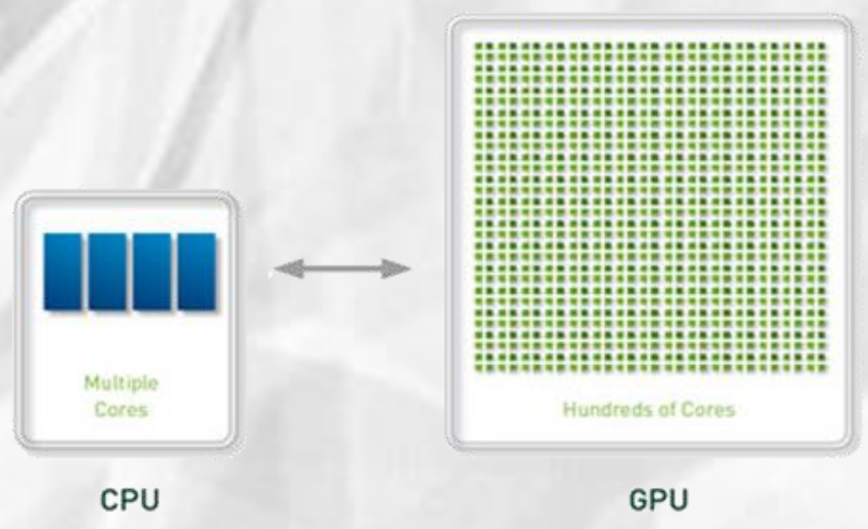
# GPU for Real Time processing in HEP trigger system

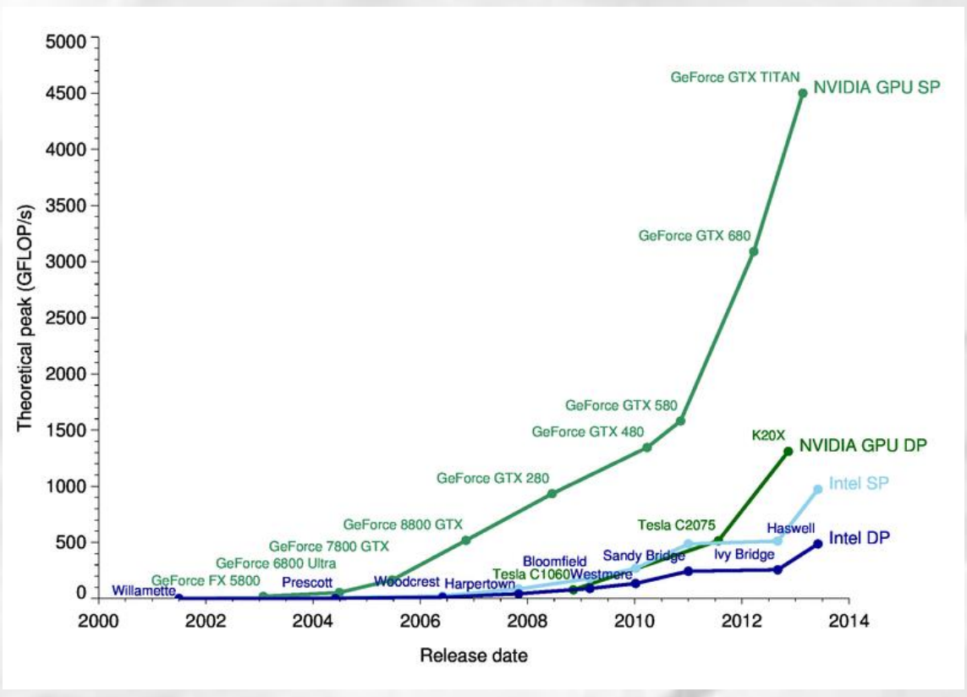
R.Ammendola [1], M.Bauce [2], A.Biagioni [3],S.Chiozzi [9], R.Fantechi [4], M.Fiorini [5], S.Giagu [2], A.Gianoli [9],E.Graverini [7], G.Lamanna [8], A.Lonardo [3], A. Messina [6], F.Pantaleo [7], R.Piandani [8] , M.Rescigno [3], F.Simula [3], M.Sozzi [7], P.Vicini [3]

(1) INFN sez.Roma-TorVergata, (2) University of Rome and INFN, (3) INFN sez. Roma-Sapienza, (4) INFN sez.Pisa and CERN, (5) University of Ferrara and INFN, (6) University of Rome and CERN, (7) University of Pisa and INFN, (8) INFN sez.Pisa, (9) INFN sez. Ferrara

## The GPU

- The use of **GPU** (Graphics processing unit) for High Performance Computing is rising in the last years.
- The main differences between **GPU** and **CPU** are due to the different resources dedicated to computing and to the parallel architecture.
- Nowadays a single GPU can deliver more than **3 TFLOPS**.
- Vectorizable algorithms could benefit from the GPU computing power.
- GPUs are to be intended as a co-processor: the data must be brought on the video card using the PCI Express bus.

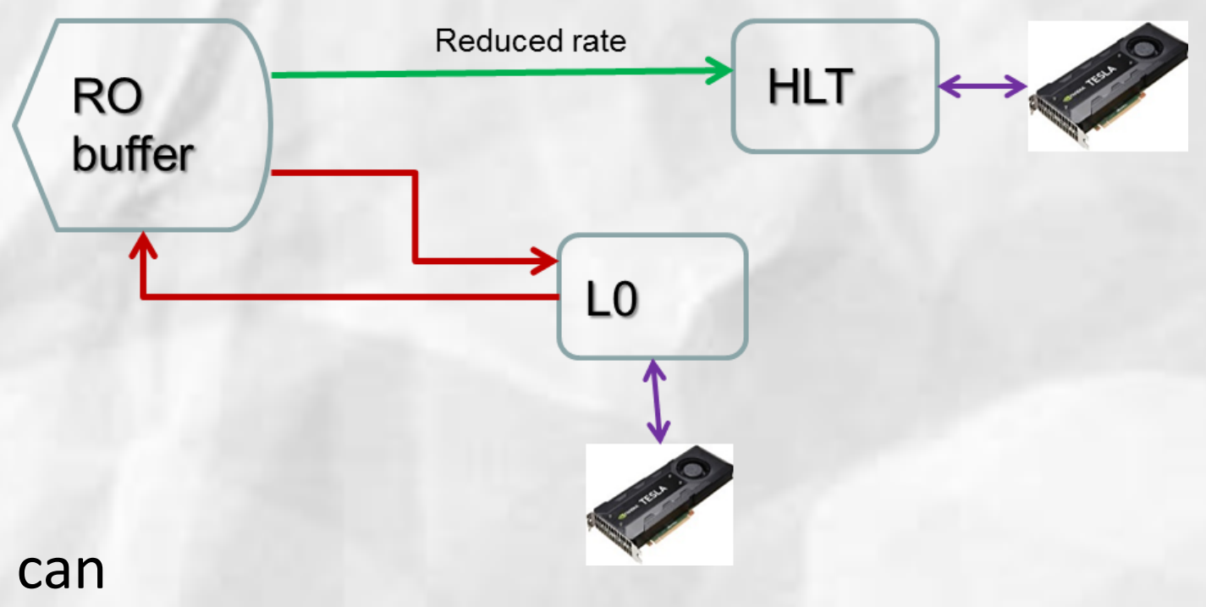




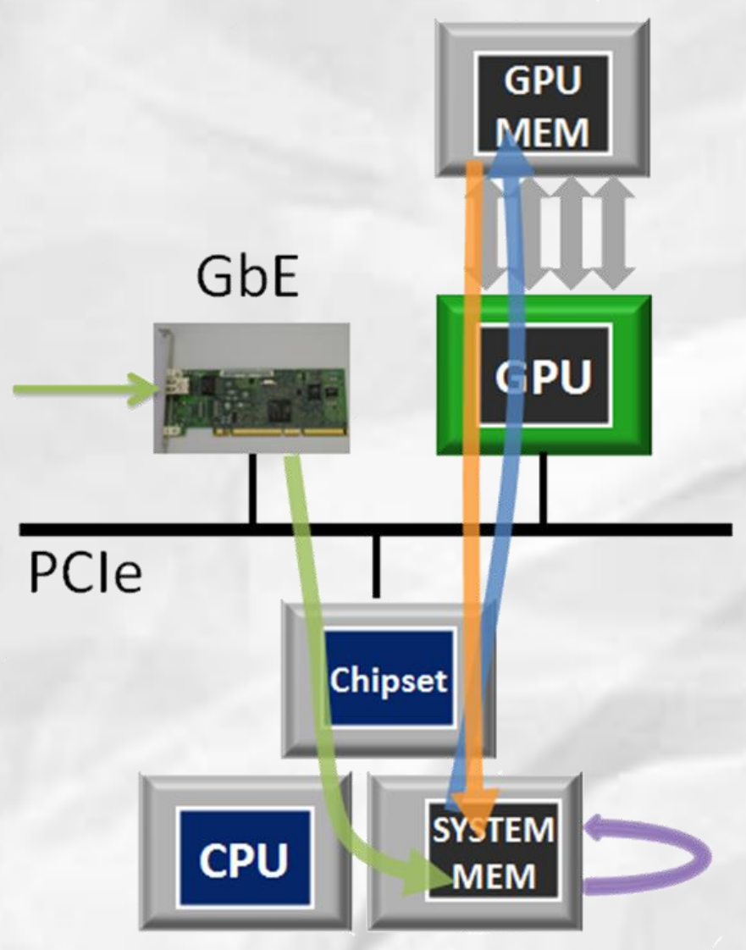
| Video Card | N. Cores | Processing Power (GFLOPS) | Bandwidth (GB/s) |
|---|---|---|---|
| NVIDIA TESLA C1060 (2009) | 240 | 933 | 102.4 |
| NVIDIA TESLA C2050 (2011) | 448 | 1288 | 144 |
| NVIDIA GTX680 (2012) | 1536 | 3090 | 192 |
| NVIDIA TESLA K20 (2013) | 2496 | 3520 | 208 |
| AMD RADEON 5870 (2010) | 1600 | 2720 | 153.6 |
| AMD RADEON 7970 (2012) | 2048 | 4096 | 288 |

## GPU in the trigger

- The **GPU** can be employed to build high selective triggers when the limit of the standard approach is due to the computing power.
- Usually the triggers are subdivided in "levels" in order to online select the interesting events:
  - The **lower level** is realized with custom electronics to apply a fast and rough selection, with low latency.
  - The **High Level Trigger (HLT)** is realized in Software allowing more precise decisions in a longer time.



- In **HLT** the use of **GPU** is quite straightforward: the video processors can increase the computing power of the online farm reducing the total number of PCs where the trigger algorithms run.
- The use of GPU in low level trigger is more difficult: the total time to have an answer from a system based on **PC+GPU** and standard links (i.e. Ethernet) is given both from the computing time and the time to bring the data in the GPU.
- For a system based on Ethernet most of the time is spent on data transfer instead of computing on **GPU** (example: packet of 1404 B with an algorithm to reconstruct ring in a Cherenkov counter, see below).
- In order to reduce the data transport latency we are investigating two ways:
  - **PFRING-DNA** Driver: a driver for fast packet capture.
  - **NANET**: direct transport of data from the **NIC** to the **GPU** without **CPU** involvement.
- In both cases a part of the latency, since we are interested in real-time systems, the fluctuations of the latency must be studied and reduced.

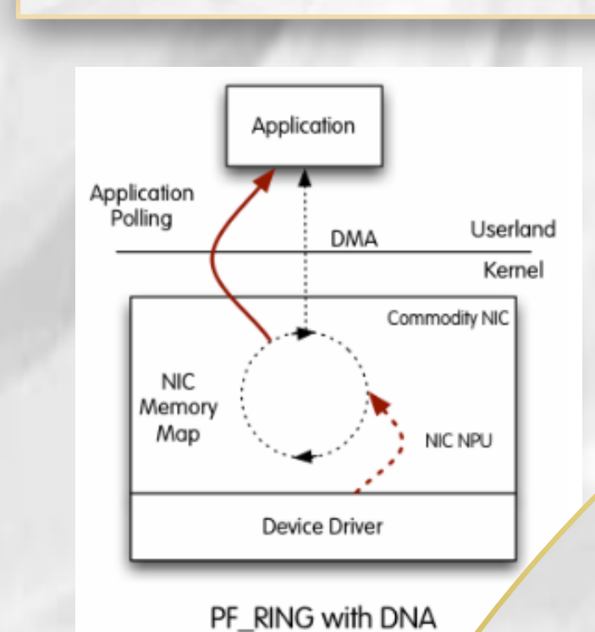


## GAP

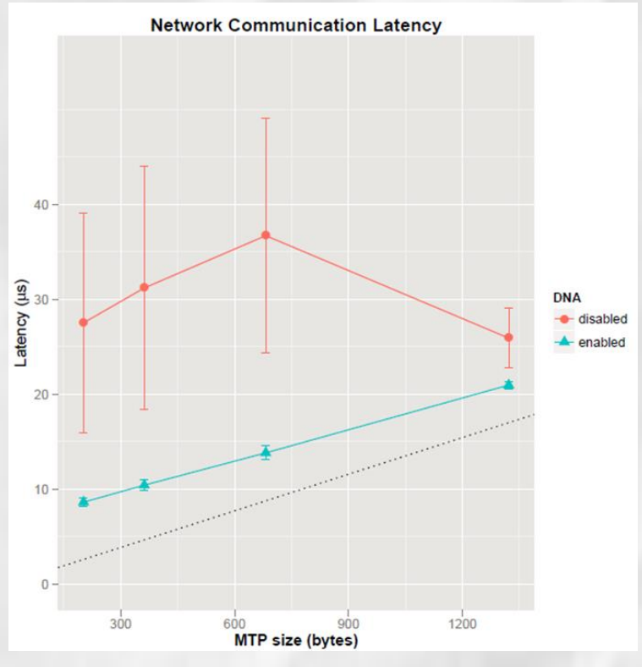
- The **GAP project** (GPU application for physics) aims at studying the use of **GPU** in real-time application.
- The main fields of study are trigger in High Energy Physics experiments and image reconstruction for medical purposes (PET, TAC and NMR).
- Three research units: INFN (Pisa and Rome Apenet+ Group), University of Ferrara and University of Rome.
- More info: **http://web2.infn.it/gap**

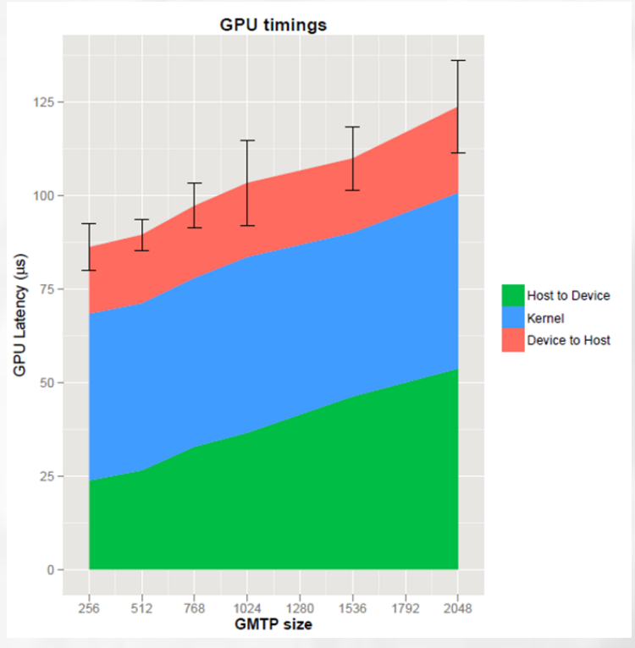
## PFRING

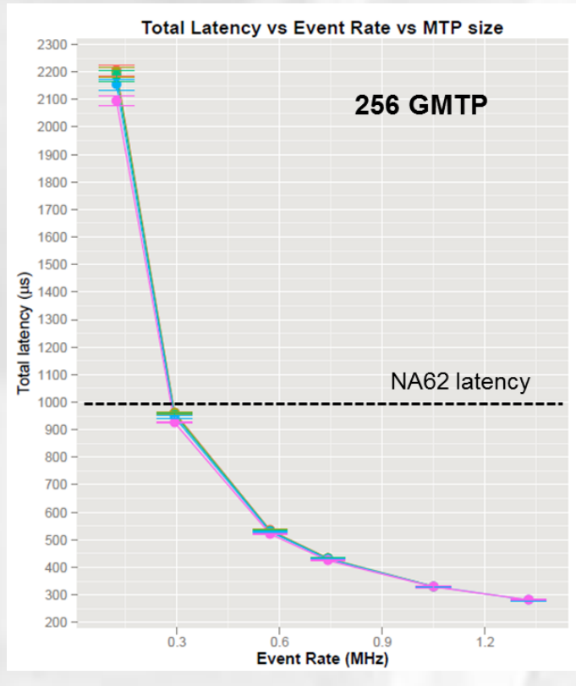
- **PFRING-DNA** is a special socket–driver, developed by **NTOP** (http://www.ntop.org) that allow to directly copy data from **NIC's FIFO** to the user memory.
- A prototype system with a readout board (**TEL62**) and a PC equipped with an **NVIDIA TESLA K20** has been used to measure latency and computing time.
- The latency has been measured with an oscilloscope by using the start of the packet in the **TEL62** as "start" and the "computation done" in the PC as "stop".



PF_RING with DNA (Direct NIC Access) drive

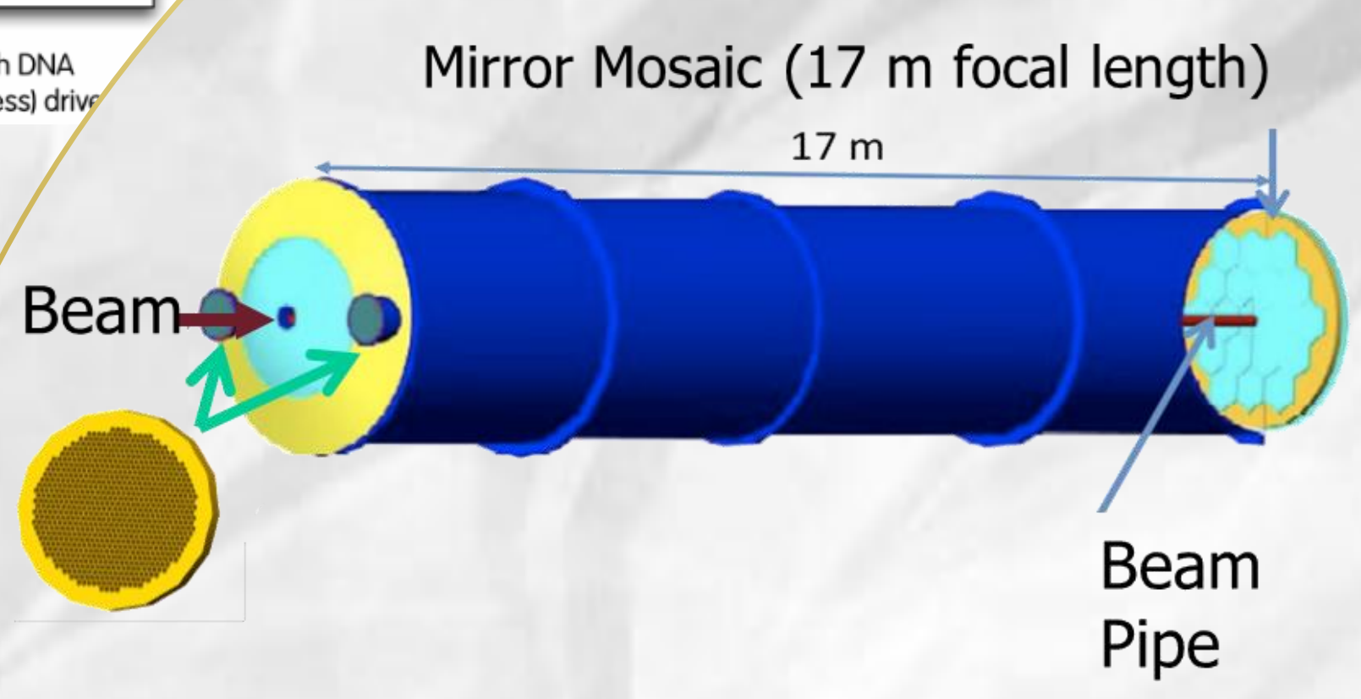




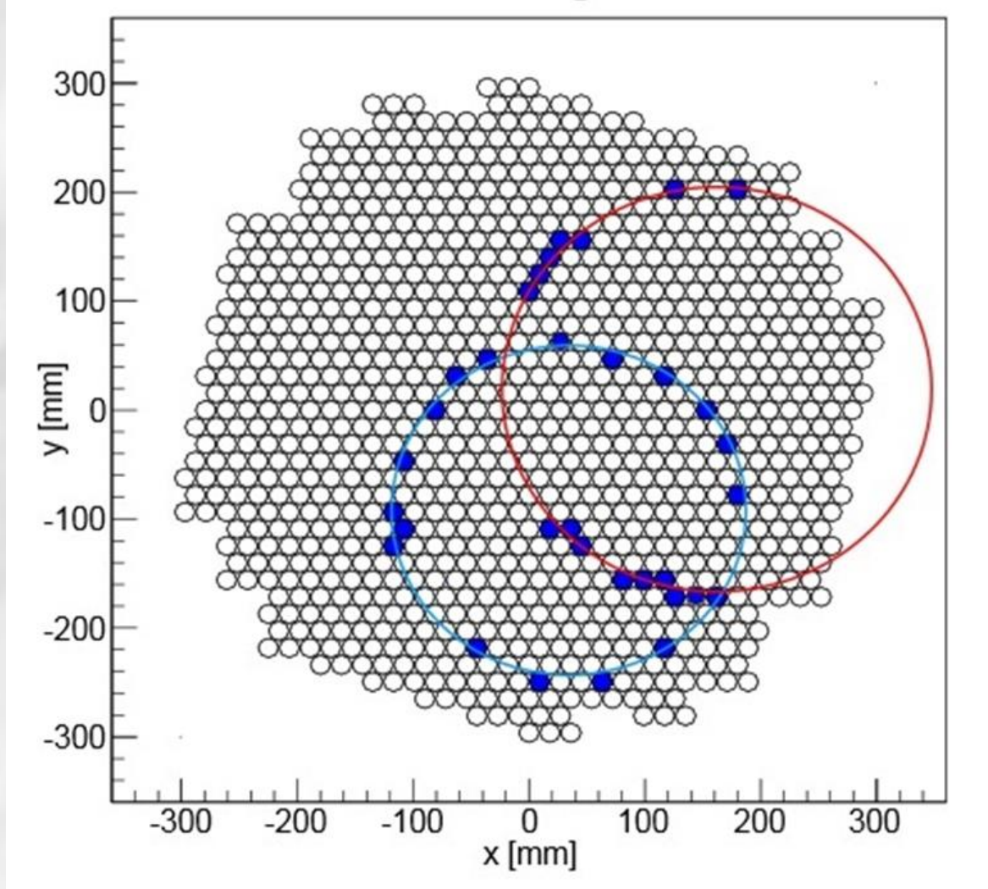


- **PFRING-DNA** allows to reduce the latency by a factor 3 and the fluctuations to a negligible level .
- The total latency is given as a function of the number of events to buffer before the start of the **GPU** computation.
- For real application the *"working point"* depends on the events rate and event dimension: for real applications the total latency (transfer time through ethernet+transfer to GPU+computing) is in the order of 100/200 us.

## NA62 physics case



- The **NA62 RICH** is a Neon (1 atm) Cherenkov detector to distinguish between pions and muons in the **15-35 GeV** range.
- The identification of rings at low trigger level is useful to build very selective trigger conditions.

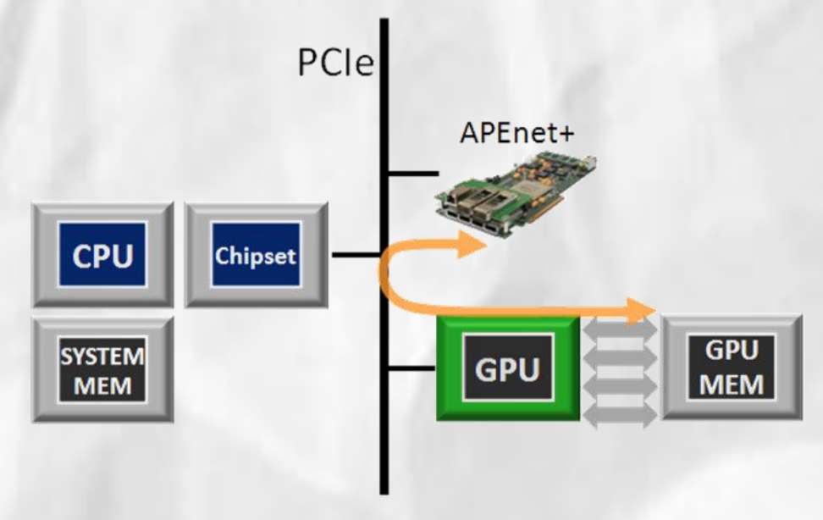


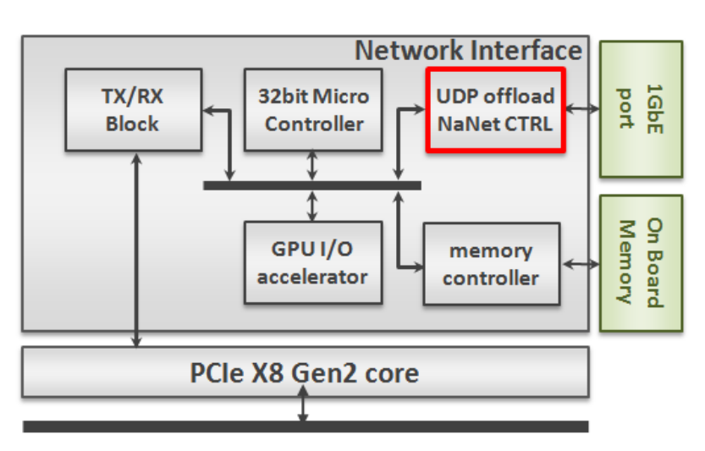
- Requirements:
  - **Fast**
  - **Offline quality**
  - **Trackless**
  - **Multi-rings**
  - **Noise tolerance**
- New parallel algorithm (called "Almagest") in two steps:
  - Pattern recognition based on parallel application of "Ptolemy's Theorem"
  - Fitting with Tobin algorithm
    - Preliminary results very encouraging:
      - ~50 ns in single ring events.
      - ~1 us in multi rings events.

## NaNet

A **FPGA** based PCIe 8x gen 2 board derived from the **apeNET+ 3D NIC** design, implementing **GPUDirect RDMA** technology over **GbE** and a **UDP** protocol offloading engine.

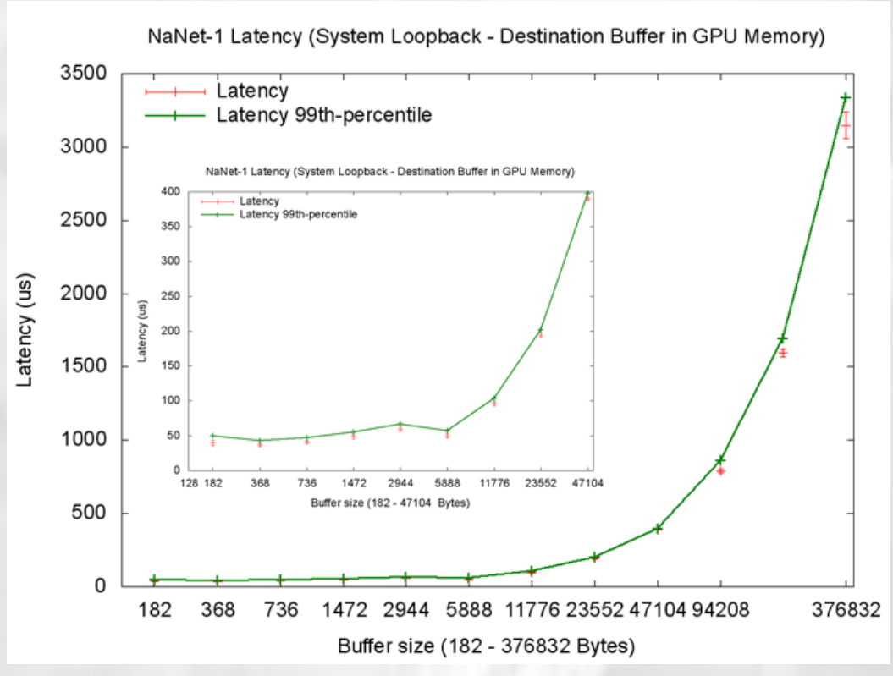
- PCIe **P2P** protocol between Nvidia Fermi/Kepler and NaNet.
  - **RDMA-style** data transfer directly from GbE or apelink into **GPU** memory w/o intermediate buffering.





- **UDP** offload: collects data coming from the GbE and redirects UDP packets into an hardware processing data path.
- **NaNet CTRL**: encapsulates the UDP payload in a newly forged APEnet+ packet.
- NaNet logic implemented on **Altera Stratix IV** development board and apeNET+ board.

- Lower and more stable **latency**, compared with **COTS NICs** and vanilla software stack (or even RTOS) approach.
- Sustained bandwidth **~119.7 MB/s**.



- For further information attend oral presentation **"NaNet: a low-latency NIC enabling GPU-based, real-time low level trigger systems"** by Alessandro Lonardo, TUE 15/10, 13:30 – 13:50, Room: Verwey Kamer.

## GPUs in HLT

- In software level triggers (**HLT**) the **GPUs** can be used to design high performance parallel algorithms for precise selection and to decrease the costs of the online farm.
- For **HLT** we are considering the **ATLAS** muon trigger as a study case for **GPU** application:
- The **ATLAS** trigger system has to cope with the very demanding conditions of the LHC experiments in terms of rate, latency, and event size.
- The increase in **LHC** luminosity and in the number of overlapping events poses new challenges to the trigger system, and new solutions have to be developed for the forthcoming upgrades (2018-2022).

## Conclusions

- The **GPU** is a specialized processor designed for fast images handling.
- The parallel **GPU's** architecture can be exploited for general purpose computation.
- The use of the **GPUs** for online applications, such as the trigger in **HEP** experiments, is very challenging.
- We are investigating two different ways to allow the use of **GPUs** in real-time: **PFRING and NANET**.
- The preliminary results on prototypes designed for the **NA62** experiment are very encouraging.
- We are investigating the use of **GPUs** in the ATLAS experiment.

## Acknowledgement

*The GAP project is supported by MIUR under grant RBFR12JF2Z ``Futuro in ricerca 2012''.*

*The NaNet project is partially supported by the EU Framework Programme 7 project EURETILE under grant number 247846.*