SI-Enhanced Reconstruction for the CLAS12 Central Tracking System

Tongtong Cao (Jefferson Lab)

CLAS12 Collaboration Meeting
Nov 17 - 20, 2026

Recent Software Updates for Conventional Reconstruction

- Sign of d_0 in seed helix parameters (PR#1519): merged; small impact
- Position and momentum calculation at a fixed radius based on seed/track helix trajectory (PR#1519): merged; small impact
- SVT sectors separation into 3 sections for seeding (PR#1517): merged; small impact
- Construction of cross combinations for seeding (PR#1597): under review; moderate impact
- Projector and H matrix in KF tracking (PR#1460): Yuri identified the issue, updated codes, and performed tests that showed a significant impact on measurement residuals. Raffaella is continuing to evaluate its impact on tracking results.
- Code optimization by Nathan to improve computational speed and reduce memory usage.

Challenges for Conventional Reconstruction



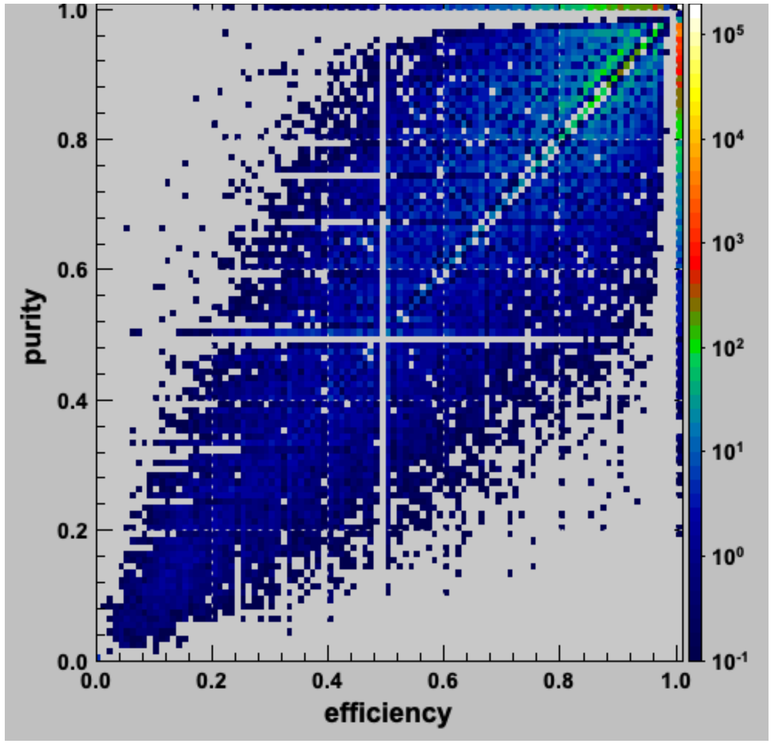
- **Impact of background:** Background affects every stage of the reconstruction pipeline, leading to track losses, increased fake or contaminated tracks, and reduced computational efficiency.

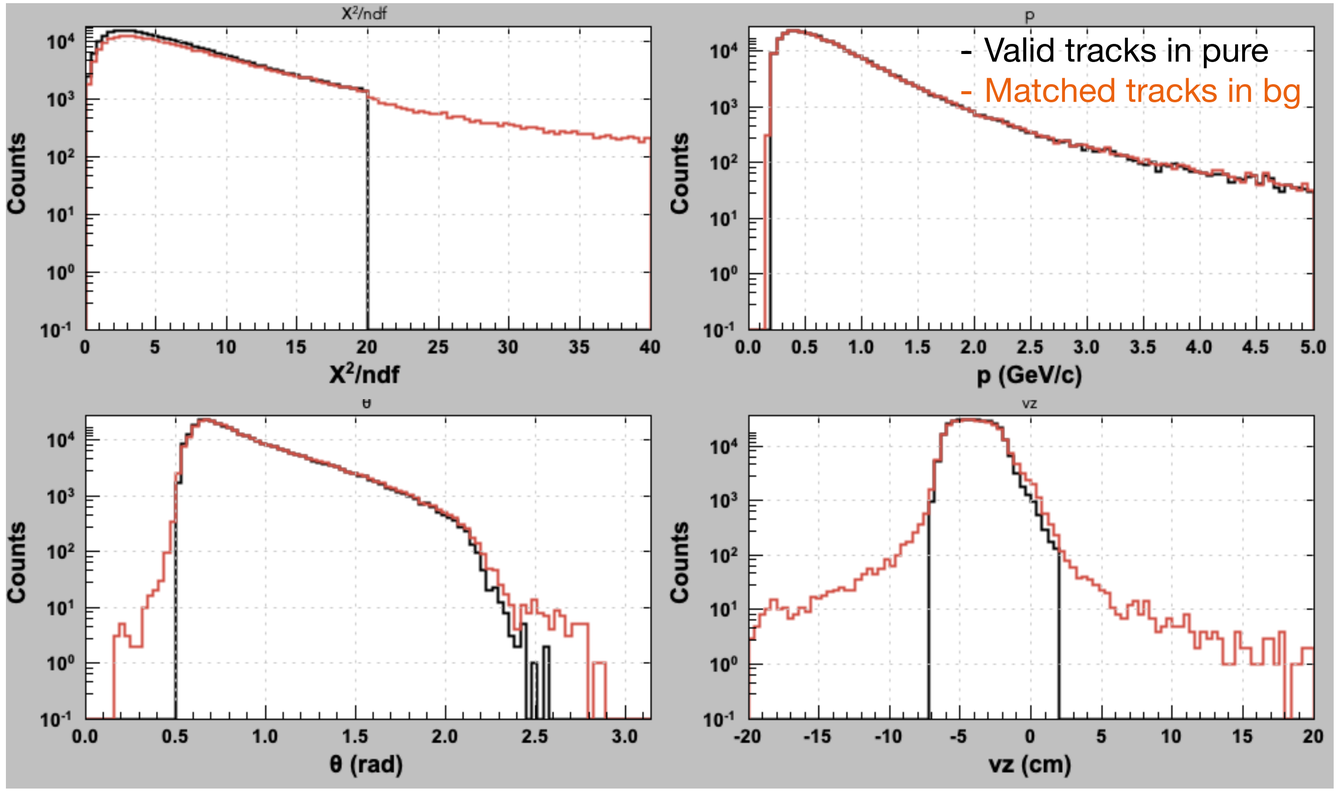
- **Limitation of conventional algorithms:** Conventional algorithms heavily rely on detector geometry, heuristic methods, and pattern-recognition procedures, making background suppression and robust reconstruction increasingly challenging.

- **Challenge in track finding:** Current algorithms for track finding entirely rely on helix fitting. However, the helix fits are very rough, with significant biases in some helix parameters.

Investigation for Background Effects

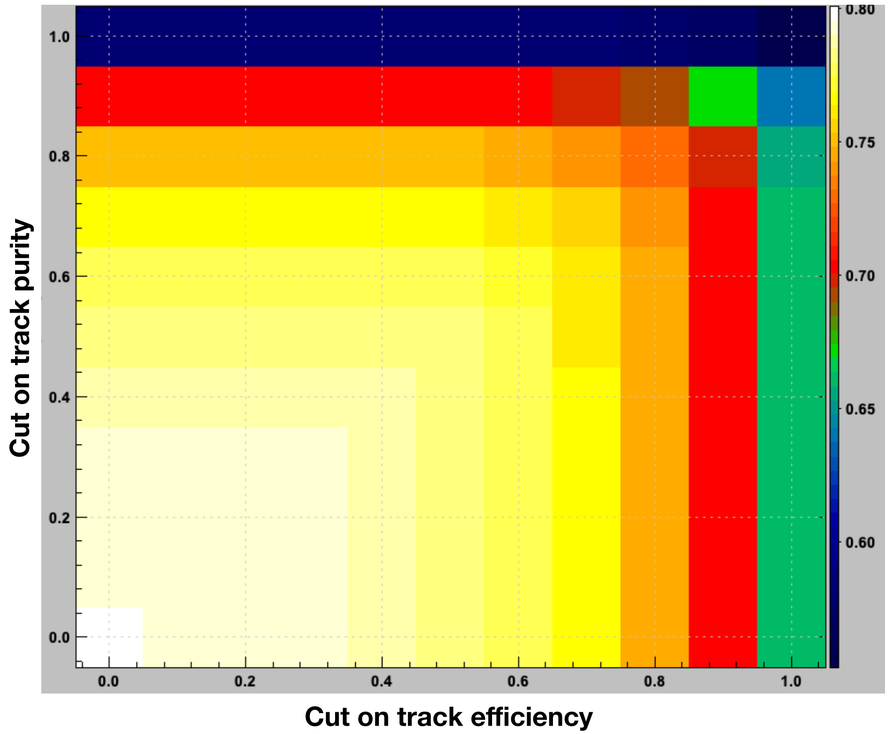
- **MC samples:** Pure & 50 nA-background-merged RGA-SIDIS
- **Study strategy:** Perform event-by-event track matching. For each valid track in the pure samples, identify the best-matched track in the background samples based on shared hits, with priority given to main-strip matches, followed by cluster purity and efficiency.
- **Valid cuts on tracks in the pure samples:** $\chi^2/NDF < 20; p > 0.2\ GeV/c; -7\ cm < V_z < 2\ cm; \theta > 30^0$
  - **Track purity:** signal hits / all hits for a track in bg samples
  - **Track efficiency:** signal hits for a track in bg samples / all signal hits for the matched track in pure samples





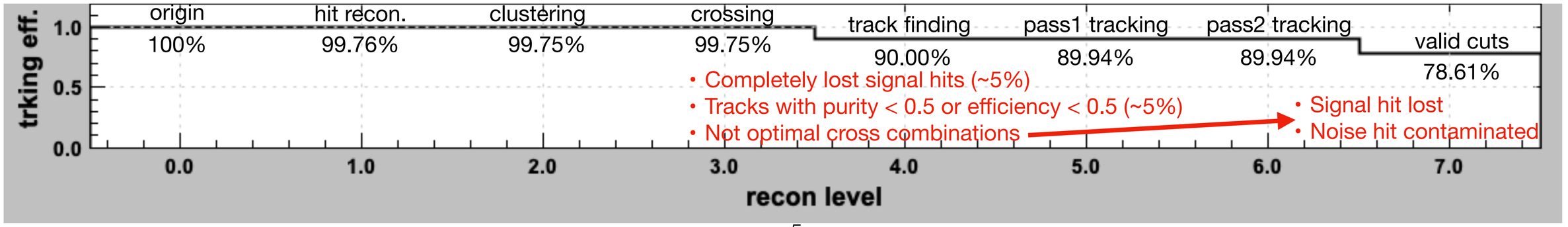
Tracking Efficiency

Tracking efficiency vs. purity & efficiency cuts



- ~5% of valid tracks in pure samples have no matched tracks in bg samples.
- ~15% of matched tracks in bg samples are eliminated with the same valid cuts for tracks in pure samples.
- Therefore, the tracking efficiency starts from ~80% without purity and efficiency cuts on valid matched tracks.

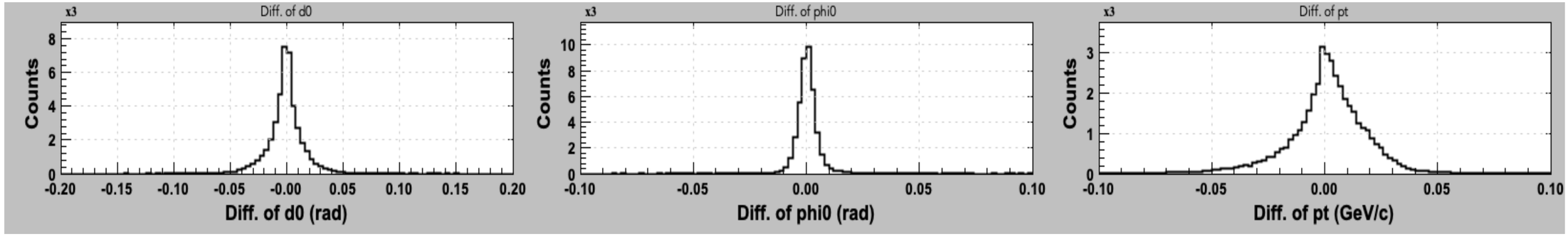
Background effects on each stage in the reconstruction pipeline for valid matched tracks with purity >= 0.5 and efficiency >= 0.5



Issue for Helix Fitting: Circular Fitting

- Algorithms for track finding entirely rely on helix fitting.
- Assuming uniform magnetic field and no energy loss, helix fitting consists of two independent components:
  - Circular fitting: Determine 3 transverse helix parameters in the x-y plane
  - Linear fitting: Determine 2 longitudinal helix parameters in the z-r plane

Differences for d_0, ϕ_0 and p_t between seeds and tracks, where the parameters are obtained from circular fits for seeds and from KF fits for tracks.



Energy loss causes bias of p_t

Issue for Helix Fitting: Linear Fitting

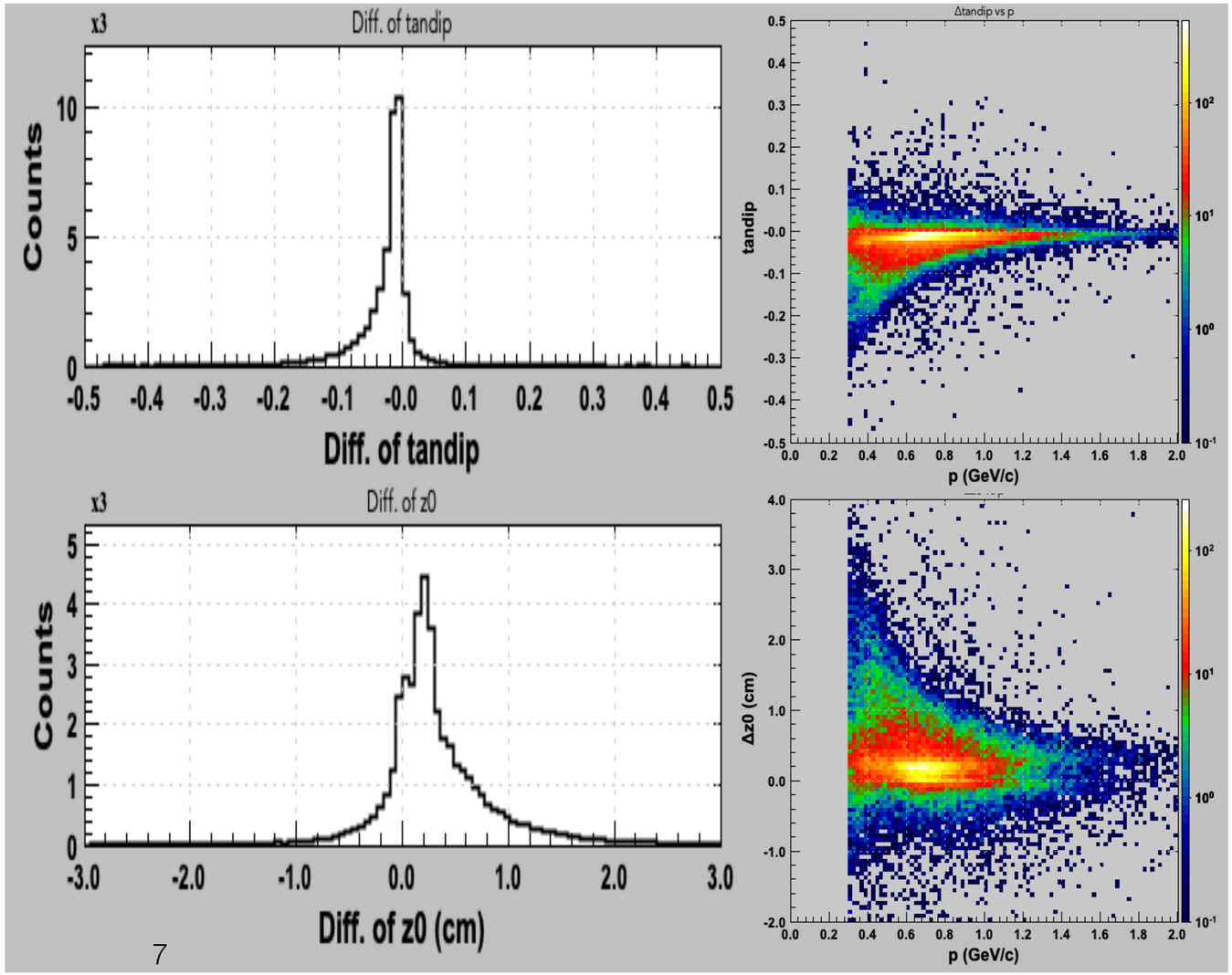
$$z = z_0 - \frac{1}{\Omega}\Delta\phi \tan\lambda$$

$$x = x_b - d_0 \sin\phi_0 - \frac{1}{\Omega}[\sin(\phi_0 + \Delta\phi) - sin\phi_0]$$

$$y = y_b + d_0 \cos\phi_0 + \frac{1}{\Omega}[\cos(\phi_0 + \Delta\phi) - cos\phi_0]$$

$$r = \sqrt{x^2 + y^2}$$

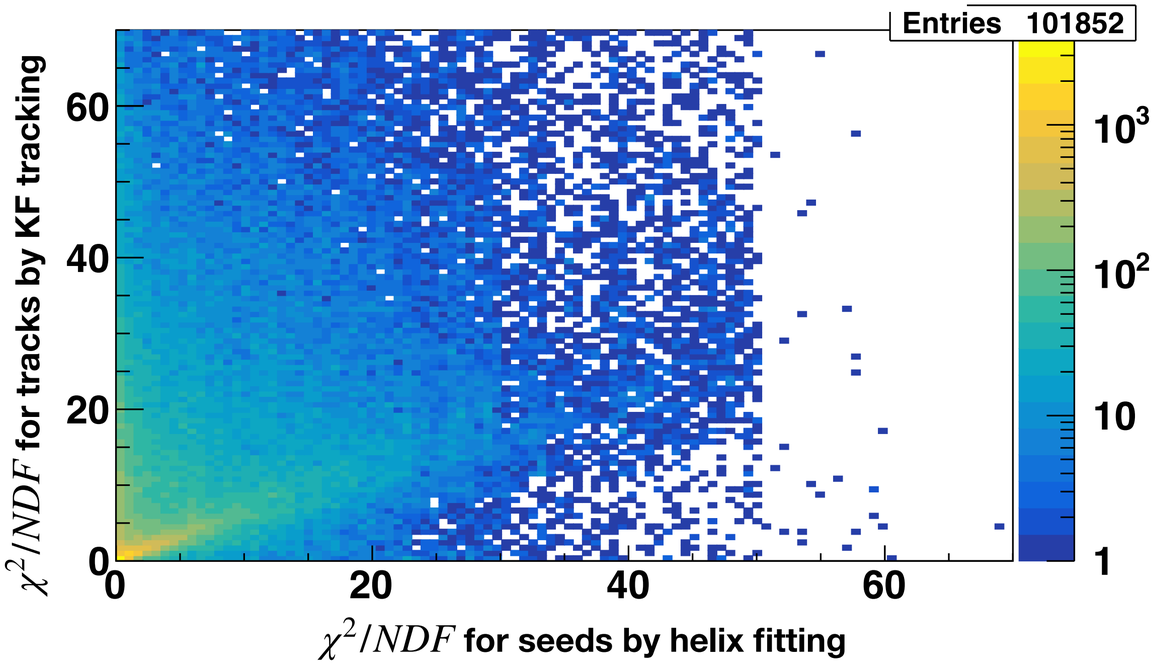
Differences for $\tan\lambda$ and z_0 between seeds and tracks, where the parameters are obtained from linear fits for seeds and from KF fits for tracks.



- Analytically, the *z-r* trajectory is not linear.
- Left pictures show that linear fits to the *z-r* trajectory introduce significant biases in the extracted parameters.
- Right pictures show that these biases become more significant at lower momentum, since curvature in the *z-r* plane is larger.

Issue for Selection of Overlapping Seeds

- When multiple seeds share clusters, a dedicated routine is used to select one based on their helix fits:
  - 1st priority: larger NDF is preferred.
  - 2nd priority: smaller chi2 is preferred.



- MC samples: 0.3 - 1.6 GeV π^+ single MC with background
- All seeds are passed to the tracking process, without applying the overlapping-track-candidate selection.
- χ^2/NDF values, separately from helix fit for seeds and KF fit for tracks, are compared. The two values do not show a direct linear relationship.
- This indicates that the selection routine for overlapping seeds based on the rough helix fit is not reliable.

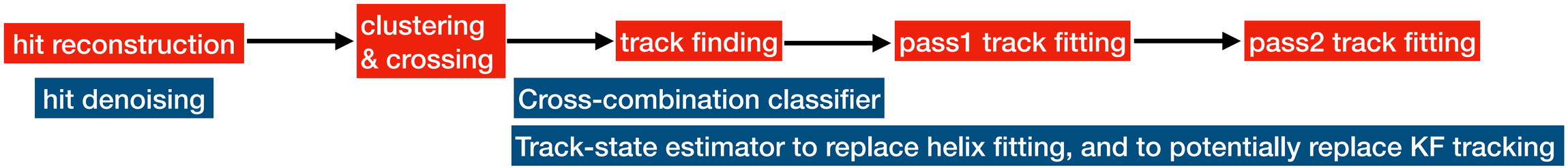
Discussion for Track Finding

- **Limitations of conventional fitting:** The simplified linear and circular fitting approaches struggle to account for complex effects, including energy loss and non-uniform magnetic fields.

- **Cross-combination construction:** Inaccurate fits can reject valid track combinations, and increase fake or contaminated combinations.

- **Overlapping track-candidate selection:** Selection among overlapping seeds based on helix-fitting results can lead to unreliable decision and incorrect track selection.

SI-Enhanced Reconstruction



- **Opportunity with SI:** SI can learn complex patterns in detector measurements and their relationships to particle trajectories, implicitly accounting for effects, such as energy loss, multi-scattering, non-uniform magnetic field, etc.
- **SI-enhanced reconstruction:** Following experience on reconstruction of the forward tracking system, SI models will be applied at three stages in the reconstruction pipeline of the central tracking system:
  - Hit denoiser: suppress noise hits
  - Cross-combination classifier: identify cross combinations as seeds for tracking
  - Track-state estimator: estimate track-state parameters
- **Expected impact:** Improve reconstruction efficiency, track quality and physics performance, while keeping the computational cost acceptable.

Extension of SI Models from Forward to Central Tracking System?

	Forward System	Central System
Detector	Drift chambers	SVT + BMTC + BMTZ
Geometry	In-parallel layers	Complex geometry with different detector types and orientations
Measurements	DOCA for DC hits	x,y,z for SVT; z for BMTC; ϕ for BMTZ
Track topologies	6-cluster and 5-cluster tracks	Arbitrary combinations of crosses, with at least 3 total crosses and at least 1 SVT cross

Due to differences of geometry, measurement representations and track topologies, some SI techniques used for the forward tracking system are not suitable for the central tracking system. Others can be extended, but neural-network architectures need to be re-designed.

Hit Denoising

- **Motivation:** Hit denoising can improve the performance and reduce computational burden in the subsequent reconstruction stages.

- **Model development:**

  - The convolutional-neural-network based auto-encoder for hit denoising of the forward tracking system is not suitable for the central tracking system due to differences in detector geometry and hit representations.

  - A GravNet/GarNet-based hit-denoising model is under development.

- **Current status:** A model trained on samples from MC particles with background has been developed and integrated into coatjava. Validation studies and SI architecture optimization are ongoing.

- **Next steps:** Develop and validate models trained on MC physics events with background and on experimental data.

- **Mathieu Ronayette will present details next.**

Plan for SI-Based Track Finding

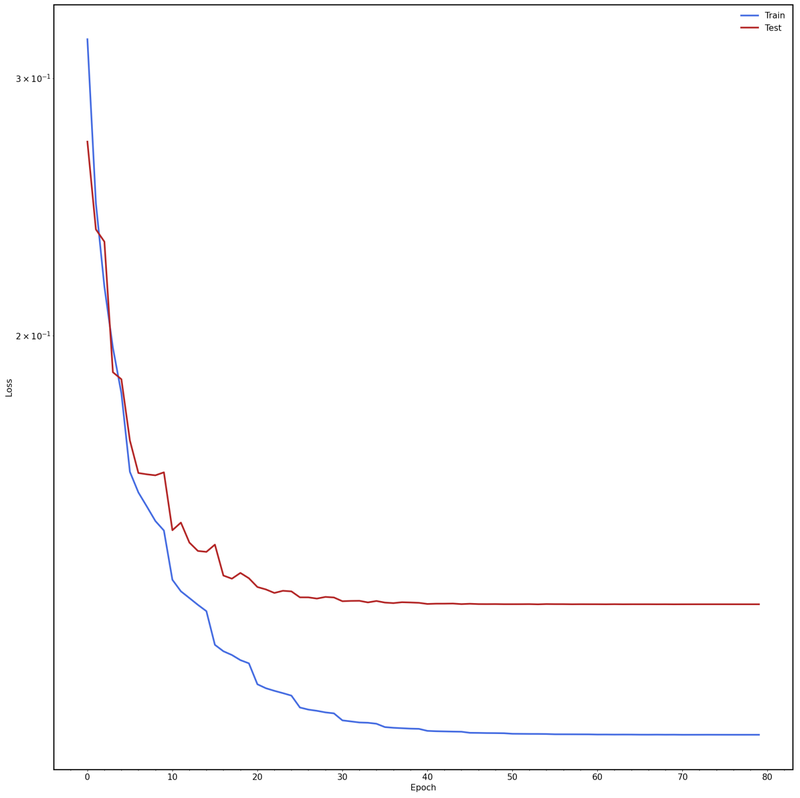
- Develop a new algorithm to construct cross combinations
  - Abandon algorithms, which rely on linear and circular fitting.
  - Explore effective pre-selection criteria to limit combinations so as to control computational cost.
- Develop a Multiple Layer Perceptron (MLP)/transformer-based binary classifier to distinguish real from fake cross combinations.

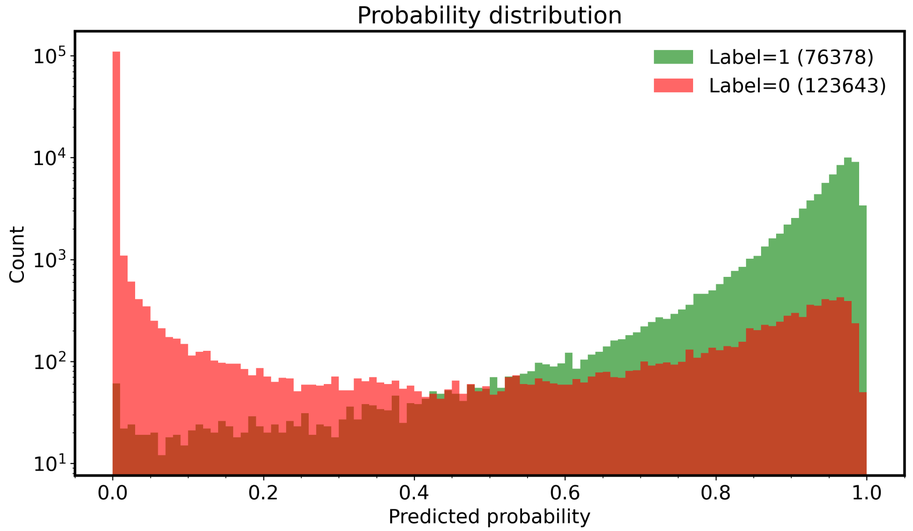
MLP-Based Track Finding

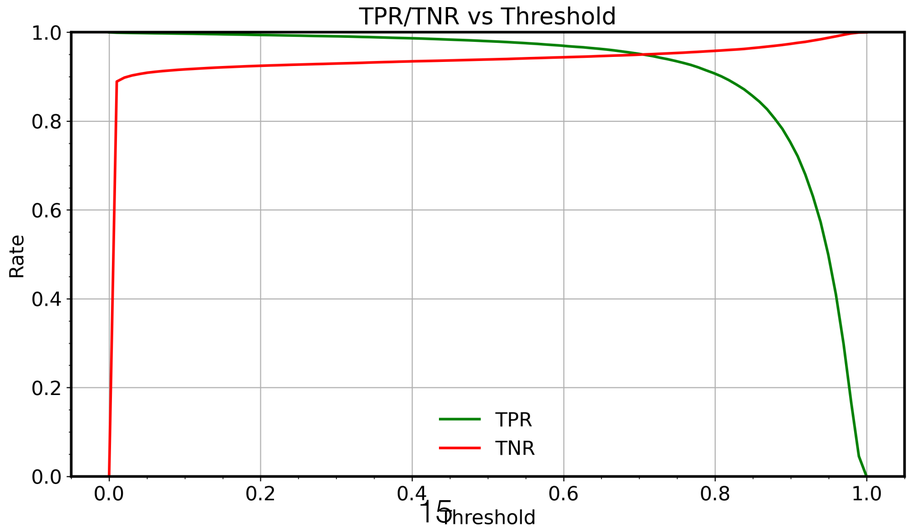
- **Model:**
  - Extension of the successful MLP-based models for track finding of the forward tracking system, an MLP binary classifier is trained to distinguish real from fake cross combinations.
  - Unlike the forward system, the central system has more complex track topologies. Therefore, the classifier must accommodate all possible cross combinations.
- **Real sample and features:** Measurements from crosses in 9 layers along valid tracks from pass2 tracking:
  - 3 SVT crosses: x, y, z, layer (= 1/2/3), mask (1/0)
  - 3 BMTC crosses: r, z, layer (= 4/7/9), mask (1/0)
  - 3 BMTZ crosses: r, φ, layer (= 5/6/8), mask (1/0)
- **Fake-sample construction:** Replace a cross in a randomly selected layer of a real combination with a randomly selected cross from the same layer.
- **Missing-layer handling:** Each detector layer has a fixed input slot. If a layer is missing, its measurement features are set to zero, while its layer identifier retains its actual value. A binary mask (1/0) indicates whether a cross is present.
- **Training samples:** Extracted from RGK run 19331 with 65nA beam. The track selection criteria are: $\chi^2/NDF < 10; p > 0.2\ GeV/c;$ $-7\ cm < V_z < 2\ cm; \theta > 30^0$

Model Performance

MLP-based Model with 3 hidden layers and 512 dimensions



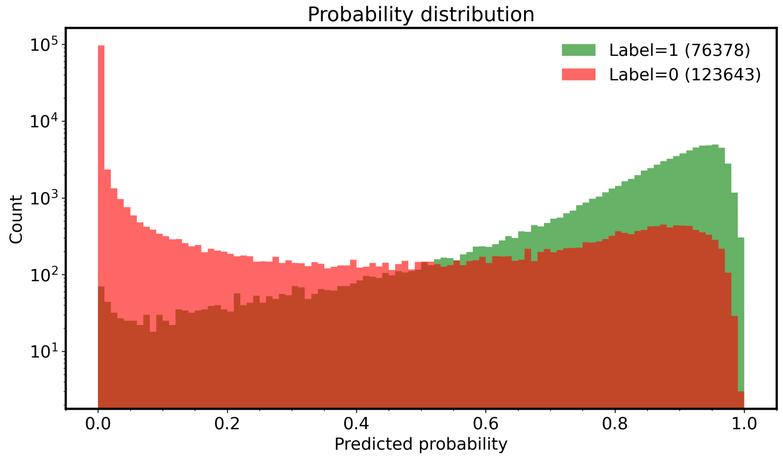


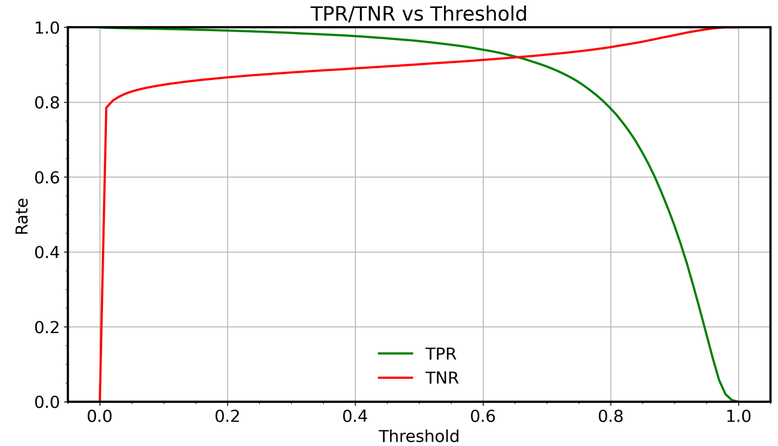


- Appropriate constraints to the random crosses are needed for construction of fake samples.
- Real samples contain part of noise/ contaminated fake tracks.
- Therefore, training samples need to be optimized with better reconstruction and further investigation for better track selection criteria.

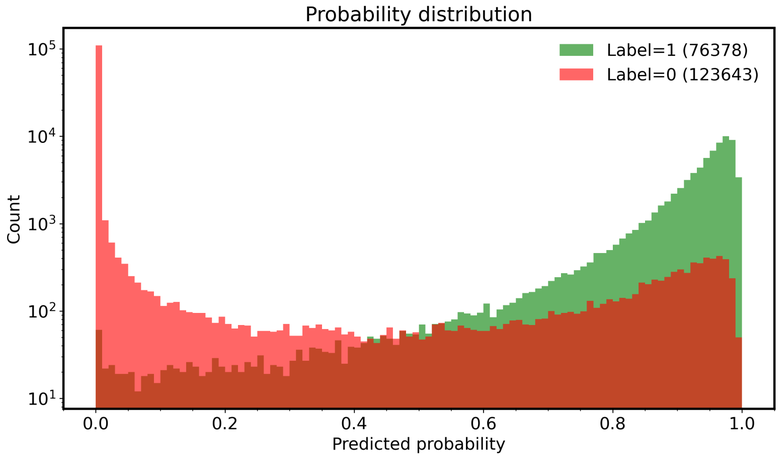
Model Performance with Different Hyper-Parameter Setting

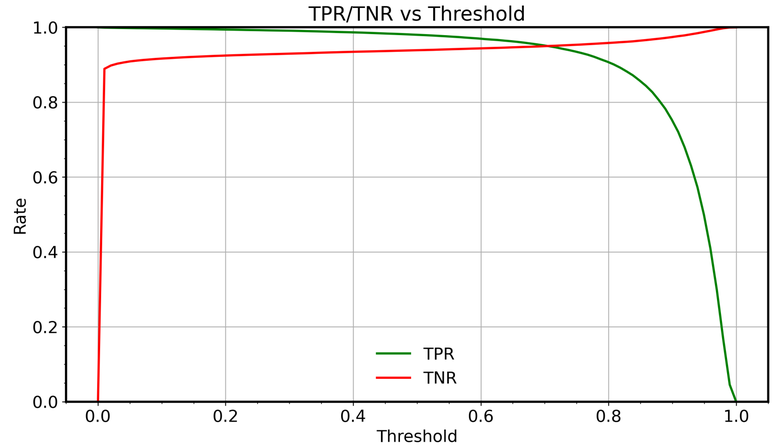
3L; 128D



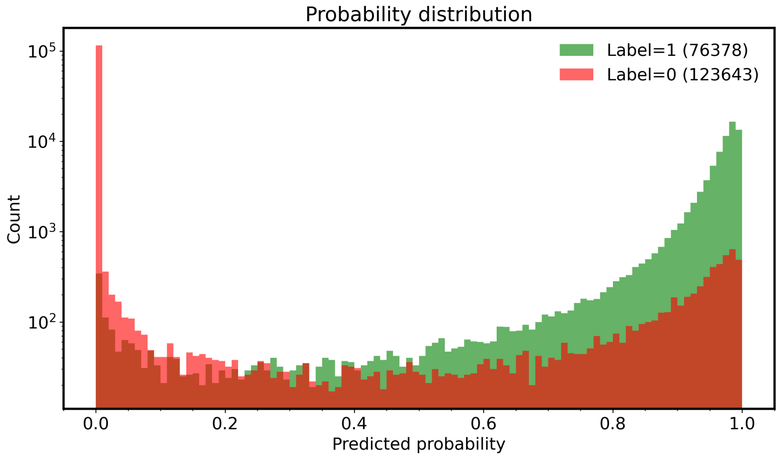


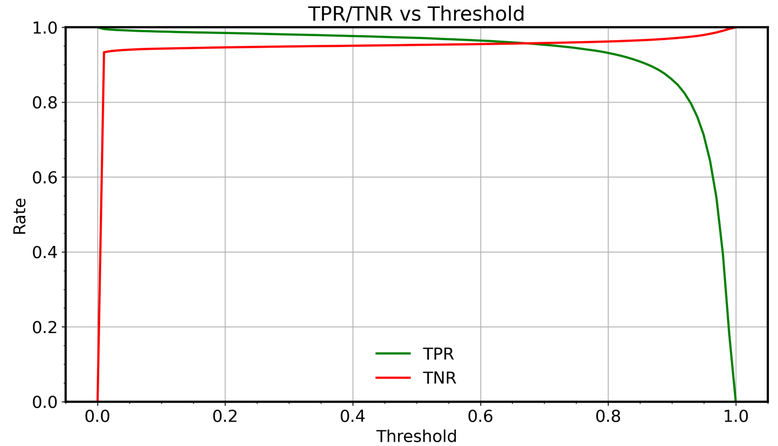
3L; 512D





7L; 512D





- More complex neural-network models show better performance.
- Further reconstruction tests are needed to determine the final model, balancing reconstruction performance and computational cost.

Plan for SI-Based Track Fitting

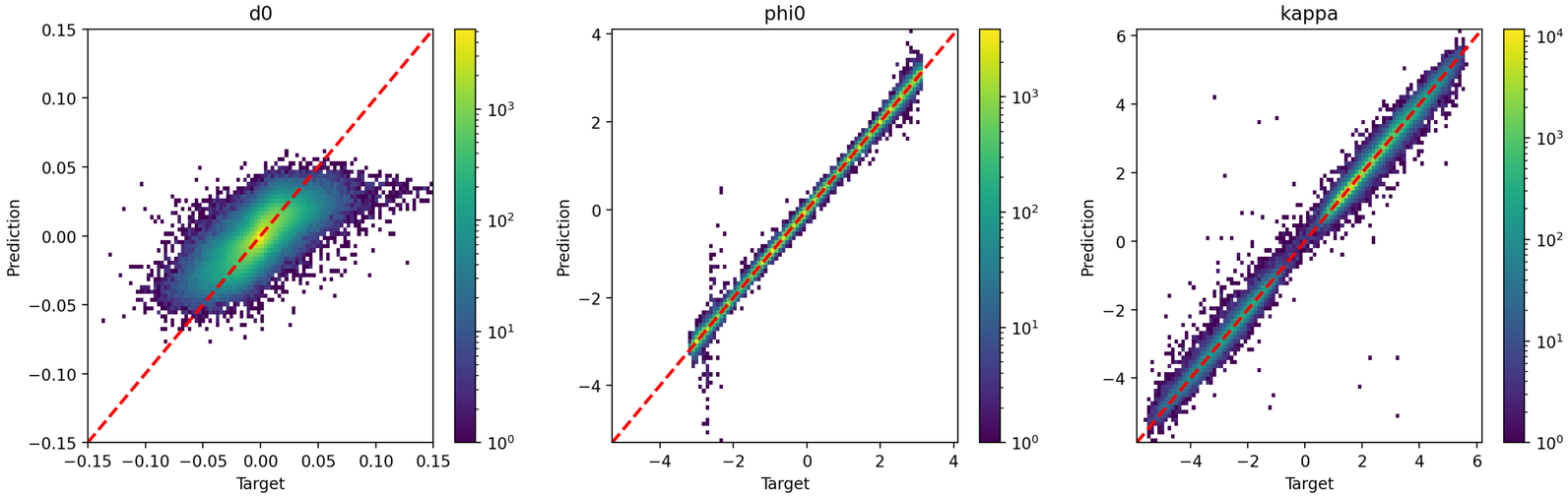
- Develop a MLP/transformer-based model for track-state estimation.
- The SI model will assist selection of overlapped seeds.
- Depending on the performance, the SI model could potentially replace the pass1 KF tracking.
- For training samples from experimental data, target is pass2 track from KF tracking. Therefore, the SI model, trained by experimental data, will not replace pass2 KF tracking.
- For training samples from simulation data, target could be truth. So the SI model, trained by simulation data, could potentially also replace pass2 KF tracking.

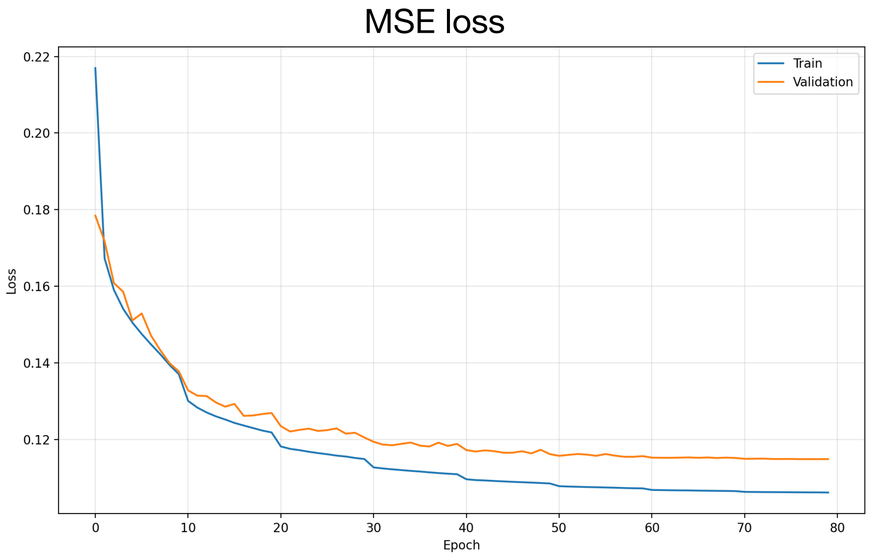
MLP-Based Track Fitting

- **Model:**
  - For the forward system, a transformer-based model was developed to estimate hist-based track states after track finding.
  - For the central system, a lightweight model is pursed to assist track finding. To minimize computational cost, an MLP-based model is explored firstly to estimate track state (d_0, ϕ_0, $\kappa = q/p_t$, z_0, $\tan\lambda$)
- **Features:** Measurements from crosses in 9 layers along valid tracks from pass2 tracking:
  - 3 SVT crosses: x, y, z, layer (= 1/2/3), mask (1/0)
  - 3 BMTC crosses: r, z, layer (= 4/7/9), mask (1/0)
  - 3 BMTZ crosses: r, φ, layer (= 5/6/8), mask (1/0)
- **Missing-layer handling:** Each detector layer has a fixed input slot. If a layer is missing, its measurement features are set to zero, while its layer identifier retains its actual value. A binary mask (1/0) indicates whether a cross is present.
- **Target:** Track state of corresponding tracks.
- **Training samples:** Extracted from RGK run 19209 with 10nA beam. The track selection criteria are: $\chi^2/NDF < 10; p > 0.2\ GeV/c;$ $-7\ cm < V_z < 2\ cm; \theta > 30^0$

Prediction Vs. Target

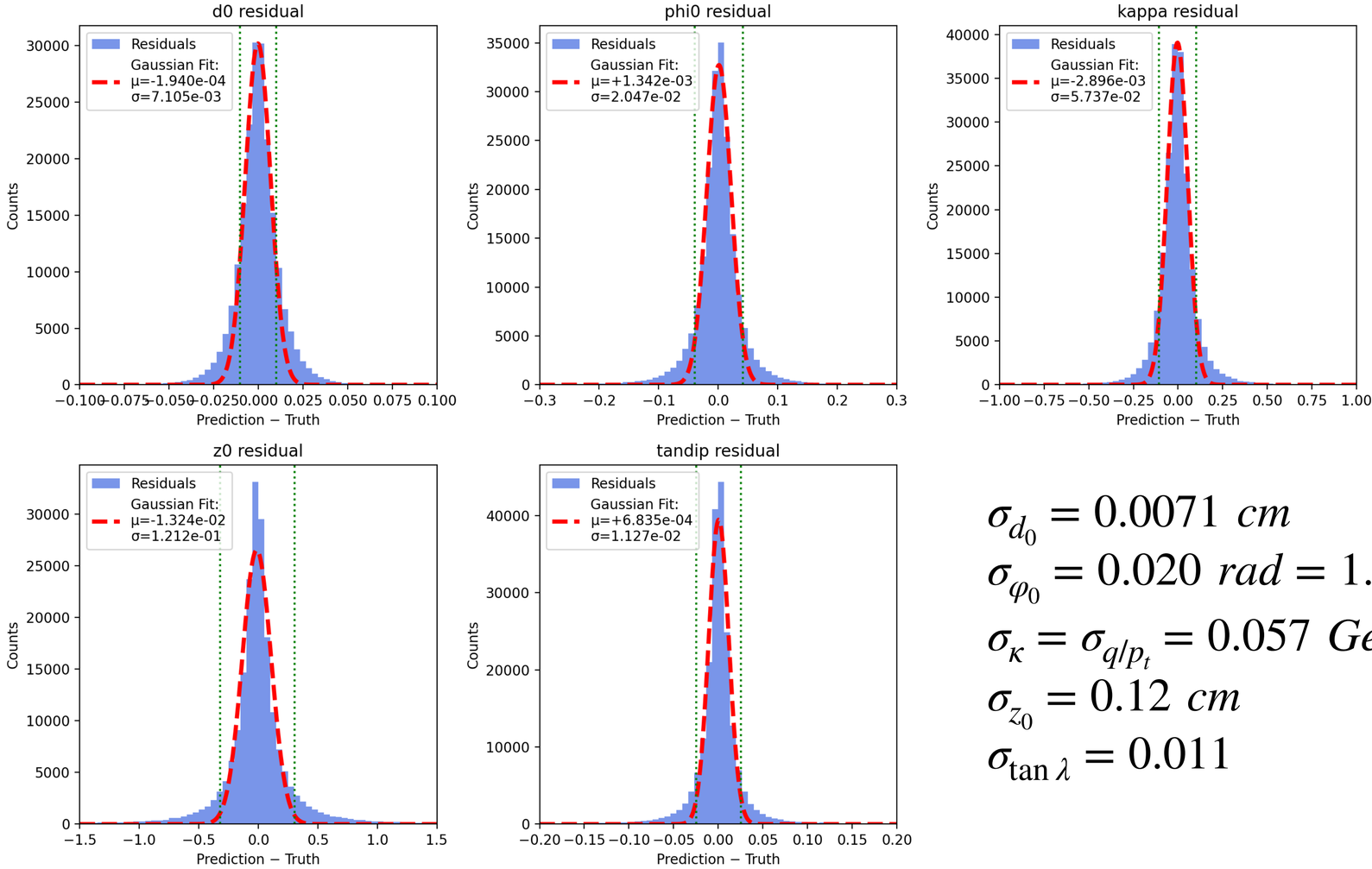
MLP-based Model with 3 hidden layers and 512 dimensions





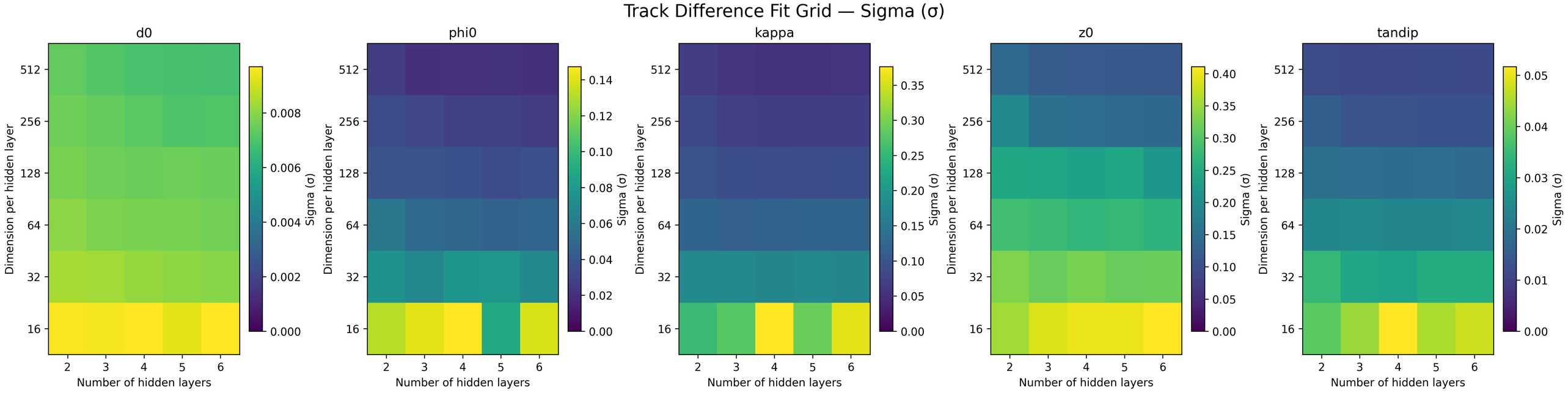
Difference and Fit

MLP-based Model with 3 hidden layers and 512 dimensions



$$\sigma_{d_0} = 0.0071\ cm$$
$$\sigma_{\varphi_0} = 0.020\ rad = 1.2\ deg$$
$$\sigma_{\kappa} = \sigma_{q/p_t} = 0.057\ GeV^{-1}$$
$$\sigma_{z_0} = 0.12\ cm$$
$$\sigma_{\tan\lambda} = 0.011$$

Hyper-Parameter Scan for Resolution of Parameters



- More complex neural-network models show better parameter-estimation quality.
- Further reconstruction tests are needed to determine the final model, balancing reconstruction performance and computational cost.

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

- Considering high occupancy and complex measurement topologies in CVT, MLP and transformer techniques offer complementary advantages and limitations.
- Both architectures will be explored for track finding and track fitting, and their performance will be compared.

Exploratory Approaches for Track Finding

- Graph Neural Network (GNN)-based model for cross-combination identification
  - Pre-processing: Use a GNN model to predict probabilities of connections between crosses.
  - Post-processing: Connect crosses and construct cross combinations based on the predicted probabilities.
- Transformer-based model for cross-to-track assignment: Assuming a maximum of N tracks per event in the central tracking system, develop a model to predict probability of each cross being associated with each of N tracks.

Workflow for SI-Enhanced Reconstruction

- **Training sample preparation:** Develop/update tools for extracting training samples separately from experimental and simulation data.
- **SI model development:** Develop/update SI software packages, and train models using samples from experimental and simulation data.
- **SI technique comparison:** Compare different SI techniques to identify the most suitable approaches for each reconstruction task.
- **Data- and simulation-based model comparison:** Compare models trained on experimental data and simulated data to evaluate their performance and determine whether a common model can be used despite differences in detector geometry.
- **Software integration and validation:** Integrate trained models into coatjava, redesign track-finding software, and perform validation to ensure reliable and consistent operation.
- **Performance evaluation:** Develop tools to quantitatively evaluate reconstruction and physics performance.
- **Model optimization:** Optimize SI architectures based on reconstruction and physics performance.
- **End-to-end pipeline optimization:** Optimize the SI-enhanced reconstruction pipeline by tuning model thresholds to balance signal acceptance and background rejection, and adjusting neural network complexity to balance reconstruction and physics performance against computational cost.
- **Iterative training and reconstruction:** Establish an iterative workflow in which improved reconstruction provides higher-quality training samples, while retrained and optimized models further improve reconstruction quality. Repeat these cycles to progressively improve overall performance.
- **Models across run groups:** Investigate model performance across different run groups to determine whether run-group-specific models are needed.

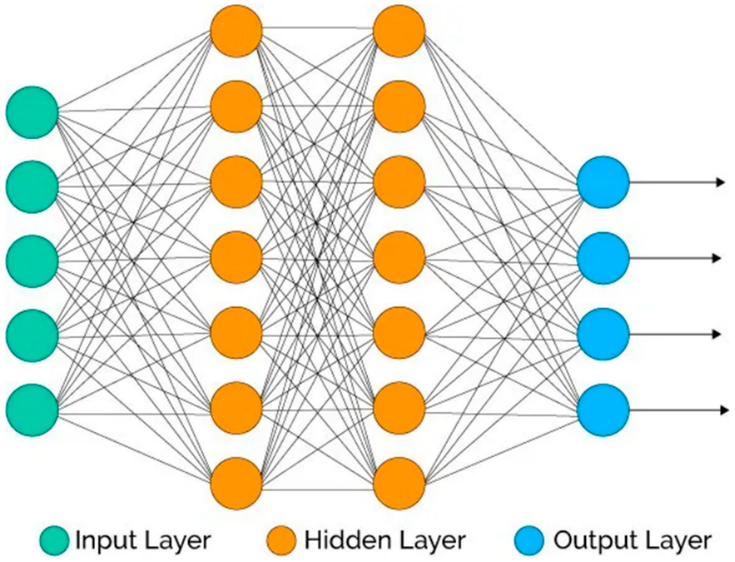
Summary

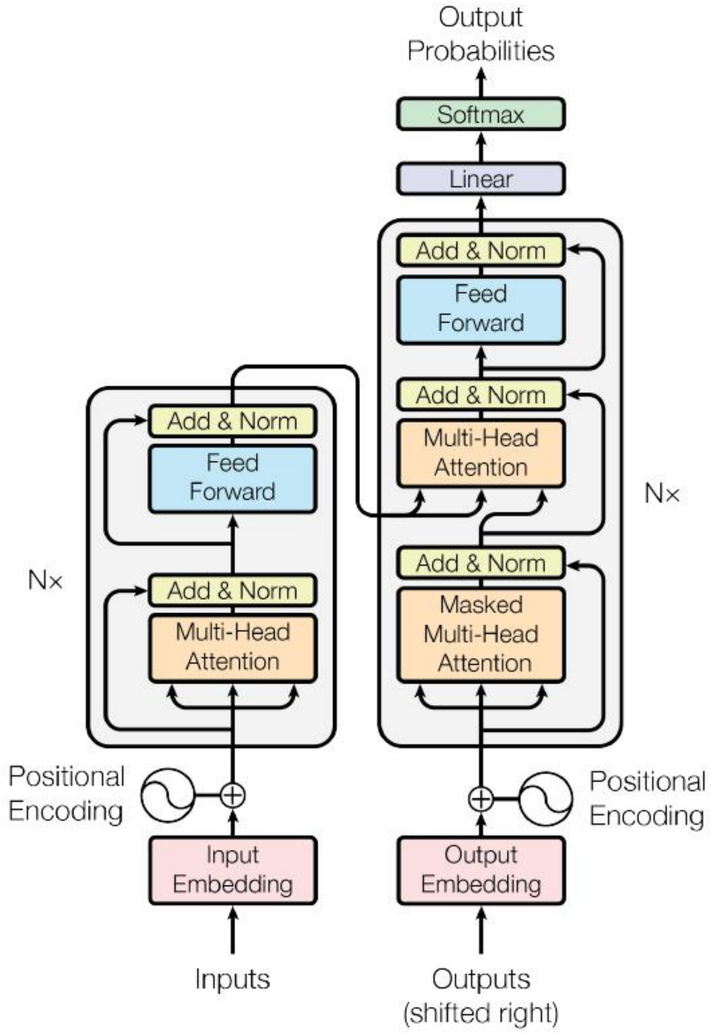
- In addition to recently identified issues that have been fixed, the main limitation for the conventional reconstruction is that track finding entirely relies on rough helix fitting.
- To mitigate background effects and address the challenges in track finding, appropriate SI models will be developed and deployed into three stages of the reconstruction pipeline: hit denoising, track finding, and track fitting.
- Integrating SI models will require to redesign part of the reconstruction software, particularly for track finding.
- SI model development is ongoing. A preliminary SI denoising model has been developed and integrated into coatjava, and validation is ongoing.
- Preliminary MLP-based models for track finding and track fitting have been developed, and transformer-based models are in development.
- Substantial efforts are needed in SI model development, reconstruction software updates, and the validation of SI-enhanced reconstruction.
- Expected timeline:
  - Complete update and optimization for conventional reconstruction algorithms by early of 2027.
  - Release the full SI-enhanced reconstruction software by end of 2027.
  - Explore additional SI approaches for track finding.

Collaboration and Personnel

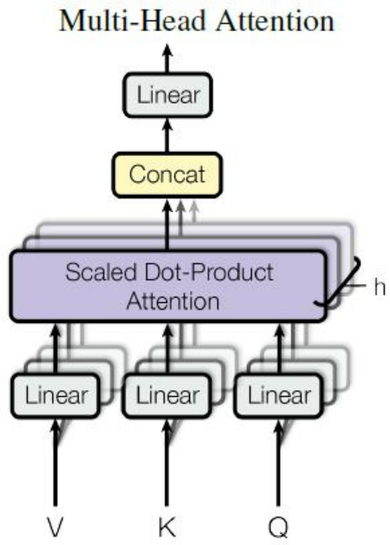
- Model development for hit denoising: Mathieu Ronayette, Pierre Chatagnon, Richard Tyson, and Tongtong Cao
- Model development for track finding and track fitting: Tongtong Cao, Uditha Weerasinghe
- Reconstruction software development: Tongtong Cao, support from the software group and CVT experts
- Model & reconstruction validation: Tongtong Cao, Nerses Gevorgyan, et al.
- Please contact me (caot@jlab.org) if you are interested into the project, and have experience with SI model development, or validation for reconstruction and physics performance.

Backup





The Transformer - model architecture.



Scaled Dot-Product Attention

