Open Accelerator Infrastructure (OAI) - Expansion Module (EXP) Base Specification r2.0 v1.0

Author: OCP OAI Workstreams

Table of Contents

1 License 4
2 Compliance with OCP Tenets 5
2.1 Openness 5
2.2 Impact 5
2.3 Scale 5
2.4 Sustainability 5
3 Acknowledgements 6
4 Overview 7
4.1 Scope 7
4.2 Acronyms 7
5 High-Level Specification for the EXP Module 8
6 EXP Mechanical Specifications 9
6.1 PCBA Form Factor 9
6.2 ExaMax2 (as reference) Connector 12
6.2.1 Connector Guidance: Housing Guide Features 12
6.2.2 Standalone Guide Modules 13
6.2.3 Mated Connector-to-Connector Via and Housing Locations 13
6.2.4 Mated Connector-to-Connector Via and Housing Locations. 14
6.2.5 Keep Out Zone 14
6.3 GenZ 2C Connector – EXP Module board Thickness – 2.36mm 16
6.3.1 Gen-Z 2C Connector 16
6.3.2 EXP Module Gold Finger 18
6.4 EXP Stiffener 19
6.5 EXP Ejector 20
7 Thermal Specification 22
7.1 Environmental Conditions 22
7.2 Single Module thermal characteristics 22
7.3 System level thermal characteristics 24
7.4 Options to improve EXP cooling capability 26
8 EXP Electrical Specification 28

8.1 EXP Connectors 28
8.2 EXP Connector Pinout Quadrants 28
8.3 EXP Pinout Description 28
8.4 EXP Power Profiles 30
8.4.1 Thermal Design Power TDP 30
8.5 System power sequencing 30
8.6 EXP Insertion Loss 32
8.7 EXP PCB Stack-Up 32
9 EXP Interconnect Topologies 35
9.1 Module ID 35
9.2 Interconnect Topology 35
9.2.1 11-Port (7+4) FC/Retimer Topology (Fully Connected + 4 Scale Out) 36
9.2.2 8-Port Switch-Based Topology 36
10 EXP reference system design 38
11 EXP Management and Security Requirements 39
11.1 Management Interface 39
11.2 Sensor Reporting 39
11.3 Error Monitoring/Reporting 40
11.4 Firmware Update 41
11.5 FRU Information 41
11.6 IO Calibration 41
11.7 EXP Security Requirements 42
11.7.1 Secure Boot 42
11.7.2 Recovery 43
11.7.3 Debug Capabilities 43
11.7.4 Attestation 43
12 Environmental 44
12.1 Environmental Requirements 44
12.2 Regulation 44
13 Revision History 45

1 License

Contributions to this Specification are made under the terms and conditions outlined in Open Web Foundation Contributor License Agreement ("OWF CLA 1.0") ("Contribution License") by:

AMD, Alibaba, Amphenol, Baidu, Broadcom, Enflame, Graphcore, H3C, Inspur, Intel, Intel Habana, Luxshare, Meta, Molex, Nallasway, Supermicro, TE, Wiwynn

Usage of this Specification is governed by the terms and conditions outlined in the Open Web Foundation Final Specification Agreement ("OWFa 1.0").

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

None.

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 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.

2 Compliance with OCP Tenets

2.1 Openness

EXP Base Specification is an open standard with joint contribution across Hyperscalers, ODMs, component suppliers, and connector vendors.

2.2 Impact

EXP Base Specification defines common form factor for multi-node scale-out and greatly cut down design cycle time and resources for companies adopting the standard.

2.3 Scale

EXP Base Specification defines power consumption up to 400W with form factor and connector selection to support future AI/HPC platform architecture.

2.4 Sustainability

EXP supports both air, liquid, or hybrid cooling for existing data center deployment and improve facility PUE and WUE when data center is equipped with full liquid cooling solution.

3 Acknowledgements

We want to acknowledge all the OCP OAI Workstream members for their contributions to this specification: The incredible collaboration between customers, accelerator manufacturers, system developers, and industry partners shows how Open Compute develops industry-standard form factors and specifications that benefit all its members.

We would especially like to thank **AMD, Amphenol, Broadcom, H3C, Intel, Meta, Microchip, TE, and Wiwynn** for their extra efforts in putting this specification together.

4 Overview

Artificial Intelligence (AI) applications are rapidly evolving and producing an explosion of new types of hardware accelerators for Machine Learning (ML), Deep Learning (DL), and High-performance Computing (HPC).

Different implementations target similar requirements for power/cooling, robustness, serviceability, configuration, programming, management, debug, inter-module communication to scale-up, and input/output bandwidth to scale out.

This base specification outlines an interoperable, modular hierarchy based on the EXP (Expansion Module) form factor to enable 32*QSFP-DD/OSFP connectors for scale-out.

We invite open contributions in the following areas:

1. Base specification (OCP Accelerator Infrastructure Sub-Project Specification)
2. Design specification (This document, detailed description of alternative, interoperable components which meet the base specification)
3. Products (schematic, layout, mechanical/thermal solutions, and firmware/software to realize the above designs)

EXP modules have two configurations with the same dimension, PHY retimer configuration and switch configuration. Two configurations are differentiated by:

- PHY retimer configuration uses 112Gbps retimer devices while switch configuration uses a switch device.
- Different cooling solution based on different configuration.
- Scale out connector can be QSFP-DD or OSFP
- PHY retimer configuration with one ExaMax2 connector while switch configuration with two ExaMax2 connectors.

4.1 Scope

The OAI-EXP base specification defines the form factor, standard specifications for expansion module, and design interfaces.

4.2 Acronyms

Acronym	Definition
ASIC	Application Specific Integrated Circuit
EXP	Expansion Module
BMC	Baseboard Management Controller
TDP	Thermal Design Power
MPN	Manufacturing Part Number
DXF	Drawing eXchange Format
PCBA	Printed Circuit Board Assembly
UBB	Universal Baseboard
OAI	Open Accelerator Infrastructure

5 High-Level Specification for the EXP Module

Module PCB Dimension	160mm (W) x 148mm (H) x 2.36mm(T)
Module Power/Input Voltage	Supports up to 400W, with 44V-59.5V DC as input power
Connectors	ExaMax2*2, GenZ 2C 2.36mm*1, QSFP-DD/OSFP*4
OAM to EXP Interconnect Links	8 Links from OAM to EXP, each link with up to x8 lanes, 112Gbps per lane

6 EXP Mechanical Specifications

This section describes the EXP form factor. 2D DXF and 3D files are included in the contribution package, with relevant reference drawings to mechanical components. Please refer to those files for further details. Also note that some EXP features are required, some others are optional for reference purpose only.

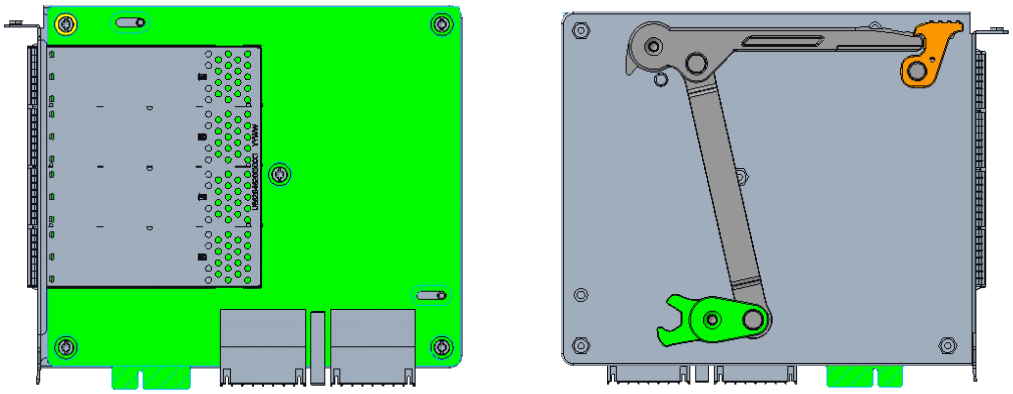


Figure 1 EXP top & bottom view

6.1 PCBA Form Factor

This section covers the required and recommended dimensions of the module PCB and component placement. Figures 2 and 3 illustrate the EXP form factor of 160mm x 148mm x 2.36mm. ExaMax2 connectors and active components are all on the top side. Five PTH mounting holes attach the module to a stiffener secured below the system PCB. These mounting holes should provide clearance for an M3.0 screw with enough thread length to secure the bottom stiffener.

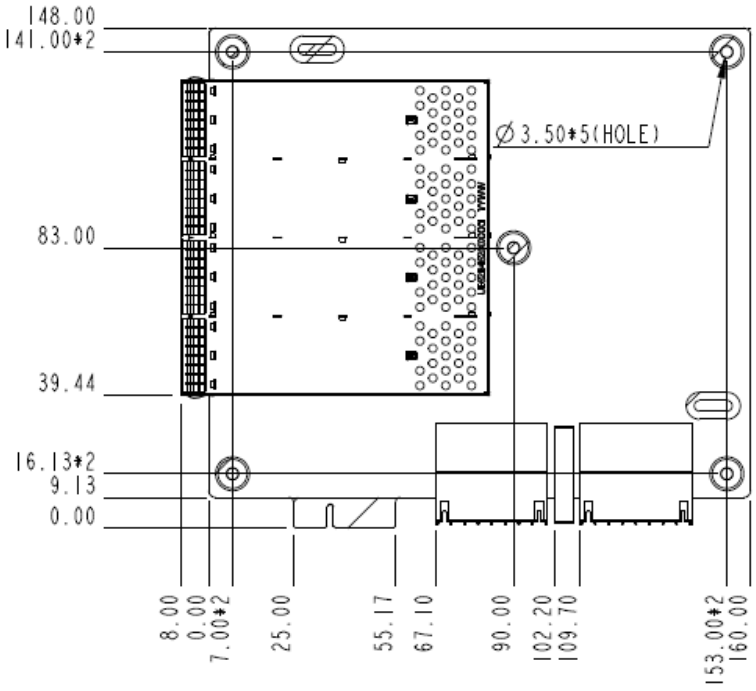


Figure 2 EXP Form Factor Dimensions, Top View

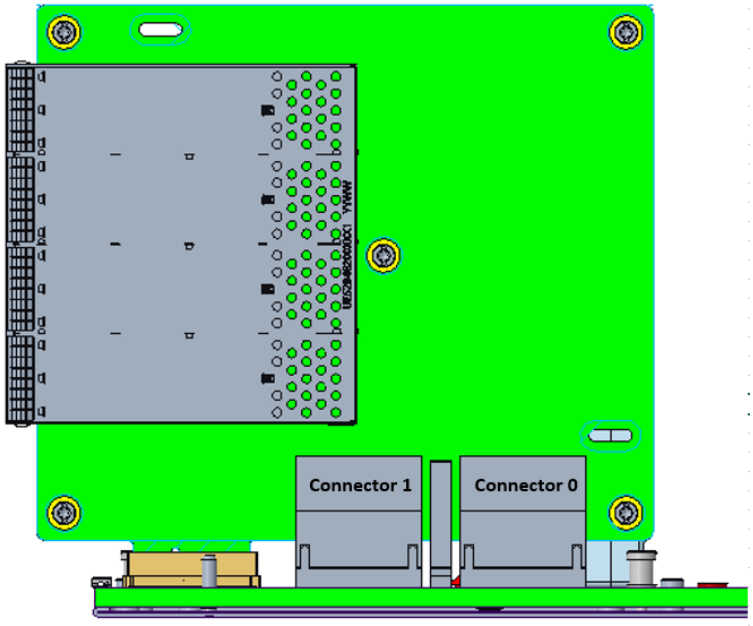


Figure 3 EXP Form Factor, Top View with UBB

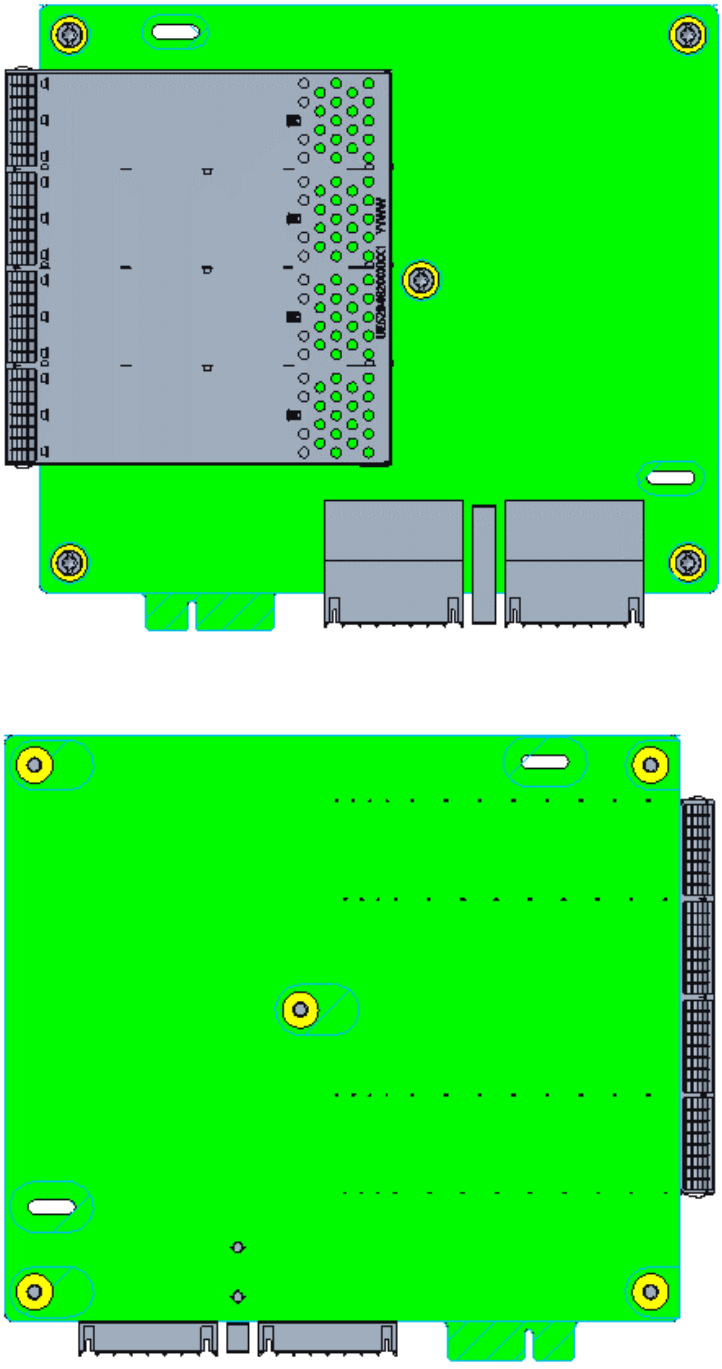


Figure 4 Top and bottom views

6.2 ExaMax2 (as reference) Connector

Amphenol ExaMax2 (as reference) product comprises the mating interface of 4*16 (64 Diff pairs) right-angle receptacle (MPN 10167059-101LF as reference) on the EXP module card side, and the 4*16 (64 Diff pairs) vertical header (MPN 10167063-101LF as reference) on the Universal Base Board.

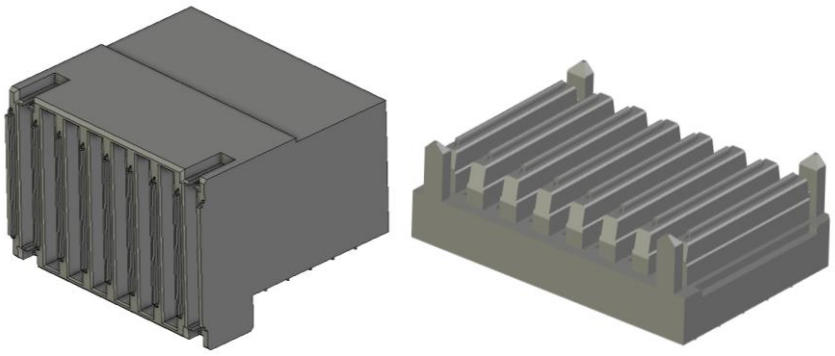


Figure 5 ExaMax2 (as reference) 4*16 Right Angle Receptacle (RAR) 10167059-101LF shown on the Left & ExaMax2 (as reference) 4*16 Vertical Header (VH) 10167063-101LF shown on the Right are provided courtesy of Amphenol.

- The force to mate a receptacle connector and compatible header shall not exceed 0.45 N per contact.
- The un-mating force shall not be less than 0.10 N per contact.
- 64 Total Differential Pairs

6.2.1 Connector Guidance: Housing Guide Features

Table below shows the amount of misalignment in each direction that can be tolerated by the guiding features on the connector housings as they are mated. It is important to note that this misalignment is not rigid misalignment. The chassis design must permit the card modules to move freely so these housing guide features can perform final alignment within clearances of separate guide modules and integrated guide pins.

Table 1 The Connector's Misalignment in Each Direction

Connector type	Nominal misalignment correction in direction parallel to connector columns	Nominal misalignment correction in direction perpendicular to connector columns
EXAMAX2 4- and 6-pair	± 1.4 mm	± 1.2 mm

The maximum acceptable angular misalignment of the EXP module relative to the UBB is ± 2°.

6.2.2 Standalone Guide Modules

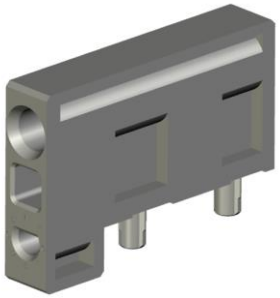


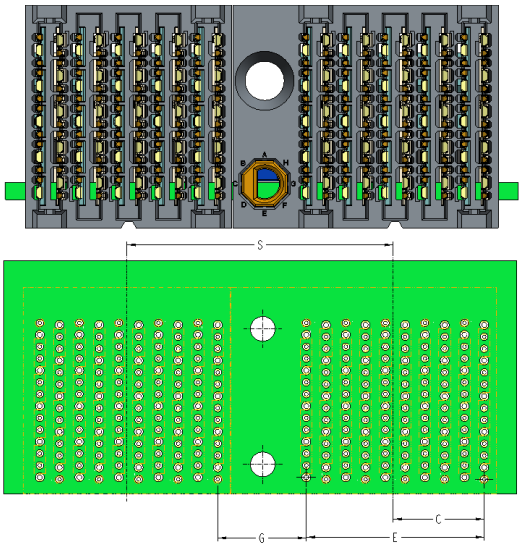
Figure 6 ExaMax2 (as reference) Guide Socket (MPN: 10158783-109LF) are provided courtesy of Amphenol.

Separate metal guide module is used to maximize a system's mechanical robustness.
Separate guide modules will allow for additional misalignment in each direction. The system consists of a round metal guide pin placed next to the male connector and a corresponding round hole with a generous lead in next to the receptacle connector. This system provides a nominal diametrical misalignment of Ø5.0mm, or ± 2.5 mm in any direction and a wipe length of 15.2 mm after the guide pin is fully engaged into the guide hole.

Table 2 The Guide module's Misalignment in Each Direction

Guide Width	Nominal misalignment correction in direction parallel to connector columns	Nominal misalignment correction in direction perpendicular to connector columns	Wipe length after guide pin is fully engaged in guide receptacle. See Table 6.
5.5 mm	±2.5 mm	±2.5 mm	15.2 mm

6.2.3 Mated Connector-to-Connector Via and Housing Locations



	Columns	2.0 Column Pitch, 4-& 6-pair			
		C	E	G	S
Option II: With 5.5mm Guide	8	7.0	14.0	8.4	22.4
	12	11.0	22.0	8.4	30.4
	16	15.0	30.0	8.4	38.4

Figure 7 ExaMax2 (as reference) Connector to Connector Spacing with Guide option

Option of guide module polarization key option as shown in below.

Table 3 Options of Guide Module

PART NUMBER	DESCRIPTION KEY
10158783-101LF	A
10158783-102LF	B
10158783-103LF	C
10158783-104LF	D
10158783-105LF	E
10158783-106LF	F
10158783-107LF	G
10158783-108LF	H
10158783-109LF	NO KEY FLAT

6.2.4 Mated Connector-to-Connector Via and Housing Locations.

The relative location of the header to the receptacle vias and the outside walls is dependent upon the inter-mating connector types. While the housing and PCB real estate varies by connector size (number of differential pairs), the locations between mated connectors remain constant as per shown in Figures 8 below.

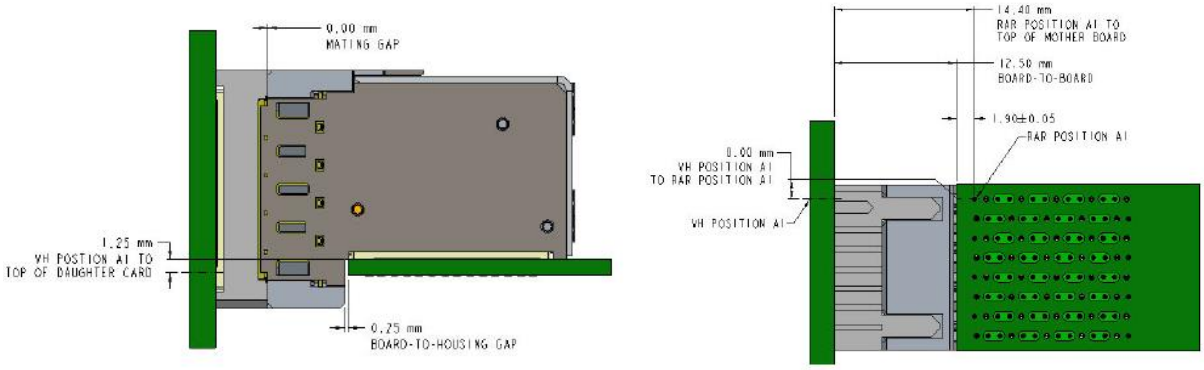


Figure 8 Backplane EXAMAX2 (as reference) VH to EXAMAX2 (as reference) RAR Mated Connector-to-Connector Locations

6.2.5 Keep Out Zone

For the removal process, it is recommended that the RAR should have keep out zones a minimum of 8.0 mm at the back of the connector with 11.0 mm on each side of the end connectors. Refer to Figure 9 and 10 for additional details.

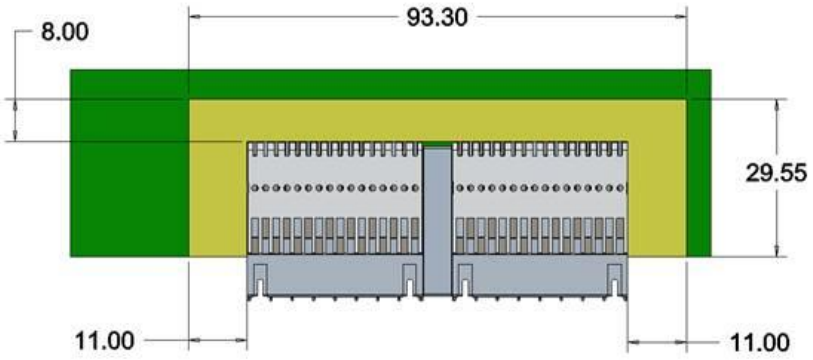


Figure 9 ExaMax2 (as reference) Keep Out Zone Requirement on EXP module side

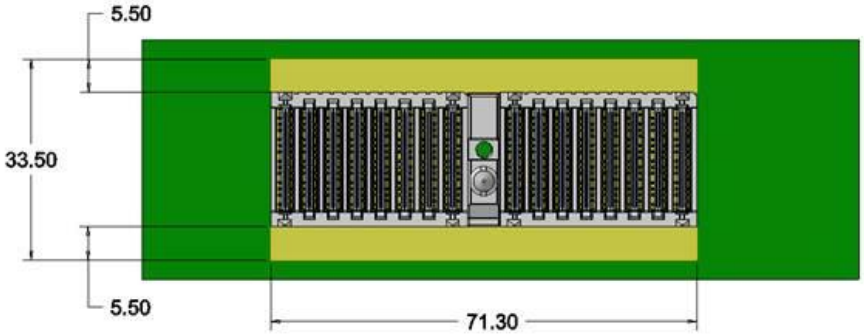


Figure 10 ExaMax2 (as reference) Keep Out Zone Requirement on UBB mating side

6.3 GenZ 2C Connector – EXP Module board Thickness – 2.36mm

The Gen-Z 2C 2.36mm connector will be published in SFF-TA-1020.

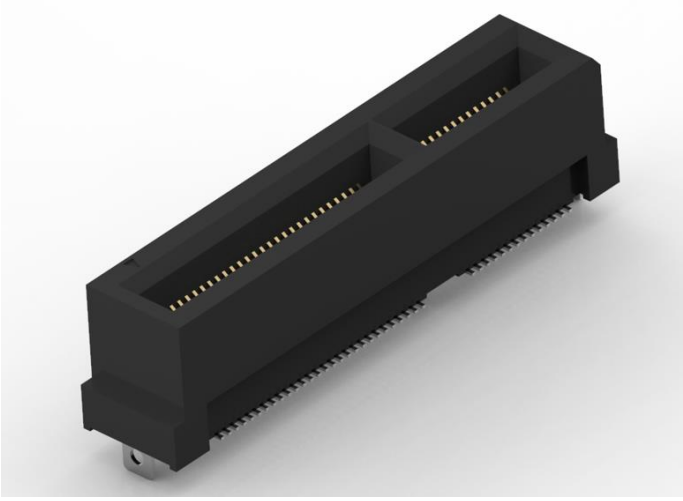
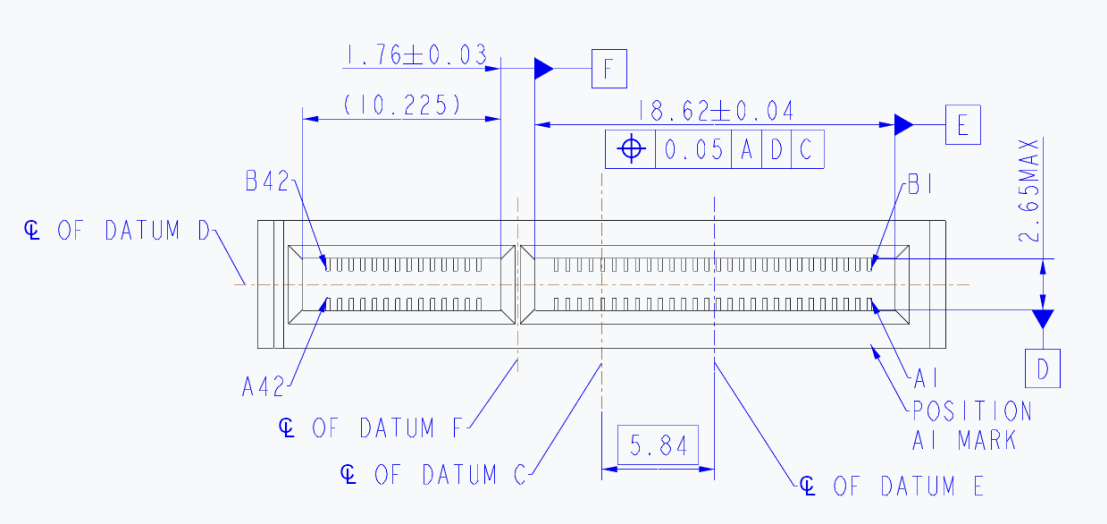
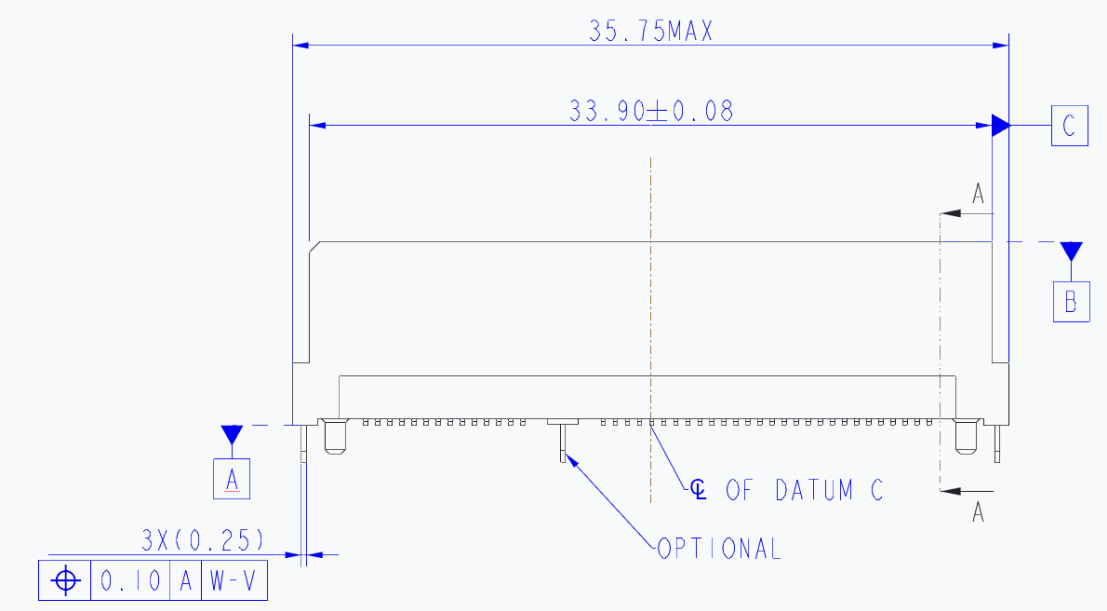
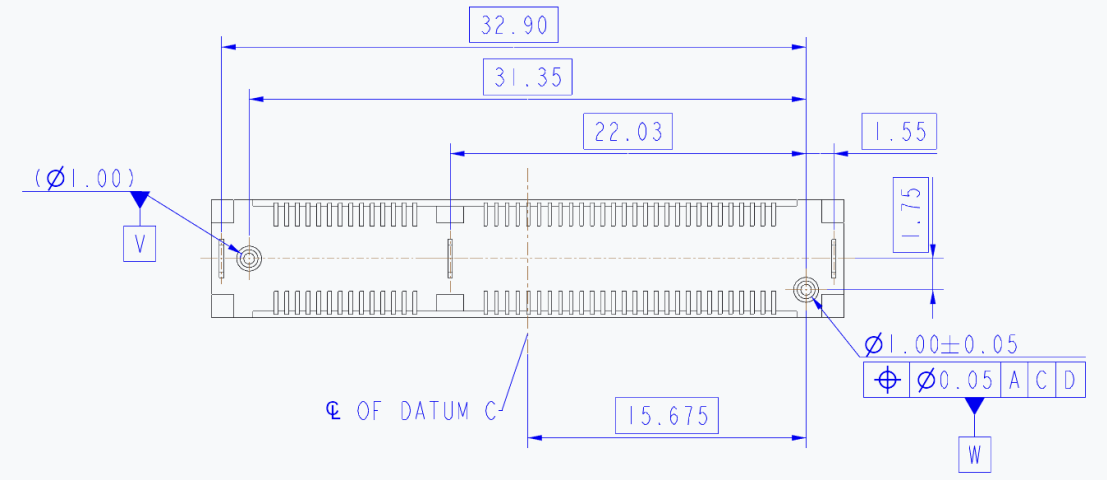


Figure 11 Gen-Z 2C for EXP sideband and power

6.3.1 Gen-Z 2C Connector




35.75MAX
33.90±0.08
C
A
B
A
3X(0.25)
0.10 A W-V
℄ OF DATUM C
OPTIONAL
A


32.90
31.35
22.03
1.55
(Ø1.00)
V
1.75
Ø1.00±0.05
Ø0.05 A C D
W
℄ OF DATUM C
15.675

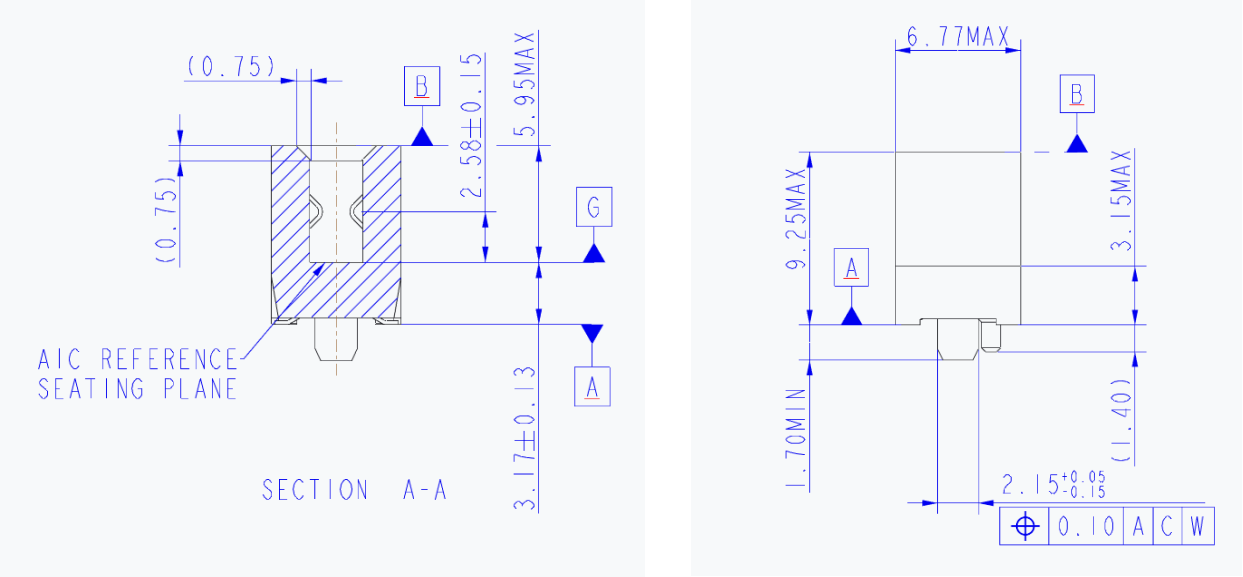
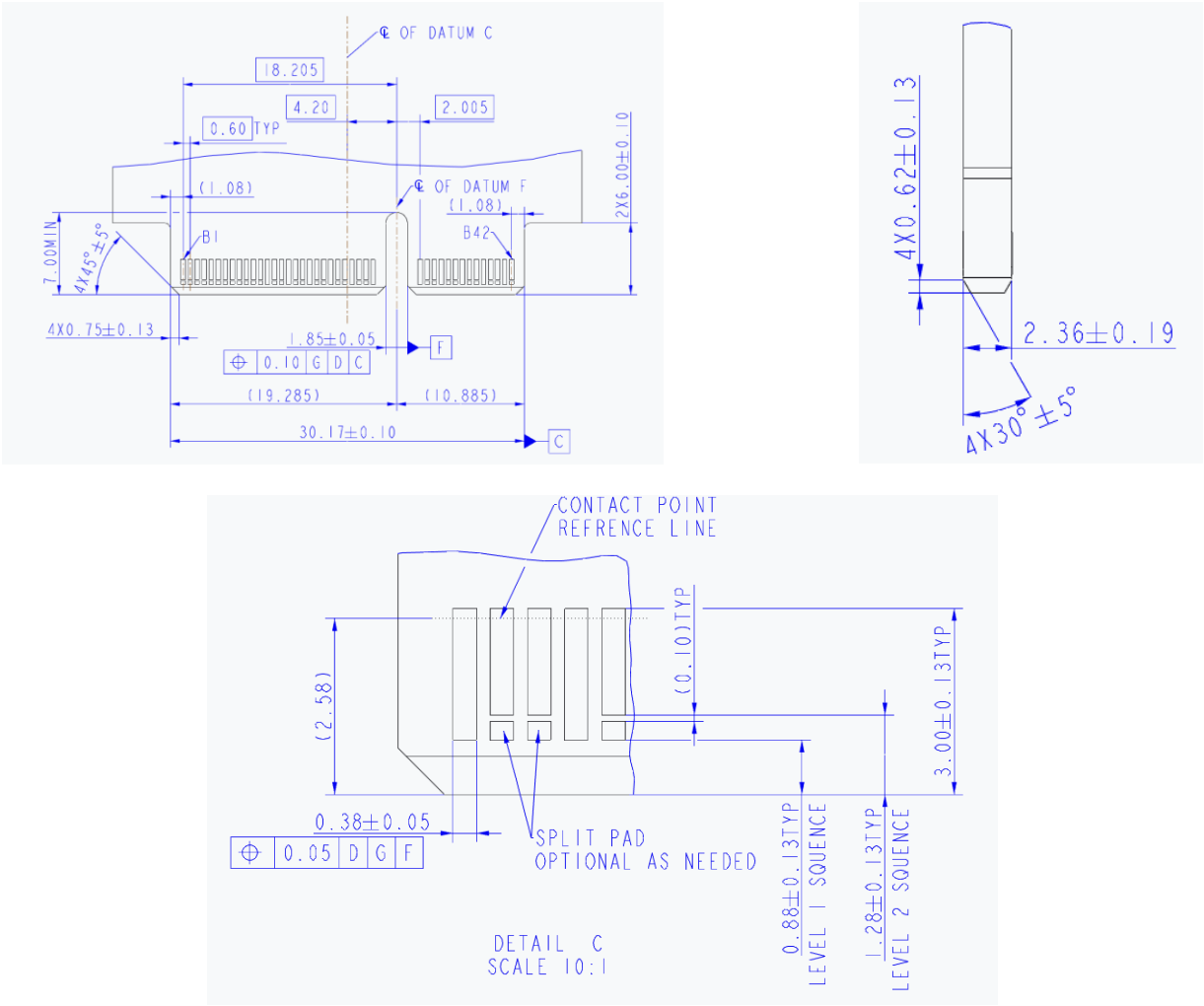


Figure 12 Gen-Z 2C connector drawing

6.3.2 EXP Module Gold Finger



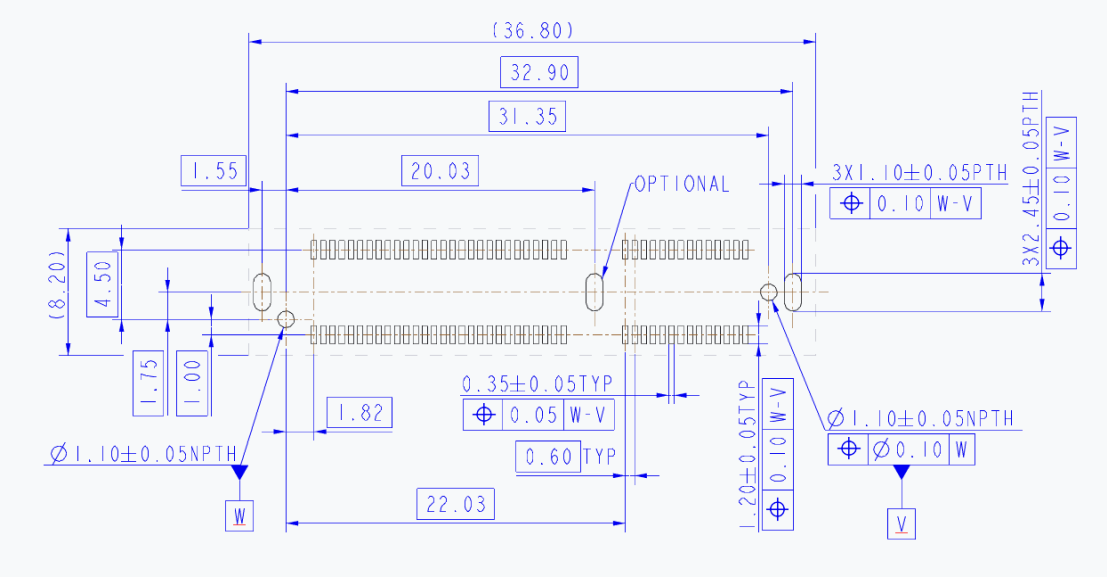


Figure 13 EXP Module Gold Finger drawing

NOTES: POSITION A1 ON OPPOSITIE SIDE OF CARD OF B1. DIMENSION FOR PAD LOCATIONS TO CENTER OF THE PAD

6.4 EXP Stiffener

EXP stiffener design is for reference only, and dimensions may be changed or adjusted to accommodate the specific application and board layout of the EXP design.

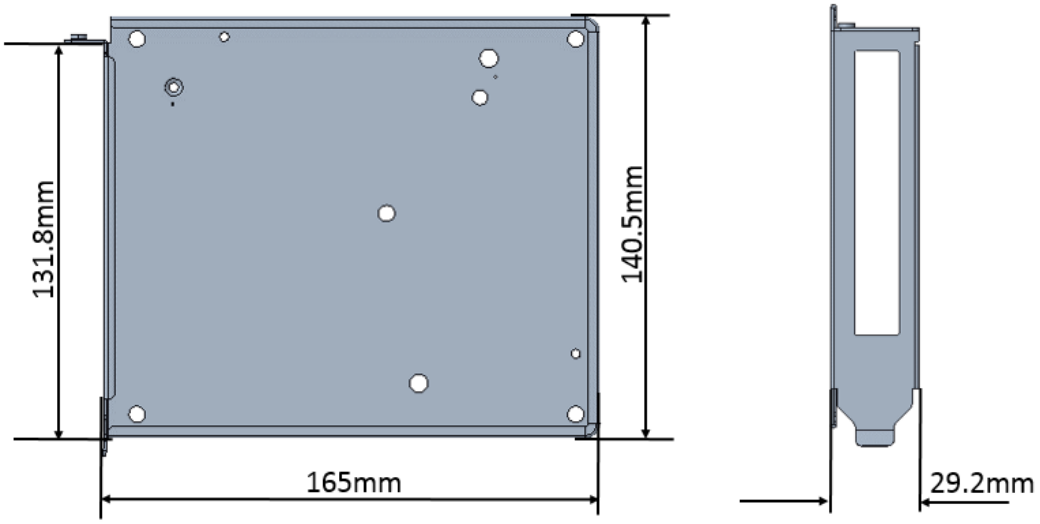


Figure 14 EXP Stiffener Dimensions,bottom & side view

6.5 EXP Ejector

EXP ejector design in Figure 15 and dimensions are shown in Figure 17 Bottom Stiffener Dimensions. Other features and dimensions of the ejector shown in the reference model are optional and can be adjusted based EXP design requirements.

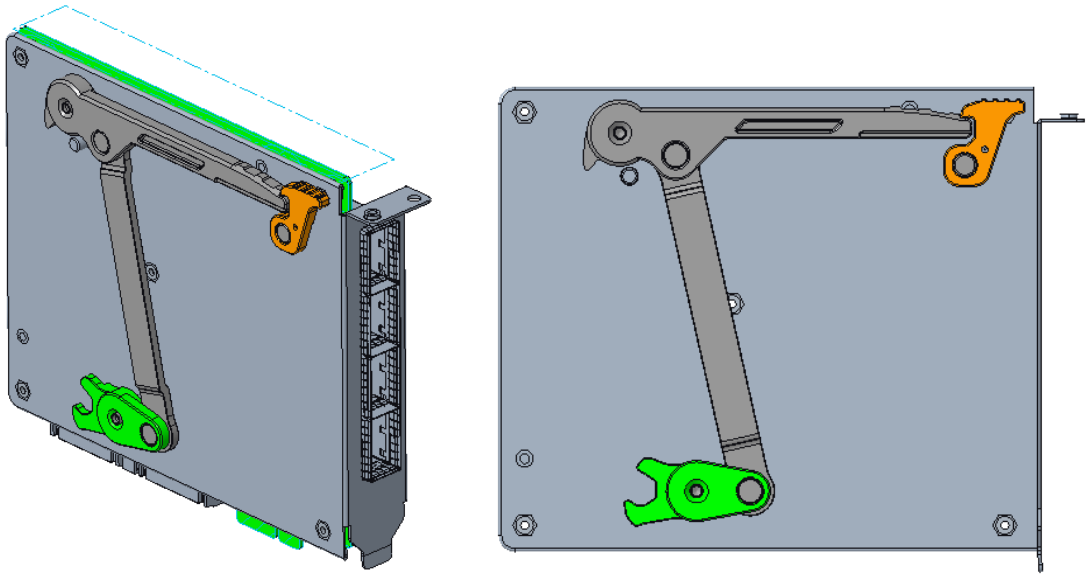


Figure 15 Ejector Reference Design

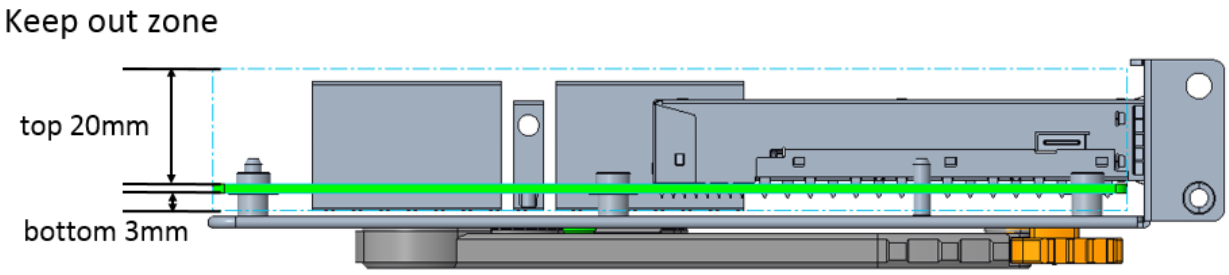


Figure 16 Ejector Reference Design of Side View

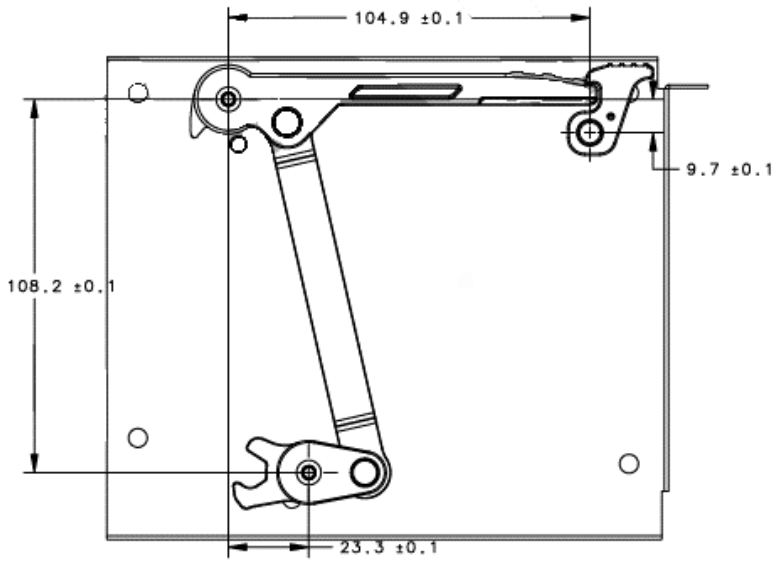


Figure 17 Bottom Stiffener Dimensions with Ejector

7 Thermal Specification

The thermal requirements are applicable to air-cooled conditions by default, except where a different cooling approach is specifically called out.

7.1 Environmental Conditions

To meet the thermal reliability requirement, the cooling solution should dissipate heat from the components when the components on EXP Module are operating at their thermal design power. EXP module should be able to work in the following environmental conditions without any throttling or thermal issues, given enough airflow supply:

- Approach Air Temperature: 5°C to 55 °C
- Altitude: sea level to 3000 ft*, without temperature derating
- Relative Humidity: 20% to 90%

*: support altitude up to 6000ft is recommended

Cold boot temperature: the module should be able to boot and operate at an initial temperature of 10°C

In addition, the module should remain unaffected at a non-operational storage temperature range of -20°C to 85°C.

7.2 Single Module thermal characteristics

The information in this section is provided as a reference guide for EXP 2.0 designers in the initial stages of the design process. It is intended to be used as a point of reference for thermal design and feasibility evaluation. However, Users should conduct further analysis based on their actual design to ensure sufficient thermal performance under real operating conditions.

The reference thermal simulation setup for the EXP module involves an OSFP MODULE located upstream and a retimer with heat sink located downstream. Each EXP module contains 4 OSFP modules,4 retimer modules and the VR module. The power consumptions of key components are listed as below.

- OSFP: 20W
- Retimer: 15W
- Power loss of VR module: 7.8W
- Maximum TDP of EXP Module: 147.8W

The reference thermal model for simulation includes the EXP module itself as well as the gap between adjacent EXP module (Figure 18). The following information provides reference results based on simulated analysis with the boundary condition as below:

- Local ambient temperature: 35°C for airflow direction 1 and 50°C for airflow direction 2
- Altitude: sea level
- Airflow rate range: 25 to 100 CFM
- Airflow direction: 2 directions, 1) from OSFP to retimer and 2) from retimer to OSFP

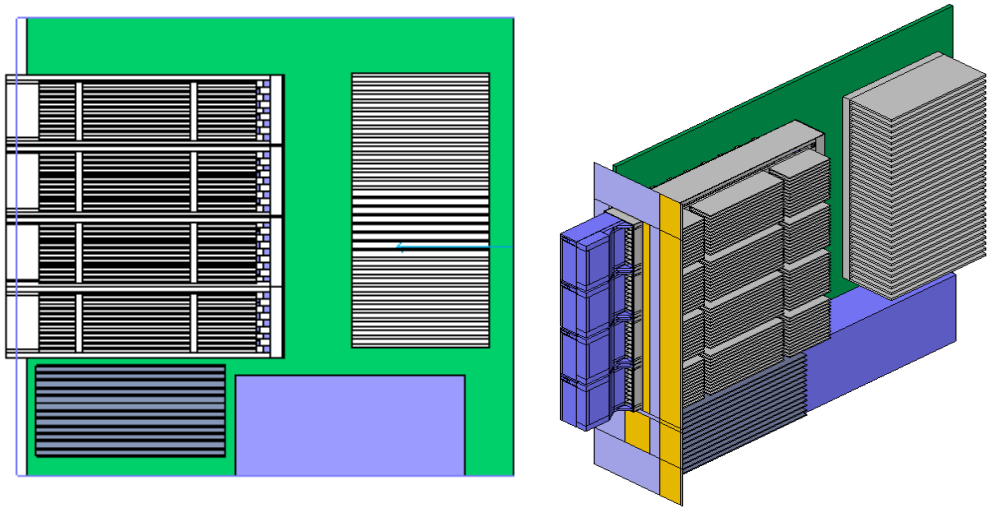
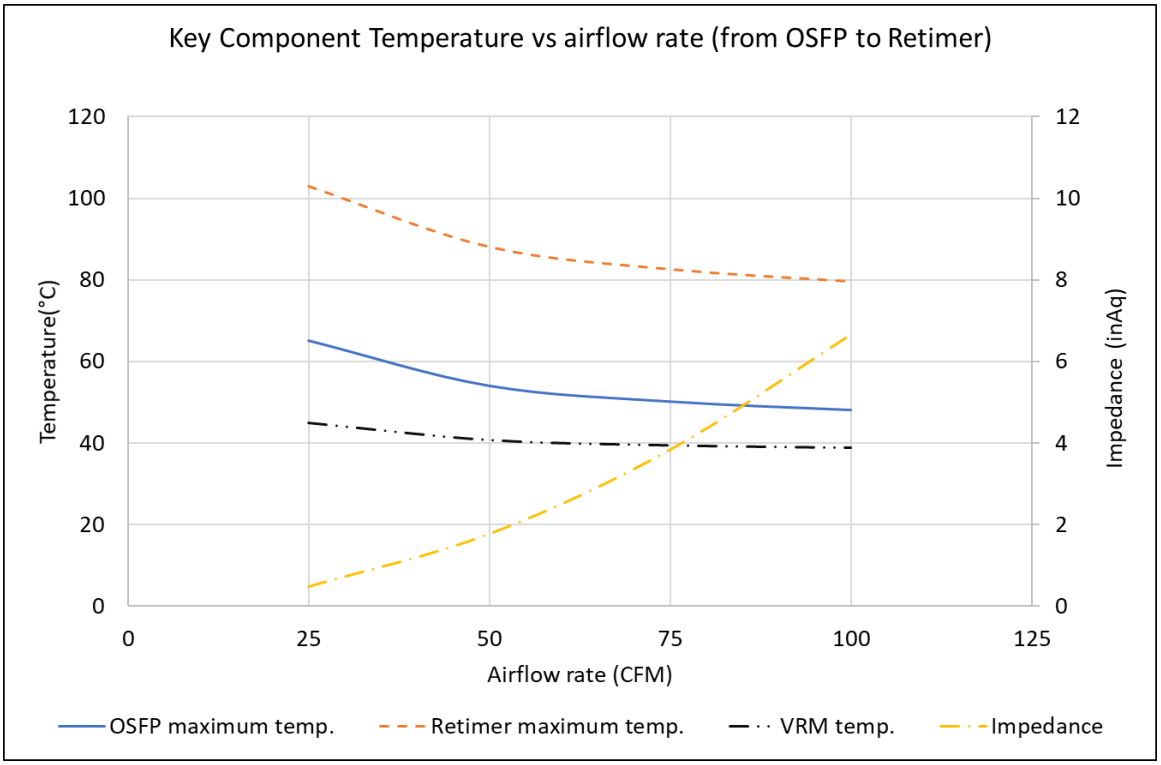


Figure 18 Layout of reference EXP module key components

Simulation results are shown as Figure 19. According to the results, OSCP modules are usually the critical component, since the suggested limit is typically at or below 70°C. At airflow direction 1), when the ambient air temperature approaching the module is at 35°C or below, 25 CFM would be sufficient for single module cooling. At airflow direction 2), if the local approach air temperature becomes at 50°C or above, however, the airflow required for OSFP cooling may lead to a pressure drop that becomes a heavy burden on the system fans. It is important to note that heatsink design and component TDP selection may impact the results significantly. Users should evaluate the cooling feasibility and airflow requirements based on their actual design.



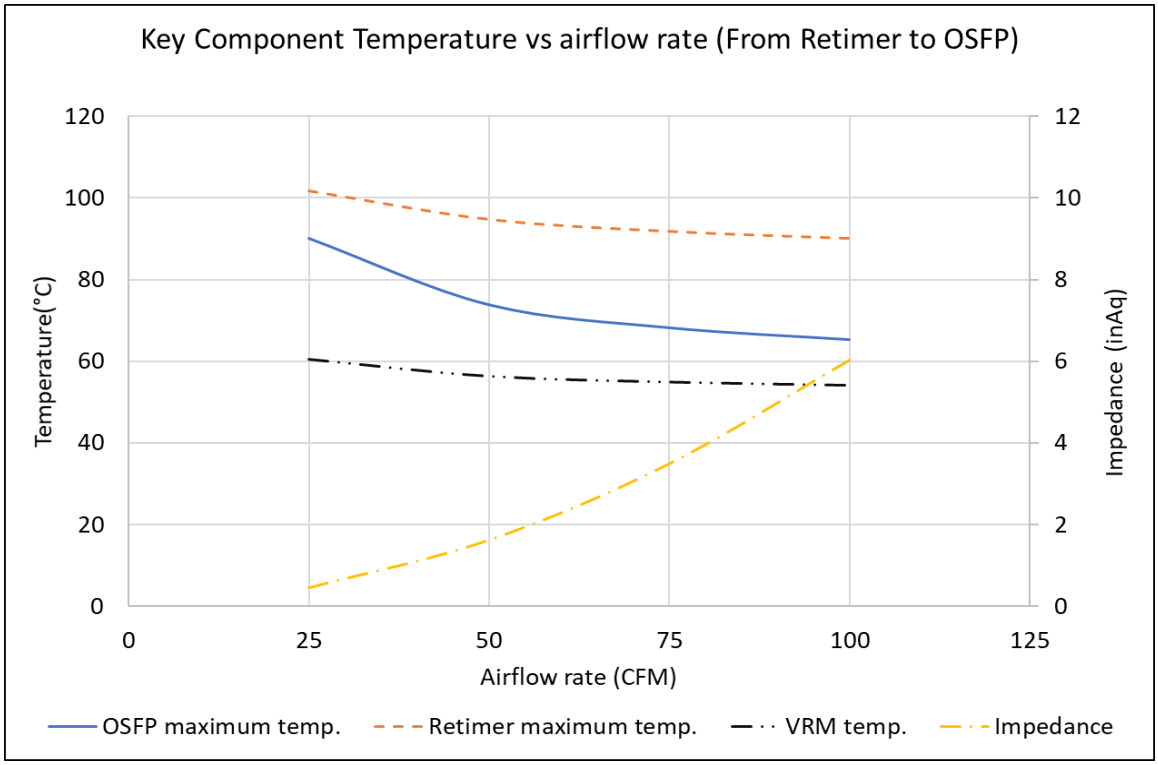


Figure 19 EXP Key component Temperature vs Airflow rate

7.3 System level thermal characteristics

In a real OAI system, multiple EXP modules will be installed to deliver the required processing and communication capabilities. The reference thermal model for UBB2.0 system includes the 8 EXP modules and 8 OAMs, and the airflow is driven by 8 80mm dual rotor fans. The following information provides reference results based on simulated analysis with the boundary condition as below:

- Ambient temperature: 35°C
- Altitude: sea level
- System height: 4OU
- Fan duty: 100%
- Airflow direction: 2 directions, 1) EXPs to OAMs; 2) OAMs to EXPs (shown as Figure 20)
- EXP module TDP: 147.8W
- OAM TDP: 600W
- MISC. power: 200W

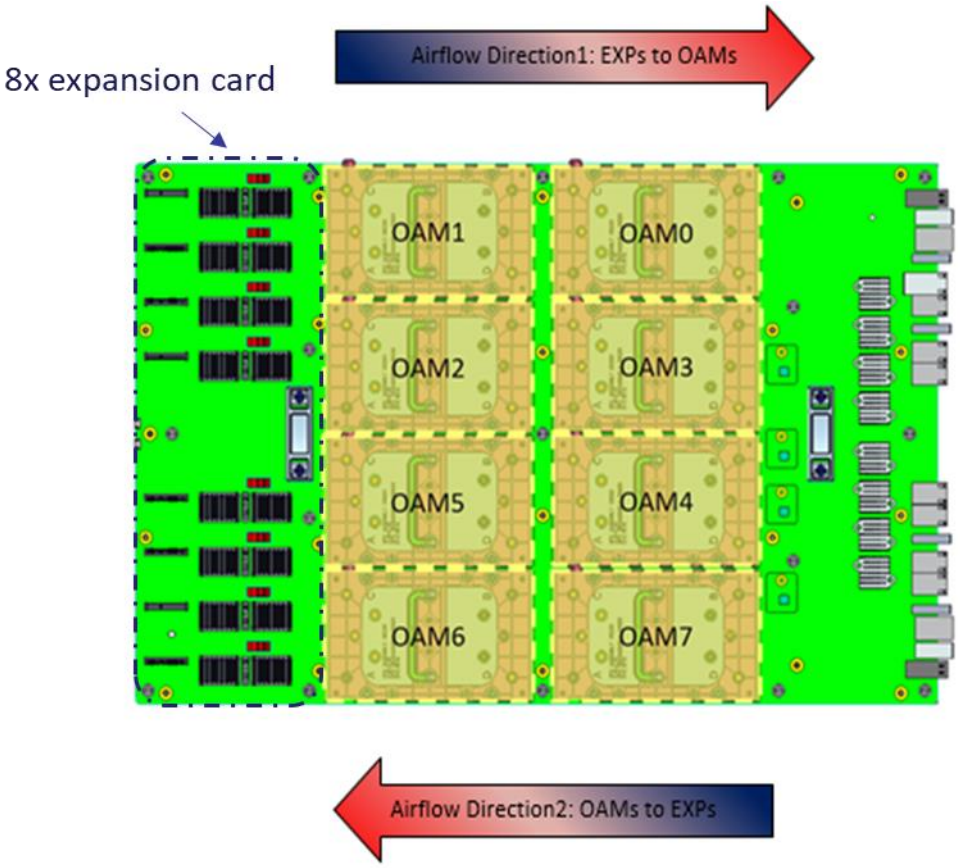


Figure 20 UBB at different airflow directions

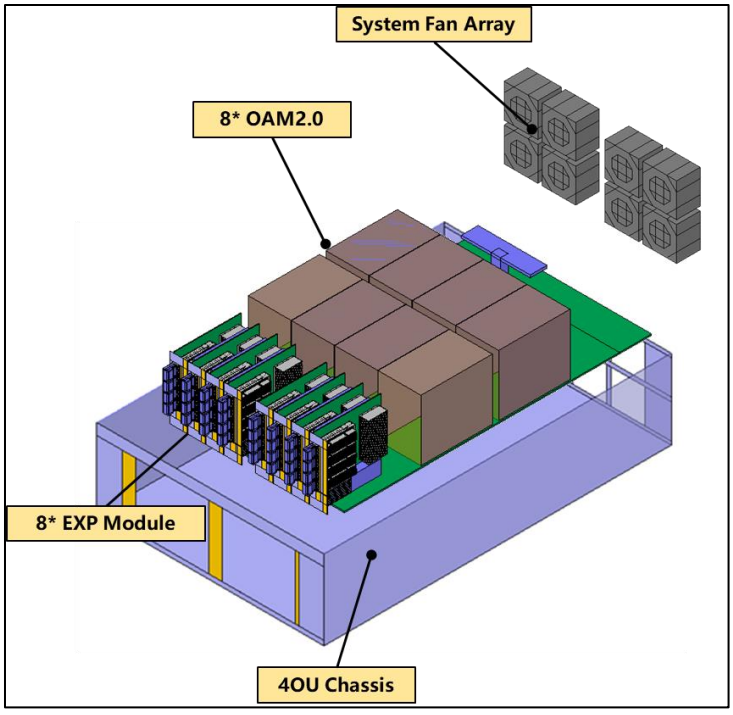


Figure 21 CFD model of the reference 4OU UBB2.0-based platform

System level simulation result is shown as Table 44. Based on the simulation results of the reference system, under airflow direction 1, each EXP module can receive airflow rates of 50 CFM or above, providing sufficient cooling capability for the key components on the EXP module. However, under airflow direction 2, the higher local ambient temperature may lead to OSFP temperature over spec. These observations align with the individual module analysis.

Table 4 Reference system level simulation result

Simulation Scenario		Airflow direction1	Airflow direction2
Hottest EXP Module airflow rate		50CFM	58 CFM
Hottest EXP Module local ambient temperature (TLA)		35°C	54°C
System airflow rate		700 CFM	625CFM
Components	TDP(W)	Temp(C)	Temp(C)
EXP module- Retimer Max. temp	15	96 (Tj)	102 (Tj)
EXP module- VRM Max. temp	7.8	42 (Tj)	62 (Tj)
EXP module- OSFP Max. temp	20	55 (Tc)	77 (Tc)

7.4 Options to improve EXP cooling capability

In summary, the thermal analysis of the EXP module highlights the critical importance of thermal management in system design. EXP module placement and airflow rate significantly impact operating temperatures. The simulation results indicate that placing the cards in a cold aisle is suggested.

In addition, there are still ways to further extend cooling capability for EXP module. One of the ways is to increase fan speed to supply more airflow rate, but fan power will be increased too. And another way is to use more fan for cooling, for example, if a 6U system include 4U UBB layer and 2U head-node as Figure 2222, whole system could share one fan wall with 12 pcs 80mm dual rotor fan, with proper airflow distribution optimization, EXP module in UBB layer could receive more airflow and extend the limitation.

Implement liquid cooling solution on OAM is another way to extend air cooling for EXP module especially for airflow direction2 as the preheat will be significantly reduced.

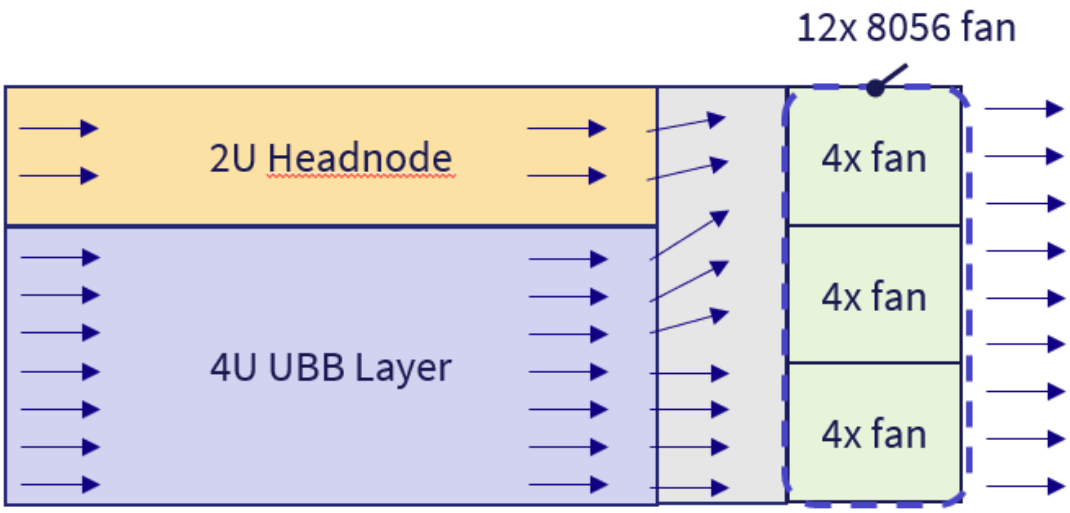


Figure 22 EXP module key component temperature vs airflow direction

8 EXP Electrical Specification

8.1 EXP Connectors

EXP supports two 4x16 ExaMax2 connectors for 112G SerDes interconnection. In PHY retimer configuration, only one ExaMax2 connector and one guide socket are needed. One 2.36mm gold finger PCB design is needed fit into Gen-Z 2C connector located on UBB, the connectivity on gold finger is for power, I2C, MDIO, SPI, JTAG, PCIe, and other side bands.

Table 5 EXP connector list

Vendor	Vendor PN	Description	Q'ty	Designation	R/V	Termination Type
Amphenol	10153370-101LF	EXAMAX2 EXP Connector 4X8 RAR	2	EXP Module 112G PAM4 signal	R	Press-Fit
Amphenol	10158783-109LF	Guide socket (EXP)	1	EXP Module Guide	R	Press-Fit + Screw

8.2 EXP Connector Pinout Quadrants

The gold finger supports the following interfaces:

- 54V/48V and 3.3V input power
- Other single-ended signals such as PRESNT#, SMBus, MDIO, SPI, JTAG, GPIOs etc.

The two ExaMax2 connectors have the following interfaces:

- In PHY retimer configuration, one connector supports up to 4x8 SerDes links
- In switch configuration, two connectors support up to 8x8 SerDes links

8.3 EXP Pinout Description

The detailed connector pin map is provided in separate spreadsheet (OAI-EXP_Pinlist_Pinmap_r2.0_v1.0.xlsx). This section only shows the pin list and description.

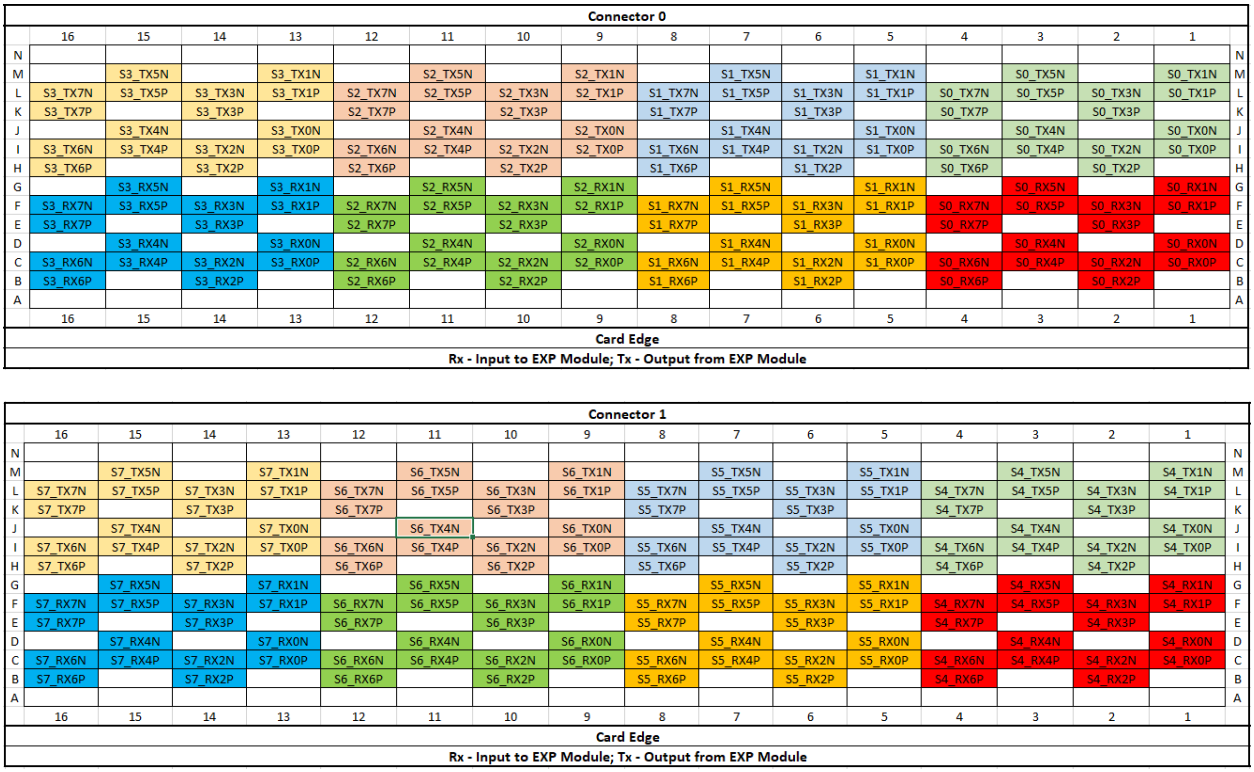
Table 6 EXP Pinouts

SFF-TA-1020						
Side B				Side A		
Position	**Signal**	**IO Type (Module Direction POV)**		**IO Type (Module Direction POV)**	**Signal**	**Position**
End Wall						
B1	PRSNT#	Output		Power Input	3.3V	A1
B2	GND	Ground		Power Input	3.3V	A2
B2	SMBCLK	Input		Ground	GND	A2
B4	SMBDAT	I/O		Input	MDC	A4
B5	MOD_ID2	Input		I/O	MDIO	A5
B6	MOD_ID1	Input		I/O	SPI_SIO3/RFU	A6
B7	MOD_ID0	Input		I/O	SPI_SIO2/RFU	A7
B8	PWRBRK#/RFU	Input		I/O	SPI_MISO/SIO1	A8
B9	JTAG_TRST	Input		I/O	SPI_MOSI/SIO0	A9
B10	JTAG_TDI	Input		Input	SPI_CS1#	A10

B11	JTAG_TDO	Output		Input	SPI_CS0#	A11
B12	JTAG_TMS	Input		Input	SPI_CLK	A12
B13	JTAG_TCK	Input		Ground	GND	A13
B14	GND	Ground		Input	PERn3	A14
B15	PETn3	Output		Input	PERp3	A15
B16	PETp3	Output		Ground	GND	A16
B17	GND	Ground		Input	PERn2	A17
B18	PETn2	Output		Input	PERp2	A18
B19	PETp2	Output		Ground	GND	A19
B20	GND	Ground		Input	PERn1	A20
B21	PETn1	Output		Input	PERp1	A21
B22	PETp1	Output		Ground	GND	A22
B23	GND	Ground		Input	PERn0	A23
B24	PETn0	Output		Input	PERp0	A24
B25	PETp0	Output		Ground	GND	A25
B26	GND	Ground		Input	PE_REFCLKn	A26
B27	PERST#/RST#	Input		Input	PE_REGCLKp	A27
B28	GND	Ground		Ground	GND	A28
Mechanical Key						
B29	54V	Power Input		Ground	GND	A29
B30	54V	Power Input		Ground	GND	A30
B31	54V	Power Input		Ground	GND	A31
B32	54V	Power Input		Ground	GND	A32
B33	54V	Power Input		Ground	GND	A33
B34	54V	Power Input		Ground	GND	A34
B35	54V	Power Input		Ground	GND	A35
B36	54V	Power Input		Ground	GND	A36
B37	54V	Power Input		Ground	GND	A37
B38	54V	Power Input		Ground	GND	A38
B39	54V	Power Input		Ground	GND	A39
B40	54V	Power Input		Ground	GND	A40
B41	54V	Power Input		Ground	GND	A41
B42	54V	Power Input		Output	PRSNT#	A42
Mechanical Key						

Table 7 GPIO Recommended Operating Conditions

	Description	MIN	MAX	Unit
V_{3V3}	IO reference voltage	3.135	3.465	V
V_{IH-3V3}	High level of IO refers to 3.3V	0.7 x V_{3V3}	3.465	V
V_{IL-3V3}	Low level of IO refers to 3.3V	-0.5	0.3 x V_{3V3}	V



Connector 0																
	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
N																
M		S3_TX5N		S3_TX1N		S2_TX5N		S2_TX1N		S1_TX5N		S1_TX1N		S0_TX5N		S0_TX1N
L	S3_TX7N	S3_TX5P	S3_TX3N	S3_TX1P	S2_TX7N	S2_TX5P	S2_TX3N	S2_TX1P	S1_TX7N	S1_TX5P	S1_TX3N	S1_TX1P	S0_TX7N	S0_TX5P	S0_TX3N	S0_TX1P
K	S3_TX7P		S3_TX3P		S2_TX7P		S2_TX3P		S1_TX7P		S1_TX3P		S0_TX7P		S0_TX3P	
J		S3_TX4N		S3_TX0N		S2_TX4N		S2_TX0N		S1_TX4N		S1_TX0N		S0_TX4N		S0_TX0N
I	S3_TX6N	S3_TX4P	S3_TX2N	S3_TX0P	S2_TX6N	S2_TX4P	S2_TX2N	S2_TX0P	S1_TX6N	S1_TX4P	S1_TX2N	S1_TX0P	S0_TX6N	S0_TX4P	S0_TX2N	S0_TX0P
H	S3_TX6P		S3_TX2P		S2_TX6P		S2_TX2P		S1_TX6P		S1_TX2P		S0_TX6P		S0_TX2P	
G		S3_RX5N		S3_RX1N		S2_RX5N		S2_RX1N		S1_RX5N		S1_RX1N		S0_RX5N		S0_RX1N
F	S3_RX7N	S3_RX5P	S3_RX3N	S3_RX1P	S2_RX7N	S2_RX5P	S2_RX3N	S2_RX1P	S1_RX7N	S1_RX5P	S1_RX3N	S1_RX1P	S0_RX7N	S0_RX5P	S0_RX3N	S0_RX1P
E	S3_RX7P		S3_RX3P		S2_RX7P		S2_RX3P		S1_RX7P		S1_RX3P		S0_RX7P		S0_RX3P	
D		S3_RX4N		S3_RX0N		S2_RX4N		S2_RX0N		S1_RX4N		S1_RX0N		S0_RX4N		S0_RX0N
C	S3_RX6N	S3_RX4P	S3_RX2N	S3_RX0P	S2_RX6N	S2_RX4P	S2_RX2N	S2_RX0P	S1_RX6N	S1_RX4P	S1_RX2N	S1_RX0P	S0_RX6N	S0_RX4P	S0_RX2N	S0_RX0P
B	S3_RX6P		S3_RX2P		S2_RX6P		S2_RX2P		S1_RX6P		S1_RX2P		S0_RX6P		S0_RX2P	
A																
	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Card Edge																
Rx - Input to EXP Module; Tx - Output from EXP Module																

Connector 1																
	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
N																
M		S7_TX5N		S7_TX1N		S6_TX5N		S6_TX1N		S5_TX5N		S5_TX1N		S4_TX5N		S4_TX1N
L	S7_TX7N	S7_TX5P	S7_TX3N	S7_TX1P	S6_TX7N	S6_TX5P	S6_TX3N	S6_TX1P	S5_TX7N	S5_TX5P	S5_TX3N	S5_TX1P	S4_TX7N	S4_TX5P	S4_TX3N	S4_TX1P
K	S7_TX7P		S7_TX3P		S6_TX7P		S6_TX3P		S5_TX7P		S5_TX3P		S4_TX7P		S4_TX3P	
J		S7_TX4N		S7_TX0N		S6_TX4N		S6_TX0N		S5_TX4N		S5_TX0N		S4_TX4N		S4_TX0N
I	S7_TX6N	S7_TX4P	S7_TX2N	S7_TX0P	S6_TX6N	S6_TX4P	S6_TX2N	S6_TX0P	S5_TX6N	S5_TX4P	S5_TX2N	S5_TX0P	S4_TX6N	S4_TX4P	S4_TX2N	S4_TX0P
H	S7_TX6P		S7_TX2P		S6_TX6P		S6_TX2P		S5_TX6P		S5_TX2P		S4_TX6P		S4_TX2P	
G		S7_RX5N		S7_RX1N		S6_RX5N		S6_RX1N		S5_RX5N		S5_RX1N		S4_RX5N		S4_RX1N
F	S7_RX7N	S7_RX5P	S7_RX3N	S7_RX1P	S6_RX7N	S6_RX5P	S6_RX3N	S6_RX1P	S5_RX7N	S5_RX5P	S5_RX3N	S5_RX1P	S4_RX7N	S4_RX5P	S4_RX3N	S4_RX1P
E	S7_RX7P		S7_RX3P		S6_RX7P		S6_RX3P		S5_RX7P		S5_RX3P		S4_RX7P		S4_RX3P	
D		S7_RX4N		S7_RX0N		S6_RX4N		S6_RX0N		S5_RX4N		S5_RX0N		S4_RX4N		S4_RX0N
C	S7_RX6N	S7_RX4P	S7_RX2N	S7_RX0P	S6_RX6N	S6_RX4P	S6_RX2N	S6_RX0P	S5_RX6N	S5_RX4P	S5_RX2N	S5_RX0P	S4_RX6N	S4_RX4P	S4_RX2N	S4_RX0P
B	S7_RX6P		S7_RX2P		S6_RX6P		S6_RX2P		S5_RX6P		S5_RX2P		S4_RX6P		S4_RX2P	
A																
	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Card Edge																
Rx - Input to EXP Module; Tx - Output from EXP Module																

Figure 23 – ExaMax2 Connector Pin map

8.4 EXP Power Profiles

This section defines EXP thermal design power (TDP) and power up/down sequences.

8.4.1 Thermal Design Power TDP

The module supports up to 400W for the switch configuration and 200W for the retimer configuration, with input voltage ranging from 44V to 59.5V. The UBB supplies power to the module through the Gen-Z 2C connector. The current capability and power status are in the table below. The power is available in state S0 only.

Table 8 Power Rails

Power Rail	Voltage Tolerance	# of pins	Power Consumption	Status
P54V/P48V	44V min to 59.5V max	14	400W	Normal power
P3.3V	3.3V±10%(max)	2	5W	Normal power

8.5 System power sequencing

System design should follow below power sequence requirement. It is recommended to check with each specific module specification to ensure the modules work correctly.

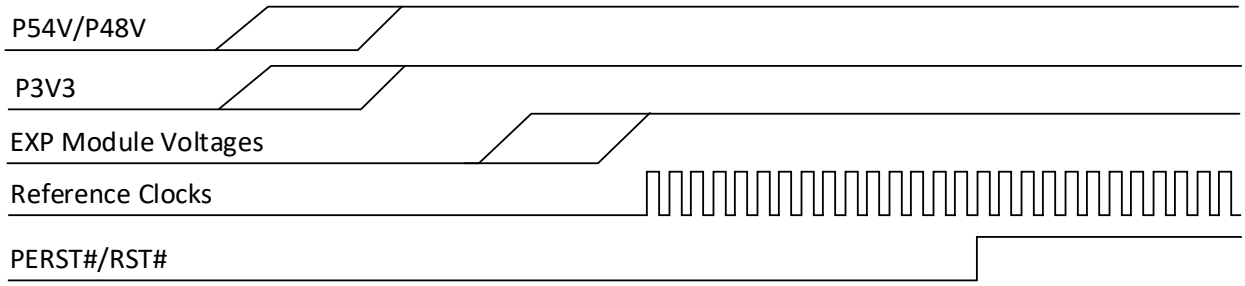


Figure 24 EXP Power Up Sequence

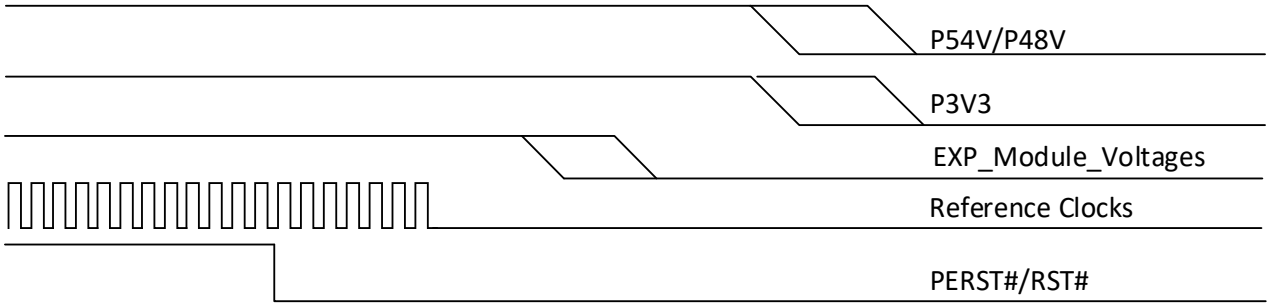


Figure 25 EXP Power-Down Sequence

Notes:

1) As the voltage planes on the module ramp up, the reference clocks from UBB will begin to turn on.
2) The RST# signal is defined for non PCIe design.

8.6 EXP Insertion Loss

The module interconnection channel total insertion loss from silicon package to mated connector should not exceed -6dB at 28GHz (and PCIe Gen5/6 @16GHz). The system integrator may contact the module supplier for details about the interconnection channel insertion loss and plan system baseboard design accordingly.

Total Tx or Rx loss on OAM + mated Mezz Connectors @28GHz (PCIe Gen5/6 @16GHz)	up to 8dB
Total Tx or Rx loss on EXP + mated BP Connectors @28GHz	up to 6dB

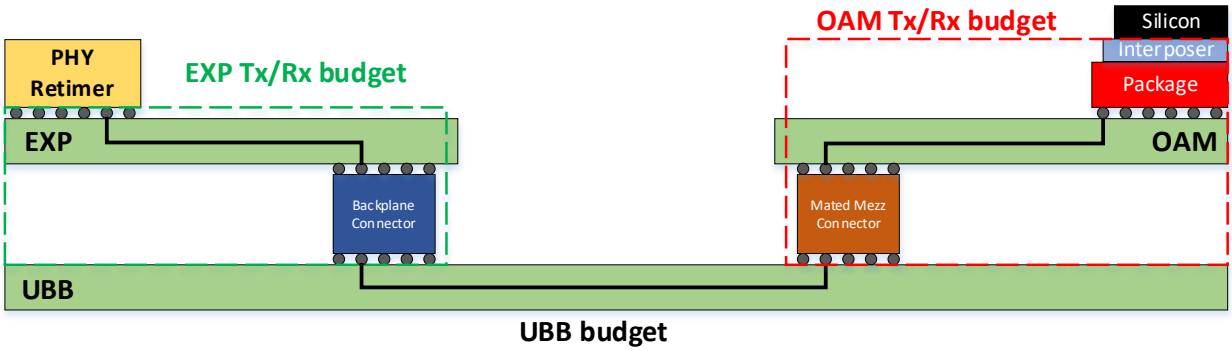


Figure 26 Channel loss budget (OAM to EXP)

Total channel loss budget = OAM Tx (die) + UBB + EXP Rx (package ball)

Example:

Total channel loss budget = 30 dB (28 GHz)

- **OAM Tx/Rx = 8 dB**
- **EXP Tx/Rx = 6 dB**
- **UBB budget = 30 dB – 8 dB – 6 dB = 16dB**

8.7 EXP PCB Stack-Up

This section describes the OAI EXP reference board stack-up and requirements. The OAI EXP reference boards use 16 PCB stack-up with 800G PCB material. To support high interconnect speed at 112G-PAM4, the PCB stack-up adheres to the following requirements:

- PCB with up to two 2oz layers to meet the required copper density
- 85 & 90 Ohms differential traces in internal signal layers as needed.
- 45 & 50 Ohms or single-ended traces as required (Depending on the chip vendor's design guide)

- PCB material depends on the maximum trace length of topology design and meets the vendor's channel loss criteria.
- Back Drilling for signals on Layer 5, 7, 10 and 12 to remove stubs from SerDes and PCI-e Gen5/6 via transitions. Limit back drilling to 1mm from the top layer to help press-fit contact.

EXP reference board uses the below stack up. Each vendor must fine-tune the width/spacing design based on the material target and impedance control table below.

Table 9 EXP reference Stack-Up

Layer	Plane Description		Copper (OZ)	Thickness (mil)
		Solder mask		0.5
L1	Top	Signal/PWR	0.5oz + plating	1.9
		PrePreg		3.5
L2	GND1	Ground	1.0	1.2
		Core (1/1)		6
L3	VCC1	Power	2.0	2.4
		PrePreg		7
L4	GND2	Ground	1.0	1.2
		Core (1/1)		4
L5	IN1	Signal/PWR	0.5	0.6
		PrePreg		5
L6	GND3	Ground	0.5	0.6
		Core (1/1)		4
L7	IN2	Signal/PWR	0.5	0.6
		PrePreg		5
L8	GND4	Ground	0.5	0.6
		Core (1/1)		4
L9	GND5	Ground	0.5	0.6
		PrePreg		5
L10	IN3	Signal/PWR	0.5	0.6
		Core (1/2)		4
L11	GND6	Ground	0.5	0.6
		PrePreg		5
L12	IN4	Signal/PWR	0.5	0.6
		Core (1/2)		4
L13	GND7	Ground	1.0	1.2
		PrePreg		7
L14	VCC2	Power	2.0	2.4
		Core (1/1)		6

L15	GND8	Ground	1.0	1.2
		PrePreg		3.5
L16	BOT	Signal/PWR	0.5oz + plating	1.9
		Solder Mask		0.5
		Total		92.2mil (with +/- 10% tolerance)

Table 10 UBB Impedance control

Trace Width (mil)	Air Gap Spacing (mil)	Impedance Type	Layer	Impedance Target (ohm)	Tolerance (+/- %)
7.2		Single-Ended	1,16	45	10%
6.0		Single-Ended	1,16	50	10%
7	6	Differential	1,16	85	10%
6.5	6.5	Differential	1,16	90	10%
6.1		Single-Ended	5,7,10,12	45	10%
5		Single-Ended	5,7,10,12	50	10%
6.3	6.7	Differential	5,7,10,12	85	10%
5.8	7.2	Differential	5,7,10,12	90	10%

9 EXP Interconnect Topologies

This section describes the recommended interconnection topology for a system with up to 8 EXPs.

9.1 Module ID

The following figure shows the MODULE_ID[2:0] strapping for the physical orientation of modules when eight EXPs are used.

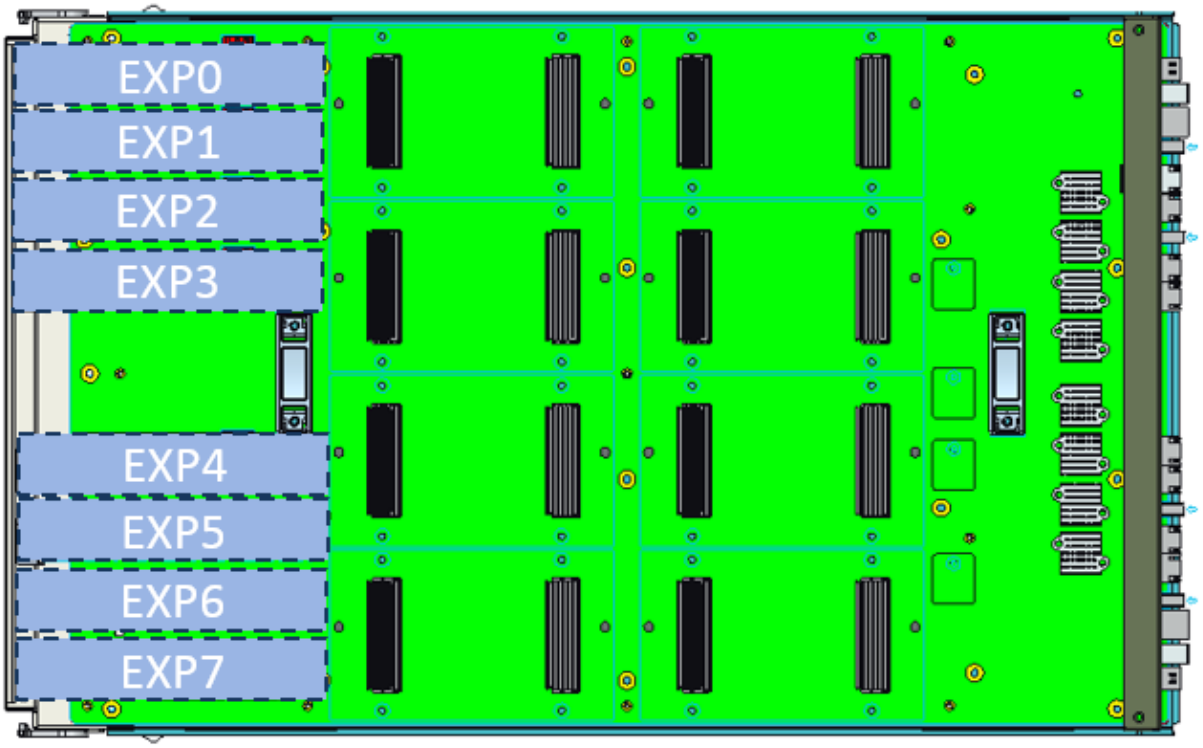


Figure 27 Required MODULE_ID[2:0] assignments for baseboards with 8 interconnected modules

Table 11 MODULE_ID List

EXP	EXP Module ID
EXP0	000
EXP1	001
EXP2	010
EXP3	011
EXP4	100
EXP5	101
EXP6	110
EXP7	111

9.2 Interconnect Topology

This section describes different interconnect topologies and routing guidance for eight EXPs with varying numbers of the port. If all seven ports are configurated and routed as x16, there is no additional port(s) for expansion. To reserve expansion port(s), we suggest limiting onboard Interconnect links up to x8. 2nd

half of port 1(x8, also referred to as 1H) is reserved for expansion by default. 2nd half of ports 4,6,7 is used in x8 based full connected topology at section 9.2.4.

9.2.1 11-Port (7+4) FC/Retimer Topology (Fully Connected + 4 Scale Out)

Figure 28 below shows an example of 11-port topology with seven SerDes ports for fully connected topology and four SerDes ports for scale out (per EXP) of 8 modules in UBB. The scale out ports are routed to EXP modules to support 32 scale out ports. Please follow port mapping to design EXP to be able to fit in the UBB.

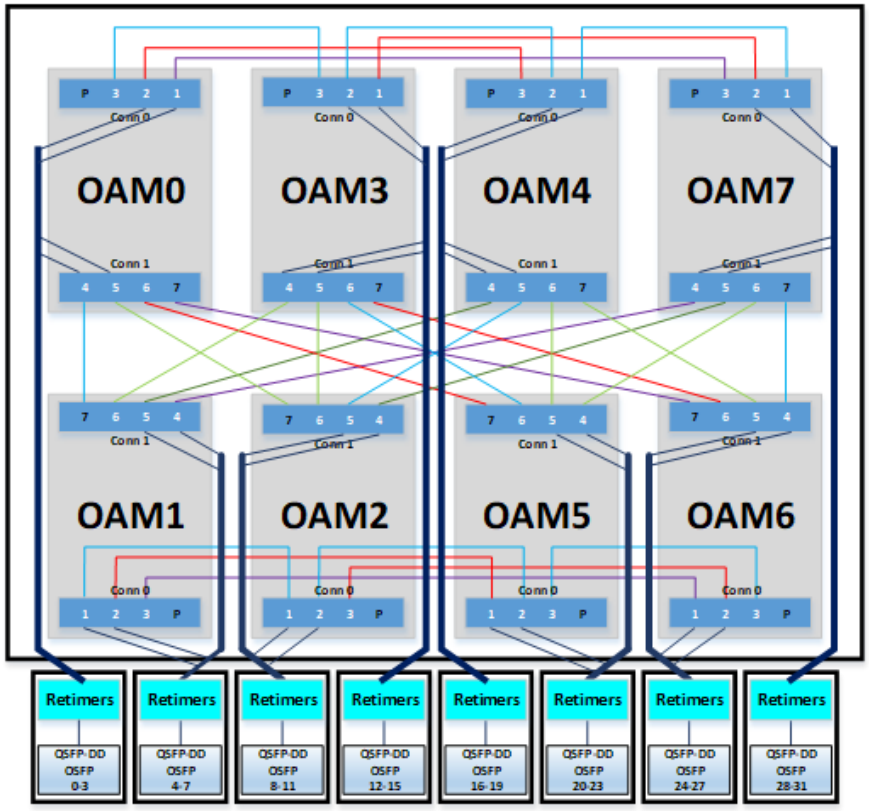


Figure 28 11-Port (7+4) FC/Retimer Topology

9.2.2 8-Port Switch-Based Topology

Figure 29 below shows an example of 8-port switch-based topology with eight SerDes ports of each EXP are connected to eight EXP modules to form fully connected topology through switch devices for scale up and out.

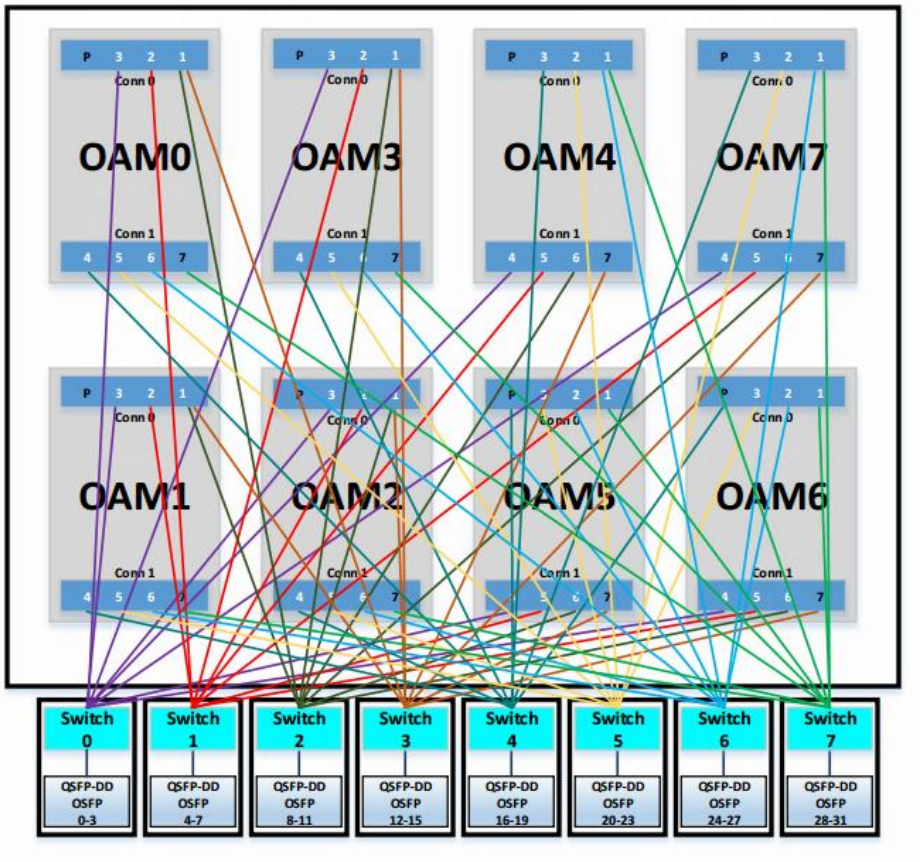


Figure 29 8-Port Switch-Based Topology

10 EXP reference system design

This section gives a system design concept as a reference. Figure 30 Reference System Design shows 8 EXP modules.

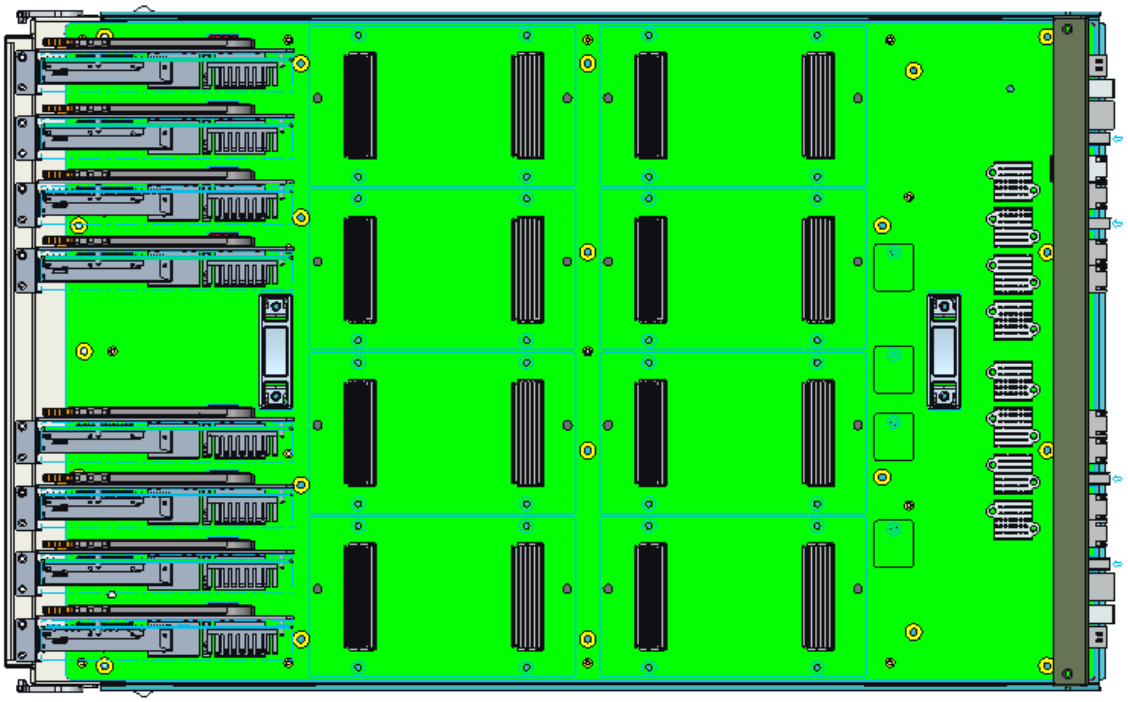


Figure 30 Reference System Design

11 EXP Management and Security Requirements

This section describes a common set of management and security requirements for EXP.

11.1 Management Interface

The EXP sideband management interface is used by a Management Controller (MC) or Baseboard Management Controller (BMC) to communicate with the EXP. Table 12 below summarizes the sideband management interfaces.

Table 12 Sideband Management Interface

Requirement	Voltage Level	Required
I2C/I3C/SMBus 2.0 compliant physical interface (Slave) *	3.3V	Yes
JTAG physical interface	3.3V	Yes
SPI physical interface	1.8V	Yes(vendor-specific)
MDC/MDIO physical interface	1.8V	

Note: EXP enters reduce/capped power mode when EXP max power is capped.

*When MCTP over SMBus is used, the BMC shall support both master and slave modes.

The EXP communicates with the Baseboard Management Controller (BMC) by using:

- SMBus:
  - SMBus supporting 1MHz mode is preferred.
  - Standard Intelligence Platform Management Bus (IPMB) and Intelligent Platform Management Interface (IPMI) commands.
  - SMBus ARP protocol.
  - Management Component Transport Protocol (MCTP) over SMBus out of band. (DMTF DSP0237)
- JTAG
  - For register dump, memory dump, debug access.
- SPI/MDIO
  - For PHY management, SPI or MDIO per vendor requirement.
- PCIe (optional)
  - To support MCTP over PCIe inband.
  - For Switch management.

11.2 Sensor Reporting

It is crucial that connected sensors of these components (voltage, current/power, temperature, etc.) can be retrieved over sideband interfaces for system management. The sensor reporting interface will only be accessible in the main power mode (S0).

The sensor reporting interface will only be accessible in the main power mode (S0).

Table 13 Sensor List

Sensor List	Remark
Power/Current	
Voltage	
ASIC Temp (hot spot and edge temp)	
ASIC Tjmax	
VR Temp	
VR Vol	
Inlet sensor	Need to define the common location
Outlet sensor	
Power State: Max power mode Reduced/Capped power mode*	

Note

Summarizes the sensors reporting list: The report from these sensors improves the system monitoring/management and allows the baseboard management device to access the critical components on the module. It is recommended that the voltage/current/power sensor reporting accuracy is within ± 2%, and the temperature reporting accuracy is within ±3°C.

EXP Module shall support Sensor discovery via IPMI or Platform Level Data Model (PLDM) for Platform Monitoring and Control (DMTF DSP0248). MC will follow mechanisms specified in DSP0248 to discover sensors supported by the EXP module and their threshold.

11.3 Error Monitoring/Reporting

System Management Controller (MC) or Baseboard Management Controller (BMC) shall be able to monitor and access the EXP through PLDM over MCTP as needed to set thresholds, clear status, determine error counts and syndromes (SW driven interrupt or an Alert pin), and identify error sources/Syndromes, etc.

EXP shall support Platform Level Data Model (PLDM) for Platform Monitoring and Control (DMTF DSP0248) for error reporting, enabling the module to define state sensors and events for health monitoring and reporting.

Table 14 Error state sensors

Sensor List	Remark
Overall EXP health status	It can be vendor-specific
Device health event	
Interconnect Link events	
Scale-out link Event	for the scale-out bus, if the different buses from the interconnect links
Other custom/Vendor-specific	

11.4 Firmware Update

The EXP should support secure boot. Detail requirements, please refer to section 11.7.1.

EXP supplier shall provide an in-band FW update (if supported) to perform the firmware update.

For out-of-band firmware updates, EXP shall support Platform Level Data Model (PLDM) for firmware update over MCTP as specified in DMTF DSP0267.

Update failure and failure types shall be communicated to BMC as specified in DSP0267.

11.5 FRU Information

System Management Controller (MC) or Baseboard Management Controller (BMC) shall access related internal registers to get module information (see Table 15).

FRU shall be accessed through Platform Level Data Model (PLDM) for FRU Data, following DMTF DSP0257.

Table 15 FRU Information

FRU Info	Remark
Manufacturing Date	
Manufacturer	
Product Name	
Serial	
Part Number	
FRU ID	
Version	
Asset Tag	
Firmware Version	
EXP Spec Version	
EXP TDP	
SerDes Link Speed	
Custom Data 1	
Custom Data 2	
Custom Data 3	
Custom Data 4	
Custom Data 5	
Custom Data 6	

11.6 IO Calibration

The system shall be able to get PCIe/interconnect training status and margin information. For in-band access, specific tools and/or API shall be provided by the EXP supplier.

11.7 EXP Security Requirements

11.7.1 Secure Boot

The device must support a secure boot. There are multiple requirements to implement secure boot. Here is a brief list:

- Hardware-based Root of Trust:
  - Immutable Root-of-Trust (e.g., OTP) required for provisioning asymmetric RoT key and other security-related critical information.
  - ROM code for minimum support needed for Secure Boot
- Boot Time Verification:
  - On every boot, the ROM code should cryptographically verify mutable firmware code using the asymmetric RoT public key.
  - All mutable code should be verified by signature authentication before allowing it to execute.
  - ROM code patching is not permitted in a production device.
  - In case of failures, please refer to the 'Recovery' section for more details.
- TOCTOU Attack Protection:
  - The attacker should not be able to modify the firmware image (Time of Use) after the signature of the firmware is verified (Time Of Check) during the boot process
- Anti-Rollback Protection:
  - The device must support Security Version Number (SVN) in the immutable memory to protect against downgrade attacks
- Key Revocation or Change of Ownership:
  - The device must support Secure revocation of Intermediate key (used for signing of firmware) in case of the Key compromise
  - For a change of ownership, the device should either support revoking the ownership key or rotating the ownership certificate (as suggested in DSP0274)
- Secure Firmware Update:
  - Signature generation for firmware payload using asymmetric RoT private key is necessary for secure boot
  - The device should be able to authenticate the signed FW payloads before booting up using the payload

For further details on implementation and a detailed set of requirements, please refer to the OCP Hardware Secure Boot document for more information.

11.7.2 Recovery

EXP must support a recovery mechanism to restore the mutable firmware code to a state of integrity if any such firmware code or critical data are detected to have been corrupted or forced to recover through an authorized mechanism.

- EXP should also support two mutable firmware (active and recovery) regions where the recovery firmware is the previously known good image.
- In case of failure of redundant copies, EXP should support recovery over the sideband interface.
    - EXP shall support Platform Level Data Model (PLDM) firmware update over MCTP as specified in DMTF DSP0267.
    - Update failure and failure types shall be communicated to BMC as specified in DSP0267.
    - Firmware updated out-of-band should still follow the boot-time verification process of secure boot.

For further details on implementation and a detailed set of requirements, please refer to the OCP Recovery document for more information.

11.7.3 Debug Capabilities

Any intrusive debug capabilities (general-purpose register contents, alter control flow), e.g., JTAG, must be disabled for remote access. If needed, they should only be re-enabled via physical access or using cryptographically authorized tokens.

11.7.4 Attestation

It is critical to verify the firmware running on the device dynamically and the device itself cryptographically; this helps establish trust between the devices. It is recommended to support device attestation for EXP.
Here is a short list of generic requirements to support the attestation for EXP:

- Keys, seeds, and device identifiers
- Provisioning Facility (Initial Provisioning Environment Operations and Equipment)
- Device Ownership Provisioning
- Authentication, Attestation, and Enrollment protocol
- Measurement collection and storage

For further details on implementation and a detailed set of requirements, please refer to the OCP Attestation for System Components v1.0 for more information.

12 Environmental

12.1 Environmental Requirements

The EXP shall meet the following environmental requirements:

- Gaseous Contamination: Severity Level G1 per ANSI/ISA 71.04-1985
- Ambient operating temperature range: 5°C to +35°C
- Operating and Storage relative humidity: 20% to 90% (non-condensing)
- Storage temperature range: -20°C to +70°C
- Transportation temperature range: -55°C to +85°C (short-term storage)
- Operating altitude with no de-ratings: 3048m (10000 feet) – recommended as this is a Facebook spec and standard for Telco operation.

12.2 Regulation

The vendor needs to provide CB reports of the EXP, which are required to have rack-level CE. The EXP should be compliant with RoHS and WEEE. The PCB should have a UL 94V-0 certificate.

13 Revision History

Author	Description	Revision	Date
Song Kok Hang (AMD) Ahmed Abou-Alfotouh (AMD) Cheng Chen (Meta) Xinxin Wang (H3C) Hongzheng Cai (H3C) Hao Zhang (H3C) Ash Liao (Wiwynn) Nick Wang (Wiwynn) Jaylen Cheng (Wiwynn) Tao Lang (Microchip) Samuel Kocsis (Amphenol) Jonathan Chng (Amphenol) Calvin Feng (TE)	Initial Release	1.0	9/14/2023