

An analog processor for real time data filtering in large detectors

Monday, 22 May 2017 16:54 (18 minutes)

A decision making process requires to evaluate the saliency of data in a time scale short enough for the decision, to be useful in the ecosystem that generated the data. Experimental High Energy Physics pioneered in facing the problem of managing smartly and in real time big data, produced by detectors in the ns scale, and was always on the cutting edge in developing fast and complex electronic trigger systems exploiting the expected data model to perform the selection. Very large volume experiment searching for rare events such as DUNE (Deep Underground Neutrino Experiment) may produce an extremely high data flow, with a very reduced possibility of setting up an effective trigger, in particular when searching for cosmological events typically having a faint signature. Removing this bottleneck is a crucial challenge to extend the discovery potential of such experiments. We propose to overcome this limitation by introducing a novel technology, the WRM (Weighting Resistive Matrix) to perform a topological data driven selection. The WRM technique was originally invented as a fast topological trigger for hadron colliders experiment, and recently implemented as a fast engine for demanding computer vision applications. By treating DUNE data as projected grays-scale image we can exploit the WRM technology to provide a fast data driven trigger-less selection allowing a smart noise suppression on raw data in real time.

Summary

Summary

The need of HEP future experiments of searching for rare events, of increasing the measurement precision through statistics and the lack of clear theoretical framework assessing a sharp selection data model for the new physics searches, implies at the same time to increase the event production rate and to implement a smart and yet inclusive selection on row data. As a consequence, the close-to-the-sensor need of computing power grows fast with and quickly become a real bottleneck for further expanding the experiment discovery potential.

A frontier is represented by very large volume experiments searching for rare interactions, such as DUNE, which is aimed not only to the detailed study of neutrinos produced at Fermilab, but being at the same time an extremely sensitive probe to investigate the physics of distant cosmological events such as supernovae and black hole fusion, which recently exploded in relevance after the discovery of gravitational waves. A possible problem arises from the fact that catching this type of events in efficiently and reliably requires a data driven trigger relying on event topology, since the event fragment signatures are too close to the noise.

The critical features of the detectors are:

- Large detector volume $\sim 2500000 \text{ m}^3$ containing $\sim 10^5$ individual readout channels
- Each channel is digitized at 12 bits with 2 MHz sampling rate $\Rightarrow \sim 0.6 \text{ TByte/s}$ in total
- Low local signal to noise ratio the noise band being overlapped to the signal
- Scintillation based trigger could be not sufficiently selective and efficient
- very different event shapes and directions and no detector symmetry exploitable

The capacity of managing and selecting data in real time with a large data throughput is therefore a critical element and presently the requirements are beyond what it can be practically possible (performed offline moment). To achieve the real time data processing speed, we propose to use an innovative technology: Weighting Resistive Matrix (WRM), conceived for discriminating vertices in real time in HEP experiments. It relies on an analog computing architecture to implement the equivalent of a probabilistic regression. It is able to extract directly from data the most likely fit parameter values, using the energy of the input signal to execute the computation in a single clock cycle independently from the input.

The global idea of the WRM is to calculate arbitrary pattern fit at the nanosecond scale with least square fit effectiveness, by means of resistive network. The WRM chip naturally diffuses voltage levels through its resistor following a Laplace distribution, which recalls a convolution that enforces signal to noise ratio. Signals are then read by means of a specific read out connection (we call them roads) between the WRM nodes and

patterns are then detected on the convoluted data using the best fit. A fundamental feature of the WRM is that the data are evaluated by projecting them along the possible roads, the outcome being a likelihood set proportional to the norm between data and roads. WRM has been already adapted as a hardware accelerator for a fast linear segment detector in a 4-year Marie Curie ITN funded project for advanced Virtual Reality VR and Augmented Reality (AR) Safety Systems for Maintenance in Extreme Environments (EDUSAFE).

In analogy to the real time digital image analysis, detector data can be treated as a normal grayscale images. Under noisy conditions, very often signal level is close to the noise level, so a simple threshold based method is definitely not an option since it will suppress both signal and noise. Since the noise is normally uncorrelated and signal is, our idea is to take advantage of the WRM chip to exalt only correlated data, in order to clearly separate it from noise, allowing this way to a threshold based method to be later applied.

State of the art and progress beyond the state of the art

The present WRM chip is based on an outdated technology but a single chip it is still capable of analysing 400 millions of digital patches/s on an 8x8 pixel matrix (equivalent to about 25 Gbit/s). The idea is to redesign the WRM fully exploiting the analog potential of the WRM concept for feeding directly the analog data; to enlarge the matrix acceptance, maximizing the S/N and speed (~30x30 matrix), to increase the chip speed to the several GHz scale; to maximize the chip internal parallelism. We target to produce a single chip capable of processing up to ~ TByte/s, abundantly sufficient process the large neutrino experiments data even including redundancy.

The solution we prospect would be at the same time a detector primary trigger and a smart zero suppression, allowing at the same time to detect and timestamp the appearance of an event fragment and to save on the disk just the smallest possible image clip including the fragment, achieving a compression factor which can easily exceed 10^6 , depending on the event occupancy.

Potential and Expected impact

If today it is said that software is eating the world (<http://www.wsj.com/articles/SB10001424053111903480904576512250915629460>), examples start to appear of hardware innovation that aims at optimizing, or even taking over tasks of software again, to breakthrough bottlenecks of performances (<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/CNN20Whitepaper.pdf>), or energy footprints (<http://static.googleusercontent.com/media/research.google.com/>). This project proposes to pursue high-performances in a high-throughput data workflow by exploiting a novel hardware architecture performing a very efficient analog computation, that performs the designed task non-algorithmically (i.e. heuristically), and by establishing a pipeline of continuing development and deployment based on FPGAs, to guarantee future flexibility and dynamic adaptation to the project.

The project is a test bed for an emerging computation paradigm, and has the potential to contribute significantly to the future of the high throughput data handling both in high-energy physics, and in industries (UAV, automotive, robotic/industrial controls, ...). In the first place, by designing and documenting its pipeline of development, it will lower the entry costs for teams interested in exploring analogous approaches to other problems. Most importantly, thanks to the liaison by CERN-Openlab and to the collaborations established by the members of the consortium, the project will contribute to a conversation between frontier research and industry, to exchange best practices, and to explore together effective and sustainable models of transfer to the market.

Primary author: AIELLI, Giulio (U)

Co-authors: Dr ABDALLAH, Ali (University and INFN of Roma Tor Vergata); Prof. CARDARELLI, Roberto (University and INFN of Roma Tor Vergata)

Presenter: AIELLI, Giulio (U)

Session Classification: R3-Backend readout structures and embedded systems

Track Classification: Backend readout structures and embedded systems