Machine Learning For Reconstruction in JUNO

Liam Francis Jones

Supervisors: Prof. Costas Andreopoulos & Prof. Xianguo Lu

The JUNO Experiment

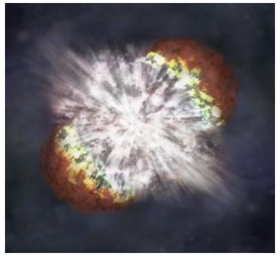
The JUNO experiment is located in the Guangdong region of southern China.

- Primary Goals
  - Determine the Neutrino Mass Ordering (NMO) → ~3σ in 6 years
  - Sub-percentage measurements of $\sin^2\theta_{12}$, Δm^2_{21} and Δm^2_{31} (world leading within 100 days!)
- Expansive complimentary physics programme

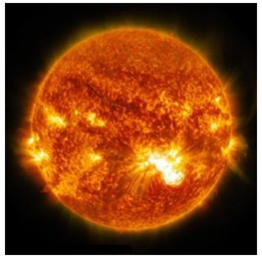
Reactor

Atmosphere

Supernova

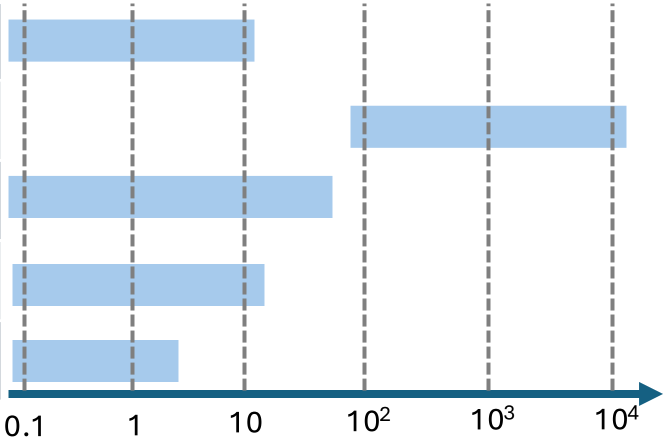


Solar



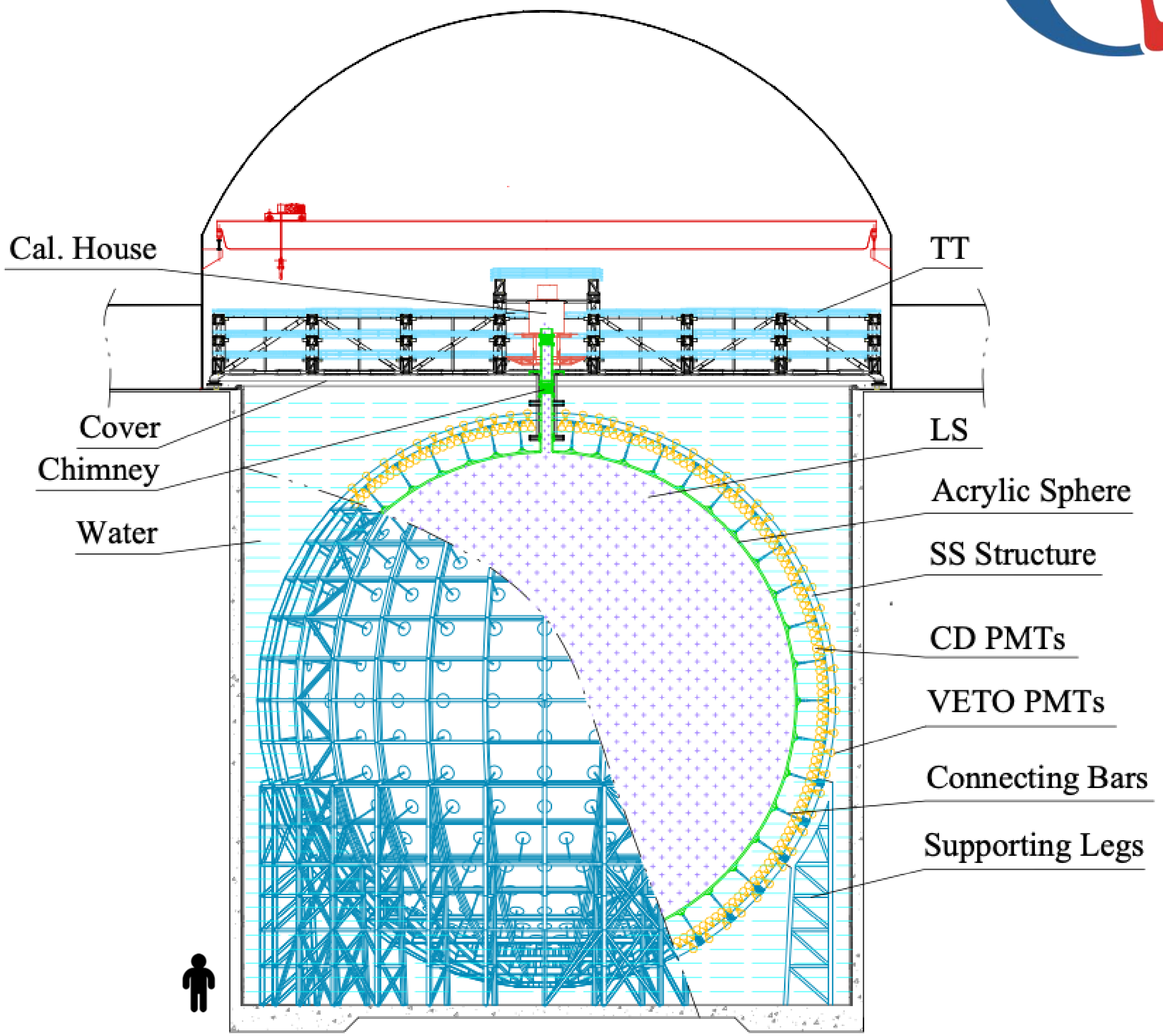
Earth

Neutrino Source	*Frequency*	*Energy Region (MeV)*
Reactor	60 / Day	
Atmosphere	100s / Year	
Supernova Burst	5000 at 10 kpc	
Sun	2-4 / Year	
Earth Crust/Mattle	400 / Year	



JUNO Detector

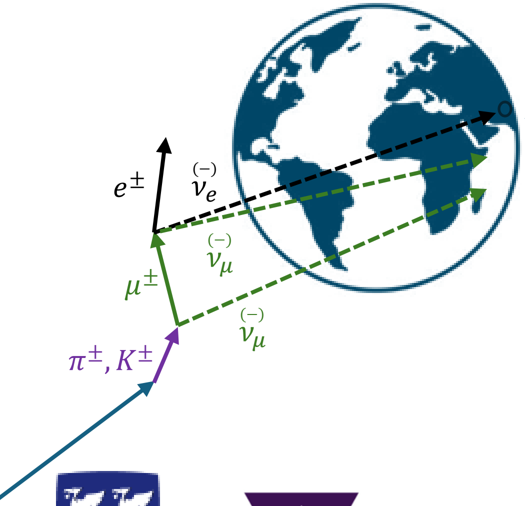
- Largest of its kind, at 20 kton of Liquid Scintillator
  - Energy Resolution of 3% at 1 MeV
  - 1400 photo-electrons per MeV of deposited energy
- Excellent photocathode coverage
  - 17,612 20-inch PMTs (75% Coverage)
  - 25,600 3-inch PMTs (3% Coverage)
- Background mitigation
  - Rock overburden of ~700m, water pool with 2,400 20-inch veto PMTs
  - Effective impurity removal

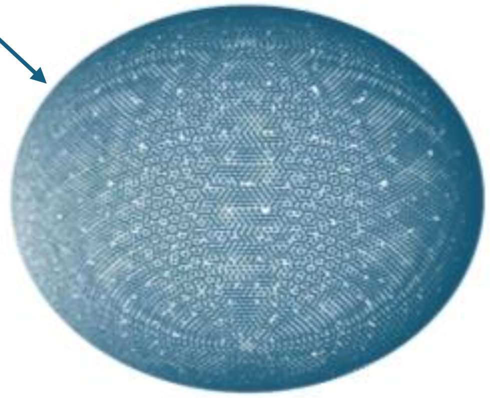


arXiv:1507.05613

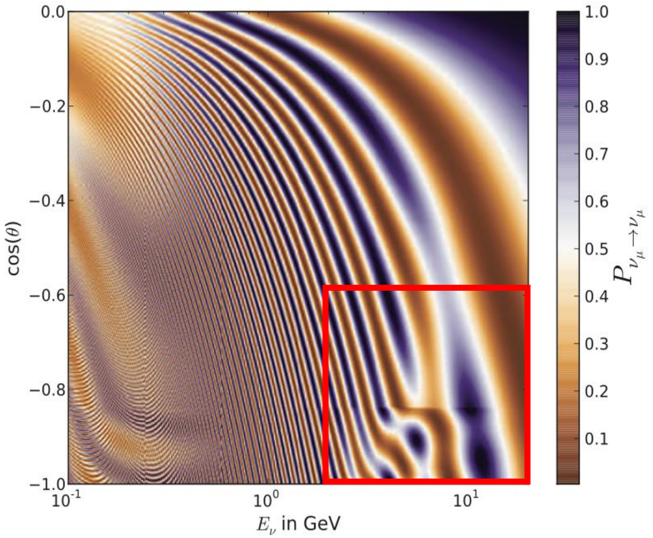
Atmospheric Neutrinos @ JUNO

Atmospheric multi-GeV neutrinos and antineutrinos offer a complementary channel to measure NMO

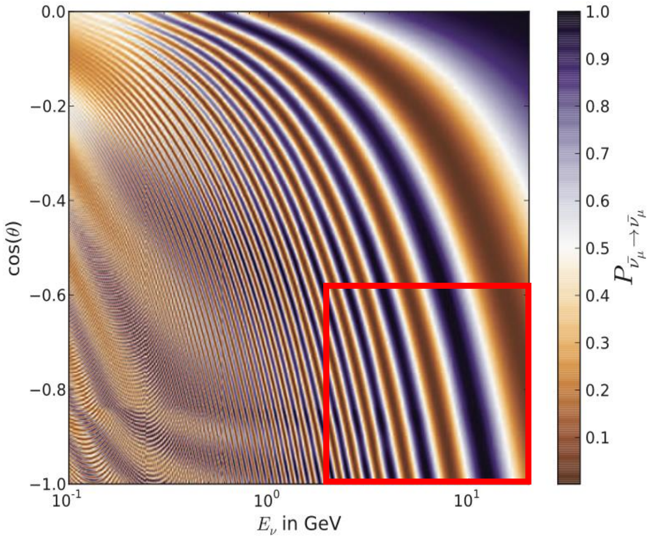




P($\nu_\mu \rightarrow \nu_\mu$) assuming NO



P($\overline{\nu}_\mu \rightarrow \overline{\nu}_\mu$) assuming NO

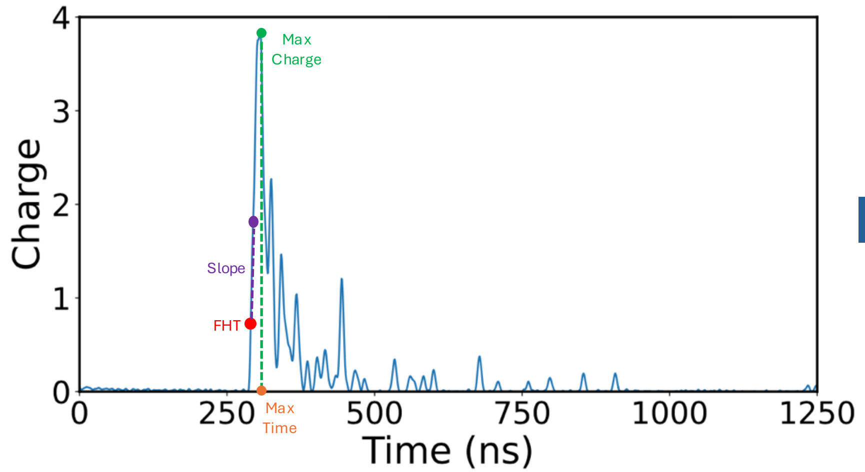


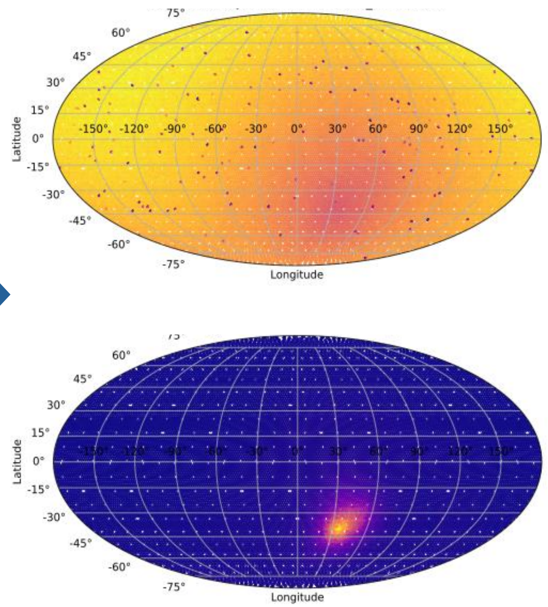
arXiv:1507.05613

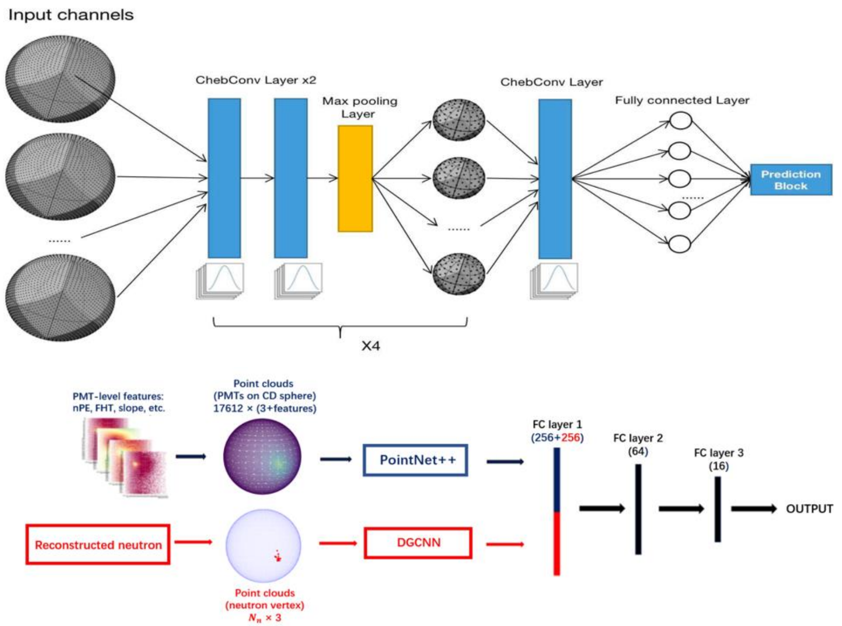
- The matter effect in atmospheric neutrinos give insight to NMO
  - If NO → Neutrinos experience stronger matter effect
  - If IO → Antineutrinos experience stronger matter effect

Reconstruction

There are three main tasks necessary for atmospheric neutrino oscillation analysis:







arXiv:2310.06281

Energy | Direction | Flavour

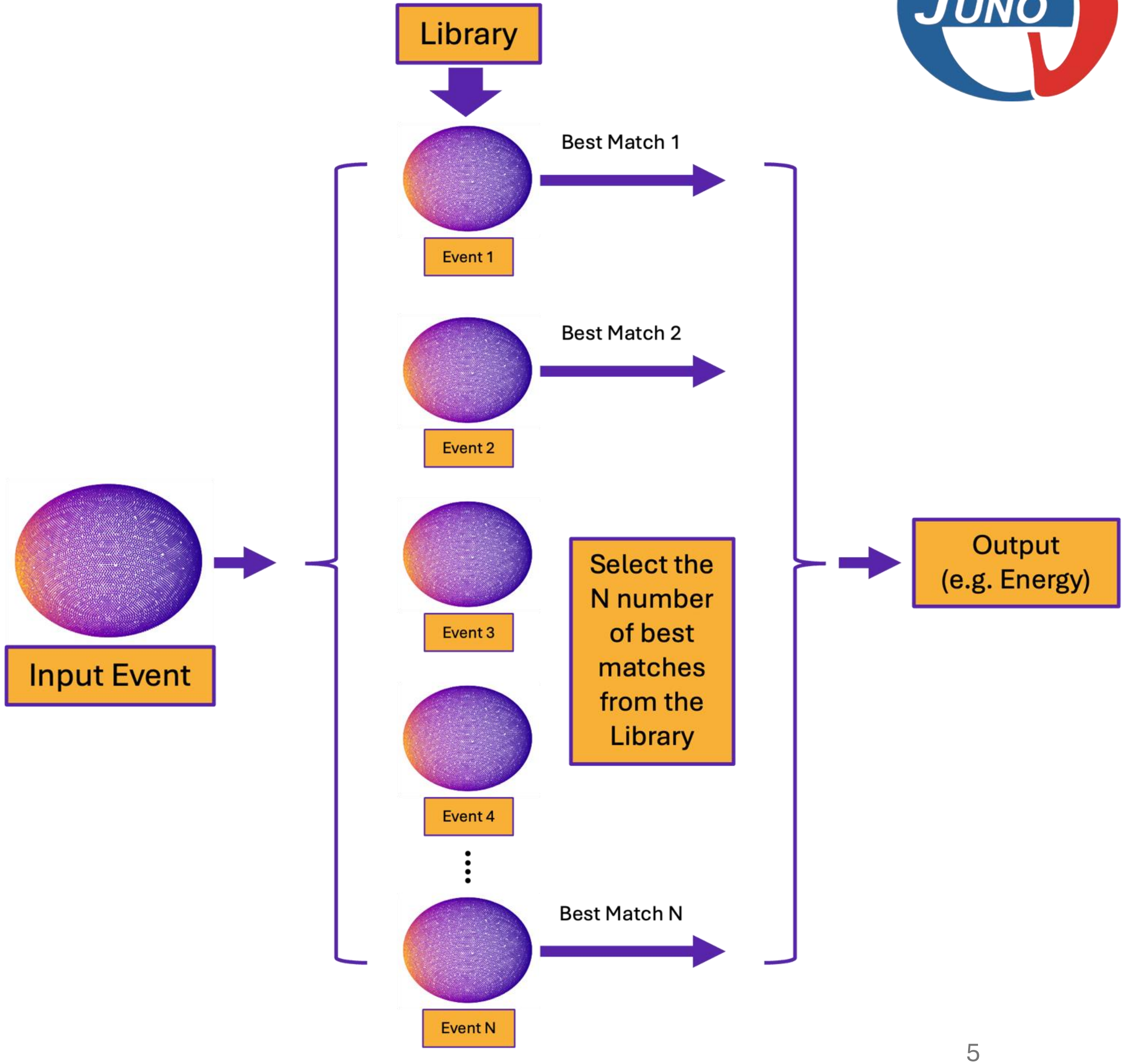
$$P_{\alpha\to\beta} \propto \sin^2\left(\frac{\Delta m^2 L}{4\,E}\right)$$

WARWICK
THE UNIVERSITY OF WARWICK

Library Event Matching

New way of classification and reconstruction:

- Compare to a large MC library
- Select the top matches
- From this, provide energy, direction and flavour with uncertainties
- Simple, interpretable and a benchmark for complex ML models

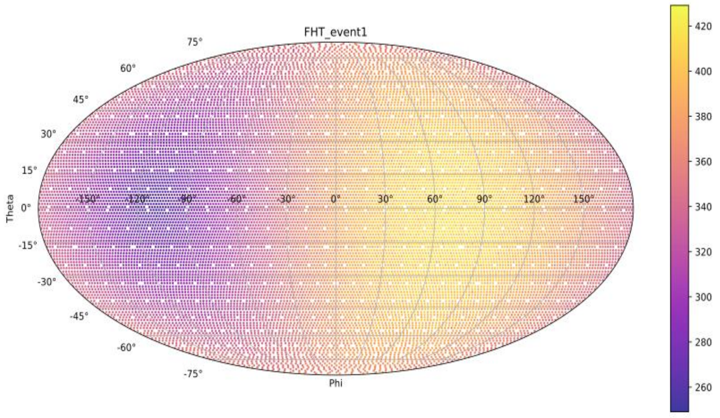


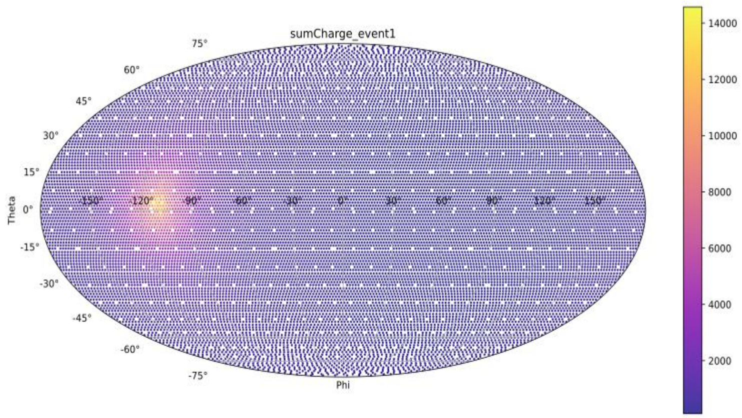
arXiv:1108.0015 arXiv:1501.00968

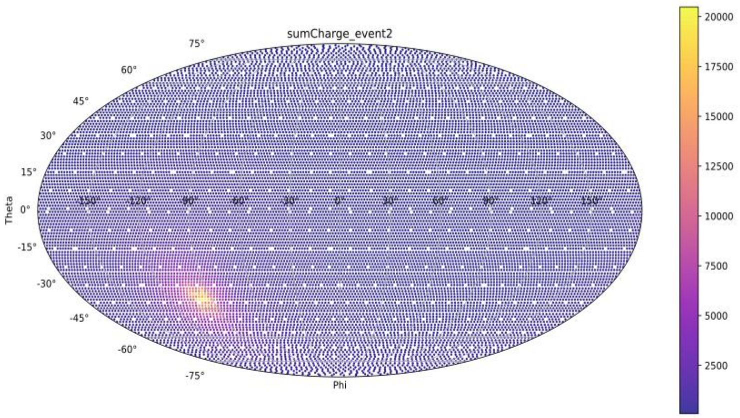
Initial Testing on ν_e Events

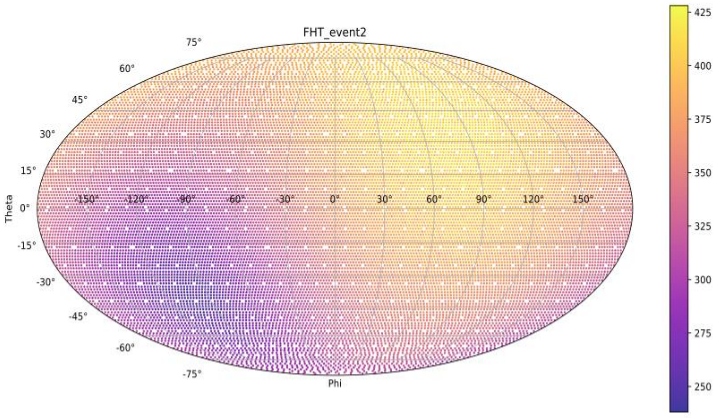
For an initial Proof-of-concept run for Library Event Matching (LEM):

- Dataset: ~3k ν_e CC events from full containment MC (1–20 GeV)
- Features: Cluster charge and angular radius
- Goal: Estimate energy and radial vertex distance



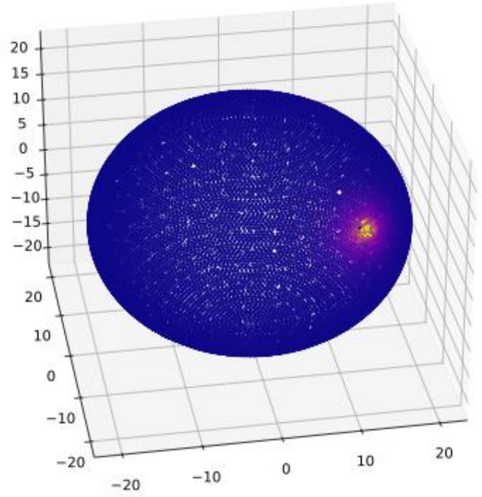


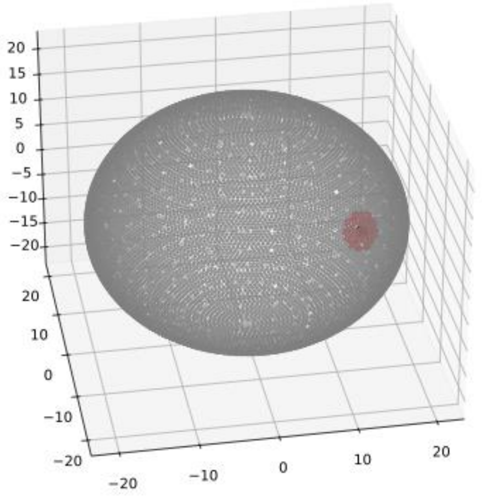




Methodology

1. Build Library
   - Simulate fully-contained atmospheric neutrino events (1–20 GeV)
2. Event Reduction
   - Cluster main activity regions in PMT data using Kernel Density Estimation (KDE)
   - Extract compact features (cluster charge and angular radius)
3. Event Comparison
   - Use k-Nearest Neighbours (KNN) to compare input event features to the library
   - Compute similarity using Manhattan/Euclidean distances
4. Output Estimation
   - Average top matches to reconstruct energy, radial vertex distance
   - Statistical uncertainty from the spread of top matches

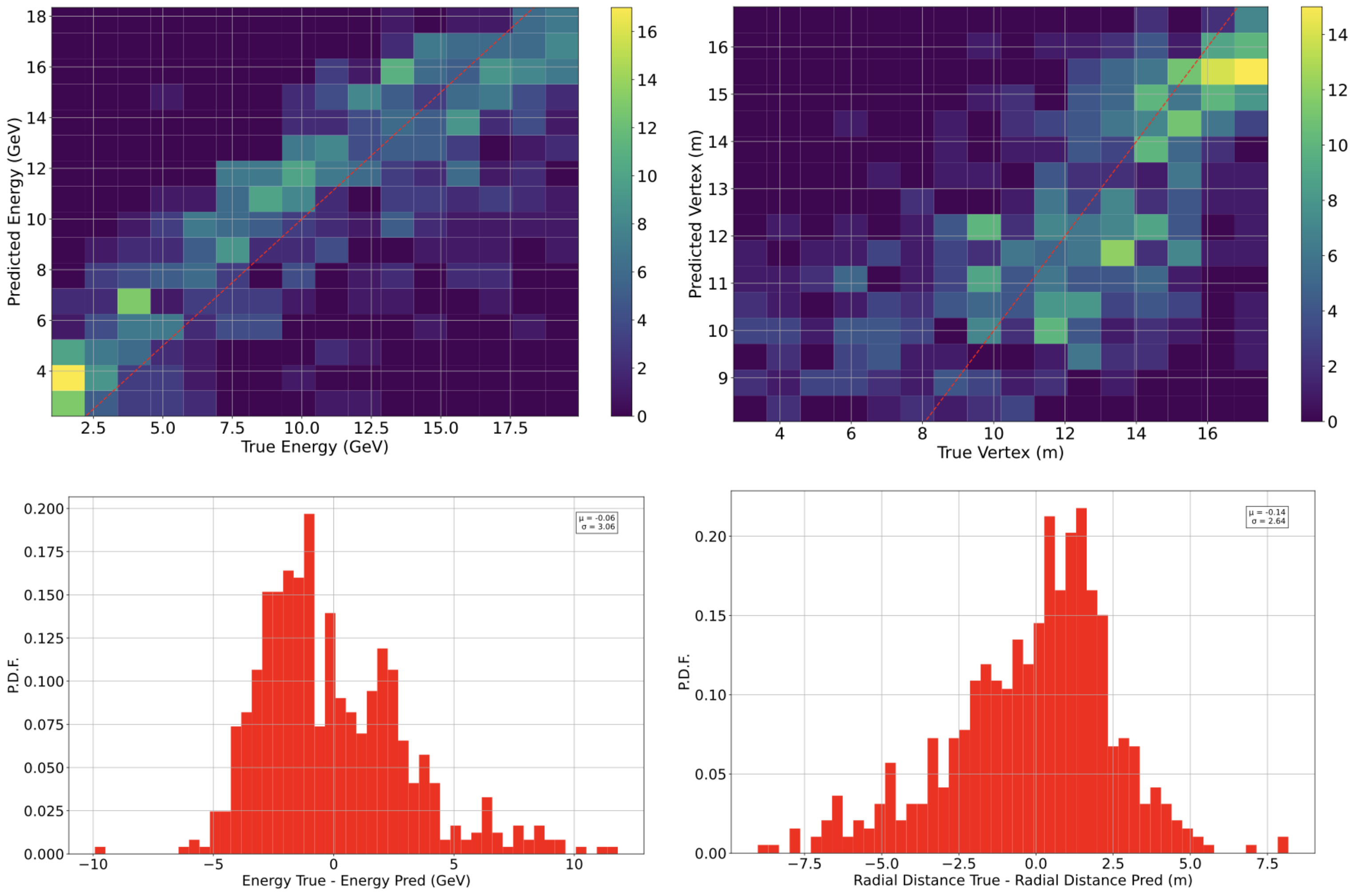




Initial Reconstruction Performance

Shows promise as an initial proof-of-concept:

- Energy and radial distance show promising trends
- The RMSE is 3.06 GeV and 2.64 m respectively
- Follows the regression line event with basic features and low stats

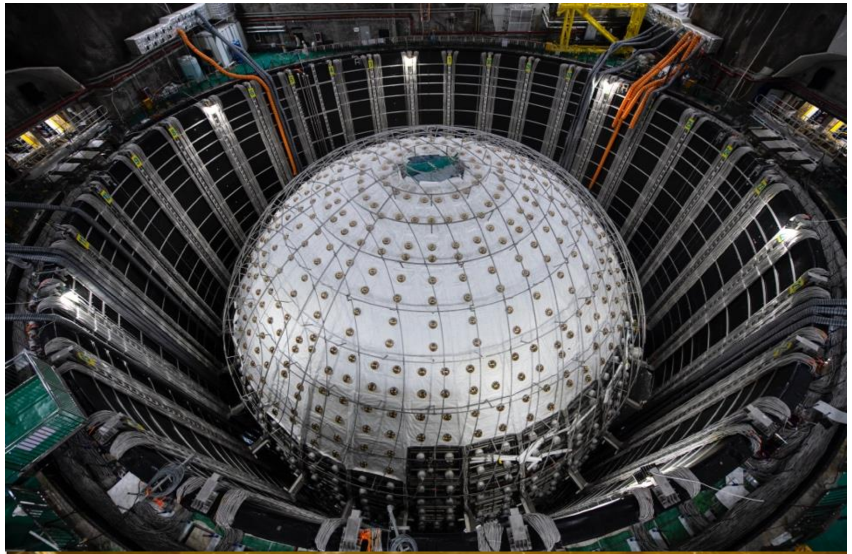


Future Developments

- Train with a larger dataset of MC samples
- Add more representative features
- Lean into deeper learning architectures
  - Use GNN models for reduced representations of events
  - Triplet loss to help?
  - Can extra features be extracted from the waveform using a NN?
- Benchmark against current JUNO ML methods
- Create a full classification and reconstruction chain

WARWICK
THE UNIVERSITY OF WARWICK

Conclusion



Work completed so far:

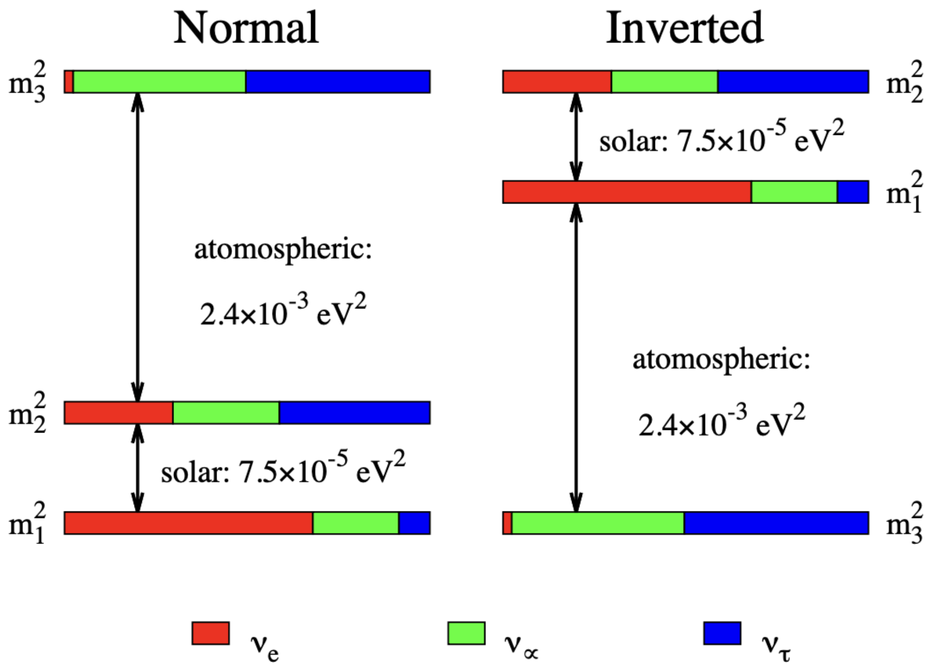
- Built a proof-of-concept LEM pipeline, cluster-based event reduction, KNN matching, and tested on ν_e CC events
- Demonstrated feasibility for basic energy and vertex estimation.
- Took part in shift work during the water filling commissioning phase

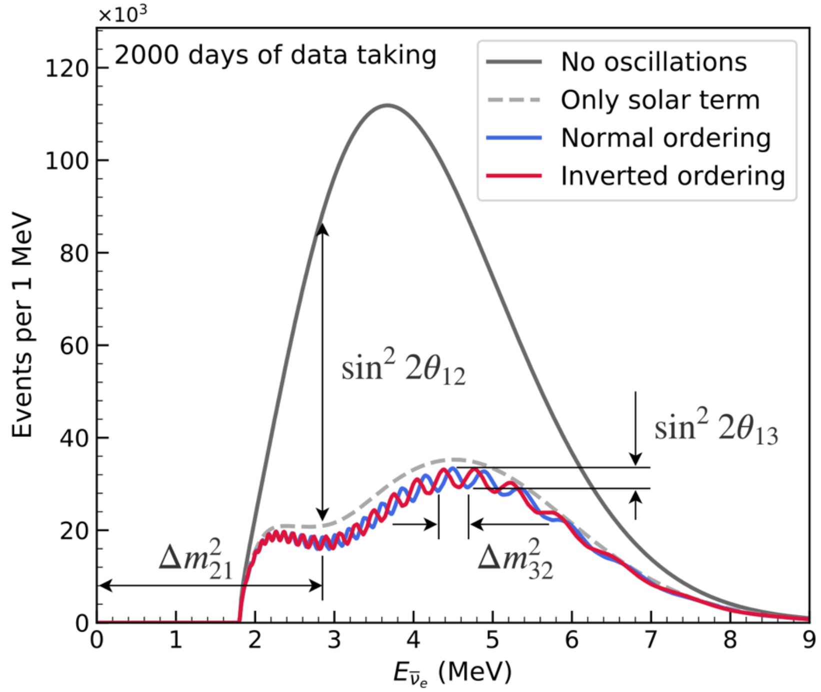
Future Work:

- Realise a full LEM model for use in JUNO
- Test with deep learning models

Back Up

NMO @ JUNO


Normal
Inverted
m_3^2
m_2^2
m_1^2
atomospheric:
2.4×10^{-3} eV^2
solar: 7.5×10^{-5} eV^2
ν_e
ν_α
ν_τ


$\times10^3$
2000 days of data taking
No oscillations
Only solar term
Normal ordering
Inverted ordering
Events per 1 MeV
$\sin^2 2\theta_{12}$
$\sin^2 2\theta_{13}$
Δm^2_{21}
Δm^2_{32}
$E_{\bar{\nu}_e}$ (MeV)

PMT Statistics

<table>
<tr><th colspan="2">Size</th><th>20-inch</th><th>20-inch</th><th>3-inch</th></tr>
<tr><td colspan="2">Brand</td><td>Hamamatsu</td><td>NNVT</td><td>HZCV</td></tr>
<tr><td colspan="2">Nº PMTs</td><td>5000</td><td>15012</td><td>25600</td></tr>
<tr><td colspan="2">Charge Collection</td><td>Dynode</td><td>MCP</td><td>Dynode</td></tr>
<tr><td colspan="2">Gain</td><td>$\sim 10^7$</td><td>$\sim 10^7$</td><td>3×10^6</td></tr>
<tr><td colspan="2">Photon Detection Efficiency</td><td>28.5%</td><td>30.1%</td><td>25%</td></tr>
<tr><td rowspan="2">Mean Dark Count Rate (kHz)</td><td>Bare</td><td>15.3</td><td>49.3</td><td rowspan="2">0.5</td></tr>
<tr><td>Potted</td><td>17.0</td><td>31.2</td></tr>
<tr><td colspan="2">Transit Time Spread (σ) [ns]</td><td>1.3</td><td>7</td><td>1.6</td></tr>
<tr><td colspan="2">Coverage</td><td colspan="2">75%</td><td>3%</td></tr>
<tr><td colspan="2">Reference</td><td colspan="2">arXiv. 2205.08629</td><td>NIM.A 1005 (2021) 165347</td></tr>
</table>

Energy Resolution Vs Other Experiments

	Target Mass (ton)	Energy Resolution at 1 MeV
Daya Bay	20	8%
Borexino	300	5%
KamLAND	1,000	6%
JUNO	20,000	3%