	156.23
L12GND	1oz		1.2															
	Core	1080*1	3.00	3.4	0.0166	3.4	0.0176											
L13SIG	1oz		1.2						6.5	40+/-4	4	50+/-5	4.5/5.5	85+/-8.5	4.1/8	93+/-9.3	3.5/9	100+/-10
	PP	2116-55%*2	9.10	4.2	0.0151	4.1	0.0161		40.1	156.23	50.8	156.23	86.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	94.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	101.5 4G:0.579dB/inch 8G:1.048dB/inch	156.23
L14SIG	1oz		1.2						6.5	40+/-4	4	50+/-5	4.5/5.5	85+/-8.5	4.1/8	93+/-9.3	3.5/9	100+/-10
	Core	1080*1	3.00	3.4	0.0166	3.4	0.0176		40.1	156.23	50.8	156.23	86.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	94.2 4G:0.579dB/inch 8G:1.048dB/inch	156.23	101.5 4G:0.579dB/inch 8G:1.048dB/inch	156.23
L15GND	1oz		1.2															
	PP	1080-67%	3.20	3.4	0.0171	3.4	0.0181											
L16BTM	1oz		1.6						8.1	40+/-4	5.1	50+/-5	5/4.5	85+/-8.5	4.5/6	93+/-9.3	4/7	100+/-10
	sm		0.7	4	0.022	4	0.022		40.8	146.21	51.4	145.58	84.8 4G:0.581dB/inch 8G:1.103dB/inch	142.99	93.9 4G:0.581dB/inch 8G:1.103dB/inch	142.77	101.1 4G:0.581dB/inch 8G:1.103dB/inch	142.77
		Remark:	97.4	Total board thickness: 2.5mm ±10%				(Including plate copper and solder mask)										
			2.47 mm															

Figure 15 Server PCB Stack-up and Impedance Control

3 Mechanical Overview

This chapter briefly introduces the server sled placement, the front panel design, and the assembly.

In the following drawing shows the mechanical structure of sled.



NO.	CHENBRO P/N	DRAWING NO.	NAME	QTY	UNIT
13					
12					
11				1	
10		RM23930-201	RM23930-201-AIR-DUCT	1	
9		RM23930-18	RM23930-18-EHE-COVER-HDIE	1	
8		RM23930-19	RM23930-19-PCIE-Top-COVER	1	
7		RM23930-15	RM23930-15-HDE-CASE-BOTTOM	1	
6		RM23930-14	RM23930-14-HDE-CASE-TOP	1	
5		RM23930-08	RM23930-08-PCIE-COVER	1	
4		RM23930-13	RM23930-13-PCIE-HDD-BKT	1	
3		MEZZ-COVER-3000		1	
2		RM23930-07	RM23930-07-IO-FRONT	1	
1		RM23930-06	RM23930-06-BOTTOM-HDIE	1	

Figure 16 Server sled structure

3.1 Main Components Placement

Server motherboard PCB size is 6.5” x 20” (165.1mm x 531.15mm). The main components placement is shown as the figure below.

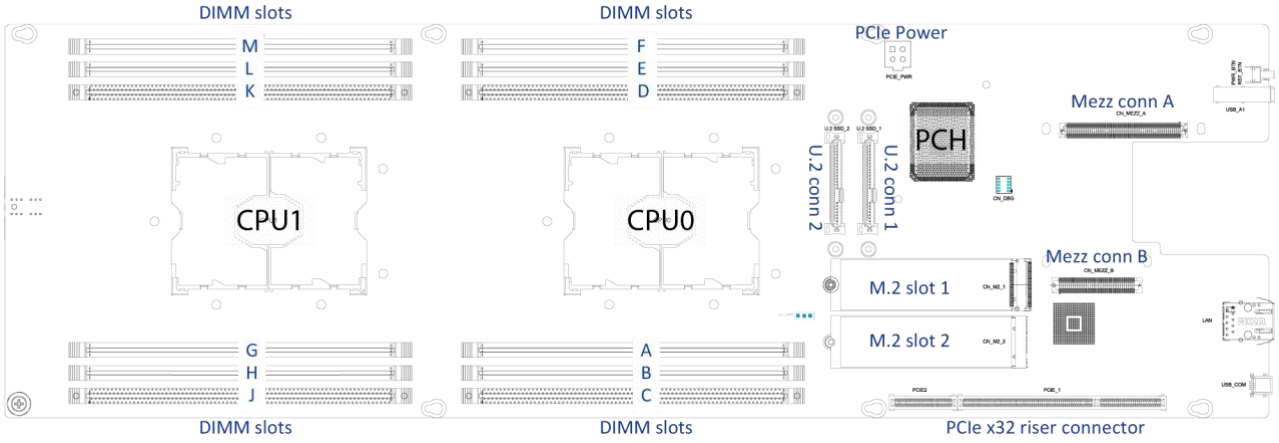


Figure 17 Server Main Components Placement

3.2 Front panel design

The figure below presents the layout of the 2OU sled front panel and the locations of the components in the front panel.

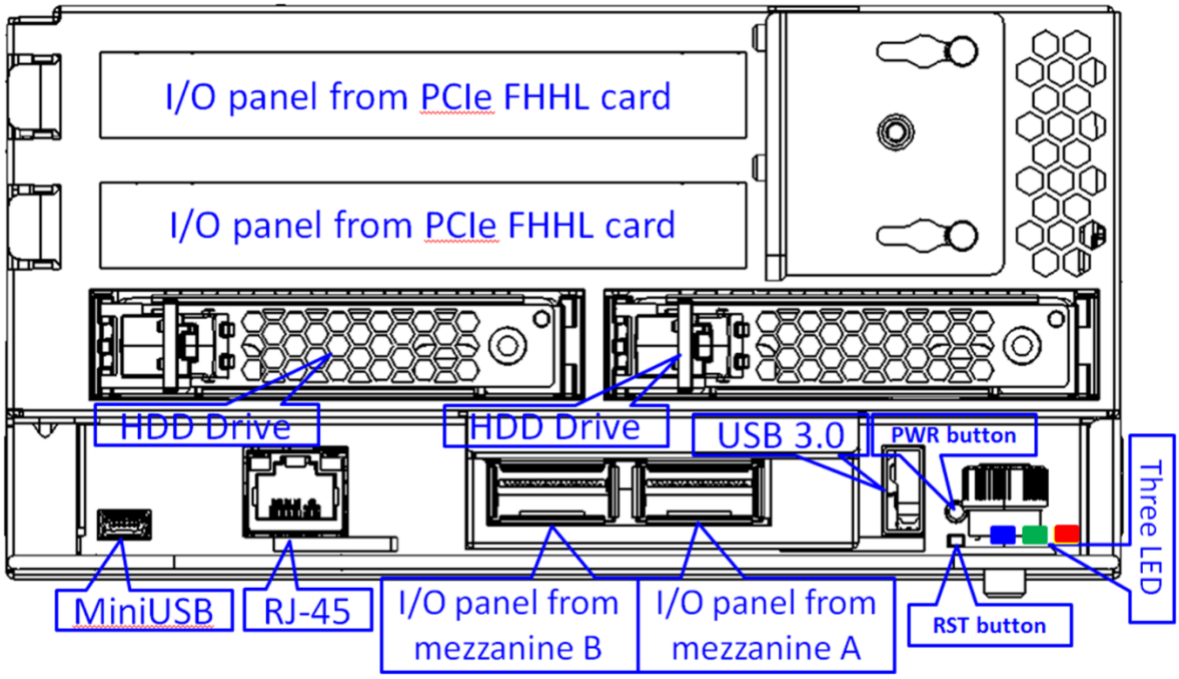


Figure 18 Front Panel Design for 2OU Server Sled

The front panel includes the following items.

- One Mini USB port for BMC debug
- One RJ-45 connector for GbE interface
- One USB 3.0 connector
- Two 2.5” HDD Drive bay from U.2 interface
- Two I/O panels from the PCIe FHHL cards
- Two I/O panels from the mezzanine card A and B
- One power button

- One reset button
- Three LEDs

USB, RJ-45, and USB 3.0

The connector type and pin list are detailed in Section 7.

HDD Drive bay

For the 2OU sled design, there are two I/O panels for the HDD drive. The HDD drive either comes from the M.2 interface or the NVMe interface.

2OU server sled front plate shall have possibility to equip local mass memory /PCIe card options:

- Two M.2 cards
- 2 local 2,5” NVMe or SATA Disks (height of disk max 15mm)
- Two x16 FHHL PCIe slots

PCIe FHHL card

For the 2OU sled design, there are two I/O panels for the PCIe FHHL cards. Both CPU0 and CPU1 through a riser card to provide a PCIe x 16 interface to support the PCIe FHHL cards.

PWR Button

When the system off, push the PWR button can boot up the system. When the system on, push the PWR button can either make the system off or enter into suspend. Depend on the BIOS configuration.

RST Button

Push the RST button can reset the system. If the reset button is pushed, the unit is held in reset as long as reset button is pressed.

LED

Three LEDs are located on the bottom right corner. The blue LED shows the Power-on status. The green LED shows the HDD activity status. The yellow LED shows the alarm events. For the details of LED, refer to section 5.2.

3.3 Fan

2 OU configuration

For 2 OU configuration, the fan model is Sunon PF80381BX.

Fan specification summary:

Size L x H x W (mm) - 80 x 80 x 38

Power source - DC 12V

Power consumption - 48W

Max speed - 14000 RPM

Air Flow - 141.9 CFM

Noise - 72.1 dB

Each fan tray includes two fans.

Below a concept diagram shows the fan placement.

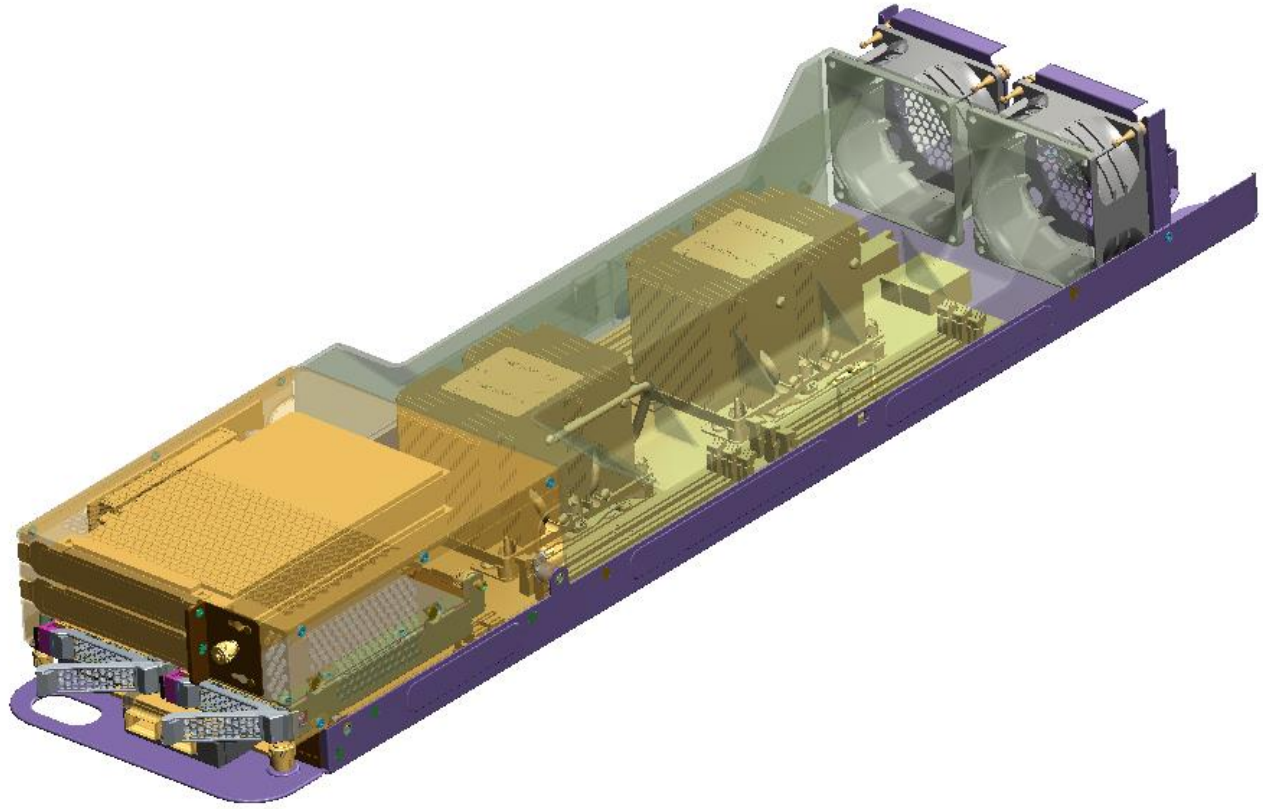


Figure 19 Server Sled Fantray for 2 OU Configuration

3.4 Cables

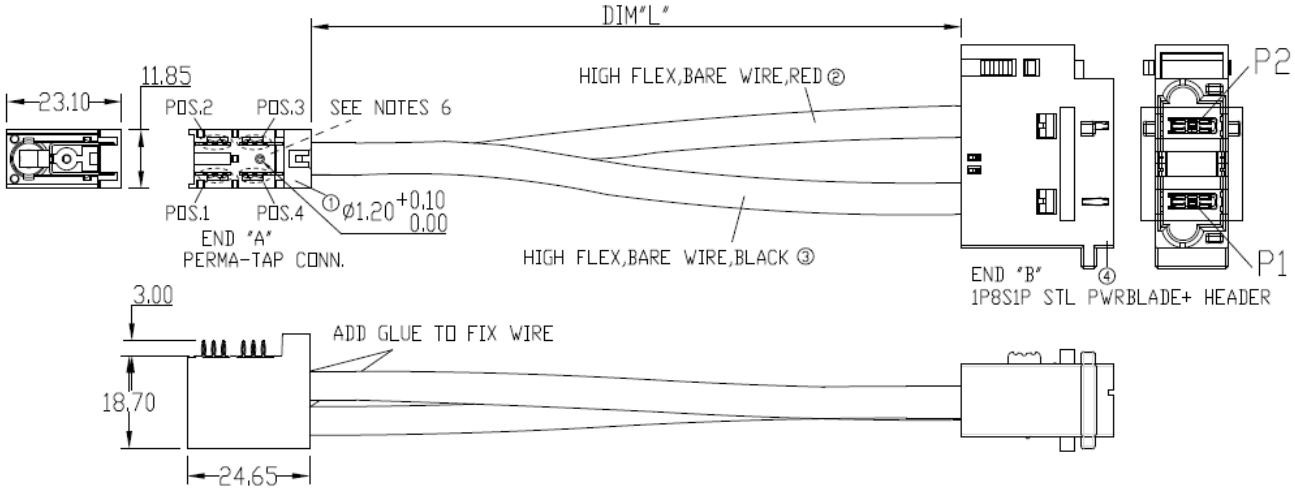
There are two kinds of cables for sled board assembly with Cubby and system rack.

Press-fit cable

This cable is press fitted onto the Server motherboard power connector to connect the motherboard with Cubby.

The connector type is AFCI 10136662-3A0026HLF.

The drawing is shown as below.



wiring list					
PIN DEFINE	END"A"	END"B"	WIRE SIZE	WIRE COLOR	DIM"L"
PWR	POS.1&POS.2	P2	8 AWG	RED	155±5
GND	POS.3&POS.4	P1	8 AWG	BLACK	145±5

Amphenol FCI

Figure 20 Server Sled Pressfit Cable Drawing

Medusa cable

This cable connects the Cubby with the system rack.

The connector type is AFCI 10130664-3C0060HLF.

The drawing is shown as below.

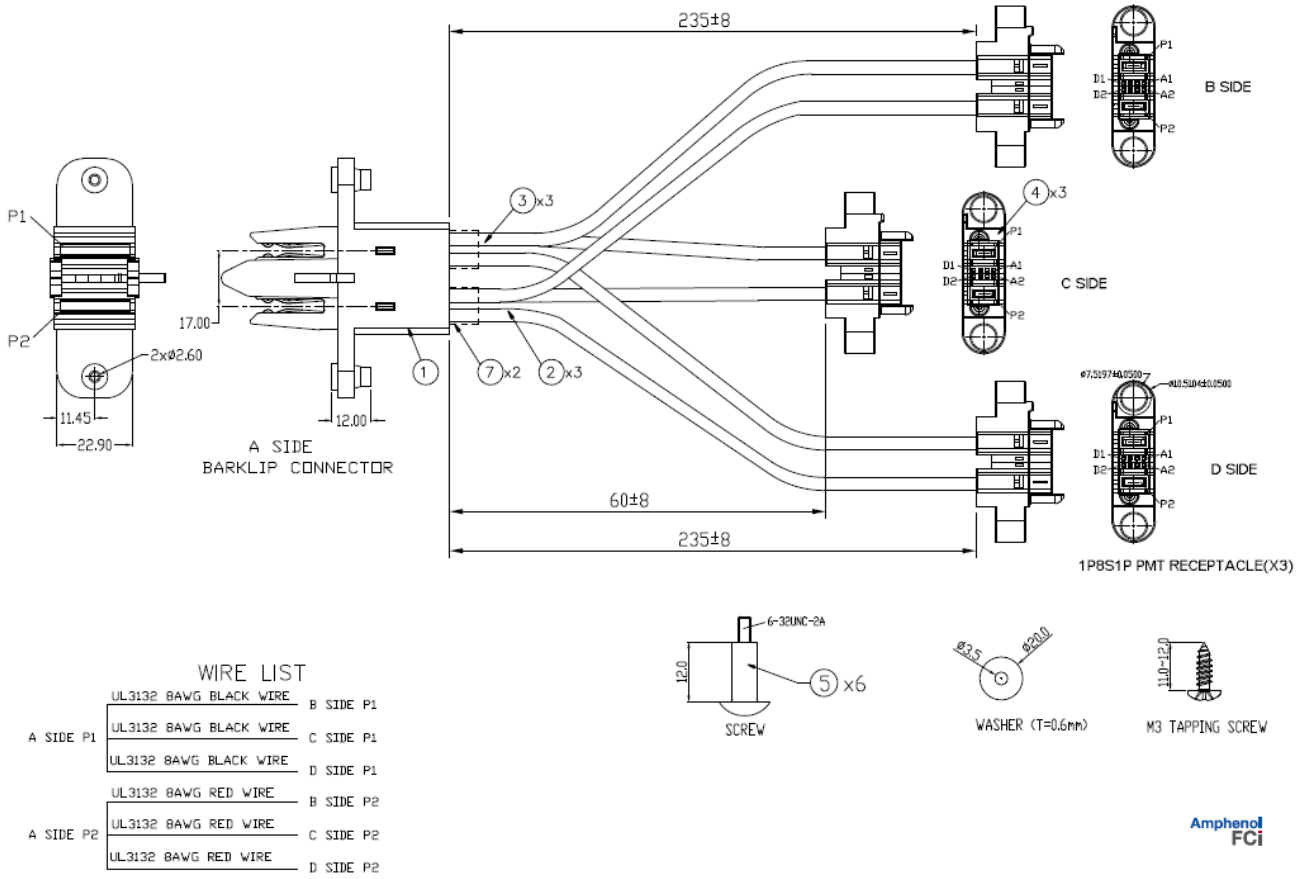


Figure 21 Cubby Medusa Cable Drawing

3.5 Board Subassembly

The board is installed into a cartridge. Inside the cartridge, there are one fan tray, one PCIe riser board, two storage bay, and one power connector.

Per the I/O capability difference, the cartridge has two types - 1OU and 2OU. The cartridge is then installed into a box called Cubby. The OCP standard defines two types for Cubby. The six-bay Cubby supports six 1OU OCP boards. The three-bay Cubby supports three 2OU OCP boards.

Below the 3D drawing shows the server sled installed with accessories.

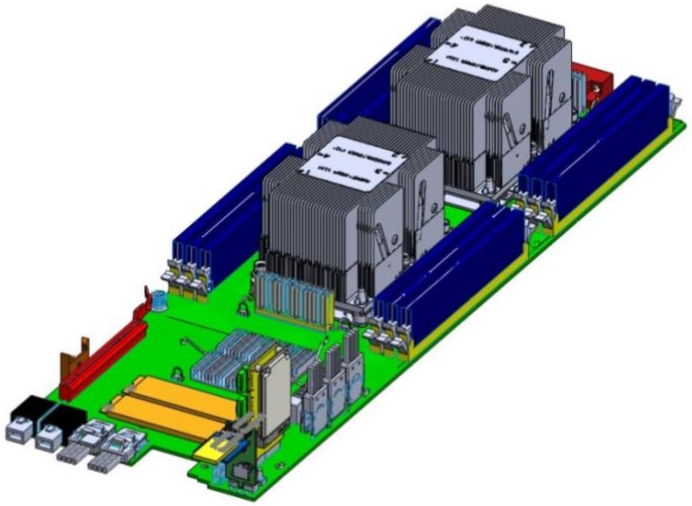


Figure 22 Server Board 3D Drawing

The figure below shows the Sled assembly with Cubby in 2 OU configuration.

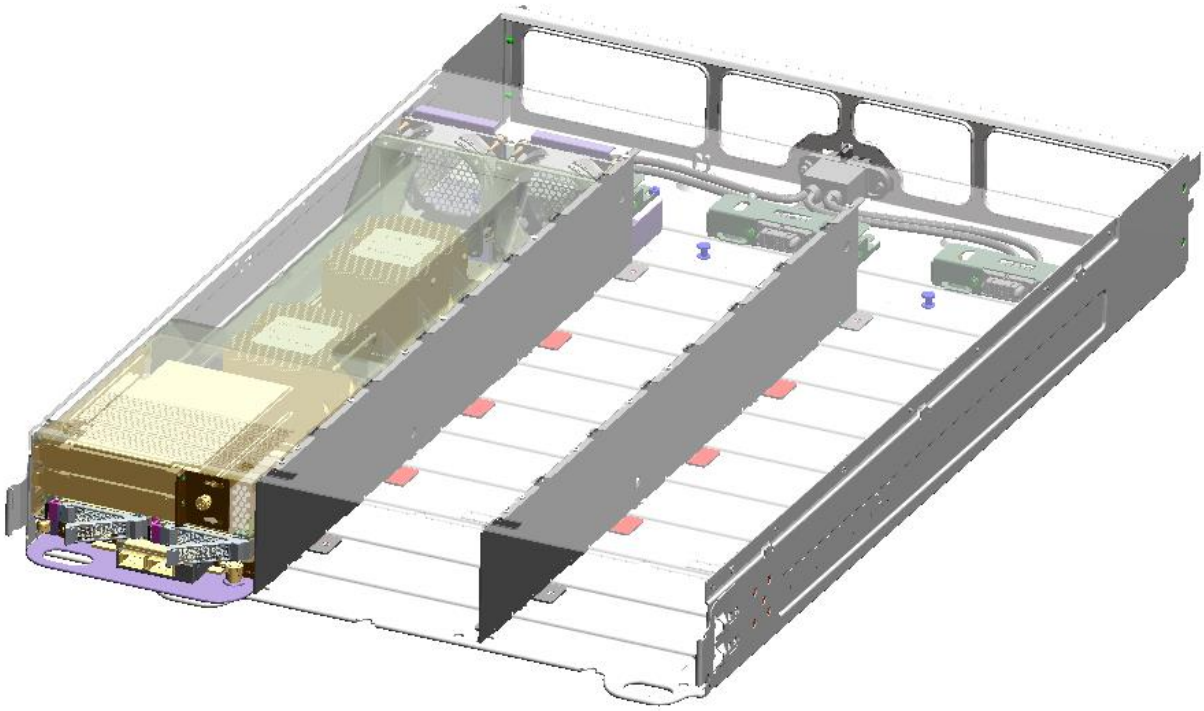


Figure 23 Server Sled Overview

3.6 Server photos

Photos of Server sled:

Figure 24 Server sled top view

Figure 25 Server base mechanics, motherboard, air guide and front assembly

4 Interfaces

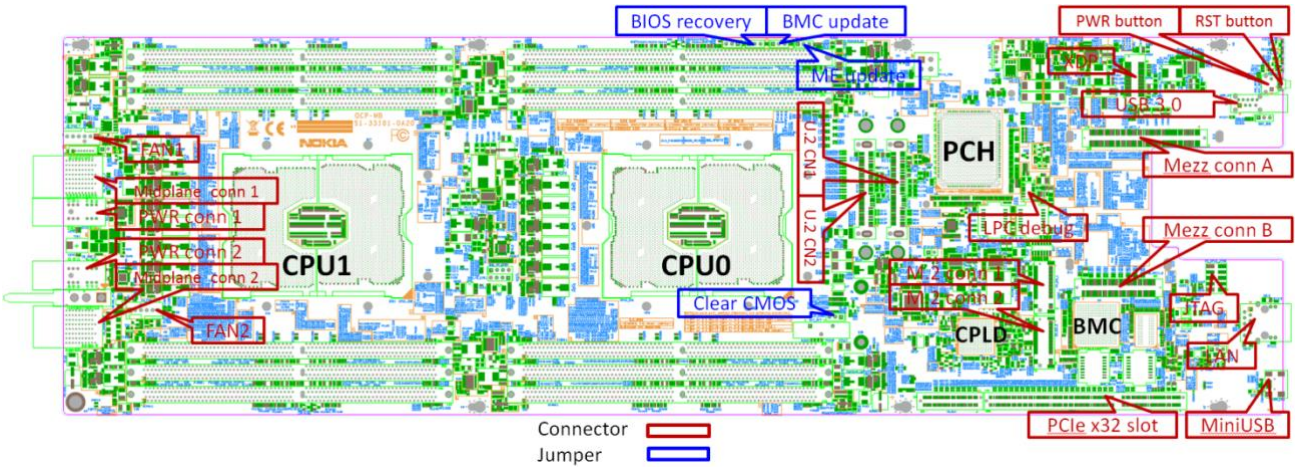
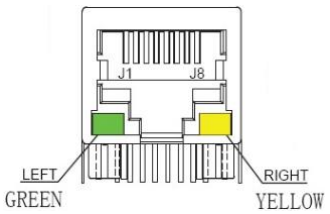


Figure 26 Connector and Jumper locations on Server main board

4.1 Interfaces at the Front Panel

4.1.1 Ethernet Port

There is a GbE network interface on the Server sled. It's routed with a RJ-45 port on the front panel.



PIN	Signal Name	In/Out
1	BI_DA+	Out
2	BI_DA-	Out
3	BI_DB+	In
4	BI_DC+	In/Out
5	BI_DC-	In/Out
6	BI_DB-	In
7	BI_DD+	In/Out
8	BI_DD-	In/Out
L1	LINK_ACT#	In/Out
L2	3.3V_STBY	In/Out
L3	LINK_100#	In/Out
L4	LINK_1000#	In/Out

Table 15 RJ-45 Connector Pinout

Two dual-color LEDs are built into the RJ-45 connector. Each LED can light on with yellow or green. The table below shows the LED status.

LED	LED color	Status
Right (Link status)	Off	Network link is not established
	Green (always on)	Link w/o activity
	Green (blinking)	Link with activity

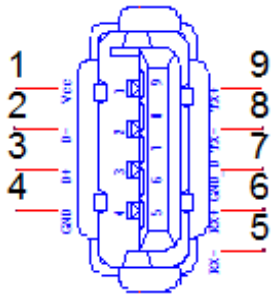
Left (Speed)	Off	10 Base-T
	Green	100 Base-T
	Yellow	1000 Base-T

Table 16 GbE port LED Status

4.1.2 USB 3.0

Server has one external Standard-A, right angle USB2.0/3/0 port on the front panel.

The connector type is Amphenol GSB4-116-344-HR.



PIN#	Signal Name	PIN#	Signal Name
1	5V	9	USB3_TX_P
2	USB2_N	8	USB3_TX_N
3	USB2_P	7	GND
4	GND	6	USB3_RX_P
		5	USB3_RX_N

Table 17 Standard-A USB 3.0 Port Pin Out

4.1.3 Mini USB

Server has one mini-B USB port on the front panel. However, this USB port is not for normal use. It's the debug port for BMC. Behind the mini USB connector, there is a USB-to-Serial controller to transfer the USB signals to UART. Then routed to the BMC.

PIN#	Signal Name	PIN#	Signal Name
1	5V	6	GND
2	USB2_N	7	GND
3	USB2_P	8	GND
4	NC	9	GND

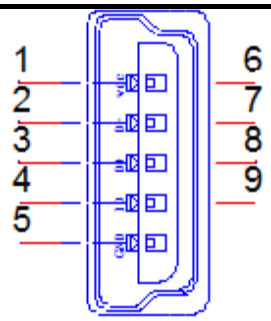
	5	GND		

Table 18 Mini-B USB Port Pin Out

The UART port is configured by BIOS to support the console emulation.

The console port configuration is detailed as the table below.

Configuration item	Option	Description	Default setting
Terminal Type	VT100, VT100+, VT-UTF8, ANSI	Configure the console emulation type	VT100
Bits per second	9600, 19200, 38400, 57600, 115200	Configure the console port baud rate	115200
Data Bits	7, 8	Configure the number of data bits in each transmitted or received serial character for both serial ports	8
Parity	None, Even, Odd	Configure if parity bit is generated (transmit data) or checked	None
Stop Bits	1, 2	Configure the number of stop bits transmitted and received in each serial character for both serial ports	1
Flow Control	None, hardware, software	Configure flow control for console redirection. Hardware flow control uses RTC/CTS. Software flow control uses XON/SOFF.	None
Terminal Size	80x24, 80x25	Set console display resolution	80x24

Table 19 Items of Console Port Configuration

4.2 Interfaces at the Rear Side

4.2.1 Power Input

An AirMax® power 2x2 connector is used for connecting the server sled with the power connector of three bay shelf (cubby). . The connector can serve 49A on 12VDC.

4.2.2 Signal Interconnect

Two AirMax® 3x8 connectors is reserved for connecting server board's high speed signals, side band signals, and management interface with midplane. In normal delivery these connectors are left unpopulated.

The connector drawing is shown as the figure below.

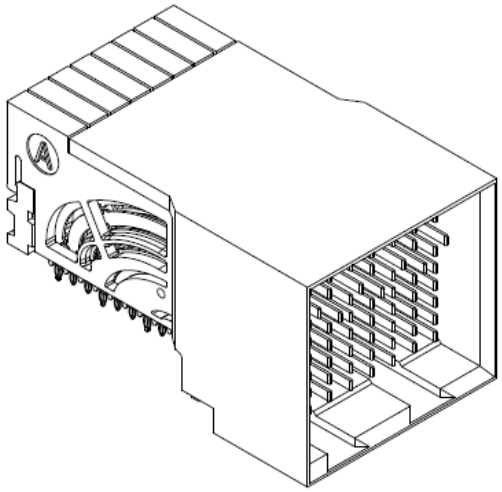


Figure 27 Middle plane Connector Drawing

	1	2	3	4	5	6	7	8
A	CPU1_PE1_RX_P0	GND	CPU1_PE1_RX_P6	FAN__TACH3	NC	PMBUS_ALERT#	CPU1_PE1_TX_P3	GND
B	CPU1_PE1_RX_N0	CPU1_PE1_RX_P3	CPU1_PE1_RX_N6	FAN__TACH2	NC	SMB_PMBUS_STBY_SDA	CPU1_PE1_TX_N3	CPU1_PE1_TX_P0
C	GND	CPU1_PE1_RX_N3	GND	FAN__TACH1	GND	SMB_PMBUS_STBY_SCL	GND	CPU1_PE1_TX_N0
D	CPU1_PE1_RX_P1	GND	CPU1_PE1_RX_P7	FAN__TACH0	NC	GND	CPU1_PE1_TX_P4	GND
E	CPU1_PE1_RX_N1	CPU1_PE1_RX_P4	CPU1_PE1_RX_N7	MATED_IN#	NC	CPU1_PE1_TX_P6	CPU1_PE1_TX_N4	CPU1_PE1_TX_P1

F	GND	CPU1_PE1_RX_N4	GND	FAN_PWM0	SLOT_ID 0	CPU1_PE1_TX_N6	GND	CPU1_PE1_TX_N1
G	CPU1_PE1_RX_P2	GND	MID_DBG_TX	GND	SLOT_ID 1	GND	CPU1_PE1_TX_P5	GND
H	CPU1_PE1_RX_N2	CPU1_PE1_RX_P5	MID_DBG_RX	CLK_100M_CN1_P	SLOT_ID 2	CPU1_PE1_TX_P7	CPU1_PE1_TX_N5	CPU1_PE1_TX_P2
I	GND	CPU1_PE1_RX_N5	MB_ON#	CLK_100M_CN1_N	PCIE_PERST#	CPU1_PE1_TX_N7	GND	CPU1_PE1_TX_N2

Table 20 Midplane Connector 1 Pin Out

	1	2	3	4	5	6	7	8
A	CPU1_PE1_RX_P8	GND	CPU1_PE1_RX_P14	NC	NC	GND	CPU1_PE1_TX_P11	GND
B	CPU1_PE1_RX_N8	CPU1_PE1_RX_P11	CPU1_PE1_RX_N14	NC	BMC_ALERT#	SMB_PEHPCPU1_STBY_SDA	CPU1_PE1_TX_N11	CPU1_PE1_TX_P8
C	GND	CPU1_PE1_RX_N11	GND	BMC_MP_GPIO	GND	SMB_PEHPCPU1_STBY_SCL	GND	CPU1_PE1_TX_N8
D	CPU1_PE1_RX_P9	GND	CPU1_PE1_RX_P15	NC	SMB_LAN_STBY_SDA	GND	CPU1_PE1_TX_P12	GND
E	CPU1_PE1_RX_N9	CPU1_PE1_RX_P12	CPU1_PE1_RX_N15	MATED_IN#	SMB_LAN_STBY_SCL	CPU1_PE1_TX_P14	CPU1_PE1_TX_N12	CPU1_PE1_TX_P9
F	GND	CPU1_PE1_RX_N12	GND	NC	NC	CPU1_PE1_TX_N14	GND	CPU1_PE1_TX_N9
G	CPU1_PE1_RX_P10	GND	NC	GND	XRC_READY_N	GND	CPU1_PE1_TX_P13	GND
H	CPU1_PE1_RX_N10	CPU1_PE1_RX_P13	NC	CLK_100M_CN2_P	XRC_PRESENT_N	CPU1_PE1_TX_P15	CPU1_PE1_TX_N13	CPU1_PE1_TX_P10
I	GND	CPU1_PE1_RX_N13	MB_ON#	CLK_100M_CN2_N	GND	CPU1_PE1_TX_N15	GND	CPU1_PE1_TX_N10

Table 21 Midplane Connector 2 Pin Out

4.3 Interfaces inside the Board

4.3.1 M.2 Connector

Server has two onboard M.2 connectors with Key ID=M and H6.7 Type. The M.2 connector has a PCIe x1 lane and an optional connection of PCIe x1 from PCH or SATA port from PCH. The M.2 connectors support 2280, 22110 card form factor with both single-side and double-side. The side band signals such as WAKE# and SATA_ACTIVITY should be connected when applied. The PERST# signal shall go active before the power on M.2 connector is removed per the PCI CEM specification.

The connector type is Foxconn/2E0BC21-S85BM-7H. The drawing is shown as the figure below.

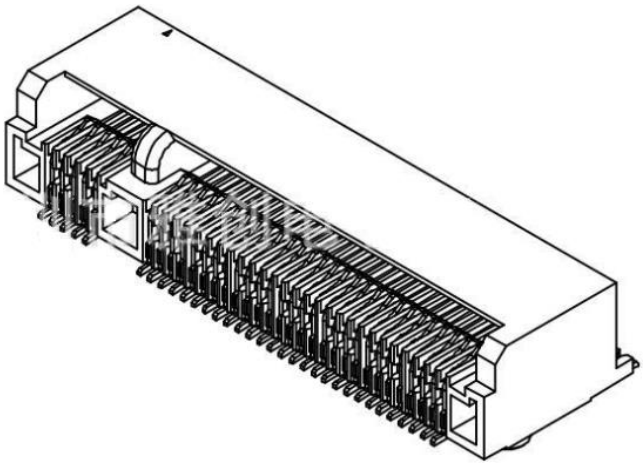


Figure 28 M.2 Connector Drawing

Signal	Pin	Pin	Signal
GND	1	2	3.3V
GND	3	4	3.3V
NC	5	6	NC
NC	7	8	NC
GND	9	10	LED_ACT#
NC	11	12	3.3V
NC	13	14	3.3V
GND	15	16	3.3V
NC	17	18	3.3V
NC	19	20	NC
GND	21	22	NC
NC	23	24	NC
NC	25	26	NC

GND	27	28	NC
P3E_PCH_RX_N	29	30	NC
P3E_PCH_RX_P	31	32	NC
GND	33	34	NC
P3E_PCH_TX_N	35	36	NC
P3E_PCH_TX_P	37	38	Test Pad
GND	39	40	NC
MUX_PCIE_SATA_RX_P	41	42	NC
MUX_PCIE_SATA_RX_N	43	44	NC
GND	45	46	NC
MUX_PCIE_SATA_TX_N	47	48	NC
MUX_PCIE_SATA_TX_P	49	50	Test Pad
GND	51	52	Test Pad
CLK_100M_N	53	54	Test Pad
CLK_100M_P	55	56	NC
GND	57	58	NC
Separation Keying			
NC	67	68	Test Pad
SSD_PEDET	69	70	3.3V
GND	71	72	3.3V
GND	73	74	3.3V
GND	75	76	

Table 22 M.2 Connector Pin Out

4.3.2 U.2 Connector

Server have two U.2 connectors. The U.2 is originally known as the SFF-8639. The connector provides more available pins for the high speed signals such as the PCIe and SATA usage. On Server sled, each U.2 connector supports PCIe x4 lanes and one SATA port.

The connector type is Molex/78757-0001. The drawing is shown as the figure below.

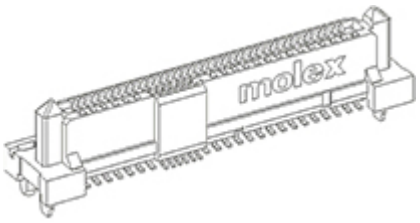


Figure 29 U.2 Connector Drawing

NOTE : The U.2 connector is a non-symmetrical connector. The pin out configuration is shown as below.

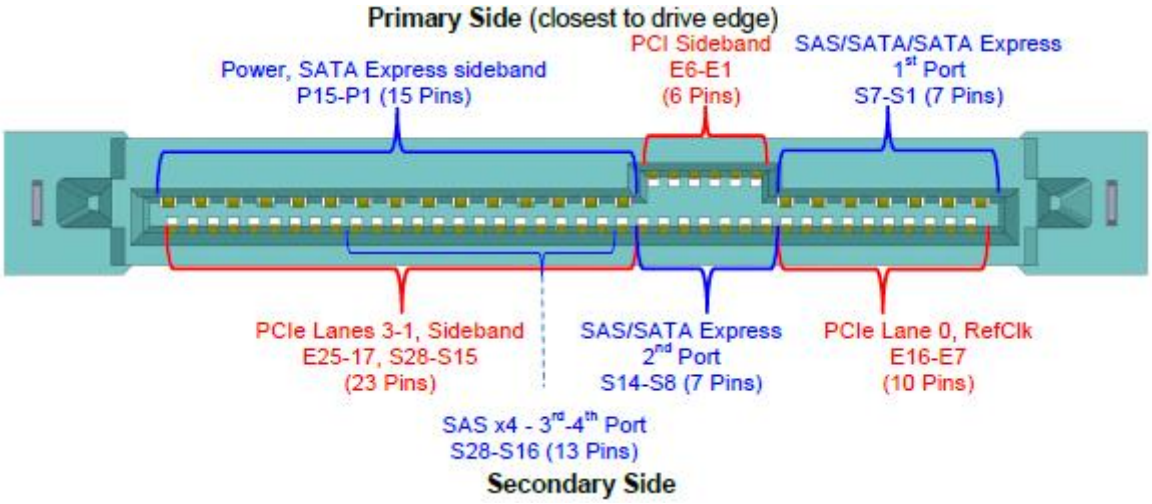


Figure 30 U.2 Connector Pin Out Configuration

Signal	Pin	Pin	Signal
GND	S1	E7	CLK1_100M_P
SATA_TX_P0	S2	E8	CLK1_100M_N
SATA_TX_N0	S3	E9	GND
GND	S4	E10	CPU0_PE1_TX_P0
SATA_RX_N0	S5	E11	CPU0_PE1_TX_N0
SATA_RX_P0	S6	E12	GND
GND	S7	E13	CPU0_PE1_RX_N0

		E14	CPU0_PE1_RX_P0
		E15	GND
		E16	NC
NC	E1	S8	GND
NC	E2	S9	NC
3.3V	E3	S10	NC
NC	E4	S11	GND
RESET#	E5	S12	NC
NC	E6	S13	NC
		S14	GND
3.3V	P1	S15	NC
3.3V	P2	S16	GND
3.3V	P3	S17	CPU0_PE1_TX_P1
GND	P4	S18	CPU0_PE1_TX_N1
	P5	S19	GND
	P6	S20	CPU0_PE1_RX_N1
5V	P7	S21	CPU0_PE1_RX_P1
5V	P8	S22	GND
5V	P9	S23	CPU0_PE1_TX_P2
NC	P10	S24	CPU0_PE1_TX_N2
LED1_ACT#	P11	S25	GND
GND	P12	S26	CPU0_PE1_RX_N2
12V	P13	S27	CPU0_PE1_RX_P2
12V	P14	S28	GND
12V	P15	E17	CPU0_PE1_TX_P3
		E18	CPU0_PE1_TX_N3
		E19	GND
		E20	CPU0_PE1_RX_N3
		E21	CPU0_PE1_RX_P3
		E22	GND
		E23	SMB_STBY_SCL
		E24	SMB_STBY_SDA

		E25	NC

Table 23 U.2 Connector 1 Pin Out

Signal	Pin	Pin	Signal
GND	S1	E7	CLK2_100M_P
SATA_TX_P1	S2	E8	CLK2_100M_N
SATA_TX_N1	S3	E9	GND
GND	S4	E10	CPU0_PE1_TX_P4
SATA_RX_N1	S5	E11	CPU0_PE1_TX_N4
SATA_RX_P1	S6	E12	GND
GND	S7	E13	CPU0_PE1_RX_N4
		E14	CPU0_PE1_RX_P4
		E15	GND
		E16	NC
NC	E1	S8	GND
NC	E2	S9	NC
3.3V	E3	S10	NC
NC	E4	S11	GND
RESET#	E5	S12	NC
NC	E6	S13	NC
		S14	GND
3.3V	P1	S15	NC
3.3V	P2	S16	GND
3.3V	P3	S17	CPU0_PE1_TX_P5
GND	P4	S18	CPU0_PE1_TX_N5
	P5	S19	GND
	P6	S20	CPU0_PE1_RX_N5
5V	P7	S21	CPU0_PE1_RX_P5
5V	P8	S22	GND

5V	P9	S23	CPU0_PE1_TX_P6
NC	P10	S24	CPU0_PE1_TX_N6
LED1_ACT#	P11	S25	GND
GND	P12	S26	CPU0_PE1_RX_N6
12V	P13	S27	CPU0_PE1_RX_P6
12V	P14	S28	GND
12V	P15	E17	CPU0_PE1_TX_P7
		E18	CPU0_PE1_TX_N7
		E19	GND
		E20	CPU0_PE1_RX_N7
		E21	CPU0_PE1_RX_P7
		E22	GND
		E23	SMB_STBY_SCL
		E24	SMB_STBY_SDA
		E25	NC

Table 24 U.2 Connector 2 Pin Out

4.3.3 Mezzanine Connector

Open Rack Server has the OCP Mezz connector A and connector B to provide up to PCIe Gen3 x16 connection for Mezzanine card. Both connectors are compliant with the OCP Mezzanine Card specification v2.0.

Connector A has 120 pin counts. The connector type is FCI/61082-122402LF.

Connector B has 80 pin counts. The connector type is FCI/61082-082402LF.

The connector drawing is shown as the figure below. Both connectors have the same aspect but different pin count.

Figure 31 Mezzanine Connector Drawing

Signal	Pin	Pin	Signal
MEZZA_PRSNT1_N	1	2	12V
5V_STBY	3	4	12V
5V_STBY	5	6	12V
5V_STBY	7	8	GND
GND	9	10	GND
GND	11	12	3.3V_STBY
3.3V_STBY	13	14	GND
GND	15	16	GND
GND	17	18	3.3V
3.3V	19	20	3.3V
3.3V	21	22	3.3V
3.3V	23	24	3.3V
3.3V	25	26	GND
BMC_RMII2_CRS_DV	27	28	IRQ_MEZZ_LAN_ALERT_N
MEZZ_50M_CLK	29	30	SMB_FRU_STBY_SCL
BMC_RMII2_TX_EN	31	32	SMB_FRU_STBY_SDA
RST_PCIE_CPU_DEV0_N	33	34	FM_PE_WAKE#
SMB_LAN_STBY_SCL	35	36	BMC_RMII2_RX_ER
SMB_LAN_STBY_SDA	37	38	GND
GND	39	40	BMC_RMII2_TXD0
GND	41	42	BMC_RMII2_TXD1
BMC_RMII2_RXD0	43	44	GND
BMC_RMII2_RXD1	45	46	GND
GND	47	48	CLK_100M_MEZZ1_P
GND	49	50	CLK_100M_MEZZ1_N

CLK_100M_MEZZ2_P	51	52	GND
CLK_100M_MEZZ2_N	53	54	GND
GND	55	56	CPU0_PCIE_PE3_TX_P7
GND	57	58	CPU0_PCIE_PE3_TX_N7
CPU0_PCIE_PE3_RX_P7	59	60	GND
CPU0_PCIE_PE3_RX_N7	61	62	GND
GND	63	64	CPU0_PCIE_PE3_TX_P6
GND	65	66	CPU0_PCIE_PE3_TX_N6
CPU0_PCIE_PE3_RX_P6	67	68	GND
CPU0_PCIE_PE3_RX_N6	69	70	GND
GND	71	72	CPU0_PCIE_PE3_TX_P5
GND	73	74	CPU0_PCIE_PE3_TX_N5
CPU0_PCIE_PE3_RX_P5	75	76	GND
CPU0_PCIE_PE3_RX_N5	77	78	GND
GND	79	80	CPU0_PCIE_PE3_TX_P4
GND	81	82	CPU0_PCIE_PE3_TX_N4
CPU0_PCIE_PE3_RX_P4	83	84	GND
CPU0_PCIE_PE3_RX_N4	85	86	GND
GND	87	88	CPU0_PCIE_PE3_TX_P3
GND	89	90	CPU0_PCIE_PE3_TX_N3
CPU0_PCIE_PE3_RX_P3	91	92	GND
CPU0_PCIE_PE3_RX_N3	93	94	GND
GND	95	96	CPU0_PCIE_PE3_TX_P2
GND	97	98	CPU0_PCIE_PE3_TX_N2
CPU0_PCIE_PE3_RX_P2	99	100	GND
CPU0_PCIE_PE3_RX_N2	101	102	GND
GND	103	104	CPU0_PCIE_PE3_TX_P1
GND	105	106	CPU0_PCIE_PE3_TX_N1
CPU0_PCIE_PE3_RX_P1	107	108	GND
CPU0_PCIE_PE3_RX_N1	109	110	GND
GND	111	112	CPU0_PCIE_PE3_TX_P0
GND	113	114	CPU0_PCIE_PE3_TX_N0
CPU0_PCIE_PE3_RX_P0	115	116	GND
CPU0_PCIE_PE3_RX_N0	117	118	GND

GND	119	120	OCP_MEZZA_PRES_N

Table 25 Mezzanine Connector A Pin Out

Signal	Pin	Pin	Signal
MEZZA_PRSNT2_N	1	2	12V_STBY
GND	3	4	12V_STBY
CPU0_PCIE_PE3_RX_P15	5	6	NC
CPU0_PCIE_PE3_RX_N15	7	8	GND
GND	9	10	CPU0_PCIE_PE3_TX_P15
GND	11	12	CPU0_PCIE_PE3_TX_N15
CPU0_PCIE_PE3_RX_P14	13	14	GND
CPU0_PCIE_PE3_RX_N14	15	16	GND
GND	17	18	CPU0_PCIE_PE3_TX_P14
GND	19	20	CPU0_PCIE_PE3_TX_N14
CPU0_PCIE_PE3_RX_P13	21	22	GND
CPU0_PCIE_PE3_RX_N13	23	24	GND
GND	25	26	CPU0_PCIE_PE3_TX_P13
GND	27	28	CPU0_PCIE_PE3_TX_N13
CPU0_PCIE_PE3_RX_P12	29	30	GND
CPU0_PCIE_PE3_RX_N12	31	32	GND
GND	33	34	CPU0_PCIE_PE3_TX_P12
GND	35	36	CPU0_PCIE_PE3_TX_N12
CPU0_PCIE_PE3_RX_P11	37	38	GND
CPU0_PCIE_PE3_RX_N11	39	40	GND
GND	41	42	CPU0_PCIE_PE3_TX_P11
GND	43	44	CPU0_PCIE_PE3_TX_N11
CPU0_PCIE_PE3_RX_P10	45	46	GND
CPU0_PCIE_PE3_RX_N10	47	48	GND
GND	49	50	CPU0_PCIE_PE3_TX_P10
GND	51	52	CPU0_PCIE_PE3_TX_N10

CPU0_PCIE_PE3_RX_P9	53	54	GND
CPU0_PCIE_PE3_RX_N9	55	56	GND
GND	57	58	CPU0_PCIE_PE3_TX_P9
GND	59	60	CPU0_PCIE_PE3_TX_N9
CPU0_PCIE_PE3_RX_P8	61	62	GND
CPU0_PCIE_PE3_RX_N8	63	64	GND
GND	65	66	CPU0_PCIE_PE3_TX_P8
GND	67	68	CPU0_PCIE_PE3_TX_N8
CLK_100M_MEZZ3_P	69	70	GND
CLK_100M_MEZZ3_N	71	72	GND
GND	73	74	CLK_100M_MEZZ4_P
RST_PCIE_CPU_DEV1_N	75	76	CLK_100M_MEZZ4_N
RST_PCIE_CPU_DEV2_N	77	78	GND
RST_PCIE_CPU_DEV3_N	79	80	OCP_MEZZB_PRES_N

Table 26 Mezzanine Connector B Pin Out

4.3.4 Fan Connector

There are two fan male connectors on server motherboard. The connector has 6 pin with pitch 2.54mm.

For 2 OU configuration, the fan female connector has 4 pins. It plugs into the fan male connector pin 1-4. For 1 OU configuration, the fan female connector has 6 pins. Plug in directly.

	PIN#	Signal Name
	1	GND
	2	12V
	3	TACH0
	4	PWM

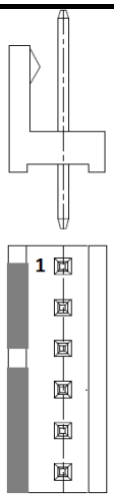
	5	TACH1
	6	NC

Table 27 Fan Connector Pin Out

4.4 PCIe Riser Slot

Open Rack Server has one PCIe x32 slot to be used by the PCIe riser cards. Both CPU0 and CPU1 route PCIe x16 lanes to the riser slot.

The x32 riser slot is assembled by one x24 slot (Samtec/HSEC8-1100-01-L-DV-A-K) and one x8 slot (Samtec/HSEC8-130-01-L-DV-A-TR) to support two types riser cards.

Type-A Riser: One slot type. Slot1 is PCIEx16 connector with x16 bus signal from CPU1 and support FHHL PCIE card.

Type-B Riser: Two slots type. Bottom slot (Slot2) is x8 PCIe connector with x8 bus signal from CPU0 and support FHHL PCIe card. TOP slot (Slot3) is x16 PCIe connector with x16 PCIe bus signal from CPU1 and also support FHHL PCIe card.

The slot drawing is shown as the figure below. Both the x24 and x8 slot have the same aspect but different pin count.

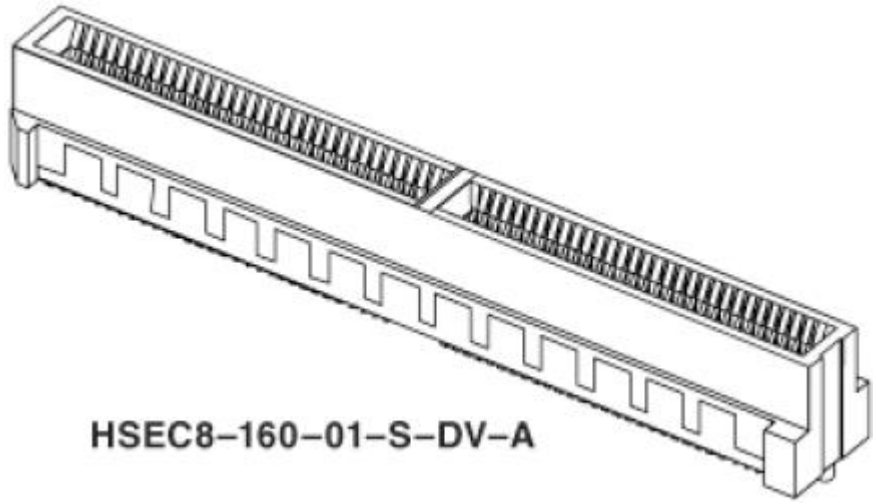


Figure 32 Riser Slot Drawing

Signal	Pin	Pin	Signal
12V	1	2	12V
12V	3	4	12V
12V	5	6	12V
12V	7	8	12V
12V	9	10	12V
12V	11	12	12V
GND	13	14	GND
BMC_SMB_SDA	15	16	SMB_SLOTX32_ALERT
BMC_SMB_SCL	17	18	PCIE_SLO1_CONFIG
PCIE_SLO0_CONFIG	19	20	PWRBRK_SLOT#
3.3V	21	22	3.3V
3.3V	23	24	3.3V
3.3V_SB	25	26	GND
SMB_HOST_STBY_ALERT	27	28	SLOTX32_WAKE#
RST_PCIE_X32_RISER#	29	30	SMB_HOST_STBY_SDA
GND	31	32	SMB_HOST_STBY_SCL
CLK_100M_PCH_SLOTX32_P0	33	34	GND
CLK_100M_PCH_SLOTX32_N0	35	36	CLK_100M_PCH_SLOTX32_P1
GND	37	38	CLK_100M_PCH_SLOTX32_N1

CLK_100M_PCH_SLOTX32_P2	39	40	GND
CLK_100M_PCH_SLOTX32_N2	41	42	CLK_100M_PCH_SLOTX32_P3
GND	43	44	CLK_100M_PCH_SLOTX32_N3
CLK_100M_PCH_SLOTX32_P4	45	46	GND
CLK_100M_PCH_SLOTX32_N4	47	48	CLK_100M_PCH_SLOTX32_P5
GND	49	50	CLK_100M_PCH_SLOTX32_N5
CPU0_PCIE_PE2_TX_N0	51	52	GND
CPU0_PCIE_PE2_TX_P0	53	54	CPU0_PCIE_PE2_RX_N0
GND	55	56	CPU0_PCIE_PE2_RX_P0
CPU0_PCIE_PE2_TX_N1	57	58	GND
CPU0_PCIE_PE2_TX_P1	59	60	CPU0_PCIE_PE2_RX_N1
GND	61	62	CPU0_PCIE_PE2_RX_P1
NC	63	64	GND
GND	65	66	NC
CPU0_PCIE_PE2_TX_N2	67	68	GND
CPU0_PCIE_PE2_TX_P2	69	70	CPU0_PCIE_PE2_RX_N2
GND	71	72	CPU0_PCIE_PE2_RX_P2
CPU0_PCIE_PE2_TX_N3	73	74	GND
CPU0_PCIE_PE2_TX_P3	75	76	CPU0_PCIE_PE2_RX_N3
GND	77	78	CPU0_PCIE_PE2_RX_P3
CPU0_PCIE_PE2_TX_N4	79	80	GND
CPU0_PCIE_PE2_TX_P4	81	82	CPU0_PCIE_PE2_RX_N4
GND	83	84	CPU0_PCIE_PE2_RX_P4
CPU0_PCIE_PE2_TX_N5	85	86	GND
CPU0_PCIE_PE2_TX_P5	87	88	CPU0_PCIE_PE2_RX_N5
GND	89	90	CPU0_PCIE_PE2_RX_P5
CPU0_PCIE_PE2_TX_N6	91	92	GND
CPU0_PCIE_PE2_TX_P6	93	94	CPU0_PCIE_PE2_RX_N6
GND	95	96	CPU0_PCIE_PE2_RX_P6
CPU0_PCIE_PE2_TX_N7	97	98	GND
CPU0_PCIE_PE2_TX_P7	99	100	CPU0_PCIE_PE2_RX_N7
GND	101	102	CPU0_PCIE_PE2_RX_P7
CPU0_PCIE_PE2_TX_N8	103	104	GND
CPU0_PCIE_PE2_TX_P8	105	106	CPU0_PCIE_PE2_RX_N8

GND	107	108	CPU0_PCIE_PE2_RX_P8
CPU0_PCIE_PE2_TX_N9	109	110	GND
CPU0_PCIE_PE2_TX_P9	111	112	CPU0_PCIE_PE2_RX_N9
GND	113	114	CPU0_PCIE_PE2_RX_P9
CPU0_PCIE_PE2_TX_N10	115	116	GND
CPU0_PCIE_PE2_TX_P10	117	118	CPU0_PCIE_PE2_RX_N10
GND	119	120	CPU0_PCIE_PE2_RX_P10
CPU0_PCIE_PE2_TX_N11	121	122	GND
CPU0_PCIE_PE2_TX_P11	123	124	CPU0_PCIE_PE2_RX_N11
GND	125	126	CPU0_PCIE_PE2_RX_P11
CPU0_PCIE_PE2_TX_N12	127	128	GND
CPU0_PCIE_PE2_TX_P12	129	130	CPU0_PCIE_PE2_RX_N12
GND	131	132	CPU0_PCIE_PE2_RX_P12
CPU0_PCIE_PE2_TX_N13	133	134	GND
CPU0_PCIE_PE2_TX_P13	135	136	CPU0_PCIE_PE2_RX_N13
GND	137	138	CPU0_PCIE_PE2_RX_P13
CPU0_PCIE_PE2_TX_N14	139	140	GND
CPU0_PCIE_PE2_TX_P14	141	142	CPU0_PCIE_PE2_RX_N14
GND	143	144	CPU0_PCIE_PE2_RX_P14
CPU0_PCIE_PE2_TX_N15	145	146	GND
CPU0_PCIE_PE2_TX_P15	147	148	CPU0_PCIE_PE2_RX_N15
GND	149	150	CPU0_PCIE_PE2_RX_P15
CPU1_PCIE_PE2_TX_N0	151	152	GND
CPU1_PCIE_PE2_TX_P0	153	154	CPU1_PCIE_PE2_RX_N0
GND	155	156	CPU1_PCIE_PE2_RX_P0
CPU1_PCIE_PE2_TX_N1	157	158	GND
CPU1_PCIE_PE2_TX_P1	159	160	CPU1_PCIE_PE2_RX_N1
GND	161	162	CPU1_PCIE_PE2_RX_P1
CPU1_PCIE_PE2_TX_N2	163	164	GND
CPU1_PCIE_PE2_TX_P2	165	166	CPU1_PCIE_PE2_RX_N2
GND	167	168	CPU1_PCIE_PE2_RX_P2
CPU1_PCIE_PE2_TX_N3	169	170	GND
CPU1_PCIE_PE2_TX_P3	171	172	CPU1_PCIE_PE2_RX_N3
GND	173	174	CPU1_PCIE_PE2_RX_P3

CPU1_PCIE_PE2_TX_N4	175	176	GND
CPU1_PCIE_PE2_TX_P4	177	178	CPU1_PCIE_PE2_RX_N4
GND	179	180	CPU1_PCIE_PE2_RX_P4
CPU1_PCIE_PE2_TX_N5	181	182	GND
CPU1_PCIE_PE2_TX_P5	183	184	CPU1_PCIE_PE2_RX_N5
GND	185	186	CPU1_PCIE_PE2_RX_P5
CPU1_PCIE_PE2_TX_N6	187	188	GND
CPU1_PCIE_PE2_TX_P6	189	190	CPU1_PCIE_PE2_RX_N6
GND	191	192	CPU1_PCIE_PE2_RX_P6
CPU1_PCIE_PE2_TX_N7	193	194	GND
CPU1_PCIE_PE2_TX_P7	195	196	CPU1_PCIE_PE2_RX_N7
GND	197	198	CPU1_PCIE_PE2_RX_P7
PCIEX32_RISE_PRSNT	199	200	GND

Table 28 PCIe x24 Slot Pin Out

Signal	Pin	Pin	Signal
GND	1	2	NC
CPU1_PCIE_PE2_TX_N8	3	4	GND
CPU1_PCIE_PE2_TX_P8	5	6	CPU1_PCIE_PE2_RX_N8
GND	7	8	CPU1_PCIE_PE2_RX_P8
CPU1_PCIE_PE2_TX_N9	9	10	GND
CPU1_PCIE_PE2_TX_P9	11	12	CPU1_PCIE_PE2_RX_N9
GND	13	14	CPU1_PCIE_PE2_RX_P9
CPU1_PCIE_PE2_TX_N10	15	16	GND
CPU1_PCIE_PE2_TX_P10	17	18	CPU1_PCIE_PE2_RX_N10
GND	19	20	CPU1_PCIE_PE2_RX_P10
CPU1_PCIE_PE2_TX_N11	21	22	GND
CPU1_PCIE_PE2_TX_P11	23	24	CPU1_PCIE_PE2_RX_N11
GND	25	26	CPU1_PCIE_PE2_RX_P11
CPU1_PCIE_PE2_TX_N12	27	28	GND
CPU1_PCIE_PE2_TX_P12	29	30	CPU1_PCIE_PE2_RX_N12

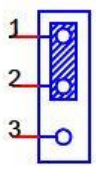
GND	31	32	CPU1_PCIE_PE2_RX_P12
CPU1_PCIE_PE2_TX_N13	33	34	GND
CPU1_PCIE_PE2_TX_P13	35	36	CPU1_PCIE_PE2_RX_N13
GND	37	38	CPU1_PCIE_PE2_RX_P13
CPU1_PCIE_PE2_TX_N14	39	40	GND
CPU1_PCIE_PE2_TX_P14	41	42	CPU1_PCIE_PE2_RX_N14
GND	43	44	CPU1_PCIE_PE2_RX_P14
CPU1_PCIE_PE2_TX_N15	45	46	GND
CPU1_PCIE_PE2_TX_P15	47	48	CPU1_PCIE_PE2_RX_N15
GND	49	50	CPU1_PCIE_PE2_RX_P15
CLK_100M_PCH_SLOTX32_P6	51	52	GND
CLK_100M_PCH_SLOTX32_N6	53	54	CLK_100M_PCH_SLOTX32_P7
GND	55	56	CLK_100M_PCH_SLOTX32_N7
USB2_PCIESLOT_N	57	58	GND
USB2_PCIESLOT_P	59	60	NC

Table 29 PCIe x8 Slot Pin Out

4.5 Jumper

There are four jumpers reserved for debug purpose. The jumper headers are all with pitch 2.54 mm and located on the component side of the PCB. The detailed setting of the jumpers is described in the subsections below.

4.5.1 ME Update



Header Position	Function
1-2	Normal operation *(DEFAULT)*
2-3	Force ME update

Table 30 ME Update Jumper Setting

4.5.2 BMC Update

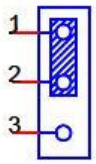
	Header Position	Function
	1-2	Normal operation *(DEFAULT)*
	2-3	Force BMC update

Table 31 BMC Update Jumper Setting

4.5.3 BIOS Recovery

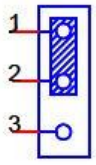
	Header Position	Function
	1-2	Normal operation *(DEFAULT)*
	2-3	Recover BIOS

Table 32 BIOS Recovery Jumper Setting

4.5.4 Clear CMOS

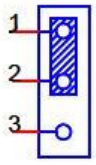
	Header Position	Function
	1-2	Normal operation *(DEFAULT)*
	2-3	Clear CMOS registers

Table 33 Clear CMOS Jumper Setting

5 Operating the unit

5.1 Unit Start-up

Open Rack Server is powered up once the 12V DC power rail is enabled by the hot swap controller. There are two regulator groups, one for standby power regulation and the other for normal power. At first the standby power group regulates the 12V to several standby power rails for CPLD, PCH, and BMC. The CPLD needs several milliseconds to initialize, then turn on all the normal power rails in a well-defined sequence. After that, the BIOS boot code starts loading.

Below a concept diagram shows the unit start-up sequence.

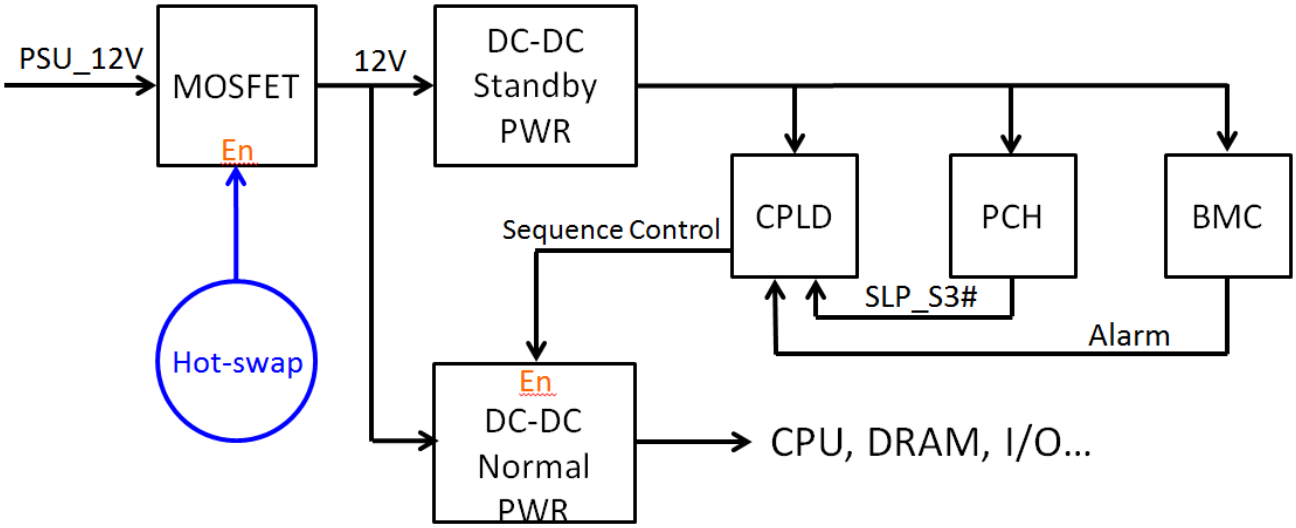


Figure 33 Server Start-Up Concept Diagram

Server is equipped with two redundant boot flashes, where BIOS images are stored, connected to the PCH SPI interface. The active BIOS flash after reset is the one that was active before the reset.

A watchdog timer is enabled along with the BIOS loading. If the BIOS fails loading and executing the boot code or the active BIOS flash is corrupted, it will not keep booting system. The boot failure leads to the watchdog timeout. Then BMC will handle this condition to do BIOS switchover.

When the BIOS code running finished, the Server will boot up into the OS. User can check the boot up screen via serial console.

5.2 LED

There are three LEDs on the front panel of server sled. The 1st one is the PWR LED for showing power status. The 2nd one is the ACT LED for showing HDD status. The 3rd one is the Alarm LED. The following table shows the LED color and the corresponding functionality.

LED color	Function	Silk Screen
Blue	Power LED Always On: Payload power is ready Blink: Payload power failed and BMC retry	PWR
Green	HDD activity Any activity on the SATA, mSATA, and M.2 interface, the LED shall be on.	HDD
Yellow	Alarm LED When pre-defined events happen, the LED shall be on. The pre-defined events such as the BIOS running error, WDT timeout, sensor out of limit, etc.	ALARM

Table 34 Front Panel LED Definition

5.3 Reset Logic

This section describes the functionality and implementation of the reset logic on server sled. The reset scheme is described below.

5.3.1 Power-up Reset

The power-up reset only happens when the Server is first powered up or if the BMC cycles the system power due to voltage monitoring event. The power-up reset completely resets all the logics on the server except the BMC.

5.3.2 Reset Button

A tact switch (SW_RESET1) is added for generating the reset. In Server Module the Reset Switch is connected to PCH SYS_RESET# Pin. Push the button will make the SYS_RESET# assertion, then the whole system will be reset except the BMC.

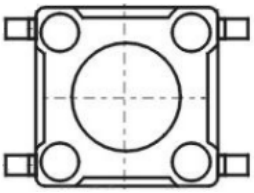


Figure 34 Reset Button

5.3.3 BMC Initiated Reset

When the following events happen, the BMC will send a reset:

- BIOS updated: The purpose is to make new BIOS take effect. BMC will reset the whole system except BMC itself.
- Watchdog timeout: BMC will issue a reset if watchdog timeout. There are three watchdog timers supported on server. They are FRB2 watchdog timer which is initiated immediately when server is ready to boot, POST watchdog timer which is enabled by BIOS to monitor the whole POST process, and OS Load watchdog timer which is enabled by BIOS to monitor the OS booting process. The watchdog timer implementation detail is shown in the table below. Once one of the watchdog timers generate timeout events, BMC will reset the server.

Watchdog Timer	Implementation	Timeout Default
FRB2 Watchdog Timer	Used for BIOS code loading check. After system boot, the FRB2 watchdog timer is started automatically by BMC. If BIOS code loading failed, FRB2 WDT bites. BMC switchover the BIOS flash then sends a reset to system.	30 seconds.
POST Watchdog Timer	Used for BIOS code execution check. If BIOS code loading finished, BIOS sends BMC command to stop FRB2 and then start POST watchdog timer. If BIOS code execution cannot pass specified checkpoints, POST WDT bites. BMC sends a reset to system.	The timeout value is related to BIOS settings. They are 3 minutes, 4 minutes, 5 minutes and 6 minutes.
OS Load Watchdog Timer	Used for OS loading check.	The timeout value is related to BIOS settings.

	If BIOS code execution done, BIOS sends commands to BMC to stop the POST WDT and starts the OS load WDT. OS Load WDT checks if everything works well in OS stage. After OS boot, BIOS doesn't have right to feed watchdog timer to prevent watchdog timeout happening. User needs to write software in OS to keep feeding watchdog timer to prevent it timeout. The application is executed on payload.	They are 5 minutes, 10 minutes, 15 minutes and 20 minutes.

Table 35 BMC Watchdog Timer Implementation

- IPMI reset command: in IPMI specification, it provides a command for payload reset. When user orders payload reset command, it will cause a reset to the server.

The reset types the BMC initiated listed as below table.

Reboot Type Number	Reboot Type Name	Description	BMC action
0x00	No Reset Action	Default	No Action
0x01	WDT FRB2	WDT timer "FRB2" timeout	Reset type : Power-up reset Action : Power off → wait for 6 seconds → Power on
0x02	WDT POST	WDT timer "POST" timeout	Reset type : Power-up reset Action : Power off → wait for 6 seconds → Power on
0x03	WDT OS Load	WDT timer "OS Load" timeout	Reset type : Power-up reset Action : Press Reset Button

0x07	OS Reset	OS sends reset by receiving instruction. E.G. “Ctrl + Alt + Del”	Reset type : depends on the command type Action : BMC just record the event into log. No action.
0x08	IPMI Command	Reset is triggered by receiving IPMI command	Reset type : depends on the command type Action : Send reset depends the command type

Table 36 Reset Type of BMC initiated Reset

If BMC updates the firmware code, BMC will send a reset to reset itself. BMC has an internal flash with 512KB for storing BMC execution code. The code update can be done by software command. Once the updating finished, BMC will issue an internal reset automatically to have the new code reloaded and take effect.

5.3.4 Reset implementation

Reset implementation is divided into BMC initiated and PCH initiated resets. Detailed implementation for the resets are listed as below.

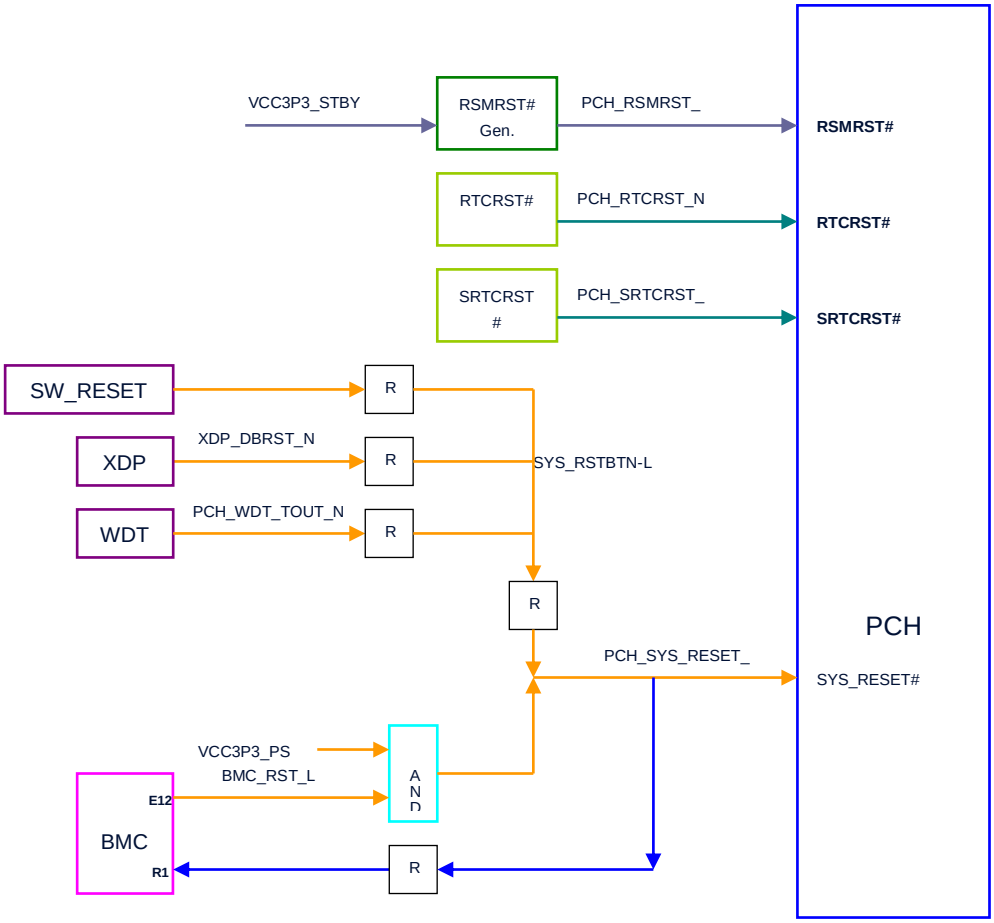


Figure 35 BMC initiated Reset Logic Diagram

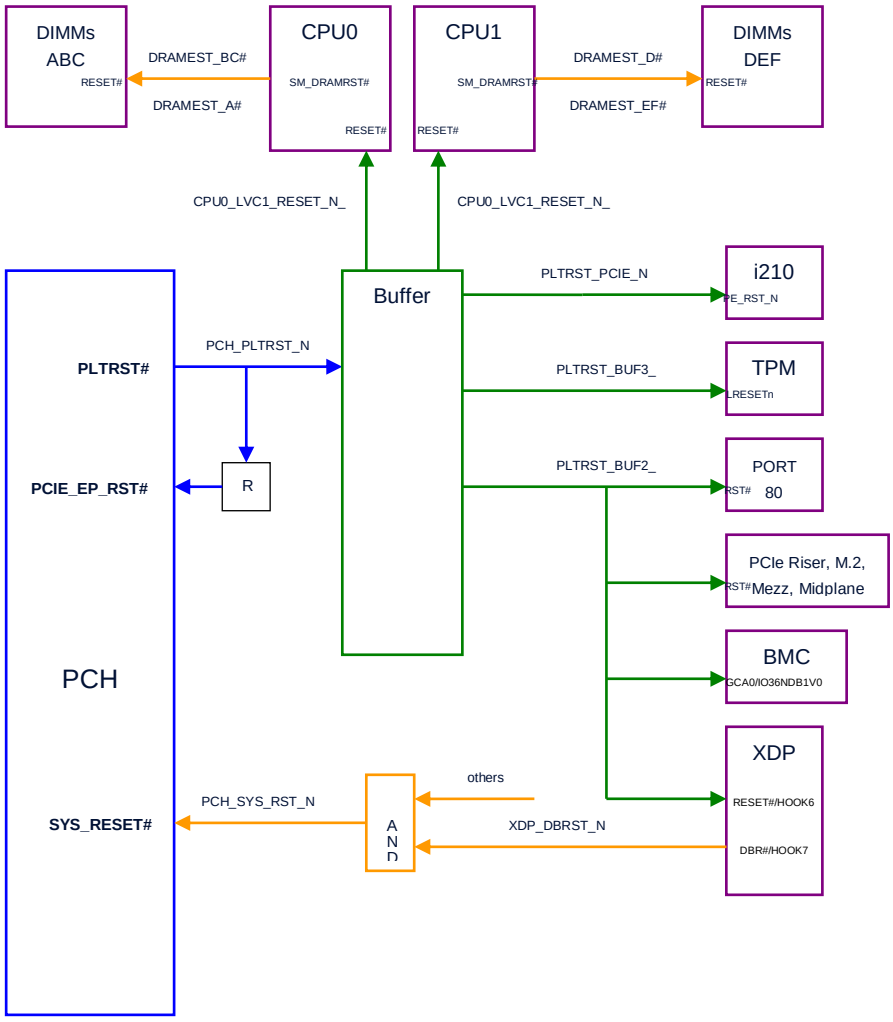


Figure 36 PCH initiated Reset Logic Diagram

5.4 System Management

On server sled, there is a BMC controller ASPEED AST2500 handling the system management including the hardware monitoring, alarming, and remote management.

The system management block diagram is shown as below.

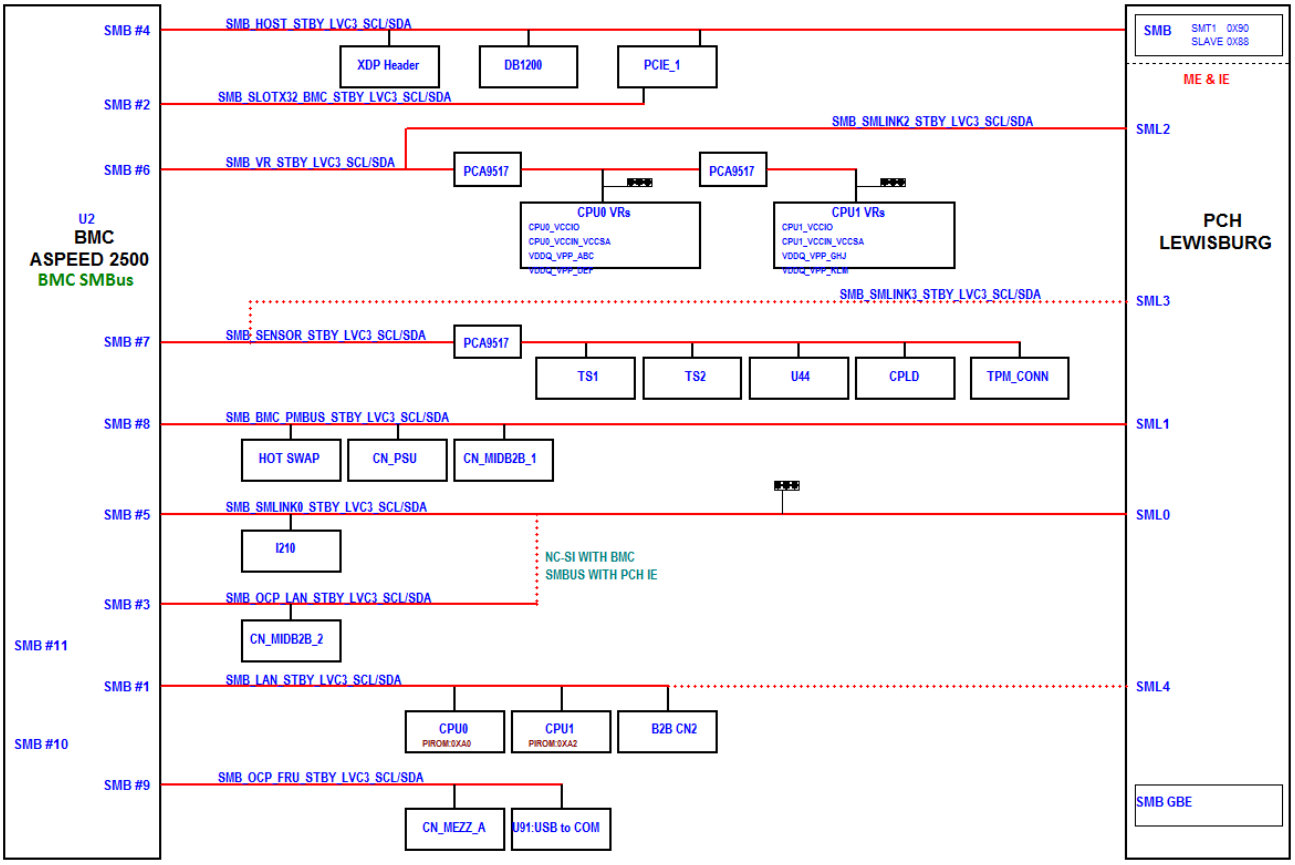


Figure 37 BMC I²C Interface

All the I2C devices I/O address are listed below.

The I² C address	
0x18h	Intel i210-IT
0xA6h	FRU EEPROM
0xAAh	SEL EEPROM
0x40h	CPU0 VR
0x5Ah	CPU1 VR
0XD0h	TPM
0x56h	Mezzanine Conn1
0x58h	Mezzanine Conn2

0x90h	Midplane Conn1
0x92h	Midplane Conn2
0x98h	LM73 Thermal Sensor (Between CPU0 and CPU1)
0X9Ah	LM73 Thermal Sensor (Nearby PCH)

Table 37 BMC SMBus I/O Address Table

5.4.1 Hardware Monitor

The BMC implements hardware monitoring to check the board status and raise an alarm if malfunction. Hardware monitor shall support the following features:

• Voltage monitoring for all payload voltages

• Temperature sensors: the monitoring of CPUs and chipset utilizing built in on-die sensors + additional temperature sensor located on the coolest area of PCB

• Emergency power shut down in case THERMTRIP# asserted

• PROCHOT# for thermal throttling

The BMC hardware monitor devices are shown as below.

1. Voltage monitoring: BMC itself supports 14 analog voltage monitoring. Refer to the section 2.1.6 for the monitored voltage planes list.

2. Thermal sensors: BMC gets thermal data from two LM73-0 thermal sensors. The LM73-0 has +/- 2°C accuracy. Refer to section 2.1.6 for the thermal sensor detail.

3. PROCHOT#

BMC detects PROCHOT# status.

The PROCHOT# goes active when the processor die temperature reaches the maximum safe operating temperature. The assertion of PROCHOT# will enable the thermal throttling to reduce the CPU cores adaptively to cool the CPU down.

4. THERMTRIP#

BMC detects THERMTRIP# status.

The processor asserts the THERMTRIP# if the processor die temperature reaches an extreme high temperature approximately 130°C. To prevent the processor from catastrophic damage, the THERMTRIP# will make the PCH drive the SLP_S3# to turn off the payload main power.

BMC controller stores the event log to SEL EEPROM, then turn off the 12V DC IN. After 6 seconds, BMC check if payload fault condition is still active. If it is still active, BMC waits until fault condition disappear. Then BMC turn on the 12V DC IN and waits WOL event to wake up the payload system.

5.4.2 Power control via IPMI over LAN

Power on command via IPMI over LAN

The server Ethernet supports the IPMI over LAN. User can send IPMI power on command to BMC via the RJ45 port. When BMC receives power on command, it will send PWRBTN signal to PCH to power on the server.

Power off command via IPMI over LAN

There are two ways to power off the server.

- OS initiated power off command
- Power off command via IPMI over LAN

When BMC receives power off command, it will send PWRBTN signal to PCH. There is a Linux ACPI daemon run on Server. When ACPI daemon receives PWRBTN event, OS will start the shutdown process to power off the system.

5.4.3 Serial over LAN

Serial over LAN is a mechanism to redirect the traffic on serial console to the network.

The server supports serial over LAN through the PCH console port UART0. Server has a multiplexer to choose the UART0 routing.

5.4.4 FRU

FRU means the Field Replaceable Unit. It’s an EEPROM located on the BMC I2C bus. The FRU saves the manufacturer, product name, board revision, part number, and board serial number. More information refers to the S/W specification.

5.4.5 SEL

The BMC implements the local System Event Log (SEL) in its non-volatile memory, i.e., an EEPROM. The EEPROM is located on the BMC I2C bus. Local SEL is useful for maintenance purposes and provides OpenHPI access to events. More Information refers to the S/W specification.

5.5 Error Handling

The Skylake processor incorporates the RASM features in accompany with the 3rd party BMC to support the error handling. The error types and the handling schemes are described below.

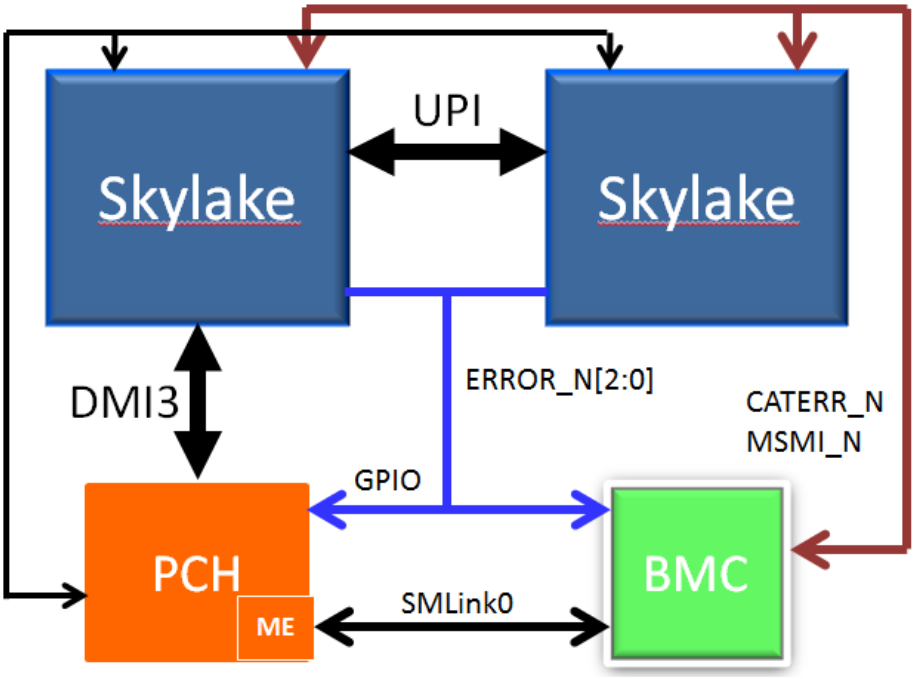


Figure 38 Error Handling Signaling Path

Two Error Reporting domains :

(1) MCA domain : Errors are originated from the CPU core and uncore(CHA, IMC, UPI, Ubox, PCU)

(2) AER domain : Errors are originated from the IIO(PCIe logic portion), PCIe, and DMI3

MCA domain

Errors in the MCA domain are reported via CATERR_N.

CATERR_N assertion indicates the system experienced a catastrophic error and cannot continue to operate. The processor will set this for unrecoverable machine check errors and other internal unrecoverable errors.

The BMC should ignore any signal from CATERR_N until PWRGOOD is asserted.

If eMCA2 is enabled, the CATERR_N signaling is replaced by the MSMI_N. The MSMI_N is used to support the enhanced SMM feature and communicates similar information as CATERR_N for an eMCA event.

No matter the eMCA2 is enabled or not, the BMC should look at both the CATERR_N and MSMI_N to capture all fatal errors.

AER domain

Errors in the AER domain are reported via ERROR_N[2..0].

ERROR_N[0]: Hardware correctable errors (no OS or FW action necessary)

ERROR_N[1]: Non-fatal error (OS or FW action required to contain and recover)

ERROR_N[2]: Fatal error (system reset likely required to recover)

The ERROR_N[2..0] signals indicate various errors from the processor I/O and indicate correctable, uncorrectable non-fatal, and fatal errors.

eMCA2 brief introduction

Besides the legacy MCA, the Skylake supports the enhanced MCA Gen 2 (eMCA2) mode. The eMCA2 is targeted towards providing an infrastructure to allow the errors to be optionally delivered to platform firmware. This means the errors signals to be delivered to the platform firmware ahead of the higher level OS or VMM software. The platform firmware can deliver richer error logs to higher level software at the time of signaling. This improves the diagnostic quality.

When eMCA2 is enabled, the CATERR_N will not be driven. The processor will drive MSMI_N for instead.

Error type logging

Once one of the CATERR_N, MSMI_N, ERROR_N[2..0] is asserted, the BMC knows an error happens in the processor. The BMC uses the PECI Proxy service of the ME to query the error types. PECI doesn’t route to the BMC. The BMC accesses the ME inside the PCH via SMLink0 to get the error type information.

6 Environmental

Open Rack Server is intended for datacenter use and shall meet the following environmental requirements.

- Ambient operating temperature range : -5°C ~ 45°C
- Operating and Storage relative humidity : 10% ~ 90% (non-condensing)
- Storage temperature range : -40°C ~ 70°C
- Compliant with EN 300 019-1-1 [20] Class 1.2 (weather protected, not temperature-controlled storage) environmental conditions
- Comply with ETSI EN300 019-1-3 Class 3.1 thermal and humidity climatogram and seismic conditions
- Compliant with EN300386 (v1.6.1) for EMC, as specified for the telecommunication centres installation environment
- Emission criteria to be compliant with CFR 47, FCC 15, CISPR 22 Class A and CISPR 24
- Compliant with EN60950-1: 2006 + A2:2013 and IEC 60950-1 for safety
- Transportation package to fulfill ETSI EN 300 019-1-2 v.2.2.1 class 2.2 Careful transportation requirements
- All components used in the unit shall fulfil the EU RoHS directive 2011/65/EU Article 7b (EN 50581 (2012))
- Compliant with TEC/EMI/TEL-001/01/FEB-09 and TEC/IR/SWN-2MB/07/MAR-10 for radiated immunity
- All materials including components, wires, and cables shall meet the ANSI T1.307-2007 and the requirements specified in GR-63-CORE chapter 4.2.3
- Meet the fire spread criteria specified in GR-63-CORE chapter 4.2.2.2 Shelf-Level Fire-Resistance Criteria
- Meet the GR-63-CORE fire resistance criterias defined in the chapter 4.2

7 Labels and Markings

The board shall include the following labels on the component side. The labels shall not be placed in a way, which may cause them to disrupt the functionality or the airflow path of the board.

Description	Type	Barcode Required?
MAC address	Adhesive label	Yes
Vendor P/N, S/N, REV	Adhesive label	Yes
Vendor Logo, Name, and Country of Origin	Silk Screen	No
PCB vendor Logo, Name	Silk Screen	No
Purchaser P/N	Adhesive label	Yes
Date Code (Industry Standard : WEEK/YEAR)	Adhesive label	Yes
RoHS compliance	Silk Screen	No
WEEE symbol:  The symbol means the board will be taken back by the manufacturer for recycle at the end of its useful life.	Silk Screen	No
CE Marking	Silk Screen	No
UL Marking	Silk Screen	No

Table 38 Labels and Markings