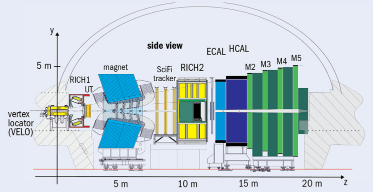
PV-finder

Rocky Garg, Michael D Sokoloff, Lauren Tompkins
(and others)

September 12, 2023

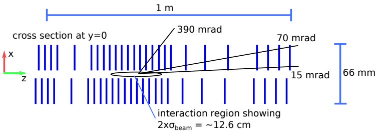
PV-finder: find primary vertices using machine learning

The Run 3 LHCb Detector



- **Primary Goal:** develop and deploy a machine learning algorithm to locates primary vertices (PVs) with high effficiency, low false positive rates, and good spatial resolution.
- **Constraints:** the algorithm should execute in both the first level software trigger (inside the purely GPU framework, `Allen`, and the second level x86 CPU trigger).

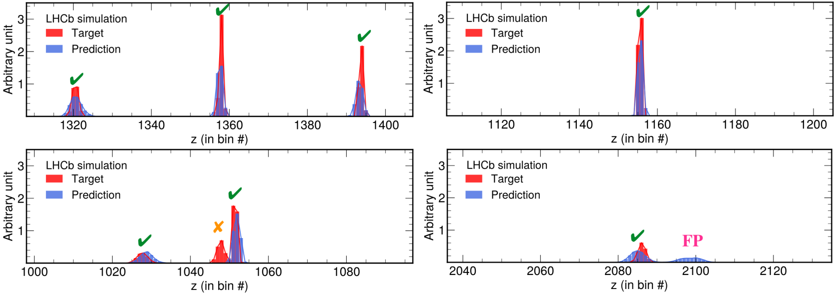
Primary Vertex Cartoon



- **Method:** tracks are reconstructed from raw hits and parameterized in terms of the error ellipsoids around their points of closest approach to the beamline (their `poca-ellipsoids`). A multistage DNN algorithm produces histograms that are interpreted to identify and locate PV candidates.
- **Challenges:** include congestion due to multiple, near-by PVs and secondary vertices (SVs) near PVs.

Longer Term Goal: adapt algorithm for use by ATLAS and CMS.

Example Histograms
LHCb

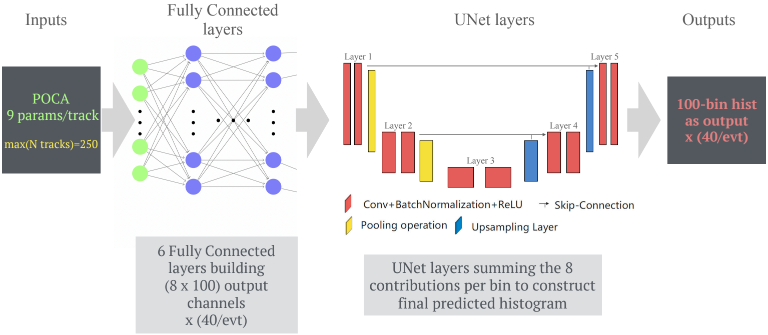


- target histograms are Gaussian distributions with heights and widths calculated from "expected" resolution (proxy PVs);
- predicted histograms are produced by the PV-finder inference engine.

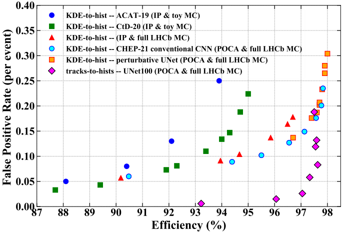
PV-finder: a new architecture for LHCb

end-to-end DNN, train using 40 × 10 mm intervals (LHCb simulation)

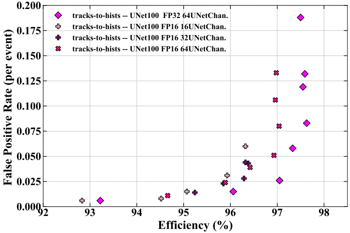
In the architecture illustrated below
- the fully connected layers are first trained as a `tracks-to-kde` modely;
- the UNet layers are similar to a `kde-to-hist` model used in the past;
- the merged model is trained with all weights floating for FP32 arithmetic.



U-Net: Convolutional Networks for Biomedical Image Segmentation



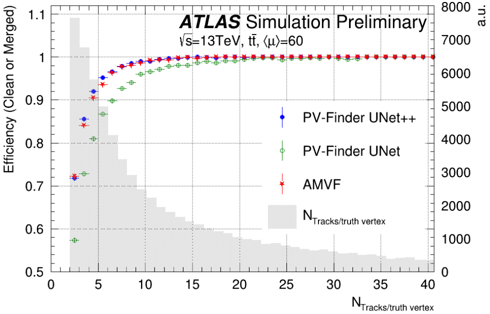
Efficiency versus false positive rate for older `KDE-to-hist` networks and for a new `tracks-to-hist` model shown in magenta diamonds. A tunable asymmetry parameter in the cost function defines the observed trajectories.



Efficiency versus false positive rate for variations of the `tracks-to-hist` model. The magenta diamonds are as in the figure on the left. The other models use `FP16` arithmetic in place of `FP32` arithmetic and smaller CNNs.

ATLAS results circa CHEP-2023

using `KDE-to-hist` models similar to LHCb's CHEP-2021 model



Efficiency: Number of truth vertices assigned to reconstructed vertices as "clean" or "merged" divided by the total number of reconstructable truth vertices

False Positive Rate: Average number of predicted vertices not matched to any truth vertex

Model	Efficiency	False Positive Rate (# / event)
PV-Finder UNet++	94.2%	1.5
PV-Finder UNet	88.7%	2.6
AMVF	93.9%	0.8

ATLAS Resolution in z: PV-finder vs. AMVF

longitudinal separation between pairs of all nearby reconstructed primary vertices

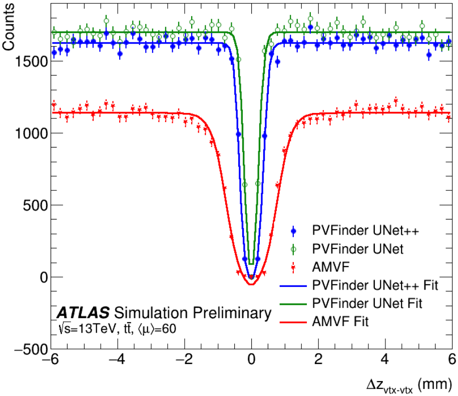
Fit Function:

$$y = \frac{a}{1 + \exp\left(b \cdot (R_{cc} - |x|)\right)} + c$$

where a, b and c are free parameters, R_{cc} is the cluster-cluster resolution, referred to as σ_{vtx}

Method	σ_{vtx} (mm)
PV-Finder UNet	0.23 ± 0.01
PV-Finder UNet++	0.37 ± 0.01
AMVF	0.76 ± 0.02

σ_{vtx}: half-width at the half-depth of the dip



see the ATLAS internal note ATL-PHYS-PUB-2023-011 for details.

Cincinnati Plans for Year 1

People working on developing PV-finder for LHCb and deploying it in `Allen`:

- IRIS-HEP personnel in Cincinnati: Simon Akar (post-doc), Mohamed Elashri (Ph.D. student), Conor Henderson (faculty), and Mike Sokoloff (faculty);
- Other Cincinnati personnel: Tom Boettcher (post-doc) and Sara Shinde (ugrad);
- Non-Cincinnati LHCb researchers: Roel Aaij (Nikhef), Daniel Cámpora (Maastricht), Dorothea vom Bruch (Marseille), and Maarten von Veghel (Nikhef) plus personnel from LNPHE (Paris).

The primary goals of the IRIS-HEP personnel are

- Mohamed will build deep neural network inference engines that will execute the pv-finder family of deep learning algorithms (and others) inside `Allen`, LHCb's GPU-resident first level software trigger. He will benchmark performance of variations of implementations including how kernels are organized, whether the convolutional layers are executed in tensor cores or CUDA cores, and the extent to which using `FP16` and `int8` quantization affects throughput and performance. He will collaborate with Tom and our non-Cincinnati colleagues.
- Simon will re-factorize the software for preparing training and validation data for the existing pv-finder family of algorithms and for defining models. He will systematically study the performances of the models at the PyTorch level, study the trade-offs between model fidelity and throughput (collaborating with Mohamed), and investigate new architectures for pv-finder using the PyTorch Geometric package. He will develop machine learning algorithms for associating individual tracks with primary vertices found by pv-finder and identify tracks that are likely to be secondary tracks. He will collaborate with LPNHE colleagues studying GNN models.

Stanford PV-finder Plans for Year 1

Personnel working on developing PV-finder for ATLAS and ACTS are:

- Rocky Garg (Stanford post-doc);
- Lauren Tompkins (Stanford faculty);
- Layan AlSarayra (IRIS-HEP Fellow);
- Ananya Singha (HSF-India Fellow);
- Cincinnati IRIS-HEP personnel, as consultants.

The primary goals for this work are

- Prepare an end-to-end `tracks-to-hist` network in the ATLAS framework and compare results with ATLAS vertex finder (AMVF).
- Prepare a `KDE-to-hist` model in the ACTS framework using heuristic KDEs (as is currently done in ATLAS framework) and compare results with ACTS-AMVF.