Open. Together.

OCP
REGIONAL
SUMMIT

OAI Overview: *An Open Accelerator Infrastructure Project for OCP Accelerator Module (**OAM**)*

Siamak Tavallaei, Principal Architect, Microsoft Azure (OCP Server Project co-Lead)
Whitney Zhao, Hardware Engineer, Facebook (OCP OAI Subproject co-Lead)

Nov 18, 2019 presentation to SC19

Preface

Recognizing the need for a standard module form factor to accommodate accelerators from different suppliers, we developed the OCP Accelerator Module (OAM) spec and contributed it to OCP in March 2019 (Facebook, Microsoft, and Baidu). After presenting the OAM spec as a group effort at 2019 OCP Global Summit, we formed a subgroup in April and encouraged other OCP members to join a team effort to build a modularly interoperable infrastructure around OAM. Many companies have joined.

Open Accelerator Infrastructure (OAI) subgroup operates under OCP Server Project.

Under a joint development agreement (OAI JDA), the scope of work at OAI subgroup for the following 9 schedules is to define the physical and logical aspects such as electrical, mechanical, thermal, management, hardware security, and physical serviceability to produce solutions compatible with existing/traditional operation systems and frameworks to run heterogeneous accelerator applications. The OAI-JDA group will contribute the resulting specification to OCP at multiple revision levels (e.g., 0.4, 0.7, 0.9, and 1.0)

1. Open Accelerator Infrastructure (**OAI**)
2. OCP Accelerator Module (OAI-**OAM**)
3. OAI Universal Baseboard (OAI-**UBB**)
4. OAI Host Interface (OAI-**HIB**)
5. OAI Power Distribution (OAI-**PDB**)
6. OAI Expansion Beyond UBB (OAI-**Expansion**)
7. OAI Security, Control, and Management (OAI-**SCM**)
8. OAI-**Tray**
9. OAI-**Chassis** (This chapter will address **air-cooled** and **liquid-cooled** aspects as well.)

The research and development in
Artificial Intelligence (AI),
Machine Learning (ML), Deep Learning (DL), and
High-Performance Computing (HPC)
are driving rapid evolution in
new types of hardware accelerators

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

Diverse Module and System Form Factors

Different Implementations

Targeting Similar Requirements!

We need an

Open Accelerator Infrastructure

for these

Complex and Expensive Systems

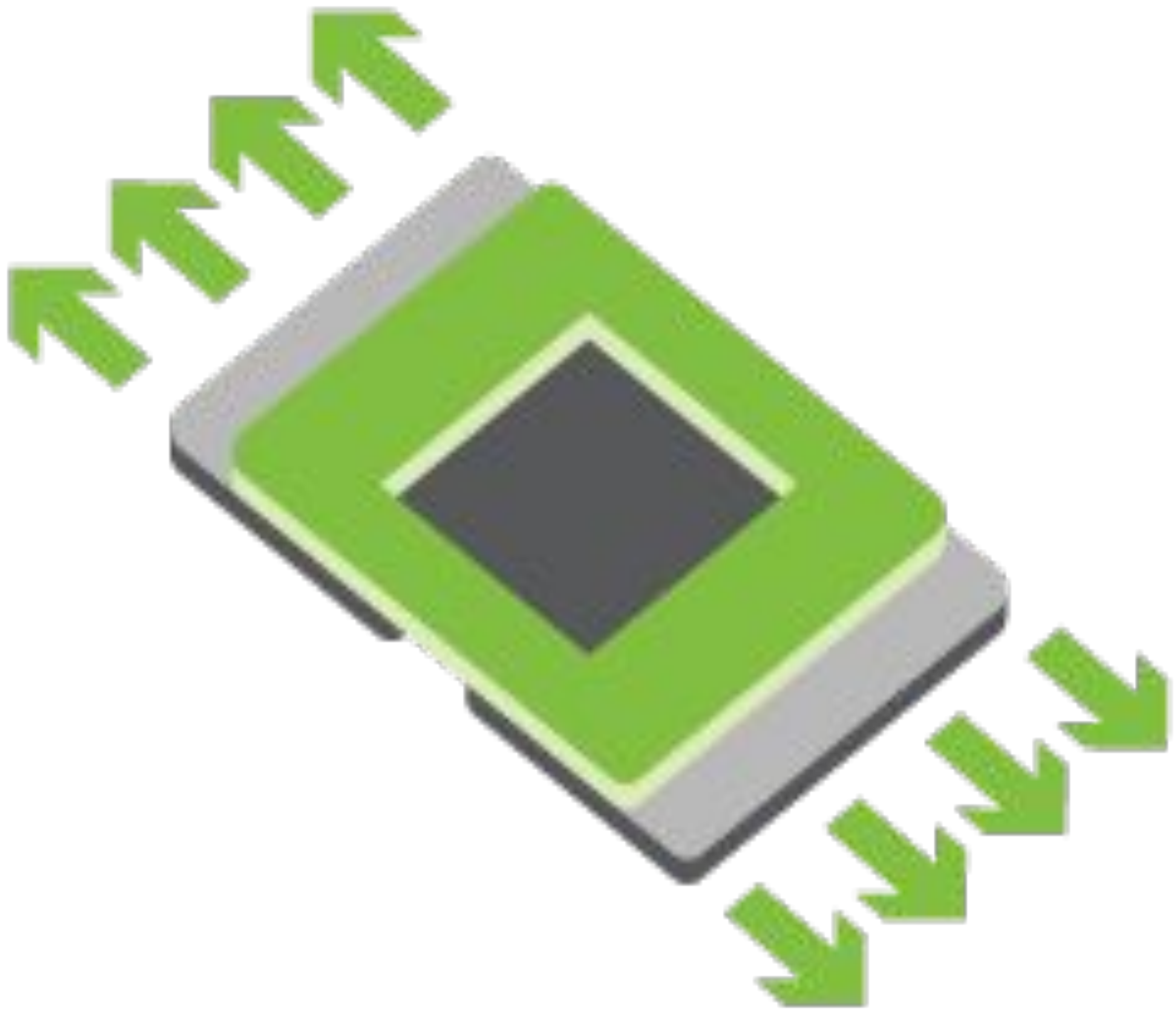
Increase Interoperability

Accelerate Innovation

Via

Modular Building Block Architecture (MBA)

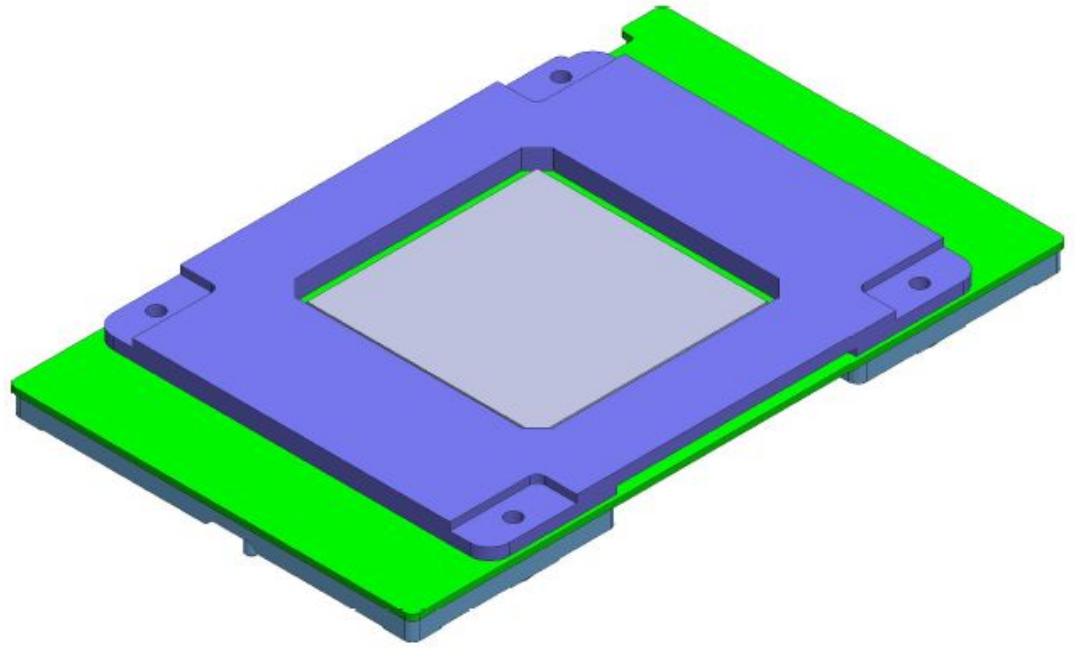
We started with *OCP Accelerator Module*(***OAM***)

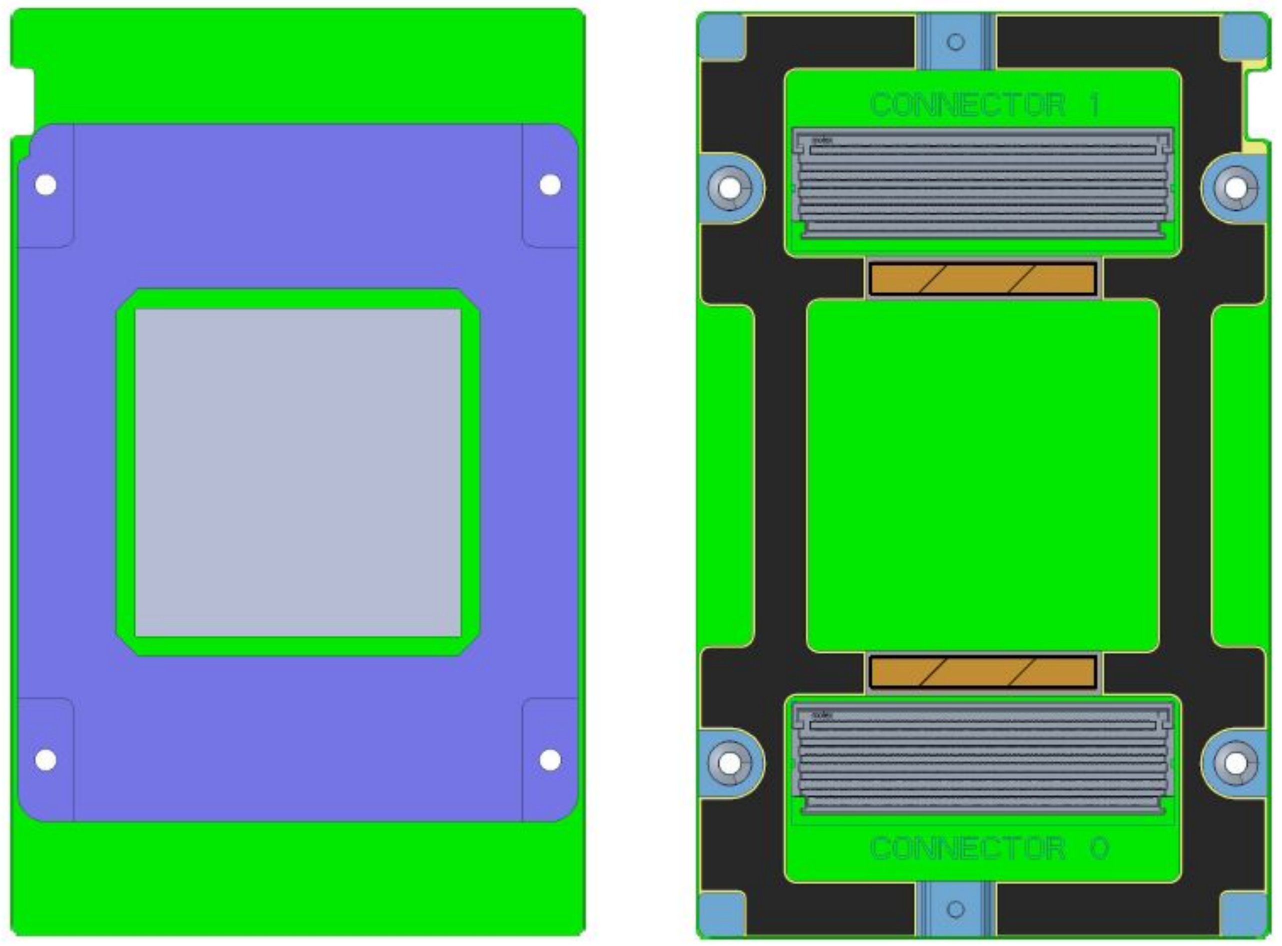


A common form factor mezzanine module for the upcoming accelerators

OCP Accelerator Module Spec

- 102mm x 165mm Module Size
- With two high-speed Mirror Mezz connectors (MPN: 2093111115)
- 12V and 48V input DC Power
- Up to 350w (12V) and up to 700w (48V) TDP
  - Up to 440W (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
- System management and debug interfaces





OCP OAI Subgroup

- Formed in March 2019 under OCP Server Project
- To build the infrastructure for fast adapting, upcoming products which meet OAM spec
- Scope: to define the physical and logical aspects such as electrical, mechanical, thermal, management, hardware security, and physical serviceability to produce solutions compatible with existing/traditional operation systems and frameworks

We are adding *Infrastructure Support*

Open & Modular

in everyway!

Hierarchical Base Specification

Well-defined boundaries
Fostering Innovation while maintaining Interoperability

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

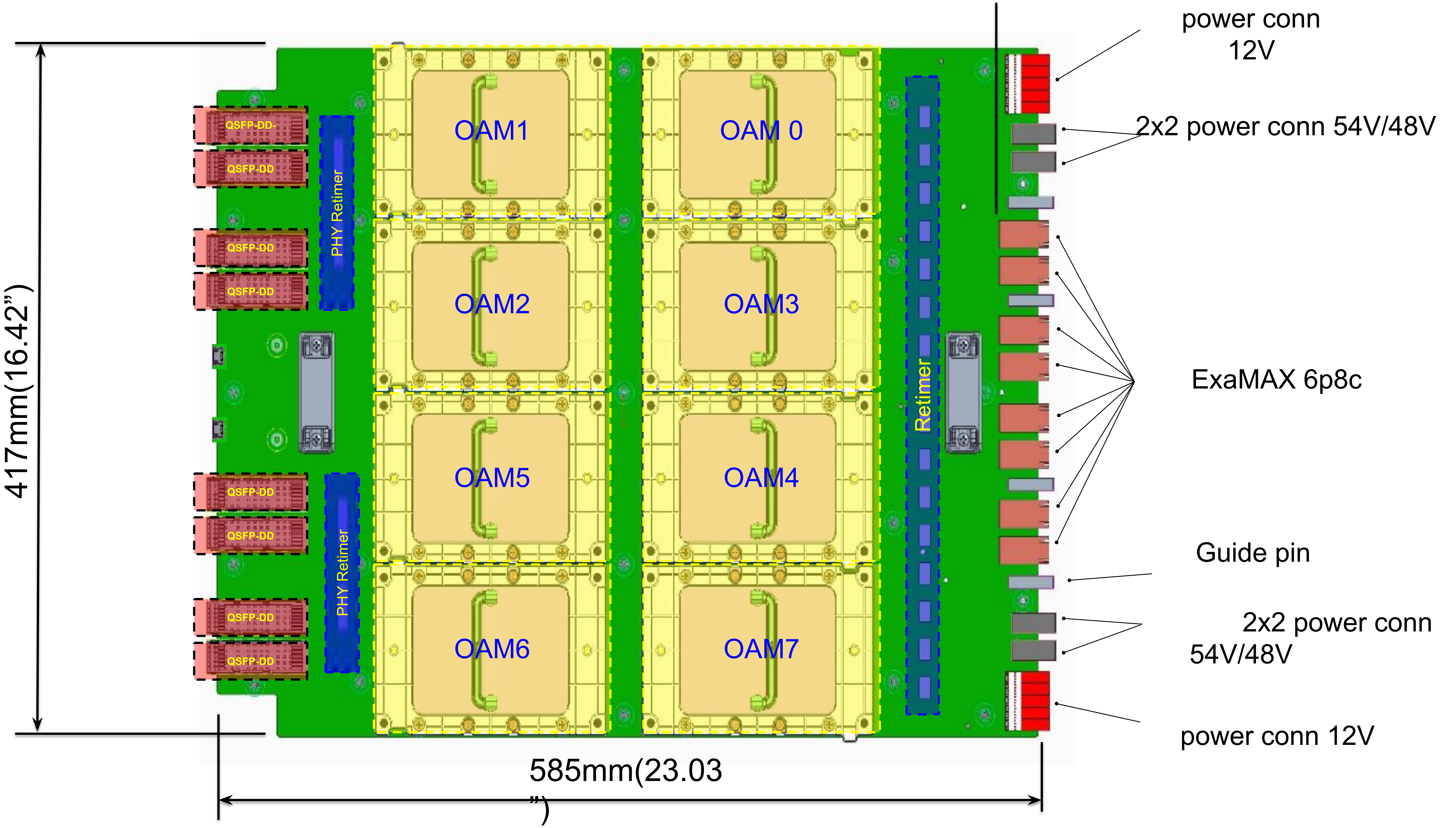
- OAM
- UBB (Interconnect Topology)
- HIB
- PDB
- Tray, Chassis
- OAI-SCM
- Expansion

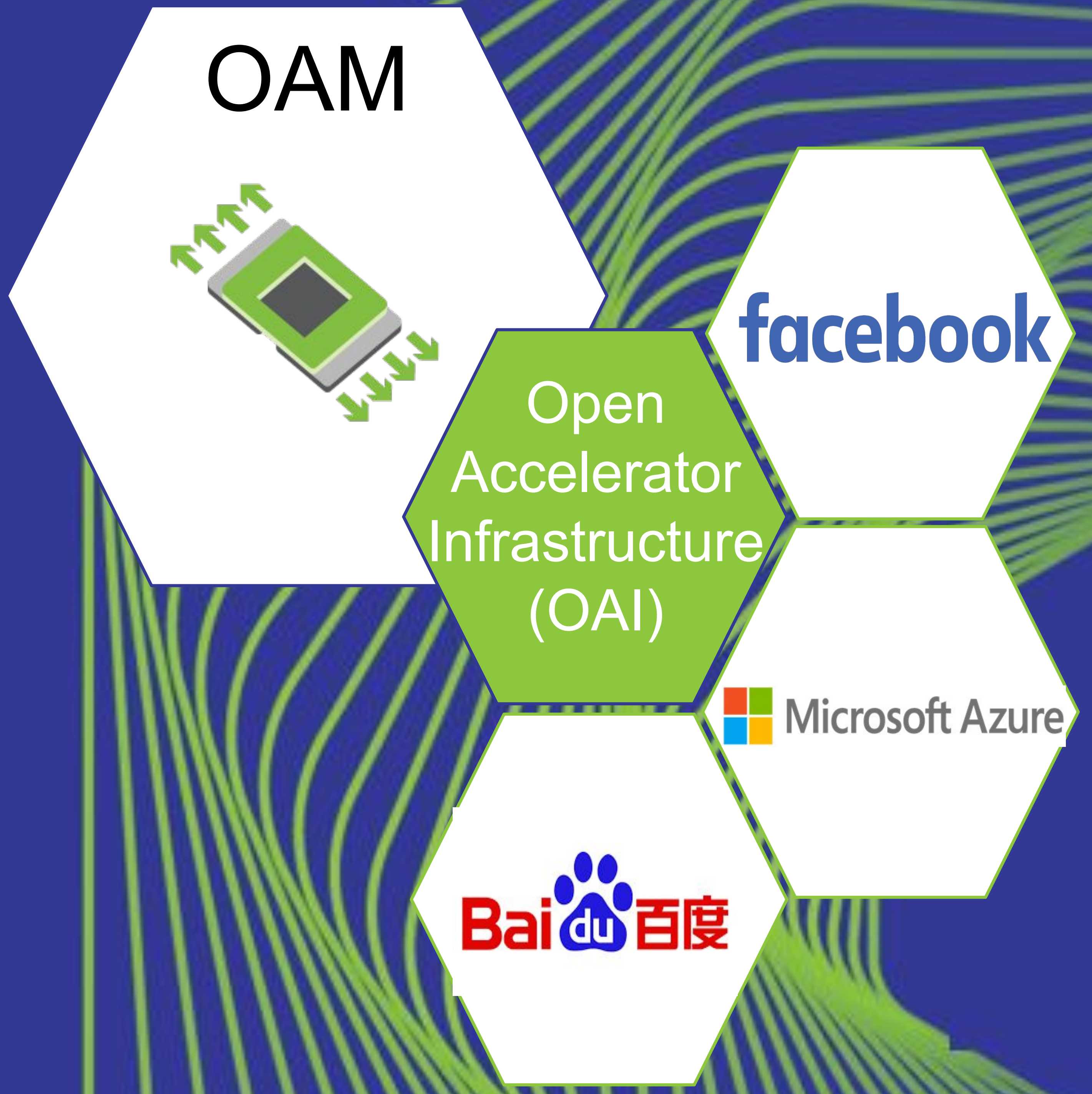
Designs and Products may be compliant to any or all specifications

Well-defined boundaries (OAI)

- Different manufacturers may offer **OAM**s with standard or propriety inter-OAM protocols
- **OAI-UBB** provides Host interface and native **Expansion** capabilities for eight OAMs
- **OAI-Tray** provides mechanical support to adapt various UBBs in 19” and 21” Chassis
- Modular power distribution allows 12V, 48V, and AC distribution to the Chassis
- **OAI-Chassis** supports Air- and Liquid-cooling in a modular way
- Rack-level Security and Baseboard Management (**OAI-SCM**)
- Each OAI Module is stateless; any FW or programmable code/logic is under RoT control
- Each OAI Module includes a FRU-ID to include vital product data (VPD)

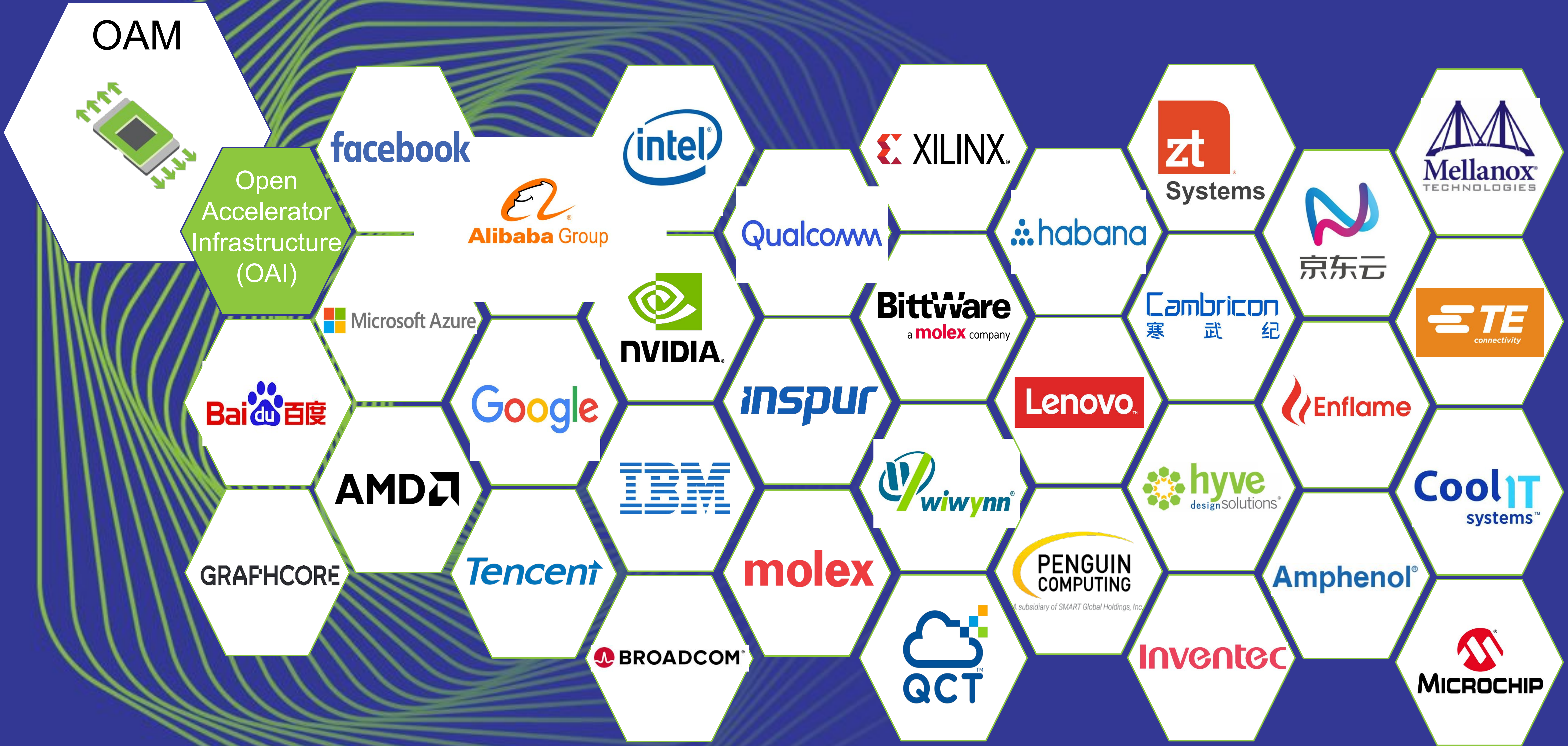
OAI-UBB: Universal Baseboard





OAM

Open Accelerator Infrastructure (OAI)



OAM

Open Accelerator Infrastructure (OAI)

facebook

Alibaba Group

Microsoft Azure

Bai du 百度

AMD

GRAPHCORE

Google

Tencent

intel

NVIDIA

IBM

BROADCOM

Qualcomm

inspur

molex

XILINX

BittWare a molex company

wiwynn

QCT

habana

Lenovo

PENGUIN COMPUTING
A subsidiary of SMART Global Holdings, Inc.

zt Systems

Cambricon 寒武纪

hyve design solutions

Inventec

京东云

Enflame

Amphenol

Mellanox TECHNOLOGIES

TE connectivity

CoolIT systems

MICROCHIP

OAI JDA Group

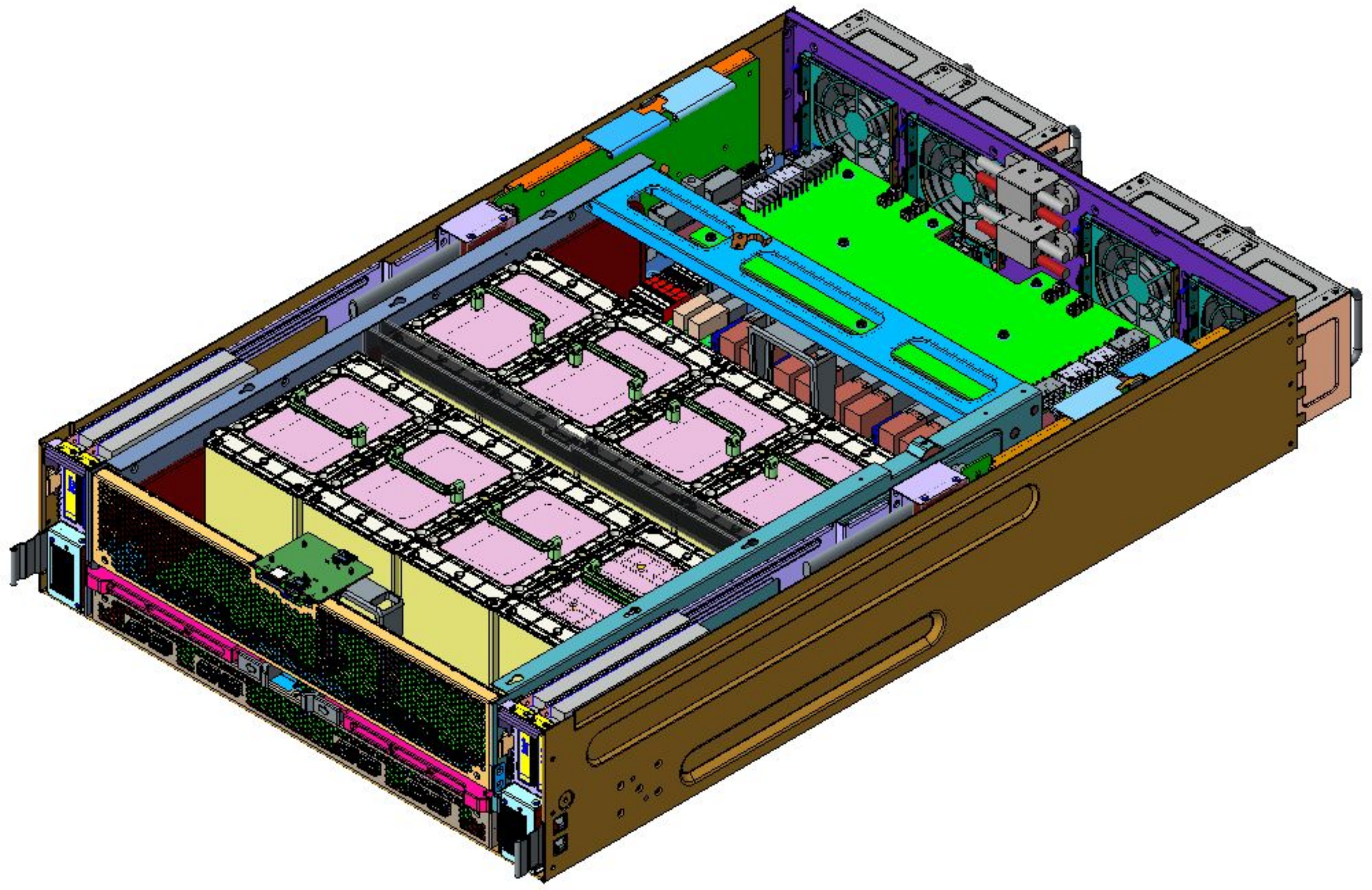
facebook | Microsoft Azure | Bai du 百度 | Tencent | 京东云 | Alibaba Group

intel | AMD | XILINX | Enflame | NVIDIA | molex

inspur | hyve design solutions | Inventec | zt Systems | Amphenol

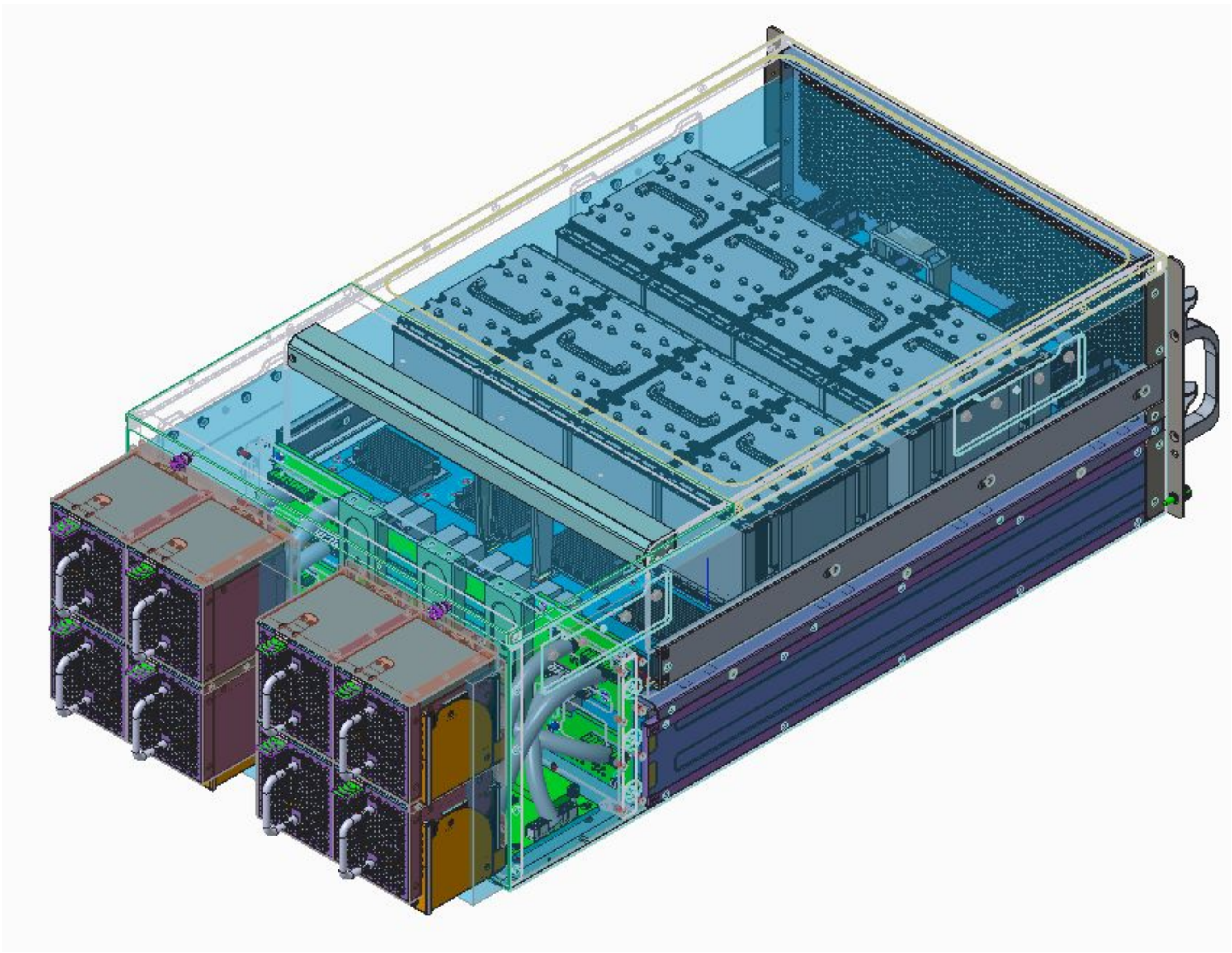
OAM Reference Designs

Inspur 21” Co-Planar system



- 21 inch 3OU, 34.6” (800mm) depth
- 8*OAMs
- UBB: Combined FC+ 6 port HCM Topology
- 4*PCIE Gen4 x16 Link to connect Hosts
- 4*PCIE Gen4 x16 Slots support 100G Infiniband or Ethernet for expansion

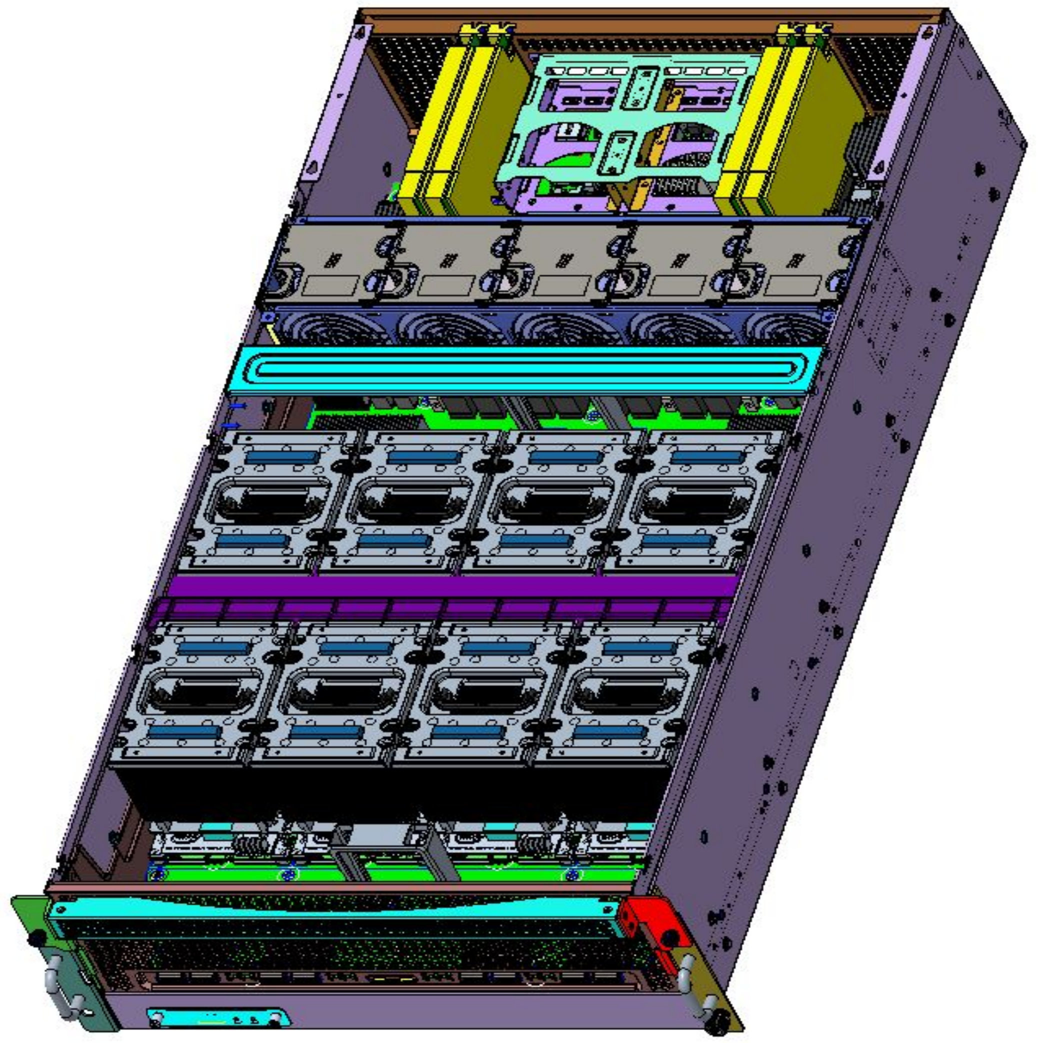
Hyve Design Solutions 19” Stacked System



- 19 inch 6RU, 30 inch (762mm) depth
- 8*OAMs
- UBB: Combined FC+ 6 port HCM Topology
- 4*PCIE Gen3x16 slots for host uplink
- 12*PCIE Gen3 x16 slots for flexible IO expansion

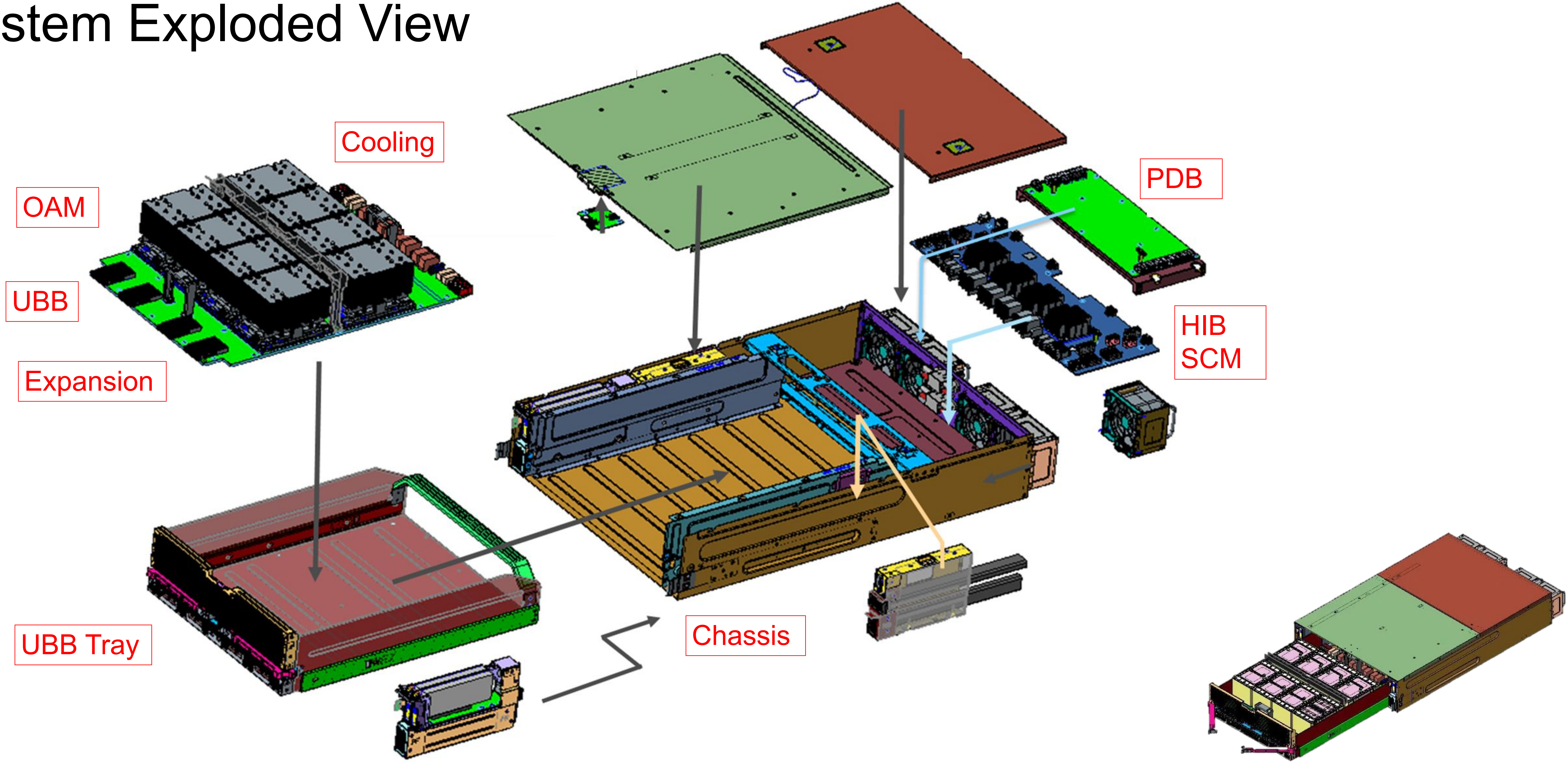
(PCIE interface will be revised to Gen4 in next release.)

ZT Systems 19” Co-Planar System



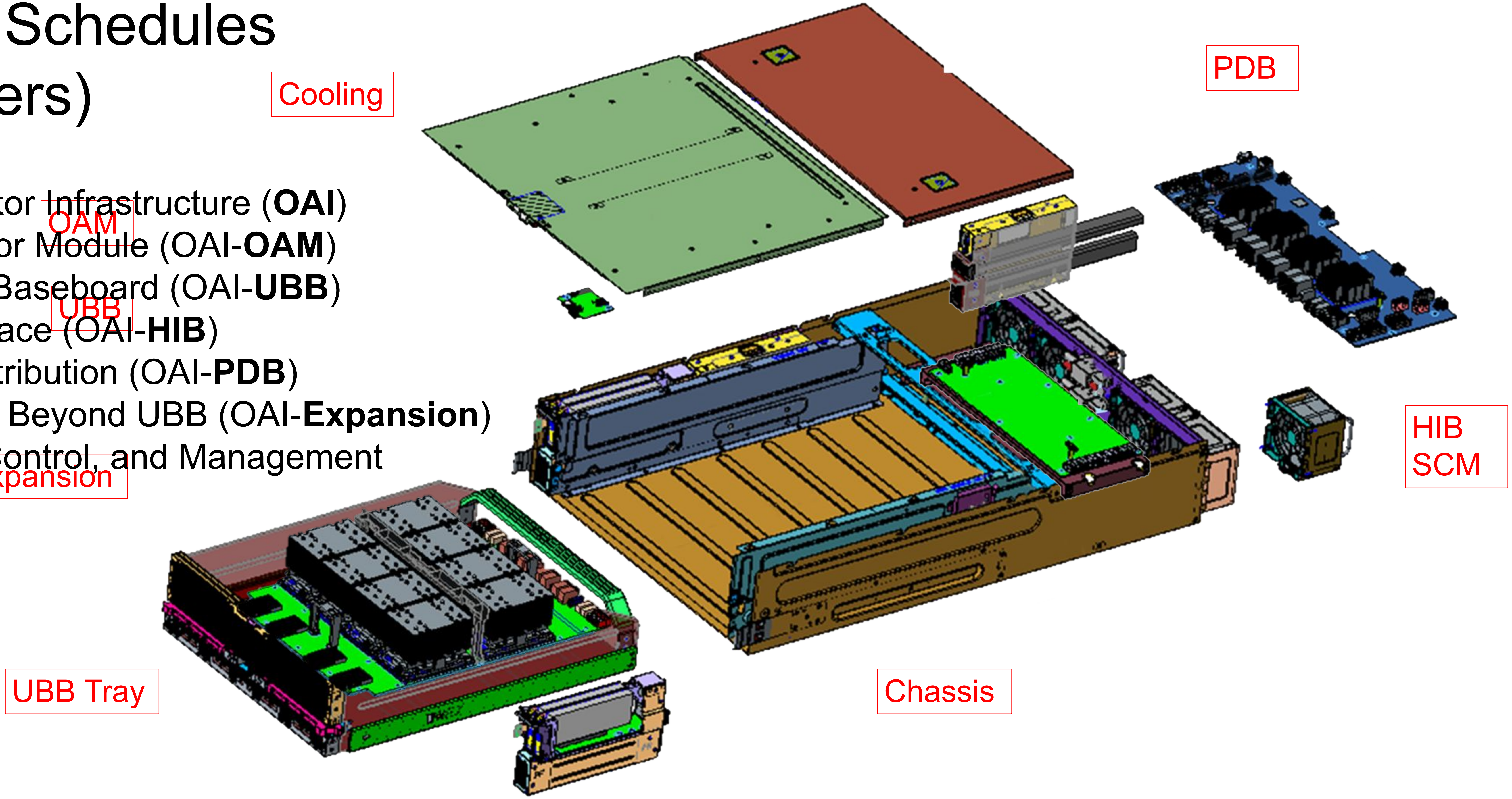
- 19 inch 4RU, 34.6” (880mm) depth
- 8*OAMs
- UBB: 8-port HCM topology
- 2*PCIE Gen4 x16 Uplinks for Multi-Host
- 4*PCIE Gen4 x16 Slots
- 4*2.5” NVME hot plug drives in front

System Exploded View



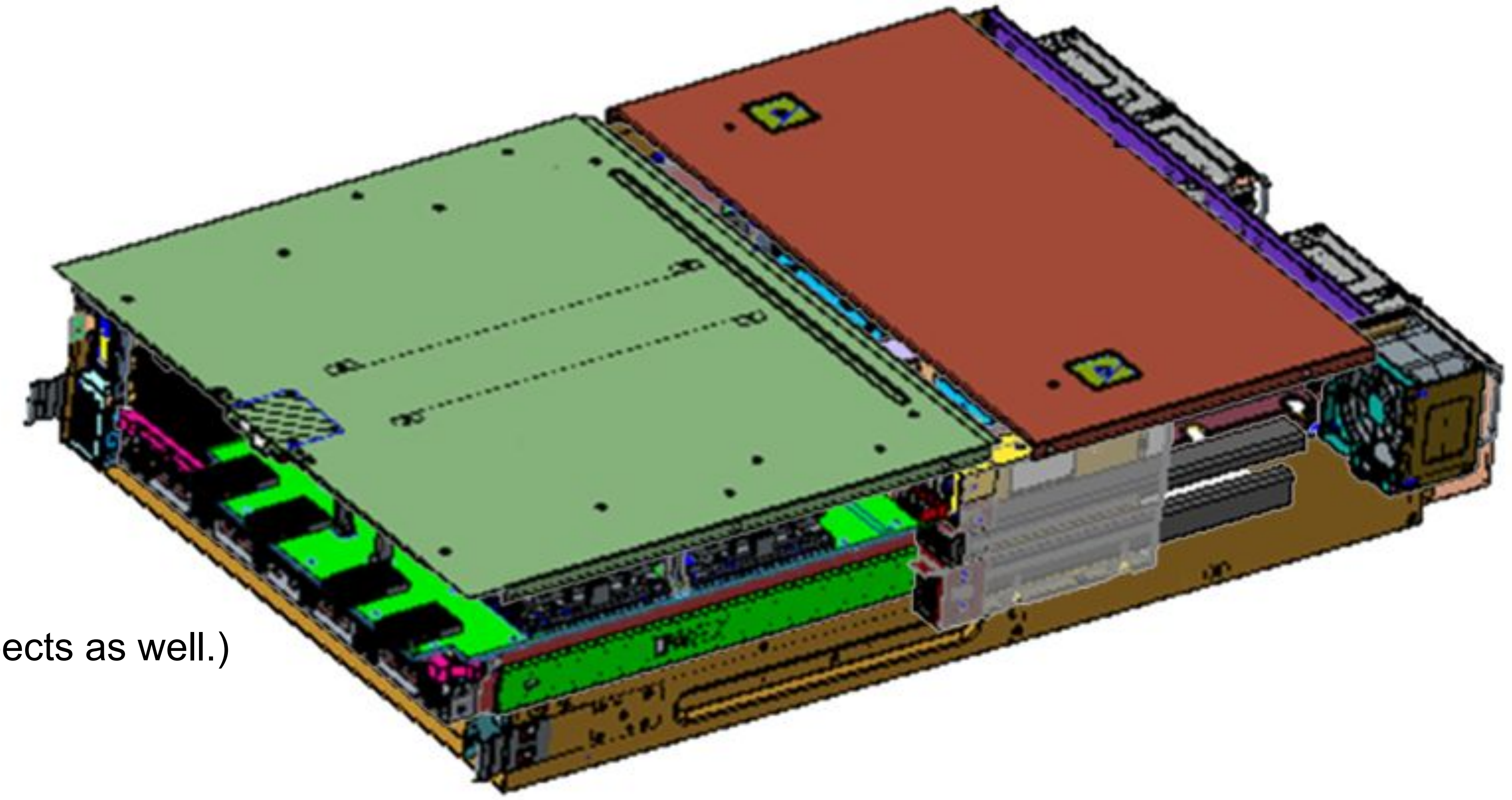
OAI Project Schedules (spec chapters)

- Open Accelerator Infrastructure (**OAI**)
- OCP Accelerator Module (OAI-**OAM**)
- OAI Universal Baseboard (OAI-**UBB**)
- OAI Host Interface (OAI-**HIB**)
- OAI Power Distribution (OAI-**PDB**)
- OAI Expansion Beyond UBB (OAI-**Expansion**)
- OAI Security, Control, and Management (OAI-**SCM**)
- OAI-**Tray**
- OAI-**Chassis**

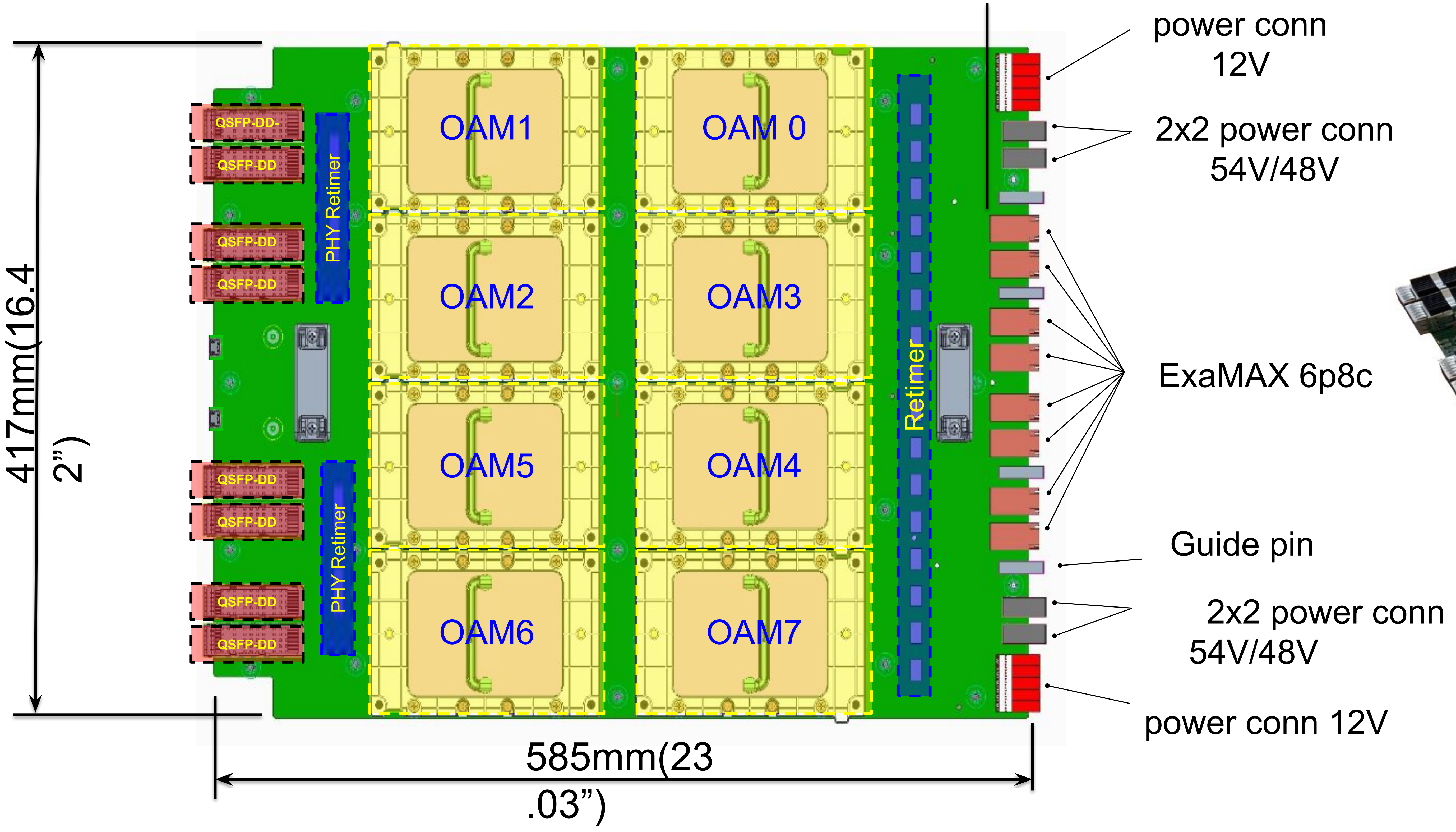


OAI Project Schedules (spec chapters)

- Open Accelerator Infrastructure (**OAI**)
- OCP Accelerator Module (OAI-**OAM**)
- OAI Universal Baseboard (OAI-**UBB**)
- OAI Host Interface (OAI-**HIB**)
- OAI Power Distribution (OAI-**PDB**)
- OAI Expansion Beyond UBB (OAI-**Expansion**)
- OAI Security, Control, and Management (OAI-**SCM**)
- OAI-**Tray**
- OAI-**Chassis** (This chapter will address **air-cooled** and **liquid-cooled** aspects as well.)



OAI-UBB: Universal Baseboard



power conn
12V

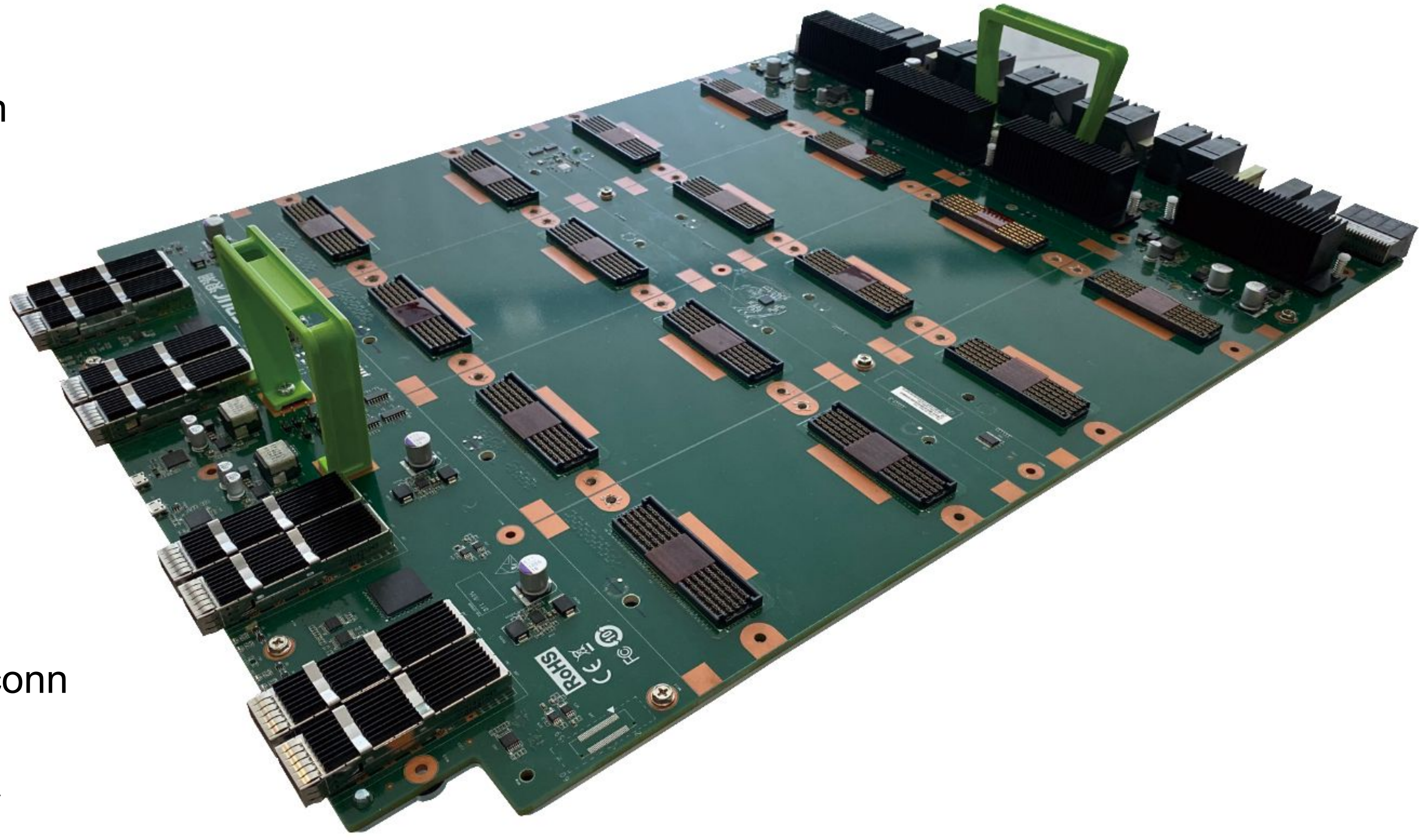
2x2 power conn
54V/48V

ExaMAX 6p8c

Guide pin

2x2 power conn
54V/48V

power conn 12V



What we define in UBB

- 8* OAMs in UBB
- Interconnect topology
- Host Interface with retimers
- Scale-out with Phy retimers
- 12V or 54V/48V power delivery
- 19’’ and 21’’ rack compatible
- Debug/management interface

Spec Overview

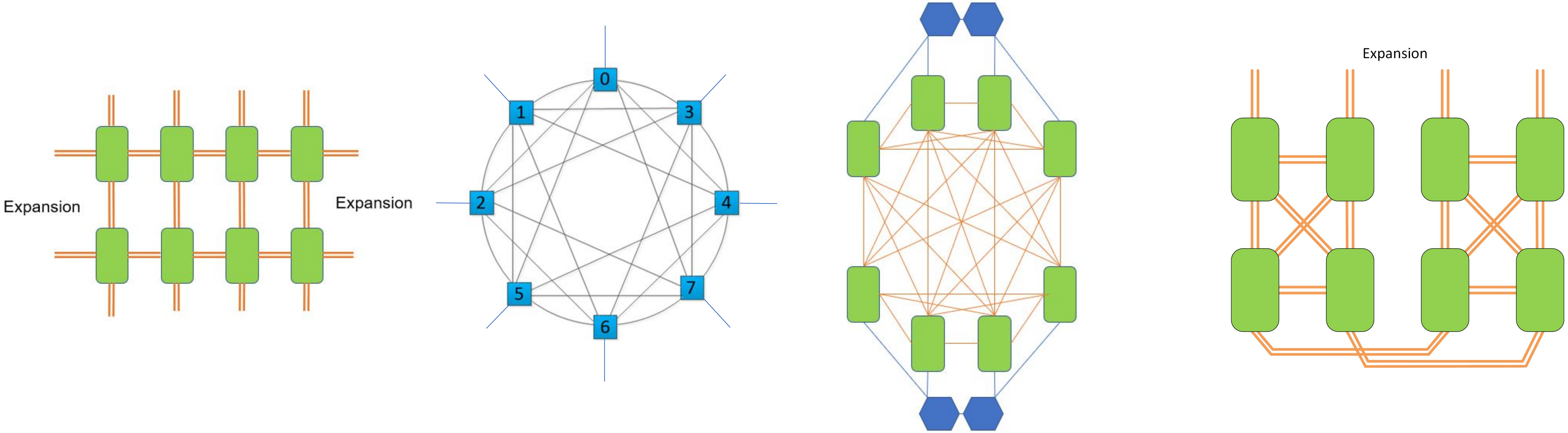
Item	Feature
UBB Dimension	585mm(L) x 417mm(W) x 3.26mm
OAM	8x OAM 12V up to 300W TDP 54V/48V up to 500W TDP
Host Interface	8 X16 Serdes, with retimers
Interconnect SerDes Speed	Up to 28Gbps NRZ or 56Gbps PAM4
Interconnect Topology	Various *UBB reference designs support FC(Fully Connected) or HCM(Hybrid Cube Mesh)
Connectors to HIB	4x 54V/48V AirMax 2x P12V PwrMax 8x ExaMax (high speed and side bands)
Scale out	8x QSFP-DD with retimers (up to 28Gbps NRZ or 56Gbps PAM4 Serdes interface)
Debug/Management/Security	JTAG/I2C/UART to microUSB2.0/Vendor proprietary

* UBB Spec v0.4 is in process to contribute to OCP

Interconnect Topology

Interconnect Topology

- UBB can support various topologies

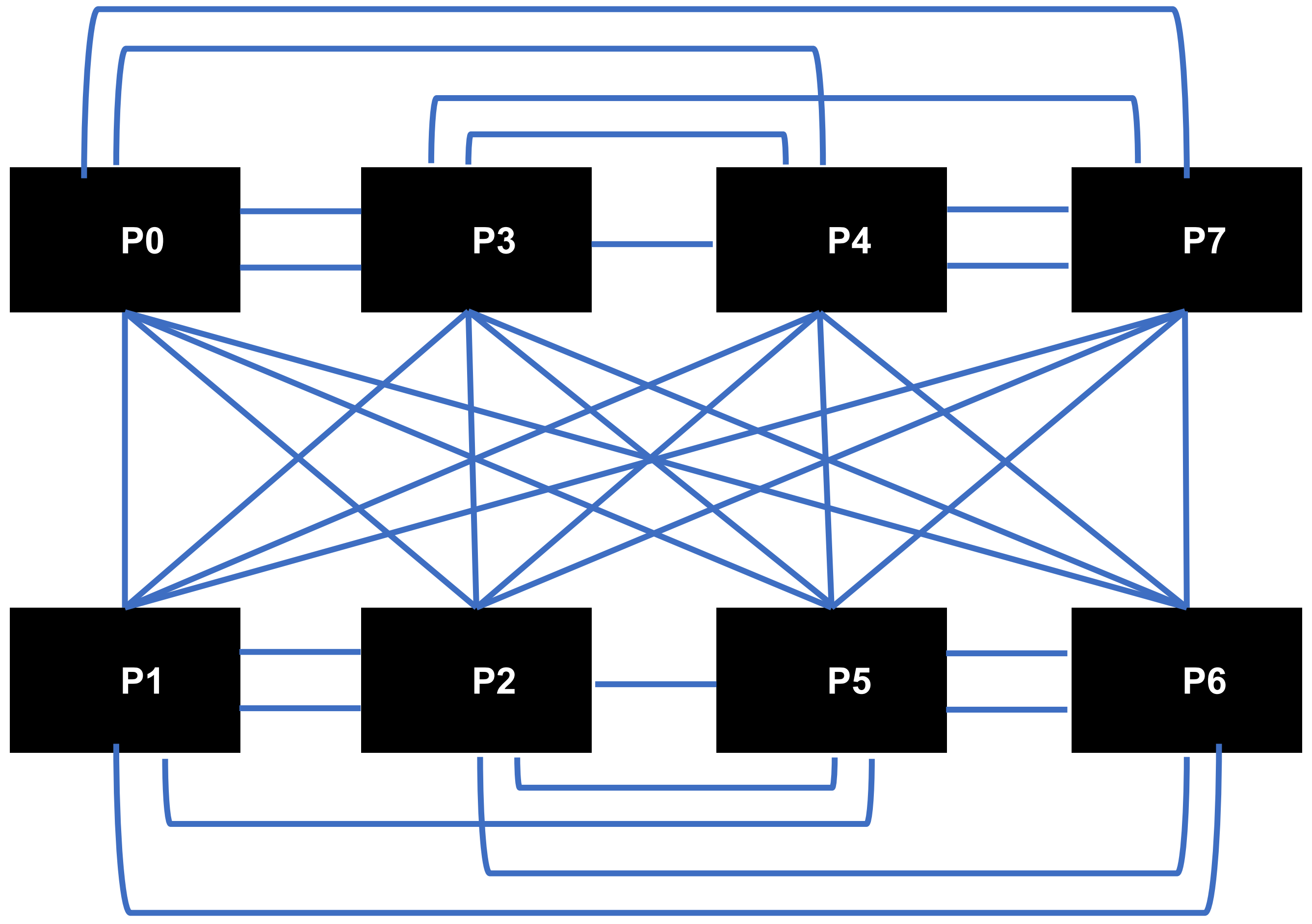


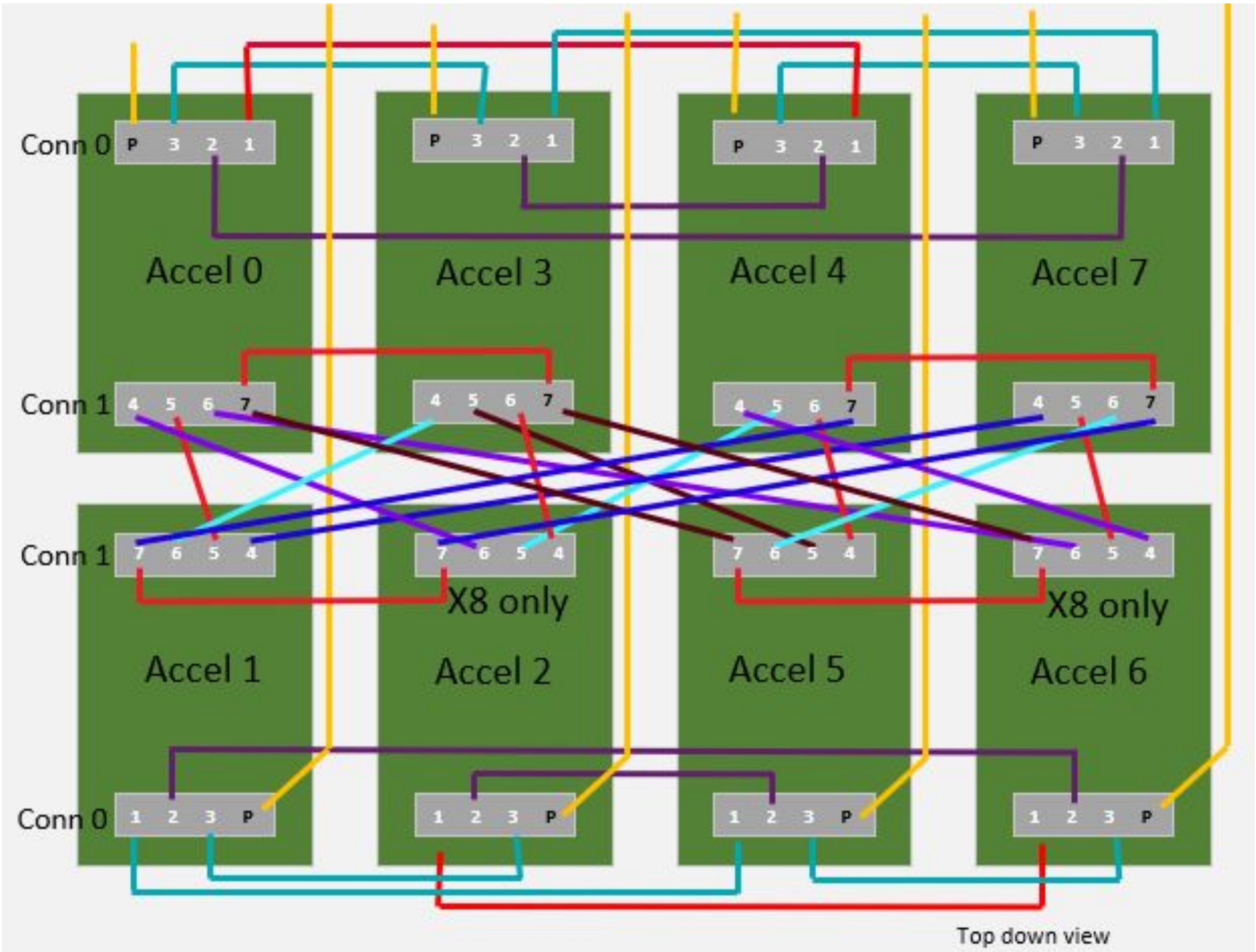
UBB Reference Boards

- OAI Group system suppliers built 3 different UBB ref boards with two topologies
  - Inspur
  - Hyve Design Solutions
  - ZT Systems/Inventec

Combined Topology (FC & 6-port HCM)

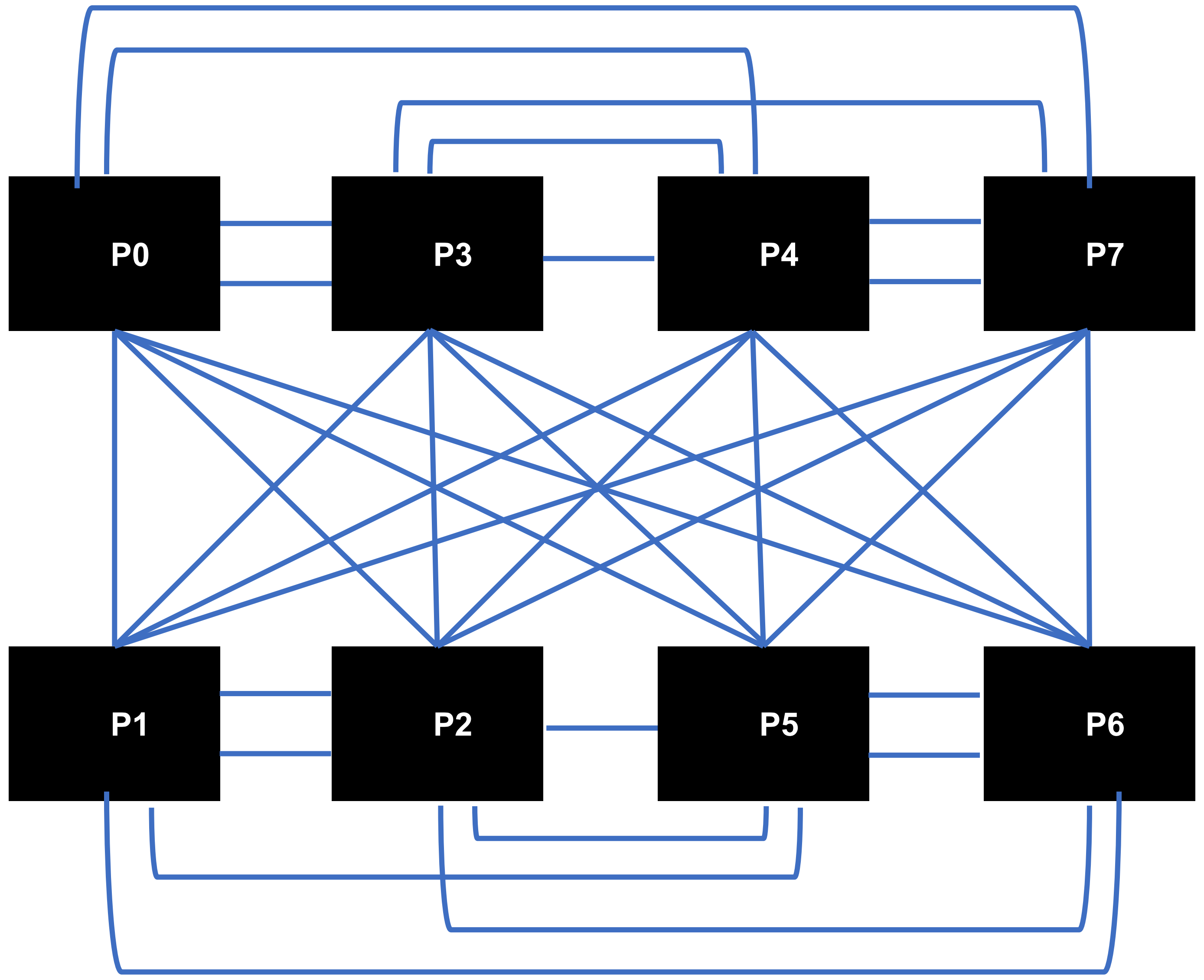
Fully-connected & Hybrid-Cube Mesh



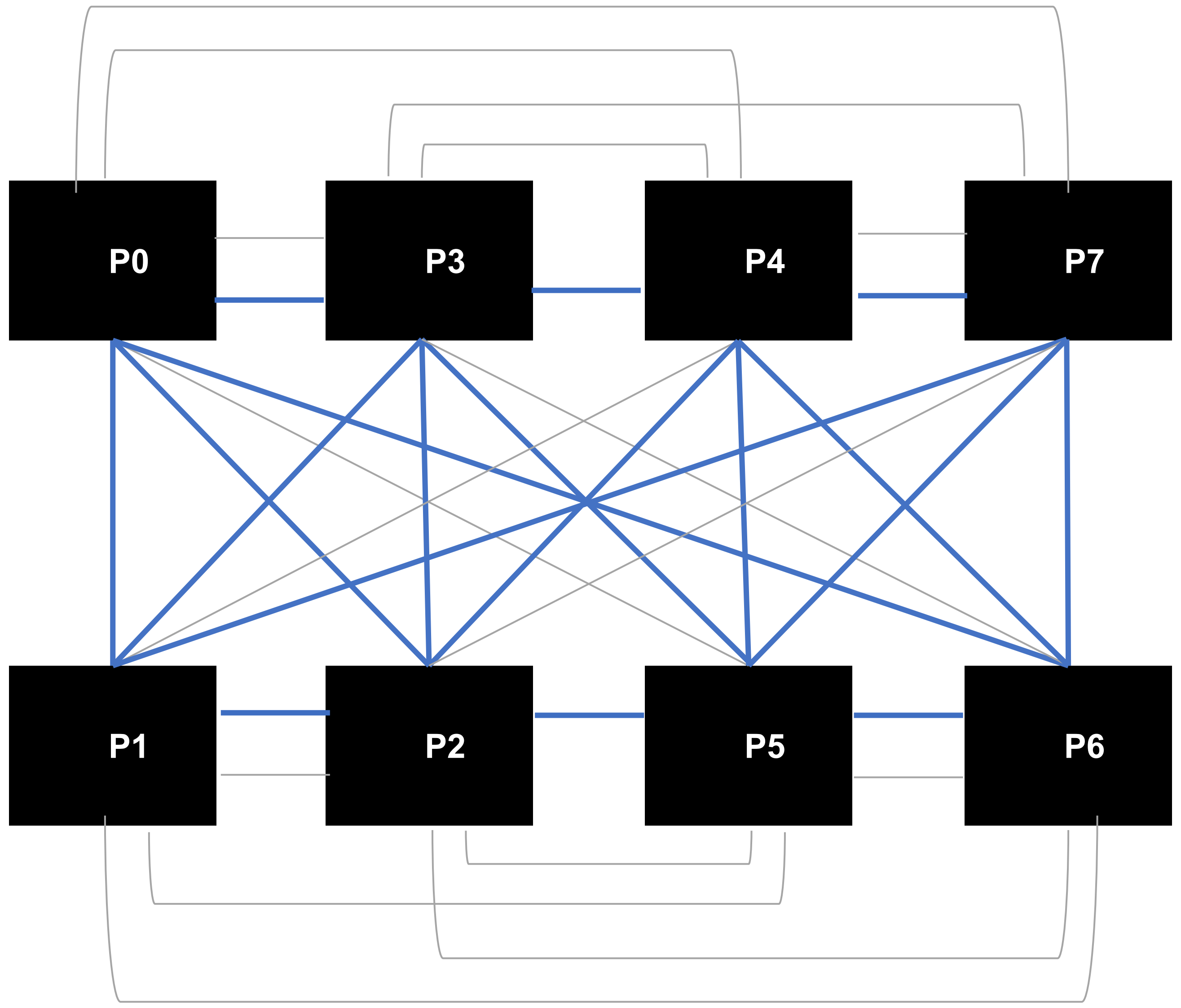


By Inspur and Hyve Design Solutions

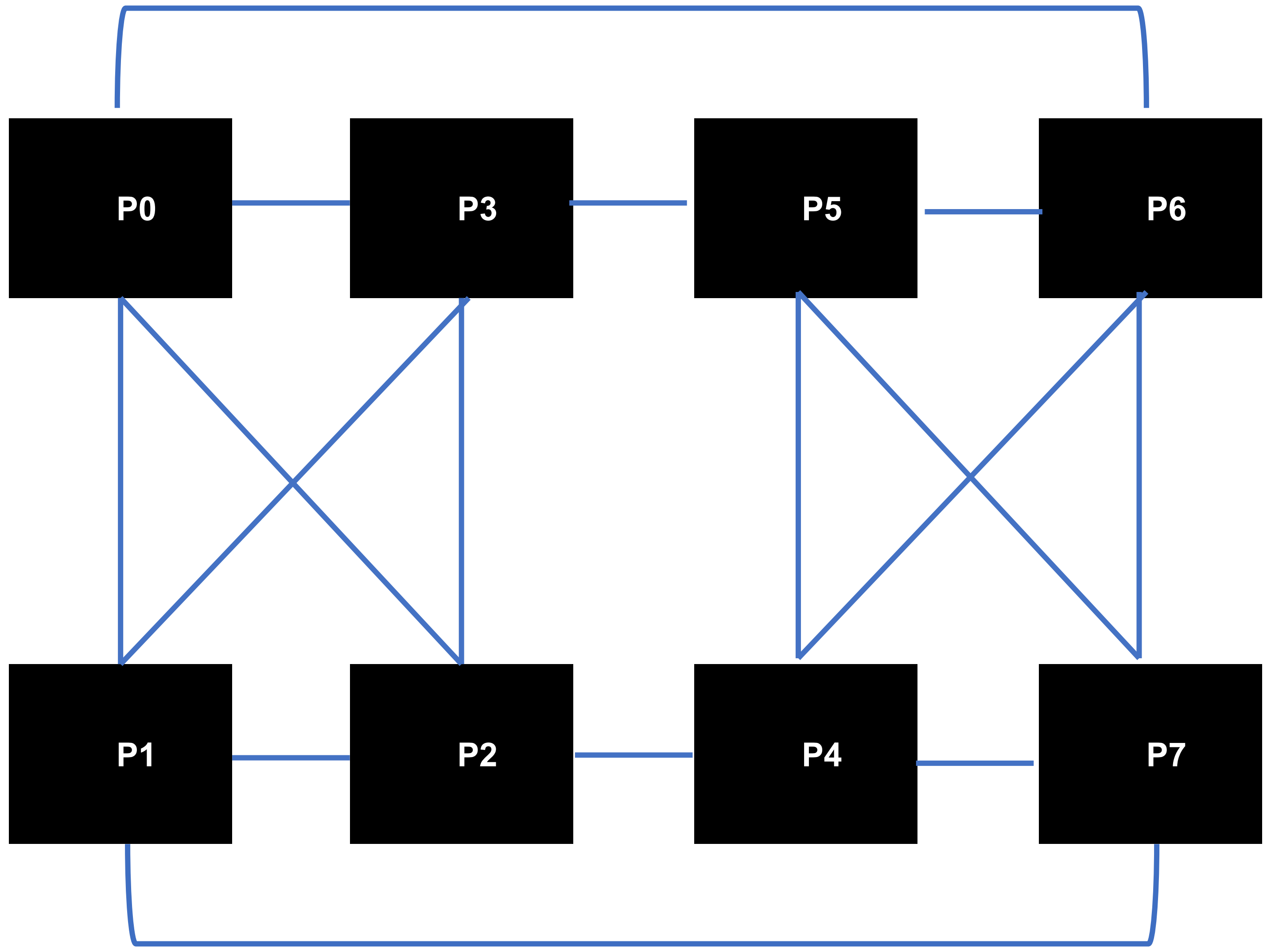
How does HCM Embedded in this topology?



Superset topology

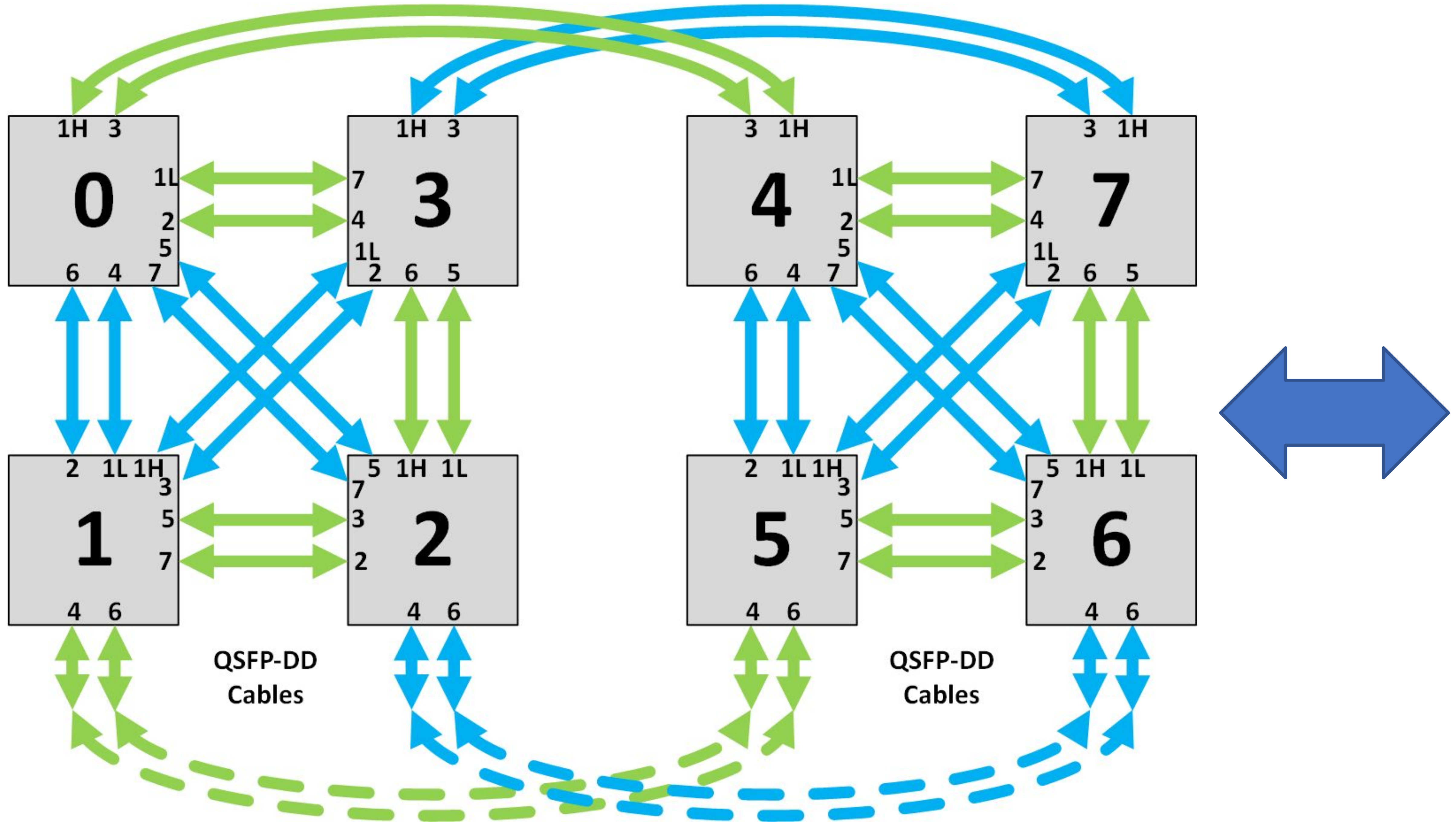


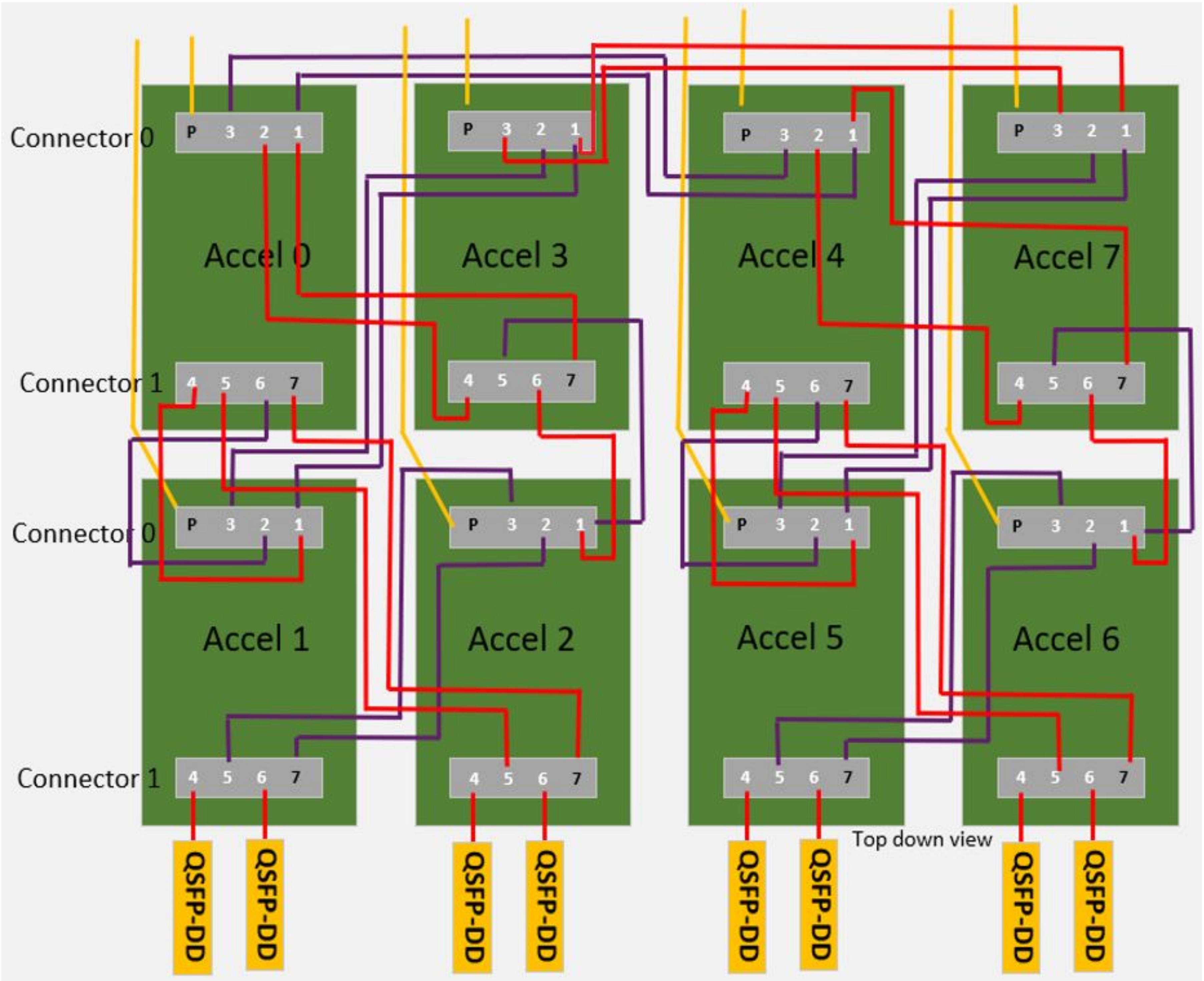
Hide unused links



Rotate 4,7,5,6 by 180°
□ HCM

8-port HCM (Hybrid-Cube Mesh)





By ZT systems/Inventec

Electrical Spec.

- UBB to HIB interfaces and detail pin lists
- Connectors and pin map
- Debug interfaces architecture
    - JTAG
    - UART
- I^2C Topology
- Power delivery block diagram
- Insertion loss and PCB stackup

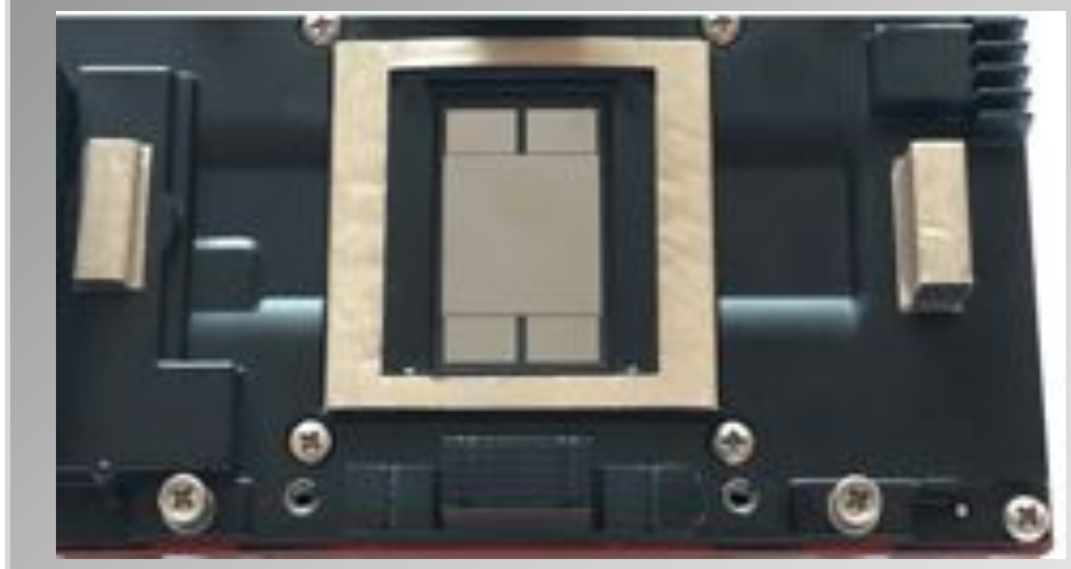
OAI-SCM: Security, System Management, and Debugging

- RoT attestation
- Sensor reporting
- Error monitoring/Reporting
- Firmware Update
- Power-capping
- FRU Information
- IO Calibration
- JTAG/I^{2}C/UART interfaces for debugging

Current OAM Status

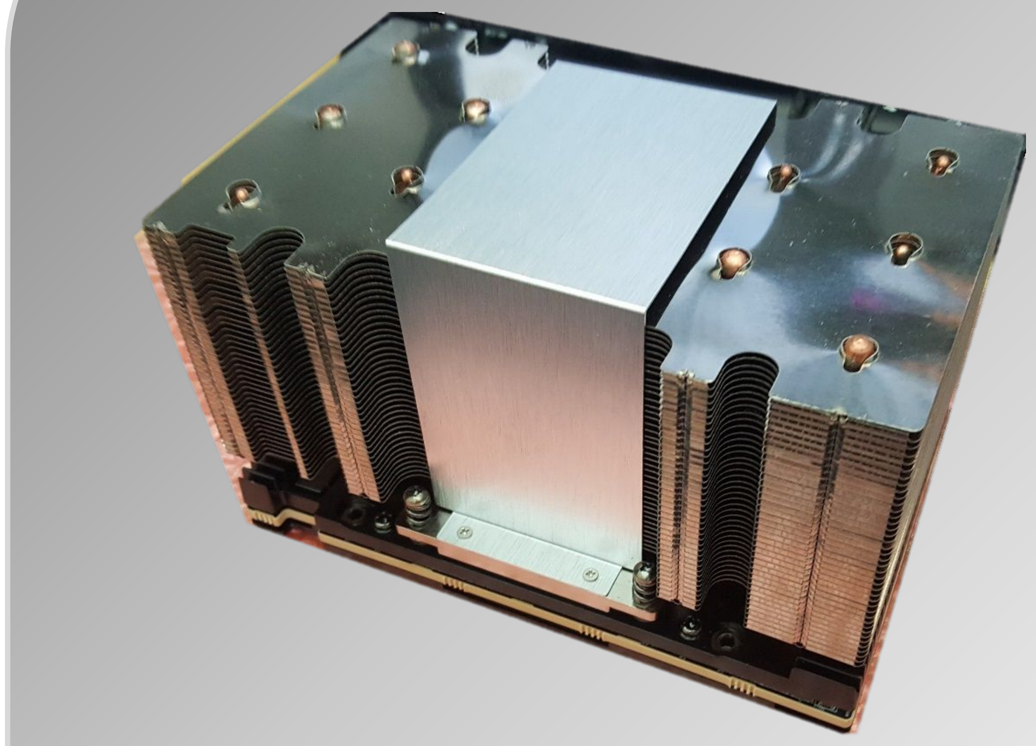
OAM Current Status

- Spec v0.85 released on March 14, 2019
- Spec v1.0 released on July 31, 2019
- We are working with accelerator suppliers to enable their OAM-based solutions



Nervana™ NNP-T OAM
Intel

Gaudi OAM
Habana

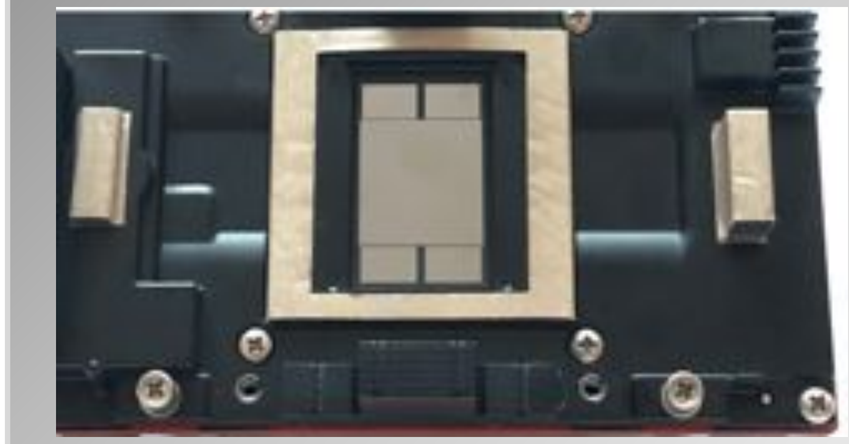


PoC OAM
AMD

V100 PoC OAM
Nvidia

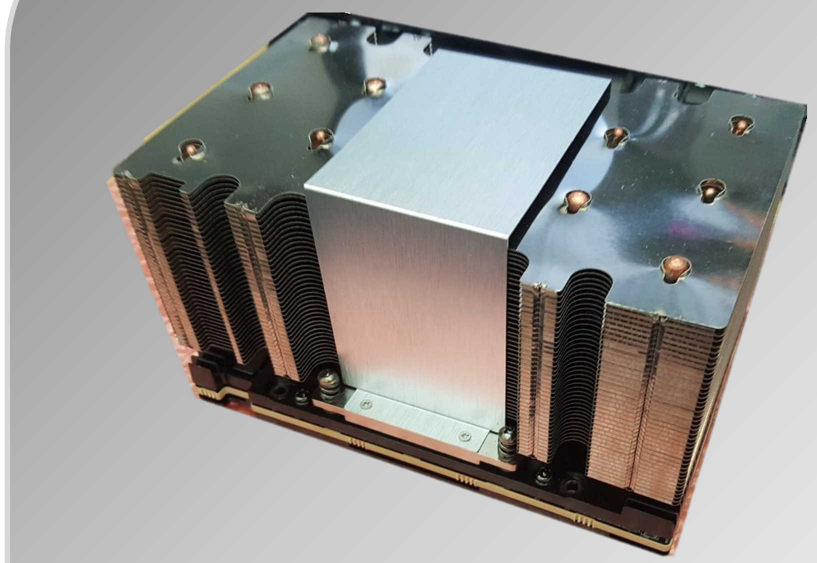
OAM Current Status

- Spec v0.85 released on March 14, 2019
- Spec v1.0 released on July 31, 2019
- We are working with accelerator suppliers to enable their OAM-based solutions



Nervana™ NNP-T OAM
Intel

Gaudi OAM
Habana

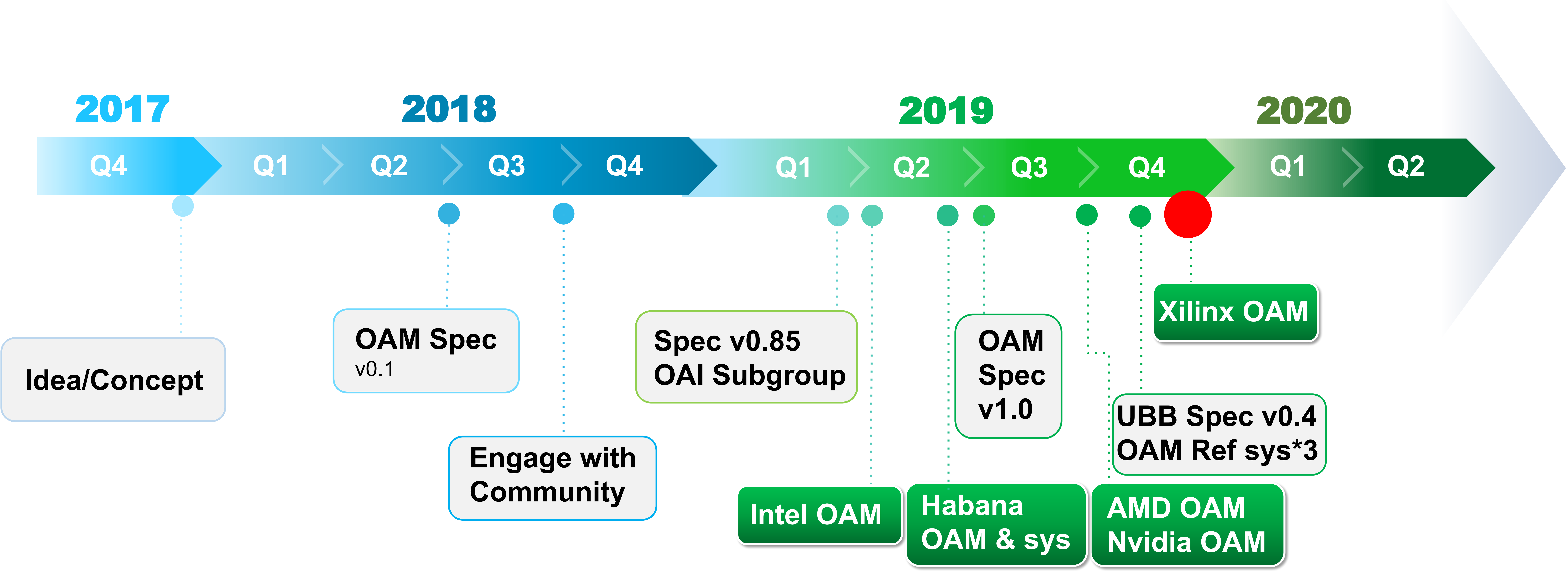


PoC OAM
AMD

V100 PoC OAM
Nvidia

OAM
Xilinx

OAI/OAM Timeline



Idea/Concept

OAM Spec v0.1

Engage with Community

Spec v0.85 OAI Subgroup

Intel OAM

Habana OAM & sys

OAM Spec v1.0

AMD OAM Nvidia OAM

UBB Spec v0.4 OAM Ref sys*3

Xilinx OAM

Next Steps

OAI Subproject Next Steps by 2020 OCP Global Summit

- OAI-OAM Spec v1.1
- OAI-UBB Spec v1.0

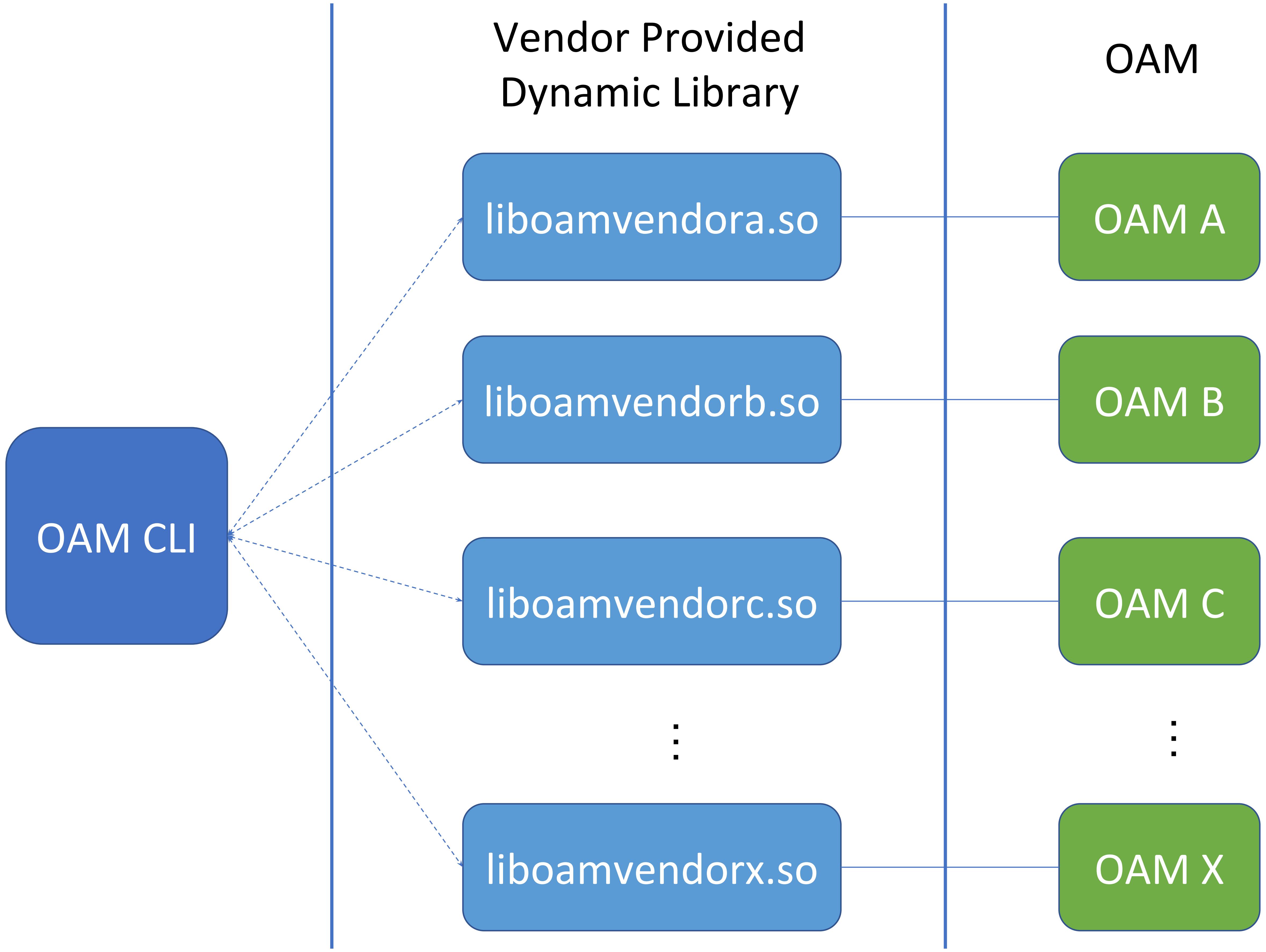
- OAI-Chassis spec with Liquid Cooling Solution
- OAMTool Spec

- OAM reference systems bring-up/validation
- OAM-based systems live demo

OAMTool

- Objectives
  - Standardizing the management of the OAMs in a vendor-agnostic way
- Scope
  - Information and status display
  - Telemetry monitoring and reporting
  - Firmware management
  - Debug log / error counter collection
  - Hardware validation (such as stress tests, HW perf)
  - Power-capping
  - Performance measurement

OAMTool Proposed Architecture



Call to Action

Get involved in the project:

OCP Server Project:	https://www.opencompute.org/projects/server
OAI subgroup:	https://www.opencompute.org/wiki/Server/OAI
OAI mailing list:	https://ocp-all.groups.io/g/OCP-OAI

OCP
REGIONAL
SUMMIT

Open. Together.

OCP Regional Summit

26–27, September, 2019

Presenters

- Siamak Tavallaei is a Principal Architect at Microsoft Azure, co-chair of OCP Server Project, and co-chair of CXL BoD Technical Task Force. Collaborating with industry partners, he drives several initiatives in research, design, and deployment of hardware for Microsoft's cloud-scale services at Azure. He is interested in Big Compute, Big Data, and Artificial Intelligence solutions based on distributed, heterogeneous, accelerated, and energy-efficient computing. His current focus is the optimization of large-scale, mega-datacenters for general-purpose computing and accelerated, tightly-connected, problem-solving machines built on collaborative designs of hardware, software, and management.

- Whitney Zhao is a seasoned hardware engineer leading AI/ML system design in Facebook. Whitney has led multiple hardware generations ranging from general purpose 2S system such as Tioga Pass to ML JBOG Big Basin systems, all of which have been contributed to OCP. She has been driving multiple hardware-software co-design initiatives across both training and inference areas, She is leading the hardware system design for Facebook's main AI workloads. She is also instrumental in bringing industry partners together to solve common infrastructure problem of bringing efficient @scale AI/ML solution for everyone to benefit from.