PCI Express Performance Measurement

User Guide

May 2020

Copyright and Legal Notice

History

The following table shows the History of this document.

Date	Version	Revision
13.02.2015	1.0	Initial Release
10.02.2016	1.1	Added chapter 4
09.12.2016	1.2	Added Intel FPGA Support
14.08.2018	2.0	New Feature : Blocklength and Completion Parameters, added Gen4 and Gen5 Link Speed Support for future compatibility
12.05.2020	2.1	Added 64-Bit support and explanation of histogram function

Table of Contents

1 Features of Smartlogic's PCIe DMA IP Core

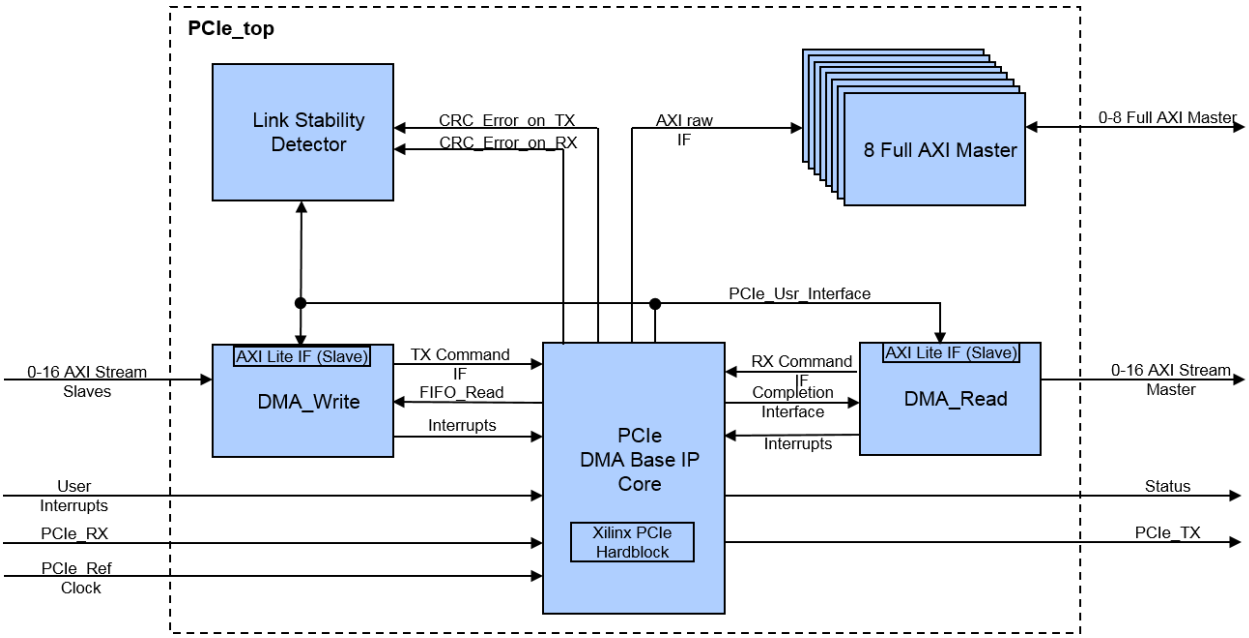
Smartlogic's PCI Express IP Core Suite allows the development of complex PCI-Express endpoints within FPGAs with only basic PCI Express protocol Know-how. The IP is built around the Xilinx/Intel PCI-Express Hard IP Macrocell and greatly simplifies the development of a powerful PCI-Express Endpoint due to standardized AXI Stream Slave and Full AXI Master Interfaces.

With this IP Core it is possible to manage up to 16 independent buffers in the Host Memory. With the powerful addressing and arbitration scheme it is possible to control the priority of the DMA data and to manage nearly any geometry in the Host Memory.

IP-Features

- PCI-SIG compliant solution
- Supports X1, X2, X4 or X8 Link Widths at Gen1, Gen2 or Gen3 speed
- Up to 8 AXI4 Masters (Burst & Single Transfers) to allow easy access of other FPGA peripherals
- Supports 32 Bit and 64 Bit DMA Addressing
- Support for either Scatter Gather or contiguous Memory
- Flexible and easy customizable Register Interface
- MSI and Legacy Interrupt Support
- Powerful customization of the Hardmacrocell regarding BARs, Reference Clock Speed and many other parameters
- Supported Technologies : Xilinx and Intel FPGAs
- Up to 16 independent AXI Stream Slaves for DMA Write transfers
- Up to 16 independent AXI Stream Masters for DMA Read transfers
- Independent Clocking of each AXI Stream channel possible
- Parametrizable Priority Control
- Memory Size up to 4 GByte per Streaming Channel
- Performance only limited by PCI-Express Bandwidth
- High Speed Simulation Mode

Block Diagram of the IP-Core:



FPGA Resources:

The following table lists different example configurations and the needed FPGA Resources:

FPGA Family	Configuration Width / Speed	User Registers[1]	FPGA-Resources[2] LUTs	FFs	BRAM	PCIE	Misc
Artix 7	64 bit Architecture X4, Gen 1 4 DMA Channels, 1 Master	15 x 32	8.803	8.521	12	1	1 MMCM
	256-Bit Architecture, X4, Gen 1 4 DMA Channels, 1 Master	15 x 32	15.934	13.156	12	1	1 MMCM

[1] The amount of user registers is design dependant

[2] These numbers include the resources of the Xilinx Endpoint Block Plus Hardcore IP

Link Speeds and Link Widths:

The following table lists the supported PCI-Express Link Speeds and Link widths for the different versions of the Demodesign:

Product Version	FPGA Family	Supported Link Speed / Link Width Gen 1, 2.5 GBit/s	Gen 2, 5.0 GBit/s	Gen 3, 8.0 GBit/s	
		X1 / X2 / X4 / X8	X1 / X2 / X4 / X8	X1 / X2 / X4	X8
64-Bit Core	Virtex 6, Kintex, Artix*, Zynq	✓	✓	--	--
256-Bit Core	Kintex, Artix*, Zynq	✓	✓	--	--
256-Bit Core	Virtex 7	✓	✓	✓	--
256-Bit Core	KintexUltrascale, VirtexUltrascale	✓	✓	✓	✓

[*] Artix does not support a x8 Link Width

2 PCI Express Performance Demodesign and Windows GUI

The PCI Express Performance Demodesign is a free evaluation application of our PCI Express IP Core which measures various significant PCI Express parameters. The measurement results will reveal the overall performance of the Host and important parameters like Maximum Payload Size (MPS), Maximum Read Request Size (MRS), Payload throughput, CRC Errors and others.

Compiled Bitstreams are available from Smartlogic for various demoboardplatforms as Xilinx AC701, KC705, VC709, Intel Cyclone V, Arria V and others for evaluation purposes at no cost.

The Demodesign is bundled with a Windows 7 Driver and a Graphical User Interface (GUI).

Supplied files and folder structure:

Folder	Remark
amd64\ amd64\sldemo_64.cat amd64\sldemo_64.inf amd64\sldemo_64.sys	Compiled Windows 7 driver 64bit
perf_test\ Perftest2.exe	Compiled Performancetest GUI for Windows

Note: For security reasons, the downloaded ZIP File is password protected. The password can be obtained at no cost by simply sending a request email to perf_eval[at]smartlogic.de . The password will be sent immediately to you. If you need a 32-Bit Windows 7 Device Driver, please get in contact with us and write an email to ip[at]smartlogic.de .

The following paragraphs provide guidelines how to install the Bitstream and the Driver and how to work with the GUI.

Installation:

The first installation step is to program the performance demo bitstream into the desired demoboard platform.

The following User Guides show in detail how to do this:

Demoboard	Xilinx/Intel User Guide
AC701	XTP227 See Sections “AC701 Setup”, “Hardware Setup” and “Program AC701 Flash with PCIe Design”
KC705	XTP197 See Sections “KC705 Setup”, “Program KC705 Flash with PCIe Design”
VC709	XTP237 See Sections “VC709 Setup”, “Program BPI Flash with PCIe Design”
KCU105	XTP350
VCU108	XTP366
Cyclone V Development Kit	MNL-01078 Cyclone V Developkit Reference Manual

Arria V GX Starter Board	MNL-01069 Arria V Starter Kit Reference Manual

In order to download these user manuals, go to the vendor's Webpage (www.xilinx.com , www.altera.com) and type in the User Guide Name in the Search field (e.g. "XTP227" for the Artix Version). You may need a Xilinx or Intel Account which can be created very easiliy at no cost.

When the Bitstream is programmed into the Demoboard, connect the board with the Host and power the system on.

After Windows has booted the new PCIe Device will be detected from Windows and you will be asked from Windows to install a device driver.

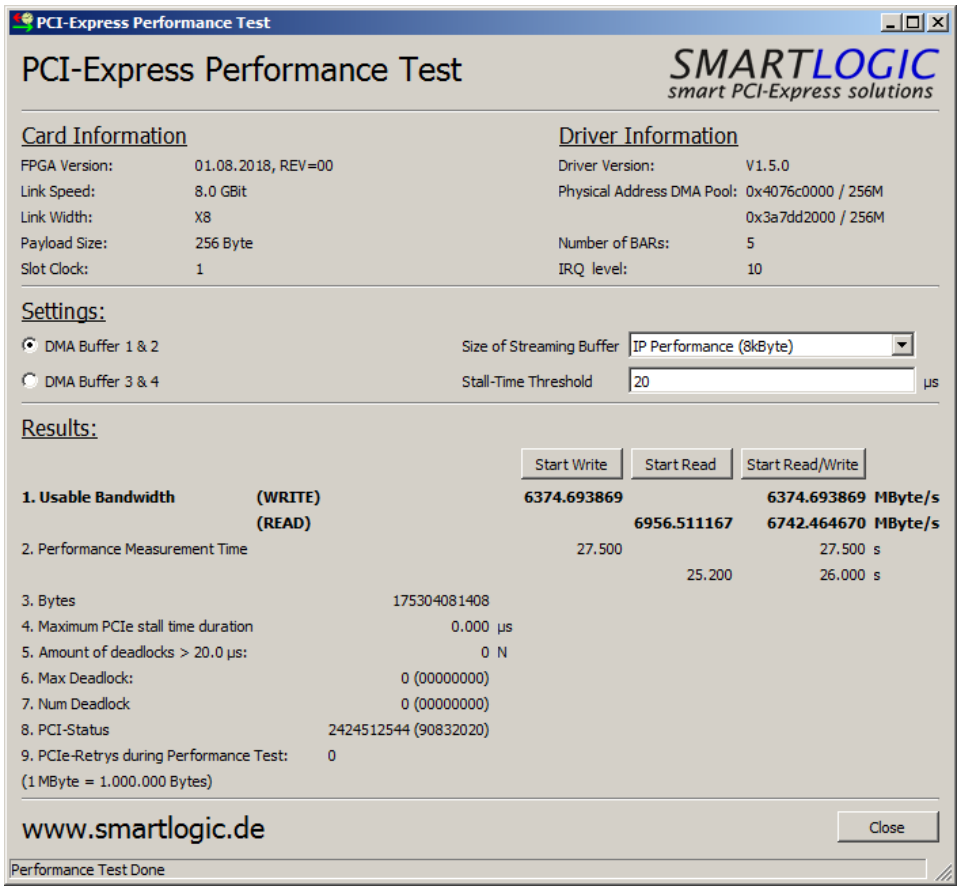
Decompress the downloaded zip folder that ontains the device driver to a local folder on your harddrive and instruct windows to install the driver from the local folder.

After successful installation re-boot Windows.

Installation of the GUI

The GUI itself does not need a specific installation procedure. Simply execute perftest.exe from the local folder. Make sure, that the two DLL Diles QTCore4.dll and QTgui4.dll reside in the same filder where perftest.exe is located.

If everything was installed correctly, the GUI should come up with the following screen:



De-Installation

In order to remove the GUI from you PC, simply delete the EXE file from the local folder.

The Driver can be removed easily from the "Device Manager". Simply click on the SL_DEMO Device and de-install the driver.

3 Working with the GUI – Measurement Results

The GUI displays several informational details about the connected endpoint in the “Card Information” Section.

In order to measure the real payload throughput (Payload = User Data without protocol Overhead) simply press one of the three buttons “Start Write”, “Start Read” or “Start Read/Write”. This will generate a well known amount of data traffic in the desired direction. The FPGA measures the time to transfer the data. After the traffic generators have sent their data, the time and the MB/s value is displayed on the GUI.

“Write” refers to the direction from the FPGA to the Host and “Read” refers to the direction from Host to FPGA.

Note: These measurements provide the real data throughputs and contains all impacts like protocol overhead, possibly poor flow control update and power management setting. The results can differ significantly on different host platforms.

Further Results after a performance measurement:

The following table lists all further informations that are available after a measurement run:

Result	Explanation
3. Bytes	This is the amount of bytes that was transferred over PCI Express for measuring the performance. The amount is calculated out of Link Width and Link Speed in order to have a test duration of approximately 30 Seconds
4. Maximum PCIe Stall time duration	This is the longest time duration during the measurement, where no data could be transferred over PCIe. Typical values are 1us to 12 us. Greater values indicate, that there are severe system impacts like power management issues. This value is important do know in order to design FIFO depths big enough without data overflow.
5. Amount of deadlocks > 50 us	This value shows how often a PCIe Stall happened with a duration of greater than the selected stall time threshold while the performance test was running.
6. Shortest Completion Dataphase	This shows the smallest datacontent of a databeat and is an indicator, if a completion was split in several smaller completions. Typically this value should be 32 for the 256 Bit Architecture and 8 for the 64-Bit architecture of the core.
7. Biggest Completion	This indicates the maximum payload of a received completion and shows, if completion combining is active.
8. PCI-Status	This value contains side information on the PCI Express Link. This is a 32-Bit value (hex in brackets) and has the following meaning: Bit 0 : Correctable Error Reporting Enable Bit 1: non fatal Error Reporting Enable Bit 2: Non_fatal reporting enable Bit 3: Unsupported Request Reporting enable Bit 4 : Enable relaxed ordering Bits 7:5 : Maximum Payload Size Bit 8: Extended Tag supported Bit 9 : Phantom Functions enable Bit 10 : Auxiliary Power PM Enable Bit 11 : Enable no snoop Bits 14:12 : Maximum Read Request size Bits 19:16: current Link Speed Bits 25:20: current Link Width Bit 31: MSI Enabled
9. PCIe Retries during Performance Test	This field reports the number of CRC Errors encountered during the measurements.

4 DMA Performance Demo – Measurement Flow without GUI

In case of non Windows based Operating Systems, it is possible to measure the performance values by programming the Endpoint Registers and reading the results directly.

However the user is responsible to reserve 2 separate 8kByte blocks of contiguous Host memory as source and destination for the tests. These 2 memory blocks should be reserved by appropriate OS system calls in order to ensure that the system will not become damaged, when the PCIe endpoint transfers data to these memory blocks.

Use the following step by step register sequence:

Step	W/R*	Register location	Value (Hex)	Description
0	R	BAR0+0x0	0x08021600	Endianess check by reading the version register. The format is DD_MM_YY_00. If 00_YY_MM_DD is read the endianess is swapped. In this case the value column hast to be swapped.
1	W	BAR1+0x4	Physical startaddress of mem block #1 32 LSBs	Important : The reserved memory block #1 must be below the 4 GByte Boundary (32-Bit Addressing). 64-Bit addresses are currently not supported for the performance test. This block is used for reading data.
2	W	BAR1+0x8	Physical startaddress of mem block #2 32 LSBs	Important : The reserved memory block #2 must be below the 4 GByte Boundary (32-Bit Addressing). 64-Bit addresses are currently not supported for the performance test. This block is used for writing data.
3	W	BAR0 + 0x8	MSBs of physical startaddress	In case the reserved memory block is above the 4 GB boundary, the 32 MSBs can be written in here: 31:0 : 32 MSBs for mem block #1 63:32 : 32 MSbs for mem block #2 Important: the 64-Bits must be written with a 64 Bit Write (2 32-Bit Writes are not allowed)
4	W	BAR1+0xC	0x2000	Bits 15:0 contain the blocklength, typical values are 0x80 up to 0x2000 Bits 31:16 contain the number of blocks that will be written (ascending) to the DMA Buffer. The block number is DMA Buffer length divided by the blocklength but limitet to 0xFFFF.
5	W	BAR2+x20		Blockamount** : Set to 0xA343C * Link Multiplier
6	W	BAR2+0x2C	0x0000_0C35	50 us Stalltime Threshold (Unit is 16 ns)
7	W	BAR0+0x10		Start Preformance Measurement: 0x0000_0001 : Only Read 0x0000_0002 : Only Write 0x0000_0003 : Read & Write in parallel
8				Wait for 30 Seconds
9		RX Result : BAR2+1C TX Result : BAR2 + 18		Poll the resultregister for a value /= 0x0 In case of a rx test, poll Bar2+0x1C In case of a tx test, poll Bar2+0x18 In case of a parallel test, both registers must be /= 0x0 The result registers contain the transmission time value in 0.1 s units for the data amount. In order to calculate the throughput, use the formula (0x2000 * Blockamount Register) / Time. If the result is 0x0, the Test has either not been started or is still running.
10	R	BAR2+0x28		Max Stall Time In order to convert this value to us, use the following formula: (Read value / 62.5)
11	R	BAR0+0x3C		RX Biterrors during Test
12	R	BAR0+0x40		TX Biterrors during Test

*W = Write, R=Read

** The Blockamount has to be calculated with the given formula in order to run the test for approximately 30 seconds. Use the values out of the table below:

Link Speed	Link Speed Multiplier	Link Width Multiplier	Example	Dataamount for rx/tx
2.5 GBit	1	1-8	Example link multiplier in case of X4 is 1*4=4	For the example the data amount is 0x2000 * 0xA343C * 4
5.0 GBit	2	1-8	Example link multiplier in case of X2 is 2*2=4	For the example the data amount is 0x2000 * 0xA343C * 4
8.0 GBit	4	1-16	Example link multiplier in case of X16 is 4*16=64	For the example the data amount is 0x2000 * 0xA343C * 64
16.0 GBit	8	1-8	Example link multiplier in case of X8 is 4*8=32	For the example the data amount is 0x2000 * 0xA343C * 32
32.0 Gbit	16	1-4	Example link multiplier in case of X4 is 16*4=64	For the example the data amount is 0x2000 * 0xA343C * 64

Optionally it is possible to poll the following register to see how the stall time over time:

	Register location	Description
R	BAR2+0x34	Stall time in 16 ms since the last read of this location