Heterogeneous computing with the NaNet framework

R. Ammendola(a), A. Biagioli(b), P. Cretaro(b), F. Lo Cicero(b), A. Lonardo(d), M. Martineilli(d), P.S. Paolucci(c), E. Pastorelli(c), L. Pontisso(c), D. Rossetti(c), F. Simula(d), P. Vicinii(d)

(a) INFN Sezione di Roma (b) INFN Sezione di Tor Vergata (c) NVIDIA Corporation

The end of the scaling of CPU operating frequencies led at first to multicore architectures and then to heterogeneous ones. GPUs clusters are the most common nowadays but FPGAs are increasing their share thanks to their communication capabilities. To enable an orchestrated combination of heterogeneous computing devices (CPUs, FPGAs and GPUs) we devised NaNet, a FPGA-based PCH-Express Network Interface Card with processing and GPU-direct capabilities, which supports multiple link technologies (110/1000E and custom ones). We have demonstrated the effectiveness of the method in the NA62 low-level trigger, implementing a real-time stream processing pipeline to generate refined physics-related primitives of the RICH detector.

NaNet Design

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:
  - CPU: 2.8 Gb/s Read, 2.5 Gb/s Write
  - CPU: 2.6 Gb/s Read & Write
- Processing data streams from the network channels on the fly (data compression/decompression, re-formatting)

GPDirect P2P/RDMA

- GPDirect allows direct data exchange on the PCIe bus with no CPU involvement.
- No bounce buffers on host memory.
- Zero copy IO.
- Latency reduction for small messages.
- NVIDIA Fermi/Kepler/Maxwell/Pascal

Support for Intel SDK for OpenCL

- Customization of vendors Board Support Packages (BSP).
- Available for Terasic DE5-NET and Bittware 55-
  - PCIe-HQ boards
- Includes just essential IP cores (UDP over 10GbE,
  - PCie and DDR memory controller) to make more room for user kernels
- Work in progress to support the new Intel Stratix 10
  - GX development board (more logic, DSPs and memory available)
- Allow the user to describe hardware through a C-like language, with the capability to leverage NaNet core

This framework has also been applied for real-time processing of Forex market data through the NaNet-T design.

NaNet-10

- Altera Stratix V Terasic DE5-NET dev board
- 4 SFP ports (Link speed up to 10 Gb/s)
- PCIe X8 Gen2/3
- GPDirect P2P/RDMA capability
- UDP offload support
- Processing stage for data decompression and events merging

Physics Case: Real-time Rings Reconstruction for the NA62 RICH Detector

The NA62 experiment at CERN aims at measuring the branching ratio of the ultra-rare decay: K⁺ → π⁺ννπ⁻K⁻ (BR =8x10⁻11)

Ring-imaging Cherenkov (RICH) detector is used: the particles generate a circular footprint radiation beam onto the light-sensitive tubes of two photomultipliers (PMT) arrays. In the standard implementation, the FPGAs on the TE620 readout boards compute simple “trigger primitives” on the fly, such as hit multiplicities, which are then sent to a central processor for matching and trigger decision, with a time budget of 800 μs.

We then added a GPU-based processing stage between the RICH detector readout and the L0 trigger processor (L0TP) with the task of generating, in real-time, physics-related primitives (i.e. centers and radii of Cherenkov rings patterns on the photomultipliers arrays), in order to boost the L0 trigger discrimination power. Results can be also sent to the High Level Trigger (L2TP) saving computing time to server farm.

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

Output synchronization

- L0TP allows a maximum jitter of 3x6.4us=19.2us for primitives timestamps inside MTP packets
- GPU alone can’t cope with such a strict requirement, therefore an hardware synchronization stage is required
- Primatives from GPU reconstruction buffer are collected and reshaped in chunks of 6.4us time span
- After a configurable delay (larger than the GPU reconstruction latency): these chunks are sent to L0TP at the right time based on their timestamp

Preliminary Results from 2018 Run

- Tested: Supermicro X9DRG-QF Intel C602 Patsburg
  - Intel Xeon E5-2602 2.0 GHz
- NaNet-10 (Terasic DE5-Net)
- nVIDIA P100
- Gathering time: 350us
- ~ 60% target beam intensity (~2x10^11 Pps)
- Histogram algorithm
- System running in parasitic mode

Contacts:
The NaNet project: http://people.cern.ch/lmertonet
Presenter Contact: paolo.cretaro@roma1.infn.it
NaNet project coordinator: alessandro.lonardo@cern.ch

NaNet is an INFN Scientific Committee E started experiment
NaNet is a collaborative research between INFN, EPFL, and the Group Research and Development department of Unicredit s.p.A.
This work was carried out within the EU‐NaMID project, under grant agreement EU‐2013TI‐1261755.