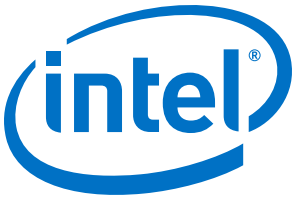


Networking Interface for Open Programmable Acceleration Engine

Intel FPGA Programmable Acceleration Card D5005

Updated for Intel® Acceleration Stack for Intel® Xeon® CPU with FPGAs: **2.0.1**

Subscribe
Send Feedback

UG-20211 | 2019.11.04
Latest document on the web: **PDF** | **HTML**

Contents

1. About this Document

This document describes how to integrate the network port feature into an Accelerator Functional Unit (AFU) and provision it from the host using the Open Programmable Acceleration Engine (OPAE) driver and tools. This document provides information about:

- Partial Reconfiguration High Speed Serial Interface (HSSI)
- Intel® Stratix® 10 Native PHY IP core parameters
- Tuning analog settings

1.1. Conventions

Table 1. Document Conventions

Convention	Description
#	If this symbol precedes a command, enter the command as a root.
$	If this symbol precedes a command, enter the command as a user.
`This font`	Indicates file names, commands, and keywords. The font also indicates long command lines. For long command lines, press Enter only if the next line starts a new command, where the # or $ character denotes the start of the next command.
<variable_name>	Indicates placeholder text that you must replace with appropriate values. Do not include the angle brackets.

1.2. Acceleration Glossary

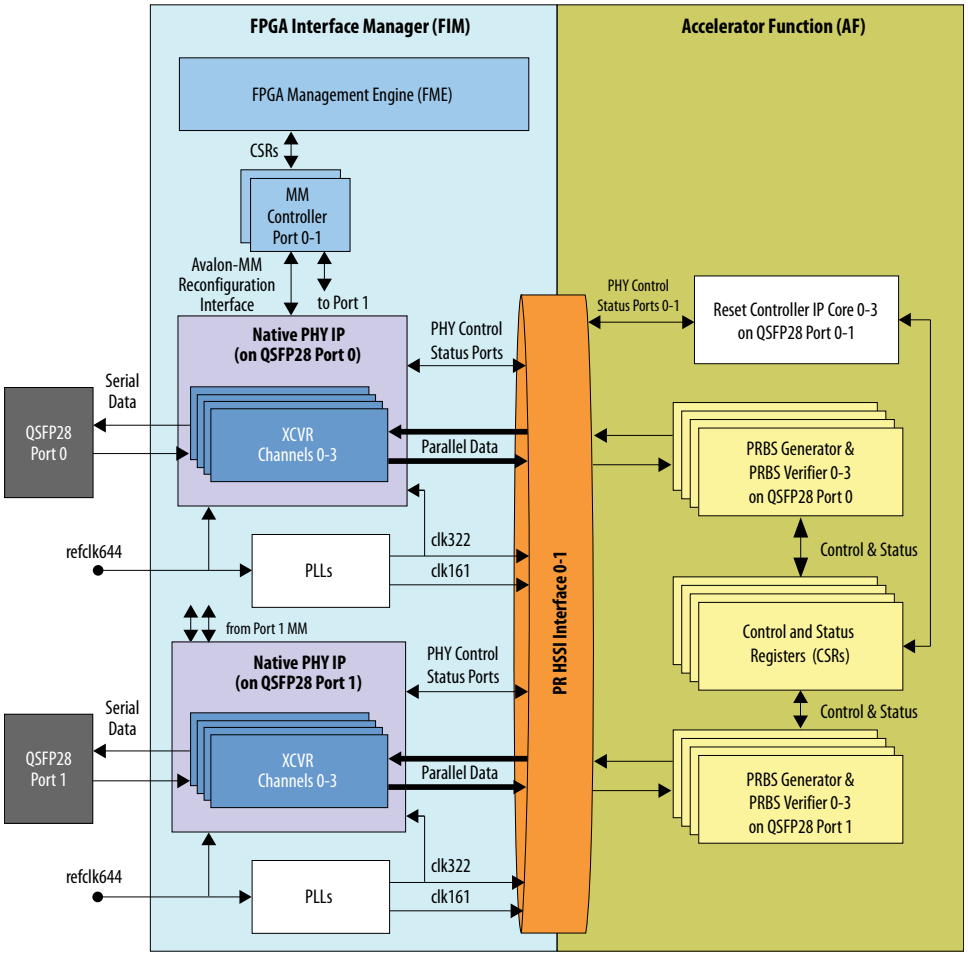
Table 2. Acceleration Stack for Intel Xeon® CPU with FPGAs Glossary

Term	Abbreviation	Description
Intel Acceleration Stack for Intel Xeon® CPU with FPGAs	Acceleration Stack	A collection of software, firmware and tools that provides performance-optimized connectivity between an Intel FPGA and an Intel Xeon processor.
Intel FPGA Programmable Acceleration Card	Intel FPGA PAC	The PCIe* accelerator card contains an FPGA Interface Manager (FIM) that pairs with an Intel Xeon processor over PCIe bus.
`OPAE_PLATFORM_ROOT`	-	A Linux shell environment variable set up during the process of installing the OPAE SDK delivered with the Acceleration Stack.
Quad Small Form Factor Pluggable 28	QSFP28	The Intel FPGA Programmable Acceleration Card D5005 has two QSFP28 cages on the I/O panel each of which supports up to 100G Ethernet. There are four TX/RX pairs per QSFP28.

2. Overview

The Intel FPGA Programmable Acceleration Card D5005 consists of two QSFP28 networking ports that can be configured for 4x10Gbps operation per port.

Figure 1. Block Diagram: Network Port Feature



The FPGA Interface Manager (FIM) instantiates two Intel Stratix 10 FPGA Transceiver Native PHY IP cores, one for each QSFP28 network port. The Native PHY IP cores are configured with four transceiver channels, enabling the Accelerator Function (AF) to instantiate an Accelerator Functional Unit (AFU) with up to 8x PRBS Generators and Verifiers, and 8x Reset Controller IP cores.

The Reset Controller IP core orchestrates analog and digital reset signaling for each transceiver channel, as required by the Intel Stratix 10 Native PHY IP core. In a real use case, along with a Reset Controller IP core, you will instantiate the 8x10G PCS and MAC IPs, as well as your user logic in the AF. The raw PHY parallel data interfaces are exposed to the Partial Reconfiguration (PR) boundary through the PR HSSI Interface. The raw PHY interface consists of 80-bit parallel data per transmit or receive direction in each transceiver, along with some sideband signals for handshaking with the Reset Controller IP core across the PR boundary.

The FIM also contains a set of PLLs for each network port. The PLLs provide all the necessary clocks for the transceivers and the AFU. The Memory-Mapped (MM) controllers instantiated in the FIM provide the ability for the software driver to have full access to the Avalon-MM Reconfiguration Interface of the Native PHY IPs through the FPGA Management Engine (FME) registers.

Note: The FIM contains only the Hard PHY in PCS-Direct mode (PMA-only). You implement your own PCS and MAC IP core in the AFU.

Table 3. Correspondence Between Acceleration Stack, FIM, and OPAE Versions

Acceleration Stack Version	FIM Version (PR Interface ID)	OPAE Version	BMC Firmware Version	BMC MAX10 Version
2.0.1	9346116d-a52d-5ca8-b06a-a9a389ef7c8d	1.1.4-8	2.0.12	2.0.6
2.0.1 Beta	8db6b54c-930e-5976-a03b-09f3c913aa95	1.1.4-8	2.0.10	2.0.4
2.0	bfac4d85-1ee8-56fe-8c95-865ce1bbaa2d	1.1.4-3	1.0.12	1.0.15

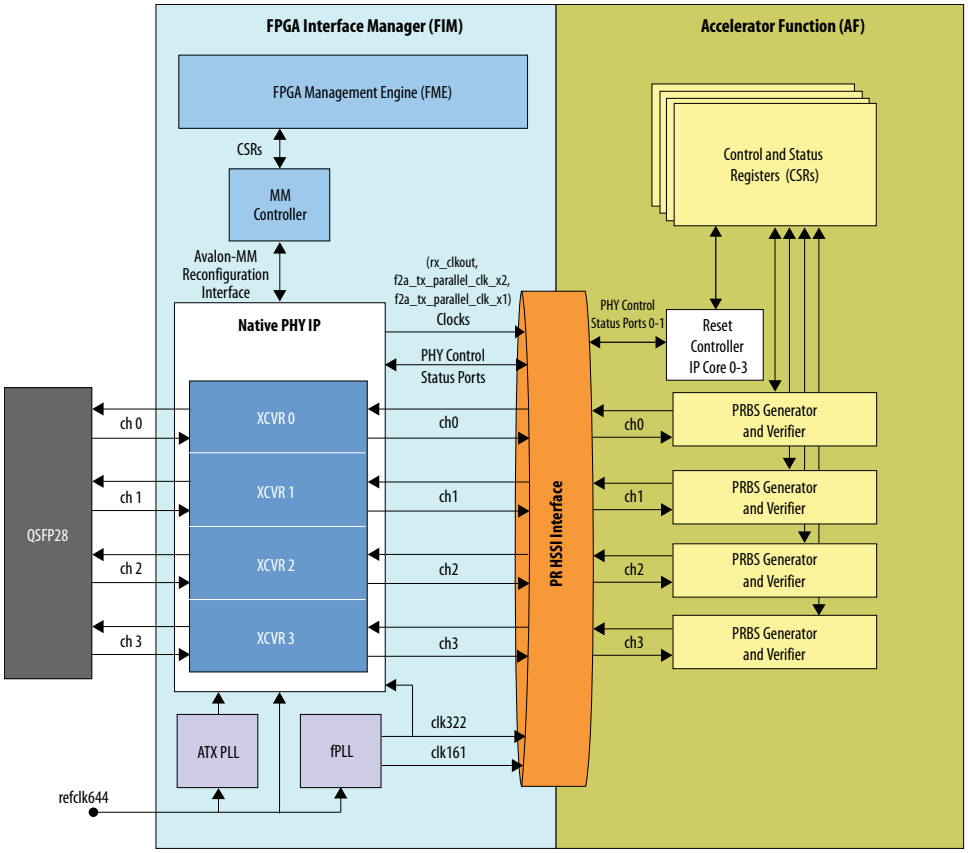
Related Information

Intel Stratix 10 L- and H-Tile Transceiver PHY User Guide (PDF)

2.1. Logical View

The FIM instantiates two PLLs that use a 644.53125MHz external reference clock to generate the necessary clocks for the Native PHY IP core and the AFU. The ATX PLL generates the high-speed serial clock for the Native PHY IP core. The fPLL generates two clocks, 322.265625MHz and 161.1328125MHz. Both of the fPLL clocks and RX clocks from the Native PHY IP core are provided to the AFU through the PR HSSI interface.

Figure 2. Logical View of the HSSI PHY



2.2. Physical View

This section depicts the hardware view of a single transceiver channel and its sub-components as part of the Native PHY IP core. The Native PHY IP core is optimized for the lowest roundtrip latency. The Native PHY IP TX/RX PCS-Core Interface FIFOs are configured as following:

- Both QSFP28 Port-0 and Port-1 TX FIFOs are in Phase Compensation mode such that TX clocks can be shared across all 4 channels per QSFP28 interface.
- QSFP28 Port-0 RX FIFO is in Phase Compensation mode.
- QSFP28 Port-1 RX FIFO is in Register mode (bypassed).

Note: In the following figures, the PCS, MAC, and User Logic blocks under AF are shown for illustration. These blocks are not provided by Intel as part of the AFU. Intel only provides an example AFU with 8xPRBS Generators and Verifiers.

Figure 3. Physical View with QSFP28 Port-0

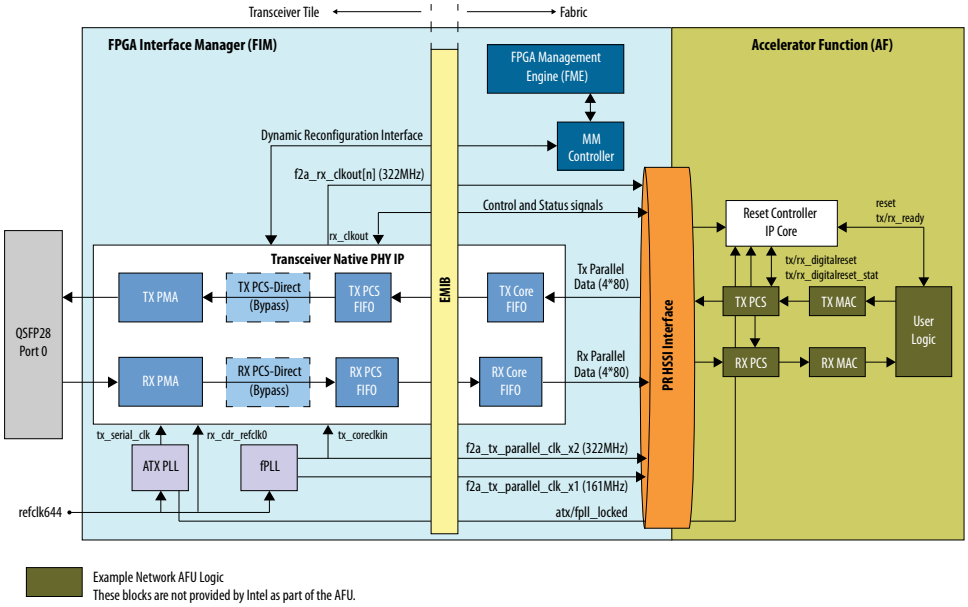
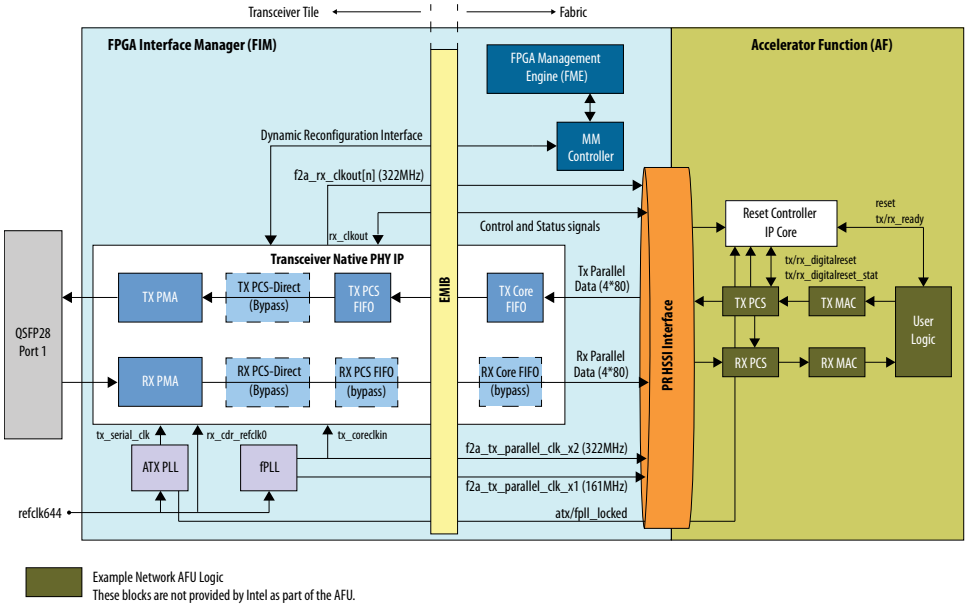


Figure 4. Physical View with QSFP28 Port-1



2.3. Clock Architecture

This section describes the clocking architecture of the Native PHY IP core.

All four channels on the TX parallel data interface are clocked by `f2a_tx_parallel_clk_x2` clock, per QSFP28 interface. Each one of the four channels on the RX parallel data interface is clocked by its own corresponding `f2a_rx_clkout[n]` clock, per QSFP28 interface.

On both the QSFP28 ports, `tx_clkout[n]` interfaces of the Native PHY IP core have no connection (NC) because the TX FIFO is in Phase Compensation mode and the `f2a_tx_parallel_clk_x2` clock is used to drive the `tx_coreclkin[n]` interfaces.

On the QSFP28 Port-0, `rx_coreclkin[n]` interfaces of the Native PHY IP core are connected to `rx_clkout[n]` interfaces because the RX FIFO is in Phase Compensation mode.

On the QSFP28 Port-1, `rx_coreclkin[n]` interfaces of the Native PHY IP core are connected to ground because the RX FIFO is in Register mode.

Figure 5. Clocking Architecture with QSFP28 Port-0

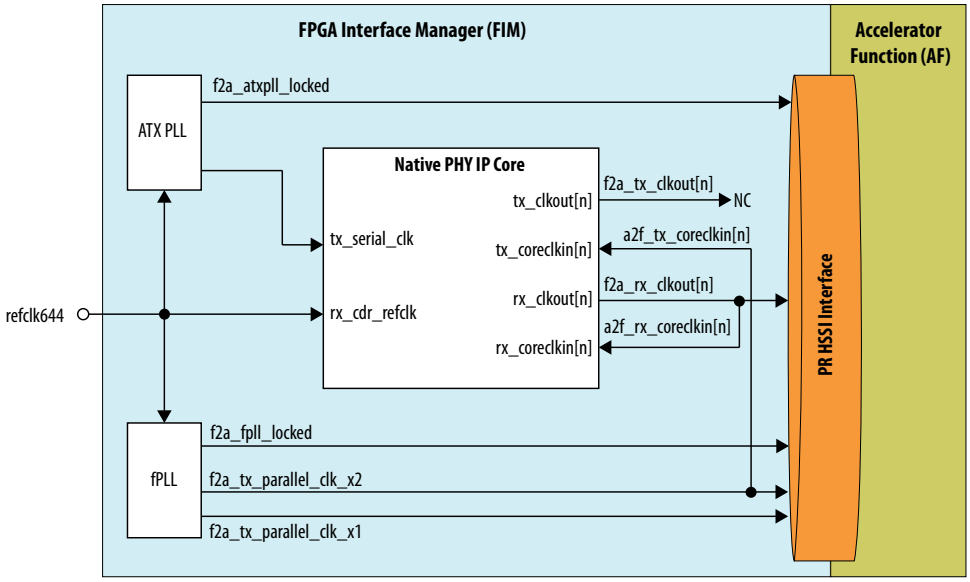


Figure 6. Clocking Architecture with QSFP28 Port-1

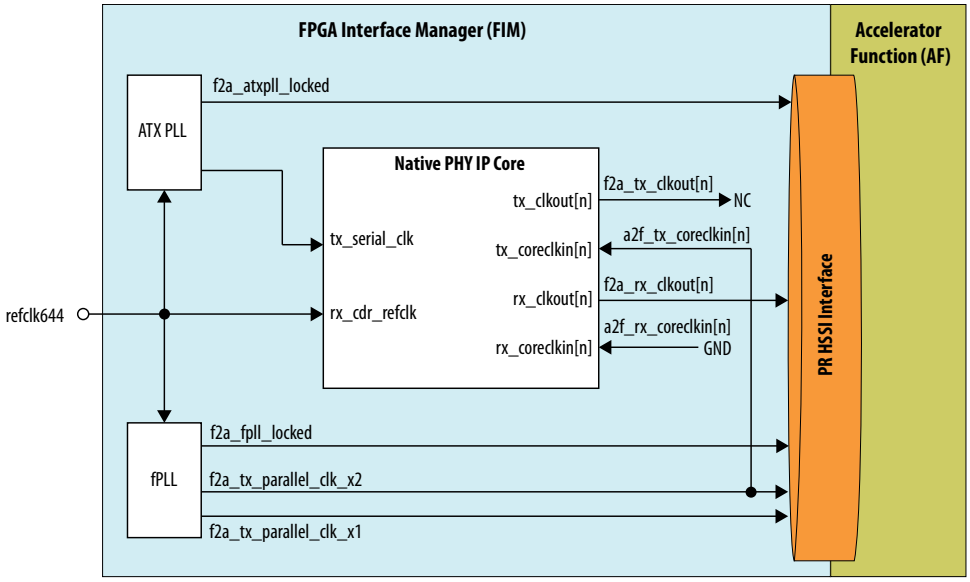


Table 4. Clock Frequencies

Clock Name	Frequency in MHz
refclk644	644.53125
rx_cdr_refclk	644.53125
tx_serial_clk	5156.25
f2a_tx_parallel_clk_x1	161.1328125

continued...

Clock Name	Frequency in MHz
f2a_tx_parallel_clk_x2	322.265625
tx_clkout[n]	322.265625
tx_coreclkin[n]	322.265625
rx_clkout[n]	322.265625
rx_coreclkin[n]	322.265625

You can access the following clocks from AF:

- `rx_clkout[n]`
- `f2a_tx_parallel_clk_x1`
- `f2a_tx_parallel_clk_x2`

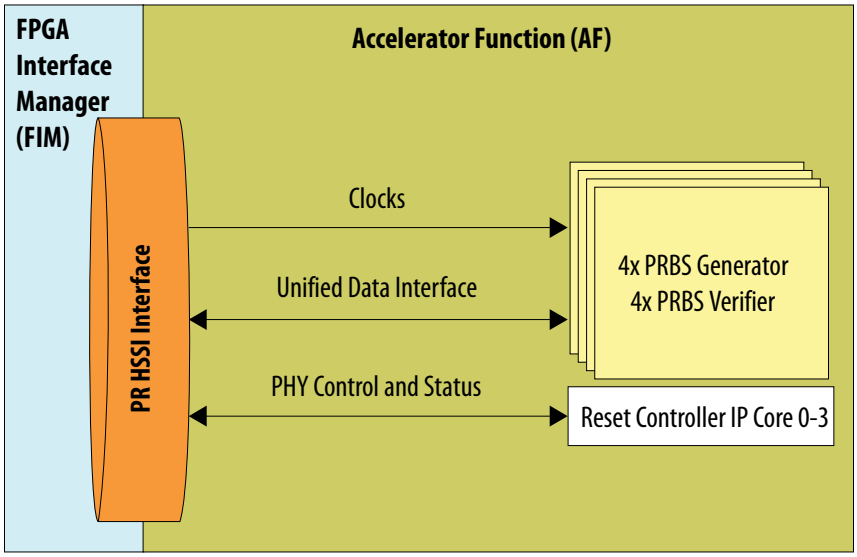
Clock Relationship

- The `refclk644`, external reference clocks , come from different sources for each QSFP28 network port. Therefore, the relationship between any given clock on network port 0 is asynchronous to any given clock on network port 1.
- The `f2a_tx_parallel_clk_x1` and `f2a_tx_parallel_clk_x2` are phase synchronous for a given QSFP28 network port.
- The `rx_clkout[n]` clocks are recovered by the Clock and Data Recovery (CDR) unit in the receiver of each channel. All the `rx_clkout[n]` clocks are asynchronous to one another.

3. Partial Reconfiguration HSSI Interface

The Partial Reconfiguration (PR) HSSI interface is a unified data interface that connects a network port to the PRBS Generators and Verifiers. The unified data interface consists of a fixed set of physical ports that are mapped to specific signaling functions. The PR HSSI interface also provides clocks for synchronization as well as control and status signals for analog and digital reset sequence orchestration between the PHY in FIM and the reset controller IP core in AF. The figure below provides a high-level block diagram for one QSFP instance.

Figure 7. PR HSSI Block Diagram



3.1. Clock Signals

The clocks of the PR HSSI Interface synchronize the unified data interface between the PRBS Generators and Verifiers, and the HSSI PHY. The signal directions listed for HSSI ports are from the perspective of the FIM. The signals listed below are identical for both QSFP28 interfaces.

Table 5. Clock Signals

Port Name	Width	Direction	Description
f2a_tx_parallel_clk_x1	1	Output	A 161.1328125 MHz clock generated by an fPLL in the HSSI PHY from a 644.53125 MHz QSFP28 external reference clock. This clock is intended to drive the user logic in the AF.
f2a_tx_parallel_clk_x2	1	Output	A 322.265625 MHz clock generated by an fPLL in the HSSI PHY from a 644.53125 MHz QSFP28 external reference clock. This clock drives the tx_coreclkin inputs of all 4 channels of the Native PHY IP core. All transmit data from AFU to HSSI PHY should be synchronous to f2a_tx_parallel_clk_x2.
f2a_rx_clkout	4	Output	A 322.265625 MHz clock at the output of the Native PHY IP core rx_clkout[n] interface. All receive data to the PRBS Verifiers from the HSSI PHY is synchronous to f2a_rx_clkout[n], per transceiver channel n.

3.2. Data Interface and Signals

The HSSI unified data interface conforms to the Intel Stratix 10 FPGA Transceiver Native PHY IP core configured in 32-bit PCS-Direct mode. It consists of generic parallel data and encoding control interfaces for transmit and receive that are mapped to specific signaling behavior as outlined in the *Intel Stratix 10 L- and H-Tile Transceiver PHY User Guide*. The unified data interface also includes flow control ports to manage passing data to and from the HSSI PHY interface.

The table below provides a cross reference from the hssi:raw_pr unified data interface signals to the Intel Stratix 10 FPGA Transceiver Native PHY IP core with enhanced PCS signal set. The HSSI PHY IP is configured in Configuration-32, PMA width-32, FPGA Fabric width-32. The TX Core FIFO is configured in Phase Compensation mode. The RX Core FIFO QSFP0 is configured in Phase Compensation mode and RX Core FIFO QSFP1 is configured in Register mode. The Simplified Data Interface is disabled. The Double-Rate Transfer is disabled. For detailed information on these signals, refer to the *Intel Stratix 10 L- and H-Tile Transceiver PHY User Guide*.

Table 6. Data Signals

Port Name	Width	Direction	Clock Domain	Native PHY IP Port Name	Reference
Transmit and Receive Data and Encoding Control Ports					
a2f_tx_parallel_data	4*80	Input	f2a_tx_parallel_clk_x2	tx_parallel_data	PCS-Core Interface Ports: PCS-Direct
f2a_rx_parallel_data	4*80	Output	f2a_rx_clkout[n]	rx_parallel_data	
Flow Control Ports					
f2a_tx_fifo_empty	4	Output	Reserved		
f2a_tx_fifo_full	4	Output	Reserved		
f2a_tx_fifo_pempty	4	Output	Reserved		
f2a_tx_fifo_pfull	4	Output	Reserved		

continued...

Port Name	Width	Direction	Clock Domain	Native PHY IP Port Name	Reference
a2f_rx_bitslip	4	Input	Reserved		
f2a_rx_fifo_empty	4	Output	Reserved		
f2a_rx_fifo_full	4	Output	Reserved		
f2a_rx_fifo_pempty	4	Output	Reserved		
f2a_rx_fifo_pfull	4	Output	Reserved		
a2f_rx_fifo_rd_en	4	Input	Reserved		

Related Information

Intel Stratix 10 L- and H-Tile Transceiver PHY User Guide (PDF)

3.3. Control and Status Signals

The PR HSSI Interface provides signals for HSSI PHY PCS status and transceiver loopback control. The signal behavior conforms to the Intel Stratix 10 FPGA Transceiver Native PHY IP core in 32-bit PCS-Direct mode. The below table cross references the HSSI port names to the Native PHY IP port names.

Table 7. Control and Status Signals

hssi Port Name	Width	Direction	Clock Domain	Native PHY IP Core Port Name	Reference
f2a_tx_ready	4	Output	Reserved		
f2a_rx_ready	4	Output	Reserved		
a2f_rx_seriallpbken	4	Input	Asynchronous	rx_seriallpbken	*Table: RX PMA Ports-PMA QPI Options* in PMA, Calibration, and Reset Ports
f2a_atxpll_locked	1	Output	Asynchronous	-	-
f2a_fpll_locked	1	Output	Asynchronous	-	-
f2a_tx_cal_busy	4	Output	Asynchronous	tx_cal_busy	*Table: User-coded Reset Controller, Transceiver PHY, and TX PLL Signals* in User-Coded Reset Controller Signals
f2a_rx_cal_busy	4	Output	Asynchronous	rx_cal_busy	*Table: User-coded Reset Controller, Transceiver PHY, and TX PLL Signals* in User-Coded Reset Controller Signals
f2a_rx_is_lockedtodata	4	Output	Synchronous to CDR	rx_is_lockedtodata	
f2a_rx_is_lockedtoref	4	Output	f2a_rx_clkout[n]	rx_is_lockedtoref	*Table: RX PMA Ports* in PMA, Calibration, and Reset Ports
a2f_tx_analogreset	4	Input	Synchronous to Reset Controller IP input clock (recommended 100-125MHz)	tx_analogreset	*Table: User-coded Reset Controller, Transceiver PHY, and TX PLL Signals* in User-Coded Reset Controller Signals

continued...

hssi Port Name	Width	Direction	Clock Domain	Native PHY IP Core Port Name	Reference
a2f_rx_analogreset	4	Input	Synchronous to the Reset Controller IP core input clock (recommended 100-125MHz)	rx_analogreset	*Table: User-coded Reset Controller, Transceiver PHY, and TX PLL Signals* in User-Coded Reset Controller Signals
f2a_tx_analogreset_stat	4	Output	Asynchronous	tx_analogreset_stat	
f2a_rx_analogreset_stat	4	Output	Asynchronous	rx_analogreset_stat	
a2f_tx_digitalreset	4	Input	Synchronous to the Reset Controller IP core input clock (recommended 100-125MHz)	tx_digitalreset	
a2f_rx_digitalreset	4	Input	Synchronous to the Reset Controller IP core input clock (recommended 100-125MHz)	rx_digitalreset	
f2a_tx_digitalreset_stat	4	Output	Asynchronous	tx_digitalreset_stat	
f2a_rx_digitalreset_stat	4	Output	Asynchronous	rx_digitalreset_stat	

3.4. Connecting the PCS to the HSSI Interface

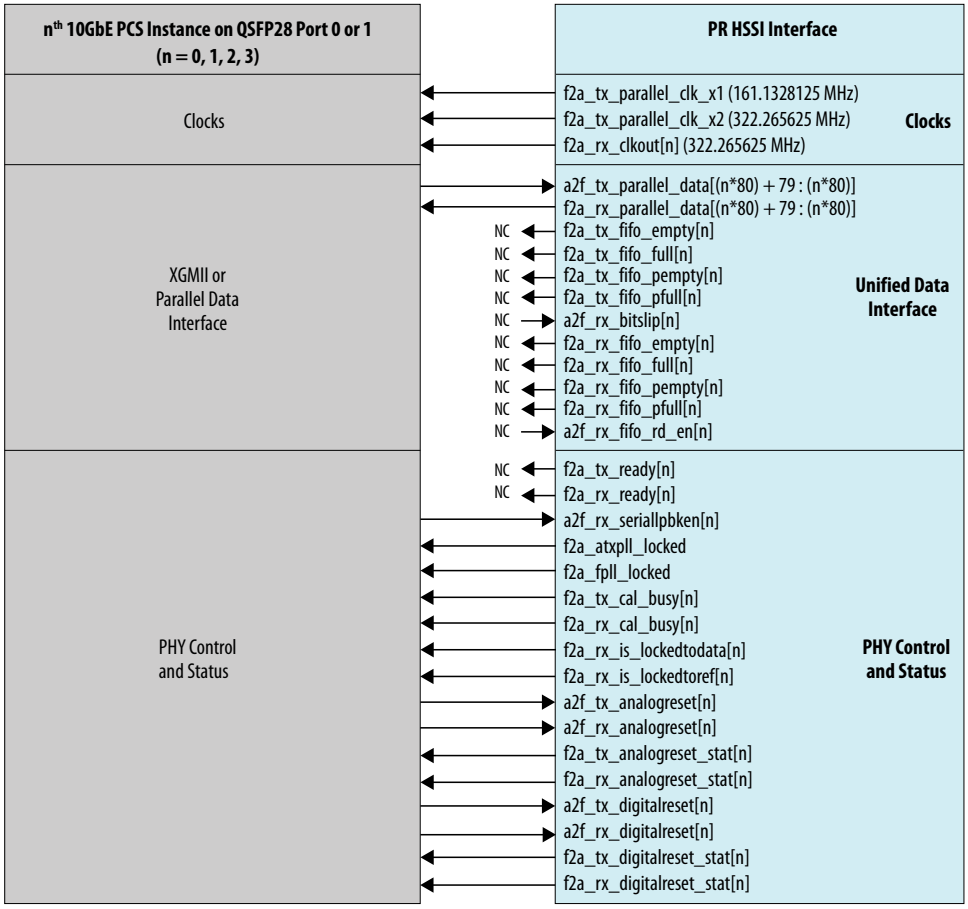
In 32-bit PCS-Direct mode, the interface between the PCS and HSSI PHY maps as following:

Table 8. Interface Mapping

TX Port Function	TX Port	RX Port Function	RX Port
Configuration-32, PMA Width-32, FPGA Fabric width-32			
data[31:0]	tx_parallel_data[31:0]	data[31:0]	rx_parallel_data[31:0]
tx_fifo_wr_en	tx_parallel_data[79]	rx_prbs_err	rx_parallel_data[35]
		rx_prbs_done	rx_parallel_data[36]
		rx_data_valid	rx_parallel_data[79]

Figure 8. Connecting the PCS to the HSSI Interface

This figure illustrates how to connect a 10GbE PCS to the HSSI PHY using the PR HSSI Interface.



4. Native PHY IP Core Parameters

During the FIM instantiation, the following IP parameters were selected for generating the PHY IP core. These parameter settings are informatory, you can not control or configure them. For more information about these parameters, refer to the *Intel Stratix 10 L- and H-Tile Transceiver PHY User Guide*.

Figure 9. Category: General

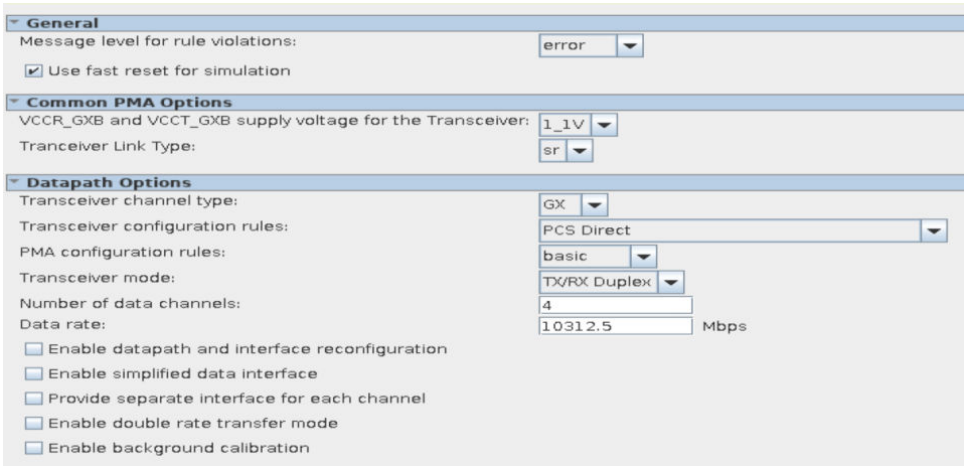


Figure 10. Category: TX PMA

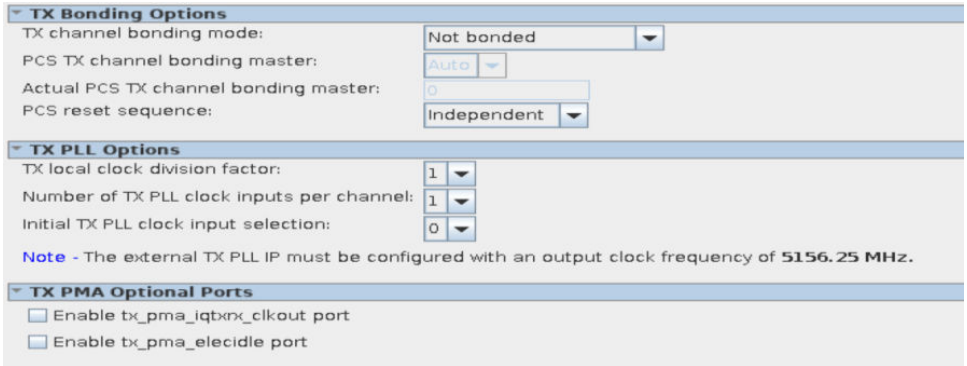


Figure 11. Category: RX PMA

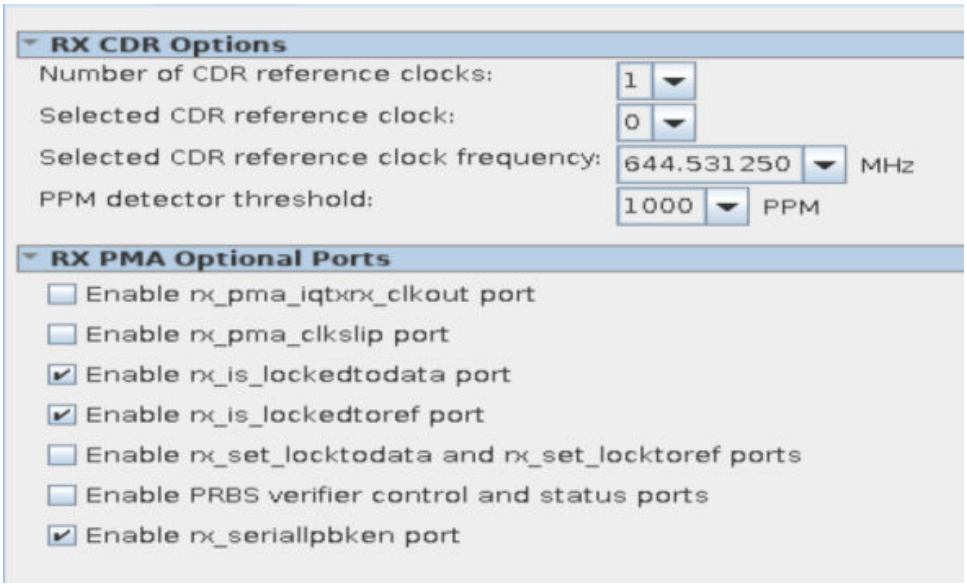


Figure 12. Category: PCS-Direct

Figure 13. Category: PCS Core Interface

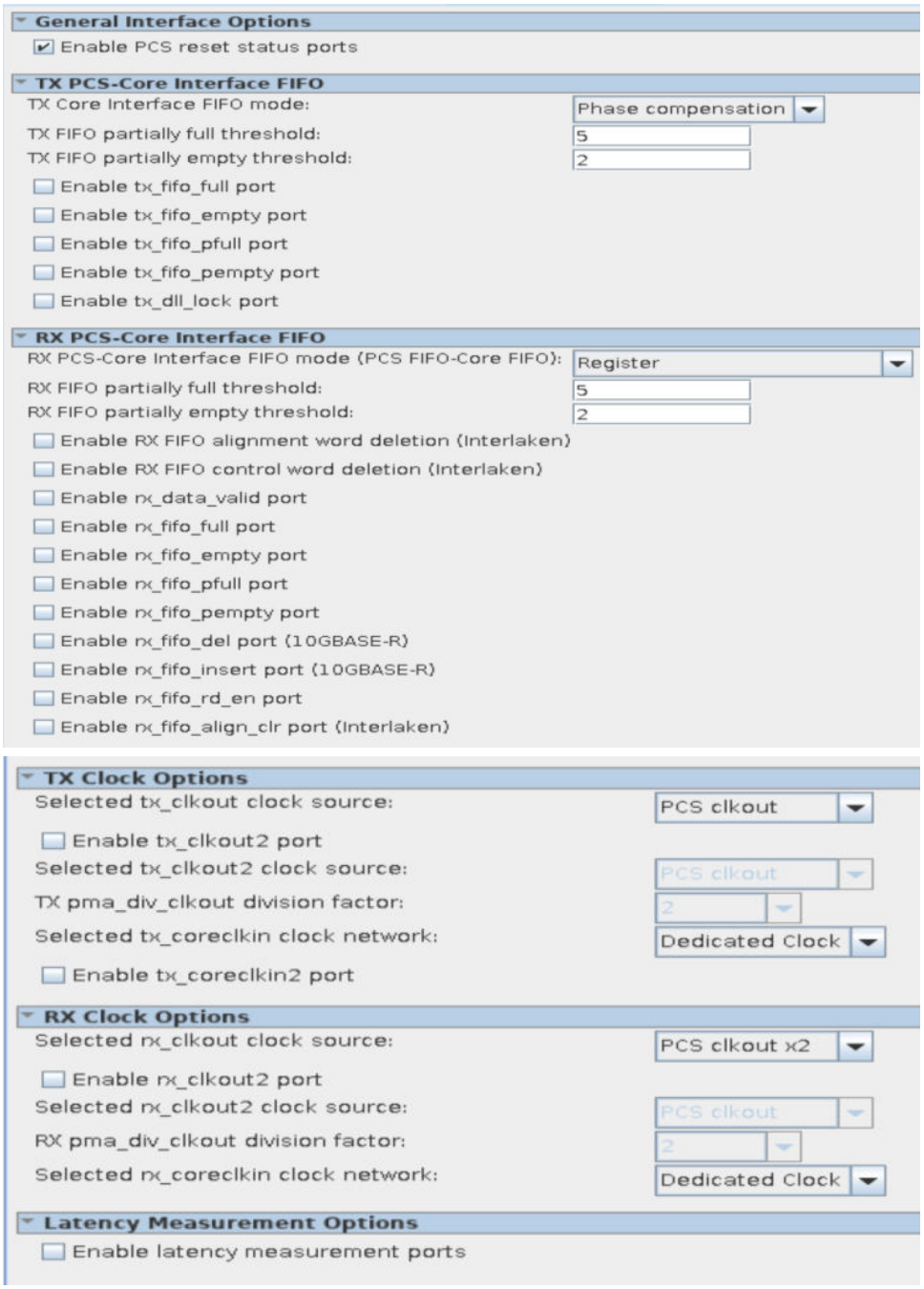


Figure 14. Category: Analog PMA Setting

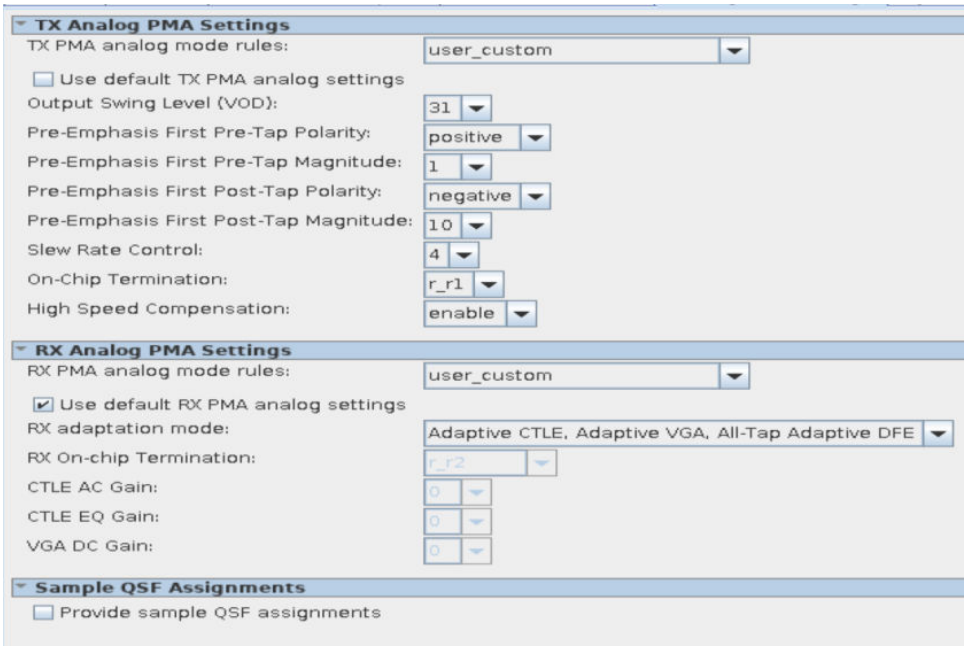
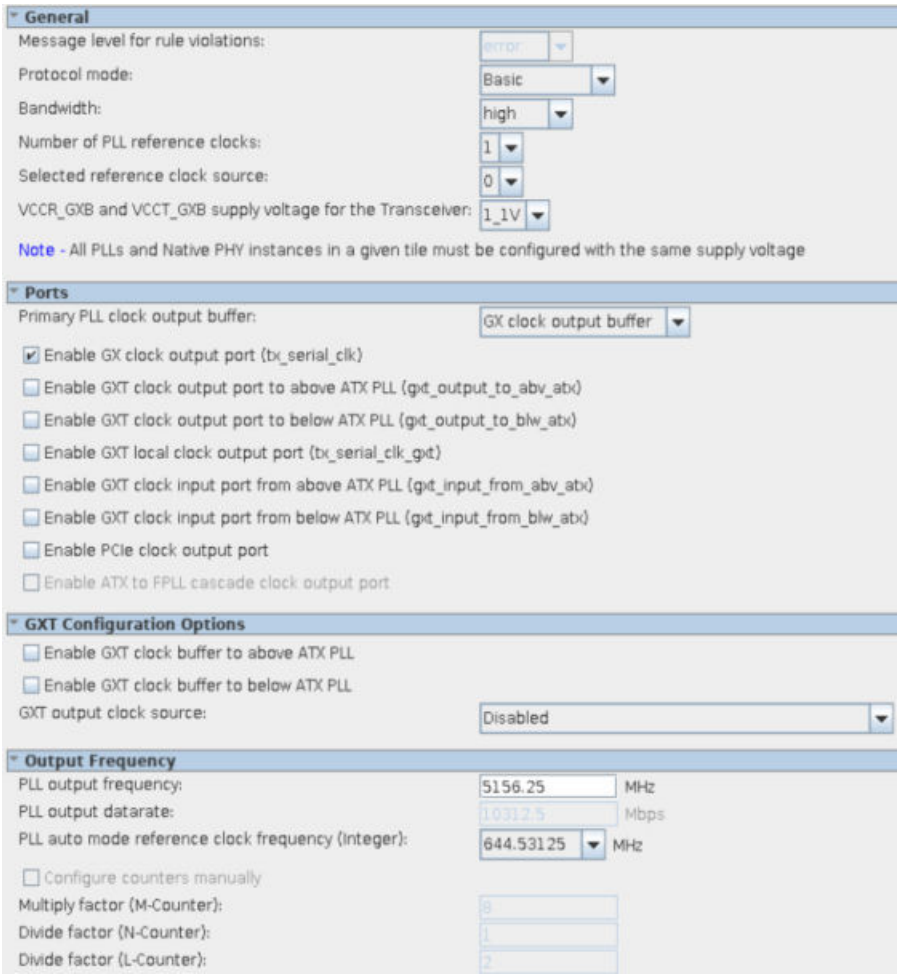
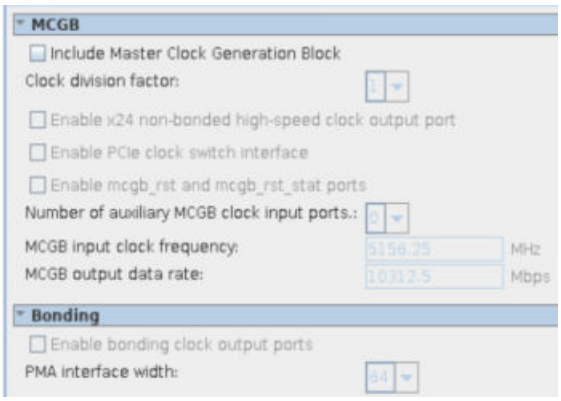


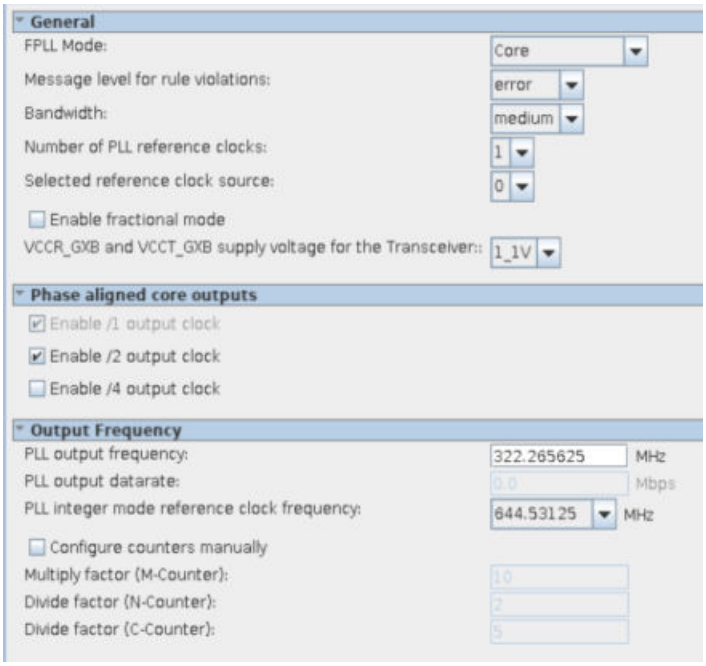
Figure 15. Category: ATX PLL IP Setting





Parameter Names	Parameter Values
datarate	10312.5 Mbps
K counter (valid in frac...	1
L cascade counter (val...	0
L cascade predivider/V...	select_vco_output
L counter (valid in non-...	2
M counter	8
N counter	1
PLL output frequency	5156.25 MHz
vco_freq	10312.5 MHz

Figure 16. Category: fPLL IP Setting



Parameter Names	Parameter Values	Parameter Units
L-Counter	1	
M-Counter	10	
N-Counter	2	
C-Counter	5	
K-Counter	0	
PLL output datarate	0.0	Mbps
PLL output frequency	322.265625	MHz
VCO frequency	6445.3125	MHz
PFD frequency	322.265625	MHz
C0 Counter Output Fr...	322265625	ps
C0 Counter Output Fr...	322265625	ps
C1 Counter Output Fr...	0	ps
C1 Counter Output Fr...	0	ps

Related Information

Intel Stratix 10 L- and H-Tile Transceiver PHY User Guide (PDF)

5. OPAE Support

The OPAE SDK includes the following support for the Intel FPGA PAC D5005 network port feature:

- OPAE kernel driver sysfs files enable configuration of the network port feature and allows access to related information on the Intel FPGA PAC D5005 from the host.
  - 128-bit UUID
  - HSSI PHY PMA analog settings

5.1. Supported Settings

The OPAE driver is capable of changing the following transmitter settings per transceiver channel:

- Pre-emphasis
- TX Output Differential Swing (VOD)
- TX Compensation

For more information, refer to the Transmitter PMA Logical Register Map.

5.2. Unsupported Settings

The OPAE driver is not capable of changing the following settings:

- Transmitter Slew Rate
- PMA Receiver Settings (VGA, CTLE, DFE, Adaptation Modes)

5.3. Tuning Information

Before you proceed further, you must install and load the OPAE driver and tools. For more information, refer to the Intel Acceleration Stack Quick Start Guide: Intel FPGA ProgrammableAcceleration Card D5005.

sysfs Tree

Sysfs entries allow reading or writing to HSSI configuration and status registers (CSR):

```
/sys/class/fpga/intel-fpga-dev.0/intel-fpga-fme.0/intel-pac-hssi.<1/2>.auto/
```

The HSSI `sysfs tree` is as follows:

- qsfp<0/1>
  - ctrl HSSI_CTRL_QSFP<0/1> CSR: allows access to control registers
  - stat HSSI_STAT_QSFP<0/1> CSR: allows access to status registers
  - chan<0/1/2/3>: analog settings of each of the 4 transceiver channels per QSFP interface
    - tx_post_tap: Pre-emphasis 1st post-tap magnitude and polarity
    - tx_pre_tap: Pre-emphasis 1st pre-tap magnitude and polarity
    - tx_vod: TX output differential swing
    - tx_comp: TX Compensation

tx_post_tap

- Use tx_post_tap sysfs entry to tune the transmitter pre-emphasis 1st post-tap magnitude and polarity.
- Valid magnitude is between -24 and 24.

To evaluate the correct setting, refer to the Intel Stratix 10 H-tile Pre-Emphasis and Output Swing Estimator.

Example:

1. Change directory to the desired QSFP interface and channel:

```
$ cd /sys/class/fpga/intel-fpga-dev.0/intel-fpga-fme.0/intel-pac-
hssi.<1/2>.auto/qsfp<0/1>/chan<0/1/2/3>
```

2. Read current tx_post_tap setting:

```
$ cat tx_post_tap
```

   Output: 0

3. Write new tx_post_tap magnitude and polarity, assume it as magnitude of 1 with positive polarity:

```
$ sudo -- sh -c 'echo +1 > tx_post_tap'
```

4. Verify that tx_post_tap:

```
$ cat tx_post_tap
```

   Output: +1

tx_pre_tap

- Use tx_pre_tap sysfs entry to tune the transmitter pre-emphasis 1st pre-tap magnitude and polarity.
- Valid magnitude is between -15 and 15.

To evaluate the correct setting, refer to the Intel Stratix 10 H-tile Pre-Emphasis and Output Swing Estimator. Also, refer to the example under tx_post_tap.

tx_vod

- Use `tx_vod` sysfs entry to tune the transmitter output differential swing.
- Valid output swing level is between 17 (600 mV) and 31 (VCCT or Transmitter Power Supply Voltage)

To evaluate the correct setting, refer to the Intel Stratix 10 H-tile Pre-Emphasis and Output Swing Estimator.

Example:

1. Change directory to the desired QSFP interface and channel:

```
$ cd /sys/class/fpga/intel-fpga-dev.0/intel-fpga-fme.0/intel-pac-
hssi.<1/2>.auto/qsfp<0/1>/chan<0/1/2/3>
```

2. Read current `tx_vod` setting:

```
$ cat tx_vod
```

Output: 31

3. Write new `tx_vod` output, assume it as 29:

```
$ sudo -- sh -c 'echo 29 > tx_vod
```

4. Verify that `tx_vod`:

```
$ cat tx_vod
```

Output: 29

tx_comp

- Use `tx_comp` sysfs entry to tune the transmitter compensation, which helps reduce the PDN induced ISI jitter when enabled.
- Valid compensation value is either 0 (off) or 1 (on)

Example:

1. Change directory to the desired QSFP interface and channel:

```
$ cd /sys/class/fpga/intel-fpga-dev.0/intel-fpga-fme.0/intel-pac-
hssi.<1/2>.auto/qsfp<0/1>/chan<0/1/2/3>
```

2. Read current `tx_comp` setting:

```
$ cat tx_comp
```

Output: 1

3. TX compensation is currently enabled, let's turn it off:

```
$ sudo -- sh -c 'echo 0 > tx_comp
```

4. Verify that `tx_comp`:

```
$ cat tx_comp
```

Output: 0

Monitor dmesg for Errors

Example: Error in setting the transmitter output differential swing to 100

```
$ echo 100 > tx_vod
bash: echo: write error: Invalid argument

Check dmesg
$ dmesg
[ 7597.306591] intel-pac-hssi intel-pac-hssi.2.auto: Max VOD is 31
```

Example: Error in setting a legal tx_vod value

```
$ echo 31 > tx_vod
bash: echo: write error: Connection timed out

Check dmesg
$ dmesg
[ 7812.184357] intel-pac-hssi intel-pac-hssi.2.auto: timeout, HSSI ack not
received

Check if the channel is held in reset
 $ cat stat
0x000f000f000f000f

Deaasert the reset
$ echo 0x0 > ctrl
$ cat stat0xf3c0f3c0f3c0f3c0
```

6. Document Revision History for Networking Interface for OPAE

Document Version	Acceleration Stack Version	Changes
2019.11.04	2.0.1 (compatible with Intel Quartus® Prime Pro Edition 19.2)	Updated the entire document to reflect: • Addition of the PHY PCS-direct mode and `hssi_PRBS` AFU support. • Removal of the 10GbE MAC AFU support.
2019.08.05	2.0 (compatible with Intel Quartus Prime Pro Edition 18.1.2)	Initial release.