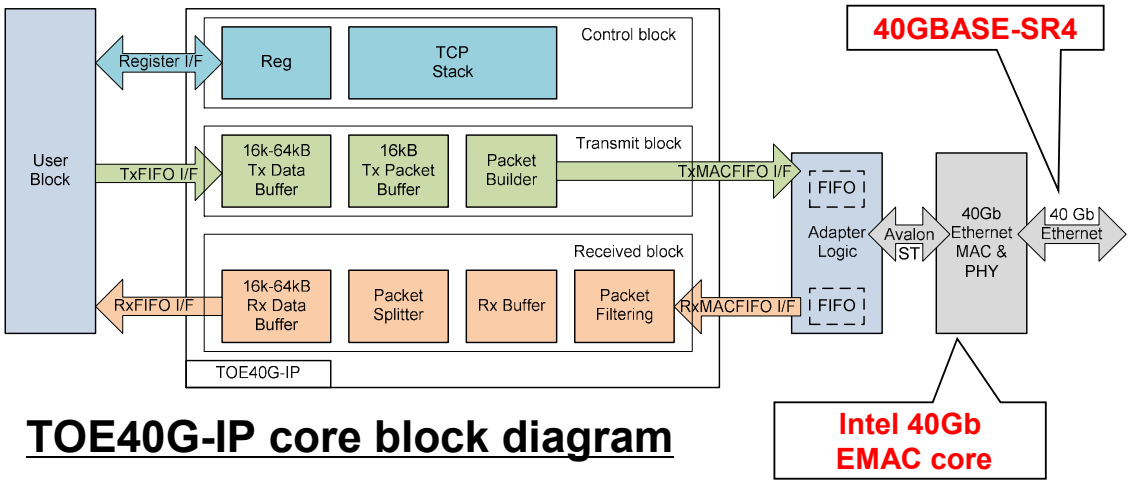
TOE40G-IP Introduction (Intel)

Ver1.0E

Realize 40GbE limit speed!

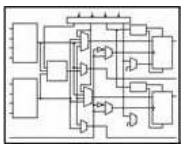
TOE40G-IP core Overview

- TCP/IP off-loading engine for 40GBASE-SR4
- Inserts between user logic and Intel 40Gb EMAC module
- Fully hard-wired TCP control for both Tx and Rx
- Supports Full Duplex communication



TOE40G-IP core block diagram

TOE40G-IP core Advantage 1



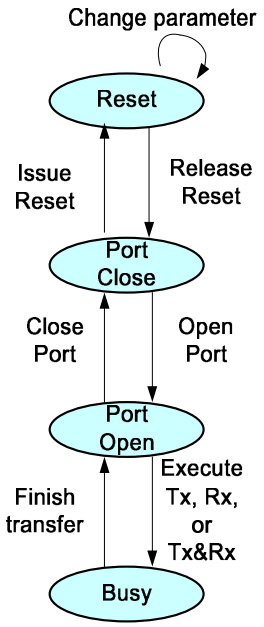
- Fully hard-wired TCP/IP protocol control
  - Possible to build CPU-less network system
  - Zero load for CPU
- Support all of Tx only, Rx only, and full-duplex
  - More than 4GByte/sec real transfer speed for half-duplex
  - Still 4GByte/sec real transfer speed for full-duplex
- Guarantee transfer data reliability
  - Tx: Automatic retry when No-ACK, Duplicate-ACK, timeout
  - Rx: Automatic ACK control by Sequence number calculation

TOE40G-IP core Advantage 2

- Selectable data buffer size
  - Selectable buffer size of memory usage vs. performance
- Compatible with Intel MAC core (Low latency 40G EMAC)
  - Real operation confirmed under 40GBASE-SR4 environment
- Many reference design on Intel evaluation board
  - Full Quartus project for standard Intel board
  - Free SOF-file for evaluation before purchase
  - All source code (except IP-core) in design project

TOE40G-IP core Operation

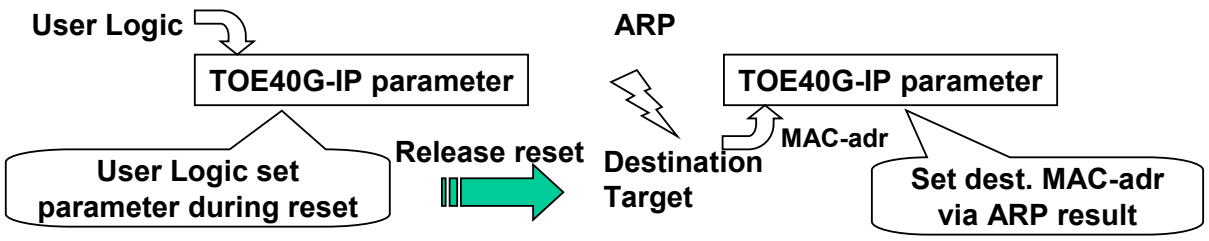
- Set parameter (IP-adr&MAC-adr, etc) during Reset
- Release Reset then initialize including ARP
- Idle state after initialization finish, wait command
- Port open by either of Active (Client) or Passive (Server) mode
- Tx and Rx operates individually (full-duplex)
- If want change parameter, move to Reset state (transfer/packet length can change except Busy)



State Diagram

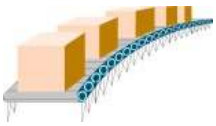
TOE40G-IP Initialization

- Set parameter to TOE40G-IP
  - User logic can set parameter during TOE40G-IP reset
  - Set IP address, MAC address, and Port number
  - Release reset after parameter setting finish
- TOE40G-IP executes ARP after reset release
  - Client mode: Issue ARP to the destination target
  - Server mode: Wait ARP from the destination target



High-Speed Tx

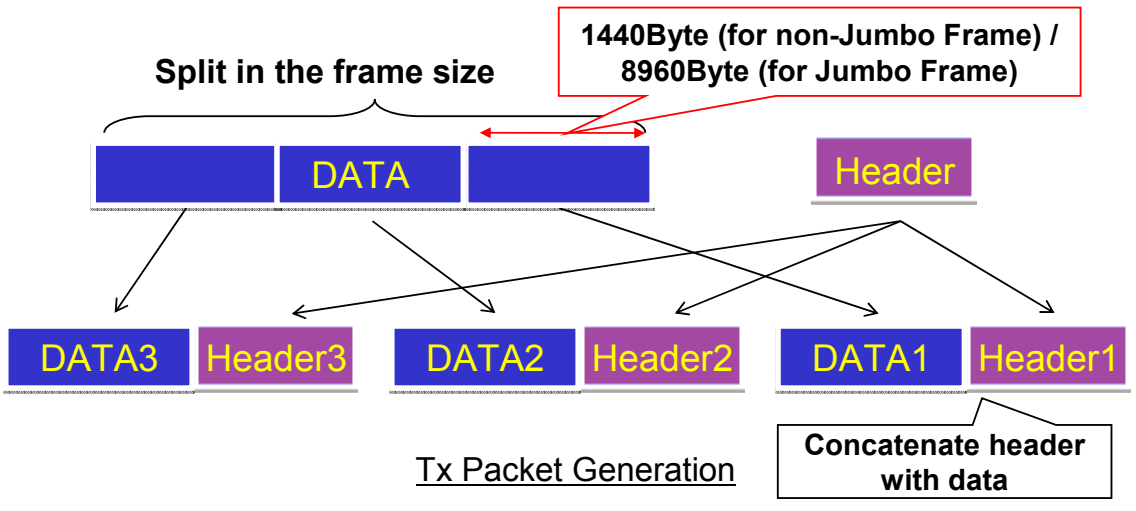
- **Tx packet generation**
  - User Logic writes Tx data to TxFIFO
  - Split Tx data in the frame size
  - Concatenate header with Tx data



- **Automatic retransmit function**
  - Check ACK reply from destination
  - Detect No-ACK, Duplicate-ACK, and Timeout
  - Resend same packet by such ACK error detection

Tx Packet Generation

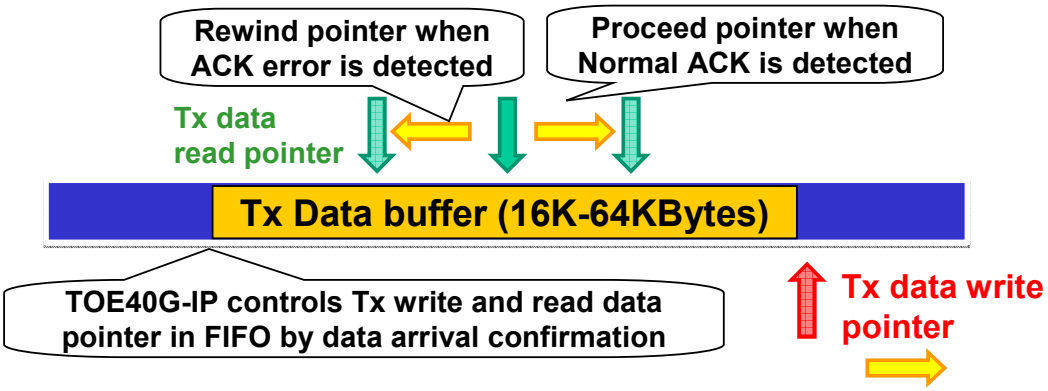
- **Generate header and concatenate it with Tx data**
  - TOE40G-IP splits Tx data in the frame size
  - Generate checksum and sequence number in TOE40G-IP



Tx Packet Generation

Automatic retransmit

- Retransmit function by dedicated FIFO
  - Proceed pointer by normal ACK reception
  - Rewind pointer by illegal ACK reception
  - TOE40G-IP controls pointer and retransmit operation



High-Speed Rx

- Rx packet header check
  - Ignore packet if destination is not TOE40G-IP or if checksum is wrong
- Data reordering
  - Reorder when sequence number skip is detected
  - Avoid retransmit request for transfer efficiency
  - If reordering is not possible, then send duplicate ACK
- Duplicate data management
  - Check duplicate data in Rx packet
  - Retrieve original data by trimming duplicate data part

Rx Packet Header Check

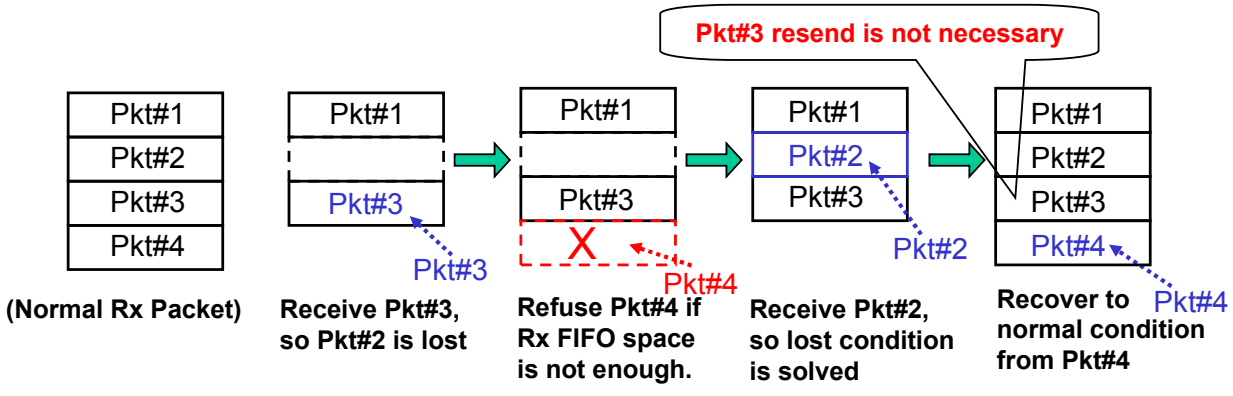
- **Verify header check sum in Rx packet**
  - **Also check following condition in TOE40G-IP**

Byte Offset	Protoco	Description	Check condition
0-5	ICMP	Destination MAC adr	Match with MAC adr set by SML/SMH register
6-11	ICMP	Source MAC adr	Match with target MAC adr set by ARP
12-13	ICMP	Type	= 0x0800 (IP packet)
14	IP	Version/Header	= 0x45 (IPv4, IP header len=20)
20	IP	Flag/Fragment OFS	= b"000000" (no fragment)
23	IP	Protocol Number	= 0x06(TCP packet)
26-29	IP	Source IP adr	Match with IP adr set by DIP register
30-33	IP	Destination IP adr	Match with IP adr set by SIP register
34-35	TCP	Source port number	Match with DPN register or extracted target port number in Passive Open
36-37	TCP	Destination port number	Match with port number set by SPN register
38-41	TCP	Sequence number	Possible value within TOE2-IP core can process this packet

Header check condition in Rx packet

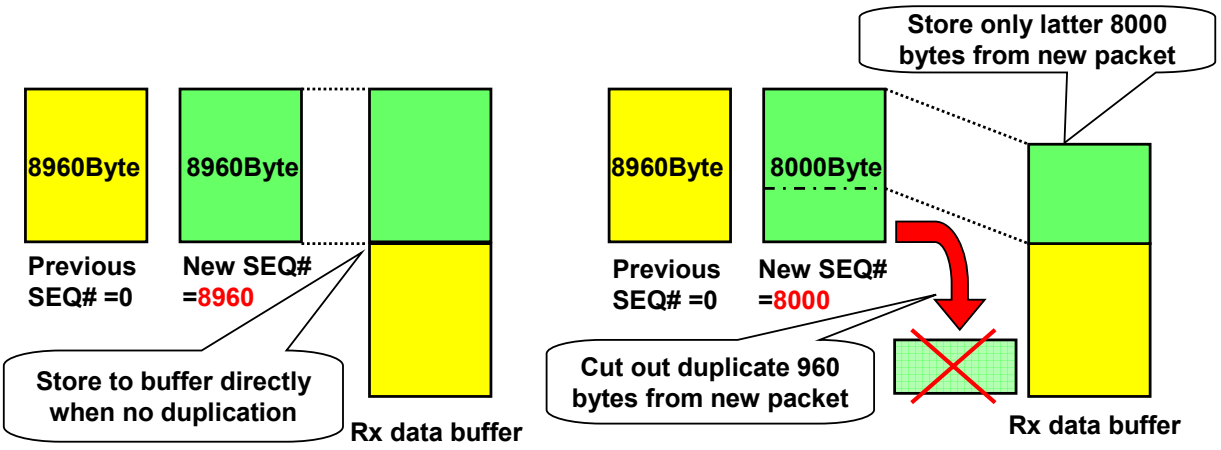
Data Reordering

- **Function when SEQ number skip is detected**
  - **Not accept any packet other than that can solve lost condition**
- **Data reordering function**
  - **Recover data contiguous from lost-solved packet**
  - **Keep performance by suppress resend request**



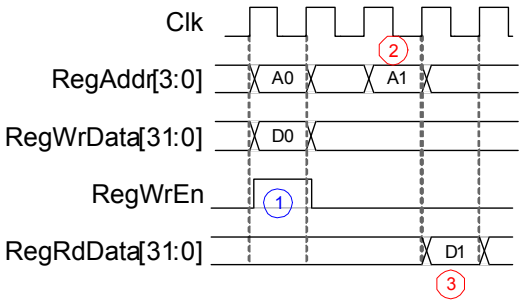
Duplicate data trimming

- **Detect data duplication and correct automatically**
  - **Detect Rx data duplication by checking sequence number**
  - **Trim duplicate block and retrieve contiguous data**



User Interface (Control)

- **3 types of Register I/F, Tx FIFO I/F, and Rx FIFO I/F**
  - **Register I/F for initial parameter setting and Tx/Rx command**
  - **Tx FIFO I/F and Rx FIFO I/F is standard FIFO interface**

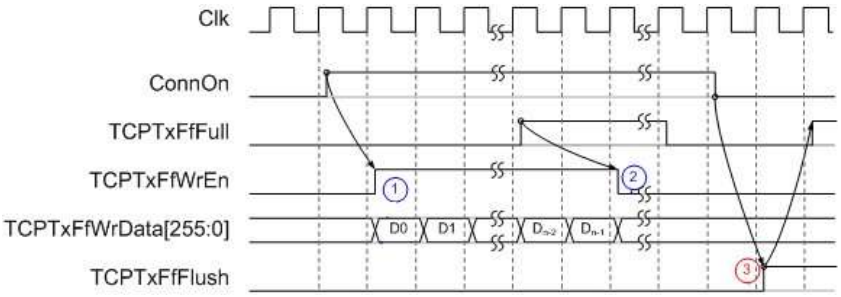


Register I/F timing

[Register Write]
(1) Assert RegWrEn with RegAddr and RegWrData

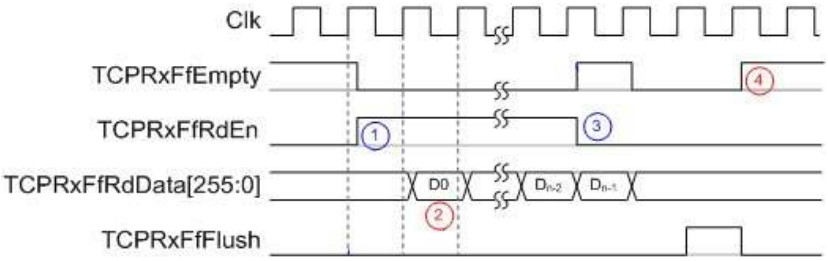
[Register Read]
(2) Set RegAddr
(3) Valid RegRdData output in the next clock

User Interface (Data)



[Tx data write]
(1) Write data with WrEn
(2) Suspend write within 4 Clk after Full assertion
(3) FIFO clear by Flush

Tx FIFO I/F timing



[Rx data read]
(1) Read by RdEn assertion when not Empty
(2) Read data after 1 Clk
(3) Read is inhibited when Empty
(4) FIFO clear by Flush

Rx FIFO I/F timing

Buffer Capacity

- **Parameterized 2 types of data buffer**
  - **(1) Tx Data Buffer: 16K/32K/64KBytes**
  - **(2) Rx Data Buffer: 16K/32K/64KBytes**
- **User can optimize resource usage and performance**

Generic Name	Range	Description
TxBufBitWidth	9-11	Set Tx data buffer size in address bit width When set to 9, size is 16KBytes, when 11, 64Kbytes for example.
RxBufBitWidth	9-11	Set Rx data buffer size in address bit width When set to 9, size is 16KBytes, when 11, 64KBytes for example.

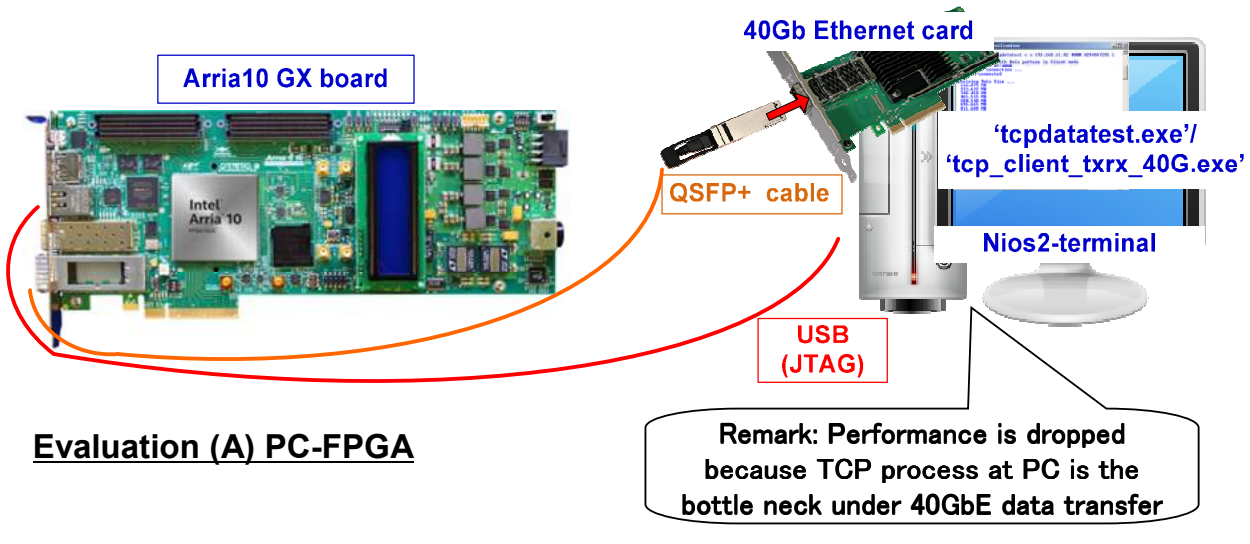
Buffer size is selectable by parameterization

2 types free SOF file for Evaluation

- 2 types of free SOF file for evaluation with Intel board
  - Communicate between (A)PC and FPGA, or (B)2 FPGA Bd.
  - For PC-FPGA connection, PC side become bottle neck (Real time TCP process by PC with 40Gbit/s is not practical)
- Measure transfer performance and data reliability
  - Supports both Half-Duplex and Full-Duplex mode
  - Data reliability check by real time data verification
- Bit file is free of charge
  - User can evaluate IP-core before purchase

Evaluation (A) PC-FPGA

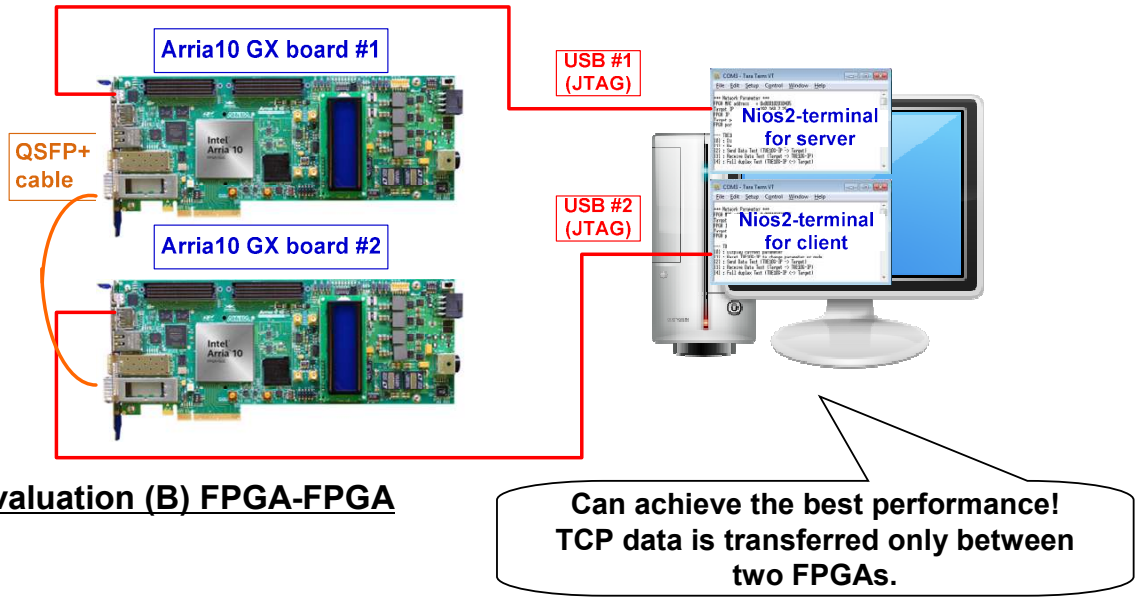
- Use of evaluation software on PC side
  - Note that it can't achieve the best performance due to the bottle neck at PC side



Evaluation (A) PC-FPGA

Evaluation (B) FPGA-FPGA

- **Communicate between two FPGA boards**
  - **External PC controls both two FPGAs via Nios2 console**

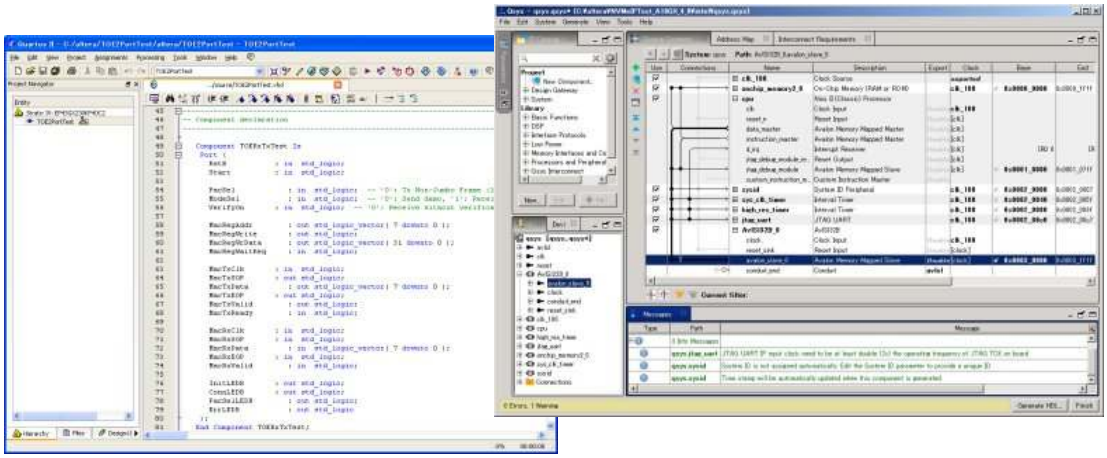


Evaluation (B) FPGA-FPGA

Can achieve the best performance! TCP data is transferred only between two FPGAs.

Reference Design Overview

- **Quartus design project for real operation**
  - **All source code (except IP-core) included in full project**
  - **Both half-duplex and full-duplex design in IP-core package**



Quartus/Qsys project in package

Effective Development on Ref. Design

- Quartus project is attached to the package
- Full source code (VHDL) except IP core
- Can save user system development duration
  - Confirm real board operation by original reference design.
  - Then modify a little to approach final user product
  - Check real operation in each modification step.

Short-term development is possible without big turn back

Resource Usage

- TOE40G-IP core standalone resource usage
  - Condition = Maximum buffer setting
    (TxDataBuf=RxDataBuf=64KB)

Family	Example Device	Fmax (MHz)	ALMs[1]	Registers[1]	Block Memory bit[2]	Design Tools
Arria 10 GX	10AX115S2F45I1SG	322	3,656	5,696	1,179,648	QuartusII18.0

TOE40G-IP core standalone compilation result

This result is based on maximum buffer size setting.
User can save memory resource by smaller buffer size setting

Performance (Half-Duplex)

```
+++ TOE40G-IP Send Mode +++
Enter transfer size (aligned to 256-bit): 32 - 0x1FFFFFFFE0 => 0x1FFFFFFFE0
Enter packet size (aligned to 256-bit)  : 32 - 8960         => 8960
Input mode                : [0] Client  [1] Server => 1
Wait Open connection ...
Connection opened
Start data sending
Send 4957 MByte Recv 0 Byte
Send 9913 MByte Recv 0 Byte
        |               |
Send 128880 MByte Recv 0 Byte
Send 133837 MByte Recv 0 Byte
Send data complete

Close connection
Connection closed

Total tx transfer size = 4294967295 (256-bit)
Total = 137.438[GB]
Time = 27726[ms] , Transfer speed = 4907[MB/s]
```

Sends 128GBytes data to another FPGA by Jumbo Frame (8960byte)

Transfer Performance = 4907MByte/sec!

Half-Duplex transfer result of two FPGAs communication

Performance (Full-Duplex)

```
+++ TOE40G-IP Full-duplex Mode +++
Enter total size (aligned to 256-bit) : 32 - 0x1FFFFFFFE0 => 0x1FFFFFFFE0
Enter packet size (aligned to 256-bit): 32 - 8960         => 8960
Enable data verification : [0] Disable [1] Enable => 1
Input mode               : [0] Client  [1] Server => 0
Open connection
Connection opened
Start data transferring
Send 4939 MByte Recv 4931 MByte
Send 9878 MByte Recv 9837 MByte
        |               |
Send 133355 MByte Recv 132505 MByte
Send 137438 MByte Recv 137419 MByte
Send data complete

Close connection
Connection closed

Total tx transfer size = 4294967295 (256-bit)
Total = 137.438[GB] , Time = 28003[ms] , Transfer speed = 4858[MB/s]

Total rx transfer size = 4294967295 (256-bit)
Total = 137.438[GB] , Time = 28003[ms] , Transfer speed = 4858[MB/s]
Input any key to stop the test
```

Transfer 128GBytes data for both Tx and Rx by Jumbo Frame (8960byte)

Transfer Performance = 4858MByte/sec!

Full-Duplex transfer result of two FPGAs communication

Conclusion

- High Speed data transfer via TOE40G-IP core
  - 4GByte/sec speed with guaranteed TCP protocol
- Fully hard-wired logic for complicated TCP process
  - Automatic re-send/re-order/duplicated data truncation
- Easy User I/F via register and FIFO
  - Simple enough to build CPU-less system
- Reference design with real operation available
  - Steady development with “design and check” in parallel

TOE40G-IP shall strongly support product development to make maximum use of 40GbE merit!

For more detail

- Detailed documents available on the web site.
  - https://dgway.com/TOE40G-IP_A_E.html
- Contact
  - Design Gateway Co,. Ltd.
  - E-mail : ip-sales@design-gateway.com
  - FAX : +66-2-261-2290

Revision History

Rev.	Date	Description
1.0E	December 14, 2018	English version initial release