SI-Enhanced Reconstruction for the RGH Tracking Systems

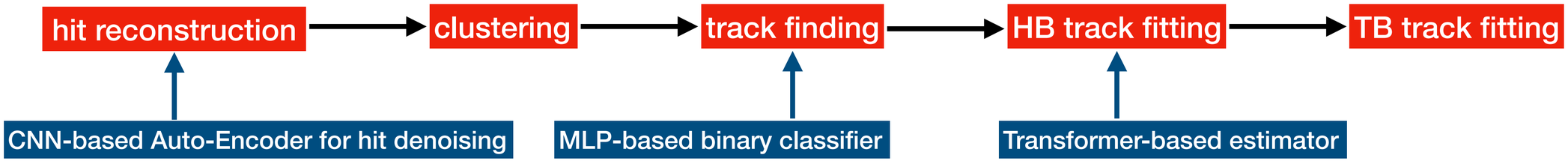
T. Cao

Nov. 13-14, 2026

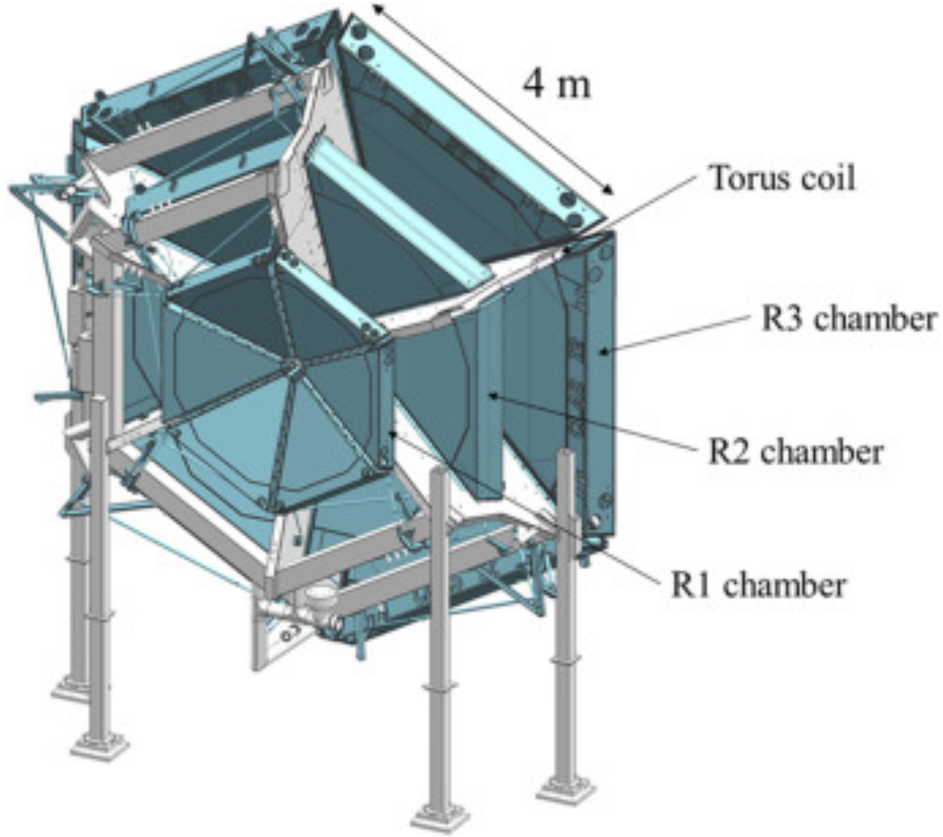
RGH Workshop

SI-Enhanced Reconstruction for the Forward Tracking System

SI-Enhanced Reconstruction for Forward Tracking

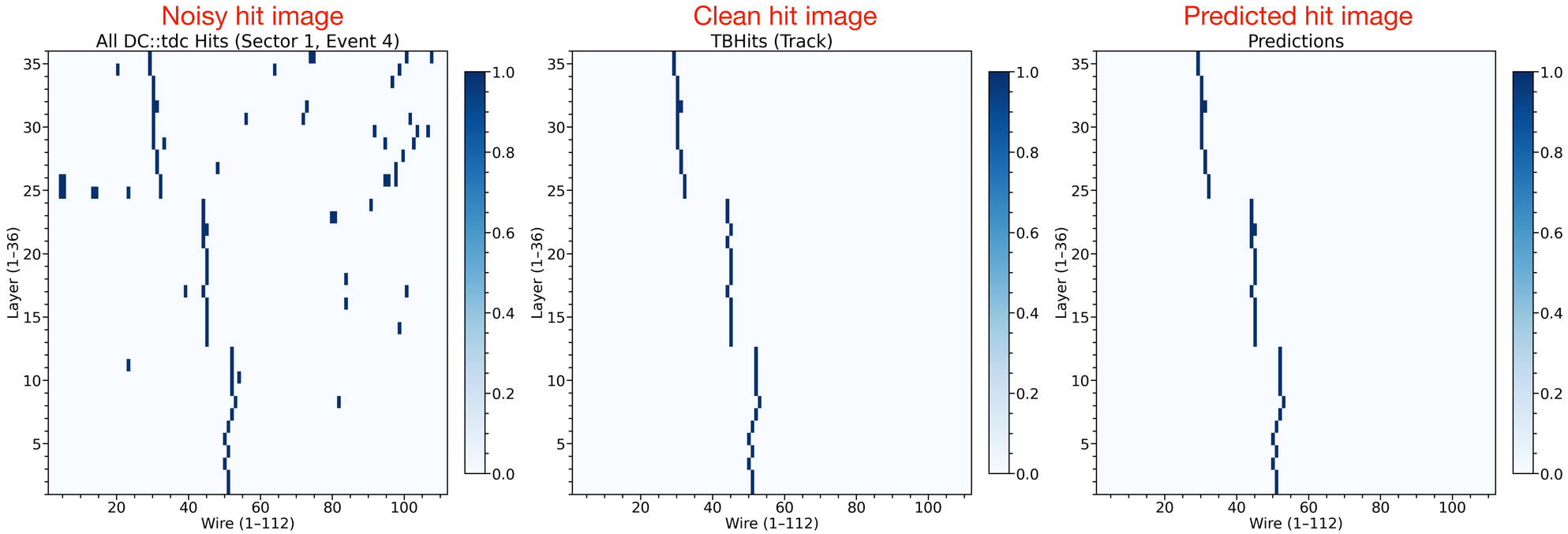


- **Benefits of SI:** SI can learn complex patterns in detector measurements and their relationships to particle trajectories, enabling effective background suppression, improved reconstruction performance, and reduced computational cost.
- **SI-enhanced reconstruction:** SI models are applied at three stages in the pipeline of the reconstruction:
  - Hit denoising: suppress noise hits
  - Track finding: identify cluster combinations as track candidates for tracking
  - HB track fitting: estimate track-state parameters
- **Software integration and application:** The SI models have been integrated into the reconstruction software and successfully applied in data cooking.

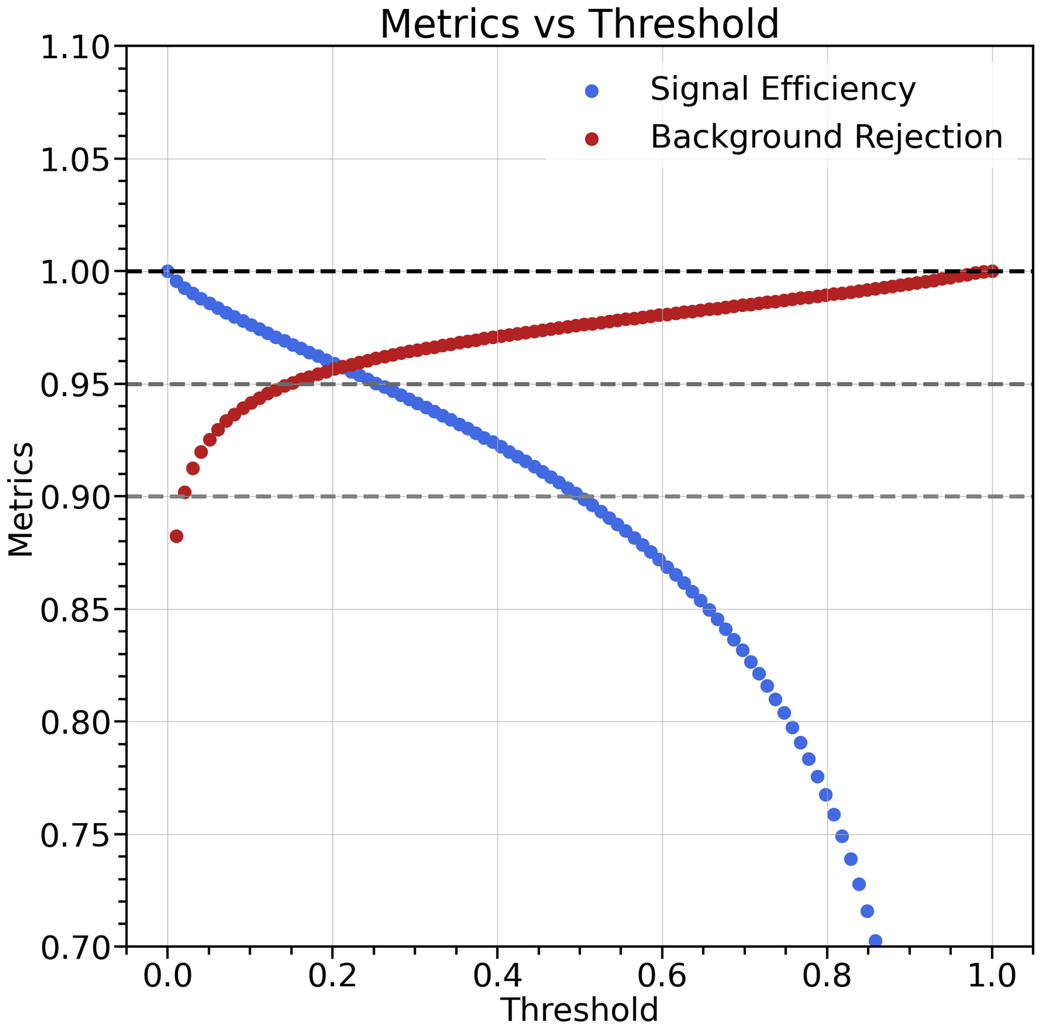


Hit Denoising by Convolutional Auto-Encoder

- CLAS12 suffers from high DC hit occupancy, especially at R1.
- A convolutional-neural-network (CNN) based auto-encoder learns global spatial correlations of tracks and suppresses non-track-like patterns automatically.
- The auto-encoder takes a noisy hit image as input, compresses it into a latent representation, and reconstructs a cleaned version of the same image.
- The CNN-based auto-encoder is ideal because DC hits live on a 2D grid (layer × wire), tracks form correlated structures, and noise is mostly spatially uncorrelated.



Model Performance



- Model's hyper-parameters:
  - Blocks: 2
  - Filters: 48
  - Kernel size: 4 × 6
- Threshold is set as 0.025 by default to obtain the best tracking efficiency.
- ~90% of noise hits are rejected, while ~99% of signal hits are retained.

Track Finding by MLP

- **Track topology:** The forward tracking system consists of six superlayers in each sector. Both 6-cluster tracks (all 6 superlayers have clusters) and 5-cluster tracks (one superlayer is missing a cluster) are reconstructed.

- **Reconstruction procedure:** 6-cluster combinations are identified firstly, and then 5-cluster combinations are identified using remaining clusters.

- **Challenges for conventional track finding:**

  - Reject fake 6-cluster combinations so that their clusters remain available for identifying 5-cluster combinations.

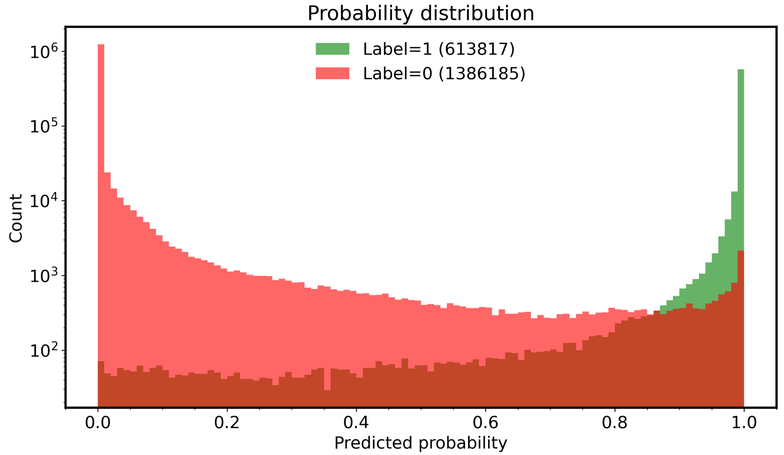
  - Resolve ambiguities when multiple cluster combinations share one or more clusters, selecting the most likely genuine combination.

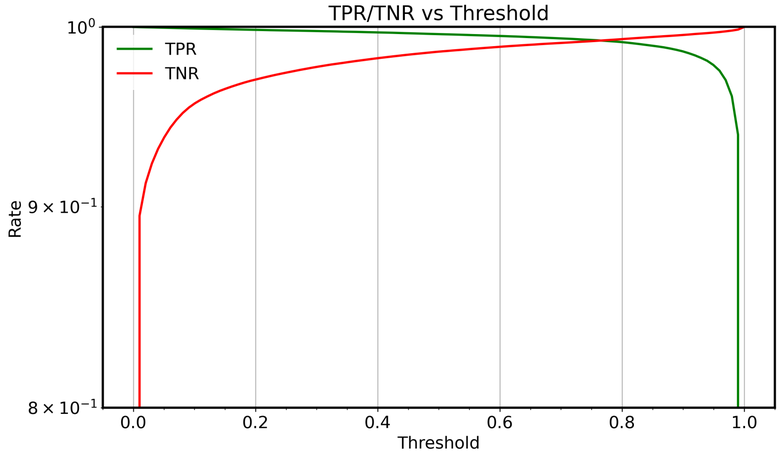
- **SI-based track finding:** Two Multi-Layer Perceptron (MLP) binary classifiers for 6-cluster and 5-cluster combinations were developed separately to distinguish genuine cluster combinations from fake combinations.

Model Performance

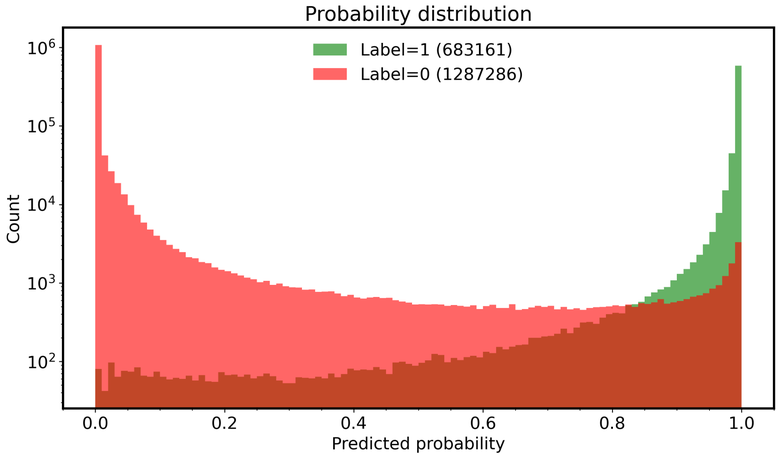
- TPR: True Positive Rate
- TNR: True Negative Rate

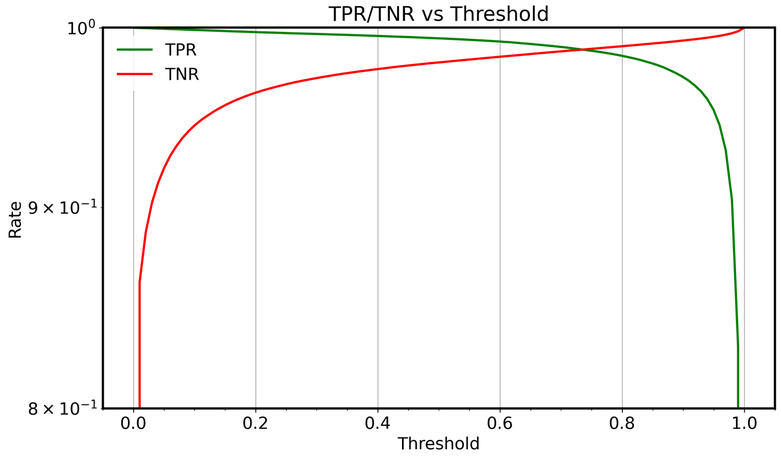
- **Model for 6-cluster combinations**
  - # of layers = 4
  - Dim. of hidden layers = 64
- Threshold is set as 0.95 by default.
- Almost all fake combinations are rejected, while ~98% of genuine combinations are retained.





- **Model for 5-cluster combinations**
  - # of layers = 3
  - Dim. of hidden layers = 64
- Threshold is set as 0.05 by default.
- ~94% fake combinations are rejected, while almost all genuine combinations are retained.

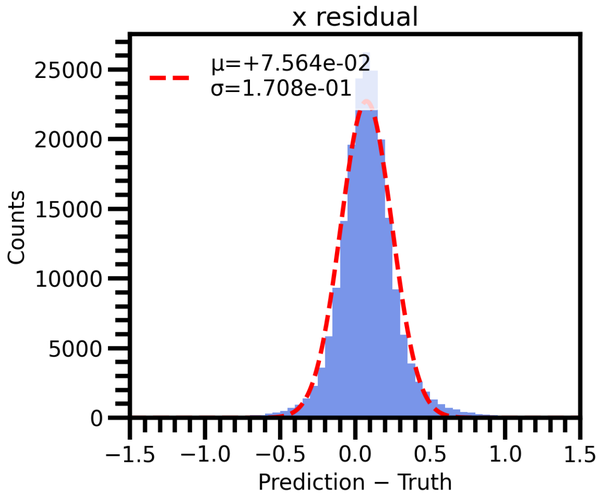


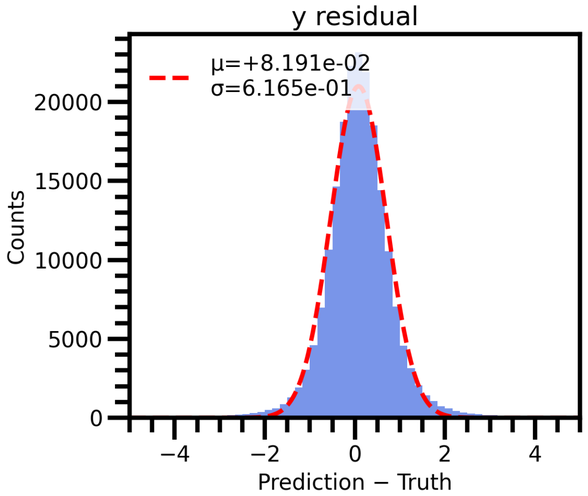


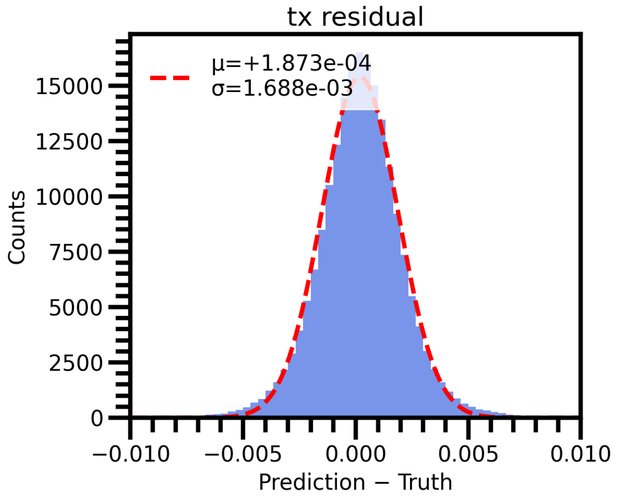
HB Track Fitting by Transformer

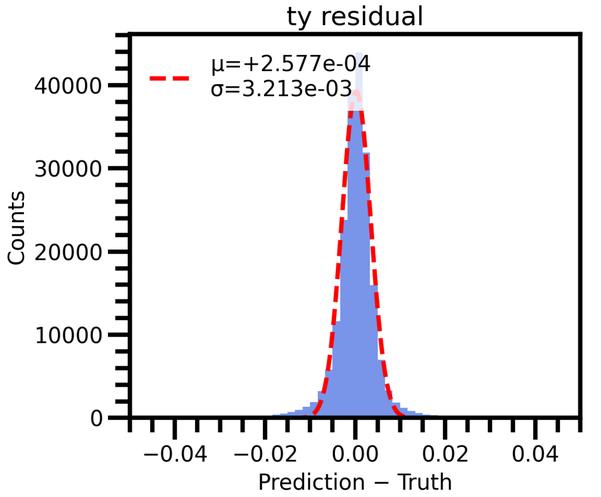
- HB tracking by conventional Kalman Filter (KF) faces challenges:
  - Approximate HB measurements: HB measurements are constrained by cell size, and do not precisely represent hit positions.
  - Left–right ambiguity: DC hits suffer from left–right ambiguity relative to wires.
  - Rough initial state estimation: Initial track parameters from the pattern recognition rely on an over-simplified quadratic fit to three crosses.
- The SI model based on the transformer technique treats track reconstruction as a sequence-to-state regression problem.
- The transformer leverages the self-attention mechanism to capture global correlations among hits from different detector layers, and it can directly estimate HB track state by hit pattern:
  $hit\ pattern \rightarrow track\ state\ (x, y, t_x, t_y, Q)$

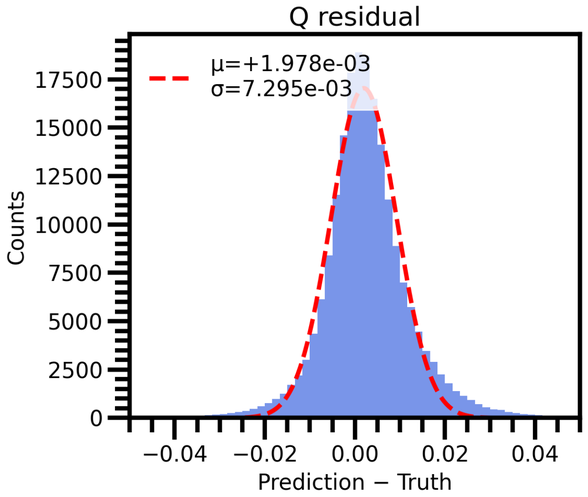
Model Performance







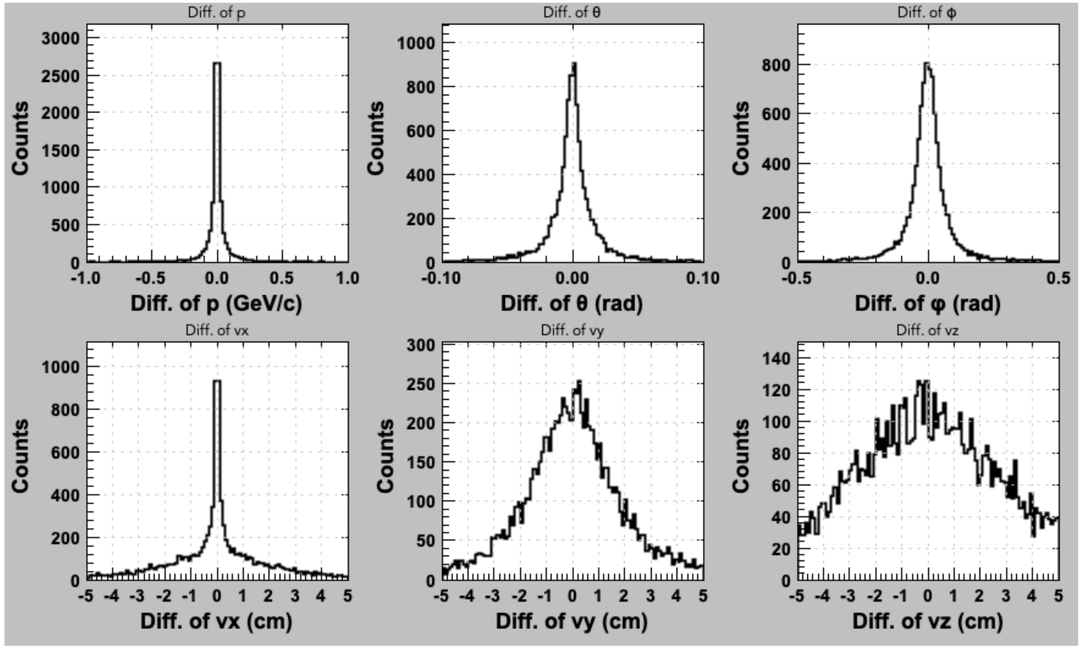




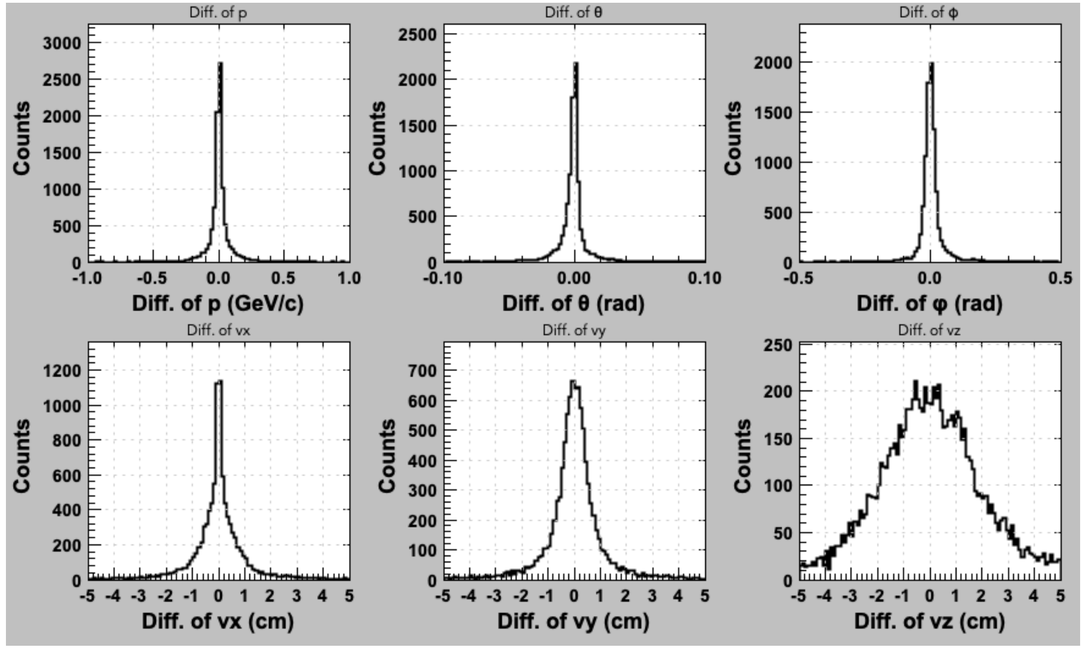
Truth is TB track state

Check HB Track Resolution with TB Tracks as Reference

HB tracking by KF



HB tracking by SI estimator



Difference between HB and TB track state at vertex

- Resolution for tracks by SI estimator is much better than HB tracking by KF.
- It will change quality for the following HB reconstruction based on HB tracks, and further affect TB reconstruction.

Discussion for SI and KF Tracking

- **KF-based HB tracking through thick tubes:** The conventional KF propagates the track state layer by layer through DC, using HB measurements with large uncertainties determined by size of DC cells.

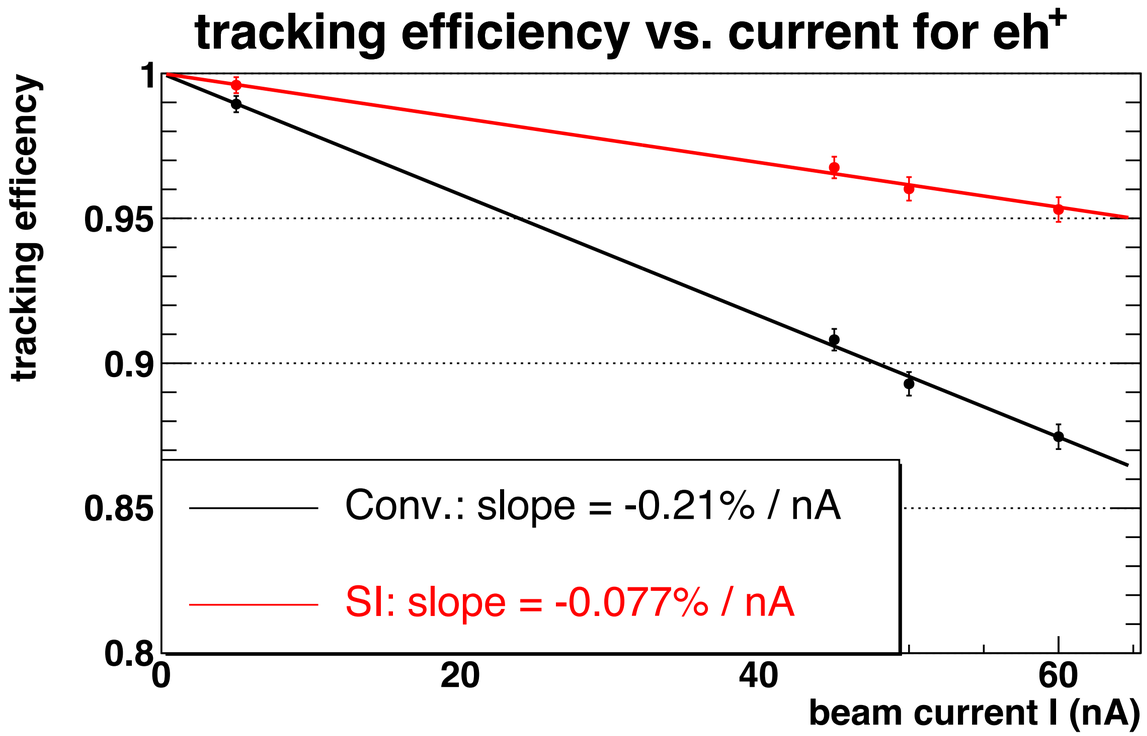
- **Global hit-pattern learning with SI:** Instead of sequentially propagating the track, SI captures global correlations among detector hits and learns a direct mapping from the hit pattern to the track state. In doing so, it can implicitly account for effects such as left–right ambiguity, magnetic-field inhomogeneity, and DC-cell geometry.

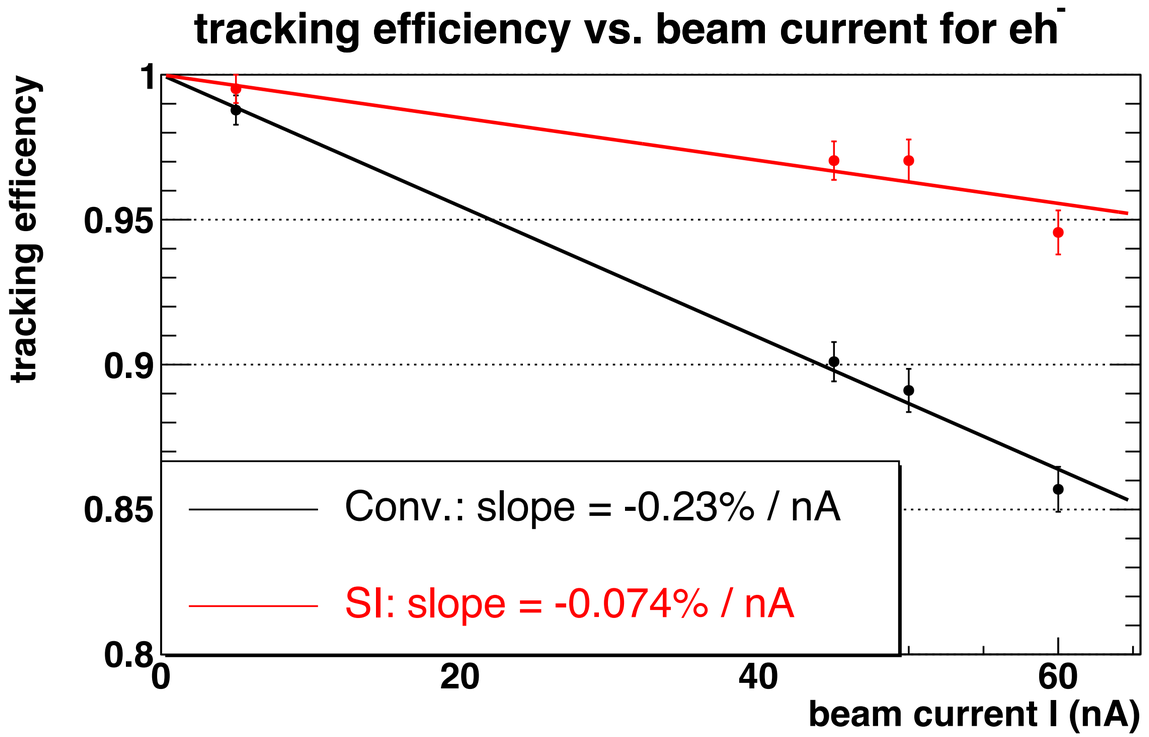
- **Improved HB tracking resolution:** The SI-based approach achieves better HB track-state resolution than the KF-based approach.

- **Training with experimental data:** To avoid discrepancies between simulation and experiment, the model is trained using experimental data, with TB tracks reconstructed by KF serving as the reference targets.

- **TB tracking:** KF-based TB tracking is not replaced by the SI model. Instead, the SI-predicted HB track state serves as the initial state for TB tracking. KF, working in conjunction with the Deterministic Annealing Filter (DAF) to significantly mitigate effects of noise hits, refines the trajectory using accurate TB hit measurements, ultimately providing the final high-precision track parameters.

Tracking Efficiency for RGA Inbending





Tracking Efficiency across Run Groups

- Tracking efficiency reduction rate for inbending runs:

Reduction rate (per nA)		Conv.	SI
RGA	eh^+	0.21%	0.077%
	eh^-	0.23%	0.074%
RGB	eh^+	-	0.11%
	eh^-	-	0.10%
RGD	eh^+	-	0.079%
	eh^-	-	0.087%

Tracking efficiency at a given beam current = 100% - reduction rate * beam current

- Tracking efficiency reduction rate for outbending runs is smaller than inbending runs.
- With comprehensive updates for forward tracking, tracking efficiency reduction rate is improved to be better than 0.1% per nA for RGA and RGD in-bending, and close to 0.1% per nA for RGB in-bending, whose tracking efficiency is relatively worse since forward tagger is on.

Physics Event Yields

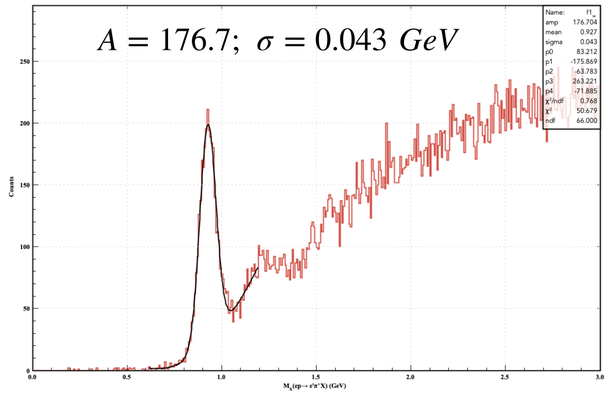
- Events of are selected with cuts for all final-state particles: V_z [-15, 5] cm, p > 0.5 GeV and |chi2pid| < 3.
- Missing mass distributions are fitted to extract events yields.

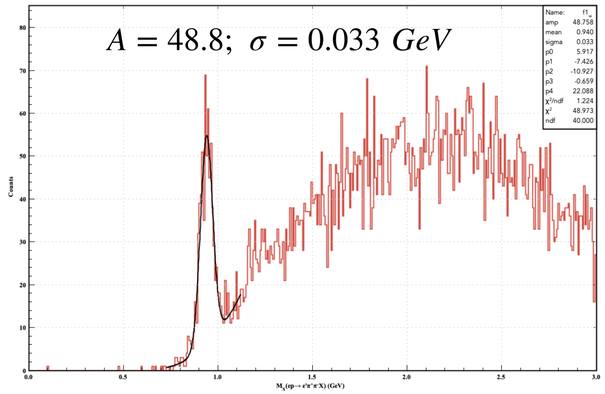
Run 5342 (50 nA)

$ep \rightarrow e'\pi^+X$

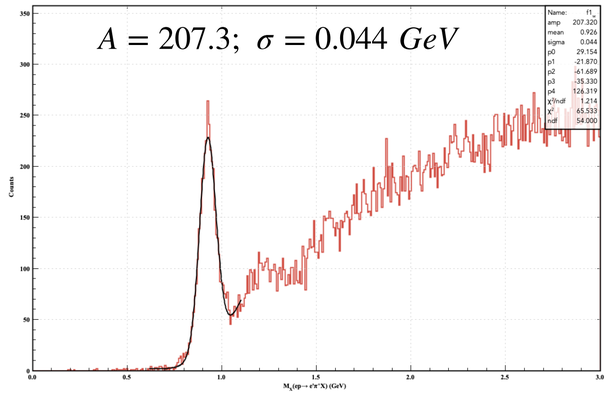
$ep \rightarrow e'\pi^+\pi^-X$

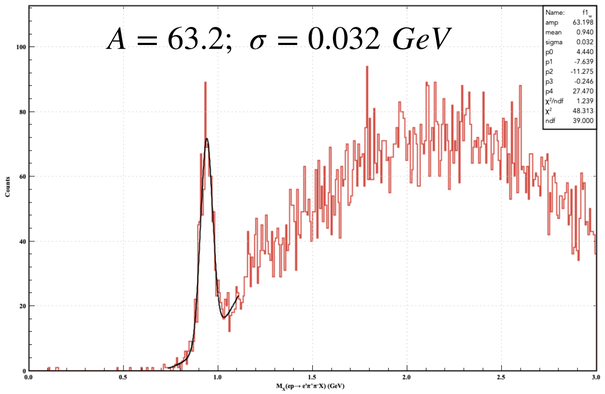
Conv.



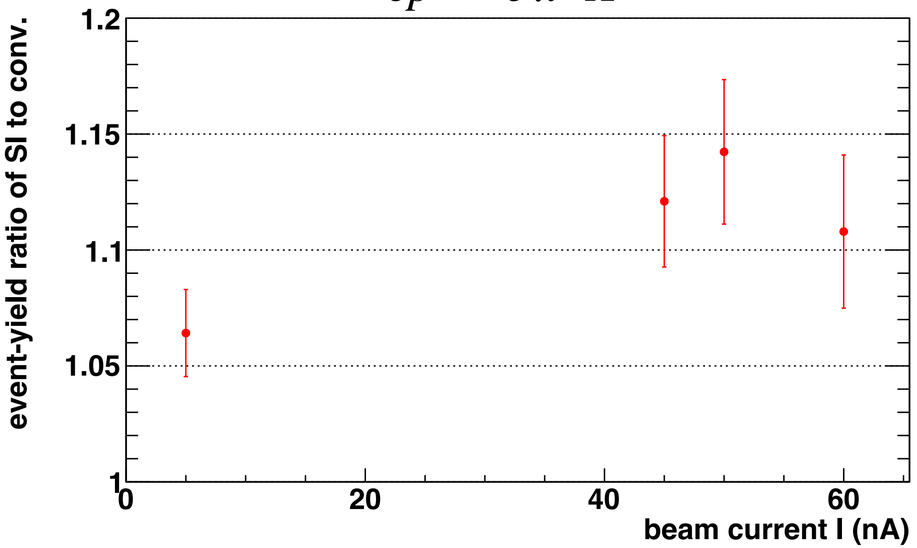


SI

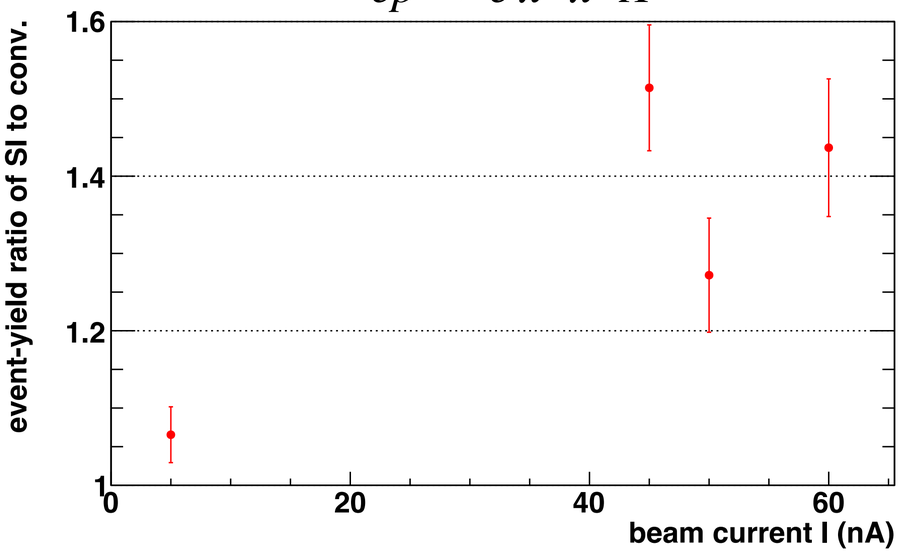




$ep \rightarrow e'\pi^+X$



$ep \rightarrow e'\pi^+\pi^-X$



Summary

- Traditional reconstruction algorithms often struggle to capture global correlations among hits or clusters along particle trajectories, particularly in the presence of inhomogeneous magnetic field and high background conditions. In contrast, SI-based models are capable of learning these global spatial correlations, thereby providing effective assistance to track reconstruction and significantly reducing background impact.

- At different stages of the reconstruction pipeline, appropriate SI techniques are selected. Model architectures and thresholds are optimized to maximize overall reconstruction performance.

- For the SI-enhanced reconstruction, the tracking efficiency reduction rate is below 0.1% per nA for RGA and RGD in-bending and approximately 0.1% per nA for RGB in-bending, with smaller losses for out-bending.

- The SI models significantly accelerate reconstruction. The processing time for the forward reconstruction is reduced by approximately 42% for production runs.

- Advantages of SI become more pronounced as luminosity increases, where increased detector occupancy and background noise make conventional reconstruction more challenging.

SI-Based Recoil Tracking

Challenges for Recoil Tracking

- μRWell trackers are located in the region with weak magnetic field, and therefore provide only the momentum direction.
- The target is in the region with magnetic field, making it difficult to reconstruct the particle trajectory between the vertex and the recoil trackers.
- Consequently, estimating the path length from the vertex to the time-of-flight (TOF) detector is challenging, limiting accuracy of momentum-magnitude estimation using TOF information.
- These challenges limit the accuracy of momentum estimation at the vertex.

SI-based Solution

- Use a SI technique to combine position measurements and time information to estimate the momentum at the vertex, implicitly learning distribution of the magnetic field.

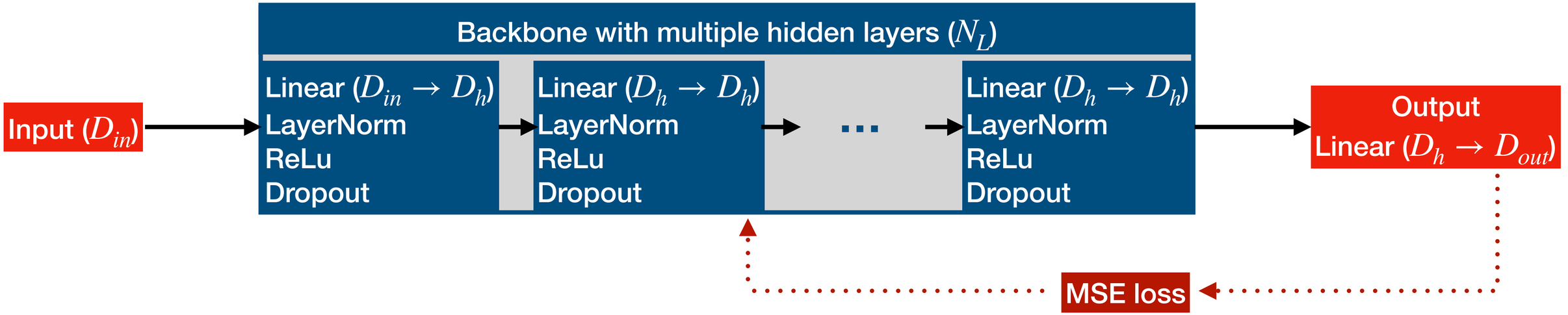
- Developed an estimator based on the MLP technique.

  - Input features: point (x, y, z) of 3 crosses and TOF from reconstruction

  - Predicted output: momentum (p, θ, φ) at vertex

  - Referenced target: momentum (p, θ, φ) at vertex from MC truth

MLP-Based Architecture

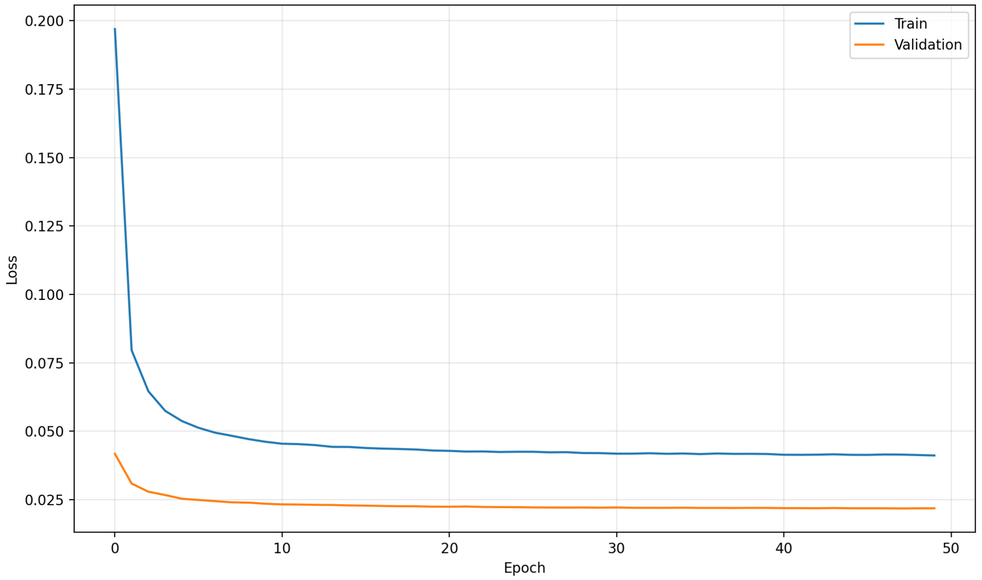


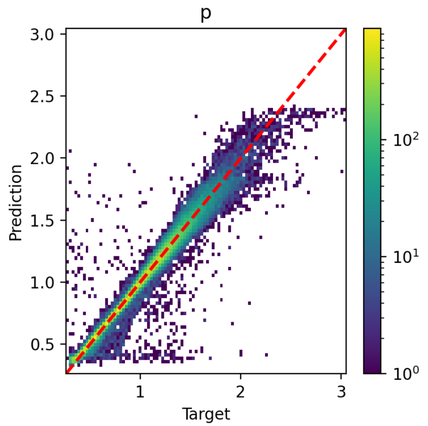
- Each hidden layer in the backbone consists of:
  - Linear: A fully connected layer to transform a set of numbers into a new set of numbers by multiplying them with learnable weights.
  - LayerNorm: Normalize the linear layer's output to make training more stable
  - ReLu: Activation functions that introduce nonlinearity into the neural network (NN)
  - Dropout: A regularization technique to reduce overfitting in NN

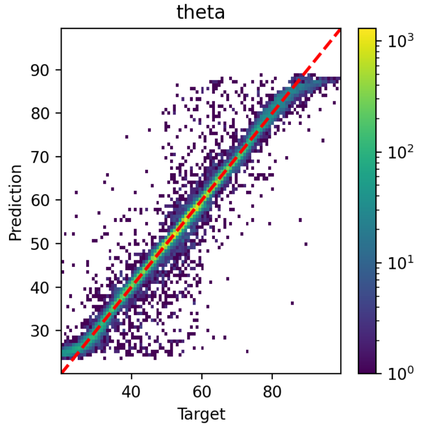
- How NN weights are updated for each input batch of training loop?
  1. Forward pass: Input features → MLP → Prediction
  2. Loss calculation: Calculate MSE between prediction and target
  3. Back-propagation: Calculate gradients of the loss with respect to trainable weights
  4. Weight update: Optimizer (Adam) updates weights using the gradients

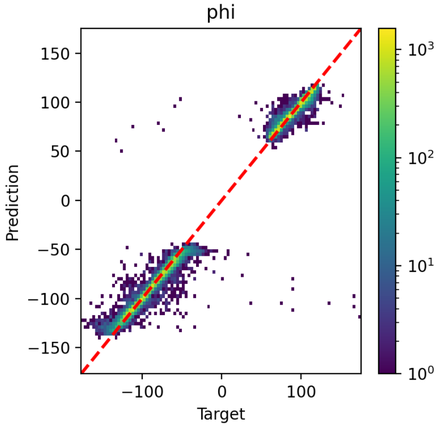
Model Performance

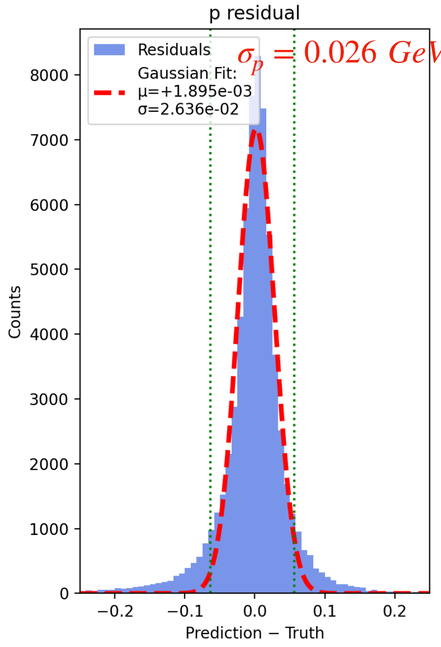
Model with $N_L = 4$ and $D_h = 64$



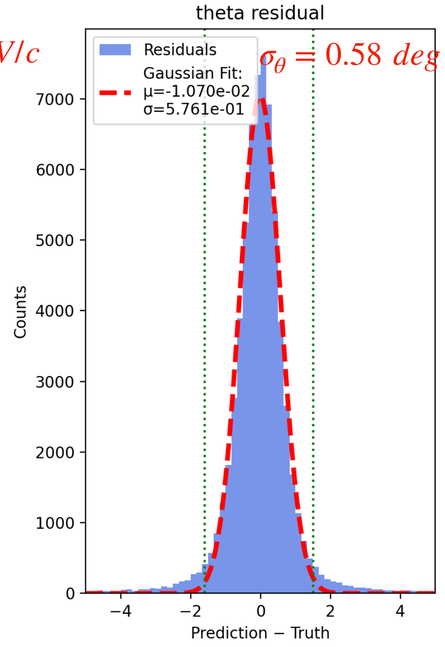




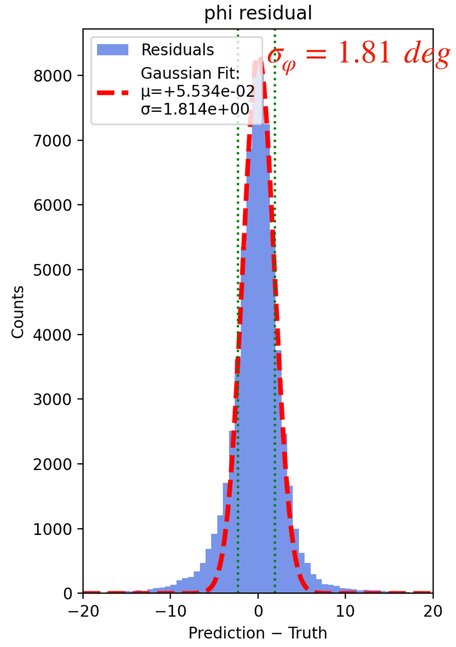




$\sigma_p = 0.026\ GeV/c$

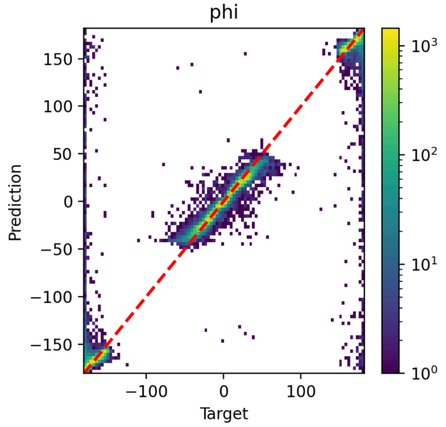


$\sigma_\theta = 0.58\ deg$



$\sigma_\varphi = 1.81\ deg$

φ is rotated by 90^0 to avoid performance degradation near the $\pm 180^0$ boundaries.



SI Software and Model Deployment

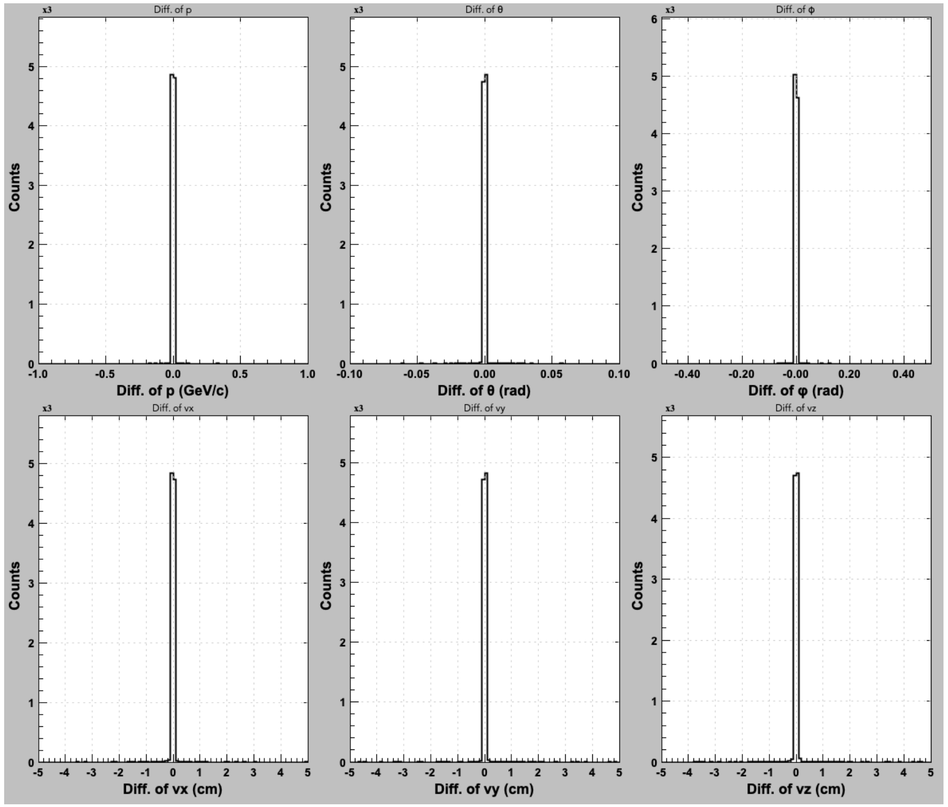
- The SI software package is designed to be easily extended to support additional utilities and functionalities with minor modifications:
  - Train a model using only two crosses and TOF.
  - Add a particle-type label (1, 2, 3, …) into training data, and train a model to predict both track state and particle type simultaneously.
  - Modify output and loss function to train a dedicated classifier specifically for particle identification.
  - Others…
- The package provides flexible options for model configuration and train settings.
- A trained model can be deployed in coatjava using the Deep Java Library (DJL).

Summary

- A MLP-based model has been developed to estimate track state for recoiled particles.
- The lightweight MLP architecture enables fast inference, making it well suited for integration into the reconstruction pipeline.
- The AI software package could be easily extended to support additional utilities and functionalities with minor modifications.

Backup Slides

Comparison of TB Tracks between Different HB Tracks as Initial State of TB Tracking



- TB KF results remain consistent, although different HB track states are used for initialization.
- Therefore, impact on TB tracking resolution is negligible when the hit resolution definition remains unchanged.
- However, changes in HB track states affect TB hit resolution. Consequently, TB tracking results are influenced when the TB hit resolution is updated.

Discuss for MLP and Transformer Techniques

	MLP	Transformer
Core mechanism	Fully connected NN layers	Self-attention + feed-forward NN layers
Input representation	Fixed-size vector	Sequence of tokens
Cross-layer relationships	Learned implicitly	Modeled explicitly through attention
Complexity and computational cost	Usually lower	Usually higher

MLP NN

$$y_i = b_{i0} + \sum_j c_{ij} ReLu(b_{ij} + \sum_k w_{ijk} x_k)$$