Found 103002 of 109733 (eff 93.87%)  
False positive rate = 0.251 per event  
Asymmetric cost function

Found 96616 of 109733 (eff 88.05%)  
False positive rate = 0.0485 per event  
Symmetric cost function

Events in sample = 20K  
Training sample = 240K

ML for PV Reconstruction

Fang, Schreiner, Sokoloff, Weisser, Williams

March 20, 2019
The changes

- 30 MHz software trigger
- 7.6 PVs per event (Poisson distribution)
- Roughly 5.5 visible PVs per event

The problem

- Much higher pileup
- Very little time to do the tracking
- Current algorithms too slow

We need to rethink our algorithms from the ground up...
Vertices and tracks

**Introduction**

**Vertices**

- Events contain $\approx 7$ Primary Vertices ($\approx 5$ visible PVs)
  - A PV should contain 5+ long tracks
- Multiple Secondary Vertices (SVs) per event as well
  - A SV should contain 2+ tracks

**Adapt to machine learning?**

- Sparse 3D data (41M pixels) $\rightarrow$ rich 1D data
- 1D convolutional neural nets
- Highly parallelizable, GPU friendly
- Opportunities to visualize learning process
A hybrid ML approach

Introduction

Machine learning features (so far)

- Prototracking converts sparse 3D dataset to feature-rich 1D dataset
- Easy and effective visualization due to 1D nature
- Even simple networks can provide interesting results

Training

Validation
Tracking procedure

- Hits lie on the 26 planes
- For simplicity, only 3 tracks shown
Tracking procedure

- Hits lie on the 26 planes
- For simplicity, only 3 tracks shown
- Make a 3D grid of voxels (2D shown)
- Note: only $z$ will be fully calculated and stored

![Graph showing tracking procedure](image-url)
Tracking procedure

- Hits lie on the 26 planes
- For simplicity, only 3 tracks shown
- Make a 3D grid of voxels (2D shown)
- Note: only $z$ will be fully calculated and stored
- Tracking (full or partial)

---

$z$ axis (along the beam)
Tracking procedure

- Hits lie on the 26 planes
- For simplicity, only 3 tracks shown
- Make a 3D grid of voxels (2D shown)
- Note: only $z$ will be fully calculated and stored
- Tracking (full or partial)
- Fill in each voxel center with Gaussian PDF
Tracking procedure

- Hits lie on the 26 planes
- For simplicity, only 3 tracks shown
- Make a 3D grid of voxels (2D shown)
- Note: only $z$ will be fully calculated and stored
- Tracking (full or partial)
- Fill in each voxel center with Gaussian PDF
- PDF for each (proto)track is combined

$z$ axis (along the beam)
Tracking procedure

- Hits lie on the 26 planes
- For simplicity, only 3 tracks shown
- Make a 3D grid of voxels (2D shown)
- Note: only z will be fully calculated and stored
- Tracking (full or partial)
- Fill in each voxel center with Gaussian PDF
- PDF for each (proto)track is combined
- Fill z "histogram" with maximum KDE value in xy
Example of $z$ KDE histogram

Note: All events from toy detector simulation

**Human learning**
- Peaks generally correspond to PVs and SVs

**Challenges**
- Vertex may be offset from peak
- Vertices interact
Target distribution

Build target distribution

- True PV position as the mean of Gaussian
- $\sigma$ (standard deviation) is 100 $\mu$m (simplification)
- Fill bins with integrated PDF within $\pm 3$ bins ($\pm 300 \mu$m)
## Neural network architecture

### Inputs

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>25</th>
<th>26</th>
<th>...</th>
<th>3998</th>
<th>3999</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convolution Width: 25 Channels: 1 → 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 25 Channels

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>15</th>
<th>16</th>
<th>...</th>
<th>3998</th>
<th>3999</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convolution Width: 15 Channels: 25 → 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Convolution

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>15</th>
<th>16</th>
<th>...</th>
<th>3998</th>
<th>3999</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convolution Width: 15 Channels: 25 → 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Convolution

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>4</th>
<th>5</th>
<th>...</th>
<th>91</th>
<th>92</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convolution Width: 91 Channels: 1 → 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Output

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>3997</th>
<th>3998</th>
<th>3999</th>
<th>4000</th>
</tr>
</thead>
</table>

### Activation Functions

- Leaky relu
- Leaky relu
- Leaky relu
- Leaky relu
- Softplus

---

**ML for PV Reconstruction**

Fang, Schreiner, Sokoloff, Weisser, Williams

March 20, 2019 8/14
Approach

- Symmetric cost function: low FP but low efficiency
- Adding asymmetry term controls trade-off for FP vs. efficiency
False Positive and efficiency rates

Results

Search for PVs
- Search ±5 bins (±500µm) around a true PV
- At least 3 bins with predicted probability > 1% and integrated probability > 20%.

Tunable efficiency vs. FP
- The asymmetry parameter controls FP vs. efficiency

ML for PV Reconstruction
Fang, Schreiner, Sokoloff, Weisser, Williams
March 20, 2019 10/14
Compare predictions with targets: When it works

Results

Comparing predictions with targets: When it works

**PV found example**

- True: 48.904 mm
- Pred: 48.954 mm
- Δ: 50 µm

**Masked (<5 tracks) example**

- Pred: 0.976 mm

Event 0 @ 48.9 mm: PV found

Event 0 @ 1.0 mm: Masked

ML for PV Reconstruction

Fang, Schreiner, Sokoloff, Weisser, Williams

March 20, 2019
Compare predictions with targets: When it fails

Results

False Positive example

Event 2 @ 65.7 mm: False positive

Pred: 65.696 mm

Kernel Density

Event 3 @ 51.9 mm: PV not found

True: 51.898 mm

Kernel Density

PV not found example

ML for PV Reconstruction

Fang, Schreiner, Sokoloff, Weisser, Williams

March 20, 2019  12/14
Proof-of-Principle established: a hybrid ML algorithm using a 1-dimensional KDE processed by a 5-layer CNN finds primary vertices with efficiencies and false positive rates similar to traditional algorithms.

Efficiency is tunable; increasing the efficiency also increases the false positive rate.

Adding information should improve performance.
- can add KDE (x,y) information to algorithm
- can associate tracks to PV candidates, then iterate.

Next steps: train with full LHCb MC and deploy inference engine in LHCb Hlt1 framework.

Beyond LHCb
- approach might work for ATLAS and CMS (in 2D?);
- algorithm is an interesting ML laboratory.
Questions?

Source code:

- [https://gitlab.cern.ch/LHCb-Reco-Dev/pv-finder](https://gitlab.cern.ch/LHCb-Reco-Dev/pv-finder)
- Runnable with Conda on macOS and Linux
  - Run: `conda env create -f environment-gpu.yml`
  - Python 3.6+ and PyTorch used for machine learning code
  - Generation now available too using the new Conda-Forge ROOT and Pythia8 packages

Supported by:

- NSF OAC-1836650: IRIS-HEP
- NSF OAC-1740102: SI2:SSE
- NSF OAC-1739772: SI2:SSE
Event 2 @ 114.6 mm: PV found

- True: 114.622 mm
- Pred: 114.597 mm
- $\Delta$: -26 $\mu$m

Event 5 @ 197.4 mm: PV found

- True: 197.461 mm
- Pred: 197.396 mm
- $\Delta$: -65 $\mu$m

 Kernel Density

Target
Predicted
Masked

xy maximum [m]

z values [mm]

ML for PV Reconstruction

Fang, Schreiner, Sokoloff, Weisser, Williams

March 20, 2019  15/14
More predictions with targets (2)

Event 5 @ 221.5 mm: PV found
- True: 221.595 mm
- Pred: 221.546 mm
- Δ: -49 µm

Event 6 @ 36.1 mm: PV found
- True: 36.068 mm
- Pred: 36.400 mm
- Δ: 332 µm
Event 6 @ 129.3 mm: PV found

- True: 129.336 mm
- Pred: 129.337 mm
- Δ: 1 µm

Event 6 @ 143.2 mm: PV found

- True: 143.224 mm
- Pred: 143.199 mm
- Δ: -25 µm
The VELO

Tracks

- Originate from vertices (not shown)
- Hits originate from tracks
- We only know the true track in simulation
- Nearly straight, but tracks may scatter in material

The VELO

- A set of 26 planes that detect tracks
- Tracks should hit one or more pixels per plane
- Sparse 3D dataset (41M pixels)
- Beam width \((x, y)\): 40 \(\mu m\) for LHCb, what is yours?
- Transverse resolution: 5–15 \(\mu m\) for LHCb depending on number of tracks, what is yours?
- Longitudinal resolution: 40–100 \(\mu m\) for LHCb depending on number of tracks, what is yours?
- Cleaning up prototracks based on IP could simplify kernel
- Can prototracking be done in the triggers?