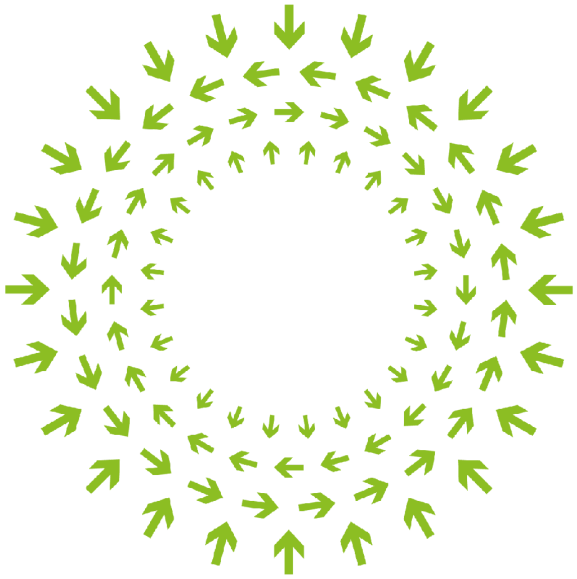


OPEN Compute Project

Network: Switch HW

Meta IceCube OCP Design Spec

1.0.0

Author: Rongchun Zhou, Ivy Wu, Julia Huynh, Kalpak Dhake, Jeremy Chinn, Joel George, Mike Choi

Date: Nov 11, 2026

Version History

Date	Version #	Author	Description
Nov 11, 2026	1.0.0	Rongchun Zhou, Ivy Wu, Julia Huynh, Kalpak Dhake, Jeremy Chinn, Joel George, Mike Choi	Initial release and reviewed/cleaned up based on OCP review feedback.

Current Template Version:

OCP Modular Design Specification Template V1.6.0

Table of Contents

Version History 2
Current Template Version: 2
1. License 4
1.1. Open Web Foundation (OWF) CLA 5
2. Compliance with OCP Tenets 6
2.1. Openness 6
2.2. Efficiency 6
2.3. Impact 6
2.4. Scale 6
2.5. Sustainability 6
3. Design Specifications 8
3.1. Description 8
3.2. User Requirements 8
3.3. Solution Architecture 8
3.4. Supported Functions 8
3.4.1 Environmental Regulatory Compliance and Design 8
3.5. Non-Functional Requirements 10
3.6. Features 11
3.7. Physical Design 11
3.8. Mechanical Design 11
3.8.1 Rack Compatibility 13
3.8.2 General Requirements 14
3.9. Electrical Requirements 14
3.10. Thermal Design Requirements 14
3.10.1 Liquid Cooling 16
3.10.2 Air Cooling 17
3.10.3 Thermal Control 17
3.11. Onboard Power 17
3.12. Interfaces 20
3.12.1 DC Power Connector 20
3.12.2 OSFP User Ports 20
3.12.3 QSFP28 Ethernet Management Port 20
3.12.4 RJ45 Ethernet Management Port 20
3.12.5 RJ45 Console Port 20
3.12.6 USB Management Port 21

3.12.7 LSC port 22
3.12.8 System LEDs 23
3.12.9 Signal List 23
3.12.10 Rear Side Power, I/O, Expansion Board and Midplane Subsystems 23
3.13. Performance 23
3.14. Security 23
3.15. Hardware Management 24
3.15.1 UART 25
3.15.2 OOB Ethernet 26
3.15.3 I2C bus tree 27
3.15.4 SPI bus tree 28
3.15.5 PIC (Physical Interface Card) 29
3.15.6 Sensors 29
3.15.7 Fan PWM and RPM 30
3.15.8 Platform Management 30
3.15.9 LEDs 31
3.16. Constraints 32
3.17. Test and Validation 32
3.18. External Software/Firmware Support (Strongly recommended) 32
3.19. Compliance 33
3.20. Arm SystemReady (only for Arm-based Systems) Requirement 34
3.21. Repository Location 34
3.22. Prescribed Materials 34
3.23. References (recommended) 34

1. License

THE UPDATED DEFAULT CONTRIBUTOR LICENSE AGREEMENT (CLA) IS **OWFa 0.9**. PLEASE VERIFY THE CORRECT CLA/FSA IS USED AND EXECUTED FOR THIS CONTRIBUTION.

1.1. Open Web Foundation (OWF) CLA

Contributions to this Specification are made under the terms and conditions set forth in **Modified Open Web Foundation Agreement 0.9 (OWFa 0.9)**. (As of October 16, 2024) ("Contribution License") by:

Meta Platforms, Inc.

Usage of this Specification is governed by the terms and conditions set forth in **Modified OWFa 0.9 Final Specification Agreement (FSA)** (As of October 16, 2024) **("Specification License").**

You can review the applicable Specification License(s) referenced above by the contributors to this Specification on the OCP website at https://www.opencompute.org/contributions/templates-agreements.

For actual executed copies of either agreement, please contact OCP directly.

Notes:

The above license does not apply to the Appendix or Appendices. The information in the Appendix or Appendices is for reference only and non-normative in nature.

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 IN ORDER 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

Please describe how this Specification complies with the OCP tenets.

A full explanation of the OCP core tenets can be seen here.

2.1. Openness

The electrical design and mechanical design of IceCube is fully open. All the PCB, BOM and chassis design files have been included in the design package that has been submitted to the OCP community.

2.2. Efficiency

The IceCube system has a PHY-less design. It uses industry hardened N+M PCB construction for the SMB (Switch Main Board) PCB. Its dimension is only 2OU, about half of the height of its previous fabric switches at Meta. The platform software design leverages the Meta's established BMC-lite and PlatformManager architecture. It has better power efficiency, lower acoustic noise level and improves operational efficiency over its predecessors.

2.3. Impact

IceCube is based on the industry latest and largest 102T switch ASIC. It introduces several groundbreaking innovations, such as using RF flyover cables for high speed in-chassis connectivity and hybrid air-liquid cooling technology. It will lead the industry to turn these innovative technologies into maturity.

2.4. Scale

IceCube's major switching role is for the back-end network applications. It can be used for the front-end network applications as well. The IceCube can be used with various optics module types, such as 2x800G-DR4 and 2x800G-FR4, either fully retimed or LRO, and at various speed modes. The IceCubes are installed in the OCP ORv3-HPR racks that are commonly available at Meta. These features allow IceCube to be deployed at scale to serve versatile switch roles.

IceCube has built-in many sensors to be monitored in the large data centers. The FWs on IceCube are field-upgradeable. These features allow IceCube to be deployed and maintained at large scale.

2.5. Sustainability

IceCube consumes less than 3,820W in the worst case for Meta's applications with 5nm optic DSP pluggable modules. It achieves a power efficiency of about 37pj/bit, which is

about 7% better than its predecessors. With 3nm DSP optics modules, the power efficiency will be further improved to less than 34pj/bit.

3. Design Specifications

3.1. Description

Meta's IceCube is a high performance 102.4T Ethernet switch. It provides 128x800GE or 512x200GE non-blocking switch capacity, and is backward compatible to 400GE, 200GE and 100GE standards. The system design aims to improve system reliability, resiliency, operational efficiency and NPI efficiency.

3.2. User Requirements

The IceCube network switch device shall be within 2OU form factor and deployed at ORv3 HPRv1 racks.

3.3. Solution Architecture

The IceCube network switch contains a Switch Main Board (SMB), a Main Carrier Board (MCB), two Physical Interface Cards (PIC), two LED panel cards, one IO card on the left and one IO card on the right.

The switch ASIC on the SMB is the most critical part of the IceCube, which provides up to 512x 200GE switch capacity.

3.4. Supported Functions

The IceCube switch supports hybrid liquid and air cooling. It has four fan trays that are in 3+1 redundancy that support field replacement.

Its E1.S SSD, RunBMC module card and memory cards are field replaceable to improve the system reliability.

3.4.1 Environmental Regulatory Compliance and Design

The following compliance reports must be from labs accredited to a current version of IEC 17025.

- CE Declaration to the following Regulatory Directives
  - EMC Directive 2014/30/EU
  - Low Voltage (LVD) Directive 2014/35/EU
- Safety Certification
  - CB certificate/report to IEC 62368-1 including all national deviations
  - CB certificate/report to IEC/EN 60950-1 including all national deviations
  - UL/CSA/IEC/EN 60950-1 with all latest amendments
  - UL/IEC 62368-1 with all latest amendments
  - CNS14336 Taiwan BSMI safety regulation
  - EN 60825-1 Safety of Laser products – part 1

    - UL 94-V0 Flammability rating
- EMC certification and report
    - FCC Part 15 (Class A)
    - ICES-003 (Canada) Class A
    - EN55032 (Europe) Class A
    - CISPR32 (International) Class A
    - AS/NZS CISPR32 (Australia and New Zealand) Class A
    - VCCI CISPR 32 (Japan) Class A
    - CNS 13438 (Taiwan) Class A
    - EN61000-3-2
    - EN61000-3-3
    - EN55035
    - EN 61000-4-2 ESD
    - EN 61000-4-3 Radiated Immunity
    - EN 61000-4-4 EFT
    - EN 61000-4-5 Surge
    - EN 61000-4-6 Low Frequency Conducted Immunity
    - EN 61000-4-11 Voltage Variations and Dips

The immunity levels are defined in the following table.

Network Equipment						
Immunity Standard	Description	Standard Criteria (Required)	Meta Criteria (Goal)	Standard Level (Required)	Meta Level (Goal)	Remarks
IEC 61000-4-2	Electrostatic Discharge (ESD)	B	A	4 kV Cont. / 8 kV Air	5.6 kV Cont. / 11.2 kV Air	Criteria A (Meta Goal) is highly desirable for any level
IEC 61000-4-3	Radiated Immunity	A	A	3 V/m	10 V/m	For communication cables >3 meters
IEC 61000-4-4	EFT	B	B	0.5 kV		Ports: Signal &Telecom (>3 meters), DC power input
		B	B	1 kV		Ports: AC power input & AC/DC Converter
IEC 61000-4-5	Surges	C	A	1 kV		Ports: Signal and Telecom connected to outdoor cables

		B	A	0.5 kV		Ports: DC power input
				1 kV L-L / 2 kV L-GND		Ports: AC power input & AC/DC Converter
IEC 61000-4-6	Conducted Immunity	A	A	3 V	10 V	Ports: Signal & Telecom (>3 meters), AC and DC power input
IEC 61000-4-8	Magnetic Field	A	A	1 A/m		Only EUT containing devices susceptible to magnetic fields
IEC 61000-4-11	Voltage Dips	B	B	>95% Reduction		Ports: AC power input & AC/DC Converter (0.5 Period)
		C	C	30% Reduction		Ports: AC power input & AC/DC Converter (25 Periods)
	Voltage Interruptions	C	C	>95% Reduction		Ports: AC power input & AC/DC Converter (250 Periods)

Table 1: Network Equipment Immunity Test Requirements

The following is the environmental compliance requirement.

- Environmental Compliance
  - RoHS 2
  - REACH (SVHC & Annex 17)
  - WEEE
  - POP Regulation
  - Prop 65
  - Batteries Directive
  - Halogen-free IEC/EN 61249-2-21 (900 ppm Br or Cl, 1500 ppm combined)
  - Phthalate (DEHP, DBP, DiBP, BBP)-free (1000 ppm)
  - Arsenic-free (1000 ppm)
- CE Declaration to the following environmental Directives:
  - RoHS Directive 2011/65/EU

3.5. Non-Functional Requirements

IceCube system design shall support hot insertion and removal for in fleet maintenance.

3.6. Features

Please refer to Hardware Management section for the hardware features supported in IceCube.

3.7. Physical Design

IceCube system design uses flyover cable technology to route the high speed RF signals to the front panels. The following rendering illustrates its internal chassis.

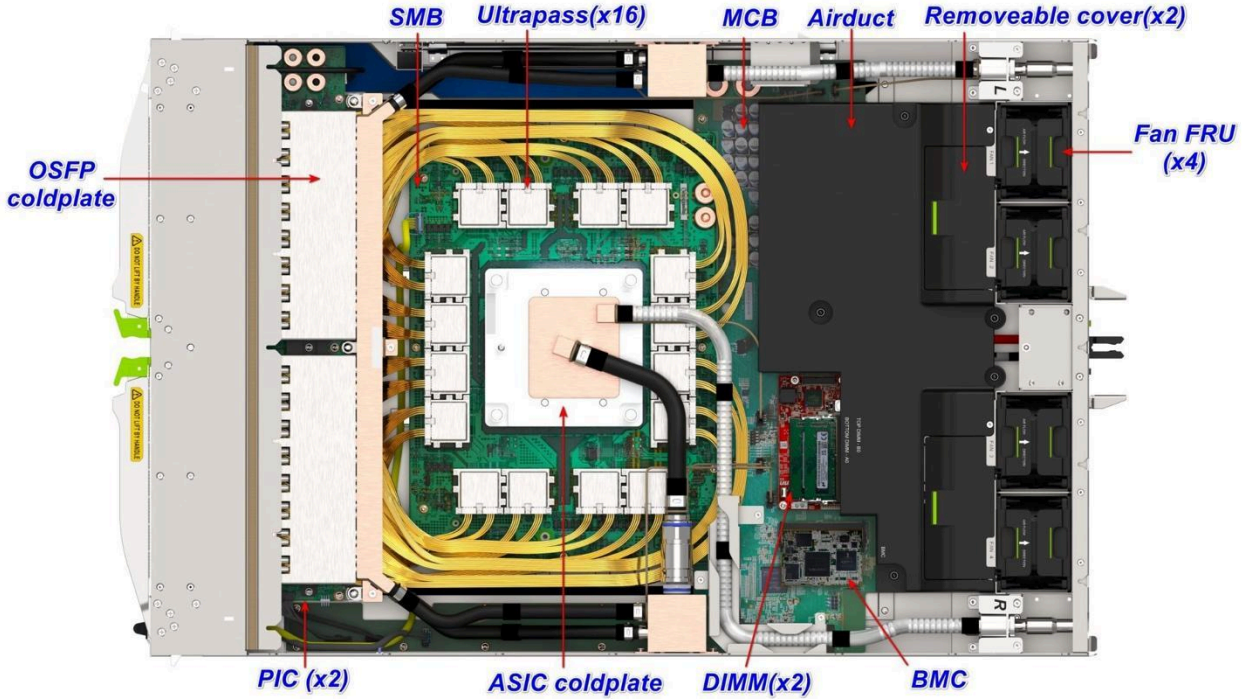


Figure 1: IceCube Internal Chassis Top View

3.8. Mechanical Design

IceCube chassis fits in the compact 2OU form factor.
Handles:

- Four handles on top of the chassis
- Two (2) ejector handles for ejection and injection
- Stop feature for operation safety

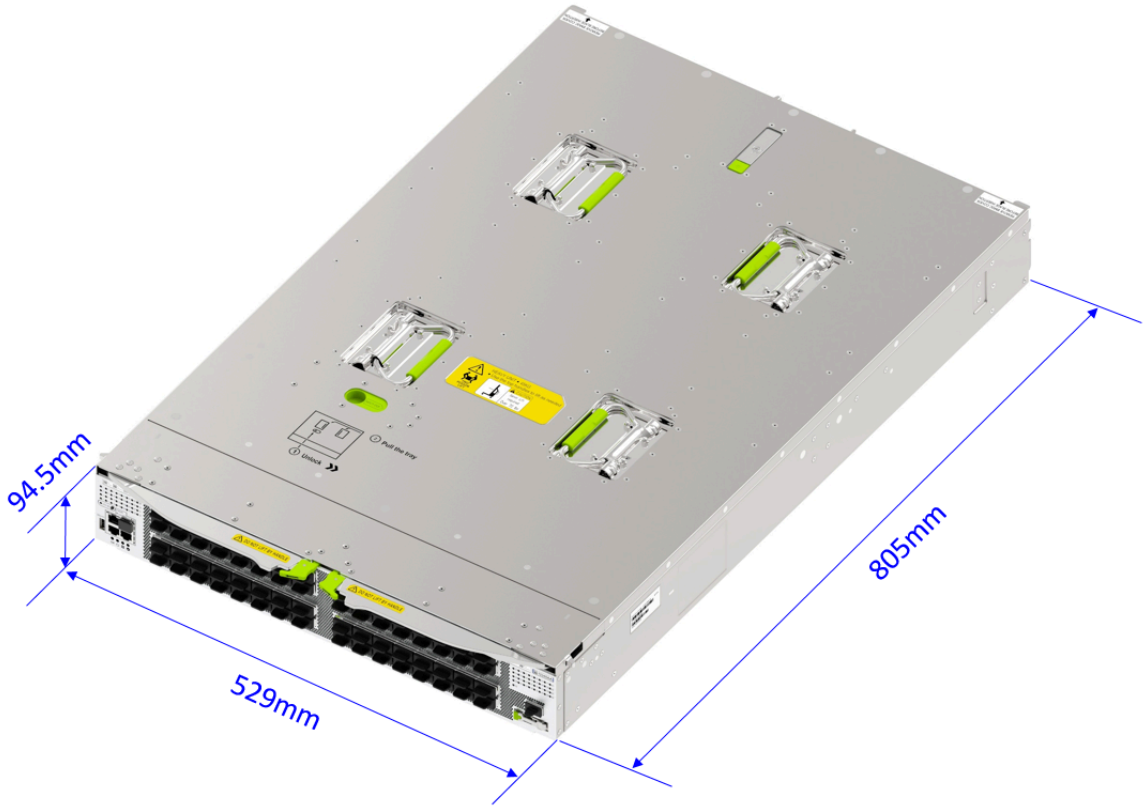


Figure 2: IceCube Chassis ISO View

There are 64 OSFPs connectors in 1x8 belly-to-belly layout on the front panel. An E1.S SSD is accessible from the front panel. A QSFP28 port on the front panel connects to the ASIC Ethernet management port. Enough ventilation should be provided for air cooling.

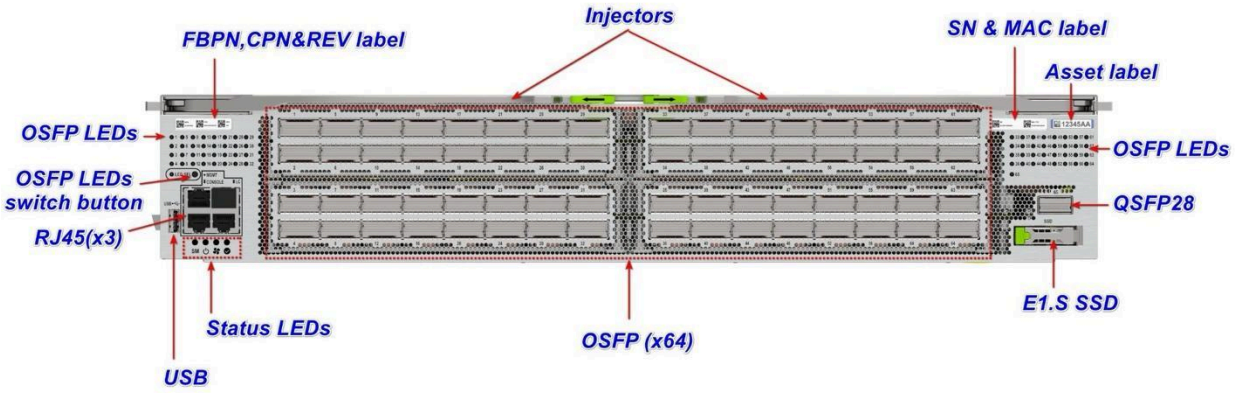


Figure 3: IceCube Chassis Front View

The IceCube has two BMQC (Blind Mated Quick Connect) connectors from the rear. One is for cold coolant input on the right (rear view), and the other is for the hot coolant return on the left (rear view). The blade receives its 51V power from the ORv3-HPR bus bar on the rear. There should be enough venting holes on the rear for air cooling.

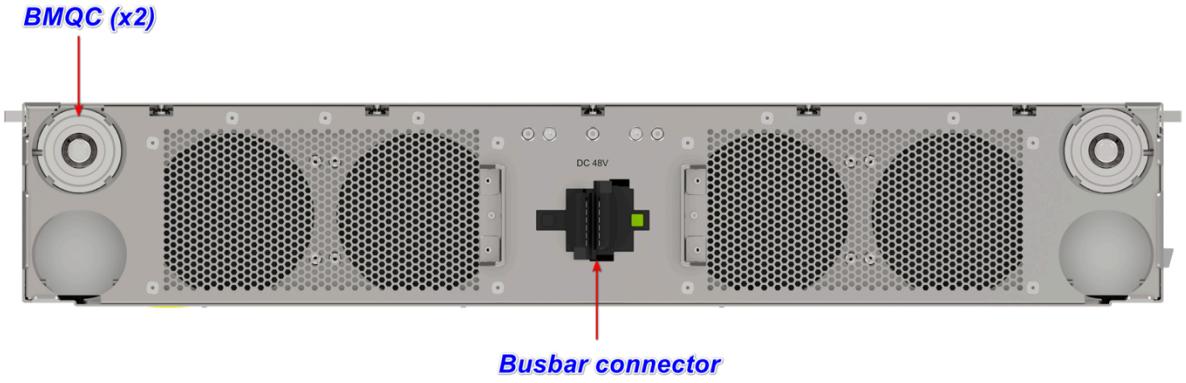


Figure 4: IceCube Chassis Rear View

3.8.1 Rack Compatibility

The IceCube shall be installed on OpenRack V3 HPRv1 racks, with a canister adapter.

3.8.2 General Requirements

Item	Requirements
Physical User Port	64x OSFP-RHS
ASIC Management Port	1x QSFP28
SSD	1x E1.S
User port LEDs	2x tri-color LEDs per port. 1x LED per port is acceptable if time multiplexed.
SYS LED	4x tri-color system LEDs
Console Port	1x RJ45
OOB Management Port	1x RJ45
Liquid Sensing Cable Port	1x RJ45
Power Input	51V ORv3-HPR power busbar
Liquid Connector	2 BMQC connectors
Weight Target	46Kg or 102Lb
Dimension Envelope	94.5mm(H) x 529mm(W) x 805mm(D)

Table 2: IceCube External Interfaces and Mechanical Requirements

3.9. Electrical Requirements

The blade power input envelope conforms to ORv3-HPR power spec: 45 - 53 Vdc (steady state), 41 - 57 Vdc (transient less than 2 ms). The blade system power shall not exceed 3,820 W for the typical Meta applications.

The IceCube is based on the Broadcom Tomahawk6C switch ASIC. The system architecture doesn't limit the switch ASIC to Broadcom TH6C. It could be an equivalent switch ASIC with 512x212GPAM4 SerDes that satisfies the similar hardware interface, thermal, and application software requirements.

3.10. Thermal Design Requirements

The following table defines the IceCube operating and non-operating environmental specs.

IceCube Environmental Spec	min	max	Note
Stressed Operating Temperature	0ºC	45ºC sea level	manufacture burn-in
Short-term Operating Temperature	0ºC	35ºC sea level	
Recommended Operating Temperature	0ºC	30ºC 6000ft	Passed -5C sea level
Recommended Operating Relative Humidity	5%	90%	Passed 93% 35C
Storage Temperature	-40º C	70ºC	
Storage Relative Humidity	5%	95%	
Non-operating Altitude		12,000meter	-40C~65C
Sound pressure		63.3dBA	30ºC 6000ft

Table 3: IceCube Operating/Non-operating Conditions

The thermal design of IceCube utilizes hybrid air-liquid cooling technology. High-power devices, specifically the OSFP optic modules and switch ASIC, are liquid cooled, whereas other system elements like the COMe/SCM, board-level components, RunBMC, QSFP28, E1.S SSD, and power stages rely on front-to-back air cooling. This section details the component temperature, liquid-loop, TIM, and cold-plate requirements alongside L10 blade-level validation criteria. The thermal design of IceCube is optimized to support a 102T-generation switch ASIC and OSFP 128x800G-DR4, FR4 and potentially LR4 or ZR optics (up to 43 W per OSFP) in a 1x8 belly-to-belly configuration.

The system liquid cooling interface is defined by the two Blind Mated Quick Connect (BMQC) couplings situated at the chassis rear. The following liquid cooling specifications are mandatory at this boundary.

The following table summarizes the liquid and air cooling parameters.

Attributes	Requirement	Notes
Coolant Type	PG25 (25% propylene glycol)	
Inlet Coolant Temperature	max 40C	Design point

Inlet Coolant Location	Left BMQC (front view)	Per rear-view definition
Outlet Coolant Location	Right BMQC (front view)	
Liquid Flow Rate [LPM]	< 1.5 * liquid cooled power [kW]	Efficiency cap
Minimum Flow Rate	[TBD] LPM	Meet all Tj/Tcase limits
Pressure Drop (dP)	≤ [TBD] kPa	Measured at BMQC
Max Working Pressure	[TBD] kPa	Static and transient
Max Ambient Temperature	35ºC	
Operating Altitude	6000 ft	
System Airflow	<0.145CFM/W (sea level), 0.181CFM/W (6000 ft)	Fully populated, max traffic

Table 4: IceCube Thermal Design Specs

The IceCube system is designed to prevent condensation on any chassis surface during operation. All surfaces carrying liquid must remain above the maximum dew point calculated from the environmental operating humidity and ambient air limits provided in the Product Specifications.

Domain	Cooled Components	Cooling Method
Liquid	Switch ASIC, OSFP optical modules (64×)	Cold plate
Air	SCM/COMe, power stages/VRMs, E1.S SSD, QSFP28, RunBMC, board devices	Front-to-back airflow

Table 5: IceCube Components Cooled by Liquid and Air

3.10.1 Liquid Cooling

At minimum, the switch ASIC and all OSFP optical modules shall be liquid-cooled.

Coverage and construction

- Cold plates shall provide thermal coverage for the switch ASIC and for the OSFP modules in the 1×8 belly-to-belly front-panel arrangement.
- Wetted materials shall be compatible with PG25 coolant for the service life of the product. Galvanic couples between dissimilar wetted metals shall be avoided.

Thermal performance

- Each cold plate shall achieve a thermal resistance sufficient to hold its target component below the limits in at the maximum inlet coolant temperature (40 °C) and minimum rated flow.
- The OSFP optics cold plate shall hold each module case temperature ≤ 70 °C across all supported optics types and speed modes.

Hydraulic performance

- Liquid flow rate to the system liquid-cooled component [LPM] shall be < 1.5* power of the liquid cooled components [kW]
- The pressure drop of each blade at rated flow shall be ≤ 10 psi.
- Flow distribution across parallel cold-plate paths shall be balanced such that the least-favored component still meets its temperature limit.

3.10.2 Air Cooling

- Airflow shall be front-to-back.
- Cooling shall be provided by 4× (80 mm × 80 mm × 56 mm) fans at the chassis rear with N+1 redundancy.
- The system shall maintain all air-cooled components within their operating limits with any single fan tray failed, and shall support indefinite service time with one fan tray removed.
- The airflow impedance path shall be designed such that total system airflow does not exceed the CFM/W 0.145 CFM/W at sea level and 0.181 CFM/W (total system airflow/ total system power) at an altitude of 6000 ft when the system is fully populated with any of the supported optics (and at maximum traffic)
- Fan PWM control shall operate at approximately 25 kHz; fan RPM shall be monitored, with an out-of-range condition reported to the microserver.

3.10.3 Thermal Control

- The microserver shall monitor, at minimum: system inlet ambient temperature, switch ASIC on-die temperature, CPU temperature, optics temperatures, and coolant inlet/outlet temperature.
- Fan speed shall be governed by a combined control law: a step function on inlet ambient temperature plus PID control on critical components (switch ASIC on-die, CPU, optics).
- The control system shall have access to all thermal sensors in the system and shall expose thermal telemetry for data-center monitoring.

3.11. Onboard Power

The system onboard power design shall meet the system operating requirements. It needs to have enough power capacity and meets the devices power-up/down sequence requirements.

The following diagram illustrates the IceCube system power tree.

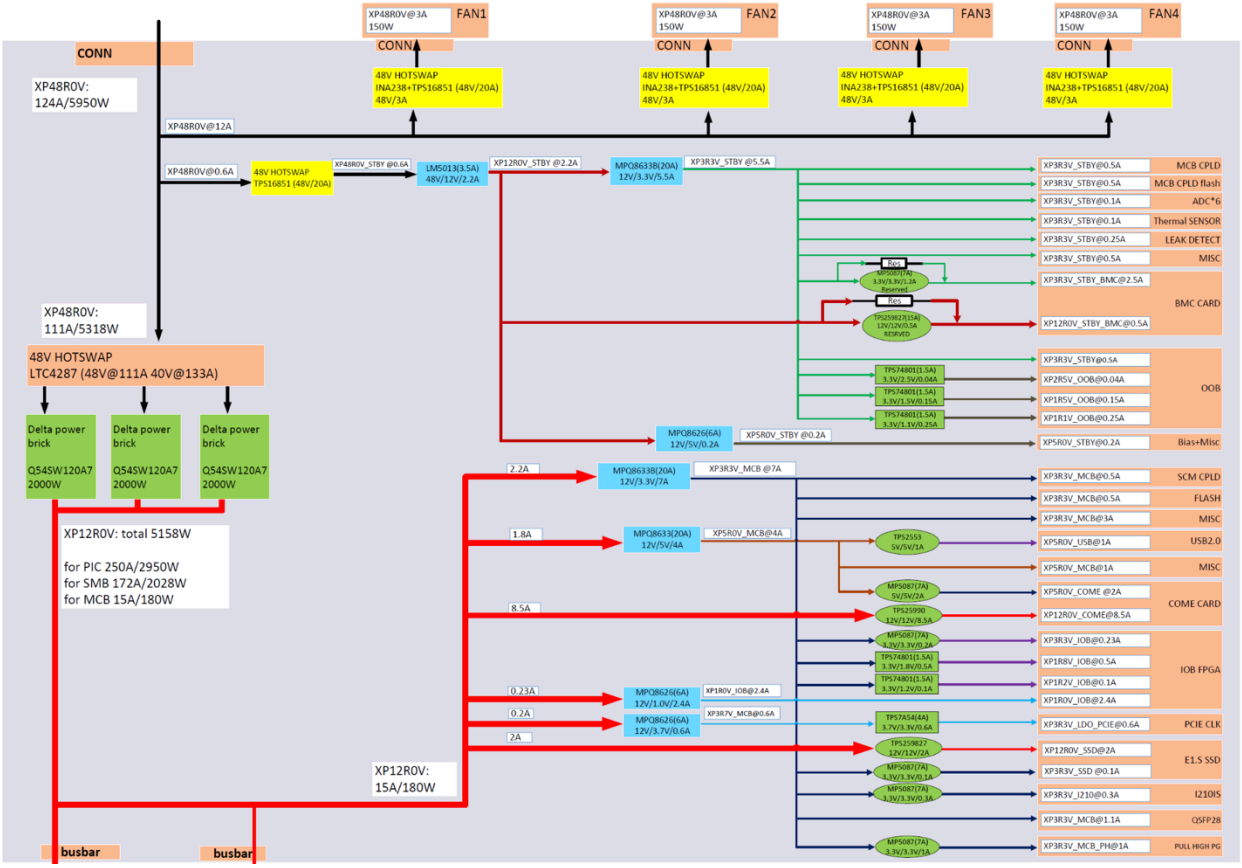


Figure 5: IceCube Power Tree on MCB

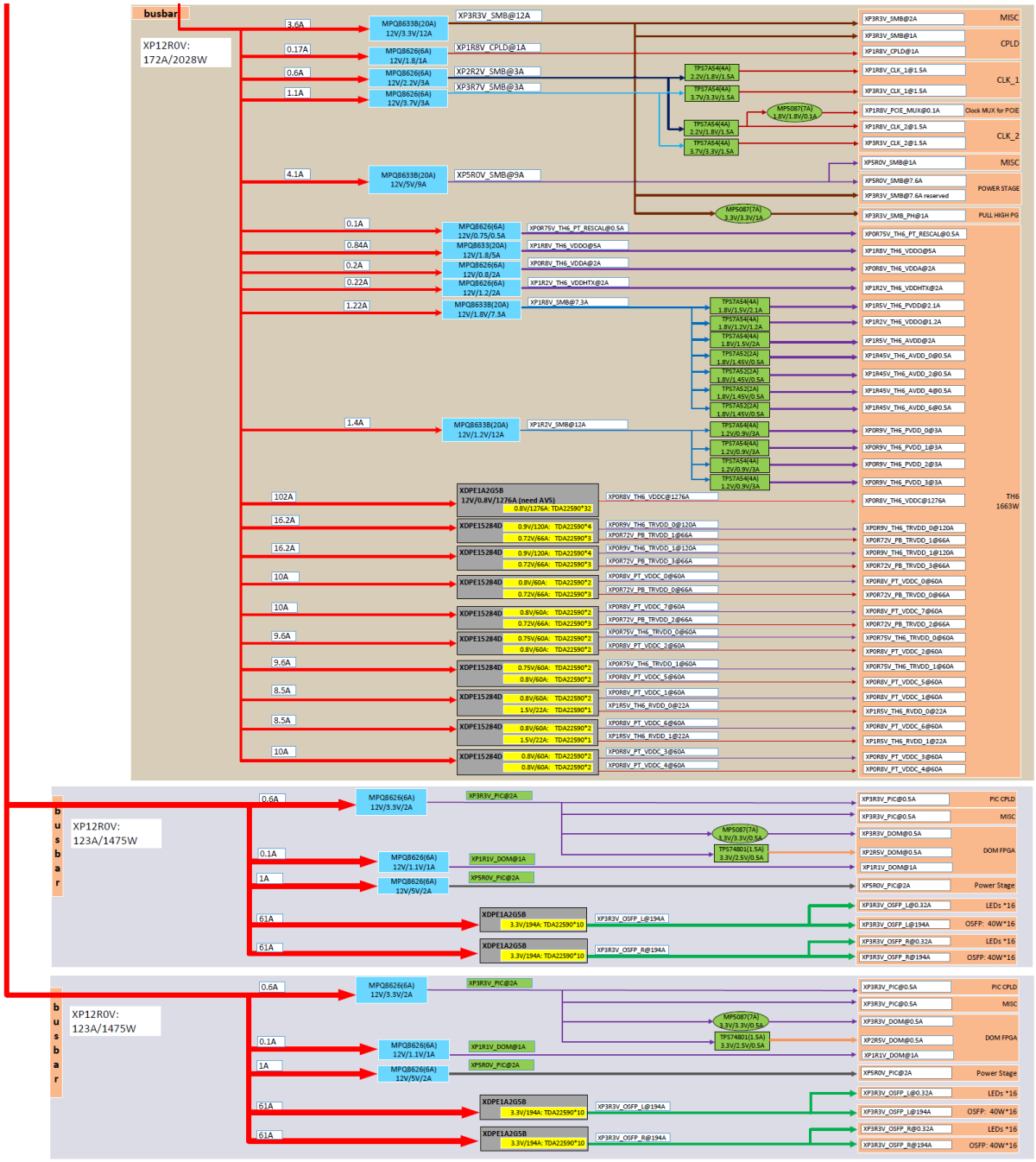


Figure 6: IceCube Power Tree on SMB and PICs

3.12. Interfaces

3.12.1 DC Power Connector

The system shall provide a DC power connector that connects to ORv3-HPR power busbar for Meta data center operations.

3.12.2 OSFP User Ports

The system shall have 64x OSFP-RHS user ports that are compliant to OSFP MSA specifications. Each OSFP port shall support up to 43W power consumption for potential ZR applications.

The OSFP user ports shall be OIF CEI-224G-VSR-PAM4 compliant to support Meta 2x800G optics modules. The OSFP user ports should also support LRO (Linear Receiver Optics) modules.

Each OSFP user port shall have two tri-color LEDs for two of its sub-ports when using Meta's 2x800G optics modules. A single LED per port is acceptable, but there should be a method to time-multiplex the LED to display the states of the two sub-ports.

3.12.3 QSFP28 Ethernet Management Port

The system shall have one QSFP28 Ethernet management port connected to the switch ASIC management port. The port shall support 100GE-CR4 and 100GE with optics modules.

50GE-CR2 is acceptable. When 50GE-CR2 is supported, lane 1 and lane 2 will be active, while lane 3 and lane 4 are unused.

With a QSFP28-to-SFP+ adaptor, the QSFP28 Ethernet management port should support 10GE SFP+.

The QSFP28 port shall have one tri-color LED for its status

3.12.4 RJ45 Ethernet Management Port

The system shall have one RJ45 Ethernet management port for 1000Base-T applications. The connector shall support CAT6/6A copper cable. The connector shall have two LEDs for link status indication.

3.12.5 RJ45 Console Port

The system shall have one RJ45 console port for RS232 applications at 9600 baud rate. The connector shall have the following pinout definition.

Pin Number	Pin Name
1	RTS
2	DTR
3	TXD
4	GND
5	GND
6	RXD
7	DSR
8	CTS

Table 6: Console port pinout definition

3.12.6 USB Management Port

The system can optionally have a USB 3.0 type-A connector. An OCP debug card can be plugged into the system USB 3.0 connector for debugging purposes. This connector should support three debugging functions: a USB 2.0 High Speed interface, a debug UART interface and a debug I2C interface. Its pinout is defined in the following table.

Pin Number	Pin Name	Note
1	VBUS	
2	USB 2.0 DP	
3	USB 2.0 DN	
4	GND	
5	I2C_SCL, 3.3V	
6	I2C_SDA, 3.3V	
7	debug_card_present	Active high to indicate OCB debug card is present to enable debug UART buffer.
8	UART_RX, 3.3V	from remote to the system

9	UART_TX, 3.3V	from the system to remote

Table 7: USB port pinout definition

3.12.7 LSC port

Leak Sensor Connector port is a RJ45 port used to interconnect with RMC (Rack Management Controller). The blade sends the blade leak events to the RMC, and RMC turns off the blade main power in case of leaks. The pinout of this connector is defined in the following table.

Pin	**Signal Name**	**Description**	**I/O**	**Tray**	**RMC**
1	Leakage_SCL	I2C interface to leakage detection logic	I	Pulled up to 3.3V_STBY with 4.7Kohm	
2	GND			Connect to GND	Connect to GND
3	Leakage_SDA	I2C interface to leakage detection logic	I/O	Pulled up to 3.3V_STBY with 4.7Kohm	
4	Small_Leak_L	Small leak signal. Once a small leak is detected, then tray logic will pull this signal low.	O	Connect to CPLD with isolator	Pulled up to 3.3V_STBY through 4.7k at RMC
5	Tray_PowerGood_L	3.3V STBY power good signal. Active Low. Tray will hold HighZ until CPLD is fully up.	O	Connect to CPLD with isolator	Pulled up to 3.3V_STBY through 4.7k at RMC
6	Tray_PowerEnable_L	Active Low. Tray will pull this pin Low. RMC will connect this pin to CPLD. Once this pin is pulled high by RMC, the tray shall turn off main power and leave STBY power on.	I	Pull low through 4.7Kohm	RMC will pull signal high to 3.3V_STBY if a leak that required trays shutdown occurs

7	Tray_Present_L	Tray Present signal. Active low.	O	Connect to GND through 0ohm resistor	Pulled up to 3.3V_STBY through 4.7k
8	Tray_Leak_L	Tray leak signal. Once a tray leak is detected at the base of the tray, then tray logic will pull this signal low.	O	Connect to CPLD with isolator	Pulled up to 3.3V_STBY through 4.7k at RMC

Table 8: LSC port pinout definition

3.12.8 System LEDs

Besides the OSFP user port and QSFP28 LEDs, the system has several management LEDs.

- SYS LED
  - A tri-color LED for the whole system
- SMB LED
  - A tri-color LED for SMB status
- PWR LED
  - A tri-color LED for system power status
- FAN LED
  - A tri-color LED for fan tray status

3.12.9 Signal List

The IceCube IO interface signal definition can be found in the previous section.

3.12.10 Rear Side Power, I/O, Expansion Board and Midplane Subsystems

The IceCube system shall support hot insertion and removal to improve the overall switch network reliability.

3.13. Performance

IceCube supports 128x800GE or 512x200GE non-blocking switch capacity.
Its power efficiency for Meta's typical use case is about 37pJ/bit with 5nm DSP optic modules.
Its acoustic noise level is 63.3dBA.

3.14. Security

The IceCube is capable of a security mechanism that is supported by the CPU processor, which can be extended and be used as the root of the trust mechanism / chain, as well as attestation.

On the BMC side, the BMC CPU is also capable of secure boot / verified boot mechanism by enabling this mechanism at u-boot, and using the locked bootloader partition in one of the BMC SPI flash chips.

3.15. Hardware Management

IceCube implements BMC-lite management architecture. The BMC manages the most critical platform management tasks, such as power-on control, fail recovery and system event logging. The microserver takes care of the other tasks such as VRM control, sensor management, fan speed control algorithm, FW programming, etc.

IceCube shall have an out-of-band management connection through RJ-45 Ethernet port.

Both microserver and BMC shall be able to reprogram the BIOS.

Both microserver and BMC shall be able to reprogram IOB FPGA. Microserver shall be able to reprogram the FPGAs and CPLD FW images.

Microserver shall have access to all the sensors in the system.

The CPU microserver in the system is Intel IceLake-D1736 CPU. The system architecture doesn’t limit the CPU selection to Intel Ice Lake-D1736. An alternative CPU is possible as long as it satisfies the hardware interface, thermal, and application software requirements. Minimum 64GB DDR4 DIMM w/ ECC.

Under the BMC-lite architecture on Minipack3, the following logic devices in the system assist BMC and CPU to execute the platform management tasks: IOB FPGA, two DOM FPGA, and a few CPLDs.

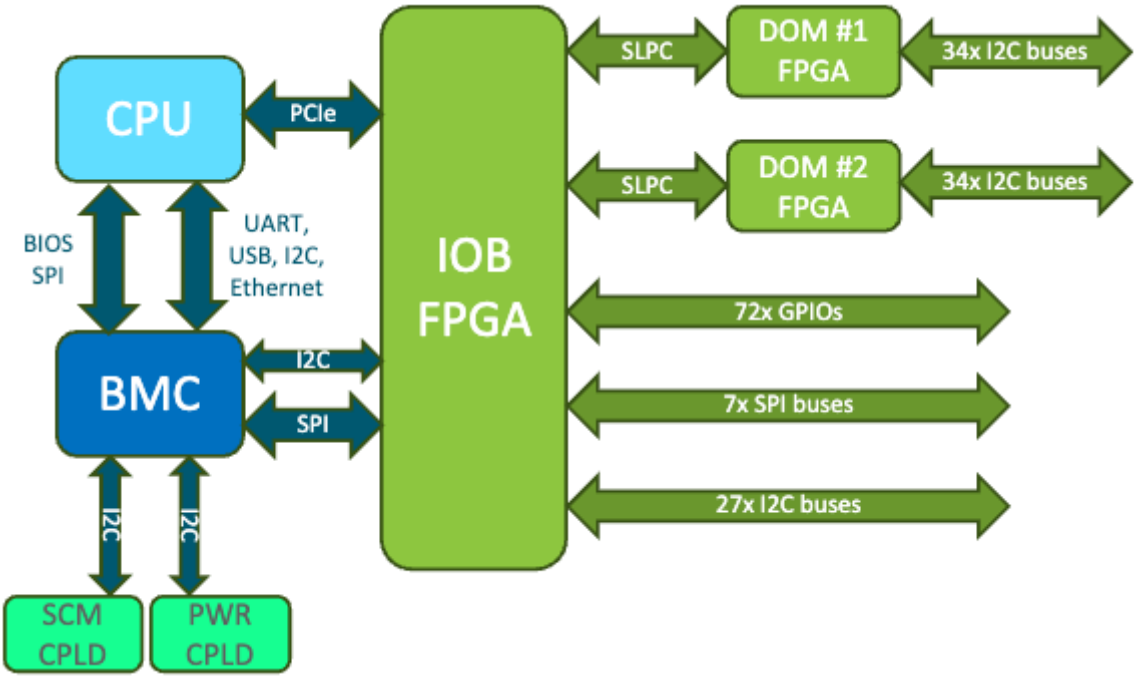


Figure 7: BMC-lite architecture

3.15.1 UART

The system has one RS232 port connected to the BMC directly. CPU UART0 connects to BMC UART5 through the SCM CPLD for SOL application. The system also has one UART port connected to the SCM CPLD. The SCM CPLD can switch between CPU UART0 and BMC UART2 for debugging purposes at the UART port. The UART connections are illustrated in the following diagram.

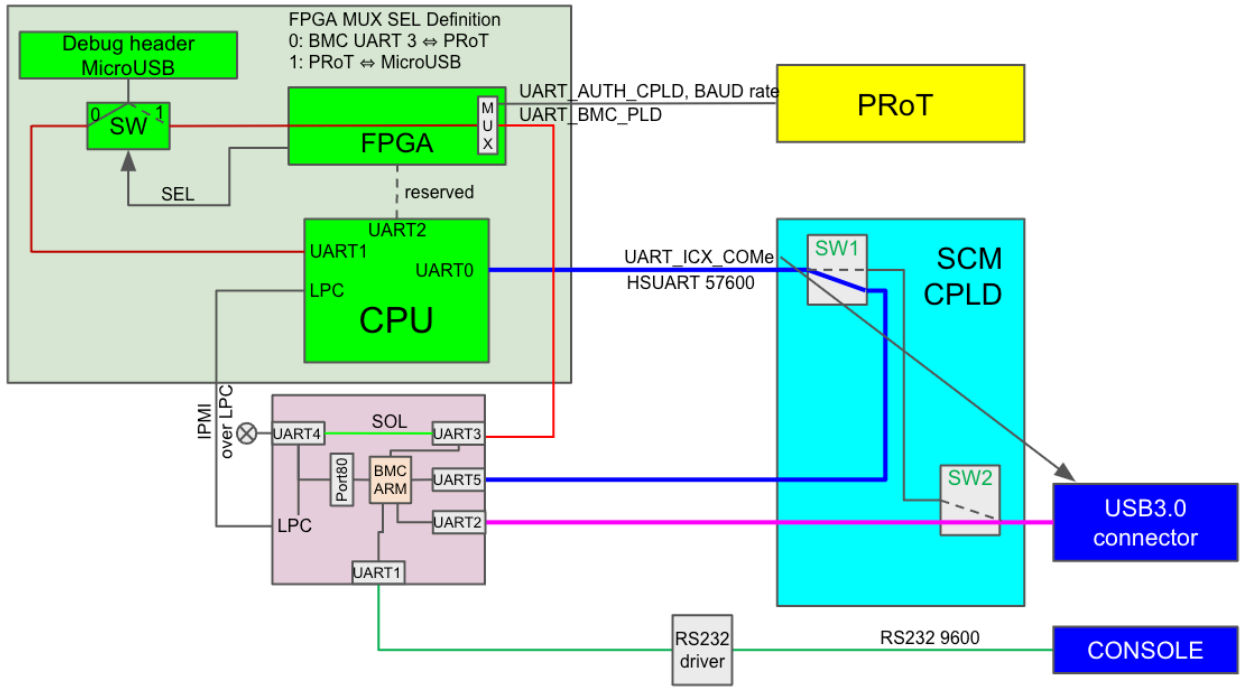


Figure 8: IceCube UART

3.15.2 OOB Ethernet

The system has one OOB management accessible through the 1000Base-T RJ45 port. As shown in the following diagram, the 1000Base-T RJ45 in blue is connected to the OOB GbE switch. The RJ45 in red is reserved in the existing design. The switch has one RGMII port connected to the BMC, and one SGMII port connected to the CPU through I210S.

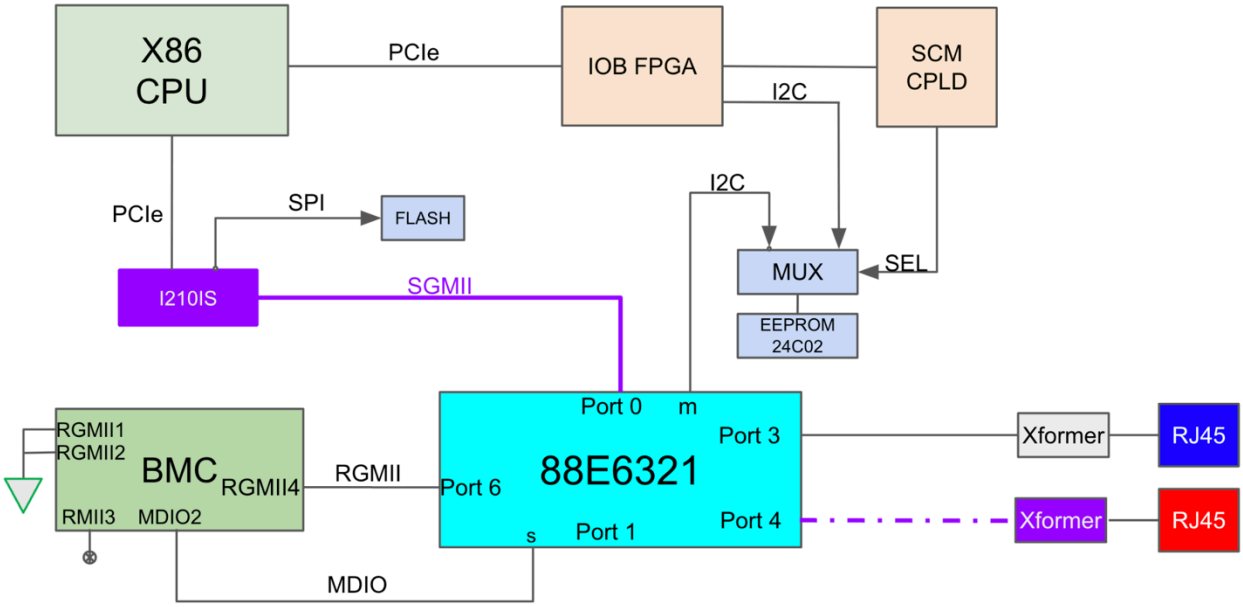


Figure 9: IceCube OOB management Ethernet

The system has another OOB Ethernet at the QSFP28 port. It can be configured as either 50GbE-CR2 or 100GbE-CR4. This port is connected to the ASIC Ethernet management port.

3.15.3 I2C bus tree

IOB FPGA is at the heart of the I2C tree in IceCube architecture. It translates the PCIe bus into numerous I2C buses directly attached to the IOB FPGA. It alleviates the I2C management workload from the BMC in the BMC-Lite architecture. An example of IceCube I2C tree is illustrated below.

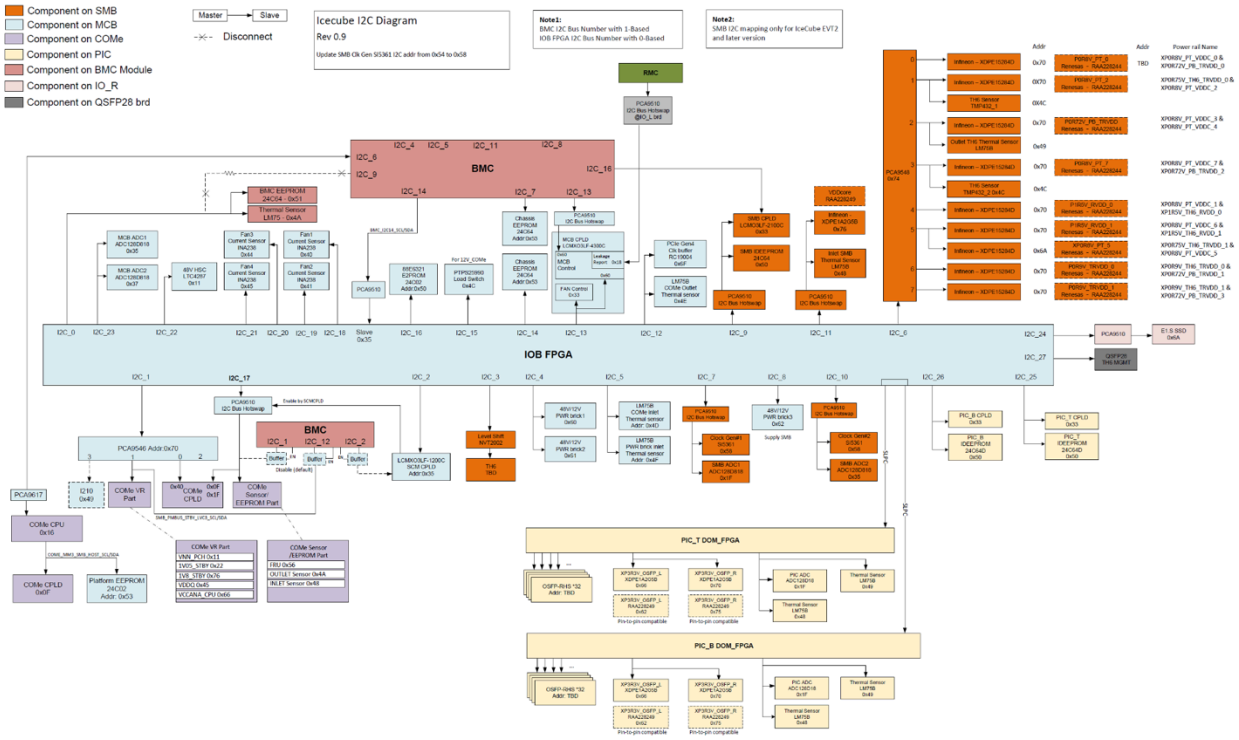


Figure 10: IceCube I2C tree

The OSFP I2C bus numbers are defined in the following table. Each OSFP I2C bus has a dedicated I2C controller in the DOM FPGA, which allows independent and parallel OSFP module I2C access.

User Port	E1	E5	E9	E13	E17	E21	E25	E29	E33	E37	E41	E45	E49	E53	E57	E61
I2C	DOM 2.0	DOM 2.2	DOM 2.4	DOM 2.6	DOM 2.8	DOM 2.10	DOM 2.12	DOM 2.14	DOM 2.16	DOM 2.18	DOM 2.20	DOM 2.22	DOM 2.24	DOM 2.26	DOM 2.28	DOM 2.30
User	E2	E6	E10	E14	E18	E22	E26	E30	E34	E38	E42	E46	E50	E54	E58	E62

Port																
I2C	DOM 2.1	DOM 2.3	DOM 2.5	DOM 2.7	DOM 2.9	DOM 2.11	DOM 2.13	DOM 2.15	DOM 2.17	DOM 2.19	DOM 2.21	DOM 2.23	DOM 2.25	DOM 2.27	DOM 2.29	DOM 2.31
User Port	E3	E7	E11	E15	E19	E23	E27	E31	E35	E39	E43	E47	E51	E55	E59	E63
I2C	DOM 1.0	DOM 1.2	DOM 1.4	DOM 1.6	DOM 1.8	DOM 1.10	DOM 1.12	DOM 1.14	DOM 1.16	DOM 1.18	DOM 1.20	DOM 1.22	DOM 1.24	DOM 1.26	DOM 1.28	DOM 1.30
User Port	E4	E8	E12	E16	E20	E24	E28	E32	E36	E40	E44	E48	E52	E56	E60	E64
I2C	DOM 1.1	DOM 1.3	DOM 1.5	DOM 1.7	DOM 1.9	DOM 1.11	DOM 1.13	DOM 1.15	DOM 1.17	DOM 1.19	DOM 1.21	DOM 1.23	DOM 1.25	DOM 1.27	DOM 1.29	DOM 1.31

Table 9: OSFP port and DOM I2C channel mapping

3.15.4 SPI bus tree

IOB FPGA offloads the SPI bus management from the BMC too. As illustrated in the following diagram, IOB has direct SPI bus access to the SPI devices in the system. CPU can manage these devices through IOB FPGA over the PCIe bus.

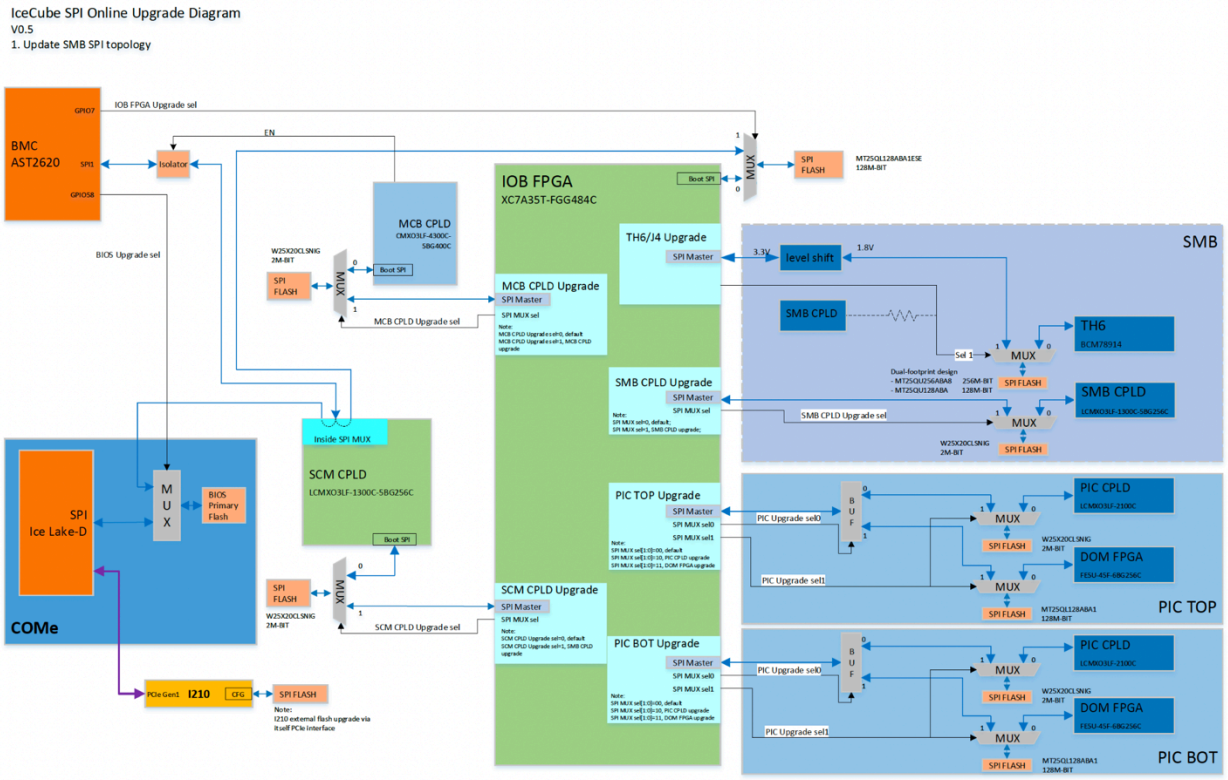


Figure 11: IceCube SPI tree

3.15.5 PIC (Physical Interface Card)

The system has two separate PIC cards to interface with the OSFP pluggable modules.

- Each PIC carries 32x OSFPs and one DOM FPGA
- There are two PICs in the IceCube

3.15.6 Sensors

There are a lot of sensors in the system to monitor the system health status. They are divided into three major categories: VRM sensors, temperature sensors and voltage sensors. Fan control algorithm running on the CPU monitors the temperature sensors all the time to adjust the fan speed in the system. VRM sensors and voltage sensors are being monitored to help the field support team in troubleshooting. Most of these sensors are connected to the FPGA and the CPLD devices directly.

Leak detection circuit shall be implemented to monitor liquid health status.

- Once a “faulty leak” event is detected, it shall be reported to the local BMC. No power-off or coolant shut-off should be executed upon “faulty leak” events.
- Once a “tray leak” is detected, it shall be reported to the local BMC immediately. The local BMC shall shut down its main power, leaving the standby power alive. The incident shall be also reported to the external Rack Management Controller (RMC) through the

front panel RJ45 GPIO LSC port. Upon receiving the leak report, the RMC shall shut off the coolant supply to the leaked blade. The leak shall be cleared by facility engineers eventually before powering up the blade again.

3.15.7 Fan PWM and RPM

There are four fan trays in the system. Fans are managed through a virtual fan CPLD in the physical MCB CPLD. The fan PWM frequency is around 25KHz. The fan RPM is monitored by fan CPLD and reported to the microserver through CPLD registers. If the fan RPM is too low, fan CPLD will report a warning in its register.

3.15.8 Platform Management

To mitigate the supply chain risk, the design has enabled multi-source parts in the hardware design. Although these multi-source parts are interchangeable from the hardware perspective, their drivers sometimes are not compatible with each other. To allow the software to automatically load the proper drivers, the system incorporates an IDEEPROM for each PCBA that requires drivers. The IDEEPROM contents define the driver information for each PCBA. The location and the address of the IDEEPROM are indexed in the platform EEPROM. FBOSS reads the platform EEPROM to create a map to access the IDEEPROMs in the system to retrieve the BSP driver information, and load the drivers accordingly. This mechanism is named as Platform Manager at Meta FBOSS.

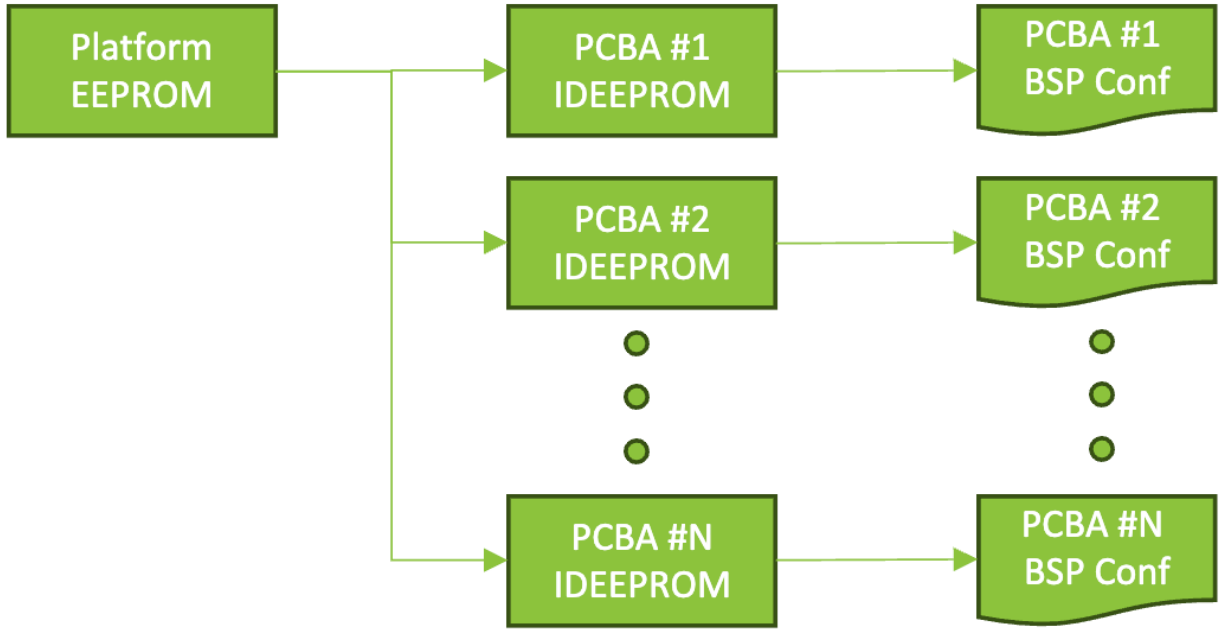


Figure 12: IceCube Platform Manager

The following diagram shows all the EEPROMs used in IceCube to self-identify the driver types in the system.

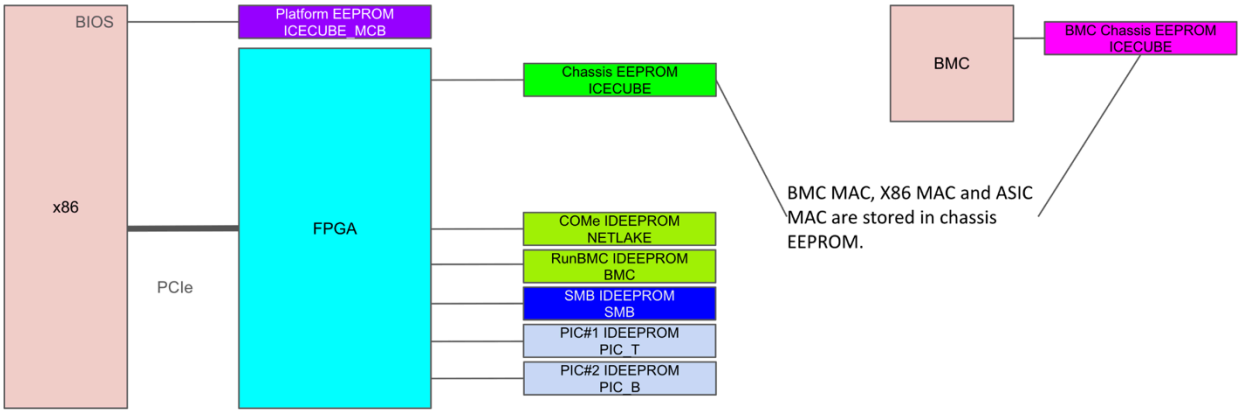


Figure 13: IceCube inventory EEPROMs

Platform EEPROM is accessible by CPU BIOS. Its content can be accessible at Linux through dmidecode command.

Chassis EEPROM contains the PRODUCT NAME information. It stores MAC addresses for BMC, CPU and ASIC ports.

Board IDEEPROM defines the list of the devices that need specific drivers. It is used by PlatformManager.

- COMe IDEEPROM is the board IDEEPROM for COMe PCBA.
- SMB IDEEPROM the board IDEEPROM for SMB PCBA.
- PIC_T IDEEPROM the board IDEEPROM for PIC top PCBA.
- PIC_B IDEEPROM the board IDEEPROM for PIC bottom PCBA.

3.15.9 LEDs

There are 64 user port tri-color LEDs on the front panel. Each user port LED illustrates one OSFP user port link status. For example, LED 1 is for OSFP port 1, LED 2 is for OPFP port 2, etc. Since each OSFP contains two subports to support 2x800G DR4/FR4 on IceCube, the user port LED toggle modes between subport 1 and subport 2, which is controlled by the LED SEL button.

When the LED SEL button is pressed, the user port LED mode travels from the default auto mode to subport 1 mode, then from subport 1 mode to subport 2 mode, and then from subport 2 mode back to the default auto mode.

- By default in the auto mode, the LED status of each OSFP port toggles between its subport 1 and subport 2 every 3 seconds. If the LED SEL LED is blue, the user port LED is for subport 1. If the LED SEL LED is green, the user port LED is for subport 2.
- In the support 1 mode, the LED SEL LED is blue. All the user port LEDs show the subport 1 status.

- In the subport 2 mode, the LED SEL LED is green. All the user port LEDs show the subport 2 status.

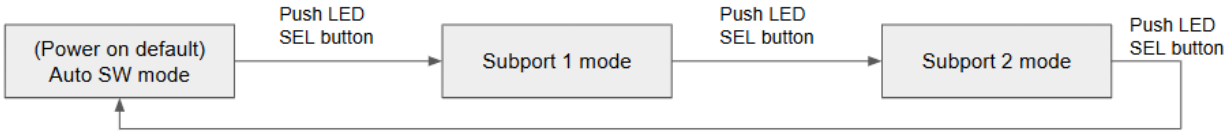


Figure 14: IceCube user port LED modes

SMB FRU LED is a tri-color LED driven by the MCB CPLD.

QSFP28 port 65 LED is a tri-color LED driven by DOM FPGA.

SCM FRU LED, SYS, PWR, SMB fan FAN LEDs are driven by SCM CPLD.

3.16. Constraints

n/a.

3.17. Test and Validation

The IceCube design shall be qualified through EDVT (electrical design verification test), MDVT (mechanical design verification test), environmental reliability test, EMC, safety, RDT (reliability demonstration test), HALT (Highly Accelerated Life Test).

3.18. External Software/Firmware Support (*Strongly* recommended)

Meta is running FBOSS SW (open-sourced) along with FBOSS platform services (also open-sourced) on IceCube microserver CPU. On the BMC side, BMC image for IceCube (open-sourced) can be used; which supports all standard features that other Meta switch systems support. Note that IceCube uses the “BMC-Lite” architecture, where most of the platform management features have been moved from BMC to microserver CPU, with the help of platform controllers (FPGAs and CPLDs.) Anyone that wants to develop SW stack on IceCube is strongly recommended to start with the open-sourced FBOSS Software stack and Meta OpenBMC software source tree.

IceCube has the following FW items that can be in-field upgraded.

- BIOS
- MCB CPLD
  - Manage system power on/off sequence, and fan tray management.
- SCM CPLD
  - Manage power on/off sequence, reset controls, UART selection etc management on the SCM.
- SMB CPLD
  - Manage power on/off sequence, reset controls on the SMB.
- PIC CPLD

    - Manage OSFP present/interrupt status management, OSFP LED data streams.
- IOB FPGA
    - IO bridge FPGA to translate PCIe commands to peripheral bus protocols, such as I2C, SPI and SLPC.
- DOM FPGA
    - FPGA to translate SLPC commands to I2C protocols for optics module management.

All these firmwares are stored in Flash devices. All the CPLD devices have golden images that will not be in-field upgraded. Only the application images stored in the external SPI Flash devices can be in-field upgraded. They all can be upgraded by the microserver CPU. BIOS and IOB FPGA can be in-field upgraded by BMC too.

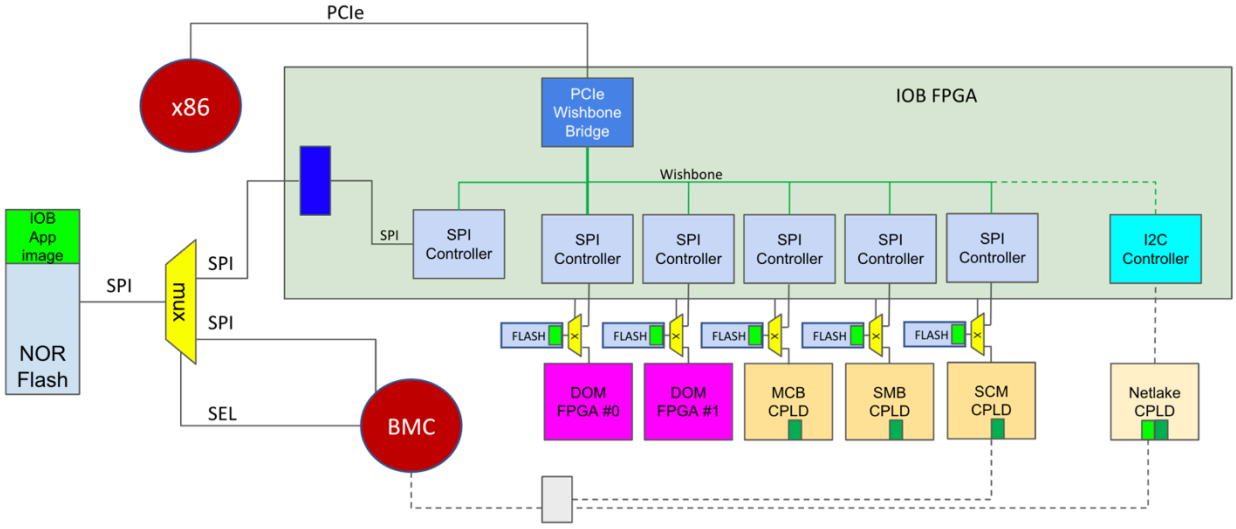


Figure 15: IceCube FW device topology

3.19. Compliance

ITEM	REQUIREMENT	REFERENCE	MANDATORY
1	Chassis shall fit in the compact 2OU form factor in an ORv3 HPRv1 rack.	3.8	Y
2	Chassis shall support 64x OSFP-RHS optic modules in 2x800G-DR4/FR4/LR/ZR.	3.12	Y
3	A QSFP28 port on the front panel connects to the ASIC Ethernet management port	3.8	Y

4	Chassis shall support 512x200GE non-blocking switch capacity.	3.13	Y
5	Chassis shall support hot swap feature.	3.5	Y
6	Chassis shall have management Ethernet port and console port for OOB management.	3.12	Y
7	FW shall support in-field upgrade.	3.18	Y
8	Chassis shall operate at 0-35C at sea level, or up to 30C 6,000ft	3.10	Y
9	Chassis shall consume less than 3,820W for Meta's typical use cases.	3.9	Y
10	Chassis shall meet all the EMC and safety compliance requirements.	3.4	Y

Table 10: Compliance Checklist

3.20. Arm SystemReady (only for Arm-based Systems) Requirement

n/a

3.21. Repository Location

n/a

3.22. Prescribed Materials

n/a

3.23. References (recommended)

n/a