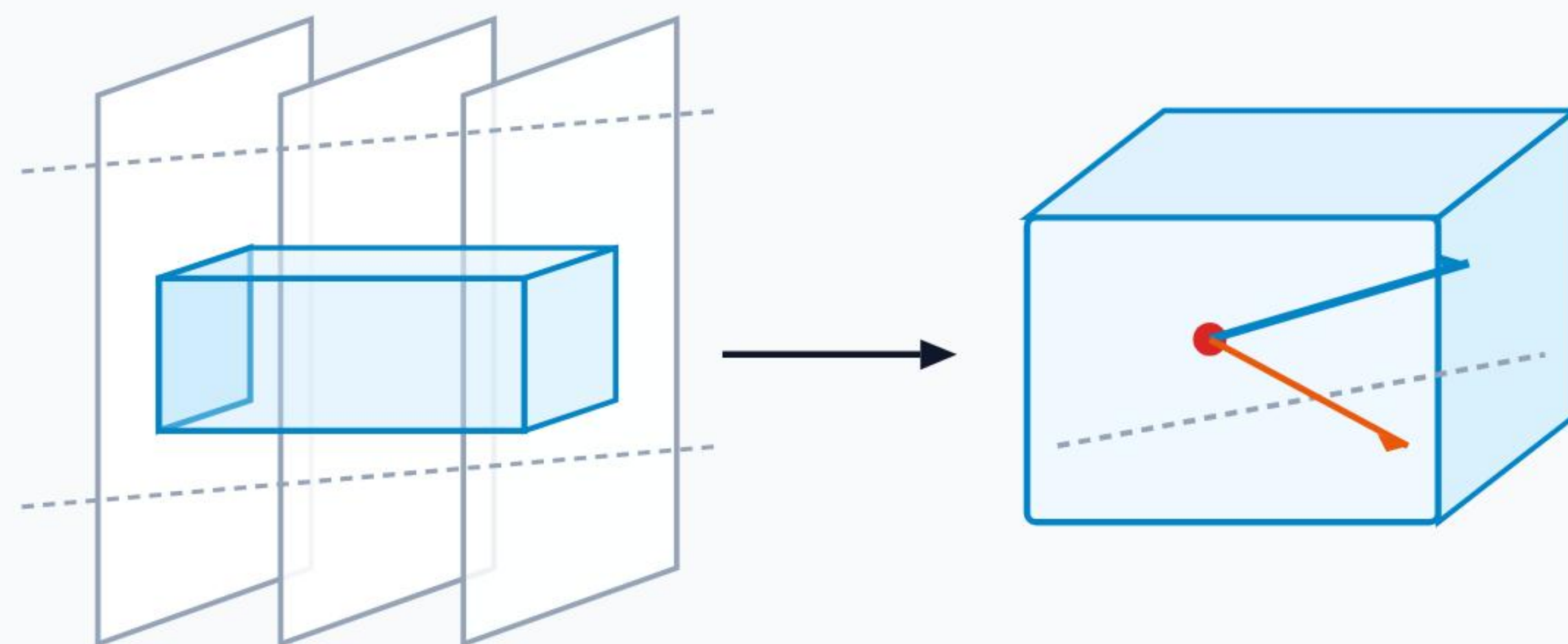


Detector Volume Sampling & Subvolume Pipeline

Base Storage: HDF5

FIGURE 1: Full Emulsion Stack & Subvolume Extraction

730 Plates // HDF5 Base Storage



Full Emulsion Stack (730 Plates)
~140 GB background in HDF5 with selectable candidate ROI cutouts.

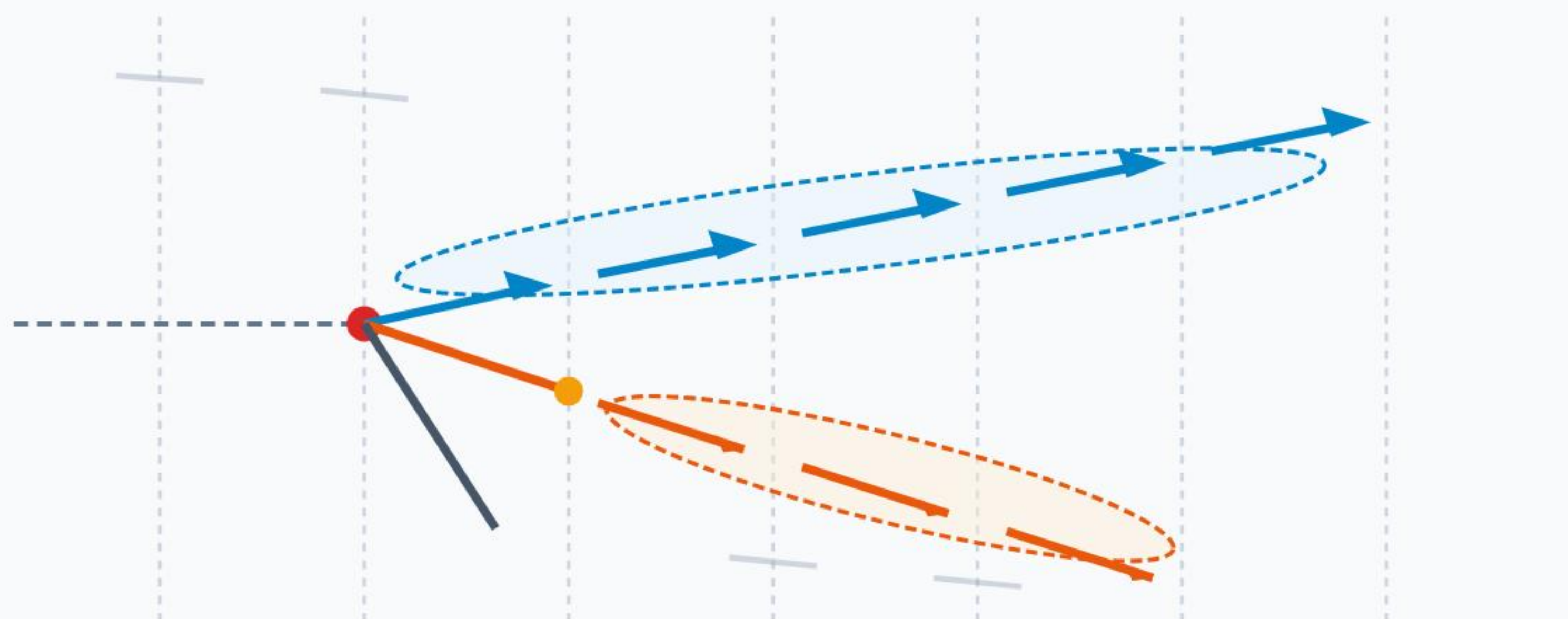
Extracted 3D Subvolume (ROI)
Target interaction overlaid onto background and streamed to data loaders.

- **Base Data Storage:** Background (~140 GB across 730 emulsion plates) and simulated signal instances are stored natively in HDF5 format.
- **Subvolume Ingestion:** Rather than loading monolithic full-detector arrays, extract localized 3D subvolume bounding boxes around candidate regions.
- **Signal vs. Background:** Keep simulated signal vertices and measured background separate; overlay interaction vertices directly onto background instances during training.
- **Background Density Scaling:** Explore stitching adjacent background samples to evaluate reconstruction robustness under elevated track densities.
- **Data Loader Integration:** High-performance ML data loaders stream and crop subvolumes on the fly directly from the base HDF5 archives.

Instance Definition & Reconstruction Targets

Tracking & Kinematics

FIGURE 2: Base Track Ensembles to Particle Flow Instance Graph & Label Topology



- **Instance Definition:** The ensemble of base tracks across plates within the subvolume that collectively form reconstructed particles.
- **Directional Linkage:** Predict directed graph connections associating base track segments across plates along the particle's direction of travel.
- **Kinematics & Coordinates:** Supervised regression for track momentum magnitude $|p|$ along with plate start/end spatial coordinates.
- **Decay Topology:** Hierarchically label secondary decay kinks and associate daughter tracks to parent short-lived particles.
- **Particle Flow Output:** Provide model-agnostic targets enabling both graph neural networks and full particle flow architectures.

Interaction Classification & Benchmark Scope

Physics & Composition

Target Channels & Composition

Neutrino CC Flavors Charged Current ν_e, ν_μ, ν_τ interactions	~10k each
Neutral Current (NC) Hadronic shower topologies without primary lepton	~10k
Dedicated Charm Topology ν CC events featuring short-lived charm decay kinks	~10k
Non-Flavor Backgrounds Neutral hadrons, charged hadrons, muon photonuclear	~10k each
Target Benchmark Scale	Scaling from ~40k up to 100k+ events

Benchmark Scope & Formulation

- Vertex Localization:** Primary interaction task predicting the spatial 3D coordinates of the interaction vertex inside the emulsion volume.
- Current & Energy:** Predict incoming neutrino energy E_ν and classify Neutral Current vs. Charged Current interactions.
- Dedicated Charm Finding:** Include labeled ν CC charm production events featuring characteristic short-lived kink signatures.
- Non-Neutrino Topologies:** Incorporate neutral hadron, charged hadron, and muon photonuclear interactions as critical false-positive background tests.
- Generator Invariance:** Provide events synthesized across multiple generator-side configurations to assess model generalization.