Enhancing low energy reconstruction and classification in KM3NeT/ORCA with transformers

28/11/2025, Caen

Iván Mozún Mateo
On behalf of the KM3NeT collaboration

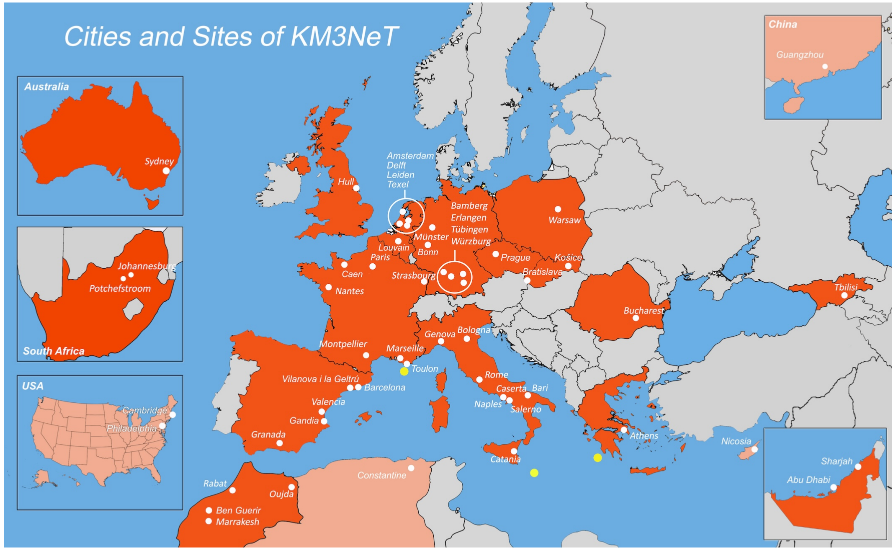
Overview

1. KM3NeT: neutrino telescopes
2. Reconstruction challenges in KM3NeT
3. Transformers for reconstruction
4. Multi-detector reconstruction in KM3NeT/ORCA
5. Conclusions

KM3NeT

KM3NeT is an **international collaboration**

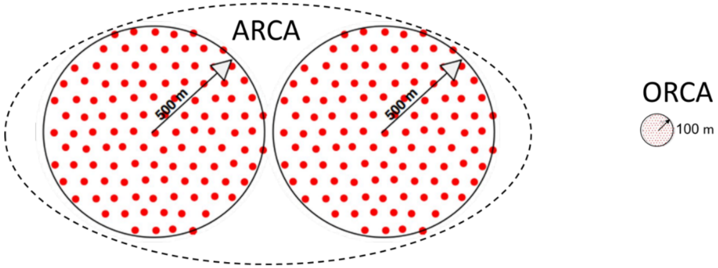
- 22 countries
- 65 partner institutes
- ~250 members



Two **undersea neutrino telescopes**

- **KM3NeT/ARCA**
  - Optimized for 1 TeV – 10 PeV
  - Identify high-energy neutrino sources in the Universe.
  - 36m vertical spacing and 90m horizontal spacing
- **KM3NeT/ORCA**
  - Optimized for 1 – 100 GeV
  - Determine the mass ordering of neutrinos.
  - 9m vertical spacing and 20m horizontal spacing

Currently under construction: **ORCA33 (29%), ARCA51 (22%)**



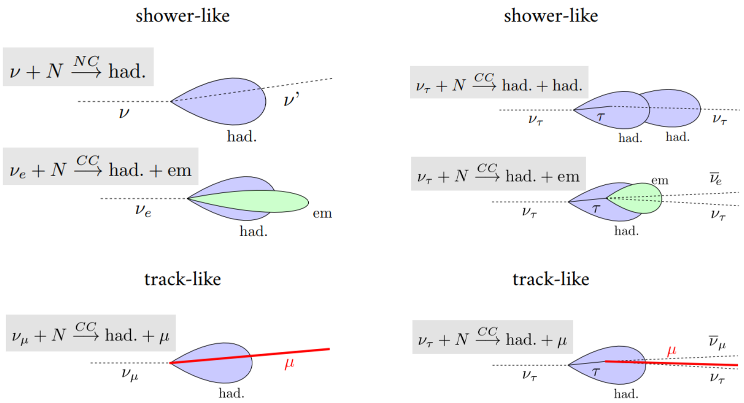
KM3NeT: neutrino telescopes

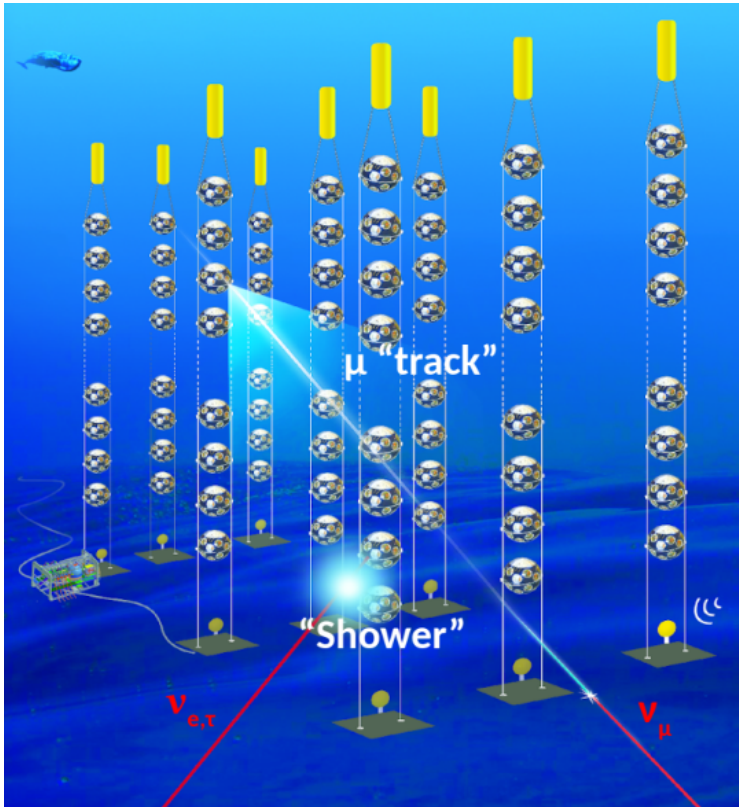
Same **technology**:

- 1 (2) building block(s) for ORCA (ARCA)
- 115 vertical detection units (DUs) per block
- 18 digital optical modules (DOMs) per DU
- 31" PMTs per DOM

Same **detection principle**:

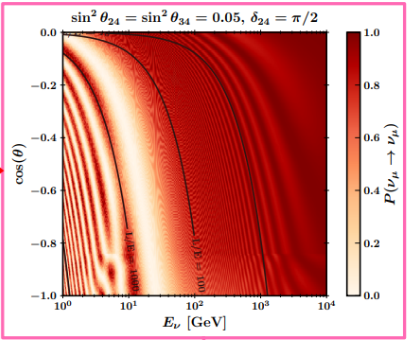
Light collection from ***Cherenkov radiation*** *emitted by particles traveling faster than the speed of light in water*

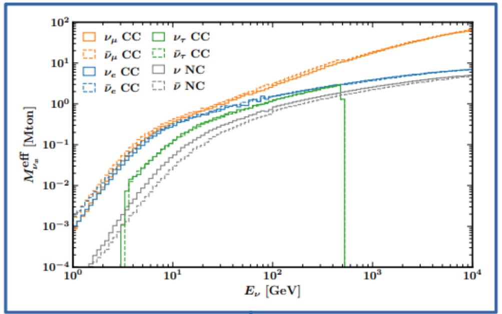




Importance of a good reconstruction

Δm^2_{ij}
θ_{ij}
δ_{ij}

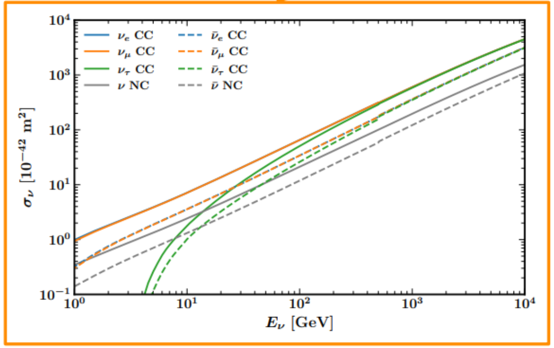


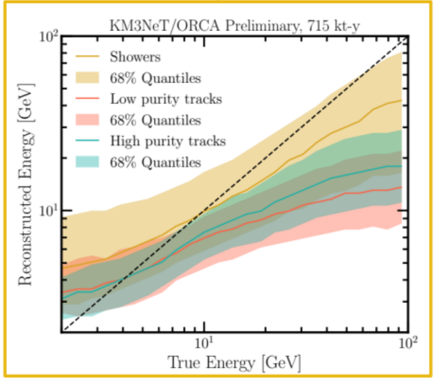


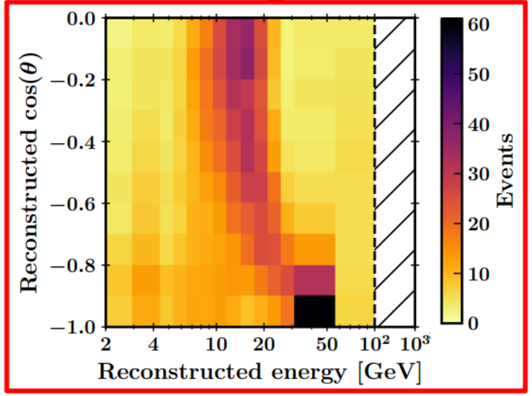
To produce good estimations of **predicted event rates** for an analysis, many pieces are considered: **fluxes**, **oscillations parameters**, **cross sections**, **effective masses** and **detector response**

One of the main sources of systematic errors in the analysis is the detector response R_i, modeled by the reconstruction algorithm.

$$\phi^{\nu_y}_{atm}(E_t,\theta_t) \times P_{\nu_y \to \nu_x}(E_t,\theta_t) \times \sigma_{\nu_x}(E_t) \times M^{\nu_x}_{eff}(E_t) \times R_i(E_t,\theta_t,\nu_x,E_r,\theta_r) \Rightarrow \mu_i(E_r,\theta_r)$$





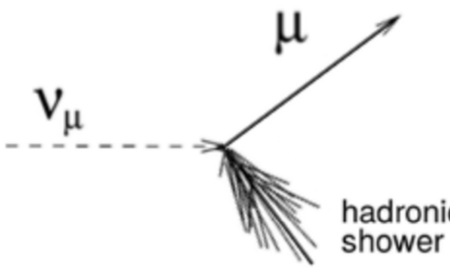


Challenges of neutrino physics in ORCA

Reconstruct neutrinos from light

Maximum Likelihood Fit (MLF) algorithms
- Reconstruct under track or shower hypothesis
- Do not reconstruct the neutrino itself

nuT model beyond MLF
- Directly reconstruct neutrino
- Simultaneously reconstruct all hypothesis

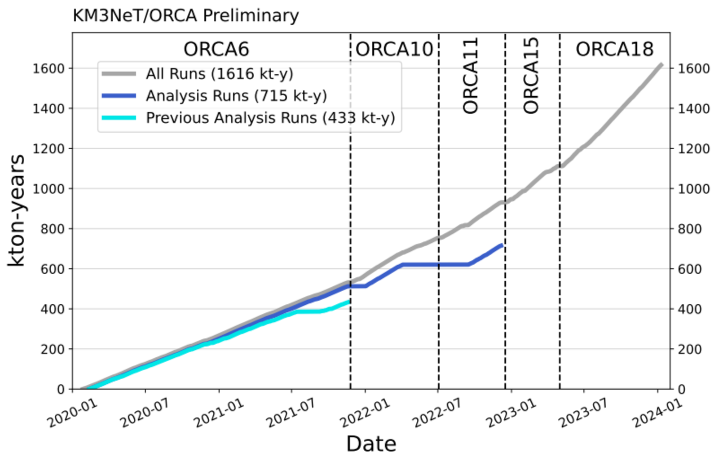


A growing detector

Realistic Monte Carlo samples are generated based on actual data-taking runs, capturing the complexities of deep-sea conditions

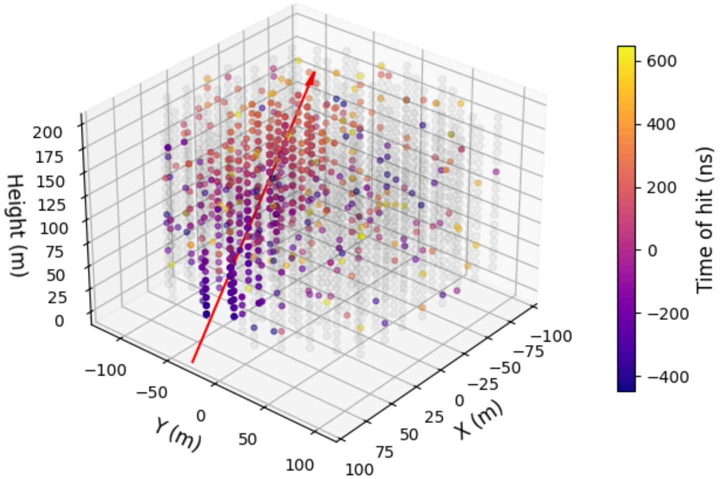
Need to use data and computing resources efficiently

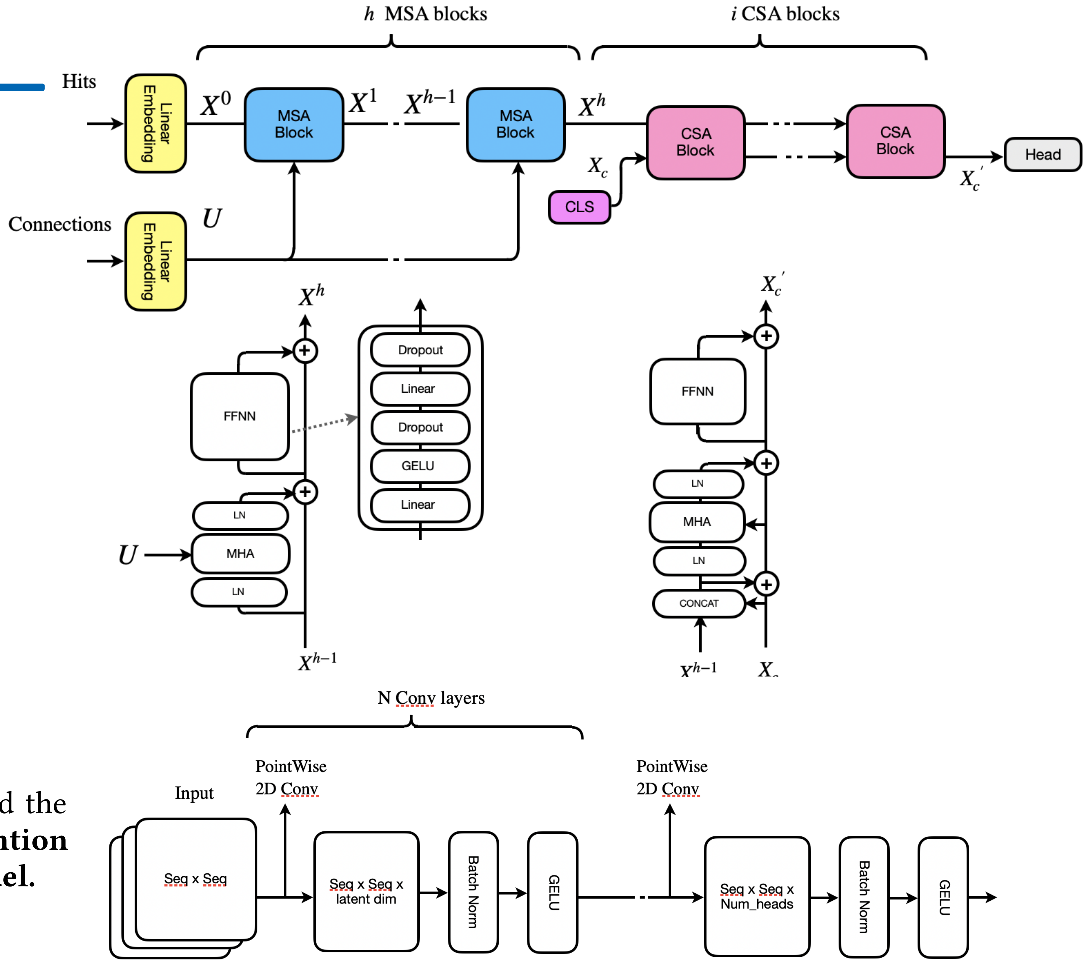
Pre-trained models are leveraged to propagate information across configurations



nuT architecture

The input data is the low-level hit information that composes the light pattern detected in the telescope, composed by **all type of hits**.





The information is processed in parallel by the transformer and the high-level information is extracted in the attention blocks. **Attention masks provide detector and physics knowledge to the model.**

The class token is an abstract representation of the event.

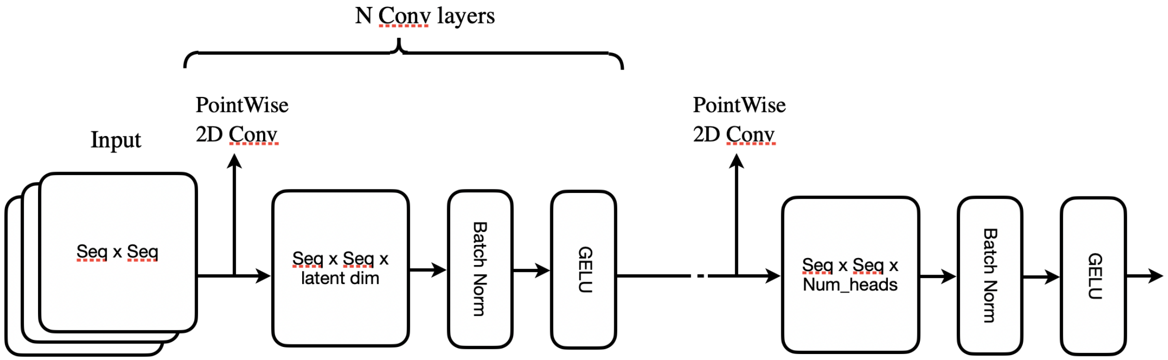
Attention masks

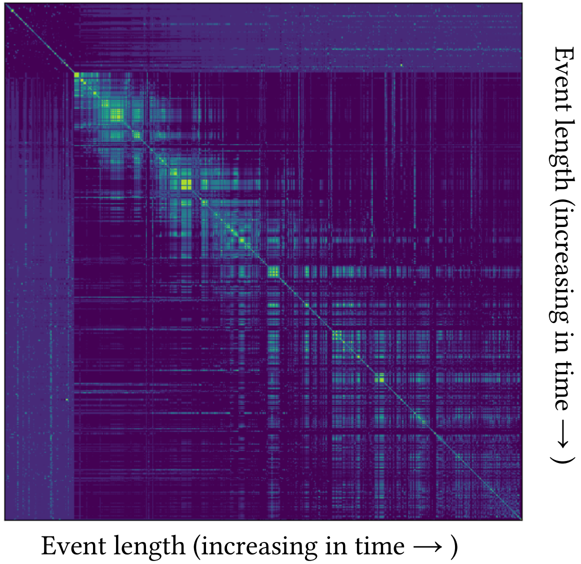
Detector and physics information introduced in the model through pairwise relationships

- **Causality:** space-time relativistic distance
- **Distance:** euclidean distance
- **Coincidence:** channel coincidence for DU, DOM and PMT

in the **self-attention mechanism**

$$Attn(Q,K,V)=SoftMax\left(\frac{Q\cdot K^T}{\sqrt{d}}+U\right)\cdot V$$





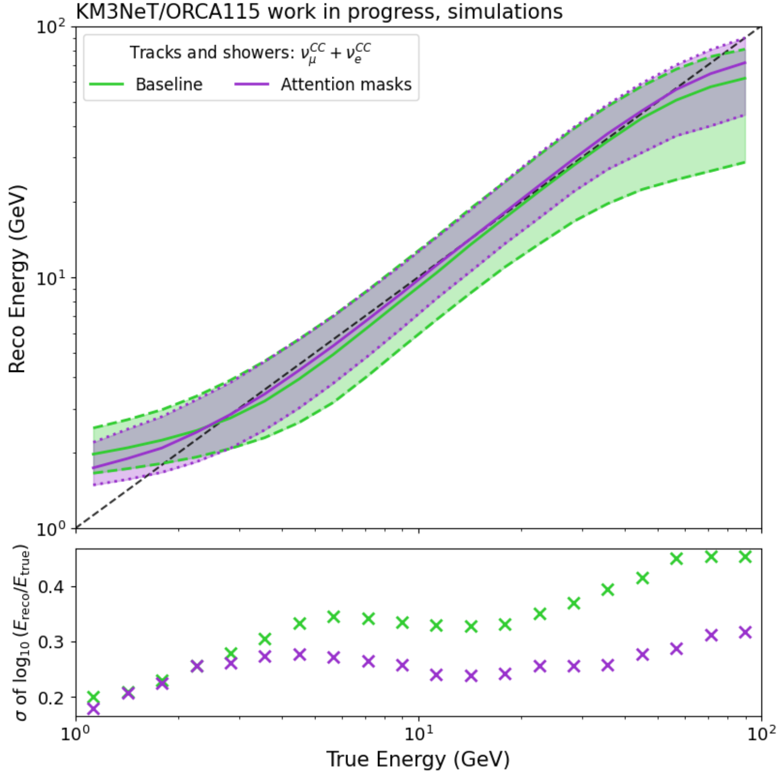
The attention mask show clear patterns of light pulses coming from a charged particle or from optical background ⟶ **The model understands the physics**

Attention masks

Detector and physics information introduced in the model through pairwise relationships in the **self-attention mechanism**.

Improvements in energy reconstruction:

- Better energy estimation at low energies.
- Better energy estimation at high energies.
- Less spreading of predicted energies.

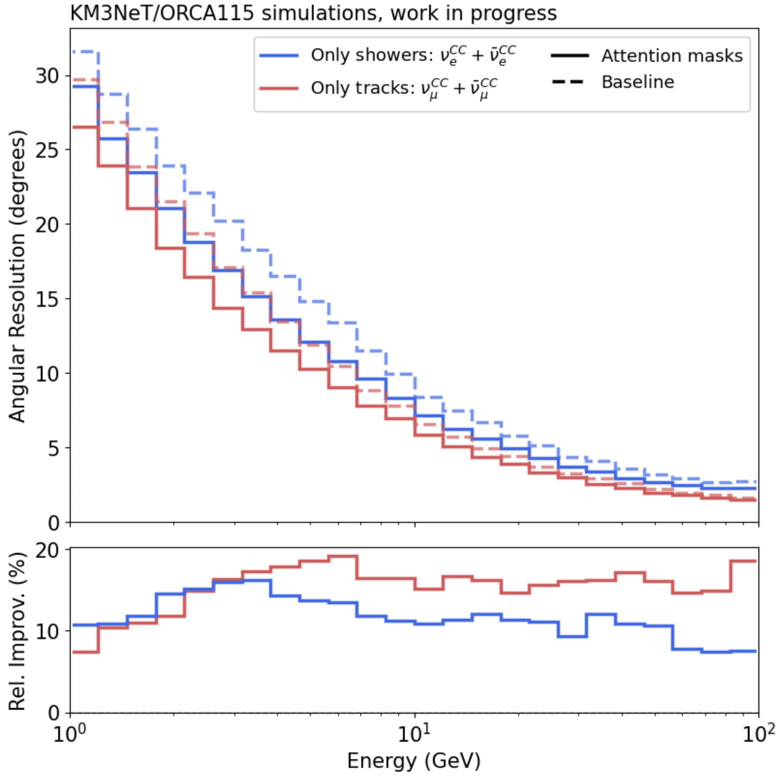


Attention masks

Detector and physics information introduced in the model through pairwise relationships in the **self-attention mechanism**.

Improvements in direction reconstruction:

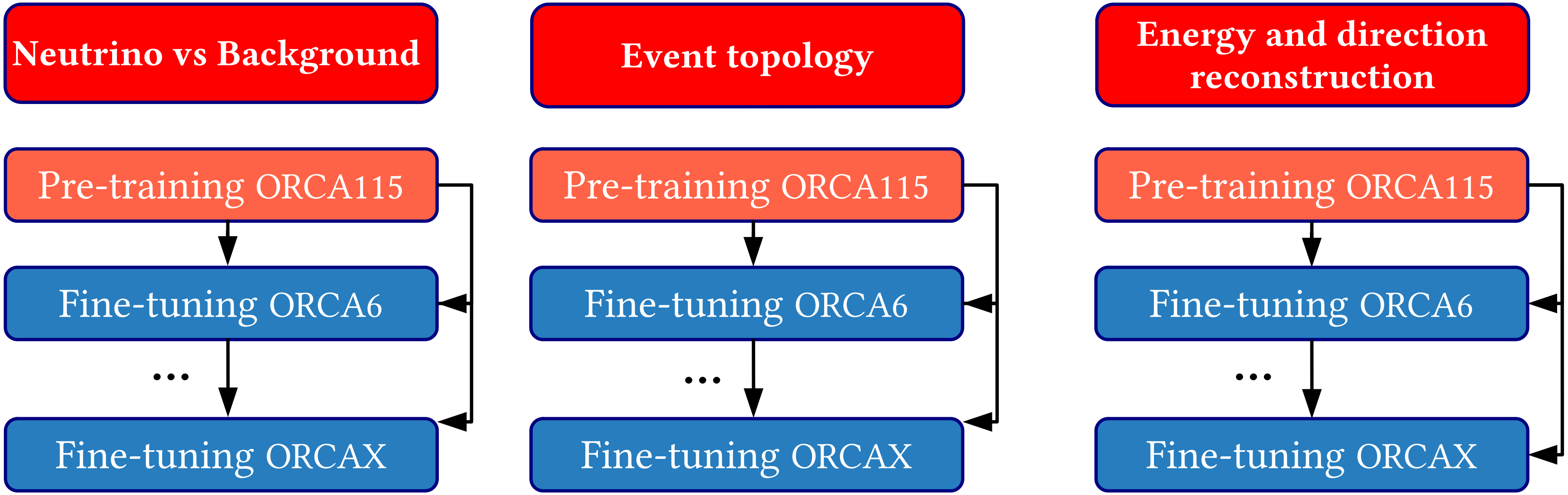
- Improvement of 10% for track-like events.
- ~10% improvement at low energies for showers.
- Above 15% improvement above 10 GeV for showers.



Learning approach for KM3NeT/ORCA

Motivation: the transformer is a language model

- KM3NeT/ORCA115 is the final detector, having all the possible neutrino physics encapsulated
- The information about KM3NeT/ORCA115 is used to understand our current detector
- All the knowledge will be retained when fine-tuning on other smaller configurations

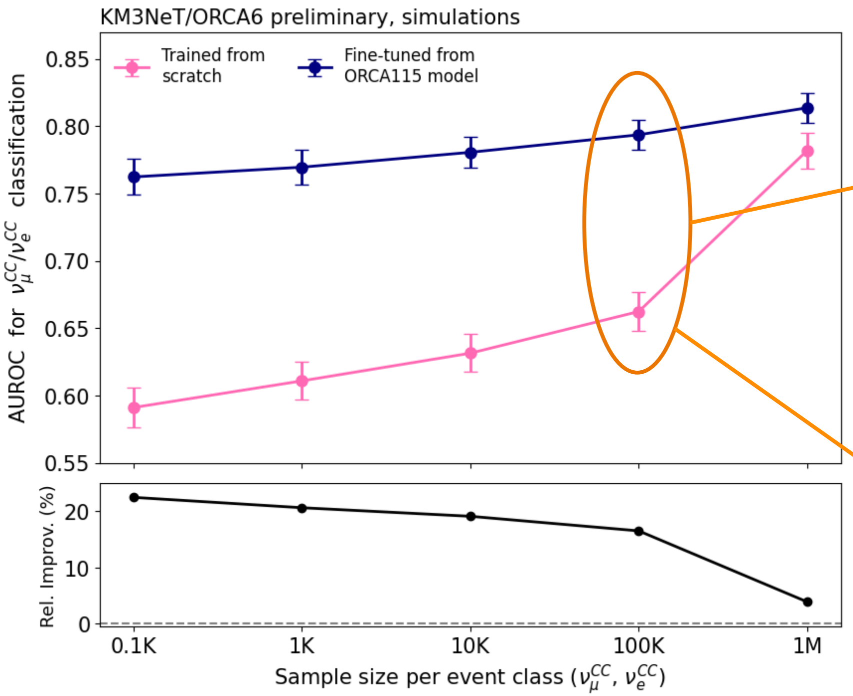


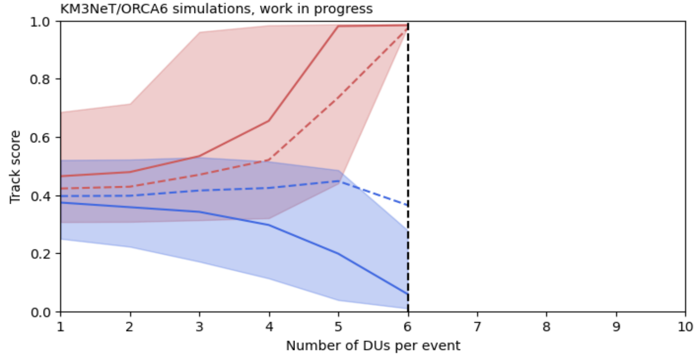
A growing detector

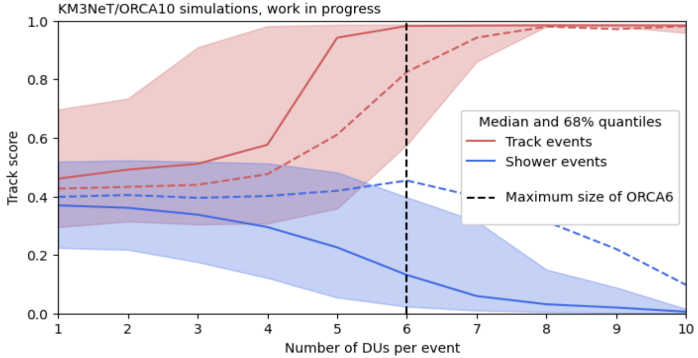
Realistic Monte Carlo samples are generated based on actual data-taking runs, capturing the complexities of deep-sea conditions

Pre-trained models are leveraged to propagate information across configurations w.r.t. **models trained from scratch**

Better discrimination power when adding more lines



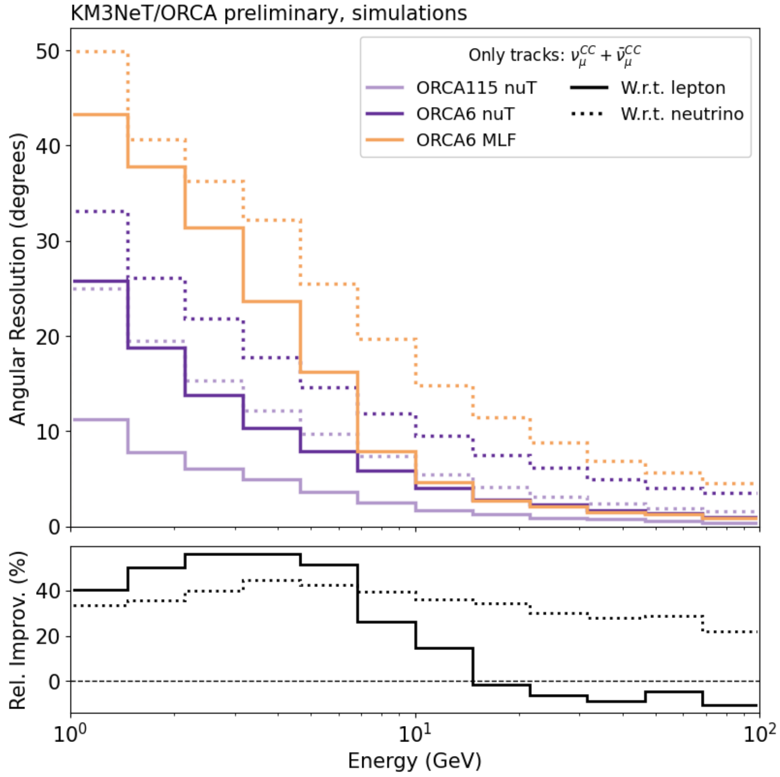


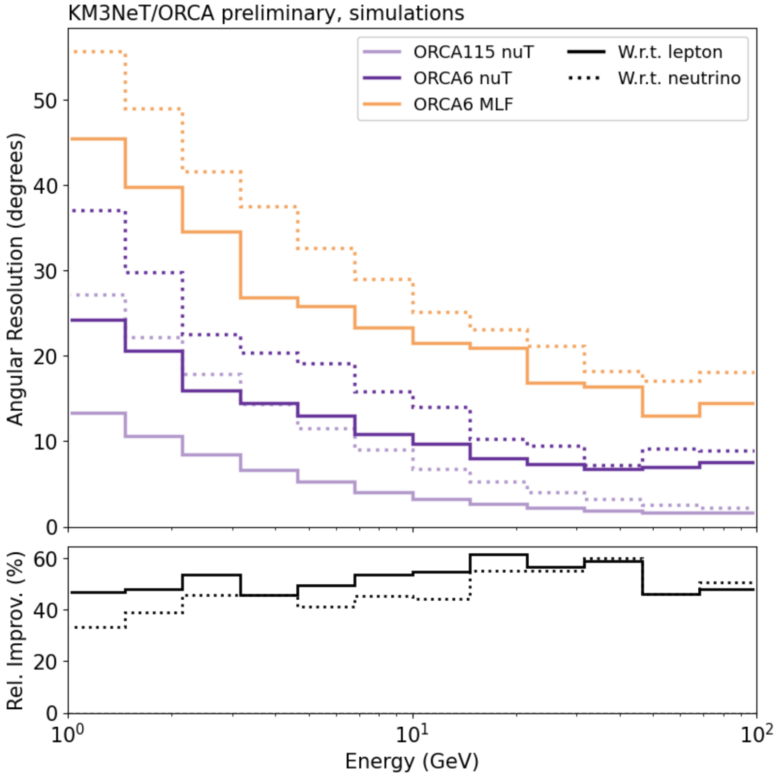


Reconstruction of neutrinos: angular resolution

MLF: <10° resolution above 10 GeV

nuT: improved resolution w.r.t. neutrino (+30% for tracks, +40% for showers)



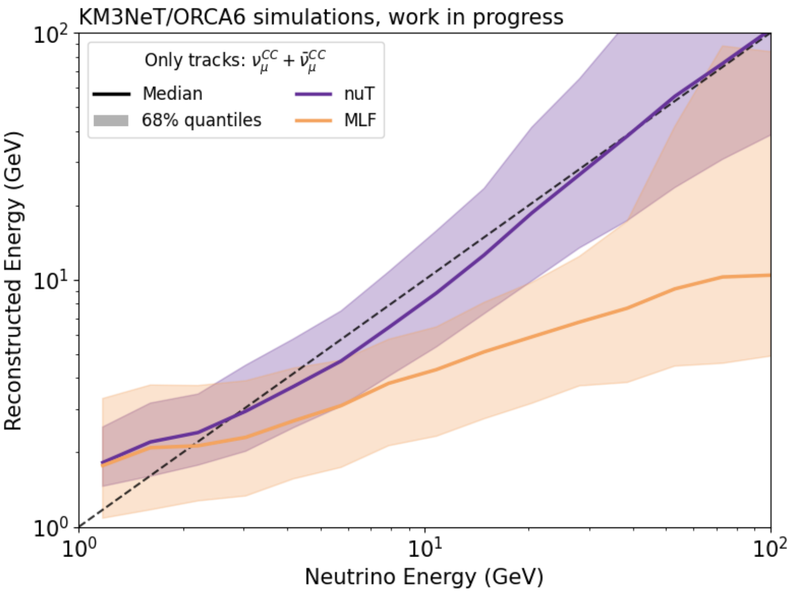


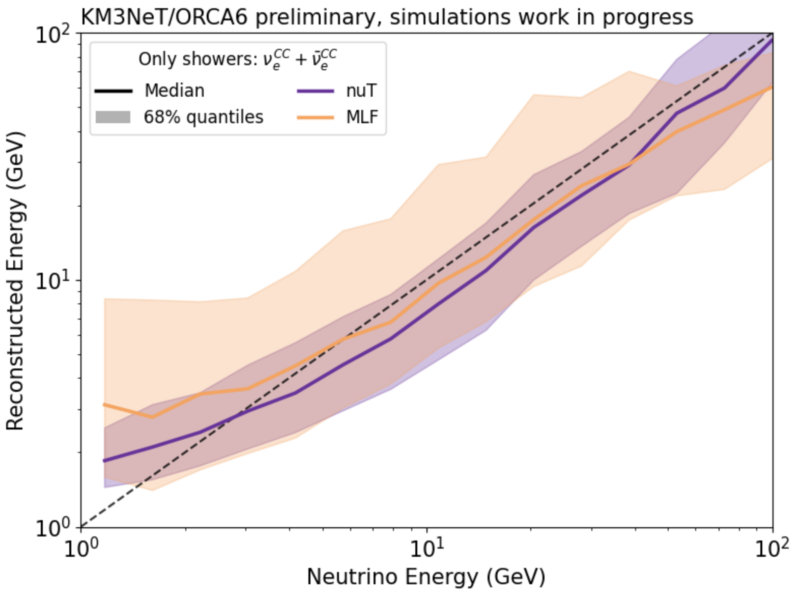
Reconstruction of neutrinos: energy estimation

MLF: underestimation in track reconstruction from missing hadronic component, compensated in pure shower events

nuT: accounts for visible and non-visible energy

Saturation effect at high energies due to event containment



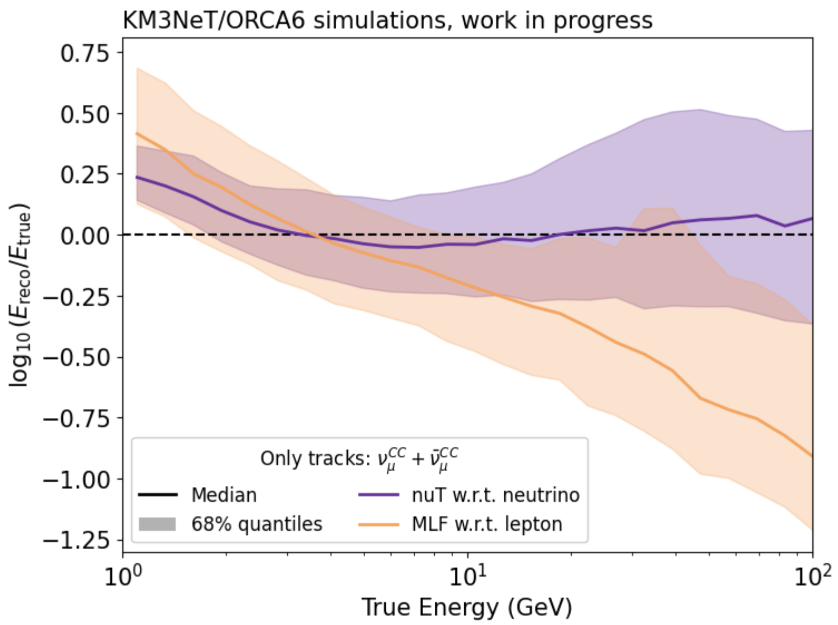


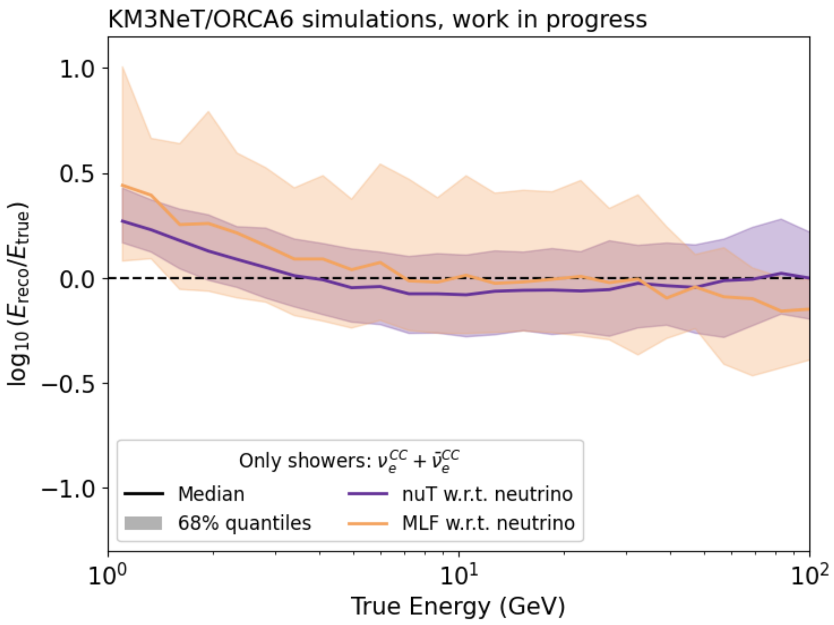
Reconstruction of neutrinos: energy resolution

MLF: underestimation in track reconstruction from missing hadronic component, compensated in pure shower events

nuT: compromise found when reconstructing tracks and showers simultaneously

Saturation effect at high energies due to event containment





Conclusions

Transformers in KM3NeT

- Reconstruct directly neutrino properties accounting simultaneously for track and shower hypothesis
- Attention masks:
  - Introduce physics and detector knowledge in the model
  - Boost training and increase performance

Propagate information across telescopes through fine-tuning

- Transformers are particularly effective to deal with small detectors and very limited data
- Information propagated across detectors
- Speeds up training and boosts model robustness

Thank you for your attention!