Product Brief

PEX8725, PCI Express Gen 3 Switch, 24 Lanes, 10 Ports

Highlights

- **PEX8725 General Features**
  - 24-lane, 10-port PCIe Gen 3 switch
    - Integrated 8.0 GT/s SerDes
  - 19 x 19mm^2, 324-pin FCBGA package
  - Typical Power: 5.4 Watts

- **PEX8725 Key Features**
  - **Standards Compliant**
    - PCI Express Base Specification, r3.0 (compatible w/ PCIe r1.0a/1.1 & 2.0)
    - PCI Power Management Spec, r1.2
    - Microsoft Vista Compliant
    - Supports Access Control Services
    - Dynamic link-width control
    - Dynamic SerDes speed control
  - **High Performance**
    - ***performance*PAK**
      - ✓ Read Pacing (bandwidth throttling)
      - ✓ Multicast
      - ✓ Dynamic Buffer/FC Credit Pool
    - Non-blocking switch fabric
    - Full line rate on all ports
    - Packet Cut-Thru with 132ns max packet latency (x8 to x8)
    - 2KB Max Payload Size
  - **Integrated DMA Engine**
    - Four DMA Channels
    - Internal Descriptor Support
    - DMA function independent from transparent switch function
    - 64-bit Addressing
    - Pre-fetch Descriptor Mode
    - Stride Mode
  - **Multi-Host & Fail-Over Support**
    - 2 Configurable Non-Transparent ports
    - Failover with Non-Transparent port
    - Up to 4 upstream/Host ports with 1+1 or N+1 failover to other upstream ports
  - **Quality of Service (QoS)**
    - Two Virtual Channels
    - Eight traffic classes per port
    - Weighted round-robin source port arbitration
  - **Reliability, Availability, Serviceability**
    - ♦ ***vision*PAK**
      - ✓ Per Port Performance Monitoring
      - ✓ SerDes Eye Capture
      - ✓ PCIe Packet Generator
      - ✓ Error Injection and Loopback
    - 3 Hot-Plug Ports with native HP Signals
    - All ports hot-plug capable thru I^{2}C
    - SSC Isolation on up to 6 ports
    - ECRC and Poison bit support
    - Data Path parity
    - Memory (RAM) Error Correction
    - Advanced Error Reporting
    - Port Status bits and GPIO available
    - JTAG AC/DC boundary scan

The ExpressLane™ PEX8725 device offers Multi-Host PCI Express switching capability enabling users to connect multiple hosts to their respective endpoints via scalable, high bandwidth, non-blocking interconnection to a wide variety of applications including ***servers, storage, communications, and graphics platforms.*** *The PEX8725 is well suited for* ***fan-out****,* ***aggregation, and peer-to-peer*** *traffic patterns.*

Multi-Host Architecture

The PEX8725 employs an enhanced version of PLX's field tested PEX8724 PCIe switch architecture, which allows users to configure the device in legacy single-host mode or multi-host mode with up to four host ports capable of 1+1 (one active & one backup) or N+1 (N active & one backup) host failover. This powerful architectural enhancement enables users to build PCIe based systems to support high-availability, failover, redundant, or clustered systems.

High Performance & Low Packet Latency

The PEX8725 architecture supports packet **cut-thru with a maximum latency of 132ns (x8 to x8).** This, combined with large packet memory, flexible common buffer/FC credit pool and non-blocking internal switch architecture, provides full line rate on all ports for performance-hungry applications such as **servers** and **switch fabrics**. The low latency enables applications to achieve high throughput and performance. In addition to low latency, the device supports a packet payload size of up to 2048 bytes, enabling the user to achieve even higher throughput.

Integrated DMA Engine

The PEX8725 boasts a versatile and powerful built-in DMA engine. The DMA engine removes the burden of having to move data between devices away from the processor – allowing the processor to perform computational tasks instead. The four DMA channels can support high data rate transfers between I/O devices connected to any of the switch's ports. Additionally, the DMA engine in the PEX8725 can be used to complement the DMA engine in the processor by providing additional DMA channels for higher performance.

Data Integrity

The PEX8725 provides **end-to-end CRC** (ECRC) protection and **Poison bit** support to enable designs that require **end-to-end data integrity**. PLX also supports data path parity and memory (RAM) error correction circuitry throughout the internal data paths as packets pass through the switch.

Flexible Configuration

The PEX8725's 10 ports can be configured to lane widths of x1, x2, x4, x8, or x16. Flexible buffer allocation, along with the device's **flexible packet flow control,** maximizes throughput for applications where more traffic flows in the downstream, rather than upstream, direction. Any port can be designated as the upstream port, which can be changed dynamically. Figure 1 shows some of the PEX8725's common port configurations in legacy Single-Host mode.

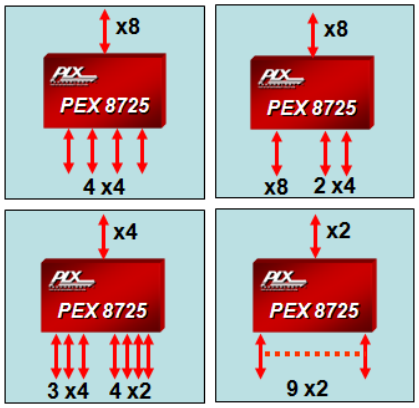

Figure 1. Single-Host Port Configurations

The PEX8725 can also be configured in Multi-Host mode where users can choose up to four ports as host/upstream ports and assign a desired number of downstream ports to each host. In Multi-Host mode, a virtual switch is created for each host port and its associated downstream ports inside the device. The traffic between the ports of a virtual switch is completely isolated from the traffic in other virtual switches. Figure 2 illustrates some configurations of the PEX8725 in Multi-Host mode where each ellipse represents a virtual switch inside the device.

The PEX8725 also provides several ways to configure its registers. The device can be configured through strapping pins, **I²C interface**, host software, or an optional serial EEPROM. This allows for easy debug during the development phase, performance monitoring during the operation phase, and driver or software upgrade.

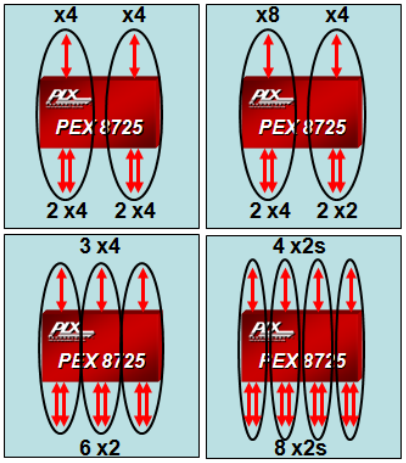

Figure 2. Multi-Host Port Configurations

Dual-Host & Failover Support

In Single-Host mode, the PEX8725 supports 2 **Non-Transparent (NT) Ports,** which enables the implementation of **dual-host systems** for redundancy and host failover capability. The NT port allows systems to isolate host memory domains by presenting the processor subsystem as an endpoint rather than another memory system. Base address registers are used to translate addresses; doorbell registers are used to send interrupts between the address domains; and scratchpad registers (accessible by both CPUs) allow inter-processor communication (see Figure 3).

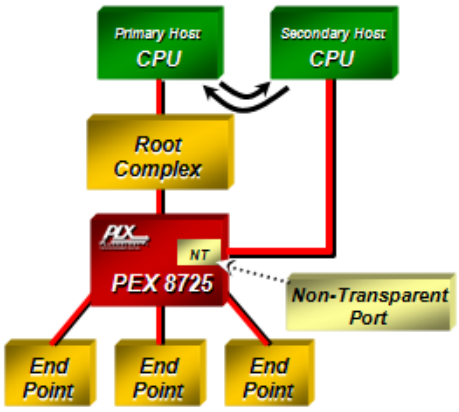

Figure 3. Non-Transparent Port

Multi-Host & Failover Support

In Multi-Host mode, PEX8725 can be configured with up to four upstream host ports, each with its own dedicated downstream ports. The device can be configured for 1+1 redundancy or N+1 redundancy. The PEX8725 allows the hosts to communicate their status to each other via special door-bell registers. In failover mode, if a host fails, the host designated for failover will disable the upstream port attached to the failing host and program the downstream ports of that host to its own domain. Figure 4a shows a two host system in Multi-Host mode with two virtual switches inside the device and Figure 4b shows Host 1 disabled after failure and Host 2 having taken over all of Host 1's end-points.

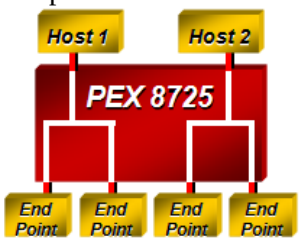

Figure 4a. Multi-Host

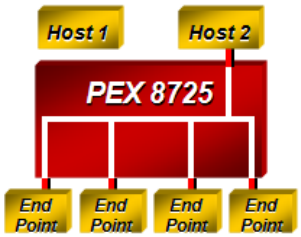

Figure 4b. Multi-Host Fail-Over

Hot-Plug for High Availability

Hot-plug capability allows users to replace hardware modules and perform maintenance without powering down the system. The PEX8725 hot plug capability feature makes it suitable for **High Availability (HA) applications**. Three downstream ports include a Standard Hot-Plug Controller. If the PEX8725 is used in an application where one or more of its downstream ports connect to PCI Express slots, each port's Hot-Plug Controller can be used to manage the Hot-Plug event of its associated slot. Every port on the PEX8725 is equipped with a hot-plug control/status register to support hot-plug capability through external logic via the I²C interface.

SerDes Power and Signal Management

The PEX8725 provides low power capability that is fully compliant with the PCIe power management specification and supports software control of the SerDes outputs to allow optimization of power and signal strength in a system. Furthermore, the SerDes block supports **loop-back modes** and **advanced reporting of error conditions**, which enables efficient management of the entire system.

Interoperability

The PEX8725 is designed to be fully compliant with the PCI Express Base Specification r3.0, and is backwards compatible to PCI Express Base Specification r2.0, r1.1,

and r1.0a. Additionally, it supports **auto-negotiation**, **lane reversal**, and **polarity reversal**. Furthermore, the PEX8725 is tested for Microsoft Vista compliance as well. All PLX switches undergo thorough interoperability testing in PLX's **Interoperability Lab** and **compliance testing at the PCI-SIG plug-fest.**

*performance*PAK™

Exclusive to PLX, *performance*PAK is a suite of unique and innovative performance features which allows PLX's Gen 3 switches to be the highest performing Gen 3 switches in the market today. The *performance*PAK features consists of the Read Pacing, Multicast, and Dynamic Buffer Pool.

Read Pacing

The Read Pacing feature allows users to throttle the amount of read requests being made by downstream devices. When a downstream device requests several long reads back-to-back, the Root Complex gets tied up in serving that downstream port. If that port has a narrow link and is therefore slow in receiving these read packets from the Root Complex, then other downstream ports may become starved – thus, impacting performance. The Read Pacing feature enhances performances by allowing for the adequate servicing of all downstream devices.

Multicast

The Multicast feature enables the copying of data (packets) from one ingress port to multiple (up to 9) egress ports in one transaction allowing for higher performance in dual-graphics, storage, security, and redundant applications, among others. Multicast relieves the CPU from having to conduct multiple redundant transactions, resulting in higher system performance.

Dynamic Buffer Pool

The PEX8725 employs a dynamic buffer pool for Flow Control (FC) management. As opposed to a static buffer scheme which assigns fixed, static buffers to each port, PLX's dynamic buffer allocation scheme utilizes a common pool of FC Credits which are shared by other ports. This shared buffer pool is fully programmable by the user, so FC credits can be allocated among the ports as needed. Not only does this prevent wasted buffers and inappropriate buffer assignments, any unallocated buffers remain in the common buffer pool and can then be used for faster FC credit updates.

*vision*PAK™

Another PLX exclusive, *vision*PAK is a debug diagnostics suite of integrated hardware and software instruments that users can use to help bring their systems to market faster. *vision*PAK features consist of Performance Monitoring, SerDes Eye Capture, Error Injection, SerDes Loopback, and more.

Performance Monitoring

The PEX8725's real time performance monitoring allows users to literally "see" ingress and egress performance on each port as traffic passes through the switch using PLX's Software Development Kit (SDK). The monitoring is completely passive and therefore has no affect on overall system performance. Internal counters provide extensive granularity down to traffic & packet type and even allows for the filtering of traffic (i.e. count only Memory Writes).

SerDes Eye Capture

Users can evaluate their system's signal integrity at the physical layer using the PEX8725's SerDes Eye Capture feature. Using PLX's SDK, users can view the receiver eye of any lane on the switch. Users can then modify SerDes settings and see the impact on the receiver eye. Figure 5 shows a screenshot of the SerDes Eye Capture feature in the SDK.

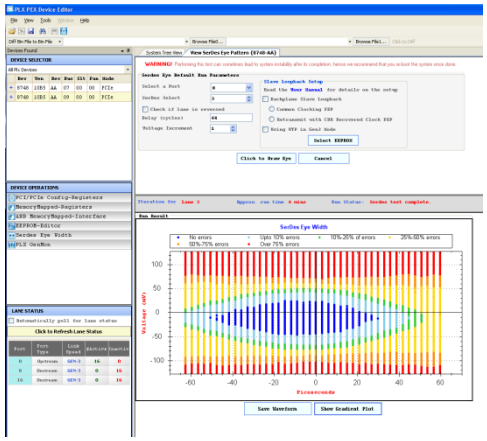

Figure 5. SerDes Eye Capture

PCIe Packet Generator

The PEX8725 features a full-fledged PCIe Packet Generator capable of creating programmable PCIe traffic running at up to Gen 3 speeds and capable of saturating a x16 link. Using PLX's Software Development Kit (www.plxtech.com/sdk), designers can create custom traffic scripts for system bring-up and debug. Fully integrated into the PEX8725, the Packet Generator proves to be a very convenient on-chip debug tool. Furthermore, the Packet Generator can be used to create PCIe traffic to test and debug other devices in the system.

Error Injection & SerDes Loopback

Using the PEX8725's Error Injection feature, users can inject malformed packets and/or fatal errors into their system and evaluate a system's ability to detect and recover from such errors. The PEX8725 also supports Internal Tx, External Tx, Recovered Clock, and Recovered Data Loopback modes.

Applications

Suitable for **host-centric** as well as **peer-to-peer traffic patterns,** the PEX8725 can be configured for a wide variety of form factors and applications.

Host Centric Fan-out

The PEX8725, with its symmetric or asymmetric lane configuration capability, allows user-specific tuning to a variety of host-centric applications. Figure 6 shows a server design where, in a quad or multi processor system, users can assign endpoints/slots to CPU cores to distribute the system load. The packets directed to different CPU cores will go to different (user assigned) PEX8725 upstream ports, allowing better queuing and load balancing capability for higher performance. Conversely, the PEX8725 can also be used in single-host mode to simply fan-out to endpoints.

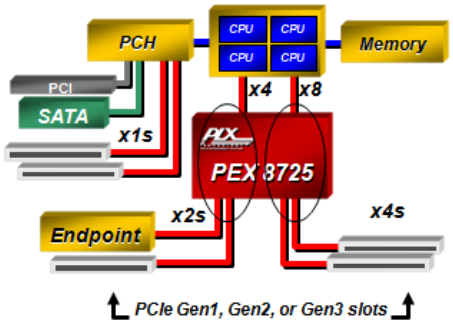

Figure 6. Host Centric Dual Upstream

Multi-Host Systems

In multi-host mode, the PEX8725 can be shared by up to four hosts in a system. By creating four virtual switches, the PEX8725 allows four hosts to fan-out to their respective endpoints. This reduces the number of switches required for fan-out, saving precious board space and power consumption. In Figure 7, the PEX8725 is being shared by four different servers (hosts) with each server is running its own applications (I/Os). The PEX8725 assigns the endpoints to the appropriate host and isolates them from the other hosts.

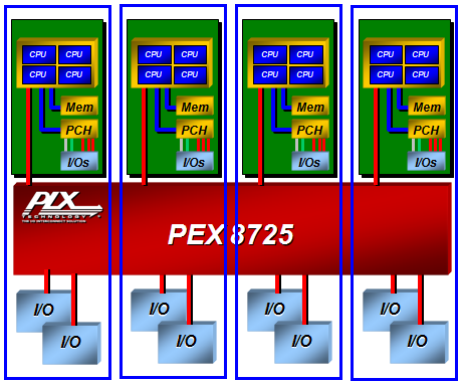

Figure 7. Multi-Host System

Host Failover

The PEX8725 can also be utilized in applications where host failover is required. In the below application (Figure 8), two hosts may be active simultaneously and controlling their own domains while exchange status information through doorbell registers or I^2C interface. The devices can be programmed to trigger fail-over if the heartbeat information is not provided. In the event of a failure, the surviving device will reset the endpoints connected to the failing CPU and enumerate them in its own domain without impacting the operation of endpoints already in its domain.

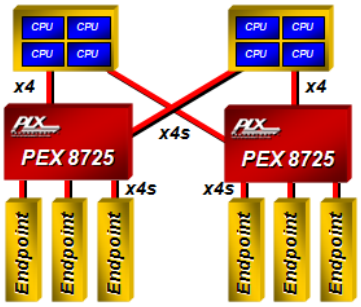

Figure 8. Host Fail-Over

N+1 Fail-Over in Storage Systems

The PEX8725's Multi-Host feature can also be used to develop storage array clusters where each host manages a set of storage devices independent of others (Figure 9). Users can designate one of the hosts as the failover-host for all the other hosts while actively managing its own endpoints. The failover-host will communicate with other hosts for status/heartbeat information and execute a failover event if/when it gets triggered.

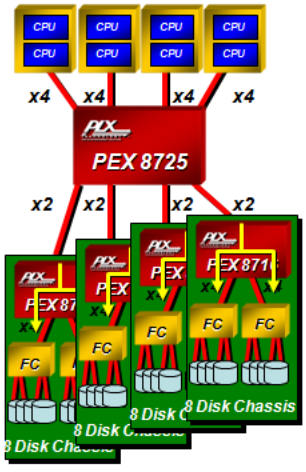

Figure 9. N+1 Failover

Software Model

From a system model viewpoint, each PCI Express port is a virtual PCI to PCI bridge device and has its own set of PCI Express configuration registers. It is through the upstream port that the BIOS or host can configure the other ports using standard PCI enumeration. The virtual PCI to PCI bridges within the PEX8725 are compliant to the PCI and PCI Express system models. The Configuration Space Registers (CSRs) in a virtual primary/secondary PCI to PCI bridge are accessible by type 0 configuration cycles through the virtual primary bus interface (matching bus number, device number, and function number).

Interrupt Sources/Events

The PEX8725 switch supports the INTx interrupt message type (compatible with PCI 2.3 Interrupt signals) or Message Signaled Interrupts (MSI) when enabled. Interrupts/messages are generated by PEX8725 for hot plug events, doorbell interrupts, baseline error reporting, and advanced error reporting.

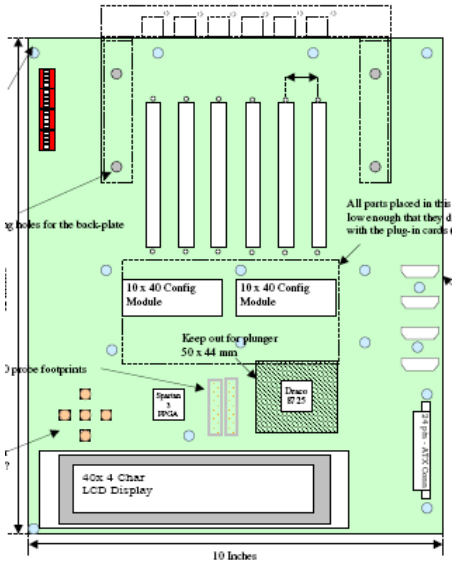

Figure 10. PEX8725 RDK

Development Tools

PLX offers hardware and software tools to enable rapid customer design activity. These tools consist of a hardware module (PEX8725 RDK), hardware documentation (available at www.plxtech.com), and a Software Development Kit (also available at www.plxtech.com).

ExpressLane PEX8725 RDK

The PEX8725 RDK (see Figure 10) is a hardware module containing the PEX8725 which plugs right into your system. The PEX8725 RDK can be used to test and validate customer software, or used as an evaluation vehicle for PEX8725 features and benefits. The PEX8725 RDK provides everything that a user needs to get their hardware and software development started.

Software Development Kit (SDK)

PLX's Software Development Kit is available for download at www.plxtech.com/sdk. The software development kit includes drivers, source code, and GUI interfaces to aid in configuring and debugging the PEX8725.

Both *performance*PAK and *vision*PAK are supported by PLX's RDK and SDK, the industry's most advanced hardware- and software-development kits.

Product Ordering Information

Part Number	Description
PEX8725-BA80BC G	24-Lane, 10-Port PCI Express Switch, Pb-Free (19x19mm^2)
PEX8725-BA RDK	PEX8725 Rapid Development Kit

Visit www.plxtech.com for more information.