



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
DIVISION OF SIGNALS, CONTROL AND ROBOTICS

Energy Reconstruction in KM3NeT/ARCA using Graph Neural Networks

DIPLOMA THESIS

of

Panagiotis Gkotsis

Supervisor: Athanasios Rontogiannis
Associate Professor NTUA

Co-supervisors: Evangelia Drakopoulou
Researcher NCSR Demokritos
Ekaterini Tzamariudaki
Director of Research NCSR Demokritos

COMPUTER VISION, SPEECH COMMUNICATION AND SIGNAL PROCESSING GROUP
Athens, October 2024



National Technical University of Athens
School of Electrical and Computer Engineering
Division of Signals, Control and Robotics
Computer Vision, Speech Communication and Signal Processing Group

Energy Reconstruction in KM3NeT/ARCA using Graph Neural Networks

DIPLOMA THESIS

of

Panagiotis Gkotsis

Supervisor: Athanasios Rontogiannis
Associate Professor NTUA

Co-supervisors: Evangelia Drakopoulou
Researcher NCSR Demokritos
Ekaterini Tzamariudaki
Director of Research NCSR Demokritos

Approved by the Examining Committee on 24th October, 2024.

.....
Athanasios Rontogiannis
Associate Professor NTUA

.....
Petros Maragos
Professor NTUA

.....
Ioannis Kordonis
Assistant Professor NTUA

Athens, October 2024

.....
PANAGIOTIS GKOTSIS
Graduate of Electrical and Computer Engineering NTUA

Copyright © – All rights reserved Panagiotis Gkotsis, 2024.

It is prohibited to copy, store and distribute this work, in whole or in part, for commercial purposes. Reproduction, storage and distribution for a non-profit, educational or research nature are permitted, provided the source of origin is indicated and the present message maintained. Enquiries regarding use for profit should be directed to the author.

The views and conclusions contained in this document are those of the author and should not be construed as representing the official positions of the National Technical University of Athens.

Abstract

Neutrinos, fundamental particles that interact only weakly with matter, are crucial to our understanding of both particle physics and a variety of astrophysical phenomena. Despite their importance, detecting neutrinos is a significant challenge due to their elusive nature. Large-scale water Cherenkov detectors, such as those of the KM3NeT experiment which are located deep in the Mediterranean Sea, are designed to capture the Cherenkov light produced when high-energy neutrinos interact with atomic nuclei in water. These interactions allow physicists to probe the properties of neutrinos and study extremely energetic astrophysical objects and processes. One of the key challenges in these experiments is accurately reconstructing the energy, direction and other characteristics of each neutrino event from the data collected by the detector optical sensors. Traditionally, event reconstruction has relied on classical algorithms, such as maximum likelihood estimation, however the rise of modern Deep Learning methods as powerful means of extracting information from data has demonstrated the potential to improve upon these techniques. In particular, Graph Neural Networks (GNNs) have emerged as a promising approach due to their ability to naturally incorporate the non-grid-like structure and sparsity of neutrino event data. This thesis investigates the application of DynEdge, a particular GNN architecture, to energy reconstruction in the KM3NeT/ARCA detector, using the GraphNeT Deep Learning framework, developed by the IceCube neutrino experiment. In this context, the KM3NeT Deep Learning data format was integrated into GraphNeT, allowing for the first use of this tool in KM3NeT and facilitating its future use within the collaboration. Various configurations of models and training setups are evaluated on two versions of our neutrino event dataset, snapshot and triggered data, in order to determine the optimal approach to energy reconstruction. The results allow for a direct comparison between DynEdge and ParticleNet –the GNN model previously used by KM3NeT– validating the performance of both models. This work also presents significant improvement in the energy reconstruction of low-energy neutrino events, achieved by the implementation of event weighting to balance the energy distribution of the input dataset. This strategy also contributes to the elimination of biases in the training process, leading to a more generalized model. Finally, the thesis explores the challenges of reconstructing high-energy (PeV scale) neutrino events, providing insights into the predictive limitations of the model, which stem from the physical constraints that apply at this energy scale.

Keywords — Neutrinos, KM3NeT, Energy reconstruction, Graph Neural Networks, Sample weighting

Περίληψη

Τα νετρίνα, θεμελιώδη σωματίδια που αλληλεπιδρούν μόνο ασθενώς με την ύλη, είναι ιδιαίτερης σημασίας για την κατανόηση τόσο της σωματιδιακής φυσικής όσο και πολλών αστροφυσικών φαινομένων. Παρόλα αυτά, η ανίχνευση των νετρίνων αποτελεί σημαντική πρόκληση λόγω της ασθενώς αλληλεπιδρώσας φύσης τους. Μεγάλης κλίμακας ανιχνευτές Cherenkov, όπως αυτοί του πειράματος KM3NeT που βρίσκονται σε μεγάλο βάθος στη Μεσόγειο Θάλασσα, έχουν σχεδιαστεί ώστε να συλλαμβάνουν την ακτινοβολία Cherenkov που παράγεται όταν νετρίνα υψηλής ενέργειας συγκρούονται με ατομικούς πυρήνες στο νερό. Αυτές οι αλληλεπιδράσεις επιτρέπουν στους φυσικούς να διερευνούν τις ιδιότητες των νετρίνων και να μελετούν αστροφυσικά αντικείμενα και διεργασίες εξαιρετικά υψηλών ενεργειών. Μία από τις βασικές προκλήσεις σε αυτά τα πειράματα είναι η ακριβής ανακατασκευή της ενέργειας, της κατεύθυνσης, αλλά και άλλων χαρακτηριστικών του κάθε γεγονότος νετρίνου, από τα δεδομένα που συλλέγονται από τους οπτικούς αισθητήρες του ανιχνευτή. Παραδοσιακά, η ανακατασκευή γεγονότων βασίζεται σε κλασικούς αλγόριθμους, όπως η εκτίμηση μέγιστης πιθανοφάνειας, ωστόσο η ανάδειξη των σύγχρονων μεθόδων Βαθιάς Μάθησης ως ισχυρών μέσων εξαγωγής πληροφοριών από δεδομένα έχει καταδείξει τη δυνατότητα βελτίωσης αυτών των τεχνικών. Ειδικότερα, τα Νευρωνικά Δίκτυα Γράφων (NΔΓ) έχουν αναδειχθεί ως μια πολλά υποσχόμενη προσέγγιση λόγω της ικανότητάς τους να ενσωματώνουν με φυσικό τρόπο τη μη πλεγματική δομή και την αραιότητα των δεδομένων γεγονότων νετρίνων. Η παρούσα εργασία διερευνά την εφαρμογή μιας συγκεκριμένης αρχιτεκτονικής NΔΓ, του μοντέλου DynEdge, στην ανακατασκευή ενέργειας στον ανιχνευτή KM3NeT/ARCA, χρησιμοποιώντας το εργαλείο Βαθιάς Μάθησης GraphNeT, που αναπτύχθηκε από το πείραμα νετρίνων IceCube. Σε αυτή τη βάση, η μορφή δεδομένων του πειράματος KM3NeT ενσωματώθηκε στο GraphNeT, επιτρέποντας την πρώτη χρήση αυτού του εργαλείου στο πείραμα KM3NeT και διευκολύνοντας τη μελλοντική του χρήση στο πλαίσιο του πειράματος. Διάφορες διαμορφώσεις μοντέλων και ρυθμίσεων εκπαίδευσης αξιολογούνται σε δύο εκδοχές του συνόλου δεδομένων γεγονότων νετρίνων, τα δεδομένα στιγμιότυπου και τα δεδομένα σκανδαλισμού, προκειμένου να προσδιοριστεί η βέλτιστη προσέγγιση για την ανακατασκευή της ενέργειας. Τα αποτελέσματα επιτρέπουν την άμεση σύγκριση μεταξύ του DynEdge και του ParticleNet –του μοντέλου NΔΓ που είχε προηγουμένως χρησιμοποιηθεί από το πείραμα KM3NeT– επιβεβαιώνοντας την απόδοση και των δύο μοντέλων. Στην εργασία αυτή παρουσιάζεται ακόμη σημαντική βελτίωση στην ανακατασκευή της ενέργειας γεγονότων νετρίνων χαμηλής ενέργειας, η οποία επιτυγχάνεται με την εφαρμογή της τεχνικής της στάθμισης γεγονότων για την εξισορρόπηση της ενεργειακής κατανομής του συνόλου δεδομένων εισόδου. Η στρατηγική αυτή συμβάλλει επίσης στον περιορισμό της μεροληψίας κατά τη διαδικασία εκπαίδευσης, οδηγώντας σε ένα πιο γενικευμένο μοντέλο. Τέλος, η εργασία διερευνά τις προκλήσεις της ανακατασκευής γεγονότων νετρίνων υψηλής ενέργειας (κλίμακα PeV), παρέχοντας πληροφορίες για τους περιορισμούς στις προβλέψεις του μοντέλου, οι οποίοι προέρχονται από τους ενυπάρχοντες φυσικούς περιορισμούς σε αυτήν την ενεργειακή κλίμακα.

Λέξεις Κλειδιά — Νετρίνα, KM3NeT, Ανακατασκευή ενέργειας, Νευρωνικά Δίκτυα Γράφων, Στάθμιση δειγμάτων

Acknowledgments

To begin with, I would like to express my gratitude to my supervisor, Athanasios Rontogiannis, who not only agreed to supervise this thesis, but also offered his valuable insights and constant encouragement, both of which proved essential for its successful completion.

This thesis was carried out at the Institute of Nuclear and Particle Physics of NCSR “Demokritos”. I am deeply grateful to my co-supervisors, Dr. Evangelia Drakopoulou & Dr. Ekaterini Tzamariudaki, as well as Dr. Christos Markou, for the opportunity to conduct this thesis within the Astroparticle Physics group. The continual, daily guidance and support from my co-supervisors were indispensable and this work would not be possible without them. I would also like to extend my heartfelt appreciation to all members of the group, for providing an excellent and enjoyable collaborative environment.

Special thanks are due to Dr. George Giannakopoulos, of the Institute of Informatics and Telecommunications of NCSR “Demokritos”, for offering his invaluable knowledge and advice to aid with the application of GNNs to the problem of energy reconstruction.

Finally, I would like to thank my friends and family, whom I love more than I can convey.

Panos Gkotsis
October 2024

Ευχαριστίες

Θα ήθελα αρχικά να εκφράσω την ευγνωμοσύνη μου στον επιβλέποντα καθηγητή μου, Αθανάσιο Ροντογιάννη, ο οποίος όχι μόνο συμφώνησε να επιβλέψει αυτήν τη διπλωματική εργασία, αλλά επίσης προσέφερε τις πολύτιμες γνώσεις και τη συνεχή ενθάρρυνσή του, οι οποίες αποδείχθηκαν απαραίτητες για την επιτυχή ολοκλήρωσή της.

Η παρούσα εργασία πραγματοποιήθηκε στο Ινστιτούτο Πυρηνικής και Σωματιδιακής Φυσικής του ΕΚΕΦΕ «Δημόκριτος». Είμαι βαθύτατα ευγνώμων προς τις συνεπιβλέπουσές μου, Δρ. Ευαγγελία Δρακοπούλου και Δρ. Αικατερίνη Τζαμαριουδάκη, καθώς και προς τον Δρ. Χρήστο Μάρκου, για την ευκαιρία να διεξαγάγω αυτήν την εργασία στο πλαίσιο της ομάδας Αστροσωματιδιακής Φυσικής. Η αδιάλειπτη, καθημερινή καθοδήγηση και υποστήριξη από τις συνεπιβλέπουσές μου ήταν αναντικατάστατες και αυτή η εργασία δεν θα ήταν δυνατή χωρίς εκείνες. Θα ήθελα επίσης να εκφράσω τις θερμές μου ευχαριστίες σε όλα τα μέλη της ομάδας, για τη δημιουργία ενός εξαιρετικού και ευχάριστου συνεργατικού περιβάλλοντος.

Ιδιαίτερες ευχαριστίες οφείλονται στον Δρ. Γιώργο Γιαννακόπουλο, του Ινστιτούτου Πληροφορικής και Τηλεπικοινωνιών του ΕΚΕΦΕ «Δημόκριτος», για την προσφορά των ανεκτίμητων γνώσεων και συμβουλών του προς την εφαρμογή των ΝΔΓ στο πρόβλημα της ανακατασκευής ενέργειας.

Τέλος, θα ήθελα να ευχαριστήσω τους φίλους και την οικογένειά μου, τους οποίους αγαπώ περισσότερο από ό,τι μπορώ να εκφράσω.

Πάνος Γκότσης
Οκτώβριος 2024

Contents

Contents	13
List of Figures	15
List of Tables	19
Εκτεταμένη Περίληψη στα Ελληνικά	21
1 Introduction	45
1.1 Overview and Motivations	46
1.2 Contributions	46
1.3 Thesis Outline	47
2 Neutrino Physics	49
2.1 Neutrinos in the Standard Model	50
2.1.1 The Standard Model of Particle Physics	50
2.1.2 Neutrinos	50
2.2 Neutrino Sources	51
2.2.1 Atmospheric neutrinos	52
2.2.2 Astrophysical neutrinos	53
2.3 Neutrino Detection	54
2.3.1 Neutrino interactions with matter	54
2.3.2 Cherenkov neutrino detection	55
2.3.3 Neutrino event topologies	55
3 KM3NeT	59
3.1 The KM3NeT Experiment	60
3.1.1 KM3NeT/ARCA	60
3.1.2 KM3NeT/ORCA	60
3.2 Detector Design	61
3.3 Data Acquisition and Triggering	62
3.4 Background Sources	63
3.4.1 Noise background	63
3.4.2 Physics background	63
3.5 Event Reconstruction	64
4 Deep Learning & Graph Neural Networks	67
4.1 Deep Learning	68
4.1.1 Architecture of Deep Neural Networks	68
4.1.2 Training of Deep Neural Networks	69
4.2 Graph Neural Networks	72
4.2.1 Graph structure	72
4.2.2 Edge convolution	73
4.2.3 Network architecture	73

4.2.4	Software tools	74
4.3	Overview of DL &GNN Applications in Neutrino Telescope Event Reconstruction	75
5	Energy Reconstruction with Graph Neural Networks	77
5.1	The KM3NeT DL Data Format	78
5.1.1	Integration of the KM3NeT DL data format in GraphNeT	78
5.2	GNN Trainings with GraphNeT in ARCA21	79
5.2.1	Dataset and data preprocessing	79
5.2.2	Model implementation	80
5.2.3	Training details ¶meters	83
5.2.4	Training configurations	84
5.3	Training Results and Comparisons	85
5.3.1	Snapshot hits	85
5.3.2	Triggered hits	91
5.3.3	Event weighting	92
5.4	Investigation of Model Predictions	94
6	Conclusion	99
6.1	Summary	100
6.2	Future Work	100
	Bibliography	103

List of Figures

1	Ακτινοβολία Cherenkov (με μπλε χρώμα) που εκπέμπεται σε γωνία Cherenkov θ από την τροχιά ενός φορτισμένου σωματιδίου, το οποίο κινείται οριζόντια με ταχύτητα $v = \beta c$ εντός μέσου με δείκτη διάθλασης n . Από [27].	22
2	Επισκόπηση των τοπολογιών γεγονότων που παράγονται στις αλληλεπιδράσεις νετρίνων – νουκλεονίων CC & NC. Για τις αλληλεπιδράσεις NC, η τοπολογία γεγονότων είναι η ίδια και για τις τρεις γεύσεις νετρίνων. Από [34].	23
3	Καλλιτεχνική απεικόνιση ενός ανιχνευτή του πειράματος KM3NeT στον θαλάσσιο πυθμένα. Διακρίνονται τα DUs, καθένα από τα οποία περιλαμβάνει 18 DOMs. Παραχώρηση του πειράματος KM3NeT.	24
4	Αριστερά: Νευρώνας με μεροληψία b και συνάρτηση ενεργοποίησης $f(\cdot)$ που δέχεται εισόδους x_i με βάρη w_i και παράγει την έξοδο του όπως στην Εξ. 4. Δεξιά: Σχηματική απεικόνιση Βαθέως ΝΔ. Από [51].	26
5	Αριστερά: Κατανομή των hits ενός γεγονότος στο πείραμα KM3NeT (θέση z συναρτήσει του χρόνου t). Δεξιά: Αναπαράσταση του ίδιου γεγονότος με χρήση γράφου. Κάθε hit είναι ένας κόμβος, με ακμές προς τους 8 πλησιέστερους γείτονές του, οι οποίοι υπολογίζονται από την Εξ. 6. Από [59].	28
6	Αρχιτεκτονική του δικτύου DynEdge. Ο γράφος εισόδου θεωρείται εδώ ότι έχει n κόμβους και 6 γνωρίσματα ανά κόμβο. Από [61].	30
7	Αριστερά: Δομή του μπλοκ EdgeConv που χρησιμοποιείται στο ParticleNet. Δεξιά: Αρχιτεκτονική του δικτύου ParticleNet όπως παρουσιάζεται στο [74].	31
8	Ενεργειακή κατανομή των νετρίνων στα σύνολα δεδομένων εκπαίδευσης, επικύρωσης και δοκιμής, κανονικοποιημένη έτσι ώστε το εμβαδόν κάτω από κάθε καμπύλη να είναι 1. Ο άξονας x είναι ο εκθέτης της ενέργειας, με κάθε bin να καλύπτει εύρος 0,1. Όλα τα σύνολα δεδομένων παρουσιάζουν την ίδια ενεργειακή κατανομή.	33
9	Διαγράμματα διασποράς της προβλεπόμενης συναρτήσει της πραγματικής ενέργειας. Η μαύρη διαγώνια γραμμή δείχνει την ιδανική ανακατασκευή. Μια δεύτερη μαύρη γραμμή υποδεικνύει τη διάμεσο της προβλεπόμενης τιμής για κάθε bin πραγματικής ενέργειας.	37
10	Διαγράμματα ποσοστημορίων της προβλεπόμενης συναρτήσει της πραγματικής ενέργειας. Σε όλα τα διαγράμματα, η κόκκινη διαγώνια γραμμή υποδεικνύει την ιδανική ανακατασκευή. Η διάμεσος και τα διαστήματα εμπιστοσύνης 68% & 95% των προβλέψεων για κάθε bin πραγματικής ενέργειας απεικονίζονται με μαύρο, μπλε και ανοιχτό γκρι αντίστοιχα.	37
11	Συγκρίσεις μεταξύ των διαφόρων διαμορφώσεων εκπαίδευσης του Πίνακα 1. Σε όλα τα διαγράμματα συγκρίνονται τα διαστήματα εμπιστοσύνης 68%.	39
12	Αριστερά: Διάγραμμα διασποράς της προβλεπόμενης συναρτήσει της πραγματικής ενέργειας για την T2. Μέσον: Διάγραμμα ποσοστημορίων της προβλεπόμενης συναρτήσει της πραγματικής ενέργειας για την T2. Δεξιά: Διάγραμμα σύγκρισης μεταξύ της εκπαίδευσης T5 με hits στιγμιότυπου και της εκπαίδευσης T2 με hits σκανδαλισμού. Συγκρίνονται τα διαστήματα εμπιστοσύνης 68% για την κάθε εκπαίδευση.	40
13	Αριστερά: Διάγραμμα διασποράς για την σταθμισμένη εκπαίδευση. Μέσον: Διάγραμμα ποσοστημορίων για την σταθμισμένη εκπαίδευση. Δεξιά: Διάγραμμα σύγκρισης μεταξύ της σταθμισμένης και της μη σταθμισμένης εκπαίδευσης (hits στιγμιότυπου T5). Συγκρίνονται τα διαστήματα εμπιστοσύνης 68%.	41

14	Ενεργειακές τομές εύρους 0,2 σε $\log_{10} E_{true}$ (μπλε) και οι αντίστοιχες προβλέψεις του ΝΔΓ (πορτοκαλί) για μη εκκινόντα γεγονότα στην ενεργειακή περιοχή $10^6 - 10^8$ GeV.	42
15	Αριθμός hits ανά γεγονός συναρτήσει της πραγματικής ενέργειας των νετρίνων (αριστερά) και της προβλεπόμενης ενέργειας από το ΝΔΓ (δεξιά). Γεγονότα με περισσότερα από 1.000 hits εμφανίζονται με πορτοκαλί χρώμα. Το δίκτυο τα αναγνωρίζει σωστά ως γεγονότα υψηλής ενέργειας.	43
16	Ενεργειακές τομές εύρους 0,2 σε $\log_{10} E_{true}$ (μπλε) και οι αντίστοιχες προβλέψεις του ΝΔΓ (πορτοκαλί) για μη εκκινόντα γεγονότα με λιγότερα από 1.000 hits στην ενεργειακή περιοχή $10^6 - 10^8$ GeV.	43
2.1	Particles in the SM of Particle Physics. The faint lines indicate which fermions (<i>i.e.</i> quarks & leptons) interact with each force-mediating boson. Antiparticles are not shown. From [2].	50
2.2	Measured and expected neutrino fluxes from different neutrino sources as a function of neutrino energy. From [11].	51
2.3	Atmospheric neutrino production in a cosmic ray cascade, caused by the interaction of the primary cosmic ray with an atmospheric nucleus, over the underground Super-Kamiokande neutrino detector. From [12].	52
2.4	Neutrino-matter interaction cross sections in the TeV and PeV regime (neutrino-nucleus CC/NC & Glashow resonance). From [25].	54
2.5	Cherenkov radiation (in blue) emitted at Cherenkov angle θ from the track of a charged particle which is moving horizontally at velocity $v = \beta c$ inside a medium with refractive index n . From [27].	55
2.6	Overview of the event signatures produced in neutrino-nucleon CC & NC interactions. For NC interactions, the event signature is the same for all three neutrino flavors. From [34].	56
3.1	Overview of the KM3NeT Collaboration member institutes' locations (white) and detector sites (yellow). Apart from the two detector sites for ARCA and ORCA, a third suitable site is shown off the coast of Pylos, Greece. Courtesy of KM3NeT.	60
3.2	A KM3NeT multi-PMT Digital Optical Module (DOM). From [28].	61
3.3	Artist's impression of a KM3NeT detector on the sea floor. The DUs, consisting of a base, 18 DOMs and a buoy, can clearly be seen. Courtesy of KM3NeT.	62
3.4	Flux of atmospheric muons (at different water depths) and atmospheric neutrino-induced muons (of different energies) as a function of the cosine of the zenith angle θ . For downward angles, <i>i.e.</i> $\cos(\theta) > 0$, the atmospheric muon contribution dominates the total flux. From [41].	64
4.1	Left: Neuron with bias b and activation function $f(\cdot)$ receiving inputs x_i through connections with weights w_i and producing its output as in Eq. 4.1. Right: Schematic depiction of a Deep Neural Network. From [51].	68
4.2	Left: Hit distribution of a KM3NeT event (position z vs. time t). Right: Graph representation of the same event. Each hit is now a graph node, with edges to its 8 nearest neighbors, calculated using Eq. 4.14. From [59].	72
4.3	Architecture of the DynEdge network. The input graph is assumed here to have n nodes and 6 features per node. From [61].	74
4.4	Left: Structure of the EdgeConv block used in ParticleNet. Right: Architecture of the ParticleNet network as presented in [74].	75
5.1	Neutrino energy distribution in the training, validation and test datasets, normalized so that the area under each curve is equal to 1. The x -axis is the energy exponent, with each bin covering a 0.1 range. All datasets exhibit the same energy distribution.	80
5.2	Comparison of the ReLU and Softplus activation functions.	81
5.3	Comparison of the MSE and Log Cosh loss functions.	82
5.4	Predicted with respect to true energy scatter plots for all entries of Table 5.3. In all plots, the black diagonal line indicates the ideal reconstruction. A second black line indicates the median predicted value for each true energy bin. The events are unweighted.	86

5.5	Quantile plots for all entries of Table 5.3. In all plots, the red diagonal line indicates the ideal reconstruction. The median and the 68% & 95% confidence intervals of the predictions for each true energy bin are shown in black, blue and light gray respectively.	87
5.6	Comparisons between the various training configurations of Table 5.1. In all plots, the 68% confidence intervals for each training are compared.	89
5.7	Comparison between GraphNeT training T6 and the best OrcaNet training. The standard reconstruction is also shown. The 68% confidence intervals are compared.	90
5.8	Predicted with respect to true energy scatter plots for all entries of Table 5.4. In all plots, the black diagonal line indicates the ideal reconstruction. A second black line indicates the median predicted value for each true energy bin. The events are unweighted.	91
5.9	Quantile plots for all entries of Table 5.4. In all plots, the red diagonal line indicates the ideal reconstruction. The median and the 68% & 95% confidence intervals of the predictions for each true energy bin are shown in black, blue and light gray respectively.	91
5.10	Comparison between snapshot hits training T5 and the best triggered hits training (T2). The 68% confidence intervals for each training are compared.	92
5.11	Scatter and quantile plots for the weighted training. The diagonal indicates the ideal reconstruction. The median predicted value for each true energy bin is shown in both plots in black. On the quantile plot, the 68% & 95% confidence intervals of the predictions for each true energy bin are shown in blue and light gray respectively.	93
5.12	Comparisons between the weighted and unweighted training (snapshot hit T5). The 68% confidence intervals for each training are compared.	93
5.13	Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events in the energy range $10^5 - 10^6$ GeV.	95
5.14	Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events in the energy range $10^6 - 10^7$ GeV.	96
5.15	Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events in the energy range $10^7 - 10^8$ GeV.	96
5.16	Number of hits per event with respect to the true neutrino energy (left) and the GNN predicted energy (right). Events with more than 1,000 hits are shown in orange. The network correctly identifies these as high-energy events.	97
5.17	Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events with fewer than 1,000 hits in the energy range $10^5 - 10^6$ GeV.	97
5.18	Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events with fewer than 1,000 hits in the energy range $10^6 - 10^7$ GeV.	98
5.19	Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events with fewer than 1,000 hits in the energy range $10^7 - 10^8$ GeV.	98

List of Tables

1	Διαμορφώσεις εκπαίδευσης (hits στιγμιοτύπου).	36
2	Διαμορφώσεις εκπαίδευσης (hits σκανδαλισμού).	36
3	Αποτελέσματα εκπαίδευσεων (hits στιγμιοτύπου). Οι έντονες τιμές υποδεικνύουν το καλύτερο αποτέλεσμα στο GraphNeT για κάθε μετρική. Τα αποτελέσματα για την τυπική ανακατασκευή («std reco») και το καλύτερο μοντέλο στο OrcaNet, στο ίδιο σύνολο δεδομένων, παρουσιάζονται επίσης για σύγκριση.	37
4	Αποτελέσματα εκπαίδευσης (hits σκανδαλισμού). Οι έντονες τιμές υποδεικνύουν το καλύτερο αποτέλεσμα για κάθε μετρική.	39
5.1	Training configurations (snapshot hits).	84
5.2	Training configurations (triggered hits).	85
5.3	Training results (snapshot hits). Bold values indicate the best GraphNeT result for each metric. The results for the standard reconstruction (“std reco”) and the best OrcaNet model in the same dataset are also shown for comparison.	85
5.4	Training results (triggered hits). Bold values indicate the best result for each metric.	91

Εκτεταμένη Περίληψη στα Ελληνικά

Φυσική των Νετρίνων

Νετρίνα στο Καθιερωμένο Πρότυπο

Τα νετρίνα είναι στοιχειώδη σωματίδια, φερμιόνια με σπιν 1/2 που ανήκουν στην οικογένεια των λεπτονίων. Υπάρχουν τρία είδη ή γεύσεις νετρίνων. Κάθε γεύση συνδέεται με το αντίστοιχο φορτισμένο λεπτόνιο (ηλεκτρόνιο, μιονίο ή ταυ). Έτσι, οι τρεις γεύσεις νετρίνων είναι το νετρίνο ηλεκτρονίου ν_e , το νετρίνο μιονίου ν_μ και το νετρίνο ταυ ν_τ . Τα αντίστοιχα αντισωματίδια τους συμβολίζονται με $\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$ αντίστοιχα. Όπως όλα τα λεπτόνια, τα νετρίνα δεν αλληλεπιδρούν με την ισχυρή δύναμη και, δεδομένου ότι είναι ηλεκτρικά ουδέτερα, δεν αλληλεπιδρούν ούτε ηλεκτρομαγνητικά. Συνεπώς, στο Καθιερωμένο Πρότυπο της Σωματιδιακής Φυσικής, τα νετρίνα συμμετέχουν μόνο στην ασθενή αλληλεπίδραση, η οποία διαδίδεται από τα μποζόνια W^\pm και Z .

Πηγές Νετρίνων

Τα νετρίνα είναι τα πιο άφθονα γνωστά σωματίδια μη-μηδενικής μάζας στο Σύμπαν. Υπάρχουν διάφορα είδη πηγών νετρίνων, όμως στην παρούσα εργασία μας ενδιαφέρουν νετρίνα υψηλής ενέργειας στην περιοχή TeV/PeV^1 , στην οποία οι σημαντικότερες συνεισφορές προέρχονται από τα ατμοσφαιρικά και τα αστροφυσικά νετρίνα.

- **Ατμοσφαιρικά νετρίνα:** Τα ατμοσφαιρικά νετρίνα παράγονται κατά τις αλληλεπιδράσεις της κοσμικής ακτινοβολίας με ατομικούς πυρήνες στη γήινη ατμόσφαιρα, ως προϊόντα των καταγισμών σωματιδίων που αυτές προκαλούν. Η ροή ατμοσφαιρικών νετρίνων είναι η κυρίαρχη ροή νετρίνων στην ενεργειακή κλίμακα των GeV και TeV [12, 13].
- **Αστροφυσικά νετρίνα:** Τα αστροφυσικά νετρίνα είναι υψηλής ενέργειας νετρίνα που παράγονται ως υποπροϊόν της επιτάχυνσης των αδρονίων στις πηγές κοσμικής ακτινοβολίας [15]. Η ροή των αστροφυσικών νετρίνων είναι υποδεέστερη της ροής των ατμοσφαιρικών νετρίνων κάτω από τα 100 TeV , ωστόσο τα αστροφυσικά νετρίνα κυριαρχούν έναντι των ατμοσφαιρικών στην ενεργειακή κλίμακα των PeV . Σε αντίθεση με την κοσμική ακτινοβολία και τα φωτόνια, τα νετρίνα ταξιδεύουν σχεδόν ανεμπόδιστα στο διάστημα. Αυτή η ιδιότητα τα καθιστά ιδανικούς κοσμικούς αγγελιοφόρους, επιτρέποντας τη δημιουργία της *αστρονομίας νετρίνων*, η οποία αποτελεί τμήμα της *αστρονομία πολλαπλών αγγελιοφόρων* [19–21].

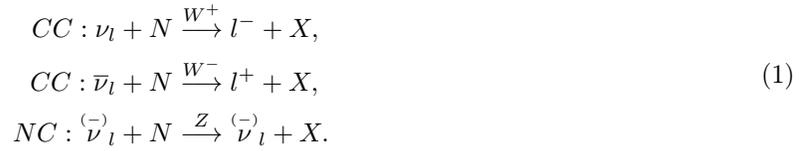
Ανίχνευση Νετρίνων

Αλληλεπίδραση νετρίνων με την ύλη

Ο κύριος μηχανισμός με τον οποίο τα νετρίνα υψηλής ενέργειας αλληλεπιδρούν με την ύλη είναι η βαθιά ανελαστική σκέδαση μεταξύ αυτών και των νουκλεονίων (δηλαδή πρωτονίων ή νετρονίων στους ατομικούς πυρήνες): το νετρίνο αλληλεπιδρά με ένα μεμονωμένο κουάρκ του νουκλεονίου, αποθέτοντάς του αρκετή ενέργεια ώστε να προκαλέσει την καταστροφή του νουκλεονίου και την παραγωγή ενός αδρονικού² καταγισμού [22]. Υπάρχουν δύο τύποι αλληλεπιδράσεων νετρίνων-νουκλεονίων: αλληλεπιδράσεις φορτισμένου ρεύματος (*charged current* – CC) μέσω των μποζονίων W^\pm και ουδέτερου ρεύματος (*neutral current* – NC) μέσω του μποζονίου Z . Αυτές οι αλληλεπιδράσεις αναπαρίστανται παρακάτω, όπου ως N αναφέρεται στο νουκλεόνιο και ως X ο αδρονικός καταγισμός, ενώ $l = e, \mu, \tau$:

¹ $1\text{eV} = 1,602 \times 10^{-19}\text{J}$

²Αδρόνια ονομάζονται τα σωματίδια που αποτελούνται από κουαρκ.

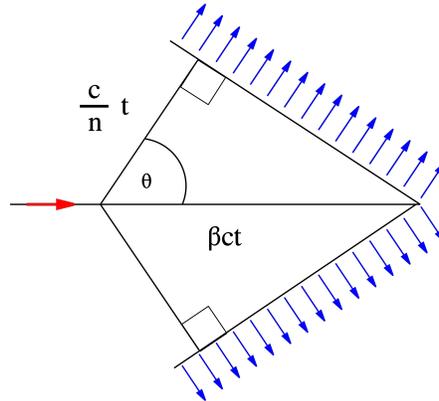


Ανίχνευση νετρίνων μέσω ακτινοβολίας Cherenkov

Η ανίχνευση νετρίνων στα παρατηρητήρια νετρίνων βασίζεται στο φαινόμενο της ακτινοβολίας Cherenkov. Η ακτινοβολία Cherenkov παράγεται από φορτισμένα σωματίδια που κινούνται εντός διηλεκτρικού μέσου ταχύτερα από την ταχύτητα φάσης του φωτός σε αυτό το μέσο [26]. Η ακτινοβολία εκπέμπεται σε ένα κωνικό μέτωπο κύματος σε χαρακτηριστική γωνία από την τροχιά του σωματιδίου (γωνία Cherenkov):

$$\cos(\theta_c) = \frac{1}{\beta n},
 \tag{2}$$

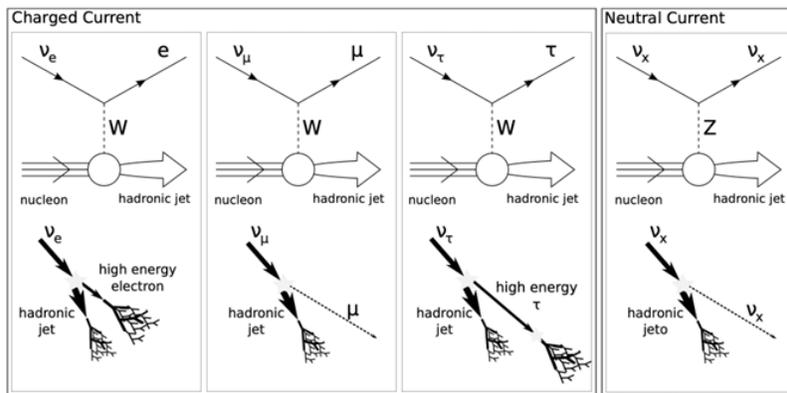
όπου n είναι ο δείκτης διάθλασης του μέσου και $\beta = \frac{v}{c}$, v η ταχύτητα του φορτισμένου σωματιδίου. Το φαινόμενο απεικονίζεται στο Σχ. 1. Για σχετικιστικά ($\beta \approx 1$) φορτισμένα σωματίδια σε νερό, $\theta_c \approx 42^\circ$. Όπως προκύπτει από την εξίσωση 1, όταν τα νετρίνα αλληλεπιδρούν με την ύλη, παράγονται φορτισμένα δευτερογενή σωματίδια. Αν αυτά ή τα προϊόντα διάσπασής τους κινούνται με υπερφωτεινές ταχύτητες στο μέσο, θα εκπέμψουν ακτινοβολία Cherenkov, η οποία μπορεί να ανιχνευθεί με συστοιχίες φωτοπολλαπλασιαστών (PMTs). Προκειμένου να ανιχνεύσουν φωτόνια Cherenkov, τα παρατηρητήρια νετρίνων πρέπει να λειτουργούν σε συνθήκες σκότους, να προστατεύονται από τα ατμοσφαιρικά μόνια, τα οποία επίσης παράγουν ακτινοβολία Cherenkov, αλλά και να αρκετά μεγάλα. Έτσι, συνήθως κατασκευάζονται υποθαλάσσια (π.χ. KM3NeT [28]), εντός των πάγων της Ανταρκτικής (π.χ. IceCube [31]) ή σε υπόγειες δεξαμενές νερού (π.χ. Super-Kamiokande [32]).



Σχήμα 1: Ακτινοβολία Cherenkov (με μπλε χρώμα) που εκπέμπεται σε γωνία Cherenkov θ από την τροχιά ενός φορτισμένου σωματιδίου, το οποίο κινείται οριζόντια με ταχύτητα $v = \beta c$ εντός μέσου με δείκτη διάθλασης n . Από [27].

Τοπολογίες γεγονότων νετρίνων

Οι διαφορετικές αλληλεπιδράσεις νετρίνων-νουκλεονίων οδηγούν σε διαφορετικά προϊόντα σωματιδίων και, επομένως, σε διαφορετικές τοπολογίες γεγονότων στους ανιχνευτές νετρίνων. Υπάρχουν δύο βασικά είδη τοπολογιών που μπορούν να αποτελέσουν μέρος ενός γεγονότος νετρίνου: οι τροχίες και οι καταγισμοί. Μια τροχιά αναφέρεται σε ένα μεμονωμένο φορτισμένο σωματίδιο που παράγει τον χαρακτηριστικό κώνο ακτινοβολίας Cherenkov. Το μόνο σωματίδιο που παράγει τοπολογία τροχιάς στους ανιχνευτές νετρίνων είναι το μόνιο. Ο καταγισμός σωματιδίων μπορεί να είναι είτε αδρονικός (αποτελείται κυρίως από αδρόνια) είτε ηλεκτρομαγνητικός (αποτελείται από ηλεκτρόνια και φωτόνια). Τα φορτισμένα σωματίδια που παράγονται σε καταγισμούς εκπέμπουν επίσης ακτινοβολία Cherenkov (Εξ. 2) όταν κινούνται με υπερφωτεινές ταχύτητες, ωστόσο το συλλογικό μοτίβο εκπομπής των σωματιδίων σε έναν καταγισμό είναι ένα σφαιρικό μέτωπο κύματος [15].



Σχήμα 2: Επισκόπηση των τοπολογιών γεγονότων που παράγονται στις αλληλεπιδράσεις νετρίνων – νουκλεονίων CC & NC. Για τις αλληλεπιδράσεις NC, η τοπολογία γεγονότων είναι η ίδια και για τις τρεις γεύσεις νετρίνων. Από [34].

Όλες οι αλληλεπιδράσεις νετρίνων-νουκλεονίων (Εξ. 1) περιέχουν μια συνιστώσα αδρονικού καταγισμού, όπως φαίνεται στο Σχ. 2. Αυτή η συνιστώσα προέρχεται από το νουκλεόνιο ως αποτέλεσμα της ενέργειας που εναποτίθεται σε αυτό κατά την αλληλεπίδραση. Για τις αλληλεπιδράσεις NC αυτή είναι η μόνη διαθέσιμη υπογραφή του γεγονότος. Στις CC αλληλεπιδράσεις (αντι)νετρίνων ηλεκτρονίου, το εξερχόμενο e^\mp θα παράγει επιπλέον έναν ηλεκτρομαγνητικό καταγισμό. Οι CC αλληλεπιδράσεις (αντι)νετρίνων ταυ δίνουν πιο πολύπλοκες τοπολογίες, οι οποίες δεν μας αφορούν στην παρούσα εργασία. Στις CC αλληλεπιδράσεις (αντι)νετρίνων μιονίου, το εξερχόμενο μ^\mp αποτελεί μια τροχιά. Στο υπόλοιπο του κειμένου, θα επικεντρωθούμε σε γεγονότα τροχιάς από αλληλεπιδράσεις $(\bar{\nu}_\mu)$ CC.

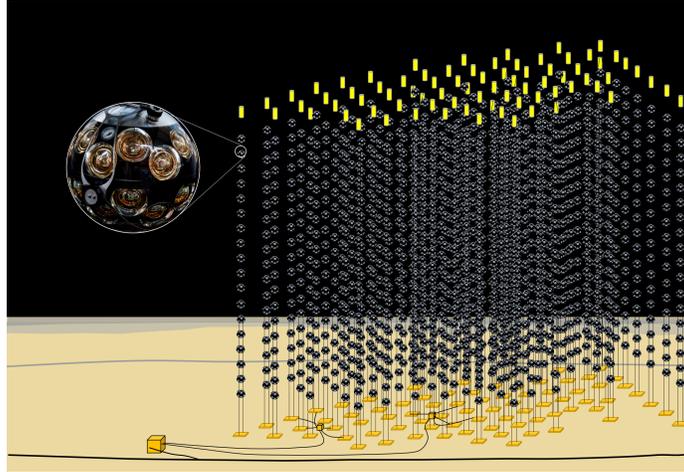
Το Πείραμα KM3NeT

Το *Cubic Kilometer Neutrino Telescope* (KM3NeT [28]) είναι μια διεθνής ερευνητική συνεργασία που αναπτύσσει ανιχνευτές νετρίνων νέας γενιάς στα βάθη της Μεσογείου. Το πείραμα KM3NeT αποτελείται από δύο υποθαλάσσιους ανιχνευτές Cherenkov, οι οποίοι χρησιμοποιούν την ίδια αρχή ανίχνευσης, αλλά ελαφρώς διαφορετική ανιχνευτική διάταξη. Οι δύο ανιχνευτές ονομάζονται KM3NeT/ARCA (*Astrophysical Research with Cosmics in the Abyss*) και KM3NeT/ORCA (*Oscillations Research with Cosmics in the Abyss*):

- **KM3NeT/ARCA:** Ο ανιχνευτής ARCA είναι ποντισμένος περίπου 100 km νοτιοανατολικά της Σικελίας σε βάθος 3.500 m. Ο επιστημονικός στόχος του ανιχνευτή ARCA είναι να παρατηρήσει νετρίνα αστροφυσικής προέλευσης. Για τον σκοπό αυτό, αποτελείται από αραίες συστοιχίες φωτοανιχνευτών, που σχηματίζουν έναν ανιχνευτή όγκου 1 GTon, ευαίσθητο στα νετρίνα στην ενεργειακή περιοχή TeV/PeV, η οποία κυριαρχείται από τα αστροφυσικά νετρίνα.
- **KM3NeT/ORCA:** Ο ανιχνευτής ORCA βρίσκεται περίπου 40 km ανοικτά της Τουλόν στη Γαλλία σε βάθος 2.450 m. Ο επιστημονικός στόχος του ανιχνευτή ORCA σχετίζεται με το φαινόμενο των ταλαντώσεων νετρίνων. Για τον σκοπό αυτό, επικεντρώνεται στα ατμοσφαιρικά νετρίνα στην ενεργειακή περιοχή GeV, χρησιμοποιώντας πολύ πυκνότερες συστοιχίες φωτοανιχνευτών από τον ανιχνευτή ARCA, οι οποίες σχηματίζουν έναν ανιχνευτή όγκου περίπου 7 MTons.

Ανιχνευτική Διάταξη

Ως ανιχνευτές Cherenkov, οι ανιχνευτές του πειράματος KM3NeT χρησιμοποιούν συστοιχίες φωτοανιχνευτών για τη σύλληψη της ακτινοβολίας Cherenkov από τα προϊόντα αλληλεπίδρασης των νετρίνων. Οι ανιχνευτές αποτελούνται από κάθετες συστοιχίες που ονομάζονται Ανιχνευτικές Μονάδες (*Detection Units – DUs*). Ο τελικός ανιχνευτής ARCA θα περιέχει 230 DUs, σε δύο τμήματα των 115, ενώ ο τελικός ανιχνευτής ORCA θα περιέχει 115 DUs σε ένα τμήμα. Κάθε DU αποτελείται από 18 Ψηφιακά Οπτικά Στοιχεία (*Digital Optical Modules – DOMs*), γυάλινες σφαίρες διαμέτρου 17 ιντσών. Κάθε DOM περιέχει 31 φωτοπολλαπλασιαστές (PMTs) τοπο-



Σχήμα 3: Καλλιτεχνική απεικόνιση ενός ανιχνευτή του πειράματος KM3NeT στον θαλάσσιο πυθμένα. Διακρίνονται τα DUs, καθένα από τα οποία περιλαμβάνει 18 DOMs. Παραχώρηση του πειράματος KM3NeT.

θετημένους προς διαφορετικές κατευθύνσεις, επιτρέποντας την ανίχνευση φωτονίων από σχεδόν οποιαδήποτε γωνία [28, 38]. Τα DU είναι στερεωμένα στο βυθό και συγκρατούνται σε όρθια θέση από σηματοδούρες. Μια καλλιτεχνική απεικόνιση ενός ανιχνευτή του πειράματος KM3NeT φαίνεται στο Σχ. 3. Η οριζόντια απόσταση μεταξύ των DUs είναι 90 m (20 m) για τον ανιχνευτή ARCA (ORCA), ενώ η κατακόρυφη απόσταση των DOMs σε κάθε DU είναι 36 m (9 m). Ο πολύ μεγαλύτερος ανιχνευτικός όγκος του ανιχνευτή ARCA είναι απαραίτητος για την παρατήρηση ικανοποιητικού αριθμού αστροφυσικών νετρίνων, δεδομένου ότι τα τελευταία βρίσκονται στο υψηλής ενέργειας και χαμηλής ροής τμήμα του ενεργειακού φάσματος των νετρίνων. Η εκδοχή του ανιχνευτή που μας ενδιαφέρει στην παρούσα εργασία είναι ο ανιχνευτής ARCA με 21 DUs, ο οποίος εφεξής αναφέρεται ως *ARCA21*, δηλαδή η τελευταία εκδοχή του ανιχνευτή ARCA για την οποία η λήψη δεδομένων είχε ολοκληρωθεί κατά τη διάρκεια της εκπόνησης της εργασίας.

Συλλογή Δεδομένων και Σκανδαλισμός

Για κάθε «χτύπημα» (εφεξής: hit) φωτονίου σε κάποιο PMT, καταγράφονται και αποστέλλονται στην ξηρά ο χρόνος άφιξης του φωτονίου, το Time-over-Threshold (ToT) και ένα αναγνωριστικό για το εν λόγω PMT. Ο χρόνος άφιξης φωτονίου ορίζεται ως η στιγμή κατά την οποία το πλάτος του σήματος στο PMT υπερβαίνει το προκαθορισμένο όριο των 0,3 φωτοηλεκτρονίων. Αντίστοιχα, το ToT είναι το χρονικό διάστημα για το οποίο το πλάτος παραμένει πάνω από το προαναφερθέν κατώφλι. Πριν την αποθήκευση των δεδομένων, απαιτείται μείωση του όγκου τους σε πραγματικό χρόνο μέσω ενός συστήματος σκανδαλισμού. Ο σκανδαλισμός (triggering) αναφέρεται στη διαδικασία εντοπισμού και επιλογής υποψήφιων γεγονότων ενδιαφέροντος από τη ροή δεδομένων του ανιχνευτή με τη χρήση κατάλληλων κριτηρίων. Το σύστημα σκανδαλισμού του πειράματος KM3NeT χρησιμοποιεί τρία επίπεδα σκανδαλισμού:

- **L0 trigger:** αναφέρεται στο κατώφλι ώστε να καταγραφεί ένας παλμός PMT ως hit. Ως εκ τούτου, L0 hits θεωρούνται όλα τα καταγεγραμμένα hits.
- **L1 trigger:** αναφέρεται σε μια τοπική σύμπτωση δύο ή περισσότερων L0 hits από διαφορετικά PMT στο ίδιο DOM εντός ενός σταθερού χρονικού παραθύρου $\Delta T = 10$ ns.
- **L2 trigger:** αναφέρεται σε πιο σύνθετα κριτήρια, με βάση και τους προσανατολισμούς των PMT.

Διαφορετικοί αλγόριθμοι σκανδαλισμού μπορούν να εφαρμοστούν για τον εντοπισμό γεγονότων τροχιάς ή καταιγισμού. Για τον ανιχνευτή ARCA, ένα γεγονός σκανδαλιζείται εάν τουλάχιστον 5 L1 hits σε διαφορετικά DOMs αναγνωριστούν ως αιτιωδώς συνδεδεμένα από τον αλγόριθμο σκανδαλισμού τροχιάς ή καταιγισμού [28]. Όλα τα L1 hits που απαρτίζουν ένα γεγονός αναφέρονται ως *hits σκανδαλισμού*. Όταν όμως ένα γεγονός σκανδαλιζείται, αποθηκεύονται όλα τα L0 hits σε ένα χρονικό παράθυρο γύρω από τα hits σκανδαλισμού. Αυτά ονομάζονται *hits στηρημοτύπου* και είναι το πλήρες σύνολο των hits στον ανιχνευτή κατά το χρονικό διάστημα που συνέβη το γεγονός.

Πηγές Υποβάθρου

Εκτός από σήματα που προκαλούνται από νετρίνα, υπάρχουν διάφορες πηγές υποβάθρου που μπορούν να προκαλέσουν hits φωτονίων στα PMT του ανιχνευτή:

- **Υπόβαθρο θορύβου:** αναφέρεται σε διεργασίες που συνεισφέρουν οπτικό θόρυβο στα PMT, ο οποίος απορρίπτεται από τους αλγόριθμους σκανδαλισμού. Προέρχεται (α) από ραδιενεργές διασπάσεις στο θαλασσινό νερό, με κυριότερη το ^{40}K , το οποίο αποτελεί την κύρια πηγή οπτικού θορύβου στους ανιχνευτές, και (β) από τη βιοφωταύγεια, δηλαδή την εκπομπή ορατού φωτός από ζωντανούς οργανισμούς, ιδιαίτερα συχνή στα θαλάσσια βάθη.
- **Φυσικό υπόβαθρο:** αναφέρεται σε φυσικές διεργασίες που παράγουν σήματα που πληρούν τις συνθήκες σκανδαλισμού και έτσι προσδιορίζονται ως γεγονότα. Στο πείραμα KM3NeT φυσικό υπόβαθρο είναι (α) τα ατμοσφαιρικά μίονια, τα οποία αποτελούν την πλειονότητα των σκανδαλισμένων γεγονότων, και (β) μόνο για τον ανιχνευτή ARCA, τα ατμοσφαιρικά νετρίνα, καθώς είναι μη διακρίσιμα από τα αστροφυσικά.

Ανακατασκευή Γεγονότων

Κάθε hit ενός αποθηκευμένου γεγονότος περιλαμβάνει την ακόλουθη πληροφορία:

- θέση- x , $-y$, $-z$
- κατεύθυνση- x , $-y$, $-z$
- χρόνος άφιξης t
- Time-over-Threshold ToT

Η θέση αναφέρεται στη θέση του PMT, ενώ οι συνιστώσες κατεύθυνσης σχηματίζουν ένα μοναδιαίο διάνυσμα στην κατεύθυνση του PMT. Το πρόβλημα που καλούμαστε να εξετάσουμε είναι αυτό της ανακατασκευής γεγονότος. Ανακατασκευή γεγονότος είναι η διαδικασία εκτίμησης των παραμέτρων ενδιαφέροντος για το εν λόγω γεγονός (όπως η ενέργεια ή η κατεύθυνση του εισερχόμενου σωματιδίου), δεδομένου του συνόλου των hits που απαρτίζουν το γεγονός. Ξεχωριστοί αλγόριθμοι ανακατασκευής εφαρμόζονται για γεγονότα τροχιάς και καταγισμών, λαμβάνοντας υπόψη τις διαφορετικές τοπολογίες τους. Επικεντρωνόμαστε στην περίπτωση των τροχιών, καθώς αυτές μελετάμε στην παρούσα εργασία.

Ο στόχος του κλασικού αλγόριθμου ανακατασκευής τροχιάς του πειράματος KM3NeT είναι να προσαρμόσει την υπόθεση μιας τροχιάς μιονίου στο μοτίβο των hits του γεγονότος. Ο αλγόριθμος επιλέγει της παραμέτρους της τροχιάς με στόχο τη μεγιστοποίηση της πιθανότητας το μοτίβο των hits να αντιστοιχεί στην προσαρμοσμένη τροχιά [28]. Καθώς τόσο η θέση όσο και η διεύθυνση του μιονίου είναι παράμετροι που πρέπει να προσαρμοστούν, το πρόβλημα είναι μη γραμμικό και αντιμετωπίζεται σε διαδοχικά βήματα, τα οποία αναφέρονται ως αλυσίδα ανακατασκευής [42]. Στο πρώτο βήμα, εκτελούνται πολλαπλές προκαταρκτικές προσαρμογές, καθεμία με διαφορετική υποτιθέμενη κατεύθυνση της τροχιάς, με βήμα της 1° . Υποθέτοντας μια κατεύθυνση τροχιάς και αγνοώντας τις επιδράσεις της σκέδασης φωτονίων, το πρόβλημα μετατρέπεται σε γραμμικό, οπότε μπορεί να εφαρμοστεί προσαρμογή βασισμένη στους χρόνους άφιξης των φωτονίων Cherenkov, προκειμένου να εκτιμηθεί η θέση και ο χρόνος του μιονίου σε ένα συγκεκριμένο σημείο της τροχιάς του. Η διαδικασία επαναλαμβάνεται για όλες τις κατευθύνσεις και οι 12 καλύτερα προσαρμοσμένες κατευθύνσεις αποθηκεύονται για χρήση στο επόμενο στάδιο. Στο επόμενο βήμα της αλυσίδας, η τροχιά του μιονίου ανακατασκευάζεται με τη μέθοδο εκτίμησης μέγιστης πιθανοφάνειας, χρησιμοποιώντας τις καλύτερες κατευθύνσεις από το προκαταρκτικό στάδιο ως σημείο εκκίνησης. Η κατεύθυνση και η θέση του μιονίου τώρα προσαρμόζονται ταυτόχρονα. Η συνάρτηση πιθανοφάνειας κατασκευάζεται από συναρτήσεις πυκνότητας πιθανότητας (ΣΠΠ), οι οποίες περιγράφουν την πιθανότητα παρατήρησης ενός PMT hit σε σχετικό χρόνο Δt σε σχέση με τον αναμενόμενο χρόνο άφιξης ενός hit από ένα μη σχεδασμένο φωτόνιο Cherenkov:

$$\mathcal{L} = \prod_{hit\ PMTs} \frac{\partial P}{\partial t}(\rho_i, \theta_i, \phi_i, \Delta t), \quad (3)$$

όπου ρ_i είναι η ελάχιστη απόσταση του μιονίου από το PMT i , οι γωνίες θ_i και ϕ_i περιγράφουν τον προσανατολισμό του PMT σε σχέση με την τροχιά, και το Δt είναι η διαφορά μεταξύ του αναμενόμενου και του μετρημένου χρόνου άφιξης του φωτός. Οι ΣΠΠ ενσωματώνουν τις επιδράσεις της σκέδασης και της διασποράς του φωτός,

καθώς και πληροφορίες όπως το ποσοστό οπτικού υποβάθρου, την χβαντική απόδοση και τη γωνιακή αποδοχή των PMT. Μόλις βρεθεί η βέλτιστη προσαρμογή, η τροχιά του μιονίου έχει προσδιοριστεί και η προκύπτουσα τιμή πιθανοφάνειας αποθηκεύεται. Στη συνέχεια, γίνεται προσαρμογή της ενέργειας του μιονίου, χρησιμοποιώντας την πιθανότητα hit/non-hit για όλα τα PMT εντός ενός κυλινδρικού όγκου γύρω από την ανακατασκευασμένη τροχιά. Η πιθανότητα εκτιμάται και εδώ μέσω ΣΠΠ. Η πιθανοφάνεια του παρατηρούμενου μοτίβου hits μπορεί στη συνέχεια να προσδιοριστεί ως συνάρτηση της ενέργειας του μιονίου. Η μεγιστοποίηση αυτής της πιθανοφάνειας αποδίδει την ανακατασκευασμένη τιμή της ενέργειας του μιονίου. Τα πολύ καλά αποτελέσματα της κλασικής αλυσίδας ανακατασκευής του πειράματος KM3NeT λειτουργούν ως εξαιρετικό σημείο αναφοράς για τις προσπάθειες βελτίωσης με τη χρήση τεχνικών Μηχανικής Μάθησης.

Βαθιά Μάθηση & Νευρωνικά Δίκτυα Γράφων

Βαθιά Μάθηση

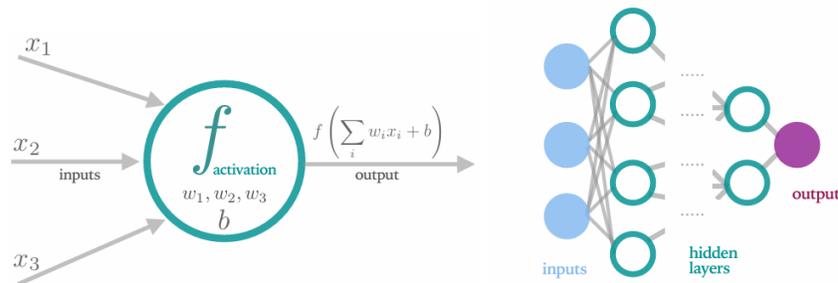
Η Βαθιά Μάθηση (BM) είναι το υποσύνολο των τεχνικών Μηχανικής Μάθησης (MM) που χρησιμοποιούν Βαθιά Νευρωνικά Δίκτυα (ΝΔ) για τον εντοπισμό μοτίβων και την εξαγωγή αναπαραστάσεων από δεδομένα. Η διάδοση της χρήσης των μεθόδων BM στη δεκαετία του 2010 επέτρεψε στα ΝΔ να συμβάλουν σε ανακαλύψεις σε διάφορους επιστημονικούς κλάδους μεταξύ των οποίων και η Φυσική, στην οποία εργαλεία BM παίζουν καθοριστικό ρόλο σε ένα ευρύ φάσμα εφαρμογών [49, 50]. Στο πλαίσιο αυτής της εργασίας, εστιάζουμε σε προβλήματα *επιβλεπόμενης μάθησης*, όπου στο σύνολο δεδομένων εκπαίδευσης οι εισόδοι x έχουν αντιστοιχιστεί με επιθυμητές *τιμές-στόχους* y . Έτσι, το δίκτυο καλείται να συμπεράνει μια συνάρτηση F έτσι ώστε $F(x) = y$. Το πρόβλημα που μας ενδιαφέρει είναι πρόβλημα *παλινδρόμησης*, όπου το μοντέλο έχει στόχο να προβλέψει μια συνεχή μεταβλητή-στόχο από τα δεδομένα εισόδου.

Αρχιτεκτονική Νευρωνικών Δικτύων

Τα Βαθιά Νευρωνικά Δίκτυα αποτελούνται από ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Κάθε επίπεδο αποτελείται υπολογιστικούς κόμβους, που ονομάζονται *νευρώνες* (βλ. Σχ. 4). Το επίπεδο εισόδου δέχεται τα δεδομένα εισόδου, ενώ στα κρυφά επίπεδα πραγματοποιείται η εξαγωγή αναπαραστάσεων. Κάθε νευρώνας συνδέεται με τους νευρώνες του προηγούμενου επιπέδου με ένα *βάρος* w . Για έναν νευρώνα i στο επίπεδο l , η έξοδος $\alpha_i^{(l)}$ δίνεται από τη σχέση:

$$\alpha_i^{(l)} = f \left(\sum_{j=1}^n w_{ij}^{(l)} \alpha_j^{(l-1)} + b_i^{(l)} \right). \quad (4)$$

Το $\alpha_j^{(l-1)}$ αντιπροσωπεύει την έξοδο του j -οστού νευρώνα του επιπέδου $l - 1$, $w_{ij}^{(l)}$ είναι τα σχετικά βάρη και $b_i^{(l)}$ μία παράμετρος που ονομάζεται *μεροληψία* (bias) του νευρώνα. Η συνάρτηση $f(\cdot)$ καλείται *συνάρτηση ενεργοποίησης* και εισάγει μη γραμμικότητα στο μοντέλο, επιτρέποντάς του να μαθαίνει πιο σύνθετα μοτίβα. Το τελικό επίπεδο του δικτύου είναι το επίπεδο εξόδου, η έξοδος του οποίου αποτελεί την απόκριση του δικτύου. Το ΝΔ που περιγράφηκε εδώ ονομάζεται multi-layer perceptron (MLP).



Σχήμα 4: Αριστερά: Νευρώνας με μεροληψία b και συνάρτηση ενεργοποίησης $f(\cdot)$ που δέχεται εισόδους x_i με βάρη w_i και παράγει την έξοδό του όπως στην Εξ. 4. Δεξιά: Σχηματική απεικόνιση Βαθύως ΝΔ. Από [51].

Εκπαίδευση Νευρωνικών Δικτύων

Προκειμένου το δίκτυο να παρέχει λογικές τιμές εξόδου στις εισόδους του, πρέπει πρώτα να εκπαιδευτεί. Κατά τη διαδικασία εκπαίδευσης, οι ρυθμιζόμενες παράμετροι του δικτύου, δηλαδή τα βάρη και οι μεροληψίες³ βελτιστοποιούνται. Αντίθετα, παράμετροι που έχουν καθοριστεί πριν από τη διαδικασία εκπαίδευσης ονομάζονται υπερπαράμετροι. Η εκπαίδευση ενός ΝΔ είναι μια επαναληπτική διαδικασία που αποσκοπεί στην ελαχιστοποίηση μιας συνάρτησης απώλειας L , η οποία ποσοτικοποιεί τη διαφορά μεταξύ των προβλέψεων και των τιμών-στόχων. Κάθε βήμα εκπαίδευσης αποτελείται από το εμπρόσθιο πέρασμα και την οπισθοδιάδοση.

Εμπρόσθιο πέρασμα & υπολογισμός απώλειας: Κατά τη διάρκεια του εμπρόσθιου πέρασματος, τα δεδομένα εισόδου περνούν από όλα τα επίπεδα του δικτύου (σύμφωνα με την Εξ. 4), από την είσοδο προς την έξοδο, για την παραγωγή προβλέψεων. Μόλις ληφθούν οι προβλέψεις, υπολογίζεται η συνάρτηση απώλειας L .

Οπισθοδιάδοση: Αφού υπολογιστεί η απώλεια, πρέπει να ρυθμιστούν τα βάρη ώστε αυτή να μειωθεί. Κατά τη διαδικασία της οπισθοδιάδοσης [52], υπολογίζεται η κλίση της συνάρτησης απώλειας σε σχέση με το κάθε βάρος, διαδίδοντας έτσι την απώλεια προς τα πίσω μέσω του δικτύου. Αυτή η κλίση δείχνει πώς πρέπει να τροποποιηθούν τα βάρη ώστε η απώλεια να μειωθεί.

Βελτιστοποίηση: Τα βάρη του δικτύου ενημερώνονται με την επαναληπτική μέθοδο της καθόδου κλίσης. Η ιδέα είναι να ενημερώνονται τα βάρη προς την κατεύθυνση μέγιστης μείωσης της συνάρτησης απώλειας. Μετά από αρκετά βήματα εκπαίδευσης, η απώλεια θα πρέπει να συγκλίνει σε ένα τοπικό ελάχιστο στον παραμετρικό χώρο των βαρών. Αν \mathbf{w} είναι το διάνυσμα όλων των βαρών του δικτύου, τότε στην κάθοδο κλίσης τα βάρη προσαρμόζονται σύμφωνα με τη σχέση:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \cdot \frac{\partial L}{\partial \mathbf{w}}. \quad (5)$$

Η παράμετρος η είναι ο ρυθμός μάθησης, μια υπερπαράμετρος που ελέγχει το μέγεθος κάθε ενημέρωσης των βαρών. Μικρό η σημαίνει ότι το δίκτυο ακολουθεί πιο πιστά το μονοπάτι της μέγιστης καθόδου, αλλά οδηγεί σε αργή μάθηση, ενώ μεγάλο η οδηγεί σε ταχύτερη μάθηση, όμως μπορεί να προκαλέσει αστάθεια της διαδικασίας εκπαίδευσης και υπερπήδηση του τοπικού ελαχίστου.

Στην πράξη, χρησιμοποιούνται αποτελεσματικότεροι αλγόριθμοι βελτιστοποίησης [54]. Ο πιο απλός είναι η Στοχαστική Κάθοδος Κλίσης (ΣΚΚ). Στη ΣΚΚ, κάθε βήμα εκπαίδευσης περιλαμβάνει μόνο έναν περιορισμένο αριθμό δειγμάτων, που ονομάζεται παρτίδα (batch). Το μέγεθος παρτίδας είναι μια υπερπαράμετρος. Η απώλεια στην Εξ. 5 είναι επομένως η απώλεια για μία μόνο παρτίδα, γεγονός που μειώνει σημαντικά τις απαιτήσεις μνήμης και υπολογισμού. Κάθε βήμα εκπαίδευσης πραγματοποιείται με νέα παρτίδα. Μία πλήρης διέλευση από ολόκληρο το σύνολο δεδομένων σε πολλαπλά βήματα εκπαίδευσης ονομάζεται εποχή. Η τυπική εκπαίδευση αποτελείται από αρκετές εποχές. Ένας ευρέως χρησιμοποιούμενος, πιο προηγμένος αλγόριθμος βελτιστοποίησης είναι ο αλγόριθμος Adam [55]. Ο Adam βελτιώνει τη ΣΚΚ ρυθμίζοντας δυναμικά το ρυθμό μάθησης για κάθε παράμετρο, επιτρέποντάς του να επιβραδύνει ή να επιταχύνει προσαρμοστικά τη διαδικασία μάθησης για διαφορετικές παραμέτρους και άρα οδηγώντας σε αποτελεσματικότερη σύγκλιση.

Γενίκευση: Για να είναι επιτυχημένο ένα μοντέλο Μηχανικής Μάθησης, πρέπει να έχει την ικανότητα να γενικεύει, δηλαδή να αποδίδει καλά και σε νέα δεδομένα. Δύο σημαντικά φαινόμενα που πρέπει να αποφεύγονται είναι η υποπροσαρμογή (underfitting) και η υπερπροσαρμογή (overfitting). Το πρώτο συμβαίνει όταν το μοντέλο δεν μπορεί να συλλάβει τα υποκείμενα μοτίβα στα δεδομένα, συχνά επειδή δεν είναι αρκετά σύνθετο ή ο χρόνος εκπαίδευσής του είναι πολύ μικρός. Το δεύτερο συμβαίνει όταν το μοντέλο αρχίζει να απομνημονεύει λεπτομέρειες των δεδομένων εκπαίδευσης, ακόμη και του θορύβου ή των ακραίων περιπτώσεων. Αυτό έχει ως αποτέλεσμα εξαιρετική απόδοση στο σύνολο εκπαίδευσης, αλλά φτωχή γενίκευση σε νέα δεδομένα. Προκειμένου να διασφαλιστεί η ικανότητα του μοντέλου να γενικεύει, το διαθέσιμο σύνολο δεδομένων συχνά χωρίζεται σε σύνολο εκπαίδευσης, επικύρωσης και δοκιμής. Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση του μοντέλου. Η απόδοση του μοντέλου ελέγχεται στο σύνολο επικύρωσης σε τακτά χρονικά διαστήματα, συνήθως στο τέλος κάθε εποχής. Όταν το μοντέλο εμφανίσει σημαντικά καλύτερη απόδοση στο σύνολο εκπαίδευσης από ό,τι στο σύνολο επικύρωσης, τότε αυτό μπορεί να αποτελεί ένδειξη υπερπροσαρμογής. Παρόλο που το δίκτυο δεν εκπαιδεύεται ποτέ στο σύνολο επικύρωσης, η επανειλημμένη χρήση του ως μέτρο απόδοσης μπορεί να δημιουργήσει μεροληψία προς μοντέλα που τυχάνει να παρουσιάζουν βελτιωμένη απόδοση σε αυτό. Για το λόγο

³ Από εδώ και στο εξής, τα βάρη και οι μεροληψίες του δικτύου θα αναφέρονται από κοινού απλά ως βάρη.

αυτό, το σύνολο δοκιμών προορίζεται για την αξιολόγηση του τελικού μοντέλου, παρέχοντας μια αμερόληπτη αξιολόγησή του.

Νευρωνικά Δίκτυα Γράφων

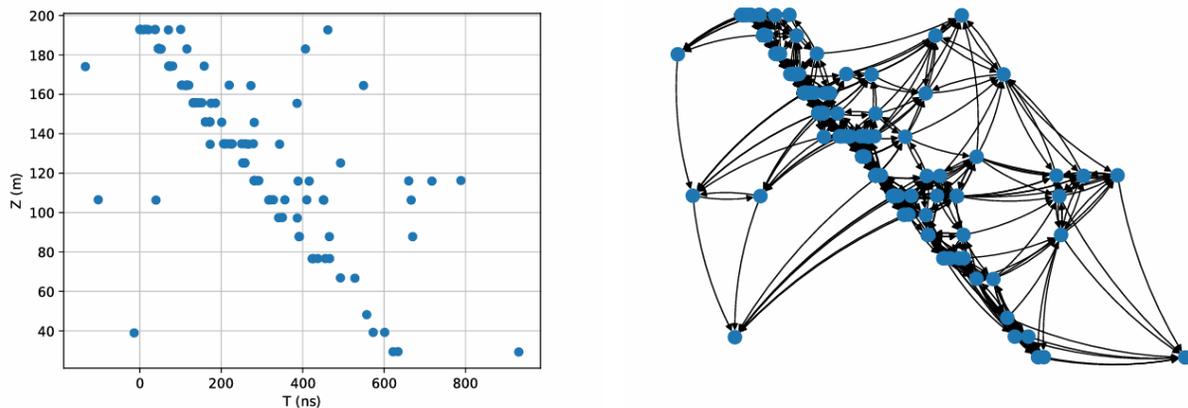
Τα πειράματα (αστρο)σωματιδιακής φυσικής, όπως το KM3NeT, παράγουν δεδομένα με δομή που μοιάζει με γράφο. Στους ανιχνευτές αυτούς, κάθε γεγονός αποτελείται από μεμονωμένες μετρήσεις που μπορούν να αναπαρασταθούν ως σημεία στο χώρο και στο χρόνο. Αυτά τα σημεία είναι αραιά διασκορπισμένα και δεν έχουν κανονική δομή πλέγματος, ενώ η κατανομή και ο αριθμός τους ποικίλλουν ανά γεγονός. Για παράδειγμα, στο KM3NeT ο αριθμός και η θέση των hits μεταβάλλονται σε κάθε γεγονός, ενώ οι ακριβείς θέσεις των DOM αλλάζουν διαρκώς λόγω των θαλάσσιων ρευμάτων. Επιπλέον, τα σημεία είναι μη ταξινομημένα, καθώς αντιπροσωπεύουν αλληλεπιδράσεις σωματιδίων χωρίς συγκεκριμένη διάταξη. Αυτά τα χαρακτηριστικά καθιστούν τους γράφους έναν φυσικό τρόπο αναπαράστασης των γεγονότων ενός ανιχνευτή σωματιδίων: *κάθε γεγονός παράγει έναν γράφο, με κόμβους που αναπαριστούν αλληλεπιδράσεις σωματιδίων (hits) και ακμές που περιγράφουν τις σχέσεις μεταξύ τους*. Συνεπώς, δεν αποτελεί έκπληξη το γεγονός ότι τα *Νευρωνικά Δίκτυα Γράφων (ΝΔΓ)* έχουν πρόσφατα συγκεντρώσει σημαντικό ενδιαφέρον στη Σωματιδιακή Φυσική, όπως φαίνεται από τις ανασκοπήσεις της χρήσης τους στα [56–58].

Δομή γράφου

Κάθε γεγονός στο πείραμα KM3NeT μπορεί να αναπαρασταθεί ως ένας κατευθυνόμενος γράφος $G = (V, E)$. Κάθε κόμβος του γράφου v_i αντιπροσωπεύει ένα PMT hit, με τις πληροφορίες του hit (θέση- x , $-y$, $-z$, κατεύθυνση- x , $-y$, $-z$, χρόνος t και ToT) να αναφέρονται ως *γνωρίσματα του κόμβου/hit*. Επομένως, εάν χρησιμοποιηθούν όλα, κάθε κόμβος θα έχει 8 γνωρίσματα. Ένα παράδειγμα του γράφου ενός γεγονότος μπορεί να δει κανείς στο Σχ. 5. Εξαιτίας του μεγάλου και μεταβλητού μεγέθους των γεγονότων, η χρήση πλήρως συνδεδεμένων γράφων δεν είναι πρακτική, ιδίως επειδή ο αριθμός των ακμών σε αυτούς αυξάνεται με το τετράγωνο του αριθμού των κόμβων. Για τον λόγο αυτό, κάθε κόμβος συνδέεται με τους k πλησιέστερους γείτονές του (*k-nearest neighbors – k-NN*), όπου k είναι μια υπερπαράμετρος. Η μετρική η οποία θα χρησιμοποιηθεί στην παρούσα εργασία για τον προσδιορισμό των πλησιέστερων γειτόνων, είναι η 4D Ευκλείδεια μετρική στον χώρο και τον χρόνο:

$$d^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 + c^2 \Delta t^2, \quad (6)$$

όπου c είναι η ταχύτητα του φωτός. Ο υπολογισμός των πλησιέστερων γειτόνων είναι υπολογιστικά ακριβός, διότι απαιτεί τον υπολογισμό της απόστασης μεταξύ κάθε ζεύγους κόμβων, καθιστώντας αυτό το βήμα $\mathcal{O}(n^2)$, όπου n ο αριθμός των κόμβων/hits στο γεγονός.



Σχήμα 5: Αριστερά: Κατανομή των hits ενός γεγονότος στο πείραμα KM3NeT (θέση z συναρτήσει του χρόνου t). Δεξιά: Αναπαράσταση του ίδιου γεγονότος με χρήση γράφου. Κάθε hit είναι ένας κόμβος, με ακμές προς τους 8 πλησιέστερους γείτονές του, οι οποίοι υπολογίζονται από την Εξ. 6. Από [59].

Συνέλιξη ακμών

Το θεμελιώδες συστατικό των ΝΔΓ που χρησιμοποιούνται στην παρούσα εργασία είναι το μπλοκ *Συνέλιξης Ακμών* ή *EdgeConv*, το οποίο εισήχθη στο [60]. Η ιδέα πίσω από το *EdgeConv* είναι η επέκταση της έννοιας της συνέλιξης από τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) στα ΝΔΓ. Στα ΣΝΔ, ένας πυρήνας (kernel) ολισθαίνει πάνω στην εικόνα, και το εσωτερικό γινόμενο των σημείων του με τα εικονοστοιχεία της εικόνας καταγράφεται σε κάθε σημείο. Στο *EdgeConv*, η έννοια του πυρήνα αντικαθίσταται από τους πλησιέστερους γείτονες, με τον αριθμό των γειτόνων k κάθε κόμβου να ισοδυναμεί με το μέγεθος του πυρήνα. Η συνέλιξη είναι τότε μια πράξη που ενεργεί στις εξερχόμενες ακμές κάθε κόμβου.

Μετά τον προσδιορισμό των πλησιέστερων γειτόνων, δηλαδή των ακμών του γραφήματος, ορίζονται τα γνωρίσματα των ακμών. Η ακμή e_{ij} από τον κόμβο v_i στον κόμβο v_j ορίζεται να έχει γνωρίσματα ακμής $(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i) \in \mathbb{R}^{2F}$, με $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^F$ να είναι τα γνωρίσματα των κόμβων v_i, v_j αντίστοιχα, ενώ F είναι ο αριθμός των γνωρισμάτων σε κάθε κόμβο. Η συνέλιξη ακμών συνίσταται στην ακόλουθη πράξη:

$$\mathbf{x}'_i = \sum_{j \in \mathcal{N}(i)} h_{\theta}(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i). \quad (7)$$

Σε αυτή την έκφραση, h_{θ} είναι ένα MLP 3 επιπέδων με παραμέτρους θ και $\mathcal{N}(i)$ είναι το σύνολο των k πλησιέστερων γειτόνων του κόμβου v_i . Το νέο διάνυσμα γνωρισμάτων κόμβου $\mathbf{x}'_i \in \mathbb{R}^{F'}$ έχει διαφορετικό αριθμό γνωρισμάτων από ό,τι πριν ($F' \neq F$). Εν ολίγοις, το *EdgeConv* δέχεται ως είσοδο έναν γράφο με n κόμβους, καθέναν με F γνωρίσματα, και εξάγει έναν γράφο με τον ίδιο αριθμό κόμβων, όπου ο καθένας έχει F' γνωρίσματα.

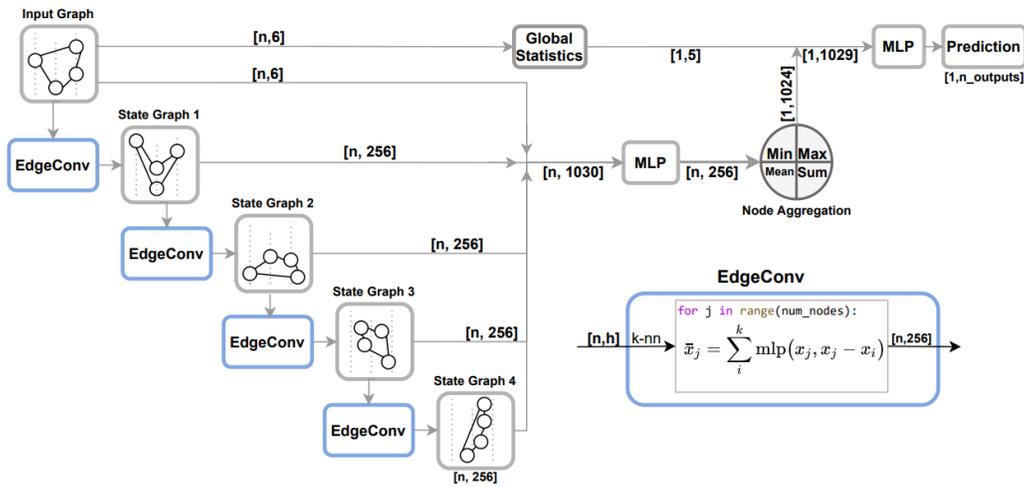
Αρχιτεκτονική δικτύου

Η αρχιτεκτονική ΝΔΓ που χρησιμοποιείται στην παρούσα εργασία ονομάζεται *DynEdge*, προτάθηκε από το IceCube στο [61] και εμφανίζεται στο Σχ. 6. Το *DynEdge* αποτελείται από 4 διαδοχικά μπλοκ *EdgeConv*. Στην έξοδο του κάθε μπλοκ παράγεται ένας γράφος με 256 γνωρίσματα ανά κόμβο. Οι ακμές του γράφου ενημερώνονται ώστε να προσδιοριστούν οι νέοι k πλησιέστεροι γείτονες, δημιουργώντας έτσι έναν νέο γράφο. Η χρήση διαδοχικών μπλοκ *EdgeConv* επιτρέπει στο δίκτυο να μάθει υψηλότερου επιπέδου χαρακτηριστικά των δεδομένων. Οι γράφοι εξόδου όλων των μπλοκ, όπως και ο αρχικός γράφος, συνδυάζονται σε έναν πίνακα $[n, 4 \cdot 256 + F]$, όπου n είναι ο αριθμός των κόμβων και F ο αριθμός των γνωρισμάτων κόμβου του αρχικού γράφου. Ο πίνακας αυτός τροφοδοτείται σε ένα MLP 3 επιπέδων, το οποίο μειώνει τη διάστασή του σε $[n, 256]$. Έπειτα, εκτελούνται πράξεις ομαδοποίησης πάνω στους κόμβους με έως 4 διαφορετικούς τρόπους: *mean*, *min*, *max* & *sum*, με την έξοδο κάθε πράξης να είναι ένας πίνακας μεγέθους $[1, 256]$. Οι πίνακες αυτοί συνενώνονται και έτσι, εάν χρησιμοποιηθούν και οι 4 επιλογές, ο πίνακας που θα προκύψει θα έχει διάσταση $[1, 4 \cdot 256] = [1, 1.024]$. Αυτός ο πίνακας συνδέεται στη συνέχεια με τα *global statistics*, ένα σύνολο 5 τιμών που υπολογίζονται από τον αρχικό γράφο και τον χαρακτηρίζουν συνολικά. Αυτές είναι ο αριθμός των κόμβων/hits και οι λόγοι ομοφιλίας για τις θέσεις $-x$, $-y$, $-z$ και το χρόνο t . Ο λόγος ομοφιλίας για ένα δεδομένο γνώρισμα κόμβου είναι το ποσοστό των ακμών του γράφου που συνδέουν κόμβους που έχουν την ίδια τιμή για το χαρακτηριστικό αυτό. Το τελικό διάνυσμα $[1, 1.029]$ περνάει σε ένα τελικό MLP 3 επιπέδων για την παραγωγή της εξόδου του δικτύου.

Εργαλεία λογισμικού

Η παρούσα εργασία κατέστη δυνατή χάρις στη χρήση εργαλείων λογισμικού ειδικά σχεδιασμένων για προβλήματα BM στο πείραμα KM3NeT και άλλα πειράματα νετρίνων:

- **GraphNeT:** Το GraphNeT [64] είναι ένα ανοιχτού κώδικα πλαίσιο BM με βάση το PyTorch [62] και το PyTorchLightning [66], που αναπτύχθηκε από το πείραμα IceCube με στόχο να παρέχει στους φυσικούς νετρίνων φιλικά προς τον χρήστη εργαλεία BM για την εκτέλεση εργασιών ανακατασκευής γεγονότων σε οποιοδήποτε πείραμα νετρίνων. Όλες οι εκπαιδεύσεις ΝΔΓ για την παρούσα εργασία πραγματοποιήθηκαν με τη χρήση του GraphNeT.
- **Εργαλεία BM του πειράματος KM3NeT:** Το πλαίσιο BM του πειράματος KM3NeT αποτελείται από δύο βιβλιοθήκες Python, που ονομάζονται *OreaSong* [67] και *OreaNet* [68].



Σχήμα 6: Αρχιτεκτονική του δικτύου DynEdge. Ο γράφος εισόδου θεωρείται εδώ ότι έχει n κόμβους και 6 γνωρίσματα ανά κόμβο. Από [61].

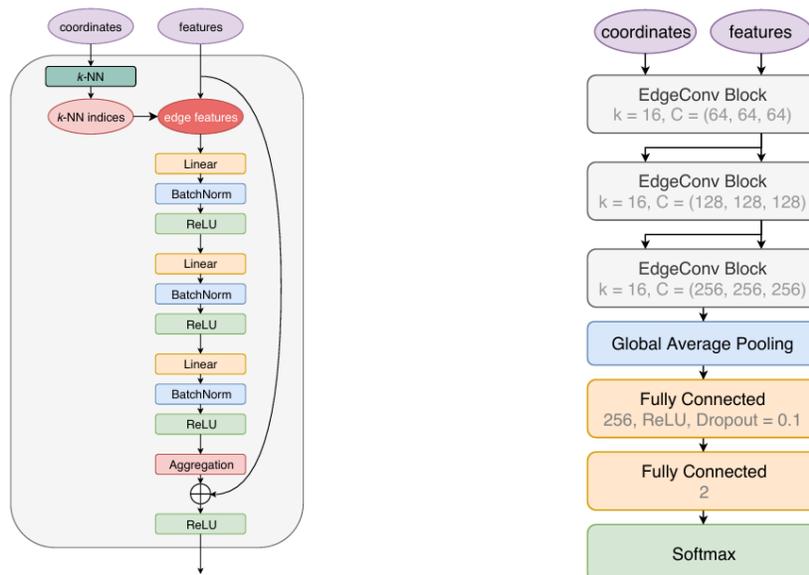
- Το OrcaSong χρησιμοποιείται για τη μετατροπή KM3NeT αρχείων σε μορφή αποτελεσματική για εφαρμογές BM. Συγκεκριμένα, μετατρέπει τα δεδομένα από τη μορφή αρχείου ROOT στη μορφή αρχείου HDF5, αφαιρώντας παράλληλα πληροφορίες μη σχετικές με την εκπαίδευση των ΝΔ. Τα προκύπτοντα αρχεία HDF5 έχουν πολύ μικρότερο μέγεθος και η εσωτερική δομή τους επιτρέπει πιο αποτελεσματικές λειτουργίες ανάγνωσης κατά την εκπαίδευση. Για το λόγο αυτό, αυτή είναι η μορφή αρχείων που χρησιμοποιήθηκε και στην παρούσα εργασία.
- Το OrcaNet είναι μια βιβλιοθήκη BM ειδική για το KM3NeT, βασισμένη στο TensorFlow [63], που χρησιμοποιείται για την εκπαίδευση ΝΔ. Αν και δεν θα χρησιμοποιηθεί άμεσα στην παρούσα εργασία, τα αποτελέσματα που λαμβάνονται με το OrcaNet θα συγκριθούν με εκείνα που προκύπτουν από εκπαιδεύσεις με το GraphNet.

Επισκόπηση Εφαρμογών BM & ΝΔΓ στην Ανακατασκευή Γεγονότων σε Τηλεσκοπία Νετρίνων

Οι κλασικοί αλγόριθμοι ανακατασκευής γεγονότων, τόσο στο πείραμα KM3NeT [42], όσο και στο πείραμα IceCube [69, 70], χρησιμοποιούν είτε προσαρμογή ελαχίστων τετραγώνων για γρήγορη, αλλά λιγότερο ακριβή γεωμετρική προσαρμογή, είτε προσεγγίσεις μέγιστης πιθανοφάνειας με πιο εξελιγμένη μοντελοποίηση της απόκλισης του ανιχνευτή για την τελική ανακατασκευή. Τα τελευταία χρόνια, η πρόοδος στη BM οδήγησε στην υιοθέτηση μεθόδων BM στα προβλήματα ανακατασκευής των περισσότερων πειραμάτων Σωματιδιακής Φυσικής. Αρχικά, χρησιμοποιήθηκαν μέθοδοι βασισμένες στην εικόνα, κυρίως ΣΝΔ, παρέχοντας συγκρίσιμα ή βελτιωμένα αποτελέσματα σε σχέση με τους κλασικούς αλγόριθμους τόσο στο πείραμα IceCube [72] όσο και στον ανιχνευτή KM3NeT/ORCA [73], προσφέροντας παράλληλα ταχύτερους χρόνους υπολογισμού.

Τα μειονεκτήματα της εφαρμογής των ΣΝΔ σε δεδομένα Σωματιδιακής Φυσικής, ιδίως η απώλεια πληροφορίας που συνδέεται με τη μετατροπή των δεδομένων σε εικόνες, οδήγησαν την κοινότητα να επιδιώξει εναλλακτικές προσεγγίσεις, όπως τα ΝΔΓ. Με βάση την εργασία των Wang *et al.* [60], η οποία εισήγαγε την έννοια της συνέλιξης ακμών σε νέφη σημείων, οι Qu και Gouskos [74] πρότειναν μια αρχιτεκτονική ΝΔΓ με την ονομασία *ParticleNet*, με εφαρμογή σε συγχρούσεις σωματιδίων στον επιταχυντή LHC. Όσον αφορά στα τηλεσκοπία νετρίνων, το IceCube εισήγαγε το μοντέλο DynEdge και το εφάρμοσε στα προβλήματα ταξινόμησης και ανακατασκευής γεγονότων προσομοιωμένων νετρίνων χαμηλής ενέργειας, επιτυγχάνοντας βελτιωμένα αποτελέσματα σε σχέση με τους κλασικούς αλγόριθμους [61]. Μια προκαταρκτική μελέτη για την ενεργειακή ανακατασκευή γεγονότων υψηλής ενέργειας έχει επίσης συζητηθεί στο [75].

Το KM3NeT έχει επίσης στραφεί στη χρήση ΝΔΓ για ανακατασκευή και ταξινόμηση γεγονότων. Για το σκοπό



Σχήμα 7: Αριστερά: Δομή του μπλοκ EdgeConv που χρησιμοποιείται στο ParticleNet. Δεξιά: Αρχιτεκτονική του δικτύου ParticleNet όπως παρουσιάζεται στο [74].

αυτό, η αρχιτεκτονική ParticleNet από το [74] έχει υλοποιηθεί στο OrcaNet, με ελάχιστες τροποποιήσεις (Σχ. 7). Όπως το DynEdge, το ParticleNet χρησιμοποιεί μπλοκ EdgeConv με MLP 3 επιπέδων για τη συνέλιξη ακμών πάνω στους k -NN. Το ParticleNet αποτελείται από 3 διαδοχικά μπλοκ EdgeConv, αυξάνοντας τον αριθμό των γνωρισμάτων ανά κόμβο από 64 σε 128 σε 256. Σε αντίθεση με το DynEdge, όπου οι έξοδοι όλων των μπλοκ EdgeConv συνενώνονται πριν την ομαδοποίηση, το ParticleNet χρησιμοποιεί συνδέσεις συντόμευσης για να αθροίσει τα γνωρίσματα κόμβων στην είσοδο κάθε μπλοκ με αυτά στην έξοδό του. Μετά τα μπλοκ EdgeConv, εκτελείται μια μόνο πράξη ομαδοποίησης (σε αντίθεση με τις έως 4 παράλληλες λειτουργίες στο DynEdge), η οποία μειώνει τη διάσταση της εξόδου από $[n, 256]$ σε $[1, 256]$. Αυτή η πράξη λέγεται «global average pooling» (αντίστοιχη με την ομαδοποίηση «mean» στο DynEdge). Το αποτέλεσμα διαδίδεται στη συνέχεια μέσω ενός μόνο κρυμμένου επιπέδου 256 νευρώνων, το οποίο τροφοδοτεί το επίπεδο εξόδου όπου παράγονται οι προβλέψεις. Τόσο στο ParticleNet όσο και στο DynEdge, ο αριθμός των εξόδων και η συνάρτηση ενεργοποίησης του επιπέδου εξόδου επιλέγονται ανάλογα με το εκάστοτε πρόβλημα. Χρησιμοποιώντας την παραπάνω αρχιτεκτονική αρκετές μελέτες ανακατασκευής έχουν πρόσφατα διεξαχθεί στο KM3NeT:

- Ανακατασκευή της κατεύθυνσης, της διαμέτρου και της πολλαπλότητας δεσμών μιονίων στον ανιχνευτή ORCA4 (ORCA με 4 DU). Η ανακατασκευή της κατεύθυνσης των μιονίων αποδείχθηκε ανταγωνιστική με τους κλασικούς αλγόριθμους, ενώ η διάμετρος της δέσμης και η πολλαπλότητα (αριθμός των μιονίων στη δέσμη) των μιονίων ανακατασκευάστηκαν για πρώτη φορά στο πείραμα KM3NeT [59, 76, 77].
- Εκτεταμένες μελέτες ανακατασκευής (ενέργεια, κατεύθυνση) και ταξινόμησης (σήμα/υπόβαθρο, τροχιά/καταιγισμός) για τον ανιχνευτή ORCA6. Τα αποτελέσματα αυτής της εργασίας ήταν πολλά υποσχόμενα, καθώς τόσο η ανακατασκευή της γωνίας ζενίθ όσο και της ενέργειας παρουσίασαν παρόμοια ή βελτιωμένη απόδοση σε σύγκριση με την κλασική ανακατασκευή [34].
- Παρόμοιες μελέτες έχουν πραγματοποιηθεί για τον ανιχνευτή ARCA6-8 [78]. Τα ΝΔΓ πέτυχαν μικρότερα σφάλματα από τον κλασικό αλγόριθμο ανακατασκευής στην ανακατασκευή της ενέργειας και της γωνίας ζενίθ των νετρίνων τόσο για τροχιές όσο και για καταιγισμούς. Παρουσιάστηκαν επίσης αποτελέσματα ταξινόμησης σήματος/υποβάθρου, καθώς και ανακατασκευής της κατεύθυνσης και πολλαπλότητας δεσμών μιονίων.

Επομένως, τα ΝΔΓ έχουν φανεί πολλά υποσχόμενα σε προβλήματα ανακατασκευής γεγονότων σε μικρότερες εκδόσεις των ανιχνευτών KM3NeT. Η παρούσα εργασία:

- Εφαρμόζει τα ΝΔΓ σε μεγαλύτερη εκδοχή των ανιχνευτών, καθώς αυτοί έχουν πλέον επεκταθεί, συγκεκριμένα στην ανακατασκευή ενέργειας για τον ανιχνευτή ARCA21.

- Δοκιμάζει περαιτέρω την απόδοση των ΝΔΓ σε νετρίνα υψηλής ενέργειας, δεδομένου ότι οι περισσότερες από τις παραπάνω εργασίες εξέτασαν μόνο νετρίνα χαμηλής ενέργειας στον ανιχνευτή ORCA.
- Εφαρμόζει για πρώτη φορά το μοντέλο ΝΔΓ DynEdge σε KM3NeT δεδομένα.
- Ενσωματώνει τα KM3NeT δεδομένα στο εργαλείο GraphNeT, καθιστώντας το διαθέσιμο προς χρήση στο πείραμα KM3NeT.
- Προσφέρει σημαντική βελτίωση στην ανακατασκευή γεγονότων σε χαμηλές ενέργειες για τον ανιχνευτή ARCA21.

Ανακατασκευή της Ενέργειας με Νευρωνικά Δίκτυα Γράφων

Μορφή Δεδομένων Βαθιάς Μάθησης στο Πείραμα KM3NeT

Για την εκτέλεση επιβλεπόμενων εργασιών BM, όπως η ανακατασκευή γεγονότων με ΝΔΓ, απαιτείται ένα σύνολο επιστημονικών δεδομένων, στο οποίο κάθε γεγονός θα συνοδεύεται από μια αντίστοιχη τιμή αλήθειας για την ενέργεια, την κατεύθυνση κ.λπ. που θα χρησιμεύσει ως τιμή-στόχος κατά την εκπαίδευση. Επομένως, αντί για τα πραγματικά δεδομένα του ανιχνευτή, των οποίων οι τιμές αλήθειας είναι άγνωστες, θα χρησιμοποιηθούν γεγονότα που παράγονται σε προσομοιώσεις *Monte Carlo*. Οι προσομοιώσεις Monte Carlo (MC) χρησιμοποιούνται στη Φυσική Υψηλών Ενέργειών για να προσομοιώνουν τις φυσικές διεργασίες ενδιαφέροντος και να μοντελοποιούν την απόκριση του ανιχνευτή. Στο πείραμα KM3NeT, οι προσομοιώσεις MC παράγουν αλληλεπιδράσεις νετρίνων σε μία περιοχή που περιλαμβάνει τον ανιχνευτή. Προσομοιώνεται η ακτινοβολία Cherenkov από τα δευτερογενή σωματίδια, καθώς και η ανίχνευσή της από τα PMT του ανιχνευτή. Αφού προσομοιωθεί και η συνεισφορά του θορύβου για το αντίστοιχο διάστημα συλλογής δεδομένων, ή *run*, του ανιχνευτή, εφαρμόζονται οι αλγόριθμοι σκανδαλισμού και τα σκανδαλισμένα γεγονότα αποθηκεύονται. Το αποτέλεσμα της παραγωγής MC είναι ένα σύνολο αρχείων, το καθένα από τα οποία περιέχει MC γεγονότα για ένα *run* του ανιχνευτή. Για κάθε γεγονός, περιλαμβάνονται τόσο το σύνολο των hits του, όσο και η πραγματική ενέργεια, κατεύθυνση, θέση της αλληλεπίδρασης κ.λπ., καθώς και οι αντίστοιχες ανακατασκευασμένες ποσότητες από τους τυπικούς αλγόριθμους ανακατασκευής. Τα δεδομένα του πειράματος KM3NeT (πειραματικά και MC) αποθηκεύονται αρχικά σε αρχεία *ROOT*, μια μορφή αρχείων που χρησιμοποιείται στο εργαλείο ανάλυσης δεδομένων *ROOT* που αναπτύχθηκε από το CERN. Ωστόσο, καθώς τα αρχεία *ROOT* του πειράματος KM3NeT δεν έχουν βελτιστοποιηθεί για εφαρμογές BM, προτού χρησιμοποιηθούν υποβάλλονται συνήθως σε προεπεξεργασία με τη χρήση του *OrcaSong*, το οποίο τα μετατρέπει στη μορφή αρχείου *HDF5*. Τα τελικά *HDF5* αρχεία έχουν ελάχιστο μέγεθος σε σύγκριση με τα αρχικά αρχεία *ROOT*, ενώ παράλληλα προσφέρουν γρήγορη προσπέλαση στις ιεραρχικές δομές δεδομένων που περιέχονται σε αυτά. Για τους λόγους αυτούς, ήταν η μορφή αρχείων που επιλέχθηκε για τις εκπαιδεύσεις στο GraphNeT.

Ενσωμάτωση των KM3NeT δεδομένων BM στο GraphNeT

Προκειμένου να χρησιμοποιηθεί το GraphNeT, είναι απαραίτητο να ενσωματωθεί η μορφή των KM3NeT δεδομένων BM σε αυτό, έτσι ώστε να μπορούν να χρησιμεύσουν ως είσοδος στο μοντέλο. Το GraphNeT προσφέρει υποστήριξη για δύο μορφές αρχείων (*Parquet* & *SQLite*), μέσα από κατάλληλες κλάσεις που χειρίζονται αιτήματα προσπελάσεων των δεδομένων για κάθε μία από τις δύο αυτές μορφές αρχείων. Για την ενσωμάτωση των αρχείων *HDF5* στο GraphNeT δημιουργήθηκε μια τρίτη τέτοια κλάση, η οποία χειρίζεται αιτήματα προσπελάσεων σε αρχεία *HDF5*.

Εκπαίδευση ΝΔΓ με το GraphNeT στον Ανιχνευτή ARCA21

Σύνολο δεδομένων και προεπεξεργασία

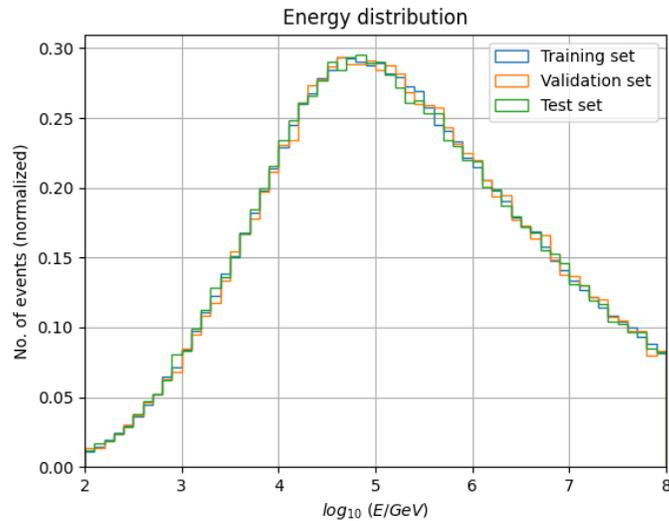
Το πρόβλημα ανακατασκευής που αντιμετωπίζεται σε αυτήν την εργασία αφορά γεγονότα τροχιάς, δηλαδή γεγονότα (αντι)νετρίνων μιονίου. Το σύνολο δεδομένων που θα χρησιμοποιηθεί αποτελείται από MC δεδομένα από 90 *runs* του ανιχνευτή. Κάθε *run* περιέχει περίπου 4.500 γεγονότα κάθε τύπου ($\nu_\mu / \bar{\nu}_\mu$), έτσι ώστε το σύνολο δεδομένων να ανέρχεται σε περίπου 800.000 γεγονότα, ισομερώς κατανομημένα μεταξύ ν_μ και $\bar{\nu}_\mu$. Από αυτά τα 90 *runs*, 75 θα χρησιμοποιηθούν για εκπαίδευση και 15 για επικύρωση (~ 83% / 17%). Ένα πρόσθετο σύνολο 19 *runs* έχει δεσμευτεί ως σύνολο δοκιμής για την αξιολόγηση των προβλέψεων. Η ενέργεια των νετρίνων

στο σύνολο δεδομένων κυμαίνεται μεταξύ $[10^2, 10^8]$ GeV. Το Σχ. 8 δείχνει την κατανομή των γεγονότων σε αυτό το ενεργειακό εύρος στα σύνολα δεδομένων εκπαίδευσης, επικύρωσης και δοκιμής. Παρατηρούμε λιγότερα γεγονότα στις υψηλότερες ενέργειες λόγω της χαμηλής ροής νετρίνων και του μικρού ανιχνευτικού όγκου του ανιχνευτή ARCA21, και επίσης λίγα γεγονότα σε χαμηλές ενέργειες, καθώς ο ανιχνευτής είναι λιγότερο ευαίσθητος στην περιοχή των λίγων TeV και κάτω.

Κάθε γεγονός στον ανιχνευτή ARCA21 περιέχει κατά μέσο όρο περίπου 1.000 hits. Υπάρχουν, ωστόσο, ακραίες περιπτώσεις με πολύ μεγαλύτερο αριθμό hits. Δεδομένου ότι ο προσδιορισμός των k -NN για έναν ολόκληρο γράφο είναι μια $\mathcal{O}(n^2)$ διεργασία, ο μέγιστος αριθμός των hits σε ένα γεγονός έχει οριστεί σε 5.000. Εάν ένα γεγονός περιείχε αρχικά περισσότερα από 5.000 hits, τότε όλα τα hits σκανδαλισμού διατηρούνται, ενώ κάποια από τα hits στιγμιότυπου παραλείπονται, ώστε να παραμείνουν συνολικά μόνο 5.000 hits. Κάθε PMT hit στα δεδομένα περιέχει τα ίδια 8 γνωρίσματα: θέση- x , $-y$, $-z$, κατεύθυνση- x , $-y$, $-z$, χρόνος t και ToT. Φυσικά, τα εύρη τιμών των γνωρισμάτων διαφέρουν σημαντικά μεταξύ τους. Η χρήση γνωρισμάτων με μεγάλες και διαφορετικές τιμές είναι προβληματική, καθώς μπορεί να δημιουργήσει μια μεροληψία προς τα γνωρίσματα με μεγαλύτερες αριθμητικές τιμές, οδηγώντας το δίκτυο να αγνοήσει τα υπόλοιπα. Για τον λόγο αυτό, κάθε γνώρισμα x μετασχηματίζεται αφαιρώντας τη μέση τιμή του μ_x και διαιρώντας με την τυπική του απόκλιση σ_x στο σύνολο των δεδομένων:

$$x \rightarrow \frac{x - \mu_x}{\sigma_x}. \quad (8)$$

Οι τρεις συντεταγμένες κατεύθυνσης δεν μετασχηματίζονται, καθώς ορίζονται έτσι ώστε να σχηματίζουν ένα μοναδιαίο διάνυσμα για κάθε hit, επομένως οι τιμές τους είναι ήδη κεντραρισμένες γύρω από το 0 και εμπίπτουν στο αποδεκτό εύρος $[-1, 1]$.



Σχήμα 8: Ενεργειακή κατανομή των νετρίνων στα σύνολα δεδομένων εκπαίδευσης, επικύρωσης και δοκιμής, κανονικοποιημένη έτσι ώστε το εμβαδόν κάτω από κάθε καμπύλη να είναι 1. Ο άξονας x είναι ο εκθέτης της ενέργειας, με κάθε bin να καλύπτει εύρος 0,1. Όλα τα σύνολα δεδομένων παρουσιάζουν την ίδια ενεργειακή κατανομή.

Υλοποίηση του μοντέλου

Οι εκπαιδεύσεις που πραγματοποιήθηκαν στην παρούσα εργασία χρησιμοποιούν το μοντέλο DynEdge. Στην ενότητα αυτή παρουσιάζονται διάφορες επιλογές υπερπαραμέτρων και λεπτομέρειες της υλοποίησης.

Αριθμός πλησιέστερων γειτόνων: Ένας μεγάλος αριθμός πλησιέστερων γειτόνων k μπορεί να αυξήσει την απόδοση του μοντέλου επιτρέποντάς του να καταγράφει σχέσεις μεταξύ πιο απομακρυσμένων κόμβων του γράφου. Ωστόσο, αυξάνει επίσης σημαντικά τον υπολογιστικό χρόνο, καθώς αυξάνει τον αριθμό των πράξεων

κατά τις συνελίξεις ακμών. Στην παρούσα εργασία επιλέχθηκε η τιμή $k = 16$ και για τα 4 μπλοκ EdgeConv του μοντέλου, τιμή διπλάσια από την αντίστοιχη του πειράματος IceCube [61] και ίση με την τιμή που χρησιμοποιήθηκε σε προηγούμενες μελέτες του ανιχνευτή ARCA [78]. Μεγαλύτερες τιμές του k δεν εξετάστηκαν καθώς (α) το υπολογιστικό κόστος ήταν ήδη υψηλό και (β) προηγούμενες εκπαιδεύσεις στο OrcaNet με $k = 32$ δεν απέδωσαν βελτιωμένα αποτελέσματα.

Ομαδοποίηση: Το DynEdge διαθέτει 4 επιλογές ομαδοποίησης («mean», «max», «min», «sum»), οι οποίες μπορούν να χρησιμοποιηθούν παράλληλα. Η χρήση πολλαπλών επιλογών πλεονεκτεί, καθώς καθιστά περισσότερες πληροφορίες διαθέσιμες στο μοντέλο. Έτσι, αρχικά χρησιμοποιήθηκαν και οι 4 επιλογές. Όμως, όταν ο αριθμός των κόμβων στον γράφο είναι μεγάλος, η πρόσθεση («sum») πάνω στους κόμβους για κάθε γνώρισμα μπορεί να οδηγήσει σε μεγάλες αριθμητικές τιμές, οι οποίες ενδεχομένως να κυριαρχήσουν έναντι των άλλων επιλογών ομαδοποίησης. Αυτή ακριβώς είναι η περίπτωση στο ARCA21, όπου κάθε γεγονός περιέχει περίπου 1.000 hits στιγμιότυπου, οδηγώντας σε γράφους με περίπου 1.000 κόμβους κατά μέσο όρο. Έτσι, για μεταγενέστερες εκπαιδεύσεις, η ομαδοποίηση «sum» αφαιρέθηκε.

Διαμόρφωση επιπέδου εξόδου: Το επίπεδο εξόδου του μοντέλου πρέπει να διαμορφωθεί ανάλογα με το εκάστοτε πρόβλημα, εν προκειμένω την ανακατασκευή της ενέργειας των νετρίνων, δηλαδή ένα πρόβλημα παλινδρόμησης, στο οποίο το μοντέλο καλείται να προβλέψει μία μοναδική τιμή, την ενέργεια του αρχικού σωματιδίου. Στην απλούστερη περίπτωση, το επίπεδο εξόδου αποτελείται από έναν μόνο νευρώνα, η έξοδος του οποίου είναι η προβλεπόμενη τιμή ενέργειας. Η συνάρτηση ενεργοποίησης αυτού του νευρώνα πρέπει να αποτρέπει μη φυσικές τιμές ενέργειας, άρα να περιορίζει το πεδίο εξόδου σε θετικές τιμές. Για το λόγο αυτό, χρησιμοποιείται η συνάρτηση *Softplus*:

$$f_{Softplus}(x) = \frac{1}{\beta} \ln(1 + e^{\beta x}), \quad (9)$$

με $\beta = 0,05$. Η συνάρτηση αυτή είναι μια ομαλή προσέγγιση της ευρέως διαδεδομένης συνάρτησης ReLU:

$$f_{ReLU}(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0. \end{cases} \quad (10)$$

Η Softplus αίρει τη μη διαφορισιμότητα της τελευταίας στο $x = 0$ και περιορίζει την έξοδο σε αυστηρά θετικές τιμές.

Συναρτήσεις απώλειας: Οι δύο πρώτες συναρτήσεις απώλειας που χρησιμοποιήθηκαν στην παρούσα εργασία είναι το ελάχιστο τετραγωνικό σφάλμα (*MSE*) ή L_2 και η απώλεια *Log Cosh* που χρησιμοποιείται επίσης στο [61]:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (11)$$

$$L_{Log\ Cosh} = \frac{1}{N} \sum_{i=1}^N \ln[\cosh(\hat{y}_i - y_i)], \quad (12)$$

όπου y_i είναι η πραγματική τιμή του i -οστού δείγματος, \hat{y}_i είναι η πρόβλεψη του δικτύου και N είναι το μέγεθος παρτίδας. Ο τετραγωνισμός στην απώλεια MSE ενισχύει τα σφάλματα, γεγονός που μπορεί να προσφέρει ταχύτερη σύγκλιση. Ωστόσο, συνεπάγεται ότι οι ακραίες τιμές στα δεδομένα έχουν μεγαλύτερη επίδραση στην απώλεια, επηρεάζοντας σε ορισμένες περιπτώσεις δυσανάλογα το μοντέλο. Η απώλεια Log Cosh είναι παρόμοια με την απώλεια L_2 για μικρά σφάλματα, αλλά μοιάζει με την απώλεια L_1 (μέσο απόλυτο σφάλμα – *MAE*) για μεγαλύτερα σφάλματα. Έτσι, εξασφαλίζει κλίση της συνάρτησης απώλειας ανάλογη του σφάλματος για μικρές τιμές σφάλματος, ώστε τα μικρά σφάλματα να ελαχιστοποιούνται ικανοποιητικά, είναι διαφορίσιμη στο 0, σε αντίθεση με την L_1 , και διαθέτει ευρωστία σε ακραίες τιμές, σε αντίθεση με την L_2 . Η τρίτη συνάρτηση απώλειας που θα εξεταστεί στην παρούσα εργασία είναι η *Log Normal* απώλεια:

$$L_{Log\ Normal} = \frac{1}{N} \sum_{i=1}^N -\ln \left[\frac{1}{\sqrt{2\pi}\hat{\sigma}_i^2} e^{-\frac{(y_i - \hat{\mu}_i)^2}{2\hat{\sigma}_i^2}} \right] = \frac{1}{N} \sum_{i=1}^N \left[\ln(\sqrt{2\pi}) + \ln(\hat{\sigma}_i) + \frac{(y_i - \hat{\mu}_i)^2}{2\hat{\sigma}_i^2} \right]. \quad (13)$$

Η συνάρτηση αυτή είναι ο φυσικός λογάριθμος μιας κανονικής κατανομής. Τόσο το $\hat{\mu}_i$ όσο και το $\hat{\sigma}_i$ είναι παράμετροι προσαρμογής, με το $\hat{y}_i \equiv \hat{\mu}_i$ να είναι η εκτίμηση του δικτύου για την ενέργεια και το $\hat{\sigma}_i$ να είναι μια εκτίμηση αβεβαιότητας για την πρόβλεψη. Η συνάρτηση απώλειας ελαχιστοποιείται στο σημείο $(\hat{\mu}_i, \hat{\sigma}_i) = (y_i, 0)$ και είναι παρόμοια με την απώλεια MSE, καθώς είναι τετραγωνική ως προς το σφάλμα. Η εκτίμηση αβεβαιότητας πραγματοποιείται με την προσθήκη ενός δεύτερου νευρώνα στο επίπεδο εξόδου του δικτύου και μπορεί να χρησιμοποιηθεί ως κριτήριο επιλογής γεγονότων στην ανάλυση δεδομένων. Η απώλεια Log Normal είναι η συνάρτηση απώλειας που χρησιμοποιείται στο OrcaNet και έτσι υλοποιήθηκε και στο GraphNeT για να χρησιμοποιηθεί στην παρούσα εργασία.

Σημειώνεται ότι, καθώς το ενεργειακό εύρος των νετρίνων στα MC δεδομένα είναι $[10^2, 10^8]$ GeV, καλύπτοντας 6 τάξεις μεγέθους, οι τιμές που υπεισέρχονται στη συνάρτηση απώλειας του δικτύου –όποια και αν είναι αυτή– αντιστοιχούν πάντα στον λογάριθμο της ενέργειας ($E \rightarrow \log(E)$). Έτσι, το εύρος τιμών-στόχων είναι $[2, 8]$, γεγονός που οδηγεί σε αυξημένη αριθμητική σταθερότητα και ομαλότερη σύγκλιση.

Λεπτομέρειες & παράμετροι εκπαίδευσης

Όλες οι εκδοχές του μοντέλου εκπαιδεύτηκαν σε κατάλληλο υπολογιστή στο Ινστιτούτο Πυρηνικής και Σωματιδιακής Φυσικής (ΙΠΣΦ) του ΕΚΕΦΕ «Δημόκριτος», εξοπλισμένο με δύο NVIDIA RTX A6000 48 GB GPU, 125 GB RAM και 48 πυρήνες CPU. Στον υπολογιστή αυτόν, κάθε εποχή στο σύνολο δεδομένων των 800.000 γεγονότων διαρκούσε περίπου 250 λεπτά.

Για τις αρχικές εκπαυσεις, χρησιμοποιήθηκε μέγεθος παρτίδας 32. Τόσο μικρά μεγέθη παρτίδας παρέχουν λιγότερο καλή προσέγγιση της συνολικής κλίσης του συνόλου δεδομένων, καθώς η κλίση είναι πιο θορυβώδης. Αυτό μπορεί να οδηγήσει σε ένα μοντέλο που συγκλίνει πιο αργά, αλλά γενικεύει καλύτερα, καθώς οι θορυβώδεις κλίσεις μπορούν να βοηθήσουν το μοντέλο να βγει από τα τοπικά ελάχιστα του παραμετρικού χώρου. Αργότερα, ωστόσο, ακολουθήθηκε η πρόταση των προγραμματιστών του GraphNeT να χρησιμοποιείται μεγάλο μέγεθος παρτίδας, δεδομένου ότι μπορεί να προσεγγίσει καλύτερα την κλίση του συνόλου δεδομένων, και το μέγεθος της παρτίδας αυξήθηκε σε 1.024, χωρίς να παρατηρηθεί απώλεια στην ικανότητα γενίκευσης.

Για όλες τις εκπαυσεις στην παρούσα εργασία χρησιμοποιήθηκε ο αλγόριθμος βελτιστοποίησης Adam, με τις προεπιλεγμένες παραμέτρους του $(\beta_1, \beta_2, \epsilon) = (0,9, 0,999, 10^{-8})$. Ο ρυθμός μάθησης ορίστηκε αρχικά σε 10^{-4} , ωστόσο, αυτή η τιμή οδηγούσε στο φαινόμενο των *εκρηκτικών κλίσεων*, όπου οι υπερβολικά μεγάλες κλίσεις προκαλούν ασταθείς ενημερώσεις στα βάρη του μοντέλου, εμποδίζοντας τη σύγκλιση του. Για το λόγο αυτό, χρησιμοποιήθηκε περιορισμός κλίσεων με τιμή 0,5. Σε μεταγενέστερες εκπαυσεις, χρησιμοποιήθηκε τμηματικά γραμμικός ρυθμός μάθησης με περίοδο προθέρμανσης μισής εποχής, ο οποίος ξεκινούσε από 10^{-5} με μέγιστο το 10^{-3} και ύστερα μειωνόταν μέχρι το 10^{-5} έως το τέλος της εκπαίδευσης, όπως περιγράφεται στο [61].

Διαμορφώσεις εκπαυσεων

Προτού συζητηθούν τα αποτελέσματα των εκπαυσεων των ΝΔΓ, πρέπει να παρουσιαστούν οι ακριβείς διαμορφώσεις εκπαίδευσης που χρησιμοποιήθηκαν. Σε όλες τις εκπαυσεις χρησιμοποιήθηκε ως τιμή-στόχος η πραγματική ενέργεια του MC νετρίνου. Οι εκπαυσεις μπορούν να χωριστούν σε δύο ομάδες, ανάλογα με το σύνολο των hits που χρησιμοποιήθηκαν κατά την εκπαίδευση (hits στιγμιότυπου ή σκανδαλισμού).

Hits στιγμιότυπου: Οι περισσότερες εκπαυσεις μας αξιοποίησαν τα hits στιγμιότυπου (δηλαδή σκανδαλισμένα και μη σκανδαλισμένα hits). Μια σύνοψη των εκπαυσεων με hits στιγμιότυπου παρουσιάζεται στον Πίνακα 1. Οι δύο πρώτες εκπαυσεις διεξήχθησαν σε μια μειωμένη εκδοχή του πλήρους συνόλου δεδομένων, που περιείχε μόνο 200.000 γεγονότα ανινετρίνων μιονίου. Σε αυτές τις εκπαυσεις, δοκιμάστηκαν οι απώλειες MSE και Log Cosh. Η εκπαίδευση no. 3 επανέλαβε την προηγούμενη με το πλήρες σύνολο δεδομένων. Μέχρι αυτό το σημείο το ToT δεν παρεχόταν ως μέρος των γνωρισμάτων των hits στο μοντέλο, δεδομένου ότι η ένταση του φωτός κάθε hit θεωρείται ότι μπορεί να συναχθεί από την πυκνότητα των hits σε γειτονικά PMT, η οποία είναι ήδη διαθέσιμη στο μοντέλο. Ωστόσο, δεδομένου ότι το γνώρισμα ToT ήταν διαθέσιμο, προστέθηκε στις επόμενες εκπαυσεις, με την εκπαίδευση no. 4 να είναι μια επανάληψη της προηγούμενης, αλλά με την προσθήκη του ToT, ώστε τα αποτελέσματα να μπορούν να συγκριθούν. Για τις υπόλοιπες εκπαυσεις, το μέγεθος της

παρτίδας αυξήθηκε από 32 σε 1.024, η επιλογή ομαδοποίησης «sum» αφαιρέθηκε και χρησιμοποιήθηκε ο τμηματικά γραμμικός ρυθμός μάθησης που αναφέρθηκε παραπάνω. Η εκπαίδευση no. 5 κάνει χρήση της απώλειας Log Cosh, επιτρέποντας την άμεση σύγκριση με τις προηγούμενες εκπαιδεύσεις. Τέλος, η εκπαίδευση no. 6 χρησιμοποιεί την απώλεια Log Normal, επιτρέποντας τη σύγκριση των διαφορετικών συναρτήσεων απώλειας.

No.	Γεγ/τα	Εποχ.	Απώλεια	Μέγ. Παρτ.	Ρυθμ. Μάθ.	Σημειώσεις
1	200.000	20	MSE	32	10^{-4}	Χωρίς ToT, 4πλη ομάδ., περιορισμός κλίσεων
2	200.000	20	Log Cosh	32	10^{-4}	Χωρίς ToT, 4πλη ομάδ., περιορισμός κλίσεων
3	800.000	10	Log Cosh	32	10^{-4}	Χωρίς ToT, 4πλη ομάδ., περιορισμός κλίσεων
4	800.000	10	Log Cosh	32	10^{-4}	Με ToT, 4πλη ομάδ., περιορισμός κλίσεων
5	800.000	30	Log Cosh	1.024	Τμημ. γραμμικός $10^{-5} \rightarrow 10^{-3} \rightarrow 10^{-5}$	Με ToT, χωρίς ομάδ. «sum»
6	800.000	30	Log Normal	1.024	Τμημ. γραμμικός $10^{-5} \rightarrow 10^{-3} \rightarrow 10^{-5}$	Με ToT, χωρίς ομάδ. «sum»

Πίνακας 1: Διαμορφώσεις εκπαίδευσης (hits στιγμιότυπου).

Hits σκανδαλισμού: Πραγματοποιήθηκαν επίσης εκπαιδεύσεις μόνο hits σκανδαλισμού. Δεδομένου ότι τα hits αυτά είναι όσα έχουν επισημανθεί από το σύστημα σκανδαλισμού ως συμπίπτοντα, τα περισσότερα hits οπτικού υποβάθρου αναμένεται να έχουν αφαιρεθεί από τα γεγονότα. Συνεπώς, η εκπαίδευση του μοντέλου μόνο με hits σκανδαλισμού θα μπορούσε να βελτιώσει την ανακατασκευή της ενέργειας. Οι εκπαιδεύσεις με τα hits σκανδαλισμού συνοψίζονται στον Πίνακα 2 και διαφέρουν ως προς τον ρυθμό μάθησης.

No.	Γεγ/τα	Εποχ.	Απώλεια	Μέγ. Παρτ.	Ρυθμ. Μάθ.	Σημειώσεις
1	800.000	15	Log Cosh	1.024	10^{-3}	Με ToT, χωρίς ομάδ. «sum», περιορισμός κλίσεων
2	800.000	30	Log Cosh	1.024	Τμημ. γραμμικός $10^{-5} \rightarrow 10^{-3} \rightarrow 10^{-5}$	Με ToT, χωρίς ομάδ. «sum»
3	800.000	20	Log Cosh	1.024	Τμημ. γραμμικός $10^{-5} \rightarrow 2 \cdot 10^{-3} \rightarrow 10^{-3}$	Με ToT, χωρίς ομάδ. «sum»

Πίνακας 2: Διαμορφώσεις εκπαίδευσης (hits σκανδαλισμού).

Αποτελέσματα και Συγκρίσεις Εκπαιδεύσεων

Ακολουθεί παρουσίαση και συζήτηση των αποτελεσμάτων των εκπαιδεύσεων ΝΔΓ που πραγματοποιήθηκαν. Εκτός από τα δύο πρώτα μοντέλα με hits στιγμιότυπου, τα οποία εκπαιδεύτηκαν και αξιολογήθηκαν σε μικρότερο σύνολο δεδομένων, όλα τα αποτελέσματα και τα διαγράμματα σε αυτή την ενότητα βασίζονται στο σύνολο δοκιμών, το οποίο περιέχει 172.484 γεγονότα.

Hits στιγμιότυπου

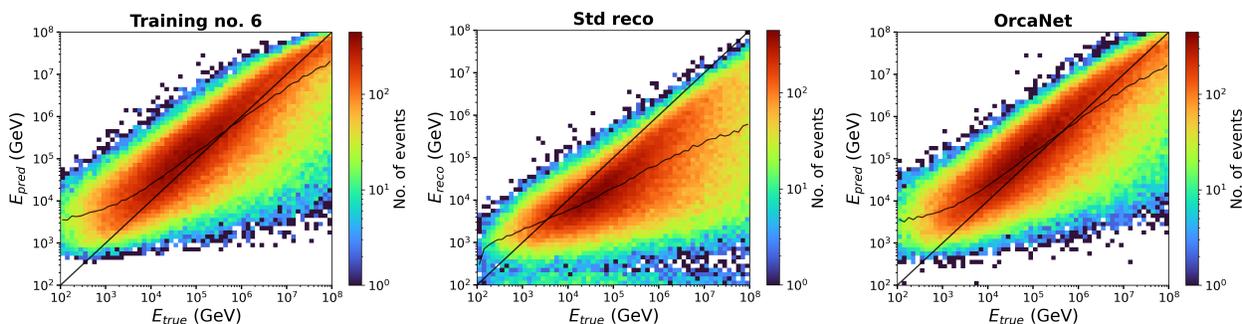
Τα αποτελέσματα των εκπαιδεύσεων hits στιγμιότυπου του Πίνακα 1 παρουσιάζονται στον Πίνακα 3. Για κάθε εκπαίδευση, το MAE, το MSE και ο συντελεστής προσδιορισμού R^2 ⁴ παρουσιάζονται, με τις εκπαιδεύσεις no. 5

⁴ Στατιστικό μέγεθος για το οποίο υψηλότερες τιμές δηλώνουν καλύτερα προσαρμοσμένο μοντέλο, με τη μέγιστη δυνατή τιμή να είναι 1.

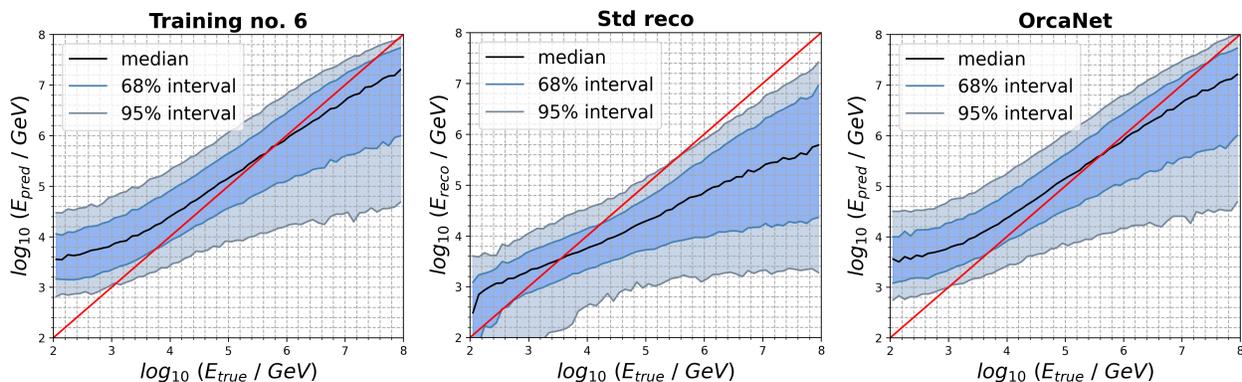
& 6 να μοιράζονται τις καλύτερες μετρικές. Τα αποτελέσματα της καλύτερης εκπαίδευσης (no. 6), καθώς και της καλύτερης εκπαίδευσης με το OrcaNet, αλλά και της τυπικής ανακατασκευής, απεικονίζονται σε διαγράμματα διασποράς και ποσοστημορίων της προβλεπόμενης συναρτήσε της πραγματικής ενέργειας στα Σχ. 9 και 10 αντίστοιχα.

No.	MAE	MSE	R^2
1	0,644	0,726	0,565
2	0,648	0,734	0,560
3	0,590	0,611	0,637
4	0,585	0,620	0,632
5	0,573	0,603	0,642
6	0,580	0,596	0,646
std reco	1,016	1,770	-0,133
OrcaNet	0,577	0,594	0,648

Πίνακας 3: Αποτελέσματα εκπαίδευσεων (hits στιγμασιτύπου). Οι έντονες τιμές υποδεικνύουν το καλύτερο αποτέλεσμα στο GraphNeT για κάθε μετρική. Τα αποτελέσματα για την τυπική ανακατασκευή («std reco») και το καλύτερο μοντέλο στο OrcaNet, στο ίδιο σύνολο δεδομένων, παρουσιάζονται επίσης για σύγκριση.



Σχήμα 9: Διαγράμματα διασποράς της προβλεπόμενης συναρτήσε της πραγματικής ενέργειας. Η μαύρη διαγώνια γραμμή δείχνει την ιδανική ανακατασκευή. Μια δεύτερη μαύρη γραμμή υποδεικνύει τη διάμεσο της προβλεπόμενης τιμές για κάθε bin πραγματικής ενέργειας.



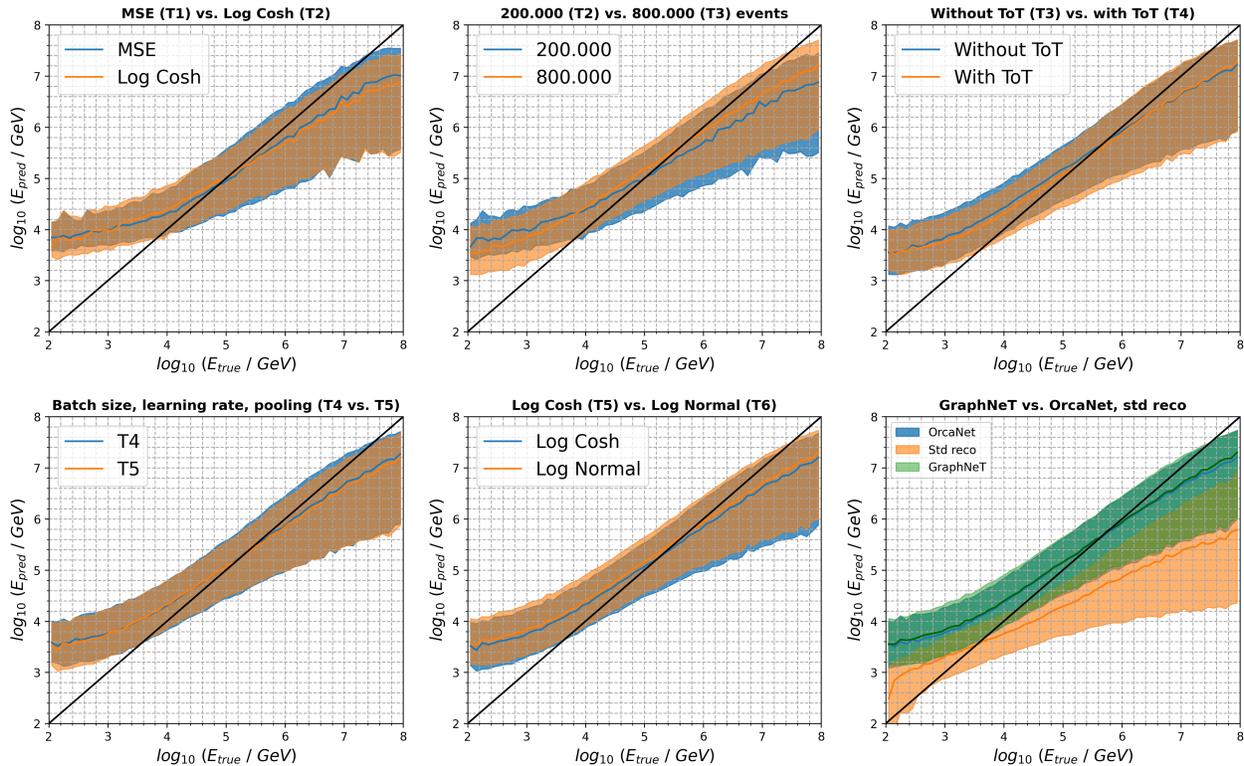
Σχήμα 10: Διαγράμματα ποσοστημορίων της προβλεπόμενης συναρτήσε της πραγματικής ενέργειας. Σε όλα τα διαγράμματα, η κόκκινη διαγώνια γραμμή υποδεικνύει την ιδανική ανακατασκευή. Η διάμεσος και τα διαστήματα εμπιστοσύνης 68% & 95% των προβλέψεων για κάθε bin πραγματικής ενέργειας απεικονίζονται με μαύρο, μπλε και ανοιχτό γκρι αντίστοιχα.

Όπως φαίνεται από τα Σχ. 9, 10, το δίκτυο παράγει τις πιο ακριβείς προβλέψεις στο μέσο του ενεργειακού εύρους, δηλαδή για E_{true} που κυμαίνεται από 10^3 έως 10^6 GeV (κλίμακα TeV). Αντίθετα, οι χαμηλότερες και

υψηλότερες ενέργειες τείνουν να υπερεκτιμώνται και να υποεκτιμώνται αντίστοιχα. Το φαινόμενο αυτό είναι πιο έντονο για τις χαμηλότερες ενέργειες, των οποίων η εκπροσώπηση στο σύνολο δεδομένων είναι ελάχιστη (βλ. Σχ. 8). Για τις υψηλότερες ενέργειες (κλίμακα PeV, $E_{true} > 10^6$ GeV), αναμένεται υποεκτίμηση, καθώς τα μόνια υψηλής ενέργειας δεν εναποθέτουν εν γένει ολόκληρο το ενεργειακό τους περιεχόμενο στον ανιχνευτή, δηλαδή δεν είναι όλη η ενέργειά τους ορατή. Το γεγονός αυτό αντικατοπτρίζεται και στο μεγάλο εύρος των προβλέψεων για τα γεγονότα αυτά. Μπορούμε επίσης να παρατηρήσουμε ότι το καλύτερο μοντέλο OrcaNet, που εκπαιδεύτηκε και αξιολογήθηκε στα ίδια σύνολα δεδομένων, παρουσιάζει την ίδια συμπεριφορά με τα δίκτυα που βασίζονται στο GraphNeT. Τέλος, η τυπική ανακατασκευή αναφέρεται στον αλγόριθμο εκτίμησης μέγιστης πιθανοφάνειας. Θα πρέπει να σημειωθεί ότι ο αλγόριθμος αυτός ανακατασκευάζει την *ενέργεια του μιονίου*. Επομένως, οι χαμηλότερες ενεργειακές προβλέψεις σε σύγκριση με τα ΝΔΓ είναι αναμενόμενες, καθώς η ορατή ενέργεια του μιονίου είναι χαμηλότερη από την ενέργεια του πρωτογενούς νετρίνου.

Συγκρίσεις εκπαίδευσεων: Στα συγκριτικά διαγράμματα του Σχ. 11 παρατίθενται τα διαστήματα 68% για διαφορετικά ζεύγη διαμορφώσεων εκπαίδευσης. Στη συνέχεια, οι εκπαιδεύσεις αναφέρονται με το γράμμα «T» ακολουθούμενο από τον αριθμό εκπαίδευσης από τον Πίνακα 1 (π.χ. «T1»). Οι περισσότερες συγκρίσεις γίνονται μεταξύ εκπαίδευσεων που διαφέρουν μόνο σε μία παράμετρο, ώστε να είναι αξιολογήσιμη η επίδρασή της:

- **MSE (T1) vs. Log Cosh (T2):** Η απώλεια MSE δίνει καλύτερες μετρικές, εξαιτίας της μεγαλύτερης ποινής στα μεγάλα σφάλματα, η οποία οδηγεί σε καλύτερη ελαχιστοποίηση των μεγάλων τιμών σφάλματος. Αν και οι επιδόσεις είναι παρόμοιες, η μπλε λωρίδα (MSE) είναι λίγο καλύτερα ευθυγραμμισμένη με τη διαγώνιο από την πορτοκαλί λωρίδα (Log Cosh) στο μεγαλύτερο μέρος του ενεργειακού εύρους.
- **200.000 (T2) vs. 800.000 (T3) events:** Η χρήση του μεγαλύτερου συνόλου δεδομένων βελτιώνει σημαντικά την απόδοση του δικτύου, με το R^2 να αυξάνεται κατά 13,8%. Τα μεγαλύτερα οφέλη στο διάγραμμα σύγκρισης εντοπίζονται στα γεγονότα χαμηλής και υψηλής ενέργειας (υποεκπροσωπούμενες περιοχές), τα οποία επωφελούνται περισσότερο από την αύξηση του μεγέθους του συνόλου δεδομένων.
- **Χωρίς ToT (T3) vs. με ToT (T4):** Αν και η προσθήκη της μεταβλητής ToT έχει ελάχιστο αντίκτυπο στις μετρικές, το διάγραμμα σύγκρισης αποκαλύπτει ότι η διάμεσος είναι ελαφρώς καλύτερα ευθυγραμμισμένη με τη διαγώνιο για την εκπαίδευση που περιλαμβάνει το ToT, ειδικά στην κλίμακα TeV. Για το λόγο αυτό, αποφασίστηκε να συνεχιστεί η χρήση του ToT.
- **Batch size, learning rate, pooling (T4 vs. T5):** Η σύγκριση περιλαμβάνει τις εξής αλλαγές: αφαίρεση της ομαδοποίησης «sum», αύξηση του μεγέθους παρτίδας σε 1.024 και χρήση τμηματικά γραμμικού ρυθμού μάθησης. Παρατηρούμε μικρή βελτίωση στις μετρικές, με το R^2 να αυξάνεται κατά 1,6%, καθώς και το χαμηλότερο MAE από όλα τα μοντέλα ανακατασκευής της ενέργειας στον ανιχνευτή ARCA21.
- **Log Cosh (T5) vs. Log Normal (T6):** Καθώς η απώλεια Log Normal είναι απώλεια τύπου L_2 , ενώ η απώλεια Log Cosh απώλεια τύπου L_1 , εύλογα η πρώτη εκτοπίζει τη δεύτερη στη μετρική MSE, ενώ το αντίθετο ισχύει για τη μετρική MAE. Οι δύο συναρτήσεις απωλειών επιδεικνύουν πολύ παρόμοιες μετρικές, ενώ στο αντίστοιχο διάγραμμα σύγκρισης οι δύο ζώνες πρακτικά ταυτίζονται. Η χρήση της απώλειας Log Normal στο OrcaNet αποδεικνύεται δικαιολογημένη, καθώς προσφέρει την ίδια απόδοση με απλούστερες συναρτήσεις απωλειών, ενώ ενσωματώνει μια εκτίμηση αβεβαιότητας στις προβλέψεις του δικτύου.
- **GraphNeT (T6) vs. OrcaNet, std reco:** Εδώ συγκρίνεται η εκπαίδευση T6 στο GraphNeT με την καλύτερη εκπαίδευση στο OrcaNet στο ίδιο σύνολο δεδομένων. Τα δύο δίκτυα διαθέτουν την ίδια συνάρτηση απωλειών, επιτρέποντας την άμεση σύγκριση των δύο αρχιτεκτονικών. Τα μοντέλα συγκρίνονται μαζί με την τυπική ανακατασκευή. Η λωρίδα της τυπικής ανακατασκευής βρίσκεται χαμηλότερα από τα μοντέλα ΝΔΓ, καθώς ανακατασκευάζει την ενέργεια του προκύπτοντος μιονίου και όχι την ενέργεια του νετρίνου. Η τυπική ανακατασκευή βρίσκεται σε καλύτερη συμφωνία με τα μοντέλα στις χαμηλότερες ενέργειες, στις οποίες μεγαλύτερο μέρος της ενέργειας των νετρίνων γίνεται ορατή. Οι λωρίδες των δύο μοντέλων επικαλύπτονται σχεδόν πλήρως, όπως υποδηλώνουν οι παρόμοιες μετρικές τους. Κατά συνέπεια, φαίνεται ότι και οι δύο αρχιτεκτονικές έχουν παρόμοιες δυνατότητες και ότι και τα δύο πλαίσια μπορούν να υποστηρίξουν με επιτυχία εκπαίδευσεις ΝΔΓ. Αυτό είναι σημαντικό αποτέλεσμα, καθώς (α) επιβεβαιώνει την ορθότητα στις επιλογές διαμόρφωσης και την ρύθμιση των υπερπαραμέτρων στην αρχιτεκτονική *ParticleNet* στο OrcaNet, και (β) παρέχει στο πείραμα *KM3NeT* ένα άμεσα διαθέσιμο εναλλακτικό εργαλείο με εξαιρετικές δυνατότητες (*GraphNeT*), το οποίο είναι εύκολο στη χρήση και μπορεί να φιλοξενήσει οποιαδήποτε επιθυμητή λειτουργικότητα.



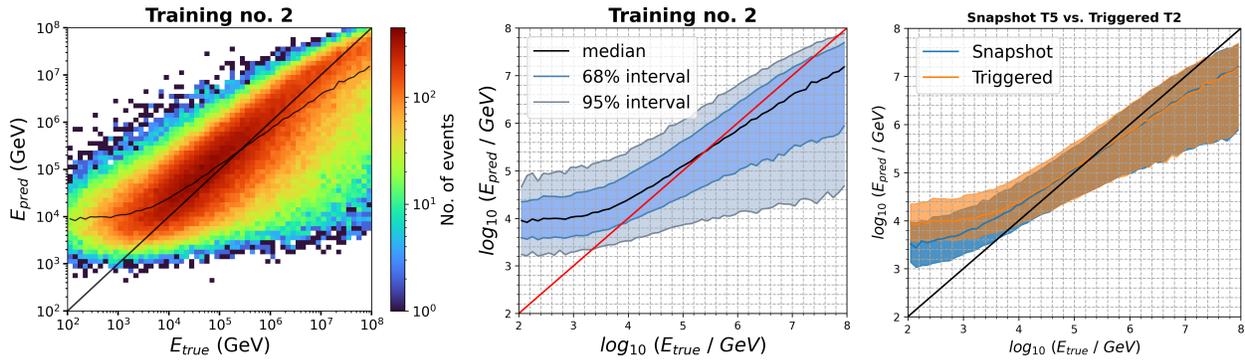
Σχήμα 11: Συγκρίσεις μεταξύ των διαφόρων διαμορφώσεων εκπαίδευσης του Πίνακα 1. Σε όλα τα διαγράμματα συγκρίνονται τα διαστήματα εμπιστοσύνης 68%.

Hits σκανδαλισμού

Τα αποτελέσματα των εκπαίδευσεων με hits σκανδαλισμού του Πίνακα 2 δίνονται στον Πίνακα 4. Παρουσιάζονται το MAE, το MSE και ο συντελεστής προσδιορισμού R^2 . Φαίνεται ότι η εκπαίδευση no. 2, με τον χαμηλότερο ρυθμό μάθησης έχει την καλύτερη απόδοση, σύμφωνα με δύο από τις τρεις μετρικές. Αυτό μπορεί να οφείλεται στις πιο σταθερές ενημερώσεις των βαρών που βοηθούν το δίκτυο να αποφύγει την υπερπήδηση τοπικών ελαχίστων. Τα διαγράμματα διασποράς και ποσοστημορίων της προβλεπόμενης συναρτήσε της πραγματικής ενέργειας για την καλύτερη εκπαίδευση (no. 2) παρουσιάζονται στο Σχ. 12. Στο ίδιο Σχήμα παρουσιάζεται και το διάγραμμα σύγκρισης μεταξύ της συγκεκριμένης εκπαίδευσης και της αντίστοιχης εκπαίδευσης με hits στιγμιότυπου (ίδια διαμόρφωση δικτύου). Συγκρίνοντας τις μετρικές στους Πίνακες 3 και 4, φαίνεται ότι οι εκπαίδευσεις με hits σκανδαλισμού έχουν σημαντικά χειρότερες επιδόσεις σε σύγκριση με αυτές με hits στιγμιότυπου. Στο διάγραμμα σύγκρισης του Σχ. 12 παρατηρούμε ότι και οι δύο εκπαίδευσεις έχουν παρόμοιες επιδόσεις στο μεσαίο και στο ανώτερο τμήμα του ενεργειακού εύρους. Παρόλα αυτά, η πρώτη έχει κατά 9% καλύτερη βαθμολογία R^2 από τη δεύτερη. Η διακύμανση της απόδοσης εντοπίζεται στις ενέργειες κάτω από 10^4 GeV, όπου η ζώνη των hits στιγμιότυπου βρίσκεται πιο κοντά στη διαγώνιο. Λόγω αυτού του ελλείμματος απόδοσης στις χαμηλές ενέργειες και της αποτυχίας να παρατηρηθούν βελτιώσεις σε άλλα τμήματα του ενεργειακού εύρους, δεν επιδιώχθηκε περαιτέρω εκπαίδευση με hits σκανδαλισμού.

No.	MAE	MSE	R^2
1	0,620	0,694	0,588
2	0,623	0,687	0,592
3	0,635	0,705	0,581

Πίνακας 4: Αποτελέσματα εκπαίδευσης (hits σκανδαλισμού). Οι έντονες τιμές υποδεικνύουν το καλύτερο αποτέλεσμα για κάθε μετρική.



Σχήμα 12: Αριστερά: Διάγραμμα διασποράς της προβλεπόμενης συναρτήσεως της πραγματικής ενέργειας για την T2. Μέσον: Διάγραμμα ποσοστημορίων της προβλεπόμενης συναρτήσεως της πραγματικής ενέργειας για την T2. Δεξιά: Διάγραμμα σύγκρισης μεταξύ της εκπαίδευσης T5 με hits στιγμιότυπου και της εκπαίδευσης T2 με hits σκανδαλισμού. Συγκρίνονται τα διαστήματα εμπιστοσύνης 68% για την κάθε εκπαίδευση.

Στάθμιση γεγονότων

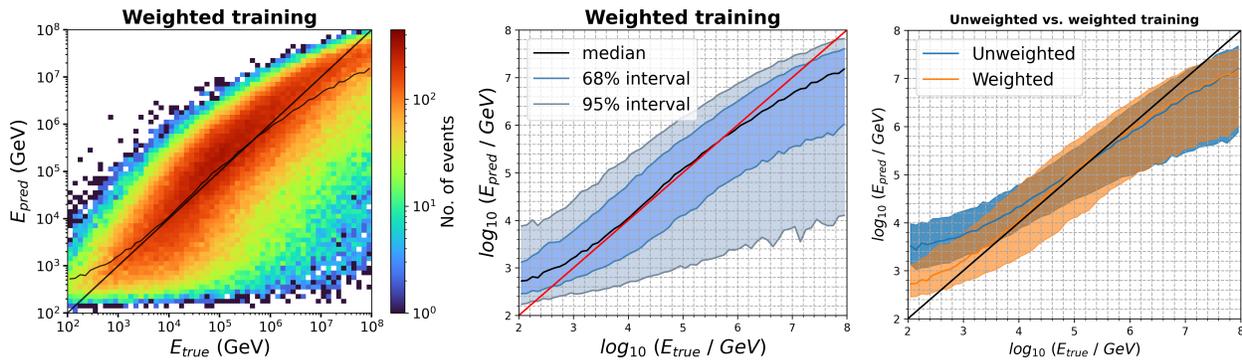
Όλες οι εκπαιδεύσεις που συζητήθηκαν μέχρι στιγμής πραγματοποιήθηκαν με την αρχική κατανομή της MC ενέργειας (Σχ. 8). Ως αποτέλεσμα, η μέση του ενεργειακού εύρους υπερεκπροσωπείται, ενώ τα άκρα –κυρίως οι χαμηλότερες ενέργειες, $E < 10^3$ GeV– υποεκπροσωπούνται. Αυτή η ανομοιόμορφη κατανομή σημαίνει ότι το μοντέλο δεν μαθαίνει αρκετά καλά τις υποεκπροσωπούμενες περιοχές του συνόλου δεδομένων, λόγω της έλλειψης δειγμάτων. Αυτό είναι εμφανές στα διαγράμματα διασποράς και ποσοστημορίων, όπου σχεδόν κανένα γεγονός με E_{true} κάτω από 1 TeV δεν ανακατασκευάζεται στη διαγώνιο. Η μη ισορροπημένη κατανομή της μεταβλητής-στόχου μπορεί επίσης να προκαλέσει μεροληψία του μοντέλου προς την πρόβλεψη τιμών στις πιο πυκνοκατοικημένες περιοχές του ενεργειακού εύρους. Αυτό θα μπορούσε να παράγει χαμηλότερες τελικές τιμές απωλειών, ωστόσο θα οδηγούσε σε προβλέψεις που δεν βασίζονται μόνο στην υποκείμενη φυσική πληροφορία. Για τους λόγους αυτούς, εφαρμόστηκε στην εκπαίδευση η **στάθμιση δειγμάτων**, προκειμένου να εκπαιδευτεί το μοντέλο σε μια ομοιόμορφη κατανομή ενέργειας. Στη στάθμιση δειγμάτων, κάθε γεγονός i έχει ένα σχετικό βάρος w_i , το οποίο είναι αντιστρόφως ανάλογο με τον αριθμό των γεγονότων στο αντίστοιχο bin της ενέργειας. Αυτά τα βάρη πολλαπλασιάζουν την απώλεια κάθε γεγονότος, έτσι ώστε η απώλεια κάθε παρτίδας να είναι ένας σταθμισμένος μέσος όρος:

$$L_{batch} = \frac{1}{N} \sum_{i=1}^N w_i L_i, \quad (14)$$

όπου L_i είναι η απώλεια για το γεγονός i . Με αυτόν τον τρόπο, δίνεται μεγαλύτερη σημασία στα γεγονότα που προέρχονται από υποεκπροσωπούμενες ενεργειακές περιοχές, βελτιώνοντας έτσι την προβλεπτική ικανότητα του μοντέλου σε αυτές και αναγκάζοντάς το να γενικεύει καλύτερα σε όλο το ενεργειακό εύρος. Η τεχνική εφαρμόστηκε στο μοντέλο T5 με hits στιγμιότυπων, η εκπαίδευση του οποίου επαναλήφθηκε και θα αναφέρεται στο εξής ως **σταθμισμένη**. Τα διαγράμματα διασποράς και ποσοστημορίων για αυτό το δίκτυο, καθώς και το διάγραμμα σύγκρισης με το αντίστοιχο μη σταθμισμένο μοντέλο T5, παρουσιάζονται στο Σχ. 13.

Η εξέταση των αποτελεσμάτων της σταθμισμένης εκπαίδευσης μας οδηγεί στα ακόλουθα συμπεράσματα:

- Η διάμεσος στη σταθμισμένη εκπαίδευση είναι σαφώς πιο κοντά στη διαγώνιο από ό,τι η διάμεσος της μη σταθμισμένης εκπαίδευσης σε όλο το εύρος της ενέργειας. Το μοντέλο τείνει να παράγει προβλέψεις γύρω από τη σωστή τιμή για όλες τις ενέργειες: είναι περισσότερο ευθυγραμμισμένο με τη συνολική ενεργειακή κατανομή και δεν μεροληπτεί προς τις υπερεκπροσωπούμενες ενεργειακές περιοχές, ήτοι γενικεύει καλύτερα.
- Οι προβλέψεις στις χαμηλότερες ενέργειες είναι οι πιο βελτιωμένες, λόγω του μεγάλου βάρους που δόθηκε σε αυτά τα γεγονότα. Αυτό αποτελεί σημαντική βελτίωση σε σχέση με όλα τα προηγούμενα μοντέλα και είναι το μοναδικό μοντέλο NΔΓ που παρουσιάζει αυτήν την προβλεπτική ικανότητα στις χαμηλές ενέργειες.



Σχήμα 13: Αριστερά: Διάγραμμα διασποράς για την σταθμισμένη εκπαίδευση. Μέσον: Διάγραμμα ποσοστημορίων για την σταθμισμένη εκπαίδευση. Δεξιά: Διάγραμμα σύγκρισης μεταξύ της σταθμισμένης και της μη σταθμισμένης εκπαίδευσης (hits στιγμιότυπου T5). Συγκρίνονται τα διαστήματα εμπιστοσύνης 68%.

- (c) Μικρή βελτίωση παρατηρείται στις υψηλές ενέργειες, παρά το γεγονός ότι έχουν και αυτές αυξημένα βάρη. Η παρατηρούμενη υποεκτίμηση των υψηλών ενεργειών είναι συνέπεια της έλλειψης φυσικώς διαθέσιμων πληροφοριών στα hits των γεγονότων αυτών.
- (d) Η διασπορά των προβλέψεων στο μέσον του ενεργειακού εύρους έχει αυξηθεί. Αυτό είναι αποτέλεσμα των μικρότερων βαρών σε αυτή την περιοχή και είναι το κόστος για ένα αμερόληπτο μοντέλο.

Διερεύνηση των Προβλέψεων του Μοντέλου

Από τη συζήτηση των αποτελεσμάτων είναι προφανές ότι οι εκπαιδευμένες διαμορφώσεις του DynEdge, καθώς και το μοντέλο OrcaNet, παρουσιάζουν ορισμένα βασικά χαρακτηριστικά στις προβλέψεις τους, κοινά για όλα τα μοντέλα. Αν χωρίσουμε το ενεργειακό εύρος που μας ενδιαφέρει ($10^2 - 10^8$ GeV) σε τρεις περιοχές (χαμηλές, μεσαίες και υψηλές ενέργειες), μπορούμε να συνοψίσουμε τα ευρήματά μας στις ακόλουθες παρατηρήσεις:

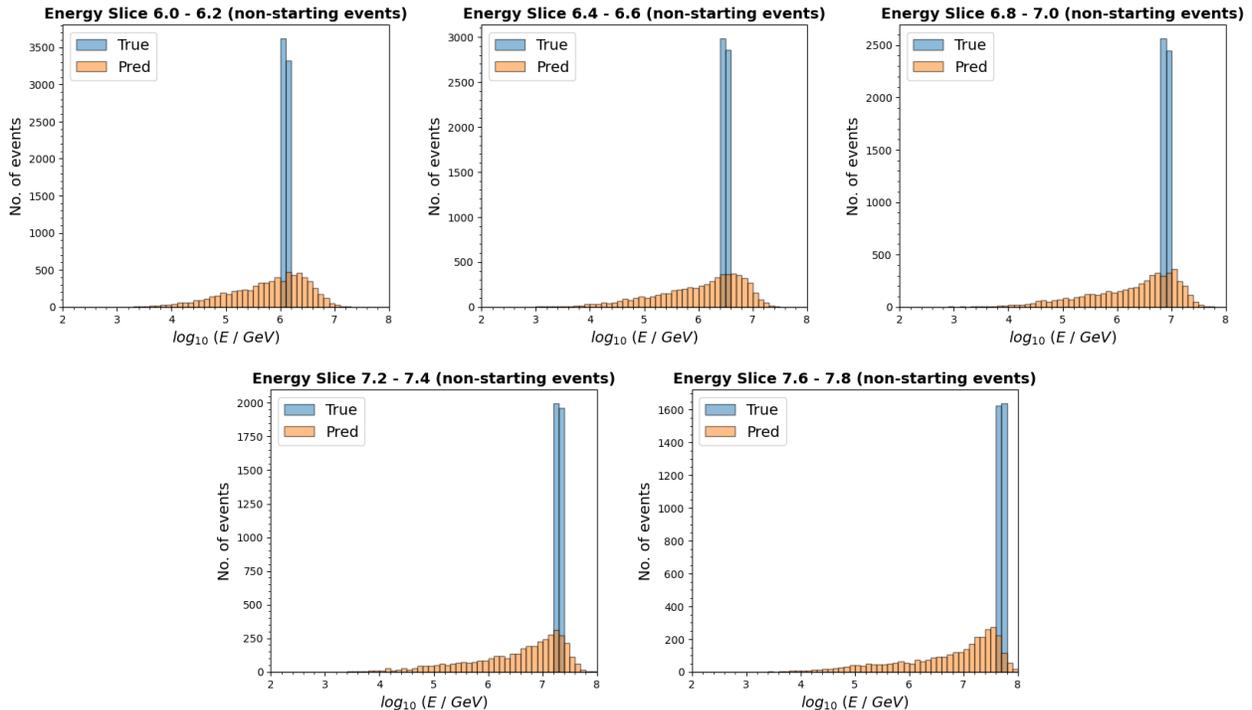
- (a) Στις χαμηλές ενέργειες ($10^2 - 10^4$ GeV) τόσο τα μοντέλα ΝΔΓ όσο και η τυπική ανακατασκευή παρέχουν υπερεκτίμηση της πραγματικής ενέργειας των νετρίνων. Σημαντική βελτίωση επιτεύχθηκε σε αυτή την περιοχή με τη σταθμισμένη εκπαίδευση, η οποία αυξάνει τη σημασία των γεγονότων χαμηλής ενέργειας, επιτρέποντας στο δίκτυο να τα μαθαίνει καλύτερα.
- (b) Στις μεσαίες ενέργειες ($10^4 - 10^6$ GeV) τα μοντέλα αποδίδουν βέλτιστα, δεδομένου ότι είναι η ενεργειακή περιοχή με τα περισσότερα δείγματα εκπαίδευσης. Οι προβλέψεις των μοντέλων σε αυτή την περιοχή είναι επικεντρωμένες γύρω από τη σωστή τιμή-στόχο. Τα νετρίνα αυτών των ενεργειών εναποθέτουν το μεγαλύτερο μέρος της ενέργειάς τους στον ανιχνευτικό όγκο, επιτρέποντας μια ακριβέστερη ανακατασκευή.
- (c) Στις υψηλές ενέργειες ($10^6 - 10^8$ GeV) όλα τα μοντέλα παρουσιάζουν εξαιρετικά παρόμοια συμπεριφορά. Καθώς η πραγματική ενέργεια των νετρίνων αυξάνεται πέραν των 10^6 GeV, τα μοντέλα αρχίζουν να την υποεκτιμούν όλο και περισσότερο. Η απόδοση αυτή δεν βελτιώνεται σημαντικά από τη σταθμισμένη εκπαίδευση, σε αντίθεση με τις χαμηλές ενέργειες. Αυτό υποδηλώνει ότι τα μοντέλα πλησιάζουν το φυσικό όριο της ακρίβειας της ανακατασκευής της ενέργειας. Αυτά τα υψηλής ενέργειας νετρίνα αφήνουν μόνο ένα μέρος του φωτός στον ανιχνευτή οδηγώντας έτσι στην υποεκτίμηση της ενέργειάς τους. Καθώς ο ανιχνευτής μεγαλώνει, η ανακατασκευή της ενέργειας στις υψηλές ενέργειες αναμένεται να βελτιωθεί.

Φαίνεται ότι στα δύο ενεργειακά άκρα, τα μοντέλα τείνουν να προβλέπουν τιμές πιο κοντά στη μέση του ενεργειακού εύρους (δηλαδή υπερεκτιμούν τα γεγονότα χαμηλής ενέργειας & υποεκτιμούν τα γεγονότα υψηλής ενέργειας). Αυτό θα μπορούσε να υποδηλώνει μια μεροληψία του μοντέλου προς τις ενέργειες που περιέχουν περισσότερα δείγματα εκπαίδευσης, ωστόσο, πρέπει να ληφθούν υπόψη δύο ελαφρυντικοί παράγοντες:

1. Η υπερεκτίμηση των χαμηλών ενεργειών είναι παρούσα τόσο στις προβλέψεις του ΝΔΓ, όσο και στην τυπική ανακατασκευή. Επομένως, το ζήτημα δεν έγκειται στη χρήση ενός συγκεκριμένου συνόλου γεγονότων για την εκπαίδευση ενός μοντέλου DL. Επιπλέον, αυτή η υπερεκτίμηση έχει μειωθεί σημαντικά με τη σταθμισμένη εκπαίδευση, η οποία έχει εξαλείψει κάθε πιθανή μεροληψία.

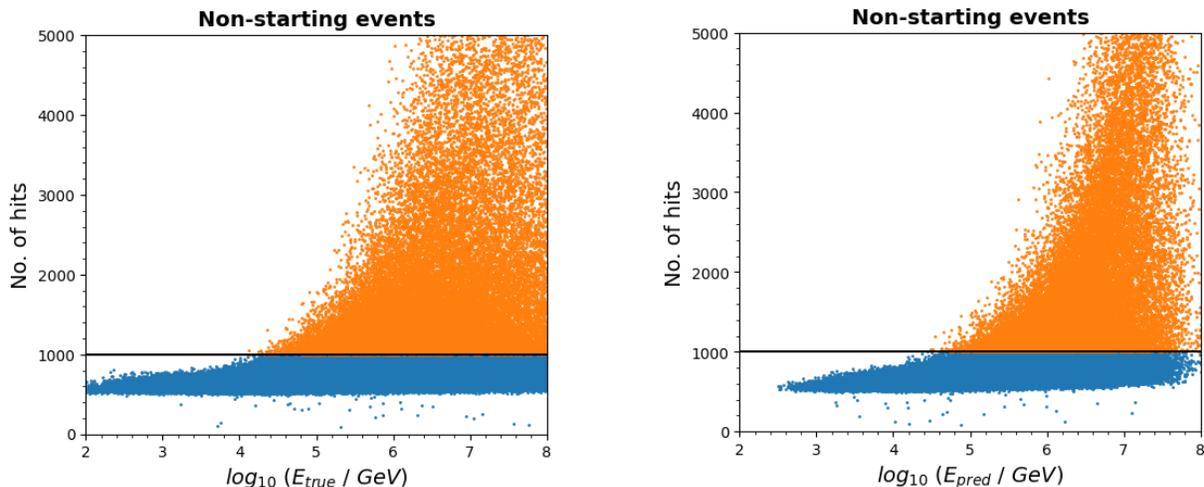
2. Η υποεκτίμηση της υψηλής ενέργειας είναι αποτέλεσμα του γεγονότος ότι δεν είναι ορατή όλη η ενέργεια κάθε νετρίνου στον ανιχνευτή. Τα γεγονότα για τα οποία το σημείο αλληλεπίδρασης βρίσκεται μακριά του ανιχνευτικού όγκου ονομάζονται *μη εκκινόντα*. Σε αυτά, (α) η συνιστώσα καταγισμού της αλληλεπίδρασης CC δεν είναι ορατή και (β) το μόνιο έχει ήδη χάσει μέρος της ενέργειάς του μέχρι να εισέλθει στον ανιχνευτή. *Επειδή η πλειονότητα των γεγονότων υψηλής ενέργειας αλληλεπιδρούν μακριά από τον ανιχνευτή, ισχύει επομένως ότι η ορατή ενέργεια είναι σημαντικά μικρότερη από την ενέργεια των νετρίνων*. Δεδομένου αυτού του γεγονότος, θα μπορούσε να υποστηριχθεί ότι η απόδοση του ΝΔΓ στις υψηλές ενέργειες είναι ακόμη καλύτερη από την αναμενόμενη, αφού η πλειονότητα των γεγονότων προβλέπεται σχετικά κοντά στην πραγματική τιμή τους, παρά τον περιορισμό αυτό.

Δεδομένου ότι ο ανιχνευτής ARCA προορίζεται κυρίως για τη μελέτη των νετρίνων υψηλών ενεργειών, προσπαθήσαμε να κατανοήσουμε καλύτερα τις προβλέψεις του μοντέλου στις υψηλές ενέργειες, εντοπίζοντας αιτιώδεις σχέσεις μεταξύ των δεδομένων εισόδου και της μεταβλητής-στόχου, τις οποίες χρησιμοποιεί το μοντέλο για να κάνει προβλέψεις. Έτσι, χωρίζουμε το σύνολο δεδομένων σε ενεργειακές τομές εύρους 0,2 σε $\log_{10} E_{true}$ και εξετάζουμε τις προβλέψεις του μοντέλου σε κάθε τομή. Περιοριζόμαστε σε μη εκκινόντα γεγονότα στην ενεργειακή περιοχή $10^6 - 10^8$ GeV, όπως φαίνεται στο Σχ. 14. Οι τομές της ενέργειας απεικονίζονται με μπλε, ενώ οι προβλέψεις για τα ίδια γεγονότα με πορτοκαλί. Για κάθε ενεργειακή τομή, η κορυφή της κατανομής των προβλέψεων βρίσκεται πολύ κοντά στην ίδια την τομή. Αυτό ισχύει ακόμη και στις υψηλότερες ενέργειες. Ωστόσο, καθώς μετακινούμαστε προς υψηλότερες ενέργειες, η κατανομή των προβλεπόμενων τιμών γίνεται σημαντικά ευρύτερη, με πολλές προβλέψεις να είναι πολύ χαμηλότερες από τις πραγματικές τιμές, διότι, όπως περιμέναμε, ένα μικρότερο ποσοστό της ενέργειας είναι ορατό κατά μέσο όρο.



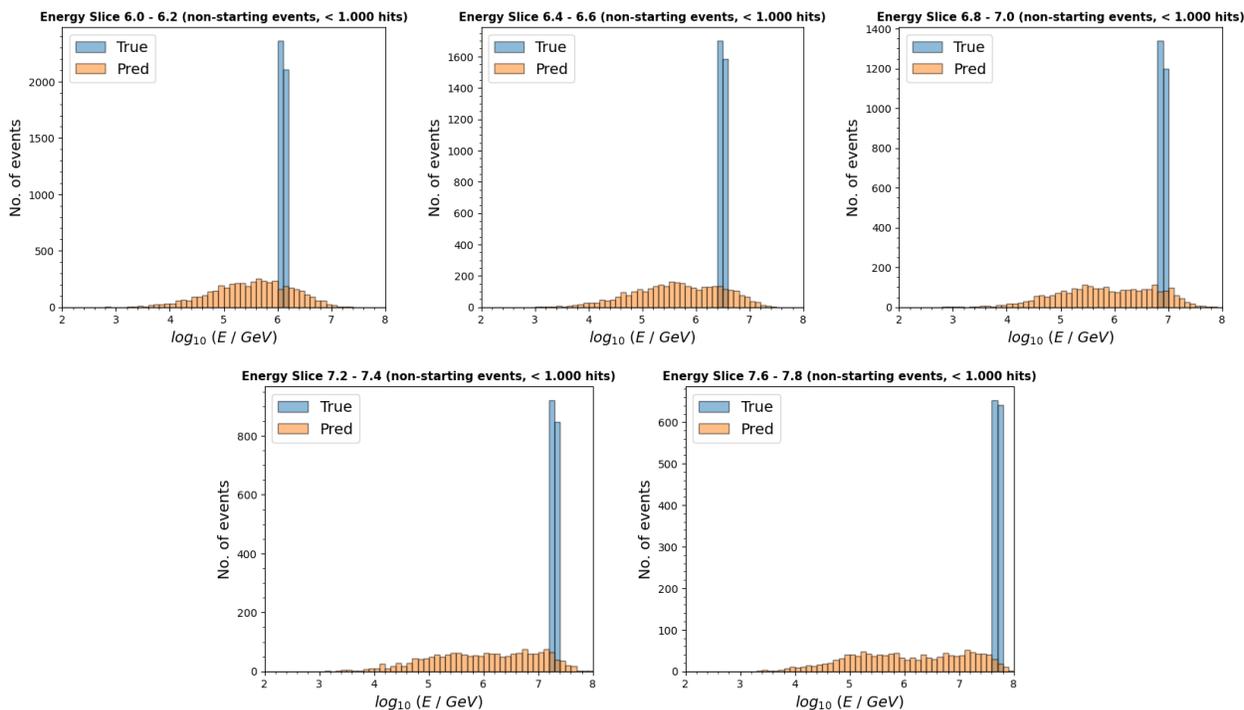
Σχήμα 14: Ενεργειακές τομές εύρους 0,2 σε $\log_{10} E_{true}$ (μπλε) και οι αντίστοιχες προβλέψεις του ΝΔΓ (πορτοκαλί) για μη εκκινόντα γεγονότα στην ενεργειακή περιοχή $10^6 - 10^8$ GeV.

Ωστόσο, η ύπαρξη κορυφών στις προβλέψεις στις υψηλότερες ενεργειακές τομές υποδηλώνει την ύπαρξη ενός ή περισσοτέρων χαρακτηριστικών που επιτρέπουν την αναγνώριση αυτών των γεγονότων ως γεγονότων υψηλής ενέργειας, παρά το γεγονός ότι είναι μη εκκινόντα. Ένα τέτοιο χαρακτηριστικό είναι ο αριθμός των hits σε κάθε γεγονός. Η απεικόνιση του αριθμού των hits σε σχέση με την ενέργεια των νετρίνων δείχνει ότι τα γεγονότα με περισσότερα από 1.000 hits βρίσκονται σχεδόν αποκλειστικά σε ενέργειες πάνω από 10^5 GeV, σχέση την οποία το μοντέλο έχει προσδιορίσει, όπως φαίνεται στο Σχ. 15. Με αφορμή αυτήν την παρατήρηση, επαναλαμβάνουμε τις ενεργειακές τομές, περιλαμβάνοντας μόνο τα μη εκκινόντα γεγονότα με λιγότερα από 1.000 hits (Σχ. 16). Προχωρώντας σε υψηλότερες ενέργειες τομή προς τομή, παρατηρούμε ότι η κατανομή των προβλέψεων υστερεί



Σχήμα 15: Αριθμός hits ανά γεγονός συναρτήσει της πραγματικής ενέργειας των νετρίνων (αριστερά) και της προβλεπόμενης ενέργειας από το ΝΔΓ (δεξιά). Γεγονότα με περισσότερα από 1.000 hits εμφανίζονται με πορτοκαλί χρώμα. Το δίκτυο τα αναγνωρίζει σωστά ως γεγονότα υψηλής ενέργειας.

ολόένα και περισσότερο από τις πραγματικές τιμές της ενέργειας, ενώ παράλληλα σταδιακά εξομαλύνεται. Στις υψηλότερες ενέργειες, η κατανομή των προβλέψεων έχει γίνει σχεδόν επίπεδη. Με άλλα λόγια, για αυτά τα γεγονότα, το μοντέλο υποεκτιμά σταθερά την ενέργεια στην κλίμακα PeV (πάνω από 10^6 GeV), όπως ακριβώς περιμέναμε.



Σχήμα 16: Ενεργειακές τομές εύρους 0,2 σε $\log_{10} E_{true}$ (μπλε) και οι αντίστοιχες προβλέψεις του ΝΔΓ (πορτοκαλί) για μη εκκινόντα γεγονότα με λιγότερα από 1.000 hits στην ενεργειακή περιοχή $10^6 - 10^8$ GeV.

Συμπεράσματα

Σε αυτή την εργασία, εξερευνούμε τις δυνατότητες των Νευρωνικών Δικτύων Γράφων στην αντιμετώπιση του προβλήματος ανακατασκευής γεγονότων σε πειράματα νετρίνων, συγκεκριμένα στην ανακατασκευή ενέργειας για γεγονότα τροχιάς στον ανιχνευτή νετρίνων KM3NeT/ARCA21, χρησιμοποιώντας το μοντέλο DynEdge. Αυτό καθίσταται εφικτό με την ενσωμάτωση της μορφής δεδομένων Βαθιάς Μάθησης του πειράματος KM3NeT στο GraphNeT, η οποία πραγματοποιήθηκε στο πλαίσιο αυτής της εργασίας. Συγκρίνουμε τις διάφορες διαμορφώσεις εκπαίδευσης και αναφέρουμε εκτενώς τα αποτελέσματα. Συνολικά, διαπιστώνουμε ότι η χρήση των hits στιγμιοτύπων ως εισόδου για την εκπαίδευση υπερτερεί της εκπαίδευσης με τα σκανδαλισμένα hits. Η αξιοσημείωτη αναφορά της εργασίας μας είναι η ενσωμάτωση της τεχνικής της στάθμισης δειγμάτων στο μοντέλο, μέσω της οποίας επιτυγχάνεται σημαντική βελτίωση της απόδοσής του στα γεγονότα χαμηλής ενέργειας. Η προκύπτουσα ανακατασκευή χαμηλής ενέργειας ξεπερνά όλα τα προηγούμενα μοντέλα, καθώς και το μοντέλο του OrcaNet και τον κλασικό αλγόριθμο ανακατασκευής του KM3NeT, με μόνο μέτρια επίπτωση στο μέσο της ενεργειακής κλίμακας. Η ενσωμάτωση της στάθμισης δειγμάτων εξαλείφει επίσης τις μεροληψίες των προηγούμενων μοντέλων λόγω της μη ομοιομορφής ενεργειακής κατανομής του συνόλου δεδομένων. Επιπλέον, η διερεύνηση της ανακατασκευής γεγονότων υψηλής ενέργειας παρέχει αιτιολόγηση των προβλέψεων του μοντέλου σε αυτό το ενεργειακό εύρος και μας επιτρέπει να εκτιμήσουμε τους φυσικούς περιορισμούς που επηρεάζουν την ανακατασκευή αυτών των γεγονότων. Τέλος, η πρώτη αυτή χρήση του πλαισίου GraphNeT στο KM3NeT αναδεικνύει τις δυνατότητες αυτού του εργαλείου Μηχανικής Μάθησης, το οποίο γίνεται πλέον διαθέσιμο στο πείραμα KM3NeT.

Μελλοντικές Επεκτάσεις

Τα ευρήματα και οι περιορισμοί που συζητούνται σε αυτήν την εργασία υποδεικνύουν ότι μπορεί να διεξαχθεί περαιτέρω έρευνα με στόχο τη βελτίωση των αποτελεσμάτων ανακατασκευής. Μελλοντικές επεκτάσεις θα μπορούσαν να περιλαμβάνουν:

- **Δημιουργία ισορροπημένου συνόλου δεδομένων:** Σε αυτήν την εργασία, η ανομοιομορφία του συνόλου δεδομένων αντιμετωπίστηκε μέσω της στάθμισης δειγμάτων. Αυτό αυξάνει τη σημασία των υποεκπροσωπούμενων περιοχών ενέργειας στο σύνολο δεδομένων, ωστόσο ο πραγματικός αριθμός δειγμάτων από αυτές τις περιοχές παραμένει χαμηλός. Μια εναλλακτική προσέγγιση θα ήταν η δημιουργία ενός ισορροπημένου συνόλου δεδομένων, με τον ίδιο αριθμό γεγονότων σε κάθε bin της ενέργειας, χρησιμοποιώντας την πλήρη παραγωγή MC δεδομένων για τον ανιχνευτή ARCA21. Αυτό θα εξέθετε το μοντέλο σε περισσότερα παραδείγματα εκπαίδευσης από τις υποεκπροσωπούμενες ενεργειακές περιοχές, επιτρέποντάς του ενδεχομένως να μάθει καλύτερες αναπαραστάσεις για αυτές.
- **Χρήση διαφορετικής μεταβλητής-στόχου:** Η μεταβλητή-στόχος για τις εκπαιδεύσεις των ΝΔΓ σε αυτήν την εργασία ήταν η MC ενέργεια του νετρίνου. Μπορεί, όμως, να είναι περισσότερο φυσικά αιτιολογημένη η χρήση μιας ποσότητας πλήρως προσδιορίσιμης από τις φυσικές πληροφορίες που είναι διαθέσιμες σε κάθε γεγονός, όπως η ορατή ενέργεια ή η ενέργεια του μιονίου σε ένα συγκεκριμένο σημείο της τροχιάς (η οποία είναι και η ποσότητα που εκτιμά ο κλασικός αλγόριθμος ανακατασκευής). Αυτό θα απαιτούσε την ενσωμάτωση της νέας μεταβλητής-στόχου στα αρχεία HDF5 του KM3NeT DL.
- **Χρήση νέων αρχιτεκτονικών δικτύου:** Παρόλο που η χρήση των ΝΔΓ έχει αποδειχθεί επιτυχής σε εργασίες ανακατασκευής για ανιχνευτές νετρίνων, υπάρχει αυξανόμενο ενδιαφέρον στην επιστημονική κοινότητα για την εξερεύνηση και άλλων τύπων μοντέλων. Ιδιαίτερα, η εφαρμογή των Μετασχηματιστών (Transformers), είτε αυτόνομα είτε σε υβριδικές αρχιτεκτονικές ΝΔΓ + Μετασχηματιστής, έχει προσελκύσει ενδιαφέρον, με τέτοιου είδους μοντέλα να έχουν ήδη εφαρμοστεί στο GraphNeT.
- **Εφαρμογή σε μεγαλύτερες εκδοχές του ανιχνευτή:** Στην τρέχουσα μορφή τους, οι ανιχνευτές του πειράματος KM3NeT έχουν μόνο ένα μέρος του προγραμματισμένου τελικού τους μεγέθους. Καθώς οι ανιχνευτές επεκτείνονται, επεκτείνονται και οι ικανότητές τους στην ανίχνευση νετρίνων, ιδιαίτερα όσον αφορά τις περιοχές υψηλής ενέργειας του φάσματος. Η εφαρμογή μοντέλων Μηχανικής Μάθησης σε εργασίες ανακατασκευής για μεγαλύτερες εκδοχές του ανιχνευτή μπορεί έτσι να προσφέρει μια πιο ακριβή αναπαράσταση των δυνατοτήτων αυτών των τεχνικών και καλύτερη σύγκρισή τους με τις κλασικές μεθόδους.

Chapter 1

Introduction

1.1	Overview and Motivations	46
1.2	Contributions	46
1.3	Thesis Outline	47

1.1 Overview and Motivations

Neutrinos are fundamental particles that play a crucial role in our understanding of the universe, yet they are notoriously difficult to detect due to their weak interactions with matter. The study of neutrinos serves various scientific goals, such as understanding the mechanisms behind high-energy astrophysical phenomena and probing the fundamental properties of neutrinos themselves. Large-scale water Cherenkov detectors such as KM3NeT/ARCA, an underwater neutrino telescope located in the Mediterranean Sea, observe neutrino interactions by detecting the Cherenkov light produced when energetic neutrinos interact with atomic nuclei in water.

One of the critical challenges in neutrino detectors is the accurate reconstruction of the physical properties of each neutrino interaction, or *event*, such as the energy and direction of the incoming particle, based on the pattern of optical hits in the detector photosensors. Reconstruction tasks in neutrino detectors have traditionally been approached using classical algorithms such as maximum likelihood estimation, which rely on well-established statistical techniques to optimize the fit between observed data and physical models of neutrino interactions. In recent years, however, there has been growing interest within the Experimental Neutrino Physics community around the use of Deep Learning methods to tackle event reconstruction tasks. In particular, Graph Neural Networks (GNNs) have lately attracted attention, since the lack of grid-like structure and the sparsity present in neutrino event data render their representation as graphs appropriate and well-motivated. Graph Neural Networks have shown promise in improving neutrino event reconstruction at only a fraction of the per-event computational cost and, for this reason, major experiments are proceeding to incorporate GNNs in event reconstruction studies. To this end, KM3NeT has utilized the ParticleNet GNN architecture, while the IceCube experiment has proposed the DynEdge model, an alternative to ParticleNet, and developed GraphNeT, a Deep Learning framework specifically designed for use in neutrino telescope event reconstruction tasks.

1.2 Contributions

This thesis explores the application of DynEdge, the Graph Neural Network architecture proposed by IceCube, to the problem of energy reconstruction in the KM3NeT/ARCA neutrino detector, using the GraphNeT Deep Learning framework. Our main contributions are as follows:

- Integration of the KM3NeT Deep Learning data format in GraphNeT, which extends the list of GraphNeT-compatible file formats to include HDF5 files and enables the first use of GraphNeT in the KM3NeT Collaboration. This tool thus becomes readily available for use in future KM3NeT endeavors.
- Evaluation of different dataset, model and training configurations with respect to their effect on the quality of the energy reconstruction and determination of the best configuration choices.
- Direct comparison of the DynEdge architecture with ParticleNet, the Graph Neural Network architecture employed in previous KM3NeT studies. The comparison validates the configuration choices and the reconstruction capabilities of both models.
- Substantial improvement of the energy reconstruction at low neutrino energies (below 10 TeV) compared to existing results, attained via the implementation of event weighting to balance the energy distribution of the input dataset. This practice additionally eliminates any bias relating to the input energy distribution, thus achieving a better-generalizing model.
- Investigation into the reconstruction of high-energy (PeV scale) neutrino events, providing justification of the model predictions in this energy range and allowing better appreciation of the physical constraints which limit the reconstruction of these events.

1.3 Thesis Outline

This thesis is organized into the following Chapters:

- **Chapter 1: Introduction**
Provides an overview of the motivations behind this study, as well as its contributions.
- **Chapter 2: Neutrino Physics**
Contains a short introduction to neutrinos, including their types, sources, detection mechanism and event topologies in underwater detectors.
- **Chapter 3: KM3NeT**
Introduces the KM3NeT experiment, including its scientific goals and detector design. The trigger system and background sources in the detectors are discussed. The problem of event reconstruction is defined and the maximum likelihood event reconstruction method is then described.
- **Chapter 4: Deep Learning & Graph Neural Networks**
Provides a brief overview of Deep Neural Networks. Graph Neural Networks are then discussed and the graph structure, the concept of graph convolutions and the DynEdge network architecture are introduced. A review of the use of Deep Learning and Graph Neural Networks in neutrino telescope event reconstruction is then provided.
- **Chapter 5: Energy Reconstruction with Graph Neural Networks**
Presents the application of the DynEdge network to the problem of energy reconstruction in KM3NeT/ARCA21. The integration of the KM3NeT data format in GraphNeT is examined. The dataset that was used, as well as various aspects of the training implementation are discussed. Finally, the findings of the trainings, including analysis and interpretation of the results, are then extensively presented, highlighting our contributions.
- **Chapter 6: Conclusion**
Provides a summary of the thesis and its findings, as well as suggestions for future work.

Chapter 2

Neutrino Physics

2.1	Neutrinos in the Standard Model	50
2.1.1	The Standard Model of Particle Physics	50
2.1.2	Neutrinos	50
2.2	Neutrino Sources	51
2.2.1	Atmospheric neutrinos	52
2.2.2	Astrophysical neutrinos	53
2.3	Neutrino Detection	54
2.3.1	Neutrino interactions with matter	54
2.3.2	Cherenkov neutrino detection	55
2.3.3	Neutrino event topologies	55

2.1 Neutrinos in the Standard Model

2.1.1 The Standard Model of Particle Physics

Much of the progress in modern physics has been driven by the quest to determine the fundamental constituents of matter and their interactions. In the second half of the 20th century, this quest materialized into a robust theoretical framework, the Standard Model (SM) of Particle Physics, which is in excellent agreement with the vast majority of experimental data [1]. In the SM, there exist 12 matter particles with half-integer spin, called *fermions*, divided into 6 *quarks* and 6 *leptons*. Quarks form composite particles, called *hadrons* (*e.g.* the proton $p(uud)$ or the neutron $n(udd)$). The 12 fermions can also be divided in three generations of 4, each generation containing heavier but otherwise similar particles to the previous one. Furthermore, for each particle, there exists a corresponding *antiparticle*, which has the same mass and spin, but opposite charges (*e.g.* electric charge). The SM also incorporates three of the four known fundamental interactions, or forces, (electromagnetic, strong & weak interactions), while gravity is not a part of the SM. Each interaction is mediated through force-carrier particles with integer spin, called *bosons*. The photon is the mediator for electromagnetism, the W^\pm and Z bosons mediate the weak force, while the strong force is mediated by gluons. The currently established SM framework is completed by the Higgs boson, whose associated quantum field is responsible for the generation of particle masses. An overview of the SM picture of the elementary particles of nature can be seen in Fig. 2.1.

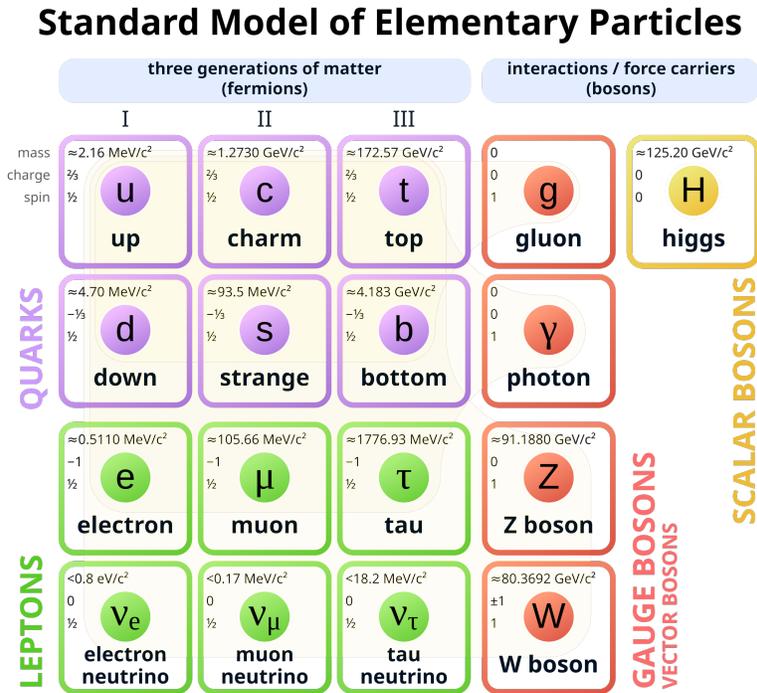


Figure 2.1: Particles in the SM of Particle Physics. The faint lines indicate which fermions (*i.e.* quarks & leptons) interact with each force-mediating boson. Antiparticles are not shown. From [2].

2.1.2 Neutrinos

Neutrinos are SM fermions of spin $1/2$ that belong in the lepton family. There are three neutrino flavors, one for each of the three fermion generations. Each neutrino flavor is associated with a corresponding charged lepton (electron, muon or tau). Thus, the three neutrino flavors are the electron neutrino ν_e , the muon neutrino ν_μ and the tau neutrino ν_τ . Their corresponding antiparticles are denoted by $\bar{\nu}_e$, $\bar{\nu}_\mu$, $\bar{\nu}_\tau$ respectively. Like all leptons, neutrinos do not interact with the strong force and, since they are electrically neutral, they do not interact electromagnetically either. Consequently, in the SM, neutrinos only participate in the weak interaction.

Neutrinos were postulated in 1930 by W. Pauli as a means to explain the continuous energy spectrum of electrons in β -decays, while ensuring energy, linear and angular momentum conservation [3]. The term “neutrino” was coined by E. Fermi, who in 1934 included the particle in his theory of β -decay [4]. The existence of the neutrino was finally established by the Cowan-Reines experiment in 1956, in which electron antineutrinos produced in a nuclear reactor were detected [5]. The muon neutrino was discovered in 1962 by L. Lederman, M. Schwartz and J. Steinberger [6], while the tau neutrino was only discovered in 2000, by the DONUT Collaboration at Fermilab [7]. The observation of neutrino flavor oscillations, a phenomenon where neutrinos produced in flavour ν_α can later be detected in flavour ν_β , by the Super-Kamiokande Collaboration in 1998 [8] and the SNO Collaboration in 2002 [9] confirmed that neutrinos possess non-zero mass. This is in contrast to the SM, according to which neutrinos are massless. The mechanism through which neutrinos gain mass, the neutrino oscillation parameters and the exact values of the neutrino masses are topics of active experimental and theoretical research [10].

2.2 Neutrino Sources

Neutrinos are the most abundant known massive particles in the Universe. There exist a number of different types of neutrino sources, producing neutrinos with energies that span many orders of magnitude, from (possibly) sub-meV up to EeV¹, as can be seen in Fig. 2.2. For the energy range of interest in this thesis (high-energy neutrinos in the TeV/PeV regime), the most important contributions to the neutrino flux come from atmospheric and astrophysical neutrinos.

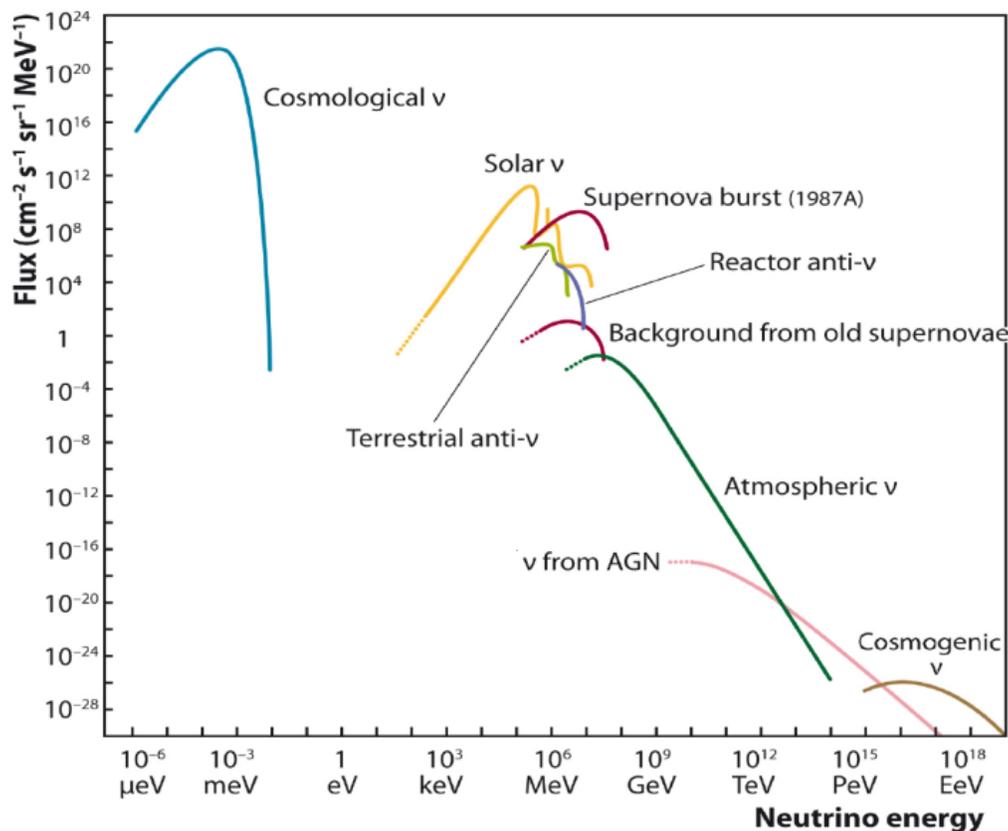


Figure 2.2: Measured and expected neutrino fluxes from different neutrino sources as a function of neutrino energy. From [11].

¹1eV = $1.602 \times 10^{-19} J$

2.2.1 Atmospheric neutrinos

Atmospheric neutrinos are produced in cosmic ray interactions with atomic nuclei in the Earth's atmosphere. Cosmic rays are high-energy charged particles of galactic and extragalactic origin, mainly protons and other atomic nuclei, that constantly bombard the Earth. Their collisions with atomic nuclei in the atmosphere result in particle cascades, as illustrated in Fig. 2.3, in which charged pions π^\pm and kaons K^\pm are produced in large numbers. These *mesons*² are unstable and can decay into muons:

$$\begin{aligned}\pi^+, K^+ &\longrightarrow \mu^+ + \nu_\mu, \\ \pi^-, K^- &\longrightarrow \mu^- + \bar{\nu}_\mu.\end{aligned}\tag{2.1}$$

Muons are themselves unstable, decaying into electrons and two neutrinos:

$$\begin{aligned}\mu^+ &\longrightarrow e^+ + \nu_e + \bar{\nu}_\mu, \\ \mu^- &\longrightarrow e^- + \bar{\nu}_e + \nu_\mu.\end{aligned}\tag{2.2}$$

This is the primary mechanism by which atmospheric neutrinos are created. These neutrinos constitute the *conventional neutrino flux* and, as evident by the above decay chain, are expected in flavor ratio $\nu_e : \nu_\mu : \nu_\tau \approx 1 : 2 : 0$. Since cosmic rays are positively charged nuclei, more π^+, K^+ than π^-, K^- are produced and therefore the ratio of neutrinos to antineutrinos is greater than 1. The ratio $\nu_e : \nu_\mu$ (i.e. $\frac{\nu_\mu + \bar{\nu}_\mu}{\nu_e + \bar{\nu}_e}$) increases with energy, since at higher energies more muons can reach the surface of the Earth before decaying, due to relativistic time dilation [12, 13]. Cosmic ray cascades also produce heavier hadrons - containing at least one charm or bottom quark- whose decays also produce atmospheric neutrinos. These neutrinos constitute the *prompt neutrino flux*, which is expected to dominate over the conventional flux for neutrino energies greater than 1 PeV [13, 14]. The atmospheric neutrino flux is the dominant neutrino flux in the GeV to TeV energy range.

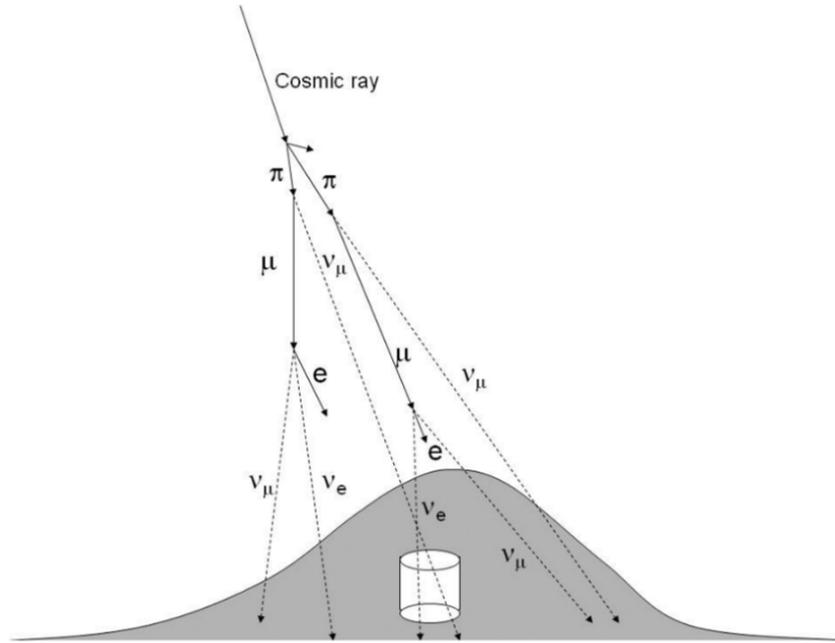


Figure 2.3: Atmospheric neutrino production in a cosmic ray cascade, caused by the interaction of the primary cosmic ray with an atmospheric nucleus, over the underground Super-Kamiokande neutrino detector. From [12].

²Hadrons containing one quark and one antiquark.

2.2.2 Astrophysical neutrinos

Cosmic rays are produced in a variety of astrophysical sources in the Universe, in which hadrons are accelerated to very high energies. As a byproduct of the acceleration of hadrons at cosmic ray sources, high-energy neutrinos (as well as γ -rays) are produced, often referred to as astrophysical neutrinos. There are two main processes contributing to neutrino production at cosmic ray sources, *proto-hadronic* and *photo-hadronic* interactions. In proto-hadronic interactions, high-energy protons interact with other high-energy protons/nuclei, resulting in the production of (among others) charged mesons, mostly pions π^\pm , and also kaons K^\pm , which eventually decay leptonically, producing neutrinos:

$$p + p/N \longrightarrow X + \pi^\pm/K^\pm, \quad (2.3)$$

$$\begin{aligned} \pi^+, K^+ &\longrightarrow \mu^+ + \nu_\mu, \\ \pi^-, K^- &\longrightarrow \mu^- + \bar{\nu}_\mu, \end{aligned} \quad (2.4)$$

$$\begin{aligned} \mu^+ &\longrightarrow e^+ + \nu_e + \bar{\nu}_\mu, \\ \mu^- &\longrightarrow e^- + \bar{\nu}_e + \nu_\mu. \end{aligned} \quad (2.5)$$

In photo-hadronic interactions, high-energy protons interact with photons, producing Δ^+ baryons³, which decay into pions and protons/neutrons:

$$\begin{aligned} p + \gamma &\longrightarrow \Delta^+ \longrightarrow p + \pi^0, \\ p + \gamma &\longrightarrow \Delta^+ \longrightarrow n + \pi^+. \end{aligned} \quad (2.6)$$

In the second decay mode, π^+ particles then decay as above, while neutrons decay into protons, thereby also producing neutrinos:

$$n \longrightarrow p + e^- + \bar{\nu}_e. \quad (2.7)$$

Since pion decay is the main mechanism which produces neutrinos in cosmic ray sources, the production flavor ratio is expected to be $\nu_e : \nu_\mu : \nu_\tau \approx 1 : 2 : 0$. However, due to neutrino oscillations, astrophysical neutrinos are expected to reach the Earth with flavor ratio $\nu_e : \nu_\mu : \nu_\tau \approx 1 : 1 : 1$ [15]. The astrophysical neutrino flux is sub-dominant to the atmospheric neutrino flux below 100 TeV, however astrophysical neutrinos dominate over atmospheric in the PeV scale.

Astrophysical neutrino sources can be of both galactic and extragalactic origin. Galactic sources are thought to include supernova remnants (SNRs), pulsar wind nebulae (PWNe), binary systems such as microquasars, as well as galactic structures like the Fermi bubbles [15, 16], while extragalactic sources, which are expected to dominate the cosmic ray spectrum at energies above 1 PeV, are thought to include Active Galactic Nuclei (AGNs), Gamma-ray bursts (GRBs), tidal disruption events (TDEs) and starburst galaxies [17, 18].

Unlike charged cosmic ray particles, which can be deflected by interstellar magnetic fields, and high-energy photons, which can be absorbed by the interstellar medium, neutrinos travel virtually unobstructed and therefore their direction of travel points to the original cosmic ray source. This property makes astrophysical neutrinos ideal cosmic messengers and has allowed for the establishment of a new astronomy sub-field, *neutrino astronomy*. Together with the emerging field of Gravitational Wave astronomy, as well as traditional astronomy (*i.e.* observations across the EM spectrum) and cosmic ray observations, they provide multiple “messengers” for studying astrophysical objects and processes. This practice of combining observations from different messengers in astronomy research is called *multi-messenger astronomy* [19–21].

³Hadrons containing three quarks or three antiquarks.

2.3 Neutrino Detection

2.3.1 Neutrino interactions with matter

The main mechanism by which high-energy neutrinos (*e.g.* atmospheric and astrophysical neutrinos) interact with matter is Deep Inelastic Scattering (DIS) between them and nucleons (*i.e.* protons or neutrons in atomic nuclei). In DIS, the neutrino interacts with an individual quark of the nucleon, depositing enough energy to the quark to cause the destruction of the original nucleon and the production of a hadronic cascade [22].

As mentioned in Sec. 2.1, neutrinos in the SM participate only in the weak interaction, which is mediated via the W^\pm and Z bosons. There are therefore two types of neutrino-nucleon interactions: *charged current* (CC) interactions mediated by W^\pm bosons and *neutral current* (NC) interactions mediated by Z bosons. These interactions are represented below, where N refers to the nucleon and X to the hadronic cascade, while $l = e, \mu, \tau$:

$$\begin{aligned} CC : \nu_l + N &\xrightarrow{W^+} l^- + X, \\ CC : \bar{\nu}_l + N &\xrightarrow{W^-} l^+ + X, \\ NC : \bar{\nu}_l + N &\xrightarrow{Z} \bar{\nu}_l + X. \end{aligned} \tag{2.8}$$

In general, the interaction of neutrinos with electrons has a much smaller cross section⁴ than the neutrino-nucleus interactions and thus can often be ignored. However, the interaction of antielectron neutrinos with electrons to produce W^- bosons exhibits a resonance at center-of-mass energy equal to the W boson mass, $\sqrt{s} = M_W = 80.4$ GeV, which, in the electron rest frame, corresponds to a neutrino energy of $E_\nu = \frac{M_W^2}{2m_e} = 6.3$ PeV. This is known as the *Glashow resonance* [23, 24]:

$$\bar{\nu}_e + e^- \rightarrow W^- \rightarrow \begin{cases} X & \sim 67.4\% \\ l^- + \bar{\nu}_l & \sim 32.6\% \text{ (all flavors)}. \end{cases} \tag{2.9}$$

where X again refers to a decay into hadrons and $l = e, \mu, \tau$. Indicative cross sections for the neutrino-matter interactions described above can be seen in Fig. 2.4.

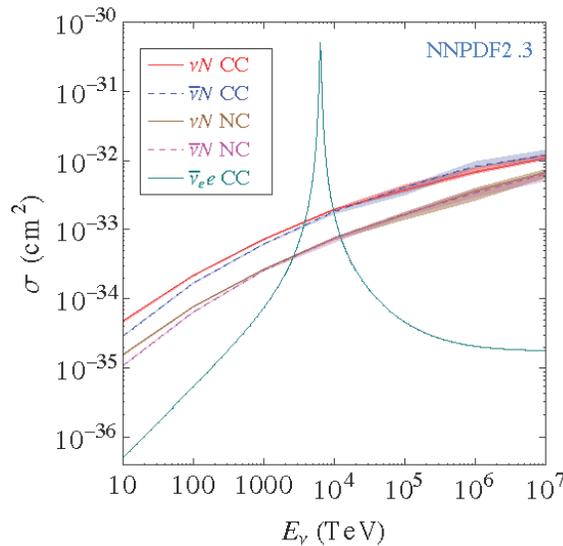


Figure 2.4: Neutrino-matter interaction cross sections in the TeV and PeV regime (neutrino-nucleus CC/NC & Glashow resonance). From [25].

⁴In physics, the cross section of an interaction is a measure of the probability that this interaction will occur between two particles.

2.3.2 Cherenkov neutrino detection

Neutrino detection in neutrino observatories relies on the phenomenon of *Cherenkov radiation*. Cherenkov radiation is produced by charged particles moving through a dielectric medium faster than the phase velocity of light in that medium [26]. This happens because charged particles will polarize the particles of the medium nearest to them. As they return to their ground state, the medium particles emit the excess energy as photons. If the velocity of the charged particles is greater than that of light in that medium, successive wavefronts overlap, interfering constructively and resulting in a conical wavefront at a characteristic angle from the particle track (Cherenkov angle):

$$\cos(\theta_c) = \frac{1}{\beta n}, \quad (2.10)$$

where n is the refractive index of the medium and $\beta = \frac{v}{c}$, v being the velocity of the charged particle. This is illustrated in Fig. 2.5. For relativistic ($\beta \approx 1$) charged particles in water, $\theta_c \approx 42^\circ$.

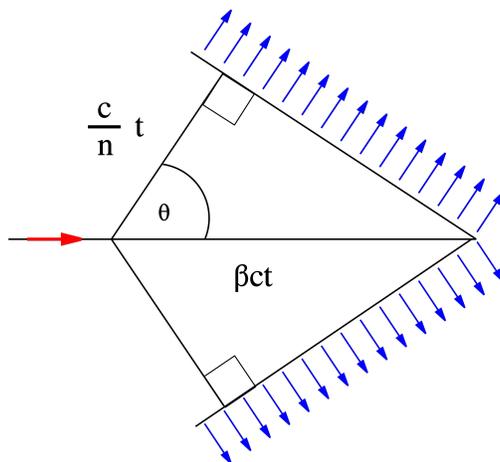


Figure 2.5: Cherenkov radiation (in blue) emitted at Cherenkov angle θ from the track of a charged particle which is moving horizontally at velocity $v = \beta c$ inside a medium with refractive index n . From [27].

As evident from Eqs. 2.8, 2.9, when neutrinos interact with matter, charged secondary particles are produced. If some of these particles –or their decay products– are moving with superluminal velocities in the medium in which they are travelling, Cherenkov radiation will be emitted, which can then be detected with arrays of suitable photodetectors, called *photomultiplier tubes* (PMTs).

In order to detect these Cherenkov photons, neutrino observatories need to operate in conditions of darkness. Furthermore, they need to be shielded against atmospheric muons, which also produce Cherenkov radiation as they pass through matter. Lastly, one should note that (a) neutrinos have extremely small cross sections, since they only interact via the weak interaction, and (b) the neutrino flux decreases rapidly with energy in the high-energy regime (see *e.g.* Fig. 2.2). It is therefore necessary for neutrino observatories to be sufficiently large and thus be based on an abundant dielectric medium, such as water. In order to meet the above conditions, neutrino observatories are usually constructed underwater (KM3NeT [28], ANTARES [29], Baikal-GVD [30]), in the Antarctic ice (IceCube [31]) or in large underground water tanks (Super-Kamiokande [32]).

2.3.3 Neutrino event topologies

The different types of neutrino-matter interactions presented in Sec. 2.3.1 lead to different particle products and therefore different event signatures, or topologies, in neutrino detectors. There exist two basic kinds of topologies that can be part of a neutrino event signature: *tracks & showers*. A track refers to a single charged particle producing the characteristic Cherenkov radiation cone. The only particle that produces a track topology in Cherenkov neutrino detectors is the muon, since high-energy muons can travel up to several kilometers in water before decaying. A shower refers to a cascade of particles and can be either hadronic or electromagnetic. Hadronic showers originate from the energy transferred to quarks and consist primarily

of hadrons. On the other hand, electromagnetic showers originate from electrons. High-energy electrons lose energy via the process of *bremstrahlung* (German for “breaking radiation”), *i.e.* photon emission. These high-energy photons then produce electron-positron⁵ pairs, which are themselves subject to energy loss via *bremstrahlung*. This process is hence repeated several times until the energy that was carried by the original electron has been dissipated, thus creating the cascade [33]. High-energy showers are generally cylindrically shaped, with length of the order of 10 m and diameter of the order of 10 cm. Individual charged particles produced in showers also emit Cherenkov radiation according to Eq. 2.10, provided they move at superluminal speeds inside the medium. However, the collective emission pattern of all the particles in a shower is not a cone but a spherical wavefront with increased intensity at the Cherenkov angle with respect to the cascade axis [15].

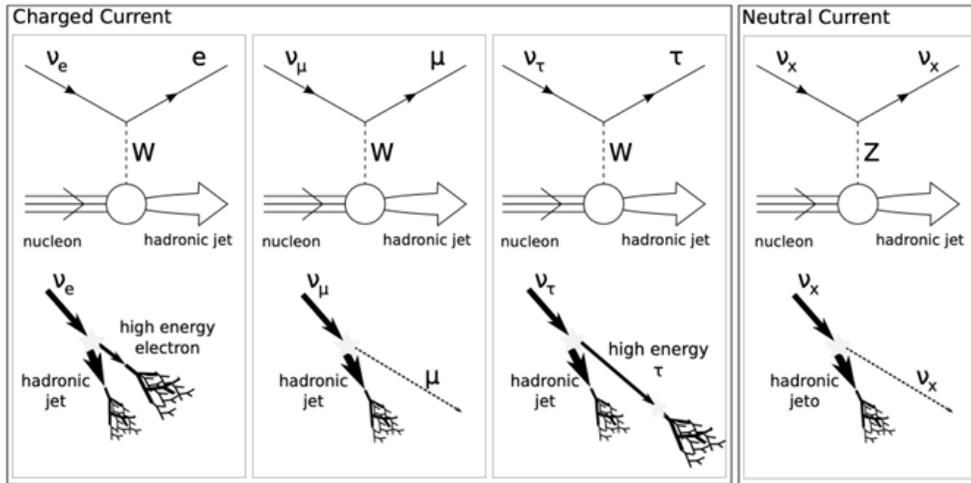


Figure 2.6: Overview of the event signatures produced in neutrino-nucleon CC & NC interactions. For NC interactions, the event signature is the same for all three neutrino flavors. From [34].

All neutrino-nucleon interactions (Eq. 2.8) contain a hadronic shower component, as can be seen in Fig. 2.6. This component originates from the nucleon and is a result of the energy deposited to it during the interaction. For NC interactions, regardless of the neutrino flavor, this is the only available event signature, since the outgoing neutrino does not interact any further. In electron (anti)neutrino interactions, the outgoing e^\mp will additionally produce an electromagnetic shower, with the same starting point as the hadronic shower. In muon (anti)neutrino interactions, the outgoing μ^\mp constitutes a particle track. NC and $(\bar{\nu}_e)$ CC events are commonly called *shower-like* events or *shower* events, while $(\bar{\nu}_\mu)$ CC events are commonly called *track-like* events or *track* events.

For $(\bar{\nu}_\tau)$ CC events the situation is slightly more complicated. Since taus have a much shorter lifetime than muons, they quickly decay, having traversed only a short distance from the interaction vertex. Tau particles decay in the following manner (shown for τ^- , similar for τ^+):

$$\tau^- \longrightarrow \begin{cases} X & \sim 64.8\% \\ e^- + \bar{\nu}_e + \nu_\tau & \sim 17.8\% \\ \mu^- + \bar{\nu}_\mu + \nu_\tau & \sim 17.4\% \end{cases} \quad (2.11)$$

where X represents a hadronic cascade. For the first two decay modes, a hadronic or electromagnetic shower will be produced respectively. The overall signature of the event will therefore consist of the initial hadronic shower associated with all neutrino-nucleon interactions, as well as a second shower, overlapping with the first but with a different interaction vertex. This is often called a *double-bang* event. Given the proximity of the two interactions vertices, double-bang events are usually difficult or even impossible to resolve. For the case of the muon decay of the tau, the event signature is a muon track instead of a second shower, making this a track-like event.

⁵Established name for the antielectron.

Finally, $\bar{\nu}_e e^-$ Glashow resonance event signatures can be inferred from Eq. 2.9. These events will usually produce a hadronic shower. For the W^- decay into an electron, an electromagnetic shower will be produced, while for the decay into a muon the outcome will be a muon track. For the tau decay of the W^- , the result will depend on the decay of the tau, as described above (2.11). Therefore, the Glashow resonance produces a track only in a little over 10% of decays.

For the remainder of this thesis, we will focus on track-like events, specifically track-like events from $(\bar{\nu}_\mu^-)$ CC interactions, since this is the main neutrino-induced track channel in neutrino detectors.

Chapter 3

KM3NeT

3.1	The KM3NeT Experiment	60
3.1.1	KM3NeT/ARCA	60
3.1.2	KM3NeT/ORCA	60
3.2	Detector Design	61
3.3	Data Acquisition and Triggering	62
3.4	Background Sources	63
3.4.1	Noise background	63
3.4.2	Physics background	63
3.5	Event Reconstruction	64

3.1 The KM3NeT Experiment

The *Cubic Kilometer Neutrino Telescope* (KM3NeT [28]) is an international research collaboration developing next-generation neutrino telescopes in the depths of the Mediterranean Sea. KM3NeT, a successor to ANTARES [29], consists of two underwater Cherenkov neutrino detectors, both utilizing the same detection principle but a different detector layout, optimised for each detector’s scientific goals. The two detectors are KM3NeT/ARCA (*Astrophysical Research with Cosmics in the Abyss*) and KM3NeT/ORCA (*Oscillations Research with Cosmics in the Abyss*). An overview of the KM3NeT member institutes and detector locations can be seen in Fig. 3.1.

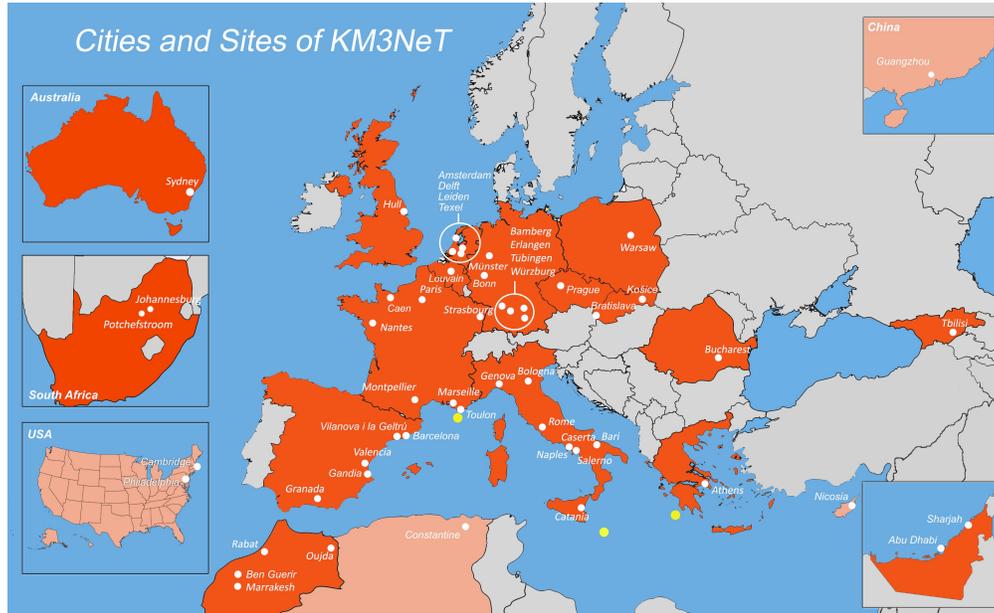


Figure 3.1: Overview of the KM3NeT Collaboration member institutes’ locations (white) and detector sites (yellow). Apart from the two detector sites for ARCA and ORCA, a third suitable site is shown off the coast of Pylos, Greece. Courtesy of KM3NeT.

3.1.1 KM3NeT/ARCA

The ARCA detector is located approximately 100 km southeast of Capo Passero in Sicily, Italy, deployed at a depth of 3,500 m. ARCA’s scientific objective is to contribute in neutrino astronomy, *i.e.* observe neutrinos of astrophysical origin. To this end, ARCA consists of sparse photodetector arrays, forming a 1 Gton detector, in order to be sensitive to neutrinos in the TeV/PeV energy range, where astrophysical neutrinos dominate the neutrino flux. ARCA aims to study the diffuse astrophysical neutrino flux, initially discovered by IceCube [35], as well as to determine astrophysical neutrino point sources of galactic and extragalactic origin. For this reason, its location on the globe is advantageous since its field of view contains the galactic centre and most of the galactic disk. Furthermore, its position in the Northern Hemisphere, *i.e.* in the opposite hemisphere to the IceCube detector in the South Pole, means that the two telescopes have complimentary fields of view.

3.1.2 KM3NeT/ORCA

The ORCA detector is located approximately 40 km offshore Toulon, France, near the site of the decommissioned ANTARES detector, at a depth of 2,450 m. ORCA’s scientific objective is related to the phenomenon of neutrino oscillations, in which neutrinos produced in flavour ν_α can later be detected in flavour ν_β , where $\alpha, \beta = e, \mu, \tau$, $\alpha \neq \beta$. This signifies the existence of three neutrino mass states ν_1, ν_2, ν_3 , with masses m_1, m_2, m_3 respectively. Using neutrino oscillations, the squared mass differences $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$, where $i, j = 1, 2, 3, i \neq j$, can be determined, therefore also determining the neutrino mass hierarchy, that is the ordering of the three neutrino masses. There exist two independent squared mass differences, say Δm_{21}^2 and

Δm_{31}^2 , since $\Delta m_{32}^2 = \Delta m_{31}^2 - \Delta m_{21}^2$. While Δm_{21}^2 has been determined to be $\Delta m_{21}^2 = 7.4 \cdot 10^{-5} \text{ eV}^2$ (*i.e.* positive, $m_2 > m_1$), the sign of Δm_{31}^2 has not yet been determined, with $|\Delta m_{31}^2| = 2.5 \cdot 10^{-3} \text{ eV}^2$ [36, 37]. As a result, there are two possible neutrino mass orderings, the *normal ordering*, where $m_1 < m_2 < m_3$, and the *inverted ordering*, where $m_3 < m_1 < m_2$. ORCA aims to determine the sign of Δm_{31}^2 and thus the neutrino mass hierarchy. To this end, ORCA focuses on atmospheric neutrinos in the GeV energy range, thereby using much denser photodetector arrays than ARCA, which form a detector of about 7 MTons.

3.2 Detector Design

As previously stated, the KM3NeT detectors are water Cherenkov detectors. This means that they employ arrays of photodetectors to capture Cherenkov radiation emitted by the neutrino interaction products, as explained in Sec. 2.3.2. The KM3NeT detectors are modular, consisting of vertical photodetector arrays called *Detection Units* (DUs). The final ARCA detector will contain 230 such DUs, arranged in two separate building blocks of 115, while the final ORCA detector will contain 115 DUs in a single building block. Each DU consists of 18 so-called *Digital Optical Modules* (DOMs), 17-inch diameter pressure-resistant glass spheres, containing the actual photodetectors and the readout electronics. The DOMs are the most fundamental components of the KM3NeT detectors. Each DOM contains 31 3-inch *photomultiplier tubes* (PMTs) facing in different directions, allowing Cherenkov light to be detected from almost any angle [28, 38]. A KM3NeT multi-PMT DOM can be seen in Fig. 3.2. Apart from the DOMs, each DU also contains an electro-optical cable for power and data transmission. The DUs are anchored to the seabed and are held upright by buoys. Note that this design entails that the DUs can be displaced horizontally due to ocean currents and so the precise position of the DOMs has to be monitored. The detector is linked to the shore via the main electro-optical cable, which transmits power to the detector and data to the shore. An artist's impression of a KM3NeT detector can be seen in Fig. 3.3.



Figure 3.2: A KM3NeT multi-PMT Digital Optical Module (DOM). From [28].

On average, the horizontal distance between DUs is 90 m for ARCA and 20 m for ORCA, while the vertical spacing of the DOMs in each DU is 36 m for ARCA and 9 m for ORCA. This gives ARCA DUs a total height of about 700 m and ORCA DUs a total height of about 200 m. Each of the two ARCA building blocks will have a radius of approximately 500 m and a volume of 0.5 km^3 , giving the full ARCA detector an instrumented volume of 1 km^3 . The full ORCA detector will be much smaller, with a radius of only about 100 m and an effective volume 150 times smaller than ARCA. ARCA's much larger instrumented volume

of seawater is necessary to observe astrophysical neutrinos at a reasonable event rate, since they sit in the high-energy, low-flux part of the neutrino energy spectrum. The detector configuration of interest for this thesis is the ARCA detector with 21 DUs deployed, hence referred to as *ARCA21*, since this is the latest version of ARCA for which data-taking was complete while work for this thesis was being carried out.

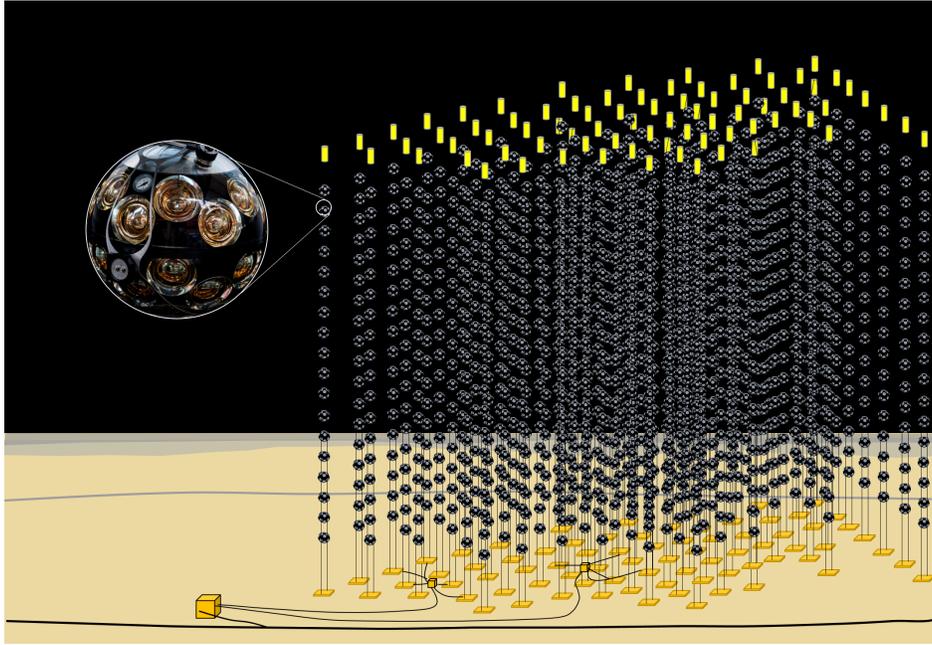


Figure 3.3: Artist’s impression of a KM3NeT detector on the sea floor. The DUs, consisting of a base, 18 DOMs and a buoy, can clearly be seen. Courtesy of KM3NeT.

3.3 Data Acquisition and Triggering

KM3NeT operates under an all-data-to-shore concept. For each PMT photon hit, the photon time of arrival, the Time-over-Threshold (ToT) and a PMT identifier are recorded and sent to shore. The photon time of arrival is defined as the moment when the amplitude of the PMT signal surpasses the predefined threshold of 0.3 photoelectrons. Similarly, the ToT is the time interval for which the amplitude remains over the aforementioned threshold.

The data rate for a complete building block of 115 DUs is expected to be roughly 25 Gb/s. Data reduction using real-time on-shore event triggering is therefore required before data can be stored on disk. Triggering refers to the process of identifying and selecting candidate events of interest from the detector data stream using appropriate criteria. The KM3NeT trigger system uses three trigger levels:

- **L0 trigger:** refers to the threshold for a PMT pulse to be recorded as a hit, which is applied offshore. L0 hits are therefore all the recorded hits.
- **L1 trigger:** refers to a local coincidence of two or more L0 hits from different PMTs in the same DOM within a fixed time window of $\Delta T = 10$ ns.
- **L2 trigger:** refers to more sophisticated coincidence filtering making use of the PMT orientations.

Different triggering algorithms can be applied to identify track-like and shower-like event topologies. For ARCA, an event is triggered if at least 5 L1 hits in different DOMs are identified as causally connected by the track or shower trigger algorithm [28]. All of the L1 hits which form an event are referred to as *triggered hits*. When an event is triggered, all the L0 hits in a time window around the triggered hits (*i.e.* starting before the first L1 hit and ending after the last L1 hit) are stored. These are called *snapshot hits* and are the complete set of hits in the detector during and around the event.

3.4 Background Sources

Apart from neutrino-induced signals, several background sources which can cause photon hits in the detector PMTs exist. These background sources can be divided into noise background and physics background.

3.4.1 Noise background

Noise background refers to processes contributing optical noise to the PMTs, which is rejected by the triggering algorithms. Noise background in the detectors stems from radioactive decays in seawater and bioluminescence.

Radioactive decays

Seawater contains trace amounts of radioactive elements, the most abundant of which is potassium-40 (^{40}K). ^{40}K has a half-life of $1.25 \cdot 10^9$ y and decays via beta decay or electron capture [39]:



With an energy around 1 MeV, the beta decay electron can produce Cherenkov radiation. Similarly, the gamma ray photon emitted from the excited ^{40}Ar nucleus in the electron capture process can scatter off electrons via Compton scattering. These electrons may then also emit Cherenkov radiation. ^{40}K decays represent the main source of noise in the KM3NeT detectors, creating a persistent optical background capable of producing L1 hits.

Bioluminescence

Bioluminescence is the emission of visible light by living organisms. It is highly common in the deep sea, where little to no sunlight is present. Bioluminescent organisms in these environments exhibit remarkable diversity, ranging from microorganisms such as bacteria to crustaceans, jellyfish and fish [40]. Larger animals usually produce short bioluminescence events, *e.g.* lasting a few seconds. On the other hand, bioluminescent microorganisms produce a more consistent, yet less intense optical background. Also, their bioluminescent activity is subject to seasonal variations. The effect of bioluminescence on the detectors can be reduced using a high-rate veto, where data from a PMT is vetoed if its hit count exceeds a certain threshold.

3.4.2 Physics background

Physics background refers to physical processes that produce signals similar to those of interest. These signals fulfill the trigger conditions and are thus identified as events. Therefore, their contribution needs to be subtracted during data analysis. The physics background of the KM3NeT detectors are atmospheric muons. For ARCA, atmospheric neutrinos are also considered part of the background.

Atmospheric muons

As evident from Eq. 2.1, the same cosmic ray showers which produce atmospheric neutrinos in the Earth's atmosphere also produce large numbers of muons. These so-called atmospheric muons do not always decay before reaching the surface and can even traverse large amounts of rock or water, gradually losing energy and emitting Cherenkov radiation in the process. Building neutrino detectors in deep underground or underwater locations therefore provides shielding against most of the atmospheric muons. Even so, the majority of triggered events in the KM3NeT detectors are atmospheric muon events. Unlike neutrinos, muons cannot pass through the Earth, meaning atmospheric muons only travel through the detector in downward angles, as can be seen in Fig. 3.4. For this reason, it is possible to remove their contribution by rejecting downgoing tracks during data analysis, leaving only upgoing or starting tracks (*i.e.* tracks that start inside the instrumented volume, indicating a neutrino interaction occurred) to be considered as neutrino events.

Atmospheric neutrinos (ARCA)

The ARCA detector, as described in Sec. 3.1.1, aims to detect astrophysical neutrinos. Therefore, specifically for ARCA, atmospheric neutrinos constitute a form of irreducible background, since they are indistinguishable from astrophysical neutrinos in the detector. However, as discussed in Sec. 2.2.2, the astrophysical neutrino flux dominates over the atmospheric flux above 100 TeV. Hence, even though differentiation of atmospheric and astrophysical neutrinos is not possible on an event-by-event basis, one can study the astrophysical neutrino flux by focusing on the higher part of the neutrino energy spectrum, where the contribution of atmospheric neutrinos becomes negligible.

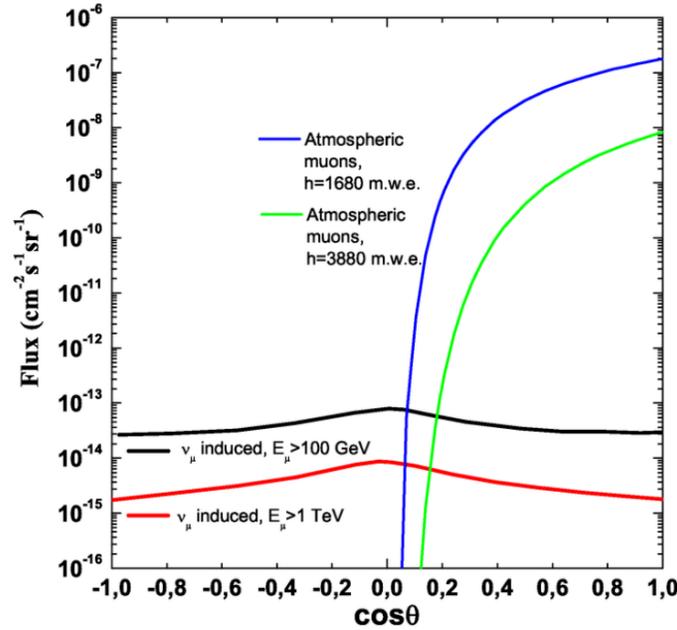


Figure 3.4: Flux of atmospheric muons (at different water depths) and atmospheric neutrino-induced muons (of different energies) as a function of the cosine of the zenith angle θ . For downward angles, *i.e.* $\cos(\theta) > 0$, the atmospheric muon contribution dominates the total flux. From [41].

3.5 Event Reconstruction

As explained in Sec. 3.3, once an event has been triggered, all of its snapshot hits are stored. Stored hits contain the following information, where position refers to the PMT position, while the direction components form a unit vector according to the PMT direction:

- position- x , $-y$, $-z$
- direction- x , $-y$, $-z$
- time of arrival t
- Time-over-Threshold ToT

As with any Physics experiment, data collection in KM3NeT serves to provide physicists with the necessary data to perform analyses aimed towards achieving the Physics goals of the experiment. Such data analyses require the use of high-level physical information characterizing each neutrino event as a whole, such as the neutrino direction or the energy deposited in the detector as a result of the event. However, raw detector data consist solely of the individual PMT hits which constitute each event. That is to say, they are low-level data, from which higher-level information needs to be inferred. Therefore, before any data analysis can take place on the stored events, it is necessary that they are *reconstructed*.

Event reconstruction is the process of estimating parameters of interest for each event in question (such as the energy or direction of a particle), given the set of hits which constitute the event. In other words, it refers to the procedure of extracting high-level event information from low-level data recorded by the detectors. Event reconstruction in KM3NeT is accomplished using dedicated reconstruction algorithms, which take as input the set of PMT hits of each event and produce as output estimates of the reconstructed quantities of that event. Separate reconstruction algorithms are applied to track and shower events, to account for their differing topologies. The present work studies $\bar{\nu}_\mu$ CC interactions, therefore we will focus on the case of tracks by describing the corresponding track reconstruction algorithm.

The goal of the classical KM3NeT track reconstruction algorithm is to fit the hypothesis of a muon track to the event data (*i.e.* the hits), thereby determining the trajectory and energy of the muon. The algorithm is based on *maximum likelihood estimation*, *i.e.* it aims to maximise the likelihood that the hit pattern corresponds to the fitted muon trajectory and energy [28]. Since both the muon position and direction are fit parameters, the problem is non-linear and is tackled in successive steps, referred to as the *reconstruction chain*. This reconstruction process is outlined in [42].

Prefit

In the first step, called the prefit, a multitude of fits is performed, each one with a different assumed direction of the track, with a step of 1° . By assuming a track direction and ignoring the effects of photon scattering, the problem is reduced to a linear one. The coordinate system is defined such that the muon travels along the z -axis. The remaining parameters of the fit are then the position (x_0, y_0) of the muon and the time t_0 at which it crosses the $z = 0$ plane. The linear fit is based on the expected Cherenkov photon arrival times, which are a function of the fit parameters (x_0, y_0) & t_0 . The parameter values are obtained by minimization of the χ^2 , which expresses the time difference between the actual and expected photon arrival times for given (x_0, y_0) & t_0 . The fit is performed using only a subset of causally connected hits, in order to reduce the contribution from the optical background. The process is repeated for all directions and the 12 best-fit directions are stored to be used in the next reconstruction stage. The best-fit directions are determined by the fit quality Q , which is defined as:

$$Q = \text{NDF} - 0.25 \frac{\chi^2}{\text{NDF}}, \quad (3.2)$$

where NDF is the number of degrees of freedom, *i.e.* the number of hits used in the fit. The inclusion of the NDF in the fit quality serves to weigh fit directions with fewer used hits less favorably than directions with a large number of hits used in the fit.

Muon trajectory fit

In the following step in the chain, the muon trajectory is reconstructed with a maximum-likelihood search, using the best-fit directions from the prefit as a starting point. The muon direction and position are now fitted simultaneously, by minimization of a likelihood ratio function. This function is based on semi-analytical probability density functions (PDFs) of the photon arrival times, describing the chance of observing a PMT hit at a relative time Δt with respect to the expected arrival time of a direct hit from an unscattered Cherenkov photon. The PDFs incorporate the effects of light scattering and dispersion, as well as information such as the optical background rate and the quantum efficiency and angular acceptance of the PMTs. These PDFs, denoted here f , are used to construct a function $\mathcal{L}(H_1)$ characterizing the likelihood that the observed hit pattern was a result of the track hypothesis H_1 :

$$\mathcal{L}(H_1) = \prod_{\text{hits}} f(\rho_i, \theta_i, \phi_i, \Delta t), \quad (3.3)$$

where ρ_i is the distance of closest approach of the muon to PMT i , θ_i and ϕ_i describe the PMT orientation with respect to the track, and Δt is the difference between the expected and measured light arrival time. A similar function $\mathcal{L}(H_0)$ representing the likelihood that the hit pattern was caused by the optical background is also constructed. Track reconstruction is then performed by minimizing the likelihood ratio:

$$\lambda = \ln \frac{\mathcal{L}(H_0)}{\mathcal{L}(H_1)}. \quad (3.4)$$

This minimization identifies the track parameter values which are most compatible with the observed hit pattern. Once those have been determined, the muon trajectory has been fitted.

Muon energy fit

Finally, the muon energy is fitted by examining photon hits in all PMTs within a cylindrical volume surrounding the reconstructed track. Specifically, for each PMT in this cylinder, an estimate of the number of hits from various sources such as ^{40}K decays, electromagnetic showers and ionisation and radiative energy losses of the muon is produced using the PDFs. As a result, the expected number of hits for each PMT is calculated as a function of the muon energy and follows a Poisson distribution. The likelihood $\mathcal{L}(E_\mu)$ of the observed hit pattern as a function of the muon energy E_μ can then be determined using the hit/no-hit probability for all PMTs within the cylinder surrounding the track. Maximisation of this likelihood yields the reconstructed muon energy value.

The classical KM3NeT reconstruction chain described above already produces very good results. It will therefore serve as an excellent benchmark for improvement efforts using Deep Learning techniques, which will be discussed in the following Chapter.

Chapter 4

Deep Learning & Graph Neural Networks

4.1	Deep Learning	68
4.1.1	Architecture of Deep Neural Networks	68
4.1.2	Training of Deep Neural Networks	69
4.2	Graph Neural Networks	72
4.2.1	Graph structure	72
4.2.2	Edge convolution	73
4.2.3	Network architecture	73
4.2.4	Software tools	74
4.3	Overview of DL &GNN Applications in Neutrino Telescope Event Reconstruction	75

4.1 Deep Learning

Deep Learning (DL) is the subset of Machine Learning (ML) techniques using Deep Neural Networks (NNs) to identify patterns and extract meaningful representations from raw data. Despite the origins of modern NNs dating back to the 1940s [43, 44], it was not until the 2010s that the use of DL methods became widespread. This surge in popularity can be attributed to increases in computing power, brought about by the development of powerful GPUs and TPUs, the availability of large datasets, the development of more complex and sophisticated NN architectures and algorithmic advances [45]. These developments in the field resulted in remarkable performance improvements in tasks such as speech recognition, computer vision and natural language processing. During the last few years, DL has contributed in advancements in a variety of scientific disciplines, such as Biology [46], Medicine [47] and Material Science [48]. Physics has not been immune to this DL frenzy, with DL tools being successfully implemented in a wide range of physics tasks [49], both in theory and experiment, while playing a crucial role in searches for new physics [50].

In the context of this thesis, we focus on *supervised learning*, *i.e.* learning tasks which utilize a labeled training dataset, where the inputs x to the network have been paired with corresponding output target values, or *labels*, y . Thus, the network is tasked to infer a function F such that $F(x) = y$. Supervised learning problems most often fall under the categories of either *classification* or *regression*. In classification problems, the model is tasked to assign each input to one of two or more different categories, or *classes*, and thus the target variables are the discrete class labels. Conversely, in regression problems, the model aims to predict a continuous target variable from features of the input data.

4.1.1 Architecture of Deep Neural Networks

The structure of Deep Neural Networks typically consists of an input layer, one or more hidden layers, and an output layer. Each layer consists of a series of nodes, called *neurons* (see Fig. 4.1). The input layer serves as the entry point for data, with each neuron corresponding to a specific feature x_i of the input vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ ⁶. The hidden layers are where the network learns to extract features from the input data. Connections exist between each hidden layer neuron and some or all of the neurons of the previous layer. Each connection is characterized by a *weight* w . The output of hidden layer neurons can be computed from the weighted sum of their inputs. Specifically, for a given neuron i in layer l , the output $\alpha_i^{(l)}$ is given by:

$$\alpha_i^{(l)} = f \left(\sum_{j=1}^n w_{ij}^{(l)} \alpha_j^{(l-1)} + b_i^{(l)} \right). \quad (4.1)$$

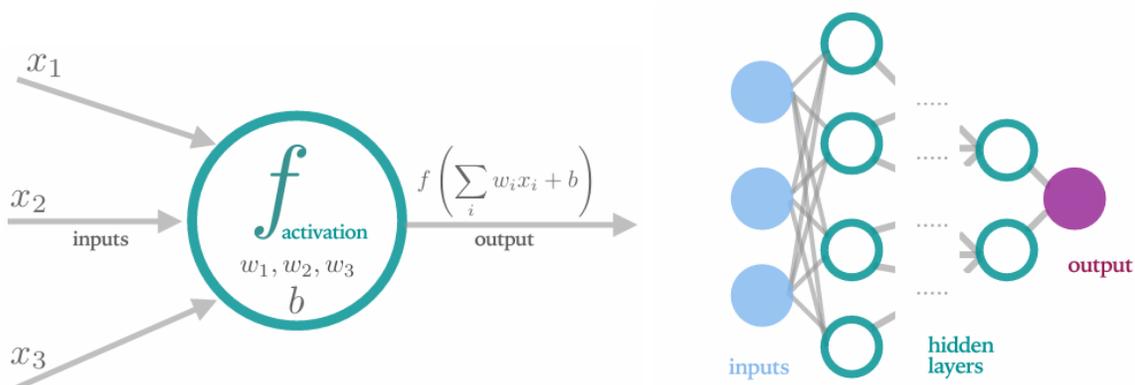


Figure 4.1: Left: Neuron with bias b and activation function $f(\cdot)$ receiving inputs x_i through connections with weights w_i and producing its output as in Eq. 4.1. Right: Schematic depiction of a Deep Neural Network. From [51].

⁶Vectors and matrices will be denoted in **bold** lowercase and capital letters respectively.

Here, $\alpha_j^{(l-1)}$ represents the output of the j -th neuron of the previous layer, $w_{ij}^{(l)}$ are the relevant weights, $b_i^{(l)}$ is known as the *bias*, a parameter which is associated with each neuron, and $f(\cdot)$ is the so-called *activation function*. Eq. 4.1 can be rewritten to refer to all the neuron outputs $\boldsymbol{\alpha}^{(l)} = [\alpha_i^{(l)}]$ of layer l , taking the simpler form:

$$\boldsymbol{\alpha}^{(l)} = f\left(\mathbf{W}^{(l)}\boldsymbol{\alpha}^{(l-1)} + \mathbf{b}^{(l)}\right), \quad (4.2)$$

where of course $\mathbf{W}^{(l)} = [w_{ij}^{(l)}]$ and $\mathbf{b}^{(l)} = [b_i^{(l)}]$. The activation function $f(\cdot)$ plays an important role by introducing non-linearity into the model, enabling the network to learn more complex patterns. Without non-linear activation functions, the network would be reduced to a linear transformation and would not be able to utilize complex relationships in the data. An example of a commonly used activation function is the Rectified Linear Unit (ReLU), defined as:

$$f_{\text{ReLU}}(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0. \end{cases} \quad (4.3)$$

The final layer of the network is the output layer. As the name suggests, its output constitutes the network response to a particular input. The network architecture which has been described here is often referred to as a multi-layer perceptron (MLP).

4.1.2 Training of Deep Neural Networks

In order for the network to provide reasonable output values to its inputs, it must first be trained. During the training process, the tunable parameters of the network, *i.e.* the *weights* and *biases*⁷ are optimized. This is in contrast to parameters such as the number and size of the network layers, which have been fixed before the training process and are called *hyperparameters*. Training a deep neural network involves an iterative process aimed at minimizing a *loss function*, which quantifies the difference between the network predictions and the actual target values. Each training step is composed of two phases: the *forward pass* and the *backpropagation*.

Forward pass & loss calculation

During the forward pass, the input data is passed through the network layer by layer to generate predictions. Each layer transforms the data according to Eq. 4.2, progressively extracting higher-level features until the final output is produced. Once the predictions are obtained, the loss function is calculated, providing a measure of the error in the network predictions. For regression tasks, which will be the focus of this thesis, examples of commonly used loss functions are L_1 and L_2 losses, *i.e.* the mean absolute error (MAE) and mean squared error (MSE):

$$L_{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|, \quad (4.4)$$

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (4.5)$$

where y_i is the true label of the i -th sample, \hat{y}_i is the network prediction and N is the number of training samples.

⁷From here on, unless otherwise noted, the weights and biases of the network will jointly be referred to simply as *weights*.

Backpropagation

After the loss has been computed, the adjustable parameters of the network, *i.e.* the weights, have to be tuned in order to decrease it. This is achieved through the process of backpropagation [52], in which the network computes the gradient of the loss function with respect to each weight, essentially propagating the error backward through the network. This gradient indicates how much the loss would change with a small adjustment to the weight, guiding the network on how to modify its parameters to reduce the loss.

If we define $\mathbf{z}^{(l)} \equiv \mathbf{W}^{(l)}\boldsymbol{\alpha}^{(l-1)} + \mathbf{b}^{(l)}$, then Eq. 4.2 becomes

$$\boldsymbol{\alpha}^{(l)} = f\left(\mathbf{z}^{(l)}\right). \quad (4.6)$$

Defining also the error of the i -th neuron in layer l as

$$\delta_i^{(l)} \equiv \frac{\partial L}{\partial z_i^{(l)}} \iff \boldsymbol{\delta}^{(l)} \equiv \frac{\partial L}{\partial \mathbf{z}^{(l)}}, \quad (4.7)$$

it can then be shown [53] that the gradients of the loss with respect to the weights and biases obey the following equations:

$$\begin{aligned} \frac{\partial L}{\partial w_{ij}^{(l)}} = \delta_i^{(l)} \alpha_j^{(l-1)} &\iff \frac{\partial L}{\partial \mathbf{W}^{(l)}} = \boldsymbol{\delta}^{(l)} \left[\boldsymbol{\alpha}^{(l-1)}\right]^T, \\ \frac{\partial L}{\partial b_i^{(l)}} = \delta_i^{(l)} &\iff \frac{\partial L}{\partial \mathbf{b}^{(l)}} = \boldsymbol{\delta}^{(l)}. \end{aligned} \quad (4.8)$$

The error term $\boldsymbol{\delta}^{(l)}$ is recursively calculated, starting from the output layer:

$$\begin{aligned} \boldsymbol{\delta}^{(l)} &= \left[\left(\mathbf{W}^{(l+1)}\right)^T \boldsymbol{\delta}^{(l+1)} \right] \odot f'(\mathbf{z}^{(l)}), \\ \boldsymbol{\delta}^{(L)} &= \frac{\partial L}{\partial \boldsymbol{\alpha}^{(L)}} \odot f'(\mathbf{z}^{(L)}), \end{aligned} \quad (4.9)$$

where \odot refers to the element-wise or Hadamard product of two matrices, $f'(\cdot)$ is the derivative of the activation function and (L) symbolises the output layer.

Optimization

Finally, the network weights are updated using the *gradient descent* iterative method. The idea behind gradient descent is to update the weights in the direction of the steepest decrease of the loss function. Over several training steps, this should lead the loss to converge to a local minimum in the parameter space of the weights. If we let \mathbf{w} be the vector of all the weights of the network, then in gradient descent the weights are adjusted according to:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \cdot \frac{\partial L}{\partial \mathbf{w}}, \quad (4.10)$$

with the components of $\frac{\partial L}{\partial \mathbf{w}}$ being calculated from Eqs. 4.8, 4.9. The parameter η is the *learning rate*, a hyperparameter that controls the step size of the weight updates. The choice of learning rate is important: a small η means the network follows the path of steepest descent more closely, but results in slow learning, requiring more iterations to converge, while a large η results in faster training, but may cause the training process to become unstable and overshoot the local minimum.

In practice, gradient descent is not the most efficient way to optimize the network weights. Hence, different optimization algorithms, or *optimizers*, are used [54]. The simplest and most well-known method is *Stochastic*

Gradient Descent (SGD). In SGD, each training step does not involve the entire training dataset, but only a limited number of samples, called a mini-batch or simply *batch*. The number of samples in each batch, known as the *batch size*, is a hyperparameter. The loss going into Eq. 4.10 is therefore the loss for a single batch, which is an approximation of the loss over the entire dataset. This greatly reduces the memory and computational requirements on the GPUs. Each training step contains a new batch of samples. One complete cycle through the entire training dataset over multiple training steps is called an *epoch*, with a typical training consisting of several epochs.

A commonly used, more advanced optimization algorithm is the *Adam* optimizer [55], which stands for “adaptive moment estimation”. Adam improves upon SGD by dynamically adjusting the learning rate for each parameter, allowing it to adaptively slow down or speed up the learning process for different parameters, leading to more efficient convergence. Defining $\mathbf{g}_t \equiv \frac{\partial L_t}{\partial \mathbf{w}_{t-1}}$ to be the gradient at training step t , Adam uses moving averages to estimate the first and second moments of the gradients:

$$\begin{aligned}\mathbf{m}_t &= \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t, \\ \mathbf{v}_t &= \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \mathbf{g}_t^2.\end{aligned}\tag{4.11}$$

The hyperparameters β_1 and β_2 are exponentially decaying rate parameters with commonly used values of 0.9 and 0.999 respectively. After a bias-correction:

$$\begin{aligned}\hat{\mathbf{m}}_t &= \frac{\mathbf{m}_t}{1 - \beta_1^t}, \\ \hat{\mathbf{v}}_t &= \frac{\mathbf{v}_t}{1 - \beta_2^t},\end{aligned}\tag{4.12}$$

the weights can finally be updated:

$$\mathbf{w}_t = \mathbf{w}_{t-1} - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t + \epsilon}}.\tag{4.13}$$

Here, the hyperparameter ϵ is a small constant added for numerical stability, with a common default value of 10^{-8} . Note that the division on the last term is element-wise.

Generalization

For a Machine Learning model to be successful, it must have the ability to *generalize*, that is perform well not only on its training data, but also on new, unseen data. In this context, two important phenomena which ought to be avoided are *underfitting* and *overfitting*. Underfitting occurs when the model cannot capture the underlying patterns in the data, leading to poor performance on both the training and unseen datasets. This is often a result of the model being too simplistic to represent the complex relationships in the data or the model training time being too short. It can usually be avoided by increasing the model complexity or training duration. On the other hand, overfitting happens when the model begins learning specific details of the training data, including noise and outliers, essentially memorizing the training set. This results in excellent performance on the training set but poor generalization to new data.

In order to ensure the ability of the model to generalize and avoid overfitting, the available dataset is often split into three parts: the *training set*, the *validation set* and the *test set*. The training set is used to fit the model, *i.e.* learn the underlying patterns and relationships in the data by adjusting its parameters to minimize the loss function, as described in the previous sections. To guard against overfitting, the validation set is used during the training process. The model performance is checked on the validation set at regular intervals, typically at the end of every epoch, and the validation loss (or other appropriate performance metric) is compared to the training loss. If the model begins performing significantly better on the training set than on the validation set, then this can be a sign of the onset of overfitting. When this is the case, *early stopping* of the training can be utilized, since the model performance is not expected to improve. Conversely, if the model performs poorly on both the training and validation sets, it may be underfitting, indicating that the model is too simple or insufficiently trained to capture the essential patterns in the data. Even though the

network is never actually trained on the validation set, its repeated usage as a measure of performance can create a bias towards models showing an improved performance on this particular set. For this reason, the test set is reserved for evaluation of the final model, providing an unbiased assessment of its generalization ability.

4.2 Graph Neural Networks

(Astro)particle physics experiments generate data with inherently graph-like structure. In the complex detectors of these experiments, such as the KM3NeT detectors, events consist of individual measurements which can be represented as points in space and time. These data points are sparsely scattered and lack a regular grid structure, while their distribution and number vary on an event-by-event basis. This is true for example in KM3NeT, where the number and location of hits varies greatly in each event, while the precise DOM positions are constantly changing due to sea currents. Furthermore, the data points are unordered, since they represent particle interactions, which themselves have no particular ordering. These attributes make graphs a natural way to represent particle detector events. *Each event produces a graph, with graph nodes representing particle interactions (e.g. detector hits) and graph edges describing relations between them.*

Graph representation is thus more conceptually justified than image-based representations, employed for example in Convolutional Neural Networks (CNNs). In images, a limited resolution in space and time has to be chosen, meaning the loss of information on the precise location of detector hits. Additionally, extensive zero-padding, which creates overhead, is needed to account for the sparsity of the data as well as irregular detector geometries. In other representations, such as sequence-based representations, e.g. Recurrent Neural Networks (RNNs), an artificial ordering of the data points has to be introduced, which may not be optimal. It is therefore no surprise that *Graph Neural Networks* (GNNs), which are not limited by these restrictions, have recently gathered significant interest in Experimental Particle Physics, as a promising tool for DL applications. GNNs are already being implemented in tasks such as particle identification and reconstruction, with reviews of their use in Particle Physics provided in [56–58]. The rest of this Section is devoted to providing an outline of the graph structure, network architecture and software tools for the GNNs used in this thesis.

4.2.1 Graph structure

Each KM3NeT event can be represented as a directed graph $G = (V, E)$. Each graph node v_i represents a PMT hit, with the hit information, i.e. position- x , $-y$, $-z$, direction- x , $-y$, $-z$, time t and ToT, being referred to as *node* or *hit features*. If all these features are used, each node will therefore have 8 node features. An example of the graph representation of an event can be seen in Fig. 4.2.

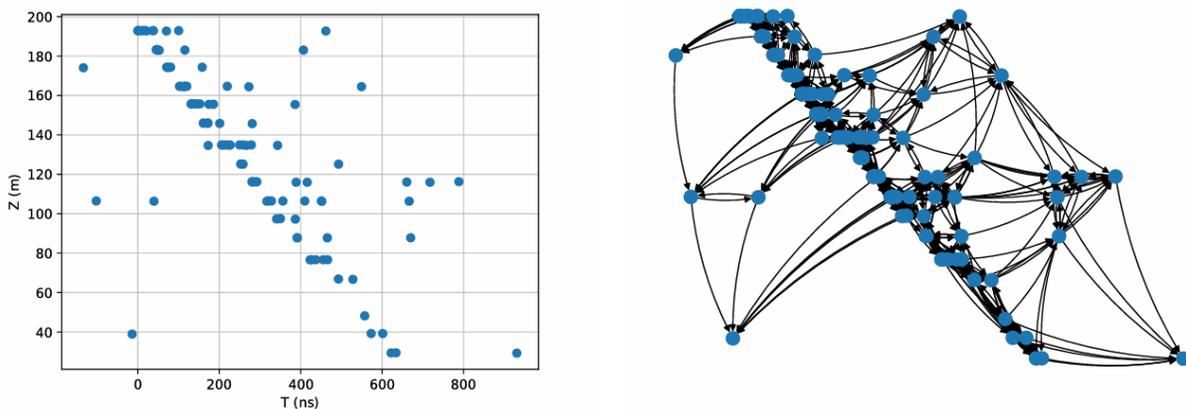


Figure 4.2: Left: Hit distribution of a KM3NeT event (position z vs. time t). Right: Graph representation of the same event. Each hit is now a graph node, with edges to its 8 nearest neighbors, calculated using Eq. 4.14. From [59].

Due to the large and variable size of graphs in KM3NeT data, the use of fully connected graphs is not practical, especially since the number of edges in a fully connected graph grows quadratically with the number of nodes. For this reason, each node is connected to its k -nearest neighbors (k NN), where k is a hyperparameter. The calculation of the nearest neighbors requires that an appropriate distance metric be defined. The most obvious choice, which will be used in this work, is to use the 4D Euclidean metric in space and time:

$$d^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 + c^2 \Delta t^2, \quad (4.14)$$

where c is the speed of light. Calculating the nearest neighbors is computationally expensive, since the calculation of the distance between every pair of nodes is required, making this step $\mathcal{O}(n^2)$, n being the number of nodes, *i.e.* hits in the event.

4.2.2 Edge convolution

The fundamental component of the GNNs used in this thesis is the *Edge Convolution* block, or *EdgeConv* block, introduced in [60] to operate on point clouds. The idea behind EdgeConv is to extend the concept of convolutions from CNNs to GNNs. In CNN convolutional layers, a kernel of weights is slid across the image, recording the dot product of the kernel with the image pixels at every point. In EdgeConv, the concept of the kernel is replaced by the nearest neighbors, with the number of neighbors k of each node being equivalent to the kernel size. The convolution is then an operation acting on the outgoing edges of each node.

After determination of the nearest neighbors, *i.e.* determination of the graph edges, the edge features are defined. Edge e_{ij} from node v_i to node v_j is defined to have edge features $(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i) \in \mathbb{R}^{2F}$, with $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^F$ being the node features of nodes v_i, v_j respectively, F being the number of features in each node. The edge features are then propagated through a 3-layer MLP and summed across the nearest neighbors. The result is the updated node feature vector for each node of the graph. The new node feature vector $\mathbf{x}'_i \in \mathbb{R}^{F'}$ has a different number of features than before ($F' \neq F$). Mathematically, the edge convolution then consists in the following operation:

$$\mathbf{x}'_i = \sum_{j \in \mathcal{N}(i)} h_{\theta}(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i). \quad (4.15)$$

In this expression, h_{θ} is the MLP with learnable parameters θ and $\mathcal{N}(i)$ is the set of the k nearest neighbors to node v_i . In short, the EdgeConv block takes as input a graph with n nodes, each with F node features, and outputs a graph with the same number of nodes, each with F' features.

4.2.3 Network architecture

The GNN architecture that will be used in this thesis is named *DnyEdge* and was put forward by IceCube in [61]. The architecture is displayed in Fig. 4.3. DynEdge consists of 4 consecutive EdgeConv blocks. At the output of each block a graph which has the same number of nodes as the original, but 256 features per node, is produced. The graph edges are then updated to determine the new k nearest neighbors, thus creating a new graph in latent space. The stacking of consecutive EdgeConv blocks allows the network to learn higher-level features of the data. The output-graph of each block, along with the original graph, are concatenated into a $[n, 4 \cdot 256 + F]$ matrix, where n is the number of nodes and F the number of features per node of the original graph. The resulting matrix is feed into a 3-layer MLP which reduces its dimensionality down to $[n, 256]$. Pooling operations are then performed over the nodes, in up to 4 distinct ways: *mean*, *min*, *max* & *sum*, the output of each operation being an array of size $[1, 256]$. The arrays are then concatenated and thus, should all 4 options be used, the resulting array will have dimension $[1, 4 \cdot 256] = [1, 1,024]$. This array is then concatenated with the *global statistics*, a set of 5 values calculated from the original graph, characterizing it as a whole. These are the number of PMT hits, *i.e.* number of nodes n in the graph, and the homophily ratios for position- x , $-y$, $-z$ and time t . The homophily ratio for a given node feature is the percentage of graph edges which connect nodes that share the same value for this feature. It is thus a measure of the number of pulses which occurred at the same PMT or at the same time. The final $[1, 1,029]$ vector is passed to a final 3-layer MLP to produce the network output.

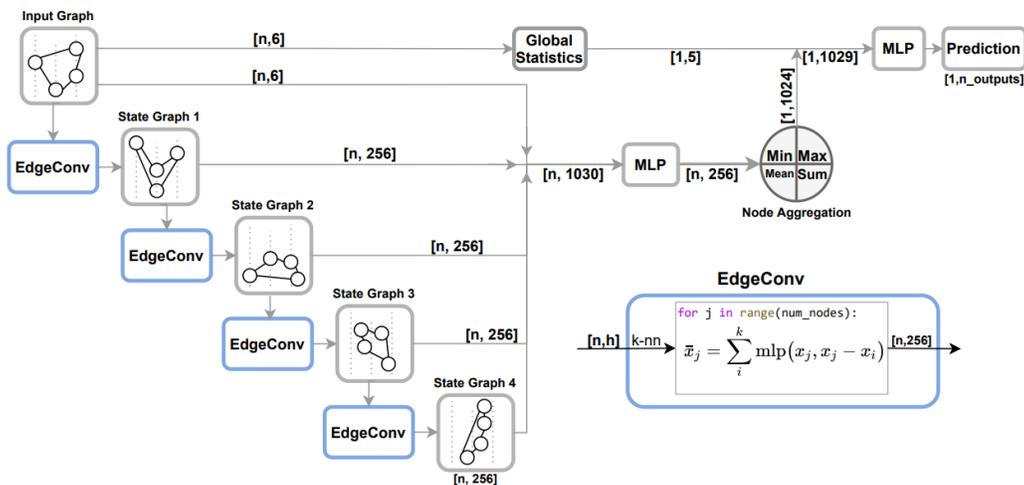


Figure 4.3: Architecture of the DynEdge network. The input graph is assumed here to have n nodes and 6 features per node. From [61].

4.2.4 Software tools

The work for this thesis was made possible by the use of software tools designed specifically for DL tasks in KM3NeT and other neutrino experiments. These tools are built upon popular ML libraries like PyTorch [62] and TensorFlow [63], but are tailored to the needs of the experiment(s) in which they will be used. In this case, the tools which are relevant are *GraphNeT* and the KM3NeT DL framework, *OrcaSong* and *OrcaNet*.

GraphNeT

GraphNeT [64] is an open-source PyTorch-based DL framework developed by the IceCube experiment, with the goal of providing neutrino physicists with DL tools to perform reconstruction tasks for any neutrino experiment. GraphNeT utilizes *PyG* (PyTorch Geometric) [65] and *PyTorch Lightning* [66] to construct and train GNNs and to provide user-friendly end-to-end functionality. All of the GNN trainings for this thesis were carried out using GraphNeT. The DynEdge model, presented above, has been implemented in GraphNeT by the IceCube Collaboration.

KM3NeT DL tools

The KM3NeT DL framework consists of two Python modules, named *OrcaSong* [67] and *OrcaNet* [68]. The former is responsible for data preprocessing, while the latter for network training:

OrcaSong

OrcaSong is used to convert KM3NeT files to a format that is efficient for DL applications. KM3NeT data are stored in ROOT files, a commonly used file format in high-energy physics. *OrcaSong* converts them into the HDF5 file format, while also removing information that is irrelevant for the NN training and the evaluation of the results. The resulting HDF5 files are much reduced in size and their internal structure allows for more efficient reading operations during training. For this reason, this is the format that was used for the input files to the GNN trainings in this work.

OrcaNet

OrcaNet is a KM3NeT-specific, TensorFlow-based module, used to carry out network trainings. The chosen network architecture in *OrcaNet* is outlined in Sec. 4.3. Although it will not be used directly in this work, results obtained with *OrcaNet* will be compared to those obtained in trainings with GraphNeT.

4.3 Overview of DL & GNN Applications in Neutrino Telescope Event Reconstruction

KM3NeT has successfully developed and used classical reconstruction algorithms, presented in [42]. As described in Sec. 3.5, these algorithms utilize least-squares fitting for a fast, but less accurate geometric fit and maximum likelihood approaches with more sophisticated modeling of the detector response for the final reconstruction. Similar reconstruction techniques have also been employed by other neutrino telescopes. In IceCube, for example, likelihood-based reconstruction algorithms adapted for both high- [69] and low-energy [70] reconstruction have been reported.

In recent years, the major advances in Deep Learning have led to the adoption of a wide range of DL methods in Particle Physics. An area in which this trend is prevalent is the reconstruction and classification tasks of most Particle Physics experiments. Drawing inspiration from image processing and computer vision, image-based methods, most notably CNNs, have been used. One of the first applications of a CNN in a neutrino experiment was for event classification in NO ν A, an accelerator neutrino experiment based in Fermilab [71]. CNNs have since been demonstrated to provide comparable or improved results to classical algorithms in neutrino telescope reconstruction tasks, both in IceCube [72] and in KM3NeT/ORCA [73], while also offering faster computation times.

The disadvantages of applying CNNs to Particle Physics data, discussed in Sec. 4.2, especially the information loss associated with the transformation of detector data into images, lead to alternative approaches, namely GNNs, to also be pursued by the community. Inspired by the paper by Wang *et al.* [60], which introduced the concept of edge convolutions on point clouds (see Sec. 4.2.2), Qu and Gouskos [74] put forward a GNN architecture named *ParticleNet*, in order to tackle jet tagging, a type of classification task in particle colliders such as the LHC. Similar GNN models have since made their way to neutrino telescopes. IceCube has introduced the DynEdge model, presented in Sec. 4.2.3, and applied it to the classification and reconstruction tasks of low-energy simulated neutrino events in their detector, achieving improved results over their classical algorithms [61]. A preliminary study on the energy reconstruction of high-energy events has also been discussed in [75].

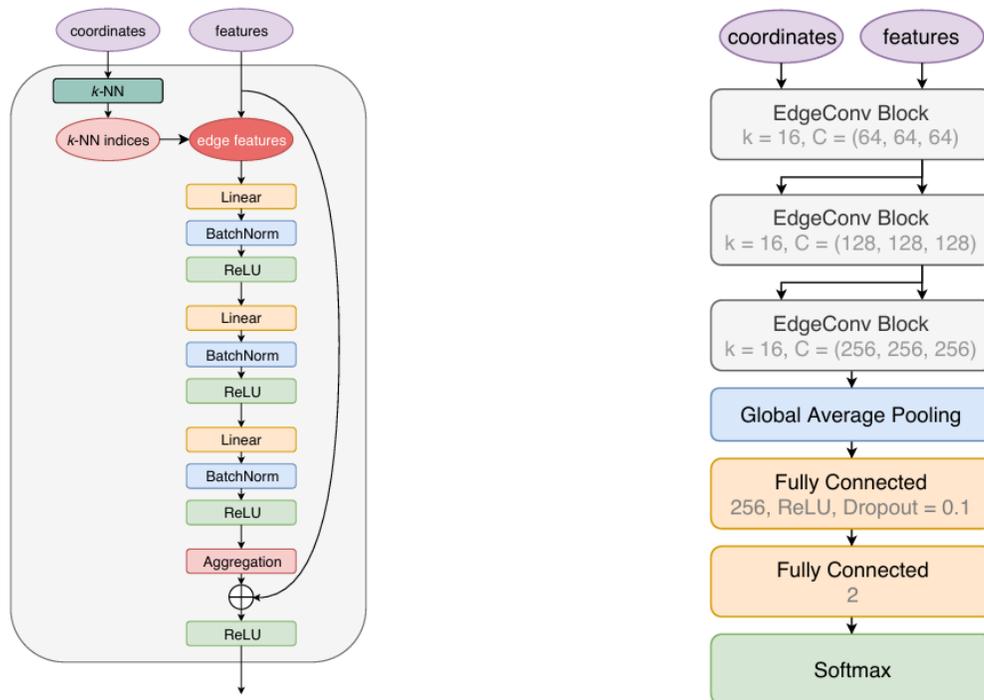


Figure 4.4: Left: Structure of the EdgeConv block used in ParticleNet. Right: Architecture of the ParticleNet network as presented in [74].

KM3NeT has also dedicated efforts to the application of GNNs to event reconstruction and classification. To this end, the ParticleNet architecture from [74] has been implemented in OrcaNet, with minimal modifications. This architecture is presented in Fig. 4.4. Similar to DynEdge, the fundamental block of ParticleNet is an EdgeConv block which performs the edge convolution over the k -NN with a 3-layer MLP. The ParticleNet architecture consists of 3 consecutive EdgeConv blocks. Each such block increases the number of features per node from 64 to 128 to 256. Unlike DynEdge, where the outputs of all EdgeConv blocks are concatenated before pooling (see Fig. 4.3), ParticleNet instead passes the input node features of each block through to its output via a shortcut connection. This is achieved using a single fully connected layer in order to bring the input node features to the dimension of the output node features, so that the two can be added. After the EdgeConv blocks, a pooling operation is performed, which reduces the dimensionality of the output from $[n, 256]$ to $[1, 256]$. ParticleNet only has a single pooling option available, unlike the 4 different operations which can be chosen in parallel in DynEdge. This option is “global average pooling”, which is functionally identical to “mean” pooling in DynEdge, *i.e.* it averages the node features across all nodes. The output vector is then propagated through a single hidden layer of 256 neurons, which feeds the output layer where the network predictions are produced. In the OrcaNet implementation of ParticleNet, just as in DynEdge, the number of outputs and the activation function of the output layer are chosen depending on the task at hand.

Using the network architecture described above, a number of reconstruction studies have recently been conducted in KM3NeT. In his PhD thesis [59], S. Reck used this model for the reconstruction of muon bundles (multiples of atmospheric muons entering the detector at the same time) in the ORCA4 detector (ORCA with 4 DUs). In this work, the muon direction, bundle diameter and muon multiplicity (number of muons in the bundle) were reconstructed. The muon direction reconstruction was shown to be competitive with classical algorithms, while the bundle diameter and muon multiplicity were reconstructed for the first time in KM3NeT [76, 77]. In another PhD thesis [34], D. Guderian performed extensive reconstruction (energy, direction) and classification (signal/background, track/shower) studies for the ORCA6 detector. The results of this work were promising, with both the zenith angle and energy reconstruction showing similar or improved performance when compared to the classical reconstruction.

In KM3NeT/ARCA, similar studies have been performed for ARCA6-8 [78]. The GNNs achieved lower errors than the standard reconstruction algorithm in both neutrino energy and zenith angle reconstruction and for both tracks and showers. Results on signal/background classification, as well as direction and multiplicity reconstruction of muon bundles were also presented.

In light of these works, it is evident that GNNs have shown promise in event reconstruction tasks in smaller versions of the KM3NeT detectors. As the detectors have since been extended, it is therefore of interest to apply GNNs to larger detector versions. This thesis achieves this by considering energy reconstruction in ARCA21, with the added benefit of further testing GNN performance on high-energy events, since most of the works above only considered low-energy events in ORCA. In terms of reconstruction results, our model achieves substantial improvement of the energy reconstruction of ARCA21 events below 10 TeV, compared to previous reconstruction efforts. Our work also provides better appreciation of the physical constraints which limit the reconstruction in higher energies. Further, the first application of DynEdge, an independently developed model, to KM3NeT data has the potential to offer a valuable comparison to the GNN model architecture already in use by the experiment, as well as highlight improvement opportunities. Lastly, the integration of KM3NeT data in the GraphNeT framework means that a well-developed alternative to OrcaNet is rendered available for use in the KM3NeT Collaboration.

Chapter 5

Energy Reconstruction with Graph Neural Networks

5.1	The KM3NeT DL Data Format	78
5.1.1	Integration of the KM3NeT DL data format in GraphNeT	78
5.2	GNN Trainings with GraphNeT in ARCA21	79
5.2.1	Dataset and data preprocessing	79
5.2.2	Model implementation	80
5.2.3	Training details & parameters	83
5.2.4	Training configurations	84
5.3	Training Results and Comparisons	85
5.3.1	Snapshot hits	85
5.3.2	Triggered hits	91
5.3.3	Event weighting	92
5.4	Investigation of Model Predictions	94

Following the discussion of Graph Neural Networks in Chapter 4, their application to the energy reconstruction of ARCA21 track-like events is now presented. In this effort, the DynEdge model, implemented within the GraphNeT framework, is used. Therefore, before proceeding to the examination of the GNN performance, the KM3NeT data format and its integration in GraphNeT are outlined. Afterwards, the training configurations that were used are presented, followed by a discussion of the training results, including comparisons to the standard reconstruction algorithm and trainings performed with OrcaNet. The Chapter concludes with an investigation on the physical justification of the model predictions, with a focus on higher energies.

5.1 The KM3NeT DL Data Format

In order to carry out supervised DL tasks, as are event reconstruction tasks with GNNs, a *labeled* dataset is needed. In other words, each event in our dataset must come with a corresponding truth value for its energy, direction, *etc.*, which will serve as the target value during the training. It is, therefore, obvious that real detector data, for which these truth values are unknown, cannot be used to train our DL models. Instead, events generated in *Monte Carlo simulations* will be used. Monte Carlo (MC) simulations are a critical component of high-energy physics experiments, as they simulate the physical processes of interest and model the detector response. In this way, they can inform physicists about the expected characteristics of physics and background signals in the detector, assist in the assessment of data analysis algorithms, help quantify systematic uncertainties, as well as validate experimental data.

MC simulations in KM3NeT are produced via the so-called simulation chain. Initially, neutrinos are generated in an area around the detector and then propagated in the seawater. If a neutrino interaction occurs, the secondary particles are propagated and the Cherenkov light emission, propagation and detection by the PMTs is also simulated. Data collection in the actual KM3NeT detectors is split into intervals of a few hours called “runs”. In each run, the detector conditions are monitored and later used in the MC production, so that effects like noise in the detector can be modeled on a run-by-run basis. After both the signal and the noise contribution have been simulated, the triggering algorithms (see Sec. 3.3) are applied and the events passing the trigger selection are stored. Finally, the reconstruction chain, presented in Sec. 3.5, is applied to the MC events, just as it is in the experimental data. The result of the MC production is a set of files, each typically containing generated events for one detector run. Along with the hit information, each event also contains the true energy, direction, interaction vertex position, *etc.* values with which it was generated, as well as the corresponding reconstructed quantities from the standard reconstruction algorithms. In a DL reconstruction task, the former will serve as the necessary truth labels, while the latter can be used to compare standard and DL reconstruction methods.

KM3NeT data (both experimental and MC) are generally stored in *ROOT* files, as mentioned in Sec. 4.2.4. The ROOT file format is a file format used within the ROOT data analysis framework, developed by CERN. ROOT is extensively used in high-energy physics for managing and analyzing the large amounts of data produced by experiments. ROOT files are an efficient way to store large amounts of data. However, KM3NeT ROOT files have not been tailored for DL applications. For this reason, KM3NeT data intended for DL applications are usually preprocessed using OrcaSong, which converts them to the *HDF5* file format. The KM3NeT HDF5 files used in DL have minimal size compared to the original ROOT files, while also offering fast random access to the hierarchical data structures within. These advantages, together with the fact that the required HDF5 files were readily available, meant that this was the file format that was chosen for the trainings in GraphNeT.

5.1.1 Integration of the KM3NeT DL data format in GraphNeT

In order to utilize the GraphNeT framework, it is necessary to integrate the KM3NeT DL data format in GraphNeT, so that KM3NeT data can serve as input to the model. GraphNeT offers support for only two file formats: Parquet & SQLite. Using GraphNeT on KM3NeT data, therefore, requires either converting HDF5 files to one of these two formats or extending GraphNeT to be able to handle HDF5 files directly. Due to the fast random access and the storage efficiency of HDF5 files, it was decided that the latter option would be selected. Internally, GraphNeT builds dataset objects in the following way: it contains an abstract `Dataset` class, which inherits from the corresponding PyTorch `Dataset` class, so that when a file is read, a `Dataset` object is created. In order to accommodate the two supported file formats, this abstract `Dataset` class serves

as the base class for two file-type-specific classes, which contain methods that handle table queries for each of the two supported file formats. Integrating HDF5 files into GraphNeT requires creating a third such class, built to handle table queries for HDF5 files. Then, HDF5 files can serve as input to GraphNeT simply by invoking instances of this class. A working demonstration of such a class had already been made by members of the KM3NeT Collaboration and was able to serve as a starting point.

At this point it is important to note that the OrcaSong module uses a two-step process to convert ROOT files to HDF5 files. The first step is a conversion from ROOT to HDF5, with the resulting files hence referred to in this text simply as “KM3NeT HDF5 files”. Then, a second conversion creates new HDF5 files, referred to within the Collaboration as “DL HDF5 files”, which contain a structure that is better suited for DL applications. In both cases, the PMT hits for all the events in the file are stored together in a single table x . Further, truth labels for all events are stored in a second table y . In KM3NeT HDF5 files, an “id” column is present in the hit table x , containing the event number in which each hit belongs. All of the hits for each event are stored in the table consecutively. Therefore, in order to access the entirety of hits in a certain event, as is often required, one must go through the table of hits x to locate the first and last hit with “id” number equal to the event number in question. This process is of course inefficient. In contrast, in DL HDF5 files, the table of hits x is accompanied by an auxiliary table which contains the location of the first hit for each event in table x , as well as the number of hits in the event, making random access in DL HDF5 files much faster than in KM3NeT HDF5 files.

The demonstration class which served as the starting point for HDF5 file integration in GraphNeT had been built to work on KM3NeT HDF5 files. Since trainings with GNNs are computationally expensive, requiring many hours per epoch, effort was devoted to modify this class to work on DL HDF5 files, in order to make the training process less time-consuming. *Ultimately, switching from KM3NeT HDF5 to DL HDF5 files led to a reduction in training time of over 50%.*

5.2 GNN Trainings with GraphNeT in ARCA21

This section contains a discussion on the technical details of the GNN models and the training configurations employed in this work.

5.2.1 Dataset and data preprocessing

As mentioned previously, the reconstruction task that is tackled in this thesis concerns track-like events, *i.e.* muon (anti)neutrino events. The dataset which will be used consists of MC data from 90 detector runs. Each run contains on average about 4,500 events of each type ($\nu_\mu / \bar{\nu}_\mu$), so that the total dataset comes up to roughly 800,000 events, equally split between muon neutrinos ν_μ and muon antineutrinos $\bar{\nu}_\mu$. Of these 90 runs, 75 will be used for training and 15 for validation ($\sim 83\% / 17\%$ split). An additional set of 19 runs has been kept aside for inference, *i.e.* as the test set for evaluation of the model predictions. The neutrino energy of the MC events in our ARCA21 dataset is in the range of $[10^2, 10^8]$ GeV. Fig. 5.1 shows the distribution of events across this energy range in the training, validation and test datasets, confirming that all datasets follow the same curve. This energy distribution is a convolution of an assumed neutrino energy spectrum and the detector response: we observe fewer events as we move to higher energies due to the decrease in the neutrino energy spectrum and the small instrumented volume of ARCA21, and also few events at low energies, since the detector is less sensitive in the sub-TeV and low-TeV region.

Each ARCA21 event contains on average roughly 1,000 hits. There exist, however, outlier events with a much larger number of hits than others. Since identifying the k -nn for an entire graph is a $\mathcal{O}(n^2)$ operation, the maximum number of hits that an event can contain in the DL HDF5 files has been set to 5,000 hits. If an event initially contains more than 5,000 hits, then all of the triggered hits are kept, while some of the non-triggered snapshot hits are omitted during conversion to DL HDF5 files with OrcaSong, so that only 5,000 hits remain in total.

Each PMT hit in the data contains the same 8 hit features: position- x , $-y$, $-z$, direction- x , $-y$, $-z$, time t and ToT. Naturally, the numerical values that the hit features can take vary greatly both across the different features and across the dataset. Using features with large and differing values during the training is problematic, as it can create a bias towards features with larger numerical values, which can dominate the

learning process and in essence cause the network to ignore other features. For this reason, it is important to implement a feature normalization step, ensuring that the input feature values are within an acceptable range. In this step, each feature x is transformed by subtracting its mean μ_x and dividing by its standard deviation σ_x across the dataset:

$$x \rightarrow \frac{x - \mu_x}{\sigma_x}. \quad (5.1)$$

The three direction coordinates are not transformed at all, since they are defined to form a unit vector for each hit and thus their values are already centered around 0 and fall within the acceptable range of $[-1, 1]$.

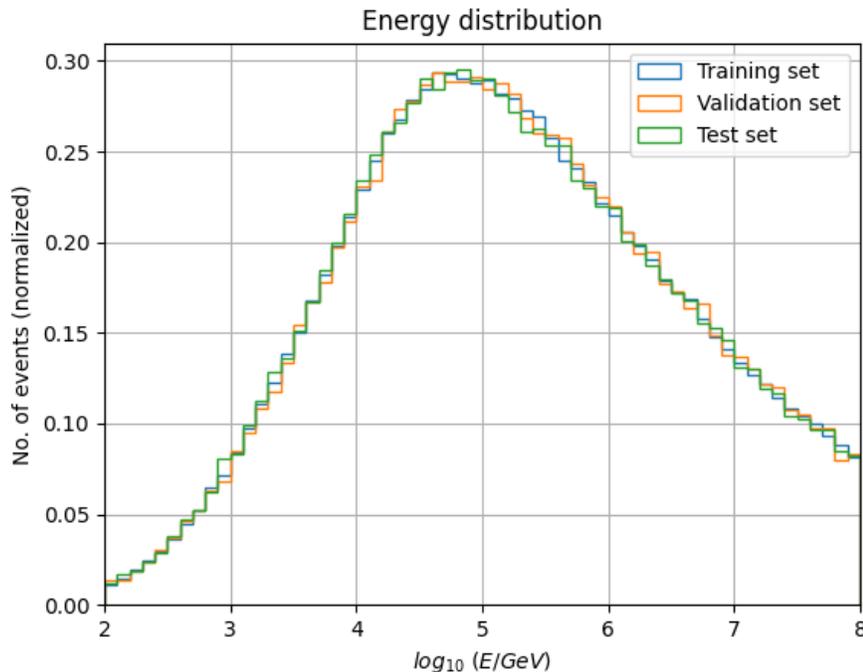


Figure 5.1: Neutrino energy distribution in the training, validation and test datasets, normalized so that the area under each curve is equal to 1. The x -axis is the energy exponent, with each bin covering a 0.1 range. All datasets exhibit the same energy distribution.

5.2.2 Model implementation

The trainings that are documented in this work utilize the DynEdge model, presented in Sec. 4.2. Various hyperparameter choices and implementation details are presented in this Section.

Number of nearest neighbors k

The number of nearest neighbors k is an important hyperparameter of the model. A large k can increase the performance of the model by allowing it to capture relationships between more distant nodes of the graph. However, it also substantially increases the computation time, since it increases the number of operations in the convolutional layers. For our trainings in GraphNeT, a value of $k = 16$ was chosen for all 4 EdgeConv blocks of the model. This is double the value that has been used by IceCube [61], which only has a single PMT per DOM and thus less complex data structure, and equal to the value used in previous ARCA studies [78]. Larger values of k were not considered due to the already high computation time of the trainings and the fact that ARCA21 trainings which had been performed in OrcaNet with $k = 32$ did not yield improved results.

Pooling schemes

Unlike the ParticleNet architecture, which only employs average pooling, DynEdge, as mentioned in Sec. 4.2.3, has 4 pooling options available (“mean”, “min”, “max” & “sum”), which can be used in parallel. Using multiple pooling options is advantageous, since it renders more information available to the model to then make the predictions. For this reason, all 4 options were used in the initial trainings. For later trainings, however, the “sum” pooling was removed, leaving only 3 pooling operations. “Sum” pooling will perform addition over the nodes for each feature, meaning that, when the number of nodes in the graph is large, this option can lead to large numerical values in the corresponding output subarray, which can then dominate over the other pooling options. This is exactly the case in ARCA21, where each event has roughly 1,000 snapshot hits, leading to graphs with about 1,000 nodes on average.

Output layer configuration

The output layer of the model needs to be configured according to the task at hand, that is the neutrino energy reconstruction, *i.e.* a regression task, in which the model is called to predict a single value, the energy of the original particle. In the simplest case, the output layer consists of a single neuron, the output of which is the predicted energy value. The activation function of this neuron needs to prevent non-physical energy values, *i.e.* restrict the output domain to positive values. For this reason, a variant of the ReLU function (Eq. 4.3), the *Softplus* function is used:

$$f_{\text{Softplus}}(x) = \frac{1}{\beta} \ln(1 + e^{\beta x}), \quad (5.2)$$

with $\beta = 0.05$. This function is a smooth approximation of the ReLU function, avoiding the non-differentiability of the latter at $x = 0$ and constraining the output to strictly positive values (see Fig. 5.2).

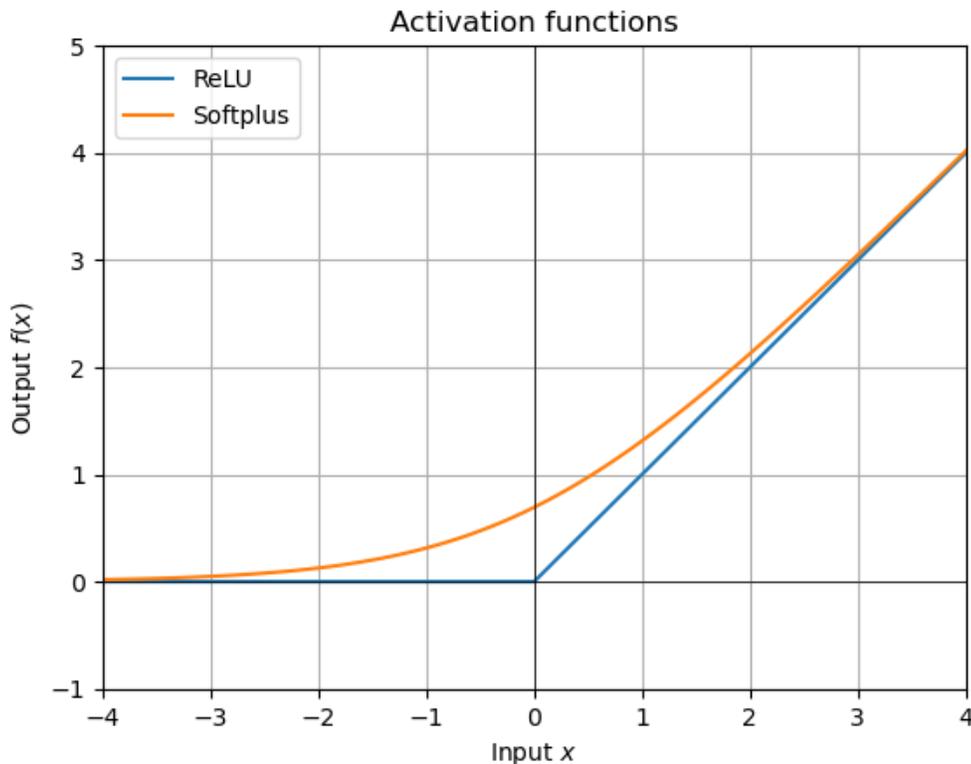


Figure 5.2: Comparison of the ReLU and Softplus activation functions.

Loss functions

In order to carry out the trainings, a sensible loss function must be chosen. The simplest loss function that was applied in this work is the MSE or L_2 loss:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (5.3)$$

where y_i is the true label of the i -th sample, \hat{y}_i is the network prediction and N is the number of batch samples, *i.e.* the batch size. The squaring operation in the MSE loss amplifies the errors, which can offer faster convergence. However, at the same time, it entails that outliers in the data have a significantly larger impact on the loss, which in some cases could affect the model disproportionately and compromise its performance and generalization ability. For this reason, the *Log Cosh* loss, used in [61], was instead extensively used in the trainings that were performed:

$$L_{Log\ Cosh} = \frac{1}{N} \sum_{i=1}^N \ln [\cosh(\hat{y}_i - y_i)]. \quad (5.4)$$

The Log Cosh loss is similar to the L_2 loss for small errors, but is akin to the L_1 loss for larger error values. This behavior combines the benefits of the L_1 and L_2 losses. In the L_1 loss, the gradient is constant and not proportional to the error, like it is in the L_2 loss. This can cause the model to minimize small errors less effectively, thus converging less smoothly. The Log Cosh loss alleviates this because the gradient is proportional to the error at small error values. Also, unlike the L_1 loss, the Log Cosh loss is differentiable at 0. At the same time, its similarity to the L_1 loss for large errors provides robustness to outliers. The behavior of the two loss functions can be seen in Fig. 5.3.

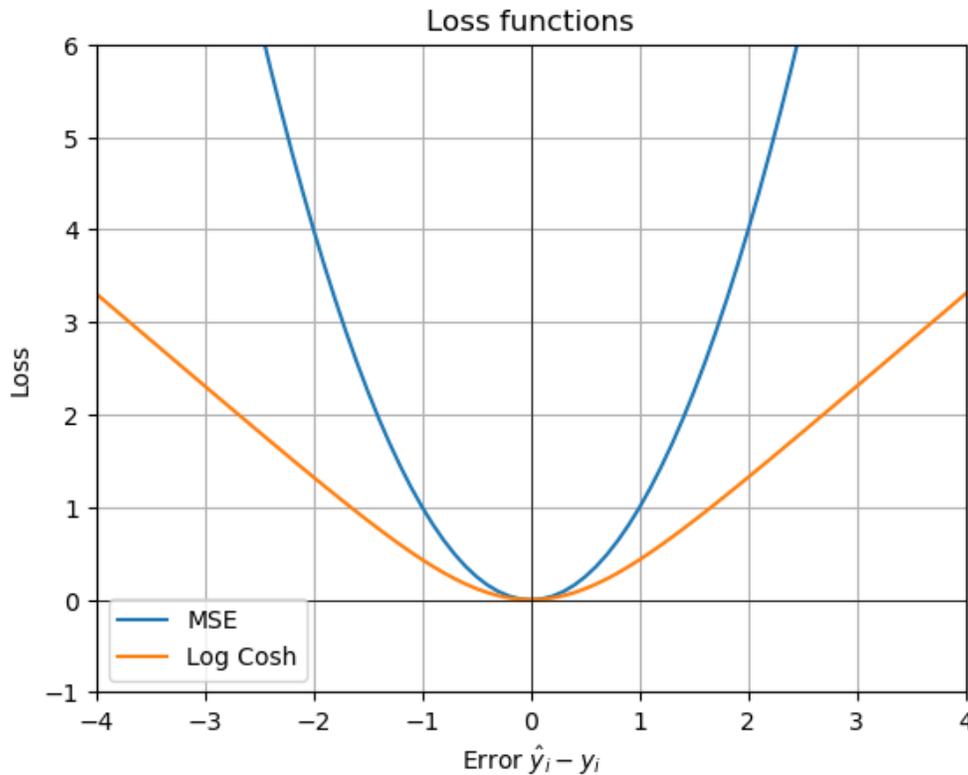


Figure 5.3: Comparison of the MSE and Log Cosh loss functions.

The third loss function that will be examined in this work is the *Log Normal* loss function:

$$L_{Log\ Normal} = \frac{1}{N} \sum_{i=1}^N -\ln \left[\frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} e^{-\frac{(y_i - \hat{\mu}_i)^2}{2\hat{\sigma}_i^2}} \right] = \frac{1}{N} \sum_{i=1}^N \left[\ln(\sqrt{2\pi}) + \ln(\hat{\sigma}_i) + \frac{(y_i - \hat{\mu}_i)^2}{2\hat{\sigma}_i^2} \right]. \quad (5.5)$$

This function is the natural logarithm of a normal distribution. In this case, both $\hat{\mu}_i$ and $\hat{\sigma}_i$ are fit parameters, with $\hat{y}_i \equiv \hat{\mu}_i$ being the network estimate for the energy and $\hat{\sigma}_i$ being an uncertainty estimate on the energy prediction. The loss function is minimized at $(\hat{\mu}_i, \hat{\sigma}_i) = (y_i, 0)$ and, as can be seen from Eq. 5.5, it is similar to the MSE loss in that it is quadratic with respect to the error. The uncertainty estimate is realized with the addition of a second neuron in the output layer of the network. It serves as a measure of the reconstruction quality of each event and is thus a helpful quantity since it can be utilized as an event selection criterion in data analysis. The Log Normal loss function is the loss function of choice in OrcaNet trainings and was thus also implemented in GraphNeT to be used in this work.

Finally, as it has already been mentioned, the neutrino energy range in our MC data for ARCA21 is $[10^2, 10^8]$ GeV, an extreme range spanning 6 orders of magnitude. For this reason the energy values that go into the network loss function –whichever that is– are always log-transformed ($E \rightarrow \log(E)$), such that both the truth and the predicted value are the logarithm of the true/predicted energy respectively. This results in a range of target values of $[2, 8]$, leading to increased numerical stability and smoother convergence.

5.2.3 Training details & parameters

All of the model configurations were trained on a local server at the Institute of Nuclear and Particle Physics (INPP) of NCSR “Demokritos”. This server was equipped with two 48 GB NVIDIA RTX A6000 GPUs, 125 GB of RAM and 48 CPU cores. On this machine, each training epoch on the 800,000 event dataset needed 250 minutes, using the training parameters described below.

In each training, the training and validation datasets are used as input and the corresponding training and validation dataloaders are created. The training dataloader needs to be shuffled after every epoch, in order to avoid order bias, where the network starts learning patterns of the order of the data. For the initial trainings, a batch size of 32 was used. Due to the small number of events per batch, such small batch sizes provide less good of an approximation of the overall gradient of the dataset, *i.e.* the gradient is more noisy. This can lead to a model that converges slower, but generalizes better, since the noisy gradients can help kick the model out of local minima in its phase space. After discussion with GraphNeT developers, however, it was suggested that a large batch size be used, since it can better approximates the gradient of the dataset. The batch size was subsequently increased to 1,024, without observing any loss in generalization ability. Such an increase would normally come with a substantial improvement in training speed. However, in this case, due to GPU memory limitations, the actual batch size could not be further increased. This large batch size was therefore set using gradient accumulation, a technique in which the actual batch remains small, but a larger effective batch size is achieved by storing gradients and performing a gradient update once every few batches.

Another parameter which can be exploited in the training configuration is the so-called “number of workers”. This PyTorch parameter defines the number of CPU subprocesses which load the data to the GPU(s) in parallel. Increasing the number of workers can greatly reduce the data loading time and therefore lead to much faster training. However, above a certain number of workers the limitations of running a large number of parallel CPU processes, not least of which is the increasing RAM consumption, outweigh the benefits. In our case, the number of workers was set equal to the number of CPU cores, which was 48. Larger values were tested, however they could not be supported by the server due to memory limitations.

For all the trainings in this work, the Adam optimizer (Sec. 4.1.2) was used, with its default parameters $(\beta_1, \beta_2, \epsilon) = (0.9, 0.999, 10^{-8})$. The learning rate was initially set at 10^{-4} . However, this learning rate value would lead to the phenomenon of *exploding gradients*, where gradients become excessively large, causing unstable updates to the model weights and preventing the network from converging. For this reason, gradient clipping was employed in initial trainings with a value of 0.5. In later trainings, a piecewise linear learning rate with a warm-up period was used. This learning rate scheduler used a warm-up period starting from 10^{-5} and reaching the maximum of 10^{-3} by the middle of the first epoch, before gradually reducing the learning rate down to 10^{-5} again until the end of the training, as described also in [61].

5.2.4 Training configurations

Before the results of the GNN trainings can be discussed, the exact training configurations that were used must first be presented. The trainings that were performed are outlined below. All trainings used the *true neutrino energy* (*i.e.* the MC neutrino energy) as a true label, that is the energy of the primary neutrino particle that caused the interaction. Trainings can be split in two groups, according to the set of hits that were used during training, *i.e.* snapshot or triggered (see Sec. 3.3).

Snapshot hits

For most of the trainings in this work, all of the snapshot hits in each event (both triggered & non-triggered) were provided to the model. A summary of the snapshot hits trainings is presented in Table 5.1. The first two trainings were conducted on a reduced version of the full dataset, containing only 200,000 muon antineutrino events. In these trainings, the MSE and Log Cosh loss functions were tested. Training no. 3 repeated the previous training with the full dataset. Up until this point the ToT was not part of the hit features provided to the model, since the light intensity of each hit is thought to be inferable from the hit density at neighboring PMTs, which is already available to the model. However, since the ToT feature was available, it was added in subsequent trainings, with training no. 4 being a repeat of the previous one, but with ToT added, so that results could be compared. For the remaining trainings, the batch size was increased from 32 to 1,024, the “sum” pooling option was removed and the piecewise linear learning rate from [61] was used, removing the need for gradient clipping. Training no. 5 makes use of the Log Cosh loss, allowing direct comparison with previous trainings. Finally, training no. 6 uses the Log Normal loss, allowing the comparison of the different loss functions.

No.	Events	Epochs	Loss	Batch size	Learning rate	Notes
1	200,000	20	MSE	32	10^{-4}	Without ToT, all pooling options, gradient clipping
2	200,000	20	Log Cosh	32	10^{-4}	Without ToT, all pooling options, gradient clipping
3	800,000	10	Log Cosh	32	10^{-4}	Without ToT, all pooling options, gradient clipping
4	800,000	10	Log Cosh	32	10^{-4}	With ToT, all pooling options, gradient clipping
5	800,000	30	Log Cosh	1,024	Piecewise linear $10^{-5} \rightarrow 10^{-3} \rightarrow 10^{-5}$	With ToT, no “sum” pooling
6	800,000	30	Log Normal	1,024	Piecewise linear $10^{-5} \rightarrow 10^{-3} \rightarrow 10^{-5}$	With ToT, no “sum” pooling

Table 5.1: Training configurations (snapshot hits).

Triggered hits

Apart from trainings with all snapshot hits, trainings with only triggered hits were also performed. Since the triggered hits are hits which have been flagged by the trigger system as coincident, optical background hits are expected to have been removed from the events. Training the model only on the triggered hits could therefore improve the energy reconstruction. The triggered hit trainings are summarized in Table 5.2. The trainings differ on the learning rate. Training no. 1 used the default Adam learning rate of 10^{-3} together with gradient clipping to avoid exploding gradients. Trainings no. 2 & 3 used the piecewise linear learning rate, with different intermediate and final learning rate values and a different epoch budget. Both trainings start from a learning rate of 10^{-5} , increasing until the middle of the first epoch. In training no. 2, the default maximum of 10^{-3} was used, with the learning rate reducing to 10^{-5} by the end of the 30th epoch. In training no. 3, the maximum was set to $2 \cdot 10^{-3}$ and the learning rate was reduced to 10^{-3} over 20 epochs.

No.	Events	Epochs	Loss	Batch size	Learning rate	Notes
1	800,000	15	Log Cosh	1,024	10^{-3}	With ToT, no “sum” pooling, gradient clipping
2	800,000	30	Log Cosh	1,024	Piecewise linear $10^{-5} \rightarrow 10^{-3} \rightarrow 10^{-5}$	With ToT, no “sum” pooling
3	800,000	20	Log Cosh	1,024	Piecewise linear $10^{-5} \rightarrow 2 \cdot 10^{-3} \rightarrow 10^{-3}$	With ToT, no “sum” pooling

Table 5.2: Training configurations (triggered hits).

5.3 Training Results and Comparisons

In this Section, the results of the GNN trainings that were performed are presented and discussed. Apart from the first two snapshot hit models, which were trained and evaluated on a smaller dataset, all of the results and plots in this Section are based on the test set, which contains 172,484 events.

5.3.1 Snapshot hits

The results of the snapshot hit trainings of Table 5.1 are documented in Table 5.3. For each training, the MAE, MSE and the R^2 score⁸ are presented, with trainings no. 5 & 6 sharing the best metrics. Further, the results of each training are visualized in the scatter plots and quantile plots of the predicted with respect to true energy in Fig. 5.4 and Fig. 5.5 respectively. The results are discussed in detail below.

No.	MAE	MSE	R^2
1	0.644	0.726	0.565
2	0.648	0.734	0.560
3	0.590	0.611	0.637
4	0.585	0.620	0.632
5	0.573	0.603	0.642
6	0.580	0.596	0.646
std reco	1.016	1.770	-0.133
OrcaNet	0.577	0.594	0.648

Table 5.3: Training results (snapshot hits). Bold values indicate the best GraphNeT result for each metric. The results for the standard reconstruction (“std reco”) and the best OrcaNet model in the same dataset are also shown for comparison.

⁸The R^2 score, or *coefficient of determination*, is a statistical measure that represents the proportion of the variation in the dependent variable that is predictable from the independent variable. It is defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \quad (5.6)$$

where SS_{res} and SS_{tot} are the sum of squares of residuals and the total sum of squares, defined as:

$$SS_{res} = \sum_i (y_i - \hat{y}_i)^2, \quad (5.7)$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2,$$

where \bar{y} is the mean of the data y_i . The higher the R^2 score, the better the model explains the observed variance in the data, with the maximum possible score being 1.

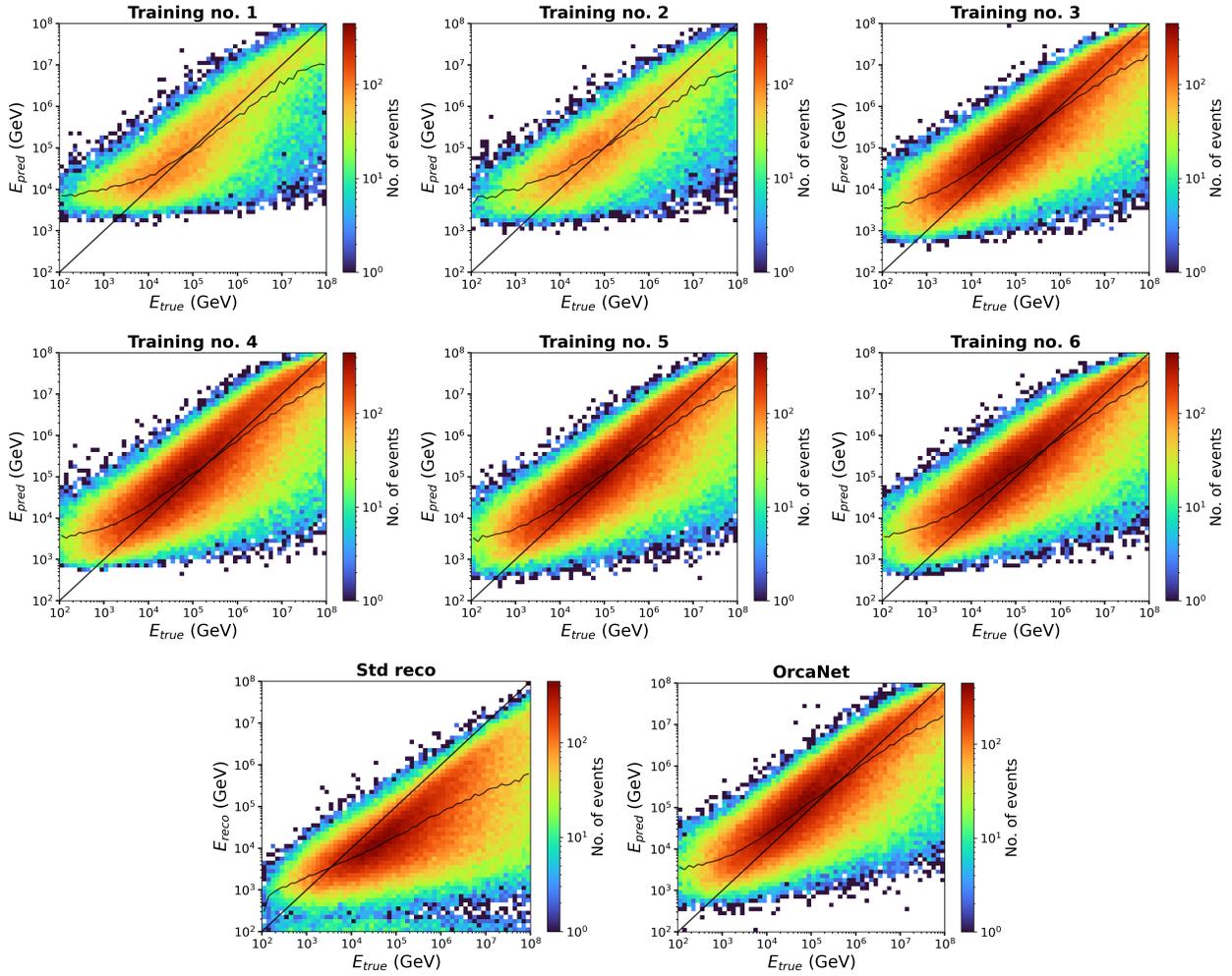


Figure 5.4: Predicted with respect to true energy scatter plots for all entries of Table 5.3. In all plots, the black diagonal line indicates the ideal reconstruction. A second black line indicates the median predicted value for each true energy bin. The events are unweighted.

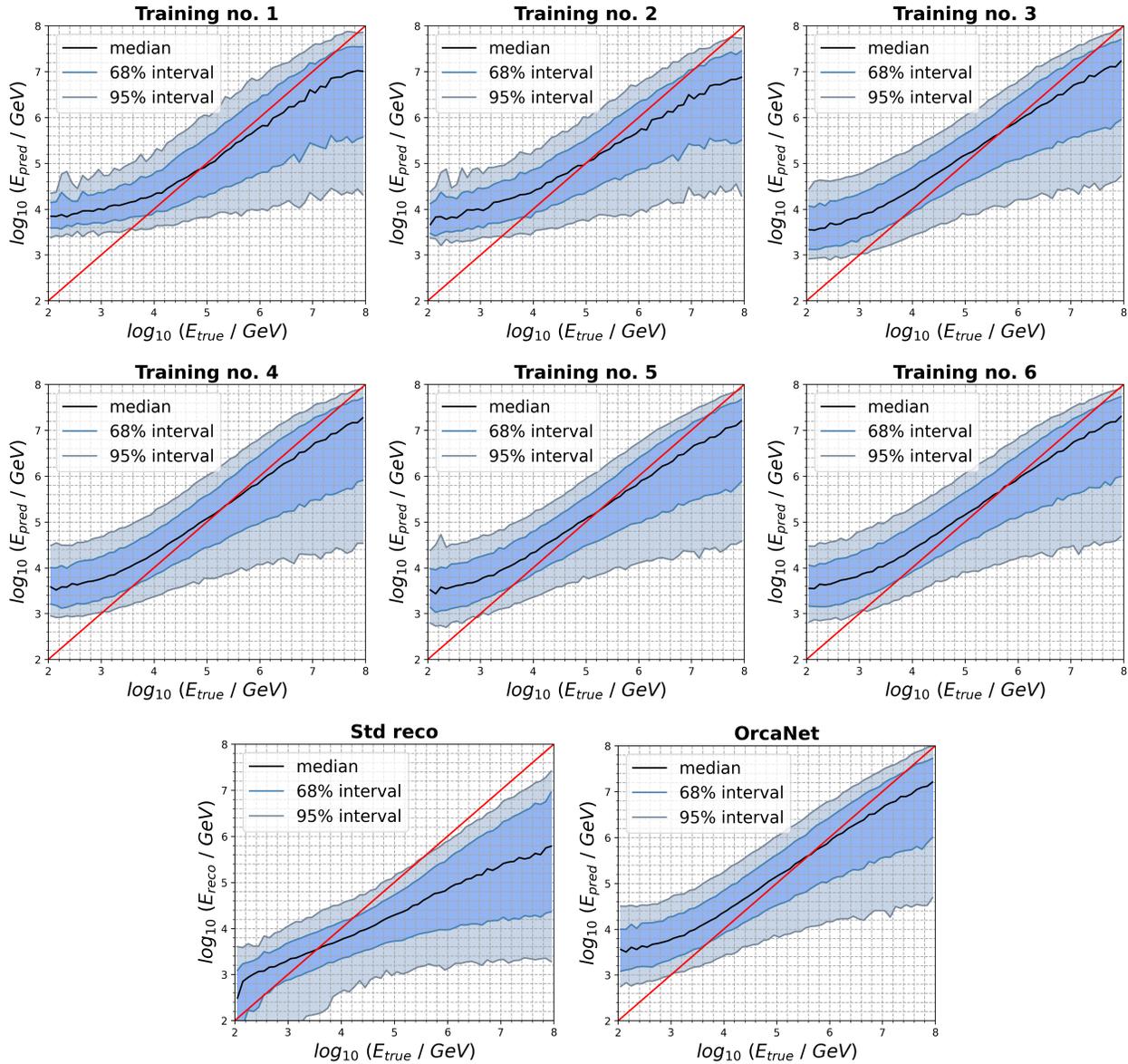


Figure 5.5: Quantile plots for all entries of Table 5.3. In all plots, the red diagonal line indicates the ideal reconstruction. The median and the 68% & 95% confidence intervals of the predictions for each true energy bin are shown in black, blue and light gray respectively.

As can be seen by the plots in Figs. 5.4, 5.5, there is consistency in the prediction behavior between the trainings. The networks generally produce the most accurate predictions in the middle of the energy range, that is for E_{true} ranging from 10^3 to 10^6 GeV (*i.e.* in the TeV scale). In contrast, it is clear that lower and higher energies tend to be over- and underestimated respectively. This effect is more pronounced for lower energies, whose representation in the dataset is minimal (see Fig. 5.1). For the highest energies (*i.e.* PeV scale, $E_{true} > 10^6$ GeV), an underestimation is expected, since high-energy muons do not generally deposit their entire energy content in the detector. In other words, not all of the muon energy is visible to the detector. This fact is also reflected in the large range of the network predictions for these events. We can further observe that the best OrcaNet model, trained and evaluated on the same datasets, exhibits the same behavior as the GraphNeT-based networks. Finally, the standard reconstruction refers to the reconstruction algorithm described in Sec. 3.5. It should be noted that the standard reconstruction algorithm reconstructs the *muon energy*. Therefore, the lower energy predictions compared to the GNNs are to be expected, since the visible energy of the muon is lower than the energy of the primary neutrino.

Training comparisons

The metrics of Table 5.3 can allow us to draw conclusions about the best training configuration. This is further aided by the comparison plots in Fig. 5.6, in which the 68% quantiles (shown also in Fig. 5.5) are overlaid for different pairs of training configurations. In the following, the trainings are referred to by the letter “T” followed by their training number from Table 5.1 (*e.g.* “T1”). Most comparisons are drawn between trainings that differ in only one parameter, so that the effect of each configuration parameter can be evaluated separately:

- **MSE (T1) vs. Log Cosh (T2):** The first two trainings, performed on the reduced dataset of 200,000 events, offer a comparison between the MSE and Log Cosh loss functions. It can be seen that the network trained with the MSE loss has the better metrics. This is because of the greater penalty applied to large errors in the MSE loss, leading to a better minimization of large prediction error values. Looking at the comparison plot in Fig. 5.6, the performances appear similar, with the MSE band (in blue) being slightly better aligned with the diagonal than the Log Cosh band (in orange) for most of the energy range.
- **200,000 (T2) vs. 800,000 (T3) events:** As expected, utilization of the larger dataset of 800,000 events greatly improves the network performance, with the R^2 score increasing by 13.8%. This improvement is also evident in the corresponding comparison plot in Fig. 5.6, with the greater performance gains being observed in the most underrepresented energy regions of the dataset (low- & high-energy events), which stand to benefit more by an increase in dataset size.
- **Without ToT (T3) vs. with ToT (T4):** Addition of the ToT variable as a node feature of the graphs has minimal impact on the training metrics, which only vary by about 1% between the two trainings, with only the MAE improving. However, the comparison plot of Fig. 5.6 reveals a median that is slightly better aligned to the diagonal for the training which includes ToT (in orange), especially in TeV energies. For this reason, it was decided to continue using the ToT as a feature in subsequent trainings.
- **Batch size, learning rate, pooling (T4 vs. T5):** This is the sole comparison in which the compared trainings contain more than one different parameters. These configuration changes were suggested by GraphNeT developers for our use case of DynEdge in ARCA21. The changes (removal of “sum” pooling, increase of the batch size to 1,024 and use of the piecewise linear learning rate) brought about a small improvement in the regression metrics, with the R^2 score increasing by 1.6%. In fact, this training features the lowest MAE of all ARCA21 energy reconstruction models. The comparison plot shows two very similar bands for the two trainings. It is worth noting that the use of a larger batch size, *i.e.* fewer weight updates, causes the network to converge more slowly. This was somewhat mitigated by the use of the piecewise linear learning rate, however the model still required an increased epoch budget compared to previous models. Since GNN trainings are computationally expensive, this fact, coupled with the relatively small observed performance gains, leads to the recommendation that a smaller batch size be used in future configuration and hyperparameter tuning tests, in order to reduce the computation time, with a larger batch size configuration then being tested for the final model.

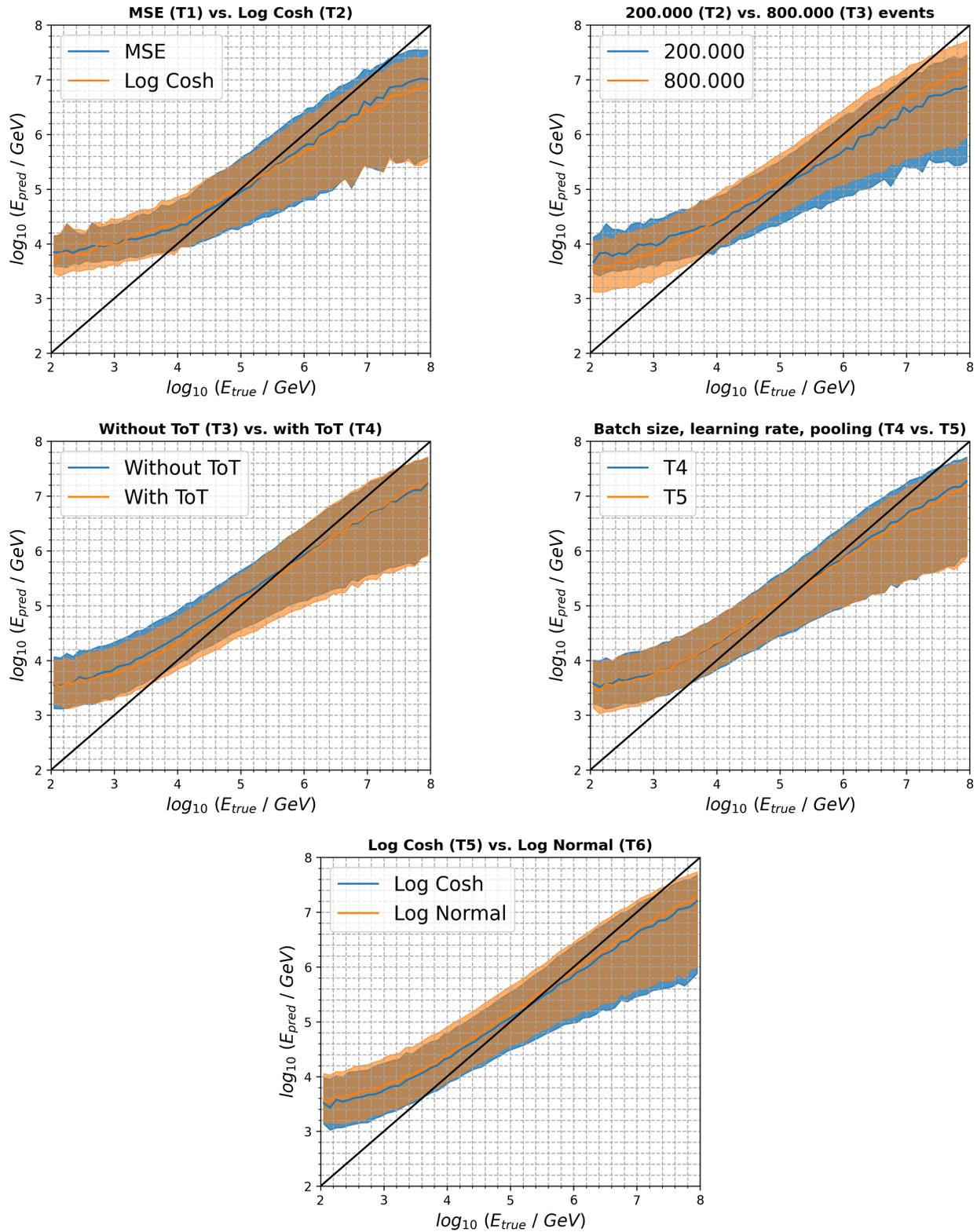


Figure 5.6: Comparisons between the various training configurations of Table 5.1. In all plots, the 68% confidence intervals for each training are compared.

- **Log Cosh (T5) vs. Log Normal (T6):** The implementation of the Log Normal loss function from OrcaNet in GraphNeT allows for a direct comparison between the loss functions of choice in the two frameworks. The Log Normal loss is a kind of L_2 loss, while the Log Cosh loss is most similar to the L_1 loss, so it is no surprise that the former outcores the latter in the MSE metric, while the opposite is true for the MAE metric. The Log Normal training achieves the best MSE and R^2 scores out of all GraphNeT trainings, scoring effectively identical scores to the best OrcaNet model. Overall, the two loss functions achieve very similar scores. Their closely matched performance is also evident in the corresponding comparison plot of Fig. 5.6, where the two bands are virtually identical, with the Log Normal network predictions (in orange) being on average slightly higher than the Log Cosh network predictions (in blue). Use of the Log Normal loss in OrcaNet is thus proven to be justified, since it offers the same performance as simpler loss functions, while incorporating an uncertainty estimate in the network predictions.

This discussion on snapshot hit trainings concludes with a comparison of GraphNeT-based trainings with the best OrcaNet-based training that was available for ARCA21 on the same dataset. Training no. 6 was chosen for this comparison, since it has the best MSE and R^2 score out of all GraphNeT trainings. It also features the same loss function as the OrcaNet model, allowing for a direct comparison of the two architectures. The two models produce nearly identical scores for all metrics in Table 5.3. The models are also compared in Fig. 5.7, along with the standard reconstruction. The standard reconstruction band lays lower than the GNN models, since it reconstructs the energy of the resulting muon and not the neutrino energy. Comparing the standard reconstruction to the GNN predictions, we can observe that the two are in better agreement in the lower parts of the energy range. This is because, on average, a greater fraction of the neutrino energy is deposited in the detector via the muon for low-energy events and so the two reconstructions are more closely matched in this region. The bands of the two GNN models are almost completely overlapping, as is suggested by their similar performance on the metrics. Consequently, it appears that both underlying architectures have similar capabilities and both frameworks can successfully support GNN trainings. This is an important result, since (a) *it validates the configuration choices in the existing OrcaNet implementation of the ParticleNet architecture and signifies that the relevant hyperparameters have been well tuned*, and (b) *it provides the Collaboration with a readily available alternative tool of excellent capabilities (GraphNeT), which has been developed to have a low entry barrier and be able to easily accommodate any desired functionality*.

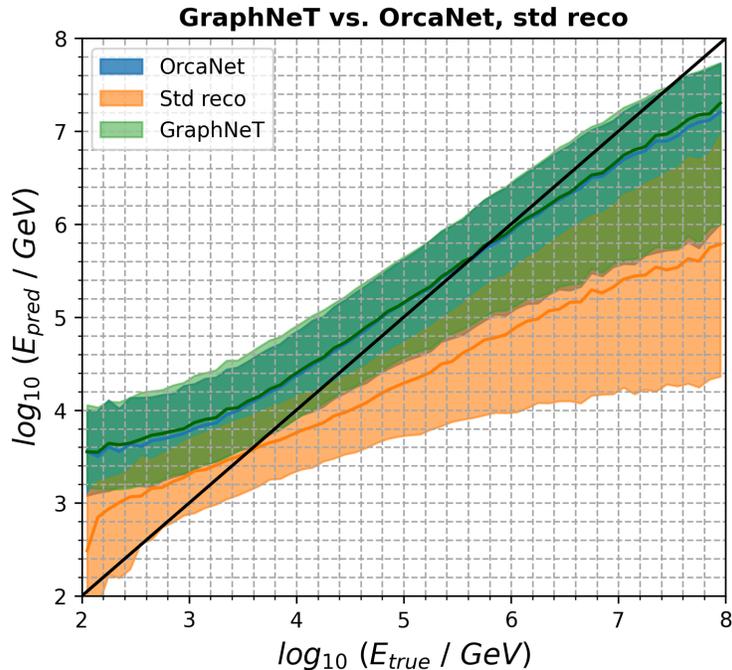


Figure 5.7: Comparison between GraphNeT training T6 and the best OrcaNet training. The standard reconstruction is also shown. The 68% confidence intervals are compared.

5.3.2 Triggered hits

The results of the triggered hit trainings of Table 5.2 are documented in Table 5.4. As in the case of the snapshot hit trainings, the MAE, MSE and the R^2 score are presented. The scatter plots and quantile plots of the predicted with respect to true energy for all three triggered hit trainings are also presented in Fig. 5.8 and Fig. 5.9 respectively.

No.	MAE	MSE	R^2
1	0.620	0.694	0.588
2	0.623	0.687	0.592
3	0.635	0.705	0.581

Table 5.4: Training results (triggered hits). Bold values indicate the best result for each metric.

Looking at the results of Table 5.4, it appears that training no. 2 is the best performing of the three trainings, according to two of the three metrics. Interestingly, this is the training with the lowest learning rate. This can be explained by the lower learning rate leading to more stable weight updates and helping the network avoid overshooting and oscillating around local minima.

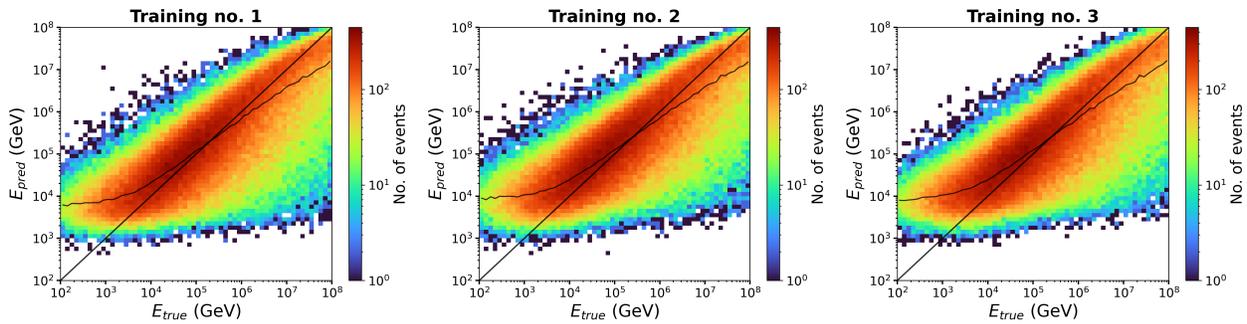


Figure 5.8: Predicted with respect to true energy scatter plots for all entries of Table 5.4. In all plots, the black diagonal line indicates the ideal reconstruction. A second black line indicates the median predicted value for each true energy bin. The events are unweighted.

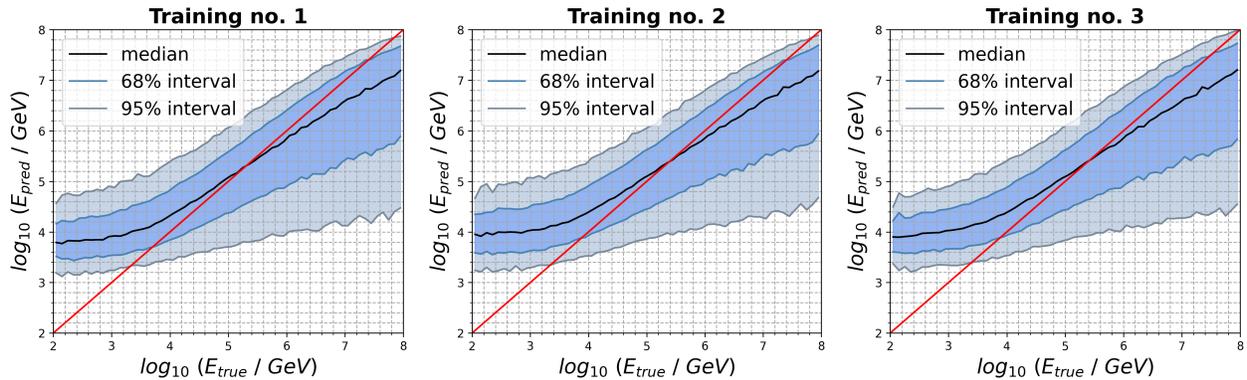


Figure 5.9: Quantile plots for all entries of Table 5.4. In all plots, the red diagonal line indicates the ideal reconstruction. The median and the 68% & 95% confidence intervals of the predictions for each true energy bin are shown in black, blue and light gray respectively.

Finally, a comparison between the snapshot and triggered hit trainings can be drawn. By comparing the training metrics in Tables 5.3 and 5.4, it can be seen that the triggered hit trainings are performing worse compared to the snapshot hit ones by a significant margin. This comparison is best showcased using snapshot hit training no. 5 with triggered hit training no. 2, since they share the same network configuration. We can see that the former has a 9% better R^2 score than the latter. The comparison plot between these two

trainings is shown in Fig. 5.10. Looking at this plot, we can observe that both trainings perform similar in the middle and upper parts of the energy range. The source of the performance variation is therefore located, in energies below 10^4 GeV, where the two bands detach, with the snapshot hits band being closer to the diagonal. Due to this performance deficit of triggered hit networks in low energies and the failure to observe improvements in other parts of the energy range, it was concluded that snapshot hits trainings are the better choice and no further triggered hit trainings were pursued.

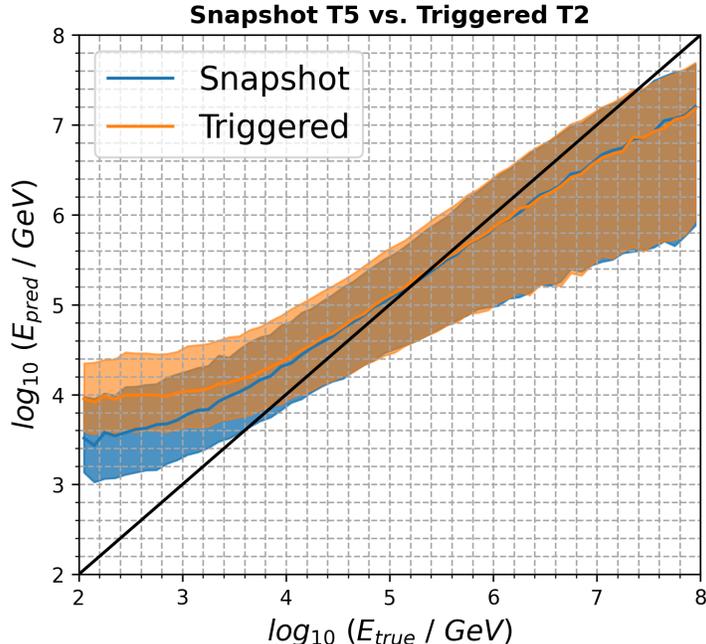


Figure 5.10: Comparison between snapshot hits training T5 and the best triggered hits training (T2). The 68% confidence intervals for each training are compared.

5.3.3 Event weighting

All trainings using snapshot and triggered hits that were discussed so far were performed using the original MC energy distribution, seen in Fig. 5.1. This means that the middle part of the energy range is overrepresented, while the outer portions –primarily the lowest energies, $E < 10^3$ GeV– are underrepresented. This uneven distribution means that the model learns the overrepresented areas of the dataset better, while predicting poorly on the underrepresented ones due to the lack of training samples. This is clearly evident in the scatter and intervals plots of Figs. 5.4, 5.5, where, for events with E_{true} below a few hundred GeV, almost no events are reconstructed on the diagonal. The imbalanced target distribution might also cause the model to become biased towards predicting values in the more densely populated regions of the energy range. This might produce lower final loss values, however it would lead to predictions that are not based on the underlying physical information alone. For these reasons, *sample* or *event weighting* was implemented into the training, in order to train the model on an effectively uniform energy distribution. In sample weighting, each event i has an associated weight w_i , which is inversely proportional to the number of events in the corresponding energy bin. During the loss calculation, these weights multiply the per-event losses, so that the loss for each batch is a weighted mean:

$$L_{batch} = \frac{1}{N} \sum_{i=1}^N w_i L_i, \quad (5.8)$$

where L_i is the loss for event i . This way, greater importance is placed on events from underrepresented energy regions, thus improving the predictive ability of the model in those areas and forcing it to generalize better across the entire energy range.

Event weighting was therefore applied to a snapshot hit training (since snapshot hit trainings outperformed the trainings using triggered hits). Apart from the application of weights to the loss function, the training configuration is otherwise identical to snapshot hit training no. 5. For simplicity, this training will from now on be referred to as *weighted*. The scatter and quantile plots for this network are presented in Fig. 5.11. The results of this training have also been overlaid with those of snapshot hits training no. 5 in Fig. 5.12.

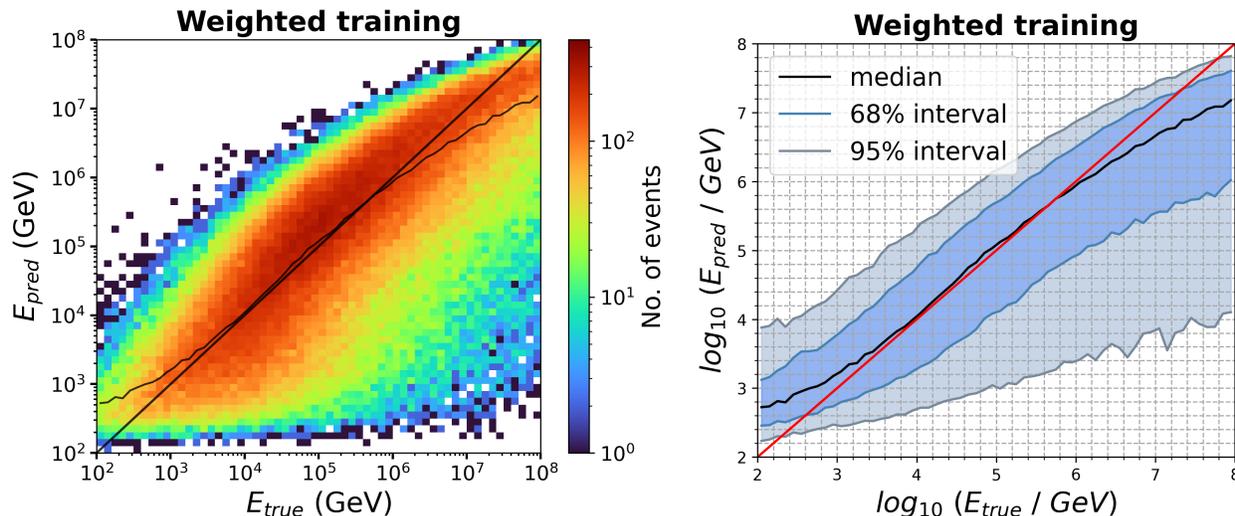


Figure 5.11: Scatter and quantile plots for the weighted training. The diagonal indicates the ideal reconstruction. The median predicted value for each true energy bin is shown in both plots in black. On the quantile plot, the 68% & 95% confidence intervals of the predictions for each true energy bin are shown in blue and light gray respectively.

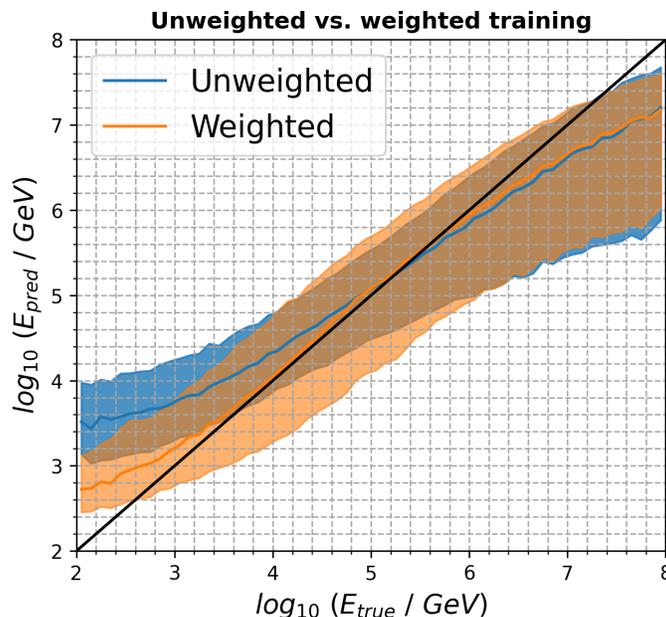


Figure 5.12: Comparisons between the weighted and unweighted training (snapshot hit T5). The 68% confidence intervals for each training are compared.

Examination of the training results, as these appear in the plots above, reveals interesting observations and allow us to reach the following conclusions:

- (a) The median for the weighted training (orange line in Fig. 5.12) is much more closely aligned to the diagonal than the unweighted training median (blue line), an observation which holds across the entire energy range. This signifies that the model tends to produce predictions around the correct value for all energies: it is more aligned with the overall energy distribution and is not biased towards the overrepresented energy regions, *i.e.* it generalizes better.
- (b) The model predictions for the lowest energies are the most improved, owing to the large importance placed on these events during training. *This represents a substantial improvement to previous models and is the only GNN model which exhibits this predictive ability at low energies.*
- (c) Little improvement is observed in the highest energies, despite them also being associated with larger weights. *This suggests that the observed high-energy underestimation is not a result of the dataset or the models, but rather a consequence of the physically available information in the hits of each event.* This is indeed the case, as we will discuss in the next Section.
- (d) The spread of the model predictions in the middle part of the energy range has been increased. This is a result of the smaller event weights in this region and is ultimately the cost required for an unbiased, better generalizing model.

5.4 Investigation of Model Predictions

The discussion of the training results, which was carried out in the previous Section, highlights the fundamental characteristics of the model predictions. It is evident that the trained DynEdge configurations, as well as the OrcaNet model, exhibit certain key features in their predictions, common across all models. If we divide the energy range of interest (10^2 - 10^8 GeV) into three regions (low, middle and high energies), we can summarize our findings in the following observations:

- (a) In *low energies* (10^2 - 10^4 GeV) both the GNN models and the standard reconstruction provide an overestimate to the true neutrino energy. Significant improvement was achieved in this region with the weighted training that was previously discussed, which increases the importance of low-energy events, allowing the network to learn those better.
- (b) The *middle energies* (10^4 - 10^6 GeV) are the region where the models perform the best. This is unsurprising given the fact that it is the energy region with the most training samples. Model predictions in this region are generally centered around the correct target value. Neutrinos in this energy range also deposit most of the light in the instrumented volume thus allowing for a more accurate energy reconstruction.
- (c) In the *high energies* (10^6 - 10^8 GeV) all models exhibit extremely similar behavior. As the true neutrino energy increases beyond 10^6 GeV, the models begin to increasingly underestimate it. In fact, all snapshot and triggered hit trainings, as well as the OrcaNet model, exhibit very similar performance in this energy region. Crucially, this performance is not significantly improved by the weighted training, contrary to what was observed for low energies. This suggests that the models are approaching the physical limit of the accuracy of the energy reconstruction. These highly energetic neutrinos leave only a fraction of the light in the detector thus leading to the underestimation of their energy. As the detector is growing, the energy reconstruction in high energies is expected to improve.

Looking at these observations, one could point out that, in the two energy extremes, the models tend to predict values closer to the middle of the energy range (*i.e.* overestimate low-energy events & underestimate high-energy events). This could potentially indicate a bias in the model predictions towards the energy region that contains more training samples, which could limit the usability of the model. However, two already mentioned mitigating factors need to be taken into account in relation to this argument:

1. The low-energy overestimation is present not only in the GNN predictions, but also in the standard reconstruction (as can be seen in Fig. 5.7). The fact that both the GNNs and classical methods produce this overestimation suggests that the issue does not lie with the use of a particular set of events to train a DL model. Further, this low-energy overestimation has been greatly reduced by the weighted training, which has eliminated any possible bias.

- The high-energy underestimation results from a physical limitation which prevents the model from producing more accurate energy predictions. This limitation stems from the fact that not all of the energy in each event is visible to the detector. In fact, for most of the events, the interaction vertex (*i.e.* the point where the neutrino CC interaction occurs) lies outside the detector volume. We call these *non-starting events*. In non-starting events, (a) the shower component of the CC interaction is not visible and (b) the muon trajectory begins outside the instrumented volume, meaning that by the time the muon has entered the detector it has already lost part of its energy. Consequently, the further away the interaction vertex is from the detector, the less of the muon energy will be visible. Furthermore, not all muon tracks stop inside the detector. If the muon carries enough energy to exit the detector, even less of its energy will be visible. *Since the majority of high-energy events interact far from the detector, it is therefore the case that the visible energy is significantly smaller than the neutrino energy.* Given this fact, it could be argued that the GNN performance on high energies is even better than expected, since the majority of events are predicted relatively near their truth value, despite this limitation.

The aforementioned points highlight the need to better understand how the model predictions can be interpreted from a physical perspective. Since ARCA is primarily intended for studying the high-energy part of the neutrino spectrum, an investigation on the high-energy event predictions will be conducted. In this effort, an attempt to identify key causal relations between the input data and target variable, which the model utilizes in order to make predictions, will be made.

To better understand the model predictions at high energies, we will split the dataset into energy slices and observe how the model predicts the events in each slice. The slices will have a width of 0.2 in $\log_{10} E_{true}$. Non-starting events in the energy ranges $10^5 - 10^6$ GeV, $10^6 - 10^7$ GeV & $10^7 - 10^8$ GeV are displayed in Figs. 5.13, 5.14 & 5.15 respectively. The true energy bins are shown in blue, while the predicted energy bins, for the same events, in orange. We can observe that, for each energy slice, the peak of the distribution of predicted values lies very near (or at) that energy slice. This remains true even in the highest energy slices. However, as we move to higher energies, the distribution of predicted values for each energy slice gets significantly wider, with many predictions being far lower than the true values. This is consistent with the fact that, for high-energy events, a smaller percentage of the energy is expected to be visible on average.

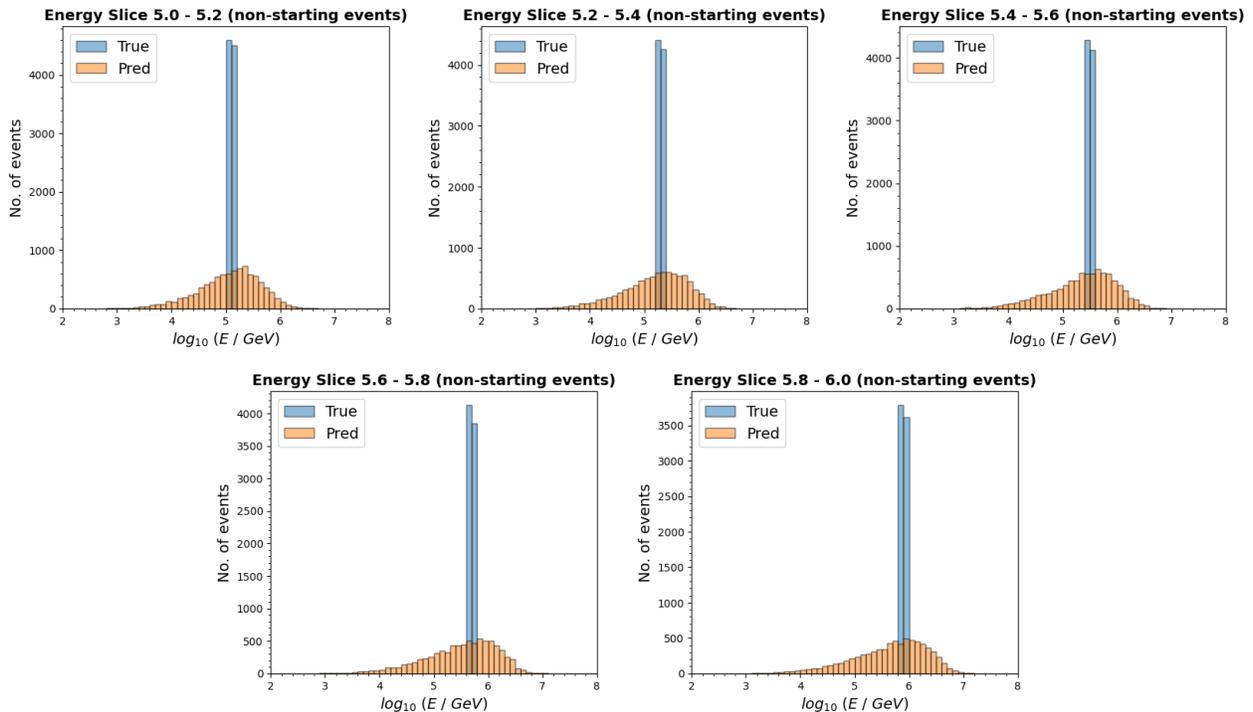


Figure 5.13: Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events in the energy range $10^5 - 10^6$ GeV.

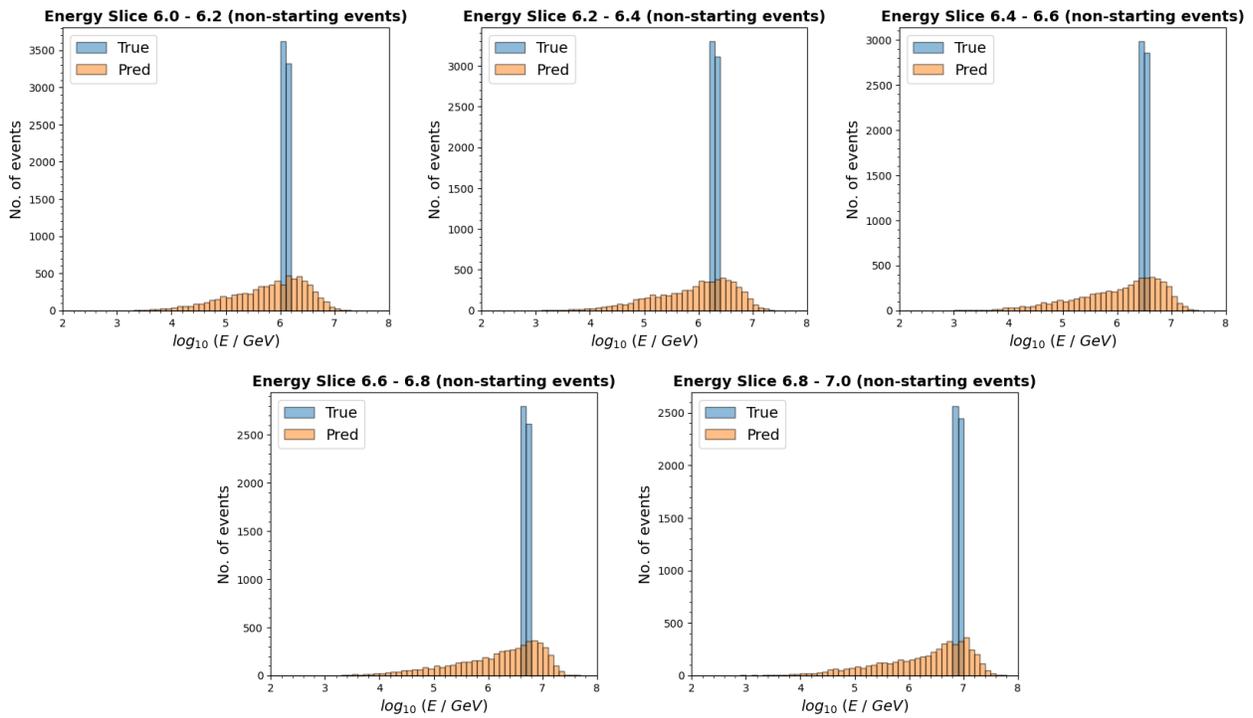


Figure 5.14: Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events in the energy range $10^6 - 10^7$ GeV.

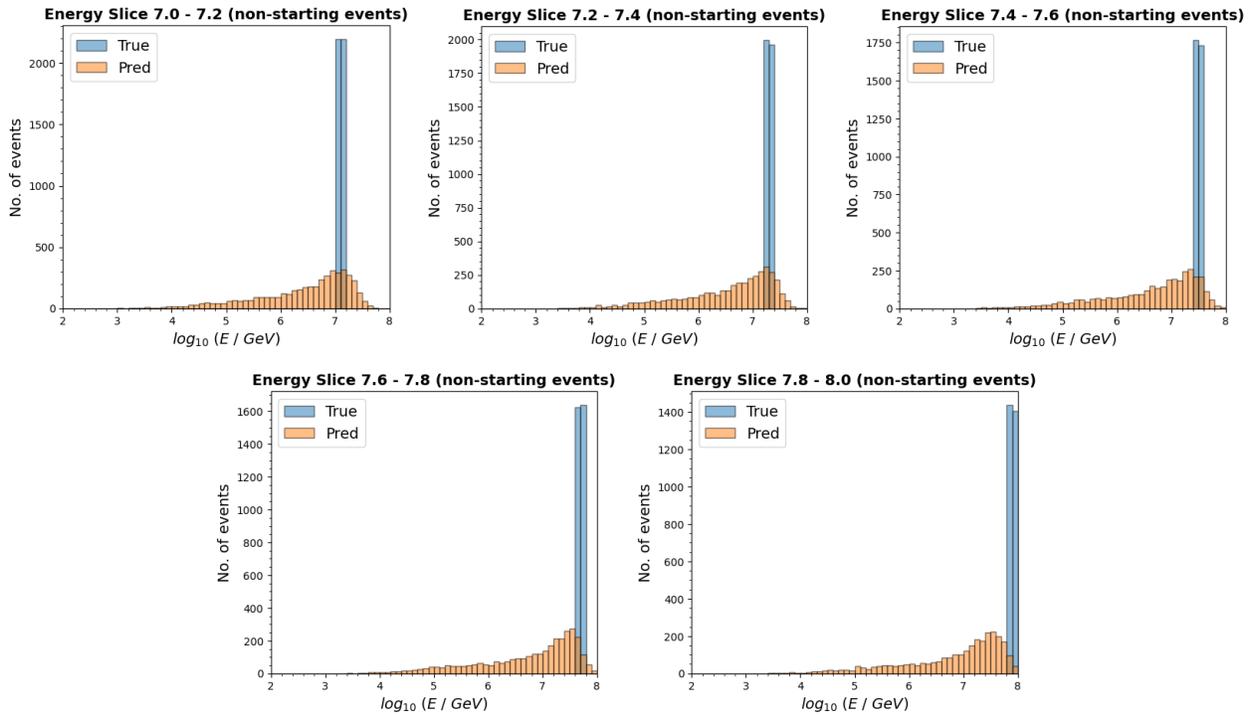


Figure 5.15: Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events in the energy range $10^7 - 10^8$ GeV.

However, the observation of peaks in the predictions of the high-energy slices suggests the existence of one or more features which allows the identification of these events as highly energetic, despite the fact that they are non-starting events. One such feature was identified to be the number of hits in each event. Plotting the number of hits with respect to the true neutrino energy (Fig. 5.16), shows that events with more than 1,000 hits are located almost exclusively in energies above 10^5 GeV. The model has correctly determined this relationship, as evident from the second plot of Fig. 5.16 (number of hits with respect to predicted energy).

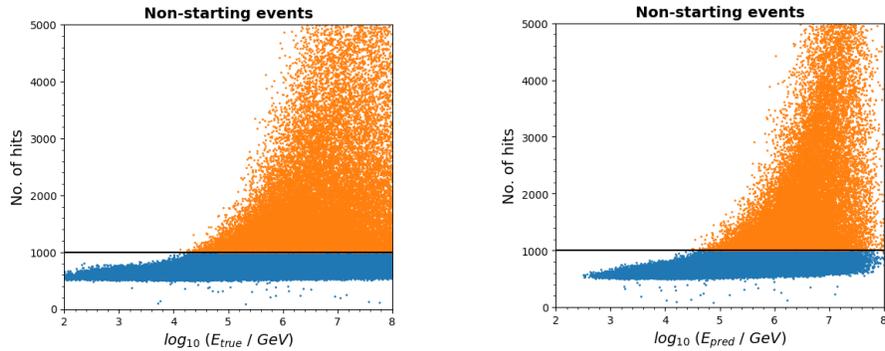


Figure 5.16: Number of hits per event with respect to the true neutrino energy (left) and the GNN predicted energy (right). Events with more than 1,000 hits are shown in orange. The network correctly identifies these as high-energy events.

Motivated by this observation, we repeat the energy slices presented above, only including non-starting events with fewer than 1,000 hits (Fig. 5.17, 5.18, & 5.19). Starting from 10^5 GeV and moving to higher energies slice by slice, we begin to notice the distribution of predictions lagging increasingly behind the true energy values, while also progressively flattening. Once we reach the highest energies, the distribution of predictions has become virtually flat. In other words, for non-starting events with fewer than 1,000 hits, *the model consistently underestimates the energy in the PeV scale (above 10^6 GeV), in accordance with expectations.*

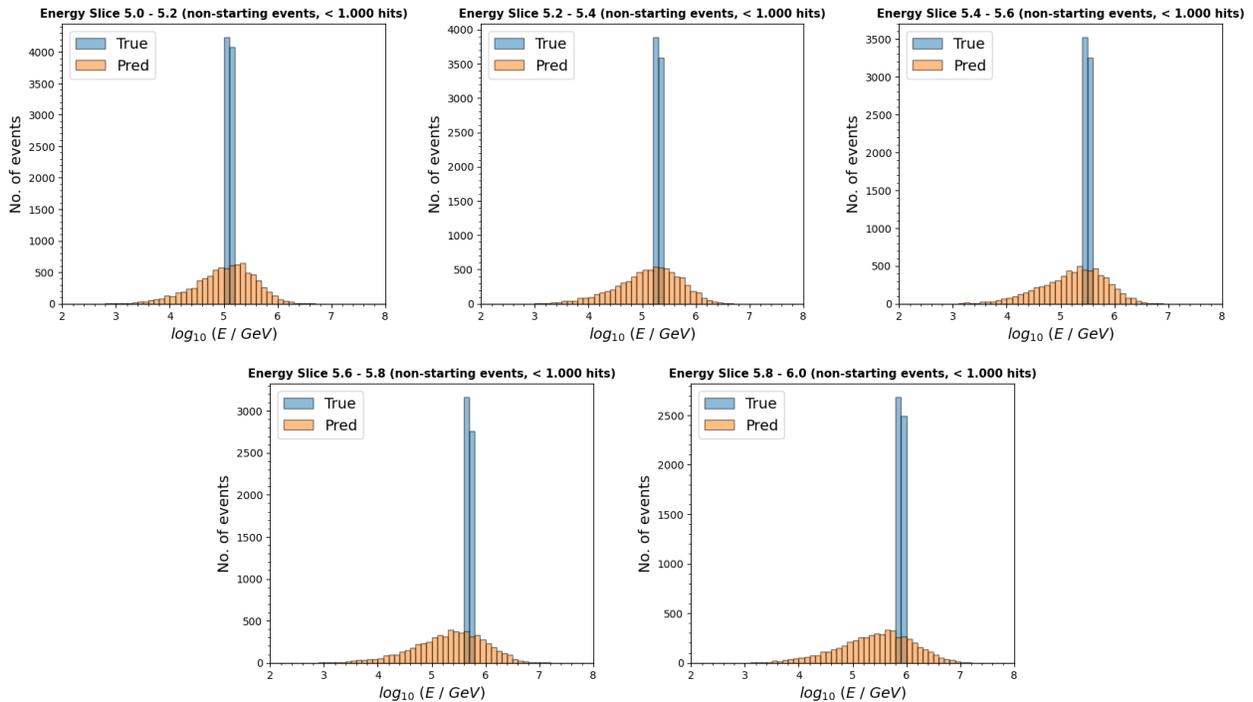


Figure 5.17: Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events with fewer than 1,000 hits in the energy range $10^5 - 10^6$ GeV.

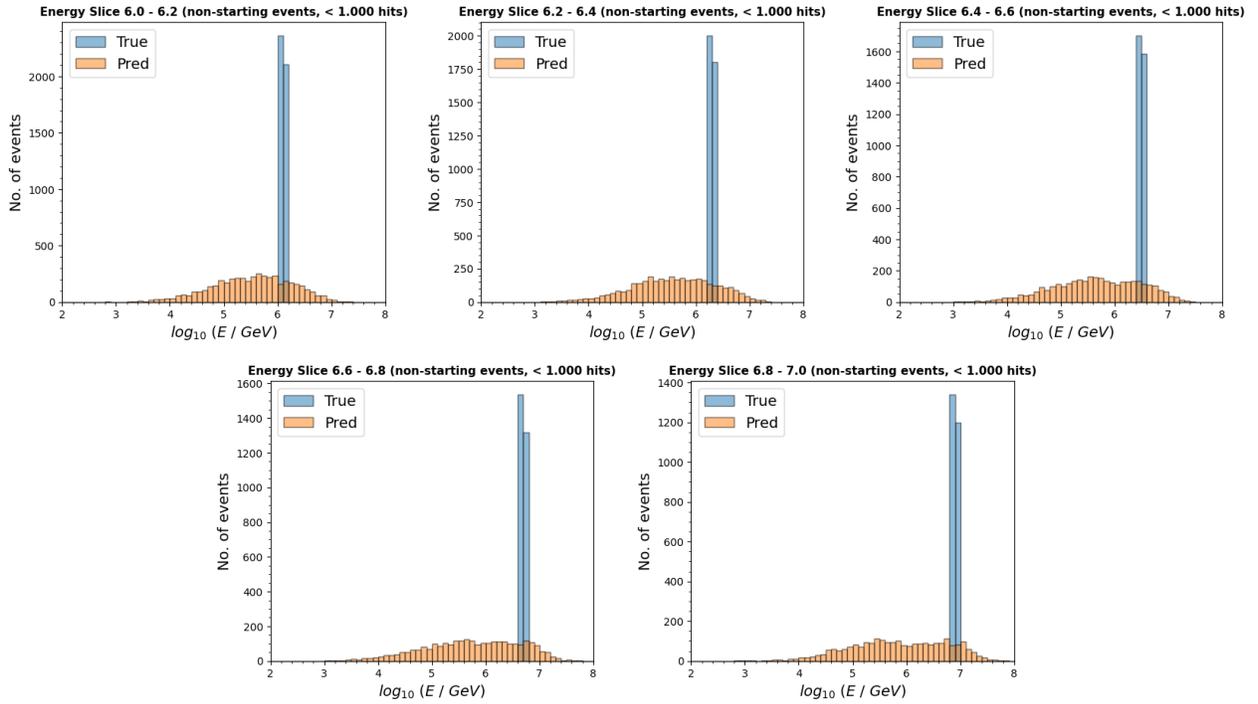


Figure 5.18: Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events with fewer than 1,000 hits in the energy range $10^6 - 10^7$ GeV.

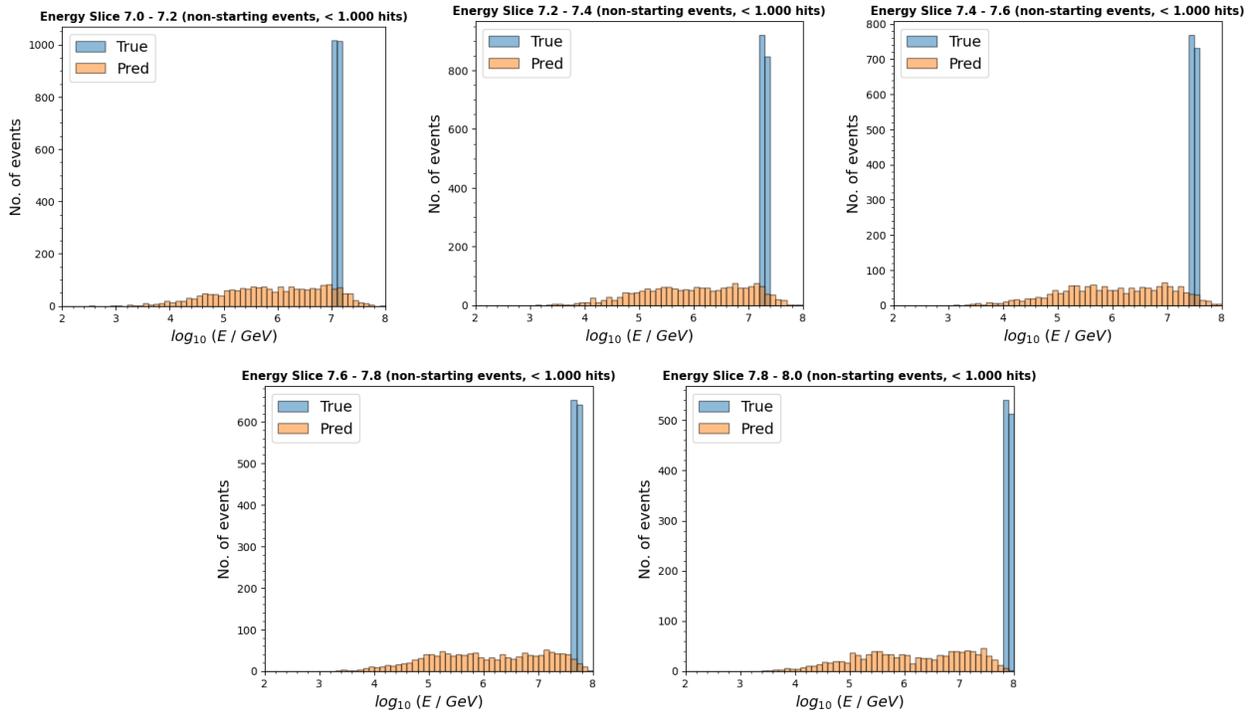


Figure 5.19: Energy slices of width 0.2 in $\log_{10} E_{true}$ (blue) and the corresponding GNN predictions (orange) for non-starting events with fewer than 1,000 hits in the energy range $10^7 - 10^8$ GeV.

Chapter 6

Conclusion

6.1 Summary	100
6.2 Future Work	100

6.1 Summary

In this work, we explored the potential of Graph Neural Networks in tackling event reconstruction tasks in neutrino experiments and specifically energy reconstruction of track-like events in the KM3NeT/ARCA21 neutrino detector. Before discussing the problem and the techniques used, a short introduction to Neutrino Physics was required. This was provided in Chapter 2, where the types of neutrino particles, as well as their sources in the TeV/PeV energy range, were presented. Further, the detection mechanism in underwater neutrino detectors, Cherenkov neutrino detection, was described and the possible event topologies were discussed.

The KM3NeT experiment was subsequently discussed in Chapter 3. After outlining the scientific goals of the experiment, the detector design is introduced, followed by a discussion of the KM3NeT trigger system and the background sources in the detectors. The Chapter concludes with the description of the problem of event reconstruction and the presentation of classical event reconstruction algorithms employed in KM3NeT, which are based on the maximum likelihood method.

Chapter 4 then introduces the main tools that are used in this thesis. After a description of Deep Neural Networks, the types of Graph Neural Networks that will be used in this thesis are then presented. We initially introduce the graph structure, followed by the concept of graph convolutions and the DynEdge network architecture, implemented in GraphNeT. Finally, we perform a review of the use of Deep Learning techniques in the event reconstruction tasks of neutrino telescopes, with a focus on the recent use of Graph Neural Networks in IceCube and KM3NeT.

Finally, in Chapter 5, we apply the DynEdge network to the problem of energy reconstruction of track-like events in the KM3NeT/ARCA21 detector. This is made possible by the integration of the KM3NeT Deep Learning data format in GraphNeT, which was carried out as part of this work. After a discussion on the dataset that was used, as well as various implementation details, the model configurations that were trained are presented. We then compare various training configurations and extensively report on the results. Overall, we find that using the snapshot hits as input for the training outperforms the trainings with triggered hits. Most notably, by incorporating the technique of sample weighting into our model, we manage to substantially enhance its performance in low-energy events. The resulting low-energy reconstruction outperforms all previous training configurations, as well as the best OrcaNet configuration and the KM3NeT standard reconstruction algorithm, with only a moderate impact in the middle of the energy range. The incorporation of event weighting also eliminates any biases of previous models present due to the non-uniform energy spectrum of our dataset. Further, an investigation into the reconstruction of high-energy events provides justification of the model predictions in this energy range and allows us to appreciate the physical constraints which limit the reconstruction of these events. Finally, this first use of the GraphNeT framework in KM3NeT showcases the potential of this valuable Deep Learning tool, which becomes readily available for the Collaboration to use in future endeavors.

6.2 Future Work

The findings and limitations discussed in this thesis regarding the application of Graph Neural Networks in KM3NeT/ARCA energy reconstruction suggest that further research in this area can be conducted, in the quest to improve reconstruction results. Future endeavors that build on this work could include:

- **Creation of a balanced dataset:** In this work, the imbalance of the ARCA21 MC dataset with respect to the target quantity was tackled through sample weighting, as has been extensively discussed. This increases the significance of underrepresented energy regions in the dataset, however the actual number of samples from these regions remains low. An alternative approach would be to create a balanced dataset, with the same number of events in each energy bin, by utilizing the full ARCA21 MC production. This would expose the model to more training examples from underrepresented energy regions, possibly allowing it to learn better representations for these regions.
- **Usage of a different target quantity:** The target energy for GNN trainings throughout this work has been the true MC neutrino energy. An argument can be made, however, for the use of a quantity which is fully inferable from the physical information present in the hits of each event, thus potentially

enhancing our confidence in the model predictions, such as the visible energy or the muon energy at a particular point on the track (the latter is also what the standard reconstruction estimates). This would require the inclusion of the new target quantity in KM3NeT DL HDF5 files.

- **Employment of novel network architectures:** Although the use of Graph Neural Networks has proven successful in neutrino detector reconstruction tasks, there has been growing interest in the community in exploring other types of models. Most notably, the application of Transformers, either in standalone or in hybrid GNN + Transformer architectures, has recently attracted attention, with such models already being implemented in GraphNeT.
- **Application on larger detector versions:** In their current form, the KM3NeT detectors are only a fraction of their planned final size. As the detectors are being extended, so are their neutrino detection capabilities, especially with regards to the high-energy regions of the neutrino energy spectrum. The application of Deep Learning models in reconstruction tasks for larger detector versions can thus provide a more accurate representation of the potential of these techniques and their comparison to classical methods.

Bibliography

- [1] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli. “The standard model of particle physics”. *Rev. Mod. Phys.* 71 (2 Mar. 1999), pp. 96–111. DOI: [10.1103/RevModPhys.71.S96](https://doi.org/10.1103/RevModPhys.71.S96). URL: <https://link.aps.org/doi/10.1103/RevModPhys.71.S96>.
- [2] Cush, via Wikimedia Commons. *Standard Model of Elementary Particles*. Accessed 05-July-2024. URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [3] W. Pauli. “Dear radioactive ladies and gentlemen”. *Phys. Today* 31N9 (1978), p. 27.
- [4] F. L. Wilson. “Fermi’s Theory of Beta Decay”. *American Journal of Physics* 36.12 (Dec. 1968), pp. 1150–1160. ISSN: 0002-9505. DOI: [10.1119/1.1974382](https://doi.org/10.1119/1.1974382). URL: <https://doi.org/10.1119/1.1974382>.
- [5] C. L. Cowan et al. “Detection of the Free Neutrino: a Confirmation”. *Science* 124.3212 (1956), pp. 103–104. DOI: [10.1126/science.124.3212.103](https://doi.org/10.1126/science.124.3212.103). URL: <https://www.science.org/doi/abs/10.1126/science.124.3212.103>.
- [6] G. Danby et al. “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos”. *Phys. Rev. Lett.* 9 (1 July 1962), pp. 36–44. DOI: [10.1103/PhysRevLett.9.36](https://doi.org/10.1103/PhysRevLett.9.36). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.9.36>.
- [7] K. Kodama et al. “Observation of tau neutrino interactions”. *Physics Letters B* 504.3 (2001), pp. 218–224. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(01\)00307-0](https://doi.org/10.1016/S0370-2693(01)00307-0). URL: <https://www.sciencedirect.com/science/article/pii/S0370269301003070>.
- [8] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: [10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [9] Q. R. Ahmad et al. “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”. *Phys. Rev. Lett.* 89 (1 June 2002), p. 011301. DOI: [10.1103/PhysRevLett.89.011301](https://doi.org/10.1103/PhysRevLett.89.011301). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.89.011301>.
- [10] Particle Data Group. “Review of particle physics”. *Phys. Rev. D* 110 (3 Aug. 2024). Chapter 14, p. 030001. DOI: [10.1103/PhysRevD.110.030001](https://doi.org/10.1103/PhysRevD.110.030001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.110.030001>.
- [11] C. Spiering. “Towards high-energy neutrino astronomy”. *The European Physical Journal H* 37.3 (Aug. 2012), pp. 515–565. ISSN: 2102-6467. DOI: [10.1140/epjh/e2012-30014-2](https://doi.org/10.1140/epjh/e2012-30014-2). URL: <https://doi.org/10.1140/epjh/e2012-30014-2>.
- [12] T. Kajita. “Atmospheric neutrinos and discovery of neutrino oscillations”. *Proceedings of the Japan Academy, Series B* 86.4 (2010), pp. 303–321. DOI: [10.2183/pjab.86.303](https://doi.org/10.2183/pjab.86.303).
- [13] K. Okumura. “Atmospheric Neutrino Flux Measurement”. *EPJ Web Conf.* 208 (2019), p. 07002. DOI: [10.1051/epjconf/201920807002](https://doi.org/10.1051/epjconf/201920807002). URL: <https://doi.org/10.1051/epjconf/201920807002>.
- [14] W. Bai et al. “Forward production of prompt neutrinos from charm in the atmosphere and at high energy colliders”. *Journal of High Energy Physics* 2023.10 (Oct. 2023), p. 142. ISSN: 1029-8479. DOI: [10.1007/JHEP10\(2023\)142](https://doi.org/10.1007/JHEP10(2023)142). URL: [https://doi.org/10.1007/JHEP10\(2023\)142](https://doi.org/10.1007/JHEP10(2023)142).
- [15] U. F. Katz and C. Spiering. “High-energy neutrino astrophysics: Status and perspectives”. *Progress in Particle and Nuclear Physics* 67.3 (2012), pp. 651–704. ISSN: 0146-6410. DOI: <https://doi.org/10.1016/j.pnpnp.2011.12.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0146641011001189>.

- [16] M. Ahlers and K. Murase. “Probing the Galactic origin of the IceCube excess with gamma rays”. *Phys. Rev. D* 90 (2 July 2014), p. 023010. DOI: [10.1103/PhysRevD.90.023010](https://doi.org/10.1103/PhysRevD.90.023010). URL: <https://link.aps.org/doi/10.1103/PhysRevD.90.023010>.
- [17] D. F. G. Fiorillo. “High-Energy and Ultra-High-Energy Neutrino Astrophysics”. *Universe* 10.3 (2024). ISSN: 2218-1997. DOI: [10.3390/universe10030149](https://doi.org/10.3390/universe10030149). URL: <https://www.mdpi.com/2218-1997/10/3/149>.
- [18] R. Abbasi et al. “Improved Characterization of the Astrophysical Muon-neutrino Flux with 9.5 Years of IceCube Data”. *The Astrophysical Journal* 928.1 (Mar. 2022), p. 50. DOI: [10.3847/1538-4357/ac4d29](https://doi.org/10.3847/1538-4357/ac4d29). URL: <https://dx.doi.org/10.3847/1538-4357/ac4d29>.
- [19] The IceCube Collaboration et al. “Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A”. *Science* 361.6398 (2018), eaat1378. DOI: [10.1126/science.aat1378](https://doi.org/10.1126/science.aat1378). URL: <https://www.science.org/doi/abs/10.1126/science.aat1378>.
- [20] R. Abbasi et al. “Evidence for neutrino emission from the nearby active galaxy NGC 1068”. *Science* 378.6619 (2022), pp. 538–543. DOI: [10.1126/science.abg3395](https://doi.org/10.1126/science.abg3395). URL: <https://www.science.org/doi/abs/10.1126/science.abg3395>.
- [21] R. Abbasi et al. “Observation of high-energy neutrinos from the Galactic plane”. *Science* 380.6652 (2023), pp. 1338–1343. DOI: [10.1126/science.adc9818](https://doi.org/10.1126/science.adc9818). URL: <https://www.science.org/doi/abs/10.1126/science.adc9818>.
- [22] J. A. Formaggio and G. P. Zeller. “From eV to EeV: Neutrino cross sections across energy scales”. *Rev. Mod. Phys.* 84 (3 Sept. 2012), pp. 1307–1341. DOI: [10.1103/RevModPhys.84.1307](https://doi.org/10.1103/RevModPhys.84.1307). URL: <https://link.aps.org/doi/10.1103/RevModPhys.84.1307>.
- [23] S. L. Glashow. “Resonant Scattering of Antineutrinos”. *Phys. Rev.* 118 (1 Apr. 1960), pp. 316–317. DOI: [10.1103/PhysRev.118.316](https://doi.org/10.1103/PhysRev.118.316). URL: <https://link.aps.org/doi/10.1103/PhysRev.118.316>.
- [24] M. G. Aartsen et al. “Detection of a particle shower at the Glashow resonance with IceCube”. *Nature* 591.7849 (Mar. 2021), pp. 220–224. ISSN: 1476-4687. DOI: [10.1038/s41586-021-03256-1](https://doi.org/10.1038/s41586-021-03256-1). URL: <https://doi.org/10.1038/s41586-021-03256-1>.
- [25] C.-Y. Chen, P. S. B. Dev, and A. Soni. “Standard model explanation of the ultrahigh energy neutrino events at IceCube”. *Phys. Rev. D* 89 (3 Feb. 2014), p. 033012. DOI: [10.1103/PhysRevD.89.033012](https://doi.org/10.1103/PhysRevD.89.033012). URL: <https://link.aps.org/doi/10.1103/PhysRevD.89.033012>.
- [26] P. A. Čerenkov. “Visible Radiation Produced by Electrons Moving in a Medium with Velocities Exceeding that of Light”. *Phys. Rev.* 52 (4 Aug. 1937), pp. 378–379. DOI: [10.1103/PhysRev.52.378](https://doi.org/10.1103/PhysRev.52.378). URL: <https://link.aps.org/doi/10.1103/PhysRev.52.378>.
- [27] A. Horvath, [CC BY-SA 2.5](https://creativecommons.org/licenses/by-sa/2.5/), via Wikimedia Commons. *Cherenkov*. Accessed 05-July-2024. URL: <https://commons.wikimedia.org/wiki/File:Cherenkov.svg>.
- [28] S. Adrian-Martinez et al. “Letter of intent for KM3NeT 2.0”. *J. Phys. G* 43.8 (2016), p. 084001. DOI: [10.1088/0954-3899/43/8/084001](https://doi.org/10.1088/0954-3899/43/8/084001). URL: <https://iopscience.iop.org/article/10.1088/0954-3899/43/8/084001>.
- [29] M. Ageron et al. “ANTARES: The first undersea neutrino telescope”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 656.1 (2011), pp. 11–38. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2011.06.103>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900211013994>.
- [30] A. D. Avrorin et al. “Baikal-GVD: status and prospects”. *EPJ Web Conf.* 191 (2018), p. 01006. DOI: [10.1051/epjconf/201819101006](https://doi.org/10.1051/epjconf/201819101006). URL: <https://doi.org/10.1051/epjconf/201819101006>.
- [31] A. Achterberg et al. “First year performance of the IceCube neutrino telescope”. *Astroparticle Physics* 26.3 (2006), pp. 155–173. ISSN: 0927-6505. DOI: <https://doi.org/10.1016/j.astropartphys.2006.06.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0927650506000855>.
- [32] S. Fukuda et al. “The Super-Kamiokande detector”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 501.2 (2003), pp. 418–462. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)00425-X](https://doi.org/10.1016/S0168-9002(03)00425-X). URL: <https://www.sciencedirect.com/science/article/pii/S016890020300425X>.
- [33] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. ISBN: 9781107034266. DOI: <https://doi.org/10.1017/CB09781139525367>.
- [34] D. Guderian. “Development of detector calibration and graph neural network-based selection and reconstruction algorithms for the measurement of oscillation parameters with KM3NeT/ORCA”. PhD thesis. Universität Münster, 2022.

-
- [35] M. G. Aartsen et al. “Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector”. *Science* 342.6161 (2013), p. 1242856. DOI: [10.1126/science.1242856](https://doi.org/10.1126/science.1242856). URL: <https://www.science.org/doi/abs/10.1126/science.1242856>.
- [36] I. Esteban et al. “The fate of hints: updated global analysis of three-flavor neutrino oscillations”. *Journal of High Energy Physics* 2020.9 (Sept. 2020), p. 178. ISSN: 1029-8479. DOI: [10.1007/JHEP09\(2020\)178](https://doi.org/10.1007/JHEP09(2020)178). URL: [https://doi.org/10.1007/JHEP09\(2020\)178](https://doi.org/10.1007/JHEP09(2020)178).
- [37] *NuFIT 5.3*. Accessed 04-August-2024. Mar. 2024. URL: <http://www.nu-fit.org/>.
- [38] S. Aiello et al. “The KM3NeT multi-PMT optical module”. *Journal of Instrumentation* 17.07 (July 2022), P07038. DOI: [10.1088/1748-0221/17/07/P07038](https://dx.doi.org/10.1088/1748-0221/17/07/P07038). URL: <https://dx.doi.org/10.1088/1748-0221/17/07/P07038>.
- [39] J. Chen. “Nuclear Data Sheets for A=40”. *Nuclear Data Sheets* 140 (2017), pp. 1–376. ISSN: 0090-3752. DOI: <https://doi.org/10.1016/j.nds.2017.02.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0090375217300169>.
- [40] S. H. D. Haddock, M. A. Moline, and J. F. Case. “Bioluminescence in the Sea”. *Annual Review of Marine Science* 2. Volume 2, 2010 (2010), pp. 443–493. ISSN: 1941-0611. DOI: <https://doi.org/10.1146/annurev-marine-120308-081028>. URL: <https://www.annualreviews.org/content/journals/10.1146/annurev-marine-120308-081028>.
- [41] B. Baret and V. Van Elewyck. “High-energy neutrino astronomy: detection methods and first achievements”. *Reports on Progress in Physics* 74.4 (Apr. 2011), p. 046902. DOI: [10.1088/0034-4885/74/4/046902](https://dx.doi.org/10.1088/0034-4885/74/4/046902). URL: <https://dx.doi.org/10.1088/0034-4885/74/4/046902>.
- [42] K. Melis, A. Heijboer, and M. De Jong. “KM3NeT/ARCA Event Reconstruction Algorithms”. *PoS ICRC2017* (2017), p. 950. DOI: [10.22323/1.301.0950](https://doi.org/10.22323/1.301.0950).
- [43] W. S. McCulloch and W. Pitts. “A logical calculus of the ideas immanent in nervous activity”. *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: [10.1007/BF02478259](https://doi.org/10.1007/BF02478259). URL: <https://doi.org/10.1007/BF02478259>.
- [44] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Taylor & Francis, 2005. ISBN: 9781135631901. URL: <https://books.google.gr/books?id=ddb4AgAAQBAJ>.
- [45] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [46] J. Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. *Nature* 596.7873 (Aug. 2021), pp. 583–589. ISSN: 1476-4687. DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2). URL: <https://doi.org/10.1038/s41586-021-03819-2>.
- [47] H.-P. Chan et al. “Deep Learning in Medical Image Analysis”. In: *Deep Learning in Medical Image Analysis : Challenges and Applications*. Ed. by G. Lee and H. Fujita. Cham: Springer International Publishing, 2020, pp. 3–21. ISBN: 978-3-030-33128-3. DOI: [10.1007/978-3-030-33128-3_1](https://doi.org/10.1007/978-3-030-33128-3_1). URL: https://doi.org/10.1007/978-3-030-33128-3_1.
- [48] K. Choudhary et al. “Recent advances and applications of deep learning methods in materials science”. *npj Computational Materials* 8.1 (Apr. 2022), p. 59. ISSN: 2057-3960. DOI: [10.1038/s41524-022-00734-6](https://doi.org/10.1038/s41524-022-00734-6). URL: <https://doi.org/10.1038/s41524-022-00734-6>.
- [49] G. Carleo et al. “Machine learning and the physical sciences”. *Rev. Mod. Phys.* 91 (4 Dec. 2019), p. 045002. DOI: [10.1103/RevModPhys.91.045002](https://doi.org/10.1103/RevModPhys.91.045002). URL: <https://link.aps.org/doi/10.1103/RevModPhys.91.045002>.
- [50] G. Karagiorgi et al. “Machine learning in the search for new fundamental physics”. *Nature Reviews Physics* 4.6 (June 2022), pp. 399–412. ISSN: 2522-5820. DOI: [10.1038/s42254-022-00455-1](https://doi.org/10.1038/s42254-022-00455-1). URL: <https://doi.org/10.1038/s42254-022-00455-1>.
- [51] F. Psihas et al. “A review on machine learning for neutrino experiments”. *International Journal of Modern Physics A* 35.33 (2020), p. 2043005. DOI: [10.1142/S0217751X20430058](https://doi.org/10.1142/S0217751X20430058). URL: <https://doi.org/10.1142/S0217751X20430058>.
- [52] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. *Nature* 323.6088 (Oct. 1986), pp. 533–536. ISSN: 1476-4687. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0). URL: <https://doi.org/10.1038/323533a0>.
- [53] M. A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. URL: <http://neuralnetworksanddeeplearning.com/>.
- [54] S. Ruder. “An overview of gradient descent optimization algorithms” (2017). arXiv: [1609.04747](https://arxiv.org/abs/1609.04747). URL: <https://arxiv.org/abs/1609.04747>.
-

- [55] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization” (2014). arXiv: [1412.6980](https://arxiv.org/abs/1412.6980). URL: <https://arxiv.org/abs/1412.6980>.
- [56] S. Thais et al. “Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges” (2022). arXiv: [2203.12852](https://arxiv.org/abs/2203.12852). URL: <https://arxiv.org/abs/2203.12852>.
- [57] J. Shlomi, P. Battaglia, and J.-R. Vlimant. “Graph neural networks in particle physics”. *Machine Learning: Science and Technology* 2.2 (Dec. 2020), p. 021001. DOI: [10.1088/2632-2153/abbf9a](https://doi.org/10.1088/2632-2153/abbf9a). URL: <https://dx.doi.org/10.1088/2632-2153/abbf9a>.
- [58] J. Duarte and J.-R. Vlimant. “Graph Neural Networks for Particle Tracking and Reconstruction”. In: *Artificial Intelligence for High Energy Physics*. Chap. Chapter 12, pp. 387–436. DOI: [10.1142/9789811234033_0012](https://doi.org/10.1142/9789811234033_0012). URL: https://www.worldscientific.com/doi/abs/10.1142/9789811234033_0012.
- [59] S. Reck. “Cosmic ray composition measurement using Graph Neural Networks in KM3NeT ORCA”. PhD thesis. FAU Erlangen-Nürnberg, 2023.
- [60] Y. Wang et al. “Dynamic Graph CNN for Learning on Point Clouds”. *ACM Trans. Graph.* 38.5 (Oct. 2019). ISSN: 0730-0301. DOI: [10.1145/3326362](https://doi.org/10.1145/3326362). URL: <https://doi.org/10.1145/3326362>.
- [61] R. Abbasi et al. “Graph Neural Networks for low-energy event classification & reconstruction in IceCube”. *Journal of Instrumentation* 17.11 (Nov. 2022), P11003. DOI: [10.1088/1748-0221/17/11/P11003](https://doi.org/10.1088/1748-0221/17/11/P11003). URL: <https://dx.doi.org/10.1088/1748-0221/17/11/P11003>.
- [62] A. Paszke et al. “PyTorch: an imperative style, high-performance deep learning library”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2019.
- [63] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <https://www.tensorflow.org/>.
- [64] A. Sogaard et al. “GraphNeT: Graph neural networks for neutrino telescope event reconstruction”. *Journal of Open Source Software* 8.85 (2023), p. 4971. DOI: [10.21105/joss.04971](https://doi.org/10.21105/joss.04971). URL: <https://doi.org/10.21105/joss.04971>.
- [65] M. Fey and J. E. Lenssen. “Fast Graph Representation Learning with PyTorch Geometric”. In: *ICLR Workshop on Representation Learning on Graphs and Manifolds*. 2019.
- [66] W. Falcon and The PyTorch Lightning team. *PyTorch Lightning*. Version 1.4. Mar. 2019. DOI: [10.5281/zenodo.3828935](https://doi.org/10.5281/zenodo.3828935). URL: <https://github.com/Lightning-AI/lightning>.
- [67] S. Reck et al. *OrcaSong*. Version 4.8.7. 2022. URL: <https://git.km3net.de/ml/OrcaSong>.
- [68] S. Reck et al. *OrcaNet*. Version 1.0.2. 2022. URL: <https://git.km3net.de/ml/OrcaNet>.
- [69] R. Abbasi et al. “A muon-track reconstruction exploiting stochastic losses for large-scale Cherenkov detectors”. *Journal of Instrumentation* 16.08 (Aug. 2021), P08034. DOI: [10.1088/1748-0221/16/08/P08034](https://doi.org/10.1088/1748-0221/16/08/P08034). URL: <https://dx.doi.org/10.1088/1748-0221/16/08/P08034>.
- [70] R. Abbasi et al. “Low energy event reconstruction in IceCube DeepCore”. *The European Physical Journal C* 82.9 (Sept. 2022), p. 807. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-022-10721-2](https://doi.org/10.1140/epjc/s10052-022-10721-2). URL: <https://doi.org/10.1140/epjc/s10052-022-10721-2>.
- [71] A. Aurisano et al. “A convolutional neural network neutrino event classifier”. *Journal of Instrumentation* 11.09 (Sept. 2016), P09001. DOI: [10.1088/1748-0221/11/09/P09001](https://doi.org/10.1088/1748-0221/11/09/P09001). URL: <https://dx.doi.org/10.1088/1748-0221/11/09/P09001>.
- [72] R. Abbasi et al. “A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory”. *Journal of Instrumentation* 16.07 (July 2021), P07041. DOI: [10.1088/1748-0221/16/07/P07041](https://doi.org/10.1088/1748-0221/16/07/P07041). URL: <https://dx.doi.org/10.1088/1748-0221/16/07/P07041>.
- [73] S. Aiello et al. “Event reconstruction for KM3NeT/ORCA using convolutional neural networks”. *Journal of Instrumentation* 15.10 (Oct. 2020), P10005. DOI: [10.1088/1748-0221/15/10/P10005](https://doi.org/10.1088/1748-0221/15/10/P10005). URL: <https://dx.doi.org/10.1088/1748-0221/15/10/P10005>.
- [74] H. Qu and L. Gouskos. “Jet tagging via particle clouds”. *Phys. Rev. D* 101 (5 Mar. 2020), p. 056019. DOI: [10.1103/PhysRevD.101.056019](https://doi.org/10.1103/PhysRevD.101.056019). URL: <https://link.aps.org/doi/10.1103/PhysRevD.101.056019>.
- [75] C. Bellenghi et al. “Extending the IceCube search for neutrino point sources in the Northern sky with additional years of data”. *PoS ICRC2023* (2023), p. 1060. DOI: [10.22323/1.444.1060](https://doi.org/10.22323/1.444.1060).
- [76] S. Reck, T. Eberl, and U. Katz. “Muon bundle reconstruction with KM3NeT/ORCA using graph convolutional networks”. *PoS ICRC2021* (2021), p. 1048. DOI: [10.22323/1.395.1048](https://doi.org/10.22323/1.395.1048).

- [77] S. Reck et al. “Graph neural networks for reconstruction and classification in KM3NeT”. *Journal of Instrumentation* 16.10 (Oct. 2021), p. C10011. DOI: [10.1088/1748-0221/16/10/C10011](https://doi.org/10.1088/1748-0221/16/10/C10011). URL: <https://dx.doi.org/10.1088/1748-0221/16/10/C10011>.
- [78] F. Filippini et al. “Data reconstruction and classification with Graph neural networks in KM3NeT/ARCA”. *PoS ICRC2023* (2023), p. 1194. DOI: [10.22323/1.444.1194](https://doi.org/10.22323/1.444.1194).