Spartan-6 FPGA Connectivity Targeted Reference Design

User Guide

UG392 (v1.5) October 5, 2010

XILINX®

Revision History

The following table shows the revision history for this document.

Date	Version	Revision
12/08/09	1.0	Initial Xilinx release.
12/17/09	1.1	Updated Figure 1-1, page 12. Revised register and LUT utilization numbers in Table 1-1 and Table 1-2. Updated Software Bring Up in Chapter 2 section. Updated the MAC address to $ ./setmac_id ethX <SP605_MAC_ID>. Added clarification to Test Setup and Payload Statistics. Edited Using Various Features including removal of Checksum Offload section and updating NIC Statistics. Updated Packet Size. Revised the following command line instruction to: `$ ./implement_1000basex.sh`. Added a note above Figure 2-23. Minor edits to Figure 3-1, Figure 3-3, Figure 3-4, Figure 3-5, Figure 3-8, Figure 3-9 and Figure 3-11. Changed Checksum Field Mapping for C2SDescUserStatus[63:32] in Table 3-3. Clarifying edits to the Clocking Strategy section. Added Checksum Offload in Chapter 5. Revised some of the TRD file/folder names. Added troubleshooting tips in Table D-1.
05/13/10	1.2	Updated resource utilization numbers in Table 1-1 and Table 1-2. Added ProjNav Based Flow. Added the `html` folder to Figure B-1 and updated the design and driver discussion.
06/14/10	1.3	Updated DDR3 interface performance to 667 Mb/s in Figure 1-1 and on page 13. Updated LUT resource utilization numbers in Table 1-1 and Table 1-2. Revised Simulation Requirements and IP Cores with TRD sections. Revised the differential clock input requirement to 333.5 MHz on page 57 and updated the PLL settings in the `memc3_infrastructure.v` file. Also updated the Clocking Strategy section. Updated to 667 MHz clock in Figure 3-9. Updated the Theoretical Calculation section.
08/11/10	1.4	Removed references to specific ISE software releases in Hardware Test Setup Requirements and Simulation Requirements. In IP Cores with TRD, changed `gig_eth_pcs_pma_v10_4` to `gig_eth_pcs_pma_v10_5` and `mig_v3_4.xco` to `mig_v3_5.xco`.
10/05/10	1.5	Updated resource utilization numbers in Table 1-1 and Table 1-2. Updated description in Simulation Requirements. Updated core/netlist names and added information about `readme.txt` in IP Cores with TRD. Added Simulating the Design with ISim. Changed "Simulating the Design" to Simulating the Design with ModelSim and updated the description. Moved User Controlled Macros and Test Selection sections. Added Design Version Register (0x8000). Updated `0x8000` to `0x8004` in UserApp Advertisement Register (`0x8004`) heading.

Table of Contents

Chapter 3: Functional Description

Chapter 4: Performance Estimation

Chapter 5: Designing with the TRD Platform

Appendix A: Register Description

About This Guide

The Spartan®-6 FPGA Connectivity Targeted Reference Design delivers all the basic components of a targeted design platform for the connectivity domain in a single package. Targeted Design Platforms from Xilinx provide customers with simple, smart design platforms for the creation of FPGA-based solutions in a wide variety of industries.

This user guide details a targeted reference design developed for the connectivity domain on a Spartan-6 FPGA. The aim is to accelerate the design cycle and enable FPGA designers to spend less time developing the infrastructure of an application and more time creating a unique value-add design.

Guide Contents

This document contains the following chapters:

- Chapter 1, Introduction to the Reference Design provides an introduction to the targeted reference design and outlines its features.
- Chapter 2, Getting Started provides a quick-start guide to help the user get started with the hardware setup and simulation.
- Chapter 3, Functional Description provides a detailed architectural description of the hardware and software driver implemented. A separate software programming manual listing the software APIs is delivered with the software source.
- Chapter 4, Performance Estimation provides a detailed report on theoretical estimation and analysis of the design performance.
- Chapter 5, Designing with the TRD Platform provides suggested variations and add-ons to the platform to implement different functions.
- Appendix A, Register Description details certain registers programmed by the software driver. This section provides a comprehensive register programming manual.
- Appendix B, Directory Structure details the directory structure and the organization of various files and folders.
- Appendix C, Setting Up a Private LAN provides instructions on setting up a private LAN connection to enable testing with Netperf.
- Appendix D, Troubleshooting provides tips on areas to look for when debugging a design.

Additional Documentation

The following documents have been used as a reference to develop this Targeted Reference Design (TRD).

- UG654, *Spartan-6 FPGA Integrated Endpoint Block for PCI Express User Guide*
- Northwest Logic Packet DMA Backend Core User Guide
  - Product Sheet
- DS537, *XPS LL TEMAC (v2.03a) Data Sheet*
- UG388, *Spartan-6 FPGA Memory Controller User Guide*
- UG155, *LogiCORE IP Ethernet 1000BASE-X PCS/PMA or SGMII v10.4 User Guide*
- DS563, *PLBV46 Master Single (v1.00a) Data Sheet*
- MARVELL PHY 88E1111 Data Sheet
- UG526, *SP605 Hardware User Guide*
- Fedora Project
  - Netperf v2.4 - Manual
  - GTK+ 2.0 - Manual

Additional Support Resources

To search the database of silicon and software questions and answers or to create a technical support case in WebCase, see the Xilinx website at:

http://www.xilinx.com/support.

XILINX®

Chapter 1

Introduction to the Reference Design

The Targeted Reference Design

The targeted reference design demonstrates the key integrated components in a Spartan®-6 FPGA, namely the integrated Endpoint block for PCI Express®, the transceivers, and the memory controller working together in an application along with additional IP cores including the third-party (Northwest Logic) Packet Direct Memory Access (DMA) engine, Xilinx® Platform Studio LocalLink Tri-Mode Ethernet MAC (XPS-LL-TEMAC), and the Xilinx Memory Interface Generator (MIG) in the CORE Generator™ tool.

This chapter introduces the targeted reference design for Spartan-6 FPGAs and outlines its features.

Introduction

The block diagram of the Targeted Reference Design (TRD) is shown in Figure 1-1, page 12. This design is a PCI Express v1.1 compliant x1 Endpoint block showcasing these independent applications.

- Network interface card (network path), providing either:
  - GMII mode using an external Ethernet PHY (typically used to connect to Copper networks) or
  - 1000BASE-X mode using the GTP transceivers on the FPGA (typically used to connect to optical fiber Ethernet networks).

  This card allows the designer to connect to an external network and run networking applications including browsing web pages, telnet, and ftp sites.
- External memory interface over PCI Express—also referred to as the memory path.

  This application showcases data movement between system memory and DDR3 SDRAM through the Spartan-6 FPGA.

The TRD uses a bus-mastering packet Direct Memory Access (DMA) engine to offload processor data transfer overhead. The DMA works in conjunction with the integrated Endpoint for PCI Express and enables high-speed data movement between system memory and the FPGA.

The TRD framework is built with the integrated Endpoint and DMA blocks forming the foundation of an entire platform. The network path and memory path are two applications developed around this foundation.

The TRD uses the XPS-LL-TEMAC IP core along with a PLBv46 master and slave on the network path. The host system is connected to the integrated Endpoint and uses the PLBv46 interface to configure the registers of the XPS-LL-TEMAC. The Ethernet applications are demonstrated in two modes:

- GMII mode using the external PHY onboard
- 1000BASE-X mode using the Spartan-6 FPGA transceivers through an additional 1000BASE-X PCS-PMA IP core

The memory path showcases the integrated memory controller on a Spartan-6 FPGA. The controller communicates to onboard DDR3 SDRAM through MIG. Using the connected components, data can be written to and read back from the external memory.

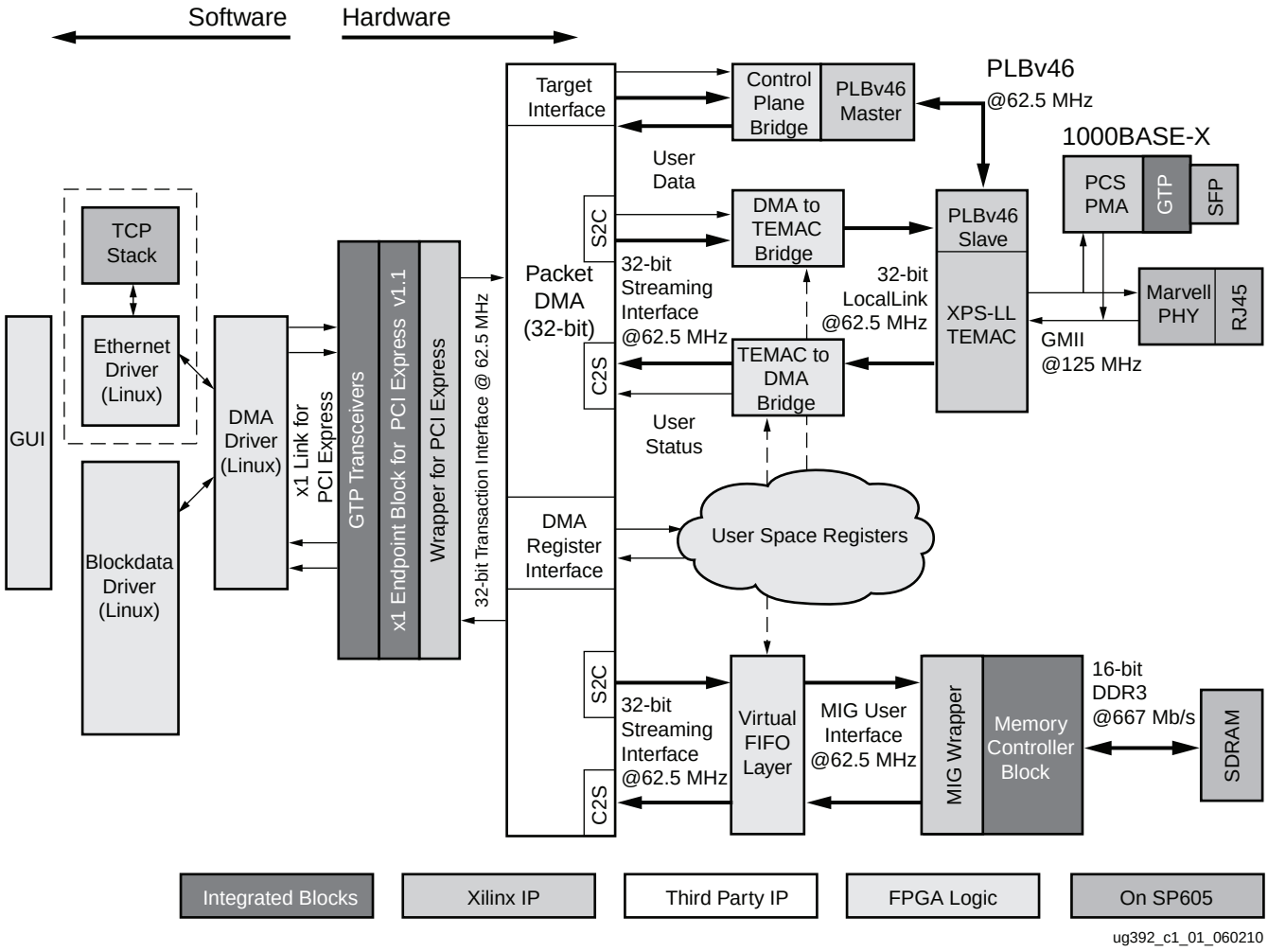


Figure 1-1: **Top-Level Design Overview**

Features

The design features are classified as either:

- Base Features: Features of the base components of the design including DMA and the integrated Endpoint block for PCI Express.
- Application Features: Features supported by each application developed on top of the base features.

Additionally, the source code of the software driver for a Linux platform is provided. The layered implementation of the software driver enables the designer to choose any layer (pertaining to specific hardware functionality) and customize it to specific requirements.

Base Features

This section lists the base features of the integrated Endpoint block for PCI Express and Packet DMA which form the backbone of the entire design.

- PCI Express v1.1 compliant x1 Endpoint block
  - One lane operation with GTP transceiver at a line rate of 2.5 Gb/s/direction
  - MSI and Legacy interrupt support
  - Transaction interface utilization engine for PCI Express
  - Optional high-performance mode utilizing extra block RAMs
- Bus Mastering Packet DMA
  - Multichannel operation
  - Packetized interface (similar to LocalLink)
  - Scatter Gather operation
  - DMA performance engine
  - Full-duplex operation (independent transmit and receive channels)

Application Features

The various application features are:

- Network Path Features
  - Ethernet address filtering
  - Ethernet statistics engine
  - TCP/UDP checksum offload
  - Auto-negotiation
  - Jumbo frames
  - Optional 1000BASE-X mode
- Memory Path Features
  - Integrated DDR3 controller providing up to 667 Mb/s performance
    - 16-bit single-component memory interface providing up to 10.672 Gb/s aggregate bandwidth
  - Reusable virtual FIFO interface with programmable depth

Resource Utilization

Table 1-1 and Table 1-2 list the resource utilization obtained from the map report during implementation phase. The XC6SLX45T-3-FGG484 is the target FPGA.

Note: The utilization numbers reported are obtained with the default options as provided in the top-level design. A change in default options will result in a change in utilization. The transceiver utilization is reported for the GTPA1_DUAL; one transceiver each is utilized from the GTPA1_DUAL (not both).

Table 1-1: **Resource Utilization—GMII Mode with External PHY**

Resource	Utilization	Total Available	Percentage Utilization
Slice registers	21,990	54,576	40
Slice LUTs	17,623	27,288	64
IOB	82	296	28
RAMB16BWER	35	116	29
RAMB8BWER	3	232	1
DCM	1	8	12
PLL_ADV	2	4	50
BUFG	10	16	62
GTPA1_DUAL	1	2	50

Table 1-2: **Resource Utilization—1000BASE-X Mode**

Resource	Utilization	Total Available	Percentage Utilization
Slice registers	22,439	54,576	41
Slice LUTs	17,662	27,288	64
IOB	58	296	19
RAMB16BWER	35	116	29
RAMB8BWER	3	232	1
DCM	1	8	12
PLL_ADV	2	4	50
BUFG	10	16	62
GTPA1_DUAL	2	2	100

The 1000BASE-X mode utilizes the PCS-PMA core, which increases the utilization of LUTs. Because the 1000BASE-X mode provides a serial interface as compared to a parallel interface with GMII mode, a corresponding reduction is seen in IOB usage. The 1000BASE-X mode uses an additional GTP transceiver.

XILINX®

Chapter 2

Getting Started

Introduction

This chapter is a quick-start guide enabling the user to test the Targeted Reference Design (TRD) in hardware with the software driver provided and also simulate it. It provides step-by-step instructions for testing the design in hardware.

Requirements

This section lists the minimum prerequisites of hardware testing and simulation.

Hardware Test Setup Requirements

The prerequisites for testing the design in hardware are:

- SP605 board with an XC6SLX45T-3 device
- ISE® software design tools
- USB JTAG for FPGA programming
- RJ45 cable and network connectivity
- PC with PCIe® v1.1 compliant slot, with at least 1GB of RAM and a 1MB cache
- Fedora 10 Linux (32-bit) OS corresponding with Linux kernel version 2.6.27
  - GTK+ v2.0 package required for GUI application
  - Netperf v2.4 (optional)

Simulation Requirements

This section lists the prerequisites for simulation:

- ISE software design tools (system or embedded edition with EDK installed)
- ModelSim v6.5c or later

Hardware Test Setup

This section details the hardware setup and use of the provided application GUI to help the user get started quickly with the design in hardware. It provides a step-by-step explanation on hardware bring-up, software bring-up, Ethernet bring-up, and using the application GUI provided.

All procedures listed require super-user access on Linux machines. When using the Fedora 10 Linux OS LiveCD provided with the kit, super-user access is granted by default due to

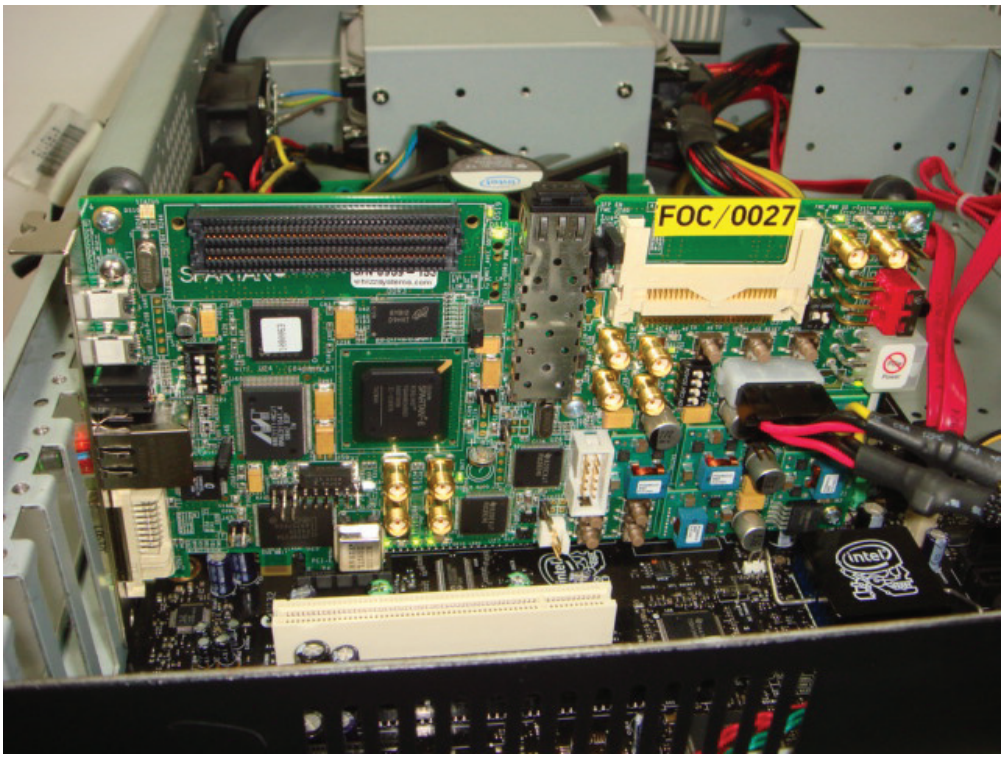
the way the kernel image is built; if not using LiveCD, contact your system administrator for super-user access.

Note: The changes made during LiveCD environment are not saved. For example, if the driver files were copied on the desktop after booting with LiveCD, they do not exist anymore after the system is shutdown and booted again with LiveCD. This is because the root file system for OS on LiveCD is created on system RAM and is not permanently on the hard disk.

Hardware Bring Up

This section lists the detailed steps for hardware bring-up.

1. With the host system switched to off, insert the SP605 board in the PCI Express® slot through the PCI Express x1 edge connector.
2. Connect the 12V ATX power supply 4-pin disk drive type connector to the board and also connect an Ethernet LAN cable in the RJ-45 slot provided. Do not power the board with both an external supply and the ATX supply at the same time.
3. Make sure the connections are secure so as to avoid loose contact problems. Power on the system.

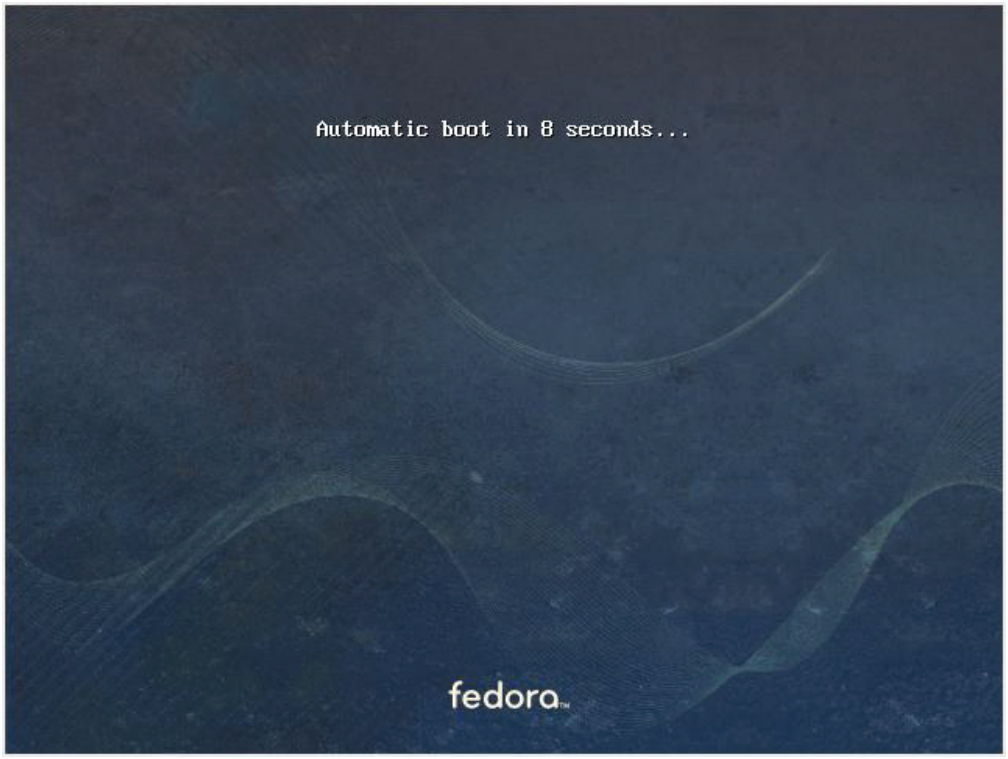


UG392_c2_01_120609

Figure 2-1: **Board Setup**

4. If using a PC with the Fedora 10 installed, proceed directly to step 7.

5. When using the Fedora 10 LiveCD provided in the SP605 kit, configure the desktop PC to boot from the CD ROM in the BIOS while the system starts. The Fedora 10 LiveCD provided in the kit is for Intel compatible PCs. The CD contains a complete bootable Fedora 10 Live environment and packages required for the targeted reference design demonstration. While the system boots from the CD, the screen is as shown in Figure 2-2.



UG392_c2_02_120609

Figure 2-2: **Fedora 10 LiveCD Boot Screen**

6. On the login screen (Figure 2-3), follow the instructions and proceed.

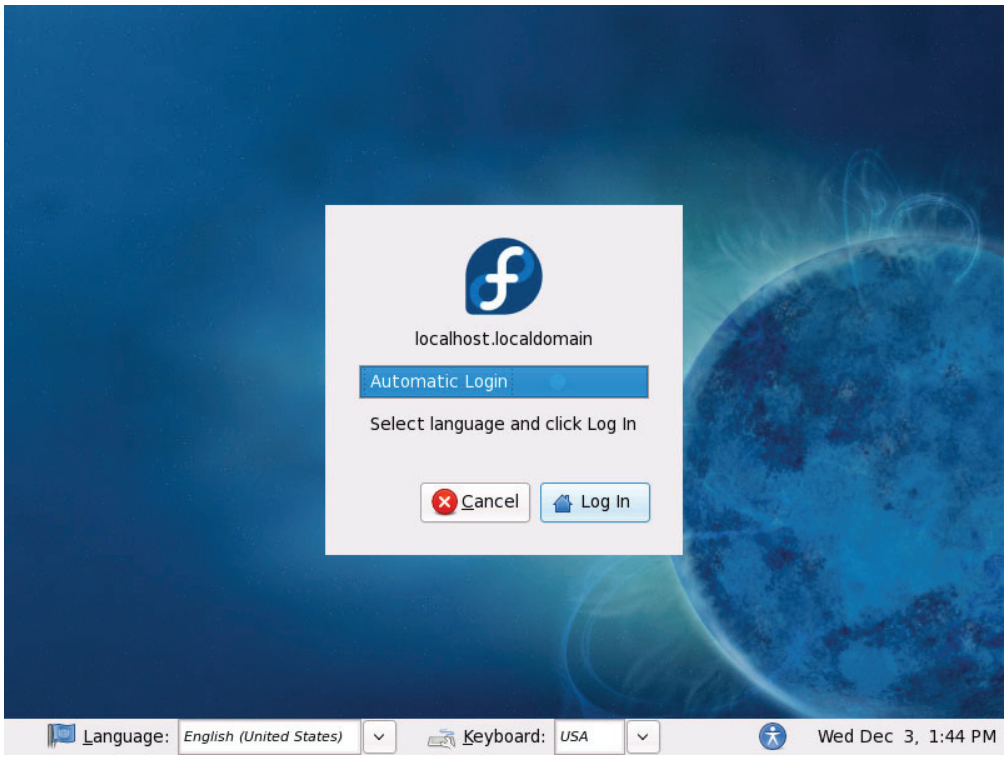


Figure 2-3: **Fedora 10 LiveCD Automatic Login**

7. Verify the hardware status through visual inspection. Make sure that the following indicators glow after powering on the system. Figure 2-4 indicates the location of the indicators.
   - PCI Express link up
   - Ethernet link status (seen on the left-hand side in Figure 2-4, side-view of the board)
   - Memory calibration complete

Figure 2-4 shows the entire design as mapped to hardware and the various components used on the SP605 board. The figure also highlights the visual indicators.

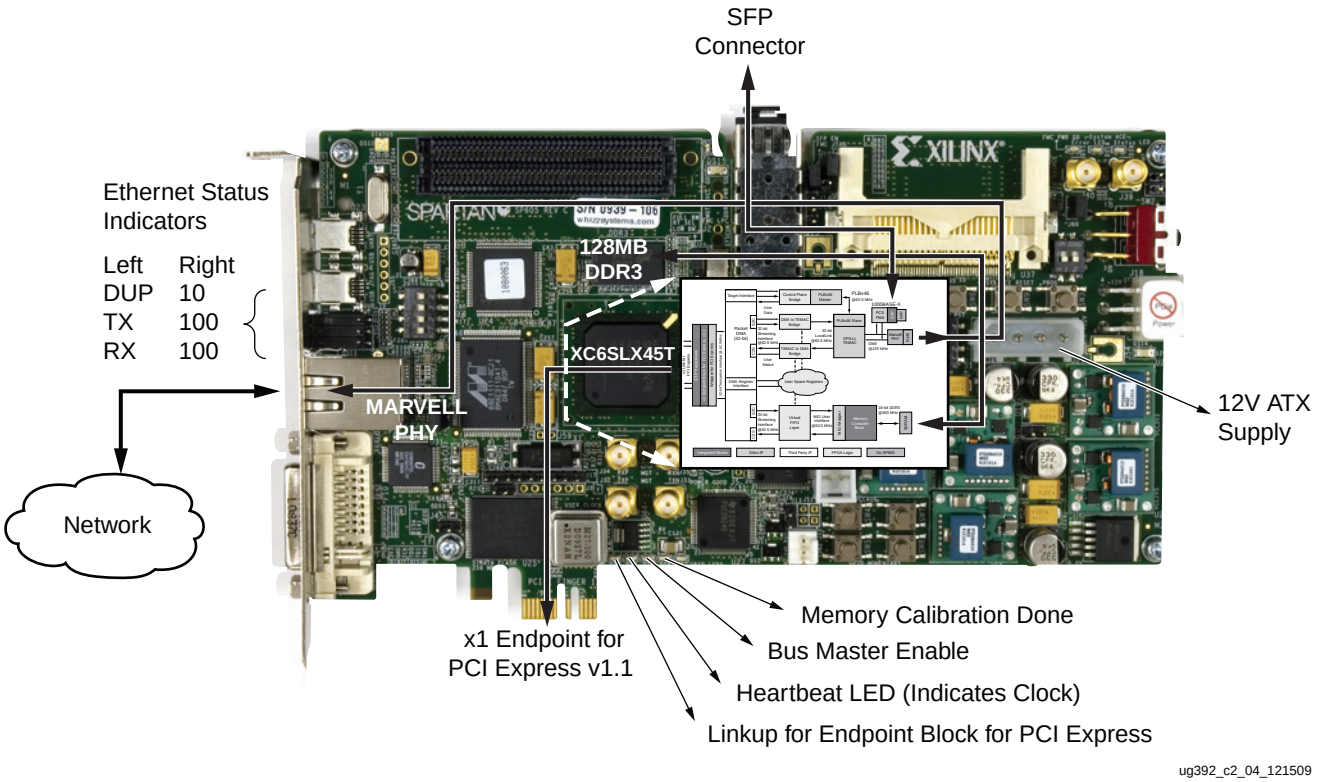


Figure 2-4: **Design on SP605 Board**

8. Check if the Endpoint block for PCI Express is detected by the system after the system boots by opening a terminal window (go to **Application→System Tools→Terminal**) and perform the following:

   At the terminal command line:

   ```
   $ lspci
   ```

   One of the entries should show the following information:

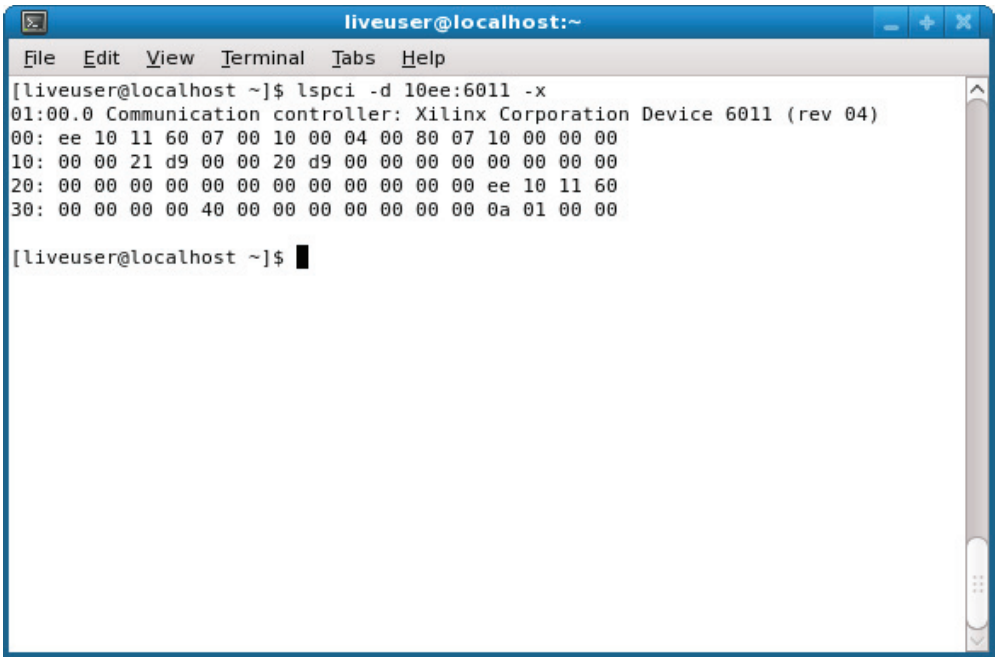
   ```
   01:00.0 Communication controller: Xilinx Corporation Device 6011 (rev 04)
   ```

   Where, `01:00.0` shows the PCI Express bus, device, and function number. As details in the output of the `lspci` command can vary from system to system, it is important that the Xilinx device is identified.

   To view details about the device configuration space, perform the following:

   ```
   $ lspci -d 10EE:6011 -v -x
   ```

   where, `10EE` is the vendor ID and `6011` is the device ID for the Endpoint. As details can vary from system to system, it is important to check that the vendor ID (`10EE`) and device ID (`6011`) match.



UG392_c2_05_120609

Figure 2-5: **Ispci Output**

To see more configuration space details type:

```
$sudo lspci -d 10EE:6011 -vvv -xxx
```

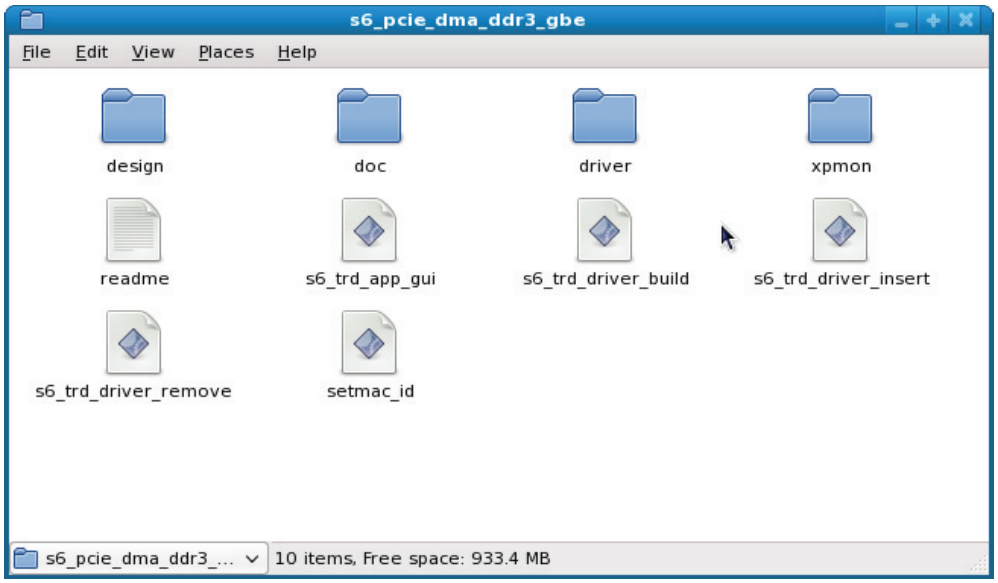
Software Bring Up

This section gives steps for bringing up the software driver, application, and using the various scripts provided.

1. After the device is detected, copy the `s6_pcie_dma_ddr3_gbe` folder from the USB flash drive provided in the Connectivity kit to the desktop (or a folder of choice). After the folder is copied, unmount and disconnect the USB drive.

   Double-click on the copied `s6_pcie_dma_ddr3_gbe` folder.

   The screen capture in Figure 2-6 shows the contents of the `s6_pcie_dma_ddr3_gbe` TRD folder.



UG392_c2_06_120609

Figure 2-6: **TRD Folder on the USB Flash Drive**

2. This step involves driver compilation and insertion of the kernel modules. Separate steps are defined for a command line user conversant with Linux or for a user preferring button-click operations.

Command Line Mode using Makefile

To compile and insert the driver, type at the command line in the terminal in the driver folder. Navigate to the `s6_pcie_dma_ddr3_gbe/driver` folder and follow the steps.

To clean the area:

```
$ make clean
```

To compile the files and build the kernel objects:

```
$ make
```

To insert the kernel object files:

```
$ make insert
```

Command Line Mode using Executable Scripts

The executable scripts provided can be run as-is on the command line:

For compilation of the driver modules:

```
$ ./s6_trd_driver_build
```

For insertion of the driver modules and to invoke the application GUI:

```
$ ./s6_trd_driver_insert
```

Mouse Click Driven Mode

To compile the driver, open the `s6_pcie_dma_ddr3_gbe` folder.

Double-click on **s6_trd_driver_build**. This cleans the area and builds the kernel objects. A window prompt appears (shown in Figure 2-7), click on **Run in Terminal** and proceed.

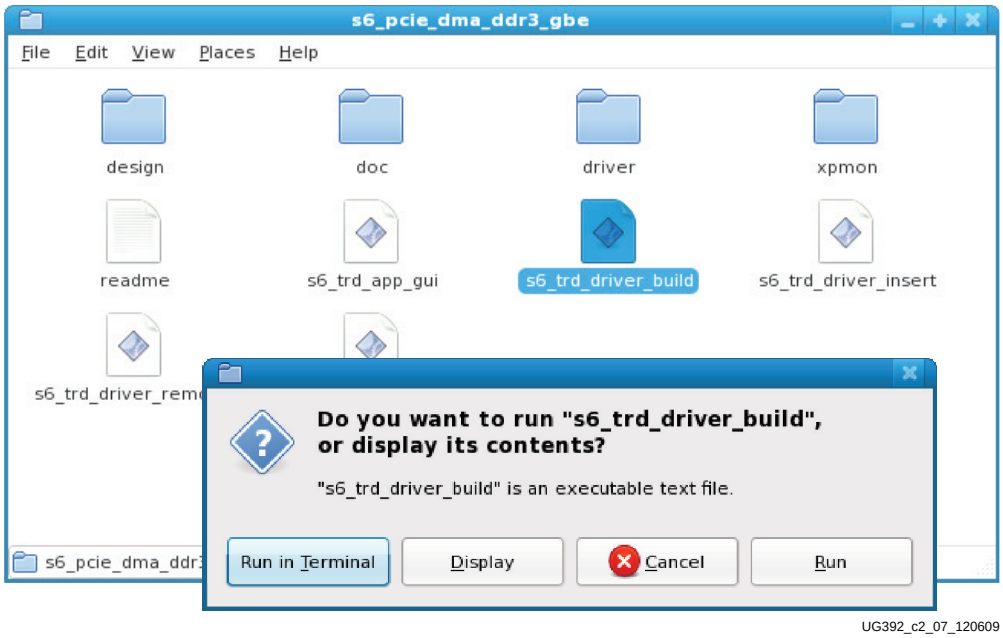


Figure 2-7: **Executing the Build Script**

Double-click on **s6_trd_driver_insert** (Figure 2-8). A Window prompt appears (shown in Figure 2-8), click on **Run in Terminal** and proceed. This inserts the driver modules into the kernel.

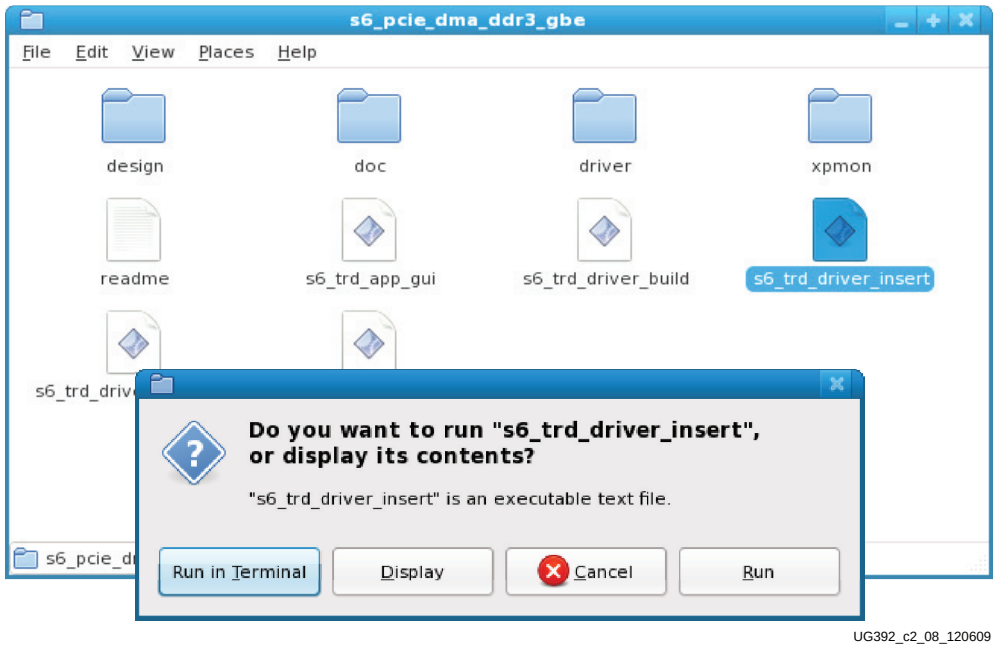


Figure 2-8: **Executing the Insert Script**

Proceed to Network Bring Up for network bring-up and GUI application usage.

3. To check if the driver modules are loaded into the kernel, try to use the `lsmod` utility in Linux. To see a list of kernel objects loaded:

   ```
   $lsmod | more
   ```

   `xblockdata_s6`, `xgbeth`, and `xdma_s6` modules are the Xilinx driver modules loaded as part of this TRD.

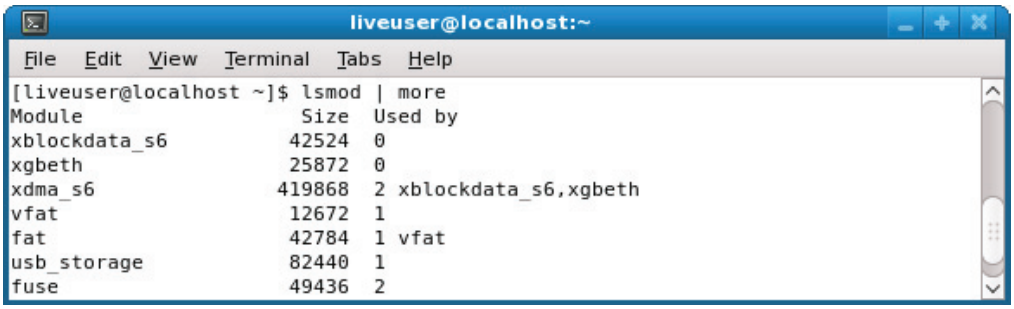


Figure 2-9: **lsmod Output**

Network Bring Up

After successful completion of the steps listed in Hardware Bring Up, page 16 and Software Bring Up, page 20, proceed to enable the Ethernet interface.

This section details Ethernet bring-up and uses the Network Configuration GUI on Linux.

1. To add a new network device, open the Network Configuration GUI (Figure 2-10)

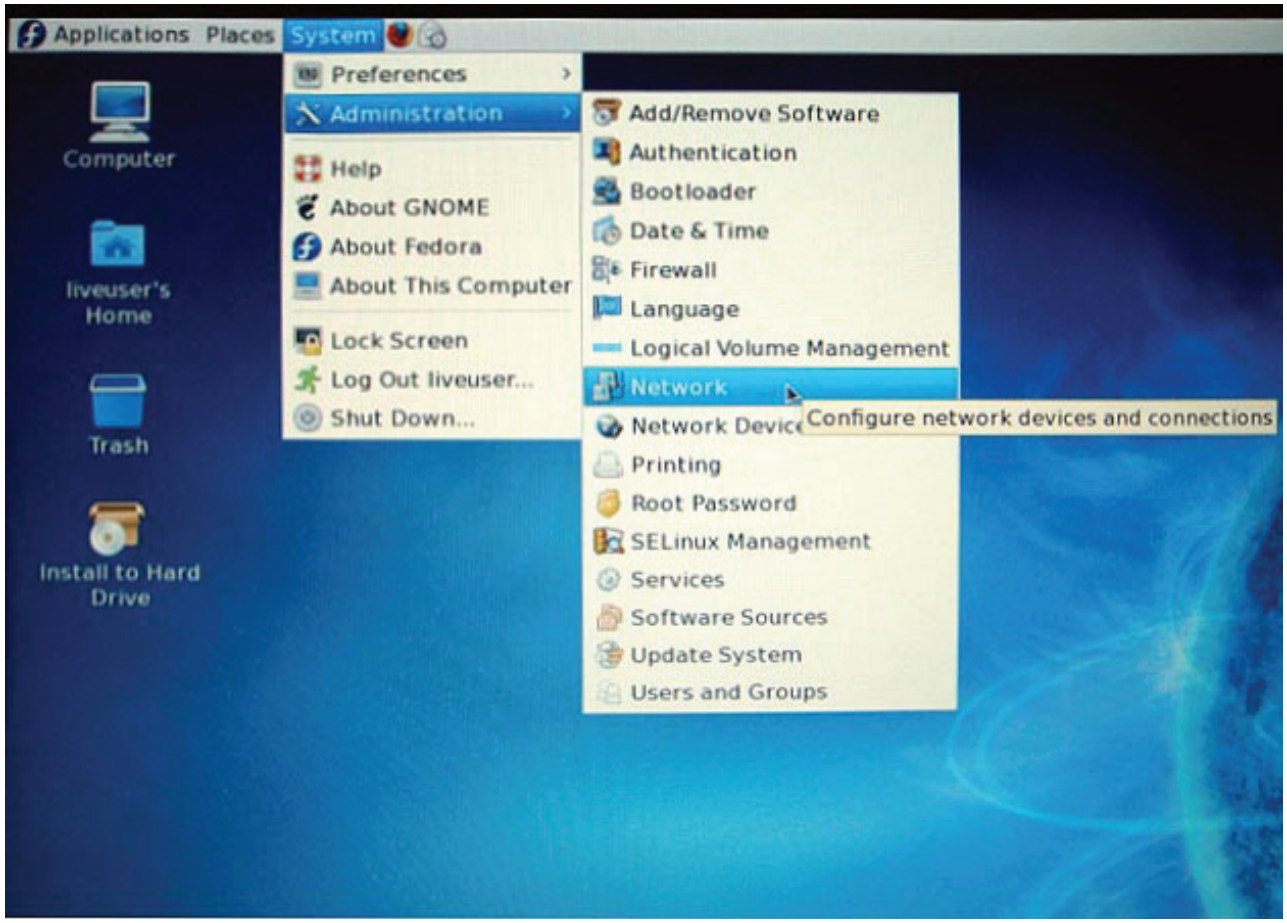


Figure 2-10: **Network Configuration GUI Path**

Note: If the LiveCD is not being used, invoking the Network Configuration GUI requires super-user password.

The GUI in Figure 2-11 appears.

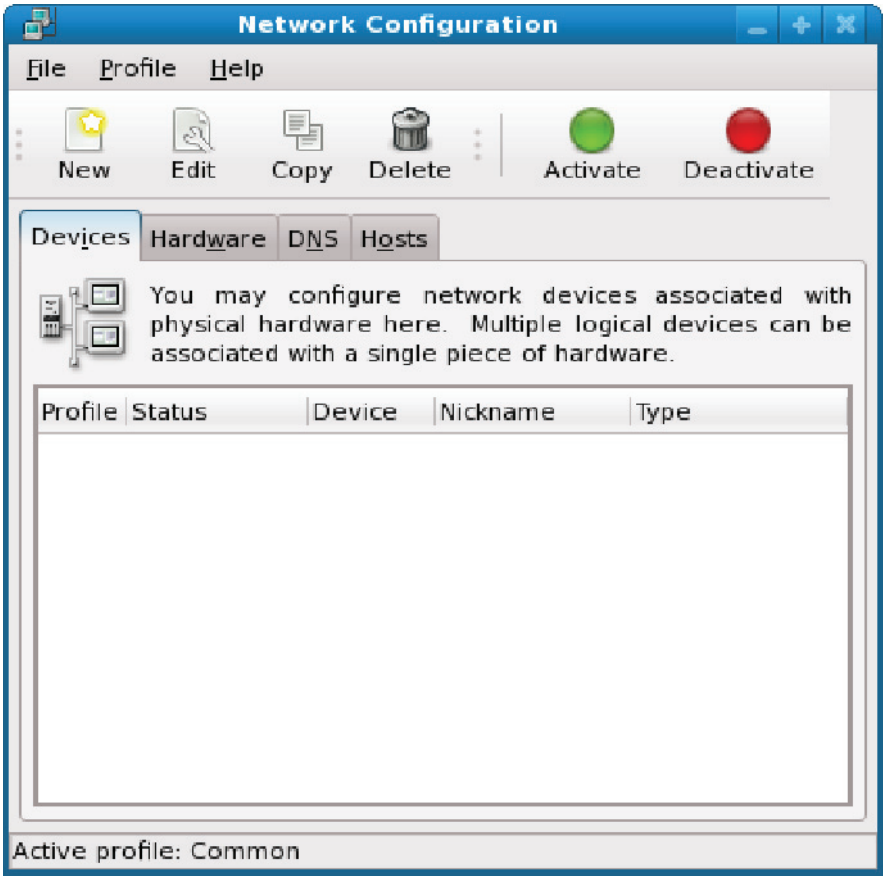


Figure 2-11: **Network Configuration GUI—Devices Tab**

With the LiveCD, the devices tab does not show anything as no interface is active. Make sure that the Ethernet cable is not connected to the existing LAN port on the PC.

Note: Activating multiple Ethernet ports on the same system requires each of the ports to be on different subnets. For this setup, do not connect any other network cable to the existing Ethernet ports on the PC. Connect the network cable only to the SP605 board RJ45 slot provided.

The **Hardware** tab shows the hardware devices present. As an example for this section, assume the NIC is detected as `eth1` as shown in Figure 2-12.

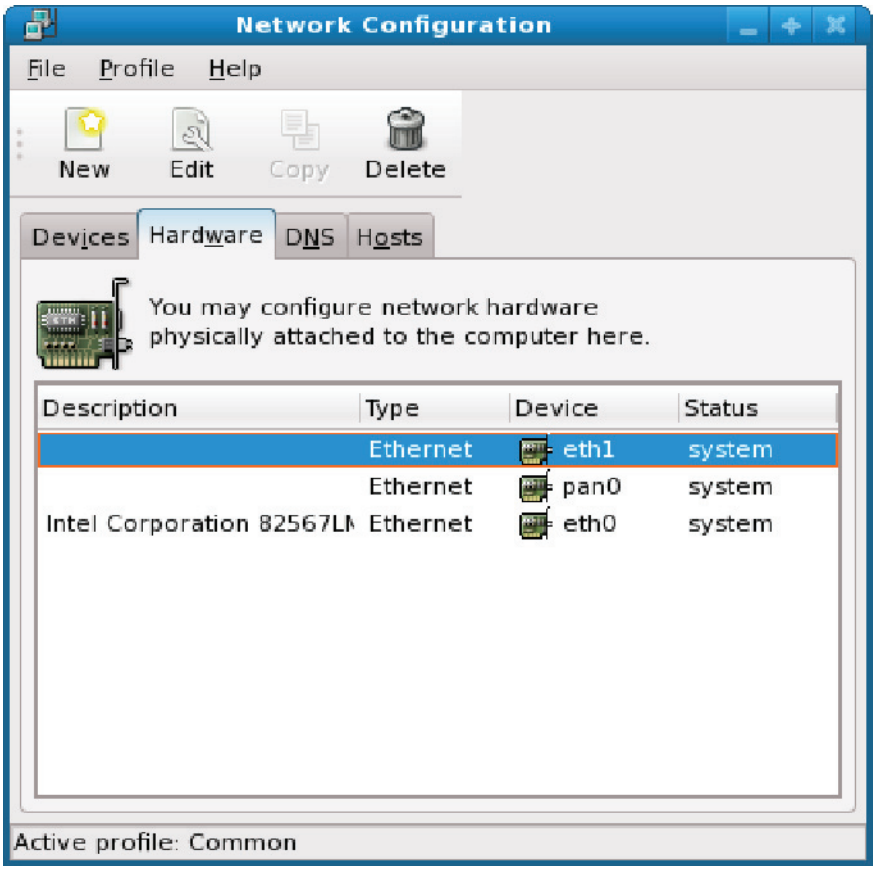


Figure 2-12: **Network Configuration GUI—Hardware Tab**

To have the SP605 NIC show up in the **Devices** tab, a new Ethernet connection needs to be created.

Click on **New→ Create New Ethernet Connection**.

Complete the setup process by following the instructions carefully on the screen. If an existing Ethernet connection(s) is available, this NIC is detected as an additional ethX interface where X would depend on the number of existing Ethernet interfaces.

The IP address assignment can be static or dynamic depending on the network setup. Contact the network administrator to understand IP address assignments on the network and also to obtain the necessary settings for network configuration.

Enter appropriate DNS settings for DHCP configuration as per the network administrator's instructions.

Once done, what is shown in Figure 2-13 should appear on the screen.

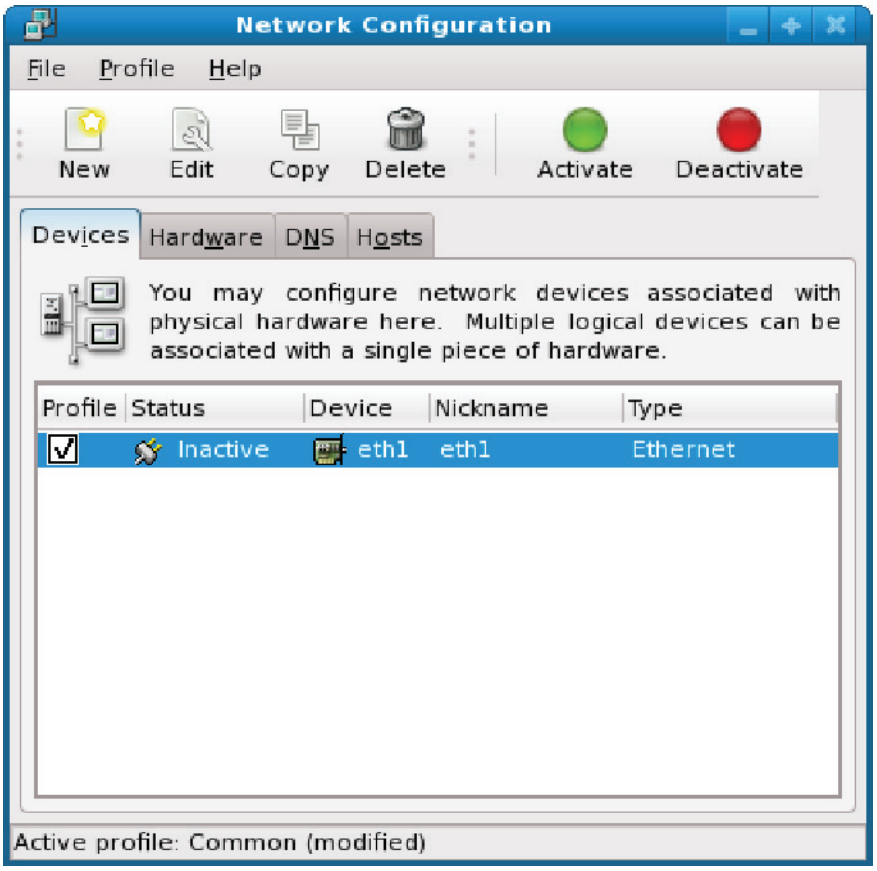


Figure 2-13: **Network Configuration after Adding Interface**

Save changes by clicking on the **File→Save Changes** option. Do not close the network configuration GUI as it will be required later to activate the interface.

2. MAC Address Assignment

   Use the `ifconfig` utility to check the device:

   ```
   $ /sbin/ifconfig eth1
   ```

   Initially the NIC is assigned an Ethernet interface `eth1` with a default MAC address of `AA:BB:CC:DD:EE:FF`.

   Make note of the MAC ID assigned to the SP605 board with your kit (Figure 2-14).

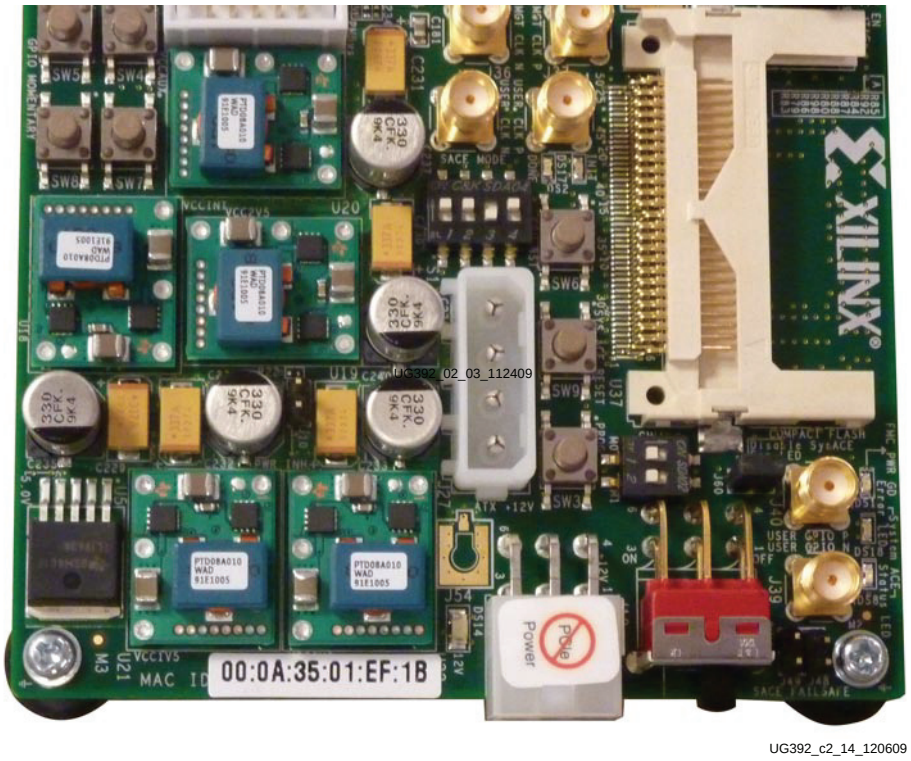


Figure 2-14: **MAC Address Label on SP605 Board**

Open a terminal and navigate to the `s6_pcie_dma_ddr3_gbe` folder. Set the MAC address for the Ethernet MAC in the design using the script provided:

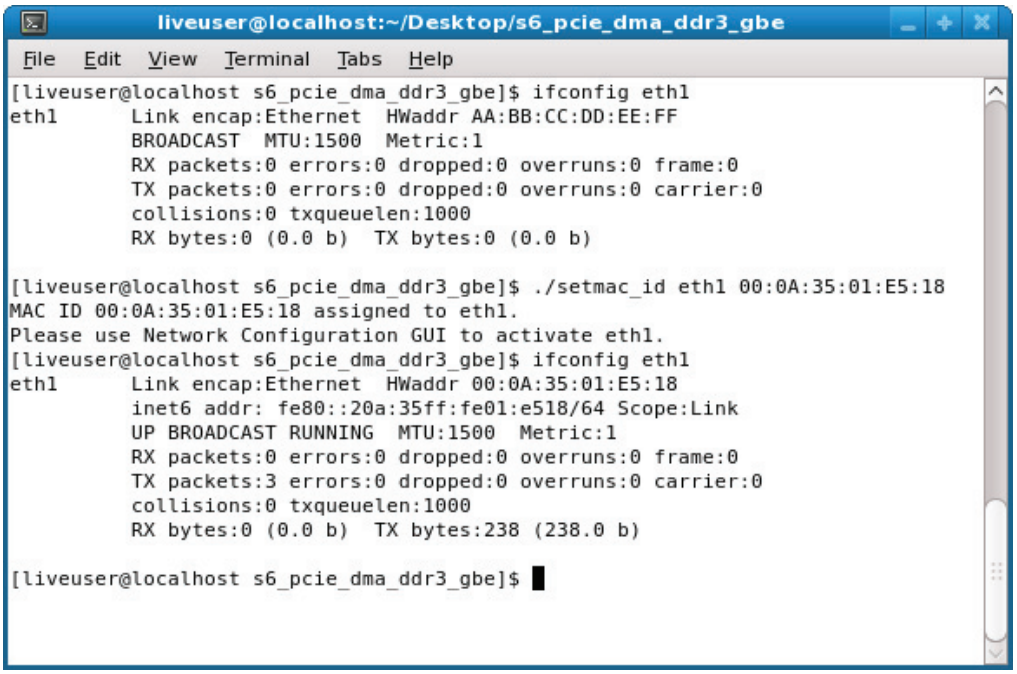
```
$ ./setmac_id ethX <SP605_MAC_ID>
```

For the MAC ID shown, the command would be:

```
$ ./setmac_id eth1 00:0A:35:01:EF:1B
```

After the MAC ID is assigned, try `ifconfig` again and the assigned MAC ID assigned to the NIC should appear.

Refer to the screen capture in Figure 2-15.



UG392_c2_15_120609

Figure 2-15: **MAC ID Assignment**

3. Invoke the Network Configuration GUI and activate the interface by clicking on the **Activate** button.
4. Invoke the browser and set the browser proxy settings as required for your network (contact your network administrator). Now online service should be available.

Using Application GUI

This section explains the application GUI provided and various features available.

1. This step involves GUI invocation. Separate steps are defined for a command line user conversant with Linux or for a user preferring button-click operations.

Command Line Mode Using Makefile

To compile and invoke the GUI navigate to the `s6_pcie_dma_ddr3_gbe/xpmon` folder, follow these steps:

To clean the area.

```
$ make clean
```

To compile the files.

```
$ make
```

To invoke the GUI.

```
$./xpmon
```

Command Line Mode Using Executable Scripts

The executable scripts provided in `s6_pcie_dma_ddr3_gbe` folder can be used as is on the command line:

```
$ ./s6_trd_app_gui
```

Mouse Click Driven Mode

Double click on **s6_trd_app_gui**. A window prompt appears (Figure 2-16), click on **Run in Terminal** and proceed.

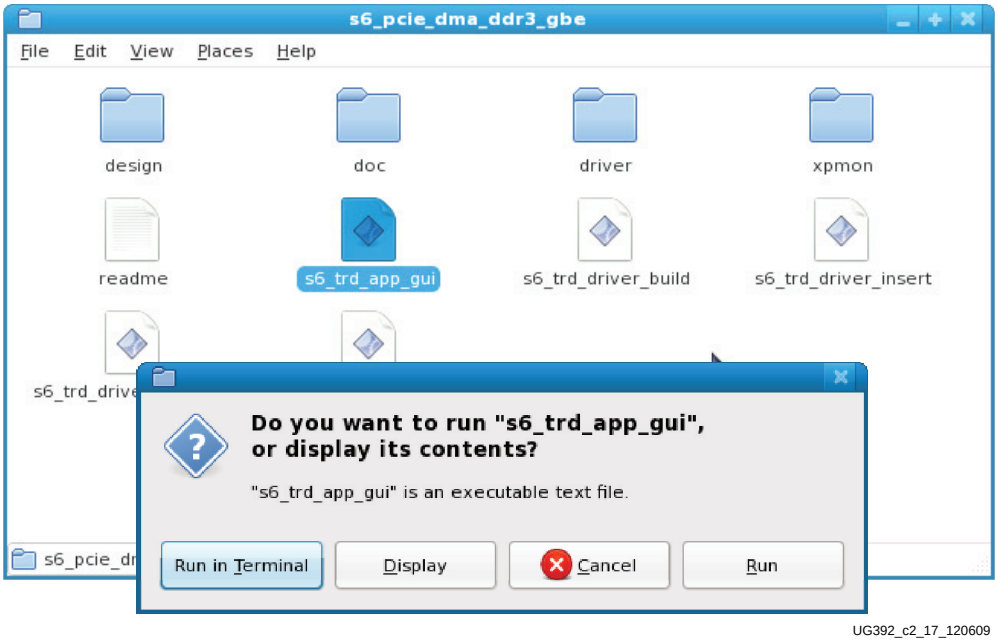


Figure 2-16: **Invoking GUI through Script**

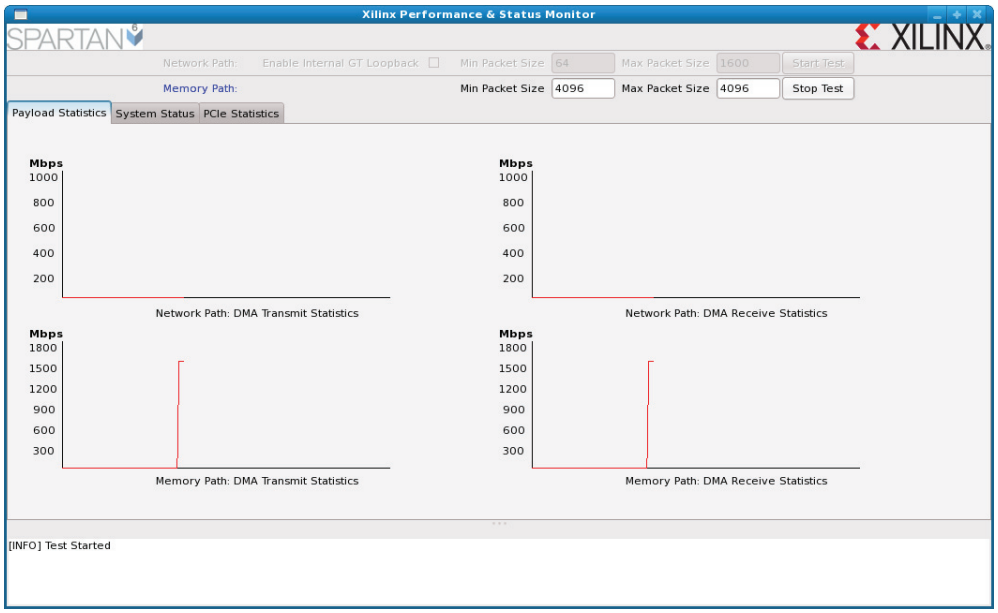
2. GUI walk-through screen-by-screen:

Test Setup and Payload Statistics

This screen defines the various test options provided for the memory path. The packet size for Ethernet can not be controlled by the application GUI as it is managed by the network (TCP/IP) stack.

There is an option to set the minimum and maximum packet size in bytes. While executing the test, the software driver builds packets of random length within the specified range. The range supported is 256-4096 bytes.

The screen in Figure 2-17 plots the data throughput obtained from the various DMA engines.

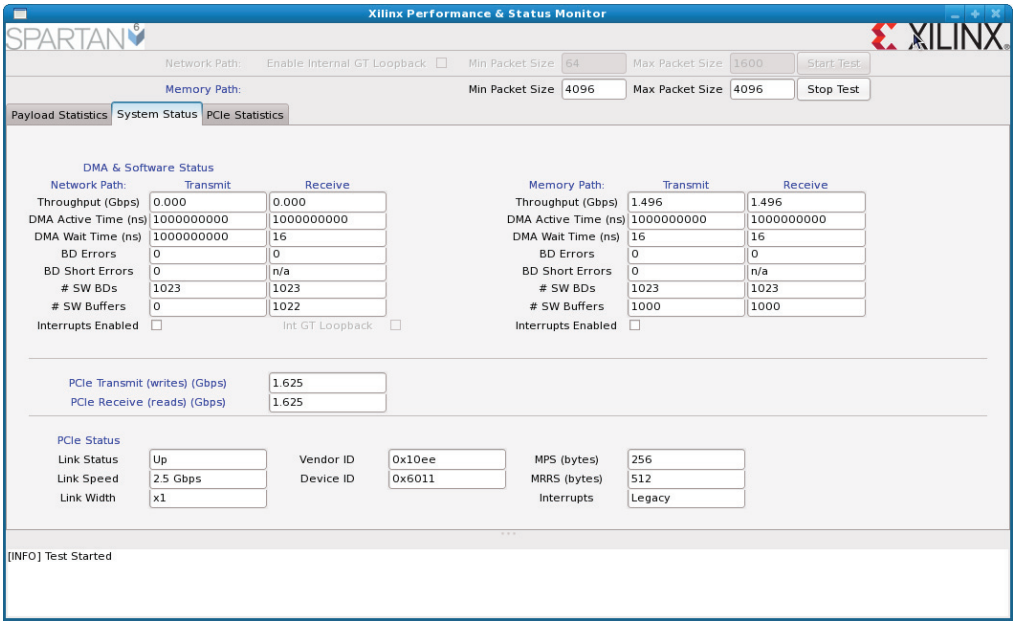


UG392_c2_18_120609

Figure 2-17: **Test Setup and Payload Statistics Screen**

System Status Screen

This screen gives the status of the PCI Express and DMA engine.

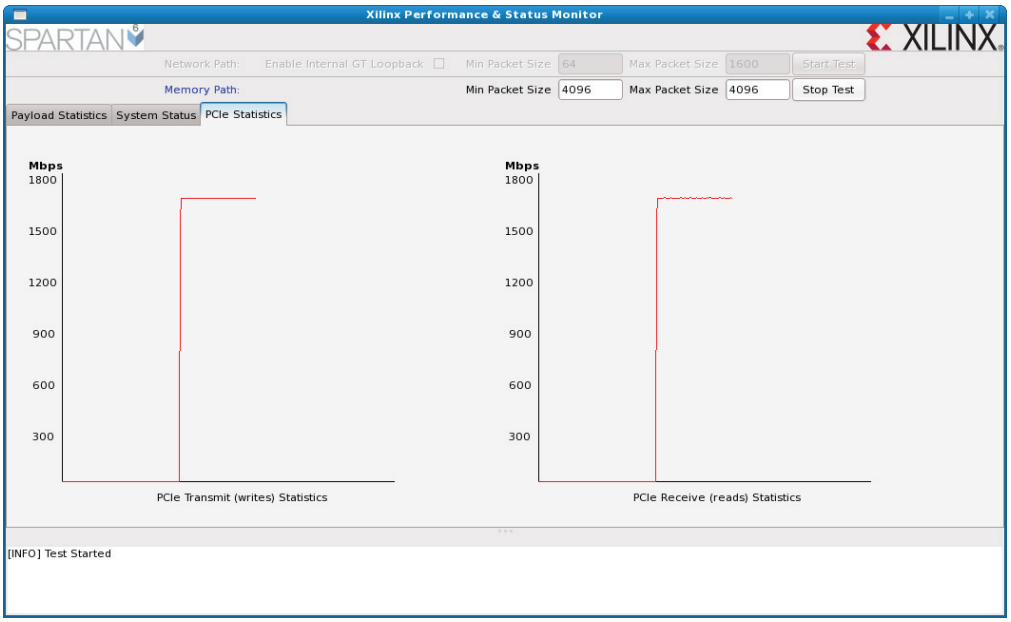


UG392_c2_19_120609

Figure 2-18: **System Status Screen**

Transaction Statistics

This screen plots the transaction utilization statistics on the PCI Express Transaction interface.



UG392_c2_20_120609

Figure 2-19: **Transaction Utilization Statistics**

Using Various Features

This section explains the various features of Ethernet configurable via standard tools or software driver macros. These features, when tested with LiveCD, will require the Linux commands given to be preceded by `sudo`.

Ethernet Specific Features

The Ethernet specific features can be exercised by using command line utilities like `ifconfig` and `ethtool` present in Linux.

The Ethernet driver provides functions which are used by `ifconfig` and `ethtool` to report information about the NIC. For reporting packet drops due to FCS errors, registers provided by the Ethernet Statistics IP are used.

The `ifconfig` utility is defined as the interface configurator and is used to configure the kernel-resident network interface and the TCP/IP stack. It is commonly used for setting an interface's IP address and netmask and disabling or enabling a given interface apart from assigning MAC address, and changing maximum transfer unit (MTU) size.

The `ethtool` utility is used to change or display Ethernet card settings. `ethtool` with a single argument specifying the device name prints the current setting of the specific device.

More information about `ifconfig` and `ethtool` can be obtained from the manual (man) pages on Linux machines.

NIC Statistics

The NIC statistics can be obtained using the `ethtool` command:

```
$ ethtool –S ethX
```

The error statistics are obtained by reading the registers provided by the Ethernet Statistics IP.

PHY registers (MARVELL PHY registers in case of GMII mode and PCS-PMA registers in case of 1000BASE-X mode) can be read using the following command:

```
$ ethtool –d ethX
```

Certain statistics can also be obtained from the `ifconfig` command:

```
$ ifconfig ethX
```

Autonegotiation

The tri-mode Ethernet MAC (TEMAC) is capable of operating at 10 Mb/s, 100 Mb/s or 1,000 Mb/s speeds. Autonegotiation is the process by which two devices capable of supporting different speeds choose a common speed and establish a link.

The following command results in the information about the Ethernet interface selected, which also lists the speed of operation currently selected.

```
$ ethtool -S ethX
```

To change the speed, use the following command and select a speed. If the device is currently operating at 1000 Mb/s, select the 100 Mb/s or 10 Mb/s options to set a different speed. The speed change should succeed if the link partner supports the newly advertised speed.

```
$ ethtool -s ethX speed [10|100|1000]
```

Promiscuous Mode

Enabling promiscuous mode disables address filtering in the TEMAC. This causes the driver to receive all the traffic which increases the CPU load.

The following command can be used to enable promiscuous mode:

```
$ ifconfig ethX promisc
```

The following command disables the promiscuous mode:

```
$ ifconfig ethX -promisc
```

Using a packet sniffer like Wireshark in promiscuous mode puts the driver and in turn the device in promiscuous mode.

Memory Application Specific Features

This section describes the feature usage specific to the memory path connected to external DDR3 SDRAM through the Spartan-6 Memory Controller block. The packet size feature is available through the GUI.

Packet Size

As there are no sidebands to store the start-of-packet and end-of-packet information in DDR3, the design builds packets of a programmed size when sending data to system memory. This packet size can be programmed by writing to a specific register in the memory path user-space register. The application GUI programs this register with an average of the minimum and the maximum packet size value entered by the user.

Shutting Down the System

The driver modules are automatically removed if the system is rebooted; however, the following steps are advised before shutting down the system for a graceful exit.

1. This step involves removal of the kernel modules. The steps are defined for both a command line user conversant with Linux and for a user preferring button-click operations.

Command Line Mode using Makefile

To remove the driver, type in the following command line in the terminal in the driver folder.

```
$ make remove
```

Command Line Mode using Executable Scripts

The executable scripts provided can be run as-is on the command line.

For removal of driver modules:

```
$./s6_trd_driver_remove
```

Mouse-Click Driven Mode

To compile the driver, navigate to the `s6_pcie_dma_ddr3_gbe` folder.

To unload the driver modules (Figure 2-20), click on **s6_trd_driver_remove**.

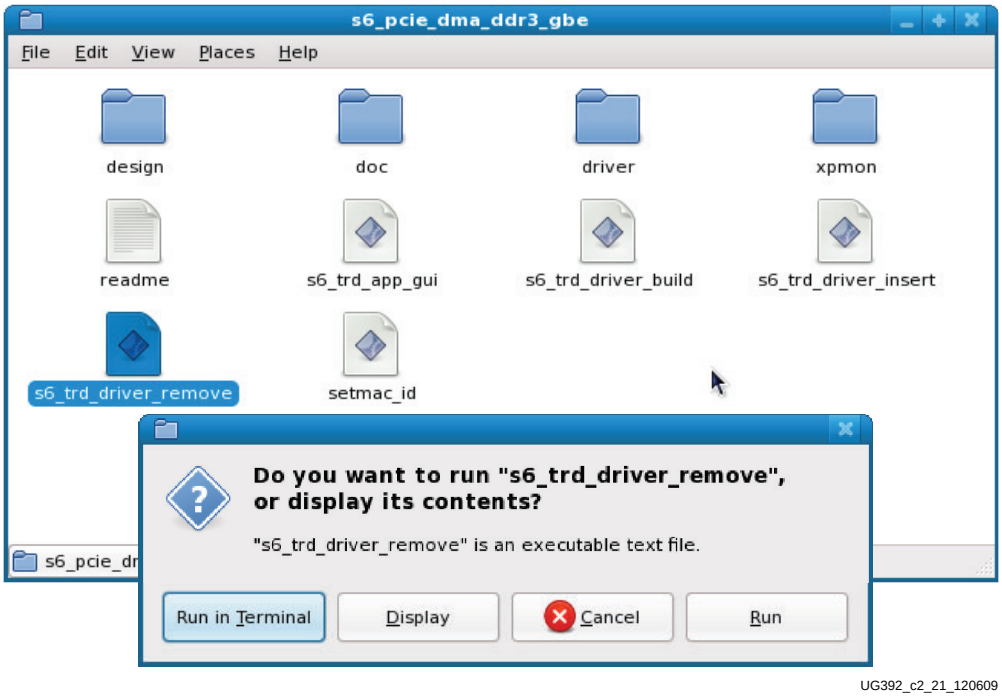


Figure 2-20: **Executing Driver Removal Script**

2. To shut down the system select the **System→ Shutdown** option. The LiveCD will be ejected. Follow the onscreen instructions. Any files saved during the LiveCD session are not accessible the next time LiveCD is run.

IP Cores with TRD

The Xilinx LogiCORE IPs required for the TRD are shipped with the TRD. The following cores/netlists are located in the `design/coregen_ip` directory:

- `s6_pcie_ip`
- `gig_eth_pcs_pma_ip`
- Netlist for FIFOs

The XPS-LL-TEMAC IP is present in the EDK install directory and is used from there.

MIG is generated from the CORE Generator tool.

Open a terminal window (on Linux) or a DOS command (on Windows) and navigate to the `design/coregen_ip` directory. Type the following on the command line:

```
coregen -b mig_ip.xco -p coregen.cgp
```

Additionally a golden set of XCO files are also provided under the `reference/xco_files` directory so that the cores can be regenerated.

Generating MIG cores overwrites the `mig_ip.xco` provided. To regenerate the core, copy the `mig_ip.xco` and `mig.prj` from the `design/reference/xco_files`.

Information on the version of IP cores used can be obtained from the `readme.txt` provided with the design directory.

A node-free hardware evaluation license is shipped with the TRD under the `design/license` directory.

Start the Xilinx License Manager and click on **Copy License** to install the license provided on your PC. For details on license installation, refer to UG665: *Spartan-6 FPGA Connectivity Kit Getting Started Guide* or http://www.xilinx.com/tools/faq.htm.

Implementing the Design

This section explains how to implement the design after making certain modifications. The implementation flow is provided in two modes, script based and ProjNav based (ISE software GUI flow).

Script Based Flow

Implementation scripts for both Linux and Windows operating systems are provided under the `design/implement` folder.

Implementation on Linux

1. Navigate to the `design/implement/lin` folder.
2. To generate a design using GMII through a MARVELL PHY, execute the following on a command line:

   ```
   $ ./implement_gmii.sh
   ```

   To generate a design with a PCS-PMA core in the 1000BASE-X mode, execute the following on a command line:

   ```
   $ ./implement_1000basex.sh
   ```
3. After completion of the implementation process, a results folder is created with the bitstream, all the reports, and intermediate implementation results.

Implementation on Windows

1. Navigate to the `design/implement/nt` folder
2. To generate a design using GMII through a MARVELL PHY, double-click on `implement_gmii.bat`
3. To generate a design with a PCS-PMA core in the 1000BASE-X mode, double-click on `implement_1000basex.bat`

After completion of the implementation process, a results folder is created with the bitstream, all the reports, and intermediate implementation results.

ProjNav Based Flow

The Spartan-6 FPGA Connectivity TRD provides files for a GUI-based flow, where designers can view the design hierarchy and source code. Execution of this flow requires a Xilinx ISE tools' installation, where both the XILINX and XILINX_EDK environment variables are defined.

- To implement a GMII design, navigate to the `design/implement/projnav_flow_gmii` folder.
- To implement a 1000BASE-X design, navigate to the `design/implement/projnav_flow_1000basex` folder.

The scripts for the GMII design flow are in `projnav_gmii.bat` and for the 1000BASE-X design flow are in `projnav_1000basex.bat`. Execution of these scripts:

- Generates the MIG IP core
- Generates the ISE design tools project files from the `Tcl` file provided (`xise` files)
- Opens the ISE software GUI and loads the relevant project

Note: Do not manually modify the `Tcl` or `xise` files.

- On Linux

```
$ ./projnav_gmii.bat
or
$ ./projnav_1000basex.bat
```

- On Windows

  Source the ISE environment and execute the `bat` file.

Once the ISE software GUI opens, a designer can manually click on the various options of synthesis, implementation, and bitstream generation, and work through the flow. Files generated during the flow are available in the same directory.

A copy of reference set of scripts is provided under the `design/reference` folder.

Programming the SP605

The SPI flash on the SP605 board is pre-programmed with the TRD file. This section provides a check on the on-board jumper settings after modification or use by another user.

Board Settings

This section shows the jumper settings required on the board (Figure 2-21) for the TRD to work when programmed in SPI x4 Flash mode.

1. Set the mode switch SW1 to `01` (M1 = 0 and M0 = 1).
2. Jumper J46 is required to be ON for the FPGA to be programmed with the onboard SPI x4 flash.
3. All other jumpers and switches are required to be in default position. Refer to UG526, *SP605 Board User Guide* for default jumper/switch settings.

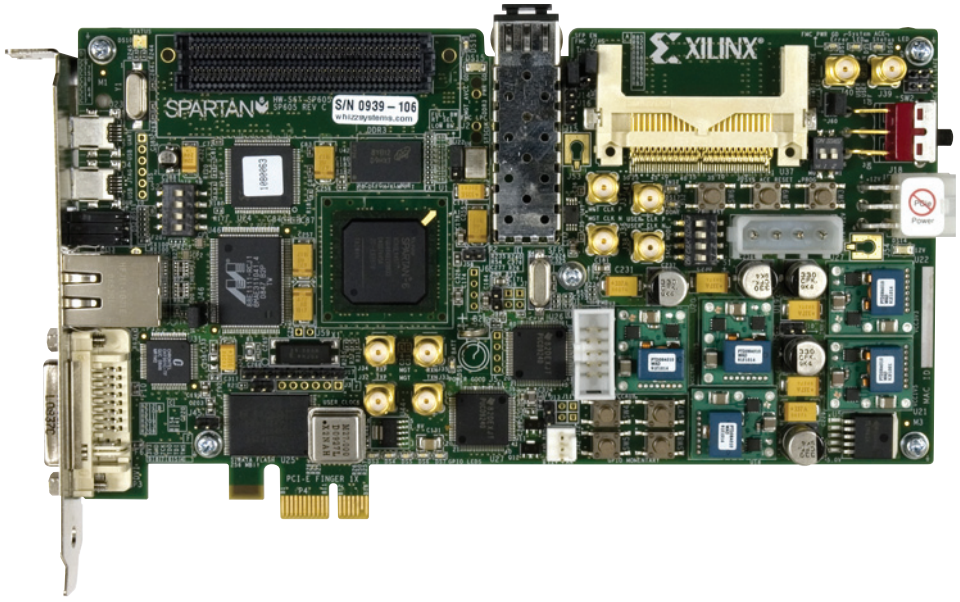


Figure 2-21: **Board Setup**

Board Programming

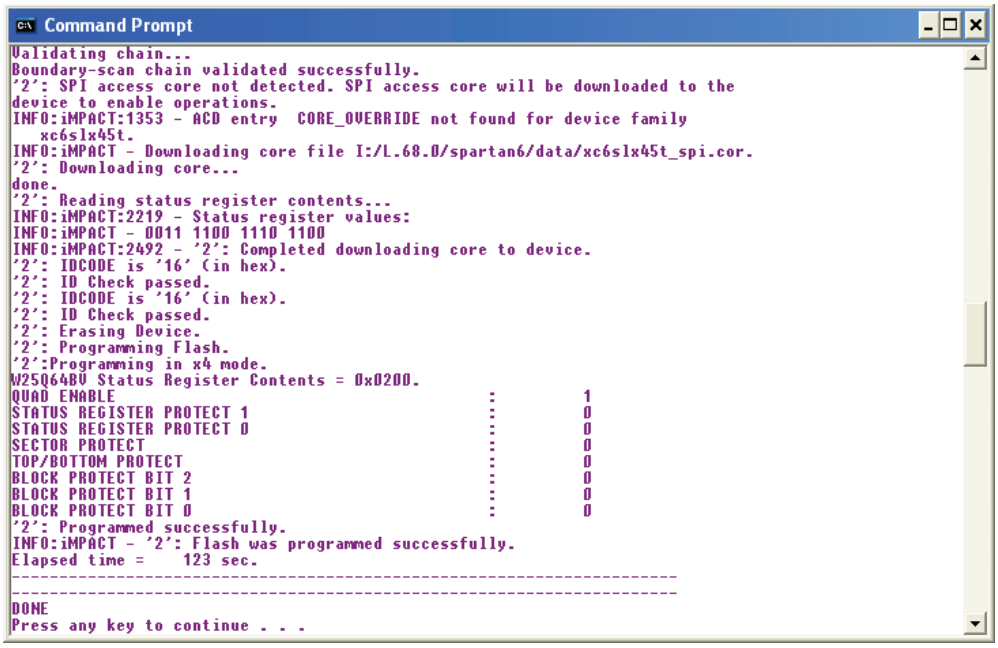
This section explains the programming of the SP605's onboard SPI x4 Flash.

1. Connect the USB download cable between the SP605 board and the PC. The SP605 board is required to be powered ON for programming. Use the power-supply brick provided with the connectivity kit for this purpose.
2. Navigate to the `design/reference/configuration` folder and copy the flash programming file to this folder.
3. Open the file `spi_program.cmd` and modify the name of the MCS file to be programmed to be the same as the one copied.

   On Linux, execute the following on the command line of a terminal window:

   ```
   $ impact -batch spi_program.cmd
   ```

On Windows, double-click on the `spi_program.bat` file. Programming on Windows opens the DOS command window as shown in Figure 2-22.



UG392_c2_23_120609

Figure 2-22: **Programming Status on Windows**

Testing 1000BASE-X Mode

The design supports the 1000BASE-X mode of operation. To test this, the additional requirements are to use an SFP to RJ45 adapter (for a copper cable) or an SFP optical adapter (for an optical cable). These items are not part of the SP605 connectivity kit.

To test the 1000BASE-X mode of operation, program the device with either `sp605_use_1000basex.bit` or `sp605_use_1000basex.mcs` provided under the `reference/configuration` folder.

Jumpers J44 and J22 are required to be in default position. Refer to UG526, *SP605 Board User Guide* for default jumper settings.

To test the 1000BASE-X mode of operation, the Ethernet driver needs to be compiled by enabling an additional macro.

- Navigate to the `driver/xgbeth` folder
- Open the **Makefile** and add **-DUSE_1000BASEX** to the end of the line defining **EXTRA_CFLAGS**

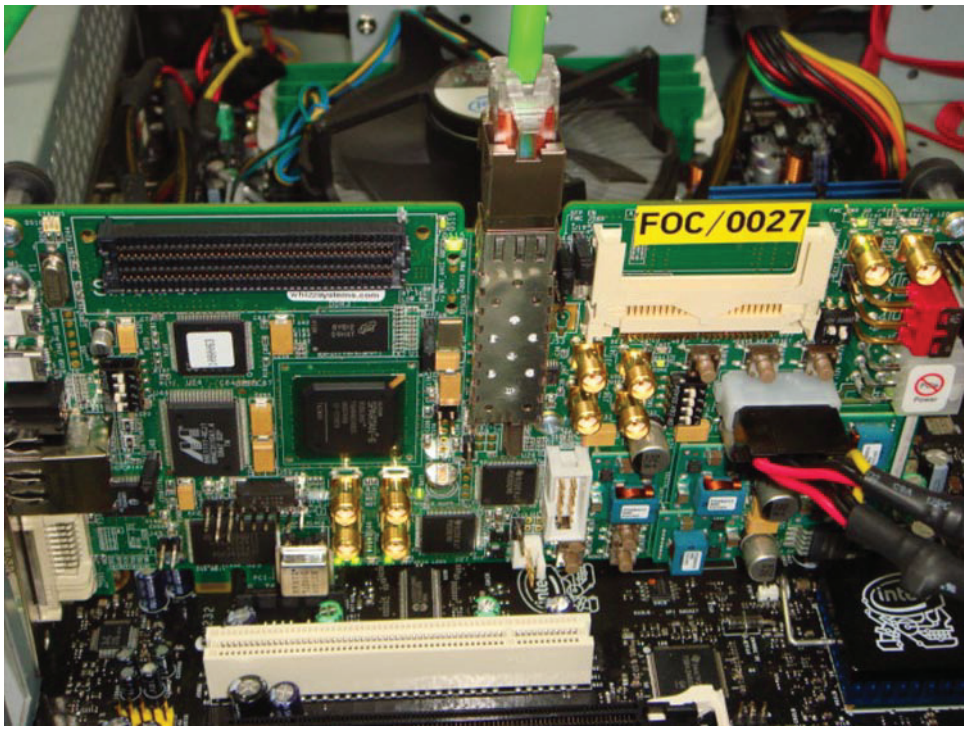
Hardware programming of the 1000BASE-X design and recompilation of drivers with this change works in the same flow as described in the Hardware Bring Up, Software Bring Up, and Network Bring Up sections.

The LEDs indicated as Ethernet status indicators in Figure 2-3 do not apply when using a 1000BASE-X design.

The setup with 1000BASE-X using an SFP to RJ45 adapter is shown in Figure 2-23.

The 1000BASE-X mode in this TRD was tested with the HP 378928-B21 Cisco Gigabit Ethernet RJ45 SFP module. This is a RJ45 SFP module that can connect to copper media.

Note: The 1000BASE-X mode of operation requires a 1 Gb/s Ethernet connection.



UG392_c2_16_120609

Figure 2-23: **Board Setup with an SFP to RJ45 Adapter**

Simulation

This section details the simulation environment provided with the design. This simulation environment is provided to show the functionality of the design. The simulation environment showcases basic traffic movement and demonstrates the functionality of the various components.

Overview

The simulation environment (Figure 2-24) consists of the design (commonly referred to as the design under test or DUT) connected to a Virtex-6 FPGA root-port model for PCI Express.

The root-port model for PCI Express is a limited test-bench environment that provides a test-program interface. The root-port model provides a source mechanism for generating downstream PCI Express traffic to stimulate the DUT and a destination mechanism for receiving upstream PCI Express traffic from the DUT in a simulation environment.

This simulation environment is built on top of the simulation environment generated by Virtex-6 FPGA integrated block for PCI Express.

The simulation environment consists of the following:

- A root-port model connected to DUT
- Transaction Layer Packet (TLP) generation tasks for various programming operations
- Test cases to generate different traffic scenarios

The simulation environment demonstrates the basic functionality of the design through various test-cases. The simulation environment shows the following:

- Configuration transactions for PCI Express
- DMA initialization
- XPS-LL-TEMAC initialization
- Virtual FIFO initialization
- Traffic movement over Ethernet
- Traffic movement over virtual FIFO
- Depending on the test case selected, it shows:
  - Interrupt handling (MSI or legacy interrupt)
  - DMA disable operation
  - Packet spanning across multiple descriptors

The simulation test bench sets up the buffer descriptors and corresponding buffers based on the compilation macros explained in detail in User Controlled Macros. The length of the packets in a network path is selected randomly and the packet size is fixed for a memory path.

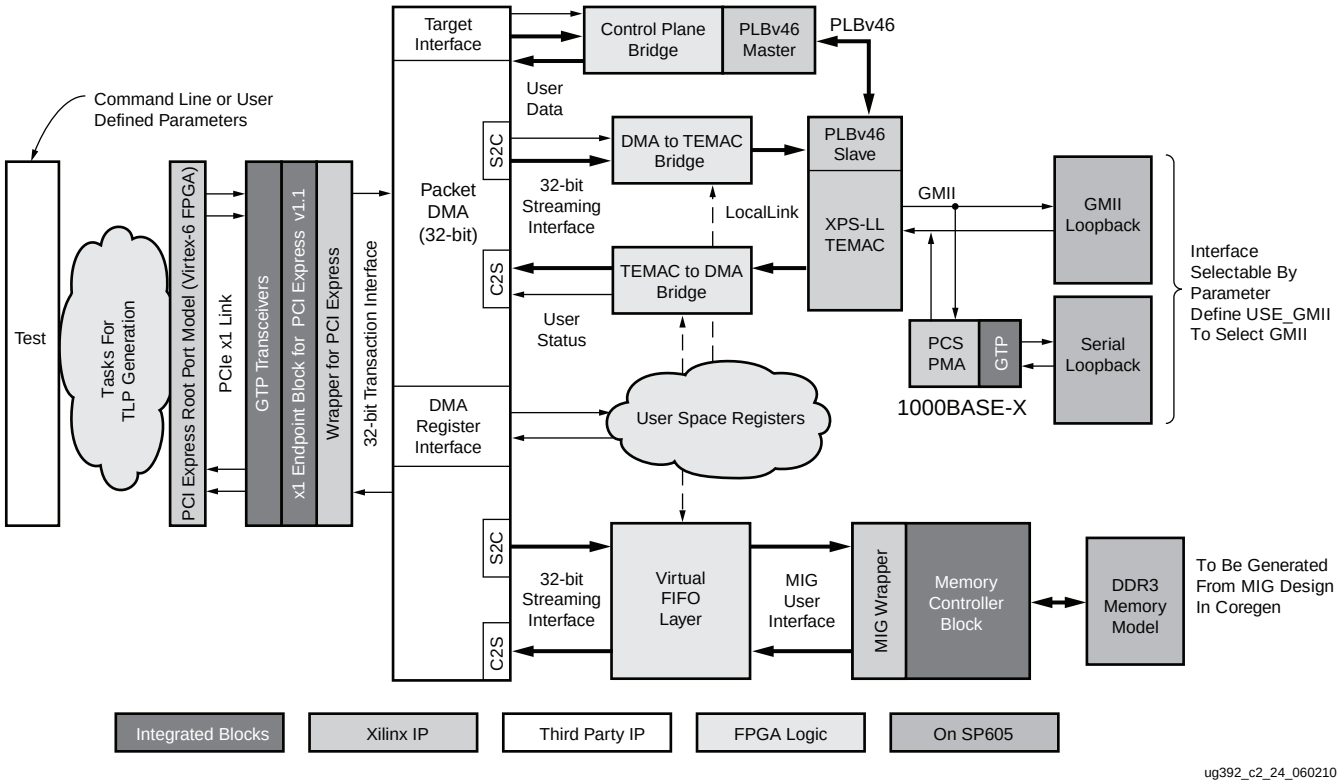


Figure 2-24: **Simulation Overview**

The simulation environment uses the Micron DDR3 memory model and connects the Ethernet interface in loopback mode (either GMII or serial-interface loopback depending on the mode selected).

The simulation environment creates log files during simulation. These log files contain a detailed record of every TLP that was received and transmitted respectively, by the root port model.

User Controlled Macros

The simulation environment allows definition of some macros controlling test bench configuration. These values can be changed in the `user_defines.v` file under the `design/sim/include` folder.

Table 2-1 describes the parameters that can be modified.

Table 2-1: **User Controlled Macro Description**

Macro Name	Default Value	Description
CH0	Defined	Enables network path initialization and traffic.
CH1	Not defined	Enables memory path initialization and traffic.
CH0_S2C_BD_COUNT	6	Number of S2C descriptors set up for network path. For basic_test, this is also the total number of packets to be transmitted. This value can be increased to a maximum of 25 for one packet per descriptor.

Table 2-1: **User Controlled Macro Description *(Cont'd)***

Macro Name	Default Value	Description
CH1_S2C_BD_COUNT	6	Number of S2C descriptors set up for memory path. For basic_test, this is also the total number of packets to be transmitted. This value can be increased to a maximum of 25 for one packet per descriptor.
USE_GMII	Not defined	Enables GMII loopback for XPS_LL_TEMAC. By default, 1000BASE-X mode is enabled.
LEGACY_INTR	Not defined	Enables legacy interrupts for PCI Express when defined, otherwise, MSI is enabled.
DETAILED_LOG	Not Defined	Enables a detailed log of each transaction.

Test Selection

The test environment generates packets of random length for network path and builds appropriate Ethernet frames. For memory path, packets of fixed length (1024 bytes) are generated.

Table 2-2 describes the various tests provided by the simulation environment.

Table 2-2: **Test Description**

Test Name	Description
basic_test	Basic Test: This test runs a defined number of packets per application. One buffer descriptor defines one full packet in this test.
packet_spanning	Packet Spanning Multiple Descriptors: This test spans a packet across two buffer descriptors.
test_interrupts	Interrupt Test: Sets the interrupt bit in descriptor and enables interrupt registers. This test also shows interrupt handling by acknowledging relevant registers.
dma_disable	DMA Disable Test: Shows a DMA disable operation sequence on a channel.

Simulating the Design with ISim

The relevant script for simulation with ISim is provided. On Linux, open a terminal and source the Xilinx environment.

- Navigate to the `design/sim/isim_lin` folder.
- Execute this script:

```
$ ./simulate_isim.sh
```

On Windows, open a command prompt and source the Xilinx environment.

- Navigate to the `design/sim/isim_nt` folder.
- Execute the `simulate_isim.bat` file.

Execution of these scripts results in compilation, elaboration, opening of the ISim GUI, running simulation, and tracing of the waveform.

Simulating the Design with ModelSim

The simulation environment provides scripts for simulation with ModelSim. Simulation of the TRD requires compilation of the Xilinx EDK Simulation Libraries for ModelSim. Xilinx provides a tool called *compxlib* for this purpose. UG628: *Command Line Tools User Guide* (Chapter 26) contains details on options specific to the ModelSim simulator version.

To run the simulation, execute the following scripts at the command prompt after setting the required environment.

ModelSim (`design/sim/mti` folder)

- On Linux

  ```
  $ ./simulate_mti.bat
  ```

- On Windows

  Execute `simulate_mti.bat`. This invokes the ModelSim GUI and runs simulation.

By default, the simulation script file specifies the Basic Test to be run using the following syntax:

```
"+TESTNAME=basic_test"
```

The test selection can be changed by specifying a different test-case argument as specified in Table 2-2 in the scripts provided.

Chapter 3

Functional Description

This chapter describes the hardware design and software driver architecture in detail.

Hardware Design Description

This section discusses the hardware design architecture in detail. Figure 3-1 shows a detailed view of the TRD. The network application is denoted by USER_APP0 and the external memory application is denoted by USER_APP1. The hardware design contains several key IP components which are stitched together with additional FPGA logic to create the entire TRD framework.

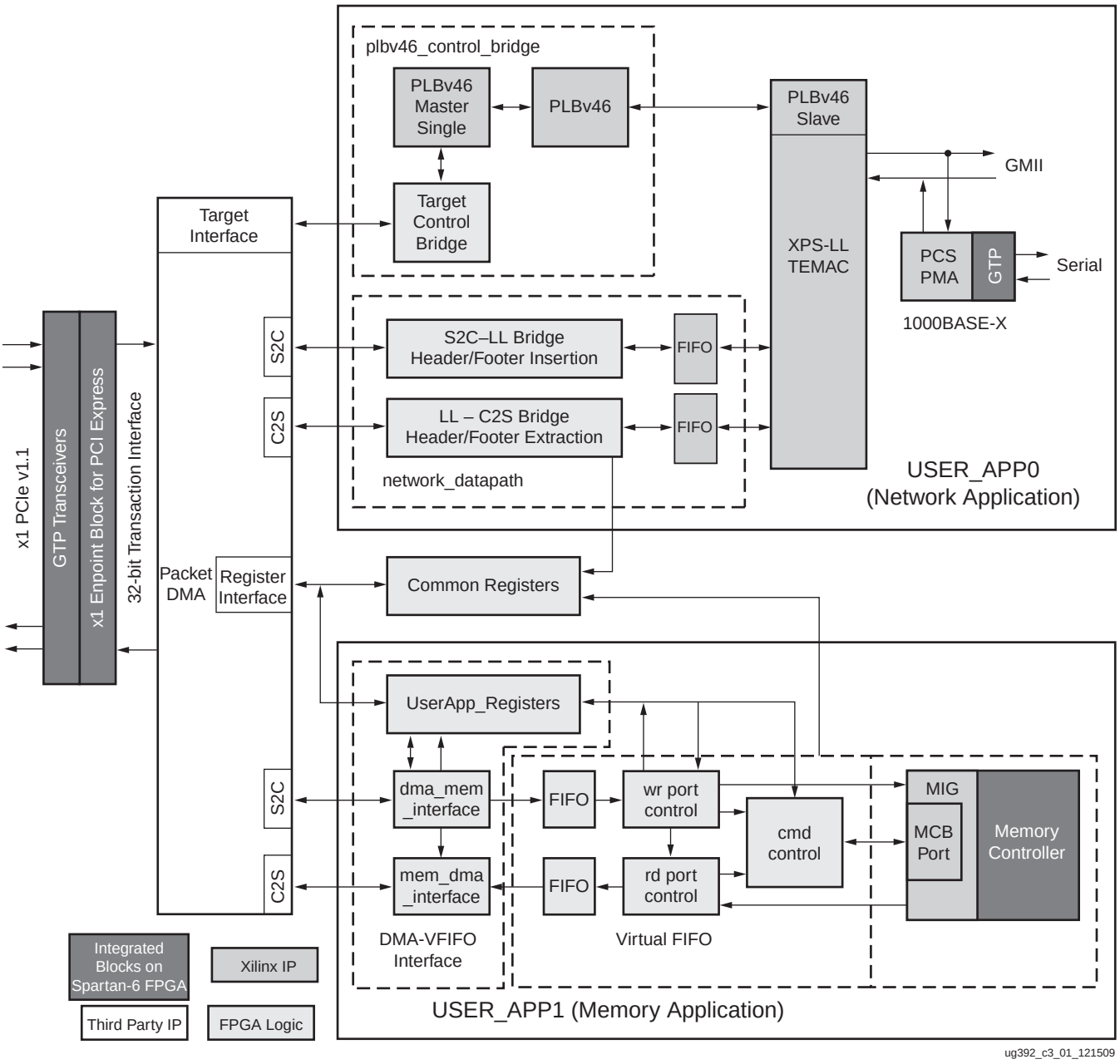


Figure 3-1: **Detailed Design Block Diagram**

The hardware architecture is described in detail under the following sections:

- Base Design Architecture: Describes the Endpoint block for PCI Express and the Packet DMA block
- Network Path Architecture: Details the network path, use of various IPs, and developed FPGA logic
- Memory Path Architecture: Details the memory path, use of various IPs, and developed FPGA logic

Base Design Architecture

The TRD is made up of two key base building blocks: the integrated Endpoint for PCI Express (PCIe) and the Packet DMA. The Endpoint provides a high-speed serial bus interface between the Spartan®-6 FPGA and a host system. Any movement of data between system memory and hardware over the PCI Express interface involves huge memory transfers, which could potentially consume a large amount of processor bandwidth if entirely managed by the processor. This transfer overhead is reduced by using a bus mastering Packet DMA controller to transfer data with minimal processor intervention. These building blocks together form the basis of the network application and external memory application detailed in this chapter.

Integrated Endpoint for PCI Express

The LogiCORE IP Spartan-6 FPGA Integrated Endpoint Block for PCI Express provides wrappers around the integrated Endpoint block for PCI Express. It is a PCI Express v1.1 compliant core supporting x1 lane width operating at 2.5 Gb/s line rate per direction. The wrapper combines the integrated Endpoint block with GTP transceivers, clocking, and reset logic to provide a simplified user interface also known as the transaction interface (TRN). The user interface data width is 32-bits operating at a 62.5 MHz clock.

For details on this core, refer to UG654, *Spartan-6 FPGA Integrated Endpoint Block for PCI Express User Guide.*

This core is generated with two 32-bit base-address registers (BAR) for this design. One BAR maps to DMA registers and another BAR maps to the XPS-LL-TEMAC and external PHY registers.

A device ID value of `6011` is used and class code is set to `07_80_00` to indicate a communication controller.

By selecting the SP605 board, a 125 MHz reference clocking scheme is automatically picked up. All other settings remain at the default values.

Packet DMA

This block is a four-channel bus-mastering packet direct memory access (DMA) controller.

This controller helps move high-speed data between the system memory and FPGA using PCI Express. It enables the simultaneous operation of two different user applications through the four channels provided. Each channel of DMA is either a system-to-card (S2C or transmit) or a card-to-system (C2S or receive). The four-channel DMA used in this design has two system-to-card (S2C) or transmit channels and two card-to-system (C2S) or receive channels.

Each DMA channel has a set of independent registers. Registers specific to this TRD are described in DMA Registers in Appendix A. Further details of various registers can be obtained from the NorthWest Logic (NWL) Packet DMA user guide.

The DMA interfaces to the Integrated Endpoint block for PCI Express through the transaction interface and provides a streaming packetized user interface. Refer to the NorthWest Logic Packet DMA User Guide for details on this user interface.

The DMA controller requires a 64 KB register space mapped to BAR0.

The DMA controller provides the following interfaces for register space mapped to BARs in the configuration space for PCI Express:

- Register interface for registers mapped to BAR0
  - All DMA registers are mapped to BAR0 from `0x0000` to `0x7FFF`. The address range from `0x8000` to `0xFFFF` is available to the user through this interface.
- Target interface for registers mapped to a BAR other than BAR0

Both these interfaces have independent read and write interfaces.

The user is required to perform read or write operations on the user implemented registers. All other aspects including building an appropriate completion TLP or extracting data from a memory write TLP are managed by the DMA controller.

Target read and target write interfaces have a command phase (transactions are granted) and a data phase (transaction data is transferred). Though the target interface can support multiple DW transactions, it is only used for one DW transactions (only register operations) in this TRD. Not all available target interface ports are used in this TRD. For details on register and target interface signals, refer to the NorthWest Logic Packet DMA user guide.

Scatter Gather Operation

A scatter gather scheme requires a common memory resident data structure that holds the list of DMA operations to be performed. DMA operations are organized as a linked list of buffer descriptors.

The term *scatter* refers to the ability to write the data into different memory locations (basically scattering data in memory). The term *gather* refers to the capability to gather data from different locations in memory and build a packet out of it.

Buffer descriptors, as the name suggests, describe the data buffer. Each buffer descriptor is eight double words (DW) in size. One DW has four bytes. Eight DWs is a total of 32 bytes. The DMA operation implements buffer descriptor chaining which allows a packet to be described by more than one buffer descriptor.

The buffer descriptor layout is slightly different for S2C and C2S directions as highlighted in Figure 3-2. The various fields are described in Table 3-1.

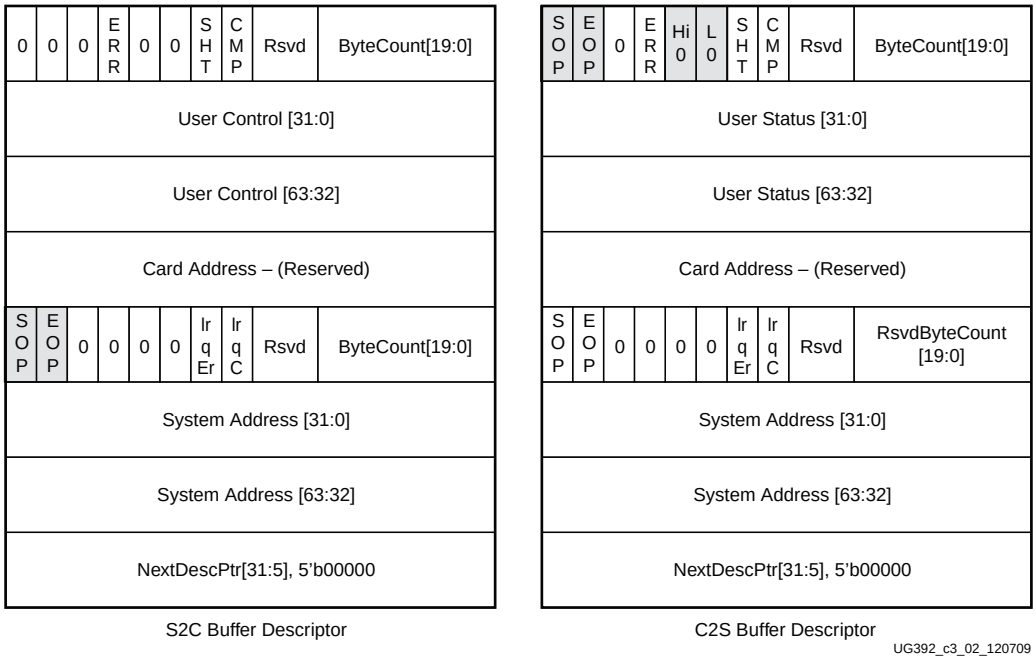


Figure 3-2: **Buffer Descriptor Layout**

Table 3-1: **Buffer Descriptor Field Description**

Descriptor Fields	Functional Description
SOP	Start of Packet: In the S2C direction, indicates the start of a new packet. In the C2S direction, DMA updates this field to indicate to software the start of a new packet.
EOP	End of Packet: In the S2C direction, indicates the end of current packet. In the C2S direction, DMA updates this field to indicate to software the end of the current packet.
ERR	Error: Set by the DMA on a descriptor update to indicate an error while executing that descriptor.
SHT	Short: Set when the descriptor is completed with a byte count less than the requested byte count. It is common for C2S descriptors having an EOP status set and should not be treated as an error in the C2S direction, but should be analyzed when set for S2C descriptors.
CMP	Complete: This field is updated by the DMA to indicate to the software the completion of an operation associated with that descriptor.
Hi 0	User status High is zero: Applicable only to C2S descriptors. This is set to indicate User Status [63:32] = 0. This bit is also used to ensure coherency to make sure that the descriptor status is valid.
L 0	User status Low is zero: Applicable only to C2S descriptors. This is set to indicate User Status [31:0] = 0. This bit is also used to ensure coherency to make sure that the descriptor status is valid.
Irq Er	Interrupt on Error: This bit indicates to the DMA to issue an interrupt when the descriptor results in error.
Irq C	Interrupt on Completion: This bit indicates to the DMA to issue an interrupt when an operation associated with the descriptor is completed.

Table 3-1: **Buffer Descriptor Field Description *(Cont'd)***

Descriptor Fields	Functional Description
ByteCount[19:0]	Byte Count: In S2C direction, this indicates to the DMA the byte count queued up for transmission. In C2S direction, the DMA updates this field to indicate the byte count updated in system memory.
RsvdByteCount[19:0]	Reserved Byte Count: In S2C direction, this is equivalent to the byte count queued up for transmission. In C2S direction, this indicates the data buffer size allocated. Depending on the packet size, the DMA might or might not utilize the entire buffer.
User Control/User Status	User Control or Status Field: In S2C direction, this is used to transport application specific data to the DMA. In C2S direction, the DMA can update application specific data in this field.
Card Address	Card Address Field: Reserved for packet DMA.
System Address	System Address: Defines the system memory address where the buffer is to be fetched from (for S2C direction) or written to (for C2S direction).
NextDescPtr	Next Descriptor Pointer: This field points to the next descriptor in the linked list. All descriptors are required to be 32-byte aligned.

Packet Transmission

Initially, the software driver prepares a ring of descriptors in system memory and writes the starting and ending addresses of the descriptors in the relevant channel registers in the DMA. Once enabled, the DMA fetches the descriptor followed by the buffer it points to. Data is fetched from the host memory and made available to the user application at the DMA S2C streaming interface. The packet interface signals (for example, start-of-packet, end-of-packet) are built from the control fields in the descriptor. The information present in the user-control field is made available during `s2c_sop`.

To indicate completion of the data fetch corresponding to this descriptor, the DMA engine updates the first DW of the descriptor by setting a completed bit by highlighting the Status and ByteCount field as shown in Figure 3-3. Software driver analyzes the completed field to free up the buffer memory and also move the descriptor back under software ownership.

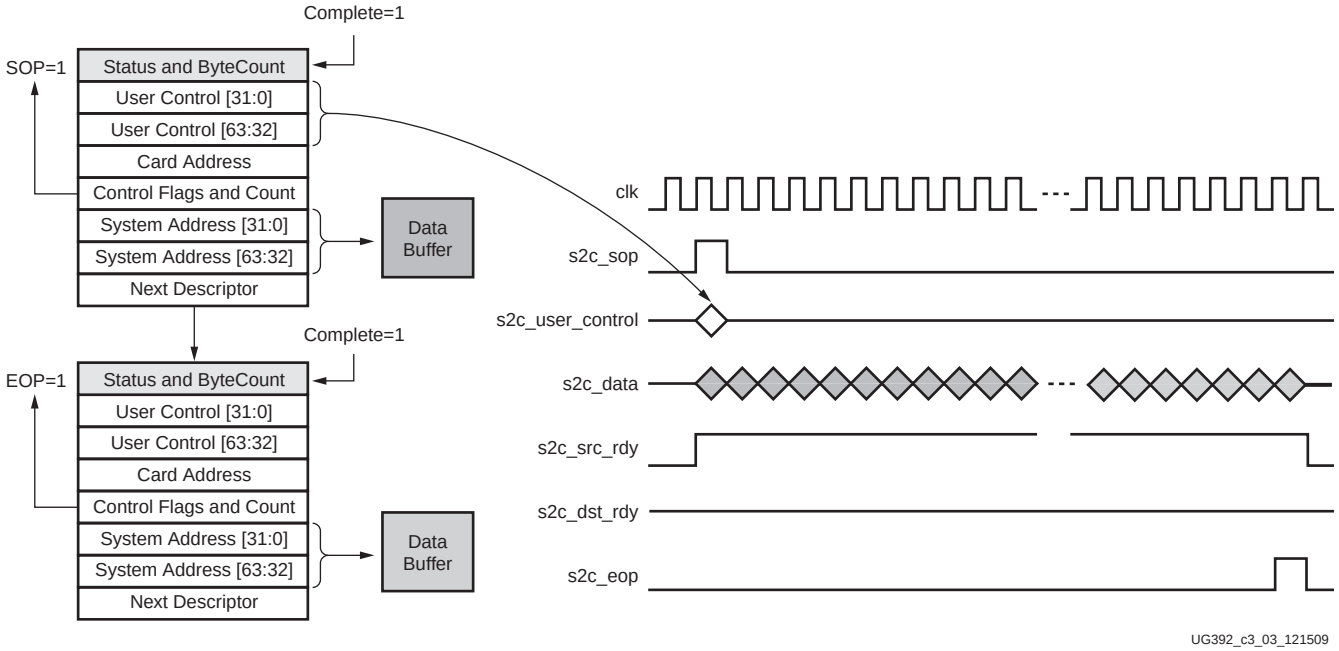


Figure 3-3: **Data Movement from System to Card**

Packet Reception

The software driver prepares a ring of descriptors with each descriptor pointing to an empty buffer and programs the starting and ending addresses in relevant DMA channel registers. The DMA reads the descriptors and waits for the user application to provide data on the C2S streaming interface (Figure 3-4). When the user application provides data, the DMA writes the data into one or more empty buffers pointed to by the pre-fetched descriptors. After the packet is written to host memory, the DMA updates the status in the descriptor. The user status field is considered valid only during `c2s_eop`. When updating EOP status, the DMA engine updates three DWs in descriptor otherwise, it only updates one DW. The completed bit in the updated status field indicates to the software driver the availability of data received from the hardware.

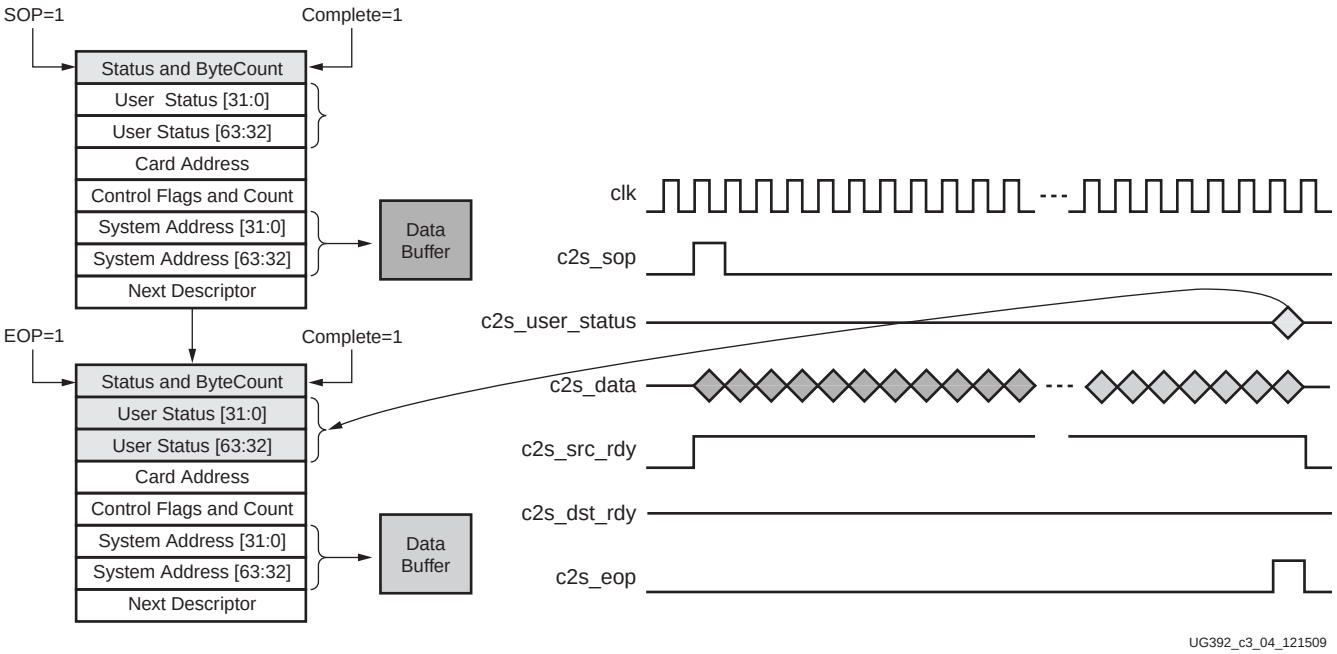


Figure 3-4: **Data Movement from Card to System**

Transmit and receive channel registers are updated in between (dependent on data availability in the transmit direction and dependent on periodic time-out or interrupts in the receive direction) to ensure uninterrupted data flow to the DMA. By analyzing the buffer descriptor, the software driver can read the status of the associated DMA transfers, fetch user information on receive channels, and determine the completion of transfer.

Network Path Architecture

A network interface card (NIC) is a device used to connect computers to a local-area network (LAN). The software driver interfaces to the networking stack (or the TCP-IP stack) and the Ethernet frames are transferred between system memory and Ethernet MAC in hardware through the base design.

The NIC application supports the following modes of operation:

- GMII mode interfacing to an external Ethernet PHY (typically used to connect to copper Ethernet networks)
- 1000BASE-X mode using the Spartan-6 FPGA GTP transceivers (typically used to connect to optical-fiber Ethernet networks)

The following sections describe the various IP cores and relevant features used.

XPS-LL-TEMAC

The XPS-LL-TEMAC IP is generated with a soft TEMAC option for the Spartan-6 FPGA. It provides options to support checksum offload and Ethernet statistics monitoring which make the XPS-LL-TEMAC core an obvious choice for this TRD framework. For the user, this simplifies and accelerates design as it eliminates the additional overhead of designing FPGA logic when connecting individual cores (such as soft TEMAC and Ethernet statistics) together in a system.

Register Access

The XPS-LL-TEMAC provides a PLBv46 slave interface to enable access to its registers. The external PHY registers are accessed through the MDIO interface in the XPS-LL-TEMAC. Certain registers are described in XPS-LL-TEMAC Registers in Appendix A. For more details, refer to DS537, *XPS LL TEMAC (v2.03a) Data Sheet* or a later version.

Use of Target Interface Instead of Register Interface

The DMA controller provides a register interface for registers mapped to BAR0. However, this interface requires that the register read data be available immediately one cycle after the read request is issued. As this one cycle access requirement is not possible with XPS-LL-TEMAC registers or with external PHY registers, these registers are mapped onto a different BAR. The DMA controller provides a target interface for accesses to user space that is not mapped to BAR0. Refer to the DMA user guide for details on a target interface.

PLBv46 master is used to connect a target interface to PLBv46 slave. Control logic, as shown in Figure 3-5, is designed to convert target interface commands to IP interconnect (IPIC) signaling. This IPIC interface is input to the PLBv46 master, which in turn drives the corresponding read or write commands onto the PLBv46 bus. Port description of IPIC and PLBv46 interface can be obtained from DS563, *PLBV46 Master Single (v1.00a) Data Sheet* or a later version.

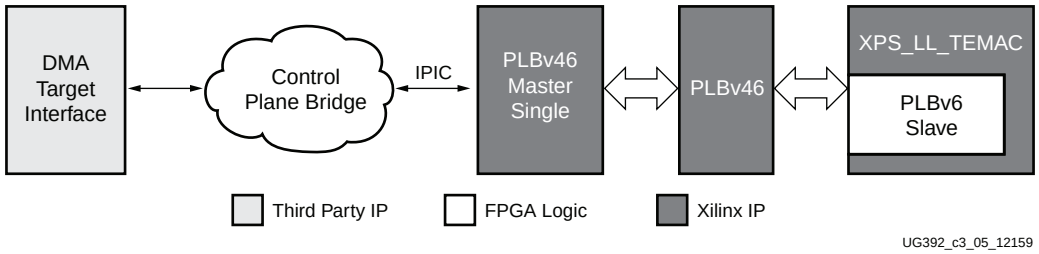


Figure 3-5: **Control Plane Logic**

Data Interface

The interface bridge logic is designed to connect the DMA streaming interface to XPS-LL-TEMAC data interface. The XPS-LL-TEMAC LocalLink interface requires header and footer in addition to payload. It does the following:

- For the S2C DMA port, the bridge builds the relevant header for the XPS-LL-TEMAC transmit LocalLink with the given user control fields and it functions as the DMA-to-LocalLink header-insert logic.
- For the C2S DMA port, the bridge extracts the relevant footer fields from the XPS-LL-TEMAC receive LocalLink and provides the DMA with a user-status field and it functions as the LocalLink-to-DMA footer-strip logic.

Checksum Offload

For TCP or UDP Ethernet protocols, data integrity is maintained by calculating and verifying checksum values over the frame data. Checksum calculation in software can be relatively slow and use significant processor utilization for large frames at high Ethernet data rates.

This design demonstrates hardware acceleration by offloading the TCP/UDP checksum calculation to the hardware. Checksum information is passed between the software and the hardware (XPS-LL-TEMAC) by user control and status fields in the buffer descriptors

which are subsequently mapped to header and footer fields of the transmit and receive LocalLink frames.

This section describes the mapping of header and footer fields in the transmit and receive directions respectively.

In the transmit direction, the 64-bit user control field provided in the buffer descriptor is used to communicate checksum related specific information to the hardware. The user control field is made available along with the start-of-packet (`s2c_sop`) signal.

Table 3-2 and Figure 3-6 shows the mapping of the user-control field to LocalLink header.

Table 3-2: **Mapping of User Control to LocalLink Header**

S2C Descriptor User Control Field	Checksum Field Mapping	Description
S2CDescUserControl[31:0]	{TX_CSCNTRL,`15'd0`,TX_CSBEGIN}	TX_CSCNTRL is the checksum control which controls insertion of checksum into the data frame. TX_CSBEGIN is the beginning offset which points to the first data byte to be included in the checksum calculation
S2CDescUserControl[63:32]	{TX_CSINSERT,TX_CSINIT}	TX_CSINSERT is the offset which points to the location where the checksum should be written into the TCP/UDP segment header. TX_CSINIT is the seed used to insert the pseudo header into the checksum calculation. If the stack inserts the pseudo header checksum into the packet, this field should be zeroed.

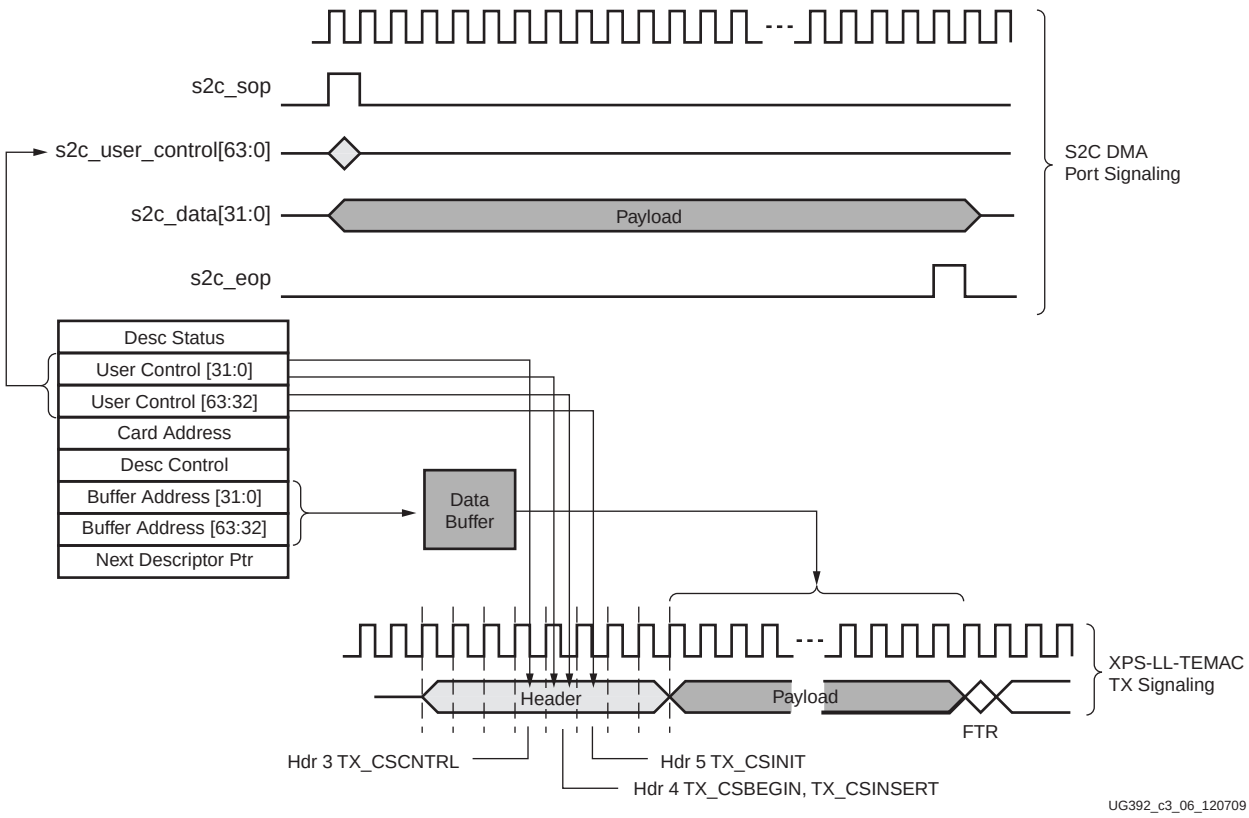


Figure 3-6: **User-Control Field Mapping to LocalLink Header**

In the receive direction, the checksum value obtained from the XPS-LL-TEMAC is transferred to the system through the 64-bit user-status field in the buffer descriptor. The user status is considered valid only during the assertion of the end-of-packet (`c2s_eop`).

Table 3-3 and Figure 3-7 shows the footer-field mapping diagrammatically.

Table 3-3: **Mapping of Footer to User Status Field**

C2S Descriptor User Status Field	Checksum Field Mapping	Description
C2SDescUserStatus[31:0]	{RX_CSRAW,`16'd0`}	RX_CSRAW is the raw checksum calculated over the entire Ethernet payload starting from the 14th byte.
C2SDescUserStatus[63:32]	{RX_BYTECNT,`16'd0`}	RX_BYTECNT is the length of the received frame in bytes. It is the number of bytes in the Ethernet frame.

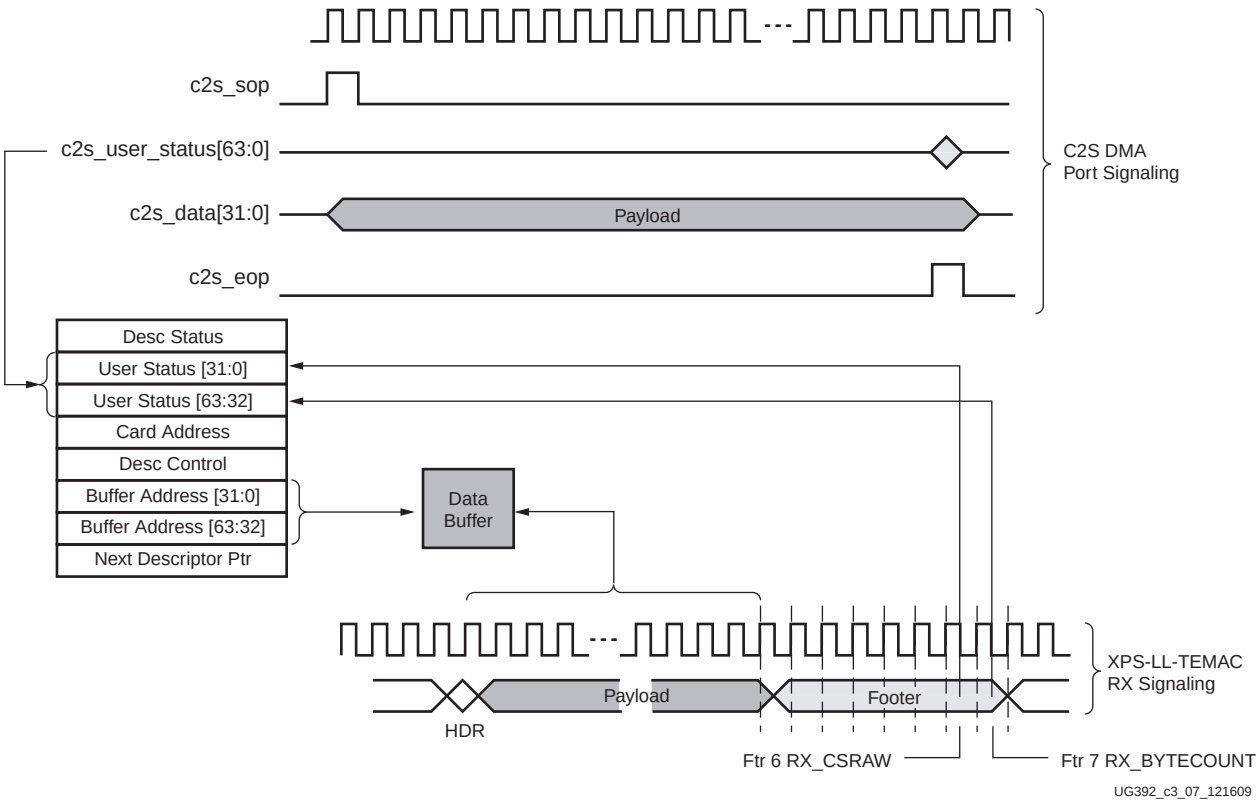


Figure 3-7: **LocalLink Footer Mapping to User Status Field**

The XPS-LL-TEMAC core also supports programmable transmit and receive block RAM size. An increased block RAM size supports jumbo frames up to 9 KB.

Use of Additional Asynchronous FIFOs in Network Datapath

The GTX_CLK, which is the GMII transmit clock, is required to be 125 MHz within a range of ±100 ppm variation as per the IEEE 802.3 specification. The PCI Express specification allows for a variation of ±300 ppm or +2800 ppm variation with spread-spectrum enabled. Therefore, a 125 MHz clock derived from a 62.5 MHz clock provided by the integrated Endpoint for PCI Express can not be used as a GTX_CLK. A 200 MHz crystal available on the SP605 board is used to source the GTX_CLK.

XPS-LL-TEMAC requires LocalLink clock to be edge aligned with the GTX_CLK. This necessitates using the 62.5 MHz clock derived from the 200 MHz on-board crystal as the LocalLink clock on the XPS-LL-TEMAC datapath. To satisfy this requirement, an asynchronous LocalLink FIFO is used in the network datapath to cross both the 62.5 MHz PCI Express and 62.5 MHz XPS-LL-TEMAC clock domains.

The asynchronous LocalLink FIFO is 512 locations deep with a 40-bit data width. 32-bits are used to store data, four bits to store the remainder (REM) field and one bit each for framing signals (SOF, SOP, EOP, and EOF).

Ethernet 1000BASE-X PCS-PMA Core

The TRD framework provides an optional 1000BASE-X mode of operation enabled by integrating Ethernet 1000BASE-X PCS/PMA core. The transceivers on Spartan-6 FPGA provide certain Physical Coding Sublayer (PCS) and Physical Medium Attachment (PMA) sub-layers for 1G Ethernet.

The Ethernet 1000BASE-X PCS/PMA core implements the PCS and PMA specific functionality as defined in IEEE 802.3 specific to 1000BASE-X operation. The core implements logic for clause 36 and 37 of the specification. For more details on the functionality implemented by this core, refer to UG155, *LogiCORE IP Ethernet 1000BASE-X PCS/PMA or SGMII v10.3 User Guide*.

In 1000BASE-X mode, instead of multiple GMII signals being exposed at the FPGA interface, only the serial transceiver interface is made visible.

The serial interface is routed to the SFP connector on the SP605 board where an optical module (using a fiber optic connection) or SFP to RJ45 converter module (using a copper connection) can be plugged in.

Memory Path Architecture

This section details the external memory connection through the integrated memory controller block on the Spartan-6 FPGA.

Additional FPGA logic implements the virtual FIFO layer around the user interface delivered by the MIG wrapper.

Memory Interface Generator

MIG provides a wrapper around the integrated memory controller and physical interface block. MIG defines all the necessary attributes required to implement a memory interface.

When selecting the memory component as MT41J64M16LA-187E DDR3 (which is the available DDR3 component on the SP605 board), the MIG wrapper automatically programs the appropriate memory attributes and timing parameters. This DDR3 component is connected to the MCB in bank 3 of the FPGA. The DDR3 SDRAM is selected as memory in bank 3 of the FPGA.

One 32-bit bidirectional port (out of five) is used at the user interface. Details of the MIG user interface signals can be obtained from UG388, *Spartan-6 FPGA Memory Controller User Guide*.

MIG Wrapper Modification for 200 MHz Clock Support

The MIG core generated from the CORE Generator tool requires a 333.33 MHz differential clock input (3000 ps time period). As the SP605 board does not have a direct 333.33 MHz differential clock source, the following changes are made to the file delivered from the MIG core to facilitate the use of a 200 MHz differential clock available on the SP605 board.

In the file `infrastructure.v`, the following changes are made:

- The input clock period to PLL is changed to five
- The PLL settings are changed to generate a 667 MHz clock and other relevant clocks for the memory controller

The user constraints file is modified suitably for locations and constraints. The MIG core generated files, which are changed for this purpose, are provided with the design.

Virtual FIFO

The DMA packetized interface is a streaming interface providing data and frame delineation signals (for example, start-of-packet). DDR3 requires an address for data access in memory.

To enable a packetized interface to direct the data into memory, a virtual FIFO wrapper is designed around the MIG core provided user interface. The virtual FIFO wrapper makes the DDR3 appear as a FIFO; the logic implemented takes care of address generation for read and write. The virtual FIFO design requires the frame size to be a multiple of four bytes, as there is no easy way to store the frame-delineator information in external memory.

The memory path logic writes the data obtained from the DMA into memory, reads it back, and presents the same data to the DMA. An overview of this block is shown in Figure 3-8.

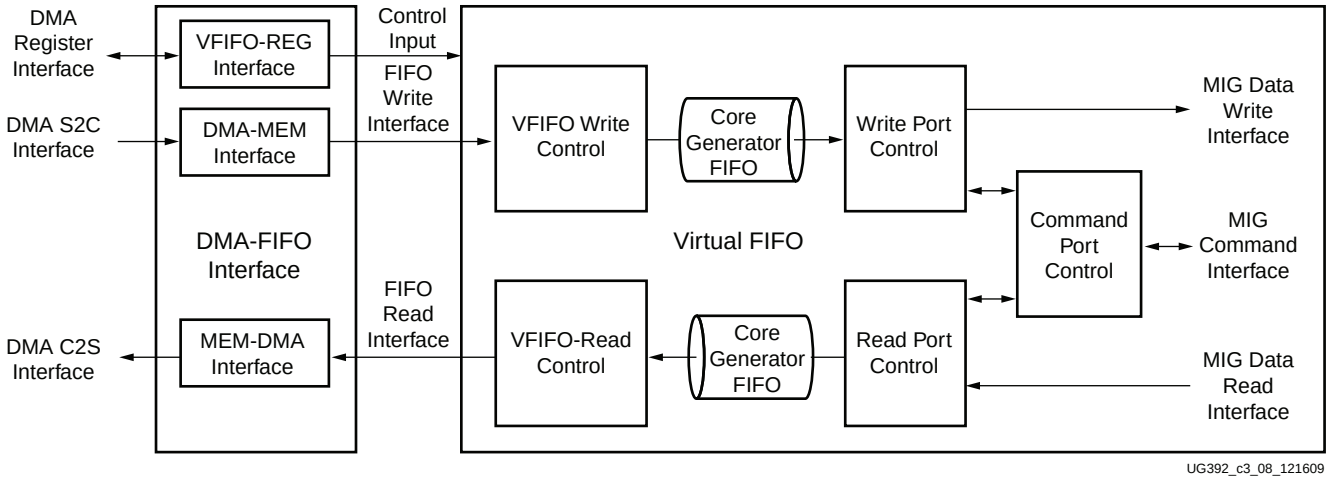


Figure 3-8: **Virtual FIFO Block Diagram**

The DMA-VFIFO interface logic connects the DMA streaming interface to the virtual FIFO interface and also implements the registers used on the memory path.

Since the packetization information is lost while storing the data into memory, a packet of a fixed size is built and presented to the DMA C2S interface. The size of the frames to be provided back to the DMA can be programmed through a register as described in Virtual FIFO Receive Packet Length Register (0x9104) in Appendix A. The MEM-DMA interface block manages packetization in the C2S direction.

The address region in the DDR3 memory being used as a virtual FIFO can be defined by programming the start and end address registers as described in Virtual FIFO Start Address Register (0x9108) in Appendix A and Virtual FIFO End Address Register (0x910C) in Appendix A.

The virtual FIFO interface is similar to general FIFO interface, and the ports available to the user are listed in Table 3-4. The virtual FIFO interfaces to the MIG core through the MIG user interface as described in UG388, *Spartan-6 FPGA Memory Controller User Guide*.

Table 3-4: **Virtual FIFO User Interface Description**

Port	Type	Functional Description
vfifo_wr_en	Input	Write enable to indicate write data availability
vfifo_wr_data[31:0]	Input	Write data to be written to memory
vfifo_wr_space[8:0]	Output	Indicates the available write buffer space
vfifo_rd_en	Input	Read enable to read data
vfifo_rd_data[31:0]	Output	Read data output from memory
vfifo_rd_space[8:0]	Output	Indicates the available read data
vfifo_empty	Output	Indicates an absence of data (FIFO is empty)
rst	Input	Reset to virtual FIFO
clk	Input	Clock to virtual FIFO
start_addr[31:0]	Input	Start address for DDR3 which is implemented in FIFO
end_addr[31:0]	Input	End address for DDR3 which wraps around the FIFO

As seen in Figure 3-8, two CORE Generator FIFOs are used at the input and output interfaces of the virtual FIFO logic to provide a continuous stream of data with adequate buffering to the user.

The depth of the CORE Generator FIFOs is 512 × 32, that is, it can store 32-bits of data with a depth of 512 locations.

A summary of the functionality of the various modules is listed in Table 3-5.

Table 3-5: **Virtual FIFO Modules**

Module	Description
VFIFO Write Control	A bridge between DMA S2C and virtual FIFO write interfaces. The payload of packets received on the DMA S2C streaming interface is written into the FIFO.
Write Port Control	Manages the write interface of the MIG core and works with the command port control block.
Command Port Control	Responsible for generating the commands to the MIG command interface. It works with the write and read modules and drives the respective commands on the MIG interface.
Read Port Control	Manages the read interface of the MIG core and works with the command port control block.
VFIFO Read Control	A bridge between the virtual FIFO read interface and the DMA C2S interface. The module manages the conversion of data read from the virtual FIFO into packets as defined by packet size and builds the corresponding LocalLink format for the entire frame.

Common Registers

A set of common registers are defined and shown in Figure 3-1. These registers include the following:

TRN Monitor Registers

The TRN monitor registers provide transaction interface utilization for PCI Express. A transaction interface utilization monitor is designed to help analyze the transaction layer utilization. The TRN monitor provides registers to count the following:

- Double words transmitted on TRN-TX, including packet headers
- Double words received on TRN-RX, including packet headers
- Transmitted memory write transactions payload count
- Received completion transactions payload count

Details of these register bits are found in TRN Monitor Registers in Appendix A.

Clocking and Reset

This section describes the clocking and reset scheme of the design.

Clocking Strategy

An overview of the clocking strategy is shown in Figure 3-9. Apart from various clock requirements of the cores, the FPGA logic in the design runs at a uniform clock rate of 62.5 MHz.

The Endpoint for PCI Express outputs a 62.5 MHz clock. DMA, FPGA logic, and data and control interfaces of the network and memory paths operate at the same speed.

Additionally, XPS-LL-TEMAC requires a 125 MHz clock for driving GMII signals and a 200 MHz reference clock for clocking the IODELAY controllers for GMII signals. These clocks are derived from the 200 MHz differential clock available on the SP605 board through a DCM. The LocalLink clock of 62.5 MHz to XPS-LL-TEMAC is also derived from this 200 MHz input clock.

The differential clock required by the MCB and the DDR3 SDRAM is also derived from the 200 MHz differential clock source available onboard.

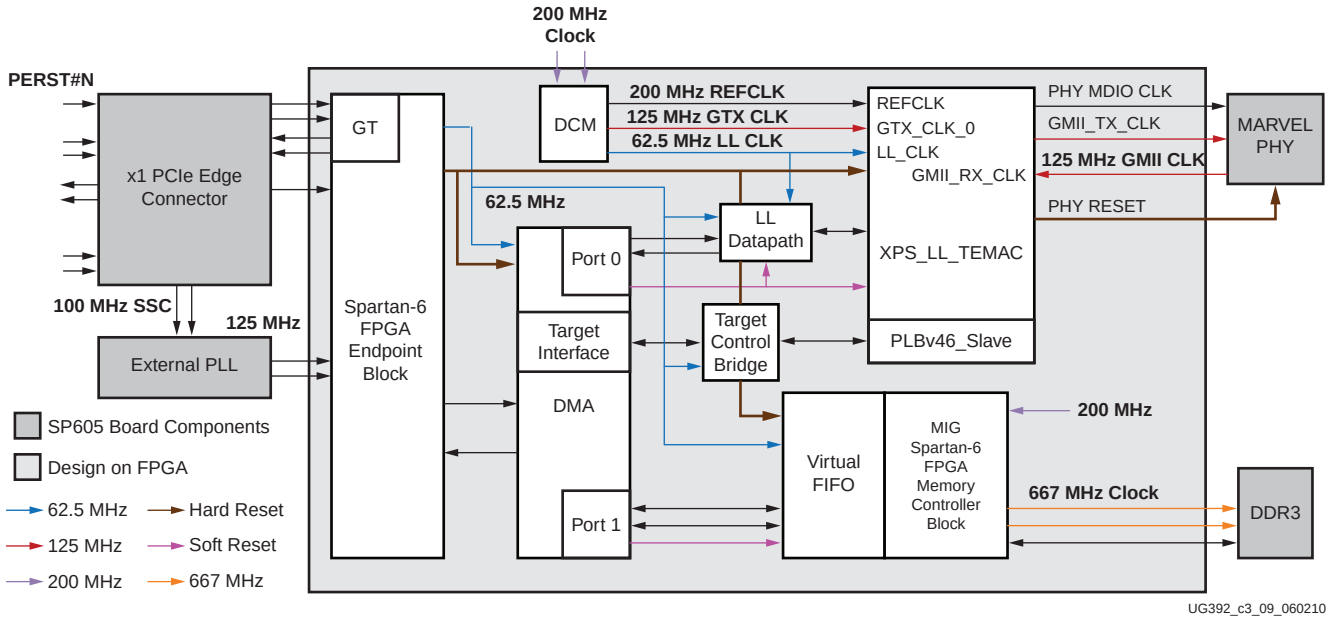


Figure 3-9: **Clocking and Reset Strategy**

Figure 3-10 shows the clocking structure when connecting to the PCS-PMA core for 1000BASE-X mode. The difference arises only with 125 MHz clock connection to XPS-LL-TEMAC. The 62.5 MHz LocalLink clock is derived from this 125 MHz clock provided. For 1000BASE-X mode, a 200 MHz clock input to XPS-LL-TEMAC is not required. The 200 MHz clock to MIG core is provided through the on-board oscillator.

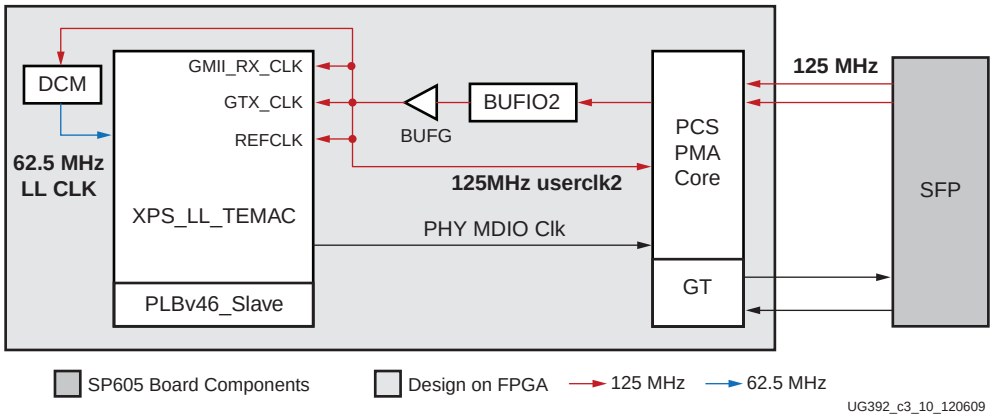


Figure 3-10: **Clocking Strategy with a PCS-PMA Core**

Reset Scheme

This section outlines the reset scheme of the design. An overview of the reset connection is shown in the Figure 3-9.

The system reset for PCI Express (PERST#N) driven by the downstream port through the edge connector is the only hard reset available to the entire design. This reset is provided as an output by the Endpoint core which is connected to the various blocks used.

In addition, a software driver programmed reset, also called a soft reset, is provided by the DMA. This reset is per DMA channel and used to drive the user logic blocks connected to the specific port being reset. The soft reset feature is used when the software driver is loading and unloading to flush out all pending transactions and prevent any stale transactions from interrupting the CPU when the hardware is not in use.

Software Design Description

The software component of the TRD framework is comprised of one or more Linux kernel-space driver modules with one user-space application, which controls the design operation.

> ***Note:*** For details on available user APIs, refer to the documentation supplied with the driver sources.

The software is designed using building blocks for scalability. Additional user-space applications can be designed with the existing blocks.

The software is designed to meet the following requirements:

- Generate adequate data to showcase the high-performance capabilities of the hardware design with specific focus on throughput on the PCI Express and Ethernet links.
  - Hardware design must demonstrate the use of PCI Express compliant DMA, 1 Gb/s Ethernet, and DDR3, while achieving best possible data throughputs in the process.
- Effectively showcase the use of the multichannel DMA transfers
- Provide an easy to use and intuitive user interface
- Provide an extensible, reusable, and customizable modular design.

The feature list of the User Application and Linux kernel-space drivers that enables these requirements is described in User-Space Application Features.

User-Space Application Features

The user-space application (which is a graphical user interface or GUI) provides the following features:

- GUI management of the driver and device for configuration control, and for status display
- GUI front-end for a graphical display of performance statistics collected at the PCI Express transaction interface and DMA engine

Standard Linux tools should be used as described in Ethernet Specific Features in Chapter 2 for control of Ethernet specific features.

Kernel-Space Driver Features

The kernel-space application provides these features:

- Configuration of the DMA engine to achieve data transfer between the hardware and main system memory
- Transfer of Ethernet packets from a Linux TCP/IP stack to the network path in hardware for transmission into the LAN, and from the network path in hardware to the Linux TCP/IP stack for handling by networking applications. This is the Ethernet data flow.

- Transfer of a block data stream from the driver to the memory path in hardware for storing in DDR3, and loopback into system. This is the memory data flow.

Data Flow Model

This section provides an overview of the datapath flows in both software and hardware.

Ethernet Data Flow

Figure 3-11 illustrates the Ethernet data flow. On the transmit path, data from the networking application (for example, an internet browser) is packetized in the TCP/IP stack, converted into Ethernet frames, and handed over to the driver for transmission. The Ethernet driver then queues up the packet for the packet DMA. The DMA fetches the packet through the Endpoint for PCI Express and transfers it to the XPS-LL-TEMAC where it is transmitted through the Ethernet link into the LAN.

On the receive side, packets arriving on the XPS-LL-TEMAC are collected by the packet DMA. The DMA pushes the packet to the driver through the Endpoint for PCI Express. The driver hands off the packet to the upper layers for further processing.

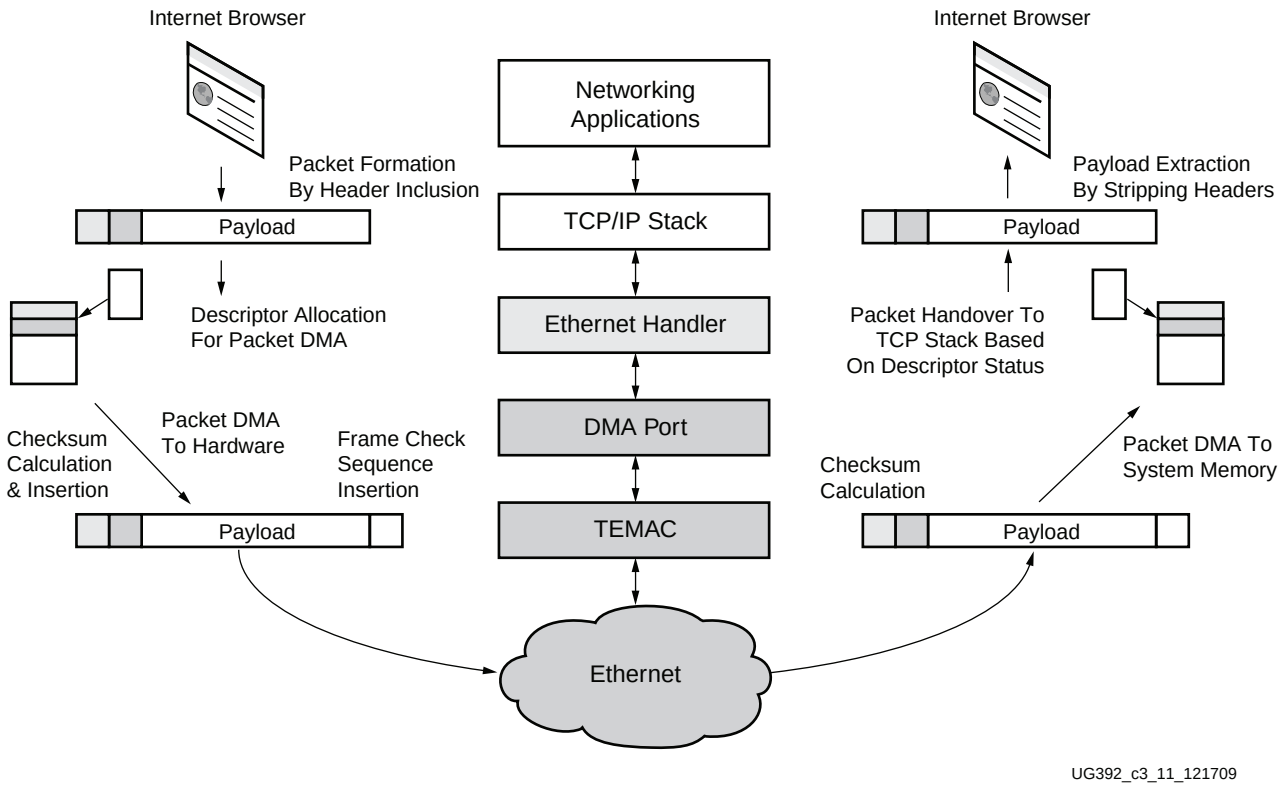


Figure 3-11: **Ethernet Data Flow**

In a typical use scenario, the user runs an application, such as a web browser, and packets are transmitted to and received from the network.

Memory Data Flow

Figure 3-12 illustrates the memory data flow. A streaming block data flow is implemented on the DDR3 side.

On the transmit side, data buffers are generated in the block data handler and queued up for transmission. The packet DMA fetches the packets through the Endpoint for PCI Express and transfers them to the Virtual FIFO which writes them into the DDR3 memory. The data written to the DDR3 is read and transferred back to the DMA creating a loopback scenario. On the receive side, the DMA pushes the packets to the software driver through the Endpoint for PCI Express. The driver receives the packets in its data buffers, verifies the data, and discards the buffers.

The driver also reads the performance statistics collected by the hardware, and makes these available to the `xpmon` GUI.

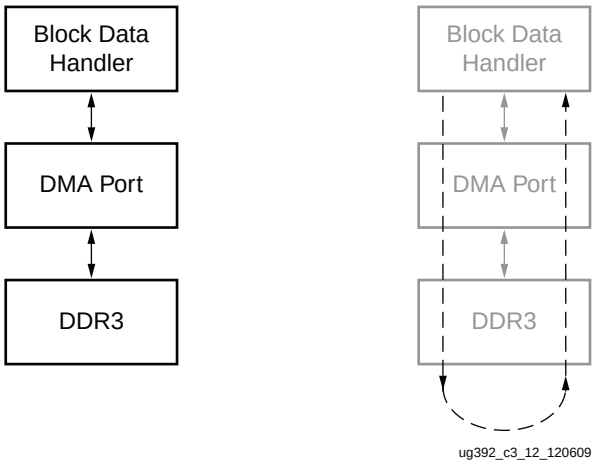


Figure 3-12: **DDR3 Memory Data Flow**

In a typical use scenario, the user starts the test through the GUI. The GUI displays the performance statistics collected during the test until the user stops the test.

Software Architecture

The software architecture has two basic data flow paths. One is the Ethernet flow and the other is the streaming block data flow. The Ethernet flow carries typical networking packets into and out of the system through a DMA port, destined for the XPS-LL-TEMAC. The block data flow carries blocks of raw data into and out of the system through another DMA port, destined for the DDR3.

Figure 3-13 shows the architecture overview. The standard software blocks already exist and the driver designed interacts with them. No modifications are done to the standard blocks; they are used as is.

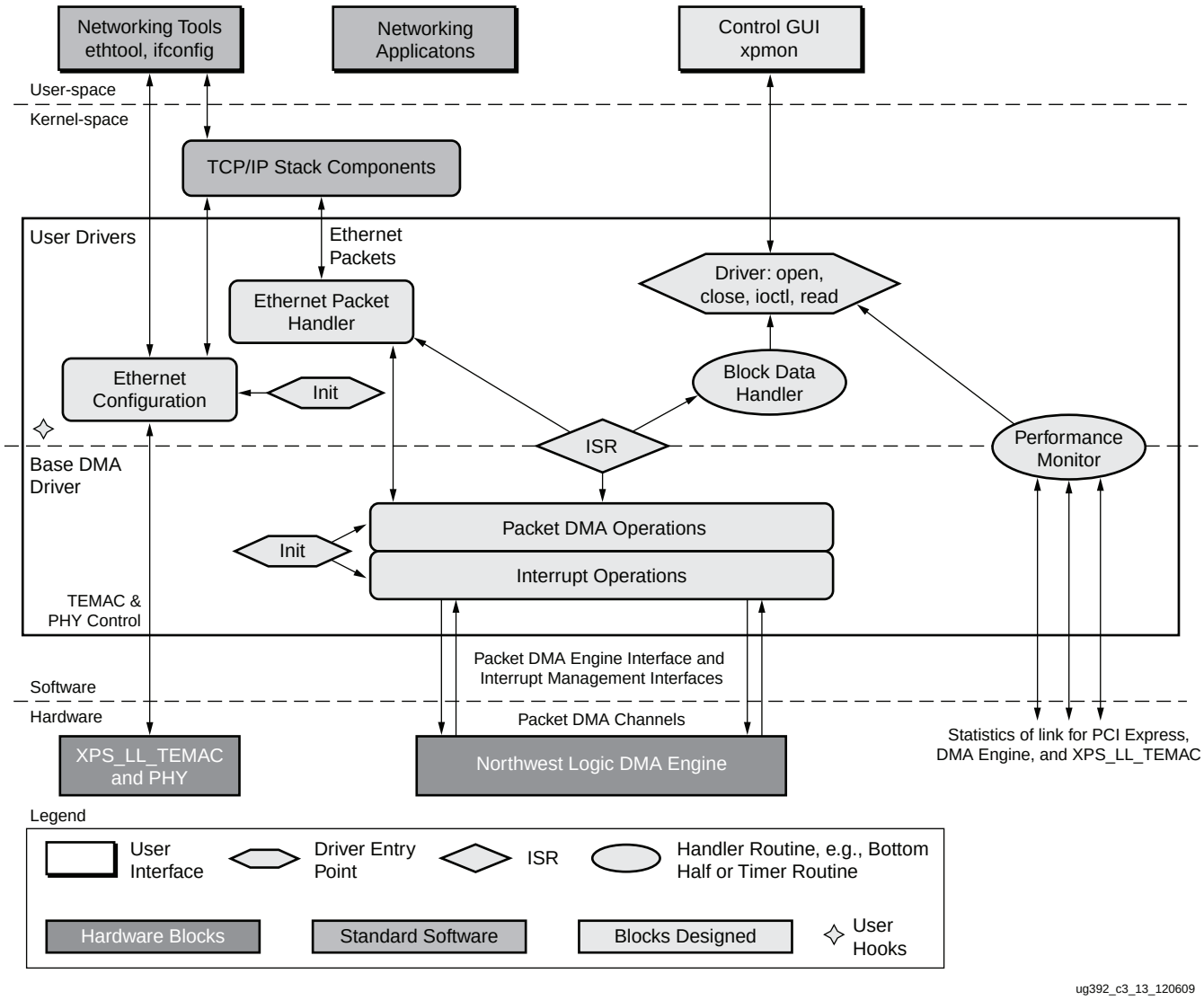


Figure 3-13: **Software Architecture Overview**

Applications

Control and Monitor GUI (xpmon)

A graphical user interface tool (xpmon) monitors device status, runs performance tests, and displays statistics. It conveys the user-configured test parameters to the block data handler in the driver, which then starts an appropriate test. Performance statistics gathered during the test are periodically conveyed to the GUI, where they are displayed in one or more graphs.

Networking Tools

Unlike the block data driver, the Ethernet functionality in the driver does not require the control and monitor GUI to be operational. Ethernet comes up with the prior configured settings. Standard Linux networking tools (for example, ifconfig and ethtool) can be used

by the system administrator when the configuration needs to be changed. The driver provides the necessary hooks which enable standard tools to communicate with it.

Networking Applications

Standard networking applications such as web browser, telnet, or Netperf can be used to initiate traffic in the Ethernet flow. The driver fits under the TCP/IP stack software, using the standard hooks provided.

Kernel Components

Driver Entry Points

The driver has several entry points, a few are described in this section. The *probe()* function is invoked by the system when a hardware match is detected after driver insertion (when the PCIe device probed by the driver is found). After reading the device's configuration space, various initialization actions are performed; initialization of the DMA engine(s), XPS-LL-TEMAC, and PHY; setting up of the receive and transmit buffer descriptor rings; and initialization of interrupts.

The other driver entry points are mainly used in the block data flow: when the GUI starts up and shuts down; when a new performance test is started or stopped; and to convey periodic performance statistics results to the GUI.

Some driver entry points are specific to Ethernet configuration, and these are invoked by the system if the user attempts to change any configuration using standard tools.

DMA Operations

For each DMA channel, the driver sets up a buffer descriptor ring. At initialization, the receive ring (associated with a C2S channel) is fully populated with buffers meant to store incoming packets, and the full receive ring is submitted for DMA. The transmit ring (associated with S2C channel) is empty. As packets arrive for transmission, they are added to the buffer descriptor ring, and submitted for DMA.

Ethernet Packet Handler

The driver is informed when a packet is queued for transmission by the upper layer. The driver adds the packet buffer to the transmit descriptor ring and submits it for DMA. Subsequently, when the driver is informed that the packet has been successfully transmitted, it removes the packet buffer from the descriptor ring, frees it, and clears the descriptor for future use.

On the receive side, the full-receive descriptor ring is submitted to the DMA engine, ready to accommodate packets on arrival. Subsequently, when the driver is informed that a packet has arrived, it removes the packet buffer from the descriptor ring, passes it to the upper layer, and clears the descriptor for future use.

Block Data Handler

The data payload for the block data flow is generated and consumed in the block data handler. When a test is started, data buffers of random or fixed sizes are generated based on user selection and then queued for the transmit DMA. The hardware design loops this data back through the DDR3 and the data buffers arrive in the system as receive DMA. The handler discards the data and returns the buffer to a free pool for future use.

Interrupt Service Routine

The interrupt service routine (ISR) manages interrupts from the DMA engine and other errors from hardware (if any).

The driver sets up the DMA engine to interrupt after every N descriptors that it processes. This value of N can be set by a compile-time macro. The ISR invokes the functionality in the block and Ethernet handler routines pertaining to handling received data, and housekeeping of completed transmit and receive descriptors.

Performance Monitor

The performance monitor is a handler which reads all the performance-related registers (link level for PCI Express, DMA engine level). Each of these is read periodically at an interval of one second.

TCP/IP Stack

The TCP/IP stack code also resides in the Linux kernel. It has well-defined hooks for the Ethernet driver to attach to, allowing configuration and communication of all standard networking tools and applications with the driver.

User Hooks

The design and code is developed to allow modification, using compile-time variables and through adaptable APIs to different applications.

DMA Descriptor Management

This section describes the descriptor management portion of the DMA operation. It also describes the data alignment requirements of the DMA engine.

The nature of traffic, especially on the Ethernet side of the design, is bursty, and packets are not of fixed sizes. For example, connect/disconnect establishment and ACK/NAK packets are small. Therefore, the software is not able to determine, in advance, the number of packets to be transferred, and accordingly set up a descriptor chain. Packets can fit in a single descriptor or can span across multiple descriptors. Also, the receive size of the actual packet can be smaller than the original buffer provided to accommodate the packet.

For proper descriptor management:

- The software and hardware must be able to independently work on a set of buffer descriptors in a supplier-consumer model
- The software must be informed of packets being received and transmitted as data flow happens. On the receive side, the software needs to know the size of the actual received packet

The rest of this section describes the driver design using the features provided by third party packet DMA IP to achieve the design objectives.

The status fields in descriptor help define the completion status, start and end-of-packet to the software driver.

Dynamic DMA Updates

This section describes how the descriptor ring is managed in the transmit or system-to-card (S2C) and receive or card-to-system (C2S) directions. It does not give details on the driver's interactions with upper software layers.

Table 3-6 presents a summary of the terminology used in this section.

Table 3-6: **Terminology Summary**

Term	Description
HW_Completed	The register containing the address of the last descriptor that the DMA has completed processing. This corresponds to the `Reg_Completed_Desc_Ptr` register in DMA.
HW_Next	The register containing the address of the next descriptor that the DMA processes. This corresponds to the `Reg_Next_Desc_Ptr` register in DMA.
SW_Next	The register containing the address of the next descriptor that software submits for DMA processing. This corresponds to the `Reg_SW_Desc_Ptr` register in the DMA.

Initially, the driver prepares descriptor rings for each DMA channel. In the reference design, the driver prepares four rings.

Transmit Initialization Phase

1. Driver initializes the *HW_Next* and *SW_Next* registers to start of ring
2. Driver resets the *HW_Completed* register
3. Driver initializes and enables the DMA engine

Transmit (S2C) Descriptor Management

In Figure 3-14, the darkened blocks indicate the descriptors under hardware control and the white blocks indicate the descriptors under software control.

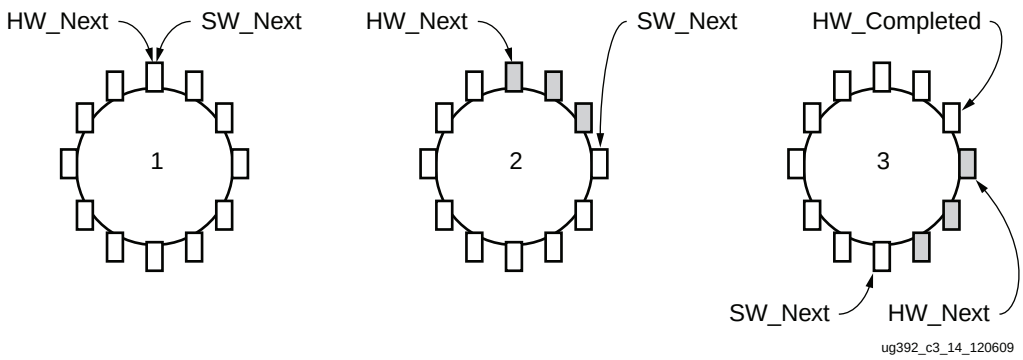


Figure 3-14: **Transmit Descriptor Ring Management**

Packet Transmission

1. The packet arrives in the Ethernet packet handler
2. The packet is attached to one or more descriptors in the ring
3. The driver marks the SOP, EOP, and IRQ_on_completion in the descriptors
4. The driver adds any user-control information (for example, checksum-related) to the descriptors
5. The driver updates the SW_Next register

Post-Processing

1. The driver checks for completion status in the descriptor
2. The driver frees the packet buffer

This process continues as the driver keeps adding packets for transmission and the DMA engine keeps consuming packets. Since the descriptors are already arranged in a ring, the post-processing of descriptors is minimal and dynamic allocation of the descriptors is not required.

Receive (C2S) Descriptor Management

In Figure 3-15, the darkened blocks indicate the descriptors under hardware control and the white blocks indicate the descriptors under software control.

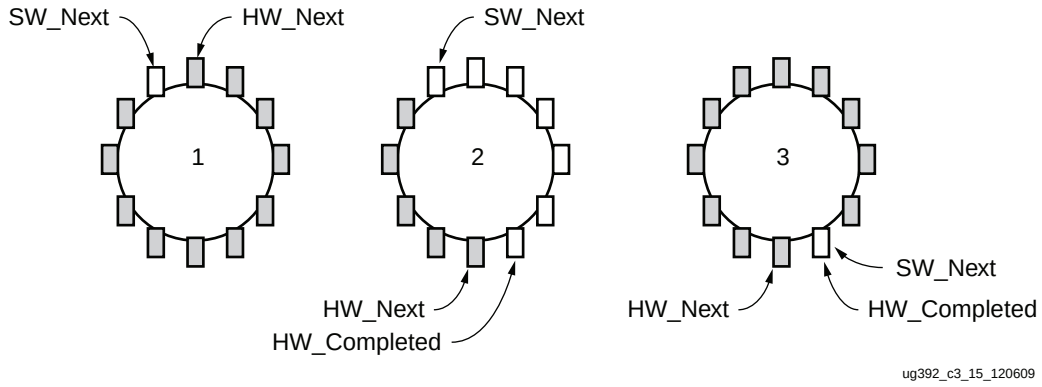


Figure 3-15: **Receive Descriptor Ring Management**

Receive Initialization Phase

1. The driver initializes the receive descriptor with an appropriate Ethernet or block data buffer.
2. The driver initializes the HW_Next register to the start of the ring and the SW_Next register to the end of the ring
3. The driver resets the HW_Completed register
4. The driver initializes and enables the DMA engine

Post-Processing after Packet Reception

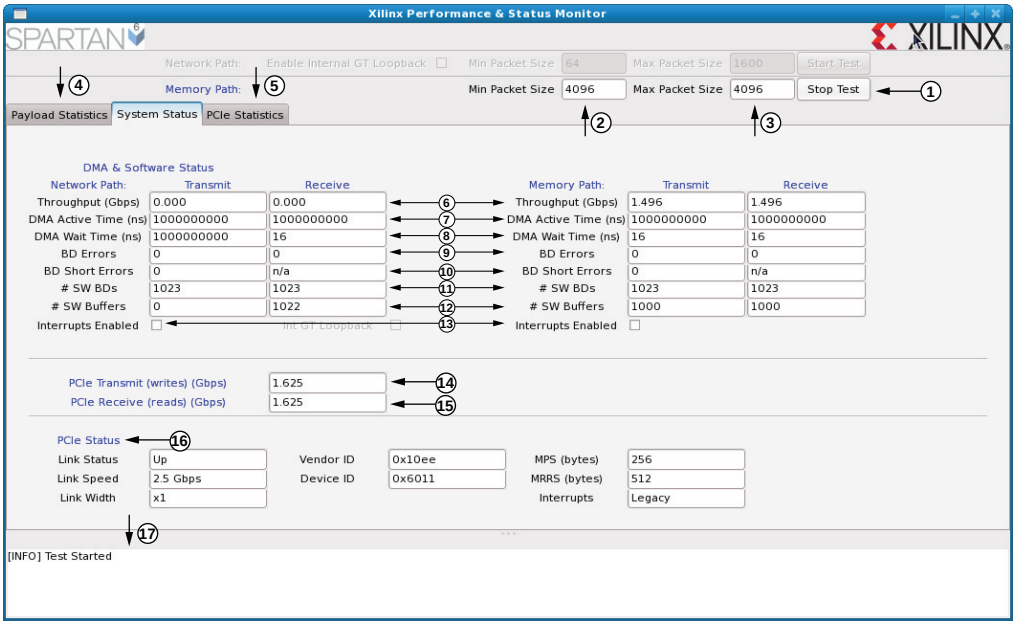
1. The driver checks for completion status in the descriptor
2. The driver checks for SOP, EOP, and user status information
3. The driver forwards the completed packet buffer(s) to the upper layer
4. The driver allocates the new packet buffer for the descriptor
5. The driver updates the SW_Next register

This process continues as the DMA engine keeps adding received packets in the ring, and the driver keeps consuming packets. Since the descriptors are already arranged in a ring, post-processing of descriptors is minimal and dynamic allocation of descriptors is not required.

User Interface—Control and Monitor GUI

The control and monitor GUI is the main application tool for the configuration, status, and statistics display. When installed, the driver appears as a device table entry in Linux. The GUI uses the typical file-handling functions (open, close, read, write, ioctl) on this device to communicate with the driver. These calls result in the invocation of the appropriate driver entry points. The I/O control function *ioctl()* is used as a driver entry point by the application GUI.

Figure 3-16 shows a screen capture of the GUI status. An explanation of the various fields referenced by numbers follows.



UG392_c3_16_121709

Figure 3-16: **Software Application Screen Capture**

1. Test start/stop control for memory application
2. Min. Packet Size: Minimum packet size selection in bytes
3. Max. Packet Size: Maximum packet size selection in bytes
4. Payload Statistics: Shows the payload statistics graphs based on DMA engine performance monitor
5. PCIe statistics: Plots the PCIe transaction interface utilization
6. Throughput: DMA payload throughput in Gb/s for each engine.
7. DMA Active Time: The time (in nanosecond) the DMA engine has been active in the last one second.
8. DMA Wait Time: The time (in nanosecond) the DMA was waiting for the software to provide more descriptors.
9. BD Errors: Indicates a count of descriptors which caused a DMA error – indicated by the error status field in descriptor update
10. BD Short Errors: Indicates short error in descriptors in the transmit direction when the entire buffer specified by length in the descriptor could not be fetched. This field is not applicable for receive direction.

11. # SW BDs: Indicates count of total descriptors set up in the descriptor ring
12. # SW Buffers: Indicates count of total data buffers associated with the ring
13. Interrupts Enabled: Indicates interrupt enable status for that DMA engine
14. PCIe Transmit: Reports the TRN transmit utilization as obtained from the transaction monitor in hardware
15. PCIe Receive: Reports the TRN receive utilization as obtained from the transaction monitor in hardware
16. PCIe Status: Reports the status of various PCIe fields as reported in the endpoint's configuration space
17. Text pane at the bottom shows up informational messages, warnings or errors.

The GUI has individual tabs for the following:

Status

- Link status for PCI Express
- DMA Engine status

The driver always maintains information on the status of the hardware. The GUI invokes *ioctl()* to read this status information and updates it every few seconds.

Statistics

- Link statistics for PCI Express provided by hardware
- Graphic display of all statistics

The driver maintains a set of arrays to hold per-second sampling points of these statistics. These statistics are periodically collected by the performance monitor handler. The arrays are managed in a circular fashion. The GUI periodically invokes an *ioctl()* to read these statistics and then displays them.

A separate test button is provided to run user tests. The test button is applicable only to the memory path in the design as the traffic for Ethernet is generated by upper layers (that is, TCP/IP) and by standard applications.

Test

- Test setup
- Start/Stop of test

When the user starts a test, the GUI informs the driver the parameters of the test:

- Minimum and maximum packet size. If these are different, the driver generates packets of random sizes within these bounds.

The driver entry point sets up the test parameters and informs the block data handler, which then starts setting up the block data buffers for transmission, reception, or both. Similarly, if the user were to abort a test, the GUI informs the driver, which sets up the abort mechanism. The test is aborted by stopping the transmit side flow and then allowing the receive side flow to drain.

The GUI programming environment is GTK+.

System Logging

The drivers also generate system messages which can be viewed via `/var/log/messages` files or the `dmesg` tool or the System Logs tool. The level of logging can be controlled by setting macros in the Makefiles, as described in Log Verbosity Level in Chapter 5. Increasing the log verbosity level affects the driver and, therefore, the system throughput.

XILINX®

Chapter 4

Performance Estimation

This chapter presents a theoretical estimation of performance on the interface for PCI Express, Ethernet interface, and the memory interface. It also presents a method to measure performance.

PCI Express Performance

PCI Express is a serialized, high bandwidth and scalable point-to-point protocol that provides highly reliable data transfer operations. The maximum transfer rate of a protocol specification v1.1 compliant core is 2.5 Gb/s. This rate is the raw bit rate per-lane per-direction and not the actual data-transfer rate. The effective data-transfer rate is lower due to protocol overheads and other system design trade-offs.

The link performance of PCI Express together with packet DMA is estimated under the following assumptions:

- Each buffer descriptor points to a 1 KB data buffer space
- Maximum Payload Size (MPS) = 128B
- Maximum Read Request Size (MRRS) = 128B
- Read Completion Boundary (RCB) = 64B
- TLPs of 3DW considered without extended CRC (ECRC): Total overhead of 20B
- One ACK assumed per TLP: DLLP overhead of 8B
- Update FC DLLPs are not accounted for but they do affect the final throughput slightly

Performance is projected by estimating the overheads and then calculating the effective throughput by deducting these overheads. Descriptor fetch/update as required for data movement through DMA is also considered as an overhead.

Independent calculations are made for each direction of a C2S or a S2C DMA engine.

The notes in Table 4-1 describe the abbreviated conventions.

Table 4-1: **PCI Express Performance Estimation with DMA**

Transaction Overhead	ACK Overhead	Comment
MRD: C2S Descriptor = 20/1024 = 2.5/128	8/1024 = 1/128	One descriptor fetch in a C2S engine for 1 KB data (TRN-TX); 20B of TLP overhead and 8 bytes of DLLP overhead
CPLD: C2S Descriptor = (20+32)/1024 = 6.5/128	8/1024 = 1/128	Descriptor reception C2S engine (TRN-RX)
MWR: C2S Descriptor = (20+12)/1024 = 4/128	8/1024 = 1/128	Descriptor update C2S engine (TRN-TX)
MWR: C2S Buffer = 20/128	8/128	MPS = 128B; Buffer write C2S engine (TRN-TX)
MRD: S2C Descriptor = 20/1024 = 2.5/128	8/1024 = 1/128	Descriptor fetch in S2C engine (TRN-TX)
CPLD: S2C Descriptor = (20+32)/1024 = 6.5/128	8/1024 = 1/128	Descriptor reception S2C engine (TRN-RX)
MWR: S2C Descriptor = (20+4)/1024 = 3/128	8/1024 = 1/128	Descriptor update S2C engine (TRN-TX)
MRD: S2C Buffer = 20/128	8/128	MRRS = 128B; Buffer fetch S2C engine (TRN-TX)
CPLD: S2C Buffer = 20/64 = 40/128	8/64 = 16/128	RCB = 64B; Buffer reception S2C engine (TRN-RX)

Notes:

1. Nomenclature for table: Memory Read transaction (MRD); Memory Write transaction (MWR); Completion with data (CPLD); Card-to-system (C2S) for receive direction DMA; System-to-card (S2C) for transmit direction DMA.

Once all overheads for each DMA engine are estimated, effective throughput is calculated. The throughput calculation for application data (packets available at the DMA streaming interface) for a x1 link is tabulated in Table 4-2.

Table 4-2: **PCI Express Throughput Estimate**

Direction	Overhead	Effective Throughput (Gb/s)
PCIe transmit (C2S only)	100 × 27.5/(128 + 27.5) = 17.68%	1.64
PCIe receive (C2S only)	100 × 16.5/(128 + 16.5) = 11.41%	1.77
PCIe transmit (S2C only)	100 × 42.5/(128 + 42.5) = 24.9%	1.5
PCIe receive (S2C only)	100 × 56.5/(128 + 56.5) = 30.6%	1.38

The S2C engine (data transmission, that is, reading data from system memory) issues read requests and receives data through completions. This engine exercises data (actual frame) traffic on the PCIe receive link giving a performance of ~1.38 Gb/s. This is the PCIe memory read performance.

The C2S engine (data reception, that is, writing data to system memory) issues write requests. This engine exercises data (actual frame) traffic on the PCIe transmit link giving a performance ~1.64 Gb/s. This is PCIe memory write performance.

PCIe receive on C2S implies a descriptor fetch and ACK-NAK, resulting in a low overhead. PCIe transmit on S2C is made up of descriptor and buffer fetch requests. Read requests do not contribute towards data throughput as they are only headers.

Ethernet Performance

The raw line rate of the Ethernet link is 1.25 Gb/s, which is commonly referred to as gigabit Ethernet after accounting for 8B/10B encoding overheads.

The performance, as seen by various Ethernet applications at different layers, is lesser than the throughput seen at the driver and the Ethernet interface. This is due to the various headers and trailers inserted in each packet by all the layers of the networking stack. Ethernet is used as a medium to carry traffic and various protocols, including TCP/UDP, to implement protocol specific header/trailer formats.

Consider transmission control protocol (TCP) as an example. The protocol header includes the following:

- TCP/IP Overhead: 20 bytes TCP header + 20 bytes IP header
- Ethernet Overhead: 14 bytes Ethernet header + 4 bytes trailer

Based on this overhead, the theoretical TCP throughput is as shown in Equation 4-1, where D is the application message size in bytes.

$$TCP\ throughput = \left(\frac{D}{D + 40 + 18}\right) \times 1000\ Mb/s \qquad \text{Equation 4-1}$$

More precisely, for application message sizes greater than 1460 bytes, the formula is shown in Equation 4-2, where M is the MTU size configured on the system.

$$TCP\ throughput = \left(\frac{D}{D + 40}\right) \times \left(\frac{M}{(M + 14)}\right) \times 1000\ Mb/s \qquad \text{Equation 4-2}$$

The calculation of theoretical throughput is tabulated in Table 4-3. The TCP/IP protocol overhead has a significant impact when the send message sizes are smaller than ~1200 bytes.

Table 4-3: **Theoretical Throughput Estimate for TCP**

Message Size in Bytes	Effectiveness (%)	Theoretical Throughput (Mb/s)
64	52.46	524.59
128	68.82	688.17
256	81.53	815.29
512	89.82	898.25
1024	94.64	946.40
1442	96.15	961.47
2048	96.92	969.21
4096	97.86	978.59

Memory Controller Performance

The Spartan-6 FPGA memory controller block, as used in this design, has a total of 16 I/Os interfacing to external DDR3 memory.

Theoretical Calculation

$$Maximum\ I/O\ Rate\ (double\ data\ rate) = 333.5\ MHz \times 2 = 667\ Mb/s \quad \text{Equation 4-3}$$

$$Maximum\ Bandwidth = (Maximum\ I/O\ rate) \times (Number\ of\ I/Os) \quad \text{Equation 4-4}$$

$$= 667\ Mb/s \times 16 = 10.627\ Gb/s \quad \text{Equation 4-5}$$

Equation 4-4 calculates the theoretical maximum bandwidth of the memory controller. An estimate of memory controller performance is as calculated in Equation 4-9:

With larger burst lengths, high efficiency is achievable. With a 32-bit port using a burst length of 32, a total of 1024 bits are transferred.

Number of bits transferred per cycle is:

$$16(bit\ width) \times 2(double\ data\ rate) = 32\ bits/\text{cycle} \quad \text{Equation 4-6}$$

Total cycles used for 1024 bits:

$$1024 / 32 = 32\ cycles/transfer \quad \text{Equation 4-7}$$

Assuming 10 cycles read to write overhead:

$$32 / 42 = 76\%\ efficiency \quad \text{Equation 4-8}$$

Assuming 5% efficiency overhead for refresh:

$$71\%\ efficiency\ at\ 667\ Mb/s\ \text{for 16-bit DDR3} = 7577\ Mb/s = 7.577\ Gb/s \quad \text{Equation 4-9}$$

The final estimated bandwidth available to the Virtual FIFO is 7.577 Gb/s.

In the current design, with x1 PCIe at 2.5 Gb/s line rate, the theoretical maximum rate at which the virtual FIFO can read and write to memory controller is calculated as 32 bits × 62.5 MHz = 2 Gb/s. The average data throughput provided by the PCIe and DMA is ~1.6 Gb/s in each direction.

Measuring Performance

This section describes methods to measure performance and presents an analysis of what to expect when different parameters are varied.

PCI Express performance is dependent on factors like maximum payload size, maximum read request size, and read completion boundary which depend on the systems chosen. With higher MPS values, performance improves as packet size increases.

Table 4-4 lists the registers provided by the hardware to aid in the software performance measurement.

Table 4-4: **Performance Registers in Hardware**

Register	Description
DMA Completed Byte Count	DMA implements a completed byte-count register per engine, which counts the payload bytes delivered to the user on the streaming interface.
TRN-TX Utilization	This register counts traffic on TRN-TX interface including TLP headers for all transactions.

Table 4-4: **Performance Registers in Hardware *(Cont'd)***

Register	Description
TRN-RX Utilization	This register counts traffic on TRN-RX interface including TLP headers for all transactions.
TRN-TX Payload	This register counts payload for memory write transactions upstream, which includes buffer write and descriptor updates.
TRN-RX payload	This register counts payload for completion transactions downstream, which includes descriptor or data buffer fetch completions.

These registers are updated once every second by hardware. Software reads them periodically at a one second interval. The value read directly gives the throughput in bytes per second.

The TRN monitor registers can be read to understand PCIe transaction layer utilization. The DMA registers provide throughput measurement for the actual payload transferred.

These registers only estimate hardware performance. A software application is required to measure application performance of both the hardware and software impact on overall throughput.

Ethernet Performance Measurement

Ethernet performance can be measured in a private LAN environment using a standard benchmarking tool like Netperf.

Netperf 2.4 can be used as the testbench for measuring outbound and inbound throughput. This is a data transfer application running on top of the TCP/IP stack. The client can be configured for different message sizes and to open a TCP connection or a UDP connection. The two systems are connected in a private LAN connection, which avoids other external LAN traffic. To measure the CPU utilization as accurately as possible, no other applications (other than the standard ones) should be run during the test.

Throughput Estimate and Analysis

It is possible, based on the theoretical estimates to obtain a throughput of up to 935 Mb/s in both inbound and outbound directions.

For jumbo frames, a higher performance is expected. Frames greater than 1514 bytes are categorized as jumbo frames. The XPS-LL-TEMAC IP used in this TRD supports jumbo frames up to 9K bytes.

Higher performance is expected with an increase in MTU size. A higher value of MTU implies that a packet has more payload and less header content.

CPU Utilization Analysis

Improved CPU utilization for larger packets is expected with checksum offload enabled. By offloading an expensive operation to hardware, such as a checksum calculation, the CPU time can be efficiently utilized elsewhere. With small packets, little improvement can be seen in CPU utilization with checksum offload as checksum computation does not contribute greatly to CPU utilization overhead, as does the protocol overhead of using small packets.

Refer to Appendix C, Setting Up a Private LAN for steps on setting up a private LAN connection.

XILINX®

Chapter 5

Designing with the TRD Platform

The Targeted Reference Design (TRD) platform is intended to be a framework for system designers to derive extensions or modify designs.

This chapter outlines various ways for a designers to evaluate, modify and re-run the TRD for the connectivity platform.

The suggested modifications are grouped under these categories:

- Software-only modifications: Only changes required to the software driver are covered. The same bitstream provided with the TRD works. Only driver recompilation is required.
- Design (top-level only) modifications: Changes to parameters in the top-level file of the design (`design/source/s6_pcie_dma_ddr3_gbe.v`). No other files require modifications. The design must be re-implemented. See Implementing the Design, page 35 for instructions on design implementation. Depending on the modification, the software driver might require changes.
- Design Changes: This modification shows the plug-n-play feature of the TRD platform. The design must be re-implemented after making the changes. The process requires licenses for additional IP cores being used (if any). The software driver must be modified accordingly.

Any change to the software driver or macros in Makefiles require re-compilation and re-building of the kernel objects. All paths for various files mentioned in this chapter are under the `s6_pcie_dma_ddr3_gbe` directory.

While describing the modifications, each section also describes the implication of the corresponding modification on the overall functionality or performance.

Software-Only Modifications

This section describes modifications to the platform done directly in the software driver. The same hardware design (bitstream) works.

Macro-Based Modifications

This section describes the modifications that result when compiling the software driver with various macro options, either in the Makefile or in the driver source code.

Descriptor Ring Size

The number of descriptors to setup in the descriptor ring are defined as a compile-time option.

Modify the DMA_BD_CNT in `driver/xdma/xdma_base.c` macro to change the size of the buffer descriptor ring used for DMA operations. Smaller rings can adversely affect throughput, which is observed by running the performance tests.

A larger descriptor ring size uses additional memory but improves performance because more descriptors can be queued to hardware.

Log Verbosity Level

Log verbosity level can be controlled by:

- Adding DEBUG_VERBOSE in the Makefiles, which causes the drivers to generate verbose logs.
- Adding DEBUG_NORMAL in the Makefiles, which causes the drivers to generate informational logs.
- Removing both these macros from the Makefiles, which causes the drivers to only generate error logs.

Changes in the log verbosity are observed when examining the system logs. Increasing the logging level also causes a drop in throughput.

Driver Mode of Operation

The base DMA driver is configured to run in either interrupt mode with MSI or Legacy interrupts, or in polled mode. Only one mode can be selected. The driver is controlled by:

- Adding TH_BH_ISR in the `driver/xdma/Makefile`, which causes the base DMA driver to run in interrupt mode.

By default, polled mode of operation is enabled.

Size of Block Data

The default amount of data being transmitted and received in the block data driver, which changes the throughput observed with this driver, is modified by:

- Modifying NUM_BUFS in `driver/xblockdata/user.c` to change the number of buffers in the free pool available to the driver.

Do not exceed the available system memory when changing these defaults.

Checksum Offload

By default, the hardware supports checksum offload as defined by the C_TEMAC0_TXCSUM and C_TEMAC0_RXCSUM parameters in `s6_pcie_dma_ddr3_gbe.v`. To inform the TCP/IP stack layers of checksum offload capability in hardware, the driver needs to be compiled with an additional flag. Navigate to the `driver/xgbeth` folder and modify the Makefile to include `-DENABLE_CSO` in the EXTRA_CFLAGS option.

Recompiling the Ethernet driver with this option enables the checksum offload feature. Checksum offload reduces CPU utilization.

Software Driver Code Modifications

This section describes modifications to the software driver code to change design behavior or performance by modifying the block data handler (`driver/xblockdata/user.c`).

Data is written into DDR3 memory in a flat, unstructured manner, with known patterns. It is possible to create a packet format, with some form of CRC, which can be verified on the receive path. Packets are generated and verified within the driver and are not conveyed to or from any real user application as data. Transferring this data between the driver and a user application requires significant changes in the driver entry points and in the driver's PutPkt() and GetPkt() routines. The data is transmitted (written) into DDR3 memory, and is looped back and received (read) from DDR3 memory.

Design Top-Level Modifications

This section describes changes to parameters in the top-level design file which can change the design behavior. Modifications to the software driver might be required based on the parameters being changed.

Hardware-Only Modifications

This section outlines the changes that only require parameter changes in the design top-level file (`source/s6_pcie_dma_ddr3_gbe.v`). No change to software is required.

PCIe High-Performance Mode

The Endpoint block for PCI Express provides an optional high-performance mode utilizing extra block RAMs which increases the credits as more packet buffering space is available. This mode can be enabled by defining PCIE_HIGH_PERF during design implementation. Enabling this option shows a change in performance.

Hardware and Software Modifications

This section outlines changes to the top-level design file (`s6_pcie_dma_ddr3_gbe.v`) which also requires software driver modifications.

Jumbo Frames

To enable jumbo frames, transmit and receive FIFO size in XPS-LL-TEMAC is varied by varying the following parameters in the top-level file:

- C_TEMAC0_TXFIFO in `design/source/s6_pcie_dma_ddr3_gbe.v` modifies the transmit FIFO depth
- C_TEMAC0_RXFIFO in `design/source/s6_pcie_dma_ddr3_gbe.v` modifies the receive FIFO depth

The corresponding change in software requires jumbo frames to be enabled in the Ethernet handler:

- Add ENABLE_JUMBO in the `driver/xgbeth/Makefile`

Larger storage for packets implies a smaller number of packets being dropped at the Ethernet receive interface which reduces retransmissions from upper layers in the TCP stack.

PCIe Vendor and Device ID

The vendor and device ID for PCI Express are changed by changing parameters in the top level file:

- CFG_VEN_ID in the file `design/source/s6_pcie_dma_ddr3_gbe.v` changes the vendor ID
- CFG_DEV_ID in the file `design/source/s6_pcie_dma_ddr3_gbe.v` changes the device ID

The corresponding change in software:

- PCI_VENDOR_ID_DMA: Change this macro in `driver/xdma/xdma_base.c`
- PCI_DEVICE_ID_DMA: Change this macro in `driver/xdma/xdma_base.c`

Architectural Modifications

This section describes architecture level changes to the functionality of the platform. These include removing and inserting a delivered application.

Aurora IP Integration

The LogiCORE IP Aurora 8B/10B implements the Aurora 8B/10B protocol using the high-speed GTP transceivers. The core is a scalable, lightweight link-layer protocol for high-speed serial communication. It is used to transfer data between two devices using transceivers. It provides an easy-to-use LocalLink compliant framing interface. This core is generated from the CORE Generator software.

A 1-lane Aurora design with 4-byte user interface data width can be connected in place of XPS-LL-TEMAC, as shown in Figure 5-1.

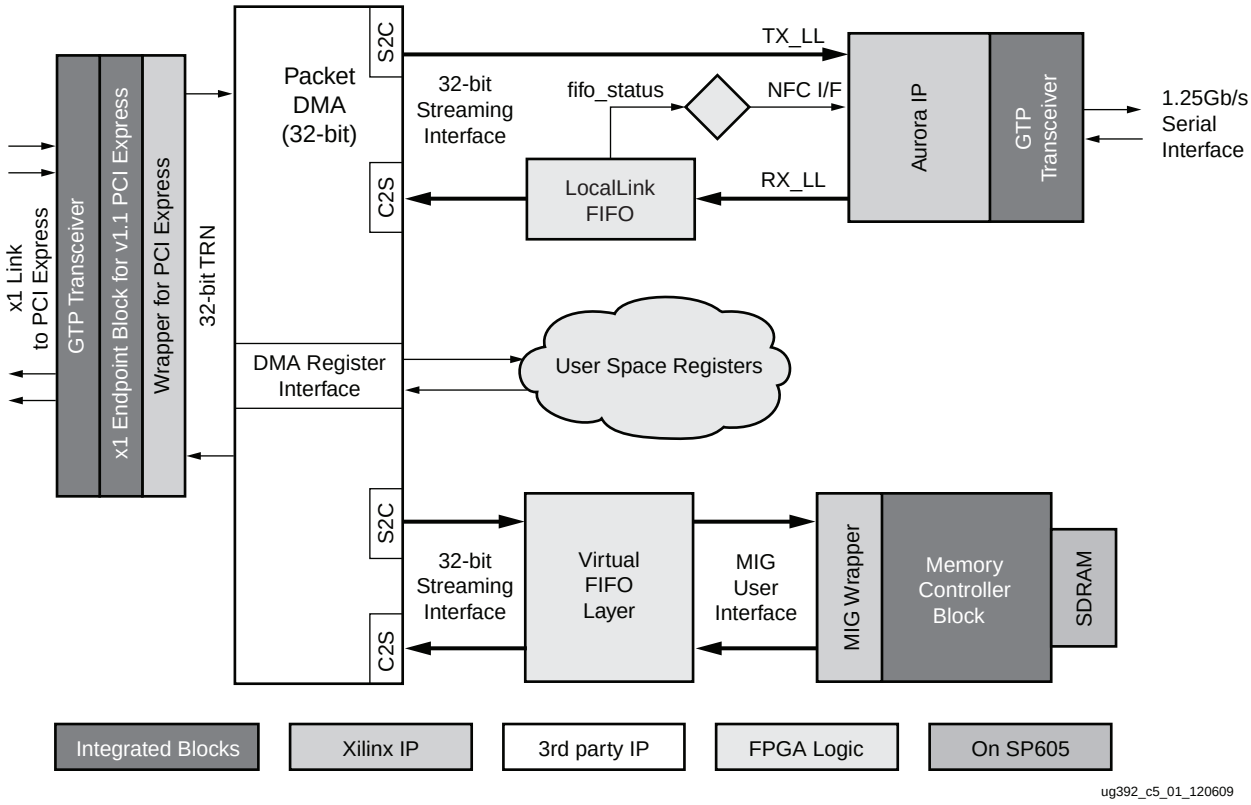


Figure 5-1: **Integrating Aurora**

The Aurora core does not support throttling in the receive direction as the core has internal buffers.

The suggested approach is to use a FIFO between the DMA and Aurora and to use the native flow control (NFC) in Aurora to prevent FIFO overflow. FIFO overflow control through NFC is a widely used option.

A LocalLink FIFO can be used for this purpose. A FIFO output indicating percentage of FIFO being occupied can be used to drive NFC.

The round trip delay through the Aurora interfaces between the NFC request and the first pause arriving at the originating channel partner must not exceed 256 symbol times.

For a 1.25 Gb/s rate, 1 symbol = 10 × 800 ps = 8 ns. With a 256 symbol time, the result is 256 × 8 ns = 2048 ns

With a 62.5 MHz clock (16 ns period), this is 128 clock cycles (which is the worst-case delay). If a LL-FIFO of depth 512 is used, then the NFC should be asserted once it is half-full.

Instead of the network driver, the same block data driver code provided is used to drive traffic over Aurora. The Aurora serial interface can be looped back externally or connected to another Aurora link partner.

Using Multiple Virtual FIFO Instances

The current design uses one virtual FIFO, which utilizes one 32-bit bidirectional port on the Spartan-6 FPGA memory controller block. As additional memory bandwidth is available, the same SDRAM can be partitioned to implement multiple FIFOs. This can be achieved by multiple instantiations of the virtual FIFO logic. The example in Figure 5-2 shows three memory controller ports. Each virtual FIFO interface has a dedicated address range within DDR3. This address range is defined by start and end address ranges for each virtual FIFO instance. Inline processing (as shown) can be added between the virtual FIFO interfaces.

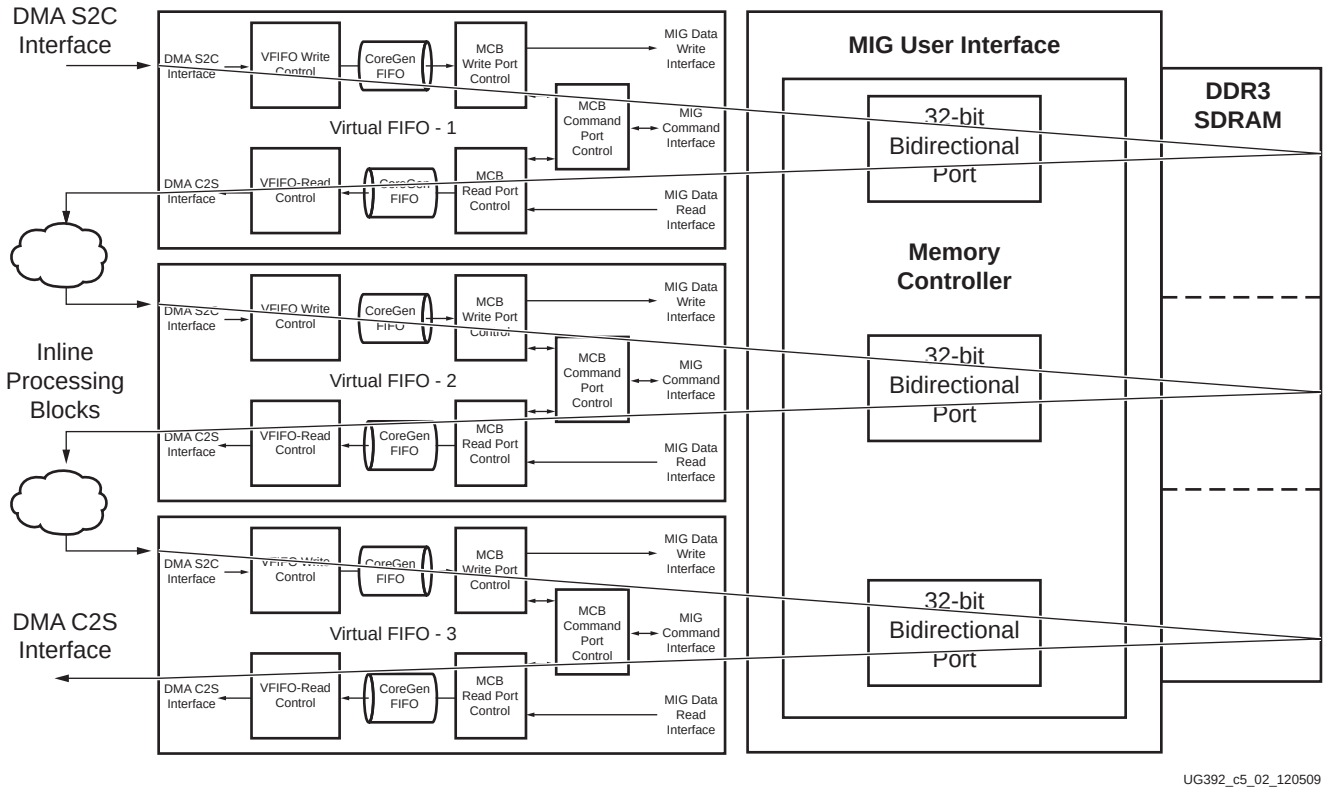


Figure 5-2: **DDR3 as Multiple Virtual FIFO**

Accordingly, a software application can be developed as detailed in Software Driver Code Modifications, page 80.

If the application developed is for image manipulation or processing, certain image processing operations (for example: image rendering or color inversion) can be offloaded to hardware as inline processing operations (shown in the blocks in Figure 5-2). Other examples of inline processing include an operation on chunks of data (CRC calculation) or some signal processing transformations (FFT or digital filtering).

XILINX®

Appendix A

Register Description

This appendix is a quick reference to describe the registers programmed by the software driver. For all registers and further details, refer to the specific user guides.

This appendix also describes the hardware registers and mapping of these registers with respect to the base address register (BAR) in PCI Express.

All DMA engine registers are mapped to BAR0. Table A-1 describes the mapping of multiple channel registers.

Table A-1: **DMA Channel Register Address**

DMA Channel	Offset from BAR0
Channel-0 S2C	0x0
Channel-1 S2C	0x100
Channel-0 C2S	0x2000
Channel-1 C2S	0x2100

Registers for interrupt handling in the DMA are grouped under a category called common registers. These registers are offset from BAR0 by 0x4000.

Figure A-1 shows the layout of registers.

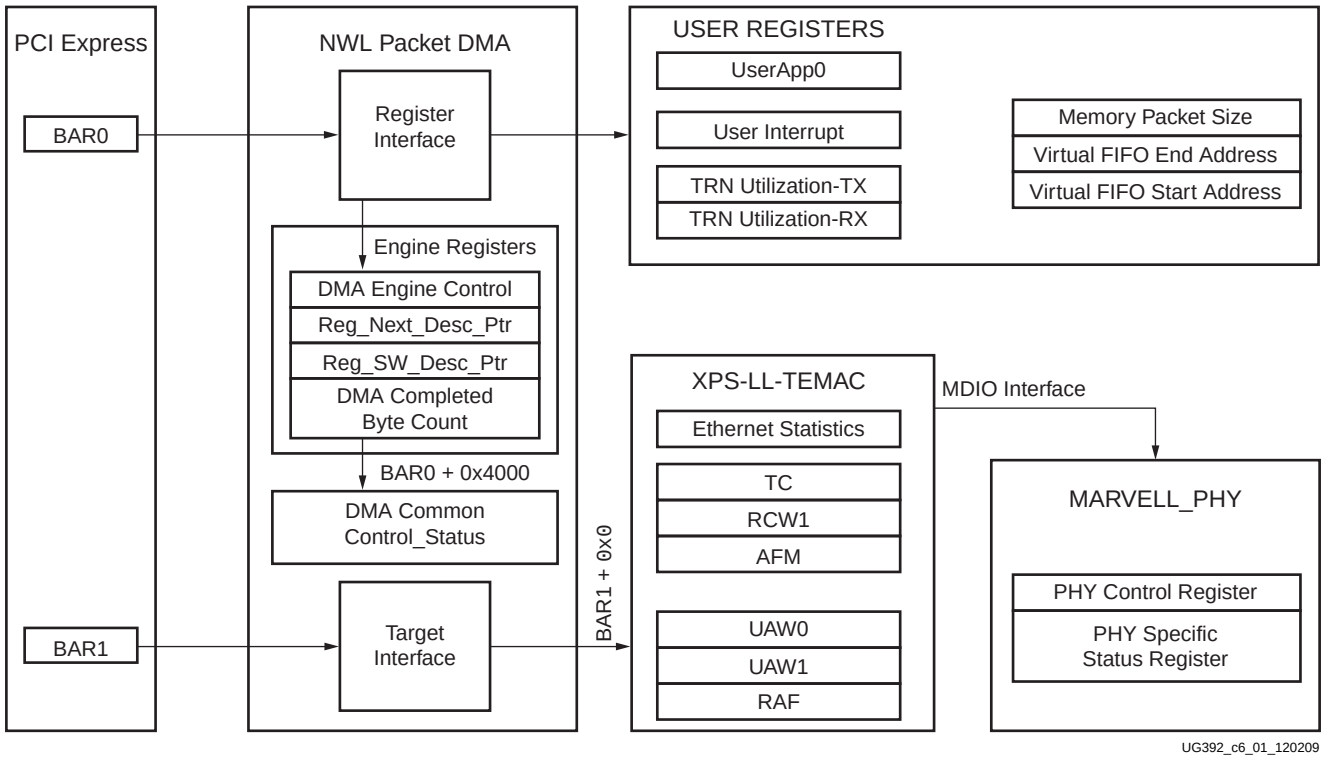


Figure A-1: **Register Map**

The user logic registers are mapped as described in Table A-2. XPS-LL-TEMAC and external PHY registers are mapped to BAR1.

Table A-2: **User Register Address Offsets**

User Logic Register Group	Range (Offset from BAR0)
User Application Advertisement Registers	0x8000–0x80FF
User Interrupt Registers	0x8100–0x81FF
TRN Utilization Registers	0x8200–0x82FF
User App0 Registers	0x9000–0x90FF
User App1 Registers	0x9100–0x91FF

DMA Registers

This section describes the prominent DMA registers frequently used by the software driver. For a detailed description of all registers available, please refer to the Northwest Logic Packet DMA Backend Core User Guide, page 10.

Channel Specific Registers

The registers described in Table A-3 through Table A-6 are present in all channels. The address of the register is the channel address offset from BAR0 plus the register offset.

DMA Engine Control (0x0004)

Table A-3: **DMA Engine Control Register (0x0004)**

Bit	Field	Mode	Default Value	Description
0	Interrupt enable	RW	0	Enables interrupt generation.
1	Interrupt active	RW1C	0	Interrupt active is set whenever an interrupt event occurs. Write a 1 to clear.
2	Descriptor complete	RW1C	0	Asserted when an interrupt on completion bit is set in the descriptor.
3	Descriptor alignment error	RW1C	0	Asserted when the descriptor address is unaligned and that DMA operation is aborted.
4	Descriptor fetch error	RW1C	0	Asserted when the descriptor fetch errors out. That is, the completion status is not successful.
5	SW_Abort_Error	RW1C	0	Asserted when the DMA operation is aborted by software.
8	DMA Enable	RW	0	Enables the DMA engine and once enabled, the engine compares the next descriptor pointer and software descriptor pointer to begin execution.
10	DMA_Running	RO	0	Indicates DMA in operation.
11	DMA_Waiting	RO	0	Indicates DMA waiting on software to provide more descriptors.
14	DMA_Reset_Request	RW	0	Issues a request to user logic connected to DMA to abort outstanding operation and prepare for reset. This is cleared when user acknowledges the reset request
15	DMA_Reset	RW	0	Asserting this bit resets the DMA engine and issues a reset to the user logic

Next Descriptor Pointer (0x0008)

Table A-4: **DMA Next Descriptor Pointer Register**

Bit	Field	Mode	Default Value	Description
[31:5]	Reg_Next_Desc_Ptr	RW	0	Next Descriptor Pointer: Writable when DMA is not enabled. It is read only when DMA is enabled. This should be written to initialize the start of a new DMA chain.
[4:0]	Reserved	RO	5'b00000	Required for 32-byte alignment

Software Descriptor Pointer (0x000C)

Table A-5: **DMA Software Descriptor Pointer Register**

Bit	Field	Mode	Default Value	Description
[31:5]	Reg_SW_Desc_Ptr	RW	0	Software Descriptor Pointer: The location of the first descriptor in the chain, which is still owned by the software
[4:0]	Reserved	RO	5'b00000	Required for 32-byte alignment

Completed Byte Count (0x001C)

Table A-6: **DMA Completed Byte Count Register**

Bit	Field	Mode	Default Value	Description
[31:2]	DMA_Completed_Byte_Count	RO	0	Completed Byte Count: Records the number of bytes that transferred in the previous second. It has a resolution of four bytes.
[1:0]	Sample Count	RO	0	Sample Count: Incremented every time a sample is taken at a one second interval.

Common Registers

The registers described in this section are common to all engines and are located at the given offsets from BAR0.

Common Control and Status (0x4000)

Table A-7: **DMA Common Control and Status Register**

Bit	Field	Mode	Default Value	Description
0	Global Interrupt Enable	RW	0	Globally enables or disables interrupts for all DMA engines.
1	Interrupt Active	RO	0	Reflects the state of the DMA interrupt hardware output considering the state of the global interrupt enable.
2	Interrupt Pending	RO	0	Reflects the state of the DM A interrupt output without considering the state of the global interrupt enable.
3	Interrupt Mode	RO	0	0: MSI mode 1: Legacy interrupt mode
4	User Interrupt Enable	RW	0	Enables generation of user interrupts.
5	User Interrupt Active	RW1C	0	Indicates an active user interrupt.
23:16	S2C Interrupt Status	RO	0	Bit[i] indicates interrupt status of S2C DMA engine[i]. If S2C engine is not present, then this bit is read as zero.
31:24	C2S Interrupt Status	RO	0	Bit[i] indicates interrupt status of C2S DMA engine[i]. If C2S engine is not present, then this bit is read as zero.

Network Path IP Registers

This section defines the commonly used XPS-LL-TEMAC and PHY registers. For a detailed explanation of all registers, please refer to the respective user guides.

XPS-LL-TEMAC Registers

The XPS-LL-TEMAC contains memory and addressable registers for read/write operations. It is assumed that only TEMAC0 is used. The memory map is divided into three types:

- Soft Registers: Registers in the XPS-LL-TEMAC wrapper; for example, the reset and statistics registers.

- Direct Registers: Registers in the soft TEMAC core
- Indirect Registers: Registers in the soft TEMAC core which are indirectly addressable or external PHY registers. These registers are accessed through directly accessible registers as detailed in the XPS-LL-TEMAC data sheet.

The register offsets are mentioned from BAR1 for directly addressable registers.

Reset and Address Filter Register (0x0)

Table A-8: **Reset and Address Filter Register**

Bit	Field	Mode	Default Value	Description
31	HtRst	RW	0	TEMAC Reset: This bit provides a means for resetting the soft TEMAC core. This bit is self clearing. 0: Normal operation, TEMAC core not reset 1: Initiate a reset of the TEMAC core
18	StatsRst	RW	0	Statistics Counters Reset: This bit provides a means for resetting the statistics counters if present. This bit is self clearing. 0: Normal operation, statistics counters not reset 1: Initiate a reset of the statistics counters

Statistics Registers

Only certain statistics registers which indicate errors are read in the design.

Table A-9: **Statistics Register**

Offset	Register	Description
0x298	FCS Errors (lower 32 bits)	A count of received frames that failed the CRC check and were at least 64 bytes in length.
0x2B8	Length/Type out of range (lower 32 bits)	A count of frames received that had length/type field not matching the number of data bytes received.
0x2F0	Underrun Errors (lower 32 bits)	A count of frames that would otherwise be transmitted but could not be completed due to FIFO underrun.

Receive Configuration Word Register (Indirect, 0x240)

This register sets the behavior of the receive TEMAC interface.

Table A-10: **Receive Configuration Word Register**

Bit	Field	Mode	Default Value	Description
31	RST	RW	0	Reset: When this bit is 1, the receiver is reset. The bit automatically resets to 0. The reset also sets all of the receiver configuration registers to default values. 0: No reset 1: Initiates a receiver reset
30	JUM	RW	1	Jumbo Frame Enable: When this bit is 1, the receiver accepts frames over the maximum length specified in the IEEE 802.3 specification. 0: Receive jumbo frames disabled 1: Receive jumbo frames enabled

Table A-10: **Receive Configuration Word Register *(Cont'd)***

Bit	Field	Mode	Default Value	Description
29	FCS	RW	1	In-Band FCS Enable: When this bit is 1, the receiver provides the FCS field with the rest of the frame data. When this bit is 0 the FCS field is stripped from the receive frame data. In either case the FCS field is verified. 0: Strip the FCS field from the receive frame data 1: Provide the FCS field with the receive frame data
28	RX	RW	1	Receive Enable: When this bit is 1, the receiver logic is enabled to operate. When this bit is 0, the receiver ignores activity on the receive interface. 0: Receive disabled 1: Receive enabled
25	LT_DIS	RW	0	Length/Type Field Valid Check Disable: When this bit is 1, it disables the Length/Type field check on the receive frame. 0: Perform Length/Type field check 1: Do not perform Length/Type field check

Transmit Configuration Word Register (Indirect, 0x280)

This register sets the behavior of the transmit path of the TEMAC.

Table A-11: **Transmit Configuration Word Register**

Bit	Field	Mode	Default Value	Description
31	RST	RW	0	Reset. When this bit is 1, the transmitter is reset. The bit automatically resets to 0. The reset also sets all of the transmitter configuration registers to their default values. 0: no reset 1: initiates a transmitter reset
30	JUM	RW	1	Jumbo Frame Enable When this bit is 1, the transmitter sends frames over the maximum length specified in IEEE 802.3 specification. 0: send jumbo frames disabled 1: send jumbo frames enabled
29	FCS	RW	0	In-Band FCS Enable. When this bit is 1, the transmitter accepts the FCS field with the rest of the frame data. When this bit is 0 the FCS field is calculated and supplied by the transmitter. 0: transmitter calculates and sends FCS field 1: FCS field is provided with transmit frame data
28	TX	RW	1	Transmit Enable. When this bit is 1, the transmit logic is enabled to operate. 0: transmit disabled 1: transmit enabled

Management Configuration Register (Indirect, 0x340)

This register programs the MDIO clock divider and successful programming of this register generates the MDIO clock used to program the PHY.

Table A-12: **Management Configuration Register**

Bit	Field	Mode	Default Value	Description
6	MDIO_EN	RW	0	MDIO Enable: When this bit is 1, the MDIO (MII Management) interface is used to access the PHY. 0: MDIO disabled 1: MDIO enabled
5:0	CLK_DIVIDE	RW	0	Clock Divide: This value is used to derive the MDC (MII Management interface clock) signal. The maximum permitted frequency is 2.5 MHz.

Address Filter Mode Register (Indirect, 0x390)

Table A-13: **Address Filter Mode Register**

Bit	Field	Mode	Default Value	Description
31	PM	RW	0	Promiscuous Receive Address Mode Enable: When this bit is 1, the receive address filtering is disabled and all destination addresses are accepted. When this bit 0, the receive address filtering is enabled. 0: address filtering enabled 1: address filtering disabled (all addresses accepted)

User Application Registers

The various user registers are described in this section. All registers are 32 bits wide. Bit fields not defined are considered reserved with a read always returning a value of zero.

Design Version Register (0x8000)

This register tracks the design version so that code maintenance is easily traceable. The software driver uses this register to associate the correct version with the hardware design.

Table A-14: **Design Version Register**

Bit Location	Field	Mode	Default Value	Description
31:28	Device	RO	0000	• 0000: Spartan-6 • 0001: Virtex-6
11:4	Version	RO	0001_0000	Defines TRD version; updated based on release versions.
3:0	Sub-version	RO	0000	Non-AXI version of design.

User Application Advertisement Registers

UserApp Advertisement Register (0x8004)

This register advertises which user application is connected to which DMA engine. This enables the software to associate appropriate descriptors with relevant DMA engines in case Ethernet path and memory path are swapped in hardware.

The following bits in the register are engine indicators:

- [31:28]: S2C Engine 0
- [27:24]: S2C Engine 1
- [15:12]: C2S Engine 0
- [11:8]: C2S Engine 1

Table A-15: **UserApp Advertisement Register**

Bit Location	Field	Mode	Default Value	Description
31:28	S2C_0	RO	0001	A value of 0001 indicates network path connected to S2C engine-0.
27:24	S2C_1	RO	0010	A value of 0010 indicates memory path connected to S2C engine-1.
15:12	C2S_0	RO	1001	A value of 1001 indicates network path connected to C2S engine-0.
11:8	C2S_1	RO	1010	A value of 1010 indicates memory path connected to C2S engine-1.

User Interrupt Registers

These registers handle the various interruptible conditions in the user application. For an interruptible condition, if the interrupt is enabled, an user interrupt is signaled to the DMA which gets converted to either MSI or a legacy interrupt message upstream depending on the interrupt mode enabled by the software driver in the configuration space for PCI Express.

User Interrupt Enable Register (0x8100)

This is the user interrupt enable register which enables/disables specific user interrupts.

Table A-16: **User Interrupt Enable Register**

Bit Location	Field	Mode	Default Value	Description
31	PLB Error Enable	RW	0	Setting this bit enables PLB error conditions to generate interrupts over PCIe.
30	TEMAC Interrupt Enable	RW	0	Setting this bit enables TEMAC interrupt conditions to generate interrupts over PCIe.

Table A-16: **User Interrupt Enable Register *(Cont'd)***

Bit Location	Field	Mode	Default Value	Description
29	MCB TX Error Enable	RW	0	Setting this enables MCB specific TX error conditions to generate interrupts to the system.
28	MCB RX Error Enable	RW	0	Setting this enables MCB specific RX error conditions to generate interrupts to the system.

The enable fields described in Table A-16 are set in combination with interruptible conditions generates a user interrupt into DMA which translates to interrupts on a PCIe link.

User Interrupt Status Register (0x8104)

This is the user interrupt status register which indicates what caused the user interrupt. Relevant bits get set on corresponding errors. Software is required to write a 1 to clear the set bits which acknowledge the user interrupt.

Table A-17: **User Interrupt Status Register**

Bit Location	Field	Mode	Default Value	Description
31	PLBError	RW	0	Indicates PLB error as cause of user interrupt. Write 1 to clear.
30	TEMAC Interrupt	RW	0	Indicates TEMAC error as the cause of user interrupt. Write 1 to clear.
29	MCB_TxErr	RW	0	Memory controller error on transmit interface (MCB TX FIFO overflow). Write 1 to clear.
28	MCB_RxErr	RW	0	Memory controller error on receive interface (MCB RX FIFO underflow). Write 1 to clear.

TRN Monitor Registers

This defines the registers implemented for measuring TRN utilization.

These registers are updated once every second by hardware. These registers have a resolution of four bytes and provide a 2-bit sample count which increments every second. The sample count provides a mechanism for software to keep track of distinct reads and also to synchronize register values across the same one second interval.

Transmit Utilization Byte Count (0x8200)

This register counts the utilization of the transmit interface of the PCIe core. It increments every clock cycle when both trn_tx_src_rdy_n and trn_tx_dst_rdy_n are asserted.

Table A-18: **Transmit Utilization Byte Count Register**

Bit Location	Field	Mode	Default Value	Description
31:2	Transmit Utilization Count	RO	0	Gives the count when TRN-TX interface was active. This register has a resolution of four bytes. Multiply the value obtained by four to get the byte count.
1:0	Sample Count	RO	0	A 2-bit sample count which increments once every second.

Receive Utilization Byte Count (0x8204)

This register counts the utilization of the receive interface of the Endpoint for PCI Express. It increments every clock cycle when both `trn_rx_src_rdy_n` and `trn_rx_dst_rdy_n` are asserted.

Table A-19: **Receive Utilization Byte Count**

Bit Location	Field	Mode	Default Value	Description
31:2	Receive Utilization Count	RO	0	This gives the count when TRN-RX interface was active. This register has a resolution of four bytes. Multiply the value obtained by four to get the byte count.
1:0	Sample Count	RO	0	2-bit sample count which increments once every second.

Upstream Memory Write Byte Count (0x8208)

This register counts the payload of memory write transactions sent upstream on the transmit interface of the PCIe core.

Table A-20: **Upstream Memory Write Byte Count**

Bit Location	Field	Mode	Default Value	Description
31:2	MWR Payload Count	RO	0	This gives the count of MWR payload bytes sent across TRN-TX. This register has a resolution of four bytes. Multiply the value obtained by four to get the byte count.
1:0	Sample Count	RO	0	2-bit sample count which increments once every second

Downstream Completion Payload Byte Count (0x820C)

This register counts the payload of completion transactions received at the Endpoint on the receive interface for PCI Express.

Table A-21: **Downstream Completion Payload Byte Count**

Bit Location	Field	Mode	Default Value	Description
31:2	CplD Payload Count	RO	0	Gives the count of the CplD payload bytes received across TRN-RX. This register has a resolution of four bytes. Multiply the value obtained by four to get the byte count.
1:0	Sample Count	RO	0	2-bit sample count which increments once every second

TRN Monitor Control (0x8210)

This is the monitor control registers which defines a software controlled reset. When asserted, this clears the counters.

Table A-22: **TRN Monitor Control Register**

Bit Location	Field	Mode	Default Value	Description
0	Monitor Reset	RW	0	Monitor Soft Reset: When 1, resets the TRN monitor counters.

User App1 Registers

This group defines the registers specific to user application connected to DMA channel-1, which is the memory application for this design.

Virtual FIFO Status Register (0x9100)

This register indicates the status of DDR3 calibration to the software driver. It enables software to determine if the hardware is ready for operation.

Table A-23: **Virtual FIFO Status Register**

Bit Location	Field	Mode	Default Value	Description
0	Calibration Status	RO	0	Calibration Done: This bit indicates calibration done status from memory controller.

Virtual FIFO Receive Packet Length Register (0x9104)

This register indicates the size of the packet in bytes to be built in the receive direction. It initializes with a default value of 1 KB.

Table A-24: **Virtual FIFO Receive Packet Length Register**

Bit Location	Field	Mode	Default Value	Description
31:0	Packet Length	RW	32'h0000_0400	DDR3 Receive Packet Length: Indicates the size of the packet (in bytes) to be built in the receive direction.

Virtual FIFO Start Address Register (0x9108)

This register indicates the start address for DDR3 partition. It initializes with the default value of zero on reset. Software programming of this register is optional.

Table A-25: **Virtual FIFO Start Address Register**

Bit Location	Field	Mode	Default Value	Description
31:0	Start Address	RW	0x0	DDR3 Start Address: Indicates the start address in DDR3 from where virtual FIFO starts.

Virtual FIFO End Address Register (0x910C)

This register indicates the end address for DDR3 partition. It initializes with the default value of 32'h07FF_FFFF on reset. Software programming of this register is optional.

Table A-26: **Virtual FIFO End Address Register**

Bit Location	Field	Mode	Default Value	Description
31:0	End Address	RW	32'h07FF_FFFF	DDR3 End Address: Indicates the end address in DDR3 where the virtual FIFO wraps around to the start address.

Virtual FIFO Error Statistics Register (0x9110)

This register is the DDR3 error statistics register which records an error count on DDR3. This register accumulates the DDR3 error count and is cleared on reset.

Table A-27: **Virtual FIFO Error Statistics Register**

Bit Location	Field	Mode	Default Value	Description
31:0	Error Stats	RW	0	DDR3 Error Statistics.

Appendix B

Directory Structure

Introduction

This section describes the directory structure and explains the organization of various files/folders.

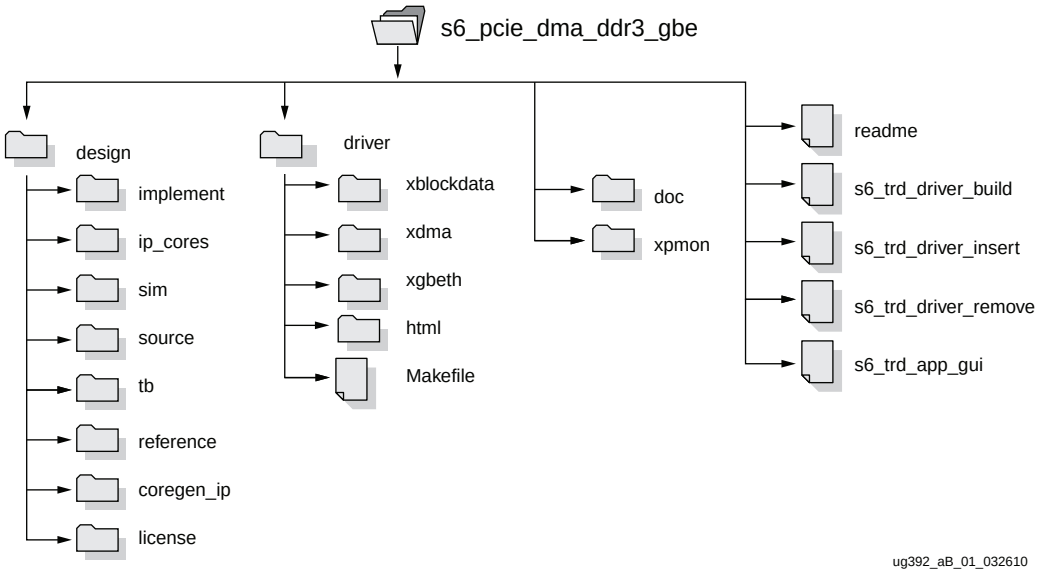


Figure B-1: **Directory Structure**

Design

The design folder contains all the hardware design deliverables.

- implement: Contains the implementation scripts for the design for both windows and Linux operating systems supporting both command line mode and the ProjNav flow.
- ip_cores: Contains the third-party DMA IP related files and Xilinx IP files modified for this TRD.
- sim: Contains the simulation scripts for supported simulators for both windows and Linux operating systems
- source: Contains the source code deliverable files
- tb: Contains the testbench related files for simulation
- reference: Bit files and MCS files for golden reference and XCO files generated by Xilinx cores. Also includes scripts for the ProjNav flow.

- coregen_ip: Includes the CORE Generator IP and netlists used in the design
- license: Includes hardware evaluation license for Xilinx IPs

Driver

The driver folder contains all the software driver and application deliverables.

- xblockdata: Contains the source code for the block data driver
- xdma: Contains the source code for the packet DMA driver
- xgbeth: Contains the source code for the Ethernet driver
- html: Contains the software driver documentation files generated by Doxygen.
- Makefile: Contains the Makefile for the software driver and application compilation

xpmon

Contains source code for the application GUI.

doc

Contains the TRD user guide.

readme

Details the use of various simulation and implementation scripts.

s6_trd_driver_build

Contains the script to build the driver and GUI modules.

s6_trd_driver_insert

Contains the script to insert the driver modules.

s6_trd_driver_remove

Contains the script to remove the driver modules.

s6_trd_app_gui

Contains the script to invoke the XPMON application GUI

Appendix C

Setting Up a Private LAN

Introduction

This section describes the steps used to set up a private LAN connection between two machines for Ethernet performance measurement.

Figure C-1 shows a private LAN connection between two machines.

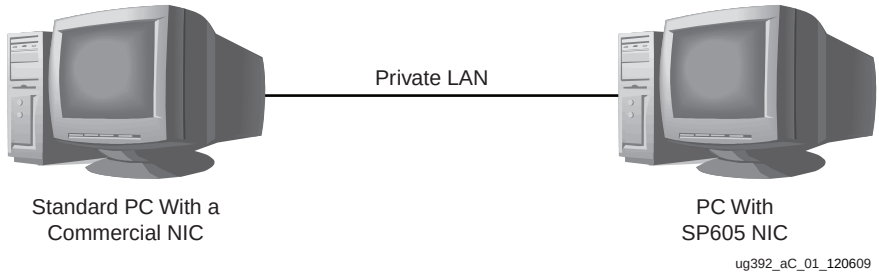


Figure C-1: **Private LAN Setup**

To set up a private LAN connection:

1. Connect the Ethernet cable between the two machines; connect as a private LAN setup. One of them is a standard machine which has a commercial NIC and the other has SP605 NIC. The machine with SP605 NIC is referred to as the unit-under-test (UUT) and the other machine, with a commercial NIC, is referred to as the standard machine.
2. Assign an IP address statically on both machines. Make sure that they have the same netmask. This can be done on a terminal in command line mode:

   ```
   $ ifconfig ethX up 172.16.64.7
   ```

   For this example, it is assumed that the standard machine is assigned an IP address of 172.16.64.9 and the UUT is assigned an IP address of 172.16.64.7.
3. After the interface is activated and after assignment of a static IP address, try a ping between the machines.
4. Install Netperf v2.4 on both machines. Netperf works with a client server model. In this setup, UUT is programmed as the client and the other standard machine as the server. On a terminal on the standard machine, invoke netserver:

   ```
   $ netserver
   ```
5. Open a terminal on the UUT and try running Netperf

   ```
   $ netperf -H <IP-address-standard machine>
   ```

   This command runs a ten second TCP_STREAM test by default and reports outbound performance. Refer to the Netperf manual for the various test options available.

XILINX®

Appendix D

Troubleshooting

Introduction

This section includes some troubleshooting tips (Table D-1). It is not meant as an exhaustive troubleshooting guide. It is based on the following assumptions:

- User has followed instructions as explained in Chapter 2, Getting Started.
- User has made sure that the PCI Express link is up and the Endpoint device is discovered by the host and can be seen with `lspci`.
- Visual indicators (LEDs) as listed on page 18 are functioning and have been checked.

Table D-1: **Troubleshooting Tips**

Issue	Possible Resolution
Activation of Ethernet interface fails with network configuration GUI	1. Check the MAC address to make sure the MAC address is programmed as provided with the SP605 Connectivity Kit. 2. If assigning an IP address statically, make sure that it does not clash with any other IP address on the network. Contact the network administrator regarding specific IP address allocation. 3. In the network configuration GUI, under the Devices tab, for the device, uncheck the **bind to MAC address** option in device properties. Device properties are invoked by double-clicking on the device.
Network is connected but webpage does not load in the browser	1. Check the browser's network proxy settings suitable for your network with your network administrator 2. Make sure that the browser is not in work-offline mode
1000BASE-X Design not working.	1. Make sure that the additional required hardware is connected as explained in Testing 1000BASE-X Mode, page 38. Make sure the Ethernet connection is 1 Gb/s. 2. Make sure that the correct design is downloaded to hardware (`sp605_use_1000basex`) 3. Make sure that `driver/xgbeth/Makefile` has **DUSE_1000BASEX** defined under **EXTRA_CFLAGS**