

Specification of NVMe interface to FPGA Fabric

Draft

Introduction

The DUNE experiment (<http://www.dunescience.org/>) store as much data possible if a potential Super Nova is detected. It is anticipated that this will be done by writing to NVMe attached storage (<https://nvmexpress.org/>). Approximately 300 devices will be written to in parallel for each 10kTonne detector module.

Assumptions

256GByte or 512GByte NVMe devices with a sustained write bandwidth of 2GByte/s. There will either be one or two NVMe devices connected to each FPGA

Specifications

- Firmware and software shall be supplied in source code form.
- Firmware and software may be released by the DUNE collaboration as Open Source (licence TBD) or retained within the collaboration
- Regardless of method of release and licencing both DUNE and BEAM Ltd will retain a permanent irrevocable licence to use and develop the firmware and software.
- Firmware will be targeted for Xilinx 7 series Ultrascale or Ultrascale Plus FPGAs
- Firmware will be targeted at programmable logic only (I.e. no use of Zynq ARM cores).
- A demonstration of the firmware and software must be supplied running on an Evaluation FPGA board. (Board and NVMe device to be agreed between BEAM and DUNE).
- Firmware and software must be as little encumbered by patent and licence restrictions as possible.
 - For example, if a soft core microprocessor is used it should be Open Source.
- Data shall be written in continuous bursts.
 - The maximum burst length (in data volume) will be known in advance.
 - The maximum burst length can either be a programmable parameter or used at synthesis time.
 - The burst length will be between 20GBytes and 200GBytes.
- The firmware should be able to cope with up to two separate bursts of data stored to NVMe
 - The third burst will overwrite the first, the fourth burst the second, etc.
- The maximum write speed should be limited by the NVMe drive, not the firmware.
- The interface for data writing will be AXI-4 Stream.
 - There will be additional user flags. These may include
 - “burst active” which goes high at the start of a burst and falls to indicate the end of a burst
 - “data valid” goes high to indicate that the data being presented should be written. May go high and low multiple times within a burst.
- The interface for data reading should be AXI-4 Lite or Wishbone.

- o The readout can be “block oriented” with a register written to indicate what block should be read and a status register to indicate that the block has been read to memory buffer.
 - o It should be possible to read data at an average speed of 100MBit/s or greater.
- The same interface can be used for setup and control and for reading data
- Functionality (for example, setting up drive, preparing lists of blocks to write, etc.) can be moved between the firmware and software but an “end-to-end” implementation should be demonstrated for setup , monitoring, writing and reading
- Documentation about the structure of the firmware and software should be provided to enable maintenance by DUNE.