Pattern Recognition Techniques for Finding Very Rare Events in the COMET Experiment

Ewen Lawson Gillies

Imperial College London
High Energy Particle Physics

May 29th, 2015

Overview

Developing an algorithm to distinguish between signal and background particles using a series of gradient boosted decision trees.

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

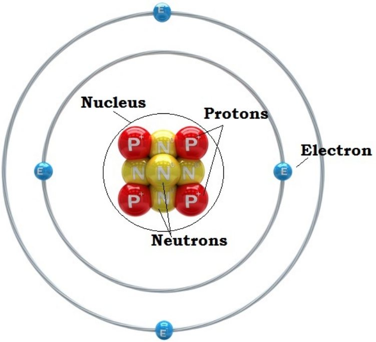
At 99% signal retention, this method removes 99.5% of background hits.

Physics

The Physics of Very Small Things [1]

Modern nuclear physics was born in the early 1900's. At this time, the smallest things looked like this:

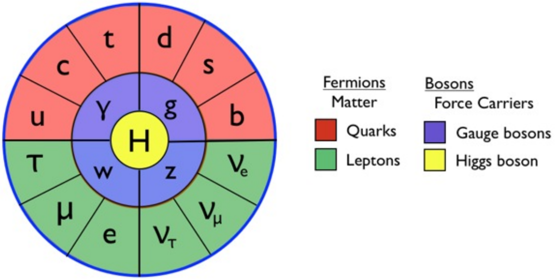
	Charge	Mass
Atom	0	$< 10^{-25}$ kg
Proton	+1	10^{-27}kg
Neutron	0	10^{-27}kg
Electron	-1	10^{-30}kg



This was a complete list of very small things, until the 1930's...

The Physics of Very Small Things [2]

The muon was discovered in 1936. This discovery destroyed the simple model of nuclear physics, but was the first step to the Standard Model.



Particles of the Standard Model

Lepton Physics

Lepton Physics [1]

Lets focus on leptons. Essentially, these objects are closely related to electrons. These are the charged leptons:

	Charge	Mass	$L(\tau)$	$L(\mu)$	$L(e)$
Tauons, τ^-	-1	10^{-27} kg	1	0	0
Muons, μ^-	-1	10^{-28} kg	0	1	0
Electron, e^-	-1	10^{-30} kg	0	0	1

These are the neutral leptons, called *neutrinos*. Note the lepton numbers for each neutrino, labelled $L(\tau)$, $L(\mu)$, and $L(e)$. These particle were only discovered to have mass in the 1990's.

	Charge	Mass	$L(\tau)$	$L(\mu)$	$L(e)$
τ-neutrino, ν_τ	0	10^{-37} kg	1	0	0
μ-neutrino, ν_μ	0	10^{-37} kg	0	1	0
e-neutrino, ν_e	0	10^{-37} kg	0	0	1

Lepton Physics [2]

All leptons have anti-particle partners. These are the same mass of their partners, but opposite in charge and lepton number. These are the charged anti-leptons:

	Charge	Mass	$L(\tau)$	$L(\mu)$	$L(e)$
Anti-Tauons, τ^+	$+1$	10^{-27} kg	-1	0	0
Anti-Muons, μ^+	$+1$	10^{-28} kg	0	-1	0
Anti-Electron, e^+	$+1$	10^{-30} kg	0	0	-1

And now for the neutral anti-leptons, the *anti-neutrinos*:

	Charge	Mass	$L(\tau)$	$L(\mu)$	$L(e)$
Anti τ-neutrino, $\bar{\nu}_\tau$	0	10^{-37} kg	-1	0	0
Anti μ-neutrino, $\bar{\nu}_\mu$	0	10^{-37} kg	0	-1	0
Anti e-neutrino, $\bar{\nu}_e$	0	10^{-37} kg	0	0	-1

Lepton Flavour Conservation [1]

Simply stated, its that all lepton numbers are conserved in an interaction. For example, for muons, this means that $L(\mu)$ in the first part of the interaction is the same as $L(\mu)$ at the end.

Example: Muon Decay

$$\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e$$

Example: Muon Capture in a Nucleus, N

$$\mu^- + N \rightarrow \nu_\mu + N'$$

Lepton Flavour Violation

Massive neutrinos break this conservation law. This is called Neutral Lepton Flavour Violation (NLFV). This violation is very small and hard to detect. Even so, this violation breaks the Standard Model. The question is:

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

This is called Charged Lepton Flavour Violation. Such a discovery would be a huge breakthrough, as big as any from the LHC. The three main places this is tested for are:

"μ to three e" : $\mu^+ \rightarrow e^+ + e^+ + e^-$
"μ to e–γ" : $\mu^+ \rightarrow e^+ + \gamma$
Muon to Electron Conversion : $\mu^- + N \rightarrow e^- + N$

Current Experimental Limits

Charged Lepton Flavour Conservation has been tested for decades. No experiments have found any sign of CLFV. They place the following *upper limits* on the process.

- $Br(\mu^+ \to e^+ + e^+ + e^-) < 1.0 \times 10^{-12}$ (SINDRUM 1988)
- $Br(\mu^+ \to e^+ + \gamma) < 5 \times 10^{-13}$ (MEG 2013)
- $B(\mu^- + Au \to e^- + Au) < 7 \times 10^{-13}$ (SINDRUM II 2006)

COMET focuses on muon to electron conversion. Without CLFV, this process can only come indirectly from NLFV. It is unimaginably rare:

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

By 2017, COMET Phase I aims to achieve the sensitivity of :

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

Lepton Physics Summary

Lepton Flavour

- Lepton Flavour is grouped into three categories, $L(\tau), L(\mu), L(e)$
- Each category has a neutral and charged particle
- Each particle has an anti-particle

Lepton Flavour Conservation

- Amount of $L(\tau), L(\mu), L(e)$ beginning $=$ Amount of $L(\tau), L(\mu), L(e)$ at the end
- Neutral leptons violate this in a very small way
- Charged leptons not observed to violate this yet.

Discovering charged lepton flavour violation would be a huge discovery, atleast as important as any recent LHC discovery. The LHC is not optimized to search for CLFV, so these results are complimentary.

COMET

COMET Basics

The goal of COMET is to create a lot of muons, have them interact with aluminium to make muonic atoms, and see if any electrons fly out.
In the Standard Model, electrons can come from muon decay in orbit.

$$\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e$$

This has a peak electron energy of 52.8 MeV.

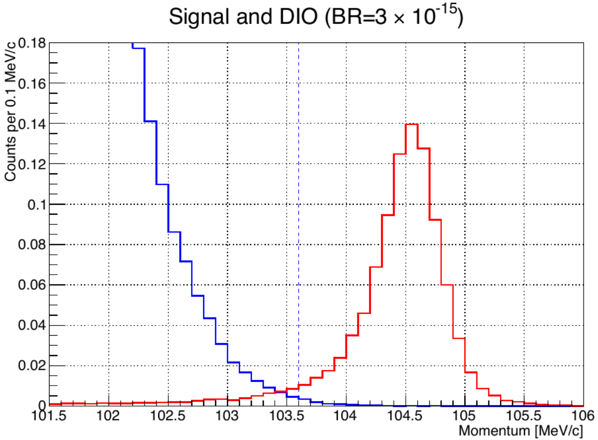
With CLFV, this can happen through muon to electron conversion.

$$\mu^- + \text{Al} \rightarrow e^- + \text{Al}$$

This create an electron of energy 105 MeV, far away from the background peak.

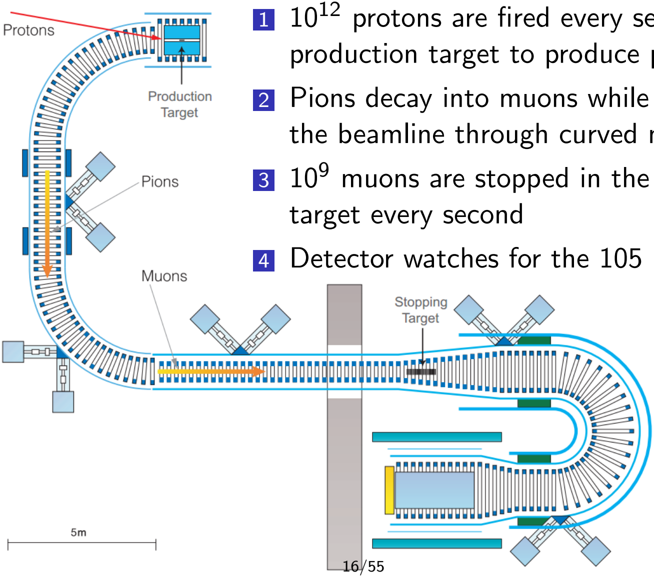
COMET Signal

Both background and signal processes will produce 105 MeV electrons. **We need to find more than background alone can produce.**



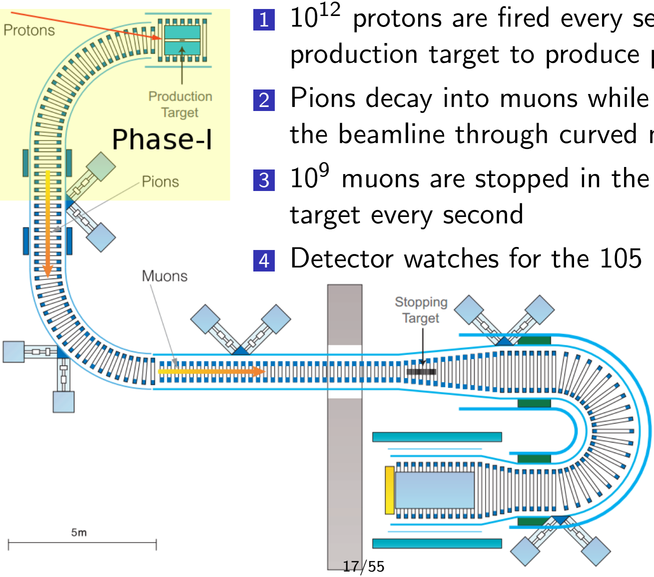
COMET Design [1]

1. 10^{12} protons are fired every second at the production target to produce pions
2. Pions decay into muons while flying down the beamline through curved magnets
3. 10^{9} muons are stopped in the aluminium target every second
4. Detector watches for the 105 MeV electrons

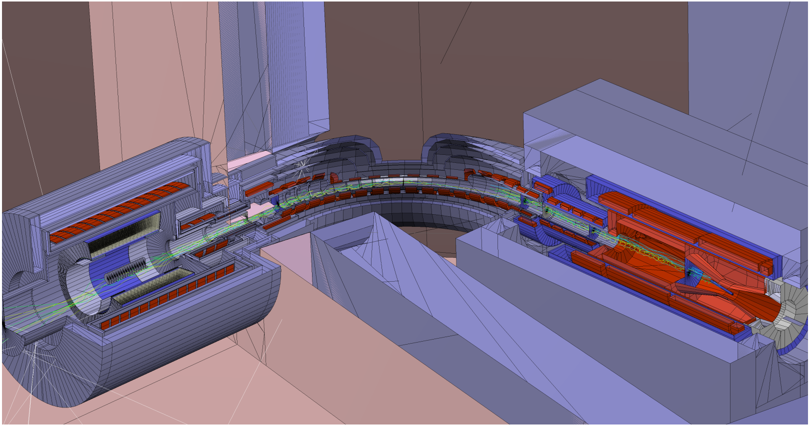


COMET Design [1]

1. 10^{12} protons are fired every second at the production target to produce pions
2. Pions decay into muons while flying down the beamline through curved magnets
3. 10^{9} muons are stopped in the aluminium target every second
4. Detector watches for the 105 MeV electrons



COMET Design [2]



Cylindrical Detector [1]

The detector measures the radius of curvature of a charged particle in a magnetic field.

- Larger transverse momentum = larger radius of curvature.
- Inner radius of detector is large, blinding it to low energy particles.
- Uses $\sim 4,400$ wires to reconstruct path, hence radius of curvature.

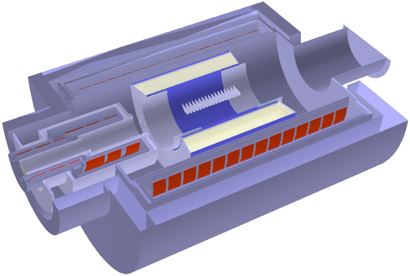
$$r = \frac{p_T}{eB}$$

r = Radius of Curvature

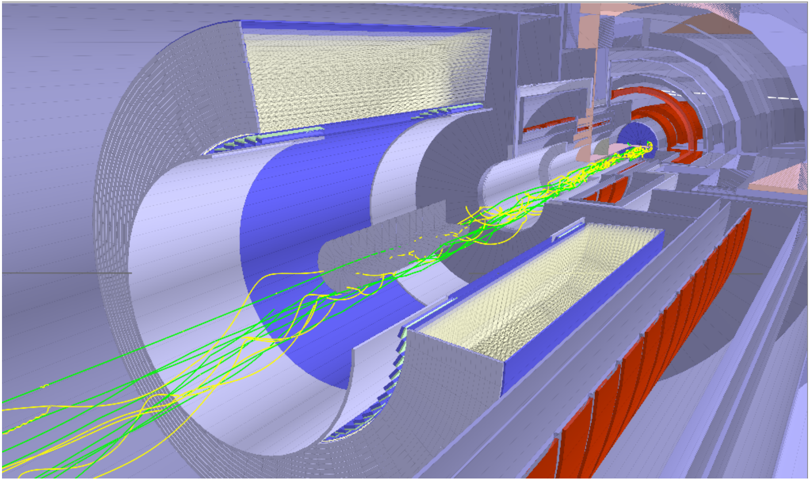
p_T = Transverse Momentum

e = Charge of Electron

B = Magnetic Field Strength

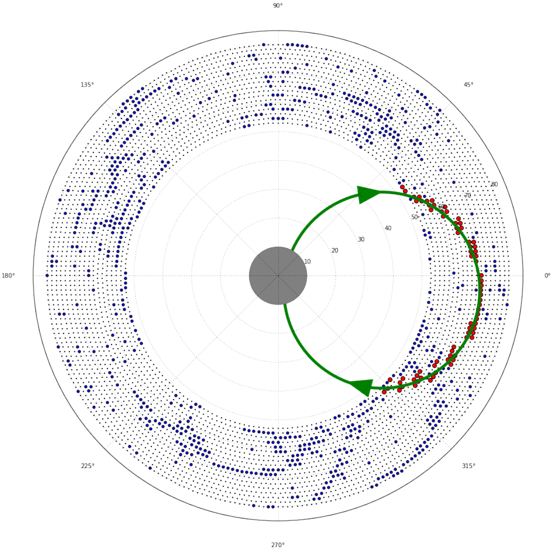


Cylindrical Detector [2]



Typical Event

- **Signal Hits** from 105 MeV electron ejected from aluminium target. Average is 80 per signal electron.
- **Background Hits** from other particles in the detector. Average is 360 hits per event in current simulations.



COMET Summary

COMET is designed to look for CLFV by:

- Producing a lot of muons
- Have them interact with aluminium
- Check if any become electrons

Muons that become electrons would have a very distinct energy. To find these electrons:

- Find a track whose path corresponds to the signal energy.
- This path is reconstructed from "hits" which occur when the electron gets close to a wire.
- We must see more electrons at the signal energy than could come from background to claim a discovery.

COMET Phase-I aims to improve the current upper limit on how often CLFV may occur by a measurement that is 100 times more sensitive.

Track Finding

Classification Problem

"Is this wire a signal hit or a background hit?" This is *not* track fitting. This is finding the points that correspond to a signal track.

Hit wires have three main features:

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

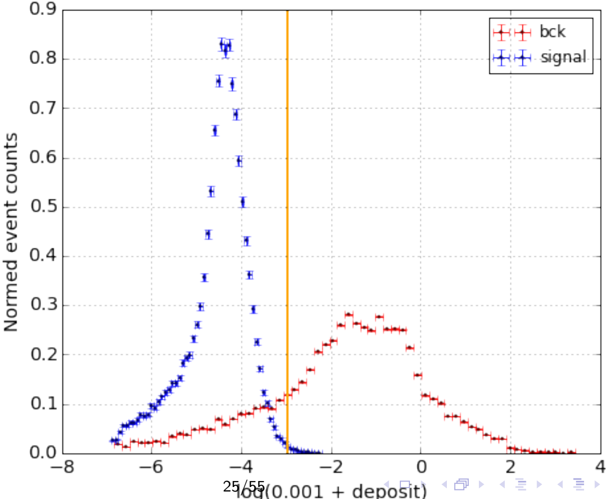
Construct a classification algorithm in layers:

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

Combine the results into a classifier, remove background hits, and define signal tracks. Test and tune this against simulated data.

Previous Classifier

Previous method used a cut on energy deposition, removing 80% of background while keeping 99.7% of signal.



Pattern Recognition Techniques for Finding Very Rare Events in the COMET Experiment

Ewen Lawson Gillies

GBDT

Gradient Boosted Decision Tree

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

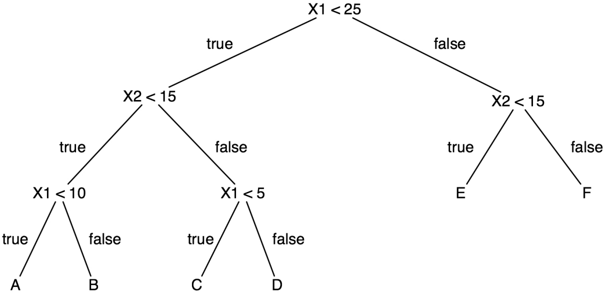


Figure: Generic tree features X1 and X2, classes A, B, C, D, E and F.

Gradient boosting takes a weighted sum of decision trees. The weights are determined to minimize a loss function that describes misclassification rate.

Wire Level GBDT

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

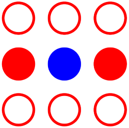
Before looking at those, we can use the wire level features to assign a probability that this wire is a signal

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

During the local level GBDT where neighbours are considered, we can use this wire-level GBDT value to check how signal-like this wire's neighbours are.

Local Level GBDT

Exploit both wire and neighbour features to form local features. The neighbours' features are summed. These sums are taken from two groups of neighbours for any given wire:



`neigh` : All Red Circles

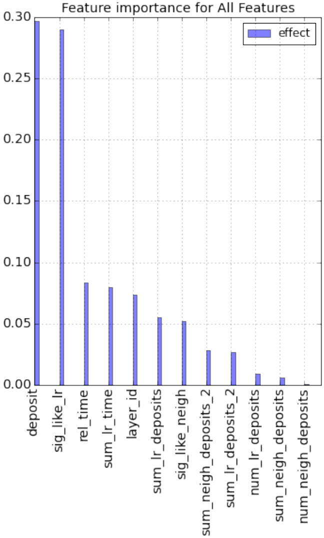
`lr` : Filled Red Circles (left/right)

Examples :

`sum_lr_time` : Sum timing of hits from left/right neighbours

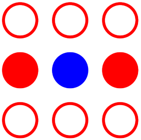
`sig_like_neigh` : Sum of wire GBDT output for all neighbours

Local Level GBDT



`neigh` : All Red Circles

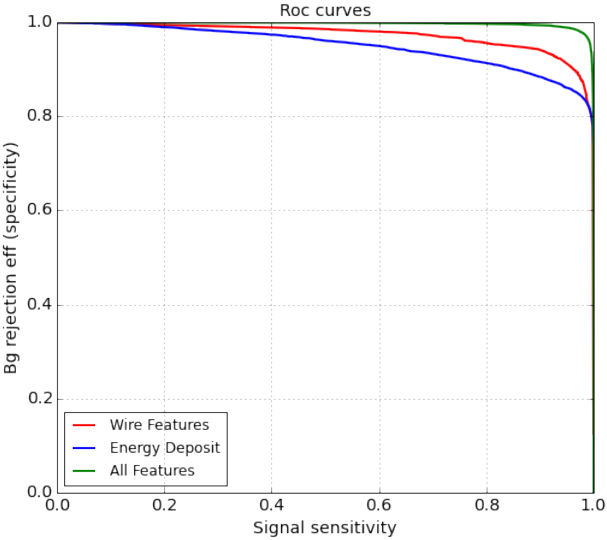
`lr` : Filled Red Circles (left/right)



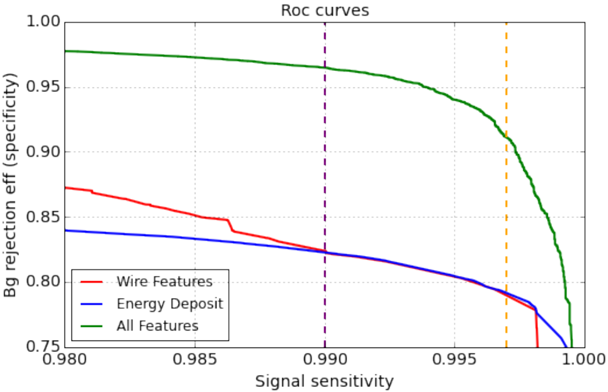
Classes of Features :

- Wire Features
- Sums of neighbouring wire features
- Sums of Wire GBDT output for neighbours

Local ROC Curve [1]



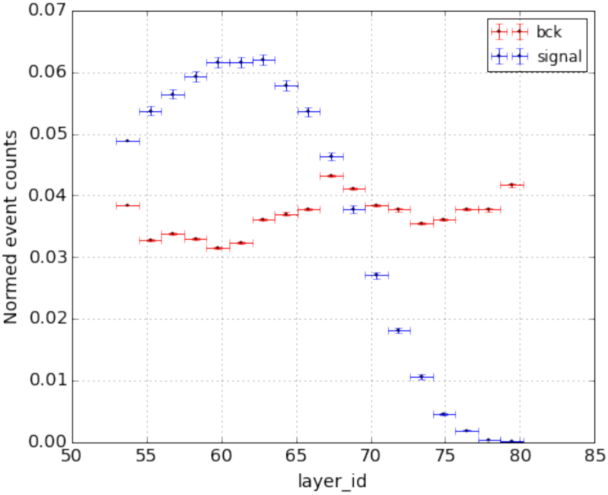
Local ROC Curve [2]



	Sig. Sen.	Bkg Rejection
5 KeV equivalent	99.7%	80%, 93%
Stable Benchmark	99%	83%, 97%

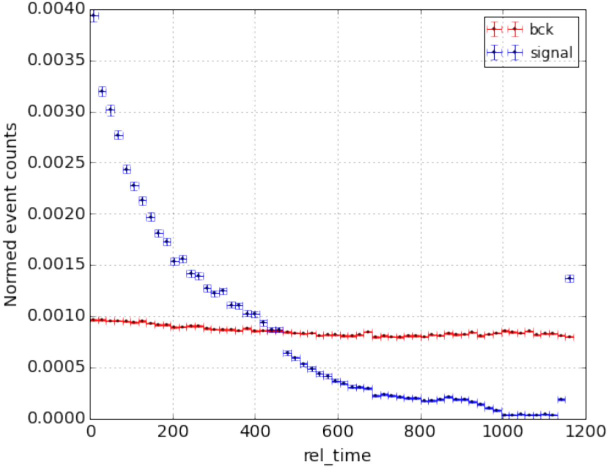
Feature : Radial Distance

May introduce some selection bias in signal, not yet considered.



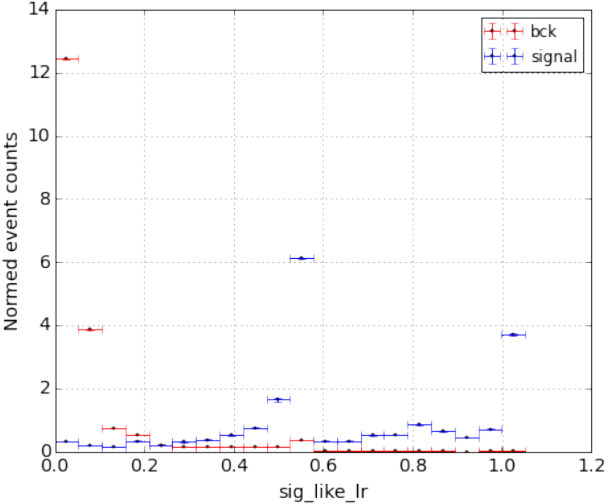
Feature : Relative Time

Timing of hit considered relative to “trigger” timing.



Feature : Signal Like LR Neighbours

Strong feature, but not new information.

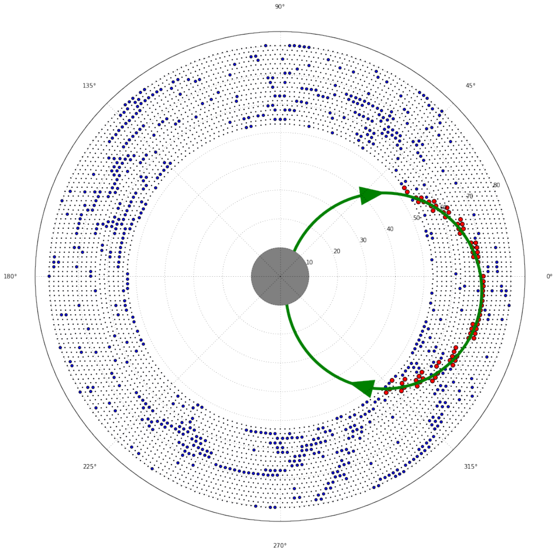


Hough Transform

Shape Feature

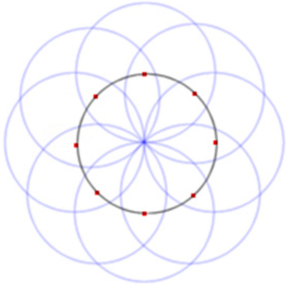
All signal hits should be part of a track that forms a helix in 3D space.

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



Circular Hough Transform

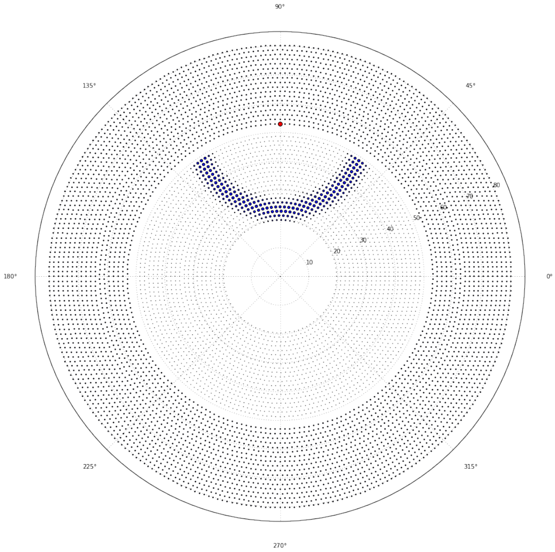
- XY-"Space" : Red points, (x, y), on desired circle
- AB-"Space" : Blue Circles, (a, b), possible centres of each red point



Interception of blue circles gives center common to all points in XY "space." **Assume radius is known beforehand**.

Cylindrical Detector Layout

- CyDet from end plate
- Dark outer dots are wires, i.e. points in XY
- Lighter central dots centres of circles, i.e. points in AB
- Red dot is hit, blue dots potential track center sized by probability.



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 [number of track centres, number wires].
- **W** is the wire vector of length [number of wires], where W_j is the output of the local GBDT.
- **C** is the track center vector of length [number of tracks centres], where $T_{ij}W_j = C_i$, which is the likelihood that there is a track centred at position $\mathbf{r}_i$.

Forward Transform

$$\underbrace{T_{ij}}_{\text{Hough}} \overbrace{W_j}^{\text{Local properties}} = \underbrace{C_i}_{\text{Track Centers}}$$

Inverse Transform

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

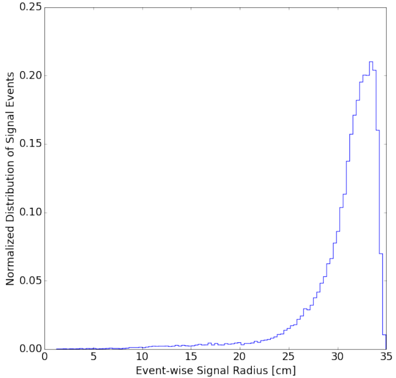
Optimizing the Hough Transform [1]

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.

$$r = \frac{p_t}{eB}$$

Take magnetic field $B = 1$ T for detector region.

e is the charge on an electron.



Optimizing the Hough Transform [2]

Fit this distribution directly to recover values for T_{ij}.

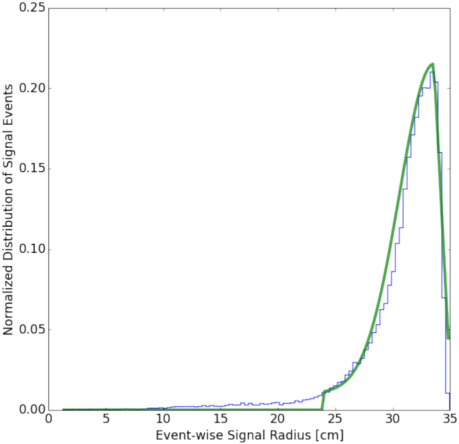
If $r_{\mathrm{min}} < r < r_{\mathrm{sig}}$:

$$T_{ij} \propto \exp\left(\frac{\left[|\mathbf{r}_i - \mathbf{r}_j| - r_{\mathrm{sig}}\right]^2}{2\sigma_{\mathrm{sig}}^2}\right)$$

If $r_{\mathrm{sig}} < r < r_{\mathrm{max}}$:

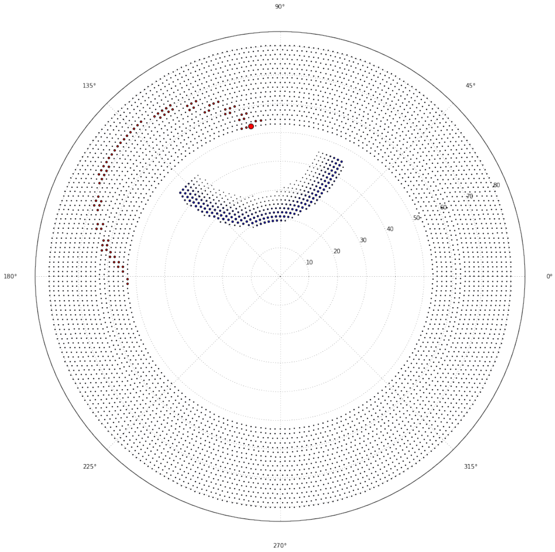
$$T_{ij} \propto 1 - \frac{r - r_{\mathrm{sig}}}{r_{\mathrm{max}} - r_{\mathrm{sig}} + 0.1}$$

$r_{\mathrm{sig}} = 33.6$ cm
$r_{\mathrm{max}} = 35$ cm
$r_{\mathrm{min}} = 24$ cm
$\sigma_{\mathrm{sig}} = 3$ cm



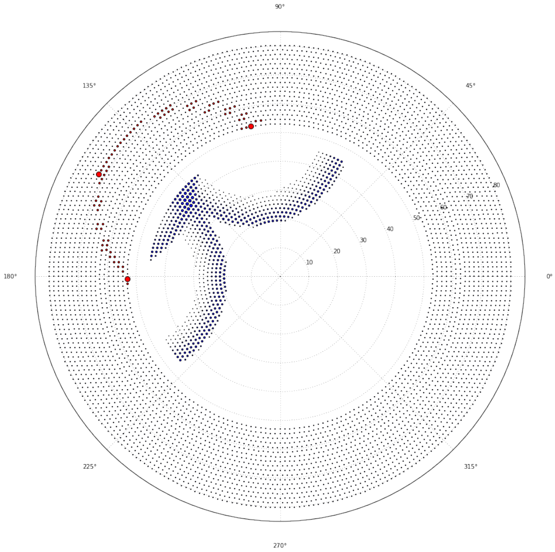
Demo of the Hough Transform [1]

Possible centres, from **one point**, on a signal track.



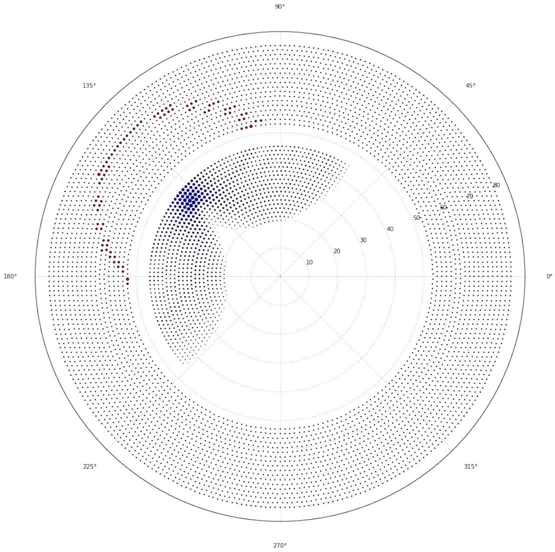
Demo of the Hough Transform [2]

Possible centres, from **three points**, on a signal track.



Demo of the Hough Transform [3]

Possible centres, from **all points**, on a signal track. [Scaling of centres sizes has been adjusted].



Executing the Hough Transform

1. Get Tracks : perform forward hough transform on GBDT output to get $C_i = T_{ij} W_j$.
2. Choose Best Tracks : reweight to highlight "best" track centres using:
$$C_i' = \exp\left(\alpha C_i\right)$$
3. Find Wires : Transform back using $W_j' = \left(T_{ij}\right)^T C_i'$.
4. Combined GBDT : using the local features plus W_j'.

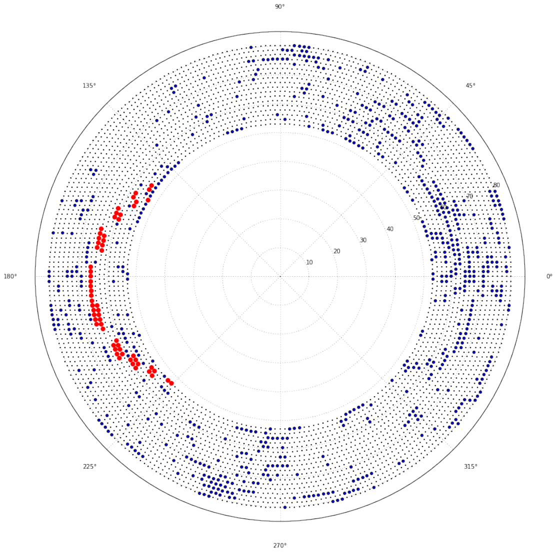
Aim:

- To select signal hit wires along track that were missed by GBDT.
- To also remove clusters of background that locally look like signal, but do not form a circle.

New parameter α has huge effect on output.

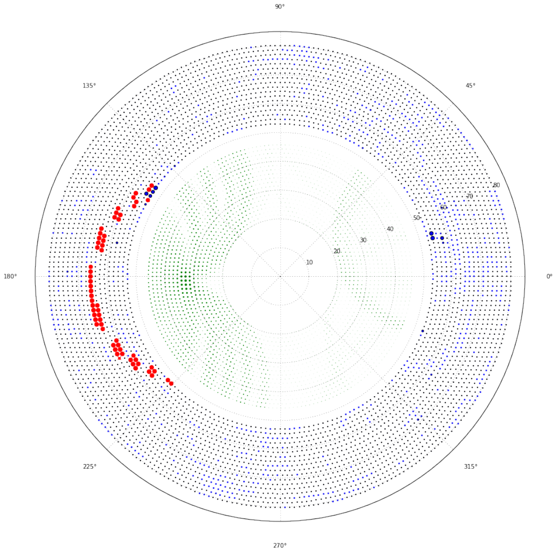
Demo of the Hough Feature [1]

Background Hits,
Signal Hits.



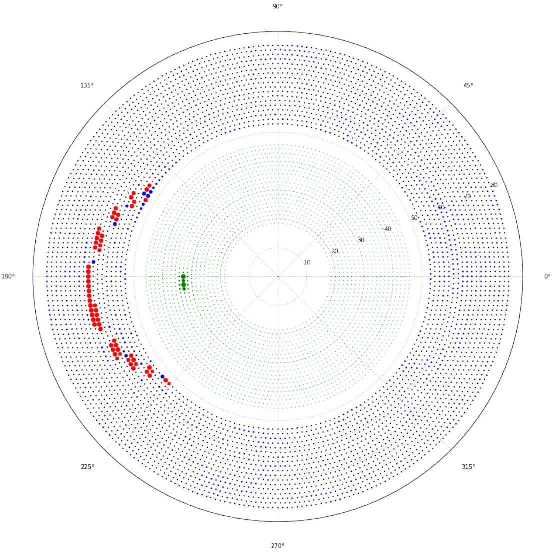
Demo of the Hough Feature [2]

- Signal hits scaled by local GBDT output W_j.
- Background hits scaled by local GBDT output W_j.
- Track centres scaled by C_i from $C_i = T_{ij} W_j$.



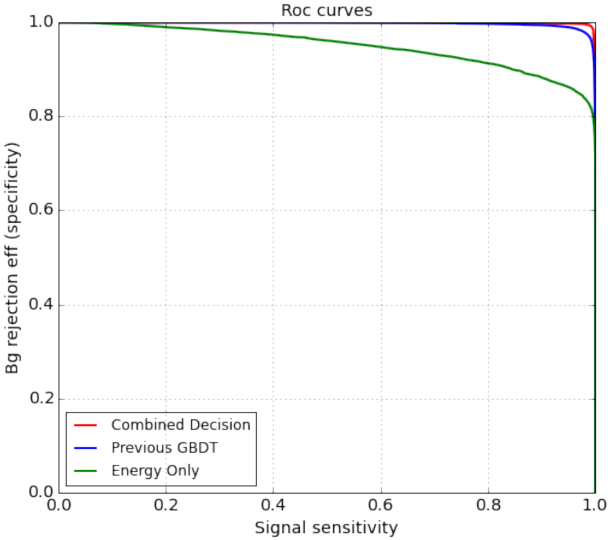
Demo of the Hough Feature [3]

- Track centres reweighted by C_i' from $C_i' = \exp(\alpha C_i)$.
- Signal hits scaled by hough inverse output W_j' from $W_j' = (T_{ij})^T C_i'$.
- Background hits scaled by hough inverse output W_j' from $W_j' = (T_{ij})^T C_i'$.

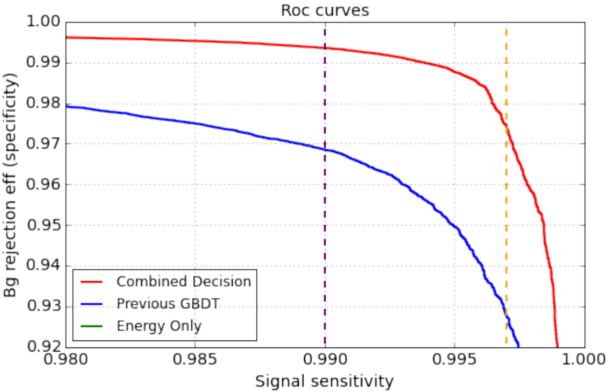


Combined GBDT

Combined ROC Curve [1]

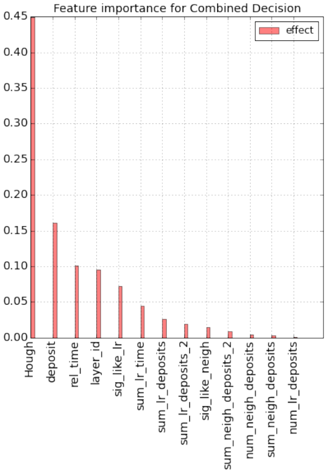


Combined ROC Curve



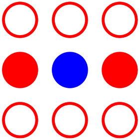
	Sig. Sen.	Bkg Rejection
5 KeV equivalent	99.7%	[80%], 93%, 97.5%
Stable Benchmark	99%	[83%], 97%, 99.5%

Combined Feature Importance



`neigh` : All Red Circles

`lr` : Filled Red Circles (left/right)



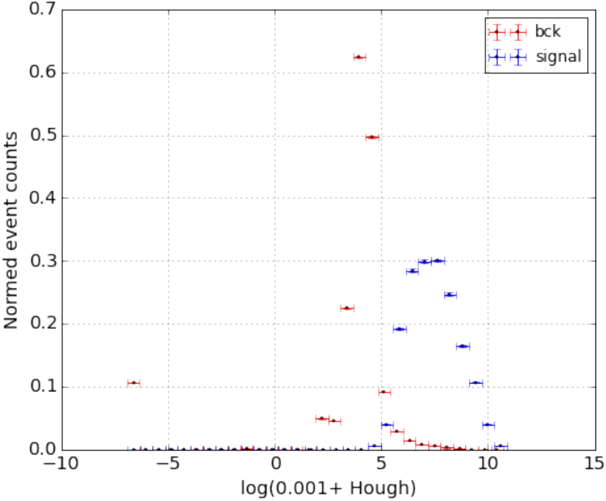
New Feature :

Hough Output W_j' from inverse hough on weighted track center C_i'

Overall : Performance improved.

Feature : Hough

Strong new feature that incorporates shape of track.



Summary

Current Status

- Full analysis chain is working in REP (Reproducible Experiment Platform).
- Local GBDT features can still be improved
- Hough is still sub-optimal, as there is a fairly large parameter space. Can be improved.

Future Development

- Using this method on better simulation data
- Optimizing existing parameters

The real detector environment will be more challenging. Currently, larger simulations are being produced, which will help determine optimal parameters.