OPEN

Compute Project

50Gb/s Ethernet Mezzanine Card

Connect. Accelerate. Outperform.™

1 Contents

2 Overview

2.1 License

As of April 7, 2011, the following persons or entities have made this Specification available under the Open Web Foundation Final Specification Agreement (OWFa 1.0), which is available at http://www.openwebfoundation.org/legal/the-owf-1-0-agreements/owfa-1-0:

Facebook, Inc.

You can review the signed copies of the Open Web Foundation Agreement Version 1.0 for this Specification at http://opencompute.org/licensing/, which may also include additional parties to those listed above.

Your use of this Specification may be subject to other third party rights. THIS SPECIFICATION IS PROVIDED "AS IS." The contributors expressly disclaim any warranties (express, implied, or otherwise), including implied warranties of merchantability, non-infringement, fitness for a particular purpose, or title, related to the Specification. The entire risk as to implementing or otherwise using the Specification is assumed by the Specification implementer and user. IN NO EVENT WILL ANY PARTY BE LIABLE TO ANY OTHER PARTY FOR LOST PROFITS OR ANY FORM OF INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES OF ANY CHARACTER FROM ANY CAUSES OF ACTION OF ANY KIND WITH RESPECT TO THIS SPECIFICATION OR ITS GOVERNING AGREEMENT, WHETHER BASED ON BREACH OF CONTRACT, TORT (INCLUDING NEGLIGENCE), OR OTHERWISE, AND WHETHER OR NOT THE OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

3 Card Features

Single-Host Card	Multi-Host Card
• ConnectX®-4 EN based • Single-Host x 50GbE port (QSFP28) • PCI-e x8 Gen 3.0 • NC-SI • Power rail supply of 5VAUX and 3.3VAUX • 16MB Flash	• ConnectX®-4 EN based • Multi-Host x 50GbE port (QSFP28) • PCI-e x8 Gen 3.0 • NC-SI • Power rail supply of 5VAUX and 3.3VAUX • 16MB Flash • 128Kb FRU EEPROM • Multi-Host support

For more detailed information, please refer to 50GbE User Manual for OCP.

3.1 Multi-Host Technology

Mellanox's ConnectX®-4 Lx Multi-Host technology enables connecting multiple hosts into a single interconnect adapter by separating the ConnectX-4 Lx PCIe interface into multiple and independent PCIe interfaces. Each interface is connected to a separate host with no performance degradation.

The connection of four fully-independent PCIe buses to four hosts lowers total cost of ownership in the data center. It reduces CAPEX requirements from four cables, four adapters (NICs), and four switch ports to only one of each. Furthermore, it reduces OPEX by cutting down on switch port management and overall power consumption.

Multi-Host technology features uncompromising independent host management, with full independent NC-SI/MCTP support to each host and to the NIC. IT managers can remotely control the configuration and power state of each host individually, such that management of one host does not affect host traffic performance or the management of the other hosts, guaranteeing host security and isolation. To further lower the total cost of ownership, ConnectX-4 Lx supports management of the multiple hosts using a single BMC, with independent NC-SI/MCTP management channels for each of the managed hosts.

The below figure provides an illustration of the Multi-Host technology that the ConnectX-4 Lx Multi-Host Evaluation Board (EVB) enables. The ConnectX-4 Lx adapter connects to the hosts over the PCI Express interface, with dedicated PCIe lanes per host.

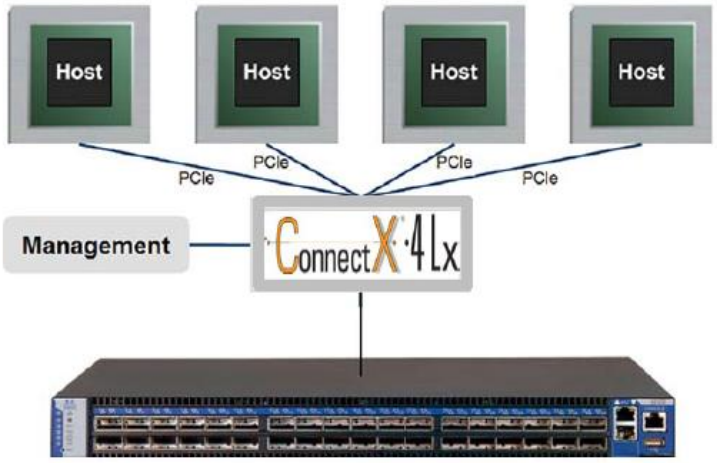


Figure 1: Multi-Host Support Feature

3.2 Form Factor

The following figures illustrate the cards' form factor and dimensions (in mm). Refer to the 2D DXF and 3D files for dimensions, tolerance, and height restrictions.

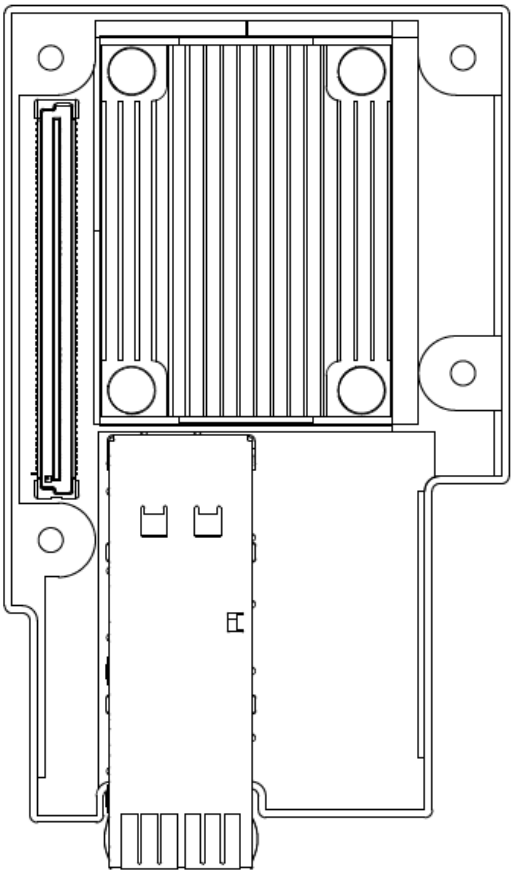


Figure 2: Form Factor (Top) – Single-Host Card

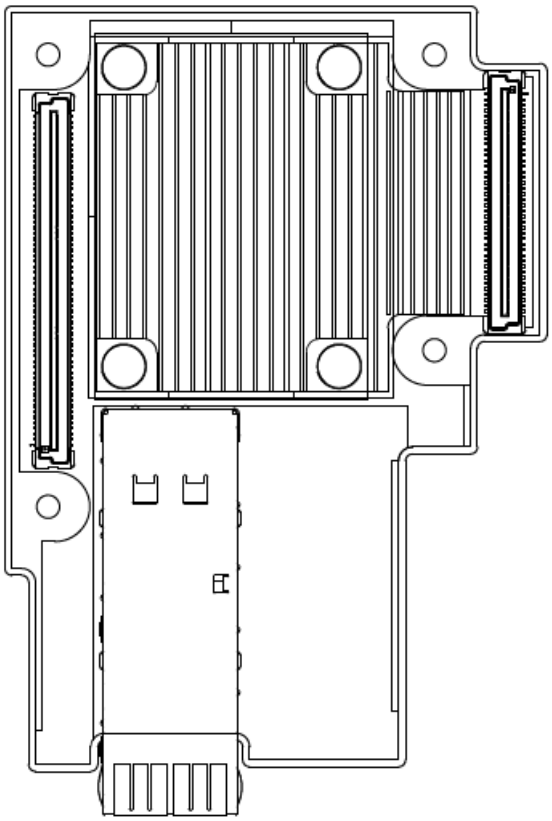


Figure 3: Form Factor (Top) – Multi-Host Card

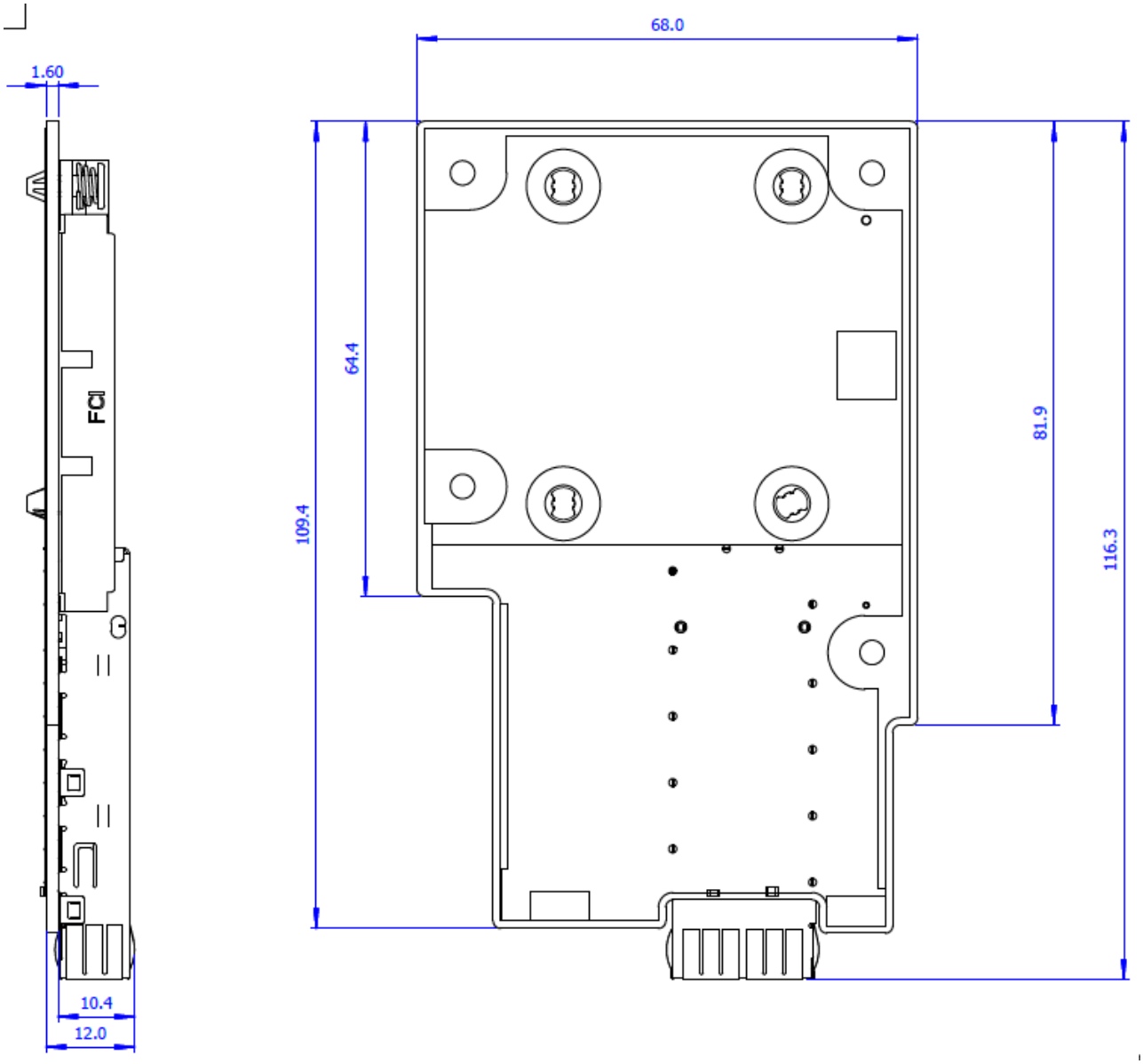


Figure 4: Single-Host Card Dimensions

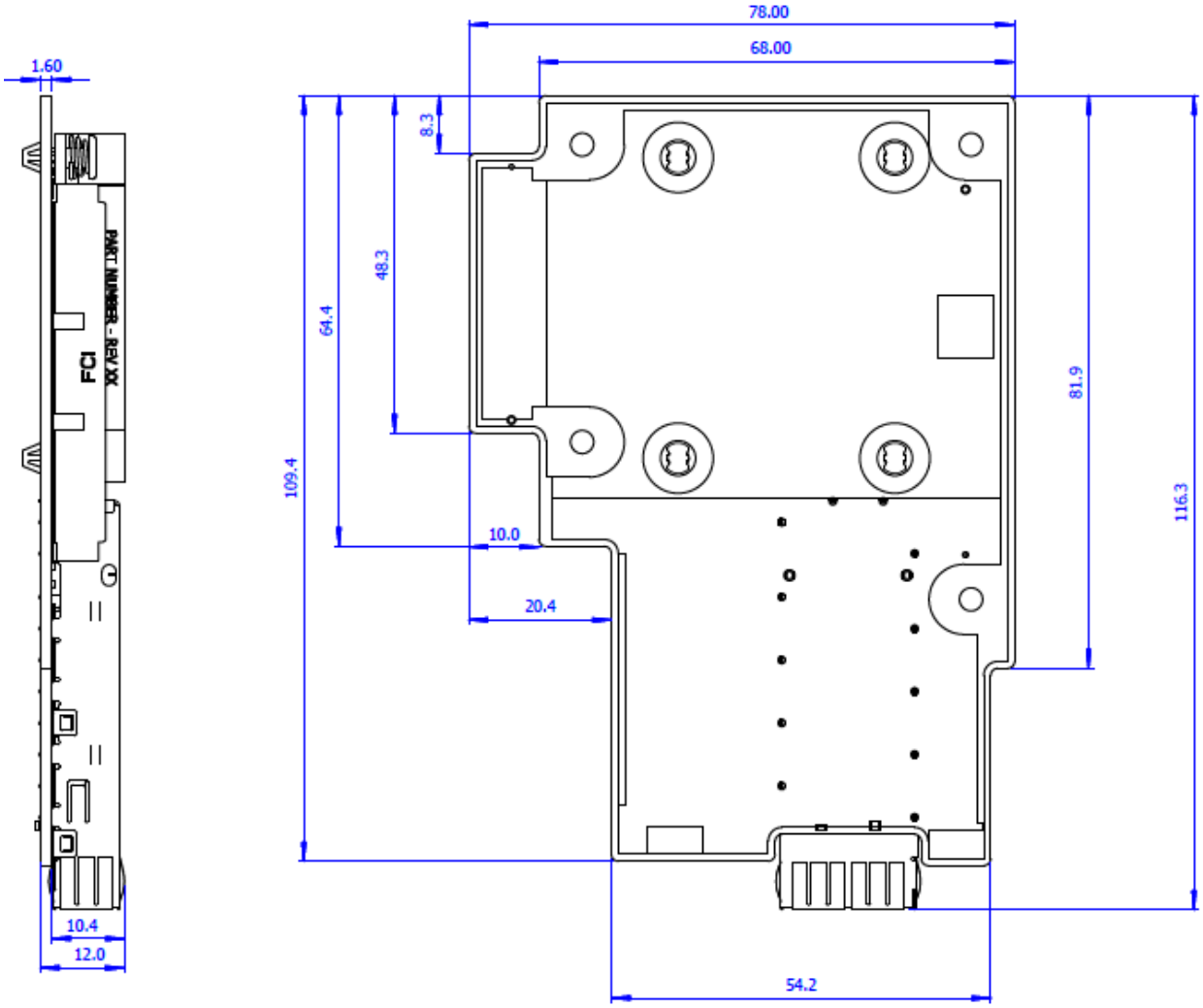


Figure 5: Multi-Host Card Dimensions

3.3 Major Components

- QSFP28
- Heat sink
- LEDs for link indication
- DC-2-DC converters
- **Single- Host Card:** FCI connector to the baseboard
  Multi-Host Card: 2 FCI connectors to the baseboard
- 156Mhz Oscillator
- CPLD for GPIOs expansion (applies only to Multi-Host card)

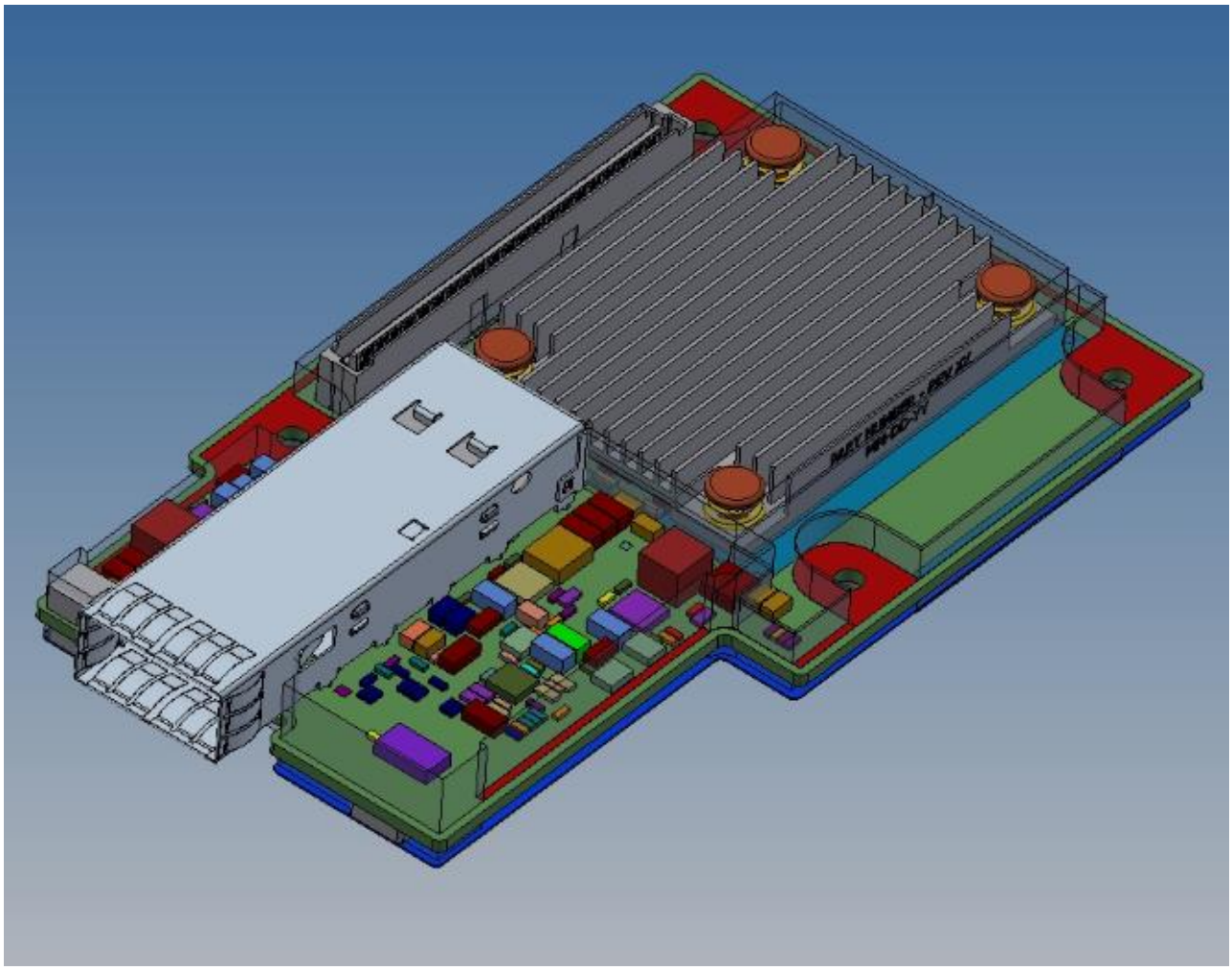


Figure 6: Single-Host Card Top View

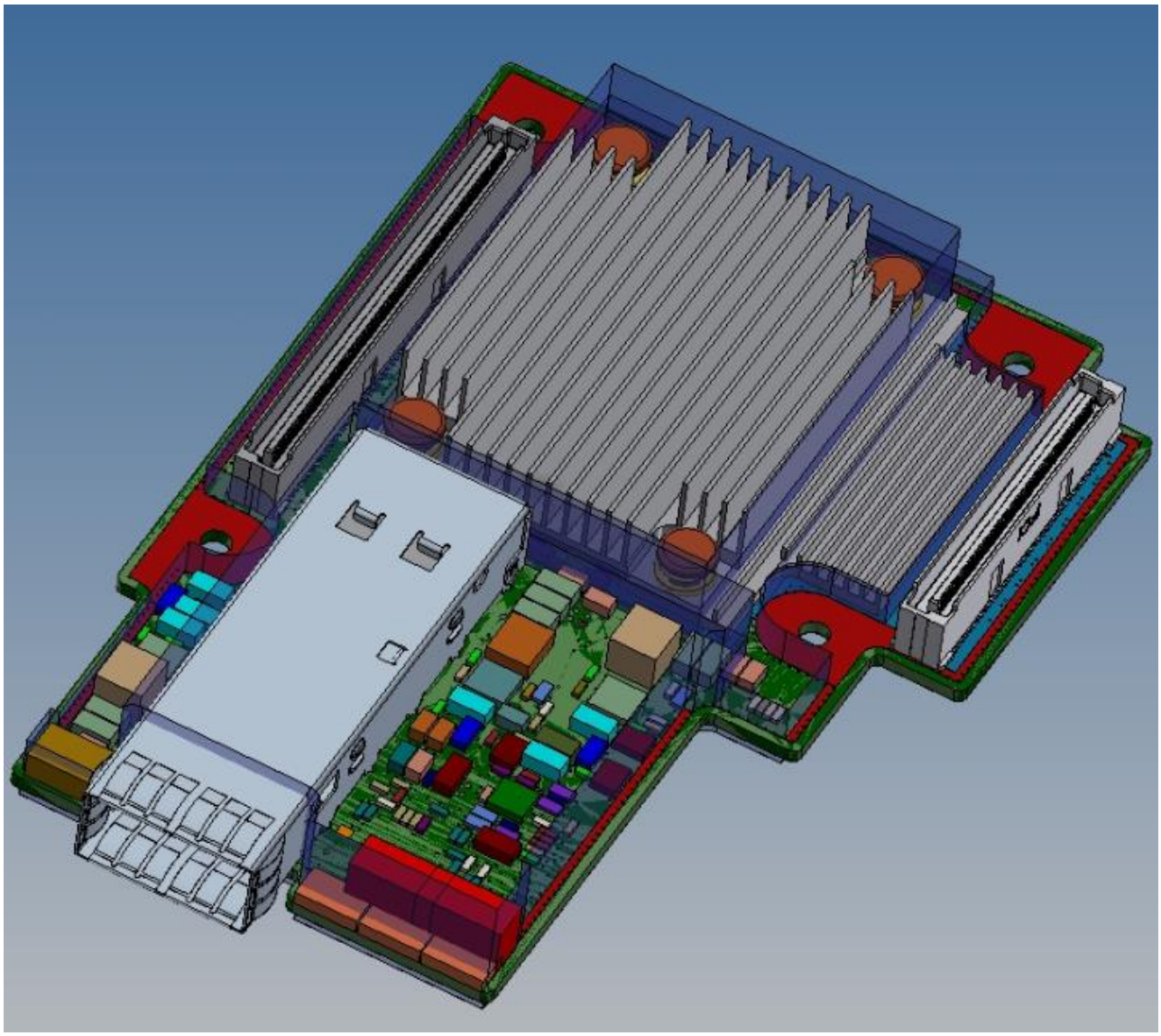


Figure 7: Multi-Host Card Top View

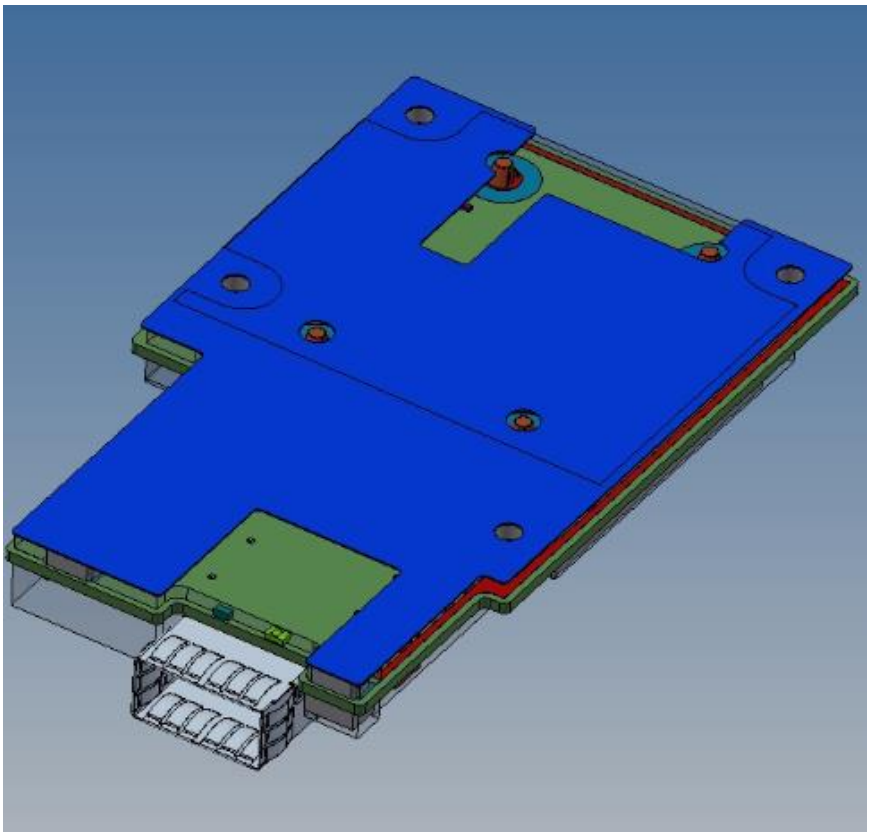


Figure 8: Single-Host Card Bottom View

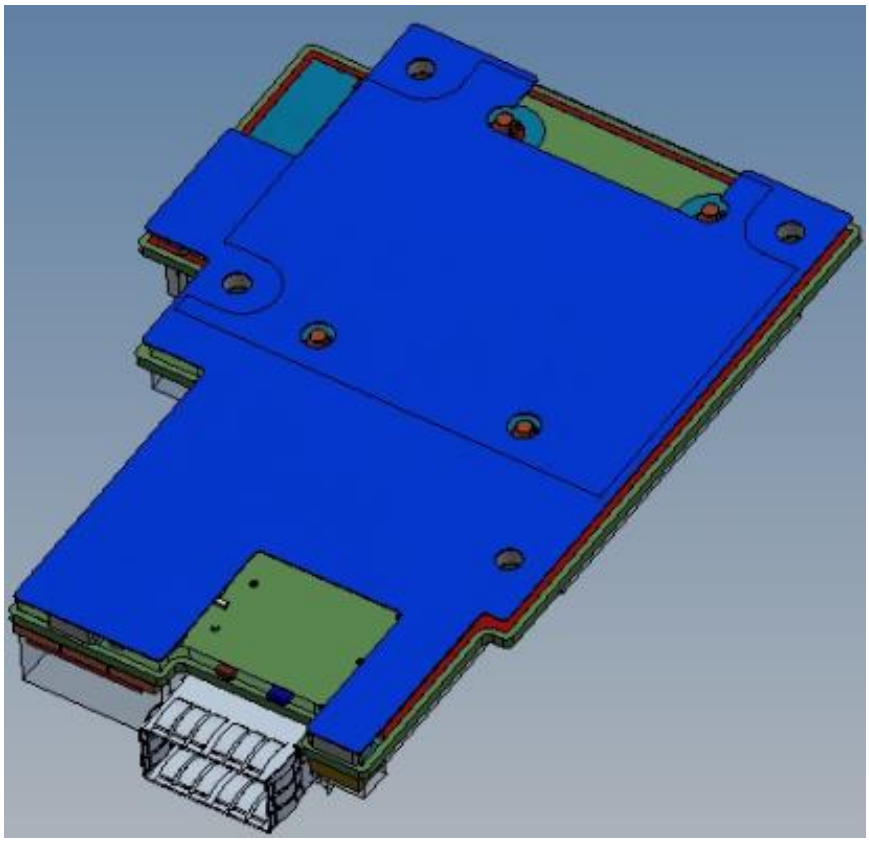


Figure 9: Multi-Host Card Bottom View

3.4 Connector

3.4.1 Single-Host Card

Single-Host card has one FCI B2B connector mounted on the mezzanine card to mate with the connector mounted on the motherboard. The Connector (FCI/61083-124402LF) is a 120 pin connector and provides x8 PCI-e lanes.

3.4.2 Multi-Host Card

Multi-Host card has two FCI B2B connectors, connector A and connector B, mounted on the mezzanine card to mate with connectors mounted on the motherboard. Connector A (FCI/61083-124402LF) is a 120 pin connector and provides x8 PCI-e lanes. Connector B (FCI/61083-084402LF) is an 80 pin connector and provides x8 PCI-e lanes.

3.4.3 Pin Definition

The mezzanine card pinout definitions of Connector A and Connector B are as follows. The directions of the signals are from the perspective of the motherboard.

- Single-Host Card: Connector A
- Multi-Host Card: Connector A and B

Table 10: Connector A Pinout Definition

Pin	Signal	Description	Pin	Signal	Description
A1	MEZZ_PRSNTA1_N/BASEBOARD_A_ID	Present pin1, short to pin120 on Mezz card	A61	P12V_AUX/P12V	AuxPower
A2	P5V_AUX	Aux Power	A62	P12V_AUX/P12V	AuxPower
A3	P5V_AUX	Aux Power	A63	P12V_AUX/P12V	AuxPower
A4	P5V_AUX	Aux Power	A64	GND	Ground
A5	GND	Ground	A65	GND	Ground
A6	GND	Ground	A66	P3V3_AUX	Aux Power
A7	P3V3_AUX	Aux Power	A67	GND	Ground
A8	GND	Ground	A68	GND	Ground
A9	GND	Ground	A69	P3V3	Power (Not Connected)

A10	P3V3	Power (Not Connected)	A70	P3V3	Power (Not Connected)
A11	P3V3	Power (Not Connected)	A71	P3V3	Power (Not Connected)
A12	P3V3	Power (Not Connected)	A72	P3V3	Power (Not Connected)
A13	P3V3	Power (Not Connected)	A73	GND	Ground
A14	NCSI_CRSDV	SE_NCSI_CRS_DV_CON (Carrier Sense/Receive Data Valid)	A74	LAN_3V3STB_ALERT_N	SMBus Alert
A15	NCSI_RCLK	SE_NCSI_REF_CLK (NCSI CLK)	A75	SMB_LAN_3V3STB_CLK	SMBus Clock
A16	NCSI_TXEN	SE_NCSI_TX_EN (Transmit Enable)	A76	SMB_LAN_3V3STB_DAT	SMBus Data
A17	RSVD	S_PERST0_L	A77	RSVD	PCIE wake up
A18	MEZZ_SMCLK	I2C SMBus Clock	A78	NCSI_RXER	SE_NCSI_RX_ERR_CON(Receive error, not supported)
A19	MEZZ_SMDATA	I2C SMBus Data	A79	GND	Ground
A20	GND	Ground	A80	NCSI_TXD0	SE_NCSI_TX_D0 (Transmit Data)
A21	GND	Ground	A81	NCSI_TXD1	SE_NCSI_TX_D1 (Transmit Data)
A22	NCSI_RXD0	SE_NCSI_RX_D0_CON (Receive Data)	A82	GND	Ground
A23	NCSI_RXD1	SE_NCSI_RX_D1_CON (Receive Data)	A83	GND	Ground
A24	GND	Ground	A84	RSVD	2nd set of 100MHz PCIE Clock
A25	GND	Ground	A85	RSVD	2nd set of 100MHz PCIE Clock
A26	RSVD	set of 100MHz PCIE Clock (Not Connected)	A86	GND	Ground
A27	RSVD	set of 100MHz PCIE Clock (Not Connected)	A87	GND	Ground
A28	GND	Ground	A88	KR_TX_DP<8>	PCIE TX signal

A29	GND	Ground	A89	KR_TX_DN_<8>	PCIE TX signal
A30	KR_RX_DP<8>	PCIE RX signal	A90	GND	Ground
A31	KR_RX_DN<8>	PCIE RX signal	A91	GND	Ground
A32	GND	Ground	A92	KR_TX_DP<9>	PCIE TX signal
A33	GND	Ground	A93	KR_TX_DN_<9>	PCIE TX signal
A34	KR_RX_DP<9>	PCIE RX signal	A94	GND	Ground
A35	KR_RX_DN<9>	PCIE RX signal	A95	GND	Ground
A36	GND	Ground	A96	KR_TX_DP<10>	PCIE TX signal
A37	GND	Ground	A97	KR_TX_DN_<10>	PCIE TX signal
A38	KR_RX_DP<10>	PCIE RX signal	A98	GND	Ground
A39	KR_RX_DN<10>	PCIE RX signal	A99	GND	Ground
A40	GND	Ground	A100	KR_TX_DP<11>	PCIE TX signal
A41	GND	Ground	A101	KR_TX_DN_<11>	PCIE TX signal
A42	KR_RX_DP<11>	PCIE RX signal	A102	GND	Ground
A43	KR_RX_DN<11>	PCIE RX signal	A103	GND	Ground
A44	GND	Ground	A104	KR_TX_DP<12>	PCIE TX signal
A45	GND	Ground	A105	KR_TX_DN_<12>	PCIE TX signal
A46	KR_RX_DP<12>	PCIE RX signal	A106	GND	Ground
A47	KR_RX_DN<12>	PCIE RX signal	A107	GND	Ground

A48	GND	Ground	A108	KR_TX_DP<13>	PCIE TX signal
A49	GND	Ground	A109	KR_TX_DN_<13>	PCIE TX signal
A50	KR_RX_DP<13>	PCIE RX signal	A110	GND	Ground
A51	KR_RX_DN<13>	PCIE RX signal	A111	GND	Ground
A52	GND	Ground	A112	KR_TX_DP<14>	PCIE TX signal
A53	GND	Ground	A113	KR_TX_DN_<14>	PCIE TX signal
A54	KR_RX_DP<14>	PCIE RX signal	A114	GND	Ground
A55	KR_RX_DN<14>	PCIE RX signal	A115	GND	Ground
A56	GND	Ground	A116	KR_TX_DP<15>	PCIE TX signal
A57	GND	Ground	A117	KR_TX_DN_<15>	PCIE TX signal
A58	KR_RX_DP<15>	PCIE RX signal	A118	GND	Ground
A59	KR_RX_DN<15>	PCIE RX signal	A119	GND	Ground
A60	GND	Ground	A120	MEZZ_PRSNT2_N	present pin2, short to pin1 on mezz card

Note: Applies only to Multi-Host card.

Table 2: Connector B Pinout Definition

Pin	Signal	Description	Pin	Signal	Description
B1	MEZZ_PRSNTB1_N/BASEBOARD_B_ID	Present pin1, short to pin80 on Mezz card	B41	P12V_AUX/P12V	AuxPower
B2	GND	Ground	B42	P12V_AUX/P12V	AuxPower
B3	KR_RX_DP<0>	PCIE RX signal	B43	RSVD	Reserved
B4	KR_RX_DN<0>	PCIE RX signal	B44	GND	Ground
B5	GND	Ground	B45	KR_TX_DP<0>	PCIE TX signal
B6	GND	Ground	B46	KR_TX_DN<0>	PCIE TX signal
B7	KR_RX_DP<1>	PCIE RX signal	B47	GND	Ground
B8	KR_RX_DN<1>	PCIE RX signal	B48	GND	Ground
B9	GND	Ground	B49	KR_TX_DP<1>	PCIE TX signal
B10	GND	Ground	B50	KR_TX_DN<1>	PCIE TX signal
B11	KR_RX_DP<2>	PCIE RX signal	B51	GND	Ground

B12	KR_RX_DN<2>	PCIE RX signal	B52	GND	Ground
B13	GND	Ground	B53	KR_TX_DP<2>	PCIE TX signal
B14	GND	Ground	B54	KR_TX_DN<2>	PCIE TX signal
B15	KR_RX_DP<3>	PCIE RX signal	B55	GND	Ground
B16	KR_RX_DN<3>	PCIE RX signal	B56	GND	Ground
B17	GND	Ground	B57	KR_TX_DP<3>	PCIE TX signal
B18	GND	Ground	B58	KR_TX_DN<3>	PCIE TX signal
B19	KR_RX_DP<4>	PCIE RX signal	B59	GND	Ground
B20	KR_RX_DN<4>	PCIE RX signal	B60	GND	Ground
B21	GND	Ground	B61	KR_TX_DP<4>	PCIE TX signal
B22	GND	Ground	B62	KR_TX_DN<4>	PCIE TX signal
B23	KR_RX_DP<5>	PCIE RX signal	B63	GND	Ground
B24	KR_RX_DN<5>	PCIE RX signal	B64	GND	Ground
B25	GND	Ground	B65	KR_TX_DP<5>	PCIE TX signal
B26	GND	Ground	B66	KR_TX_DN<5>	PCIE TX signal
B27	KR_RX_DP<6>	PCIE RX signal	B67	GND	Ground
B28	KR_RX_DN<6>	PCIE RX signal	B68	GND	Ground
B29	GND	Ground	B69	KR_TX_DP<6>	PCIE TX signal
B30	GND	Ground	B70	KR_TX_DN<6>	PCIE TX signal

B31	KR_RX_DP <7>	PCIE RX signal	B71	GND	Ground
B32	KR_RX_DN <7>	PCIE RX signal	B72	GND	Ground
B33	GND	Ground	B73	KR_TX_DP<7>	PCIE TX signal
B34	GND	Ground	B74	KR_TX_DN<7>	PCIE TX signal
B35	RSVD	3nd set of 100MHz PCIE Clock	B75	GND	Ground
B36	RSVD	3nd set of 100MHz PCIE Clock	B76	GND	Ground
B37	GND	Ground	B77	RSVD	4nd set of 100MHz PCIE Clock
B38	RSVD	S_PERST1_L	B78	RSVD	4nd set of 100MHz PCIE Clock
B39	RSVD	S_PERST2_L	B79	GND	Ground
B40	RSVD	S_PERST3_L	B80	MEZZ_PRSNTB2_ N	Present pin80, short to pin1 on Mezz card

3.5 Power Capability and Status on Connector

Baseboard supplies power to power pins on Mezzanine card connectors. The current capability and power status is provided in Table 3.**Error! Reference source not found.** ormal power is available on state S0 only. Auxiliary power is available at all power states including hibernate state S4 or off state S5.

Note: 50GbE is fed from 5VAUX and 3.3VAUX power rails only.

Table 3: Power Pins for Connector

Power Rail	Voltage Tolerance	# Pins	Current Capability	Status
P12V_AUX/P 12V	±8%(max)	3	2.4A	Normal power or auxiliary power
P5V_AUX	±9%(max)	3	2.4A	Auxiliary power
P3V3_AUX	±5%(max)	2	1.6A	Auxiliary power
P3V3	±5%(max)	8	6.4A	Normal power

3.6 Installation in Chassis

3D View of shows the 3D view of the mezzanine card installed in an Open Computer chassis.

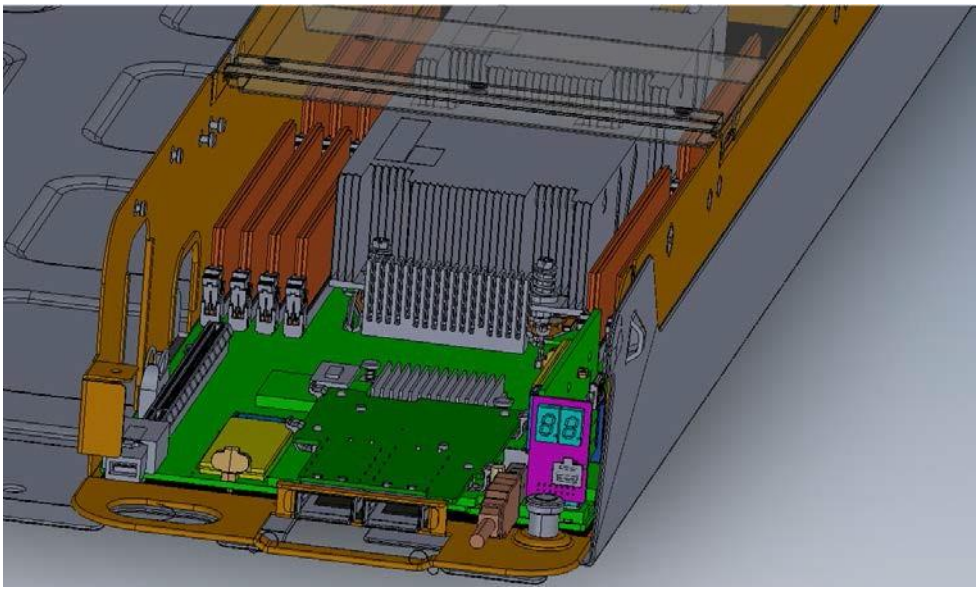


Figure 11: 3D View of Mezzanine Card installed in OCP Chassis - Example

4 Interfaces

4.1 Ethernet Interface

The network ports of the ConnectX®-4 Lx adapter card are compliant with the IEEE 802.3 Ethernet standards. Ethernet traffic is transmitted through the cards' QSFP28 connectors.

4.2 PCI Express Interface

The ConnectX®-4 Lx EN network interface cards support PCI Express 3.0 (1.1 and 2.0 compatible). The device can be either a master initiating the PCI Express bus operations or a slave responding to PCI bus operations. The following lists the PCIe interface features:

- PCIe Gen 3.0, compatible with 2.0 and 1.1
- 2.5GT/s, 5.0 GT/s, or 8.0GT/s link rate x8

- Auto-negotiates to x8, x4, or x1

4.3 LED Interface

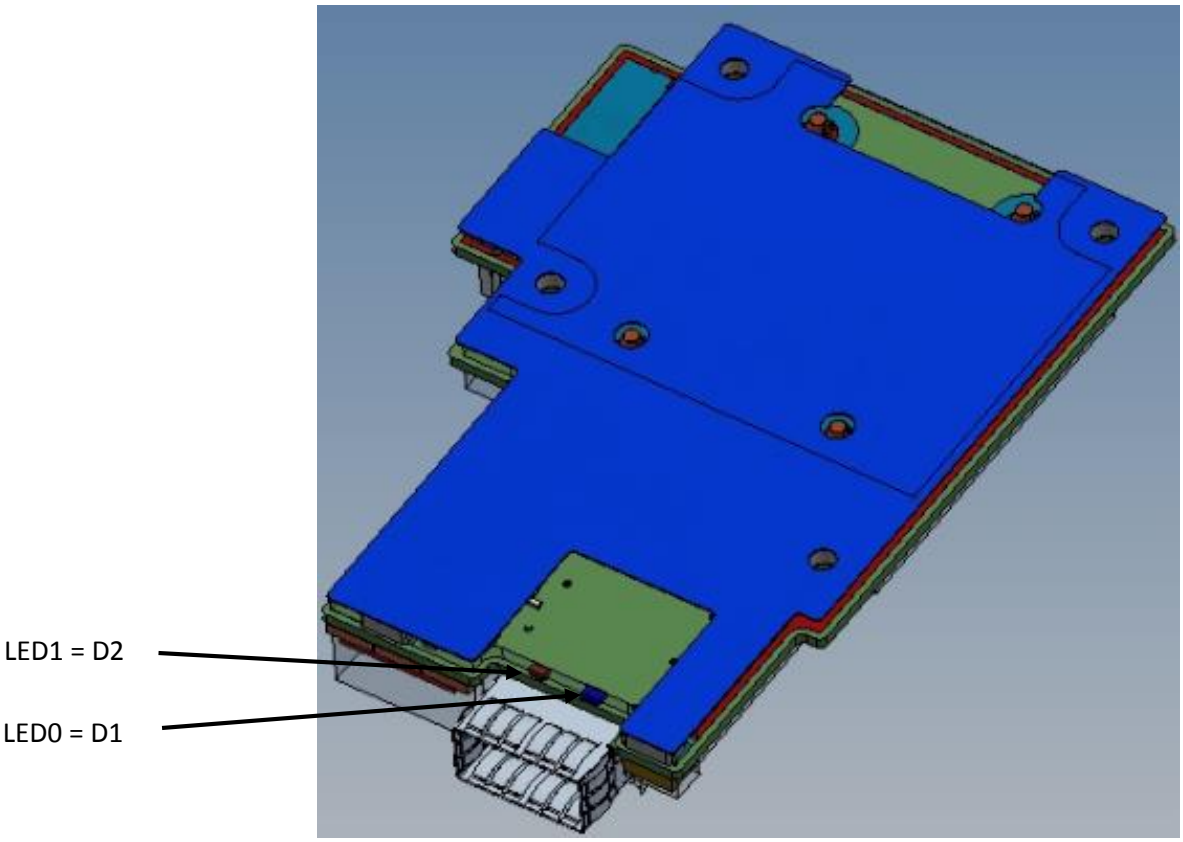


Figure 12: Mezzanine Card LED Location

There are two I/O LED per port to indicate link status and speed. See Table 4.

Table 4: Physical and Logical Link Indications

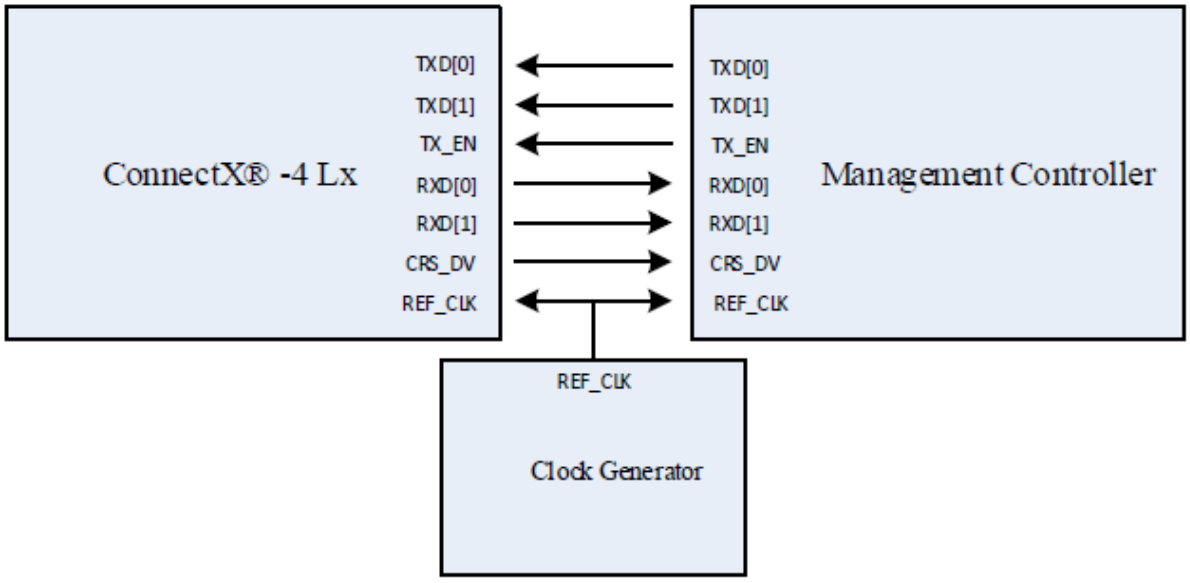
LED	Function	LED Signals
LED0 - Physical link speed	• A constant Green indicates physical link with the highest rated speed. • A constant Yellow indicates physical link with degraded speed. • If LED0 is off, then the physical link has not been established.	Port 0: LED0=D1, LED1=D2
LED1 - Logical link/activity	• A constant Green indicates a valid logical (data activity) link without data transfer. • A blinking Green indicates a valid logical link with data transfer. • If LED1 is off, then the logical link has not been established.	

4.4 Management Interfaces

4.4.1 Network Controller Sideband Interface (NC-SI)

The adapter supports a slave Network Controller Sideband Interface (NC-SI) that can be connected to a BMC. The adapter's NC-SI implementation supports all mandatory NC-SI commands specified in the *Network Controller Sideband Interface (NC-SI) Specification, Rev. 1.1.0.*

The traffic between the BMC and the network goes through a dedicated data-path, and is agnostic to the host operating state. Additionally, through special configuration, local communication between the BMC and the host can be enabled/disabled. When local-loopback is enabled, additional configuration defines if the traffic between the local host and the BMC is visible on the network port or not.



5 Specifications

5.1 Single-Host Card Specifications

Table 5: Single-Host Card Specifications Table

Protocol Support	**Ethernet:** 50GBASE-R4, 40GBASE-CR4, 40GBASE-KR4, 40GBASE-SR4,40GBASE-LR4, 40GBASE-ER4, 40GBASE-R2, 25GBASE-R, 20GBASE-KR2,1000BASE-CX, 1000BASE-KX, 10GBASE-SR, 10GBASE-LR,10GBASE-ER,10GBASE-CX4, 10GBASE-KX4, 10GBASE-CR, 10GBASE-KR, SGMII
	Data Rate: 1/10/25/40/50 Gb/s– Ethernet
	PCI Express Gen-3: SERDES @ 8.0GT/s, 8 lanes (2.0 and 1.1 compatible)
Power and Environmental	**Voltage:** 5V_AUX, 3.3V_AUX
	Typical Power[1]: Passive Cables - 9.07W
	Maximum Power: Passive Cables - 10.82W 3.5 Active Cables - 14.32W
	Maximum power available through QSFP28 port: 3.5W
	Temperature: Operational: 0°C to 35°C Non-operational: -40°C to 70°C
	Humidity: 90% relative humidity[2]
	Air Flow[3]: Passive/Active Cables - 250LFM (port to heat sink)

[1] Typical power for ATIS traffic load.

[2] For both operational and non-operational states.

[3] Simulated inside air tunnel 15.2 mm high and 73 mm wide. Airflow direction - QSFP28 to ConnectX-4 Lx.

5.2 Multi-Host Card Specifications

Table 6: Multi-Host Card Specifications Table

Protocol Support	**Ethernet:** 50GBASE-R4, 40GBASE-CR4, 40GBASE-KR4, 40GBASE-SR4,40GBASE-LR4, 40GBASE-ER4, 40GBASE-R2, 25GBASE-R, 20GBASE-KR2,1000BASE-CX, 1000BASE-KX, 10GBASE-SR, 10GBASE-LR,10GBASE-ER,10GBASE-CX4, 10GBASE-KX4, 10GBASE-CR, 10GBASE-KR, SGMII
	Data Rate: 1/10/25/40/50 Gb/s– Ethernet
	PCI Express Gen-3: SERDES @ 8.0GT/s, 8 lanes (2.0 and 1.1 compatible)
Power and Environmental	**Voltage:** 5V_AUX, 3.3V_AUX
	Typical Power[4]: Passive Cables – 11.2W
	Maximum Power: Passive Cables – 13.5W 3.5 Active Cables – 17W
	Maximum power available through QSFP28 port: 3.5W
	Temperature: Operational: 0°C to 35°C Non-operational: -40°C to 70°C
	Humidity: 90% relative humidity[5]
	Air Flow[6]: Passive/Active Cables - 250LFM (port to heat sink)

[4] Typical power for ATIS traffic load.

[5] For both operational and non-operational states.

[6] Simulated inside air tunnel 15.2 mm high and 73 mm wide. Airflow direction - QSFP28 to ConnectX-4 Lx.

5.3 MAC Address Label Requirements

5.3.1 Placement Rules

MAC address label(s) must be scannable when Mezzanine card is installed in server, rack, etc. by system vendor, rack integrator, and DC user without interrupt of normal operation.

For 2x MAC addresses, 2x 2D bar codes and 2x human readable texts for MAC address need to be placed within 10mm from Mezzanine card PCB edge as shown Figure 12: Mac Address Label Placement.

For 3x MAC addresses, 3x 2D bar codes and 3x human readable texts for MAC address need to be placed within 18mm from Mezzanine card PCB edge.

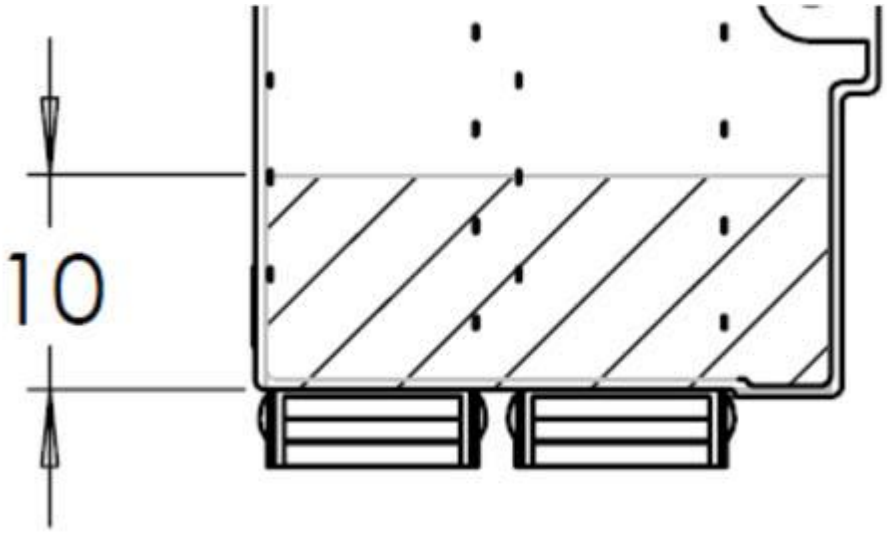


Figure 12: Mac Address Label Placement

The scanned bar code should not include ".". Example: "AA.BB.CC.00.11.20" should scan as "AABBCC001120". Implementation example is shown in Figure 133: MAC Address Label Implementation Example.

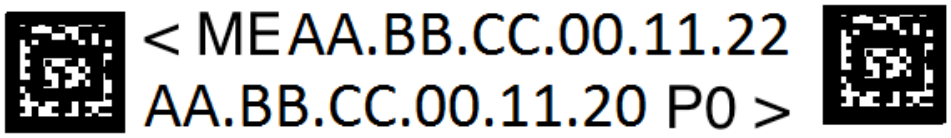


Figure 133: MAC Address Label Implementation Example

5.3.2 Barcode Requirements

A header is required for the label. For example " Po:hh.hh.hh.hh.hh.hh". The font size should be larger that 5 points. If there is more than one MAC address per unit, a human readable text header is required and must differentiate between the MAC addresses. For example "ME:hh.hh.hh.hh.hh.hh" or "Po:hh.hh.hh.hh.hh.hh" or "P1: hh.hh.hh.hh.hh.hh".

5.3.3 Label Permanence and Technology

The MAC address label must adhere for at least 3-4 years in Open Compute usage and thermal environment. Materials can be PCB silkscreen, polyester, and polyamide with acrylic adhesive.

5.3.4 Label Size

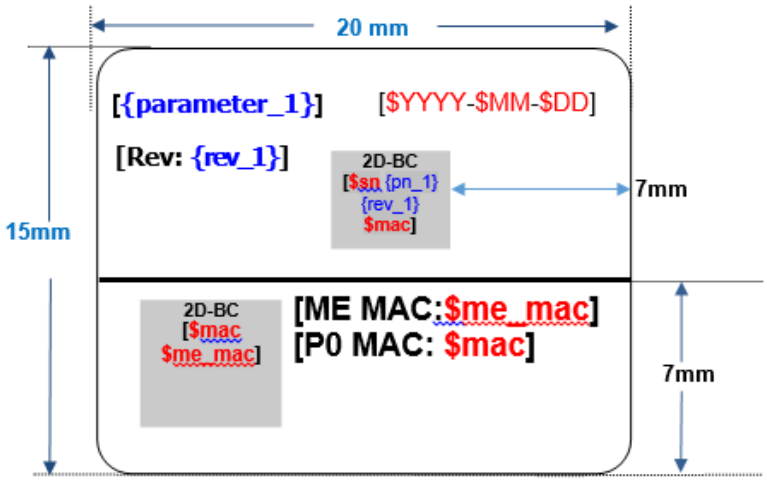


Figure 14: Single-Host Board Label Size

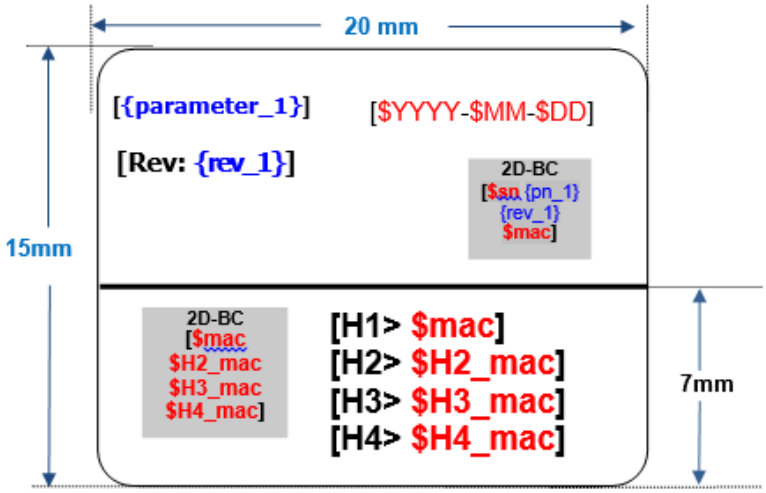


Figure 15: Multi Host Board Label Size

6 Environmental

6.1 Environmental Requirements

This Mezzanine card shall meet the same requirements specified in the OCP systems that the Mezzanine cards is installed in. The OCP system that uses OCP Mezzanine card shall define air flow direction, inlet air temperature, air flow (or speed) to the local area where Mezzanine card is at, and simulation boundary.

Please refer to Section 6: Specifications for Airflow direction, inlet air temperature and air flow (or speed) to the local area where Mezzanine card is at.

6.2 Regulation

This mezzanine card meets KCC, CE, CTUV-US, RCM, ROHS, CB and FCC class A.

7 Revision History

Rev 1.0 – March 2016

- First release