AI GOES MAD²

Machine Learning in KM3NeT

AI goes MAD-2

15 october 2024

Jorge Prado González (jprado@km3net.de)
Instituto de Física Corpuscular, Valencia, Spain

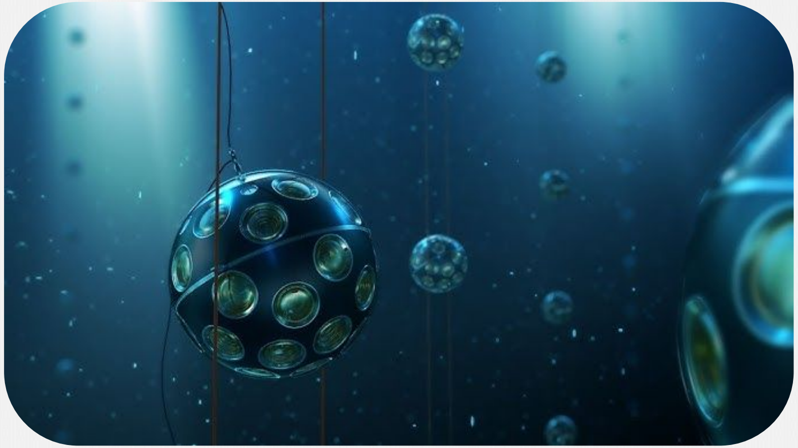
Work in collaboration with Iván Mozún Mateo

Index

- Neutrino telescope: KM3NeT.
- Machine Learning in KM3NeT.
- Model performance comparison.

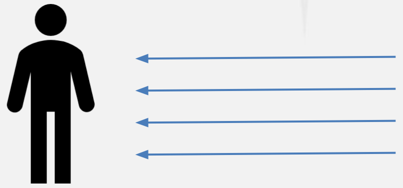
PART I

Neutrino Telescope: KM3NeT

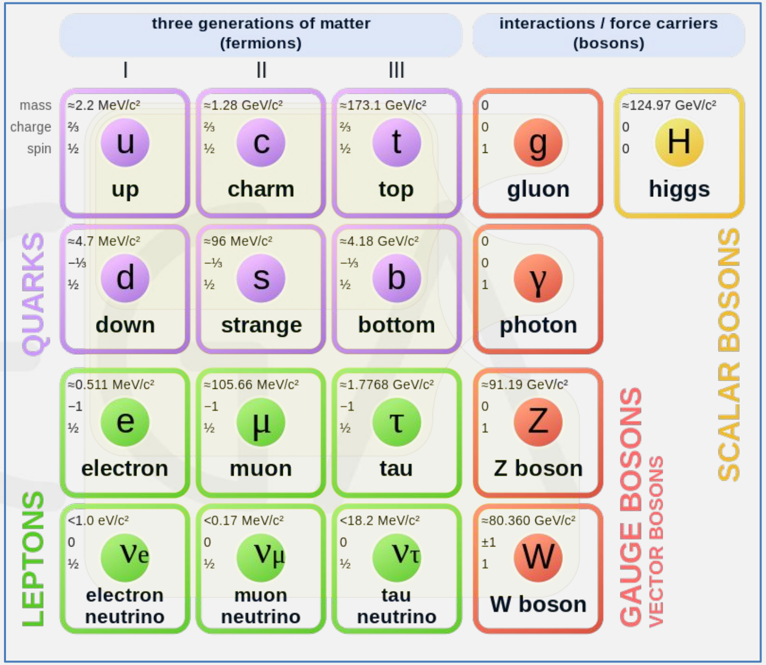


Neutrinos

- Neutrinos are **particles of the Standard Model**.
- Neutrinos are **the second most abundant particles** in the universe (after photons).
- They only interact via weak force with extremely small cross sections.
- Neutrinos can enlighten the path and help **answering some unsolved questions about particle physics**, as well as to **understand some astrophysical objects**.

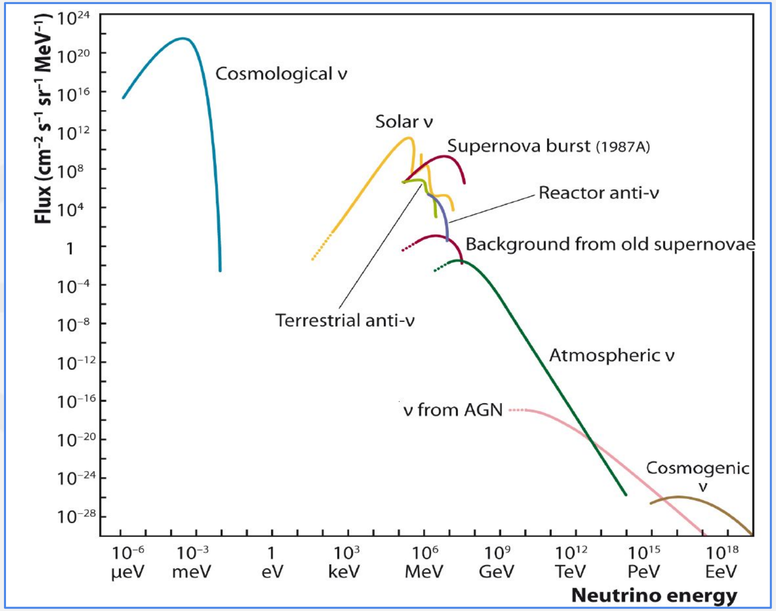


$$100\ trillion\ {}^{\nu}/_{s}$$

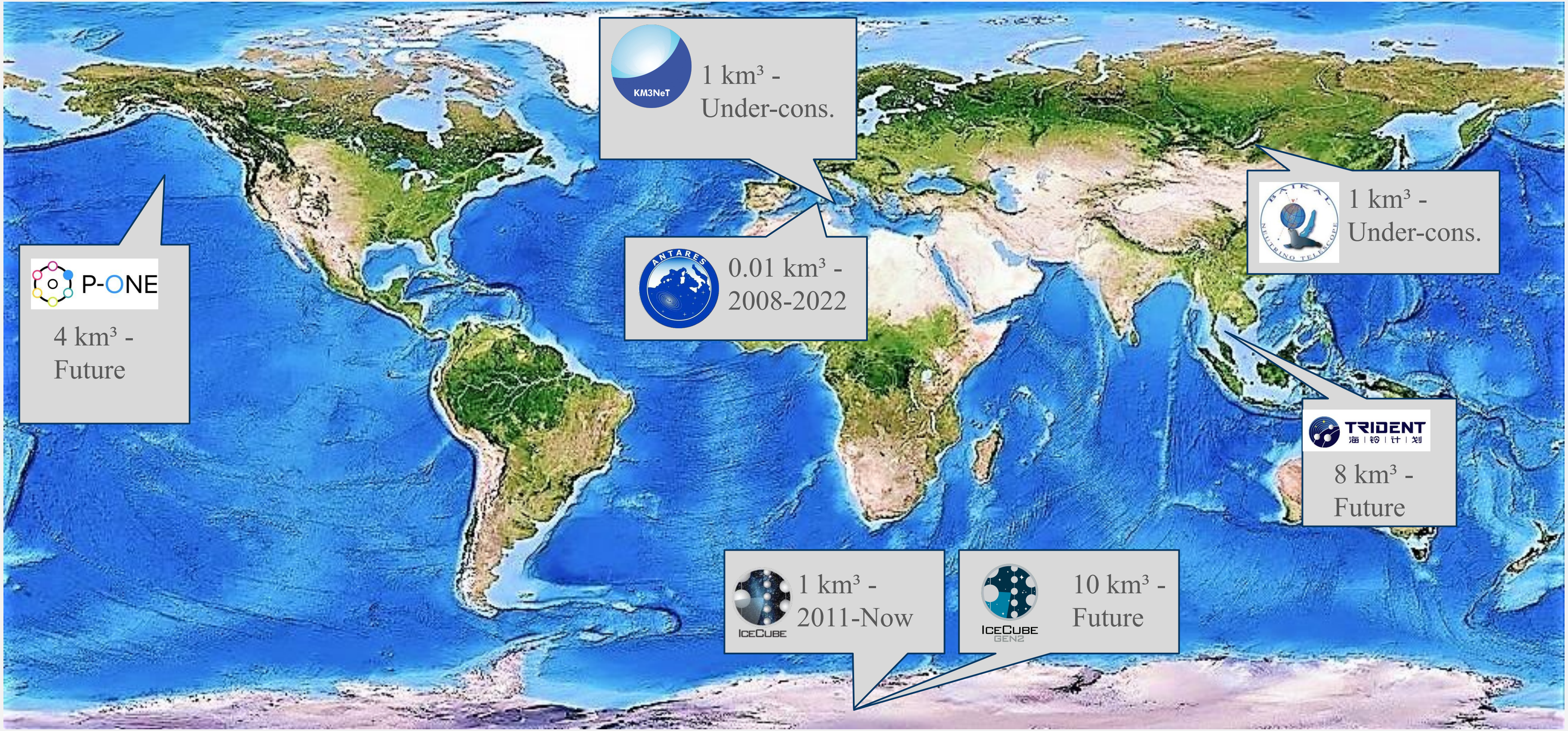


Neutrinos

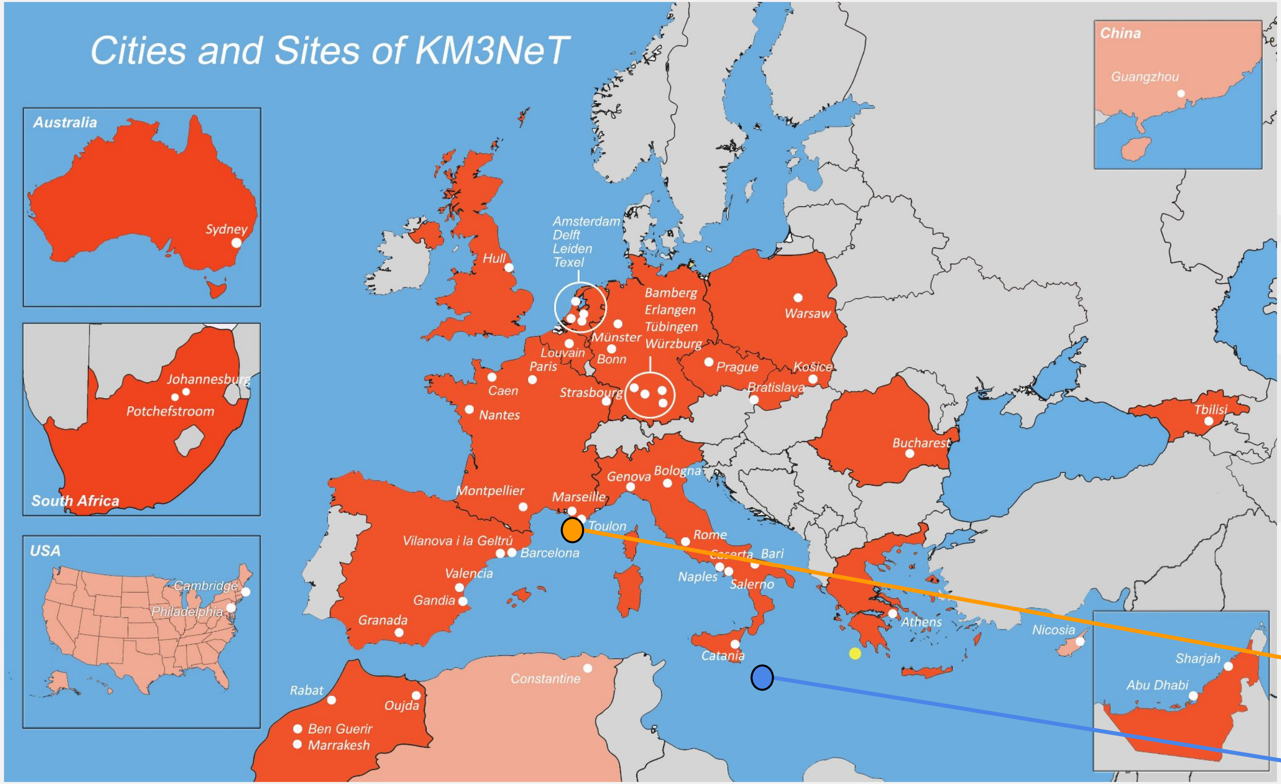
- Neutrinos are **particles of the Standard Model**.
- Neutrinos are **the second most abundant particles** in the universe (after photons).
- They only interact via weak force with extremely small cross sections.
- Neutrinos can enlighten the path and help **answering some unsolved questions about particle physics**, as well as to **understand some astrophysical objects**.
- Neutrinos coming from very different sources:
  - ➢ The Sun.
  - ➢ Supernovae.
  - ➢ AGNs.
  - ➢ Nuclear reactors.
  - ➢ Atmospheric neutrinos.



Neutrino telescopes



KM3NeT



- International collaboration with
  - ~250 members.
  - 65 partner institutes.
  - Over 22 countries.
- Two detectors in different sites: KM3NeT/ORCA and KM3NeT/ARCA:
  - Same technology.
  - Same data processing.
  - Same software and common dataformats.
  - Different size and granularity.

KM3NeT/ORCA

KM3NeT/ARCA

KM3NeT - ARCA and ORCA

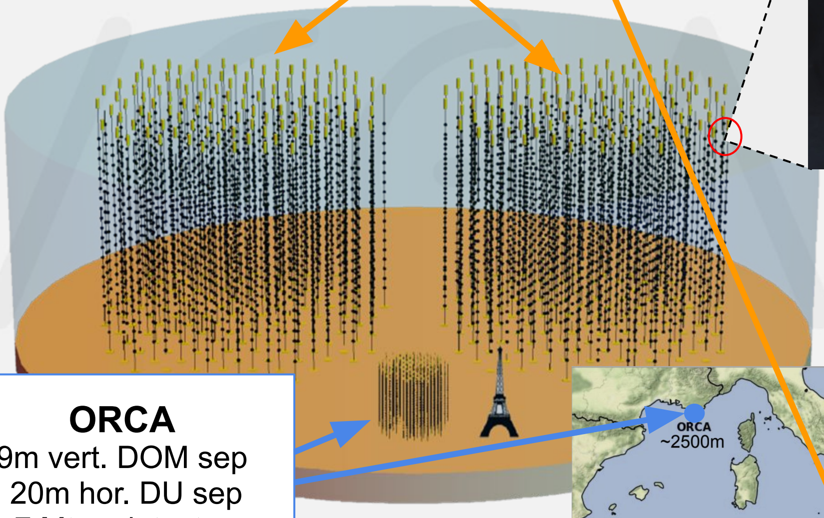
- **KM3NeT/ORCA:**
  - ➢ Low energies (~few GeV to hundreds of GeVs).
  - ➢ Fundamental neutrino property studies (mainly).
  - ➢ **Full ORCA:** 115 DUs, 18 DOMs per DU.
  - ➢ **Current ORCA:** 23 DUs deployed.
- **KM3NeT/ARCA:**
  - ➢ High energies (sub-TeV to few PeV).
  - ➢ Astrophysical studies (mainly).
  - ➢ **Full ARCA:** 230 DUs, 18 DOMs per DU.
  - ➢ **Current ARCA:** 28 DUs deployed.

DU: Detection Unit. String of 18 DOMs.

DOM: Digital Optical Module.

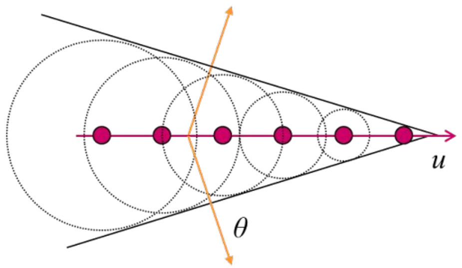
ARCA
36m vert. DOM sep.
90m hor. DU sep.
1 Gton detector

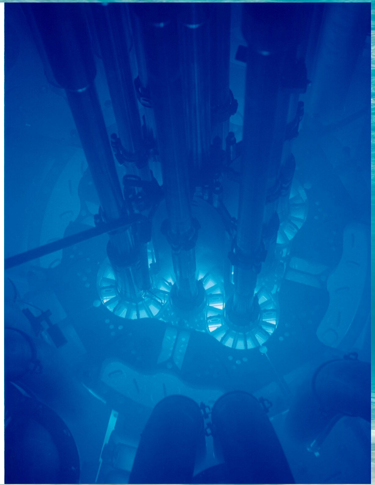
DOM
31x3" PMTs

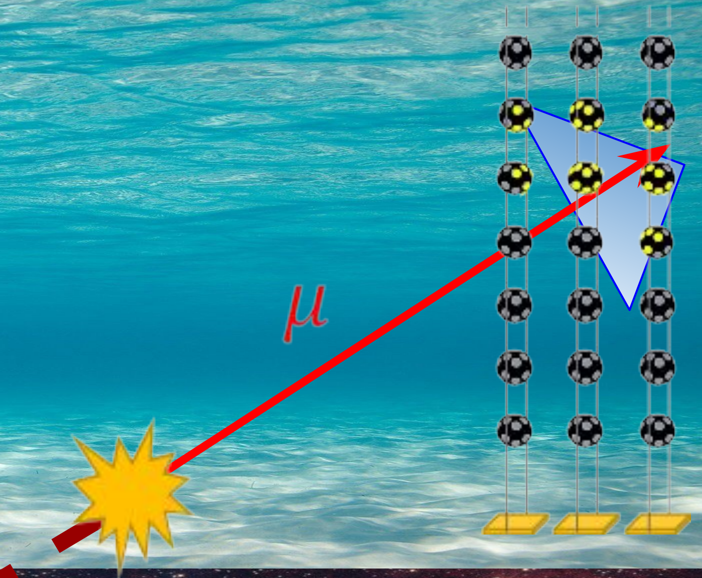


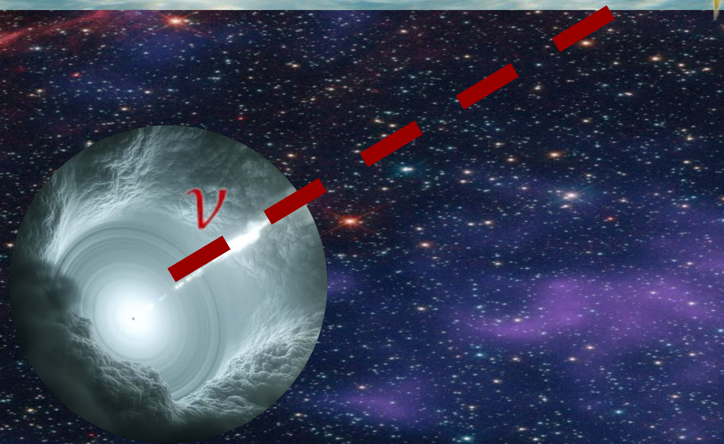
ORCA
9m vert. DOM sep
20m hor. DU sep
7 Mton detector

Cherenkov light





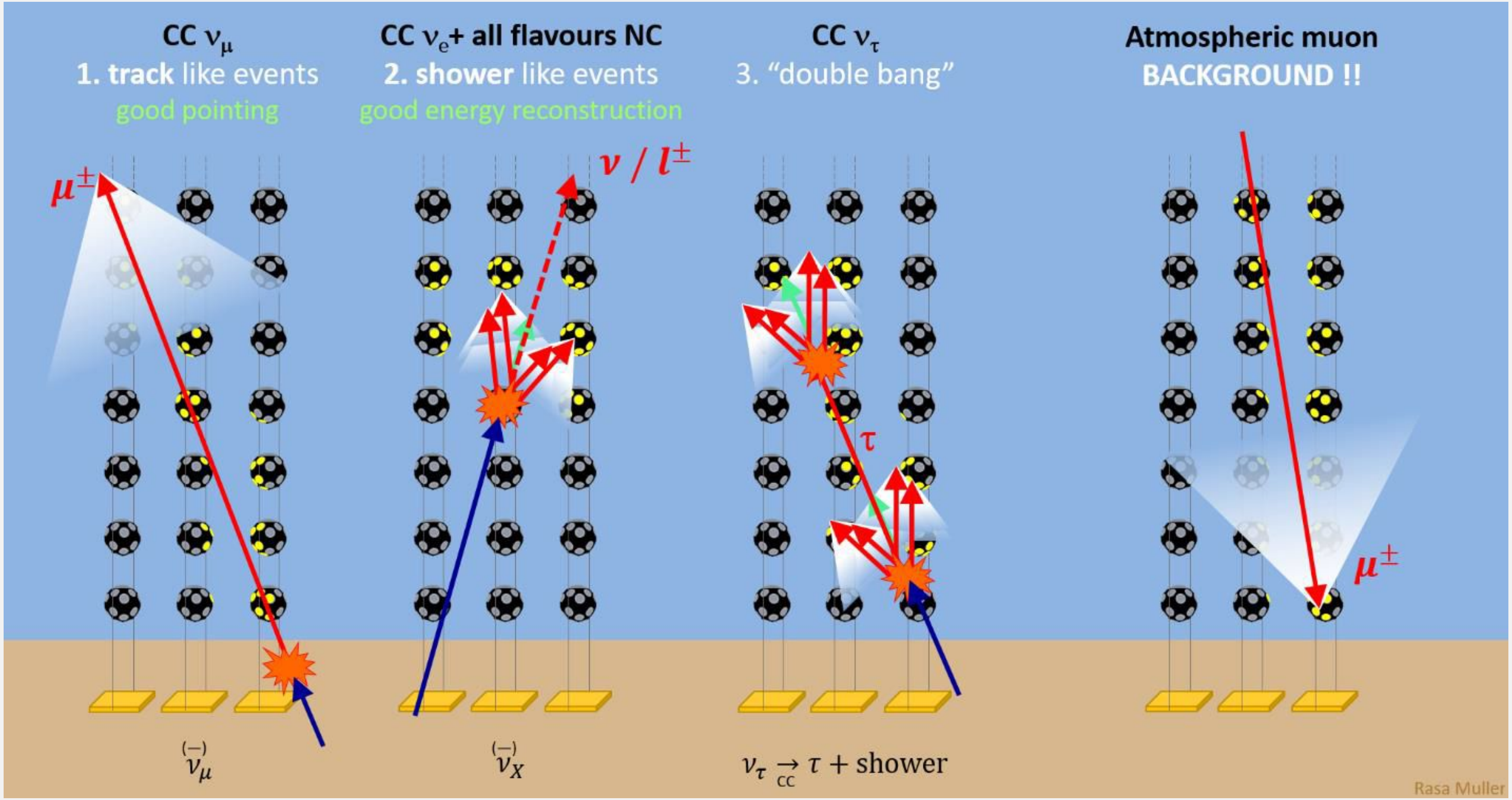




We can detect neutrinos!

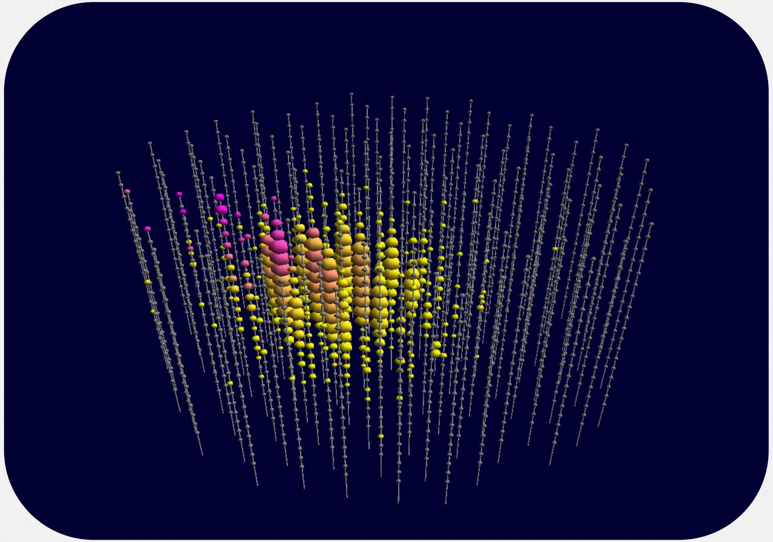
Video

KM3NeT - Detection principle. Event topology



PART II

Machine Learning in KM3NeT



Deep Learning Projects in KM3NeT

CNNs:

- Event reconstruction for KM3NeT/ORCA using convolutional neural networks (M. Moser, KM3NeT)
- Event Classification and Energy Reconstruction for ANTARES using Convolutional Neural Networks (N. Geißelbrecht, ANTARES)
- Deep learning reconstruction in ANTARES (J. García-Méndez et al., ANTARES)
- Dark matter search towards the Sun using Machine Learning reconstructions of single-line events in ANTARES (J. García-Méndez et al., ANTARES)

Fully-connected NNs:

- Deep Neural Networks for combined neutrino energy estimate with KM3NeT/ORCA6 (S. Peña Martínez, KM3NeT)

Several different ML-based projects being already part of physics analyses (BDTs, RFs…):

- ParamPID: t/s, nu/noise and nu/mu classifier with XGBoost (A. Lazo & L. Maderer, KM3NeT)
- CR composition measurement: Atm. muon bundle reconstruction using RFs (P. Kalaczynski, KM3NeT)
- BoostTauID: identify GeV tau neutrinos in ORCA with XGBoost/ParamPID (N.Geißelbrecht, KM3NeT)

Deep Learning Projects in KM3NeT

GNNs:

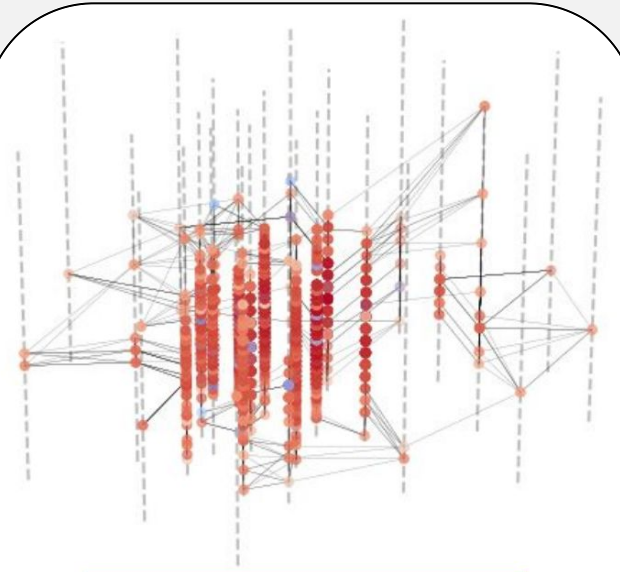
- Development of detector calibration and graph neural network-based selection and reconstruction algorithms for the measurement of oscillation parameters with KM3NeT/ORCA (D. Guderian, KM3NeT)
- Data reconstruction and classification with graph neural networks in KM3NeT/ARCA6-8 (F. Filippini et al., KM3NeT)
- Cosmic ray composition measurement using Graph Neural Networks for KM3NeT/ORCA (S. Reck, KM3NeT)
- Optimisation of energy regression with sample weights for GNNs in KM3NeT/ORCA (B.Setter, KM3NeT)
- Tau neutrino identification with Graph Neural Networks in KM3NeT/ORCA (L. Hennig, KM3NeT)
- Energy reconstruction in ARCA21 using GNNs (E. Tragia, P. Gkotsis, E. Drakopoulou, KM3NeT)
- Particle ID classification, energy, direction and interaction vertex position reconstruction in KM3NeT/ORCA using Dynedge (J. Prado, KM3NeT)
- Neutrino Selection using GNNs for ARCA28 (A. Veutro, KM3NeT)
- Heavy neutral lepton signal identification using DYENDGE in KM3NeT/ORCA (J. Prado, KM3NeT)

Transformers/Foundation models:

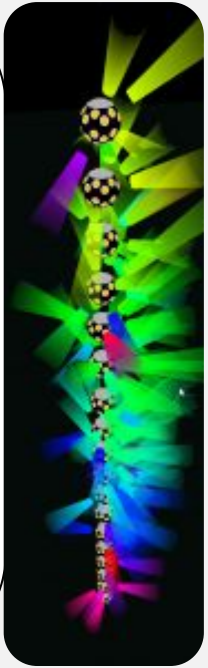
- Transformer based classification and reconstruction in KM3NeT/ORCA (I. Mozún, KM3NeT)
- Transfer learning in KM3neT/ORCA with transformers (I. Mozún, KM3NeT)

Representing data with graphs

- *"It is preferable not to shape the problem to the tool, but the tool to the problem" [My ML professor].*
- GNNs using graphs as input **capture the irregular geometries** of our events with **no underlying assumption of on the geometry.**
- Using **nodes** with inputs:
  - ➢ Position where the hit happens **(x,y,z)**
  - ➢ The direction of the PMT collecting the hit **(dx,dy,dz)**
  - ➢ The time when the hit happens **(t)**
  - ➢ The time that the PMT is collecting over 3 PE **(ToT)**
- Connected to its N-nearest neighbors in an euclidean space.

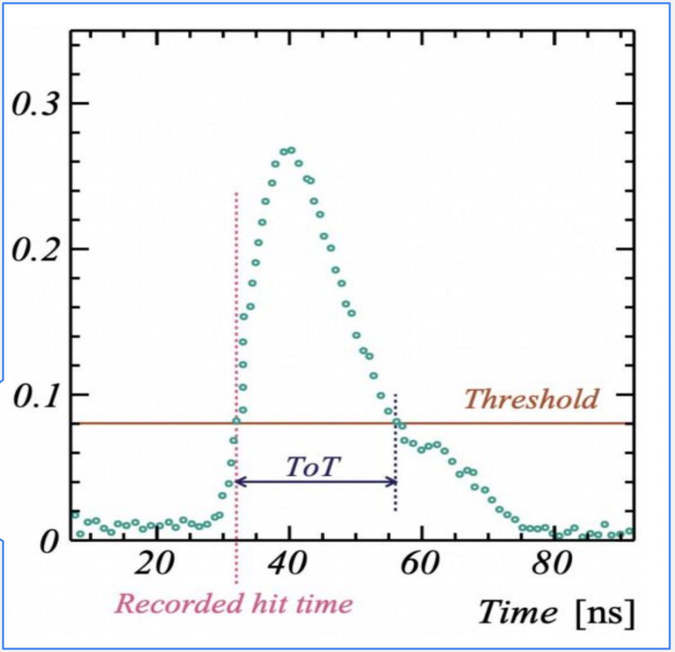


3.4 TeV ν_e event as graph with k = 8



Representing data with graphs

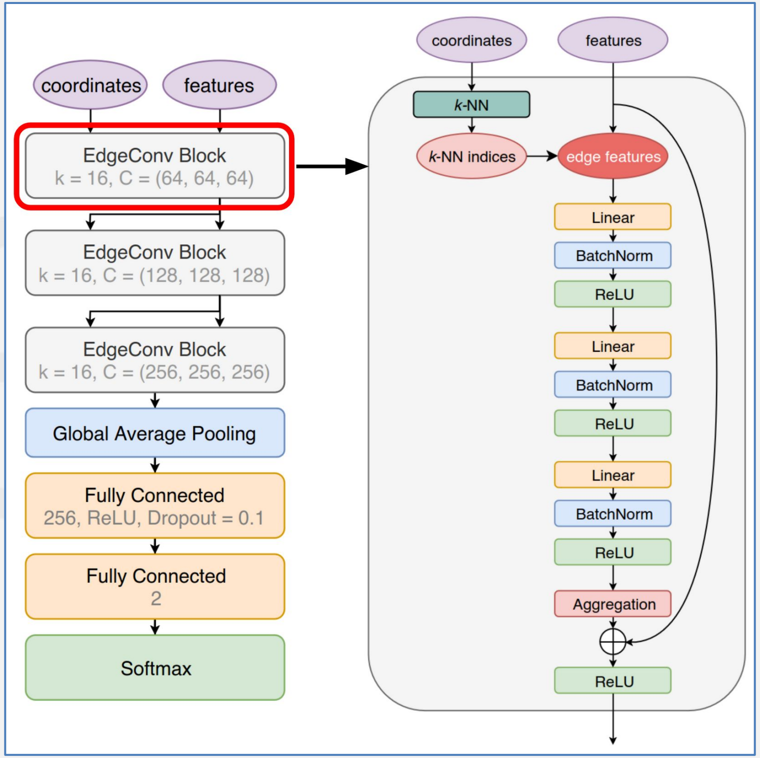
- *"It is preferable not to shape the problem to the tool, but the tool to the problem" [My ML professor].*
- GNNs using graphs as input **capture the irregular geometries** of our events with **no underlying assumption of on the geometry.**
- Using **nodes** with inputs:
  - ➢ Position where the hit happens **(x,y,z)**
  - ➢ The direction of the PMT collecting the hit **(dx,dy,dz)**
  - ➢ The time when the hit happens **(t)**
  - ➢ The time that the PMT is collecting over 3 PE **(ToT)**
- Connected to its N-nearest neighbors in an euclidean space.



ParticleNeT Model

- **ParticleNeT is a GNN** that was originally designed to jets at LHC.
- **Adapted and used in ORCA** under the name of ORCANeT since some years ago.
- **3 convolutional blocks** connected after to some **average pooling** to summarize the learned information and connect to a **fully connected perceptron** to learn more complex relations between node features.
- **Features a hierarchical structure**, early layers focused on learning low level features while deeper layers will capture more abstract, global patterns.
- Model with **521k trainable parameters***.

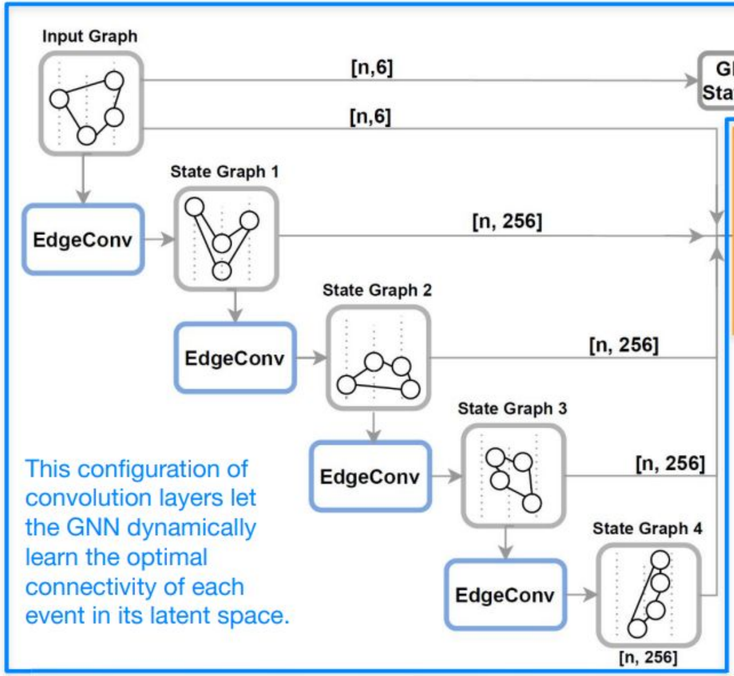
(*) Original model of ParticleNeT uses "mean" as the only node aggregation. Here it has been changed to ["mean", "max", min","sum"] changing the number of trainable parameters of the model from ~370k to this 512k.



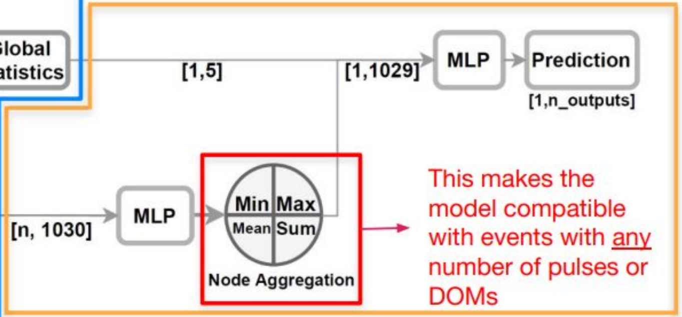
More complex GNNs - DYNEDGE

Graph Convolutional Layers

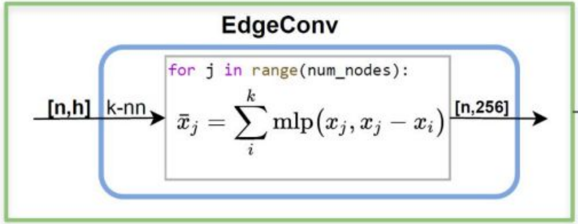
Classical Neural Networks



This configuration of convolution layers let the GNN dynamically learn the optimal connectivity of each event in its latent space.



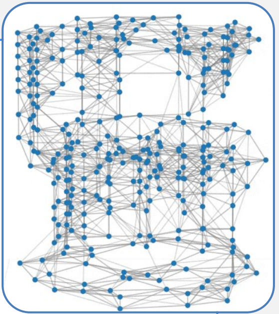
This makes the model compatible with events with any number of pulses or DOMs



Our Choice in Convolution

(https://arxiv.org/pdf/1801.07829.pdf)

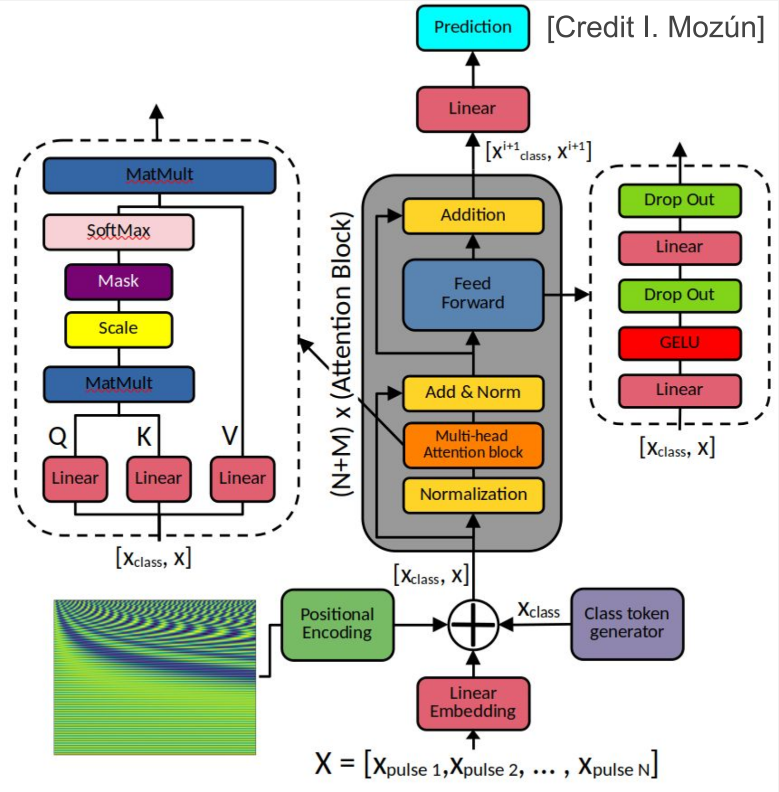
Captures globally relevant features in local areas by considering the difference of a node and the neighbours it's connected to. Only the neighbouring nodes contributes to the convolution!



1.4M trainable parameters

Transformers in KM3NeT

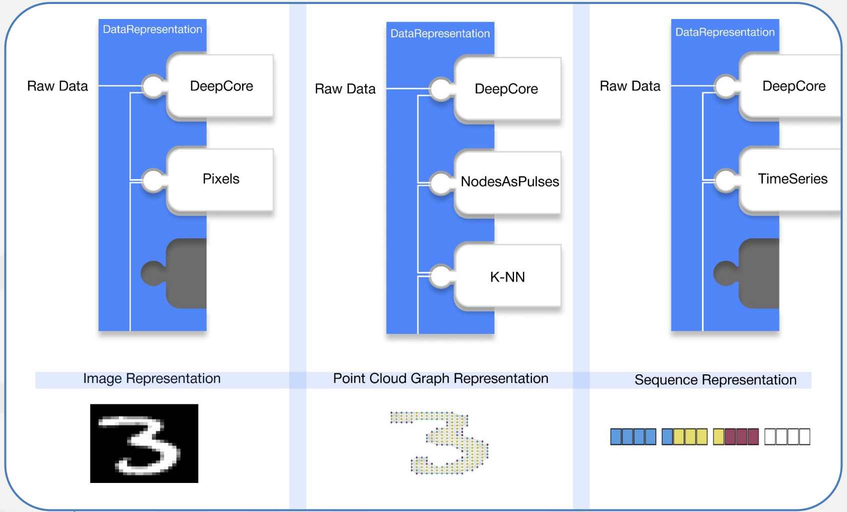
- Transformer model are **inspired in Natural Language Processing** tasks. They are starting to overtake but they have **a lot of learnable parameters that require a lot of data** for an optimal performance.
- The model arranges **light pulses in tokens** with information **[pos, dir, t, ToT]** in a sequence and learns relationships among them to then perform different tasks.
- Model with **1.6M trainable parameters**.
- Offers potential for **transfer learning**: fine-tuning pre-trained models for different detector configurations.

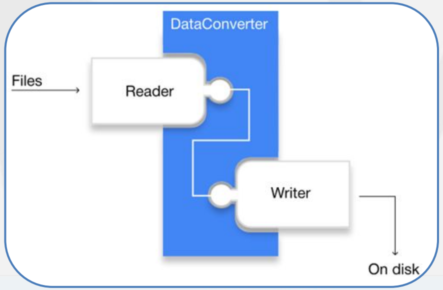


More details about the model and transfer learning in the follow up talk by Iván Mozún!

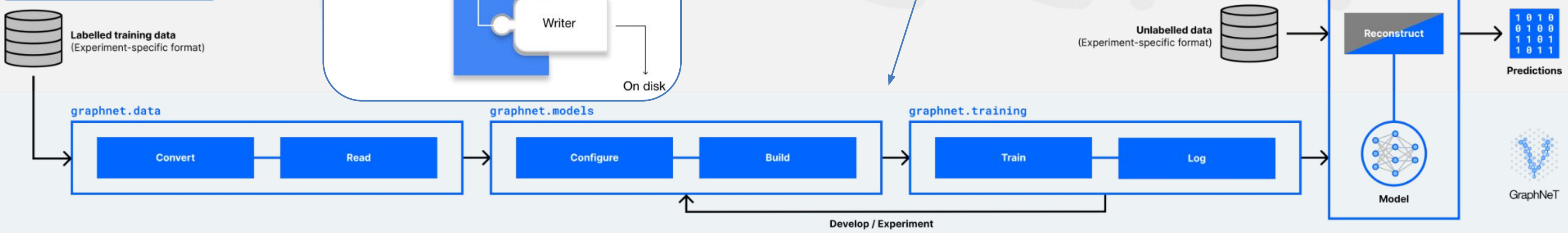
Comparing models in KM3NeT

- **GraphNeT** is a common framework for DL projects.
- It has a **modular structure** that makes it very **easy to embed your detector and your models** in the software.
- Dedicated instructions to include your detector.
- **Many models implemented**. ParticleNeT, DYNEDGE, RNNs and top 3 kaggle solution (see description here)
- Working on **SQLite** or **Parquet**.



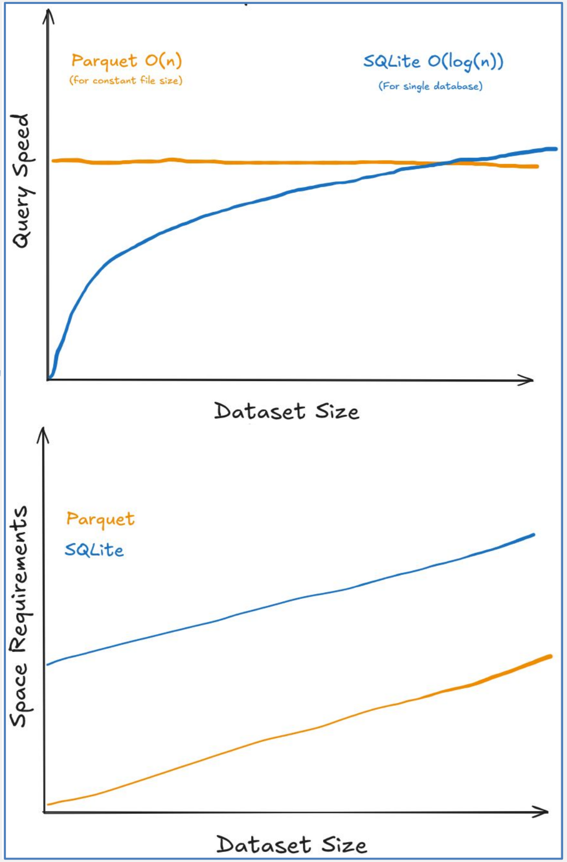


GraphNeT
Deep Learning for Neutrino Telescopes



Comparing models in KM3NeT

- **GraphNeT** is a common framework for DL projects.
- It has a **modular structure** that makes it very **easy to embed your detector and your models** in the software.
- Dedicated instructions to include your detector.
- **Many models implemented**. ParticleNeT, DYNEDGE, RNNs and top 3 kaggle solution (see description here)
- Working on **SQLite** or **Parquet**.



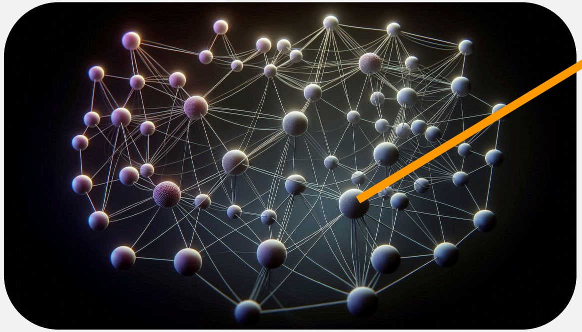
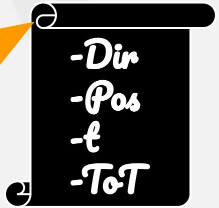
PART III

Preliminary model comparison

KM3NeT

Comparing Models

- Comparison between **ParticleNeT**, **DYENDGE** and the **transformer** model training, validating and testing on the exact same events.
- Trained for: **classification between tracks and showers**, **reconstruction of the neutrino energy** and the incoming **direction of the neutrino**.
- Using electron (anti-)neutrino as showers and muon (anti-)neutrinos as tracks in the energy range 1-100 GeV.

	Tracks	Showers
Training	1.92M	1.92M
Validation	480k	480k
Test	278k	278k

	Trainable parameters
ParticleNeT	521k
Dynedge	1.4M
Transformer	1.6M

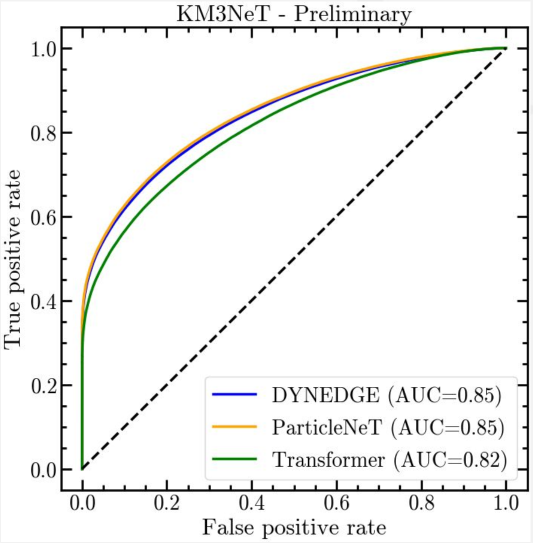
ParticleNeT and DYNEDGE trained 1 GPU Tesla V100-SXM2-32GB

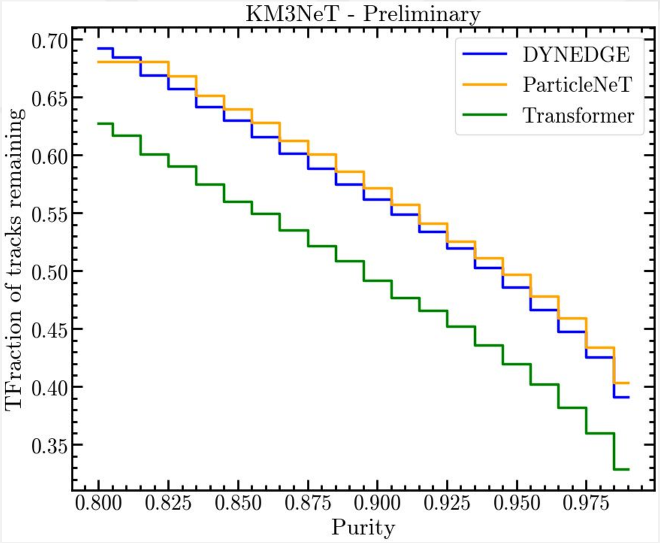
Transformer trained in 8 GPUs NVIDIA A100-SXM4-80GB

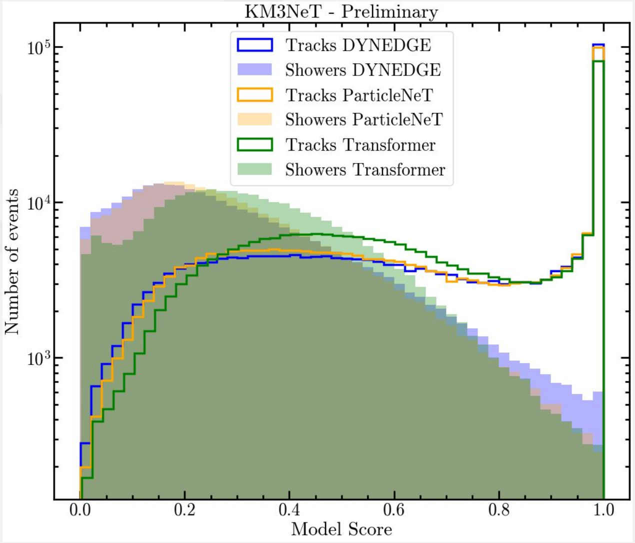
Track-Shower Classification

- Transformer performing worse DYNEDGE or ParticleNeT.
- The **simplest model is performing the best**. Lack of training events to extract the full potential of DYNEDGE or the transformer?

	N-epochs	Loss function	Time per epoch
ParticleNeT	38	BinaryCrossEntropy	~26min
DYNEDGE	28	BinaryCrossEntropy	~28min
Transformer	52	BinaryCrossEntropy	~15min

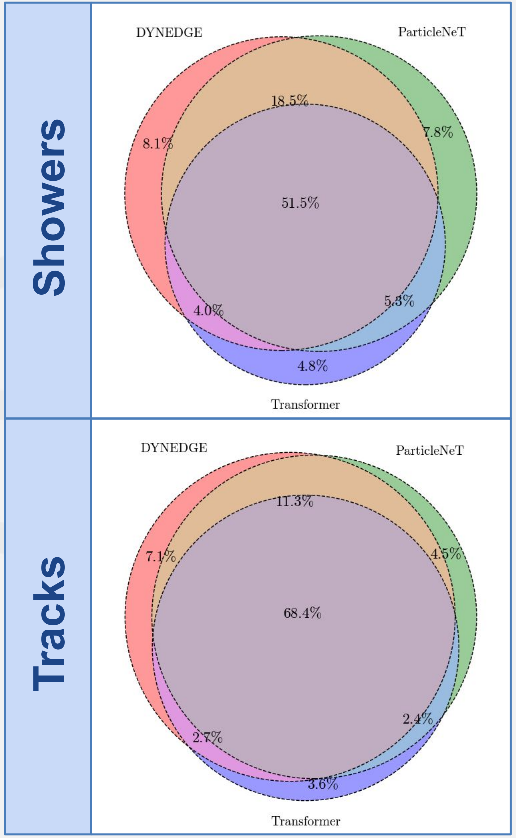






Track-Shower Classification

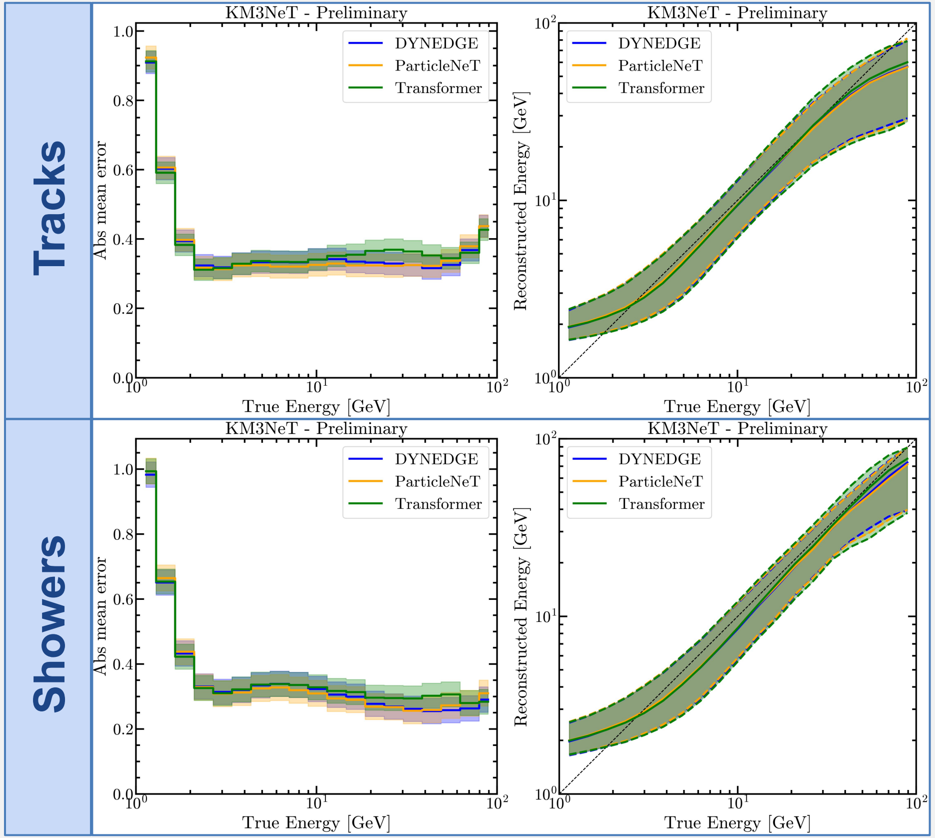
- Check if the models are targeting different events. If, for instance, one model finds it easy to classify a track as a track while the other struggles, while for a different track the opposite happens, **then the combination of both predictions** might be valuable.
- Here it has been considered events classified as a shower for a model those with a score smaller than 0.3 and tracks the ones with a score greater than 0.7.
- Most of the events are understood the same way by the three models.
- Tried a BDT with features: the three scores + the three energy reconstructions (see in a moment) but very little improvement in the classification (AUC=0.86).



Energy Reconstruction

- Very **similar performances of the three** models in the entire energy range.
- For tracks and showers all the three models **struggle with events below 3 GeV**.
- For tracks **over 60 GeV the three models have difficulties** in the prediction as well.

	N-epochs	Loss function	Time per epoch
ParticleNeT	29	LogCosh	~22mins
DYNEDGE	22	LogCosh	~23mins
Transformer	58	LogCosh	~15mins

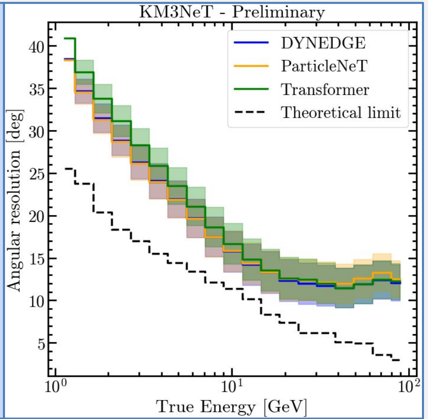


Direction Reconstruction

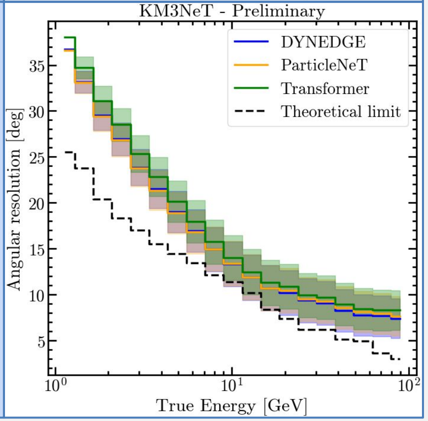
- **Better performance of ParticleNeT and DYNEDGE** over the transformer.
- Can reach **resolutions below 15º for tracks and for showers over 10 GeV**.
- VonFisherMises loss function outputs **as well a measurement of how certain the model is of a prediction**. Can be used to achieve a sample of over 30% of the events but with angular resolutions close to the theoretical limit. (See this talk at NPML2024)

	N-epochs	Loss function	Time per epoch
ParticleNeT	45	VonfisherMises	~27mins
DYNEDGE	25	VonFisherMises	~31mins
Transformer	56	GaussianNegativeLogLikelihood	~28mins

Showers



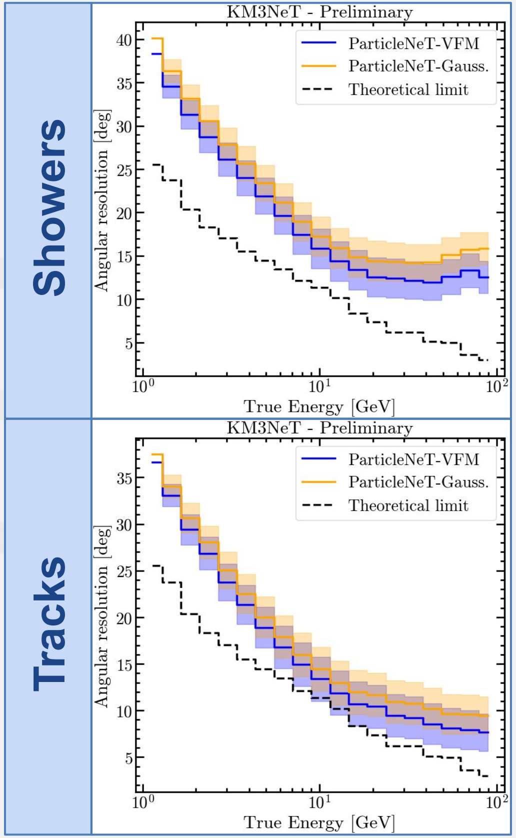
Tracks



Trying new loss functions?

- **ParticleNeT** in KM3NeT and the **transformer** model used **GaussianNegativeLogLikelihood** (GNLL) as loss function for direction reconstructions while **DYENDGE** for IceCube direction reconstruction used **VonFisherMises** (VFM). Can any of the models perform better by using a different loss function?
- **ParticleNeT** trained with the exact same hyperparameters shows an **improvement when changing from the GNLL to the VFM** in every energy bin.
- Transformer might enhance its performance as well (under study).

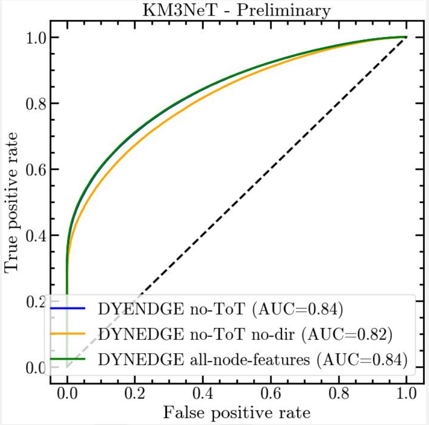
	N-epochs	Time per epoch
ParticleNeT Gauss.	29	~28mins
ParticleNeT VFM	45	~27mins

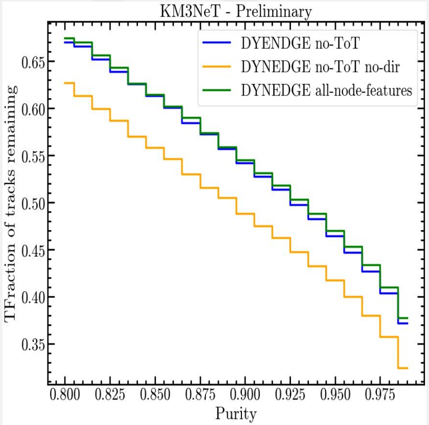


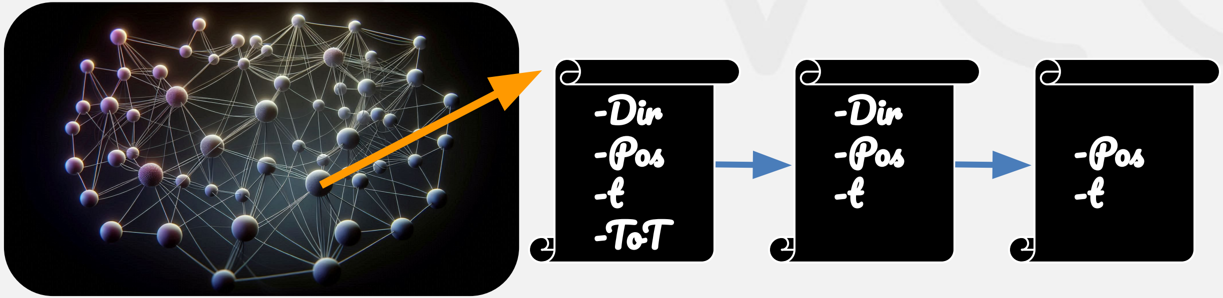
Node features studies

- How well the model **performs with less node information?**
- **ToT does not affect** the performance but **PMT direction does.**
- Need to check per task. (work in progress).

	N-epochs	Time per epoch
DYN. all	26	~27mins
DYN. no-ToT	26	~22mins
DYN. no_ToT,dir	38	~20mins



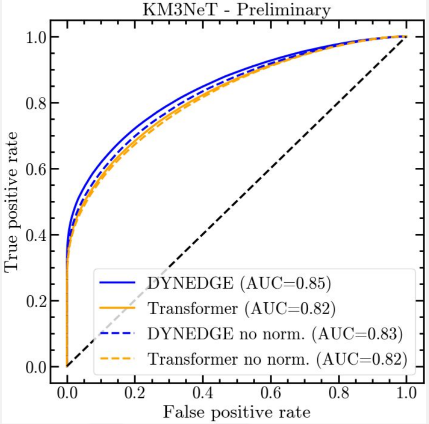


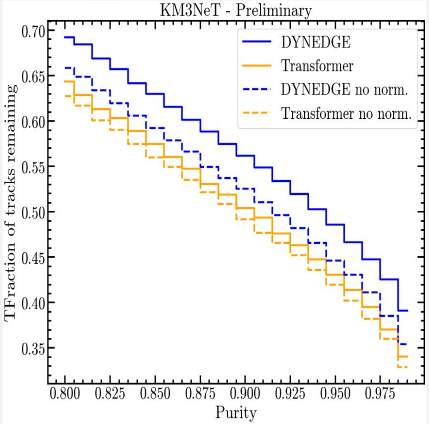


Node Inputs Normalization Effect

- In principle, **the model should be able to find out the range of values for each feature and learn how to deal with it**.
- Providing a **normalization so that the features are in the range [-10, 10]** (as in here) **enhances the performance and improves the training times**.
- Need to study and optimize the normalization? Normalizing x, y and z so that they lay within -10 and 10 can not distort the geometry? Is this possible that this also affects to the choice of neighbours?
- For sure it affects, but the optimization is work in progress.

	N-epochs	Time per epoch
DYNEDGE (Norm / No-norm)	26 / 30	21 mins / 24mins
Transformer (Norm / No-norm)	32 / 52	15.3mins / 14.9mins





Conclusions

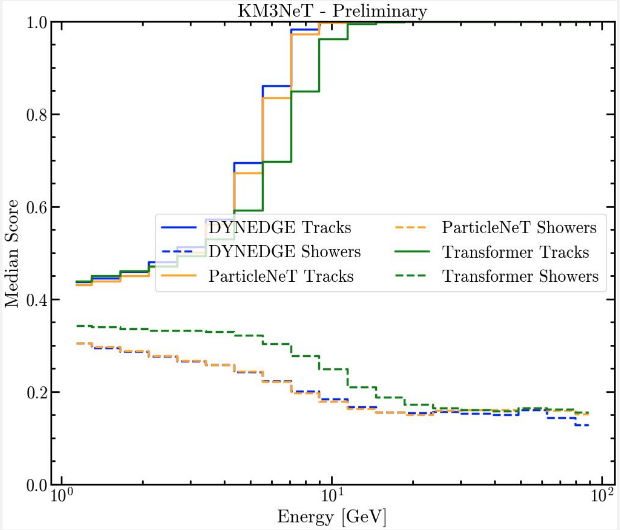
- **Deep learning** techniques are **very suitable techniques to study neutrino events in KM3NeT** detector as they capture information that with BDT or likelihood based methods for study are missing.
- **Need to perform further comparisons** and then study how well is the agreement between the predictions on simulation and data.
- It is key to see the **adaptability of models** for different detector configurations as we are working on a growing project. **See next talk about Transfer Learning!**
- IceCube, KM3NeT, P-One, Baikal-GVD…very similar detectors. Would it be nice to study how well the techniques used in one detector work in another. (work in progress, **see paper soon**)

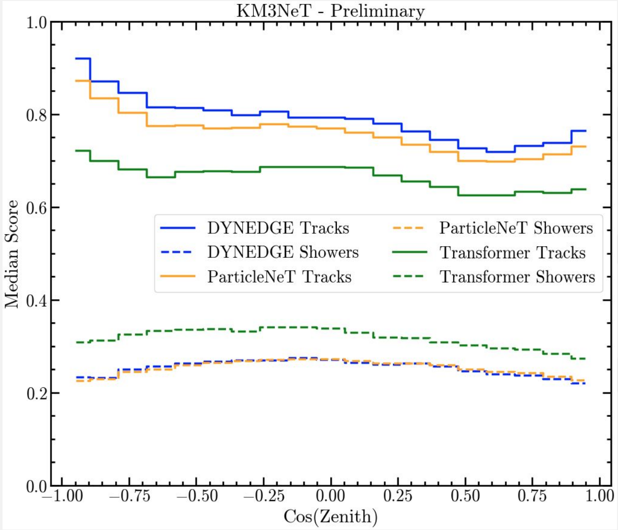
Thank you!

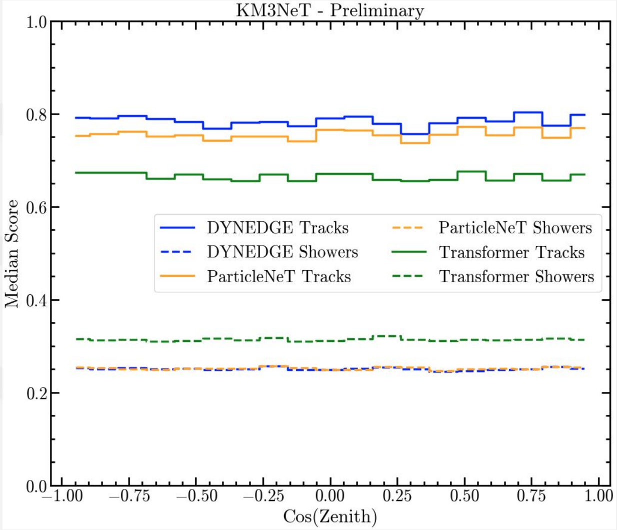
Models trained in:

Artemisa

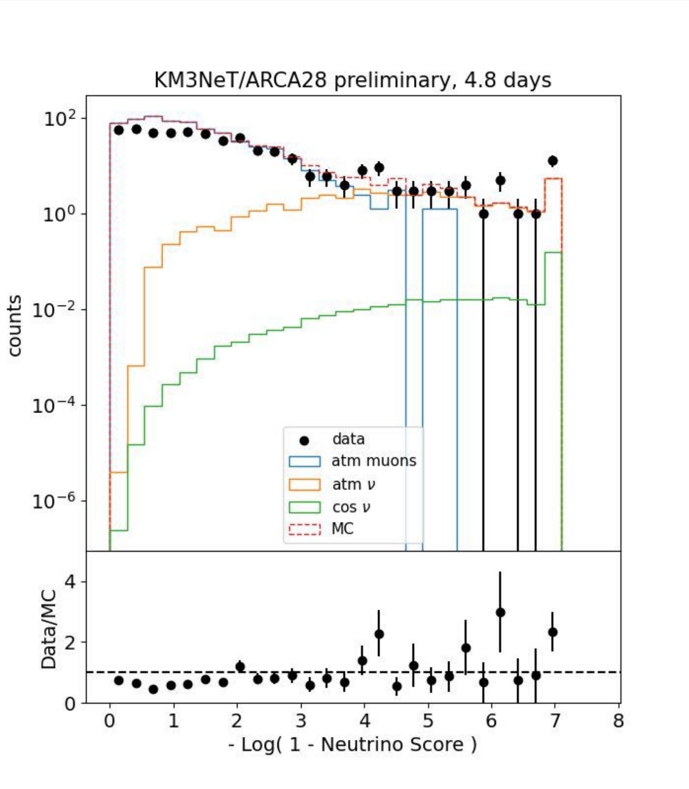
Appendix - TS score over E, Zen and Az



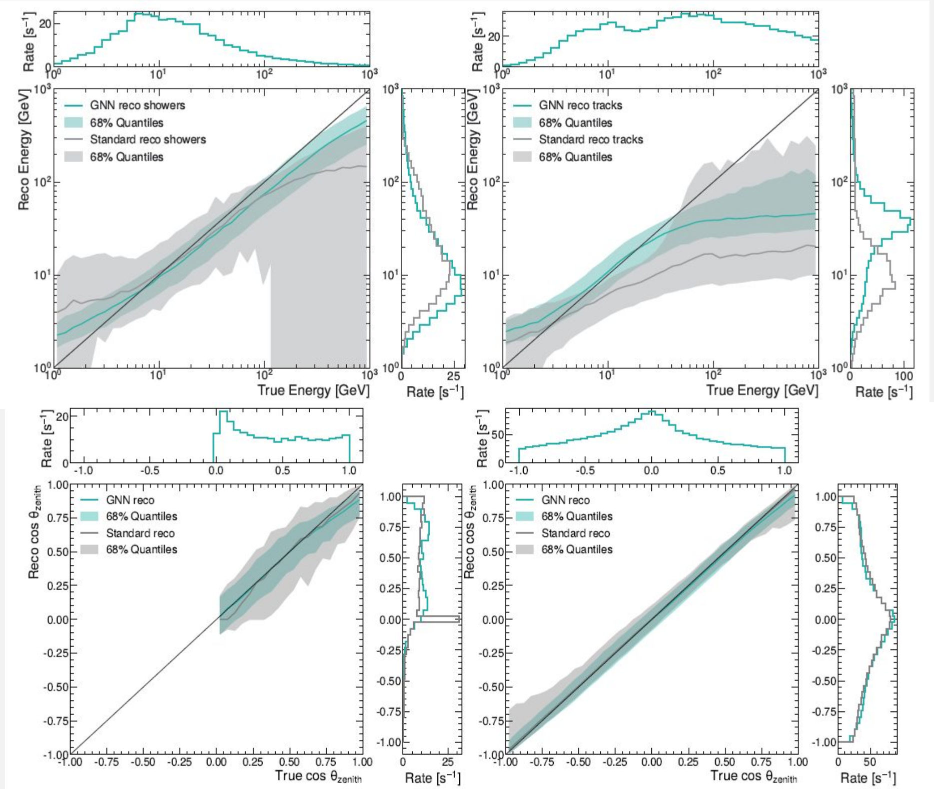




ParticleNeT in KM3NeT



[Credit A. Veutro]



[Credit S. Peña Martínez]