Innovative Algorithms

Area leads:
Heather Gray (UC-Berkeley/LBNL),
David Lange (Princeton)

From the CWP to IRIS-HEP

iris-hep

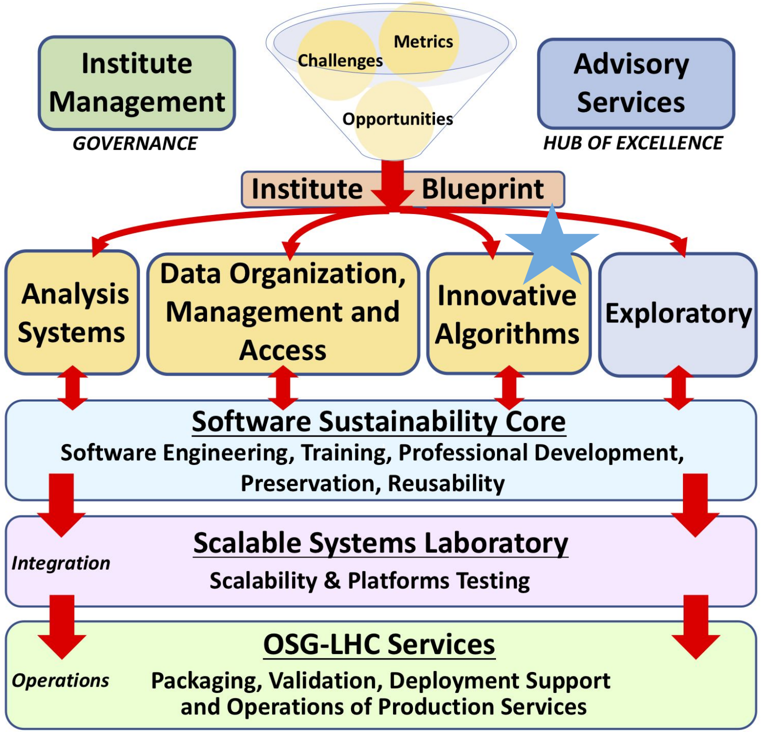
A Roadmap for HEP Software and Computing R&D for the 2020s

HEP Software Foundation[1]

ABSTRACT: Particle physics has an am
for the coming decades. This progran
hardware, either to build new facilities a
Similarly, it requires commensurate inv
manage, process, and analyse the shear
for the HL-LHC in particular, it is critic
agree on the software goals and priorities
In this spirit, this white paper describes
this software upgrade.

Contents

CWP

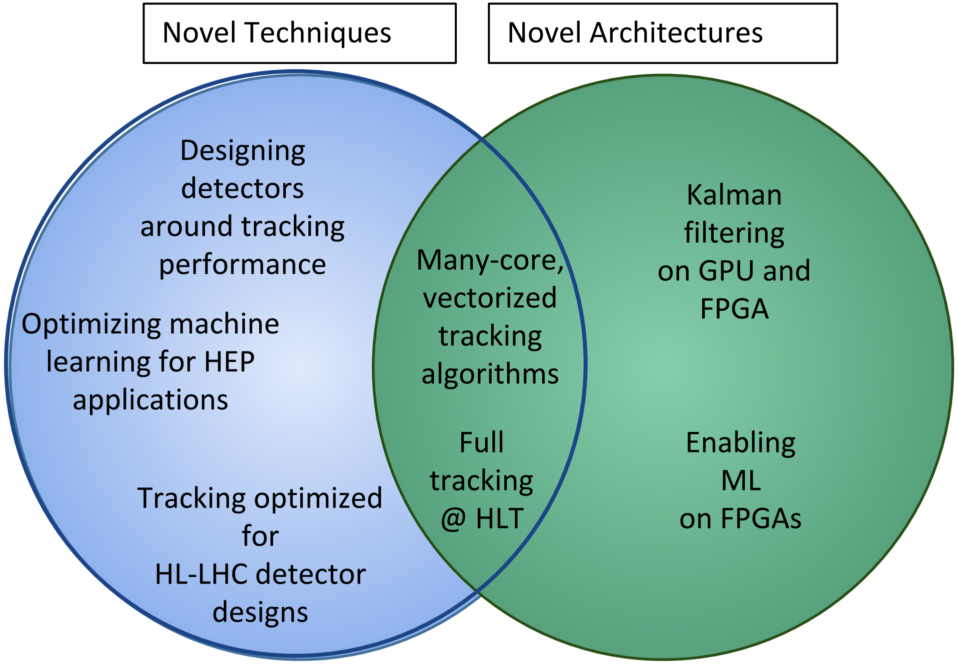


Scope of Innovative Algorithms (IA)

- Algorithms for real-time processing of detector data in the software trigger and offline reconstruction are critical components of HEP's computing challenge.
- **These algorithms face a number of new challenges during HL-LHC:**
  1. Upgraded accelerator capabilities, with more collisions per bunch crossing (pileup)
  2. Detector upgrades, including new detector technologies and capabilities
  3. Increased event rates to be processed
  4. Emerging computing architectures

Innovative Algorithms will employ a wide range of strategies to address these challenges and ensure that experiments are ready for HL-LHC physics

Initial activities will form around two themes: Novel Techniques and Novel Architectures



Given the HL-LHC timescale, projects must strive to advance best practices for software development in HEP

PIs in Innovative Algorithms

Peter Wittich, Cornell

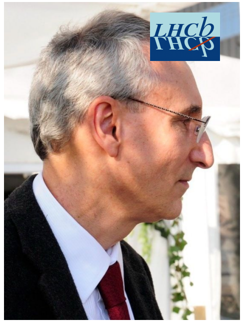
Mike Williams, MIT

Mark Neubauer, UIUC

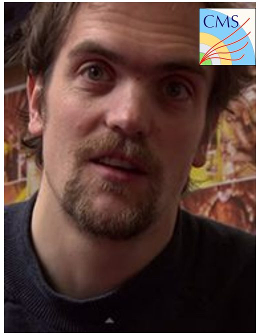
Avi Yagil, San Diego

David Lange, Princeton

Lauren Tompkins, Stanford



Mike Sokoloff, Cincinnati



Phil Harris, MIT

Kyle Cranmer, NYU

Heather Gray, Berkeley

Innovative Algorithms Projects

- **MKFit**: Parallel Kalman-filter tracking
  - Peter Wittich, Avi Yagil, Pete Elmer, Slava Krutelyov, Steve Lantz, Mario Masciovecchio, Dan Riley, Matevz Tadel, Bei Wang
- **ACTS**: Experiment-independent, inherently parallel track reconstruction
  - Heather Gray, Lauren Tompkins, Xiaocong Ai, Nick Cinko, Rocky Garb (Jan 2020)
- **FastPID**: Fast PID simulation for LHCb
  - Mike Williams, Daniel Craik
- **ML on FPGAs**: Fast inference of deep neural networks on FPGAs
  - Mark Neubauer, Philip Harris, Daniel Craik, Dylan Rankin
- **ML4Jets**: Machine learning for jets
  - Kyle Cranmer, Sebastian Macaluso, Irina Espejo
- **ML4Vertexing**: Machine learning for vertexing
  - Mike Sokoloff, Mike Williams, Henry Schreiner, Marian Stahl, Gowtham Atluri, Sarah Carl

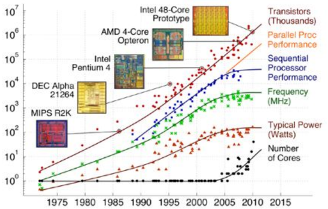
Groups are focused on answering 2 questions

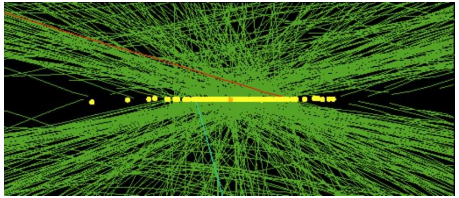
How to redesign tracking algorithms for HL-LHC?

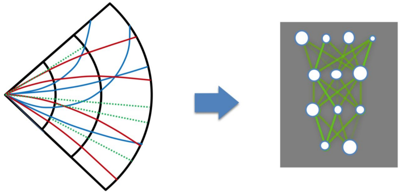
- Determination of charged-particle trajectories (“tracking”) is largest component of event reconstruction
- IRIS-HEP investigations
  - More efficient algorithms
  - More performant algorithms
  - Use of hardware accelerators

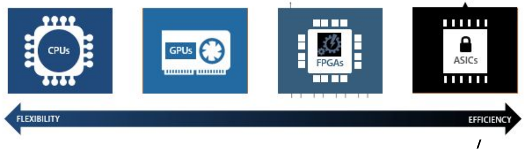
How to make use of major advances in machine learning (ML)?

- Use of 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









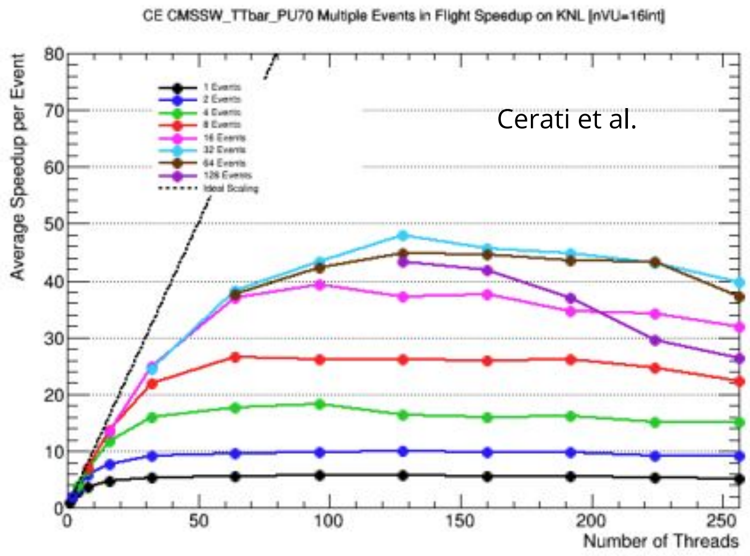
MKFit - Parallel Tracking (Cornell, Princeton, UCSD)

Aim to develop track finding/fitting implementations that work efficiently on many-core architectures (vectorized and parallelized algorithms)

Collaborators:

- Fermilab and University of Oregon (DOE SciDAC4),
- USCMS Ops Program
- CMS software (CMSSW) and trigger groups

http://trackreco.github.io/



https://arxiv.org/pdf/1811.04141.pdf

MKFit: Progress and Plans

- Primary focus is code integration with CMSSW
  - Initial version integrated and to be included in next CMS production release
- Production release of Matriplex expected soon
- R&D evaluations underway
  - GPU demonstrators
  - Methods to streamline data conversions
- Presentations: ACAT, Connecting the Dots, and IRIS-HEP topical meeting

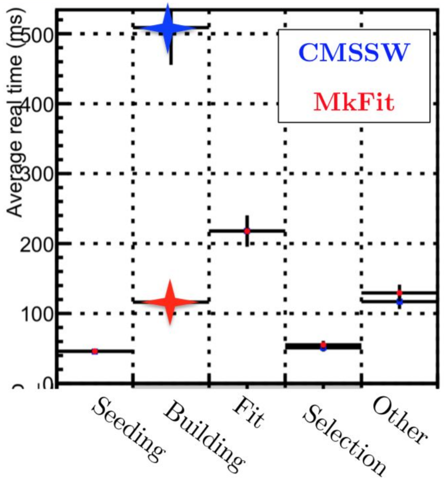
Integrated Timing Performance

Technical Details

- Run mkFit within CMSSW
- mkFit used for building only
- Single-thread test using TTBar PU 50

Results

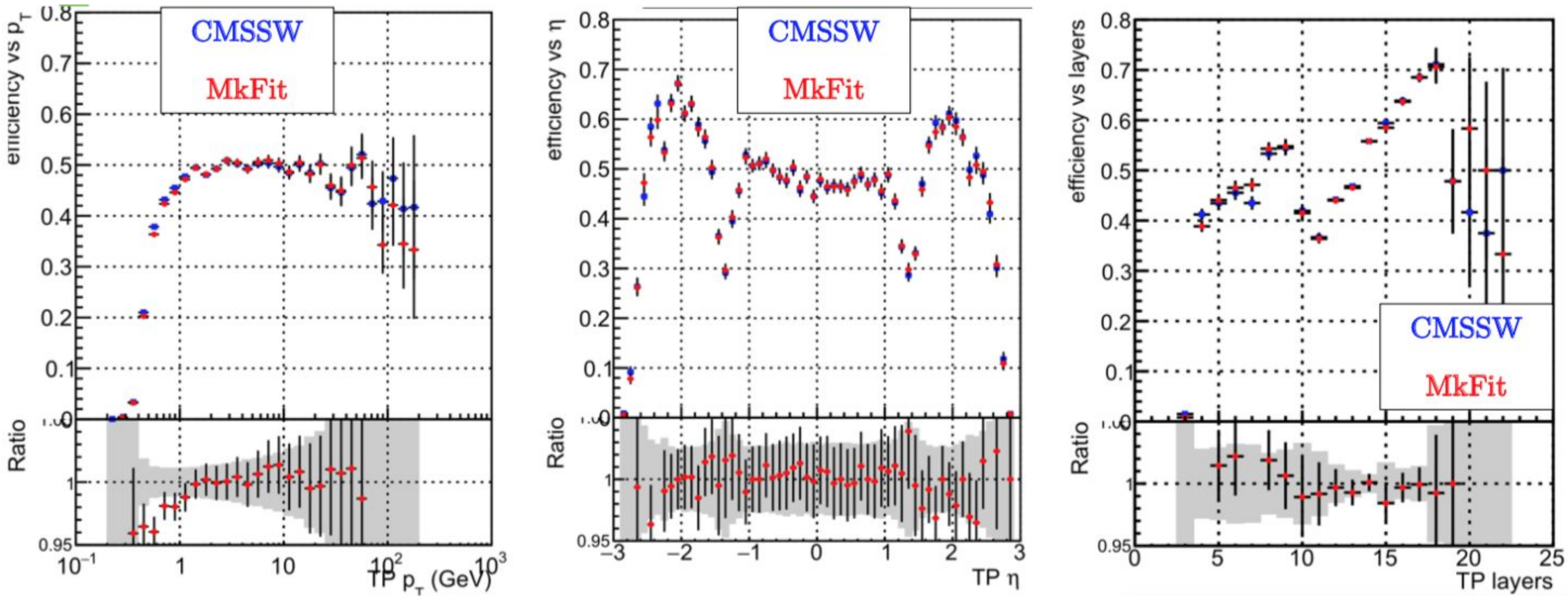
- Track building is **4.3x faster**
- 40% of time is spent in data format conversions – actual track finding is **7x faster**
- Track building now takes **less time than track fitting**
- Even larger potential speedups if multiple threads are used



Hall et al.

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

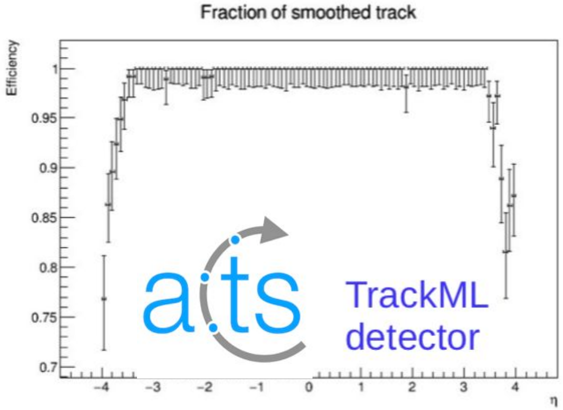
Tracking performance in CMS simulation (high-PU)



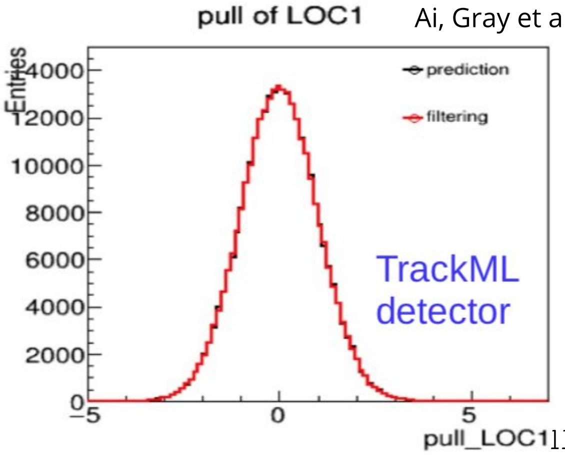
Efficiency benchmarks match CMSSW performance in apples-to-apples comparison with realistic geometry and expected detector conditions

ACTS (UCB, Stanford): Overview

- Open-source software project for multi-experiment track reconstruction built on the extensive experience in track reconstruction in the ATLAS experiment.
- Also being pursued for Belle-II and FCC
- Discussions with JLab, EIC, LDMX, NuStar about potential applications
- Initial IRIS-HEP contributions:
  pattern recognition, ambiguity resolution, GPU demonstrators
- **Collaborators**: CERN, KIT, LBNL

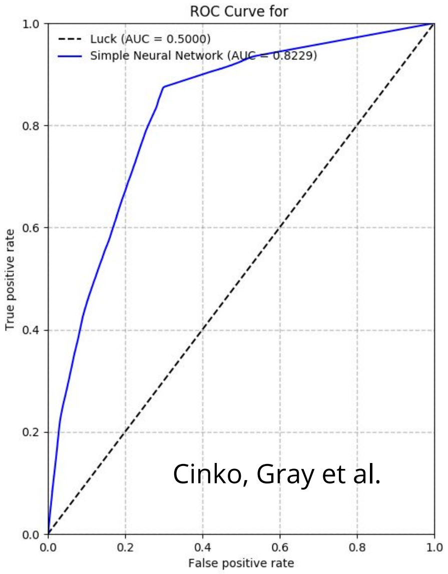


Ai, Gray et al.



ACTS (UCB, Stanford): Progress and Plans

- Hosted Berkeley Tracking Workshop (hackathon) in January 2019
- IRIS-HEP contributions to ACTS
    - Kalman Filter prototype algorithm implemented: performance and validation studies underway
    - Ambiguity resolution algorithm implementec ML studies ongoing
    - Track following implementation to begin
    - NERSC GPU-hackathon: prototype seeding code implementation on GPU
- Presentations: DPF 2019, USATLAS Annual meeting



ML on FPGAs (MIT, UIUC)

- HLS4ML is a machine learning inference package for FPGAs. Creates firmware implementations of ML algorithms using high level synthesis language (HLS)
- Initial IRIS-HEP contribution: Identify specific use cases and operational scenarios for use of FPGA-based algorithms in experiment software trigger, event reconstruction or analysis algorithms

- **Collaborators**: FNAL, MIT, CERN, Florida, UIC, UW

hls4ml

https://github.com/hls-fpga-machine-learning/hls4ml

ML on FPGAs

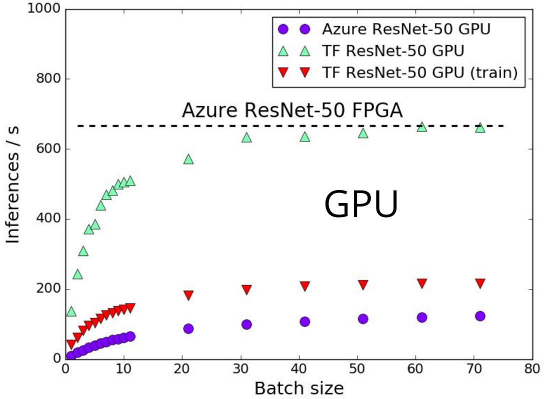
- Presentations include ACAT, CTD, HOW2019, IRIS-HEP topical meeting
- FastML workshop (Partly IRIS-HEP blueprint) @FNAL starting tomorrow
- Paper submitted to Computing and Software for Big Science

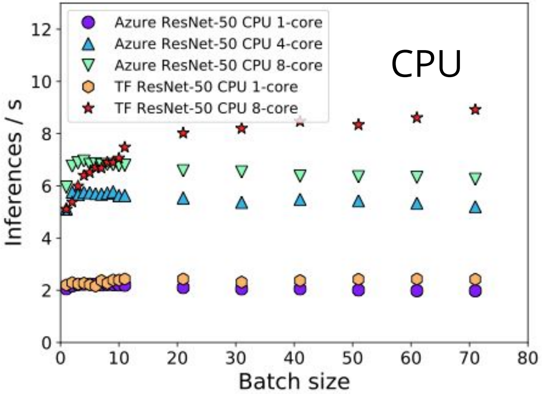
FPGA-accelerated machine learning inference as a service for particle physics computing

Javier Duarte · Philip Harris · Scott Hauck · Burt Holzman · Shih-Chieh Hsu · Sergo Jindariani · Suffian Khan · Benjamin Kreis · Brian Lee · Mia Liu · Vladimir Lončar · Jennifer Ngadiuba · Kevin Pedro · Brandon Perez · Maurizio Pierini · Dylan Rankin · Nhan Tran · Matthew Trahms · Aristeidis Tsaris · Colin Versteeg · Ted W. Way · Dustin Werran · Zhenbin Wu

https://arxiv.org/pdf/1904.08986.pdf

Duarte et al.





Applications for R&D and plan forward

Local calorimetric reconstruction to demonstrate physics-grade machine learning algorithms used in core online/offline reconstruction

- A major contribution to the overall HLT timing budget (15-20%, as algorithms run on essentially every event)
- Using CMS HCAL as initial example for developmentInitial evaluations are using NN regression algorithm to do cluster reconstruction

Current work has seeded two recent NSF awards (HDR, CSSI) to demonstrate FPGA use at scale and to broaden the set of ML algorithms easily ported to FPGAs

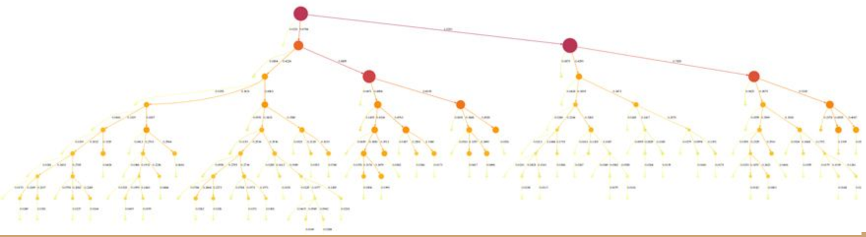
- IRIS-HEP deliverables to focus on demonstrating that the “physics” performance of ML approaches does (or does not) outperform that of current approaches in calorimetric reconstruction

ML4Jets (NYU)

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

Progress and plans:

- Co-organized KITP Conference (Feb 2019)
- Co-organizing Hammers & Nails Workshop (July 2019),
  IPAM Workshop (October, 2019), ML4Jets workshop (January 2020)
- Community engagement/workshops on topics such as
  - Fast simulation techniques for detector and reconstruction objects
  - Establishing/curating common metrics, datasets, and other ingredients for event reconstruction algorithm development. Eg. Top Tagging ↓



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

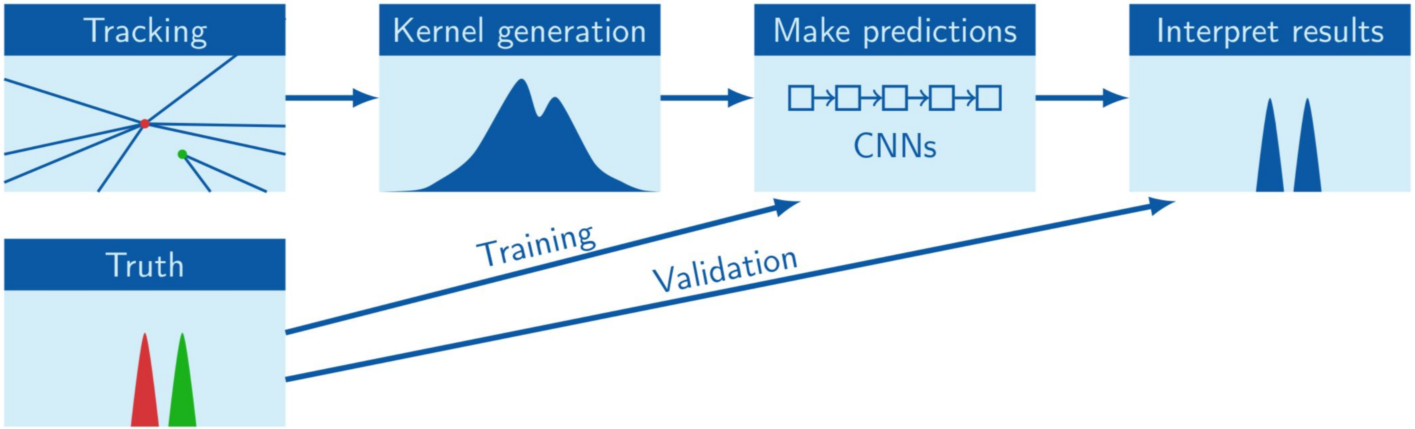
Tree Neural network approach demonstrated on reference dataset

	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

ML4Vertexing (Cincinnati, Princeton)

Develop novel primary vertex algorithm using hybrid Machine Learning

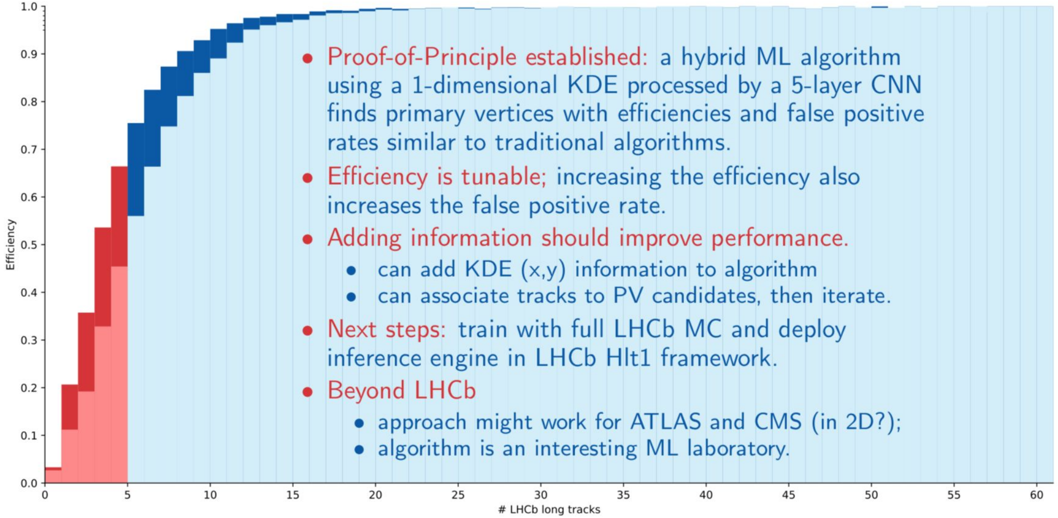
- Motivation: Run 3 luminosity increase for LHCb means that algorithms must be robust and efficient enough to find 5 vertices per event at 30 MHz data rate



ML4Vertexing: Progress and Plans

- Presentations at ACAT, CTD and HOW2019
- Initial software version released. Now working to fit into the LHCb software and production system environments
- Recent algorithmic improvements include using multidimensional information as would be necessary to adopt this approach in higher pileup environments

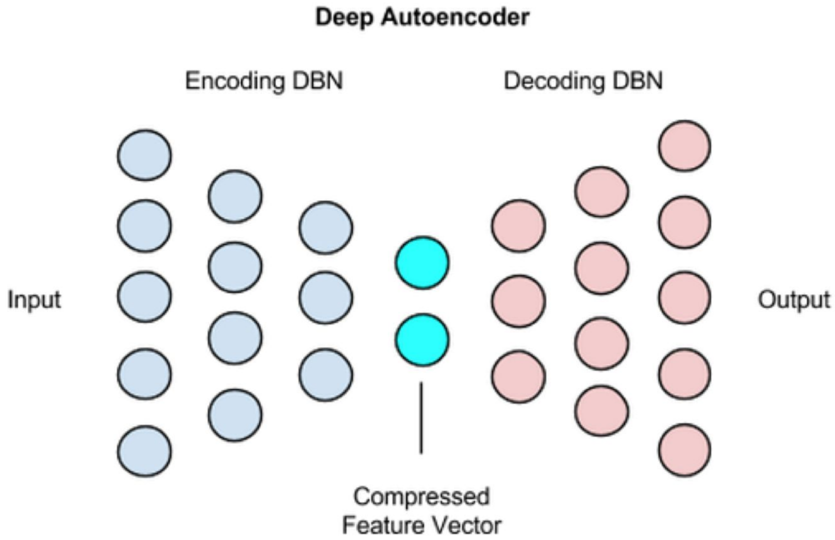
Schreiner et al.



FastPID (MIT)

Goal: Improve particle ID using machine learning techniques

- Current R&D aims to evaluate autoencoder approach for particle identification in LHCb environment
- **Collaborators**: Universite de Paris VI, Yandex School of Data Science



Progress and plans:

- Developed working version of a VAE
- Simulate PID distributions with good fidelity
- Working to document and release results this fall

Primary IA Goals for IRIS-HEP design phase

Novel algorithm demonstrations

- Effectiveness of GAN/autoencoder approach for PID
- Performance benchmarks for KalmanFilter in CMSSW for trigger/reconstruction
- Performance benchmarks for ACTS components on GPUs
- Identify promising operational scenarios FPGA use in reconstruction/HLT.
    - Performance assessments for FPGA-based reconstruction/HLT algorithms
- Effectiveness of machine learning track ambiguity resolution algorithms
- Assessment of parallel algorithm implementations for regionally based pattern recognition

Software products developed and released to HEP community

- Matriplex package release (**Now included in CMSSW via mkFit integration**)
- ML vertexing algorithm release (**Initial versions done**)
- ML on FPGAs release
- ACTS v0 release

Community engagement including workshops on tracking (CTD2020 @ Princeton), machine learning (ML4Jets @ NYU) and machine learning (FastML @ Fermilab)