Machine Learning Techniques for Neutrino Reconstructions in IceCube

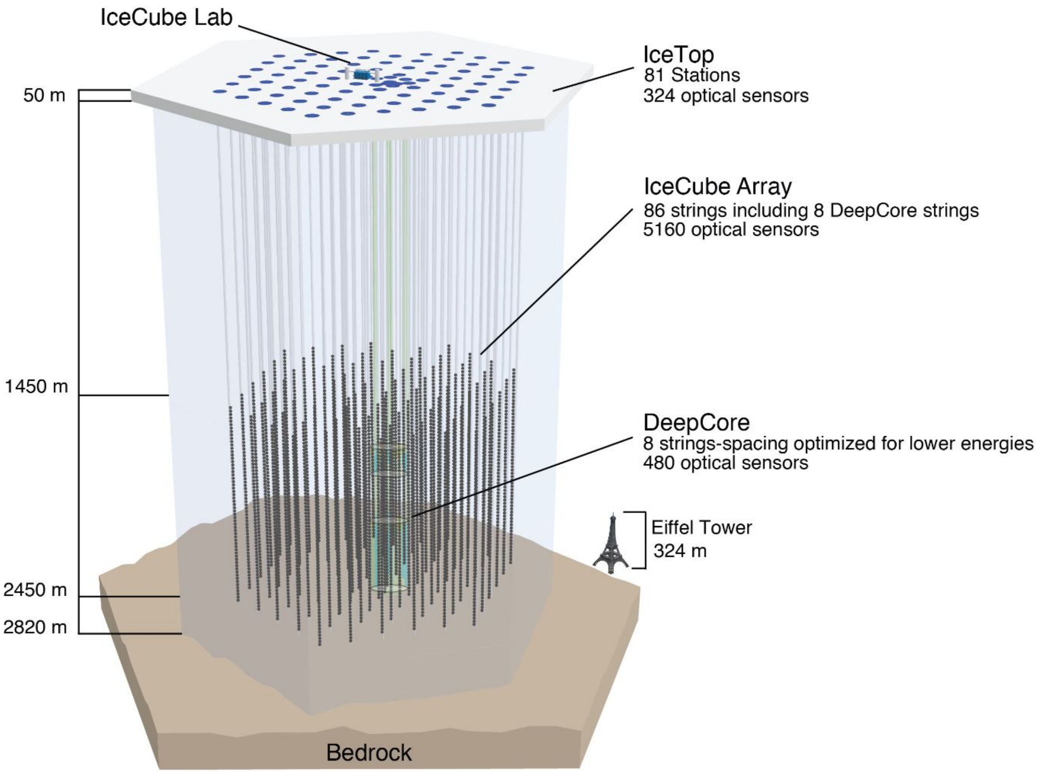
Philip Weigel
for the IceCube Collaboration
(pweigel@mit.edu)

Workshop on Machine Learning for Analysis of High-Energy Cosmic Particles
January 28, 2025

The IceCube Neutrino Observatory

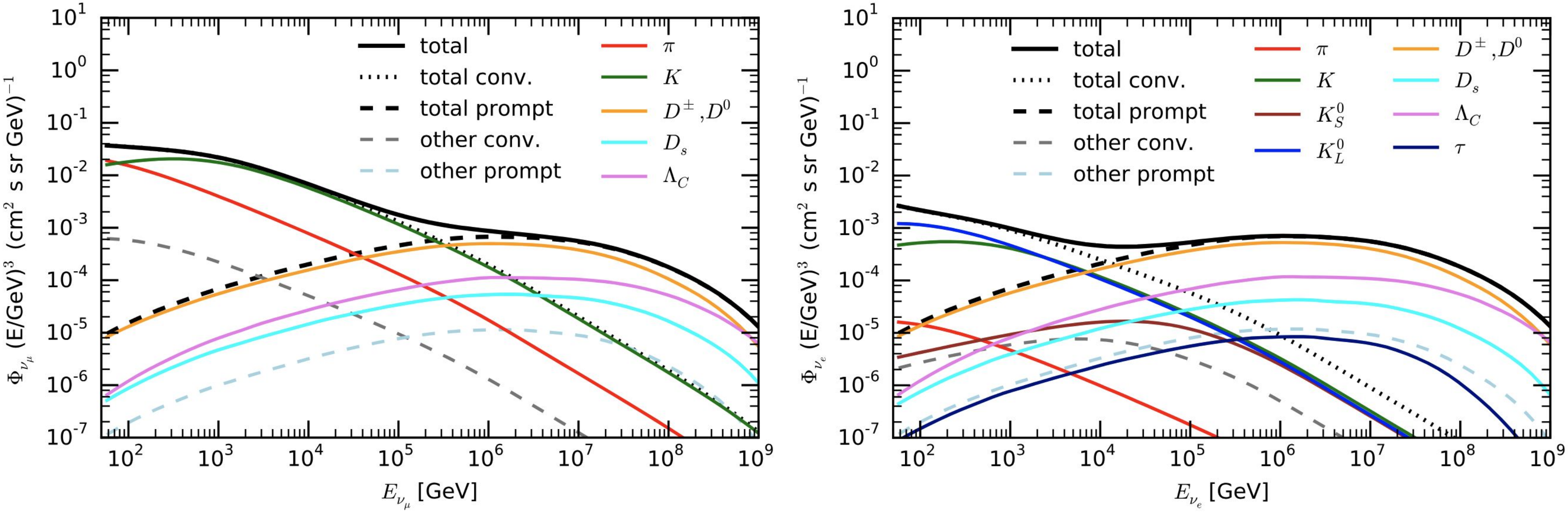
Detector Design

- 1 gigaton of instrumented ice
- 5,160 light sensors, or digital optical modules (DOMs), digitize and time-stamp signals
- 1 square kilometer surface array, IceTop, with 324 DOMs
- 2 nanosecond time resolution
- IceCube Lab (ICL) houses data processing and storage and sends 100 GB of data north by satellite daily



Where do our neutrinos come from?

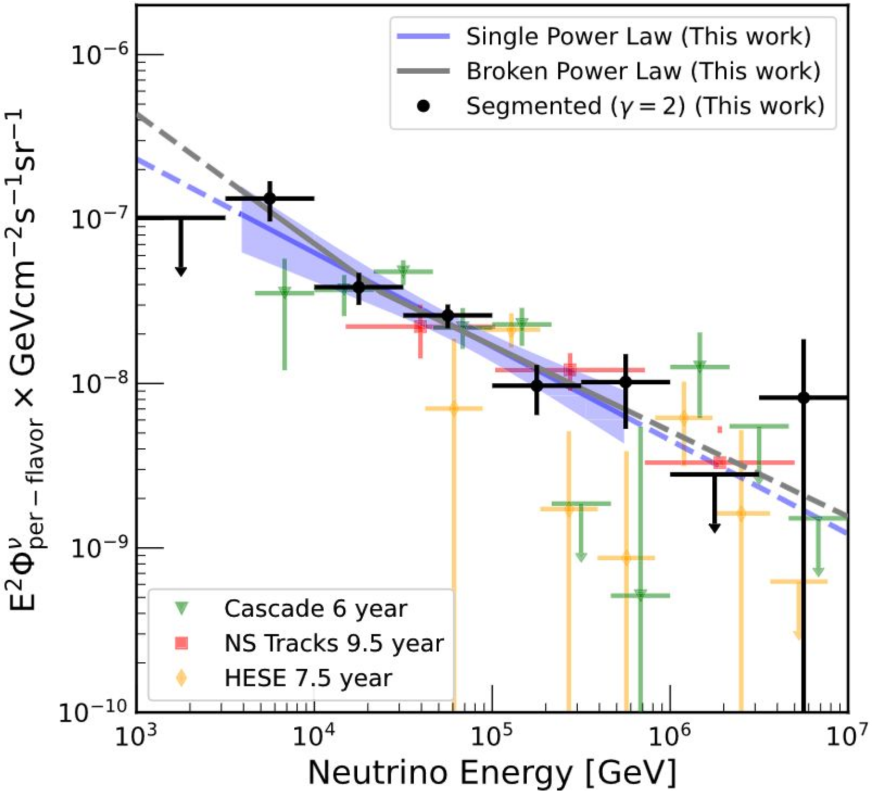
- Neutrinos are produced by cosmic ray interactions in the atmosphere
  - Primarily pion and kaon decay, small component from charmed mesons/baryons



EPJ Web Conf. 99 (2015) 08001

Where do our neutrinos come from?

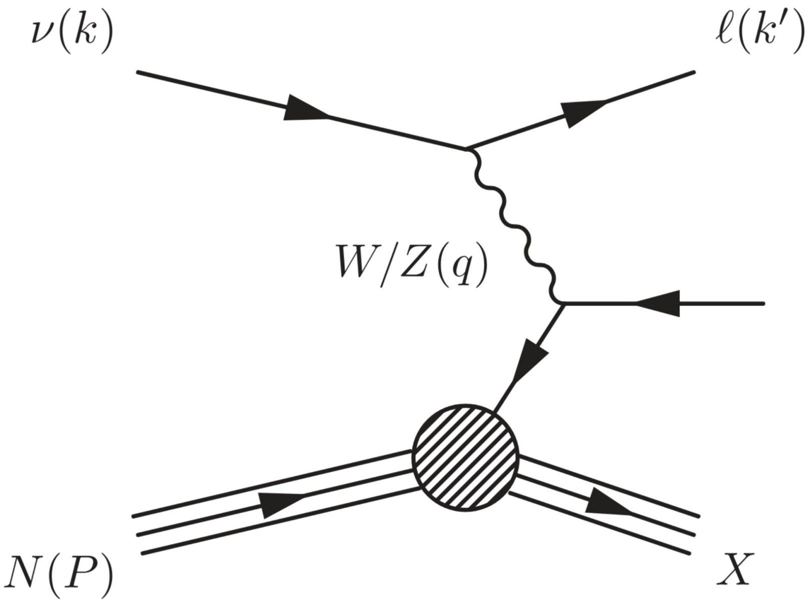
- At high energies, a larger fraction are of astrophysical origin
- Lots of interesting physics
  - Neutrino sources
  - Diffuse flux/flavor measurements
  - Beyond-the-Standard-Model physics
  - …and more!
- Low statistics, so accurate measurements of the neutrino properties are very important



Phys.Rev.D 110 (2024) 2, 022001

Neutrino Interactions

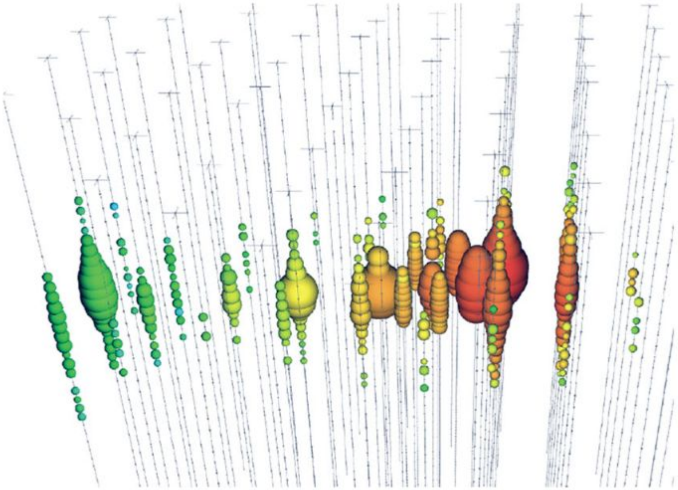
- Almost all of our events are neutrino deep inelastic scattering
  - Neutral current (NC) → out lepton = neutrino
  - Charged current (CC) → out lepton = e/μ/τ
- Neutrino energy cannot be directly measured, but inferred from the secondary particles
  - Light produced by the hadronic shower and outgoing lepton (if CC)
- The inelasticity is defined as:
  - y = hadronic energy / neutrino energy



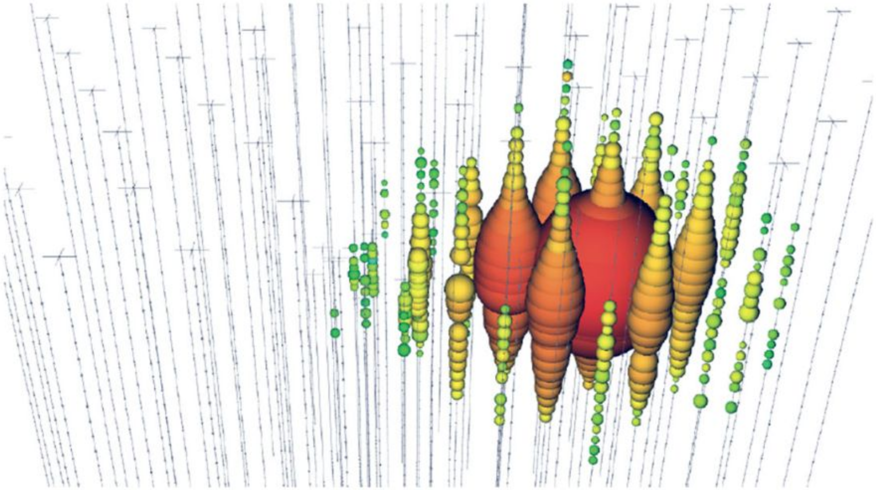
Phys.Rev.D 109 (2024) 11, 113001

Event Morphologies

- Most events fall into two classes: tracks and cascades
  - Others exist: starting tracks, double cascades, etc.



Atmospheric muons
ν_μ CC



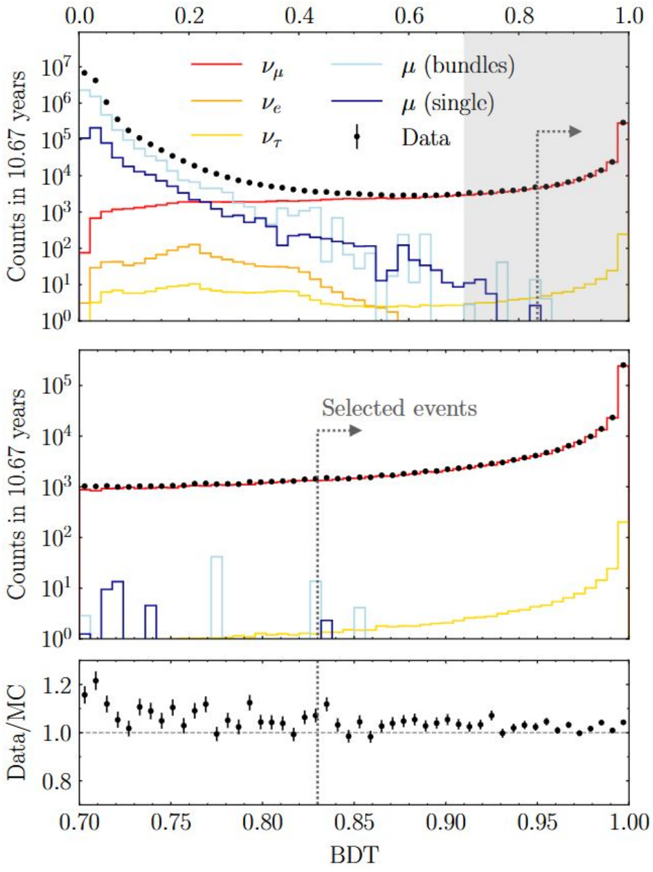
ν_e CC, ν_τ CC, ν_ℓ NC,
Glashow Resonance

Neutrino Reconstructions

- What are the quantities that we're interested in?
  - Neutrino energy
  - Energy losses
  - Direction
  - Inelasticity
  - Particle ID/event morphology
  - Vertex position
- Traditional maximum likelihood estimation-based methods can be very slow and rely on approximations
  - ML-based reconstructions → significantly faster
- Today, I will show some of the techniques we have used in recent results and new/ongoing developments
  - Not an exhaustive list!

Boosted Decision Trees

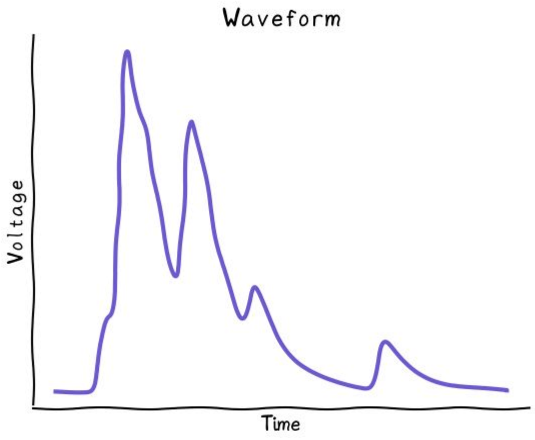
- "Classical" machine learning tool often used to remove background events
- Example: latest 3+1 sterile neutrino analysis event selection
  - Uses high-level reconstructions and low-level event statistics as inputs
  - Trained on a large sample of atmospheric muon, bundle, and neutrino events
- Powerful discriminator against atmospheric muon backgrounds and cascade events → pure track sample
  - >99.9% muon neutrino purity, ~350k events in 10.7 years of data



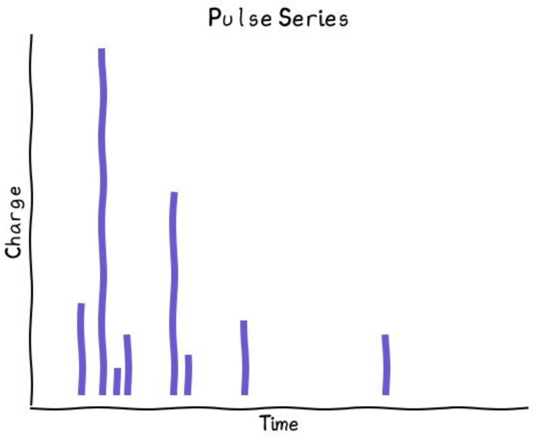
Phys.Rev.D 110 (2024) 9, 092009

DNN Input Features

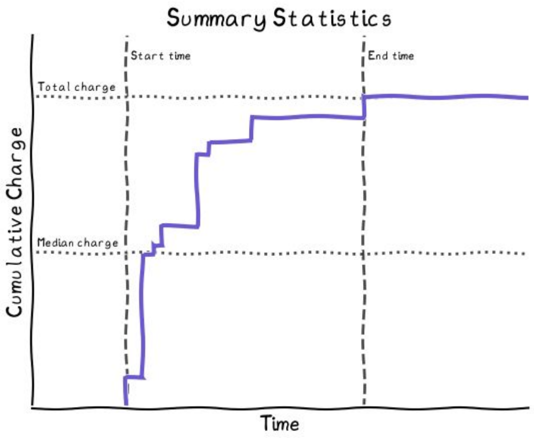
- Need to choose how you want to input your data into a network
- Three main options:



Very long, variable length



Variable length (unless padded)

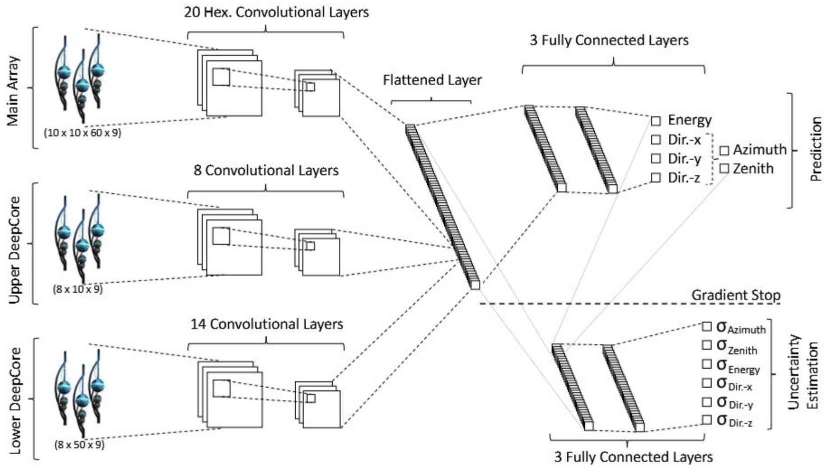


Small, fixed size for every DOM

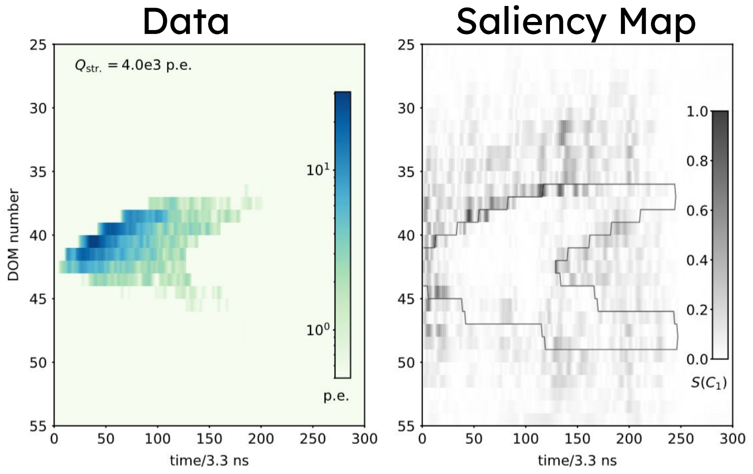
Less information

Convolutional Neural Networks

- CNNs require fixed input sizes
  - Absence of data → pad with zeros
  - Can use modified convolutions to exploit detector symmetries
- Different implementations used in the observation of neutrinos from the galactic plane and the observation of astrophysical tau neutrinos



JINST 16 (2021) P07041

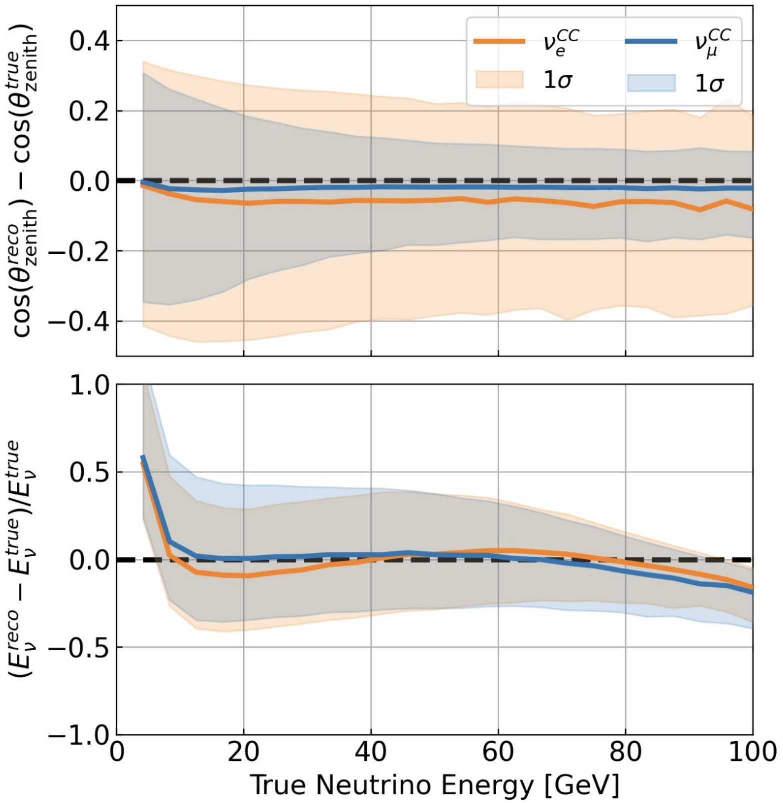


Phys.Rev.Lett. 132 (2024) 151001

Convolutional Neural Networks

- Recent neutrino oscillations result with DeepCore leveraged CNN-based reconstructions

- Largest improvements in the lowest energy bins (E < 40 GeV)
  - Important for resolving the oscillation maximum!

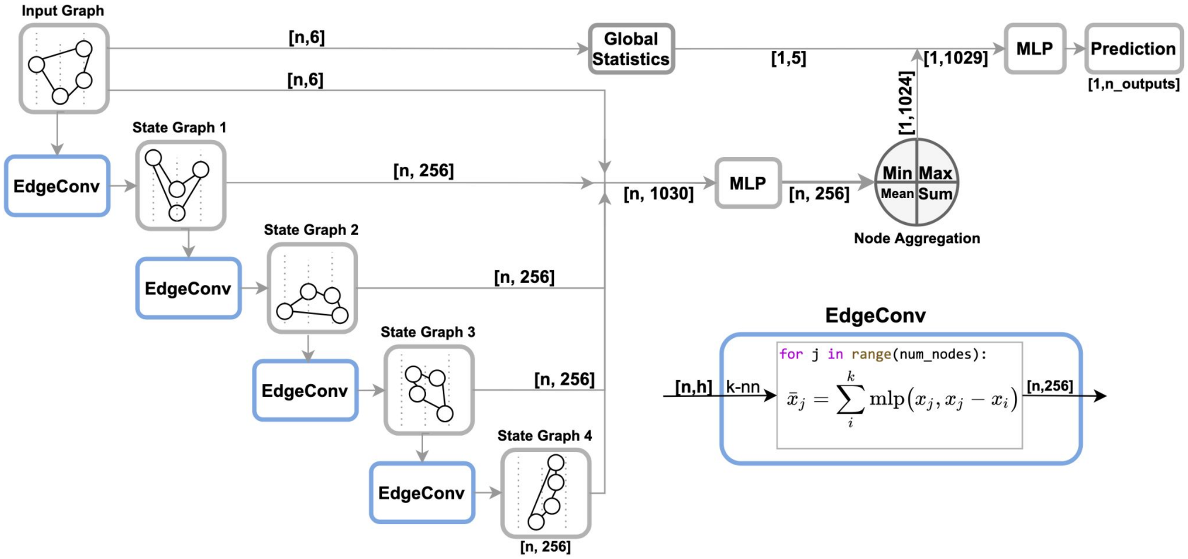


arXiv:2405.02163

Graph Neural Networks: DynEdge

- DynEdge is a graph neural network (GNN) model
  - Construct a graph representation of data, perform edge convolutions, and combine with global event information

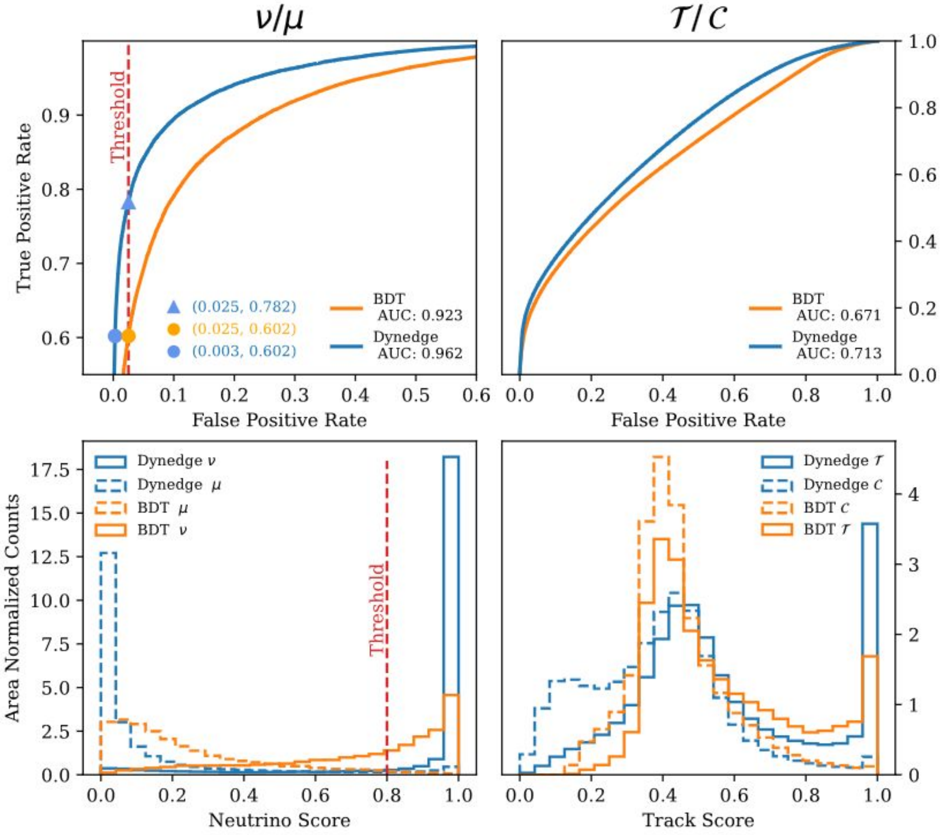
kNN Graph →



JINST 17 (2022) 11, P11003

Classification: DynEdge vs. BDT

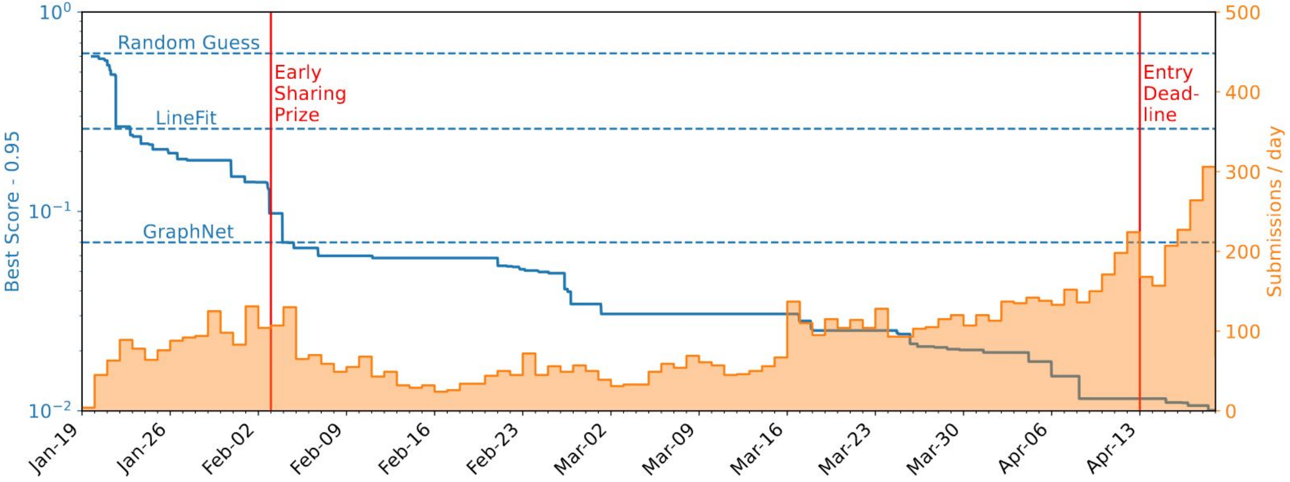
- At low energies, DynEdge has been shown to outperform BDTs at classification tasks
  - Neutrinos vs. muons
  - Tracks vs. cascades
- Improvements over LLH methods for reconstruction tasks
  - Up to 20% improvements in energy and direction reconstructions
- Graph inputs are constructed using summary statistics for each DOM as a node
  - 8 nearest neighbors for edge connections



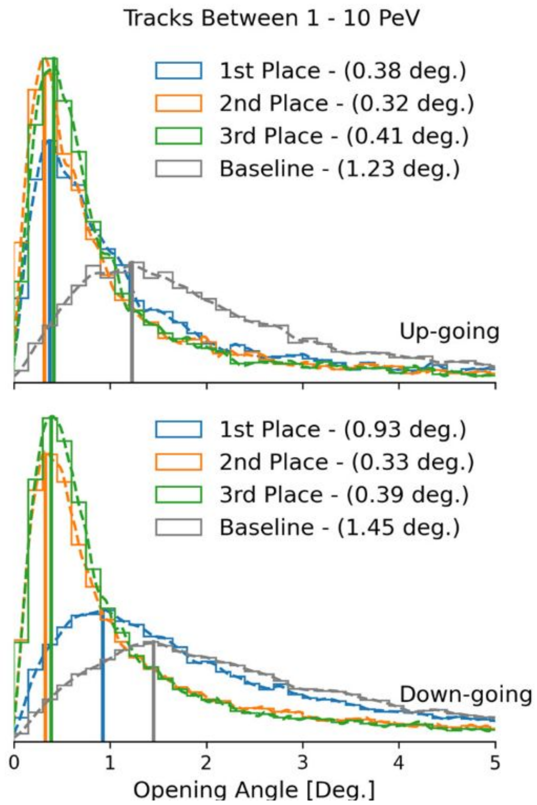
[JINST 17 (2022) 11, P11003]

IceCube Kaggle Competition

- Public competition (direction reco) with monetary prizes:
  - A large sample of IceCube simulation was provided
  - https://www.kaggle.com/competitions/icecube-neutrinos-in-deep-ice
- Many different techniques and DNN architectures with interesting results



arxiv:2307.15289

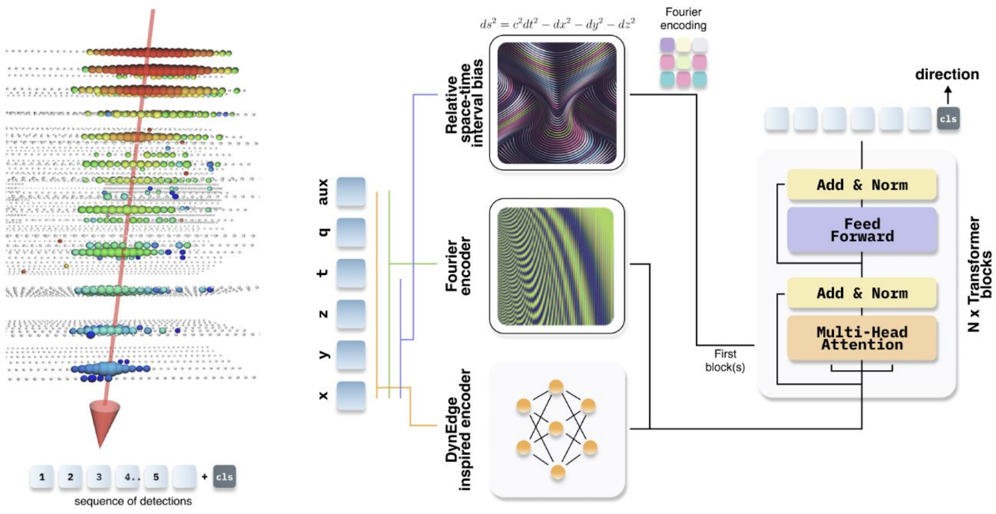


Eur.Phys.J.C 84 (2024) 6, 646

Transformer Reconstructions

- Transformer-based models were a large portion of the best-scoring solutions in the Kaggle competition
- Basic architecture:
  - Graph/positional encoding
  - Multi-head attention
  - MLP
- Reconstructed quantity can be extracted from a learnable token or a combined output sequence
- These techniques have been applied to other IceCube reconstruction tasks

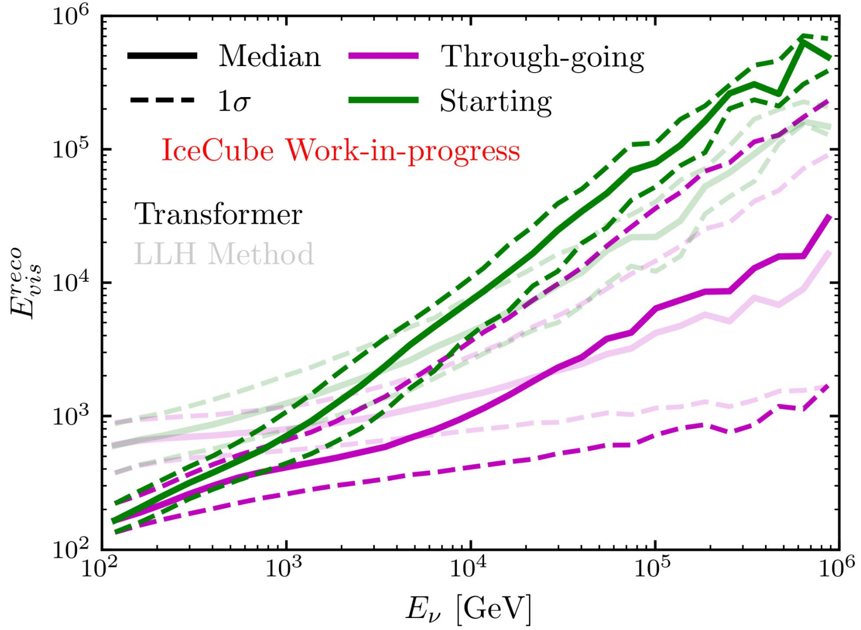
Kaggle 2nd Place Solution:

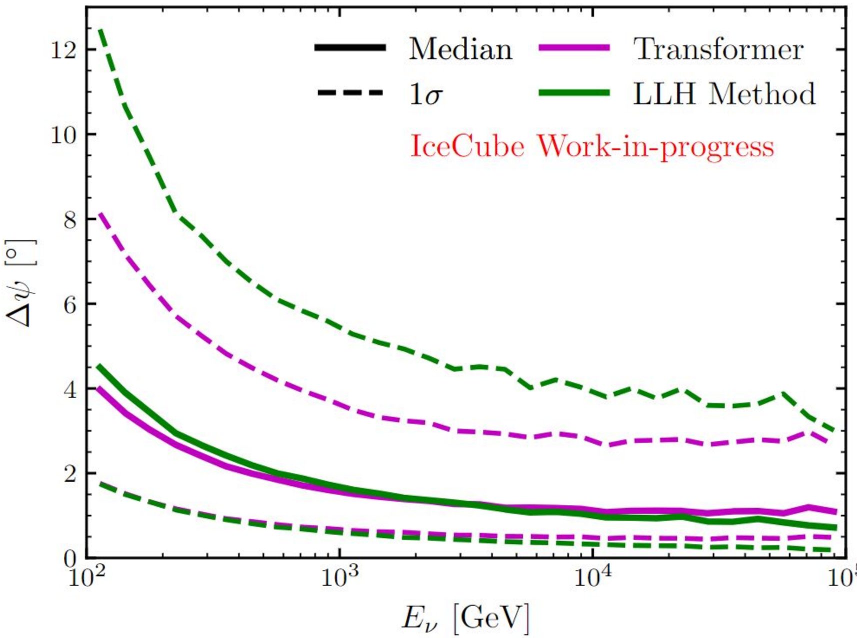


arXiv:2310.15674

Transformer Reconstructions

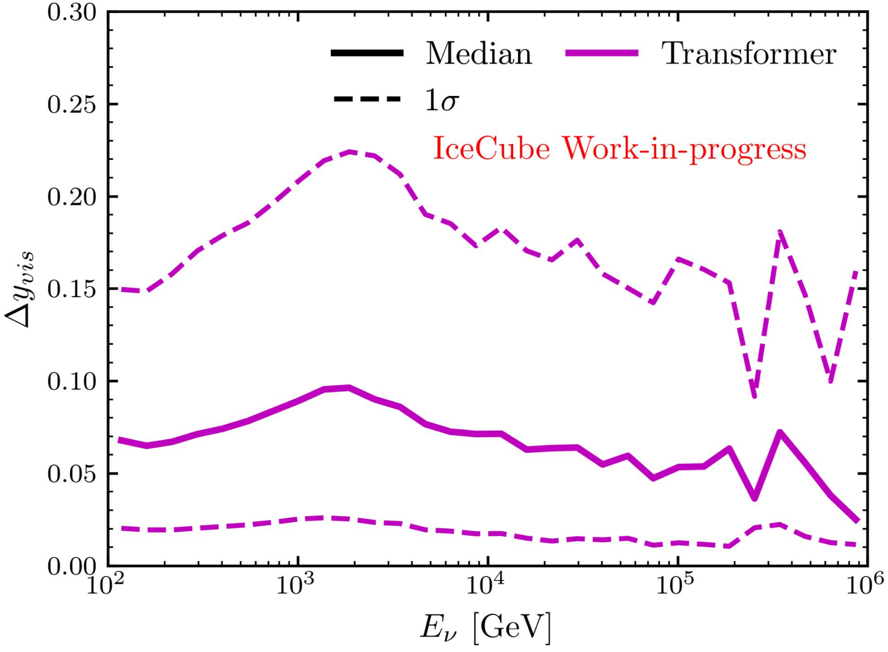
- Strong improvements for energy reconstruction, less so for the median directional error → related to training strategy





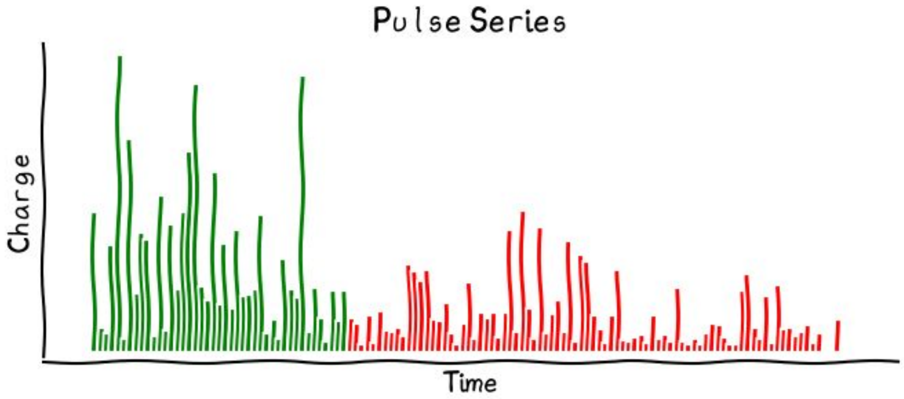
Transformer Inelasticity Reconstruction

- An exciting application is "visible" inelasticity reconstruction
  - Proxy variable using the detectable energy in the detector
  - Statistical nu/nubar separation, cross section measurements, tau neutrinos
- Outperforms previous architectures for the same task:
  - Random forest RMSE ~0.19
  - CNN RMSE ~0.17
  - Transformer RMSE ~0.13



Dealing with long sequences

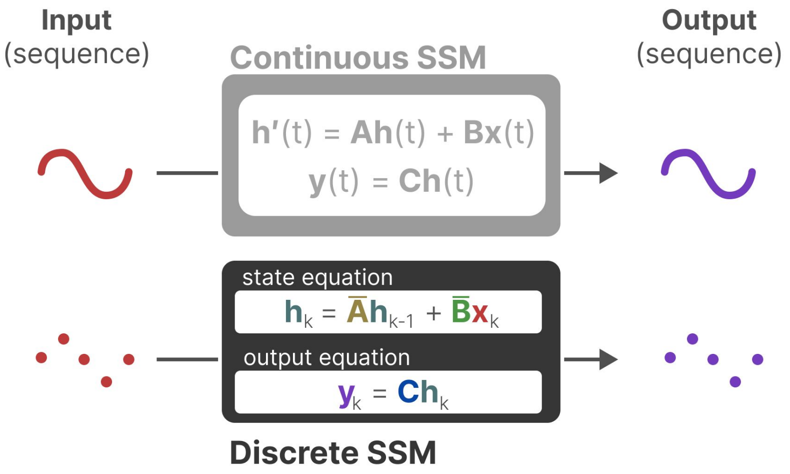
- For very large sequences, mainly high energy events, it is difficult to keep every pulse since the memory requirement scales quadratically
  - Naive implementation is to truncate after some number of pulses
  - Better methods exist, but removing any pulse is throwing away event information

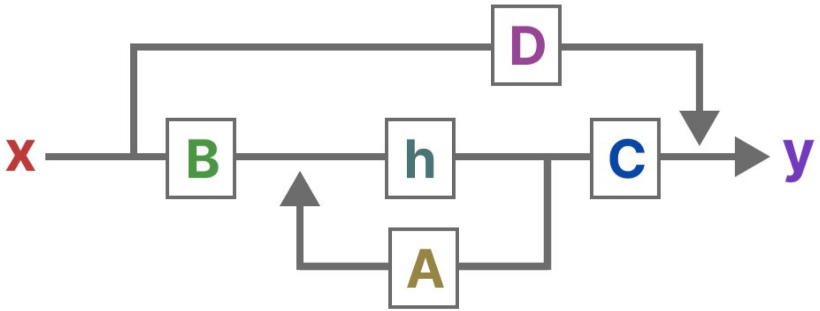


- Are there alternatives for long-sequence data?

State Space Models

- Stateful sequence-to-sequence model from classical control theory
  - Discretized with learnable parameters
- Has both a recurrent and convolutional representation
  - Fast training and fast inference
- Input has an ordering, does not require any positional encoding
- Generally, fewer parameters than transformer-based models for similar performance

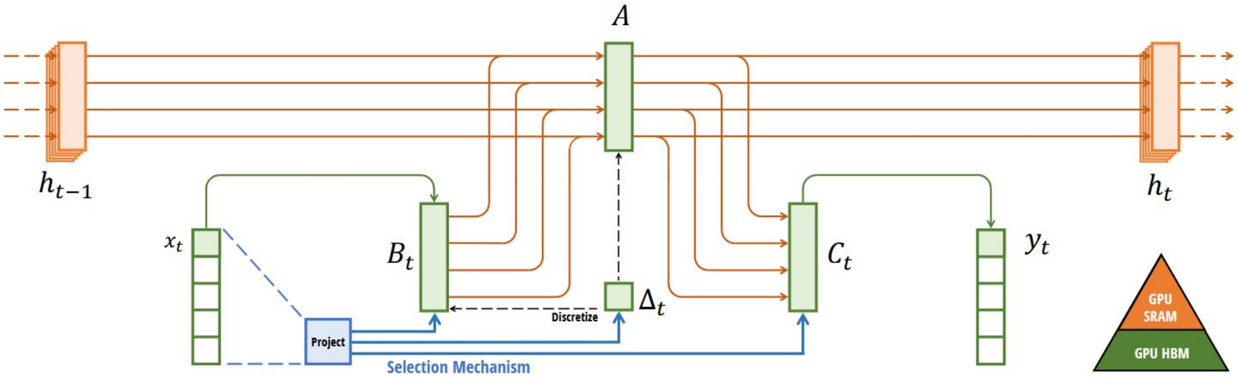


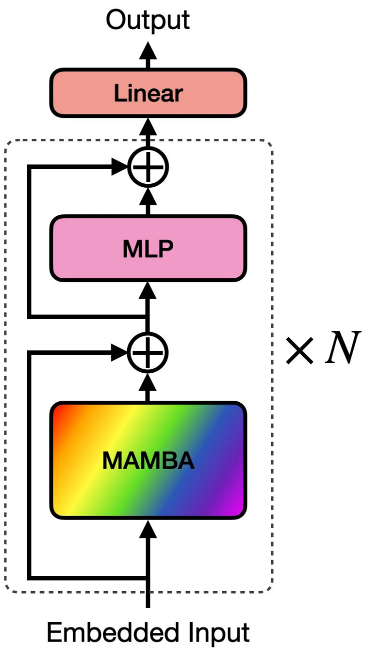


Figures from [M. Grootendorst]

MAMBA

- SSM+selection and hardware-aware algorithms
  - Selection mechanism → input-dependent sequence interactions
  - Very fast inference (scales linearly with sequence length)
- Good backbone architecture for long-sequence data
  - Nearly a drop-in replacement for MHA in a transformer model
- Does MAMBA work for neutrino reconstructions?

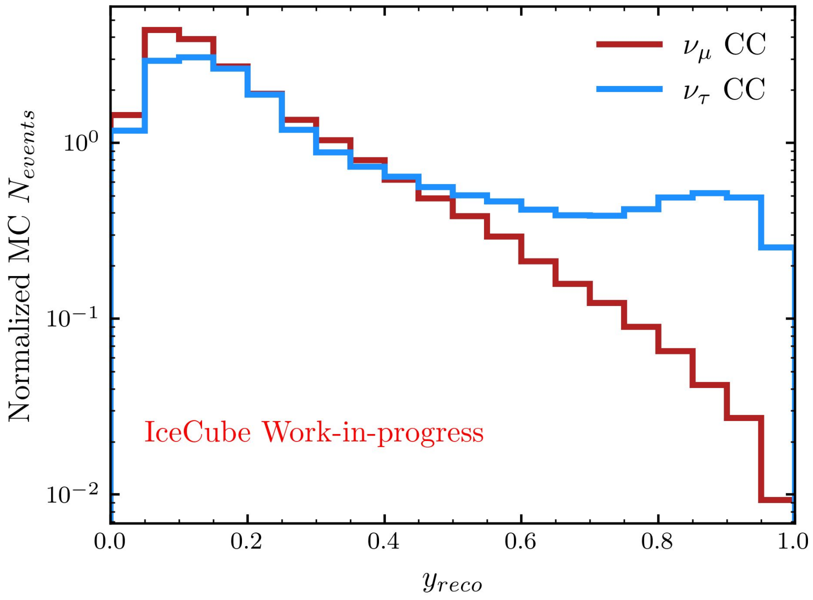


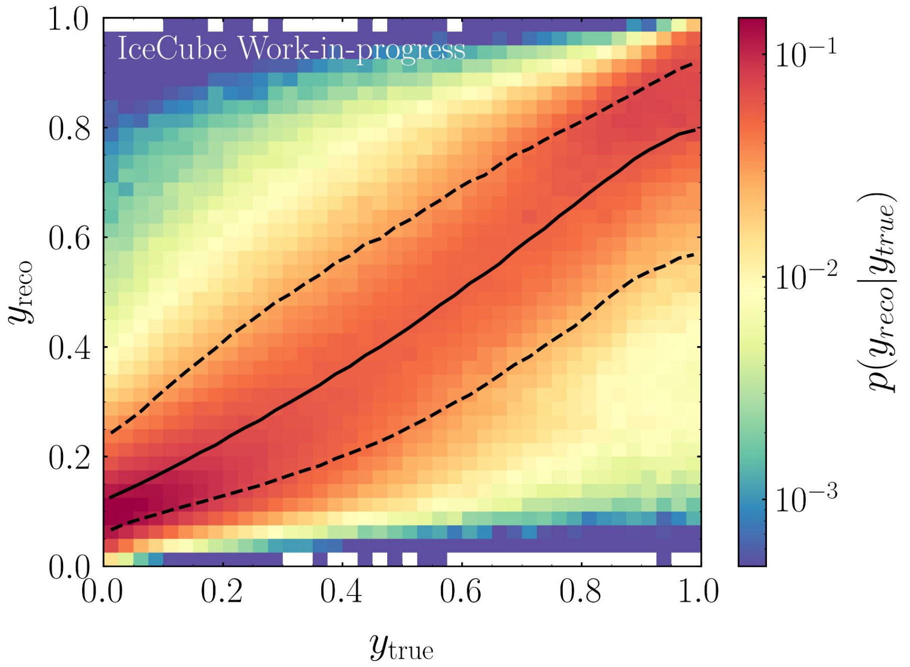


arXiv:2312.00752

Example: MAMBA Inelasticity Reconstruction

- ~40M parameter MAMBA model trained on CC muon neutrino events
  - Leverages fine-grained pulse series information without truncation
  - Comparable performance to transformers, ~5x less GPU memory, ~800 Hz inference





GraphNeT

GraphNeT

- The machinery developed for DNN-based reconstructions does not need to be specialized to each experiment
- The same technique employed by one experiment could be adapted to another experiment quite easily
  - A case for an open-source, cross-experiment collaborative effort

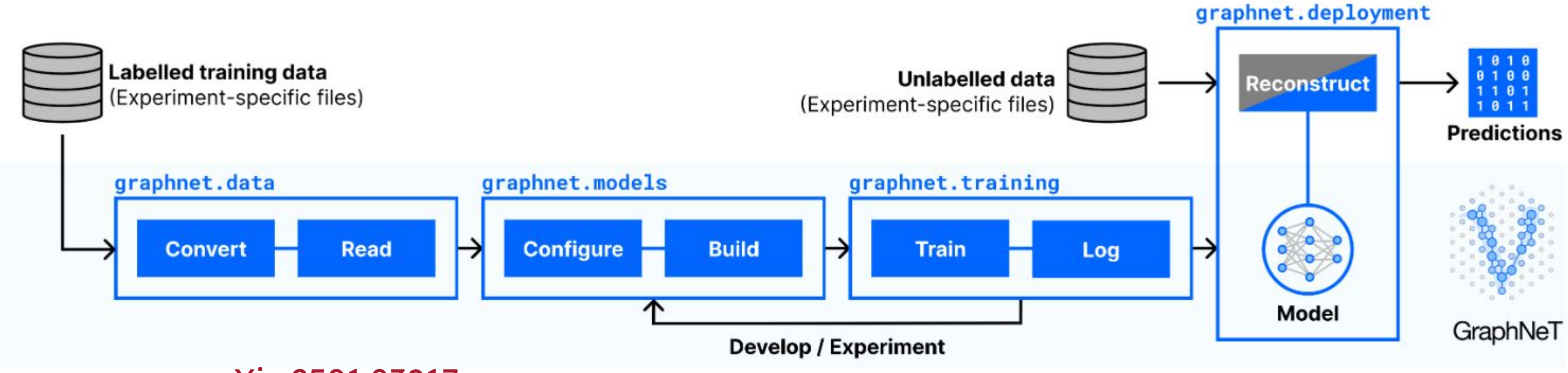
GraphNeT

Deep Learning for Neutrino Telescopes

https://github.com/graphnet-team/graphnet

GraphNeT Workflow

- Construct a model using the library of detectors, models, tasks
  - e.g. DynEdge + Direction Reconstruction
- Train the model using a labeled MC training sample
- Model can be applied to data using deployment modules, which can be integrated into different processing chains



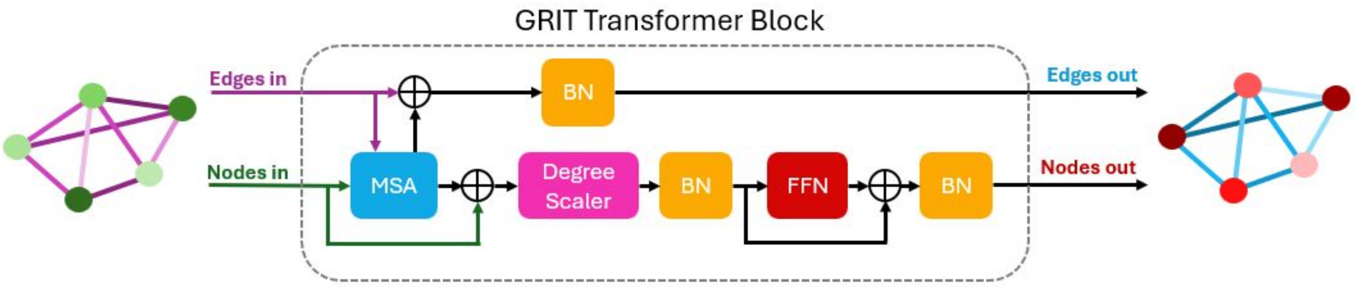
arXiv:2501.03817

A Few Implemented Architectures

- DynEdge
  - Graph convolutional neural network
  - DynEdge+Transformer model also available (Kaggle 1st place solution)
- IceMix
  - Transformer with sinusoidal position encoding, space+time attention bias
  - Implementation of Kaggle 2nd place solution
- ParticleNet
  - Graph convolutional neural network (based on arXiv:1902.08570)
- GRIT
  - Graph transformer model (based on arXiv:2305.17589)
- Normalizing flows
  - Implementation of models from jammy_flows

GRIT

- Graph transformer model that incorporates edge information into MHA and updates edge values
  - Based on the paper “Graph Inductive Biases in Transformers without Message Passing” (arXiv:2305.17589)
- Can incorporate different methods of absolute/relative position encoding (e.g. relative random walk encodings)
  - Encoding is not required, but expected to give a boost in performance
  - These methods may require significantly more GPU memory (larger graphs)

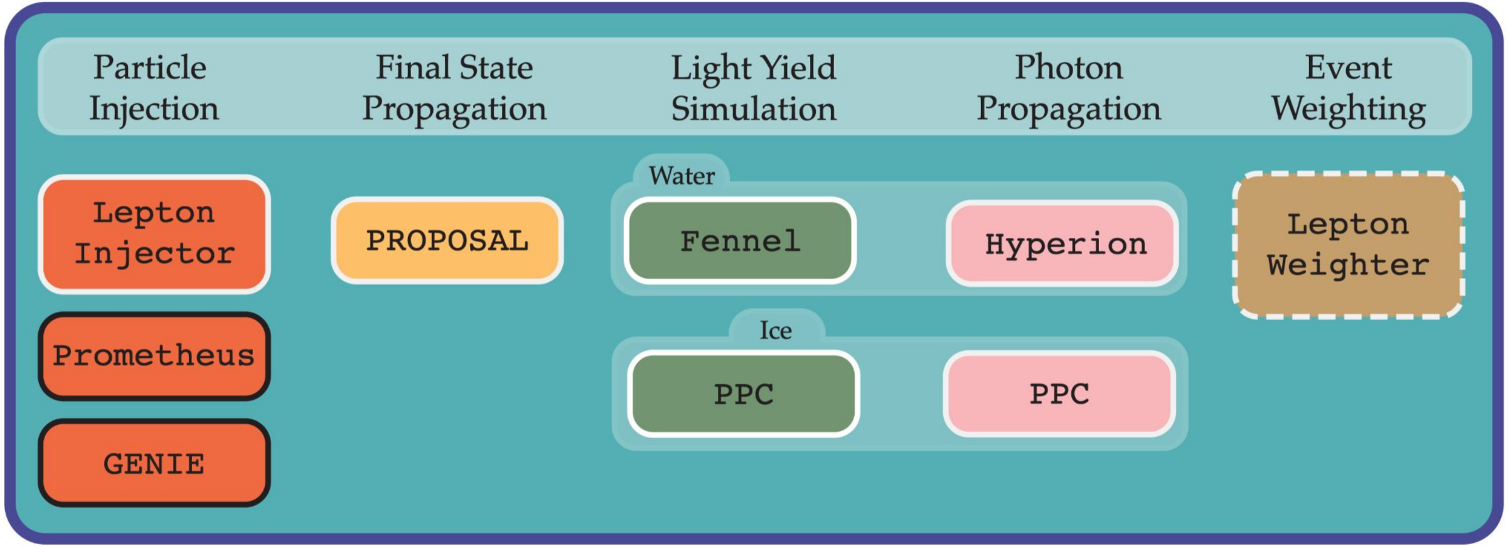


Which models are the best?

- No model is likely the "best" at everything, but there will be performance differences depending on the data
- Lots of choices beyond just the architecture
  - What is the best way to construct a graph of spatio-temporal data?
  - Can you use the full pulse series, or do you need to use summary statistics?
- Ongoing effort to benchmark these different architectures against several datasets → apples to apples comparison
  - Datasets generated using PROMETHEUS for different detector geometries
  - O(10M) events per dataset, neutral- and charged-current interactions
  - Simulation is simplified, does not contain every detector effect

PROMETHEUS: Open-source simulation

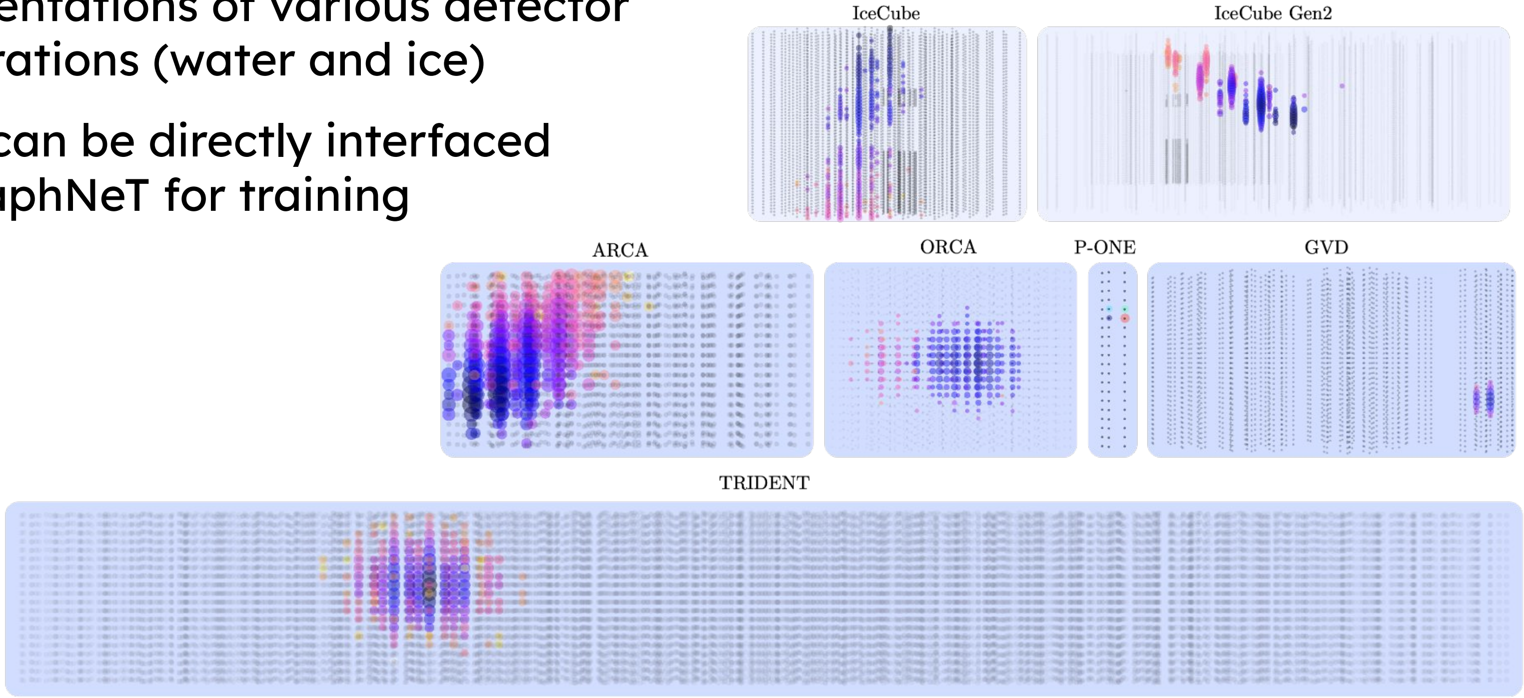
https://github.com/Harvard-Neutrino/prometheus



Comput.Phys.Commun. 304 (2024) 109298

PROMETHEUS: Open-source simulation

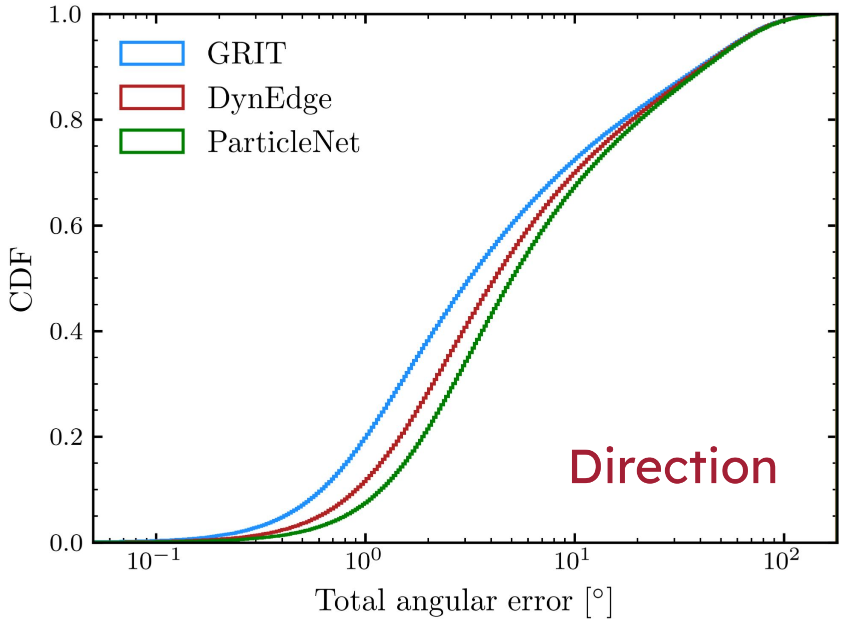
- Implementations of various detector configurations (water and ice)
- Output can be directly interfaced with GraphNeT for training

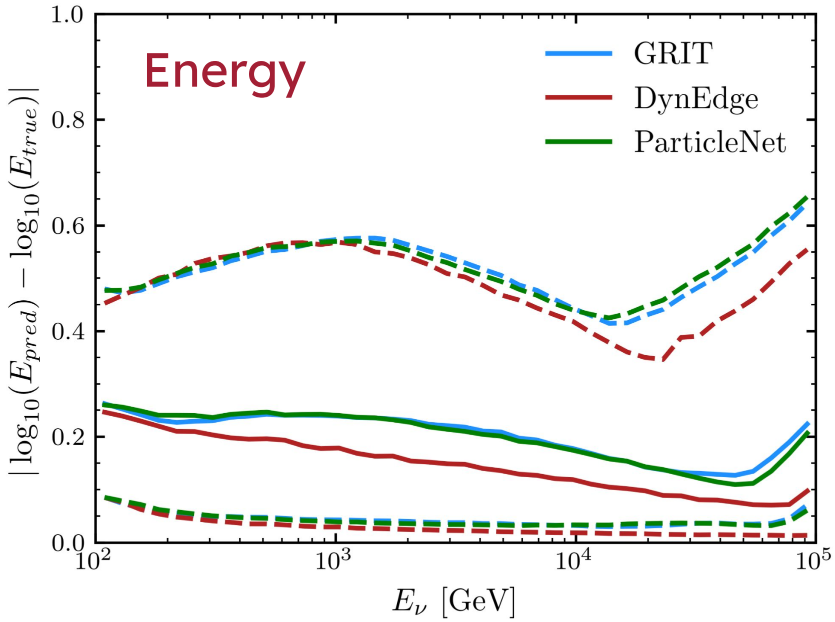


Comput.Phys.Commun. 304 (2024) 109298

Preliminary Model Comparisons

- Active effort to evaluate the performance of each architecture on different tasks and different detector configurations
  - Still a work-in-progress, results may change!





Conclusions

- Many analyses in IceCube are now leveraging advances in ML-based reconstruction and classification techniques
  - Showed only a small selection of results here, there are many more applications of these methods that I did not have time to show!
- The state-of-the-art continues to evolve quickly
  - New architectures and techniques pop up nearly every day
- There is an active effort to develop and maintain an open-source and cross-experiment machine learning framework: GraphNeT
  - Consider implementing your experiment!

Thank you for listening!