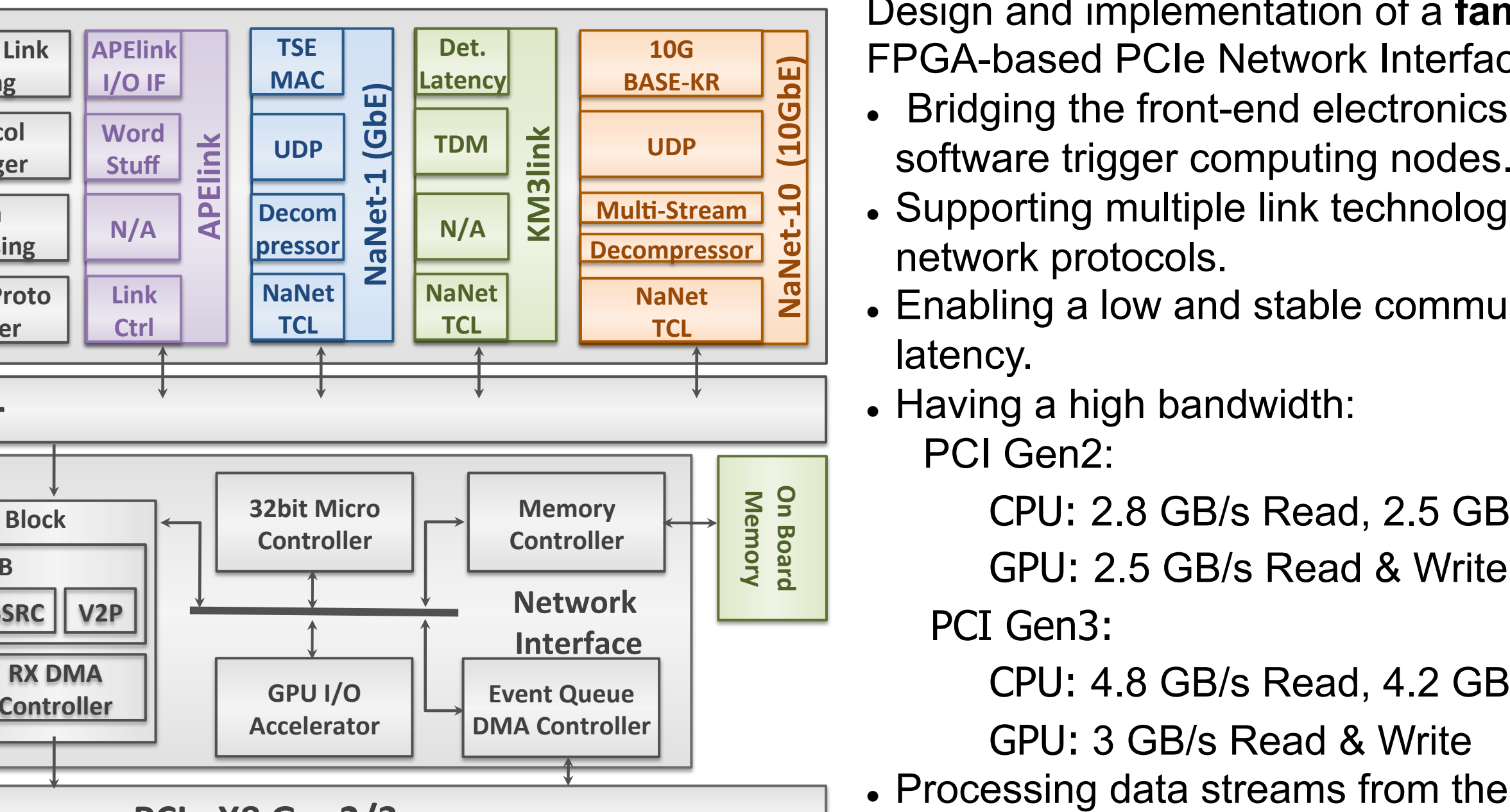


TWEPP 2017
UC Santa Cruz
September 11-15 2017

(a) INFN Sezione di Roma (b) INFN Sezione di Roma Tor Vergata (c) Università di Pisa (d) INFN Sezione di Pisa (e) NVIDIA Corporation

NaNet Design



The diagram illustrates the NaNet Design architecture, which is a multi-layered system. At the top is the **I/O Interface**, which consists of several functional blocks: Physical Link Coding, Protocol Manager, Data Processing, and APEnet Proto Encoder on the left; APElink I/O IF, Word Stuff, and Link Ctrl in the middle; TSE MAC, UDP, Decompressor, and NaNet TCL on the right; Det. Latency, TDM, N/A, and NaNet TCL in the center; and 10G BASE-KR, UDP, Multi-Stream Decompressor, and NaNet TCL on the far right. These blocks are connected to a central **Router**. Below the Router is the **PCIe X8 Gen2/3 core**, which contains a **TX Block** (with TX DMA Controller 2 and TX DMA Controller 1), a **RX Block** (with TLB, BSRC, V2P, and RX DMA Controller), a **32bit Micro Controller**, a **Memory Controller**, a **GPU I/O Accelerator**, and a **Network Interface** (with Event Queue DMA Controller). The **On Board Memory** is connected to the Memory Controller. The entire system is connected to the **PCIe X8 Gen2/3 core**.

Design and implementation of a **family of** FPGA-based PCIe Network Interface Cards :

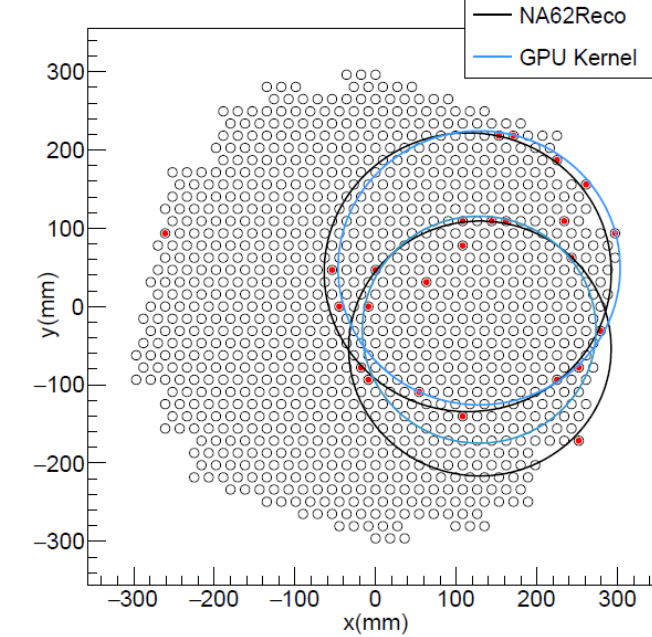
- Bridging the front-end electronics and the software trigger computing nodes.
- Supporting multiple link technologies and network protocols.
- Enabling a low and stable communication latency.
- Having a high bandwidth:
 - PCI Gen2:
 - CPU: 2.8 GB/s Read, 2.5 GB/s Write
 - GPU: 2.5 GB/s Read & Write
 - PCI Gen3:
 - CPU: 4.8 GB/s Read, 4.2 GB/s Write
 - GPU: 3 GB/s Read & Write
- Processing data streams from the network channels on the fly (data compression/decompression, re-formatting ...).

The diagram illustrates a hardware architecture connected via a vertical black line labeled "PCIe". On the left side of the PCIe line, there are two stacked components: a "CPU" (blue box) and "SYSTEM MEM" (grey box). To the right of the CPU is a "Chipset" (grey box). An orange curved arrow originates from the Chipset and points to a green circuit board labeled "NaNet | APEnet+". Below this board is a green box labeled "GPU". A series of four horizontal white arrows point from the GPU to a grey box labeled "GPU MEM" on the far right.

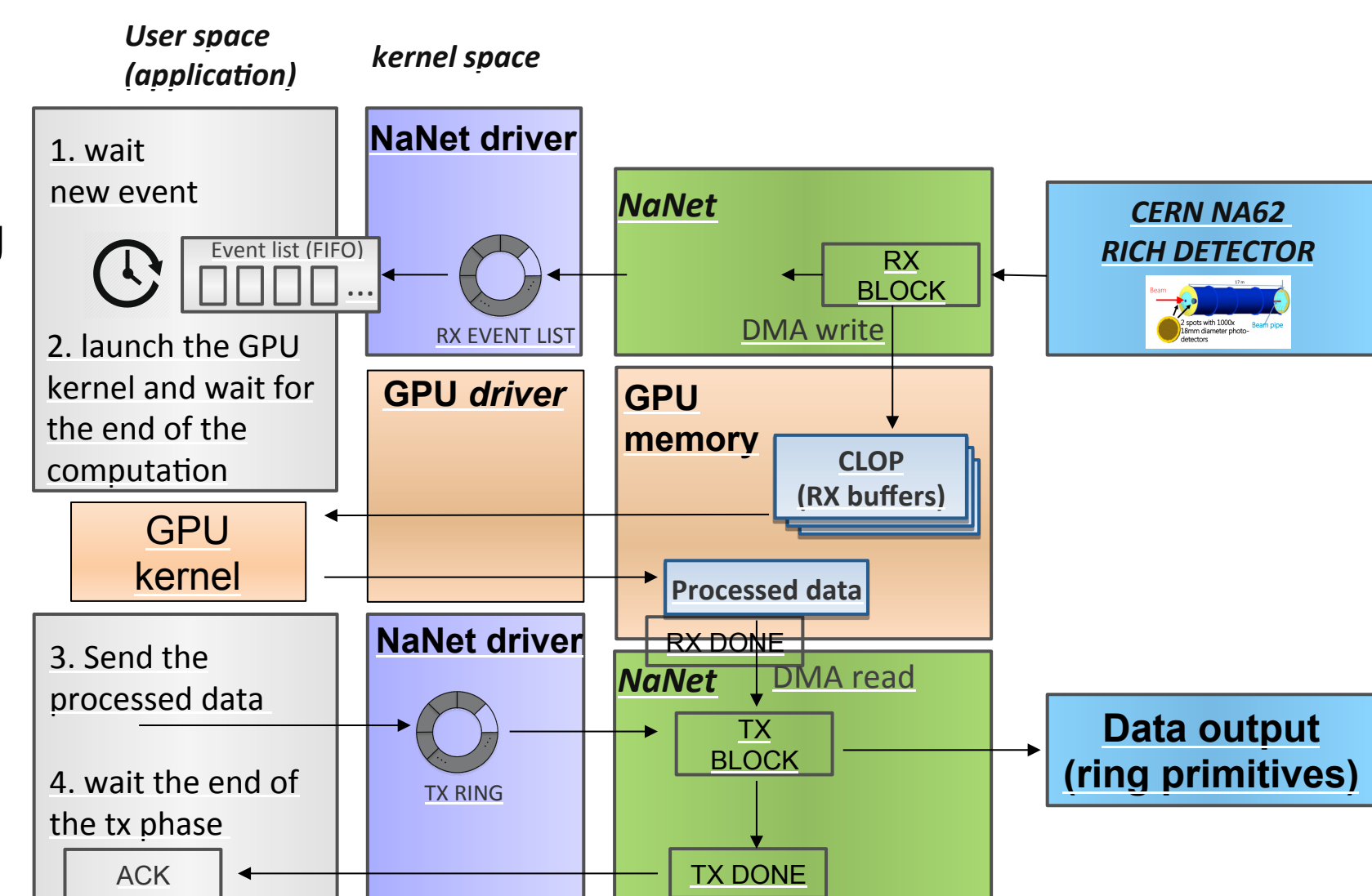
-

-

- Specific histogram-based algorithm developed for trackless, fast, and high resolution ring fitting
- Detection of particle speed (radius) and direction (center)



- NaNet NIC DMA-writes a “receiving done” event in a memory region called “event queue”
- trapped by a kernel-space device driver
- notified to the user application which launches a CUDA kernel to process the data using the GPU

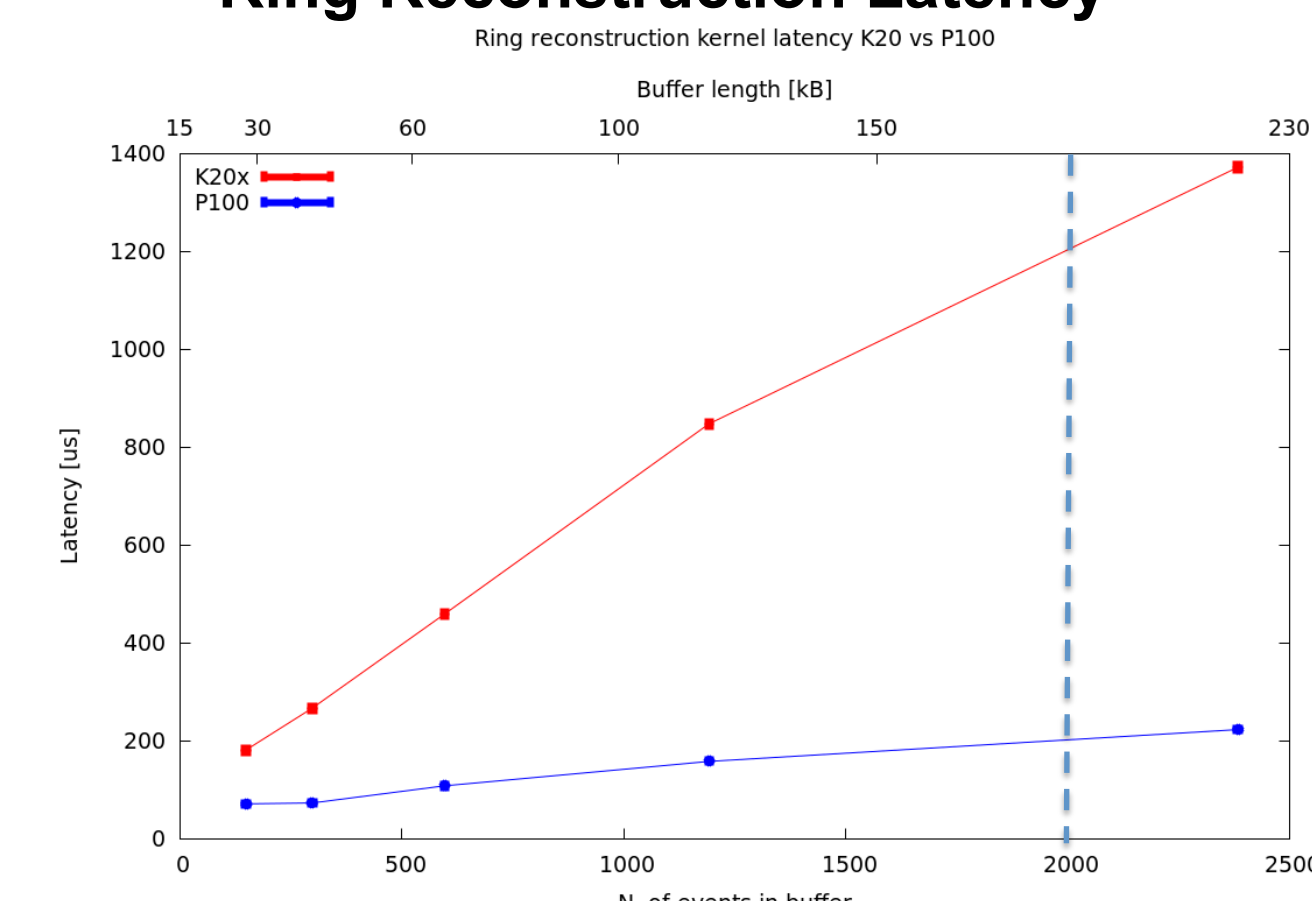
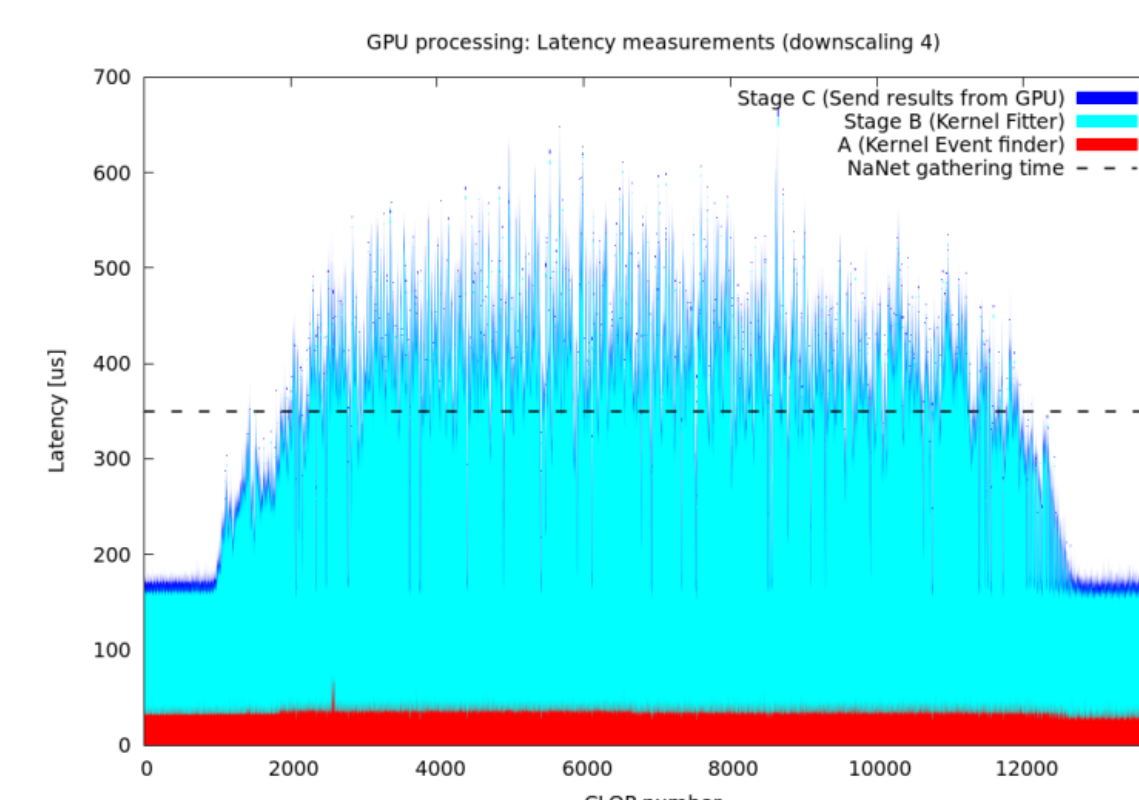


- Results of the processing – i.e. number and kind (electron, pion, muon) of rings – is eventually sent via NaNet board to the trigger system that collects data from all detectors:
 - data are DMA-read directly from GPU memory;
 - the kernel device driver (invoked by the user application on HOST) instructs the NIC by filling a “descriptor” into a dedicated, DMA-accessible memory region called “TX ring”;
 - the presence of new descriptors is notified to NaNet by writing on a doorbell register over PCIe;
 - NaNet NIC issues a “tx done” completion event in the “event queue”.

-

- ❑ The latency of event indexing in the GPU CLOP buffer is ~ constant (red)
- ❑ The latency of the ring reconstruction GPU kernel increases with the number of events in the CLOP buffer (cyan)
- ❑ Real-time stream processing: processing time \leq events gathering time
- ❑ Reaching a downscaling factor of 4 (i.e. processing only every fourth event), the latency of GPU processing exceeds the events gathering time:

we need to speed-up the ring reconstruction!



NVIDIA Pascal P100 is a potential solution.