Development of Network Interface Cards for TRIDAQ Systems with the NaNet Framework

TWEPP2016
Karlsruhe, Germany
September 26-30, 2016

R. Ammendola[1], A. Biagioni[2], P. Cretaro[2], S. Di Lorenzo[4], O. Frezza[2], G. Lamanna[3], F. Lo Cicero[2], A. Lonardo[2], M. Martinelli[2], P. S. Paolucci[2], E. Pastorelli[2], R. Piandani[4], L. Pontisso[4], D. Rossetti[5], F. Simula[2], M. Sozzi[4], P. Valente[2], P. Vicini[2]

[1]Sezione di Tor Vergata, Istituto Nazionale di Fisica Nucleare, Rome, Italy, [2]Sezione di Roma, Istituto Nazionale di Fisica Nucleare, Rome, Italy,
[3]Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, Frascati (Rome), Italy,[4]Sezione di Pisa, Istituto Nazionale di Fisica Nucleare, Pisa, Italy, [5]nVIDIA Corp, Santa Clara, CA, USA

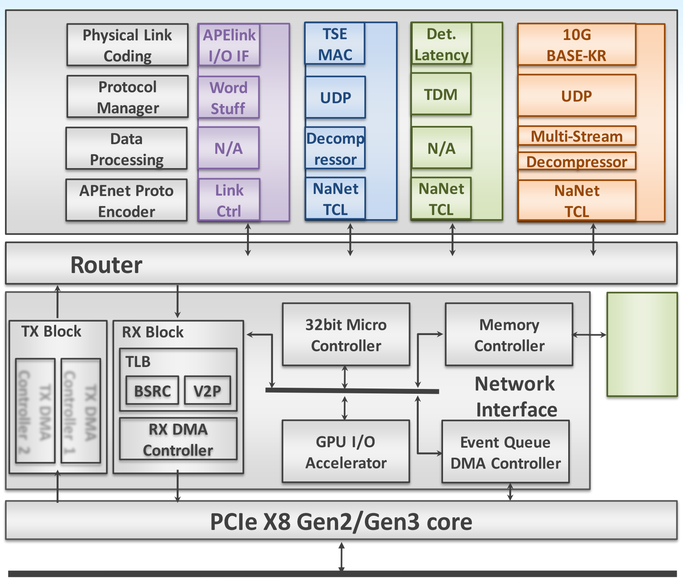
Abstract

NaNet is a framework for the development of FPGA-based PCI Express (PCIe) Network Interface Cards (NICs) with real-time data transport architecture that can be effectively employed in TRIDAQ systems.

Key features of the architecture are the flexibility in the configuration of the number and kind of the I/O channels, the hardware offloading of the network protocols stack, the stream processing and the zero-copy RDMA (for both CPU and GPU) capabilities.

Three NIC designs have been developed with the NaNet framework for the CERN NA62 L0 trigger and for the KM3NeT-IT underwater neutrino telescope DAQ system.

NaNet Design

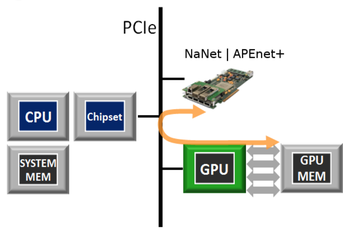


- I/O Interface
  - Multiple link.
  - Multiple network protocols.
    - Off-the-shelf: 1GbE, 10GbE
    - Custom: APElink (34 gbps/QSFP), KM3link
- Router
  - Dynamically interconnects I/O and NI ports.
- Network Interface
  - Manages packets TX/RX from and to CPU/GPU memory.
  - TLB & Nios II Microcontroller
    - Virtual memory management
- PCIe X8 Gen2 Core
  - CPU BW: 2.8 GB/s Read ÷ 2.5 GB/s Write
  - GPU BW: 2.5 GB/s Read & Write.
- Finalizing PCIe X8 Gen3 Core

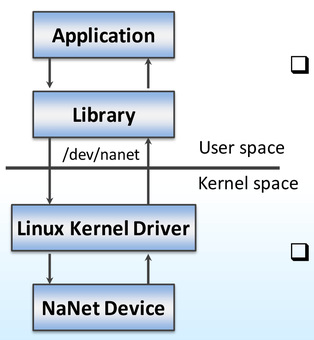
GPUDirect P2P/RDMA

GPUDirect allows direct data exchange on the PCIe bus with no CPU involvement (zero copy)

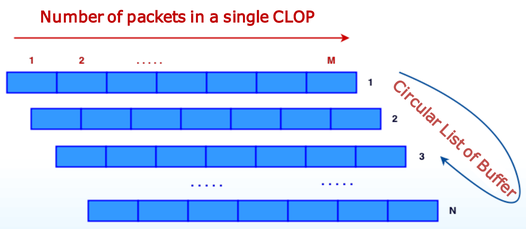
-> Latency reduction for small messages



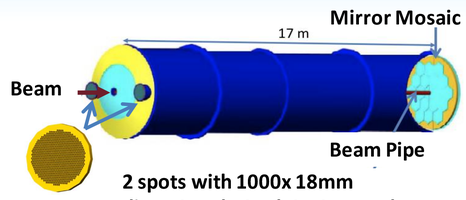
NaNet Software



- Host
  - User Space Application
  - User space Library (Open/Close, CLOP management,…)
  - Linux Kernel Device Driver
- NaNet Device
  - Nios II Microcontroller: single process software (bare metal) performing system configuration & initialization tasks

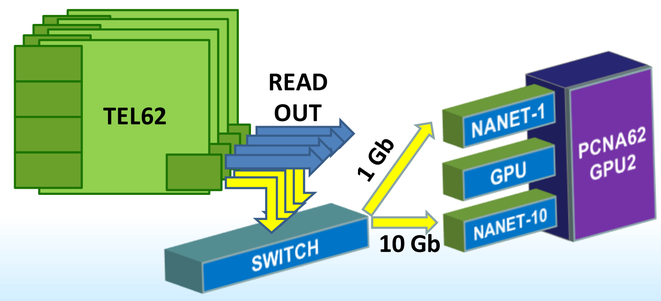


Case Study: NA62 RICH Detector



Ring-imaging Čerenkov detector

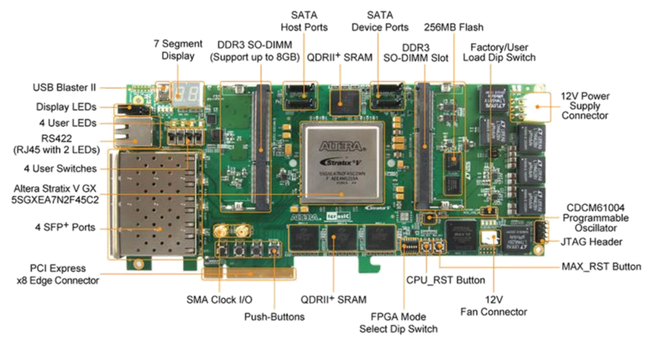
- Pion-Muon discrimination
- 70 ps time resolution
- **10 MHz** event rate
- **20** photons detected on average per single ring event (**hits** on photo-detectors)
- **40 Byte** per event



- 4 TEL62 for RICH detector
  - 8×1GbE links for data r/o
  - 4×1GbE trigger primitives
  - 4×1GbE GPU trigger
- Events rate: 10 MHz
- L0 trigger rate: 1 MHz
- Max Latency: 1 ms

NaNet-10 (four 10GbE SFP+ Ports)

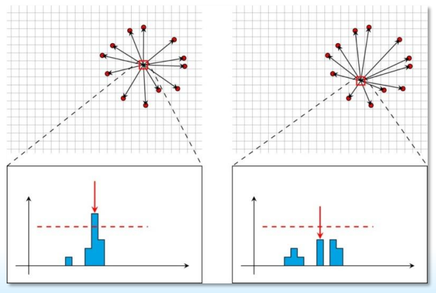
- ALTERA Stratix V Terasic DE5-NET dev board
- 4 SFP+ ports (Link speed up to 10 Gb/s)
- PCIe X8 Gen2/3
- GPUDirect P2P/RDMA capability
- UDP offload support
- Real-time decompressing and event merging capability
- Planned 40GbE development



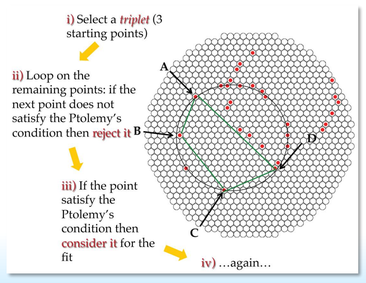
Rings reconstruction on GPU

Pattern recognition: Histogram

- XY plane divided into a grid
- An histogram is created with distances from these points and hits on the physics event
- Rings are identified looking at distance bins whose contents exceed a threshold value



A new multi-ring algorithm: Almagest



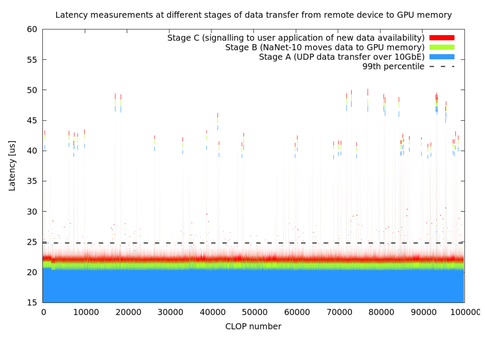
- Based on Ptolemy's theorem
- Several triplets run in parallel
- Several events at the same time
- **<0.5 µs per event** (multi-rings) for large buffers on Tesla K20c

Results

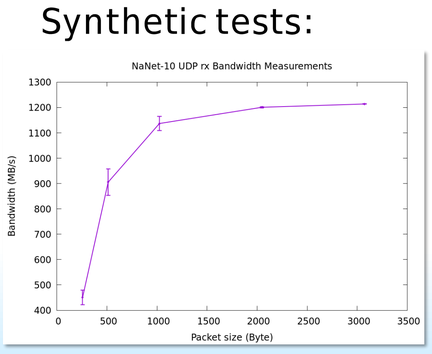
Synthetic tests

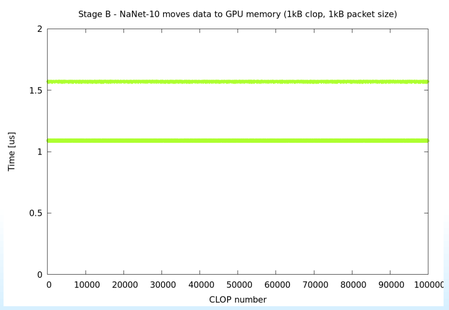
Test bed:

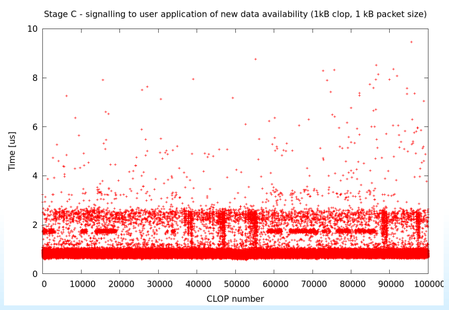
- Supermicro X9DRG-HF Intel C602 Patsburg , INTEL Xeon E5-2630 2.6 GHz 64 GB DDR3 nVIDIA K20Xm
- 1GbE NIC → switch → NaNet-10 for latency measurements
- 10GbE NIC → NaNet-10 for bandwidth



Synthetic tests:



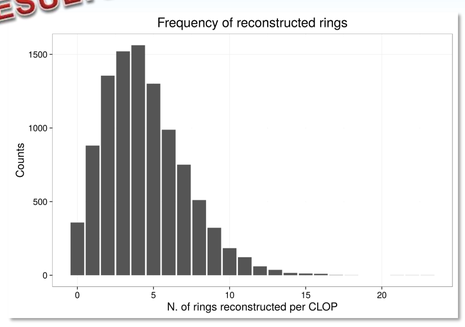


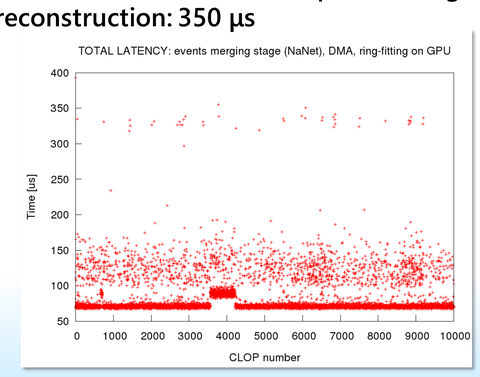


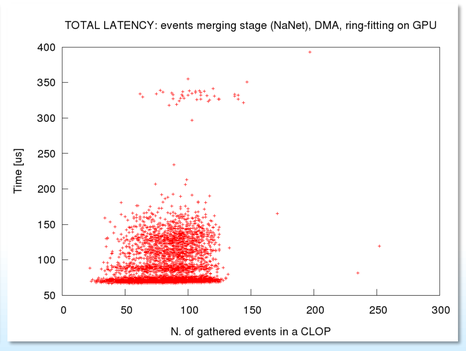
NA62 2016 Run Experimental results

PRELIMINARY RESULTS

- Supermicro X9DRG-QF Intel C602 Patsburg , INTEL Xeon E5-2602 2.0 GHz 32 GB DDR3 nVIDIA K20c
- ~25% target beam intensity ($9*10^{11}$ Pps)
- 1/16 downscaling factor
- 8 CLOP, each 32 kB
- Timeout on data collection prior to ring reconstruction: 350 µs







Contacts:
NaNet project: http://apegate.roma1.infn.it/nanet
APE project: http://apegate.roma1.infn.it/APE
Presenter Contact: paolo.cretaro@roma1.infn.it
NaNet coordinator: alessandro.lonardo@roma1.infn.it