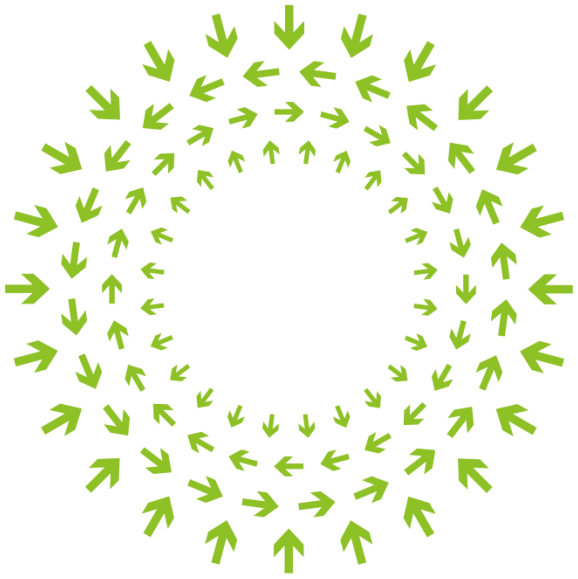


OPEN

Compute Project

Open CloudServer network mezzanine I/O specification V1.1

Authors:

Mark Shaw, Director of Hardware Engineering, Microsoft

Martin Goldstein, Principal Systems Architect, Microsoft

1 Revision History

Date	Name	Description
1/28/2014		Version 1.0
6/5/2014	MASHAW	Version 1.1
		Section 4, Figure 1 – Updated for NCSI Support
		Section 5 – Added 40GbE Future Support
		Section 5.1, Table 1 – Added 40GbE Future Support
		Section 5.3, Table 3 – Adjusted signals to represent dual 10GbE configuration
		Section 5.3, 5.4 – Added NCSI Support
		Section 5.4, Table 4 – Fixed pinout. Swapped location of NWK1/NW2 I2C and Present signals.
		Section 5.5 – Added trace length budgets

2 Scope

This document focuses on the Open CloudServer system network mezzanine input/output (I/O).

3 Contents

4 Overview

This document outlines specifications for the Open CloudServer mezzanine network interface card (NIC). The NIC interfaces to the processor via PCI-Express (PCIe) Gen3 x8 I/O bus and dual 10 Gbit Ethernet. Single and dual 10 Gbit Ethernet are supported. NCSI and Dual 40 Gbit Ethernet pins have been defined for potential future use.

Figure 1 shows a block diagram of the network mezzanine I/O for a 2 x 10GbE configuration.

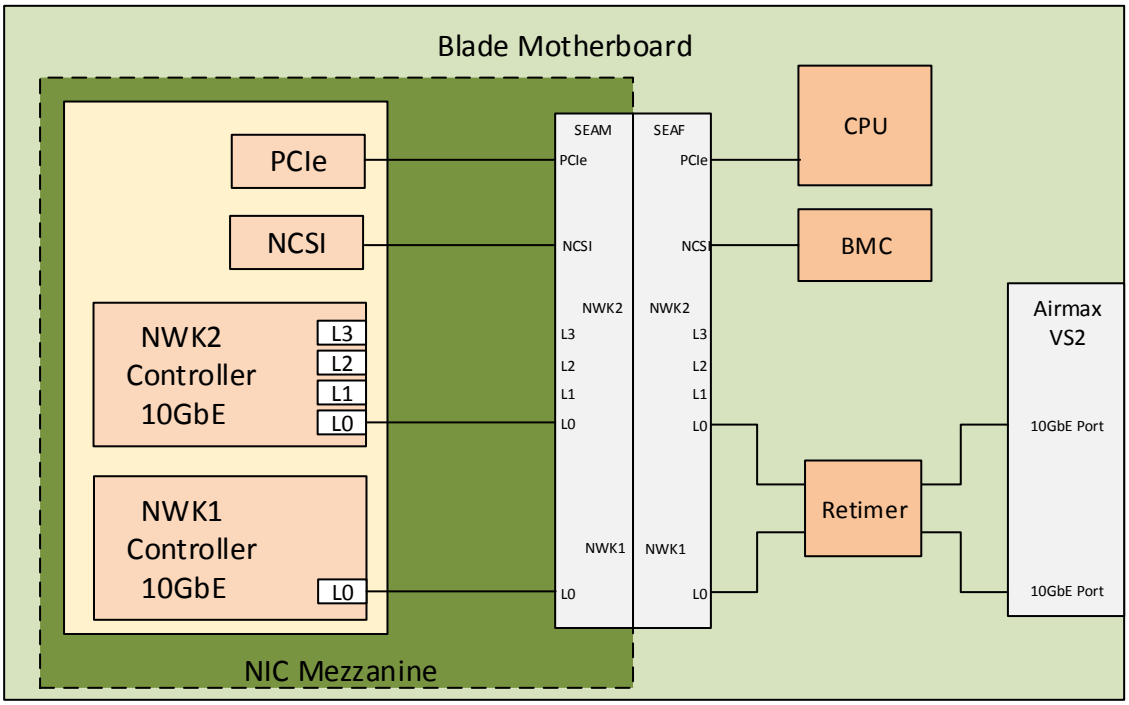


Figure 1. Network Mezzanine I/O connectivity

- The compute blade uses the following retimer:
  - 4-channel: TI DS110RT410; 7x7mm 48-pin quad-flat no-leads (QFN) package
- NWK2 is reserved for optional future 40GbE

5 Signaling Interface

The networking mezzanine interface has been defined to provide high bandwidth and a flexible interface. The card receives a PCI-Express Gen3 x8 bus from the CPU.

The supported output networks configurations include:

- 1 x 10G Ethernet
- 2 x 10G Ethernet
- 1 x 10G and 1 x 40G Ethernet – Future Support

5.1 Ethernet Port Mapping

Table 1 shows the mappings of the networking signals to the backplane.

Table 1. Network Mezzanine Port to Connector Mapping

Networking signals	10G blade signal names	Description
NWK_1_TX/RX[0]	ENET10G_1_TD/RD	10Gb Ethernet Port
NWK_2_TX/RX[0]	ENET10G_2_TD/RD	10Gb Ethernet Port
NWK_2_TX/RX[1:3]	RSVD	Can be combined with NWK_2_TX/RX0 for 40GbE support

5.2 Connectors

Table 2 shows the connector part numbers for the network mezzanine card.

Table 2. Connector Part Numbers, Network Mezzanine Card

Manufacturer	Card connector MPN	Motherboard connector MPN
Samtec	SEAM-20-03.5-L-08-2-A-K-TR	SEAF-20-06.5-L-08-2-A-K-TR
Molex	45970-2385	45971-2385

The stackup height of the SEAM is 10mm with a 6.5mm SEAF and a 3.5mm SEAM connector. In this configuration, it is expected that taller components will be place on the top side of the printed circuit board (PCB), as shown in Figure 2.

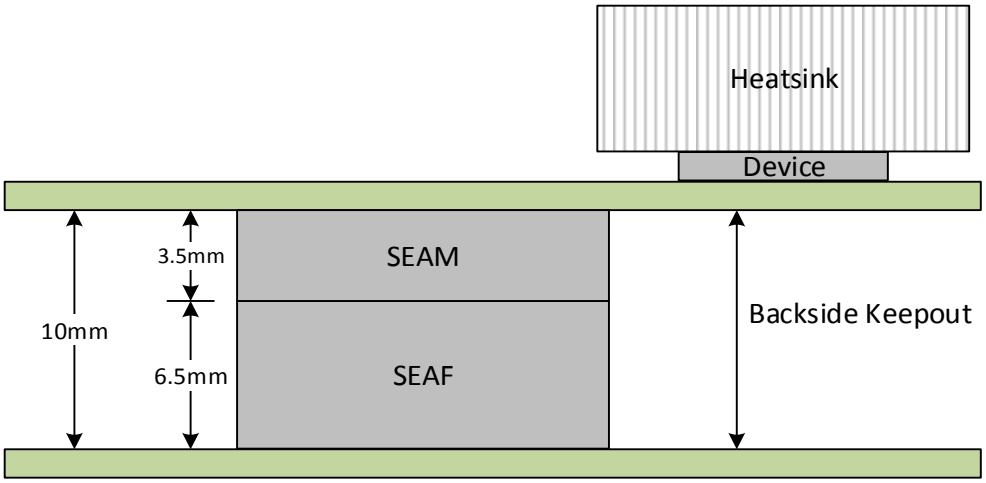


Figure 2. Connector stackup

5.3 Signal Definitions

Table 3 defines the signals used in the NIC mezzanine interface for the dual 10GbE configuration.

Table 3. NIC Mezzanine Connector Signal Definitions

Bus type	I/O	Logic	Definition
P3E_CPU1_LAN_RX_DP/N[7:0]	O	CML	PCIe Gen3 from the NIC Mezz to the CPU
P3E_CPU1_LAN_TX_DP/N[7:0]	I	CML	PCIe Gen3 from the CPU to the NIC Mezz
CLK_100M_NIC_PE_DP/N	I	CML	100MHz PCIe Clock
PCIE_RESET_N	I	3.3V	PCIe Reset
MEZZ_PRESENT_N	O	3.3V	Mezz Present – Should be GND on Mezzanine
NWK_1_TX[0]P/N	O	CML	Port 1 10GbE Transmit from Mezz to Motherboard
NWK_1_RX[0]P/N	I	CML	Port 1 10GbE Receive from Motherboard to Mezz
NWK_2_TX[0]P/N	O	CML	Port 2 10GbE Transmit from Mezz to Motherboard
NWK_2_RX[0]P/N	I	CML	Port 2 10GbE Receive from Motherboard to Mezz
NCSI_TXD[1:0]	O	LVTTL	NCSI Transmit Data[1:0 from NIC to BMC
NCSI_RXD[1:0]	I	LVTTL	NCSI Receive Data[1:0] from BMC to NIC
NCSI_TX_EN	O	LVTTL	NCSI Transmit Data Valid from BMC to NIC

Bus type	I/O	Logic	Definition
NCSI_CLK_IN	I	LVTTL	NCSI Input Clock
NCSI_CRS_DV	I	LVTTL	NCSI Receive Data Valid. Connects to Management Controller Transmit Enable
NCSI_RX_ER	I	LVTTL	NCSI Receive Error
NWK1_ACT_LED	O	3.3V	Port 1 Activity LED
NWK1_LINK_LED	O	3.3V	Port 1 Link LED
NWK2_ACT_LED	O	3.3V	Port 2 Activity LED
NWK2_LINK_LED	O	3.3V	Port 2 Link LED
SMB_ALERT_N	O	3.3V	I2C Alert from NIC Mezz to BMC
NWK1_PRESENT_N	I	3.3V	Port 1 Cable Present Indicator
NWK1_I2C_SDA	I/O	3.3V	Port1 I2C Data to Cable
NWK1_I2C_CLK	O	3.3V	Port 1 I2C Clock to Cable
NWK2_PRESENT_N	I	3.3V	Port 2 Cable Present Indicator
NWK2_I2C_SDA	I/O	3.3V	Port 2 I2C Data to Cable
NWK2_I2C_SCL	O	3.3V	I2C to QSFP+ Cable
PCIE_WAKE_N	O	3.3V	PCIe Wake
SMB_SCL	I	3.3V	I2C to BMC
SMB_SDA	I/O	3.3V	I2C to BMC
NIC_MEZZ_ID[1:0]	O	3.3V	NIC Mezz ID – Connected to BMC on motherboard NIC_MEZZ_ID[1:0] definition is: TBD
P3V3	I	3.3V	3.3V Input Power
P3V3_AUX	I	3.3V	3.3V Aux Input Power
P12V_AUX	I	12V	12V Input Power
Ground			Ground Pins

5.4 Connector Pinout

Table 4 lists the pinout for the 160-pin connector on the tray mezzanine card that interfaces to the tray backplane.

Table 4. NIC Mezzanine Connector PinoutPower

1	GND	PCIE_RESET_N	GND	PCIE_WAKE_N	NIC_MEZZ_ID0	SMB_ALERT_N	GND	CLK_100M_NIC_PE_DP	8
9	P3E_CPU1_LAN_TX_DP<0>	GND	P3E_CPU1_LAN_RX_DP<7>	GND	SMB_SCL	NWK2_PRESENT_N	GND	CLK_100M_NIC_PE_DN	16
17	P3E_CPU1_LAN_TX_DN<0>	GND	P3E_CPU1_LAN_RX_DN<7>	GND	SMB_SDA	GND	MEZZ_PRESENT_N	GND	24
25	GND	P3E_CPU1_LAN_TX_DP<1>	GND	P3E_CPU1_LAN_RX_DP<6>	GND	NCSI_TXD_0	GND	NWK_2_TX0P	32
33	GND	P3E_CPU1_LAN_TX_DN<1>	GND	P3E_CPU1_LAN_RX_DN<6>	GND	NCSI_TXD_1	GND	NWK_2_TX0N	40
41	P3E_CPU1_LAN_TX_DP<2>	GND	P3E_CPU1_LAN_RX_DP<5>	GND	NCSI_RXD_0	GND	NWK_2_TX1P	GND	48
49	P3E_CPU1_LAN_TX_DN<2>	GND	P3E_CPU1_LAN_RX_DN<5>	GND	NCSI_RXD_1	GND	NWK_2_TX1N	GND	56
57	GND	P3E_CPU1_LAN_TX_DP<3>	GND	P3E_CPU1_LAN_RX_DP<4>	GND	NCSI_TX_EN	GND	NWK_2_TX2P	64
65	GND	P3E_CPU1_LAN_TX_DN<3>	GND	P3E_CPU1_LAN_RX_DN<4>	GND	NCSI_CLK_IN	GND	NWK_2_TX2N	72
73	P3E_CPU1_LAN_TX_DP<4>	GND	P3E_CPU1_LAN_RX_DP<0>	GND	NCSI_CRS_DV	GND	NWK_2_TX3P	GND	80
81	P3E_CPU1_LAN_TX_DN<4>	GND	P3E_CPU1_LAN_RX_DN<0>	GND	NCSI_RX_ER	GND	NWK_2_TX3N	GND	88
89	GND	P3E_CPU1_LAN_TX_DP<5>	GND	P3E_CPU1_LAN_RX_DP<1>	GND	NWK_1_ACT_LED	GND	NWK_2_RX0P	96
97	GND	P3E_CPU1_LAN_TX_DN<5>	GND	P3E_CPU1_LAN_RX_DN<1>	GND	NWK_1_LINK_LED	GND	NWK_2_RX0N	104
105	P3E_CPU1_LAN_TX_DP<6>	GND	P3E_CPU1_LAN_RX_DP<2>	GND	NWK_2_ACT_LED	GND	NWK_2_RX1N	GND	112
113	P3E_CPU1_LAN_TX_DN<6>	GND	P3E_CPU1_LAN_RX_DN<2>	GND	NWK_2_LINK_LED	GND	NWK_2_RX1P	GND	120
121	GND	P3E_CPU1_LAN_TX_DP<7>	GND	P3E_CPU1_LAN_RX_DP<3>	GND	NWK_1_TX0P	GND	NWK_2_RX2N	128
129	NWK2_I2C_SDA	P3E_CPU1_LAN_TX_DN<7>	GND	P3E_CPU1_LAN_RX_DN<3>	GND	NWK_1_TX0N	GND	NWK_2_RX2P	136
137	NWK2_I2C_SCL	GND	NIC_MEZZ_ID1	GND	NWK_1_RX0P	GND	NWK_2_RX3N	GND	144
145	P12V_AUX	P3V3_AUX	P3V3	GND	NWK_1_RX0N	GND	NWK_2_RX3P	NWK1_I2C_SDA	152
153	P12V_AUX	P3V3_AUX	P3V3	P3V3	GND	NWK1_PRESENT_N	GND	NWK1_I2C_SCL	160

Table 5 specifies the local area network (LAN) mezzanine power ratings for the rail.

Table 5. LAN Mezzanine Power Ratings

Power rails	Amps/pin (at 40oC)	Total number of pins	Power load capacity by connector pins (W)	Motherboard limited power budget (W)
12V_AUX	2A	2	43.2W	25.2W
3.3V_AUX	2A	2	11.88W	1.2375W
3.3V	2A	3	17.82W	9.9W
Total power budget (per mezzanine card)				25W

Figure 3 shows the LAN mezzanine power-up sequence.

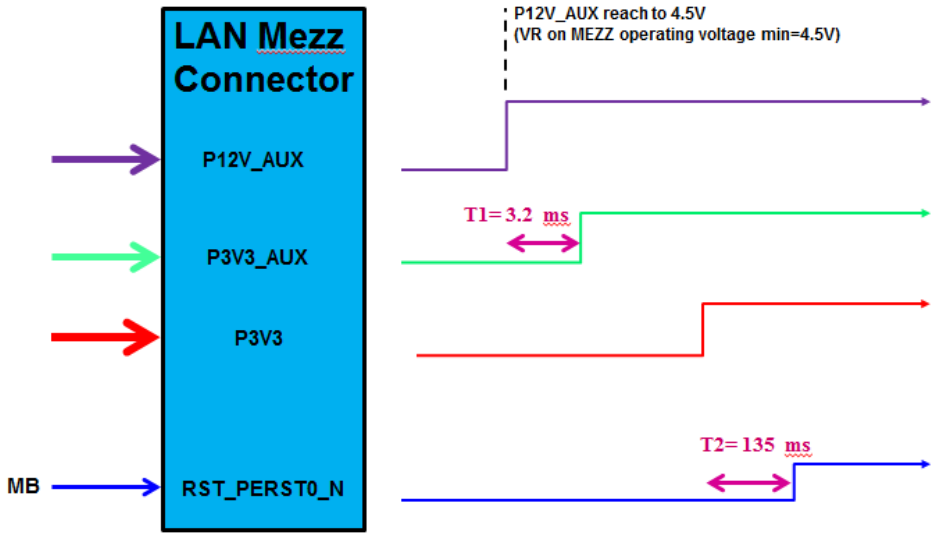


Figure 3. LAN mezzanine power-up sequence

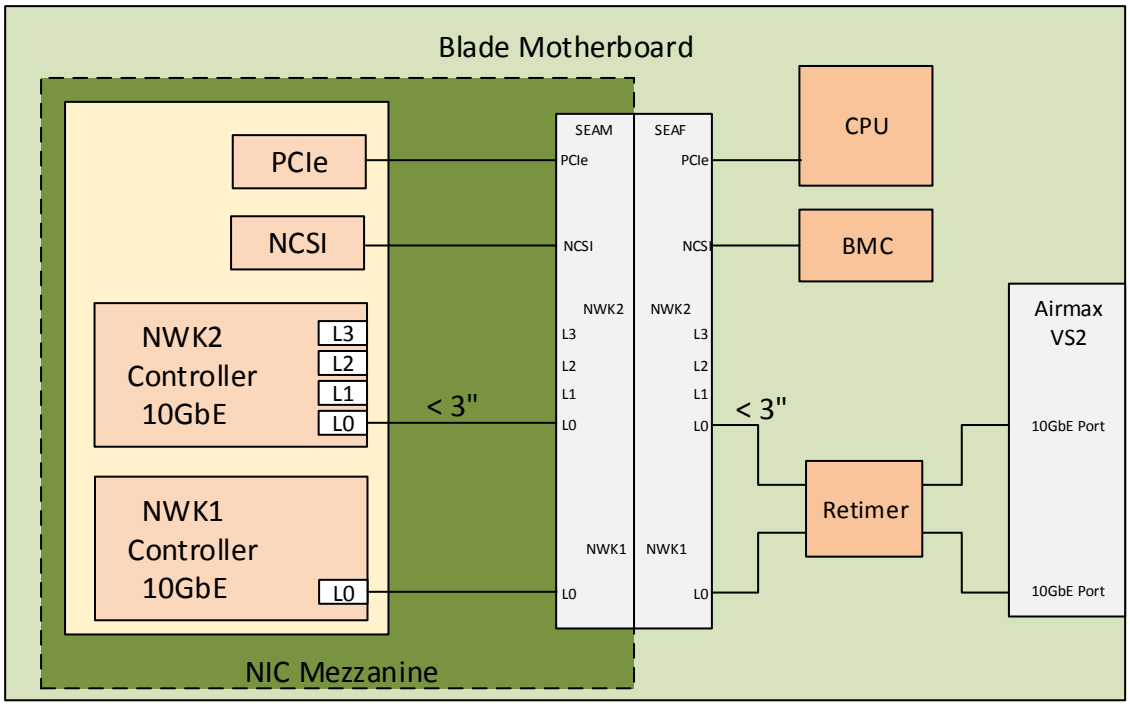
Note that the maximum power consumption of the mezzanine cards is 25W. The reset signal conforms to the PCI Express reset specifications.

5.5 Electrical Budget

Table 6 below shows the budgetary trace lengths for critical signals in the NIC Mezzanine assuming standard FR-4 PCB material. This is intended to communicate the current topology expectation for compatibility with the existing design blade motherboard design and is not intended to replace simulation.

Table 6. Budgetary Trace Lengths

	Trace Length (NIC Mezz)
PCIe	< 3”
10GbE	< 3”
NCSI	< 3”
I2C	< 3”



6 Mechanical

Figure 4 shows the dimensions of the network mezzanine. Note that PCB board thickness can vary as long as the keep-in volumes are not violated.

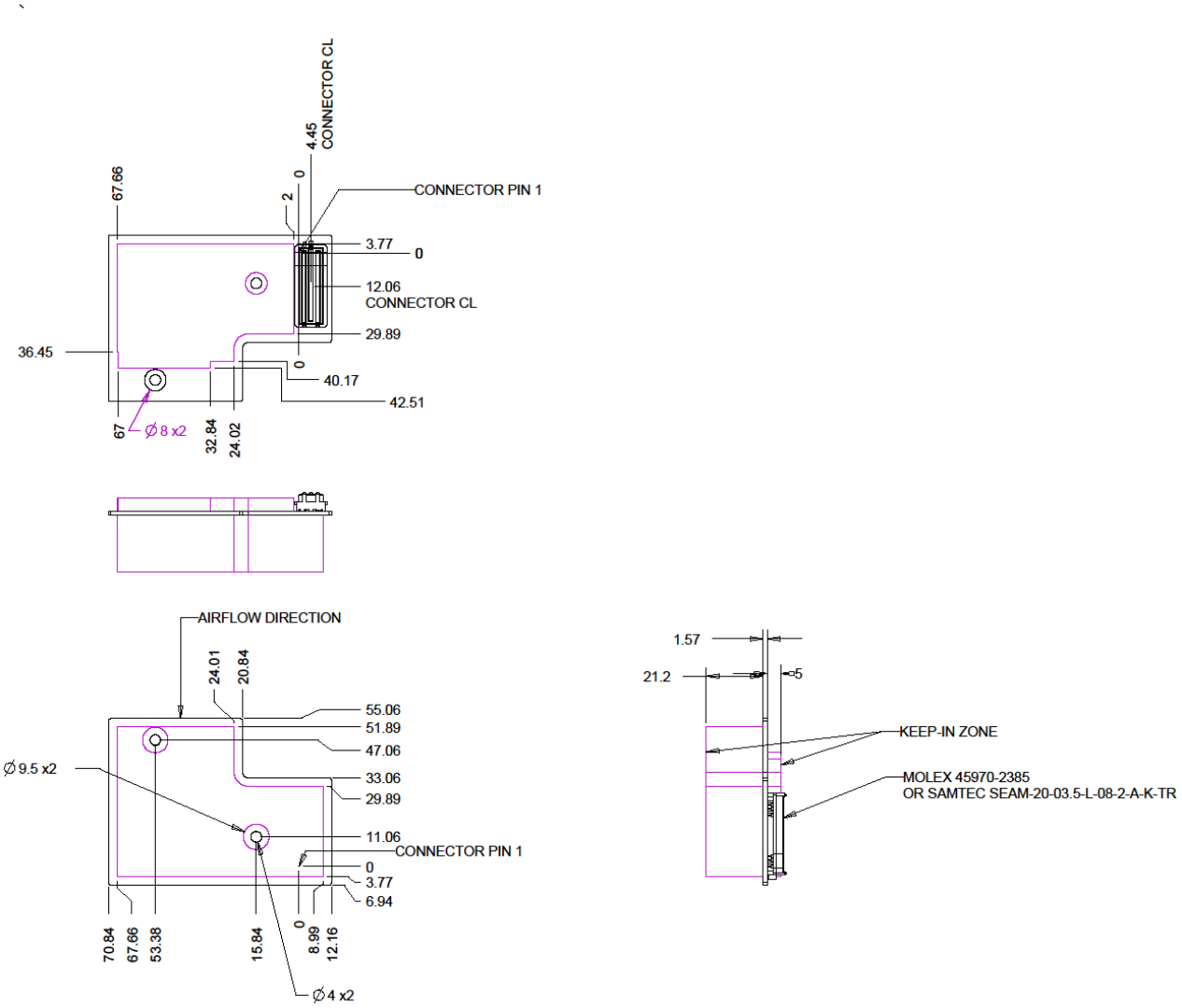


Figure 4. Network mezzanine dimensions

7 Thermal

The network mezzanine cards are designed to be located at the extreme downstream position of the server, so the air is heated by all of the upstream components before reaching the cards. The direction of air flow is shown in Figure 4.

Table 7 shows the worst-case environmental conditions that can be expected at the network mezzanine card inlet. The thermal solution and component selection should be sufficient for these conditions.

Table 7. Environmental Operating Conditions, Network Mezzanine Card

Variable	Worst case operating condition
Approaching airflow rate	200 ft/min, uniform

Maximum allowable pressure drop across card	0.040 "H2O max at 200 ft/min
Approaching airflow temperature	66°C

8 Appendix: Commonly Used Acronyms

This section provides definitions of acronyms used in the system specifications.

ACPI – advanced configuration and power interface

AHCI – advanced host controller interface

AHJ – authority having jurisdiction

ANSI – American National Standards Institute

API – application programming interface

ASHRAE – American Society of Heating, Refrigerating and Air Conditioning Engineers

ASIC – application-specific integrated circuit

BCD – binary-coded decimal

BIOS – basic input/output system

BMC – baseboard management controller

CFM – cubic feet per minute (measure of volume flow rate)

CM – Chassis Manager

CMOS – complementary metal–oxide–semiconductor

COLO – co-location

CTS – clear to send

DCMI – data center manageability interface

DDR3 – double data rate type 3

DHCP – dynamic host configuration protocol

DIMM – dual inline memory module

DPC – DIMMs per memory channel

DRAM – dynamic random access memory

DSR – data set ready

DTR – data terminal ready

ECC – error-correcting code

EEPROM – electrically erasable programmable read-only memory

EIA – Electronic Industries Alliance

EMC – electromagnetic compatibility

EMI – electromagnetic interference

FRU – field replaceable unit

FTP – file transfer protocol

GPIO – general purpose input output

GUID – globally unique identifier

HBI – high business intelligence

HCK – Windows Hardware Certification Kit

HMD – hardware monitoring device

HT – hyperthreading

I^2C – inter-integrated circuit

IBC – international building code

IDE – integrated development environment

IEC – International Electrotechnical Commission

IOC – I/O controller

IPMI – intelligent platform management interface

IPsec – IP security

ITPAC – IT pre-assembled components

JBOD – "just a bunch of disks"

KCS – keyboard controller style

L2 – layer 2

LAN – local area network

LFF – large form factor

LPC – low pin count

LS – least significant

LUN – logical unit number

MAC – media access control

MDC – modular data center containers

MLC – multi-level call

MTBF – mean time between failures

MUX – multiplexer

NIC – network interface card

NUMA – non-uniform memory access

OOB – out of band

OSHA - Occupational Safety & Health Administration

OTS – off the shelf

PCB – printed circuit board

PCIe – peripheral component interconnect express

PCH – platform control hub

PDB – power distribution backplane

PDU – power distribution unit

Ph-ph – phase to phase

Ph-N – phase to neutral

PNP – plug and play

POST – power-on self-test

PSU – power supply unit

PWM – pulse-width modulation

PXE – preboot execution environment

QDR – quad data rate

QFN – quad flat package no-lead

QPI – Intel QuickPath Interconnect

QSFP – quad small form-factor pluggable

RAID – redundant array of independent disks

REST - representational state transfer

RM – rack manager

RMA – remote management agent

ROC – RAID-on-chip controller

RSS – receive-side scaling

RTS – request to send

RU – rack unit

RxD – received data

SAS – serial-attached small computer system interface (SCSI)

SATA – serial AT attachment

SCK – serial clock

SCSI – small computer system interface

SDA – serial data signal

SDR – sensor data record

SFF – small form factor

SFP – small form-factor pluggable

SMBUS – systems management bus

SMBIOS – systems management BIOS

SOL – serial over LAN

SPI – serial peripheral interface

SSD – solid-state drive

TB – tray backplane

TDP – thermal design power

TM – tray midplane

TOR – top of rack

TPM – trusted platform module

TxD – transmit data

U – rack unit

UART – universal asynchronous receiver/transmitter

UEFI – unified extensible firmware interface

UL – Underwriters Laboratories

UPS – uninterrupted power supply

Vpp – voltage peak to peak

WMI – Windows Management Interface