GNN Track Reconstruction of Non-helical BSM Signatures

Qiyu Sha, Daniel Murnane, Daniel Whiteson, Max Fieg

Institute of High Energy Physics CAS

University of California Irvine

Lawrence Berkeley National Lab

Quantum Computing and Machine Learning Workshop 2024

Quirk Introduction

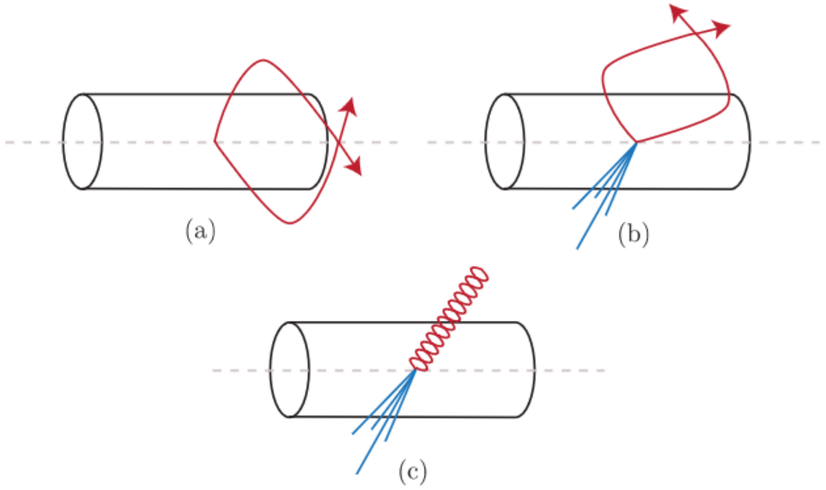
Quirks are stable BSM particles that are charged under an unbroken non-Abelian gauge force which confines at low energies:

- Used in models of dark matter, little Higgs scenarios, folded SUSY…

Quirks are characterized by a hidden QCD-like confinement scale $\boldsymbol{\Lambda}$ and mass m_Q with:

$$\boldsymbol{\Lambda} << m_Q$$

- Once produced quirks are separated by a QCD-like color-string which keep the quirk pair neutral
- But as opposed to the SM, the small energy stored in the string is insufficient to produce a quirk pair and thus preventing hadronization



0805.4642

Quirks are subjected to a restoring force with the scale $\boldsymbol{\Lambda}^2$ and exhibit oscillations on the scale(γ is the Lorentz boost of the quirk pair)

$$d_{cm} \approx 2\ \mathrm{cm}(\gamma - 1)\left(\frac{m_Q}{100\ \mathrm{GeV}}\right)\left(\frac{\mathrm{keV}}{\Lambda}\right)^2$$

What’s the plan

- In ATLAS, searching for tracks among particles hits is a huge combinatorial problem, usually aided by making assumptions about particle trajectories.
- Those assumptions limit our ability to discover particles which violate those assumptions, such as the quirk which has the oscillations rather than standard helices. Writing a dedicated quirk tracker would require a complete rewrite.
- Use the more flexible ML-based tracking algorithm to learn maybe a good way to find quirks.

1. Does the GNN tracking work for non-helical tracks?

2. When it is trained on SM (i.e. mostly helical) tracks, can it work on non-helical?

3. When it is trained on non-helical, can it work on non-helical?

4. When we mix the SM and non-helical tracks and training on them, can the trained model work on search non-helical in the mixed dataset?

5. Will the reconstruction efficiency be different for quirks of different mass and scale?

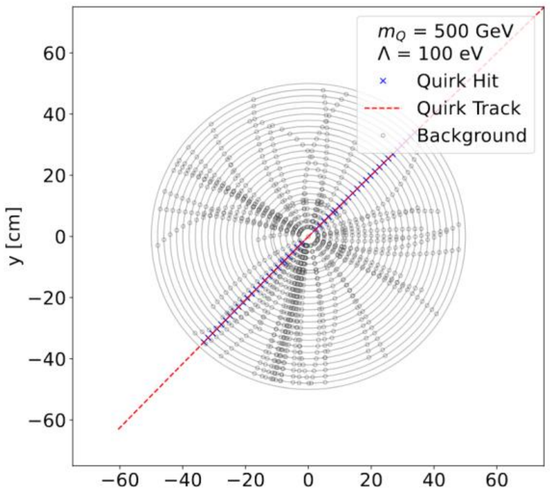
Dataset

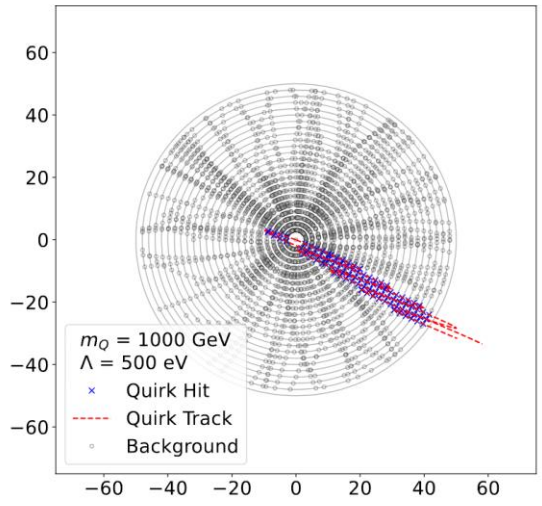
- Use MG5 generate samples through pp $\rightarrow Q\bar{Q}$ + j:
  - Quirk：Collect Quirk and through a simplistic model of the ATLAS detector which consists of 8 layers/25layers of trackers.

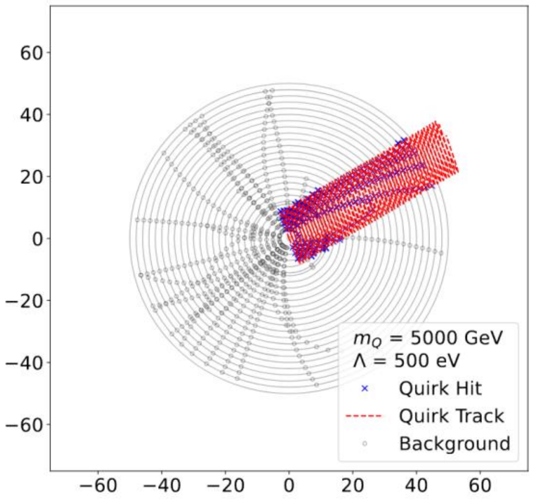
1708.02243

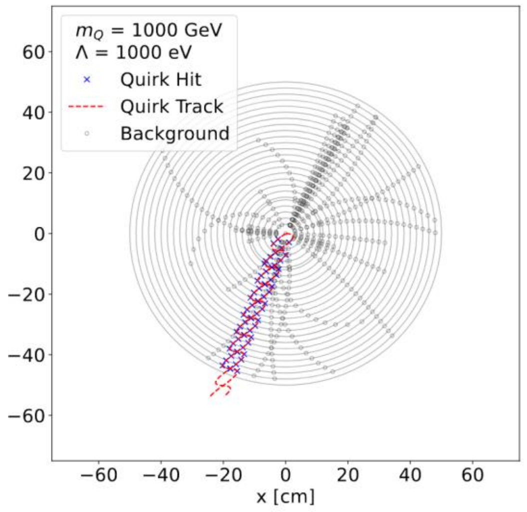
- For 8 layers, 500 GeV quirk pair with the string tension (Lambda) = 500 eV
- For 25 layers, we generate samples of simulated quirks in the mass range [100,5000] GeV and Lambda range[100, 5000] eV.
  - Background：Jet (~100 particles for one event)

- The quirk track becomes more complex and crazy when mass/scale larger.



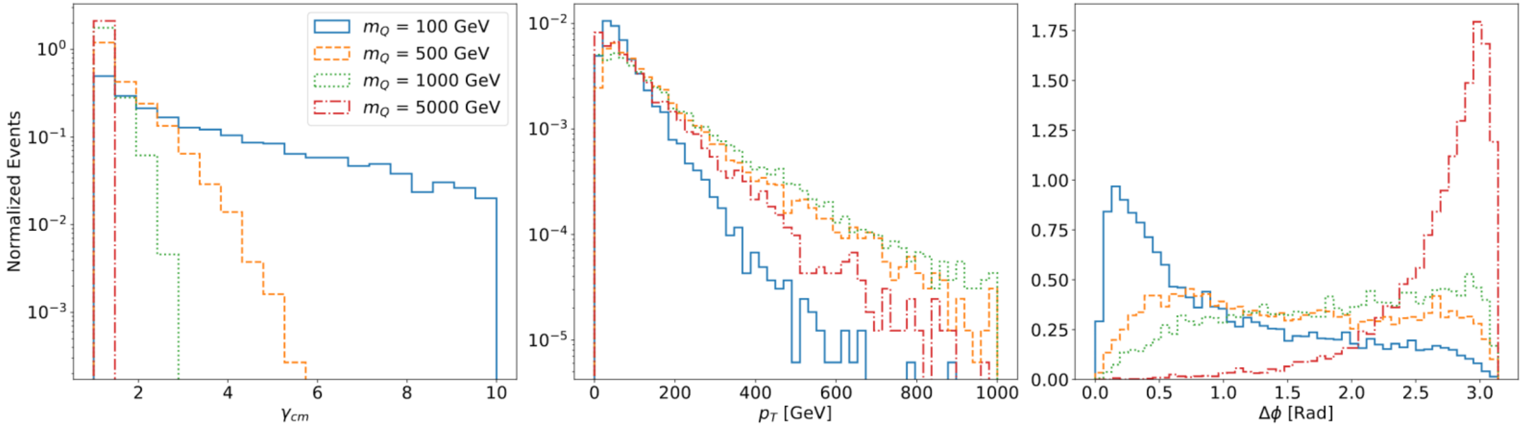


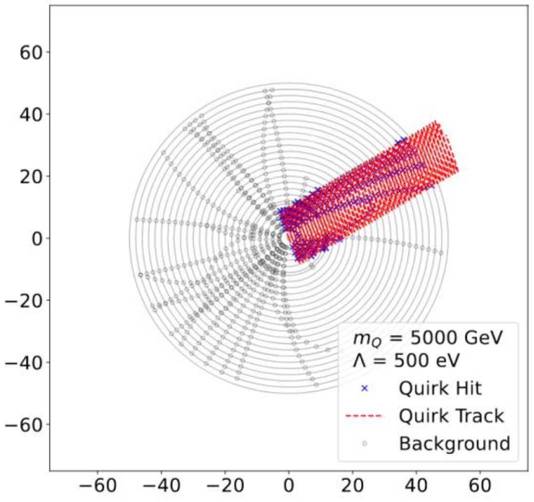


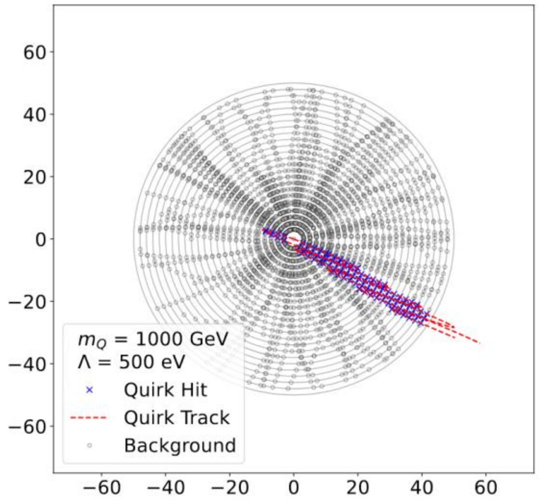


Quirk Dataset

The Lorentz factor γ, the transverse momentum P_T and opening angle $\Delta\phi$ of the quirk pair for different samples.







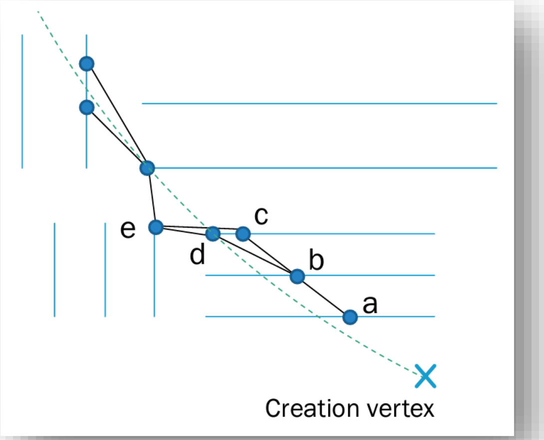
To avoid the crazy tracks (Which in-out one layer repeatedly) We focus on one simple category of quirk tracks initially (This is the first study for "well-behaved" quirks), so we do the simple selection on Quirk dataset:

- $N_{hit} < 3^* \ N_{layer}$
- The eff of this "well-behaved" selection is 40%-90% for our dataset.

Pipeline

Construct graph → Metric Learning → Classify edges → GNN → Identify connect components

Based on Exa.Trkx.



Defining “True neighbors”:

- For each particle, order hits by increasing the “hit_id”
- Group by shared module ID
- Connect all combinations from layer L_i to L_{i+1}

Metric Learning:

- For all hits in detector, embed features into N-dimensional space.
- Associate neighboring hits as close in N-dimensional distance.
- Score each “neighbour” hit within embedding neighborhood against the “target” hit at center.

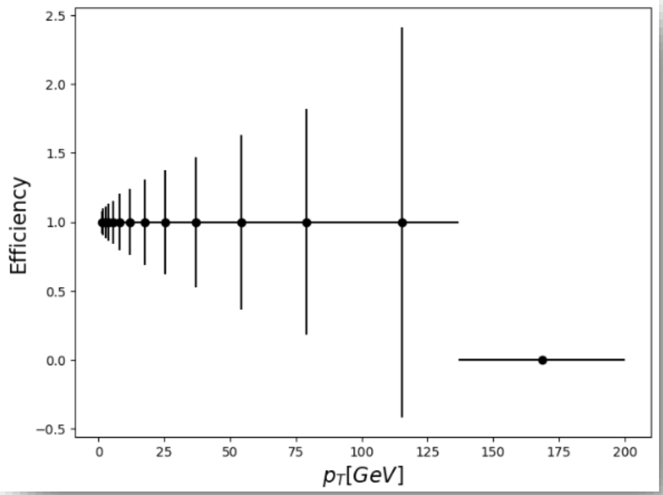
Training details – 8layers background training

- Training on the background dataset(SM tracks)
- 8layers, 500 GeV quirk pair with the string tension (Lambda) = 500 eV
- 1k events to train on.

The track definition
$\|\eta\| < 4$
$n_{track}^{hits} \geq 5$
$n_{particle}^{hits} \geq 7$
Double-majority matching

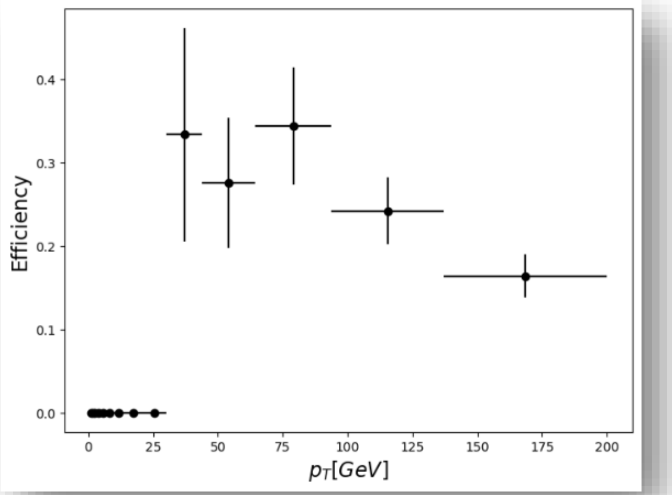
- background inference:

97.9% reconstructed efficiency



- quirk inference (8layers, mass= 500GeV, Lambda = 500eV):

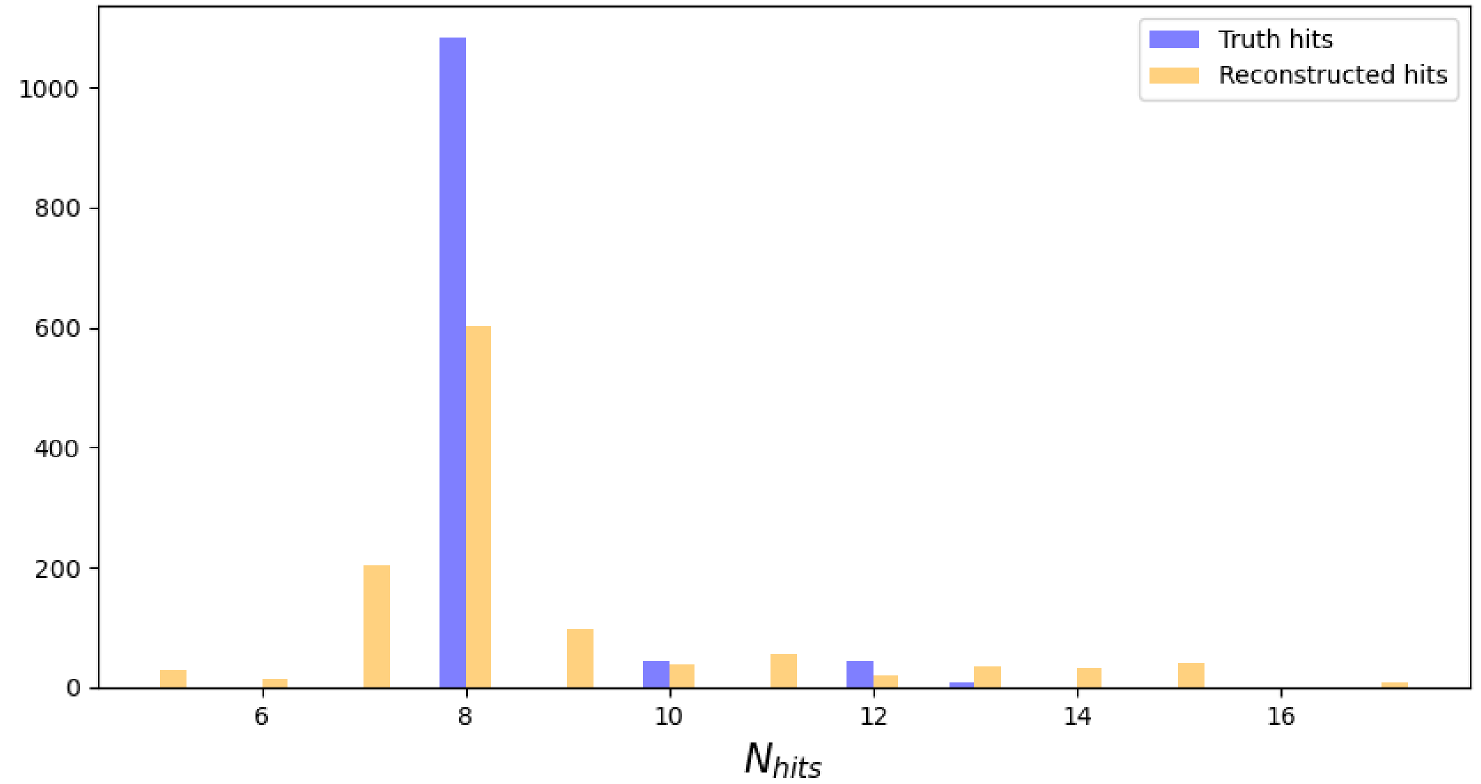
10.2% reconstructed efficiency



Distribution of reconstructed quirks

The distribution of reconstructed quirks' information: (quirk inference)

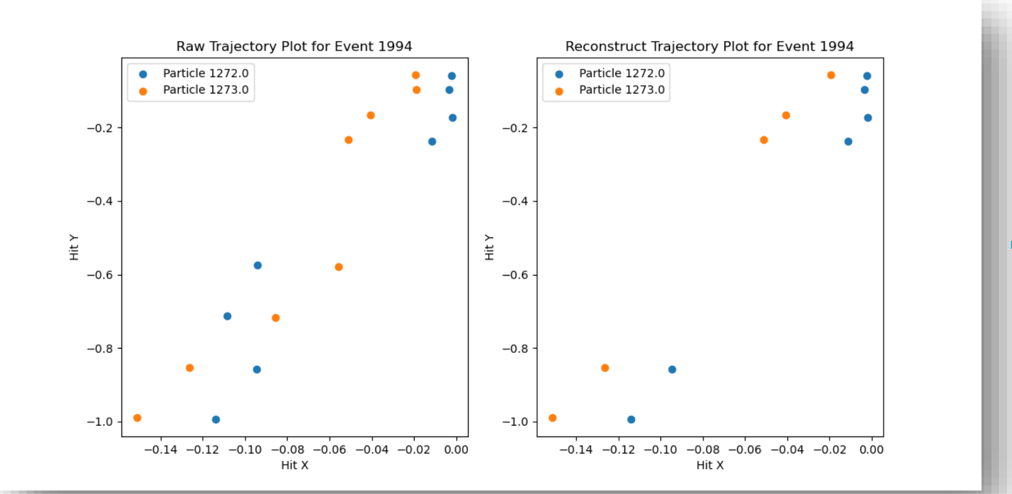
- $r, \phi, z(cm)$ are truth information of hits. r is scaled to (0,1). The plots are shown in the backup.
- n_{reco}^{hits} is the number of reconstructed hits, n_{truth}^{hits} is the number of truth hits.

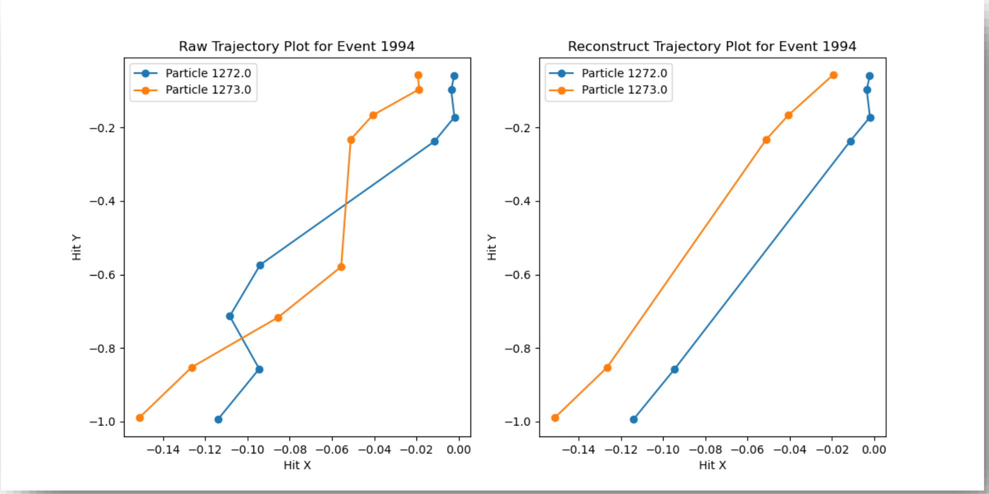


Reconstructed hits of quirk

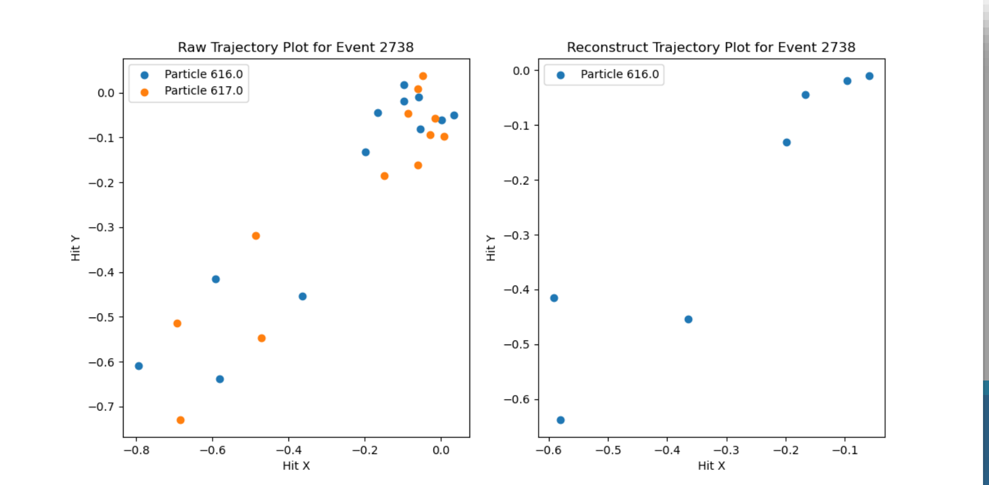
With same event (use the reconstructed event information):

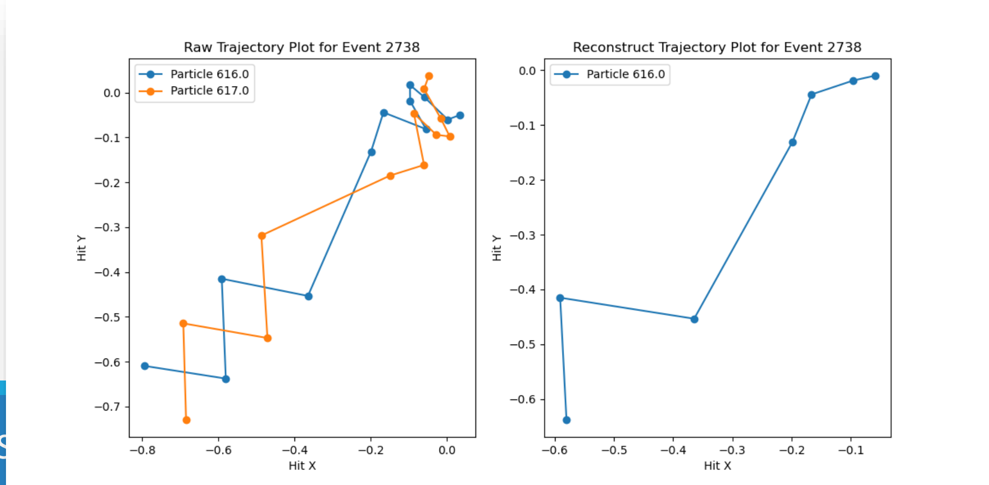
- Some $\mathrm{hits}_{\mathrm{reco}}$ are the part of truth quirk track.





- Only reconstruct simple and smooth track. (The particle 617 is failed to be reconstructed)





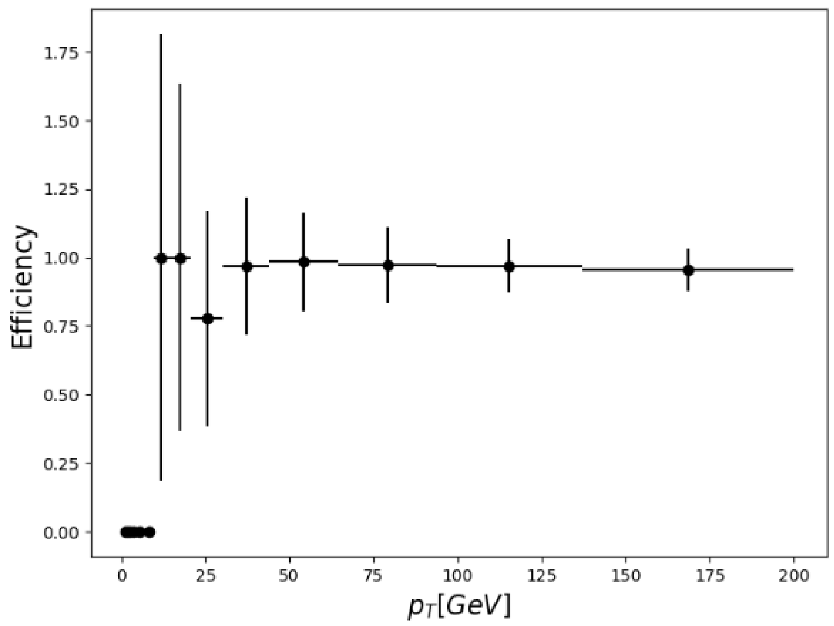
Results: Quirk training, quirk inference

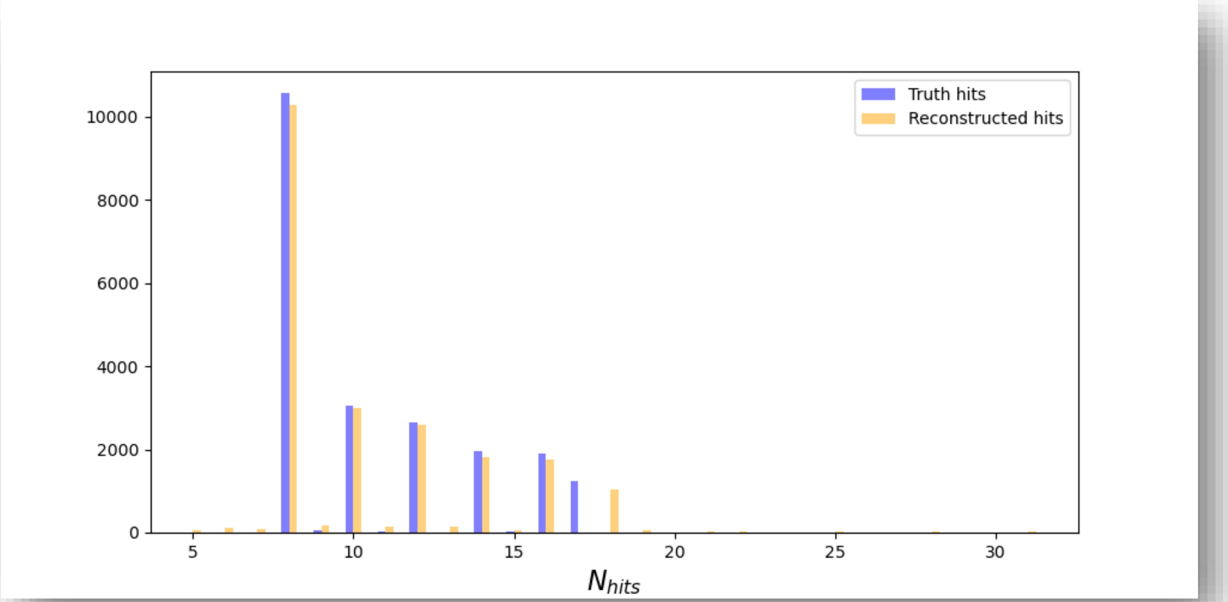
- 8layers, 500 GeV quirk pair with the string tension (Lambda) = 500 eV
- 1k events to train on.

Well-behaved Quirk training, quirk inference: 91.5% reconstructed efficiency

The distribution of reconstructed quirks' information:

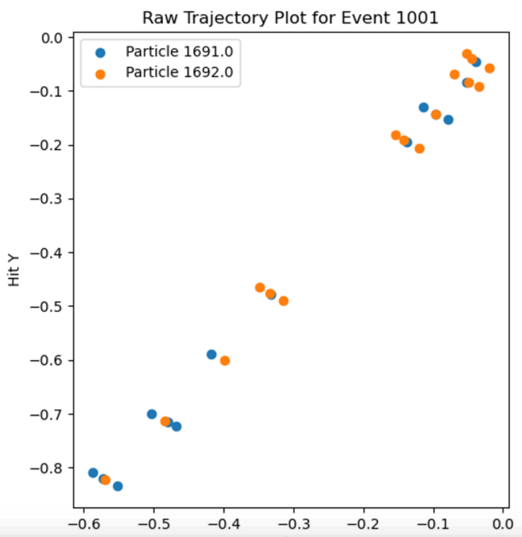
- $r, \phi, z(cm)$ are truth information of hits. r is scaled to (0,1). The plots are shown in the backup.
- n_{reco}^{hits} is the number of reconstructed hits, n_{truth}^{hits} is the number of truth hits.

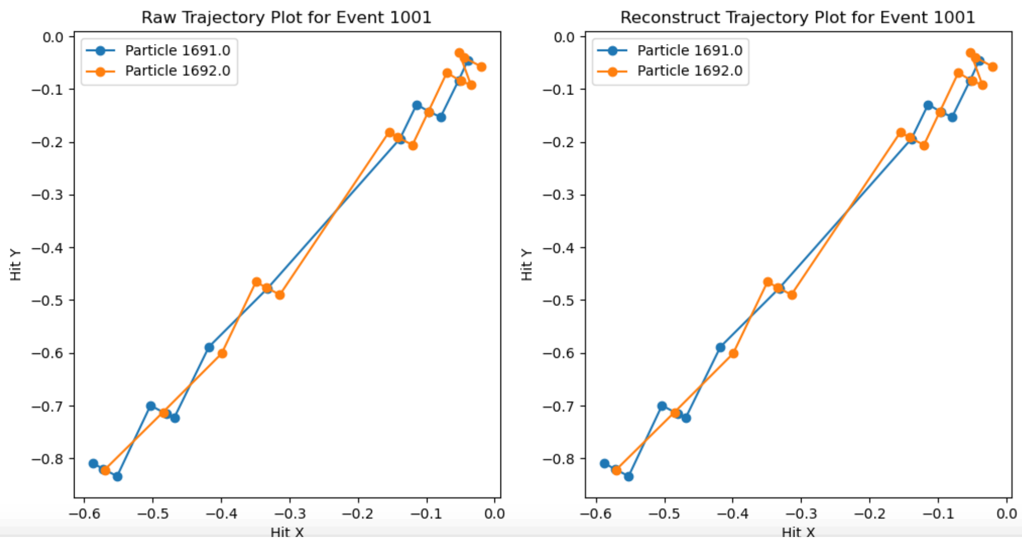


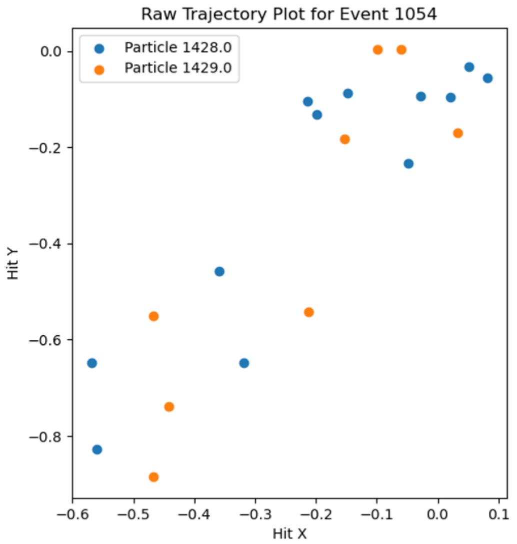


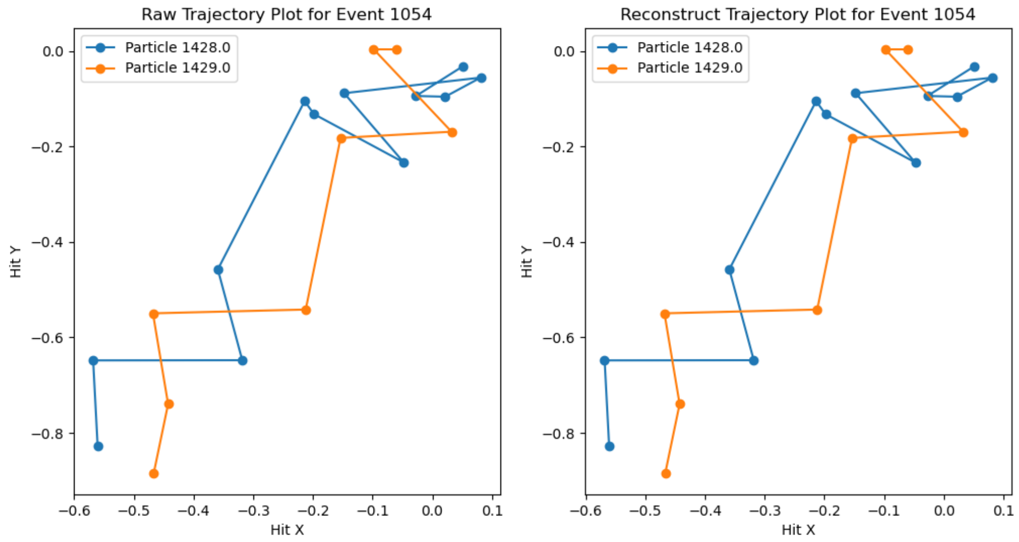
Reconstructed hits of quirk

All of well-behaved quirks are reconstructed well even though the dot plot looks chaos:

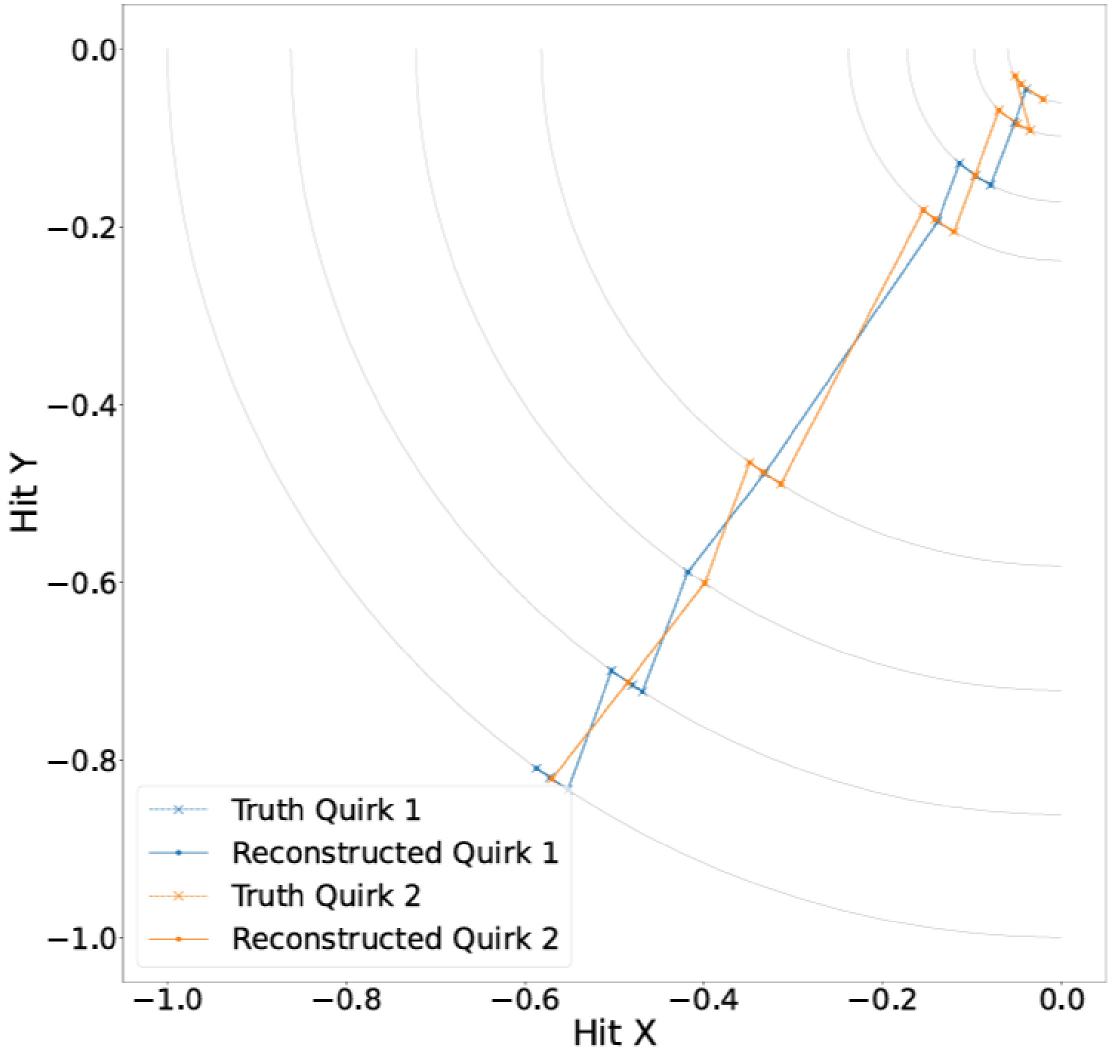


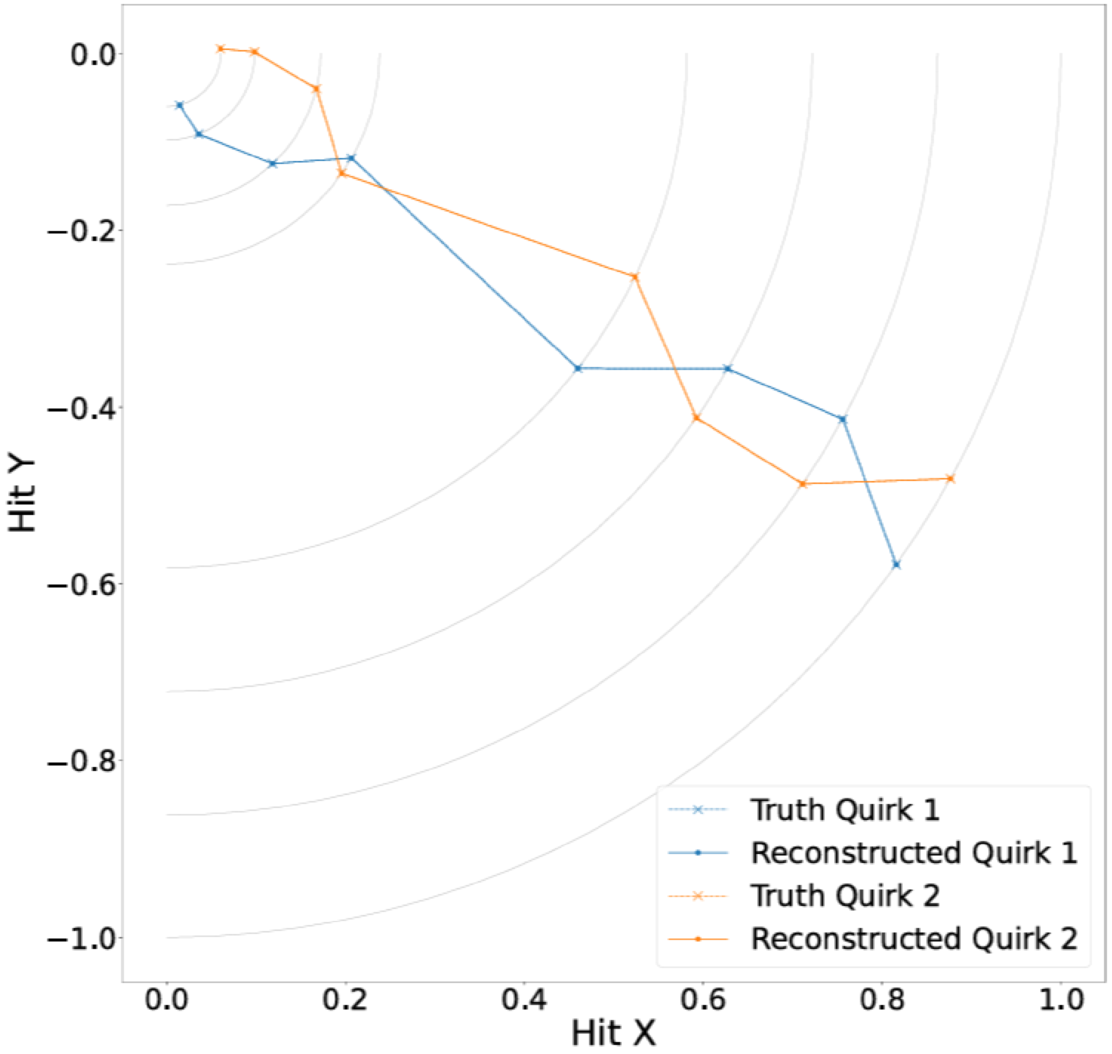






Reconstructed hits of quirk





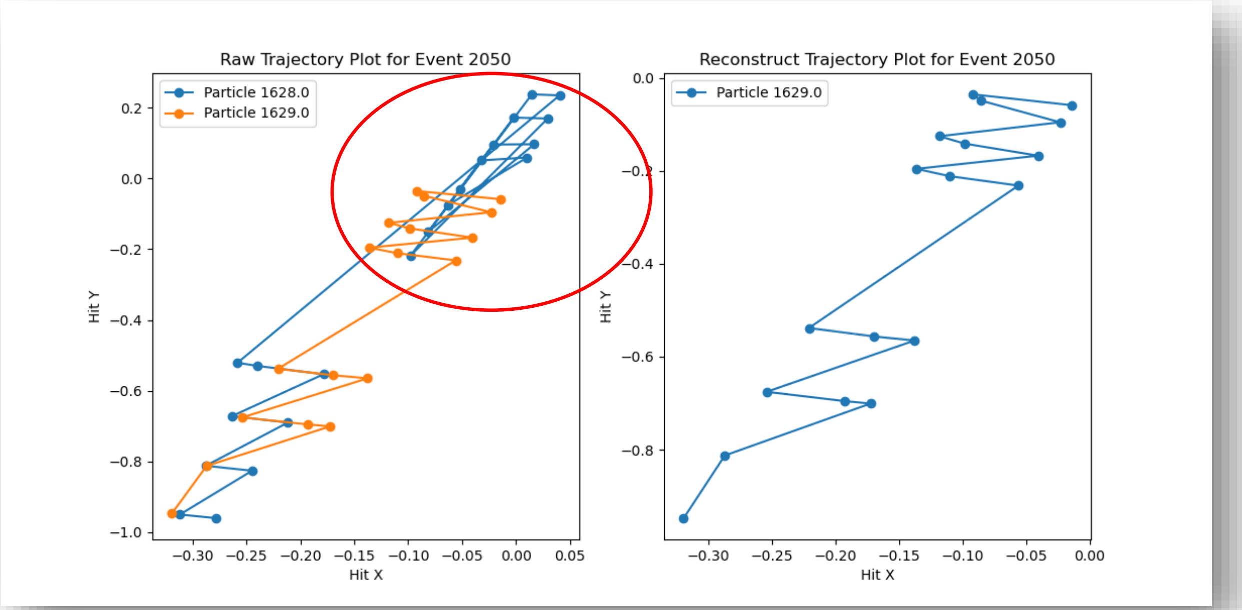
Qiyu Sha qsha@cern.cn

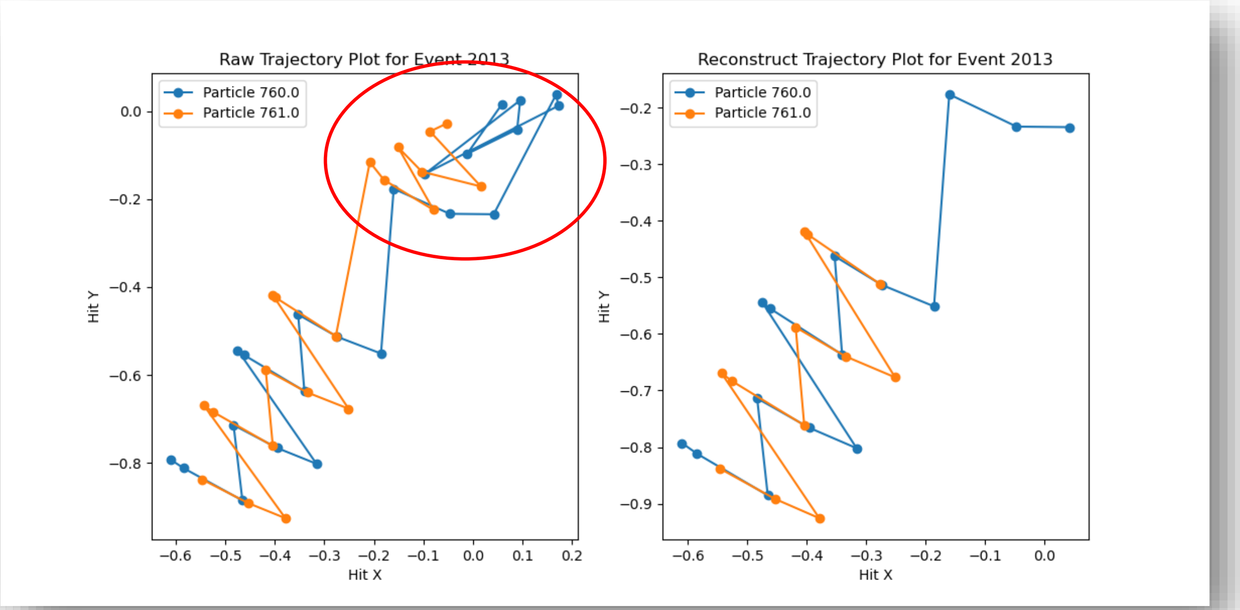
Results: All Quirk training, quirk inference

We training on all quirks without pre-selection, the performance has dropped significantly:
56.3% reconstructed efficiency

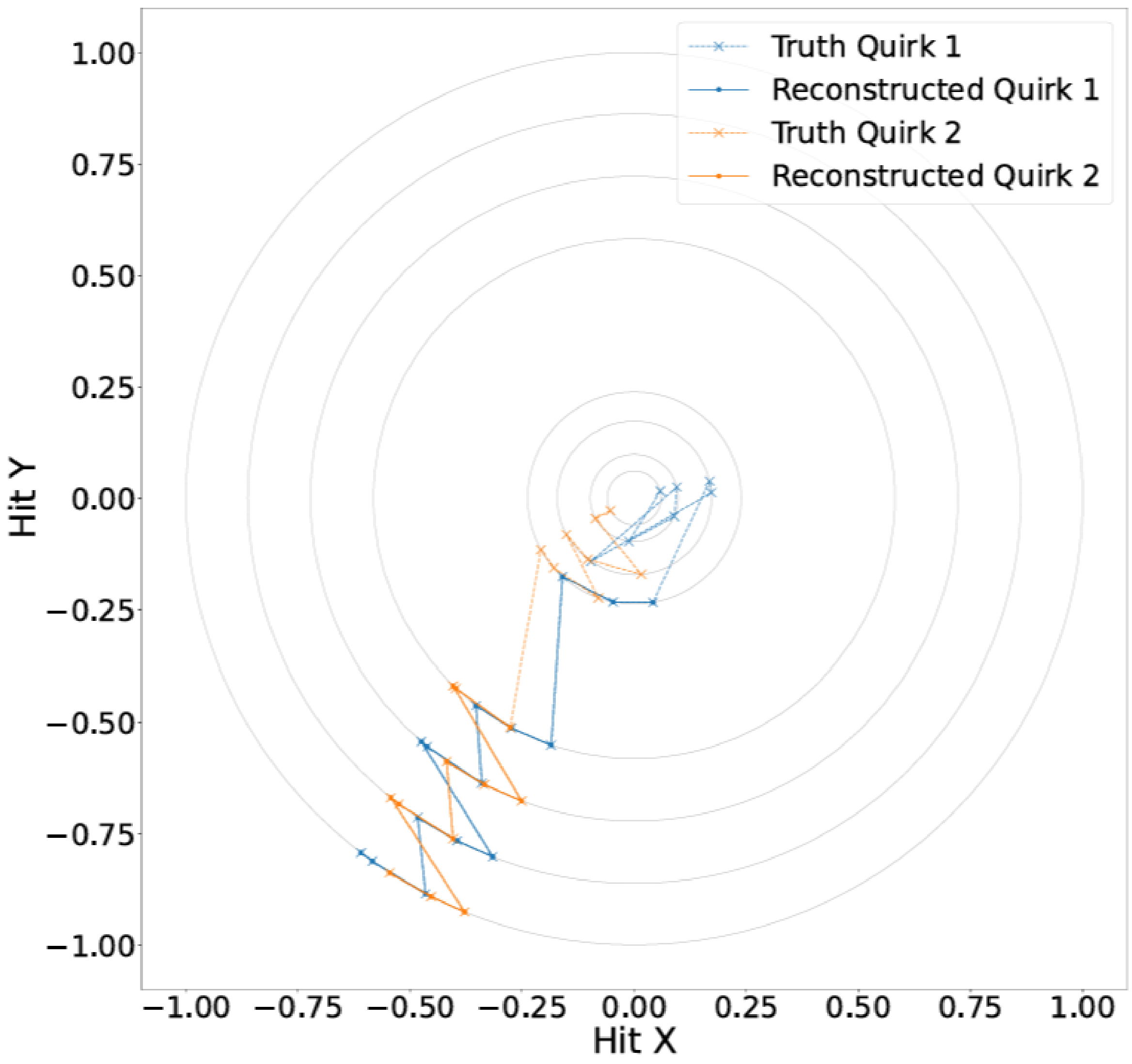
Also, When tracks become crazy with lots of hits and in-out layers, the reconstructed performance is bad:

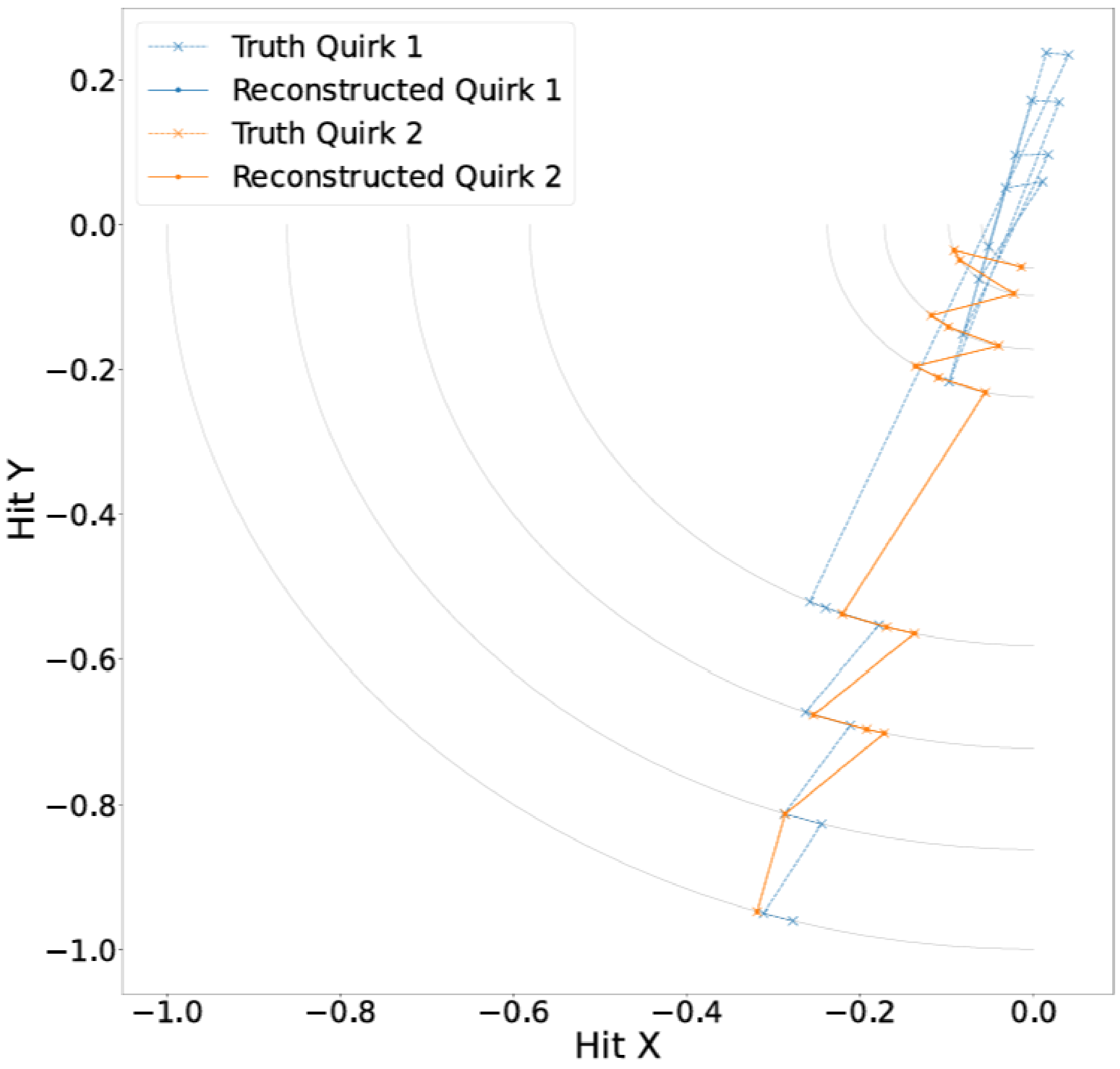
- “Well-behaved” selection is useful.





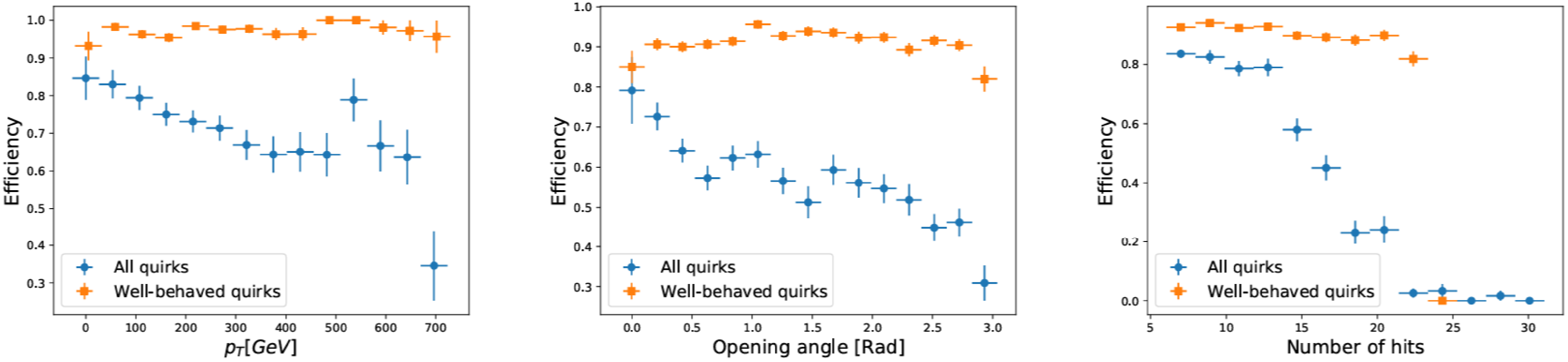
Results: All Quirk training, quirk inference





Results: All Quirk training, quirk inference

Efficiency versus P_T, quirk opening angle(center), and number of true hits(right) for well-behaved quirks or all quirks.

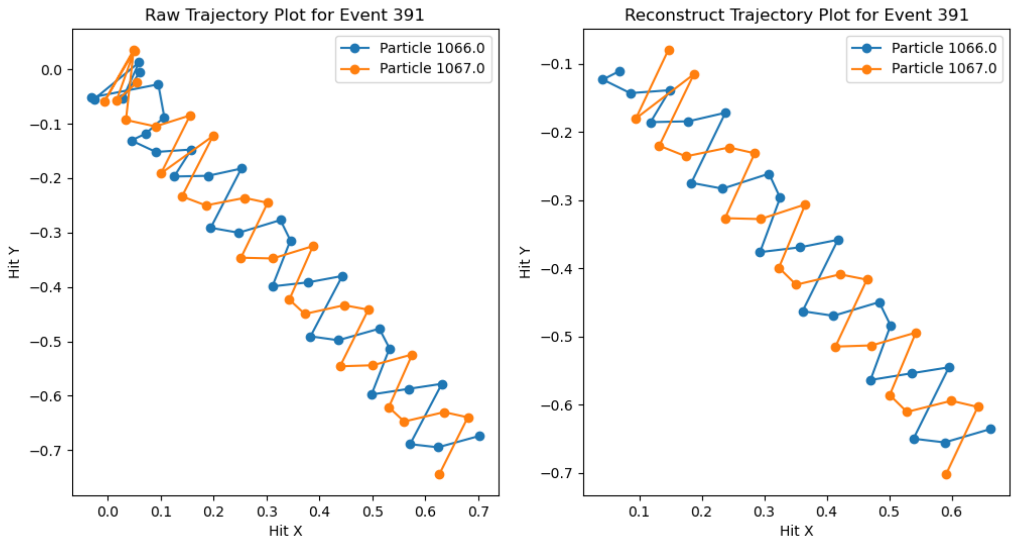


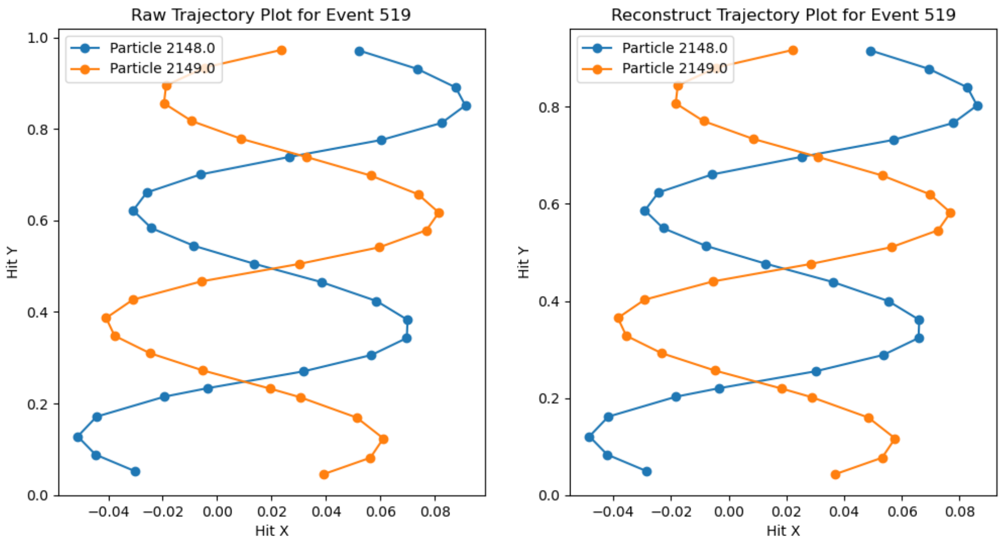
Results: Mix training, mix inference

- Considering the realistic applications of non-helical tracking will require identifying such tracks among many helical tracks from background processes. We mix the SM tracks and well-behaved track, then training on them.
- In this training, the quirk with positive labels as well as SM tracks with negative labels.

The 61.4% efficiency we get in the 8 layers, 500 GeV quirk pair with the 500eV string tension.
To explore the dependence of the number of tracking layers, the study was repeated in the 25 layers setting, yielding an efficiency of 79% which is better than 8 layers.

The reconstructed performance in the 25 layers:





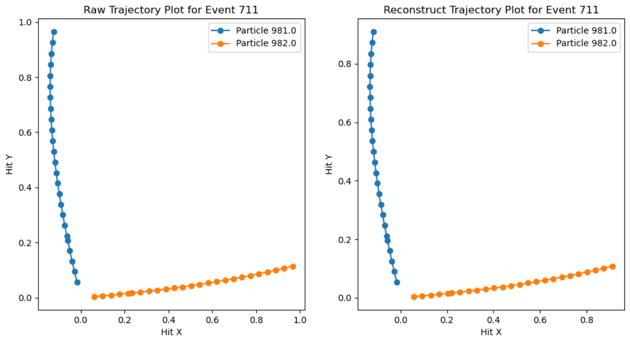
More datasets, more results

How oscillation length(d) affect the reconstruction efficiency?

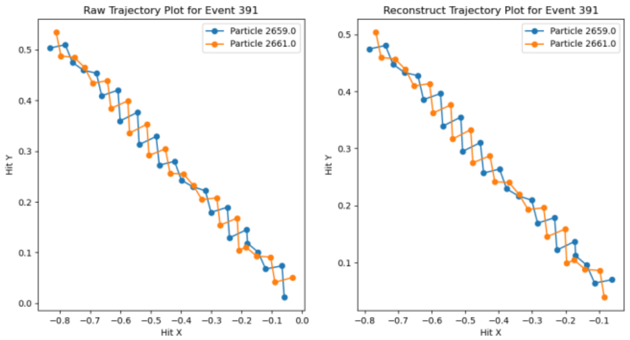
A scan of quirk parameters (m_Q, λ) is applied:

- The efficiency is smaller when oscillation length is smaller.
- More reconstruction performance are shown in backup.

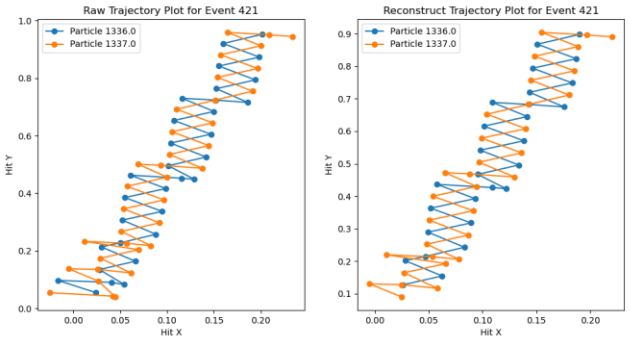
Mass_100_Lambda_100:



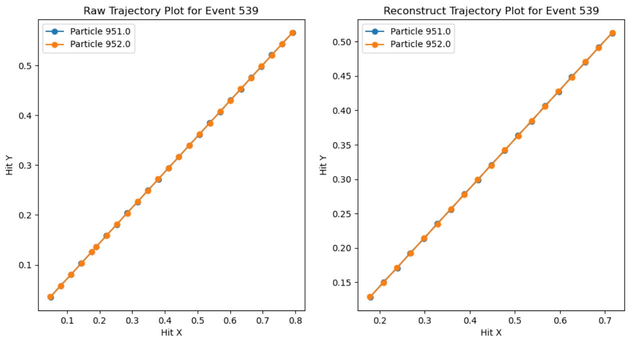
Mass_100_Lambda_1000:



Mass_100_Lambda_500:



Mass_100_Lambda_5000:



m_Q (GeV)	Λ(eV)	$\bar{\gamma}$	σ_γ	d [cm]	Efficiency	*Well-behaved* fraction
100	100	4.4	3.4	670	91.0%	88.3%
	500	3.7	2.7	21.6	82.8%	77.0%
	1000	3.1	2.05	4.2	77.8%	79.7%
	2000	3.0	2.0	1.0	60.4%	83.4%
	3000	3.0	1.9	0.44	35.6%	83.6%
	4000	2.9	1.8	0.24	34.5%	84.5%
	5000	2.9	1.7	0.15	24.2%	85.0%
500	100	1.9	0.7	896	92.0%	82.3%
	500	1.8	0.6	31	79.0%	51.2%
	1000	1.7	0.6	7.3	64.3%	53.1%
	2000	1.7	0.5	1.7	60.9%	59.9%
	3000	1.7	0.5	0.8	59.6%	62.3%
	4000	1.6	0.5	0.4	42.6%	63.0%
	5000	1.6	0.5	0.7	39.2%	63.8%
1000	100	1.5	0.3	950	92.7%	80.5%
	500	1.4	0.3	32	63.6%	40.2%
	1000	1.4	0.3	7.6	62.7%	42.3%
	2000	1.4	0.3	1.8	69.2%	48.6%
	3000	1.3	0.2	0.8	54.1%	51.7%
	4000	1.3	0.2	0.4	59.7%	52.9%
5000	100	1.04	0.03	420	84.8%	40.2%
	500	1.03	0.02	11	69.8%	32.6%
	1000	1.03	0.02	2.7	65.3%	35.2%
	2000	1.03	0.02	0.7	49.2%	39.6%
	3000	1.03	0.02	0.3	36.4%	40.8%
	4000	1.03	0.02	0.2	34.6%	41.2%

Some tests

- **Generalizability:**

Mix samples with different (m_Q, λ), then training on them and inference one interpolative point(the point which don't in the training mixed dataset, Mass_100_Lambda_1000.

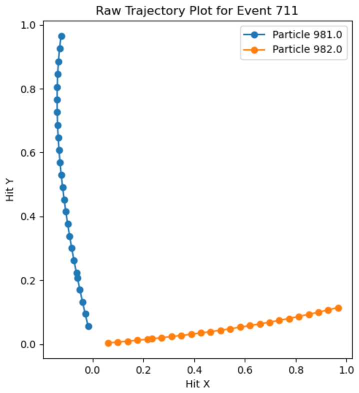
We can get the 52.6% efficiency.

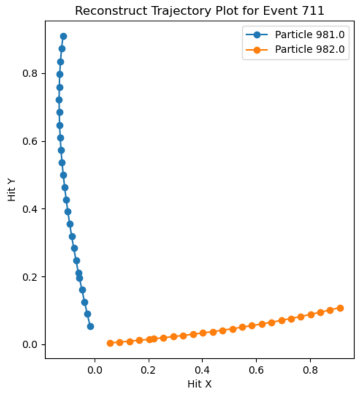
This result is worse than 79.7% which comes from training on Mass_100_Lambda_1000, but still work.

- **SM tracks training, quirk inference:**

Only λ=100eV have a good performance due to the quirk track in 100eV is similar as the SM tracks.

In other point, the reconstructed efficiency is worse than training on the mixed dataset.





m_Q (GeV)	Λ(eV)	Efficiency
100	100	81.3%
	500	19.0%
	1000	18.8%
	2000	15.4%
	3000	18.6%
	4000	17.0%
	5000	29.8%
500	100	89.9%
	500	7.0%
	1000	12.8%
	2000	11.8%
	3000	32.2%
	4000	21.0%
1000	100	93.2%
	500	5.4%
	1000	4.8%
	2000	31.2%
	3000	24.6%
	4000	29.6%
5000	100	80.5%
	500	2.2%
	1000	13.0%
	2000	8.0%
	3000	15.2%
	4000	18.8%

m_Q (GeV)	Λ(eV)	$\bar{\gamma}$	σ_γ	d [cm]	Efficiency	*Well-behaved* fraction
100	100	4.4	3.4	670	91.0%	88.3%
	500	3.7	2.7	21.6	82.8%	77.0%
	1000	3.1	2.05	4.2	77.8%	79.7%
	2000	3.0	2.0	1.0	60.4%	83.4%
	3000	3.0	1.9	0.44	35.6%	83.6%
	4000	2.9	1.8	0.24	34.5%	84.5%
	5000	2.9	1.7	0.15	24.2%	85.0%
500	100	1.9	0.7	896	92.0%	82.3%
	500	1.8	0.6	31	79.0%	51.2%
	1000	1.7	0.6	7.3	64.3%	53.1%
	2000	1.7	0.5	1.7	60.9%	59.9%
	3000	1.7	0.5	0.8	59.6%	62.3%
	4000	1.6	0.5	0.4	42.6%	63.0%
	5000	1.6	0.5	0.7	39.2%	63.8%
1000	100	1.5	0.3	950	92.7%	80.5%
	500	1.4	0.3	32	63.6%	40.2%
	1000	1.4	0.3	7.6	62.7%	42.3%
	2000	1.4	0.3	1.8	69.2%	48.6%
	3000	1.3	0.2	0.8	54.1%	51.7%
	4000	1.3	0.2	0.4	59.7%	52.9%
5000	100	1.04	0.03	420	84.8%	40.2%
	500	1.03	0.02	11	69.8%	32.6%
	1000	1.03	0.02	2.7	65.3%	35.2%
	2000	1.03	0.02	0.7	49.2%	39.6%
	3000	1.03	0.02	0.3	36.4%	40.8%
	4000	1.03	0.02	0.2	34.6%	41.2%

Conclusion

- We show that ML-based tracking can learn to reconstruct non-helical tracks with high efficiency when training on non-helical tracks and mixed tracks. That will allow for powerful new quirk searches and open the door to other weird-track searches

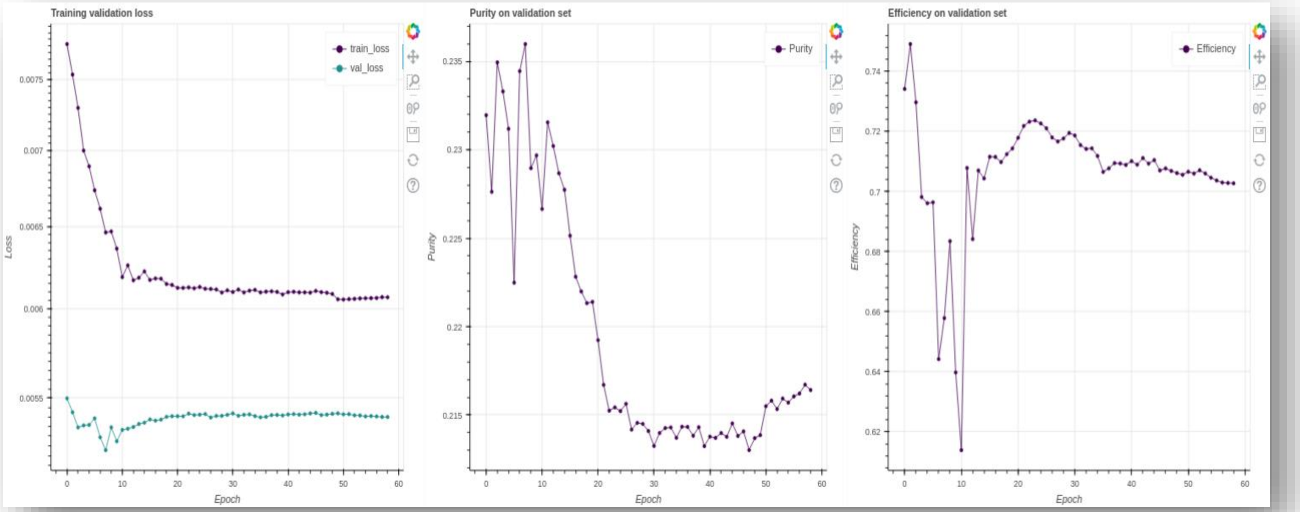
- Could use non-helical tracks as a tool to understand GNN reconstruction on helical tracks, or hard-to-reconstruct SM particles

- Hope this tool could help us find the BSM particles with non-helical tracks.

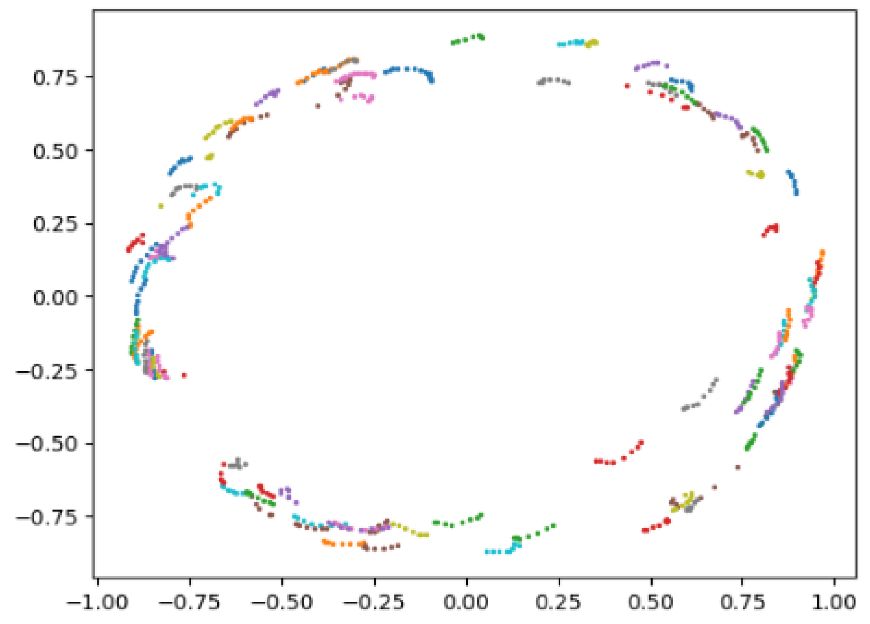
Backup

Metric Learning: Background training, quirk inference

Use metric learning to reduce the dimension: Embedding the space points on to graphs.

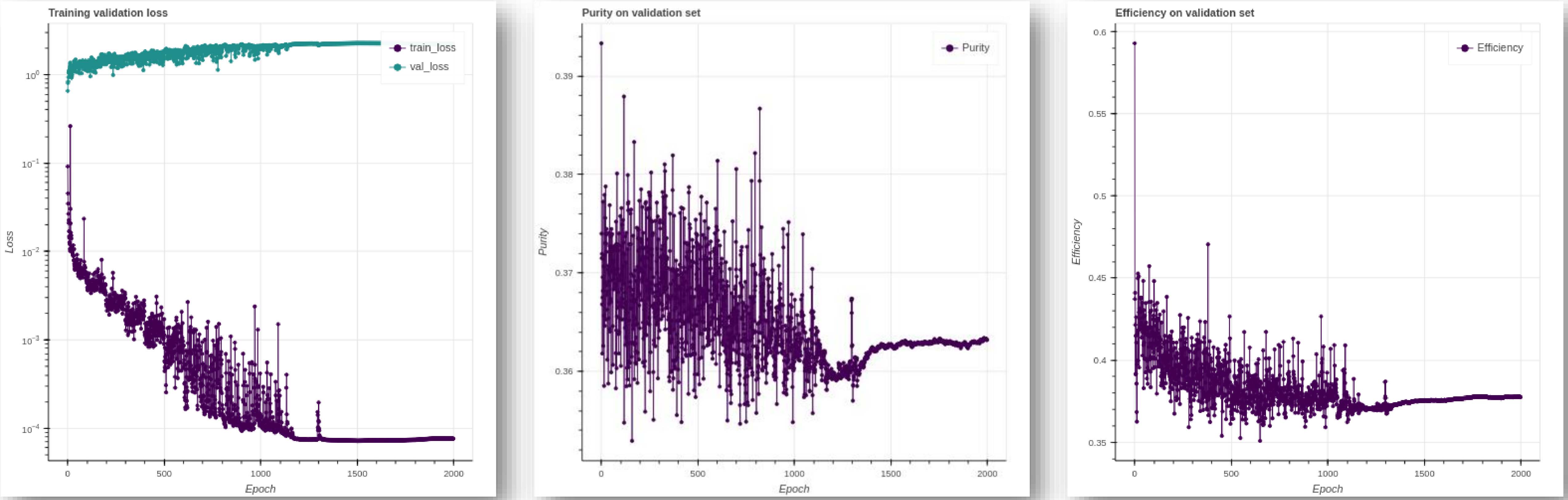


After Embedding:



GNN: Background training, quirk inference

Train GNN to classify edges as either “true” (belonging to the same track) or “false” (not belonging to the same track)



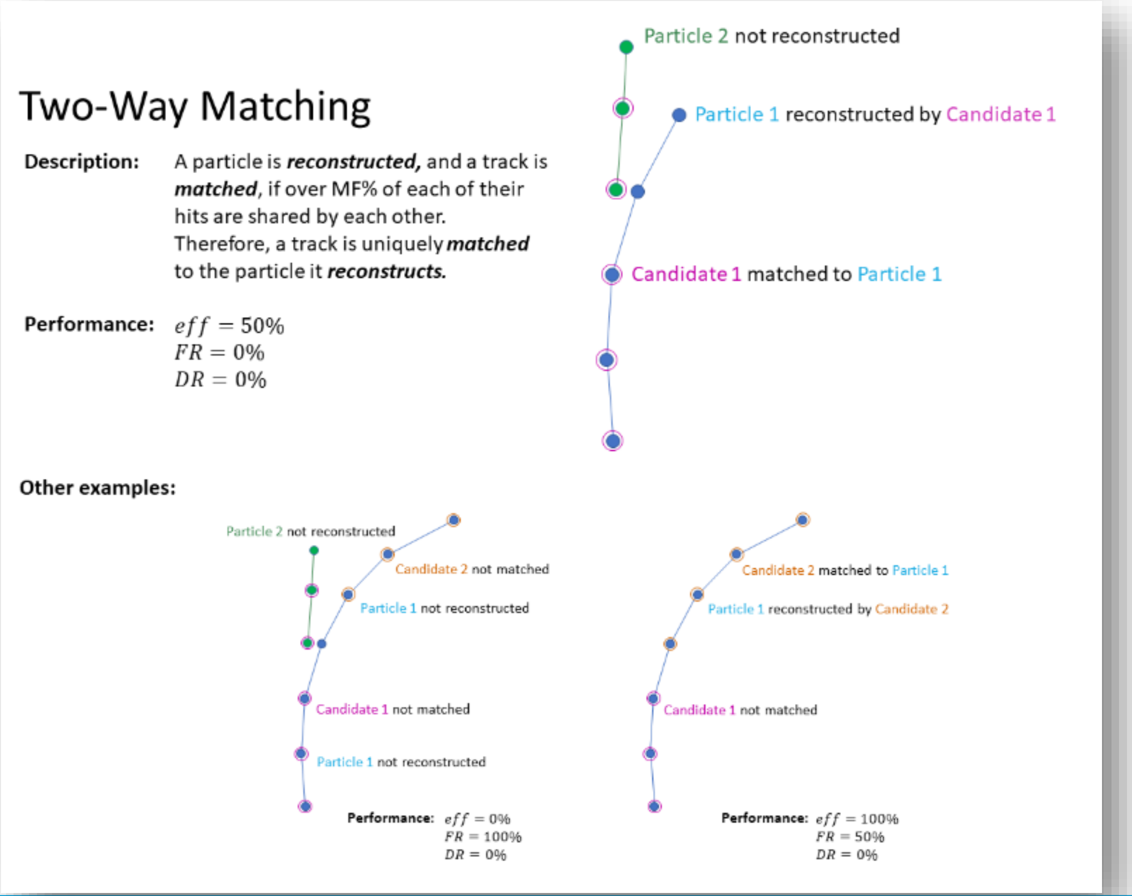
Performance – Tracking definitions

Physics cuts: $\{|\eta| < 4\}$

Some selection for reconstructed particles: For bkg, we have 8 true hits for each particles, for quirk, we have ≥ 8 true hits.

- min_reco_length: 5 (Reconstructable)
- min_truth_length: 7

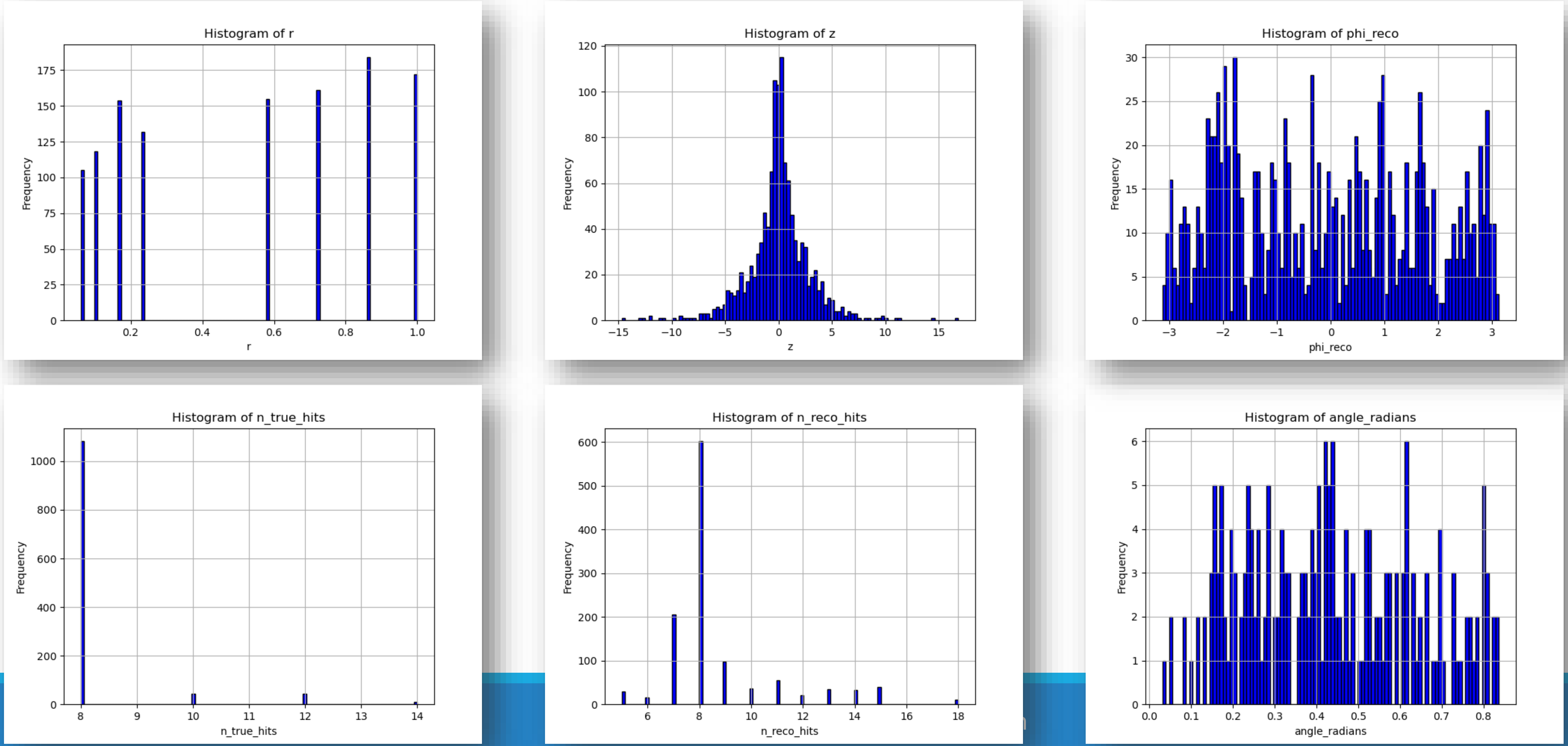
- Matching style: Two_way



Distribution of reconstructed quirks

The distribution of reconstructed quirks' information:

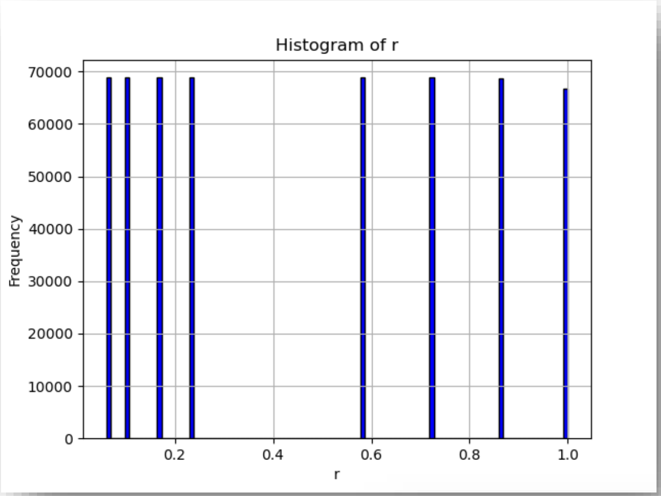
- $r, \phi, z(cm)$ are truth information of hits. r is scaled to (0,1).
- n_reco_hits is the number of reconstructed hits, n_true_hits is the number of truth hits.

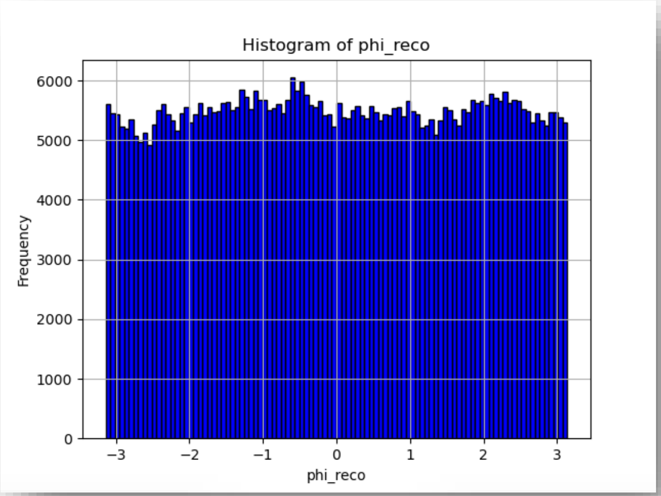


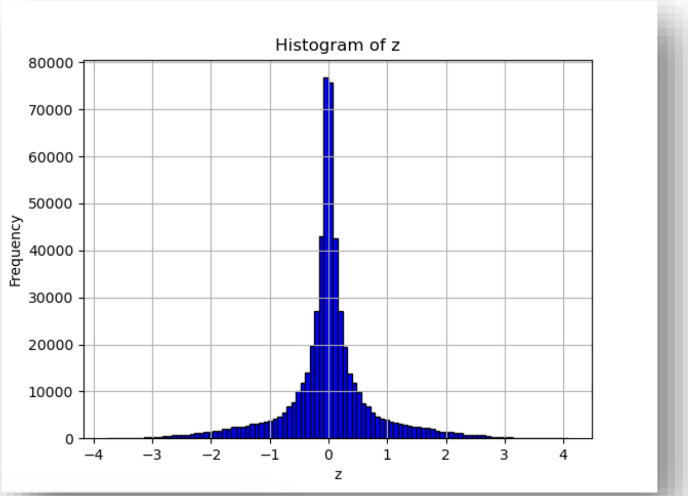
Distribution of reconstructed background

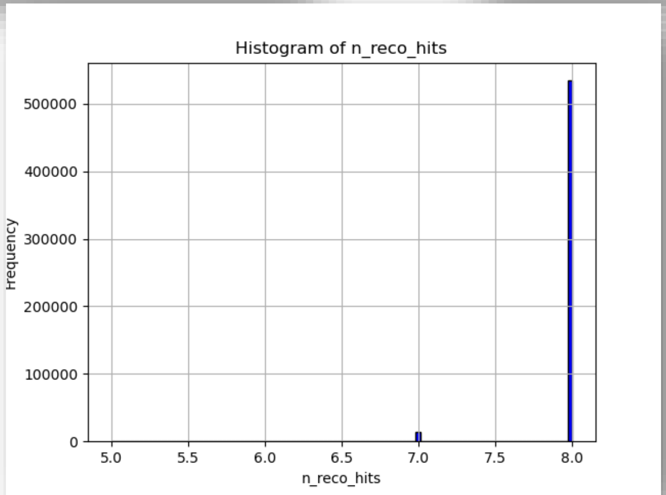
The distribution of reconstructed bkg(SM)s’ information:

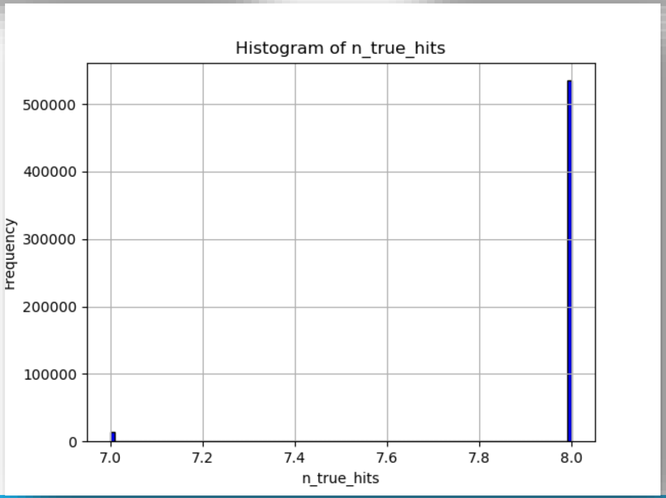
- The reconstructed information is similar as the truth information (n_hits)







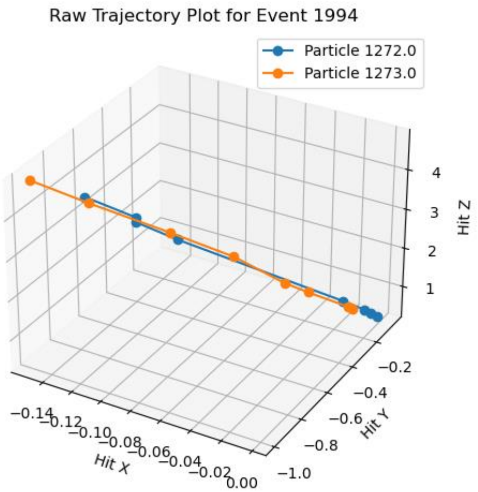


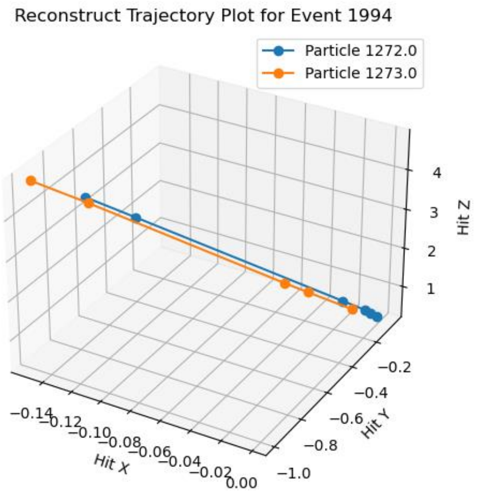


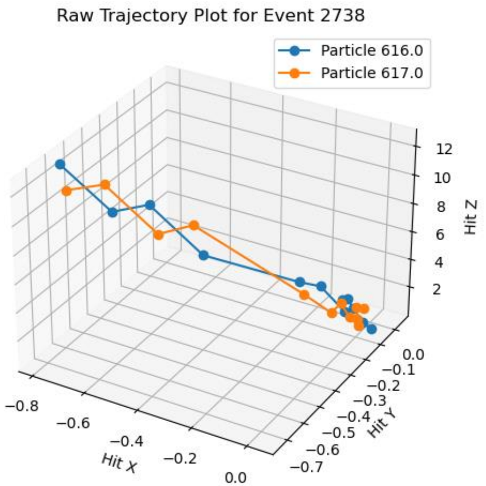
Reconstructed hits of quirk

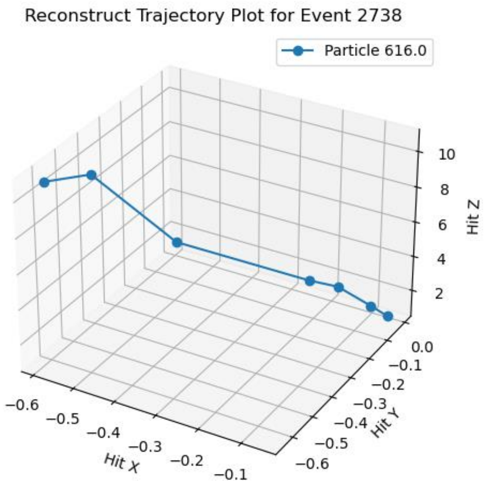
With same event (use the reconstructed event information):

- Some $\mathrm{hits}_{\mathrm{reco}}$ are the part of truth quirk track.
- Only reconstruct simple and smooth track.



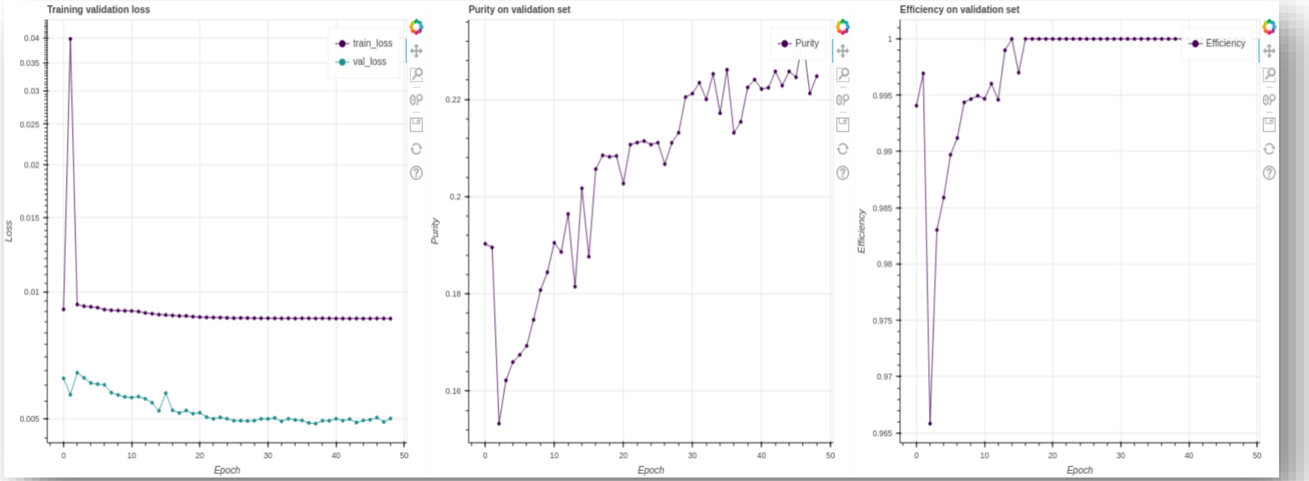




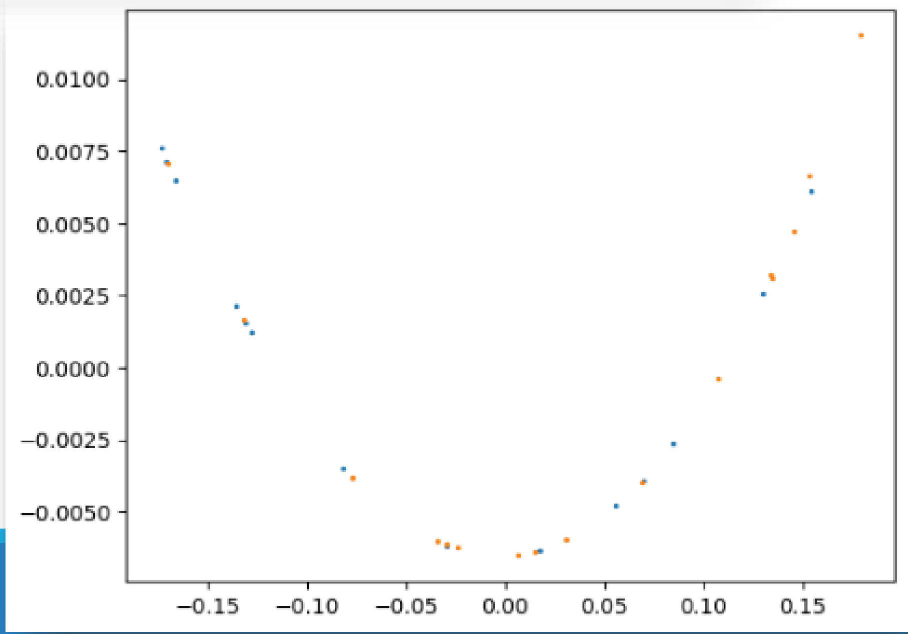


Metric Learning : Quirk training, quirk inference

Use metric learning to reduce the dimension: Embedding the space points on to graphs.

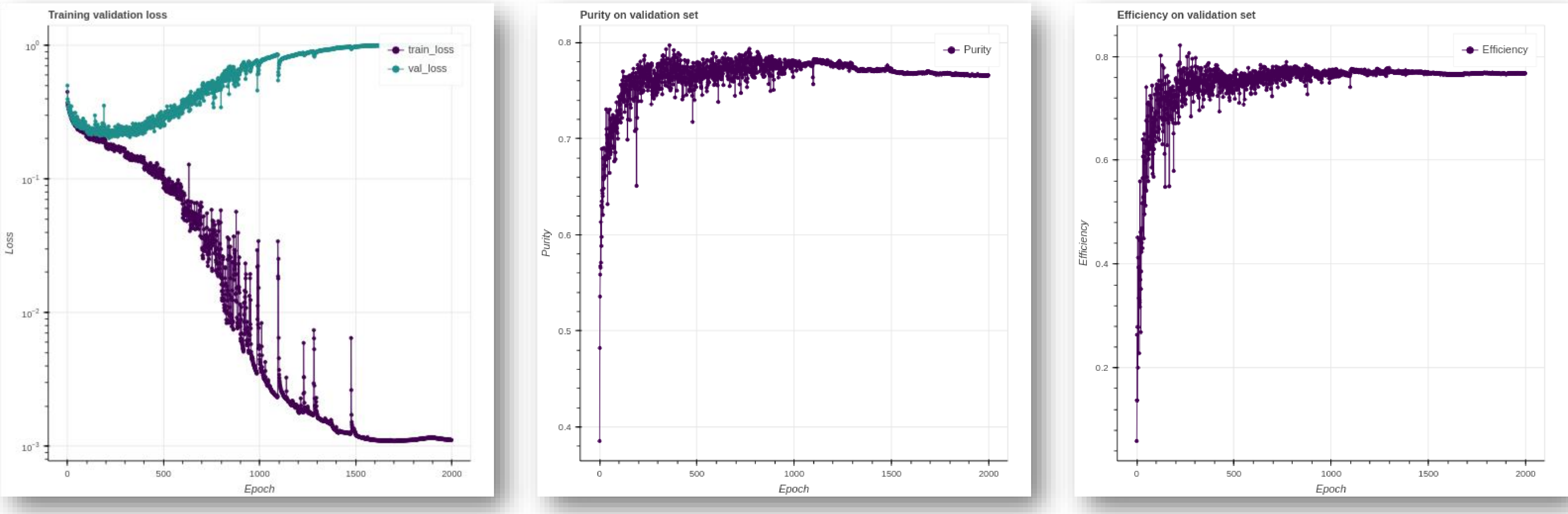


After Embedding:



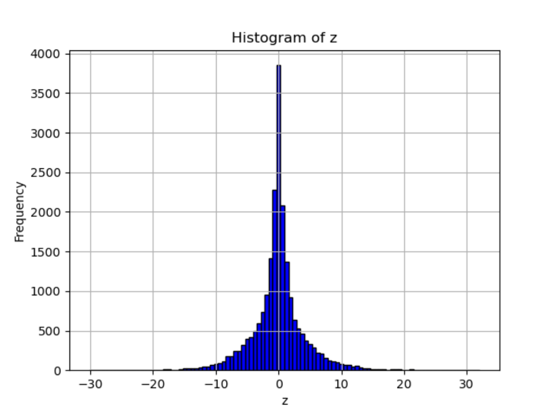
GNN : Quirk training, quirk inference

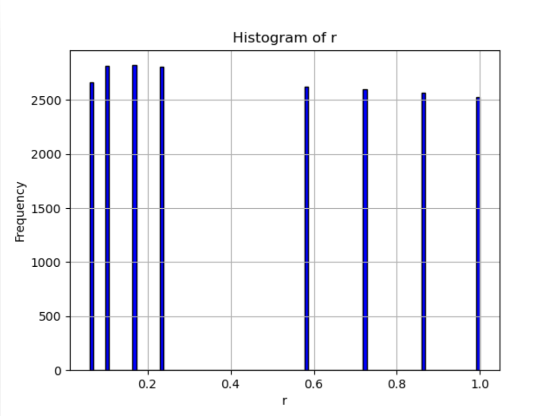
Train GNN to classify edges as either “true” (belonging to the same track) or “false” (not belonging to the same track)

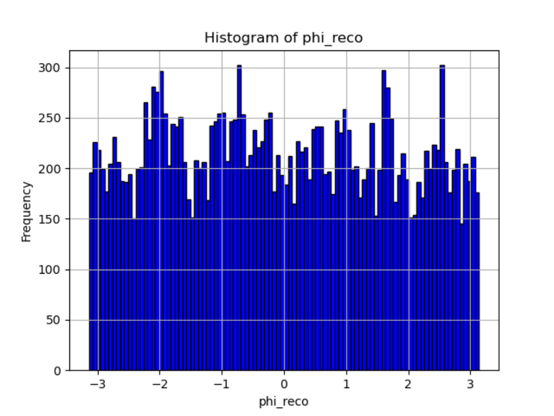


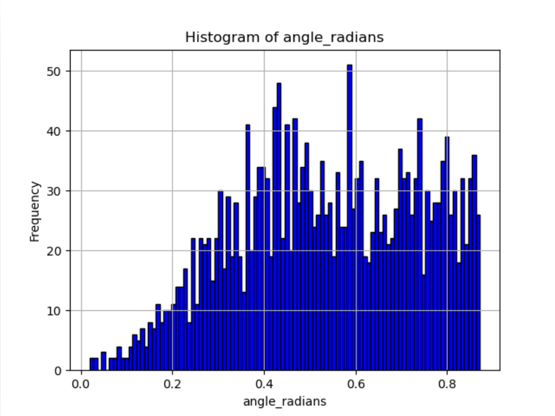
Results: Quirk training, quirk inference

Distribution of reconstructed quirks:



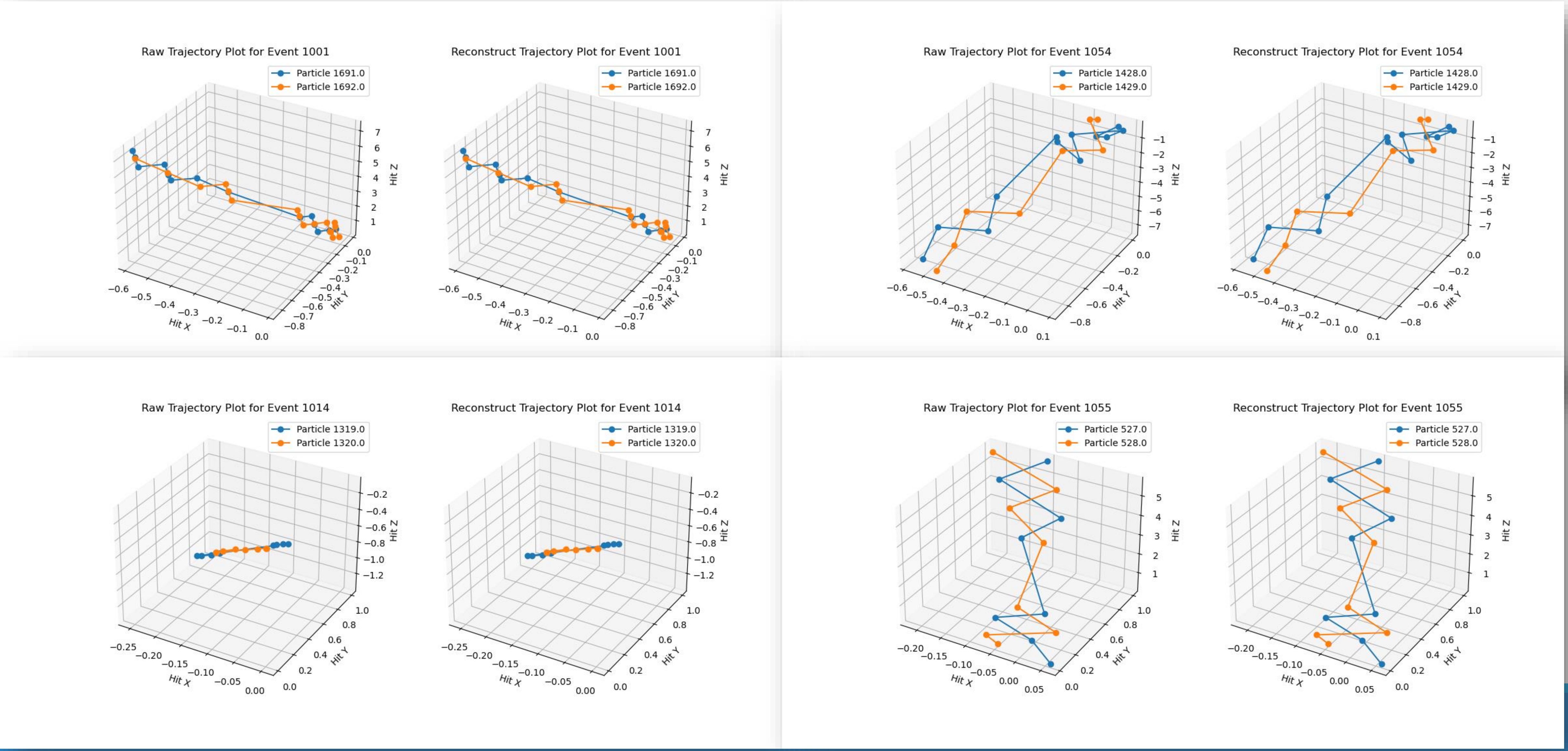






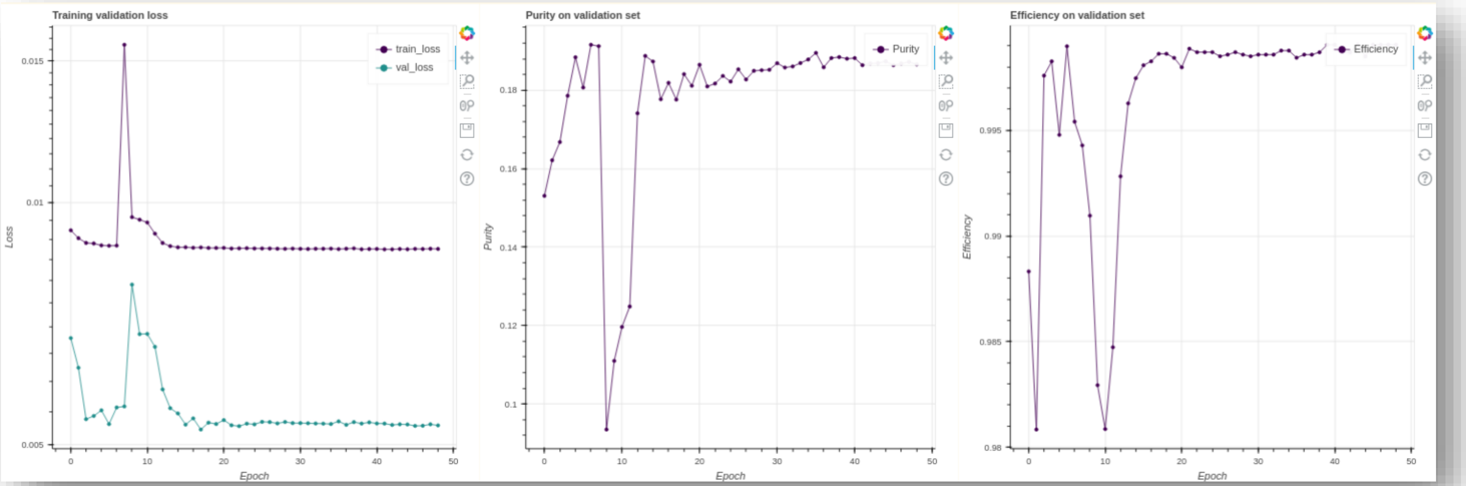
Reconstructed hits of quirk

All of well-behaved quirks are reconstructed well:

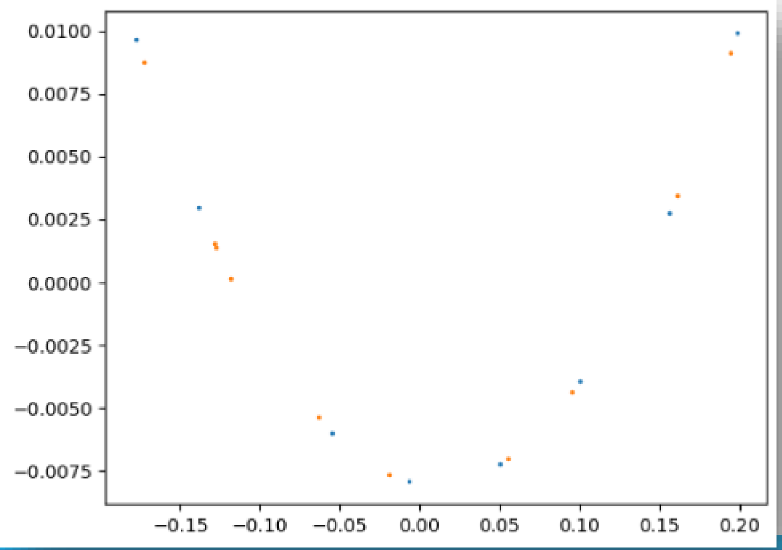


Metric Learning : All Quirk training, quirk inference

Use metric learning to reduce the dimension: Embedding the space points on to graphs.



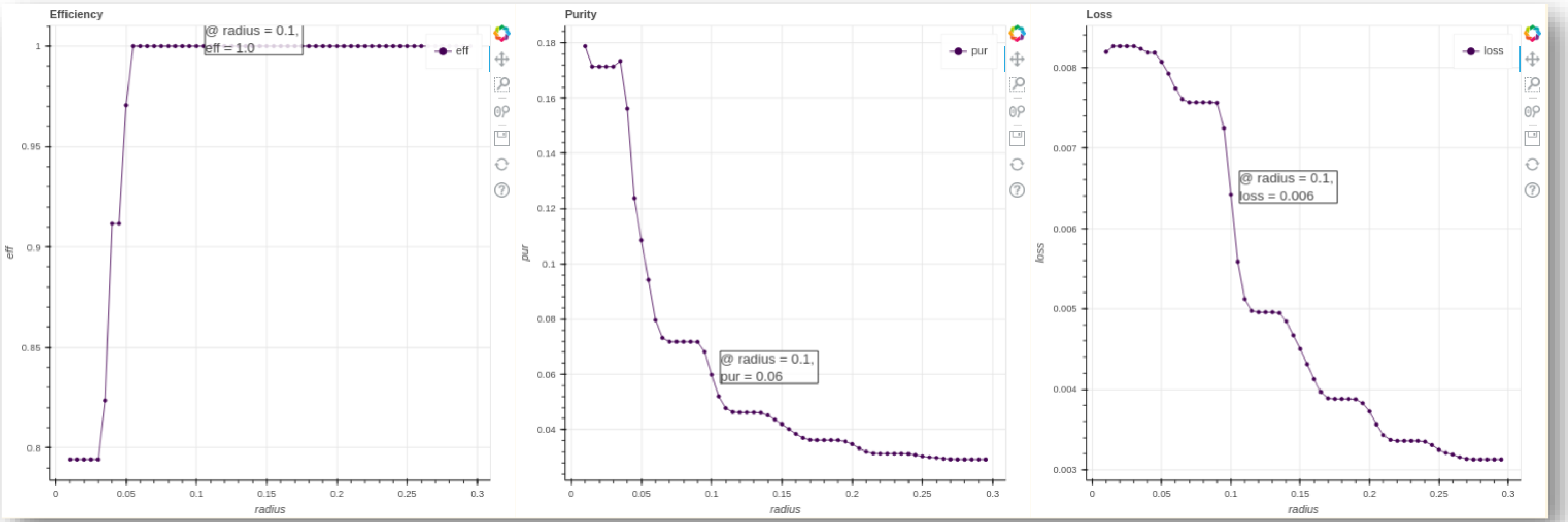
After Embedding:



Qiyu Sha qsha@cern.ch

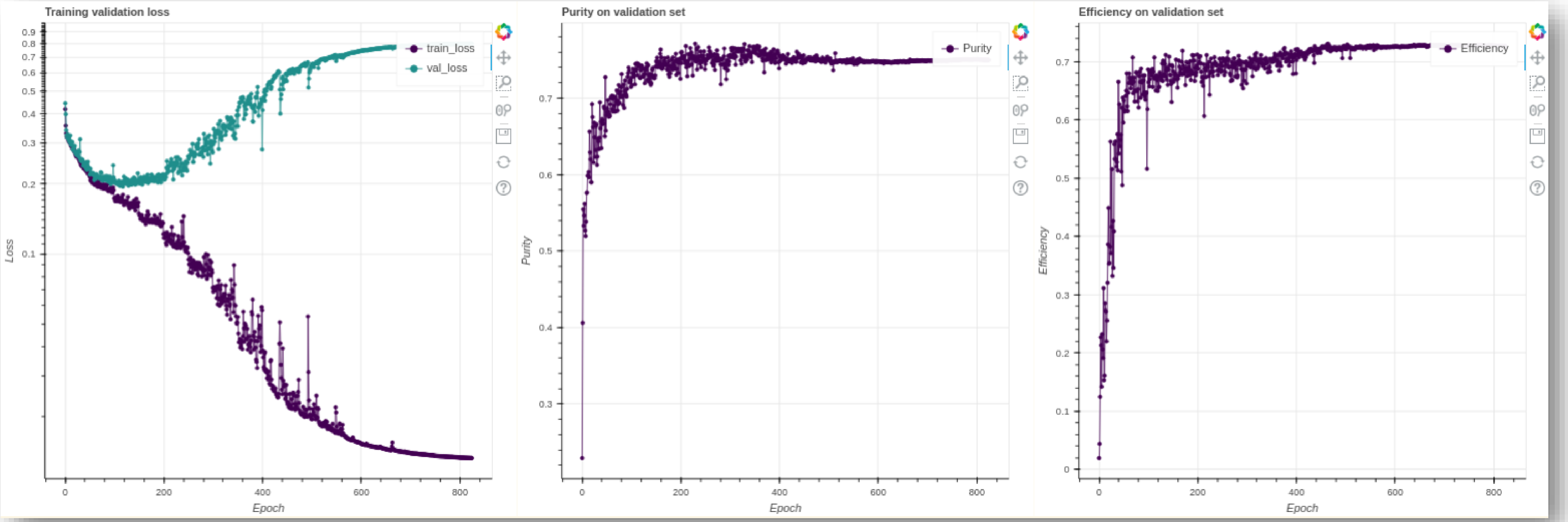
Metric Learning : All Quirk training, quirk inference

Evaluate the model performance on one test data sample to see how the efficiency and purity change with the embedding radius.



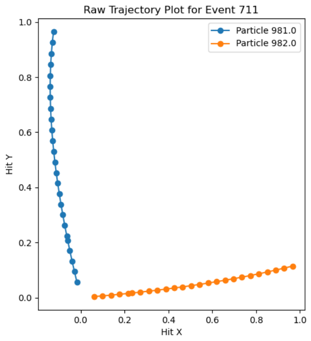
GNN : All Quirk training, quirk inference

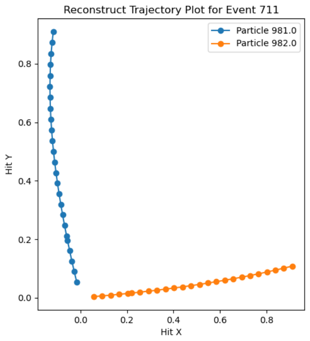
Train GNN to classify edges as either “true” (belonging to the same track) or “false” (not belonging to the same track)

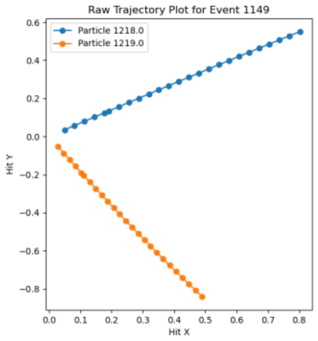


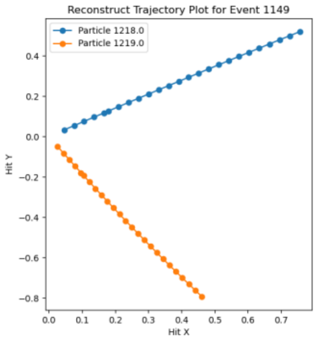
Reconstruction performance

Mass_100_Lambda_100

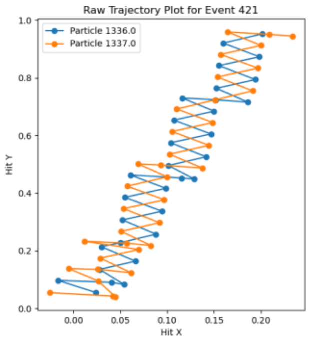


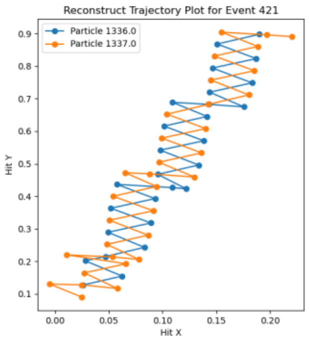


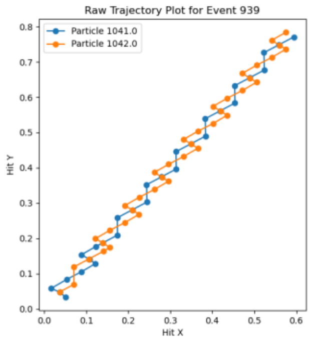


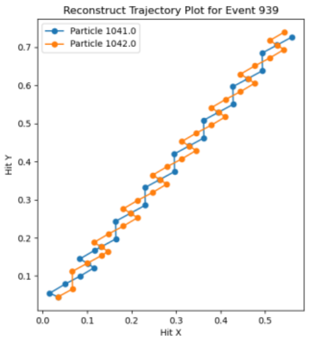


Mass_100_Lambda_500



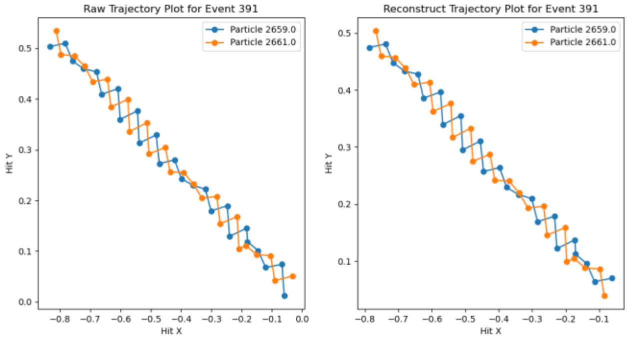


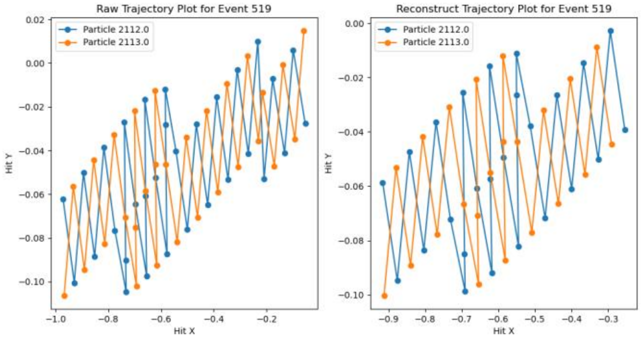




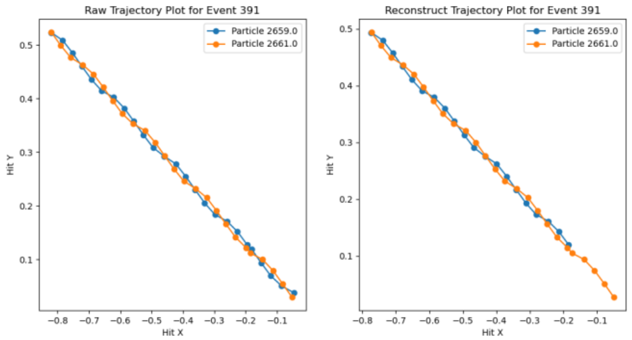
Reconstruction performance

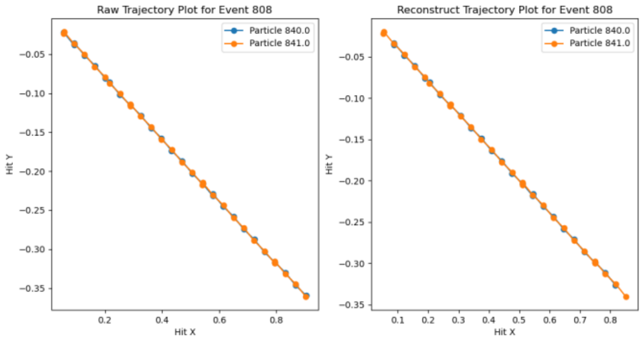
Mass_100_Lambda_1000



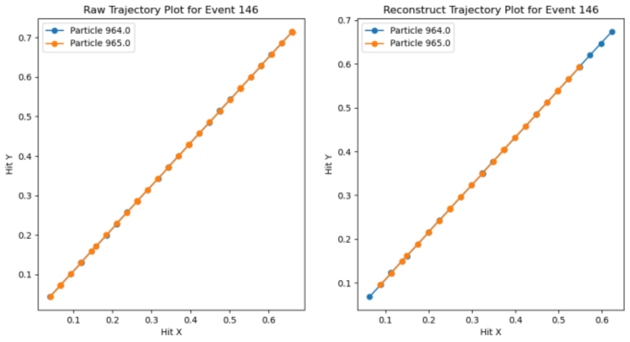


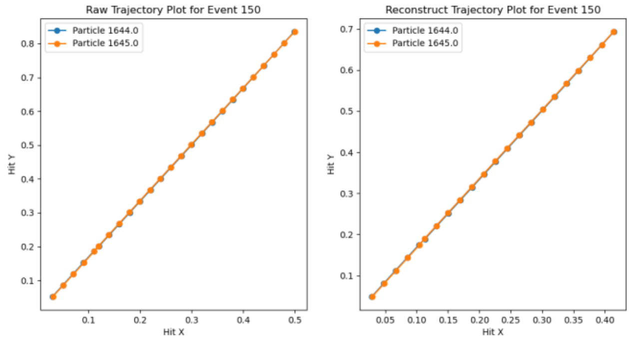
Mass_100_Lambda_2000





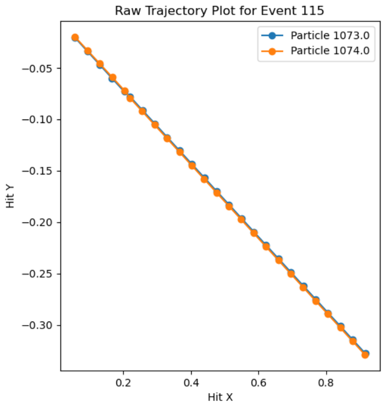
Mass_100_Lambda_3000

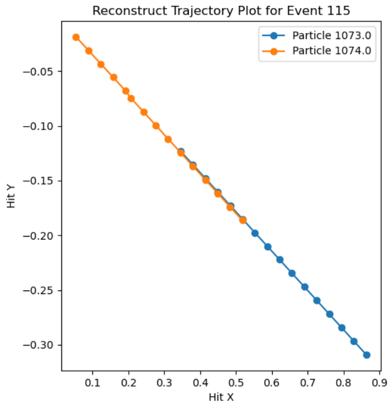


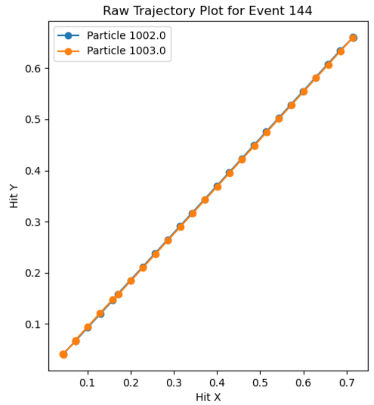


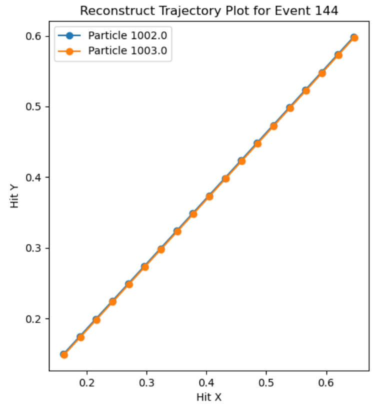
Reconstruction performance

Mass_100_Lambda_4000

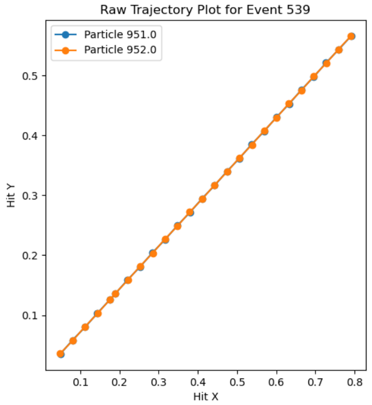


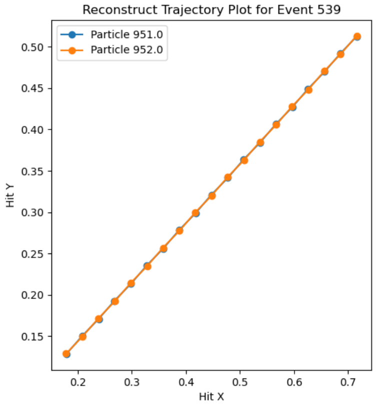


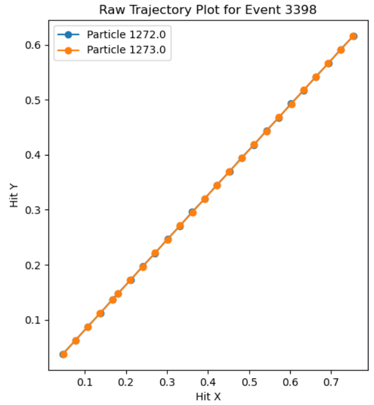


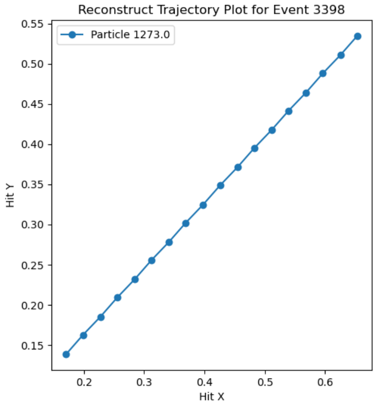


Mass_100_Lambda_5000



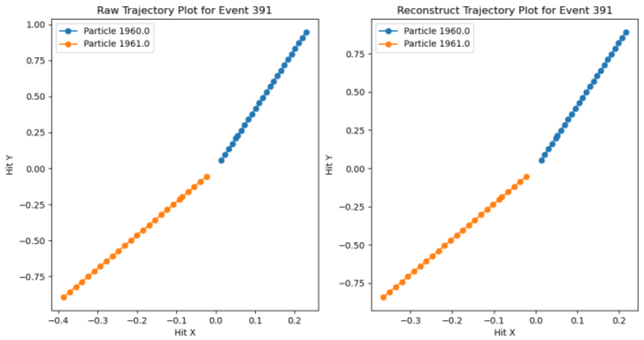


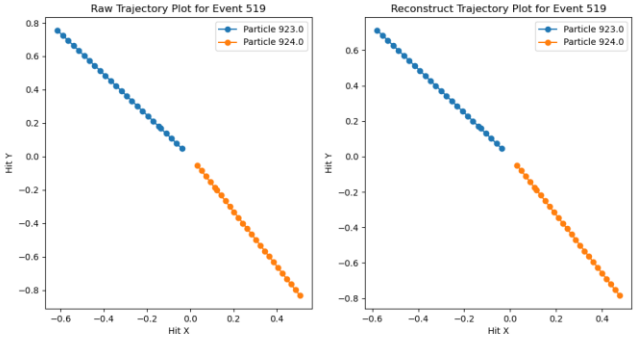




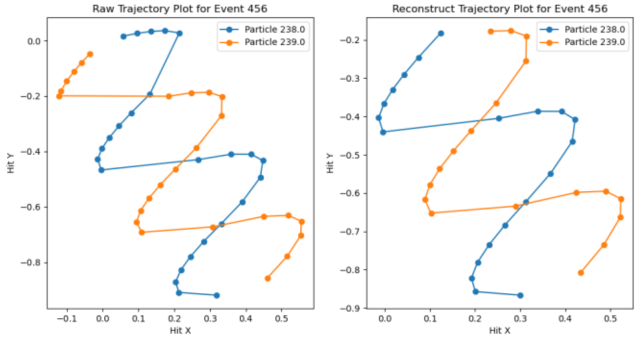
Reconstruction performance

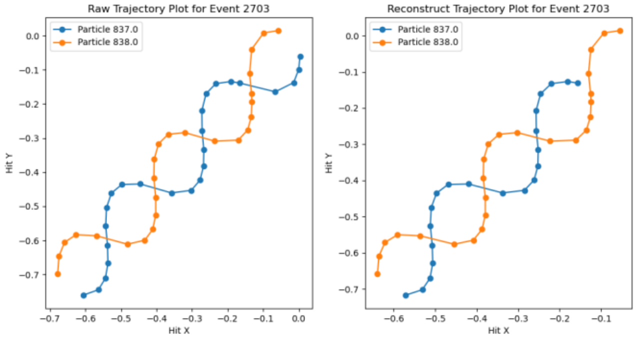
Mass_1000_Lambda_100



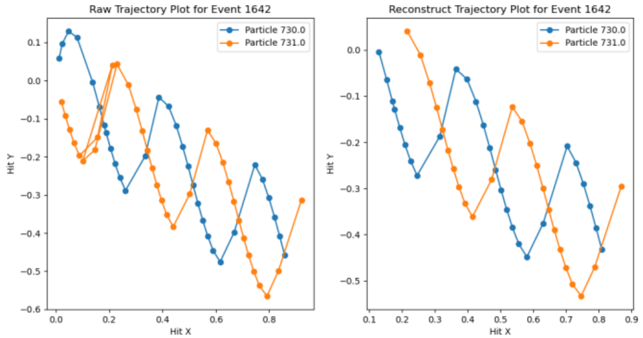


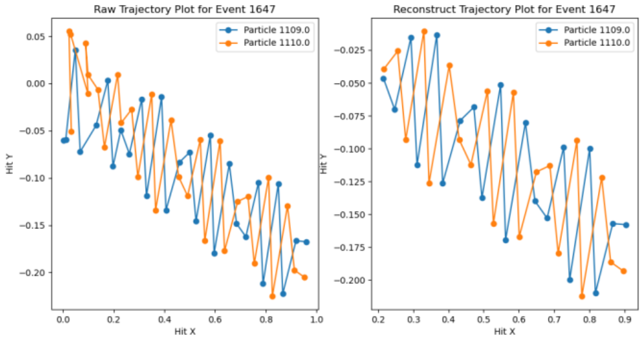
Mass_1000_Lambda_500





Mass_1000_Lambda_1000

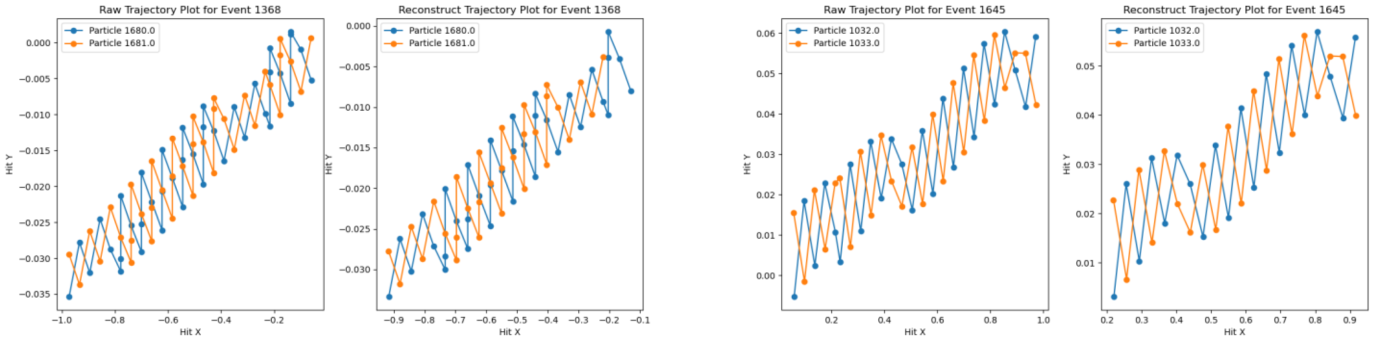




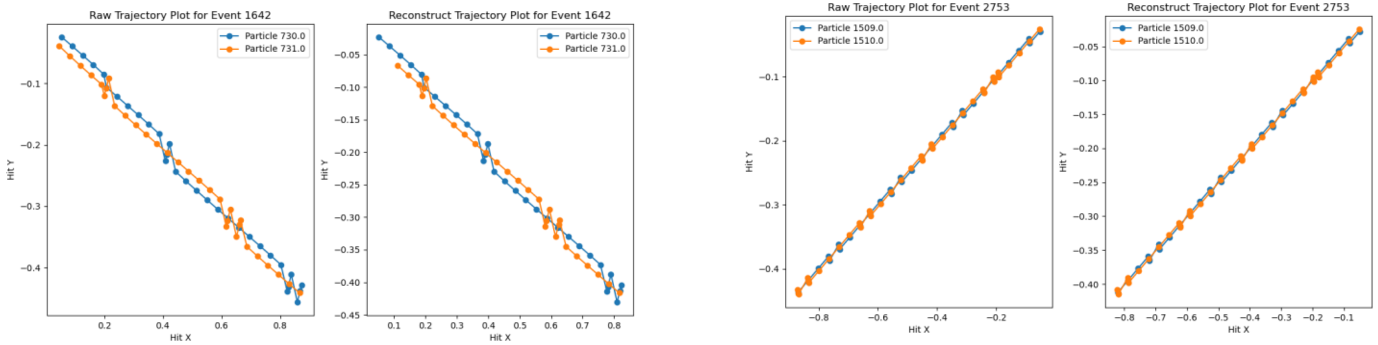
Reconstruction performance

Mass_1000_Lambda_2000

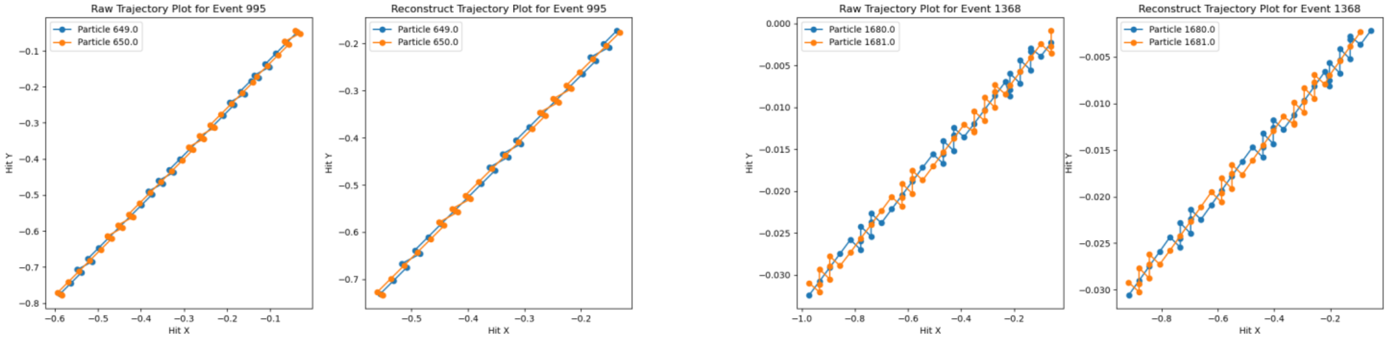
Not completely overlapped, like m100, that is why we can still get a good eff.



Mass_1000_Lambda_3000



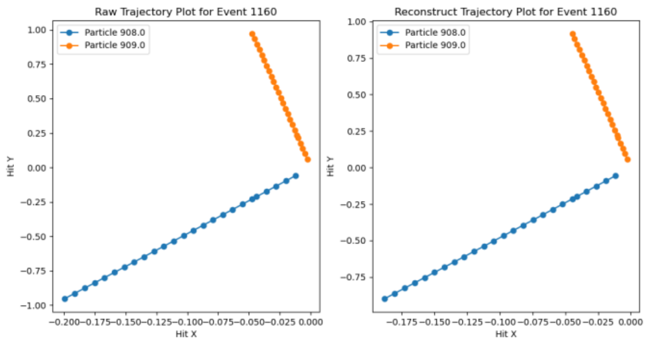
Mass_1000_Lambda_4000



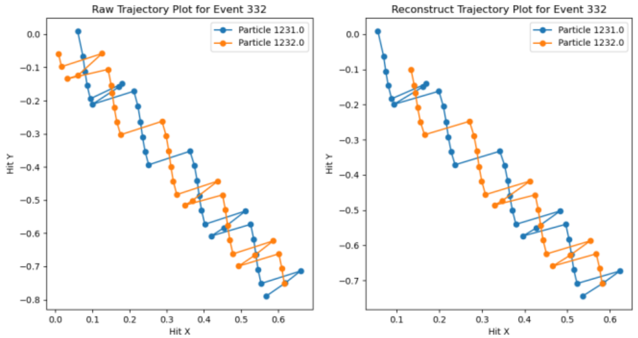
m_Q (GeV)	Λ(eV)	$\bar{\gamma}$	$\sigma(\gamma)$	d [cm]	Efficiency
100	100	3.7	3.4	540	91.0%
	500			21.6	82.8%
	1000			5.4	77.8%
	2000			1.4	49.4%
	3000			0.6	12.8%
	4000			0.34	4.0%
	5000			0.22	1.0%
500	100	1.8	0.6	800	92.0%
	500			32	79.0%
	1000			8	64.3%
	2000			2	53.4%
	3000			0.9	54.4%
	4000			0.5	25.1%
	5000			0.3	16.6%
1000	100	1.4	0.3	800	92.7%
	500			32	56.4%
	1000			8	55.1%
	2000			2	65.4%
	3000			0.9	41.0%
	4000			0.5	51.4%
5000	100	1.03	0.003	300	84.8%
	500			12	69.8%
	1000			3	65.3%
	2000			0.8	38.7%
	3000			0.3	14.6%
	4000			0.1	3.1%

Reconstruction performance

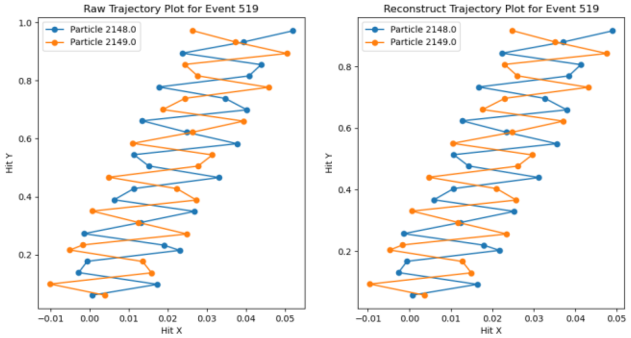
Mass_500_Lambda_100



Mass_500_Lambda_500

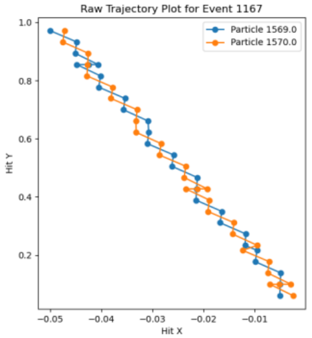


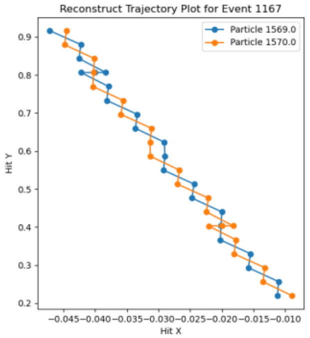
Mass_500_Lambda_1000

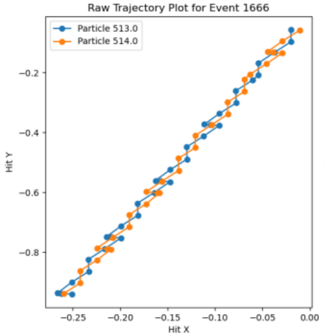


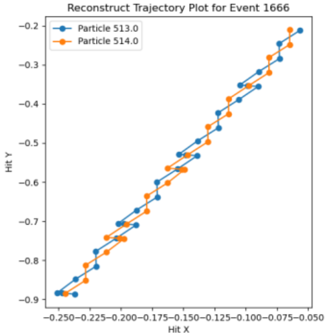
Reconstruction performance

Mass_500_Lambda_2000

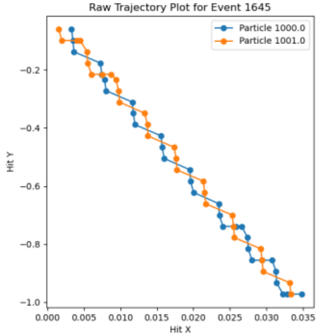


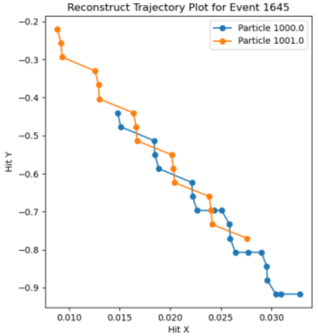


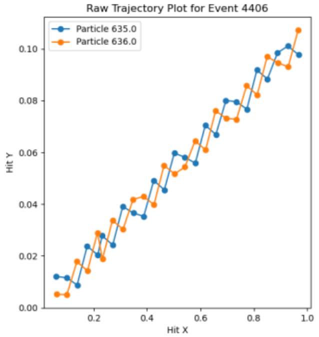


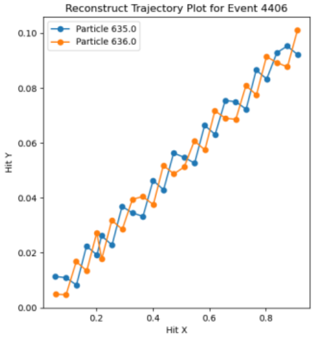


Mass_500_Lambda_3000



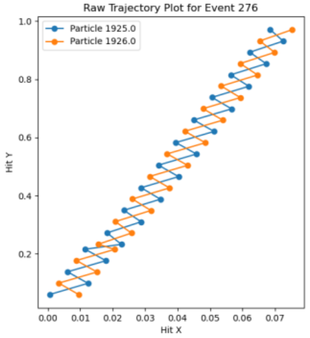


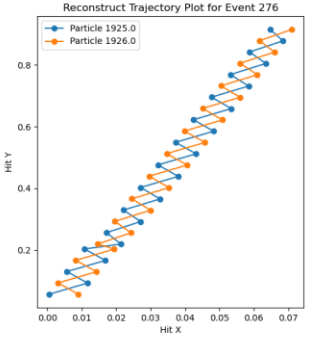


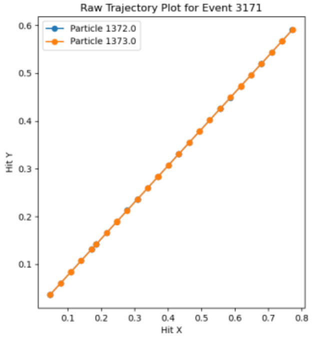


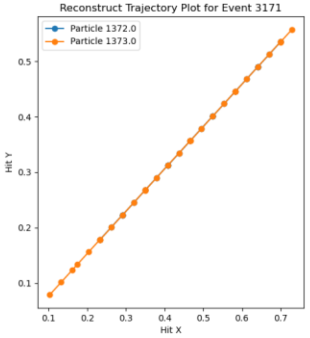
Mass_500_Lambda_4000

More and more overlaped



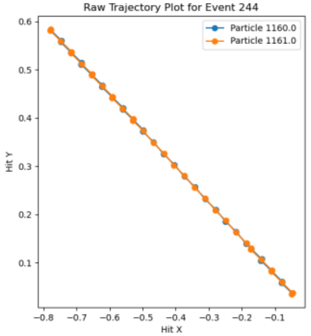


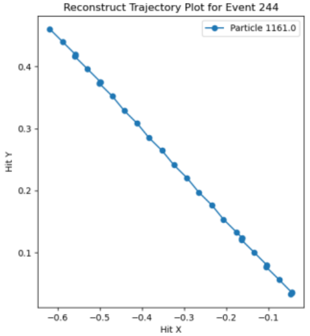


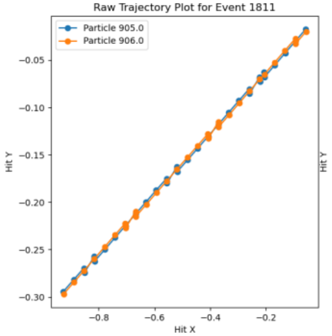


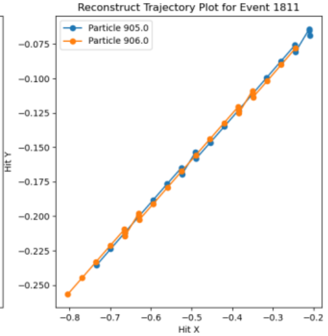
Reconstruction performance

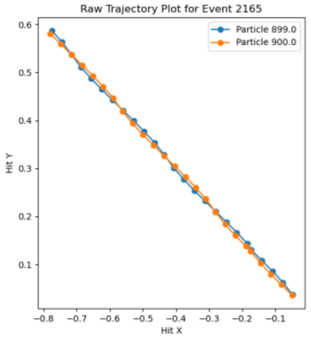
Mass_500_Lambda_5000 (Most of them can only be reconstructed as one track)

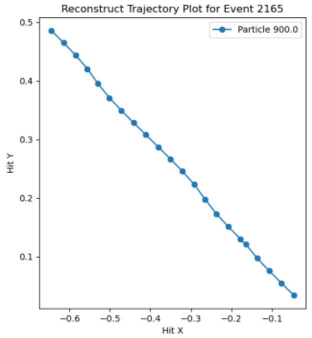


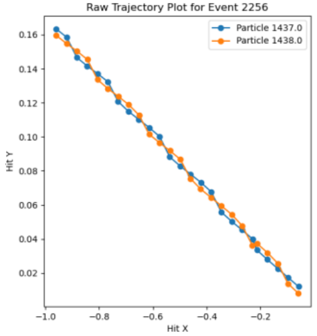


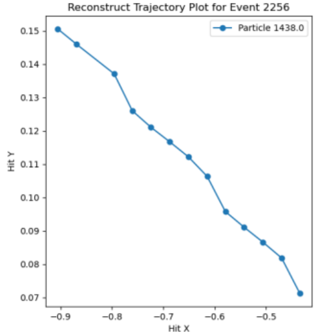












m_Q (GeV)	Λ(eV)	$\bar{\gamma}$	$\sigma(\gamma)$	d [cm]	Efficiency
100	100	3.7	3.4	540	91.0%
	500			21.6	82.8%
	1000			5.4	77.8%
	2000			1.4	49.4%
	3000			0.6	12.8%
	4000			0.34	4.0%
	5000			0.22	1.0%
500	100	1.8	0.6	800	92.0%
	500			32	79.0%
	1000			8	64.3%
	2000			2	53.4%
	3000			0.9	54.4%
	4000			0.5	25.1%
	5000			0.3	16.6%
1000	100	1.4	0.3	800	92.7%
	500			32	56.4%
	1000			8	55.1%
	2000			2	65.4%
	3000			0.9	41.0%
	4000			0.5	51.4%
5000	100	1.03	0.003	300	84.8%
	500			12	69.8%
	1000			3	65.3%
	2000			0.8	38.7%
	3000			0.3	14.6%
	4000			0.1	3.1%