Feasibility Study of a Graph-Neural-Network-Based Cosmic Muon Trigger for the Mu3e Experiment

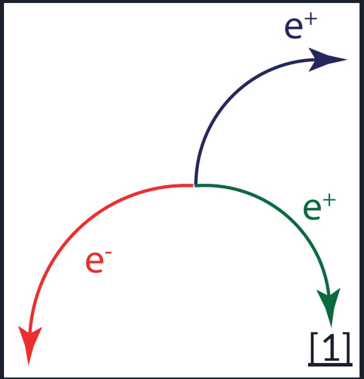
David Karres
On behalf of the Mu3e collaboration
Physikalisches Institut - Heidelberg University
CHEP 2024 - Kraków
October 19.-25. 2024

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

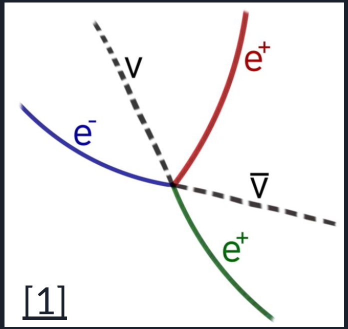
The Mu3e Experiment

- Located at the Paul Scherrer Institute (PSI) in Switzerland
- Planned to search for the lepton flavour violating decay $\mu^+ \rightarrow e^+e^-e^+$
  - Standard Model branching ratio $<10^{-54}$
  - Observation would point towards physics beyond the SM
- Goal:
  - Directly observe the decay, **or**
  - Set a new upper limit for the branching ratio of $<10^{-16}$
- Background suppression requirements:
  - Momentum resolution <0.5MeV/c
  - Precise vertex finding

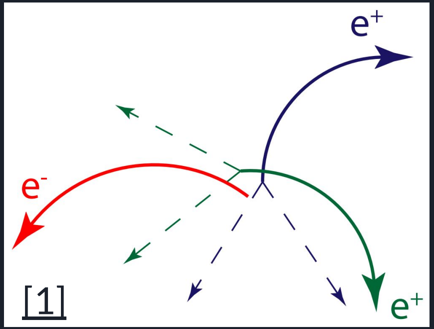
Signal topology



Internal conversion background ($\mu^+ \rightarrow e^+e^-e^+\nu\bar{\nu}$)

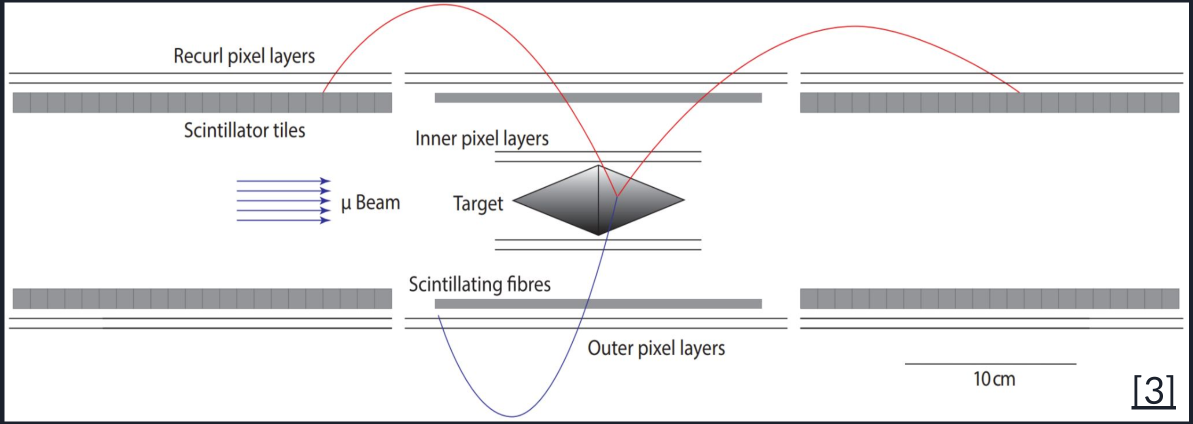


Combinatorial background Michel decay ($\mu^+ \rightarrow e^+\nu\bar{\nu}$) + Bhabha



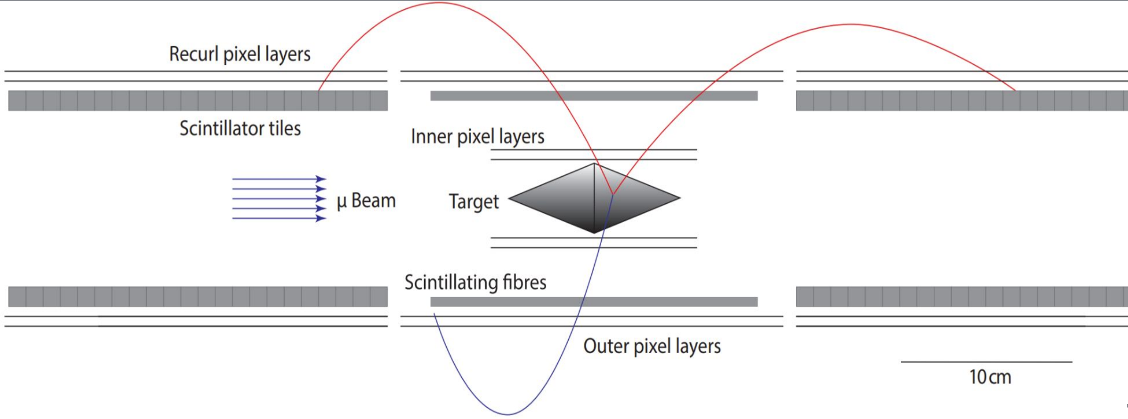
The Mu3e Pixel Tracking Detector

- Muon beam with 10^8 muons/second for phase I
  - Muons decay at rest on the target
- Ultra-thin silicon pixel sensors (MuPix)
- Solenoidal magnetic field, B = 1T
- Gaseous helium cooling



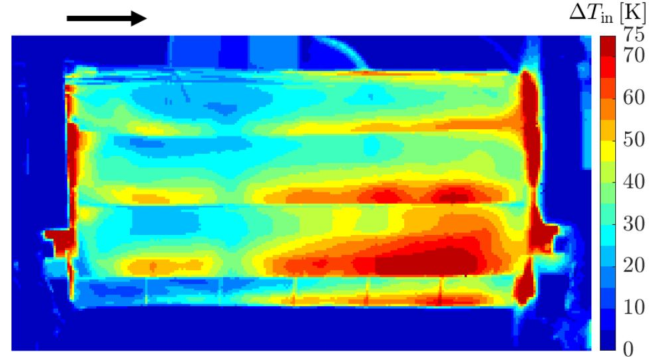
The Mu3e Pixel Tracking Detector

- Muon beam with 10^8 muons/second for phase I
  - Muons decay at rest on the target
- Ultra-thin silicon pixel sensors (MuPix)
- Solenoidal magnetic field, B = 1T
- Gaseous helium cooling



[3]

- Time-dependent misalignment factors:
  - Thermal expansion
  - Helium flow
  - Magnetic field
- Online detector alignment system required

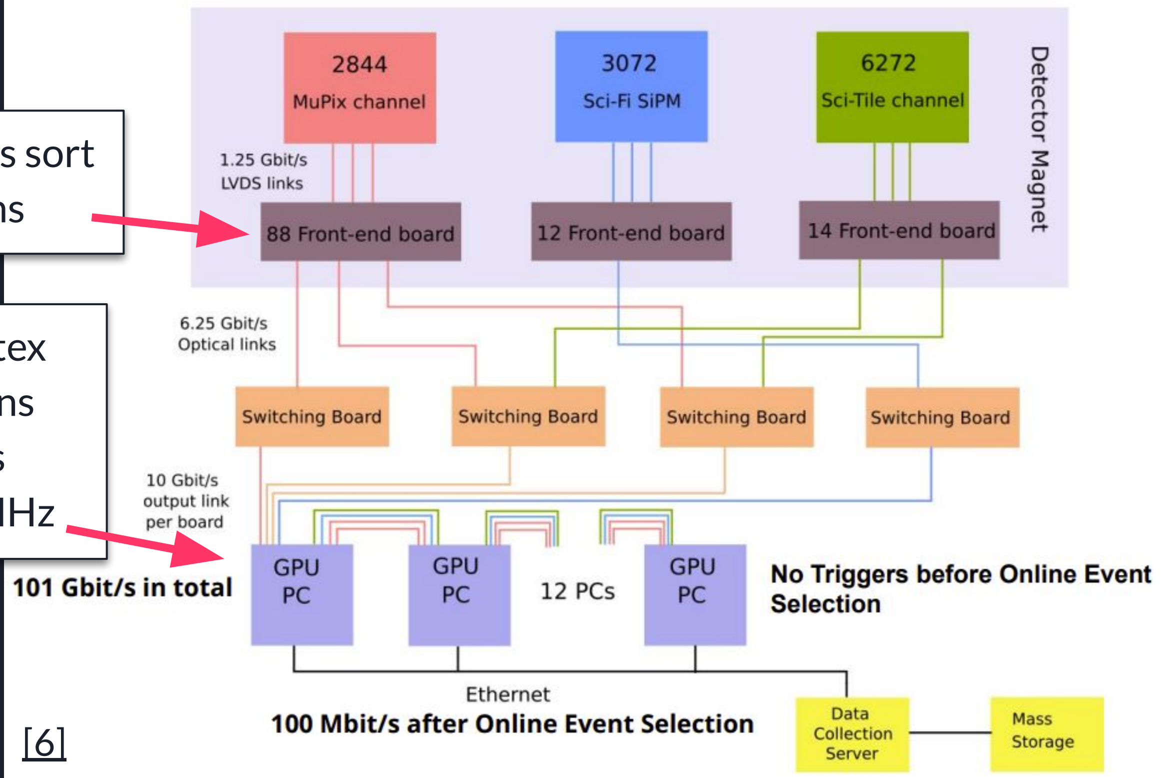


[3]

Data Acquisition at Mu3e

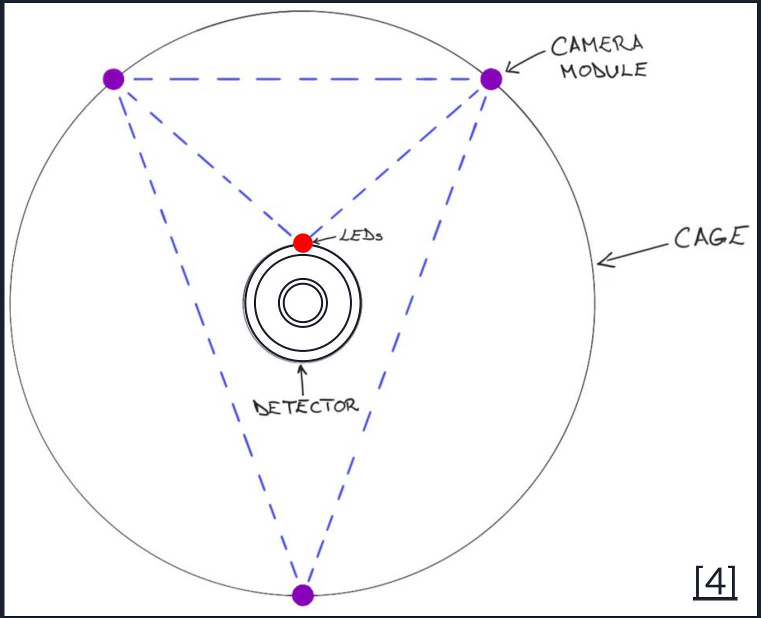
Front-end FPGAs sort data into 8 ns bins

Online track and vertex reconstruction of 64ns time frames on GPUs
⇔ Frame rate 15.6 MHz



Online Alignment at Mu3e

- Camera tracking system
  - Outer layer is tracked with cameras
  - Provide reference layer for alignment

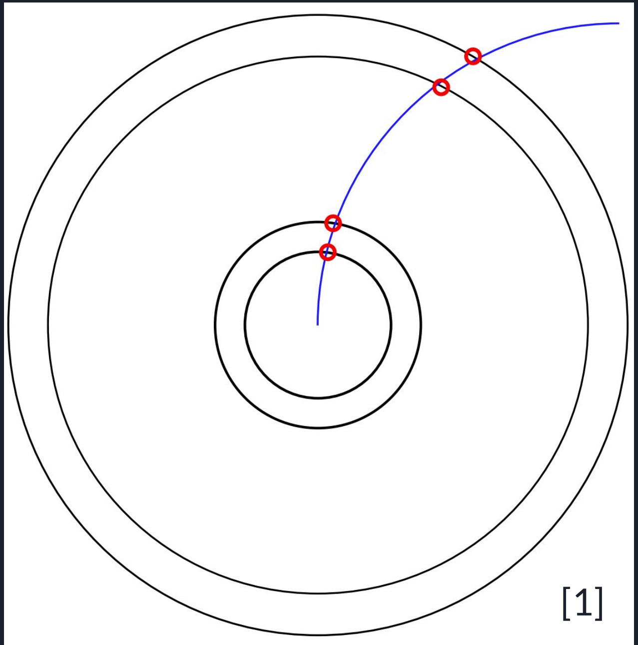


Online Alignment at Mu3e

- Camera tracking system
  - Outer layer is tracked with cameras
  - Provide reference layer for alignment

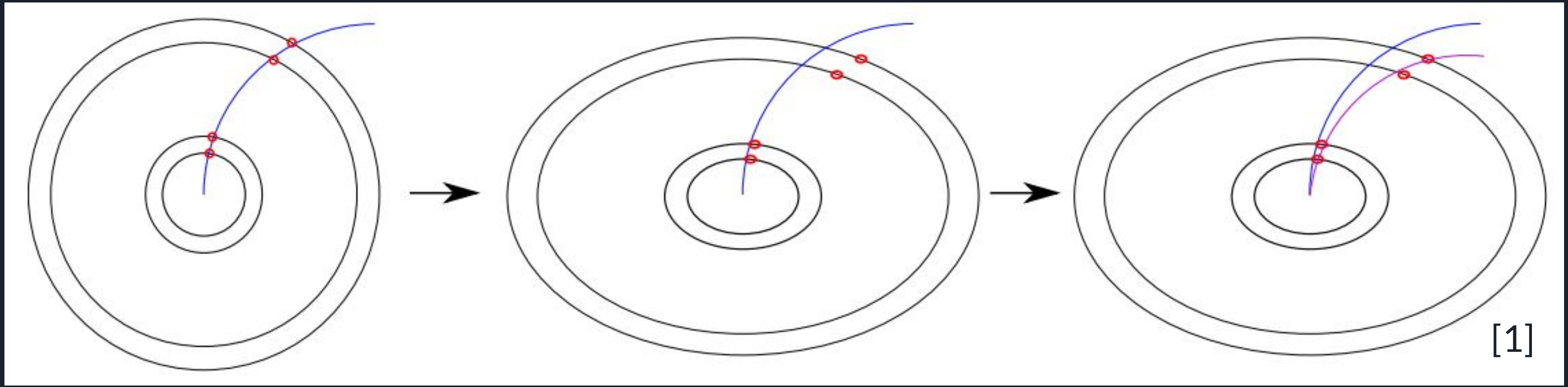
- Track-based alignment
  - Use tracks originating from the target in the central barrel
  - Minimize Chi2 by adjusting track and alignment parameters

$$\chi^2(\mathbf{q}_j, \mathbf{p}) = \sum_j^{\text{tracks}} \sum_i^{\text{hits}} \left(\frac{m_{ij} - f(\mathbf{q}_j, \mathbf{p})}{\sigma_{ij}} \right)^2$$



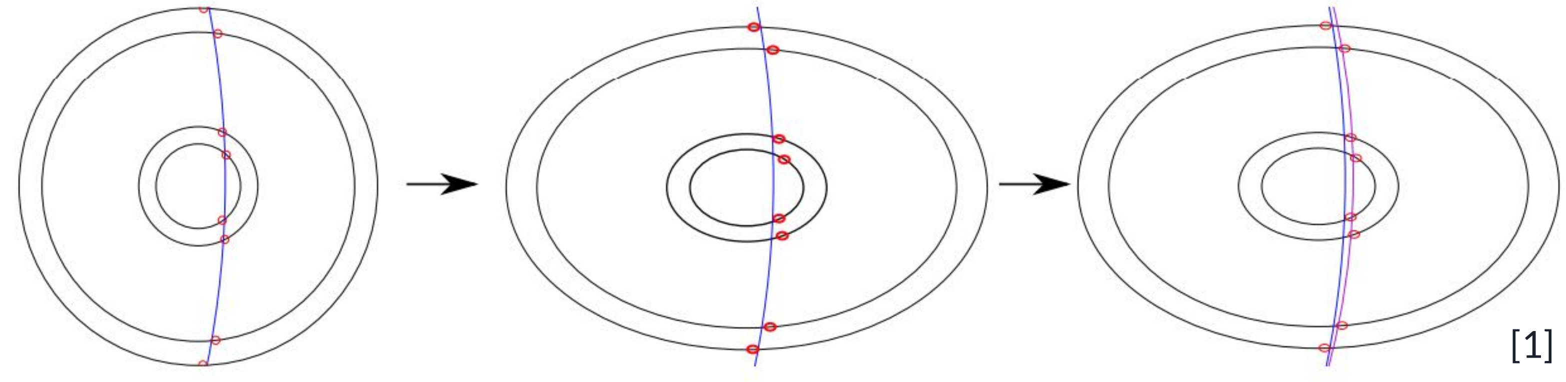
Weak Modes

- Problem of track-based alignment: Weak modes
    - Global tracker deformations like bowing, twisting, etc.
    - Able to fit a track with equal Chi2 as for the aligned case



Weak Modes

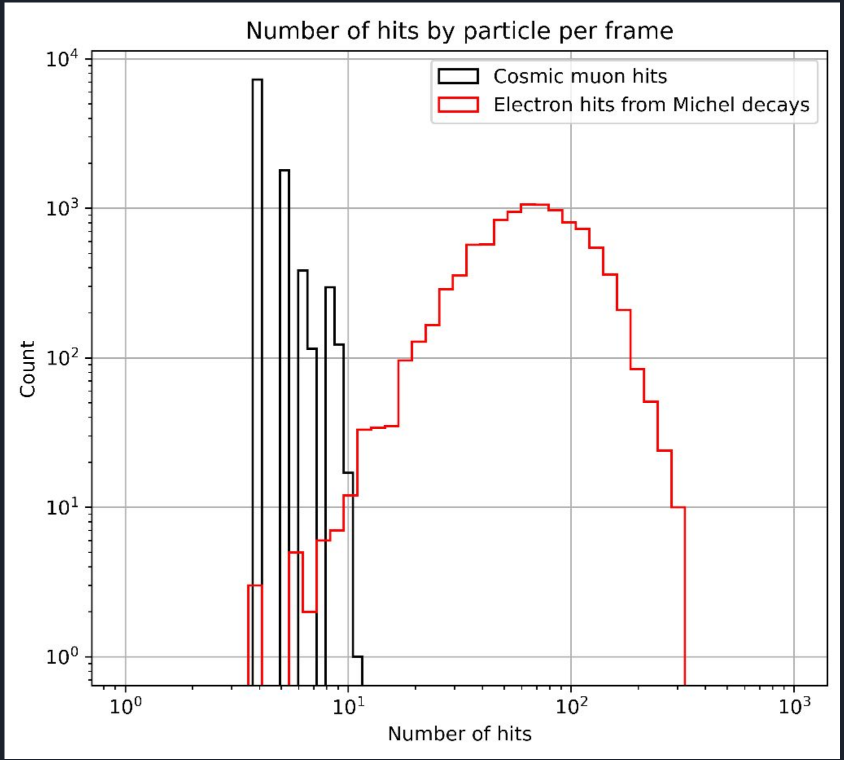
- Problem of track-based alignment: Weak modes
  - Global tracker deformations like bowing, twisting, etc.
  - Able to fit a track with equal Chi2 as for the aligned case
- Solution: Cosmic ray muons
  - Fundamentally different track topology
  - High momentum ⇒ Almost no deflection inside magnetic field
  - Even distribution over whole detector



[1]

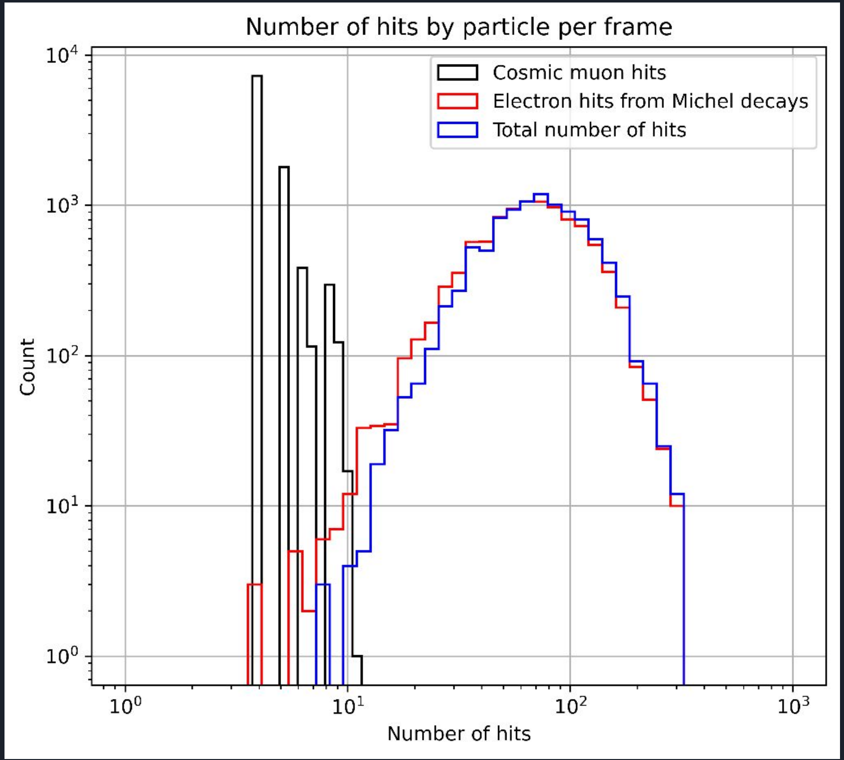
Challenges for a Cosmic Muon Trigger

- Cosmics rate 10 Hz ↔ Beam rate 100 MHz:
  Only 1 in 10^6 frames (64 ns) contain a cosmic muon
- Background suppression: 10^{-3} - 10^{-4} frames
- High trigger efficiency
- Most cosmics leave ~4 hits among
  O(100) background hits per frame



Challenges for a Cosmic Muon Trigger

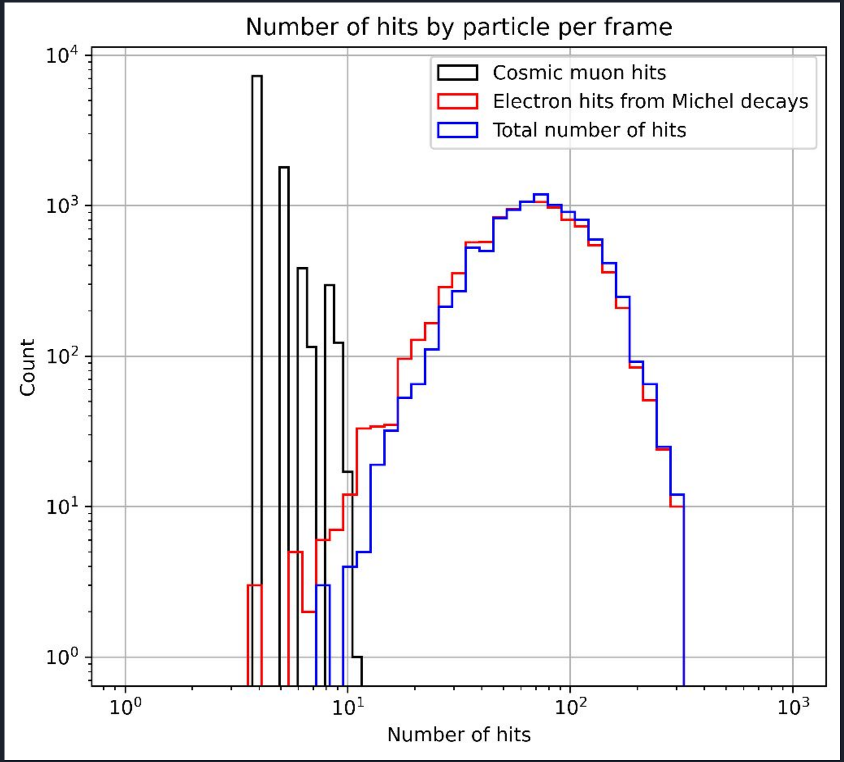
- Cosmics rate 10 Hz ↔ Beam rate 100 MHz:
  Only 1 in 10^6 frames (64 ns) contain a cosmic muon
- Background suppression: 10^{-3} - 10^{-4} frames
- High trigger efficiency
- Most cosmics leave ~4 hits among
  O(100) background hits per frame



Challenges for a Cosmic Muon Trigger

- Cosmics rate 10 Hz ↔ Beam rate 100 MHz:
  Only 1 in 10^6 frames (64 ns) contain a cosmic muon
- Background suppression: 10^{-3} - 10^{-4} frames
- High trigger efficiency
- Most cosmics leave ~4 hits among
  O(100) background hits per frame

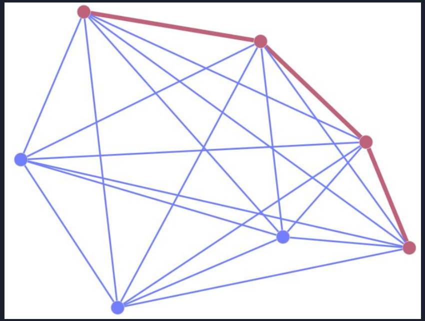
- Previously studied: Hardware-based pattern recognition with associative memory chips [5]
- This study: **Graph neural networks (GNNs) for cosmic muon track reconstruction**



Track Reconstruction with Graph Neural Networks

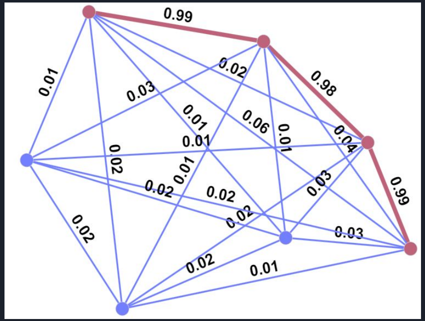
Graph Construction

- Represent frames as graphs
- Hits ⇔ Nodes
- Track segments ⇔ Edges
- Frames contain O(10^2) hits
  ⇒ Fully connected graphs
  ⇒ O(10^4) edges/graph



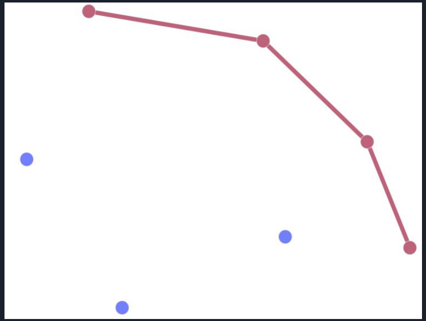
Edge Labeling

- Infer edge-classifying GNN to obtain edge scores ∈ [0,1]
- High score ⇔ Muon edge
- Low score ⇔ Fake/electron edge



Track Reconstruction

- Cut edges below threshold
- Run connected components algorithm to obtain track candidates
- Keep candidates with ≥4 hits



Edge Labeling with Graph Neural Networks

Encoders	Interaction Network	Decoder
Two encoding Multi Layer Perceptrons (MLPs)	Edge and node update blocks	Single edge decoding MLP
Embed edge and node features into higher dimensional latent space	Perform message passing to learn geometric patterns within the graph	Transform latent space representation into single edge score

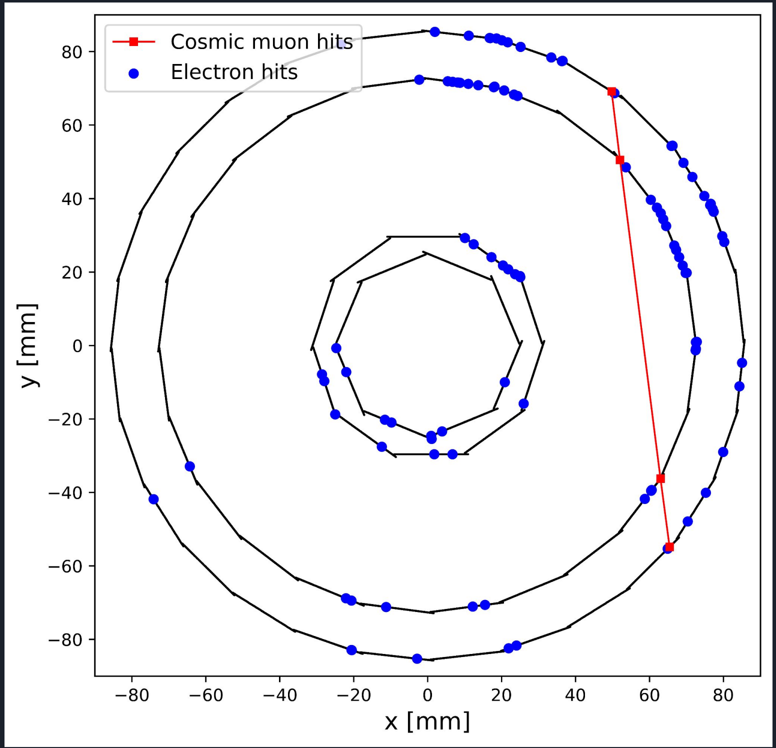
Preliminary model size optimization:

- 24 dimensional edge- and node feature embeddings
- 4 message passing steps
- 2 hidden layers per MLP

Based on the Acorn framework [7]

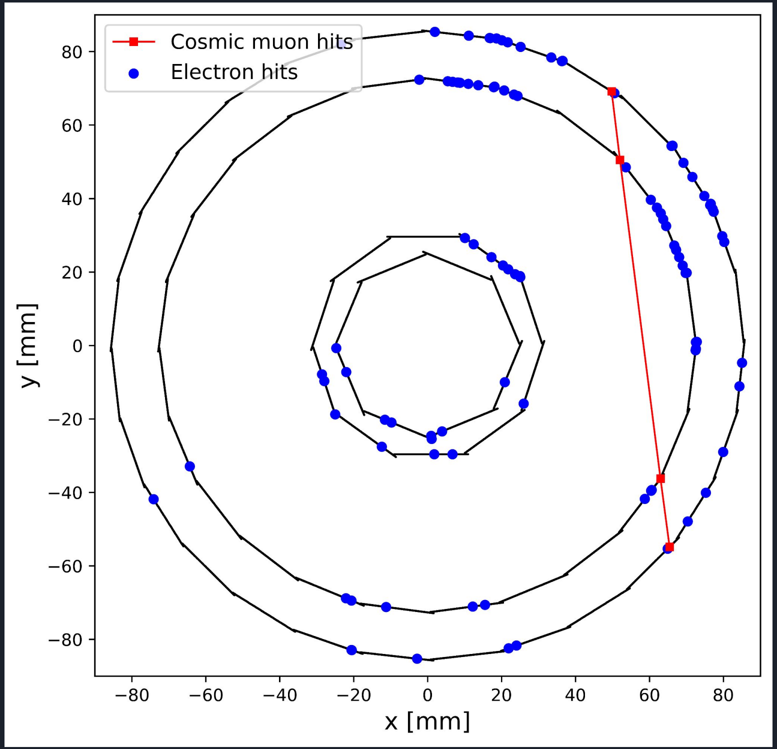
Data Samples for Training and Inference

- Simulated samples for training and inference via Mu3e simulation package (based on Geant4)
- Data samples:
  - Cosmic muons mixed with Michel decays (signal)
  - Michel decays only (background)
- Train on:
  - 100% Cosmics with Michel
  - 50% Cosmics with Michel + 50% Michel only (boost background rejection)
  - 10k training frames/epoch
  - 1k frames/epoch for validation and testing



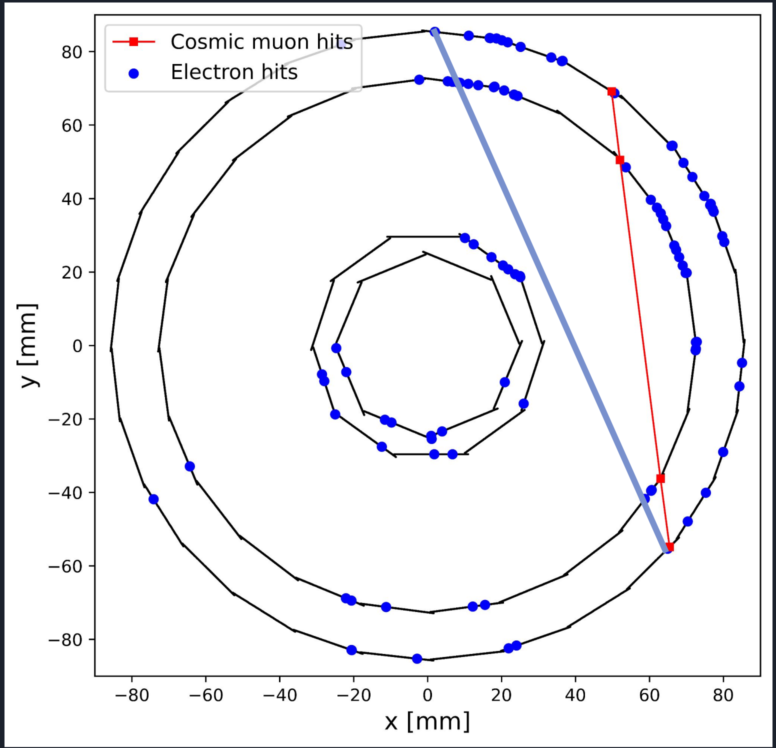
Track Reconstruction: Definitions

- True track: All hits from cosmic muon
- Fake track: At least one non-muon hit
  ⇒ Trigger any frame with at least one reconstructed cosmic muon track
- Track reconstruction efficiency:
  - Evaluate on cosmics with Michel
  - #triggers/ #signal frames
- Background acceptance:
  - Evaluate on Michel only
  - #triggers / #background frames



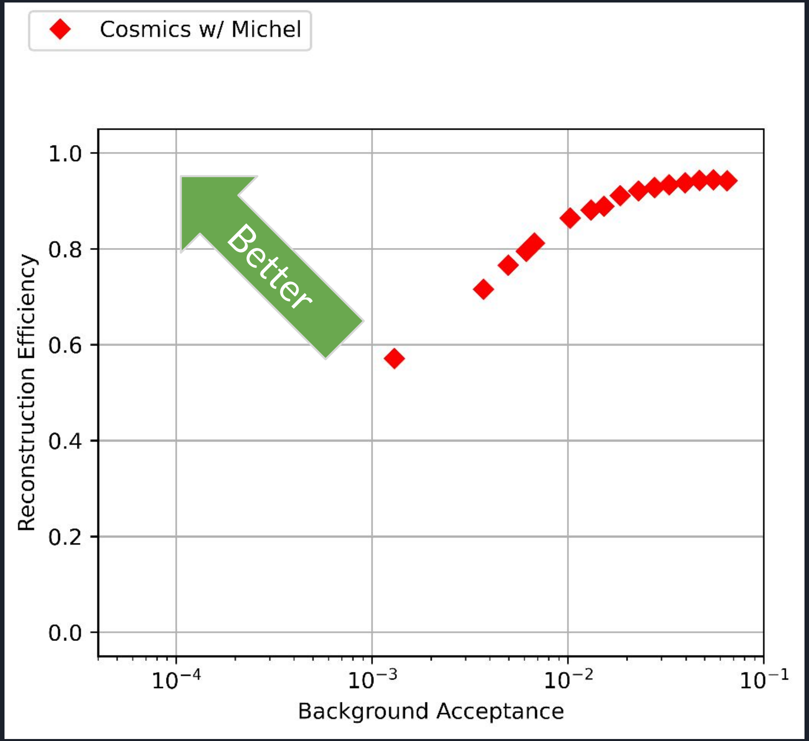
Track Reconstruction: Definitions

- True track: All hits from cosmic muon
- Fake track: At least one non-muon hit
  ⇒ Trigger any frame with at least one reconstructed cosmic muon track
- Track reconstruction efficiency:
  - Evaluate on cosmics with Michel
  - #triggers/ #signal frames
- Background acceptance:
  - Evaluate on Michel only
  - #triggers / #background frames



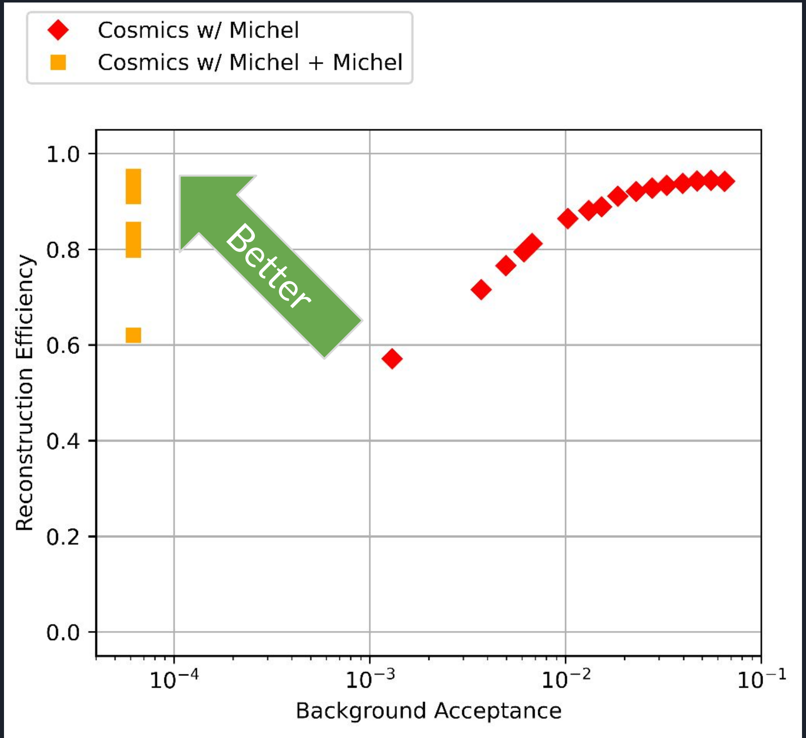
Track Reconstruction: Evaluation

- Inference on 16k Michel frames and 1k cosmics with Michel frames
- Baseline model
  - Reconstruction efficiency ≤95%
  - Efficiency drops below 80% at background acceptance ~6×10^{-3}



Track Reconstruction: Evaluation

- Inference on 16k Michel frames and 1k cosmics with Michel frames
- Baseline model
  - Reconstruction efficiency ≤95%
  - Efficiency drops below 80% at background acceptance ~6×10^{-3}
- Adding Michel only frames decreases background acceptance by orders of magnitude
  - Reconstruction efficiency ~95% at background acceptance ≤6.2×10^{-5}
  - Only one fake track observed ⇒ Statistically limited



Summary

- Utilization of graph neural networks for cosmic muon track reconstruction are studied for the Mu3e experiment
- Massive boost in performance by training on Cosmics with Michel + Michel only
  - Track reconstruction efficiency: 95.2% at background acceptance: $\leq 6.2\times10^{-5}$
  - Full reconstruction rate of <1000 frames/s required
- Online trigger deployment on GPU or FPGA is targeted
  - Throughput (and optimization) has to be studied

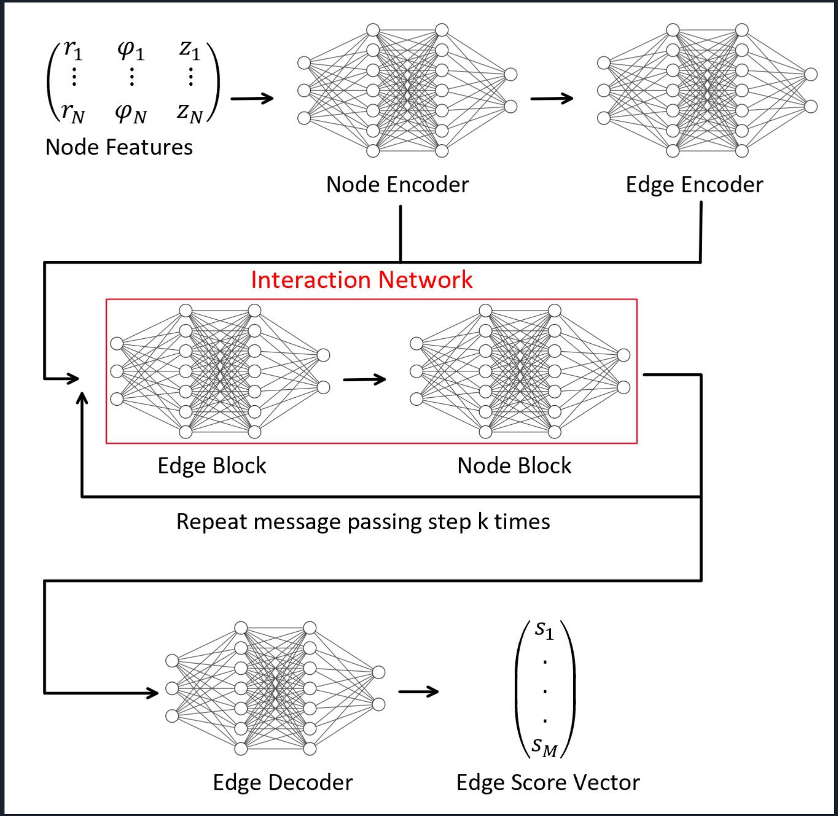
References

- [1] U. Hartenstein, Track Based Alignment for the Mu3e Pixel Detector, 2019, Link
- [2] Mu3e Collaboration Website
- [3] K. Arndt et al., Technical design of the phase I Mu3e experiment, 2020, published in Nucl.Instrum.MethA. 1014, 165679, 2021
- [4] G. Stanic, A Camera Alignment System for the Mu3e Experiment, 2021, Link
- [5] K. Neureither, Towards an Online Reconstruction of Cosmic Muons for Mu3e using Hardware-Based Pattern Recognition, 2020, Link
- [6] H. Murugan, Online Track Reconstruction for the Mu3e Experiment, DPG Spring Meeting, 2024, Link
- [7] Git Repository of the Acorn framework, Link

Backup

Edge Labeling with GNNs

- Want to classify edges as true or false
  - Assign edge scores and label all edges below a score cut threshold as false and the rest as true
- Node features (cylindrical hit coordinates) are embedded into a latent space by the node encoder
- Edge encoder takes both embedded node features and generates edge features
- Both embeddings are passed to the interaction network to perform k message passing steps
- Edge decoder takes the transformed embedding and outputs a single number - the edge score



Message Passing in an Interaction Network

- Graph with embedded node and edge features is fed into interaction network
- Edge block MLP takes embedded edge feature and the connected node features and updates it
- Node block MLP takes embedded node feature and an aggregation (e.g. a sum) of the updated edge features to output node update
- Repeat k times to extract higher level geometric patterns