



Notes

By Bryan Le

Introduction

This document provides system design guidelines for IDT's PCI Express® 2.0 base specification compliant System Interconnect switch device family. Information provided in this document is applicable to the following devices: 89HPES64H16[A]G2, 89HPES48H12[A]G2, 89HPES34H16G2, 89HPES22H16G2, and 89HPES32H8G2. In this document, the PES64H16G2 is used as the primary reference. The letters "G2" within the device names indicate that these devices are capable of GEN2 (5.0 GT/S) serial data speeds. The PES64H16G2 device offers 64 PCIe lanes divided into 16 ports of 4 lanes each. The PES48H12G2 device offers 48 PCIe lanes divided into 12 ports of 4 lanes each, and so on.

This document also describes the following device interfaces and provides relevant board design recommendations:

- 1) PCI Express Interface
- 2) Reference Clock (REFCLK) Circuitry
- 3) Reset (Fundamental Reset) Schemes
- 4) SMBus Interfaces
- 5) GPIO and JTAG pins
- 6) Power and Decoupling Scheme

PCI Express Interface

Port Configuration

Each of the sixteen ports of the PES64H16G2 is statically allocated 4 lanes with ports labeled from 0 through 15. In a default configuration, SWMODE[3:0] = 0x0, Port 0 is always the upstream port while the remaining ports are always downstream ports. In a Multi-partition configuration, SWMOE[3:0] = 0xC, or a Multi-partition with Serial EEPROM initialization configuration, SWMODE[3:0] = 0xD, all ports come up as unattached. Through a Serial EEPROM or Slave SMBus interface, any port can be configured as an upstream port or as downstream ports. All ports can operate at a maximum link width of x4 (i.e. 4 lanes) and support both 2.5 GT/S (Gen1) and 5.0 GT/S (Gen2) speeds.

Per the PCIe® specification, each switch port is viewed as a virtual PCI-PCI bridge device. In the PES64H16G2, PCI device numbering follows the port numbering. Port 0 corresponds to Device 0 on the upstream bus. Port 1 corresponds to Device 1 on the PES64H16G2 virtual PCI bus, Port 2 to Device 2, and so on.

Note: Unused PCIe TX and RX lanes are not required to have a termination and can be left open.

Notes

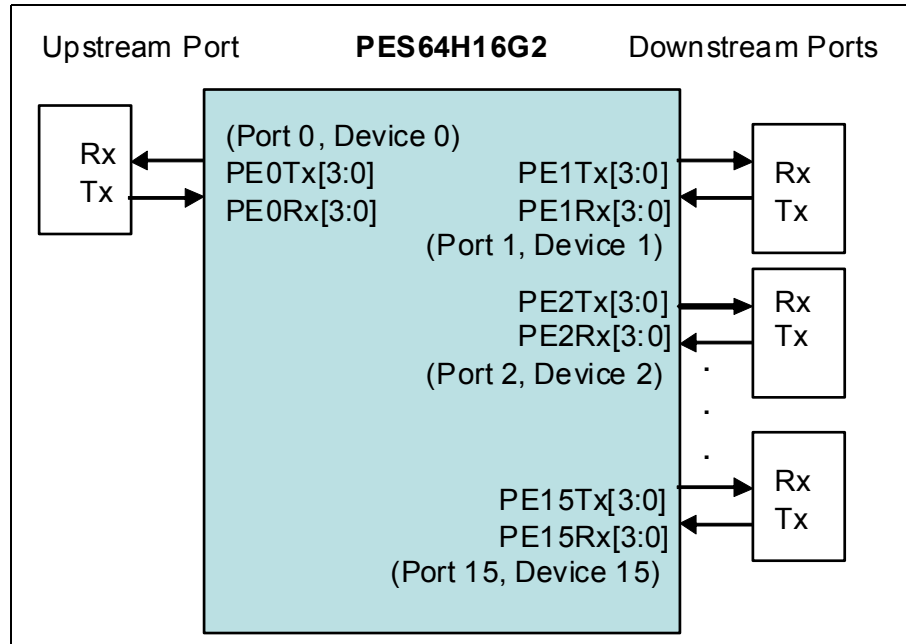


Figure 1 Port Numbering and Device Numbering

Link Width

During link training, link width is automatically negotiated. Each port is capable of independently negotiating to a x4, x2, or x1 link width. Thus, the PES64H16G2 may be used virtually in any sixteen port switch configuration (e.g., {x4, x4, x4, x4}, {x4, x4, x1, x1}, {x4, x1, x1, x1}, etc.).

Lane Reversal

The PES64H16G2 supports automatic lane reversal outlined in the PCIe specification. This allows trace routing flexibility to avoid crossovers and potentially reduces the number of trace vias required for signal routing. Lane reversal must be done for both the transmitter and the receiver of a port.

Lane reversal mappings for the various non-trivial x4 maximum link width configurations are illustrated in Figures 7.1 in the IDT 89HPES64H16G2 Device User Manual.

Polarity Inversion

Each port of the PES64H16G2 supports automatic polarity inversion defined by the PCIe specification. This allows trace routing flexibility to avoid crossovers and potentially reduces the number of trace vias required for signal routing. Polarity inversion is a function of the receiver and not the transmitter. The transmitter never inverts its data. During link training, the receiver examines symbols 6 through 15 of the TS1 and TS2 ordered sets for inversion of the PExRP[n] and PExRN[n] signals. If an inversion is detected, then logic for the receiving lane automatically inverts received data. Polarity inversion is a lane function and not a link function. Therefore, it is possible for some lanes of link to be inverted and for others not to be inverted.

Notes

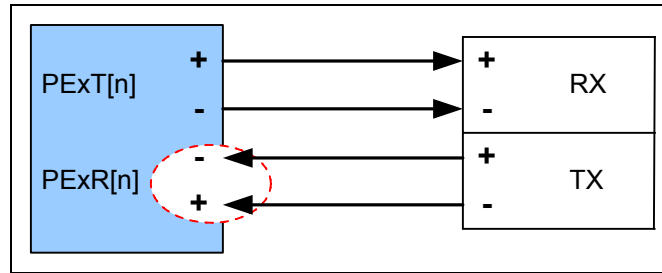


Figure 2 Polarity Inversion

AC Coupling

PCI Express signaling requires AC coupling between the transmitter and the receiver. The capacitor value must be between 75 nF and 200 nF. The 0402 package size and 100 nF value are recommended. The AC coupling capacitors should be placed near the transmitter of the device or the connector to minimize discontinuity effects.

Routing Differential Pairs

The switch includes 50 Ohm resistor on-die terminations on both the transmit and the receive pins. No external termination is required. Individual traces within a given differential pair (positive and negative) must be matched in length to a tolerance of 5 mils. Length matching within a differential pair should occur on a segment-by-segment basis, as opposed to length matching across the total distance of the overall route. In addition, the spacing between traces of adjacent pairs must be at least 20 mils edge-to-edge to reduce crosstalk effects.

Note that trace length matching between pairs is not required because the PCIe 2.0 specifications allow up to 8ns of skew between differential pairs. AC coupling capacitors are associated with the Tx differential pairs and should be located symmetrically on the top or bottom layer between the switch and the PCIe connectors/devices.

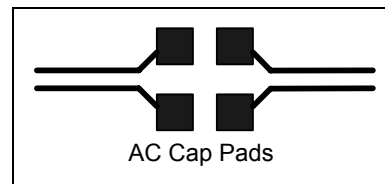


Figure 3 AC Capacitor Placement

Every effort should be made to avoid vias on the PCIe differential pairs since they can result in a signal loss of up to 0.25 dB. When a via is unavoidable, its pad size should be less than 25 mils, its hole size should be less than 14 mils, and its anti-pads should be 35 mils or smaller. No extra vias should be added over and above those needed for IC pads or a connector. Vias in a differential pair should always be at the same relative location and placed in a symmetrical fashion along the differential pair as shown in Figure 4.

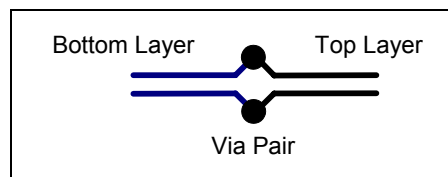


Figure 4 Via Placement

Notes

Avoid 90-degree bends or turns on traces. Wherever possible, the number of left and right bends should be matched as closely as possible. Alternating left and right turns helps to minimize length skew differences between each signal of a differential pair.

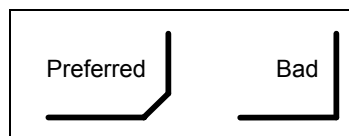


Figure 5 Acceptable Bends and Unacceptable Bends

Depending on the system topology and the maximum targeted trace length, regular FR4 material is appropriate dielectric material. In case of a backplane type of application, higher quality, lower loss material, such as Nelco 4000-13, may be used.

A HSPICE simulation kit for the switch can be requested by emailing ssdhelp@idt.com.

Serdes Reference Resistor Pins

The switch has one Serdes Reference Resistor pin per port. The 3.0K +/- 1% reference bias resistor should be located as close to these pins as possible and should be tapped immediately to the GROUND plane. This resistor must be isolated from any source of noise injection. One way to achieve this is to place the resistor on the back-side of the board, directly underneath the device. No bypass capacitors must be placed on these pins.

Reference Clock (REFCLK) Circuitry

The switch has two differential global reference clock inputs (GCLKP[1:0]/GCLKN[1:0]) that are used to generate all of the clocks required by the internal switch logic and the SerDes. The differential clock inputs require the signal source to drive a 0V common-mode and the REFCLK signal must meet the electrical specifications defined in the PCI Express Card Electromechanical Specification. AC coupling is not required on reference input clocks.

The reference clock inputs support spread spectrum clocking (SSC) for reducing EMI. The required method is to adjust the spread technique to not allow for modulation above the nominal frequency. This technique is often called “down-spreading.” If SSC is used, all clocks must come from a single source. This includes the clock for the switch itself, the clock for the devices connected to the downstream ports of the switch, and the clocks for the root complex chipset or other devices (switch or bridge) connected to the upstream port of the switch. If SSC is not used, multiple clock sources are allowed for each PCI Express device in the tree.

Global Reference Clock Selection

The frequency of the global reference clock inputs can be either 100 MHz or 125 Mhz, and the Global Clock Frequency Select (GCLKFSEL) input is used to indicate the choice of frequency. The PCIe CEM specification requires a nominal frequency of 100 MHz for the reference clock pair. Thus, in the majority of applications, the 100 MHz clock input should be selected.

- For the 100 MHz clock input, GCLKSEL input pin must be asserted low.
- For the 125 MHz clock input, GCLKSEL input pin must be tied to a pull-up resistor of 3.3V power.

Notes

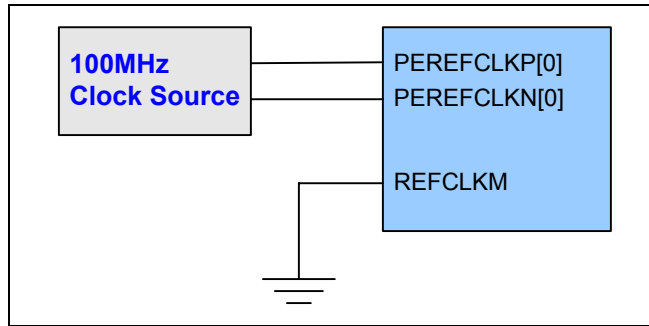


Figure 6 100 MHz Global Reference Clock Selection

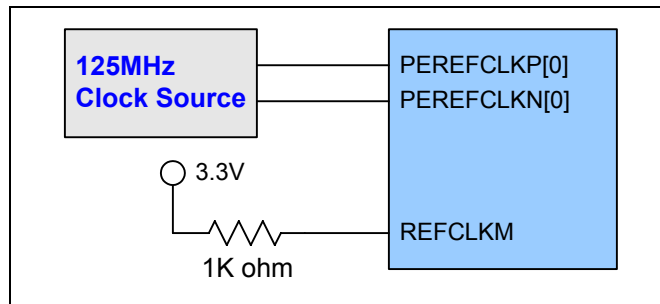


Figure 7 125 MHz Global Reference Clock Selection

The differential clock for the switch and the fifteen downstream devices can be derived from the clock buffer/generator such as ICS9DB803. System designers can use other Gen2-compatible clock buffers/generators. The ICS9DB803 device is used on the IDT evaluation boards.

The switch provides two clock operation modes for each side of the switch: Global Clock and Local Port Clock. System designers must configure the CLKMODE[2:0] pins depending on which mode is chosen for the upstream side and the downstream side of the switch. For details related to each mode, please refer to the PES64H16G2, PES48H12G2, or PES32H8G2 Device User Manuals. The Spread Spectrum Clock must be disabled when the non-common clock is used on either the upstream port or downstream port.

Note: The downstream port's reference clock must be controlled by the power good signal or an appropriate control signal if downstream slots support hot-plug operation.

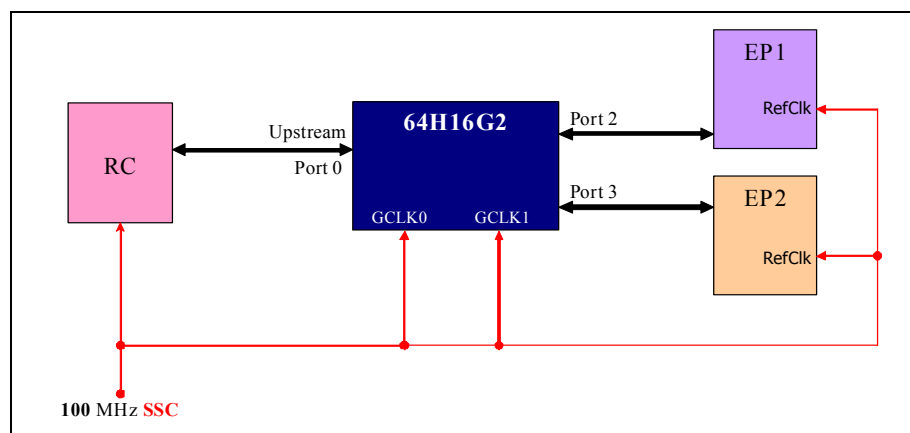


Figure 8 Example of Common Clock Mode with SSC

Notes

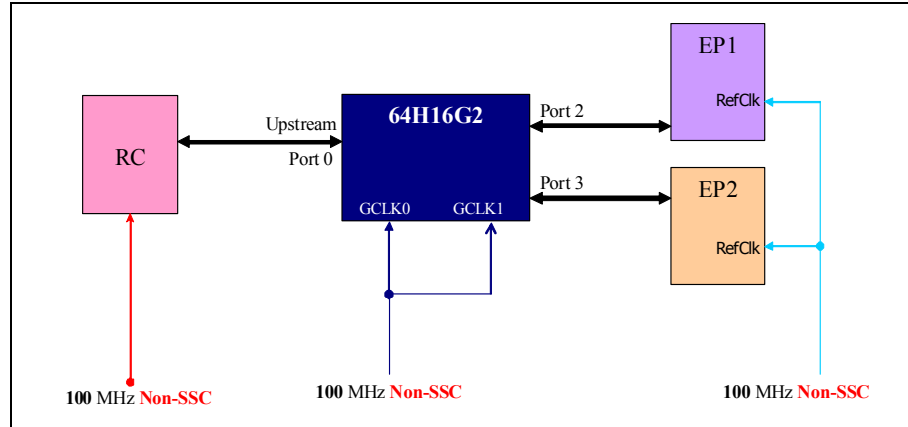


Figure 9 Example of Non-Common Clock Mode

Local Port Clocked Mode

The switch has a differential reference clock input (PxCLK) per port. Refer to the switch user manual for the number of supported PCLK input pins. This allows support for Spread Spectrum Clocking (SSC) either globally or independently on a per port basis. Figures 10 shows implementation of Local Port Clock mode with SSC on root complex via P0CLK.

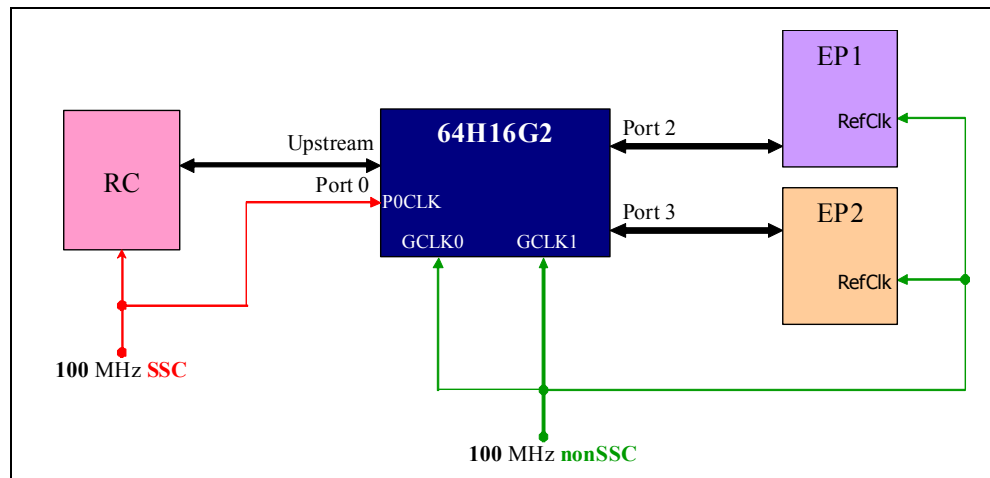


Figure 10 Example of Local Port Clock mode with SSC on Root Complex

Reset (Fundamental Reset) Schemes

The PERSTN pin is used to reset all logic inside the switch and is a Schmitt Trigger Input which can be connected to the PERST# from the system or a power-on reset circuit. In a system, the values of Tpvperl and Tperst-clk depend on the mechanical form factor in which the switch is used. For example, the PCIe Card Electromechanical Specification, Revision 2.0, specifies the minimum value of Tperst-clk=100µs and Tpvperl=100ms.

Notes

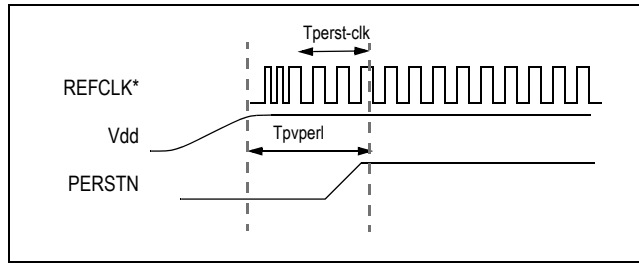


Figure 11 Fundamental Reset

For the reset signals to downstream ports, reset schemes listed below can be implemented.

- Simplified Reset Scheme
- Reset Scheme for Hot Plug Support

Simplified Reset Scheme

If Hot Plug support is not required, the simplified reset scheme can be implemented as shown in Figure 12. Add a buffer on the PERST# signal if the output from system is not able to drive a number of fan-out loads for all downstream ports.

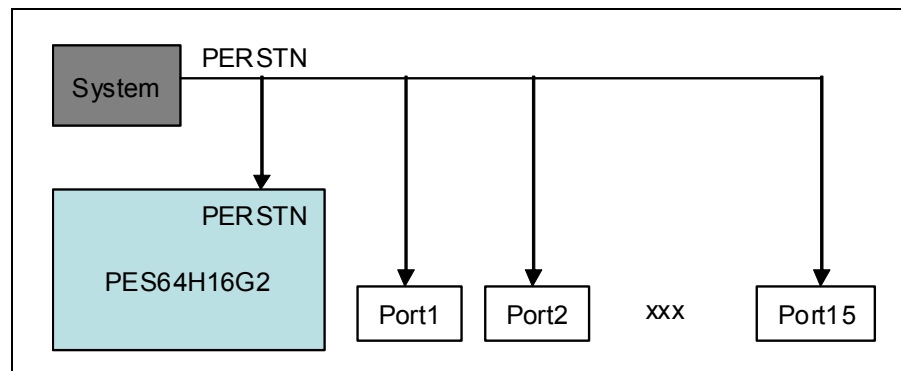


Figure 12 Simplified Reset Scheme

Reset Scheme for Hot Plug Support

Figure 13 shows an implementation where downstream endpoints have independent fundamental reset. This scheme should be used if Hot-Plug support is needed selectively on the downstream ports.

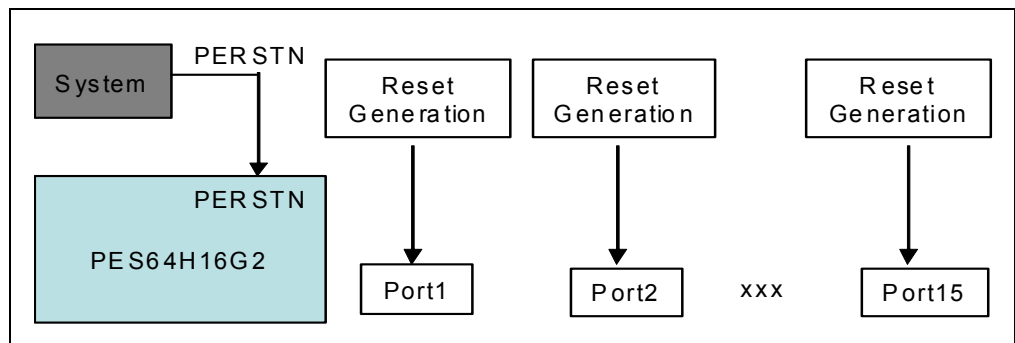


Figure 13 Reset Scheme for Hot Plug Support

Notes

RSTHALT

When this signal is asserted high during a PCI Express fundamental reset, the switch continuously returns Configuration Request Retry Completion Status (CRS) to Configuration Requests during the enumeration process. This allows the system BIOS via the SMBus to access internal registers before normal device operation begins. The device exits the RSTHALT state when the RSTHALT bit is cleared in the SWCTL register by a SMBus master. This RSTHALT mode is not required in most applications. The RSTHALT pin should be pulled down externally if the application does not use a SMBus master to initialize internal registers.

SMBus Interfaces

The switch provides two SMBus interfaces.

- The Master SMBus interface provides connection for an external serial EEPROM used for initialization and optional external I/O expanders.
- The slave SMBus interface provides full access to all software visible registers in the PES64H16G2, allowing every register in the device to be read or written by an external SMBus master. The slave SMBus may also be used to preload the serial EEPROM used for initialization.

Six pins make up each of the two SMBus interfaces. These pins consist of an SMBus clock pin, an SMBus data pin, and 4 SMBus address pins. In the slave interface, these address pins allow the SMBus address to which the device responds to be configured. In the master interface, these address pins allow the SMBus address of the serial configuration EEPROM from which data is loaded to be configured. The SMBus address is set up on negation of PERSTN by sampling the corresponding address pins.

Configuring Master SMBus clock frequency

The master SMBus clock's frequency can be configured by MSMBSMODE input pin. When this pin is pulled up, the master SMBus operates at 100 KHz. When this pin is pulled down, the master SMBus operates at 400 KHz.

Note that the master SMBus clock will be in the active state during loading of the EEPROM contents or while accessing the IO Expanders. Once the operation is completed, the master SMBus clock will be in the inactive state, which is high.

Configuring a serial EEPROM address

During a fundamental reset, a serial EEPROM is required to initialize any software visible register within the device. Serial EEPROM loading occurs if the Switch Mode (SWMODE [2:0]) field selects an operating mode that performs serial EEPROM initialization (e.g., Normal switch mode with Serial EEPROM initialization). The address used by the SMBus interface to access the serial EEPROM is specified by the MSMBADDR [4:1] signals, as shown in Table 1.

Bit	PES64H16G2	EEPROM
1	MSMBADDR[1]	A0
2	MSMBADDR[2]	A1
3	MSMBADDR[3]	A2
4	MSMBADDR[4]	0
5	1	1
6	0	0
7	1	1

Table 1 Serial EEPROM SMBus Address

Notes

Any serial EEPROM compatible with those listed in Table 2 can be used to store switch initialization values. EEPROM space may not be fully utilized because some of these devices are larger than the total available PCI configuration space that can be initialized in the switch.

Serial EEPROM	Size
24C32	4 KB
24C64	8 KB
24C128	16 KB
24C256	32 KB
24C512	64 KB

Table 2 PES64H16G2 Compatible Serial EEPROMs

If a serial EEPROM address is assigned to 0x50, for example, all of MSMBADDR[4:1] pins should be asserted low as shown in Figure 14.

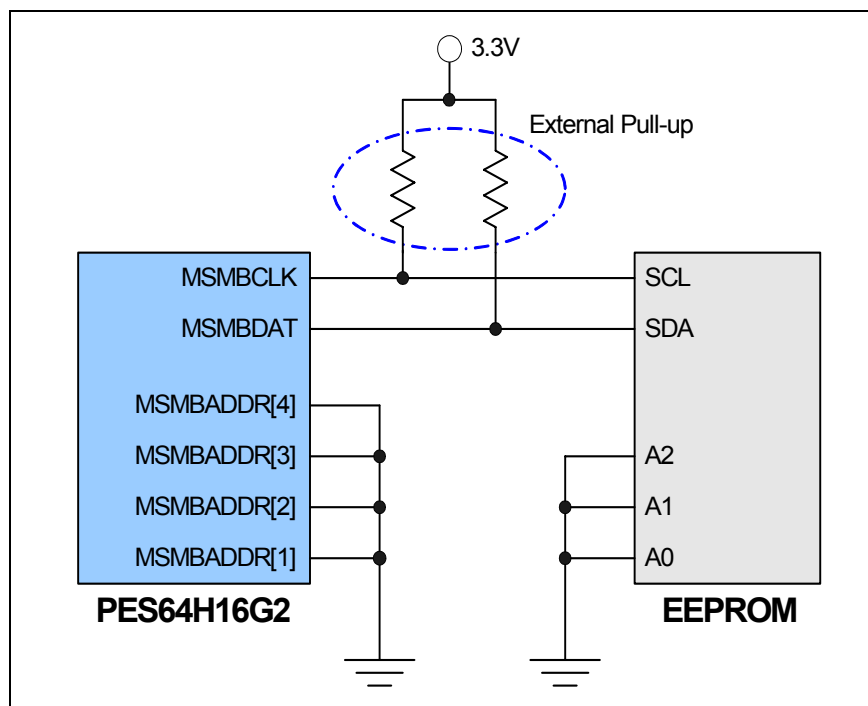


Figure 14 Example of EEPROM Address Configuration for 0x50

Notes

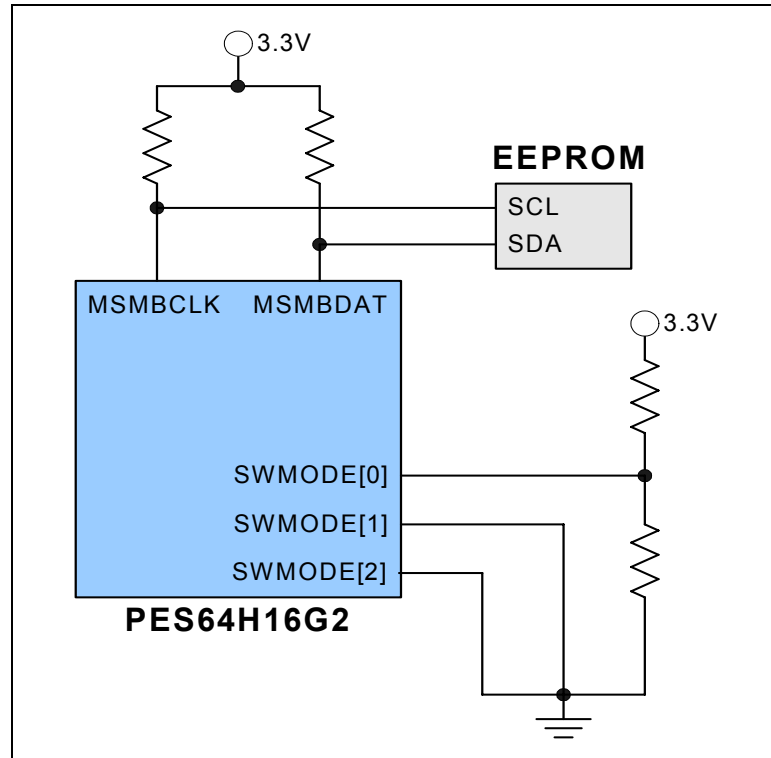


Figure 15 A Typical Implementation for SWMODE Selection and EEPROM Interface

Configuring the I/O Expander Address

The switch utilizes external SMBus / I²C-bus I/O expanders connected to the master SMBus interface for hot-plug and port status signals. The switch is designed to work with Phillips PCA9555 compatible I/O expanders (i.e., PCA9555, PCA9535, and PCA9539). See the Phillips PCA9555 data sheet for details on the operation of this device. For applications that require more than 8 I/O expanders, the MAX7311 is recommended since it is compatible with the Phillips PCA9555 and supports 64 slave addresses.

The switch supports up to fourteen external I/O expanders numbered 0 to 13. Refer to the device user manual for the allocation of functions to I/O expanders.

Figure 16 illustrates an example of the interface for I/O expanders 0, 2, and 12. During switch initialization, the SMBus/I²C-bus address allocated to each I/O expander used in that system configuration should be written to the corresponding I/O Expander Address (IOE[0,2,12]ADDR) field. The IOE[0,2]ADDR fields are contained in the I/O Expander Address 0 (IOEXPADDR0) register while the IOE[12]ADDR fields are contained in the SMBus I/O Expander Address 3 (IOEXPADDR3) register.

Hot-plug outputs and I/O expanders may be initialized via serial EEPROM. Since the I/O expanders and serial EEPROM both utilize the master SMBus, no I/O expander transactions are initiated until serial EEPROM initialization completes.

Notes

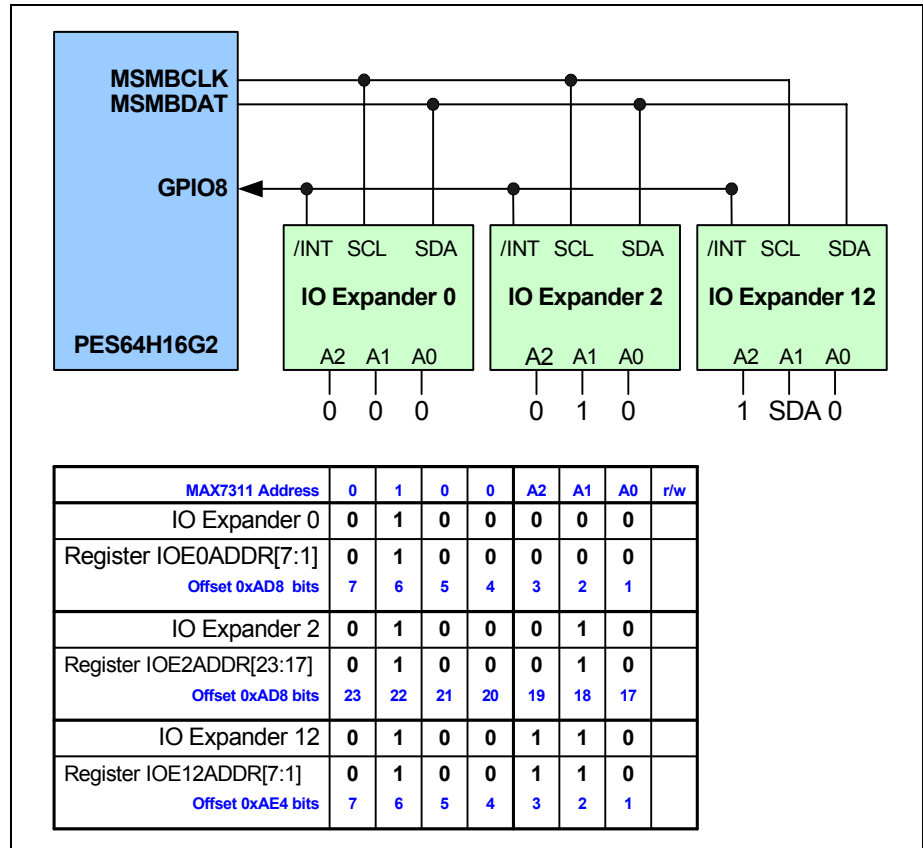


Figure 16 Example of I/O Expander Interface

Slave SMBus address interface

The slave SMBus interface provides the switch with a configuration, management, and debug interface. Using the slave SMBus interface, an external master can read or write any software visible register in the device. The address is specified by the SSMBADDR[5,3:1] signals as shown in Table 3.

Address Bit	Address Bit Value
1	SSMBADDR[1]
2	SSMBADDR[2]
3	SSMBADDR[3]
4	0
5	SSMBADDR[5]
6	1
7	1

Table 3 Slave SMBus Address

Notes

Power and Decoupling Scheme

The switch has five different types of power supply pins:

1. $V_{DD}CORE$ (1.0V) powers the digital core of the switch.
2. $V_{DD}PEA$ (1.0V) power the SERDES core and analog circuits. $V_{DD}PEA$ should have no more than 25mVpeak-peak AC power supply noise superimposed on the 1.0V nominal DC value.
3. $V_{DD}PEHA$ (2.5V) power the SERDES core and analog circuits. $V_{DD}PEHA$ should have no more than 50mVpeak-peak AC power supply noise superimposed on the 2.5V nominal DC value.
4. $V_{DD}PETA$ (1.0V) is the termination voltage used on the SERDES TX lanes. $V_{DD}PETA$ can be adjusted to modify the TX common mode voltage as well as the voltage swing.
5. $V_{DD}I/O$ (3.3V or 2.5V) powers the low speed IOs of the switch.

Note: The PES64H16G2 can work with 2.5V or 3.3V on the $V_{DD}I/O$. However, 3.3V is preferable.

$V_{DD}CORE$, $V_{DD}PEA$, and $V_{DD}PETA$ can be derived from the same voltage source with appropriate bypass capacitors and a ferrite bead. If all voltages can not be handled by a single voltage regulator, they can be derived from separate voltage regulators.

A ferrite bead can be used to attenuate the power noise and improve the analog circuit performance in a noisy environment. The following three parameters should be considered when you select a ferrite bead for power rails:

- Very low DC resistance
- Impedance of 50 ~ 120 ohms at 100 MHz
- Provides enough DC current

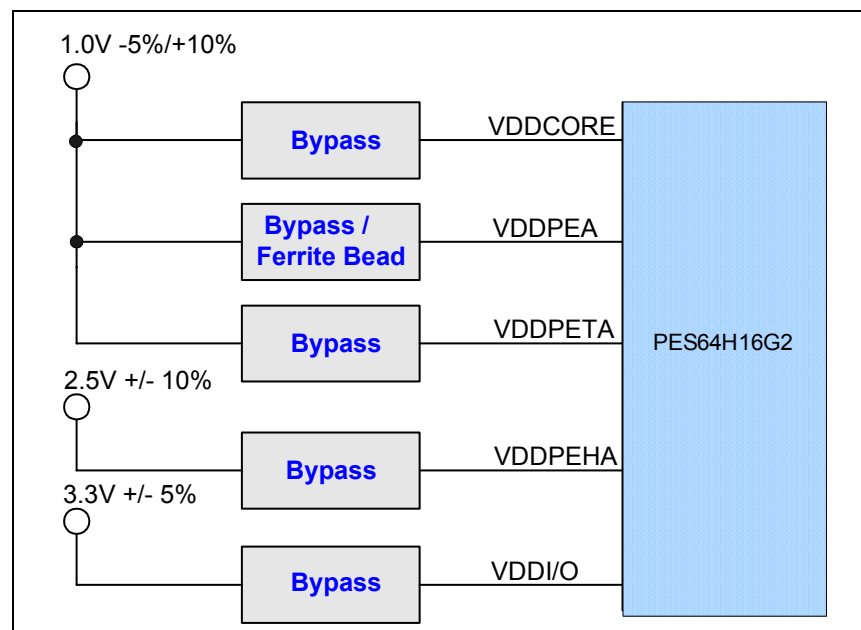


Figure 17 Board Power Supply Arrangement

Power Consumption

The typical and maximum power consumption can be found in the appropriate switch data sheet (see Reference Documents at the end of this guide).

Power-Up/Power-Down Sequence

During power supply ramp-up, $V_{DD}I/O$ must remain at least 1.0V above $V_{DD}CORE$ at all times. If $V_{DD}I/O$ is brought up first, then there is no problem. There are no other power-up sequence requirements for the various operating supply voltages. The power-down sequence can occur in any order.

Notes

Decoupling Scheme

1) One bypass capacitor per power pin is recommended if board layout allows. 0402 package ceramic capacitors are recommended for 0.1 μ F and 0.01 μ F capacitors.

2) Bypass Capacitors must be placed as close to the device pins as possible based on space availability. Note that some of the vias need to be shared in order to create space for placing a capacitor next to a pin.

3) A bigger capacitor should be used to filter out low frequency noise. Larger 1 μ F and 47 μ F capacitors should be added around the part. Two bigger capacitors per voltage supply are appropriate. One option is to spread out the big capacitors at four corners, top and bottom layers of the chips.

4) Short and wide traces should be used to minimize resistance and inductance.

5) Prioritize the bypass capacitors in the following order for each power supply:

1. $V_{DD}CORE$
2. $V_{DD}PEA / V_{DD}PETA$
3. $V_{DD}PEHA$
4. $V_{DD}I/O$

GPIO and JTAG pins

GPIO Pins

The switch has a number of General Purpose I/O (GPIO) pins that may be individually configured as general purpose inputs, general purpose outputs, or alternate functions. GPIO pins are controlled by the General Purpose I/O Function (GPIOFUNC), General Purpose I/O Configuration (GPIOCFG), and General Purpose I/O Data (GPIOD) registers in the upstream port's PCI configuration space. Please refer to the device data sheet for additional details.

The internal pull-up resistors value for the GPIO pins under typical condition is about 92K ohm.

JTAG Pins

The switch provides the JTAG Boundary Scan interface to test the interconnections between integrated circuit pins after they have been assembled onto a circuit board. For details of the interface, please refer to the appropriate switch user manual (see Reference Documents below).

The JTAG_TRST_N pin must be asserted low when the switch is in normal operation mode (i.e. drive this signal low with an external pull-down or control logic on the board if the JTAG interface is not used).

Switch Partitioning

The logical view of a PCI Express switch is shown in Figure 18.

Notes

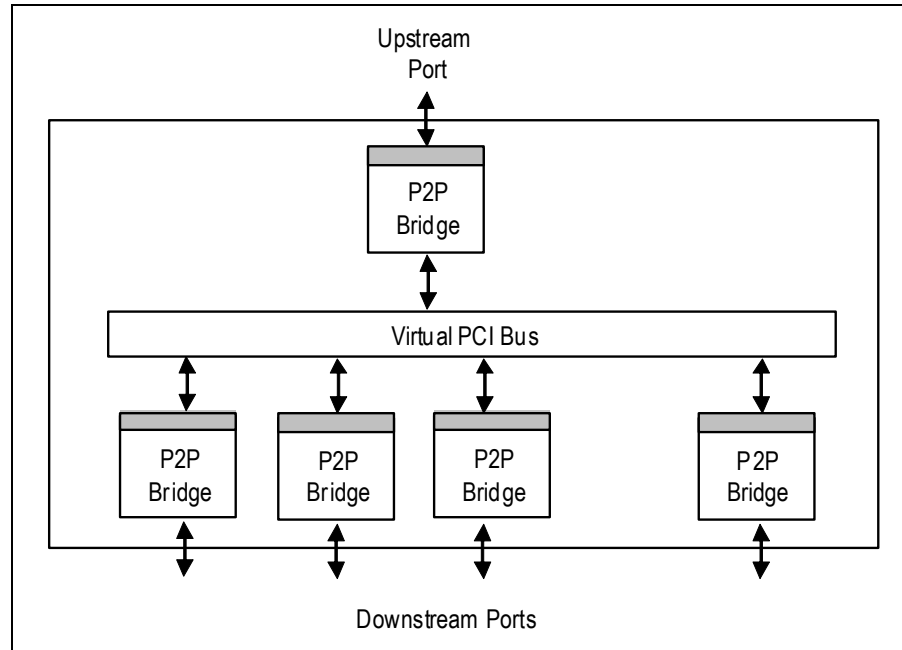


Figure 18 Transparent PCI Express Switch

The switch is a partitionable switch. This means that in addition to operating as a standard PCI express switch, the switch ports may be partitioned into groups that logically operate as completely independent PCI Express switches. Figure 19 illustrates a three partition configuration.

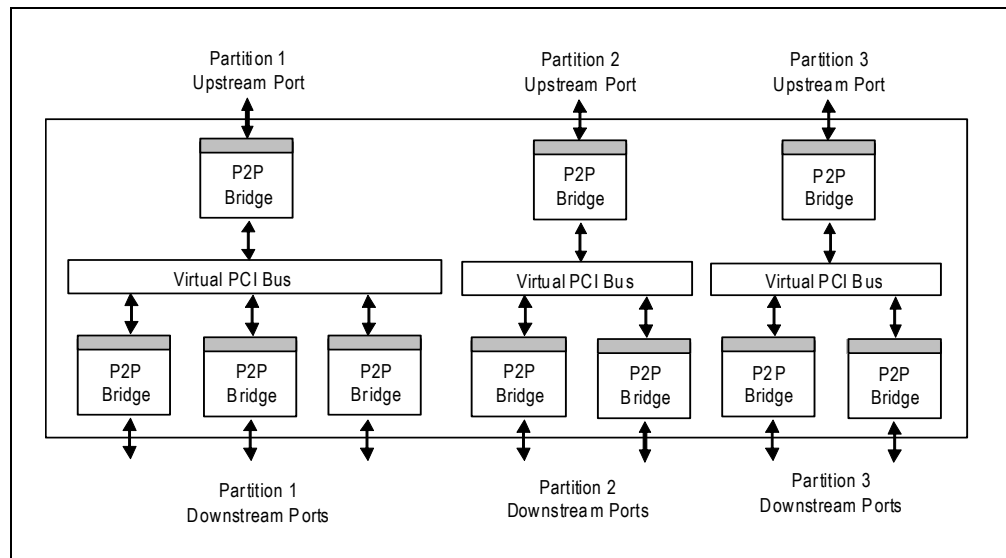


Figure 19 Example of Switch Partitioning

The configuration of partition states and port modes may be done statically (via serial EEPROM) or dynamically (via SMBus) during the fundamental reset sequence. However, static configuration is recommended.

Notes

Static Reconfiguration

The following is a sample EEPROM sequence to configure two partitions. Partition 0 has one upstream port (Port 0) and seven downstream ports (Ports 1 to 7). Partition 1 has one upstream port (Port 8) and 3 downstream ports (Ports 9, 10, and 11). Other ports are disabled. The sequence assumes the SWMODE signal in the boot vector is set to 0xD (i.e., Multi-partition with serial EEPROM initialization). In this mode, all partitions are initially disabled and all ports are initially unattached.

1. Set the following timer registers to a value of 0x0.
 - Side Effect Delay Timer (SEDELAY register)
 - Port Operating Mode Change Drain Delay Timer (POMCDELAY register)
 - Reset Drain Delay Timer (DRAINDELAY register)
 - Upstream Secondary Bus Reset Delay (USBRDELAY register)
2. Change the state of partition 0 to “Active” by setting the state field in the SWPART0CTL register.
3. Change the state of partition 1 to ‘Active’ by setting the state field in the SWPART1CTL register.
4. Add downstream ports to the partitions using the following sequence for each port.
 - Change the port operating mode by setting the following fields in the SWPORTxCTL register. This causes the port to be added to the selected partition
 - MODE field to ‘Downstream switch port’
 - PART field to the appropriate partition (e.g., 0 or 1)
 - OMA field to ‘no action’
 - Do a full link retrain on the port by setting the FLRET bit in the port’s PHYSTATE0 register. This will cause the port’s link to retrain from the Detect state.
5. Add the upstream port to each partition using the following sequence for each port.
 - Change the port operating mode by setting the following fields in the SWPORTxCTL register. This causes the port to be added to the selected partition.
 - MODE field to ‘Upstream switch port’
 - PART field to the appropriate partition (e.g., 0 or 1)
 - OMA field to ‘no action’
6. Disable the unused ports by using the following sequence for each port.
 - Change the port operating mode by setting the following fields in the SWPORTxCTL register. This causes the port to be disabled.
 - MODE field to ‘Disabled’
 - OMA field to ‘no action’
 - PART field to any value (this field is irrelevant for this port operating mode change)
7. Set the PCI Express capabilities and extended capabilities list for the upstream ports. This is done by configuring the Next Pointer (NXTPTR) field appropriately in the capability header register of the capabilities that form the two lists. This step is not required for the downstream ports.
 - Specifically, the following register fields must be set appropriately.
 - NXTPTR field in the PCI Power Management Capabilities (PMCAP) register
 - NXTPTR field in the PCI Express VC Extended Capability Header (PCIEVCCAP) register
8. If the application requires partition reset control via the PARTxPERSTN input signal, enable the appropriate GPIO alternate function as described in Chapter 17 of the PES64H16G2 User Manual.
9. Set the following timer registers to their default values.
 - Side Effect Delay Timer (SEDELAY register)
 - Port Operating Mode Change Drain Delay Timer (POMCDELAY register)
 - Reset Drain Delay Timer (RDRAINDELAY register)
 - Upstream Secondary Bus Reset Delay (USSBRDELAY register)
 - The above sequence must complete before the PCI Express hierarchy is enumerated (i.e., in less than 1 second from the de-assertion of the switch fundamental reset signal (PERSTN)). Refer to section Switch Fundamental Reset in Chapter 5 of the PES64H16G2 User Manual for details regarding switch fundamental reset timings.

Notes

Dynamic Reconfiguration

Dynamic reconfiguration refers to the reconfiguration of the switch partitions after the switch fundamental reset sequence completes (i.e., run-time reconfiguration). Possible partition reconfigurations are listed below.

- A downstream port is added or removed from a partition
- An upstream port is added or removed from a partition
- The operating mode of the upstream port is modified.

Partition reconfiguration may be initiated by software through modification of the operating mode of a port. A system may require software notification when a partition reconfiguration occurs. If the reconfiguration results in the addition, removal, or change in operating mode of the upstream port associated with the partition, then the system may be notified of the reconfiguration by a link down event detected by the component upstream of the partition (i.e., the root or switch downstream port). This form of notification requires that the OMA field in the SWPORTxCTL register be set to reset or hot reset.

Reference Documents

PES32H8G2, PES48H12[A]G2, PES22H16G2, PES34H16G2, PES64H16[A]G2 Data Sheets and Device User Manuals

PCI Express Base Specification, Revision 1.1 & 2.0, PCI-SIG

PCI Express Card Electromechanical Specification Revision 1.1 & 2.0, PCI-SIG

PCI to PCI Bridge Architecture Specification, Revision 1.2, PCI-SIG

SMBus Specification, Revision 2.0

Intel PCI Express Electrical Interconnect Design

Revision History

August 14, 2009: Initial publication.