Machine Learning in IceCube

A3D3 Seminar / 12 September 2022

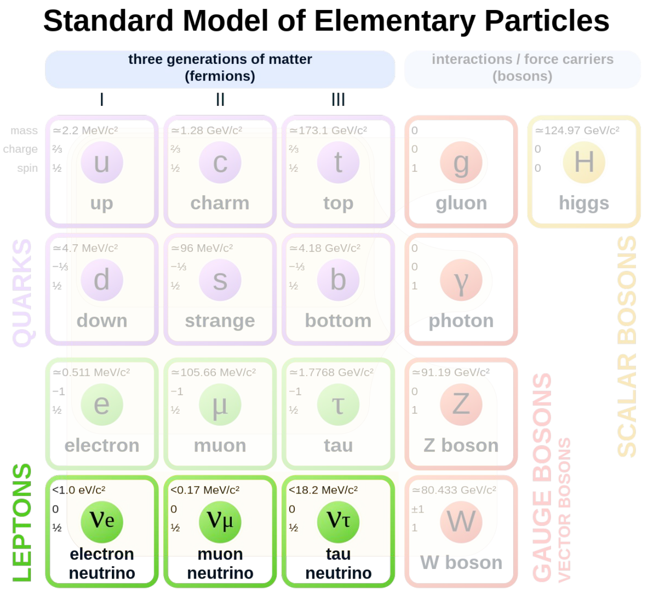
Andreas Søgaard

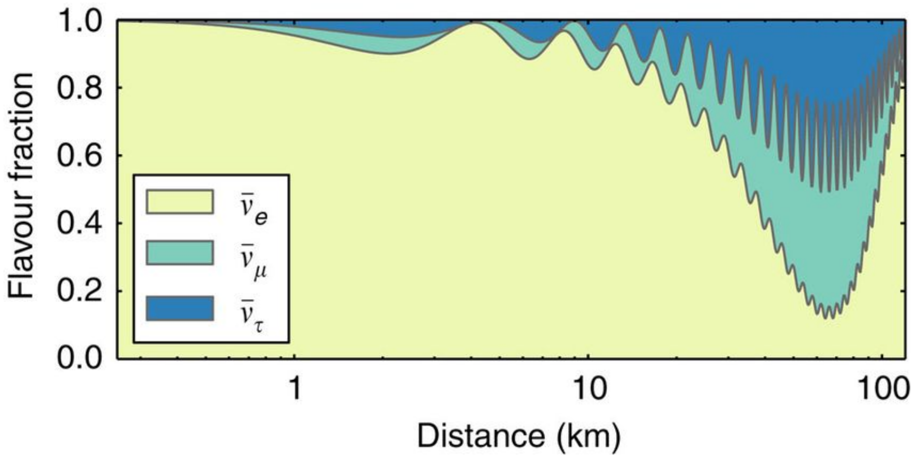
Niels Bohr Institute, University of Copenhagen

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 890778.

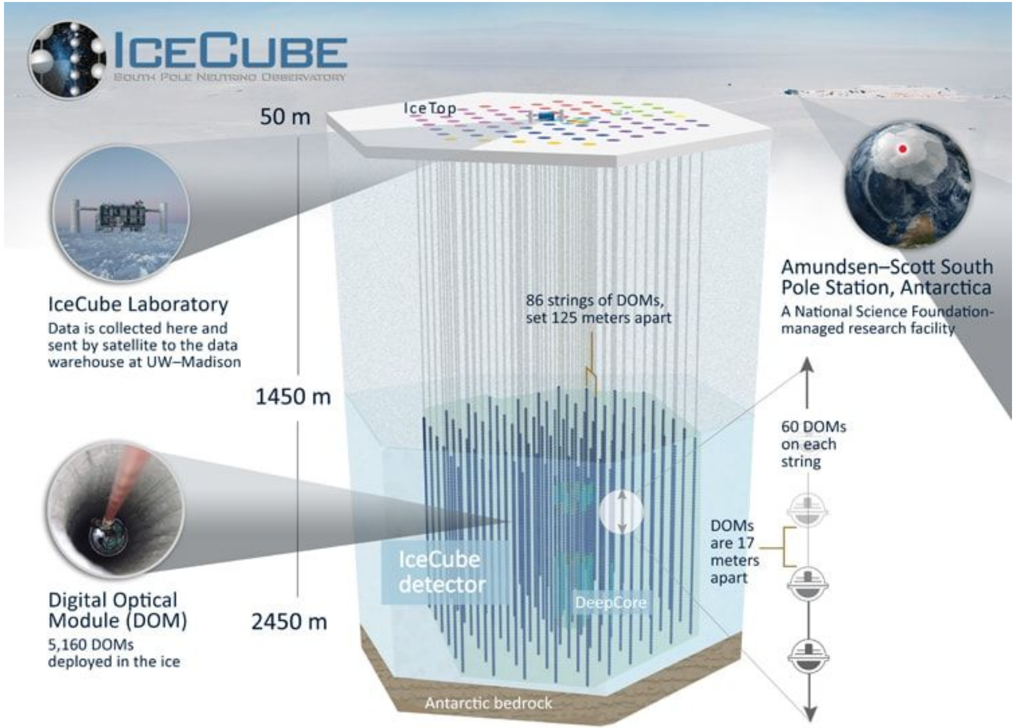
1. Neutrino physics and IceCube

Physics motivation

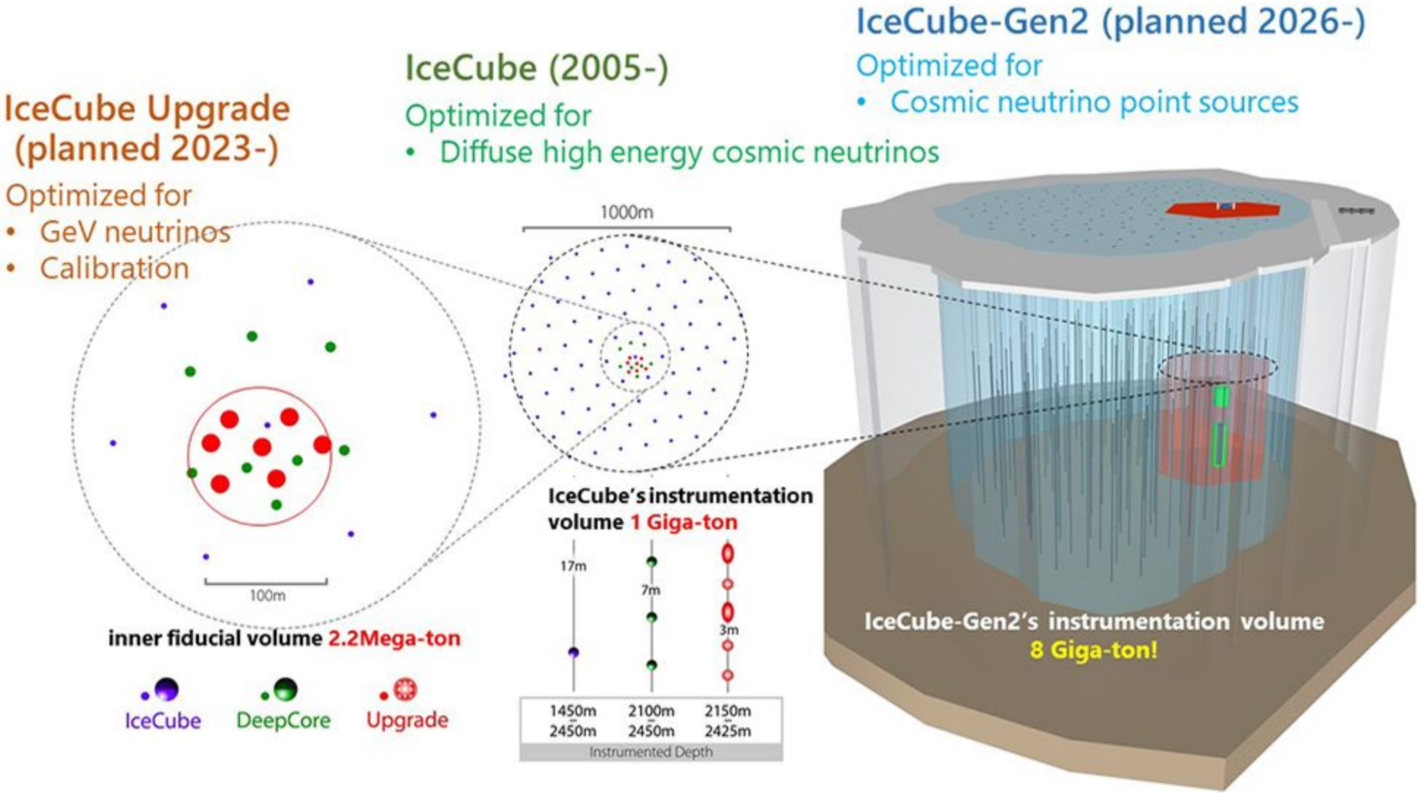




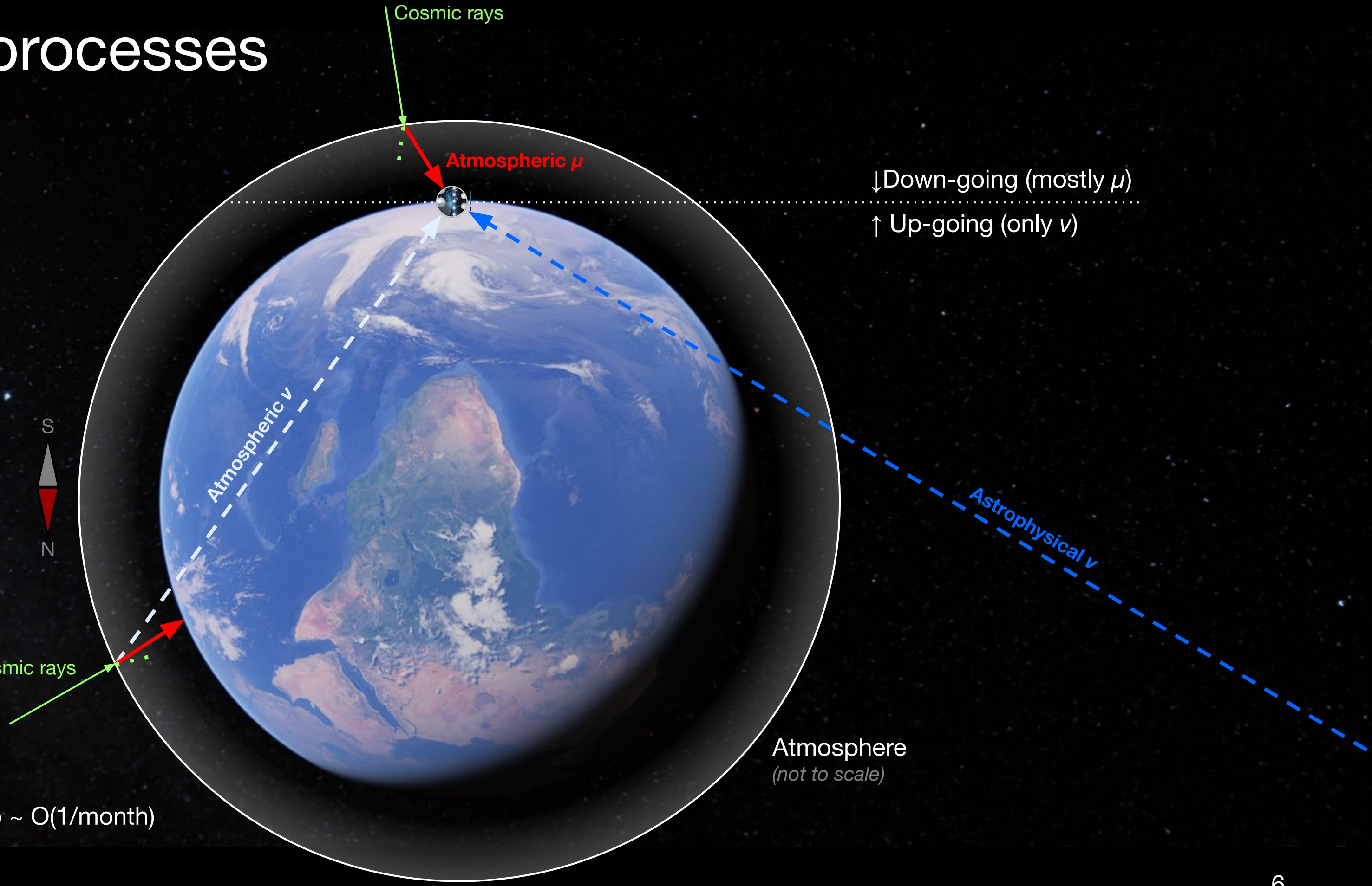
IceCube detector



Beyond IceCube



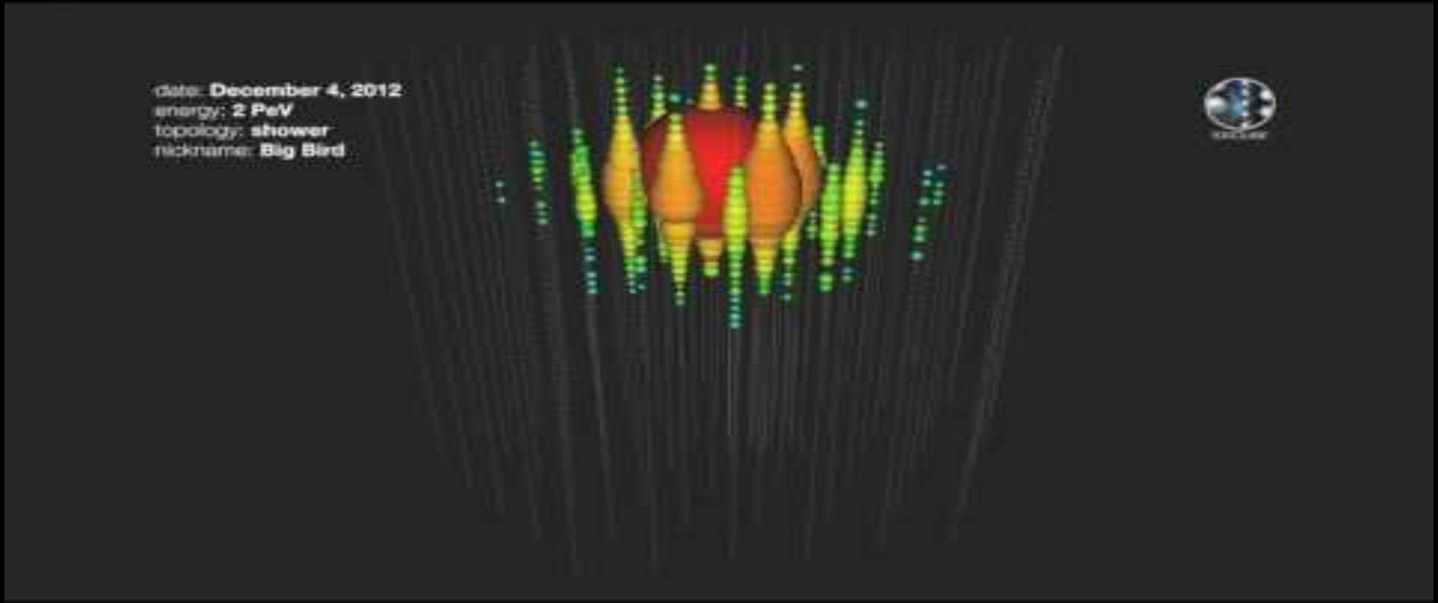
Generating processes



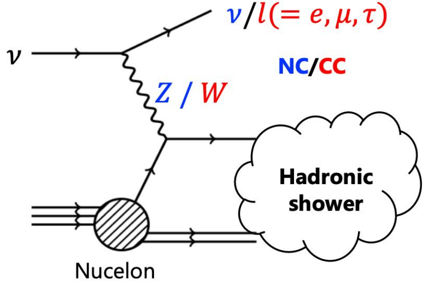
Rates

- Atmospheric μ ~ O(10^3 Hz)
- Atmospheric ν ~ O(10^{-3} Hz)
- Astrophysical ν ~ O(10^{-7} Hz) ~ O(1/month)

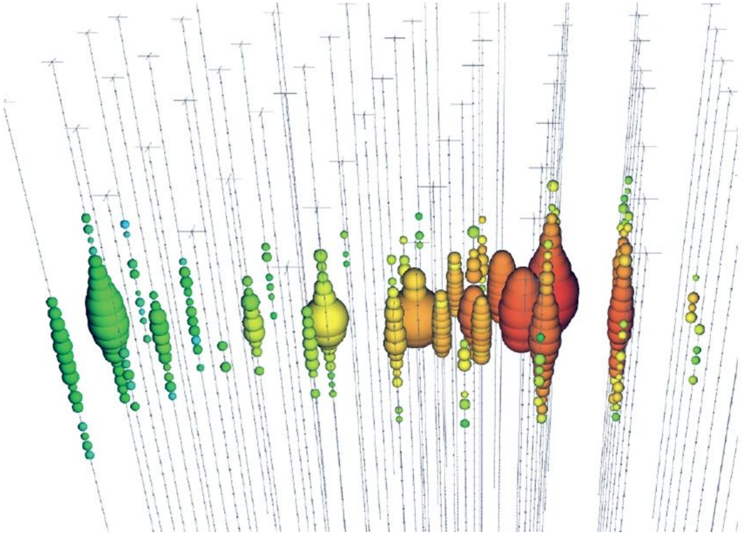
IceCube events



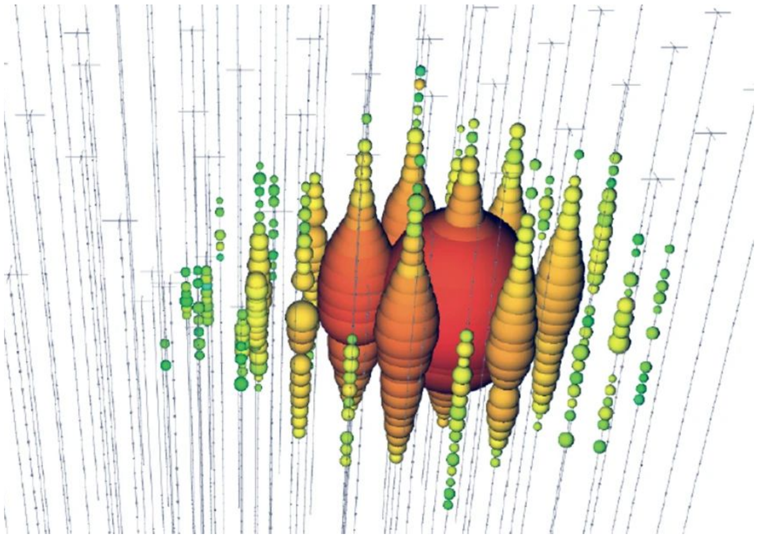
Event topologies



"Track-like"

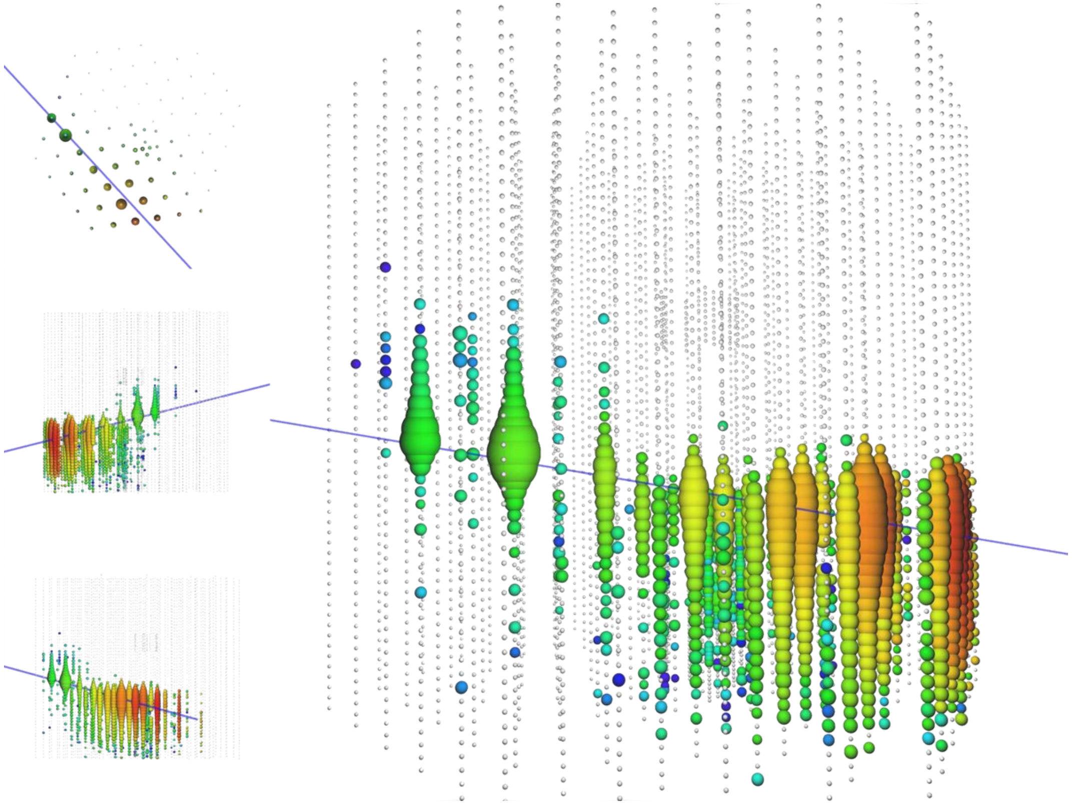


"Cascade-like"



IceCube's data is unique

- High rates, O(kHz) × O(10 years)
- Often sparse charge deposits
- Non-trivial geometry
- Geometrically heterogeneous, nested sub-detectors
- Similar detection method in all sub-detectors



Classical approaches to reconstruction in IceCube

RETRO — reverse table reconstruction [2203.02303]

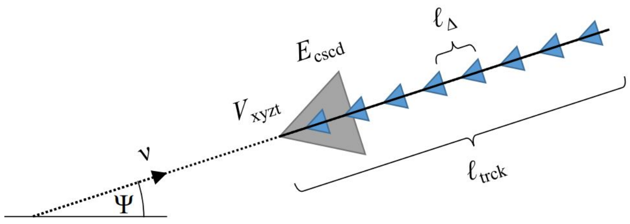
Sum of explicit per-DOM likelihood:

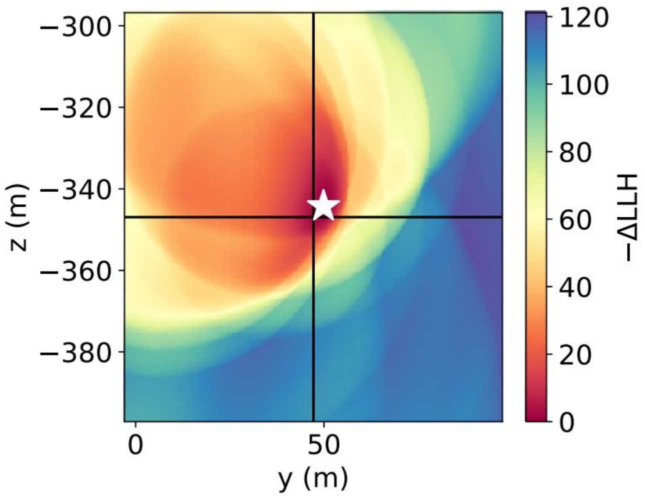
$$\log \mathcal{L}(\vec{\theta}) = \sum_i \log P(t_i|\vec{\theta}) - \Lambda(\vec{\theta}) - N$$

for finding pulses at times t_i given 8-dimensional parameter set θ comprising interaction vertex, pointing (zenith and azimuth), and “track” resp. “cascade” energy depositions.

Optimised directly for each event (2D slice →)

Inference speed ≈ O(40 sec./event)





2. Machine Learning in IceCube

Machine learning in IceCube is maturing

Approx. 50 literature results covering:

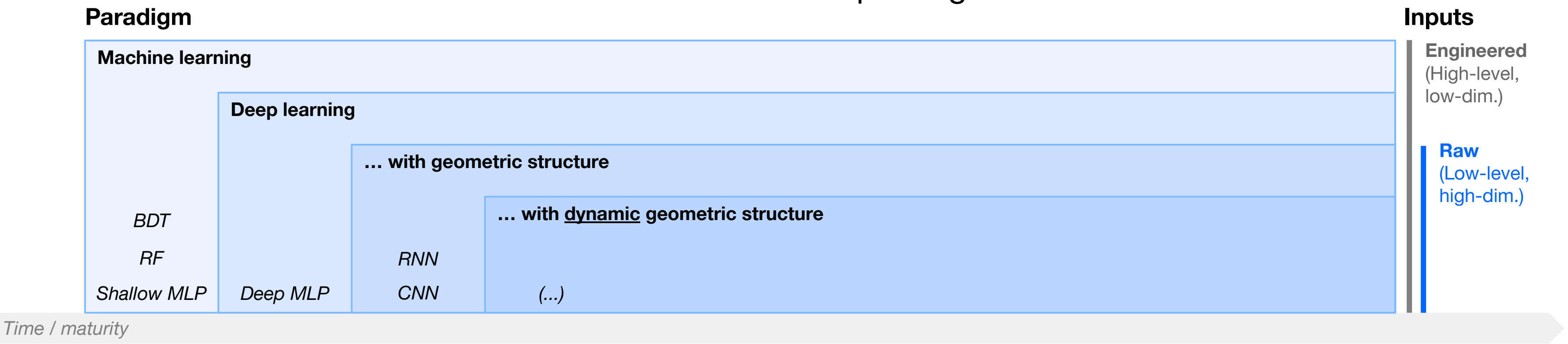
- Boosted decision trees [1705.08103]
- Random forests [2006.05215]
- Deep neural networks [1906.04317]
- Convolutional neural networks [2101.11589]
- *Etc.*

\+ Many not-yet-public efforts

Arc of ML in physics (not just IceCube):

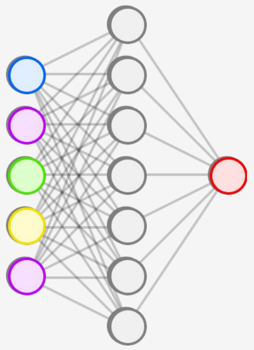
From Simple, analysis-level models on high-level features
Towards Complex, multi-purpose models on low-level features

→ Focus on **neural networks** as a highly flexible ML paradigm



Common types of neural networks

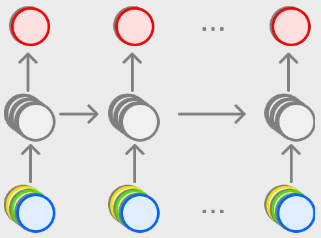
(Deep) Neural network — (D)NN
Occasionally: Multi-layer perceptron (MLP)



Structure
No inherent geometric structure

Example
Engineered, high-level event features

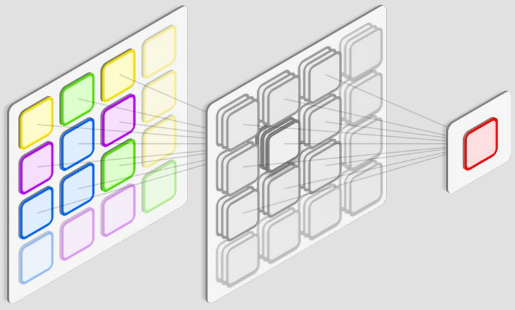
Recurrent neural network (RNN)



Structure
Sequential

Example
Time-series

Convolutional neural network (CNN)



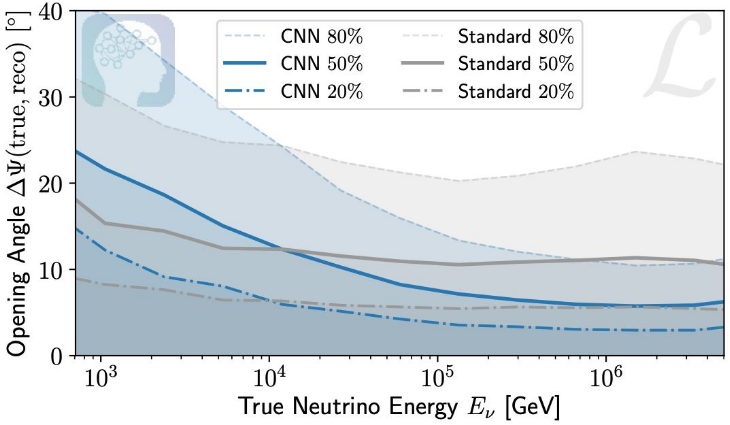
Structure
Orthogonal data, translation invariance

Example
Images

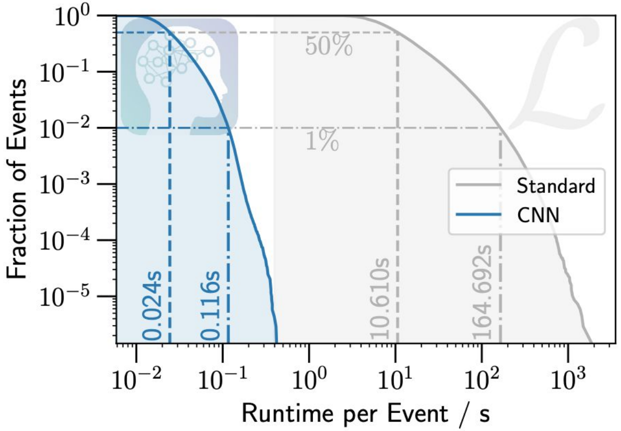
Machine Learning shows vast potential in IceCube

For instance, high-energy cascade reconstruction using **CNNs**:

50% improvement in resolution at high energies



2-3 orders of magnitude reduction in reconstruction time[1]



[1] CNN run on GPU.

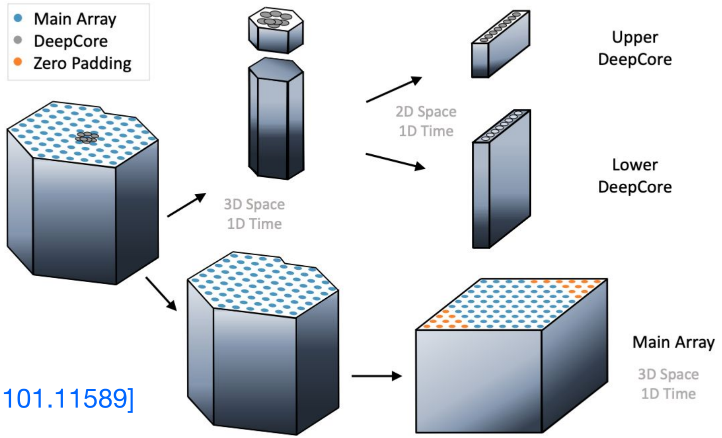
But CNNs face conceptual challenges

Requirement

- Splitting up sub-detectors
- Distorting geometry
- Pulses must be “summarised” to DOM-level
- Highly specialised kernels

Risk

- ❌ Weakening local correlations
- ❌ Losing geometric details
- ❌ Losing granular information
- ❌ Reducing generalisation potential



Unifying "zoo" of architectures

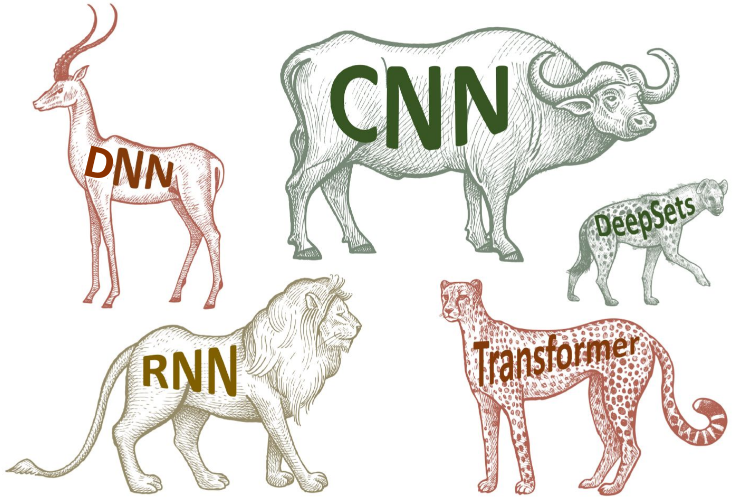
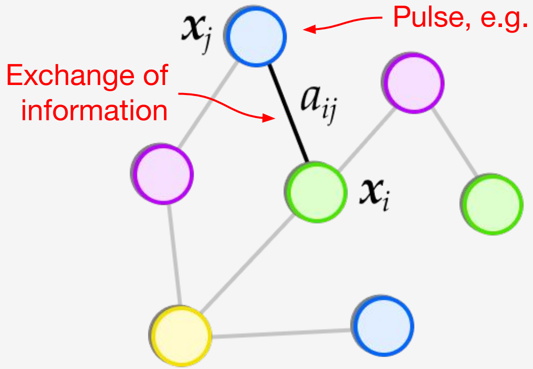


Image adapted from:
Bronstein, Geometric foundations of deep learning (2021)

Graph Neural network — GNN

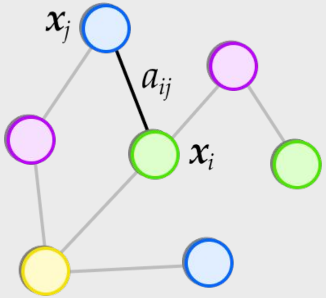
Representation
Data $\boldsymbol{x}_i$ on nodes a graph; nodes connected by edges $\boldsymbol{a}_{ij}$

Structure
Any that can be encoded through *adjacency* of nodes



Most neural network architectures can be seen as special cases of the GNN with added structure
(Bronstein et al., 2021)

Why GNNs are a natural choice in IceCube



Challenge

- ⛰️ Sparse charge deposits
- ⛰️ Non-trivial geometry
- ⛰️ Nested sub-detectors

Addressed

- ✅ Only ingest hit PMTs in each event
- ✅ No requirements on structure
- ✅ Information transfer among all nodes

Plus:

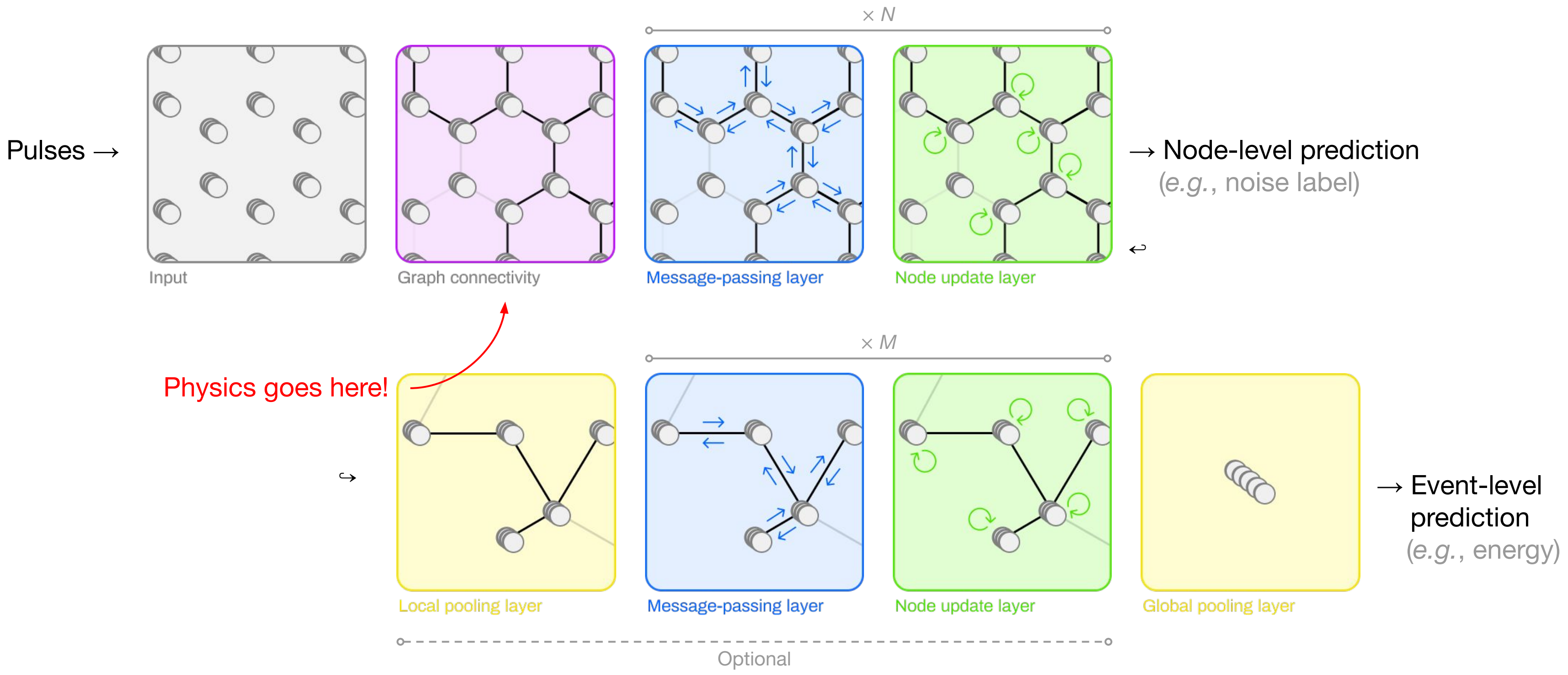
- ✅ No need to "summarise" pulses to DOM-level
- ✅ No need for specialised kernels
- ✅ Ability to *encode* physics, material properties, etc. into structure, e.g.,

$$a_{ij} \sim \frac{\mathrm{sign}(\Delta t_{ij})}{1 + \| x_i^{\mu} - x_j^{\mu} \|^2}$$

… with the regular DL benefit:

- ✅ Fast inference

Anatomy of a graph neural network



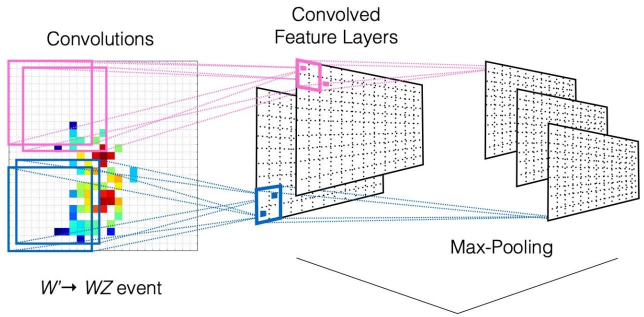
But…

… are graphs then always
the right solution?

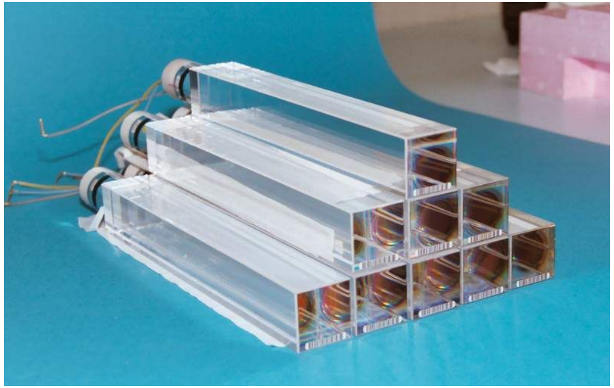
Detour: General-purpose exps. at CERN

Electromagnetic and hadronic calorimeters *are* mostly "pixelated" in azimuth and pseudorapidity with a few depth layers (≈ colour channels)

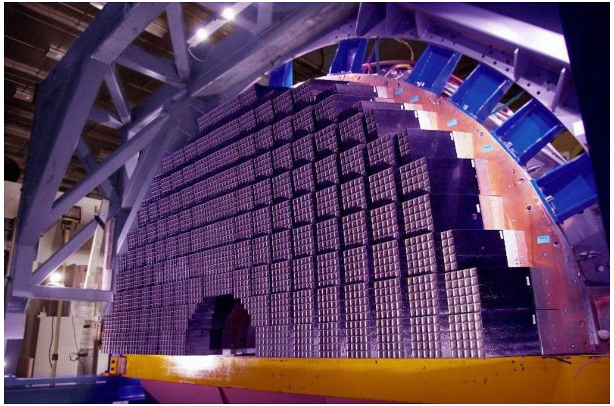
→ **CNNs** are a very reasonable and effective paradigm (~x2 in background rejection vs. simple combinations of analytically calculated jet substructure moments)



Source: [1511.05190]



Source: CMS ECAL Endcaps photographs

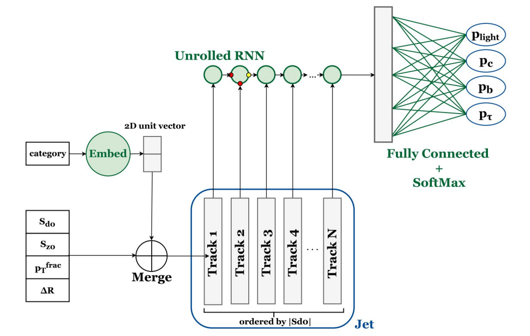


Source: CMS ECAL Endcaps photographs

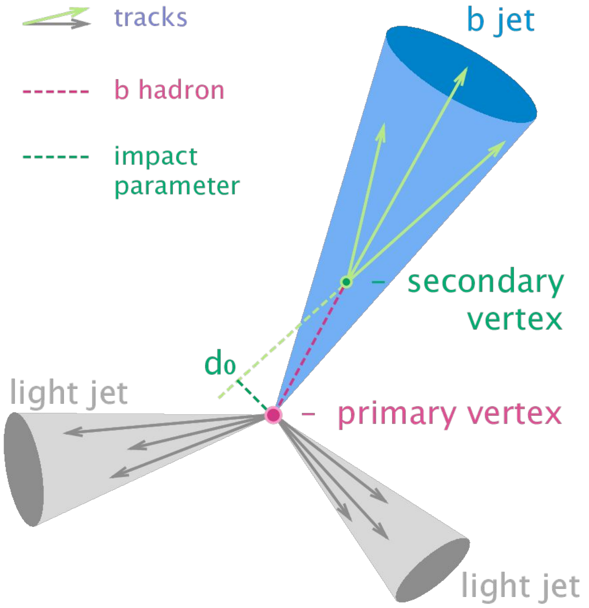
Detour: General-purpose exps. at CERN

"RNN based b-tagging algorithm can exploit the spatial and kinematic correlations between tracks which are initiated from the same b-hadron" (ATL-PHYS-PUB-2017-003) which alternative methods don't easily allow for, and the notion of an impact parameter provides a natural track ordering for the task.

→ **RNNs** are a very reasonable and effective paradigm (~x2.5 in light-jet background rejection vs. comparable, analytical algorithm)



Source: ATL-PHYS-PUB-2017-003



Source: Wikimedia / Nazar Bartosik

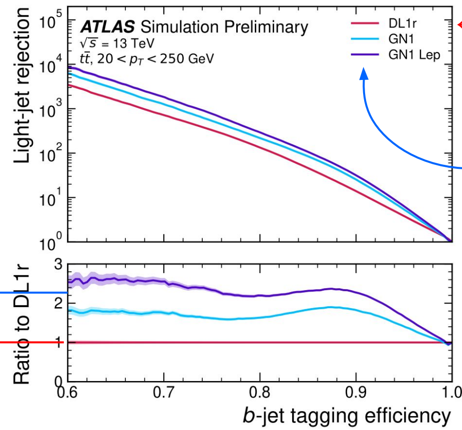
Still, GNNs providing large performance gains…

Top-jet identification (CMS's ParticleNet, 2020)

			$1/\varepsilon_b$ at $\varepsilon_s = 30\%$
2D CNN →	ResNeXt-50		1147 ± 58
	P-CNN		759 ± 24
	PFN	(...)	888 ± 17
	ParticleNet-Lite		1262 ± 49
GNN →	ParticleNet		$\mathbf{1615 \pm 93}$

+40%

b-jet tagging (ATLAS's GN1, 2022)



← Comb. RNN
+ DNN
+ Expert vars.

Single
multi-objective GNN

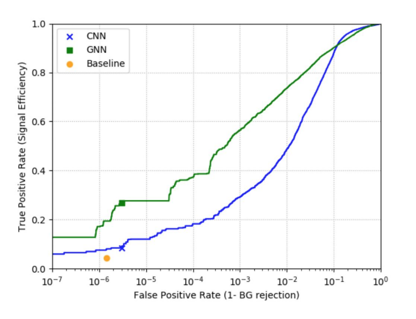
+2-300%

… plus tracking, secondary vertex finding, pile-up mitigation, and more report marked performance improvements from GNNs vs. other, seemingly well-motivated approaches.

Take-away: Despite other paradigms being seemingly good fits for the underlying structure of specific problems, GNNs can be considered a more fundamental paradigm that generalises the others.

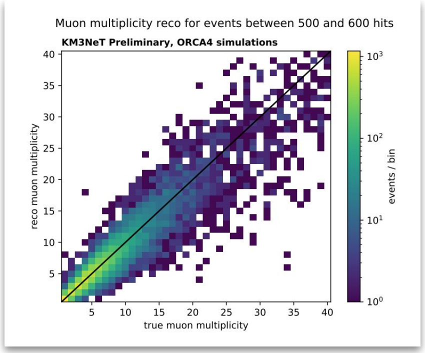
Graph neural networks in IceCube and beyond, so far

Choma et al., IceCube [1809.06166]



μ vs. ν_μ classification, out-performing CNN and line-fit

Reck et al., KM3NeT [2107.13375]

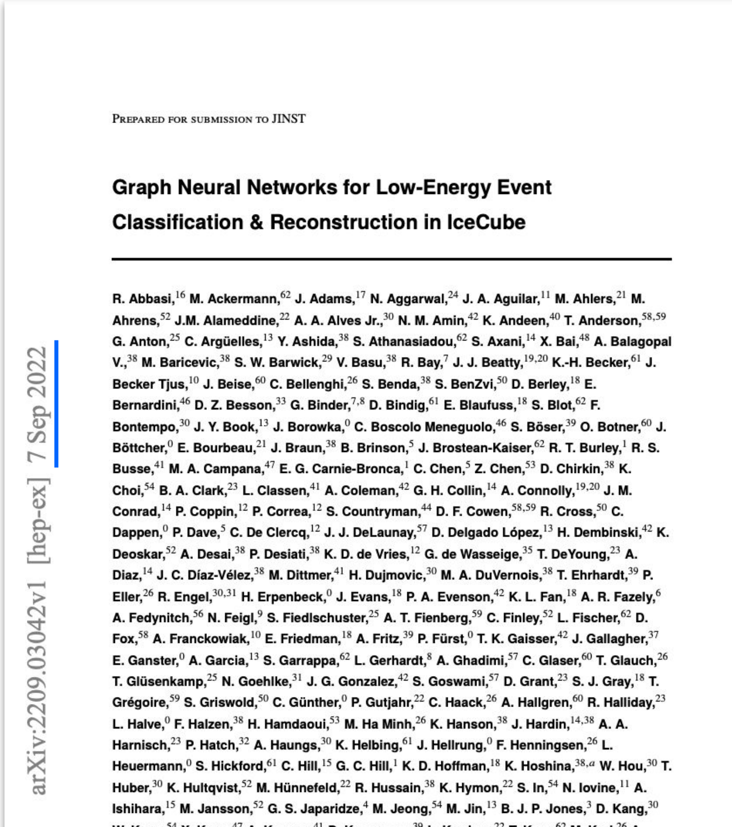


Muon bundles multiplicity and neutrino PID

🥁 … and … 🥁

Hot-off-the-press result 🔥

[2209.03042]



arXiv:2209.03042v1 [hep-ex] 7 Sep 2022

Prepared for submission to JINST

Graph Neural Networks for Low-Energy Event Classification & Reconstruction in IceCube

R. Abbasi,[16] M. Ackermann,[62] J. Adams,[17] N. Aggarwal,[24] J. A. Aguilar,[11] M. Ahlers,[21] M. Ahrens,[52] J.M. Alameddine,[22] A. A. Alves Jr.,[30] N. M. Amin,[42] K. Andeen,[40] T. Anderson,[58,59] G. Anton,[25] C. Argüelles,[13] Y. Ashida,[38] S. Athanasiadou,[62] S. Axani,[14] X. Bai,[48] A. Balagopal V.,[38] M. Baricevic,[38] S. W. Barwick,[29] V. Basu,[38] R. Bay,[7] J. J. Beatty,[19,20] K.-H. Becker,[61] J. Becker Tjus,[10] J. Beise,[60] C. Bellenghi,[26] S. Benda,[38] S. BenZvi,[50] D. Berley,[18] E. Bernardini,[46] D. Z. Besson,[33] G. Binder,[7,8] D. Bindig,[61] E. Blaufuss,[18] S. Blot,[62] F. Bontempo,[30] J. Y. Book,[13] J. Borowka,[0] C. Boscolo Meneguolo,[46] S. Böser,[39] O. Botner,[60] J. Böttcher,[0] E. Bourbeau,[21] J. Braun,[38] B. Brinson,[5] J. Brostean-Kaiser,[62] R. T. Burley,[1] R. S. Busse,[41] M. A. Campana,[47] E. G. Carnie-Bronca,[1] C. Chen,[5] Z. Chen,[53] D. Chirkin,[38] K. Choi,[54] B. A. Clark,[23] L. Classen,[41] A. Coleman,[42] G. H. Collin,[14] A. Connolly,[19,20] J. M. Conrad,[14] P. Coppin,[12] P. Correa,[12] S. Countryman,[44] D. F. Cowen,[58,59] R. Cross,[50] C. Dappen,[0] P. Dave,[5] C. De Clercq,[12] J. J. DeLaunay,[57] D. Delgado López,[13] H. Dembinski,[42] K. Deoskar,[52] A. Desai,[38] P. Desiati,[38] K. D. de Vries,[12] G. de Wasseige,[35] T. DeYoung,[23] A. Diaz,[14] J. C. Díaz-Vélez,[38] M. Dittmer,[41] H. Dujmovic,[30] M. A. DuVernois,[38] T. Ehrhardt,[39] P. Eller,[26] R. Engel,[30,31] H. Erpenbeck,[0] J. Evans,[18] P. A. Evenson,[42] K. L. Fan,[18] A. R. Fazely,[6] A. Fedynitch,[56] N. Feigl,[9] S. Fiedlschuster,[25] A. T. Fienberg,[59] C. Finley,[52] L. Fischer,[62] D. Fox,[58] A. Franckowiak,[10] E. Friedman,[18] A. Fritz,[39] P. Fürst,[0] T. K. Gaisser,[42] J. Gallagher,[37] E. Ganster,[0] A. Garcia,[13] S. Garrappa,[62] L. Gerhardt,[8] A. Ghadimi,[57] C. Glaser,[60] T. Glauch,[26] T. Glüsenkamp,[25] N. Goehlke,[31] J. G. Gonzalez,[42] S. Goswami,[57] D. Grant,[23] S. J. Gray,[18] T. Grégoire,[59] S. Griswold,[50] C. Günther,[0] P. Gutjahr,[22] C. Haack,[26] A. Hallgren,[60] R. Halliday,[23] L. Halve,[0] F. Halzen,[38] H. Hamdaoui,[53] M. Ha Minh,[26] K. Hanson,[38] J. Hardin,[14,38] A. A. Harnisch,[23] P. Hatch,[32] A. Haungs,[30] K. Helbing,[61] J. Hellrung,[0] F. Henningsen,[26] L. Heuermann,[0] S. Hickford,[61] C. Hill,[15] G. C. Hill,[1] K. D. Hoffman,[18] K. Hoshina,[38,a] W. Hou,[30] T. Huber,[30] K. Hultqvist,[52] M. Hünnefeld,[22] R. Hussain,[38] K. Hymon,[22] S. In,[54] N. Iovine,[11] A. Ishihara,[15] M. Jansson,[52] G. S. Japaridze,[4] M. Jeong,[54] M. Jin,[13] B. J. P. Jones,[3] D. Kang,[30]

GNNs for Low-Energy Event Classification & Reconstruction

Scope

Monte Carlo-based proof-of-concept paper on the use of GNNs in IceCube for a variety of physics tasks targeting the low-energy region (1 GeV – 1 TeV; most events in 50 – 200 GeV) → Potential to improve sensitivity to oscillation params.

Event selection

Leveraging existing 7-level event selection process designed for oscillation analyses, which was aimed at reducing the approx. 3 kHz trigger rate to approx. 1 mHz (i.e., > $\times 10^6$ reduction) required for the application of (slow) state-of-the-art reconstruction such as RETRO.

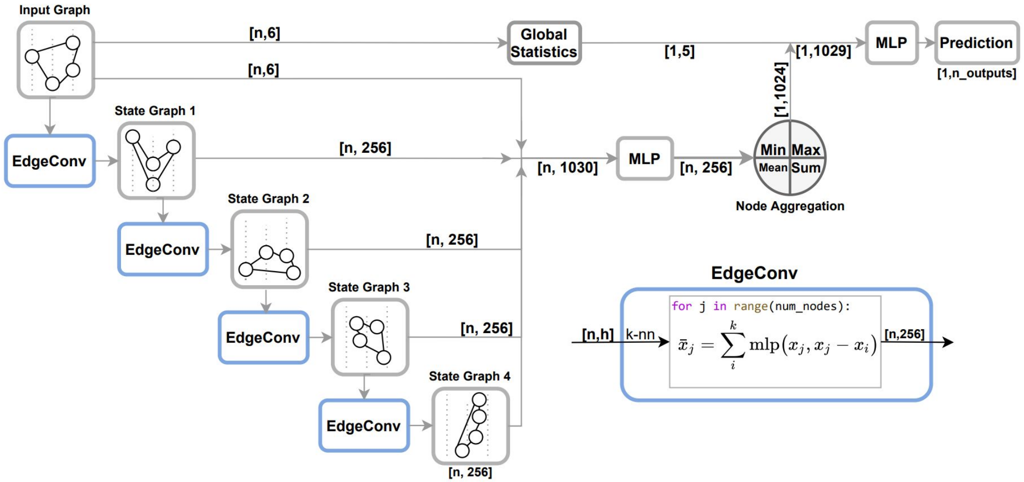
Implications

Performance gains from GNNs applied at the 7th level of a reconstruction chain based on simpler approaches (out of necessity) are **lower bounds on the impact of GNNs on physics analyses,** as their fast inference times allows for leveraging high-precision reconstruction much earlier in the reconstruction.

The "DynEdge" model

Anatomy:

- Connecting 8 nearest pulses
  - *xyz*-space for first convolution
  - Dynamical edge connections for each of four subsequent GNN layers ("DynEdge")
- Single-layer MLP-based "edge convolutions" operation
- Skip connection between convolutional layers
- Four global poolings + high-level features (homophily + num. pulses)
- Single-layer readout MLP
- All models single-task



Performance studies

1. Neutrino vs. muon classification

Neutrino sample purity

2. Track vs. cascade classification
3. Neutrino energy reconstruction
4. Neutrino zenith reconstruction

Oscillation analysis binning

5. Impact on oscillation contours

Oscillation analysis sensitivity

6. Robustness to systematic uncertainties

Robustness of ↑

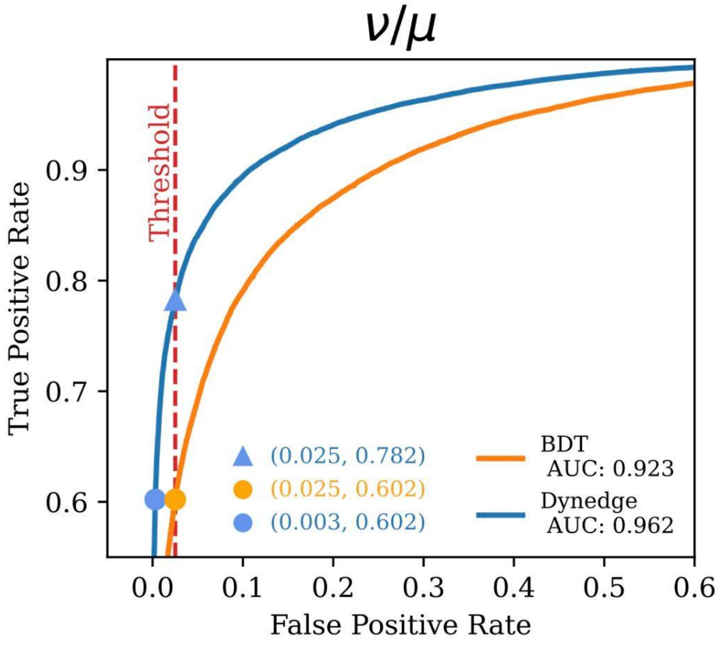
Neutrino vs. muon classification

Motivation
Atmospheric muons constitute the largest background. Identifying neutrinos among these are essential to a pure signal.

ML task
Binary classification of muon vs. neutrino events.

Result

- At fixed background rejection, DynEdge **improves signal efficiency by 18%**, vs. in-use BDT.
- At signal efficiency, DynEdge increases muon **background rejection by > ×8.**

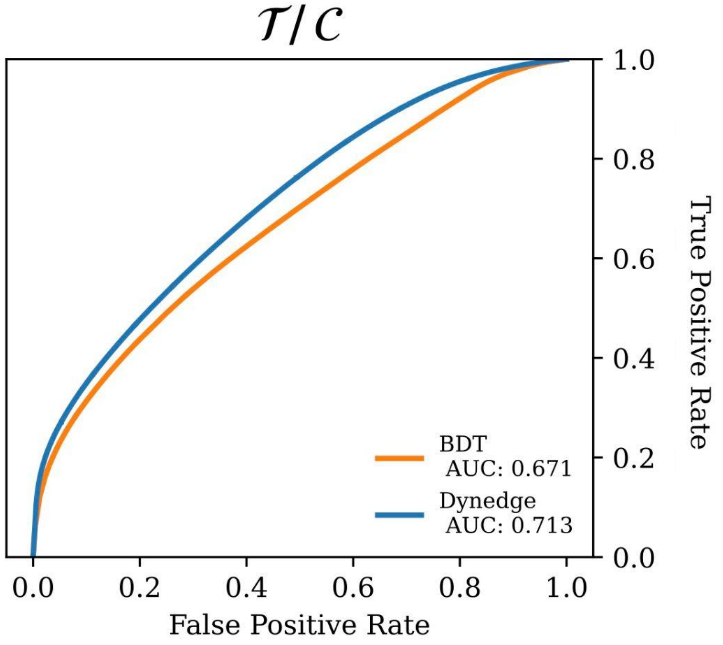


Tracks vs. cascade classification — 2/6

Motivation
IceCube is not sensitive to individual neutrino flavours, but uses “track” and “cascade”-like events and proxy categories. These are crucial for oscillation measurements, which rely on flavour identification to measure, e.g., muon neutrino disappearance.

ML task
Binary classification of track- vs cascade-like events.

Result
About **6% improvement** in ROC AUC, cf. in-use BDT, yielding significantly cleaner “PID” bins.

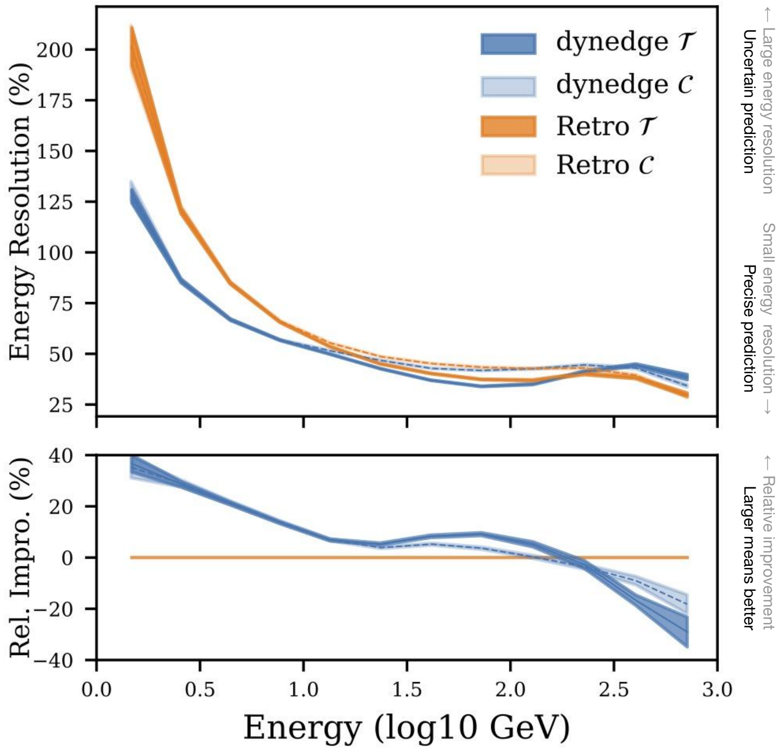


Neutrino energy reconstruction

Motivation
Oscillation analyses are binned in energy, meaning that improved energy resolutions leads to sharper oscillation measurements.

ML task
1D regression using log-cosh loss in $\log_{10}(E/\mathrm{GeV})$. Relative improvements vs. RETRO are measured according to the 68%–inter-percentile range of the residuals in E.

Result
Average **improvement in resolution of around 20%** in the energy range relevant for oscillation measurements
(1 GeV < E < 100 GeV, or $0 < \log_{10}(E/\mathrm{GeV}) < 2$)

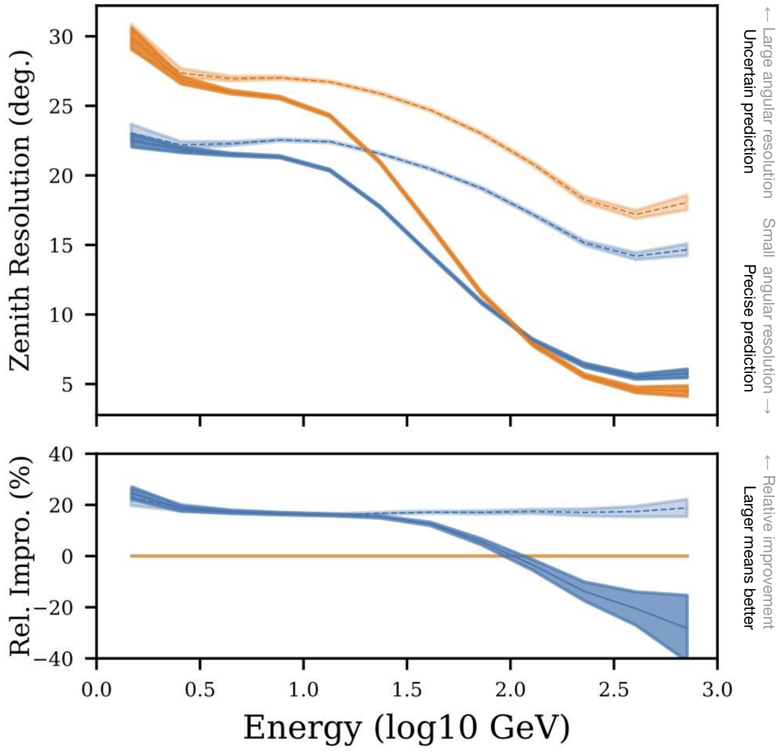


Neutrino zenith reconstruction

Motivation
Oscillation analyses are binned in distance, which for atmospheric neutrinos corresponds to the zenith angle (i.e., angle wrt. the horizon). This means that improved zenith resolutions leads to sharper oscillation measurements.

ML task
1D regression using von Mises-Fisher loss to quantify uncertainty through Gaussian approximation.

Result
Average **improvement in resolution of around 20%** in the energy range relevant for oscillation measurements (E < 100 GeV, or $\log_{10}(E/\mathrm{GeV}) < 2$)



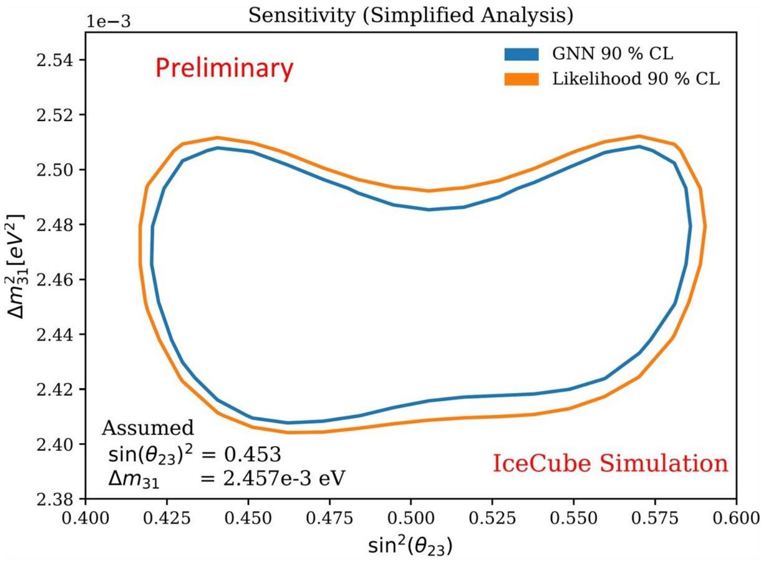
Impact on oscillation contours — 5/6

Background

IceCube uses an analysis software called PISA for oscillation analyses. By providing **energy**, **zenith** and **track vs. cascade** predictions, PISA can produce the corresponding oscillation contours. This can be used to produce simplified oscillation contour that ignore systematic uncertainties, thereby isolating the impact of different reconstruction algorithms.

Result

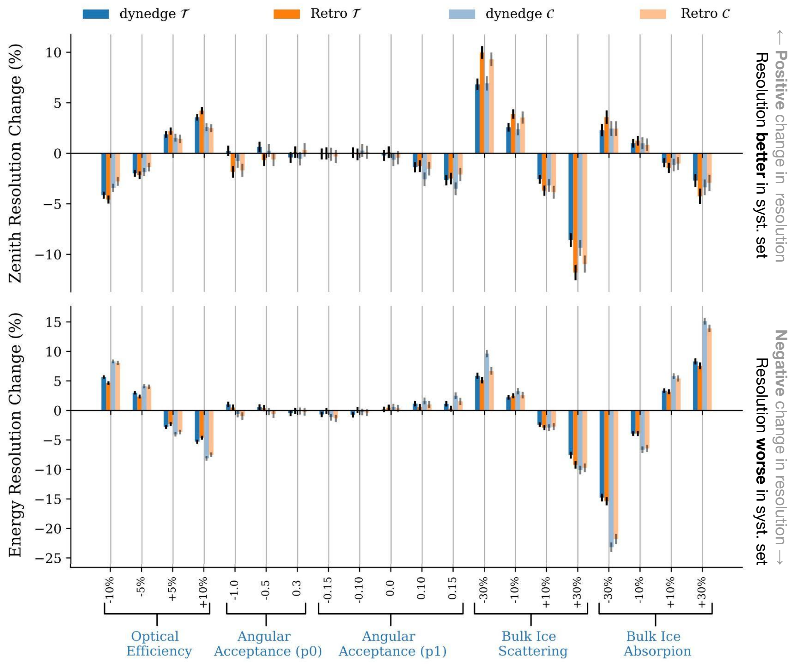
Improvements seen on this plot corresponds to an **additional 2.5 detector-years**, or 20% in terms of area. Lower estimate of GNN impact on physics results.



Robustness to systematic uncertainties — 6/6

Motivation
With complex deep learning models acting on low-level inputs, there may be a (healthy) fear that they “overfit” patterns in the training data, and do not extrapolate well to unseen data (e.g., variations under systematic uncertainties).

Result
Almost identical behaviour to RETRO, which is entirely based on physically inspired modelling. This suggests that the GNNs does not seem to introduce a greater, or less intuitive, dependence on systematic uncertainties.



Recap

- GNNs provide significant improvement in performance over alternative approaches, both likelihood-based and ML-based.

  → Increase sensitivity of e.g. oscillations analyses.

- GNNs provide speed-ups over likelihood-based reconstruction by several orders of magnitude.

  → Deploy high-resolution reco. at near-real time, early in analysis chain

- Seamless to extend to other detector / DOM configurations using same set of components.

  → Can support any upgrades to, and extensions of, the IceCube experiment.

Future research directions

Currently done:

- Per-event classification and regression
- Per-node (pulse, OM) classification and regression

Future directions:

- Physics-informed GNNs
- Segmentation of overlapping events
- Adversarial training / domain adaptation for mitigation of data/MC differences and parametrised systematic uncertainties
- GNN explainability
- Anomaly detection
- *Etc.*

3. Ways of working with ML in physics

🌶️ Spicy take 🌶️

In experimental particle physics,
machine learning only has real value
when used on experimental data.

The challenge with machine learning in physics

Challenge

⛰️ Developers may have little ML experience

⛰️ Siloed development, often from scratch

Risk

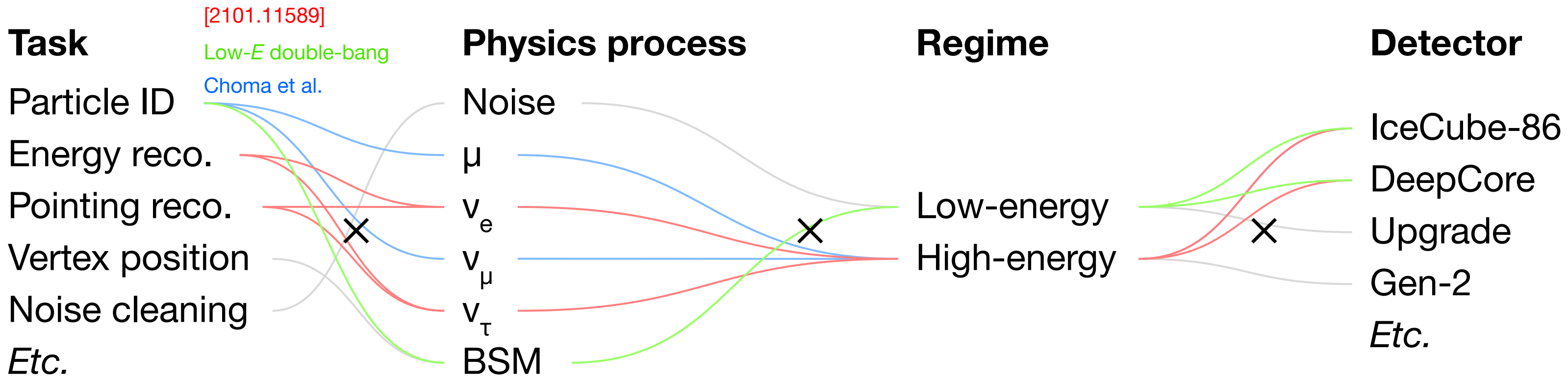
❌ Brittle, suboptimal solutions

❌ Time spent on “boilerplate” instead of physics

```
from tensorflow.keras import *
```

We're working on very related problems

Similar detectors, data, physics processes, deployment setting, end-users, etc.



"Zoo" of use cases not solved holistically

Potential for new ways of working with ML in physics

Potential

- Address “zoo” of ML use cases holistically
- Collaboration between individual use cases
- Using validated, best-practices code
- Efficient software/ML development workflows

Outcome

- More time for physics
- Better, more reliable results
- Contributions of individual ML developers has a broader, lasting impact in the collaboration

GraphNeT
Graph Neural Networks for
Neutrino Telescope Event Reconstruction

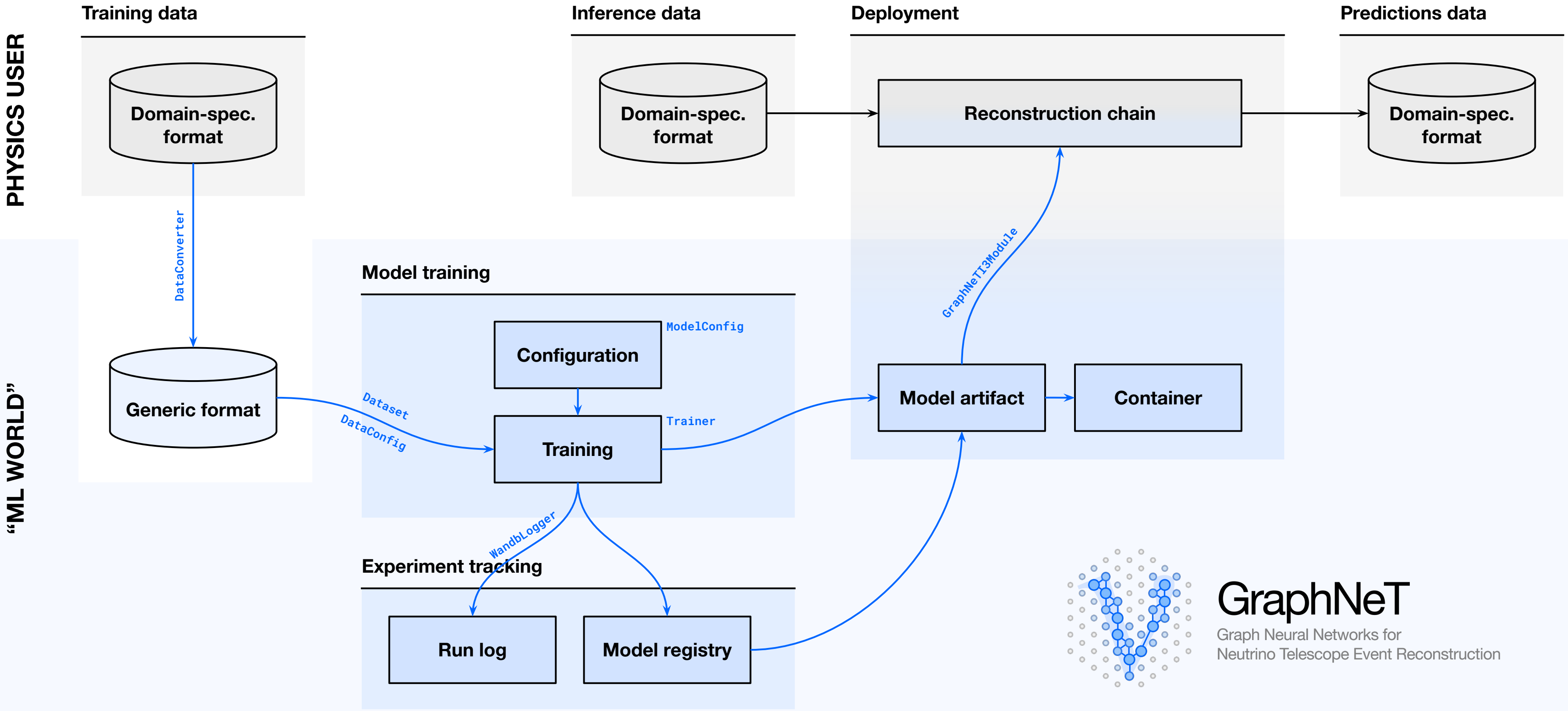
icecube/graphnet

graphnet-team

Proposition

- Reusable GNN components for plug-and-play ML
- All components for end-to-end ML pipeline (data → prod.)
- Validated code, following best practices
- Applicable across all of IceCube **+ other experiments**

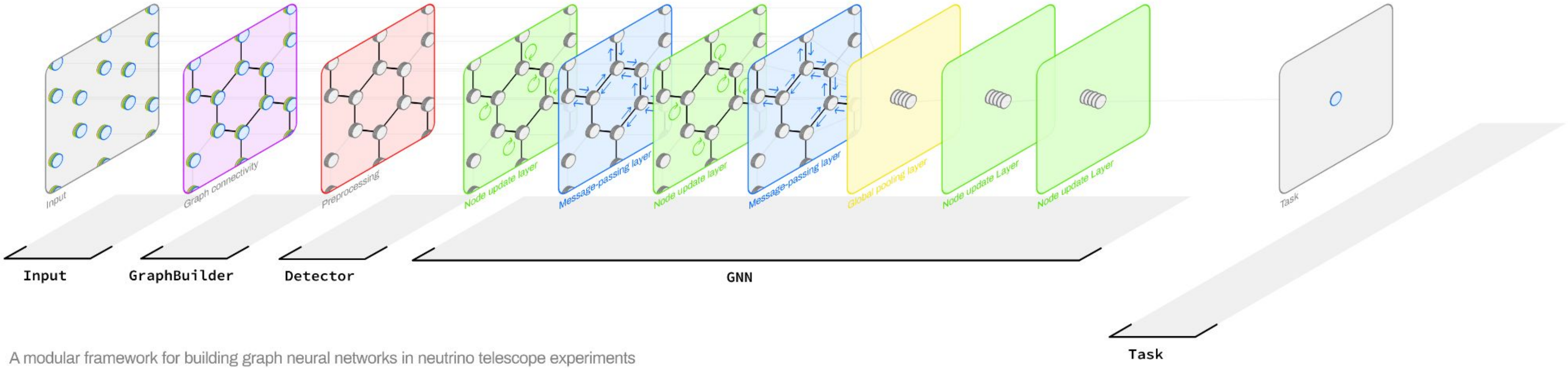
Factoring out ML from physics



GraphNeT
Graph Neural Networks for
Neutrino Telescope Event Reconstruction

GraphNeT in a nutshell

Modularised, plug-and-play ML components for any use case.



A modular framework for building graph neural networks in neutrino telescope experiments

GraphNeT in a nutshell

```python
from graphnet import (
  EuclideanGraphBuilder as GraphBuilder,
  IceCubeUpgrade as Detector,
  DynEdge_V2 as GNN,
  EnergyReconstruction as Task,
)

# Go do physics!
```

…ideally

```python
from graphnet import Model

model = Model.from_pretrained(
    "icecube-low-energy-neutrino-v2"
)

# Go do physics!
```

Collaboration impact

Summary

- What IceCube is, what it does, and how it has traditionally operated.
- How more standard ML has been used in IcCube and to what effect.
- How effective GNNs are in IceCube and similar experiments
- How to optimise the impact of ML on physics through new ways of working.

Appendix

Why a platform approach to ML may work for IceCube

	General-purpose exps. at CERN	Optical neutrino telescopes like IceCube
Detector	Several different detection principles and detectors used	Same basic detection principle + devices used across all sub-detectors
Reconstruction tasks	Myriad of reco tasks: PID and properties for most SM particles + dedicated BSM reco., each possibly leveraging multiple sub-detectors	Few reconstruction tasks: PID + few properties only for neutrinos and muons
Potential for synergy	The various performance groups operate on different inputs with different end goals. Hard to unify efforts across these groups. **However,** with the advent of particle or unified flow objects, which tries to provide a unified representation of all "particles" at the reconstruction-level, perhaps this could become (more) feasible — and perhaps solvable within a GNN paradigm.	Simplistically, all analysis rely on a large sample of high-purity neutrino events (excl. atm. muons + noise), with precise flavour ID (track/cascade), energy, and pointing. Large physics impact (no. analysis) from improving central reconstruction. Pre-trained models have high utility.