Silicom Connectivity Solutions

PE3100G2DBIRM Content Director Server Adapter

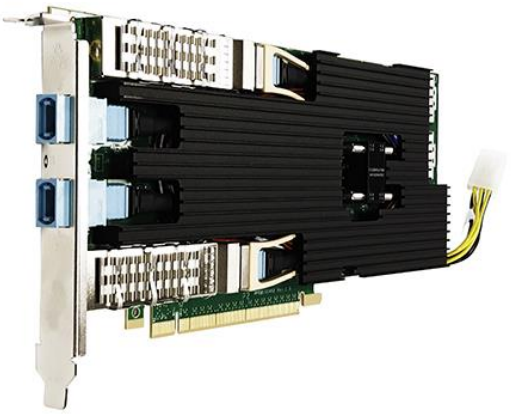
Dual Port Fiber 100 Gigabit Intel® FM10840 Based

Product Description

Silicom's 100 Gigabit Ethernet PCI Express content aware director Bypass server adapters for multi-host platform connectivity is designed for servers and high-end appliances. The Silicom content aware director server adapter is designed with an on-board smart routing architecture that enables packets to be redirected or dropped based on defined rules.

The Silicom's 100 Gigabit Ethernet content aware packet director for multi-host platform connectivity reduces host system process since only packets that are defined to be targeted to the host systems are routed to the host; other packets can be routed to the other port or can be dropped by the content aware hardware routing architecture.

The Silicom's 100 Gigabit Ethernet content aware packet director for multi-host platform connectivity is targeted to network applications that needs to process, monitor or bypass packets based on defined rules. The adapter supports three main modes of operation: Content Aware Bypass, Content Aware TAP and content Aware filtering NIC.



Content Aware Bypass

Silicom's 100 Gigabit Ethernet content aware director for multi-host platform connectivity provides intelligent packet redirection capability where rules specify which packets are directed to the host system and which packets are directed to the other port (Bypass).

Content Aware TAP

Silicom's 100 Gigabit Ethernet content aware director for multi-host platform connectivity provides intelligent packet redirection capability where all packets are directed to the other port (Bypassed) and rules specify which packets are copied to the host system (TAP).

Content Aware Filtering NIC

Silicom's 100 Gigabit Ethernet content aware for multi-host platform connectivity provides intelligent packet redirection capability where rules specify which packets are directed to the host or dropped.

The Silicom 100 Gigabit Ethernet PCI Express content director server adapter is based on Intel FM10840 Ethernet controller and a L3 switch router. Silicom's 100 Gigabit Ethernet PCI Express adapter is based on standard L2 driver and with the content director engine reduces CPU host system processing.

Key Features

Content Aware Director:

- Provides intelligent packet redirection capability where rules specify which packets are directed to the host system and which packets are directed to the other port (Bypass).
- Provides intelligent packet redirection capability where all packets are directed to the other port (Bypassed) and rules that specify which packets are copied to the host system (TAP).
- Provides intelligent packet filtering / drop capability where rules specify which packets are directed to the host or dropped.
- Provides redirection rules that can be defined using source IP/ destination IP / Source Port / Destination Port / VLAN tuples.
- Redirection and packet filtering / drop are performed by the hardware itself in wire speed and do not require any software and CPU host system power processing.
- Intelligent redirect mechanism is controllable via software.
- Intelligent routing mechanism is controllable via software.
- Support 2x100G / QSFP28 ports.
- 100G QSFP28 Ports support 100GBase-SR4 and 100GBase-LR4.

Bypass / Disconnect:

- Bypass / Disconnect Ethernet ports on Power Fail, System Hangs or Software Application Hangs.
- Software programmable Bypass Disconnect or Normal Mode.
- On Board Watch Dog Timer (WDT) Controller.
- Software programmable time out interval.
- Software Programmable WDT Enable / Disable counter.
- Software programmable Bypass Capability Enable / Disable.
- Software Programmable Disconnect Capability Enable / Disable.
- Software Programmable mode (Bypass, Normal or Disconnect mode) at Power up.
- Software Programmable mode (Bypass, Normal mode) at Power off.
- Independent Bypass operation in every two ports.
- Emulates standard NIC

Common Key features:

- PCI Express Multi-Host Interfaces:
  - Bifurcation mode of PCIe x8x8 lanes from/to host via gold fingers of edge card
  - Two additional custom PCIe connectors on board supporting maximum PCIe x8 lanes on each of connector
  - Support PCI Express Base Specification Revision 3.0, 8GT/s, 5GT/s or 2.5GT/s
  - Up to 200Gbps in multi-host connections of 4x 50Gbps 8-lane PCIe interfaces

Intel FM10840 Features:

- Single element 4MB shared memory
- L2/L3/L4/OpenFlow forwarding & ACLs
- Stateless load balancing to CPUs
- Datacenter Bridging (lossless Ethernet)
- 32K 40-bit TCAM entries
- 16K MAC & NextHop tables
- Up to 200Gbps High-bandwidth CPU interface
- Support up to 24,576 frames on receive and transmit queues per port
- Up to 2x100G (4 x 25G)

- 300ns network latency (100GbE)
- 1000nS host-network latency

- **LAN Features:**
  256 queues per PCIe x8 interface
  SR-IOV (64 VFs per PCIe x8 interface)
  IP/TCP/UDP checksum
  Receive side scaling (RSS)
  TCP segmentation offload (TSO/LSO)
  LEDs indicator for link/Activity.

Technical Specifications

Bypass Specifications:	
WDT Interval (Software Programmable):	3,276,800 mSec (3,276.8 Sec): Maximum 100 mSec (0.1 Sec) : Minimum WDT Interval = (2^wdt_interval_parameter)*(0.1) sec. wdt_interval_parameter: { Valid Range: 0-15}
– ZS4: Fiber 100GBASE-SR4 Ethernet Technical Specifications:	
IEEE Standard / Network topology:	Fiber Gigabit Ethernet, 100GBase-SR4 (850nM)
Data Transfer Rate:	103.125GBd
Cables and Operating distance: **Up to:**	Multimode fiber: 62.5um, (OM4) 50m * * Defined as half of max distance
– ZL4: Fiber 100GBASE-LR4 Ethernet Technical Specifications:	
IEEE Standard / Network topology:	Fiber 100Gigabit Ethernet, 100GBASE-LR4 (1310nM)
Data Transfer Rate:	103.1GBd
Cables and Operating distance: **Up to:**	Single-Mode: 5km* * Defined as half of max distance
Operating Systems Support	
Operating system support:	Linux

General Technical Specifications	
Interface Standard:	PCI-Express Base Specification Revision 3.0(8 GTs)
Board Size:	Standard height long add-in card: 241.3mm X 111.15mm (9.5”X 4.376”)
PCI Express Card Type:	X16 Lane (PCIe bifubrication 2×8)
PCI Express Voltage:	+12V ± 8%
External Voltage from external PW jack:	+12V ± 8%
PCI Connector:	Gold Finger: X16 Lane, PCIe bifubrication 2×8 ; 2 x ULTRAPORT SLIMSAS™ (SlimLine) R/A receptacle
Controller:	Intel FM10840
Holder:	Metal Bracket
Operating Humidity:	0%–90%, non-condensing
Operating Temperature:	0°C – 40°C (32°F – 104°F), Air flow requirement 200FLM
Storage:	-40°C–65°C (-40°F–149°F)
EMC Certifications:	Card shall meet CE, FCC Class B, ROHS requirements.
LEDs	
LEDs:	(1) Link/Act 100Gbps LED per port : Link : Turns on Yellow Activity : Blinks Yellow KINGBRIGHT, P/N KPBA-3010SYKCGKC, or compatible. λd : 590 nm (1) Bi-color LED per segment (2 ports): Normal : Off Bypass : Turns on Green KINGBRIGHT, P/N KPBA-3010SYKCGKC, or compatible. λd : 570 nm Disconnect : Turns on Yellow (KINGBRIGHT, P/N KPBA- 3010SYKCGKC, or compatible. λd : 590 nm)
LEDs location:	LEDs are located on the PCB, visible via holes in the metal bracket. Each green Link/Act 100G LED (1 LED per port) is located below its own connector port. The bi-color LED for Bypass/Disconnect is located between the 2 ports.

Connector:	(2) MTP/MTP (SR4) (2) LC/LC (LR4) (2) ULTRAPORT SLIMSAS™ (SlimLine) 8X24 R/A receptacle: Amphenol, P/N U10-A074-260T

Functional Description

Director Director – Content Aware Bypass

Silicom's 100 Gigabit Ethernet content aware director Provides intelligent packet redirection capability where rules specify which packets are directed to the host system and which packets are directed to the other port (Bypass)

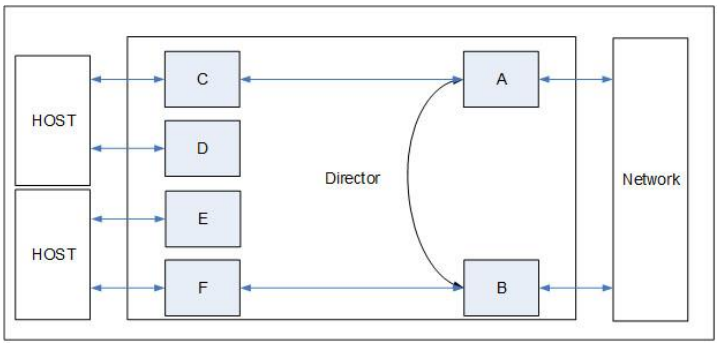


Figure 1: Content Aware Bypass Functional Block Diagram

Figure 1 illustrates example of functional block diagram of content aware Bypass:

Packets received in port A and meet rule are directed to port C and D, other packets are directed to port B (Bypass).

Packets received in port B and meet rule are directed to ports E and F, other packets are directed to port A (Bypass).

Director – Content Aware TAP

Silicom's Ethernet content aware director Provides intelligent packet redirection capability where all packets are directed to the other port (Bypassed) and rules specify which packets are copied to the host system (TAP).

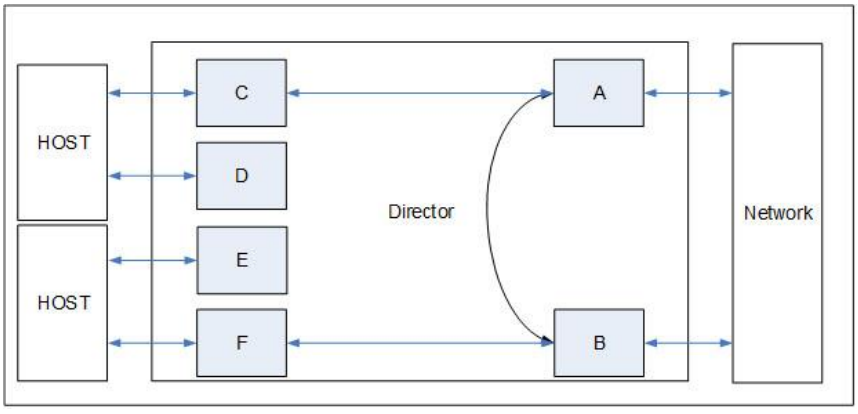


Figure 2: Content Aware TAP Functional Block Diagram

Figure 2 illustrates example of functional block diagram of content aware TAP:

Packets received in port A and meet rule are directed to ports B nad C (TAP), other packets are directed to port B (Bypass).

Packets received in port B and meet rule are directed to ports F and A (TAP), other packets are directed to port A (Bypass).

Director – Content Filtering NIC

Silicom's Ethernet content aware provides intelligent packet redirection capability where rules specify which packets are directed to the host or dropped.

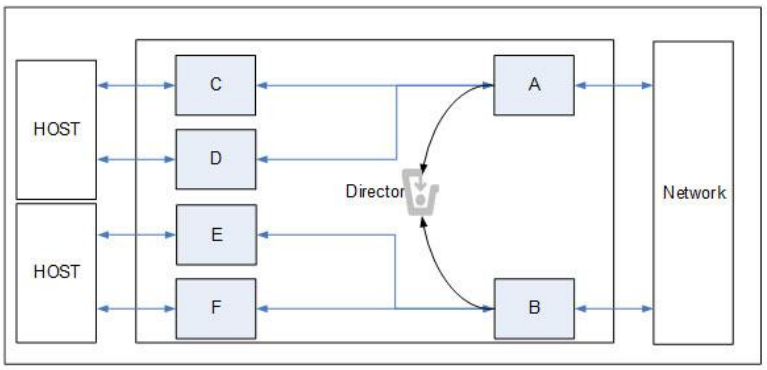


Figure 3: Content Aware Filtering NIC Functional Block Diagram

Figure 3 illustrates functional block diagram of content aware TAP:

Packets received in port A and meet rule, direct to port C and D. Packets received in port A and do not meet rule are dropped.

Packets received in port B and meet rule, direct to ports E and F. Packets received in port B and do not meet rule are dropped.

Director – Load balancing (*Future Support)

Silicom's Ethernet content aware director provides a load balancing of the traffic coming from 2 100G external ports (A, B). The traffic is balanced, based on a defined hash configuration, to the 4x 50G internal interfaces (C,D,E and F) that are going to the host

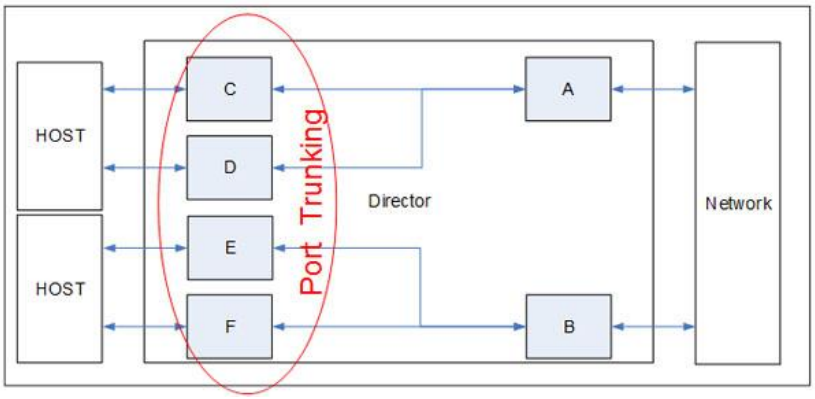


Figure 4: Load Balancing Functional Block Diagram

If there is external port that is heavily loaded the traffic will go into the 4 internal interfaces with balanced load. An ISL tag is added to all incoming packets, it enables the host to know the source port, and the ISL tag is removed from packets that are sent back from the host.

Director – Tagged In-Line rule aware mode (*Future Support)

Silicom's 100 Gigabit Ethernet content aware director Provides intelligent packet redirection capability where rules specify which packets are directed to the host system and which packets are directed to the other ports (Bypass) but at the same time it will get these bypassed traffic into the host with a ISL tag marking that these packets are bypassed, per the rules that the host will issue to the Silicom's 100 Gigabit Ethernet content aware director.

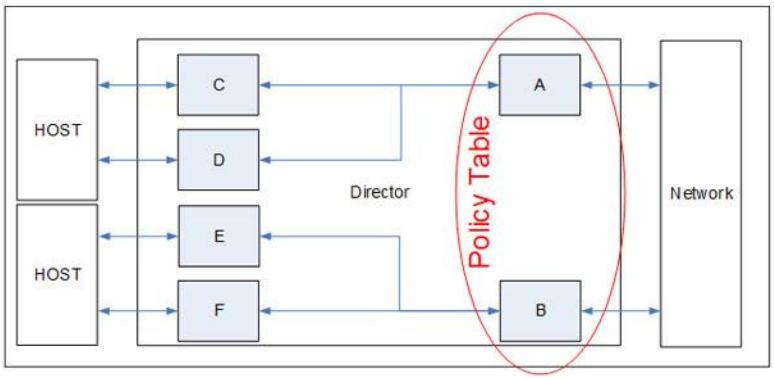


Figure 5: Tagged In-Line rule aware mode Block Diagram

Figure 5 illustrates functional block diagram of Tagged In-Line rule aware mode:
Port Group (A,C,D), (B,E,F) configured as VLAN groups.
Packets received in the 2 x 100G external ports (A and B) and meet and are directed to the other ports of the Vlan group with ISL tag that notify that match found. The original packet is sent to the "Output Port of Switch" in the rule matched policy
Packets received in the 2 x 100G external ports (A and B) and do not meet the rule are directed to the other ports of the Vlan group taged with ISL tag that notify that no-match found.

Director: Rules Classification and capabilities

The Redirector supports the following capabilities:

Maximum total number of rules is 16K.

Each of the 16K rules can be defined to any port the on board multi-layer switch.

Each rule refers to incoming packet.

Rules are executed per order. First rule that matches will be executed.

Rules can be added and removed on the fly.

Each rule can include one or more classification fields. A rule match will be when all fields defined are match.

Each field can have a bit masking to check part of the classification field.

Per port statistics can be read, like packets count, errors, VLAN, and more.

Rules and action are done in wire speed at any packet size.

Rules classification fields

Rules classification is done based on the first 128 bytes of the packets. The following list provides rules classification fields.

MAC address, source & destination

IPv4 – source & destination IP

IPv6 – source & destination IP

L4 Port – source & destination port

Ethernet Protocol – ethertype

IP Protocol num.

VLAN ID tagging

User defined fields

DSCP – match the different services code point – the six most significant bits of the Type of Service octet (IPv4) or Traffic Class octet (IPv6).*

IPv6 Flow Label*

IP length*

ISL Frame Type*

ISL USER*

Source & destination port range*

VLAN priority*

VLAN tag type*

TCP flags*

TOS – match Type of Service octet (IPv4) or Traffic Class octet (IPv6)*

TTL field in a IPv4 header or Hop Limit in a IPv6 header*

*Future SW supports.

Execution per rule

The following Executions per rule are supported:

Drop – when a rule matches the packet will be dropped

Redirect – when a rule matches redirect the packet to the defined destination port

Mirror – when a rule matches copy packet also to a defined destination port

Director Advanced Features:

Port trunking between the different Intel ports to Intel ports connected to it for load balancing between the different Intel ports

Port trunking between the different External Intel port to the external switch connected to it for load balancing between the different external ports.

Session balancing with L3/L4 hashing or other mechanism

ISL (Inter Switch Link) Tagging per port can be added to the packets per configuration

ISL Tagging can be removed and can be forward to specific port per the ISL index.

Quality of Service support with the following features:

Priority levels: 16 internal “switch” priorities, 8 or 16 VLAN priorities (optional use of CFI bit as an extra VLAN priority bit)

Arbitrary mapping of ingress VLAN priority to an internal VLAN priority

Arbitrary mapping of an internal VLAN priority to egress VLAN priority

Arbitrary mapping of internal VLAN priority to switch priority

Arbitrary mapping of DSCP to switch priority, configurable priority source selection.

Scheduler: 8 traffic classes, arbitrary mapping of switch priorities to traffic class, deficit weighted round-robin or strict priority.

Notification: Two congestion notifications can be supported;

Virtual output queue congestion notification (VCN) and Intel proprietary backward congestion notification (FCN).

Open Flow support (consistent with OpenFlow protocol standard)

sFlow support

User defined Packet transmission with two optional modes: 1. Simple mode – transmit on specific port. 2. Switched mode – where switch determines destination port/ports, or with specific information such as whether or not egress processing rules should be applied.

Storm Control Management – Switch can support a variety storm controller. Each storm controller can be programmable to define rat, condition (like unicast ICMP frames whose TTL is at most 1), frame type (can be OR’ed), ingress & egress port ports. Actions: do nothing, drops frames to port (according to filter)

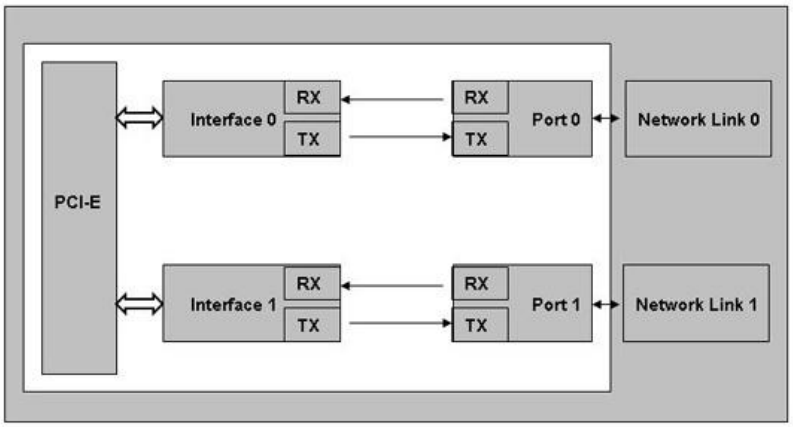
*Future SW supports.

Bypass/Disconnect Mode of operation

Silicom's Bypass adapter supports the following mode states: Normal/inline, Bypass/Fail-To-Wire and Disconnect modes.

Normal/Inline mode

In Normal mode, the ports are independent interfaces (see Figure 6: Normal mode, one Bypass pair is illustrated).



Normal Mode Functional Block Diagram

Bypass/Fail-To-Wire mode

In Bypass mode, the connections of the Ethernet network ports are disconnected from the interfaces and switched over to the other port to create a crossed connection loop-back between the Ethernet ports. The connections of the interfaces are left unconnected. (See Figure 7: one Bypass pair illustrated)

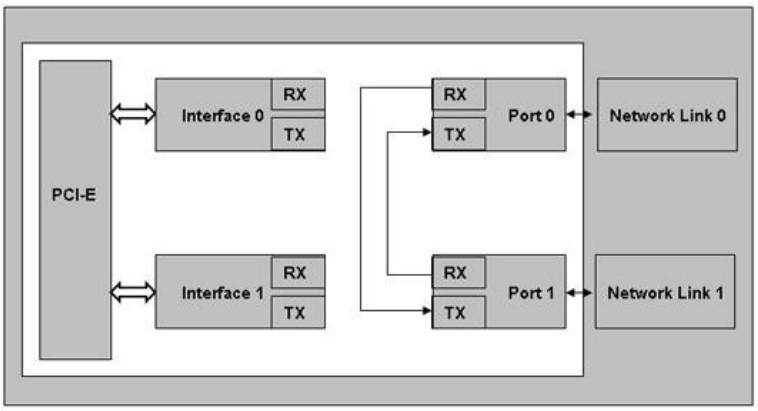


Figure 5: Bypass Mode Functional Block Diagram

Disconnect mode

In Disconnect mode, the transmit connections of the interfaces are disconnected from the ports. The switch / router connected to the adapter does not detect link partner (See Figure 8):

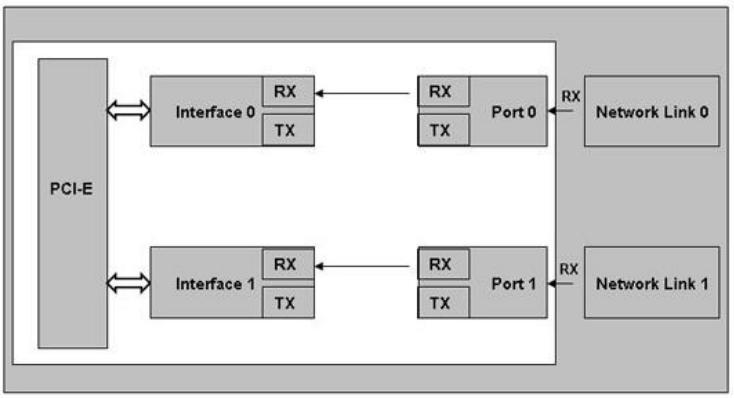


Figure 6: Disconnect Mode Functional Block Diagram

Bypass/Disconnect Features

Silicom's Bypass server adapter supports software programmable to select Normal, Bypass or Disconnect modes. Silicom's Bypass adapters supports Disable Bypass, Disable Disconnected capabilities; hence, if those adapters receive Disable Bypass capability / Disable Disconnect commands, the adapter does not Bypass / does not Disconnect its Ethernet ports, The Disable Bypass Capabilities are reserved also after power off. This feature enables to emulate a standard NIC.

Key Features:

All modes of operation are fully software configured with a well-defined and easy to use API.

Disable bypass/Disconnect for a standard NIC emulation and operation, retained even with power cycle

Range of Watchdog timer timing, configurable for detecting appliance failure.

Watchdog timer reset function by application or self-reset by the product bypass driver

Easy to read product status and product features capabilities

Fast switching between mode with less than 10mS of transit time

Compatible with all Silicom bypass products, one driver set and command set for all products.

PE3100G2DBIRM Block Diagram

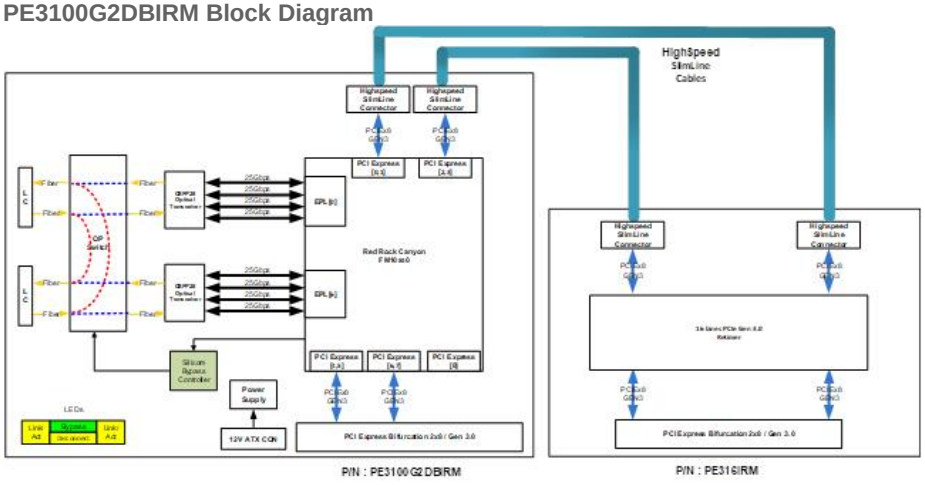


Figure 7: PE3100G2DBIRM Block Diagram

Block Diagram Overview:

The PE3100G2DBIRM Server Bypass adapter is based on FM10840 Ethernet Switch of Intel.

Two Ethernet Port Logic (EPL) which are connected to EPL0 and EPL6 of the FM10840.

Card has four x8 PCIe End Points (PEPs) interfaces which are connected to:

PEP[0,1], PEP[2,3], PEP[4,5], and PEP[6,7].

PEP[4,5] and PEP[6,7] are connected to gold fingers to PCIe connector x16 lanes of host supporting PCIe bifurcation 2×8.

PEP[0,1] and PEP[2,3] are connected on High speed connectors for multi-host connectivity purpose.

To enable this kind of connections Silicom providing a custom PCIe cables and retimer card (P/N PE38IRM and PE316IRM)

Order Information

P/N	Description	Notes	Notes
PE3100G2DBIRM-ZS4	PE3100G2DBIRM-ZS4	Dual port (QSFP28) 100 Gigabit Ethernet PCIe Content Director Bypass Server Adapter, for Multi-Host Platform Connectivity	RoHS Compliant, 2×8 Gen 3, based on Intel FM10840
	PE316IRM	PCI Express x16 Gen 3.0 Host Connector to Dual ULTRAPORT SLIMSAS™ high speed Cable Adapter	RoHS Compliant, 2×8 Gen 3
	CBL000074	SlimSAS x8 Cable, 0.4m	2 units
PE3100G2DBIRM-ZL4	PE3100G2DBIRM-ZL4	Dual port (QSFP28) 100 Gigabit Ethernet PCIe Content Director Bypass Server Adapter, for Multi-Host Platform Connectivity	RoHS Compliant, 2×8 Gen 3, based on Intel FM10840
	PE316IRM	PCI Express x16 Gen 3.0 Host Connector to Dual ULTRAPORT SLIMSAS™ high speed Cable Adapter	RoHS Compliant, 2×8 Gen 3
	CBL000074	SlimSAS x8 Cable, 0.4m	2 units

1V0