Heterogeneous Computing with GPUs for Trigger Decision in CMS Experiment at the CERN-LHC

Aravind T S, Kajari Mazumdar, Andrea Bocci

Tata Institute of Fundamental Research, Mumbai
Plan

- Introduction to LHC and its future
- CMS Experiment : a brief overview
- Challenges at high instantaneous luminosity
- Trigger in CMS Experiments
- **Heterogeneous Computing using Accelerators**
- GPU at HLT for CMS
- Summary

Reminder: physics mandates of LHC:

1. Discover Higgs boson → **achieved in 2012 by CMS and ATLAS experiments**
2. Probe physics at TeV energy scale
   
   and many more ...
LHC to continue for 2 more decades, with more vigour (intensity) over time.
Preparation started already→ upgrades

We are here

Very successful operations since 2010. Have delivered only few % of total data volume expected by 2040.

Currently analysing data collected so far & preparing for Run3 and future upgrades.
$O(10^8)$ front-end electronic channels, dominated by tracking system (66M pixels + 9.6M Silicon microstrips)

Weighs $14 \times 10^6$ Kg (heaviest detector built to date)
CMS experiment: some typical numbers

One of the 2 multipurpose experiments, with focus on Higgs & discovery of physics at TeV energy scale.

- At LHC, # of bunches/beam: 2556, Number of protons/bunch: $1.1 \times 10^{11}$
- Bunch crossing rate: 40MHz $\rightarrow$ front-end electronics ready every 25 nano-sec.
- p-p total interaction cross section: $\sim 80$ mb, compare with Higgs production: $50$ pb @13 TeV
- Expect rate of new physics, eg. other possible heavy particle production, several-many orders lower
- Instantaneous luminosity $10^{34}$/cm$^2$/s $\rightarrow$event rate: $10^9$Hz (dominated by mundane physics)
  size of each event: $\sim 1$-2 MB
- Storing all events for further analysis would require data flow of around 60TB/s (!!).
- Technology for permanent storage, ie, tape writing capability drives amount of data which can be actually stored for offline analysis ($\sim$ few Gbps) $\Rightarrow$ store few kHz events for detailed analysis).

Compromise: select potentially the MOST interesting events online: Trigger
Challenge of event reconstruction at high instantaneous luminosity

- Average number of interactions during a single bunch crossing (Pile Up : PU) for nominal luminosity of $10^{34} \text{ /cm}^2\text{/s}$, PU $\sim 20$
- This causes information about multiple interactions recorded simultaneously in the detector as a single event eg., at future instantaneous luminosity of $10^{35} \text{ /cm}^2\text{/s}$, PU = 200

- PU increases measured energy in the calorimeter and the density of charged particle tracks in the tracking system.
- This overhead on reconstruction hinders the trigger decision to be faster and efficient.

*Increase in the number of hits in the detector layers exponentially increases the combinatorics involved in the reconstruction of charged particle tracks.*
2-level Trigger of CMS

**Level1 (L1) trigger**: i) implemented close to the detector back-end
   ii) reasonably coarse checks for simple logics
   iii) executed via custom electronics
   iv) Decision latency: $3.8 \mu s$

This selects ~ 100 kHz events (in future: 750 kHz, latency 12 $\mu s$)

**High Level Trigger (HLT)**
   i) implemented via computer farm
   ii) fast complete reconstruction of the event
   iii) decision based on complex signatures of interesting events.
   iv) Decision within few hundred msec.

This selects ~ 1kHz events (in future: 7.5 kHz)

Selection of interesting events capped by processing, communication and storage capabilities at each level.

~1000 Hz
Heterogeneous Computing using Accelerators

- CPUs are highly optimized for sequential workflows
- CPUs are not the optimal way to go if large data volume needs to be processed with a single instruction (Single Instruction Multiple Data : SIMD).
- **Heterogeneous Computing** uses offloading of tasks to **specialized hardware** designed to accelerate the execution of such instructions less usage of CPU time and power!
- **Heterogenous combination of GPUs, FPGAs, ASICs, DSP blocks etc..**

FPGAs are specialized in Pipelining

- Divides the whole logic in smaller submodules to execute limited no. of instructions at a time
- Modules are linked in sequence and data flows from one end to other.
- Each module could take next data item just after execution of a previous one: **Pipelining**
- These submodules are implemented at logic gate levels.

**Result: massive acceleration in speed of execution**
GPU at HLT for CMS -1

- GPUs are massively parallel accelerators to which we can offload SIMD jobs.
- From the 2022, CMS will use GPUs to evaluate event properties partly at HLT → Baseline
- R&D and benchmarking for the offloading has been progressing and has accepted as the baseline for Run3.
- The GPUs will be deployed at two instances for the HLT

1. GPUs for event rebuilding in a subsystem
   - After positive L1 decision, info for a given event has to be collated from all parts of detector
   - GPU will map the raw detector signals to the geometry of the detector
   - For a given subsystem (Tracker/Calorimeter) GPU will assemble a complete global description of the event.
2) GPUs for global event reconstruction

Recently revamped algorithms scale better to increasing event complexity.

Reconstruction on calorimeters redesigned to use the massive parallelism offered by GPUs.

Reconstruction in tracker is affected the most with increasing instantaneous luminosity due to the combinatorics.
Hadronic and Electromagnetic calorimeter reconstruction already benefited the most by offloading to GPU.

The time consumed for track reconstruction in Silicon tracker is also considerably reduced.

But tracks produced in GPU have to be converted to CPU-compatible form for overall event reconstruction. Hence, at present, the acceleration or gain of using GPU is partially masked.
GPU+CPU workflows: spare time can be used for new trigger logics in HLT.

The new approaches give options to accelerate existing algorithms for event reconstruction.

- **Old workflow**: tracks reconstructed in the outer detectors (e.g., muon chamber) were to be matched to reconstructed hits in the tracker in an iterative method.
- **New workflow**: hit combinatorics can be handled fast to reconstruct tracks within the tracking volume and then matched with the passage/flow of particles into outer detectors (calorimeter, muon chamber).

**New approach**

**Old approach**

1. **Reconstruction of tracks from outer detectors**
2. **Remove the corresponding tracks from Silicon tracker**
3. **Reconstruct good tracks in Si Tracker**
4. **Good Inner tracks**
5. **Repeat**
6. **Refit**
7. **Final Tracks**

**New approach**

1. **Reconstruction of tracks from outer detectors**
2. **Reconstruction of tracks in Si Tracks**
3. **Matching**
4. **Refit**
5. **Final Tracks**
Summary

- GPUs will be used for the High Level Trigger decision in CMS experiment from the next Run of LHC (2022 onward).
- Integrating GPU based software model to existing setup (*conversion*) is a limitation now.
- **Better ways to redesign the workflow of HLT decision process has opened up.**
- Substantial speedup is obtained for the reconstruction. More R&D will improve it further.
- For future upgrades, along with GPUs, FPGAs are also being explored as options for compute operations at HLT.
References

- LHC machine configuration 2018
- The CMS experiment at the CERN LHC
- A brief review of Triggering in CMS
- Heterogeneous Computing for trigger in CMS
- Public Plots for the GPU based tracking in CMS
- HL-LHC and plans
Backup
Illustration for event reconstruction in CMS

Highly compute intensive

Detector data is used to identify tracks of charged particles, infer its momentum, energy deposited by charged and neutral particles and finally identify the global picture of an event.
Plans for the Future upgrade

CMS is planned to operate after a major *Phase 2* upgrade, in 2027. The instantaneous luminosity at the beamspot is going to 7.5 time the existing value. *The trigger rates at L1 and HLT are targeted to be 750 kHz and 7.5 kHz* (current values 100kHz and 1kHz).

To meet the challenge its projected to offload 90% of the workload onto Heterogeneous accelerators. FPGAs, an electronic chip which can be programmed to the logic gate level to implement algorithms, is the current forerunner to be considered for HLT upgrade (along with GPUs).

The challenge for managing a heterogeneous analysis framework is the code portability and compatibility for scaling. Various initiatives have been supported by CMS to study and develop compatibility interfaces. *Non existence of open source development from the industry* has been a major hurdle in progressing in these area.
Si Tracker and Pixel Detector

- Biased PN junction diodes (made into sheets)
- Passage of charged particles induces current: SIGNAL
- Very small latency and precise localization of signal

Pixel cells: 100 μm × 150 μm, 125M channels
[channel occupancy ~ 0.01 %, 360μm Thick]

Si Strip cell: 10 cm × 80 μm, 9.5M channels
[channel occupancy ~ 1-2 %, With 540μm]
(both of them provides 3-D hits)
Total Si surface area of around 200m² (~ a tennis court)

Current tracker inner barrel

Pixel Upgrade plans

Current density 1ns after a charged particle passes through the biased Si cell. The current collection is complete <8ns after the impact.
Tracking performance for the GPU algorithms

Reference: Patatrack Results, Patatrack Result Plots Page
GPU Workflow and Computing Performance for tracker reconstruction

Reference: Slides from Exa.TrkX kickoff meeting
CMS Integrated Luminosity, pp, 2018, $\sqrt{s} = 13$ TeV

Data included from 2018-04-17 10:54 to 2018-10-26 08:23 UTC

- LHC Delivered: 67.86 fb$^{-1}$
- CMS Recorded: 63.67 fb$^{-1}$

CMS Preliminary

**Event rebuilding**

100m

**surface**

Good pic will come