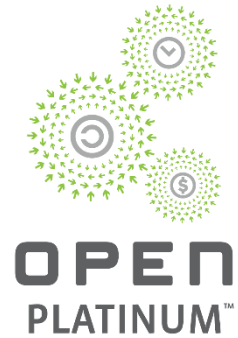
OPEN POSSIBILITIES.

Inspur MX1 system introduction

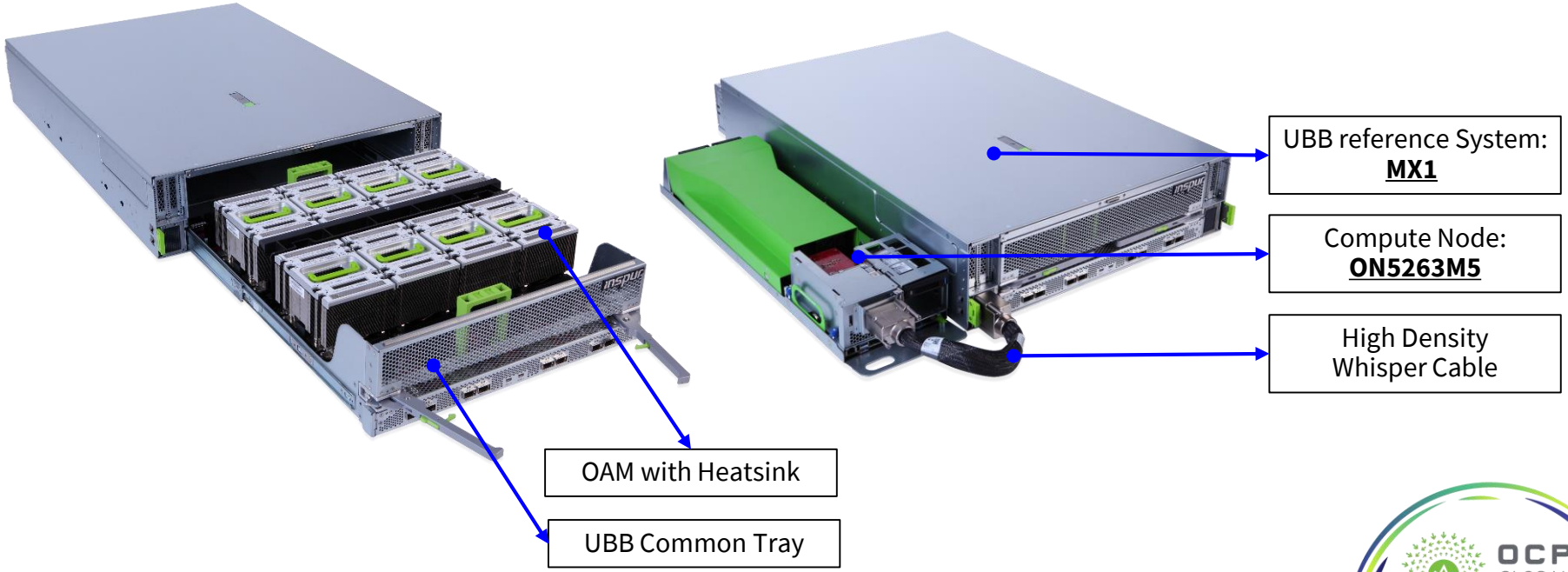
OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

System Overview

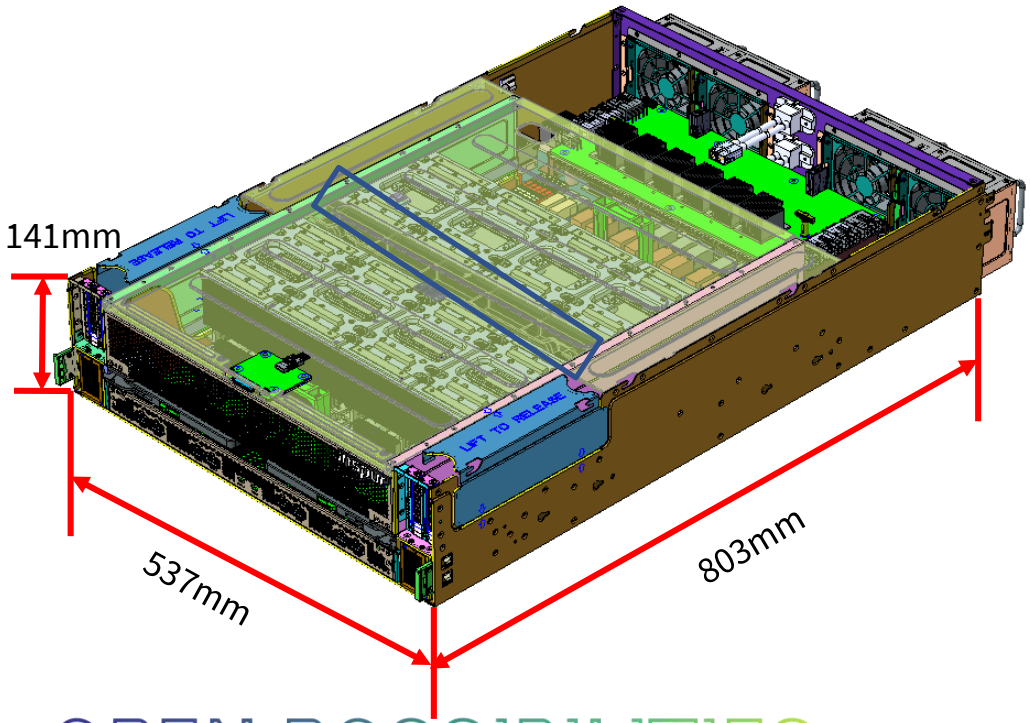


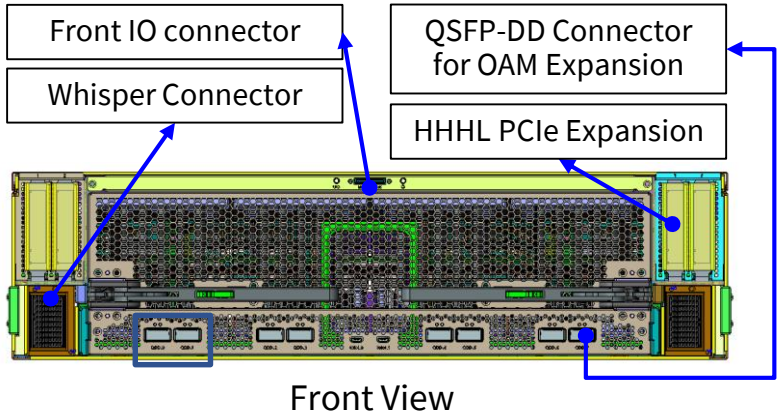
OPEN POSSIBILITIES.

Inspur MX1 System



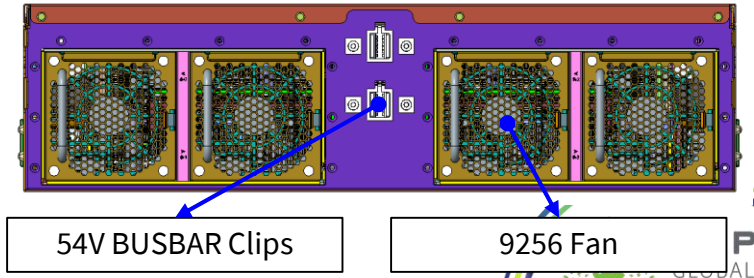
Inspur System Overview





Front View

Rear View



OPEN POSSIBILITIES.

Inspur OAI System Features

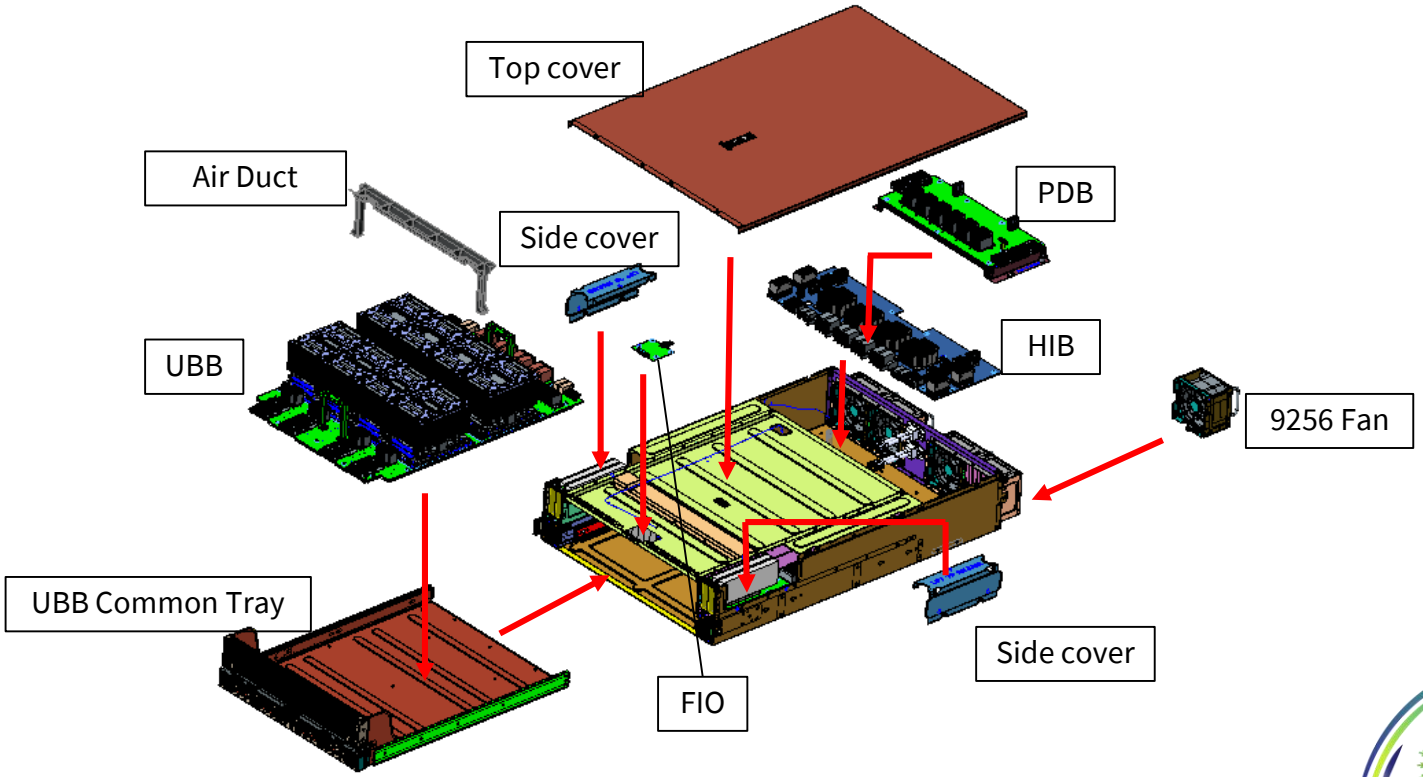
Product model: MX1	
Chassis	21” 3OU Rack mount
Dimensions	537W*141H*803D (mm)
Connection with Compute node	Up to PCIe Gen4 x32
OAM	Support Max 8pcs 48~54V OAM(up to 550W each); Support Max 8pcs 12V OAM (up to 350W each)
Power without OAM	1570W
PCIe Switch	Support PCIe Gen4 (100lanes/chip)
PCIe re-timer	Support PCIe Gen4 x16
Phy re-timer	56Gbps PAM-4 or 10/28Gbps NRZ x16
Expansion slots	Up to 4 x PCIe Gen4 x16 low profile standard card
BMC	AST2520
I/O	Dongle connector for dedicate NIC and UBS, UID/PWR Button with LED , QSFDDx8 for OAM scale out, micro USBx2 for OAM debug
Ambient Working Temperature	5-35 ℃

OPEN POSSIBILITIES.

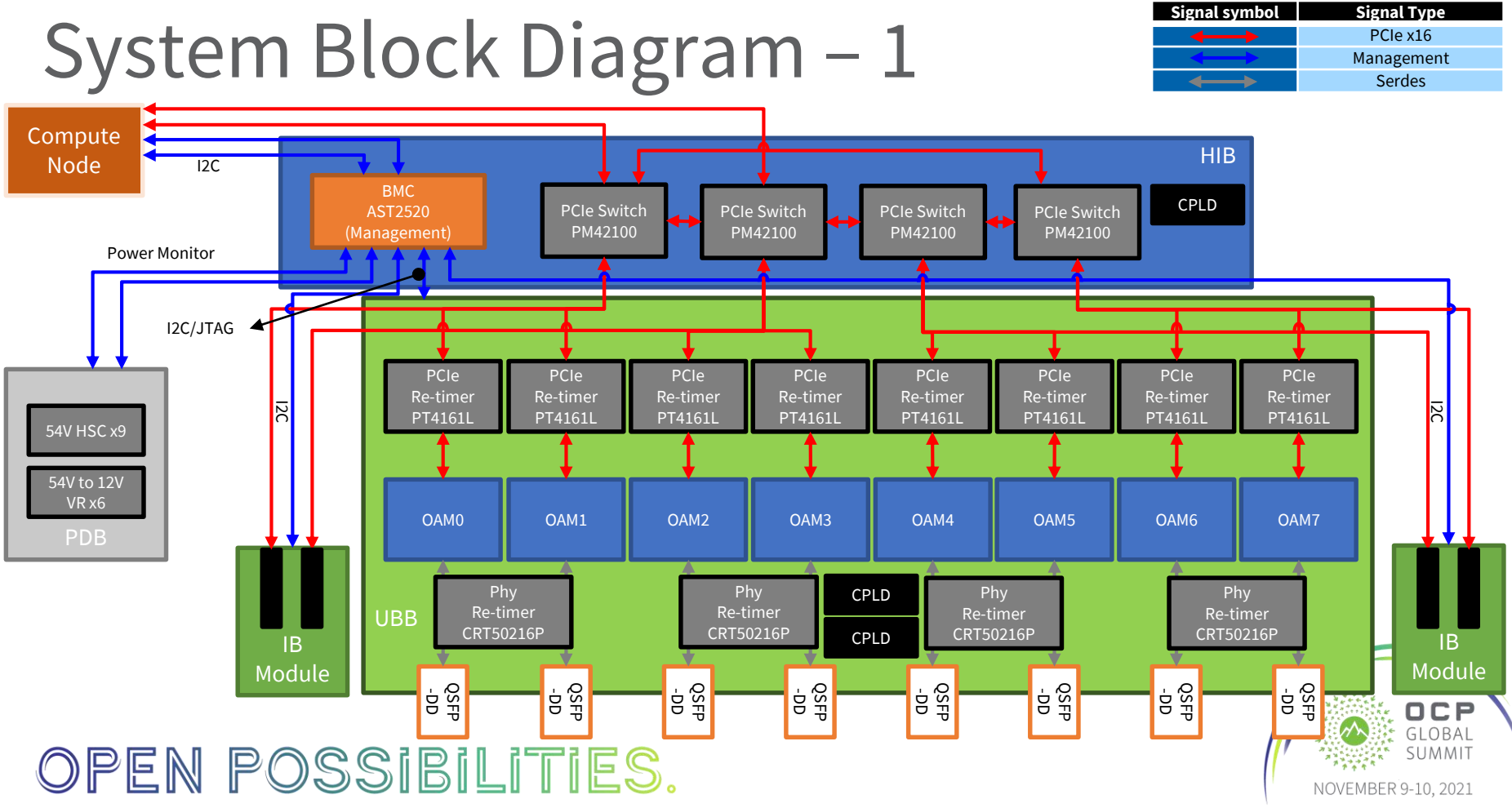
OPEN Compute Project
SOLUTION PROVIDER®

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

System Explode View

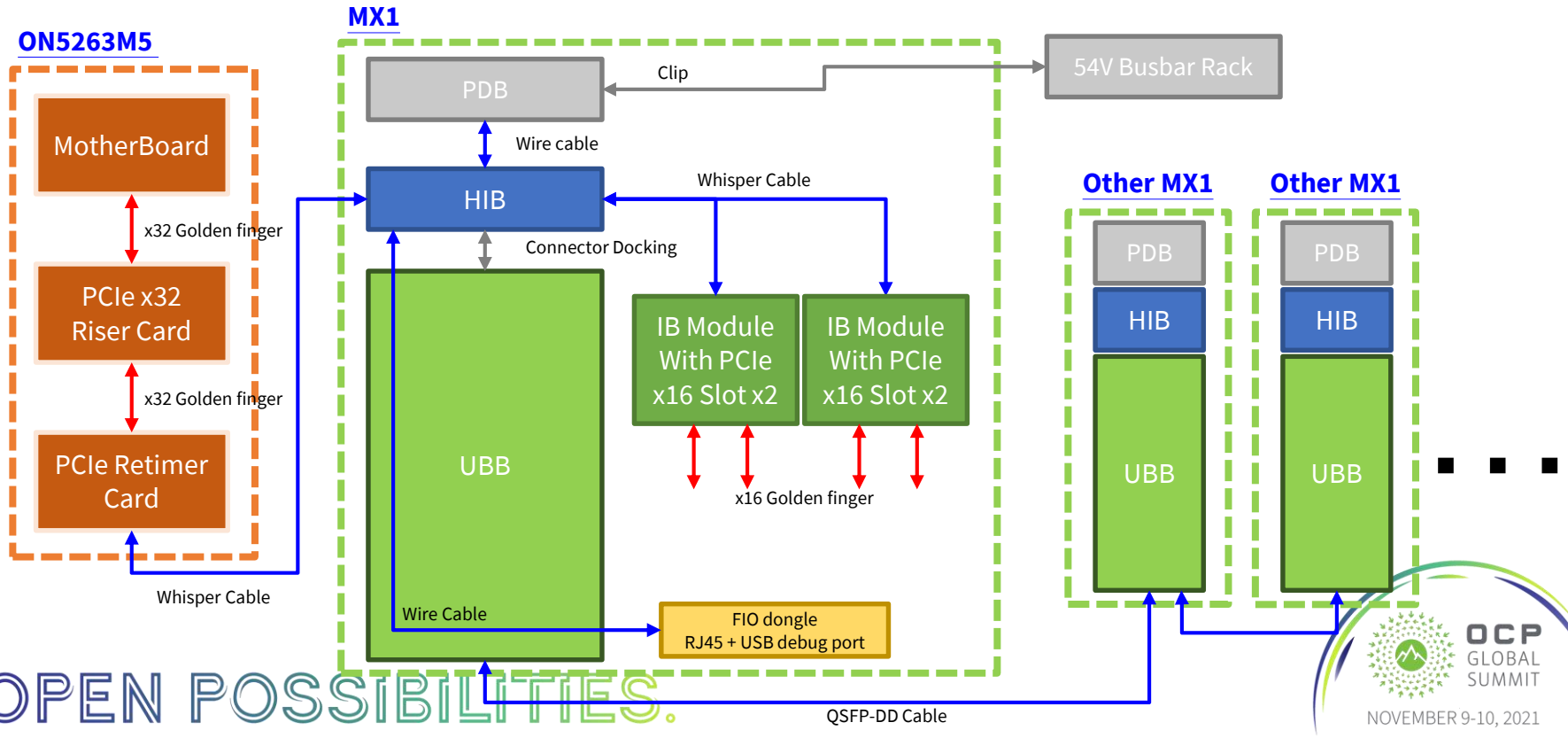


System Block Diagram – 1

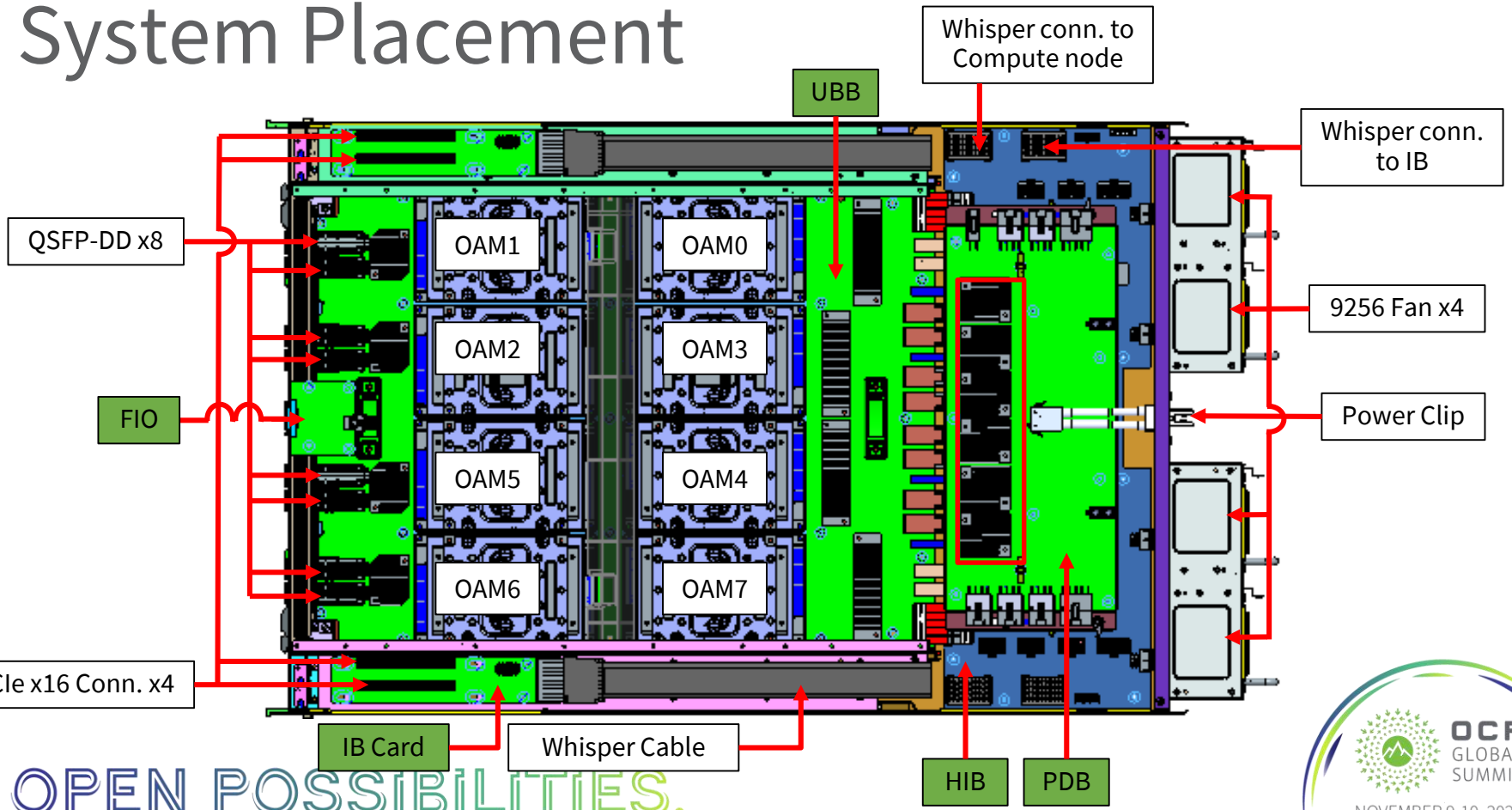


System Block Diagram – 2

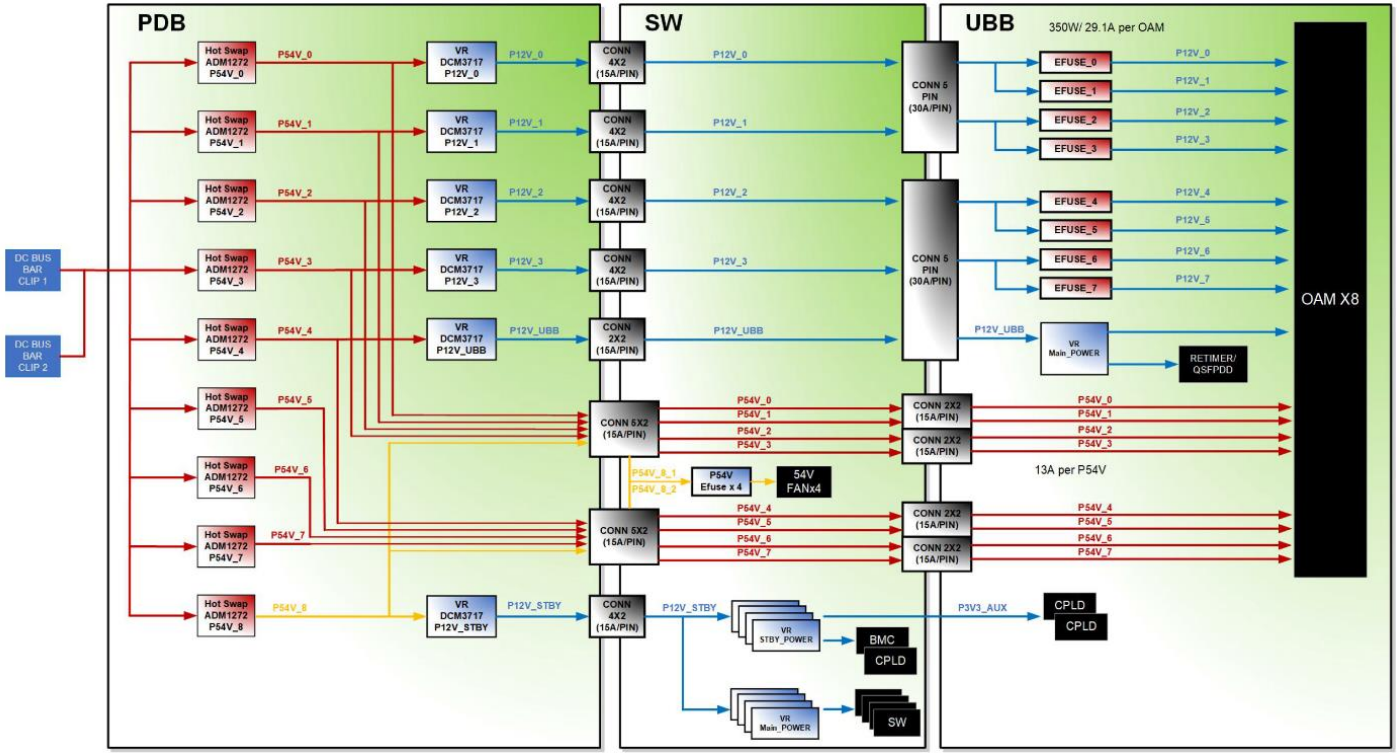
Connection symbol	Connection Type
	Golden Finger
	Cable
	Connector Docking



System Placement



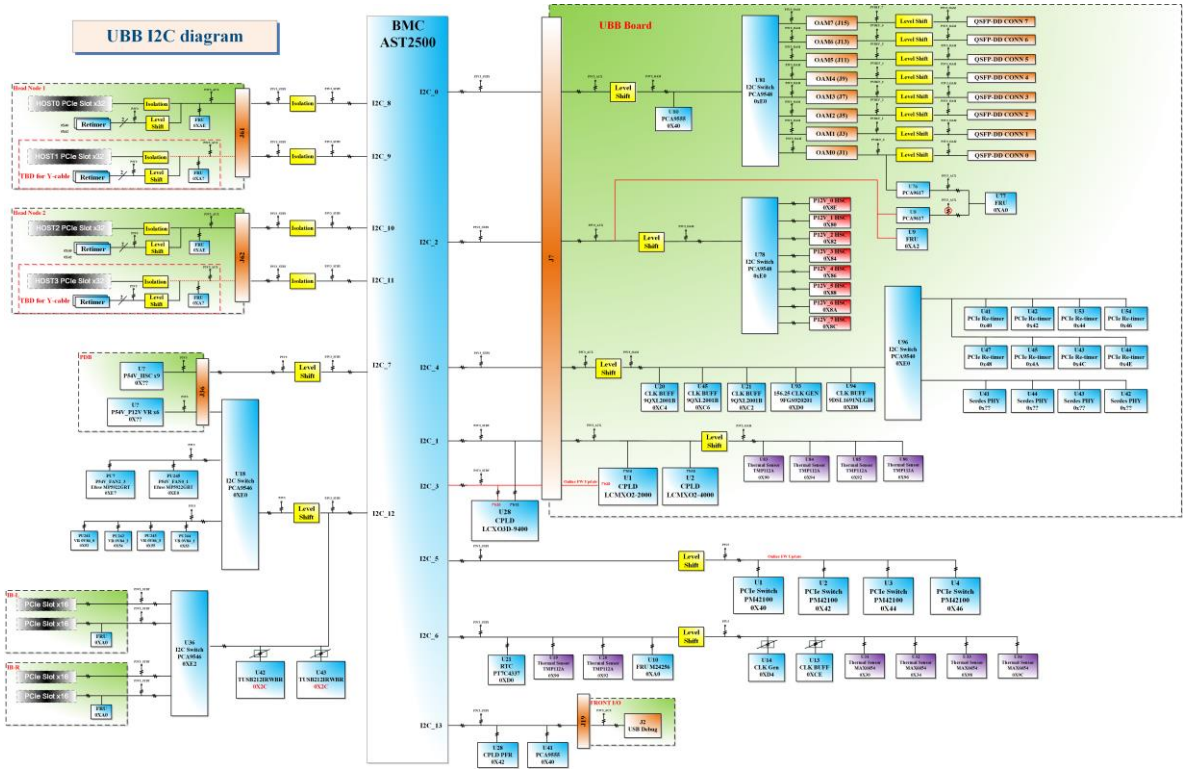
System Power Map



OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

System I2C Map

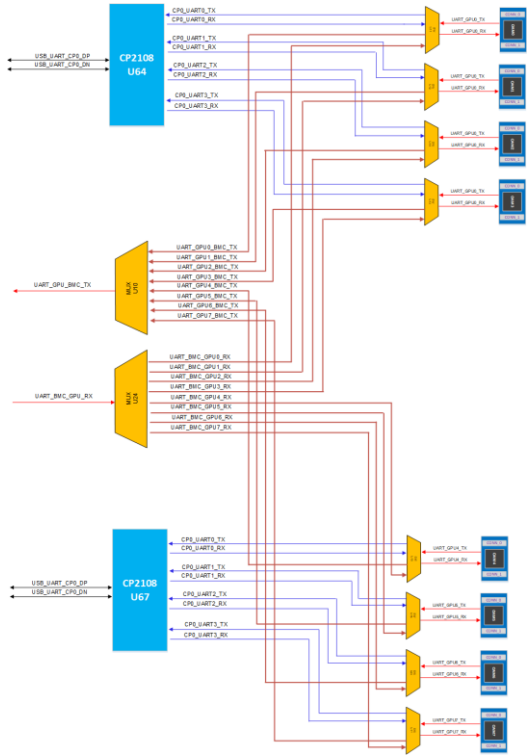


For HW point of view, UBB MX1 has verified different OAM as below

1. INTEL NERVANA (EOL)
2. HABANA
3. Enflame
4. Cambricon

OPEN POSSIBILITIES.

System Management Map



For HW point of view, UBB MX1 has verified different OAM as below

1. INTEL NERVANA (EOL)
2. HABANA
3. Enflame
4. Cambricon

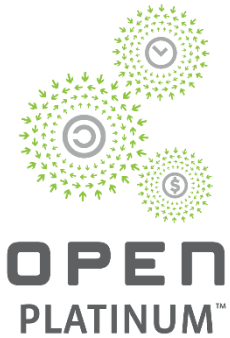
System Power Budget (700W OAM)

Component	Quantity	Power per item (W)	Design Power (W)	Utilization	Usage Power (W)
UBB					
OAM (48V input, no VR loss)	8	700	5600	100%	5600
PCIe Retimer	8	9.6	76.8	80%	61.44
PHY Retimer	4	18.7	74.8	80%	59.84
QSFP-DD	8	12.5	100	80%	80
HIB					
PCIe Switch	4	44	176	80%	140.8
BMC (AST2500)	1	1.76	1.76	80%	1.408
Infinity Band Board					
PCIe slot	4	75	300	80%	240
FAN					
9256 FAN	4	140	560	100%	560
VR loss					
Main Chip VR loss (90% efficiency)	1	22.36	22.36	100%	22.36
Total power					6765.85

System Power Budget (350W OAM)

Component	Quantity	Power per item (W)	Design Power (W)	Utilization	Usage Power (W)
UBB					
OAM (12V input, with VR loss)	8	350	2800	100%	2800
PCIe Retimer	8	9.6	76.8	80%	61.44
PHY Retimer	4	18.7	74.8	80%	59.84
QSFP-DD	8	12.5	100	80%	80
HIB					
PCIe Switch	4	44	176	80%	140.8
BMC (AST2520)	1	1.76	1.76	80%	1.408
Infinity Band Board					
PCIe slot	4	75	300	80%	240
FAN					
9256 FAN	4	140	560	100%	560
VR loss					
Main Chip VR loss (90% efficiency)	1	333.48	333.48	100%	333.48
Total power					4276.96

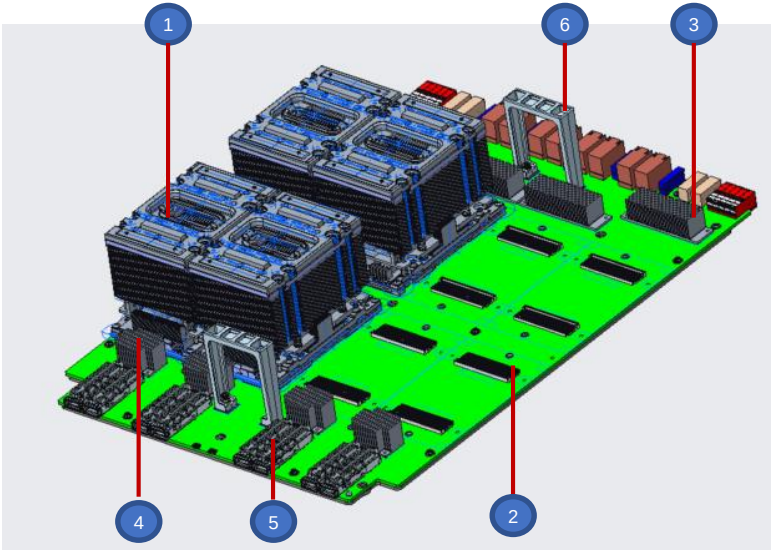
PCBA List



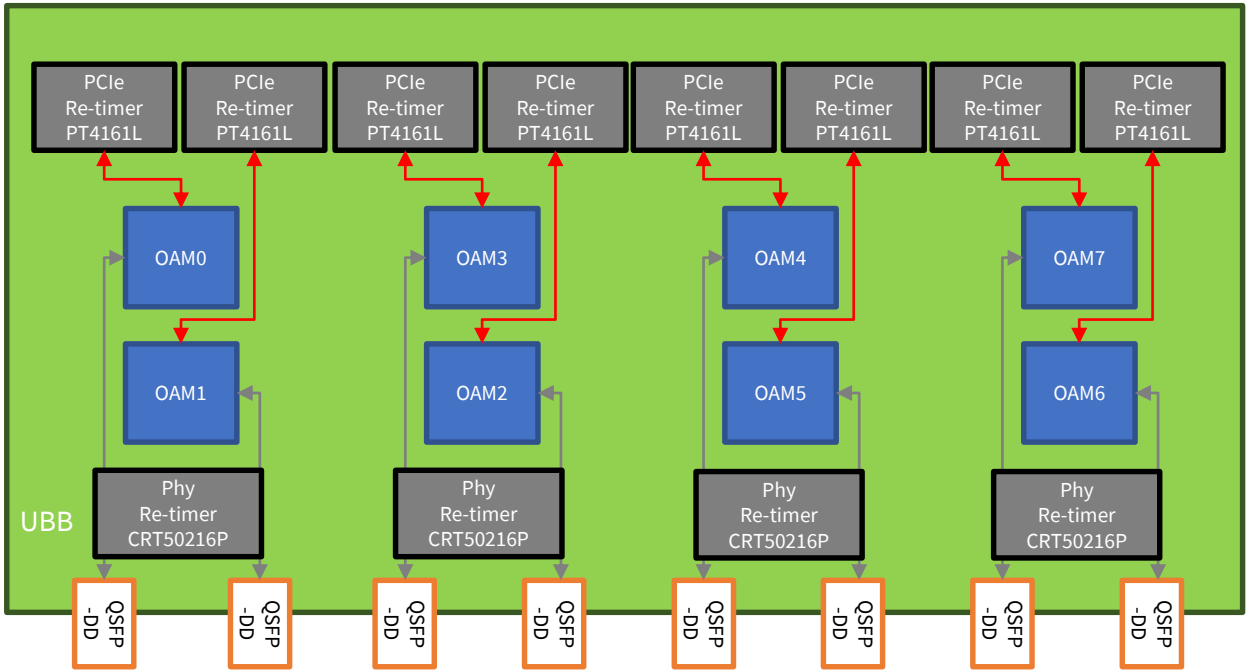
OPEN POSSIBILITIES.

Universal Baseboard Placement

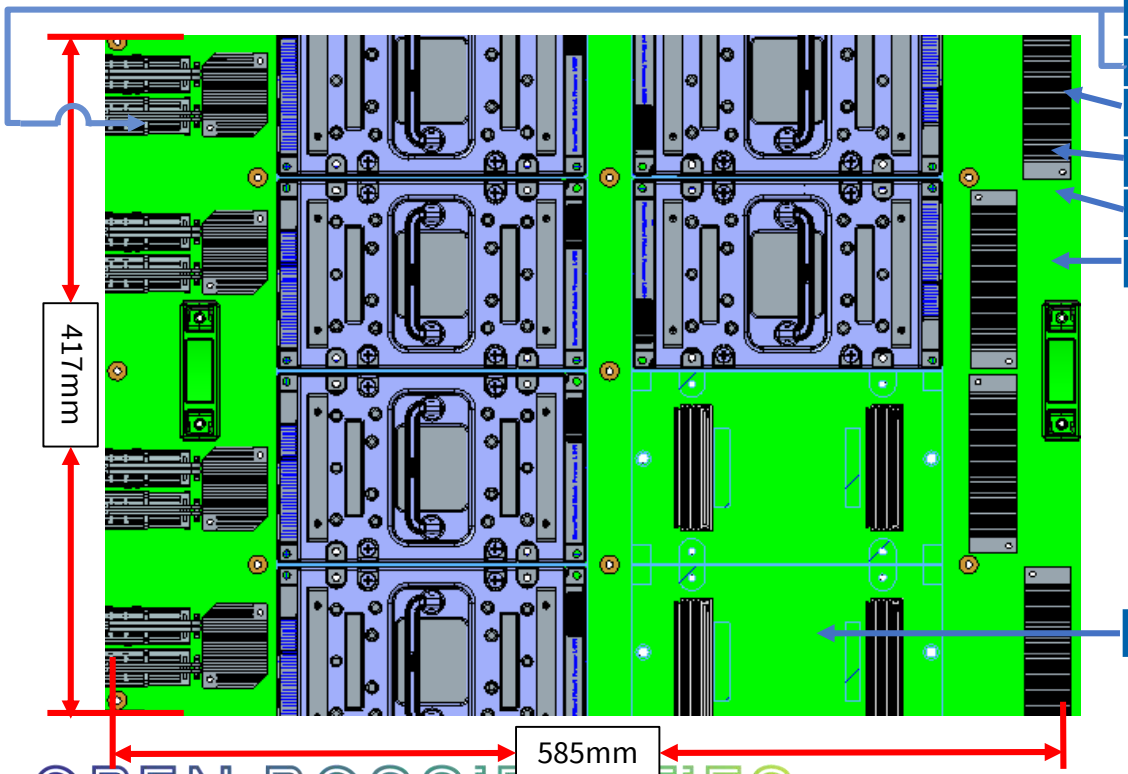
1	8* OAM Module
2	16* OAM Connector
3	8* PCIe Gen4 Retimer
4	4* PAM-4 Retimer
5	8* QSFPDD Connector
6	2* Handle



Universal Baseboard Placement

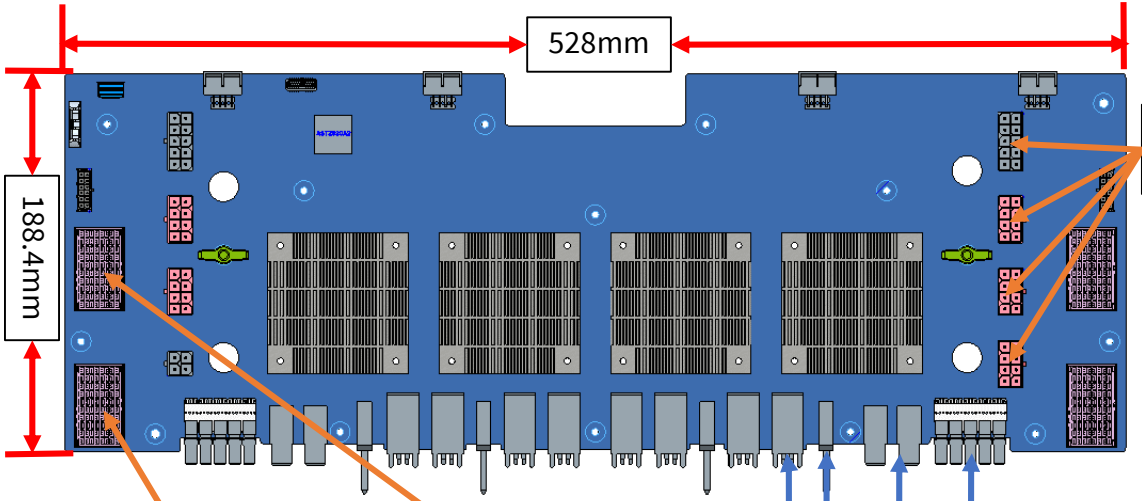


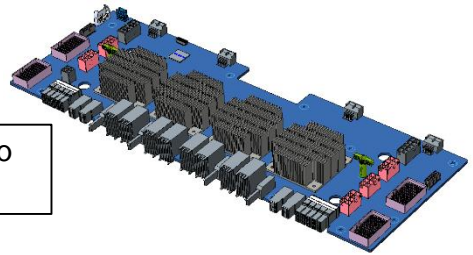
Universal Baseboard Placement



QSFP-DD Connector	Amphenol	UE36-A 1070-3000T
QSFP-DD Cage	Amphenol	UE36-B 16221-06A5A
12V Connector	Amphenol	JX410-513xx
54V Connector	Amphenol	10028917-001LF
Guide Pin	Amphenol	10037909-101LF
High Density Connector	Amphenol	10131762-101LF
Mezzanine Connector	Molex	2093111115

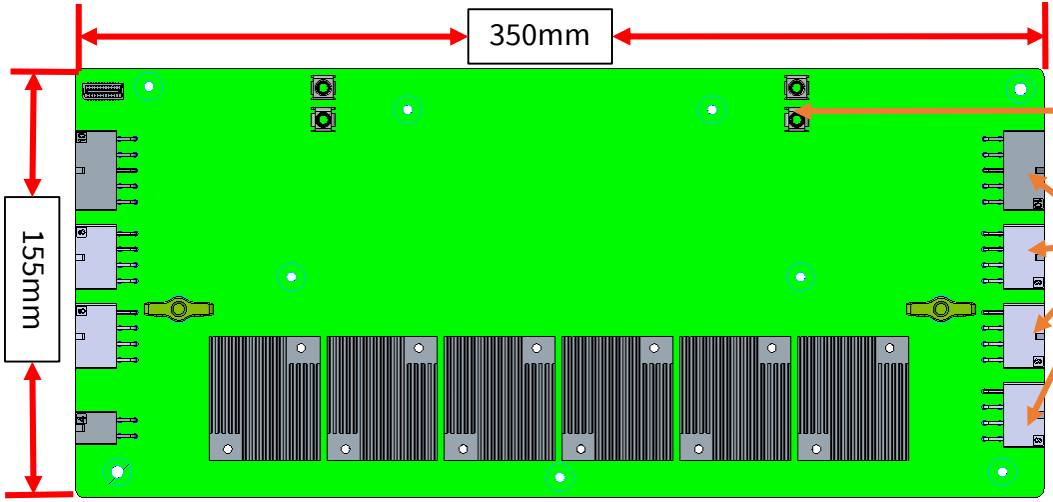
High-speed Interface Board Placement





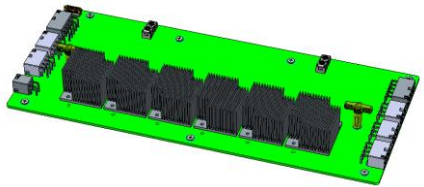
12V Connector	Amphenol	JX412-50432
54V Connector	Amphenol	10052620-5544P00LF
Guide Pin	Amphenol	10044314-101LFC
High Density Connector	Amphenol	10124558-101LF

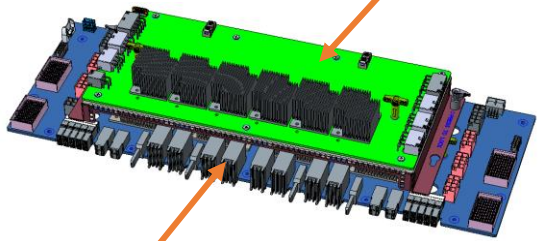
Power Delivery Board Placement



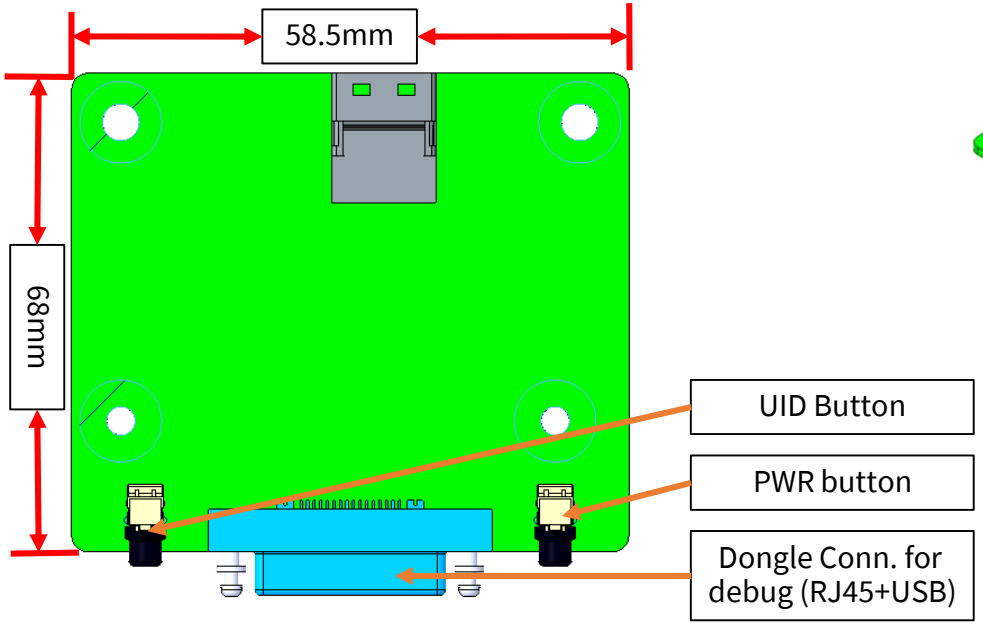
Power Conn. to 54V Busbar

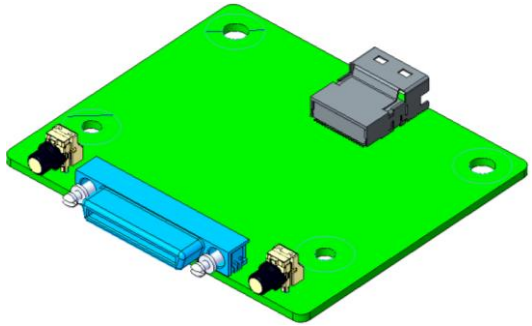
Power Conn. to HIB



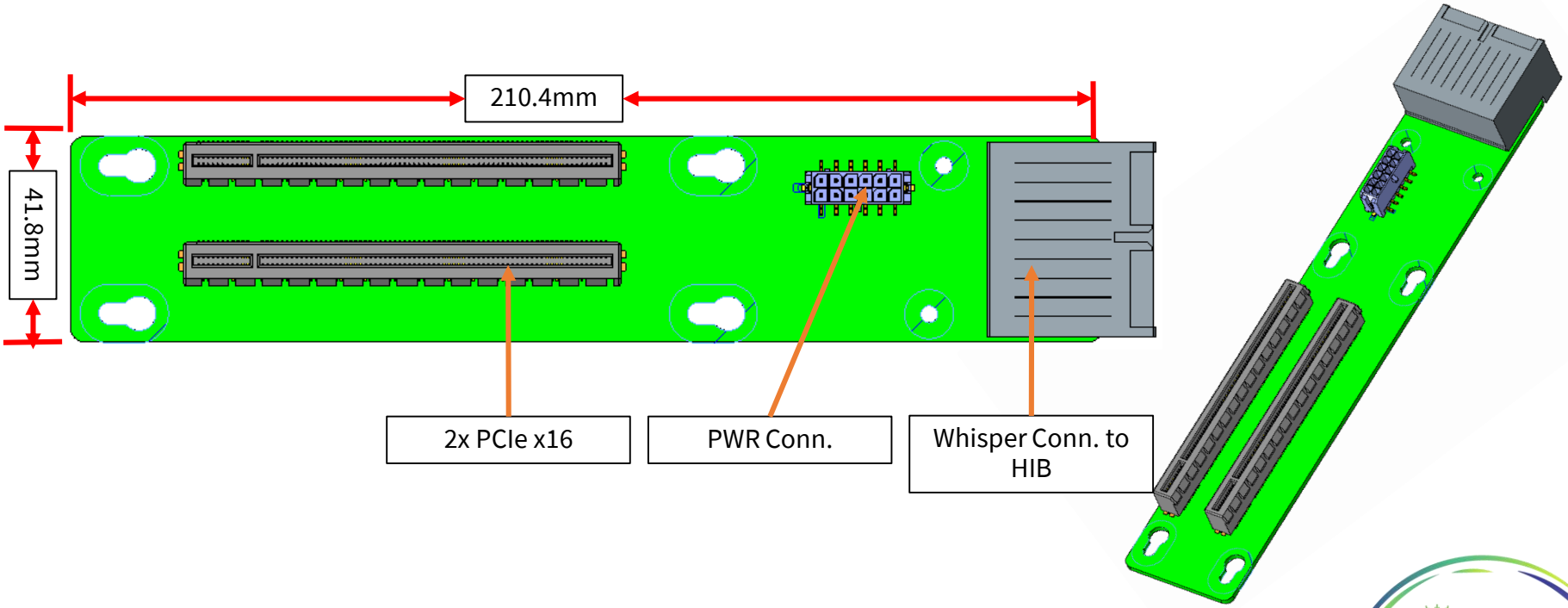


Front IO Board Placement

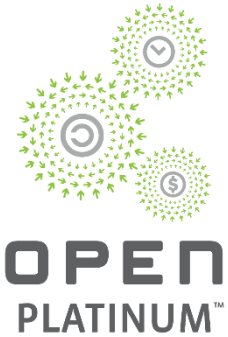




Infinity Band Board Placement

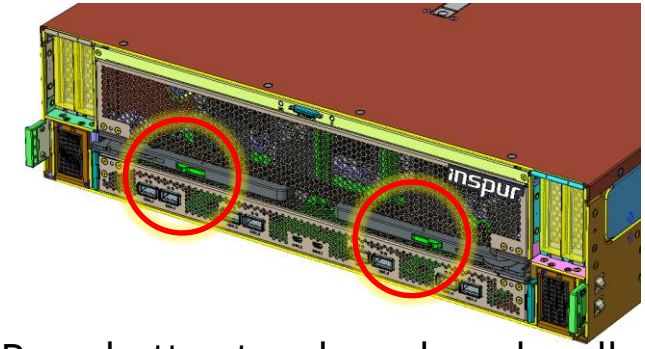


Mechanical Design

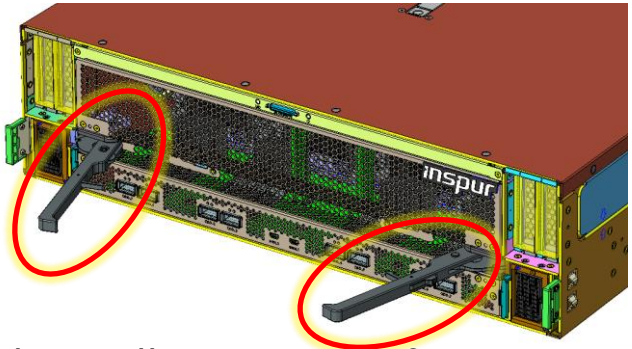


OPEN POSSIBILITIES.

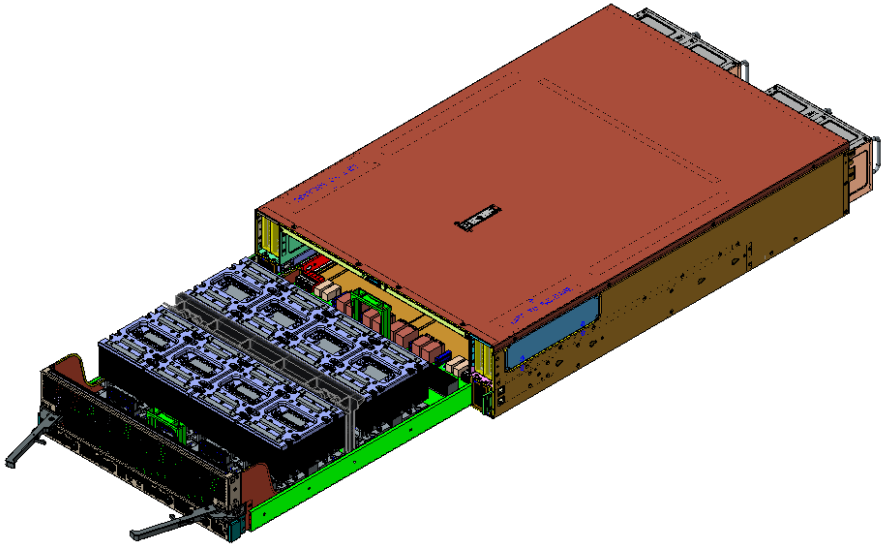
System Access Solution – UBB tray



Press button to release lever handle.



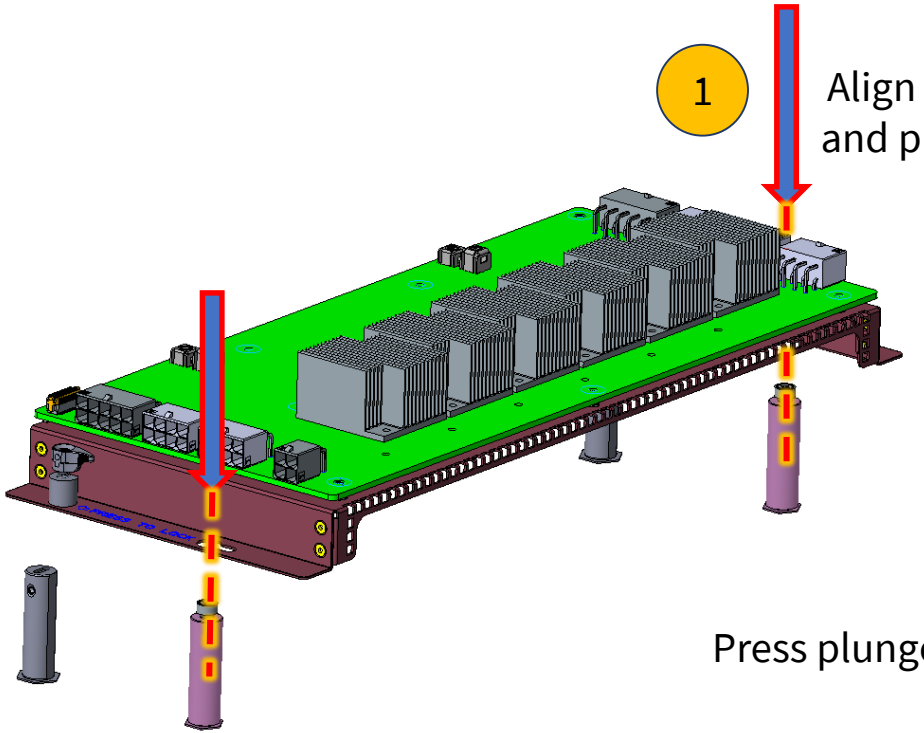
Then pull out UBB tray from system.



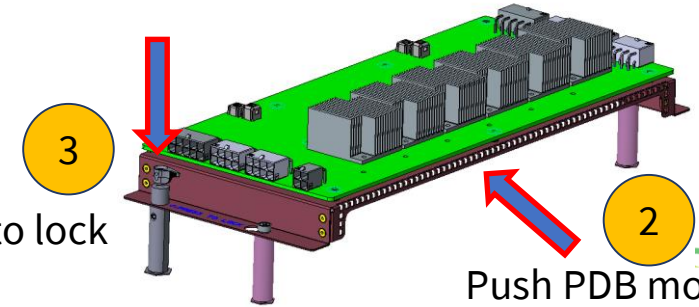
OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

System Access Solution – PDB



1 Align bottle holes to standoffs and put down the PDB module



3 Press plunger to lock

2 Push PDB module to system rear side

OPEN POSSIBILITIES.

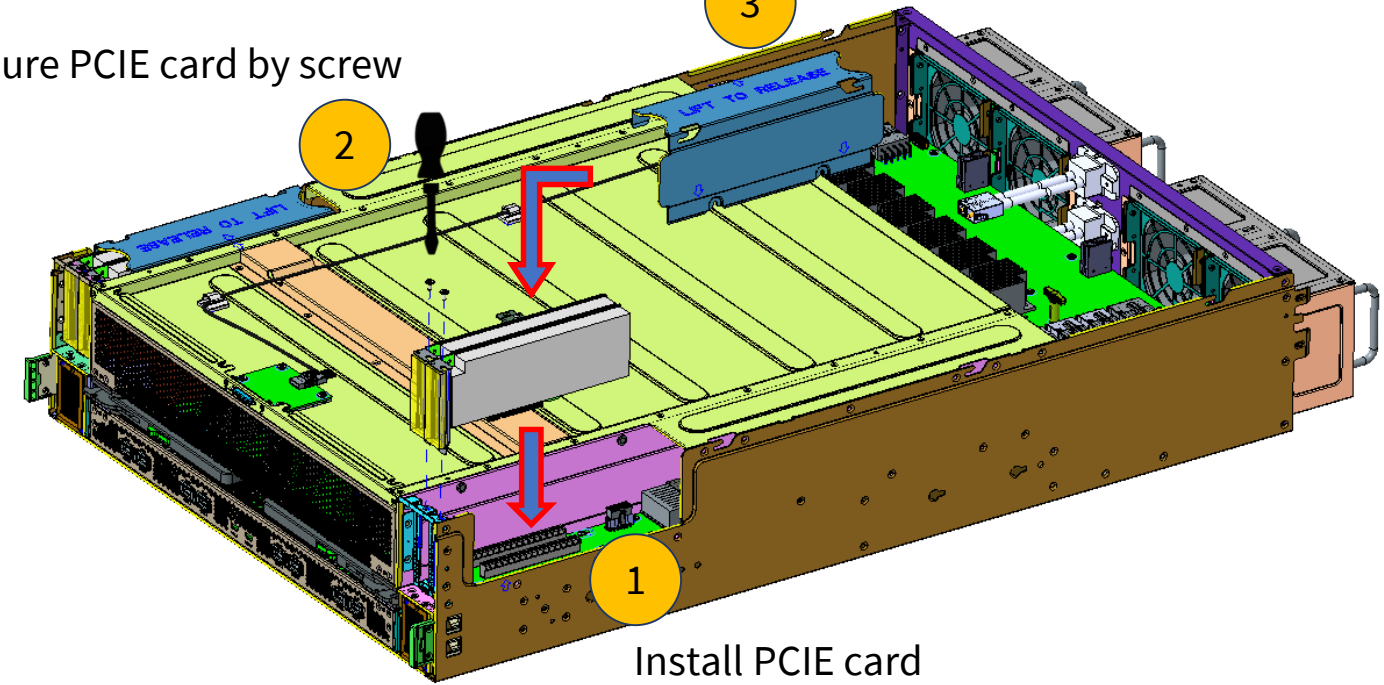
System Access Solution – PCIe cards

Install side cover in place

3

To secure PCIE card by screw

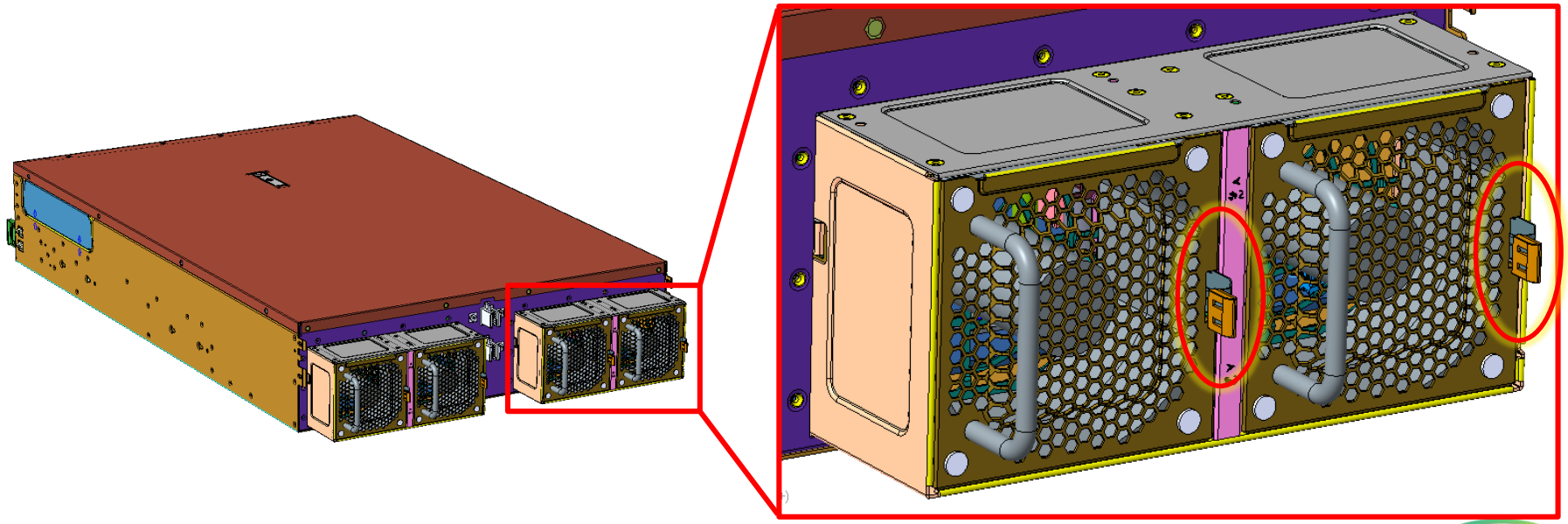
2



1

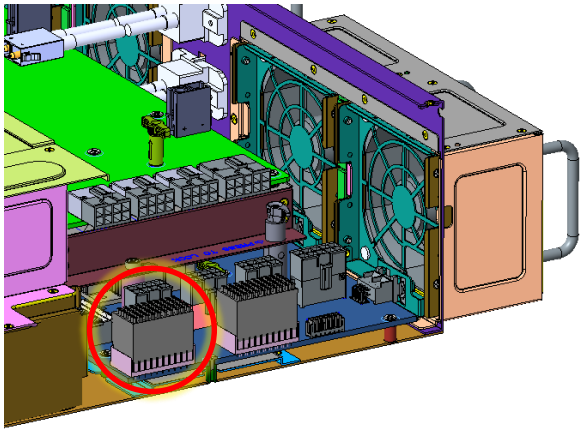
Install PCIE card

System Access Solution – Fan module

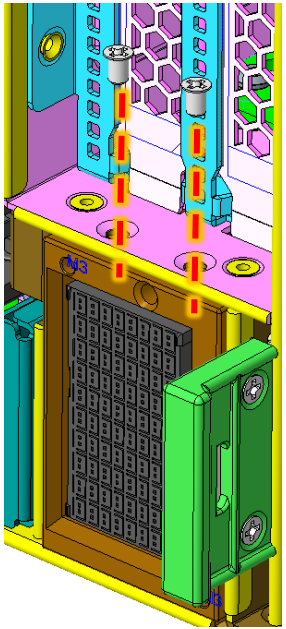


Press latch to remove fan module.

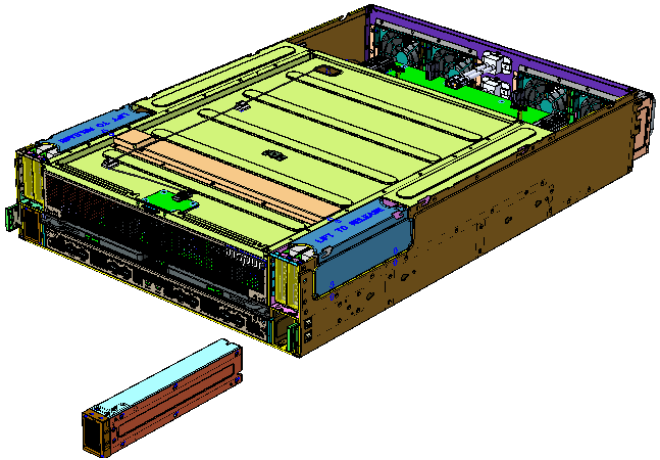
System Access Solution – Whisper cable



Unplug Whisper cable from switch board



Unfasten mounting screws

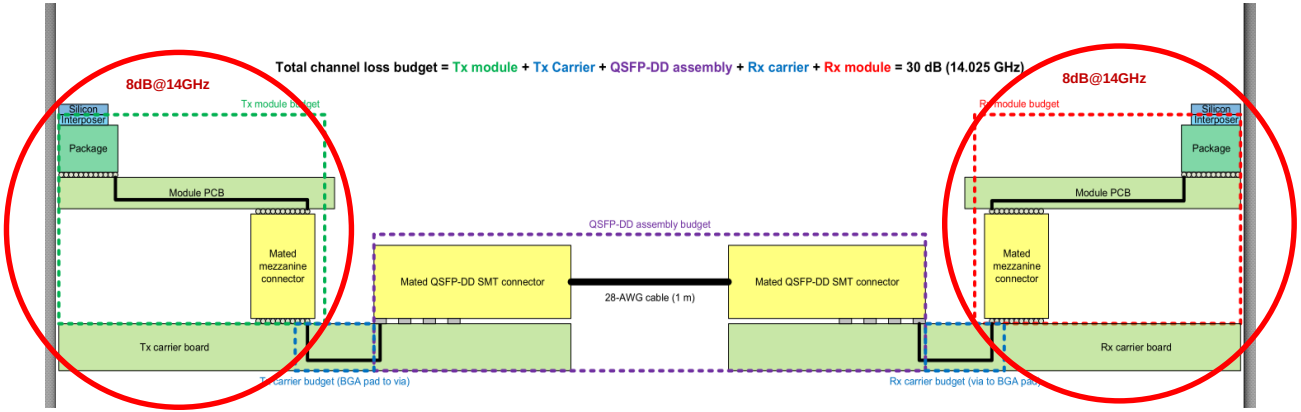


Draw out Whisper cable module

ICL Channel Loss Estimation and Measurement

OPEN POSSIBILITIES.

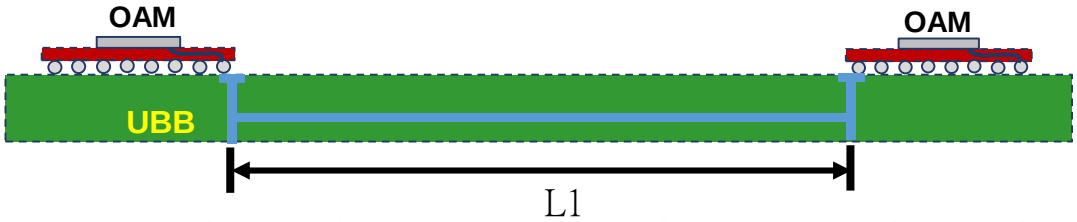
ICL 28Gb Topology



Here is an example for a system base board loss budget:

Die to Die IL Budget from Module supplier	30dB
Total loss on Module @28Gbps	up to 8dB
Base Board IL Budget @28Gbps	30-8*2=14dB

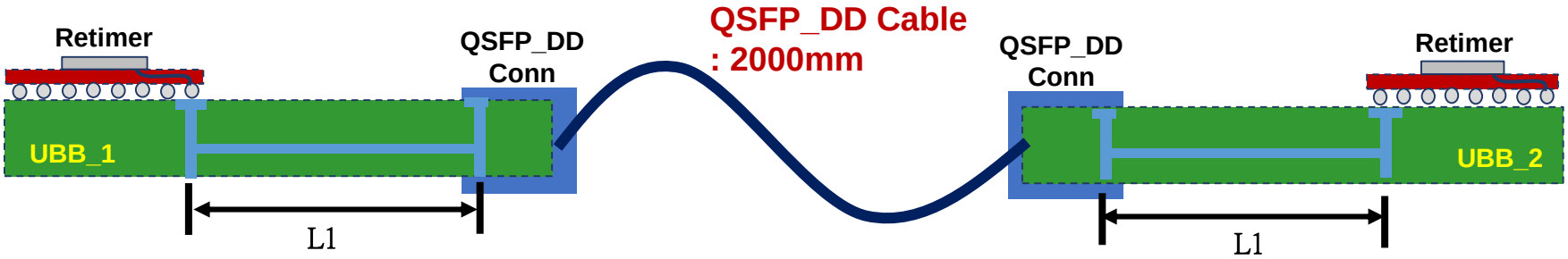
ICL 28Gb NRZ Loss Estimation (I) : OAM to OAM on UBB



ICL NRZ @14.025GHz	OAM PKG	Via (backdrill& Via Stub < 20 mil)	L1	Via (backdrill& Via Stub < 20 mil)	OAM PKG	Total
Ultra Low (inch)	----	----	16.0	---	----	16 inch
Ultra Low (dB)	8.0	0.5	12.96	0.5	8.0	29.96 dB
Channel Loss Spec.	----	----	----	----	----	30 dB

- Assumptions :
- (1) All traces route on inner layer with 90ohm (4 mil-core and 5-mil pp) and the insertion loss of trace is 0.75*1.08dB/inch at 14.025GHz with ultra low loss material (IT-988G). High temperature effect with 8% is included into channel loss estimation.
- (2) Optimization for transition via is necessary. (backdrill, anti-pad control, need to check aspect ratio) Via loss (stub length < 20 mil) is 0.5dB at 14.025GHz.
- (3) OAM package loss is 8dB at 14.025GHz.
- (4) Loss budget is 30 dB for Intel ICL 28Gbs NRZ SPEC.
- **Summary : When routing length with 16 inch on inner layer between OAMs, loss budget margin meets Intel ICL loss criteria. Channel loss evaluation is Middle Risk.**

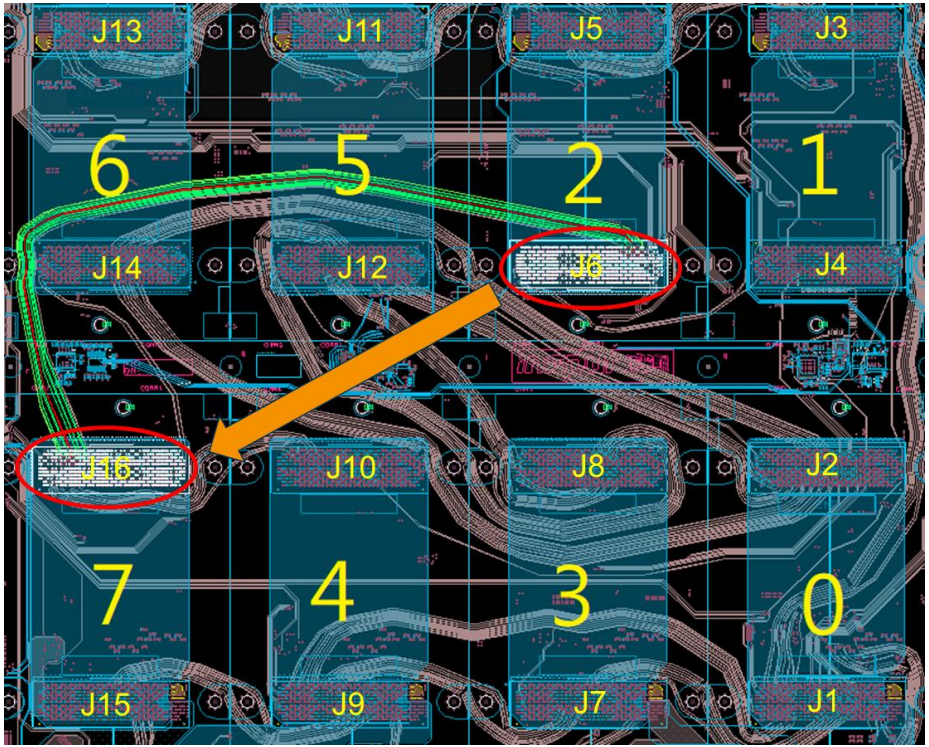
ICL 28Gb NRZ Loss Estimation (II) : QSFP DD Cable Between UBB



ICL NRZ @14.025GHz	Retimer PKG	Via (backdrill& Via Stub < 20 mil)	L1	Via (backdrill& Via Stub < 20 mil)	QSFP_DD Conn_A	QSFP_DD Cable+Paddle	QSFP_DD Conn	Via (backdrill& Via Stub < 20 mil)	L2	Via (backdrill& Via Stub < 20 mil)	Retimer PKG	Total
Ultra Low (inch)	----	----	2.0	---	---	78.74	---	---	2.0	---	----	82.74 inch
Ultra Low (dB)	3.0	0.5	1.62	0.5	1.0	16.72	1.0	0.5	1.62	0.5	3.0	29.96 dB
Channel Loss Spec.	----	----	----	----	----	----	----	----	----	----	----	30 dB

- Assumptions :
- (1) All traces route on inner layer with 90ohm (4 mil-core and 5-mil pp) and the insertion loss of trace is 0.75*1.08dB/inch at 14.025GHz with ultra low loss material (IT-988G). High temperature effect with 8% is included into channel loss estimation.
- (2) Optimization for transition via is necessary. (backdrill, anti-pad control, need to check aspect ratio) Via loss (Backdrill and Via stub length < 20 mil) is 0.5dB at 14.025GHz.
- (3) Credo PHY Retimer package worse case loss is about 3.0dB at 14.025GHz.
- (4) QSFP_DD cable assembly = Conn A+ Paddle Card A + 30AWG Raw Cable + Paddle Card B + Conn B;
  Conn A/B loss is 1.0dB at 14.025GHz ; Paddle Card A/B loss is 0.8dB at 14.025GHz ;
  30AWG raw cable loss is 7.0*1.08dB/m (0.1778*1.08dB/inch) at 14.025GHz. High temperature effect with 8% is included into channel loss estimation.
- (5) Loss budget is 30dB for Intel ICL 28Gbs NRZ SPEC.
- **Summary : When routing length with 2 inch on inner layer and 2-meter cable, loss budget margin meets ICL loss criteria. Channel loss evaluation is Middle Risk.**

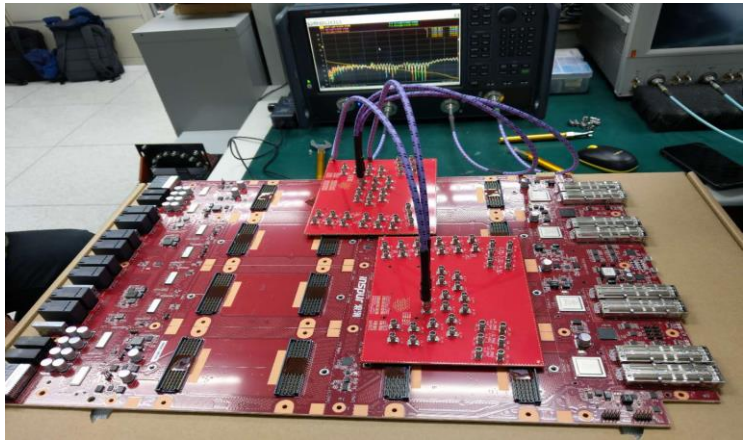
ICL 28Gb NRZ Loss : Test Worse Case with 15.312”

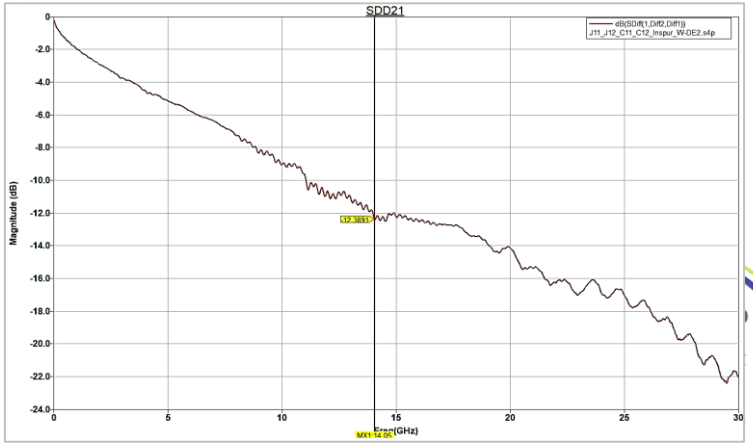


Worse Case	J1: J11, J12 J2: C11, C12	
Net	Pin	Length (mil)
SDS_GPU2_GPU7_S7S7_DN<6>	J16.C11:J6.J11	15312
SDS_GPU2_GPU7_S7S7_DP<6>	J16.C12:J6.J12	15312

ICL 28Gb NRZ Loss : Test Result

- Equipment：
  - N5247A PNA-X 微波網路分析儀，67 GHz
- Test Setting：
  - Frequency: 10MHz – 40GHz
  - Step: 10MHz (4000 points)
  - Port Direction: Odd-Even (1_3_2_4)
- Insertion Loss ：
  - 12.389 dB at 14.05GHz without test fixture



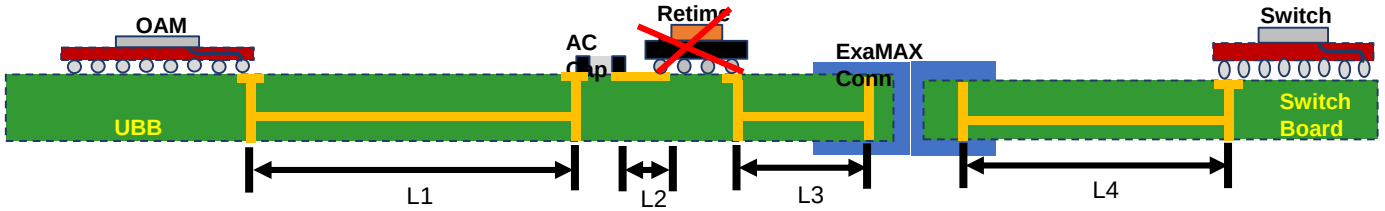


OPEN POSSIBILITIES.

PCIe 4.0 Channel Loss Estimation

OPEN POSSIBILITIES.

PCIE 4.0 Loss Estimation (I) : UBB to HIB without Retimer

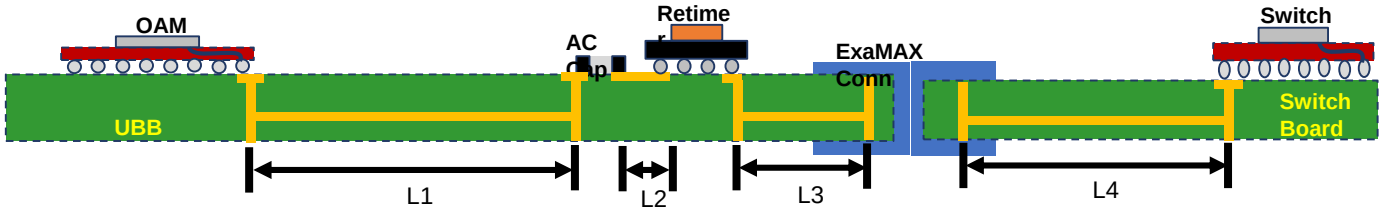


PCIe 4.0 @8GHz	OAM PKG	Via	UBB L1	Via	CAP	Via	UBB L2	Via	UBB L3	Via	ExaMAX Conn	Via	SW L4	Breakout	Via	SW PKG	Total
Ultra Low (inch)	----	----	19.0	----	---	---	1.0	---	4.0	----	----	---	8.5	0.5	---	----	33 inch
Ultra Low (dB)	5.0	0.5	11.90	0.5	1.0	0.5	0.63	0.5	2.51	0.5	1.5	0.5	5.32	0.53	0.5	3.0	34.9 dB
Channel Loss Spec.		----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	28 dB

- Assumptions :
- (1) All traces route on inner layer with 85ohm and the insertion loss of trace is 0.58*1.08 dB/inch at 8GHz with ultra low loss material.
  High temperature effect with 8% is included into channel loss estimation.
- (2) Via/Capacitor loss are 0.5/1.0 dB per one and ExaMAX Conn loss is 1.5 dB at 8GHz.
- (3) Breakout loss 0.99*1.08 dB/inch at 8GHz for ultra low loss material.
- (4) OAM package loss is 5.0dB and Switch package loss is 3.0dB at 8GHz.
- (5) Loss budget is 28 dB for PCIE Gen4.

- Summary : When routing length without Retimer, loss budget margin don't meets PCIe 4.0 loss criteria.
- Channel loss evaluation is High Risk.

PCIE 4.0 Loss Estimation (II) : UBB to HIB with Retimer

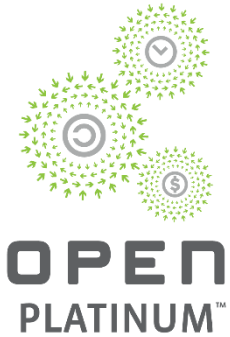


PCIe 4.0 @8GHz	OAM PKG	Via	UBB L1	Via	CAP	Via	UBB L2	RE PKG	Via	UBB L3	Via	ExaMAX Conn	Via	SW L4	Breakout	Via	SW PKG	Total
Ultra Low (inch)	----	----	19.0	----	---	---	1.0	---	----	----	----	----	----	----	----	----	----	20 inch
Ultra Low (dB)	5.0	0.5	11.90	0.5	1.0	0.5	0.63	2.0	----	----	----	----	----	----	----	----	----	22.03 dB
Ultra Low (inch)	----	----	----	----	----	----	----	---	---	4.0	----	----	---	8.5	0.5	---	----	13 inch
Ultra Low (dB)	----	----	----	----	----	----	----	2.0	----	2.51	0.5	1.5	0.5	5.32	0.53	0.5	3.0	16.37 dB
Channel Loss Spec.		----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	28 dB

- Assumptions :
- (1) All traces route on inner layer with 85ohm and the insertion loss of trace is 0.58*1.08 dB/inch at 8GHz with ultra low loss material.
  High temperature effect with 8% is included into channel loss estimation.
- (2) Via/Capacitor loss are 0.5/1.0 dB per one and ExaMAX Conn loss is 1.5 dB at 8GHz.
- (3) Breakout loss 0.99*1.08 dB/inch at 8GHz for ultra low loss material.
- (4) OAM package loss is 5.0dB and Switch package loss is 3.0dB at 8GHz.
- (5) Loss budget is 28 dB for PCIE Gen4.

- Summary : When routing length with Retimer, loss budget margin meets PCIe 4.0 loss criteria on before Retimer side and after Retimer side.
- Channel loss evaluation is Low Risk.

Thermal Simulation

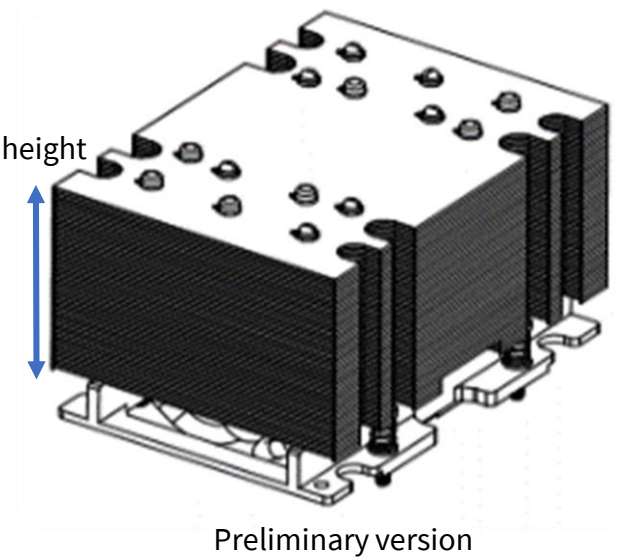


OPEN POSSIBILITIES.

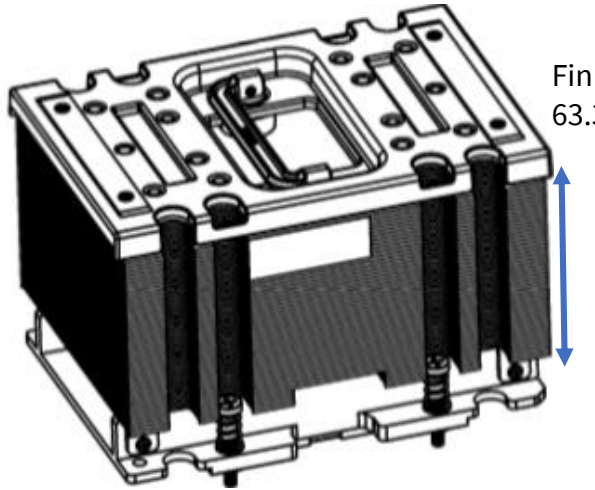
OAM Heatsink Design For Common Tray

- ✓Add top cover & HS module handle for accessing
- ✓Reduce fin numbers to meet common UBB tray specification

Fin area height
66.1mm



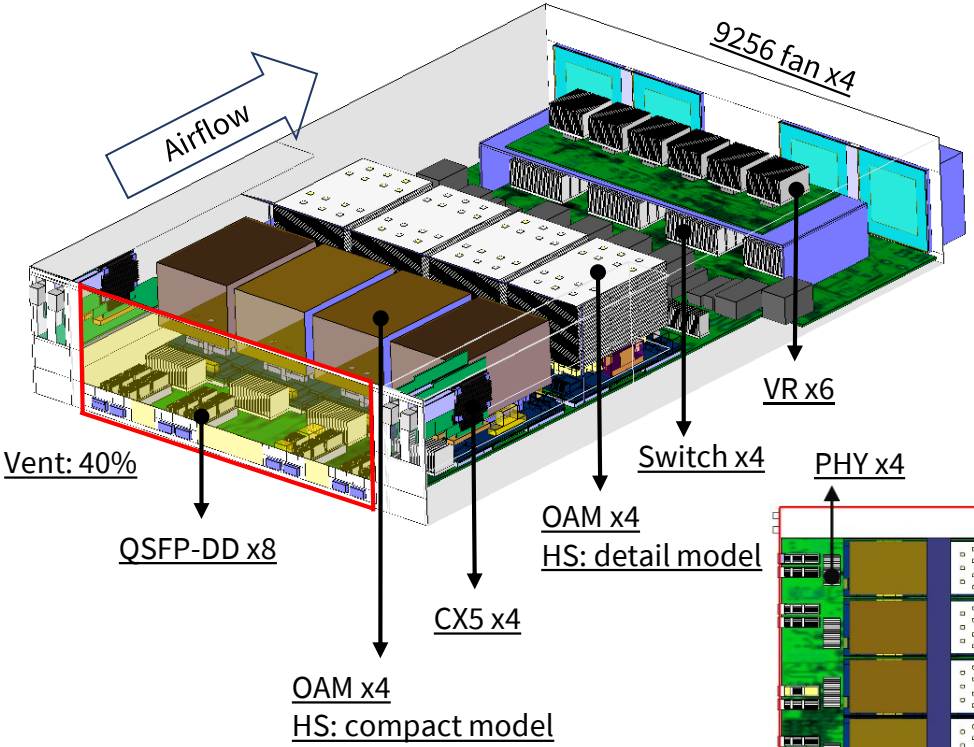
Preliminary version

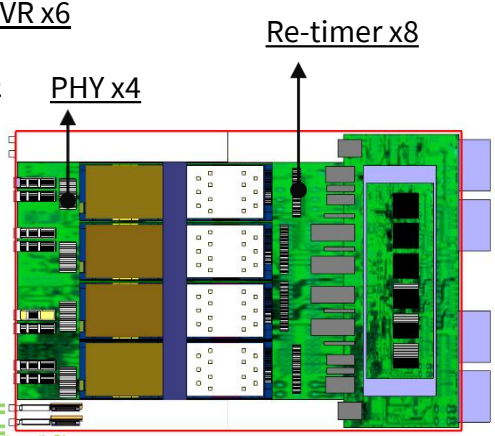


Fin area height
63.3mm

Current version

Power Deliver Board Placement





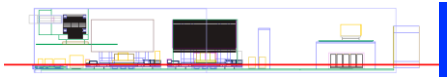
Top view

Thermal Design Power			
Component	Unit Power	Q'ty	Power (W)
OAM	450	8	3600
Re-timer	9.6	8	76.8
Switch	44	4	176
PHY	18.7	4	74.8
QSFP-DD	12.5	8	100
CX5	18.1	4	72.4
VR	25	6	150
9256 Fan	140	4	560
Total Power (watt)			4810

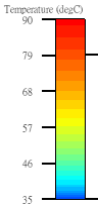
Simulation Results (Temperature Field)

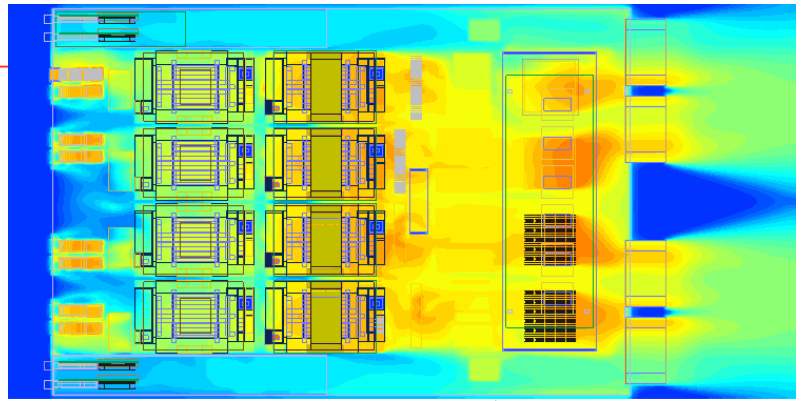
Simulation Boundary Condition Setup

- ✓ TDP at 35C ambient
- ✓ Fan normal
- ✓ Sea-level
- ✓ Compact model for upper stream OAM HS
- ✓ Compact model HS for VR
- ✓ Symmetric placement for switch HS two for detail model; two for compact model



Cut plane: y=17.5mm





Airflow

Simulation Result			
Location		Spec (C)	Temp.
OAM	ASIC	115.0	76.5
	HBM	95.0	93.3
PCIe Re-timer		85.0	79.5
PCIe Switch		115.0	90.8
PHY Re-timer		125.0	78.7
QSFP-DD		70.0	66.9
CX5		105.0	94.1
VR		125.0	108.7

Test by Nervana OAM 280W x8pcs,

Thank you!