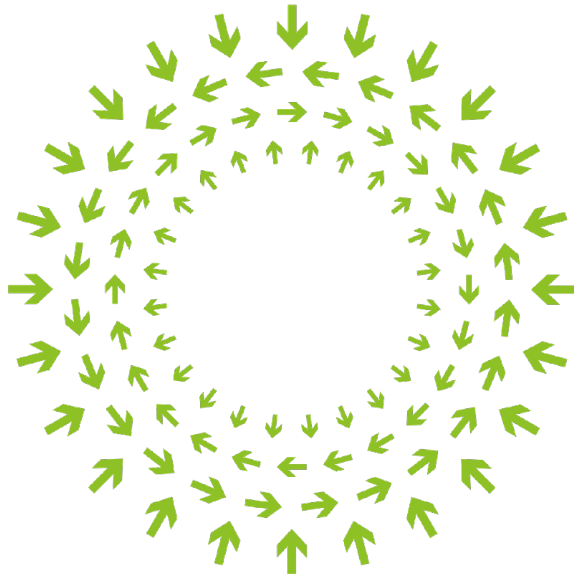




OPEN

Compute Project

100/400G Large EMUX Router Specifications

Revision 1.0

Version 1.0

Base Specification

Effective June 30, 2023

Author: Rehan Karim, Director Whitebox Design & Engineering - AT&T Labs

Author: Douglas Olsen, Principal Member of Technical Staff - AT&T Labs

Author: Shivakumar Mysore, Lead Member of Technical Staff - AT&T Labs

Table of Contents

1. License	5
1.1. OPTION A: OCP CLA	5
1.2 Acknowledgements	6
2. Compliance with OCP Tenets	7
2.1. Openness	7
2.2. Efficiency	7
2.3. Impact	7
2.4. Scale	7
2.5. Sustainability	8
3. Change Log	8
4. Scope	9
5. Overview	10
5.1 Network Topology Diagram	10
5.2 Large EMUX Key Requirements	10
5.3 Large EMUX High Level Block Diagram	11
6. Rack Compatibility	12
7. Physical Specifications	13
7.1 Physical Design Constraints	13
7.2 Common Design Concepts	13
7.3 X86 Considerations	14
7.3.1 Trusted Platform Module	14
7.3.2 Coin-Cell Lithium Battery	14
7.4 Network Timing Module Implementation	15
8. Thermal Design Requirements	15
8.1 Thermal Shutdown	15
9. I/O System	15
9.1 Craft/Management Interfaces	15
9.2 Network Interfaces	16
10. Rear Side Power, I/O, Expansion Board and Midplane Subsystems	17
10.1 Fan Module Specifications	17
10.2. Power Supply Specifications	17

11. Intel® Virtual RAID on CPU (Intel® VROC) for Storage	18
11.1. Description	18
11.2. Software and Firmware Components	19
11.3. Configuration Options	19
11.4. RAID Volume States	21
11.5. System Requirements	21
11.6. Error Handling Details	23
12. Mechanical	24
12.1. Recessed Reset Button Behavior	24
12.2. Silk Screen implementation	24
12.3. LED Operations Recommendations	24
12.4. Port Numbering Specifications	25
13. Motherboard Power System	26
13.1. Watchdog Implementation	26
13.2. CPLD logic and Controls	26
13.3. Dying Gasp Guidance	26
13.4. Supported Optics	26
13.5. Number of MAC addresses and Address Constraints	27
13.6. HW BMC Specification	27
13.7. Detection of insertion and removal of optical/DAC pluggable modules	28
13.8. Resets	28
13.9. ONIE	28
13.10. BMC Software	28
14. Environmental Regulations and Requirements	29
15. Prescribed Materials	29
16. System Firmware	29
16.1. Firmware upgrade process	30
17. Hardware Management	30
18. Security	30
19. References (recommended)	30
20. Tables and Figures	31

Appendix A - Checklist for IC approval of this Specification (to be completed by contributor(s) of this Spec)	32
Appendix B - Contribution Process FAQs	33

1. License

1.1. OPTION A: OCP CLA

Contributions to this Specification are made under the terms and conditions set forth in Open Compute Project Contribution License Agreement (“OCP CLA”) (“Contribution License”) by:

AT&T

You can review the Contributor License(s) for this Specification on the OCP website at <https://www.opencompute.org/legal-documents>. For actual executed copies of either agreement, please contact OCP directly.

Usage of this Specification is governed by the terms and conditions set forth in: **Open Compute Project Hardware License – Permissive (“OCPHL Permissive”)** also known as a “Specification License”.

Notes:

- 1) 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.

1.2 Acknowledgements

The Contributors of this Specification would like to acknowledge the following companies for their feedback: **UfiSpace, Intel, Broadcom**

List all companies or individuals who may have assisted you with the specification by providing feedback and suggestions but did not provide any IP.

2. Compliance with OCP Tenets

This Specification complies to the following OCP tenets.

A full explanation of the OCP core tenets can be seen [here](#).

2.1. Openness

The goal of the Large EMUX specification is to encourage openness and maximize reuse of common hardware through disaggregation of hardware from software. This specification will try to keep the requirements to a set that can be used in a variety of use cases and environments. Open hardware specifications enable designs from multiple ODMs and software implementations from different NOS vendors.

2.2. Efficiency

The Large EMUX is built on a modular unitized white box design that can be put together to address larger scale-out network configurations. By enabling central hub architecture, new access points can be built out more efficiently from the hub, saving on associated costs and improving the network's scalability. The convergence of multiple access traffic also makes it easier to implement virtualized and software defined networking. Improved Multi updater to perform upgrades of various software components such as BIOS, BMC, CPLD, SSD on both CI and KGI will reduce the operator error and upgrade time. Embedded Diagnostic tool will aid in isolating ODM HW issues v/s NOS issues. Eye Pattern response without NOS will help in troubleshooting of Optics in production.

2.3. Impact

Disaggregation of hardware and software enables service providers to reuse the exact same hardware with a different NOS vendor to meet the requirements of different use cases. Building platforms on merchant silicon allows for a larger ecosystem of vendors to build the hardware and develop innovative solutions that can benefit service providers. Monopoly of one ODM and /or NOS vendors reduces the cost and bring healthy competition.

2.4. Scale

The Large EMUX come with remote management features that allow service providers to deploy large scale open network infrastructures for various network applications. The firmware/software inside the box are 100% remote upgradeable, easy to maintain and can be upgraded rapidly. Furthermore, the Large EMUX is capable of being deployed using a spine and leaf architecture, which enables large scale and rapid deployment.

2.5. Sustainability

The results of our designs reduce the amount of space, power and cooling needed to not only sustain, but expand their network services.

3. Change Log

Date	Version #	Author	Description
6/27/2023	1.0	Shivakumar Mysore, AT&T Douglas Olsen, AT&T Ken Chen, UfiSpace Andrew Lui, UfiSpace	Initial draft

4. Scope

The purpose of this document is to define the high-level technical specifications for the Large EMUX 100G/400G platform that AT&T is submitting to the Open Compute Project. The specifications list key requirements and constraints that the design must meet but leaves room for innovation for the hardware manufacturer in the design/implementation/manufacturing process.

5. Overview

The data rates in packet processing silicon have outpaced the hardware chassis face plate port density since 2010. The industry standard architecture is to use low-cost merchant silicon platforms to efficiently pre-aggregate traffic (EMUX) into high-speed PE router ports and initiate a centralized hub for both fixed and mobile traffic.

As service providers continue to build out next generation L2 and L3 networks network based on merchant silicon white box hardware the Ethernet transmission speeds continue to increase resulting in the need for higher rate ethernet multiplexing from 100GE to 400GE, this triggers the need for Next Gen EMUX Platform 100G/400G.

5.1 Network Topology Diagram

The figure below is the topology diagram for the Large EMUX.

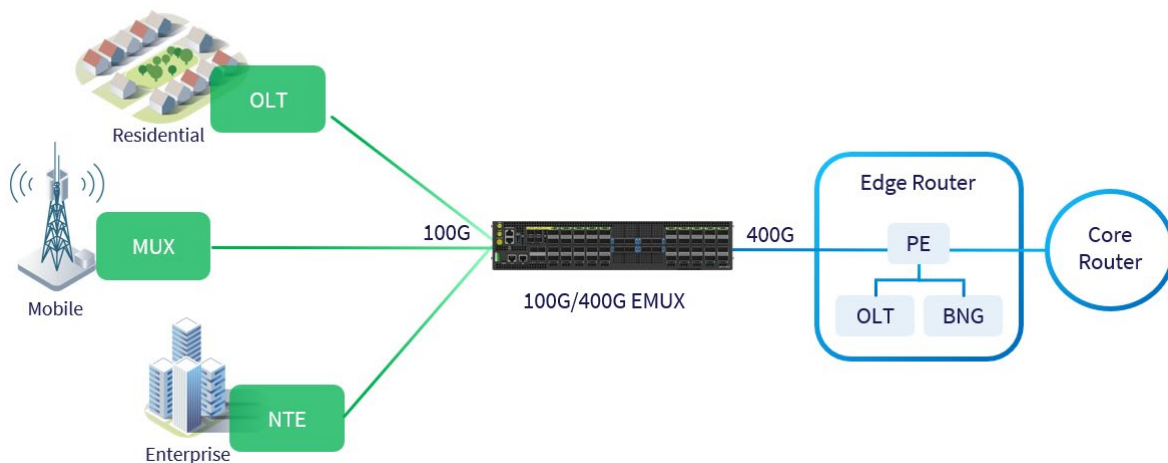


Figure 1 Large EMUX Network Topology Diagram

5.2 Large EMUX Key Requirements

- Class-C NTM module with Stratum 3E OCXO to support Class-C timing is required for this design.
- This box should be 2RU in physical design. The option depends on how much ZR/ZR+ optics need to be supported in this box.
- Port-0 100GE should have MUX configurable for native 4x 10GE/25GE SFP28 and support for SFP-DD.
- For this specification, a single Large EMUX need to support 100% of the QSFP-DD ports with ZR/ZR+ optics.
- Needs to support GNSS Receiver and T-GM function.

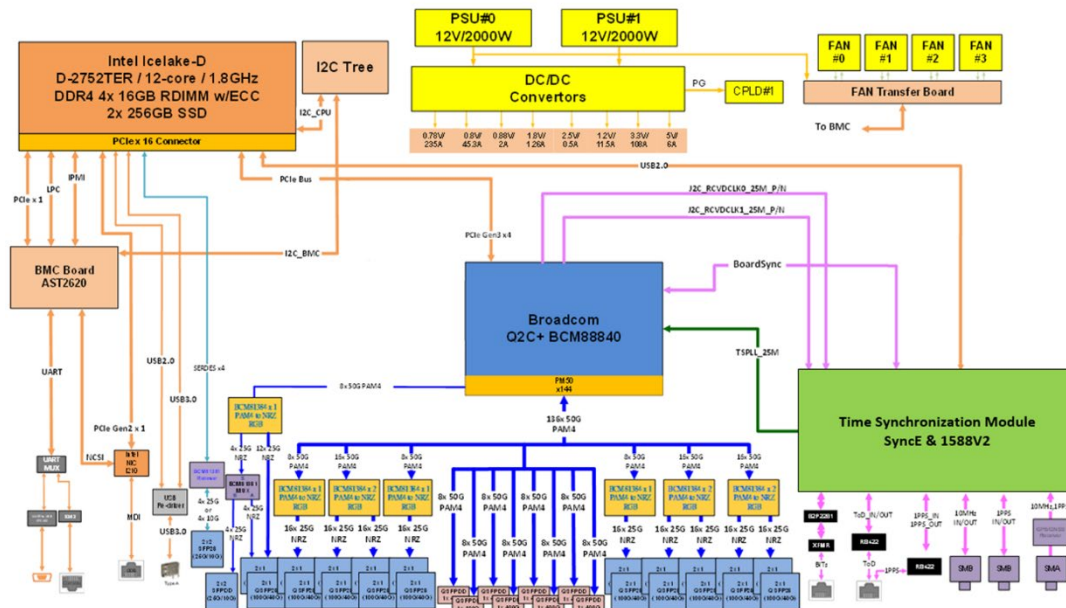
- 10Mhz, 1PPS, BITS and TOD support are also required for the design for some use cases that may require them.
- Implementation of future generations of NPUs is acceptable and it is understood that network features may change due to future developments.

Feature	Description
Network	40x100GE QSFP28 8x400GE QSFP-DD
Craft	RJ45 serial console Micro USB serial console USB 3.0 Type-A
Network management	1x1G RJ45 Mgt 4x10G SFP+ Mgt
Timing	SMA GNSS Input SMB 1PPS In/Out SMB 10MHz In/Out RJ45 ToD & BITS Input
NPU	BRCM Q2c+ BCM88840
CPU	Daughter Board: Intel Icelake-D Generation or later with VROC enablement for various RAID options to support NVME SSD or SATA RAID
BMC	Aspeed AST2620 with Redfish support
Synchronization	Stratum 3E OCXO, SyncE IEEE 1588v2, T-TC, T-BC/OC, T-GM
LED	Power and System status LEDs per unit Link and Activity LEDs per port FAN Status LED PSU Status LED per PSU
Air Flow	Redundant Hot Swap Fan, Port to Power (F2B)
PSU	Redundant Hot Swap AC or DC

Table 1: Large EMUX Datasheet

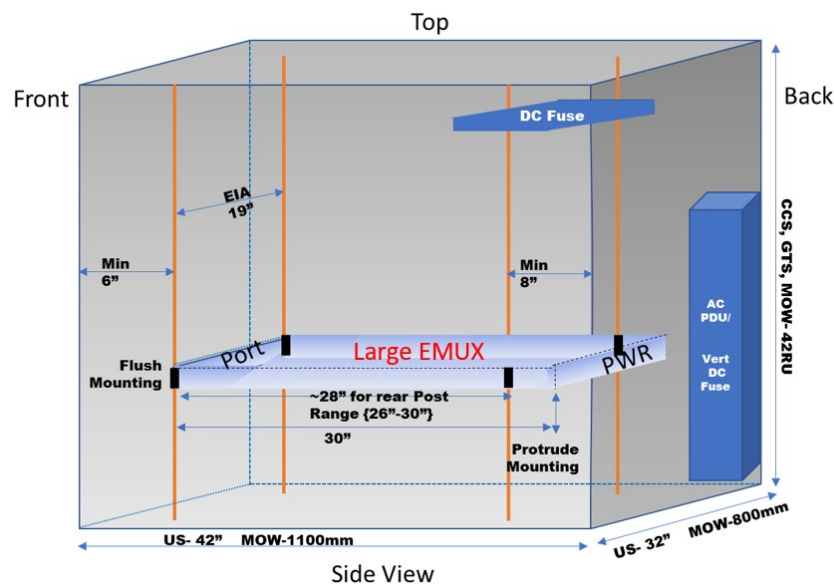
5.3 Large EMUX High Level Block Diagram

The figure below provides a high level block diagram for the Large EMUX.



6. Rack Compatibility

The white boxes specified in this need to comply to the following:



This cabinet dimensions are used in the three environmental spaces {CCS, GTS, MOW} that is targeted for these use cases. The figure illustrates the typical cabinet and minimum mounting clearances required to allow for cabling and power cables and PDUs. For example, the front rails need to be located minimum 6 inches back from the front of the cabinet. For this use case, it may require 7 inches to allow for sufficient fiber bend radius or fiber densities for example. The rear mounting rails need to have 8 inches minimum. For a 30" deep piece of equipment, this would mean the front mounting ears are flushed and the rear mounting rails needs to be able to adjust to 27 inches $\{= 42-7-8\}$. That is why the rear mounting rails need to be adjustable from $\{26" - 30"\}$.

7. Physical Specifications

7.1 Physical Design Constraints

The objective of the designs for the Large EMUX components is to be able to mount them in the AT&T standard cabinet which are EIA- 19", 42" deep and 42RU high.

- Width: 19" rack mount EIA cabinet standard with 4 post mounting.
- Depth: Maximum 30" depth to allow for Cabling and Power clearance in a 42" deep cabinet.
- Height: 2RU
- Air Flow: Front to back.
- Access: Front Access for fibers and cables. Rear Access for Power and Fan FRUs.
- Temperature: Range $\{0C \text{ to } + 45C\}$ Ambient. Typical 26C.
- Environmental Spaces: AT&T -{CCS, GTS, MOW Tele-Houses}

7.2 Common Design Concepts

The white box design should adopt modularity and reuse where possible to achieve scale and minimize sparing complexity. The common components in the white box are the following:

- PSU FRU
- Fan FRU
- The Craft, Management, and Console interfaces
- BMC and Redfish
- X86 host CPU enabled with Intel VROC HW License key for NVME RAID
- Network Timing Module

7.3 X86 Considerations

For the Large EMUX Specification, the x86 CPU needs to meet the following design requirements:

1. Needs to be designed as a modular daughter board that can be reused in multiple white box configurations where CPU cores, DRAM, SSD are factory installable options.
2. Needs to be designed with sufficient signals to be connected to the Main switch board that will allow it to accommodate a couple generations of CPUs. (Current is Icelake-D)
3. Enough signals between CPU daughter board and main switch board to accommodate "PUNT-Path" from MAC to CPU directly.
4. Needs to be designed to accommodate a range of CPU sizes from 4 cores up to max 16 cores from minimal memory to max DRAM memory.
5. CPU daughter board needs to support RAID-0 SSD.
6. Redundant BIOS Flash
7. TPM 2.0
8. Operation without Coin-Cell Battery for TELCO and with Coin-Cell as other providers deem fit.
9. All NIC must be supported by DPDK (refer to <http://dpdk.org/doc/nics>).
10. VROC function reduces the need to separate HW modules for HW RAID functionality.

7.3.1 Trusted Platform Module

The latest Trusted Platform Module (2.0 or greater) shall be used for secure storage of keys and certificates in a hardware chip and is an integral part of creating a Secure Boot environment so that the device cannot be easily taken over, such as by booting from a USB drive.

This is a future proof design specification because it is dependent on the availability of ONIE secure boot. The initial software release will operate with TPM disabled in BIOS and use regular ONIE Boot process.

7.3.2 Coin-Cell Lithium Battery

Typical X86 design specifies the use of a Coin-Cell Lithium battery to maintain the RTC. However, the presence of this battery conflicts with the TP76200/TP76450 requirements. As such, the Large EMUX design does not have a Coin-Cell battery.

7.4 Network Timing Module Implementation

Network timing module is used to enable support for ITU-T Synchronous Ethernet (SyncE) IEEE 1588v2 (Default Profile, G.8265.1 G8275.1, G.8275.2 profiles), T-TC, T-BC/OC, and T-GM. The minimum timing accuracy should be Class C as per ITU-T Recommendation G.8273.2.

8. Thermal Design Requirements

The components used in the Large EMUX needs to meet the following requirements.

- AT&T TP76200 & TP76450 for Level 3.
- Copies of this document and general information about AT&T's environmental equipment standards can be found at <https://ebiznet.sbc.com/sbcnebs/>

8.1 Thermal Shutdown

NEBS/TP76200 compliant equipment should have the ability to be configured to shut down when the thermal threshold is exceeded or continue to operate until the equipment fails completely. Configurable means that the “user” can select the thermal overload behavior. This must be set through software. The default should always be to implement equipment shutdown in a thermal event.

Shutdown means that all non-essential functions of the chassis are powered off and only temperature monitoring capability remain such that, if the thermal event ends, the chassis will autonomously reboot and restore service. One way of accomplishing this is to have the management hardware command the power supplies to shut off their main outputs but maintain an auxiliary power bus that powers the management/monitoring functionality.

9. I/O System

9.1 Craft/Management Interfaces

The following table lists the craft and management interfaces that are needed on the Large EMUX devices.

Craft Type	QTY	Purpose
Micro USB Serial	1	Console
RJ-45 Serial	1	Console
USB 3.0 Port	1	USB data access

RJ-45 10M/100M/1G	1	Ethernet OOB Powered on Standby Power rail
SFP+ 10G	4	For management to CPU

Console:

Only one Serial input can be active for the Console. RJ45 RS232 is higher priority than Micro-USB. RJ45 RS232 will be active if both interfaces are connected by user. The Serial console needs to support default selectable between the BMC or the X86 CPU.

OOB Management:

The RJ-45 OOB Ethernet management port needs to be operational even when the system is in the shutdown mode. As such it needs to be designed using the standby power rail. It also needs to provide simultaneous connectivity to the X86 CPU and the BMC. The Intel I-210 NIC is specified to use for the RJ-45 OOB Ethernet Management. Design should allow to shutdown Ethernet OOB access to the BMC as needed.

USB 3.0 port:

USB port can be used to provide access to external USB drive for initial system setup or system rebuild. However, for security reasons, once the system is up and running under the control of the NOS, the hardware/firmware of the box needs to provide a mechanism to turn off access to this external USB port. This lock needs to persist even after a cold boot, warm boot, other reset mechanism. To unlock this would require authentication to access the utility to set the registers to unlock it.

SFP+ ports:

The four 10G SFP+ ports provide connections to the X86 Host CPU.

9.2 Network Interfaces

The following table summarizes the minimum support of third-party optical transceivers for the Large EMUX.

10/25G	100G	200G	400G
SFP28 AOC	QSFP28 AOC	QSFP28 AOC	QSFP-DD AOC
SFP28 SR	QSFP28 SR4	QSFP56 SR4	QSFP-DD SR8
SFP28 LR	QSFP28 CWDM4	QSFP56 FR4	QSFP-DD FR4
	QSFP28 LR4		QSFP-DD LR4
	QSFP28 ER4		QSFP-DD ER4
	QSFP28 ZR4		QSFP-DD ZR
	QSFP28 ZR		QSFP-DD OpenZR+

	QSFP28 DR1 QSFP28 FR1 QSFP28 LR1		QSFP-DD DR4 QSFP-DD DR4+ QSFP-DD PLR4
--	--	--	---

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

10.1 Fan Module Specifications

The Large EMUX environment will have rear access. Fans need to be redundant, field replaceable and hot swappable. Fans are accessible from rear panel.

10.2. Power Supply Specifications

- DC PSU accept nominal -48V DC Power.
- AC PSU operates from 100-240V, 50-60Hz input range.
- The power supplies shall meet the 80 Plus Gold or better for high efficiency.
- There shall be redundant, field replaceable, hot-swappable power supplies.
- The power supplies should support loss of power indicators to facilitate BMC
- Two grounding screws on rear panel of the system.

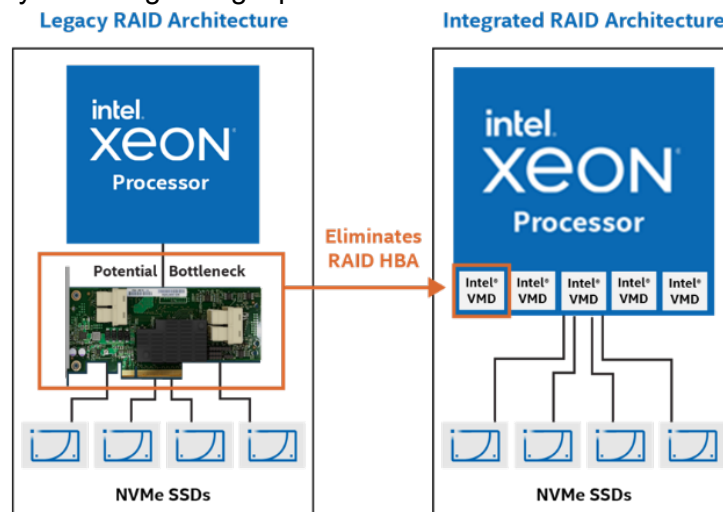
Table below shows the PSU efficiency under different loads for different 80-Plus standards.

	20% Load	50% Load	100% Load
80 Plus	80%	80%	80%
80 Plus Bronze	82%	85%	82%
80 Plus Silver	85%	88%	85%
80 Plus Gold	87%	89%	87%
80 Plus Platinum	90%	92%	89%

11. Intel® Virtual RAID on CPU (Intel® VROC) for Storage

11.1. Description

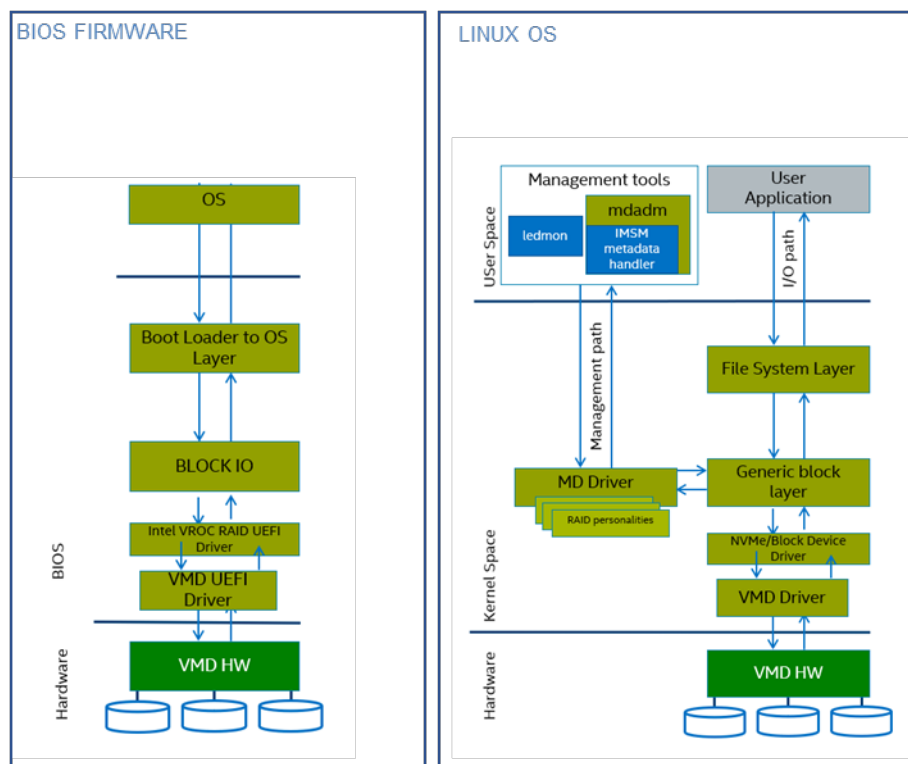
Intel® VROC is an integrated RAID solution on Intel® Xeon® Processors that provides redundancy, availability, and serviceability (RAS) for NVMe storage. Intel® VROC improves storage performance, lowers cost and improves power efficiency by eliminating the need of an external hardware RAID card. Other benefits are reduced complexity, space savings, and improved reliability by removing a single point of failure



Intel VROC supports RAID 0, 1, 5, and 10. RAID level 1 can be selected for Boot. For this application RAID level 1 is recommended for boot. RAID 1 requires 2 devices. RAID 1 boot provides a mirrored copy of the Operating System on each NVMe device. This feature allows for long term serviceability for storage. In the event one device fails to respond, or completely fails, the system continues with servicing IO on the remaining healthy device with an exact copy of the Operating System.

11.2. Software and Firmware Components

Intel VROC includes both Linux OS and BIOS firmware components. The BIOS firmware include the Intel VMD UEFI driver and the Intel UEFI VROC driver to manage NVMe device enumeration and RAID handling. For Linux, the Intel VROC driver uses open source VMD and mdadm with a proprietary metadata, along with its management components.



11.3. Configuration Options

RAID 1 Boot on 2 NVMe devices can be customizable in size. The RAID 1 partition can be large enough to service the Operating system, but leave space on the devices for other purposes, such as logging and caching, and can implement Data RAID level 0 for increased capacity and performance.

The below tables illustrate the possible use cases with Intel VROC using only 2 NVMe devices. RAID 0 is mentioned due to the increased capacity and performance.

Use Case #1

Exposed Block Device	redundant	Purpose
RAID 1 on partial size of NVMe capacity	yes	Boot and Operating system redundancy
Logging	no	Section of NVMe to store logging data
Caching	no	Section of NVMe to cache data

Use Case #2

Exposed Block Device	redundant	Purpose
RAID 1 using entire capacity of NVMe	yes	Boot and Operating system redundancy
RAID 1 Logging	yes	RAID 1 Partition to store logging data
RAID 1 Caching	yes	RAID 1 Partition to cache data

Use Case #3

Exposed Block Device	redundant	Purpose
RAID 1	yes	Boot and Operating system redundancy
RAID 0 Logging	no	2 nd RAID 0 on same 2 NVMe – striped for performance and increased capacity for logging data
RAID 0 Caching	no	2 nd RAID 0 on same 2 NVMe – striped for performance and increased capacity for caching data

11.4. RAID Volume States

RAID volume will be in a “normal” state at creation. If one device fails, the state of the RAID volume will show as “degraded” until a new device is plugged into the system. The degraded state can continue without service and does not affect functionality or performance but is an indicator that a replacement device is needed for a redundant state to be maintained.

RAID Volume State	Description
Clean	RAID volume healthy, data is in sync
Clean, resyncing	RAID volume is initializing
Clean, Degraded	RAID volume has no redundancy due to one failed device
Clean, degraded, recovering	RAID volume detects newly inserted device and data is being copied across both devices

11.5. System Requirements

Intel® Volume Management Device (Intel® VMD) is an HBA-Like Controller on the uncore of the Intel Xeon Processor to manage NVMe. Intel VMD is the foundation for Intel VROC, and must be enabled on the platform to support Intel VROC.

11.5.1. BIOS Components:

BIOS	Purpose
Intel VMD - enabled	Allows VMD to manage NVMe
Intel VROC UEFI Driver	Allows Bootable RAID
Intel VROC HII Menu	GUI inside BIOS to configure/manage RAID

11.5.2. System OS components:

Intel VROC Linux is up streamed into the Linux community. Mdadm is used with Intel Proprietary metadata that is written to each device at the time of RAID creation. This metadata is used for management within the Operating System and within the BIOS environment. The Intel VMD driver is another Linux kernel component helping to manage the NVMe devices.

Linux OS Components for Intel® Xeon Gen 4 (Ice Lake)

Component	Kernel version or Component version
VMD	Kernel version – starting in 4.11 through 5.17
mdadm	4.2
Linux Kernel	<same as above>

Intel VROC Link to list of supported Linux Operating systems and Configurations.

<https://www.intel.com/content/www/us/en/support/articles/000030310/memory-and-storage/datacenter-storage-solutions.html>

11.5.3. Remote Deployment and provisioning

Intel VROC UEFI Driver packages include tools for deployment and provisioning RAID outside of the OS and BIOS menu.

The following tools can be used remotely to configure RAID and can be scripted into .nsh files for provisioning the platform with RAID, as well as running validation testing.

Component	Purpose
RcfgVROC.efi	This tool must match the version of the UEFI drivers being used. Configures RAID, and can mimic failing scenarios for testing purposes. NSH scripts for remote deployment can be created, using this tool.
RcmpVROC.efi	Use same version as driver version. Debugging purposes to ensure system compliance.
HWKeyChecker.efi	Same version as driver version. Debugging purposes, to ensure correct Hardware RAID key is installed and found by UEFI VROC drivers.

11.6. Error Handling Details

Intel VROC creates a robust system that is protected from kernel Panics and system hangs are avoided. In a RAID 1 configuration, in the event a device becomes unresponsive, Intel VROC will handle the error, without disruption to service, and the operation will continue on the next healthy device.

Intel VMD as a foundation of Intel VROC handles Advanced Error Reporting (AER) and provides logging within Linux, therefore providing protection to the rest of the PCIe bus, preventing whole system shut down. In Linux, the PCIe Bus driver or hotplug driver handles AER and logging. Upon receiving an interrupt at MSI-X 0 for each root port, the Intel® VMD driver uses a sophisticated algorithm to determine the precise device to apply error handling based severity. The actions taken are specific to the error on the device, to preserve the remaining eco system. <add that it goes into syslog>

Intel VMD will log the following errors. If a non-fatal error is encountered, the Intel VMD driver will log the error and standard recovery efforts are taken. If a fatal error is encountered, Intel VMD will treat the error as a “device surprise hot removed”. But system shall continue to operate without hang, reboot, or crash.

Advanced Uncorrectable Error	Severity Register
Poisoned TLP	Non-Fatal (0b).
Completion Timeout	Non-Fatal (0b).
Completer Abort	Non-Fatal (0b).
Unexpected Completion	Non-Fatal (0b).
ECRC Error	Non-Fatal (0b).
Unsupported Request Error	Non-Fatal (0b).
ACS Violation	Non-Fatal (0b).
Atomic Op Egress Blocked	Non-Fatal (0b).
TLP Prefix Blocked	Non-Fatal (0b).
MC Blocked TLP	Non-Fatal (0b).

12. Mechanical

12.1. Recessed Reset Button Behavior

When the recessed reset button is depressed for less than equal to 10 seconds and released, then this should cause a warm reset of the whole white box. If >10 sec has no action.

12.2. Silk Screen implementation

It is a strong recommendation that the manufacturer choose Pantones, Contrast, and Font Size that MAXIMIZE visibility to the field technicians working in low light, tight spaces, and crowded cabling conditions.

12.3. LED Operations Recommendations

Refer to AT&T Hardware Common Systems Requirements for recommendations on system and interface LED colors and operations.

The indicator lamps (LEDs) must convey the information described in the table below. The number, colors, and flash behaviors are desired but not mandatory.

LED Name	Description	State
SYNC	LED to indicate System timing/clock synchronization status	Green - System timing/clock is synchronized from external timing/clock reference. Amber - System timing/clock synchronization is in free-run or holdover mode. Off –System timing/clock synchronization is disable
SYS	LED to indicate system diagnostic test results	Green – Normal Amber – Fault detected Off – No Power
FAN	LED to indicate the status of the system fans	Green – All fans work normal Amber Flashing – One or more fan fault Off – No Power
PS0	Led to indicate status of PSU0	Green - Normal Amber Flashing – Fault detected Off – No Power

PS1	Led to indicate status of PSU1	Green - Normal Amber Flashing – Fault detected Off – No Power
ID	LED to indicate Location of switch in Central Office	Blue Flashing – Set by management to locate switch Slow flashing – System is in standby state Off – Function not active
OOB LED	LED to indicate link status of 10/100/1000 RJ45 management port	On /Flashing – Port up (flashing indicates activity) Green – Highest Supported Speed (1G) Amber – Lower Supported Speed (100M/10M) Off – No Link/Port down
SFP(28)- LEDS	Led to indicate status of SFP28	Green –Highest Supported Speed (25G) Amber –Lower Supported Speed (10G) On /Flashing – Port up (flashing indicates activity) Off – No Link/Port down
QSFP(28) LEDS	Led to indicate status of QSFP28	Four Green On/Flashing – 100G speed and individual 25G breakout port up (flashing indicates activity). Four Amber On/Flashing – 40G speed and individual 10G breakout port up (flashing indicates activity) Off – No Link/Port down
QSFP-DD LEDS	Led to indicate status of QSFP-DD	Green –Highest Supported Speed (400G) Amber –Lower Supported Speed (100G) On /Flashing – Port up (flashing indicates activity) Off – No Link/Port down

Table 2: LED Indicators

12.4. Port Numbering Specifications

AT&T numbering standard for white boxes starts from zero {0,1,2,...}, increasing from Left to Right. This applies to Ports, FAN and PSU and other FRU (Field Replaceable Units). Numbering starting from Zero also applies to break out ports which is more dependent upon software implementation as opposed to hardware and silk-screening implementations.

Manufacturer has a degree of freedom for the numbering with respect to vertical grouping. For examples, the following schemes are acceptable.

NOTE: The port numbering illustration shown in these tables does not reflect the actual number ports specified for the Large EMUX. It is just to illustrate the acceptable numbering scheme.

0	2	4	6	8	10	12	14	16	18
1	3	5	7	9	11	13	15	17	19

20	22	24	26	28	30	32	34	36	38
21	23	25	27	29	31	33	35	37	39

Table 3: 2 grouping numbering schemes: Upper/lower port grouping. Sequentially Numbered Upper grouping then followed by Lower grouping.

13. Motherboard Power System

13.1. Watchdog Implementation

Design needs to include a watchdog mechanism prevent the system from being stuck in a “hanged” state. This can be done via one or combination of the following:

1. Dedicated watchdog circuitry.
2. Intel TCO watchdog
3. Or some kind of watchdog implementation between BMC and X86 host.

The choice of implementation is left to the manufacturer as long as this is documented and appropriate drivers or mechanisms to leverage this capability from the NOS is provided.

13.2. CPLD logic and Controls

The specification leaves freedom to manufacturer to use any combination of discreet logic/PLD/CPLD/FPGA necessary to implement of the specification. It is recommended to use current practice and provide I2C interface to the Host for control of PSU, FAN, Optics, and any other key components such as reading of registers or erasing and flashing of NVRAM, EEPROM, Flash, etc. The manufacturer must provide the instructions and drivers to access these components as part of the BSP- Baseboard Support Package, so that NOS development can access and control these components as necessary.

13.3 Dying Gasp Guidance

Dying Gasp is not a required feature for this application mainly because of the environment in which these boxes will be used. They are AT&T or Leased facilities with redundant power feeds and back-up power.

13.4. Supported Optics

The white boxes specified for Large EMUX should have 25G, 100G, and 400G interfaces. The white box design needs to be able to support optics from {DAC, AOC, SR, FR, LR, ER, ZR/ZR+}. For the {ZR, ZR+} support the design need to support power and cooling for up to

100% optics operating at {400G ZR/ZR+}. This is the minimum specification. Full PMD list needed here

13.5. Number of MAC addresses and Address Constraints

Each Large EMUX should reserve a minimum block of 1024 MAC addresses.

13.6. HW BMC Specification

The system will be designed with Baseboard Management Controller (BMC) to allow for remote lights out operations, management, and access.

The most important requirement for the BMC is that it must be secure. The BMC will be connected to WAN and/or Internet connections. As shown in the System Block Diagram, the BMC has an Ethernet connection which is a shared external connection to RJ45 OOB Management Ethernet Port with the X86 host.

The long-term goal is to use Open BMC when Open BMC supports the required features needed for the operational/business mode in a secure way. In the interim, a commercial BMC implementation is acceptable. Regardless of the open or commercial implementation, the firmware must be capable of disabling this Ethernet access when needed based upon the use case or operating environment.

The following requirements are specified for the BMC.

- Dual Flash memory to support remote reliable in-band BMC Software upgrade.
- Power management: On/Off of control system, Host CPU, and MAC where it makes sense. That is, it does not make sense if a component is powered off which completely cuts off access to the BMC to power it back on. In this case, this component should always be powered on the standby power.
- Temperature monitoring
- Voltage monitoring
- Fan control
- Reset control
- Host CPU boot up status
- Serial number / unique identifier
- Board revision ID
- I2C interfaces to Host CPU, USB, temperature sensors, and voltage controllers.
- Monitoring detect signals – including loss of power from the power supplies.
- Must support Redfish Implementation.
- Must support IPMI 2.0 host mode to provide the following capability via the IPMI interface to allow NOS transition development:

- ✓ temperature reading and alarms at 3 levels (minor, major, critical) for Processor modules, Chassis, power supply, fans, Broadcom chipset.
- ✓ status information for fans, power supply, interface modules, processor modules, fan tray

13.7. Detection of insertion and removal of optical/DAC pluggable modules

The white box CPLD design should be able to detect insertion and removal of pluggable optical modules on the NIF ports and notify the x86 host with an interrupt and a vector code.

When defective SFP is detected, a quarantine mechanism is implemented to isolate the issue. Notification of defective SFP sent to NMS via NOS with the defective port identification and the maintain the port as a defective status until a new SFP has been replaced. After confirming the SFP is no longer defective, there should be a procedure to notify NMS to remove it from quarantine.

13.8. Resets

RESETS in the Design of all white boxes

- Software reset of BMC
 1. Reset BMC from x86 host via IPMI command
 2. Reset BMC via Redfish
- Software reset of x86
 1. SSH into x86 and issue reset or reboot
 2. Reset x86 host from BMC via IPMI command
 3. Reset x86 host from BMC via Redfish
- Hardware Level reset of BMC via setting on the CPLD on the board
- Hardware Level reset of x86 via CPLD settings.
- Hardware Level reset of x86 or BMC via watchdog timer
- Hardware Level reset of components supplied by 12-volt power rail.
- Hardware Level reset of entire white box via IPMI OEM Command.

13.9. ONIE

Fulfillment of the ONIE hardware specification as laid out here:

https://opencomputeproject.github.io/onie/design-spec/hw_requirements.html

13.10. BMC Software

ODM needs to start support of Open BMC with Redfish implementation with the new white box specifications. However, in some deployments, with Commercial BMC, IPMI 2.0 is the current implementation, therefore IPMI 2.0 needs to be continued to be supported to give time to work

with NOS vendors to migrate to Redfish. Therefore, Large EMUX needs to have support for both IPMI and Redfish implementation.

14. Environmental Regulations and Requirements

The components used in the Large EMUX needs to meet the following requirements.

- AT&T TP76200 & TP76450 for Level 3.
- Copies of this document and general information about AT&T's environmental equipment standards can be found at <https://ebiznet.sbc.com/sbcnebs/>

15. Prescribed Materials

The components used in the Large EMUX needs to meet the following requirements. AT&T TP76200, TP76450 for Level 3, and TP76201.

Copies of this document and general information about AT&T's environmental equipment standards can be found at <https://ebiznet.sbc.com/sbcnebs/>

Compliance to material regulation directives should be from accredited labs and meet the following regulatory directives.

- REACH
- RoHS
- WEEE

16. System Firmware

To reduce firmware upgrade complexity of multi-vendor white boxes and provide the capability to upgrade the firmware on white boxes independently from the NOS version, the Large EMUX must be designed with multi-updater features.

The multi-updater defines a method that decouples the firmware upgrade sequence logic from the NOS while allowing NOS to manage the firmware upgrade. Moreover, the defined method may be used for firmware upgrade by any other management systems throughout all stages of the white box device provisioning. Starting from the ODM release, going through the devices setup at VAR site and ending with device deployment at the customer site.

16.1. Firmware upgrade process

The firmware upgrade process and definitions have been contributed to OCP by AT&T in 2022. Refer to Section 18 of the contributed document titled: [DNX-based Distributed Disaggregated Chassis Routing System Evolution \(V2\): Specification of Hardware and Capabilities](#).

17. Hardware Management

See Section 12.

18. Security

All AT&T White Box initiatives have integrated TPM 2.0 hardware modules installed. (As indicated in section 7.3.1) Although hitherto they have been deployed in a 'disabled' state; AT&T is preparing to leverage this built-in security hardware, along with security elements of Unified Extensible Firmware Interface (UEFI) to combat successful malicious attacks on our White Box network.

TPM has the following capabilities:

1. Performing public key cryptographic operations
2. Computing hash functions
3. Key management and generation
4. Secure storage of keys and other secret data
5. Random number generation
6. Integrity measurement
7. Attestation
8. Hard disk encryption

With TPM 2.0 enabled, vendors can setup inbound checks against PCR measurements from a remote or local verifiers; together with other security firmware interfaces such as, UEFI secure boot, TPM minted EK certificates and a host of other hardware security modules (HSM), to ensure the security of their disaggregated white box environment.

19. References (recommended)

1. "Title", publication year, publication journal/conference/standard, volume, pages, link to publication if available

20. Tables and Figures

Figure 1 Large EMUX Network Topology Diagram	10
Figure 2: Large EMUX High Level System Block Diagram	12
Figure 3: AT&T Standard 42RU Cabinet, Mounting, Clearance Guidelines	12
Table 1: Large EMUX Datasheet	11
Table 2: LED Indicators	25
Table 3: Grouping numbering schemes	26

Appendix A - Checklist for IC approval of this Specification (to be completed by contributor(s) of this Spec)

Complete all the checklist items in the table with links to the section where it is described in this spec or an external document .

Item	Status	Link to detailed explanation
Has this contribution been presented to an OCP Project group during a project call or engineering workshop?	Yes	To be presented in August Telco project call
Approval by Project Leads	Yes	To be approved prior to August Telco project call
Is this contribution entered into the OCP Contribution Portal?	Yes	To be contributed prior to August Telco project call
Was it approved in the OCP Contribution Portal?	Yes	To be approved prior to IC call

Appendix B - Contribution Process FAQs

As a contributor to a hardware specification, here are some questions that often come up.

Q1. What type of specification am I contributing to OCP?

- a. **The base specification** for a de-facto standard (ex: interface type)
- b. The base specification for a product <product type> (product may be coming but within the next 1-2 years)
- c. Modification of an existing <type> specification (state which existing spec is being modified) resulting in a revised specification.
- d. **Design specification** (based on an existing base specification) with more refined design details (product coming in 12-15 months)
- e. A detailed **Product specification** for a <product type> for a very specific product being available in 3-6 months of approval of this Spec
- f. If none of the above, please contact OCP Staff for better direction.

Q2. How do I know if what I am contributing will be accepted by OCP?

- a. Before contributing any specifications, please contact either OCP Staff (Rob Coyle, Michael Schill) or the Project Lead for the Project that best represents your contribution. They will guide you as to what's the best form for your contribution. Project List [here](#).

Q3. What is the contribution process for my hardware spec?

- a. Follow the flow for your spec type [here](#).

Q4. What if my spec is not developed yet and I want to collaborate with other companies?

- a. Please contact either OCP Staff (Rob Coyle or Michael Schill) or the Project Lead for the Project that best represents your contribution. They will help you find other collaborators and help you with the contribution process for a multi-party contribution.

Q5. I have a question about the Contribution License Agreement (CLA).

- a. Please contact OCP Staff and we can help you with questions.

Q6. Do I need to have a product in order to make a contribution?

- a. Please see Q1. Some types of contributions do not result in a product. Some examples are whitepapers, case studies, OCP Ready Assessment, etc.. Please work with the OCP Staff on the better direction on your specification type.