Weekly report

Qiyu Sha

Institute of High Energy Physics CAS, Beijing

Zprime to mumu

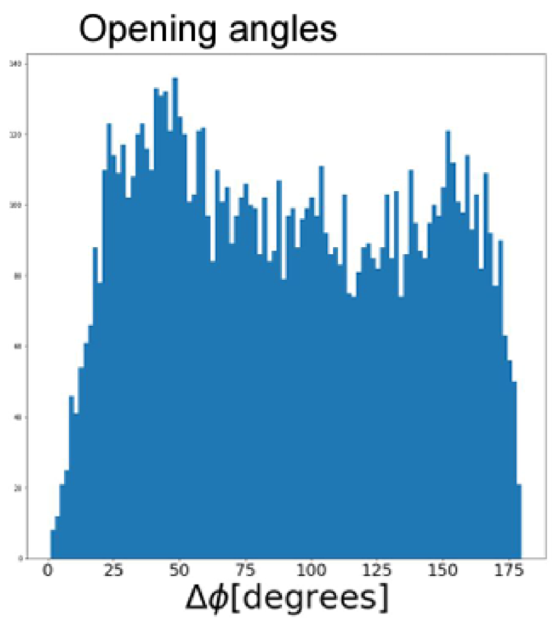
- Have requested EB

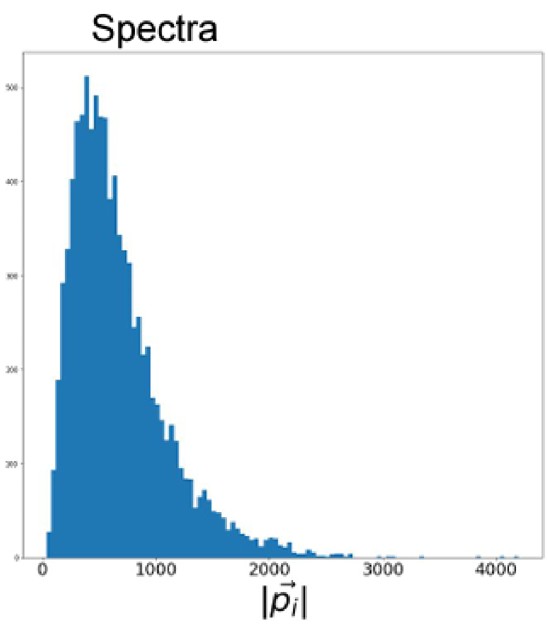
Di-higgs yyML

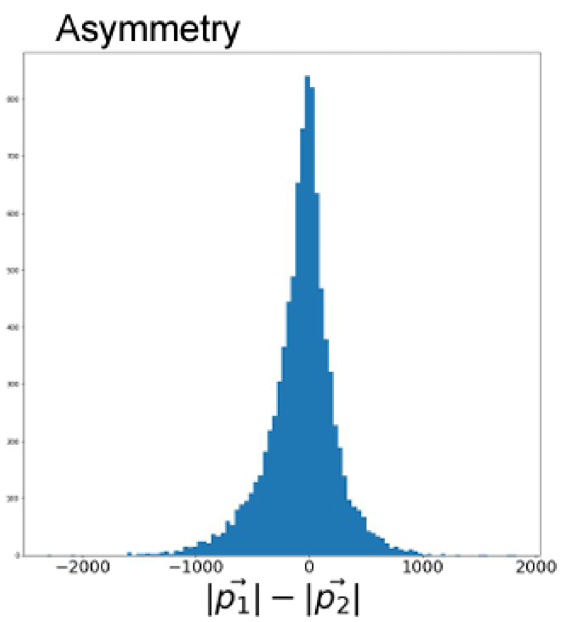
- DAOD/MxAOD sample done.
- Will finish the klamda scan next week.

Quirk track ML

- Do some check on the dataset:







<u>Takeaways</u>:

Uniform opening angle

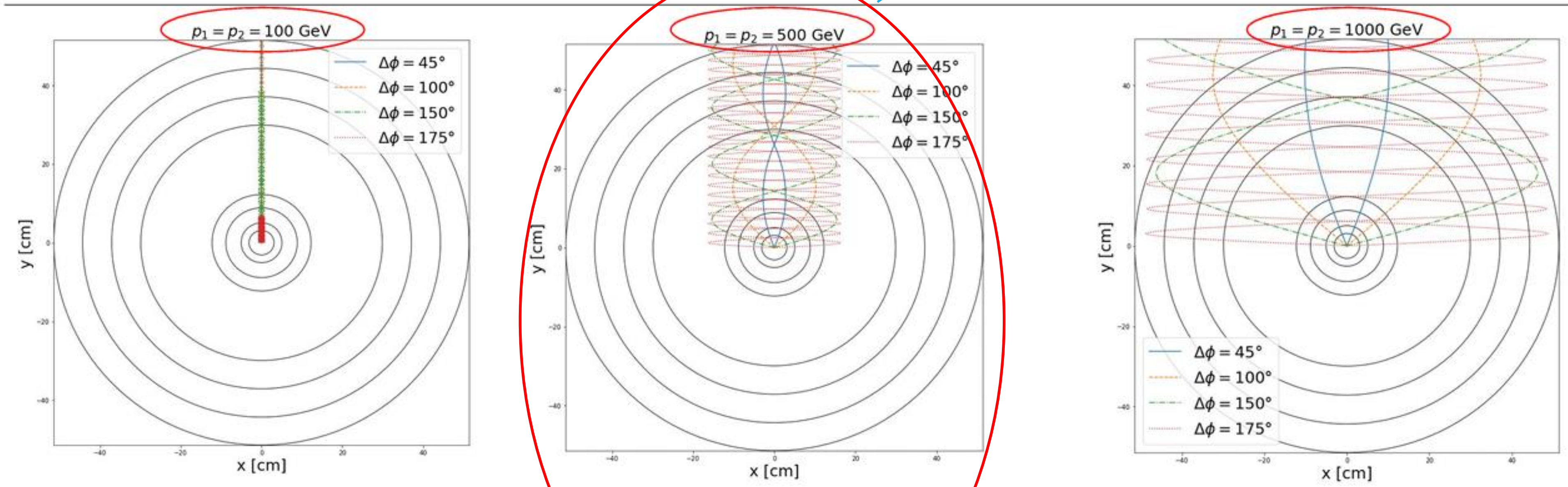
~500 GeV momenta

Some asymmetry, but typical case is p1 = p2. So I'll idealize and resimulate

Quirk track ML

Quirks come out at a variety of momenta with different opening angles. Let's set p1 = p2, and look at trajectories for different opening angles centered on y axis

Use now.



<u>100 GeV</u> quirks→
too soft too separate →
8 hits per quirk

<u>500 GeV</u> quirks→
Start to separate →
Can rack up many hits

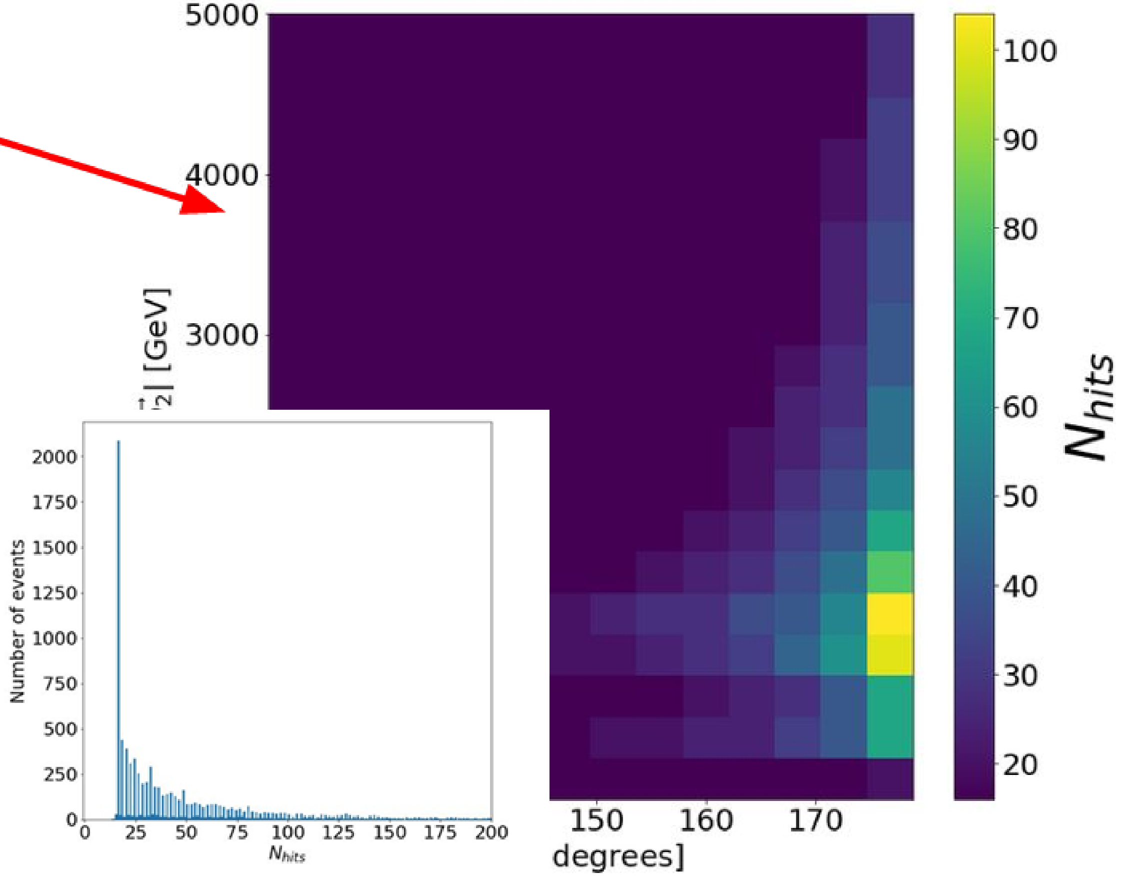
<u>1000 GeV</u> quirks→
For phi<=100, can still have only 8 hits, otherwise very many

Quirk track ML

Let’s put it all together and map the number of hits as a function of opening angle and momenta

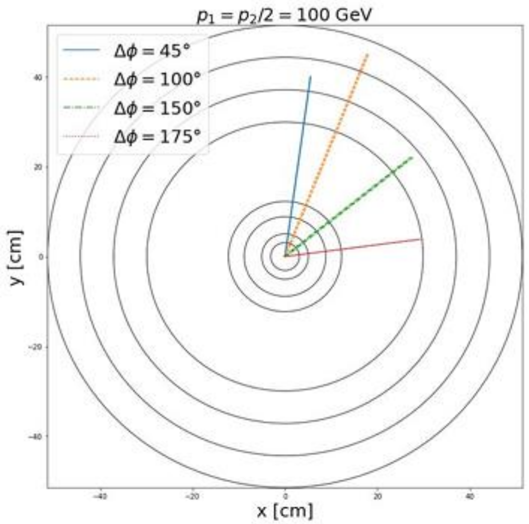
- We rack up many hits for only large opening angles, but typically get ~20+
    - This is not our actual dataset but is representative

- Looking at the actual data, median of 26 hits (13 for each quirk)

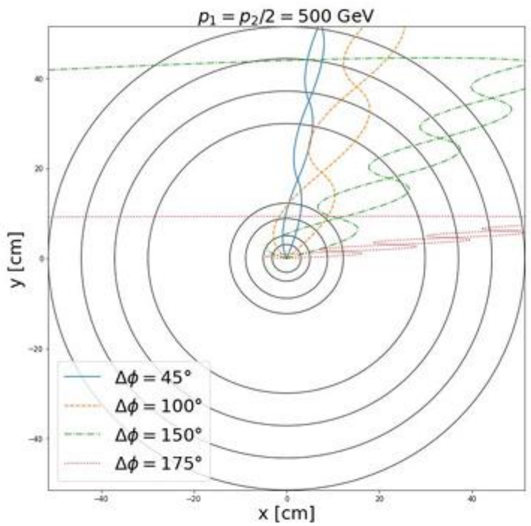


Quirk track ML

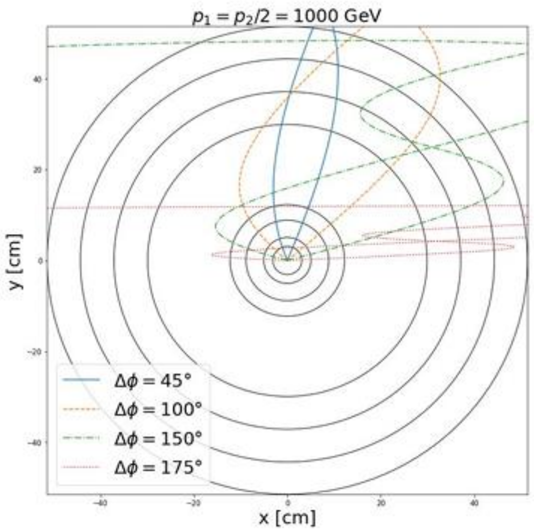
Now let's study some asymmetry and set p1 = p2 /2.
A coupled pendulum…Crazy tracks…



100,200 GeV quirks→
Same story. Q2 drags Q1
but they don't separate →
8hits



500,1000 GeV quirks→
Start to separate →
Only at large opening angles
do we see crazy tracks



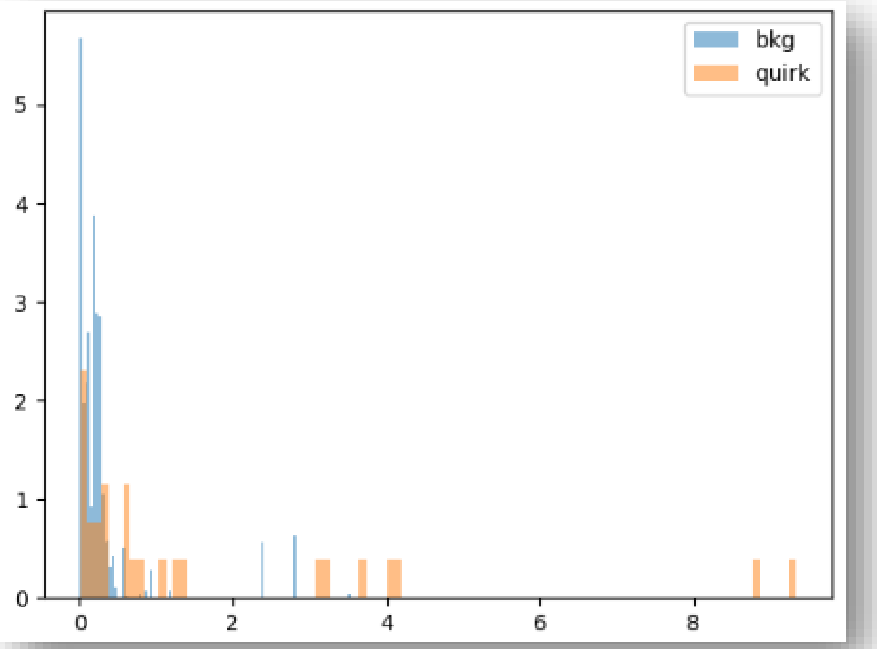
1000,2000 GeV quirks→
You get the idea…

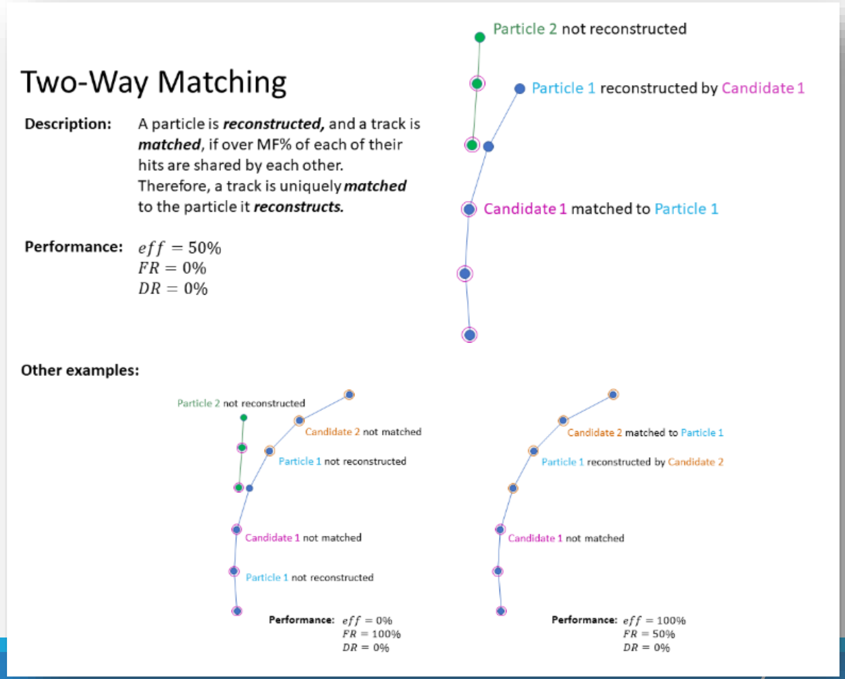
Quirk track ML

- Calculate the Chi-square value of track fit on the reco hits.
  - Choose the chi-square value of (r,z) fit:
    - Quirk tracks have bad fit performance as expected.

- Choose a new matching style with tighter filters to cut the wrong reconstructed Quirk tracks.

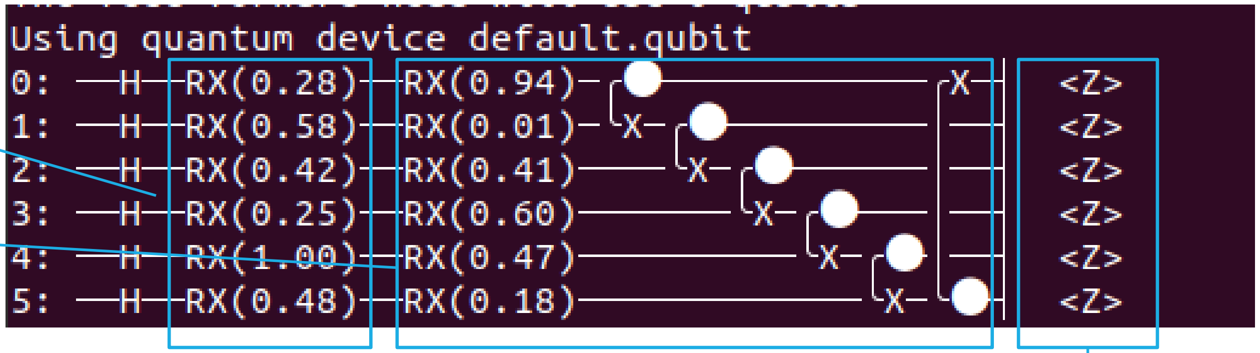
- Redo the training with more Quirk events.
  - Draw plots and do more analysis this week.





QPT

- Quantum Transformer works now. (Use Pennylane)
  - Use the same dataset as QSVM analysis. (just for test, 25k bkg and 36k Sig dataset)
  - Simulator: Pennlylane default device.
  - Circuit:
    - AngleEmbedding for embedding the dataset
    - BasicEntanglerLayers for entangle.
    - Expval(Pauliz) for measurement.
      - This is different from the “counts” in QSVM (Results from random tests --- Such as get the probabilities of “00000” in 10000 tests)
      - Measure the expected value of the supplied observable: $\langle \varphi | Pauliz | \varphi \rangle$
      - We don’t care the measurement method, just a mapping, we can treat them as an arbitrary function. Use “Expval” save the time.



  - We use the QML to do the self-attention.
  - Time consuming: O(n), ~80 mins for 10k dataset with one epoch and one block(Q_layer).

QPT

- More details for transformer part.
  - After the self-attention, we need to do the training in the classical transformer. (Will add the particle transformer part if we change the dataset.)
  - Decided the batch-size first which depends on the GPU memory in IHEP (Need to check later)
  - Optimize the hyperparameter. (To do list)
    - Learning rate
    - Embed_dim in classical transformer. (Fixed to qubits number in Quantum part)
    - ffn_dim (Hidden layer dimension of feedforward networks)
    - Blocks(Q_layer): Affect the time consuming in the Quantum part.
    - Dropout_rate: Regularization method to prevent the overfitting, for small dataset as 10k, we can use 0,1 or 0.
    - Epoch: Changeable in the classical part and fixed to 5 or 10 in the Quantum part to save the time.
  - The tool for optimization: GridSearch(easy but waste time), Quasi-random(Google experts suggest), Hugging face (Visualization is done well).

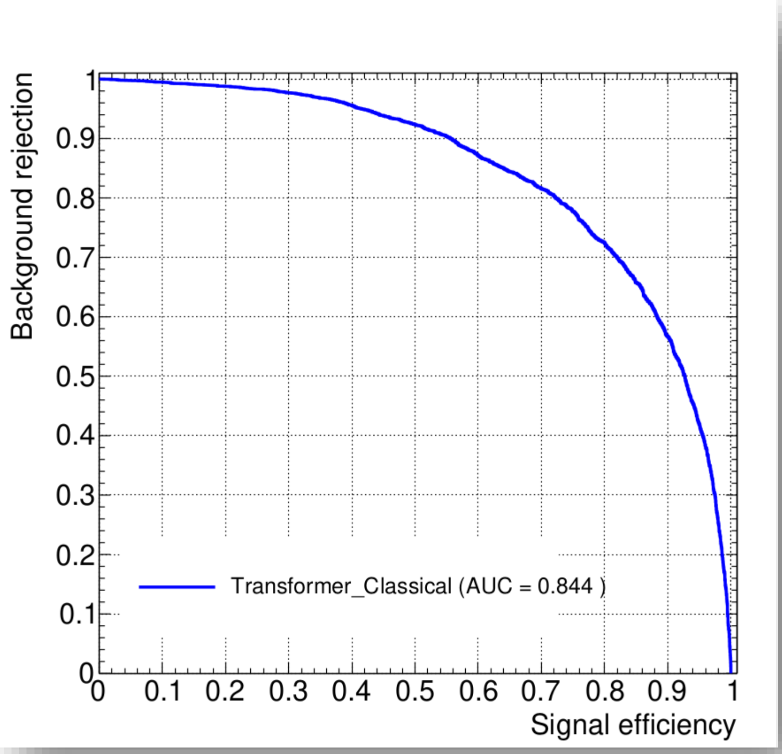
QPT

- Current results (Use CPU): ~76% acc on validation dataset both in Quantum transformer and classical transformer in 20k dataset (10k train, 10k val).
  - Quantum:

```
Epoch 8/30
Info in <TCanvas::Print>: pdf file ./plot/ROCs/ROC_10000_train.pdf has been created
Info in <TCanvas::Print>: pdf file ./plot/ROCs/ROC_10000_val.pdf has been created
Epoch: 08 | Epoch Time: 140m 18s
        Train Loss: 0.510 | Train Acc: 76.24%
         Val. Loss: 0.500 |  Val. Acc: 76.52%
```

  - Classical:

```
Epoch 27/30
Info in <TCanvas::Print>: pdf file ./plot/ROCs/ROC_20000_train_Classical.pdf has been created
Info in <TCanvas::Print>: pdf file ./plot/ROCs/ROC_20000_val_Classical.pdf has been created
Epoch: 27 | Epoch Time: 0m 48s
        Train Loss: 0.498 | Train Acc: 76.56%
         Val. Loss: 0.491 |  Val. Acc: 76.59%
```



  - By the way, the AUC of SVM is ~0.94 in 20k. In the case of small dataset, the performance of transformer is inevitably inferior to SVM and we need to make drastic changes to fit small dataset for same performance as SVM. (SVM is not suitable for large dataset because of O(n^2))