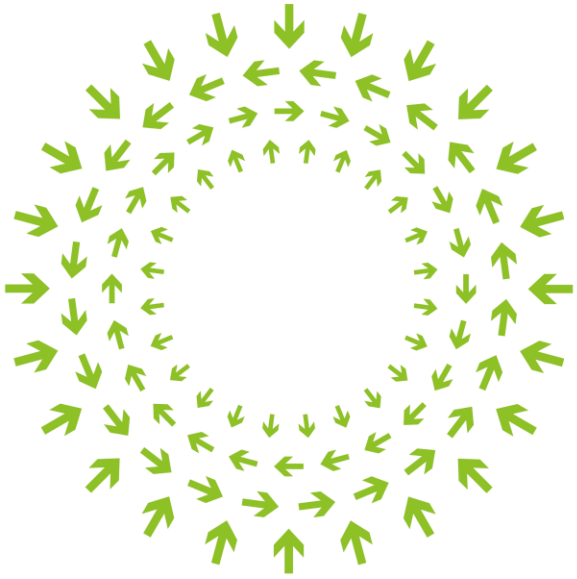


OPEN

Compute Project

Project Olympus Intel® Xeon® Scalable Processor 3U Server Specification

Rev 1.0

Author: Mark D. Chubb

License

Contributions to this Specification are made under the terms and conditions set forth in Open Compute Project Contribution License Agreement (“OCP CLA”) (“Contribution License”) by: **ZT Group Int’l, Inc, dba ZT Systems**

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

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

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

Note: The following clarifications, which distinguish technology licensed in the Contribution License and/or Specification License from those technologies merely referenced (but not licensed), were accepted by the Incubation Committee of the OCP:

NOTWITHSTANDING THE FOREGOING LICENSES, THIS SPECIFICATION IS PROVIDED BY OCP "AS IS" AND OCP EXPRESSLY DISCLAIMS ANY WARRANTIES (EXPRESS, IMPLIED, OR OTHERWISE), INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, FITNESS FOR A PARTICULAR PURPOSE, OR TITLE, RELATED TO THE SPECIFICATION. NOTICE IS HEREBY GIVEN, THAT OTHER RIGHTS NOT GRANTED AS SET FORTH ABOVE, INCLUDING WITHOUT LIMITATION, RIGHTS OF THIRD PARTIES WHO DID NOT EXECUTE THE ABOVE LICENSES, MAY BE IMPLICATED BY THE IMPLEMENTATION OF OR COMPLIANCE WITH THIS SPECIFICATION. OCP IS NOT RESPONSIBLE FOR IDENTIFYING RIGHTS FOR WHICH A LICENSE MAY BE REQUIRED IN ORDER TO IMPLEMENT THIS SPECIFICATION. THE ENTIRE RISK AS TO IMPLEMENTING OR OTHERWISE USING THE SPECIFICATION IS ASSUMED BY YOU. IN NO EVENT WILL OCP BE LIABLE TO YOU FOR ANY MONETARY DAMAGES WITH RESPECT TO ANY CLAIMS RELATED TO, OR ARISING OUT OF YOUR USE OF THIS SPECIFICATION, INCLUDING BUT NOT LIMITED TO ANY LIABILITY FOR LOST PROFITS OR ANY CONSEQUENTIAL, INCIDENTAL, INDIRECT, SPECIAL OR PUNITIVE DAMAGES OF ANY CHARACTER FROM ANY CAUSES OF ACTION OF ANY KIND WITH RESPECT TO THIS SPECIFICATION, WHETHER BASED ON BREACH OF CONTRACT, TORT (INCLUDING NEGLIGENCE), OR OTHERWISE, AND EVEN IF OCP HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Table of Contents

Table of Figures

Table of Tables

Scope

This document defines the technical specifications for the Project Olympus Intel Xeon Scalable 3U PCIe Expansion Server. The ZT model number for this system is the XPO200 3UN, which will be used to identify the 3U Server throughout this Specification.

This XPO200 3UN 3U Server implementation is compatible with existing Project Olympus building blocks. Other Project Olympus elements which are utilized in this system are outlined in their respective specifications. These elements include, but are not limited to, the Power Supply Unit (PSU), Power Management Distribution Unit (PMDU), Universal Motherboard, Server Rack, and Rack Manager (RM). Specifications for the overall Project Olympus Rack are posted here…

http://www.opencompute.org/wiki/Server/ProjectOlympus

Table 1: Project Olympus Specifications

Specification title	Description
Project Olympus Server Rack Specification	Describes the mechanical rack hardware used in the system
Project Olympus Server Mechanical Specification	Describes the mechanical structure for the server used in the system.
Project Olympus Universal Motherboard Specification	Describes the server motherboard general requirements.
Project Olympus PSU Specification	Describes the Power Supply Unit (PSU) used in the server
Project Olympus Power Management Distribution Unit Specification	Describes the Power Management Distribution Unit (PMDU).
Project Olympus Rack Manager Specification	Describes the Rack Manager PCBA used in the PMDU.

Overview

This specification defines a Project Olympus based 2 Socket, 3U Server (XPO200 3UN) that is intended to support multiple single-wide and double-wide PCIe Cards. This server design is based on the Project Olympus 3U Server Base Specifications and fits within the Project Olympus OCP framework. The OCP Approved Mount Olympus Motherboard based on the Intel Purley platform can be used in this system along with various PCIe Riser Options to provide an IO rich platform for up to 6 x Double-wide Full-Length Cards plus 1 x Single-wide Half-Length Card or up to 12 x Single-wide Full-Length Cards plus 1 x Single-wide Half-Length Card. Due to the increased power requirements in this XPO200 3UN Server, three of the Project Olympus 3-phase 1kW Power Supply Silver Boxes will also be used in this system. Each Silver Box includes 3 internal Power Supply Units (PSUs) providing a total of 9 PSUs in an 8 + 1 redundancy scheme. Each Project Olympus 1kW Power Supply will be stacked vertically in the 3U Chassis allowing use of the Project Olympus vertical Power Managed Distribution Unit (PMDU).

1. Rack Compatibility

This XPO200 3UN PCIe Expansion Server is compatible with the Project Olympus 19" EIA310-D Rack. Due to the Server's weight, it requires custom slide rails, which allow for smoother travel during Serviceability when sliding in and out of the Rack.

2. System Requirements

The system requires a Mount Olympus Intel Purley generation Motherboard, which can be combined with several different PCIe Riser Options for system configuration flexibility. It includes a Power Distribution Board (PDB) to allow load sharing of the 3 Project Olympus 1kW PSUs. These components reside within a 3U chassis and are compatible with the Project Olympus ecosystem.

2.1 Server Configurations

Below are the targeted XPO200 3UN Compute Node Server SKUs known at the time of this System Architecture Specification release. All XPO200 3UN Server Configurations are subject to change based on Customer requirements. Additional Configurations may be added at a later time based on Customer requirements.

Table 2: Supported System Configuration Options

Feature	Qty	Description
Chassis	1	3U, 19" EIA310-D Compliant supporting Project Olympus PMDU connections
Motherboard	1	Mount Olympus 2-Socket Intel Xeon Scalable Motherboard (Purley)
Processor	2	Intel® Xeon® Platinum 8168 processor (24 core, 2.7 Ghz, 205W)
Memory	12	32GB DDR4, DR, 2667 R-DIMMs; Total System Memory: 384GB
PCIe Riser 3 & 5	2	4-Slot Active PCIe x16 Riser Card with 96-lane PCIe Switch
PCIe Riser 4	1	5-Slot Active PCIe x16 Riser Card with 96-lane PCIe Switch
GPU Card	12	Nvidia Tesla T4 GPU, LP, 75W PCIe x16 Card
Ethernet	1	10G Single port SFP+ PCIe 2.0 x8 5GT/s
HDD/SSD	1	M.2 960GB NVMe SSD, PCIe x4 110mm (sourced from CPU)
Security	1	TPM2.0 SPI Module
System Fans	6	60mmx56mm Dual Rotor Fans
Power Supply	3	Project Olympus 1020W 3-Phase, non-LES PSU
Power Distribution	1	PDB and Cable Harnesses to support 12V Power to MB, Risers, PCIe Cards, and System Fans

2.2 Motherboard

The motherboard used in the XPO200 3UN System is the Mt. Olympus Universal Motherboard (MB), which is the computational element in the Project Olympus Server. This is a Dual Socket Purley generation server board defined by Microsoft.

NOTE: The information in this section is for reference only. For the latest detailed Spec information please refer to the Mt Olympus MB Spec posted here…

http://www.opencompute.org/wiki/Server/ProjectOlympus

- Processor:
  - 2 Socket Spread Core Design using Xeon (Skylake-SP/Cascade Lake-SP) Processors
  - Supports up to 205W TDP
  - Includes 1U Remote Heatsink
- Memory: 24 DDR4 DIMMs, 2 DIMMs per Channel
- PCH: Lewisburg
- PCIe Slots/Connectors:
  - PCIe x8 Connector (Slot #1) – CPU0
  - PCIe x8 Connector (Slot #2) – CPU0
  - PCIe x16 Riser Connector (Slot #3) – CPU0
  - PCIe x16 Riser Connector (Slot #4) – CPU1
  - PCIe x16 Riser Connector (Slot #5) – CPU1
  - PCIe x4 M.2 Connector (M.2 #1) – PCH
  - PCIe x4 M.2 Connector (M.2 #2) – PCH
  - PCIe x4 M.2 Connector (M.2 #3) – CPU1
  - PCIe x4 M.2 Connector (M.2 #4) – CPU1
  - PCIe x8 OCuLink Connector – CPU1
- SATA Connectors:
  - 4 x SATA 7-pin Connectors (SATA[3:0]) – PCH
  - 2 x SATA MiniSAS HD Connectors (SATA[7:4]/PCIe[15:12] & SATA[11:8]/PCIe[19:16]) – PCH
- BMC: ASPEED AST1250
  - All PCIe Slots are connected to BMC I2C Buses for PCIe Card telemetry
  - I2C MUXes are used to avoid I2C Address contention
- Security: SPI TPM2.0 Module
- Front IO Ports:
  - 1 x BMC Dedicated Management NIC Port
  - 2 x USB3.0 Ports
  - Power Button (Pre-Production Only)
  - Reset Button (Pre-Production Only)
  - 1 x Video Port (EMPTY)
  - 1 x 10GbE SFP+ Connector (EMPTY)

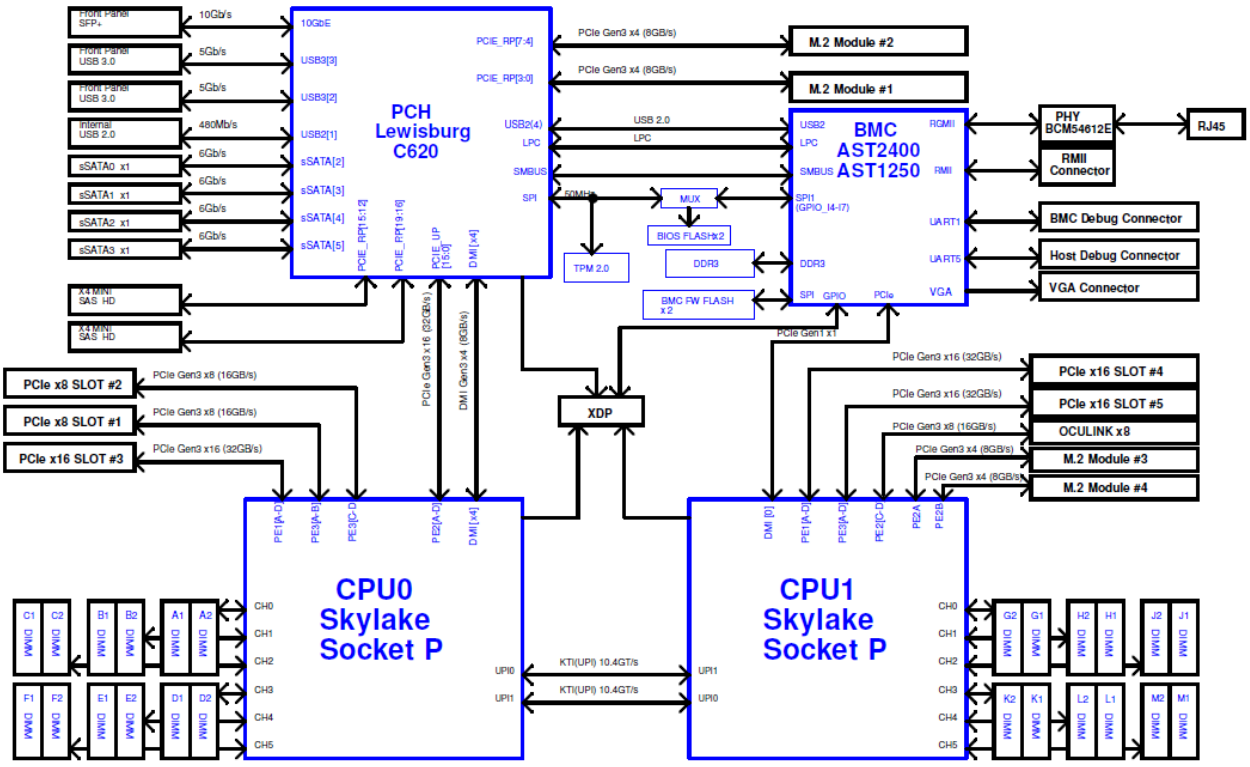


Figure 1: Motherboard Block Diagram

2.3 PCIe Configuration

2.3.1 PCIe Port Mapping

The PCIe ports from each processor are mapped as shown in the Figure and Table below:

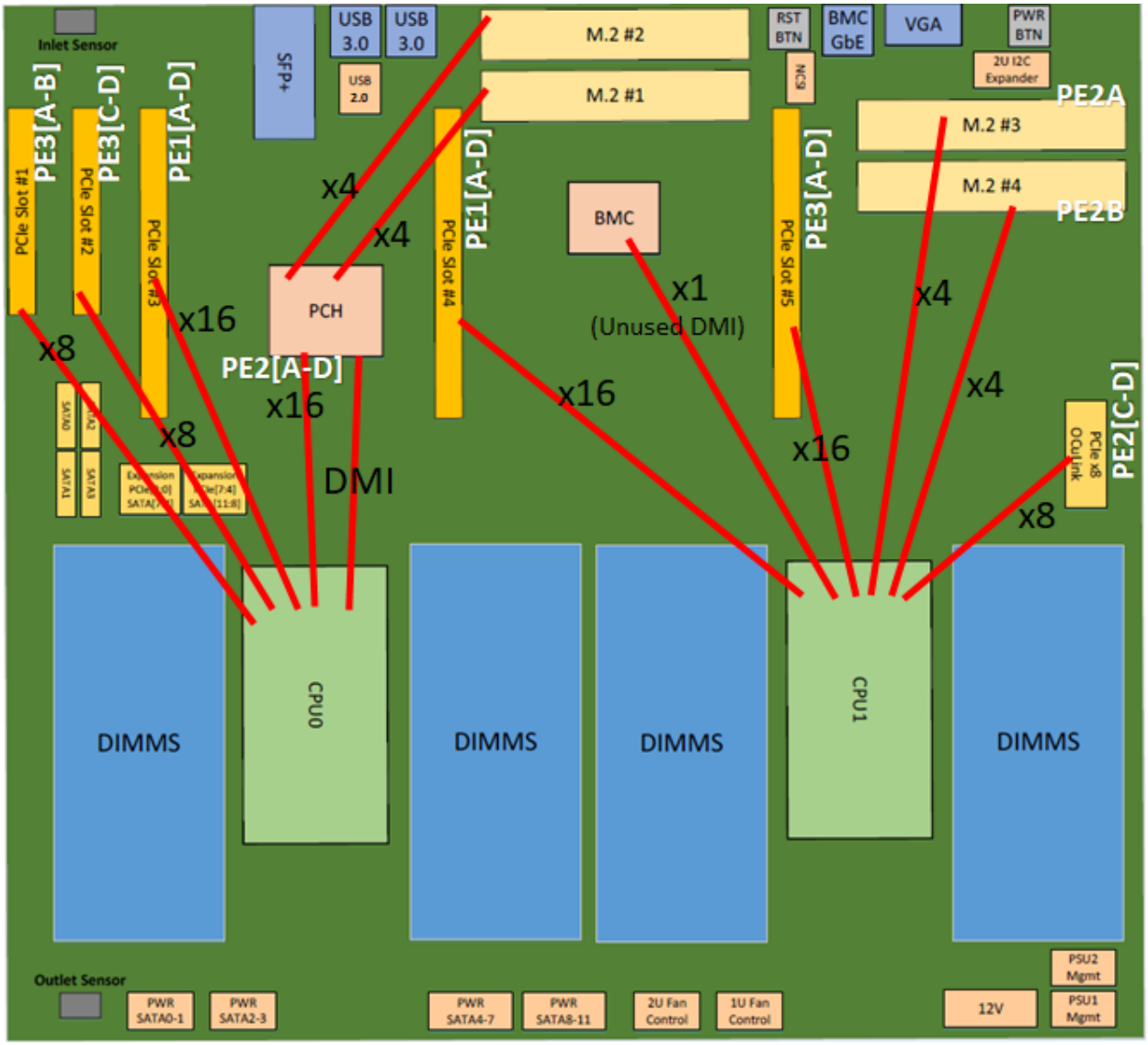


Figure 2: Motherboard PCIe Mapping Layout

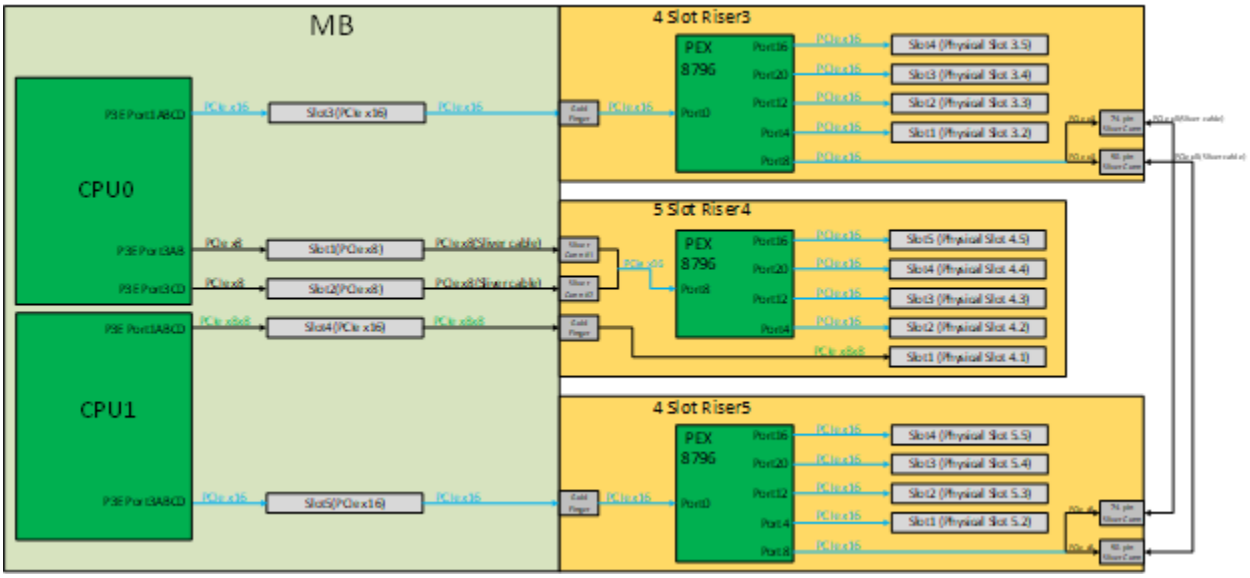
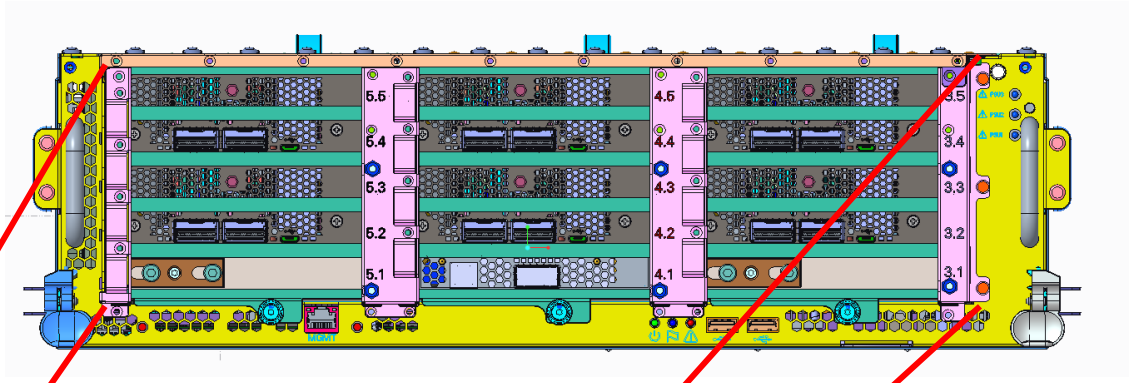


Figure 3: XPO200 3UN PCIe Bus Connectivity for 5-Slot Active & 4-Slot Active Riser Config

2.3.2 PCIe Card Physical Numbering

Below is the PCIe Card Chassis Level Numbering scheme for the XPO200 3UN System configurations as viewed from the front of the Chassis.



4-Slot Active Riser	5-Slot Active Riser	4-Slot Active Riser
PCIe Card#5.5	PCIe Card#4.5	PCIe Card#3.5
PCIe Card#5.4	PCIe Card#4.4	PCIe Card#3.4
PCIe Card#5.3	PCIe Card#4.3	PCIe Card#3.3
PCIe Card#5.2	PCIe Card#4.2	PCIe Card#3.2
PCIe Card#5.1 (Invalid)	PCIe Card#4.1	PCIe Card#3.1 (Invalid)

Double-Wide capable slots in Blue
Single-Wide only slots in Green

Figure 4: Physical Card Numbering for 13 Single-Wide PCIe Cards

2.4 4-Slot Active x16 PCIe Riser Card

This is a PCIe x16 Active Riser Card with Broadcom/PLX 96-lane Gen3 PCIe Switch to 4 x16 PCIe Slots. There can be two 4-Slot Active Risers per system, each plugged into MB Riser Slots #3 & #5.

2.4.1 Major Component Placement

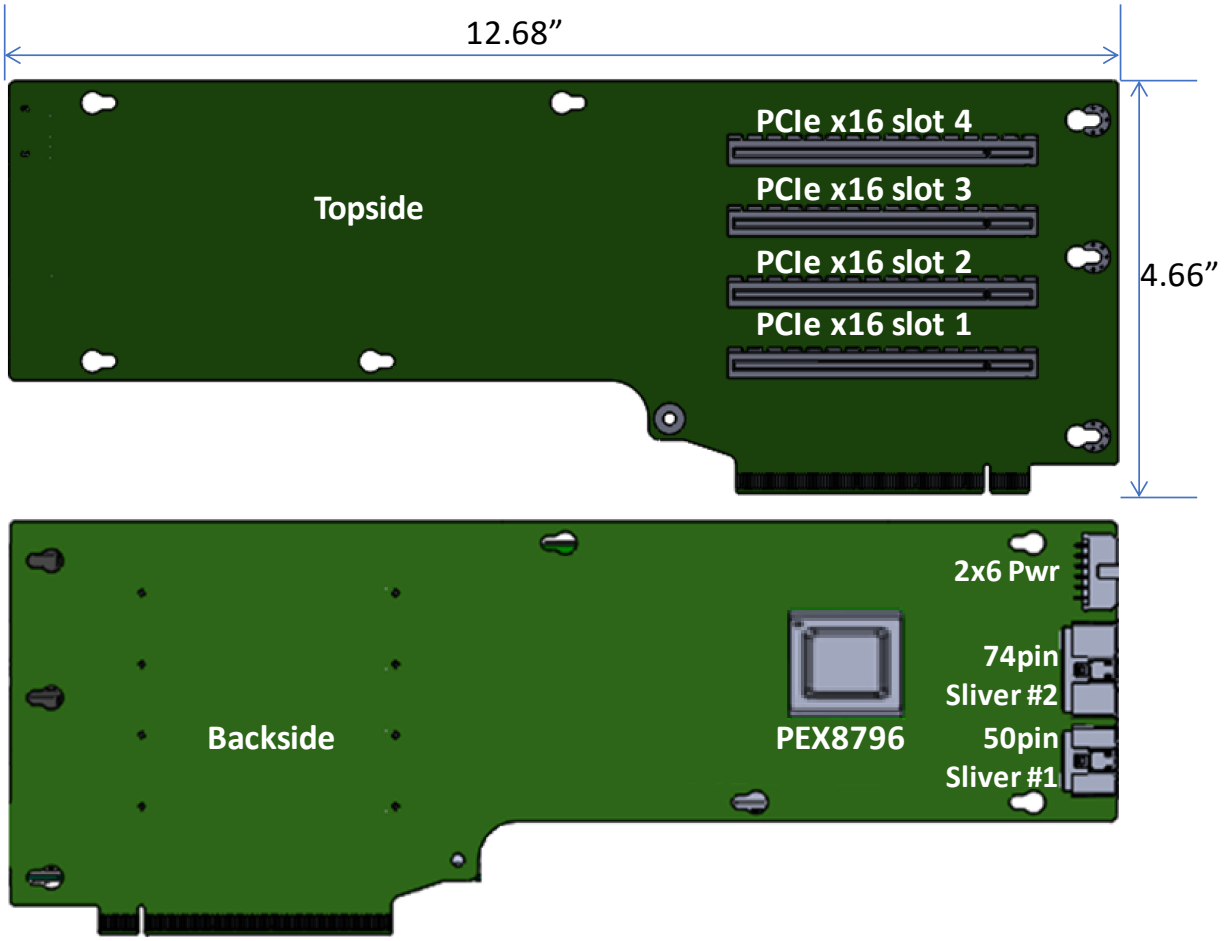


Figure 5: 4-Slot Active Riser Layout (Top and Backside)

2.4.2 Block Diagram

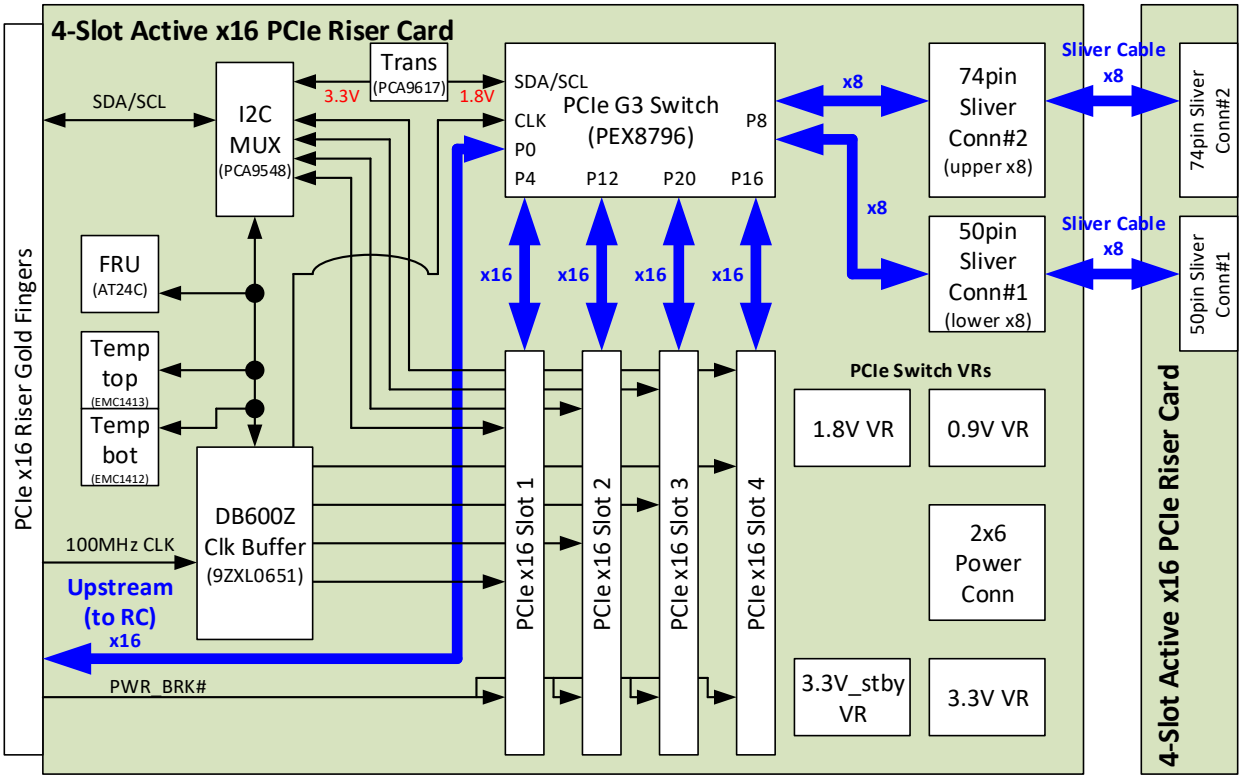


Figure 6: 4-Slot Active Riser Block Diagram

2.4.3 PCIe Riser Configuration

- 4 x PCIe x16 slots supporting (Slots #1-4, Bottom to Top):
  - 4 x FHFL Single-Wide PCIe cards OR
  - 2 x FHFL Double-Wide PCIe cards
  - Requires PWRBRK_N (pin B30) signal from MB CPLD routed to each Slot for High Power PCIe Card throttling
    - This feature can be used for system power capping
    - This signal is typically only supported on High Power PCIe Cards (i.e. Cards greater than 75W)
- PCIe Port Mapping and Routing Layers are shown in the figure below. PCIe Connectors (not shown) are on Topside:

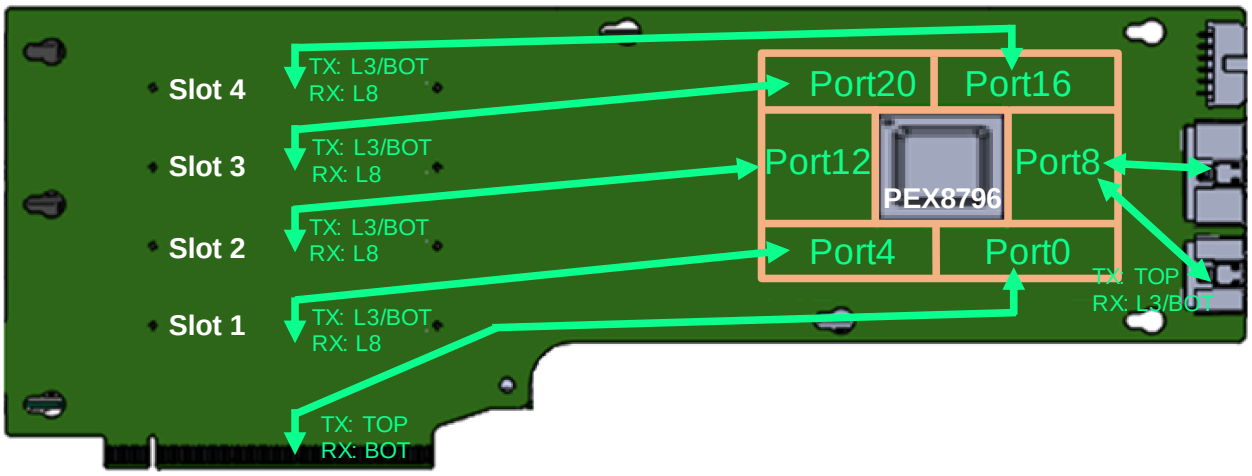


Figure 7: 4-Slot Active Riser PCIe Port Mapping and Routing Layers (View from Backside)

Table 3: 4-Slot Active Riser Stack-up

Dielectric and Copper Properties											
Layer			Copper		Dielectric Material					Layer	
Name	Description	Type	Weight	Type	Type	Er	Df	Actual Dk	Actual Df/1MH	Design Thickness (mil)	Actual Thickness (mil)
Top_SM	Top Soldermask	Soldermask			Soldermask	3.5	0.030	3.70		0.50	0.50
L1top	Signal_L1top	Signal	0.5 oz. plated	STD						1.90	2.10
PP_L1top_L2	Prepreg	PP			1080*1	3.6	0.009	3.50	0.007	2.70	2.72
L2	GND_L2	GND	0.5 oz.	RTF						0.65	0.66
Core_L2_L3	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L3	Signal_L3	Signal	0.5 oz.	RTF						0.65	0.66
PP_L3_L4	Prepreg	PP			1080*2	3.6	0.009	3.50	0.007	5.50	5.11
L4	GND_L4	GND	0.5 oz.	RTF						0.65	0.66
Core_L4_L5	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L5	Signal_L5	Signal	0.5 oz.	RTF						0.65	0.66
PP_L5_L6	Prepreg	PP			7628+1080	3.6	0.009	3.60	0.007	9.50	9.61
L6	PWR_L6	PWR	1.0 oz.	RTF						1.30	1.25
Core_L6_L7	Core	Core			2116X1	3.8	0.009	3.60	0.007	4.00	4.20
L7	PWR_L7	PWR	1.0 oz.	RTF						1.30	1.25
PP_L7_L8	Prepreg	PP			7628+1080	3.6	0.009	3.60	0.007	9.50	9.61
L8	Signal_L8	Signal	0.5 oz.	RTF						0.65	0.66
Core_L8_L9	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L9	GND_L9	GND	0.5 oz.	RTF						0.65	0.66
PP_L9_L10	Prepreg	PP			1080*2	3.6	0.009	3.50	0.007	5.50	5.11
L10	Signal_L10	Signal	0.5 oz.	RTF						0.65	0.66
Core_L10_L11	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L11	GND_L11	GND	0.5 oz.	RTF						0.65	0.66
PP_L11_L12bot	Prepreg	PP			1080*1	3.6	0.009	3.50	0.007	2.70	2.72
L12bot	Signal_L12bot	Signal	0.5 oz. plated	STD						1.90	2.10
Bot_SM	Bot Soldermask	Soldermask			Soldermask	3.5	0.030	3.70		0.50	0.50
					with soldermask, total thickness:					64.00	64.04
					without soldermask, total thickness:					63.00	63.04

- Broadcom/PLX 96-Lane PCIe Gen3 Switch (PEX8796)

- Switch Bifurcated into 5 x16 Ports (1 x16 Port to each x16 Slot and 1 x16 Port to Sliver conns)
- Depending on the PCIe Card requirement, the Switch can be bifurcated into different lane width configurations. Custom Switch Firmware (FW) is required to change the Switch bifurcation.
- Switch is located on Backside as required by Chassis Thermals
- PCIe Switch Thermal Design Power Requirements:
  - 2 x Double-Wide Card Config can use up to 64-lanes and requires 32.5W
  - 4 x Single-Wide Card Config can use up to 96-lanes and requires 35.8W
  - Heatsink is required
- Please refer to the PEX8796 Datasheet and User Manual for additional design requirements

- 2 x PCIe x8 Sliver connectors supporting:
  - PCIe x16 Peer-to-Peer (P2P) connectivity between two PCIe Switches on two different 4-Slot Risers (Risers #3 & #5)
  - P2P connectivity supports direct transactions between PCIe Switches/PCIe Cards without using the CPU Root complex to move the data or store in main memory, thus reducing overall PCIe communication latency.
  - 2 different Sliver Connectors/Cables are used from TE Connectivity:
    - 50 pin Connector/Cable supports lower x8 PCIe lanes
    - 74 pin Connector/Cable supports upper x8 PCIe lanes
    - Different Sliver Connector sizes are to ensure proper cabling from Riser to Riser
    - 85-ohm characteristic impedance

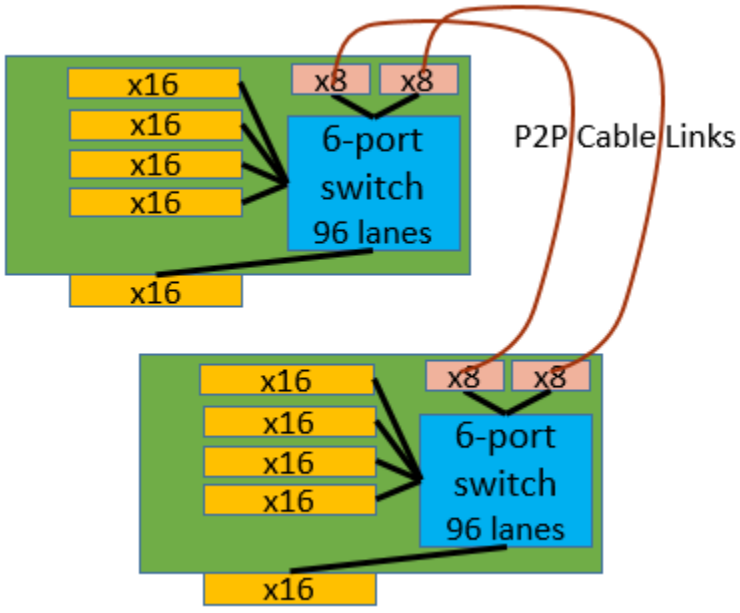


Figure 8: 4-Slot Active Riser Peer-to-Peer Cable Connectivity

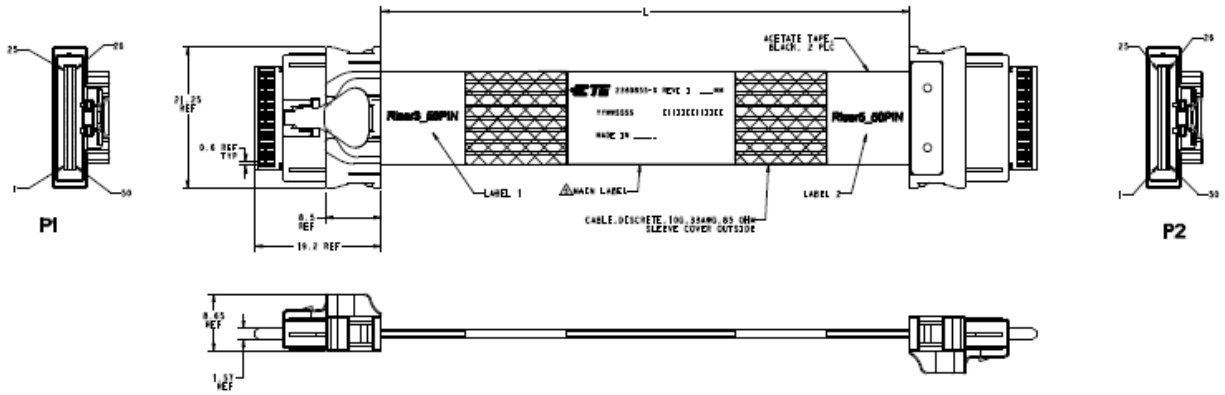


Figure 9: 50 pin Sliver Cable

Table 4: 50 pin Sliver Cable Wire Mapping

	P1		P2
	1	GND	50
PAIR 1	2	◁⟷▷	49
	3	◁⟷▷	48
	4	GND	47
PAIR 2	5	◁⟷▷	46
	6	◁⟷▷	45
	7	GND	44
PAIR 3	8	◁⟷▷	43
	9	◁⟷▷	42
	10	GND	41
PAIR 4	11	◁⟷▷	40
	12	◁⟷▷	39
	13	GND	38
PAIR 5	14	◁⟷▷	37
	15	◁⟷▷	36
	16	GND	35
PAIR 6	17	◁⟷▷	34
	18	◁⟷▷	33
	19	GND	32
PAIR 7	20	◁⟷▷	31
	21	◁⟷▷	30
	22	GND	29
PAIR 8	23	◁⟷▷	28
	24	◁⟷▷	27
	25	GND	26

	P1		P2
	26	GND	25
PAIR 9	27	◁⟷▷	24
	28	◁⟷▷	23
	29	GND	22
PAIR 10	30	◁⟷▷	21
	31	◁⟷▷	20
	32	GND	19
PAIR 11	33	◁⟷▷	18
	34	◁⟷▷	17
	35	GND	16
PAIR 12	36	◁⟷▷	15
	37	◁⟷▷	14
	38	GND	13
PAIR 13	39	◁⟷▷	12
	40	◁⟷▷	11
	41	GND	10
PAIR 14	42	◁⟷▷	9
	43	◁⟷▷	8
	44	GND	7
PAIR 15	45	◁⟷▷	6
	46	◁⟷▷	5
	47	GND	4
PAIR 16	48	◁⟷▷	3
	49	◁⟷▷	2
	50	GND	1

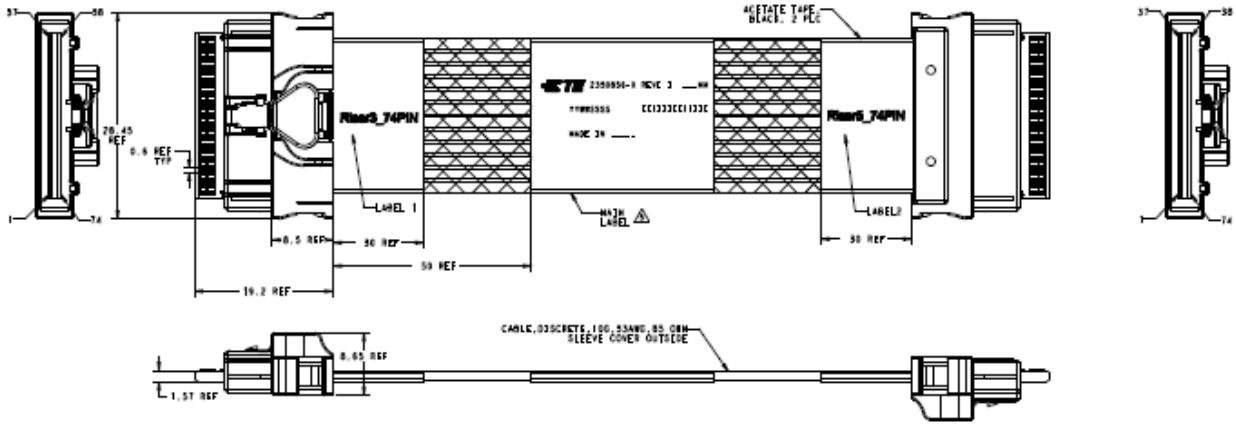


Figure 10: 74 pin Sliver Cable

Table 5: 74 pin Sliver Cable Wire Mapping

	P1		P2
	1	GND	74
PAIR 1	2	◁—▷	73
	3	◁—▷	72
	4	GND	71
PAIR 2	5	◁—▷	70
	6	◁—▷	69
	7	GND	68
PAIR 3	8	◁—▷	67
	9	◁—▷	66
	10	GND	65
PAIR 4	11	◁—▷	64
	12	◁—▷	63
	13	GND	62
PAIR 5	14	◁—▷	61
	15	◁—▷	60
	16	GND	59
PAIR 6	17	◁—▷	58
	18	◁—▷	57
	19	GND	56
PAIR 7	20	◁—▷	55
	21	◁—▷	54
	22	GND	53
PAIR 8	23	◁—▷	52
	24	◁—▷	51
	25	GND	50
PAIR 9	26	◁—▷	49
	27	◁—▷	48
	28	GND	47
PAIR 10	29	◁—▷	46
	30	◁—▷	45
	31	GND	44
PAIR 11	32	◁—▷	43
	33	◁—▷	42
	34	GND	41
PAIR 12	35	◁—▷	40
	36	◁—▷	39
	37	GND	38

	P1		P2
	38	GND	37
PAIR 13	39	◁—▷	36
	40	◁—▷	35
	41	GND	34
PAIR 14	42	◁—▷	33
	43	◁—▷	32
	44	GND	31
PAIR 15	45	◁—▷	30
	46	◁—▷	29
	47	GND	28
PAIR 16	48	◁—▷	27
	49	◁—▷	26
	50	GND	25
PAIR 17	51	◁—▷	24
	52	◁—▷	23
	53	GND	22
PAIR 18	54	◁—▷	21
	55	◁—▷	20
	56	GND	19
PAIR 19	57	◁—▷	18
	58	◁—▷	17
	59	GND	16
PAIR 20	60	◁—▷	15
	61	◁—▷	14
	62	GND	13
PAIR 21	63	◁—▷	12
	64	◁—▷	11
	65	GND	10
PAIR 22	66	◁—▷	9
	67	◁—▷	8
	68	GND	7
PAIR 23	69	◁—▷	6
	70	◁—▷	5
	71	GND	4
PAIR 24	72	◁—▷	3
	73	◁—▷	2
	74	GND	1

2.4.4 PCIe Riser Power Requirements

- Each PCIe slot supports up to 75W as defined for a standard PCIe x16 card.

- A 12V Auxiliary connector (2x6 pin cabled from PDB) is required on the Riser to support PCIe Slot power, PEX8796 VRs, 3.3V VR, and 3.3V_Stby VR.
    - The required 12V Current for this Riser is ~30A (assume 35A for conn derating).
    - 12V AUX Connector uses 5 x 12V pins: 35A/5 pins = 7A per pin.
    - 12V AUX Connector uses 1 x 12V_PSU pin to generate 3.3V_Stby Power.
- Additional 12V power required by high power PCIe cards will cable directly from the PDB to the cards themselves.
- Different Wire Harnesses can be used depending on the PCIe Card's 12V Auxiliary Power connector
- 3.3V Slot power is required for all PCIe Cards:
    - Per the PCIe Spec, 3.3V requires 3.0A per slot (4 x 3A = 12A), which cannot be pulled from only three Riser Slot 3.3V Pins.
    - 3.3V must be sourced from an on-board VR
- 3.3V_Stby may be required by some PCIe Cards:
    - Per the PCIe Spec, 3.3V_Stby requires 375mA per slot (4 x 375ma = 1.5A), which cannot be pulled from only a single Riser Slot 3.3V_Stby pin.
    - A small VR (generated from the 12V_PSU rail) will be added to the design to support the required current.

Table 6: 4-Slot Active Riser Power Budget

4-Slot Active Riser 12V Budget	**Voltage (V)**	**Current (A)**	**Qty**	**Total Current (A)**	**Power (W)**	**VR eff**	**12V Power Req'd (W)**	
PCIe 12V	12	5.5	4	22	264	100%	264	
PCIe 3.3V	3.3	3	4	12	39.6	90%	44	
DB600Z 3.3V	3.3	0.081	1	0.081	0.2673	90%	0.30	
Misc Logic/Components	3.3	0.1	1	0.1	0.33	90%	0.37	
PEX8796 VDD09	0.9		1		25.53	90%	28.36	
PEX8796 VDD09A	0.9		1		7.97	90%	8.86	
PEX8796 VDD18A	1.8		1		2.24	85%	2.63	**12V Current (A)**
PEX8796 VDD18	1.8		1		0.04	85%	0.05	
						Totals	**348**	**29**

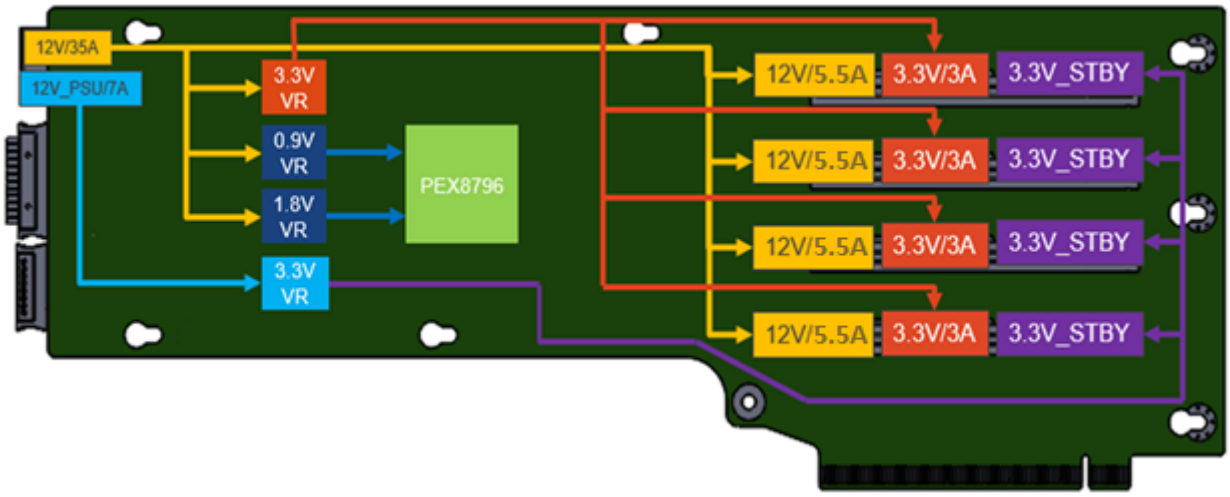


Figure 11: 4-Slot Active Riser Power Delivery Diagram

2.4.5 PCIe Riser I2C Requirements

All PCIe Connectors connect to the BMC I2C for PCIe Card telemetry. Use the PCIe designated pins for I2C (B5 & B6). To avoid I2C Address contention between PCIe Cards, an I2C MUX is needed on each riser. The design must add appropriate voltage translation/isolation between different I2C voltage domains.

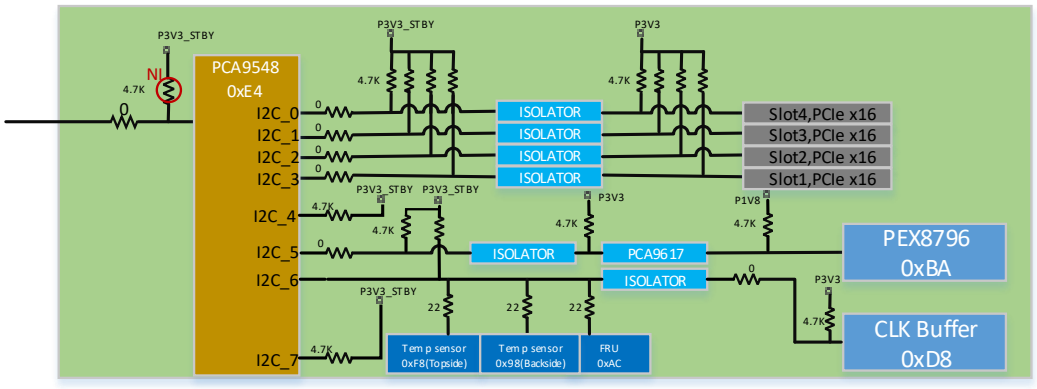


Figure 12: I2C Block Diagram for 4-Slot Active PCIe Riser

2.4.6 PCIe Riser Clock Requirements

A Differential Clock Buffer (DB600Z) is used to fanout additional PCIe Clocks. The Clock Buffer is set to PLL Bypass Mode by default. Recommend including resistor stuffing Options to change the PLL Bandwidth Mode if needed to improve Clock Jitter performance:

- PLL High BW Mode
- PLL Low BW Mode
- PLL Bypass Mode (default)

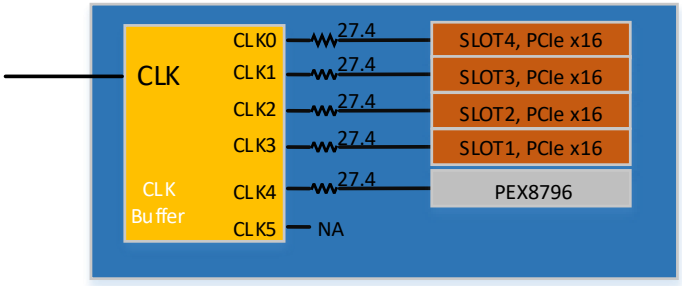


Figure 13: Clock Block Diagram for 4-Slot Active PCIe Riser

2.5 5-Slot Active x16 PCIe Riser Card

This is an active x16 Riser Card with 5 x16 PCIe Slots supported in Riser Slot #4. The Bottom x16 PCIe Slot is sourced from the x16 Riser gold-fingers and 4 additional x16 PCIe Slots are sourced from a Broadcom/PLX 96-lane Gen3 PCIe Switch. The PCIe Switch is, in turn, sourced from 2 x8 Sliver Connectors, which are cabled to 2 x8 PCIe Slots on the MB.

2.5.1 Major Component Placement

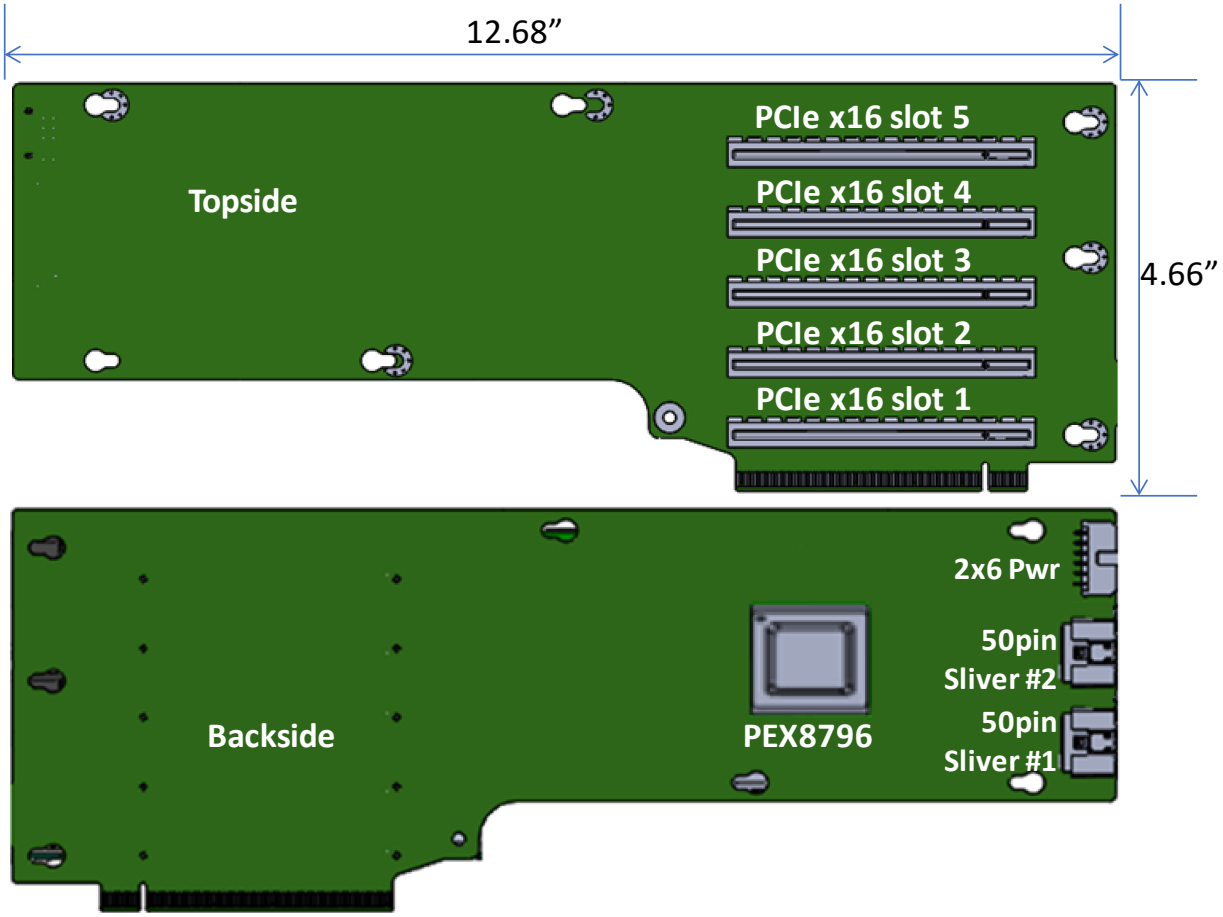


Figure 14: 5-Slot Active Riser Layout (Top and Backside)

2.5.2 Block Diagram

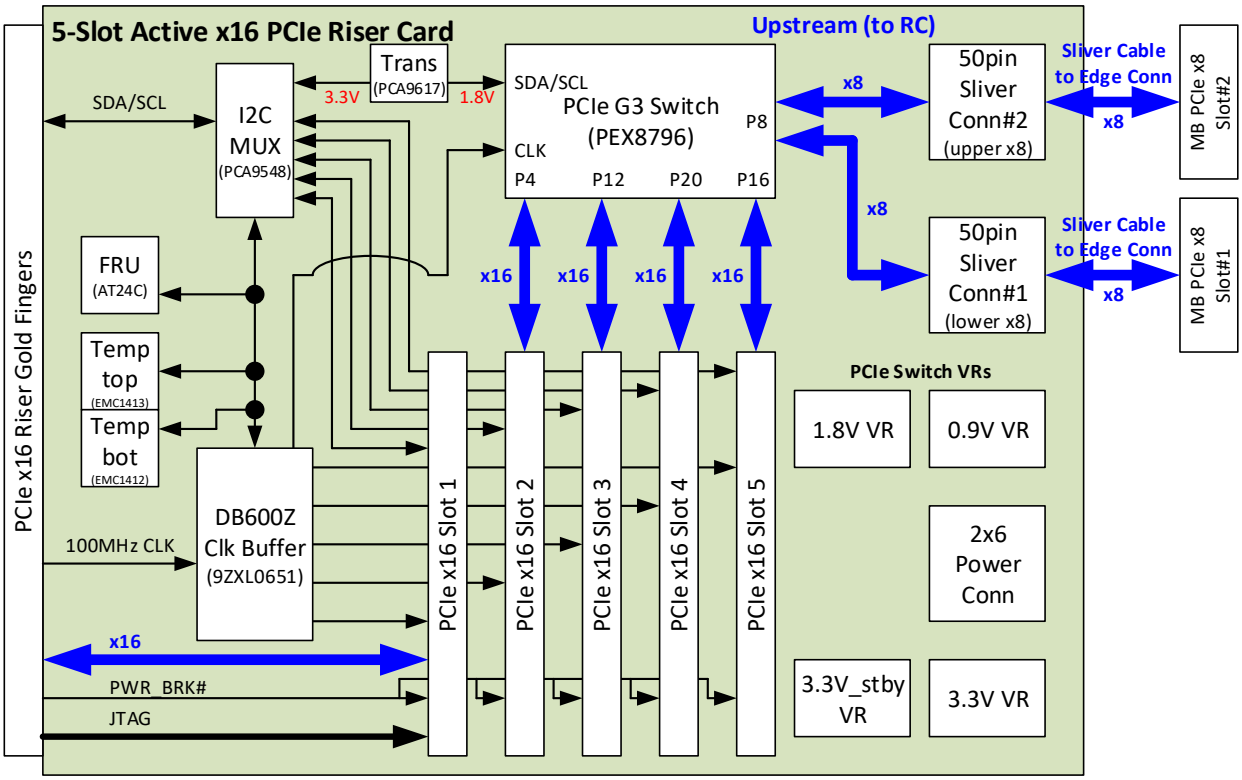


Figure 15: 5-Slot Active Riser Block Diagram

2.5.3 PCIe Riser Configuration

- 1 x PCIe x16 slot supporting (Slot #1, Bottom):
  - 1 x FHHL Single-Wide PCIe card
  - Designated for a Network Card
  - JTAG signals routed through PCIe JTAG pins of Slot #1 for debug purposes
  - All x16 PCIe lanes are sourced from MB PCIe Riser Slot #4
- 4 x PCIe x16 slots supporting (Slots #2-5, Bottom to Top):
  - 4 x FHFL Single-Wide PCIe cards OR
  - 2 x FHFL Double-Wide PCIe cards
  - Requires PWRBRK_N (pin B30) signal from MB CPLD routed to each Slot for High Power PCIe Card throttling
    - This feature can be used for system power capping
    - This signal is typically only supported on High Power PCIe Cards (i.e. Cards greater than 75W)
- PCIe Port Mapping and Routing Layers are shown in the figure below. PCIe Connectors (not shown) are on Topside:

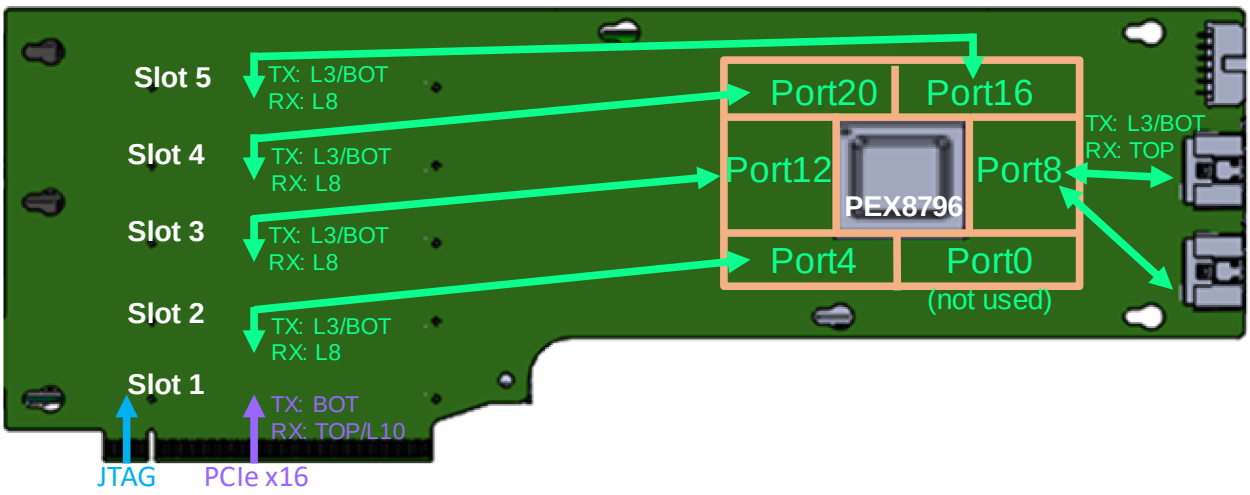


Figure 16: 5-Slot Active Riser PCIe Port Mapping and Routing Layers (View from Backside)

Table 7: 5-Slot Active Riser Stack-up

Dielectric and Copper Properties											
Layer			Copper		Dielectric Material					Layer	
Name	Description	Type	Weight	Type	Type	Er	Df	Actual Dk	Actual Df/1MH	Design Thickness (mil)	Actual Thickness (mil)
Top_SM	Top Soldermask	Soldermask			Soldermask	3.5	0.030	3.70		0.50	0.50
L1top	Signal_L1top	Signal	0.5 oz. plated	STD						1.90	2.10
PP_L1top_L2	Prepreg	PP			1080*1	3.6	0.009	3.50	0.007	2.70	2.72
L2	GND_L2	GND	0.5 oz.	RTF						0.65	0.66
Core_L2_L3	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L3	Signal_L3	Signal	0.5 oz.	RTF						0.65	0.66
PP_L3_L4	Prepreg	PP			1080*2	3.6	0.009	3.50	0.007	5.50	5.11
L4	GND_L4	GND	0.5 oz.	RTF						0.65	0.66
Core_L4_L5	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L5	Signal_L5	Signal	0.5 oz.	RTF						0.65	0.66
PP_L5_L6	Prepreg	PP			7628+1080	3.6	0.009	3.60	0.007	9.50	9.61
L6	PWR_L6	PWR	1.0 oz.	RTF						1.30	1.25
Core_L6_L7	Core	Core			2116X1	3.8	0.009	3.60	0.007	4.00	4.20
L7	PWR_L7	PWR	1.0 oz.	RTF						1.30	1.25
PP_L7_L8	Prepreg	PP			7628+1080	3.6	0.009	3.60	0.007	9.50	9.61
L8	Signal_L8	Signal	0.5 oz.	RTF						0.65	0.66
Core_L8_L9	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L9	GND_L9	GND	0.5 oz.	RTF						0.65	0.66
PP_L9_L10	Prepreg	PP			1080*2	3.6	0.009	3.50	0.007	5.50	5.11
L10	Signal_L10	Signal	0.5 oz.	RTF						0.65	0.66
Core_L10_L11	Core	Core			1086X1	3.6	0.009	3.50	0.007	3.00	3.00
L11	GND_L11	GND	0.5 oz.	RTF						0.65	0.66
PP_L11_L12bot	Prepreg	PP			1080*1	3.6	0.009	3.50	0.007	2.70	2.72
L12bot	Signal_L12bot	Signal	0.5 oz. plated	STD						1.90	2.10
Bot_SM	Bot Soldermask	Soldermask			Soldermask	3.5	0.030	3.70		0.50	0.50
					with soldermask, total thickness:					64.00	64.04
					without soldermask, total thickness:					63.00	63.04

- Broadcom/PLX 96-Lane PCIe Gen3 Switch (PEX8796)
    - Switch Bifurcated into 5 x16 Ports (1 x16 Port to each x16 Slot and 1 x16 Port to Sliver conns)
    - Depending on the PCIe Card requirement, the Switch can be bifurcated into different lane width configurations. Custom Switch FW is required to change the Switch bifurcation.
    - Switch is located on Backside as required by Chassis Thermals
    - PCIe Switch Thermal Design Power Requirements:
        - 2 x Double-Wide Card Config can use up to 48-lanes and requires 32.5W
        - 4 x Single-Wide Card Config can use up to 80-lanes and requires 35.8W
        - Heatsink is required
    - Please refer to the PEX8796 Datasheet and User Manual for additional design requirements
- 2 x PCIe x8 Sliver connectors supporting:
    - Delivery of x16 lanes to Broadcom/PLX 96-Lane PCIe Gen3 Switch
    - x8 Edge Connector to x8 Sliver Connector Cable
    - 85-ohm characteristic impedance
    - 2 different Sliver Connectors/Cables are used from TE Connectivity:
        - 50 pin Connector/Cable supports lower x8 PCIe lanes, connected to MB PCIe Slot #1
        - 50 pin Connector/Cable supports upper x8 PCIe lanes, connected to MB PCIe Slot #2

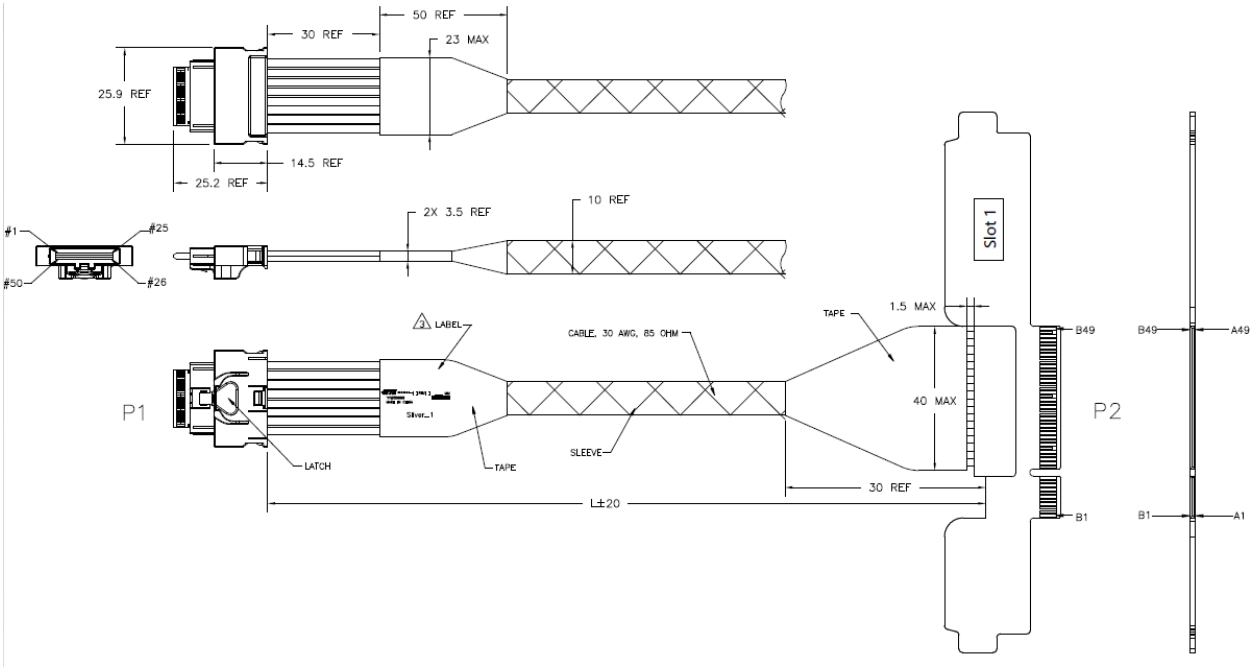
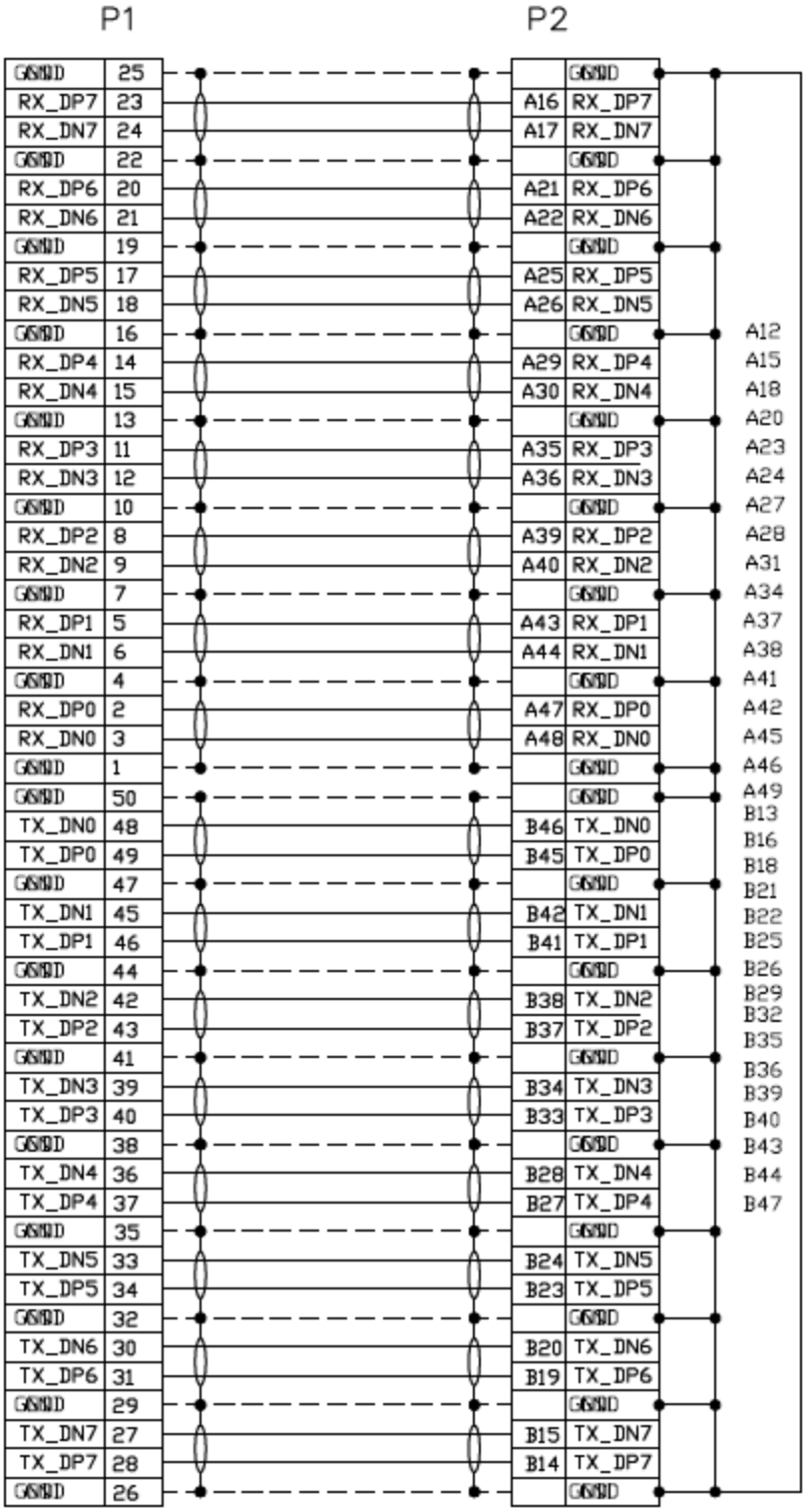


Figure 17: 50 pin x8 Sliver Connector to PCIe x8 Edge Connector Cable (Slot #1)

Table 8: 50 pin x8 Sliver Connector to PCIe x8 Edge Connector Cable Wire Mapping



2.5.4 PCIe Riser Power Requirements

- Each PCIe slot supports up to 75W as defined for a standard PCIe x16 card.
- 12V and 3.3V to PCIe Slot #1 is supported from the MB Riser Slot
- A 12V Auxiliary connector (2x6 pin cabled from PDB) is required on the Riser to support PCIe Slots #2-5 power, PEX8796 VRs, 3.3V VR, and 3.3V_Stby VR.
    - The required 12V Current for this Riser is ~30A (assume 35A for conn derating).
    - 12V AUX Connector uses 5 x 12V pins: 35A/5 pins = 7A per pin.
    - 12V AUX Connector uses 1 x 12V_PSU pin to generate 3.3V_Stby Power.
- Additional 12V power required by high power PCIe cards will cable directly from the PDB to the cards themselves.
- Different Wire Harnesses can be used depending on the PCIe Card's 12V Auxiliary Power connector
- Only Slots #2 and #4 can support a High Power, double-wide PCIe Card
- 3.3V Slot power is required for all PCIe Cards:
    - Per the PCIe Spec, 3.3V requires 3.0A per slot (5 x 3A = 15A), which cannot be pulled from only three Riser Slot 3.3V Pins.
    - 3.3V must be sourced from an on-board VR
- 3.3V_Stby may be required by some PCIe Cards:
    - Per the PCIe Spec, 3.3V_Stby requires 375mA per slot (5 x 375ma = 1.875A), which cannot be pulled from only a single Riser Slot 3.3V_Stby pin.
    - A small VR (generated from the 12V_PSU rail) will be added to the design to support the required current.

Table 9: 5-Slot Active Riser Power Budget

5-Slot Active Riser 12V Budget	**Voltage (V)**	**Current (A)**	**Qty**	**Total Current (A)**	**Power (W)**	**VR eff**	**12V Power Req'd (W)**	
PCIe 12V	12	5.5	4	22	264	100%	264	
PCIe 3.3V	3.3	3	4	12	39.6	90%	44	
DB600Z 3.3V	3.3	0.081	1	0.081	0.2673	90%	0.30	
Misc Logic/Components	3.3	0.1	1	0.1	0.33	90%	0.37	
PEX8796 VDD09	0.9		1		25.53	90%	28.36	
PEX8796 VDD09A	0.9		1		7.97	90%	8.86	
PEX8796 VDD18A	1.8		1		2.24	85%	2.63	**12V Current (A)**
PEX8796 VDD18	1.8		1		0.04	85%	0.05	
						Totals	**348**	**29**

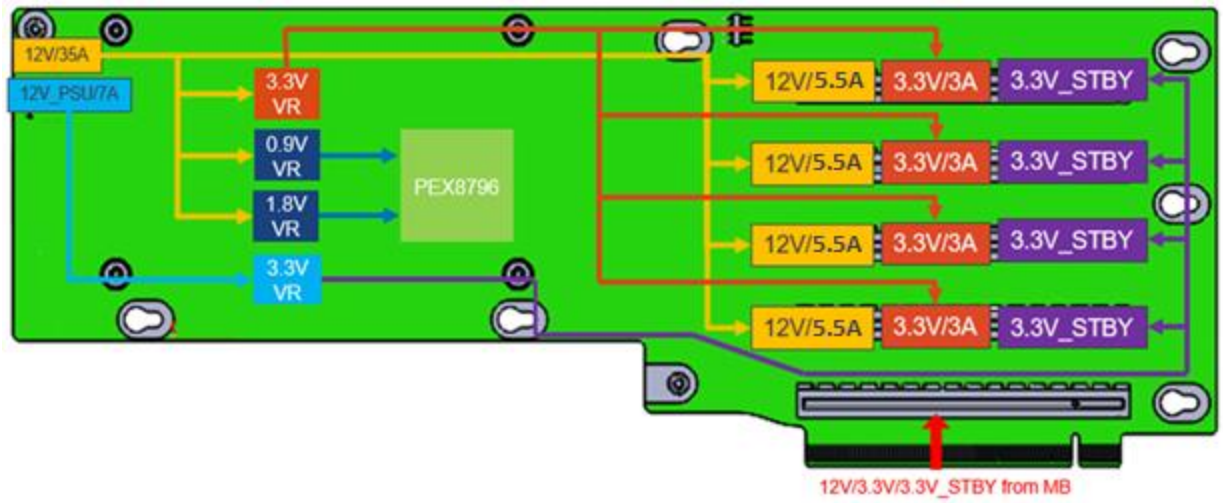


Figure 18: 5-Slot Active Riser Power Delivery Diagram

2.5.5 PCIe Riser I2C Requirements

All PCIe Connectors connect to the BMC I2C for PCIe Card telemetry. Use the PCIe designated pins for I2C (B5 & B6). To avoid I2C Address contention between PCIe Cards, an I2C MUX is needed on each riser. The design must add appropriate voltage translation/isolation between different I2C voltage domains.

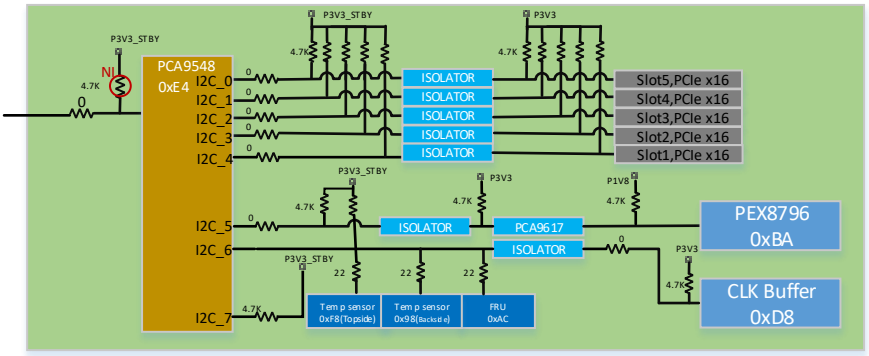


Figure 19: I2C Block Diagram for 5-Slot Active PCIe Riser

2.5.6 PCIe Riser Clock Requirements

A Differential Clock Buffer (DB600Z) is used to fanout additional PCIe Clocks. The Clock Buffer is set to PLL Bypass Mode by default. Recommend including resistor stuffing Options to change the PLL Bandwidth Mode if needed to improve Clock Jitter performance:

- PLL High BW Mode
- PLL Low BW Mode
- PLL Bypass Mode (default)

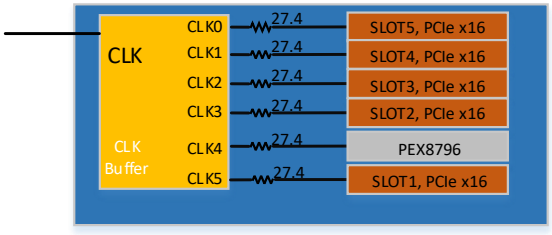


Figure 20: Clock Block Diagram for 5-Slot Active PCIe Riser

2.6 PCIe Card Requirements

This section provides any special requirements for the PCIe Cards in order to support the System features of the Project Olympus server.

2.6.1 Power Brake (PWRBRK_N)

Power Brake (PWRBRK_N; Pin B30) is an optional feature in the PCIe Spec that allows the system to throttle the PCIe Card. This feature is useful for System Power Capping when the 12V current in the system exceeds a predefined Over Current (OC) threshold in the PSU. With the Power Brake feature, the System has the ability to throttle PCIe Cards in order to avoid shutting down and ride through the OC event. It is required that all high-power PCIe Cards used in the XPO200 3UN support the Power Brake function in order to support the Project Olympus Universal MB Power Capping feature.

The response time of the PCIe Card throttling relative to a PWRBRK_N signal assertion must be approximately 10ms or less. The PCIe Card should throttle down to at least 50% of its max thermal design power (TDP). The PCIe Card must return to full speed after PWRBRK_N has been de-asserted.

NOTE: This feature is not required for lower power PCIe Cards (< 150W) but is strongly recommended.

2.6.2 IPMI Capable I2C

Any PCIe Cards used in the XPO200 3UN should support IPMI based I2C commands for BMC access to I2C devices and sensors.

3. Server Management Requirements

The following section provides a high-level overview of the XPO200 3UN Server Management Requirements. Additional Server Management Requirements can be found in the Project Olympus BMC FW Specification.

3.1 BMC Requirements

- Must generate the SHA-256 (Secure Hash Algorithm-256bit) Hash of the Signed BMC FW
- Rack Manager supported in the PMDU on the Project Olympus Rack will assign a “U” number to the Server BMC through PSU1 NODE_ID/SLOT_ID Bits only. PSU2 and PSU3 NODE_ID/SLOT_ID signals are not connected.
- System Fan Speed Control (FSC) mainly based on CPU and PCIe Card Temperature Sensors
- PSU Fan Speed Relationship with System Fan Speed developed to boost PSU fan speed relative to different system fan speeds:
    - If System Fans are less than 80% PWM, PSU can use its own FSC
    - If System Fans are greater than or equal to 80% PWM, PSU Fans must be boosted to 100% to avoid air recirculation through PSUs
- 1U Project Olympus Power Supply FW Updates from the BMC must be performed one PSU (Silver Box) at a time. The PSU will ensure that each of its internal 340W Modules are updated in sequence. This flow will ensure that only one internal 340W module out of nine is powered down/reset across the three Silver Boxes.
- Due to Battery capacity limitations, no Battery related features will be supported on this system.
- When a PSU ALERT# signal is asserted, this is an indication that a critical event has occurred in one or more of the PSUs. The BMC FW must read all 3 of the PSU status registers and log a SEL Event.
- Must support the Remote Debug-At-Scale feature to access CPU debug information through the BMC dedicated management port.

3.2 I2C Block Diagram

The MB BMC must be able to access all PCIe Cards through I2C. Below are the I2C Device Mappings for the XPO200 3UN Server Configuration.

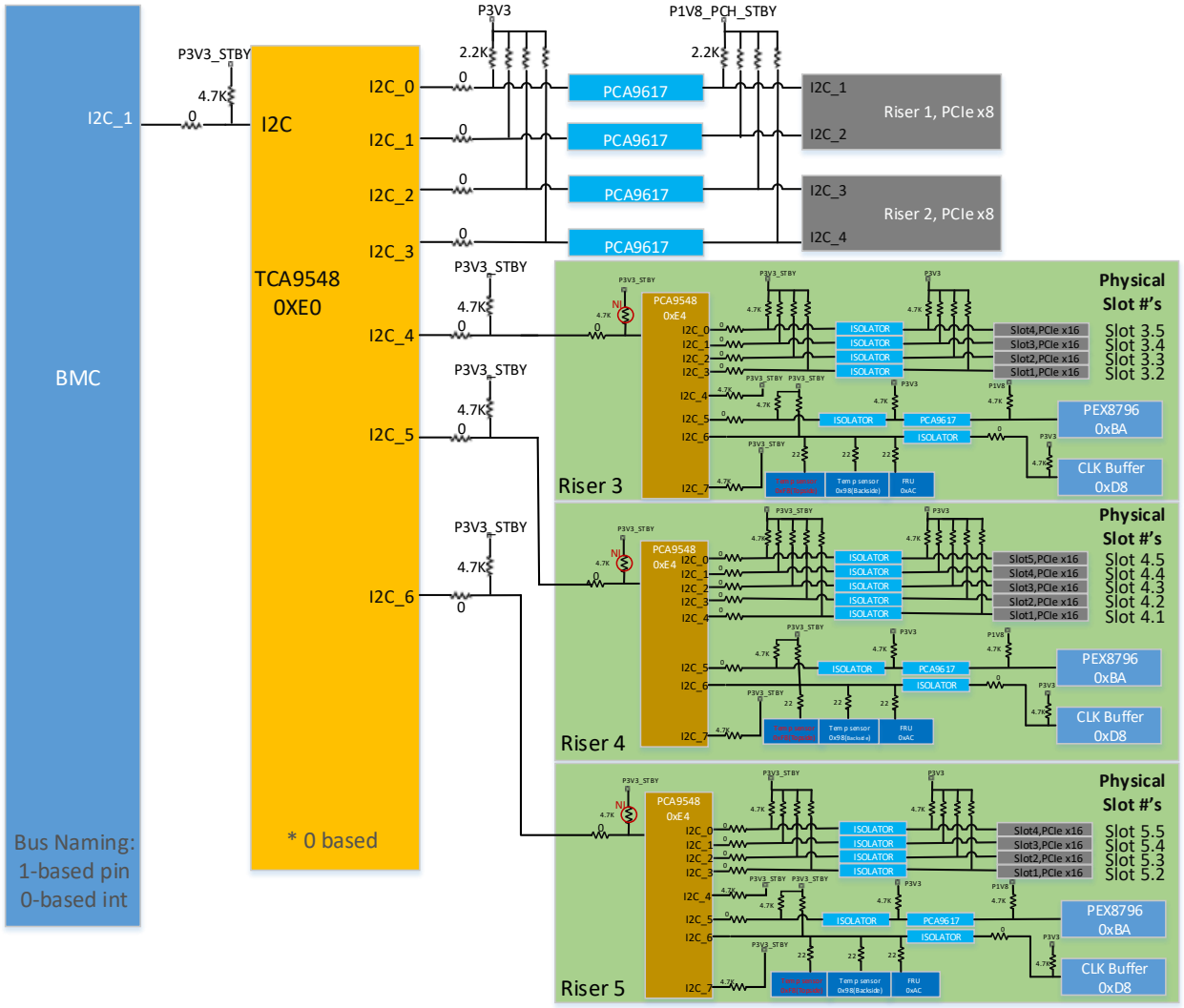


Figure 21: I2C Block Diagram for 13 Single-Wide PCIe Cards

Table 10: I2C Bus Mapping for 13 Single-Wide PCIe Cards

	Valid/Invalid	PCIe Riser Slot # (Silkscreen on Riser)	PCIe Physical Slot # (1 to 5 bottom to top)	I2C Bus (From BMC side, 0-based)	I2C Channel on First I2C MUX (0-based)	I2C Channel on Second I2C MUX (0-based)
PCIe Riser in Slot 3	Valid	4	3.5	0	4	0
	Valid	3	3.4	0	4	1
	Valid	2	3.3	0	4	2

	Valid/Invalid	PCIe Riser Slot # (Silkscreen on Riser)	PCIe Physical Slot # (1 to 5 bottom to top)	I2C Bus (From BMC side, 0-based)	I2C Channel on First I2C MUX (0-based)	I2C Channel on Second I2C MUX (0-based)
	Valid	1	3.2	0	4	3
	Invalid	N/A	3.1	N/A	N/A	N/A
PCIe Riser in Slot 4	Valid	5	4.5	0	5	0
	Valid	4	4.4	0	5	1
	Valid	3	4.3	0	5	2
	Valid	2	4.2	0	5	3
	Valid	1	4.1	0	5	4
PCIe Riser in Slot 5	Valid	4	5.5	0	6	0
	Valid	3	5.4	0	6	1
	Valid	2	5.3	0	6	2
	Valid	1	5.2	0	6	3
	Invalid	N/A	5.1	N/A	N/A	N/A

4. Thermal Design Requirements

The following section provides a high-level overview of the XPO200 3UN Server Thermal Design Requirements.

4.1 Thermal Design Requirements

- Mechanically Supports up to 8 Dual Rotor 60mm x 56mm System Fans (6 Supported Electrically)
- Supports 158 CFM/kW @ 25C, 30C and 35C Ambient
- Supports Single-Fan Zone for System Components (not including PSUs)
- Supports two rotor Fan-Fail scenario: N+2 condition (CFM/kW Spec not required during Fan-Fail)
- The system must operate at full functionality in an “N” Fan condition at the highest workload and highest ambient defined.
- For dual rotor Fan numbering: “a” Fans on inlet side and “b” Fans on exhaust side
- Rear of chassis supports plastic Fan Flappers for Fan-Fail condition to prevent air recirculation.

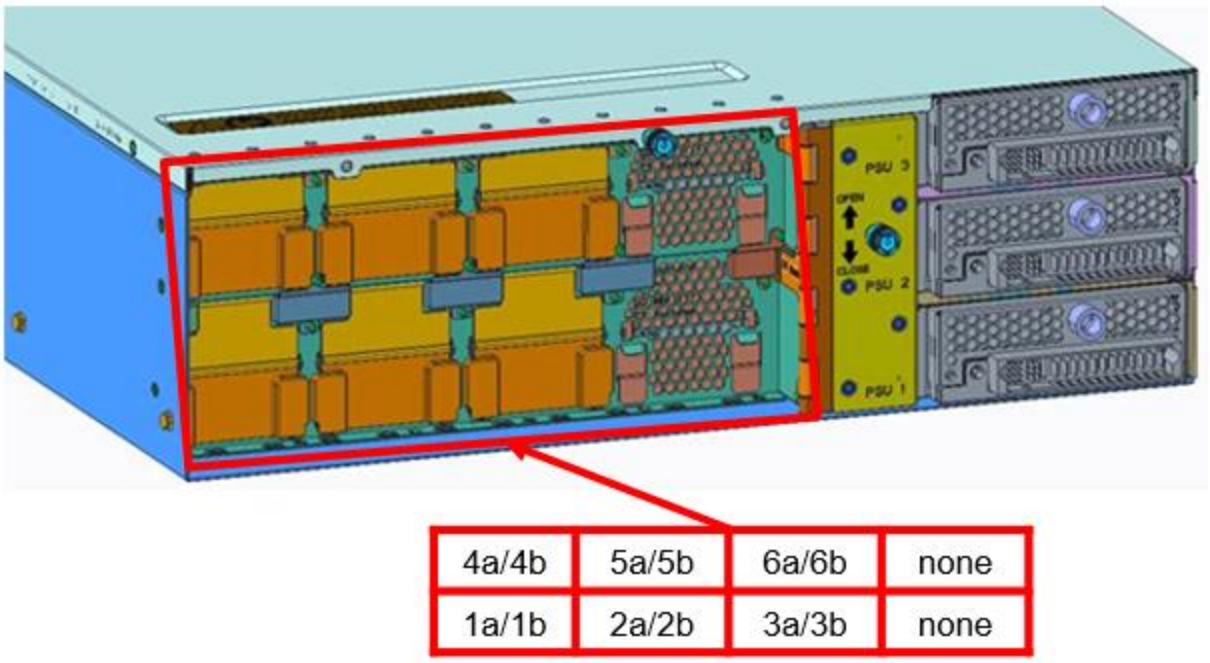


Figure 22: 3U Chassis System Fan Position Diagram with Closed Fan Flappers

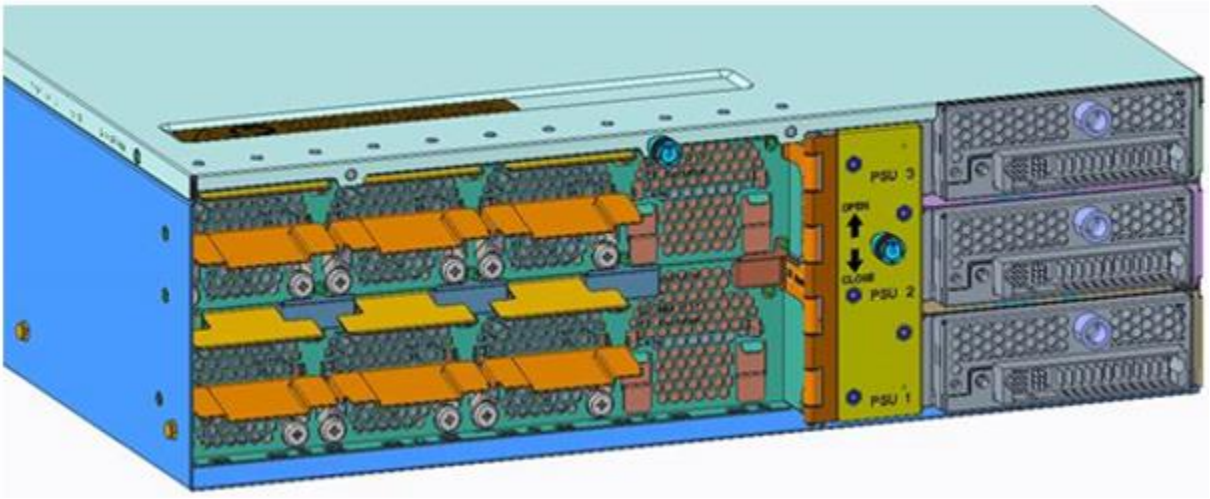


Figure 23: 3U Chassis with Open Fan Flappers

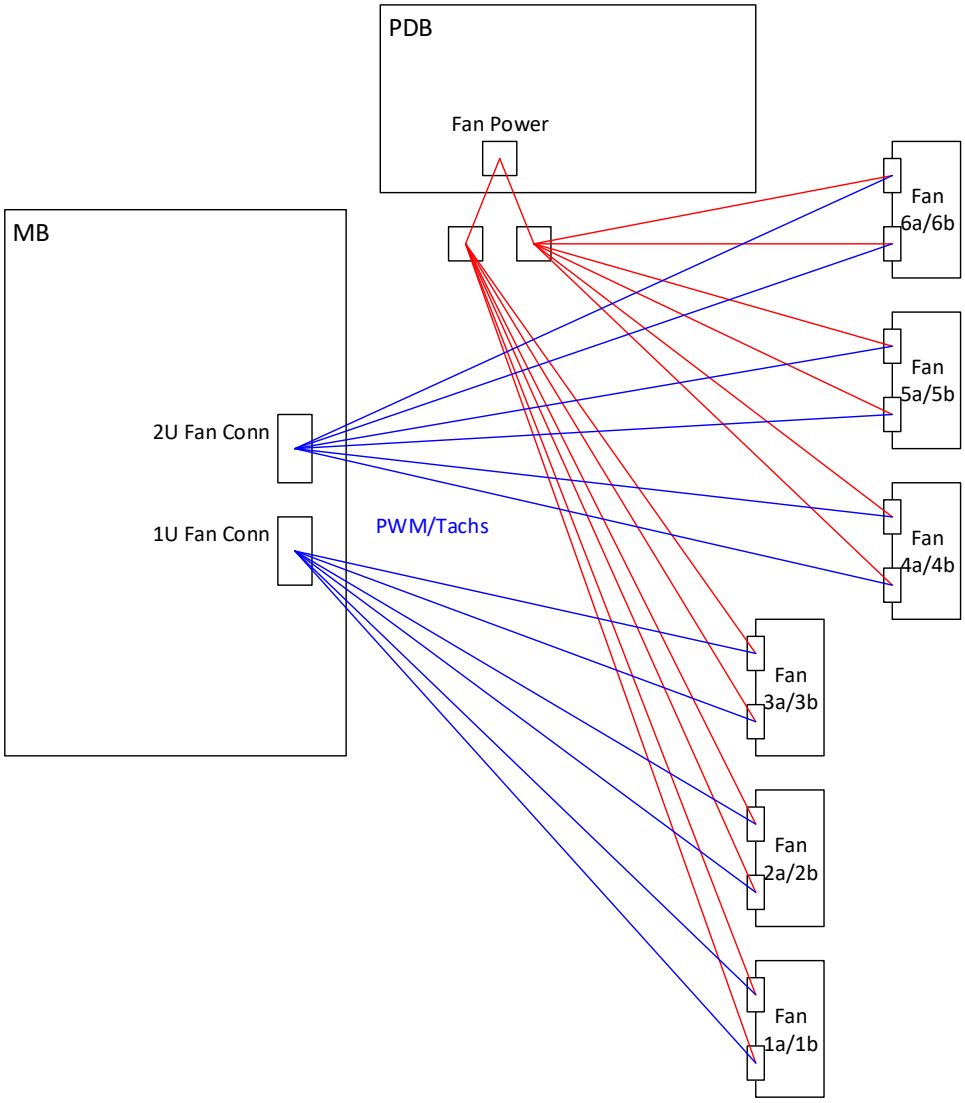


Figure 24: System Fan Connectivity Diagram

- Supports isolated PSU Fan Zone with sheet metal and plastic ducting
- For lightly loaded configs where a limited number of PCIe cards are plugged into PCIe Riser 4 (Middle Riser) a plastic baffle can be used to block airflow through the middle of the Chassis. The baffle forces more airflow across the high powered PCIe Cards on the outer Riser Slots.

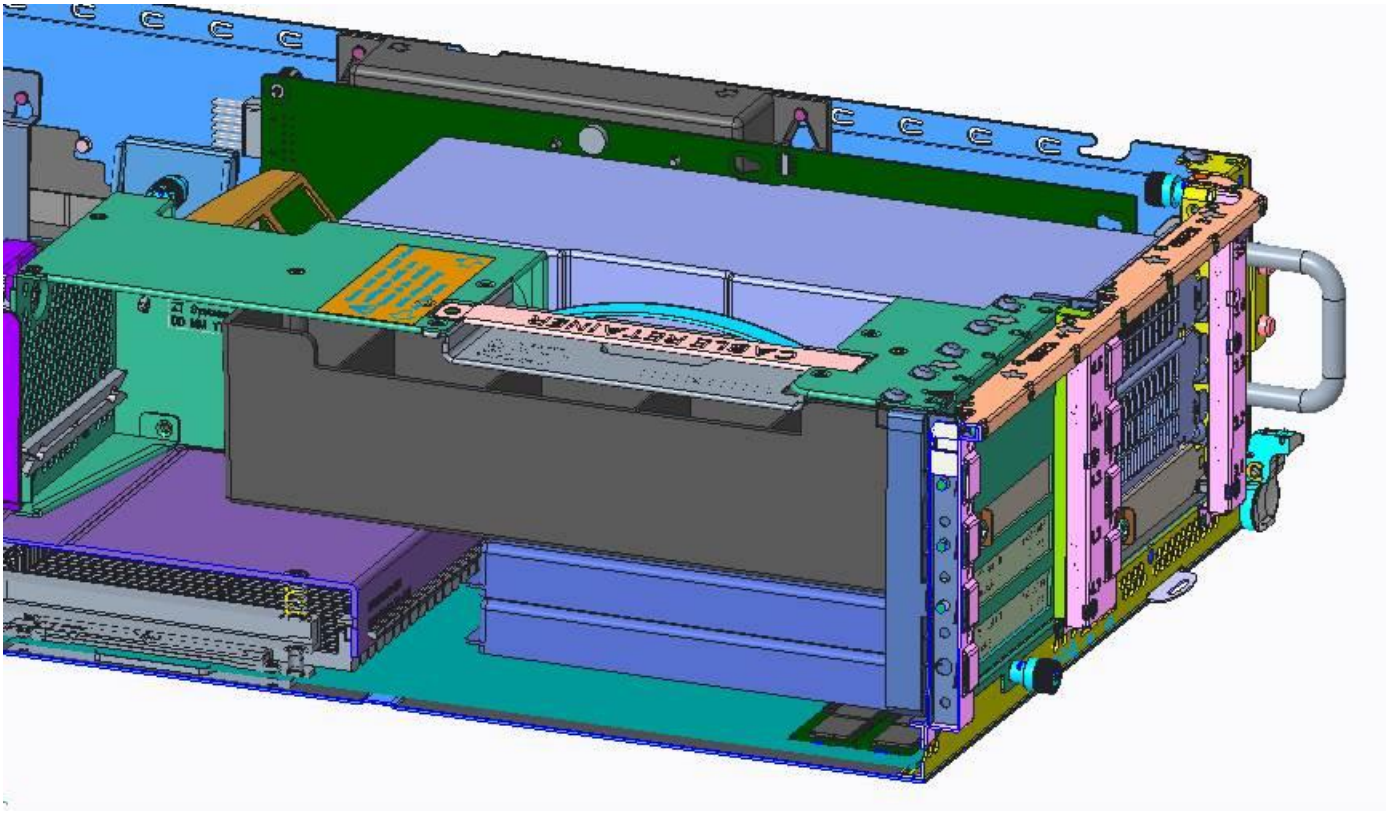


Figure 25: 3U Chassis Cross Section Showing Middle Riser Baffle (to block airflow)

4.2 PCIe Riser Temp Sensor Locations

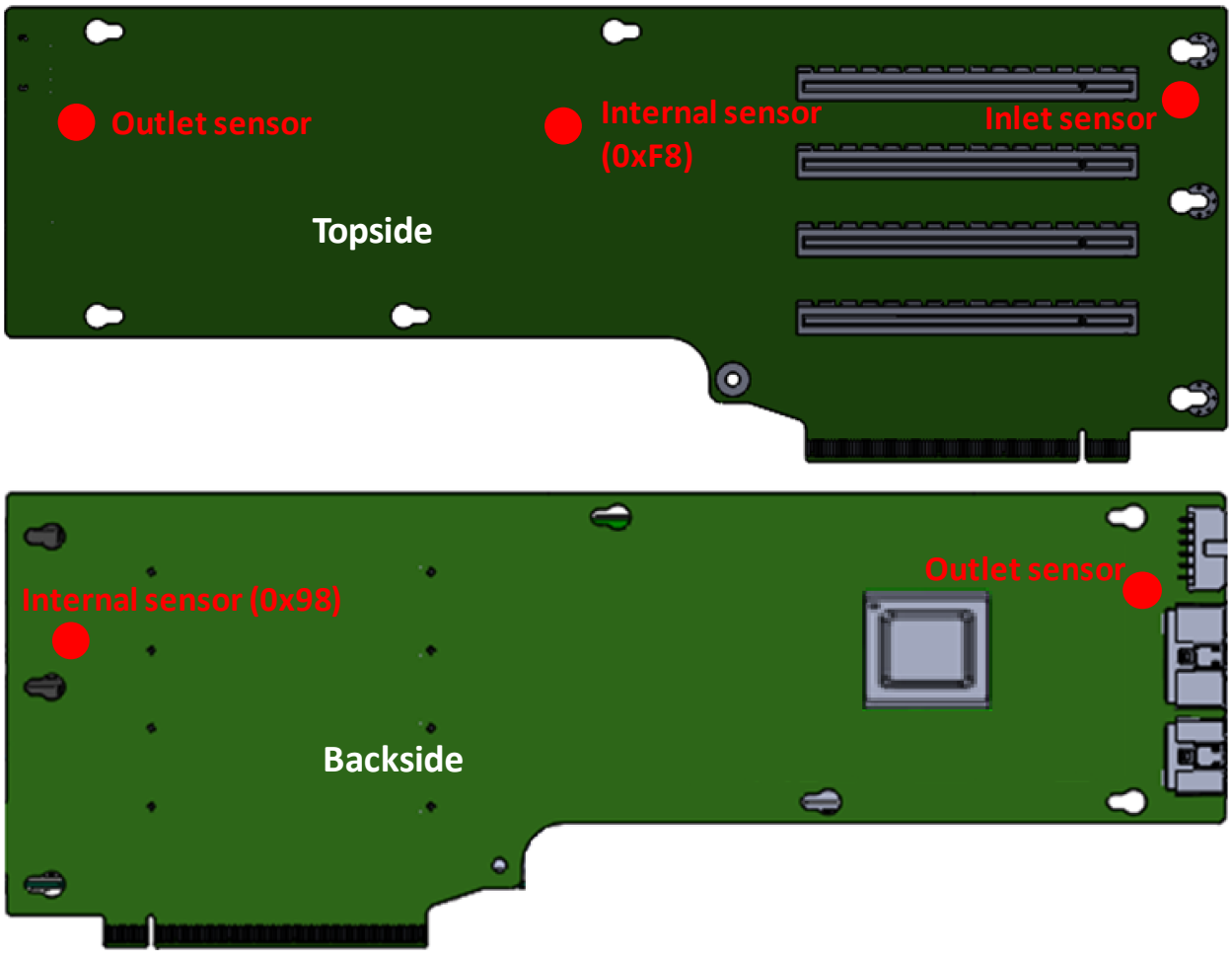


Figure 26: 4-Slot Active Riser Temp Sensor Locations

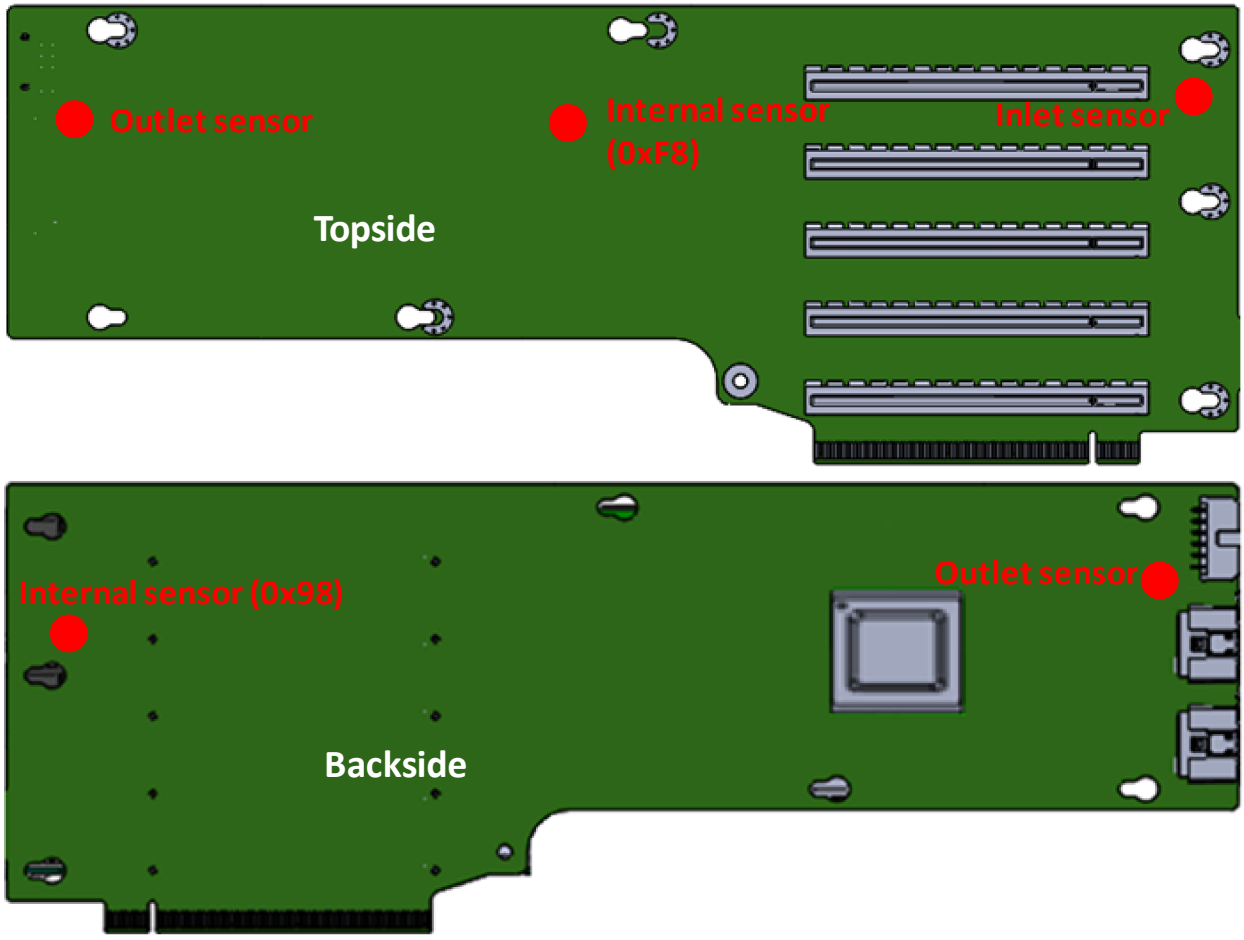


Figure 27: 5-Slot Active Riser Temp Sensor Locations

5. Mechanical Requirements

5.1 Chassis Requirements:

- 3U 19” EIA310-D Compliant Chassis
- Supports Project Olympus 42U or 48U Racks
- Supports Custom King Slide Rails (Does NOT support Project Olympus T-Pin Tracks)
- Supports Project Olympus PMDU power connections
- Supports OCP Approved Project Olympus Motherboards (with changes to BIOS & BMC FW)
- Supports quick release front latches for Blade Hot Swap from Rack
- Supports Split Top Cover for convenient Serviceability of components in front MB Bay
- Supports clean cable management with strategically placed cable tie mounts
- Supports Power Distribution Board on elevated shelf in Rear Bay
- Supports 2.5” or 3.5” HDD/SSD on elevated shelf in Rear Bay
- Supports 3 x Project Olympus 1kW PSUs

- Supports up to 8 x Dual Rotor 60mm Fans (6 fans supported electrically)
- Supports 3 x Riser Cages and retention guides/brackets for PCIe Risers and single-wide/double-wide cards respectively

5.2 Chassis Views

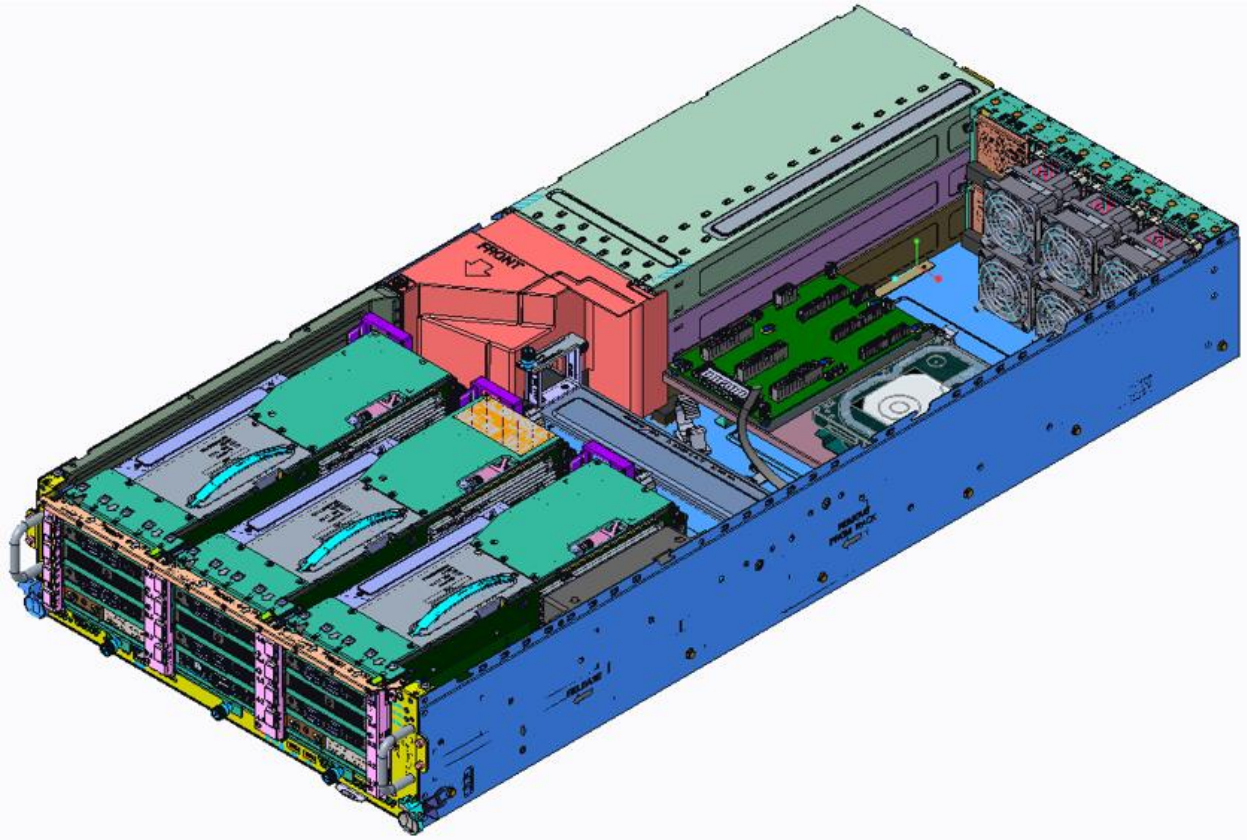


Figure 28: 3U Chassis/System Layout (Iso View)

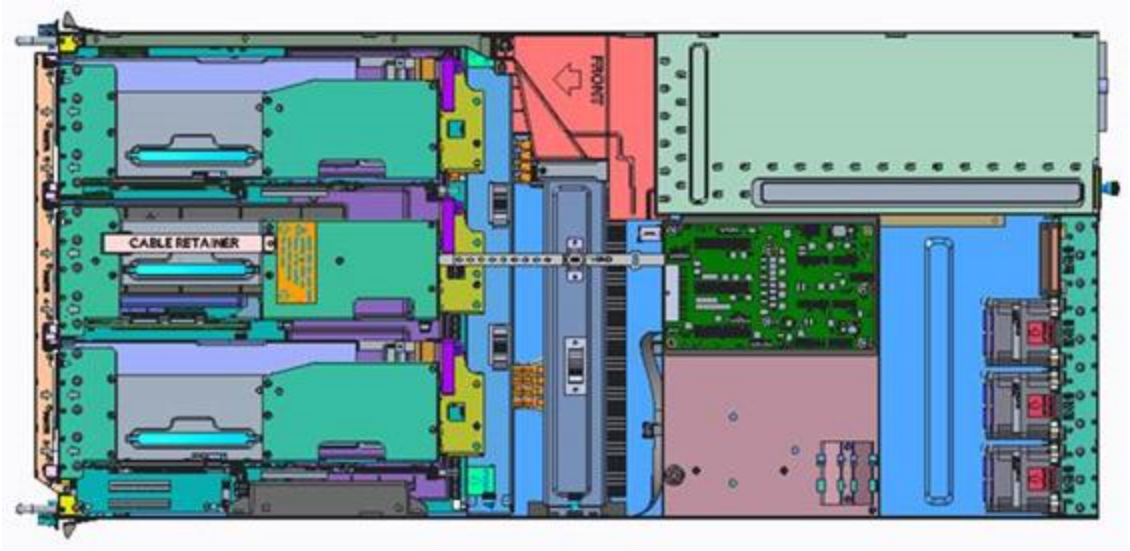


Figure 29: 3U Chassis/System Layout (Top View)

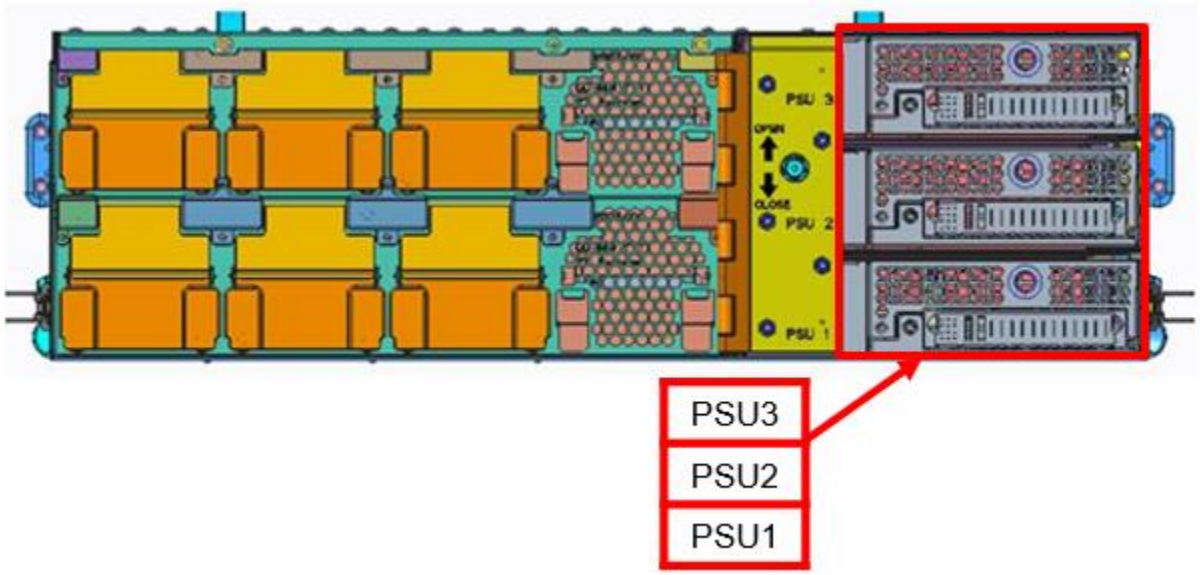


Figure 30: 3U Chassis/System and PSU Layout (Rear View)

Due to the weight of the XPO200 3UN Server (90lbs), the standard Project Olympus T-Pin Slide Rails will not allow the Chassis to pull in and out smoothly. As a result, a custom King Slide Rail Kit is used to support easier travel during Servicing. However, since the width of the chassis is 17.36" (44.10 cm) thinner Slide rails were required to fit into the 19" Project Olympus Rack, which limits how far the XPO200 3UN Server can be pulled out of the Rack. The XPO200 3UN Server can only be pulled out about 67%, with 650mm-travel ball bearing sliding rails. For this reason, the Top cover is split into two sections (Front

and Rear). The Front Top Cover section allows access to most of the System hardware, leaving only the System Fans, PDB and Power Cables under the Rear Top Cover without access.

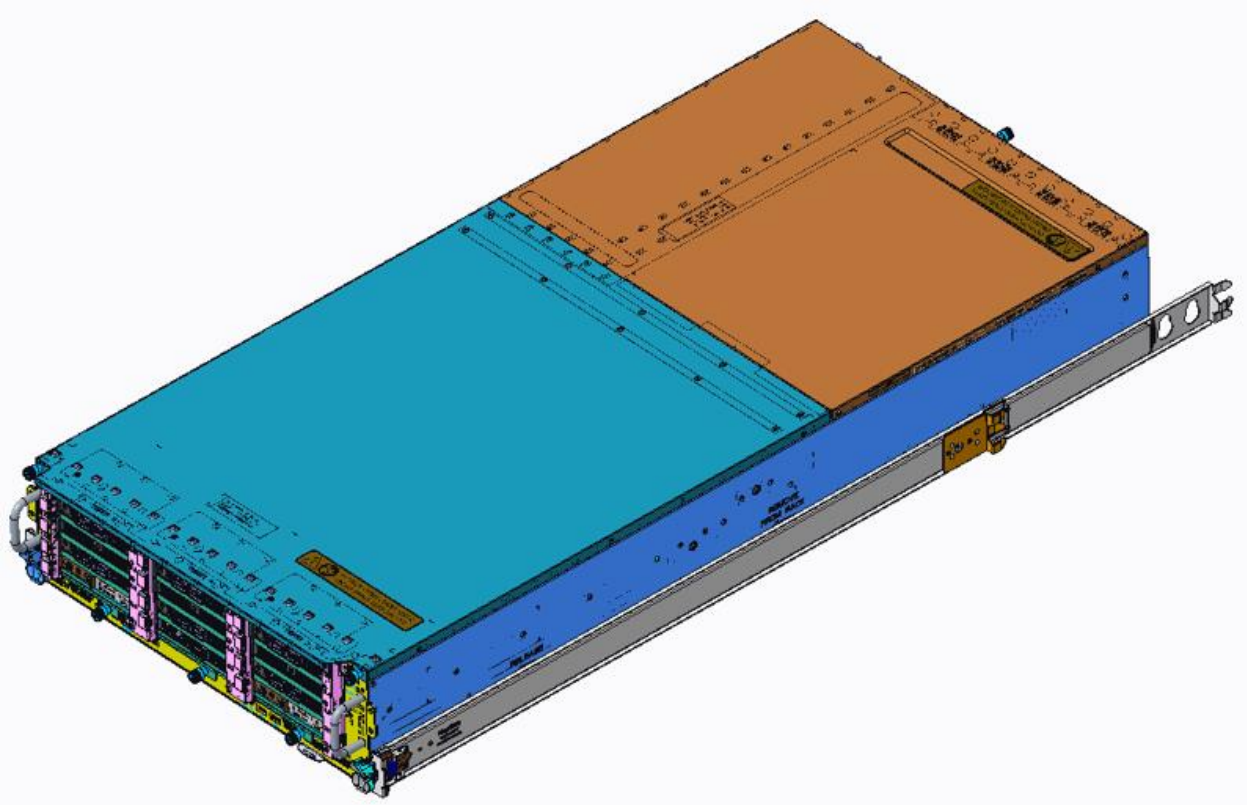


Figure 31: 3U Chassis/System with Split Top Cover King Slide Rails

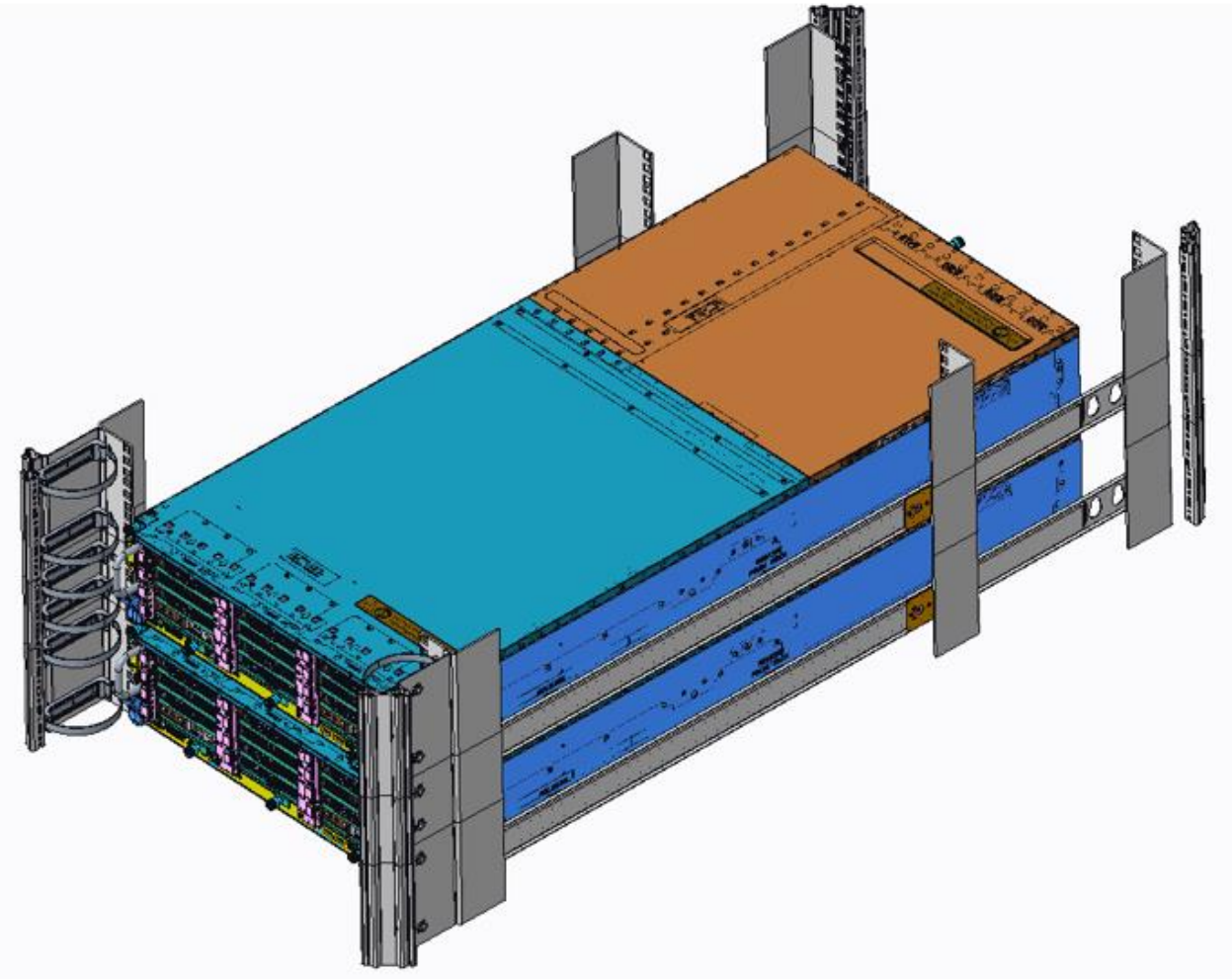


Figure 32: 3U Chassis/System in 19” Project Olympus Rack

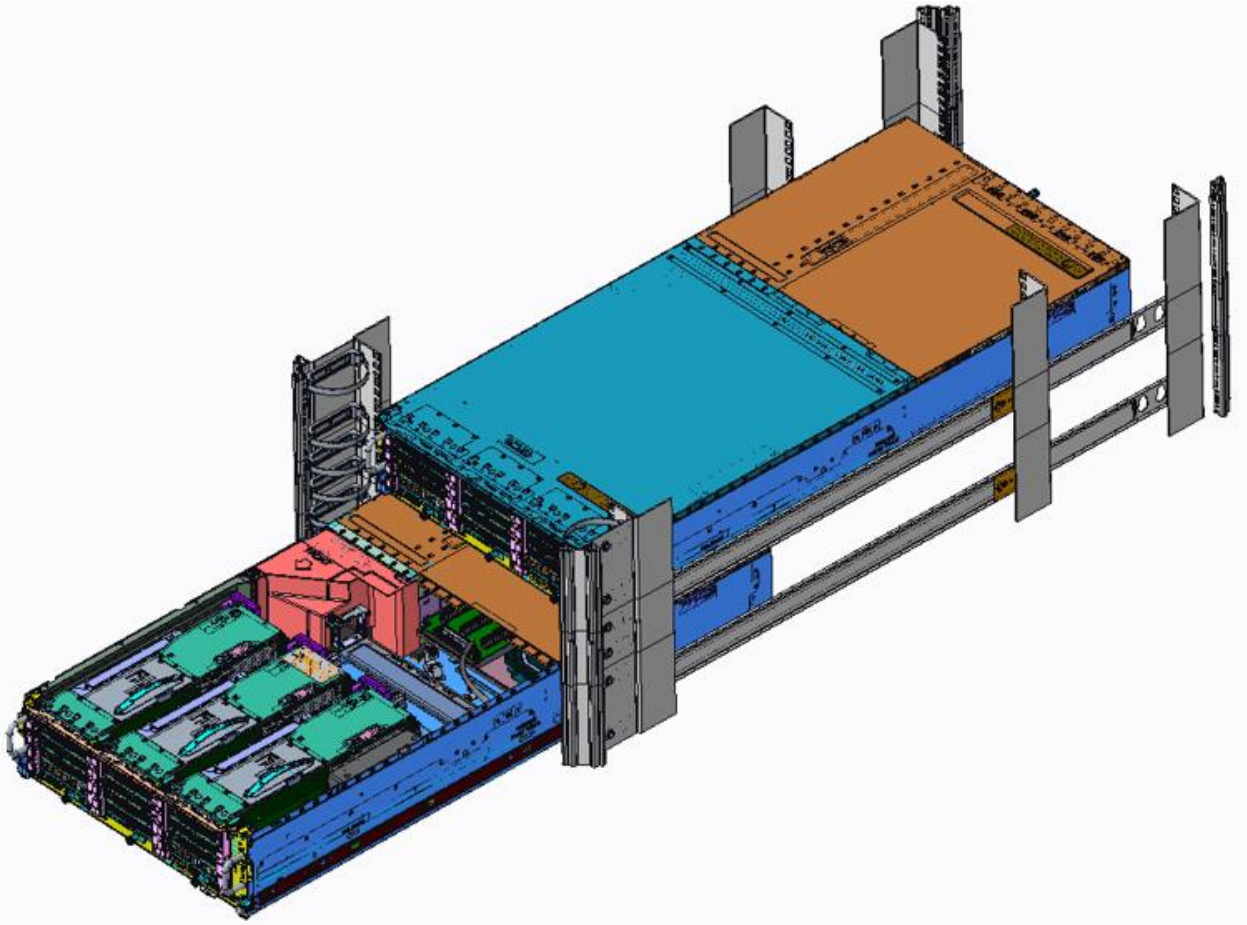


Figure 33: 3U Chassis/System in 19” Project Olympus Rack (Extended 67%)

6. System Power Requirements

The XPO200 3UN Server will support the 1U Project Olympus 1kW Power Supplies. Due to System Power Requirements, multiple Power Supplies in parallel will be required. This section defines how the power delivery sub-system will be designed, controlled and monitored.

6.1 System Power Budget

The following section provides System Power Budgets for the 3U PCIe Expansion Server Configurations. Specific System SKUs will require separate Power Budget analysis to ensure they remain within the Max TDP and Max Peak loading envelope.

Table 11: XPO200 3UN Server Power Budget

Function	Device		Power (W)	VR Eff.	Util.	Qty.	Subtotal (W)
CPU	Intel Skylake (SKX) 205W	TDP	205	90%	100%	2	455.6
		PMAX	348.5	90%	100%	2	774.4
System Memory	32GB DDR4, 2667MHz, 2Rx4 RDIMM		7.54	90%	100%	24	201.1
PCH	Lewisburg PCH		6	90%	100%	1	6.7
PCIe Riser Card	Riser #3 / #5 4 slots active 4SW		33.9	90%	100%	2	75.3
PCIe Riser Card	Riser #4 5 slots active 4SW		33.9	90%	100%	1	37.7
GPU Card Single-Wide	Nvidia T4 GPU Card	TDP	75	100%	100%	12	900.0
		PMAX	204	100%	100%	12	2448.0
Network	10G Single port SFP+ PCIe 2.0 x8 5GT/s		15	90%	100%	0	0.0
BMC	AST1250		1.7	90%	100%	1	1.9
USB	USB3.0		2.5	90%	100%	2	5.6
SSD	M.2 Samsung PM963 960GB PCIe SSD		7.5	90%	100%	7	58.3
System Fans	Delta 6056 GFC0612DSA01XXX-REVX00		32.4	100%	100%	6	194.4
Other	Drivers, logic, pull-ups, etc …		10	90%	100%	1	11.1
				Total (TDP)			2023
		Turbo mode		Total (PMAX)			3890
				PSU Efficiency			93.5%
			Total AC Power				**2163**

6.2 Power Distribution Board

This is a Power Distribution Board (PDB) designed for the XPO200 3UN Server, which supports load sharing of 12V Power from the 3 x Project Olympus Power Supplies (PSUs) to the entire system. The PDB also supports the pass through and combining of various management signals required for Project Olympus Rack Manager & BMC identification, control and monitoring.

6.3 Major Component Placement

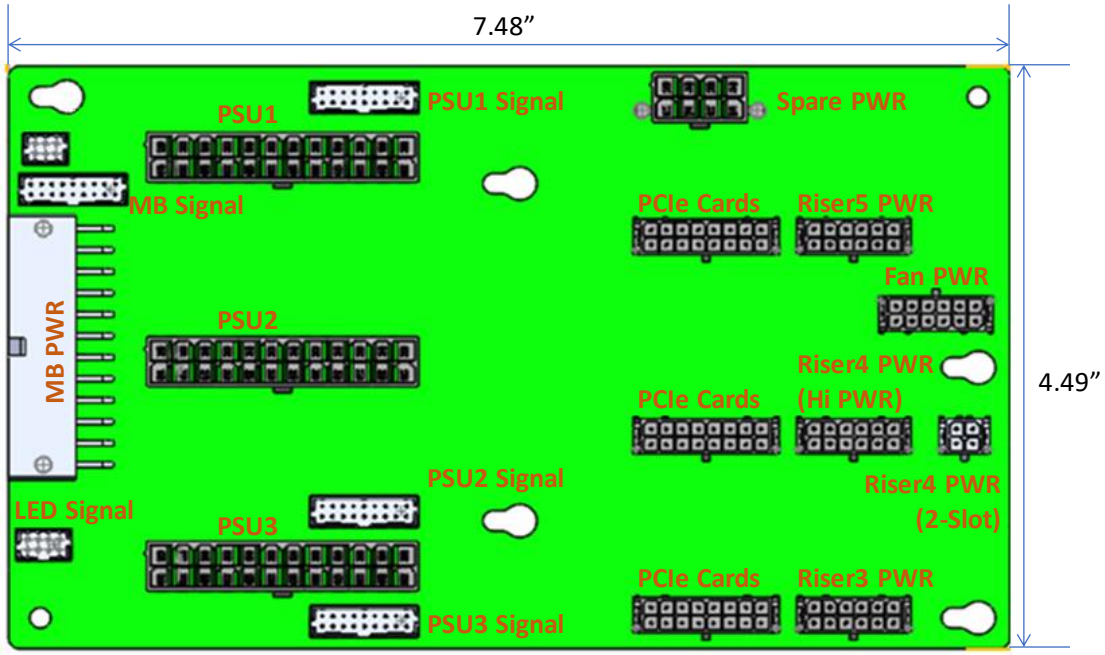


Figure 34: Power Distribution Board Layout

6.4 PDB Power Block Diagram

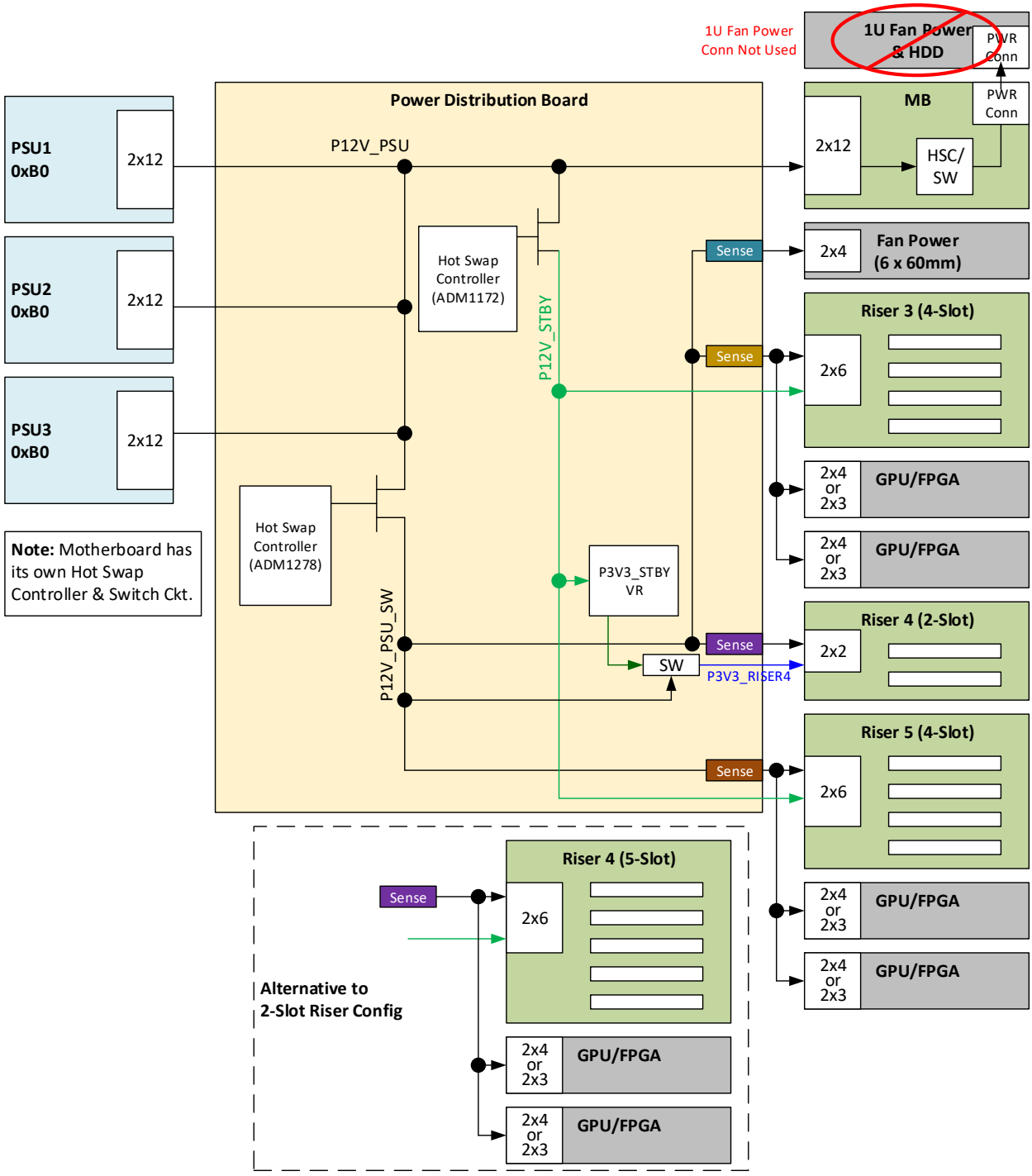


Figure 35: Power Distribution Board Power Block Diagram

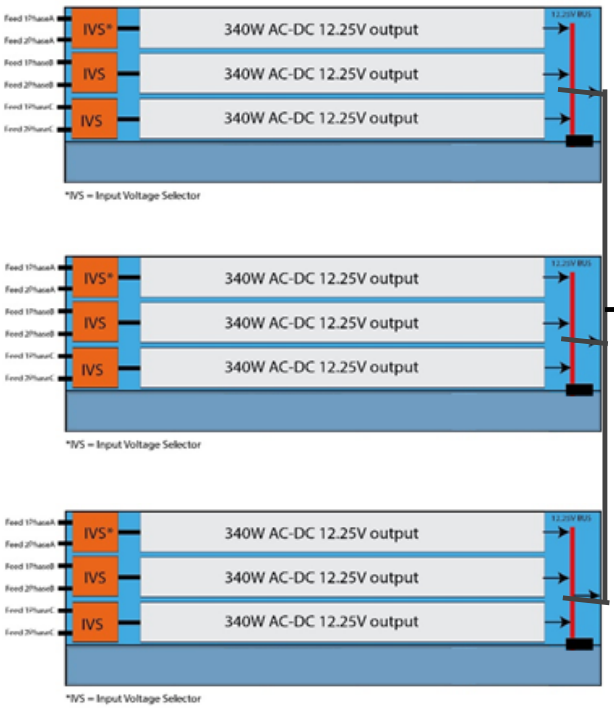
NOTE: XPO200 3UN uses the 5-Slot Riser Option in Riser Slot 4 and does not use Double-wide PCIe Cards with 12V Aux power connections. The diagram above is showing optional configurations that can be supported by the 3U Server.

6.5 PDB Configuration

The XPO200 3UN Server will use the PDB to Load Share across 3 x Project Olympus PSUs. The PDB uses Passive Droop Sharing between PSUs, meaning no V_Sense or I_Share signals are required. The PSU Setpoint should be 12.35V. The Project Olympus PSUs support 3 x internal, phase-balanced 340W PSUs that also support Passive Droop Sharing.

Since the XPO200 3UN Server supports 3 PSUs with 3 internal PSU Modules, this equals 9 total PSU Modules per Server. As a result, the Server supports an 8+1 redundancy model, where the system will ride through any single PSU module failure. Depending on the loading conditions a System may be able to ride through losing 2 PSU Modules. To maintain 8+1 PSU Module redundancy with margin, the max supported System Power Budget is approximately 2400W (DC).

Project Olympus PSUs typically support an integrated battery pack (680W max). However, the combined battery power of 3 x 680W packs does not meet the XPO200 3UN Server max System Power requirements. Consequently, the Non-Battery PSU will be used in this system and therefore, no Battery related features will be supported.



Max Supported Power		
Loading	Power (W)	Current (A)
TDP	2400	200
Peak	2719	227

Figure 36: XPO200 3UN 3U Server Power Supply Load Sharing Diagram (Max Supported Power)

6.6 PDB Power Delivery Requirements

The PDB supports a Hot Swap Controller for Power Telemetry and isolation if the current sense circuits are tripped. A Hot Swap Controller that can be used for this application is the ADM1278. All P12V System components are behind this Hot Swap Controller except for the MB, since the MB already has its own Hot Swap Controller. The PDB Hot Swap Controller sits on the same BMC I2C bus as the MB Hot Swap Controller, but with a different I2C address.

A 12V current sense chip that can be used for each power delivery branch, except for the MB power delivery branch, is the INA301:

- Branch P12V_A: Riser #3 and 2 x High Powered PCIe Cards
- Branch P12V_B: Riser #4 and 2 x High Powered PCIe Cards
- Branch P12V_C: Riser #5 and 2 x High Powered PCIe Cards
- Branch P12V_D: Fan Power

NOTE: All Fan Power comes directly from the PDB, so the Fan power connectors from the MB are not used.

Below is a list of additional components and features supported on the PDB:

- IO Expander: Device that supports exposing additional GPIOs from the PDB to the BMC through I2C.
- FRU: Contains identification information about the PDB.
- Temp Sensor: Supports Temperature Sensor readings that can be used by the BMC to ensure Thermal Levels remain within an acceptable range.
- 12V Vsense Pins (Reserved for Future Use): Used to support 12V remote sense for more optimal load sharing between Power Supplies. Not used on current version of the 3U PCIe Expansion Server or Project Olympus PSUs.

6.7 PDB Miscellaneous IO Requirements

This section describes the miscellaneous power related IO signals that traverse the PDB from the Project Olympus Rack Manager to the Project Olympus MB and corresponding BMC.

- PMBUS: Supports communication channels from BMC to PSUs and PDB I2C devices
    - PSU1_PMBUS has a private connection to BMC and PCH
    - PSU2_PMBUS & PSU3_PMBUS share a connection through an I2C MUX to avoid address contention
- PSU_ALERT_N: PSU output that represents a change in status in any one of the three PSUs. Although each PSU signal is Open Drain, they include 1kOhm internal pull-ups. These pull-ups in combination with the 4.7K pull-up on the MB create a strong equivalent pull-up and make it difficult for the PSU drivers to drive the PSU_ALERT_N signal low enough to meet the Vil_max of the receivers on the MB. As a result, each PSU_ALERT_N signal has been isolated by its own Open Drain Buffer. If any PSU asserts its PSU_ALERT_N signal, the BMC should read the status registers of all three PSUs (through PMBUS) to find out what happened. The PDB PSU_ALERT_N signal feeds both PSU1_ALERT_N and PSU2_ALERT_N pins to the BMC and on-board Throttle logic:
    - PSU1_ALERT_N (thru 1U Conn): Feeds the CPU PROCHOT_N pins for immediate CPU throttling.
    - PSU2_ ALERT_N (thru 2U Conn): Feeds the PCIe Slot PWRBRK_N pins for immediate PCIe Card throttling.

- PS_ON_N: PSU input used to Power-on/Power-off the PSUs. Each signal is Wire-OR'd together so that all three PSUs can be Powered-on/Powered-off at the same time.
- BLADE_EN_N: Pass-through signal from Rack Manager (RM) that supports initiating a Power-On/Power-Off event through assertion/de-assertion of PS_ON_N signal. Only BLADE_EN_N from PSU1 is tied to PS_ON_N. BLADE_EN_N from PSU2 and PSU3 can be monitored by the BMC (through I2C IO Expander) as a provision in case separate System actions are required; but are not currently supported in the 3U PCIe Expansion Server BMC FW.
- BLADE_THROTTLE: Pass-through signal from RM that supports immediate throttling of the System. Only PSU1 BLADE_THROTTLE is supported/required on the PDB since it ties directly into the PROCHOT_N and PWRBRK_N inputs of the CPU and GPU cards respectively.
- NODE_ID[5:0]/SLOT_ID[5:0]: Pass-through signal from RM that assigns a unique ID number to each Server (stored in its BMC). Each ID corresponds to a "U" location in the Rack. Only PSU1 passes the NODE_ID bits to a given 3U Server/BMC. The other PSU NODE_ID bits are not needed and subsequently left floating on the PDB.
- PSU_LED[1:0]: PSU output that provides status indications for its corresponding PSU. Since the MB only supports 2 sets of PSU_LEDs, the PDB supports a separate PSU_LED header/cable, which houses 3 x Bi-Color LEDs; one for each PSU. See following section for more details on the PSU_LED Cable.
- BLADE_PRESENT_N: Passthrough signal from BMC to RM that ensures that each Server is present. In the case of this system, only BLADE_PRESENT_N from PSU1 is used. PSU2 & PSU3 BLADE_PRESENT_N signals are left floating.

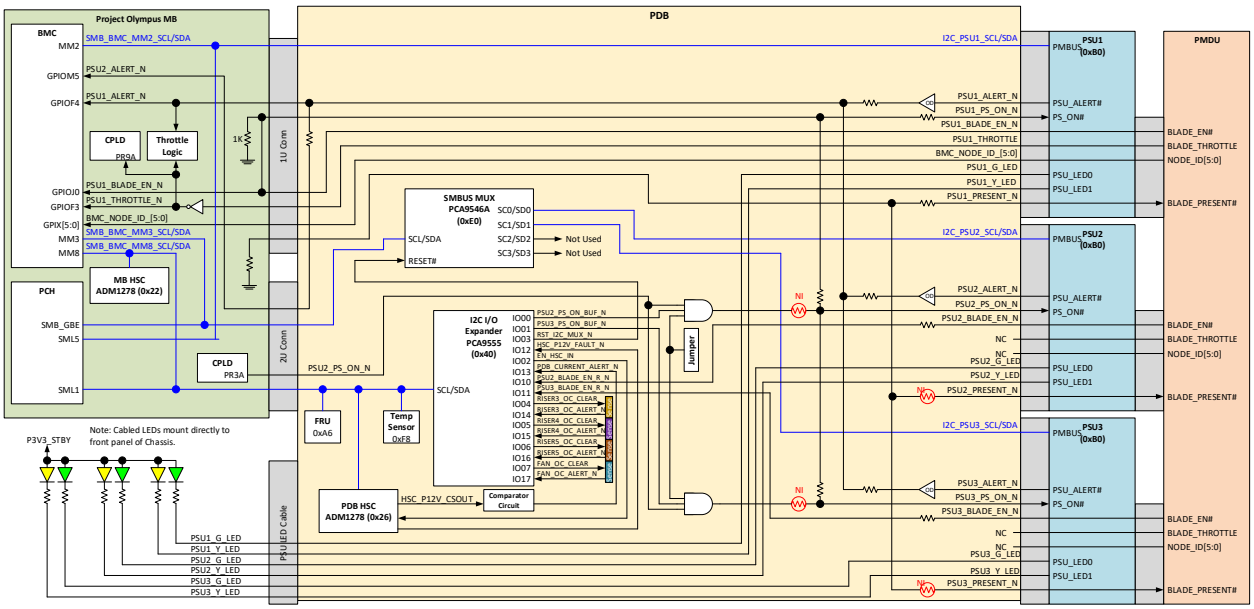


Figure 37: Power Distribution Board Miscellaneous IO Block Diagram

6.8 Split Cable Power Harness for PCIe Cards

The section below describes the cable power harnesses used for the high power PCIe cards (150W, 225W, 300W). Depending on the pinout of the power connector on the PCIe Add-In card, a different power harness may be required. 12V Cable Power Harnesses are not required for standard 75W PCIe Cards.

Different 12V Split Cable Power Harnesses will be designed to support delivering power from the PDB to two high power PCIe Cards per PCIe Riser. Each high power PCIe Card will include one or more of 2x3 or 2x4 12V Auxiliary connectors to support additional power above the base 75W PCIe Slot power. Below is an example of a 12V Split Cable Power Harness for a high power PCIe Card that supports two 2x4 12V Auxiliary connectors.

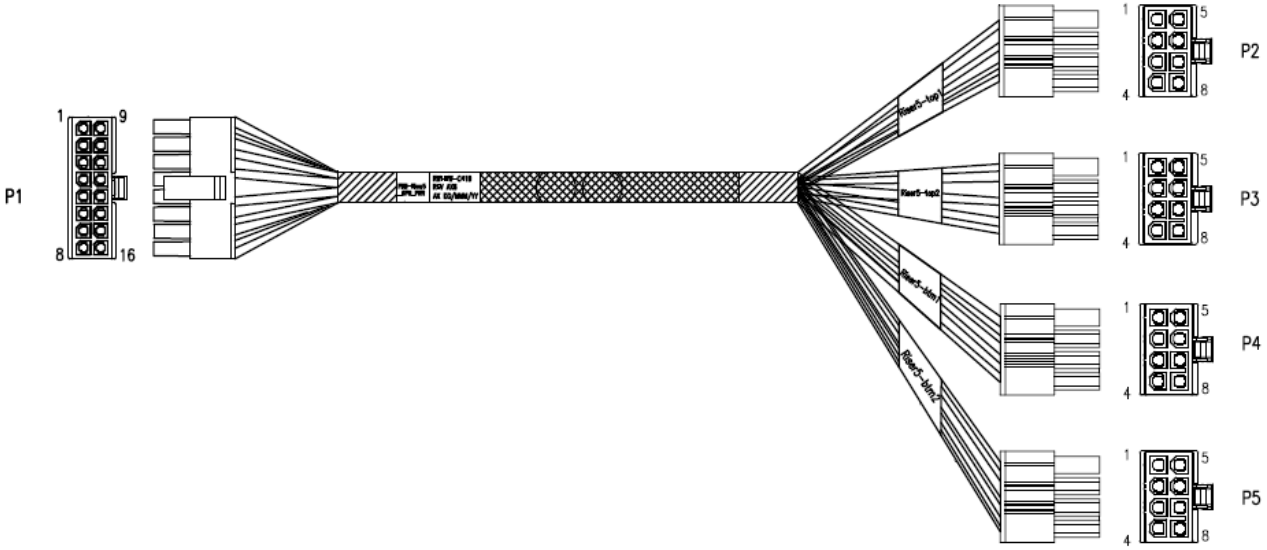


Figure 38: Split Cable Power Harness for 300W High Power PCIe Card (Example)

6.9 Front Panel PSU LED Cable

A small Front Panel LED Cable will support 3 x Bi-Color LEDs to provide status on the 3 different Project Olympus PSUs. The Front Panel PSU LED Cable will be attached to the PDB and snap into apertures in the front right side of the 3U Chassis. All System diagnostic LEDs must be visible from the Front of the Chassis (cold aisle).

Table 12: PSU Status LED Descriptions

LED Signal Name	Color	Description
PSUx_GREEN_LED	Green	Solid On = AC and DC Power Good Blinking = Battery Power Good
PSUx_YELLOW_LED	Yellow	Solid On = Failure of 1 PSU Phase Blinking = Failure of 2 PSU Phases

NOTE: Since this system only supports non-LES PSUs, the Battery LED states are not applicable.

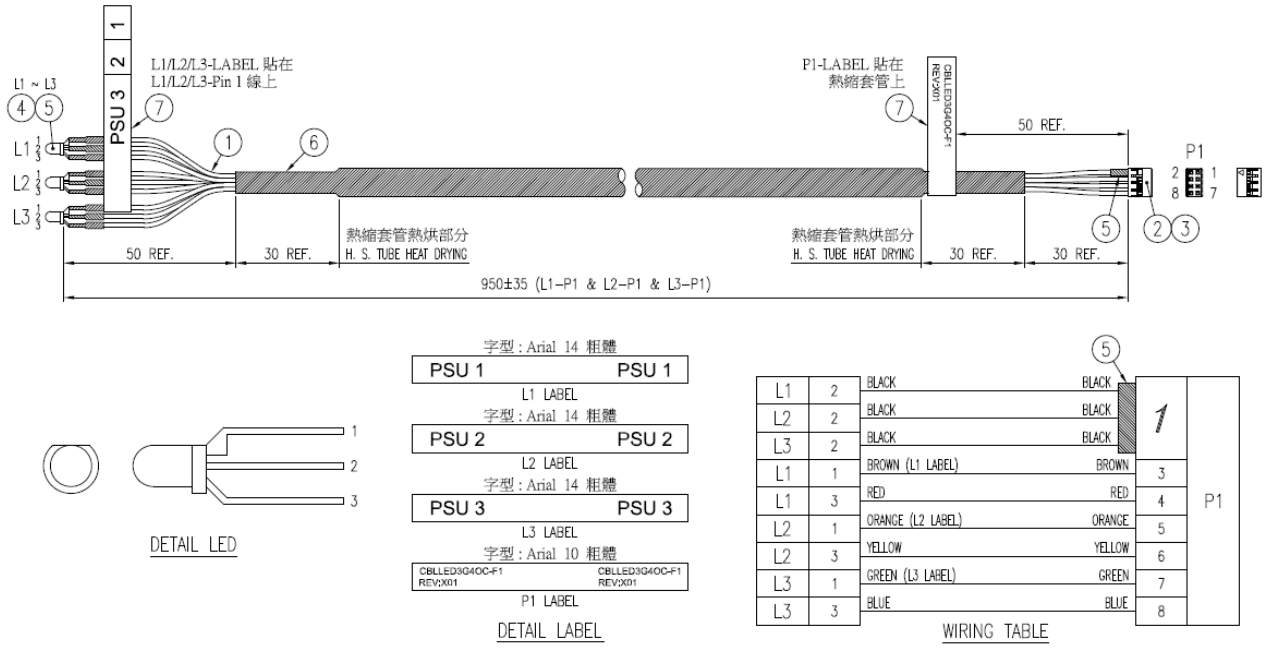


Figure 39: Front Panel PSU LED Cable

6.10 System Power Capping ("Quick Response")

The Project Olympus 3U Server supports Power Capping as an extension of the Project Olympus Universal MB feature. For MB Power Capping triggers and flows, please refer to the Project Olympus Universal MB Specifications. This section only covers PSU and PDB level triggers of the 3U Server, which are intended to support the "Quick Response" Power Capping feature (i.e. less than 10ms).

The main trigger of a "Quick Response" System Power Capping event in the 3U Server is when a 12V Over Current (OC) event (Fault or Warning) is indicated in the STATUS_IOUT command of any one of the three Project Olympus PSUs. In most cases, the Power Capping feature will not be needed, but it is intended as a precaution to avoid shutting down the System, until the event is over (if temporary), or the System can be properly serviced (if permanent). If an OC Event should happen, one of the three PSUs will assert its PSU_ALERT_N pin as a notification to the System to take action. As described in the PDB Miscellaneous IO Section of this Specification, the PDB combines the PSU_ALERT_N pins of all three Project Olympus PSUs and cables both the PSU1_ALERT_N (drives CPU_PROCHOT_N pins) and PSU2_ALERT_N (drives PCIe PWRBRK_N pins) pins on the MB for instant CPU and PCIe Card throttling. The CPU and PCIe Cards are two of the highest power consumers in the system. So, by throttling these devices, the overall 12V current should be reduced enough for the System to ride through the OC Event without shutting down, thus maintaining high availability. While the BMC monitors this System Power Capping Event, it is not involved in the flow. For System Power Capping to help the 3U Server ride through an OC Event, the response time of the CPU and PCIe Card throttling must be approximately 10ms or less. The Power Capping flow through the BMC would not fall within this 10ms window, therefore the use of the immediate hardware PSU_ALERT_N signal is needed, or the System would shut down.

There are various causes of an OC Event that occur when System Devices drive up the power based on peak workloads, failed components, or some combination of these. One such condition that can trigger an OC Event is when one or more PSU Modules fail, and the remaining PSU modules are expected to carry the load of the System. The 3U Server supports three PSUs, each with three internal PSU Modules for a total of 9 x PSU Modules. The 3U Server supports an 8+1 redundancy model, but under lighter loading conditions could also support a 7+2 or even a 6+3 redundancy model. Whatever the case may be, any time a PSU module is lost, the current draw on the remaining PSUs will increase. Depending on the System Loading conditions, or how many PSU Modules fail, the current can increase enough to cross the PSU OC Limit, thereby asserting the PSU_ALERT_N signal and forcing the System Power to be reduced.

NOTE: There is no PSU_ALERT_N signal assertion from the Project Olympus PSU when one or more PSU Modules fail. The PSU_ALERT_N signal is only asserted for OC events. It is not asserted for any other PSU failure events (over temp, Vin fault, Fan failure, etc…). If one or more PSU Modules fail, a SEL Event is logged by the BMC FW (during 1 second polling intervals), but there is no System Power Capping triggered; provided the remaining PSU Modules can handle the load. System Power Capping is only triggered when the PSU OC limit is tripped, which could potentially be caused by losing one or more PSU Modules depending on the System loading conditions. This means Power Capping is only used when the System needs it to protect the PSUs from shutting down in an over load situation.