



NaNet-10: a 10GbE Network Interface Card for the GPU-based Low-Level Trigger of the NA62 RICH Detector.

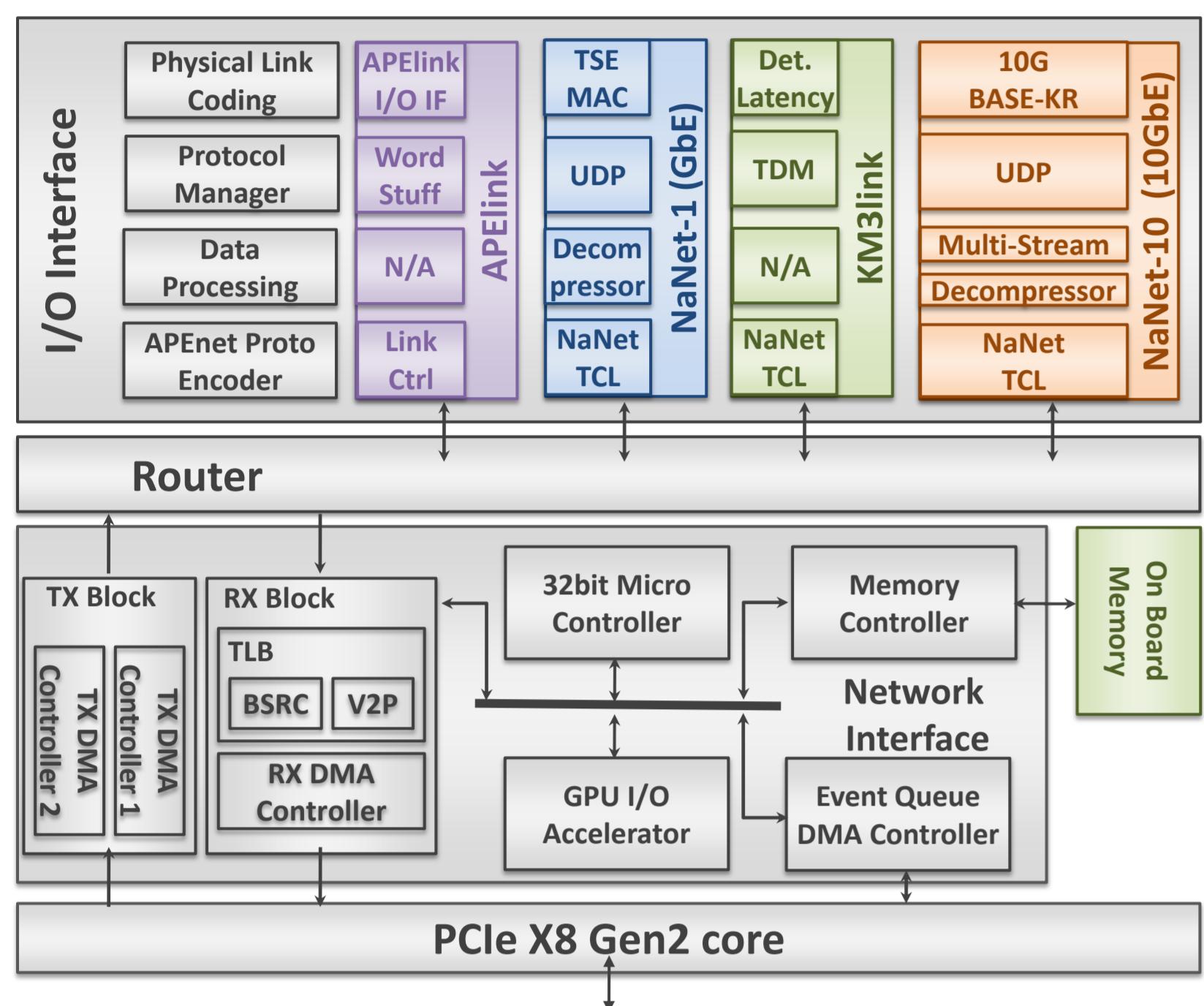
R. Ammendola^(a), A. Biagioni^(b), M. Fiorini^(c), O. Frezza^(b), G. Lamanna^(d), F. Lo Cicero^(b), A. Lonardo^(b), M. Martinelli^(b), I. Neri^(c), P.S. Paolucci^(b), E. Pastorelli^(b), L. Pontisso^(f), D.Rossetti^(e), F. Simula^(b), M. Sozzi^(f), L. Tosoratto^(b) and P. Vicini^(b)
 (a) INFN Sezione di Roma Tor Vergata (b) INFN Sezione di Roma (c) Università di Ferrara e INFN Sezione di Ferrara
 (d) INFN LNF and CERN (e) NVIDIA Corp. USA (f) INFN Sezione di Pisa and CERN

TWEPP2015
Lisbon, Portugal
28 Sep - 02 Oct
2015

NaNet Project Objectives

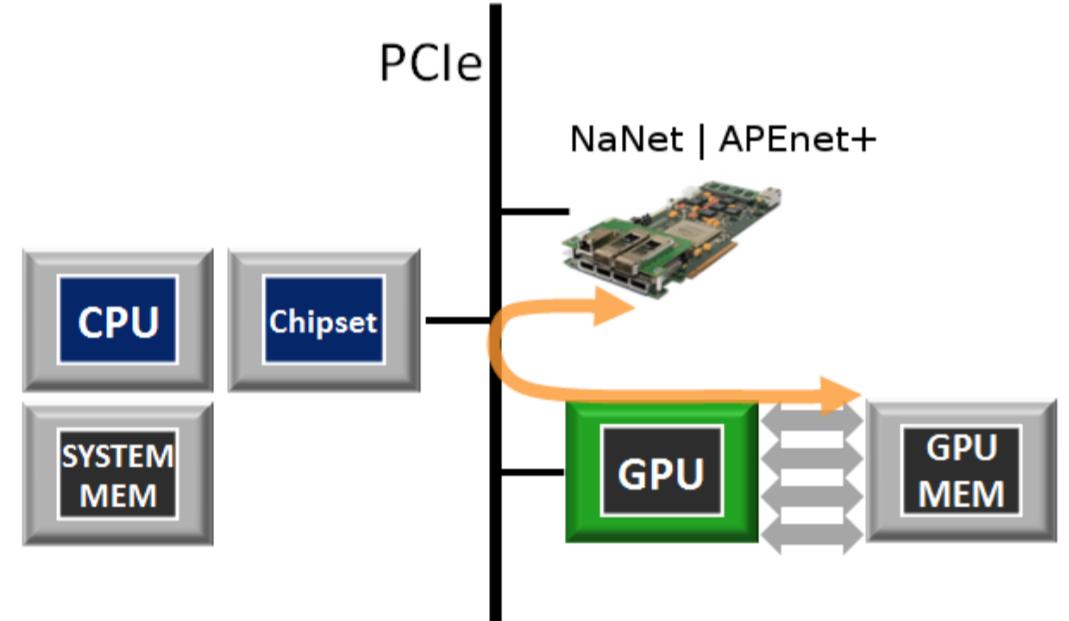
- Design and implementation of a family of FPGA-based PCIe Network Interface Cards:
- Bridging the front-end electronics and the software trigger computing nodes.
 - Supporting multiple link technologies and network protocols.
 - Enabling a low and stable communication latency.
 - Having a high bandwidth.
 - Processing data streams from detectors on the fly
 - Optimizing data transfers with GPU accelerators.

NaNet Modular Design



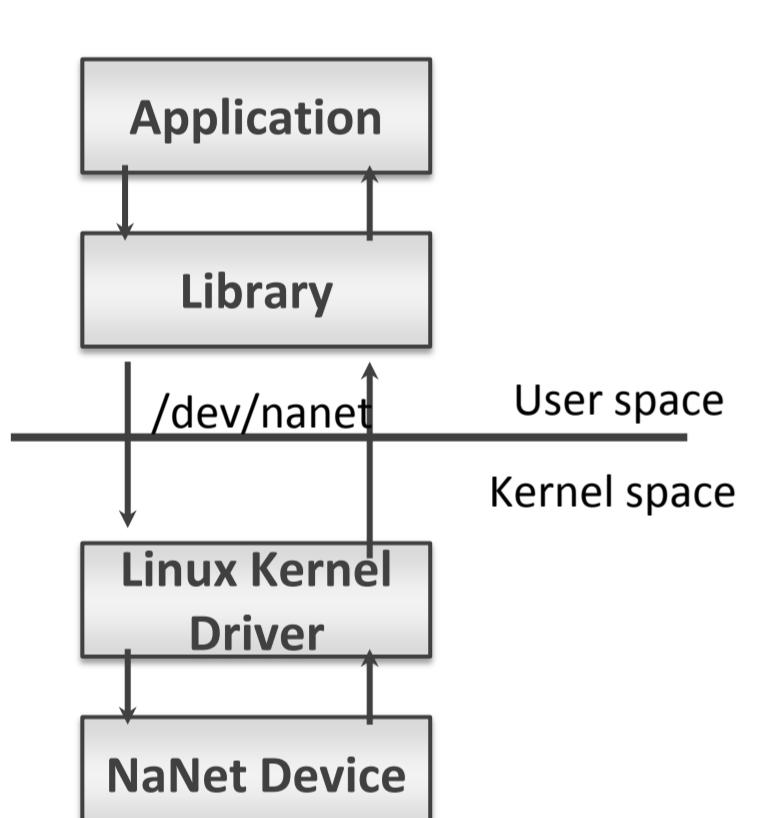
- I/O Interface
 - Multiple link.
 - Multiple network protocols.
 - Off-the-shelf: 1GbE, 10GbE
 - Custom: APEnet (34gbps/QSFP), KM3link
- Router
 - Dynamically interconnects I/O and NI ports.
- Network Interface
 - Manages packets TX/RX from and to CPU/GPU memory.
 - TLB & Nios II Microcontroller
 -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

NaNet Design – GPUDirect P2P/RDMA



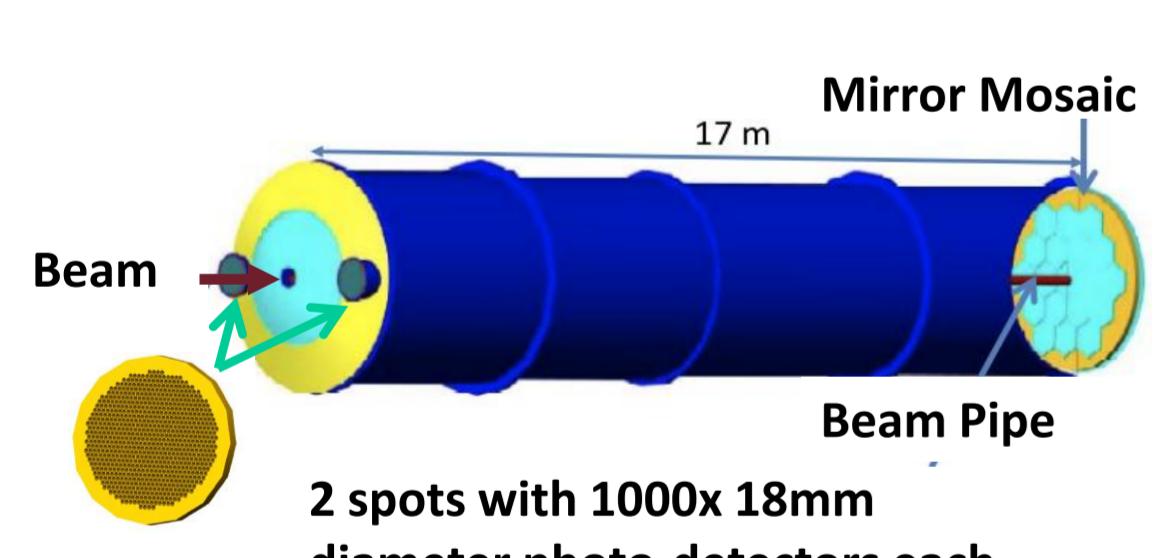
- GPUDirect allows direct data exchange on the PCIe bus with no CPU involvement.
- No bounce buffers on host memory.
- Zero copy I/O.
- Latency reduction for small messages.
- nVIDIA Fermi/Kepler/Maxwell

NaNet Software



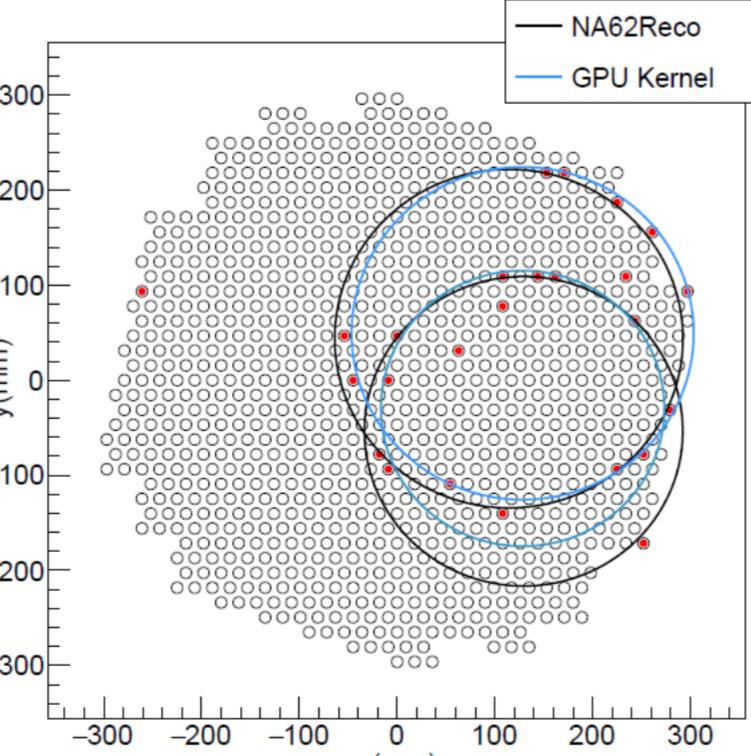
- Host
 - Linux Kernel Driver
 - User space Library (open/close, buf reg, wait recv evts, ...)
- Nios II Microcontroller
 - Single process program performing System Configuration & Initialization tasks.

Case Study: NA62 RICH Detector



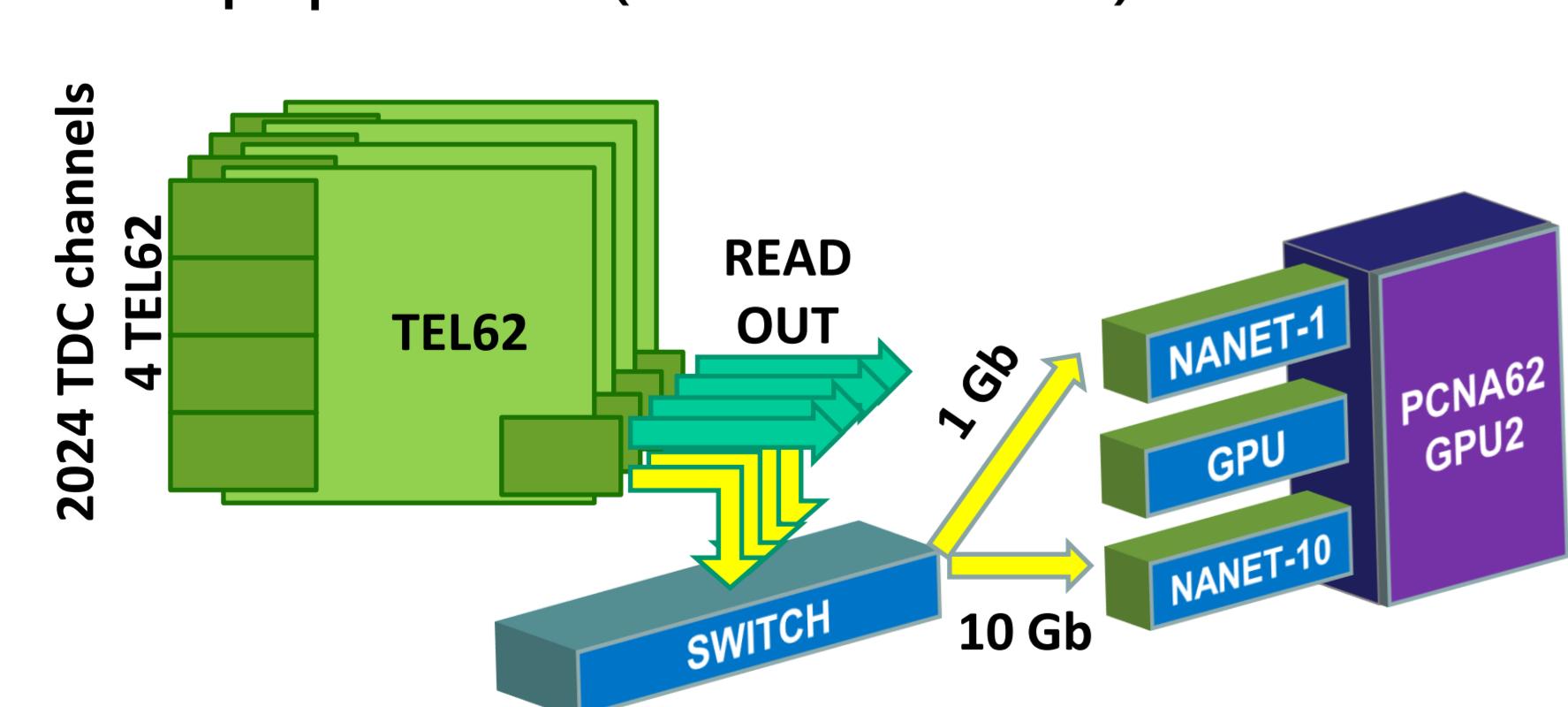
Ring-imaging Čerenkov detector

- 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



Rings pattern recognition and fit also performed on GPU:

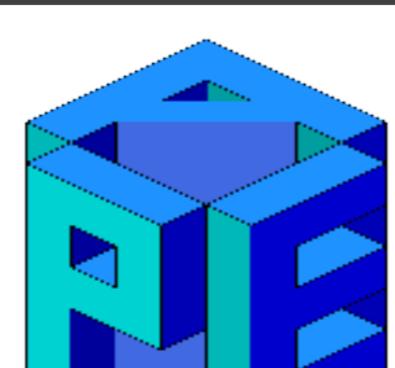
- New algorithm ("Almagest") developed for trackless, fast, and high resolution ring fitting.
- Rough detection of particle speed (radius) and direction (centre).
- few μ s per event (on NVIDIA K20x).



- 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

Contacts

<http://apegate.roma1.infn.it>
<http://euretile.roma1.infn.it>
<http://na62.web.cern.ch/na62/>

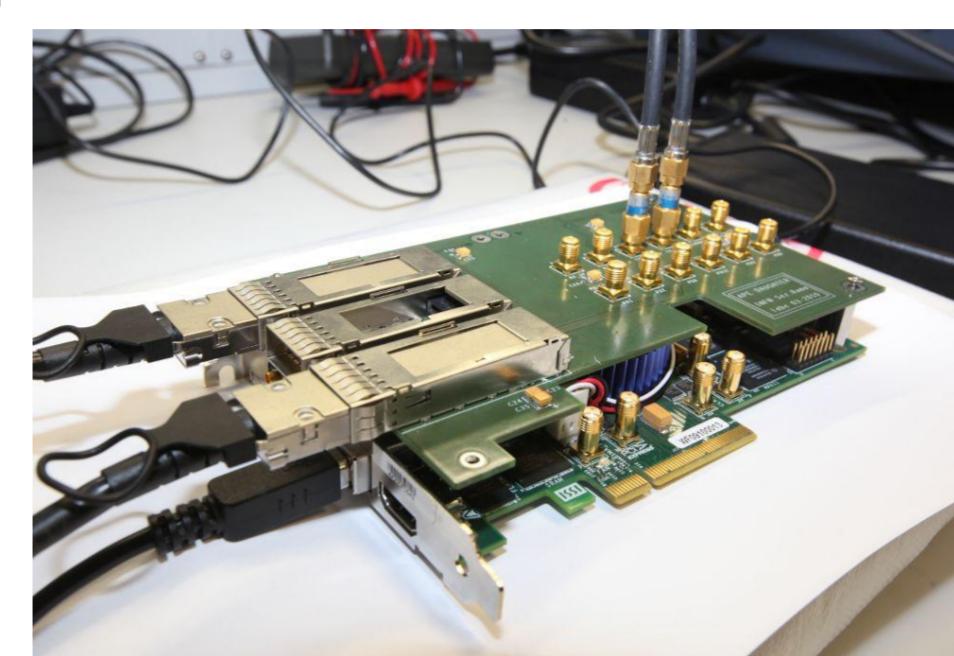


alessandro.lonardo@roma1.infn.it
 piero.vicini@roma1.infn.it
 Presenter: andrea.biagioni@roma1.infn.it

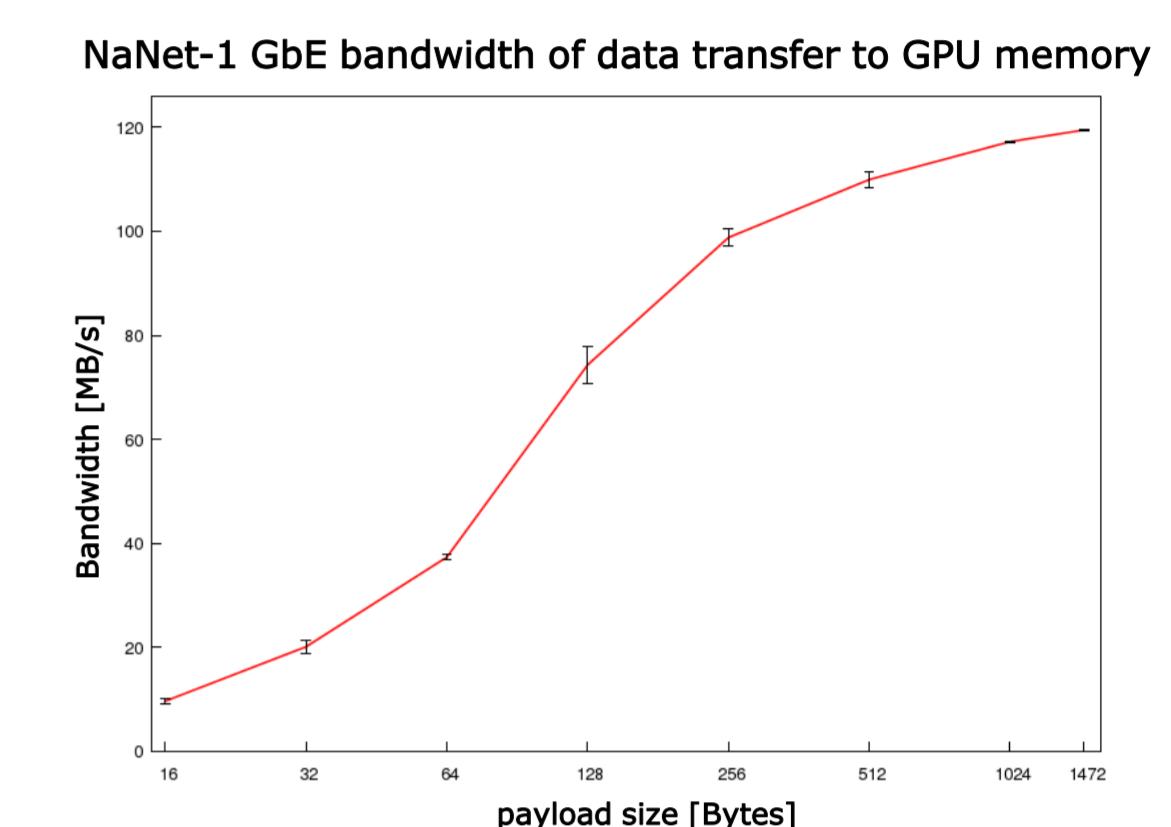
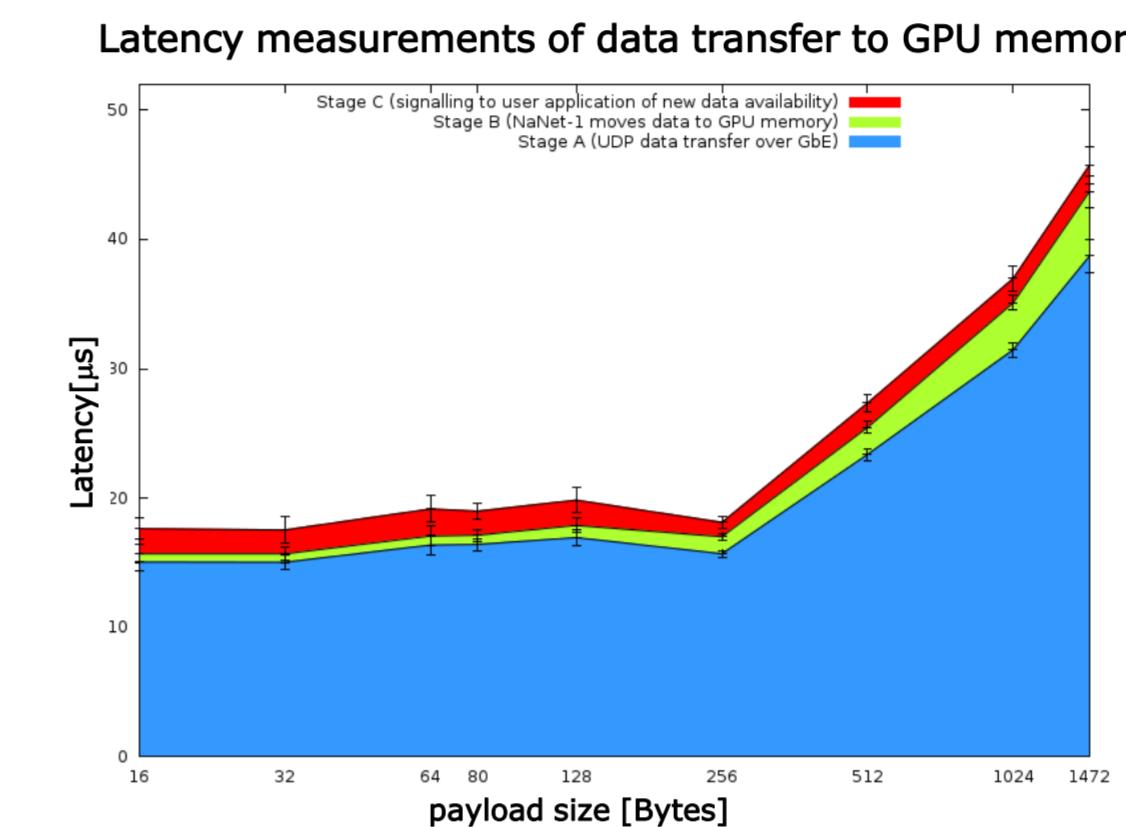


NaNet-1

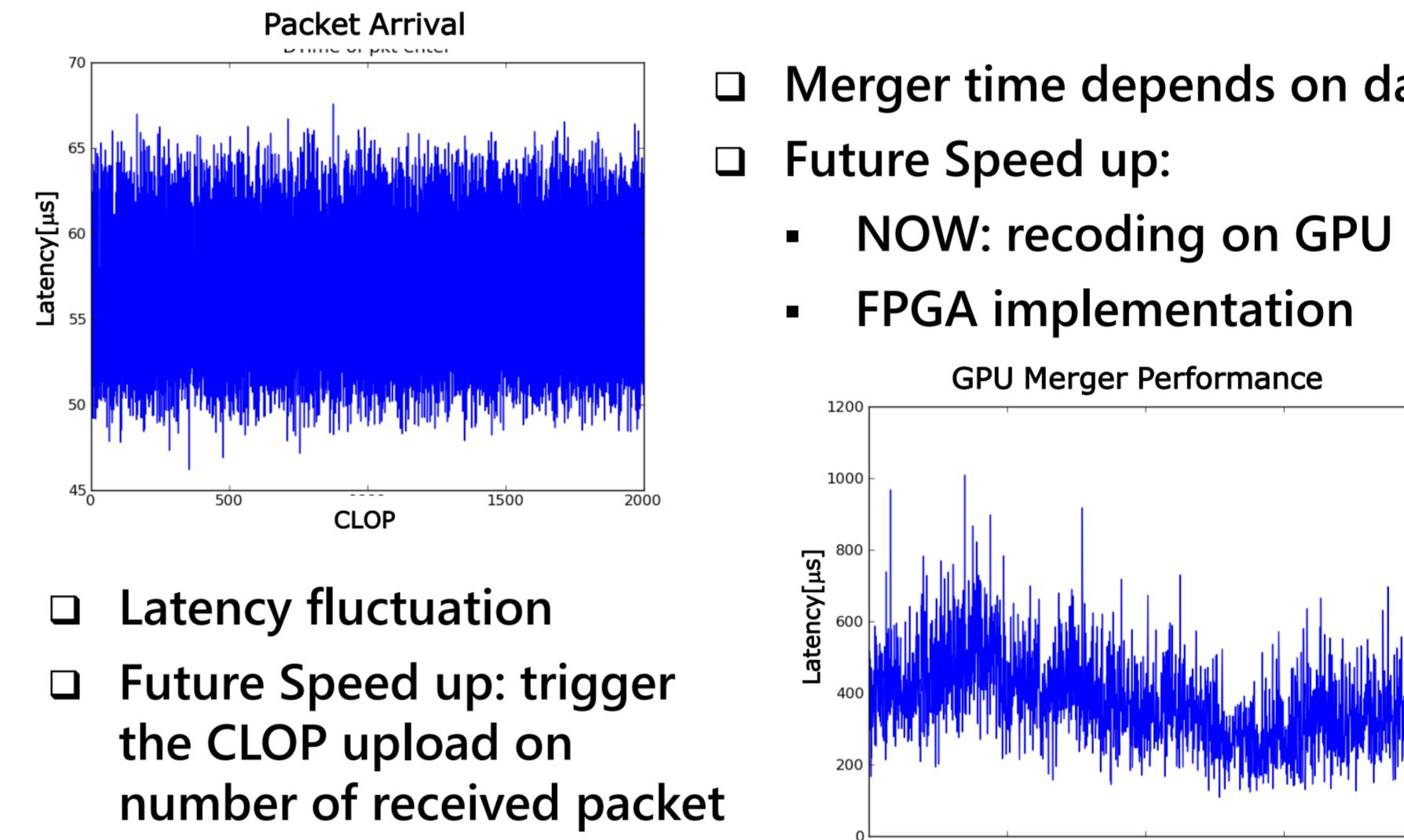
- Implemented on Altera Stratix IV dev board (EP4SGX230KF40C2)
- 1GbE PHY Marvell 88E1111
- TTC daughtercard with HSMC connector for timing (clock, SOB/EOB) and trigger signals
- Supports three additional APElink channels (20 Gb/s each) with HSMC daughtercard



NaNet-1 benchmark



NaNet-1 in RICH low level trigger processor

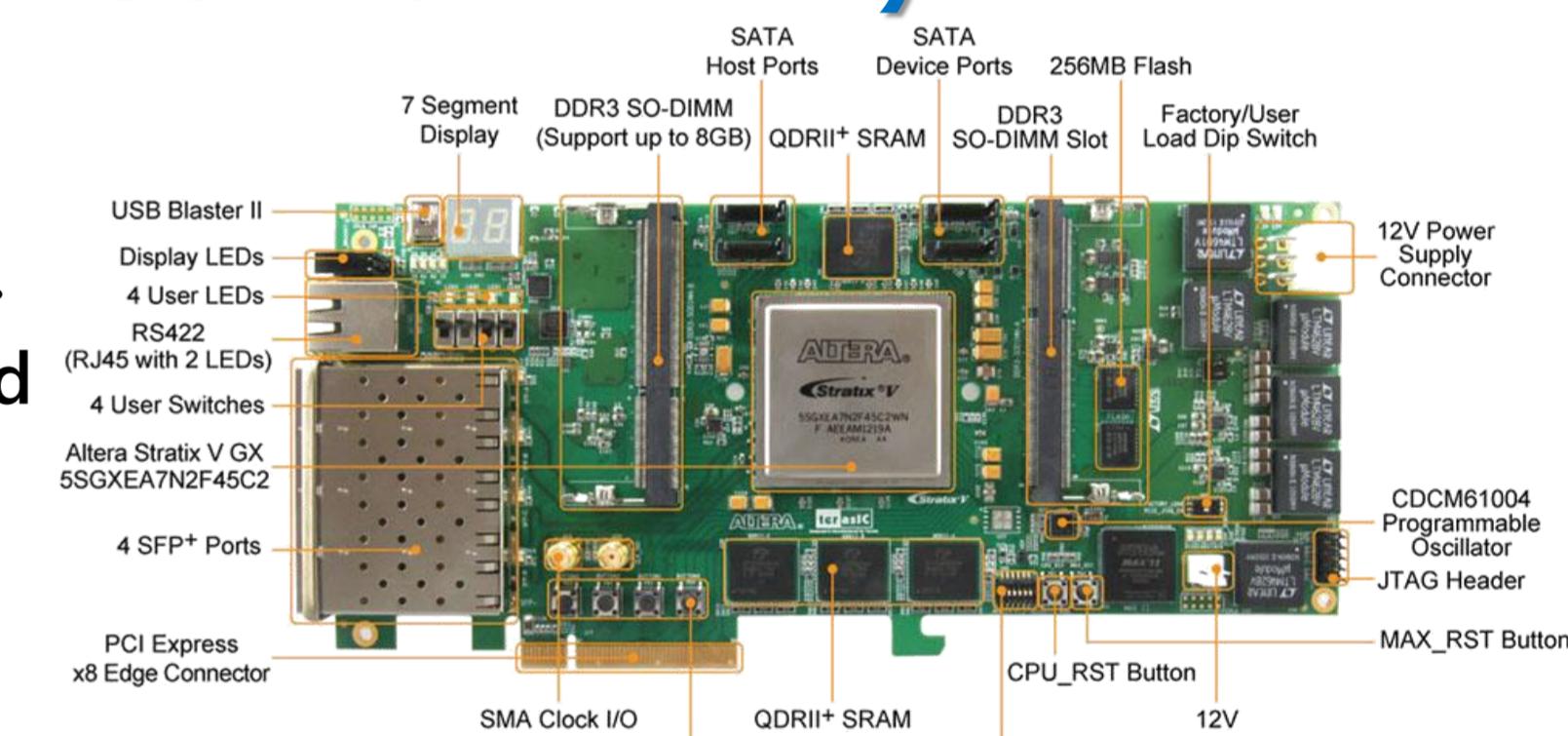


- Merger time depends on data size.
- Future Speed up:
 - NOW: recoding on GPU
 - FPGA implementation
- Latency fluctuation
- Future Speed up: trigger the CLOP upload on number of received packet

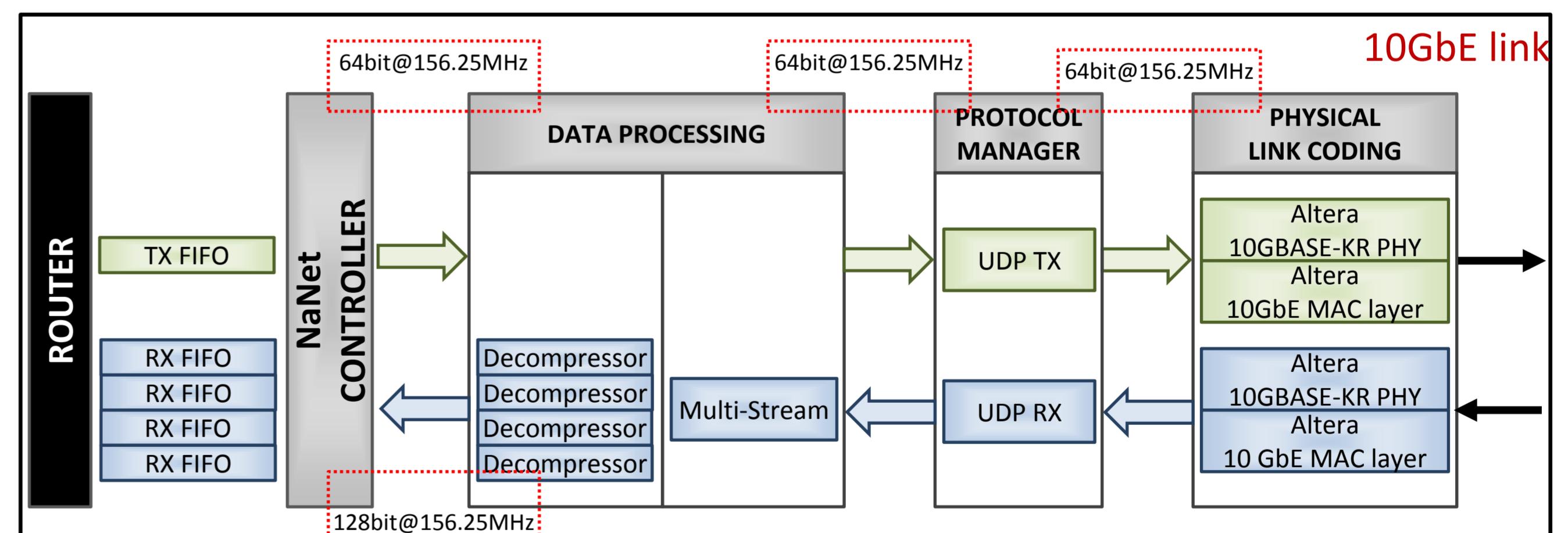
- Computing time (C2070):
 - OLD: 80 μs per event
 - NEW: 1 μs per event

NaNet-10 (four 10GbE SFP+ Ports)

- ALTERA Stratix V dev board.
- PCIe x8 Gen3 (8 GB/s).
- 4 SFP+ ports (Link speed up to 10Gb/s).
- Implemented on Terasic DE5-NET board
- GPUDirect P2P/RDMA capability
- UDP offloads supports
- Available 4Q2015.
- Planned 40GbE development.



NaNet-10 link



- A fully UDP/IP 10GbE Link (IEEE 802.3-2005 compliant) implemented in hardware.
 - Altera 10GBASE-KR PHY block; Altera 10 GbE MAC layer
 - 10 Gbit UDP/IP Core adapted from 1 Gbit UDP/IP core at www.opencores.org
 - AXI-lite data interface working @ 156.25MHz
 - Fully customizable IP and MAC address
 - ARP level functionalities. 256 entry cache for IP-to-MAC address translation
 - PHY block tested with an optical cable 3m long
- NaNet Transmission Control Logic
 - TX path: APEnet/AXI/UDP protocol translation
 - RX path: UDP/AXI/APEnet protocol translation. Virtual Address Generation. Data flow Manager.

NaNet-10 benchmark

