Machine Learning Applications for Track Fitting on the STAR Forward Tracking System

Colby Smith of Abilene Christian University for the STAR Collaboration

ABILENE CHRISTIAN UNIVERSITY

www.acu.edu

Solenoidal Tracker at RHIC

The Solenoidal Tracker at RHIC (STAR) Experiment seeks to study the makeup and structure of nuclear particles through heavy ions collisions and polarized proton-proton collisions. Until recently, only mid-rapidites could be measured based on the structure of the STAR solenoid. This left out crucial information about what happens at forward rapidities. Tracking in the forward rapidity region is crucial in measuring potential asymmetries caused by the Collins effect.

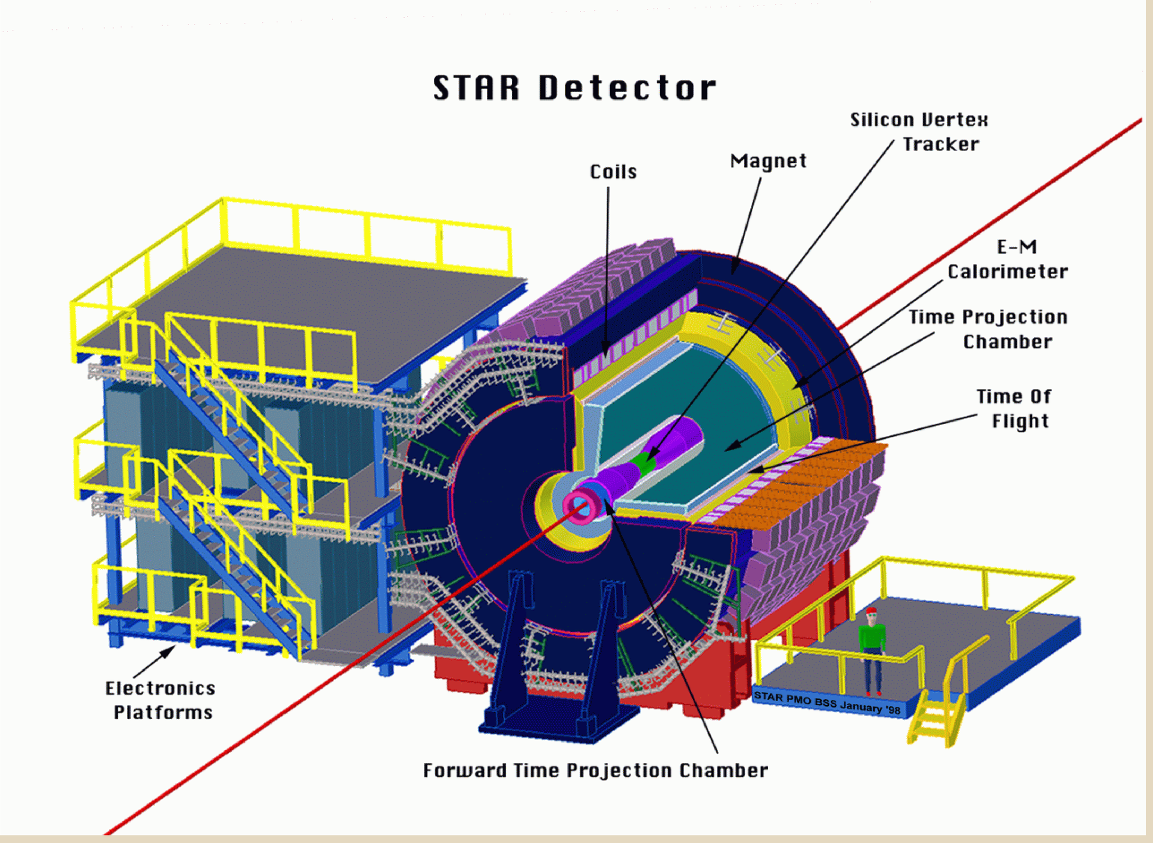

Figure 1: The STAR Detector, prior to its Forward Upgrade

Forward Upgrade

The Forward Upgrade extends the coverage of STAR to rapidities between $2.5 \leq \eta \leq 4$. The upgrade consists of two systems: the Forward Tracking System (FTS) and the Forward Calorimeter System (FCS) pictured below.

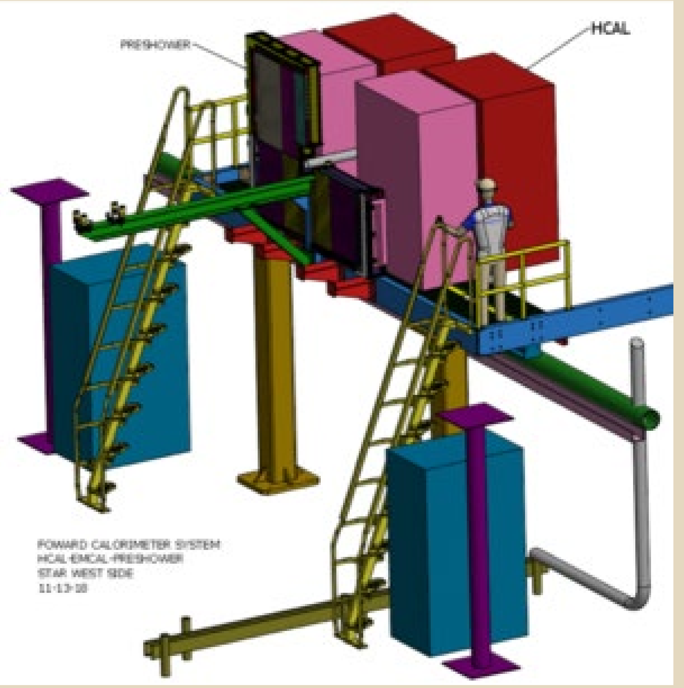

Figure 2: 3D Model of the Forward Calorimeter System

Forward Tracking System

The FTS is comprised of two detectors: the small-strip Thin Gap Chamber (sTGC) and the Forward Silicon Tracker (FST). These detectors are oriented on the forward face of the solenoidal TPC. The FST is made up of three circular silicon plates, while the sTGC is made up of four pentagonal thin wire chambers. Each detector plane records hits from charged particles that pass through after a collision. Figure 3 below shows the orientation of both detectors relative to STAR.

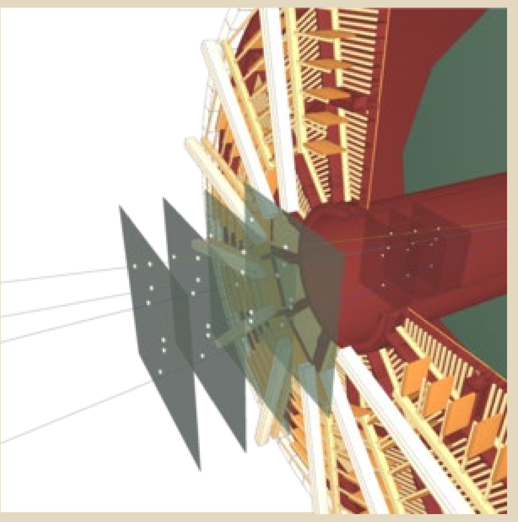

Figure 3: 3D Model of the sTGC (green) and FST (red)

Track Finding

Track finding is the process of reconstructing charged particle trajectories from hits recorded in the planes of the FTS. Previously, this had been accomplished by "cutting" the data according to certain criteria. The process begins with taking all pairs of hits in the sTGC between two consecutive planes. Then the criteria are calculated for each pair using . This process of removing hits pairs that are deemed "fake" is repeated for several other criteria on hit pairs, then all possible three point combinations are made and the criteria cutting continues. Eventually, all that is left are three hit combinations that are physically realistic. From here, it is easy to group two combinations to find a track through the sTGC. Once a track has been found, it is traced back to the FST. If a hit in the FST corresponds with this track, it is updated, and the process continues through all three planes.

Machine Learning Motivation

The process of classifying "real" and "fake" pairs by cutting over criteria can be improved using machine learning. A Multi-Layer Perceptron (MLP) Classifier model from scikit-learn was trained and tested on raw pair coordinate data with no cuts. This allows for the algorithm to infer patterns in the data about whether the points between two planes could realistically form a "real" hit.. Additionally, this method can be used for the FST and sTGC separately, then combined in the end to improve the overall rate at which tracks are found.

Efficiency vs Purity

When finding "real" or "fake" hit pairs, efficiency and purity of the results becomes important. Efficiency is defined as the ratio of "real" pairs kept vs. total "real" pairs. Purity is defined as the ratio of "real" pairs kept vs. total pairs kept. In general, efficiency is more important than purity because it is more important to keep as many "real" pairs in the result as possible. Any "fake" pairs that slip through can be found and removed at a later step. As shown in Figure 4, as more pairs are included the efficiency increases while the purity decreases.

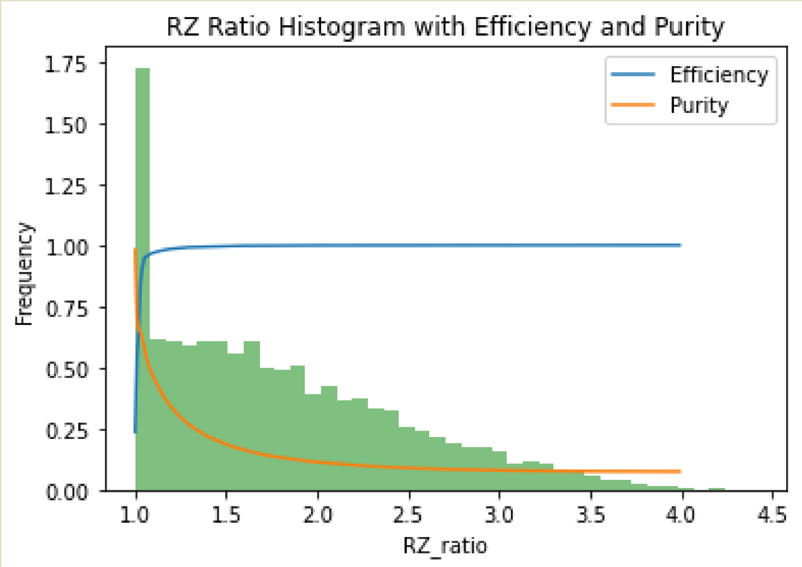

Figure 4: RZ Ratio Criteria, Efficiency, and Purity

The RZ ratio criteria is the ratio between the total distance between two points, 'r', and the distance between the two planes each point was on, 'z'.

Preliminary Results

To test the viability of the MLP classifier, it was tested and trained over simulation data in the FST. The efficiency and purity were calculated over several different input parameters, as shown in Table 1. Each model was run 10 times with random states to ensure model stability.

Parameters	Efficiency [%]	Purity [%]
XYZ Pair	0.8723±0.0042	0.7690±0.0040
Rho, Phi, Z Pair	0.9052±0.0033	0.7846±0.0174
Previously used Criteria	0.8979±0.0034	0.7849±0.0058

Table 1: Comparison of Parameters in the FST MLP Classifier

For the FST, the rho, phi, and z data points achieved the greatest efficiency. This is likely caused by the physical makeup of the FST being spherical in nature. Overall, the cylindrical coordinates performed better than rectangular coordinates and the criteria used in previous track finding. When moving to track reconstruction, these parameters should be used to train the model for the FST.

Conclusion

In conclusion, the MLP Classifier from scikit-learn has a high efficiency in identifying "real" hit pairs in the FST. More research must be done to further expand machine learning to complete the track finding between the FST and sTGC, and to prove whether this method is viable. However, with an initial efficiency of >90%, machine learning has promising applications for track finding in the STAR Forward Upgrade.

Acknowledgements

Thank you to Dr. Jim Drachenberg, Dr. Daniel Brandenburg, and Noah Mitchell

U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY