raNVMe-IP Introduction for Intel

Random Access to SSD without CPU

Random Write
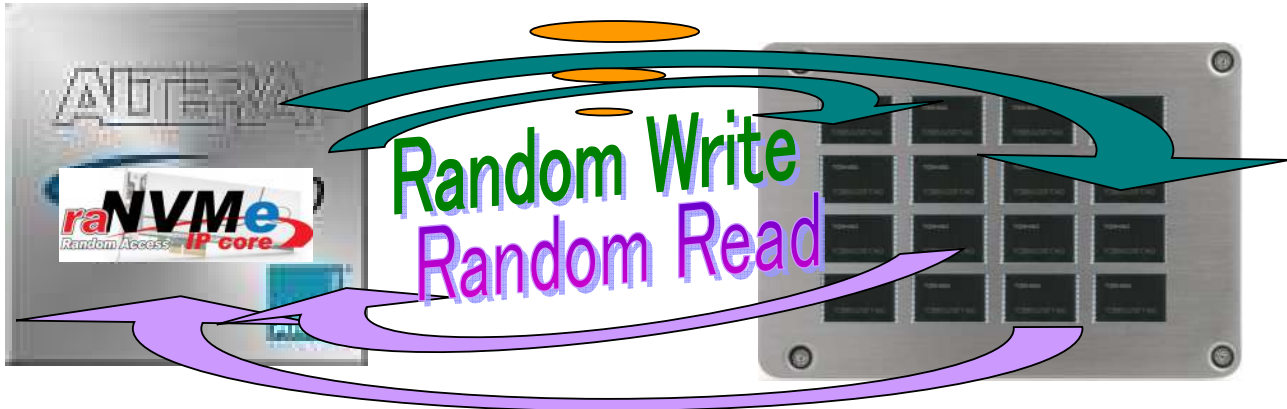Random Read

The Very Best Solution for Database Search Application!

What's raNVMe-IP

- **What's raNVMe-IP?** -> **Directly connect NVMe SSD with FPGA and execute random read or write access**
- **Advantage** -> **No need for CPU, its F/W, External Memory**
- **Application1** -> **Database search engine using SSD system**
- **Application2** -> **Video/Sensor data recording by Start-Stop operation.**

Simple User I/F

IP Core Body
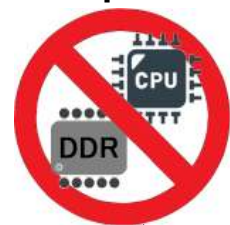
No Need for CPU and External DDR memory
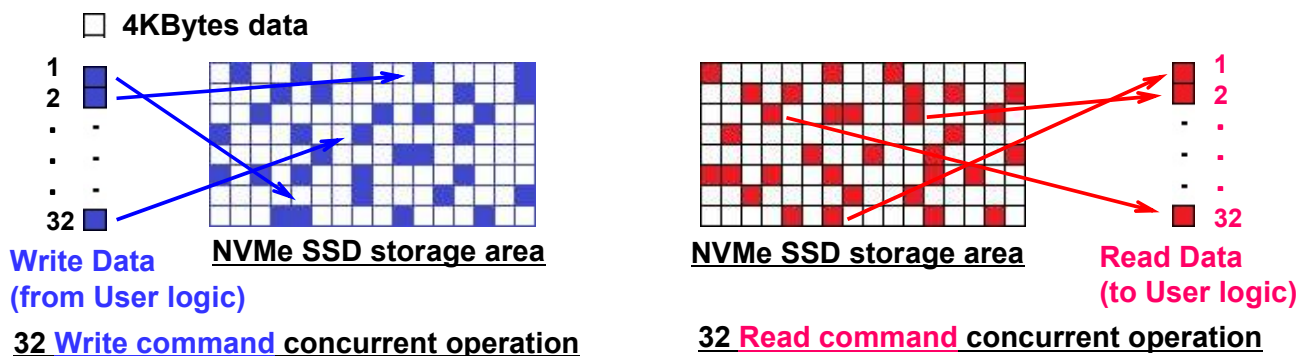
PCIe Hard IP in FPGA
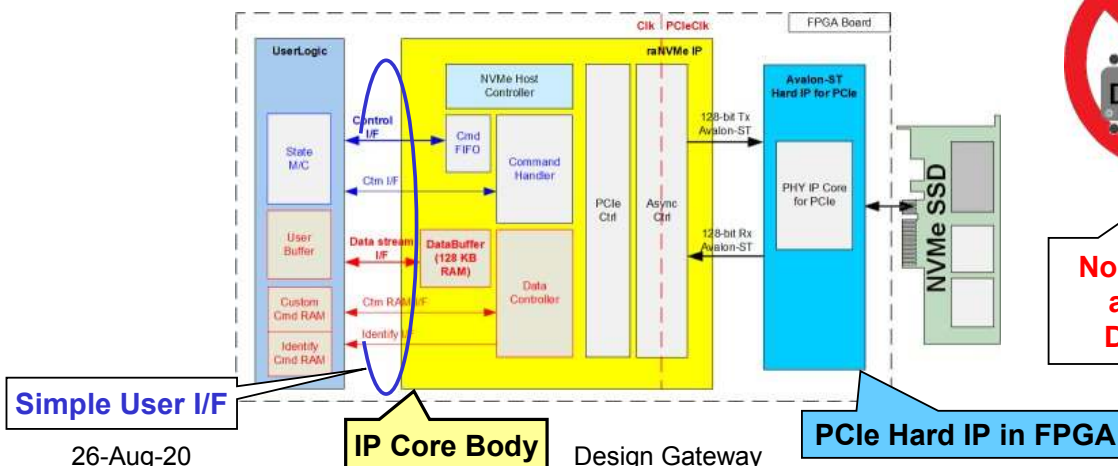
raNVMe-IP Merit

1. **Execute automatic SSD random write/read**
   - **Either write or read 32 commands at maximum in parallel.**
   - **Write or read data per one command is fixed to 4KBytes.**
2. **High Performance and Compact size**
   - **Write=592KIOPs, Read=226KIOPs**
   - **IP-Core Size=2016ALMs, Block Memory=136KBytes**
3. **Interface: Simple and easy connection**
   - **Direct connection to Intel PCIe Hard IP core**
   - **User I/F control is parameter with pulse, data is FIFO-like**
   - **Use BRAM for data buffer (external DDR memory not required)**
4. **Rich Features: Custom command in addition to Read/Write**
   - **Supports SMART/FLUSH/Shutdown custom command**
5. **Environment: Full reference design project**
   - **Full Quartus project with real board operation in the package**
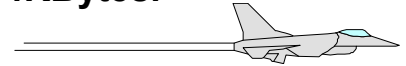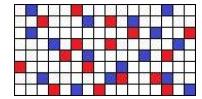
Merit1: Core Operation

- **User can select either Write or Read operation**
- **Executes 32 commands at maximum concurrently with different (random) address.**
- **Write or read data per one command is fixed to 4KBytes.**

☐ **4KBytes data**

Write Data (from User logic)

NVMe SSD storage area

NVMe SSD storage area

Read Data (to User logic)

32 Write command concurrent operation

32 Read command concurrent operation

raNVMe-IP concurrent command operation image

Merit2: Performance (Write)

```
+++ Write Command selected +++

Please input [Start Address] and [Length] in unit of 8
Select Transfer Mode             : [0] Sequential [1] Random => 0
Enable Data Verification         : [0] Disable    [1] Enable => 1
Enter Start Address (512 Byte)   : 0x0 - 0x3B9E12A8 => 0
Enter Length (512 Byte)          : 0x8 - 0x3B9E12A8 => 0x4000000
Selected Pattern [0]Inc32 [1]Dec32 [2]All_0 [3]All_1 [4]LFSR => 4
```

In Progress **Progress status**
100%

```
Total =  34.359 [GB] , Time = 14663 [ms]
Transfer speed = 2343 [MB/s], 599K [IOPS]
```
Output performance

Sequential Write Performance:
2343MB/s

```
+++ Write Command selected +++

Please input [Start Address] and [Length] in unit of 8
Select Transfer Mode             : [0] Sequential [1] Random => 1
Enable Data Verification         : [0] Disable    [1] Enable => 0
Enter Start Address (512 Byte)   : 0x0 - 0x3B9E12A8 => 0
Enter Length (512 Byte)          : 0x8 - 0x3B9E12A8 => 0x4000000
```

In Progress **Progress status**
100%

Random Write Performance:
592,000IOPS

```
Total =  34.359 [GB] , Time = 14839 [ms]
Transfer speed = 2315 [MB/s], 592K [IOPS]
```

Evaluation condition:
SSD: Samsung SSD 970PRO 512GB

Write Performance Evaluation Result of raNVMe-IP core

Merit2: Performance (Read)

```
+++ Read Command selected +++

Please input [Start Address] and [Length] in unit of 8
Select Transfer Mode             : [0] Sequential [1] Random => 0
Enable Data Verification         : [0] Disable    [1] Enable => 1
Enter Start Address (512 Byte)   : 0x0 - 0x3B9E12A8 => 0
Enter Length (512 Byte)          : 0x8 - 0x3B9E12A8 => 0x4000000
Selected Pattern [0]Inc32 [1]Dec32 [2]All_0 [3]All_1 [4]LFSR => 4
```

In Progress **Progress status**
100%

```
Total =  34.359 [GB] , Time = 14969 [ms]
Transfer speed = 2295 [MB/s], 587K [IOPS]
```
Output performance

Sequential Read Performance:
2295MB/s

```
+++ Read Command selected +++

Please input [Start Address] and [Length] in unit of 8
Select Transfer Mode             : [0] Sequential [1] Random => 1
Enable Data Verification         : [0] Disable    [1] Enable => 0
Enter Start Address (512 Byte)   : 0x0 - 0x3B9E12A8 => 0
Enter Length (512 Byte)          : 0x8 - 0x3B9E12A8 => 0x4000000
```

In Progress **Progress status**
100%

Random Read Performance
226,000IOPS

```
Total =  34.359 [GB] , Time = 38893 [ms]
Transfer speed = 883 [MB/s], 226K [IOPS]
```
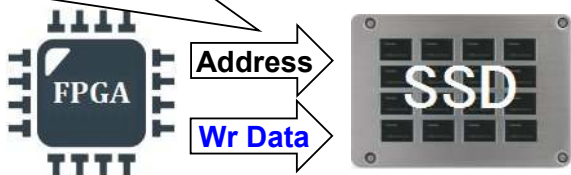
Evaluation condition:
SSD: Samsung SSD 970PRO 512GB

Read Performance Evaluation Result of raNVMe-IP core

Merit2: SSD performance description

1. **Write performance is high for both sequential and random**
    - SSD uses internal cache for write command to increase apparent speed.
    - SSD notify write completion even real data write to NAND memory is not finished (write back operation).
    - Can reduce NAND block erase count by grouping write destinations.
2. **Sequential read is fast but random read speed is dropped**
    - Pre-fetch is possible on the SSD side for sequential read.
    - Random read destination is unpredictable, so individual read is unavoidable.
    - SSD must start read access after destination info. arrival by read command.

Address and data direction is the same in Write cmd.
-> Pipeline parallel process is possible.

Address and data direction is opposite in Read cmd.
->Real read is necessary in each operation.

FPGA → Address → SSD
FPGA → Wr Data → SSD

Signal direction in Write command

FPGA → Address → SSD
FPGA ← Rd Data ← SSD

Signal direction in Read command

Merit2: Compact Size

- **Optimized size with minimum resource consumption**
    - Implements dedicated and optimized logic for NVMe SSD control
- **FPGA internal block memory for data buffer**
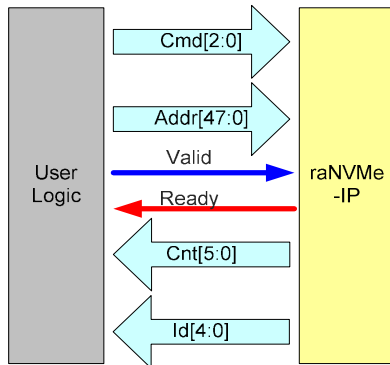   - Internal memory can minimize access overhead

Family	Example Device	Fmax (MHz)	ALMs	Registers	Block Memory bit	Design Tools
Arria10 GX	10AX115S2F45I2SG	300	2016	3437	1,114,176	QuartusII 16.0

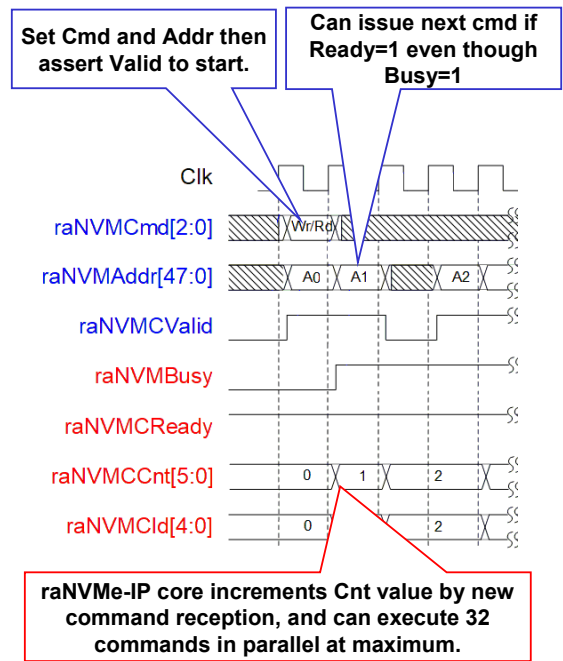**raNVMe-IP Core standalone resource usage**

A10AX1150 total ALM count is 427,200, raNVMe-IP core occupies 2016 = 0.47% logic utilization.

Merit3: Command I/F

- **Simple user interface**
  - Set Command/Address parameters
  - Start command (up to 32 cmds concurrently) with Valid pulse by monitoring Ready signal.
- **Cnt/Ld value for command progress check**
  - Cnt value shows concurrent command count in progress.
 - Id value indicates command in execution.
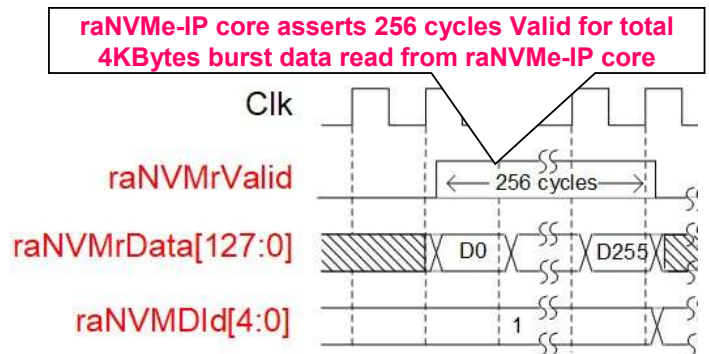

Basic Command I/F Signals

Set Cmd and Addr then assert Valid to start.

Can issue next cmd if Ready=1 even though Busy=1

raNVMe-IP core increments Cnt value by new command reception, and can execute 32 commands in parallel at maximum.

Command I/F waveform

Merit3: Data I/F

- **Data interface is 128bit bus for read/write individually**
  - 256 cycles burst transfer with Valid assertion for 4KBytes data xfer

User logic asserts 256 cycles Valid for total 4KBytes burst data write to raNVMe-IP core
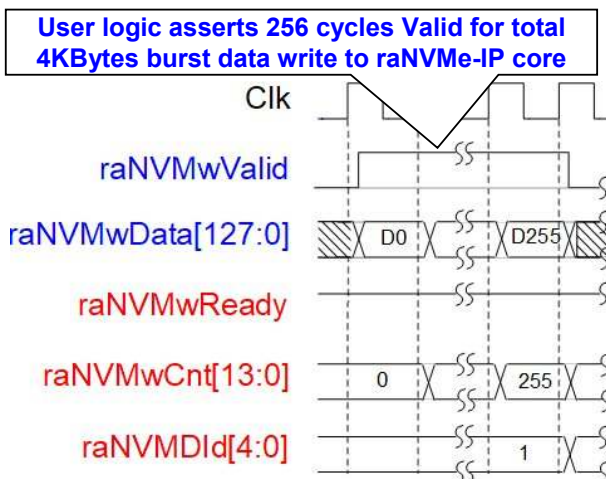
Write data to raNVMe-IP core

raNVMe-IP core asserts 256 cycles Valid for total 4KBytes burst data read from raNVMe-IP core

Read data fromraNVMe-IP core

Merit4: Rich Features

- **SMART command for SSD health condition check**
  - Can monitor internal temperature, total write size, etc.
- **FLUSH command to force cache flush operation**
  - User can adjust trade-off between performance and data evacuation
- **Safe Shutdown before SSD power down**
  - IP-core executes safe shutdown by user request
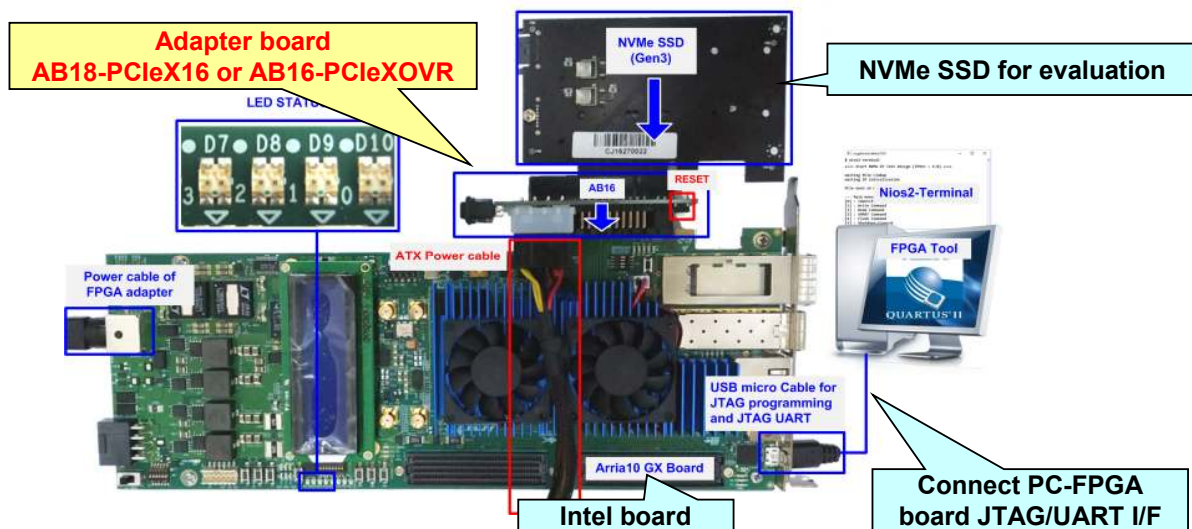
> User logic can monitor SSD status such as temperature or total written data count

```
<< SMART Log Information >>
Temperature              : 32 Degree Celsius
Total Data Read          : 47469 GB
Total Data Written       : 65373 GB
Power On Cycles          : 3991 Times
Power On Hours           : 79 Hours
Unsafe Shutdowns         : 220 Times
```

SMART command result example

Merit5: Environment

- **Real operation check with Intel evaluation board**
- **Free sof-file for evaluation before IP-core purchase**

Adapter board AB18-PCIeX16 or AB16-PCIeXOVR

NVMe SSD (Gen3)

NVMe SSD for evaluation

LED STATUS

Nios2-Terminal

FPGA Tool

Power cable of FPGA adapter

ATX Power cable

AB16

RESET

USB micro Cable for JTAG programming and JTAG UART

Arria10 GX Board

Intel board

Connect PC-FPGA board JTAG/UART I/F

raNVMe-IP evaluation environment (Arria10GX evaluation example)

Merit5: Reference Design

- **Quartus project attached with raNVMe-IP deliverables**
- **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

---

raNVMe-IP Application example1

- **Multiple user (host) sharing single SSD system**
  - Multiple user logic (sensor data etc.) shares single SSD.
  - Insert Arbiter between multiple user and raNVMe-IP core
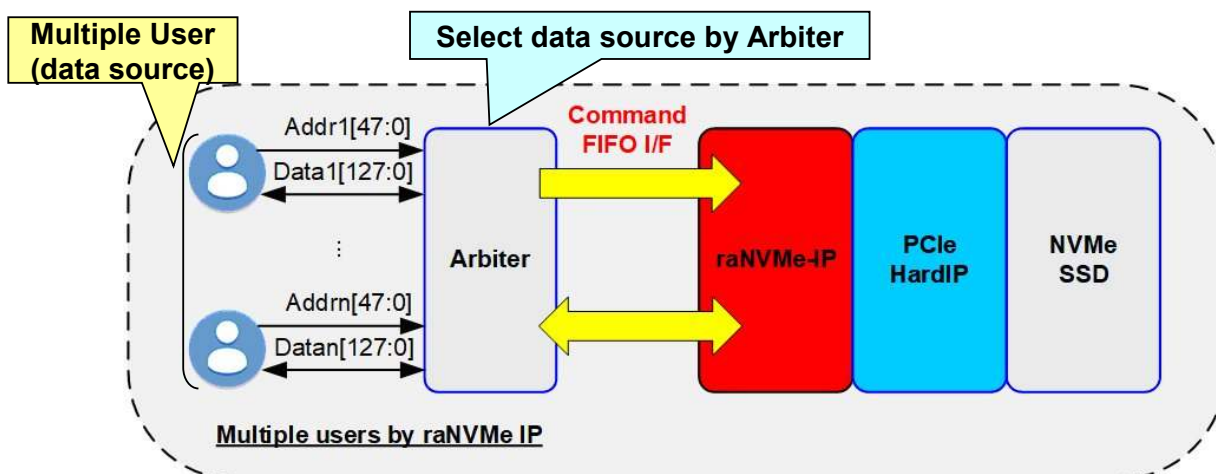
raNVMe-IP Application example2

- **Data Stream Application** (Video recording Start-Stop operation)
  - Suitable such application that total recording volume is unpredictable.
  - Recording by 4KBytes unit, fill dummy data (0-fill) for the last data.

Generate successive address
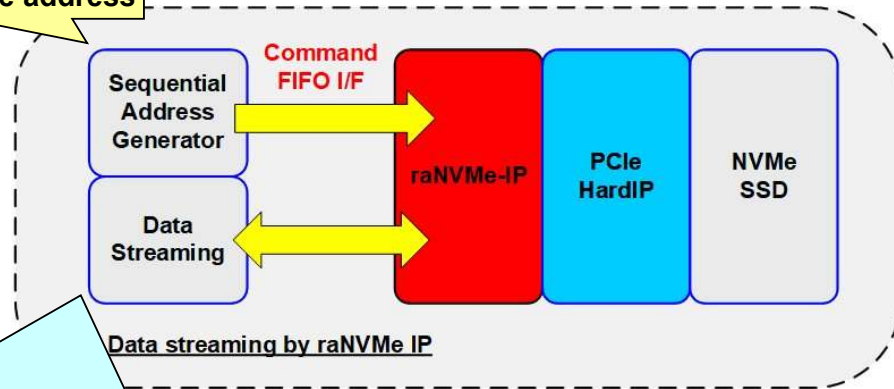
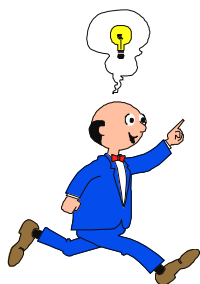**Recording data source (4KBytes unit, fill dummy data for extra data after stop)**

For more detail

- **Detailed technical information available on the web site.**
  - https://dgway.com/raNVMe-IP_A_E.html
- **Contact**
  - **Design Gateway Co,. Ltd.**
  - **sales@design-gateway.com**
  - **FAX: +66-2-261-2290**

Revision History

Rev.	Date	Description
1.0E	26th-Aug-2020	English version initial release