Track Finding in the COMET Experiment Using Boosted Decision Trees

Ewen Lawson Gillies

Imperial College London
High Energy Particle Physics

Connecting the Dots 2018
March 21st, 2018

Overview

COMET is a next generation, high intensity experiment looking for new physics.

1. New Physics: Charged Lepton Flavor Violation
2. New Designs: The **Co**herent **M**uon to **E**lectron **T**ransition (COMET) experiment
3. New Techniques: **G**radient **B**oosted **D**ecision **T**rees (GBDT) and Hough Transforms in Track Finding

New Physics & CLFV

Lepton Flavor Violation

Lepton flavor is conserved in the Standard Model.

Muon Decay: $\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e$
Muon Capture: $\mu^- + N \rightarrow \nu_\mu + N'$

Do the charged leptons, (τ, μ, e), violate this conservation law of the Standard Model?

Current Experimental Limits

Charged Lepton Flavor Conservation has been tested for decades. Upper limits for muonic search channels:

- $\text{Br}(\mu^+ \to e^+ + e^+ + e^-) < 1.0 \times 10^{-12}$ (SINDRUM 1988)
- $\text{Br}(\mu^+ \to e^+ + \gamma) < 4.2 \times 10^{-13}$ (**MEG 2016**)
- $\text{B}(\mu^- + \text{Au} \to e^- + \text{Au}) < 7 \times 10^{-13}$ (SINDRUM II 2006)

COMET focuses on muon to electron conversion. Without CLFV, this process can only come indirectly with processes involving neutrinos:

$$\text{B}(\mu^- + N \to e^- + N) \sim 10^{-52}$$

In 2018, COMET Phase I aims to achieve the sensitivity of:

$$\text{B}(\mu^- + \text{Al} \to e^- + \text{Al}) < 7.2 \times 10^{-15}$$

Signal Process: μ-e Conversion

μ-e Conversion:

$$\mu^- + N(A,Z) \rightarrow e^- + N(A,Z)$$

Momentum of Signal Electron:

$$E_e = m_\mu - B_\mu - E_{\mathrm{recoil}}$$

For Aluminum (COMET):

$$E_e = 104.9\ \mathrm{MeV}$$

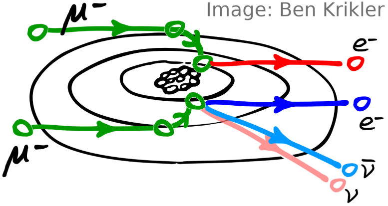
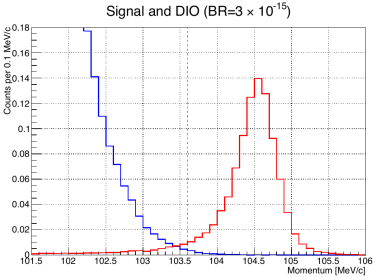


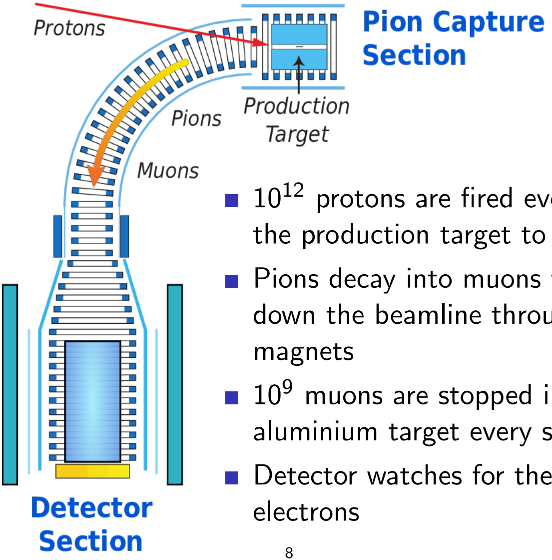
Image: Ben Krikler



COMET Design Principles

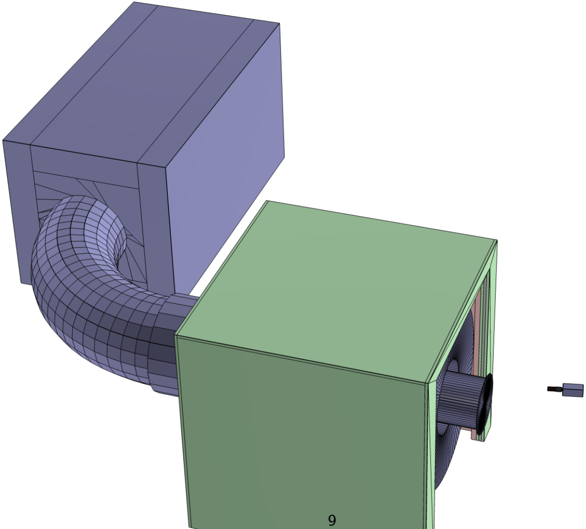
COMET Phase I Design

Yandex Data Factory
Imperial College London

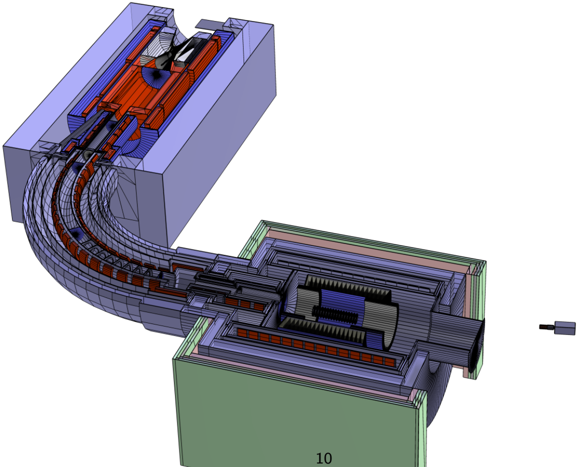


- 10^{12} protons are fired every second at the production target to produce pions
- Pions decay into muons while flying down the beamline through curved magnets
- 10^{9} muons are stopped in the aluminium target every second
- Detector watches for the 105 MeV electrons

Phase I Geometry



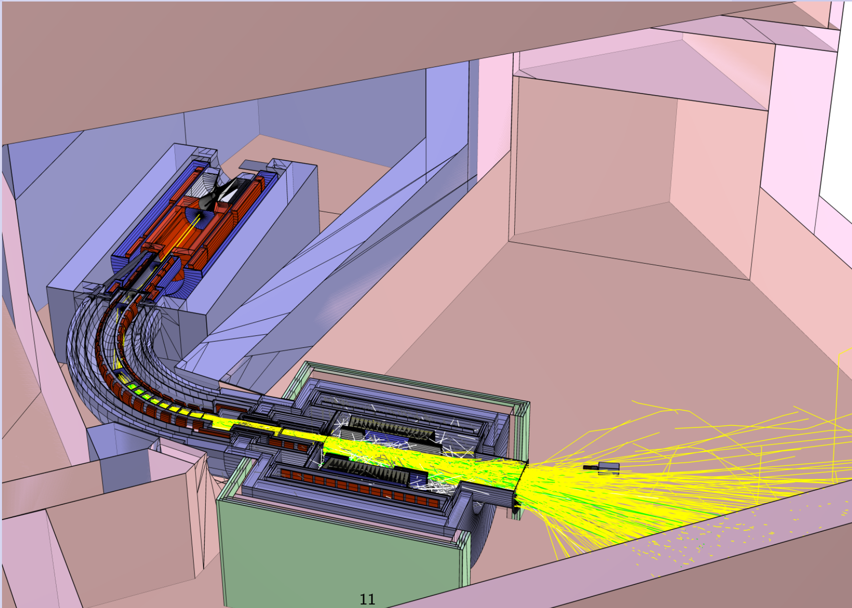
Phase I Geometry



Phase I Geometry

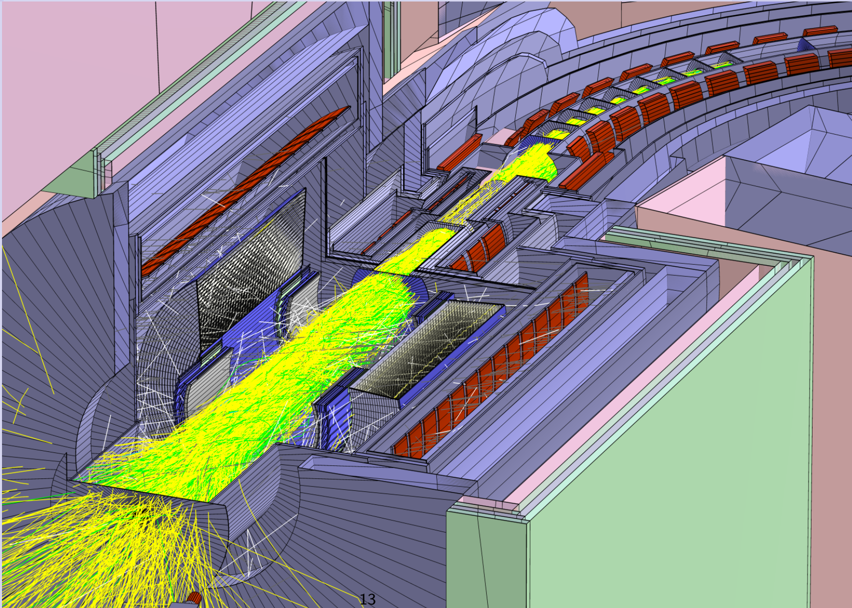
Decisions and COMET

Ewen Gillies

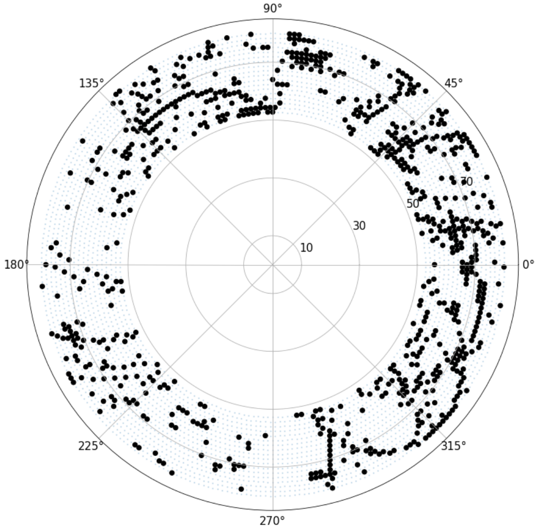


New Tracking Techniques

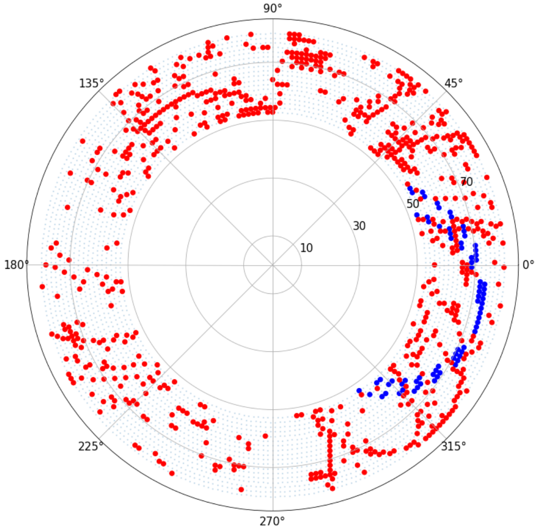
Cylindrical Detector



Typical Event [1]



Typical Event [2]



Classification Problem

"Is this wire a signal hit from a signal track". Algorithm developed with Dr. Alex Rogozhnikov when he was at Yandex.

Hit wires have three main features

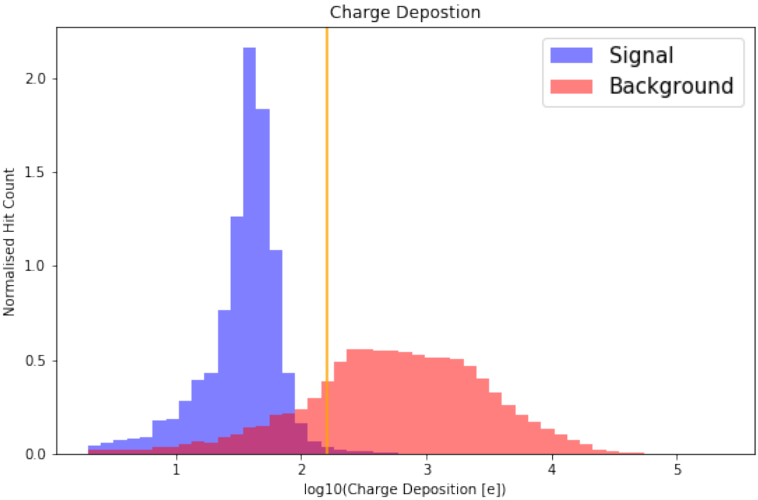
- Radial distance from centre.
- Energy deposited by charged particle.
- Timing of energy deposition.

Define categories of features:

1. "Local" Features: Features on the wire itself
2. "Neighbour" Features: Features of adjacent wires
3. "Shape" Features: Check if the wire forms a circle with other hit wires

Previous Classifier: Cutting on Charge

Cut removes 80% of background while keeping 99% of signal.



Neighbour-Level GBDT

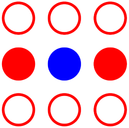
Local and Neighbour Features

Yandex Data Factory

Imperial College London

Signal hits are often grouped in local clusters, meaning neighbouring wire features are extremely important.

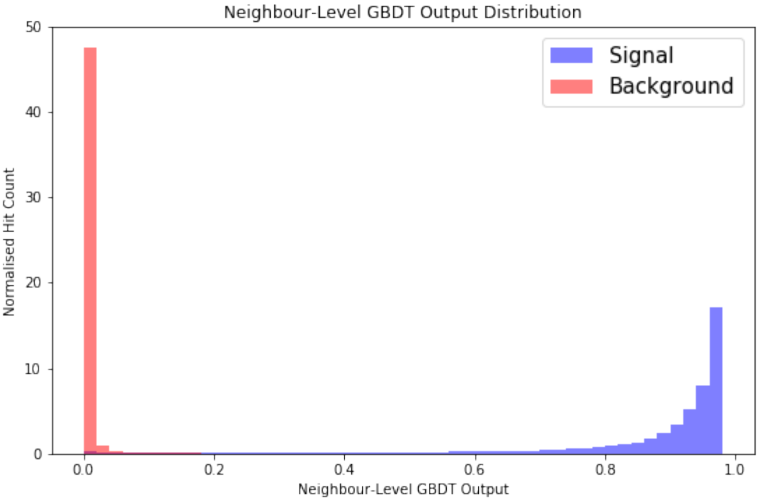
Neighbour-Level Features:



- Radial distance from centre, same for wire and LR neighbours (1 feature)
- Energy deposited on wire, left neighbour, and right neighbour (3 features).
- Timing of hit on wire, left neighbour, and right neighbour (3 features).

Neighbour-Level GBDT Output Distribution

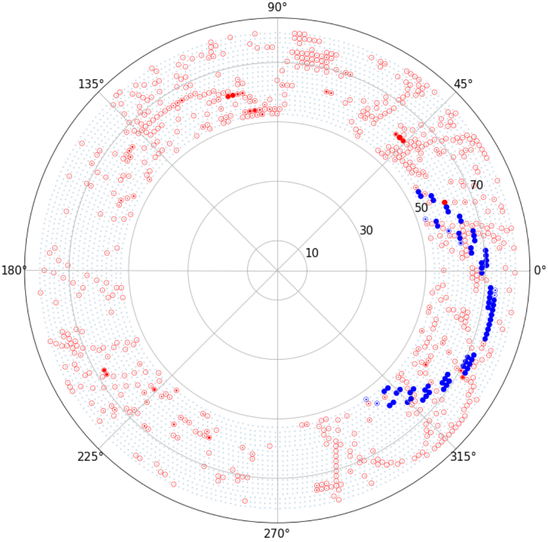
Output of GBDT trained on local and neighbour features.



Output of Neighbour-Level GBDT

Yandex Data Factory
Imperial College London

- Open circles are original hit locations
- **Signal Hits** and **Background Hits** are scaled to the output of the neighbour level GBDT.



Hough Transform

Circular Hough Transform

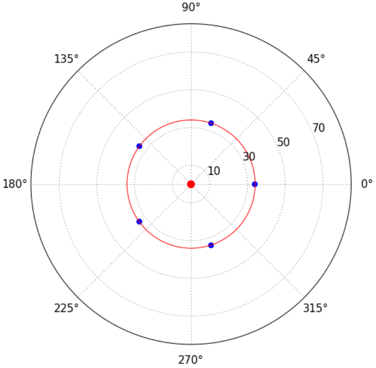


Figure 1: Points in (x, y) space, blue, thought to be on a circle, red, whose centre lies at the origin, orange.

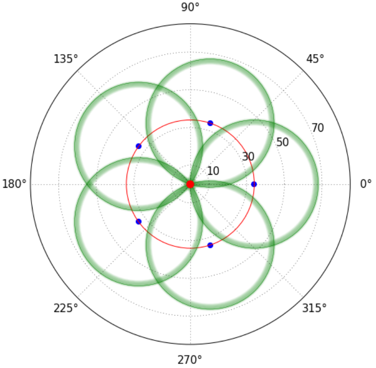
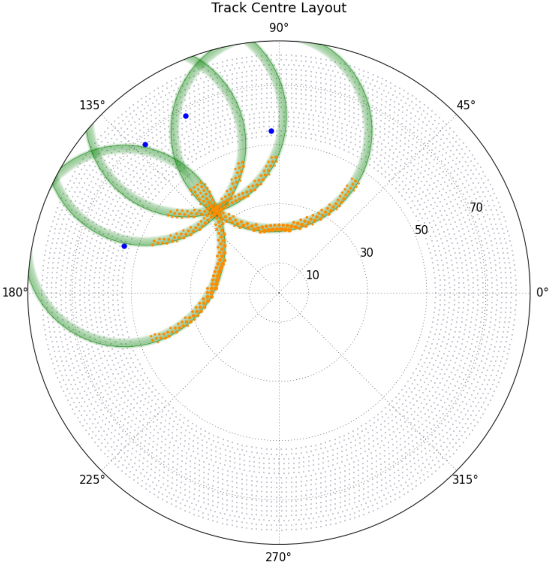


Figure 2: A mapping from the points in (x, y) space, blue, to possible circle centers in (a, b) space, green.

Hough Implementation

Yandex Data Factory
Imperial College London

- **Hits** with corresponding hough contributions
- **Track centers** scaled by contributions from hit points.



Executing the Hough Transform

Weight wire j's contribution by its GBDT output:

$$W_j = y_{\text{Grad.}}\left(f_1^{(j)}, \ldots, f_N^{(j)}\right) \text{ for N features}$$

Apply hough transform between wire j and track center i:

$$\underbrace{T_{ij}}_{\text{Hough}} \overbrace{W_j}^{\text{GBDT Score}} = \underbrace{C_i}_{\text{Track centre}}$$

Reweight the results to highlight maxima:

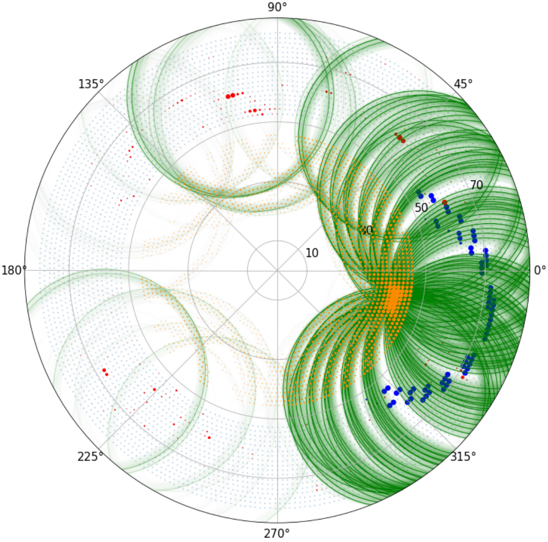
$$C_i \to C_i'(\alpha) = \exp(\alpha C_i)$$

Invert the transform:

$$\underbrace{(T_{ij})^{-1}}_{\text{Inv. Hough}} \overbrace{C_i'}^{\text{Reweighted Track centers}} = \underbrace{W_j'}_{\text{New Feature}}$$

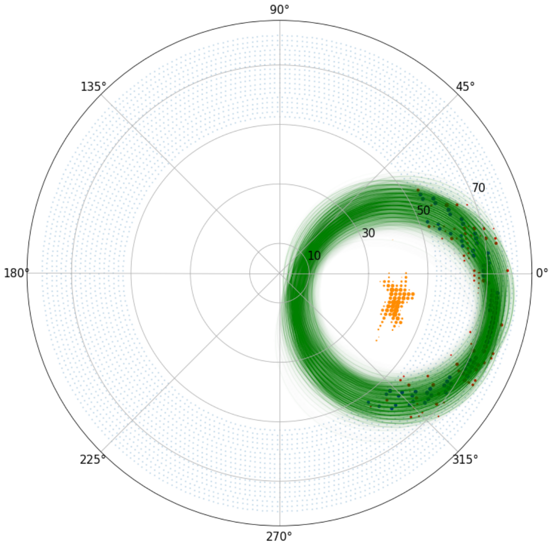
Forward Hough Transform on Neighbour Level GBDT Output

- Signal hits scaled by neighbour level GBDT output W_j.
- Background hits also scaled by W_j.
- Hough transform scaled by W_j of corresponding hit.
- Track centers scaled by C_i from $C_i = T_{ij} W_j$.



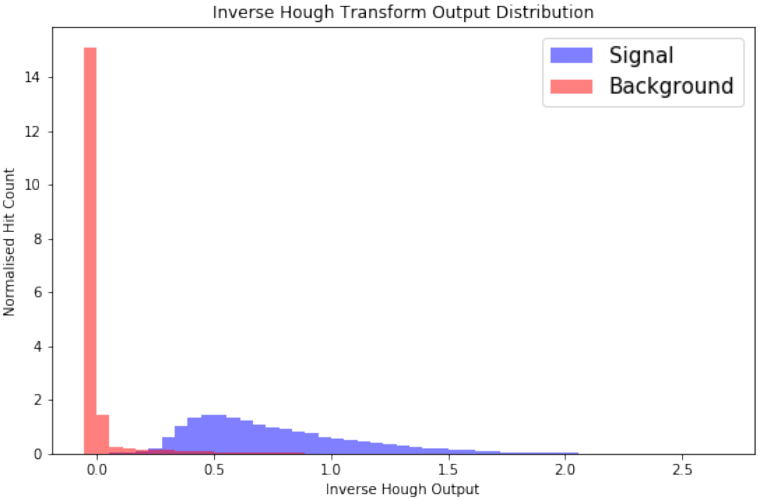
Inverse Reweighted Hough Transform Output

- Signal hits scaled by reweighted inverse Hough output W'_j.
- Background hits scaled by W'_j.
- Track centers scaled by C'_i.
- Inverse Hough transform scaled by C'_i of corresponding centre.



Inverse Hough Output Feature

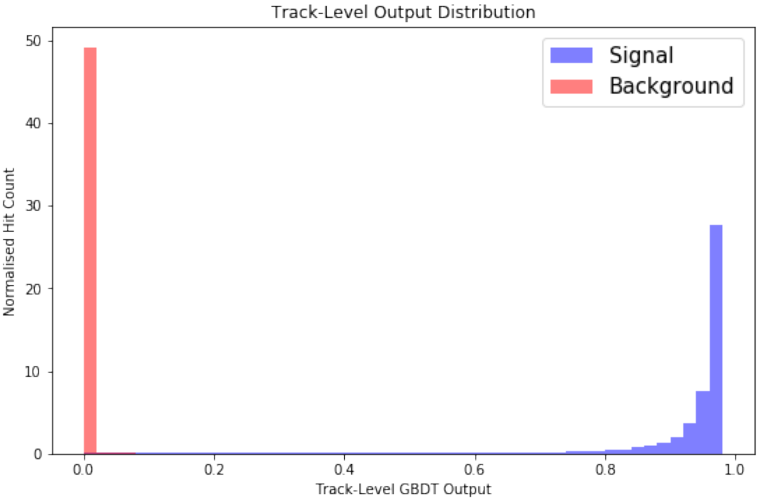
New feature rewards hits that form a track with signal-like hits.



Track-Level GBDT

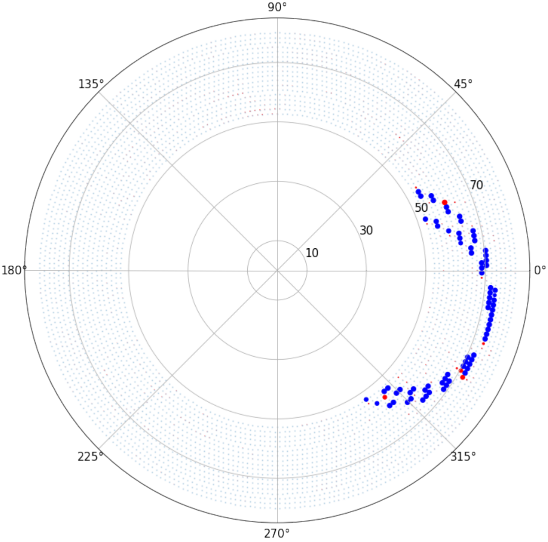
Track-Level GBDT Output Distribution

GBDT trained on local, neighbour, and new track features.



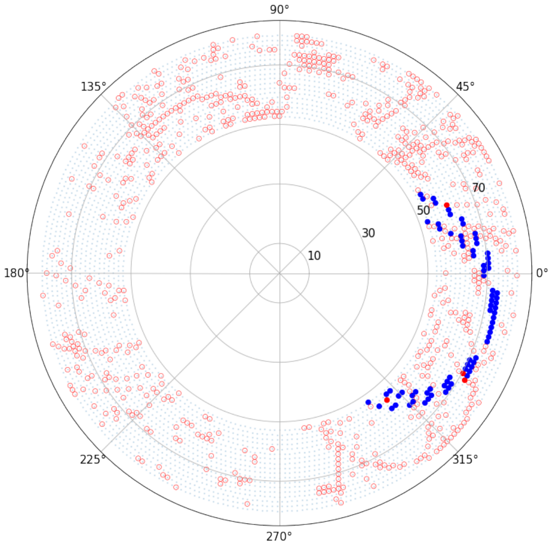
Inverse Reweighted Hough Transform Output

- Signal hits and Background hits scaled by output of track-level GBDT.
- Note: No cuts are placed on scaling of these outputs, this is the full response of the track-level GBDT.



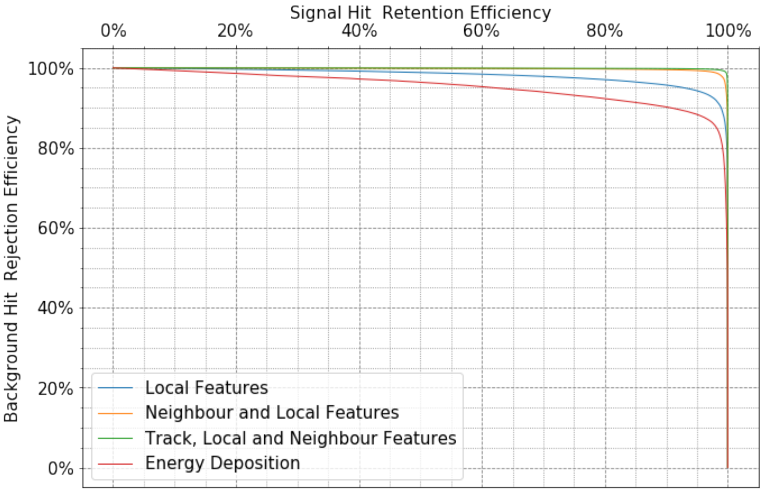
Inverse Reweighted Hough Transform Output

- Cut placed on GBDT output that preserves 99% of signal hits.
- Signal hits and Background hits are filled if they pass the cut.



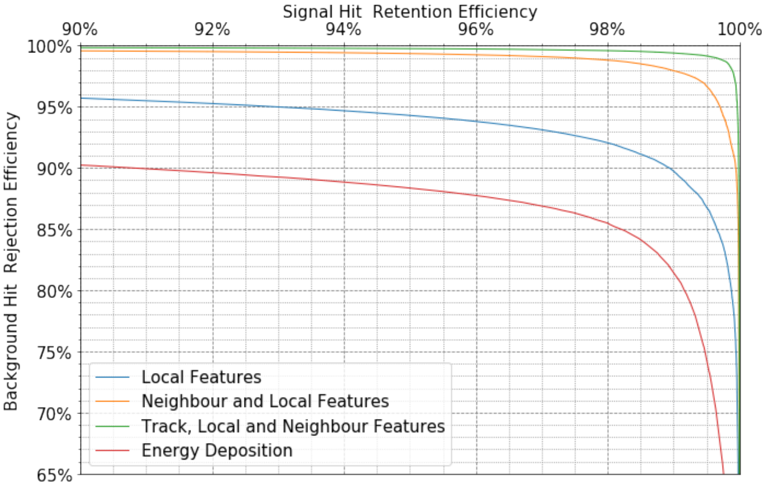
ROC Curves [1]

Comparison of cut-based classifier vs GBDT methods.



ROC Curves [2]

Zoomed ROC curves, note the axes.



Track Finding Summary

The track finding algorithm developed with Dr. Alex Rogozhnikov at Yandex is successful.

- This is the first time BDTs have been used in track finding (so far as I know).
- Further development still needed to define tracks as collections of filtered hit points.

Further work: Track Trigger

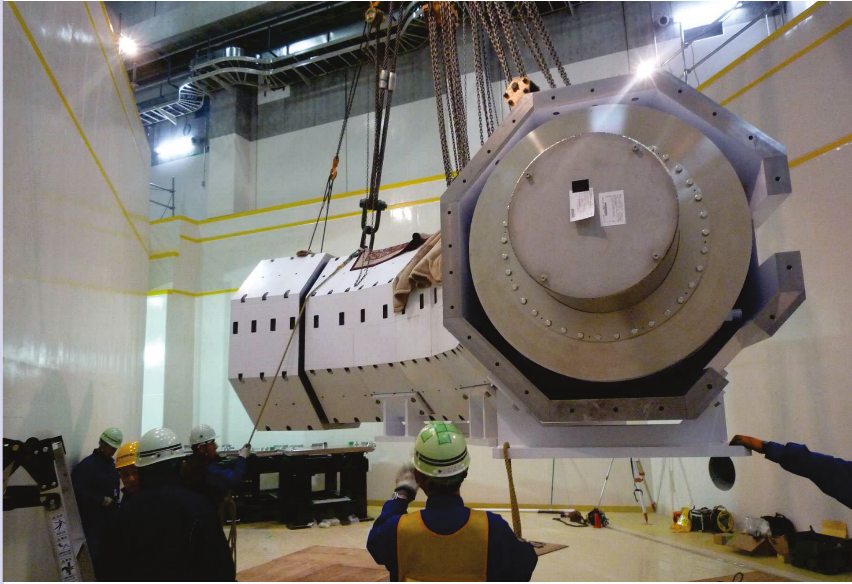
- Algorithm has been developed with Yandex.
- FPGA firmware developed on similar principles in Japan.
- Implementation is underway

https://github.com/ewengillies/track-finding-yandex

Thanks!

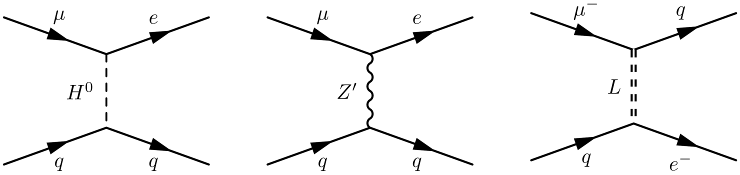
Yandex Data Factory

Imperial College London

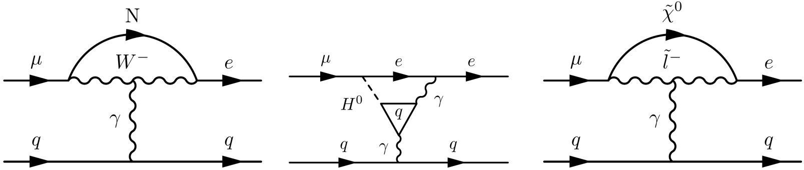


Backup

Possible Channels for Signal



(a) Exotic Higgs **(b)** Z-prime **(c)** Leptoquarks

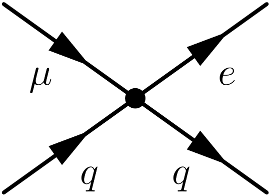


(d) Heavy Neutrinos **(e)** Exotic Higgs **(f)** Supersymmetry

Possible Channels for New Physics

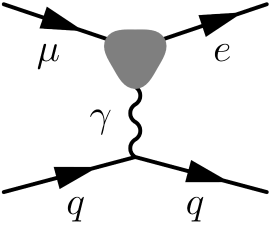
Four-Fermi contact:

- Increased sensitivity for μ-e conversion
- Model-independent search

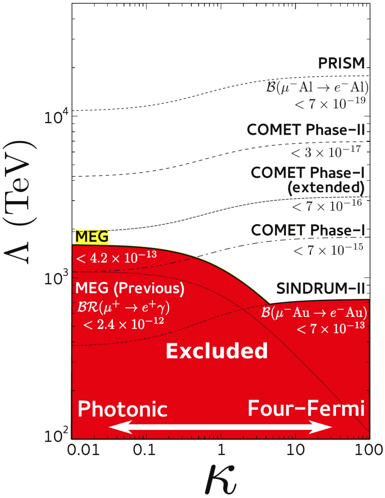


Photonic:

- Still accessible in μ-e conversion search.
- Less sensitive than dedicated μ-e gamma experiments (like MEG).



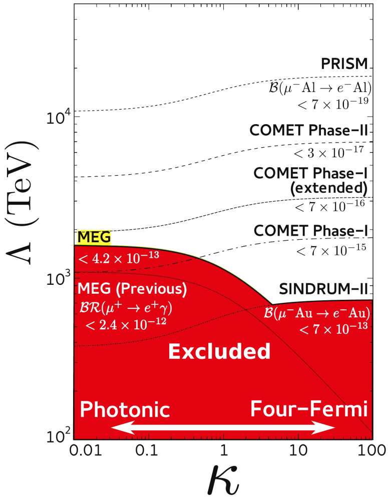
Complementary Searches



- Relative sensitivity to Four Fermi and Photonic interactions is model dependent.
- Highly complimentary to MEG search

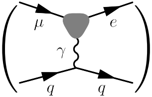
$$\mathcal{L} = \frac{1}{1+\kappa}\frac{m_\mu}{\Lambda^2}\left(\bar{\mu}_R \sigma^{\mu\nu} e_L F_{\mu\nu}\right) + \frac{\kappa}{1+\kappa}\frac{1}{\Lambda^2}\left(\bar{\mu}_L \gamma^\mu e_L\right)\left(\bar{q}_L \gamma_\mu q_L\right)$$

Complementary Searches

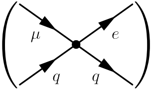


- Relative sensitivity to Four Fermi and Photonic interactions is model dependent.
- Highly complimentary to MEG search

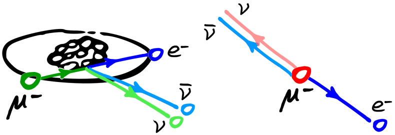
$$\mathcal{L} = \frac{1}{1+\kappa} \frac{m_\mu}{\Lambda^2} \left(\quad \right)$$



$$+ \frac{\kappa}{1+\kappa} \frac{1}{\Lambda^2} \left(\quad \right)$$

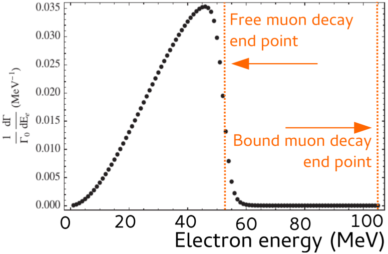


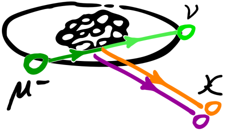
Background Processes



Bound Muon Free Muon

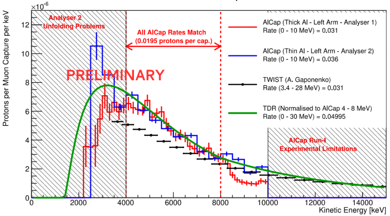
Muon Decay Kinematics





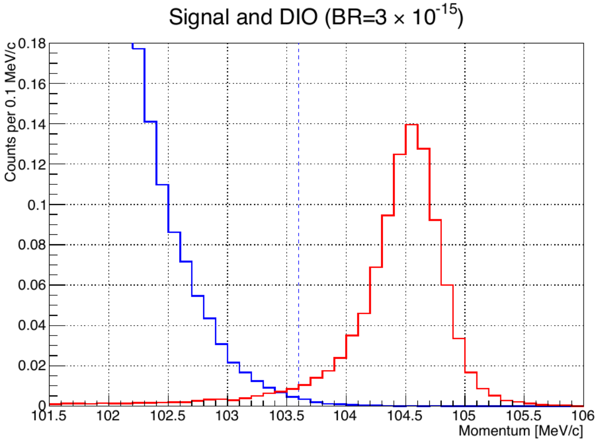
Nuclear Muon Capture

A. Edmonds Proton Emission Spectrum AlCAP Experiment



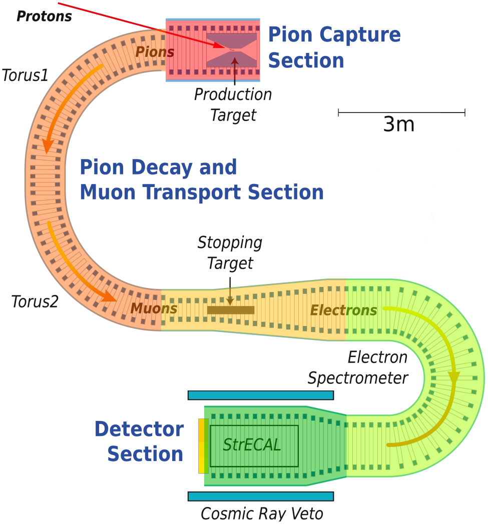
COMET Signal

Both background and signal processes will produce 105 MeV electrons.

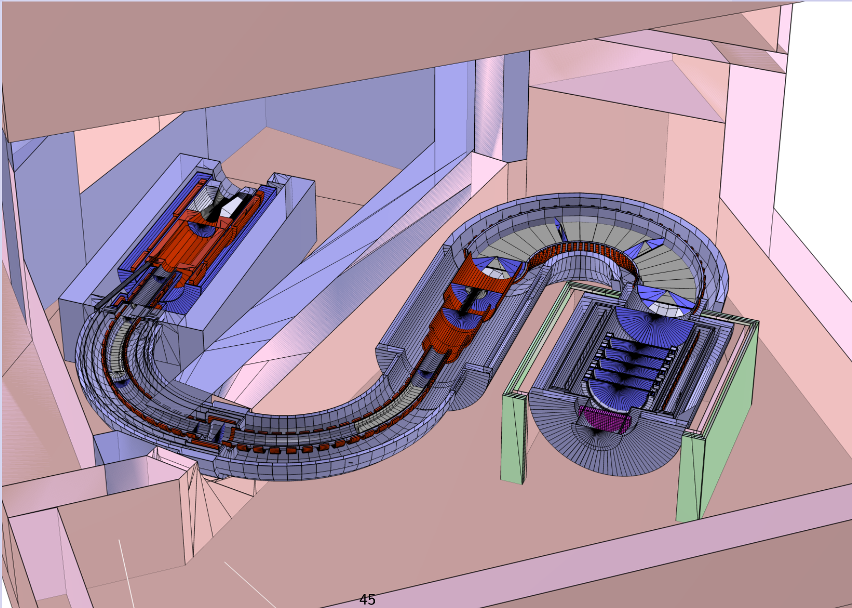


The COMET Experiment: Phase II

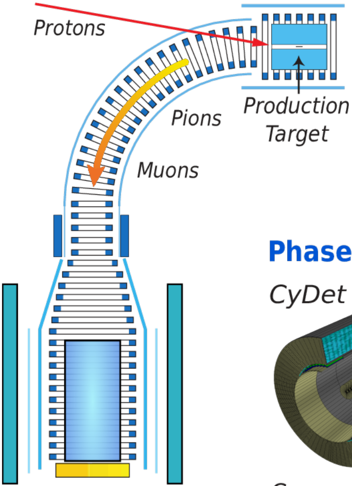
- Capture backwards scattered pions from proton beam.
- Bent solenoids select low momentum muons.
- Muons stopped in target, conversion occurs here!
- Bent solenoids select high momentum electrons.
- Detector waits for offset fiducial time window.



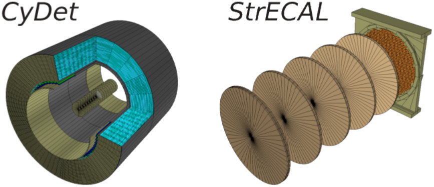
Phase II Geometry



COMET Phase I Design



Phase-I Detectors



Cylindrical Detector: Stereometry

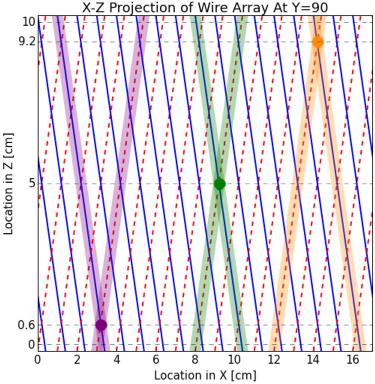


Figure 4: A projection of a wire array with alternating stereo angles from above.

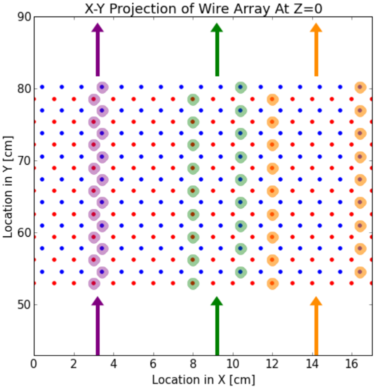


Figure 5: A projection at $Z = 0$ of a wire array with alternating stereo angles from along the beamline.

Decision Tree

Yandex Data Factory
Imperial College London

Sample is split by series of threshold cuts. At each stage, cut is taken that improves the "purity" of classification at next node.

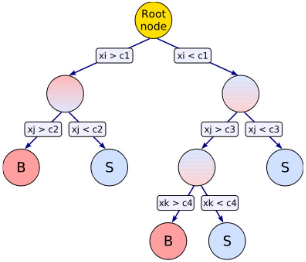


Figure 6: A decision tree, where the features are labelled as $\{xi, xj, xk\}$. The first cut is on xi at value $xi = c1$. This process is continued until some stopping criteria is reached. The leaf nodes are labelled as background, B, or signal, S.

Gradient Boosted Decision Tree

Gradient boosting takes a weighted sum of decision trees. The weights are determined to minimize a loss function that describes misclassification rate. For a hit with a vector of features $\mathbf{f}$:

Decision Tree i: $h_i(\mathbf{f}) = +1 \text{ or } -1$

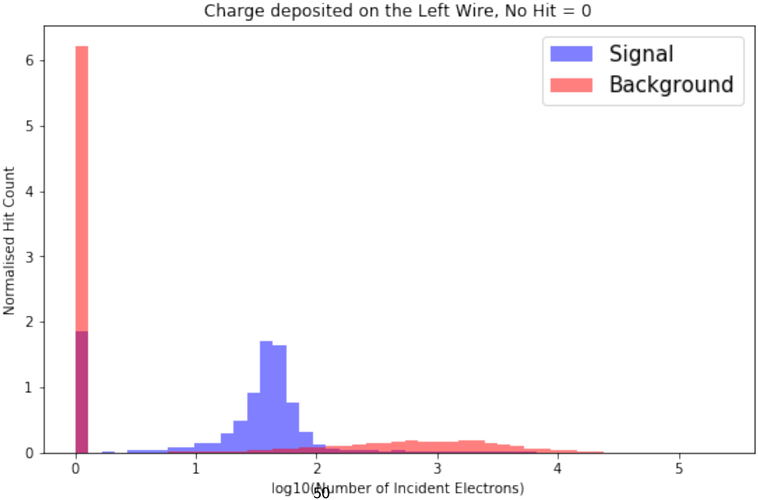
GBDT: $y_{\text{Grad}}(\mathbf{f}, \mathbf{b}) = \sum_{i=0}^{N_{\text{trees}}} b_i h_i(\mathbf{f})$

Loss Function: $F(y_{\text{Grad}}, y) = -2\left[y \cdot y_{\text{Grad}} + \ln\left(1 + e^{y_{\text{Grad}}}\right)\right]$

Minimising this function with respect to the weights $\mathbf{b}$ fully determines the GBDT.

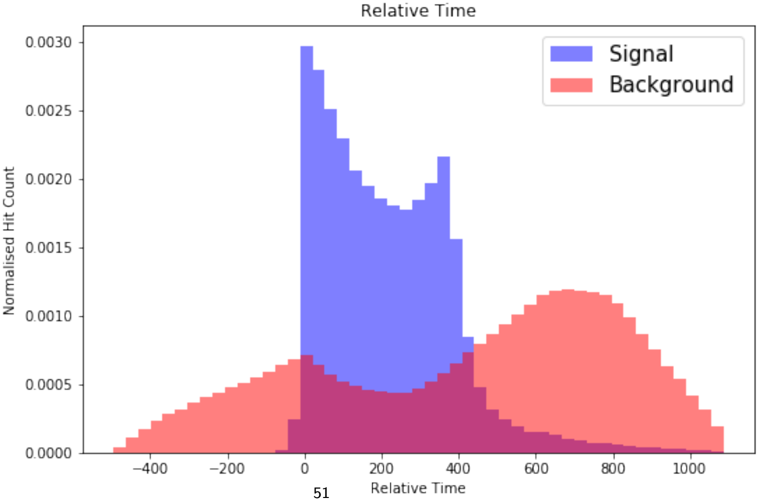
Left Neighbour Deposit

The overflow bin near 0 shows that background hits are far more likely to have no neighbouring hit.



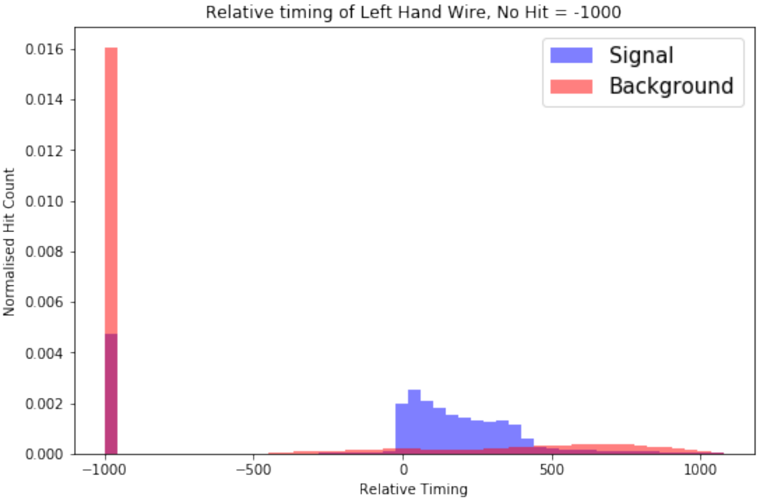
Relative Timing Distribution

This is the timing of the hit relative to the timing of the trigger signal.

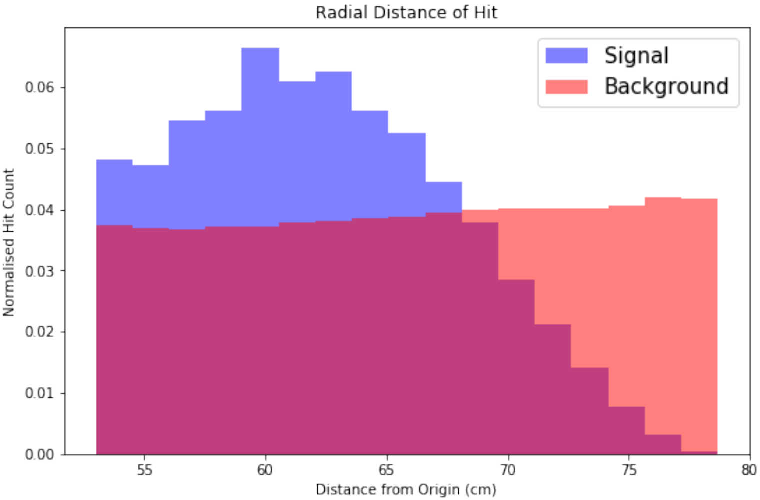


Left Neighbour Relative Timing Distribution

Wires with no hits get a very negative time.



Radial Distance Distribution

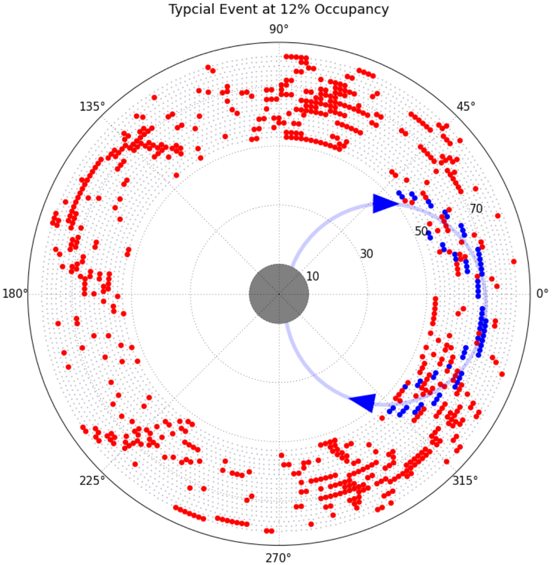


Shape Feature

All signal hits should be part of a track that forms a helix in 3D space.

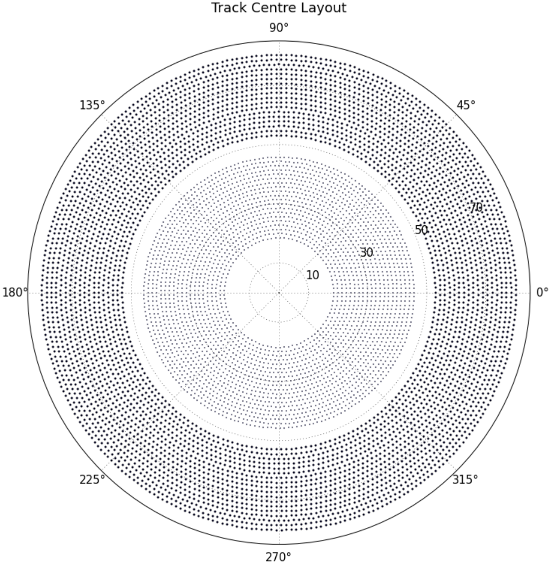
Projecting the track onto a slice of the cylindrical detector gives a **circular shape**.

Stereo angles of the wire array causes displacement of circle between even and odd layers.



Track Centre Layout

- Dark outer dots are wires, i.e. points in (x, y).
- Lighter central dots track centers, i.e. points in (a, b).
- Location of track centers is dictate by geometry, spacial resolution taken to match wire spacing.



Defining the Hough Transform

Define likelihood that a track centred at position $\mathbf{r}_i$ contains a hit wire j at position $\mathbf{r}_j$ as T_{ij}.

- **T** is the Hough Transform matrix of shape $[N_{\text{tracks}}, N_{\text{wires}}]$
- **W** is the hit wire vector of length $[N_{\text{wires}}]$, i.e. $W_j = 1$ for a hit and $W_j = 0$ for no hit.
- **C** is the track center vector of length $[N_{\text{tracks}}]$, where C_i is the likelihood that a signal track exists at track centre i.

Forward Transform

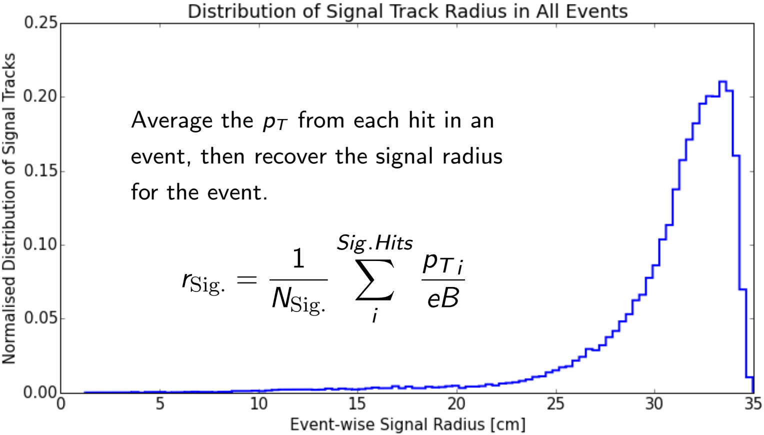
$$\underbrace{T_{ij}}_{\text{Hough}} \overbrace{W_j}^{\text{Hit property}} = \underbrace{C_i}_{\text{Track centers}}$$

Inverse Transform

$$\underbrace{(T_{ij})^{-1}}_{\text{Inv. Hough}} \overbrace{C_i}^{\text{Track property}} = \underbrace{W_j}_{\text{Wire Hits}}$$

Optimizing the Hough Transform [1]

How do we define T_{ij}? Recover the distribution of the radii of signal tracks directly from simulation. Each track has an associated particle, with transverse momentum p_T.



Average the p_T from each hit in an event, then recover the signal radius for the event.

$$r_{\text{Sig.}} = \frac{1}{N_{\text{Sig.}}} \sum_{i}^{Sig.Hits} \frac{p_{T\,i}}{eB}$$

Optimizing the Hough Transform [2]

Fit this distribution directly to recover values for T_{ij}. For distance $d_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ between track centre j and wire i:

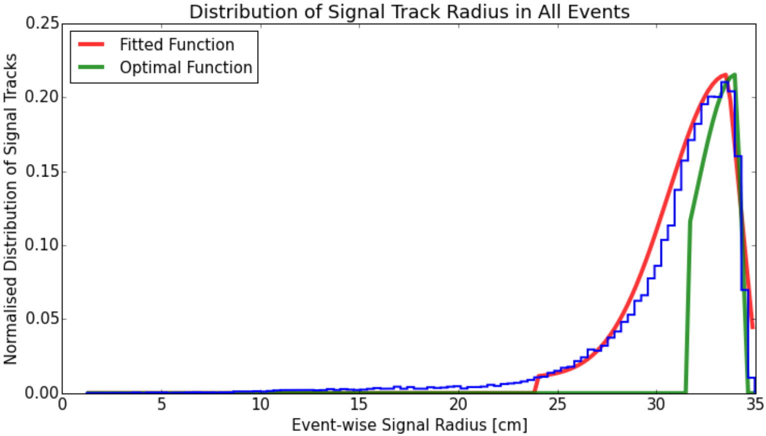
$$T(d_{ij}) = T_{ij} \propto \begin{cases} \exp\left(\frac{\left[d_{ij} - r_{\rm sig}\right]^2}{2\sigma_{\rm sig}^2}\right) & : \ r_{\rm min} < d_{ij} < r_{\rm sig} \\ 1 - \frac{d_{ij} - r_{\rm sig}}{r_{\rm max} - r_{\rm sig} + 0.1} & : \ r_{\rm sig} < d_{ij} < r_{\rm max} \\ 0 & : \ \text{else} \end{cases}$$

This is half a Gaussian centred around the signal radius for smaller radii and a linear drop off for larger radii.

The parameters are the signal radius, $r_{\rm sig}$, the spread for lower values, $\sigma_{\rm sig}$, and the minimal and maximal radii considered, $r_{\rm min}$ and $r_{\rm max}$.

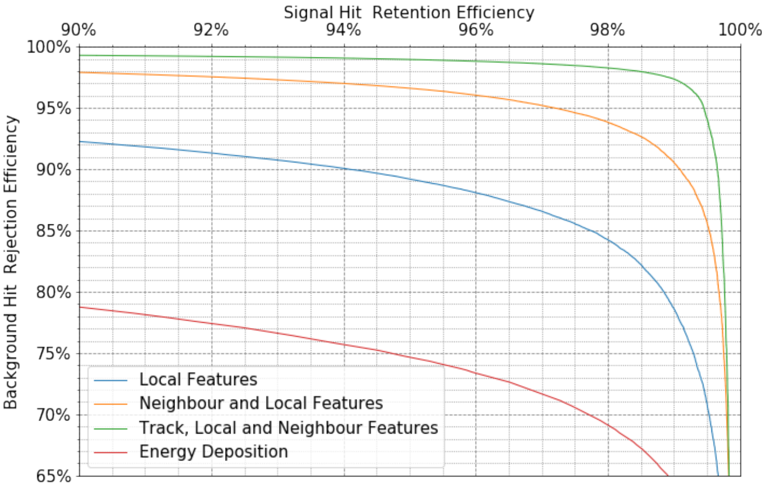
Optimizing the Hough Transform [3]

Curve	r_{min}	r_{sig}	r_{max}	σ_{sig}
Fitted Function	24 cm	33.6 cm	35 cm	3 cm
Optimal Function	31.5 cm	34 cm	34.5 cm	2 cm



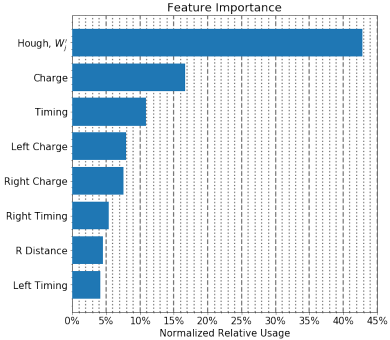
ROC Curve: Previous Result

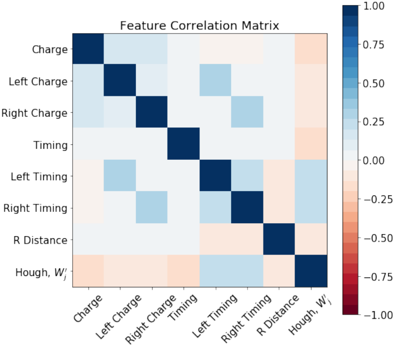
Zoomed ROC curves for previous sample



Feature Evaluation

The feature importance evaluates how often a feature was used to split a node.





COTTRI Board