Development in DAQ and Triggering

Jinlong Zhang
Argonne National Laboratory

TIPP’17, May 22-26, 2017, Beijing
Disclaimer

• This may not be the presentation you originally expected, not even the conference originally planned
• The theme is LHC experiment abundant
• Some statements are my personal bias
• I was not able to cite the references properly for the materials included

My apology
Introduction

• Trigger/DAQ system overview

• Developing in trigger
  — Triggerless scheme
  — Specific aspects (track, global, timing)

• Developing in DAQ
  — Accessing commodity (PCIe)
  — Storage evolution

• Trends
  — Accelerator (GPU, CPU+FPGA)
  — Common platform

• Summary
Collider Experiment Examples

Event Size (byte)

High Level-1 Trigger (1 MHz)

High No. Channels High Bandwidth (1000 Gbit/s)

High Data Archives (PetaBytes)

LHCb

ATLAS

CMS

BEML

II

Global DAQ Design

~0.1M chan.
~500 COPPERs
50 R/O PCs

0(100 Pbits)

0(100 Gb/s)

~ 160 Gb/s

~ 25 Gb/s

~ 1.5 Gb/s

Event Manager

Builder Network

Computing Services

Control and Monitor

Filter Systems

Readout Systems

Connections

CMS

ATLAS
Neutrino Experiment Examples

Trigger not necessarily as complicated as collider experiments, but data throughput comparable.
Cosmologic Instrument Examples

CTA Camera

LSST 4000 3.2Gbps

Data throughput comparable to LHC experiments, or even larger
Over-simplified Requirements

- Customized ASICs to handle the detector signals (FE electronics) in the upstream of the Trigger/DAQ
- Powerful hardware (FPGA based, GPU, CPUs and/or combinations) and software algorithms to perform data reduction (trigger)
- High speed links, huge computing capacity and storage space to handle the event data (DAQ)
- Enabled by
  - Moore's Law (CPUs, also FPGAs and GPUs)
  - Link technology (transceivers, networking)
  - Storage technology
Link to Upstream

http://www.xilinx.com

<table>
<thead>
<tr>
<th>Type</th>
<th>Max Performance</th>
<th>Max Transceivers</th>
<th>Peak Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtex UltraScale+</td>
<td>GTY</td>
<td>32.75</td>
<td>128</td>
</tr>
<tr>
<td>Kintex UltraScale+</td>
<td>GTH/GTY</td>
<td>16.3/32.75</td>
<td>44/32</td>
</tr>
<tr>
<td>Virtex UltraScale</td>
<td>GTH/GTY</td>
<td>16.3/30.5</td>
<td>60/60</td>
</tr>
<tr>
<td>Kintex UltraScale</td>
<td>GTH/GTY</td>
<td>16.3/16.3</td>
<td>64</td>
</tr>
<tr>
<td>Virtex-7</td>
<td>GTX/GTH/GTZ</td>
<td>12.5/13.1/28.05</td>
<td>56/96/16</td>
</tr>
<tr>
<td>Kintex-7</td>
<td>GTX</td>
<td>12.5</td>
<td>32</td>
</tr>
<tr>
<td>Artix-7</td>
<td>GTP</td>
<td>6.6</td>
<td>16</td>
</tr>
<tr>
<td>Zynq UltraScale+</td>
<td>GTR/GTH/GTY</td>
<td>6.0/16.3/32.75</td>
<td>4/44/28</td>
</tr>
<tr>
<td>Zynq-7000</td>
<td>GTX</td>
<td>12.5</td>
<td>16</td>
</tr>
</tbody>
</table>

- Readout system will utilize these serDes speeds or faster, so
- High speed radiation hard link need be developed
  - lpGBT modest
**Link in Downstream**

- Network for hundreds of 100 GBE links not a problem soon
- PCIe Gen4 expected in later 2017
Changing Paradigms

- No trigger (triggerless) or less trigger levels
- Online Offline fusion
- Better physics performance or enhancing physics capability
- Less/common effort

Your favorite detector/instrument

- L1/HW trigger
  - Event Filter
    - Offline reco
      - Analysis
  - Event Filter
    - Offline reco
      - Analysis
- L1/HW trigger
  - Event Filter
    - Offline reco
      - Analysis
- L1/HW trigger
  - Event Filter
    - Offline reco
      - Analysis

Go Triggerless

ALICE in Run 3

3.6 TByte/s into PC farm

O² (Online Offline) System
Partial calibration and reconstruction online, replacing the original raw data with compressed data

LHCb in Run 3

PCle40
writes in RAM

EVENT BUILDER PC

~ 9000 optical links

x 500 nodes

~ 18 links

100 Gb/s

100 Gb/s

100 Gb/s

EVENT BUILDER SWITCH

EVENT FILTER FARM

Acq rate:
Pb-Pb 50 kHz
pp and p-Pb up to 200 kHz

Complete change in detector readout
• continuous
• triggered

New DAQ - HLT - OFFLINE systems.

STORAGE
90 GB/s

2015
Event size
65 KB
Event rate
1 MHz
Aggregate bandwidth
520 Gb/s

2018
100 KB
40 MHz
32 Tb/s
Triggerless Not Yet Possible

**ATLAS in Run4**

- ATLAS seeded, regional @ 1 MHz and full event @ 100 kHz
- CMS self seeded, @ 40 MHz and latency of ~4 μs

**CMS in Run 4**

- Track trigger
  - ATLAS seeded, regional @ 1 MHz and full event @ 100 kHz
  - CMS self seeded, @ 40 MHz and latency of ~4 μs
- Global Event Processing with more/finer input
- Possible to use precise timing
Track Trigger

Also others ways to perform tracking: HEP track finding with Micron Automata Processor, Artificial Retina processor, ... ...

And the buzzword: Deep learning
AM Approach

Use hardware to perform the global tracking in two steps: pattern recognition and track fit.

- Single Hit
- Road
- SuperStrip (bin)

Pattern recognition in coarse resolution (superstrip → road)

- Prestored patterns (10^9)
- Content-addressable Memory (CAM)
- Fast pattern recognition

Track fit in full resolution (hits in a road): F(x_1, x_2, x_3, ...) ≈ a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + ... = 0

- ATLAS FTK (Phase-I upgrade) with:
  - <1 billion of patterns
  - AMChip06 (~128K pattern)
  - <100 µs

- ATLAS hardware trigger in Phase-II upgrade:
  - ~10 billion of patterns
  - ~512k patterns per chip
FPGA Approach

• The reference option for CMS Phase-II Upgrade
  — Hough Transform and Kalman Filter in FPGA

Geometric Processor (GP) - pre-processes stub data, and divides the octant into 36 finer sub-sectors

Hough Transform (HT) - A highly parallelised first stage track-finder that identifies groups of stubs consistent with a track in the r-\(\phi\) plane

Kalman Filter (KF) - A candidate cleaning and precision fitting algorithm

Duplicate Removal (DR) - Uses precise fit information to remove duplicate tracks generated by the HT

Source

Detector octant 1 (right)
Detector octant 2 (left)

36 links

One box = one MP7

36 links

72 links

KF + DR

Sink

36 links

12 links

KF + DR

Track Finder Processor

36 links

72 links

Detector Octant 1 (left)
Detector Octant 1 (right)
Detector Octant 2 (left)
Detector Octant 2 (right)

N BOARDS = 8 × TMP

\(x N, \text{ where } N = \text{TM period}\)
Global Event Processing

- Data transfers are time multiplexed within the system
  - Increases flexibility
  - Simplifies evolution
  - Maximizes physics
Precise Timing

- Using precise timing information in trigger for pileup rejection
- Challenging to achieve the time resolution as a sizeable detector
- Huge data throughput (pixel detector after all)
DAQ in General

• PC-based data aggregation
  - Ethernet or InfiniBand
  - PCIe
• Network bandwidth becoming very affordable
  - Revisiting the philosophy of “move minimal amount of data”
  - Capability for high event building rate (even decouple from event filtering or other data processing)
• Heterogeneous computing resource (ASIC/FPGAs, GPGPUs, ... )
• Tight integration with offline
  - From the blur boundary to the full fusion
  - Better utilization of (online) resources
I/O Card Utilizing PCIe

ATLAS FELIX

CaRIBou as a multi-chip modular DAQ system

- COTS hardware may require tweaks to be used, but still could be game changer, particularly for small experiments
Storage Evolution

• Throughput is the real challenge

• Real world example exists with current Technology for a system with capacity of ~50PB and throughput of ~5 TB/s

• We should look at storage technologies 10 years from now

• Evolution of existing technologies
  — Consumer NAND drive getting cheaper than spinning drive
  — Lustre and GPFS

• New technologies
  — 3D XPoint

• Innovations in the storage stack
  — Seagate Kinetic, ...
ALICE TPC track reconstruction got a factor 2-3 speedup and saved 0.5M USD during Run 1

- Performance highly dependent on workload
- Could also integrate with other components (NIC) for serious data processing
- Comparison need consider hardware, power, cooling, and effort, ...
CPU + FPGA

LHCb RICH PID algorithm

• Acceleration of factor up to 35 with Intel® Xeon®-FPGA with respect to single Intel® Xeon® thread
• Theoretical limit of photon pipeline: a factor 64 for Stratix V FPGA, for Arria 10 FPGA a factor ~ 300
• Bottleneck: Data transfer bandwidth to FPGA
Common Platform

- Sharing a hardware unit with powerful FPGA(s) and high speed links
  - ATCA/xTCA, PCIe, etc
- Leaving the intelligence differences for firmware and software

State of the Art: ATLAS gFEX

- 30 layer PCB
- 3 Virtex Ultrascale+
- 1 Zynq Ultrascale+
- 35 minPODs
DES-GW program using DECam at Chile to perform optical followup of gravitational wave signals from LIGO/Virgo
Recommendations From CPAD

- Encourage the development of high-bandwidth radiation hard optical links (>10Gb/s)
- Encourage the development of scalable DAQ system to enable the transition from custom hardware to commodity networking and computing as early as possible
- Encourage the development in hybrid CPU-FPGA, GPGPU, storage, high speed optical and electrical communication
- Encourage studies of the impact of timing information in the trigger at ATLAS/CMS
- Encourage focus on emerging technologies such as photonics and wireless communication