

Innovative Algorithms

Area leads:

Heather Gray (UC-Berkeley/LBNL),

David Lange (Princeton)



From the CWP to IRIS-HEP

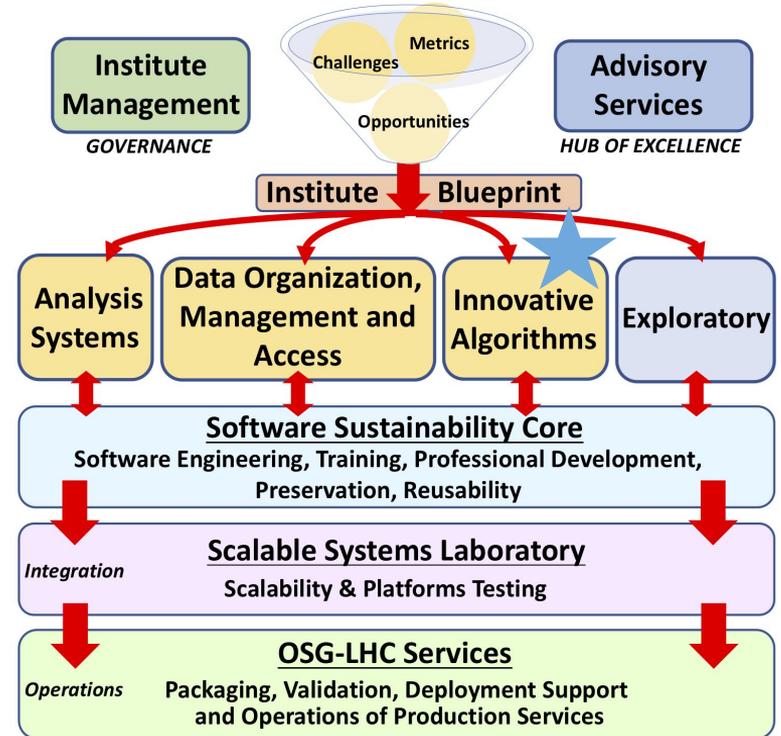
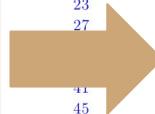
A Roadmap for HEP Software and Computing R&D for the 2020s

HEP Software Foundation¹

ABSTRACT: Particle physics has an aim for the coming decades. This program 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

1	Introduction	2
2	Software and Computing Challenges	5
3	Programme of Work	11
3.1	Physics Generators	11
3.2	Detector Simulation	15
★ 3.3	Software Trigger and Event Reconstruction	23
★ 3.4	Data Analysis and Interpretation	27
3.5	Machine Learning	
3.6	Data Organisation, Management and Access	
3.7	Facilities and Distributed Computing	41
3.8	Data-Flow Processing Framework	45
3.9	Conditions Data	48
3.10	Visualisation	51
3.11	Software Development, Deployment, Validation and Verification	53
3.12	Data and Software Preservation	58
3.13	Security	61
4	Training and Careers	65
4.1	Training Challenges	66
4.2	Possible Directions for Training	67
4.3	Career Support and Recognition	68
5	Conclusions	69
	Appendix A List of Workshops	72
	Appendix B Glossary	74
	References	80

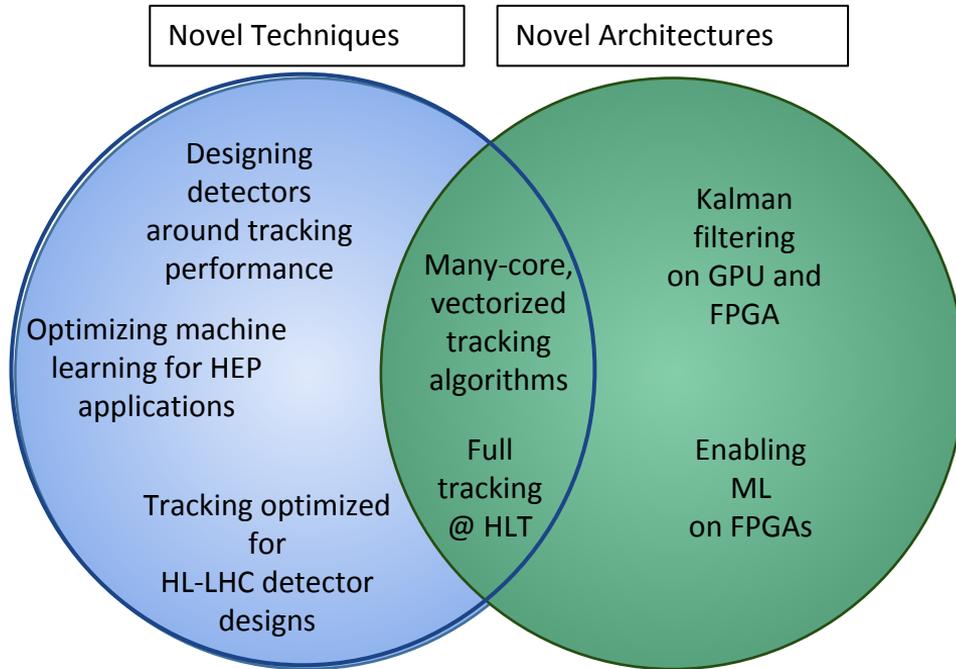


Scope of Innovative Algorithms (IA)

- 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

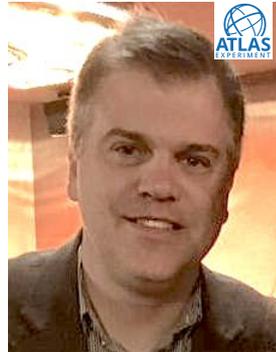
PIs in Innovative Algorithms



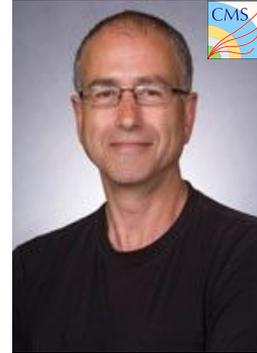
Peter Wittich, Cornell



Mike Williams, MIT



Mark Neubauer, UIUC



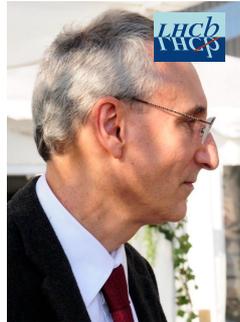
Avi Yagil, San Diego



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, PD TBD
- **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, Dylan Ranking, PD TBD
- **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



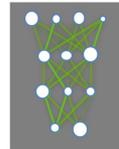
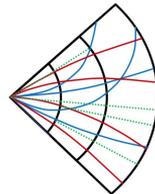
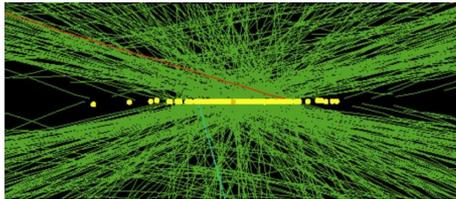
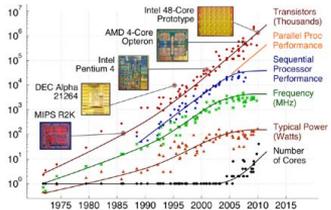
Groups are focused on answering 2 questions

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



FLEXIBILITY

EFFICIENCY

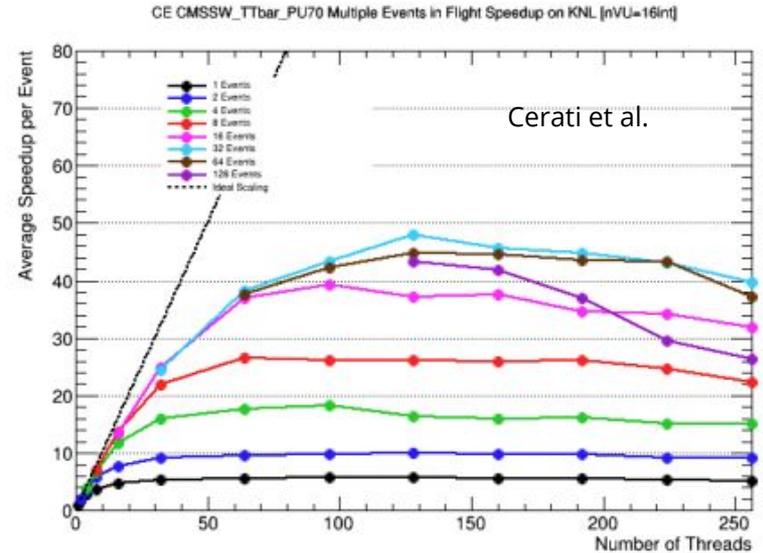
MKFit - Parallel Tracking (Cornell, Princeton, UCSD)

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

<http://trackreco.github.io/>

Collaborators:

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



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

MkFit: Progress and Plans

- Recent community presentations: ACAT conference, Connecting the Dots workshop, and IRIS-HEP topical meeting
- Primary focus is code integration with CMSSW
- Production release of Matriplex expected soon
- GPU evaluations underway
- Methods to streamline data conversions understudy

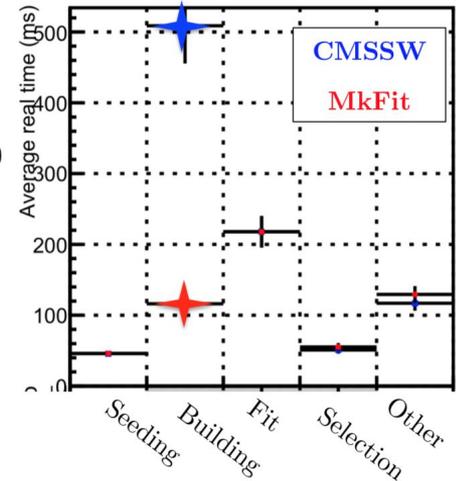
Integrated Timing Performance

Technical Details

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

Results

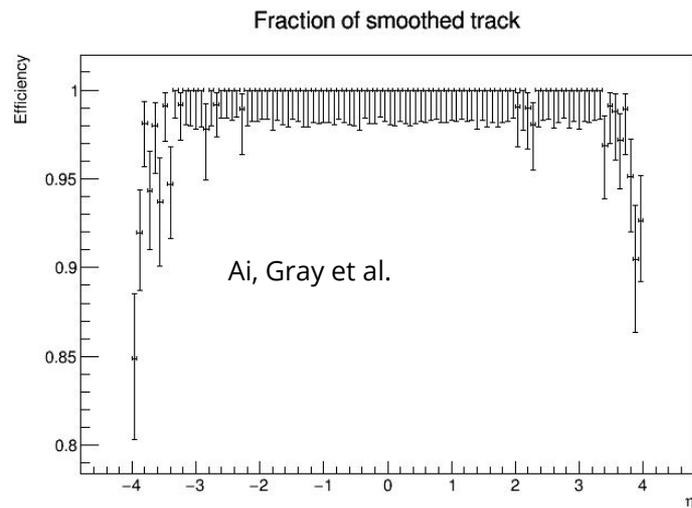
- Track building is **4.3x faster**
- 40% of time is spent in data format conversions – actual track finding is **7x faster**
- Track building now takes **less time than track fitting**
- Even larger potential speedups if multiple threads are used



Hall et al.
* Measured on SKL, mkFit compiled with AVX-512, turbo boost disabled

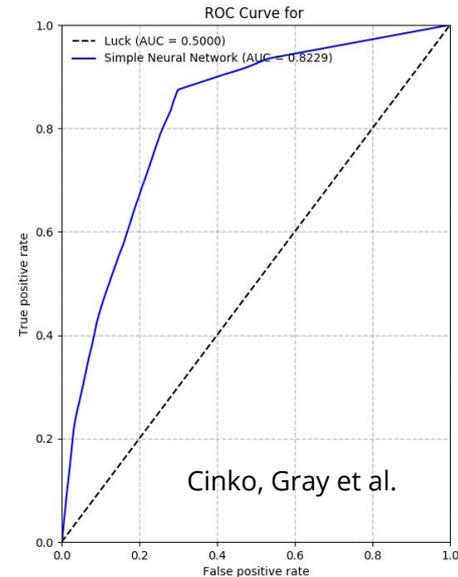
ACTS (UCB, Stanford): Overview

- Open-source software project for multi-experiment track reconstruction built on the extensive experience in track reconstruction in the ATLAS experiment
- Initial IRIS-HEP contributions: pattern recognition, ambiguity resolution, GPU demonstrators
- **Collaborators:** CERN, KIT, LBNL



ACTS (UCB, Stanford): Progress and Plans

- Hosted [Berkeley Tracking Workshop](#) (hackathon) in January 2019
- Kalman Fitter algorithm implemented: performance studies underway
- Ambiguity resolution algorithm implemented: Studies using ML approach are ongoing
- Track following implementation in progress
- GSOC19 student will study alignment in test beam geometries this summer



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



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

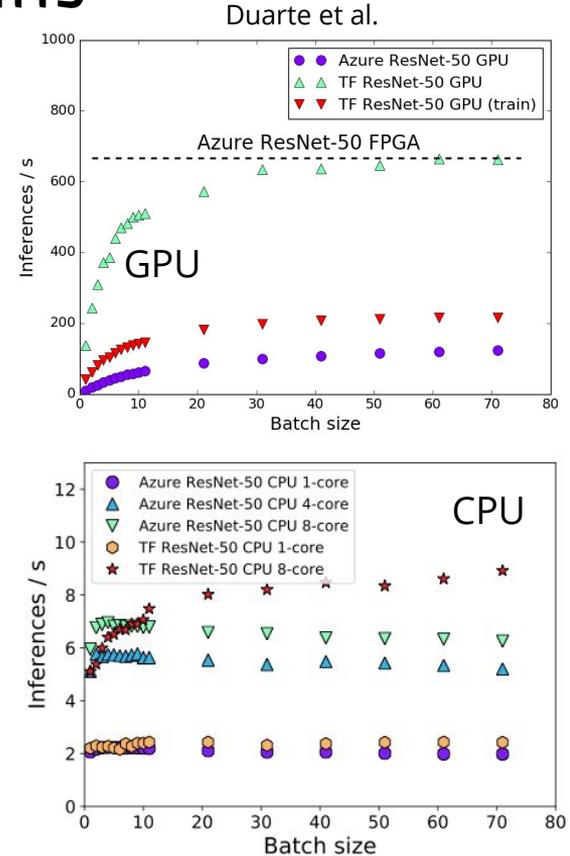
ML on FPGAs: Progress and Plans

- Recent IRIS-HEP topical meeting presentation
- ACAT, CTD and HOW2019 presentations
- 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 · Sergio 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>

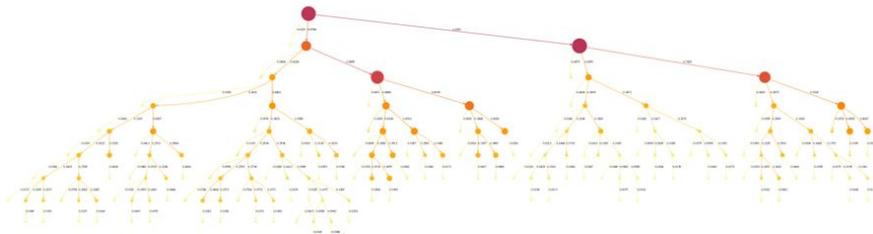


ML4Jets (NYU)

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

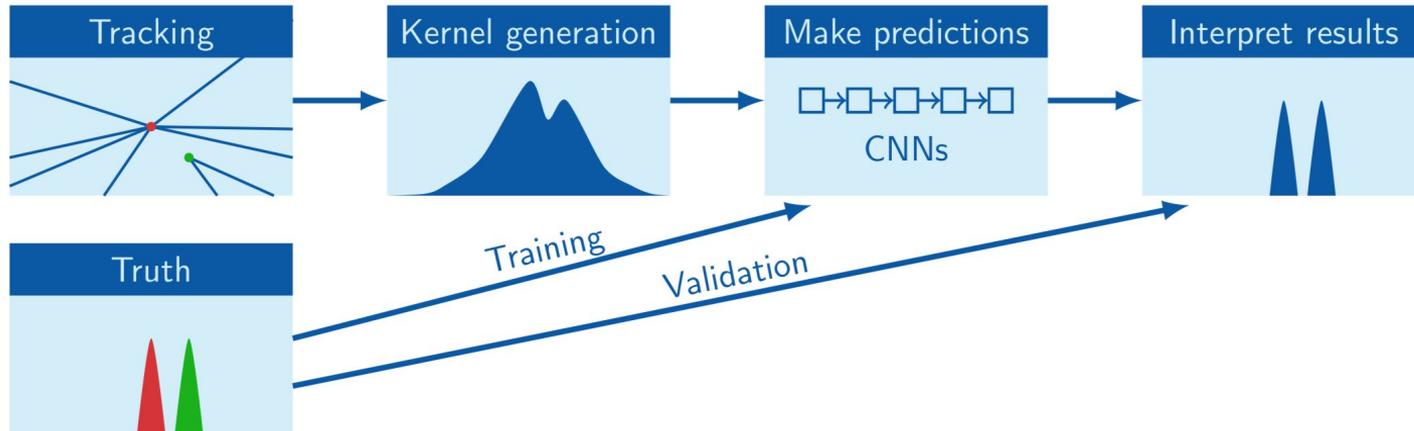
G. Kasieczka (ed)¹, T. Plehn (ed)², A. Butter², K. Cranmer³, D. Debnath⁴, M. Fairbairn⁵,
W. Fedorko⁶, C. Gay⁶, L. Gouskos⁷, P. T. Komiske⁸, S. Leiss¹, A. Lister⁶, S. Macaluso^{3,4},
E. M. Metodiev⁸, L. Moore⁹, B. Nachman^{10,11}, K. Nordström^{12,13}, J. Pearkes⁶, H. Qu⁷,
Y. Rath¹⁴, M. Rieger¹⁴, D. Shih⁴, J. M. Thompson², and S. Varma⁵

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

ML4Vertexing (Cincinnati, Princeton)

Develop novel primary vertex algorithm using hybrid Machine Learning

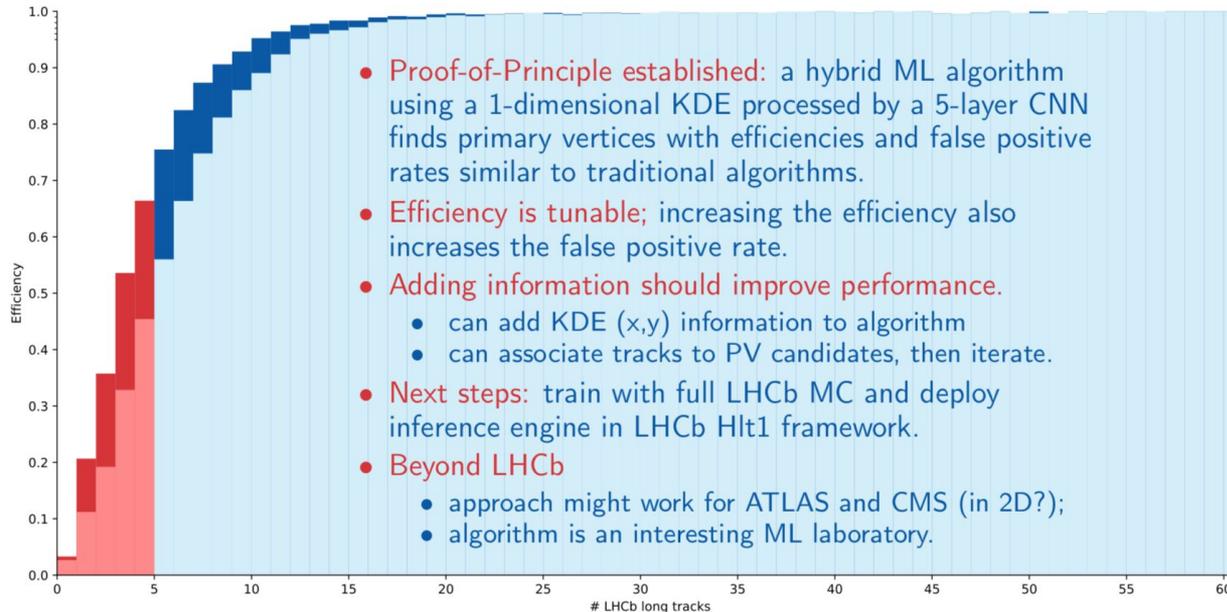
Motivation: Run 3 luminosity increase for LHCb means that algorithms must be robust and efficient enough to find 5 vertices per event at 30 MHz data rate



ML4Vertexing: Progress and Plans

- Presentations at ACAT, CTD and HOW2019
- Software released. Now evaluating how to fit into the LHCb software and production system environments

Schreiner et al.



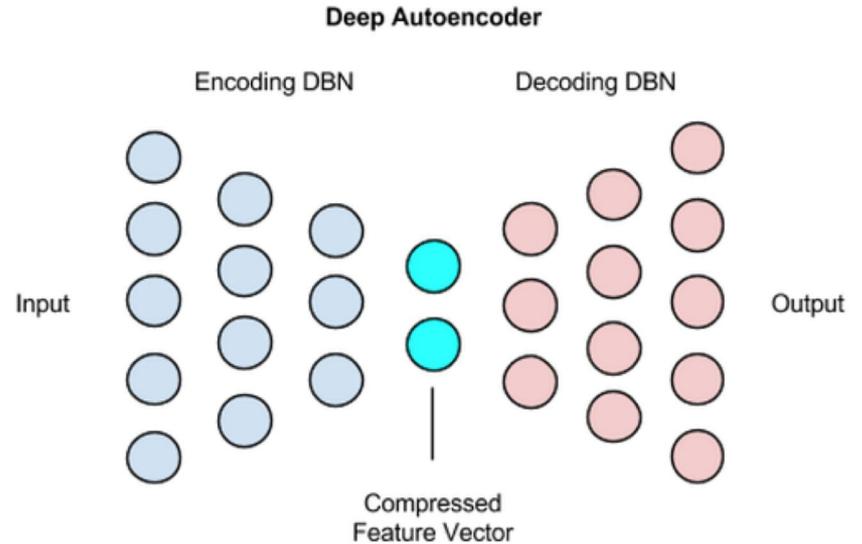
FastPID (MIT)

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



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
- ML vertexing algorithm release
- ML on FPGAs release
- ACTS v1 release

Community engagement including hosting workshops on tracking (CTD2020 @Princeton and machine learning (ML4Jets2020 @NYU)