

NaNet: Design of FPGA-Based Network Interface Cards for Real-Time Trigger and Data Acquisition Systems in HEP Experiments

NSS 2015
San Diego, CA
1-7 november 2015

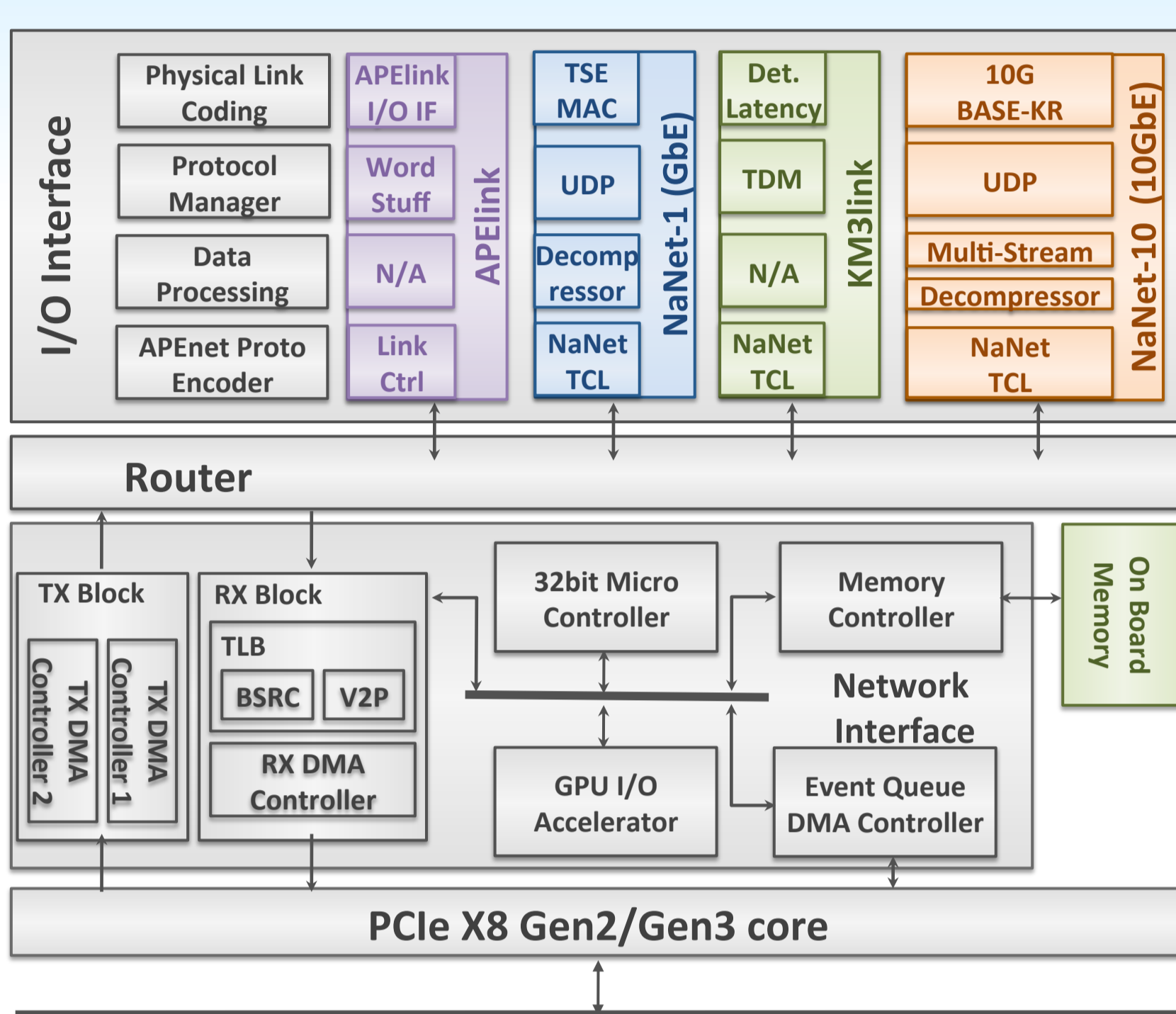
R. Ammendola¹, A. Biagioni², O. Frezza², G. Lamanna^{3,4}, F. Lo Cicero², A. Lonardo², M. Martinelli², P. S. Paolucci², E. Pastorelli², L. Pontisso⁵, D. Rossetti⁶, F. Simula², M. Sozzi^{3,5}, L. Tosoratto², P. Vicini²

¹Sezione di Tor Vergata, Istituto Nazionale di Fisica Nucleare, Rome, Italy, ²Sezione di Roma, Istituto Nazionale di Fisica Nucleare, Rome, Italy, ³CERN, Geneva, Switzerland, ⁴Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, Frascati (Rome), Italy, ⁵Sezione di Pisa, Istituto Nazionale di Fisica Nucleare, Pisa, Italy, ⁶NVIDIA Corp, Santa Clara, CA, USA

Abstract

NaNet is a modular design of a family of FPGA-based PCIe Network Interface Cards specialized for low-latency real-time applications with intensive I/O tasks involving the CPU and/or the GPU accelerators. NaNet features a Network Interface module that implements RDMA-style communications both with the host (CPU) and the GPU accelerators memories (GPUDirect RDMA) relying on the services of a high performance PCIe Gen3 x8 core. NaNet I/O Interface is highly flexible and is designed for low and predictable communication latency: a dedicated stage manages the network stack protocol in the FPGA logic offloading the host operating system from this task and thus eliminating the associated process jitter effects. Between the two above mentioned modules, stand the data processing and switch modules: the first implements application-dependent processing on streams, e.g. performing compression algorithms, while the second routes data streams between the I/O channels and the Network Interface module. This general architecture has been specialized up to now into three configurations, namely NaNet-1, NaNet³ and NaNet-10.

NaNet Design

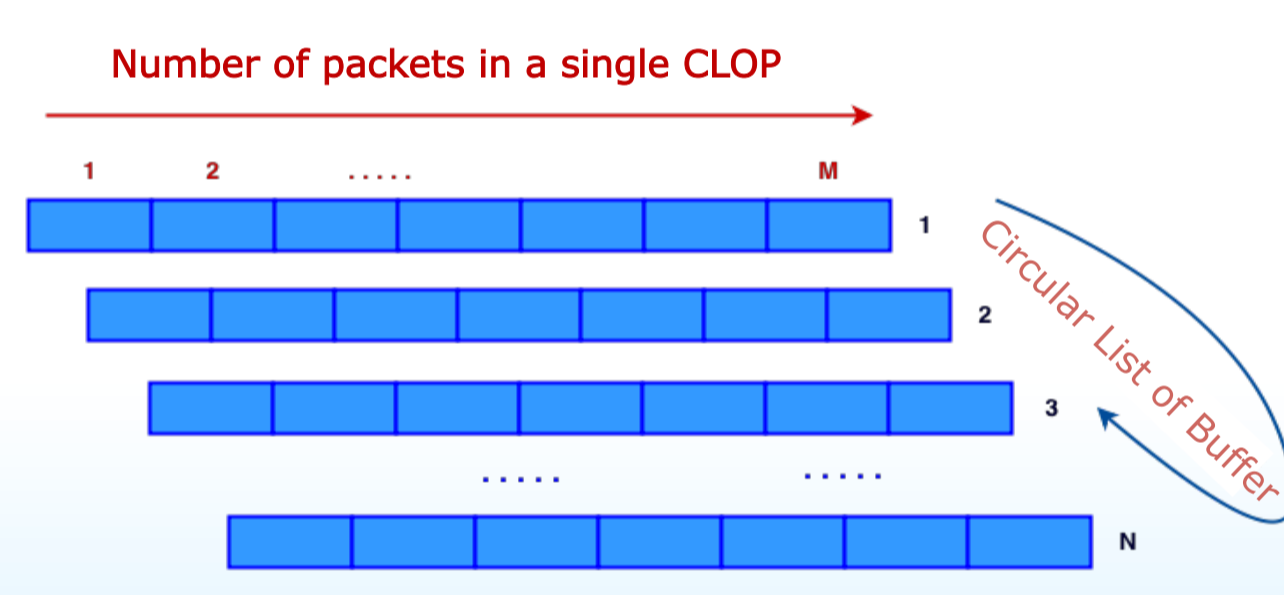


GPUDirect P2P/RDMA

GPUDirect allows direct data exchange on the PCIe bus with no CPU involvement (zero copy) -> Latency reduction for small messages

NaNet Software

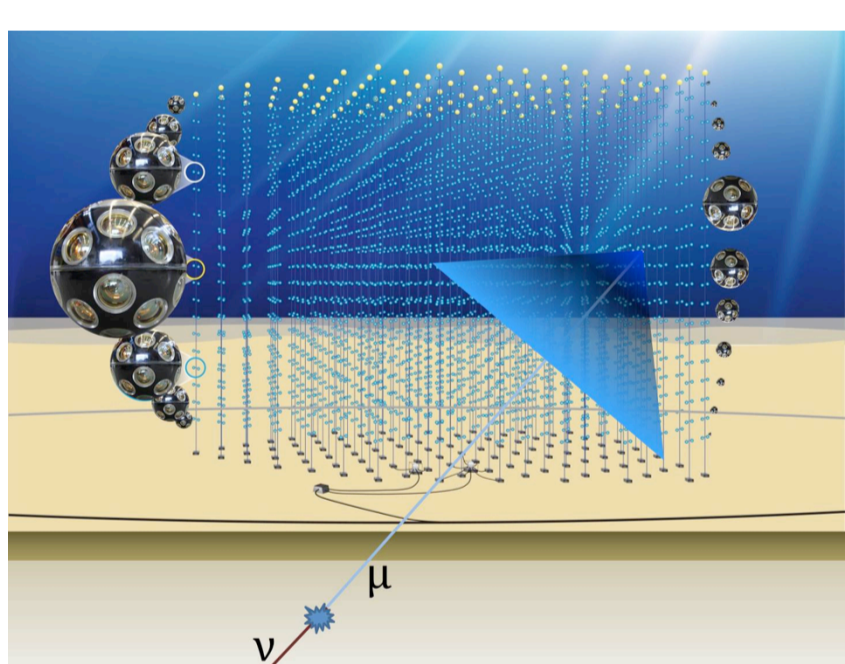
- Host
 - User Space Application
 - User space Library (Open/Close, CLOP management,...)
 - Linux Kernel Device Driver
- NaNet Device
 - Nios II Microcontroller: single process software (bare metal) performing system configuration & initialization tasks



KM3NeT-Italia experiment

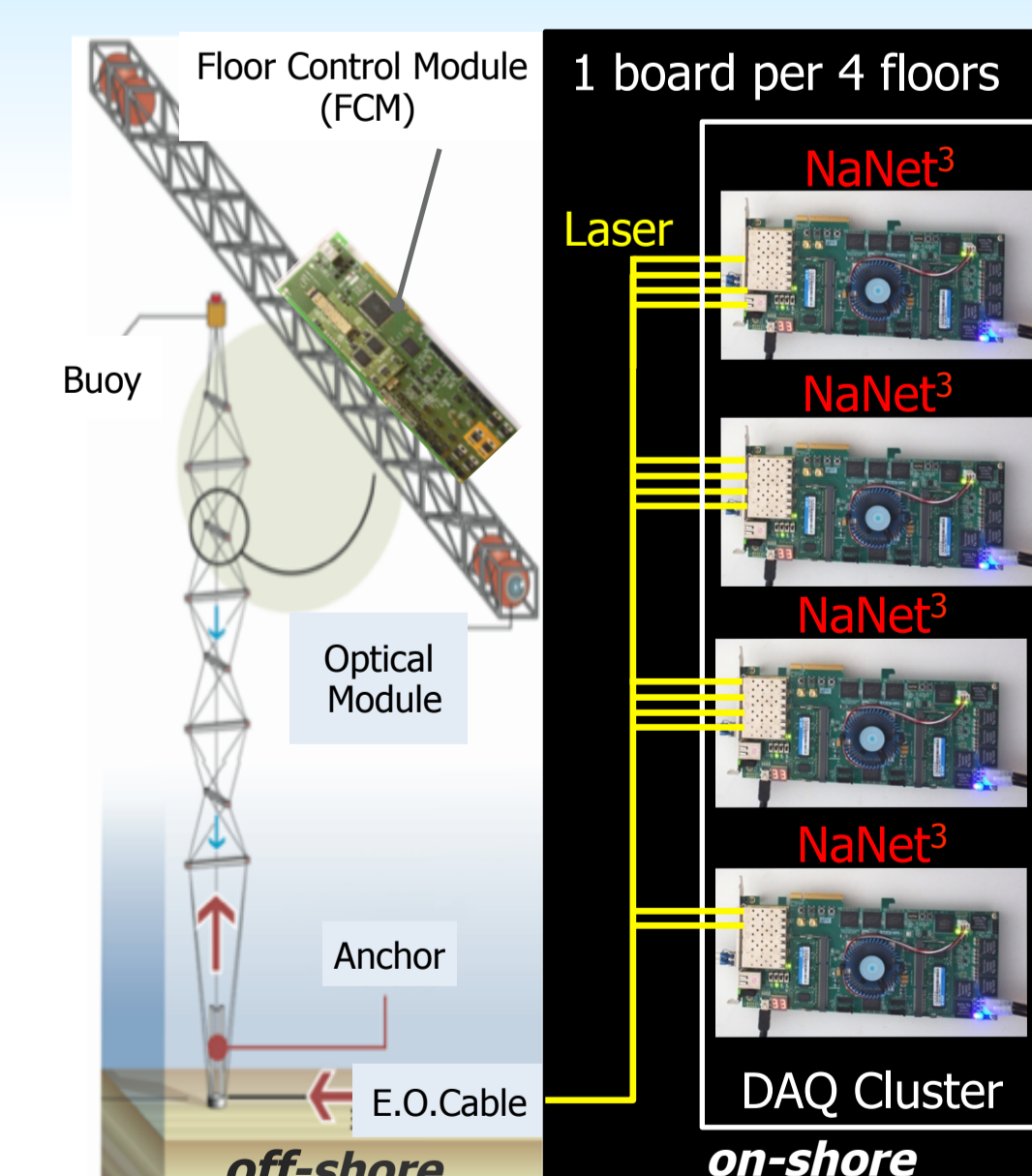
European deep-sea research infrastructure hosting a new generation of a neutrino telescope with a volume of several cubic kilometres located at the bottom of the Mediterranean Sea.

- Data produced by OMs, hydrophones, and instruments, are collected by an electronic board contained in a vessel at the centre of the floor (FCM board)
- FCM manages communication between the on-shore lab and the underwater devices, also distributing the timing information (GPS clock) and signals received from the on-shore equipment
- Deterministic latency links are required to obtain a common timing and known delay for the spatially distributed readout



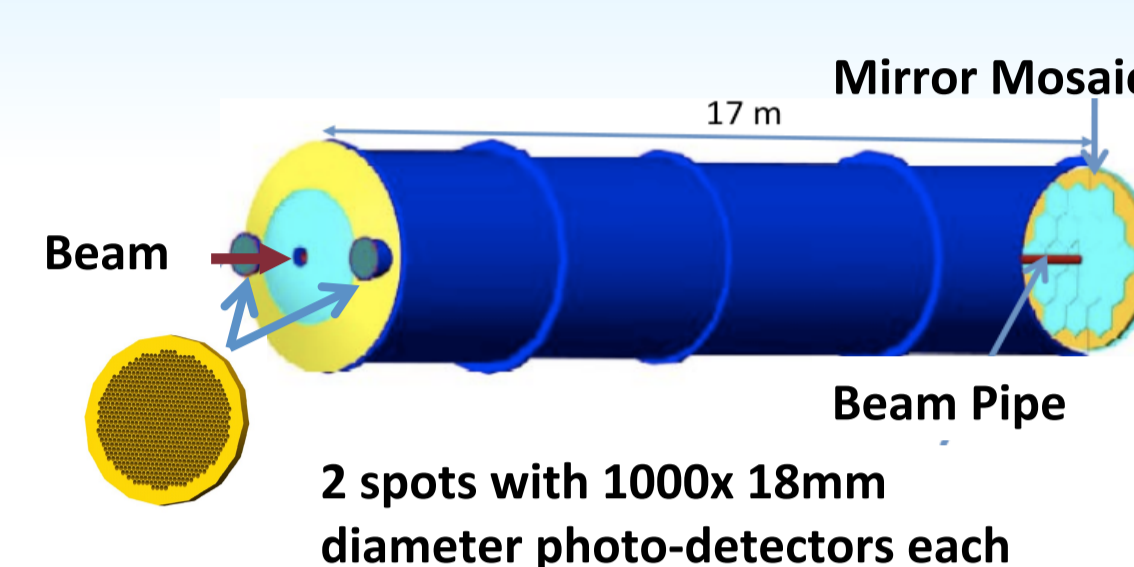
NaNet³

Is the counterpart for 4 FCM KM3NeT-it experiment boards.



- Implemented on the Terasic DE5-NET Stratix V FPGA dev board
- 4 custom 2.5 Gbps deterministic latency optical channels
- Link speed up to 10 Gb/s
- GPUDirect P2P/RDMA capability
- Deterministic latency link: employs Altera Deterministic Latency Transceivers with an 8B/10B encoding scheme as Physical Link Coding and Time Division MultiPlexing (TDMP) data transmission protocol

Case Study: NA62 RICH Detector

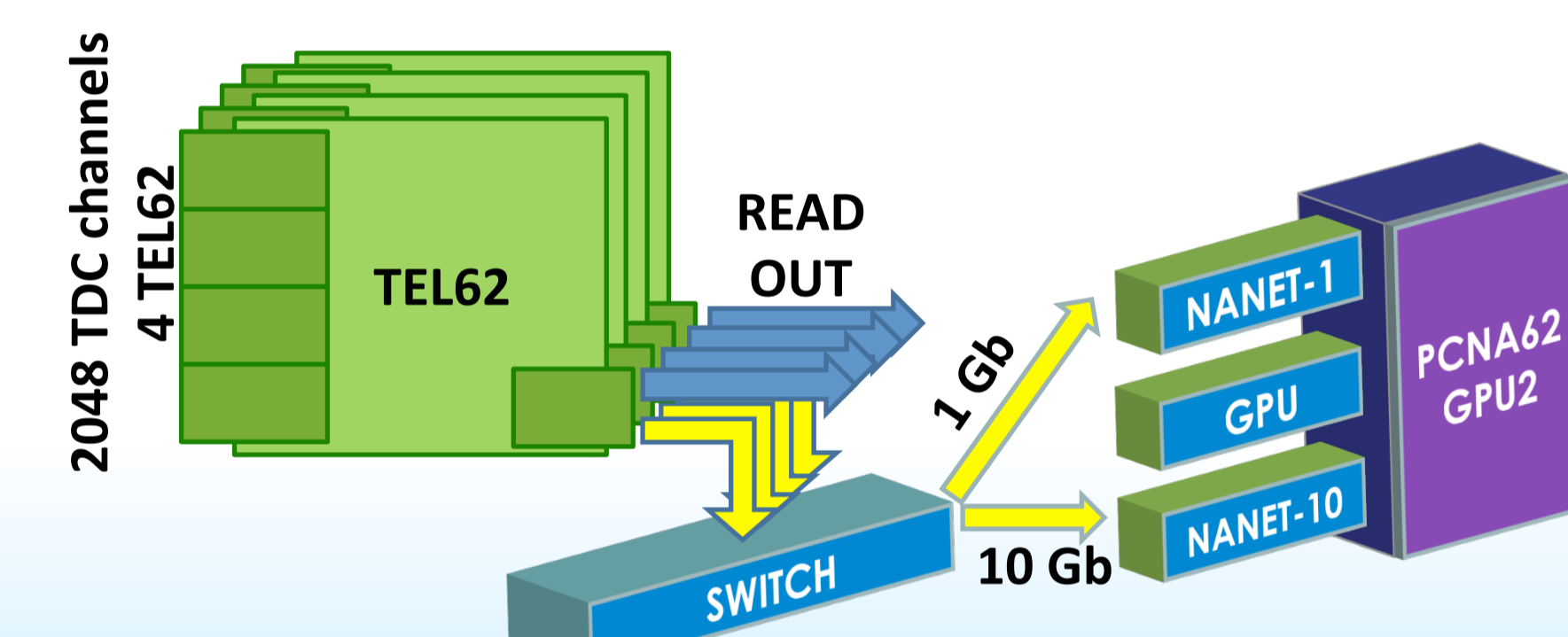
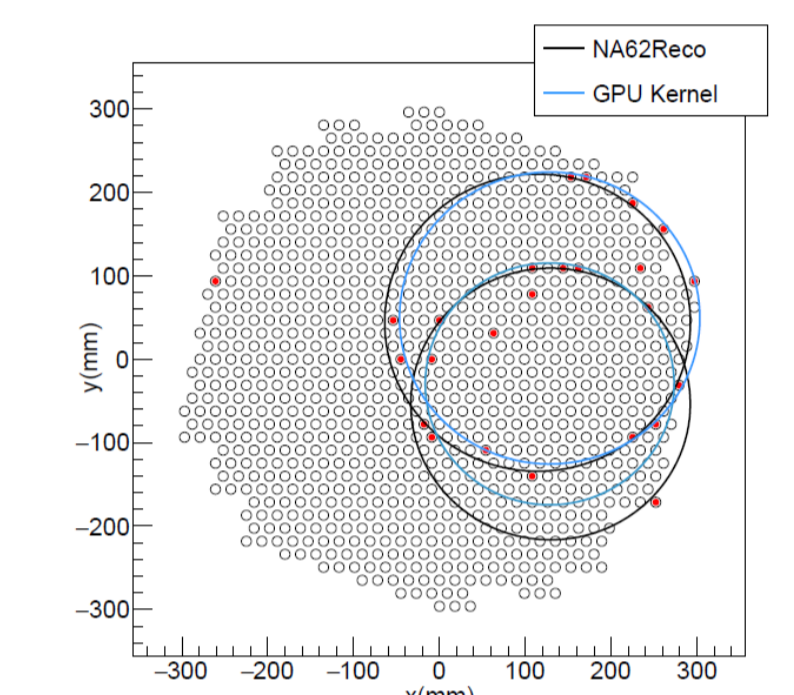


Ring-imaging Čerenkov detector

- 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

Rings pattern recognition and fit also performed on GPU:

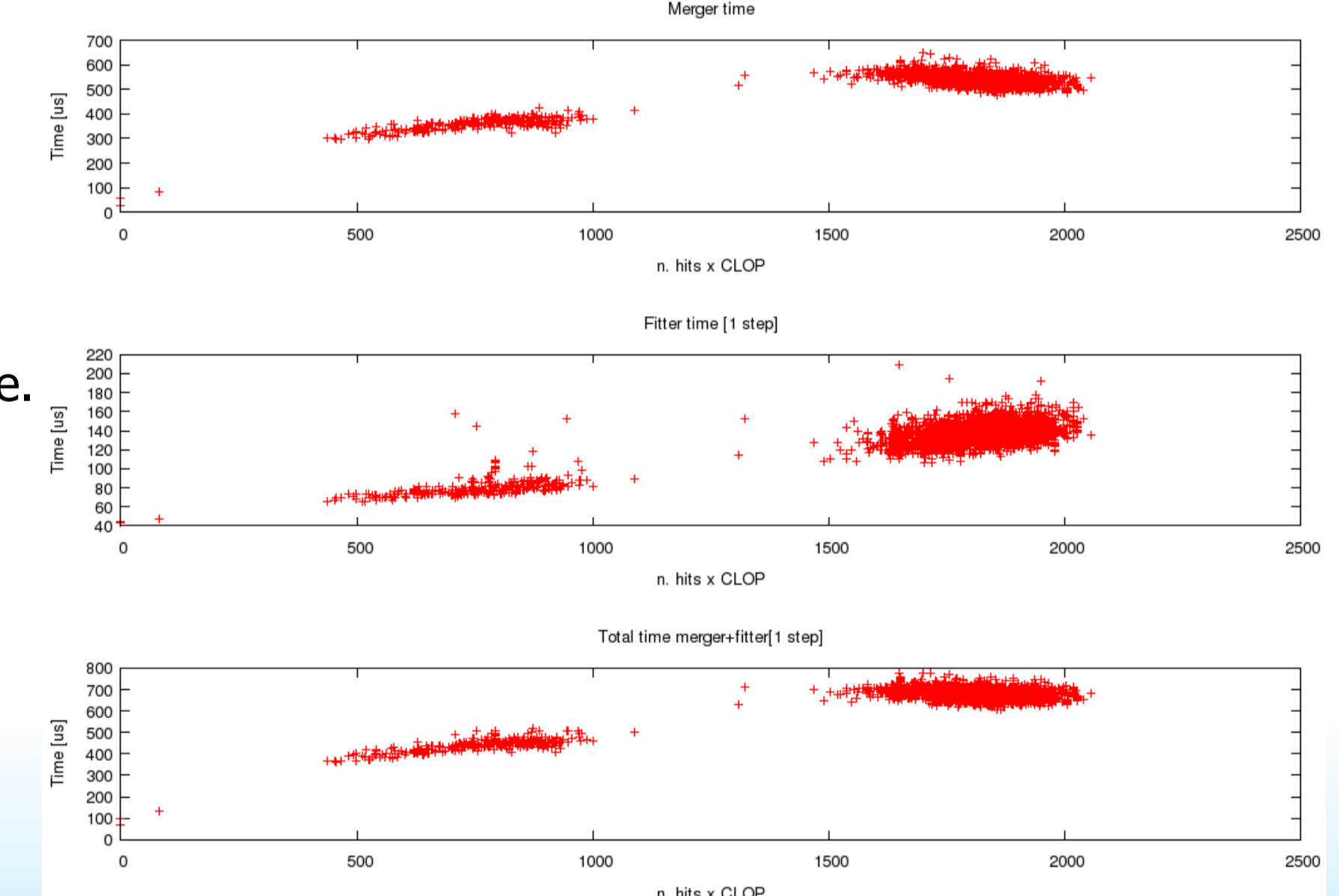
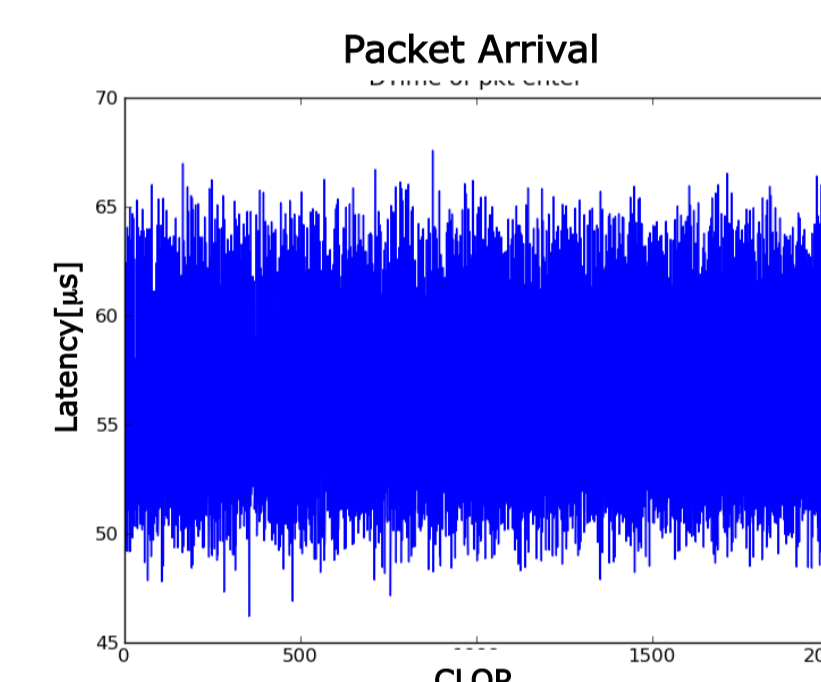
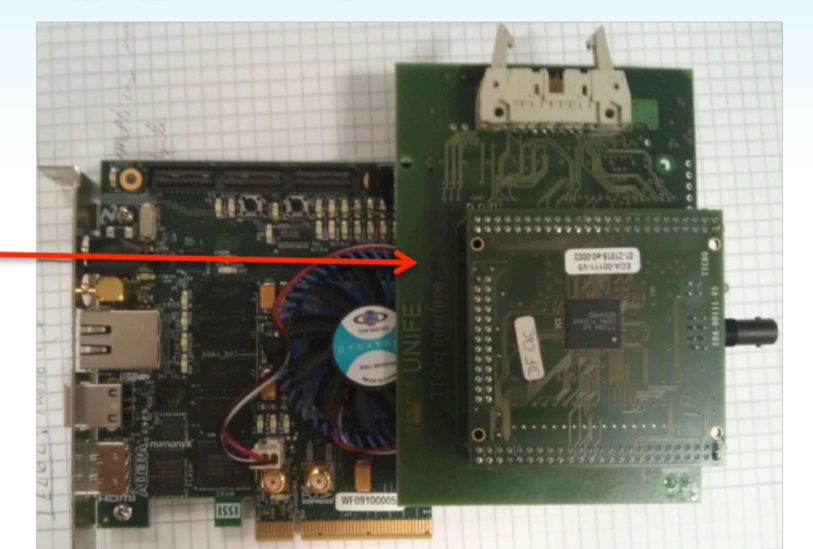
- New algorithm ("Almagest") developed for trackless, fast, and high resolution ring fitting
- Rough detection of particle speed (radius) and direction (centre).
- few μ s per event (on NVIDIA K20x)



- 4 TEL62 for RICH detector
 - 8x1GbE links for data r/o
 - 4x1GbE trigger primitives
 - 4x1GbE GPU trigger
- Events rate: 10 MHz
- L0 trigger rate: 1 MHz
- Max Latency: 1 ms

NaNet-1 in RICH low level trigger processor

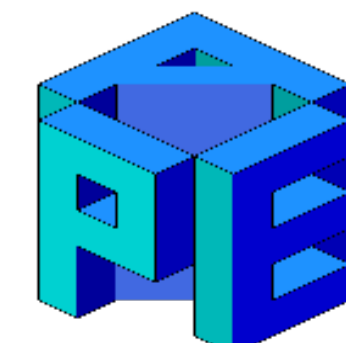
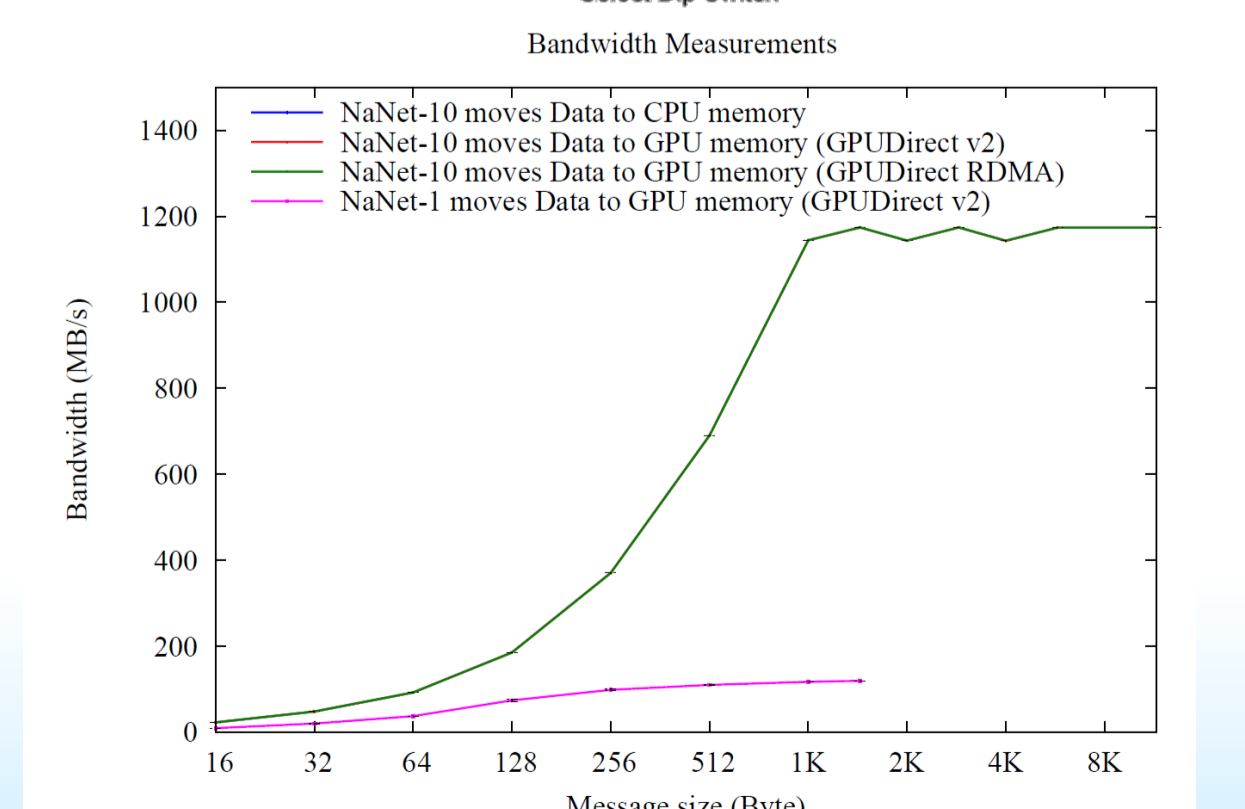
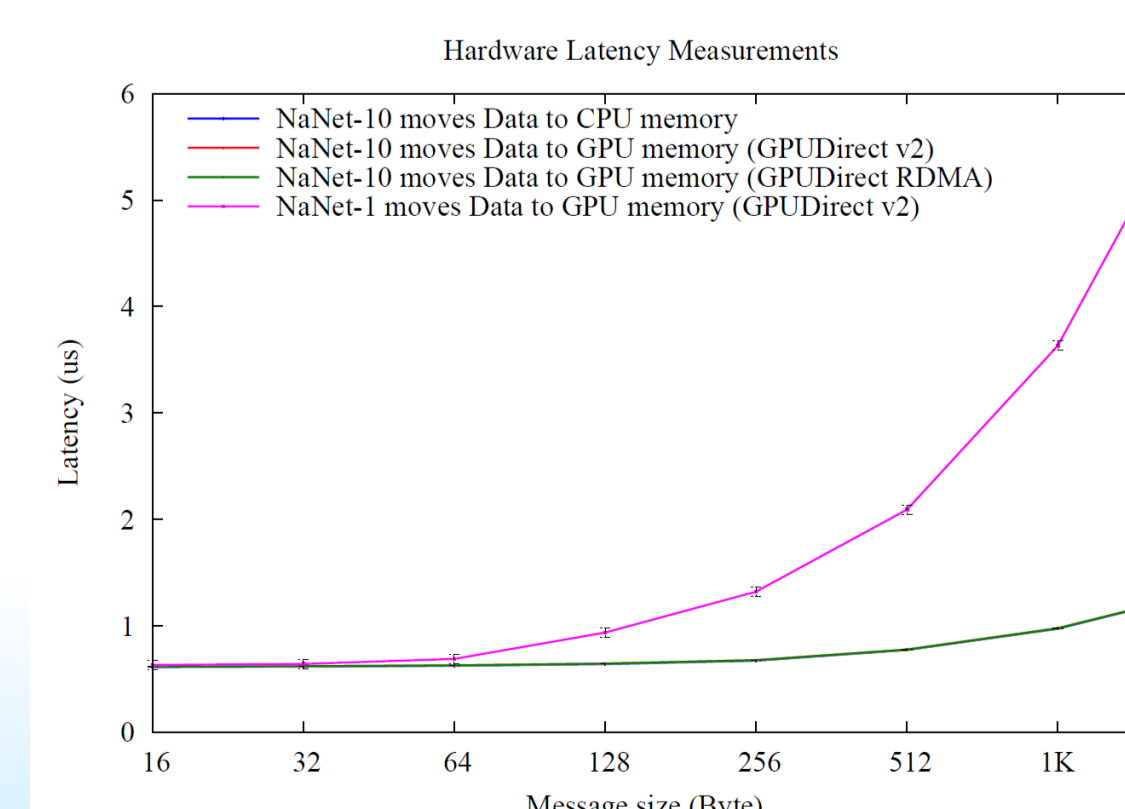
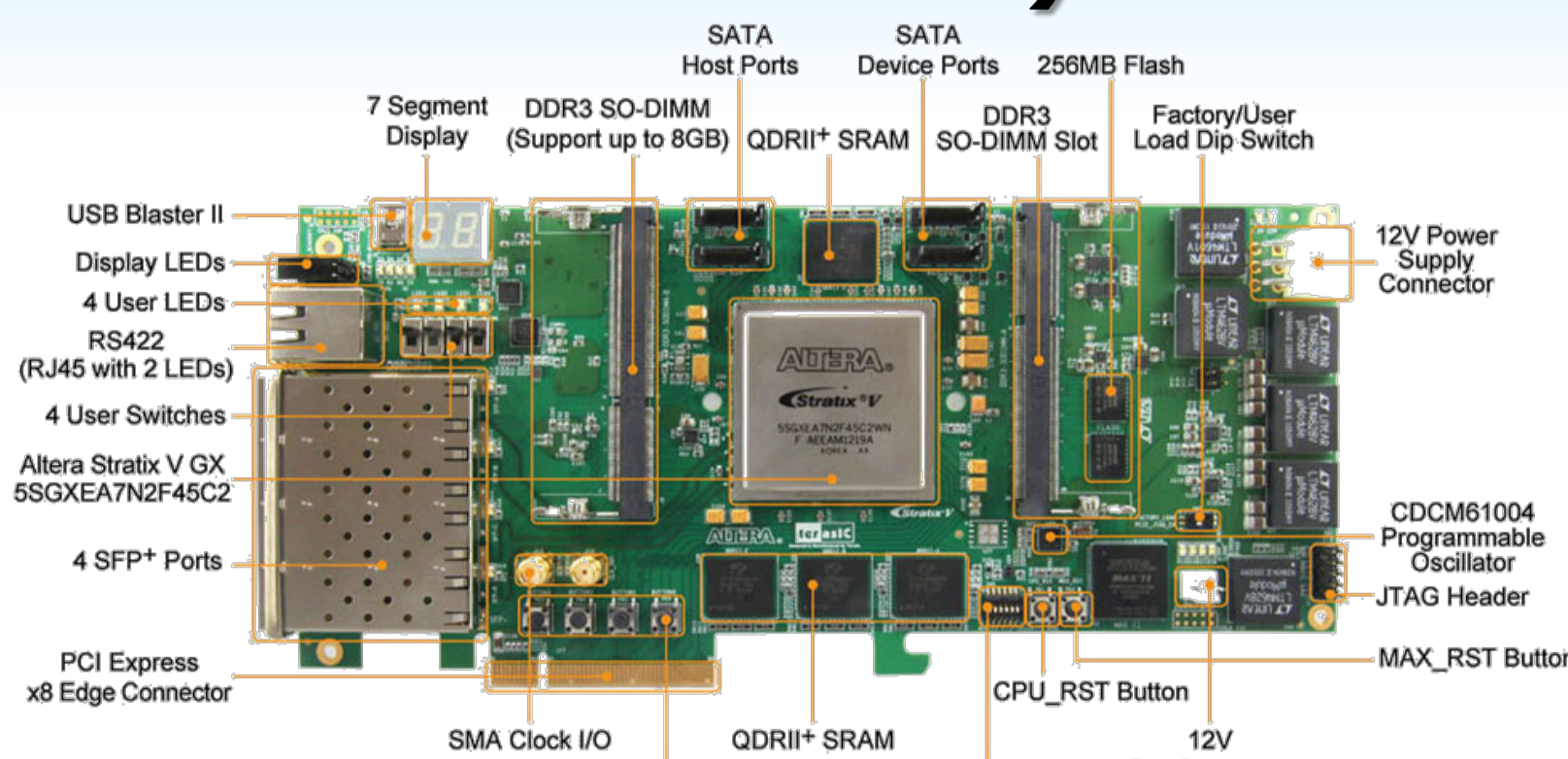
- Implemented on Altera Stratix IV dev board
- TTC daughtercard with HSMC connector for timing (clock, SOB/EOB) and trigger signals



- Merge time depends on data size. Future Speed up:
 - NOW performed on GPU
 - Working on FPGA implementation
- Computing time (K20c):
 - ~1 μ s per event

NaNet-10 (four 10GbE SFP+ Ports)

- ALTERA Stratix V Terasic DE5-NET dev board
- 4 SFP+ ports (Link speed up to 10 Gb/s)
- Implemented on Terasic DE5-NET board
- GPUDirect P2P/RDMA capability
- UDP offload supports
- Planned 40GbE development



Contacts:

The APE project: <http://apegate.roma1.infn.it/APE>
The EURETILE project: <http://euretile.roma1.infn.it/>
Presenter Contact: michele.martinelli@roma1.infn.it
This project was partially funded by the EURETILE european FP7 grant 247846.