Data/MC adaptation with adversarial training in KM3NeT

João Coelho

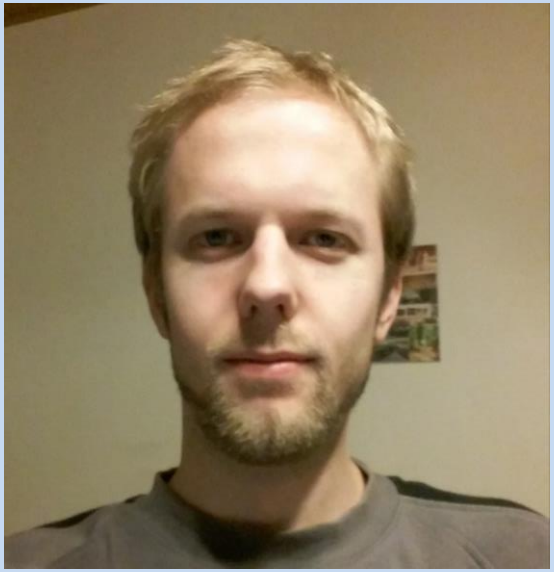
28 September 2022

This work is supported by the Data Intelligence Institute of Paris (diiP), IdEx Université de Paris, ANR-18-IDEX-0001

Collaborators

Shen Liang
Univ. Paris Cité
Postdoc (diiP)
DANN Implementation

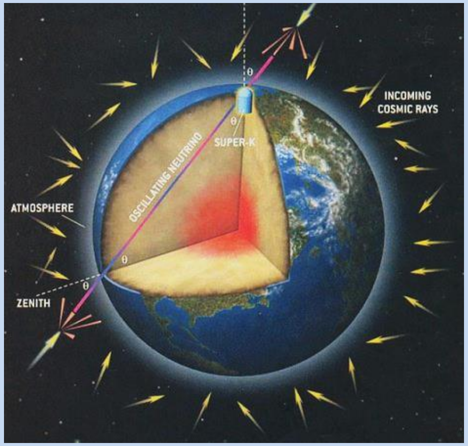
Santiago Peña Martínez
APC Laboratory
PhD Student
GNN Development

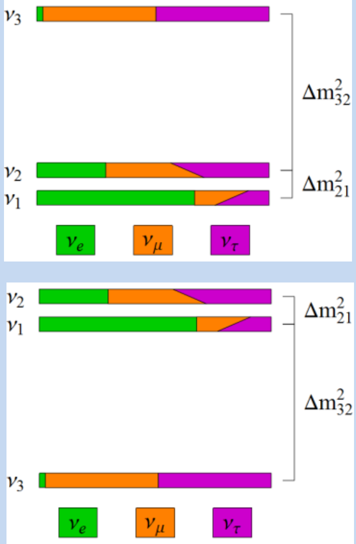


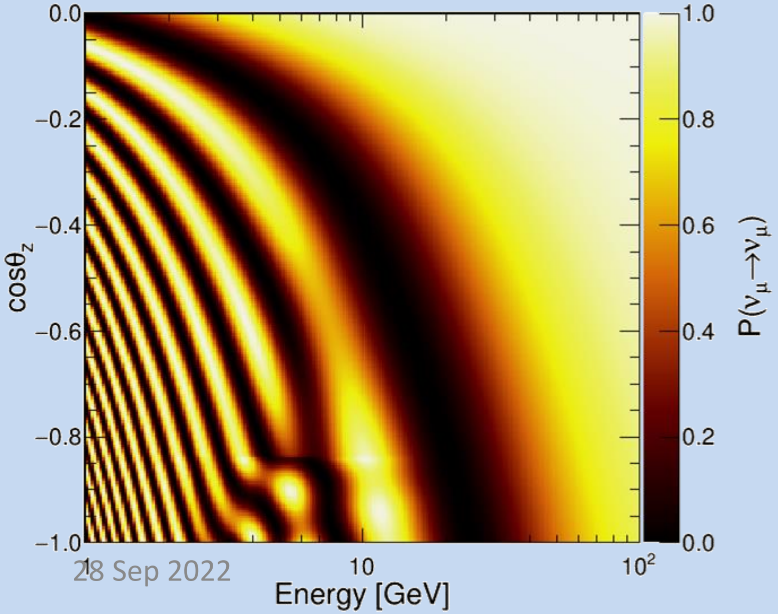
Daniel Guderian
Univ. of Münster
PhD Student (Graduated)
GNN Development
PhD Thesis

KM3NeT Experiment

Measure atmospheric neutrino oscillations

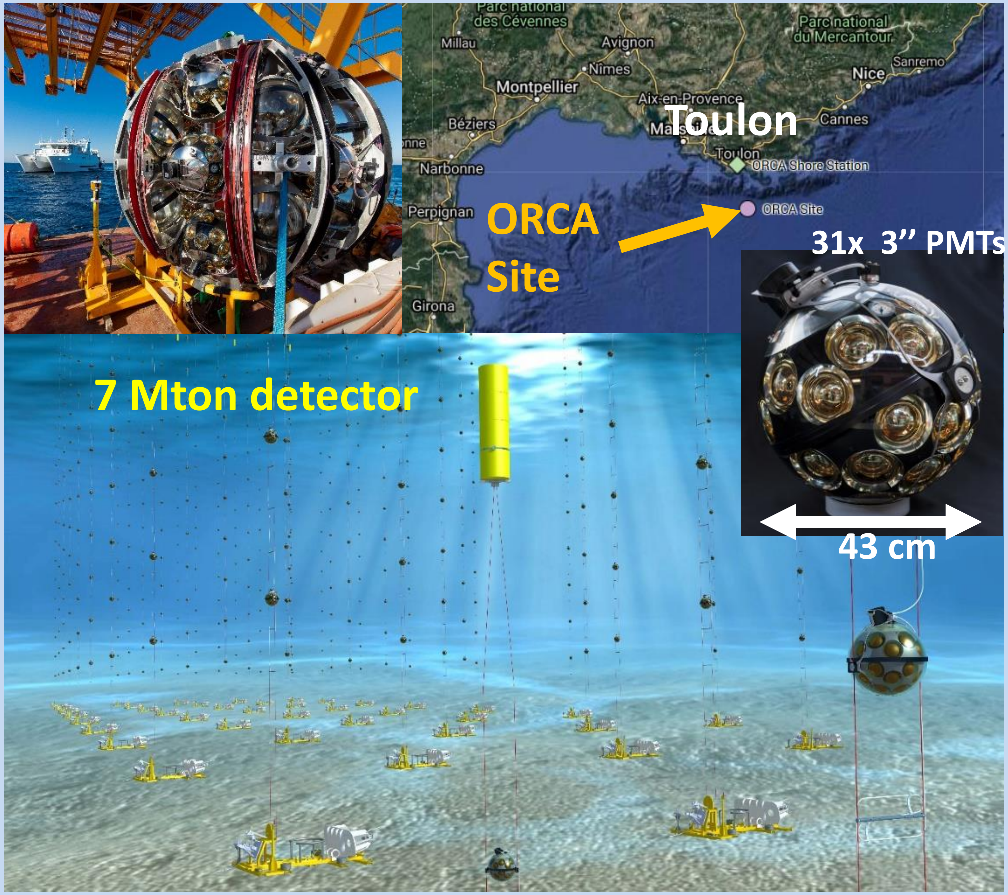






Main goal*:

What is the Neutrino Mass Ordering?



Toulon

ORCA Site

31x 3'' PMTs

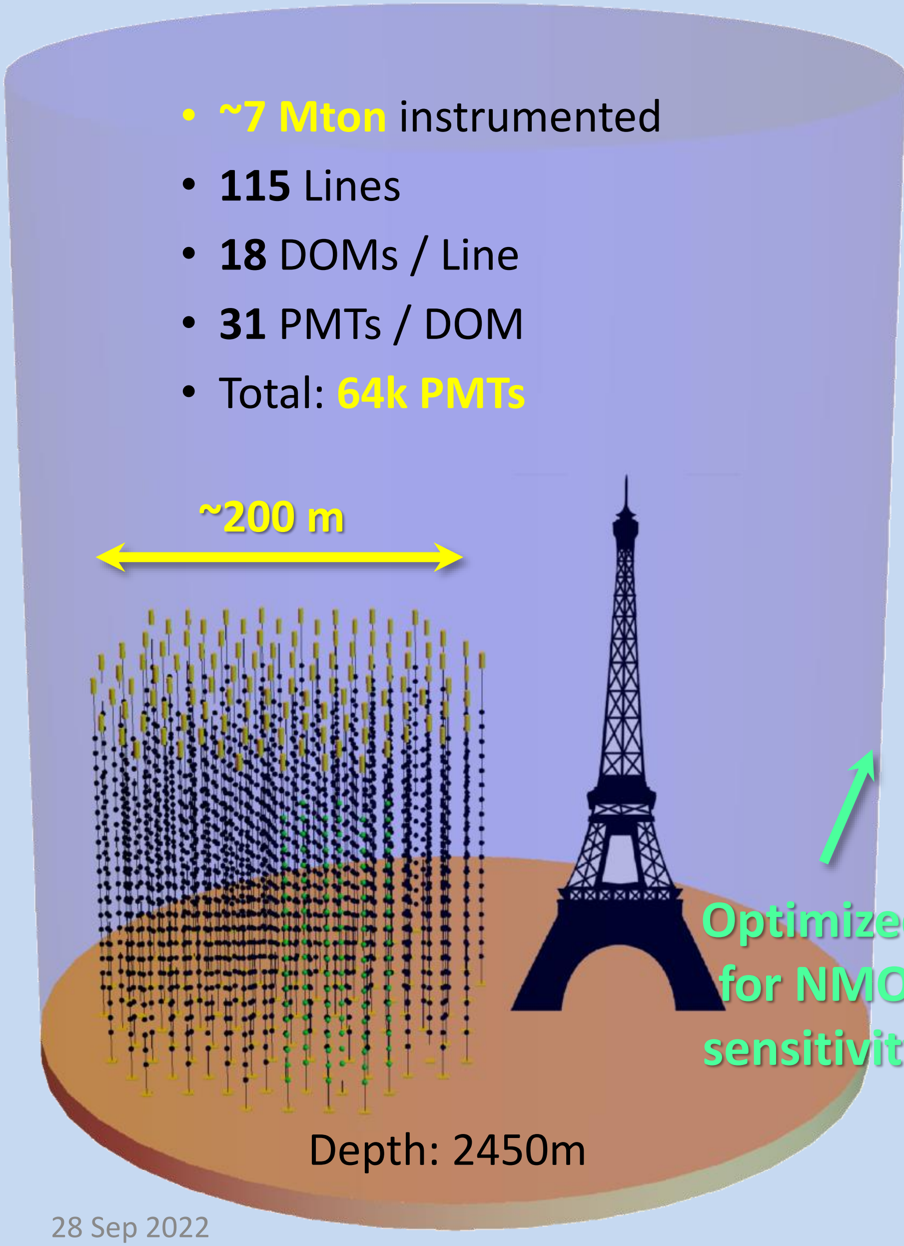
43 cm

7 Mton detector

*Also contains a neutrino astronomy branch

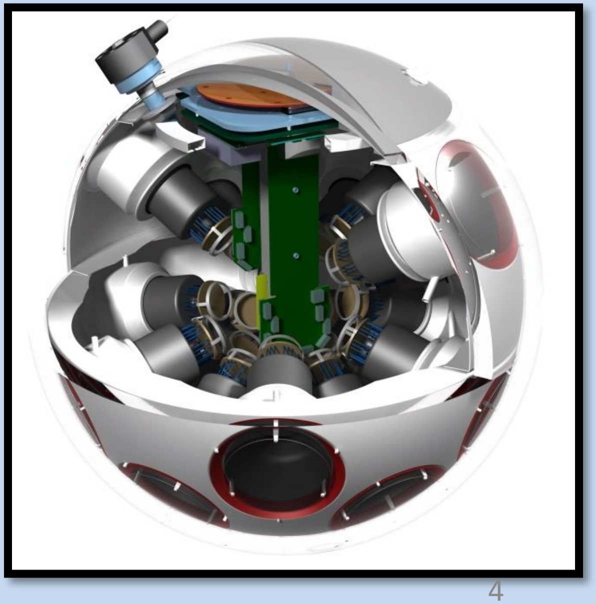
The KM3NeT/ORCA Detector

- **~7 Mton** instrumented
- **115** Lines
- **18** DOMs / Line
- **31** PMTs / DOM
- Total: **64k PMTs**



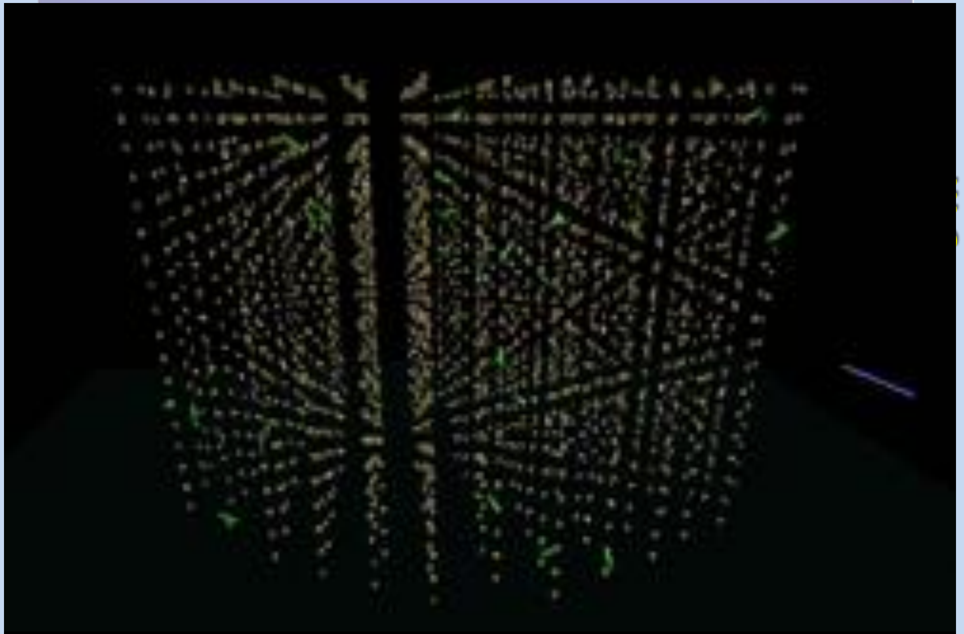
Depth: 2450m

Optimized for NMO sensitivity

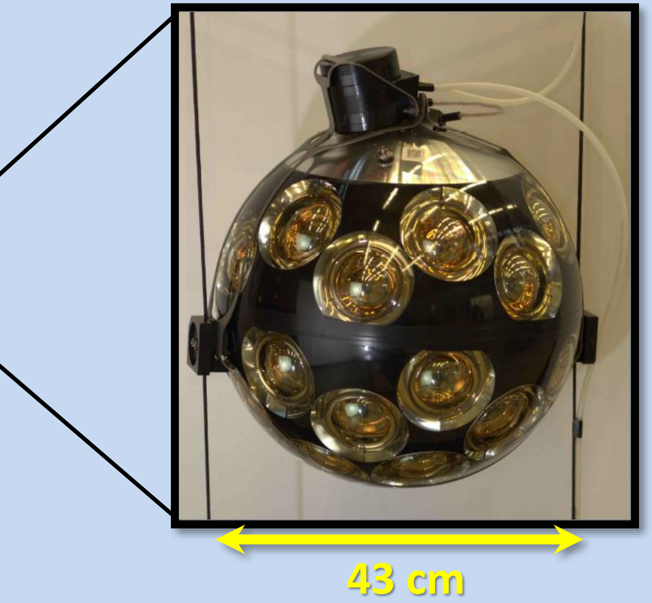


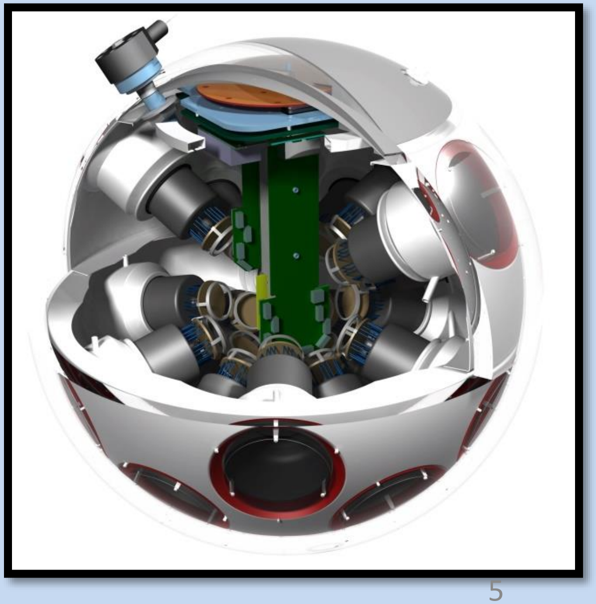
The KM3NeT/ORCA Detector

- **~7 Mton** instrumented
- **115** Lines
- **18** DOMs / Line
- **31** PMTs / DOM
- Total: **64k PMTs**

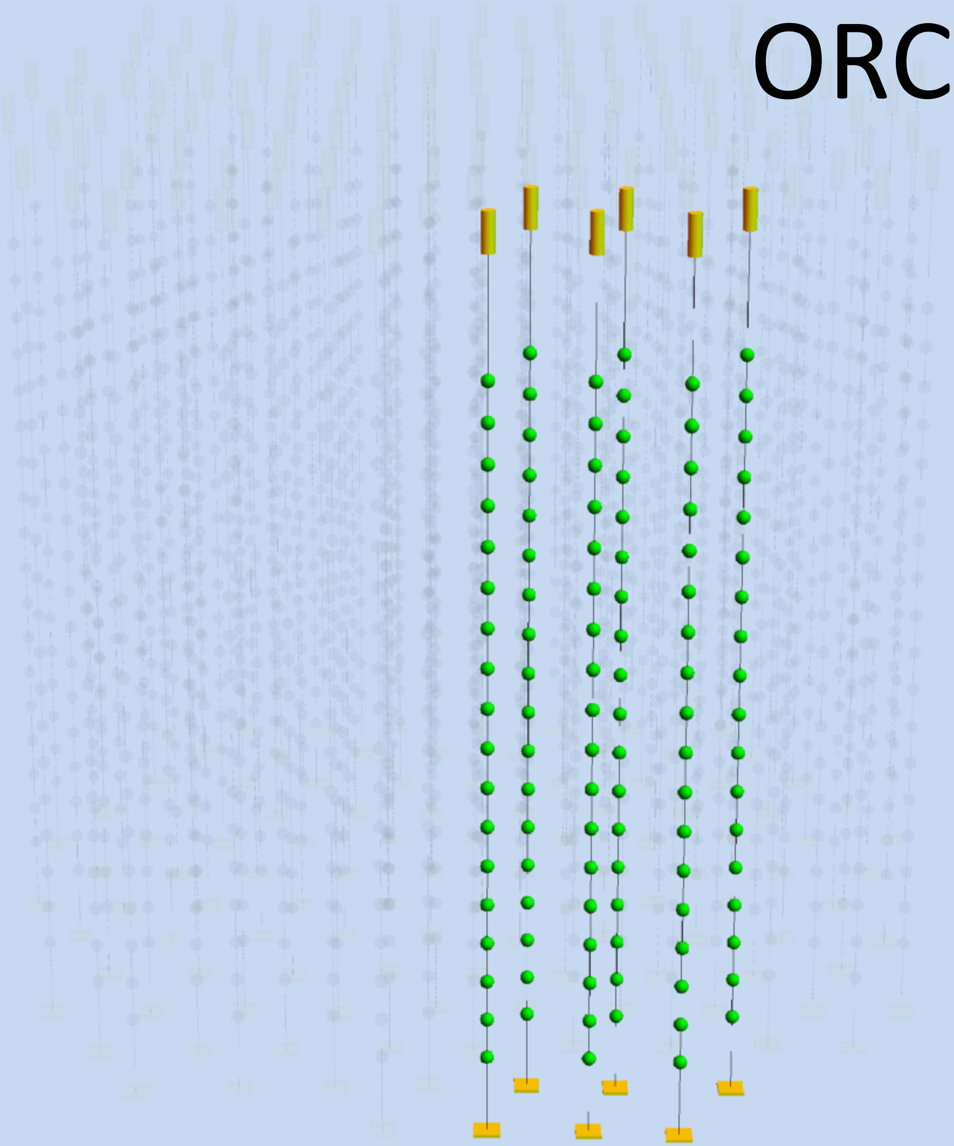


http://www.cherenkov.nl/



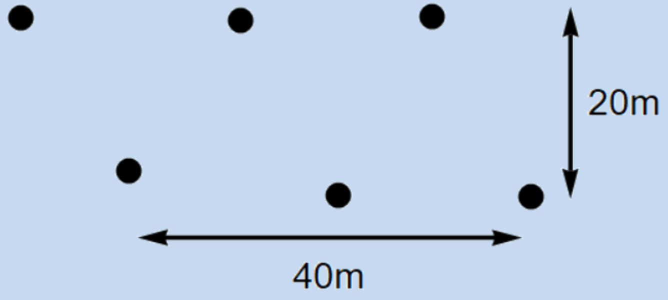


Current Work Focus: ORCA6

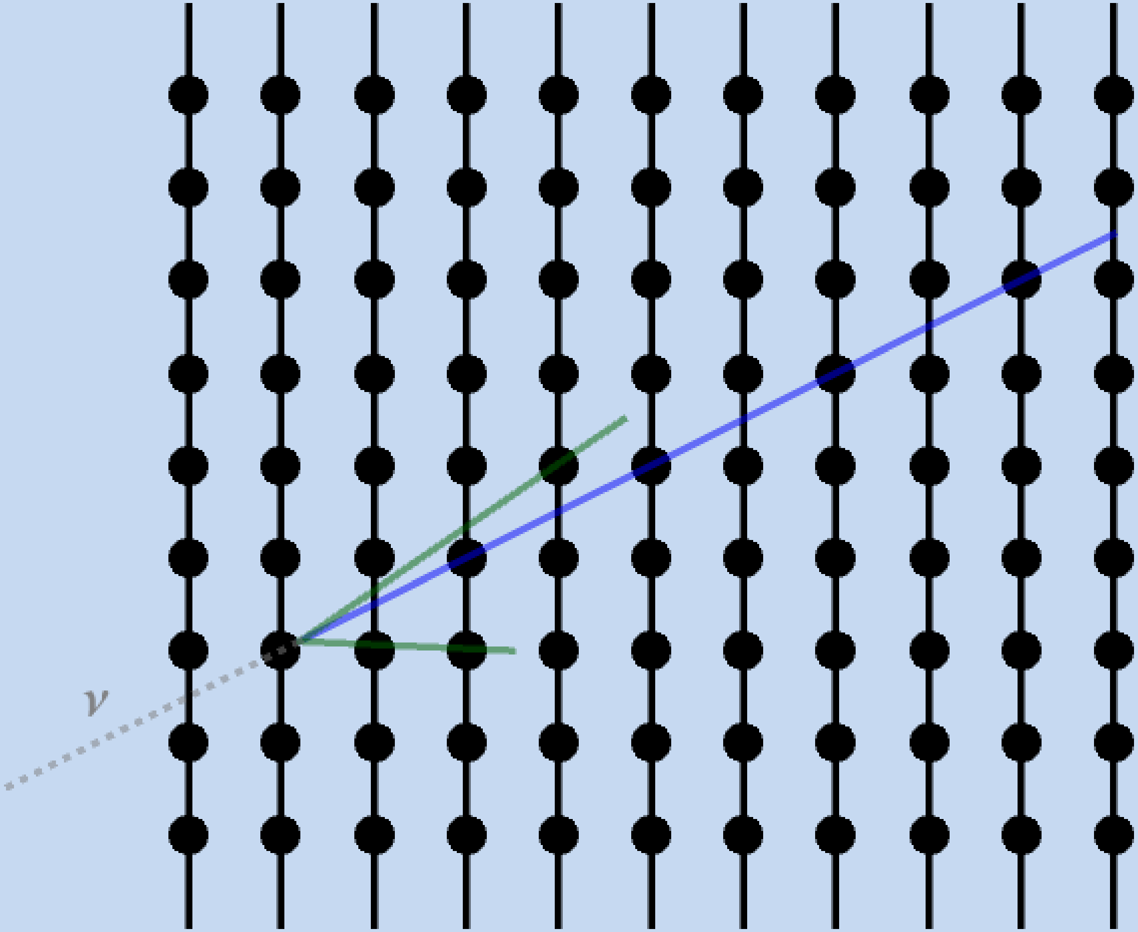


- Took data from Jan 2020 to Nov 2021
- Total of ~560 days available
- Most work done on 355 days sample
- Instrumented mass: 364 kton
- Total of 3348 PMTs

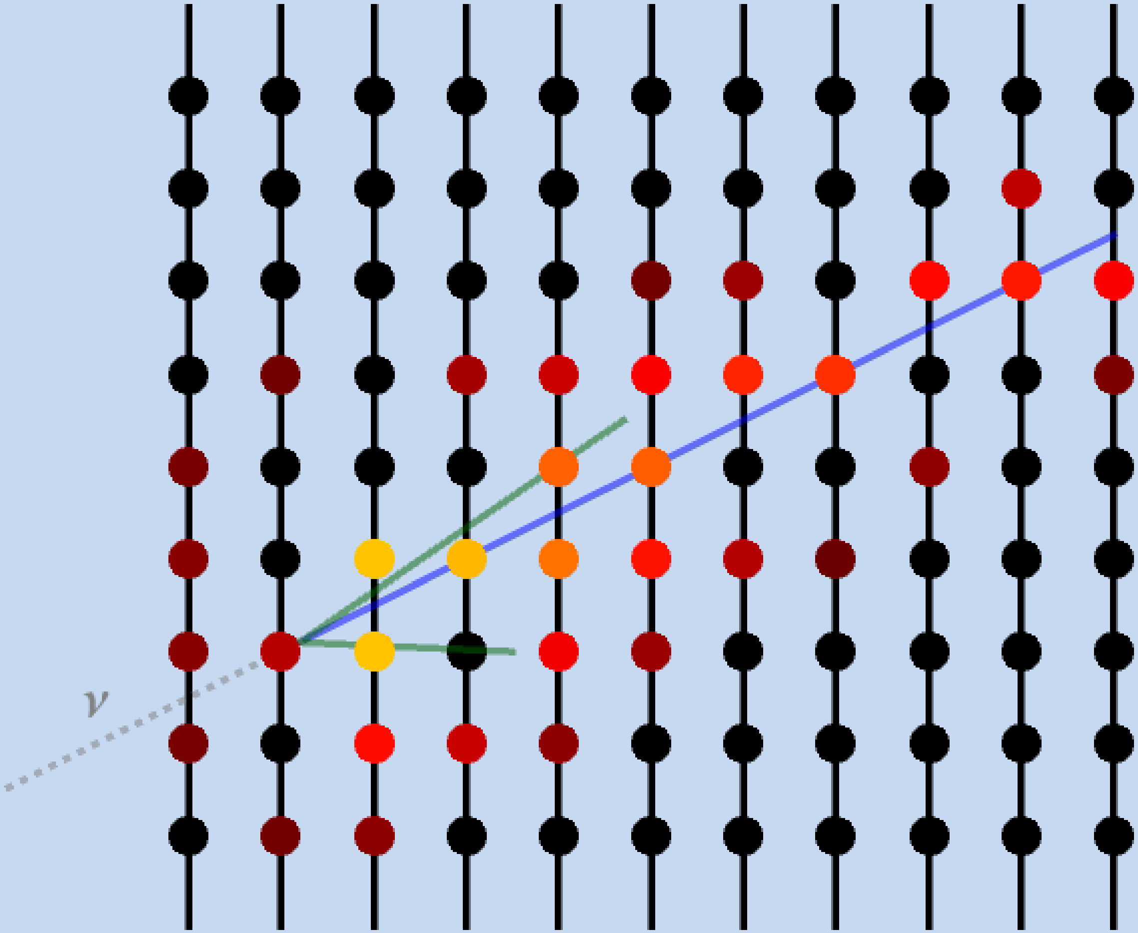
Footprint



Neutrino Events



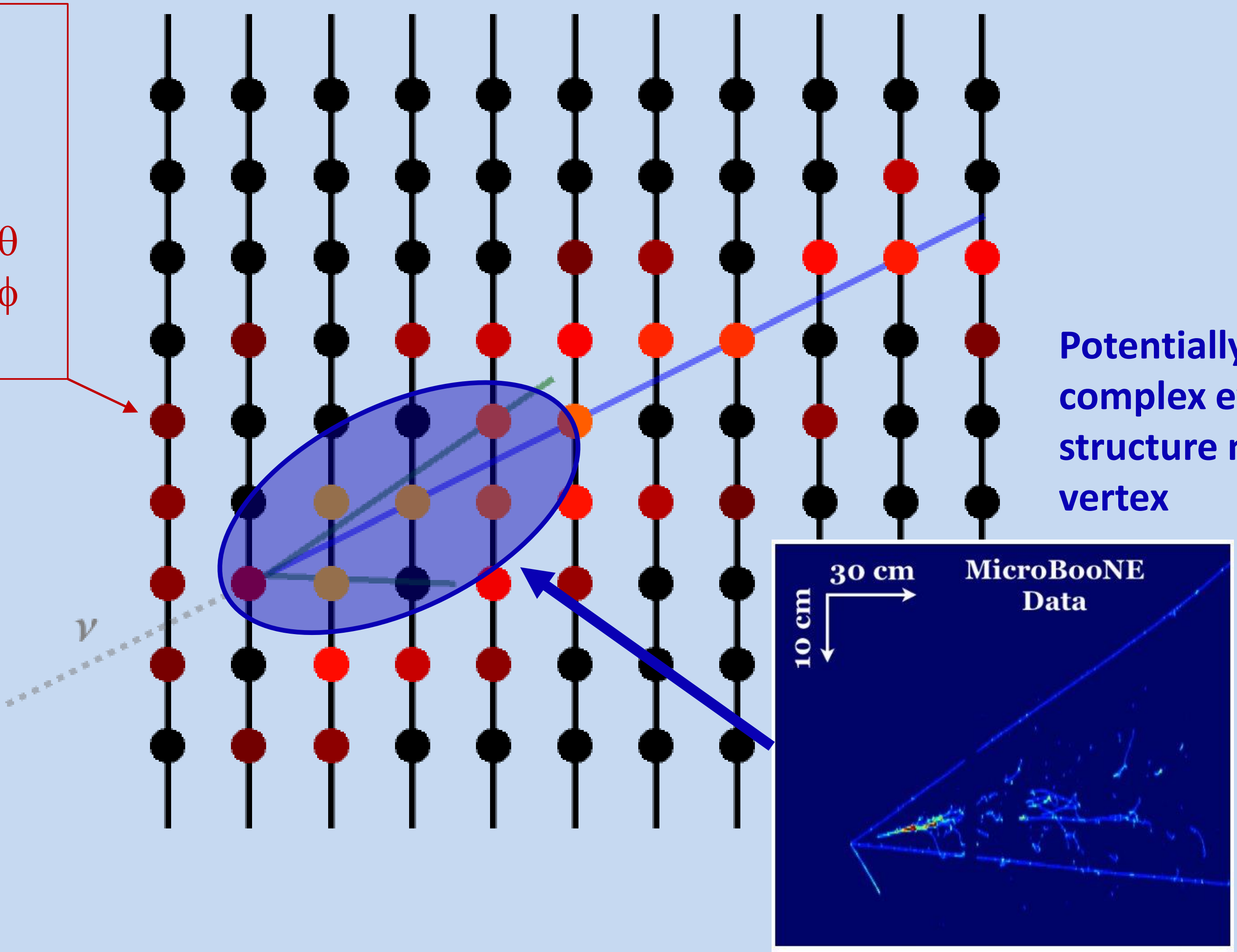
Neutrino Events



Neutrino Events

7D Object

- Charge
- Position X
- Position Y
- Position Z
- PMT Angle θ
- PMT Angle ϕ
- Time

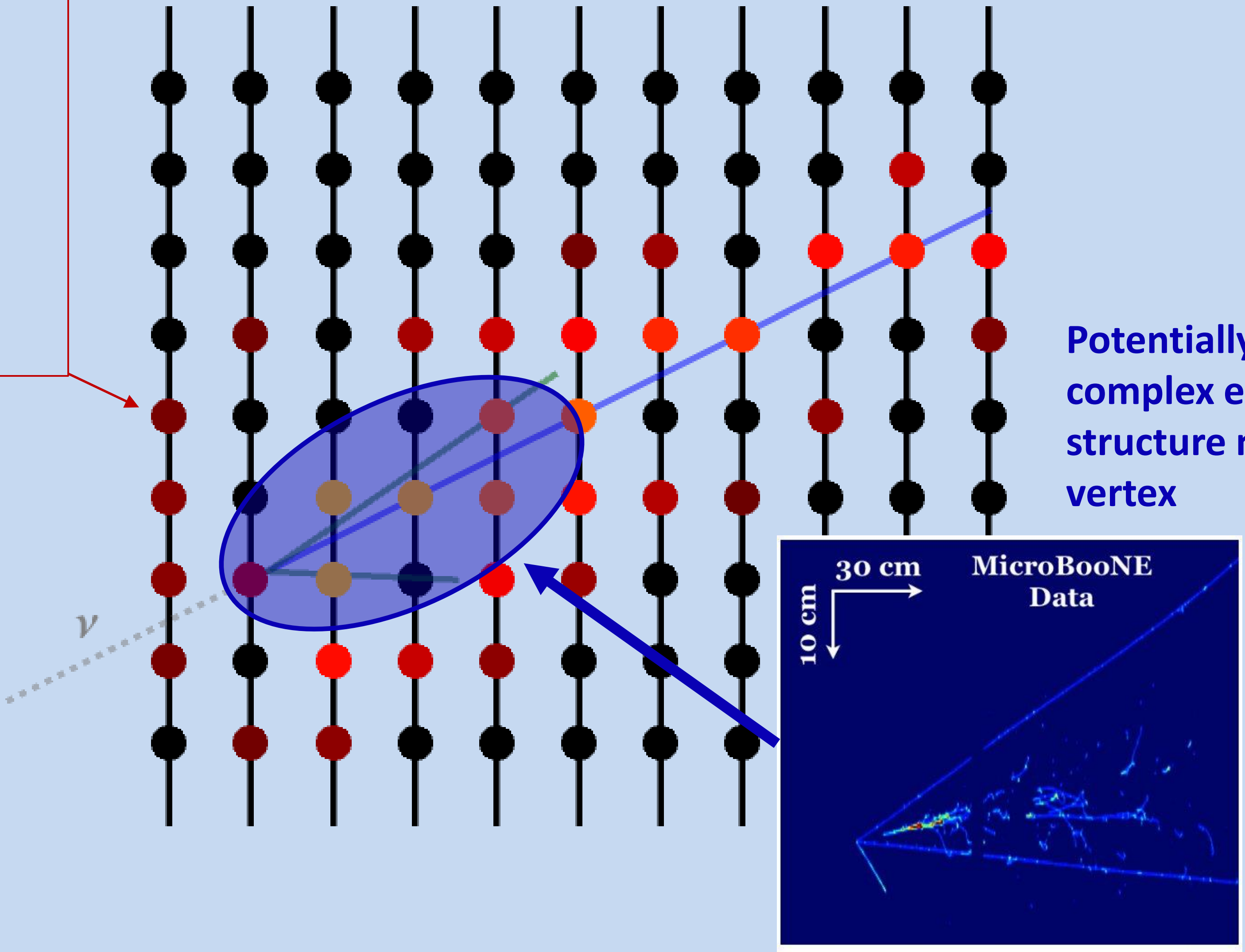


Potentially complex event structure near vertex

Neutrino Events

Current Implementation

- ~~Charge~~
- Position X
- Position Y
- Position Z
- PMT Dir. X
- PMT Dir. Y
- PMT Dir. Z
- Time

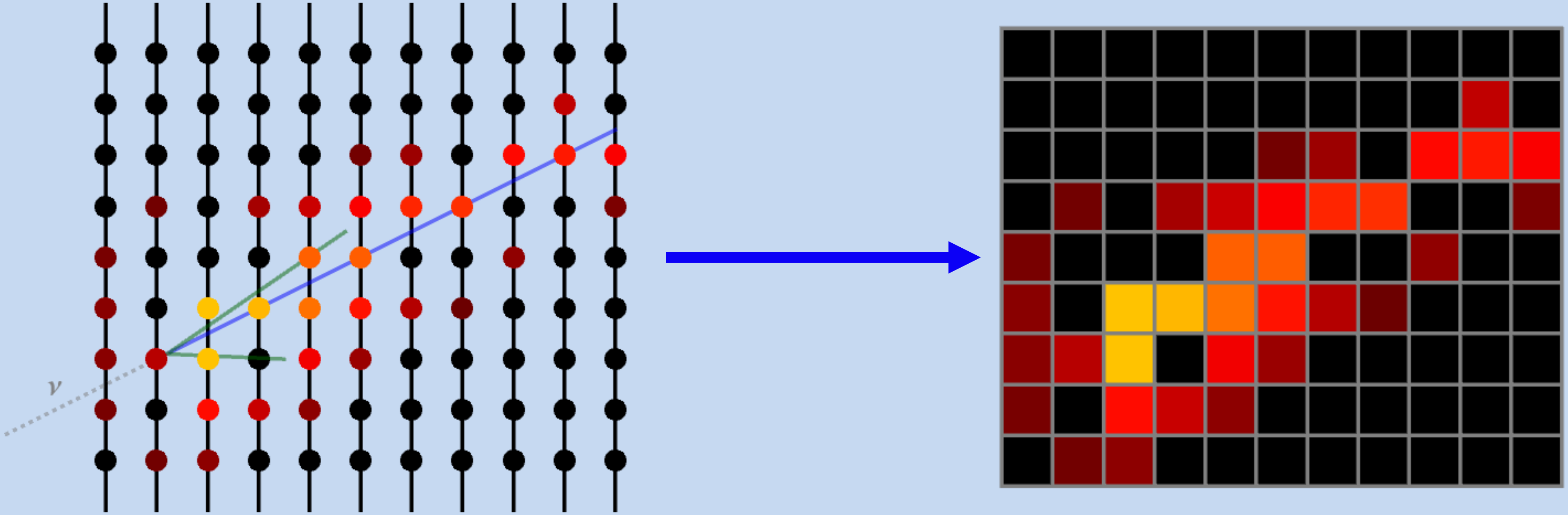


Potentially complex event structure near vertex

Deep Learning Reco

Initial work:
Convolutional NN

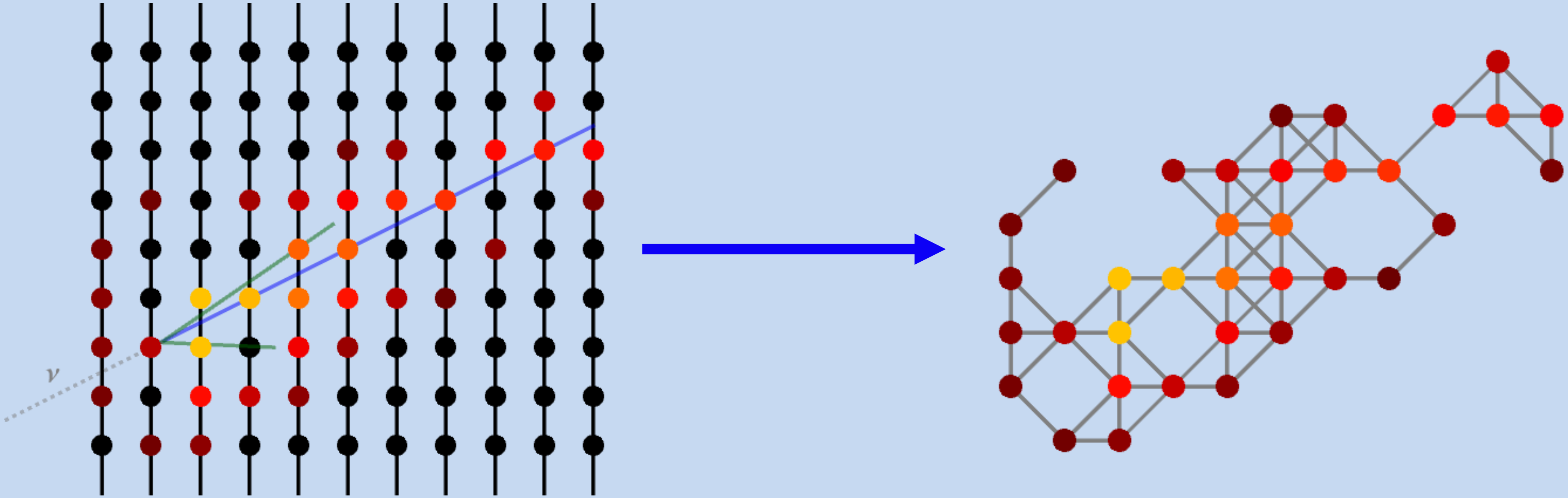
- Fast
- Translation invariance
- 2D or 3D input
- Loss of information (usually)
- Rigid grid structure
- Natural treatment of empty pixels



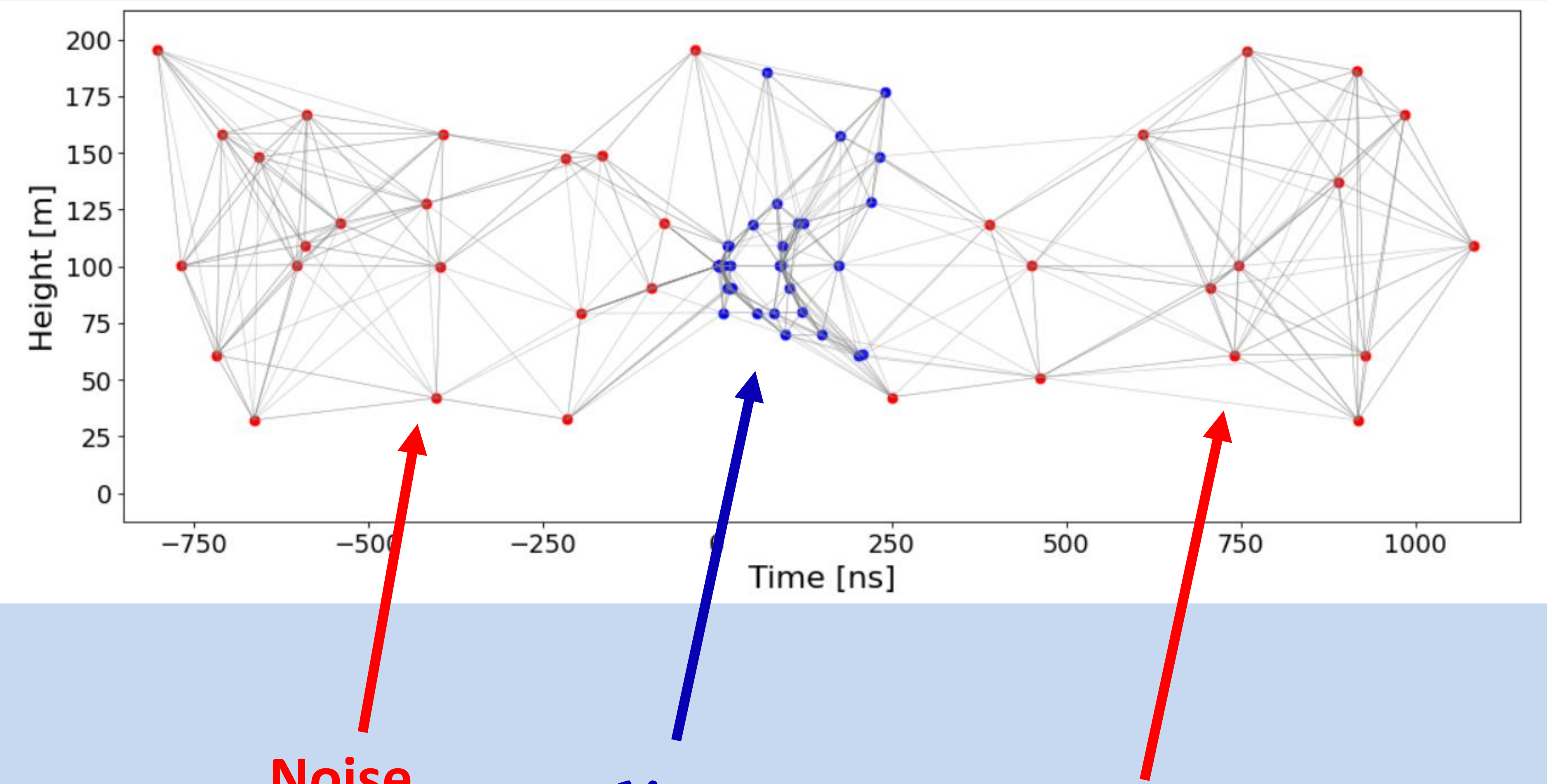
Deep Learning Reco

Better representation:
Graph NN

- Slower
- Translation and rotation invariance
- N-D input
- No loss of information
- Flexible structure



Graph Example



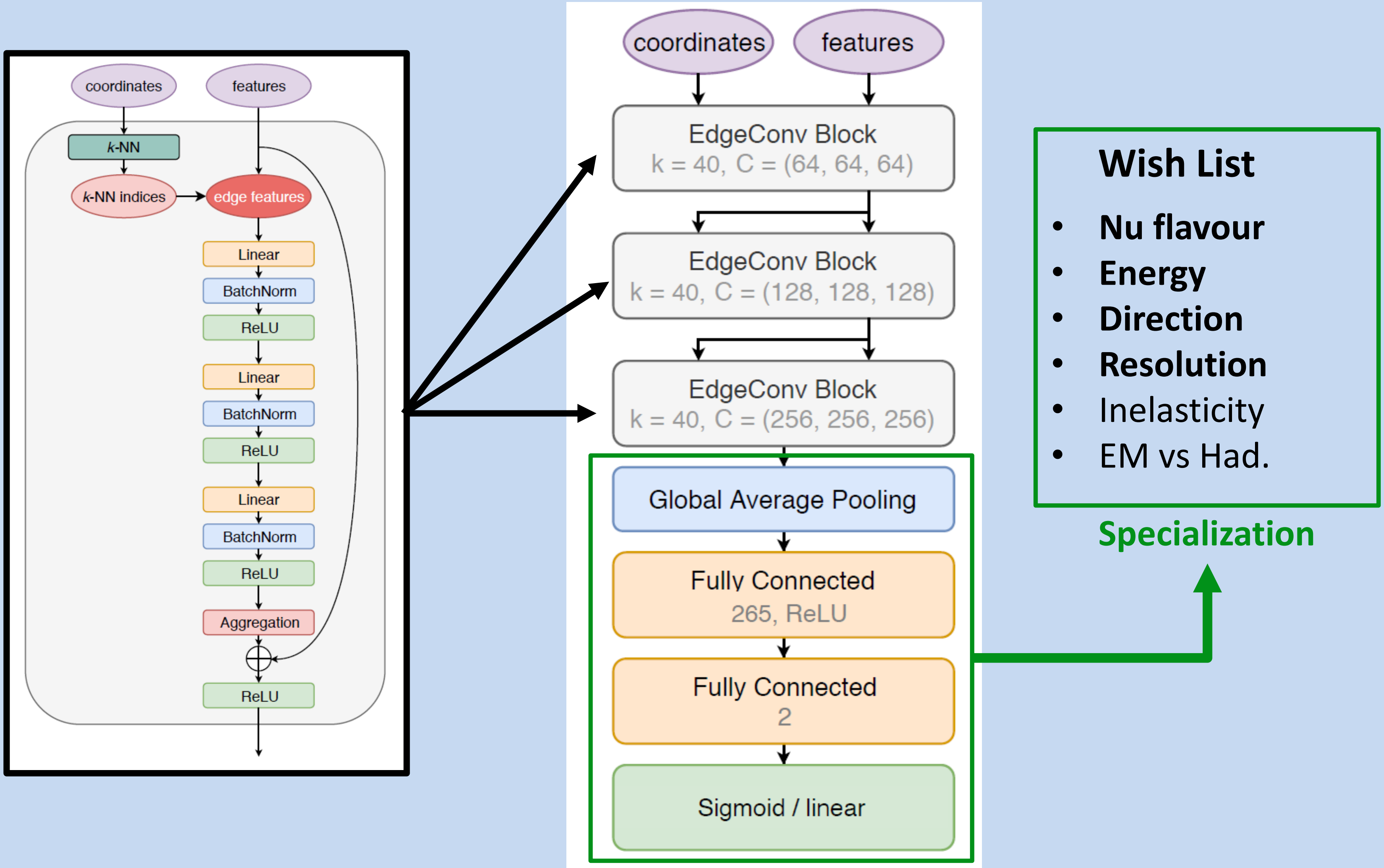
Noise

ν_μ

Noise

Graph Convolution

- Core layer implemented from [ParticleNet: PRD 101, 056019 (2020)]



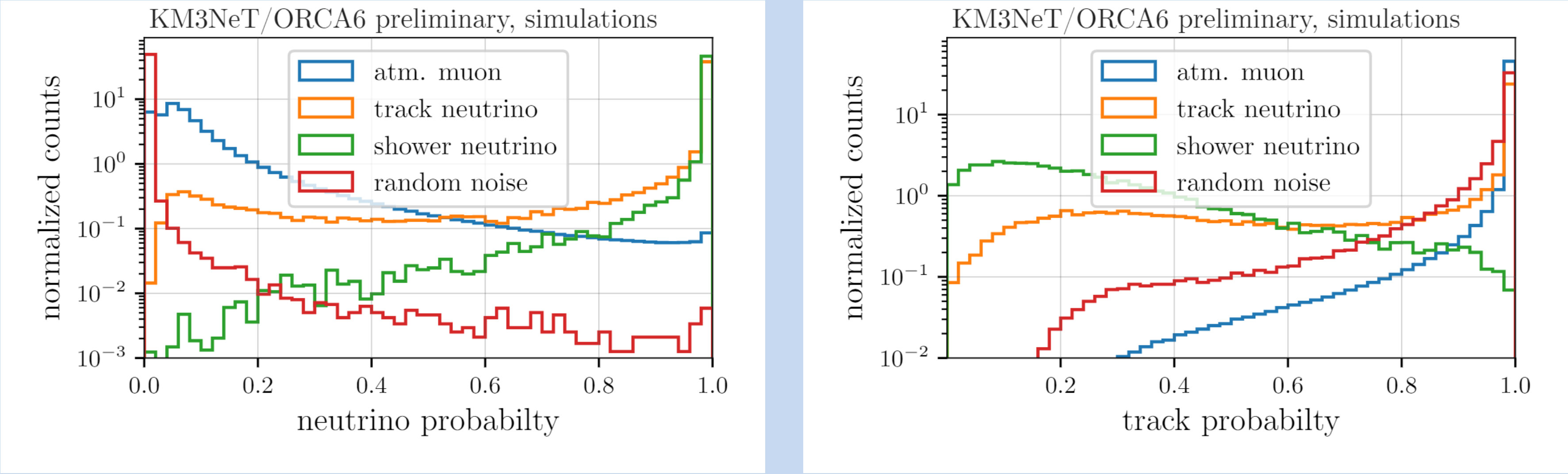
Wish List

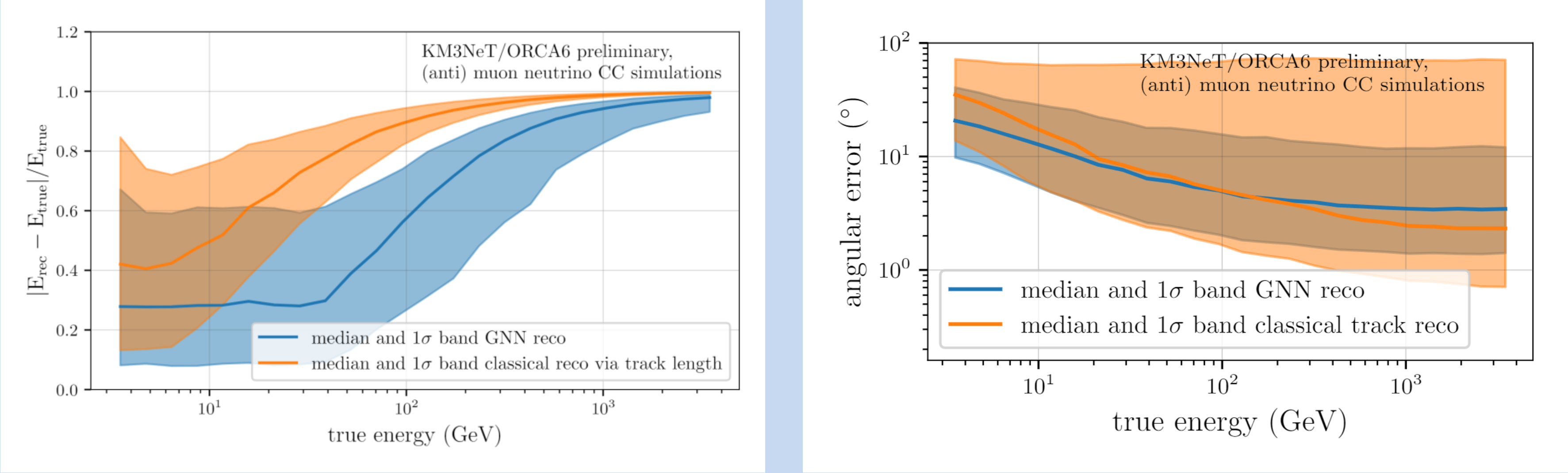
- **Nu flavour**
- **Energy**
- **Direction**
- **Resolution**
- Inelasticity
- EM vs Had.

Specialization

Reconstruction Results

Classification

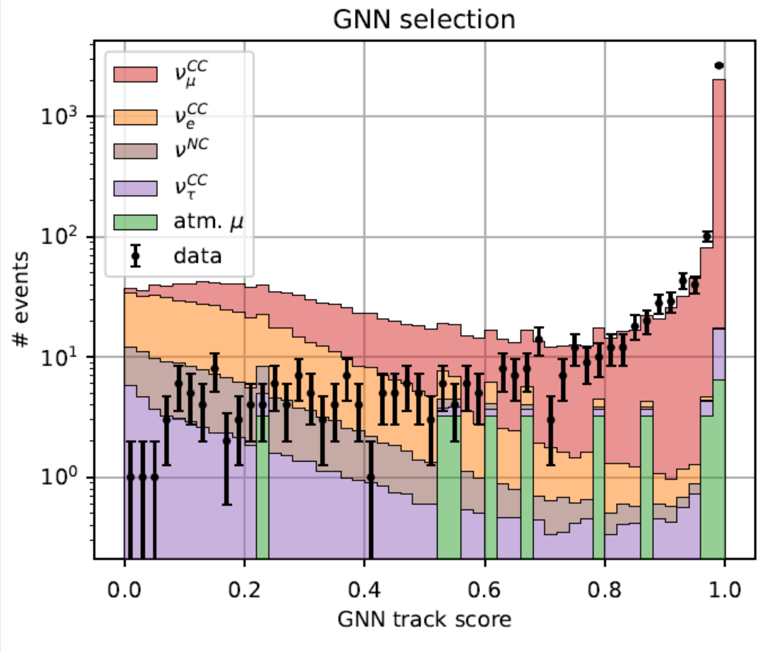


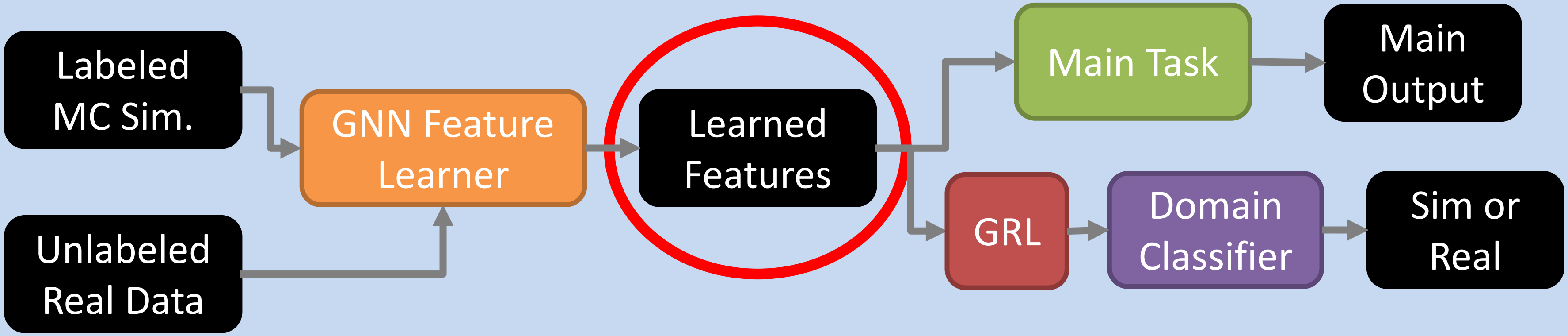


Regresssion

MC is always wrong

- Mostly we find the GNN prediction to show reasonable agreement between data and MC, but some tasks show disagreement
- Working on implementing the idea from the DANN paper: [JMLR 2016, vol. 17, p. 1-35]
- Adversarial training to push GNN to ignore differences in the input distributions between data and MC





Prevent learning Data/MC differences

Labeled MC Sim.

Unlabeled Real Data

GNN Feature Learner

Learned Features

Main Task

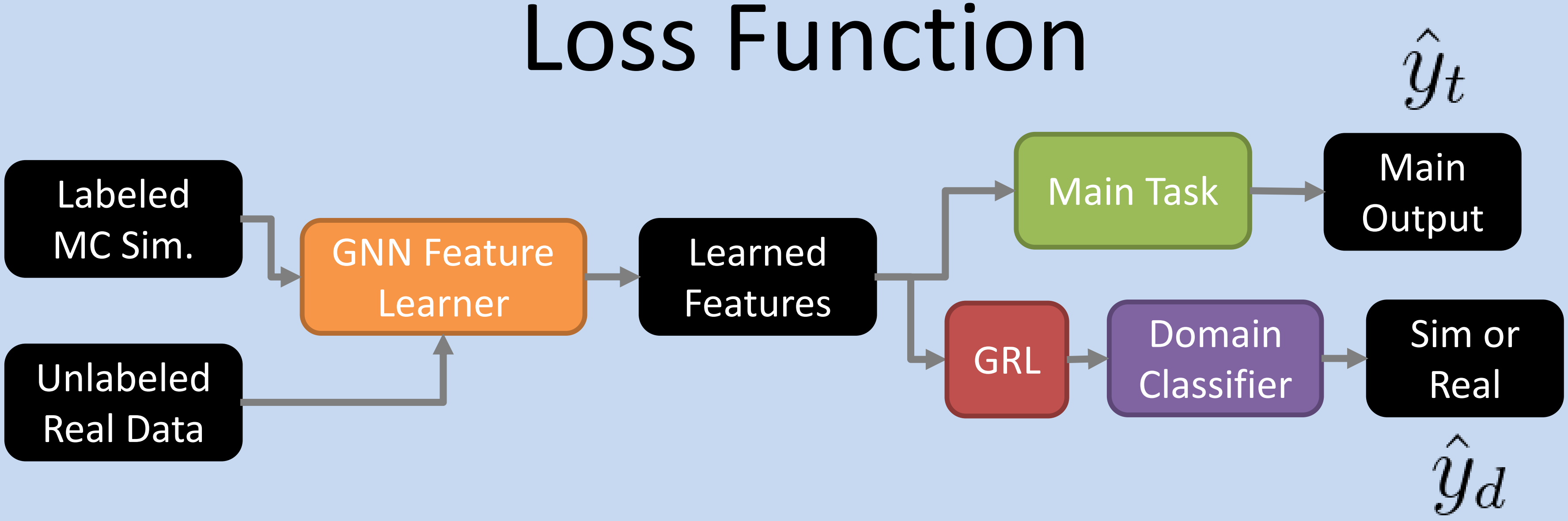
Main Output

GRL

Domain Classifier

Sim or Real

Loss Function



$\hat{y}_t$

Labeled MC Sim.

Unlabeled Real Data

GNN Feature Learner

Learned Features

Main Task

Main Output

GRL

Domain Classifier

Sim or Real

$\hat{y}_d$

Zero if unlabeled

minimize

minimize

Not available for real data

$$y_d \; \mathcal{L}_t(G_t(G_f(x;\boxed{\theta_f});\boxed{\theta_t}), y_t) +$$

$$\lambda \; \mathcal{L}_d(G_d(\boxed{R}(G_f(x;\boxed{\theta_f}));\boxed{\theta_d}), y_d)$$

minimize

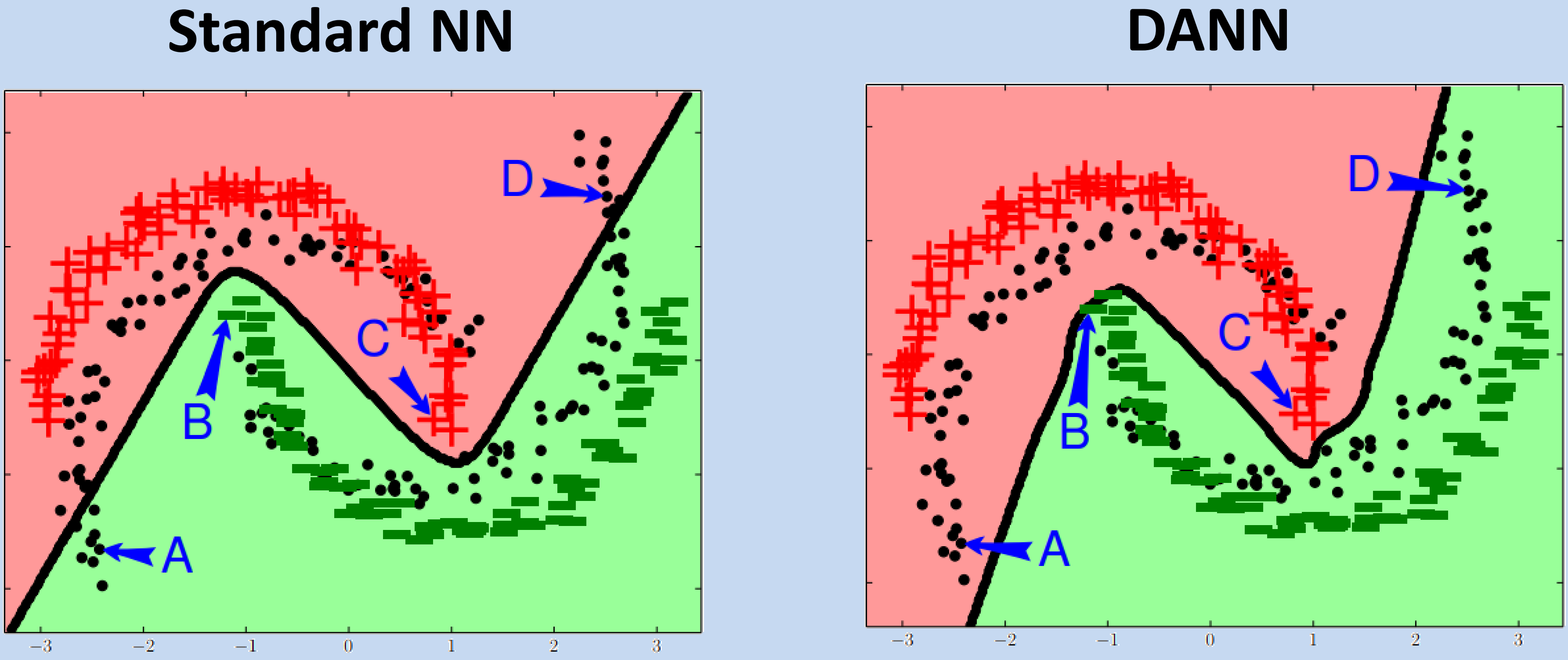
Control domain regularization

$$R(x) = x$$

maximize

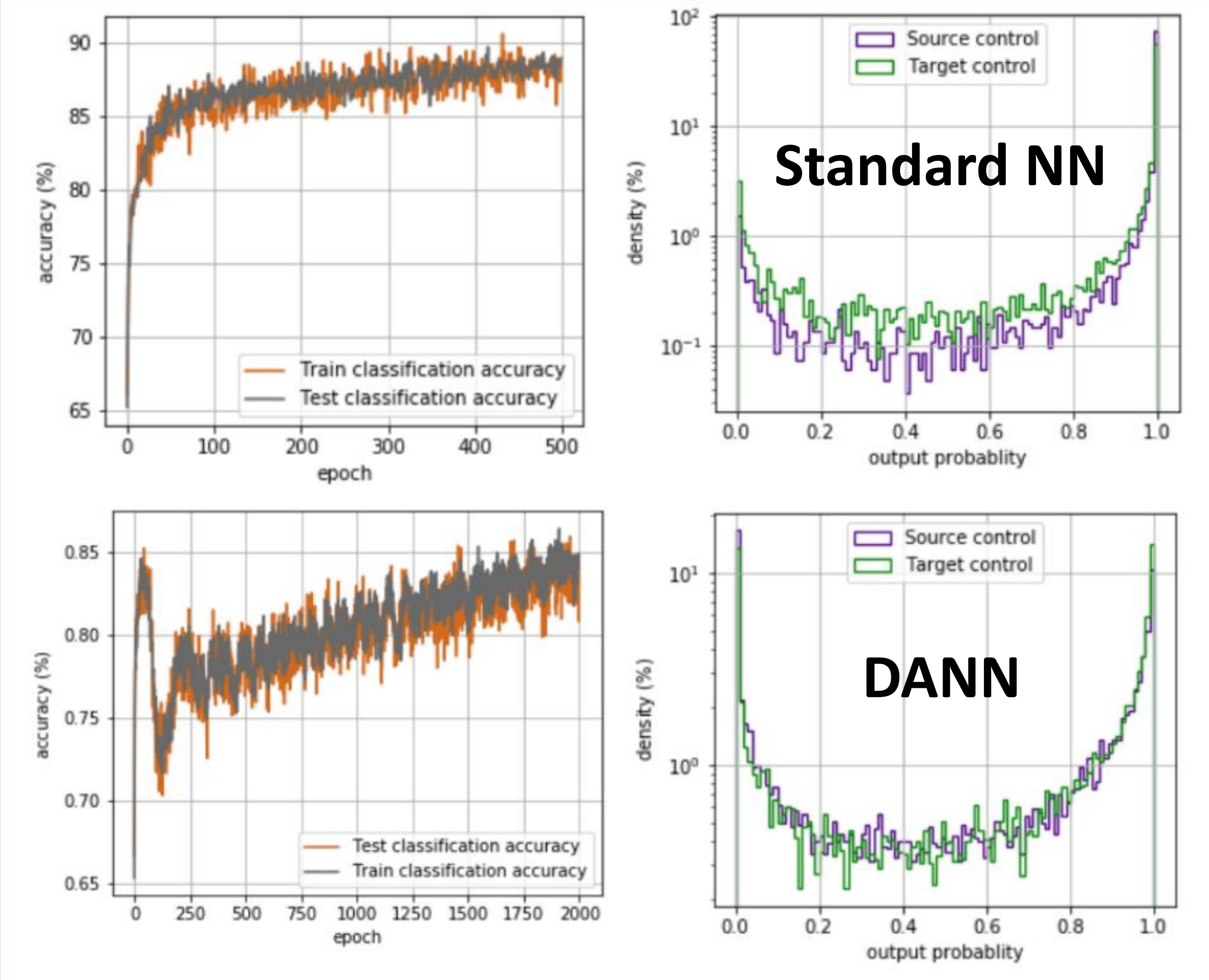
$$\frac{dR}{dx} = -1$$

Examples



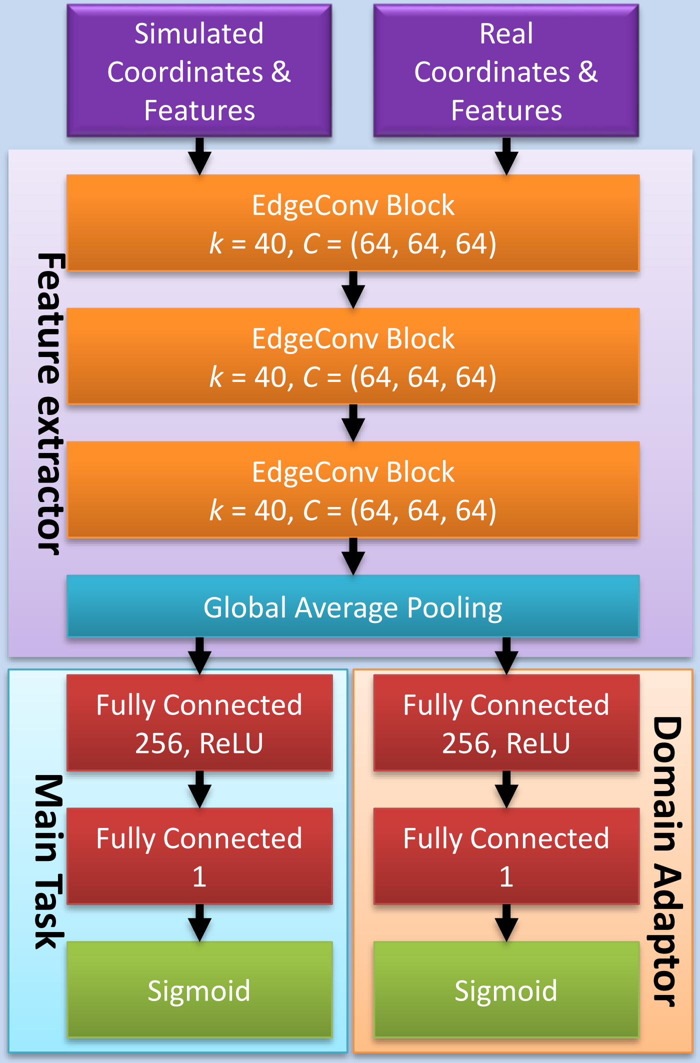
DANN Paper: JMLR 2016, vol. 17, p. 1-35

Examples



Sim-to-real (LHCb): arXiv:1912.08001 [cs.LG]

Our Implementation Details



Simulated Coordinates & Features

Real Coordinates & Features

Feature extractor

EdgeConv Block
$k = 40$, $C = (64, 64, 64)$

EdgeConv Block
$k = 40$, $C = (64, 64, 64)$

EdgeConv Block
$k = 40$, $C = (64, 64, 64)$

Global Average Pooling

Main Task

Fully Connected
256, ReLU

Fully Connected
1

Sigmoid

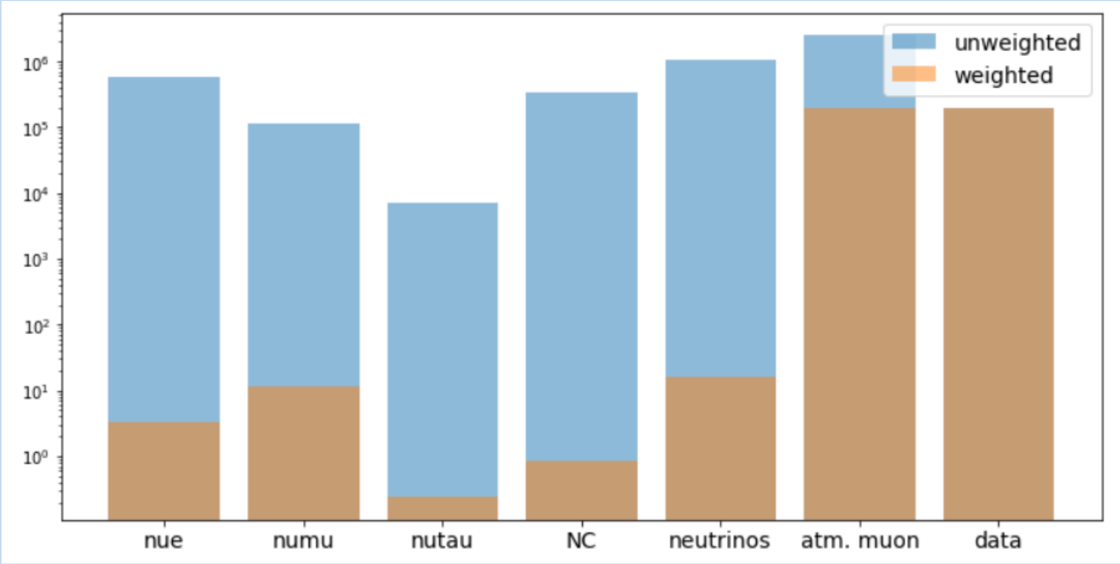
Domain Adaptor

Fully Connected
256, ReLU

Fully Connected
1

Sigmoid

- Batch size: 32
- Learning rate: 0.025
- Trained for 25 epochs
- Training sample:
  - 192k data examples
  - 3.5M MC examples
- Balancing may be important
- Real data: $\sim 10^4$ μ for each ν

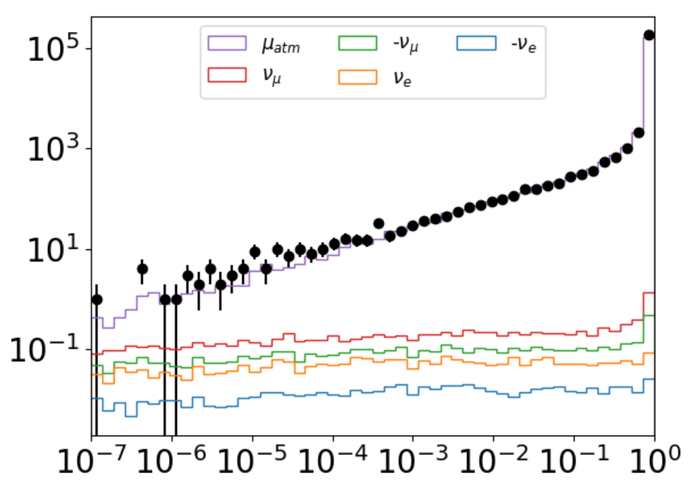


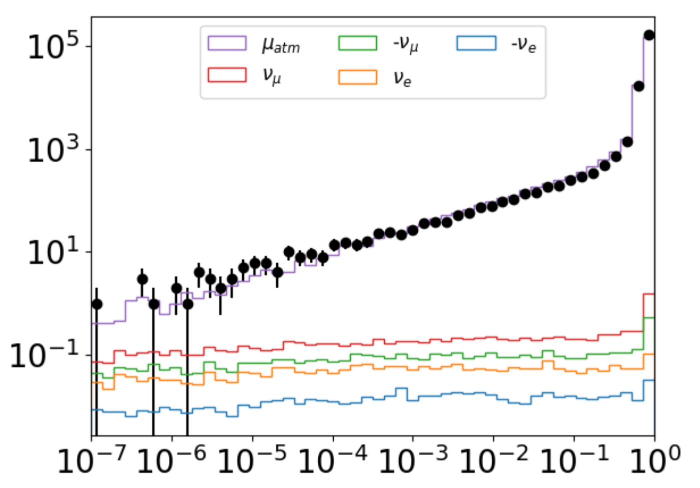
Initial Results (Nu Classifier)

Standard NN

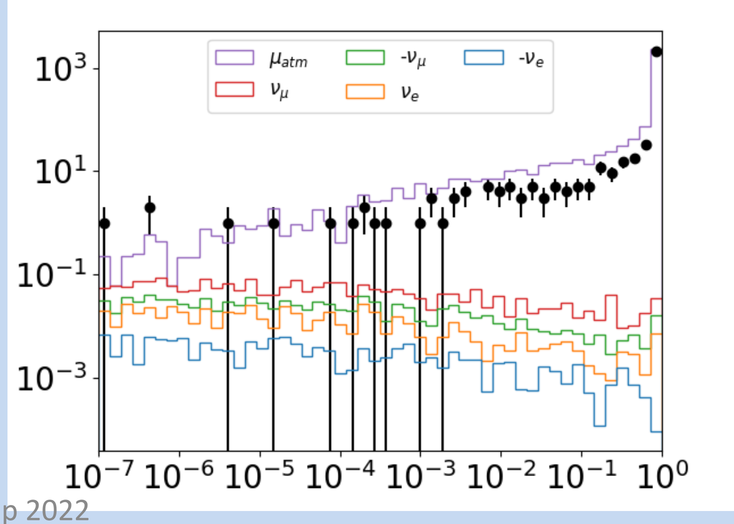
DANN

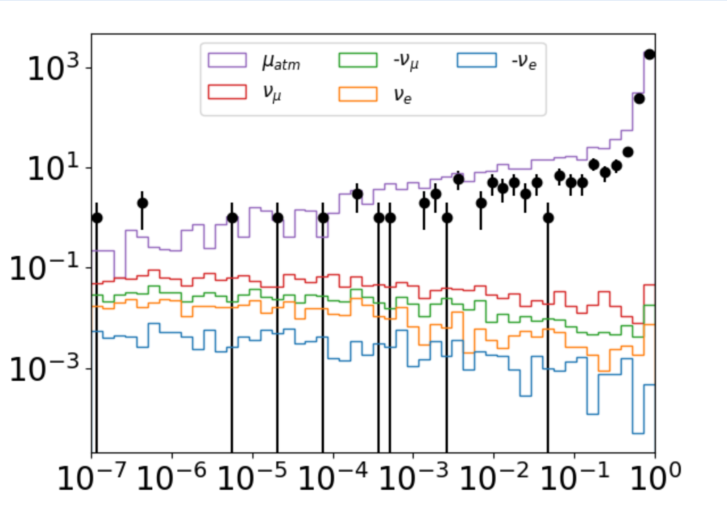
All Events





Preselection





Conclusions

- Supervised learning is performed in MC, but needs to be applied to unlabelled real data

- Domain-Adversarial Neural Networks (DANN) can be used to train models that can reduce domain shift

- KM3NeT has a successful program of implementing GNNs for event reconstruction and classification, but significant data/MC discrepancies seen in some cases

- Starting to train a DANN on these neutrino reconstruction tasks
- Results still in very early stages

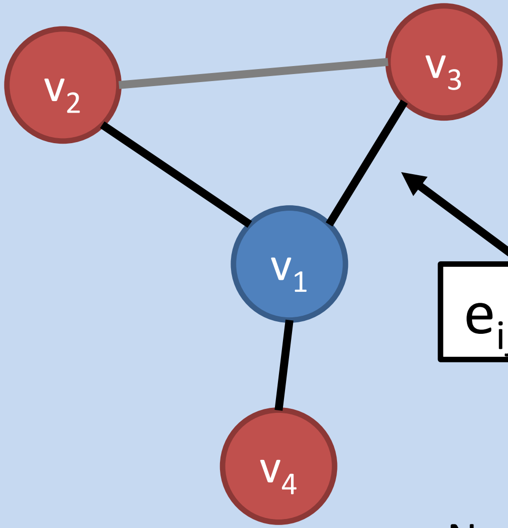
Backup

Graph Convolution

- Core layer implemented from ParticleNet: PRD 101, 056019 (2020)

Coordinates

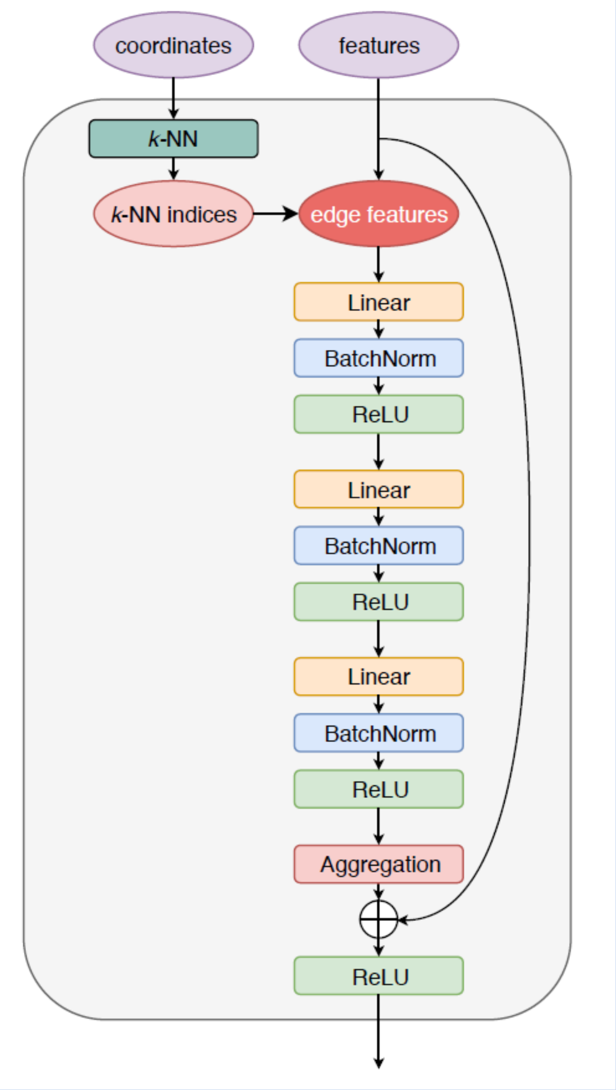
$v_j = \mathbf{x, y, z, t}, d_x, d_y, d_z$



$e_{ij} = (v_i, v_j - v_i)$

Neural Network

$$v_i \to \sigma(h_0(v_i) + \langle h(e_{ij}) \rangle)$$

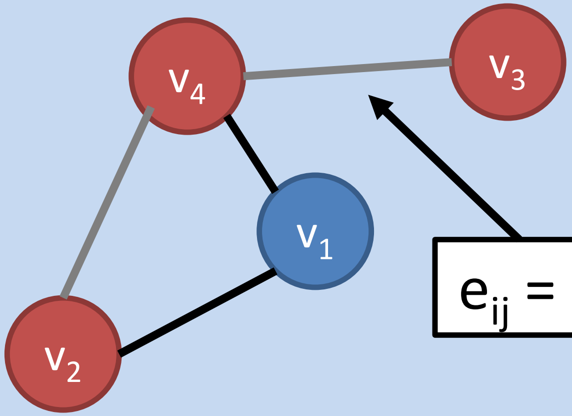


Graph Convolution

- Core layer implemented from ParticleNet: PRD 101, 056019 (2020)

Coordinates

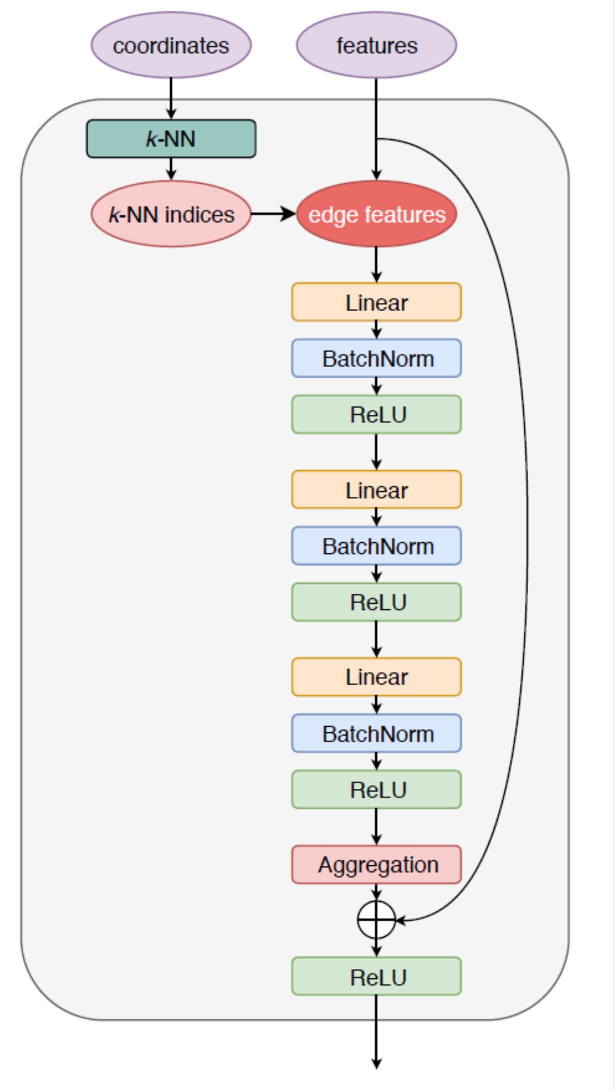
$v_j = \mathbf{x, y, z, t}, d_x, d_y, d_z$



$e_{ij} = (v_i, v_j - v_i)$

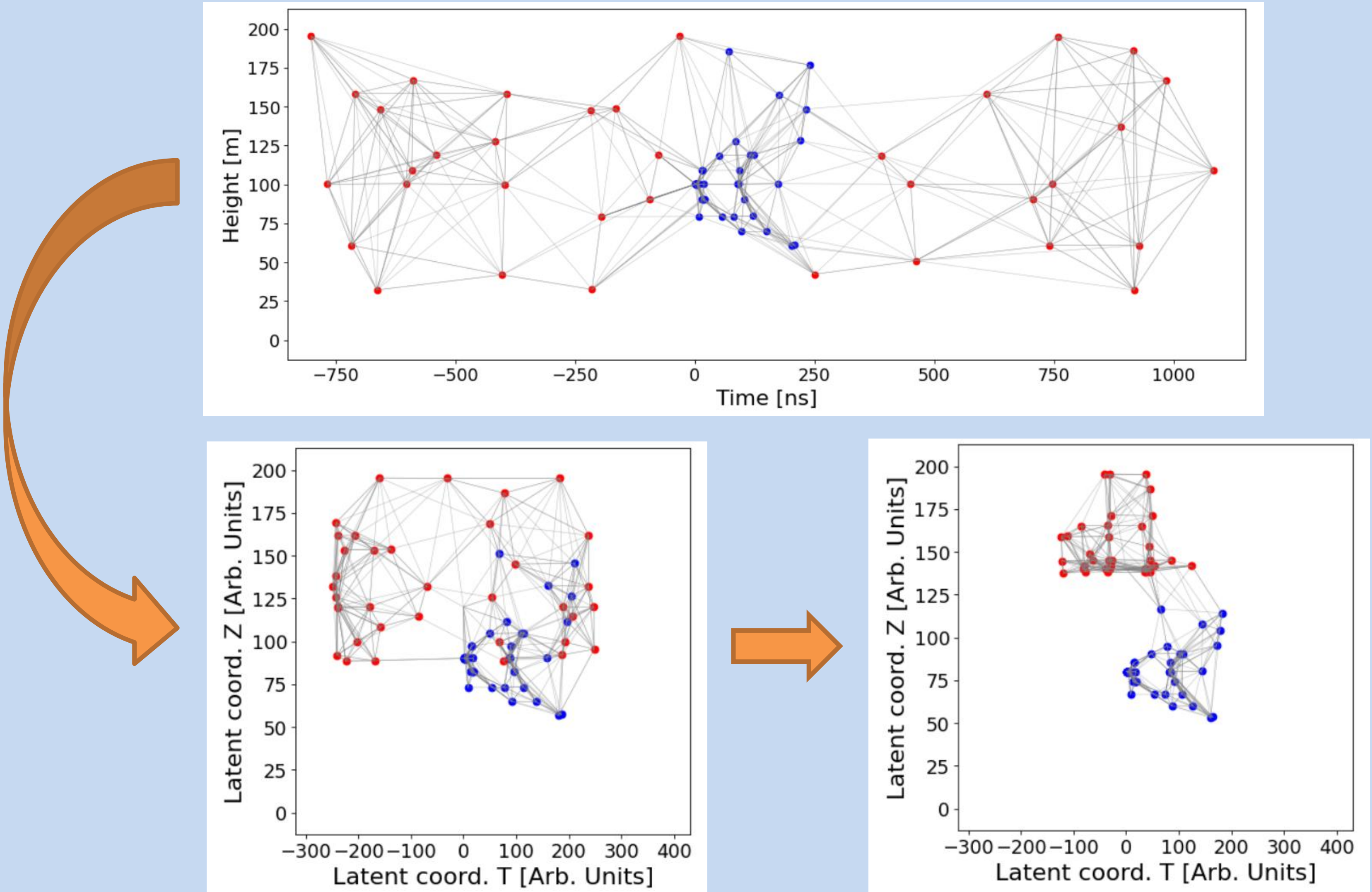
Neural Network

$$v_i \rightarrow \sigma(h_0(v_i) + \langle h(e_{ij}) \rangle)$$



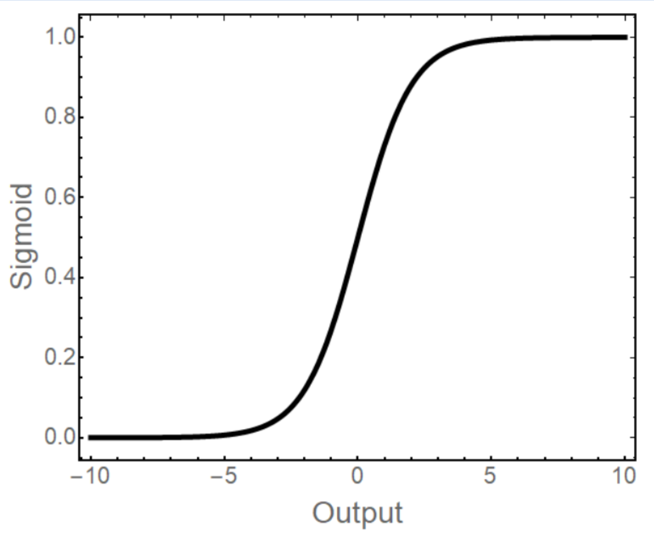
Graph Convolution

- Core layer implemented from ParticleNet: PRD 101, 056019 (2020)

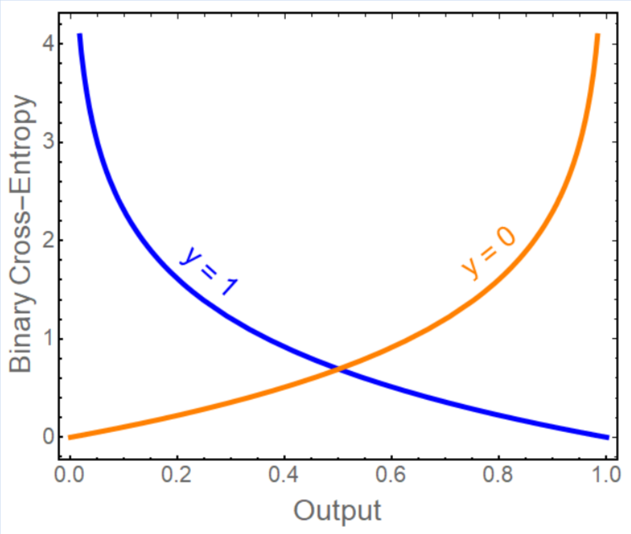


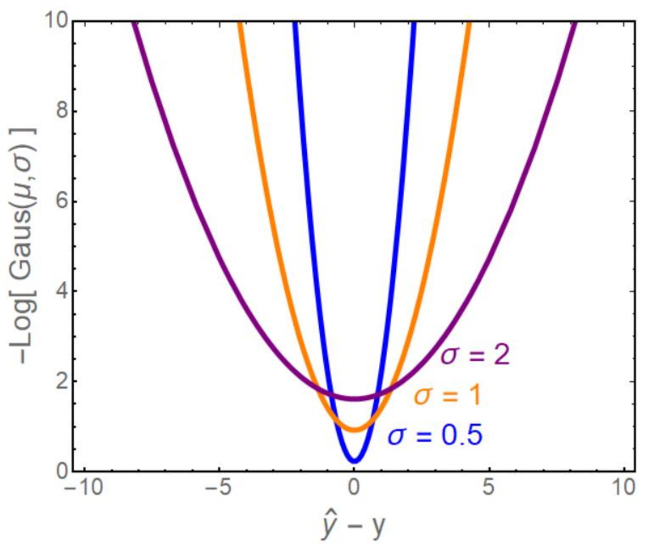
Loss Functions

Classification

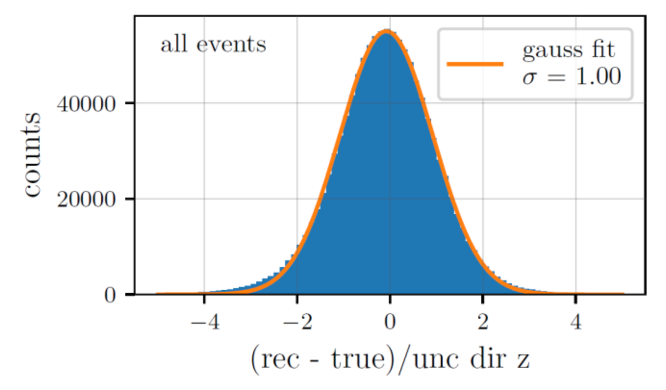


Standard Pair Validated with hyperparam. tuning





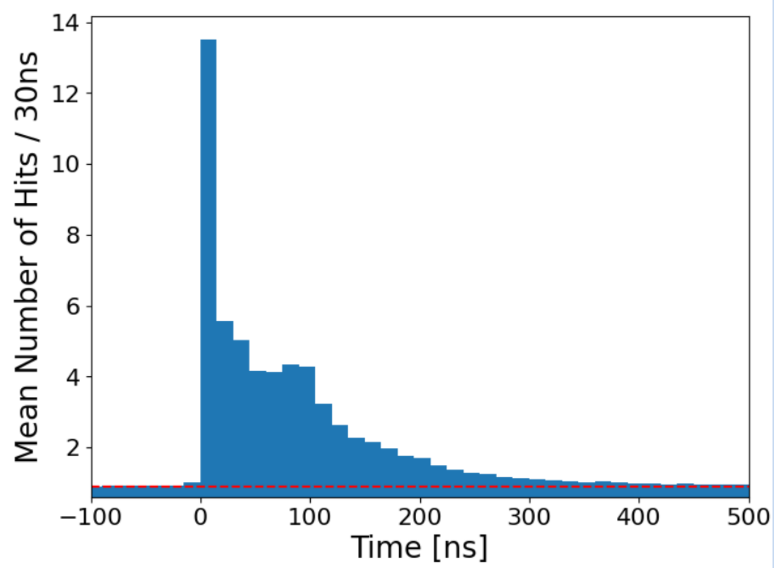
- Output both mean and std. dev. of target
- Allows network to estimate uncertainty

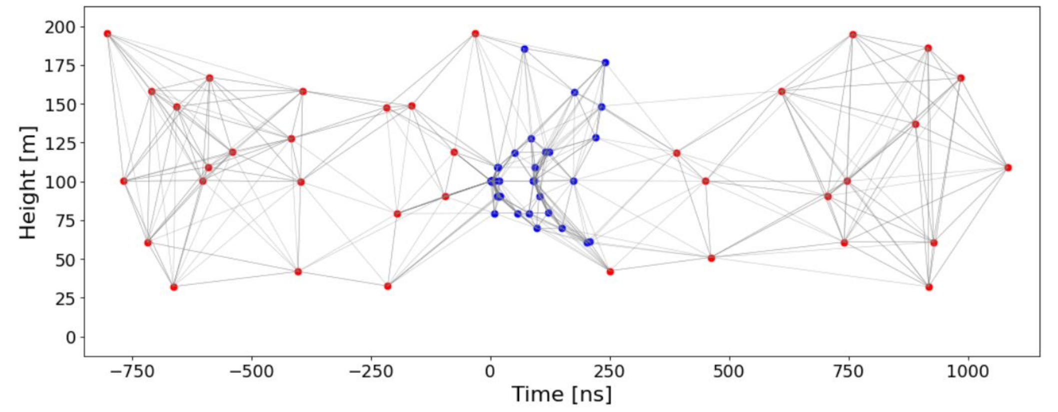


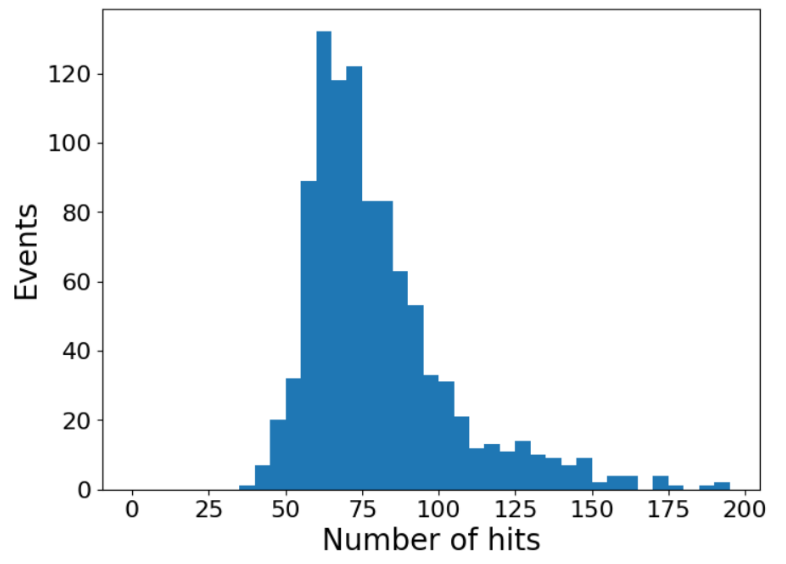
Regresssion

Graph Statistics

- Noise rate: ~8 kHz / PMT (radioactivity ^{40}K)
- ORCA6: 6×18×31 = 3348 PMTs ~ 30 MHz
- ORCA115: 64k PMTs ~ 500 MHz
- Event length ~ 100m / c ~ 300 ns
  - ORCA6: ~10 noise hits
  - ORCA115: ~150 noise hits
- On average ν_μ events will produce ~25 hits
- Overall, relatively large graphs
- Mostly noise dominated, but spatial-temporal structure is relatively clean



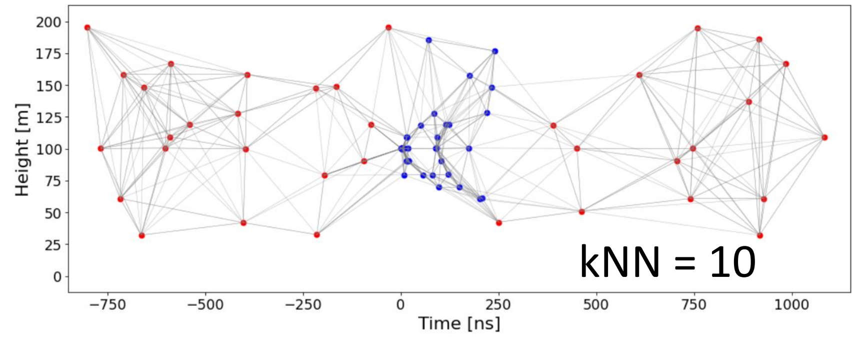




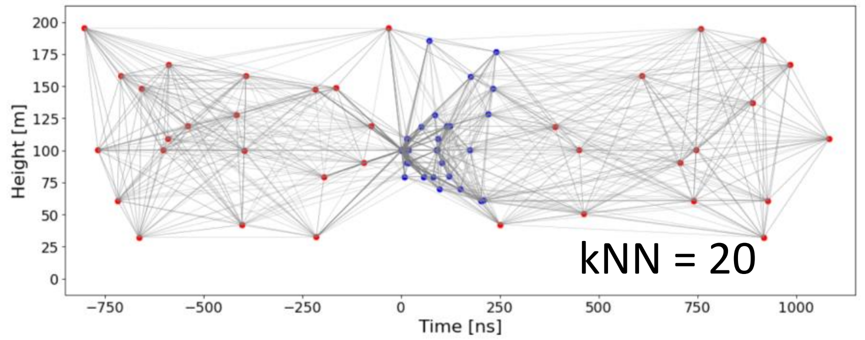
Currently keeping 2μs timeslice → ~80 hits

Connectivity

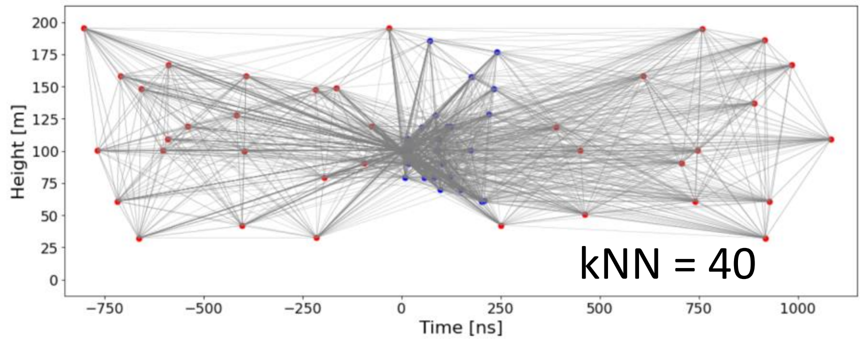
- Number of neighbors is a hyperparameter



kNN = 10

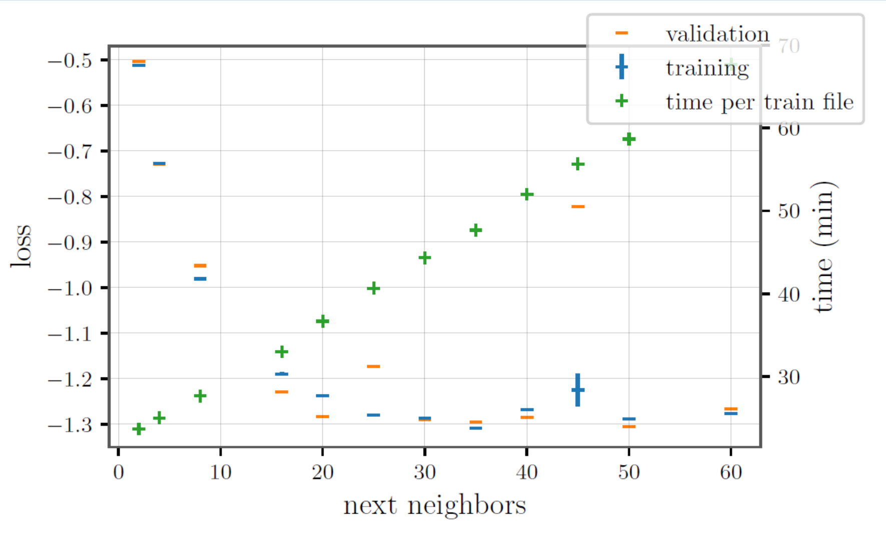


kNN = 20



kNN = 40

- Current default: kNN = 40
- Performance degraded below kNN = 20
- Only 3 hidden layers, so information does not propagate enough to process whole graph
- Deeper network with fewer neighbours may be an option



Data Balancing

- Currently data is balanced by including similar numbers of events in each category
- Weights are ignored, does it matter? No significant impact seen so far
- What should be the correct balance when atm. muons are 10^4 times larger?

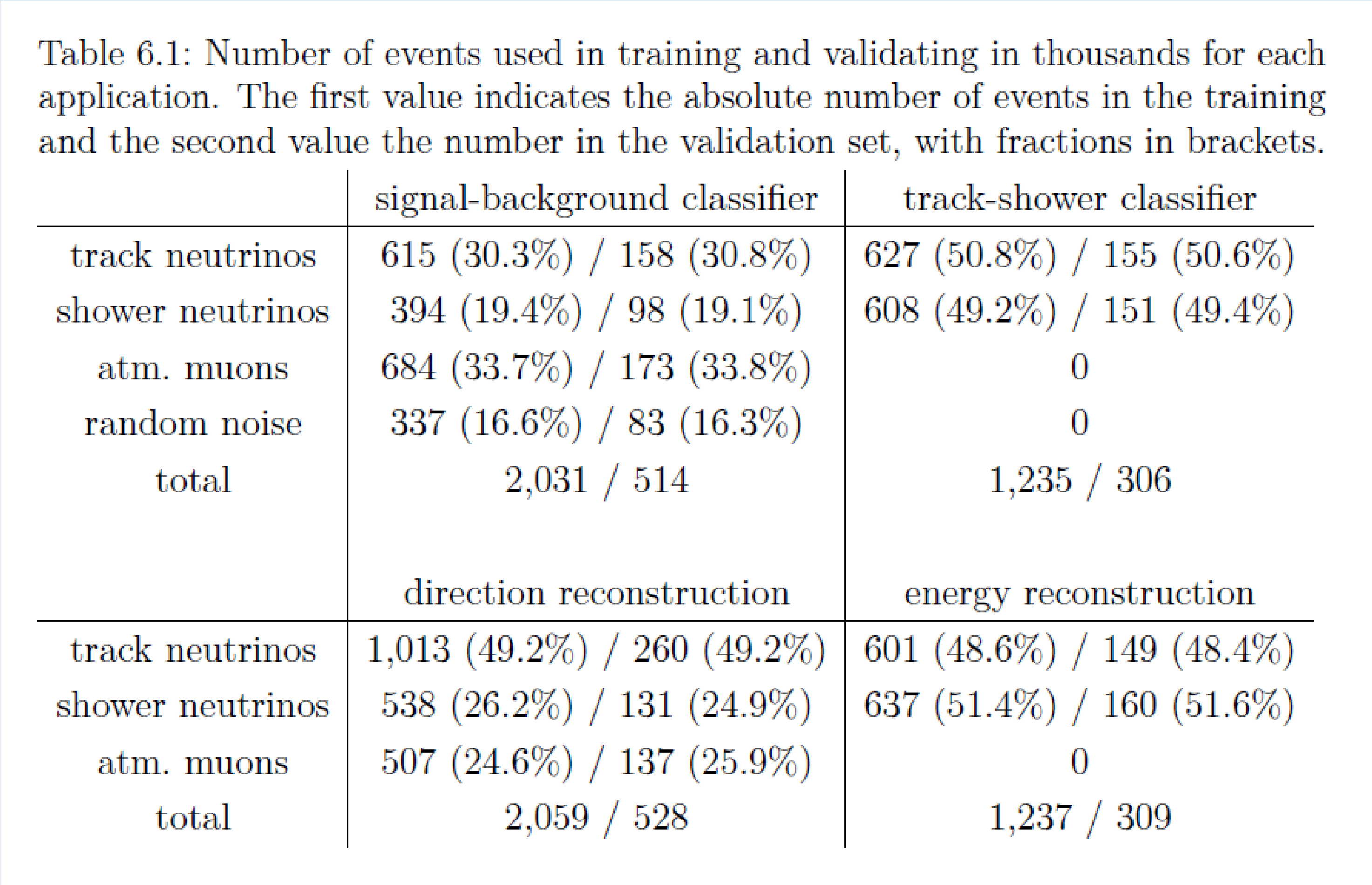


Table 6.1: Number of events used in training and validating in thousands for each application. The first value indicates the absolute number of events in the training and the second value the number in the validation set, with fractions in brackets.

	signal-background classifier	track-shower classifier
track neutrinos	615 (30.3%) / 158 (30.8%)	627 (50.8%) / 155 (50.6%)
shower neutrinos	394 (19.4%) / 98 (19.1%)	608 (49.2%) / 151 (49.4%)
atm. muons	684 (33.7%) / 173 (33.8%)	0
random noise	337 (16.6%) / 83 (16.3%)	0
total	2,031 / 514	1,235 / 306

	direction reconstruction	energy reconstruction
track neutrinos	1,013 (49.2%) / 260 (49.2%)	601 (48.6%) / 149 (48.4%)
shower neutrinos	538 (26.2%) / 131 (24.9%)	637 (51.4%) / 160 (51.6%)
atm. muons	507 (24.6%) / 137 (25.9%)	0
total	2,059 / 528	1,237 / 309