ALICE
A JOURNEY OF DISCOVERY

UNIVERSITY OF MALTA
L-Università ta' Malta

TPC looper finder status

J. A. Briffa, G. Valentino
University of Malta

Thanks to M. Ivanov, K. Schweda, R. Shahoyan, J. Wiechula for valuable discussions

ALICE Offline week, March 2017

Outline

- Introduction
- Ground truth definitions
- Looper finder algorithm based on HT
- Results
- Conclusions

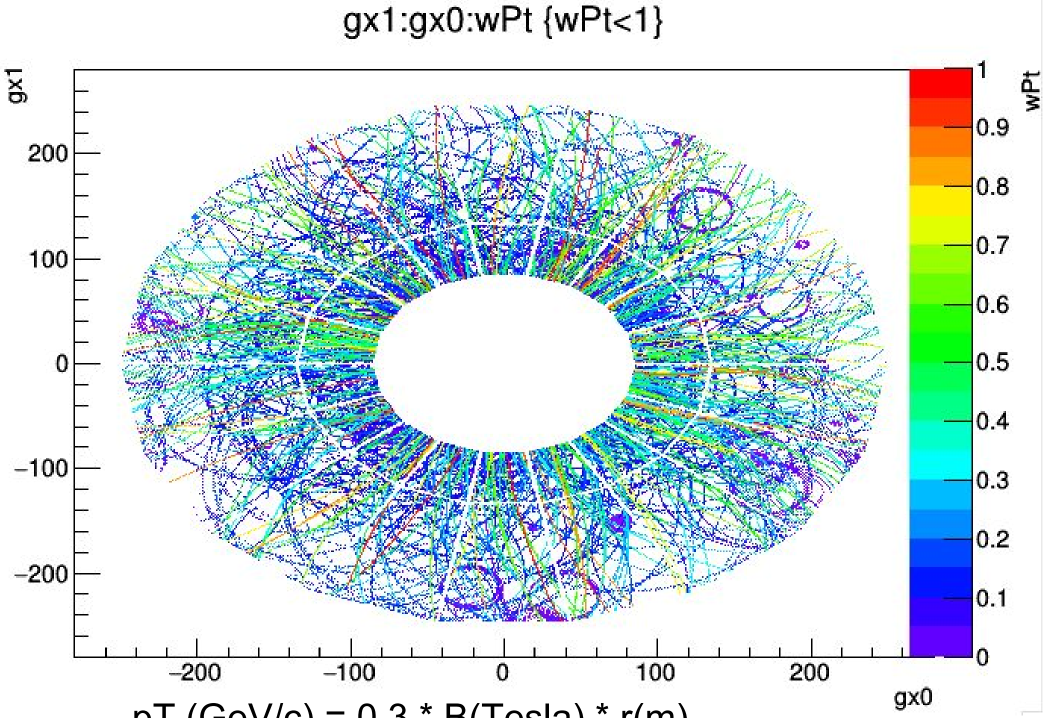
Introduction

- The output from the TPC is currently ~5 PB/year, and due to increase in Run 3.

- Very low momentum tracks < 50 MeV are not used for physics analysis, called loopers.

- Depending on beam conditions, loopers make up to 80% of the raw data in pp collisions

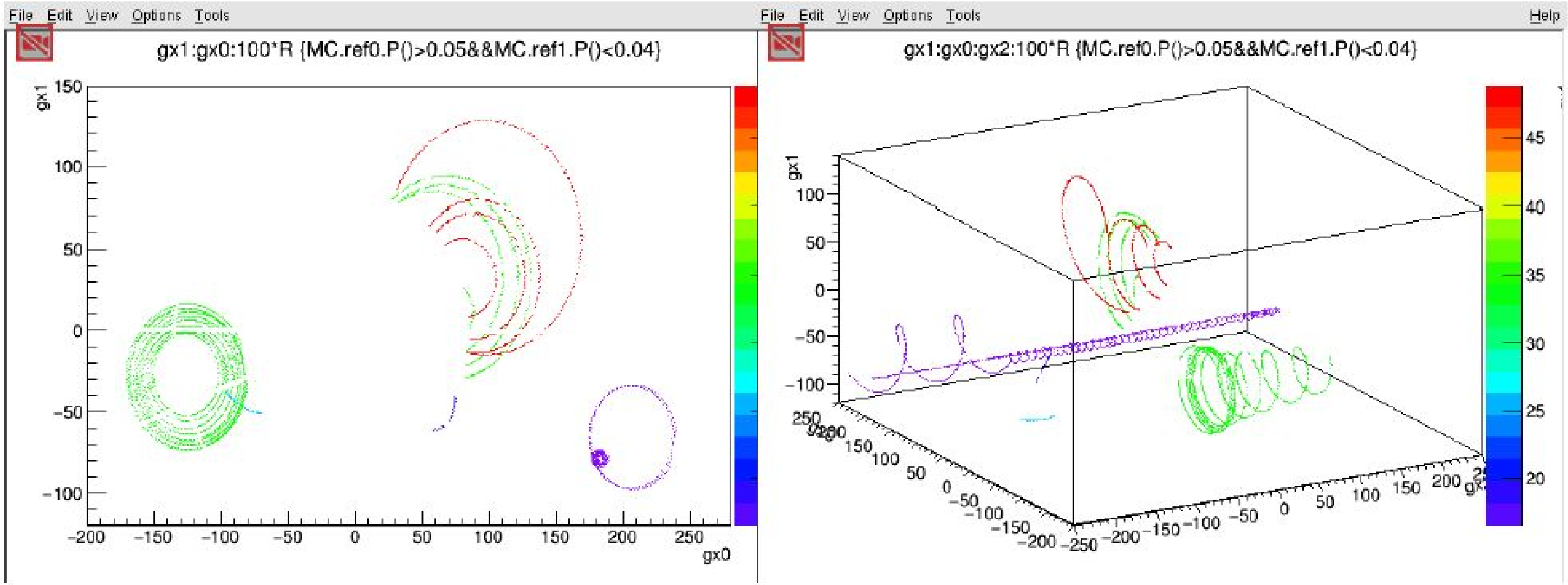
- It is desirable to remove these loopers in order to achieve better data compression.



pT (GeV/c) = 0.3 * B(Tesla) * r(m)

Introduction

- In reality, looper helix parameters may change along the trajectory due to dE/dx, tan φ, multiple scattering, distortion.



M. Ivanov

Ground truth - theoretical

Definitions, assuming MC data has perfect truth:

- Looper: any portion of a track with pT < 40 MeV (~10% in the available MC data)
- Physics track: any portion of a track with pT > 50 MeV
- True positive: a cluster correctly tagged as belonging to a looper
- False positive: a physics track cluster wrongly tagged as being part of looper

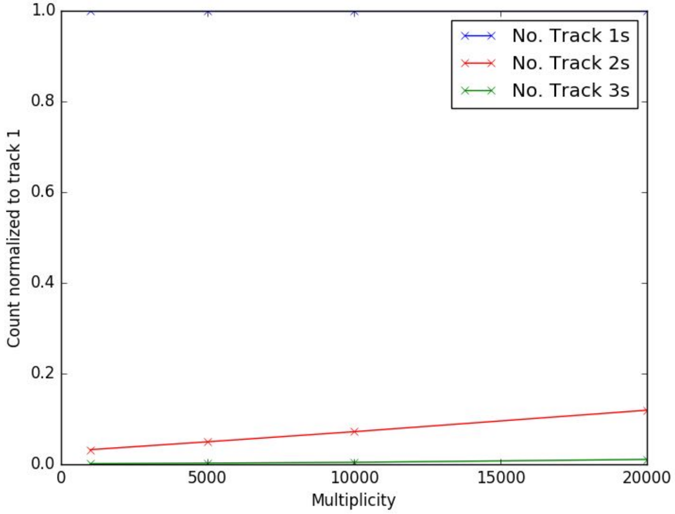
In reality, uninteresting data (junk) can be either of:

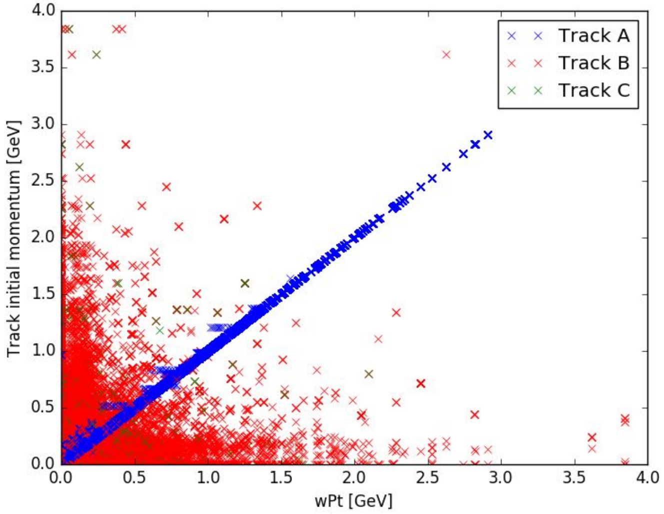
1) Noise: coming from electronics, not part of any track
2) Clusters belonging to low pT (< 40 MeV) loopers
3) Clusters from loopers of physics tracks with too large inclination angle above the pad row. These are not good for tracking.

The proposed looper finder is only looking for (2), which is 70-80% of junk.

Ground truth - in practice

- Each cluster can be assigned to 1-3 tracks, which may be low-pT or high-pT.
- We currently have access to a “weighted pT” (individual cluster momentum) corresponding to the first track.
- Up to 10% of clusters can have be part of a second track and may have a lower pT, therefore reducing TPR.





Looper finder algorithm - 2D

- The Hough transform accumulator space A is computed using a Gaussian kernel, and normalizing for circle arc length and radial distance from main beam axis:

$$A(x,y,r) = \sum_{i=1}^{n} G(r - \sqrt{(x_i - x)^2 + (y_i - y)^2}, \sigma) \times \frac{(x_i^2 + y_i^2)^{0.75}}{2\pi r}$$

where: $$G(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

- The accumulator space is discretized with a resolution of 1cm in *x,y,r*
- The limits of *x,y* correspond to the detector geometry
- The upper limit of *r* corresponds to the user-defined threshold on Pt, while the lower limit is fixed at 1 cm
- In general we use a Gaussian kernel with fixed σ = 1 cm, truncated to ±2 cm
- Each entry in A represents the likelihood for a looper with axis at *x,y* and radius *r*

Looper finder algorithm - 2D

- Then, the label or likelihood for a particular cluster is computed as:

$$\lambda_i = \max_{x,y,r} A(x, y, r) G(r - \sqrt{(x_i - x)^2 + (y_i - y)^2}, \sigma)$$

- This represents the likelihood that the given cluster forms part of any looper within the parameter range of the accumulator space.
- Note that the algorithm does not associate a cluster to a specific looper, but only computes the aggregate likelihood over all considered loopers.
- This allows to carry all information until the final threshold decision is taken per cluster.

Looper finder algorithm - 3D

- In the 3D version, after computing the 2D accumulator space A, we select the set of candidate circles as the parameter tuples (*x,y,r*) that correspond to the highest contribution from at least one cluster.
- This is determined by:

$$U(\arg\max_{x,y,r} A(x,y,r) G(r - \sqrt{(x_i - x)^2 + (y_i - y)^2}, \sigma)) = 1 \quad \forall i = 1...n$$

- This step effectively maps a cluster to a 2D candidate looper.
- Next, we iterate over the parameter space for tuples where *U*(*x,y,r*) = 1; for each:
  - We extract the list of clusters that fit within an annulus of ±2 cm
  - We sort these in increasing *z* and calculate the finite difference in *z* and θ for adjacent clusters, where θ is the radial angle of the looper.
  - We fit a straight line to the relationship of δ*z* and δθ to determine the best fit helix pitch and starting angle

Looper finder algorithm - 3D

- Finally, we label each cluster with its contribution to this best fit helix.
- Each cluster is given the highest contribution over all helices [one for each *x,y,r* tuple where *U*(*x,y,r*) = 1]
- This is computed as:

$$\lambda_i = \max_{x,y,r} A(x,y,r)G(r - \sqrt{(x_i - x)^2 + (y_i - y)^2)}, \sigma) \times G(\theta(z_i) - tan^{-1}\frac{y_i - y}{x_i - x}, \sigma)$$

- where $\theta(z_i)$ is the fitted helix's angular position at the cluster's *z* position.

Results

- Receiver-Operating Characteristic (ROC) curves (true +ve vs false +ve) for:
  - Different ground truths
  - Fixed vs variable sigma
  - 2D vs 3D hough transform
  - Slicing in z

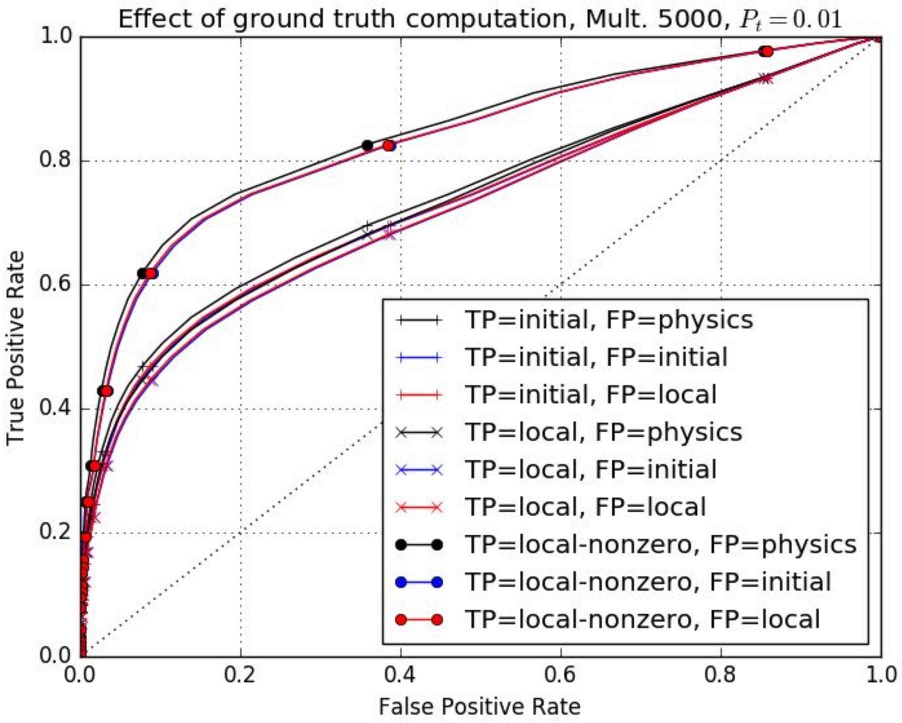
- Understanding the false positives

Results - different ground truths

- ROC curve for same algorithm but different interpretations of available ground truth.

TP	
initial	any(Pt < threshold)
local	wPt < threshold
local-nonzero	wPt > 0 AND wPt < threshold

FP	
physics	isUsed
initial	any(Pt > threshold)
local	wPt > threshold



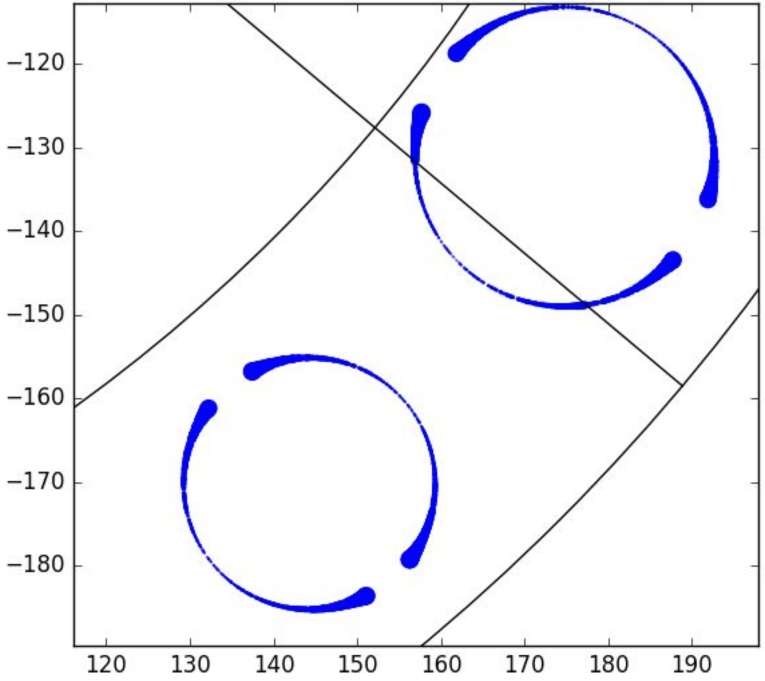
Results - fixed vs variable sigma

- **Fixed σ:**
  - Initial attempt has σ=1

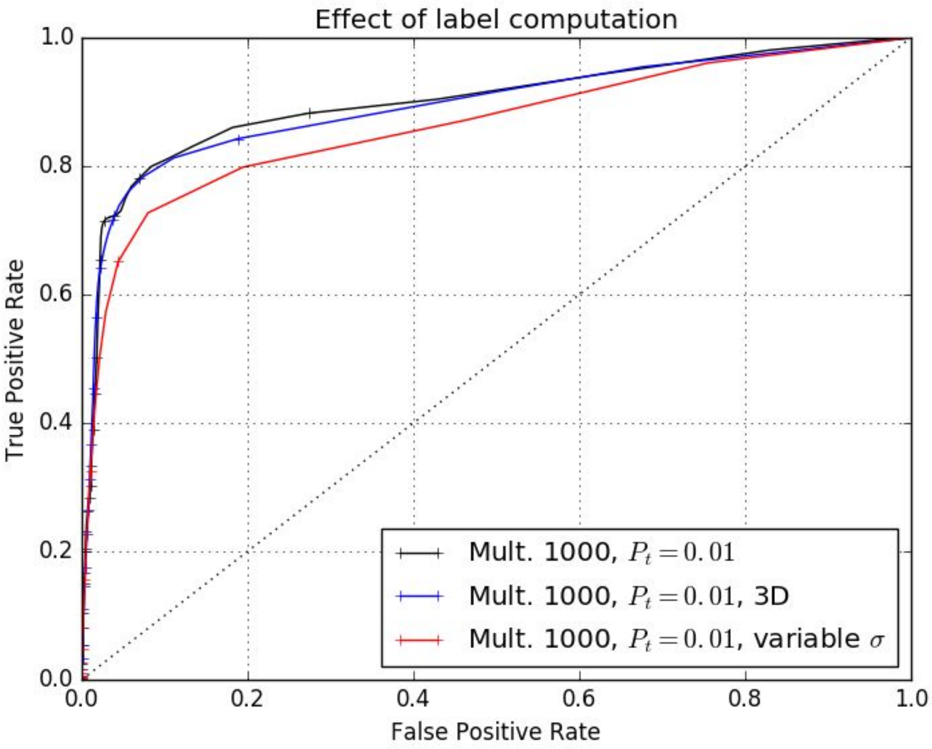
- **Variable σ:**
  - attempts to model physical uncertainty
  - Does not yet model quantization in Hough space
  - Figure shows for σ<1

- In both cases, a bounding box of 2 cm x 2 cm is applied.

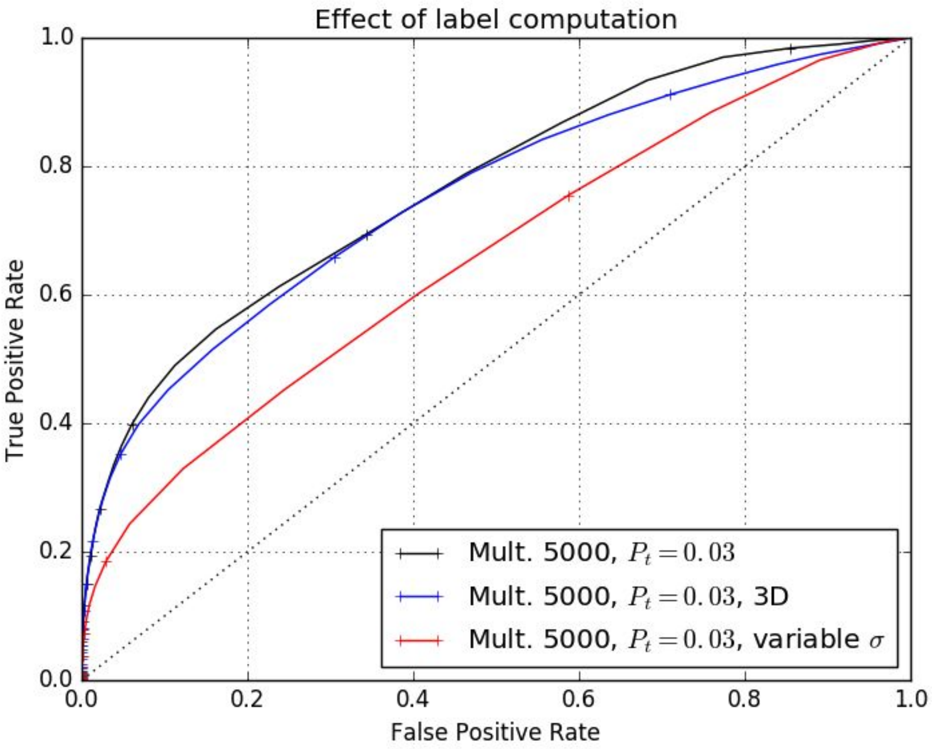
$$s=\lfloor \arctan(y/x)/(\pi/9) \rfloor$$
$$\alpha=(s+1/2)\times\pi/9$$
$$\phi=\arctan(dy/dx)-\alpha+\pi/2$$
$$\sigma=\sqrt{0.03^2+L\times0.006^2+(0.15\times L\times\tan(\phi))^2}$$



Results - fixed (2D vs 3D) vs variable sigma



Results - fixed (2D vs 3D) vs variable sigma

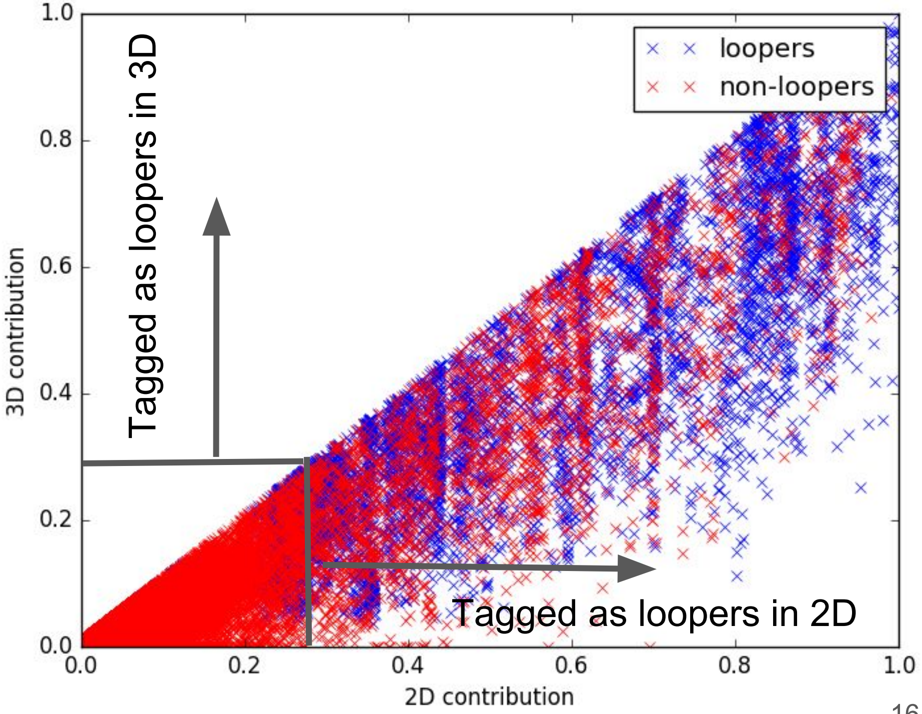


Results - fixed sigma, 2D vs 3D

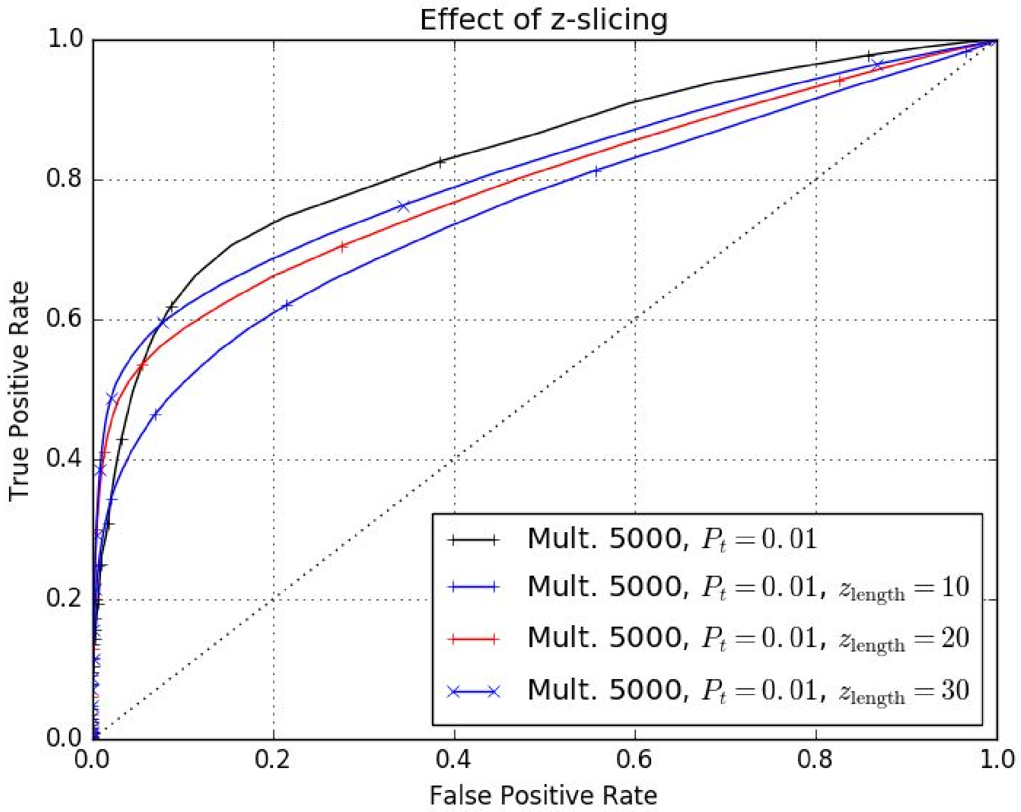
- With 3D, we are increasing the FPR with no gain in the TPR
  - We believe that the signal-to-noise ratio remains the same for 2D vs 3D.

- Plot shows the normalized accumulator space value achieved with 3D method as a function of 2D method.

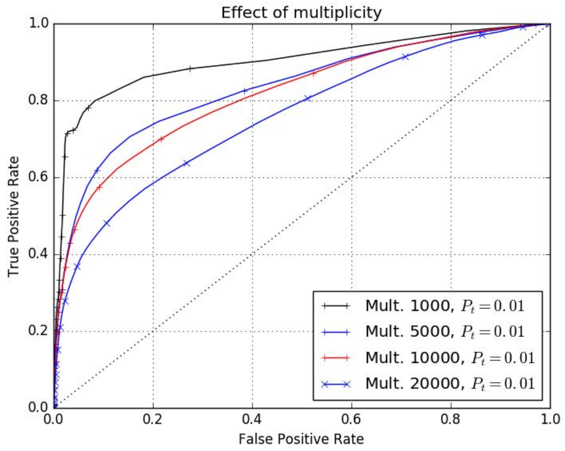
- The wPt ground truth is then used to distinguish between loopers and non-loopers.

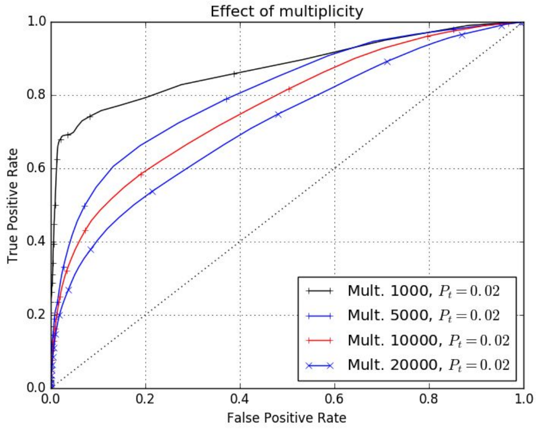


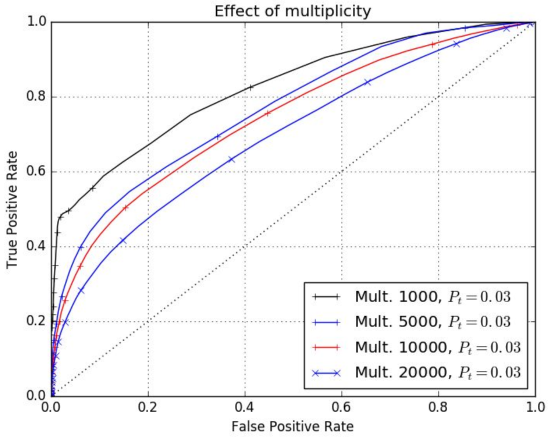
Results - Slicing in z

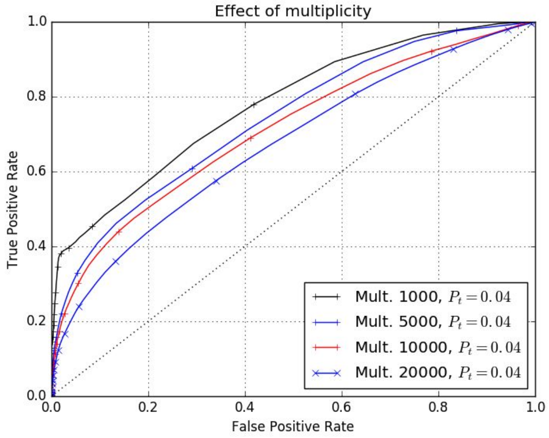


Results - by multiplicity for different pT









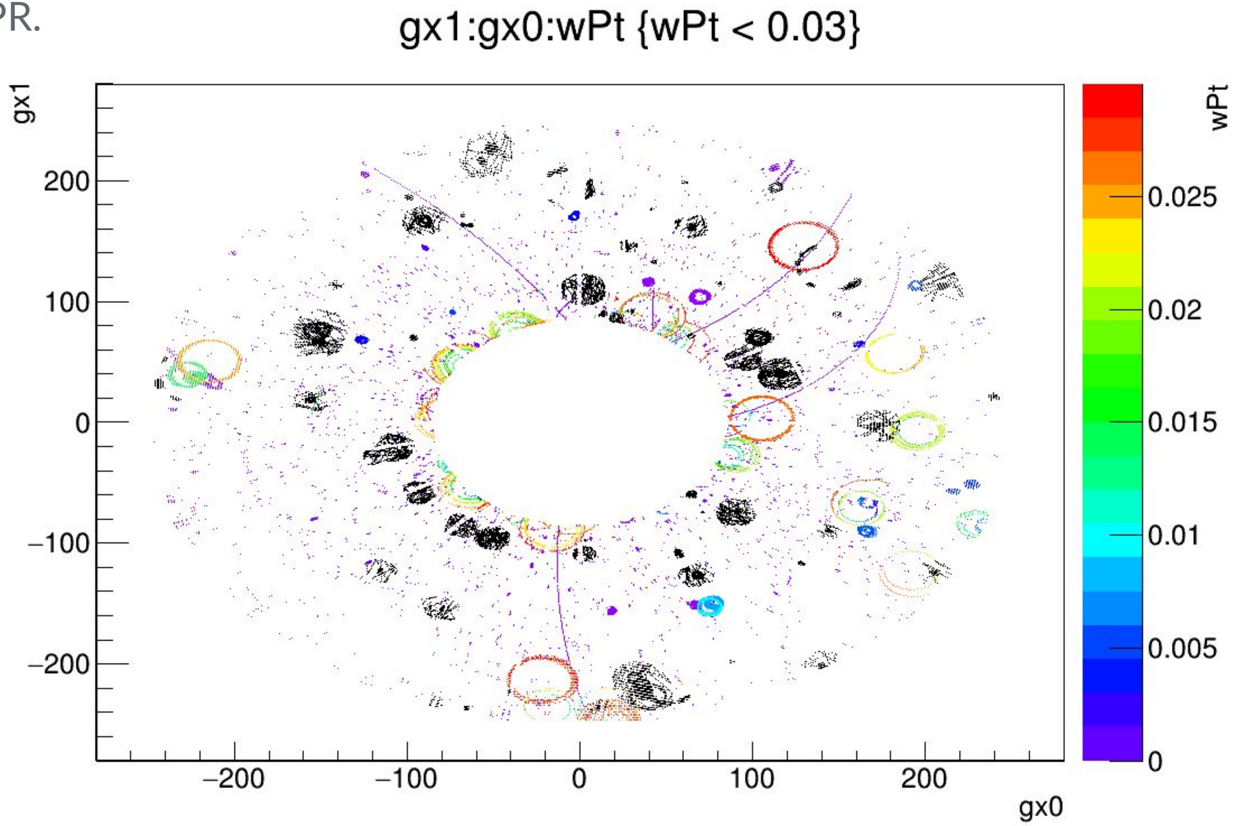
Results - timings

- Results are for most time-consuming case of pT < 50 MeV.
- Timings on:
  - CPU: Intel Xeon E5-1650 @ 3.5GHz
  - GPU: NVIDIA GTX Titan Black
- Still not fully optimized for GPU but for now more to have a working implementation.

Multiplicity	**2D (fit circles + label clusters)**	**3D (fit circles + identify circles + label clusters)**
1000	8s (GPU) + 1s (GPU)	3s (GPU) + 1s (GPU) + 112s (CPU)
5000	20s (GPU) + 2s (GPU)	18s (GPU) + 2s (GPU) + 901s (CPU)
10000	38s (GPU) + 4s (GPU)	38s (GPU) + 4s (GPU) + 2308s (CPU)
20000	74s (GPU) + 7s (GPU)	74s (GPU) + 7s (GPU) + 5457s (CPU)

HT classifier - 2D visualization

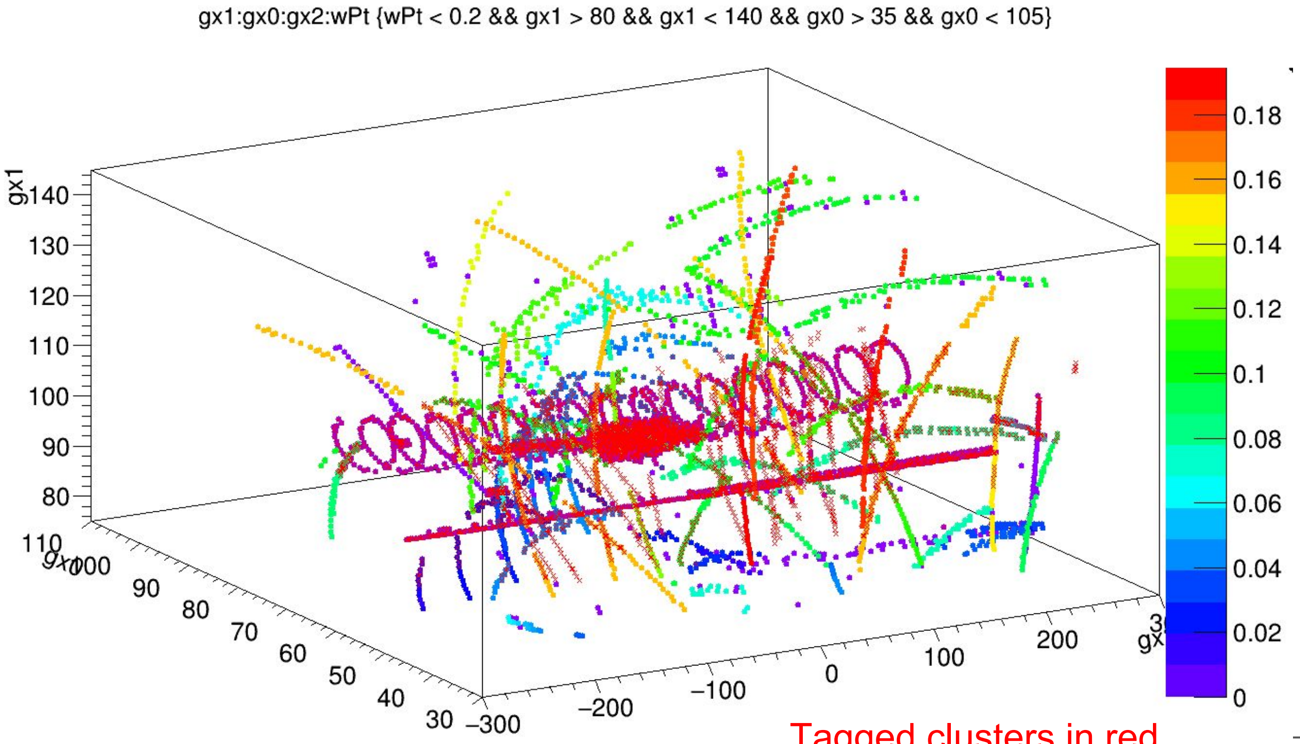
- Mult 1000: ground truth + tagging up to pT = 10 MeV @ ROC operating point: 81% TPR, 10% FPR.



Tagged clusters in black

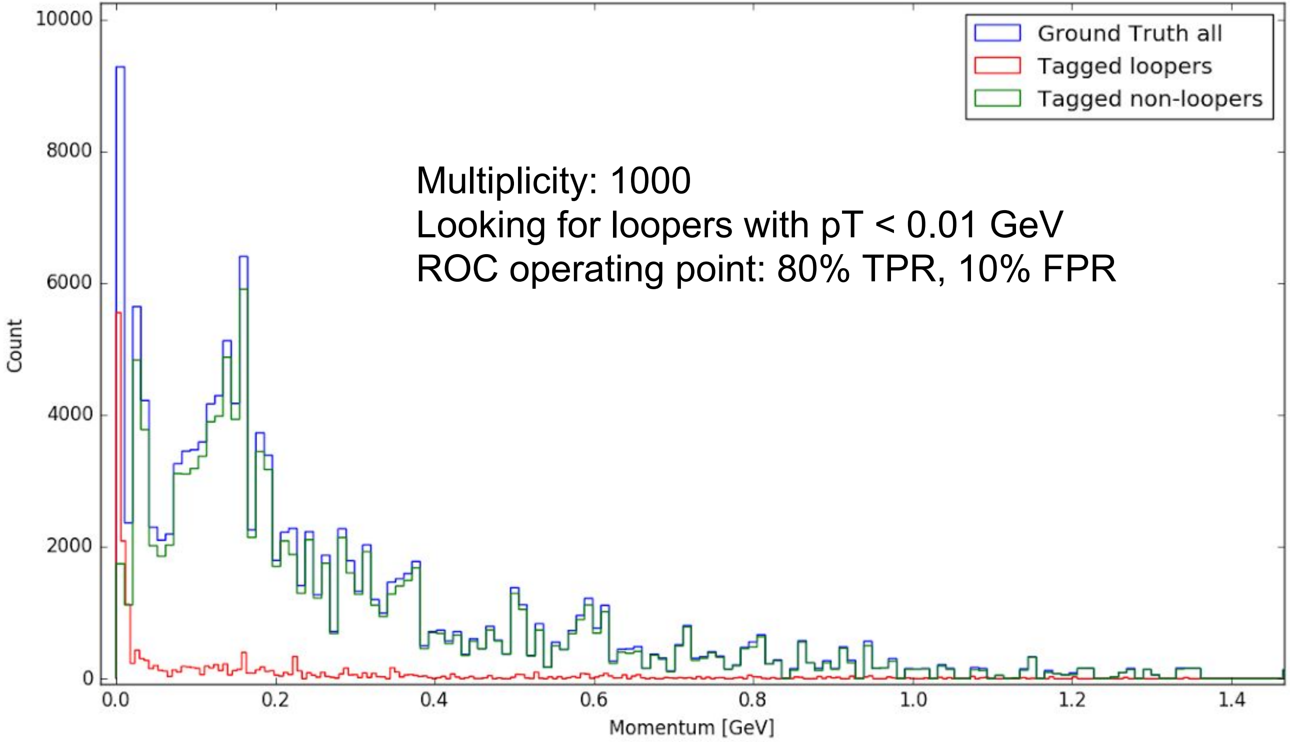
HT classifier - 3D visualization

- Mult 1000: ground truth + tagging up to pT = 10 MeV @ ROC operating point: 81% TPR, 10% FPR.



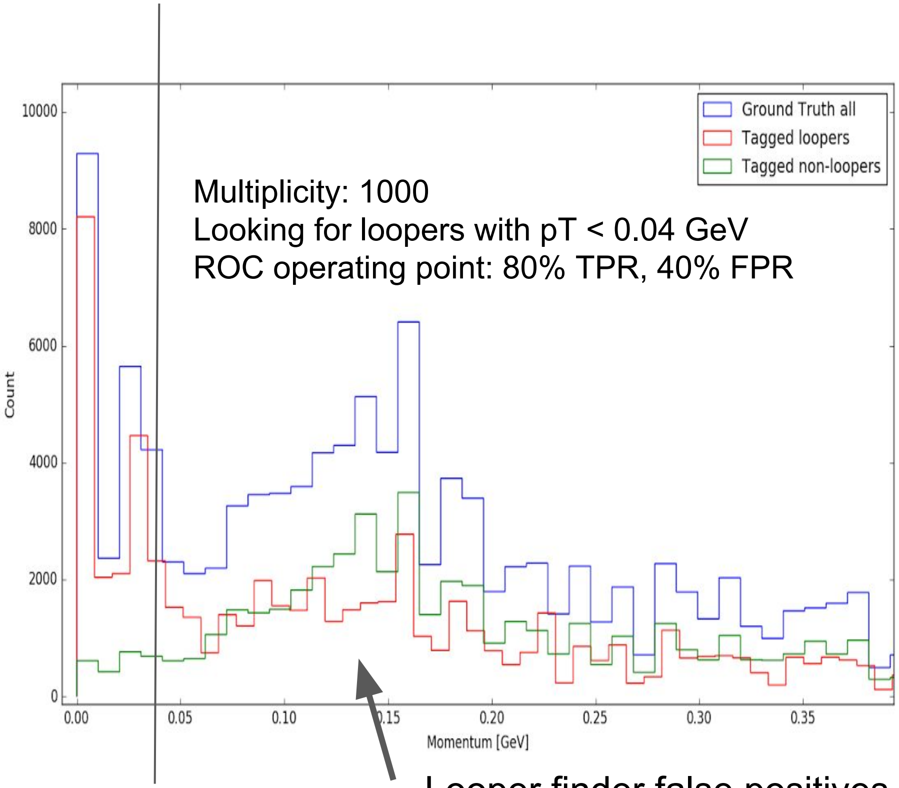
Tagged clusters in red

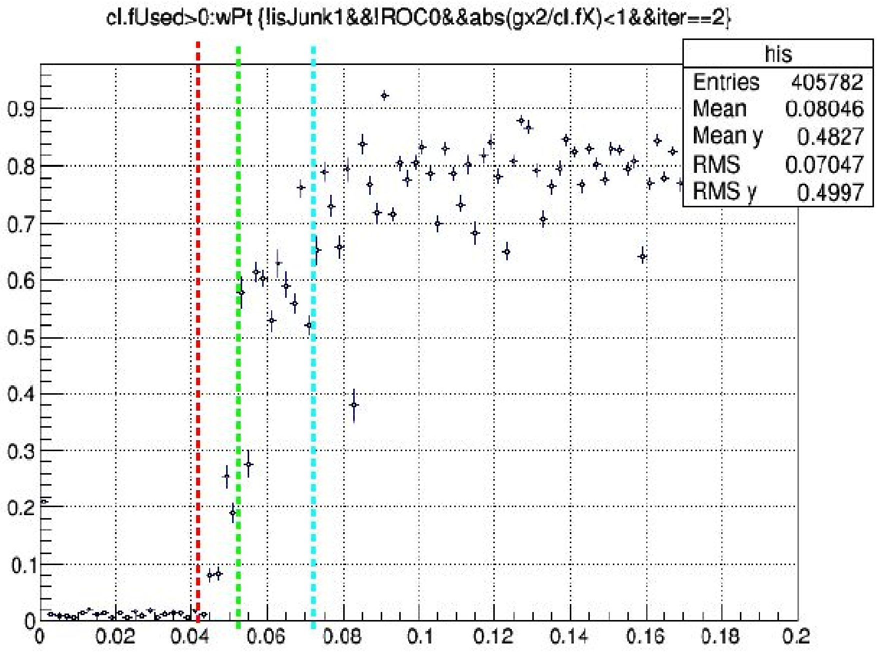
Results - understanding the false positives



Multiplicity: 1000
Looking for loopers with pT < 0.01 GeV
ROC operating point: 80% TPR, 10% FPR

Results - combining looper finder + offline tracker





M. Ivanov

Looper finder false positives which could be avoided with offline tracker

Conclusions

- Currently awaiting complete ground truth, i.e. pT per cluster per track: this should show a more correct (and probably better) ROC curves further if the tracks are ordered by decreasing pT.
- Implement sliding window in z for dE/dX and multiple scattering effects.
- Try to reduce false positives -> quantize the radius into fixed intervals, instead of searching for all clusters belonging to any looper with momentum < X GeV.