OPEN POSSIBILITIES.

Introduction of Meta network hardware

Wedge 400: 16x400G + 32x200G ToR Switch

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Introduction of Wedge 400

Lingjun Wu, Hardware Engineer, Meta

Vignesh Vijayanath, Technical Program Manager, Meta

Johnson Liu, Engineering Director, Celestica

OPEN POSSIBILITIES.

Motivation for Wedge 400

- Increased adoption of 100G NICs on Compute and flash storage platforms
- Need for higher front panel port density
- Enabling AI/ML platforms with higher NIC speeds
- Enabling future uplink bandwidth expansion

Key Improvements over Wedge 100s

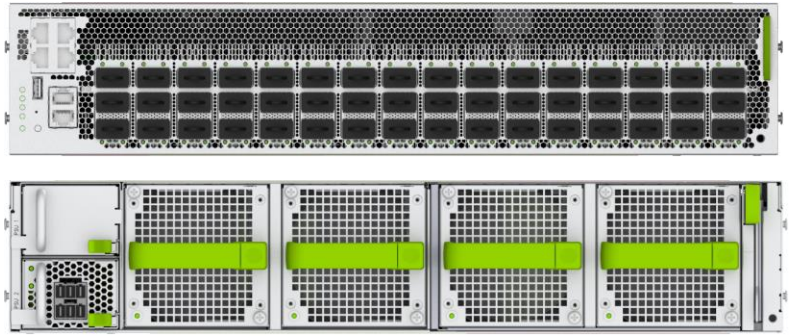
- 2RU 48 Port Switch
  - 32 QSFP56 ports support up to 200G per port
  - 16 QSFP-DD ports support up to 400G per port
- 4X the switching capacity (3.2T → 12.8T)
- 8X burst absorption performance
- RS-FEC support for lower speeds
- Field replaceable CPU sub-system

ToR Evolution



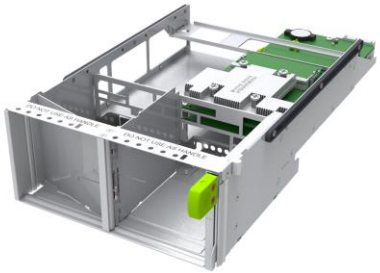
Switch Platform	Wedge 40	Wedge 100	Wedge 400
Year Introduced	2015	2017	2020
ASIC	Trident 2 (BRCM)	Tomahawk (BRCM)	Tomahawk 3 (BRCM)
Switching Capacity	1.28 Tbps	3.2 Tbps	12.8 Tbps
Supported Speeds	10GE/40GE	10/25/40/50/100 GE	10/25/40/50/100/200/400 GE
Physical Port Config	16x 40G QSFP	32 x 100G QSFP28	16 x 400G QSFP-DD + 32 x 200G QSFP56

Product Overview

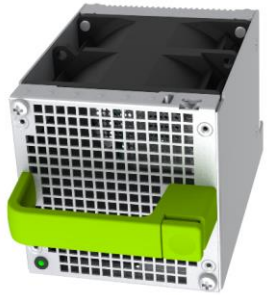


AC PSU

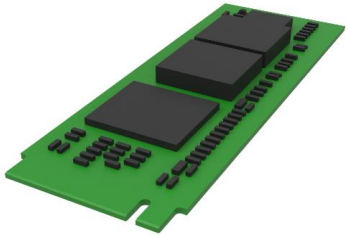
DC PEM



Switch Control Module
(FRU)



Fan Module
(FRU)



BMC Storage Module
(FRU)

- 21 inch ORv2 Tray Adaptor

OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Wedge 400 System Introduction

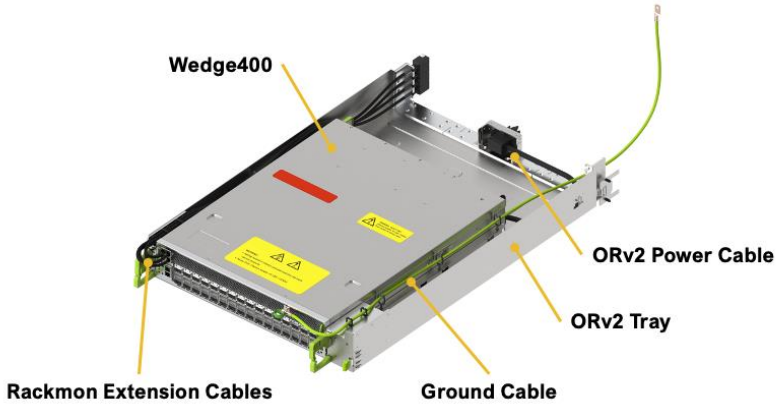
NETWORKING

- Wedge400 HW Description
- Wedge 400 Top/Side View
- Wedge400 Front/Rear View
- SWE (Switch Element)
- Out-of-Band Ethernet
- PCIe Assignment
- UART and Rackmon
- SPI Diagram
- BSM (BMC Storage Module)
- Power Evaluation

Wedge 400: HW Description

NETWORKING

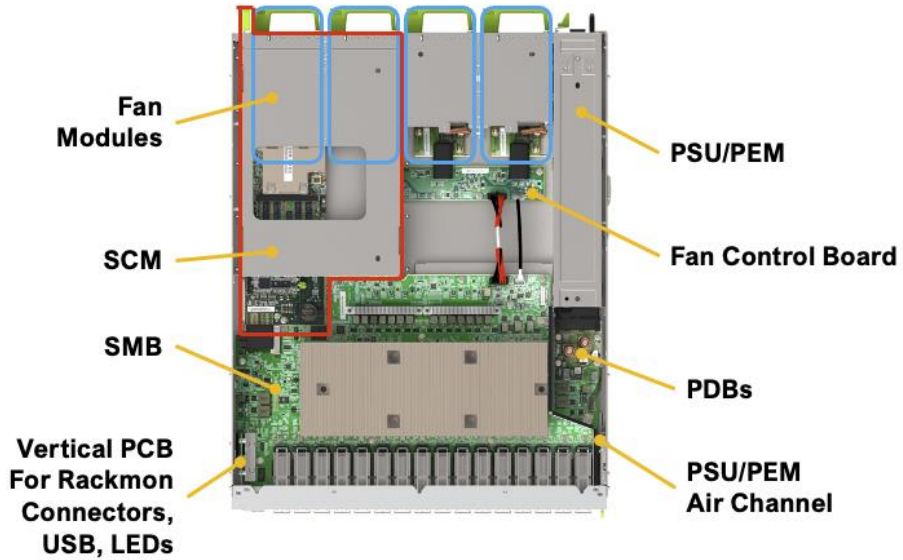
- 19-inch rack, 2RU
- ORv2 adaptor tray
- Uplinks: 16xQSFP-DD (400/200/100G)
- Downlinks: 32xQSFP56(200/100/50/25/10G)
- Switch Software: FBOSS
- Single stage / Single ASIC (TH3)
- Minilake COMe Module (OCP)
- BMC (AST2520)

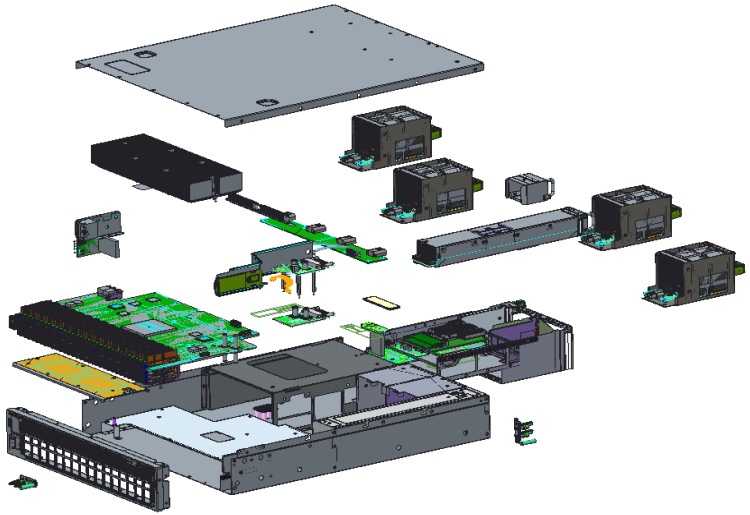


Wedge 400 Top/Side View

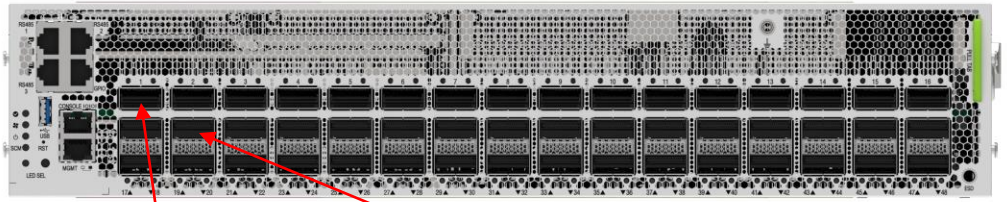
NETWORKING

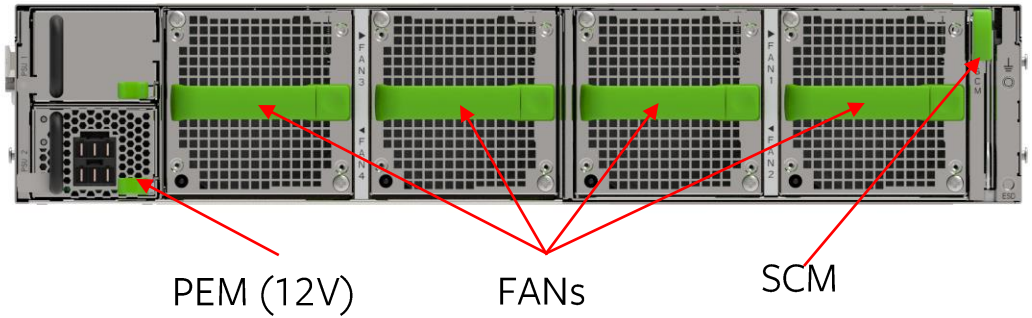
- FRU: SCM, FAN modules, PSU/PEM.





Wedge 400 Front/Rear View





NETWORKING

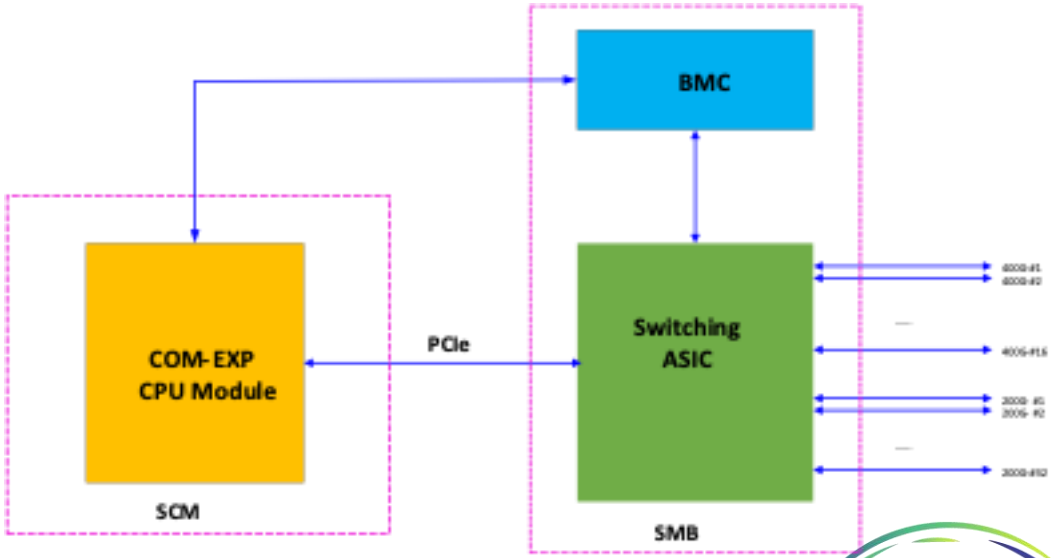
OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Switch Element (SWE)

NETWORKING

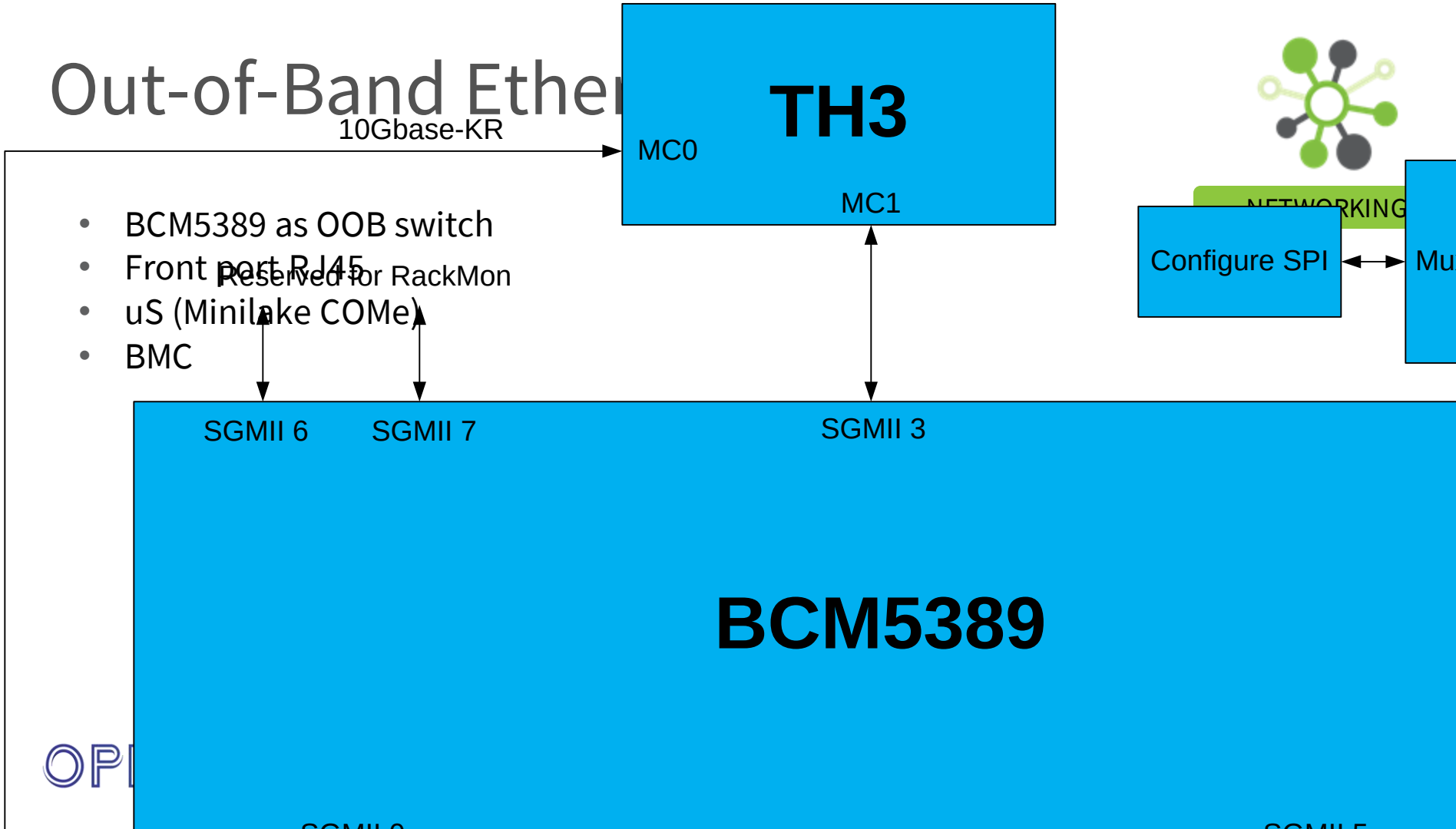
- Switch element consists of three components
- TH3 switch ASIC
- Minilake COMe module
- BMC
- Single switch element



OPEN POSSIBILITIES.

Out-of-Band Ether

- BCM5389 as OOB switch
- Front port RJ45
- uS (Minilake COMe)
- BMC

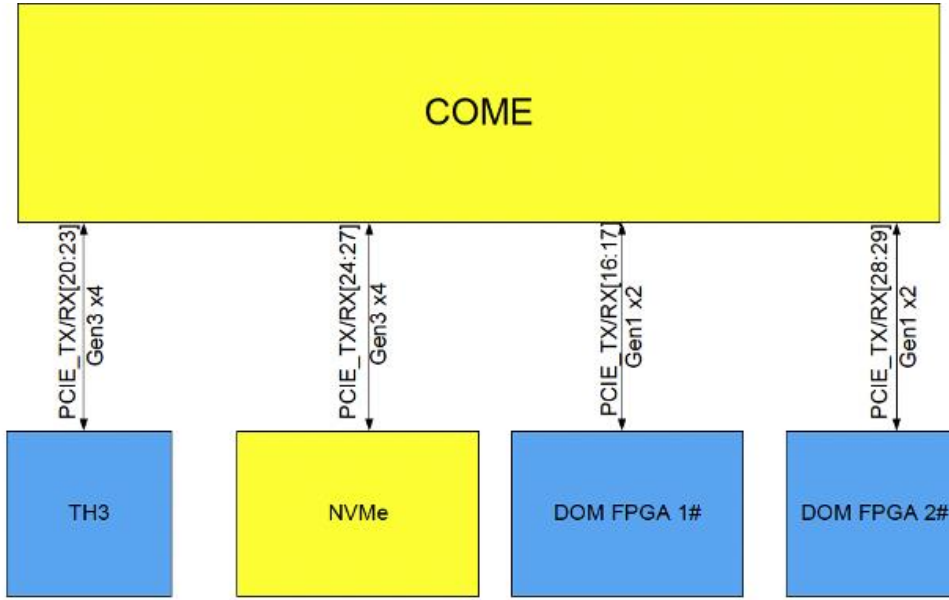


PCIe Assignment

NETWORKING

- PCIe Gen3
- x4 lanes to TH3 (switch ASIC)
- x4 lanes to NVMe SSD

- PCIe Gen1
- x2 lanes to DOM FPGA#1
- x2 lanes to DOM FPGA#2



mon

COME board

LPC

UART

- BMC access
- Sol.sh to uS
- OCP debug card access

SCM_UART5_SEL/SCM_UART2_SEL

- Rackmon
- Manage PSU/BBU inside rack

LPC

UART2

or uS

UART5

BMC

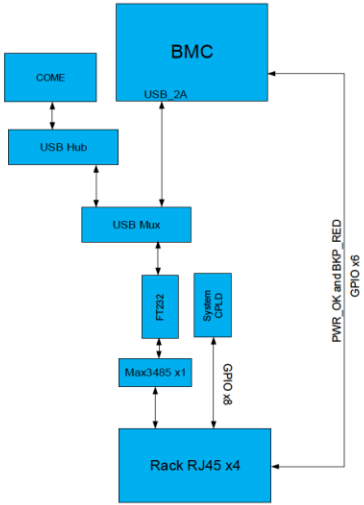
System CPLD

Mux

TIES.

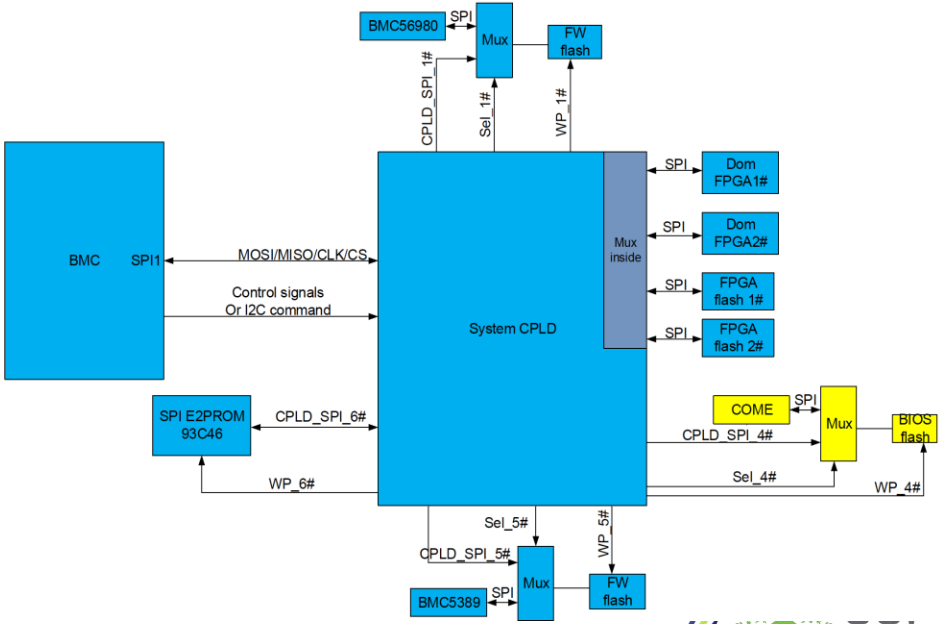
g Uart

NETWORKING



SPI Diagram

- BMC Upgrading FWs
- TH3 FW flash
- FPGA flashes
- Secondary BIOS
- OOB switch config EEPROM



NETWORKING

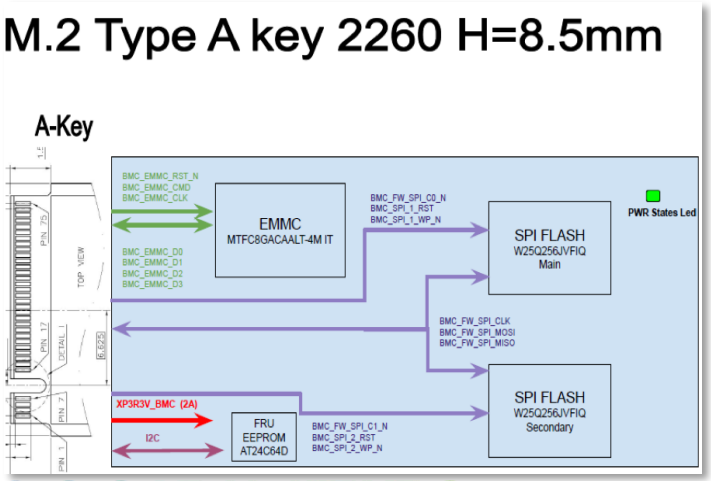
OPEN POSSIBILITIES.

GLOBAL SUMMIT
NOVEMBER 9-10, 2021

BSM (BMC Storage Module)

NETWORKING

- Meta defined BMC storage module
- Primary & secondary flashes
- eMMC and EEPROM



OPEN POSSIBILITIES.

Power Consumption

NETWORKING

Rack Type	Downlink NIC Speed	NIC Qty.	Uplinks	W400 Power (Typical)
Ref #1	50G	16	8*100G CWDM4	240W
			4*100G CWDM4	222W
Ref #2	100G	4	4*100G CWDM4	212W
		8	8*100G CWDM4	242W
Ref #3	100G	32	8*100G CWDM4	285W

Use cases inside Meta

NETWORKING

- Deployed in production since July 2020
- Onboarded onto all new rack platforms (Compute, Storage, AI/ML)

Optics Supported

- 100G (Native, 40G & 10G mode)
- 200G (Native, 100G mod)
- 400G

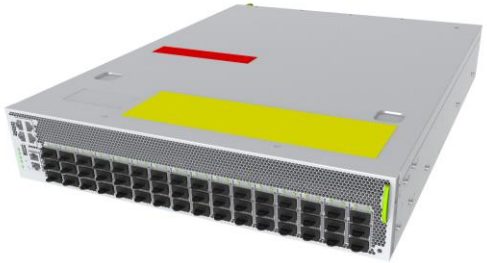
Physical Media Supported

- DAC Q-Q (Multi-host & Single host)
- DAC Q-2Q
- DAC Q-4S

Celestica's Role in OCP

NETWORKING

- Celestica is a leading hardware platform solutions provider to both service provider and enterprise markets, with a focus on leading technologies such as 400G
- We provide both standard solutions or customized solutions by collaboratively engaging with customers like Meta to develop, deliver, and support their platforms such as **Wedge400**



OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Collaboration Model

NETWORKING

HW Development

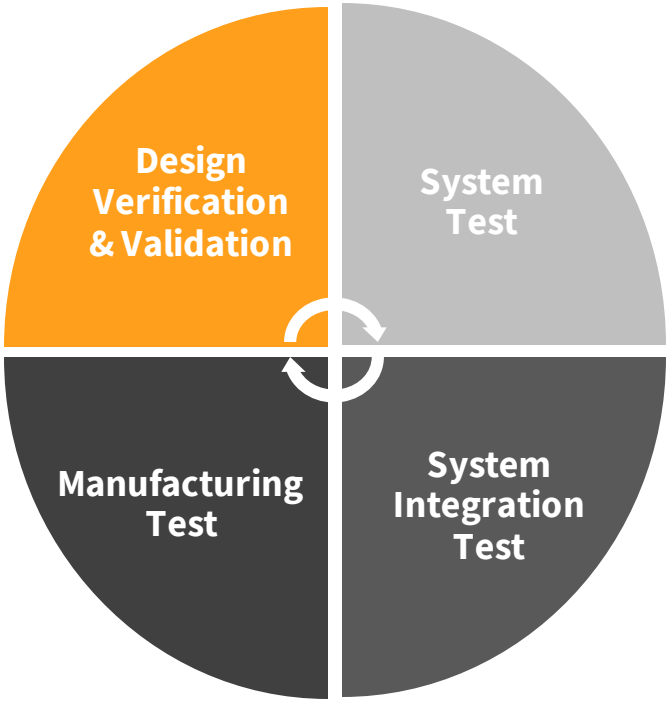
- Celestica aligns with Meta on requirements and finalizes system level architecture.
- Celestica responsible for HW design, and complete prototype bring-up and the following functional/reliability tests.
- Meta will review Celestica engineering design and test reports.

SW Development

- Joint design and source sharing
- From third-party or community, like ASIC SDK/BIOS from vendor, and standard open source Linux distribution.
- Celestica responsible for diagnostic software development.

Wedge 400 Test Strategy

- **Comprehensive test coverage:**
  board level beyond 90% (ICT plus Boundary test) and 100% test coverage for function
- **Closed loop and continuous improvement:**
  test case mapping to product requirement, RCCA process
- **Management system:**
  issue management, issue lifecycle, DI criteria, test progress transparency
- **Automation:**
  efficiency, quality, and software release lead time improvement, real time closed loop, large scale rack orchestration test
- **High test standard:**
  stress test (margin, cycle, long duration), sample size (>35), CICT with dedicated test setup
- **Manufacturing test:**
  criteria for each test phase, design and MFG co-design MFG test process, customized diagnostics



Call to Action

- Timeline for Contribution Availability
  - Wedge400 specs and design package for OCP Accepted in Nov. 2021
- Timeline for Product Availability
  - To be provided by Celestica
- Where to find additional information
  - For more information, please visit: https://www.celestica.com/our-expertise/hardware-platform-solutions/overview
  - For further inquiries, please contact: ccsinnovation-cls@celestica.com

Thank you!

NOVEMBER 9-10, 2021

Open Discussion