Innovative Algorithms

Presenter: David Lange

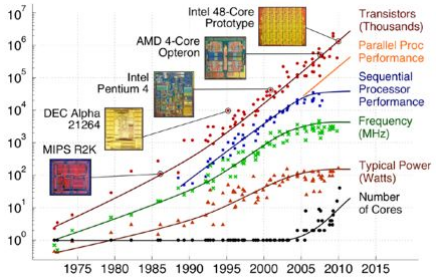
Major Activities in IA

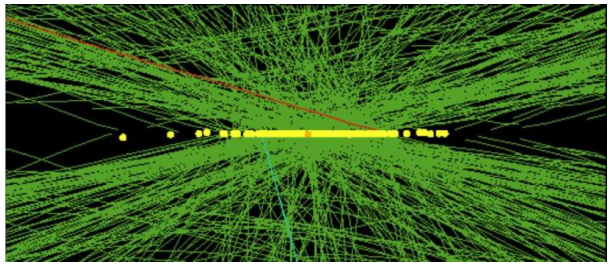
Redesigning tracking algorithms for HL-LHC

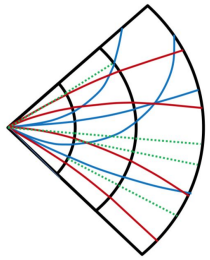
- Determination of charged-particle trajectories ("tracking") is largest component of CPU for event reconstruction
- IRIS-HEP investigations
  - *More efficient algorithms*
  - *More performant algorithms*
  - *Use of hardware accelerators*

Exploiting major advances in machine learning (ML)

- ML in HEP may be a major opportunity
  - *Capitalize on industry and data science techniques and tools*
  - *Could reduce CPU needs*
  - *Could lead to wider use of accelerators*
- IRIS-HEP investigations
  - *New HEP applications of ML*
  - *Use of new ML techniques*
  - *ML on accelerators in realistic HEP apps*

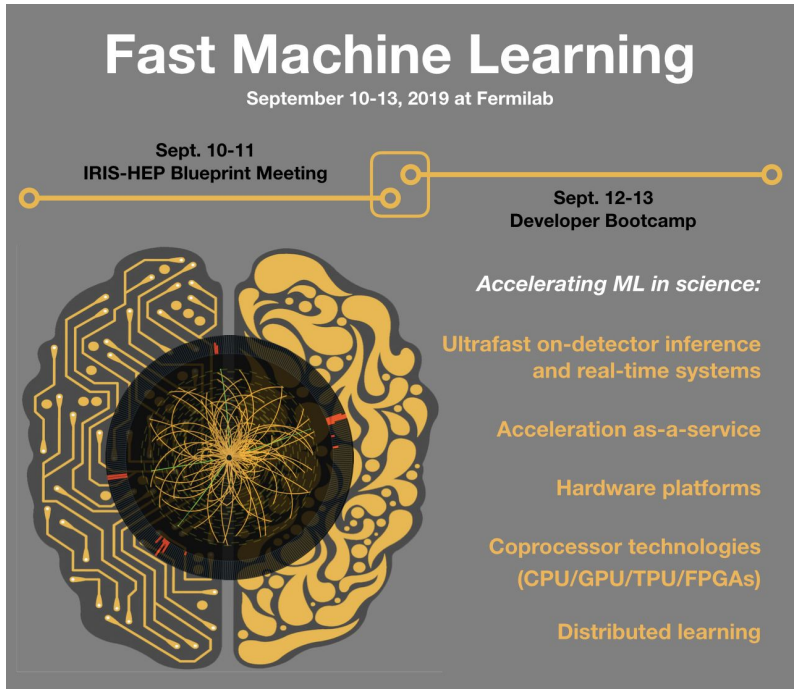






ML on FPGAs contributions to HLS4ML / FastML

- IRIS-HEP goals were to identifying specific use cases and operational scenarios for use of FPGA-based algorithms in experiment software trigger, event reconstruction or analysis algorithms
- The FastML workshop at FNAL attracted nearly 200 people and partly served as a blueprint workshop to build a roadmap towards accelerated ML applications and demonstration goals
- The IRIS-HEP team has recently started two related (but more broadly applicable) NSF projects: Advancing Science with Accelerated Machine Learning and Machine learning and FPGA computing for real-time applications in big-data physics experiments

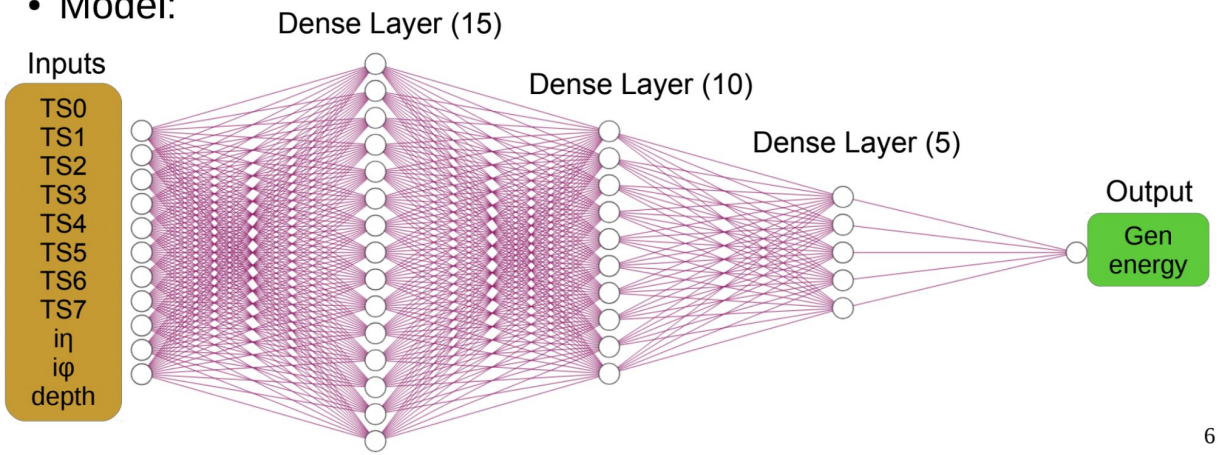


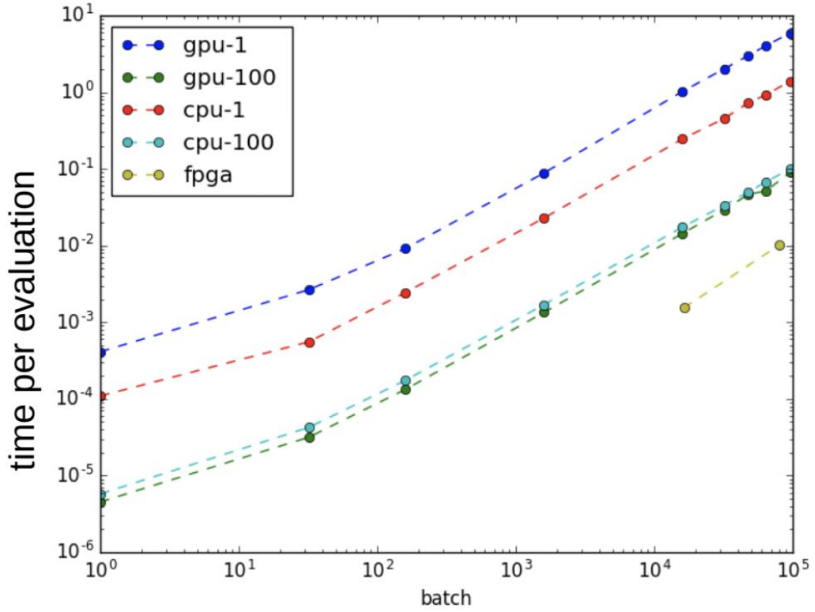
New focus for IRIS-HEP in this area is on developing and demonstrating applications

We aim to demonstrate the physics performance of FPGA-friendly machine learning based algorithms relative to current production algorithms

- CMS HCAL local reconstruction - determination of the in-time and out-of-time charge for reconstruction hits.
- Pixel clustering using graph neural networks

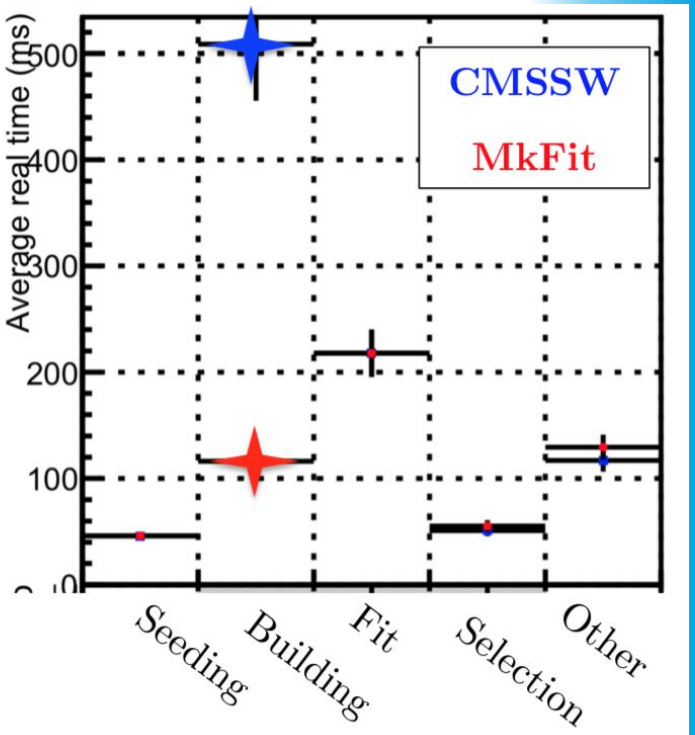
- Model:



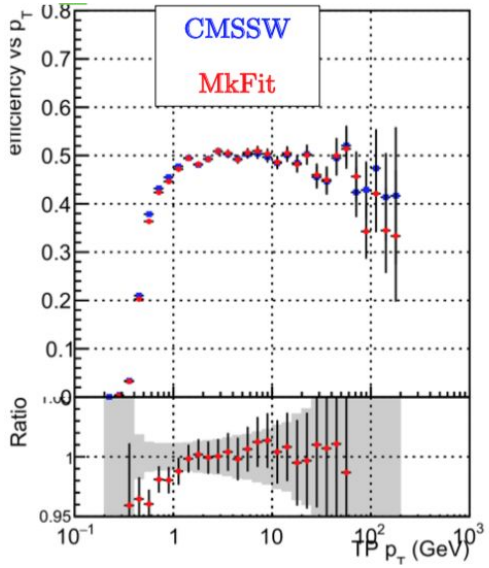


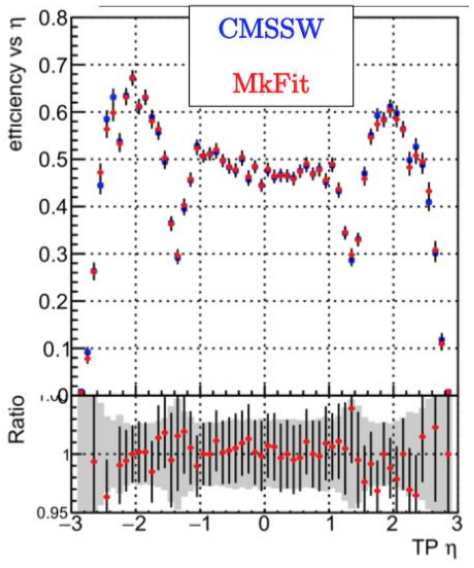
Parallel Tracking contributions to mkFit project

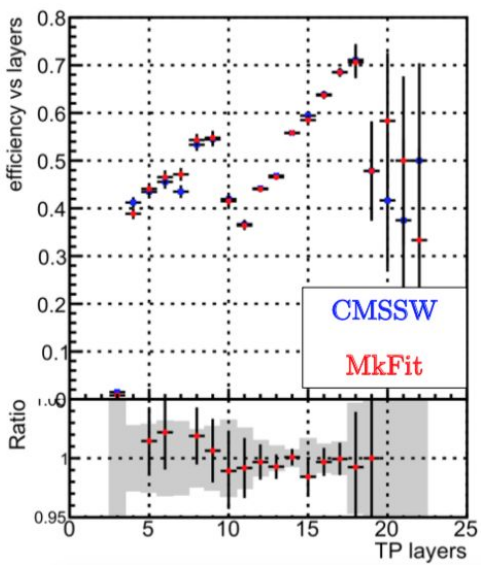
- Develop track finding/fitting implementations that wor many-core architectures (vectorized and parallelized al faster track building w/ similar physics performance in benchmark comparisons
  - *Recently integrated into CMSSW*



* Measured on SKL, mkFit compiled with AVX-512, turbo boost disabled







mkFit post CMSSW integration

Substantial progress towards first draft of mkFit publication

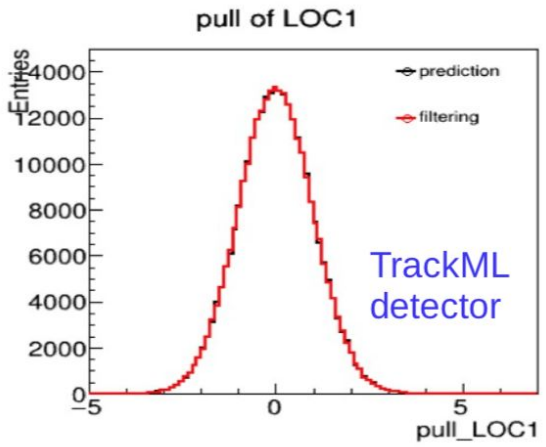
Looking at GCC vs ICC performance comparisons
(higher levels of vectorization achieved by ICC vs GCC)

Continued improvement in physics performance achieved (using HLT-like configuration as first use case)

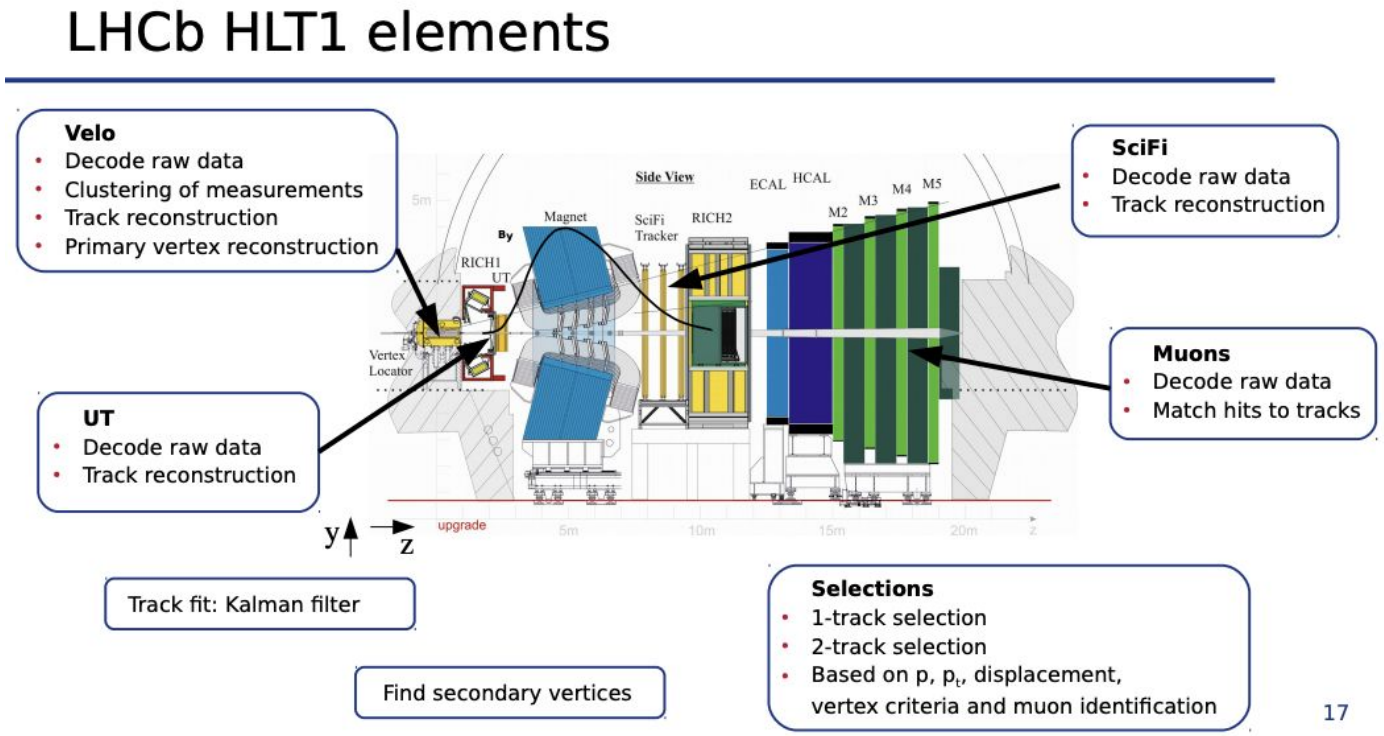
Investigating GPU demonstrators, including strip tracker unpacking

Tracking contributions to ACTS Project

- Open-source software project for multi-experiment track reconstruction built on the extensive experience in track reconstruction in the ATLAS experiment.
- IRIS-HEP contributions include
  - *Development of the Kalman Filter*
  - *Porting ACTS seeding code to run on GPUs*
  - *Ongoing development of Combinatorial Kalman Filter*
  - *Developing connections with other experiments (e.g. Belle-2, JLAB) who may be interested in using ACTS*



Engagement with the Allen project



IRIS-HEP contributing to

- Enabling run-time algorithm configurability
- Adding asynchronous monitoring capabilities

Upcoming workshops

ML4Jets2020

15-17 January 2020
America/New_York timezone

Search...

- Overview
- Call for Abstracts
- Timetable
- Registration
- Participant List
- LHCOlympics2020
- Slack channel
- Getting to NYU
- Accomodation

Machine learning has become a hot topic in particle physics over the past several years. In particular, there has been a lot of progress in the area of particle and event identification, reconstruction, fast simulation and others. One significant area of research and development has focused on jet physics. In this workshop, we will discuss current progress in this area, focusing on new breakthrough ideas and existing challenges. The ML4Jets workshop will be open to the full community and will include LHC experiments as well as theorists and phenomenologists interested in this topic. This year's workshop is hosted at NYU in New York City.

This workshop follows successful workshops in 2017 and 2018.

Registration is now closed - please email the organizing committee if you are still interested in attending the workshop in person.

International Organizing Committee:
Kyle Cranmer (NYU), Ben Nachman (LBNL), Maurizio Pierini (CERN), Tilman Plehn (Heidelberg), Jesse Thaler (MIT)

Local Organizing Committee:
Kyle Cranmer (NYU) and Sebastian Macaluso (NYU)

Starts 15 Jan 2020, 09:00
Ends 17 Jan 2020, 19:00
America/New_York

https://indico.cern.ch/event/809820/overview

Connecting The Dots 2020

22-24 April 2020
Princeton University
America/New_York timezone

Search...

- Overview
- Scientific Program
- Organizing Committees
- Call for Abstracts
- Local Information
  - CTD 2020 Venue
  - Getting to CTD2020
  - Campus logistics
  - Accommodation
  - Visas

Local organizers
ctd2020-loc@iris-hep.org

The Connecting The Dots workshop series brings together experts on track reconstruction and other problems involving pattern recognition in sparsely sampled data. While the main focus will be on High Energy Physics (HEP) detectors, the Connecting The Dots workshop is intended to be inclusive across other scientific disciplines wherever similar problems or solutions arise.

Princeton University (Princeton, New Jersey, USA) will host the 2020 edition of Connecting The Dots. It is the 6th in the series after: Berkeley 2015, Vienna 2016, LAL-Orsay 2017, Seattle 2018, and Valencia 2019.

Registration: CTD2020 will be 2.5 days of plenary sessions, starting Wednesday morning April 22, and finishing around lunch time Friday April 24. The workshop is open to everyone and proposals for talks are being accepted until December 6. More information available on the Scientific Program and Call for Abstracts pages.

Workshop registration will be available soon. The regular registration fee will be 270 USD and 185 USD for students (either graduate or undergraduate student as of January 1, 2020). After February 22, the registration fee will increase by 25 USD (for non-students only). This fee covers local support, morning and afternoon coffee breaks, two lunches, the welcome reception and workshop dinner.

This workshop is partially supported by National Science Foundation grant OAC-1836650 (IRIS-HEP), the Princeton Physics Department and the Princeton Institute for Computational Science and Engineering (PICSciE).

http://cern.ch/CTD2020

Community Building

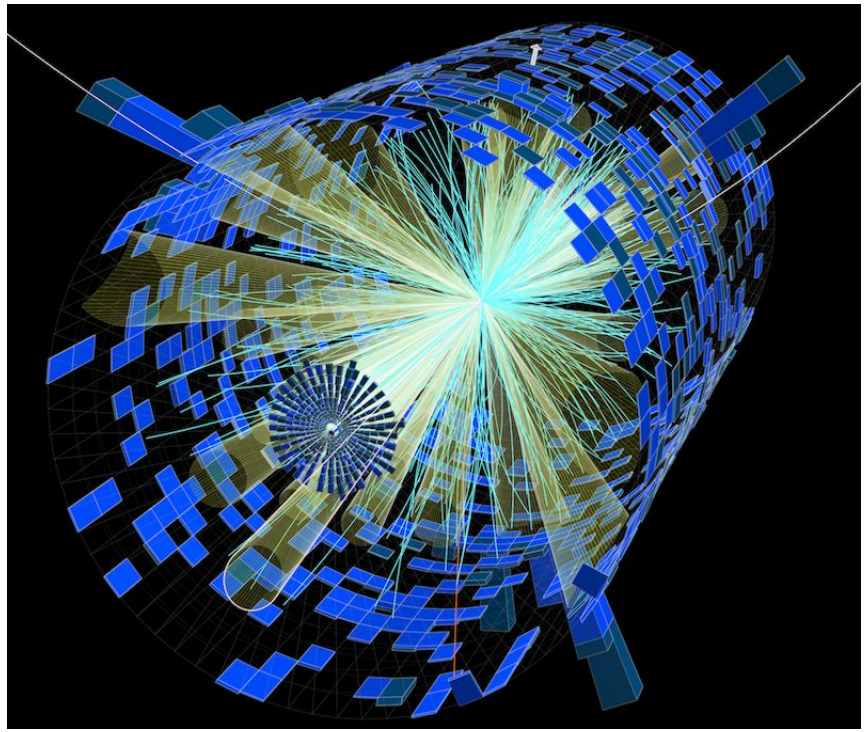
- IA projects have established a consensus around tracking and use of machine learning techniques as critical challenges and are working to address them first
- Some IA projects are already being integrated with LHC experimental software applications
  - *Completed: MkFit into CMS*
  - *Ongoing: ACTS into ATLAS , ML4Vertexing into LHCb*
- IA members have numerous community leadership roles, including:
  - *HSF working group leadership (software trigger and reconstruction, simulation))*
  - *Organizing community workshops and conferences*
    - Tracking workshop around ACTS at Berkeley and DESY (May 2020), FASTML blueprint workshop at FNAL, ML4Jets (January 2020, NYU), *Workshop at NeuroIPS 2019, Connecting the Dots (April 2020, Princeton)*
- IA developments impact the broader community
  - *Developing connections with other experiments exploring the use of ACTS (including Belle-2, JLAB, LDMX, eIC)*
  - *Industry connections (e.g. Microsoft) from HL4ML*
  - *MkFit co-development with DUNE, University of Oregon*

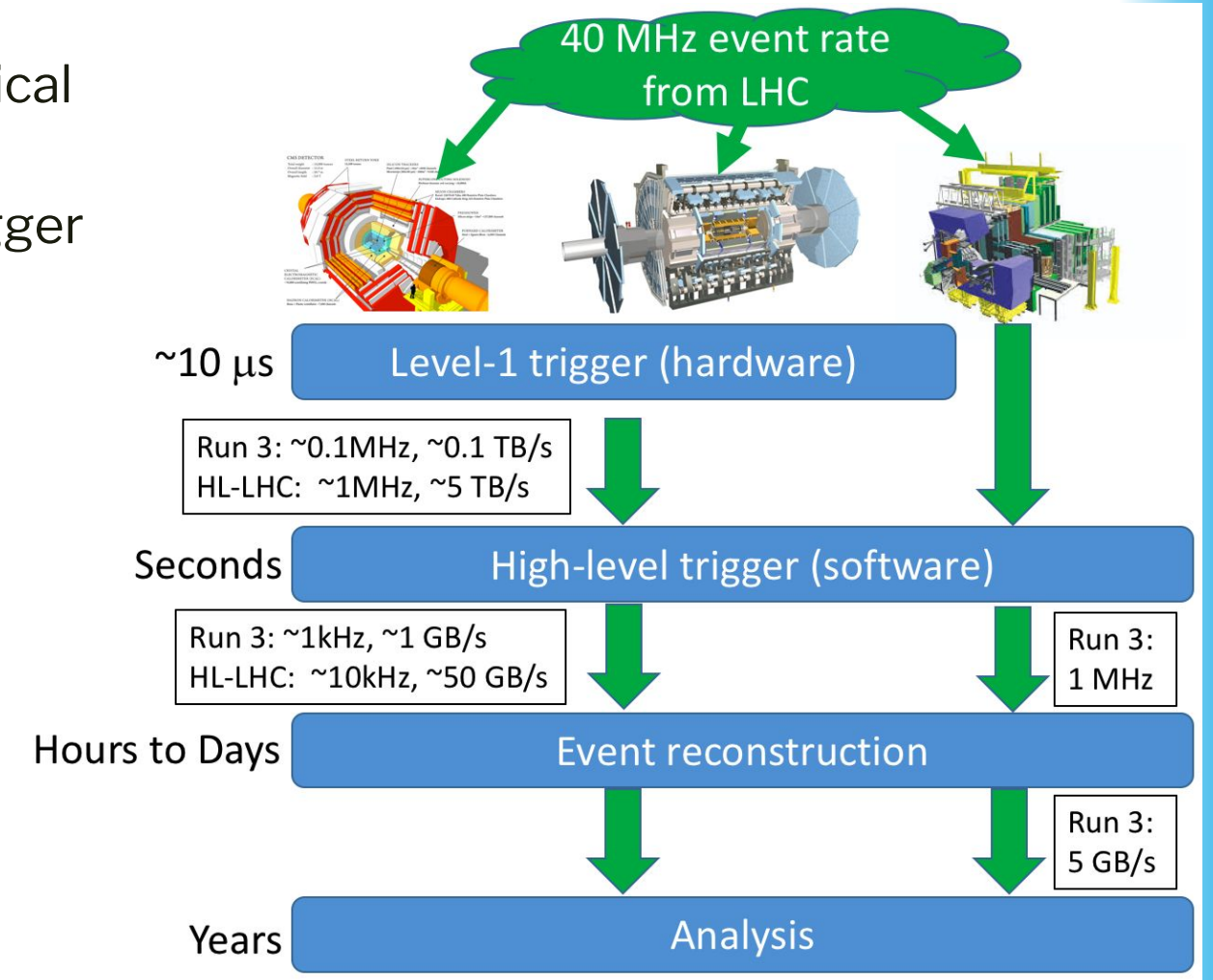
Backups

Focus Area: Innovative Algorithms

Developing and reengineering algorithms for critical HEP software applications. Primary focus areas:

- real-time data processing in the software trigger
- offline reconstruction





ML4Jets project to establish and curate common metrics and datasets for algorithm development

Crossover project to connect with diverse segments of machine learning community.
Strong connections with theoretical community interested in jet physics

- Tree Neural network approach demonstrated on reference dataset

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [30]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], M. Fairbairn[5], W. Fedorko[6], C. Gay[6], L. Gouskos[7], P. T. Komiske[8], S. Leiss[1], A. Lister[6], S. Macaluso[3,4], E. M. Metodiev[8], L. Moore[9], B. Nachman,[10,11], K. Nordström[12,13], J. Pearkes[6], H. Qu[7], Y. Rath[14], M. Rieger[14], D. Shih[4], J. M. Thompson[2], and S. Varma[5]

https://arxiv.org/pdf/1902.09914.pdf