DOCTORAT MATIERE BRETAGNE MOLECULES LOIRE ET MATERIAUX

UNIVERSITÉ DE NANTES

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE NANTES

ÉCOLE DOCTORALE N° 596
Matière, Molécules, Matériaux
Spécialité : *Physique des particules*

Par
Léonard Imbert

Precision measurement of solar neutrino oscillation parameters with the JUNO small PMTs system and test of the unitarity of the PMNS matrix

Thèse présentée et soutenue à Nantes, le Too soon and too early at the same time
Unité de recherche : Laboratoire SUBATECH, UMR 6457

Rapporteurs avant soutenance :

Someone	Very high position, Somewhere
Someone else	Another high position, Somewhere else

Composition du Jury :

Président :	Jacques Chirac	French president, France
Examinateurs :	Yoda	Master, Dagobha
Dir. de thèse :	Frédéric Yermia	Professeur, Université de Nantes
Co-dir. de thèse :	Benoit Viaud	Chargé de recherche, CNRS

Invité(s) :

Victor Lebrin	Docteur, CNRS/SUBATECH

List of Abbreviations

ACU	**A**utomatic **C**alibration **U**nit
BDT	**B**oosted **D**ecision **T**ree
CD	**C**entral **D**etector
CLS	**C**able **L**oop **S**ystem
CNN	**C**onvolutional **NN**
DN	**D**ark **N**oise
DNN	**D**eep **NN**
FCDNN	**F**ully **C**onnected **D**eep **NN**
GNN	**G**raph **NN**
GT	**G**uiding **T**ube
IBD	**I**nverse **B**eta **D**ecay
IO	**I**nverse **O**rdering
JUNO	**J**iangmen **U**nderground **N**eutrino **O**bservatory
LPMT	**L**arge **PMT**
LS	**L**iquid **S**cintillator
MSE	**M**ean **S**quared **E**rror
MC	**M**onte **C**arlo simulation
ML	**M**achine **L**earning
NMO	**N**eutrino **M**ass **O**rdering
NN	**N**eural **N**etwork
NO	**N**ormal **O**rdering
NPE	**N**umber of **P**hoto **E**lectron
OSIRIS	**O**nline **S**cintillator **I**nternal **R**adioactivity **I**nvestigation **S**ystem
PE	**P**hoto **E**lectron
PMT	**P**hoto-**M**ultipliers **T**ubes
PReLU	**P**arametrized **Re**ctified **L**inear **U**nit
ROV	**R**emotely **O**perated under-LS **V**ehicle
ReLU	**Re**ctified **L**inear **U**nit
SPMT	**S**mall **PMT**
TAO	**T**aishan **A**ntineutrino **O**servatory
TR Area	**T**otal **R**eflexion **Area**
TTS	**T**ime **T**ransit **S**pread
TT	**T**op **T**racker
WCD	**W**ater **C**herenkov **D**etector

Contents

Remerciements

Introduction

Chapter 1

Neutrino physics

The neutrino, or ν for the close friends, a fascinating and invisible particle. Some will say that dark matter also have those property but at least we are pretty confident that neutrinos exists.

1.1 Standard model

Decrire le m
Regarder th
Kochebina
Limite du r
Interessant,
les neutrino
CP ? Pb des

1.1.1 Limits of the standard model

1.2 Historic of the neutrino

First theories

Discovery

Milestones and anomalies

1.3 Oscillation

1.3.1 Phenomologies

1.4 Open questions

Chapter 2

The JUNO experiment

"Ave Juno, rosae rosam, et spiritus rex". It means nothing but I found it in tone.

The first idea of a medium baseline (∼52 km) experiment, was explored in 2008 [1] where it was demonstrated that the Neutrino Mass Ordering (NMO) could be determined by a medium baseline experiment if $\sin^2(2\theta_{13}) > 0.005$ without the requirements of accurate knowledge of the reactor antineutrino spectra and the value of Δm^2_{32}. From this idea is born the Jiangmen Underground Neutrino Observatory (JUNO) experiment.

JUNO is a neutrino detection experiment under construction located in China, in Guangdong proving, near the city of Kaiping. Its main objectives are the determination of the mass ordering at the 3-4σ level in 6 years of data taking and the measurement at the sub-percent precision of the oscillation parameters Δm^2_{21}, $\sin^2\theta_{12}$, Δm^2_{32} and with less precision $\sin^2\theta_{13}$[2].

FIGURE 2.1 – **On the left:** Location of the JUNO experiment and its reactor sources in southern china. **On the right:** Aerial view of the experimental site

For this JUNO will measure the electronic anti-neutrinos ($\bar{\nu}_e$) flux coming from the nuclear reactors of Taishan, Yangjiang, for a total power of 26.6 GW$_{th}$, and the Daya Bay power plant to a lesser extent. All of those cores are the second-generation pressurized water reactors CPR1000, which is a derivative of Framatome M310. Details about the power plants characteristics and their expected flux of $\bar{\nu}_e$ can be found in the table 2.1. The distance of 53 km has been specifically chosen to maximize the disappearance probability of the $\bar{\nu}_e$. The data taking is scheduled to start early 2025.

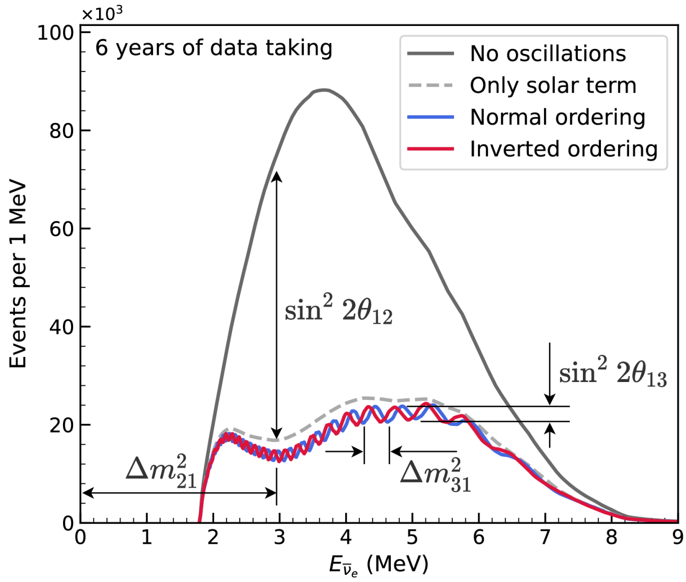


FIGURE 2.2 – Expected number of neutrinos event per MeV in JUNO after 6 years of data taking. The black curve shows the flux if there was no oscillation. The light gray curve shows the oscillation if only the solar terms are taken in account (θ_{12}, Δm^2_{21}). The blue and red curve shows the spectrum in the case of, respectively, NO and IO. The dependency of the oscillation to the different parameters are schematized by the double sided arrows. We can see the NMO sensitivity by looking at the fine phase shift between the red and the blue curve.

2.1 Neutrinos physics in JUNO

Even if the JUNO design detailed in section 2.2 was optimized for the measurement of the NMO, its large detection volume, excellent energy resolution and background level and understanding make it also an excellent detector to measure the flux coming from other neutrino sources. Thus the scientific program of JUNO extends way over reactor antineutrinos. The following section is an overview of the different physics topic JUNO will contribute in the coming years.

2.1.1 Reactor neutrino oscillation for NMO and precise measurements

Previous works [1, 3] shows that oscillation parameters and the NMO can be observed by looking at the $\bar{\nu}_e$ disappearance energy spectrum coming from medium baseline nuclear reactor. This disappearance probability can be expressed as [2] :

$$P(\bar{\nu}_e \to \bar{\nu}_e) = 1 - \sin^2 2\theta_{12} c_{13}^4 \sin^2 \frac{\Delta m^2_{21} L}{4E} - \sin^2 2\theta_{13} \left[c_{12}^2 \sin^2 \frac{\Delta m^2_{31} L}{4E} + s_{12}^2 \sin^2 \frac{\Delta m^2_{32} L}{4E} \right]$$

Where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$, E is the $\bar{\nu}_e$ energy and L is the baseline. We can see the sensitivity to the NMO in the dependency to Δm^2_{32} and Δm^2_{31} causing a phase shift of the spectrum as we can see in the figure 2.2. By carefully adjusting a theoretical spectrum to the data, one can extract the NMO and the oscillation parameters. The statistic procedure used to adjust the theoretical spectrum is reviewed in more details in the section 2.7. To reach the desired sensitivity, JUNO must meet multiple requirements but most notably:

1. An energy resolution of $3\%/\sqrt{E(\mathrm{MeV})}$ to be able to distinguish the fine structure of the fast oscillation.
2. An energy precision of 1% in order to not err on the location of the oscillation pattern.
3. A baseline of 53 ± 0.5 km to maximise the $\bar{\nu}_e$ oscillation probability.
4. At least $\approx$ 100,000 events to limit the spectrum distortion due to statistical uncertainties.

$\bar{\nu}_e$ flux coming from nuclear power plants

To get such high measurements precision, it is necessary to have a very good understanding of the sources characteristics. For its NMO and precise measurement studies, JUNO will observe the energy spectrum of neutrinos coming from the nuclear power plants Taishan and Yangjiang's cores, located at 53 km of the detector to maximise the disappearance probability of the $\bar{\nu}_e$.

Reactor	Power (GW_{th})	Baseline (km)	IBD Rate (day^{-1})	Relative Flux (%)
Taishan	9.2	52.71	15.1	32.1
Core 1	4.6	52.77	7.5	16.0
Core 2	4.6	52.64	7.6	16.1
Yangjiang	17.4	52.46	29.0	61.5
Core 1	2.9	52.74	4.8	10.1
Core 2	2.9	52.82	4.7	10.1
Core 3	2.9	52.41	4.8	10.3
Core 4	2.9	52.49	4.8	10.2
Core 5	2.9	52.11	4.9	10.4
Core 6	2.9	52.19	4.9	10.4
Daya Bay	17.4	215	3.0	6.4

TABLE 2.1 – Characteristics of the nuclear power plants observed by JUNO. The IBD rate are estimated from the baselines, the reactors full thermal power, selection efficiency and the current knowledge of the oscillation parameters

The $\bar{\nu}_e$ coming from reactors are emitted from β-decay of unstable fission fragments. The Taishan and Yangjiang reactors are Pressurised Water Reactor (PWR), the same type as Daya Bay. In those type of reactor more the 99.7 % and $\bar{\nu}_e$ are produced by the fissions of four fuel isotopes ^{235}U, ^{238}U, ^{239}Pu and ^{241}Pu. The neutrino flux per fission of each isotope is determined by the inversion of the measured β spectra of fission product [4–8] or by calculation using the nuclear databases [9, 10].

The neutrino flux coming from a reactor at a time t can be predicted using

$$\phi(E_\nu, t)_r = \frac{W_{th}(t)}{\sum_i f_i(t) e_i} \sum_i f_i(t) S_i(E_\nu) \tag{2.1}$$

where $W_{th}(t)$ is the thermal power of the reactor, $f_i(t)$ is the fraction fission of the ith isotope, e_i its thermal energy released in each fission and $S_i(e_\nu)$ the neutrino flux per fission for this isotope. Using this method, the flux uncertainty is expected to be of an order of 2-3 % [11].

In addition to those prediction, a satellite experiment named TAO[12] will be setup near the reactor core Taishan-1 to measure with an energy resolution of 2% at 1 MeV the neutrino flux coming from the core, more details can be found in section 2.4.1. It will help identifying unknown fine structure and give more insight on the $\bar{\nu}_e$ flux coming from this reactor.

One the open issue about reactor anti-neutrinos flux is the so-called neutrino anomaly [13], an unexpected surplus of neutrino emission in the spectra around 5 MeV. Multiples scientists are trying to explain this surplus by advanced recalculation of the nuclei model during beta decay [14, 15] but no consensus on this issue has been reached yet.

Background in the neutrinos reactor spectrum

Considering the close reactor neutrinos flux as the main signal, the signals that are considered as background are:

- — The geoneutrinos producing background in the $0.511 \sim 2.7$ MeV region.
- — The neutrinos coming from the other nuclear reactors around Earth.

In addition to all those physics signal, non-neutrinos signal that would mimic an IBD will also be present. It is composed of:

- — The signal coming from radioactive decay (α, γ, β) from natural radioactive isotopes in the material of the detector.
- — Cosmogenic event such as fast neutrons and activated isotopes induced by muons passing through the detector, most notably the spallation on ^{12}C.

All those events represent a non-negligeable part of the spectrum as shown in figure 2.3.

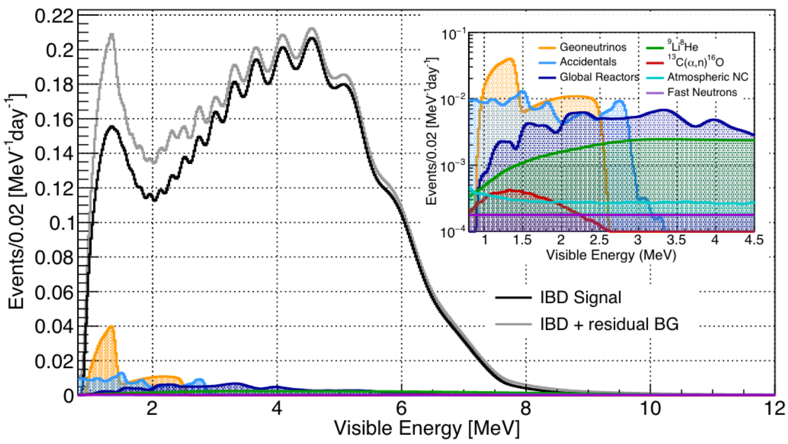


FIGURE 2.3 – Expected visible energy spectrum measured with the LPMT system with (grey) and without (black) backgrounds. The background amount for about 7% of the IBD candidate and are mostly localized below 3 MeV [11]

Identification of the mass ordering

To identify the mass ordering, we adjust the theoretical neutrino energy spectrum under the two hypothesis of NO and IO. Those give us two χ^2, respectively χ^2_{NO} and χ^2_{IO}. By computing the difference $\Delta\chi^2 = \chi^2_{NO} - \chi^2_{IO}$ we can determine the most probable mass ordering and the confidence interval: NO if $\Delta\chi^2 > 0$ and IO if $\Delta\chi^2 < 0$. Current studies shows that the expected sensitivity the mass ordering would be of 3.4σ after 6 years of data taking in nominal setup[2]. More detailed explanations about the procedure can be found in the section 2.7.

Precise measurement of the oscillations parameters

The oscillations parameters θ_{12}, θ_{13}, Δm^2_{21}, Δm^2_{31} are free parameters in the fit of the oscillation spectrum. The precision on those parameters have been estimated and are shown in table 2.2. Wee see that for θ_{12}, Δm^2_{21}, Δm^2_{31}, precision at 6 years is better than the reference precision by an order of magnitude [11]

	Central Value	PDG 2020	100 days	6 years	20 years
$\Delta m^2_{31} (\times 10^{-3} \mathrm{eV}^2)$	2.5283	$\pm$0.034 (1.3%)	$\pm$0.021 (0.8%)	$\pm$0.0047(0.2%)	$\pm$0.0029(0.1%)
$\Delta m^2_{21} (\times 10^{-3} \mathrm{eV}^2)$	7.53	$\pm$0.18 (2.4%)	$\pm$0.074 (1.0%)	$\pm$0.024(0.3%)	$\pm$0.017(0.2%)
$\sin^2 \theta_{12}$	0.307	$\pm$0.013 (4.2%)	$\pm$0.0058 (1.9%)	$\pm$0.0016(0.5%)	$\pm$0.0010(0.3%)
$\sin^2 \theta_{13}$	0.0218	$\pm$0.0007 (3.2%)	$\pm$0.010 (47.9%)	$\pm$0.0026(12.1%)	$\pm$0.0016(7.3%)

TABLE 2.2 – A summary of precision levels fir the oscillation parameters. The reference value (PDG 2020 [16]) is compared with 100 days, 6 years and 20 years of JUNO data taking.

2.1.2 Other physics

While the design of JUNO is tailored to measure $\bar{\nu}_e$ coming from nuclear reactor, JUNO will be able to detect neutrinos coming from other sources thus allowing for a wide range of physics studies as detailed in the table 2.3 and in the following sub-sections.

Research	Expected signal	Energy region	Major backgrounds
Reactor antineutrino	60 IBDs/day	0–12 MeV	Radioactivity, cosmic muon
Supernova burst	5000 IBDs at 10 kpc	0–80 MeV	Negligible
	2300 elastic scattering		
DSNB (w/o PSD)	2–4 IBDs/year	10–40 MeV	Atmospheric ν
Solar neutrino	hundreds per year for ^{8}B	0–16 MeV	Radioactivity
Atmospheric neutrino	hundreds per year	0.1–100 GeV	Negligible
Geoneutrino	$\approx$ 400 per year	0–3 MeV	Reactor ν

TABLE 2.3 – Detectable neutrino signal in JUNO and the expected signal rates and major background sources

Geoneutrinos

Geoneutrinos designate the antineutrinos coming from the decay of long-lived radioactive elements inside the Earth. The 1.8 MeV threshold necessary for the IBD makes it possible to measure geoneutrinos from ^{238}U and ^{232}Th decay chains. The studies of geoneutrinos can help refine the Earth crust models but is also necessary to characterise their signal, as they are a background to the mass ordering and oscillations parameters studies.

Atmospheric neutrinos

Atmospheric neutrinos are neutrinos originating from the decay of π and K particles that are produced in extensive air showers initiated by the interactions of cosmic rays with the Earth atmosphere. Earth is mostly transparent to neutrinos below the PeV energy, thus JUNO will be able to see neutrinos coming from all directions. Their baseline range is large (15km $\sim$ 13000km), they can have energy between 0.1 GeV and 10 TeV and will contain all neutrino and antineutrinos flavour. Their studies is complementary to the reactor antineutrinos and can help refine the constraints on the NMO [2].

Supernovae burst neutrinos

Neutrinos are crucial component during all stages of stellar collapse and explosion. Detection of neutrinos coming for core collapse supernovae will provide us important informations on the mechanisms at play in those events. Thanks to its 20 kt sensible volume, JUNO has excellent capabilities to detect all flavour of the $\mathcal{O}$(10 MeV) postshock neutrinos, and using neutrinos of the $\mathcal{O}$(1 MeV) will give informations about the pre-supernovae neutrinos. All those informations will allow to disentangle between the multiple hydro-dynamic models that are currently used to describe the different stage of core-collapse supernovae.

Diffuse supernovae neutrinos background

Core-collapse supernovae in our galaxy are rare events, but they frequently occur throughout the visible Universe sending burst of neutrinos in direction of the Earth. All those events contributes to a low background flux of low-energy neutrinos called the Diffuse Supernovae Neutrino Background (DSNB). Its flux and spectrum contains informations about the red-shift dependent supernovae rate, the average supernovae neutrino energy and the fraction of black-hole formation in core-collapse supernovae. Depending of the DSNB model, we can expect 2-4 IBD events per year in the energy range above the reactor $\bar{\nu}_e$ signal, which is competitive with the current Super-Kamiokande+Gadolinium phase [17].

Beyond standard model neutrinos interactions

JUNO will also be able to probe for beyond standard model neutrinos interactions. After the main physics topics have been accomplished, JUNO could be upgraded to probe for neutrinoless beta decay ($0\nu\beta\beta$). The detection of such event would give critical informations about the nature of neutrinos, is it a majorana or a dirac particle. JUNO will also be able to probe for neutrinos that would come for the decay or annihilation of Dark Matter inside the sun and neutrinos from putative primordial black hole. Through the unitary test of the mixing matrix, JUNO will be able to search for light sterile neutrinos. Thanks to JUNO sensitivity, multiple other exotic can be performed on neutrino related beyond standard model interactions.

2.2 The JUNO detector

The JUNO detector is a scintillator detector buried 693.35 meters under the ground (1800 meters water equivalent). It consist of Central Detector (CD), a water pool and a Top Tracker (TT) as showed in figure 2.4a. The CD is an acrylic vessel containing the 20 ktons of Liquid Scintillator (LS). It is supported by a stainless steel structure and is immersed in that water pool that is used as shielding from external radiation and as a cherenkov detector for the background. The top of the experiment is partially covered by the Top Tracker (TT), a plastic scintillator detector which is use to detect the atmospheric muons background and is acting as a veto detector.

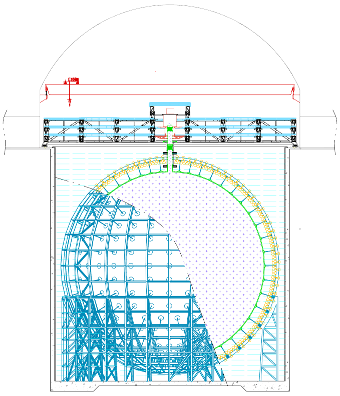
The top of the experiment also host the LS purification system, a water purification system, a ventilation system to get rid of the potential radon in the air. The CD is observed by two system of Photo-Multipliers Tubes (PMT). They are attached to the steel structure and their electronic readout is submersed near them. A third system of PMT is also installed on the structure but are facing outward of the CD, instrumenting the water to be cherenkov detector. The CD and the cherenkov detector are optically separated by Tyvek sheet. A chimney for LS filling and purification and for calibration operations connects the CD to the experimental hall from the top.

The CD has been dimensioned to meet the requirements presented in section 2.1.1:

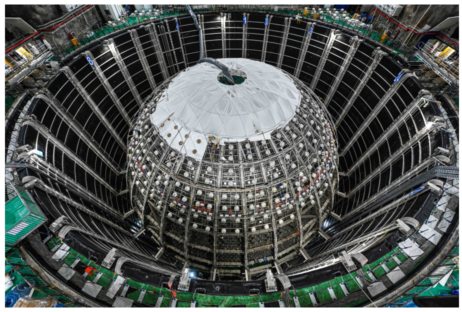
— Its 20 ktons monolithic LS provide a volume sizeable enough, in combination with the expected $\bar{\nu}_e$ flux, to reach the desired statistic in 6 years. Its monolithic nature also allow for a full containment of most of the events, preventing the energy loss in non-instrumented parts that would arise from a segmented detector.

— Its large overburden shield it from most of the atmospheric background that would pollute the signal.

— The localization of the experiment, chosen to maximize the disappearance with a 53km baseline and in a region that allow two nuclear power plant to be used as sources.



(A) Schematics view of the JUNO detector.



(B) Top down view of the JUNO detector under construction

FIGURE 2.4

This section cover in details the different components of the detector and the detection systems.

2.2.1 Detection principle

The CD will detect the neutrino and measure their energy mainly via an Inverse Beta Decay (IBD) interaction with proton mainly from the ^{12}C and H nucleus in the LS:

$$\bar{\nu}_e + p \to n + e^+$$

Kinematics calculation shows that this interaction has an energy threshold for the $\bar{\nu}_e$ of $(m_n + m_e - m_p) \approx 1.806$ MeV [18] where m_λ is the mass of the λ particle. This threshold make the experiment blind to very low energy neutrinos. The residual energy $E_\nu - 1.806$ MeV is be distributed as kinetic energy between the positron and the neutron. The energy of the emitted positron E_e is given by [18]

$$E_e = \frac{(E_\nu - \delta)(1 + \epsilon_\nu) + \epsilon_\nu \cos\theta \sqrt{(E_\nu - \delta)^2 + \kappa m_e^2}}{\kappa} \tag{2.2}$$

where $\kappa = (1 + \epsilon_\nu)^2 - \epsilon_\nu^2 \cos^2\theta \approx 1$, $\epsilon_\nu = \frac{E_\nu}{m_p} \ll 1$ and $\delta = \frac{m_n^2 - m_p^2 - m_e^2}{2m_p} \ll 1$. We can see from this equation that the positron energy is strongly correlated to the neutrino energy.

The positron and the neutron will then propagate in the detection medium, the Liquid Scintillator (LS), loosing their kinetic energy by exciting the molecule of the LS (more details in section 2.2.2). Once stopped, the positron will annihilate with an electron from the medium producing two 511 KeV gamma. Those gamma will themselves interact with the LS, exciting it before being absorbed by photoelectrical effect. The neutron will be captured by an hydrogen, emitting a 2.2 MeV gamma in the process. This gamma will also deposit its energy before being absorbed by the LS.

The scintillation photons have frequency in the UV and will propagate in the LS, being re-absorbed and re-emitted by compton effect before finally be captured by PMTs instrumenting the acrylic sphere. The analog signal of the PMTs digitized by the electronic is the signal of our experiment.

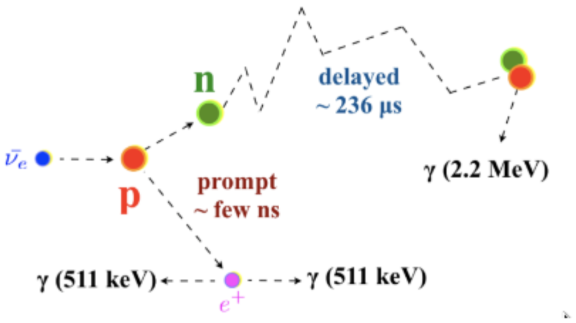


FIGURE 2.5 – Schematics of an IBD interaction in the central detector of JUNO

The signal produced by the positron is subsequently called the prompt signal, and the signal coming from the neutron the delayed signal. This naming convention come from the fact that the positron will deposit its energy rather quickly (few ns) where the neutron will take a bit more time ($\sim 236\ \mu$s).

2.2.2 Central Detector (CD)

The central detector, composed of 20 ktons of Liquid Scintillator (LS), is the main part of JUNO. The LS is contained in a spherical acrylic vessel supported by a stainless steel structure. The CD and its structural support are submerged in a cylindrical water pool of 43.5m diameter and 44m height. We're confident that the water pool provide sufficient buffer protection in every direction against the rock radioactivity.

Acrylic vessel

The acrylic vessel is a spherical vessel of inner diameter of 35.4 m and a thickness of 120 mm. It is assembled from 265 acrylic panels, thermo bonded together. The acrylic recipes has been carefully tuned with extensive R&D to ensure it does not include plasticizer and anti-UV material that would stop the scintillation photons. Those panels requires to be pure of radioactive materials to not cause background. Current setup where the acrylic panels are molded in cleanrooms of class 10000, let us reach a uranium and thorium contamination of <0.5 ppt. The molding and thermoforming processes is optimized to increase the assemblage transparency in water to >96%. The acrylic vessel is supported by a stainless steel structure via supporting node (fig 2.6). The structure and the nodes are designed to be resilient to natural catastrophic events such as earthquake and can support many times the effective load of the acrylic vessel.

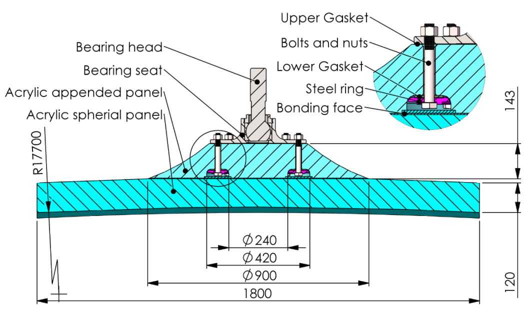


FIGURE 2.6 – Schematics of the supporting node for the acrylic vessel

Liquid scintillator

The Liquid Scintillator (LS) has a similar recipe as the one used in Daya Bay [19] but without gadolinium doping. It is made of three components, necessary to shift the wavelength of emitted photons to prevent their reabsorption:

1. The detection medium, the *linear alkylbenzene* (LAB). Selected because of its excellent transparency, high flash point, low chemical reactivity and good light yield. Accounting for $\sim$ 98% of the LS, it is the main component with which ionizing particles and gamma interact. Charged particles will collide with its electronic cloud transferring energy to the molecules, gamma will interact via compton effect with the electronic cloud before finally be absorbed via photoelectric effect.
2. The second component of the LS is the *2,5-diphenyloxazole* (PPO). A fraction of the excitation energy of the LAB is transferred to the PPO, mainly via non radiative process [20]. The PPO molecules de-excites in the same way, transferring their energy to the bis-MSB. The PPO makes for 1.5 % of the LS.
3. The last component is the *p-bis(o-methylstyryl)-benzene* (bis-MSB). Once excited by the PPO, it will emit photon with an average wavelength of $\sim$ 430 nm (full spectrum in figure 2.7) that can be detected by our photo-multipliers systems. It amount for $\sim$ 0.5% of the LS.

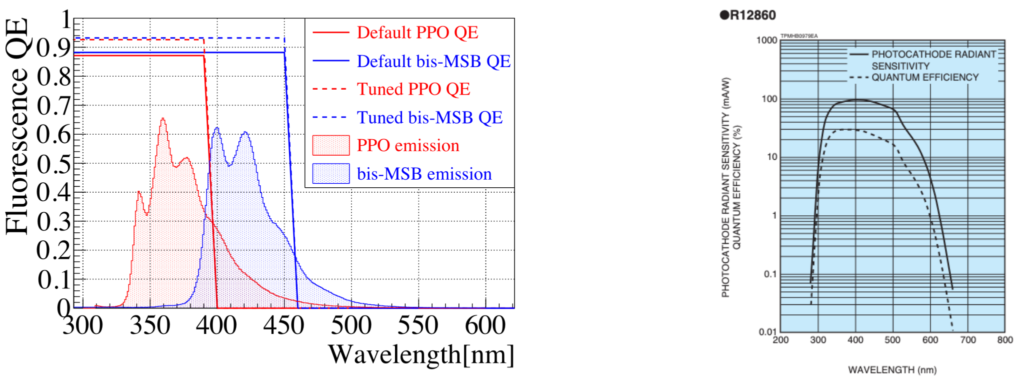


FIGURE 2.7 – **On the left:** Quantum efficiency (QE) and emission spectrum of the LAB and the bis-MSB [19]. **On the right:** Sensitivity of the Hamamatsu LPMT depending on the wavelength of the incident photons [21].

This formula has been optimized using dedicated studies with a Daya Bay detector [19, 22] to reach the requirements for the JUNO experiment:

— A light yield / MeV of the amount of 10^4 photons to maximize the statistic in the energy measurement.

— An attenuation length comparable to the size of the detector to prevent losing photons during their propagation in the LS. The final attenuation length is 25.8m [23] to compare with the CD diameter of 35.4m.

— Uranium/Thorium radiopurity to prevent background signal. The reactor neutrino program require a contamination fraction $F < 10^{-15}$ while the solar neutrino program require $F < 10^{-17}$.

The LS will frequently be purified and tested in the Online Scintillator Internal Radioactivity Investigation System (OSIRIS) [24] to ensure that the requirements are kept during the lifetime of the experiment, more details to be found in section 2.4.2.

Large Photo-Multipliers Tubes (LPMTs)

The scintillation light produced by the LS is then collected by Photo-Multipliers Tubes (PMT) that transform the incoming photon into an electric signal. As described in figure 2.8, the incident photons interact with the photocatode via photoelectric effect producing an electron called a Photo-Electron (PE). This PE is then focused on the dynodes where the high voltage will allow it to be multiplied. After multiple amplification the resulting charge - in coulomb [C] - is collected by the anode and the resulting electric signal can be digitalized by the readout electronics from which the charge and timing can be extracted.

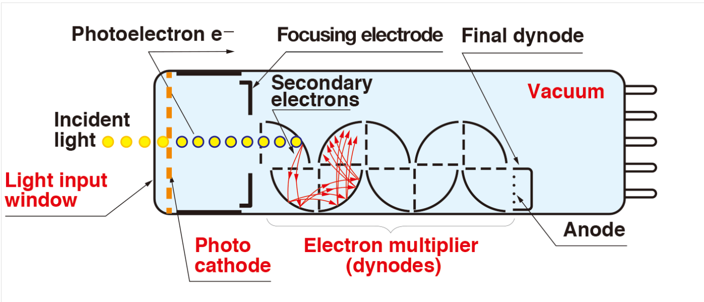


FIGURE 2.8 – Schematic of a PMT

The Large Photo-Multipliers Tubes (LPMT), used in the central detector and in the water pool, are 20-inch (50.8 cm) radius PMTs. $\sim$ 5000 dynode-PMTs [21] were produced by the Hamamatsu© company and $\sim$ 15000 Micro-Channel Plate (MCP) [25] by the NNVT© company. This system is the one responsible for the energy measurement with a energy resolution of $3\%/\sqrt{E}$, resolution necessary for the mass ordering measurement. To reach this precision, the system is composed of 17612 PMTs quasi uniformly distributed over the detector for a coverage of 75.2% reaching $\sim$ 1800 PE/MeV or $\sim$ 2.3% resolution due to statistic, leaving $\sim$ 0.7% for the systematic uncertainties. They are located outside the acrylic sphere in the water pool facing the center of the detector. To maintain the resolution over the lifetime of the experiment, JUNO require a failure rate $<$ 1% over 6 years.

The LPMTs electronic are divided in two parts. One "near", located underwater, in proximity of the LPMT to reduce the cable length between the PMT and early electronic. A second one, outside of the detector that is responsible for higher level analysis before sending the data to the DAQ.

The light yield per MeV induce that a LPMT can collect between 1 and 1000 PE per event, a wide dynamic range, causing non linearity in the PMT response that need to be understood and calibrated, see section 2.3 for more details.

Small Photo-Multipliers Tubes (SPMTs)

The Small PMT (SPMTs) system is made of 3-inch (7.62 cm) PMTs. They will be used in the CD as a secondary detection system. Those 25600 SPMTs will observe the same events as the LPMTs, thus sharing the physics and detector systematics up until the photon conversion. With a detector coverage of 2.7%, this system will collect $\sim$ 43 PE/MeV for a final energy resolution of $\sim$ 17%. This resolution is not enough to measure the NMO, θ_{13}, Δm^2_{31} but will be sufficient to independently measure θ_{12} and Δm^2_{21}.

Due to the low PE rate, SPMTs will be running in photo-counting mode in the reactor range and thus will be insensitive to non-linearity effect. Using this property, the intrinsic charge non linearity of the LPMTs can be measured by comparing the PE count in the SPMTs and LPMTs [26]. Also, due to their smaller size and electronics, SPMTs have a better timing resolutions than the LPMTs. At higher energy range, like supernovae events, LPMTs will saturate where SPMTs due to their lower PE collection will to produce a reliable measure of the energy spectrum.

The Data Acquisition System (DAQ) is designed to support the event rate of IBD, background, dark noise and supplementary storage buffers are present in the LPMT electronics to withstand the event rate during supernovae burst.

2.2.3 Veto detector

The CD will be bathed in constant background noise coming from numerous sources : the radioactivity from surrounding rock and its own components or from the flux of cosmic muons. This background needs to be rejected to ensure the purity of the IBD spectrum. To prevent a big part of them, JUNO use two veto detector that will tag events as background before CD analysis.

Cherenkov in water pool

The Water Cherenkov Detector (WCD) is the instrumentation of the water buffer around the CD. When high speed charged particles will pass through the water, they will produced cherenkov photons. The light will be collected by 2400 MCP LPMTs installed on the outer surface of the CD structure. The muons veto strategy is based on a PMT multiplicity condition. WCD PMTs are grouped in ten zones: 5 in the top, 5 in the bottom. A veto is raised either when more than 19 PMTs are triggered in one zone or when two adjacent zones simultaneously trigger more than 13 PMTs. Using this trigger, we expect to reach a muon detection efficiency of 99.5% while keeping the noise at reasonable level.

Top tracker

The JUNO Top Tracker (TT) is a plastic scintillator detector located on the top of the experiment (see figure 2.9). Made from plastic scintillator from OPERA [27] layered horizontally in 3 layers on the top of the detector, the TT will be able to detect incoming atmospheric muons. With its coverage, about 1/3 of the of all atmospheric muons that passing through the CD will also pass through the 3 layer of the detector. While it does not cover the majority of the CD, the TT is particularly effective to detect muons coming through the filling chimney region which might present difficulties from the other subsystems in some classes of events.

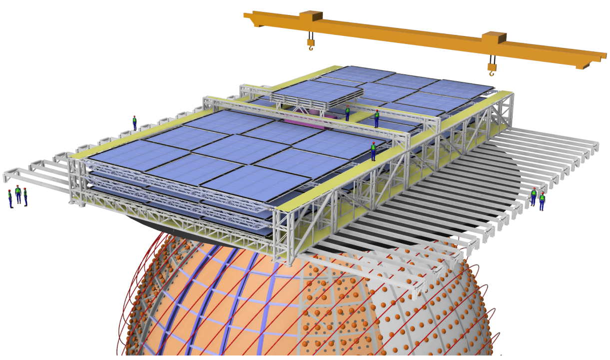


FIGURE 2.9 – The JUNO top tracker

2.3 Calibration strategy

The calibration is a crucial part of the JUNO experiment. Because we are looking at civil reactor neutrino it might be impossible to run measurement without signal, it would need to shut down every reactor from the Taishan and Yangjiang power plants which is realistically impossible. Because of this continuous rate, low frequency signal event, we need high frequency, recognisable sources in the energy range of interest : [0-12] MeV for the positron signal and 2.2 MeV for the neutron capture. It is expected that the CD response will be different depending on the type of particle, due to the interaction with LS, the position on the event and the optical response of the acrylic sphere (see section 2.6). We also expect a non-linear energy response of the CD due to the LS properties [19] but also due to the saturation of the LPMTs system when collecting a large amount of PE [26].

2.3.1 Energy scale calibration

While electrons and positrons sources would be ideal, for a large LS detector thin-walled electrons or positrons sources could lead to leakage of radionucleides causing radioactive contamination. Instead, we consider gamma sources in the range of the prompt energy of IBDs. The sources are reported in table 2.4.

Sources / Processes	Type	Radiation
^{137}Cs	γ	0.0662 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 (12C*)
^{241}Am-^{13}C	n, γ	neutron + 6.13 MeV (16O*)
$(n, \gamma)p$	γ	2.22 MeV
$(n, \gamma)^{12}$C	γ	4.94 MeV or 3.68 + 1.26 MeV

TABLE 2.4 – List of sources and their process considered for the energy scale calibration

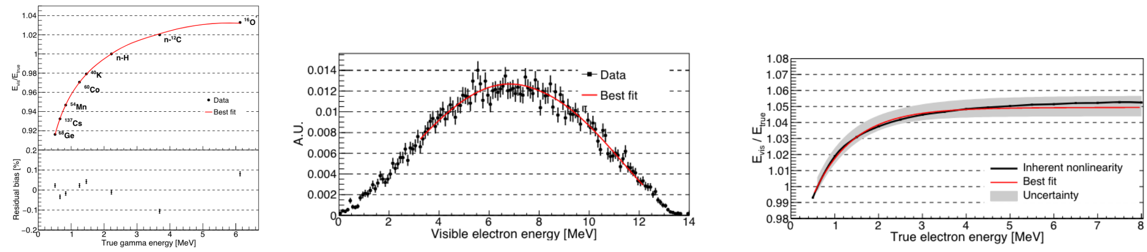
For the ^{68}Ge source, it will decay in ^{68}Ga via electron capture, which will itself β^+ decay into ^{68}Zn. The positrons will be absorbed by the enclosure so only the annihilation gamma will be released. In addition, (α, n) sources like ^{241}Am-Be and ^{241}Am-^{13}C are used to provide both high energy gamma and neutrons, which will later be captured in the LS producing the 2.2 MeV gamma.

From this calibration we call E_{vis} the "visible energy" that is reconstructed by our current algorithms and we compare it to the true energy deposited by the calibration source. The results shown in figure 2.10 show the expected response of the detector from calibration sources. The non-linearity is clearly visible from the E_{vis}/E_{true} shape. See [28] for more details.

2.3.2 Calibration system

The non-uniformity due to the event position in the detector (more details in section 2.6) will be studied using multiples systems that are schematized in figure 2.11. They allow to position sources at different location in the CD.

— For a one-dimension vertical calibration, the Automatic Calibration Unit (ACU) will be able to deploy multiple radioactive sources or a pulse laser diffuser ball along the central axis of the CD through the top chimney. The source position precision is less than 1cm.



(A) Gamma non-linearity (B) Boron non-spectrum (C) Electron non-linearity

FIGURE 2.10 – Fitted and simulated non linearity of gamma, electron sources and from the ^{12}B spectrum. Black points are simulated data. Red curves are the best fits

— For off-axis calibration, a calibration source attached to a Cable Loop System (CLS) can be moved on a vertical half-plane by adjusting the length of two connection cable. Two set of CSL will be deployed to provide a 79% effective coverage of a vertical plane.
— A Guiding Tube (GT) will surround the CD to calibrate the non-uniformity of the response at the edge of the detector
— A Remotely Operated under-LS Vehicle (ROV) can be deployed to desired location inside LS for a more precise and comprehensive calibration. The ROV will also be equipped with a camera for inspection of the CD.

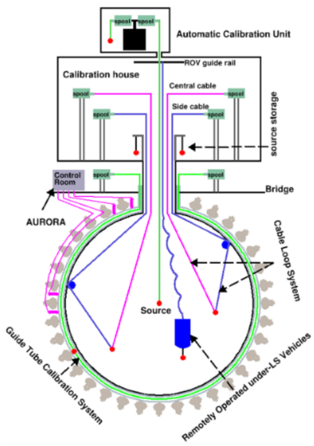


FIGURE 2.11 – Overview of the calibration system

The preliminary calibration program is depicted in table 2.5.

2.4 Satellite detectors

As introduced in section 2.1.1 and section 2.2.2, the precise knowledge and understanding of the detector condition is crucial for the measurements of the NMO and oscillation parameters. Thus two satellite detectors will be setup to monitor the experiment condition. TAO to monitor and understand the $\bar{\nu}_e$ flux and spectrum coming from the nuclear reactor and OSIRIS to monitor the LS response.

Program	Purpose	System	Duration [min]
Weekly calibration	Neutron (Am-C)	ACU	63
	Laser	ACU	78
Monthly calibration	Neutron (Am-C)	ACU	120
	Laser	ACU	147
	Neutron (Am-C)	CLS	333
	Neutron (Am-C)	GT	73
Comprehensive calibration	Neutron (Am-C)	ACU, CLS and GT	1942
	Neutron (Am-Be)	ACU	75
	Laser	ACU	391
	^{68}Ge	ACU	75
	^{137}Cs	ACU	75
	^{54}Mn	ACU	75
	^{60}Co	ACU	75
	^{40}K	ACU	158

TABLE 2.5 – Calibration program of the JUNO experiment

2.4.1 TAO

The Taishan Antineutrino Observatory (TAO) [12, 29] is a ton-level gadolinium doped liquid scintillator detector that will be located near the Taishan-1 reactor. It aim to measure the $\bar{\nu}_e$ spectrum at very low distance (< 30m) from the reactor to measure a quasi-unoscillated spectrum. TAO also aim to provide a major contribution to the so-called reactor anomaly [13]. Its requirement are to the level of 2 % energy resolution at 1 MeV.

Detector

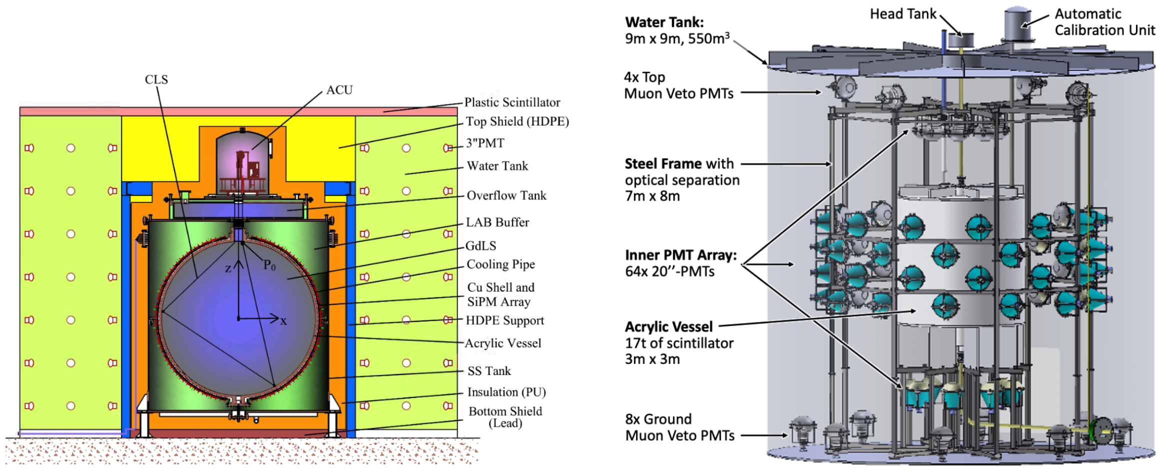
The TAO detector is close, in concept, to the CD of JUNO. It is composed of an acrylic vessel containing 2.8 tons of gadolinium-loaded LS instrumented by an array of silicon photomultipliers (SiPM) reaching a 95% coverage. To efficiently reduce the dark count of those sensors, the detector is cooled to -50 °C. The $\bar{\nu}_e$ will interact with the LS via IBD, producing scintillation light, that will be detected by the SiPMs. From this signal the $\bar{\nu}_e$ energy and the full spectrum reconstructed. This spectrum will then be used by JUNO to calibrate the unoscillated spectrum, most notably the fission product fraction that impact the rate and shape of the spectrum. A schema of the detector is presented in figure 2.12a.

2.4.2 OSIRIS

The Online Scintillator Internal Radioactivity Investigation System (OSIRIS) [24] is an ultralow background, 20 m^3 LS detector that will be located in JUNO cavern. It aim to monitor the radioactive contamination, purity and overall response of the LS before it is injected in JUNO. OSIRIS will be located at the end of the purification chain of JUNO, monitoring that the purified LS meet the JUNO requirements. The setup is optimized to detect the fast coincidences decay of $^{214}\mathrm{Bi} -^{214}\mathrm{Po}$ and $^{212}\mathrm{Bi} -^{212}\mathrm{Po}$, indicators of the decay chains of U and Th respectively.

Detector

OSIRIS is composed of an acrylic vessel that will contains 17t of LS. The LS is instrumented by a PMT array of 64 20 inch PMTs on the top and the side of the vessel. To reach the necessary background level required by the LS purity measurements, in addition to being 700m underground in the experiment cavern, the acrylic vessel is immersed in a tank of ultra pure water. The water is itself instrumented by another array of 20 inch PMTs, acting as muon veto. A schema of the detector is presented in figure 2.12b.



(A) Schematic of the TAO satellite detector

(B) Schematic of the OSIRIS satellite detector

FIGURE 2.12

2.5 Software

The simulation, reconstruction and analysis algorithms are all packaged in the JUNO software, subsequently called the software. It is composed of multiple components integrated in the SNiPER [30] framework:

- Various primary particles simulators for the different kind of events, background and calibration sources.
- A Geant4 [31–33] Monte Carlo (MC) simulation containing the detectors geometries, a custom optical model for the LS and the supporting structures of the detectors. The Geant4 simulation integrate all relevant physics process for JUNO, validated by the collaboration. This step of the simulation is commonly called *Detsim* and compute up to the production of photo-electrons in the PMTs. The optics properties of the different materials and detector components have been measured beforehand to be used to define the material and surfaces in the simulation.
- An electronic simulation, simulating the response waveform of the PMTs, tracking it through the digitization process, accounting for effects such as non-linearity, dark noise, Time Transit Spread (TTS), pre-pulsing, after-pulsing and ringing if the waveform. It's also the step handling the event triggers and mixing. This step is commonly referenced as *Elecsim*.
- A waveform reconstruction where the digitized waveform are filtered to remove high-frequency white noise and then deconvoluted to yield time and charge informations of the photons hits on the PMTs. This step is commonly referenced as *Calib*.
- The charge and time informations are used by reconstruction algorithms to reconstruct the interaction vertex and the deposited energy. This step is commonly reported as *Reco*. See section 2.6 for more details on the reconstruction.
- Once the singular events are reconstructed, they go through event pairing and classification to select IBD events. This step is named Event Classification.

— The purified signal is then analysed by the analysis framework which depend of the physics topic of interest.

The steps Reco and Event Classification are divided into two category of algorithm. Fast but less accurate algorithms that are running during the data taking designated as the *Online* algorithms. Those algorithm are used to take the decision to save the event on tape or to throw it away. More accurate algorithms that run on batch of events designated *Offline* algorithms. They are used for the physics analysis. The Offline Reco will be one of the main topic of interest for this thesis.

2.6 State of the art of the Offline IBD reconstruction in JUNO

The main reconstruction method currently run in JUNO is a data-driven method based on a likelihood maximization [34, 35] using only the LPMTs. The first step is to reconstruct the interaction vertex from which the energy reconstruction is dependent. It is also necessary for event pairing and classification.

2.6.1 Interaction vertex reconstruction

To start the likelihood maximization, a rough estimation of the vertex and of the event timing is needed. We start by estimating the vertex position using a charge based algorithm.

Charge based algorithm

The charge-based algorithm is basically base on the charge-weighted average of the PMT position.

$$\vec{r}_{cb} = a \cdot \frac{\sum_i q_i \cdot \vec{r}_i}{\sum_i q_i} \tag{2.3}$$

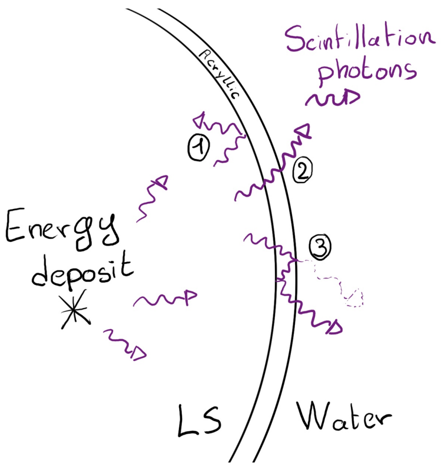
Where q_i is the reconstructed charge of the pulse of the ith PMT and $\vec{r}_i$ is its position. $\vec{r}_0$ is the reconstructed interaction position. a is a scale factor introduced because a weighted average over a 3D sphere is inherently biased. Using calibration we can estimate $a \approx 1.3$ [36]. The results in figure 2.13b shows that the reconstruction is biased from around 15m and further. This is due to the phenomena called "total reflection area" or TR Area.

As depicted in the figure 2.13a the optical photons, given that they have a sufficiently large incidence angle, can be deviated of their trajectories when passing through the interfaces LS-acrylic and water-acrylic due to the optical index difference. This cause photons to be lost or to be detected by PMT further than anticipated if we consider their rectilinear trajectories. This cause the charge barycenter the be located closer to the center than the event really is.

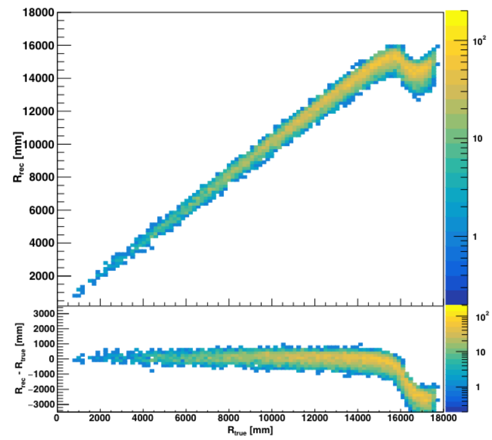
It is to be noted that charge based algorithm, in addition to be biased near the edge of the detector, does not provide any information about the timing of the event. Therefore, a time based algorithm needs to be introduced to provide initial values.

Time based algorithm

The time based algorithm use the distribution of the time of flight corrections Δt (Eq 2.4) of an event to reconstruct its vertex and t_0. It follow the following iterations:



(A) Illustration of the different optical photons reflection scenarios. **1** is the reflection of the photon at the interface LS-acrylic or acrylic-water. **2** is the transmission of the photons through the interfaces. **3** is the conduction of the photon in the acrylic.



(B) Heatmap of R_{rec} and $R_{rec} - R_{true}$ as a function of R_{true} for 4MeV prompt signals uniformly distributed in the detector calculated by the charge based algorithm

FIGURE 2.13

1. Use the charge based algorithm to get an initial vertex to start the iteration.
2. Calculate the time of flight correction for the ith PMT using

$$\Delta t_i(j) = t_i - \text{tof}_i(j) \tag{2.4}$$

   where j is the iteration step, t_i is the timing of the ith PMT, and tof_i is the time-of-flight of the photon considering an rectilinear trajectory and an effective velocity in the LS and water (see [36] for detailed description of this effective velocity). Plot the Δt distribution and label the peak position as Δt^{peak} (see fig 2.14a).
3. Calculate a correction vector $\vec{\delta}[\vec{r}(j)]$ as

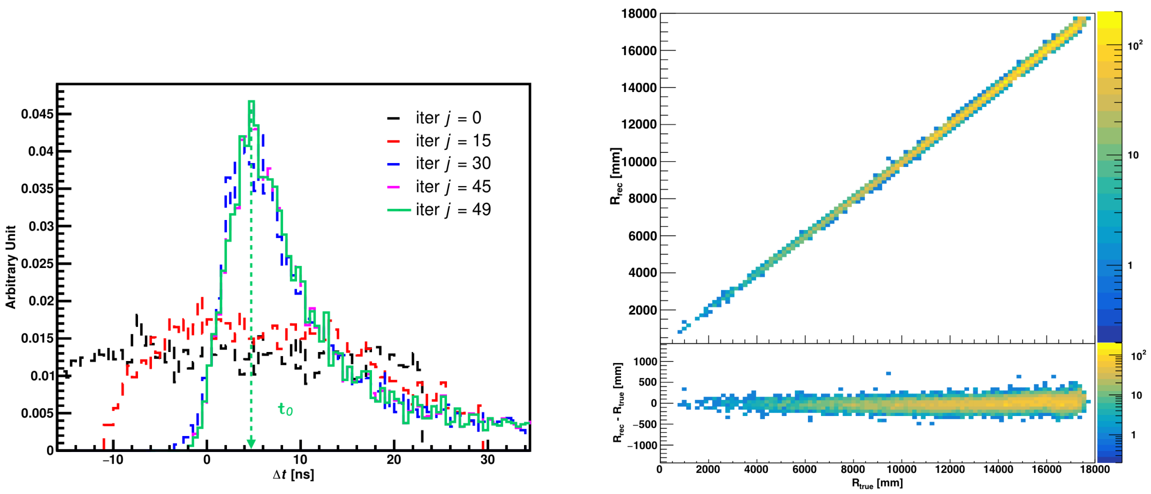
$$\vec{\delta}[\vec{r}(j)] = \frac{\sum_i \left(\frac{\Delta t(j) - \Delta t^{\text{peak}}(j)}{\text{tof}_i(j)} \right) \cdot (\vec{r}_0(j) - \vec{r}_i)}{N^{\text{peak}}(j)} \tag{2.5}$$

   where $\vec{r}_0$ is the vertex position at the beginning of this iteration, $\vec{r}_i$ is the position of the ith PMT. To minimize the effect of scattering, dark noise and reflection, only the pulse happening in a time window (-10 ns, +5 ns) around Δt^{peak} are considered. N^{peak} is the number of PE collected in this time-window.
4. if $\vec{\delta}[\vec{r}(j)] < 1\text{mm}$ or $j \geq 100$, stop the iteration. Otherwise $\vec{r}_0(j+1) = \vec{r}_0(j) + \vec{\delta}[\vec{r}(j)]$ and go to step 2.

However because the earliest arrival time is used, t_i is related to the number photoelectrons N_i^{pe} detected by the PMT [37–39]. To reduce bias in the vertex reconstruction, the following equation is used to correct t_i into t'_i:

$$t'_i = t_i - p_0/\sqrt{N_i^{\text{pe}}} - p_1 - p_2/N_i^{\text{pe}} \tag{2.6}$$

The parameters (p_0, p_1, p_2) were optimized to (9.42, 0.74, -4.60) for Hamamatsu PMTs and (41.31, -12.04, -20.02) for NNVT PMTs [36]. The results presented in figure 2.14b shows that the time based



(A) Δt distribution at different iterations step j

(B) Heatmap of R_{rec} and $R_{rec} - R_{true}$ as a function of R_{true} for 4MeV prompt signals uniformly distributed in the detector calculated by the time based algorithm

FIGURE 2.14

algorithm provide a more accurate vertex and is unbiased even in the TR area. This results $(\vec{r}_0, t_0)$ is used as initial value for the likelihood algorithm.

Time likelihood algorithm

The time likelihood algorithm use the residual time expressed as follow

$$t^i_{\mathrm{res}}(\vec{r}_0, t_0) = t_i - \mathrm{tof}_i - t_0 \tag{2.7}$$

In a first order approximation, the scintillator time response Probability Density Function (PDF) can be described as the emission time profile of the scintillation photons, the Time Transit Spread (TTS) and the dark noise of the PMTs. The emission time profile $f(t_{\mathrm{res}})$ is described like

$$f(t_{\mathrm{res}}) = \sum_k \frac{\rho_k}{\tau_k} e^{\frac{-t_{\mathrm{res}}}{\tau_k}}, \sum_k \rho_k = 1 \tag{2.8}$$

as the sum of the k component that emit light in the LS each one characterised by it's decay time τ_k and intensity fraction ρ_k. The TTS component is expressed as a gaussian convolution

$$g(t_{\mathrm{res}}) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(t_{\mathrm{res}}-\nu)^2}{2\sigma^2}} \cdot f(t_{\mathrm{res}}) \tag{2.9}$$

where σ is the TTS of PMTs and ν is the average transit time. The dark noise is not correlated with any physical events and considered as constant rate over the time window considered T. By normalizing the dark noise probability $\epsilon(t_{\mathrm{res}})$ as $\int_T \epsilon(t_{\mathrm{res}}) dt_{\mathrm{res}} = \epsilon_{\mathrm{dn}}$, it can be integrated in the PDF as

$$p(t_{\mathrm{res}}) = (1 - \epsilon_{\mathrm{dn}}) \cdot g(t_{\mathrm{res}}) + \epsilon(t_{\mathrm{res}}) \tag{2.10}$$

The distribution of the residual time t_{res} of an event can then be compared to $p(t_{\mathrm{res}})$ and the best fitting vertex $\vec{r}_0$ and t_0 can be chosen by minimizing

$$\mathcal{L}(\vec{r}_0, t_0) = -\ln\left(\prod_i p(t^i_{\text{res}})\right) \tag{2.11}$$

The parameter of Eq. 2.10 can be measured experimentally. The results shown in figure 2.15 used PDF from monte carlo simulation. The results shows that $R_{rec} - R_{true}$ is biased depending on the energy. While this could be corrected using calibration, another algorithm based on charge likelihood was developed to correct this problem.

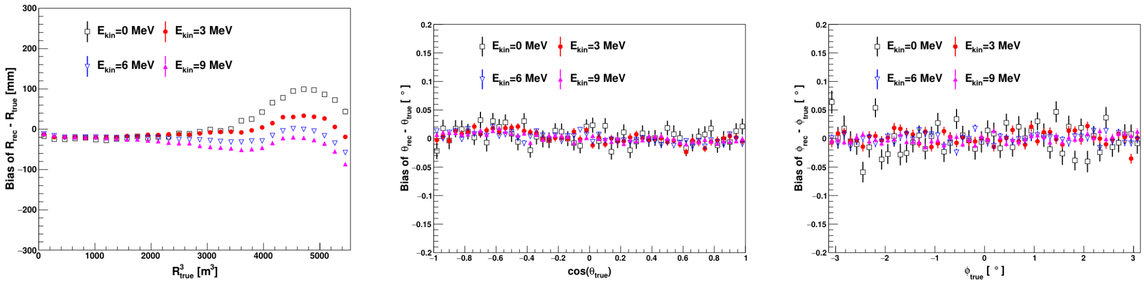


FIGURE 2.15 – Bias of the reconstructed radius R (left), θ (middle) and ϕ (right) for multiple energies by the time likelihood algorithm

Charge likelihood algorithm

Similarly to the time likelihood algorithms that use a timing PDF, the charge likelihood algorithm use a PE PDF for each PMT depending on the energy and position of the event. With $\mu(\vec{r}_0, E)$ the mean expected number of PE detected by each PMT, the probability to observe N_{pe} in a PMT follow a Poisson distribution. Thus

- The probability to observe no hit ($N_{pe} = 0$) in the jth PMT is $P^j_{nohit}(\vec{r}_0, E) = e^{-\mu_j}$
- The probability to observe $N_{pe} \neq 0$ in the ith PMT is $P^i_{hit}(\vec{r}_0, E) = \frac{\mu^{N^i_{pe}} e^{-\mu_i}}{N^i_{pe}!}$

Therefore, the probability to observe a specific hit pattern can be expressed as

$$P(\vec{r}_0, E) = \prod_j P^j_{nohit}(\vec{r}_0, E) \cdot \prod_i P^i_{hit}(\vec{r}_0, E) \tag{2.12}$$

The best fit values of $\vec{R}_0$ and E can then be calculated by minimizing the negative log-likelihood

$$\mathcal{L}(\vec{r}_0, E) = -\ln(P(\vec{r}_0, E)) \tag{2.13}$$

In principle, $\mu_i(\vec{r}_0, E)$ could be expressed

$$\mu_i(\vec{r}_0, E) = Y \cdot \frac{\Omega(\vec{r}_0, r_i)}{4\pi} \cdot \epsilon_i \cdot f(\theta_i) \cdot e^{-\sum_m \frac{d_m}{\zeta_m}} \cdot E + \delta_i \tag{2.14}$$

where Y is the energy scale factor, $\Omega(\vec{r}_0, r_i)$ is the solid angle of the ith PMT, ϵ_i is its detection efficiency, $f(\theta_i)$ its angular response, ζ_m is the attenuation length in the materials and δ_i the expected number of dark noise.

However Eq. 2.14 assume that the scintillation light yield is linear with energy and describe poorly the contribution of indirect light, shadow effect due to the supporting structure and the total reflection effects. The solution is to use data driven methods to produce the pdf by using the calibrations sources and position described in section 2.3. In the results presented in figures 2.16, the PDF was produced using MC simulation and 29 specific calibrations position [36] along the Z-axis of the detector. We see that the charge likelihood algorithm show little bias in the TR area and a better

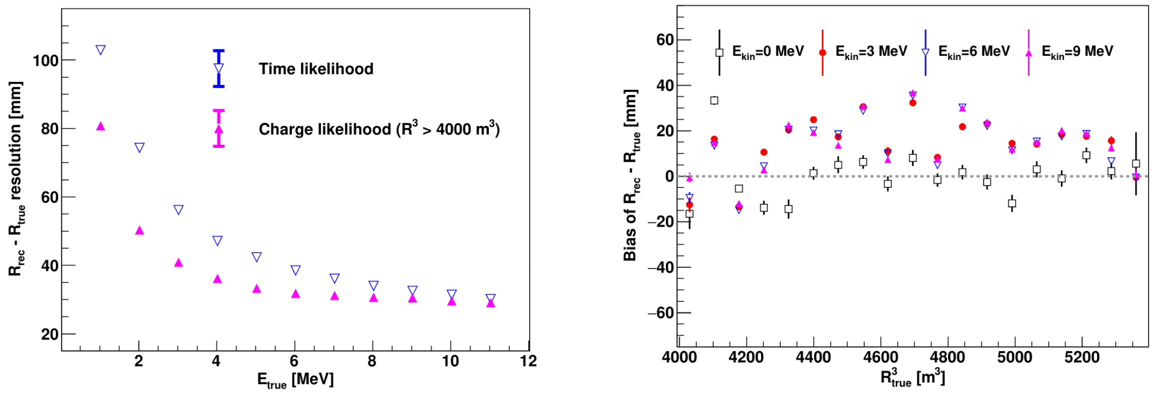


FIGURE 2.16 – **On the left:** Resolution of the reconstructed R as a function of the energy in the TR area ($R^3 > 4000\text{m}^3 \equiv R > 16m$) by the charge and time likelihood algorithms. **On the right:** Bias of the reconstructed R in the TR area for different energies by the charge likelihood algorithm

resolution than the time likelihood. The figure 2.17 shows the radial resolution of the different algorithm presented for this section, we can see the refinement at each step and that the charge likelihood yield the best results.

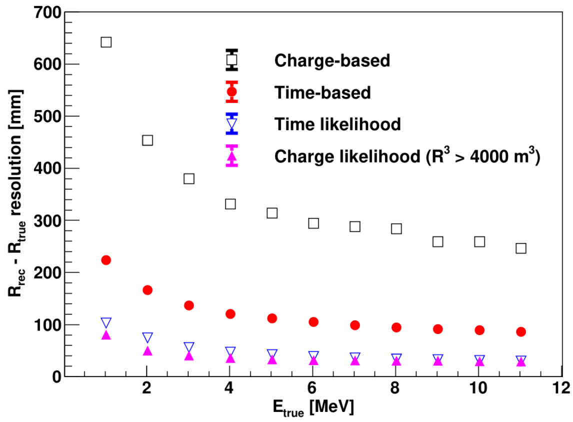
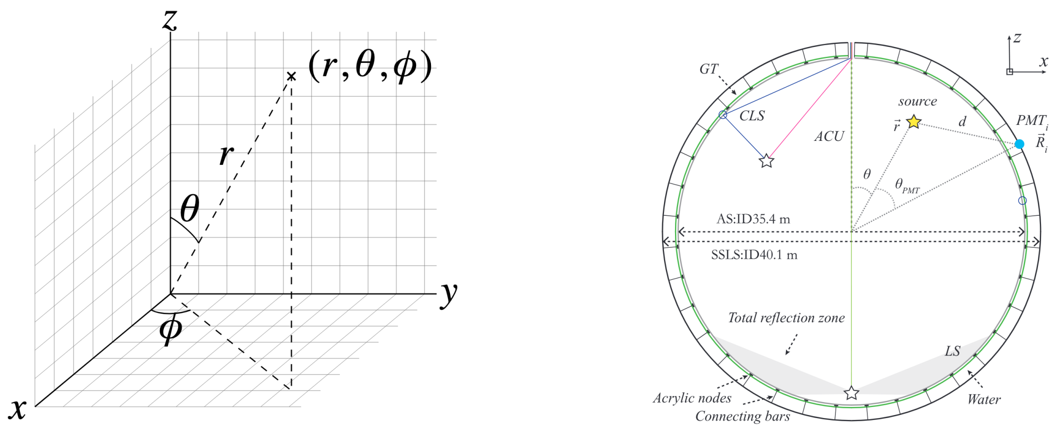


FIGURE 2.17 – Radial resolution of the different vertex reconstruction algorithms as a function of the energy

The charge based likelihood algorithms already give use some information on the energy as Eq. 2.13 is minimized but the energy can be further refined as shown in the next section.

2.6.2 Energy reconstruction

As explained in section 2.1.1, energy resolution is crucial for the NMO and oscillation parameters measurements. Thus the energy reconstruction algorithm should take into consideration as much detector effect as possible. The following method is a data driven method based on calibration samples inspired by the charge likelihood algorithm described above [40].



(A) Spherical coordinate system used in JUNO for reconstruction

(B) Definition of the variables used in the energy reconstruction

FIGURE 2.18

Charge estimation

The most important element in the energy reconstruction is $\mu_i(\vec{r}_0, E)$ described in Eq. 2.14. For realistic cases, we also need to take into account the electronics effect that were omitted in the previous section. Those effect will cause a charge smearing due to the uncertainties in the N_{pe} reconstruction. Thus we define $\hat{\mu}^L(\vec{r}_0, E)$ which is the expected N_{pe}/E in the whole detector for an event with visible energy E_{vis} and position $\vec{r}_0$. The position of the event and PMTs are now defined using $(r, \theta, \theta_{pmt})$ as defined in figure 2.18b.

$$\hat{\mu}(r, \theta, \theta_{pmt}, E_{vis}) = \frac{1}{E_{vis}} \frac{1}{M} \sum_{i}^{M} \frac{\frac{\bar{q}_i}{\hat{Q}_i} - \mu_i^D}{\mathrm{DE}_i}, \; \mu_i^D = \mathrm{DNR}_i \cdot L \tag{2.15}$$

where i runs over the PMTs with the same θ_{pmt}, DE_i is the detection efficiency of the ith PMT. μ_i^D is the expected number of dark noise photoelectrons in the time window L. The time window have been optimized to $L = 280$ ns [40]. $\bar{q}_i$ is the average recorded photoelectrons in the time window and $\hat{Q}_i$ is the expected average charge for 1 photoelectron. The N_{pe} map is constructed following the procedure described in [35].

Time estimation

The second important observable is the hit time of photons that was previously defined in Eq. 2.7. It is here refined as

$$t_r = t_h - \mathrm{tof} - t_0 = t_{LS} + t_{TT} \tag{2.16}$$

where t_h is the time of hit, t_{LS} is the scintillation time and t_{TT} the transit time of PMTs that is described by a gaussian

$$t_{TT} = \mathcal{N}(\overline{\mu_{TT} + t_d}, \sigma_{TT}) \tag{2.17}$$

where μ_{TT} is the mean transit time in PMTs, σ_{TT} is the Transit Time Spread (TTS) of the PMTs and t_d is the delay time in the electronics. The effective refraction index of the LS is also corrected to take into account the propagation distance in the detector.

The timing PDF $P_T(t_r|r, d, \mu_l, \mu_d, k)$ can now be generated using calibration sources [40]. This PDF describe the probability that the residual time of the first photon hit is in $[t_r, t_r + \delta]$ with r the radius of the event vertex, $d = |\vec{r} - \vec{r}_{PMT}|$ the propagation distance, μ_l and μ_d the expected number of PE and dark noise in the electronic reading window and k is the detected number of PE.

Now let denote $f(t, r, d)$ the probability density function of "photoelectron hit a time t" for an event happening at r where the photons traveled the distance d in the LS

$$F(t, r, d) = \int_t^L f(t', r, d)dt' \tag{2.18}$$

Based on the PDF for one photon $k = 1$, one can define

$$P_T^l(t|k = n) = I_n^l[f_l(t)F_l^{n-1}(t)] \tag{2.19}$$

where the indicator l means that the photons comes from the LS and I_n^l a normalisation factor. To this pdf we add the probability to have photons coming from the dark noise indicated by the indicator d using

$$f_d(t) = 1/L,\ F_d(t) = 1 - \frac{t}{L} \tag{2.20}$$

and so for the case where only one photon is detected by the PMT ($k = 1$)

$$P_T(t|\mu_l, \mu_d, k = 1) = I_1[P(1, \mu_l)P(0, \mu_d)f_l(t) + P(0, \mu_l)P(1, \mu_d)f_d(t)] \tag{2.21}$$

where $P(k_\alpha, \mu_\alpha)$ is the Poisson probability to detect k_α PE from $\alpha \in \{l, d\}$ with the condition $k_l + k_d = k$.

Now that we have the individual timing and charge probability we can construct the charge likelihood referred as QMLE:

$$\mathcal{L}(q_1, q_2, ..., q_N|\vec{r}, E_{vis}) = \prod_{j\in\text{unfired}} e^{-\mu_j} \prod_{i\in\text{fired}} \left(\sum_{k=1} P_Q(q_i|k) \cdot P(k, \mu_i) \right) \tag{2.22}$$

where $\mu_i = E_{vis}\hat{\mu}_i^L + \mu_i^D$ and $P(k, \mu_i)$ is the Poisson probability of observing k PE. $P_Q(q_i|k)$ is the charge pdf for k PE. And we can also construct the time likelihood referred as TMLE:

$$\mathcal{L}(t_{1,r}, t_{2,r}, ..., t_{N,r}|\vec{r}, t_0) = \prod_{i\in\text{hit}} \frac{\sum_{k=1}^K P_T(t_{i,r}|r, d, \mu_i^l, \mu_i^d, k) \cdot P(k, \mu_i^l + \mu_i^d)}{\sum_{k=1}^K P(k, \mu_i^l + \mu_i^d)} \tag{2.23}$$

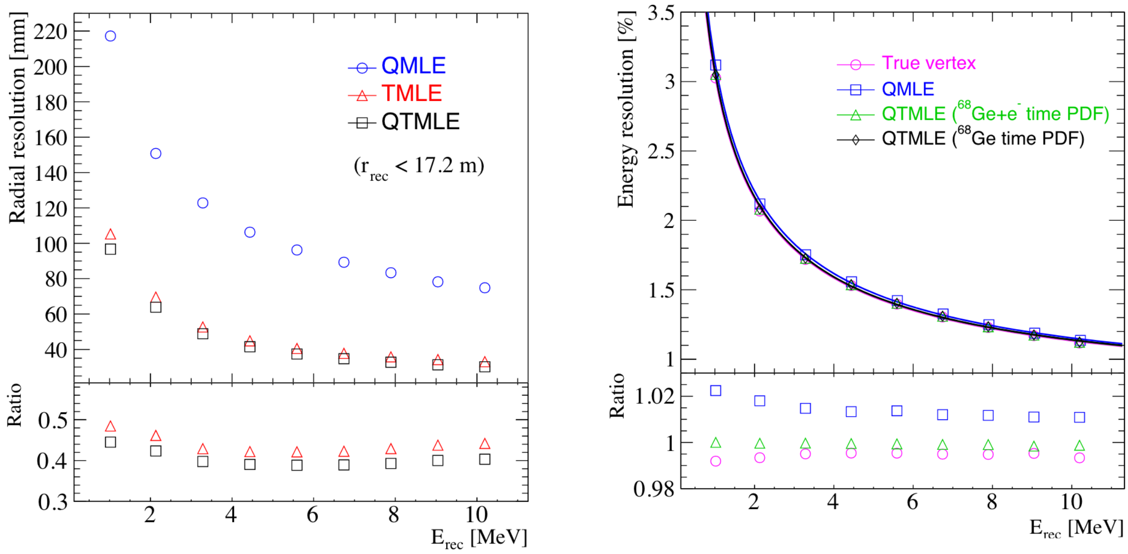
where K is cut to 20 PE and hit is the set of hits satisfying $-100 < t_{i,r} < 500$ ns.

Merging those two likelihood give the charge-time likelihood QTMLE

$$\mathcal{L}(q_1, q_2, ..., q_N; t_{1,r}, t_{2,r}, ..., t_{N,r}|\vec{r}, t_0, E_{vis}) = \mathcal{L}(q_1, q_2, ..., q_N|\vec{r}, E_{vis}) \cdot \mathcal{L}(t_{1,r}, t_{2,r}, ..., t_{N,r}|\vec{r}, t_0) \tag{2.24}$$

The radial and energy resolutions of the different likelihood are presented in figure 2.19 (from [40]). We can see the improvement of adding the time information to the vertex reconstruction and that an increase in vertex precision can bring improvement in the energy resolution, especially at low energies.

Data driven methods prove to be performant in the energy and vertex reconstruction given that we have enough calibrations sources to produce the PDF. In the next section, we'll see another type of



(A) Radial resolutions of the likelihood-based algorithm TMLE, QMLE and QTMLE

(B) Energy resolution of QMLE and QTMLE using different vertex resolutions

FIGURE 2.19

data-driven method based on machine learning.

2.6.3 Machine learning for reconstruction

Machine learning (ML) is family of data-driven algorithms that are inferring behavior and results from a training dataset. A overview of methods and detailed explanation of the Neural Network (NN) subfamily can be found in Chapter 3.

The power of ML is the ability to model complex response to a specific problem. In JUNO the reconstruction problematic can be expressed as follow: knowing that each PMT, large or small, detected a given number of PE Q at a given time t and their position is x, y, z where did the energy was deposited and how much energy was it, modeling a function that naively goes:

$$\mathbb{R}^{5\times N_{pmt}} \longmapsto \mathbb{R}^4 \tag{2.25}$$

It is worth pointing that while this is already a lot in informations, this is not the rawest representation of the experiment. We could indeed replace the charge and time by the waveform in the time window of the event but that would lead to an input representation size that would exceed our computational limits. Also, due to those computational limits, most of the ML algorithm reduce this input phase space either by structurally encoding the information (pictures, graph), by aggregating it (mean, variance, ...) or by exploiting invariance and equivariance of the experiment (rotational invariance due to the sphericity, ...).

For machine learning to converge to performant algorithm, a large dataset exploring all the phase space of interest is needed. For the following studies, data from the monte carlo simulation presented in section 2.5 are used for training. When the detector will be finished calibrations sources will be complementarily be used.

Boosted Decision Tree (BDT)

On of the most classic ML method used in physics in last years is the Boosted Decision Tree (see chapter 3.1). They have been explored for vertex reconstruction [41] et for energy reconstruction [41, 42].

For vertex and energy reconstruction a BDT was developed using the aggregated informations presented in 2.6.

Parameter	description
$nHits$	Total number of hits
$x_{cc}, y_{cc}, z_{cc}, R_{cc}$	Coordinates of the center of charge
ht_{mean}, ht_{std}	Hit time mean and standard deviation

TABLE 2.6 – Features used by the BDT for vertex reconstruction

Its reconstruction performances are presented in figure 2.21.

A second and more advanced BDT, subsequently named BDTE, that only reconstruct energy use a different set of features [42]. They are presented in the table 2.7

Neural Network (NN)

The physics have shown a rising for Neural Network (NN) in the past years for event reconstruction, notably in the neutrino community [43–46]. Three type of neural networks have explored for event reconstruction in JUNO Deep Neural Network (DNN), Convolutional Neural Network (CNN) and Graph Network (GNN). More explanation about those neural network can be found in chapter 3.

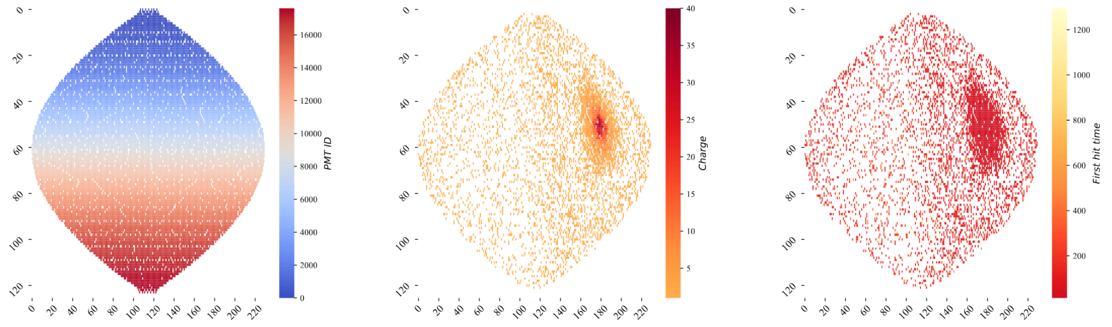
The CNN are using 2D projection of the detector representing it as an image with two channel, one for the charge Q and one for the time t. The position of the PMTs is structurally encoded in the pixel containing the information of this PMT. In [41], the pixel is chosen based on a transformation of θ and ϕ coordinates to the 2D plane and rounded to the nearest pixel. A sufficiently large image has been chosen to prevent two PMT to be located in the same pixel. An example of this projection can be found in figure 2.20. The performances of the CNN can be found in figure 2.21.

Using 2D have the upside of encoding a large part of the informations structurally but loose the rotational invariance of the detector. It also give undefined information to the neural network (what is a pixel without PMT ? What should be its charge and time ?), cause deformation in the representation of the detector (sides of projection) and loose topological informations.

One of the way to present structurally the sphericity of JUNO to a NN is to use a graph: A collection of objects V called nodes and relations E called edges, each relation associated to a couple v_1, v_2

AccumCharge	$ht_{5\%-2\%}$
R_{cht}	pe_{mean}
z_{cc}	J_{cht}
pe_{std}	ϕ_{cc}
nPMTs	$ht_{35\%-30\%}$
$ht_{kurtosis}$	$ht_{20\%-15\%}$
$ht_{25\%-20\%}$	$pe_{35\%}$
R_{cc}	$ht_{30\%-25\%}$

TABLE 2.7 – Features used by the BDTE algorithm. *pe* and *ht* reference the charge and hit-time distribution respectively and the percentages are the quantiles of those distributions. *cht* and *cc* reference the barycenters of hit time and charge respectively



(a) Map of PMTs. **(b)** Charge channel. **(c)** First hit time channel.

FIGURE 2.20 – Projection of the LPMTs in JUNO on a 2D plane. (a) Show the distribution of all PMTs and (b) and (c) are example of what the charge and time channel looks like respectively

forming the graph $G(E, V)$. Nodes and edges can hold informations or features. In [41] the nodes, are geometrical region of the detector as defined by the HealPix [47]. The features of the nodes are aggregated informations from the PMTs it contains. The edges contains geographic informations of the nodes relative positions.

This data representation has the advantages to keep the topology of the detector intact. It also permit the use of rotational invariant algorithms for the NN, thus taking advantage of the symmetries of the detector.

The neural network then process the graph using Chebyshev Convolutions [48]. The performances of the GNN are presented in figure 2.21.

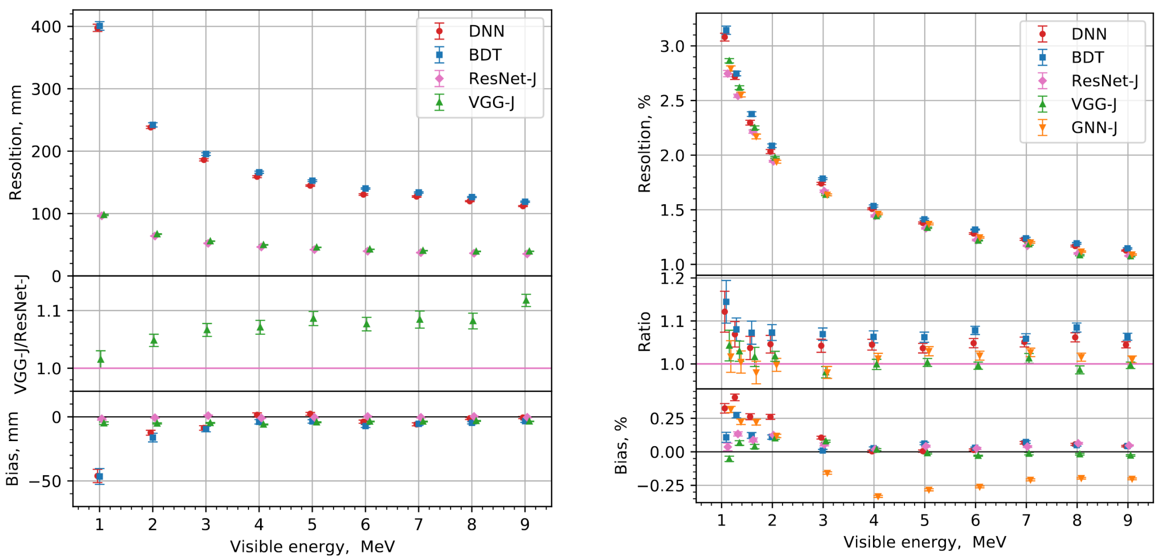


FIGURE 2.21 – Radial (left) and energy (right) resolutions of different ML algorithms. The results presented here are from [41]. DNN is a deep neural network, BDT is a BDT, ResNet-J and VGG-J are CNN and GNN-J is a GNN.

Overall ML algorithms show similar performances as classical algorithms in term of energy reconstructions with the more complex structure CNN and GNN showing better performances than BDT and DNN. For vertex reconstruction, the BDT and DNN show poor performance while CNN are on the level of the classical algorithms.

2.7 JUNO sensitivity to NMO and precise measurements

Now that the event have been reconstructed, selected and that the non-IBD background have been rejected, we have access to the measured energy flux from JUNO. We consider two spectra, the one measured by the LPMT system and the one measured by the SPMT system. This give rise to three possible analysis: A LPMT only analysis, a SPMT only analysis and a joint analysis. This joint analysis is the subject of the chapter 7 of this thesis.

The following details about JUNO measurement is common to the three analysis. The details and specific of the joint analysis are detailed in chapter 7.

2.7.1 Theoretical spectrum

To extract the oscillation parameters and the NMO from the measured spectrum, it is compared to a theoretical spectrum. This theoretical spectrum is produced based on the theory of the three flavour oscillation (see section 1.3), the measurements of the calibration and satellite experiments and Monte Carlo simulation:

- — The absolute flux and the fission product fraction calibrated by TAO.
- — The estimation of the neutrinos flux from other sources, such as the geoneutrinos, by theoretical model.
- — The computed cross-section of $\bar{\nu}_e$ and the LS.
- — The estimation of mislabelled event, such as fast neutron events from cosmic muons, using Monte Carlo simulation.
- — The measured bias and resolution of the LPMT and SPMT system by the calibration.
- — The time dependent reactor parameters (age of fuel, instantaneous power of the reactors, etc...)

These systematics parameters come with their uncertainties that need to be taken into account by the fitting framework. This theoretical spectrum will, in the end, depend of the oscillation parameters of interest $\theta_{13}, \theta_{12}, \Delta m^2_{21}, \Delta m^2_{31}$. Noise parameters can be included in the parameters spectrum such as the earth density ρ between the power plants and JUNO.

2.7.2 Fitting procedure

The theoretical and measured spectra are represented as two histograms depending on the energy. The theoretical spectrum is adjusted with the data using a χ^2 minimization where χ^2 is naively defined as

$$\chi^2 = \sum_i \frac{(N^i_{th} - N^i_{data})^2}{\sigma_i^2} \tag{2.26}$$

where N^i_{th} is the number event in the ith bin of the theoretical spectrum, N^i_{data} is the number of event in the ith bin of the measured spectrum and σ_i is the uncertainty of this bin. Two classic statistic test exist Pearson and Neyman where the difference is the estimation of σ_i parameters.

This σ_i is composed of the systematics uncertainties discussed above but also from the statistic uncertainty of the spectrum. Considering a Poisson process, the statistic uncertainty is estimated as $\sigma^i_{stat} = \sqrt{N^i}$. In a Pearson test, $N^i \equiv N^i_{th}$ whereas in a Neyman test $N^i \equiv N^i_{data}$. Under the assumption that the content of each bin follow a Gaussian distribution (a Poisson with high enough statistic), the two test are equivalent. But studies on Monte Carlo spectrum showed that the Pearson and Neyman statistic are biased in opposite direction. It is easily visible where, for the same data, Pearson will prefer a higher N^i_{th} to reduce the ration $\frac{1}{N^i_{th}}$ whereas Neyman will prefer a lower N^i_{th} to reduce the $(N^i_{th} - N^i_{data})$ term.

This problematic can be circumvented by summing the two test, yielding the CNP statistic test and/or by adding a term

$$\chi^2 = \sum_i \frac{(N^i_{th} - N^i_{data})^2}{\sigma_i^2} - \ln|\mathbf{V}| \tag{2.27}$$

where V is the covariance matrix of the theoretical spectrum yielding the PearsonV and CNPV statistic test.

The χ^2 is minimized by exploring the parameter phase space via gradient descent.

2.7.3 Physics results

The oscillation parameters are directly extracted from the minimization procedure and the error can be estimated directly from the procedure. For the NMO, the data are fitted under the two assumption of NO and IO. The difference in χ^2 give us the preferred ordering and the significance of our test. Latest studies show that the precision on oscillation parameters after six year of data taking will be of 0.2%, 0.3%, 0.5% and 12.1% for Δm^2_{31}, Δm^2_{21}, $\sin^2\theta_{12}$ and $\sin^2\theta_{13}$ respectively [11]. The expected sensitivity to mass ordering is 3σ after 6 years [49].

2.8 Summary

JUNO is one the biggest new generation neutrino experiment. Its goal, the measurements of oscillation parameters with unprecedented precision and an NMO preference at the 3 sigma confidence level, needs an in depth knowledge and understanding of the detector and the physics at hand. The characterisation and calibration of the detector are of the utmost importance and the understanding of the detector response in its resolution and bias is capital to be able to correctly carry the high precision physics analysis of the neutrino oscillation.

In this thesis, I explore the usage of data-driven reconstruction methods to validate and optimize the reconstruction of IBD events in JUNO in the chapters 4, 5 and 6 and the usage of the dual calorimetry in the detection of possible mis-modelisation in the theoretical spectrum 7.

Chapter 3

Machine learning and Artificial Neural Network

> *"I have the shape of a human being and organs equivalent to those of a human being. My organs, in fact, are identical to some of those in a prosthetized human being. I have contributed artistically, literally, and scientifically to human culture as much as any human being now alive. What more can one ask?"*
>
> *Isaac Asimov, The Complete Robot*

Machine Learning (ML) and more specifically Neural Network (NN) are families of data-driven algorithm. They are used to model complex distributions from a finite dataset to extract a generalist behavior. They learn, adapt their intrinsic parameters, interactively by computing its performance or *loss* on those dataset. They take advantage of simple microscopic operation such as *if condition* or non-continuous but differentiable function like *ReLU*. Through optimizers and the combination of a lot of those microscopic operations, they can obtain complex and precise behaviours.

They are now widely used in a wide variety of domain including natural language processing, computer vision, speech recognition and, the subject of this thesis, scientific studies.

We found them in particle physics, either as the main algorithm or as secondary algorithm, for event reconstruction, event classification, waveform reconstruction, etc..., domains where the underlying physic and detector process is complex and highly dimensional. Physicists have traditionally been forced to use simplifications or assumptions to ease the developmment of algorithms or equations (a good example is the algorithm presented in section 2.6) where machine learning could refine and take into account those effects, provided that they have enough data and computing power.

This chapter present an overview of the different kind of machine learning methods and neural networks that will be discussed in this thesis.

3.1 Boosted Decision Tree (BDT)

One of the most classic machine learning algorithm used in particle physics is Boosted Decision Tree (BDT) [50] (or more recently Gradient Boosting Machine [51]). The principle of a BDT is fairly simple : based on a set of observables, a serie of decisions, represented as node in a tree, are taken by the algorithm. Each decision point, or node, takes its decision based on a set of trainable parameters leading to a subtree of decision. The process is repeated until it reach the final node, yielding the prediction. A simplistic example is given in figure 3.1.

The training procedure follow a simple score reward procedure. During the training phase the prediction of the BDT is compared to a known truth about the data. The score is then used to

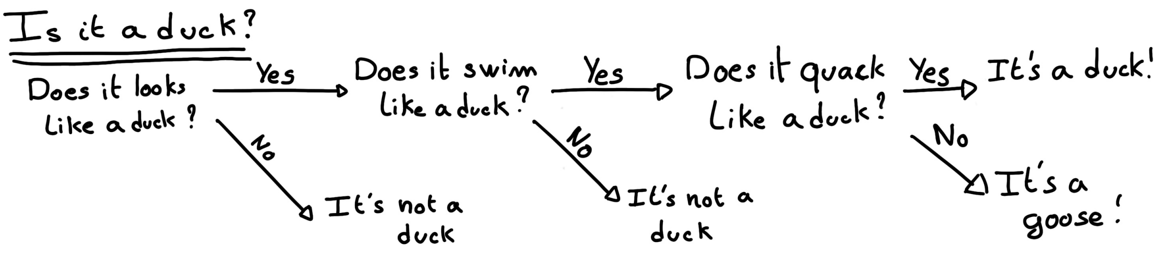


FIGURE 3.1 – Example of a BDT that determine if the given object is a duck

backpropagate corrections to the parameters of the tree. Modern BDT use gradient boosting where the gradient of the loss is calculated for each of the BDT parameters. Following the gradient descent, we can reach the, hopefully, global minima of the loss for our set of parameters.

3.2 Artificial Neural Network (NN)

One other big family of machine learning algorithm is the artificial Neural Networks (NN). The idea of developing automates which component mimic, in a simplistic way, the behavior of biological neurons emerge in 1959 with the paper "*What the Frog's Eye Tells the Frog's Brain*" [52]. They develop an automate where each component possess an *activation function*. Each one of those component then transmit its information to the other following a certain efficiency or *weight*. Those works influenced scientist and notably Frank Rosenblatt who published in 1958 what is considered the first neural network model the Perceptron [53].

Modern neural network still nowadays use the neuron metaphor to represent neural network, but approach them as a graph where the nodes are neurons possessing an activation function and edges holding the weights between those nodes. Most of the modern neural network work with the principle of neurons layers. Each neurons belong to a layer and takes input from the preceding layer and forward it result to next layer. For example the most basic set layer is the fully connected layer where each of its neurons is connected to every other neurons of the precessing layer. All the neurons posses the same activation function F. The connection between two the two layers is expressed as a tensor T^i_j where i is the index of the precedent layer and j the index of the current layer. The propagation from the layer I to J is then described as

$$J_j = F_J(T^i_j I_i + B_j) \tag{3.1}$$

where the learning parameters are the tensor T^i_j and the bias tensor B_j. This is the fundamental component of the Fully Connected Deep NN (FCDNN) family presented in section 3.2.1. Most of the modern neural networks use gradient descent to optimize their parameters, i.e. the gradient of the parameter θ in respect of the loss function $\mathcal{L}$ is subtracted to it

$$\theta_{i+1} = \theta_i - \frac{\partial \theta}{\partial \mathcal{L}} \tag{3.2}$$

i being the training iteration index. This needs the expression of $\mathcal{L}$ dependent of θ to be differentiable, thus the layer and their activation function also need to be differentiable. This simple gradient descent, designated as Stochastic Gradient Descent (SGD), can be completed with first and second order momentum like with the Adam optimizer [54].

This description of neural networks as layer introduced the principle of *depth* and *width*, the number of layers in the NN and the number of neurons in each layer respectively.

3.2.1 Fully Connected Deep Neural Network (FCDNN)

Fully Connected Deep Neural Network (FCDNN) architecture is the natural evolution of the Perceptron. The input data is represented as a first order tensor I_j and then fed forward to multiple fully connected layers (Eq 3.1) as presented in the figure 3.2a. Most of the time, the classic ReLU function

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

is used as activation function. Prelu and Sigmoid are also popular choices:

$$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{3.4}$$

$$\text{PReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{otherwise} \end{cases} \tag{3.5}$$

The reasoning behind ReLU and PReLU is that with enough of them, you can mimic any continuous function as illustrated in figure 3.2b. Sigmoid is more used in case of classification, its behavior going hand in hand with the Cross Entropy loss function used in classification problems.

(A) Schema of a FCDNN

(B) Illustration of a composition of ReLU matching a function

FIGURE 3.2

Due to its simplicity, FCDNN are also used as basic pieces for more complex architectures such as the CNN and GNN that will be presented in the next section.

3.2.2 Convolutional Neural Network (CNN)

Convolutional Neural Networks are a family of neural networks that use discrete convolution filters to process the input data. They have the advantage to be translation invariant by construction, this mean that they are capable of detecting oriented features independently of their location on the image. The learning parameters are located in the filters, the network thus learn the optimal filters extract the desired feature. 2D CNN, where the filters are second order tensors that span over third order tensors, are commonly used in image recognition [55] for classification or regression problematics.

The convolution layers are commonly chained [56], reducing the input dimension while increasing the number of filters. The idea behind is that the first layers will process local informations and the latest layers will process more global informations. To try to preserve the amount of information, we tend double the numbers of filters for each division of the input data. The results of the convolution filters is commonly then flattened and feed to a smaller FCDNN which will process the filters results to yield the desired output.

FIGURE 3.3 – Illustration of a CNN convolution filter

3.2.3 Graph Neural Network (GNN)

Graph neural network is a family of neural network where the data is represented as a graph $G(\mathcal{N}, \mathcal{E})$ composed of vertex or node $n \in \mathcal{N}$ and edges $e \in \mathcal{E}$. The edges are associated to two nodes $(u, v) \in \mathcal{N}^2$, "connecting" them. The node and the edges can hold features, commonly represented as vector $n \in \mathbb{R}^{k_n}$, $e \in \mathbb{R}^{k_e}$. We can thus define a graph using two tensors A^{ij}_{ϵ} the adjacency tensors that hold the features ϵ of the edge connecting the node i and j and the tensor N^i_ν that hold the features ν of a node i.

To efficiently manipulate such object we need to structurally encode their property in the neural network architecture: each node is equivalent (as opposite to ordered data in a vector), each node has a set of neighbours, ... One of this method is the message passing algorithm presented historically in "Neural Message Passing for Quantum Chemistry" [57]. In this algorithm, with each layer of message passing a new set of features is computed for each node following

$$n_i^{k+1} = \phi_u(n_i^k, \Box_j \phi_m(n_i^k, n_j^k, e_{ij}^k));\ n_j \in \mathcal{N}_i' \tag{3.6}$$

where ϕ_u is a differentiable update function, $\Box_j$ is a differentiable aggregation function and ϕ_m is a differentiable message function. $\mathcal{N}_i' = \{n_j \in \mathcal{N} | (n_i, n_j) \in \mathcal{E}\}$ is the set of neighbours of n_i, i.e. the nodes n_j from which it exist an edge $e_{i,j} \to (n_i, n_j)$. k is the layer on which the message passing algorithm is applied. $\Box$ need also a few other property if we want to keep the graph property, most notably the permutational invariance of its parameters (example: mean, std, sum, ...).

The edges features can also be updated, either by directly taking the results of ϕ_m or by using another message function ϕ_e.

Message passing is a very generic way of describing the process of GNN and it can be specialized for convolutional filtering [48], diffusion [58] and many other specific operation. GNN are used in a wide variety of application such as regression problematics, node classification, edge classification, node and edge prediction, ...

It is a very versatile but complex tool.

3.2.4 Adversarial Neural Network (ANN)

The adversarial machine learning, Adversarial Neural Networks (ANN) in the case of neural network, is a family of unsupervised machine learning algorithms where the learning algorithm (generator) is competing against another algorithm (discriminator). Taking the example of Generative Adversarial Networks, concept initially developed by Goodfellow et al. [59], the discriminator goal is to discriminate between data coming from a reference dataset and data produced by the generator. The generator goal, on the other hand, is to produce data that the discriminator would not be able to differentiate from data from the reference dataset. The expression of duality between the two models is represented in the loss where, at least a part of it, is driven by the results of the discriminator.

Chapter 4

Image recognition for IBD reconstruction with the SPMT system

As explained in chapter 2, JUNO is an experiment composed of two systems, the Large Photomultiplier (LPMT) and the Small Photomultiplier (SPMT). Both of the system observe the same physics event inside of the same medium but they differ in their photo-coverage, respectively 75.2% and 2.7%, their dynamic range (see section 2.2.2, a thousands versus a few dozen, and their front-end electronics (see section 2.2.2).

They are complementary in their strengths and weaknesses and support each other. One important point is their differences in expected resolution, the LPMT system outperform largely the SPMT system but is subject to effects such as charge non linearity [28] that could bias the reconstruction, effect that the SPMT system is impervious to. This topic will be studied in more detail in chapter 7. Also, due to the dynamic range of the LPMT, in case of high energy and high density event such as core-collapse supernova, the LPMT system could saturate and the lower photo-coverage become a benefit.

Thus, although event reconstruction algorithm and physics analysis combines both LPMT and SPMT systems, individual approach are key studies to understand the detector and ensure their reliability. This topic will also be studied in more details in chapter 7. The subject of this chapter is to propose a machine learning algorithm for the SPMT reconstruction based on Convolutional Neural Network (CNN).

4.1 Motivations

As explained in chapter 3, Machine Learning (ML) algorithms shine when modeling highly dimensional data from a given dataset. In our case, we have access to complete monte-carlo simulation of our detector to produce arbitrary large datasets that could represent multiple years of data taking. Ideally ML algorithms would be able to consider the entirety of the information in the detector and converge on the best parameters to yield optimal results, while classical methods where the algorithms could be biased by the prior knowledge of the detector and physics processes. To study this potential phenomena, we will compare our machine algorithm to a classical reconstruction method developed for energy and vertex reconstruction [60].

We have access to a very detailed simulation of the detector (section 2.5) that will allow us to simulate arbitrary large dataset of data while giving access to the all the physics parameters of the event. Those parameters include the target of our reconstruction algorithms: the vertex and position at with the event happened. As introduced above, we hope that the ML algorithm will be able to used all the informations in the event, meaning that potential mismodelings in our simulation could be exploited by the algorithm. This specific subject will be studied in chapter 6.

4.2 Method and model

One of simplest way to look at JUNO data is to consider the detector as an array of geometrically distributed sensors on a sphere. Their repartition is almost homogeneous, on this sphere surface providing an almost equal amount of information per unit surface on this sphere. It is then tempting to represent the detector as a spherical image with the PMT in place of pixel. Two events with two different energy or position would produce two different images.

The most common approach in machine learning for image processing and image recognition is the Convolutional Neural Network (CNN). It is widely used in research and industry [56, 61–63] due to its strengths (see section 3.2.2) and has proven its relevance in image processing.

Some CNN are developed to process spherical images [64] but for the sake of simplicity and as a first approach we decided to go with a planar projection of the detector, approach that has proven its efficiency using the LPMT system (see section 2.6.3). The details about this planar projection will be discussed in section 4.2.3.

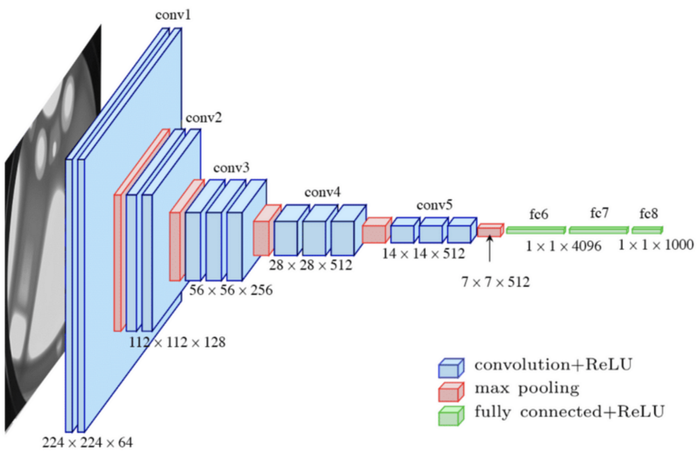


FIGURE 4.1 – Graphic representation of the VGG-16 architecture, presenting the different kind of layer composing the architecture.

4.2.1 Model

The architecture we use is derived from the VGG-16 architecture [56] illustrated in figure 4.1. We define a set of hyperparameters that will define the size, complexity and computational power of the NN. The chose hyperparameters are detailed below and their values are presented in table 4.1.

- **$\mathbf{N_{blocks}}$**: the number of convolution blocks, a block being composed of two convolutional layers with 3×3 filters using ReLU activation function, a 3×3 max-pooling layer (except for the last block) and a dropout layer.
- **$\mathbf{N_{channels}}$**: The number of channels in the first block. The number of channels in the subsequent blocks are calculated using $N^i_{channels} = 2^i * N_{channels}, i \in [1..N_{blocks}]$.
- **FCDNN configuration**: The result of the last convolution layer is flattened then fed to a FCDNN. Its configuration is expressed as a sequence of fully connected linear layer using the PReLU activation function. For example $2 * 1024 + 2 * 512$ is the sequence of 2 layers with a width of 1024 followed by 2 other layers with a width of 512. Finally the last layer is a 4 neurons wide linear layers without activation function. Each neurons of the last layer represent a component of the interaction vertex: Energy, X, Y, Z.
- **Loss**: The loss function. In this work we study two different loss function $(E + V)$ and $(E_r + V_r)$ detailed below.

We explore in this work two different activation functions

$$(E+V)(E,x,y,z) = \left\langle (E-E_{true})^2 + 0.85 \sum_{\lambda \in [x,y,z]} (\lambda - \lambda_{true})^2 \right\rangle \tag{4.1}$$

$$(E_r+V_r)(E,x,y,z) = \left\langle \frac{(E-E_{true})^2}{E_{true}} + \frac{10}{R} \sum_{\lambda \in [x,y,z]} (\lambda - \lambda_{true})^2 \right\rangle \tag{4.2}$$

where R is the radius of the CD. With the energy in MeV and the distance in meters, we use the factor 0.85 and 10 to equilibrate the two term of the loss function so they have the same magnitude.

— The loss function $(E+V)$ is close to a simple Mean Squared Error (MSE). MSE is one of the most basic loss function, the derivative is simple and continuous in every point. It is a strong starting point to explore the possibility of CNNs.

— (E_r+V_r) can be see as a relative MSE.

The idea is that: due to the inherent statistic uncertainty over the number of collected Number of Photo Electrons (NPE), the absolute resolution $\sigma(E-E_{true})$ will be larger at higher energy than at low energy. But we expect the *relative* energy resolution $\frac{\sigma(E-E_{true})}{E_true}$ to be smaller at high energy than lower energy as illustrated in figure 2.19. Because of this, by using simple MSE the most important part in the loss come from the high energy part of the dataset whereas with a relative MSE, the most important become the low energy events in the dataset. We hope that by using a relative MSE, the neural network will focus on low energy events where the reconstruction is considered the hardest part of the dataset.

N_{blocks}	$\{2, 3, 4\}$
$N_{channels}$	$\{32, 64, 128\}$
FCDNN configurations	2 * 1024 2 * 2048 + 2 * 1024 3 * 2048 + 3 * 512 2 * 4096
Loss	$\{E+V, E_r+V_r\}$

TABLE 4.1 – Sets of hyperparameters values considered in this study

Each combination of those hyperparameters (for example $(N_{blocks} = 2, N_{channels} = 32, \text{FCDNN} = (2*1024), \text{Loss} = (E+V))$), subsequently designated as configurations, is then tested and compared to each other over an analysis sample. We cannot use the mean loss because we consider multiple loss functions, there is no guarantee that comparison of their numerical value will be meaningful. We use multiple observables to rank the performances of each configuration:

— The mean absolute energy error $\langle E \rangle = \langle |E - E_{true}| \rangle$. It is an indicator of the energy bias of our reconstruction.

— The standard deviation of the energy error $\sigma E = \sigma(E - E_{true})$. This the indicator on our precision in energy reconstruction.

— The mean distance between the reconstructed vertex and the true vertex $\langle V \rangle = \langle |\vec{V} - \vec{V}_{true}| \rangle$. This an indicator of the bias and precision of our vertex reconstruction.

— The standard deviation of the distance between the true and reconstructed vertex $\sigma V = \sigma|\vec{V} - \vec{V}_{true}|$. This is an indicator if the precision in our vertex reconstruction.

For the vertex indicator, it is interesting noting that $\langle V \rangle$ is not expected to be zero and so is unfit to be used for the loss.

TODO: REMOVE BEFORE END Passage en francais pour poser les mots : en parler avec benoit. En gros, en utilisant la MSE et au vu des fluctuations statisque auxquelles sont soumis nos données, on s'attend a ce que toute somme quadratique de nos observables $(\lambda_{rec} - \lambda_{true})$ suivent une loi du chi2. C'est interessant car ca veut dire que la "perfection" de $\langle \sum \rangle = 0$ n'est pas réalisable -> AKA

MSE = 0 n'est pas realisable. Cela vient de la fluctation de nos donnees. En effet tout l'aleatoric de nos données dans la simulation (et pb pd dans la vrai vie) sont régie par des comportment quasi gaussien. Ainsi tout estimateur de nos observables auront leur erreur gaussienne. Cela soulève une analyse intérssante de l'interpretation du nb de dg de liberté des lois du chi2 produites. En effet, si il n'existe aucune correlation entre erreurs de nos observables (et que gaussienne tout ca tout ca), alors la loi du chi2 possede un $k = N_{obs}$ si ce n'est pas le cas, dans ce cas $k < N_{obs}$ et donc les différents estimateurs peuvent etre exprimés par moins d'observables.

4.2.2 Data characteristics

4.2.3 Data representation

This data is represented as 240×240 images, equivalent to third order tensor, with a charge Q channel and a time t channel. The SPMTs are then projected on the plane as illustrated in figure 4.2a. The x position is proportional to θ and the y position is defined by $\phi \sin\theta$ in spherical coordinates. $\theta = 0$ is defined as being the top of the detector and $\phi = 0$ is defined as an arbitrary direction in the detector. In practice, this is the $\phi = 0$ given by the MC simulation.

$$x = \left\lfloor \frac{\theta \cdot H}{\pi} \right\rfloor, \; \theta \in [0, \pi] \tag{4.3}$$

$$y = \left\lfloor \frac{(\phi + \pi)\sin\theta \cdot W}{2\pi} \right\rfloor, \; \phi \in [-\pi, \pi], \; \theta \in [0, \pi] \tag{4.4}$$

where H is the height of the image, W the width of the image and $(0, 0)$ the top left corner of the image.

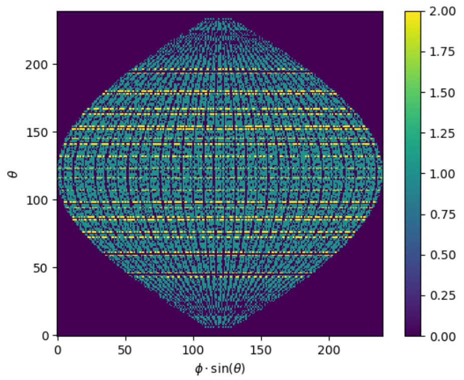
When two SPMTs are in the same pixel, the charges are summed and the lowest of the hit-time is chosen. The SPMTs being located close to each other, we expect the time difference between two successive physics signals, two photons being collected, to be small. The first hit time is chosen because it can be considered as the relative propagation time of the photons that went the "straightest", i.e. that went under the less perturbation of the two. The only potential problem in using this first time come from the Dark Noise (DN). Its time distribution is uniform over the signal and could come before a signal hit on the other SPMT in the pixel. In that case, the time information in the pixel become irrelevant and we lose the timing information for this part of the detector. As illustrated in figure 4.2a the dimension have been chosen optimized so that at most two SPMTs are in the same pixel while keeping the number of empty pixels relatively low to prevent this kind of issue.

While it could be possible to use larger images (more pixel) to prevent overlapping, keeping image small images gives multiple advantages:

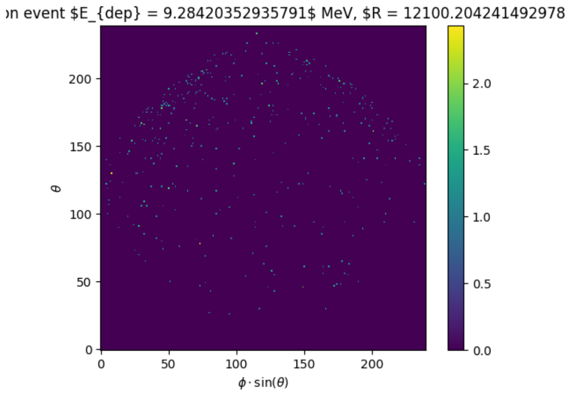
- As presented in section 4.2.1, the convolution filter we use are 3×3 convolution filter, meaning that if SPMTs would be separated by more that one pixel, the first filter would only see one SPMT per filter. This behavior would be kind of counterproductive as the first convolution block would basically be a transmission layer and would just induce noise in the data.
- It keep the network relatively small, while this do not impact the convolution layers, the flatten operation just before the FCDNN make the number parameters in the first layer of it dependent on the size of the image.
- It reduce the number of empty pixel in the image.

The question of empty pixel is an important question in this data representation. Their is two kind of empty pixel in the data.

The first kind is pixel that contain a SPMT but the SPMT did not get hit nor registered any dark noise during the event. In this case, the charge channel is zero, which have a physical meaning but then



(A) Repartition of SPMTs in the image projection. The color scale is the number of SPMTs per pixel



(B) Example of the charge channel of an event as seen by the CNN **(Need to redo the title)**

FIGURE 4.2

come the question of the time layer. One could argue that the correct time would be infinity (or the largest number our memory allows us) because the hit "never" happened, so extremely far from the time of the event. This cause numerical problem as large number, in the linear operation that are happening in the convolution layers, are more signifiant than smaller value. We could try to encode this feature in another way but no number have any significance due to our time being relative to the trigger of the experiment so -1 for example is out of question. Float and Double gives us access to special value such as NaN (Not a Number) [65] but the behavior is to propagate the NaN which leaves us with NaN for energy and position. We choose to keep the value 0 because it's the absorbing element of multiplication, absorbing the "information" of the parameter it would be multiplied by. It also can be though as no activation in the ReLU activation function.

The second kind of pixel is pixel that do not represent parts of the detector such as the corners of the images. The question is basically the same, what to put in the charge and the time channel. The decision is to set the charge and time at 0 following the reasoning presented above. Its important to keep in mind that the fact that a part of the detector that has not been hit is also an information: There is no signal in this part of the detector. This problematic will be explored in more details in chapter 5.

Another problematic that happens with this representation, and this is not dependent of the chosen projection, is the deformation in the edges of the image and the loss of the neighbouring information in the for the SPMTs at the edge of the image $\phi \sim 180°$. This deformation and neighbouring loss could be partially circumvented as explained in section 4.4

4.2.4 Dataset

In this study we use one million events coming from the full JUNO official monte-carlo simulations J23.0.1-rc8.dc1 (released the 7th January 2024). To put in perspective, the expected IBD rate in JUNO is 47 / days. Taking into account the calibration time, and the source reactor shutdown, it amount to $\sim 94'000$ IBD events in 6 years. With this million of event, we are training the equivalent of ~ 10 years of data. With this amount we reach a density of $4783 \frac{\text{event}}{\text{m}^3 \cdot \text{MeV}}$, meaning our dataset is representative of the multiple event scenarios that could be happening in the detector.

While we expect and hope the monte-carlo simulation to give use a realistic representation of the detector, there could be effect, even after the fine-tuning on calibration data, that the simulation cannot handle. Thus, once the calibration will be available, we will need to evaluate, and if needed retrain, the network on calibration data to establish definitive performances.

The data used during this analysis is monte carlo data using the official JUNO simulation software (see section 2.5 for details). The simulated data is composed of positron events, uniformly distributed in the CD volume and in kinetic energy over $E_k \in [0; 9]$ MeV producing a deposited energy $E_{dep} \in [1.022; 10.022]$ MeV. This is done to mimic the signal produced by the IBD prompt signal. Uniform distribution are used so that the CNN does not learn a potential energy distribution, favoring some part of the energy spectrum instead of other.

Those events can be considered as "optimistic". There is no pile-up with potential background or other IBD. The neutron is also not simulated, and even if we expect the main neutron signal coming from its capture with a hydrogen nucleus to come well after the time window, there is still potential energy from the neutrons that could be deposited during this time window while it's still thermalizing in the LS.

The dark noise and time window come from the monte-carlo simulation of the electronic and trigger system. The dark noise rate used in this study is coming from studies and calibration of the PMTs outside of the experimental setup [25, 66].

4.3 Results

Before presenting the results, lets discuss the different observables.

The event are considered point like in this study. The target truth position, or vertex, is the mean position of the energy deposits of the positron and the two annihilation gammas. Due to the symmetries of the detector, we mainly considered and discuss the bias and precision evolution depending of the radius R but we will still monitor the performances depending of the spheric angle θ and ϕ. From the detector construction and effect we expect relative important dependencies in radius thanks to the TR area effect presented in section 2.6 and the possibility for the positron or the gammas to escape from the CD for near the edge events. We also expect dependence in θ, the top of the experiment being non-instrumented due to the filling chimney. It is also to be noted that the events in the dataset are uniformly distributed in the CD, and so are uniformly distributed in R^3 and ϕ. The θ distribution is not uniform and we will have more event for $\theta \sim 90°$ that $\theta \sim 0°$ or $\theta \sim 180°$.

We define multiple energy in JUNO:

- E_ν: The energy of the neutrino.
- E_k: The kinetic energy of the resulting positron from the IBD.
- E_{dep}: The deposited energy of the positron and the two annihilation gammas.
- E_{vis}: The equivalent visible energy, so E_{dep} after the detector effect such as the absorption of scintillation photons by the LS and the LS response non-linearity.
- E_{rec}: The reconstructed energy by the reconstruction algorithm. The expected value depend on the algorithm we discuss about. For example the algorithm presented in section 2.6 is reconstructing E_{rec} while the ones presented in section 2.6.3 reconstruct E_{dep}.

In this study, we will set E_{rec} as our target for energy reconstruction. This choice is motivated by the ease with which we can retrieve this information in the monte-carlo data while E_{vis} is less trivial to retrieve.

4.4 Prospect

4.5 Conclusion

Intoduction next chapter

Chapter 5

Graph representation of JUNO for IBD reconstruction with the LPMT system

Chapter 6

Reliability of machine learning methods

"Psychohistory was the quintessence of sociology; it was the science of human behavior reduced to mathematical equations. The individual human being is unpredictable, but the reactions of human mobs, Seldon found, could be treated statistically"

Isaac Asimov, Second Foundation

Chapter 7

Joint fit between the SPMT and LPMT spectra

Chapter 8

Conclusion

Bibliography

[1] Liang Zhan, Yifang Wang, Jun Cao, and Liangjian Wen. “Determination of the Neutrino Mass Hierarchy at an Intermediate Baseline”. *Physical Review D* 78.11 (Dec. 10, 2008), 111103. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.78.111103. eprint: 0807.3203[hep-ex,physics:hep-ph]. URL: http://arxiv.org/abs/0807.3203 (visited on 09/18/2023).

[2] Fengpeng An et al. “Neutrino Physics with JUNO”. *Journal of Physics G: Nuclear and Particle Physics* 43.3 (Mar. 1, 2016), 030401. ISSN: 0954-3899, 1361-6471. DOI: 10.1088/0954-3899/43/3/030401. eprint: 1507.05613[hep-ex,physics:physics]. URL: http://arxiv.org/abs/1507.05613 (visited on 07/28/2023).

[3] Liang Zhan, Yifang Wang, Jun Cao, and Liangjian Wen. “Experimental Requirements to Determine the Neutrino Mass Hierarchy Using Reactor Neutrinos”. *Physical Review D* 79.7 (Apr. 14, 2009), 073007. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.79.073007. eprint: 0901.2976[hep-ex]. URL: http://arxiv.org/abs/0901.2976 (visited on 09/18/2023).

[4] A. A. Hahn, K. Schreckenbach, W. Gelletly, F. von Feilitzsch, G. Colvin, and B. Krusche. “Antineutrino spectra from 241Pu and 239Pu thermal neutron fission products”. *Physics Letters B* 218.3 (Feb. 23, 1989), 365–368. ISSN: 0370-2693. DOI: 10.1016/0370-2693(89)91598-0. URL: https://www.sciencedirect.com/science/article/pii/0370269389915980 (visited on 01/16/2024).

[5] Th A. Mueller et al. “Improved Predictions of Reactor Antineutrino Spectra”. *Physical Review C* 83.5 (May 23, 2011), 054615. ISSN: 0556-2813, 1089-490X. DOI: 10.1103/PhysRevC.83.054615. eprint: 1101.2663[hep-ex,physics:nucl-ex]. URL: http://arxiv.org/abs/1101.2663 (visited on 01/16/2024).

[6] F. von Feilitzsch, A. A. Hahn, and K. Schreckenbach. “Experimental beta-spectra from 239Pu and 235U thermal neutron fission products and their correlated antineutrino spectra”. *Physics Letters B* 118.1 (Dec. 2, 1982), 162–166. ISSN: 0370-2693. DOI: 10.1016/0370-2693(82)90622-0. URL: https://www.sciencedirect.com/science/article/pii/0370269382906220 (visited on 01/16/2024).

[7] K. Schreckenbach, G. Colvin, W. Gelletly, and F. Von Feilitzsch. “Determination of the antineutrino spectrum from 235U thermal neutron fission products up to 9.5 MeV”. *Physics Letters B* 160.4 (Oct. 10, 1985), 325–330. ISSN: 0370-2693. DOI: 10.1016/0370-2693(85)91337-1. URL: https://www.sciencedirect.com/science/article/pii/0370269385913371 (visited on 01/16/2024).

[8] Patrick Huber. “On the determination of anti-neutrino spectra from nuclear reactors”. *Physical Review C* 84.2 (Aug. 29, 2011), 024617. ISSN: 0556-2813, 1089-490X. DOI: 10.1103/PhysRevC.84.024617. eprint: 1106.0687[hep-ex,physics:hep-ph,physics:nucl-ex,physics:nucl-th]. URL: http://arxiv.org/abs/1106.0687 (visited on 01/16/2024).

[9] P. Vogel, G. K. Schenter, F. M. Mann, and R. E. Schenter. “Reactor antineutrino spectra and their application to antineutrino-induced reactions. II”. *Physical Review C* 24.4 (Oct. 1, 1981). Publisher: American Physical Society, 1543–1553. DOI: 10.1103/PhysRevC.24.1543. URL: https://link.aps.org/doi/10.1103/PhysRevC.24.1543 (visited on 01/16/2024).

[10] D. A. Dwyer and T. J. Langford. “Spectral Structure of Electron Antineutrinos from Nuclear Reactors”. *Physical Review Letters* 114.1 (Jan. 7, 2015), 012502. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.114.012502. eprint: 1407.1281[hep-ex,physics:nucl-ex]. URL: http://arxiv.org/abs/1407.1281 (visited on 01/16/2024).

[11] JUNO Collaboration et al. “Sub-percent Precision Measurement of Neutrino Oscillation Parameters with JUNO”. *Chinese Physics C* 46.12 (Dec. 1, 2022), 123001. ISSN: 1674-1137, 2058-6132. DOI: 10.1088/1674-1137/ac8bc9. eprint: 2204.13249[hep-ex]. URL: http://arxiv.org/abs/2204.13249 (visited on 08/11/2023).

[12] JUNO Collaboration et al. *TAO Conceptual Design Report: A Precision Measurement of the Reactor Antineutrino Spectrum with Sub-percent Energy Resolution*. May 18, 2020. DOI: 10.48550/arXiv.2005.08745. eprint: 2005.08745[hep-ex,physics:nucl-ex,physics:physics]. URL: http://arxiv.org/abs/2005.08745 (visited on 01/18/2024).

[13] G. Mention, M. Fechner, Th. Lasserre, Th. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau. “Reactor antineutrino anomaly”. *Physical Review D* 83.7 (Apr. 29, 2011). Publisher: American Physical Society, 073006. DOI: 10.1103/PhysRevD.83.073006. URL: https://link.aps.org/doi/10.1103/PhysRevD.83.073006 (visited on 03/05/2024).

[14] V. Kopeikin, M. Skorokhvatov, and O. Titov. “Reevaluating reactor antineutrino spectra with new measurements of the ratio between ^{235}U and ^{239}Pu β spectra”. *Physical Review D* 104.7 (Oct. 25, 2021), L071301. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.104.L071301. eprint: 2103.01684[hep-ph,physics:nucl-ex,physics:nucl-th]. URL: http://arxiv.org/abs/2103.01684 (visited on 01/18/2024).

[15] A. Letourneau et al. “On the origin of the reactor antineutrino anomalies in light of a new summation model with parameterized β^- transitions”. *Physical Review Letters* 130.2 (Jan. 10, 2023), 021801. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.130.021801. eprint: 2205.14954[hep-ex,physics:hep-ph]. URL: http://arxiv.org/abs/2205.14954 (visited on 01/16/2024).

[16] Particle Data Group et al. “Review of Particle Physics”. *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 14, 2020), 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. URL: https://doi.org/10.1093/ptep/ptaa104 (visited on 12/04/2023).

[17] Super-Kamiokande Collaboration et al. “Diffuse Supernova Neutrino Background Search at Super-Kamiokande”. *Physical Review D* 104.12 (Dec. 10, 2021), 122002. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.104.122002. eprint: 2109.11174[astro-ph,physics:hep-ex]. URL: http://arxiv.org/abs/2109.11174 (visited on 02/28/2024).

[18] Alessandro Strumia and Francesco Vissani. “Precise quasielastic neutrino/nucleon cross section”. *Physics Letters B* 564.1 (July 2003), 42–54. ISSN: 03702693. DOI: 10.1016/S0370-2693(03)00616-6. eprint: astro-ph/0302055. URL: http://arxiv.org/abs/astro-ph/0302055 (visited on 01/16/2024).

[19] Daya Bay et al. *Optimization of the JUNO liquid scintillator composition using a Daya Bay antineutrino detector*. July 1, 2020. DOI: 10.48550/arXiv.2007.00314. eprint: 2007.00314[hep-ex,physics:physics]. URL: http://arxiv.org/abs/2007.00314 (visited on 07/26/2023).

[20] J. B. Birks. “CHAPTER 3 - THE SCINTILLATION PROCESS IN ORGANIC MATERIALS—I”. *The Theory and Practice of Scintillation Counting*. Ed. by J. B. Birks. International Series of Monographs in Electronics and Instrumentation. Jan. 1, 1964, 39–67. ISBN: 978-0-08-010472-0. DOI: 10.1016/B978-0-08-010472-0.50008-2. URL: https://www.sciencedirect.com/science/article/pii/B9780080104720500082 (visited on 02/07/2024).

[21] *Photomultiplier tube R12860 | Hamamatsu Photonics*. URL: https://www.hamamatsu.com/eu/en/product/optical-sensors/pmt/pmt_tube-alone/head-on-type/R12860.html (visited on 02/08/2024).

[22] Yan Zhang, Ze-Yuan Yu, Xin-Ying Li, Zi-Yan Deng, and Liang-Jian Wen. “A complete optical model for liquid-scintillator detectors”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 967 (July 2020), 163860. ISSN: 01689002. DOI: 10.1016/j.nima.2020.163860. eprint: 2003.12212[physics]. URL: http://arxiv.org/abs/2003.12212 (visited on 02/07/2024).

[23] Hai-Bo Yang et al. “Light Attenuation Length of High Quality Linear Alkyl Benzene as Liquid Scintillator Solvent for the JUNO Experiment”. *Journal of Instrumentation* 12.11 (Nov. 27, 2017), T11004–T11004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/11/T11004. eprint: 1703.

01867[hep-ex,physics:physics]. URL: http://arxiv.org/abs/1703.01867 (visited on 07/28/2023).

[24] JUNO Collaboration et al. *The Design and Sensitivity of JUNO's scintillator radiopurity pre-detector OSIRIS*. Mar. 31, 2021. DOI: 10.48550/arXiv.2103.16900. eprint: 2103.16900[physics]. URL: http://arxiv.org/abs/2103.16900 (visited on 02/07/2024).

[25] Angel Abusleme et al. "Mass Testing and Characterization of 20-inch PMTs for JUNO". *The European Physical Journal C* 82.12 (Dec. 24, 2022), 1168. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-11002-8. eprint: 2205.08629[hep-ex,physics:physics]. URL: http://arxiv.org/abs/2205.08629 (visited on 02/08/2024).

[26] Yang Han. "Dual Calorimetry for High Precision Neutrino Oscillation Measurement at JUNO Experiment". AstroParticule et Cosmologie, France, Paris U. VII, APC, June 2021.

[27] R. Acquafredda et al. "The OPERA experiment in the CERN to Gran Sasso neutrino beam". *Journal of Instrumentation* 4.4 (Apr. 2009), P04018. ISSN: 1748-0221. DOI: 10.1088/1748-0221/4/04/P04018. URL: https://dx.doi.org/10.1088/1748-0221/4/04/P04018 (visited on 02/29/2024).

[28] JUNO collaboration et al. "Calibration Strategy of the JUNO Experiment". *Journal of High Energy Physics* 2021.3 (Mar. 2021), 4. ISSN: 1029-8479. DOI: 10.1007/JHEP03(2021)004. eprint: 2011.06405[hep-ex,physics:physics]. URL: http://arxiv.org/abs/2011.06405 (visited on 08/10/2023).

[29] Hans Th J. Steiger. *TAO – The Taishan Antineutrino Observatory*. Sept. 21, 2022. DOI: 10.48550/arXiv.2209.10387. eprint: 2209.10387[physics]. URL: http://arxiv.org/abs/2209.10387 (visited on 01/16/2024).

[30] Tao Lin et al. "The Application of SNiPER to the JUNO Simulation". *Journal of Physics: Conference Series* 898.4 (Oct. 2017). Publisher: IOP Publishing, 042029. ISSN: 1742-6596. DOI: 10.1088/1742-6596/898/4/042029. URL: https://dx.doi.org/10.1088/1742-6596/898/4/042029 (visited on 02/27/2024).

[31] S. Agostinelli et al. "Geant4—a simulation toolkit". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 1, 2003), 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688 (visited on 02/27/2024).

[32] J. Allison et al. "Geant4 developments and applications". *IEEE Transactions on Nuclear Science* 53.1 (Feb. 2006). Conference Name: IEEE Transactions on Nuclear Science, 270–278. ISSN: 1558-1578. DOI: 10.1109/TNS.2006.869826. URL: https://ieeexplore.ieee.org/document/1610988?isnumber=33833&arnumber=1610988&count=33&index=7 (visited on 02/27/2024).

[33] J. Allison et al. "Recent developments in Geant4". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (Nov. 1, 2016), 186–225. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.06.125. URL: https://www.sciencedirect.com/science/article/pii/S0168900216306957 (visited on 02/27/2024).

[34] Wenjie Wu, Miao He, Xiang Zhou, and Haoxue Qiao. "A new method of energy reconstruction for large spherical liquid scintillator detectors". *Journal of Instrumentation* 14.3 (Mar. 8, 2019), P03009–P03009. ISSN: 1748-0221. DOI: 10.1088/1748-0221/14/03/P03009. eprint: 1812.01799[hep-ex,physics:physics]. URL: http://arxiv.org/abs/1812.01799 (visited on 07/28/2023).

[35] Guihong Huang et al. "Improving the energy uniformity for large liquid scintillator detectors". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1001 (June 11, 2021), 165287. ISSN: 0168-9002. DOI: 10.1016/j.nima.2021.165287. URL: https://www.sciencedirect.com/science/article/pii/S0168900221002710 (visited on 03/01/2024).

[36] Ziyuan Li et al. "Event vertex and time reconstruction in large volume liquid scintillator detector". *Nuclear Science and Techniques* 32.5 (May 2021), 49. ISSN: 1001-8042, 2210-3147. DOI: 10.1007/s41365-021-00885-z. eprint: 2101.08901[hep-ex,physics:physics]. URL: http://arxiv.org/abs/2101.08901 (visited on 07/28/2023).

[37] Gioacchino Ranucci. "An analytical approach to the evaluation of the pulse shape discrimination properties of scintillators". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 354.2 (Jan. 30, 1995), 389–399. ISSN: 0168-9002. DOI: 10.1016/0168-9002(94)00886-8. URL: https://www.sciencedirect.com/science/article/pii/0168900294008868 (visited on 03/07/2024).

[38] C. Galbiati and K. McCarty. "Time and space reconstruction in optical, non-imaging, scintillator-based particle detectors". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 568.2 (Dec. 1, 2006), 700–709. ISSN: 0168-9002. DOI: 10.1016/j.nima.2006.07.058. URL: https://www.sciencedirect.com/science/article/pii/S0168900206013519 (visited on 03/07/2024).

[39] M. Moszyński and B. Bengtson. "Status of timing with plastic scintillation detectors". *Nuclear Instruments and Methods* 158 (Jan. 1, 1979), 1–31. ISSN: 0029-554X. DOI: 10.1016/S0029-554X(79)90170-8. URL: https://www.sciencedirect.com/science/article/pii/S0029554X79901708 (visited on 03/07/2024).

[40] Gui-Hong Huang, Wei Jiang, Liang-Jian Wen, Yi-Fang Wang, and Wu-Ming Luo. "Data-driven simultaneous vertex and energy reconstruction for large liquid scintillator detectors". *Nuclear Science and Techniques* 34.6 (June 17, 2023), 83. ISSN: 2210-3147. DOI: 10.1007/s41365-023-01240-0. URL: https://doi.org/10.1007/s41365-023-01240-0 (visited on 08/17/2023).

[41] Zhen Qian et al. "Vertex and Energy Reconstruction in JUNO with Machine Learning Methods". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1010 (Sept. 2021), 165527. ISSN: 01689002. DOI: 10.1016/j.nima.2021.165527. eprint: 2101.04839[hep-ex,physics:physics]. URL: http://arxiv.org/abs/2101.04839 (visited on 07/24/2023).

[42] Arsenii Gavrikov, Yury Malyshkin, and Fedor Ratnikov. "Energy reconstruction for large liquid scintillator detectors with machine learning techniques: aggregated features approach". *The European Physical Journal C* 82.11 (Nov. 14, 2022), 1021. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-11004-6. eprint: 2206.09040[physics]. URL: http://arxiv.org/abs/2206.09040 (visited on 07/24/2023).

[43] R. Abbasi et al. "Graph Neural Networks for low-energy event classification & reconstruction in IceCube". *Journal of Instrumentation* 17.11 (Nov. 2022). Publisher: IOP Publishing, P11003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/11/P11003. URL: https://dx.doi.org/10.1088/1748-0221/17/11/P11003 (visited on 04/04/2024).

[44] S. Reck, D. Guderian, G. Vermariën, A. Domi, and on behalf of the KM3NeT collaboration on behalf of the. "Graph neural networks for reconstruction and classification in KM3NeT". *Journal of Instrumentation* 16.10 (Oct. 2021). Publisher: IOP Publishing, C10011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/16/10/C10011. URL: https://dx.doi.org/10.1088/1748-0221/16/10/C10011 (visited on 04/04/2024).

[45] The IceCube collaboration et al. "A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory". *Journal of Instrumentation* 16.7 (July 2021). Publisher: IOP Publishing, P07041. ISSN: 1748-0221. DOI: 10.1088/1748-0221/16/07/P07041. URL: https://dx.doi.org/10.1088/1748-0221/16/07/P07041 (visited on 04/04/2024).

[46] DUNE Collaboration et al. "Neutrino interaction classification with a convolutional neural network in the DUNE far detector". *Physical Review D* 102.9 (Nov. 9, 2020). Publisher: American Physical Society, 092003. DOI: 10.1103/PhysRevD.102.092003. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.092003 (visited on 04/04/2024).

[47] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". *The Astrophysical Journal* 622 (Apr. 1, 2005). ADS Bibcode: 2005ApJ...622..759G, 759–771. ISSN: 0004-637X. DOI: 10.1086/427976. URL: https://ui.adsabs.harvard.edu/abs/2005ApJ...622..759G (visited on 04/04/2024).

[48] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. *Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering*. Feb. 5, 2017. DOI: 10.48550/arXiv.1606.09375.

eprint: 1606.09375[cs, stat]. URL: http://arxiv.org/abs/1606.09375 (visited on 04/04/2024).

[49] JUNO Collaboration et al. “JUNO Physics and Detector”. *Progress in Particle and Nuclear Physics* 123 (Mar. 2022), 103927. ISSN: 01466410. DOI: 10.1016/j.ppnp.2021.103927. eprint: 2104.02565[hep-ex]. URL: http://arxiv.org/abs/2104.02565 (visited on 09/18/2023).

[50] Leo Breiman, Jerome Friedman, R. A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. New York: Chapman and Hall/CRC, Oct. 25, 2017. 368 pp. ISBN: 978-1-315-13947-0. DOI: 10.1201/9781315139470.

[51] Jerome H. Friedman. “Greedy function approximation: A gradient boosting machine.” *The Annals of Statistics* 29.5 (Oct. 2001). Publisher: Institute of Mathematical Statistics, 1189–1232. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1013203451. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full (visited on 04/29/2024).

[52] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts. “What the Frog’s Eye Tells the Frog’s Brain”. *Proceedings of the IRE* 47.11 (Nov. 1959). Conference Name: Proceedings of the IRE, 1940–1951. ISSN: 2162-6634. DOI: 10.1109/JRPROC.1959.287207. URL: https://ieeexplore.ieee.org/document/4065609 (visited on 05/06/2024).

[53] F. Rosenblatt. “The perceptron: A probabilistic model for information storage and organization in the brain”. *Psychological Review* 65.6 (1958). Place: US Publisher: American Psychological Association, 386–408. ISSN: 1939-1471. DOI: 10.1037/h0042519.

[54] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Jan. 29, 2017. DOI: 10.48550/arXiv.1412.6980. eprint: 1412.6980[cs]. URL: http://arxiv.org/abs/1412.6980 (visited on 05/13/2024).

[55] Olga Russakovsky et al. *ImageNet Large Scale Visual Recognition Challenge*. Jan. 29, 2015. DOI: 10.48550/arXiv.1409.0575. eprint: 1409.0575[cs]. URL: http://arxiv.org/abs/1409.0575 (visited on 05/17/2024).

[56] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Apr. 10, 2015. DOI: 10.48550/arXiv.1409.1556. eprint: 1409.1556[cs]. URL: http://arxiv.org/abs/1409.1556 (visited on 05/17/2024).

[57] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. *Neural Message Passing for Quantum Chemistry*. June 12, 2017. DOI: 10.48550/arXiv.1704.01212. eprint: 1704.01212[cs]. URL: http://arxiv.org/abs/1704.01212 (visited on 05/22/2024).

[58] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. *Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting*. Feb. 22, 2018. DOI: 10.48550/arXiv.1707.01926. eprint: 1707.01926[cs, stat]. URL: http://arxiv.org/abs/1707.01926 (visited on 05/22/2024).

[59] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. *Generative Adversarial Networks*. June 10, 2014. DOI: 10.48550/arXiv.1406.2661. eprint: 1406.2661[cs, stat]. URL: http://arxiv.org/abs/1406.2661 (visited on 05/29/2024).

[60] Victor Lebrin. “Towards the Detection of Core-Collapse Supernovae Burst Neutrinos with the 3-inch PMT System of the JUNO Detector”. These de doctorat. Nantes Université, Sept. 5, 2022. URL: https://theses.fr/2022NANU4080 (visited on 05/22/2024).

[61] Dan Cireşan, Ueli Meier, and Juergen Schmidhuber. *Multi-column Deep Neural Networks for Image Classification*. version: 1. Feb. 13, 2012. DOI: 10.48550/arXiv.1202.2745. eprint: 1202.2745[cs]. URL: http://arxiv.org/abs/1202.2745 (visited on 06/27/2024).

[62] R. Abbasi et al. “A Convolutional Neural Network based Cascade Reconstruction for the IceCube Neutrino Observatory”. *Journal of Instrumentation* 16.7 (July 1, 2021), P07041. ISSN: 1748-0221. DOI: 10.1088/1748-0221/16/07/P07041. eprint: 2101.11589[hep-ex]. URL: http://arxiv.org/abs/2101.11589 (visited on 06/27/2024).

[63] D. Maksimović, M. Nieslony, and M. Wurm. “CNNs for enhanced background discrimination in DSNB searches in large-scale water-Gd detectors”. *Journal of Cosmology and Astroparticle Physics* 2021.11 (Nov. 2021). Publisher: IOP Publishing, 051. ISSN: 1475-7516. DOI: 10.1088/

1475-7516/2021/11/051. URL: https://dx.doi.org/10.1088/1475-7516/2021/11/051 (visited on 06/27/2024).

[64] Taco S. Cohen, Mario Geiger, Jonas Koehler, and Max Welling. *Spherical CNNs*. Feb. 25, 2018. DOI: 10.48550/arXiv.1801.10130. eprint: 1801.10130[cs,stat]. URL: http://arxiv.org/abs/1801.10130 (visited on 07/13/2024).

[65] "IEEE Standard for Floating-Point Arithmetic". *IEEE Std 754-2019 (Revision of IEEE 754-2008)* (July 2019). Conference Name: IEEE Std 754-2019 (Revision of IEEE 754-2008), 1–84. DOI: 10.1109/IEEESTD.2019.8766229. URL: https://ieeexplore.ieee.org/document/8766229 (visited on 07/03/2024).

[66] Chuanya Cao et al. "Mass production and characterization of 3-inch PMTs for the JUNO experiment". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1005 (July 2021), 165347. ISSN: 01689002. DOI: 10.1016/j.nima.2021.165347. eprint: 2102.11538[hep-ex,physics:physics]. URL: http://arxiv.org/abs/2102.11538 (visited on 02/08/2024).