# SMALL THINKS BIG
# Transfer learning in KM3NeT/ORCA with transformers

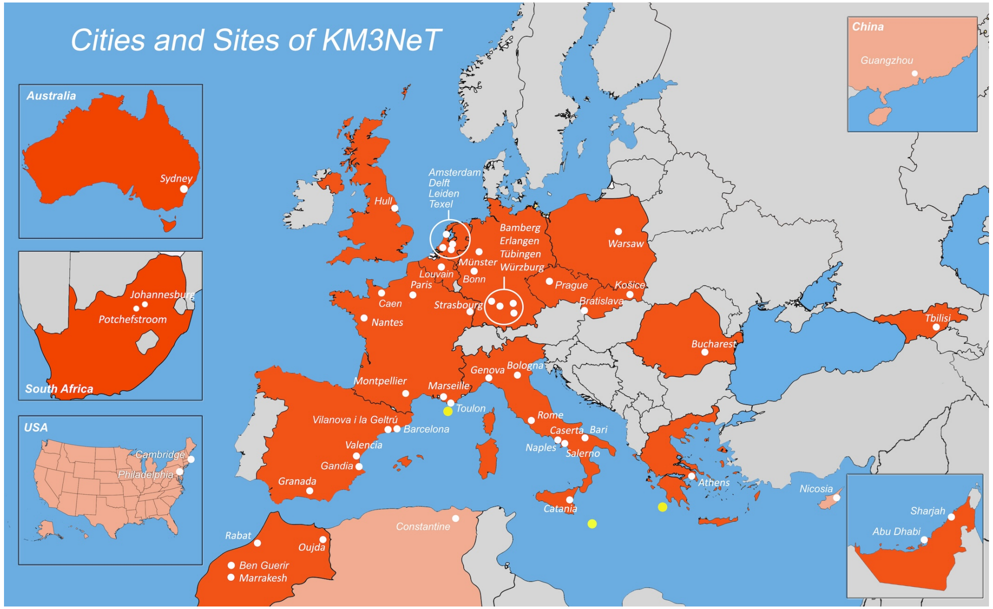
30/09/2024, Caen

Iván Mozún Mateo, Antonin Vacheret
On behalf of the KM3NeT collaboration

# Overview

# KM3NeT

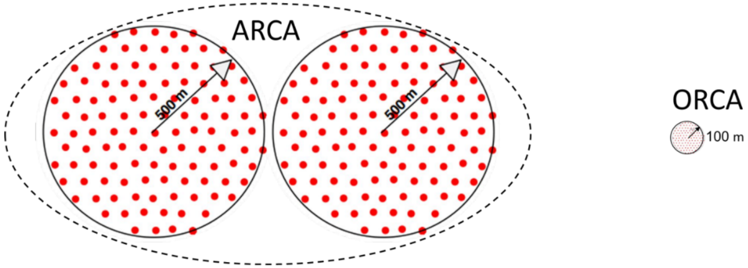
KM3NeT is an **international collaboration**
- 22 countries
- 65 partner institutes
- ~250 members




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: ORCA23 (20%), ARCA28 (12%)



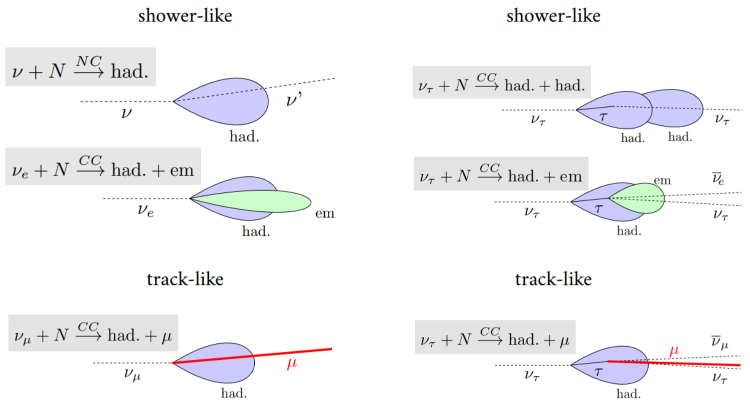
# KM3NeT: neutrino telescopes

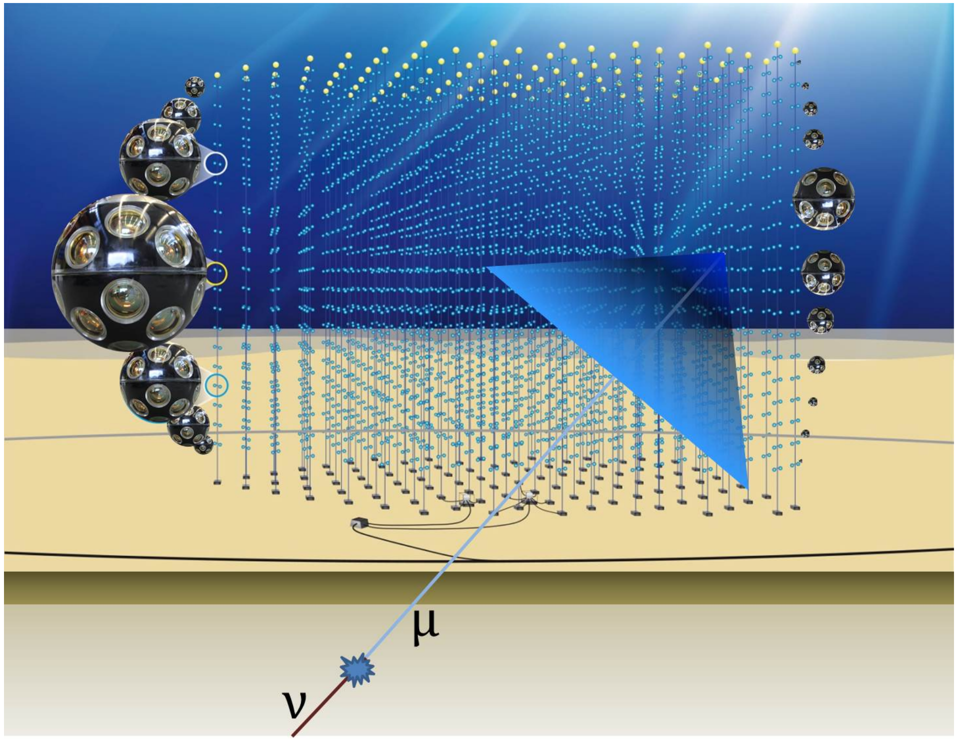
Same **technology**:

- 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**:

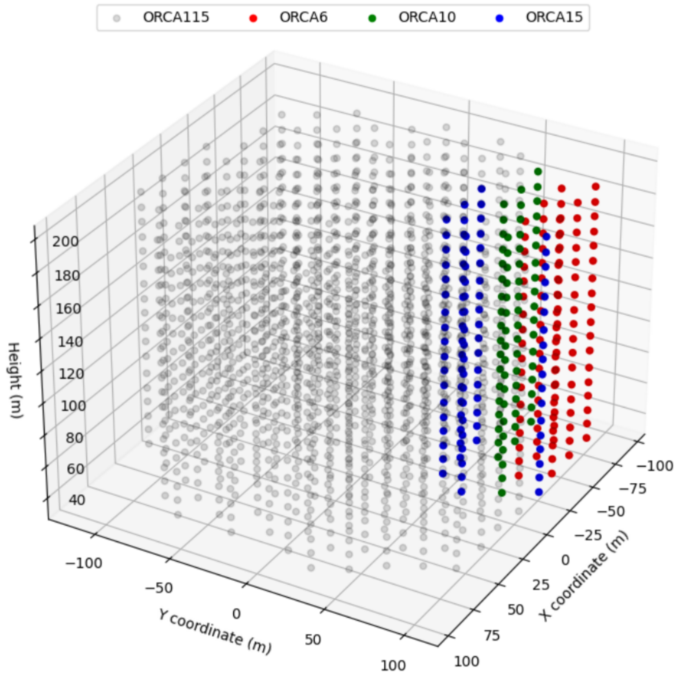
*Light collection from **Cherenkov radiation** emitted by particles traveling faster than the speed of light in water*

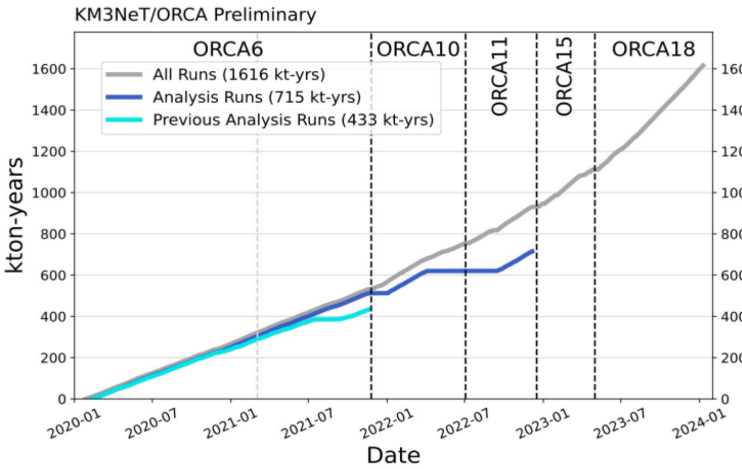


# Building the detectors

KM3NeT telescopes collect, process and analyze data as they are being built.



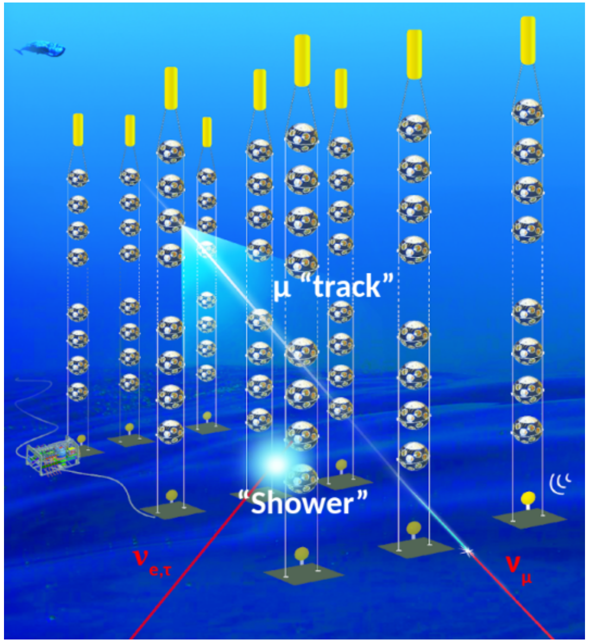

**Foundation model in KM3NeT?**

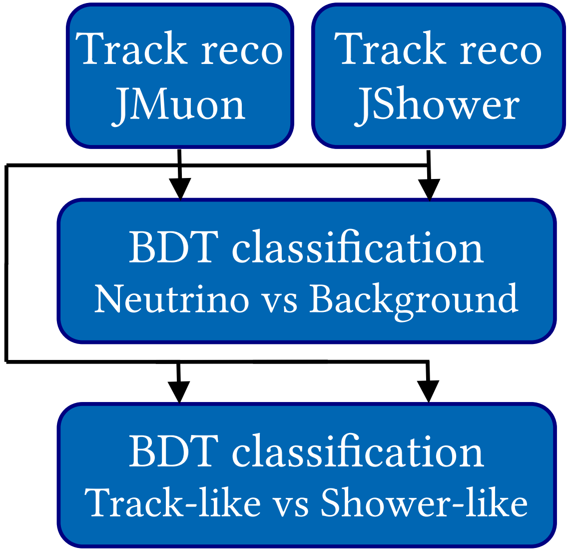
- learns as the detectors grow
- can handle multiple geometries
- can handle both KM3NeT/ORCA and KM3NeT/ARCA
- classification ↔ reconstruction

# Reconstructing neutrino physics

AI SS.AI
AI for science, science for AI

The **official KM3NeT pipeline for reconstruction and classification** relies on algorithms that are applied separately for track-like event or shower-like events. Then, simple BDTs are applied on the reconstructed variables for classification tasks.

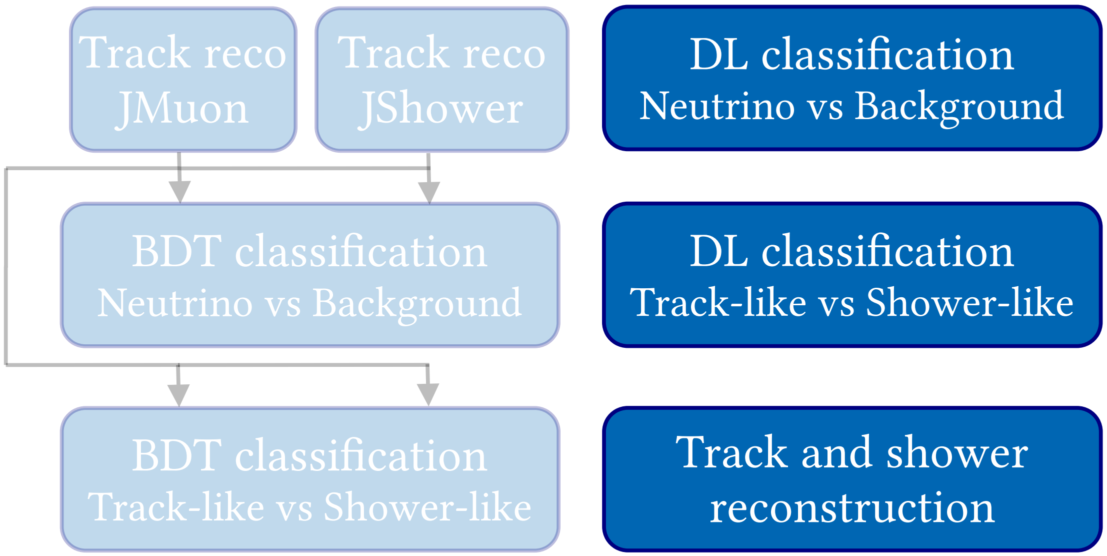



**Classic approach**

# Reconstructing neutrino physics

The **official KM3NeT pipeline for reconstruction and classification** relies on algorithms that are applied separately for track-like event or shower-like events. Then, simple BDTs are applied on the reconstructed variables for classification tasks.




Classic approach

**Deep Learning representation**

Novel **deep learning** techniques use low-level information from the detector, i.e. light pulses to

1. Let the model decide the features to use
2. Generalise over a large input domain dimensions
3. Perform different tasks

Classification and reconstruction are performed **independently**, and for any type of event.

Large DL models needs **huge amounts of very diverse data** to generalize and interpolate, improving the performances of existing algorithms.

# KM3NeT Deep Learning Outreach

AI for science, science for AI

Various DL models tested. So far, no one is considered for official analysis.

**Convolutional Neural Networks**

- 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., A NTARES)

**Deep Neural Networks**

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

**Graph Neural Networks:**

- Development of detector calibration and graph neural network-based selection and reconstruction algorithms for the measurement of oscill ation 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)

**More details here:** A Comprehensive Insight into Machine Learning Techniques in KM3NeT (J. Prado)

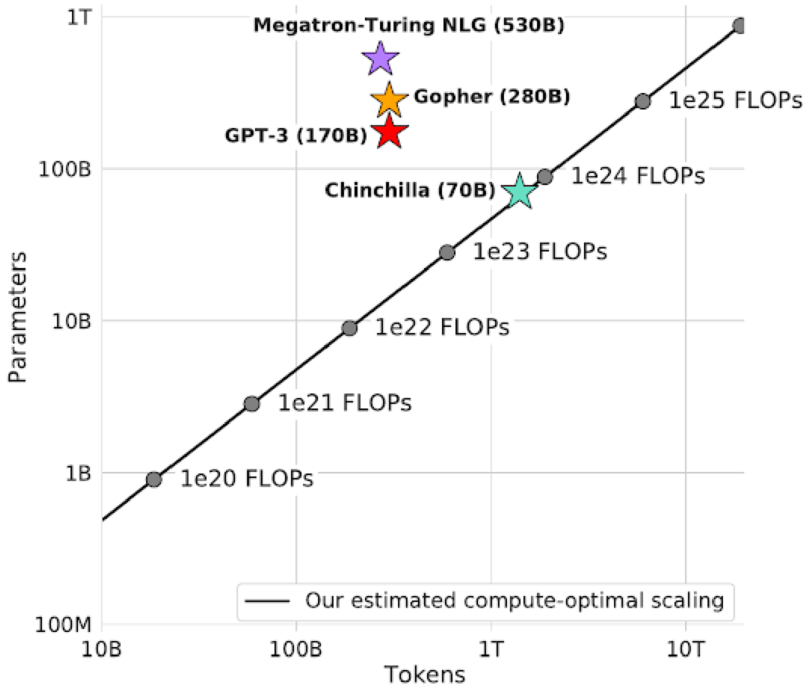
# Need for data in large language models

Neutrino telescope data is described as a set of spatial points with timing & charge information (point-cloud data), hence, most developed DL architectures are based on GNNs.

**Language models are starting to overtake but…**

- **lot of trainable parameters**
- **lot of training data**

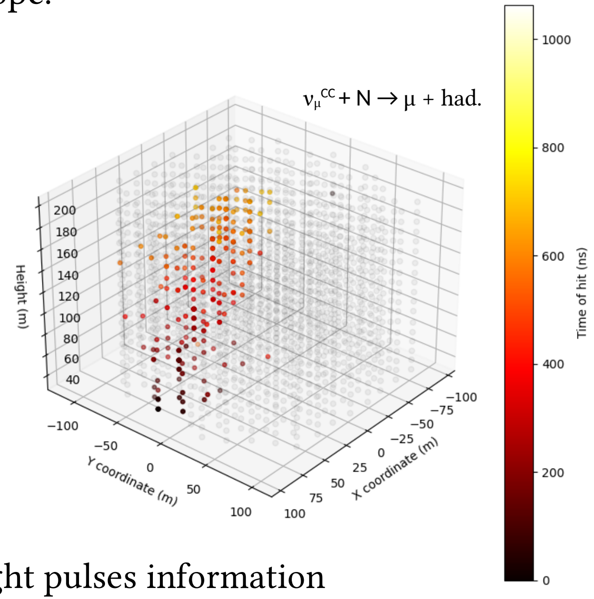
The data is too complex and requires a lot of computing resources to be produced and to encapsulate all the physics ⟶ **we must be efficient**




Scaling law for trainable parameters and tokens for large language models
arxiv.org/abs/2203.15556

# Transformer architecture

The input data is the low-level hit information that composes the light pattern detected in the telescope.

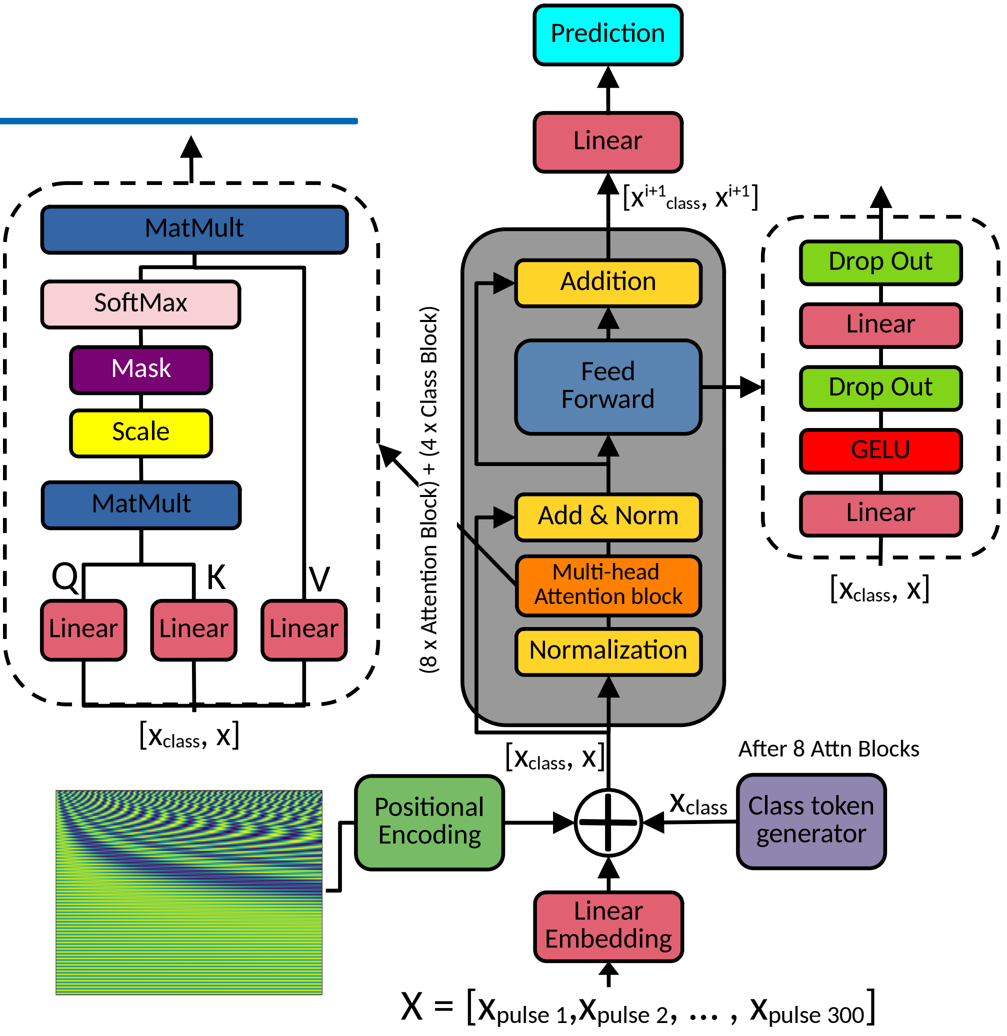



The light pulses information

$$\mathbf{X_{pulse}} = [\mathbf{pos_x}, \mathbf{pos_y}, \mathbf{pos_z}, \mathbf{dir_x}, \mathbf{dir_y}, \mathbf{dir_z}, \mathbf{t}, \mathbf{ToT}]$$

is processed in parallel by the transformer and the high-level information is extracted in the attention blocks.

**Model has ~1.6M trainable parameters.**




**Neutrinos in Deep Ice (Kaggle Competition):** arxiv.org/abs/2310.15674
**Particle Transformer for Jet Tagging:** arxiv.org/abs/2202.03772

# Transfer learning studies

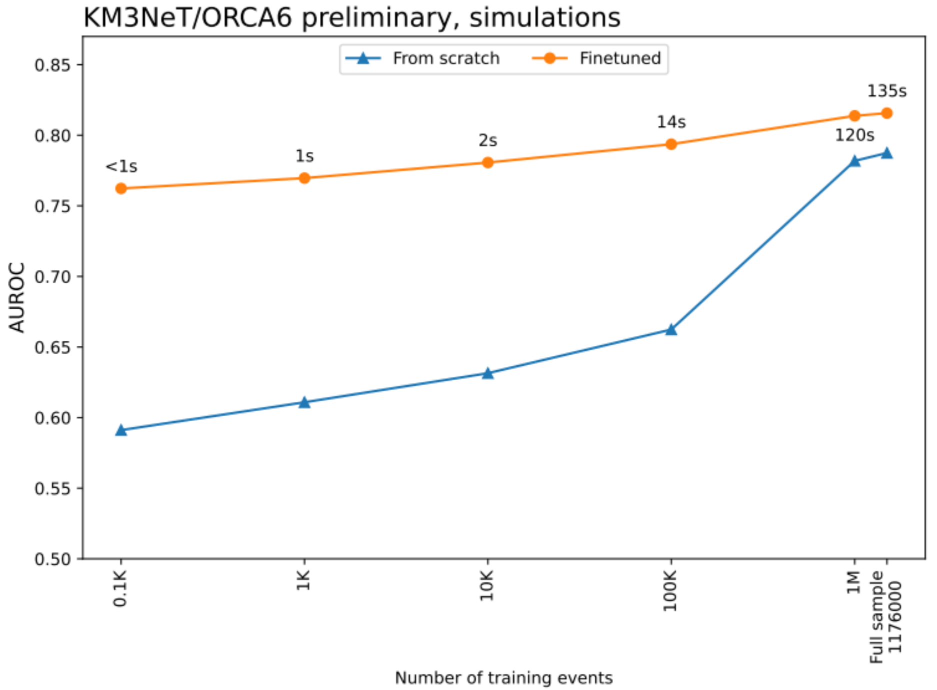
**Multiple tasks**
- Classification and reconstruction done together
- Test the capacity of the model

**Multiple detector geometries**
- A single model for all detectors: ORCA115 ⟶ ORCAX
- Run-by-Run approach:
  - MC based on data to reduce discrepancies
  - **Not enough data** to train models for every detector

**Efficient use of resources**
- Saves time and increases performance
- The information is propagated across detectors

**Multi-detector study**
ORCA6 (Feb20 – Nov21): 4075 runs
ORCA10 (Dec21 – May22): 1889 runs
100k tracks & 100k showers

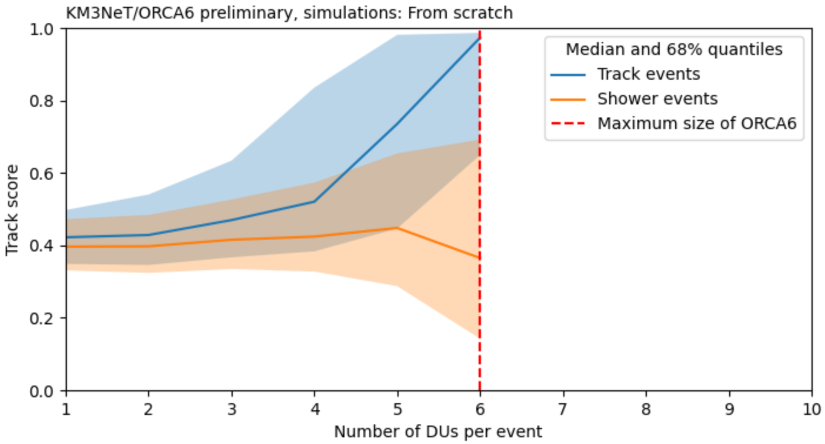
**Multi-task study**
ORCA115 dataset 1: track-shower
ORCA115 dataset 2: energy
850k tracks & 850k showers each



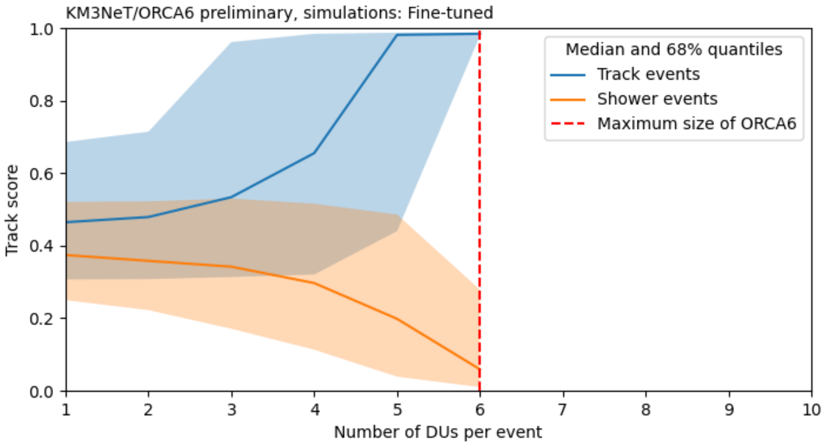

AUROC value for track-shower classification with KM3NeT/ORCA6 data.
The AUROC curves are shown as function of the training size sample.

The model is able to **interpolate** to **non-existing DUs information** because it pre-learned the full geometry.

**From scratch**




**Fine-tuned**



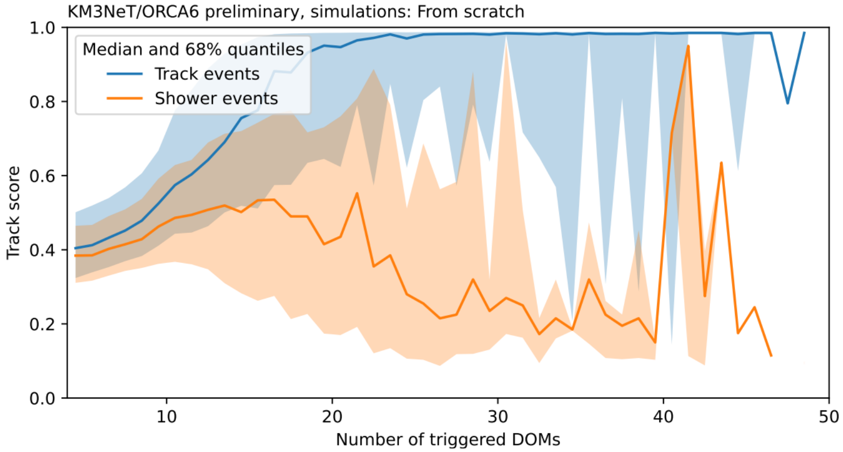

**Track-shower classification:**

- Limited data: 200k events for a detector with 6 lines is not enough to do a proper separation
- Performance: fine-tuned model works way better than the scratch one
- High dependence on event geometry: not enough discrimination with few lines

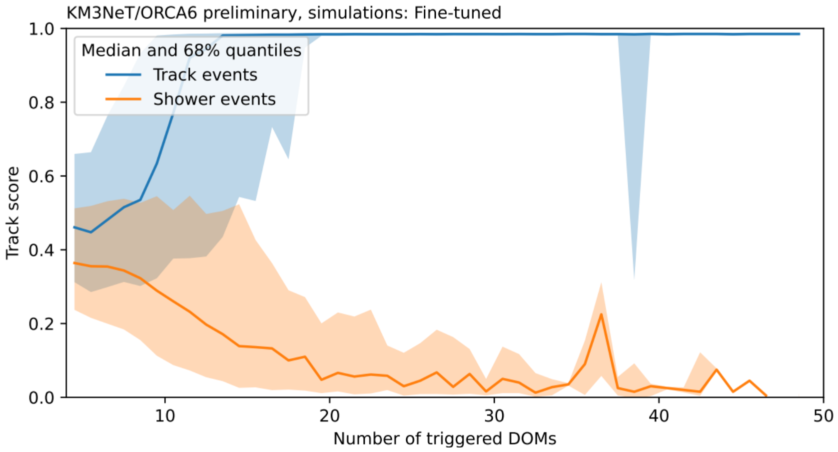
# Preliminary results on multi-geometries for KM3NeT/ORCA6

The model is able to **interpolate** to **non-existing DUs information** because it pre-learned the full geometry.

From scratch




Fine-tuned



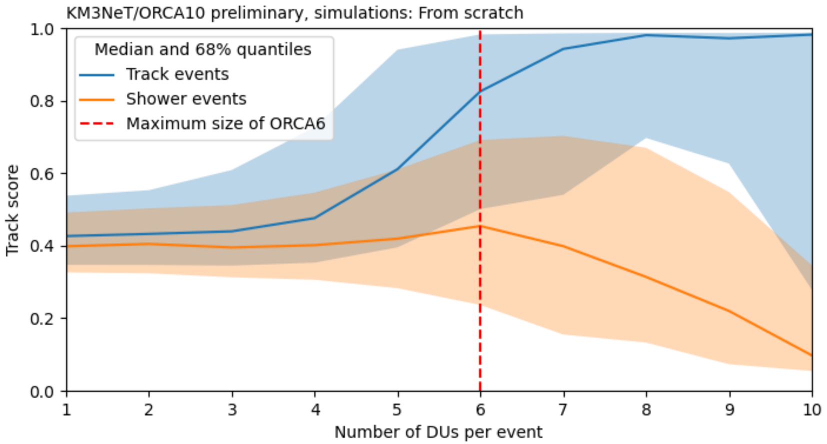

**Track-shower classification:**

- Limited data: 200k events for a detector with 6 lines is not enough to do a proper separation
- Performance: fine-tuned achieves separation in events with above 10 triggered DOMs
- Peak at ~40 triggered DOMs: fine-tuned model compensates the low statistics

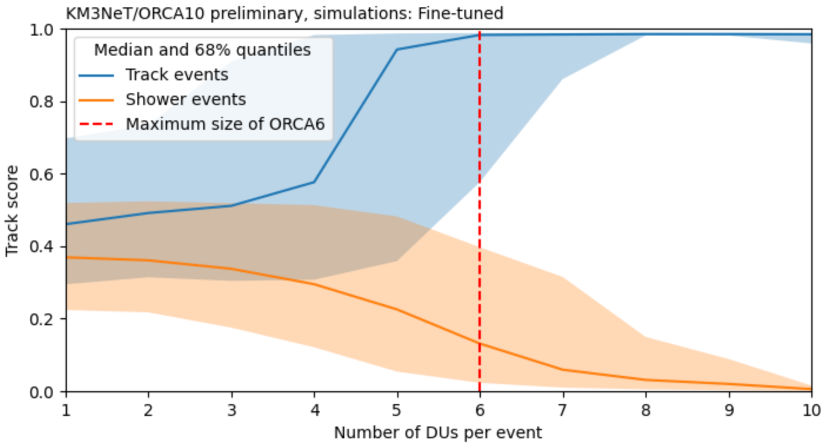
# Preliminary results on multi-geometries for KM3NeT/ORCA10

The model is able to **interpolate** to **non-existing DUs information** because it pre-learned the full geometry.

**From scratch**




**Fine-tuned**



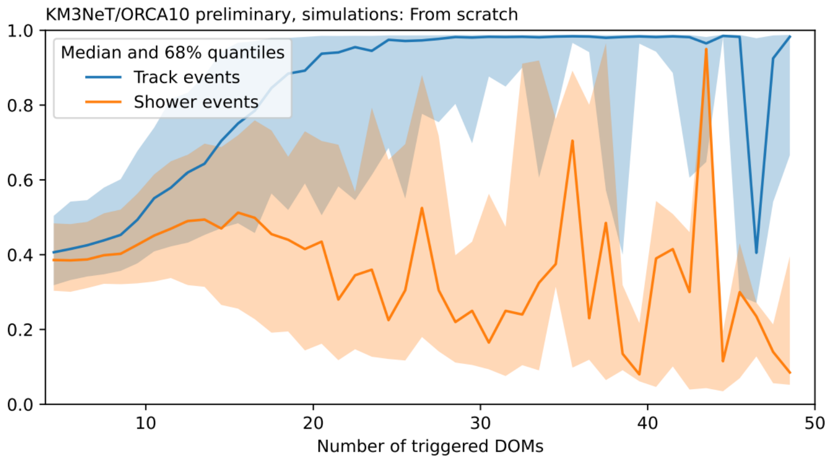

**Track-shower classification:**

- Limited data: 200k events for a detector with 6 lines is not enough to do a proper separation
- Performance: fine-tuned model works way better than the scratch one
- High dependence on event geometry: not enough discrimination with few lines
- **Major improvement with increasing detector size ↔ better event containment**

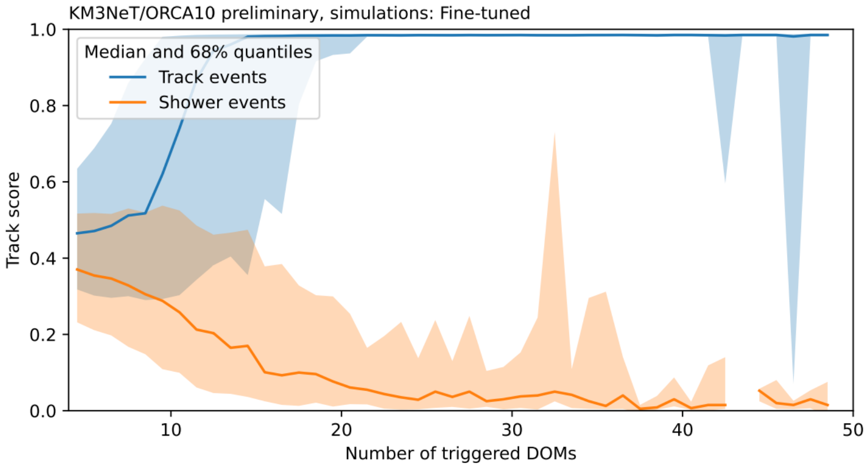
# Preliminary results on multi-geometries for KM3NeT/ORCA10

The model is able to **interpolate** to **non-existing DUs information** because it pre-learned the full geometry.

From scratch




Fine-tuned



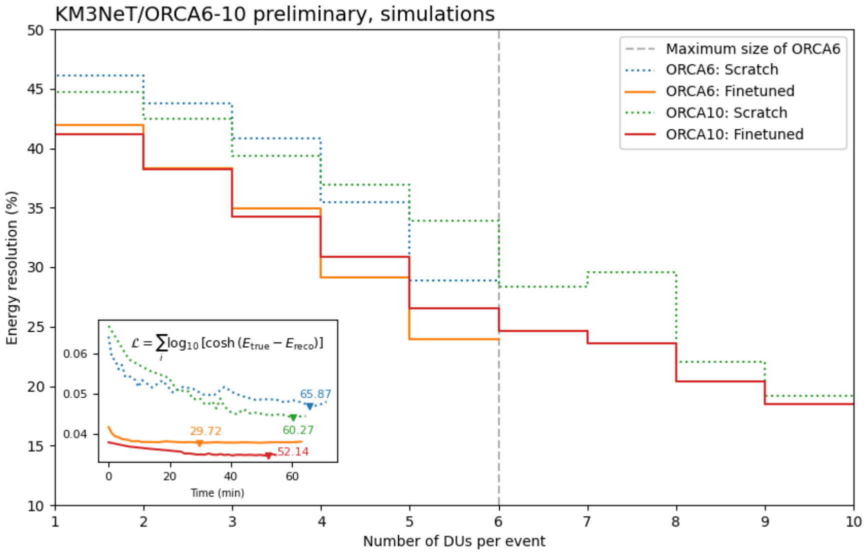

**Track-shower classification:**

- Limited data: 200k events for a detector with 6 lines is not enough to do a proper separation
- Performance: fine-tuned achieves separation in events with above 10 triggered DOMs
- Multiple peaks: fine-tuned model compensates the low statistics

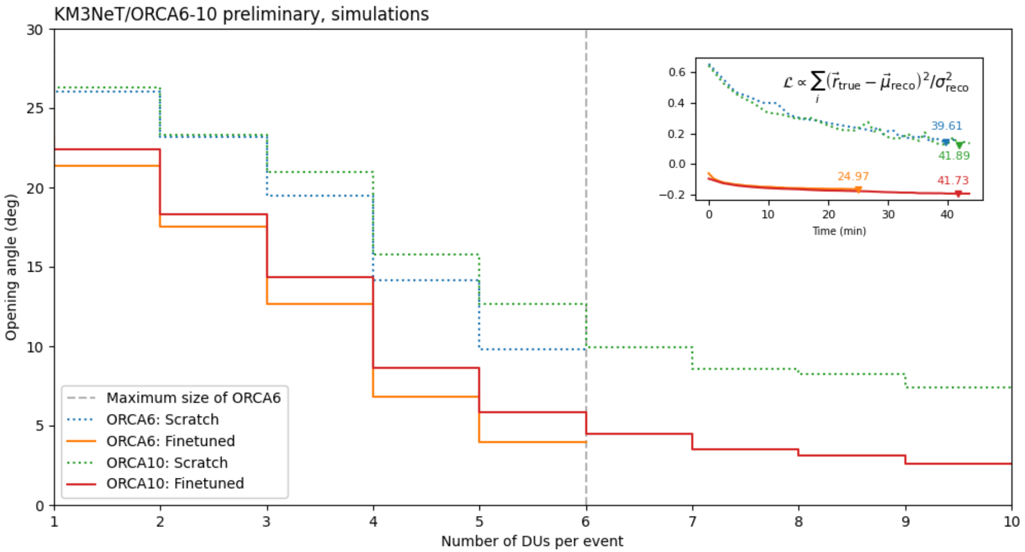
# Preliminary results on multi-geometries for KM3NeT/ORCA6-10

The model is able to **interpolate** to **non-existing DUs information** because it pre-learned the full geometry.

Energy reconstruction




Direction reconstruction




**Energy and direction reconstruction:**

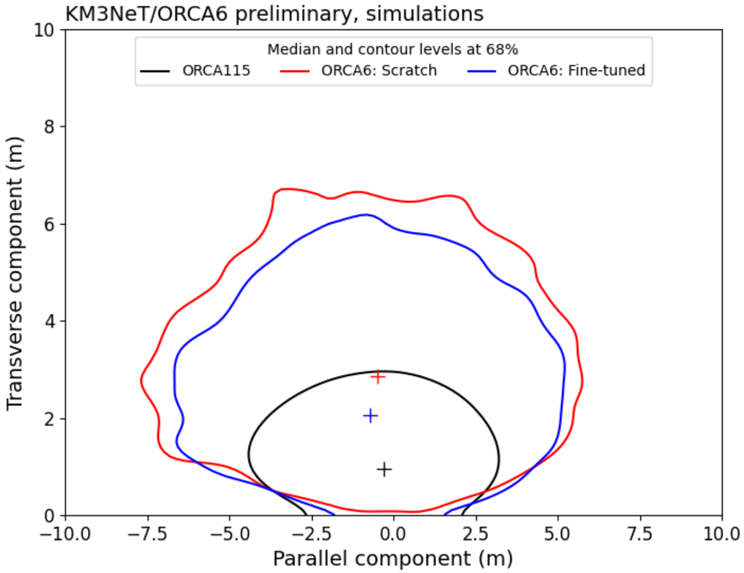
- Loss curves reveal fine-tuning's performance boost
- Similar resolution in energy reconstruction, but in less time!
- Direction reconstruction improved significantly

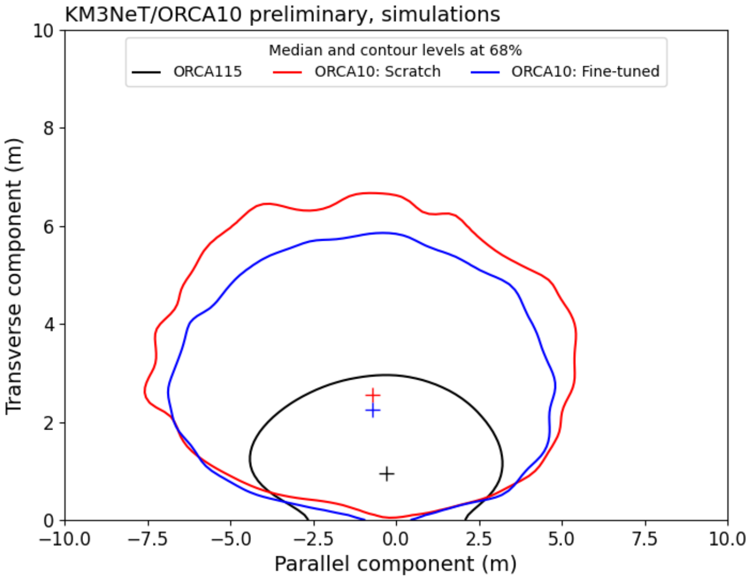
# Preliminary results on multi-geometries for KM3NeT/ORCA

**Interaction reconstruction vertex:**

- The hardest task
- Dynamic detector coordinates from RbR approach
- Small fiducial volume burdens the reconstruction

**Better data representation?**
Build a latent space to effectively accommodate dynamic coordinates and detector conditions & geometries




Interaction vertex reconstruction at KM3NeT/ORCA6 (left) and KM3NeT/ORCA10 (right) projected over the neutrino direction for 1-100 GeV atmospheric neutrinos.

# Preliminary results on multi-geometries for KM3NeT/ORCA

**Interaction reconstruction vertex:**

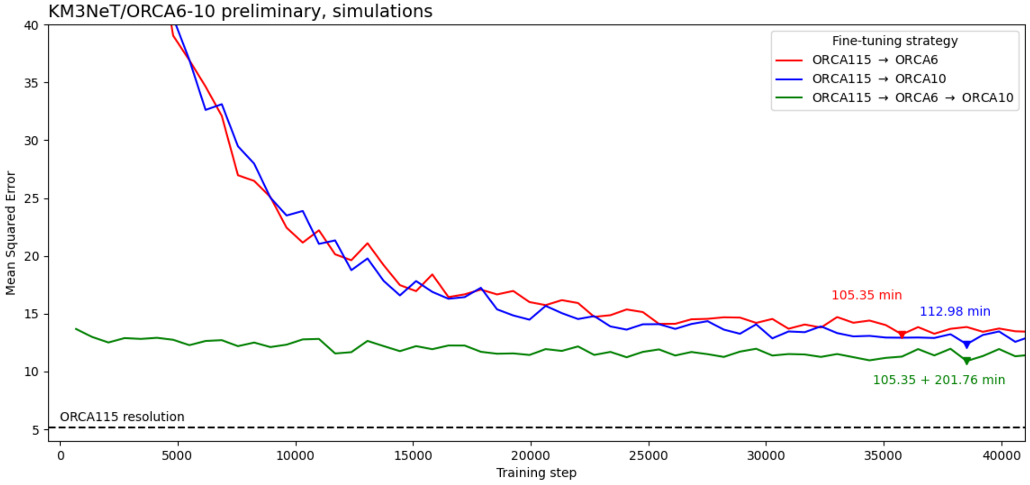
- The hardest task
- Dynamic detector coordinates from rbr approach
- Small fiducial volume burdens the reconstruction

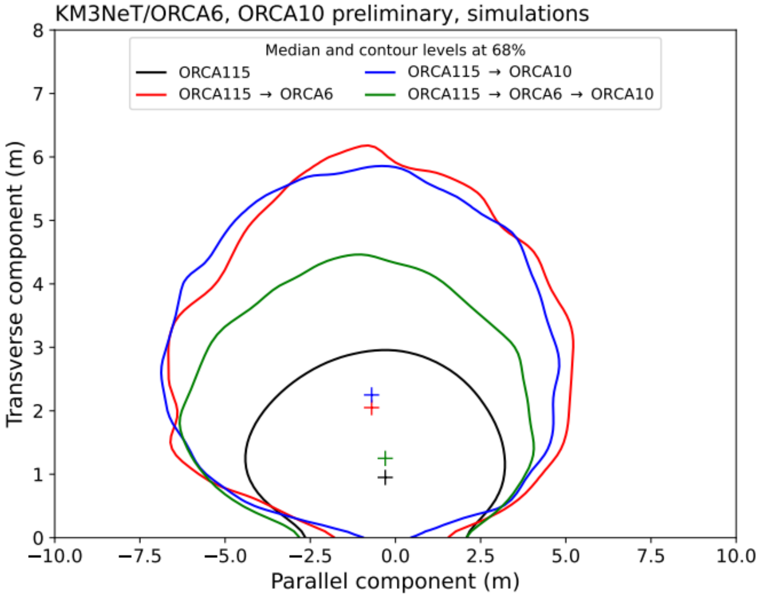
**From big to small...**

ORCA115 ⟶ ORCA6, ORCA10, etc.

**...and vice versa**

ORCA115 ⟶ ORCA6 ⟶ ORCA10 ⟶ etc.



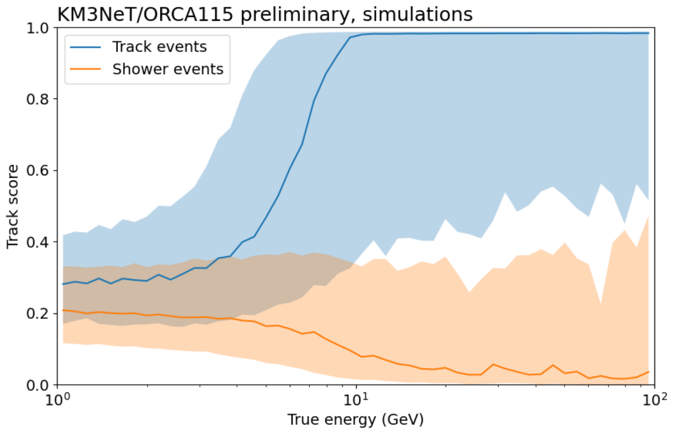

**We need to propagate the knowledge between detectors!**

# Preliminary results on multi-task for KM3NeT/ORCA

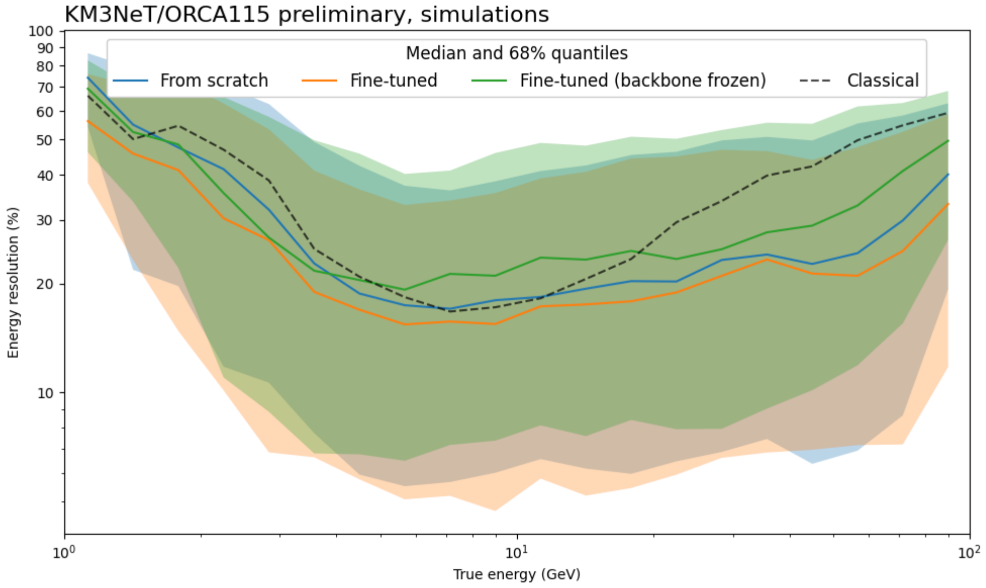
**Classification to energy reconstruction**

- A model that can handle multiple tasks
- A dataset 1 unrelated to dataset 2 (different detectors or water properties or atmospheric muons), helps a model into performing another task and makes it more robust

**Multi-task study**
ORCA115 dataset 1: track-shower
ORCA115 dataset 2: energy
850k tracks & 850k showers each




Track-score in function of neutrino energy in KM3NeT/ORCA115 for **dataset 1**




Energy resolution in function of neutrino energy in KM3NeT/ORCA115 for tracks from **dataset 2**

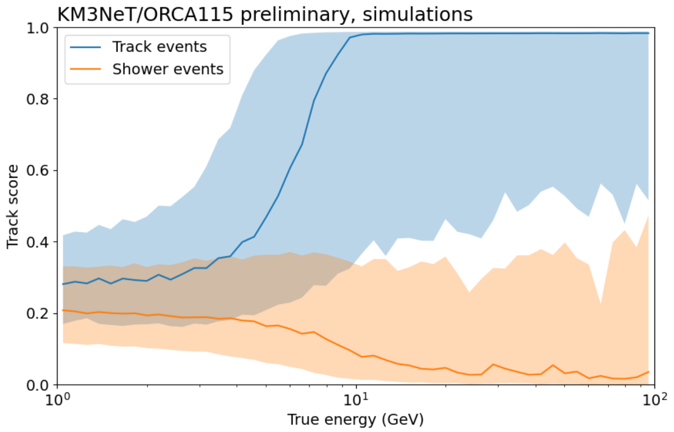
# Preliminary results on multi-task for KM3NeT/ORCA

AI SSAI
AI for science, science for AI

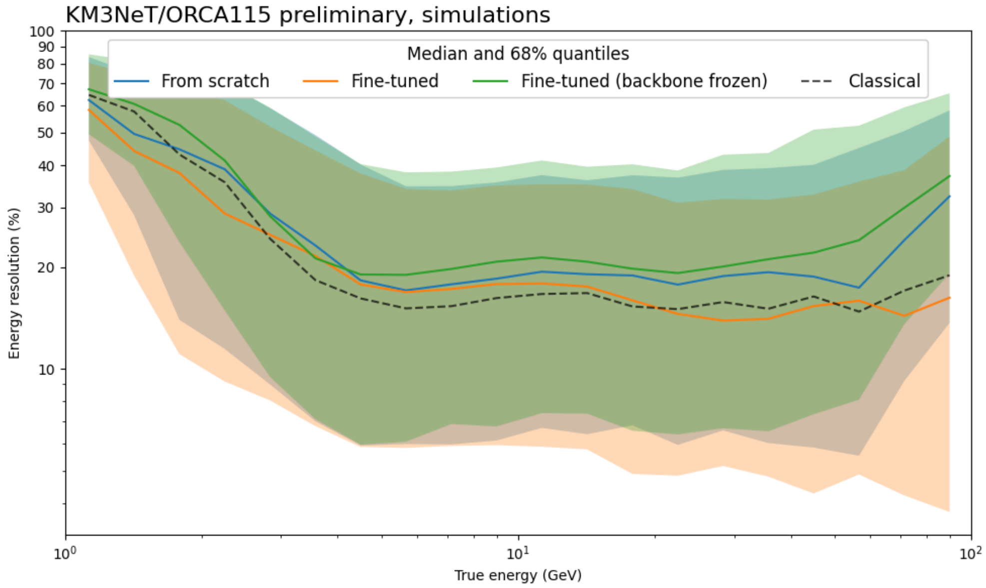
**Classification to energy reconstruction**

- A model that can handle multiple tasks
- A dataset 1 unrelated to dataset 2 (different detectors or water properties or atmospheric muons), helps a model into performing another task and makes it more robust

**Multi-task study**
ORCA115 dataset 1: track-shower
ORCA115 dataset 2: energy
850k tracks & 850k showers each




Track-score in function of neutrino energy in KM3NeT/ORCA115 for **dataset 1**



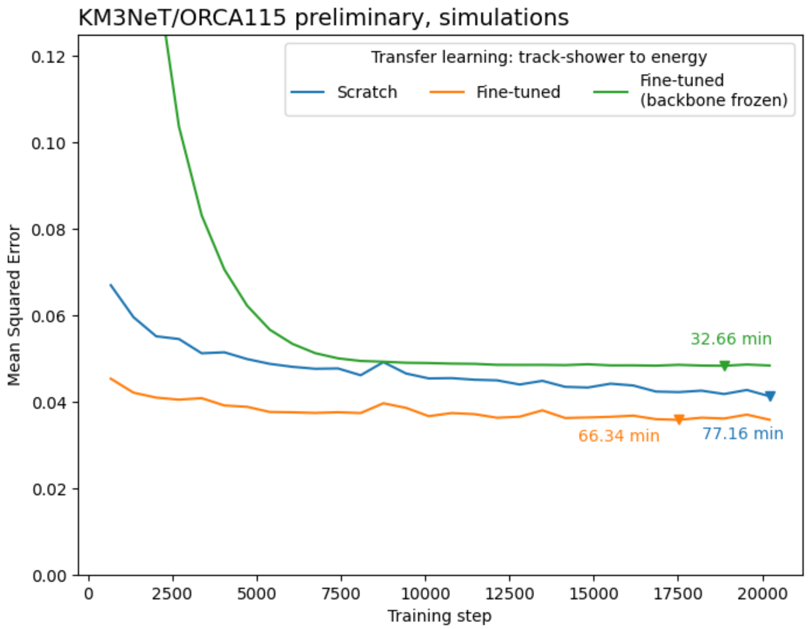

Energy resolution in function of neutrino energy in KM3NeT/ORCA115 for showers from **dataset 2**

# Preliminary results on multi-task for KM3NeT/ORCA

**Classification to energy reconstruction**

- A model that can handle multiple tasks
- A dataset 1 unrelated to dataset 2 (different detectors or water properties), helps a model into performing another task and makes it more robust

**Multi-task study**
ORCA115 dataset 1: track-shower
ORCA115 dataset 2: energy
850k tracks & 850k showers each




Fine-tuned model shows **faster convergence and efficiency**.

Freezing the backbone has a **trade-off** between training speed-up and accuracy $\longrightarrow$ suboptimal feature representation

Overall, the three cases achieve show improvements with respect to classical reconstruction methods.

# Summary

### Transfer Learning in multiple-detectors

- Transformers are particularly effective to deal with small detectors and very limited data
- Further optimization is needed in vertex reconstruction

### Transfer Learning for multi-task

- Speeds up training and boosts model robustness
- Leverages knowledge from different tasks

### The road ahead

- **From simulations to data:** ensure consistency and accuracy when transitioning to real detector data
- **Robustness tests & uncertainties:** validate model reliability across different conditions and detectors
- Estimate improvements as the detector grows to optimize **scalability**
- Develop common benchmark with state-of-the-art models
- Implement any deep learning reconstruction in the official data processing pipeline
- Start testing generative models with neutrino telescope data

# Preliminary results on KM3NeT/ORCA115

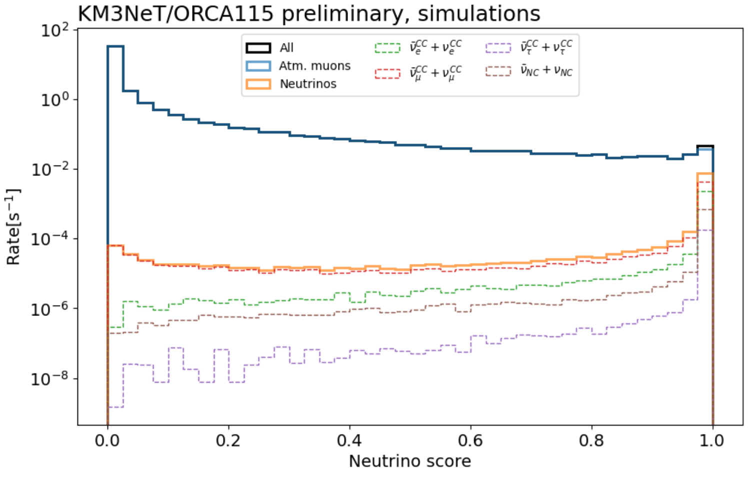
**Motivation:** the transformer is a language model

- KM3NeT/ORCA115 is the final detector, having all the possible neutrino physics encapsulated
- We can think of other configurations as similar languages to learn
- The information about KM3NeT/ORCA115 is used to understand our current detector

DL classification
Neutrino vs Background

DL classification
Track-like vs Shower-like

Track and shower
reconstruction




Event rate for neutrino score (0 for atmospheric muons, 1 for neutrinos).

**Purpose:** reject background data (atm. muon) from neutrino signal.

Atmospheric muons are more energetic, having their starting & ending points in most of the cases, out of the fiducial volume.

The model easily isolates **neutrino events** as they are mostly **fully contained** in the detector.

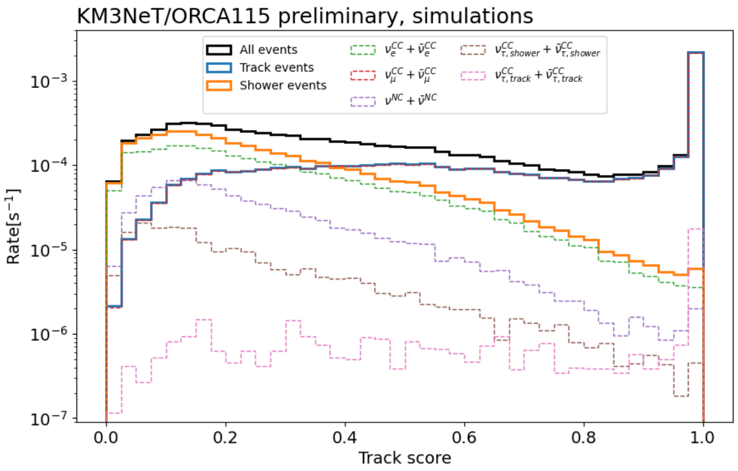
# Preliminary results on KM3NeT/ORCA115

**Motivation:** the transformer is a language model
- KM3NeT/ORCA115 is the final detector, having all the possible neutrino physics encapsulated
- We can think of other configurations as similar languages to learn
- The information about KM3NeT/ORCA115 is used to understand our current detector

DL classification
Neutrino vs Background

**DL classification**
**Track-like vs Shower-like**

Track and shower
reconstruction




Event rate for track score (0 for showers, 1 for tracks) for 1-100 GeV atmospheric neutrinos.

**Purpose:** separate the two neutrino event topologies, track-like and shower-like.

Enough separation power below 10 GeV (AUROC = 0.82)

High separation power above 10 GeV (AUROC = 0.91).

Low energy events do not contain enough pulses to properly separate these two categories

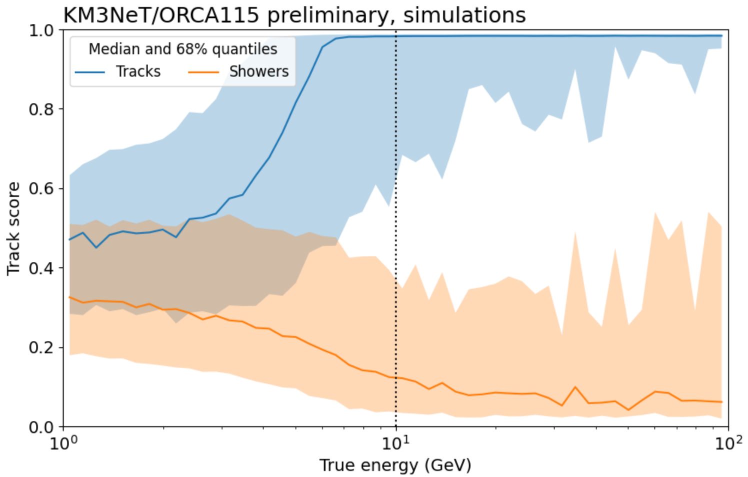
# Preliminary results on KM3NeT/ORCA115

**Motivation:** the transformer is a language model
- KM3NeT/ORCA115 is the final detector, having all the possible neutrino physics encapsulated
- We can think of other configurations as similar languages to learn
- The information about KM3NeT/ORCA115 is used to understand our current detector

DL classification
Neutrino vs Background

**DL classification**
**Track-like vs Shower-like**

Track and shower
reconstruction




Track score (0 for showers, 1 for tracks) as funtion of neutrino energy for 1-100 GeV atmospheric neutrinos.

**Purpose:** separate the two neutrino event topologies, track-like and shower-like.

Enough separation power below 10 GeV (AUROC = 0.82)

High separation power above 10 GeV (AUROC = 0.91).

**Low energy events do not contain enough pulses to properly separate these two categories**

# Preliminary results on KM3NeT/ORCA115

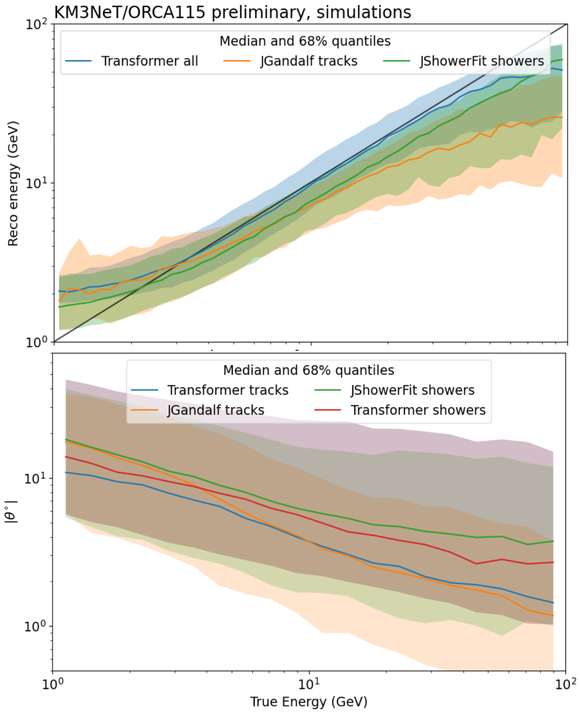
**Motivation:** the transformer is a language model

- KM3NeT/ORCA115 is the final detector, having all the possible neutrino physics encapsulated
- We can think of other configurations as similar languages to learn
- The information about KM3NeT/ORCA115 is used to understand our current detector

DL classification
Neutrino vs Background

DL classification
Track-like vs Shower-like

**Track and shower reconstruction**




True and reconstructed neutrino energy and zenith angle
Resolution for 1-100 GeV atmospheric neutrinos.

**Purpose:** reconstruct neutrino energy and neutrino direction.

Reconstruction done simultaneously for both track-like and shower-like events.

Saturation at high energies due to event containment.

Underestimation at low energies due to limited number of pulses.