Yury Malyshkin
(Joint Institute for Nuclear Research)

for the JUNO Collaboration

Application of machine learning techniques for event reconstruction in JUNO

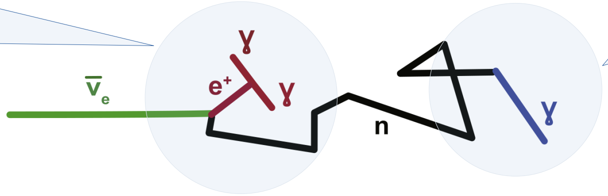
NuFact 2021: The 22nd International Workshop on Neutrinos from Accelerators
September 6-11, 2021 (Cagliari / online)

Introduction

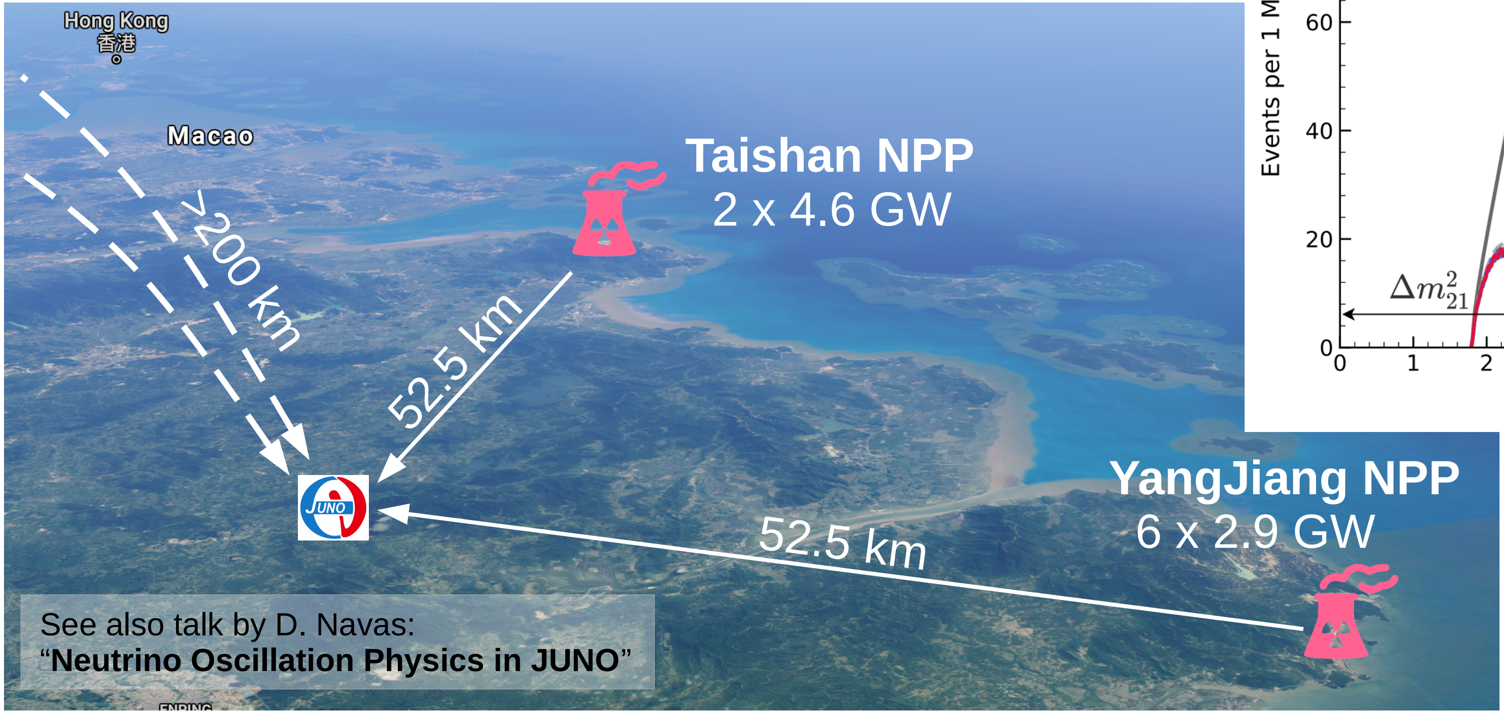
- JUNO detector is a complex apparatus suitable for multiple neutrino studies and other topics
- New machine learning (ML) are under development for data-processing at different stages:
  - trigger
  - waveform reconstruction
  - particle type discrimination
  - muon track reconstruction
  - lower energy particles vertex/energy reconstruction
  - …
- ML algorithms are usually much faster than the traditional ones. They may complement and/or replace traditional approaches. However, their reliability is still a subject of investigation.
- This talk is focused only on vertex and energy reconstruction of events from reactor cores (i.e. 0-10 MeV energy range) – the main signal in JUNO

Neutrino Energy Spectrum at JUNO

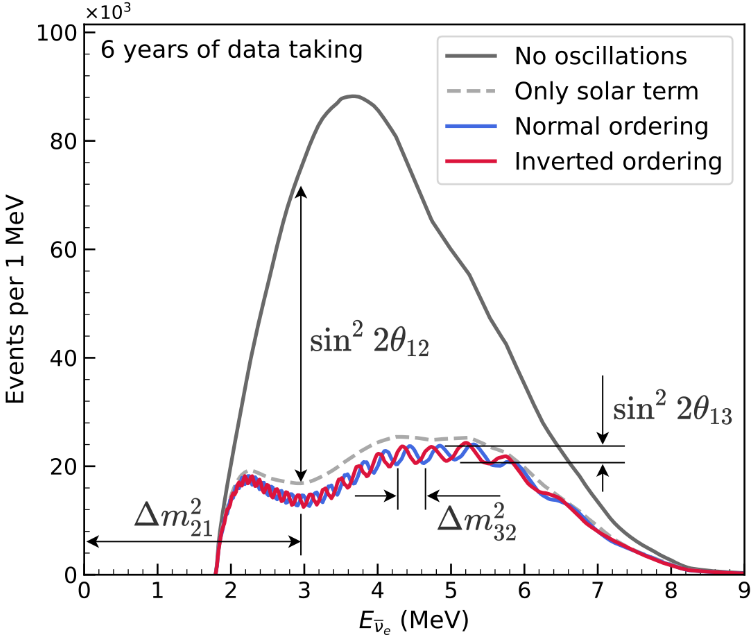
Handle for neutrino energy



IBD tag:
2.2 MeV within
~200 μs



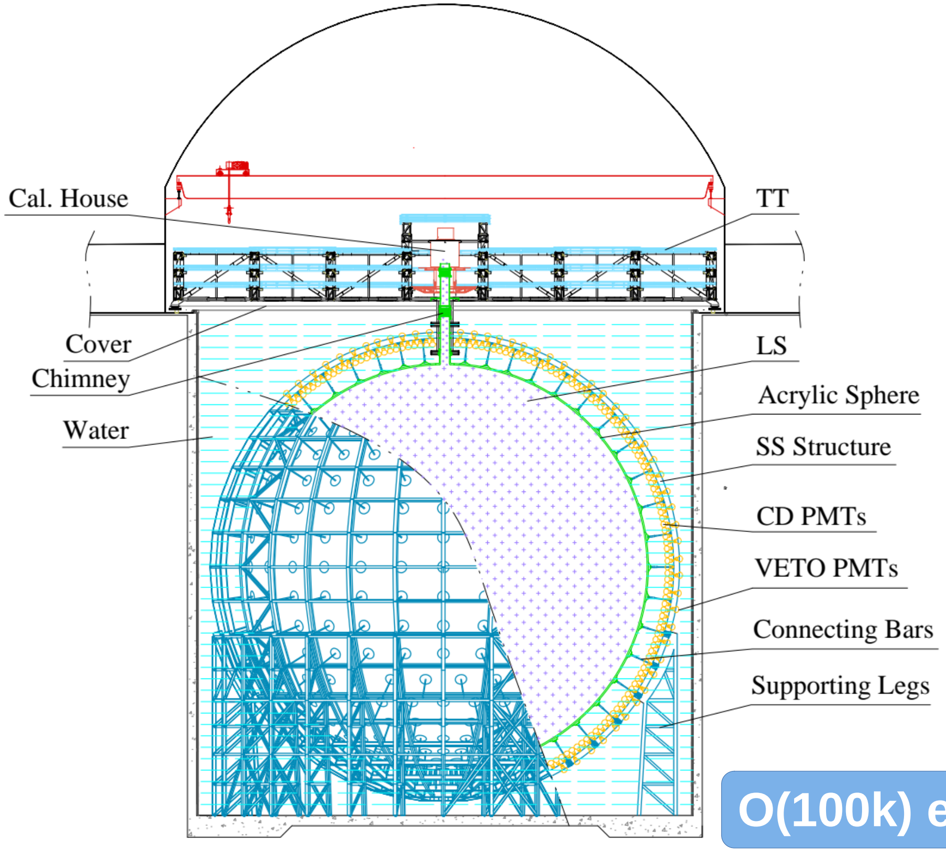
See also talk by D. Navas:
"**Neutrino Oscillation Physics in JUNO**"



Precise energy measurement is the key for neutrino oscillation physics in JUNO

JUNO Detector

See also talk by A. Paoloni:
"**Status of JUNO Experiment**"



- **Source:** 8 reactor cores (2 NPPs)
  - Powerful and relatively well understood
- **Baseline:** 52.5 km
  - Optimized for resolving NMO
- **Overburden:** ~700 m
  - Cosmic background suppression
- **Detection channel:** $\bar{\nu}_e + p \rightarrow e^+ + n$
  - Time coincident signal
  - Positron brings energy information
- **Target:** 20 kton of LAB-based liquid scintillator
  - Optimized admixtures for high light yield and transparency
  - ~1300 p.e. / MeV
- **Light detection:** 18000 20'' PMTs + 25600 3'' PMTs
  - >75% photo-coverage
  - two independent PMT systems

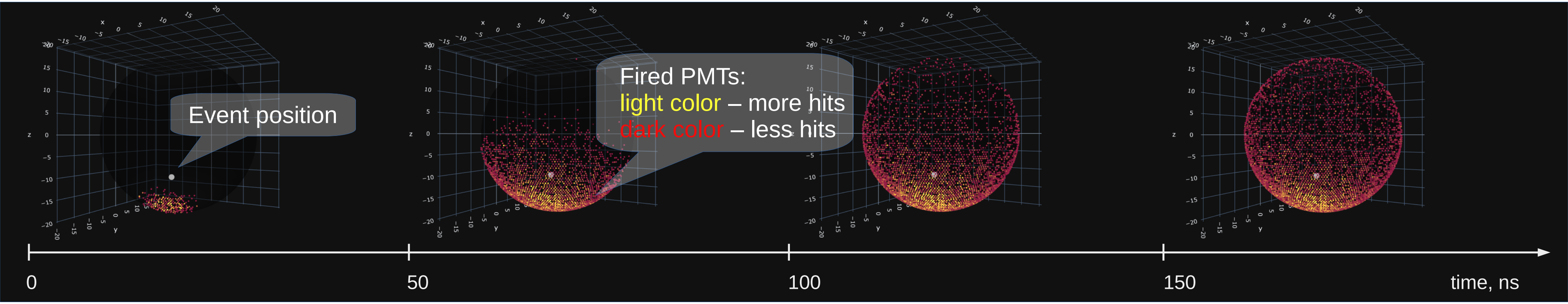
O(100k) events in 6 years

3% energy resolution @ 1 MeV

The Task

INPUT: detected signal (PMT hits) evolving in time

- from Monte Carlo: can provide virtually unlimited statistics and coverage (IN THIS TALK)
- from calibrations: real data but limited statistics and coverage (FUTURE WORK)



OUTPUT:
- Event **vertex** *(needed for event selection and as input for other reconstruction algorithms)*
- Event **energy** *(directly used in physics analysis)*

IMPORTANT:
- Get the most from available information
- High reconstruction speed

How to deal with the input data?

~18000 PMT channels arranged on spherical surface

Each PMT channel provides: (1) charge, (2) first hit time

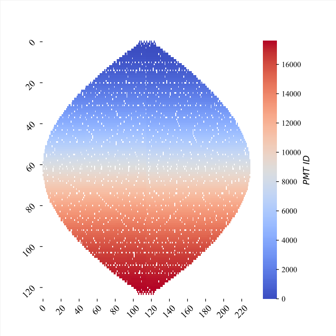
Aggregated features

Reduce input by aggregating information into a small amount of features

Train Deep Neural Network (DNN) or Boosted Decision Tree (BDT) model

2D projection -> CNN

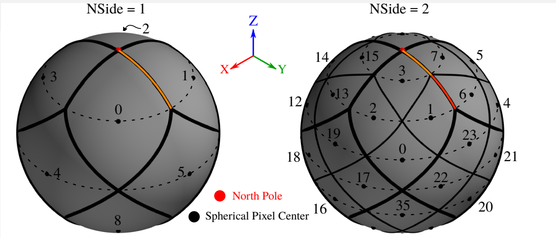
Project charge onto 2D plane, with two channels for charge and first hit time



Train Convolutional Neural Network (CNN)

HEALPix -> Graphs

Partition sphere using HEALPix pixelization, with two channels for charge and first hit time



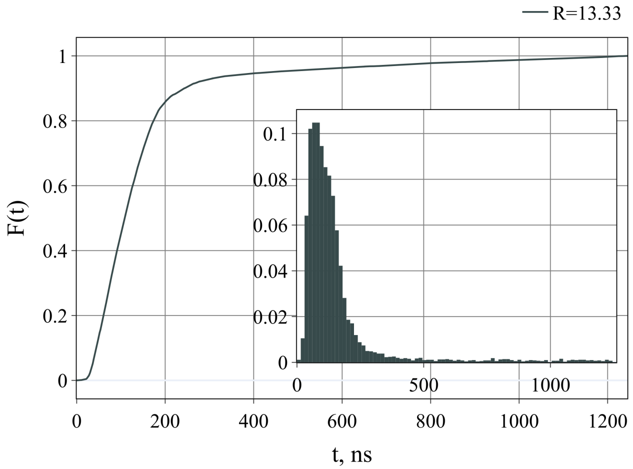
Assign pixels as node of graphs and train Graph Neural Network (GNN)

Aggregated Features

One may extract different quantities from the full signal, like

- total number of hits ($\texttt{nHits}$)
- coordinates of center of charge $\mathbf{r}_{\text{cc}} = \dfrac{\sum_i^{N_{\text{PMTs}}} \mathbf{r}_{\text{PMT}_i} q_i}{\sum_i^{N_{\text{PMTs}}} q_i}$
- hit time profile percentiles ($\texttt{ht\_1p}$, $\texttt{ht\_2p}$, …)
- ...

and use them as features in simple models (DNN or BDT*)

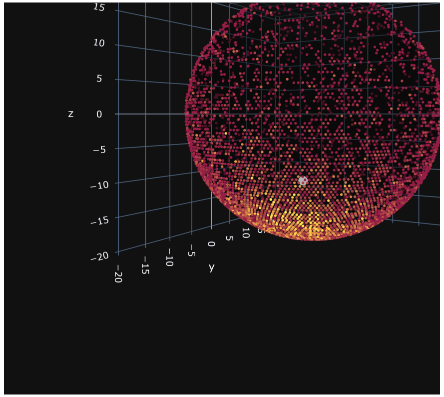


It turns out [arXiv:2106.02907] that the following subset of features performs almost as good as the full set (ordered by importance):

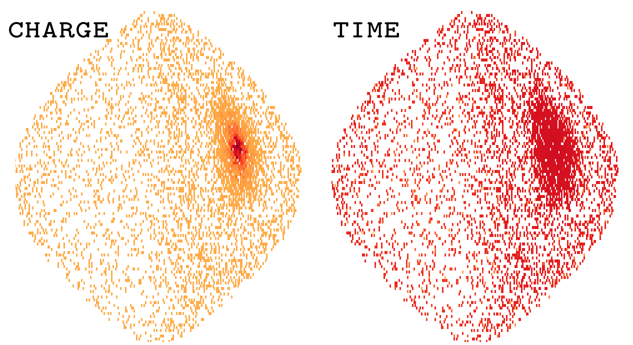
nHits, ht_20p, rho_cc, ht_2p, ht_35p, R_cc, ht_75p

* we use Boosted Decision Trees here

2D projection -> Convolutional Neural Network



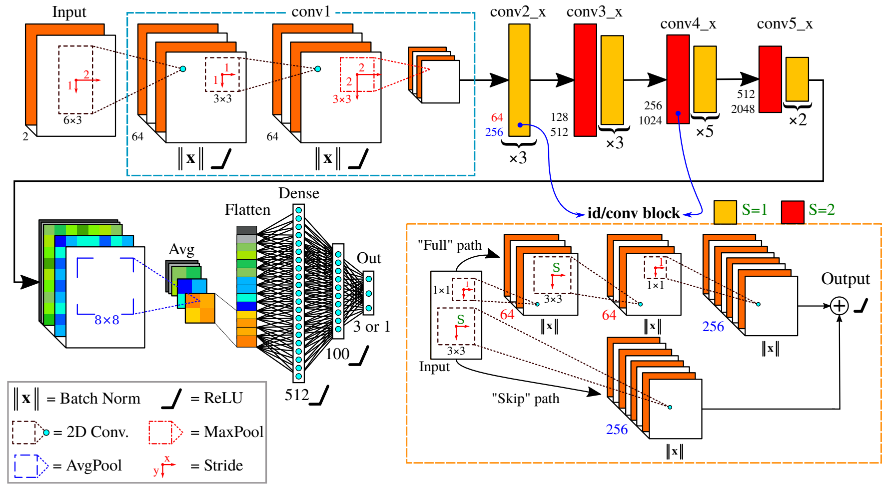
projection to 2-channel 2D image with one-to-one PMT-pixel correspondence



feed CNN

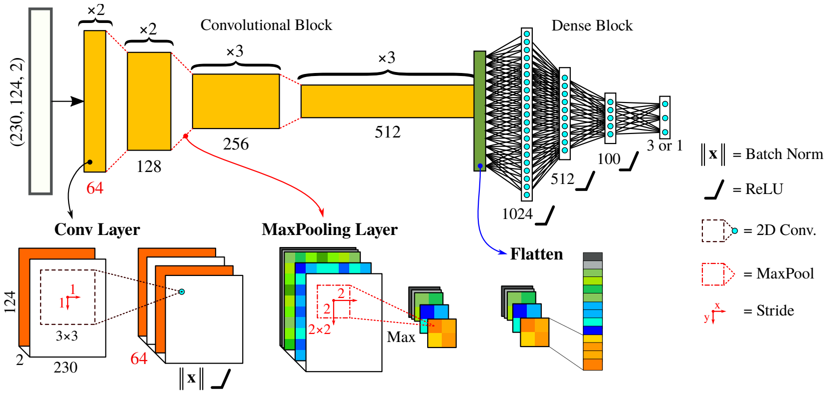
ResNet-J

(inspired by ResNet-50)

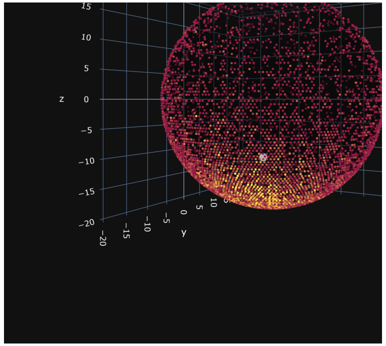


VGG-J

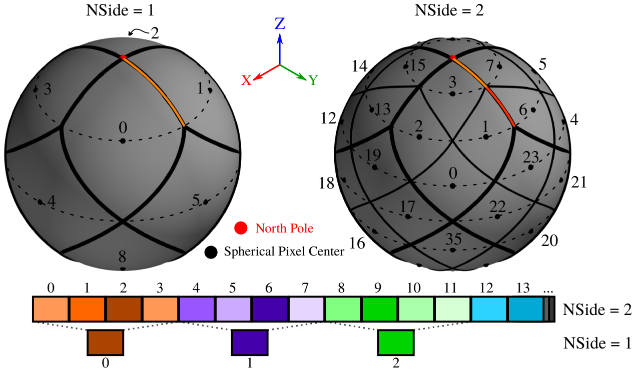
(inspired by VGG)

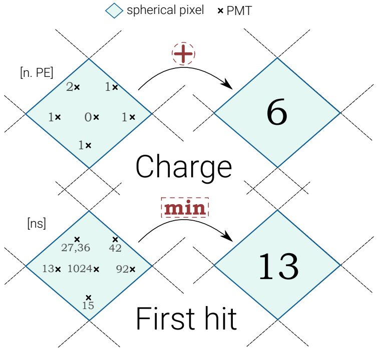


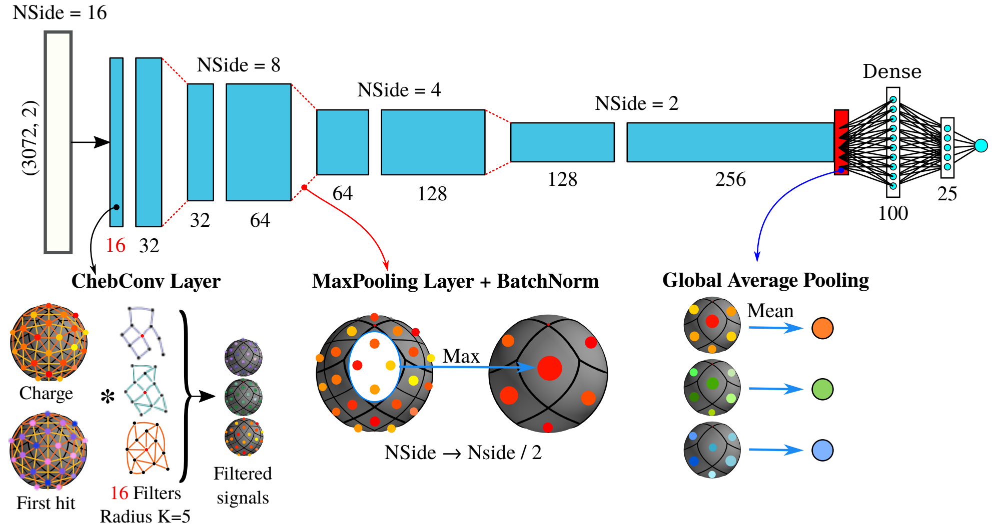
HEALPix -> Graph Neural Network



- divide the sphere surface into 3072 pixels (using HEALPix with Nside=16)
- define total charge and first hit time in each pixel





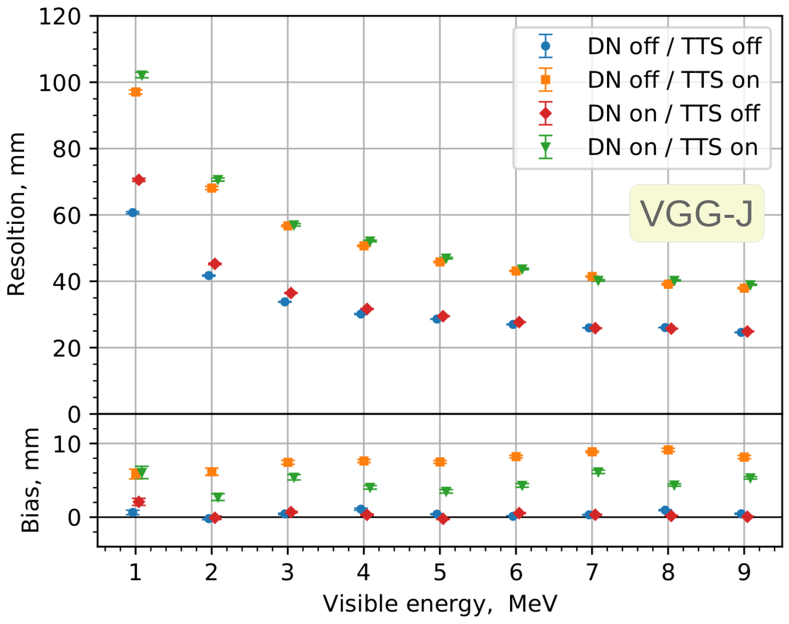


- define pixels as graph nodes
- define neighborship conditions
- define convolution operations
- feed data as 2-channel input

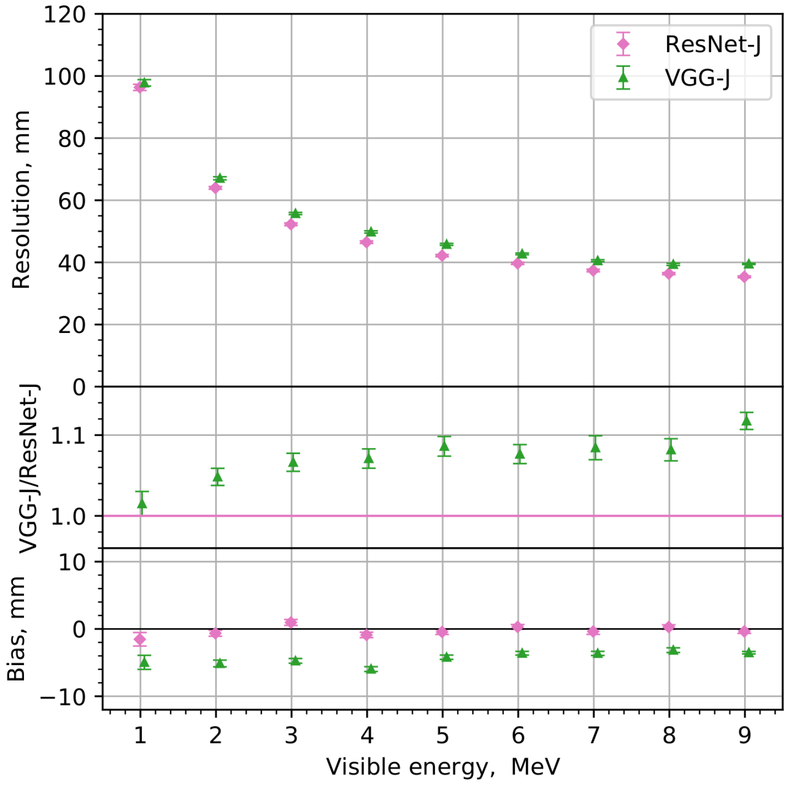
Reconstruction of Event Vertex

Electronics effects:
DN (dark noise) – stochastic hits in PMT channels
TTS (transition time spread) – smearing of first hit time time

Effect of TTS dominates – **Position is mainly defined by time information**



Simple models (BDT and DNN) provide much worse performance (not shown)



We also failed to get good resolution from Graph Neural Network (GNN-J)

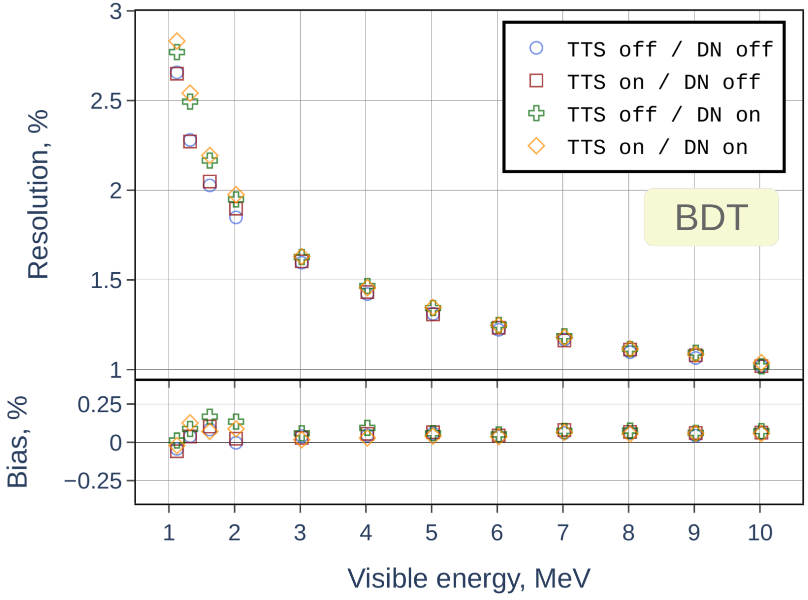
ResNet-J slightly outperforms VGG-J both in terms of resolution and bias

[arXiv:2101.04839]

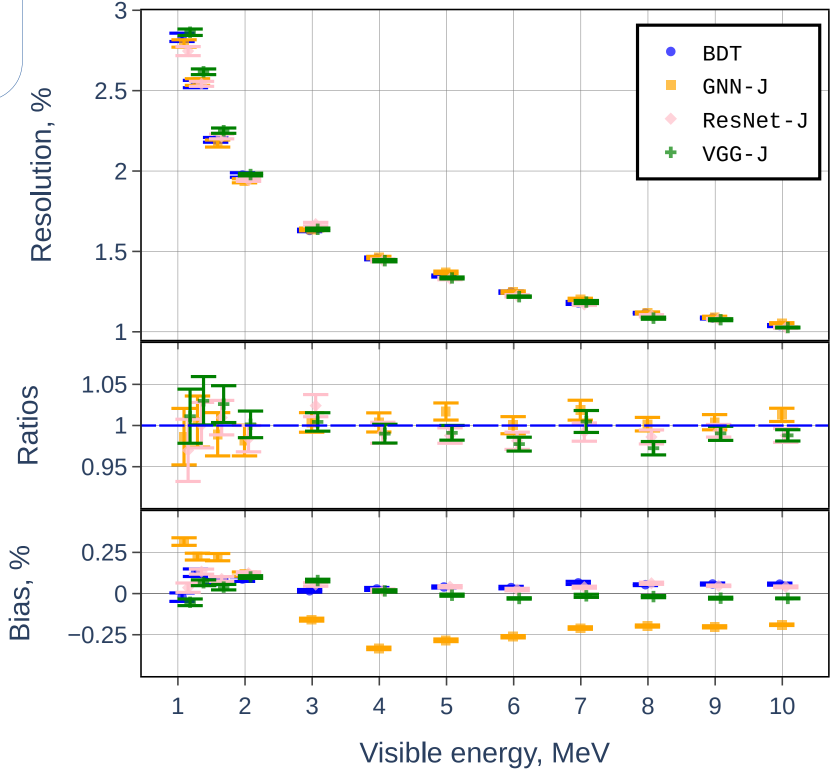
Reconstruction of Event Energy

Electronics effects:
DN (dark noise) – stochastic hits in PMT channels
TTS (transition time spread) – smearing of first hit time time

Effect of DN dominates – **energy is mainly defined by charge information**



All the models, simple and complex ones, perform nearly equally, differing mainly below 3 MeV.
(DNN, not shown, is comparable with BDT)



So, we may use DNN or BDT which are much faster!

[arXiv:2106.02907]

What if MC is not accurate?

Monte Carlo simulation has to be used for training to provide enough representativity.

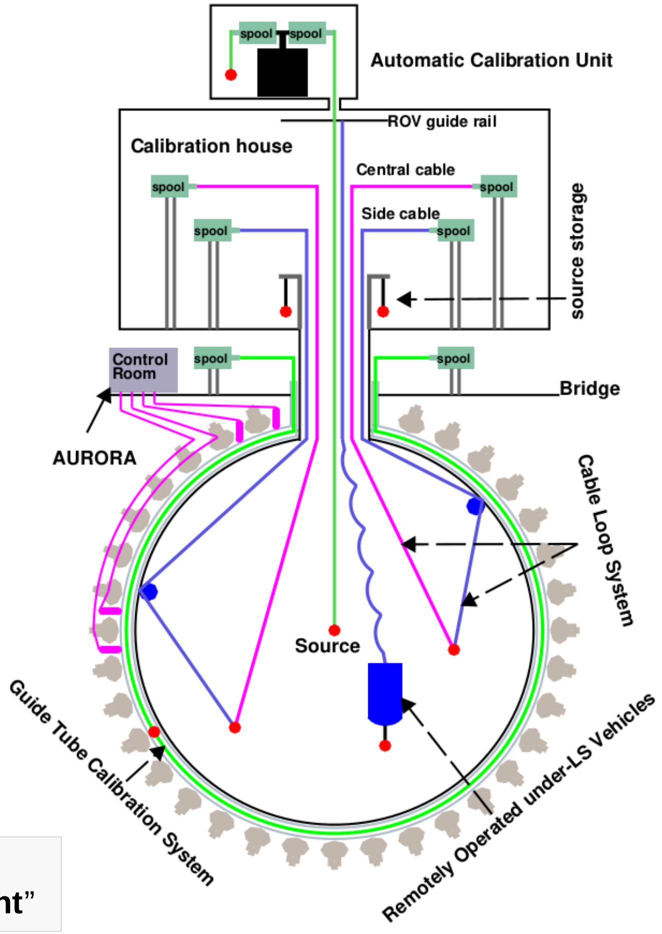
Suppose our MC is not accurate enough (although a lot of attention is paid to its precision in JUNO). Then the model **prediction may be biased**.

Calibration data are to be used to:

- understand how the reconstruction output is sensitive to MC inaccuracy
- fine-tune the models

ON-GOING WORK

Sources/Processes	Type	Radiation
^{137}Cs	γ	0.662 MeV
^{54}Mn	γ	0.835 MeV
^{60}Co	γ	1.173 + 1.333 MeV
^{40}K	γ	1.461 MeV
^{68}Ge	e^+	annihilation 0.511 + 0.511 MeV
^{241}Am-Be	n, γ	neutron + 4.43 MeV (^{12}C*)
^{241}Am-^{13}C	n, γ	neutron + 6.13 MeV (^{16}O*)
(n,γ)p	γ	2.22 MeV
(n,γ)^{12}C	γ	4.94 MeV or 3.68 + 1.26 MeV



See also talk by Yue Meng:
"**Calibration Strategy of the JUNO Experiment**"

[arXiv:2011.06405]

Summary

- A lot of ML activities in JUNO are not covered in this talk
- Vertex and energy reconstruction discussed:
  - JUNO has enormous number of channels and require **faster** methods for data-processing
  - ML algorithms seem to be able to provide excellent precision at high speed:
    - for **vertex reconstruction** complex networks with granular input are needed
    - for **energy reconstruction** simple models with a small set of aggregated features work as accurate as the complex ones but much faster
- Reliability of ML models trained on synthetic data and ways of their tuning is under investigation
- Working hard to be prepared to process real-data!

Thank you for your attention!

See also: Allesandro Paoloni “Status of JUNO” and other talks