OCP Accelerator Module and The Infrastructure

ODSA Project Workshop

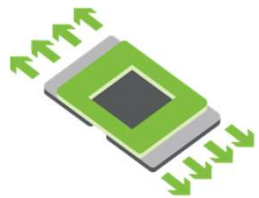
March 28, 2019

Consume. Collaborate. Contribute.

Outline

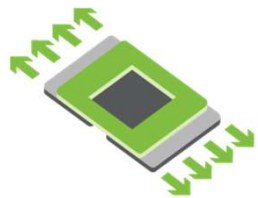
- Motivation
- Approach
- Examples
- Requesting Participation and Feedback

Motivation

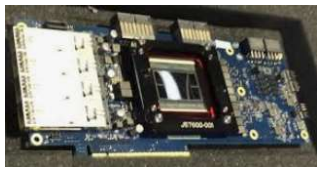


AI's rapid evolution is producing an explosion of new types of hardware accelerators for Machine Learning (ML) and Deep Learning (DL)

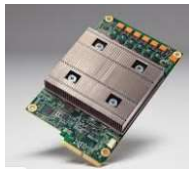
GPU | FPGA | ASIC | NPU | TPU | NNP | IPU | xPU…

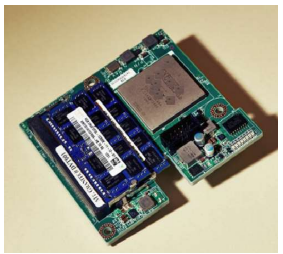


Varied Module and System Form Factors



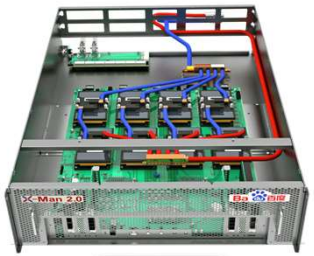
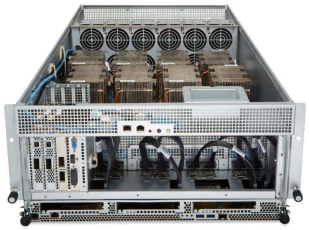
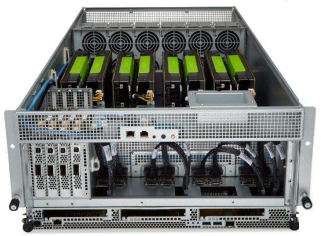










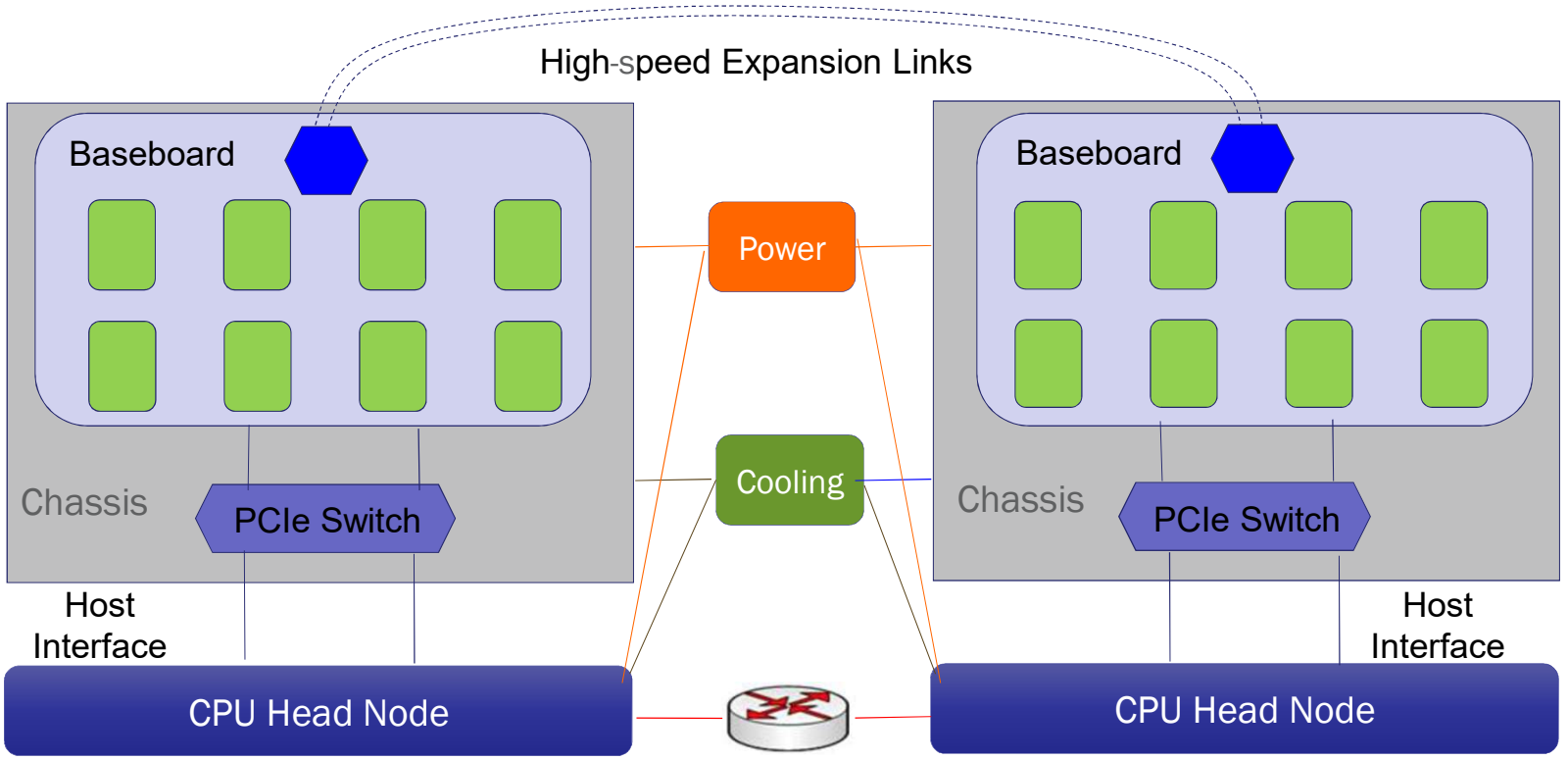





HPC

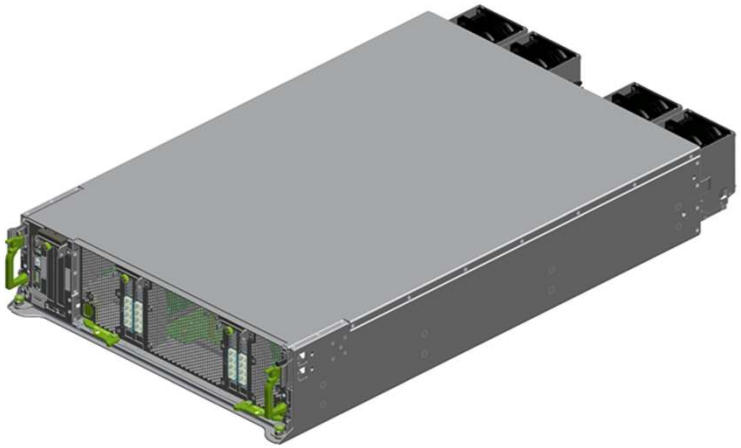
Different Implementations

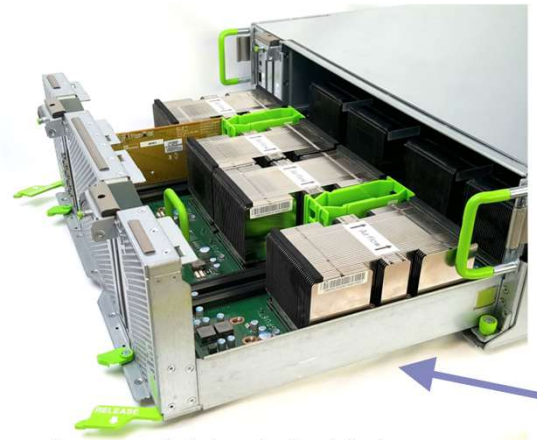
Targeting Similar Requirements!

Logical Components for AI Hardware System



Facebook Big Basin System



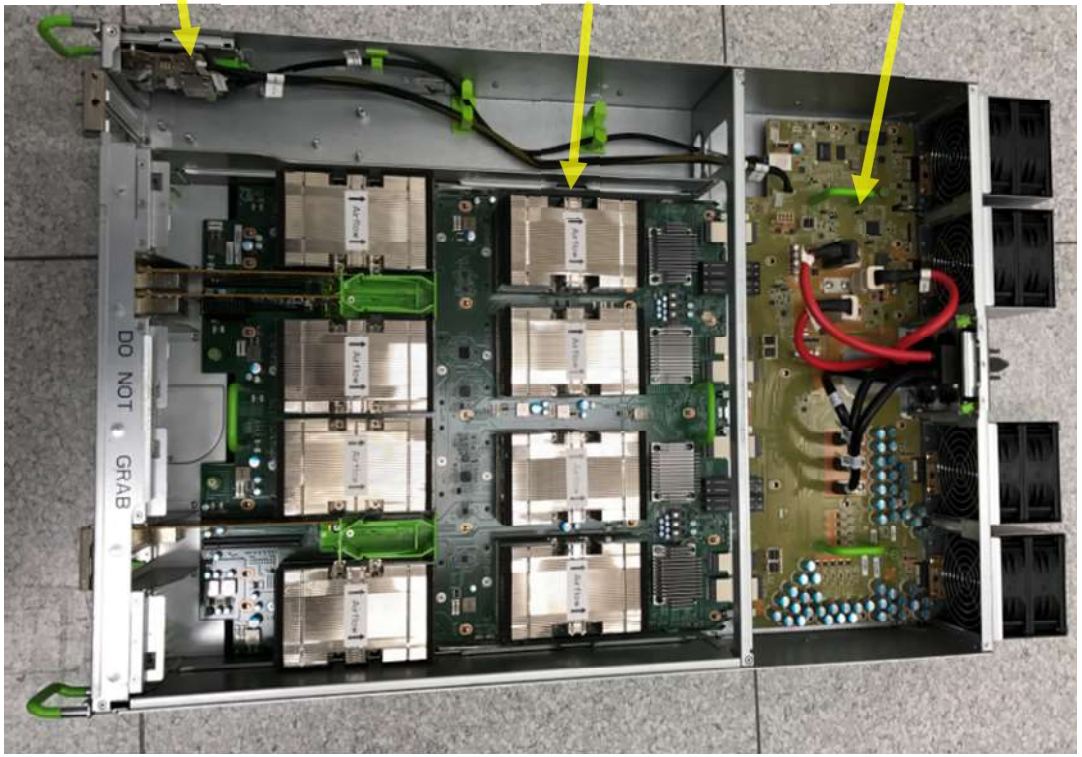


Baseboard on sliding tray

IO board

Baseboard

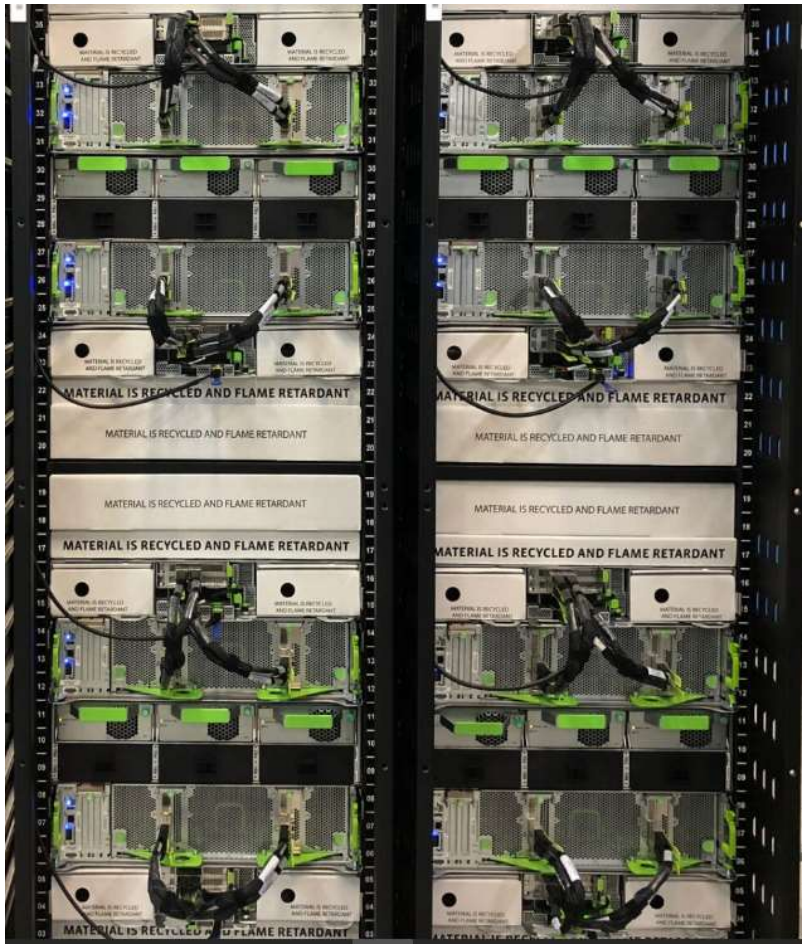
Midplane board



Big Basin Rack View

- 2S server Tioga Pass as head node
- Open Rack v2, 12.6kw
- 4 Big Basin per Rack

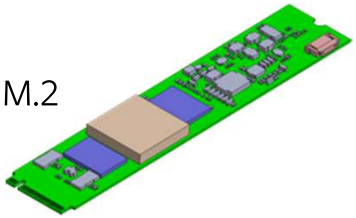
Head node Tioga Pass



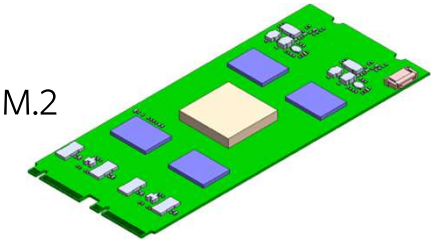
Facebook Inference/Video Accelerator Common System

Inference Module

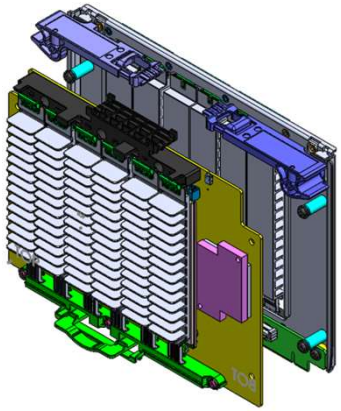
Single M.2



Dual M.2

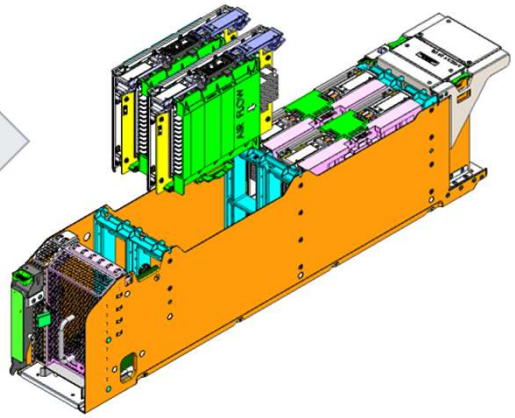


Carrier Card



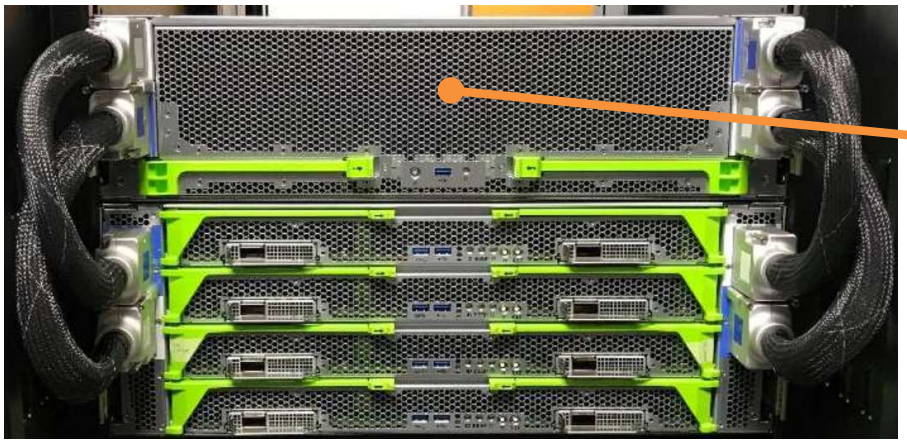
Up to 12x single /
6x dual M.2

Yosemite V2



2x Twin Lakes with 2x carrier cards

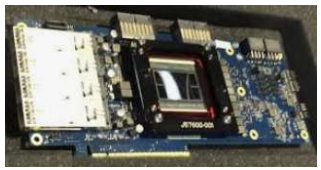
Facebook Training System



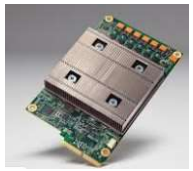
8* Socket system with 8* Accelerators

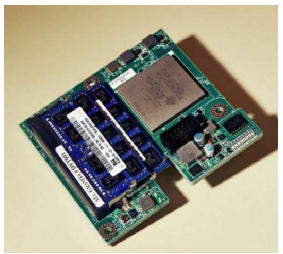
8* OCP Accelerator Modules

Varied Module and System Form Factors



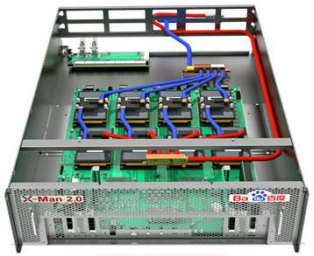
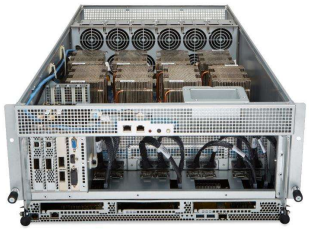
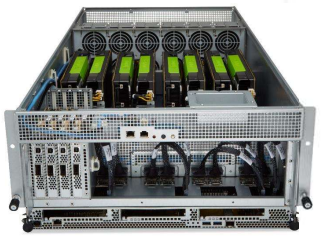











Common Requirements for Accelerator System

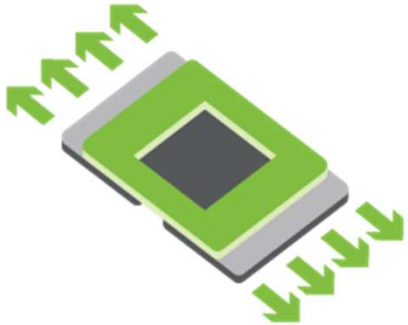
- Flexibility
- Robustness & Serviceability
- Configuration, Programming, & Management
- Power & Cooling
- Inter-module Communication to Scale Up
- Input / Output Bandwidth to Scale Out

"If you want to go *Fast*, go *Alone*;
If you want go *Far*, go *Together*"

We have done *Fast for Short-term result*;

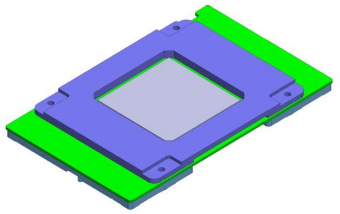
It is time to go *Far* at OCP for
Long-term gain!

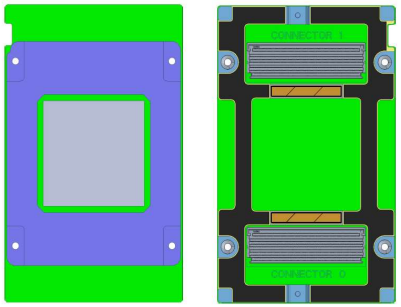
We Started from OAM.



OCP Accelerator Module(OAM) Spec

- 102mm x 165mm Module Size
- With two high-speed Mirror Mezz connectors
- 12V and 48V input DC Power
- Up to 350w (12V) and up to 700w (48V) TDP
  - Up to 450W (air-cooled) and 700W (liquid-cooled)
- Support single or multiple ASIC(s) per Module
- Up to eight x16 Links (Host + inter-module Links)
  - Support one or two x16 High speed link(s) to Host
  - Up to seven x16 high speed interconnect links
- Up to 8* Modules per Baseboard
- System management and debug interfaces



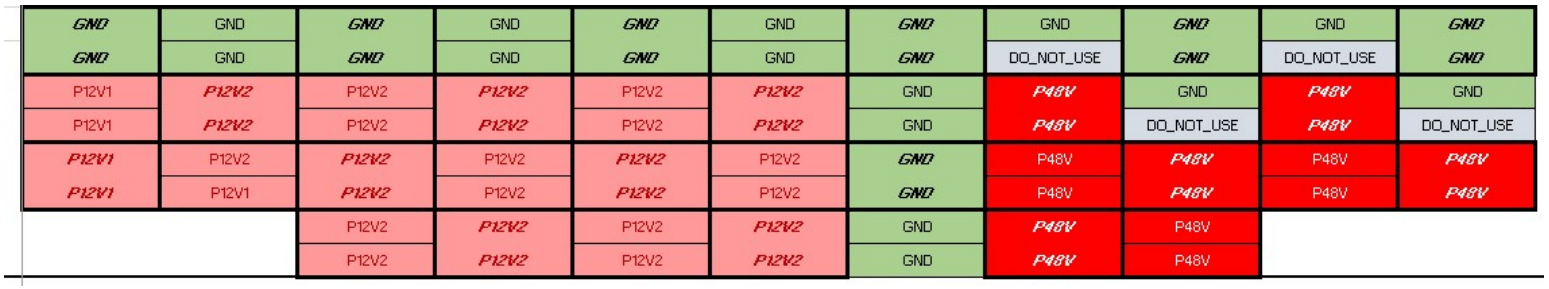



NEW PROJECT
OCP ACCELERATOR MODULE
facebook
Microsoft Azure
intel
Qualcomm
XILINX
habana
NVIDIA
BittWare a molex company
Baidu 百度
Google
inspur
Lenovo
AMD
IBM
wiwynn
Alibaba Group
Tencent
molex
PENGUIN COMPUTING
A subsidiary of SMART Global Holdings, Inc.
GRAPHCORE
HUAWEI
QCT

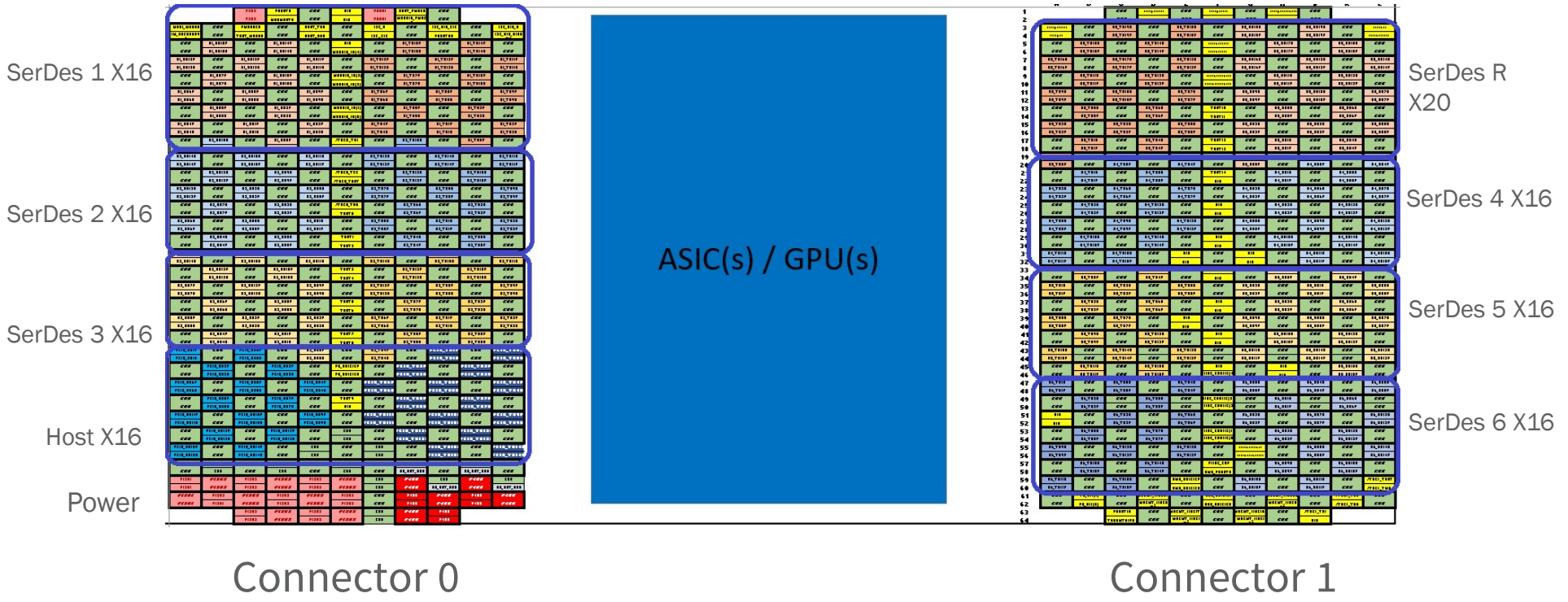
OAM Power

- Support both 12V and 48V as input
  - 12V to support up to 350w TDP
  - 48V to support up to 700w TDP

Power Rail	Voltage Tolerance	# of pins	Current Capability	Status
P12V	11V min to 13.2V max	27	27A (when at 11V)	Normal Power
P12V Mandatory	11V min to 13.2V max	5	5A (when at 11V)	Normal Power
P48V	44V min to 60V max	16	16A (when at 44V)	Normal Power
P3.3V	3.3V±10% (max)	2	2A	Normal Power

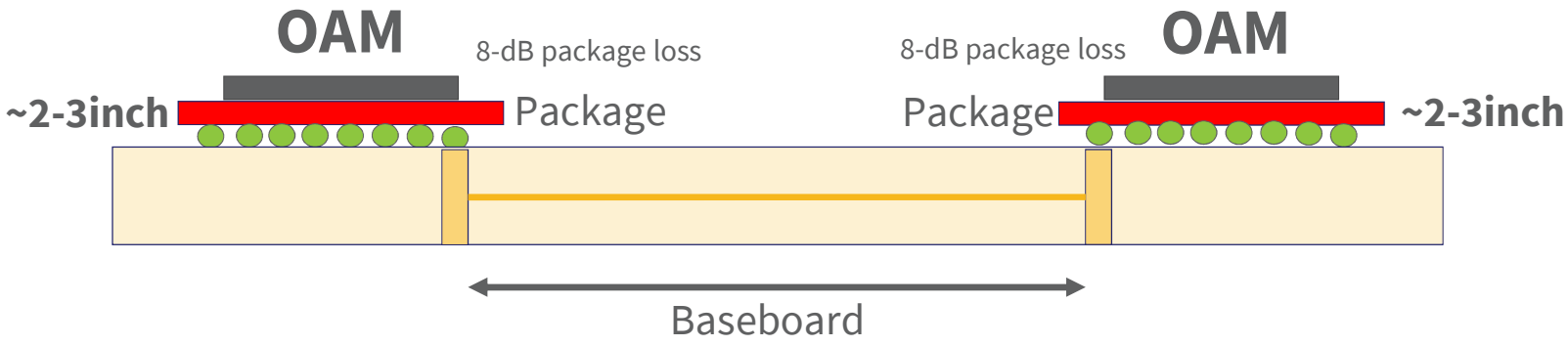


OAM Pin Map


SerDes 1 X16
SerDes 2 X16
SerDes 3 X16
Host X16
Power
ASIC(s) / GPU(s)
SerDes R X20
SerDes 4 X16
SerDes 5 X16
SerDes 6 X16
Connector 0
Connector 1

Interconnect end-to-end Channel Loss

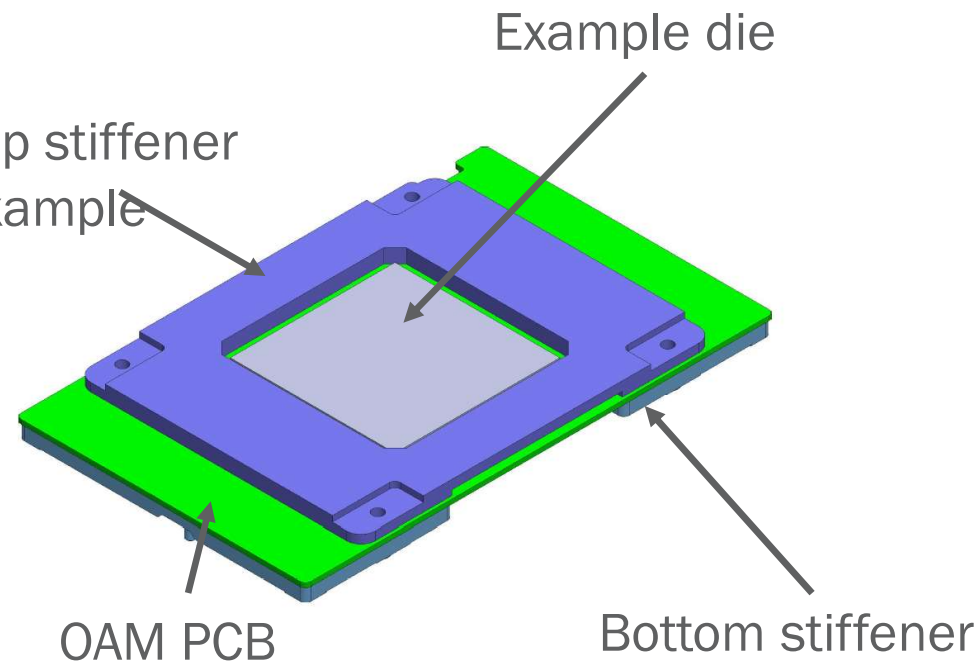
- The module interconnection channel total insertion loss @28Gbps should not be over -8dB
- System baseboard IL budget = Die to Die IL from each OAM supplier – 16dB

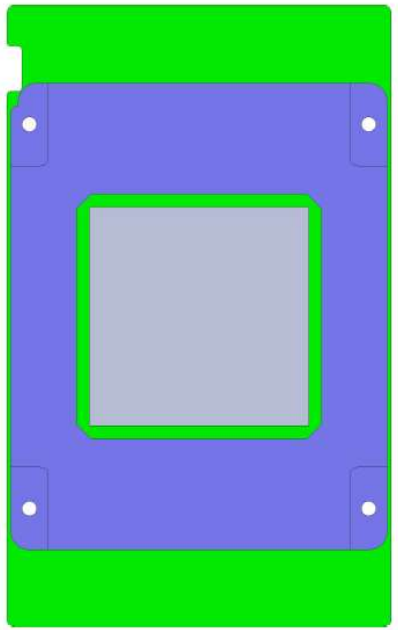


System Management/Debugging

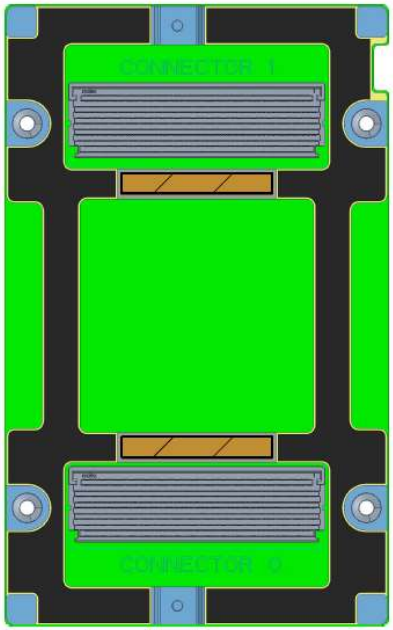
- Sensor reporting
- Error monitoring/Reporting
- Firmware Update
- Power Capping
- FRU Information
- IO Calibration
- JTAG/I2C/UART interfaces for debugging

Overview OAM: Mechanical/Thermal





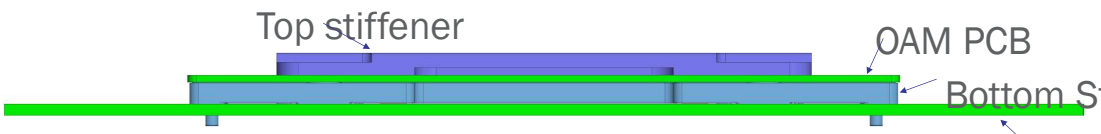
TOP VIEW



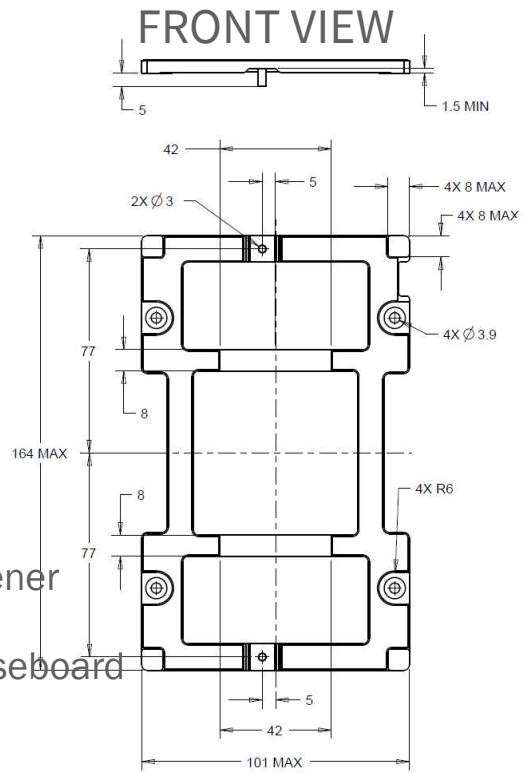
BOTTOM VIEW

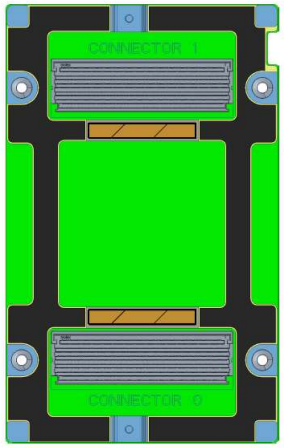
Mech Requirements – OAM Bottom Stiffener

- 5±0.15mm stiffener as required by connector
- 3mm alignment pins that extend 10mm below OAM PCB surface
- Die spring (rectangular profile coil spring) to provide unmate force
- EMI gaskets for grounding to baseboard



SIDE VIEW, FULL ASSEMBLY
(BOTTOM STIFFENER IN BLUE)



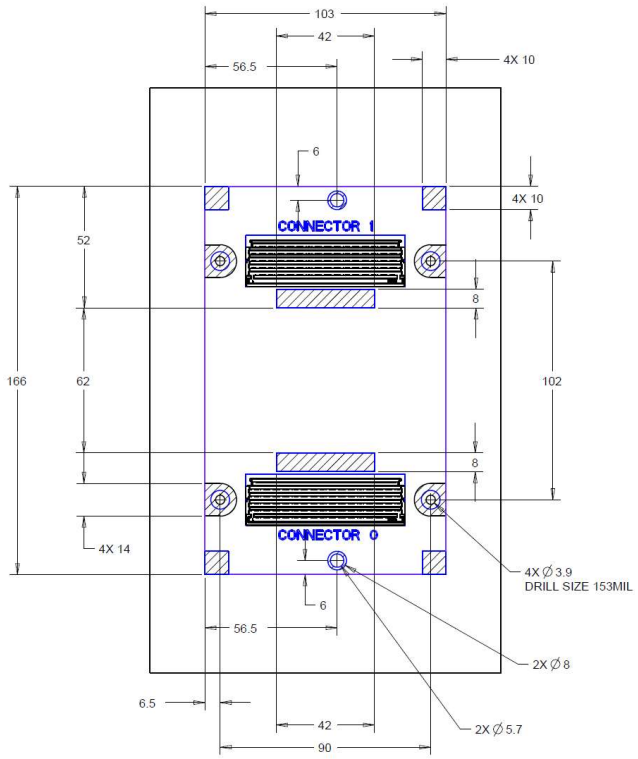


BOTTOM VIEW

Mech Requirements – System Baseboard

- Component KOZ 103 x 166mm: 0mm height
- Cross-hatched locations: Grounding Pads
- EMI grounding pads located north and south of the connectors
- 4x Mounting Holes for M3.5 screws
- 2x SMT nuts used as alignment features

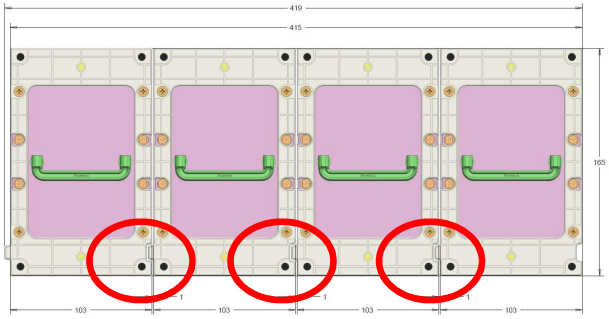
TOP VIEW



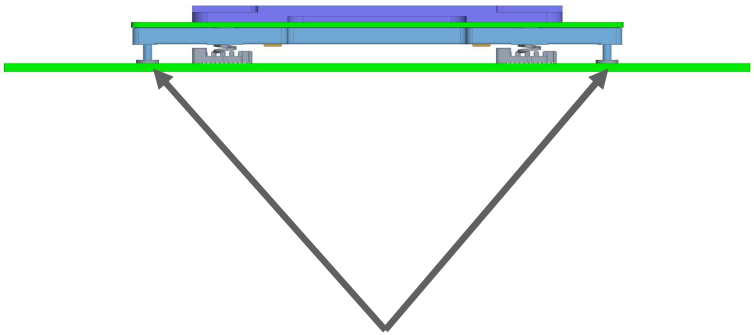
Mech Recommendations – Alignment Features

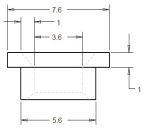
- Notch provides orientation and keying (OPTIONAL, BUT RECOMMENDED)

TOP VIEW



SIDE VIEW



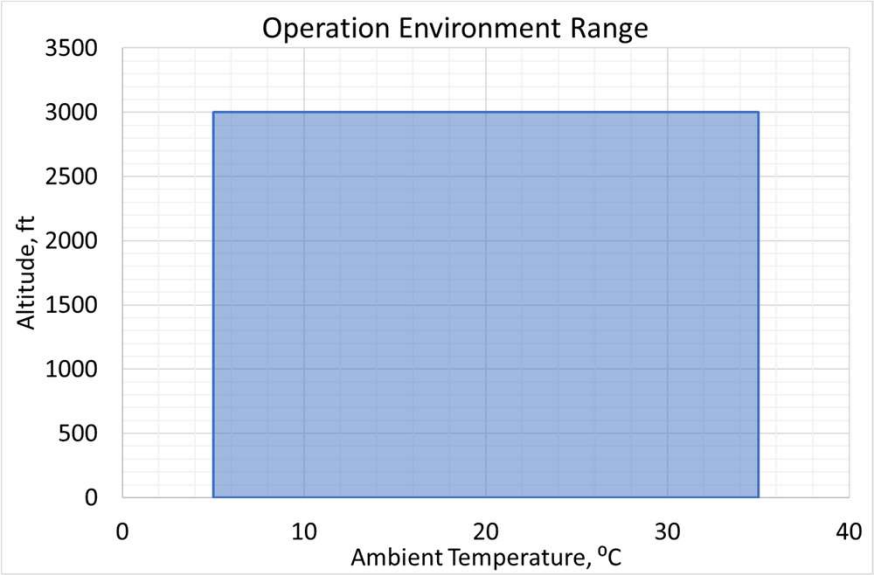


SMT nuts on baseboard

Molex Mirror Mezz Connector Gatherability: 0.76mm

Thermal Requirements – Operation Environment

- Ambient Temp: 5°C to 35°C
  - Approach Temp: 5°C to 48°C
- Altitude: sea level to 6000ft
- Humidity: 20% to 90%
- Cold boot temp limit: TBD
- Storage temp: -20°C to 85°C



- No ambient temp compensation/de-rating for altitude

Now we have a industry standard OAM spec, what's the next?

We need an

Open

Accelerator Infrastructure

Hierarchical Base Specification for OAI

Well-defined boundaries
Fostering Innovation

- OAM
- UBB (Interconnect Topology)
- Switch Board
- SCM
- Tray
- Chassis

- Power and Cooling
- Mechanical
- Electrical
- Security & Management

Designs and Products may be compliant to any or all specifications

The Universal Baseboard (UBB)

Consume. Collaborate. Contribute.

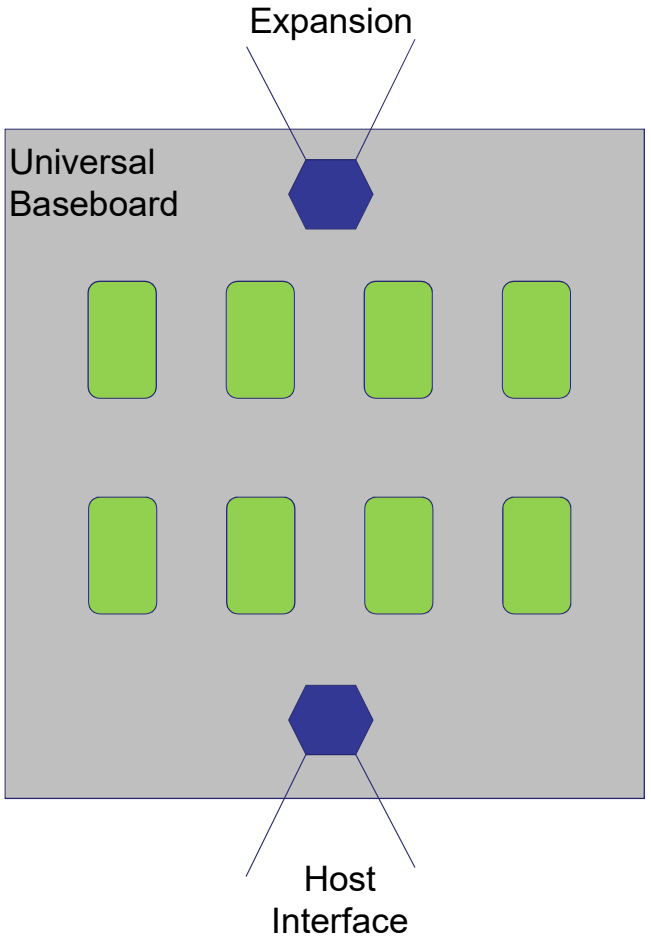
Different Neural Networks and Frameworks for Model or Data Parallelism

Benefit from different

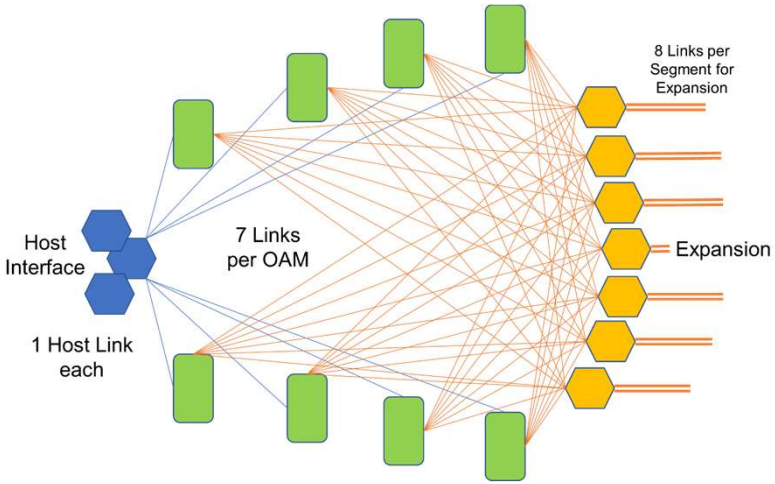
Interconnect Topologies

Universal Baseboard (UBB)

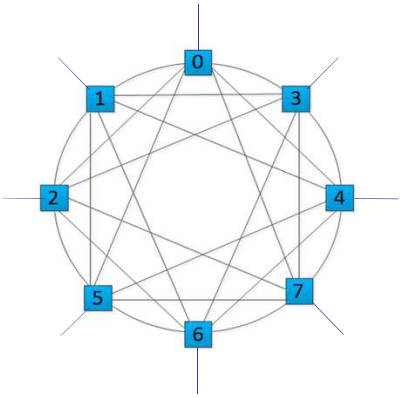
- Consider a Grid of Planar OAM sites
- Standard Volumetric
- Protocol Agnostic Interconnects
- *Wires are Wires!*



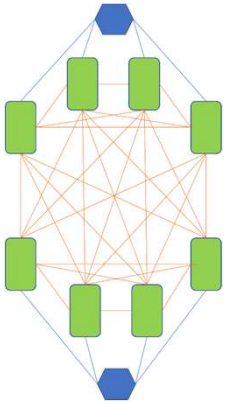
With different interconnect topologies



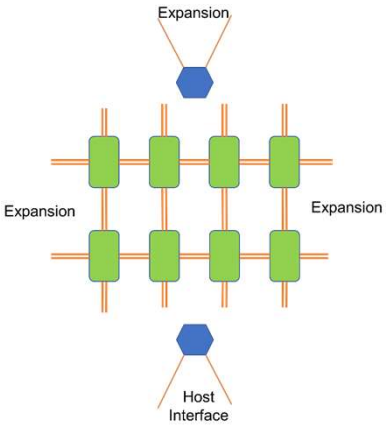
A Grid of interconnected OAMs,
Max Bisection BW
One Hop Away
Ready for Expansion



With six inter-OAM Links and one Host Link



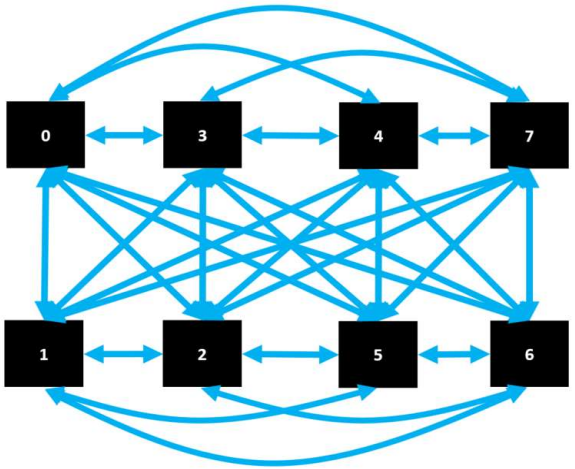
With seven inter-OAM Links and one Host Link

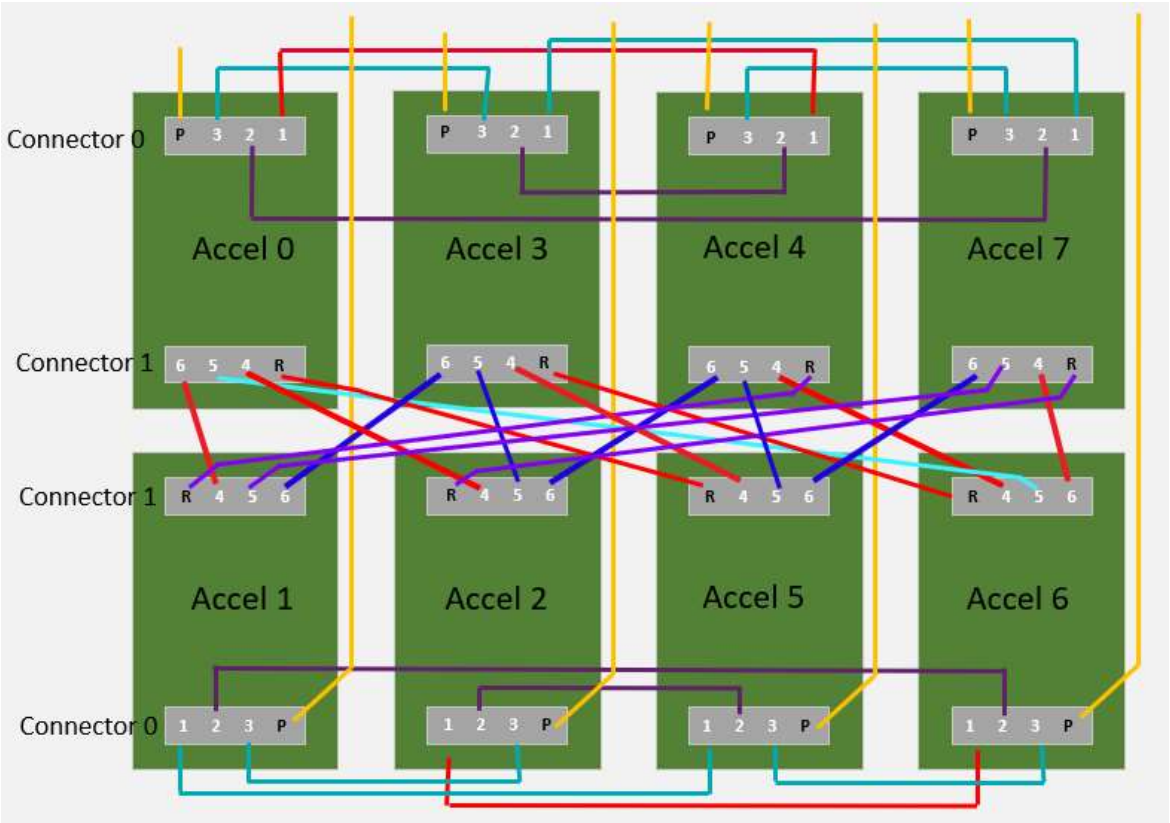


Six inter-module Links may create a 3D Mesh or Torus

OAM Topology Examples

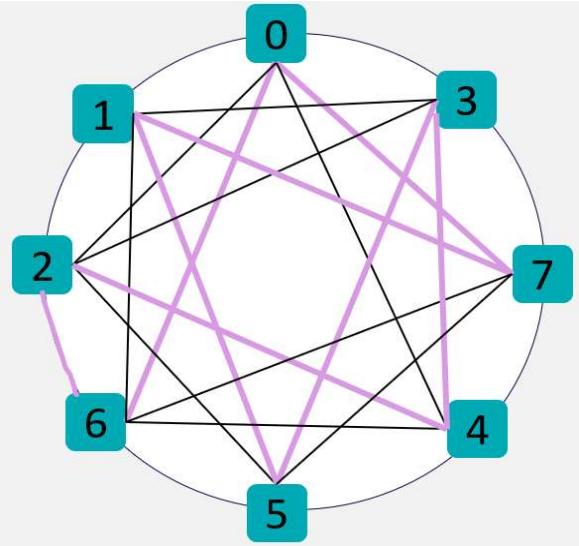
Fully Connected w/ 7 links

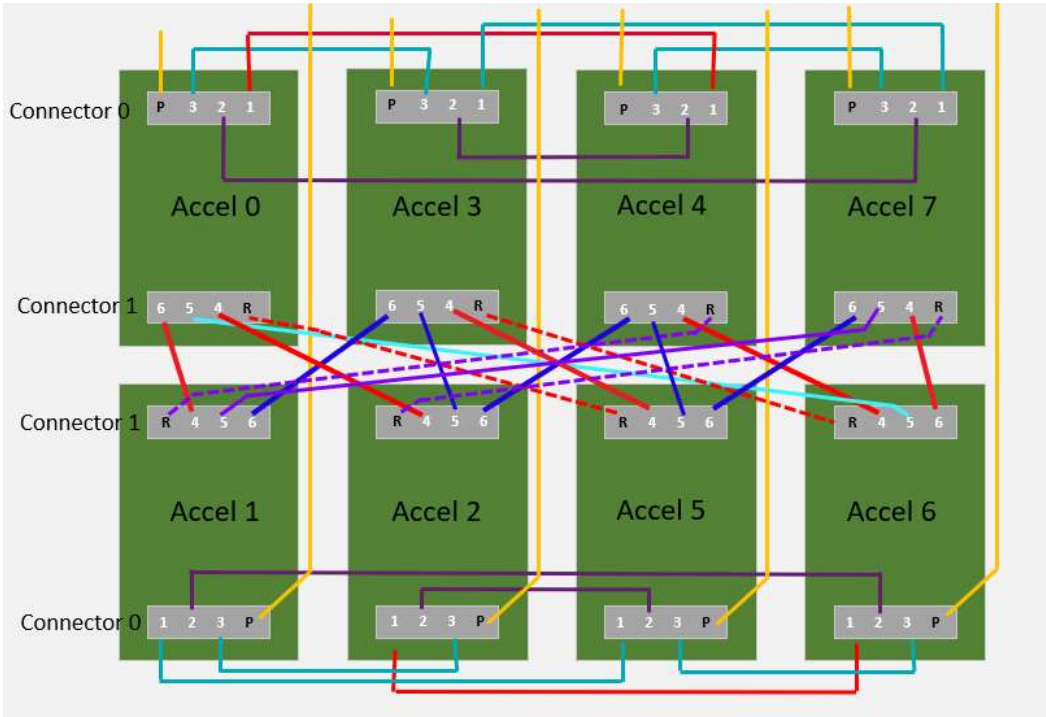




OAM Topology Examples

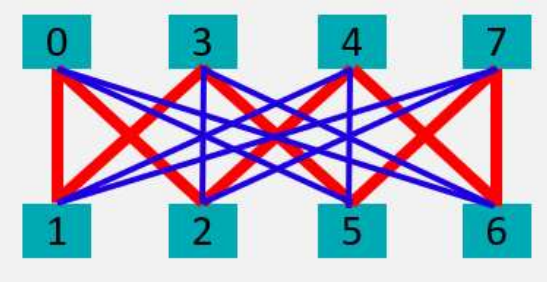
Almost Fully Connected w/ 6 links



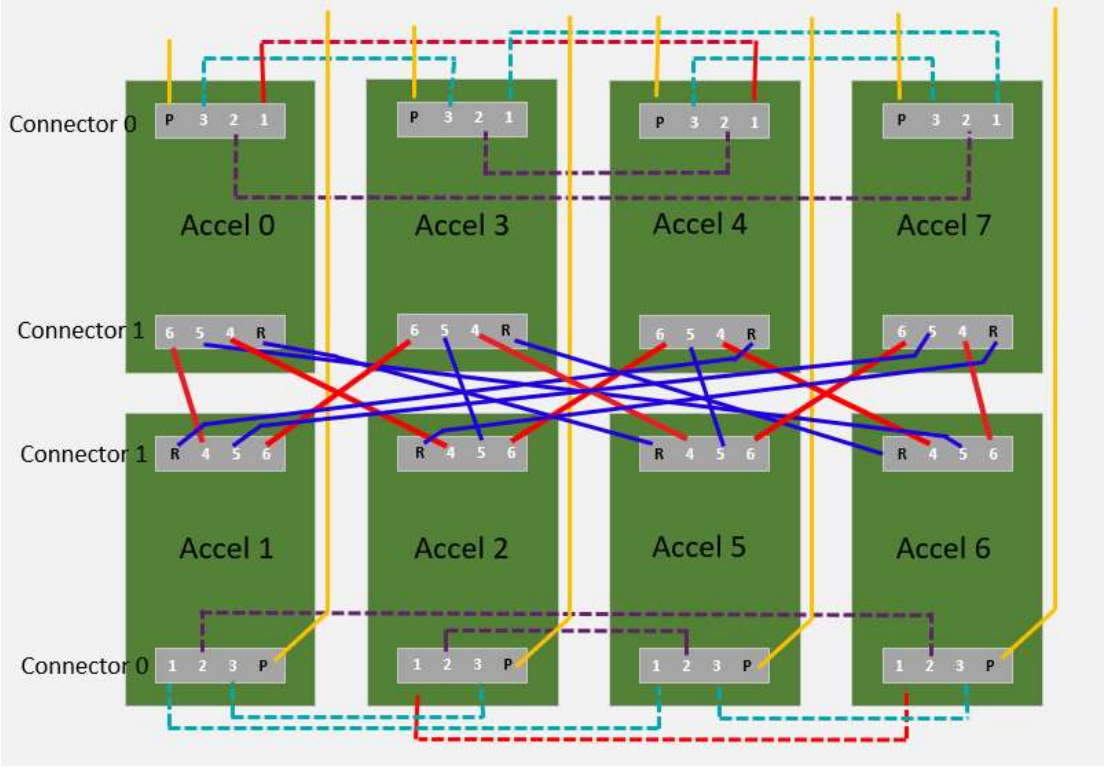


OAM Topology Examples

Rings w/ 4 links

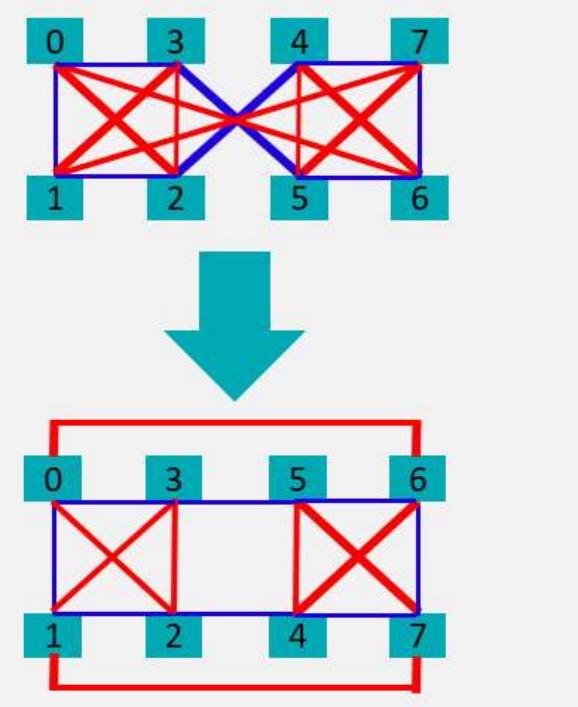


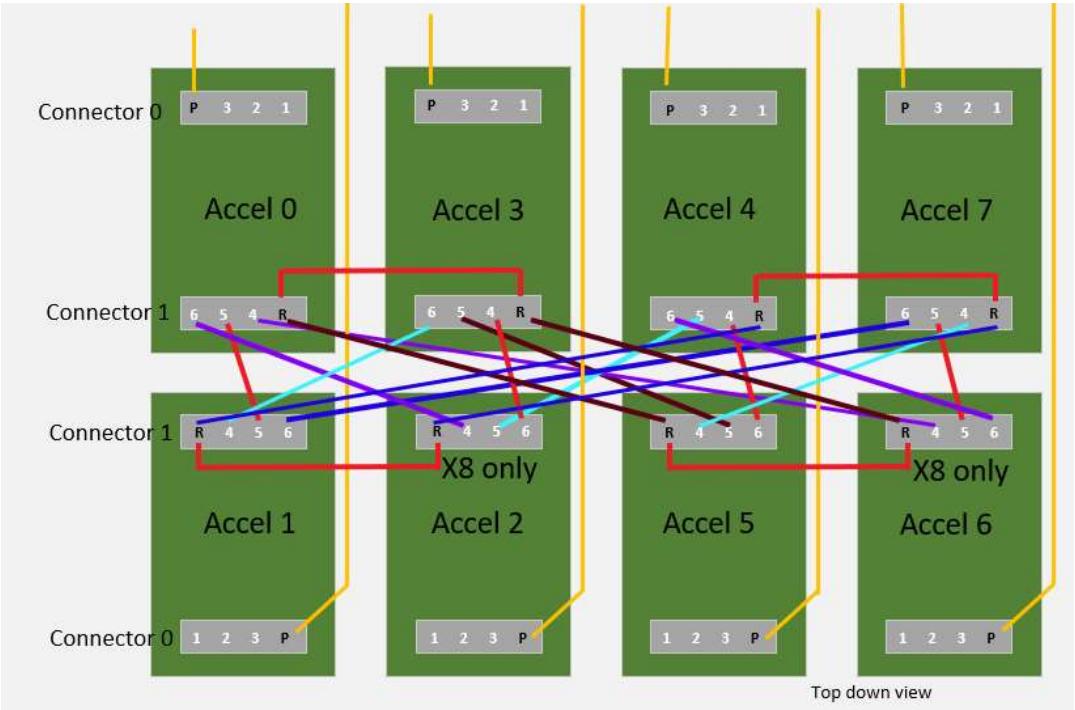
Port 4/5/6/R for AISC which has 4 links on Conn1 Only



OAM Topology Examples

Hybrid Cube Mesh w/ 4 links

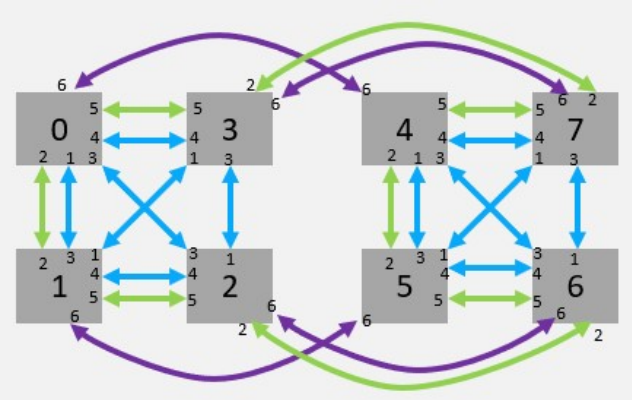


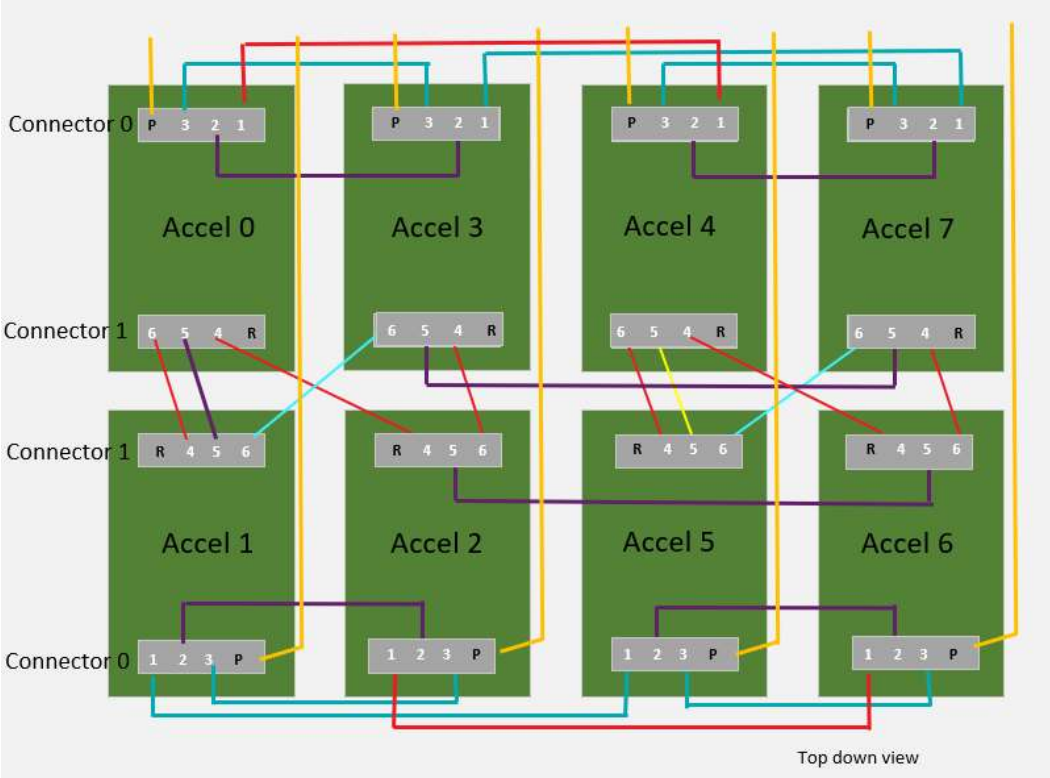


Port 4/5/6/R for AISC which has 4 links on Conn1 Only

OAM Topology Examples

Hybrid Cube Mesh w/ 6 links





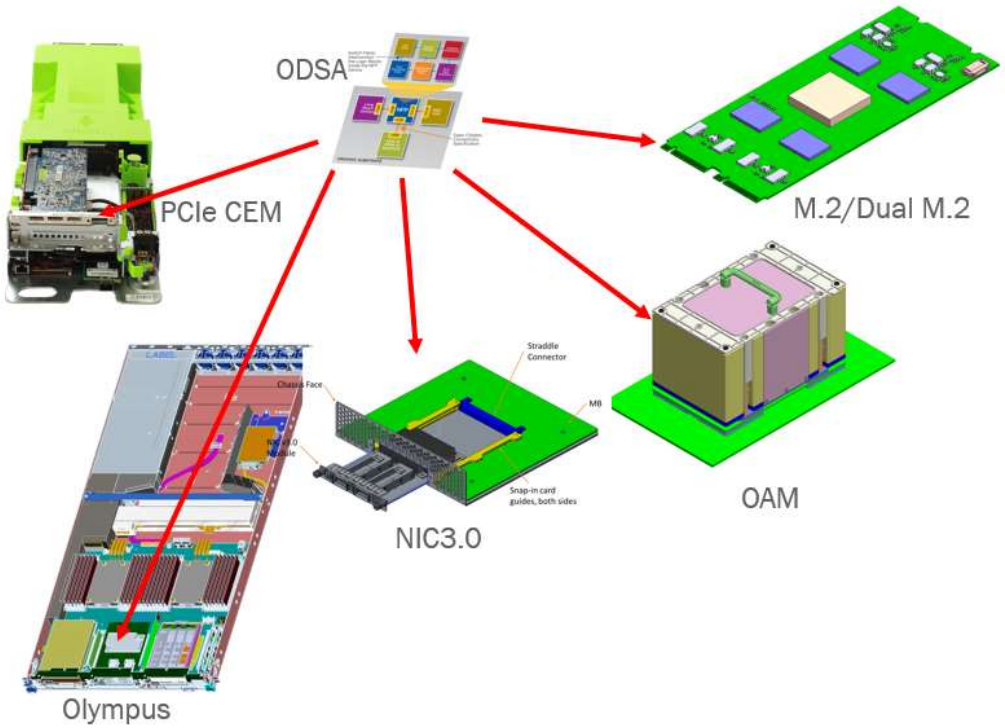
Summary for OAI/OAM

- Rev 0.85 of the OAM spec is available in OCP Wiki
- Join the Project and further develop interoperable Modules for an Open Accelerator Infrastructure(UBB, PSB, SCM, Tray, Chassis…)
- We invite you to join the OAI subgroup for further collaboration:

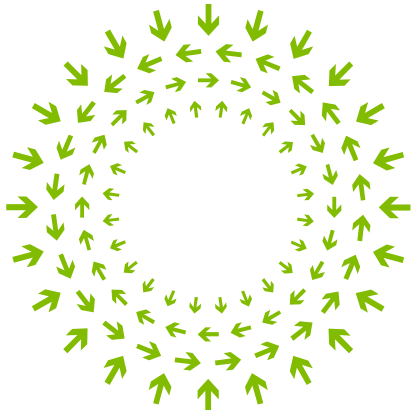
Register for the Mailing List:
https://ocp-all.groups.io/g/OCP-OAI

Wiki under OCP Server Project:
https://www.opencompute.org/wiki/Server/OAI

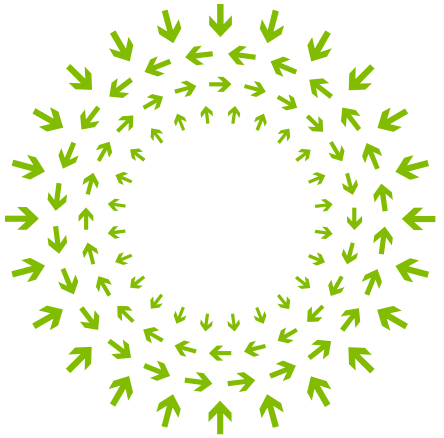
Accelerator Form Factors



- Different Form Factors
  - PCIe CEM
  - OAM
  - M.2/Dual M.2
  - OCP NIC
  - Others
- Different Accelerator Targets
  - Training
  - Inference
  - Video
  - Others



OPEN
Compute Project



OPEN
Compute Project

Consume. Collaborate. Contribute.