Using Machine Learning to identify slow pion hits in SVD and PXD and the road to full VXD tracking

Christian Wessel
DESY, 01.10.2024

christian.wessel@desy.de

HELMHOLTZ

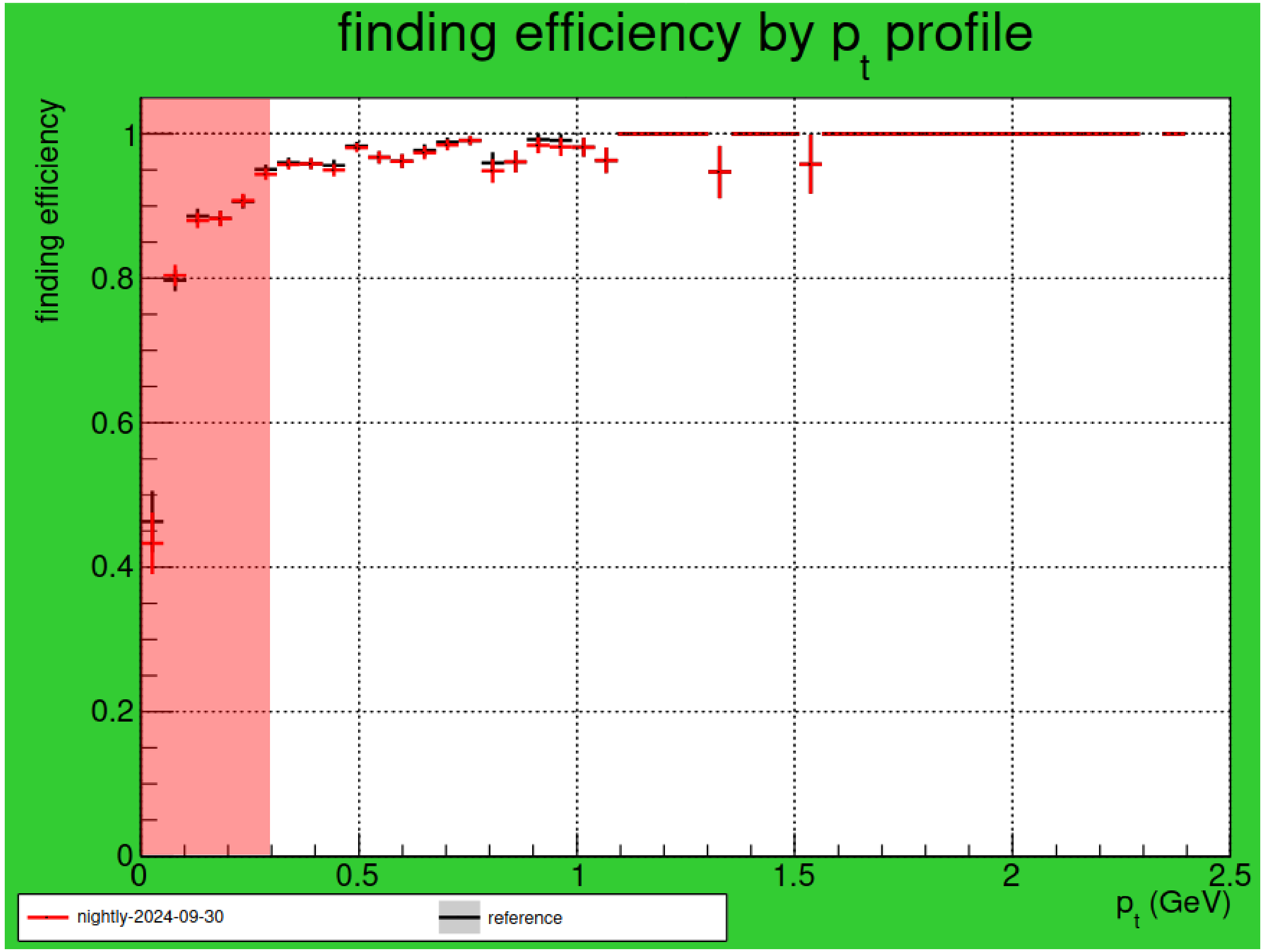
Belle II

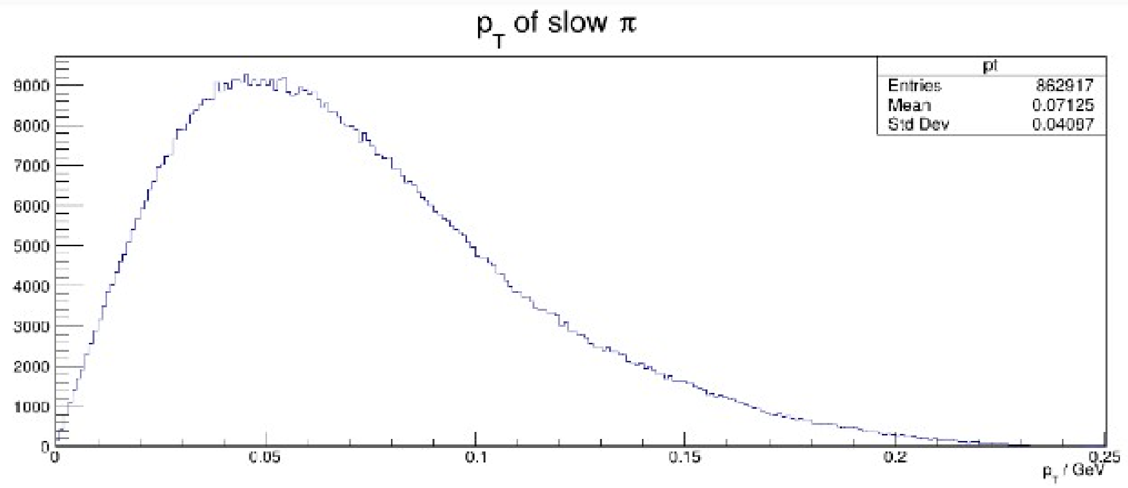
DESY.

The issue

Why?

- We lose a lot of tracks in the (very) low p_T region
- Many analyses suffer from this, for instance and in particular $B \to D^{*+} [\to D^0 \pi_s^+] X$ decays, where the D^{*+} can't be reconstructed if the resulting slow pion is missing
- But not limited to these processes
- Can we improve low p_T tracking?

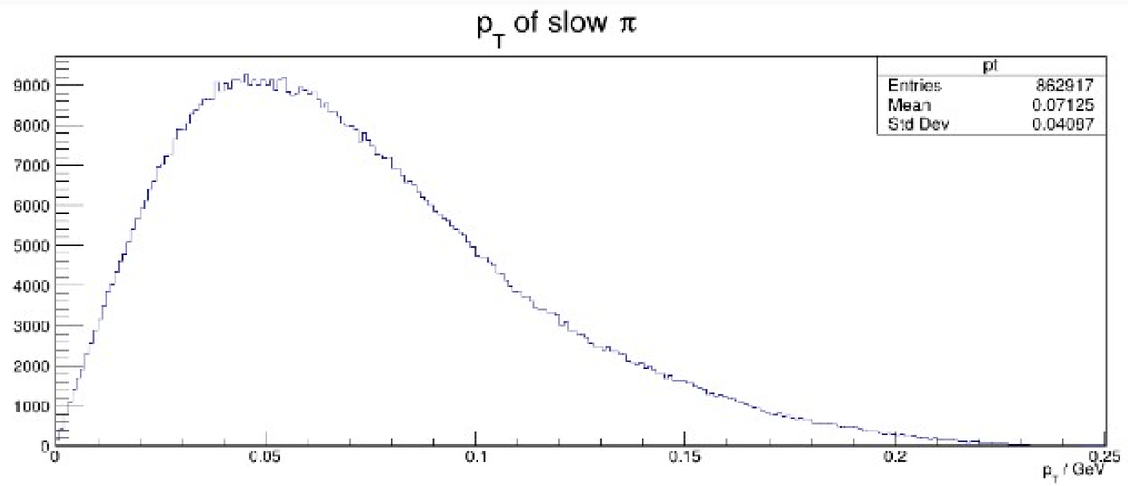


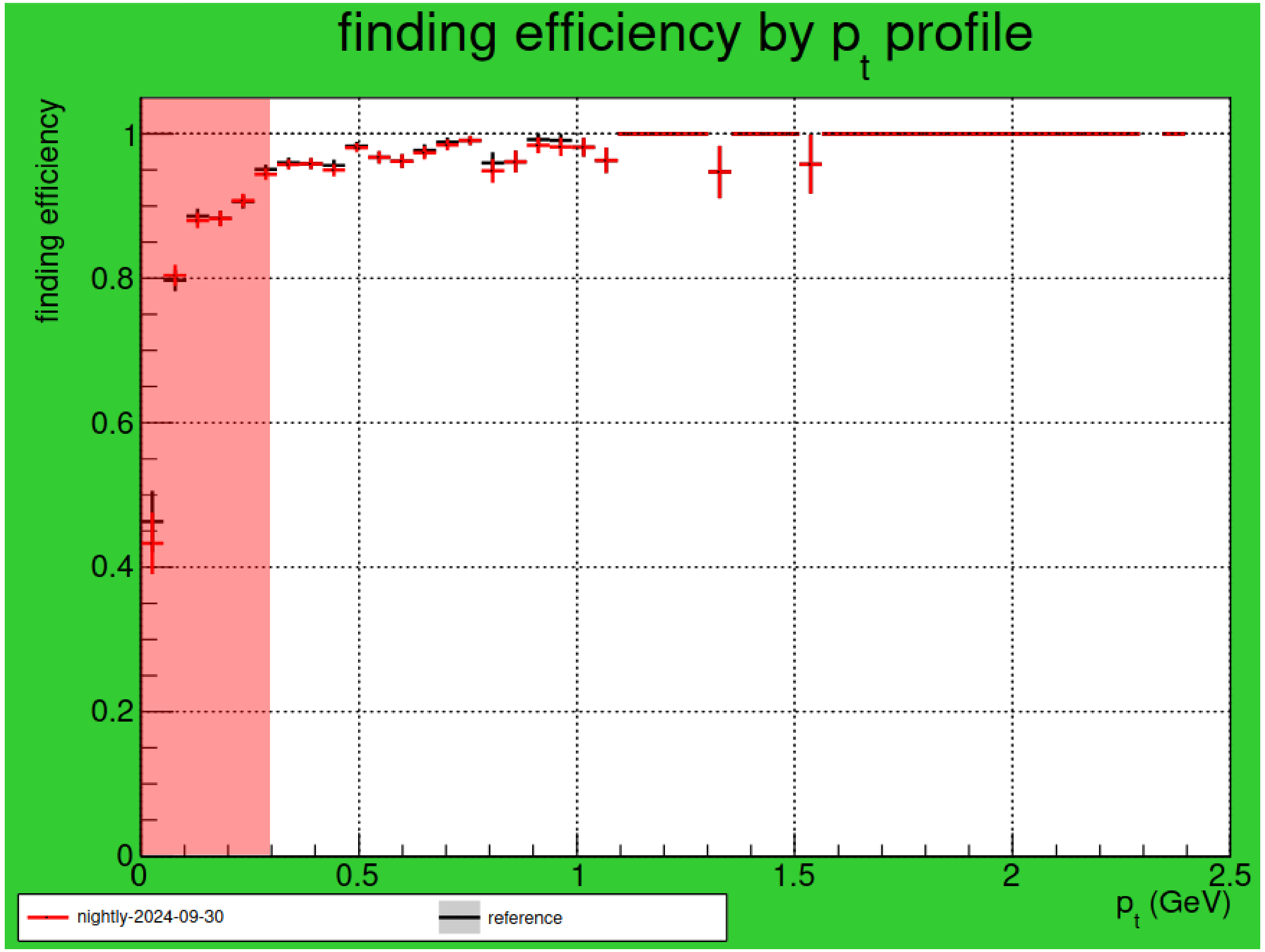


The issue

Why?

- We lose a lot of tracks in the (very) low p_T region
- Many analyses suffer from this, for instance and in particular $B \to D^{*+} [\to D^0 \pi_s^+]X$ decays, where the D^{*+} can't be reconstructed if the resulting slow pion is missing
- But not limited to these processes
- Can we improve low p_T tracking?
- I think we can!



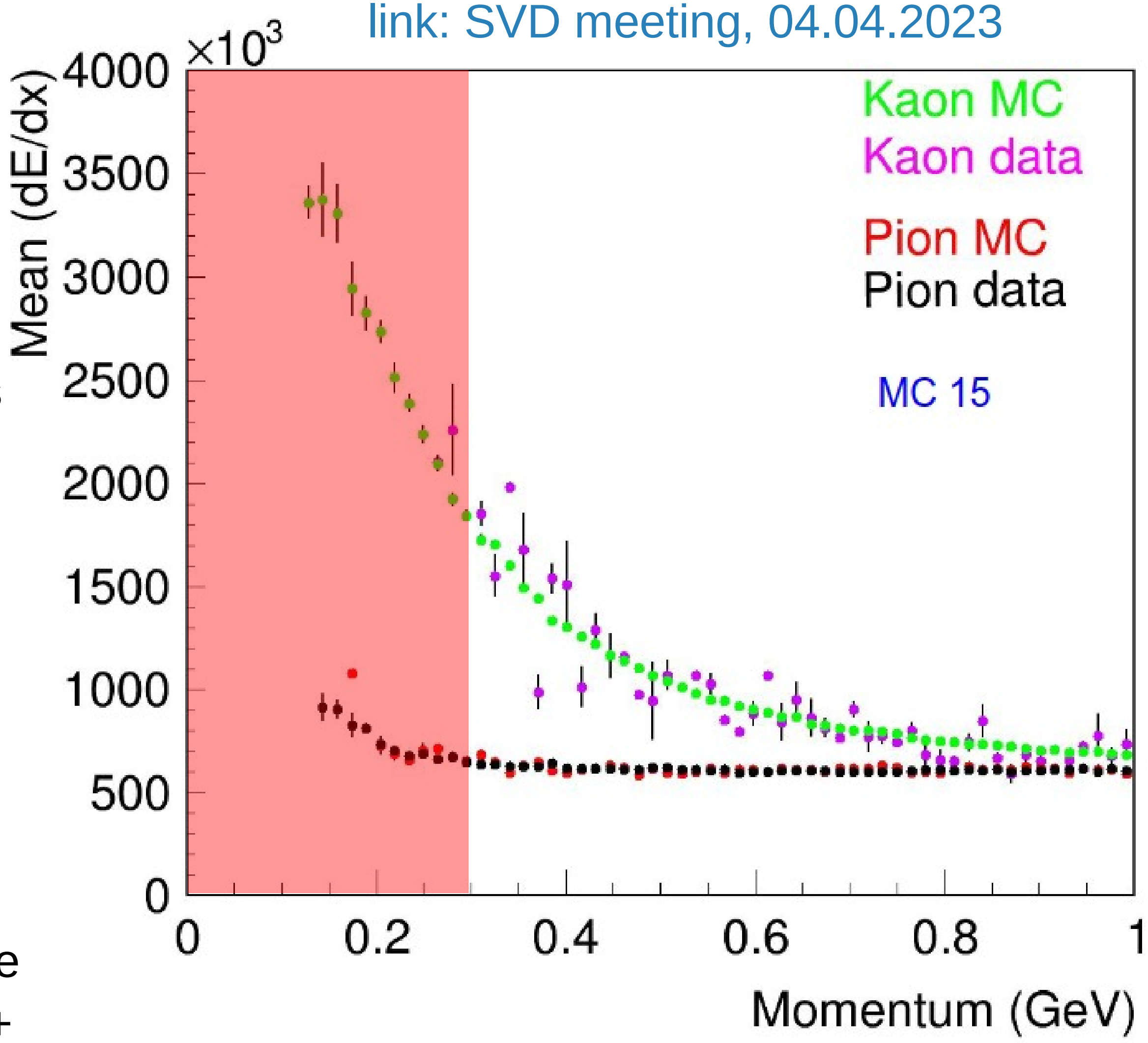


The idea

How?

- Slow pions, but also other low p_T particles, should have higher than average energy loss in SVD
- Can we use these information to identify slow pion SVDClusters / SVDSpacePoints?
- Train MVA (FBDT) with SpacePoints in B –> D^{*+} X events
- Only ~4 SpacePoints per event are from slow pion, but O(900) SpacePoints are from background particles (nominal BG, excluding SpacePoints from other signal particles)
  - Train on all slow pion SpacePoints
  - Only use a small fraction of background (0.7%) SpacePoints to obtain a balanced training data set
- Information used from clusters (u + v): (seed) charge, time + time error, size, signal-to-noise ratio, local coordinates + errors, layer, ladder, sensor

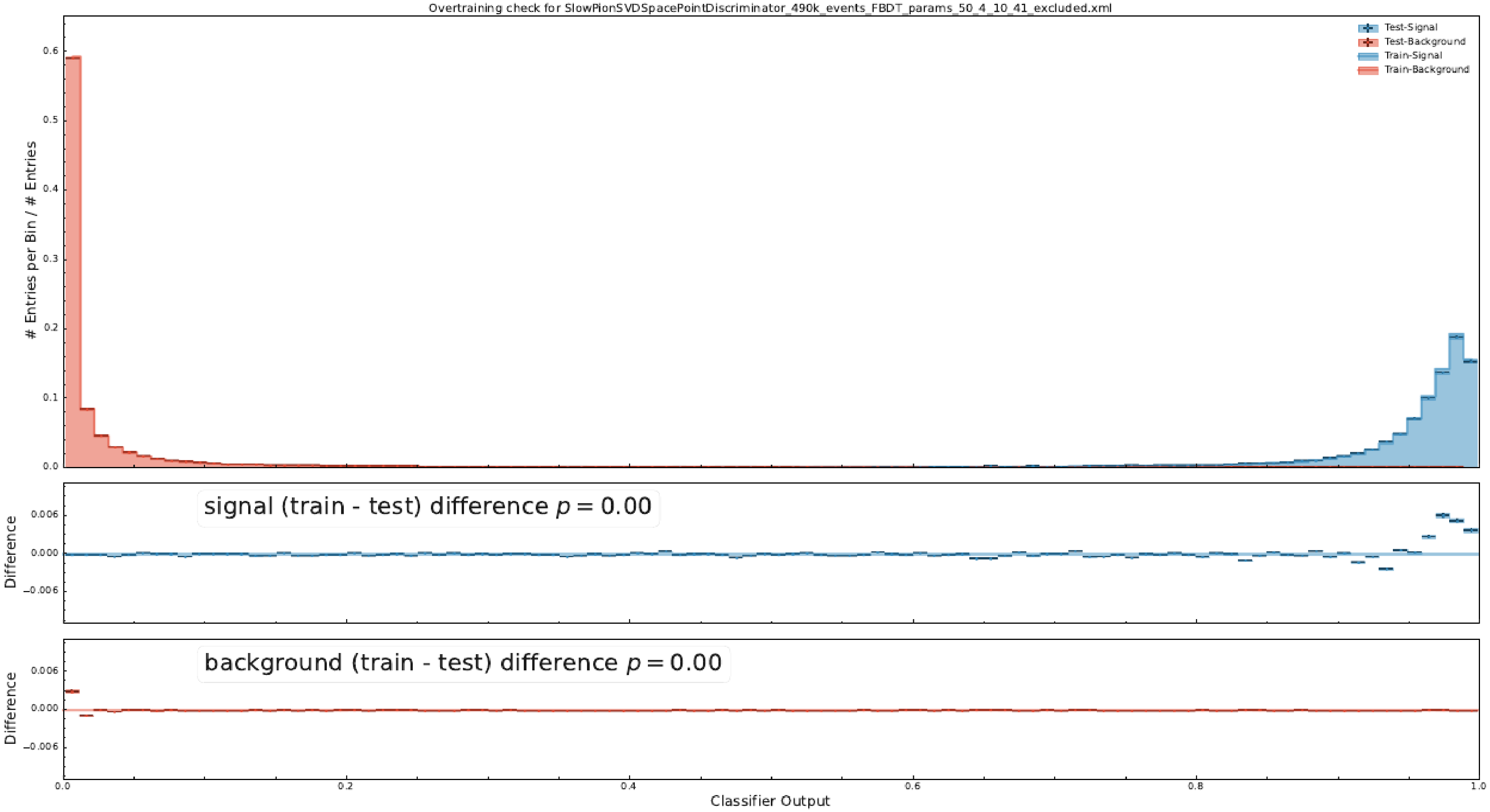
link: SVD meeting, 04.04.2023



Training results for SVD

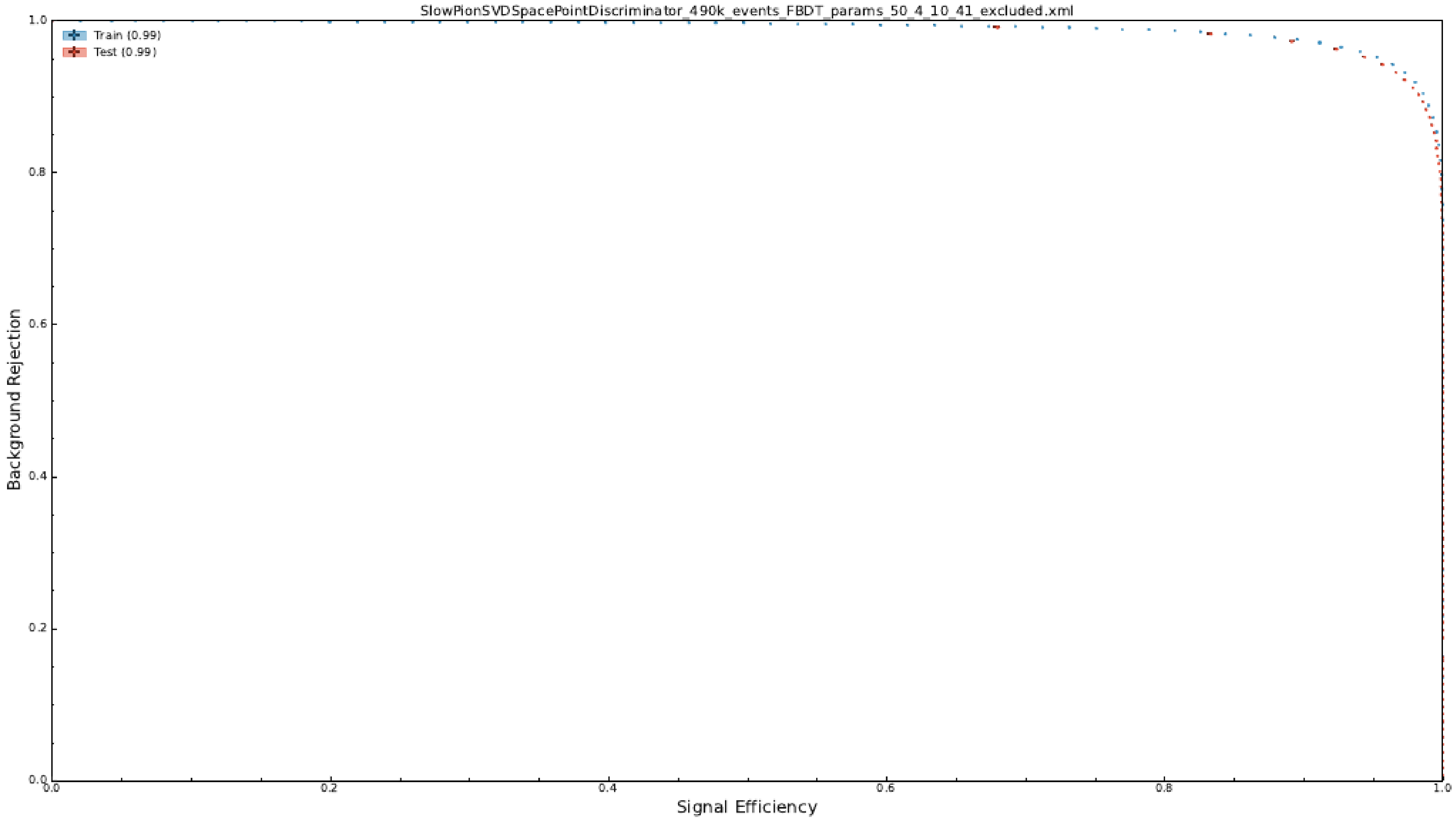
Classifier output distribution

- Near perfect separation between signal and background



Training results for SVD

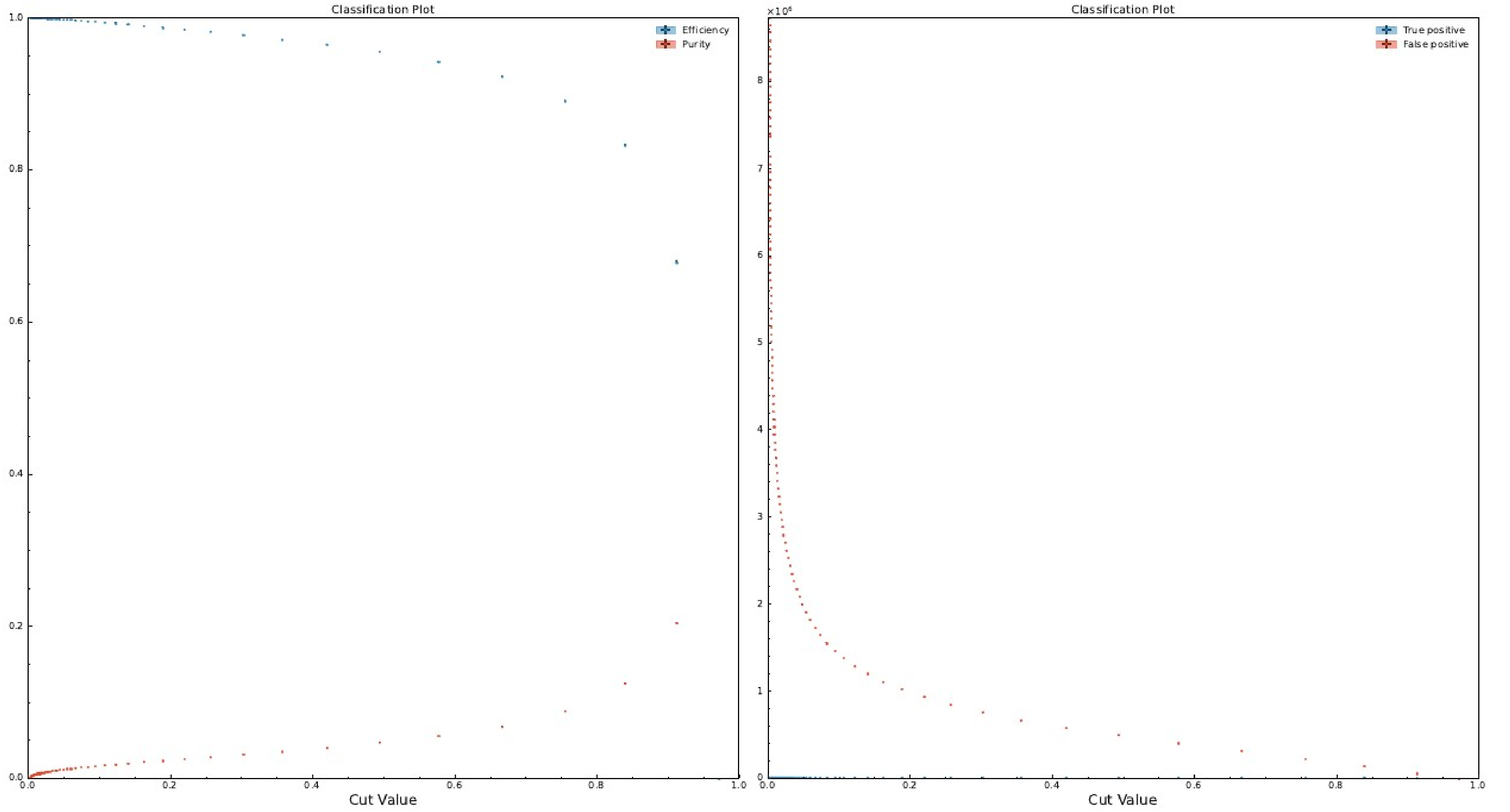
ROC



Training results for SVD

Efficiency and purity

- Background still “overwhelms” after removing 90% of it



Cluster Selection in PXD

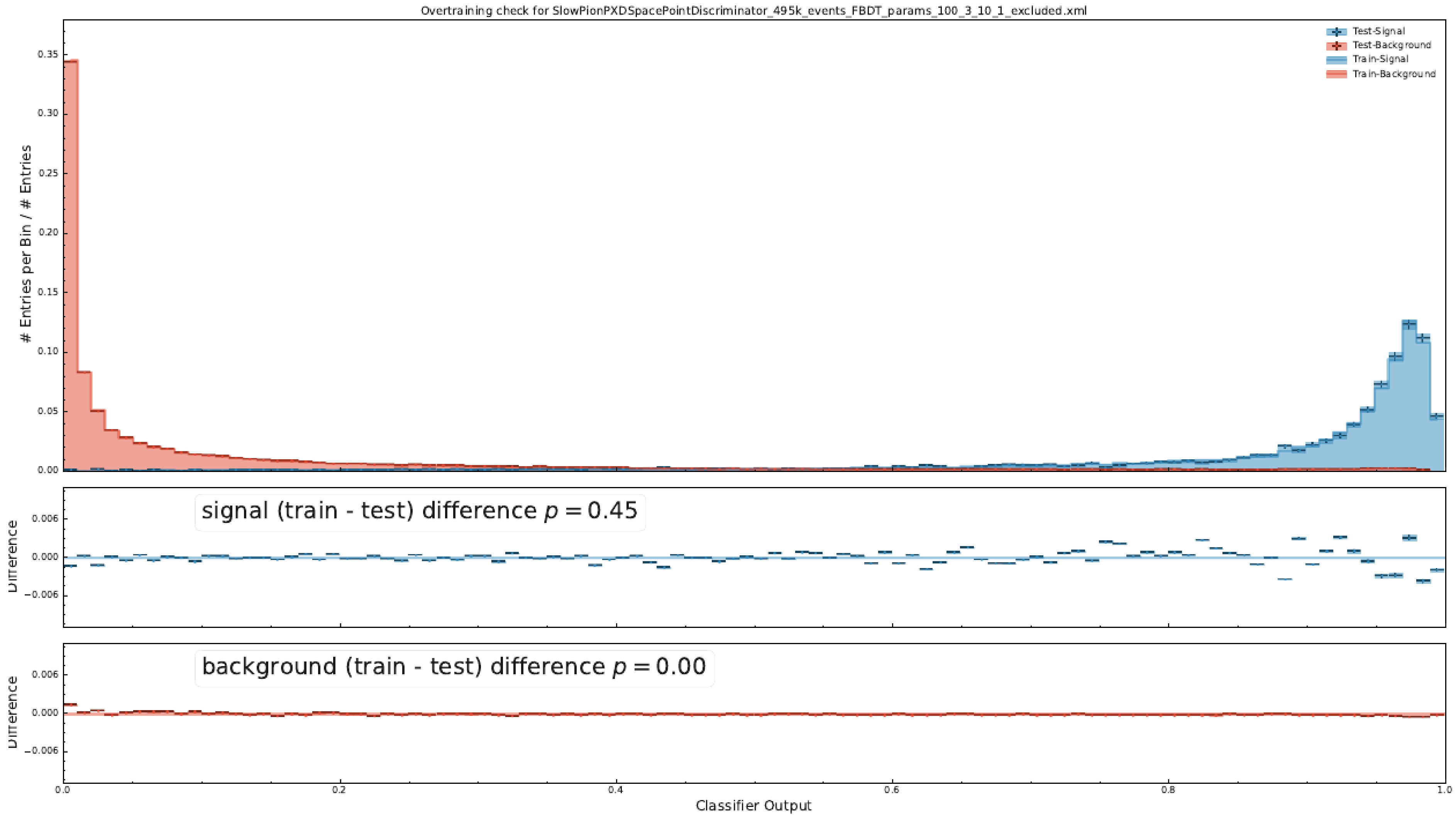
Does it work with a BDT?

- In simulation classification seems to work for SVD SpacePoints to distinguish slow pion Sps vs background
- PXD is much thinner than SVD –> separation based on cluster energies likely is much more challenging
- Background fraction much larger compared to SVD: ~2 slow pion cluster per event vs O(>10k) background clusters at nominal luminosity
- Similar inputs as for SVD: cluster positions, size (total, u, v), (seed) charge, layer, ladder, sensor
- Use all signal hits in training, but select only 0.11% of background hits to obtain balanced training data set
- FastBDT parameters for plots in following slides:
  - 100 trees
  - 3 cuts
  - 10 levels –> distributions divided in 2^{10} = 1024 bins for equal frequency binning
  - Shrinkage of 0.1
  - Smaller BDTs (less trees, less bins) yield similar but slightly worse results

Training results for PXD

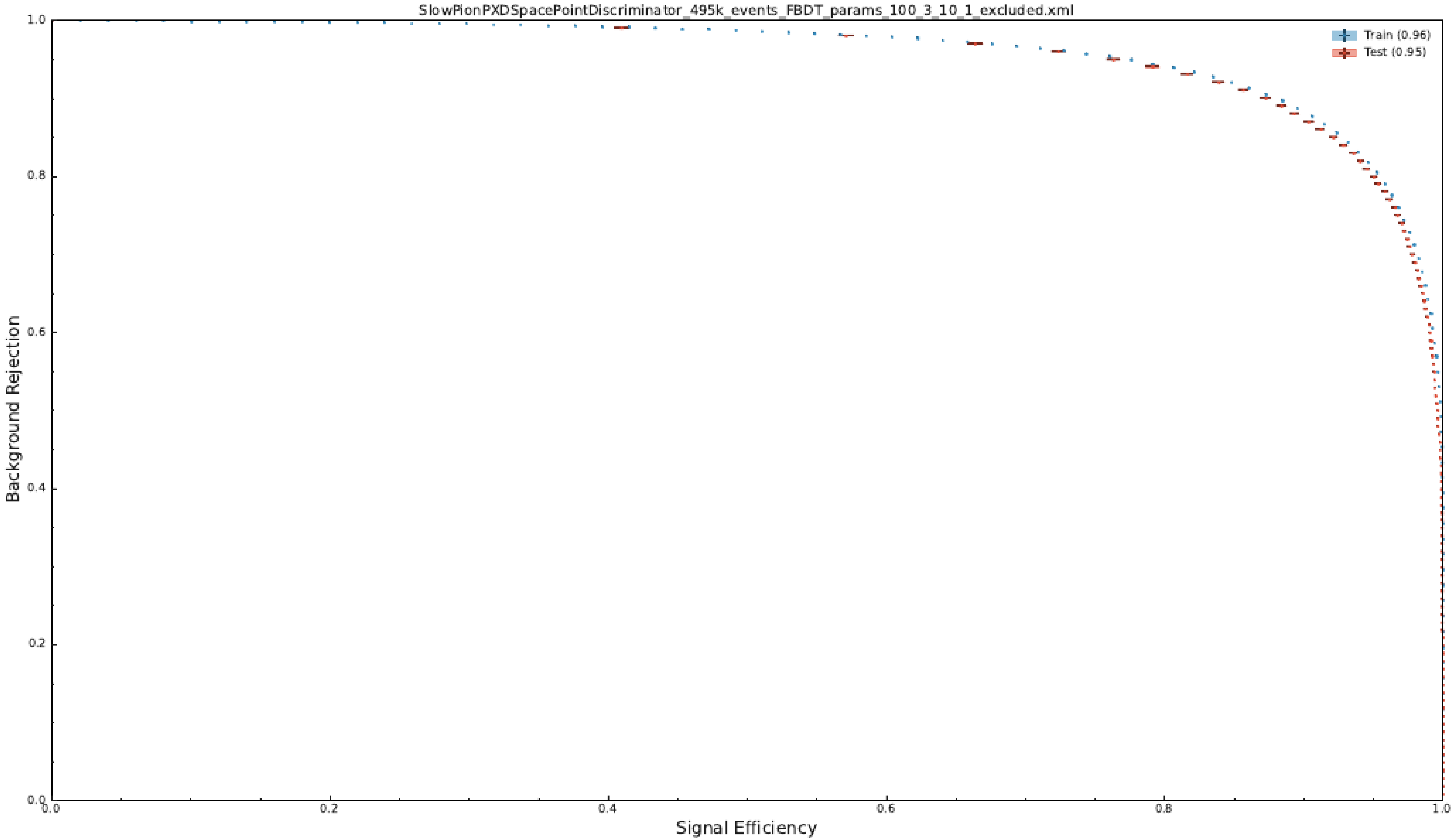
Classifier output distribution

- Good separation between signal and background



Training results for PXD

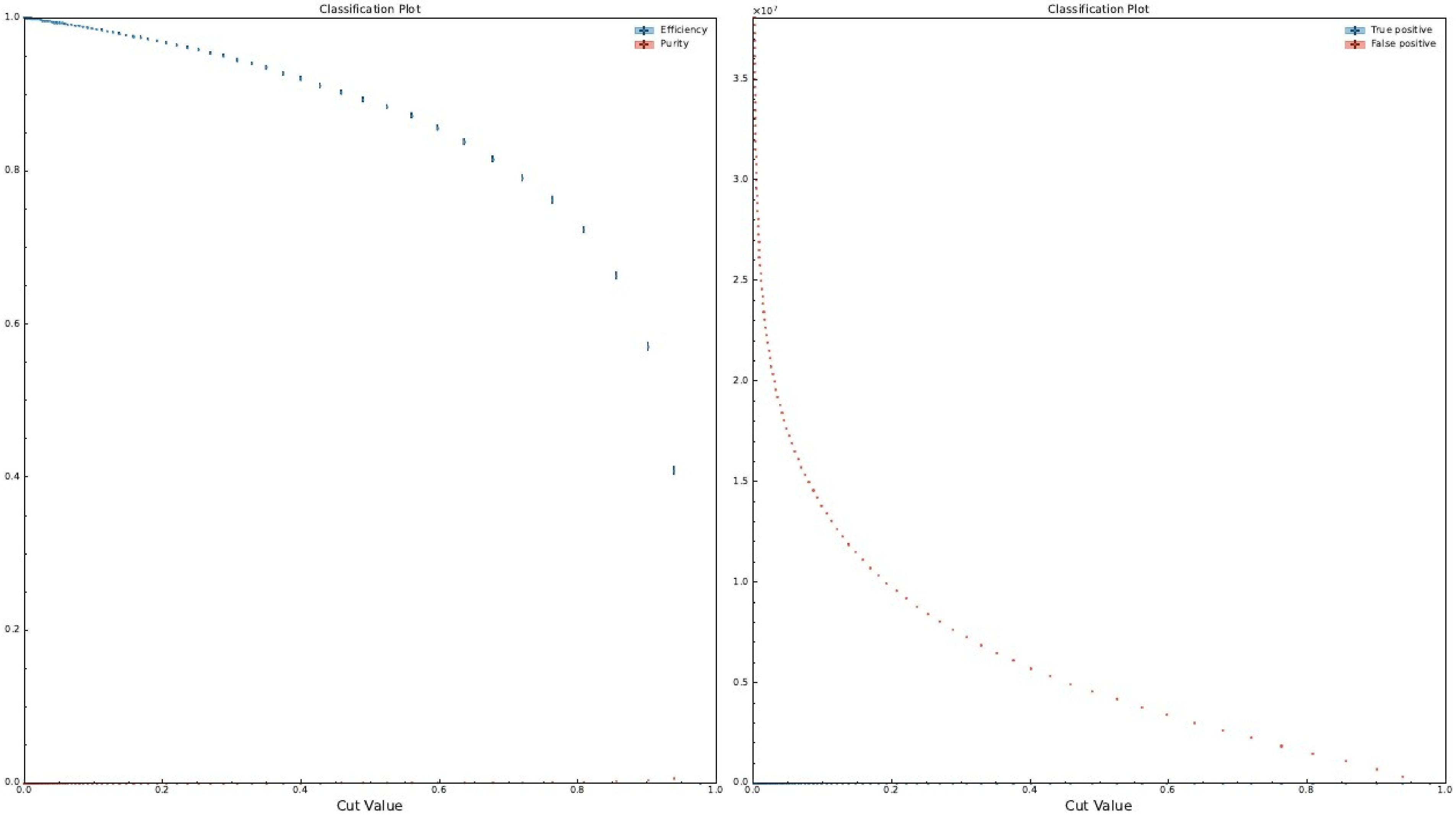
ROC



Training results for PXD

Efficiency and purity

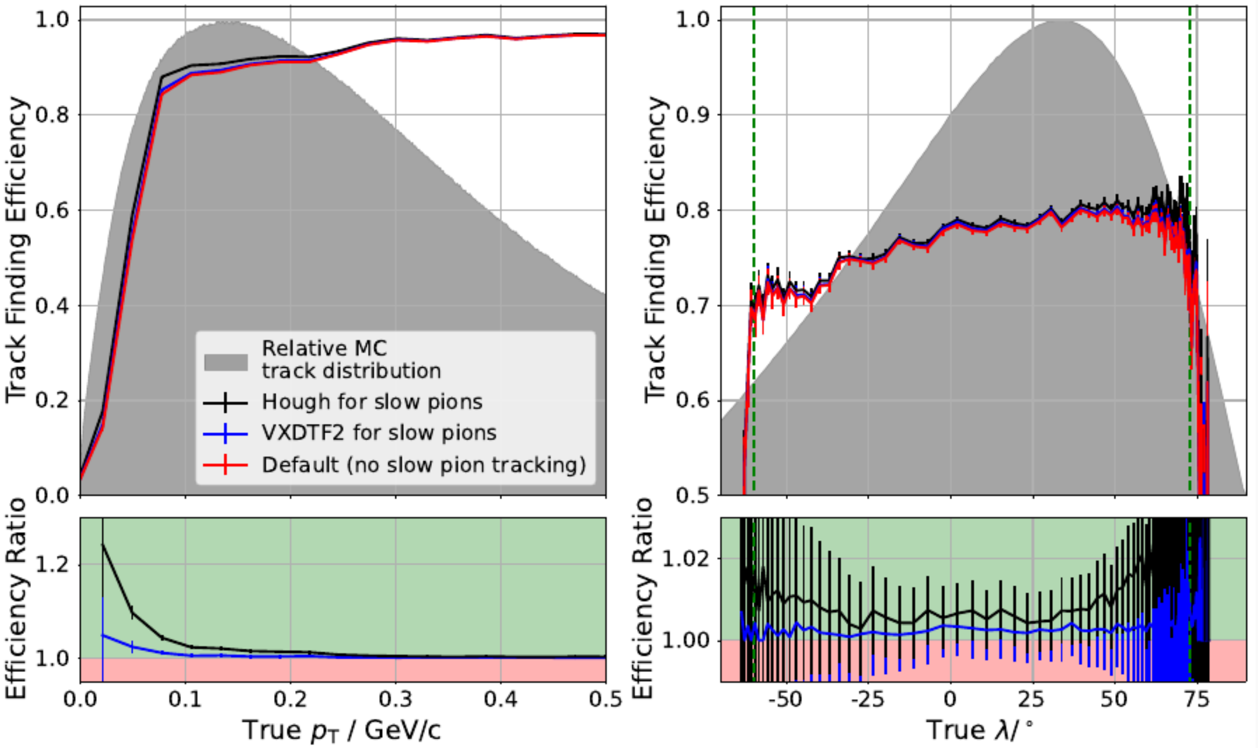
- Even with 90% background rejection the slow pion PXD clusters are still a needle in the hay stack

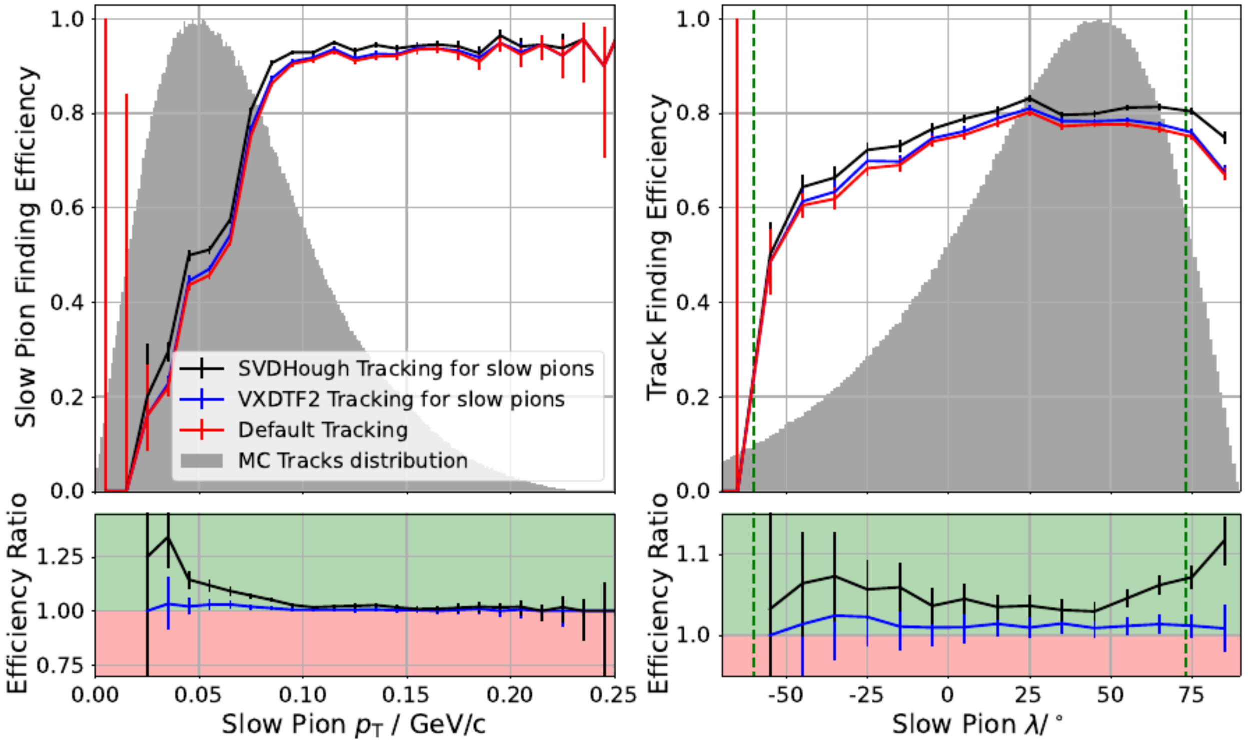


Slow Pion tracking

Find slow pions in SVD and PXD

- Proven to work for SVD-only tracking using VXDTF2 and HoughTracking
- For now full Hough Space used, will extend this to avoid high pt region –> focus on low p_T particles only
- Significant improvement in general (up to +0.9% on average, left) and slow pion (up to +3.7% on average, right) track finding efficiency
  - Below p_T of ~70 MeV still significant room for improvement –> need SVD+PXD tracking

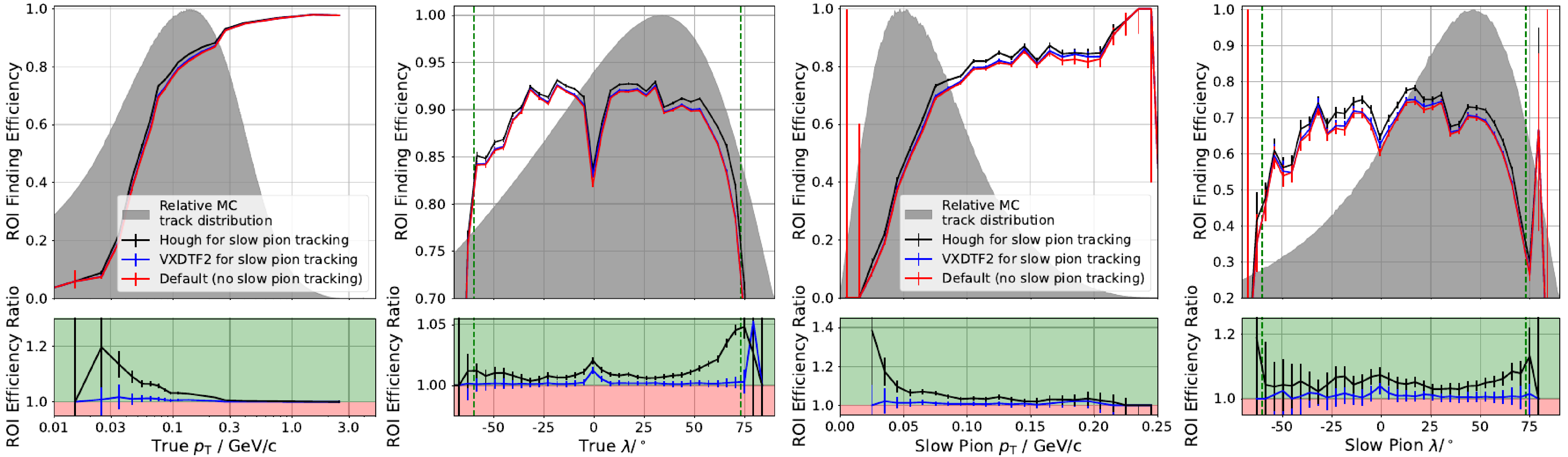




Slow Pion tracking

Find slow pions in SVD and PXD

- Additional ROI created from the additional tracks, ROI efficiency improved
  - Total up to +1% on average (left), slow pions up to 3.2% on average (right)
  - Below 70 MeV still significant room for improvement
    –> ROI based on SVD hits on two layers + IP constraint?



Slow Pion tracking

Future plans

- Create additional tracks from just two SVD hits and IP constraint for more ROI
  - Should increase ROI efficiency for low pt particles even more
- Extend currently existing Hough tracking to full VXD tracking including PXD hits, or use VXDTF2 with dedicated training to include PXD in tracking in addition to existing ToPDXCKF
  - Both for tracks from 3 SVDSpacePoints and tracks with 2 SVDSpacePoints + IP constraint
- Attempt to extend high track finding efficiency range down to p_T of ~40-50 MeV
- Concept of MVA selection can be extended to other, heavier low pt particles that lose even more energy in detector
- Need to evaluate performance on data and carefully monitor efficiency, fake rate, and data reduction

Summary and Outlook

Summary

- Slow Pion hit identification in SVD and PXD can be achieved with ML based on FastBDT
  - Reasonable performance even with small number of trees
- Complementary to approach of other groups
  - Try to find slow pion hits in SVD first instead of PXD first
    - PXD is very thin –> low energy deposit
      PXD is close to IP –> highest amount of background hits
- Tracking in SVD with identified SVD SpacePoints shown to work
- Extend tracking to include PXD
- Potentially extend to higher mass low p_T particles
- Maybe: full MVA selection of SVD signal hits to remove background hits before tracking
  - Potentially dangerous as this could remove signal hits

Thank you for your attention!

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Christian Wessel
Belle II
christian.wessel@desy.de