IA: Year 3 Plans, Year 4 & 5 Goals

David Lange, Princeton
Heather Gray, UC Berkeley/LBNL

Major Activities and Goals

Developing tracking algorithms for HL-LHC

Determining charged-particle trajectories ("tracking") requires most CPU in reconstruction

- Develop more efficient algorithms
- Develop more performant algorithms

Hardware accelerators are the way forward to speed up and reduce infrastructure cost

- Use of hardware accelerators for tracking
- ML on accelerators in realistic HEP apps

Re-engineering algorithms for hardware accelerators

Exploiting major advances in machine learning (ML)

Capitalize on industry and data science techniques and tools

- Investigate new HEP applications of ML
- Apply new ML techniques to HEP

IA projects in a slide..

1. PVFinder : ML approach to finding primary vertices (UC, MIT): Working to establish to what extent a better PV performance (physics, technical) can be achieved via an ML approach?
2. Allen framework: Configuration and monitoring of GPU framework (MIT). Due to the Allen project results, LHCb recently adopted a GPU-based first-level trigger (HLT1) for Run 3.
3. mkFit: Efficient track finding on modern architectures (UCSD, Princeton, Cornell). Has achieved large speed ups in track building. Focused on deployment into CMS HLT for Run 3.
4. FPGAs and ML for calorimetric reconstruction (MIT). Identified use case algorithms and prototypes for ML algorithms using FPGAs “as a service”
5. ACTS tracking (UC-Berkeley, Stanford). CPU/GPU implementations of Kalman filter, seed finding, ambiguity resolution in ACTS framework.
6. GNNs for tracking (UIUC, Princeton): Recently started R&D on applying graph neural networks to (pixel) tracking. Moving towards an increased collaboration with ExaTrkX.
7. ML for Jet algorihtms (NYU): Hosted ML4Jets; Developing common benchmarks

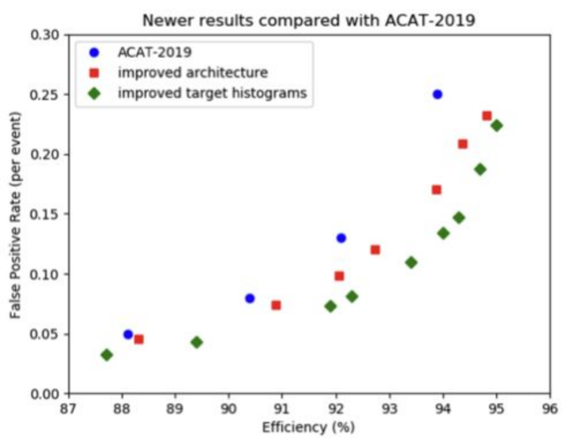
Thursday's session

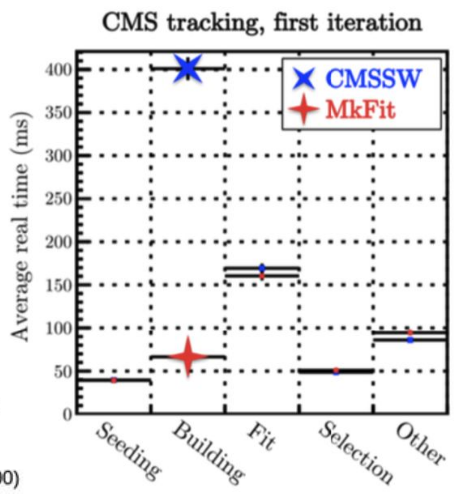
Mainly organized to discuss Year 3 goals and to gather input for the PEP.
Nevertheless, we had lots of interaction on the scientific results from each group.

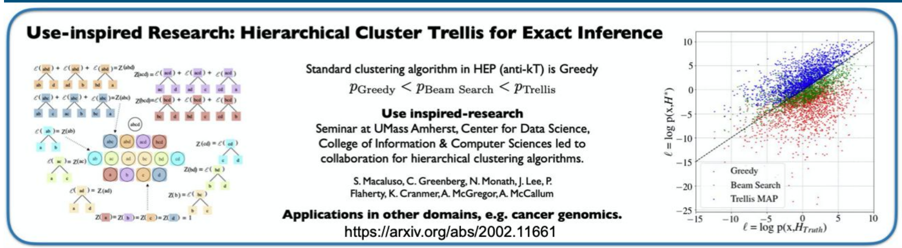
Outcomes:

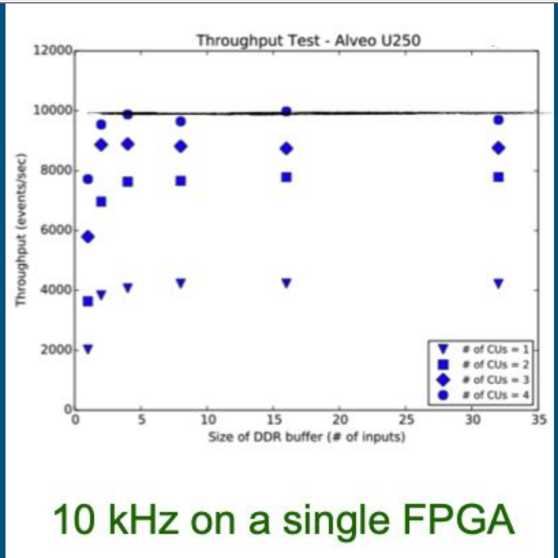
- We have captured numerous exciting results and clear R&D directions
- We have a start on what is needed for the PEP. "Homework" to be requested as a follow up for cases where not all information is in hand.
- We identified a number of topical meetings to have over the next months
- As proposed on Tuesday, cross-Institute (or at least cross-IA) discussions on FPGA R&D outcomes/directions are clearly needed.
- Two suggestions for regular meetings (beyond topical meetings). Please give us more input on your interest (knowing that these would have to happen between 11-2pm EDT which is likely full for most of the team...)
- Ideas around formulating quantitative metrics are still illusive.

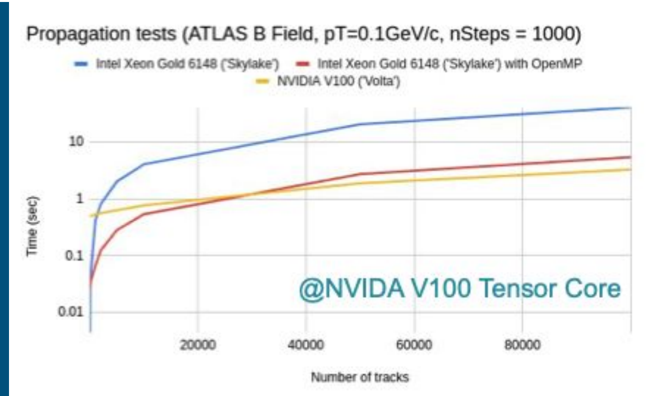
Some highlights

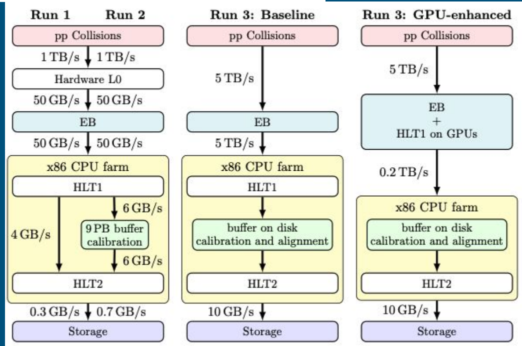




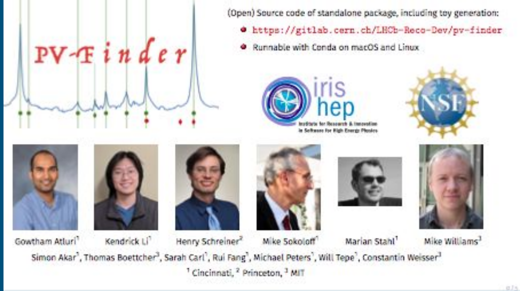








PVFinder

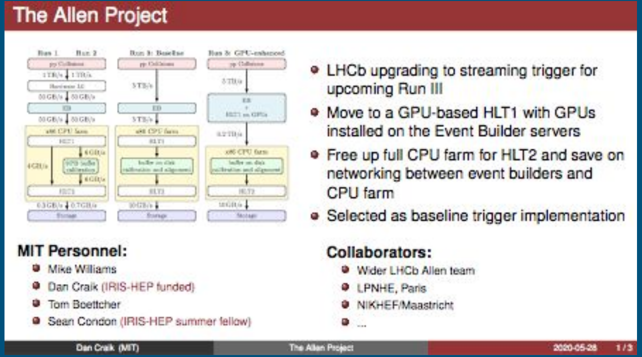


Year 3 goals

- Benchmark performance with standard LHCb simulation and software, and compare to the current baseline PV finding algorithm.
- Re-train the algorithm using full LHCb simulation in place of toy simulation.
- Deploy the algorithm in Allen.
- Develop an algorithm to assign tracks to PVs probabilistically.
- KDE generation too slow
  - develop fast ML algorithm for KDE generation (or alternative approach)
- Prune ML algorithms if necessary to fit into the HLT time budget.
- Improve existing algorithm by first characterizing PVs in categories of multiplicity, track variance, PV proximity and SVs; Develop targeted ML approaches to tackle inefficiencies.

Homework for next week: Trim Y3 goals for PEP, suggested project metric(s) and Year 4+5 outcomes

Allen



- Year 3
  - Working on defining milestones now in consultation with the Allen team
  - Likely to focus on monitoring and data quality
- Longer term
  - Allen gives us access to all experimental info from every single event at LHCb
    - Access to low- and high-level objects in single stage
  - Implementing automatic (experimental) anomaly detection in Allen would be an interesting – and fun! – (grand?) challenge
- Expected outcomes
  - Develop and deploy algorithms to configure and monitor fully GPU-based HLT application – and have it all running at LHCb

Homework for next week: Complete Year 3 goals, factoring in alignment with IRIS-HEP goals, suggest project metric(s)

mkFit

- Year 3
  - Re-write vectorized KF for the new HL-LHC tracker geometry and L1 4.1 trigger segments (w support from US-CMS OPS program)
  - Develop an alternative pattern recognition algorithm (segment-linking) to KF for the new HL-LHC tracker geometry. Naturally parallelizable and vectorized. (w support from US-CMS OPS program)
- Longer term
  - Integrate segment-linking tracking into the CMS software environment 4.1, 4.2 and benchmark (w support from US-CMS OPS program)
- Proposed outcomes
  - During Run3: Demonstrate usage by CMS in data taking
  - During HL-LHC (Phase-2 geometry of tracker): Run adapted version of mkFit, deployment and usage of Segment-Linking tracking algorithm

Homework for next week: Identify timescale for HL-LHC MC demonstration/integration. Identify what can be part of the CMS computing TDR? (eg, during IRIS-HEP Y4)

ML w/ FPGAs

- Year 3
  - Aim to demo FPGA algorithms within the HLT (at partial or full scale)
  - Include more detector information (eg, depth segmentation in CMS HCAL)
  - Evaluate other ML models / techniques
  - Evaluate on EM calorimeter data
- Longer term
  - Establish infrastructure for use of FPGAs in HLT and/or offline.
  - Possible challenge problem: Demonstrate X% of HLT or RECO on FPGA/GPU

<u>**Homework for next week**</u>: Suggest project metric(s)

- Stanford
  - Lauren Tompkins (PI), Rocky Garg (postdoc), Marcelo Vicente (FPGA Engineer), Peter Chatain (Undergraduate Summer 2020), Cole Mocyemba (Undergraduate Summer 2020)
- UC Berkeley
  - Heather Gray (PI), Xiaocong Ai (postdoc), Ralf Farkas (fellow, also Bonn),
- Collaborators, some examples
  - ACTS (Andreas Salzburger, Paul Gessinger-Befurt, Moritz Kiehn, ...)
  - Charles Leggett, Beomki Yeo (LBL)
  - Nick Styles, Georgiana Mania (DESY)
  - EIC (Barbara Jacek, Markus Diefenthaler)
  - sPhenix (Joe Osborne, Anthony Frawley)
  - Belle-2 (Florian Bernlochner, Ralf Farkas)

ACTS Tracking

- Year 3
  - ACTS demonstration on ATLAS and/or ATLAS-ITK detector
    - Comparison to current ATLAS algorithms
    - Integration and comparison of ACTS algorithms in Athena
  - Kalman Filter prototype (for track fitting) on GPU
  - Alignment implementation in ACTS
  - ACTS Paper, ACTS Workshop (2021)
  - Evaluation of novel tracking algorithms (Snowmass whitepaper)
  - Benchmark I/O and sparse memory access for Alveo (FPGA accelerator)
- Proposed Outcomes
  - Complete, modern and experiment-independent track reconstruction package
    - Deployed within the ATLAS athena framework for HL-LHC
    - Used by other high-energy and nuclear physics experiments
  - Performance evaluation and optimization (e.g. GPU or FPGA accelerated) of tracking algorithms

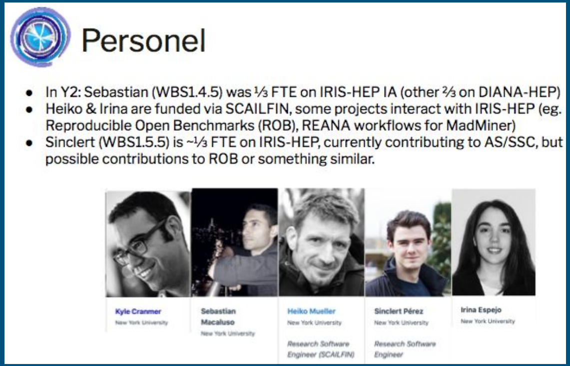
Homework for next week: Trim Y3 milestones for PEP, suggest project metric(s)

GNNs for Tracking

- Our group is aiming to improve charged-particle tracking in ATLAS & CMS using GNNs and accelerators (main focus FPGAs)
- Year 3 Milestones & Deliverables
  - Explore implementation of traditional (non-ML) tracking algorithms on FPGAs
  - Establish an efficient and 'acceleratable' baseline graph-based tracking pipeline
    - Graph design (embeddings, transformations, edge constructions), GNN architecture, Track construction/post-processing
  - Snowmass white paper
- Year 4/5 Goals
  - Demonstration and benchmarking of a graph-based tracking accelerator within an HLT-like environment
  - Accelerated tracking work gets integrated into the HL-LHC planning for one or more experiment

Homework for next week: Schedule early on demonstration of adequate physics performance; suggest project metric(s)

ML4Jets



- Potential Y3 Milestones:
  - Performance of sparse hierarchical trellis algorithm (compared to greedy, beam search, exact trellis in terms of finding maximum likelihood clustering and marginal likelihood)
  - Snowmass white paper contributions about potential to unify Monte Carlo generation
    - (eg. parton shower) and inference (eg. jet tagging) with probabilistic programming
  - Exploratory:
    - Exploration of hierarchical trellis to speed up the bottleneck in the CKKW-L matrix element / parton shower model (investigation of generalization to 2->3 splittings)
    - Exploration of probabilistic programing for MC generation in tails of jet phase space
- Proposed project outcomes:
  - Roadmap / white paper outlining what new capabilities would be enabled by enabling Monte Carlo generators (focusing on Jets) with probabilistic programming functionality and what interventions in code are needed to get there:
    - Improved MC tuning, remove CKKW-L bottle neck, more efficient MC generation in tails of jet substructure, better jet clustering and tagging, etc.

Homework for next week: Identify if/where experiment connections can be made more explicit (perhaps via algorithm implementations for jets or generators outcomes?)

New Proposal: FastMC

Enable fast (> 10x faster than standard) MC for Phase-2 for early, approximate physics studies, signal parameter scans, evaluation of physics driven systematic uncertainties and potentially more.

Lower processing **and** storage needs focusing on derived data formats (Mini/NanoAOD or DAOD_PHYS/PHYSLITE)

Program of work in 2 Phases

CMS focused and starting from FastSim

- 1) Implementation of Phase-2 detector simulation and reconstruction, specifically for HGCAL
- 2) Validation of physics performance
- 3) Expanding usage for physics in Run 3
- Milestones for end of 2020
  - first Phase-2 release
  - demonstration of Run 2 analysis

IRIS-HEP oriented, i.e. not necessary limited to CMS

- 1) FastMC tuning with ML
- 2) Fast reconstruction using ML - see Phil's talk
- 3) Calorimeter showering with ML
- Milestones for Summer 2021
  - first ML based tuning for Run 3
  - test and give feedback on 2)
  - evaluate performance of 3)

Markus Klute (MIT)