OPEN
Compute Project

Facebook Multi-Node Server Platform: Yosemite V2 Design Specification

V 0.4

Authors:

Yan Zhao, Hardware Engineer, Facebook

Jon Ehlen, Mechanical Engineer, Facebook

Jarrod Clow, Thermal Engineer, Facebook

Sai Dasari, Software Engineer, Facebook

Copyrights and Trademarks

1 Scope

This specification describes the design of the Yosemite V2 Platform that hosts four One Socket (1S) servers, or two sets of 1S server/device card pairs.

2 Contents

3 Overview

This document describes Facebook's multi-node server platform (code name: Yosemite V2) and the design requirements to integrate the platform into Open Rack V2.

The Yosemite V2 Platform is a next generation multi-node server platform that hosts four Open Compute Platform (OCP) compliant One Socket (1S) server cards, or two sets of 1S server card and device card pairs in a sled that can be plugged into an OCP vCubby chassis, which is a new 4OU form factor design to easily service and accommodate thermal challenges with higher power 1S servers and device cards.

There is a horizontally installed baseboard inside the chassis to hold 1S server cards or device cards vertically. With this new architecture, we could leverage 1S server cards that use a higher-power, higher-performance System On a Chip (SoC). The Yosemite V2 Platform fully uses space inside the sled to add additional performance and functionality. The modular design makes this platform flexible so that it is possible to adapt SoC-agnostic 1S servers from different vendors.

On the network side, the Yosemite V2 Platform has various solutions to provide network access to the 1S servers. All 1S servers have their own independent network interface, virtually. To simplify cabling, only a single network cable is allowed to connect a Yosemite V2 Platform to a top-of-rack (TOR) switch.

1S servers with integrated 10GBase-KR Ethernet controllers can use an OCP 2.0 compliant KR mezzanine card to provide 4x 10G links to a TOR switch through a single QSFP+ cable.

For the 1S servers that do not have or do not use integrated network controllers, a PCI Express (PCIe) based multi-host network mezzanine card can be used on the Yosemite V2 Platform to provide 40Gbps, 50Gbps, or 100G connectivity for the whole platform on the line side to a TOR switch. On the host side, every 1S server connects to the multi-host network mezzanine card through a x4 PCIe Gen3 link and sees the mezzanine card as if it were its own network interface card.

A device card in 1S server form factor can be used on the Yosemite V2 Platform. A device carrier card in 1S server form factor is designed to host a standard X16 full-height half-length PCIe card with maximum power of 75W.

A Baseboard Management Controller (BMC) on a baseboard is used to manage all 1S servers, device cards and the Yosemite V2 Platform. The BMC shall support both in-band and out-of-band (OOB) management so that the BMC can be accessed from 1S servers on the same Yosemite V2 platform or from an external server on the network.

All 1S servers get single 12V power from the Yosemite V2 Platform, while the Yosemite V2 Platform gets 12V power from the rack. There is a 12V power switch in front of every 1S server slot under the BMC's control. Therefore, the BMC can do full AC power cycling to 1S servers and

device cards when needed. The BMC monitors the health status of the Yosemite V2 Platform and all 1S servers and device cards (e.g., power, voltage, current, temperature, fan speed, etc.) and will take action when failures occur.

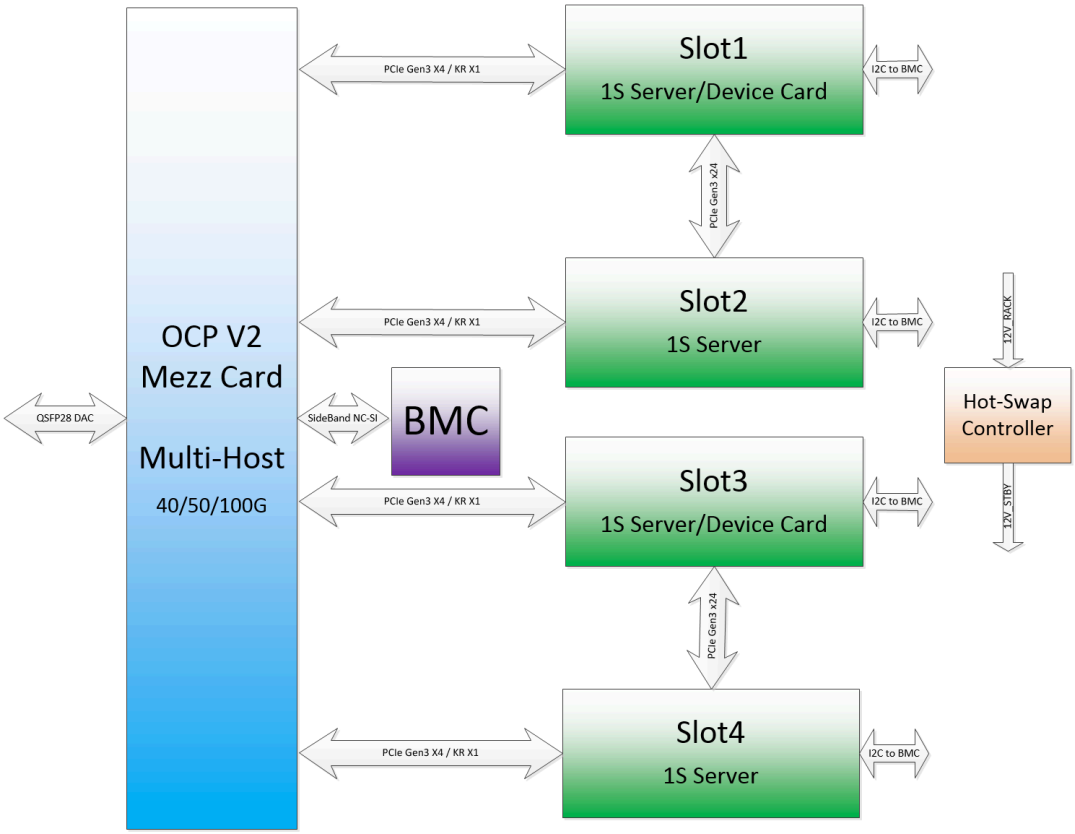
The Yosemite V2 Platform is compatible with the OCP Open Rack V2 specification. Please refer to the corresponding OCP Open Rack V2 documentation for more details about the rack. The Yosemite V2 Platform in a vCubby chassis can be safely inserted or removed to/from an Open Rack. You can find more details in the mechanical section of this document.

4 License

5 Yosemite V2 Platform Features

5.1 Platform Block Diagram

Figures 5-1 and 5-2 illustrate the functional block diagram and design details of the Yosemite V2 Platform.



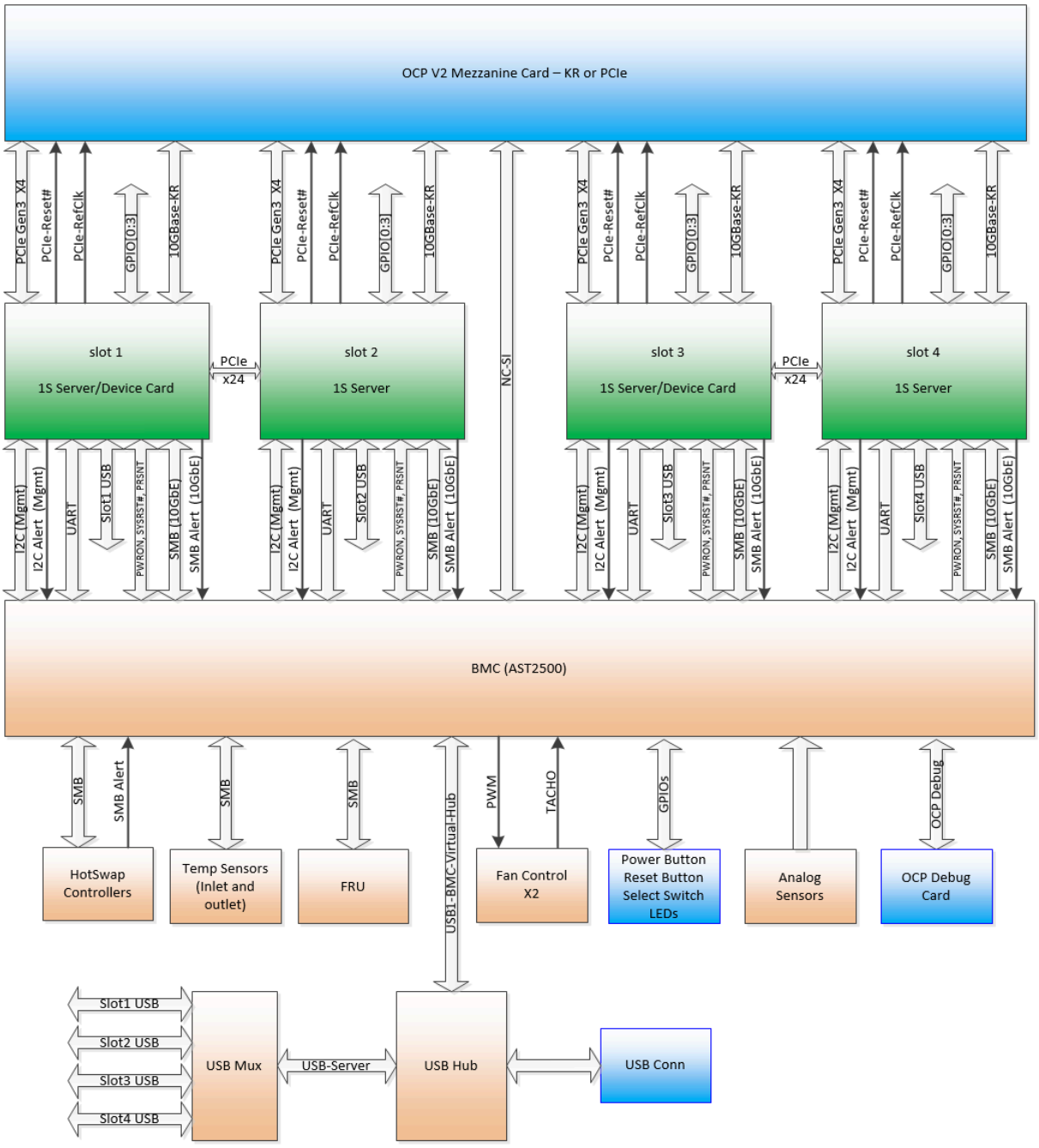


Figure 5-1: Yosemite V2 Platform Block Diagram

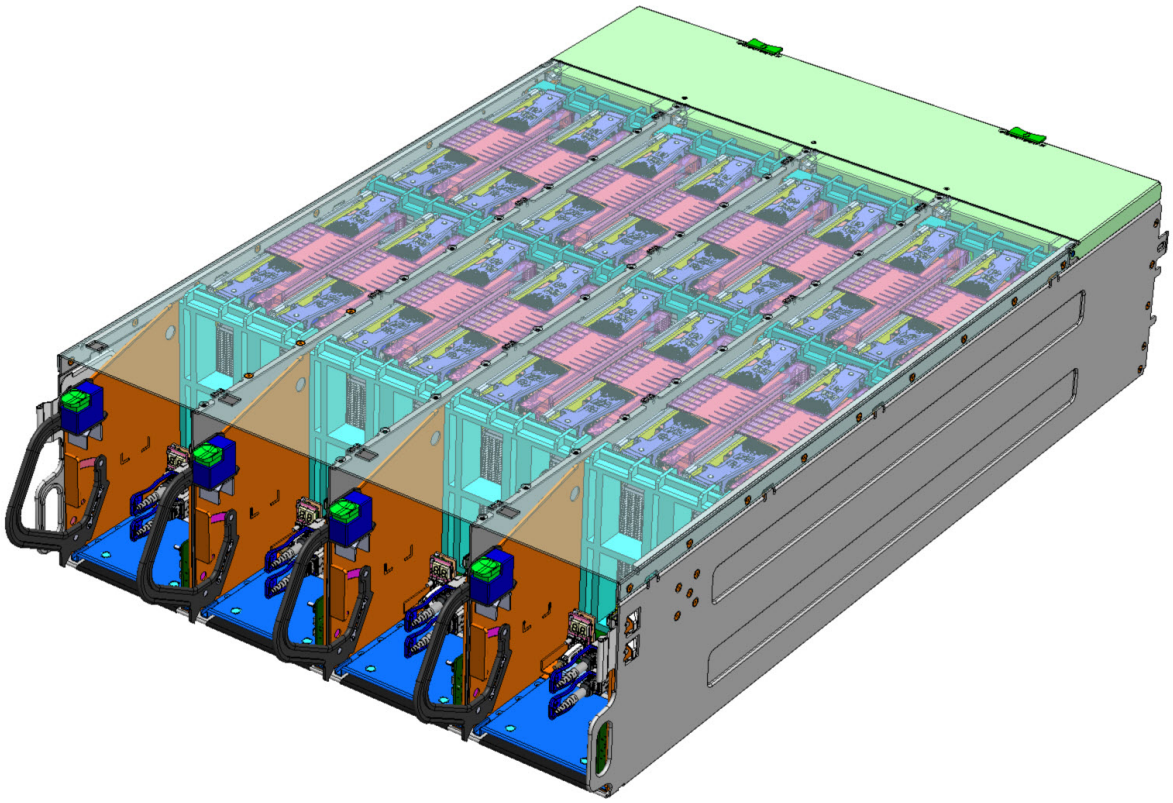


Figure 5-2: Yosemite V2 Platform

5.2 Yosemite V2 Platform Baseboard

The Yosemite V2 Platform could have two configurations: first, it can host four 1S server cards in all four slots as in the Yosemite Platform; the other option is to have two 1S servers in slot 2 and slot 4 with device cards in slot 1 and slot 3, which are connected to a 1S server in slot 2 and slot 4 through 6 x4 PCIe Gen3 links, as shown in the block diagram above, respectively. The device card is a PCIe device card to the corresponding 1S server and it can be a flash card, a GPU card, a FPGA card, and so on.

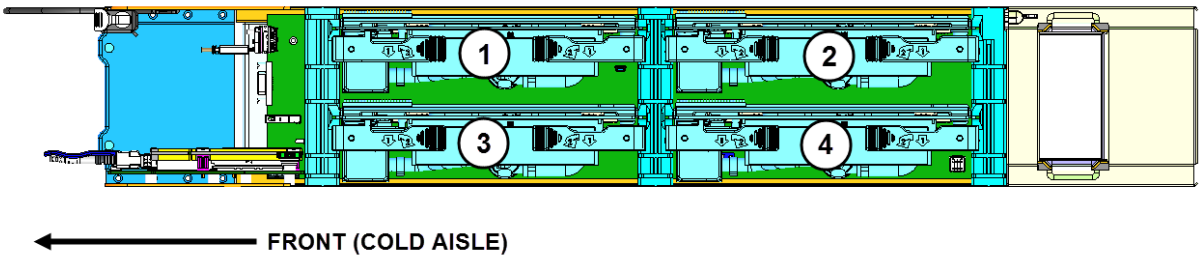


Figure 5-2: Yosemite V2 Sled Slot Positions

As there are two possible configurations for the Yosemite V2 Platform, to simplify provisioning we want the system to configure itself automatically. The BMC needs to collect system information and determine the configuration during system initialization. All the 1S servers should disable PCIe ports by default and inquire BMC to get the system level configuration to set up its PCIe ports properly. This way we can avoid collisions, such as two 1S servers trying to use each other as a device and ending up with an invalid state.

The Yosemite V2 Platform contains a baseboard to hold all connectors and common infrastructure pieces, including primary and extension connectors to host the 1S server or device card, OCP V2 Mezzanine card through an adapter card, a 12.5V inlet power connector (from the vCubby chassis), a BMC section, fan connectors, a hot-swap controller, and a front panel. Both primary and extension connectors defined in the 1S server specification will be used on the Yosemite V2 Platform.

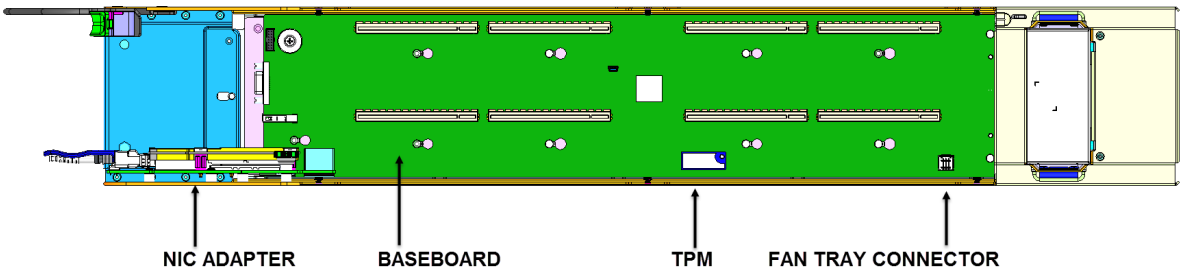


Figure 5-2A: Baseboard

The baseboard is installed horizontally on the bottom of a vCubby chassis. OCP-compliant 1S server cards with a height of 110mm or 160mm can be installed vertically to the baseboard with a proper holder.

The baseboard shall be implemented as a low-maintenance, robust platform to reduce the need of service. A BMC (ASPEED AST2500) is the main control unit on the baseboard. The Yosemite V2 Platform uses an adapter card at the front of the sled as a carrier board for OCP 2.0 mezzanine cards. The OCP 2.0 mezzanine connectors on the adapter cards have been carefully designed in a hybrid way to take a PCIe-based multi-host OCP V2 mezzanine 40GbE/50GbE/100GbE card, or a 4x10Gb capable KR retimer Mezzanine card that connects to the 1S server's built-in 10GBase-KR Network Interface Controller, as the Ethernet interface to the external world. Either way, the Network Interface Card (NIC) will be used as a shared NIC, so that a BMC can be accessed via the OOB of the NIC, Network Controller Sideband Interface (NC-SI), or System Management Bus (SMBus).

By sampling the sensors on the Yosemite V2 Platform periodically, the BMC continuously monitors the system's health status from function, power and thermal perspectives. The BMC shall implement sophisticated algorithms to control system environment accordingly.

The BMC is connected to a hot-swap controller through an Inter-Integrated Circuit (I^2C) bus so that it can get system-wide power consumption and maintain a healthy status. The BMC also controls 12.5V power to each 1S server/device card slot. It is possible to let the BMC completely shut down 12.5V to a 1S server/device card when the server/device needs an AC-level cold reboot.

A power button, a reset button, a USB connector and an OCP debug card header are provided on the front panel of the baseboard; they all belong to the current active server. The user could turn a rotary switch to select a 1S server as the current active server.

The USB connections from all 1S servers are connected to the BMC's virtual hub port through a USB multiplexer so that a user could upgrade the BMC firmware via a USB interface from a 1S server. This method is much faster than going through the OOB.

There are two fan connectors and one 12.5V inlet power connector on the backside of the baseboard to provide cooling and power to the system.

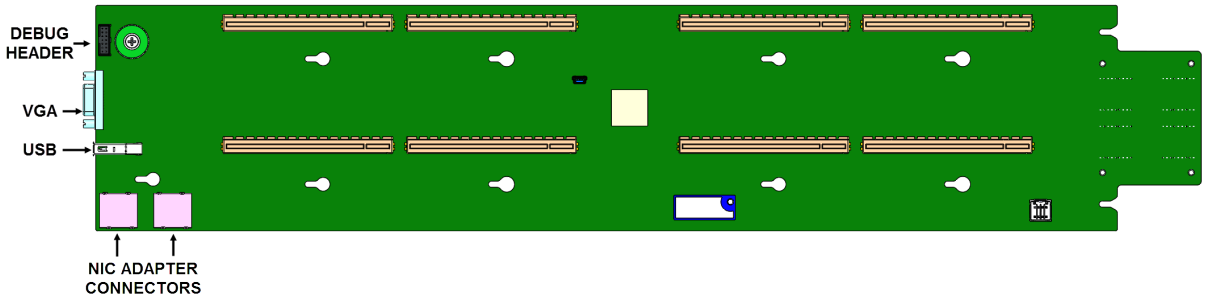


Figure 5-2B: Baseboard I/O

To improve serviceability, slot-level hot-service is required while other 1S server and/or device cards inside the same sled shall keep their normal operation without interruptions. Please note none of the 1S server and/or device card supports hot swap when power still applies. Hot-service replaces a failed 1S server or a pair of 1S server and device cards after the hosting slot(s) are completely powered off, without any impact to the rest of the platform. The Yosemite V2 Platform is designed to support continuous 12.5V power when the sled is being pulled out of vCubby chassis during hot-service.

Hot-service starts with a request to the BMC to replace a failed unit inside the Yosemite V2 chassis. If the platform is configured to host four 1S servers, the BMC shall only turn off 12.5V power to the affected slot. However, if the request is for a Yosemite platform that hosts two 1S server/device pairs, then the BMC shall shut off the 12.5V power to both slots that host this pair to avoid an unexpected hot-plug event.

Once the slot(s) power down is done and confirmed by the BMC, a user can then pull out the Yosemite V2 sled and check the failed unit to make sure a blue power LED on this unit is indeed off. Now the user can replace the failed unit with a new one, and then push the sled back to its bay. After that, the user can turn 12.5V to the affected slot(s) back on through the BMC or power button on the front panel, which should bring the affected system back to service.

During the entire hot-service process, the BMC shall monitor the thermal condition closely. The BMC shall speed up the fan accordingly, and may throttle other unaffected live 1S servers or device cards when necessary.

5.3 Yosemite V2 Platform Power Delivery

Figure 5-3 illustrates the power delivery block diagram of the Yosemite V2 Platform.

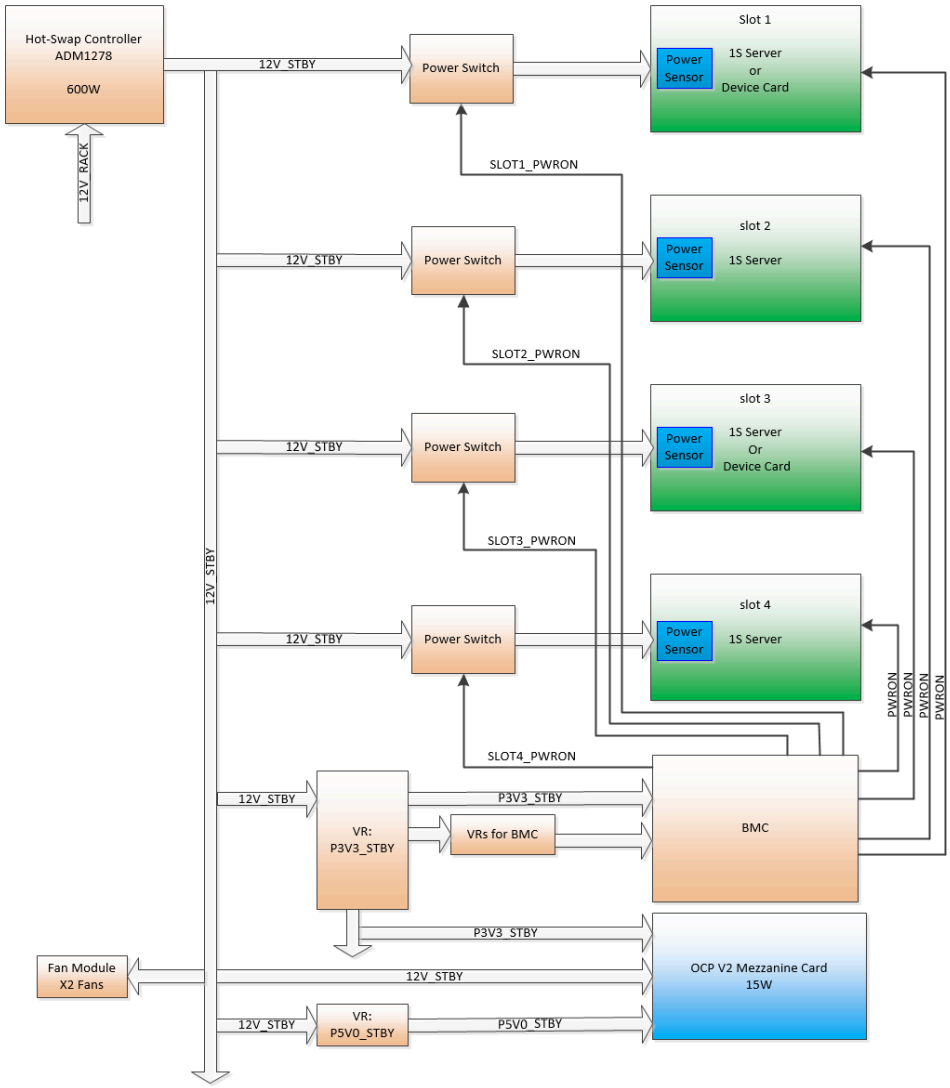


Figure 5-3: Yosemite V2 Platform Power Delivery Block Diagram

The Yosemite V2 Platform is designed to host four 1S servers, or two 1S server and device card pairs. With two connectors supported, each 1S server or device card could consume up to 192W total power. Considering power loss and the baseboard's power consumption, the Yosemite V2 Platform could support a maximum of 600W total power.

The Yosemite V2 Platform system gets 12.5V power from the vCubby chassis through a Powerbar PCB, which transmits power from the rack power bus bar cables to the baseboard through a set of sliding brushes. These brushes continuously transmit power during the sled slide in/out for the service event.

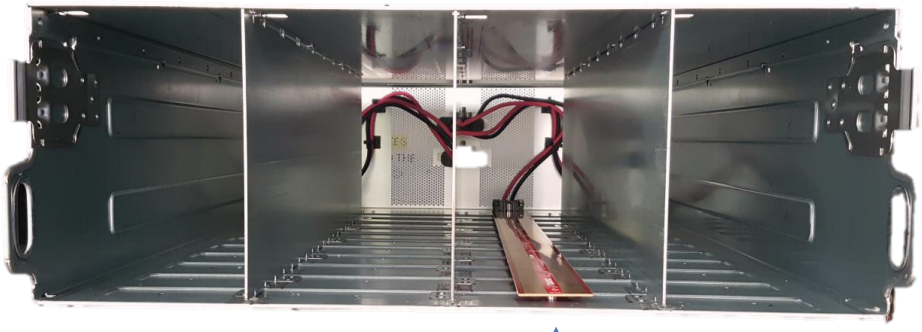


Figure 5-3A: Yosemite V2 Powerbar PCB

Figure 5-3B: Yosemite V2 Power Brushes

A hot-swap controller (ADM1278) is used to protect rack power from a voltage drop due to a transient current draw when the Yosemite V2 Platform is inserted or removed from a live rack. The hot-swap controller shall set the power limit accordingly and shut the platform's 12.5V power down immediately whenever over-current, over-voltage, under-voltage, or over-temperature conditions occur.

The BMC uses standby rails P3V3_STBY and other voltage rails to power BMC circuits and its DDR4 memory. 12.5V_STBY, P5V0_STBY, and P3V3_STBY are provided at the OCP V2 mezzanine card connectors to power an OCP V2 compliant mezzanine card.

Every 1S server shall have a power sensor right at the 12.5V input side of the card. The SoC or the Bridge IC of the 1S server needs to sample the power sensor periodically and calculate a 1 second average from those samples. The BMC shall be able to access this power sensor via the Bridge IC on the 1S server. As a debugging feature, the Bridge IC shall be able to sample the power sensor as fast as 10ms.

The device card shall also implement a power sensor right at the 12.5V input side of the card. However, as there is no intelligent controller on a device card, the BMC needs to monitor these power sensors in the same manner as 1S server.

By default, these 12.5V_STBY power switches to 1S servers and device cards should always be on unless the BMC turns them off on purpose. Thus, the integrated Ethernet controllers on the 1S server always get standby power to keep side-band traffic alive even when 1S servers are in standby mode.

There is a power switch on the 12.5V_STBY going to every slot. The BMC could power cycle a 1S server and/or a device card by toggling the corresponding 12.5V_STBY power switch. This is useful when the 1S server/device card needs a cold reset or AC power cycling. Tri-state buffers on GPIOs between 1S servers/device cards and the Yosemite V2 Platform are required to avoid leakage from the Yosemite V2 Platform to 1S servers when they are in power-off or stand-by state.

The BMC has a dedicated power-on signal for each 1S server. Depending on the power policy, the BMC enables power to 1S servers upon request. The BMC shall drive power-on signals as a power button function as defined in the Advanced Configuration and Power Interface (ACPI).

5.4 SMBus Block Diagram

Figure 5-4 illustrates the Yosemite V2 Platform SMBus block diagram.

Each 1S server has one SMBus from its integrated NIC connected to the BMC as a sideband if a KR mezzanine card is being used as the network interface for the platform. A Bridge IC is connected to the BMC on each 1S server through a dedicated I^2C bus as the management interface between a 1S server and the BMC. When a PCIe based multi-host network interface card is being used, the BMC shall use NC-SI as the sideband as it is much faster than SMBus.

A device card only uses the management SMBus. It will present all its I2C devices, such as thermal sensor, I2C mux, and FRU EEPROM, on this bus. BMC can access these I2C devices directly.

Depending on the Mezzanine card type, the BMC could connect to a Management Data Clock (MDC)/Management Data Input/Output (MDIO) or I^2C to LAN_SMB port on the Mezzanine card to configure the Mezzanine card or use the NIC's SMBus as OOB. A dedicated SMBus is also connected to MEZZ_SMB of the mezzanine card as a management path, so that the BMC can read Field Replaceable Unit (FRU) data from the mezzanine card or perform other management tasks.

The BMC can access thermal sensors, the hot-swap controller and the FRU via a separate SMBus, as shown in Figure 5-4 below.

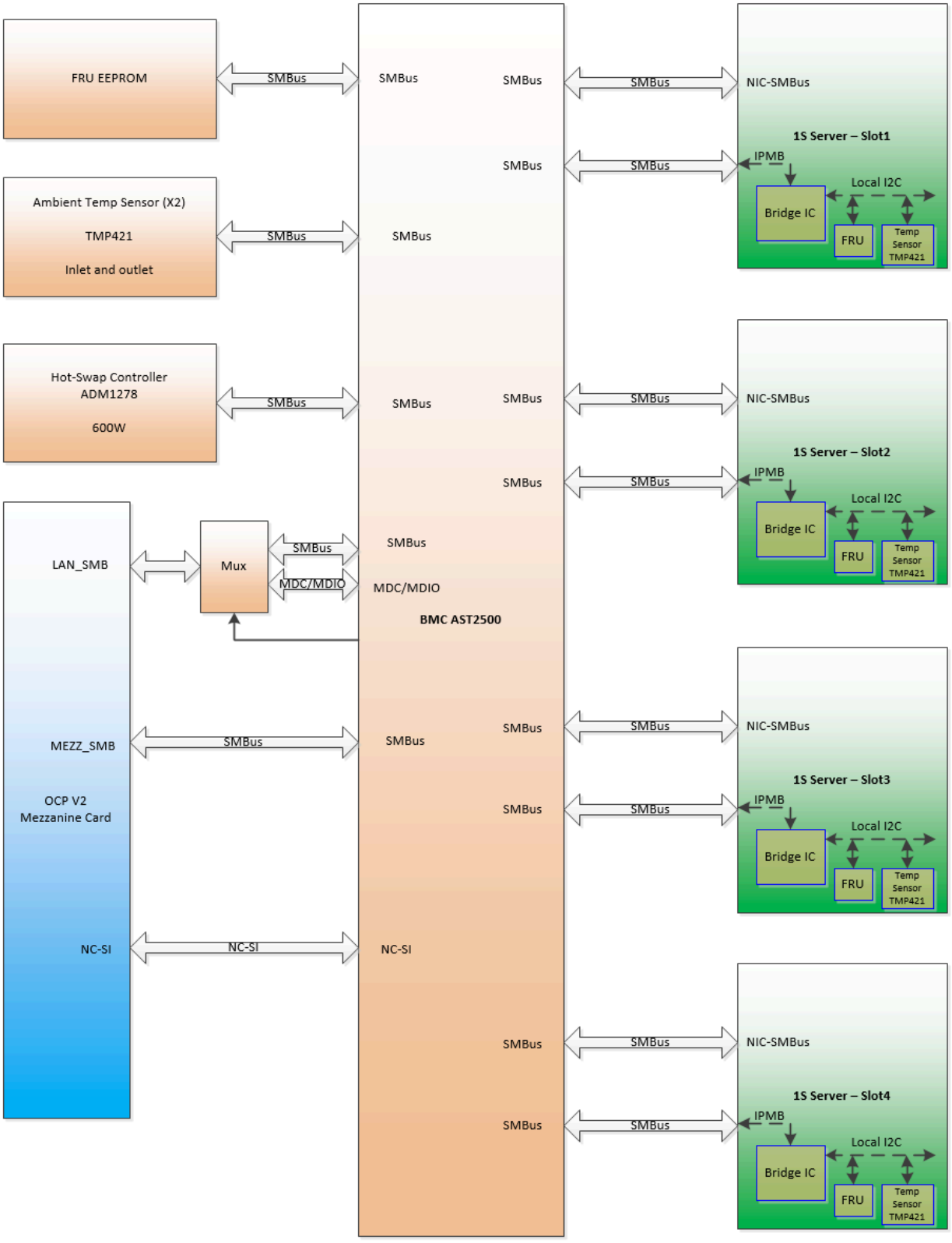


Figure 5-4: Yosemite V2 Platform SMBus Block Diagram

5.5 1S Server

5.5.1 Overview

The Yosemite V2 Platform has four slots that can host four OCP compliant 1S servers, or two pairs of OCP compliant 1S servers and device cards.

5.5.2 1S Server Connectors

The 1S server specification defined two x16 PCIe-like connectors, primary connector and extension connectors. The Yosemite V2 Platform uses both primary and extension connectors.

The X16 PCIe connector is defined in the PCIe specification. However, a completely different pinout is defined for 1S server applications.

5.5.3 1S Server Slot Pinout Definition on Yosemite V2 Platform

The Yosemite V2 Platform uses the primary and extension pinout as specified in chapter 5.6 of OCP 1S server spec, which is similar to first generation 1S server pinout definition. However, we have following deviations:

1) KR4 lanes on the extension connector have been converted to a 7th PCIe Gen3 x4 link. Please note the 7th PCIe Gen3 X4 link does not have its own PCIe reference clock and PCIe reset signal, so on the Yosemite V2 Platform this port shares a PCIe reference clock and PCIe reset signal with other ports with necessary buffers.
2) 4 GPIO pins have been repurposed for Yosemite V2 platform specifically.
3) The SATA link has been converted to a PCIe x1 link for VGA support. This link also has no dedicated PCIe reference clock and reset signal, so on the Yosemite V2 Platform this port shares a PCIe reference clock and PCIe reset signal with other ports with necessary buffers.

Table 1: Yosemite V2 1S Server Primary OCP Edge Connector A Pin-Out

Default Pin-Out			
Pin Name	**B Side**	**A Side**	**Pin Name**
P12V	1	1	PRSNT_A#
P12V	2	2	P12V
P12V	3	3	P12V
GND	4	4	GND
I2C_SCL	5	5	SVR_ID0/GPIO0
I2C_DATA	6	6	SVR_ID1/GPIO1
GND	7	7	COM_TX
PWR_BTN#	8	8	COM_RX
USB_P	9	9	SVR_ID2/GPIO2

USB_N	10	10	SVR_ID3/GPIO3
SYS_RESET#	11	11	PCIE0_RESET#
I2C_ALERT#	12	12	GND
GND	13	13	PCIE0_REFCLK_P
GND	14	14	PCIE0_REFCLK_N
PCIE0_TX0_P	15	15	GND
PCIE0_TX0_N	16	16	GND
GND	17	17	PCIE0_RX0_P
GND	18	18	PCIE0_RX0_N
PCIE0_TX1_P	19	19	GND
PCIE0_TX1_N	20	20	GND
GND	21	21	PCIE0_RX1_P
GND	22	22	PCIE0_RX1_N
PCIE0_TX2_P	23	23	GND
PCIE0_TX2_N	24	24	GND
GND	25	25	PCIE0_RX2_P
GND	26	26	PCIE0_RX2_N
PCIE0_TX3_P	27	27	GND
PCIE0_TX3_N	28	28	GND
GND	29	29	PCIE0_RX3_P
GND	30	30	PCIE0_RX3_N
PCIe_VGA_TX_P	31	31	GND
PCIe_VGA_TX_N	32	32	GND
GND	33	33	PCIe_VGA_RX_P
GND	34	34	PCIe_VGA_RX_N
PCIE1_REFCLK_P	35	35	GND
PCIE1_REFCLK_N	36	36	GND
GND	37	37	PCIE2_REFCLK_P
GND	38	38	PCIE2_REFCLK_N
PCIE1_RESET#	39	39	GND
PCIE2_RESET#	40	40	GND
GND	41	41	FAST_THROTTLE_N

GND	42	42	NIC_SMBUS_ALERT#
NIC_SMBUS_SCL	43	43	GND
NIC_SMBUS_SDA	44	44	GND
GND	45	45	KR0_RX_P
GND	46	46	KR0_RX_N
KR0_TX_P	47	47	GND
KR0_TX_N	48	48	GND
GND	49	49	PCIE1_RX0_P
GND	50	50	PCIE1_RX0_N
PCIE1_TX0_P	51	51	GND
PCIE1_TX0_N	52	52	GND
GND	53	53	PCIE1_RX1_P
GND	54	54	PCIE1_RX1_N
PCIE1_TX1_P	55	55	GND
PCIE1_TX1_N	56	56	GND
GND	57	57	PCIE1_RX2_P
GND	58	58	PCIE1_RX2_N
PCIE1_TX2_P	59	59	GND
PCIE1_TX2_N	60	60	GND
GND	61	61	PCIE1_RX3_P
GND	62	62	PCIE1_RX3_N
PCIE1_TX3_P	63	63	GND
PCIE1_TX3_N	64	64	GND
GND	65	65	PCIE2_RX0_P
GND	66	66	PCIE2_RX0_N
PCIE2_TX0_P	67	67	GND
PCIE2_TX0_N	68	68	GND
GND	69	69	PCIE2_RX1_P
GND	70	70	PCIE2_RX1_N
PCIE2_TX1_P	71	71	GND
PCIE2_TX1_N	72	72	GND
GND	73	73	PCIE2_RX2_P
GND	74	74	PCIE2_RX2_N

PCIE2_TX2_P	75	75	GND
PCIE2_TX2_N	76	76	GND
GND	77	77	PCIE2_RX3_P
GND	78	78	PCIE2_RX3_N
PCIE2_TX3_P	79	79	GND
PCIE2_TX3_N	80	80	GND
GND	81	81	P12V
GND	82	82	P12V

Table 2: Yosemite V2 1S Server Extension OCP Edge Connector B Pin-Out

Default Pin-Out			
Pin Name	**B Side**	**A Side**	**Pin Name**
P12V	1	1	PRSNT_B#
P12V	2	2	P12V
P12V	3	3	P12V
GND	4	4	GND
NCSI_TXEN	5	5	NCSI_RCLK
NCSI_TXD0	6	6	NCSI_RXD0
NCSI_TXD1	7	7	NCSI_RXD1
NCSI_CRSDV	8	8	GND
NCSI_RXER	9	9	PCIE4_REFCLK_P
GND	10	10	PCIE4_REFCLK_N
PCIE3_RESET#	11	11	GND
PCIE4_RESET#	12	12	GND
PCIE5_RESET#	13	13	PCIE5_REFCLK_P
GND	14	14	PCIE5_REFCLK_N
PCIE6_TX0_P	15	15	GND
PCIE6_TX0_N	16	16	GND
GND	17	17	PCIE6_RX0_P
GND	18	18	PCIE6_RX0_N
PCIE6_TX1_P	19	19	GND
PCIE6_TX1_N	20	20	GND

GND	21	21	PCIE6_RX1_P
GND	22	22	PCIE6_RX1_N
PCIE6_TX2_P	23	23	GND
PCIE6_TX2_N	24	24	GND
GND	25	25	PCIE6_RX2_P
GND	26	26	PCIE6_RX2_N
PCIE6_TX3_P	27	27	GND
PCIE6_TX3_N	28	28	GND
GND	29	29	PCIE6_RX3_P
GND	30	30	PCIE6_RX3_N
PCIE3_REFCLK_P	31	31	GND
PCIE3_REFCLK_N	32	32	GND
GND	33	33	PCIE3_RX0_P
GND	34	34	PCIE3_RX0_N
PCIE3_TX0_P	35	35	GND
PCIE3_TX0_N	36	36	GND
GND	37	37	PCIE3_RX1_P
GND	38	38	PCIE3_RX1_N
PCIE3_TX1_P	39	39	GND
PCIE3_TX1_N	40	40	GND
GND	41	41	PCIE3_RX2_P
GND	42	42	PCIE3_RX2_N
PCIE3_TX2_P	43	43	GND
PCIE3_TX2_N	44	44	GND
GND	45	45	PCIE3_RX3_P
GND	46	46	PCIE3_RX3_N
PCIE3_TX3_P	47	47	GND
PCIE3_TX3_N	48	48	GND
GND	49	49	PCIE4_RX0_P
GND	50	50	PCIE4_RX0_N
PCIE4_TX0_P	51	51	GND
PCIE4_TX0_N	52	52	GND
GND	53	53	PCIE4_RX1_P

GND	54	54	PCIE4_RX1_N
PCIE4_TX1_P	55	55	GND
PCIE4_TX1_N	56	56	GND
GND	57	57	PCIE4_RX2_P
GND	58	58	PCIE4_RX2_N
PCIE4_TX2_P	59	59	GND
PCIE4_TX2_N	60	60	GND
GND	61	61	PCIE4_RX3_P
GND	62	62	PCIE4_RX3_N
PCIE4_TX3_P	63	63	GND
PCIE4_TX3_N	64	64	GND
GND	65	65	PCIE5_RX0_P
GND	66	66	PCIE5_RX0_N
PCIE5_TX0_P	67	67	GND
PCIE5_TX0_N	68	68	GND
GND	69	69	PCIE5_RX1_P
GND	70	70	PCIE5_RX1_N
PCIE5_TX1_P	71	71	GND
PCIE5_TX1_N	72	72	GND
GND	73	73	PCIE5_RX2_P
GND	74	74	PCIE5_RX2_N
PCIE5_TX2_P	75	75	GND
PCIE5_TX2_N	76	76	GND
GND	77	77	PCIE5_RX3_P
GND	78	78	PCIE5_RX3_N
PCIE5_TX3_P	79	79	GND
PCIE5_TX3_N	80	80	GND
GND	81	81	P12V
POWER_FAIL_N	82	82	P12V

Table 3:Detailed Pin Definitions

Pin	Direction	Required/ Configurable	Pin Definition
P12V	Input	Required	12VAUX power from platform
I2C_SCL	Input/Output	Required	I^2C clock signal. I^2C is the primary sideband interface for server management functionality. 3.3VAUX signal. Pull-up is provided on the platform.
I2C_SDA	Input/Output	Required	I^2C data signal. I^2C is the primary sideband interface for server management functionality. 3.3VAUX signal. Pull-up is provided on the platform.
I2C_ALERT#	Output	Required	I^2C alert signal. Alerts the BMC that an event has occurred that needs to be processed. 3.3VAUX signal. Pull-up is provided on the platform.
NIC_SMBUS_SCL	Input/Output	Required	Dedicated SMBus clock signal for network sideband traffic between the BMC and the NIC. 3.3VAUX signal. Pull-up is provided on the platform.
NIC_SMBUS_SDA	Input/Output	Required	Dedicated SMBus data signal for network sideband traffic between the BMC and the NIC. 3.3VAUX signal. Pull-up is provided on the platform.
NIC_SMBUS_ALERT#	Output	Required	Dedicated SMBus alert signal for network sideband traffic between the BMC and the NIC. 3.3VAUX signal. Pull-up is provided on the platform.
NCSI_RCLK	Input	Required	NC-SI reference clock for NIC
NCSI_CRSDV	Output	Required	Carrier Sense/Receive Data Valid from NIC to BMC.
NCSI_RXER	Output	Required	Receive error from NIC to BMC
NCSI_TXEN	Input	Required	Transmit enable from BMC to NIC
NCSI_RXD[0:1]	Output	Required	Receive data from NIC to BMC
NCSI_TXD[0:1]	Input	Required	Transmit data from BMC to NIC
PWR_BTN#	Input	Required	Power on signal. When driven low, it indicates that the server will begin its power-on sequence. 3.3VAUX signal. Pull-up is provided on the platform. If PWR_BTN# is held low for greater than 4 seconds, then this indicates a soft (graceful) power off. Otherwise, a hard shutdown is initiated.

SYS_RESET#	Input	Required	System reset signal. When driven low, it indicates that the server will begin its warm reboot process. 3.3VAUX signal. Pull-up is provided on the platform.
PRSNT_A#	Output	Required	Present signal. This is pulled low on the card to indicate that a card is installed. 3.3VAUX signal. Pull-up is provided on the platform.
PRSNT_B#	Output	Required	Extension edge connector Present signal. This is pulled low on the card to indicate that a card is installed. 3.3VAUX signal. Pull-up is provided on the platform.
COM_TX	Output	Required	Serial transmit signal. Data is sent from the 1S Server module to the BMC. 3.3VAUX signal. On BMC side this signal shall be pulled up as logic 1.
COM_RX	Input	Required	Serial receive signal. Data is sent from the BMC to the 1S Server module. 3.3VAUX signal. On 1S server side this signal shall be pulled up as logic 1.
SVR_ID0/1/2/3 (GPIO0/1/2/3)	Input/Output	Required	GPIO0: is defined as CARD_TYPE which is an output from 1S server or device card. 1S server shall pull this signal low while device card shall pull this signal high to 1.2V for a PCIe device carrier card or 1.7V for a customized PCIe card in 1S server form factor. This signal is a quick way for BMC to identify card types on the Yosemite V2 platform. GPIO1: is defined as POWER_EN, 3.3VAUX signal, active high. This signal is an output on a 1S server and the 1S server shall only drive this signal high when the 1S server is powering up. On a device card this signal is an input with a pull-down on the device card. The device card shall use this signal to turn on payload power on the device card. GPIO2: is defined as EJECTOR_LATCH_DETECT_N, 3.3VAUX signal, active low. This signal is driven by ejector detection circuits on 1S server and device card. The purpose of this signal is to prevent surprise insertion/removal and protect 1S server and device card. This signal is driven low when the ejector is closed, high when the ejector is open. Yosemite V2 platform uses this signal to turn off 12.5V to this particular slot when the ejector is open with a default pull-up. GPIO3: is defined as RESET_BMC, 3.3VAUX signal, active high with a default pull-down on platform side. This signal is driven by 1S server to reset BMC on the platform.

KR0_TX_P/N	Output	Required	Primary 10GBase-KR Ethernet transmit signal. Data is sent from the 1S Server module to the platform.
KR0_RX_P/N	Input	Required	Primary 10GBase-KR Ethernet receive signal. Data is sent from the platform to the 1S Server module.
PCIE0_RESET#	Output	Required	PCIe reset signal. If a PCIe bus is connected, this signal provides the reset signal indicating the card VRs and clocks are stable when driven high to 3.3V.
PCIE0_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE0_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
PCIE0_REFCLK_P/_N	Output	Required	PCIe reference clock. This signal may or may not be connected on the platform.
PCIE1/2_RESET#	Output	Required	PCIe reset signals. If a PCIe bus is connected, this signal provides the reset signal indicating the card VRs and clocks are stable when driven high to 3.3V.
PCIE1_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE1_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
PCIE1_REFCLK_P/_N	Output	Required	PCIe reference clock. These signals may or may not be connected on the platform.
PCIe_VGA_TX_P/N	Output	Required	PCIe 2.0 or 3.0 transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE2_RESET#	Output	Required	PCIe reset signals. If a PCIe bus is connected, this signal provides the reset signal indicating the card VRs and clocks are stable when driven high to 3.3V.
PCIE2_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE2_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
PCIE2_REFCLK_P/_N	Output	Required	PCIe reference clocks. These signals may or may not be connected on the platform.

PCIE3_RESET#	Output	Required	PCIe reset signal. If a PCIe bus is connected, this signal provides the reset signal indicating the card VRs and clocks are stable when driven high to 3.3V.
PCIE3_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE3_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
PCIE3_REFCLK_P/_N	Output	Required	PCIe reference clocks. These signals may or may not be connected on the platform.
PCIE4_RESET#	Output	Required	PCIe reset signal. If a PCIe bus is connected, this signal provides the reset signal indicating the card VRs and clocks are stable when driven high to 3.3V.
PCIE4_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE4_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
PCIE4_REFCLK_P/_N	Output	Required	PCIe reference clock. These signals may or may not be connected on the platform.
PCIE5_RESET#	Output	Required	PCIe reset signal. If a PCIe bus is connected, this signal provides the reset signal indicating the card VRs and clocks are stable when driven high to 3.3V.
PCIE5_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE5_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
PCIE5_REFCLK_P/_N	Output	Required	PCIe reference clock. These signals may or may not be connected on the platform.
PCIE6_TX0/1/2/3_P/N	Output	Required	PCIe x4 bus transmit signals. Data is sent from the 1S Server module to the platform. These signals may or may not be connected on the platform.
PCIE6_RX0/1/2/3_P/N	Input	Required	PCIe x4 bus receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.

PCIe_VGA_RX_P/N	Input	Required	PCIe 2.0 or 3.0 receive signals. Data is sent from the platform to the 1S Server module. These signals may or may not be connected on the platform.
USB_P/N	Input/Output	Required	USB 2.0 differential pair.
FAST_THROTTLE_N	Input	Required	Active low open drain signal with pull-up on 1S server. Platform generates this signal and uses it as a big hammer to throttle 1S server down to lowest possible power state as fast as possible.
POWER_FAIL_N	Input	Required	Active low open drain signal with pull-up on 1S server. When this signal is asserted by platform, it informs 1S server that base system is going to cut 12V power to 1S server in certain amount of time, which is pre-defined by base system. It is possible for 1S server to perform graceful shutdown based on this signal.

5.6 PCIe Device Cards

As we have discussed previously, Yosemite V2 supports PCIe based device cards, such as standard PCIe add-in cards or customized PCIe cards in 1S server form factor.

5.6.1 Crane Flat Device Carrier

Figure 5-7 illustrates the block diagram of the Crane Flat device carrier that can host a standard, off-the-shelf PCIe add-in card.

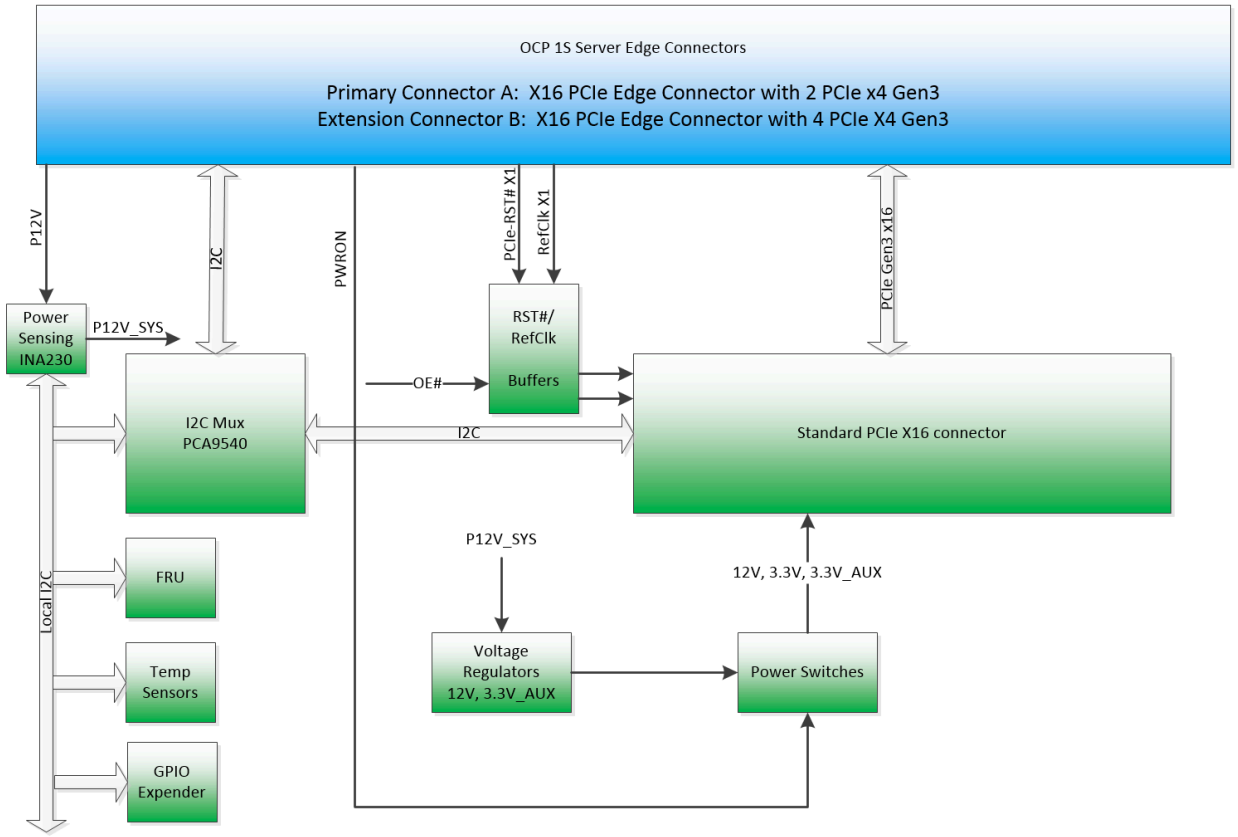


Figure 5-7: Crane Flat Device Carrier Block Diagram

The Crane Flat device carrier converts x16 PCIe lanes on 1S server extension connector to a standard x16 right angle PCIe connector. A standard, off-the-shelf PCIe add-in card can be used directly on the Yosemite V2 system through this Crane Flat device carrier, which will save a lot of design and validation time and expenses.

The Crane Flat device carrier shall support standard x16, X8, X4, X1 full-height half-length PCIe add-in cards with maximum power of 75W.

The BMC manage the Crane Flat through the management SMBus. There are FRU EEPROM, I2C GPIO expender, INA230 power sensor, the PCIe add-in card, and two thermal sensors.

The Crane Flat device carrier gets 12V from the edge connectors and converts it to 3.3V standby power, and 12V/3.3V payload power for the PCIe add-in card through power switches. The device carrier shall keep all payload power switches off until its paired 1S server issues POWER_EN. A power sensor INA230 is used to measure total power consumed by the Device Carrier.

Figure 5-8 illustrates the placement of the Crane Flat device carrier

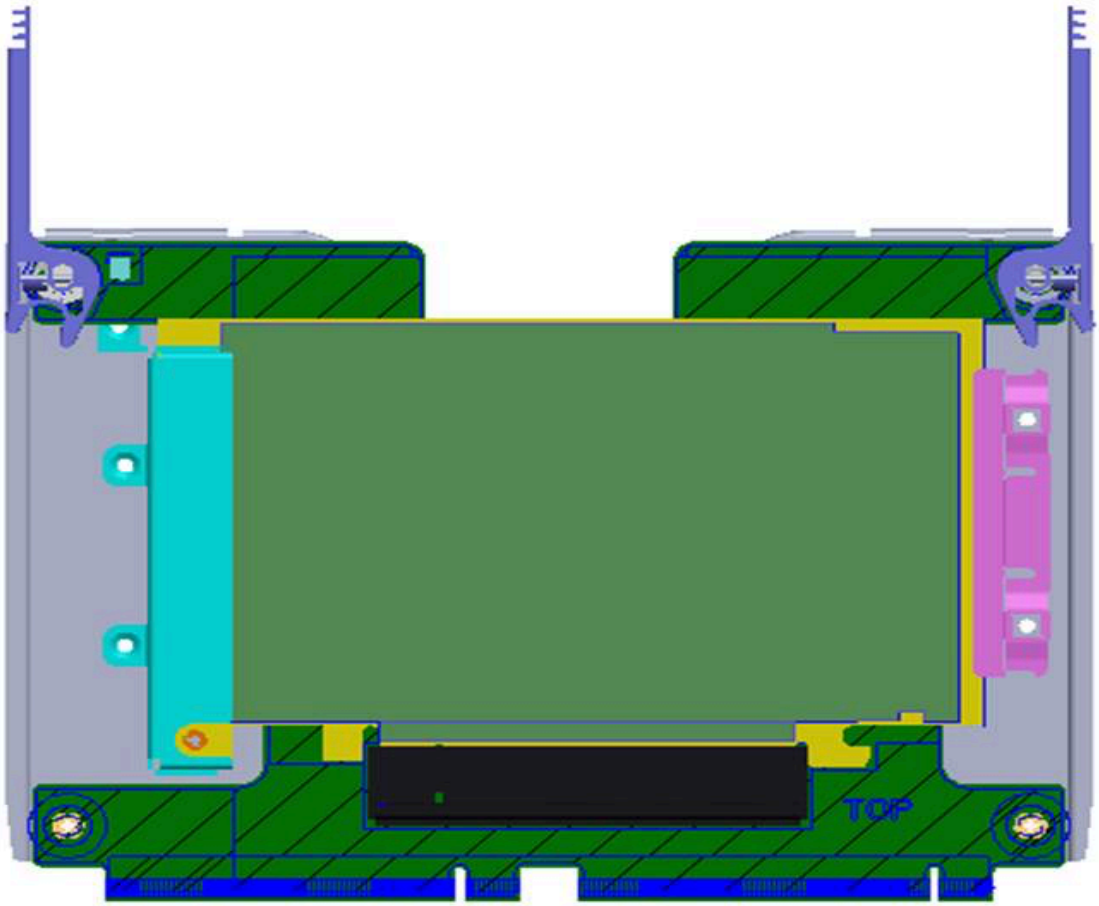


Figure 5-8: Crane Flat Device Carrier

5.6.2 Glacier Point Flash Card

Figure 5-9 illustrates the block diagram of the Glacier Point flash card.

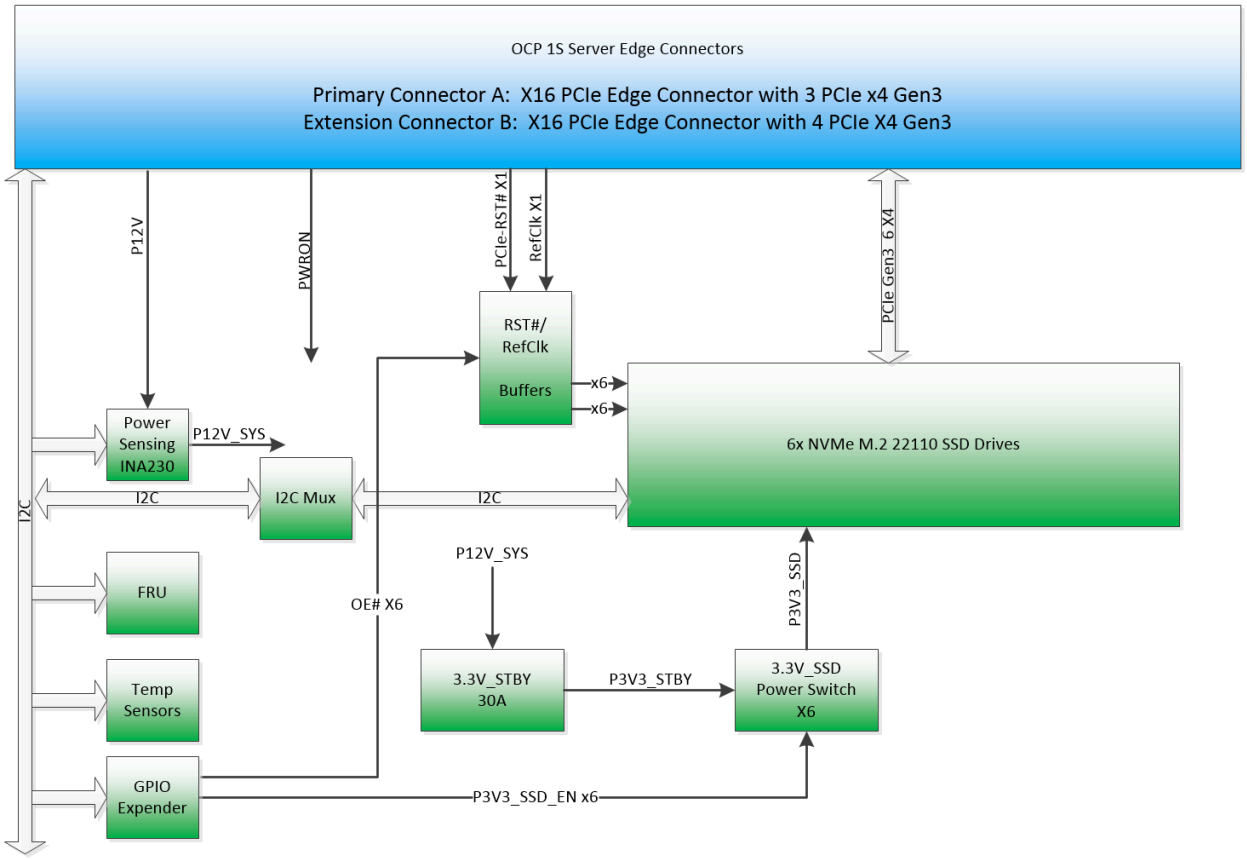


Figure 5-9: Glacier Point Flash Card Block Diagram

The Glacier Point flash card is a customized PCIe Flash card in 1S server form factor and it hosts 6 NVMe M.2 SSD drives in 22110 or 2280 form factor. Each drive uses an x4 PCIe Gen3 link connecting to edge connectors as shown in the block diagram. The flash card gets one PCIe reset signal (PCIE1_RESET#) and one PCIe reference clock (PCIE1_REFCLK_P/_N) from the primary connector, and uses buffers to provide PCIe reset and reference clocks to all NVMe SSD drives with open enable signals that are controlled by the BMC.

Similar to the Crane Flat device carrier card, the BMC shall manage the Glacier Point flash card through the management SMBus. On this SMBus, there are a FRU EEPROM, two thermal sensors, a power sensor INA230, I2C GPIO expender, and a mux to access M.2 NVMe SSD drives.

The Glacier Point flash card gets 12V from the edge connectors and converts it to 3.3V standby power, and 3.3V power for all SSD devices through power switches. Each payload power switch shall be off until its paired 1S server issues a POWER_EN signal. A power sensor INA230 is used to measure total power consumed by the Glacier Point flash card.

Figure 5-10 illustrates the placement of a Glacier Point flash card.

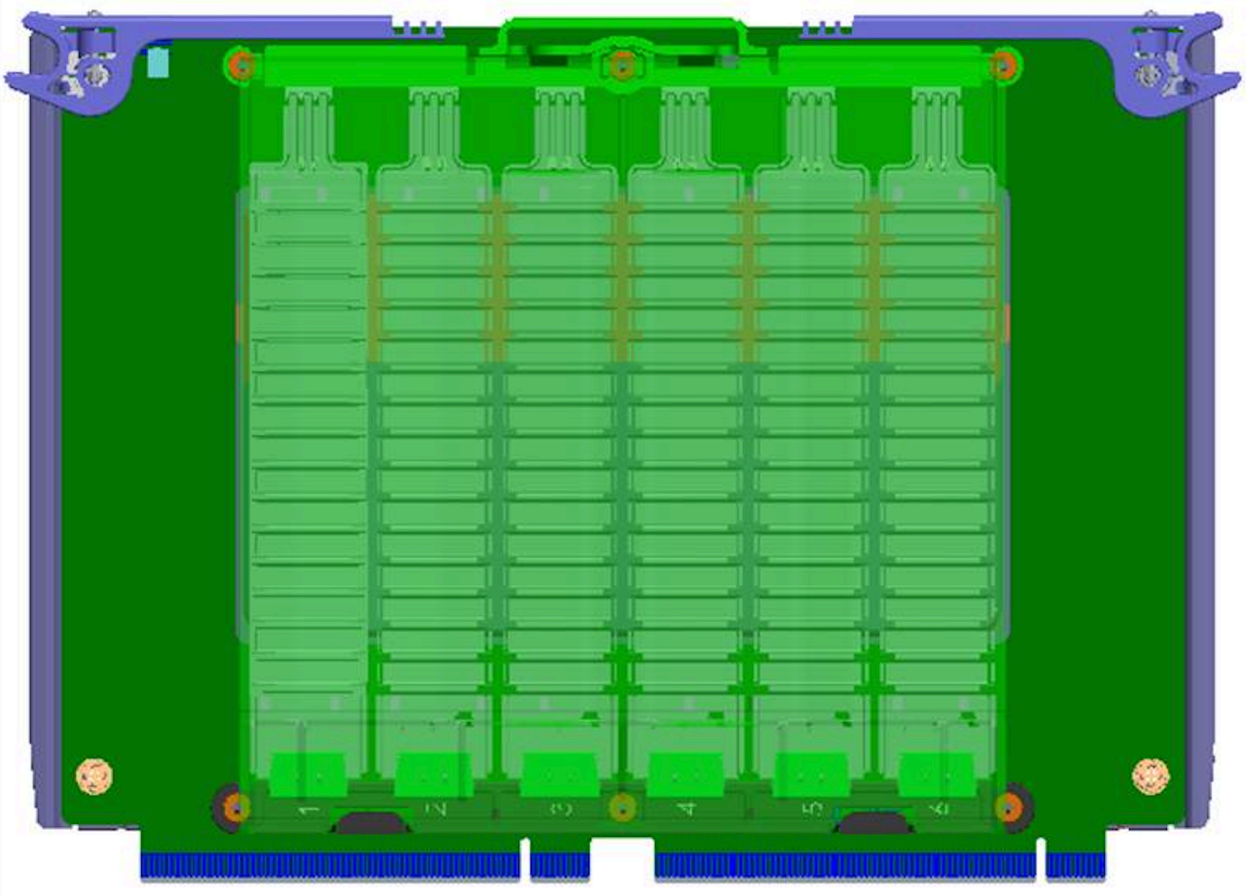


Figure 5-10: Glacier Point Flash Card

5.7 Network Options

The Yosemite V2 Platform provides various network options that support a 40GbE or 100GbE top of rack (TOR) switch with a single network cable.

To support various network adapter cards defined in the OCP 2.0 Mezzanine spec, two kinds of Adapter Cards have been designed to accommodate this requirement. A PCIe link with four lanes and a 10GBase-KR link from each 1S server, an NC-SI interface and SMbus between the BMC and management SMBus from BMC, and other related signals are routed to two OCP 2.0 Mezzanine connectors on the adapter cards through two AirMax connectors on the baseboard.

Adapter Card Type I supports following OCP 2.0 mezzanine card compliant network cards with a PCIe/10GBase-KR hybrid pinout:

- 4x10G KR-retimer mezzanine card
- Multi-Host 40G/50G NIC mezzanine card

Adapter Card Type II supports multi-host 100G NIC OCP 2.0 mezzanine cards.

5.7.1 OCP 2.0 PCIe-KR Hybrid Mezzanine Connector Pinout for Adapter Card Type I

Pin definitions of the hybrid mezzanine card for Adapter Card Type I are shown in the table below. The directions of the signals are from the perspective of the baseboard.

Table 4: OCP 2.0 Hybrid Mezzanine Connector Pinout

Connector A				Connector B			
Signal	Pin	Pin	Signal	Signal	Pin	Pin	Signal
P12V_AUX	A61	A1	MEZZ_PRSNTA1_N /BASEBOARD_A_ID	P12V_AUX	B41	B1	MEZZ_PRSNTB1_N /BASEBOARD_B_ID
P12V_AUX	A62	A2	P5V_AUX	P12V_AUX	B42	B2	GND
P12V_AUX	A63	A3	P5V_AUX	NC	B43	B3	KR_RX_DP<0>
GND	A64	A4	P5V_AUX	GND	B44	B4	KR_RX_DN<0>
GND	A65	A5	GND	KR_TX_DP<0>	B45	B5	GND
P3V3_AUX	A66	A6	GND	KR_TX_DN<0>	B46	B6	GND
GND	A67	A7	P3V3_AUX	GND	B47	B7	KR_RX_DP<1>
GND	A68	A8	GND	GND	B48	B8	KR_RX_DN<1>
P3V3	A69	A9	GND	KR_TX_DP<1>	B49	B9	GND
P3V3	A70	A10	P3V3	KR_TX_DN<1>	B50	B10	GND
P3V3	A71	A11	P3V3	GND	B51	B11	KR_RX_DP<2>
P3V3	A72	A12	P3V3	GND	B52	B12	KR_RX_DN<2>
GND	A73	A13	P3V3	KR_TX_DP<2>	B53	B13	GND
LAN_3V3STB_ALERT_N	A74	A14	NCSI_RCSDV	KR_TX_DN<2>	B54	B14	GND
SMB_LAN_3V3STB_CLK	A75	A15	NCSI_RCLK	GND	B55	B15	KR_RX_DP<3>
SMB_LAN_3V3STB_DAT	A76	A16	NCSI_TXEN	GND	B56	B16	KR_RX_DN<3>
PCIE_WAKE_N	A77	A17	PERST_N0	KR_TX_DP<3>	B57	B17	GND
NCSI_RXER	A78	A18	MEZZ_SMCLK	KR_TX_DN<3>	B58	B18	GND
GND	A79	A19	MEZZ_SMDATA	GND	B59	B19	NC
NCSI_TXD0	A80	A20	GND	GND	B60	B20	NC
NCSI_TXD1	A81	A21	GND	NC	B61	B21	GND
GND	A82	A22	NCSI_RXD0	NC	B62	B22	GND
GND	A83	A23	NCSI_RXD1	GND	B63	B23	NC
CLK_100M_MEZZ0_DP	A84	A24	GND	GND	B64	B24	NC
CLK_100M_MEZZ0_DN	A85	A25	GND	NC	B65	B25	GND
GND	A86	A26	CLK_100M_MEZZ1_DP	NC	B66	B26	GND
GND	A87	A27	CLK_100M_MEZZ1_DN	GND	B67	B27	NC
PCIe_MEZZ0_TX_DP_C<0>	A88	A28	GND	GND	B68	B28	NC
PCIe_MEZZ0_TX_DN_C<0>	A89	A29	GND	NC	B69	B29	GND
GND	A90	A30	PCIe_MEZZ0_RX_DP<0>	NC	B70	B30	GND
GND	A91	A31	PCIe_MEZZ0_RX_DN<0>	GND	B71	B31	NC
PCIe_MEZZ0_TX_DP_C<1>	A92	A32	GND	GND	B72	B32	NC
PCIe_MEZZ0_TX_DN_C<1>	A93	A33	GND	NC	B73	B33	GND
GND	A94	A34	PCIe_MEZZ0_RX_DP<1>	NC	B74	B34	GND
GND	A95	A35	PCIe_MEZZ0_RX_DN<1>	GND	B75	B35	CLK_100M_MEZZ2_DP
PCIe_MEZZ1_TX_DP_C<0>	A96	A36	GND	GND	B76	B36	CLK_100M_MEZZ2_DN

PCIe_MEZZ1_TX_DN_C<0>	A97	A37	GND	CLK_100M_MEZZ3_DP	B77	B37	GND
GND	A98	A38	PCIe_MEZZ1_RX_DP<0>	CLK_100M_MEZZ3_DN	B78	B38	PERST_N1
GND	A99	A39	PCIe_MEZZ1_RX_DN<0>	GND	B79	B39	PERST_N2
PCIe_MEZZ1_TX_DP_C<1>	A100	A40	GND	MEZZ_PRSNTB2_N	B80	B40	PERST_N3
PCIe_MEZZ1_TX_DN_C<1>	A101	A41	GND				
GND	A102	A42	PCIe_MEZZ1_RX_DP<1>				
GND	A103	A43	PCIe_MEZZ1_RX_DN<1>				
PCIe_MEZZ2_TX_DP_C<0>	A104	A44	GND				
PCIe_MEZZ2_TX_DN_C<0>	A105	A45	GND				
GND	A106	A46	PCIe_MEZZ2_RX_DP<0>				
GND	A107	A47	PCIe_MEZZ2_RX_DN<0>				
PCIe_MEZZ2_TX_DP_C<1>	A108	A48	GND				
PCIe_MEZZ2_TX_DN_C<1>	A109	A49	GND				
GND	A110	A50	PCIe_MEZZ2_RX_DP<1>				
GND	A111	A51	PCIe_MEZZ2_RX_DN<1>				
PCIe_MEZZ3_TX_DP_C<0>	A112	A52	GND				
PCIe_MEZZ3_TX_DN_C<0>	A113	A53	GND				
GND	A114	A54	PCIe_MEZZ3_RX_DP<0>				
GND	A115	A55	PCIe_MEZZ3_RX_DN<0>				
PCIe_MEZZ3_TX_DP_C<1>	A116	A56	GND				
PCIe_MEZZ3_TX_DN_C<1>	A117	A57	GND				
GND	A118	A58	PCIe_MEZZ3_RX_DP<1>				
GND	A119	A59	PCIe_MEZZ3_RX_DN<1>				
MEZZ_PRSNTA2_N	A120	A60	GND				

5.7.2 PCIe Mezzanine Card for Adapter Card Type II Pin Description

The PCIe mezzanine card pin description is shown in the table below. The signal direction is in the perspective of the baseboard.

Table 5: OCP 2.0 Mezzanine Card Pin Description

Signals on Connector A	Type	Description
GND	Ground	Ground return; total 51 pins on Connector A
P12V_AUX	Power	12V Aux power; total 3 pins on Connector A
P5V_AUX	Power	5V Aux power; total 3 pins on Connector A
P3V3_AUX	Power	P3V3 Aux Power; total 2 pins on Connector A
P3V3	Power	P3V3 power; total 8 pins on Connector A
MEZZ_PRSNTA1_N/BASEBOARD_ID_A	Output	Connector A Present Pin; connect to MEZZ_PRSNTA2_N on Mezz with 0 Ohm; Use as baseboard ID during power up
MEZZ_PRSNTA2_N	Input	Connector A Present Pin; connect to MEZZ_PRSNTA1_N on Mezz with 0 Ohm

LAN_3V3STB_ALERT_N	Input	SMBus Alert for OOB management; 3.3V AUX rail
SMB_LAN_3V3STB_CLK	Output	SMBus Clock for OOB management; 3.3V AUX rail; Share with thermal reporting interface
SMB_LAN_3V3STB_DAT	Bidirectional	SMBus Data for OOB management; 3.3V AUX rail; Share with thermal reporting interface
NCSI_RXER	Input	NC-SI for OOB management
NCSI_RCSDV	Input	NC-SI for OOB management
NCSI_RXD[1..0]	Input	NC-SI for OOB management
NCSI_RCLK	Output	NC-SI for OOB management
NCSI_TXEN	Output	NC-SI for OOB management
NCSI_TXD[1..0]	Output	NC-SI for OOB management
PCIE_WAKE_N	Input	PCIe wake up signal
PERST_N0	Output	PCIe reset signal or Node 0 PCIe reset signal
MEZZ_SMCLK	Output	PCIe SMBus Clock for Mezz slot/EEPROM; 3.3V AUX rail; Share with thermal reporting interface
MEZZ_SMDATA	Bidirectional	PCIe SMBus Data for Mezz slot/EEPROM; 3.3V AUX rail; Share with thermal reporting interface
CLK_100M_MEZZ[1..0]_DP/N	Output	PCIe Reference clock from Node [1..0]
PCIe_MEZZ0_TX_DP/N_C<1..0>	Output	PCIe TX from Node 0; total up to 2 lanes on Connector A
PCIe_MEZZ0_RX_DP/N<1..0>	Input	PCIe RX to Node 0; total up to 2 lanes on Connector A
PCIe_MEZZ1_TX_DP/N_C<1..0>	Output	PCIe TX from Node 1; total up to 2 lanes on Connector A
PCIe_MEZZ1_RX_DP/N<1..0>	Input	PCIe RX to Node 1; total up to 2 lanes on Connector A
PCIe_MEZZ2_TX_DP/N_C<1..0>	Output	PCIe TX from Node 2; total up to 2 lanes on Connector A
PCIe_MEZZ2_RX_DP/N<1..0>	Input	PCIe RX to Node 2; total up to 2 lanes on Connector A
PCIe_MEZZ3_TX_DP/N_C<1..0>	Output	PCIe TX from Node 3; total up to 2 lanes on Connector A
PCIe_MEZZ3_RX_DP/N<1..0>	Input	PCIe RX to Node 3; total up to 2 lanes on Connector A

Signals on Connector B	Type	Description
GND	Ground	Ground return; total 36 pins on Connector B
P12V_AUX	Power	12V Aux power; total 2 pins on Connector B

MEZZ_PRSNTB1_N/ BASEBOARD_ID_B	Output	Connector B Present Pin; connect to MEZZ_PRSNTB2_N on Mezz with 0 Ohm Use as baseboard ID during power up
MEZZ_PRSNTB2_N	Input	Connector B Present Pin; connect to MEZZ_PRSNTB1_N on Mezz with 0 Ohm
PERST_N[3..1]	Output	PCIe reset signal from Node[3..1]
CLK_100M_MEZZ[3..2]_DP/N	Output	PCIe Reference clock from Node [3..2]
KR_TX_DP/N<3..0>	Output	KR TX from Node[3..0]
KR_RX_DP/N<3..0>	Input	KR RX to Node[3..0]
NC	Open	These signals are not connected on adapter card.

5.7.3 OCP 2.0 PCIe-KR Hybrid Mezzanine Connector Pinout for Adapter Card Type II

Pin definitions of the hybrid Mezzanine card for Adapter Card Type I are shown in the table below. The directions of the signals are from the perspective of the baseboard.

Table 4: OCP 2.0 Hybrid Mezzanine Connector Pinout

Connector A				Connector B			
Signal	Pin	Pin	Signal	Signal	Pin	Pin	Signal
P12V_AUX	A61	A1	MEZZ_PRSNTA1_N /BASEBOARD_A_ID	P12V_AUX	B41	B1	MEZZ_PRSNTB1_N /BASEBOARD_B_ID
P12V_AUX	A62	A2	P5V_AUX	P12V_AUX	B42	B2	GND
P12V_AUX	A63	A3	P5V_AUX	NC	B43	B3	KR_RX_DP<0>
GND	A64	A4	P5V_AUX	GND	B44	B4	KR_RX_DN<0>
GND	A65	A5	GND	KR_TX_DP<0>	B45	B5	GND
P3V3_AUX	A66	A6	GND	KR_TX_DN<0>	B46	B6	GND
GND	A67	A7	P3V3_AUX	GND	B47	B7	KR_RX_DP<1>
GND	A68	A8	GND	GND	B48	B8	KR_RX_DN<1>
P3V3	A69	A9	GND	KR_TX_DP<1>	B49	B9	GND
P3V3	A70	A10	P3V3	KR_TX_DN<1>	B50	B10	GND
P3V3	A71	A11	P3V3	GND	B51	B11	KR_RX_DP<2>
P3V3	A72	A12	P3V3	GND	B52	B12	KR_RX_DN<2>
GND	A73	A13	P3V3	KR_TX_DP<2>	B53	B13	GND
LAN_3V3STB_ALERT_N	A74	A14	NCSI_RCSDV	KR_TX_DN<2>	B54	B14	GND
SMB_LAN_3V3STB_CLK	A75	A15	NCSI_RCLK	GND	B55	B15	KR_RX_DP<3>
SMB_LAN_3V3STB_DAT	A76	A16	NCSI_TXEN	GND	B56	B16	KR_RX_DN<3>
PCIE_WAKE_N	A77	A17	PERST_N0	KR_TX_DP<3>	B57	B17	GND
NCSI_RXER	A78	A18	MEZZ_SMCLK	KR_TX_DN<3>	B58	B18	GND

GND	A79	A19	MEZZ_SMDATA	GND	B59	B19	NC
NCSI_TXD0	A80	A20	GND	GND	B60	B20	NC
NCSI_TXD1	A81	A21	GND	NC	B61	B21	GND
GND	A82	A22	NCSI_RXD0	NC	B62	B22	GND
GND	A83	A23	NCSI_RXD1	GND	B63	B23	NC
CLK_100M_MEZZ0_DP	A84	A24	GND	GND	B64	B24	NC
CLK_100M_MEZZ0_DN	A85	A25	GND	NC	B65	B25	GND
GND	A86	A26	CLK_100M_MEZZ1_DP	NC	B66	B26	GND
GND	A87	A27	CLK_100M_MEZZ1_DN	GND	B67	B27	NC
PCIe_MEZZ0_TX_DP_C<0>	A88	A28	GND	GND	B68	B28	NC
PCIe_MEZZ0_TX_DN_C<0>	A89	A29	GND	NC	B69	B29	GND
GND	A90	A30	PCIe_MEZZ0_RX_DP<0>	NC	B70	B30	GND
GND	A91	A31	PCIe_MEZZ0_RX_DN<0>	GND	B71	B31	NC
PCIe_MEZZ0_TX_DP_C<1>	A92	A32	GND	GND	B72	B32	NC
PCIe_MEZZ0_TX_DN_C<1>	A93	A33	GND	NC	B73	B33	GND
GND	A94	A34	PCIe_MEZZ0_RX_DP<1>	NC	B74	B34	GND
GND	A95	A35	PCIe_MEZZ0_RX_DN<1>	GND	B75	B35	CLK_100M_MEZZ2_DP
PCIe_MEZZ1_TX_DP_C<0>	A96	A36	GND	GND	B76	B36	CLK_100M_MEZZ2_DN
PCIe_MEZZ1_TX_DN_C<0>	A97	A37	GND	CLK_100M_MEZZ3_DP	B77	B37	GND
GND	A98	A38	PCIe_MEZZ1_RX_DP<0>	CLK_100M_MEZZ3_DN	B78	B38	PERST_N1
GND	A99	A39	PCIe_MEZZ1_RX_DN<0>	GND	B79	B39	PERST_N2
PCIe_MEZZ1_TX_DP_C<1>	A100	A40	GND	MEZZ_PRSNTB2_N	B80	B40	PERST_N3
PCIe_MEZZ1_TX_DN_C<1>	A101	A41	GND				
GND	A102	A42	PCIe_MEZZ1_RX_DP<1>				
GND	A103	A43	PCIe_MEZZ1_RX_DN<1>				
PCIe_MEZZ2_TX_DP_C<0>	A104	A44	GND				
PCIe_MEZZ2_TX_DN_C<0>	A105	A45	GND				
GND	A106	A46	PCIe_MEZZ2_RX_DP<0>				
GND	A107	A47	PCIe_MEZZ2_RX_DN<0>				
PCIe_MEZZ2_TX_DP_C<1>	A108	A48	GND				
PCIe_MEZZ2_TX_DN_C<1>	A109	A49	GND				
GND	A110	A50	PCIe_MEZZ2_RX_DP<1>				
GND	A111	A51	PCIe_MEZZ2_RX_DN<1>				
PCIe_MEZZ3_TX_DP_C<0>	A112	A52	GND				
PCIe_MEZZ3_TX_DN_C<0>	A113	A53	GND				
GND	A114	A54	PCIe_MEZZ3_RX_DP<0>				
GND	A115	A55	PCIe_MEZZ3_RX_DN<0>				
PCIe_MEZZ3_TX_DP_C<1>	A116	A56	GND				
PCIe_MEZZ3_TX_DN_C<1>	A117	A57	GND				
GND	A118	A58	PCIe_MEZZ3_RX_DP<1>				
GND	A119	A59	PCIe_MEZZ3_RX_DN<1>				
MEZZ_PRSNTA2_N	A120	A60	GND				

5.7.4 Hybrid Mezzanine Card for Adapter Card Type I Pin Description

The Hybrid Mezzanine card pin description is shown in the table below. The signal direction is in the perspective of the baseboard.

Table 5: OCP 2.0 Mezzanine Card Pin Description

Signals on Connector A	Type	Description
GND	Ground	Ground return; total 51 pins on Connector A
P12V_AUX	Power	12V Aux power; total 3 pins on Connector A
P5V_AUX	Power	5V Aux power; total 3 pins on Connector A
P3V3_AUX	Power	P3V3 Aux Power; total 2 pins on Connector A
P3V3	Power	P3V3 power; total 8 pins on Connector A
MEZZ_PRSNTA1_N/BASEBOARD_ID_A	Output	Connector A Present Pin; connect to MEZZ_PRSNTA2_N on Mezz with 0 Ohm; Use as baseboard ID during power up
MEZZ_PRSNTA2_N	Input	Connector A Present Pin; connect to MEZZ_PRSNTA1_N on Mezz with 0 Ohm
LAN_3V3STB_ALERT_N	Input	SMBus Alert for OOB management; 3.3V AUX rail
SMB_LAN_3V3STB_CLK	Output	SMBus Clock for OOB management; 3.3V AUX rail; Share with thermal reporting interface
SMB_LAN_3V3STB_DAT	Bidirectional	SMBus Data for OOB management; 3.3V AUX rail; Share with thermal reporting interface
NCSI_RXER	Input	NC-SI for OOB management
NCSI_RCSDV	Input	NC-SI for OOB management
NCSI_RXD[1..0]	Input	NC-SI for OOB management
NCSI_RCLK	Output	NC-SI for OOB management
NCSI_TXEN	Output	NC-SI for OOB management
NCSI_TXD[1..0]	Output	NC-SI for OOB management
PCIE_WAKE_N	Input	PCIe wake up signal
PERST_N0	Output	PCIe reset signal or Node 0 PCIe reset signal
MEZZ_SMCLK	Output	PCIe SMBus Clock for Mezz slot/EEPROM; 3.3V AUX rail; Share with thermal reporting interface
MEZZ_SMDATA	Bidirectional	PCIe SMBus Data for Mezz slot/EEPROM; 3.3V AUX rail; Share with thermal reporting interface
CLK_100M_MEZZ[1..0]_DP/N	Output	PCIe Reference clock from Node [1..0]
PCIe_MEZZ0_TX_DP/N_C<1..0>	Output	PCIe TX from Node 0; total up to 2 lanes on Connector A
PCIe_MEZZ0_RX_DP/N<1..0>	Input	PCIe RX to Node 0; total up to 2 lanes on Connector A
PCIe_MEZZ1_TX_DP/N_C<1..0>	Output	PCIe TX from Node 1; total up to 2 lanes on Connector A

PCIe_MEZZ1_RX_DP/N<1..0>	Input	PCIe RX to Node 1; total up to 2 lanes on Connector A
PCIe_MEZZ2_TX_DP/N_C<1..0>	Output	PCIe TX from Node 2; total up to 2 lanes on Connector A
PCIe_MEZZ2_RX_DP/N<1..0>	Input	PCIe RX to Node 2; total up to 2 lanes on Connector A
PCIe_MEZZ3_TX_DP/N_C<1..0>	Output	PCIe TX from Node 3; total up to 2 lanes on Connector A
PCIe_MEZZ3_RX_DP/N<1..0>	Input	PCIe RX to Node 3; total up to 2 lanes on Connector A

Signals on Connector B	Type	Description
GND	Ground	Ground return; total 36 pins on Connector B
P12V_AUX	Power	12V Aux power; total 2 pins on Connector B
MEZZ_PRSNTB1_N/ BASEBOARD_ID_B	Output	Connector B Present Pin; connect to MEZZ_PRSNTB2_N on Mezz with 0 Ohm Use as baseboard ID during power up
MEZZ_PRSNTB2_N	Input	Connector B Present Pin; connect to MEZZ_PRSNTB1_N on Mezz with 0 Ohm
PERST_N[3..1]	Output	PCIe reset signal from Node[3..1]
CLK_100M_MEZZ[3..2]_DP/N	Output	PCIe Reference clock from Node [3..2]
KR_TX_DP/N<3..0>	Output	KR TX from Node[3..0]
KR_RX_DP/N<3..0>	Input	KR RX to Node[3..0]
NC	Open	These signals are not connected on adapter card.

5.7.5 OCP 2.0 4x10G KR-Retimer Mezzanine Card Design

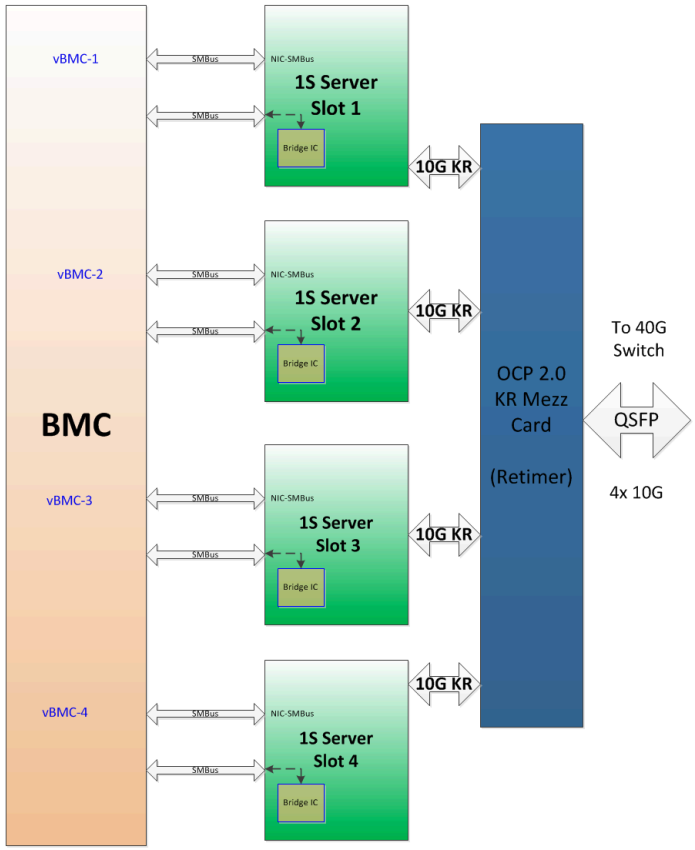


Figure 5-11: Yosemite V2 with 4X10G KR-Retimer Mezzanine Card

For the 1S servers with built-in 10GBase-KR network controllers, a 4x10G KR-retimer mezzanine card can be used to carry out four KR links in a single QSFP+ cable to a 40G TOR switch port that is configured in 4x10G mode. Figure 5-10 above illustrates a Yosemite V2 Platform that uses a 4x10G KR-retimer mezzanine card.

A KR-aware retimer, such as Inphi's CS4223 or Semtech's GN2007, is used to boost signal quality and compensate for the channel loss of four independent 10GbE links from the 1S servers to the mezzanine connectors. On the host side, the retimer shall fully support auto-negotiation and link training of 10GBase-KR protocol so that it can establish 10GbE links with the 1S servers. On the network side, the retimer shall support a single QSFP port, which can use QSFP copper or fiber cables.

The built-in 10GBase-KR network controller is used as a shared-NIC and its SMBus is connected to the BMC as the sideband. The BMC shall configure the network controller properly so that the network controller can bypass management traffic to the corresponding virtual BMC through the sideband.

Figure 5-12 illustrates the details of a 4x10G KR-retimer Mezzanine card.

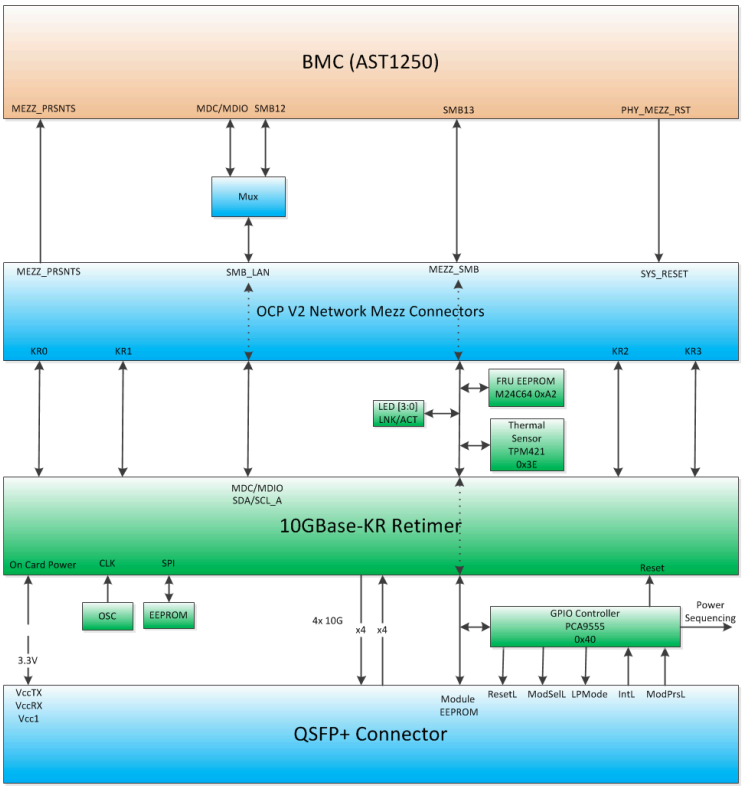


Figure 5-12: 4X10G KR-Retimer Mezzanine Card Block Diagram

The BMC manages the KR-retimer mezzanine card through MEZZ_SMB. Before powering up a mezzanine card, the BMC shall read the FRU EEPROM on the card to determine card type, configuration method, and functions. With the information in the FRU EEPROM, the BMC needs to flip the multiplexer on the baseboard properly to set SMB_LAN to either SMBus or MDC/MDIO on the BMC. After that, the BMC uses the GPIO expender on MEZZ_SMB to enable power supplies on the card, reset the retimer, and configure the retimer accordingly through SMB_LAN.

The BMC uses the GPIO expender to interact with a QSFP module as illustrated above. When there is a QSFP cable plugged into the card, the BMC shall read the EEPROM on the module to collect configuration information and then adjust the retimer's transceiver parameters accordingly.

A thermal sensor, preferably TPM421, resides on the MEZZ_SMB or SMB_LAN. The BMC monitors the thermal status of the card through this thermal sensor and controls the fans along with other thermal sensors in the platform. If a retimer offers an on-chip thermal sensor, it shall be accessible to the BMC through MEZZ_SMB or SMB_LAN.

It is possible to configure the retimer through an EEPROM instead of going through a BMC. However, the BMC must be able to update the firmware on the EEPROM if this configuration method is used.

The 4x10 KR-retimer mezzanine card must be compliant with mechanical and thermal requirements defined in the OCP 2.0 Mezzanine specification. A 15W maximum total card power

is strongly recommended to accommodate the Yosemite V2 Platform's power and thermal restrictions.

There are four LEDs on the card to indicate link and activity status of all 10GBase-KR interfaces. When the link is active, the LED shall be on. Where there is activity, the LED shall blink. It is preferred to have the retimer control these blue LEDs. However, if the retimer does not support LED control function, the BMC can drive the LEDs through an I2C LED controller on MEZZ_SMB, while the BMC gets link and status information from the Bridge IC on the 1S Servers. All LEDs shall be placed on front side of the sled, visible to the operators.

5.7.6 OCP 2.0 Multi-Host 40G/50G/100G NIC Mezzanine Card Design

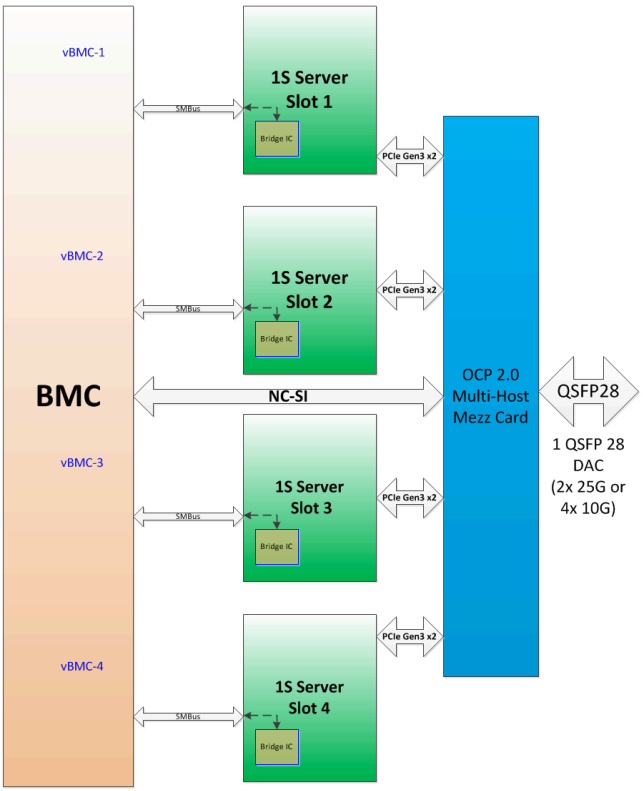


Figure 5-13: Yosemite V2 with multi-host 40G/50G PCIe Mezzanine Card

The network controller vendors offer Network Interface Cards (NICs) that support multi-host functions. These multi-host NICs provide network connectivity to multiple servers through a PCIe interface. For example, Mellanox's ConnectX-4 has PCIe interfaces to up to four independent servers with single or dual network interfaces to a TOR switch.

The Yosemite V2 Platform fully supports OCP 2.0 40G/50G/100G multi-host NICs through PCIe interfaces. As shown in Figure 5-8, there is a PCIe Gen3 link with four lanes from every 1S Server connected to the multi-host NIC through OCP 2.0 Mezzanine connectors. Every 1S server also provides its own PCIe reference clock, and PCIe reset to the multi-host NIC. With this configuration, every 1S server can operate its portion of the NIC independently regardless of any other 1S server's status.

On the network side, the multi-host NIC implements a single QSFP28 port, which can be configured automatically to 40G mode with four 10G lanes, or 50G mode with two 25G lanes, or 100G mode with four 25G lanes with proper TOR switch settings. To meet the strict signal integrity requirement of the whole channel between the multi-host NIC and a TOR switch in 25Gbps per lane speed, the channel loss from the NIC's network transceivers to the QSFP28 module must be less than 5dB. A channel loss of 3dB or less is strongly recommended. The NIC shall support both copper and fiber cables of various lengths.

There is a BMC on the Yosemite V2 Platform as the management controller. Depending on software implementation, the BMC can be virtualized so that every virtual BMC is assigned to a 1S server, or the BMC can manage all servers on a per slot basis. The multi-host NIC is used as a shared NIC on the platform and its NC-SI interface is used as the BMC's sideband. As there is only one NC-SI interface between the multi-host NIC and the BMC, it can be virtualized to provide dedicated sideband to every virtual BMC, or just provide one sideband to the whole BMC.

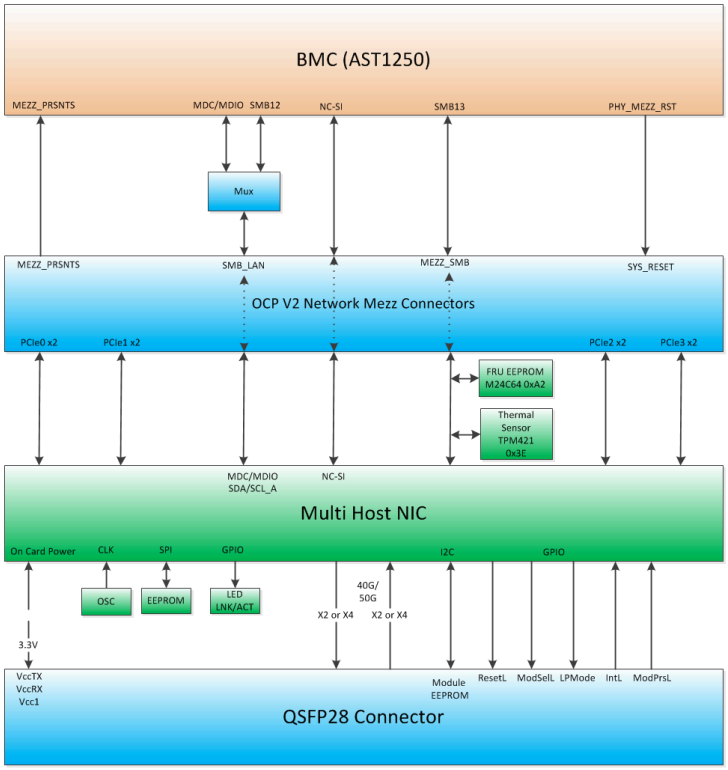


Figure 5-14: Multi-host 40G/50G PCIe Mezzanine Card Block Diagram

Figure 5-14 illustrates the details of a multi-host 40G/50G/100G NIC mezzanine card.

The BMC manages the multi-host NIC mezzanine card through MEZZ_SMB. As the multi-host NIC is an intelligent device, it will manage power, reset, and LEDs on its own. The multi-host is powered on when the Yosemite V2 Platform is powered on. The BMC shall read the FRU EEPROM on the card to determine card type, functions and configure the NC-SI interface.

The multi-host NIC also manages the QSFP28 module as illustrated above. If there is a QSFP28 cable plugged into the card, the multi-host NIC shall read the EEPROM on the module to collect configuration information and then adjust its transceiver parameters accordingly.

A thermal sensor, preferably TPM421, resides on MEZZ_SMB or SMB_LAN. The BMC monitors the thermal status of the card through this thermal sensor and controls the fans along with other thermal sensors in the platform. If there is an internal thermal sensor on the multi-host network controller IC, it shall be accessible to the BMC through MEZZ_SMB or SMB_LAN.

The multi-host NIC's firmware and configuration EEPROM can be updated by any of the 1S servers or the BMC. However, the NIC must implement a locking mechanism to eliminate race conditions such as two or more 1S servers and/or the BMC trying to access the firmware and/or configuration EEPROM at same time.

The multi-host NIC mezzanine card must be compliant with the mechanical and thermal requirements defined in the OCP 2.0 Mezzanine specification. 15W maximum total card power is strongly recommended to accommodate the Yosemite V2 Platform's power and thermal restrictions.

There are two LEDs on the card to indicate link and activity status of the link between the multi-host NIC and TOR switch. When the link is up, the link LED shall be on. Where there is activity, the activity LED shall blink. The NIC shall control this LED independently without the BMC's involvement. The LED shall be placed on front side of the sled, visible to the operators.

6 Baseboard Management Controller

The Yosemite V2 Platform uses a BMC for various platform management services and interfaces with hardware and BIOS firmware. The proposed BMC is ASPEED's AST2500.

The BMC should be a stand-alone system in parallel to the 1S servers (and/or device carrier cards). The health status of the 1S servers (and/or device carrier cards) should not affect the normal operation and network connectivity of the BMC.

6.1 1S Server I^2C Connections

There is a Bridge IC on each 1S server as a satellite management controller. The proposed Bridge IC is TI's Tiva micro-controller. The Intelligent Platform Management Bus Communications (IPMB) (I^2C) connection from the Bridge IC on the 1S server to the BMC is the primary management interface for the 1S server. Each 1S server's I^2C connection must be a separate port on the BMC to ensure a dedicated connection with no conflicting traffic. The speed of the I^2C connection is 1.0MHz if an AST2500 is used as the BMC and Texas Instruments' Tiva micro-controller is used as the Bridge IC. If a different BMC or Bridge IC is used, a maximum speed of 400kHz is expected, as per the I^2C specification.

The I^2C alert signal from each 1S server slot must be connected to the BMC. It provides an interrupt mechanism for the BMC. If the alert signal is asserted, the BMC must read the 1S server card and determine the source and cause of the interruption. If action is required, the BMC must respond in a timely fashion.

6.1.1 1S Server Command Interface

The BMC and the Bridge IC on the 1S server communicate with each other through the Intelligent Platform Management Bus (IPMB) protocol.

6.2 1S Server Serial Connections

All serial ports on the 1S server slots are connected to the BMC directly. The BMC shall implement Serial-Over-LAN (SOL) functionality to allow a user to access a 1S server remotely.

The BMC also shall redirect a 1S server's serial port to an OCP debug card on the front panel to allow local debugging. A user can use a switch to select which 1S server is connected to the debug card. By default, the BMC enables SOL to all 1S servers. When a OCP debug card is connected and activated on the selected 1S server, the BMC shall provide full access to the serial console for debug purposes and make the any existing SOL session to that server as a read-only session to avoid possible data collisions.

6.3 1S Server Discovery Process

6.3.1 Initial Discovery

The BMC can detect that a 1S server card is installed using the PRSNT# pin on the connector. If the signal is low, it means the BMC has detected a card and it has initiated the discovery process. The discovery sequence is defined as follows:

1. The BMC collects the FRU information from the Bridge IC
2. The BMC sensor tables are updated from the Bridge IC

3. The pin assignment tables are loaded in to the 1S server via the Bridge IC
4. The card is powered on based on the user input or as defined by the power policy configuration (e.g. always-off, always-on, last-power-state)

6.3.2 Pin Assignment Tables

As defined in the 1S server specification, (http://www.opencompute.org/wiki/Motherboard/SpecsAndDesigns, V0.7), a table of pin assignments must be provided to each 1S server. The table below describes the capabilities of the Yosemite V2 Platform.

Table 4: Yosemite V2 Platform Pin Assignment Table

Byte #	Byte value	Note
0	0x03	A13/A14 PCIe0 RefClk
1	0x02	A17/A18 PCIe0 Lane 0, Gen3
2	0x02	A21/A22 PCIe0 Lane 1, Gen3
3	0x02	A25/A26 PCIe0 Lane 2, Gen3
4	0x02	A29/A30 PCIe0 Lane 3, Gen3
5	0x05	A33/A34 SATA0, Gen3
6	0x03	A37/A38 PCIe2 RefClk
7	0x02	A49/A50 PCIe1 Lane 0 Gen3
8	0x02	A53/A54 PCIe1 Lane 1 Gen3
9	0x02	A57/A58 PCIe1 Lane 2 Gen3
10	0x02	A61/A62 PCIe1 Lane 3 Gen3
11	0x02	A65/A66 PCIe2 Lane 0 Gen3
12	0x02	A69/A70 PCIe2 Lane 1 Gen3
13	0x02	A73/A74 PCIe2 Lane 2 Gen3
14	0x02	A77/A78 PCIe2 Lane 3 Gen3
15	0x02	B15/B16 PCIe0 Lane 0, Gen3
16	0x02	B19/B20 PCIe0 Lane 1, Gen3
17	0x02	B23/B24 PCIe0 Lane 2, Gen3
18	0x02	B27/B28 PCIe0 Lane 3, Gen3
19	0x05	B31/B32 SATA0, Gen3
20	0x03	B35/B36 PCIe1 RefClk
21	0x02	B51/B52 PCIe1 Lane 0 Gen3
22	0x02	B55/B56 PCIe1 Lane 1 Gen3
23	0x02	B59/B60 PCIe1 Lane 2 Gen3
24	0x02	B63/B64 PCIe1 Lane 3 Gen3
25	0x02	B67/B68 PCIe2 Lane 0 Gen3
26	0x02	B71/B72 PCIe2 Lane 1 Gen3
27	0x02	B75/B76 PCIe2 Lane 2 Gen3
28	0x02	B79/B80 PCIe2 Lane 3 Gen3

The BMC must write the pin assignment to the EEPROM on the 1S server card via the Bridge IC. This ensures the BIOS can properly configure the SOC or turn off I/O that is incompatible. Before writing the pin assignment table, the Bridge IC shall check if the new table is different than the current one to avoid unnecessary updating.

6.4 1S Server Power-on Sequence

The BMC will de-assert the PWR_BTN# and SYS_RESET# signals to the 1S server to initiate power-on. The BMC will then poll the Power Good status from the 1S server to confirm if the 1S server card has powered on successfully. It will then update the power status.

6.5 Network Interface

The BMC has two possible network paths. First, if a PCIe based multi-host 40/50/100GbE NIC mezzanine card is used, the BMC can use its built-in media access controller (MAC) to transfer management traffic through an NC-SI interface with a TOR switch.

Second, the Yosemite V2 Platform may only have a KR PHY card on the Mezzanine slot and use the 1S server's built-in NICs. In this case, the BMC will use the NIC SMBus connections going from the BMC to each 1S server slot for OOB management traffic.

The mezzanine card needs to provide a Field Replaceable Unit ID (FRUID) as per the OCP 2.0 mezzanine card specification. The BMC shall use this FRUID to identify the card type and configure network paths accordingly. All unused interfaces and devices shall be disabled so that they will not interfere with the activated management interface and device.

The BMC FW needs to support both IPv4 and IPv6.

6.6 BMC Multi-Node Requirements

Since there are four 1S servers managed by a single physical BMC, the BMC shall provide virtualized BMC (vBMC) functionality to manage each server. The vBMC is responsible for providing local and remote management for a particular server. Depending on software implementation, each vBMC may have a unique IP address so that the remote clients can access and manage the server, or they may share a single IP address from the sideband.

6.7 Local Serial Console and Serial-Over-LAN

The BMC needs to support two paths to access a serial console:

- A local serial console on a debug header
- A SOL console

These must be supported through the management network described in Section **Error! Reference source not found.**. It is preferred that both interfaces are functional at all stages of system operation. When there is a legacy limitation that allows only one interface to be functional, the default is set to SOL. The BMC needs to be able to switch console connection between SOL and Local on the fly, based on the input of the Serial-Console-Select signal on the front panel.

During system booting, POST (Power On Self-Test) codes will be sent to Port 80 and decoded by the BMC to drive the LED display. POST codes should be displayed in the SOL console during system POST. Before the system displays the first screen, POST codes are dumped to – and displayed in – the SOL console in sequence (e.g., "[00] [01] [02] [E0]," etc.) After the system

shows the first screen in the SOL console, the last POST code received on Port 80 is displayed in the lower right corner of the console.

6.8 Graphics and GUI

The Yosemite V2 Platform does not require the BMC to support graphic, KVM or GUI features. All the BMC features need to be available in command-line mode by in-band and OOB IPMI command, or by SOL.

6.9 Remote Power Control and Power Policy

The vendor should implement the BMC firmware to support remote 1S server card power on/off/cycle and warm reboot through an in-band or out-of-band.

The vendor should implement the BMC firmware to support the power-on policy to be last state, always on, and always off. The default setting is last state. The change of power policy should be supported and take effect without cold resetting the BMC firmware or rebooting the 1S server system.

If AC power is applied to the BMC, it should take less than three seconds for the BMC to process the Power Button signal and power up the system for POST. It must not wait for the BMC to become ready (which will take about 90 seconds) before processing the Power Button signal.

In order to accommodate the requirement to process the Power Button signal in less than three seconds, the BMC shall enable a pass-through mode in the very early booting stages. This mode must make signals like Power Button, Reset, Universal Asynchronous Receiver/Transmitter (UART), POST Code, etc., available. Once the BMC boots completely (approximately 90 seconds), it shall also take over the control of these signals from the pass-through mode smoothly without any glitches.

6.10 POST Codes

The Bridge IC on the 1S server will pass POST codes to the BMC. The BMC should enable the POST code display to drive 8-bit HEX general Purpose Input/Output (GPIO) data to the OCP debug card on front panel. The BMC post function needs to be ready before the 1S server system BIOS starts to send the first POST code to the corresponding port. The POST codes should also be sent to the SOL so that the POST process can be monitored remotely.

6.11 Power and System Identification LEDs

The Yosemite V2 Platform combines a Power LED and a System Identification LED into a single bicolor blue-yellow LED per card. A total of 4×LEDs will be placed along the front edge of the board in a grid. The grid will be 2×rows of 2×LEDs to match the layout of the card slots.

When the Power LED is on, it defines the readiness of all power rails on the card. It also indicates the status of the card (overall health).

When the Yosemite V2 Platform is being identified by the BMC, all four yellow LEDs blink at 2.5Hz simultaneously, with 50% duty cycle (e.g., 200mSec ON and 200mSec OFF). All four blue LEDs shall be off, regardless of the power status of the cards. During this operation, all identification requests for an individual card inside the sled are ignored by the BMC. The identification operation shall continue until the user withdraws the identification request. This operation has the highest priority.

When an individual card is being identified by the BMC, its corresponding yellow LED blinks at 2.5Hz, with 50% duty cycle (e.g. 200mSec ON and 200mSec OFF). Its blue LED shall be off regardless of its power state. The identification operation shall continue until the user withdraws the identification request. This operation has second level priority.

When the selector knob is turned to the BMC position, all four blue LEDs blink at 1HZ simultaneously, with 50% duty cycle (e.g. 500mSec ON and 500mSec OFF). All four yellow LEDs shall be off. This operation has third level priority. All requests other than identification are ignored by the BMC.

If a card is not selected as the current server and is not being identified by the BMC, its corresponding LED shall operate according to the server's status. When the card is powered off, both the blue LED and the yellow LED shall be off. If the card is powered on and the server is operating normally, the blue LED shall be on and not blinking. The yellow LED shall be off. If the card is powered on and the server operates abnormally, such as a bad power state or if critical errors have been logged, the yellow LED shall be on and not blinking. The blue LED shall be off. This operation has the lowest priority.

If a card is selected as the current server but is not being identified by the BMC, its corresponding LED shall operate according to the server's status. When the card is powered off, the blue LED blinks at 1Hz with 10% duty cycle (e.g. 100mSec ON and 900 mSec OFF). The yellow LED shall be off. If the card is powered on and the server operates normally, the blue LED blinks at 1Hz with 90% duty cycle (e.g. 900mSec ON and 100mSec OFF). The yellow LED shall be off. If the card is powered on and the server operates abnormally, such as a bad power state or critical errors have been logged, the yellow LED blinks at 1Hz with 90% duty cycle (e.g. 900mSec OFF and 100mSec ON). The blue LED shall be off. This operation has lowest priority.

The Power LED blinks in different ways (varying colors and times) to convey various system statuses. Defined below are six different situations to make the LED operation more user friendly.

Situation 1: When the Yosemite V2 sled is being identified by the BMC, the BMC shall simultaneously blink all four yellow LEDs in 2.5 HZ (200ms on, 200ms off) while keeping all blue LEDs off. In this situation, the BMC shall ignore identification requests for individual servers inside this platform.

Situation 2: When the selector knob is turned to the BMC position, the BMC shall blink all four blue LEDs in 1HZ and 50% duty cycle (500ms on, 500ms off) simultaneously while keeping all yellow LEDs off.

Situation 3: When a card is not selected as the current server by the knob, but it is being identified by the BMC, regardless of its power status, the BMC shall blink this server's yellow LED in 2.5HZ frequency and 50% duty cycle while keeping this server's blue LED off.

Situation 4: When a card is not selected as the current server by the knob, and it is not being identified by the BMC, the BMC shall control the LEDs according to card's power and health status as below:

- If the card is in power off state, the BMC shall turn off both this server's blue and yellow LEDs.
- If the card is in power on state and operates, the BMC shall turn this server's blue LED solid on while keeping the yellow LED off.

- If the card is in power on state but operates abnormally, the BMC shall turn this server's yellow LED solid on but keep the blue LED off.

Situation 5: When a card is the current server selected by the knob and it is being identified by the BMC, regardless of this server's power status, the BMC shall blink this server's yellow LED in 2.5HZ and 50% duty cycle while keeping the blue LED off.

Situation 6: When a card is the current server selected by the knob and it is not being identified by the BMC, the BMC shall control the LEDs according to the card's power and health status as below:

- If the server is in the power off state, the BMC shall blink this server's blue LED in 1HZ frequency 10% duty cycle while keeping the yellow LED off (100ms on, 900ms off).
- If the server is in the power on state and operates normally, the BMC shall blink this server's blue LED in 1HZ frequency 90% duty cycle while keeping the yellow LED off (900ms on, 100ms off).
- If the server is in the power on state but operates abnormally, the BMC shall blink this server's yellow LED in 1HZ frequency 90% duty cycle while keeping the blue LED off (900ms on, 100ms off).

6.12 Time Sync

Since the Yosemite V2 Platform system has no CMOS battery backup, the BMC time sync should be from the Network Time Protocol (NTP) server instead of the Yosemite V2 Platform system.

The BMC should sync its clock from the NTP server as soon as its network interface is up and running.

The BMC should sync its clock from the NTP server periodically.

Since there is no battery on the BMC, the 1S server BIOS shall not issue IPMI Get System Event Log (SEL) Time command to sync its system clock during POST. The BMC should reject this command if its internal time is not properly synced up with NTP server.

6.12.1 NTP Time Sync Flow

1. BMC first time power on
2. The BMC tries to sync its time with one of the server cards as the server card might have battery backed up RTC
3. BMC firmware image might contain the default NTP IP address and NTP retry configuration.
4. Provisioning server will Set NTP IP Address command to the BMC
5. BMC network interface up and running
6. BMC queries date/time for the BMC clock using the configured NTP IP address
7. A default date/time (e.g., either time from one of the server cards) will be used for any event log until date is properly set
8. Once the BMC date/time is synced, the BMC will log the event and start using new time for any events that happen later
9. The BMC will sync its date/time from the NTP server periodically with an interval

6.13 Power and Thermal Monitoring, and Power Limiting

The BMC firmware shall support platform power monitoring. Enabling power monitoring for the 1S servers requires an accurate power sensor on 12.5V to the 1S server. This function should be able to access through in-band and OOB.

The BMC firmware shall support thermal monitoring, including 1S server SOCs, 1S server memory, and inlet/outlet air temperatures. To ensure accuracy, a TI TMP421 with an external PN junction is preferred to detect inlet and outlet temperatures. Take caution when implementing inlet air sensors. It is important to avoid preheating nearby components and to reduce the amount of heat conducted through the printed circuit board (PCB).

The BMC firmware shall support a power-limiting feature to make sure the platform is not drawing more power than allocated. The BMC will monitor the power consumption of each 1S server and use a SOC-specific management controller interface to limit the SOC's power consumption (e.g., P-State control).

6.14 Sensors

Both analog and discrete sensors may reside on the baseboard and on the cards. The BMC must provide a way to read all sensors across platform, e.g., sensors on the baseboard, sensors on a given server, sensors on a device carrier card, and sensors on the OCP mezzanine card.

6.14.1 Analog Sensors

The BMC has access to all analog sensors on the Yosemite V2 Platform directly or through the 1S server management connection.

Some of the required analog sensors include (but are not limited to):

- Outlet Temp
- Inlet Temp
- Slot Current
- SoC Thermal Margin
- SoC VR Temp
- SoC DIMM VR Temp
- Hot Swap Controller's power/current/voltage
- SoC TjMax
- Airflow
- System Fan Speed

6.14.2 BIOS/ME generated Sensors

Sometimes when the BIOS/ME detects a failure, it generates an SEL entry to be logged in the BMC. Some of the required event only sensors include (but are not limited to):

- Firmware health
- POST errors
- Power errors
- ProcHOT
- Machine Check errors
- PCIe errors

- Memory errors, etc.

6.15 Event Log

The vendor should implement the BMC to support storing events/logs from each 1S server, baseboard, device carrier card, and mezzanine card.

6.15.1 Logged Errors

6.15.1.1 CPU Error

Both correctable ECC errors and uncorrectable ECC errors should be logged into the Event log. Error categories include Link and L3 Cache.

6.15.1.2 Memory Error

Both correctable ECC errors and uncorrectable ECC errors should be logged into the Event log. The Error log should indicate the location of the DIMM (if applicable), channel #, and slot #.

6.15.1.3 PCI-E Error

All errors, which have a status register, should be logged into the Event log, including root complex, endpoint devices, and any switch upstream/downstream ports if available. Link disable on errors should also be logged. The error classifications Fatal, Non-fatal, or Correctable follow the 1S server vendor's recommendation.

6.15.1.4 POST Error

All POST errors, which are detected by BIOS during POST, should be logged into the Event log.

6.15.1.5 Power Error

Two power errors should be logged. One is a 12.5V DC input power failure that causes all power rails on the baseboard to lose power, including standby power. The other is an unexpected system shutdown during system S0/S1 while the 12.5V DC input is still valid.

6.15.1.6 MEMHOT# and SOCHOT#

Memory hot errors and processor hot errors should be logged. The Error log should identify the error source as internal, coming from the processor or memory, or an external error coming from the voltage regulator.

6.15.1.7 Fan Failure

Fan failure errors should be logged if the fan speed reading is outside expected ranges between the lower and upper critical thresholds. The Error log should also identify which fan fails.

6.15.1.8 PMBus Status Error

The PMBus status sensors check the PMBus controller's health status and log an error if an abnormal value is detected. The PMBus controller can be a DC Hot Swap Controller (HSC) or a PMBus AC to DC power supply unit.

For all above error logging and reporting, the user may select to enable or disable each logging option.

6.15.2 Error Threshold Setting

Enable the error threshold setting for both correctable and uncorrectable errors. Once a programmed threshold is reached, the system should trigger an event and log it.

- Memory Correctable ECC: Suggest setting the threshold value to be [1,000] in the mass production stage and [4] for the evaluation, development, and pilot run stage, with options of 1, 4, 10, and 1,000. When the threshold is reached, the BIOS should log the event, including DIMM location information and the output DIMM location code, through the debug card.
- ECC Error Event Log Threshold: Defines the maximum number of correctable DIMMs. ECC is logged in the same boot. The default value is 10, with options of Disable, 10, 50, and 100.
- PCIe Error: Follow the 1S server vendor's suggestion.

6.16 Fan Speed Control in BMC

The vendor should enable Fan Speed Control (FSC) on the BMC. The BMC samples thermal related analog sensors in real time. The FSC algorithm processes these inputs and drives two pulse width modulation (PWM) outputs in optimized speed.

6.16.1 Fan Speed Control Specification

The FSC implementation in the BMC must refer to the OCP's FSC specification.

6.16.2 Data gathering for FSC

The BMC needs to gather data as input of the FSC. The required data is described in the table below.

Table 4: Required FSC Data

Type of data	Data to be used for FSC input
Temperature	1S server SOC temperature from all slots
Temperature	1S server DIMM temperature from all slots (if available)
Temperature	Inlet and outlet air
Temperature	1S server VR of SOC and DIMM from all slots (if available)
Temperature	Hot Swap Controller
Temperature	Switch temperature

Power	Platform power from HSC
Fan speed	2 Fan tachometer inputs

6.16.3 Fan Speed Controller in BMC

The BMC should support FSC in both proportional–integral–derivative (PID) and step mode. The BMC should support both in-band and OOB FSC configuration updates. Updates should take effect immediately without rebooting. The BMC should support fan boost during fan failure.

6.16.4 Fan Connection

The Yosemite V2 Platform baseboard has one fan header on the motherboard.

6.16.5 Fan Tray

Yosemite has a cold-swap fan tray, which is comprised of 2x 80mm fans + a cable set to blind-mate interface with the baseboard.

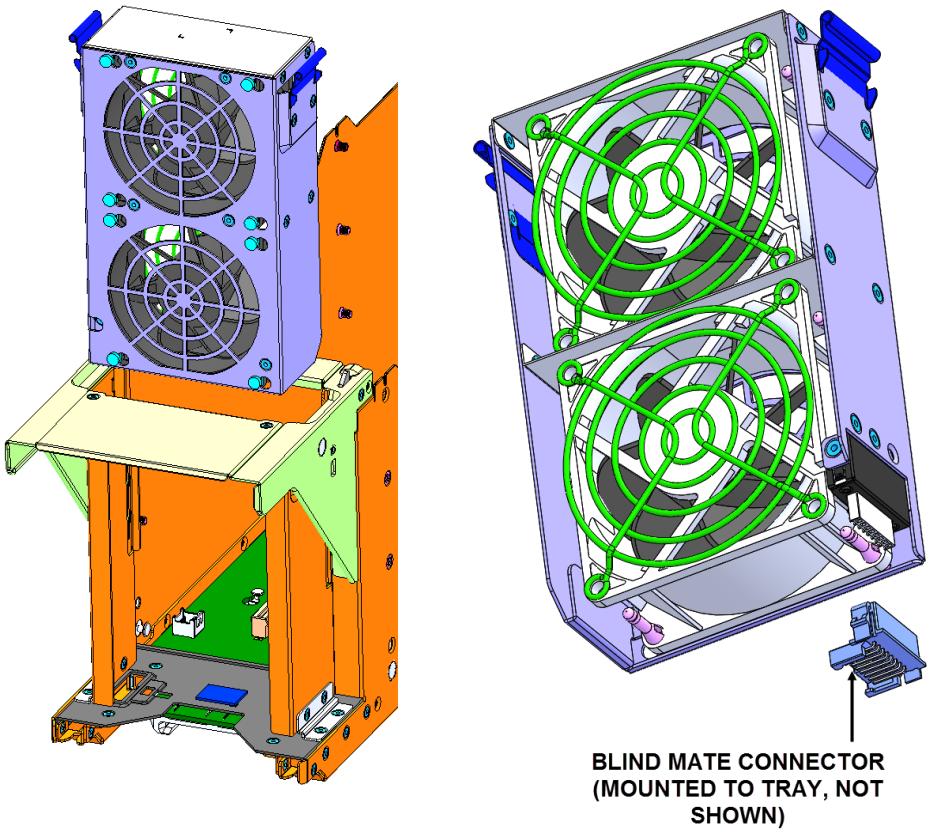


Figure 6.16.5: Fan Tray Service Mode

6.17 BMC Firmware Update

Vendors should provide tool(s) to implement a remote BMC firmware update, which will not require any physical input. This remote update can occur either through OOB via the management network or by logging into the local OS (CentOS) via the data network. Tool(s) shall support CentOS.

A remote BMC firmware update may take five minutes (maximum) to complete. The BMC firmware update process and BMC reset process do not require the host system to reboot or power down. It should have no impact to the normal operation of the host system. The BMC needs to be fully functional, with updated firmware after the update and reset, without any further configuration.

The default update should recover the BMC to the factory default settings. Options need to be provided to preserve the SEL and configuration. The MAC address should not be cleared with the BMC firmware update.

6.18 Server to Device Card Association

The Yosemite V2 Platform supports the concept of server to device card association. For example, a server in slot #1 and device card in slot #3 can be configured as a single pair. The BMC should take care when doing power control for the paired cards. For example, when a user requests power off for a device carrier card, the BMC might have to inform corresponding paired servers before actually powering off the server.

6.19 Hot Service Support

The Yosemite V2 Platform supports hot service of any card in the system while keeping all other cards in service. The BMC shall detect these hot insertions and/or removals and update its database (FRUID information, sensor information, etc.). Since the newly inserted card could possibly be of a different kind, the BMC should be able to detect the new card and configure different services, e.g., sensor monitoring might need restart for that slot to reflect the new hardware.

6.20 OpenBMC

OpenBMC refers to open source implementation of BMC functionality described in the above sections. This specification does not prevent alternate implementations that can meet similar functionality. The source code for OpenBMC is available at https://github.com/facebook/openbmc for reference.

7 Mechanical

The Yosemite V2 Platform is an Open Rack V2 compatible compute platform via the vCubby 4–bay shelf for Open Rack V2. Each VCubby can hold up to 4x compute sleds.

7.1 vCubby Chassis

vCubby is a power-mechanical shelf distributing power from the rack bus bars to up to four sled bays per shelf. Figure 7-1 shows a vCubby with the maximum available sled volume (yellow).

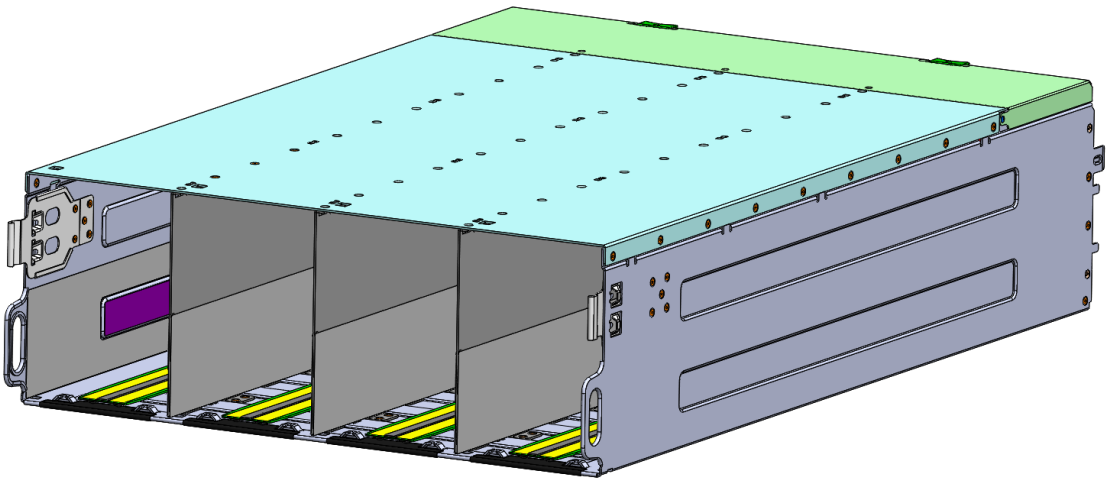


Figure 7-1: vCubby Chassis

7.2 Sled Chassis

A sheet metal and plastic sled serves as the mechanical interface between the Yosemite V2 Platform and the vCubby chassis. It also provides mechanical retention for the components inside the sled, such as the power cable assembly, fan, mezzanine card, baseboard, and 1S server cards. The combination of sled, baseboard and other components assembled in the chassis is a Yosemite V2 Platform sled.

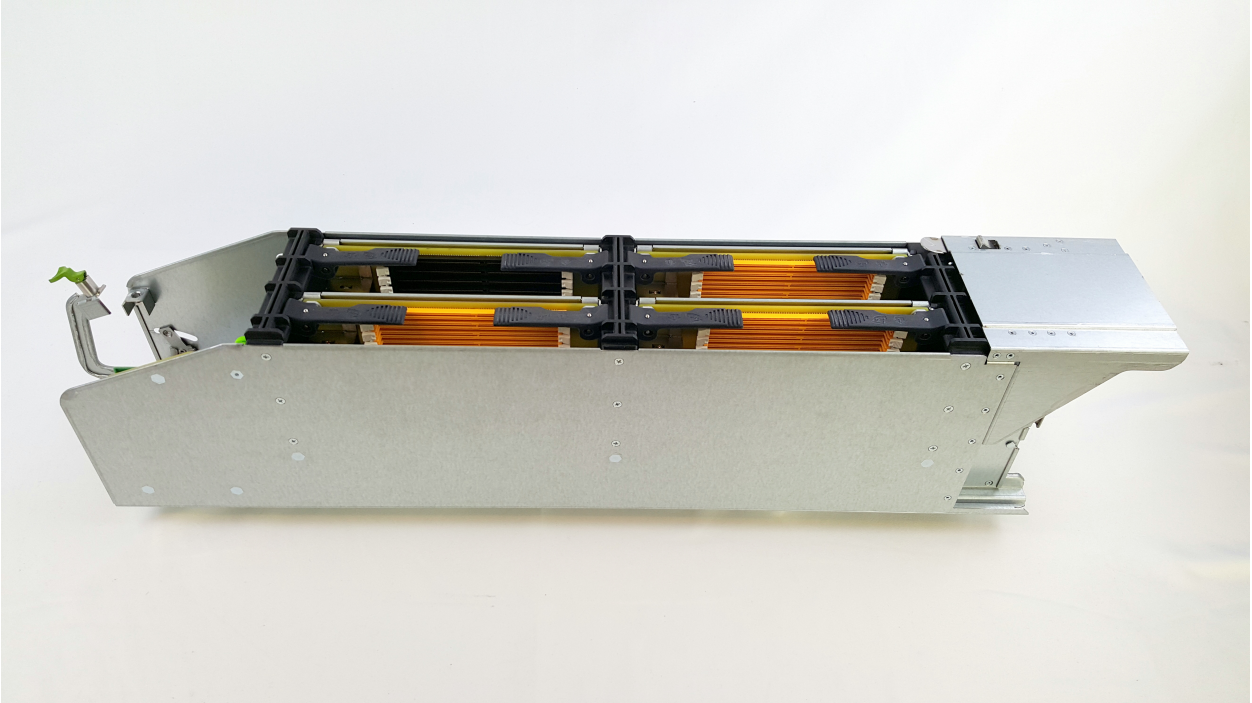
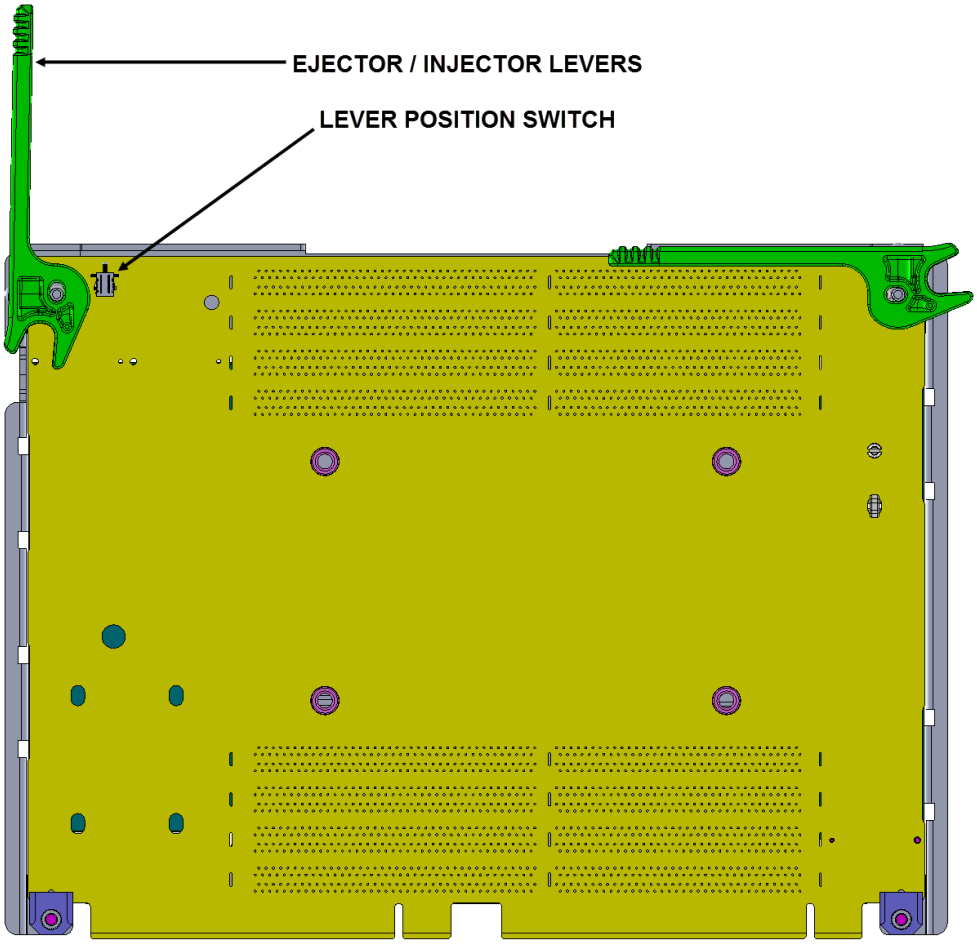


Figure 7-2: Yosemite V2 Platform Chassis, Populated with 4x 1S Server

7.3 1S Server Card Construction and Retention/Extraction

The server card is mounted to a sheet metal carrier, which slides between cardguides mounted inside the sled. The card assembly is installed/removed from the sled by rotating ejector levers to overcome the significant insertion force required by the PCI connectors.

There is a limit switch on one lever to detect when it is opened, such that the server is aware of the service event and can shut down accordingly if the technician has not already done so. This is a backup failsafe mechanism.


EJECTOR / INJECTOR LEVERS
LEVER POSITION SWITCH

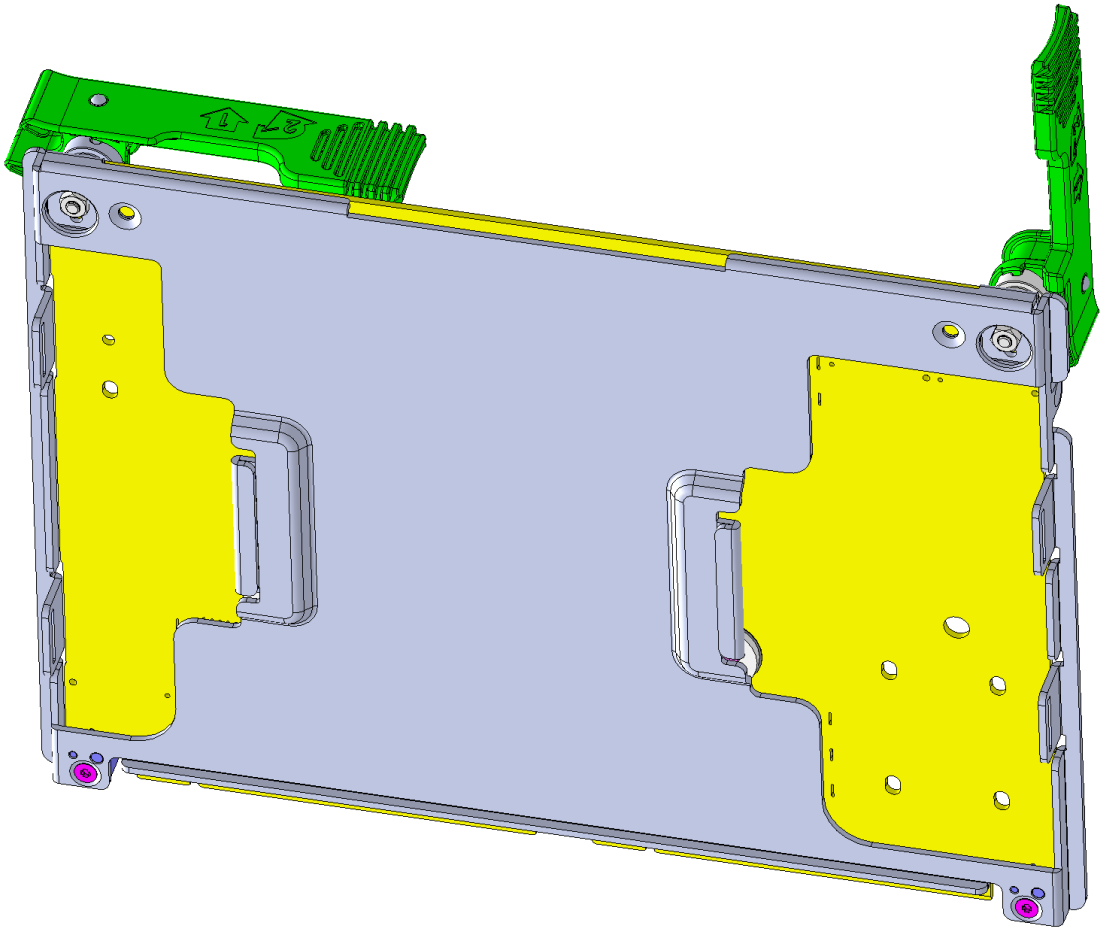


Figure 7-3: Server Card Assembly

7.4 Silkscreen

Silkscreens will be white in color and include labels for the components listed below. Additional items required on the silkscreen are listed in Section 12.

- Micro-server slots
- Fan connectors

LEDs

- Switches as PWR and RST.
- Keep-out area (see the General Card Specification drawing above)

7.5 Sled Retention

At the interface between The vCubby and the sled, a there is a “tool required” latch that snaps upward when the sled is pulled out to its furthest service position. The purpose of this latch is to prevent the sled from dropping. To completely remove the sled, a “tool” such as a pen or screwdriver must be used to disengage the latch. This is to comply with IEC 60950-1 safety standards.

When the sled is in its “home” position within the vCubby, a rotating combination pull/push handle engages with the bottom of the vCubby to ensure it is held firmly in place.

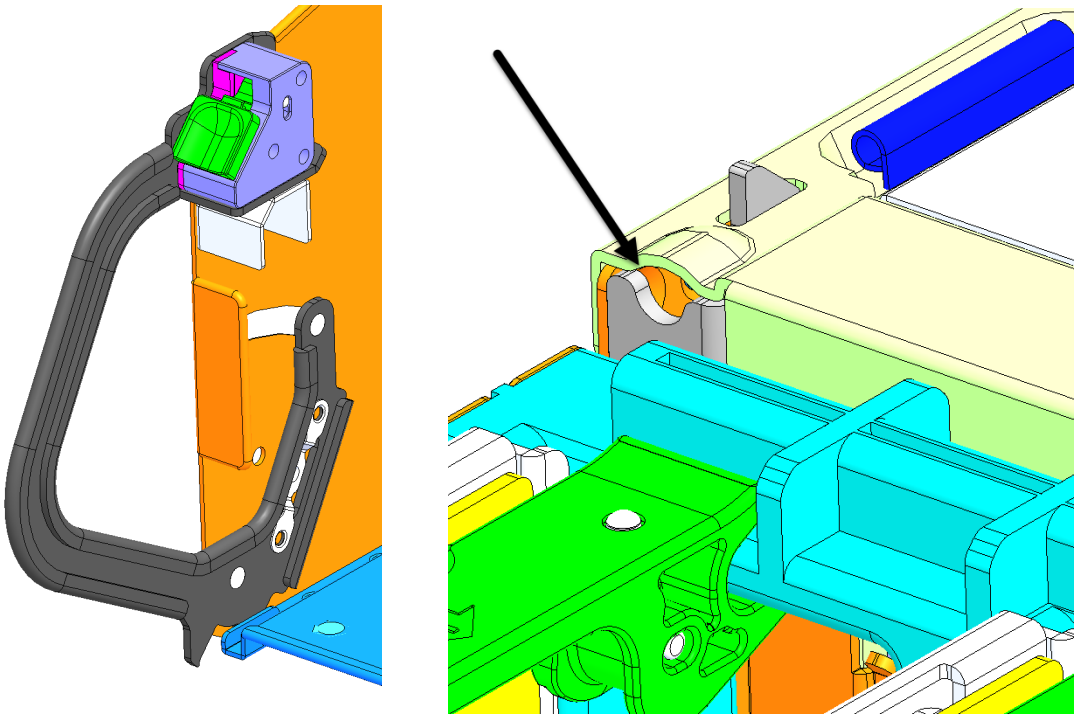


Figure 7-11: Sled Pull Handle and Service Position Latch

Connector 1 is the vCubby chassis connector. Connectors 2-3 are the cabled sled connectors. Connectors 4 or 5 will be mounted on the side-plane. The table below shows the part numbers. The table is subject to expansion as more connectors are qualified.

Table 5: Connector Part Numbers

Connector	Part Number
1	1-1892903-2
2	1-1892933-1
3	1-1892820-1
4	6450824-5

5	6450844-2
6	3-6450840-6

The baseboard will support up to four 1S server cards of fixed width (210mm) and discrete length (110mm and 160mm) via generic vertical PCIe x16 connectors (Figure 7-12). A sample acceptable reference part number is TE PN 7-1734774-3.

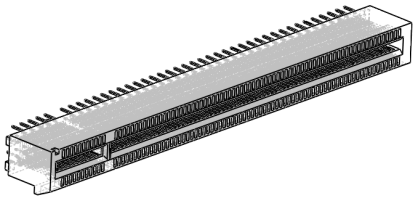


Figure 7-12: Generic vertical x16 PCIe connector

The baseboard will support the OCP Mezzanine 2.0 form factor (PCB with keepouts and connectors attached). The I/O port(s) (at least 1 QSFP or QSFP+) will face the front of the sled (Figure 7-13).

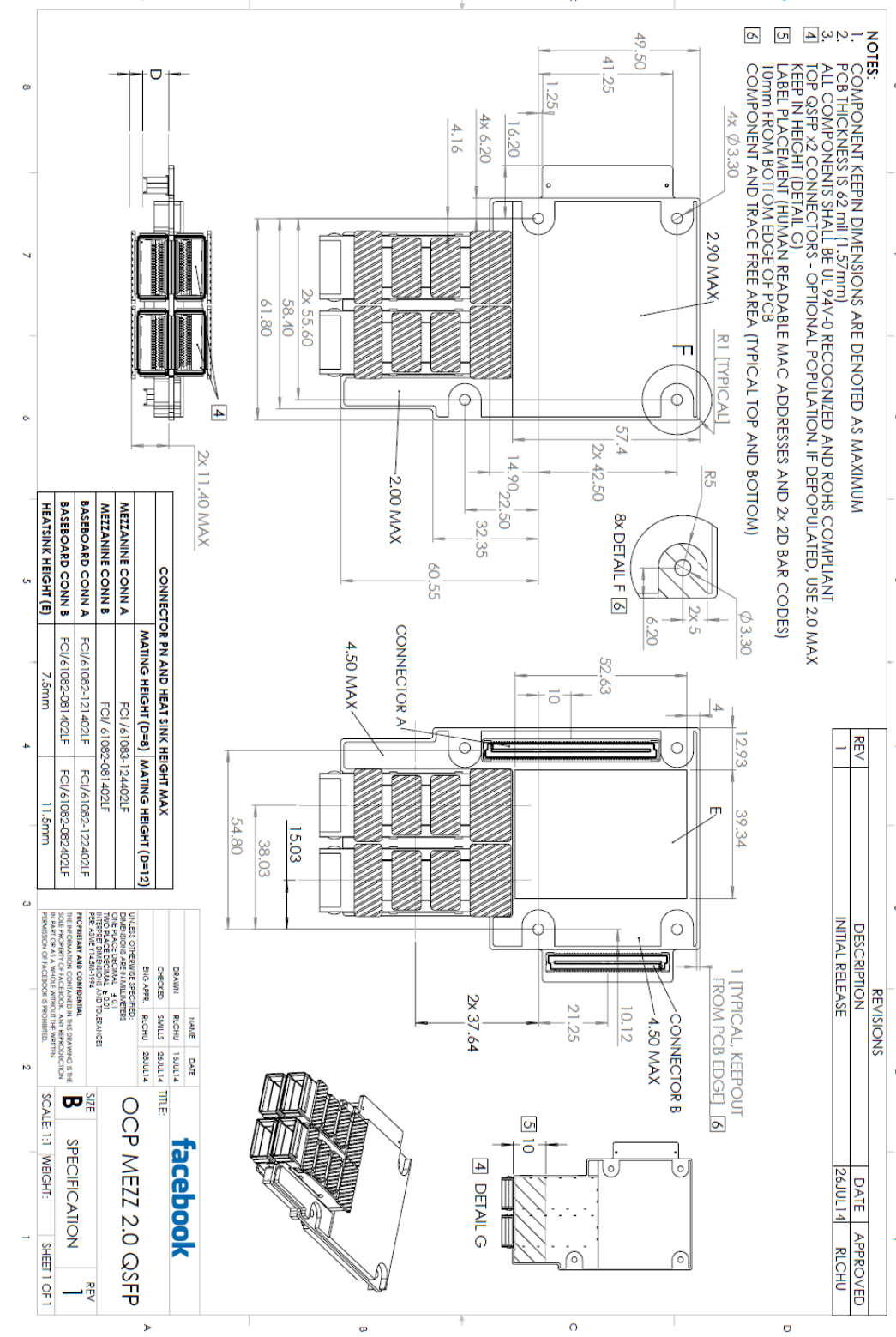


Figure 7-13: OCP mezzanine 2.0 QSFP specification drawing

8 Thermal

To meet thermal reliability requirements, the cooling solution should dissipate heat from the components when the system is operating up to its maximum thermal power. The final thermal solution of the system should be optimized and energy efficient under data center environmental conditions with the lowest capital and operating costs. The thermal solution should not allow any overheating issue for any components in system.

8.1 Data Center Environmental Conditions

This section outlines Facebook data center operational conditions.

8.1.1 Location of Data Center/Altitude

Maximum altitude is 6,000 ft above sea level. Any variation of air properties or environmental difference due to the high altitude needs to be deliberated into the thermal design.

8.1.2 Cold-Aisle Temperature

Data centers generally maintain cold aisle temperatures between 18°C and 30°C (65°F to 85°F). The mean temperature in the cold aisle is usually 25°C with 3°C standard deviation. The cold aisle temperature in a data center may fluctuate minutely depending to the outside air temperature. Every component must be cooled and must maintain a temperature below its maximum specification temperature in the cold aisle.

8.1.3 Cold-Aisle Pressurization

Data centers generally maintain cold aisle pressure between 0 inches H_2O and 0.005 inches H_2O. The thermal solution of the system should consider the worst operational pressurization possible, which generally is 0 inches H_2O and 0.005 inches H_2O with a single fan (or rotor) failure.

8.1.4 Relative Humidity

Data centers usually maintains a relative humidity between 20% and 90%.

8.2 Server Operational Conditions

8.2.1 Inlet Temperature

The inlet air temperature will vary. The cooling system in the Yosemite V2 Platform should be able to cover inlet temperatures including 20°C, 25°C, 30°C, and 35°C. Cooling above 30°C is beyond the Facebook operational condition, but used during validation to demonstrate the thermal reliability and design margin. Any degraded performance is not allowed over the validation range 0°C-35°C.

8.2.2 Pressurization

Except for the condition when one rotor or one fan in a server fan fails, the thermal solution should not consider extra airflow from data center cooling fans. If and only if one rotor or one

fan in a server fan fails, the negative or positive DC pressurization can be considered in the thermal solution in the hot aisle or the cold aisle, respectively. The maximum pressurization is 0.005 inches H_2O which is in inlet to system.

8.2.3 Fan Redundancy

The server fans at N+1 redundancy should be sufficient for cooling server components to temperatures below their maximum specification to prevent server shut down or to prevent either CPU or memory throttling. An N+1 fan redundancy in the Yosemite V2 Platform is preferred when the system is operating under normal conditions.

8.2.4 Delta T

The Delta T is the air temperature difference across the system, or the temperature difference between the outlet air temperature and the inlet air temperature. The Delta T must be greater than 13.9°C (25°F) at the rack level when the server is running within the data center operational condition. The desired server level Delta T is greater than 17°C (31°F) when the inlet air to the system is equal to or lower than 30 °C.

8.2.5 System Airflow or Volumetric Flow

The unit of airflow (or volumetric flow) used for this spec is cubic feet per minute (CFM). The CFM can be used to determine the thermal expenditure or to calculate the approximate Delta T of the system. The thermal expenditure is quantified by the metric CFM/W, which is calculated by the following formula:

$$\text{Thermal Expenditure} = \frac{\text{System airflow}}{\text{Total system power consumption, including fans}} \ [\text{CFM/W}]$$

At sea level, the maximum allowable airflow per watt in a Yosemite V2 rack is 0.13 at 30 °C inlet temperature under the normal load or 9kW rack power. The cooling solution in the system level should consider 20% reduction due to the TOR and PSU.

8.2.6 Thermal Margin

The thermal margin is the difference between the maximum theoretical safe temperature and the actual temperature. Unless specified, the system should operate at an inlet temperature of 35°C (95°F) outside of the system with a minimum 4% thermal margin or 7% thermal margin for inlet temperatures up to 30°C.

8.2.7 Thermal Sensor

The maximum allowable tolerance of thermal sensors in the Yosemite V2 Platform is ±2°C.

8.2.8 System Loading

The power consumption of individual components in the system motherboard varies by use. The total power consumption of the whole Yosemite V2 Platform also may vary with use. Please see the summary below.

- System loading: idle to 100%
- Mezzanine card: 20W maximum

A unified thermal solution that can cover up to 100% system loading is preferred. However, an original design manufacturer (ODM) can propose a non-unified thermal solution if there is an alternative way to provide cost benefits. At minimum, the air-duct design should be unified for all SKUs. Further, it is required that under idle condition the mezzanine card be adequately cooled.

8.2.9 Fan Speed Controller

The fan speed controller (FSC) must be optimized to provide the necessary cooling for all key components while aiming to maximize thermal efficiency or minimize the CFM/W. The FSC may be a combination of linear, non-linear and PID control and must be set based on sensor readings for all of the key components. The overshoot of temperature should be minimized and must be less than 2°C from the stabilized temperature. The FSC must be able to maintain the thermal margin for all components while operating within the data center environmental conditions. If required, the FSC must have separate FSC tables to accommodate 0 ft, 3,000 ft and 6,000 ft elevations.

8.3 Thermal Kit Requirements

Thermal testing must be performed at up to 35°C (95°F) inlet temperature to guarantee high temperature reliability.

8.3.1 Heat Sinks

Heat sinks must have a thermally optimized design at the lowest cost. There must be no more than three heat pipes in the heat sink. Installation must be simple and uncomplicated. Heat sinks must not block debug headers or connectors.

8.3.2 System Fan

The system fan must be highly power-efficient with dual bearings. The propagation of vibration caused by fan rotation should be minimized and limited. The minimum frame size of a fan is 60mm × 60mm and the maximum frame size is 80mm × 80mm. An ODM can propose a larger frame size than 80mm × 80mm if and only if there is an alternative way to provide cost benefits. The maximum fan thickness should be less than 38mm. Each rotor in the fan should have a maximum of five wires. Except for the condition when one fan (or one rotor) fails, the fan power consumption in system should not exceed 5% of total system power, excluding the fan power.

System fans should not have backrush currents in all conditions. System fans should have an inrush current of less than 1A on a 12.5V per fan. When there is a step change on the fan PWM

signal from low PWM to high PWM, there should be less than 10% of overshoot or no overshoot for the fan input current. The system should stay within its power envelope per Open Rack V1/V2 power specification in all conditions.

9 I/O System

This section describes the Yosemite V2 Platform's I/O requirements.

9.1 PCIe Slots

The Yosemite V2 Platform can have two standard PCIe x16 cards or a customized PCIe card in 1S server form factor on two slots.

9.2 Network

9.2.1 Data Network

The Yosemite V2 Platform uses an OCP 2.0 mezzanine card on the front panel as its primary data network interface. It could be a 4x10G KR-retimer card, or a multi-host 40G/50G/100G network interface card. Please refer to Section 5 for more details.

9.2.2 Management Network

The management network on the Yosemite V2 Platform uses the sideband of the network controller of the data network, either SMBus or NC-SI interface. Please refer to Section 5 for more details.

9.3 1S Server Slots Assignment

The Yosemite V2 Platform defines 1S Server slot ID assignment and order in the table below, which is a side view of a Yosemite V2 sled.

Table 6: Slot ID Assignment and Order

Front Side	Slot 1 (Left)	Slot 2 (Left)	Rear Side
Cold Aisle	Slot 3 (Right)	Slot 4 (Right)	Hot Aisle

9.4 Front Panel

On the Adapter Card of a Yosemite V2 sled, there is a power button, a reset button, an OCP debug card and a USB port attached to the current selected 1S server. The selected server is determined by the position indicated on the selector knob. There are four blue LEDs placed on the baseboard in the same order as 1S server slots to indicate server status.

9.4.1 Selector Knob

A user can turn the selector knob to select a 1S server and a BMC. When a 1S server is selected, it owns the power button, reset button, OCP debug card and USB port on the front panel. The LED associated with the active 1S server blinks as visual feedback to the user. When a BMC is selected, all four LEDs blink as visual feedback to the user. The BMC owns the OCP debug card, but not the power button, reset button or USB port.

9.4.2 Power Button and Reset Button

A red power button and a black reset button are on the front panel. They belong to the currently selected 1S server.

When the power button is pressed for less than four seconds and then released, the currently selected 1S server receives a Power Management event. This event will power on the 1S server (if it was off). However, if the current selected 1S server is already on but a user presses the power button for more than four seconds, the 1s Server will perform a hard power off.

If the reset button switch is pressed for any duration of time, and the currently selected 1S server is on, it shall perform a hard reset.

A label on the baseboard's silkscreen will indicate the functionality of each button.

9.4.3 LED

There are four dual colored blue/yellow LEDs on the front panel. These LEDs are used to indicate power and to identify which 1S server is currently selected. These LEDs are placed in a grid (two rows of two LEDs each) and represent each 1S server's power status. The placement and silkscreen label must match the 1S server slot ID assignment.

9.4.4 USB Connector

The Yosemite V2 Platform has one USB 2.0 port located at the front panel of the baseboard. It belongs to the currently selected 1S server.

The BIOS should support the following devices attached to the USB port:

- USB Keyboard and mouse
- USB flash drive (bootable)
- USB hard drive (bootable)
- USB optical drive (bootable)

On the baseboard, a USB mux is used to connect all four 1S servers to the USB port. The BMC will control the mux based on the position of the selector knob. In addition, a BMC's virtual hub port is connected to the 1S server through a hub. This enables any 1S server to update the BMC firmware through this path.

Newly designed OCP USB3 debug card is supported on Yosemite V2 platform.

9.4.5 OCP Debug Header

A standard OCP debug header is at the front panel of the baseboard. Through this debug header, an OCP debug card can provide serial port access to 1S servers and the BMC, as well as to the POST code display. The Reset button on the OCP debug card behaves exactly like the Reset button on the front panel.

The debug header is a 14-pin, shrouded, vertical, 2mm pitch connector. Figure 9.1 is an illustration of the header. The debug card should have a key to match with the notch to avoid pin shift when plugging in.

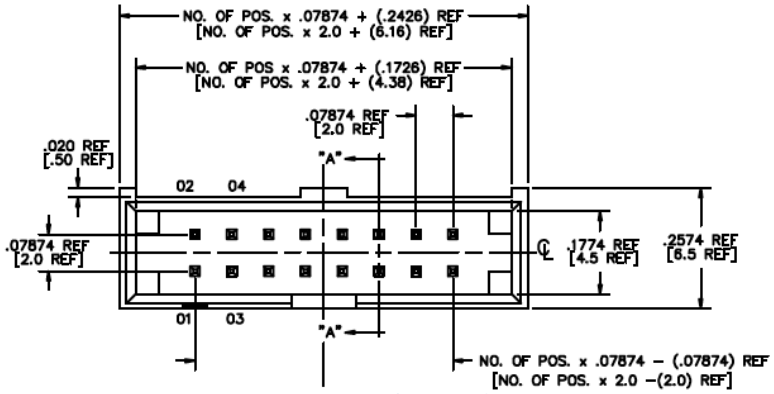


Figure 9-1: Debug Header

Table 7: Debug Header Pin Definitions

Pin (CKT)	Function
1	Low HEX Character [0] Least Significant Bit
2	Low HEX Character [1]
3	Low HEX Character [2]
4	Low HEX Character [3] Most Significant Bit
5	High HEX Character [0] Least Significant Bit
6	High HEX Character [1]
7	High HEX Character [2]
8	High HEX Character [3] Most Significant Bit
9	Serial Transmit (Motherboard Transmit)
10	Serial Receive (Motherboard Receive)
11	System Reset
12	UART Channel Selection
13	GND
14	VCC (+5VDC)

9.4.6 POST Codes

During POST, the BIOS should output POST codes onto the OCP debug card through Bridge IC and the BMC. When an SOL session is available during POST, the remote console should show the POST code.

During the boot sequence, the BIOS shall initialize and test each DIMM module. If a module fails to initialize or fails the BIOS test, the following POST codes should flash on the debug card to indicate which DIMM has failed.

Table 8: DIMM Error Code Table

	Code	Result
CPU (Channel 0 ~ 3)	A0	Channel 0 DIMM 0 (Upper furthest) Failure
	A1	Channel 0 DIMM 1 Failure
	B0	Channel 1 DIMM 0 Failure
	B1	Channel 1 DIMM 1 (Upper closest) Failure
	C0	Channel 2 DIMM 0 (Lower furthest) Failure
	C1	Channel 2 DIMM 1 Failure
	D0	Channel 3 DIMM 0 Failure
	D1	Channel 3 DIMM 1 (Lower closest) Failure

The first hex character indicates the channel of the DIMM module. The second hex character indicates the number of the DIMM module. The POST code will also display the error major code and minor code from the Intel memory reference code. The display sequence will be “00”, DIMM location, Major code and Minor code with a one-second delay for every code displayed. The BIOS shall repeat the display sequence indefinitely. The DIMM number count starts at the furthest DIMM from the CPU.

9.4.7 Serial Console

The output stage of the system’s serial console shall be contained on the debug card. The TX and RX signals from the system UART shall be brought to the debug header at the chip logic levels (+3.3V). The debug card will contain a mini-USB connector with the pin definition shown in the table below. A separate convertor is needed to provide an RS-232 transceiver and a DB9 connector.

Table 9: Debug Card Mini-USB UART Pin Definitions

Pin	Function
1	VCC (+5VDC)
2	Serial Transmit (motherboard transmit)
3	Serial Receive (motherboard receive)
4	NC
5	GND

By default, the Yosemite V2 Platform performs console redirection through the SOL. When the debug card is plugged in, debug card pin 12 shall be used to select console redirection between the SOL and the local serial port on the card, as described above.

9.5 VGA support

The Yosemite V2 Platform supports a VGA interface. The original SATA interface on 1S server interface has been repurposed to be a x1 PCIe link, which is connected to the BMC’s on-chip

VGA controller through a mux as shown below. On the baseboard, there is a mux to route the current selected 1S server's PCIe bus to this VGA controller under BMC's control. BMC turns the mux towards the current 1S server based on knob selection. Please note when the mux is switching from one to another, this is a PCIe hot-plug event to the "old" current 1S server and "new" current 1S server, as well as the BMC itself as a PCIe device. Both 1S server and BMC need to support PCIe hot-plug event to make this VGA scheme work.

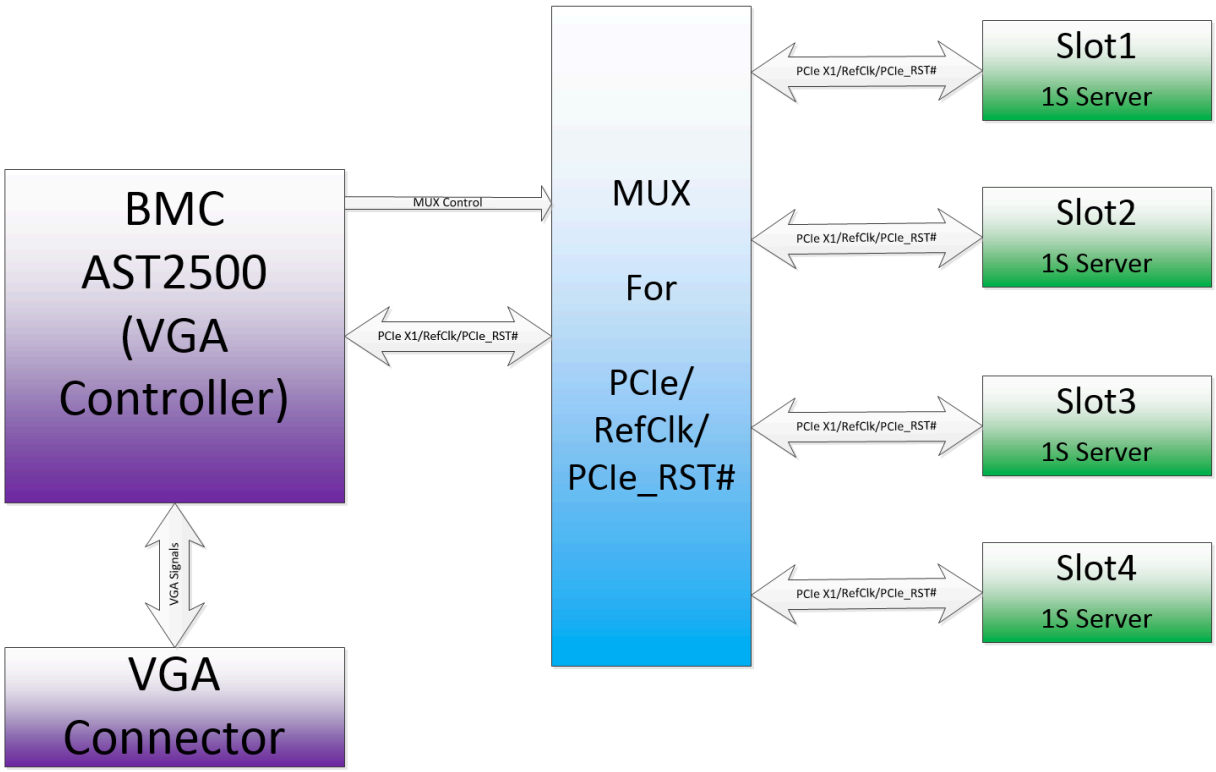


Figure 9-2: VGA support

9.6 Fan Connector

The Yosemite V2 Platform motherboard has an 8-pin fan connector that supports two fans. Every fan has its own PWM input to control the fan speed and tachometer output so that the BMC can measure the fan speed. All fans are powered by the system's 12V power supply and should be on at full speed before the BMC can control it.

Table 5: Fan Connector Pin Definition

Pin	Description
1	Second fan's PWN input
2	First fan's PWM input
3	Second fan's TACHO output
4	First fan's TACHO output

5	Second fan's power 12V
6	First fan's Power 12V
7	GND
8	GND

9.7 Power

9.7.1 Input Voltage Level

The nominal input voltage delivered by the power supply is 12.5 VDC. The voltage has a range of 11.5V to 13.5V. The motherboard shall accept and operate normally with an input voltage tolerance range between 11.25V and 13.75V.

The total power of the Yosemite V2 Platform shall be 600W or lower.

9.7.2 Capacitive Load

To ensure compatibility with the system power supply, the baseboard may not have a capacitive load greater than 4000µF. The capacitive load of the platform should not exceed the maximum value of 4000µF under any operating condition as defined in Section 10.

9.8 Hot Swap Controller Circuit

In order to have better control of the 12.5V DC power input to each platform, an HSC (ADI ADM1278) is used on the baseboard. An HSC circuit provides the following functions:

- Inrush current control when the Yosemite V2 Platform is inserted and powered up.
- Current limiting protection for over current and short circuit. The over current trip point should be able to be set to 54A.
- Safe operating area protection when the MOSFET turns on and off.
- PMBus interface to enable the following BMC actions:
    - Report server input power and log an event if it triggers the upper critical threshold.
    - Report input voltage (up to 1 decimal point) and log an event if it triggers either the lower or the upper critical threshold.
    - Log a status event based on the hot swap controller's status register.
- Provide a fast overcurrent sense alert with a resistor option to disable.

The voltage drop on the HSC current-sense resistor should be less than or equal to 25mV at full loading. The hot-swap controller should have the SMBus address set to 0x20 (7-bit format).

The power reporting of the hot-swap controller must be better than 2%, from 50W to full loading at room temperature.

9.9 1S Server Power Management

The Yosemite V2 Platform supplies single 12V power to all 1S server slots. There is a power switch for each 1S server slot under the BMC's control. These 12V power switches should be on by default unless the BMC turns them off on purpose. It is a useful feature to implement AC on, off, or cycling through the BMC.

The BMC can sample total platform power consumption from the hot-swap controller via an SMBus. As specified in the OCP 1S server specification, every 1S server shall implement a power sensor to monitor total 1S server power consumption. These power sensors are accessible to the BMC via the Bridge IC on the 1S server card. The BMC shall implement a sophisticated power management algorithm based on total platform power consumption and the power consumption of individual 1S servers.

A fast throttle feature is implemented on the platform. It enables you to throttle an individual 1S server or all 1S servers down to lowest power state in the shortest possible time period. The hot-swap controller could trigger this signal when a platform-level over-current condition happens, which will throttle down all the 1S servers. The BMC can also throttle particular 1S servers as needed.

9.10 System VRM Efficiency

High efficiency VRMs shall be used for the Yosemite V2 Platform with 91% efficiency over the 30% to 90% load range.

9.11 Power Policy

The power policy of 1S server cards on the Yosemite V2 Platform can be set by the BMC to Always On or Last Power State. When the power policy is Always On, the 1S servers will be powered on automatically regardless of their last power state. When the power policy is Last Power State, the 1S servers will restore the last power state before AC cycling.

10 Environmental Requirements and Other Regulations

10.1 Environmental Requirements

The motherboard shall meet the following environmental requirements:

- Gaseous contamination: Severity Level G1 per ANSI/ISA 71.04-1985
- Ambient operating temperature range: 0°C to +35°C
- Operating and storage relative humidity: 10% to 90% (non-condensing)
- Storage temperature range: -40°C to +70°C*
- Transportation temperature range: -40°C to +70°C (short-term storage)*

The full system shall meet the following environmental requirements:

- Gaseous contamination: Severity Level G1 per ANSI/ISA 71.04-1985
- Ambient operating temperature range: 0°C to +35°C
- Operating and storage relative humidity: 10% to 90% (non-condensing)
- Storage temperature range: -40°C to +70°C*
- Transportation temperature range: -40°C to +70°C (short-term storage)*
- Operating altitude with no de-ratings: 6000 ft

*NOTE: Liquid cooling can meet requirements in pack

10.2 Vibration and Shock

The motherboard shall meet all shock and vibration requirements according to IEC specifications IEC78-2-(*) and IEC721-3-(*) Standard & Levels. Testing requirements are listed in the table below. The motherboard shall comply fully with the specification without any electrical discontinuities during the operating vibration and shock tests. No physical damage or limitation of functional capabilities (as defined in this specification) shall occur to the motherboard during the non-operating vibration and shock tests.

Table 6: Vibration and Shock Requirements

	Operating	Non-Operating
Vibration	0.5g, 2 to 500 to 2 Hz per sweep, 10 sweeps at 1 octave/minute, test along three axes	1.2g, 2 to 500 to 2 Hz per sweep, 10 sweeps at 1 octave/minute, test along three axes (If shaker can't perform 1.2g with 2Hz, go with 5-500-5 Hz)
Shock	6g, half sine, 11ms, 5 shocks, test along three axes (Test for 7g shock only after system passes 6g shock)	12g, half sine, 11ms, 10 shocks, test along three axes

10.3 Regulations

The vendor needs to provide certification body reports of the Yosemite V2 Platform motherboard and tray at the component level.

11 Prescribed Materials

11.1 Disallowed Components

The following components are not used in the design of the motherboard:

- Components disallowed by the European Union's Restriction of Hazardous Substances Directive (RoHS 6)
- Trimmers and/or potentiometers
- Dip switches

11.2 Capacitors and Inductors

The following limitations apply to the use of capacitors:

- Only aluminum organic polymer capacitors made by high-quality manufacturers are used; they must be rated 105°C.
- All capacitors have a predicted life of at least 50,000 hours at 45°C inlet air temperature, under the worst conditions.
- Tantalum capacitors using manganese dioxide cathodes are forbidden.
- SMT ceramic capacitors with case size > 1206 are forbidden (size 1206 are still allowed when installed far from the PCB edge and with a correct orientation that minimizes the risk of cracking).
- Ceramic material for SMT capacitors must be X7R or better (COG or NP0 type are used in critical portions of the design). Only SMT inductors may be used. The use of through-hole inductors is disallowed.

11.3 Component De-rating

For inductors, capacitors, and FETs, de-rating analysis is based on at least 20% de-rating.

12 Labels and Markings

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

Table 7: Lables and Markings

Description	Type	Barcode Required?
Safety Markings	Silkscreen	No
Vendor P/N, S/N, REV (Revision would increment for any approved changes)	Adhesive label	Yes
Vendor Logo, Name & Country of Origin	Silkscreen	No
PCB Vendor Logo, Name	Silkscreen	No
Date Code (Industry Standard: Week / Year)	Adhesive label	Yes
RoHS Compliance	Silkscreen	No
WEEE Symbol. The motherboard will have the crossed out wheeled bin symbol to indicate that the manufacturer will take it back at the end of its useful life. This is defined in the European Union Directive 2002/96/EC of January 27, 2003 on Waste Electrical and Electronic Equipment (WEEE) and any subsequent amendments.	Silkscreen	No
CE Marking	Silkscreen	No
UL Marking	Silkscreen	No

13 Revision History

Author	Description	Revision	Date
Yan Zhao	▪ Initial draft	0.1	12/5/2016
Jon Ehlen Jarrod Chow Sai Dasari Yan Zhao	▪ Updated mechanical, thermal, BMC and other sections	0.2	12/15/2016
Yan Zhao	▪ Updated spec with received comments.	0.3	1/12/2017
Jarrod Chow Yan Zhao	▪ Updated thermal and environmental requirements sections. ▪ Updated VGA support section. ▪ Changed license to OCPHL-P. ▪ Incorporated review comments from various sources.	0.4	2/1/2017