

**Neutrinos, Neurons and Neutron Stars: Applications of new
statistical and analysis techniques to particle and
astrophysics**

by

Gabriel L.W.H. Collin

Ph.B., Australian National University (2011)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Signature redacted

Author

.....
Department of Physics
May 24, 2018

Signature redacted

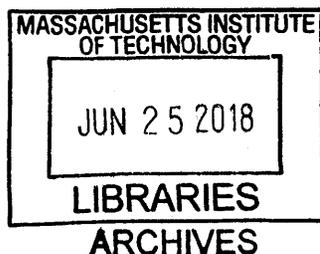
Certified by....

.....
Janet Conrad
Professor
Thesis Supervisor

Signature redacted

Accepted by.

.....
Scott A. Hughes
Interim Associate Department Head





77 Massachusetts Avenue
Cambridge, MA 02139
<http://libraries.mit.edu/ask>

DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

The images contained in this document are of the best quality available.

Neutrinos, Neurons and Neutron Stars: Applications of new statistical and analysis techniques to particle and astrophysics

Gabriel L.W.H. Collin

Submitted to the Department of Physics
on May 24, 2018, in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

ABSTRACT



The IceCube detector opens a new window into our universe; valuable for both astronomy and particle physics. This thesis spans a wide range of topics that are bound together by a common theme: the development and application of new statistical and computational methods for analysing data from particle and astrophysics experiments.

Sterile neutrinos are a hypothetical fourth kind of neutrino, which are motivated by anomalies observed in various short base-line neutrino experiments. These experiments have published results that are not mutually compatible. This thesis presents a global fit to many short base-line datasets with the addition of the recent IceCube sterile neutrino search, constraining the full $3+1$ mixing matrix for the first time. The global fit strongly favours the sterile neutrino hypothesis, although significant tension still remains within the datasets.

The origin of the observed astrophysical neutrino flux at IceCube remains elusive. Current methods, using a hot-spot model, have seen no significant clustering of events. This thesis presents a new test for point sources of neutrinos, based on the non-Poissonian Template Fitting technique. Constraints on population models for neutrino points sources are shown for the first time.

Atmospheric neutrinos form a background for astrophysical analyses on IceCube, but also serve as the signal in particle physics analyses such as the sterile neutrino search. The first comprehensive study of the effect of global atmospheric temperature variations on atmospheric neutrino fluxes is provided.

This thesis also presents two studies on using new computational methods for simulation and reconstruction on IceCube. Convolutional neural networks have been used to classify low-level waveform data, with the goal of identifying tau-neutrinos. Metropolis light transport, a rendering technique used in the CGI industry, has been extended to simulate the transport of light inside the IceCube experiment. Both show promising results, exceeding existing algorithms in their test cases.

Thesis Supervisor: Janet Conrad
Title: Professor

ACKNOWLEDGEMENTS



First, I am indebted to my parents, Mary Hinchey and Graeme Collin. The sacrifices you have made so that I could pursue my dreams have not gone un-noticed. Without your support, this thesis, and life on the other side of the world, would simply not be possible.

To my sister, Julia, thank you for putting up with my ramblings on science, doting on me when I visit, and checking for brotherly existence across the internet. To Hayden, thank you for hosting me when I visit Melbourne, I look forward to our next escape room and trivia night.

My friends in Australia (and those who escaped); Chris, Dario, Andrew, Byron, Charlotte, John, McB, Evan, Kim, Adelaide, Anna, Alex Raupach, Alex Thorman, Dante, Tash, David, Eman, Jack, Millie, and Sam: thank you for keeping me rooted in my home-away-from-home. The time you give to me, and the company you provide when I visit is much appreciated.

To my international support group; Alex Leder, Lina, and Nancy: words cannot express how important your friendship has been over the past years, this thesis is as much yours as it is mine. To John: thank you for putting up with my mad ravings on programming languages and other topics, few people appreciate my interests as much as you do.

My Boston friends; Alex Ji, Cody, Olivia, Jenny, Tom, Tom, Mukund, Jason, Keaton, Nick, Matt, James, Janos, Kiaran, Murphy, Helena and Ted: the thesis may stop, but the board games will continue. I will forever treasure the time we spent together. With a special mention to Ian: the greatest house-mate of all time.

To my current and future LNS colleagues; Charles, Stan, Tom, Yunjie, Joe, Field, among many others: you're not quite rid of me yet! And my IceCube collaborators; Ali, Blake, Chris, Dima, Donglian, Ibrahim, Logan, Shivesh, and Teppei: until our next fateful meeting in the hurricane room!

Peter Vassiliou: you planted the seed in my mind to come to America, no doubt some of the best advice I will ever receive. To my Honours advisor, Greg Lane: I will always be grateful for the time and effort you put into me as a student, you helped me find my way in a forest of many paths.

To my first post-doc, Matt Toups: you showed great patience and humour in showing me the ropes of a new field. And to my last post-doc, Carlos Argüelles: your sense of humour, depth of

knowledge, and dedication to your craft is extraordinary; we'll find a way to collaborate in the future, I'll make sure of it.

Kim Kreiger, Nicole Laraia and Cathy Modica: behind any successful enterprise is an excellent administrator, I am grateful for all the hard work you put into organising us and making our lives easier. To the MIT staff who cannot be listed here: thank you for your kindness, support and understanding over the years.

To the current and former members of our group; Len, Kazu, Josh, Jose, Ami, Zander, Allie, Marjon, Spencer, Jarrett, Alex, Lauren, Thomas, John, Loyd, and Joe: we work well together for a reason. With a special mention to Taritree, for diving into deep learning with me, and providing the machine that we train all our models on; Daniel, it was a pleasure to share our office through so many years; Christina, you helped me get started in phenomenology, and helped me through the dark days of quals; and Ben Jones, you welcomed me to not just one, but two collaborations, and your council in my projects over these last years is of great value to me. To Ben Safdi: thank you for knocking on our door with the idea of applying new statistical tools to point source searches in IceCube, it is a topic that I have now become deeply interested in.

Mike Shaevitz: I have been honoured to have you as an advisor in my graduate studies, your enthusiasm for statistics set me on this path, and forged a vision for my research. Francis Halzen and Peter Fisher: thank you for supporting me in taking my next steps as a scientist.

And, to Janet Conrad: I'm not sure, in the history our universe, that a better advisor has ever existed. It has been a roller-coaster of a thesis, for sure; but, your vision, your determination, your receptivity is an inspiration that shines through. You gave me the freedom, and encouragement, to find my own way. I look forward to my graduation, as I will no longer call you my boss; but, instead, my friend.



Contents

Introduction	17
1 Neutrinos in the standard model	21
1.1 The physics of neutrinos	21
1.1.1 Flavour physics	24
1.1.2 Neutrino oscillation	25
1.2 Production and detection	28
1.2.1 Low energy	29
1.2.2 Medium energy	32
1.2.3 High energy	34
1.2.4 Neutrino-lepton scattering	36
1.2.5 Detector signatures	37
2 Statistical inference	41
2.1 Information	42
2.2 Inference	44
2.2.1 Frequentist methods	45
2.2.2 Bayesian inference	47
2.3 Sampling	50
2.3.1 Re-weighting and re-sampling	52
2.4 Markov Chain Monte-Carlo	53
2.4.1 Affine invariant Markov Chain Monte-Carlo	55
2.4.2 Parallel tempering	56
2.4.3 Reversible jump	57
2.5 Nested sampling	58
3 Global fits to sterile neutrino models	61
3.1 Anomalies	61
3.1.1 LSND	62
3.1.2 MiniBooNE	62
3.1.3 Nuclear reactors	63
3.1.4 Electron capture calibration	64
3.2 Sterile neutrinos	64
3.2.1 Incorporation into the ν SM	65
3.3 Global fit	66
3.3.1 Inference framework	67
3.3.2 MCMC implementation	68

3.3.3	Likelihood function	69
3.3.4	Predictions	71
3.3.5	Presentation of results	71
3.3.6	Experimental data sets	72
3.4	Results	73
3.4.1	3 + 1 model	73
3.4.2	3 + 2 model	76
3.5	Incorporating the IceCube sterile neutrino search	76
3.5.1	Predictions	78
3.5.2	Inference framework	79
3.5.3	Likelihood	80
3.5.4	Results	80
3.6	Summary	82
4	Detecting neutrinos with IceCube	85
4.1	The IceCube detector	86
4.1.1	IceTop	87
4.1.2	Detector signatures	87
4.2	Digital Optical Modules	89
4.2.1	Photo-multiplier tube	90
4.2.2	Digitising electronics	90
4.3	The ice	91
4.3.1	Light model	91
4.3.2	Bulk ice model	92
4.3.3	Hole ice	94
4.4	Simulation and reconstruction	94
4.4.1	Event generation	94
4.4.2	Detector simulation	95
4.4.3	Track reconstruction	96
4.4.4	Cascade reconstruction	96
5	Atmospheric neutrino flux	99
5.1	Shower development	100
5.2	Matrix cascade approach	102
5.3	Atmospheric Density Model	103
5.3.1	Existing models	103
5.3.2	Atmospheric data	105
5.3.3	Identifying the Height Range of Interest	106
5.4	Estimation of uncertainties	108
5.5	Summary	109
6	Astrophysical neutrinos	111
6.1	The origin of cosmic rays	112
6.2	Production of astrophysical neutrinos	115
6.3	Neutrino astronomy	116

7	Search for neutrino point sources	119
7.0.1	Application to IceCube	122
7.1	Data sample	122
7.1.1	Data sets	123
7.1.2	Background model	124
7.1.3	Simulation	125
7.1.4	Binning	125
7.1.5	Low energy cut	125
7.2	Detector effects	125
7.2.1	Effective area	126
7.2.2	Point Spread Function	126
7.3	Point source injection	128
7.4	Inference framework	129
7.4.1	Injection cross-checks	130
7.5	Sensitivity	131
7.6	Results and summary	133
8	Artificial neural networks	135
8.1	Neural classifiers	137
8.1.1	Two layers	139
8.1.2	Multi-layer neural networks	140
8.2	Automatic differentiation	143
8.2.1	Forward mode	143
8.2.2	Reverse mode	144
8.2.3	Backpropagation	146
8.3	Regularisation	146
9	Search for tau neutrinos in IceCube	149
9.1	Double pulse signature	149
9.2	Convolutional neural networks	152
9.2.1	Other common operations	154
9.2.2	Training high-dimensional data	155
9.2.3	Application to the IceCube tau search	156
9.3	Training data	157
9.4	Neural network design	158
9.5	Results	161
9.6	Summary and outlook	163
10	Metropolis light transport	165
10.1	Path tracing	166
10.2	Path integration	167
10.3	The path probability distribution	168
10.3.1	Initial vertex	169
10.3.2	Intermediate vertex	169
10.3.3	Final vertex	170

10.3.4	Jump proposal	170
10.3.5	Coordinate transform	170
10.4	Implementation	171
10.4.1	Comparison of timing distribution	171
10.4.2	Comparison of angular distributions	173
10.5	Discussion	174
10.6	Summary	175
	Conclusion	177
	List of contributions	179
	Bibliography	181
	Index	207

List of Figures

1.1	The “periodic table” of particle physics.	22
1.2	Muon decay diagram.	22
1.3	Z boson exchange diagram.	23
1.4	Electron-positron annihilation.	23
1.5	The Z boson decay width.	24
1.6	β -decay diagram.	24
1.7	Flux spectrum of natural neutrinos over a wide energy scale.	28
1.8	An example of deep inelastic scattering.	35
1.9	Measurement of high-energy neutrino cross section.	35
1.10	Theoretical high-energy neutrino cross sections.	36
1.11	t -channel neutrino-lepton scattering.	36
1.12	s -channel neutrino-lepton scattering.	36
1.13	A muon in the MicroBooNE detector.	37
1.14	A simulated electromagnetic shower in Super-K.	38
2.1	Two confidence regions.	46
2.2	Two credible regions.	49
2.3	Random samples from a normal distribution.	51
2.4	An example of rejection sampling.	52
2.5	A point in the product space of \mathbb{R}^n and \mathbb{R}^{m-n} is mapped to a point in the space \mathbb{R}^m	57
3.1	The LSND anomalous event excess.	62
3.2	The MiniBooNE anomalous event excess.	62
3.3	Confidence and credible regions for a $3 + 1$ sterile neutrino model in various scenarios.	74
3.4	Confidence regions for a $3 + 2$ sterile neutrino model.	77
3.5	Comparison of 90% CL limits for muon flavor disappearance in IceCube, MINOS, CDHS, and MiniBooNE-SciBooNE.	77
3.6	Confidence regions for a $3 + 1$ sterile neutrino model with IceCube data.	81
3.7	Confidence regions for a $3 + 1$ sterile neutrino model with IceCube data, in $ U_{e4} $, $ U_{\mu4} $, and $ U_{\tau4} $	82

3.8	Comparison of the IceCube sterile neutrino limit with two extremes of θ_{34}	83
4.1	The IceCube detector.	86
4.2	A Digital Optical Module (DOM) in IceCube . . .	87
4.3	An illustration of various interaction signatures in IceCube.	88
4.4	Comparison of absorption and scattering parameters in IceCube ice models.	93
5.1	A cosmic ray induced atmospheric shower.	100
5.2	Calculation of atmospheric neutrinos fluxes per parent particle.	101
5.3	The AIRS instrument.	105
5.4	Production of neutrinos as a function of height in the atmosphere.	107
5.5	Comparison of muon and muon neutrino fluxes to the prediction from the US Standard atmosphere model.	108
5.6	The ratio of muon neutrino flux to muon anti-neutrino flux.	110
5.7	The effect of the AIRS systematic and statistical uncertainty.	110
6.1	Energy spectrum for cosmic rays.	112
6.2	Maps of the relative intensity for cosmic rays in the southern sky.	113
6.3	Previous IceCube hotspot point source search. . . .	116
7.1	The distribution of angular error in degrees for both the northern and southern sky.	126
7.2	Fractional flux distribution for both the north and the south sky.	127
7.3	Two examples of simulated data.	131
7.4	Sensitivity and expectations for four choices of templates.	132
7.5	The results of the point source search.	133
8.1	A set of examples with their matching classifications denoted in colour.	137
8.2	The sigmoid function.	138
8.3	A data set requiring two boundaries.	139
8.4	Example of a second layer acting on a data set. . . .	139
8.5	A multi-layer neural network.	140
8.6	Expression graph for forward mode automatic differentiation.	143

8.7	Expression graph for reverse mode automatic differentiation.	145
9.1	A double cascade in IceCube.	150
9.2	A simulated tau-neutrino interaction in IceCube.	151
9.3	A fully connected neural network layer.	152
9.4	A neural network layer with restricted connections.	153
9.5	A convolutional neural network layer.	153
9.6	The Rectified Linear Unit $\text{ReLU}(x)$	154
9.7	The horizontal geometry of IceCube.	157
9.8	The loss and accuracy as a function of epoch.	162
9.9	Expected number of events per year.	163
10.1	An example of a path of light.	169
10.2	The bispherical coordinate system used by the transition function.	171
10.3	The distribution of path lengths generated by the path sampler.	172
10.4	The distribution of arrival angles generated by the path sampler.	174

List of Tables

3.1	Data sets used in the sterile neutrino global fit analysis.	73
3.2	The χ^2 values for various sterile neutrino scenarios.	75
3.3	The best fit points for each sterile neutrino scenario.	75
3.4	The best fit points for a 3 + 1 sterile neutrino model with and without IceCube data.	80
3.5	Confidence intervals for a 3 + 1 sterile neutrino model with IceCube data.	81
7.1	The data sets used in the point source search.	124
7.2	The priors for the point source search.	129
7.3	The results of running the NPTF analysis on the measured data.	133
9.1	The design for the first stage of the convolutional neural network.	159
9.2	The design for the second stage of the convolutional neural network.	160
9.3	The hyper-parameters used during training.	162

Introduction

*It was a dark and stormy night on the 22nd of September, 2017...
...would that the Sun not risen that day — for the first time in six months.*

Instead, it was a dim and stormy dawn at the Amundsen-Scott South Pole Station, with temperatures at a brisk -78°C . At 8:54 AM local time, a neutrino interacted with the glacial ice deep below the station, producing a flash of visible light as a muon – produced in the interaction – travelled through the one billion ton IceCube south pole neutrino telescope.

This flash was unusually large, depositing around the same amount of light as a laser pointer when flashed for one second. Immediately, simulation and reconstruction algorithms began crunching the numbers, to find where this neutrino had come from. This estimated that the initiating neutrino has an energy of¹ $\gtrsim 180$ TeV, with an incoming direction pointing back to an origin at 77.43° right ascension, 5.72° declination – around the constellation of orion. With this information in hand, an alert was immediately sent out to astronomer’s telegram service to announce this new high energy event.

The Fermi Gamma Ray Telescope started analysing data both past and present from the origin location and discovered a blazar that was currently flaring. A blazar is a class of active galactic nuclei: galaxies with a large, highly active accretion disc surrounding their central black hole. While most in-falling matter is swallowed by the black hole, some is focussed outward by electromagnetic fields² forming a jet of relativistic particles. If this jet contains baryons – such as protons and neutrons – then collisions with interstellar gas can create high energy neutrinos that would be observable in IceCube. These blazars are quite volatile in nature, and their luminosity can vary significantly in time. When the blazar is at its brightest, it is deemed to be flaring. The coincidence between this high energy neutrino in IceCube, and a blazar flaring suggests that the blazar may be the origin of this neutrino. If true, it would be the first identification of a source of high energy neutrinos from outside our solar system.

Not every neutrino that IceCube detects is astrophysical in origin. The most common source of neutrinos is the atmosphere of the Earth. When cosmic rays strike the atmosphere, they produce showers of

¹ Approximately the same kinetic energy as a large jumping spider in flight [139].

² Another possible process is acceleration through a gravitational frame-dragging effect called the Penrose process.

exotic particles that subsequently decay into neutrinos. As neutrinos do not carry any information about how they were created, it is very difficult to determine if any single neutrino is astrophysical. However, we can distinguish collections of neutrinos at a statistical level. IceCube has observed that there is a flux of high energy neutrinos that cannot be explained by atmospheric neutrinos alone.

When we look at the night sky, we see a field of darkness with small bright points that mark the locations of stars. In physical terms, these are called point sources of light, as the stars are so far away that they can be approximated by mathematical points. When we look at the sky in neutrinos, we don't see any such point sources; almost as if we are looking up into a night sky that is just a uniform gray.

These neutrinos must be produced somewhere, and there is no known mechanism in the standard model of physics that can produce a purely diffuse distribution of neutrinos at these high energies and quantities³. It is likely, then, that these neutrinos are being produced by point sources, but those sources are too numerous and too faint to be distinguished individually. To give a comparison, the angular resolution of a human eye with good vision is around 0.02° . This means that if two stars are more than 0.02° apart on the sky, the eye can tell them apart. IceCube has an angular resolution of approximately 1° when looking at muon neutrinos. If two neutrino point sources are less than 1° apart⁴, then IceCube has great difficulty distinguishing them as two point sources instead of one.

Just as we can rarely tell, on an individual basis, if a neutrino is astrophysical in origin, but can identify a host of neutrinos as astrophysical by their energy spectrum; we may be able to discern the properties of the population of neutrino point sources statistically. This requires the creation of a statistical test which can give the probability that a distribution of neutrinos follows that expected if the neutrinos are created by a population of point sources. The application of such a test forms one of the main results of this thesis, as presented in chapter 7. This method can test a spatial distribution for the population, and three such distributions within our galaxy were considered. While attention has typically focussed on neutrino sources outside our galaxy, it is now being turned sources closer to home — specifically supernova remnants and pulsar wind nebulae [125].

Along with direction and energy, neutrinos carry one other kind of information: the type of neutrino; either electron neutrino, muon neutrino, or tau neutrino. The proportion of these three kinds of neutrino could give more information about the physical processes that created them. Muon neutrinos are easy to identify, they create long straight lines of light in IceCube. Electron and tau neutrinos both produce balls of light, and so it is nearly impossible to tell

³ The night sky looks very different if you look at it in microwave energies instead of visible light. In microwaves, the sky is mostly uniform with some distortions. This is called the cosmic microwave background, which was produced by the big bang. There is an analogous cosmic neutrino background, but it is around 15 orders of magnitude less energetic than the neutrinos being discussed here and cannot be seen by IceCube.

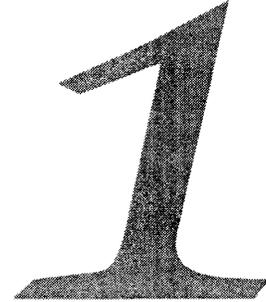
⁴ Around twice the width of the moon in the sky, as viewed from Earth.

the difference between them. This is a difficult problem in pattern recognition, but neural networks are currently being developed in computer science to solve a similar problem: pattern recognition in photographs. The application of a convolutional neural network to classifying electron and tau neutrinos is presented in chapter 9.

These analysis and reconstruction tools are reliant on simulation to validate their techniques and estimate systematic errors. The most computationally consuming simulation task in IceCube is the propagation of light inside the detector. This is currently solved using path tracing, where photons are individually marched through the ice until they are either extinguished or detected. However, IceCube is so large that most photons do not survive the journey to a detector element before they are lost to the ice and absorbed. In chapter 10, a new approach is presented, where the entire light path is specified and constrained up-front in the form of a path integral. This path integral is then sampled using statistical inference algorithms.

Understanding the flux of atmospheric neutrinos is important for both analyses where these neutrinos form a background, and those where they are the signal. Particle physics analyses probe the laws of physics by measuring differences between atmospheric neutrinos that travel various paths through the center of the Earth. These paths are determined by the location of the shower on the Earth's surface, and the behaviour of these showers depends, in turn, on the local properties of the atmosphere. Chapter 5 contains a calculation of the effect of global temperature variations on atmospheric neutrino flux calculations.

The sterile neutrino search is one such example of a particle physics result from IceCube. Sterile neutrinos are a hypothetical fourth kind of neutrino which do not interact with matter. They are motivated by anomalies in various neutrino experiments in many different energy regimes. However, some experiments see no evidence of their existence. Making sense of these conflicting results requires a comprehensive analysis that tests a single model to all the available data simultaneously. This global fit is presented in chapter 3.



Neutrinos in the standard model

From an experimentalist's point of view, the neutrino's lack of electromagnetic charge is both its most interesting, and most difficult quality. Within the standard model of particle physics, the neutrinos are the only electrically uncharged fermions — the other uncharged particles: the photon, gluon, and Higgs, are all bosons as shown in figure 1.1. Instead, the neutrino only interacts with other matter through the weak nuclear force and gravity.

The uncharged neutrino is an ideal candidate to probe the behaviour of the weak force without the interference of electromagnetism, which is factor of $\sim 10^{13}$ stronger. This large disparity in the relative interaction strength implies that the interaction probability of the neutrino is very low compared to charged particles. Hence, the neutrino poses a difficult experimental challenge in collecting enough data to make strong scientific claims.

1.1 The physics of neutrinos

The weak nuclear force is mediated by the W^\pm and Z bosons. These *gauge* – or force carrying – bosons participate as *virtual particles*. A virtual particle exists only temporarily, to facilitate the interaction, and does not have to obey the usual energy-momentum condition⁵ $E = p$. This allows neutrino interactions to happen at energies too low for a free W^\pm or Z boson – that have masses of 80 and 91 GeV – to be created. The virtual nature of this exchange boson introduces a term into the scattering cross section that is proportional to [200,

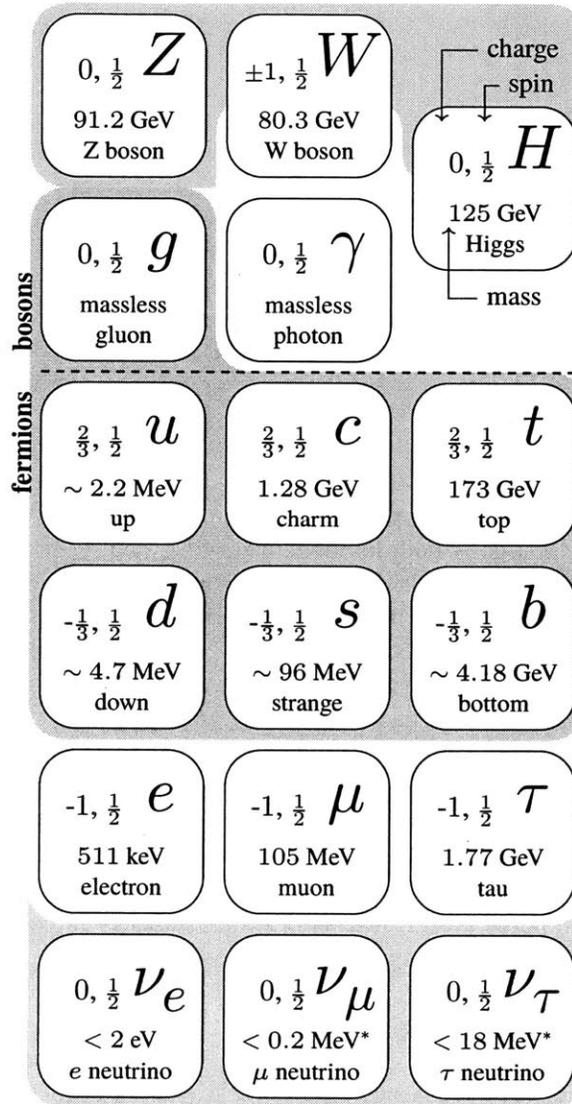
⁵ In natural units, where $c = 1$.

$$\left| \frac{1}{p^2 - M_{W/Z}^2} \right|^2, \quad (1.1)$$

in a leading order approximation, where p is the 4-momentum of the boson, and $M_{W/Z}$ is the mass of either the W^\pm or Z .

Figure 1.1:

The “periodic table” of particle physics. Properties from Patrignani et al. [198]. * μ and τ neutrino mass limits are from direct measurements, indirect limits from neutrino oscillation would be similar to the electron neutrino.



This squared 4-momentum goes by different names depending on the kind of interaction. When the boson transfers momentum between particles in a *t-channel process*, it is often denoted as $p^2 = t$ and called the “momentum transfer”. Neutrino scattering experiments are always⁶ performed with a fixed target, so that the particle that the neutrino scatters off – such as an electron or nucleus – is at rest. In this situation, the momentum transfer scales as

$$\sqrt{t} \approx \sqrt{2M_T \Delta E}, \tag{1.2}$$

where M_T is the mass of the target particle and ΔE is the energy transferred to the fixed target. When two particles annihilate to pro-

⁶ Neutrino colliders are possible in theory; however, the intensity of neutrino beams are too low for this to be practical.

duce a boson in the complementary *s-channel process*, the squared 4-momentum is denoted $p^2 = s$ and is called the “center of mass energy”. There is a similar *s-channel* relation for fixed targets:

$$\sqrt{s} \approx \sqrt{2M_T E_\nu}, \quad (1.3)$$

where E_ν is the energy of the incoming neutrino. This is in contrast to collider experiments, where the center of mass energy scales directly with the particle energy, not as the square root.

When the energy scale of the interaction is much less than $M_{W/Z}$, equation 1.1 is $\approx M_{W/Z}^{-4}$ and the interaction is highly suppressed. Unlike in electromagnetic interactions – where the cross section decreases with increasing interaction energy – the weak interaction cross section increases with energy. When the interaction energy is comparable to $M_{W/Z}$, the weak interaction is no longer “weak” relative to electromagnetism and beyond the electroweak unification scale of $\sim 10^{2.5}$ GeV the weak and electromagnetic forces are unified as a single force [198]. Examples of weak interactions are shown in figures 1.2 and 1.3. In figure 1.2, a muon decays, emitting a muon neutrino and a W^- boson. The W^- then decays into an electron and an electron anti-neutrino. This kind of interaction is called “charged current”, as the W^- carries the negative charge of the muon to the produced electron.

In contrast to this, interactions involving the uncharged Z boson are called “neutral current”. For this reason, neutral current interactions do not usually⁷ describe decays of other particles, but rather scattering processes such as the example shown in figure 1.3.

The Z boson itself is of interest for neutrino physics. When a Z boson decays, it will create a pair of fermions⁸ that are anti-particles of each other, as shown in figure 1.4. The lifetime of the Z is related to the total phase space of the decay products, of which the number of neutrinos that couple to the Z is one contributing factor. The lifetime of such a heavy particle is far too short to time directly, but can be accessed through production of Z bosons in a collider. When the center of mass energy of the collision is close to M_Z , equation 1.1 approaches infinity; an artefact of the leading order approximation used. Higher order corrections can be incorporated using the *Breit-Wigner parameterisation*⁹ [200]:

$$\left| \frac{1}{p^2 - M_{W/Z}^2} \right|^2 \rightarrow \frac{1}{(p^2 - M_Z^2)^2 + M_Z^2 \Gamma_Z^2}. \quad (1.4)$$

Thus, at $p^2 = M_Z^2$, the cross section undergoes a resonant enhancement. The width of this resonance is proportional to the *decay width*

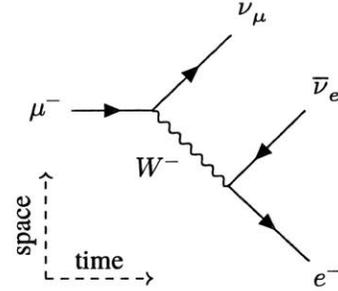


Figure 1.2:

A muon decay is an example of a neutrino interaction involving a W boson.

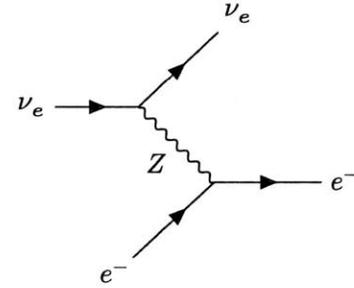


Figure 1.3:

A neutrino scattering off an electron is an example of Z boson exchange.

⁷ Neutral particles – such as the composite positronium (e^+e^-) [202] – can decay through a neutral current.

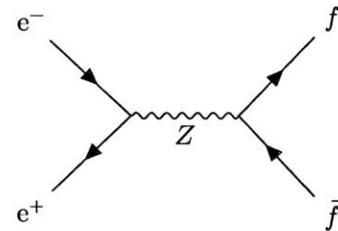


Figure 1.4:

A colliding electron and positron annihilate and produce a Z boson, which decays into a fermion-anti-fermion pair marked as $f\bar{f}$.

⁸ Only the top quark is excluded from the list of Z decay products, as its mass is larger than $M_Z/2$.

⁹ This is not, in itself, an exact treatment, and applies when the resonance is narrow with no other resonances nearby [198].

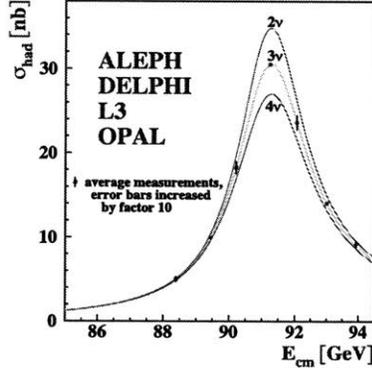


Figure 1.5:

Measurements by experiments at the LEP collider of the hadron production cross-section. The resonant enhancement is clearly visible near the Z boson mass. The continuous lines show the predicted cross-section for two, three, and four standard model neutrinos. Figure included from Collaboration et al. [67].

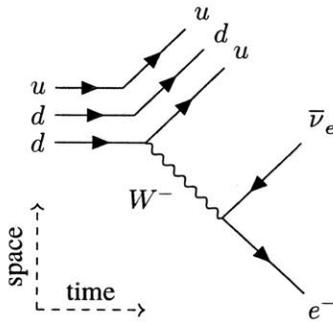


Figure 1.6:

In β -decay, a W^- boson converts a down quark into an up quark.

parameter Γ_Z , which in turn is inversely proportional to the decay lifetime. Using this connection, physicists at the LEP collider determined that the number of neutrinos that participate in the Z decay is $N_\nu = 2.994 \pm 0.011$ [190], in exact agreement with the standard model prediction of one neutrino for each generation of particles. A measurement of the width is shown in figure 1.5.

1.1.1 Flavour physics

Just as the charged nature of the W^\pm boson allows it to change charged leptons into neutrinos, it also enables the conversion of up-type quarks into down-type quarks. This is precisely what occurs in β -decay, shown in figure 1.6, where a down quark in a neutron (composition udd) is converted to an up quark, and thus changing the neutron into a proton (composition uud).

The W^\pm boson couples to the quarks in the interaction basis of the weak force, also known as the *flavour basis*. Each quark has a respective state in this basis denoted by primes, e.g. u' for an up quark. In this basis, the W^\pm boson allows the following transitions:

$$u' \leftrightarrow d', \quad (1.5)$$

$$c' \leftrightarrow s', \quad (1.6)$$

$$t' \leftrightarrow b'. \quad (1.7)$$

In this basis, *flavour* is said to be conserved, as quarks can not transition across generations.

Particles propagate through free space in the *mass basis*, as it is their mass, not flavour, that governs their kinetics. These mass states are denoted without primes, e.g. u for an up quark. The two bases can be related through a coordinate transformation matrix

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \mathbf{V}_d \begin{bmatrix} u \\ s \\ d \end{bmatrix}, \quad (1.8)$$

with a similar matrix \mathbf{V}_u for the up-type quarks.

The effect of the weak interaction in the mass basis can be viewed by taking the products of these matrices:

$$\mathbf{V}_u \mathbf{V}_d = \mathbf{V}_{\text{CKM}} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}. \quad (1.9)$$

The unitary 3×3 matrix \mathbf{V}_{CKM} is called the ‘‘Cabibbo-Kobayashi-Maskawa mixing matrix’’. As the mass and flavour bases do not perfectly align, this matrix is not entirely diagonal, which allows the W^\pm

boson to change quarks between generations in the mass basis — so called "flavour changing" interactions. Thus, hadrons like the kaon¹⁰ can decay through emission of a muon and muon anti-neutrino, mediated by a W^- boson. The elements of \mathbf{V} can be determined by measuring these flavour changing interactions. On the other hand, elements of \mathbf{V}_u and \mathbf{V}_d cannot be directly observed, and so there is some degree of freedom in parameterising them. The convention is to set \mathbf{V}_u to be diagonal, so that $\mathbf{V}_d = \mathbf{V}$ — in essence, flavour physics is considered a property of the down quarks.

¹⁰ Denoted as K^- with composition $\bar{u}s$.

1.1.2 Neutrino oscillation

One might suspect that an analogous process might happen with the leptons. Like with the quarks, we will assume that flavour physics is a property of the neutrinos. A similar transformation matrix is then introduced to connect the mass basis of the neutrinos — with states ν_1, ν_2 and ν_3 — to the flavour basis with states ν_e, ν_μ and ν_τ :

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = \mathbf{U}_{\text{PMNS}} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix}. \quad (1.10)$$

The unitary 3×3 matrix \mathbf{U}_{PMNS} is called the "Pontecorvo-Maki-Nakagawa-Sakata mixing matrix".

To see what effect this might have on neutrino physics, consider a beam of neutrinos prepared in a flavour state $|\nu_i\rangle$. To understand how these particles travel freely after production, the state must be transformed into the mass basis:

$$|\nu'_i\rangle = \mathbf{U}_{\text{PMNS}}^{-1} |\nu_i\rangle. \quad (1.11)$$

In the mass basis, their flight through free space is governed by the unitary operator $e^{-i(\hat{E}t - \hat{\mathbf{p}} \cdot \vec{x})/\hbar}$.

$$|\nu'_i(t, \vec{x})\rangle = e^{-i(\hat{E}t - \hat{\mathbf{p}} \cdot \vec{x})/\hbar} |\nu'_i\rangle \quad (1.12)$$

$$= e^{-i(\hat{E}t - \hat{\mathbf{p}} \cdot \vec{x})/\hbar} \mathbf{U}_{\text{PMNS}}^{-1} |\nu_i\rangle. \quad (1.13)$$

As interactions occur in the flavour basis, the state must be transformed back to the flavour basis in order to compute the probability of finding the neutrino in state $|\nu_f\rangle$:

$$|\langle \nu_f | \nu_i(t, \vec{x}) \rangle|^2 = |\langle \nu_f | \mathbf{U}_{\text{PMNS}} |\nu'_i(t, \vec{x}) \rangle|^2 \quad (1.14)$$

$$P(\nu_i \rightarrow \nu_f, t, \vec{x}) = \left| \langle \nu_f | \mathbf{U}_{\text{PMNS}} e^{-i(\hat{E}t - \hat{\mathbf{p}} \cdot \vec{x})/\hbar} \mathbf{U}_{\text{PMNS}}^{-1} |\nu_i\rangle \right|^2. \quad (1.15)$$

Applying the unitary evolution operator gives

$$P(\nu_i \rightarrow \nu_f, t, \vec{x}) = \left| \sum_{\alpha, \beta \in \{e, \mu, \tau\}} \sum_{i \in \{1, 2, 3\}} [\nu_f]_\beta U_{\beta i} e^{-i(E_i t - \vec{p}_i \cdot \vec{x})/\hbar} U_{\alpha i}^* [\nu_i]_\alpha \right|^2, \quad (1.16)$$

where E_i and p_i are the energy and momentum of mass state i .

The Super-Kamiokande experiment confirmed that neutrinos do oscillate, by measuring the flux of muon neutrinos coming from above and below the detector [74]. These neutrinos were generated in the Earth's atmosphere, which implies neutrinos coming from below the detector must have travelled the diameter of the Earth before being detected. Super-Kamiokande observed a significant decrease in the flux coming from below the detector, compared to the no-oscillation prediction, as the muon neutrinos oscillated into other flavour states over this large distance.

From these oscillation equations one important fact is already evident: if all the neutrino mass states are massless, or if they all have the same non-zero mass, then $E_1 = E_2 = E_3$, also $p_1 = p_2 = p_3$, and equation 1.15 can be simplified to

$$P(\nu_i \rightarrow \nu_f, t, \vec{x}) = |\langle \nu_f | \nu_i \rangle|^2, \quad (1.17)$$

showing that there is no observable difference in the flavour composition of the state as a function of time. Thus, for any dynamic evolution in the flavour composition, at least one or more of the neutrino mass states must have a distinctly different mass to the others; and, in consequence, at least one neutrino must have mass. This is in contrast to the minimal standard model prediction that all neutrinos are massless, that there should be no neutrino oscillation, and flavour should be conserved in the lepton sector.

In the regime where the energy and momentum of the neutrinos is much greater than their mass, $E \gg mc^2$ and the momentum can be Taylor expanded to first order:

$$p_i = \sqrt{\frac{E_i^2}{c^2} - m_i^2 c^2} \quad (1.18)$$

$$\approx \frac{E_i}{c} \left(1 - \frac{1}{2} \frac{m_i^2 c^4}{E_i^2} \right). \quad (1.19)$$

In addition, if we only consider positions $\vec{x} = L\hat{p}$ in the direction of the neutrino momentum, then $\vec{p}_i \cdot \vec{x} = p_i L$. In the $E \gg mc^2$ regime, the neutrino is ultra-relativistic, so $L \approx ct$, and

$$p_i L = E_i t - \frac{1}{2} \frac{m_i^2 c^3 L}{E_i}. \quad (1.20)$$

Thus, equation 1.16 can be written as

$$P(\nu_i \rightarrow \nu_f, t, \vec{x}) \approx \left| \sum_{\alpha, \beta \in \{e, \mu, \tau\}} \sum_{i \in \{1, 2, 3\}} [\nu_f]_{\beta} U_{\beta i} \exp\left(-\frac{i}{2\hbar} \frac{m_i^2 c^3 x}{E_i}\right) U_{\alpha i}^* [\nu_i]_{\alpha} \right|^2. \quad (1.21)$$

For a pure state that begins in flavour α , the probability to measure it in a pure flavour state of β can be written, by expanding the modulus, and making the equal energy approximation $E \approx E_i$ as [114]:

$$P_{\alpha \rightarrow \beta}(L) \approx \delta_{\alpha\beta} - 4 \sum_{i>j} \text{Re} [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin^2 \left(\frac{\Delta m_{ij}^2 c^3 L}{4\hbar E} \right) \quad (1.22)$$

$$+ 2 \sum_{i>j} \text{Im} [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin \left(\frac{\Delta m_{ij}^2 c^3 L}{2\hbar E} \right), \quad (1.23)$$

where $\Delta m_{ij}^2 = m_i^2 - m_j^2$ is the *squared mass difference*, also called the “squared mass splitting”. Due to the sinusoidal terms in this equation, this flavour changing effect is called “neutrino oscillation”.

The equations for anti-neutrino oscillation can be found by making the replacement $\mathbf{U}_{\text{PMNS}} \rightarrow \mathbf{U}_{\text{PMNS}}^\dagger$ in equation 1.15. Indeed, the difference between neutrino and anti-neutrino oscillation probabilities is

$$P(\nu_\alpha \rightarrow \nu_\beta) - P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta) = 4 \sum_{i>j} \text{Im} [U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*] \sin \left(\frac{\Delta m_{ij}^2 c^3 L}{4\hbar E} \right). \quad (1.24)$$

This second sum over the imaginary values thus determines the amount of CP violation present in neutrino oscillation.

It is instructive to consider a two neutrino oscillation formalism. In this case, there are only two flavour states and two mass states. The mixing matrix can be parameterised as a simple rotation matrix:

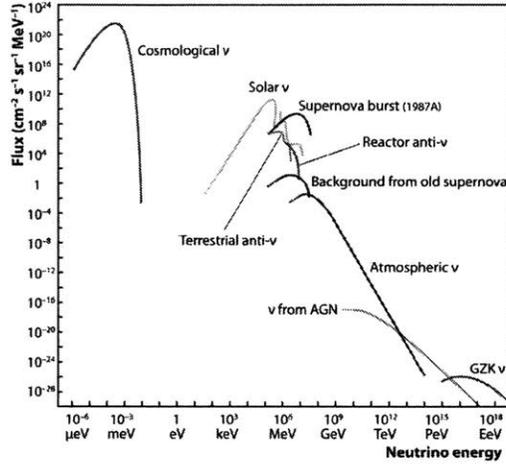
$$\mathbf{U}_{\text{PMNS}} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. \quad (1.25)$$

Equation 1.23 becomes:

$$P_{\alpha \rightarrow \beta}(L) \approx \delta_{\alpha\beta} - (2\delta_{\alpha\beta} - 1) \sin^2(2\theta) \sin^2 \left(\frac{\Delta m_{ij}^2 c^3 L}{4\hbar E} \right). \quad (1.26)$$

Figure 1.7:

Flux spectrum of natural neutrinos over a wide energy scale. The flux from nuclear reactors is shown for scale. Figure included from *Astroparticle Physics Roadmap - Phase I* [35].



When working in units of eV^2 for the squared mass splitting, GeV for the neutrino energy, and km for the distance, the physical constants in this equation can be simplified to

$$P_{\alpha \rightarrow \beta}(L) \approx \delta_{\alpha \beta} - (2\delta_{\alpha \beta} - 1) \sin^2(2\theta) \sin^2\left(\left[1.27 \frac{\text{GeV}}{eV^2 \text{ km}}\right] \frac{\Delta m_{ij}^2 L}{E}\right). \quad (1.27)$$

Hidden in this derivation is the assumption that the neutrino wavefunction is a plane wave. In general, this is not true, and neutrinos will be generated as wave packets. A thorough treatment of neutrino oscillation is outside the scope of this introduction; however, for any experiment with neutrino detectors and sources based on Earth, a plane wave approximation is sufficient [152]. At larger distances, especially those at play with astrophysical neutrinos, the wave packets of each individual mass state can become well separated in space, due to their different momenta. At this point, the mass states have *decohered*, and oscillation ceases. Instead, a detector would measure isolated mass states, which would result in a fixed fraction of the flavour states depending on the relevant column of U_{PMNS} .

1.2 Production and detection

Historically, the most important source of neutrinos was through the decay of radioactive nuclei. This was in part due to the ease of access to large quantities of unstable elements. However, the energy of these neutrinos are limited to less than ~ 10 MeV; so, today neutrinos produced using particle accelerators play an equally important role.

In addition, the universe is awash with naturally-produced neutrinos from a large variety of sources as illustrated in figure 1.7. These

neutrinos can provide us with information about naturally occurring phenomena such as our Sun, distant galaxies, the big bang, and the center of the Earth. Natural sources often reach energy scales, length scales, and interaction environments that exceed human capabilities, making them valuable for particle physics.

The mechanisms by which natural and artificial neutrinos are produced and detected is best organised by three broad energy scales of interactions on nuclei: low energies, below a few hundreds of MeV neutrinos usually initiate nuclear reactions; medium energies, from hundreds of MeV to a few tens of GeV neutrinos can eject nucleons from their parent nuclei; and high energies, from tens of GeV and beyond, neutrinos can resolve individual quarks, whose disturbance produces complex hadronic showers. This is followed by a discussion on interactions of neutrinos with leptons at all energy scales.

1.2.1 Low energy

The lowest energy source of neutrinos is the *neutrino sources|cosmic background*. Similar to the cosmic microwave background, these neutrinos were produced during the big bang. After billions of years of cooling¹¹, their average energy is hypothesised to be $\sim 10^{-4}$ eV [93]. Thus, cosmic background neutrinos have the lowest energy of neutrinos from any source. Although they are the most abundant of all natural sources of neutrinos, their very low energy make them the most difficult of all neutrinos to detect — beyond the reach of current particle physics technology.

Nuclear fission and decay

The chief radioactive process for creating neutrinos is called “beta (β) decay” — named after the at-that-time unknown electron produced in the decay. Here, a neutron in an unstable nucleus lowers its energy by decaying into a proton. This must also conserve charge through the creation of a negatively charged particle, of which the electron is the only candidate of small enough mass. Originally, it was thought that only the β particle and nucleus were involved in β -decay, as they were the only participants observed. It was soon noticed, however, that the energy of the electron was not equal to the mass lost by the nucleus.

In 1930, to rescue conservation of energy, Pauli proposed [199] a new particle to carry away the missing energy in the reaction¹². Later, Amaldi and Fermi coined the name “neutrino” for this particle [25], meaning “little neutral one”. With the discovery of the neutron, the full picture of β decay was revealed:



¹¹ As the universe expands, it also stretches out the wave function of particles, increasing their wavelength and thus reducing their energy.

¹² In addition, this particle would also allow momentum and angular momentum to be conserved.

In 1934, Frédéric and Irène Joliot-Curie discovered that the positron could also be produced in a nuclear reaction [151]:



Both reactions became known as β^- and β^+ decay, after the charge of the electron or positron produced.

This was some of the first evidence of a more general principle in particle physics, where new reactions can be discovered by applying a *crossing symmetry*. This symmetry allows a particle on one side of the reaction to be replaced by its CP -conjugate on the other side of the reaction. Thus, from β^+ decay, a crossing symmetry tells us that there must also exist a reaction which requires an anti-neutrino:



which is called “inverse β -decay”.

The experimental confirmation of the neutrino would not come until 1956. Cowan and Reines used inverse β -decay to detect electron anti-neutrinos being generated from the Hanford and Savannah reactors [81]. The neutrino’s low cross section necessitated the use of reactors, which continuously produce unstable nuclei as part of the fission process. These nuclei undergo repeated α and β decays¹³ until they reach a stable nuclear configuration. The emission of a neutrino at each of these β -decays creates an incredibly intense source of neutrinos.

In the Cowan and Reines experiment, a positron is produced when a neutrino from the reactor interacts with the detector. This positron quickly finds an electron and annihilates to produce two distinctive gamma rays of 511 keV energy¹⁴. This converts a problem of finding neutrinos, to one of finding gamma rays. Cowan and Reines placed scintillators inside their detector; when a gamma ray Compton scatters a recoiling electron can fly through the scintillator, producing visible light. The visible light was detected and recorded electronically using photo-multiplier tubes, which are discussed further in chapter 4.

While reactors provide an abundance of neutrinos, the fission products tend to be highly neutron rich [214]. Nuclei that have many neutrons are poorly understood, both due to the difficulty in creating them in a controlled environment, and their highly complicated decay process¹⁵.

For this reason, some neutrino experiments use pure, electron-capture isotopes as neutrino sources for calibration. Electron capture is another crossing symmetry of β^+ decay:



¹³ The α decay involves the emission of a Helium nucleus – two protons and two neutrons – from the nucleus, but does not create neutrinos.

¹⁴ Additionally, by using water as a neutrino target, hydrogen nuclei would be converted to free neutrons that would be captured by cadmium, creating a second distinctive gamma ray.

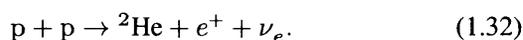
¹⁵ This is known as the “pandemonium effect” [128, 214], where a high density of nuclear energy levels near the continuum generates a wide, soft spectrum of low energy γ -rays that are difficult to accurately measure.

This reaction occurs when a nucleus has too many protons, and the conversion of one proton to a neutron is energetically favourable. To conserve charge, an electron is captured from the bound electrons of the nuclei's atom. There are only two products in this reaction, so conservation of energy and momentum demands that the neutrino be mono-energetic, making it an ideal source for energy calibration.

Intense sources of natural radioactivity also create neutrinos. The most famous example of these are the $\mathcal{O}(10\text{MeV})$ neutrinos created by the 1987A supernova. Most of the energy in a supernova is converted into neutrinos [148]. This created a sudden increase in the rate of measured neutrinos over a 13 second time period in various detectors around the world [24, 140], which was later correlated to the optical sighting of the 1987A supernova in the Large Magellanic Cloud. The Large Magellanic Cloud is a dwarf galaxy that is technically separate from our own, but still much closer than the next closest major galaxy, Andromeda. It is also worth briefly mentioning that the radioactive decay of heavy elements in the Earth's mantle will also produce neutrinos called "geo-neutrinos".

Nuclear fusion

The next most abundant source of natural neutrinos – after the cosmic neutrino background – is the Sun. Solar neutrinos are produced as one part of the Sun's major fusion reaction¹⁶:



Fusion occurs at the center of the Sun, where the pressure is greatest. The gamma rays produced by other mechanisms of the multi-process fusion reaction are quickly absorbed by the surrounding gas, and their energy is converted to heat. This obscures the fusion reaction gamma ray energy spectrum from observers on Earth, preventing the validation of solar models.

Neutrinos, on the other hand, travel unhindered through the Sun, allowing a potential window into its inner workings. Davis set out to use neutrinos for exactly this purpose in the famous Homestake experiment. In this experiment, instead of measuring the low mass products of nuclear reactions – such as electrons and gamma rays – the radioactive isotopes themselves can be measured. Davis used a crossing symmetry of β^- -decay called neutrino capture,



to convert ${}^{37}\text{Cl}$ to ${}^{37}\text{Ar}$ — a reaction that is not otherwise energetically allowed. By bubbling helium through a tank containing chlorinated fluid, the resultant radioactive argon was collected and

¹⁶ While this is the primary neutrino producing reaction in the Sun, secondary reactions such as electron capture on ${}^7\text{Be}$ and ${}^8\text{B}$ β -decay also produce neutrinos in smaller quantities.

concentrated in a cooling trap where the subsequent nuclear decay could be counted.

His experiment measured a flux of electron neutrinos that was one third of the standard solar model predictions [130], a discrepancy known as the *solar neutrino problem*. Davis had unintentionally witnessed the previously unknown particle physics effect of neutrino oscillation, which implied that neutrinos have mass.

1.2.2 Medium energy

When first discovered, it was unknown if the neutrino was of a unique type, or if there were many kinds of neutrino. A hint came in the form of muon decays, which were observed to produce an electron with a continuous energy spectrum. This led physicists to hypothesise that – like β decay – this decay produces three products. As the two additional products could not be observed, they must be neutrinos, and it was thought that they might be two different kind of neutrinos¹⁷:

¹⁷ If there were only one kind of neutrino, $\mu^\pm \rightarrow e^\pm + \gamma$ decays would be common [98].

$$\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e, \quad (1.34)$$

$$\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e. \quad (1.35)$$

Elastic and quasi-elastic scattering

This hypothesis was confirmed by Lederman et al. in 1962, by producing muon neutrinos using a particle accelerator [83]. Protons were accelerated into a beryllium target, where hadronic interactions during the collision produced pions, which subsequently decayed into muons and muon neutrinos.

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu, \quad (1.36)$$

$$\pi^+ \rightarrow \mu^+ + \nu_\mu. \quad (1.37)$$

These neutrinos were aimed into a 13 meter thick steel shield. Their detector was sat behind the shield, where it was protected from other beam related backgrounds.

At neutrino energies of order a few to tens of GeV, the neutrino is still able to produce conversions between protons and neutrons:

$$\nu_l + n \rightarrow p + l^- \quad \bar{\nu}_l + p \rightarrow n + l^+. \quad (1.38)$$

Most of the neutrino's energy goes into the kinetic energy of the recoiling nucleon, hence these are called “Charged Current Quasi-Elastic” (CCQE) interactions. Although the probability is small, quark mixing allows anti-neutrino CCQE interactions to promote an up quark to a strange quark, resulting in a nucleon being converted to a Λ^0 , Σ^- or Σ^0 hyperon [106].

Lederman's muon neutrinos interacted with the detector through these charged current exchanges, producing a clear muon. The ratio of electrons to muons seen in the detector was small, leading to the conclusion that there is in fact distinct kinds of neutrino with associated leptons.

The neutral current equivalent to charged current quasi-elastic scattering is called "neutral current elastic scattering". In this interaction, the neutrino scatters off the nucleon, transferring some momentum to it.

$$\nu_l/\bar{\nu}_l + p \rightarrow \nu_l/\bar{\nu}_l + p, \quad \nu_l/\bar{\nu}_l + n \rightarrow \nu_l/\bar{\nu}_l + n. \quad (1.39)$$

Both neutral current and charged current scattering can excite a nucleon into a higher energy level called a "baryon resonance" [106]. This excited state decays back to a proton or neutron through the emission of a pion. This allows charged current quasi-elastic scattering of (anti-)neutrinos off (neutrons) protons:

$$\nu_l + p \rightarrow p + l^- + \pi^+ \quad \bar{\nu}_l + n \rightarrow n + l^+ + \pi^- \quad (1.40)$$

In addition, neutral current interactions can also change a proton to a neutron or vice versa:

$$\nu_l/\bar{\nu}_l + p \rightarrow \nu_l/\bar{\nu}_l + n + \pi^+ \quad \nu_l/\bar{\nu}_l + n \rightarrow \nu_l/\bar{\nu}_l + p + \pi^- \quad (1.41)$$

It also possible for more than one pion to be produced, or for kaons to be produced [106].

Particle accelerators

Lederman's experiment was the first to use particle accelerators for the purpose of generating neutrinos [161]. Particle accelerators are a necessity when muon neutrinos are desired, as the creation of their partner muons require 105 MeV of energy — far above the $\mathcal{O}(1 \text{ MeV})$ of available energy in radio-isotope decay.

Modern accelerator based neutrino sources operate in much the same manner to Lederman's experiment, with one addition: a *neutrino horn*. This large electromagnet is placed behind the beryllium target, so that the resulting π^+ and π^- mesons pass through it. The horn can operate in one of two polarities, depending on the direction of the current passing through it. Its purpose is to create an intense magnetic field that will focus one kind of pion (e.g. π^-) while defocusing the other kind (e.g. π^+). As the π^- produces muon anti-neutrinos, and the π^+ produces muon neutrinos, the horn allows the generation of a beam that preferentially contains either neutrinos or anti-neutrinos¹⁸.

¹⁸ The muon produced in the decay will itself decay to produce neutrinos; however, the kinematics of the muon decay naturally defocusses the muon's decay products.

Although it may first appear that the muon neutrino produced in the pion decay should be mono-energetic, this is only true in the pion's rest frame. The complicated hadronic interactions with the beryllium target will cause the pions to fall on a spectrum of kinetic energies, and the resulting muon neutrinos will inherit part of the pion's kinetic energy. Precisely understanding this spectrum is a difficult prospect, but can be side-stepped by allowing the pions to *decay at rest*. Rather than allowing the pion to *decay in flight*, this technique forgoes the neutrino horn to enlarge the target such that the charged pions slow down and come to a stop before exiting the target material. The π^- mesons are captured by the nuclei of the material, while the π^+ mesons decay to produce a mono-energetic neutrino.

Accelerators may also be used to drive nuclear reactions. By producing a specific low-mass isotope, they avoid the complicated decay chains created in nuclear reactors.

Atmospheric neutrinos

Atmospheric neutrinos are generated when cosmic rays – typically protons – strike the Earth's atmosphere. This is analogous to an accelerator source, where the cosmic rays take the place of the beam and the atmosphere is the target. Like in the accelerator, hadronic interactions between the impinging cosmic ray and the atmosphere produce a large number of unstable mesons which subsequently decay into muon neutrinos and muons – that in turn decay into additional muon neutrinos and electron neutrinos. Atmospheric neutrinos are a valuable source of medium to high energy neutrinos, and their production is discussed further in chapter 5. These neutrinos were instrumental in confirming neutrino oscillation [74], and their role in searching for new kinds of neutrinos is expanded upon in chapter 3.

1.2.3 High energy

The highest energy neutrinos are astrophysical — produced in sources that are outside of our solar system. Astrophysical neutrinos with energies greater than 10 TeV have been observed in IceCube, but their source is as yet unknown. The study of these neutrinos, and the objects that might be creating them, is further discussed in chapter 6. The experimental details of detecting these neutrinos is covered in chapter 4.

Deep inelastic scattering

Beyond a few tens of GeV, the neutrino has sufficient energy to resolve and interact exclusively with a single quark inside the nucleus.

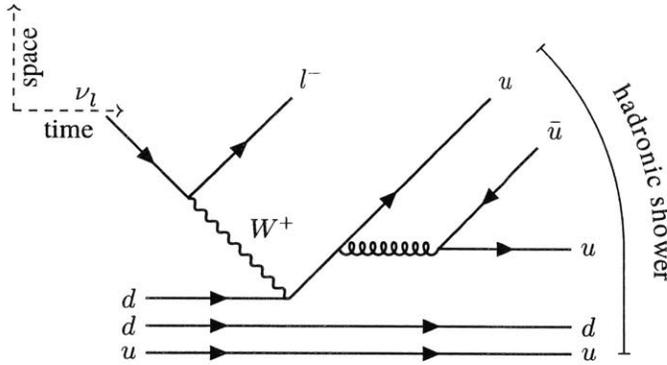


Figure 1.8:
An example of deep inelastic scattering. After a charged current interaction delivers momentum to a down quark, the quark recoils, radiating a gluon to create more quarks. This process forms a hadronic shower.

In these *deep inelastic* interactions, the quark recoils outside the nucleus. As nature abhors a naked quark, hadronic interactions quickly pull new quarks out of the vacuum as shown in figure 1.8. These new quarks themselves can diverge, necessitating the creation of more quarks to keep the final state of the system colourless. This cascade of hadrons is called a “hadronic shower”, and the energy needed to create these particles is the reason the process is deemed inelastic.

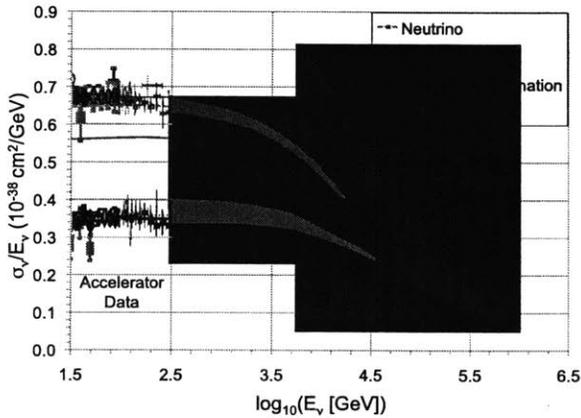


Figure 1.9:
A measurement of the deep inelastic scattering cross section as a function of the neutrino energy with the IceCube experiment. The standard model predictions for $\nu/\bar{\nu}$ are shown in blue and green respectively. As IceCube cannot distinguish neutrinos from anti-neutrinos, the red line shows the appropriately weighted combined standard model prediction. The black line shows the measured result. The yellow section shows lower energy data from accelerator experiments. Figure included from Collaboration et al. [72].

For historical reasons, deep inelastic scattering is modelled as a neutrino interacting with a *parton*: an abstraction coined by Feynman before the discovery of quarks. The deep inelastic cross section at ultra-high energies is shown in figure 1.9. In the yellow “Accelerator Data” region, the cross section for deep inelastic scattering is several orders of magnitude higher than the neutrino-lepton scattering cross section [183]. As the neutrino energy E_ν approaches 6.3 PeV, the momentum transfer and center of mass energies approach the W^\pm mass, and the neutrino-lepton scattering cross section is resonantly enhanced to be a few orders of magnitude larger than the deep inelastic scattering cross section [106]. This is known as

the Glashow resonance, and the predicted cross sections are shown in figure 1.10, as conclusive evidence of its existence has yet to be demonstrated.

Figure 1.10:

Theoretical cross sections for deep inelastic scattering (black and green), charged current neutrino-lepton scattering (red and purple), and neutral current neutrino-lepton scattering (orange and blue). Figure included from Formaggio and Zeller [106].

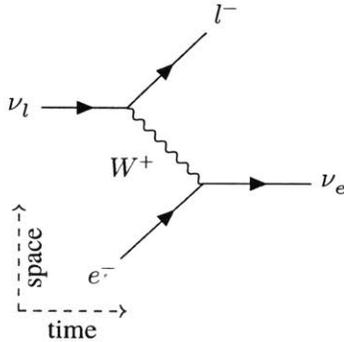
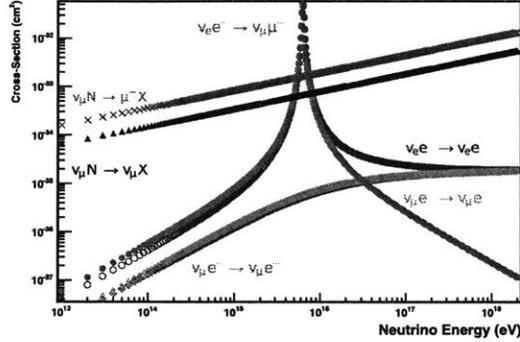


Figure 1.11:

A neutrino with flavour l scatters from an electron, producing an l lepton.

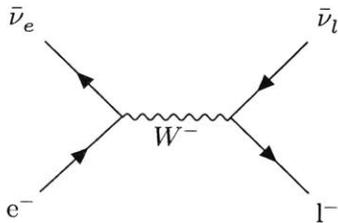


Figure 1.12:

An anti-neutrino can also scatter from an electron through an s -channel process, producing an l -neutrino pair.

1.2.4 Neutrino-lepton scattering

In charged current scattering, a neutrino of flavour l is converted into its corresponding lepton as shown in figure 1.11:

$$\nu_l + e^- \rightarrow l^- + \nu_e. \quad (1.42)$$

This scattering process has a threshold energy determined by the mass of the produced lepton [106]:

$$E_{\text{thresh}} = \frac{m_l^2 - m_e^2}{2m_e}. \quad (1.43)$$

The corresponding anti-neutrino process only occurs for electron neutrinos as shown in figure 1.12:

$$\bar{\nu}_e + e^- \rightarrow \bar{\nu}_l + l^-, \quad (1.44)$$

where l can be any of the charged leptons. The cross section for this process is approximately three times lower than for neutrinos, due to the helicity of the anti-neutrino [106]. Charged current muon or tau anti-neutrino-lepton scattering would require the presence of positrons, muons, or taus in the neutrino target.

Neutral current scattering processes are possible for all types of neutrinos. In this case, the (anti-)neutrino simply transfers momentum to an electron through a Z boson as shown in figure 1.12:

$$\nu_l/\bar{\nu}_l + e^- \rightarrow \nu_l/\bar{\nu}_l + e^-. \quad (1.45)$$

1.2.5 Detector signatures

In the previous sections, the products of various neutrino interactions have been discussed, but not how these products are identified in a detector.

Muons

The simplest and most distinctive signature is that of a muon. In most experiments, muons are *minimum ionising* particles in the sense that they deposit the minimum energy per unit length in the material they travel through. As such, they are highly penetrating particles that travel long distances before this material bleeds off enough kinetic energy to slow them to a stop. Compact neutrino detectors usually need a dedicated sub-system for identification of muons. This system sits downstream of the neutrino source, behind shielding designed to block less penetrating particles. In large format detectors, such a dedicated sub-system is not required, as the bulk of the detector naturally acts in the role of the shielding.

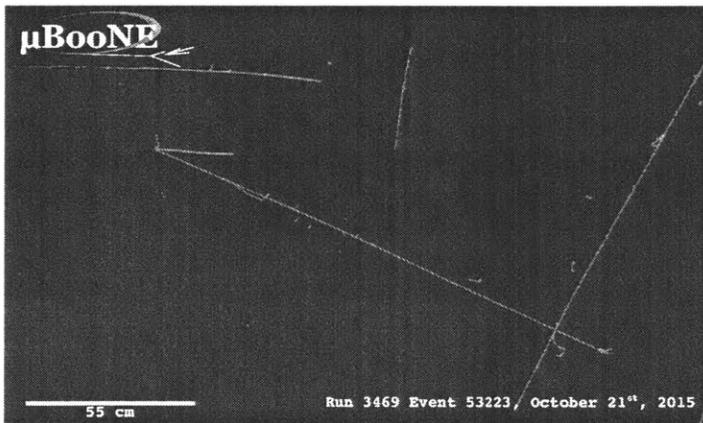


Figure 1.13:

A muon is produced when a muon neutrino, entering from the left, interacts with the MicroBooNE detector. The shorter line is created by the recoiling proton. Figure included from Woodruff and Collaboration [231].

In detectors that are either instrumented throughout their volume, or are sensitive to ionisation – such as drift chambers – the muon will leave a clean straight line as shown in figure 1.13. Other neutrino detectors – like the Super-Kamiokande experiment that confirmed neutrino oscillation – are instrumented only around their outside surface. These are usually *Cherenkov radiation* based, recording the cone of light generated when particles travel faster than the phase velocity of light: c/n where n is the refractive index of the bulk material. In these detectors, the muon will be seen as a clean circle of light where this cone intersects the detector boundary.

Electrons

Electrons, on the other hand, are only minimum ionising at very low energies, due to their light mass. At these low energies, an electron will simply leave a short track of ionisation, while a positron will also leave a short track before additionally annihilating to produce two photons with energy equal to the electron mass. At energies greater than a few MeV, the electron or positron will radiate photons that subsequently produce pairs of electron-positrons. These electrons and positrons will then repeat the process, exponentially increasing the number of particles involved to create an *electromagnetic shower*. The shower halts when the products drop below the pair-production energy threshold.

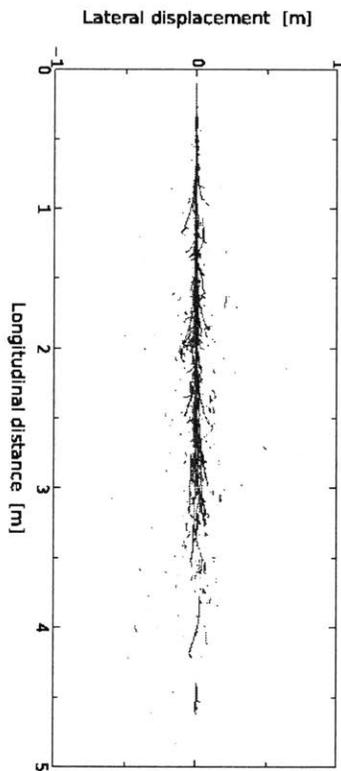


Figure 1.14:

A simulation of an electromagnetic shower in the Super-Kamiokande experiment. A 10 GeV electron is incident on the bulk water material from the top. The red lines show electrons and the blue lines are positrons. Figure included from Li and Beacom [166].

In volumetrically instrumented, or ionisation sensitive detectors, the electromagnetic shower will appear as an elongated cone of particles the length of which is proportional to the logarithm of the initial electron energy. In Cherenkov radiation based detectors, light cones are produced by the electrons and positrons in the shower. As these particles are not all moving in the same direction, the shower forms a set of fuzzy rings where the multitudes of light cones intersect the detector boundary.

Hadrons

Recoiling nucleons are usually low velocity, due to their high mass. In detectors sensitive to ionisation, they will leave short tracks. In volumetrically instrumented detectors, they may not be visible at all if the pixelisation is too coarse. In Cherenkov radiation based detectors, they likely will not produce light as their velocity is often below the threshold of c/n . Mesons, on the other hand, can be light enough to produce a significant amount of light.

Without the dedicated hadronic calorimeters common in collider experiments, distinguishing purely electronic from hadronic interactions is difficult. Both produce similar looking showers, especially as hadronic showers usually also have a strong electromagnetic component. The *critical energy*, below which showering becomes unfavourable is ~ 1 GeV for hadronic showers – compared to 80 MeV for electrons – allowing some discrimination based on energy [166]. Muons may also be present in the hadronic shower from the decay of pions and kaons, which may aid in the discrimination.

Tau neutrinos

Charged current tau neutrino interactions will produce a tau inside the detector. This tau has a lifetime of ~ 290 femtoseconds [198],

and thus decays immediately. The probability of the decay producing a muon or electron are $\sim 17\%$ and $\sim 18\%$ respectively, with hadronic products accounting for the remainder. As the tau is quite heavy, it is not usually highly relativistic, and will only travel a few millimeters or less before decaying. This is too short for all but the most sensitive neutrino detectors, and so the tau itself is essentially invisible. As such, tau neutrinos are easy to mistake for muon or electron neutrinos.

The DONUT experiment used photographic emulsion plates to provide enough spatial resolution to see the tau directly. After exposure to a tau neutrino beam, the plates were developed, scanned, and searched for the minute track that would be indicative of a tau being produced by a neutrino interaction. In 2001, the experiment announced the discovery of the tau neutrino [160], confirming the standard model prediction that the tau – itself discovered 25 years earlier – had an associated neutrino. Even today, due to the difficulty in their identification, only a handful of known tau neutrinos have been recorded across all of particle physics.

Statistical inference



Nothing can be known for sure, but some things can be known with greater certainty. Statistical inference is the process of quantifying this certainty, and selecting physical models based on measured data. As random¹⁹ phenomena are common in nature, the language of probability forms the foundation of statistics.

However, the probability distributions of data alone do not form a coherent picture of inference. The connection between probability and the degree of uncertainty in a model must be made. Originally constructed to describe communication channels, information theory provides the necessary tools to motivate frequentist and Bayesian approaches to statistics. The concepts introduced by information theory are also used as optimisation targets for the neural networks discussed in chapter 8 and 9.

Both frequentist methods and Bayesian inference form the basis of the global fit for sterile neutrinos presented in chapter 3, and the search for point sources of astrophysical neutrinos in chapter 7. Markov chain Monte-Carlo methods are employed to explore the high dimensional parameter spaces used in these analyses. These algorithms are general samplers, and are thus useful for simulation by posing the problem as a path integral, as shown in chapter 10. The theory of sampling is important for general Monte-Carlo based simulation, as detailed for IceCube in chapter 4. These topics are only covered briefly, and the interested reader is encouraged to review Applebaum [27], Bickel and Doksum [42], Burnham and Anderson [44], Casella and Berger [47], MacKay [176], and Murphy [188].

¹⁹ Either truly or effectively.

2.1 Information

The study of information theory is rooted in message encoding: the description of a code that minimises the amount of information which needs to be sent through a communication channel. These communication channels are described as sending messages about the occurrences of events which follow a certain probability distribution. An optimal code requires $-\log_2 P_k$ bits to communicate an event k that occurs with probability P_k [176].

For example, if the event is guaranteed to happen – so that $P_k = 1$ – then no bits are required, as the receiver of the message already knew that k would occur. If the distribution P has two events with equal probability, then $P_k = 1/2$ and one bit is required. For three events, with $P_1 = 1/2$, $P_2 = 1/4$, and $P_3 = 1/4$ the number of bits required to code each event is one, two, and two respectively. This code is optimal, as a message of single bit length is reserved for the more common event 1, while the longer two bit messages are used for the less common events 2 and 3, minimising the amount of information sent on average.

The average amount of information required to send messages that follow a distribution P_k can be found by taking the expectation of this negative logarithm [27]:

$$H_2(P) = \langle -\log_2 P_k \rangle_P = - \sum_k P_k \log_2 P_k. \quad (2.1)$$

By using a base 2 logarithm, the amount of information is measured in bits, and can be interpreted as the average number of yes/no binary questions required to identify an event from a set of possibilities [176]. A change in the base of the logarithm also changes the units of information — in particular, the natural logarithm yields information in *nats*:

$$H(P) = \langle -\ln P_k \rangle_P. \quad (2.2)$$

This formula is also known as the Shannon entropy, or the self-information of the distribution P .

From a practical point of view, it is likely that the message sender does not know the true distribution of events exactly. In this case, while the sender uses an encoding based on Q , the expectation of the amount of information is taken over the true distribution P . This defines a more general form of the Shannon entropy, known as the *cross-entropy* [188]:

$$H(P; Q) = \langle -\ln Q_k \rangle_P = - \sum_k P_k \ln Q_k. \quad (2.3)$$

The encoding based on Q is always sub-optimal, in the sense that the expected amount of information that will be transferred is always greater than if the encoding was based on the true distribution P .

Thus, the difference between the cross-entropy $H(P; Q)$ and the Shannon entropy $H(P)$ will always be a positive quantity, called the Kullback-Leibler (KL) divergence [44]:

$$D_{\text{KL}}(P||Q) = H(P; Q) - H(P) = \sum_k P_k \ln \frac{P_k}{Q_k}, \quad (2.4)$$

which gives the expected amount of additional information that will be sent when a scheme for Q is used in place of one devised for P .

In statistics, continuous probability distributions are common, while the Shannon entropy was only defined for discrete distributions over a set of events. As the cardinality of this set approaches infinity – as it does for a continuous distribution – this expectation of the amount of information also diverges. In addition, continuous distributions are defined in terms of probability density functions, and so a $\ln p(x)$ type measure – where $p(x)$ is a continuous distribution – would be dimensionally incorrect.

Instead, the entropy can be defined relative to a reference measure distribution $m(x)$, and after removal of the term that diverges[149, 150] becomes:

$$H(p) = - \int p(x) \ln \frac{p(x)}{m(x)} dx. \quad (2.5)$$

It follows that the continuous equivalent to the code length – $-\ln P_k$ is the point-wise information measure:

$$h(x) = - \ln \frac{p(x)}{m(x)}. \quad (2.6)$$

The problematic qualities of the continuous distribution have been removed by defining the entropy relative to $m(x)$, and so it should not be surprising that the functional form is equivalent to $-D_{\text{KL}}(p||m)$, where D_{KL} is the extension of the KL divergence to continuous functions:

$$D_{\text{KL}}(p||q) = \int p(x) \ln \frac{p(x)}{q(x)} dx. \quad (2.7)$$

In statistics, the entropy is interpreted as the amount of uncertainty in a distribution. For example, the entropy of a Kronecker delta²⁰ distribution is zero, as it is maximally certain. In comparison, the entropy of a uniform distribution is the maximum possible for any bounded distribution, as it represents the least certainty in the random variable in question. This interpretation can be generalised by

²⁰ $P_l = 1$ for some l , and $P_k = 0$ for all $k \neq l$.

considering the information measure $\ln P_k$ to be the amount of *surprise* of the value k . From this, the KL divergence $D_{\text{KL}}(P||Q)$ is interpreted as the expected additional excess *risk* of using a distribution Q in place of P .

2.2 Inference

Data is stochastic in nature, so we consider reality to be described by a probability distribution $f(\vec{d})$ where \vec{d} is the data measured by the experiment. The task of inference is to find and quantify our certainty in a model that approximates f with some tuneable parameters $\vec{\theta}$. The probability of the data under this model is denoted by the *likelihood function*²¹ $\mathcal{L}(\vec{d}|\vec{\theta})$. Ideally, this model is chosen to minimise the risk associated with using it to approximate reality. This is formalised in terms of minimising the KL divergence between f and \mathcal{L} :

$$\vec{\theta} = \underset{\vec{\theta}}{\operatorname{argmin}} D_{\text{KL}}(f||\mathcal{L}) = \underset{\vec{\theta}}{\operatorname{argmin}} \int f(\vec{d}) \frac{f(\vec{d})}{\mathcal{L}(\vec{d}|\vec{\theta})} d\vec{d}. \quad (2.8)$$

The true form of f is strictly unknown, but can be approximated using the set of measured data²² \mathcal{D} :

$$\tilde{f}(\vec{d}) = \frac{1}{|\mathcal{D}|} \sum_{\vec{d}' \in \mathcal{D}} \delta(\vec{d} - \vec{d}'). \quad (2.9)$$

Using this, the minimisation of the risk leads to a maximisation of the log-likelihood on the set of measured data:

$$\vec{\theta}_{\text{best}} = \underset{\vec{\theta}}{\operatorname{argmax}} \sum_{\vec{d} \in \mathcal{D}} \ln \mathcal{L}(\vec{d}|\vec{\theta}), \quad (2.10)$$

where $\vec{\theta}_{\text{best}}$ is known as the maximum likelihood estimate, or best-fit point.

The likelihood is usually implemented using standard probability distributions, but these distributions are often parameterised in terms of moments of the data variable, not the physical model parameters themselves. For example, a radioactive decay counting experiment can describe the statistical variation using a Poisson distribution $\text{Poiss}(\lambda)$ parameterised by the mean. This requires a *prediction function* $\Psi(\vec{\theta})$ that connects the data moments to the model parameters. In this example, $\Psi(\vec{\theta})$ would be chosen to compute the expected number of decays, if physics behaved according to the model parameters $\vec{\theta}$.

²¹ Some prescriptions denote the probability of the data as $p(\vec{d}|\vec{\theta})$, while reserving the likelihood as specifically a function over the model parameters $\mathcal{L}(\vec{\theta})$. Here, \mathcal{L} is simply used to distinguish the probability associated with the model.

²² As will be seen in section 2.3, f is approximated through its samples: the measured data.

2.2.1 Frequentist methods

Likelihood maximisation is one of the cornerstones of frequentist statistics, and forms the basis of the likelihood ratio test. This test is designed to discriminate between two hypotheses denoted H_0 and H_1 , often respectively called the *null* and *alternative* hypothesis. In the likelihood ratio test, the models associated with each hypothesis are required to be nested: the likelihood of the null hypothesis \mathcal{L}_0 should be a special case of the alternative, i.e. $\mathcal{L}_0(\vec{d}) = \mathcal{L}_1(\vec{d}|\vec{\theta}_0)$ for some θ_0 . The Neyman-Pearson lemma [189] then states that the ratio between the maximal likelihoods,

$$\Lambda(\vec{d}) = \frac{\mathcal{L}_0(\vec{d})}{\max_{\vec{\theta}'} \mathcal{L}_1(\vec{d}|\vec{\theta}')}, \quad (2.11)$$

is the most powerful *test statistic*.

If the test statistic lies above a significance threshold c , then H_0 is accepted, otherwise H_0 is rejected in favour of H_1 . This threshold is chosen such that the probability of falsely rejecting H_0 is lower than the significance level, denoted by α :

$$P(\Lambda(\vec{d}) < c | H_0) = \alpha. \quad (2.12)$$

It remains to find the probability distribution over Λ — a difficult proposition in general. Instead, the distribution can be approximated in the local neighbourhood of the best-fit point where the zeroth and first order terms of a Taylor expansion are zero:

$$\ln \Lambda(\vec{d}) \approx -\frac{1}{2}(\vec{\theta}_0 - \vec{\theta}_{\text{best}})^T \Omega(\vec{d})(\vec{\theta}_0 - \vec{\theta}_{\text{best}}). \quad (2.13)$$

Wilks' theorem [230] shows in the asymptotic case where the number of data points approaches infinity, that this expression holds exactly and $\vec{\theta}_{\text{best}}$ is normally distributed with mean $\vec{\theta}_0$ and covariance $\Omega(\vec{d})^{-1}$. Thus it follows that $-2 \ln \Lambda$ is approximately a sum of independent normally distributed variables with unit variance, and the probability distribution for this sum²³ is called the χ^2 -distribution [42].

The χ^2 is useful in itself as a test statistic. For observed data²⁴ \vec{d} and predictions $\vec{\Psi}(\vec{\theta})$, the χ^2 is defined as

$$\chi^2(\vec{\theta}) = (\vec{d} - \vec{\Psi}(\vec{\theta}))^T \Sigma^{-1} (\vec{d} - \vec{\Psi}(\vec{\theta})), \quad (2.14)$$

where Σ is the covariance matrix of the data. The χ^2 -distribution is parameterised by the number of *degrees of freedom*, defined by the number of independent normally distributed variables that were summed together. Pearson defined the χ^2 goodness-of-fit test by minimising the χ^2 over the parameter space. This best-fit χ^2 is still

²³ The number of degrees of freedom of this χ^2 -distribution is equal to the number of model parameters.

²⁴ Here it is assumed that \vec{d} is normally distributed; however, observed data is often actually Poisson distributed. When the number of observed events is large, the Poisson distribution approximates the normal distribution.

distributed according to the χ^2 -distribution, but the number of degrees of freedom is equal to the dimensionality of \vec{d} minus the number of parameters — as the minimisation constraint reduces the number of independent random variables that are being summed together.

As it happens, the normal distribution log-likelihood is proportional to the χ^2 , and so the log-likelihood ratio can be defined as

$$\ln \Lambda(\vec{d}) = -\frac{1}{2} (\chi_0^2(\vec{d}) - \chi_1^2(\vec{d})) = -\frac{1}{2} \Delta\chi^2(\vec{d}), \quad (2.15)$$

where χ_0^2 and χ_1^2 are, respectively, the minimum χ^2 for the null and alternative hypothesis. Thus, the $\Delta\chi^2$ can be used as a model comparison test statistic, and is χ^2 distributed with a number of degrees of freedom equal to the number of model parameters. The significance threshold for a given significance level is set by using the inverse cumulative distribution function of the χ^2 -distribution.

Confidence regions

The likelihood ratio test – as shown above – informs only on the best-fit point. Greater information can be derived from the data by creating a set of all points in parameter space for which the test statistic threshold is reached. This is formalised for a likelihood ratio test by first defining a ratio

$$\Lambda(\vec{d}; \vec{\theta}) = \frac{\mathcal{L}_1(\vec{d}|\vec{\theta})}{\max_{\vec{\theta}'} \mathcal{L}_1(\vec{d}|\vec{\theta}')}, \quad (2.16)$$

which compares a hypothesis $H(\vec{\theta})$, with a given set of parameters $\vec{\theta}$, to the best-fit hypothesis H_1 . As in model comparison to the null hypothesis discussed previously, $H(\vec{\theta})$ is accepted if Λ is greater than a significance threshold c . The set of all models that are accepted by this metric is

$$\mathcal{C}(\alpha) = \{ \vec{\theta} : \Lambda(\vec{d}; \vec{\theta}) > c(\alpha) \}. \quad (2.17)$$

This is known as a $1 - \alpha$ *confidence region*, and has the important property of *coverage* [47]: the fraction of all confidence regions that cover the true value of $\vec{\theta}$ should be equal to $1 - \alpha$. It should be noted that this does not imply that the true value is contained within any given region with probability $1 - \alpha$, as in the frequentist interpretation, the confidence regions are random while the true model parameter is a constant. Thus, given any single confidence region there is no room for a probabilistic interpretation, it either contains the true value or it does not.

For a χ^2 test statistic, the $1 - \alpha$ confidence region can be more concretely specified as

$$\mathcal{C}(\alpha) = \left\{ \vec{\theta} : \chi^2(\vec{d}, \vec{\theta}) - \chi_{\text{best-fit}}^2(\vec{d}) < \text{CDF}_{\chi^2}^{-1}(k, 1 - \alpha) \right\}, \quad (2.18)$$

where CDF_{χ^2} is the χ^2 Cumulative Distribution Function for k degrees of freedom, and k is the number of model parameters. An example of finding the confidence regions of a χ^2 statistic is shown in figure 2.1. For this particular choice of $c = \text{CDF}_{\chi^2}^{-1}(k, 1 - \alpha)$, two regions are produced.

If the number of model parameters exceeds three, it can become difficult to graphically display the confidence regions. The dimensionality of the problem can be reduced through a process called *profiling*²⁵. In this method, the χ^2 is replaced by a χ^2 minimised over the unwanted parameters $\vec{\phi}$:

$$\chi^2(\vec{d}; \vec{\theta}, \vec{\phi}) \rightarrow \chi^2(\vec{d}; \vec{\theta}) = \min_{\vec{\phi}} \chi^2(\vec{d}; \vec{\theta}, \vec{\phi}) \quad (2.19)$$

Confidence regions can then be constructed; however, the number of degrees of freedom of the χ^2 -distribution is no longer k . Instead, if n parameters were removed through profiling, the distribution has $k - n$ degrees of freedom, as n additional constraints have been placed on the χ^2 due to the minimisation inherent in the profiling.

2.2.2 Bayesian inference

A frequentist model comparison can determine the probability that a hypothesis will be rejected based on a significance threshold; but, to make a statement on the degree of certainty in a model, the model itself must be incorporated into the framework of probability. At first glance the concept of assigning a probability to a model does not seem to make sense; a model is either the correct description of reality or it is not, there is no “maybe”. Instead, in this framework the probability represents the degree of certainty in the model, in the same way that the entropy represents the uncertainty in a probability distribution.

This is formalised by a joint distribution $p(\vec{d}, \vec{\theta})$ over both the data and model parameters, in which the support lent to the model by the data is encoded in the statistical dependence between \vec{d} and $\vec{\theta}$. The amount of statistical dependence can be measured by the *mutual information*:

$$I(\vec{d}, \vec{\theta}) = - \int p(\vec{d}, \vec{\theta}) \ln \frac{p(\vec{d}, \vec{\theta})}{p(\vec{d})p(\vec{\theta})} d\vec{d}d\vec{\theta}. \quad (2.20)$$

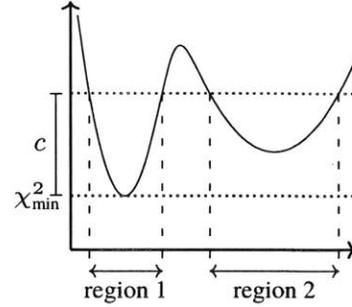


Figure 2.1: Confidence regions as defined by a χ^2 statistic. Two regions are produced by this particular choice of the significance threshold c .

²⁵ While profiling is defined more generally in terms of likelihood functions, the discussion here concentrates on χ^2 distributions due to the effect on the degrees of freedom.

For example, if the data and model are statistically independent, then $p(\vec{d}, \vec{\theta}) = p(\vec{d})p(\vec{\theta})$ and $I(\vec{d}, \vec{\theta}) = 0$, stating there is no mutual information. As marginalisation over all possible data is usually not practical or desired, the point-wise mutual information measure is more useful:

$$\text{pmi}(\vec{d}, \vec{\theta}) = -\ln \frac{p(\vec{d}, \vec{\theta})}{p(\vec{d})p(\vec{\theta})} = -\ln \frac{p(\vec{d}|\vec{\theta})}{p(\vec{d})}. \quad (2.21)$$

The conditional formulation makes the relationship clear: if the data is less likely under the model, the p.m.i is positive. If the data is more likely under the model, the p.m.i is negative²⁶.

²⁶ Only the expectation of the p.m.i, I , is required to be positive. As a point-wise measure, the p.m.i itself can be negative.

Intuitively, we should expect that if the data supports a model, the uncertainty in the model should diminish. In addition, the information gained from the data should add to the information that already exists on the model. Thus, we expect the point-wise information of the model, conditioned on the data, to be

$$h(\vec{\theta}|\vec{d}) = \text{pmi}(\vec{\theta}, \vec{d}) + h(\vec{\theta}), \quad (2.22)$$

so that the entropy of the model decreases when the p.m.i is negative, and increases when it is positive. Taking the exponential of both sides recovers the familiar Bayes' theorem:

$$p(\vec{\theta}|\vec{d}) = \frac{p(\vec{d}|\vec{\theta})p(\vec{\theta})}{p(\vec{d})} = \frac{\mathcal{L}(\vec{d}|\vec{\theta})\pi(\vec{\theta})}{\mathcal{L}(\vec{d})}. \quad (2.23)$$

The objective of Bayesian inference is to calculate $p(\vec{\theta}|\vec{d})$, known as the *posterior*. This requires the likelihood function $\mathcal{L}(\vec{d}|\vec{\theta})$, also used in section 2.2.1. Further, the *marginal likelihood*, $\mathcal{L}(\vec{d})$, must be found for a direct calculation; although, as will be seen in section 2.4, this can be side-stepped by defining an algorithm that uses a ratio of posteriors — cancelling out the marginal likelihood.

The *prior*, $\pi(\vec{\theta})$, must also be defined. This encapsulates any previous knowledge we have about the model. For example, it could be the results of a previous experiment, or based on theoretical considerations. When no such knowledge exists, the prior is often chosen to be a uniform, exponential, or normal distribution; they are the maximum entropy distributions for a distribution over a finite interval, with a specified mean, and with a specified mean and variance, respectively²⁷. A maximum entropy distribution is desired, as it possesses the most uncertainty in the model — an objectively conservative choice.

²⁷ The prior may seem like a subjective inconvenience, but it is actually a basic requirement for inference in this con-

Credible regions

The posterior can be visualised using *credible regions*. A credible region $\mathcal{C}(\beta)$ is an area of parameter space that contains β of the total probability of the distribution:

$$\int_{\mathcal{C}(\beta)} p(\vec{\theta}|\vec{d})d\vec{\theta} = \beta. \tag{2.24}$$

This does not uniquely specify the set of all points in the region \mathcal{C} , and – even in 1D – there are multiple sub-types of credible regions. The most useful definition when working in high dimensions is the Highest Posterior Density (HPD) credible region. Here, the interval is the (possibly disjoint) set of points whose posterior probability density meets a minimum threshold t :

$$\mathcal{C}(\beta) = \{\vec{\theta} : p(\vec{\theta}|\vec{d}) > t(\beta)\}, \tag{2.25}$$

where $t(\beta)$ is constrained by equation 2.24. An example of finding the credible regions is shown in figure 2.2. Intuitively this can be seen as an interval, which starting at the mode of the posterior²⁸, grows to include an area whose integrated probability is exactly β and where all points inside the interval have higher probability density than all points outside the interval.

Model comparison

So far, we have only considered the posterior for a model as a function of its parameters. This distinction is important, as the posterior for the model itself can also be defined using Bayes’ theorem:

$$p(H_0) = \frac{\mathcal{L}(\vec{d}|H_0)\pi(H_0)}{p(\vec{d})}, \tag{2.26}$$

where H_0 is a label for the model. This requires the likelihood of the model $\mathcal{L}(\vec{d}|H_0, \vec{\theta})$ marginalised over $\vec{\theta}$. It also requires the likelihood, $\mathcal{L}(\vec{d}|H)$, marginalised over all possible models H , denoted by $p(\vec{d})$.

While the first quantity can be difficult to calculate, the second is essentially impossible. It is removed by considering a posterior probability ratio between two models H_0 and H_1 :

$$\frac{p(H_0)}{p(H_1)} = \frac{\mathcal{L}(\vec{d}|H_0)\pi(H_0)}{\mathcal{L}(\vec{d}|H_1)\pi(H_1)} = \mathcal{B}_{H_0/H_1} \frac{\pi(H_0)}{\pi(H_1)}, \tag{2.27}$$

where \mathcal{B}_{H_0/H_1} is called the Bayes’ factor. The Bayes’ factor thus measures the relative evidence for each model, and can be reported in a manner independent of the prior on the hypotheses.

struction. To compute the probability of – and thus our belief in – the model using only measured data, a prior must be involved.

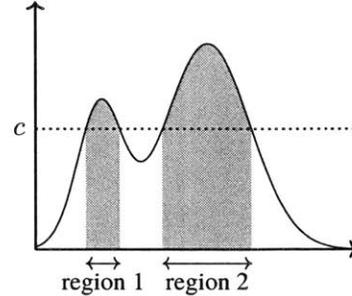


Figure 2.2: The highest posterior density credible regions. The value is c is chosen so that the total area of the shaded regions defined by c is equal to β .

²⁸ Equivalent to the best-fit point when a flat prior is used.

2.3 Sampling

The posterior is rarely amenable to an analytic treatment, and so numerical methods are required. It is most instructive to consider the numerical treatment of a probability distribution in terms of its expectation values, as these define the distribution uniquely. The expectation of a function $f(x)$ is defined as an integral over the probability space, weighted by the probability distribution $p(x)$, is:

$$\langle f(x) \rangle = \int f(x)p(x)dx. \quad (2.28)$$

To compute this expectation numerically, the integral must be approximated through a finite number of evaluations of f and p . The most basic method of defining this approximation is to consider a set of equally spaced points in probability space. The continuous functions f and p are evaluated at these points, and the weighted sum forms the approximate integral.

This can be formalised by using the Dirac delta distribution identity:

$$\langle f(x) \rangle = \int f(x) \int p(x')\delta(x - x')dx' dx. \quad (2.29)$$

Let the set of these evaluation points be \mathcal{S} , the integral over dx' can now be approximated by a sum over the points in \mathcal{S} :

$$\langle f(x) \rangle \approx \int f(x) \sum_{x' \in \mathcal{S}} p(x')\delta(x - x')\Delta x' dx \approx \int f(x)\tilde{p}(x)dx, \quad (2.30)$$

where $\Delta x'$ is the inter-point spacing, and $\tilde{p}(x)$ is the approximation to the probability distribution p :

$$\tilde{p}(x) = \sum_{x' \in \mathcal{S}} p(x')\delta(x - x')\Delta x'. \quad (2.31)$$

Each of these points is individually called a sample of the distribution, and it carries an associated sample weight. In this case, the sample weight is the product $p(x')\Delta x'$, and is required to ensure the expectation is evaluated correctly.

Given a set of these weighted samples, the expectation can be found by applying these Dirac delta distributions to the integral in equation 2.30:

$$\langle f(x) \rangle \approx \sum_{x' \in \mathcal{S}} f(x')p(x')\Delta x'. \quad (2.32)$$

This formula can also be recognised as a basic method of integration²⁹.

Like integration, it suffers from poor performance in high dimensional spaces due to most of the samples being placed in the tails of the distribution, leaving relatively few samples around the mode of the distribution where $p(x)$ is largest. To combat this, a new scheme of selecting the sample locations can be devised, where samples are clustered closer to the regions of highest probability. Consider placing the samples not at equal spacing in x , but at equal spacing in probability:

$$\int_{x_i}^{x_{i+1}} p(x) dx = \frac{1}{|\mathcal{S}|}, \quad (2.33)$$

where x_i and x_{i+1} are two consecutive samples, and $|\mathcal{S}|$ is the total number of samples. If the samples are finely spaced, then

$$p(x_i) \Delta x_i = p(x_i)(x_{i+1} - x_i) \approx |\mathcal{S}|^{-1}, \quad (2.34)$$

and the approximate distribution from equation 2.31 can be written as

$$\tilde{p}(x) = \frac{1}{|\mathcal{S}|} \sum_{x' \in \mathcal{S}} \delta(x - x'). \quad (2.35)$$

Note that in this case, all samples have the same weight of $|\mathcal{S}|^{-1}$ and so this particular scheme is called equally-weighted sampling.

While the approximate distribution \tilde{p} will now compute the expectation efficiently in multi-dimensional problems, the actual generation of the samples is no longer practical, as it requires integrating the probability distribution to find the step sizes. In fact, it is considerably easier to generate equally-weighted samples by drawing them randomly³⁰ from p . This can be done by transforming samples drawn from a uniform distribution to the desired distribution p , through the CDF of p . Let \mathcal{S}' be a set of equally-weighted samples drawn from the uniform distribution $u \sim U(0, 1)$, such that the approximate distribution \tilde{p}' for U is

$$\tilde{p}'(u) du = \frac{1}{|\mathcal{S}'|} \sum_{u' \in \mathcal{S}'} \delta(u - u') du. \quad (2.36)$$

The points $u \in \mathcal{S}'$ are mapped to the desired probability space through the inverse CDF:

$$x = \text{CDF}_p^{-1}(u), \quad (2.37)$$

²⁹ And there is no benefit to using equally spaced samples over advanced methods of integration such as quadrature.

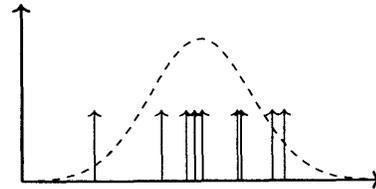


Figure 2.3:

A set of equally weighted samples, shown as arrows, for the normal distribution, shown by the dashed line.

³⁰ Indeed, this is why they are named *samples*, as they are commonly sampled from the distribution itself.

which forms a new set \mathcal{S} . After the change in variables, this yields equally-weighted samples from p :

$$\tilde{p}'(x)dx = \frac{1}{|\mathcal{S}'|} \sum_{x' \in \mathcal{S}'} \frac{\delta(x - x')}{p(x')} p(x) dx = \frac{1}{|\mathcal{S}'|} \sum_{x' \in \mathcal{S}'} \delta(x - x') dx. \quad (2.38)$$

An example for samples from the normal distribution is shown in figure 2.3.

Known as the inverse CDF theorem, this allows the efficient generation of samples when a closed form expression for the CDF is known. Although this demonstration was one-dimensional, samples from multi-dimensional distributions can be generated by successive samples from a chain of conditional distributions. When no efficient method of finding the CDF is known, more advanced methods are required.

2.3.1 Re-weighting and re-sampling

These usually generate a set of samples \mathcal{Q} from a proposal distribution q . This distribution is chosen to be easy to sample³¹ from – such as a normal distribution – and to roughly follow the desired distribution p . Then, an expectation over the desired distribution can be approximated by

$$\langle f(x) \rangle_p = \int f(x) p(x) dx = \int f(x) \frac{p(x)}{q(x)} q(x) dx \quad (2.39)$$

$$\approx \int f(x) \frac{p(x)}{q(x)} \tilde{q}(x) dx, \quad (2.40)$$

where $\tilde{q}(x)$ is the approximate distribution of q defined by equation 2.35 using the samples from \mathcal{Q} . The distribution p can thus be approximated by

$$\tilde{p}(x) = \frac{1}{|\mathcal{Q}|} \sum_{x' \in \mathcal{Q}} \frac{p(x')}{q(x')} \delta(x - x') \quad (2.41)$$

The samples from q have been *re-weighted* by a factor of $p(x)/q(x)$, and, while they are no longer equally-weighted, they can now be used to compute expectations over p . As the samples are placed in an important region, defined by q , this method is called “importance sampling”.

If equally-weighted samples are required, the approximate distribution can be re-sampled to remove these weighting factors. Suppose that a process π is applied to the samples, where the probability

³¹ In the sense that it should have a known, closed form, or easy to approximate CDF.

of the process occurring is $\omega(\pi|x)$. If the process is actually performed with distribution $\psi(\pi|x)$, then the samples undergoing the process must be re-weighted³²:

$$w \rightarrow w \frac{\omega(\pi|x)}{\phi(\pi|x)}, \quad (2.42)$$

where w is the initial weight of the sample. In this case, the process is simply the identity transformation and thus $\omega(\pi|x) = 1$. However, with probability

$$\psi(T|x) = \frac{p(x)}{q(x)}, \quad (2.43)$$

the sample is removed from the set. Thus, the surviving samples must be re-weighted by a factor of $q(x)/p(x)$ which leaves equally-weighted samples. This algorithm is called “rejection sampling”, and to ensure that $\psi(T|x)$ is a well defined probability, q is rescaled³³ by a factor M so that $\psi(T|x) < 1$ for all x . An example of $Mq(x)$ and $p(x)$ are shown in figure 2.4.

2.4 Markov Chain Monte-Carlo

Estimation of the posterior using rejection sampling is generally not feasible, as explicit calculation of the posterior probability – and thus the marginal likelihood – is still required by the rejection step. Even if the marginal likelihood were known, the selection of the proposal distribution can be complicated by an intricate, multi-modal, or simply unknown posterior. The Metropolis-Hastings Markov Chain Monte-Carlo (MCMC) algorithm neatly side-steps this first issue by defining an algorithm which depends only on probability ratios [129]:

$$\frac{p(\vec{\theta}'|\vec{d})}{p(\vec{\theta}|\vec{d})} = \frac{\mathcal{L}(\vec{d}|\vec{\theta}')\pi(\vec{\theta}')}{\mathcal{L}(\vec{d}|\vec{\theta})\pi(\vec{\theta})}, \quad (2.44)$$

such that the marginal likelihood $\mathcal{L}(\vec{d})$ is cancelled out.

This is done by first considering a Markov Chain in a continuous state space, with a probability $\sigma(x)$ having state x , and transition probability $T(x \rightarrow x')$ of moving from state x to state x' . These transitions are directed by a proposal distribution g , deliberately chosen to be localised around the current state x . Presumably the Markov Chain will eventually find itself in regions of high probability, and so samples drawn from the proposal distribution will also likely be in regions of high probability, addressing the second issue.

The state space of the Markov Chain are the model parameters $\vec{\theta}$, and the goal for the algorithm is for the probability $\sigma(\vec{\theta})$ to match

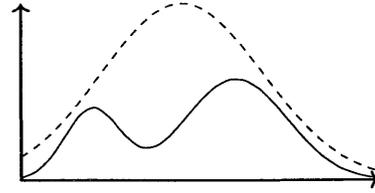


Figure 2.4:

Samples are drawn from the proposal distribution q shown as the dashed line, and are accepted or rejected based on the ratio between the desired distribution p , shown as the solid line, and q .

³² So far, this is simply another form of re-weighting as described above.

³³ This introduces a factor of M into the weights; however, this is removed by a subsequent normalisation of the approximate distribution. Although the re-sampling is random, on average M is the correction factor needed to normalise the new set of samples.

the posterior $p(\vec{\theta}|\vec{d})$ after some warm-up procedure. After this point, more samples can be generated from the posterior by simply applying the transition rule repeatedly. To achieve this goal, the Markov Chain must obey the condition of detailed balance:

$$\sigma(\vec{\theta})T(\vec{\theta} \rightarrow \vec{\theta}') = \sigma(\vec{\theta}')T(\vec{\theta}' \rightarrow \vec{\theta}), \quad (2.45)$$

which is to say that the Markov Chain will not preferentially pile up in any one direction, biasing the results.

The transition proceeds by first drawing a proposed new state $\vec{\theta}'$ from the proposal distribution $\gamma(\vec{\theta}'|\vec{\theta})$. As this new state may or may not be a likely sample from the posterior, the transition must first be accepted with probability $A(\vec{\theta}'|\vec{\theta})$. Thus, the overall transition probability is

$$T(\vec{\theta} \rightarrow \vec{\theta}') = \gamma(\vec{\theta}'|\vec{\theta})A(\vec{\theta}'|\vec{\theta}), \quad (2.46)$$

and detailed balance is maintained if

$$\frac{A(\vec{\theta}'|\vec{\theta})}{A(\vec{\theta}|\vec{\theta}')} = \frac{\sigma(\vec{\theta}')}{\sigma(\vec{\theta})} \frac{\gamma(\vec{\theta}|\vec{\theta}')}{\gamma(\vec{\theta}'|\vec{\theta})}, \quad (2.47)$$

which is known as the ‘‘Metropolis-Hastings rule’’.

As we wish the Markov Chain to eventually relax to the posterior distribution, we set $\sigma(\vec{\theta}) = p(\vec{\theta}|\vec{d})$, and choose a proposal distribution, such as a normal distribution. The Markov Chain is then initialised into a random location in model space θ_0 , and proceeds as shown in algorithm 1. At each step, a proposal is drawn from g and the likelihood computed at this proposed point. The acceptance factor,

$$A(\vec{\theta} \rightarrow \vec{\theta}') = \frac{A(\vec{\theta}'|\vec{\theta}) \mathcal{L}(\vec{d}|\vec{\theta}') \pi(\vec{\theta}')}{A(\vec{\theta}|\vec{\theta}') \mathcal{L}(\vec{d}|\vec{\theta}) \pi(\vec{\theta})} \frac{\gamma(\vec{\theta}'|\vec{\theta}_i)}{\gamma(\vec{\theta}_i|\vec{\theta}')}, \quad (2.48)$$

is computed, and the proposal is accepted with probability

$$\alpha = \min(1, A(\vec{\theta} \rightarrow \vec{\theta}')). \quad (2.49)$$

If the proposal is accepted, the current position of the chain is updated to the location of the proposal. Otherwise, the chain remains at the current location, and a new proposal is drawn for the next step.

The initialisation location of the Markov Chain will generally be in a region of low probability. As the chain moves towards the higher probability regions, it will necessarily generate samples in the vicinity of this low probability region. This introduces a bias into the samples, which must be taken into account. The simplest method of removing this bias is to simply discard the initial trajectories of the chain, usually chosen to be some significant fraction of the total number of iterations. After this *burn-in* period, the MCMC should have relaxed into the posterior distribution.

Algorithm 1 Metropolis-Hastings Markov Chain Monte Carlo

```

1:  $\vec{\theta}_0 \leftarrow$  random location in  $\theta$ -space.
2:  $\mathcal{L}_0 \leftarrow \mathcal{L}(\vec{d}|\vec{\theta}_0)$ 
3: for  $i \leftarrow 0, \dots, i_{\max}$  do
4:    $\vec{\theta}' \leftarrow$  random sample from  $\gamma(\vec{\theta}|\vec{\theta}_i)$ 
5:    $\mathcal{L}' \leftarrow \mathcal{L}(\vec{d}|\vec{\theta}')$ 
6:    $\alpha \leftarrow \min(1, A(\vec{\theta}_i \rightarrow \vec{\theta}'))$ 
7:    $u \leftarrow$  uniform random number between 0 and 1
8:   if  $u < \alpha$  then
9:      $\vec{\theta}_{i+1} \leftarrow \vec{\theta}'$ 
10:     $\mathcal{L}_{i+1} \leftarrow \mathcal{L}'$ 
11:   else
12:      $\vec{\theta}_{i+1} \leftarrow \vec{\theta}_i$ 
13:      $\mathcal{L}_{i+1} \leftarrow \mathcal{L}_i$ 

```

2.4.1 Affine invariant Markov Chain Monte-Carlo

The Metropolis-Hastings rule forms the foundation of all MCMC algorithms, but its direct application has two disadvantages.

First, the proposal distribution g must be carefully selected, or else the MCMC will very often make bad proposals that have no chance of being accepted. Ideally, the proposal distribution would match the local structure of the posterior, but hand tuning this distribution for every use-case is an undesirable quality.

The affine invariant MCMC algorithm [119] – implemented by the *emcee* software package [105] – uses information about the posterior contained in the already-drawn samples to define the proposal distribution. Rather than using the history of the Markov Chain, this algorithm defines an *ensemble* of chains running in parallel. A proposal for chain $\vec{\theta}_i^\alpha$ is then made by drawing a random chain $\vec{\theta}_i^\beta$ where $\beta \neq \alpha$ from the ensemble, then moving the current chain towards $\vec{\theta}_i^\beta$:

$$\vec{\theta}_i^{\alpha'} = \vec{\theta}_i^\beta + Z(\vec{\theta}_i^\alpha - \vec{\theta}_i^\beta), \quad (2.50)$$

where Z is a random sample from the distribution

$$h(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \frac{1}{a} \leq z \leq a \\ 0 & \text{otherwise} \end{cases}, \quad (2.51)$$

and a is a scaling parameter that typically needs little adjustment. Assuming that the other chains in the ensemble have relaxed into regions of high probability, the affine invariant algorithm should propose coordinates that are also high probability.

The Metropolis-Hasting rule for this proposal function is [105]:

$$\alpha = \min \left(1, Z^{N-1} \frac{\mathcal{L}(\vec{d}|\vec{\theta}_i^{\alpha'})\pi(\vec{\theta}_i^{\alpha'})}{\mathcal{L}(\vec{d}|\vec{\theta}_i^{\alpha})\pi(\vec{\theta}_i^{\alpha})} \right) \quad (2.52)$$

where N is the dimensionality of θ -space.

2.4.2 Parallel tempering

Secondly, chains can get caught in local modes of the posterior distribution. Although the probabilistic nature of the algorithm guarantees that it will explore the entire parameter space in the limit of an infinite number of iterations, in practice it is possible for an MCMC to stay only within a single mode of a distribution.

A similar use of ensembles can be used to inform chains in one mode of the existence of other modes in the posterior, in a scheme known as “Parallel Tempering” [90]. Each member of the ensemble³⁴ is defined to have a modified likelihood function. The modification is made in analogy to thermodynamics, with the energy of a member determined by the log-probability:

$$E(\vec{\theta}) = -\ln \mathcal{L}(\vec{d}|\vec{\theta}) - \ln \pi(\vec{\theta}). \quad (2.53)$$

Each member has an associated inverse-temperature β_k , so that the function which is actually sampled is defined by the Boltzmann distribution,

$$e^{-\beta_k E_k(\vec{\theta})}. \quad (2.54)$$

Only the members with $\beta_k = 1$ will sample the true posterior, but other members with $\beta_k > 1$ will be *tempered* – with sharp features flattened out.

These higher temperature members can freely explore the posterior, and information is exchanged between members using the *replica exchange* step. Instead of proposing to move a chain to a new region of parameter space, replica exchange proposes to swap the locations of two members with different temperatures. These participants are chosen randomly, and a proposal for a swap between $\vec{\theta}_i$ and $\vec{\theta}_j$ is accepted with probability

$$\alpha = \min \left(1, \frac{e^{-\beta_k E(\vec{\theta}_j)} e^{-\beta_l E(\vec{\theta}_i)}}{e^{-\beta_k E(\vec{\theta}_i)} e^{-\beta_l E(\vec{\theta}_j)}} \right) \quad (2.55)$$

$$= \min \left(1, \exp(\beta_k - \beta_l) (E(\vec{\theta}_i) - E(\vec{\theta}_j)) \right). \quad (2.56)$$

The replica exchange step is not performed at every step in the chain, rather it is customary to perform the exchange with a certain probability — often chosen to be around 10%.

³⁴ When combined with the affine invariant algorithm, this becomes an ensemble of ensembles.

2.4.3 Reversible jump

So far we have only considered models with a fixed number of parameters; however, it is possible to define models where the number of parameters is part of the inference problem. Known as *non-parametric* models, one example is a so-called “Gaussian mixture model”³⁵ used to describe n clusters in a dataset:

$$p(x|n, \{\eta_i, \mu_i, \sigma_i\}) = \sum_{i=1}^n \eta_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}. \quad (2.57)$$

This requires a MCMC that can not only explore the parameter space for a given number of parameters, but can also transition across the number of dimensions.

The Reversible Jump Markov Chain Monte-Carlo algorithm [122] is once such trans-dimensional sampler. A probability distribution with n parameters $p_n : \mathbb{R}^n \rightarrow \mathbb{R}$ cannot be directly compared to a probability distribution with m parameters $p_m : \mathbb{R}^m \rightarrow \mathbb{R}$ when $n \neq m$. When $m > n$, p_n can be augmented with an additional distribution $q : \mathbb{R}^{m-n} \rightarrow \mathbb{R}$ so that

$$p_n \times q : \mathbb{R}^m \rightarrow \mathbb{R} \quad (2.58)$$

has the same dimensionality as p_m . Then, after sampling, the extra dimensions introduced by q can be effectively marginalised out to recover the original distribution p_n :

$$p_n(\vec{\theta}) = \int p_n(\vec{\theta})q(\vec{z})dz \quad (2.59)$$

Although the dimensionality of $p_n \times q$ and p_m are equal, the parameters of each are members of different vector spaces. Creating a transition from one coordinate space to the other requires a coordinate transform function $g : \mathbb{R}^n \otimes \mathbb{R}^{m-n} \rightarrow \mathbb{R}^m$ as shown in figure 2.5. This function is an arbitrary choice, but to ensure efficient proposals this choice requires some care. As the chain will presumably be located in regions of high probability density in $p_n \times q$, this function should map these regions to corresponding regions of high probability density in p_m .

The jump proposal proceeds by drawing a random sample \vec{z} from q . Using \vec{z} and the current position of the chain $\vec{\theta}$, the proposal location in the higher dimensional coordinate system $\vec{\phi}$ is computed from g :

$$\vec{\phi} = g(\vec{\theta}, \vec{z}). \quad (2.60)$$

³⁵ A similar construction can be used to model an unknown number of stars in a field of view [85].

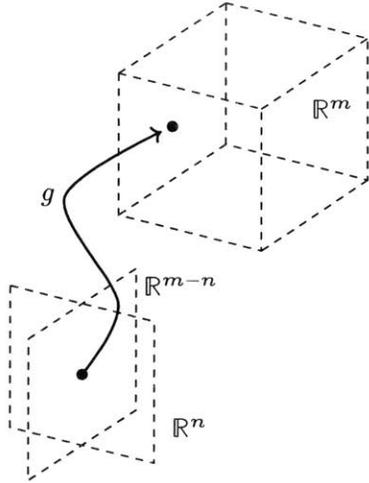


Figure 2.5: A point in the product space of \mathbb{R}^n and \mathbb{R}^{m-n} is mapped to a point in the space \mathbb{R}^m .

The acceptance factor of this proposal is given by

$$A_{\text{jump}}(n, \vec{\theta}, \vec{z} \rightarrow m, \vec{\phi}) = \frac{p_n(\vec{\theta}) p(m \rightarrow n)}{p_m(\vec{\phi}) p(n \rightarrow m) q(\vec{z})} \left| \frac{\partial g(\vec{\theta}, \vec{z})}{\partial \vec{\phi}} \right|, \tag{2.61}$$

where $|\partial g / \partial \vec{\phi}|$ is the Jacobian factor for the coordinate transform g . The probabilities $p(m \rightarrow n)$ and $p(n \rightarrow m)$ reflect the rate of which moves are proposed from m to n dimensions and vice-versa. The jump proposal is then accepted with probability

$$\alpha = \min(1, A_{\text{jump}}). \tag{2.62}$$

The reverse move of reducing the number of dimensions is performed by deterministically computing the proposal coordinates \vec{x} and \vec{z} from the current position of the chain \vec{y} via the inverse coordinate transform g^{-1} . The acceptance factor is computed again according to equation 2.61, and the jump is accepted with probability

$$\alpha = \min(1, A^{-1}). \tag{2.63}$$

In Bayesian inference, the function p_n will be the product of the likelihood for model n , the prior on the model- n parameters, and the prior on the model- n itself:

$$p_n(\vec{\theta}) = \mathcal{L}(\vec{d}|n, \vec{\theta})\pi(\vec{\theta}|n)\pi(n). \tag{2.64}$$

The sampler is initialised into a random location in one of the model spaces – usually the lowest dimensional model – and then proceeds as shown in algorithm 2. After completion, the burn-in period is discarded.

2.5 Nested sampling

The Markov Chain Monte-Carlo algorithm is designed to produce samples from the posterior distribution, and does not directly calculate the marginal likelihood³⁶. It seems plausible that the integration of the joint distribution, $\mathcal{L}(\vec{d}|\vec{\theta})\pi(\vec{\theta})$, could be estimated from the samples and their associated likelihood and prior values. However, one such method, the harmonic mean estimator, is notorious for poor performance at this task [107].

Nested sampling [220] seeks to define the marginal likelihood as a direct product of the sampling process. This is achieved by redefining the marginal likelihood expression,

$$\mathcal{L}(\vec{d}) = \int \mathcal{L}(\vec{d}|\vec{\theta})\pi(\vec{\theta})d\vec{\theta} = \int \mathcal{L}(\vec{d}|\vec{\theta})d\pi, \tag{2.65}$$

³⁶ It should be noted that Reversible Jump MCMC can directly calculate the Bayes' factor for two models.

Algorithm 2 Reversible Jump Markov Chain Monte Carlo

```

1:  $n_0 \leftarrow 1$ 
2:  $\vec{\theta}_0 \leftarrow$  random location in  $\theta$ -space for model  $n$ 
3:  $\mathcal{L}_0 \leftarrow \mathcal{L}(\vec{d}|n, \vec{\theta}_0)$ 
4: for  $i \leftarrow 0, \dots, i_{\max}$  do
5:   if  $\text{uniform}(0, 1) < 0.2$  then  $\triangleright$  With 20% chance, attempt a
   jump move
6:   if  $\text{uniform}(0, 1) < 0.5$  then  $\triangleright$  With 50% chance,
   propose to move up in  $n$ 
7:      $\vec{z} \leftarrow$  random sample from  $q(\vec{z})$ 
8:      $\vec{\phi} \leftarrow g(\vec{\theta}_i, \vec{z})$ 
9:      $m \leftarrow n_i + \Delta n$ 
10:     $\mathcal{L}' \leftarrow \mathcal{L}(\vec{d}|m, \vec{\phi})$ 
11:     $\alpha \leftarrow \min(1, A_{\text{jump}}(n_i, \vec{\theta}_i, \vec{z} \rightarrow m, \vec{\phi}))$ 
12:    else  $\triangleright$  Propose to move down in  $n$ 
13:       $\vec{z}, \vec{\phi} \leftarrow g^{-1}(\vec{\theta}_i)$ 
14:
15:       $m \leftarrow n_i - \Delta n$ 
16:       $\mathcal{L}' \leftarrow \mathcal{L}(\vec{d}|m, \vec{\phi})$ 
17:       $\alpha \leftarrow \min(1, A_{\text{jump}}(n_i, \vec{\theta}_i \rightarrow m, \vec{\phi}, \vec{z}))$ 
18:    else  $\triangleright$  Otherwise, attempt a standard MCMC move
19:       $\vec{\phi} \leftarrow$  random sample from  $\gamma(\vec{\theta}|\vec{\theta}_i)$ 
20:       $m \leftarrow n_i$ 
21:       $\mathcal{L}' \leftarrow \mathcal{L}(\vec{d}|m, \vec{\phi})$ 
22:       $\alpha \leftarrow \min(1, A(\vec{\theta}_i \rightarrow \vec{\phi}))$ 
23:     $u \leftarrow \text{uniform}(0, 1)$ 
24:    if  $u < \alpha$  then
25:       $\vec{\theta}_{i+1}, n_{i+1}, \mathcal{L}_{i+1} \leftarrow \vec{\phi}, m, \mathcal{L}'$ 
26:    else
27:       $\vec{\theta}_{i+1}, n_{i+1}, \mathcal{L}_{i+1} \leftarrow \vec{\theta}_i, n_i, \mathcal{L}_i$ 

```

in terms of the integrated prior mass measure $d\pi$. The prior mass, $P(\lambda)$, is calculated through the integral of the prior over a region of parameter space where the likelihood meets a threshold λ :

$$P(\lambda) = \int_{\mathcal{L}(\vec{d}|\vec{\theta}) > \lambda} \pi(\vec{\theta}) d\vec{\theta}. \quad (2.66)$$

By doing so, the marginal likelihood is effectively transformed from a Riemann integral to a Lebesgue integral.

The `MultiNest` software library [101] implements this algorithm. The prior mass is estimated by a number of ellipsoids, determined by the current set of samples and a clustering algorithm. At each iteration, these ellipsoids are shrunk by the removal of the lowest like-

likelihood sample, and a replacement is sampled. The current estimate of the marginal likelihood is then updated based on the prior mass volume and the lowest likelihood sample. As part of the sampling process, equally weighted samples from the posterior distribution are also generated and can be retrieved.

3

Global fits to sterile neutrino models

The well-established discoveries of neutrino mass and three-active-flavor mixing can be phenomenologically incorporated into the Standard Model as seen in chapter 1. This results in a model – called the “ ν SM” – that successfully predicts neutrino oscillations in many experiments. However, the masses and mixings have been incorporated in an ad hoc manner in analogy to similar physics in the quark sector. This naturally leads one to consider if there is more “new physics” in the neutrino sector that is yet to be discovered that can give us a clearer picture of the underlying theory.

3.1 Anomalies

Just as experimental results of solar neutrino mixing and atmospheric neutrino oscillation led to the ν SM, the path to this new sector may lie in unexplained anomalies in modern neutrino experiments.

So-called “short base-line” (SBL) neutrino experiments have the most promise in this regard. From the two-neutrino approximation oscillation formula,

$$P_{\alpha \rightarrow \beta}(L) \approx \delta_{\alpha \beta} - (2\delta_{\alpha \beta} - 1) \sin^2(2\theta) \sin^2 \left(\left[1.27 \frac{\text{GeV}}{\text{eV}^2 \text{km}} \right] \frac{\Delta m_{ij}^2 L}{E} \right), \quad (3.1)$$

we can see that the first oscillation maximum is at

$$\left[1.27 \frac{\text{GeV}}{\text{eV}^2 \text{km}} \right] \frac{\Delta m_{ij}^2 L}{E} = \frac{\pi}{2}. \quad (3.2)$$

For a given E , a large value of Δm_{ij}^2 results in an oscillation maximum at a comparatively smaller L . Thus, experiments with a short base-lines explore energy scales outside of the well mapped region of atmospheric and solar squared mass splittings. In comparison, *long base-line* experiments are designed to observe standard ν SM three-neutrino oscillations.

3.1.1 LSND

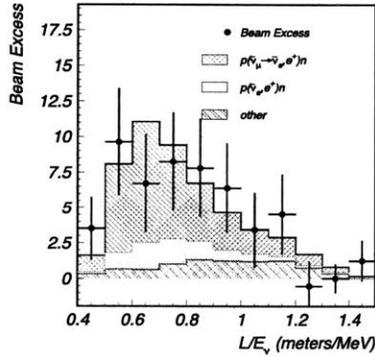


Figure 3.2:

The number of excess electron anti-neutrino events seen by LSND as a function of energy. The red and the green histograms show beam related backgrounds. The blue histogram shows a fit to a $3 + 1$ model. Figure included from Aguilar et al. [21].

The Liquid Scintillator Neutrino Detector (LSND) was an experiment located on the Los Alamos Meson Physics Facility (LAMPF) beam-line at Los Alamos National Laboratory. This detector was designed and constructed before atmospheric neutrino oscillations were confirmed by Super-Kamiokande [36, 172]. LAMPF was a decay at rest neutrino source, producing muon neutrinos, muon anti-neutrinos, and electron neutrinos, but critically, no electron anti-neutrinos. The detector is itself designed to look for inverse β -decay, with the recoiling positron producing light in the scintillator. As inverse β -decay is initiated by an electron anti-neutrino, the experiment was designed to observe muon anti-neutrinos produced by the beam-line oscillating into electron anti-neutrinos — an *appearance* experiment.

The energy for a LAMPF muon anti-neutrino lies on a spectrum from 20 to 52 MeV, governed by the muon decay. LSND was located 30 meters from the beam target, so that for squared mass splittings of $\mathcal{O}(1 \text{ eV}^2)$, the first oscillation maximum would be near the detector. The experiment observed a 3.8σ excess of $87.9 \pm 22.4 \pm 6.0$ electron anti-neutrino events above their background [21], for which the best fit for an oscillation hypothesis gave a squared mass splitting of 1.2 eV^2 .

3.1.2 MiniBooNE

The MiniBooNE experiment was designed to test the LSND anomaly using a completely different set of detector and beam technologies. The detector sits in the booster neutrino beam-line at Fermi National Accelerator Laboratory. The beam-line follows the general description given in chapter 1, allowing the selection of either muon neutrinos, or muon anti-neutrinos from the decay in flight of charged pions. MiniBooNE was designed as a mineral oil Cherenkov detector, with a much greater volume compared to LSND.

MiniBooNE searched for anomalous appearance of electron neutrinos and anti-neutrinos from the muon neutrino and anti-neutrino beam. As the detector is also sensitive to muons, the collaboration searched for anomalous disappearance of muon neutrinos and anti-neutrinos. If the flavour mixing of neutrinos is unitary, any appearance of electron-neutrinos must have a corresponding disappearance of muon neutrinos³⁷.

The experiment observed an excess of low energy electron neutrino events in both neutrino and anti-neutrino beam modes [16, 17, 19]. However, no anomalous disappearance of the muon neutrinos or anti-neutrino was found [66]. The electron neutrino and electron anti-neutrino anomalies were in marginal and good agreement with the LSND result, respectively.

3.1.3 Nuclear reactors

Nuclear reactors provide an incredibly intense source of electron anti-neutrinos, which makes them promising for observing neutrino oscillations. These experiments look for an oscillation induced deficit in the prediction of the expected number of neutrino events. Some reactor experiments also have a *near* detector close to the reactor itself, to measure the neutrino flux before any oscillation has occurred.

Most reactor experiments are long base-line, and are not optimised for high squared mass splittings. For a fixed L , the oscillation probability is governed by a sinusoidal term:

$$\sin^2\left(\omega\frac{1}{E}\right) \quad \omega = \left[1.27\frac{\text{GeV}}{\text{eV}^2\text{km}}\right] \Delta m_{ij}^2 L, \quad (3.3)$$

with frequency ω . If L is large, so is the frequency, and oscillations in the energy spectrum will become rapid. For high squared mass splittings, these oscillations have a smaller period than the energy resolution of the detector, and so the sinusoidal term is effectively averaged out:

$$\left[1.27\frac{\text{GeV}}{\text{eV}^2\text{km}}\right] \frac{\Delta m_{ij}^2 L}{E} \rightarrow \frac{1}{2}. \quad (3.4)$$

Thus, long base-line reactor experiments can constrain high squared mass splittings by their total event rate only. Prior to 2011, this rate was slightly below predictions based on β -decay electron spectra, but within the theoretical uncertainty. A re-analysis of the neutrino flux predictions increased the expected event rate, widening the gap to a statistically significant level. As a fraction of the predictions, the measured rate dropped from 0.976 ± 0.024 to 0.943 ± 0.023 [184] and is known as the *reactor anomaly*.

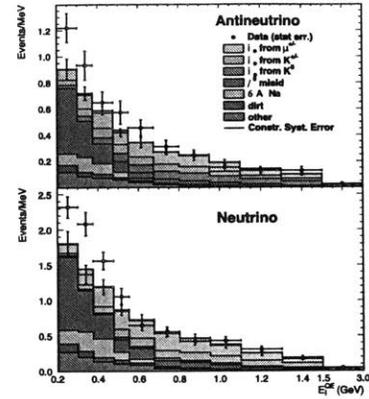


Figure 3.2:

The number of excess electron neutrino and anti-neutrino events seen by MiniBooNE as a function of energy. The stacked histograms show various backgrounds. A clear excess of events above these backgrounds is observed at low energy. Figure included from Aguilar-Arevalo and others [18].

³⁷ Or tau neutrinos, but none are present in this beam-line.

³⁸ As all were located with different base-lines.

While it is tempting to attribute this to potentially new physical processes, there are some important caveats to consider. As discussed in chapter 1, nuclear reactors produce neutrinos through long chains of decaying isotopes. This process involves ~ 6000 nuclear transitions [131], many of which traverse the poorly explored neutron rich isotopes. The most recent evidence of this uncertainty was the appearance of an excess of neutrino events at ~ 5 MeV in three long base-line neutrino experiments. The common energy to all three detector implies this cannot be a product of neutrino oscillation³⁸, and more detailed calculations of fissile isotopes and β -decay processes are beginning to resolve the issue [87, 146].

Bugey was a neutrino oscillation experiment that made use of the Bugey nuclear power plant. It was a true short base-line experiment, with detectors placed 15, 40, and 95 meters from the reactor core. Although the experiment collaboration reported no evidence of neutrino oscillations [10], modern analyses suggest a weak anomalous disappearance of electron anti-neutrinos at ~ 1 eV²).

3.1.4 Electron capture calibration

GALLEX and SAGE were solar neutrino experiments in the same style as the Homestake experiment, but using ⁷¹Ga instead of ³⁷Cl.

The clean, mono-energetic electron capture neutrino spectrums of ⁵¹Cr and ³⁷Ar made them ideal for calibration of neutrino detectors.

In GALLEX, only the ⁵¹Cr source was used, and was placed at two different locations inside the detector [156]. In SAGE, both electron-capture isotopes were used [9].

A theoretical expectation for the number of interacting neutrinos was constructed from predictions of the neutrino scattering cross section. When the position of the ⁵¹Cr source was changed in GALLEX, a significant reduction was observed in the ratio of measured flux to predicted flux. A similar phenomenon were observed in SAGE when the higher energy ⁵¹Cr source was switched with the lower energy ³⁷Ar. These results can be interpreted as anomalies, due to the relative simplicity of the test. A fit for an oscillation hypothesis gives a squared mass splitting of ~ 2 eV².

3.2 Sterile neutrinos

The previously discussed anomalies point toward a potential oscillation effect with a squared mass splitting of ~ 1 eV². This is in stark contrast with the solar and atmospheric squared mass splittings that are $\mathcal{O}(10^{-5}$ eV²) and $\mathcal{O}(10^{-3}$ eV²), respectively. This would require revision of the ν SM, as three neutrino mass states cannot have

three distinct squared mass splittings. A fourth neutrino mass state is required if we accept the additional squared mass splitting hypothesis. This new mass state in turn requires an additional flavour state.

However, the measurement of the Z boson decay width discussed in chapter 1 is conclusive that three flavour states participate in the Z decay. This new mass state cannot be excluded from the Z decay on kinematic arguments, as it will have a mass of ~ 1 eV based on present limits of the electron neutrino mass. Thus, it must be excluded from the weak interaction altogether.

This *sterile neutrino* would only interact with other particles via gravity and Yukawa couplings³⁹ to the Higgs boson. This addition of one sterile neutrino flavour state to the existing three active neutrino flavour states is called a “3 + 1” model.

³⁹ As the sterile neutrino acquires mass from its interaction with the Higgs boson.

3.2.1 Incorporation into the ν SM

The additional mass and flavour states extend the PMNS matrix to a unitary 4×4 matrix [79]:

$$\mathbf{U}_{3+1} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ \vdots & & \vdots & U_{\mu 4} \\ \vdots & & \vdots & U_{\tau 4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{bmatrix}. \quad (3.5)$$

This introduces seven new matrix elements, of which four, U_{s1}, \dots, U_{s4} , cannot be directly constrained by experiment due to the non-interacting nature of the fourth sterile flavour state. The matrix is assumed to be unitary, and the magnitude of the new elements can be constrained by the current measurements of unitarity of the PMNS matrix[197]. The new degrees of freedom can be parameterised by introducing three new neutrino mixing angles θ_{i4} and two new CP -violating phases. Equation 1.23 still holds in describing oscillations, but now the indices i, j run up to 4.

The effect of the two small solar and atmospheric splittings on an experiment designed to look for $O(1 \text{ eV}^2)$ scale splitting will be negligible. Therefore, we can assume them to be degenerate:

$$\Delta m_{21}^2 = \Delta m_{32}^2 = 0. \quad (3.6)$$

This is called the “short base-line approximation”.

The oscillation probability formula for $\nu_\alpha \rightarrow \nu_\beta$ in the 3 + 1 model then reduces to

$$P_{\alpha \rightarrow \beta} = \delta_{\alpha\beta} - 4(\delta_{\alpha\beta} - U_{\alpha 4} U_{\beta 4}^*) U_{\alpha 4}^* U_{\beta 4} \sin^2 \left(\left[\frac{1.27 \text{ GeV}}{\text{eV}^2 \text{ km}} \right] \frac{\Delta m_{41}^2 L}{E} \right), \quad (3.7)$$

where Δm_{41}^2 is the squared mass splitting associated with the new, sterile neutrino induced, neutrino oscillations. With any particular selection of α and β , this equation can be seen to be equivalent to a simple two neutrino model with a mixing amplitude of $\sin^2 2\theta_{\alpha\beta} = 4(\delta_{\alpha\beta} - U_{\alpha 4}U_{\beta 4}^*)U_{\alpha 4}^*U_{\beta 4}$.

More generally, for a $3 + N$ model incorporating N sterile neutrinos, the complex phases of \mathbf{U} must be taken into account. Let

$$\Phi_{\alpha\beta ij} = \arg(U_{\alpha i}U_{\beta i}^*U_{\alpha j}^*U_{\beta j}). \quad (3.8)$$

These are the CP -violating phases: as in three neutrino ν SM oscillation, a CP transformation of $\nu \rightarrow \bar{\nu}$ causes $\Phi \rightarrow -\Phi$, allowing a difference between neutrino and anti-neutrino oscillations. The probability of oscillation for a $3 + N$ model can then be written as

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \delta_{\alpha\beta} \\ &- 4 \sum_{j>3} \left(\delta_{\alpha\beta} - \sum_{i \geq j} |U_{\alpha i}| |U_{\beta i}| \cos \Phi_{\alpha\beta ij} \right) |U_{\alpha j}| |U_{\beta j}| \sin^2 \left(\left[\frac{1.27 \text{ GeV}}{\text{eV}^2 \text{ km}} \right] \frac{\Delta m_{ij}^2 L}{E} \right) \\ &+ 2 \sum_{j>3} \left(\delta_{\alpha\beta} - \sum_{i \geq j} |U_{\alpha i}| |U_{\beta i}| \sin \Phi_{\alpha\beta ij} \right) |U_{\alpha j}| |U_{\beta j}| \sin \left(\left[\frac{2.54 \text{ GeV}}{\text{eV}^2 \text{ km}} \right] \frac{\Delta m_{ij}^2 L}{E} \right). \end{aligned} \quad (3.9)$$

For $N > 1$ sterile neutrinos, the SBL experiments are sensitive to the ordering of the mass states through the non-squared sine term.

3.3 Global fit

The experiments discussed in section 3.1 are by no means an exhaustive list. Like the MiniBooNE muon neutrino disappearance analysis, other short base-line experiments have failed to detect anomalies. For example, CDHS was a neutrino detector placed in the ~ 1 GeV CERN Super Proton Synchrotron neutrino beam-line. With both a near and far detector, it was able to constrain the initial flux of the neutrino beam and then measure any potential oscillations after traveling 750 meters. KARMEN is an experiment of very similar design to LSND. Like LSND, it used a decay at rest beam of neutrinos to measure potential muon anti-neutrino to electron anti-neutrino oscillations. Neither CDHS or KARMEN observed anomalies, placing limits on the oscillation parameters in the high squared mass splitting regime.

These experiments present very disparate results, some do not see anomalies, and some do. Even among those that do see anomalies, the results of a model fit to a neutrino oscillation hypothesis are not completely mutually compatible.

To make sense of this picture, short-baseline neutrino oscillation results can be combined into a *global fit*. This fit considers only one hypothesis, and computes the likelihood that the observed experimental data from all experiments match this hypothesis. A combined fit allows those results with more statistical significance to drive the final result, and naturally reduces the influence of weaker results.

This analysis builds on the work from Conrad et al. [79]. The following additions have been made:

- All software was re-written from the original Fortran into C++.
- Software design encapsulates common tasks in shared routines, improving maintainability.
- Oscillation model was abstracted to allow the addition of more exotic models.
- Model space is explored using an affine invariant parallel tempering hybrid MCMC, allowing for Bayesian and frequentist interpretations.
- The addition of the SciBooNE/MiniBooNE joint muon neutrino disappearance data set.
- Improved description of the LSND experiment.

3.3.1 Inference framework

The goal is to test two hypotheses, with one and two sterile neutrinos:

- The 3+1 model has three parameters which can be constrained by our present data: U_{e4} , $U_{\mu4}$ and Δm_{14}^2 . The third mixing matrix element $U_{\tau4}$ has constraints in only the parameter space of $\Delta m_{14}^2 > 10 \text{ eV}^2$, which is outside our window of interest, so this element is ignored⁴⁰.
- The 3 + 2 model has seven parameters: two squared mass splittings, four mixing matrix elements, and one CP -violating phase.

The null hypothesis – where no sterile neutrinos exist – can be considered a nested special case of either the 3+1 or 3+2 models, where all the mixing matrix elements are set to zero: $U_{ei} = U_{\mu i} = 0$ for all i .

A mainly frequentist approach is used, with χ^2 test statistics and confidence regions are constructed. This is motivated by a desire

⁴⁰ Effectively, this parameter has been profiled out in the frequentist sense, or marginalised out in the bayesian construction.

to make direct comparisons to previous sterile neutrino global fit analyses, which use frequentist methods. However, evaluating the likelihood function over a seven dimensional grid to sufficient accuracy would require more computational resources than even modern super-computers can employ. Thus, the Markov chain Monte-Carlo sampling algorithm discussed in chapter 2 is used. Posterior samples are a direct consequence of this algorithm, and so the posterior distribution can also be shown.

3.3.2 MCMC implementation

The posterior is sampled using the parallel tempering affine invariant hybrid MCMC described in sections 2.4.1 and 2.4.2. Five temperature ensembles were used, each containing an ensemble of 20 individual chains. Replica exchanges were performed at a rate of 10%, and each temperate ensemble was a factor of $\sqrt{2}$ higher in temperature than the previous ensemble. Samples were drawn only from the $\beta = 0$ ensemble, and the first 200 samples were discarded as a burn-in period.

Although the vanilla Metropolis-Hastings MCMC is a purely sequential algorithm, these ensemble extensions permit a degree of parallelisation. Each member of the temperature ensembles can be updated independently of each other, with synchronisation only required for the occasional replica exchange. The affine invariant algorithm requires slightly more care, as only half of the ensemble can be updated independently [105]; the other half can only be updated sequentially after the first.

The Message Passing Interface (MPI) [108] is a common software library for parallelisation in scientific computing. It provides functionality for simple communication tasks between processors, allowing, for example, MCMC chains running on one CPU to request the current state of chains on another CPU. This interface extends beyond message passing on a single machine, and seamlessly enables communication between CPUs on multiple machines over a network.

MPI maintains a decentralised approach, with no main controlling process. This is ideal for workloads where individual CPUs can be mapped to localised regions of the problem space, the CPUs generally communicate to neighbouring CPUs, and high bandwidth communication requirements make a centralised controller impractical. These benefits come with the downside that all CPUs that will participate in the computation must be present and accounted for when the computation starts. This requirement is naturally met by most super-computer environments, but does not map well to cluster computing centers.

In a cluster, CPUs are allocated to a computational task opportunistically as they finish their previous work-loads. Thus, to use MPI, the cluster must reserve a pool of CPUs, and keep them idle until enough CPUs are available to begin the task. As this is wasteful of computational resources, MPI functionality is often disabled in computer clusters.

Instead, parallelisation of the MCMC across a computer cluster was achieved with a custom centralised work queue. These queues are more often found in web services, where the number of clients making requests varies continuously. The MCMC is run by a single master process, which populates a queue with the likelihood evaluations that are required for the chain to progress, and the cluster is instructed to run worker tasks. As the cluster allocates CPUs to this worker task, they connect to the master process and announce their presence. As the likelihood is a stateless function – taking only the model parameters as inputs and returning only the likelihood⁴¹ as output – the next likelihood evaluation in the queue can be assigned to the next available worker.

This queue was implemented with the zeromq distributed messaging library [86] to handle network communication. The msgpack library [185] was used to serialise the model parameters for transmission using zeromq.

3.3.3 Likelihood function

In all experiments under consideration here, the data is measured as a number of events that land in a bin. Thus, the prediction function $\Psi(\vec{\theta})$ is chosen to compute the expected number of events that would land in a bin given the sterile neutrino model parameters $\vec{\theta}$.

The experiments under consideration here come in two varieties. The first kind are high statistics experiments, where the data will be approximately normally distributed. In this case, the log-likelihood for experiment ρ , with measured data vector \vec{d}_ρ of bin contents, is the normal distribution:

$$\ln \mathcal{L}_\rho(\vec{d}_\rho | \vec{\theta}) = -\frac{1}{2} [\vec{d}_\rho - \vec{\Psi}_\rho(\vec{\theta})]^T \Sigma^{-1} [\vec{d}_\rho - \vec{\Psi}_\rho(\vec{\theta})] - \frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma|, \quad (3.10)$$

where $\vec{\Psi}_\rho$ is the prediction function for experiment ρ , N_ρ is the number of bins⁴², and Σ is the covariance matrix of the measured data.

The second kind are low statistics experiments, where the number of counts per bin is usually less than 20. In this case, the log-likelihood for experiment ρ is the Poisson distribution:

⁴¹ And χ^2 values.

⁴² And thus N_ρ is also the number of dimensions in \vec{d}_ρ and $\vec{\Psi}_\rho$.

$$\ln \mathcal{L}_\rho(\vec{d}_\rho|\vec{\theta}) = - \sum_i^N \left([\vec{\Psi}_\rho(\vec{\theta})]_i - [\vec{d}_\rho]_i \ln([\vec{\Psi}_\rho(\vec{\theta})]_i) + \ln \Gamma(1 + [\vec{d}_\rho]_i) \right), \quad (3.11)$$

where $[\vec{d}_\rho]_i$ and $[\vec{\Psi}_\rho(\vec{\theta})]_i$ are the i^{th} components of the data vector, and prediction function, respectively.

The global log-likelihood is then

$$\ln \mathcal{L}(\vec{d}|\vec{\theta}) = \sum_\rho \ln \mathcal{L}_\rho(\vec{d}_\rho|\vec{\theta}). \quad (3.12)$$

In this analysis, the χ^2 values are also computed and saved alongside the likelihood values. For high statistics experiments, this is the standard χ^2 :

$$\chi_\rho^2(\vec{\theta}) = [\vec{d}_\rho - \vec{\Psi}_\rho(\vec{\theta})]^T \Sigma^{-1} [\vec{d}_\rho - \vec{\Psi}_\rho(\vec{\theta})]. \quad (3.13)$$

For low statistics experiments, a saturated Poisson distribution was substituted for the χ^2 :

$$\chi_\rho^2(\vec{\theta}) = -2 \left(\ln \mathcal{L}_\rho(\vec{d}_\rho|\vec{\theta}) - \ln P(\vec{d}_\rho|\vec{d}_\rho) \right). \quad (3.14)$$

where $P(\vec{x}|\vec{\lambda})$ is the Poisson distribution.

The global χ^2 is thus

$$\chi^2(\vec{\theta}) = \sum_\rho \chi_\rho^2(\vec{\theta}). \quad (3.15)$$

Priors

All priors were chosen to be uniform in logarithmic parameter space. As the model parameters are continuous, this choice corresponds to an exponentially falling prior in linear parameter space. This prior was chosen to allow the global fit to explore a wide scale of Δm^2 and matrix element values.

The parameter space was limited as follows: The matrix elements were required to lie within the space of unitary matrices and be larger than 10^{-6} . The phases were required to be less than 2π . Large Δm^2 parameters require much more computing time to evaluate, which slows down the entire ensemble. Therefore, the Δm^2 parameters are required to be between 10^{-4} eV² and 10^4 eV² for $3 + 1$. In the case of additional sterile neutrinos, this was narrowed to 10^{-3} eV² and 10^3 eV². Proposed steps outside these listed boundaries will have a prior of zero.

3.3.4 Predictions

The predicted number of events for each experiment were calculated according to the methods prescribed in their respective publications. Many experiments use isotropic neutrino sources, which admit analytic formulae of the flux of un-oscillated⁴³ neutrinos at the detector for some energy and position inside the detector:

$$\Phi \propto \frac{P(\nu_\alpha \rightarrow \nu_\beta)}{L^2}. \quad (3.16)$$

The total number of events seen in an energy bin E is then related to the integral of this flux:

$$\Psi \propto \int_E^{E+\Delta E} \int_L^{L+\Delta L} \Phi dE' dL'. \quad (3.17)$$

As the oscillation probability is itself a function of a squared sinusoid, the length integration has an analytic solution in terms of exponential integrals. The remaining energy integral is performed numerically, along with an additional integral – if necessary – to account for detector energy resolution.

Neutrino beam-lines have no such analytic expressions for their flux, and these experiments use a pre-computed flux at the detector. In the case of MiniBooNE and SciBooNE, a list of simulated beam events were provided⁴⁴. Each event – which contains a neutrino type, energy and propagation length – is individually re-weighted according to the oscillation probability and then binned.

3.3.5 Presentation of results

All χ^2 values that were computed (including those for rejected proposals and the burn-in period) are saved in a set \mathcal{X} . The posterior samples (after removal of the burn-in period) are saved in a set \mathcal{S} .

The posterior can be estimated from these samples according to the construction in chapter 2:

$$\tilde{p}(\vec{\theta}|\vec{d}) = \frac{1}{|\mathcal{S}|} \sum_{\vec{\theta}' \in \mathcal{S}} \delta(\vec{\theta}' - \vec{\theta}). \quad (3.18)$$

To plot the posterior in a 2D plane, all but two parameters must be marginalised out of the distribution. With loss of generality, let the two desired parameters be θ_a and θ_b , with the remainder as θ_c, \dots . The marginalised posterior is thus

$$p(\theta_a, \theta_b|\vec{d}) = \left[\prod_{x=c, \dots} d\theta_x \pi(\theta_x) \right] p(\theta_a, \theta_b, \theta_c, \dots|\vec{d}). \quad (3.19)$$

⁴³ That is, the flux before considering oscillation effects. In the case of an $\nu_\mu \rightarrow \nu_e$ appearance experiment, this is the flux of muon neutrinos.

⁴⁴ This representation of the flux takes into account differing oscillation lengths due to varying propagation lengths due to where the neutrinos are generated in the decay pipe and where they interact inside the detector.

From this, it is clear that the marginalised posterior estimated from the samples is

$$\tilde{p}(\theta_a, \theta_b | \vec{d}) = \frac{1}{|S|} \sum_{\vec{\theta}' \in S} \delta(\theta'_a - \theta_a) \delta(\theta'_b - \theta_b). \quad (3.20)$$

From this, 90% and 99% credible regions are created.

Model selection

The best fit χ^2 is found from the samples:

$$\chi_{\min}^2 = \inf \{ \chi^2 \in \mathcal{X} \}. \quad (3.21)$$

Then, given χ_{null}^2 – the value of the χ^2 computed for the null-hypothesis – the model comparison $\Delta\chi^2$ between the best fit and the null hypothesis is given by

$$\Delta\chi_{\text{null-min}}^2 = \chi_{\text{null}}^2 - \chi_{\min}^2. \quad (3.22)$$

The $1 - \alpha$ confidence region is given by the subset of all points that satisfy

$$\mathcal{C} = \{ \vec{\theta} \mid \forall \chi_{\vec{\theta}}^2 \in \mathcal{X} : \Delta\chi^2(\vec{\theta}) < \text{CDF}_{\chi^2}^{-1}(k, 1 - \alpha) \}. \quad (3.23)$$

where k is the number of degrees of freedom. Regions of 90% and 99% confidence are used, and the 90% region is displayed on top of the 99% region. This profiles the χ^2 , and so the number of degrees of freedom k is two for a 2D plot.

3.3.6 Experimental data sets

The full list of experiments included in this study is provided in table 3.1. Most data sets used in the past analysis (Conrad et al. [79]) have been incorporated into this analysis, however the atmospheric data set was removed as it provided only a weak constraint and the implementation of the likelihood required a non-standard parameterisation for the sterile neutrino oscillation model.

The MiniBooNE disappearance data sets that were used previously have been replaced by the MiniBooNE/SciBooNE joint disappearance analyses, which are more restrictive. The MiniBooNE/SciBooNE data sets in neutrino mode [178] and anti-neutrino mode [48] were taken from the public data release for each analysis. However, for the neutrino data set, an updated covariance matrix was used, along with a cosmic background data set omitted from the data release at that time [20].

This analysis does not include data from cosmology because the cosmic microwave background and large scale structure constraints

Tag	Process	ν vs. $\bar{\nu}$	Type	N_{bins}
LSND [21]	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	$\bar{\nu}$	App	5
KARMEN [155]	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	$\bar{\nu}$	App	9
KARMEN/LSND(xsec) [78]	$\nu_e \rightarrow \nu_e$	ν	Dis	11
BNB-MiniBooNE- ν [16, 19]	$\nu_\mu \rightarrow \nu_e$	ν	App	19
BNB-MiniBooNE- $\bar{\nu}$ [17, 18]	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	$\bar{\nu}$	App	19
NuMI-MB(ν_{app}) [13]	$\nu_\mu \rightarrow \nu_e$	ν	App	10
Bugey [10, 184]	$\bar{\nu}_e \rightarrow \bar{\nu}_e$	$\bar{\nu}$	Dis	60
Gallium [9, 156]	$\nu_e \rightarrow \nu_e$	ν	Dis	4
BNB-MiniBooNE/SciBooNE- ν [178]	$\nu_\mu \rightarrow \nu_\mu$	ν	Dis	48
BNB-MiniBooNE/SciBooNE- $\bar{\nu}$ [48]	$\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$	$\bar{\nu}$	Dis	42
NOMAD [34]	$\nu_\mu \rightarrow \nu_e$	ν	App	30
CCFR84 [222]	$\nu_\mu \rightarrow \nu_\mu$	ν	Dis	18
CDHS [88]	$\nu_\mu \rightarrow \nu_\mu$	ν	Dis	15
MINOS-CC [174, 175]	$\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$	$\bar{\nu}$	Dis	25

on the presence of a fourth neutrino are model dependent, including the assumption of a standard thermal history for the Universe [7]. Sterile neutrino thermalisation can also be suppressed in a number of plausible scenarios as discussed in Abazajian [6], Bento and Berezhiani [40], Chu and Cirelli [50], Dasgupta and Kopp [84], Foot and Volkas [104], Gelmini, Palomares-Ruiz, and Pascoli [112], Hamann et al. [126], Hannestad, Hansen, and Tram [127], and Ho and Scherrer [141]. Thermalisation may not occur when one considers models with four-neutrino mixing [8], and introducing a very weak pseudo-scalar interactions that are unobservable in the SBL experiments can not only resolve the apparent disagreement between the $3+1$ models and cosmic microwave background, but also predict a Hubble constant in agreement with local measurements [28]. Based on these points, it is more informative to consider cosmological constraints separately and then make comparisons to the results derived from this analysis.

3.4 Results

3.4.1 $3+1$ model

Confidence regions for a $3+1$ model are shown in figure 3.3. The quality of the fits is described in table 3.2 and the parameters of the best fit points are provided in table 3.3.

The top row shows fits for appearance ($\nu_\mu \rightarrow \nu_e$) and disappearance – including both muon and electron flavour – data sets separately, as a function of $\sin^2 2\theta_{\mu e}$ and Δm^2 plane. Note that there is no overlap between the 90% (red) or 99% (blue) confidence level

Table 3.1:

Data sets used in this analysis, including the relevant oscillation process, neutrino vs. antineutrino type, appearance vs. disappearance type, and the number of bins in each dataset. See Conrad et al. [79] and Ignarra [147] for additional details on each experiment identified by the “tag.”

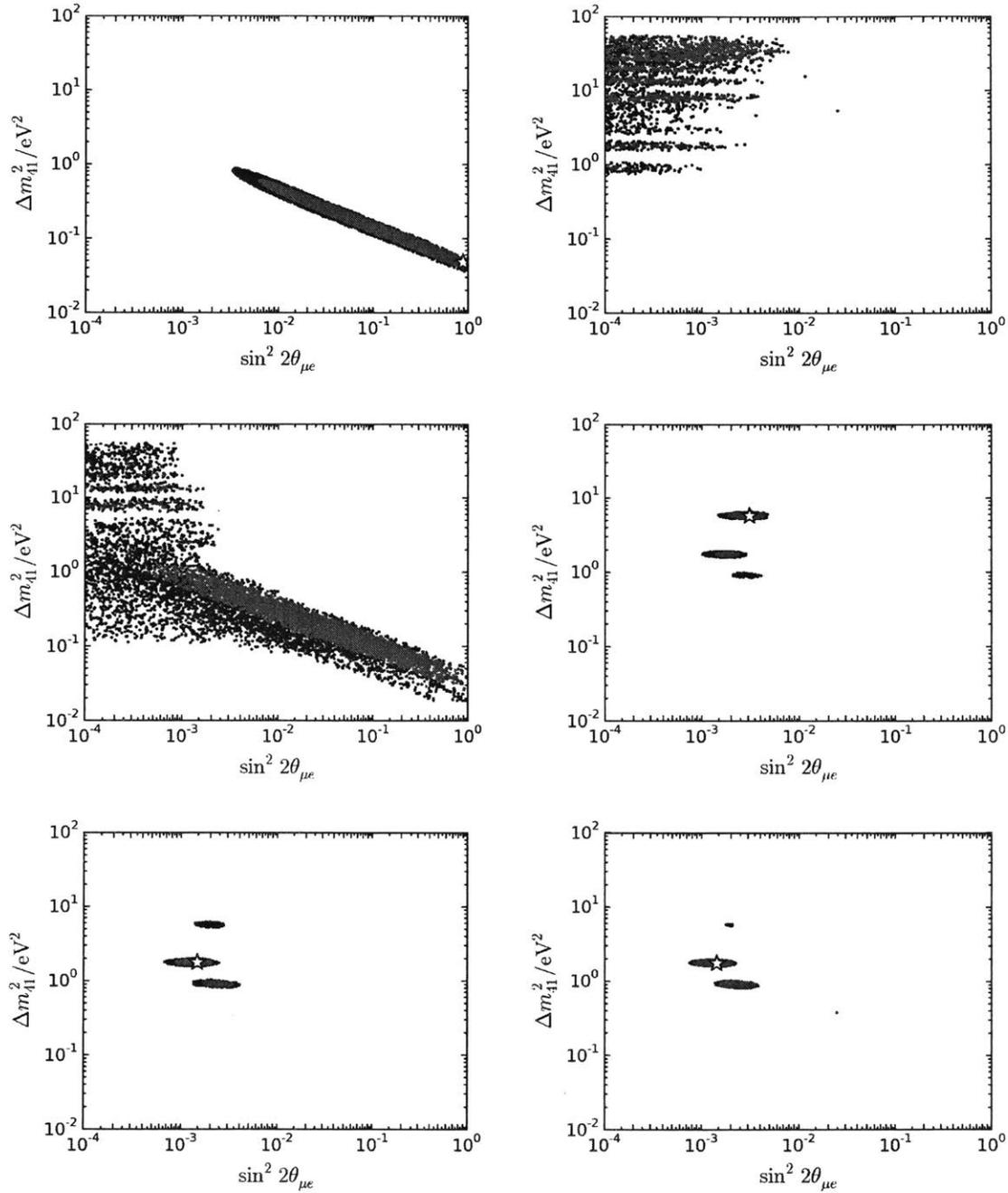


Figure 3.3: Frequentist confidence regions for a 3 + 1 model using appearance only data (top left), disappearance data (top right), neutrino data (mid left), anti-neutrino data (mid right), and global data (bottom left). The Bayesian credible intervals for 3 + 1 global data are shown bottom right. In these plots, $\sin^2 2\theta_{\mu e} = 4|U_{e4}|^2|U_{\mu4}|^2$. Red indicates 90% regions and blue indicates 99% regions.

	N_{bins}	χ_{min}^2	χ_{null}^2	$\Delta\chi_{\text{null-min}}^2$ (dof)
3+1				
All	315	306.81	359.15	52.34 (3)
App	92	88.04	150.84	62.80 (2)
Dis	223	195.84	208.32	12.48 (3)
ν	155	153.18	164.57	11.39 (3)
$\bar{\nu}$	157	138.79	194.59	55.8 (3)
3+2				
All	315	302.16	359.15	56.99 (7)

3+1	Δm_{41}^2	$ U_{e4} $	$ U_{\mu4} $
All	1.75	0.163	0.117
App	4.75×10^{-2}	0.743	0.638
Dis	7.79	0.217	2.94×10^{-2}
ν	7.71	0.248	5.67×10^{-2}
$\bar{\nu}$	5.73	0.199	0.140

3+2	Δm_{41}^2	Δm_{51}^2	$ U_{e4} $	$ U_{\mu4} $	$ U_{e5} $	$ U_{\mu5} $	ϕ_{54}
All	0.475	0.861	0.120	0.177	0.141	0.111	0.0662π

(CL) regions when the data sets are divided in this manner. Therefore, the appearance and disappearance data sets have little compatibility, producing tension in the model.

The middle row shows the neutrino (left) and anti-neutrino (right) data sets fit separately within a 3 + 1 model. Dividing the data in this manner, there is overlap between the two data sets, however the anti-neutrino data sets are highly restrictive.

The global fit for all data sets is shown on the bottom, and has two 90% allowed regions. This is in contrast to the single 90% allowed region shown in the previous analysis of Conrad et al. [79]. Both share a region at $\sim 1 \text{ eV}^2$, while the new result has an additional region at $\sim 1.7 \text{ eV}^2$. This new region is a consequence of the improved description of LSND. Compared to the previous analysis, the best fit has moved to the $\sim 1.7 \text{ eV}^2$ region. This was caused by the addition of the SciBooNE/MiniBooNE disappearance data sets.

The credible intervals of the Bayesian fit are shown in the bottom right of figure 3.3. The 90% Bayesian credible intervals largely overlap with the 90% frequentist confidence intervals shown in the bottom left plot.

Table 3.2:

The χ^2 values, degrees of freedom (dof) and probabilities associated with the best-fit and null hypothesis in each scenario: all datasets (All), appearance only (App), disappearance only (Dis), neutrino only (ν), anti-neutrino only ($\bar{\nu}$).

Table 3.3:

The oscillation parameter best-fit points in each scenario considered. The values of Δm^2 shown are in units of eV^2 .

3.4.2 3 + 2 model

To relieve tension in the 3 + 1 model, we can use the additional descriptive freedom granted by a 3 + 2 model. The frequentist global fit for this result is shown in figure 3.4. This model has seven parameters, and so we select some examples to illustrate the allowed parameter space. The left hand side of figure 3.4 shows the parameter space of the two squared mass splittings. The best fit solution is detailed in table 3.3, and was found to have both splittings less than 1 eV^2 .

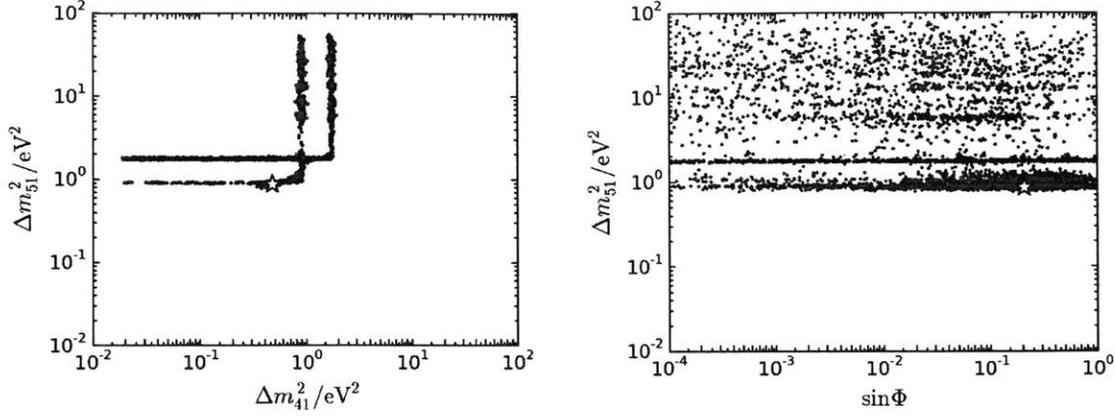
In the region of $\Delta m_{41}^2 \sim 1 \text{ eV}^2$, there is a large sweeping degeneracy in Δm_{51}^2 , which switches to Δm_{41}^2 when $\Delta m_{41}^2 \approx \Delta m_{51}^2$. This kink in the plot is purely an artefact of the $\Delta m_{51}^2 > \Delta m_{41}^2$ constraint applied in the fitting software to reduce the size of the model space. This degeneracy should be interpreted as the fit highly constraining one squared mass splitting to be equal to either of the allowed regions observed in the 3 + 1 model, while the second splitting is unconstrained at the 99% CL level.

Thus, while this result might appear at first glance to be a dramatic change from the 3 + 2 model results shown in Conrad et al. [79], which found best fit values for 3 + 2 of $\Delta m_{41}^2 = 0.92 \text{ eV}^2$ and $\Delta m_{51}^2 = 17 \text{ eV}^2$, in fact, this is actually a product of the poorly constrained model space. The previous best fit from Conrad et al. [79] remains within the allowed region.

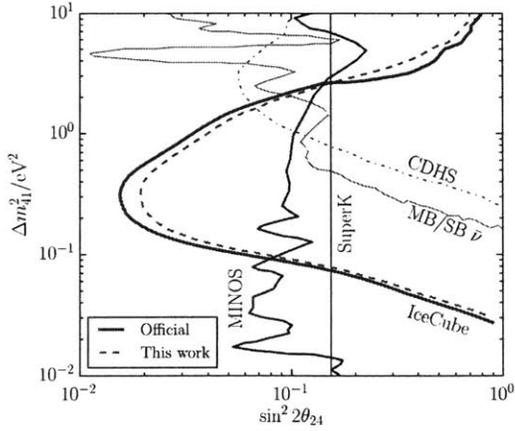
The right hand side of figure 3.4 shows the value of Δm_{51}^2 as a function of the CP -violating phase. This shows that the CP -violating phase can shift over a wide range to accommodate many solutions in Δm_{41}^2 and Δm_{51}^2 . In addition, from table 3.2, the difference in $\Delta\chi_{null-min}^2$ for 3 + 1 versus 3 + 2 models is about four, while four degrees of freedom were added. Thus, introducing the CP -parameter does not improve the fit beyond what is to be expected from the additional degrees of freedom.

3.5 Incorporating the IceCube sterile neutrino search

The IceCube experiment is very different from short base-line detectors. The gigaton scale detector is discussed in chapter 4, and particle physics analyses on this experiment typically use the atmospheric neutrinos as discussed in chapter 5. Despite the scales being dramatically larger in all regard – measuring events with energies of 400 GeV to 20 TeV, with atmospheric base-lines that span the entire diameter of the Earth – the oscillation scale of L/E is approximately 1 km GeV^{-1} — similar to short base-lines.


Figure 3.4:

Confidence regions for a 3 + 2 model using global data. **Left:** The parameter space projected into the plane of the two squared mass splittings. **Right:** Δm_{51}^2 vs. the CP -violating phase, Φ . Red indicates 90% CL and blue indicates 99% CL.


Figure 3.5:

Comparison of 90% CL limits for muon flavor disappearance of IceCube 2016, MINOS 2016, CDHS and MiniBooNE-SciBooNE. The reconstruction of the IceCube result used in this analysis is indicated by the dashed line.

A simple neutrino oscillation analysis is not practical on IceCube, in part due to the energy resolution and poorly constrained atmospheric flux normalisation. As it happens, an MSW-like matter effect causes a resonant depletion of the number of up-going atmospheric muon anti-neutrinos. This only happens in the presence of a sterile neutrino with $\Delta m^2 \sim 1 \text{ eV}^2$, and the resonance is located at an energy of

$$E_{\text{res}} = \frac{\Delta m^2 \cos 2\theta}{\sqrt{2}G_F N}, \quad (3.24)$$

where θ is an effective two flavour active-to-sterile neutrino mixing angle and N is the electron number density. In most neutrino experiments, this effect is negligible, but over the large length scales involved in IceCube it can become apparent.

This matter effect is the result of differing cross sections for coherent forward scattering between the three active flavour states. While

all three flavours can coherently forward scatter in neutral current interactions, only the electron neutrino can coherently scatter with charged current interactions, as the Earth is made of a large number of electrons. In the case of a $3 + 1$ neutrino mass ordering, as opposed to a $1 + 3$ ordering, the matter-induced resonance will appear in the anti-neutrino events rather than the neutrino events. For these reasons, a sterile neutrino search in IceCube is both interesting and can potentially give a powerful result. Figure 3.5 shows the result of one such analysis [2].

From equation 3.24 it follows that detectable effects will lie in the range $0.01 \leq \Delta m^2 \leq 10 \text{ eV}^2$ — the region of interest for short base-line anomalies. These matter effects also break degeneracies in global fits, allowing the constraint of $U_{\tau 4}$. This, in addition to the constraints on $U_{e 4}$ and $U_{\mu 4}$ given by the data sets discussed above, the constrains of the 3×3 U_{PMNS} sub-matrix from νSM three neutrino oscillation, leaves only the $U_{s 1}$ to $U_{s 4}$ elements. While these cannot be measured directly, the assumption that U_{3+1} is unitary provides four more constraints that remove these final degrees of freedom. Thus, with the addition of the IceCube sterile neutrino analysis, the $3 + 1$ mixing matrix can be completely constrained for the first time.

3.5.1 Predictions

The calculation of the predicted flux at the IceCube detector cannot be written in terms of analytic formulae as in section 3.2.1. This analysis, as well as Aartsen and others [2], both use the nuSQuIDS quantum differential equation solver to propagate the fluxes through the Earth [30, 31]. The short base-line approximation is incompatible with the matter effect being considered; thus, this analysis uses the νSM values of [118]:

$$\Delta m_{\text{sol}}^2 = 7.5 \times 10^{-5} \text{ eV}^2, \quad (3.25)$$

$$\Delta m_{\text{atm}}^2 = 2.3 \times 10^{-3} \text{ eV}^2. \quad (3.26)$$

which further result in two additional CP -violating phases. The Honda-Gaissner model is used for the initial atmospheric neutrino flux [144], while systematic and statistical variations of the atmospheric fluxes are provided by the results of chapter 5, published in Collin [77]. For the Earth electron density, nuSQuIDS makes use of the Preliminary Reference Earth Model (PREM) [89] parameterisation. The neutrino propagation implementation solves the neutrino evolution master equation that accounts for absorption, regeneration and oscillations as detailed in equations 29-30 in Gonzalez-Garcia, Halzen, and Maltoni [116]. The neutrino nucleon cross sections

use the perturbative QCD calculation from Arguelles et al. [32] and Cooper-Sarkar, Mertsch, and Sarkar [80]. The propagated flux hypothesis is used to re-weight the IceCube simulated events available in Collaboration [54].

As U_{3+1} is unitary, it can be decomposed into a series of rotations, which is required by the nuSQuIDS implementation:

$$U_{3+1} = R_{34}R_{24}R_{14}R_{23}R_{13}R_{12}, \quad (3.27)$$

where R_{ij} is a rotation by an angle of θ_{ij} about the plane defined by the i, j unit vectors. From this, the fourth column of U_{3+1} is

$$\begin{bmatrix} U_{e4} \\ U_{\mu4} \\ U_{\tau4} \\ U_{s4} \end{bmatrix} = \begin{bmatrix} \sin \theta_{14} \\ \cos \theta_{14} \sin \theta_{24} \\ \cos \theta_{14} \cos \theta_{24} \sin \theta_{34} \\ \cos \theta_{14} \cos \theta_{24} \cos \theta_{34} \end{bmatrix}, \quad (3.28)$$

which defines a coordinate transformation that relates the mixing angles used in section 3.4 to these new mixing angles:

$$\sin^2 2\theta_{ee} = \sin^2 2\theta_{14}, \quad (3.29)$$

$$\sin^2 2\theta_{\mu e} = \sin^2 2\theta_{14} \sin^2 \theta_{24}, \quad (3.30)$$

$$\sin^2 2\theta_{\mu\mu} = 4 \left(1 - \cos^2(\theta_{14}) \sin^2(\theta_{24})\right) \cos^2(\theta_{14}) \sin^2(\theta_{24}). \quad (3.31)$$

From here, it is easy to see that the ability to constrain $U_{\tau4}$ comes from the ability to constrain θ_{34} . This, in turn, comes from a critical dependence between θ_{34} and the presence of a matter effect induced resonance [168], as the effect vanishes when θ_{34} is maximal. However, maximal θ_{34} is disfavoured by induced distortions in the atmospheric $\nu_{\mu} \rightarrow \nu_{\tau}$ neutrino oscillation [92]. The CP -violating phases have been set to zero, as they have only a small effect in comparison to θ_{34} effect [92].

To validate the reproduction of the IceCube sterile neutrino analysis, the muon neutrino disappearance constraint has been reproduced and plotted along-side the official collaboration result in figure 3.5.

3.5.2 Inference framework

The high computational cost of propagating neutrino fluxes through the Earth with nuSQuIDS prevents the analysis from being directly included into the Markov Chain Monte-Carlo likelihood. Instead, the $3 + 1$ global fit was used to sample the most probable regions of parameter space. The 60,000 *parameter-set points* with the lowest χ^2 were used, and from these every 40th point was selected. This gave a fine sampling of the global fit near the minima. Of the remaining

$\sim 140,000$ parameter-set points that are far from the minima, every 60th was selected. Combining the fine sample and the coarse sample yields 4,000 points.

These 4,000 selected parameter-set points correspond to points in θ_{14} , θ_{24} , and Δm_{41}^2 only. In order to incorporate the effects of θ_{34} , ten values of the θ_{34} angle were chosen for each parameter point, resulting in a total of 40,000 parameter-set points. These points were fed into the IceCube analysis likelihood and the resulting χ^2 value was combined with the respective global fit χ^2 .

This assumes that the effect of IceCube on the global fit is a small perturbation, which is reasonable given that the IceCube-only $\Delta\chi^2$ shown in table 3.4 is small compared to the SBL only global fit $\Delta\chi^2$.

3.5.3 Likelihood

The IceCube likelihood is given by [2, 29, 153]:

$$\ln \mathcal{L}(\vec{x}|\vec{\theta}) = \min_{\eta} \left(\sum_{i=1}^N [x_i \ln \mu_i(\vec{\theta}; \eta) - \mu_i(\vec{\theta}; \eta) - \ln \Gamma(x_i + 1)] + \frac{1}{2} \sum_{\eta} \frac{(\eta - \bar{\eta})^2}{\sigma_{\eta}^2} \right), \quad (3.32)$$

where x_i are the number of events in the i^{th} bin, μ_i is the simulated expectation in the same bin, and $\vec{\theta}$ are the oscillation parameters.

The nuisance parameters η are specified in [2, 29, 54, 153], and the likelihood is maximised as a function of flux variants at each parameter point.

To use the inference scheme outlined in the previous section, the IceCube likelihood must be converted to a χ^2 . For this, the saturated Poisson defined in equation 3.14 was used.

3.5.4 Results

3+1	Δm_{41}^2	$ U_{e4} $	$ U_{\mu 4} $	$ U_{\tau 4} $	N_{bins}	χ_{min}^2	χ_{null}^2	$\Delta\chi^2$ (dof)
SBL	1.75	0.163	0.117	-	315	306.81	359.15	52.34 (3)
SBL+IC	1.75	0.164	0.119	0.00	524	518.23	568.84	50.61 (4)
IC	5.62	-	0.314	-	209	207.11	209.69	2.58 (2)

Table 3.4:

The oscillation parameter best-fit points for 3 + 1 for: the combined short base-line and IceCube data sets, IceCube-only, and short base-line alone. The units of Δm^2 are eV^2 .

The combined IceCube and short base-line global fit confidence regions are shown in figure 3.6, as a function of Δm_{41}^2 and $\sin^2 2\theta_{\mu e}$. The IceCube data excludes the solution at $\sim 1 \text{ eV}^2$ at 90% CL; although, the solution persists at 99% CL. This has important implications for future sterile neutrino searches designed to address the 1 eV^2 allowed region. The results are presented as a function of the mixing matrix elements in figure 3.7. The

$\Delta m^2/\text{eV}^2$	$ U_{e4} $	$ U_{\mu 4} $	$ U_{\tau 4} $	θ_{34}
6	[0.17,0.21]	[0.10,0.13]	[0.00,0.05]	$< 6^\circ$
2	[0.13,0.20]	[0.09,0.15]	[0.00,0.70]	$< 80^\circ$

$|U_{\tau 4}|$ result is presented on a linear scale because one sample – the preferred solution – has $|U_{\tau 4}| = 0$.

The constraint on $|U_{\tau 4}|$ produced by these results is shown in addition to the other mixing matrix elements in table 3.5. At $\Delta m^2 \sim 6 \text{ eV}^2$, this limit improves the bound of $\theta_{34} < 25^\circ$ at 90% C.L. from the MINOS experiment [12] by a factor of four. This new result on $|U_{\tau 4}|$ allows us to have the first complete picture of the extended lepton mixing matrix:

$$|U| = \begin{bmatrix} 0.79 \rightarrow 0.83 & 0.53 \rightarrow 0.57 & 0.14 \rightarrow 0.15 & 0.13 (0.17) \rightarrow 0.20 (0.21) \\ 0.25 \rightarrow 0.50 & 0.46 \rightarrow 0.66 & 0.64 \rightarrow 0.77 & 0.09 (0.10) \rightarrow 0.15 (0.13) \\ 0.26 \rightarrow 0.54 & 0.48 \rightarrow 0.69 & 0.56 \rightarrow 0.75 & 0.0 (0.0) \rightarrow 0.7 (0.05) \\ \dots & \dots & \dots & \dots \end{bmatrix}. \quad (3.33)$$

In the above, “...” represents parameters constrained by the unitarity of the 4×4 matrix. The ranges in the matrix correspond to the 90% confidence intervals. The entries in the last column correspond to the results of this analysis and are given for $\Delta m^2 \sim 2 \text{ eV}^2$ ($\Delta m^2 \sim 6 \text{ eV}^2$). The intervals shown in each entry for the standard 3×3 sub-matrix were obtained from Parke and Ross-Lonergan [196], and are independent of this fit.

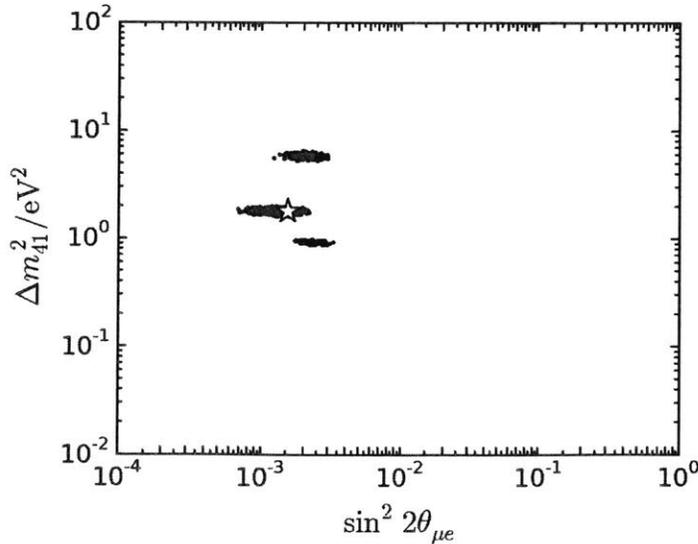


Table 3.5:

The 90% CL intervals for matrix elements and the upper limit on θ_{34} for the two allowed regions in Δm^2 . For $\Delta m^2 = 1 \text{ eV}^2$ there are no allowed regions at 90% CL.

Figure 3.6:

3 + 1 global fit confidence regions for the short base-line experiments with Ice-Cube data. Red indicates 90% CL, and blue indicates 99% CL.

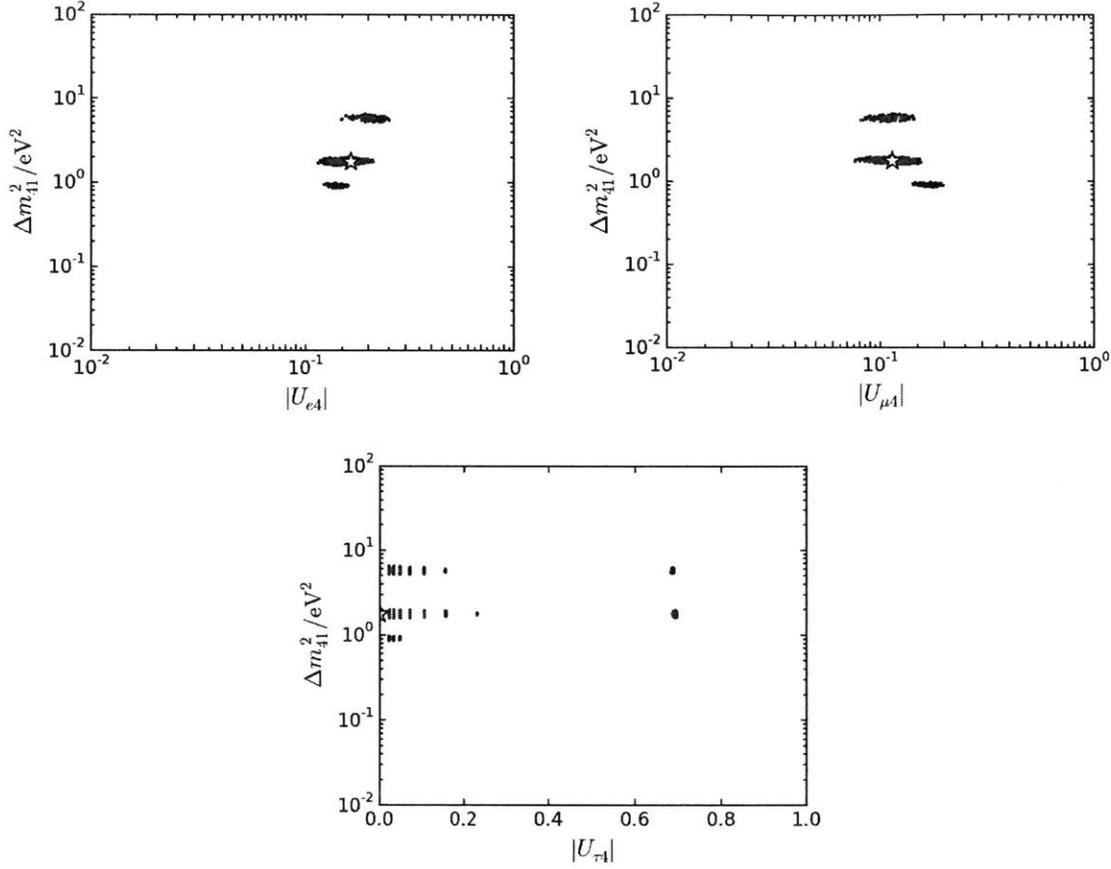


Figure 3.7:

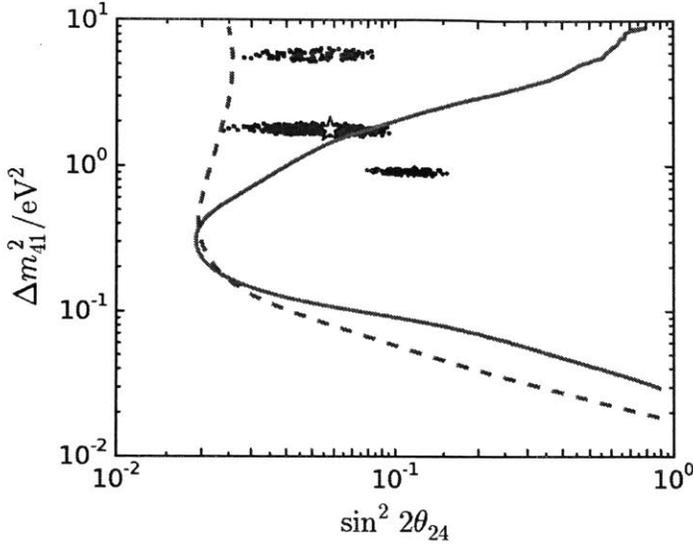
3 + 1 global fit confidence regions for the short base-line experiments with IceCube data, shown as a function of matrix element: **Left:** $|U_{e4}|$ **Right:** $|U_{\mu 4}|$ **Bottom:** $|U_{\tau 4}|$. Red indicates 90% CL, and blue indicates 99% CL.

As a check of consistency, our values in the fourth column can be compared with the upper bounds from the 3×3 non-unitarity analysis in Parke and Ross-Lonergan [196], which gave limits of $|U_{e4}| < 0.27$, $|U_{\mu 4}| < 0.73$, and $|U_{\tau 4}| < 0.623$ at 90% CL. The results in equation 3.33 are fully compatible with these upper limits, which are based on standard 3-neutrino oscillation measurements exclusive of any sterile neutrino search data.

The effect of θ_{34} on the IceCube 90% CL limit is shown in figure 3.8. This limit is modified at high Δm^2 if θ_{34} is nonzero, as illustrated by assuming $\theta_{34} = 15^\circ$, shown by the dashed line.

3.6 Summary

In a global analysis of short base-line data, this analysis finds a 3 + 1 model best fit of $\Delta m_{41}^2 = 1.75 \text{ eV}^2$ with a $\Delta\chi^2$ of 52.34 (for 3 degree of freedom) with respect to the null hypothesis. This solu-

**Figure 3.8:**

The solid line represents the 90% C.L. IceCube limit when calculated with $\theta_{34} = 0^\circ$. The dashed line is for $\theta_{34} = 15^\circ$. The result of the short baseline and IceCube global fit is overlaid. Red indicates 90% CL, and blue indicates 99% CL.

tion has interesting implications for the immediate future of sterile neutrino studies. MicroBooNE [26] – which is now taking data – is located on the Booster neutrino beamline with a peak ν_μ energy of 700 MeV. The 170 ton detector is located at a distance of 470 meters from the beam-line target. MicroBooNE is directly upstream of the 800 ton MiniBooNE experiment, which is located at 540 meters from the same target. If the presented 2 eV^2 solution of a $3 + 1$ model is correct, then MicroBooNE sits closer to oscillation maximum than MiniBooNE, thus predicting a higher signal in MicroBooNE than simple scaling for solid angle and tonnage assumes.

The combined fit of short base-line and IceCube data results in a best fit of $\Delta m_{41}^2 = 1.75 \text{ eV}^2$ with $\Delta\chi_{null-min}^2$ of 50.61 (for 4 degrees of freedom). Thus, the IceCube data further lessens the likelihood of the $\sim 1 \text{ eV}^2$ allowed region by a substantial amount. The combined global fit is sensitive to $|U_{\tau 4}|$, providing improved constraint on θ_{34} of $< 80^\circ (< 6^\circ)$ at 90% C.L. for $\Delta m_{41}^2 \approx 2(6) \text{ eV}^2$. This result, combined with constraints from fits to atmospheric and solar data sets, has been used to fill all components of the $3 + 1$ mixing matrix for the first time.

4

Detecting neutrinos with IceCube

Although the cross section of neutrinos grows roughly linearly with energy until the Glashow resonance [106], the prospects of detecting high energy neutrinos does not. Natural sources are required to reach beyond a few TeV in energy, and they follow inverse power law spectra $E^{-\gamma}$ where $\gamma \gtrsim 2$. Thus, observing the highest energy neutrinos – those that come from outside the solar system – requires the largest particle detector ever built⁴⁵.

Although the need for large detectors was known as far back as 1960 [181], the sheer scale required was not appreciated until the late 1970s when the Deep Underwater Muon and Neutrino Detector (DUMAND) was proposed [221]. With a design specification of a cubic kilometer of bulk material to act as a target for the neutrinos, no completely man-made construction was practical. Instead, the detector was proposed to make use of a natural source of such material: the ocean water off the coast of Hawaii. At a depth of 5 kilometers – sitting on the ocean floor – the detector would be isolated from potential background sources.

Unfortunately, the design was too ambitious for the then-nascent field of neutrino astronomy, and the final design – called “DUMAND-II” – was approved with a volume of $\sim 0.002 \text{ km}^3$. While construction moved at an impressive rate by the standards of oceanographers, the project was cancelled in 1995 as it was judged as lacking progress by the criteria of particle accelerators [221].

Instead, attention was already focusing on the Antarctic ice sheet. Antarctic ice absorbs less light than sea water, making energy reconstruction easier, but has greater scattering due to dust trapped in the

⁴⁵ In terms of instrumented volume. Collider experiment detectors are larger by measure of active element mass.

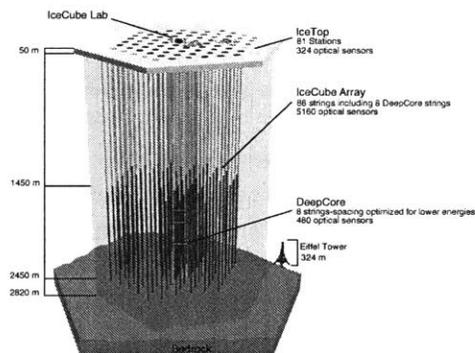
ice, leading to worse angular resolution for inferring the direction of the detected neutrinos. Nonetheless, the Antarctic ice was more practical: the ice sheet provides a solid surface on which installation equipment can be seated on and the U.S. south pole base provides an excellent place to stage the installation. While the weather can be poor at the south pole, it does not effect the subsurface of the glacier, making a detector stable once installed [221].

The Antarctic Muon and Neutrino Detector Array (AMANDA) started construction in late 1993, located just a few hundred meters from the Amundsen-Scott south pole station. The ice in the top-most layers of the glacier are formed by the accumulation of snow, and thus has a large amount of trapped air, which scatters light to a high degree. It was thought that below 800 meters, this air would form an ultra-clear clathrate due to the immense pressure of the ice above it. However, AMANDA discovered that this transition actually occurs between 1500 meters and 2000 meters below the surface.

4.1 The IceCube detector

Figure 4.1:

The IceCube detector, located ~ 2 kilometers under the south pole glacier. Figure courtesy of the IceCube collaboration.



AMANDA operated until 2009, when it was switched off. It collected ~ 7000 neutrinos, but was unable to positively identify any point sources of neutrinos, or measure a diffuse flux of astrophysical neutrinos; although, it placed the strongest limits of its time. To find astrophysical neutrinos, and perform neutrino astronomy, a larger detector was required.

Thus, a one cubic kilometer detector made from the glacial ice at the south pole – called IceCube – was proposed. IceCube is arranged into 86 “strings” of Digital Optical Modules (DOMs), which are the light collection devices that detect the Cherenkov radiation emitted

from neutrino interactions. Each string has 60 DOMs spread over 1 kilometer of cable, with additional cable length allowing the lowering of the entire assembly into holes drilled in the glacier. The cable carries both the power supply and communication connection to the surface. The holes are drilled to a depth of 2450 kilometers using a pressurised hot water drill [64].

The primary array consists of 78 strings arranged on a triangular grid with an average inter-string distance of 125 meters. This spacing was chosen to optimise the sensitivity of IceCube to neutrinos with an energy of 1 to 1000 TeV [64]. The remaining 8 strings are placed in a tighter arrangement at the center of the detector, with an average spacing of 75 meters — to allow detection of lower energy events. Called “DeepCore”, these strings have a non-standard inter-DOM spacing when compared to the strings of the primary array. The deepest 50 DOMs have a vertical spacing of 7 meters, forming the main detection bulk for DeepCore, with the other 10 DOMs above having a spacing of 10 meters to form a veto layer for background rejection [64].

4.1.1 IceTop

Situated above IceCube is a surface array called “IceTop”. This array of 162 tanks filled with clear ice was designed to extend IceCube’s capabilities to cosmic ray physics by measuring showers that develop in the atmosphere overhead. The tanks are arranged in pairs, occupying a common triangular grid to the IceCube detector.

The array also serves a secondary purpose as a surface veto for IceCube analyses. If a neutrino event seen in IceCube was initiated by a neutrino produced in a cosmic ray shower above the detector, this shower would also create a coincident event in IceTop. By utilising IceTop, IceCube is able to increase its acceptance by 10% [58].

4.1.2 Detector signatures

Two basic building blocks form events in IceCube: tracks and cascades.

Tracks

Most muons leave long single lines of light – called “tracks” – inside the detector, an illustration of which is shown on the top left in figure 4.3. As muons have relatively low energy losses, they are highly penetrating particles that can travel kilometers in rock or ice.

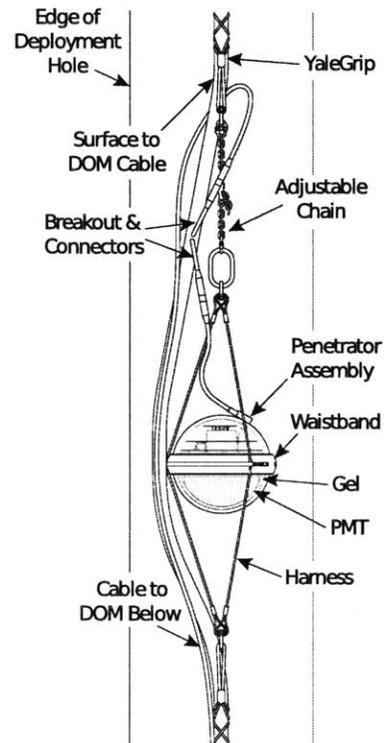


Figure 4.2: A Digital Optical Module (DOM) attached to its parent string. Figure included from Collaboration et al. [64].

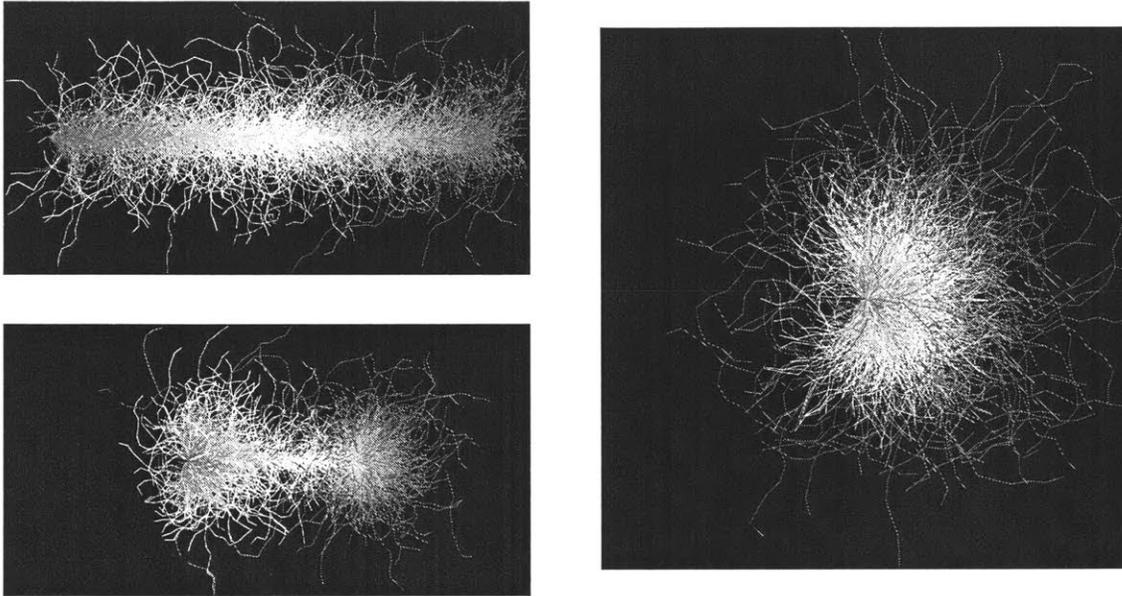


Figure 4.3:

An illustration of various interaction signatures in IceCube. Time is colour coded from early (red) to late (blue). Clockwise from top left: **A:** A track. **B:** A cascade. **C:** A double bang. Figures included from Coenders [52].

Thus, IceCube sees many muons that were created by muon neutrinos interacting in the material around, above and under the detector, increasing the sensitivity of the detector dramatically compared to the other neutrino flavours. Unfortunately, the same properties apply to the muons generated by cosmic ray showers in the atmosphere above the detector, and these muons often penetrate the glacial overburden to leave tracks in the detector that form a background for most analyses.

The elongated nature of the track makes it ideal for estimating the direction of the underlying muon. Although, it can be difficult to estimate the energy of the underlying muon when the start and end of the track lie outside the detector, as only an unknown fraction of the total deposited energy can be measured. Due to these reasons, tracks are often employed in analyses where angular resolution is more important than energy information, such as astrophysical point source searches. The increased sensitivity to tracks also makes them good choices for any analysis where statistics are a serious issue — such as the sterile neutrino search described in chapter 3.

Cascades

Cascades are the product of both electromagnetic and hadronic showers in IceCube, for which an illustration is shown on the right in figure 4.3. Unlike tracks, they are compact and often nearly

spherically symmetric, as the length scales of the shower are short compared to the optical scattering length of the ice. They are most often found at the interaction vertex of a neutrino, where they mark the hadronic cascade produced from deep inelastic scattering.

With the small footprint of the cascade being entirely contained within the detector, it is comparatively easy to reconstruct the energy of the cascade by totalling the deposited energy. However, the near spherical symmetry impairs directional reconstruction. For these reasons, cascades are most often used when energy resolution is of utmost importance, such as in diffuse astrophysical flux measurements.

Topologies

Interactions in IceCube are then further classified into various event *topologies* based on the combinations of tracks and cascades observed.

Most muon neutrinos leave a single track inside the detector. If the neutrino, and thus generated muon, are very high energy, stochastic radiative losses can take over, leaving distinct and separated cascade where the muon has induced a shower. If the muon neutrino interacts inside the detector, both a track and cascade will be left in the detector, both with a common origin at interaction vertex.

Electron neutrinos are always marked by a single cascade, as the resulting electron immediately induces an electromagnetic shower. Tau neutrinos, will – like electron neutrinos – often create single cascades, as the tau decays immediately, creating a hadronic shower that overlaps with the original interaction vertex shower. However, with $\sim 17\%$ probability the tau will decay into a muon, which will leave a track. Along with the original shower, this mimics the signature for a muon neutrino. The muon decay mode of the tau – much like the muon neutrino – also increases the reach of the detector, and thus IceCube's sensitivity to tau neutrinos.

Extremely high energy taus will be highly time dilated, delaying the lepton's decay, and separating the second cascade from the first. This leaves two distinct depositions in the detector, as can be seen in the illustration on the bottom left of figure 4.3.

4.2 Digital Optical Modules

The DOM is housed in a 13 inch diameter spherical glass vessel, with a small penetrator where a tap from the main cable enters the housing. The majority of the volume inside the DOM is occupied by a 10 inch Photo-Multiplier Tube (PMT), which is optically coupled by gel to the lower surface of the housing.

4.2.1 Photo-multiplier tube

The Hamamatsu R7081-0 PMT was selected for its time and charge resolution for the detection of single photons [71]. When such a photon strikes the optically thin conducting *photo-cathode* on the main collecting area of the tube, the photo-electric effect causes an electron to be ejected. This *photo-electron* is attracted to the first *dynode* in the dynode chain, which is an electrode held at a positive voltage compared to the photo-cathode. The acceleration caused by this electric field causes the electron to strike the dynode with enough energy to liberate more electrons, which are in turn attracted to the next dynode in the chain, repeating the process. The IceCube PMTs are operated at a voltage sufficient to produce a gain factor of 10^7 from this cascade of electrons [64]. The collection of the electrons on the last stage of the chain – called the *anode* – creates a pulse in voltage large enough for data collection electronics to measure.

As the photo-cathode is optically thin, photons can pass through the cathode to strike the first dynode directly. This can cause the ejection of an electron that initiates a similar pulse from the output of the PMT. If the photon is part of a larger burst of light, it will reach the dynode before the photo-electrons created by other photons interacting with the photo-cathode. The time difference of this *pre-pulse* compared to the primary photo-cathode induced pulse is characteristic of the PMT geometry and driving voltage. For the IceCube PMTs, this is ~ 31.8 nanoseconds, and the amplitude of the pre-pulse is typically $\sim 6.6\%$ of the primary pulse [232].

Electrons striking the first dynode may elastically backscatter, inducing *late-pulses* after they return to the dynode. The kinetics of the scattering process also give these late-pulses a characteristic time difference of around two photo-electron transit lengths [232]. Inelastic scattering is also possible, but results in a broader timing distribution. If a remnant ion inside the PMT is ionised by a photon, it too can induce a highly delayed *after-pulse* in the PMT [232].

4.2.2 Digitising electronics

The PMT power supply and main-board sit above the PMT. The main-board contains a custom digitiser called the “Analog Transient Waveform Digitizer” (ATWD). Before entering the ATWD, the PMT pulses are routed through three amplifiers of gain 16, 2, and 0.25 to capture the full dynamic range of interactions in IceCube – from single photons to high energy cascades [69].

The ATWD has three channels to handle these gain settings. As the pulses enter the ATWD, they are sampled at 300MHz – corresponding to ~ 3.33 nanosecond sampling period – and stored in 128

capacitor analogue storage [159]. This storage is digitised into waveforms over 29 microseconds, so tandem ATWDs exist on the main-board such that one can capture pulses while the other is digitising its storage [69].

The waveform from a channel is only sent to the FPGA if the next highest gain channel is saturated. As the limited analogue storage of the ATWD allows for only a 427 nanosecond recording window, a separate lower speed digitiser is also present. This digitiser samples continuously, and is programmed to save data in a 6.4 microsecond window after the main ATWD digitisation [64].

4.3 The ice

Accurate simulation and reconstruction of neutrino events in IceCube demands a thorough understanding of the bulk ice in which the detector is embedded. The two fundamental quantities of interest to IceCube are the scattering and absorption lengths of the ice.

4.3.1 Light model

Absorption leads to an attenuation in the intensity of the light I , governed by Beer's law:

$$I(\tau) = I_0 e^{-\tau_a}, \quad (4.1)$$

where I_0 is the initial light intensity, and τ is the optical absorption depth. The amount of absorption is quantified by the absorption length $a(\vec{x})$, which can be defined as an intrinsic point-wise property of the ice. The optical depth is the integrated absorption length over a path \mathcal{P} :

$$\tau_a = \int_{\mathcal{P}} a(\vec{x}(s)) ds. \quad (4.2)$$

The geometric scattering length defines the probability of a photon scattering at a position s along a path \mathcal{P} :

$$p(\vec{x}(s)) = b(\vec{x}(s)) e^{-\tau_b(s)}, \quad (4.3)$$

where τ_b is the geometric optical scattering depth along the path

$$\tau_b(s) = \int_{\mathcal{P}} b(\vec{x}(s)) ds. \quad (4.4)$$

When the photon scatters, it changes direction based on an *angular scattering distribution*, also often called the *phase function*.

The scattering length is also wavelength dependent [59]:

$$b(\lambda) = b_0 \left(\frac{\lambda}{\lambda_0} \right)^{-\alpha}. \quad (4.5)$$

The absorption length shows a similar power law scaling in wavelength, with the addition of a temperature dependence:

$$a(\lambda, T) = a_0 \left(\frac{\lambda}{\lambda_0} \right)^{-\kappa} + a' \left(\frac{T}{T'} - 1 \right) e^{-\lambda'/\lambda}. \quad (4.6)$$

The angular scattering distribution is best modelled by Mie scattering, as relatively small⁴⁶ dust grains are the dominant cause of scattering [11]. To avoid the complex calculations required for the application of Mie scattering [11], the scattering distribution is approximated by a mixture [59],

$$p(\cos \theta) = f_{\text{SL}} p_{\text{SL}}(\cos \theta) + (1 - f_{\text{SL}}) p_{\text{HG}}(\cos \theta), \quad (4.7)$$

of a simplified Liu [171] and Henyey-Greenstein [137] distribution:

$$p_{\text{SL}}(\cos \theta) = \frac{1}{2} \frac{1+g}{1-g} \left[\frac{1+\cos \theta}{2} \right]^{\frac{2g}{1-g}}, \quad (4.8)$$

$$p_{\text{HG}}(\cos \theta) = \frac{1}{2} \frac{1-g^2}{(1+g^2-2g \cos \theta)^{3/2}}. \quad (4.9)$$

where $g = \langle \cos \theta \rangle$ is the mean cosine scattering angle, which parameterises both distributions. These distributions are desirable as they have analytic inverse cumulative distribution functions.

$$\cos \theta = 2P_{\text{SL}}^{\frac{1-g}{1+g}} - 1, \quad (4.10)$$

for the simplified Liu model, where P_{SL} is a uniform random number; and similarly for the Henyey-Greenstein model,

$$\cos \theta = \frac{1}{2g} \left[1 + g^2 - \left(\frac{1-g^2}{1+g(2P_{\text{HG}}-1)} \right)^2 \right]. \quad (4.11)$$

4.3.2 Bulk ice model

The absorption, scattering and other properties of the ice as a function of position inside the detector are incorporated into an *ice model*. Many such models have been developed for use on IceCube over the years. One of the most preferred—due to its simplicity—is called SPICE Mie.

⁴⁶ With sizes roughly comparable to the wavelength of light.

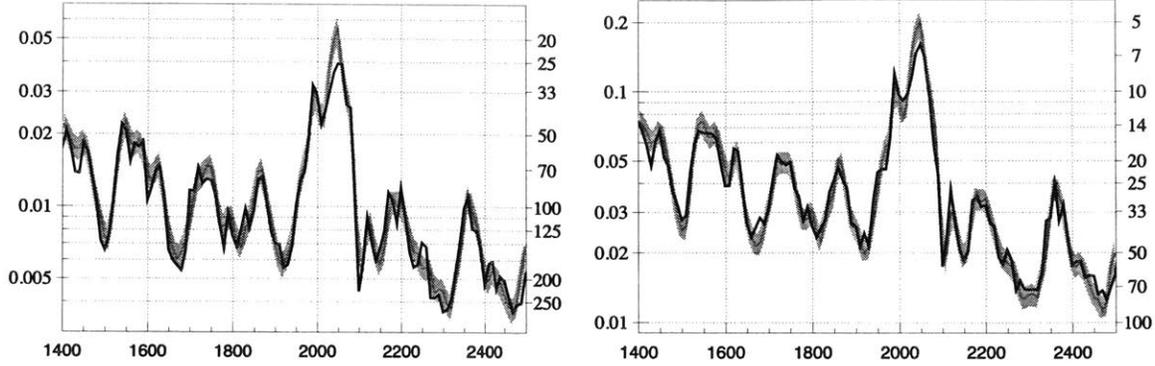


Figure 4.4:

Comparison of SPICE Mie (purple) to SPICE Lea (black). **Left:** The absorption parameter a [m^{-1}] as a function of depth in meters. **Right:** The effective scattering parameter $b_e = b(1 - \langle \cos \theta \rangle)$ [m^{-1}] as a function of depth in meters. Figure included from Halzen [124].

This model breaks the ice depth-wise into 10 meter thick layers. Each layer has an associated value of absorption and scattering parameters a_0 and b_0 , while the remaining parameters of α , κ , a' , T' , λ_0 , f_{SL} and g are considered to be constant across the entire detector [59].

The temperature of the ice is further parameterised in depth as a quadratic [59]:

$$T(z) = T_0 + T_1 z + T_2 z^2. \quad (4.12)$$

While the phase and group refractive indices are parameterised by a quartic and a relative cubic in wavelength, respectively [59]:

$$n_p(z) = \sum_{k=0}^4 n_{p,k} z^k, \quad n_g(z) = n_p(z) \sum_{k=0}^3 n_{g,k} z^k. \quad (4.13)$$

These parameters are estimated by comparison to calibration data using the “flasher“ LED system installed on each DOM. This is a series of LEDs mounted to the circuit-boards inside the DOM, and can be activated to inject known quantities of light into the detector at a wavelength of 399 ± 14 nanometers [64]. By flashing these LEDs with precise timing, the absorption and scattering parameters can be disentangled through the scattering induced time delay of light traveling between or up and down strings.

The more advanced ice model known as SPICE Lea builds upon SPICE Mie by adding directional dependence into the scattering distribution — motivated by the measurement of directional anisotropies in the transmitted light, possibly due to dust grains preferentially aligning in the direction of the ice flow [65]. This re-parameterises the scattering distribution in terms of the incoming and outgoing directions \vec{n}_{in} and \vec{n}_{out} :

$$p(\cos \theta) \rightarrow p(\hat{k}_{\text{in}} \cdot \hat{k}_{\text{out}}), \quad \vec{k} = \mathbf{A}\vec{n}, \quad (4.14)$$

for a diagonal anisotropy matrix A .

The most advanced model, SPICE³, introduces per-DOM parameters to describe LED emission and PMT efficiency properties.

4.3.3 Hole ice

In the process of drilling a hole for a detector string, hot water is transported from the surface to the drill head deep in the ice. This water has dissolved gasses present, and during re-freezing of the ice, these gasses form bubbles in the ice — creating a column of very high scattering. The static pressure will eventually force this gas back into a clathrate form, but not within the time-scales of the experiment. This column is estimated to be ~ 15 centimeters in diameter, and will not generally be aligned with the axes of the DOMs. Studies have suggested that the effect of this *hole ice* can be taken into account by a modified angular efficiency distribution [64].

4.4 Simulation and reconstruction

Simulation is the basis on which the detector is understood, and reconstruction is the method by which the neutrino properties are inferred from the measured light inside the detector. All estimates of the efficacy and accuracy of the analyses performed on IceCube require simulated data, where the “true” properties of the interacting neutrino are known. This section provides a brief summary of these complex topics, and the interested reader is referred to Coenders [52], Jones [154], Weaver [228], and Xu [232] for more detailed discussion.

4.4.1 Event generation

At the high energies available in IceCube, the Earth starts to become opaque to neutrinos. To account for this, simulation begins with a neutrino initialised at a random location over the entire surface of the Earth, pointing toward the IceCube detector. As they travel through Earth, these neutrinos may scatter on the mantle or core via charged current interactions. The resulting electron or muon will quickly come to a stop, removing the neutrino from consideration, while tau neutrinos decay almost instantly, preserving much of the energy of the tau neutrino in its *regenerated* tau neutrino product. Throwing away neutrinos which interact with the Earth would be an inefficient waste of the computational effort invested in them. Instead, these events carry a weight, which – as discussed in chapter 2 – is modified by any stochastic losses they undergo.

This weighting scheme is most crucial for the distribution of neutrino energies. IceCube analysis projects are most often interested in the highest energy events, so simulating events that follow the measured energy spectrum of $\propto E^{-2}$ or steeper results in a large number of low energy events that are subsequently removed in background-rejection stages of the analysis. The neutrinos are generated according to an E^{-1} spectrum, to increase the proportion of the interesting high energy events while still maintaining a power law energy distribution. However, unlike in the general formula for re-weighting, the weights are only updated with the simulated probability distribution $Q(E)$, and not with the true distribution $P(E)$. This biased weight – known as a `OneWeight` to the collaboration – for an event undergoing a series of physical processes \mathcal{P} is thus:

$$w_o = \frac{1}{Q(E)} \prod_{k \in \mathcal{P}} \frac{P(k)}{Q(k)}, \quad (4.15)$$

which allows the end user of the simulation to recover the unbiased event weight for any energy distribution $P(E)$ that they desire:

$$w = w_o P(E). \quad (4.16)$$

After propagation through the Earth, the neutrino is forced to interact with the detector – or foreground mass in the case of muon and tau neutrinos – with an appropriate re-weighting by the cross section. The products of the interaction are sampled from a pre-calculated table of final states, and passed to the particle propagator.

4.4.2 Detector simulation

As the interaction products travel through the detector or surrounding material, they lose energy through ionisation and radiative processes. The *particle propagator* simulates this energy loss, as well as the generation and simulation of any secondary particles such as tau decay products or pair production. If the particles are near enough to the detector, the Cherenkov radiation generated by their passage is saved and passed to the photon propagator.

The photon propagator handles the transport of light from charged particles to the DOMs. This is currently performed using either path tracing⁴⁷ with the light and ice physics described in section 4.3.2, or semi-analytic approximation models that can be appropriate for some cascade simulations. The details of path tracing based light propagation is further discussed in chapter 10.

Once simulated photons strike the DOM, the simulation of the PMT response is handed over to phenomenological models. These

⁴⁷ Also known colloquially as ray tracing.

add appropriate amounts of noise, and account for pre-pulsing, late-pulsing and after-pulsing. Further simulation steps handle the electronics and digitiser responses, including saturation effects and triggering.

4.4.3 Track reconstruction

Track reconstruction usually proceeds in three steps of increasing computational complexity. An initial guess for the location and direction of the track is made by a simple linear least squares regression. This minimises the time difference between the predicted and observed timing of the first photon detected by each DOM. The predicted timing is based on a simple direct time of flight for the photon between the track and DOM. Although scattering is substantial in the ice, the angular scattering distribution is highly forward peaked, so the first observed photon has likely taken a near direct route.

The next step takes the scattering into account through use of an analytic approximation to the time distribution of the scattered light, called a “Pandel function” [68]. Taking the result from the first step of the seed, this step provides a more accurate estimate of the muon properties, although it is still limited to the first observed photon. The final step takes this result as an input seed, and repeats the fit using a Panel function which is adjusted to account for the statistical bias of only selecting the first observed photon.

Due to the high background rate of atmospheric muons, there is small but significant probability of more than one muon present in the event under reconstruction. One method of estimating the presence of multiple tracks is to test if observed photon timings have the light-like separation that would be characteristic of a single muon traveling at near light-speed. Other methods divide the data sample into two pieces, and fit them separately to gauge if this division improves the quality of the fit.

Energy reconstruction is based on estimating the total light produced by the muon based on the observed light. This must take into effect the scattering and absorption, as well as the general geometric factor from diverging light rays.

4.4.4 Cascade reconstruction

Cascade reconstruction can be performed using the same Pandel based analytic time distribution as in the track reconstruction, but with a point source instead of a line source. More advanced methods tabulate, interpolate and match pre-simulated templates for light deposition in the detector.

Reconstruction usually focuses on continuous neutrino properties such as direction and energy, but simple discrete classification of neutrino type is also useful for some analyses. Chapter 9 presents a study in using deep convolutional neural networks to classify tau neutrinos from other cascades. Other studies are presently on-going in IceCube to evaluate how these machine learning based methods can improve direction and energy reconstruction.

5

Atmospheric neutrino flux

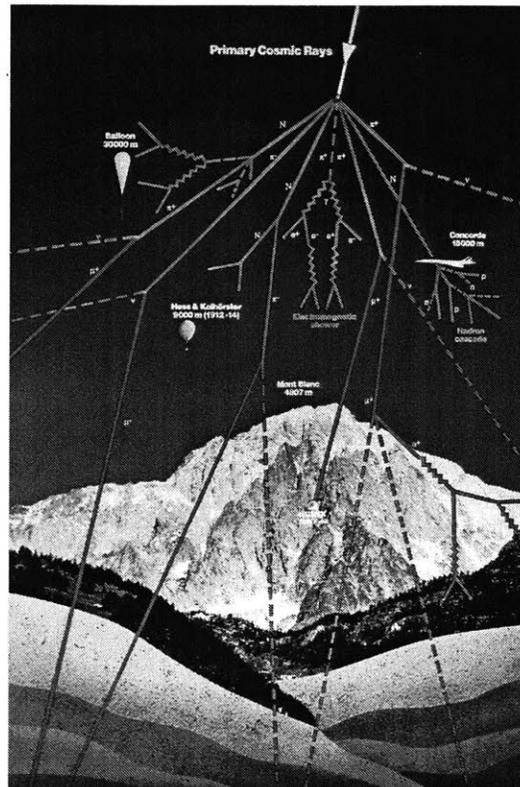
Atmospheric neutrinos are generated when cosmic rays strike the atmosphere. The interaction evolves much like a very high energy particle beam striking a target. Deep inelastic scattering is the dominant mode of interaction for these cosmic rays. Much like in neutrino deep inelastic scattering, this process creates a hadronic shower — illustrated in figure 5.1.

The calculation of the neutrino flux from these atmospheric showers has a long history. Originally motivated by the pre-cursors to IceCube described in chapter 4, these calculations used analytic approximations [234] — limited to energies much greater than 1 GeV, and unable to account for complex particle interaction or atmospheric models. The need for accurate calculations at low energies, where detectors like Super-Kamiokande are most sensitive, drove the adoption of numerical calculation techniques [109]. Full three dimensional Monte-Carlo simulations were employed to account for the effects of the Earth's magnetic field [143], and the CORSIKA [135] software package was developed specifically for this purpose.

A determination of the neutrino flux was necessary to confirm that the observed asymmetry in the upward and downward going neutrino flux measured by Super-Kamiokande was, in fact, due to neutrino oscillation. Today, the IceCube experiment requires knowledge of the neutrino flux at high energy in order to perform searches for additional beyond-the-standard-model effects.

Figure 5.1:

When a cosmic ray strikes the atmosphere, it creates a multitude of secondary products, shown in green. Many of these are unstable, or further interact in the atmosphere to produce muons and muon neutrinos, shown in red. Figure courtesy of CERN.

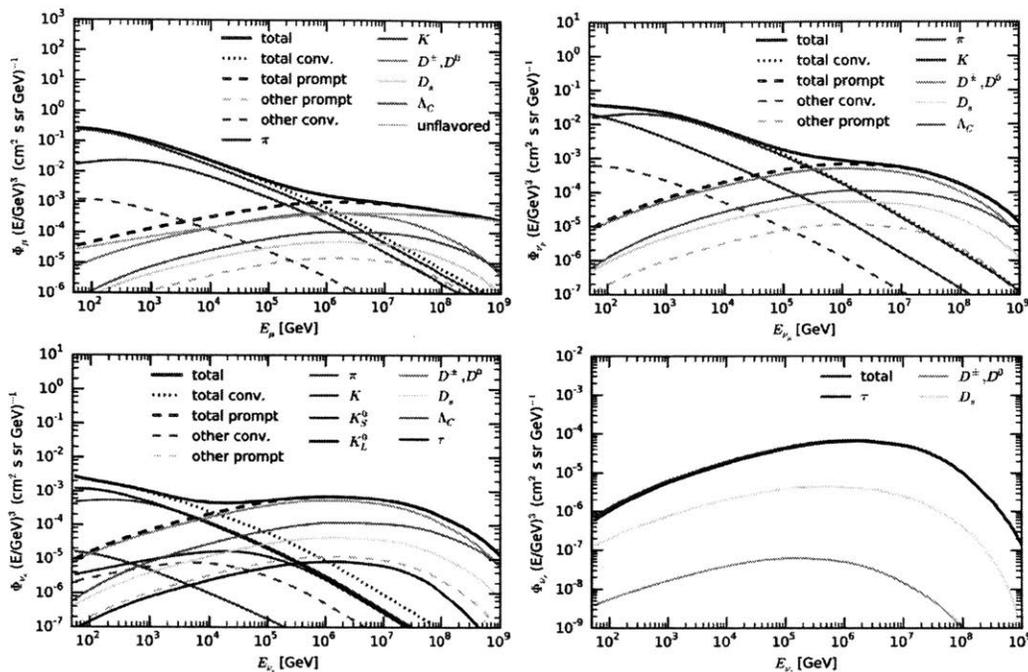


5.1 Shower development

The neutrinos created in a cosmic-ray induced, atmospheric hadronic shower are grouped into two broad categories: *conventional* and *prompt*. This classification is based on the type of particle that produced the neutrino. The familiar charged pions and kaons are considered to be sources of conventional neutrinos, and dominate the atmospheric neutrino flux at low energies, as seen in figure 5.2 More exotic particles, such as charmed mesons, are sources of prompt neutrinos. These are named for the short lifetimes of these mesons, which plays an important role in the development of the shower.

The atmosphere is unusual as far as particle beam targets go. Viewing the atmosphere from above, it has an exponentially increasing density with an extremely sharp discontinuity at the surface of the planet. As a meson or muon travels through the atmosphere, it loses energy. The longer the particle stays in the atmosphere, the less energy it will have to pass on to the neutrino it will decay into.

Very high energy pions and kaons tend to penetrate deep into the atmosphere, and muons will even make it as far as the surface of the Earth. Inside the high density of the earth, they quickly lose energy



and decay into neutrinos that are too low in energy for IceCube to detect. Prompt mesons, on the other hand, decay on a much faster time scale, avoiding the Earth and passing on most of their energy to their daughter products. The distance that these mesons will travel before they decay is governed by their relativistic boost factor — the amount of time dilation which multiplies the decay lifetime. For this reason, neutrinos from these exotic particles are the dominant component of the atmospheric flux at high energies, as pions, kaons and muons will be sufficiently boosted to reach the surface of the Earth.

The atmosphere is also a continuously changing environment. It goes through daily heating and cooling cycles, as well as long term trends in seasonal dependence, and even greater scales of climate change. As the temperature of the atmosphere is directly related to its density and pressure, these variations can translate into changes in the atmospheric neutrino flux. The energy deposition rate of a particle depends on the density of the material it is moving through, thus lower densities allow particles to retain their energy before decaying. Higher pressures also cause the atmosphere to expand, so that the incident cosmic ray interacts at a greater height, and has longer to travel before striking the surface of the Earth. This potential effects are particularly important to understand for particle physics analyses on IceCube, which use the atmospheric flux as a signal. This chapter

Figure 5.2:

Calculation of atmospheric neutrino fluxes broken down by parent particle. Figure included from Fedynitch et al. [97]. “conv.” refers to conventional neutrinos.

details a calculation of the atmospheric flux primarily designed to fit the needs of the IceCube sterile neutrino search [2].

5.2 Matrix cascade approach

Atmospheric neutrinos can be generated anywhere on the surface of Earth, and so a calculation of the flux for one or even a handful of atmospheric density profiles would fail to capture the variance in the data. Three dimensional Monte-Carlo methods – such as CORSIKA – propagate individual particles through the atmosphere, simulating their kinematics and interactions. This level of detail is typically only required when a low energy⁴⁸ atmospheric neutrino flux is desired, as geo-magnetic effects start to come into play in this regime. For this study, the use of such a computationally expensive method would severely limit the number of variations that could be computed.

The MCEq software package was developed to solve the flux calculation as a system of coupled differential equations [96, 97]. The cosmic ray shower is reduced to a 1D problem, which is a good approximation in the high energy regime⁴⁹. Let Φ_E^h be the flux of particles with energy E and type h at any point along this shower. Particles will be lost to inelastic scattering on the atmosphere and decay at a rate proportional to the interaction length $\lambda_{\text{int}, E}^h$ and decay length $\lambda_{\text{dec}, E}^h$, respectively,

$$\frac{\Phi_E^h}{\lambda_{\text{int}, E}^h}, \quad \frac{\Phi_E^h}{\lambda_{\text{dec}, E}^h}. \quad (5.1)$$

New particles of type l and energy E' will be produced by the processes, according to the same interaction and decay lengths. In addition, the new particle creation rates will be proportional to the interaction and decay coefficients $c_{l, E' \rightarrow h, E}$ and $d_{l, E' \rightarrow h, E}$, respectively,

$$c_{h, E \rightarrow l, E'} \frac{\Phi_E^h}{\lambda_{\text{int}, E}^h}, \quad d_{h, E \rightarrow l, E'} \frac{\Phi_E^h}{\lambda_{\text{dec}, E}^h}. \quad (5.2)$$

The interaction length will depend on the density of the atmosphere, while the decay length depends on the boost factor for E . Instead of defining the differential equation over the distance along the cascade z , the differential distance element is scaled by the local atmospheric density ρ :

$$dX = \rho(z) dz \quad \Rightarrow \quad X(z) = \int_0^z \rho(z') dz', \quad (5.3)$$

⁴⁸ Below 1 TeV.

⁴⁹ Interaction products are highly focused in the forward direction, so the high energy part of the shower forms a highly elongated cone.

where X is called the *slant depth*. In this coordinate, the interaction length only varies slowly with energy according to the inelastic cross-section, while the decay length is now a function of the energy and slant depth [97].

The differential matrix cascade equation governing the change in flux is then [97]

$$\frac{d\Phi_{E_i}^h}{dX} = -\frac{\Phi_{E_i}^h}{\lambda_{\text{int}, E_i}^h} - \frac{\Phi_{E_i}^h}{\lambda_{\text{dec}, E_i}^h(X)} + \sum_{E_j \geq E_i} \sum_l \left(c_{l, E_j \rightarrow h, E_i} \frac{\Phi_{E_j}^l}{\lambda_{\text{int}, E_j}^l} + d_{l, E_j \rightarrow h, E_i} \frac{\Phi_{E_j}^l}{\lambda_{\text{dec}, E_j}^l} \right) \quad (5.4)$$

The software package has a wide variety of pre-programmed inputs. The cosmic ray flux model can be one of: Hillas Gaisser [110], poly-gonato [142], Zatsepin-Sokolskaya/PAMELA [14, 235], Gaisser-Honda [109] and a combined Hillas-Gaisser/Gaisser-Honda model [96]. The hadronic model, which controls the coefficients c , can be one of: QGSJET-II-04 [194], SIBYLL 2.1 [22], SIBYLL 2.3 RC1 [91], or DPMJET-2.55 [206]. The decay coefficients d were estimated by sampling the Pythia 8 event generator [218, 219].

The flux is efficiently calculated by casting the coupled differential equations in a matrix format. Solving for one spectrum requires a few minutes of computation on a modern processor. Repeating this for each zenith angle seen by the detector results in a total computation time of a few hours. As a result, this offers a much faster method to explore the systematic effects from atmospheric models than through a Monte Carlo system.

5.3 Atmospheric Density Model

The centrifugal force exerted by the rotation of the Earth pulls the atmosphere to higher altitudes at the equator. Meanwhile, the lower intensity of sunlight at the poles causes the atmosphere to cool, making it denser. Small scale weather and large scale climate effects also come into play, including a factor due to the topography of the earth. This makes for a complicated picture that is difficult to model from first principles. Instead, an empirical model is more practical.

5.3.1 Existing models

There is a long history of such models, as a good understanding of the atmosphere is critical to aerospace engineering. The first widely accepted series of models were called the U.S. standard atmospheres [225]. The last update to this model was in 1976, although it is still used today in some applications. An important limitation to this

model is that it has no latitude or longitude dependence – it is defined for only a single location in North America.

Mass Spectrometer and Incoherent Scatter Radar models

Today, the most widely employed model is the U.S. Naval Research Laboratory *Mass Spectrometer and Incoherent Scatter* (MSIS) Radar based model from ground to *Exosphere*, published in 2000 (NRLMSISE-00) [201]. This is the last update to a series of MSIS based models.

It is a sophisticated model, that takes the following parameters as inputs:

- Date and time of day
- Latitude/Longitude
- Altitude
- Solar flux (as a function of 10.7 centimeter radiation)
- Geomagnetic index.

And gives

- Density of composite species ($N_2/N/O_2/O/Ar/He/H$)
- Total mass density
- Temperature

The model is a hybrid of physical laws and fitted climate data. For example, the altitude dependence of the temperature of the thermosphere is modelled by an approximation called the Bates temperature profile. While in the lower atmosphere, the temperature is modelled by splines over height and Legendre polynomials over latitude and longitude. The density is then computed from the temperature by approximating to a hydrostatic equilibrium [136].

Despite this sophistication, the model has some important limitations. It does not provide errors on the calculated quantities, or provide variances on the underlying fitted model parameters. There exists overall standard deviations for the residuals of the fit, but this is not sufficient for a comprehensive study of systematics.

Additionally, the model is now quite out of date. The warming trend of climate change is not expected to have much difference, however the knock-on effects of shifting weather systems could have an impact on the latitude and longitude dependence of the result. This effect would be difficult to quantify.

Earth-GRAM 2010

Instead, a more modern model is sought. The Earth Global Reference Atmosphere Model (Earth-GRAM) 2010 is the latest in a series of models published by NASA. This model was developed for modelling ballistic trajectories and orbits, and so it has significant defense applications. For this reason it is export controlled, and access is only granted to U.S. persons with a valid proposed application (but is otherwise open source and free of charge). Unfortunately, access could not be negotiated for this analysis, and so this model could not be used.

5.3.2 Atmospheric data

The most up to date data on the atmosphere can be obtained directly from measurements. Additionally, they are the most accurate source of information on the atmosphere, and they can capture the complete spectrum of variability in the atmosphere. There are two major types of data collection: balloon based, and satellite based.

Balloon data

Balloon borne measurements provide the most accurate and detailed source of data on the atmosphere. Barometers and thermometers can measure the actual pressure and temperature of the atmosphere continuously as function of height. However, balloons have very limited geographical and temporal coverage. Balloon data was not considered for this analysis as these variables were deemed important.

Satellite data

Satellites have the advantage over balloons in that they are long term instruments whose coverage spans the entire Earth. However, as they are located in space, they must take indirect measurements of atmospheric temperature. The AQUA satellite is an example of one such satellite. Its primary role is to provide real time data for weather forecasting, but NASA also provides a high level data archive for use in climate and other science. The Atmospheric Infrared Sounder (AIRS) instrument [224], shown in figure 5.3, collects the bulk of the data. This device uses the pressure and temperature dependence of the spectral absorption of CO_2 to infer the temperature profile of the atmosphere as a function of height [157]. A major complication is the presence of cloud cover, which blocks infrared emission. To combat this effect, data from the Advanced Microwave Sounding

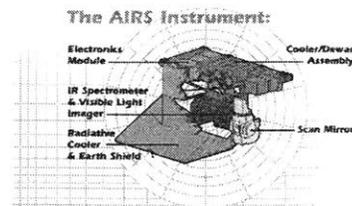


Figure 5.3:
The Atmospheric Infrared Sounder (AIRS) instrument.

Unit (AMSU) on the same satellite is used to augment the temperature reconstruction [37].

Temperature profiles are publicly released by NASA, and contain data at various processing stages [193]. The level 3 data set contains either daily, 8-day, or monthly averages over a grid of latitude and longitude. Of most importance is that this data set includes both statistical errors on all measurements, and systematic errors on the temperature measurements. This allows the statistical and systematic uncertainties of the atmosphere to be quantified.

5.3.3 Identifying the Height Range of Interest

The AIRS data is only provided to a geo-potential height of ≈ 45 kilometers. In order to determine if this coverage is sufficient, MCEq was run using the SIBYLL 2.3 RC1 interaction model over a few atmospheric models at the north pole. The results are shown in figure 5.4, over various species of particles where solid lines represent a vertically incident⁵⁰ cosmic ray – producing the shortest possible path through the atmosphere – and dashed lines represent a horizontal incidence — producing the longest possible path. This shows that the major interactions are confined to a region below 50 kilometers over a range of seasons and track angles. Thus, the AIRS data is sufficient for predicting the neutrino flux.

⁵⁰ As measured by the zenith angle of the resultant cascade makes as observed from the ground.

Use of the AIRS satellite data

⁵¹ This orbit is arranged such that it precesses through one complete revolution around the Earth each year. By doing so, the orbit is always locked into the same position relative to the sun, so that it passes over any given location on the Earth at the same local solar time each day.

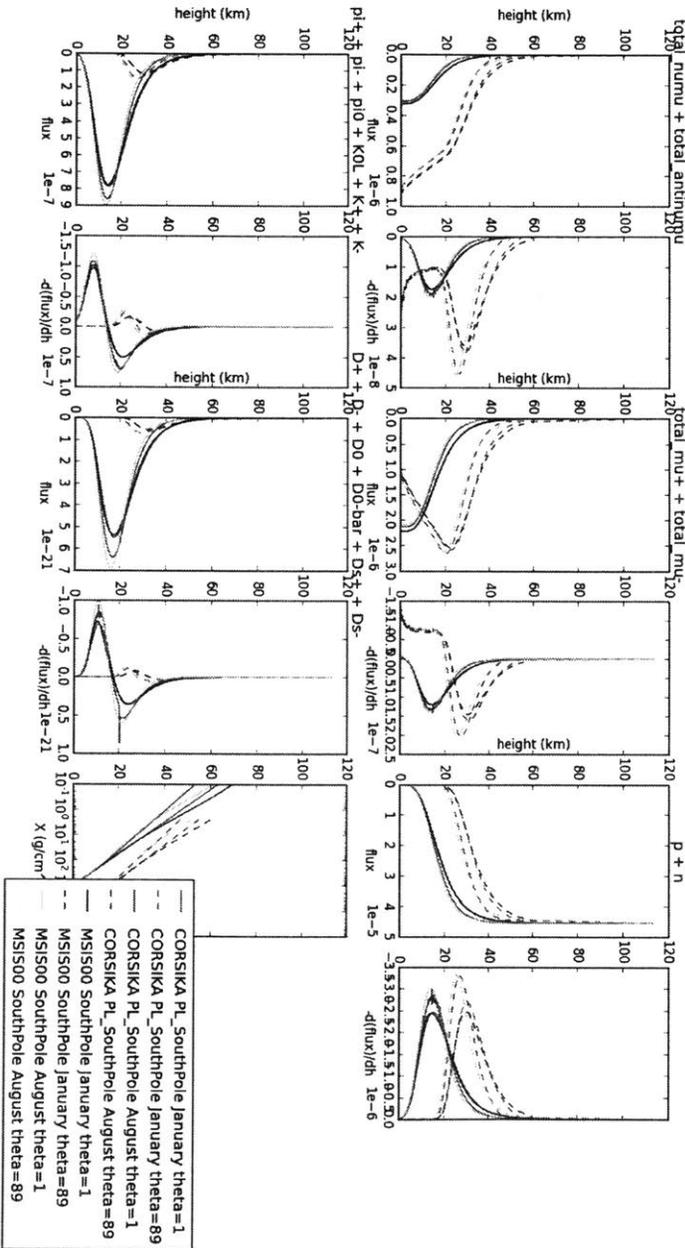
The AQUA satellite is in a helio-synchronous orbit⁵¹, which is divided into an ascending segment – facing the sun – and descending segment — in the sun shadow. The satellite crosses the equator once in each segment, and does so at the same solar time: 1:30 PM during the ascending segment, and 1:30 AM when descending. The result is that the local solar time under the satellite is only a function of the latitude of the satellite's current location. The monthly average data set was chosen for this analysis, which reflects the seasonal variations while averaging daily meteorological changes. Use of the daily data set would be complicated by missing data near the equator, from locations outside the field-of-view on that day. The 8-day dataset does not contain gaps, but increases the number of variations beyond the computational limits of this analysis.

The data provided by the satellite is arranged on a 180×360 grid, with each element representing a $1^\circ \times 1^\circ$ area on the surface of the Earth. Each of these elements contain 24 values, which correspond to measurements at 24 fixed pressure levels in the atmosphere. There are also grids of the surface skin temperature⁵² and forecasted surface pressure. These are divided into ascending and descending sets,

⁵² Air temperature at the surface.

Figure 5.4:

Production as a function of height. Flux and differential flux ($d\Phi/dh$) are plotted in pairs, side by side. These pairs correspond to 5 different groups of particles. From top left to right: 1) $\pi^+ + \pi^- + K_L^0 + K^+ + K^-$, 2) $\nu_\mu + \bar{\nu}_\mu$, 3) $D^+ + D^- + D^0 + \bar{D}^0 + D_s^+ + D_s^-$, 4) $\mu^+ + \mu^-$, 5) slant depth as a function of height, and 6) $p + n$.



in order to preserve the diurnal signal. Using a map of topography provided by the public data repository, the pressure levels below the surface of the Earth were masked out. The density of the air is calculated from the temperature using the ideal gas law at each remaining pressure level. This gives a set points of density and height $\{(\rho, h)\}$ which are interpolated in logarithmic space using an upper boundary of 10^{-25} g/cm^2 at ≈ 100 kilometers.

5.4 Estimation of uncertainties

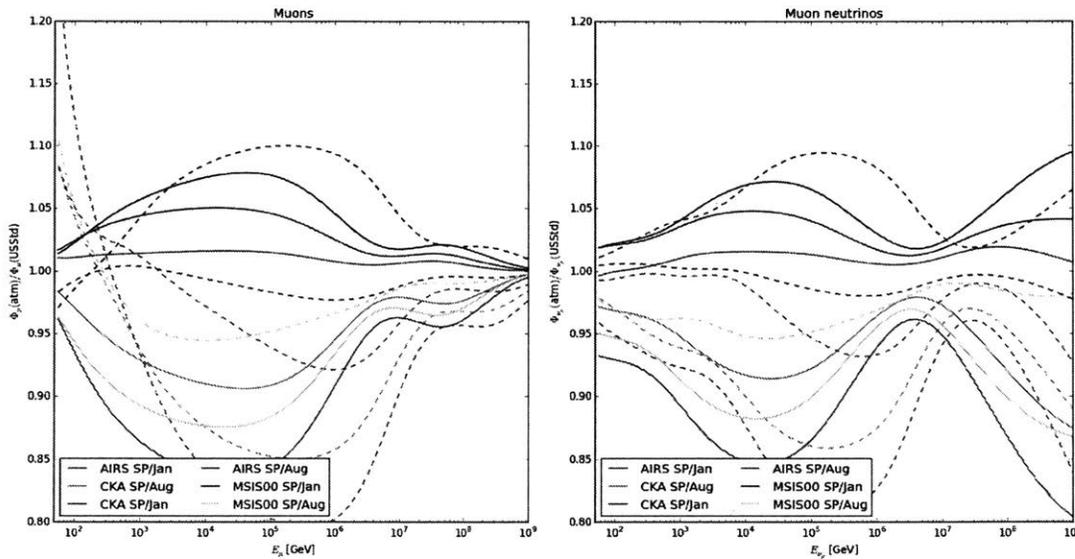


Figure 5.5: Muon and muon neutrino flux as a ratio to the US Standard atmosphere at the south pole for various atmospheric models and seasons listed in the legend. CKA is the CORSIKA south pole atmosphere parameterisation. The solid lines correspond to vertical neutrinos, and the dashed to horizontal neutrinos.

The statistical nuisance parameters are:

- Time of year (by month): The IceCube sterile neutrino search data set was recorded from 2011/05 to 2012/05. The live time of the experiment was not uniform over this period, and the time-of-year parameter was chosen using a probability distribution based on the recorded live time. The effect of seasonal variation can be seen in figure 5.5.
- Time of day (either day or night): The AIRS data has only either day or night values. Interpolating between these values would require a model of atmospheric cooling and heating. Instead, this parameter can only take the values of night or day, which should provide a conservative over-estimate of the daily variation.

- **Longitude:** The final flux spectrum for IceCube particle physics analyses are integrated over longitudes, and thus the longitudinal variation in the atmosphere becomes a source of uncertainty.
- **Statistical variations in the AIRS data:** The monthly average has a standard deviation, which is reported for each data point. Covariances are not provided by the public data.

The systematic nuisance parameters are:

- **Interaction model:** Reference [96] investigated a number of hadronic interaction and shower models. As that publication recommends, DPMJET is eliminated for this analysis, which has a poor fit to data from the LHC. This leaves the following models which we compare to determine the systematics on hadronic production and showering:
 - QGSJET-II-04,
 - SIBYLL 2.3 RC1,
 - SIBYLL 2.3 RC1 Point-like
- **Cosmic ray flux model.**
- **Systematic shifts in the AIRS data:** Each temperature data point includes a systematic error. A random z-score is chosen, and all data points are shifted by this amount according to their mean and reported systematic error.

Atmospheric neutrino fluxes were generated with all combinations of the above parameters, with the exception of the statistical and systematic AIRS parameters which were instead sampled randomly from the error estimates provided by the AIRS data.

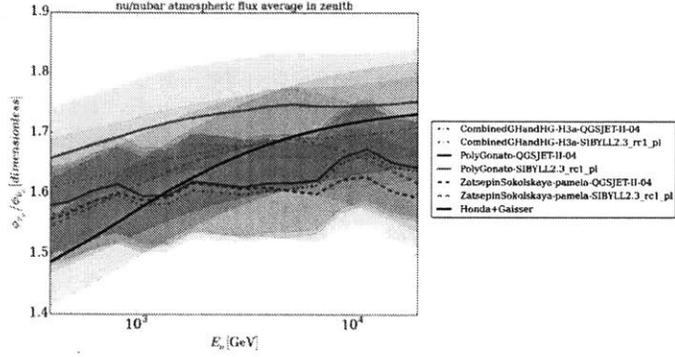
5.5 Summary

These atmospheric fluxes were of three main benefits to the IceCube sterile neutrino analysis. It was unknown at that time if seasonal variations in the atmosphere could create a localised depletion in the number of atmospheric neutrinos. If this were possible, it could mimic a sterile neutrino signal, potentially harming the analysis. These fluxes confirmed that, while significant depletions do happen over the polar regions during some seasons, the time averaged flux over a single year does not.

The sterile neutrino signal is only present in muon anti-neutrinos, as discussed in section 3.5, so the ratio of muon neutrino to anti-neutrinos, as shown in figure 5.6 is important to gauging the sensitivity of the analysis. The statistical and systematic errors in the satellite

Figure 5.6:

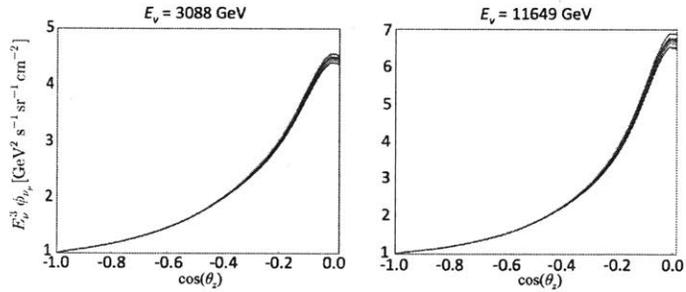
The ratio of muon neutrino flux to muon anti-neutrino flux for various choices of cosmic ray and hadronic models. An uncertainty of 5% was chosen for this ratio in the sterile neutrino analysis, and is shown as the bands around each flux. Figure included from Jones [153].



itself could be encapsulated as a continuous nuisance parameter by estimating the variability of the computed fluxes. This variability is shown in figure 5.7. Both the neutrino to anti-neutrino ratio and the atmospheric variability nuisance parameter were included in the sterile neutrino search likelihood description, providing a conservative treatment of the systematic error introduced by the neutrino fluxes. In addition, the likelihood minimised over discrete combinations of the interaction model and primary cosmic ray model calculated using this system.

Figure 5.7:

The effect of the AIRS systematic and statistical uncertainty at two example energies. The black lines show the variations caused by the satellite. The blue lines show the range of a continuous parameterisation used to capture the effect of these variations. Figure included from Jones [153].



After compilation into a database, the flux variations were published in reference [77] so that they were available outside of the IceCube collaboration. This allows external analyses of the publicly released data, of which the incorporation of the sterile neutrino search into the global fits described in chapter 3 is one example.

Astrophysical neutrinos



For over a millennium, astronomers have viewed the universe through visible light. This preoccupation with visible light is one born through practicality; the only measurement device available to the ancients was the human eye. Today, we understand that visible light is only one small component of a wider spectrum of electromagnetic radiation. The advent of electronics in the 20th century opened new views for astronomy: radio, infrared, x-ray and gamma rays; cosmic rays; neutrinos; and most recently gravitational waves. We now live in an era of *multi-messenger* astronomy, whose disparate production mechanisms each reveal unique information about their common source which would not be available through a single kind of messenger alone.

Cosmic rays are composed of the charged matter particles: electrons, positrons and atomic nuclei — of which protons are the primary constituent. These charged particles bend their trajectories when traversing the magnetic fields that span interstellar space, and scatter when crossing the boundaries between magnetic fields. Thus, the arrival direction of the cosmic rays carries almost no information about where the cosmic ray was created.

Cosmic rays encompass all the stable matter particles with the exception of the neutrino. It stands to reason that where cosmic rays are accelerated, high energy neutrinos might also be produced. The neutrino is uncharged, so its arrival direction aligns exactly with the astronomical process which created it, potentially pinpointing the objects that are generating cosmic rays. High energy neutrinos are the most tantalising of these astronomical messengers; despite the

discovery of neutrinos that come from outside our solar system, what creates them is still a mystery.

6.1 The origin of cosmic rays

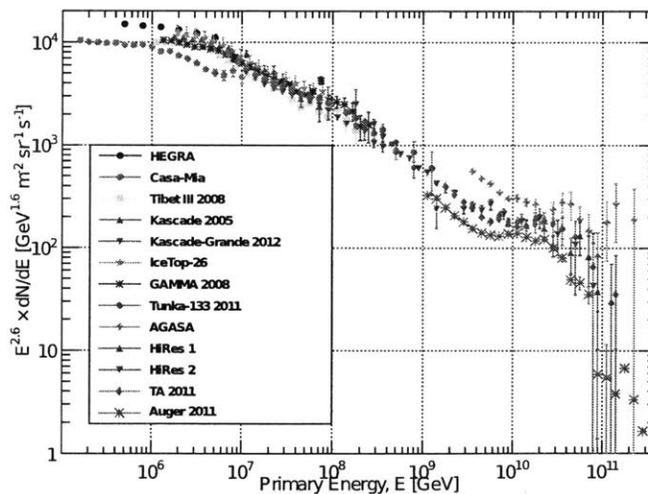
⁵³ The term has also been used to refer to particles emitted from the Sun, and accelerated in interplanetary space.

The term “cosmic ray” refers in this work to the relativistic charged particles that originate from sources outside the solar system⁵³. Victor Hess discovered these particles by measuring the production of ions inside a container during a high altitude balloon flight [138]. Later, in 1927, Jacob Clay found evidence of a dependence between cosmic ray intensity and latitude, suggesting the cosmic rays were charged particles being deflected by the Earth’s magnetic field [51].

In the 1930s, Geiger counters were used in coincidence to show that the angular distribution of cosmic rays had an azimuthal asymmetry that favoured rays arriving from the west. This showed that most cosmic rays were positively charged. While performing these experiments, Bruno Rossi observed that multiple pairs of coincidence counters would trigger at a rate greater than expected by accident [211]. He correctly inferred that that he was measuring the secondary products from a shower of particles, and Pierre Auger later added that these showers develop when a cosmic ray interacts with the atmosphere.

Figure 6.1:

Energy spectrum for cosmic rays, weighted by $E^{2.6}$. Figure included from Gaisser, Stanev, and Tilav [111].



Modern approaches use advanced balloon and space borne particle detectors to measure cosmic rays directly, and arrays of detectors on the Earth’s surface that measure the showers caused by cosmic rays. Approximately 80% of cosmic rays are protons, and 70% of

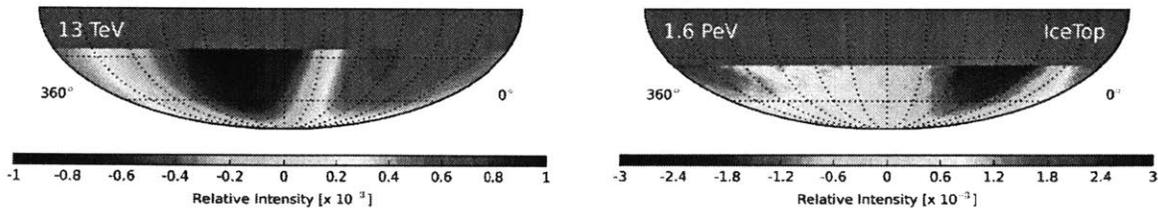
the remainder are Helium nuclei [198]. The rest are heavier elemental nuclei, with the exception of positrons and electrons that comprise $\mathcal{O}(300)$ ppm [1]. The kinetic energy spectrum is heavily peaked at around 1 GeV. Above this energy, the flux of cosmic rays falls very nearly like $E^{-2.7}$ [198]. Figure 6.1 shows that an inflection point called the “knee” is located at ~ 1 PeV where the spectrum steepens until a second inflection point call the “ankle” at 10^5 PeV where it flattens out again.

The paths through space of these cosmic rays are bent by the magnetic fields that permeate our galaxy. The gyro-radius of a cosmic ray between the knee and ankle is approximately

$$r \sim \frac{p}{|q|B} \sim 100 \text{ ly} \tag{6.1}$$

or near to the thickness of the disk of the Milky Way⁵⁴. Cosmic rays with less than this energy will be magnetically confined into helical paths inside the galaxy, while those of higher energy can escape into the intergalactic medium. This suggests that the ankle may mark a transition from a large trapped population to a population of cosmic rays that originate from outside the galaxy.

⁵⁴ Approximately 1000 light years.



At all but the highest energies, cosmic rays have lost the directional information of their origin due to the influence of the galactic magnetic fields. The relative anisotropies of the directional distribution of cosmic rays is of the order of 0.1%. In figure 6.2, the distribution shows an overall dipole anisotropy which undergoes a transition of $\sim 140^\circ$ in right ascension at ~ 100 TeV. This could indicate a change of the dominant populations of sources from the Orion arm at low energy, to the galactic center at higher energies; however, simulations predict that the anisotropy should be larger if this was the case [56]. At energies above 8,000 PeV, the anisotropy moves to point away from the galactic center, which may indicate the transition to extra-galactic cosmic rays [73].

The power law energy spectrum is not characteristic of a thermal process. Instead, a process of discrete accelerations naturally leads to a power law distribution. When a cosmic ray is deflected by a

Figure 6.2:

Maps of the relative intensity for cosmic rays in the southern sky in equatorial coordinates, as measured by IceCube. Left: At 13 TeV, a dipole anisotropy is observed. Right: At 1.6 PeV, this anisotropy rotates $\sim 140^\circ$ in right ascension. The right plot uses data from the IceTop array. Plots by IceCube collaboration *et al.* [56].

magnetic field, it gains a relative fraction of energy ζ , so that after n deflections, the energy of the cosmic ray is

$$E = (1 + \zeta)^n E_0 \quad (6.2)$$

where E_0 is the starting energy. If the cosmic ray has a probability P_{esc} of escaping the mechanism at each deflection, then the probability of a cosmic ray surviving until it has energy E is a power law:

$$P(E) \propto (1 - P_{\text{esc}})^{n(E)} = \left(\frac{E}{E_0}\right)^{-\gamma} \quad (6.3)$$

where $\gamma = -\ln(1 - P_{\text{esc}})/\ln(1 + \zeta)$. This mechanism is called ‘‘Fermi acceleration’’ [99] and can occur when a cosmic ray enters a shock front. As the ray magnetically diffuses it gains energy from the shock front before being ejected.

When a star with large mass undergoes a supernova explosion, it blows material out into the interstellar medium (ISM). This supernova remnant forms a shock front between the ejecta and the ISM, where particles entering from the ISM can be deflected. These particles can gain energy through repeated crossings, forming the Fermi acceleration mechanism described above [43].

The power required to sustain the observed flux of cosmic rays is approximately 3×10^{40} erg s⁻¹. Assuming a period of 50 years between supernovae, the time averaged power output of supernovae is approximately 10^{40} to 10^{41} erg s⁻¹ [113]. These numbers are only accurate to a few orders of magnitude, so it seems possible that supernovae could supply most of the observed cosmic ray flux, if only a few percent of the explosion-power output is converted to cosmic rays.

A pulsar can form at the center of a core collapse supernova. The interaction between the ejected material and this pulsar creates a pulsar wind nebula. These nebulae likely accelerate leptonic cosmic rays [45, 53, 192]. However, any ions present in the pulsar wind will also be accelerated to high energies [165], and an argument has been made that these ions may, in fact, carry the majority of the wind’s energy [195].

Extra-galactic neutrinos may be formed in ‘‘starburst galaxies’’. These galaxies have regions of intense star formation, and thus will also have large numbers of supernovae. The acceleration process – and maximum cosmic ray energy – would be identical to the supernovae in the Milky Way.

The highest energy cosmic rays might be accelerated in ‘‘active galactic nuclei.’’ A galactic nucleus is the region at the center of a galaxy that can harbour extremely massive black holes. The magnetic

field around one of these black holes is able to accelerate particles to energies of 10^{18} eV or more, potentially explaining the cosmic ray spectrum beyond the ankle.

6.2 Production of astrophysical neutrinos

The production of astrophysical neutrinos can be seen in analogy to neutrinos produced in man-made particle accelerators. As discussed in chapter 1, these devices create charged pions which decay to produce muon neutrinos and muons. The muons subsequently decay into muon-neutrinos and electron-neutrinos, through the following decay chain:

$$p + X \rightarrow Y + \pi^+ + \pi^- + \pi^0 \quad (6.4)$$

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \quad (6.5)$$

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \quad (6.6)$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (6.7)$$

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \quad (6.8)$$

In an example of an astrophysical process, high energy cosmic rays strike molecular clouds near the supernova where they were produced – creating a kind of “astrophysical beam dump”. This results in the production of charged pions, as above, and thus also neutrinos. The flavour ratio of these neutrinos is $(\nu_e : \nu_\mu : \nu_\tau) = (1 : 2 : 0)$; however, it should be noted that the long propagation distances involved cause the neutrino state to lose coherence, as discussed in chapter 1. Thus, the actual observed flavour ratio at Earth will be those of the individual mass states, weighted by the relative production of the mass states at the origin.

Neutral pions are also produced in these reactions. They decay electromagnetically into two photons:

$$\pi^0 \rightarrow \gamma + \gamma. \quad (6.9)$$

Evidence from the Fermi Gamma Ray Telescope points to gamma rays being produced by neutral pions around supernova remnants [237]. While this strongly supports the hypothesis that cosmic rays are produced by supernovae, it is not conclusive. Positive identification of neutrinos originating from remnants would be the smoking gun for hadronic interactions, and thus the production of cosmic rays.

Another method for neutrino production is by cosmic rays traversing the interstellar medium. These neutrinos would follow a diffuse distribution correlated with gamma ray emission from the galactic disk. Gamma rays are also observed from large regions called “Fermi bubbles” that lie below and above the galactic center [223].

While the mechanism of gamma ray production in the bubbles still unknown, a hadronic process would also create neutrinos [173].

Extra-galactic neutrinos can be produced by starburst galaxies in a similar fashion, while cosmic rays from active galactic nuclei can scatter off photons nearby to produce pions [180]:

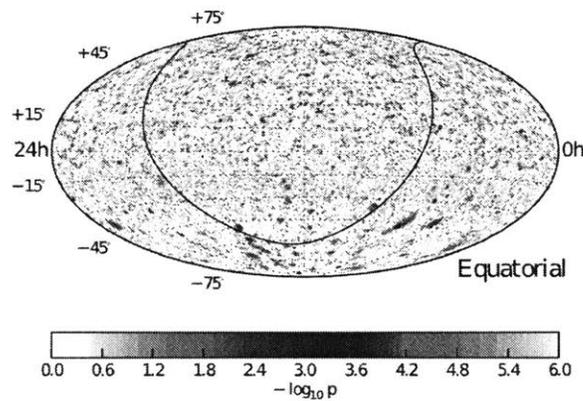


which decay to produce gamma rays and neutrinos, respectively.

6.3 Neutrino astronomy

Figure 6.3:

Unbinned likelihood test as a function of the tested source position in equatorial (J2000) coordinates. Shown is the negative logarithm of the pre-trial corrected p-values. The galactic plane is shown as a black line. Figure included from Collaboration et al. [55].



Neutrinos are neutral particles that do not participate in electromagnetic interactions. This makes them excellent candidates for astronomy, as the direction of the arriving neutrinos will point directly back to their site of production. If these sites are small in comparison to the angular resolution of the detecting equipment, such as with supernova remnants, then they are called “point sources”. The goal of neutrino astronomy is to locate and categorise these point sources. The resulting catalog can then be cross-compared with those developed from other areas of astronomy to learn more about our universe.

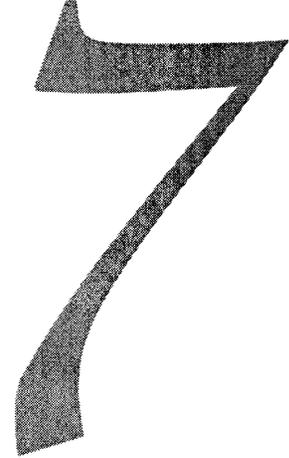
Measuring these neutrinos requires a large detector to compensate for the neutrino’s small cross section. A number of particle detectors have been constructed for this task, such as Baikal, ANTARES, AMANDA, and the IceCube experiment described in chapter 4. IceCube is by far the largest of these detectors by volume, and so has the best chance at collecting enough neutrinos to locate point sources.

The IceCube collaboration has performed a search for point sources using 7 years of data. This analysis uses an un-binned likelihood method, that tests the presence of a single point source

in a given location in the sky. This test is repeated in a grid across the sky, to search for a hotspot of neutrino emission that might indicate a point source. After correcting the test for trial factors⁵⁵, the analysis gives a p-value of 29% and 17% for the northern and southern sky respectively. These values are not significant enough to claim a discovery of a neutrino point source [55].

The flavour ratio of astrophysical neutrinos can also provide a test of the physics of long range neutrino propagation, which may be modified by beyond standard model effects [33]. The flavour ratio measured at Earth is expected to be between $(0.6 : 1.3 : 1.1)$ to $(1.6 : 0.6 : 0.8)$ for production ratios of $(0 : 1 : 0)$ to $(1 : 0 : 0)$ at the source, respectively, based on the current best global fit for three neutrino oscillation [57, 117]. Notably, the proportion of tau neutrinos does not change much between these scenarios, providing a prediction that is only weakly dependent on the nature of the production mechanism. The most recent measurement of the flavour ratio in IceCube is consistent with the $(1 : 1 : 1)$ prediction [57]; although, the ratio of electron to tau neutrinos is very weakly constrained due to the extreme difficulty – discussed in section 1.2.5 and chapter 9 – of distinguishing the two in IceCube. As of yet, no definite single identification of a tau neutrino event has been made in IceCube data [4].

⁵⁵ When many statistical trials are repeated, there is a good chance of a false positive due to a statistical fluctuation. Either the threshold for discovery must be raised, or the p-value corrected using a trial factor, to account for the number of trials performed.



Search for neutrino point sources

The basic premise behind Non-Poissonian Template Fitting (NPTF) is to differentiate point sources from diffuse emission based on the statistical distribution of counted neutrinos. The method has a history in radio and X-ray astronomy where photon counts are large enough to use Gaussian statistics. Malyshev and Hogg [179] were the first to apply this method to gamma ray astronomy, for which the number of collected photons is smaller.

Small-count statistics requires the use of Poisson distributions. Consider the number of neutrinos that might land in a single spatial bin, due to point sources. First, the number of point sources, N^{PS} , located in this bin is drawn from a Poisson distribution. Then, the number of neutrinos that land in the bin is also Poisson distributed, but with a mean that depends on N^{PS} . In this sense, the “non-Poissonian” distribution that underlies the method is more akin to a doubly Poissonian distribution.

A derivation of the distribution starts with the Poisson probability of finding m number of neutrinos in a bin:

$$p_m(S) = \frac{S^m}{m!} e^{-S}, \quad (7.1)$$

where S is the expected number of neutrinos.

The total number of sources, N^{PS} , can be written in terms of a *differential source count function*:

$$N^{\text{PS}} = \int_0^\infty \frac{dN}{dS} dS. \quad (7.2)$$

Then, the mean number of sources that produce m detected neutrinos in a bin is given by

$$x_m = \int_0^\infty p_m(S) \frac{dN}{dS} dS \tag{7.3}$$

$$= \int_0^\infty \frac{S^m}{m!} e^{-S} \frac{dN}{dS} dS. \tag{7.4}$$

The number of sources, N_m , that produce m detected neutrinos is a Poisson distributed random number with mean x_m . The total number of detected neutrinos, k , is then a weighted sum of these random numbers:

$$k = \sum_{m=1}^\infty m \cdot N_m. \tag{7.5}$$

The probability of k total detected neutrinos is an infinite number of sums over each of probability distributions of N_m . To make this approach more tractable, the problem can be posed in terms of generating functions. The generating function for a probability distribution p_k is given by:

$$P(t) = \sum_{k=0}^\infty p_k t^k, \tag{7.6}$$

where t is an auxiliary variable. The original distribution can be recovered from this function through differentiation:

$$p_k = \frac{1}{k!} \left. \frac{d^k P(t)}{dt^k} \right|_{t=0}. \tag{7.7}$$

A key feature of the generating function is that the generating function for the sum of two random numbers is the product of the generating functions for each individual random number:

$$(P + G)(t) = P(t) \cdot G(t). \tag{7.8}$$

⁵⁶ Indeed, the generating function is simply a special case of a z-transform.

This can be seen in an analogy to z-transforms⁵⁶: the probability density function for the sum of two continuous random numbers is a convolution of their individual probability density functions, and a convolution can be posed as a product in z-space.

First, the total number of detected neutrinos is written as:

$$k = \sum_{m=1}^\infty k^{(m)}, \tag{7.9}$$

where $k^{(m)} = m \cdot N_m$. The probability distribution of $k^{(m)}$ is:

$$p_{k,m} = \begin{cases} p_{N=\lfloor k/m \rfloor, m} & k \equiv 0 \pmod{m} \\ 0 & \text{otherwise} \end{cases}, \quad (7.10)$$

where $p_{N,m}$ is the poisson probability of finding N sources with a mean of x_m .

The generating function of $p_{k,m}$ is, therefore: [187]

$$P^{(m)}(t) = \exp(x_m(t^m - 1)). \quad (7.11)$$

So the generating function of p_k , the probability of k total detected neutrinos is the product of these:

$$P(t) = \prod_{m=1}^{\infty} P^{(m)}(t), \quad (7.12)$$

$$= \exp\left(\sum_{m=1}^{\infty} x_m(t^m - 1)\right). \quad (7.13)$$

With this, the probability of finding k neutrinos in a single bin can be calculated from the differential source count function. The value of x_m can be found analytically if the differential source count function follows a broken power law [187]:

$$\frac{dN}{dS} = A \begin{cases} \left(\frac{S}{S_b}\right)^{-n_1} & S > S_b \\ \left(\frac{S}{S_b}\right)^{-n_2} & \text{otherwise} \end{cases}, \quad (7.14)$$

where S_b is the location of the break, at which the index of the power law changes from n_2 to n_1 . In this analysis, only a single break is considered, although the expression for x_m can be generalised to an arbitrary number of breaks.

When considering a point source population of galactic origin, the total number of sources can change as a function of position in the sky. Mishra-Sharma, Rodd, and Safdi [187] captured this by defining a template that specifies the overall normalisation, A , of the differential source count function for each spatial bin.

When the measured data has a diffuse background – such as atmospheric neutrinos – the number of neutrinos associated with this background component will be Poisson distributed. The total number of counts in a bin will then be a sum of the number of neutrinos produced by point sources and the number produced by the diffuse background. This can be handled in a similar way as above, by multiplying the generating functions of each distribution.

The total likelihood of seeing a certain number of neutrinos in each bin (\vec{d}) is the product of the probabilities for each individual bin:

$$\mathcal{L}(\vec{d}) = \prod_{i=0}^{N_{\text{bins}}} p_{k=\vec{d}_i}. \quad (7.15)$$

The calculation of x_m for each bin and the total likelihood is performed by the `NPTFit` software package [187]. This software can handle an arbitrary number of templates, each of which can either be non-Poissonian or Poissonian distributed. For each template, a spatial distribution is given to `NPTFit`, and the software fits an overall normalisation for the template to the data. The other parameters for the fit are: the break in the power law S_b , and the indexes n_1 and n_2 for each non-Poissonian template.

7.0.1 Application to IceCube

Using Non-Poissonian Template fitting, Lee *et al.* have shown that the observed excess of gamma rays in Fermi Gamma Ray Telescope data coming from the galactic center are better described by a population of faint point sources, rather than a diffuse emission, which would be expected if the gamma rays are produced by the annihilation of dark matter [164]. The search for point sources in IceCube is a problem similar in nature. There is an observed flux of neutrinos that have an origin outside the solar system, but it is unknown if these neutrinos are being produced by point sources or a diffuse emission. While, up until now, NPTF has only seen application in gamma ray astronomy, the statistical distribution is agnostic to what is being counted.

Nonetheless, applying this to IceCube does bring new challenges. Unlike Fermi, IceCube's sensitivity varies by orders of magnitude across the sky. Also, IceCube's angular resolution is much poorer than Fermi's, with a much larger long-tail component.

7.1 Data sample

The search for point sources is fundamentally a spatial one, and so angular resolution is of paramount importance. For this reason, we consider only muon neutrinos for this analysis, as the median angular error is typically less than 1° for energies greater than 1 TeV.

While this analysis considers neutrinos that come from all directions, the distribution of these neutrinos is very different when they originate from the northern sky in comparison to the southern sky. The north sky is defined to be all declinations $> -5^\circ$, and the south

is the complement. This definition ensures that all “northern” neutrinos appear to be coming from below the horizon from the point of view of the south pole. These neutrinos are called “up-going” and must travel through tens to thousands of kilometers of rock to reach IceCube. As such, the background of cosmogenic muons is negligible due to this natural shielding. While neutrinos are generally assumed to be very penetrating, their probability of interaction grows with energy and for neutrinos with greater than 100 TeV the cross section is large enough that the Earth becomes opaque.

The detector is sensitive to these high energy neutrinos if they are “down-going”, meaning they come from the southern sky above the experiment. However, the down-going neutrinos are obscured by a large background of muons, as the 1.5 kilometers of ice above the detector provides relatively little shielding. These muons are largely contained to low energies, due to being products of a steeply falling spectrum of cosmic rays. To combat this background, aggressive cuts remove most events below 100 TeV. In addition, the event selection for the southern sky is chosen to produce approximately the same number of events as the northern sky. As the spectrum of astrophysical neutrinos is expected to also be a steeply falling function, the fraction of background events, and thus background to signal ratio is larger than in the northern sky.

The result is a data sample that is split between low energy events in the northern sky, and high energy events in the southern sky. The center of the galaxy is located just below 5° in declination, with side of the galaxy extending into the northern sky region. This allows IceCube to probe both the high and low energy regions of an emission spectrum.

7.1.1 Data sets

The IceCube collaboration organises data into sets that span one year. Each data set is given a name based on the configuration of the experiment during that year. For example, IC40 is the data set that was recorded when the detector had 40 strings installed. After the year 2011, detector construction finished with 86 strings installed and the experiment no longer changed configuration, so these data sets are also suffixed by the year the data was taken.

While each data set spans one year of calendar time, the detector is not taking data for this entire time period. The detector needs to be periodically taken offline for maintenance and calibration duties, and for construction while the detector was still under construction. The actual amount of time the detector is online and measuring is given by the “live-time”. Each data set used in this analysis, and their respective live-time is shown in table 7.1.

Table 7.1:

The data sets used in the analysis, and their respective live-time measured in days.

Name	Dates	Live-time
IC40[70]	2008/04/05 to 2009/05/20	375.5
IC59[61]	2009/05/20 to 2010/05/31	348.1
IC79[61]	2010/05/31 to 2011/05/12	315.5
IC86-2011[62]	2011/05/13 to 2012/05/15	332.6
IC86-2012[55]	2012/05/15 to 2013/05/02	330.3
IC86-2013[55]	2013/05/02 to 2014/05/06	359.9
IC86-2014[55]	2014/05/06 to 2015/05/18	367.2

7.1.2 Background model

Estimating the background of atmospheric neutrinos and cosmogenic muons is challenging. When cosmic rays strike the atmosphere, the resulting products from the interaction are tightly focused toward forwards — the same direction as the cosmic ray. The physics involved in these forward reactions is not well understood. Collider experiments are the main source of data for high energy reactions; however, forward reaction products travel down the beam pipe of the collider where detectors cannot be placed. Thus, there is a substantial amount of uncertainty involved in simulating the background component through Monte-Carlo techniques.

Fortunately, it is possible to estimate the background from the data alone. In a search for point sources, the background is all events that did not originate from a point source. Removing the point sources from the data cannot be done exactly⁵⁷, but can be approximated. This relies on the azimuthal symmetry of the detector’s effective area, and near exact alignment of the azimuthal axis with the Earth’s own rotational axis. As the Earth spins, so, too, does the detector; but the response of the detector is very nearly unchanged. Thus, the azimuthal information of events can be removed by *scrambling* them: assigning them random right ascension values. Point sources show up as hot spots, so any point sources in the data will be smeared along a line of constant declination.

The background model should be an estimate of the mean of a Poisson distribution. However, it will have some statistical noise as it is computed by binning data events. The scrambling procedure can be repeated many times and the result averaged, drastically reducing the amount of statistical noise in the right ascension dimension. Some small amount of noise will still remain when the model is viewed in declination. To remove this, the model is convolved with a Kent distribution with mean 1.08° , which is the median angular resolution at ~ 1 TeV.

⁵⁷ If it were, we would already have the answer we seek.

7.1.3 Simulation

Understanding the performance this statistical method applied to IceCube requires the simulation of neutrinos interacting with the detector. Each data set from IC40 to IC86-2011 has a complementary set of simulated events. The IC-2012 to IC-2014 data sets use a common set of simulated events, as they use the same cuts and reconstruction algorithms.

These events are generated according to an E^{-1} spectrum, and are subsequently re-weighted to a flat spectrum, with a correction applied for the survival probability of traveling through the Earth. Each event is then reconstructed using the same algorithms used for the data sample, and both the true and reconstructed quantities (such as energy, direction, etc.) are stored. The same cuts used in the data sample are then applied, so that the simulated events are a representative sample of the fraction of incident neutrinos that are measured in the detector. Each set of simulated events is then weighted by the live-time of its corresponding data set(s) and concatenated into one single set of events.

7.1.4 Binning

The data is binned into a single reconstructed energy bin, as the NPTF statistical test does not make use of energy information. The reconstructed directions of events are binned into a healpix map with an n_{side} of 64. Healpix [120] is an equal area pixelisation of spherical coordinates commonly used in astrophysics and astronomy. The pixel size is determined by the number of pixels, which is equal to $12 \times n_{\text{side}}^2$ – a value chosen to optimise the sensitivity shown in section 7.5.

7.1.5 Low energy cut

The low energy part of data distribution is dominated by background events. Much background can be removed by removing low energy events. While the data sets discussed above already include cuts at low energy, the NPTF analysis prefers a slightly higher low-energy cut. An exploration of various values of an additional low energy cut showed that the sensitivity in section 7.5 was optimised for a cut of $10^{3.5}$ GeV.

7.2 Detector effects

IceCube is far from a perfect measuring device, as discussed in chapter 4. It has significant uncertainties in energy and direction, as well

a detection efficiency that varies in energy and declination due to the effects discussed in the previous section. Uncertainties in energy reconstruction are not relevant to this analysis due to the choice of a single energy bin. The uncertainty in reconstructed direction are handled by the Point Spread Function discussed in section 7.2.2.

7.2.1 Effective area

The detection efficiency can be summarised in terms of an “effective area”, which also goes by the name of “exposure” in the astrophysics community⁵⁸. This quantity also incorporates the cross section of neutrino interactions, and has units of cm^2 so that it can be multiplied by a flux (units of $\text{cm}^{-2} \text{s}^{-1}$) to yield a rate of measured events.

⁵⁸ A term that derives from optical astronomy, where a light sensitive detector would be exposed to a certain region of the sky for a period of time.

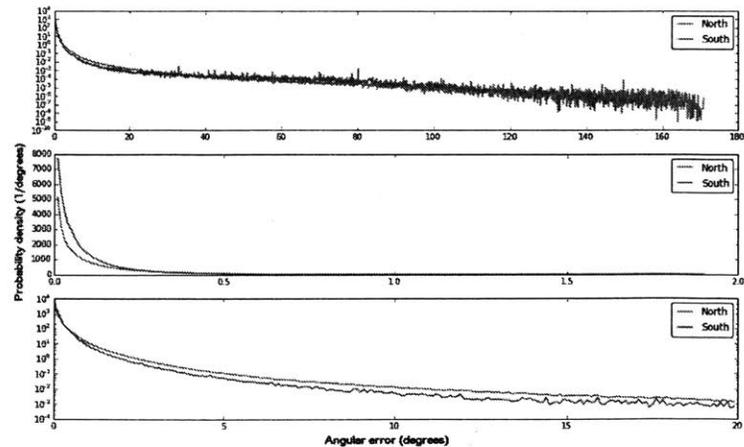
The effective area is calculated by summing the weights of the simulated neutrino events. Each weight is inversely proportional to the total number of generated events (before cuts), thus the sum of these weights will be proportional to the fraction of events that survive the cuts. These event weights have units of $\text{GeV cm}^2 \text{sr s}$ so that multiplication by an energy and angular flux distribution gives a total number of expected measured events.

In this analysis, simulated events are first re-weighted according to an assumed energy distribution of E^{-2} . These re-weighted events are then binned in energy and galactic coordinates to form an effective area map. This map will contain a small amount of statistical noise due to the finite number of simulated events used. This noise is removed by using the same scrambling and smoothing procedure described in section 7.1.2.

7.2.2 Point Spread Function

Figure 7.1:

The distribution of angular error in degrees for both the northern and southern sky. Top: The distribution over the full range of angles. Middle: The distribution zoomed into $< 2^\circ$. Bottom: The distribution zoomed into $< 20^\circ$.



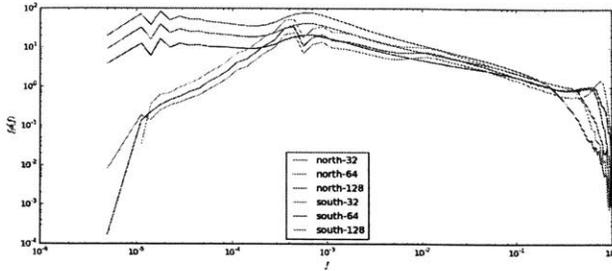


Figure 7.2:
Fractional flux distribution for both the northern and the southern sky with various choices of healpix n_{side} .

The angular resolution is summarised in terms of a Point Spread Function (PSF). This is calculated by binning the angular difference between the true and reconstructed directions of simulated events.

The PSF is shown in figure 7.1. It has a long tail, so an event can potentially be reconstructed as coming from a direction very far from the true direction of the neutrino. As a consequence, a bin where a point source is located will not contain 100% of the flux of the source. Instead it will only contain a fraction of the flux, and the remainder will be distributed amongst other bins in the map. This must be accounted for when fitting a non-Poissonian template, as it “softens” the probability distribution of counts away from being very sharply peaked at zero counts. The NPTF method uses a Fractional Flux Distribution, $\rho(f)$, to describe the probability of a bin having a certain fraction of flux, f . While this can be derived from the PSF, it is more accurate to generate the FFD from the simulated events directly.

The FFD is similar to a probability distribution; however, it is normalised as follows:

$$\int_0^1 f \rho(f) df = 1, \quad (7.16)$$

to ensure that the total amount of flux is conserved.

Generation of the FFD is performed by injecting point sources – described in section 7.3 – into an empty healpix map. The map is divided by the total number of injected events, so that each bin represents a fraction of the total flux. Then all non-zero bins are histogrammed, and the histogram is normalised using equation 7.16.

The FFD for this analysis is shown in figure 7.2. It is incorporated into the non-Poissonian distribution by modifying the expression for x_m to integrate over the distribution of fractional counts:

$$x_m = \int_0^\infty \left[\int_0^1 \rho(f) \frac{(fS)^m}{m!} e^{-fS} df \right] \frac{dN}{dS} dS. \quad (7.17)$$

This can also be posed as a modification of the source count distribution:

$$\frac{dN}{dS}(S) \rightarrow \frac{d\tilde{N}}{dS}(S) = \int_0^1 \frac{\rho(f)}{f} \frac{dN}{dS}(S/f) df. \quad (7.18)$$

The NPTFit software package can take a FFD as input and incorporate it into the calculation of x_m . A separate FFD is generated for each template discussed in section 7.5, as each template constrains events to different locations in the sky, and the effective area varies by location.

7.3 Point source injection

To gauge the correctness and effectiveness of this NPTF analysis, point sources are simulated and injected into the data. Each test assumes a template for the spatial distribution of the sources, and a differential source count function.

The number of point sources to be injected is drawn from a Poisson distribution with mean

$$N_{\text{ps}} = \int_0^\infty \frac{dN}{dS} dS. \quad (7.19)$$

The flux for each point source is sampled from a probability density derived from the differential source count function:

$$p(S) = \frac{1}{N} \frac{dN}{dS}(S). \quad (7.20)$$

The spatial position of each point source is sampled from the template.

Next, for each point source, the appropriate number of simulation events are selected, based on the point source's flux. This is done as follows. The detector response is not uniform in right ascension, so a band is placed at the point source location, spanning the entire range in right ascension and a small range in $\sin(\text{declination})$. All simulation events whose true direction fall into this band are selected, and re-weighted according to the assumed point source energy spectrum and flux. Finally, these events are re-weighted according to the solid angle of the band.

The weight of each event is now unitless, and gives the mean number of counts that this event is expected to contribute to the flux of this point source. A number of counts, n , is drawn from a Poisson distribution for each event according to this mean. A new sample of events is created by replicating each event a number of times equal to its respective value of n . For most events, this will be zero, removing them from the new sample.

This new sample forms the events for the point source; however, at this point they are distributed over a large range in right ascension. To re-position them, the events have their reconstructed coordinates rotated by the difference between their true coordinates and the point source's location. Finally, the reconstructed directions of the events are histogrammed into a healpix map, and this map is added to a background map generated from scrambled data events.

7.4 Inference framework

A Bayes' factor is used to test if the measured data is better described by a point source hypothesis, or a diffuse emission only hypothesis:

$$\mathcal{B}_{\text{NP/P}} = \frac{\mathcal{L}_{\text{NP}}(\vec{d})}{\mathcal{L}_{\text{P}}(\vec{d})}, \quad (7.21)$$

where \mathcal{L}_{NP} is the marginal likelihood including a non-Poissonian template, and \mathcal{L}_{P} is the marginal likelihood with Poisson-only statistics.

Parameter	Lower bound	Upper bound
A	$2.6 \times 10^5 \text{ GeV cm}^2 \text{ s}$	$2.6 \times 10^{19} \text{ GeV cm}^2 \text{ s}$
F_b	$10^{-15} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ s}^{-1}$	$10^{-4} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ s}^{-1}$
n_1	2	10
n_2	-7	1

Table 7.2:

The priors for each parameter of the non-Poissonian model.

The Bayes' factor is calculated using the MultiNest software package [100, 102, 103] as described in chapter 2. The priors on each parameter are flat in linear space for n_1 and n_2 and log space for the remainder, with bounds shown in table 7.2. The prior on the power law break (F_b)⁵⁹ was chosen to bracket a range between a population of sources that contribute no counts to the data, to one that dominates the data. The prior on the normalisation (A) was chosen to allow a fit of up to $\mathcal{O}(10^6)$ sources. The prior on the power law indices was chosen to cover the range of physical models⁶⁰.

Rather than using a confidence region, which is difficult to profile over such a large dimensional space using the samples generated from MultiNest, a probability ratio

$$\mathcal{M}(d; \phi) = \frac{\mathcal{L}_{\text{NP}}(\vec{d}|\phi)\pi(\text{NP})}{\mathcal{L}_{\text{P}}(\vec{d})\pi(\text{P}) + \mathcal{L}_{\text{NP}}(\vec{d})\pi(\text{NP})} \quad (7.22)$$

is plotted over the coordinate range $\phi = N^{\text{PS}}, \bar{F}^{\text{PS}}$. The prior probabilities of the Poissonian and non-Poissonian hypothesis are taken to be equal: $\pi(\text{P}) = \pi(\text{NP}) = 0.5$.

⁵⁹ The prior on the power law break is specified in units of flux (F_b), then converted to counts (S_b) by NPTFit.

⁶⁰ For the number of sources in the distribution to converge, it is necessary that $n_1 > 2$ and $n_2 < 1$.

Here, the likelihood of ϕ is computed by marginalising over the remaining model parameters θ :

$$\mathcal{L}_{\text{NP}}(\vec{d}|\phi) = \int \mathcal{L}_{\text{NP}}(\vec{d}|\phi, \theta)\pi(\theta)d\theta \quad (7.23)$$

This ratio is equal to the point-wise posterior model probability ratio when the Poissonian model is strongly preferred:

$$\mathcal{M}(d; \phi) \approx \frac{\mathcal{L}_{\text{NP}}(\vec{d}|\phi)}{\mathcal{L}_{\text{P}}(\vec{d})} \frac{\pi(\text{NP})}{\pi(\text{P})} \quad \mathcal{B}_{\text{NP/P}} \ll 1 \quad (7.24)$$

$$= \mathcal{B}_{\text{NP/P}}(\phi) \frac{\pi(\text{NP})}{\pi(\text{P})} \quad (7.25)$$

$$= \frac{p(\text{NP}|\phi, \vec{d})}{p(\text{P}|\vec{d})} \quad (7.26)$$

and is equal to the posterior over prior probability ratio when the non-Poissonian model is strongly preferred:

$$\mathcal{M}(d; \phi) \approx \frac{\mathcal{L}_{\text{NP}}(\vec{d}|\phi)}{\mathcal{L}_{\text{NP}}(\vec{d})} \quad \mathcal{B}_{\text{NP/P}} \gg 1 \quad (7.27)$$

$$= \frac{p(\phi|\vec{d}, \text{NP})}{\pi(\phi)} \quad (7.28)$$

This is motivated by the need for a metric, similar to the frequentist likelihood ratio, which marginalises over the parameters which are not being plotted, instead of profiling them as in a confidence region. The posterior is not suitable for this, as in the case where the null hypothesis is strongly preferred, the regions will be highly prior dependent. The point-wise Bayes factor is prior-independent, but is difficult to control when there is a strong signal present. By using equation 7.22, the metric naturally scales between the point-wise Bayes factor when no signal is present, to the posterior-prior ratio when a signal is present.

The quantity also has another interpretation, as it is related to the mutual information between the data and model parameters. In fact, the point-wise mutual information is simply

$$\text{pmi}(d; \phi) = \ln \mathcal{M}(d; \phi) + h(\text{NP}) \quad (7.29)$$

7.4.1 Injection cross-checks

To test the method, a population of point sources was injected on top of the background model. The NPTF model was then sampled from using this simulated data, to test if the injected model parameters can be recovered.

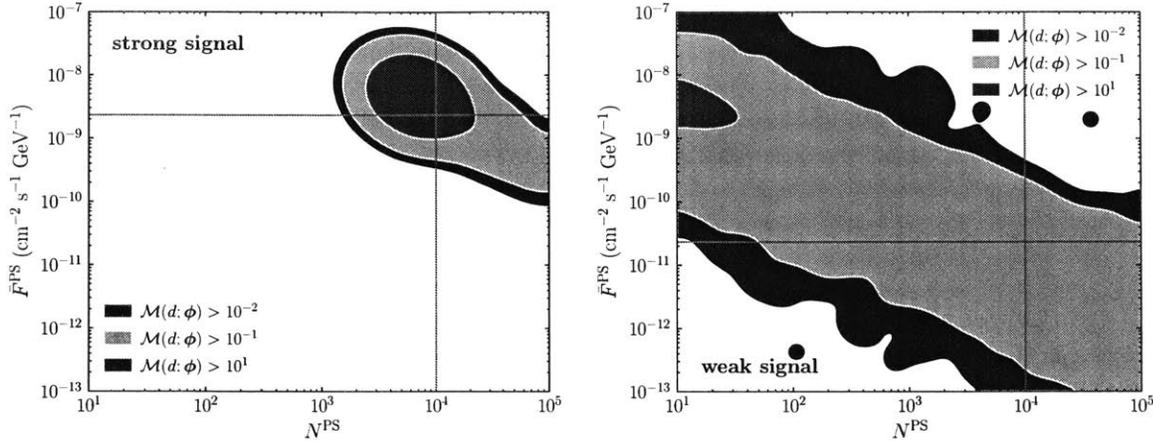


Figure 7.3 shows two injected populations using the full sky isotropic template. Both have the same expected number of sources of 10^4 , and source count function shape of $n_1 = -n_2 = 2.5$. The strong signal example – shown on the left – uses a break in the source count function of $10^{-9} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$. The weak signal example – shown on the right – uses a break of $10^{-11} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$.

The strong signal is recovered correctly. The weak signal shows a large degeneracy between the number of sources and the average flux per source, which is to be expected when the population is barely detectable.

7.5 Sensitivity

To determine the significance of a given value of $\mathcal{B}_{\text{NP/P}}$, the NPTF analysis is run on many random scrambles of the data with no point sources injected. The resulting set of Bayes' factors forms the null distribution of $\mathcal{B}_{\text{NP/P}}$, from which a p-value can be calculated – essentially treating the Bayes' factor as a test statistic.

A similar distribution can be generated for a signal hypothesis. For a given power law break F_b , normalisation A , and indexes n_1 and n_2 , point sources are injected into a random scrambling of the data. The resulting set of Bayes' factors form a signal distribution of $\mathcal{B}_{\text{NP/P}}$. The sensitivity is defined as the values of S_b , A , n_1 and n_2 where the 90% of the signal distribution has a p-value of 0.5 or less⁶¹. In addition, a distribution of $\mathcal{M}(\vec{d}; \phi)$ values can be generated, and from this, a 10%, 50%, and 90% quantile are found for the expectation of the $\mathcal{M}(\vec{d}; \phi) = 10^{-1}$ contour.

The sensitivity and expected $\mathcal{M}(\vec{d}; \phi)$ contour for four choices of templates are shown in figure 7.4. These templates are:

Figure 7.3:

The $\mathcal{M}(\vec{d}; \phi)$ contour plot for two examples of simulated data. In both, the intersection of the red lines shows where the injected population lies in parameter space. **Left:** A population with $N^{\text{PS}} = 10^4$ and $F_b = 10^{-9} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$ was injected onto the background model. **Right:** A population with $N^{\text{PS}} = 10^4$ and $F_b = 10^{-11} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$ was injected onto the background model.

⁶¹ This is equivalent to where the 10th percentile of the signal distribution is equal to the 50th percentile of the null distribution.

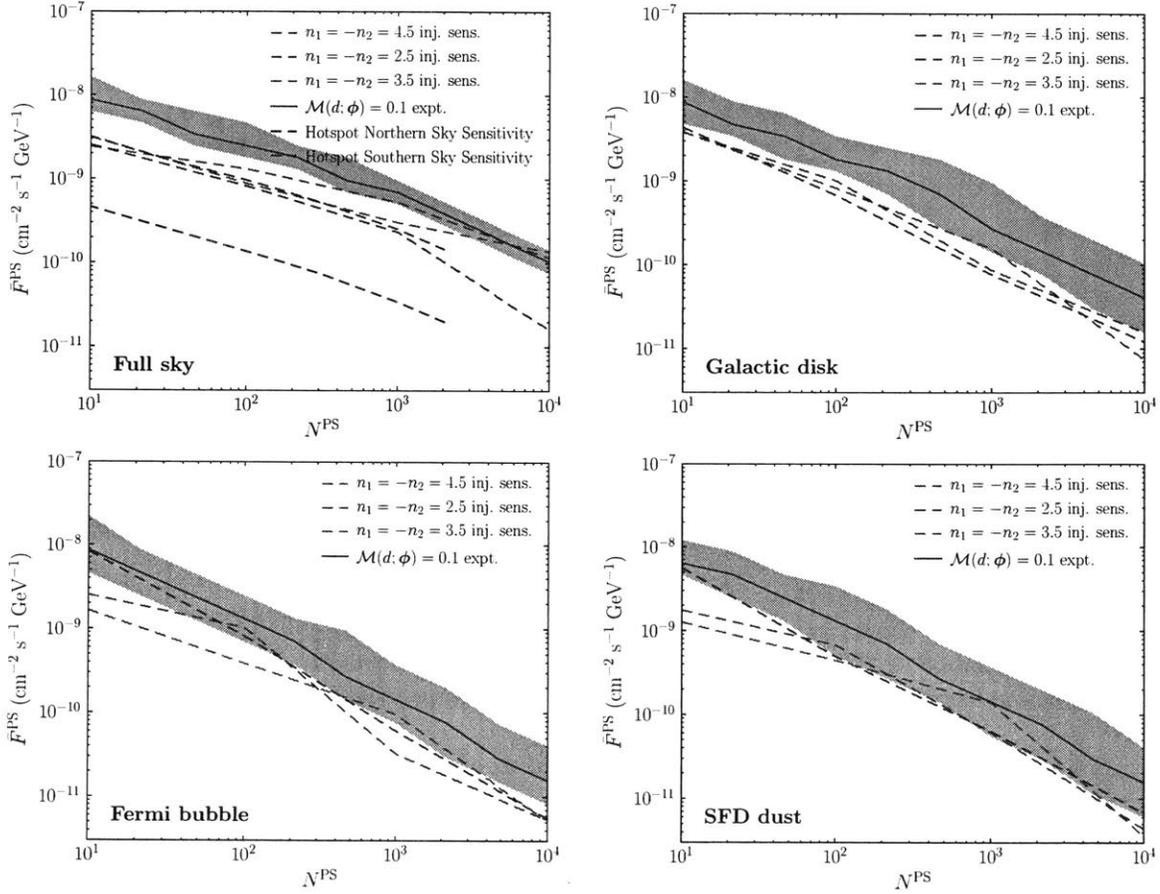
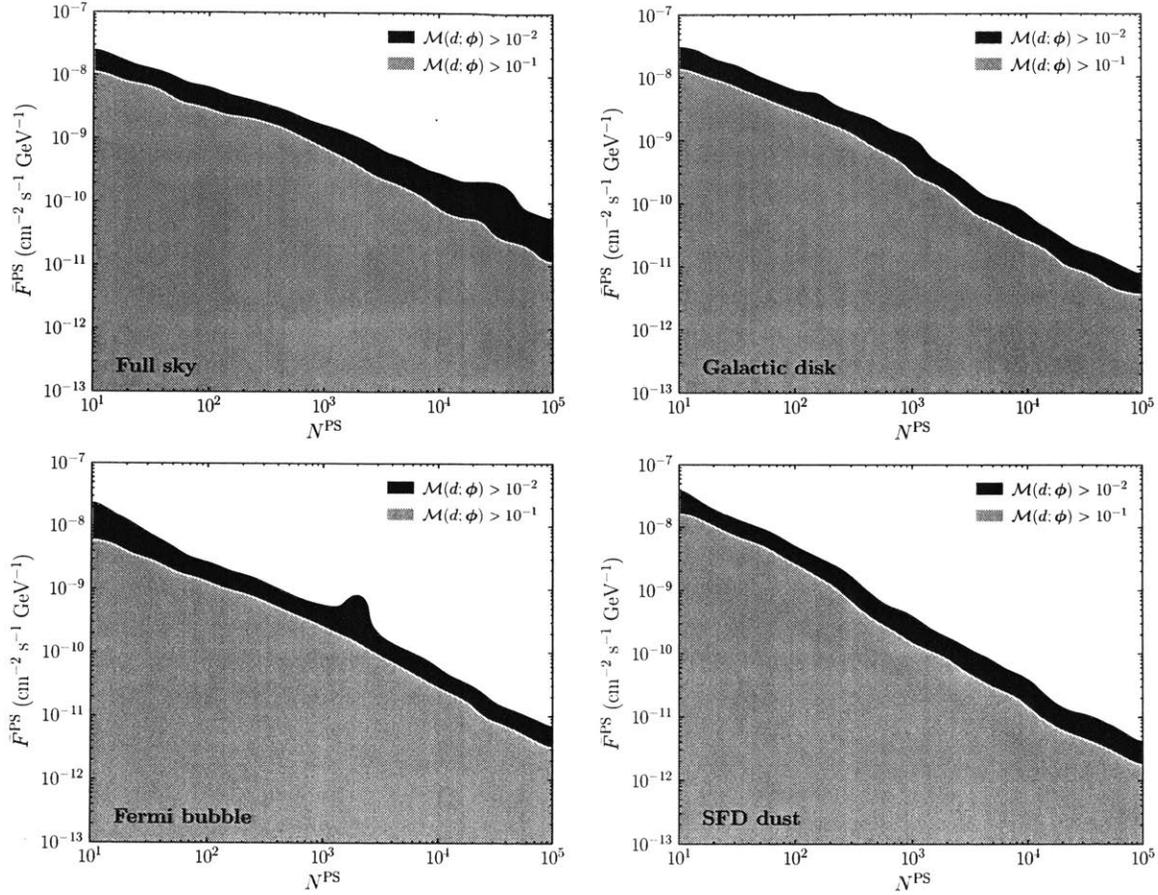


Figure 7.4:

Sensitivity and expectations for four choices of templates. Three sensitivity lines are drawn, each for injecting a different choice of the power law indices n_1 and n_2 . The vertical axis gives the average flux per source normalisation for the population of point sources. The horizontal axis gives the average number of sources in the population. Both numbers are derived from the differential source count function. The black line shows the sensitivity of the latest published all sky search for point sources [55]. The blue line shows the median expected $\mathcal{M}(\vec{d}; \phi) = 0.1$ contour, and the band shows the 10% and 90% quantiles.

- Full sky: An isotropic template.
- Galactic disk: A general model for the galactic plane used in the Fermi gamma ray excess study.
- Fermi bubble: A model of the Fermi bubbles.
- SFD dust: A map of dust in the galaxy prepared by Schlegel, Finkbeiner, and Davis [216]. This template is used as a proxy for the distribution of gas. As discussed in section 6, neutrinos are expected from supernova remnants located near molecular gas.

The northern and southern sky sensitivities for the most recently published all sky hotspot search [55] are also shown in these sensitivity plots.



7.6 Results and summary

Template	$\mathcal{B}_{\text{NP/P}}$	p-value
Full sky	0.452	66%
Galactic disk	0.377	74%
Fermi bubble	0.389	45%
SFD dust	0.400	33%

The Bayes' factors and p-values are detailed in table 7.3. The $\mathcal{M}(\vec{d}; \phi)$ contours are shown in figure 7.5.

All of the Bayes' factors and p-values are in agreement with the null hypothesis. The $\mathcal{M}(\vec{d}; \phi) = 10^{-1}$ contours lie within the expected limits from the sensitivity analysis. Curiously, the Fermi bubble template shows a small bump in the 10^{-2} contour in figure 7.5. This is likely due to a fluctuation in the sampling, as a possible signal should also show a similar bump in the 10^{-1} contour.

Figure 7.5:

The $\mathcal{M}(\vec{d}; \phi) = 10^{-1}$ and 10^{-2} contour for the NPTF analysis on measured data.

Table 7.3:

The results of running the NPTF analysis on the measured data.

The sensitivity of this NPTF analysis at low numbers of sources is approximately an order of magnitude less than the previous Hotspot all sky search [55] shown in figure 7.4. Less sensitivity in this region should be expected, as the point source population model that the NPTF analysis fits is more general than the single source model that the previously published result uses. It should also be noted that the previously published result used an un-binned likelihood method that can account for the reconstruction uncertainty on an event by event basis. Previous internal studies in the IceCube collaboration show that this additional information significantly improves the sensitivity.

Nevertheless, this is the first time that a complete population model has been tested and constrained on astrophysical neutrino data. Future directions of this topic may include the construction of an un-binned statistical test that takes a population model as an input, much as the NPTF does. This would require the incorporation of spatial correlations into the statistical distribution, which would likely improve the sensitivity of the method. At the present data shows no signs of yielding strong steady state emission sources in the near future, new analysis techniques that better model weak, below threshold populations should be a high priority.

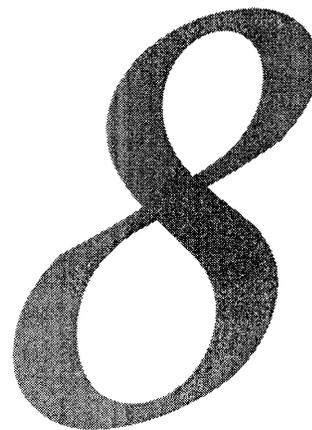
Artificial neural networks

The concept of an *artificial neural network* was first formalised by McCulloch and Pitts [182] in 1943. In that paper, the authors provided a model for a neuron, although its implementation was not feasible with the bulky vacuum tube technology of the era [236]. While McCulloch and Pitts showed that any arithmetic or logical function could be computed by neural networks, the concept of learning was not formalised until 1949 when Hebb published *The Organization of Behavior* [134]. In this book, Hebb proposed a scheme for training a neural network that we now call *Hebbian learning*. In its most famous incantation, this law reads:⁶²

“Cells that fire together, wire together.”

This heralded the creation of a new field of statistics and computer science called *machine learning*, where predictions and decision rules are derived algorithmically from data instead of following static pre-programming devised by a software developer.

The most famous example of a machine learning algorithm from this founding era is the *perceptron*, invented by Rosenblatt in 1957 [210]. The perceptron is characterised by a series of *weights*, which describe a decision boundary in the multi-dimensional space of its input variables. The choice of weights to use is posed as an optimisation problem: the perceptron is given an input example⁶³ and the output of the perceptron is computed – a value which is either 0 or 1, depending on the side of the decision boundary the input lands on. This output is then compared to the desired output of the perceptron⁶⁴; and, if the output is not correct, the weights are ad-



⁶² This is paraphrased from Hebb's original words: “When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.”

⁶³ Terminology which means a single data point from a data set.

⁶⁴ Often called the “true” value, each example is paired with a known output value that is assumed to be correct.

justed so that the perceptron will produce the correct output in the future.

This process of continuously adjusting the weights based on an optimisation objective is called “training” in machine learning parlance. When examples are paired with true output values – like in the perceptron – the additional qualification of *supervised* training is used. This is intended to invoke an image of the machine learning practitioner supervising the algorithm and correcting it when it makes a mistake.

This idea was further extended by Widrow and Hoff in the development of the ADALINE system [229]. This algorithm is similar to the perceptron, with the addition of a new weight update rule based explicitly on least squares optimisation. While ADALINE enjoyed some early successes, more complicated problems remained unapproachable by these simple linear classifiers, and artificial neuron based machine learning research began to stagnate [236]. The concept of chaining the output of one set of artificial neurons into the inputs of another to form a multi-layered network of neurons with greater expressibility was considered [217]; however, this created an optimisation problem that was intractable with the methods and hardware available at the time. Possibly the final death knell came from Minsky and Papert in their book “Perceptrons” [186], which explored the limitations of neural networks and left the impression that research in this area had reached a dead end [233].

This began a period of time that is now often called the “AI winter”⁶⁵, where interest, and thus funding, in AI and machine learning began to wane. Research did continue, and some important foundations for the field were laid down in this period. In particular, Seppo Linnainmaa published a method for efficiently calculating the gradient of multi-dimensional scalar function [169, 170]. This technique is crucial for training modern neural networks with millions of weights, and will be discussed in section 8.2.

Interest in neural networks began to rise again in the late 1980s to early 1990s with the emergence of *convolutional neural networks*. Invented by Yann Le Cun et al. in 1990 [163], these neural networks are structured to perform convolutions, which are a common filtering operation found in image processing. Organised into five to seven different layers of neural convolutions, these networks were considered “deep” by the standards of neural networks of that time. They found application in handwriting recognition, most famously in automated post-code reading systems for mail sorting.

The depth of a neural network has a large effect on how easy it is to train. While computer processors of that period were following a Moore’s law exponential increase in processing power, the processors were primarily designed to compute highly complicated

⁶⁵ It should be mentioned that while artificial intelligence and machine learning have a large overlap, they are not synonymous. Machine learning specifically refers to algorithms whose behaviour is derived from data, while artificial intelligence is any system that appears to express intelligent behaviour. A linear regressor is a kind of machine learning algorithm that is usually not considered AI, while the pre-programmed behaviour of a computer opponent in a video game is an AI algorithm that is not considered part of machine learning – although this latter case will likely cease to remain a valid example in the future.

algorithms⁶⁶. Neural networks are comparatively simple: they usually only involve multiplications and additions, with very simple branches if any. They are also parallelisable, as the outputs of the neurons in each layer can be computed independently of each other.

Processors that could handle such simple parallel operations were generally rare and expensive enough to be outside the reach of the typical academic. As it happens, computer graphics is similar in terms of its processor requirements; but, with the added benefit that consumer interest in computer games drives industry investment in cost efficient parallel processors. These *graphical processing units* (GPUs) were initially, in the 1990s, highly specialised to graphical rendering. In the early 2000s, GPU manufacturers added some general purpose computation ability due to increasing complexity in computer game graphics. There were some initial forays into using these more flexible GPUs for training neural networks [191]; but, the real turning point happened in 2008 when NVIDIA introduced the first truly general purpose graphical processing unit (GPGPU)⁶⁷. Recognising the potential demand for these new processors in science and industry, NVIDIA created and launched a programming environment soon after.

Scientists were quick to start using GPGPUs for neural network training [205]. In 2012 Krizhevsky and Hinton showed that convolutional neural networks could outperform all existing machine learning methods in image recognition by a significant margin [162], which was followed by super human performance in 2015 [133].

8.1 Neural classifiers

Although the invention and much of the development of artificial neural networks was inspired by biology, they can also be motivated from the ground up on practical terms. Suppose we have a dataset, where each data point in this set is called an example. These examples all have a corresponding *classification* associated with it. For example, in physics each example could either be signal or background, or medicine it could be positive or negative for a certain disease. This can be written as a set of pairs

$$\mathcal{S} = \{(\vec{d}_i, c_i) : i \in \text{dataset}\}, \quad (8.1)$$

where \vec{d} are the examples, and c are the matching classifications.

Suppose now, we have a new set of examples $\mathcal{T} = \{\vec{d}_i\}$ where the correct classification is unknown. The goal of machine learning is to derive a classification function f from \mathcal{S} , which can then be used to predict the correct classification for the examples in \mathcal{T} .

⁶⁶ In terms of the frequency of branch points.

⁶⁷ At that time, computer game graphical requirements had become so heterogeneous that it became impossible to effectively allocate specialised processing units. Some units would be sitting idle while the workload was bottlenecked on other units that performed a different task. Creating a completely general purpose processor meant that the entire processor could be utilised no matter what the workload was.

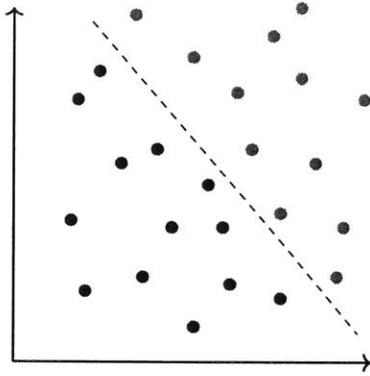


Figure 8.1:
A set of examples with their matching classifications denoted in colour. A line can easily divide the data set.

If \mathcal{S} is simple, the examples can be divided by a line as shown in figure 8.1. A prediction for the correct classification of the examples in \mathcal{T} can then be created by finding which side of this line the examples fall on. This can be done mathematically by defining the line in terms of a unit vector \hat{w} and offset b . The distance of any example \vec{d} from the line is

$$y = \vec{d} \cdot \hat{w} + b. \tag{8.2}$$

The sign of y gives which side of the line the example is on. If the binary classification of each example is $c \in \{0, 1\}$, then the predicted classification can be found by feeding y into a Heaviside theta function:

$$f_{\hat{w},d}(\vec{d}) = \Theta(\vec{d} \cdot \hat{w} + b), \quad \Theta(x) = \begin{cases} 0 & x < 0 \\ 1 & \text{otherwise} \end{cases}. \tag{8.3}$$

It remains to find \hat{w} and b to optimally divide the data set. This can be done by defining a function to be optimised, which is typically called the *loss function* in machine learning parlance. A reasonable choice for the purposes of illustration is a least squares metric, taken over all the examples in \mathcal{S} :

$$L(\hat{w}, b) = \sum_{\vec{d}, c \in \mathcal{S}} (c - f_{\hat{w},d}(\vec{d}))^2. \tag{8.4}$$

The optimal choice of \hat{w} and b minimises this metric:

$$\hat{w}_{\text{opt}}, b_{\text{opt}} = \operatorname{argmin}_{\hat{w}, b} L(\hat{w}, b). \tag{8.5}$$

This minimisation can be done with a number of methods. The most effective is gradient decent; however, the gradient of the Heaviside theta is always zero. Instead of using the Heaviside theta directly, we can use a similar looking function with a non-zero gradient, such as a sigmoid function shown in figure 8.2:

$$f_{\hat{w},d}(\vec{d}) = \sigma(\vec{d} \cdot \hat{w} + b), \quad \sigma(x) = \frac{1}{1 + e^{-x}}. \tag{8.6}$$

With this choice, the output of f is no longer a binary 0 or 1. We can use this to represent the confidence in the classification⁶⁸. For example, a data point located on the boundary will have an output value of 0.5, signifying even odds on being part of one class or the other. The slope of this function, and thus the degree of uncertainty in the classification, can be controlled through a multiplicative factor A in the argument to the sigmoid function:

$$f_{\hat{w},d}(\vec{d}) = \sigma(A(\vec{d} \cdot \hat{w} + b)). \tag{8.7}$$

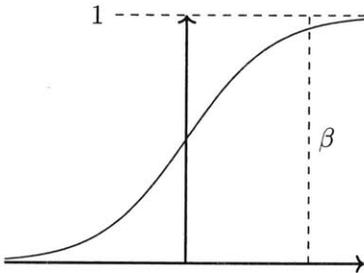


Figure 8.2:
The sigmoid function $\sigma(\beta x)$.

⁶⁸ In fact, the sigmoid function is a special case of the logistic function, which is the statistical distribution for a classification variable.

As \hat{w} is a unit vector, this factor can be absorbed into \hat{w} by promoting it to a non-unit vector \vec{w} :

$$f_{\vec{w},d}(\vec{d}) = \sigma(\vec{d} \cdot \vec{w} + b). \tag{8.8}$$

8.1.1 Two layers

We now have a working classifier based on a single decision boundary; although, we should expect for it to be rare that a data set can be well separated by one boundary. For example, the data set shown in figure 8.3 requires two boundaries working together. The parameters of the first boundary are \vec{w}_1 and b_1 in our notation, and the parameters of the secondary boundary are \vec{w}_2 and b_2 correspondingly.

The distance from each line can be organised as a vector, with \vec{w}_1 and \vec{w}_2 forming a matrix:

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &= \begin{bmatrix} \text{---}\vec{w}_1\text{---} \\ \text{---}\vec{w}_2\text{---} \end{bmatrix} \vec{d} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ \vec{y} &= \mathbf{W}_0 \vec{d} + \vec{b}_0. \end{aligned}$$

These distances can then be fed into the same sigmoid function as in equation 8.8, where it is understood that the sigmoid is applied element-wise to the input vector:

$$\vec{f}_{\mathbf{W}_0, \vec{b}_0}(\vec{d}) = \begin{bmatrix} \sigma(y_1) \\ \sigma(y_2) \end{bmatrix} = \sigma(\mathbf{W}_0 \vec{d} + \vec{b}_0). \tag{8.9}$$

These two outputs state the confidence on which side of each line the example is on. The function $\vec{f}_{\mathbf{W}_0, \vec{b}_0}$ is called the first layer. Interpreting these outputs to decide the class of the example requires a second layer, which takes the first layer as an input. In this case, the second layer can take the form of a single decision boundary, as shown in figure 8.4. To keep notation consistent, the parameters for this boundary will be written as a matrix \mathbf{W}_1 and vector \vec{b}_1 :

$$f_{\mathbf{W}_1, \vec{b}_1}(\vec{d}) = \sigma(\mathbf{W}_1 \vec{f}_{\mathbf{W}_0, \vec{b}_0}(\vec{d}) + \vec{b}_1). \tag{8.10}$$

Now, the loss function acts on the output of the second layer, which – due to the choice in using a single boundary – is a scalar:

$$L(\mathbf{W}_0, \mathbf{W}_1, \vec{b}_0, \vec{b}_1) = \sum_{\vec{d}, c \in \mathcal{S}} (c - f_{\mathbf{W}_1, \vec{b}_1}(\vec{d}))^2. \tag{8.11}$$

This can be minimised, as in the previous single layer example.

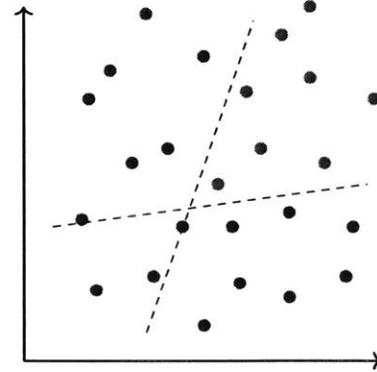


Figure 8.3: This set of data points cannot be divided by a single line, but two lines, working in tandem, can.

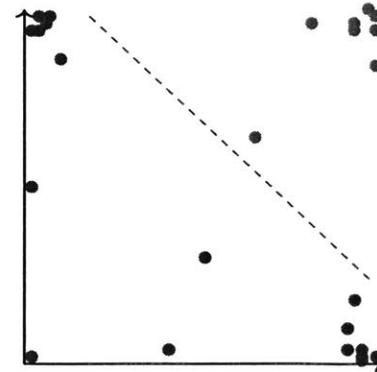
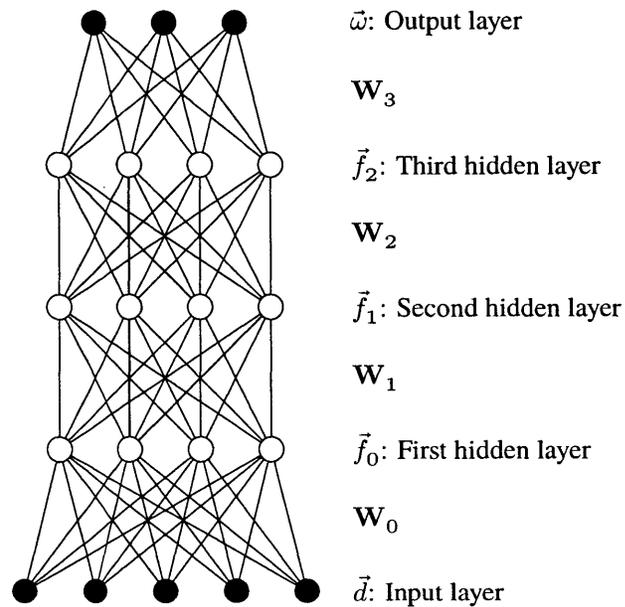


Figure 8.4: After the first layer of the classifier, the outputs are now separable by a single line.

8.1.2 Multi-layer neural networks

The astute reader might recognise the Heaviside theta and sigmoid function as looking similar to the “all-or-nothing” firing mode of a neuron. This non-linear transformation is called the “activation function” in analogy to how biological neurons are activated after some input threshold. If we assume that the outputs of these activation functions represent the outputs of neurons, then the weight vectors \vec{w} represent the strength of connections between neurons, completing the biological analogy. The b parameters then act as bias terms for each neuron.

Figure 8.5:
A multi-layer neural network.



A multi-layer neural network can be represented by a graph shown in figure 8.5. The nodes in this graph come in three forms: neurons, input nodes, and output nodes. The input and output nodes do not have the activation function applied to them⁶⁹. The connections between these nodes represent the weights in the weight matrices \mathbf{W} .

The input nodes form the input layer, and their values are given by the data vector \vec{d} . The neurons are organised into their own layers, with values given by the vector valued functions \vec{f}_i , where

$$\vec{f}_i = \vec{f}_{\mathbf{W}_i, \vec{b}_i} \tag{8.12}$$

in the notation of section 8.1.1. Each neuronal layer is called a *hidden layer*, as these values are not exposed. Finally, the value of the

⁶⁹ While the network described in section 8.1.1 did not have an explicit output node, it can still be placed in this framework by creating a single output node that connects to $\vec{f}_{\mathbf{W}_1, \vec{b}_1}$ with weight 1.

output layer is given by multiplying a weight matrix by the last hidden layer \vec{f}_n :

$$\vec{\omega} = \mathbf{W}_3 \vec{f}_n. \quad (8.13)$$

When a neural network has more than one hidden layer, it is deemed to be a *deep neural network*. Greater numbers of hidden layers give the neural network more representational power. This is usually quantified in terms of the number of neurons required to approximate a function to a certain degree of accuracy. While the universal approximation theorem [82, 145] shows⁷⁰ that a neural network with only a single hidden layer can approximate any function, the number of neurons required grows linearly with the degree of accuracy desired [208]. In contrast, when multiple hidden layers are allowed, the number of neurons only grows logarithmically with the desired accuracy⁷¹.

Multiple output nodes are used when more than one class should be discriminated by the network. For example; in physics, there may be two different kinds of signals along with background; in computer vision, there may be pictures of cats, dogs, and cars. In this case, the neural network has one output node for each class. The confidence of the input example belonging to class k is typically found by interpreting the output vector $\vec{\omega}$ through the soft-max function⁷²:

$$p_k = \frac{e^{\omega_k}}{\sum_j e^{\omega_j}}. \quad (8.14)$$

The class which the neural network guesses for the input example is the class with the highest confidence:

$$k_{\text{guess}} = \operatorname{argmax}_k (p_k). \quad (8.15)$$

The *accuracy* of the neural network is the percentage of examples that the network guesses correctly. This measurement is usually only used for human interpretation, as it is not differentiable and thus cannot be used as a loss function for optimisation.

While it is perfectly valid to use the least squares loss function described in section 8.1.1, a more motivated approach makes use of the interpretation of the class confidence p_k as a probability. This requires a definition of a “true” or target confidence \hat{p}_k . Ideally, the neural network would always guess the correct class, so this target confidence is defined as

$$\hat{p}_k = \begin{cases} 1 & k = \hat{k} \text{ is the correct class} \\ 0 & \text{otherwise} \end{cases}. \quad (8.16)$$

⁷⁰ With some mild assumptions on the activation function.

⁷¹ For multivariate functions, shallow networks grow exponentially with the number of variables, while deep networks only grow linearly.

⁷² The soft-max function is a generalisation of the logistic distribution to multiple classes.

It is common in training neural networks to use the cross entropy for the loss, and substituting the form of \hat{p} into equation 2.3 gives

$$L = H(\hat{p}; p) = -\ln p_{\hat{k}} = -\omega_{\hat{k}} + \ln \left(\sum_k e^{\omega_k} \right) \quad (8.17)$$

Thus minimising L will maximise the neural network output $\omega_{\hat{k}}$ associated with the correct class while minimising all outputs in the output vector $\vec{\omega}$.

The minimisation can be performed with a variety of methods, such as evolutionary, Monte-Carlo, or coordinate descent; however, gradient descent is most often the method of choice as it is the most efficient at finding a minimum for functions with large numbers of parameters. At a glance, for a function $L(\vec{P})$ where \vec{P} are the parameters to minimise over, the current set of parameters \vec{P}_i at iteration i in the algorithm are updated according to the gradient of L :

$$\vec{P}_{i+1} = \vec{P}_i - \eta \nabla L(\vec{P}_i) \quad (8.18)$$

where η – called the “learning rate” – controls the size of the steps in parameter space. Larger values of η cause faster convergence to the minimum, but if η is too large the algorithm can skip over or oscillate around the minimum, preventing convergence.

For gradient descent to be used, the gradient of L must be calculated. Two examples of ways to do this are numerical and symbolic differentiation. In numerical differentiation, the function L is evaluated not just at \vec{P}_i , but also in the neighborhood around \vec{P}_i . Taking the difference between these points gives a numerical estimate of the gradient. The advantage of this method is that it is easy to implement, and it works for all well behaved functions. The major disadvantage is that it requires n additional evaluations of L , where n is the number of parameters to minimise over. For deep neural networks, the number of parameters can be in the millions, making numerical differentiation impractical.

In symbolic differentiation, the derivative of L is found symbolically and implemented as a new function L' . Finding the gradient now only requires one evaluation of L' – which is more practical for large numbers of parameters – and L' can also be simplified before implementation, potentially saving some additional computational time. However, L' must be kept manually updated with any changes in implementation to L , creating a maintenance problem. Further, evaluation of L' can repeat expensive computations already performed in evaluating L . Take, for example, the function $g(x) = e^{Ax}$ so that $g'(x) = Ae^{Ax}$. Evaluation of g requires the relatively expensive computation of the exponential function with argument Ax , and evaluation of g' repeats this. Ideally, such repeated values could be re-used when evaluating the derivative.

8.2 Automatic differentiation

Automatic differentiation is a third method of calculating the derivative. It sits somewhere on the spectrum between numeric and symbolic differentiation, combining advantages of both. There are two distinct methods of doing automatic differentiation: forward mode, and reverse mode. Forward mode is better suited to functions with few parameters, such as vector valued functions. Reverse mode better handles functions with many parameters and scalar outputs, of which deep neural networks are a good example. Despite this, it is worth briefly describing forward mode as it is the more straightforward and intuitive of the two.

8.2.1 Forward mode

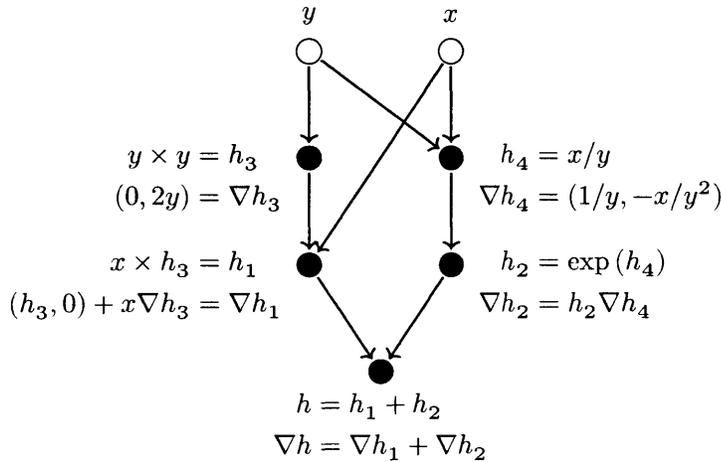


Figure 8.6:

Expression graph for $h(x, y) = xy^2 + e^{x/y}$. In forward mode automatic differentiation, derivatives are propagated forward from the input variable to the output variable.

Suppose we have a function $h(x, y) = xy^2 + e^{x/y}$. The evaluation of this function can be represented by a directed graph as shown in figure 8.6. Each black node of this graph represents a basic operation such as addition, multiplication, division, or an elementary function such as sin or exp. The input variables are denoted by white nodes. The edges of the graphs represent inputs to these basic operations, either in terms of the input variables to the function, or the results of other basic operations. In this case, the evaluation has been broken up into five basic operations labeled h_1 to h_4 along with the final result h .

In forward mode, the gradient is calculated by propagating derivatives down this graph in the same manner that the function is evaluated. This process is performed at the same time as the function evaluation, so that intermediate values can be reused – for

example the gradient of h_2 reuses the result of the exponential of h_4 . The gradients of previous nodes can be combined using the product rule.

At first, it may seem that this has only foisted the implementation and maintenance of the gradients into the implementation of h ; however, the actual propagation of derivatives can be encapsulated into library functions that provide the basic computational primitives. For example, a user can write `mul(x, exp(div(x/y)))` where `mul`, `exp`, and `div` are functions provided by an automatic differentiation library which transparently shuffle the derivatives around behind the scenes. Indeed, many programming languages allow for custom definitions of basic mathematical operators, which allow users to write functions that look like standard code they would write in absence of automatic differentiation.

Note that at every node in figure 8.6, a gradient over all input variables needs to be computed and stored. This implies that this method of calculating the gradient will require a multiplicative factor of $\mathcal{O}(n)$ additional computation and storage requirements over the original evaluation, where n is the input dimensionality. For deep neural networks, where the loss function is typically of the form $L : \mathbb{R}^n \rightarrow \mathbb{R}$, and n is very large, this cost is prohibitive. In contrast, forward mode is very efficient for functions of the form $\mathbb{R} \rightarrow \mathbb{R}^m$.

8.2.2 Reverse mode

Reverse mode attempts to solve this issue by reformulating the problem in a manner closer to the ideal case for forward mode. While forward mode propagates gradients, i.e. the vector

$$\nabla h_i = \left(\frac{\partial h_i}{\partial x}, \frac{\partial h_i}{\partial y}, \dots \right), \quad (8.19)$$

which are derivatives of the value of node h_i with respect to the input variables, reverse mode propagates *adjoints*, i.e. the scalar

$$\bar{h}_i = \frac{\partial h}{\partial h_i}, \quad (8.20)$$

which are derivatives of the output variable with respect to the value of the node. In comparison, only a single derivative⁷³ needs to be stored at each node in reverse mode, compared to the vector of derivatives in forward mode.

First, the function is evaluated in a forward pass without calculating derivatives. Once this is complete, the adjoints are propagated back up the computation graph to the input variables in a reverse

⁷³ In the case where multiple outputs are involved, such as a vector output, this adjoint also needs to be a vector of derivatives of each output variable w.r.t. the value of the node.

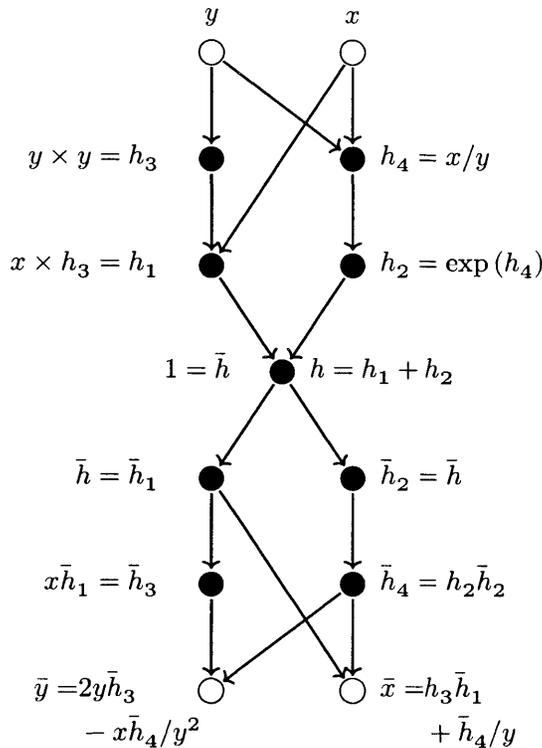


Figure 8.7:
 In reverse mode automatic differentiation, after evaluation of h , adjoints are propagated back up the evaluation tree to the input variables.

pass. This second pass is shown in figure 8.7 by a set of nodes mirrored around the output node h . Each of these mirrored nodes corresponds to a node in the original computation graph, and stores a single scalar value for the adjoint at that node.

While reverse mode is significantly more computationally efficient than forward mode for functions of the form $\mathbb{R}^n \rightarrow \mathbb{R}$, it does come with some drawbacks. The automatic construction of the adjoint propagation graph is a difficult prospect. Most programming languages do not have the capability to invert a computation in this way during compile-time, so the construction of the graph must be performed at run-time. As the nodes are evaluated in the forward pass, instructions are added to a data structure which describe the reverse pass. This means that compile-time time optimisations of this reverse pass cannot be made; although, in practice this opportunity cost is not too great, and the full forward and reverse pass is only a few times slower than a forward pass alone [46].

The second major disadvantage is that all the intermediate values (h_i) must be stored in the forward pass for access in the reverse pass. When the amount of data being processed is significant – as is often in the case in deep neural networks – the available computer memory can place a hard limit on the total complexity of the function.

This burden can be alleviated somewhat by a technique known as checkpointing [227], at the cost of additional processing time. Despite these limitations, reverse mode automatic differentiation is by far the most efficient method of calculating the derivative of a loss function for deep neural networks, due to the large number of parameters involved.

8.2.3 Backpropagation

The most widely used algorithm for training neural networks is called *backpropagation*, and is a specialised form of reverse mode automatic differentiation. The name derives from the **backward propagation of errors**, where **error** refers to the difference between the desired and actual output of the neural network – encapsulated in the loss function. While the automatic differentiation concepts behind backpropagation had already existed for many years in the literature, the name was coined and the method popularised by Rumelhart et al. in 1986 [213].

After the calculation of the error – or loss – the derivatives are propagated back up the computational graph in the reverse pass. The weights and biases are then updated according to the final adjoint of each weight and bias, and the learning rate according to the rule

$$w_{i,j} \rightarrow w_{i,j} - \eta \bar{w}_{i,j}, \quad (8.21)$$

$$b_i \rightarrow b_i - \eta \bar{b}_i, \quad (8.22)$$

where $w_{i,j}$ represent the elements of the weight matrices, and b_i the elements of the bias vectors.

8.3 Regularisation

If a neural network has too many free parameters, it can fail to generalise to the task it is being trained for. Instead, it will begin to memorise the examples it is being trained on. This is a phenomena called “over-training”, where the high degree of flexibility will lead the weights to capture the exact properties of the training examples, including any noise present. It is closely related to the concept of over-fitting in non-linear regression. In both cases, the model being optimised is no longer able to extrapolate to data it has never seen before.

Computing the exact number of free parameters needed is not practical, so *regularisation* is introduced in order to remove weak predictors – such as those patterns that only correspond to the noise in a single training example – from the neural network. Mathematically, this is usually introduced as either a l_1 or l_2 norm of the

weights, called “lasso“ and “ridge“ regression, respectively. This norm, scaled by the regression parameter λ is then incorporated into the loss:

$$L \rightarrow L + \lambda \sum_i \|w_i\|_p, \quad (8.23)$$

where w_i are the set of all weights in the model, and p is either 1 or 2 for each kind of regression.

This term adds a constant pressure for all the weights to be close to zero, and any pattern in the training data must be a strong enough predictor to overcome this pressure. The choice of norm usually depends on the degree of sparsity desired. An l_1 -norm allows for very large weights, as long as there is an even greater number of very small weights to keep the average of all weights close to zero. In contrast, the l_2 -norm tends to force the weights to be close to the same value, reducing the sparsity by disfavours models that have a few very large weights.



Search for tau neutrinos in IceCube

In a 2078 day sample – spanning six years – the IceCube experiment has detected 49 events with energy greater than 60 TeV [63]. In a (1 : 1 : 1) neutrino flavour ratio scenario, 13 of these would be expected to be tau neutrinos. Despite this, IceCube has yet to positively identify a single tau neutrino event, due to the experimental difficulties outlined in section 1.2.5.

IceCube has one advantage in identifying tau-flavour interactions over other neutrino experiments: the astrophysical neutrinos are hundreds of times higher in energy than accelerator beams. At these energies, the tau produced in a charged current interaction will undergo extreme relativistic time dilation, increasing its lifetime as seen by the detector. For example, a tau with an energy of 100 TeV will travel a mean distance of 4.9 meters over 16 nanoseconds before it decays. As IceCube's position and time resolution are on the order of tens of meters and nanoseconds, this provides an opportunity to observe the tau.

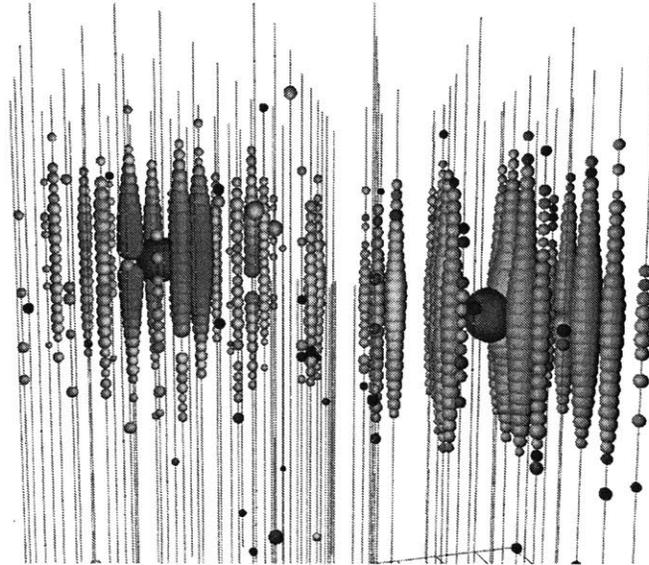
9.1 Double pulse signature

The initial tau neutrino interaction is very highly favoured to be a deep inelastic scattering. A neutral current scattering will be indistinguishable from those caused by muon or electron neutrinos, and forms a background for the search. A charged current interaction produces the tau lepton required for identification of the tau neutrino, as well as a hadronic shower from the recoiling quark. IceCube is not sensitive enough to distinguish the subtle differences between

hadronic and electromagnetic showers, and so this hadronic shower registers as a cascade inside the detector.

Figure 9.1:

Two cascades form in IceCube from a simulated interaction of a tau neutrino and subsequent decay of the tau. Figure included from Collaboration et al. [63].



In the most common decay mode, the tau produces high energy hadrons. The next most common decay mode has the tau decaying into an electron. The hadrons will initiate a hadronic shower, and the electron an electromagnetic shower, creating a second cascade as shown in figure 9.1 for a simulated event with an energy of ~ 10 PeV. In both cases, a substantial portion of the tau's energy will go into the tau neutrino produced in the decay. The last major decay mode of the tau lepton is the production of a muon. Inside IceCube, this mode will appear nearly identical to a charged current muon neutrino interaction.

A somewhat more realistic case, of a neutrino with energy 385 TeV, is shown on the left in figure 9.2. The upper gray sphere indicates where this simulated down-going tau neutrino interacted with the ice, depositing 111 TeV of energy. The tau lepton then travels 12 meters, before decaying at the location of the second gray sphere, depositing 273 TeV of energy. At these deposited energy scales, large amounts of light are generated in both showers, creating cascades which are highly extended compared to the tau decay length. For this reason, discriminating these kinds of double cascades is not possible with position information alone.

The right of figure 9.2 shows the timing information of this simulated event for two nearby DOMs. A double peaked structure is clearly visible, and this kind of signature is called a “double pulse”.

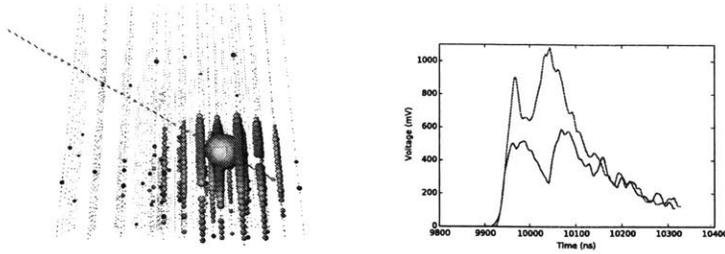


Figure 9.2:

A simulated tau-neutrino interaction in IceCube. **Left:** Two cascades that are not well separated enough in space to be seen as distinct. **Right:** Waveform level data from the event on the left for two nearby DOMs on the same string. The two cascades are well separated enough in time to be resolved. Figures included from Collaboration et al. [63].

Multi-peaked structures are not unique to tau decay events. For muon energies beyond ~ 300 GeV, radiative losses become the dominant method of energy loss [198]. As a stochastic processes [39], they produce isolated hadronic or electromagnetic showers along the muon trajectory, forming a background for tau induced double cascades. Astrophysical charged current muon neutrino interactions and cosmogenic muons penetrating through the ice above the detector are both possible sources of such high energy muons. While conventional atmospheric muon neutrinos are highly suppressed at these energies, prompt neutrinos may be an additional source.

Detector effects such as pre-pulsing and late-pulsing may also create multi-peaked structures. As discussed in section 4.2.1, these typically have a characteristic time associated with them, providing a potential path to eliminating their influence. After-pulsing occurs over a time scale of hundreds of nanoseconds, well beyond the scales being considered for tau decays.

A previous analysis has developed a derivative based method for identifying peak structures in a DOM waveform [4]. While this approach works well for smooth waveforms, derivatives enhance noise in data, and performance can suffer for borderline cases. The analysis expected 0.54 tau neutrino events with 0.35 background events in three years of data, but identified none. The astrophysical tau neutrino spectrum may be steeper than expected, suggesting that future searches should concentrate on methods that handle the lower-energy borderline cases well.

Pattern recognition may be a fruitful approach for dealing with noisy data. For example, edge finding is similar to performing derivatives, but averages over a larger data range. Often times a well versed practitioner can make a positive identification by simply looking at the data, when more traditional algorithms struggle. Machine learning provides a method for the automatic selection of patterns to look for, based on simulated training data.

Cross correlation between nearby DOMs is another potential avenue. Many otherwise below threshold waveforms can be considered in unison to make a positive identification. As there are over 5000 DOMs in IceCube, and each DOM records over a hundred data points

per waveform, this transforms the problem into one with hundreds of thousands of dimensions.

9.2 Convolutional neural networks

Convolutional neural networks have shown exceptional performance on very high dimensional pattern recognition problems with spatial correlations. These neural networks were inspired by the human visual cortex, but like the *fully connected* neural networks discussed in chapter 8, can be argued for from more fundamental principles.

Consider a digital photograph with dimensions of 256×256 pixels. Classifying the contents of this image is a $\sim 65,000$ dimensional problem. Assuming that the dimensionality of the first hidden layer is 16 times smaller than the input, the first weight matrix in a fully connected neural network has $\sim 260,000,000$ elements — considerably more than is required to solve the problem, which will lead to over-training.

A key insight on how to rectify this problem comes from the spatial locality present in the images. Consider a pixel that is located on a leaf in the image, such that it is green. The neighbouring pixels are very likely to also be green, and are thus highly correlated to the original pixel. In contrast, a pixel located very far away could be any one of a number of colours, and so there is little to no correlation over large distance scales. For this reason, patterns that encompass large numbers of pixels, and large distance scales, are unlikely to be useful.

Restricting the size of the patterns that the neural network is able to learn can be accomplished by constraining the weight matrix. In this formulation, each neuron is associated with a neighbourhood of pixels, and the output of this neuron is based only on this neighbourhood. This is equivalent to forcing the non-neighbour elements in the weight matrix to zero, drastically reducing the number of free parameters. For example, if the neighbourhood size is 5×5 pixels, then the number of free parameters for the same 256×256 image is now $256^2 \times 5^2 \sim 1,600,000$.

While this construction generalises to any number of dimensions, the mathematics is more straight-forward when considering only one dimension. The application of the weight matrix \mathbf{W} in a fully connected neural network can be written as:

$$\vec{y} = \mathbf{W}\vec{d} + \vec{b} \quad y[i] = \sum_j W[i, j]d[j] + b[i], \quad (9.1)$$

where the row and column coordinates of the vectors and matrices are written in square brackets for convenience. For this example, we

consider a neighbourhood of size 3. The restricted zero elements of \mathbf{W} can be removed from the equation:

$$y_i = \sum_{j=-1,0,1} W[i, i - j]d[i - j] + b_i. \quad (9.2)$$

If we define $K[i, j] = W[i, i - j]$, then this can be written as

$$y_i = \sum_{j=-1,0,1} K[i, j]d[i - j] + b_i. \quad (9.3)$$

An example of a fully connected neural network layer is shown in figure 9.3 for an input data vector \vec{d} . The restriction of this fully connected layer to a local neighbourhood is shown in figure 9.4, where the weights that have been forced to zero have their corresponding edges removed from the figure.

The second key insight is that important patterns – often called “features” – can be located anywhere in the image. In this sense, the problem has a translational invariance: the weights associated with neurons on one side of the image should be equal to their counterparts for neurons on the opposite side. This equality constraint – also known as *tying* the weights together – reduces the dimensionality of the image example to just 5^2 , removing the dependence on the actual size of the image entirely.

These two constraints reduce the weight matrix such that it is equivalent to performing a convolution on the image with a small kernel size⁷⁴. Hence, this kind of neural network is called a “Convolutional Neural Network” (CNN).

Mathematically, the translational invariance removes any dependence on i in the row coordinate of the weight matrix: $K[i, j] \rightarrow K[j]$, so that

$$y[i] = \sum_{j=-1,0,1} K[j]d[i - j] + b[i]. \quad (9.4)$$

Written this way, it is clear that a convolution operation with kernel K is being applied to d . Figure 9.5 shows the act of tying the weights of figure 9.4 by colouring edges that represent shared weights with the same colour.

As an aside, it should be noted that the signal processing operation for finding patterns in data is called *cross-correlation*. The convolution is related to the cross-correlation by a parity transform on the kernel $K[x] \rightarrow K[-x]$. Although most convolutional neural networks actually implement cross-correlations, the moniker “convolutional neural network” has stuck. In any case, the remainder of

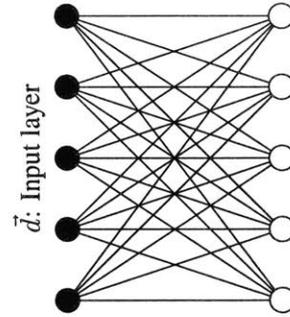


Figure 9.3: An example of a fully connected neural network, with input data vector represented by the nodes on the left, and weights by the edges between nodes.

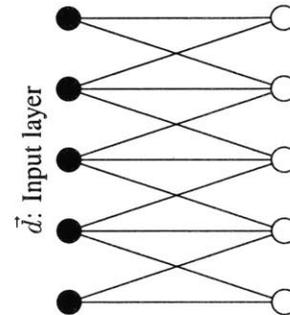


Figure 9.4: An example of a neural network similar to figure 9.3; but the weights have been restricted to be non-zero in a local neighbourhood of size 3.

⁷⁴ Specifically, the translational invariance is characteristic to convolutions. The restriction on the kernel size makes the previous illustration clear, and is used in all CNNs.

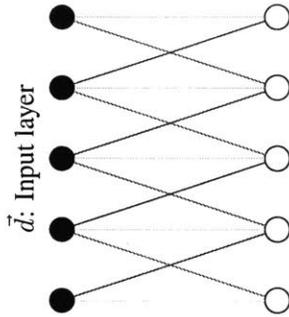


Figure 9.5:

A convolutional neural network layer. This is similar to figure 9.4; but the weights have now been tied together. This is represented by the coloured edges, where all edges with the same colour share the same weight.

this thesis will refer to these operations as convolutions in deference to the standard nomenclature.

A CNN with only a single hidden layer can only capture to most basic features of the data, such as edges and textures. Like the fully connected neural networks in chapter 8, additional hidden layers can be added, each capturing a higher level representation of the data. As the first hidden layer finds the basic patterns in the data, the output values of the hidden neurons can be interpreted as another image. The next hidden layer finds the patterns in this new image – such as basic shapes or changes in textures – effectively finding the patterns in the patterns. For photograph classification, approximately ten or more hidden layers are usually required for good performance, so CNNs are usually also deep neural networks.

So far the discussion has centered around a single kernel per layer. For a layer of a CNN to be able to learn, and identify, multiple patterns, it must be parameterised by multiple kernels. Each kernel is associated with its own set of neuron outputs, called a “channel”:

$$y_k[i] = \sum_{j=-1,0,1} K_k[j]d[i-j] + b_k[i], \quad (9.5)$$

where k is the channel number.

The next hidden layer needs to operate on this set of channels. The convolution operation can be generalised to data with channels by increasing the dimensionality of K with an input channel coordinate l :

$$y'_k[i] = \sum_{j=-1,0,1} \sum_l K_{k,l}[j]y_l[i-j] + b_k[i]. \quad (9.6)$$

In essence, K_k is now itself a set of kernels, each one of which is convolved with the input channels. Thus, while CNNs perform convolutions over spatial coordinates (denoted here in square brackets), they are fully connected over channel coordinates (denoted in subscripts).

The dimensions of the output can be reduced by using *strided convolutions*. A convolution with stride m will skip every m^{th} output, reducing the size of the output by a factor of m .

9.2.1 Other common operations

The sigmoid activation function described in equation 8.6 has a few disadvantages when working with deep, high dimensional neural networks. It requires computing an exponential and an inverse, which are two of the most computationally expensive operations for

processors. The gradient also vanishes as the argument to the function moves away from the origin, reducing the speed of convergence for the neural network.

Hahnloser et al. proposed a *Rectified Linear Unit* (ReLU) activation function on biological and mathematical grounds [123]. This function, shown in figure 9.6 rectifies its input in the positive domain:

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} x & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9.7)$$

While it may seem odd to replace a function with a vanishing gradient in the limit of large $|x|$ with one that explicitly has zero gradient in the negative domain, it is the linear nature of this activation function that aids in training [115].

Pooling is also often used to reduce the dimensionality of the data. In pooling, a single simple operation – such as taking the maximum value or the average value – is applied to a neighbourhood defined by the kernel size, and – much like strided convolutions – will skip over elements in the output by using a stride.

9.2.2 Training high-dimensional data

The loss L as defined in equation 8.17 is for a single example with true classification \hat{p} . With a corpus of training data $\mathcal{T} = \{(\vec{d}, \hat{p})\}$, the loss is computed for each example and combined by summing:

$$L(\vec{P}, \mathcal{T}) = \sum_{\vec{d}, \hat{p} \in \mathcal{T}} L(\vec{P}, \vec{d}, \hat{p}) \quad (9.8)$$

where \vec{P} is the vector of parameters for the neural network.

As the memory footprint for gradient descent on the loss for single example roughly scales with the input dimensionality times the depth of the network, the required memory for performing gradient descent on the combined loss for a typical training set with order 10^5 examples can be in the hundreds of gigabytes or more. This memory requirement is far too great for most modern computer systems, so the task must be subdivided.

Suppose that only a subset $\mathcal{S} \subset \mathcal{T}$ of the training examples are chosen randomly. If this subset is chosen to be sufficiently small, the gradient vector of this subset $\nabla L(\mathcal{S})$ is practical to compute. While this vector will not point in the same direction as the true gradient $\nabla L(\mathcal{T})$, it will be approximately correct. An update to the set of parameters for the neural network can then be performed using this approximate gradient:

$$\vec{P}_{i+1} = \vec{P}_i - \eta \nabla L(\vec{P}_i, \mathcal{S}) \quad (9.9)$$

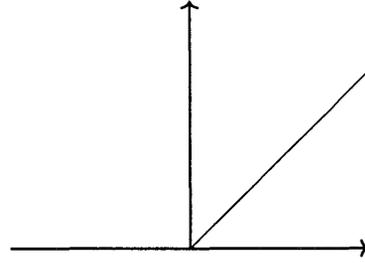


Figure 9.6:
The Rectified Linear Unit $\text{ReLU}(x)$.

After this update, a new subset is chosen from \mathcal{J} , and the process is repeated. Over many iterations, the effect of the approximate gradients should average out, leading the optimiser to the minimum. Due to the random nature of the gradient, this is known as “stochastic gradient decent” and the subsets are called “batches”.

Although this process will converge to a minimum, each update will have a high degree of variance that will reduce the efficiency of the overall minimisation procedure. This can be avoided by updating the parameters with a gradient vector \vec{v} derived from an exponential average of all previous gradient vectors [212]:

$$\vec{v}_i = \mu \vec{v}_{i-1} + \nabla L(\vec{P}_i, \mathcal{S}) \quad (9.10)$$

$$\vec{P}_{i+1} = \vec{P}_i - \eta \vec{v}_i \quad (9.11)$$

where μ is a tuneable parameter called the “momentum”.

Typically, a choice of the learning rate η which is too large will lead to over-corrections and oscillations in the path of the minimiser. On the other hand, if η is too low, performance will suffer as the minimiser takes unnecessarily small steps. There are various method for adjusting η during training, such as Adadelata, Adam, and RMSProp [212]. These schemes all adjust the learning rate on a per-parameter basis:

$$\vec{P}_{i+1} = \vec{P}_i - \boldsymbol{\eta} \vec{v}_i \quad (9.12)$$

where $\boldsymbol{\eta}$ is a diagonal matrix of learning rates. The base learning rate η is adjusted based on a function of the gradient:

$$\boldsymbol{\eta}_{i,i} = \frac{\eta}{f(\vec{v}_i)} \quad (9.13)$$

This function is often chosen to be a square root of the squared gradient, so that when the slope of the loss function is steep, the minimiser takes small, conservative steps; and, when the slope is shallow, the minimiser takes larger steps.

9.2.3 Application to the IceCube tau search

Many neutrino detectors are pixelated or segmented such that the data they produce has very similar properties to images — namely the locality and translational invariance. It stands to reason that CNNs — which sport super-human performance in photographic image classification — will also be effective in analysing data from neutrino experiments.

IceCube is effectively segmented in volume by the grid of DOMs. The timing information can be thought of an extra spatial dimension,

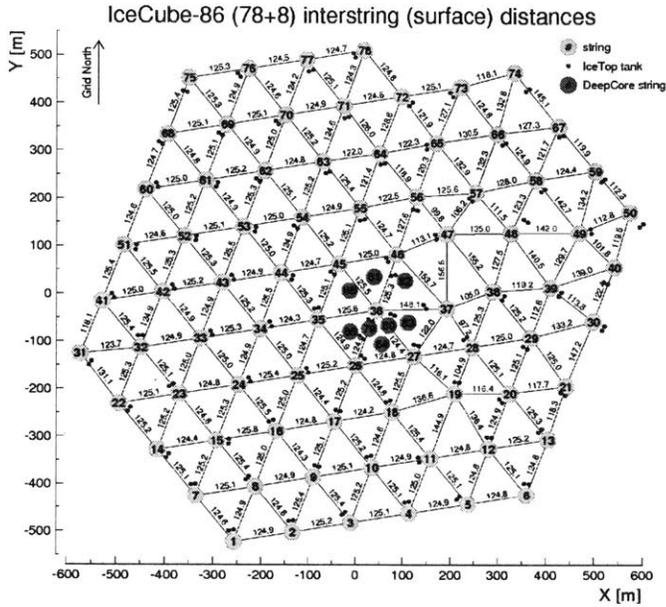


Figure 9.7:
The horizontal geometry of IceCube. Figure courtesy of the IceCube collaboration.

creating a 4D image. Locality and translational invariance are both properties of the physical interactions being observed.

IceCube also brings some unique challenges. Primarily, the horizontal layout of the detector is not completely regular. It closely follows a triangular grid, but a few strings – such as 37 and 47 – are offset from their regular position as shown in figure 9.7. DeepCore, which are strings 79 to 86, are not part of the grid entirely. As efficient convolutions must be performed on a regular grid, this requires some special handling.

While the detector is largely laterally homogenous, the amount of light absorbed and scattered by the ice varies with depth, as discussed in chapter 4. With the exclusion of the dust layer, figure 4.4 shows that this effect is less than 50% variation over the depth of the detector, and so it can be safely ignored for the present exploratory study.

9.3 Training data

The training set consists of two types of interactions:

- Charged current tau neutrino interactions: These include all decay modes of the tau lepton, in proportions according to their relative decay probabilities.
- Single cascade events: Neutral current interactions form the main background to discriminate against. Neutral current tau

neutrino interactions are used to stand in for muon and electron neutrino neutral current interactions, as they produce the same signature in the detector. Charged current electron neutrino interactions also form a smaller background, and the same tau neutrino neutral current interactions stand in for these as they both produce single cascades inside the detector.

Future studies may include charged current electron neutrino interactions directly, as there can be differences in the cascades at the Glashow resonance.

These interactions are simulated inside the detector, as described in chapter 4, producing waveforms for each DOM. These waveforms contain 128 discrete measurements of the quantity of light detected by the DOM in 3.3 nanosecond intervals. The entire event is represented as a three dimensional array of $86 \times 64 \times 128$ elements over the strings in IceCube, DOMs on each string, and time ticks in each waveform. While IceCube nominally has 60 DOMs on each string, the array is padded with zeros to 64 to make it a power of two.

The simulation includes digitiser effects, producing up to three waveforms with different simulated gains. Of these, the highest gain waveform that is not saturated is selected.

The neural network is trained without the associated weights of the simulated events. This choice was made for two reasons. Firstly, the weights have a very large range, spanning many orders of magnitude. If the loss function were adjusted according to these weights, one training example would often dominate the batch so thoroughly that the remaining examples would effectively not exist. The second reason is that weighting the training data would make the neural network pay more attention to events that are more common in the real data set. This is not necessarily useful, as the power law spectrum of the neutrino flux leads to the most common event being low energy, which is also the hardest kind of event to discriminate from the background. Training performance may then suffer, as the neural network attempts to discriminate data that is effectively indistinguishable.

9.4 Neural network design

Convolutional neural network design is constrained by the boundary conditions imposed by the problem at hand. The input for this neural network must be an array with nearly a million elements representing a single event. On the other hand, the output must be two numbers: the classification score for the event being coming from a tau neutrino interaction, and the classification score for the event coming from a background process. The problem is factorised into two stages. The first stage treats each string individually, forming a 2D image

by stacking the waveforms from all DOMs on the string next to each other.

These images are of size 64×128 , and go through an initial input convolution with asymmetric stride of 1×2 to reduce the size of the image to a square 64×64 . Successive applications of convolution + neuronal activation layers – as shown in table 9.1 – reduce the image further to a size of 2×2 . Finally, a pooling layer is applied that averages over these four pixels to produce an output with size 1×1 .

Type	Kernel	Stride	Channels	Image size
input	—	—	1	64×128
conv.	3×5	1×2	8	64×64
conv. + relu	3×3	2×2	8	32×32
conv. + relu	3×3	1×1	8	32×32
conv. + relu	3×3	2×2	8	16×16
conv. + relu	3×3	1×1	8	16×16
conv. + relu	3×3	2×2	8	8×8
conv. + relu	3×3	1×1	8	8×8
conv. + relu	3×3	2×2	8	4×4
conv. + relu	3×3	1×1	8	4×4
conv. + relu	3×3	2×2	8	2×2
ave. pool	2×2	2×2	8	1×1

Table 9.1:

The design for the first stage of the convolutional neural network. Layers are ordered from top to bottom, with the output of each layer feeding into the input of the next.

The original $86 \times 64 \times 128$ array has been reduced to $86 \times 1 \times 1$ with an additional 8 channels to make $86 \times 1 \times 1 \times 8$. This is exactly the same size as a 86×8 array, and can be considered as such.

At this point, the data is arranged in order of increasing string number, and so spatial locality is not present in the first dimension of this array. Before starting the second stage, the array needs to be transformed to a coordinate system where this locality is present.

This can be achieved by defining a continuous convolution function:

$$t_k(x', y') = \sum_{i,l} \kappa_{k,l}(x' - x[i], y' - y[i]) f_l[i], \quad (9.14)$$

where $x[i]$ and $y[i]$ are the lateral coordinates of each string, and $f_l[i]$ is the output of the first stage for string i and channel l . The convolution kernel is built from a basis of Bessel functions:

$$\kappa_{k,l}(x, y) = \sum_{n,m} C_{k,l} J_n \left(\alpha_{n,m} \frac{r(x, y)}{r_0} \right) [A_{n,m} \sin(n\theta(x, y)) + B_{n,m} \cos(n\theta(x, y))], \quad (9.15)$$

where $\alpha_{n,m}$ is the m^{th} zero of the n^{th} Bessel function J_n , $r(x, y) = \sqrt{x^2 + y^2}$ and $\theta(x, y) = \arctan \frac{y}{x}$.

The function $t(x, y)$ is evaluated at discrete points on a 16×16 , square kilometer sized grid that spans the horizontal layout of the detector. As this grid and the IceCube geometry are not learnable parameters, this kernel can be represented as a tensor:

$$\kappa_{l,i}^{k,\sigma,\tau} = \beta_{l,i,n,m}^{k,\sigma,\tau} A^{n,m} + \gamma_{l,i,n,m}^{k,\sigma,\tau} B^{n,m}, \quad (9.16)$$

with

$$\beta_{l,i,n,m}^{k,\sigma,\tau} = J_n \left(\alpha_{n,m} \frac{r(x'_\sigma - x[i], y'_\tau - y[i])}{r_0} \right) \sin(n\theta(x'_\sigma - x[i], y'_\tau - y[i])), \quad (9.17)$$

$$\gamma_{l,i,n,m}^{k,\sigma,\tau} = J_n \left(\alpha_{n,m} \frac{r(x'_\sigma - x[i], y'_\tau - y[i])}{r_0} \right) \cos(n\theta(x'_\sigma - x[i], y'_\tau - y[i])), \quad (9.18)$$

such that

$$T^{k,\sigma,\tau} = \kappa_{l,i}^{k,\sigma,\tau} F^{l,i}, \quad (9.19)$$

where T is now a 2D image, with indices σ and τ , and channel number k ; and $F^{i,l}$ is the output array of the first stage over string number i with channel number l .

Now that the data is again in the form of a 2D image with channels, the second stage of the neural network acts to find the lateral patterns and reduce the data down to the require two classification scores as shown in table 9.2. This proceeds much like the first stage, with a series of 2×2 strided convolutions interspersed with un-strided convolutions. Finally, the output of the last convolution and neural activation layer is fed into a single fully connected layer with two outputs — referred to here as channels.

Table 9.2:

The design for the second stage of the convolutional neural network. Layers are ordered from top to bottom, with the output of each layer feeding into the input of the next.

Type	Kernel	Stride	Channels	Image size
Bessel conv.	—	—	8	16×16
conv. + relu	3×3	2×2	8	8×8
conv. + relu	3×3	1×1	8	8×8
conv. + relu	3×3	2×2	8	4×4
conv. + relu	3×3	1×1	8	4×4
conv. + relu	3×3	2×2	8	2×2
fully connected	—	—	2	—

Initially, a neural network architecture called “residual neural networks” [132] was implemented. This kind of architecture is common in convolutional neural networks designed for photographic classification. However, a switch to the standard architecture shown about with a *looks linear* initialisation scheme [38] improved the convergence time by an order of magnitude.

In this scheme, a standard neural activation followed by a generalised tensor product,

$$\vec{y} = \mathbf{W}\sigma(\vec{x}), \quad (9.20)$$

is replaced by

$$\vec{y} = \mathbf{W}\sigma(\vec{x}) - \mathbf{W}'\sigma(-\vec{x}), \quad (9.21)$$

where \mathbf{W} and \mathbf{W}' are two separate weight tensors, which are initialised to the same value, but allowed to diverge during training. Thus, when the network is first initialised, the operation becomes linear when ReLU activation functions ($\sigma(x) = \text{ReLU}(x)$) are used:

$$\vec{y} = \mathbf{W} [\text{ReLU}(\vec{x}) - \text{ReLU}(-\vec{x})] \quad (9.22)$$

$$= \mathbf{W}\vec{x}. \quad (9.23)$$

A equivalent statement of the same concept, is to symmetrise and anti-symmetrise these weight matrices:

$$2\mathbf{M} = \mathbf{W} + \mathbf{W}', \quad 2\mathbf{M}' = \mathbf{W} - \mathbf{W}'. \quad (9.24)$$

Again, assuming a ReLU activation function, equation 9.21 becomes:

$$\vec{y} = (\mathbf{M} + \mathbf{M}')\text{ReLU}(\vec{x}) - (\mathbf{M} - \mathbf{M}')\text{ReLU}(-\vec{x}) \quad (9.25)$$

$$= \mathbf{M} [\text{ReLU}(\vec{x}) - \text{ReLU}(-\vec{x})] + \mathbf{M}' [\text{ReLU}(\vec{x}) + \text{ReLU}(-\vec{x})] \quad (9.26)$$

$$= \mathbf{M}\vec{x} + \mathbf{M}'\text{abs}(\vec{x}), \quad (9.27)$$

where $\text{abs}(\vec{x})$ is the element-wise absolute value. The initialisation constraint is now simply that $\mathbf{M}' = 0$, while \mathbf{M} is initialised as normal.

An initially linear neural network is desired, as it leads to a well behaved calculation of the gradient. Non-linearities can cause the adjoints – as they are propagating back up through the computation graph – to vanish or explode. In fact, it is possible to train arbitrarily deep neural networks with back-propagation when the network is both linear, and the weights are initialised to be orthogonal matrices⁷⁵ [215].

9.5 Results

Along with the neural network design discussed in the previous section, the hyper-parameters used are listed in table 9.3. The neural network was implemented using the TensorFlow deep learning library

⁷⁵ Orthogonality can be generalised to convolutions through their matrix representation

Table 9.3:
The hyper-parameters used during training.

Type	Value
Batch size	16
Regularisation (λ)	0.001
Regularisation type	l_2 -norm
Base learning rate (η)	0.005
Adaptive learning scheme	Adam [158]

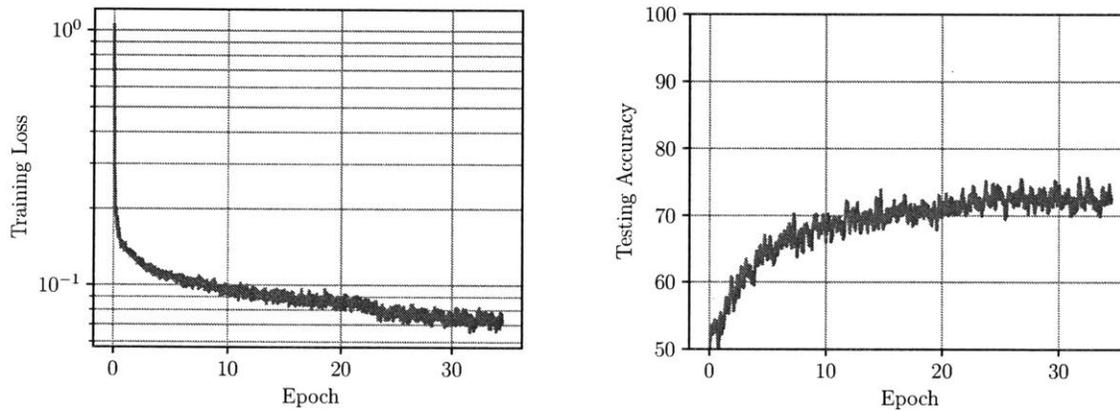


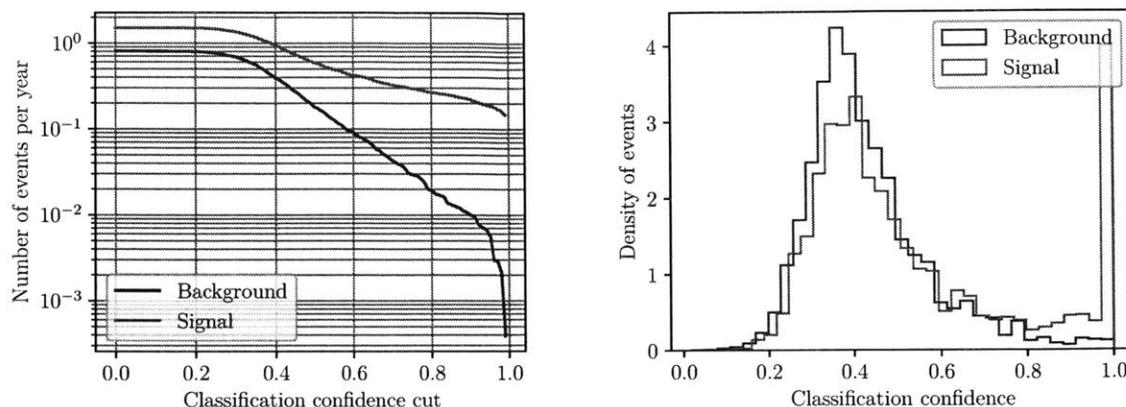
Figure 9.8:

Left: The loss computed on the training set as a function of the current epoch. **Right:** The accuracy computed on the testing set as a function of the current epoch. Both figures are a moving average over 1280 examples.

[5], and the network was trained on an NVIDIA Titan X GPU. It was trained on 28,747 simulated charged current tau neutrino interactions and 35,695 simulated neutral current tau neutrino interactions as background. Rather than setting a specific maximum number of training iterations, the neural network was left to train indefinitely, then manually stopped once the average loss had shown to have converged.

The loss curve is shown on the left in figure 9.8 as a function of the current *epoch*: the current fraction of total training sets which the neural network has been trained on. To validate the results, 20% of the examples in the training data were isolated into a testing set. The neural network was trained on the remaining 80%, then the accuracy of the neural network was computed on the testing set. As the neural network was not trained on any of the examples in the testing set, this accuracy gives an objective gauge on the performance of the neural network, without the possibility of over-training on specific examples. The test accuracy curve is shown on the right in figure 9.8 as a function of the current epoch.

Once training is completed, the trained neural network is used to classify all of the events in the testing set. These events are then re-



weighted according to their simulation weights, and an astrophysical energy spectrum of [60]:

$$\frac{d\Phi}{dE} = \left[9 \times 10^{-9} \frac{\text{events}}{\text{GeV s sr cm}^2} \right] \left(\frac{E}{100 \text{ TeV}} \right)^{-2.13}. \quad (9.28)$$

The neutral current tau neutrino event weights are additionally multiplied by a factor of five to represent the total flux from all three flavours of neutral current interactions as well as charged current electron neutrino interactions.

The cumulative event rate as a function of a classification confidence cut is shown on the left in figure 9.9. The cut is placed so that only events that score higher than the cut are counted in the event rate. The distribution of classification confidences is shown on the right in figure 9.9.

9.6 Summary and outlook

The neural network demonstrates the ability to discriminate tau neutrino events from background events. This is most clearly seen on the right of figure 9.9; a large proportion of signal events have a classification confidence of very close to 1. A confidence cut of 0.9 gives a similar tau neutrino event rate as the previously used derivative based method, but with half the single cascade background rate [4].

There are a number of potential avenues for continuing this work. In particular, the Bessel function based convolution operation is essentially interpolating grid locations between DOMs, which may be smoothing out important features. Although IceCube is an irregular triangular grid, it is possible to define a mapping between this geometry and a regular square grid slanted at 30° . While this mapping map not be exact in terms of preserving distances between DOMs on

Figure 9.9:

Left: The number of signal and background events per year that pass a cut placed on the classification confidence. **Right:** The distribution of events by their classification confidence, given their true classification class. Both figures are computed on the weighted testing sample.

different strings, it would no longer require the Bessel convolution. The eight DeepCore strings do not approximate any kind of regular geometry, and would need to be included in a separate sub-network.

Instead of factorising the network into two stages of 2D convolutions, the neural network could implement 4D convolutions over all spatial dimensions and time. Although the correlation between nearby DOMs on adjacent strings is weaker than between nearby DOMs on the same string, this may provide some benefit when an interaction happens in the space midway between strings.

As the tau lepton produced in a charged current interaction travels at near light speed at these energies, the spacing in time between light pulses can vary between DOMs. Those that sit toward the direction the tau neutrino was travelling will see a reduced spacing, while those behind the interaction will see an increased spacing. The removal of this variability through the application of an appropriate transformation may lead to a more homogenous, and thus easier to learn features. The parameters of this transformation, such as the interaction location and direction, could be estimated by another neural network, before being applied to the data for the main classification task.

While this particular exploratory study did not progress to application on measured data, there remains the question of the reliability of machine learning algorithms trained on simulation. These algorithms may learn features that are only present in simulations, due to imperfect modelling of the underlying physical processes. Application of these trained models could lead to erroneous results if this effect is not properly considered. Ideally, measured data would be used for the training dataset; however, obtaining a set with known classes of a size suitable for training is not practical⁷⁶. Much as the l_1 and l_2 norm regularisation were introduced to prevent the neural network from over-training on features specific to single training examples, measured data may be able to be incorporated as an unsupervised component into the training scheme to regularise and prevent the over-training of the neural network on features that are only present in simulated data.

This problem is not unique to particle or astrophysics. Large, well labelled datasets are hard to come by in medical imaging, and attention in this field is now being turned to training machine learning algorithms on simulated data. To bridge the gap between simulated and measured data, these medical imaging researches are developing adversarialy trained neural networks that transform measured data into real data [177].

⁷⁶ In many cases, if system were in place to obtain such a dataset, then the original problem the neural network was designed to address would have been solved by such a system.

Metropolis light transport

10

Large volume neutrino detectors rely on Cherenkov radiation to detect particles traversing their bulk. This radiation is produced when a charged particle travels through a material with a velocity that is greater than the phase velocity of light in that material. Unlike scintillation – which requires specific properties – Cherenkov radiation is generated in any material that has a significant refractive index.

As this light travels through the detector, it will interact with the bulk material, as well as any dust or impurities that are suspended within it. Impurities usually absorb light, leading to an exponential attenuation in the intensity of the light travelling through the material. Small particulate matter – such as dust grains – will often scatter the light, removing information about where the light came from.

Both of these effects complicate the task of inferring the properties of the the charged particle from the pattern of light it leaves on the detector elements. The energy of the charged particle is inferred by the total amount of light it produces — a quantity that absorption can drastically modify. The direction of the particle is inferred both by the geometric and timing pattern that the light leaves in the detector. Scattering removes sharp features from these geometric patterns, creating indistinct blobs. The scattered light also must travel further before it strikes a photo-sensitive detector element, which delays and smooths out the timing distribution of the light. As all large neutrino detectors use photo-sensitive elements with nanosecond timing resolution, and an increase in average path length of a few tens of meters will translate into tens of nanoseconds lost.

For this reason, the accurate determination of these material properties is crucial for utilising a detector to its fullest degree. In Ice-Cube, the scattering and absorption of the ice is measured in two main steps. First, a known quantity of light is injected into the detector using the flasher LEDs described in section 4.3.2. These LEDs

are mounted on each DOM, and the light propagates outwards from the string, scattering as it goes. DOMs on nearby strings measure this light, and the results are saved as part of a calibration dataset.

Second, this same process is repeated inside a simulation of the detector. This attempts to accurately model every detail of the detector operation, including the physics of light transport through the bulk as described in section 4.3.2. The amount of scattering and absorption inside the detector is iteratively adjusted, tuning the simulation toward the expectation given by the calibration dataset.

10.1 Path tracing

In IceCube, the simulation of this light transport is performed using a path tracer. The path tracer effectively solves the equations of motion for light, keeping track of its current position and direction and moving it forward through the detector.

Beer's law states that a photon with position \vec{x} and direction \hat{n} has an exponentially decreasing probability of traveling a distance L without being absorbed:

$$P_a(L|\vec{x}, \hat{n}) = e^{-\tau_a(L; \vec{x}, \hat{n})}, \quad (10.1)$$

where $\tau_a(L)$ is the absorption depth for a distance L :

$$\tau_a(L; \vec{x}, \hat{n}) = \int_0^L a(\vec{x} + s\hat{n}) ds, \quad (10.2)$$

and $a(\vec{x})$ is the absorption length of the material.

The same equations govern the probability of traveling a distance L without being scattered, with geometric scattering depth $\tau_b(L; \vec{x}, \hat{n})$ and scattering length $b(\vec{x})$. Once scattered, the photon follows a new direction according to an angular scattering distribution $\sigma(\cos\theta)$.

These principles can be used to perform path tracing as described in algorithm 3. A photon is initialised according to some desired starting location and direction. Then, a scattering depth is drawn from a logarithmic distribution, and the equivalent distance inside the detector is calculated from the scattering length. A geometric calculation determines if the current direction of the photon points towards a photo-detector, and if it is, calculates the distance to that detector element. If the photon would hit the element before the next scatter, then the path is guaranteed to terminate. All that remains is to check that the photon does not get absorbed before it reaches the detector element. If the next scatter would occur before the photon hits the detector, the absorption length is also checked to make sure that the photon does not get absorbed before the next scatter. In this

case, if no absorption happens, the photon is moved to the next scattering location, and the direction is rotated according to the angular scattering distribution σ .

Algorithm 3 Path tracer

```

1:  $\vec{x} \leftarrow$  initial location.
2:  $\vec{n} \leftarrow$  initial direction
3: for  $i \leftarrow 0, \dots, i_{\max}$  do
4:    $\tau_b \leftarrow -\ln(\text{uniform}(0, 1))$ 
5:    $\Delta L \leftarrow$  solve  $\int_0^{\Delta L} b(\vec{x} + s\hat{n})ds = \tau_b$ 
6:    $d \leftarrow$  straight line distance to target
7:   if  $d < \Delta L$  then
8:      $\tau_a = \int_0^d a(\vec{x} + s\hat{n})ds$ 
9:     if  $-\ln(\text{uniform}(0, 1)) < \tau_a$  then reject path
10:    else accept path
11:   else
12:      $\tau_a = \int_0^{\Delta L} a(\vec{x} + s\hat{n})ds$ 
13:     if  $-\ln(\text{uniform}(0, 1)) < \tau_a$  then reject path
14:     else
15:        $\vec{x} \leftarrow \vec{x} + \Delta L\hat{n}$ 
16:        $\cos \theta \leftarrow$  sample from  $\sigma(\cos \theta)$ 
17:        $\phi \leftarrow \text{uniform}(0, 2\pi)$ 
18:        $\hat{n} \leftarrow \text{rotate}(\hat{n}, \cos \theta, \phi)$ 

```

This algorithm is simple enough that it can be efficiently evaluated on a GPU. If too many photons are being absorbed before they are detected, the algorithm can be modified to keep track of a photon weight, which is reduced when the photon should have scattered, instead of terminating the path entirely. The resultant paths then form a set of weighted samples.

However, if the detector is exceptionally large, as in the case of IceCube [3, 23] or Antares [15], then scattering can become the primary issue. The solid angle to the nearest light sensitive element in these detectors can be as low as $\mathcal{O}(10^{-6})$. Thus, many photons get lost in the bulk, terminating their tracing without reaching a light sensitive element. This leads to a large inefficiency in simulation, and inferring the characteristics of the bulk can require upwards of 10,000 GPU hours [49].

10.2 Path integration

This problem is best characterised by a highly constrained initial and final state for the light rays. Path tracers perform poorly because they can constrain either the initial or final state, but not both. To impose

both constraints, the entire history of the light ray must be specified up front, forming a path. The intensity of the light that reaches the light sensitive elements can then be found by solving a path integral.

The use of path integration for the simulation of light propagation was first developed for the rendering of computer generated images, where it is known as Metropolis Light Transport (MLT) [226]. The scenes that MLT was developed to render are dominated by reflections off hard surfaces, and scattering is usually treated as a perturbation [203, 204].

Path integration can be approached numerically by specifying the path, \mathbf{f} , as a series of straight line segments connected by common vertices \vec{f}_n :

$$\int_{\Omega} e^{-S[\mathbf{f}]} D\mathbf{f} \approx \int_{\Omega} p(\vec{f}_0, \vec{f}_1, \dots) D\mathbf{f}. \quad (10.3)$$

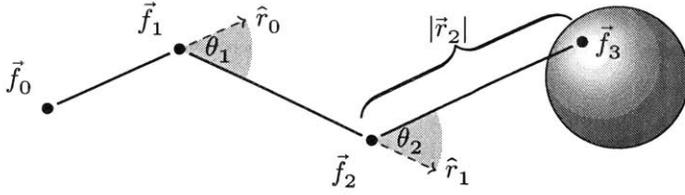
Each vertex – except the first and last – is a location where the light ray scatters, and changes direction. Instead of deriving the action, $S[\mathbf{f}]$, explicitly, the problem is framed in terms of a probability distribution $p(\dots)$. Thus the integral can be sampled using Markov Chain Monte-Carlo tools, which were developed for statistical inference.

The set of all paths, Ω , is a union of multidimensional vector spaces, as the number of scattering points is a variable quantity. For example, when the scattering length is long, the most probable path will involve few scattering vertices, while a shorter scattering length requires more vertices. The sampling algorithm must be able to change dimensionality, to adapt the number of vertices to the problem at hand. Trans-dimensional sampling can be performed using the Reversible Jump Markov Chain Monte-Carlo algorithm described in chapter 2.

10.3 The path probability distribution

It may seem most natural to express the path probability distribution in spherical coordinates. However, using the example shown in figure 10.1, the position of vertex \vec{f}_2 depends not just on $|\vec{r}_1|$ and θ_1 , but also on $|\vec{r}_0|$ and θ_0 . By using Cartesian coordinates, the probability distribution can be factorised into distributions for the initial (starting) vertex, the intermediate vertices, and the final vertex:

$$p(\vec{f}_0, \vec{f}_1, \dots, \vec{f}_n) = p_i(\vec{f}_0, \vec{f}_1) \left[\prod_{k=1}^{n-2} p_v(\vec{f}_{k-1}, \vec{f}_k, \vec{f}_{k+1}) \right] p_f(\vec{f}_{n-2}, \vec{f}_{n-1}). \quad (10.4)$$

**Figure 10.1:**

An example of a path specified as a series of four vertices connected by straight line segments.

10.3.1 Initial vertex

The initial vertex distribution captures the profile of emitted light. In this derivation, a point source is assumed, but other geometries can be derived in analogy to the final vertex. The pair of 3D coordinates \vec{f}_0 and \vec{f}_1 give the outgoing direction from the point source located at \vec{f}_0 . The initial vertex probability is

$$p_i(\vec{f}_0, \vec{f}_1) = \varepsilon(\hat{r}_0), \quad (10.5)$$

where ε is the *angular emission distribution*.

10.3.2 Intermediate vertex

The intermediate vertex distribution is composed of two factors: the probability of the light traveling from the previous vertex, \vec{f}_{k-1} , to this vertex, \vec{f}_k , along a straight line path

$$\vec{v}_k(s) = \vec{f}_{k-1} + s\vec{r}_{k-1}, \quad (10.6)$$

where $\vec{r}_{k-1} = \vec{r}_k - \vec{r}_{k-1}$, and then scattering at \vec{f}_k . The probability of traveling this distance is given by the product of two distributions: the probability density function of the ray scattering at \vec{f}_k , and the cumulative distribution function of the ray not being absorbed along this journey, of which both distributions are exponential. The second factor is the probability of the light changing direction, given by the angular scattering distribution σ . The intermediate vertex probability is

$$p_v(\vec{f}_{k-1}, \vec{f}_k, \vec{f}_{k+1}) = \frac{b(\vec{f}_k)e^{-\tau(k)}}{|\vec{r}_{k-1}|^2} \sigma(\cos \theta_k), \quad (10.7)$$

where $\cos \theta_k = \hat{r}_{k-1} \cdot \hat{r}_k$, and $\tau(k)$ is the total optical depth:

$$\tau(k) = \tau_a(k) + \tau_b(k), \quad (10.8)$$

$$\tau_a(k) = \int_0^1 a(\vec{v}_k(s)) ds, \quad (10.9)$$

$$\tau_b(k) = \int_0^1 b(\vec{v}_k(s)) ds. \quad (10.10)$$

10.3.3 Final vertex

Light sensitive elements generally have a 2D topology, and in this derivation a spherical element is assumed. The light ray must travel the final segment without scattering or absorption, so the probability of each is given by a cumulative distribution function. The constraint that the final vertex must be located on the element surface introduces a $\cos \theta / r^2$ geometric term, where θ is the angle of the incoming ray to the surface normal. The final vertex probability is

$$p_f(\vec{f}_{n-2}, \vec{f}_{n-1}) = \frac{-\hat{r}_{n-2} \cdot \vec{\nu}(\vec{f}_{n-1})}{|\vec{r}_{n-2}|^2} \rho(\vec{f}_{n-1}), \quad (10.11)$$

where $\vec{\nu}(\vec{f}_{n-1})$ is the surface normal at \vec{f}_{n-1} , and ρ is the probability of detection conditioned on the vertex location \vec{f}_n .

10.3.4 Jump proposal

Only jumps that add or remove a single vertex are considered, and are often called birth/death moves in the literature [94, 207]. A new vertex, \vec{f}' , is added to the path by placing it between two already existing adjacent vertices. A pair of adjacent vertices, \vec{f}_k and \vec{f}_{k+1} , is chosen with probability proportional to its respective scattering depth:

$$p(n \rightarrow n+1, k) = \frac{\tau_b(k)}{\sum_{l=0}^{n-1} \tau_b(l)}. \quad (10.12)$$

A vertex is removed by reconnecting its adjacent vertices. The initial and final vertices cannot be removed, and intermediate vertices are chosen with equal probability:

$$p(n \rightarrow n-1, k) = \frac{1}{n-2}. \quad (10.13)$$

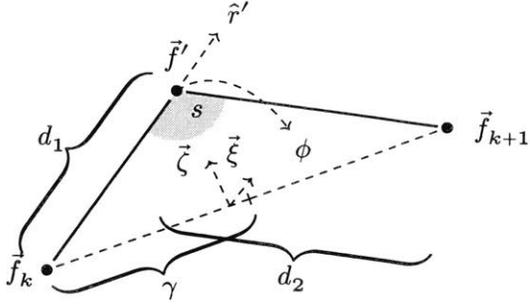
10.3.5 Coordinate transform

The new vertex is placed at a position governed by the distribution q . In practice, equation 10.7 is dominated by the highly forward peaked angular scattering distribution. Thus, it is most natural to let q be a distribution over bispherical coordinates s, t, ϕ . Figure 10.2 shows that the new vertex, \vec{f}' , is placed between the existing vertices \vec{f}_k and \vec{f}_{k+1} which both form the foci for the coordinate system.

$$\vec{f}'(s, t, \phi) = \vec{f}_k + \gamma \frac{\hat{r}_k \sinh t + \sin s (\hat{\xi} \cos \phi + \hat{\zeta} \sin \phi)}{\cosh t - \cos s}, \quad (10.14)$$

where both $\hat{\xi}$ and $\hat{\zeta}$ are mutually perpendicular unit vectors to \hat{r}_k , and $\gamma = |\vec{r}_k|/2$. The Jacobian factor for this transformation is

$$J = \left(\frac{\gamma}{\cosh t - \cos s} \right)^3 \sin s, \quad (10.15)$$


Figure 10.2:

A new vertex \vec{f}' is inserted using a bi-spherical coordinate system s , ϕ and $t = \ln(d_1/d_2)$.

and the $q(s, t, \phi)$ distribution is factorable into three independent distributions:

$$q(s) = \frac{\beta e^{-\beta \cos s}}{2 \sinh \beta}, \quad (10.16)$$

$$q(\phi) = \frac{1}{2\pi}, \quad (10.17)$$

$$q(t) = (2 + 2 \cosh t)^{-1}, \quad (10.18)$$

where β is a tuneable parameter.

10.4 Implementation

The path sampler and path tracer were both written in C++, targeting the CPU. The path tracer propagates the ray by a distance drawn from an exponential distribution, then changes direction based on the angular scattering distribution. If the ray strikes the light sensitive sphere, then the length of the ray is recorded along with a weight which reflects the conditional detection probability ρ .

10.4.1 Comparison of timing distribution

A synthetic test case was constructed to compare the path sampling method against a path tracer. In this test, both the angular emission distribution, ε , and angular scattering distribution, σ , were chosen to be Kent distributions:

$$\varepsilon(\hat{r}_0) = \frac{\kappa e^{-\kappa \hat{r}_0 \cdot \hat{\varepsilon}}}{4\pi \sinh \kappa}, \quad \sigma(\cos \psi) = \frac{\kappa e^{-\kappa \cos \psi}}{4\pi \sinh \kappa}, \quad (10.19)$$

with a κ of 20, and mean direction of $\hat{\varepsilon} = (0, 0, 1)$ for the emission distribution. The initial vertex was constrained to the origin, while the final vertex was constrained to a sphere of radius 30 centimeters

at (0, 0, 120) meters. The conditional detection probability, ρ , was chosen to be exponentially distributed:

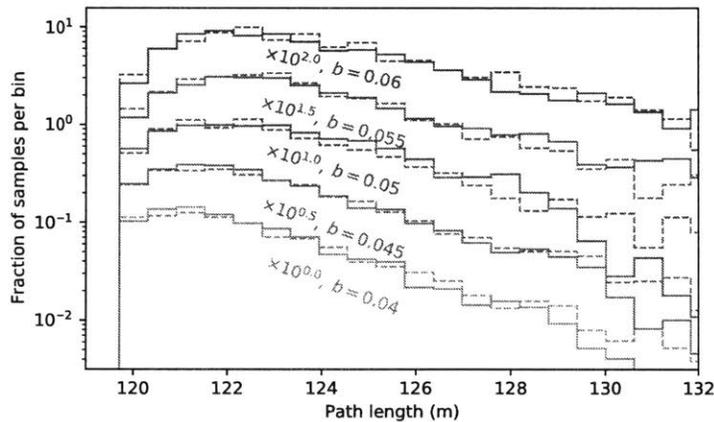
$$\rho(\vec{f}_{n-1}) = \exp\left(\kappa\left(\hat{\rho} \cdot \frac{\vec{f}_{n-1} - \vec{\eta}}{|\vec{f}_{n-1} - \vec{\eta}|} - 1\right)\right), \quad (10.20)$$

with a κ of 3, a mean direction of $\hat{\rho} = (0, -1, 0)$, and a end point of $\vec{\eta} = (0, 0, 120)$ meters.

In between jump moves, the vertex positions are updated using standard Metropolis-Hastings step moves. The proposal distribution for the intermediate vertices are normal distributions with variance proportional to $1/(n + 2)$. The proposal distribution for the final vertex is a Kent distribution. The starting vertex is not updated in the step move, as it is constrained to a point in this example.

Figure 10.3:

The distribution of paths lengths for various values of the scattering parameter. The solid and dashed lines are generated by the path sampler and path tracer respectively. Each pair of distributions are offset by the stated value for visual presentation.



The total amount of light received by a light sensitive element is related to the total amount of absorption and scattering. The relative amount of scattering to absorption can be measured using the time delay of the light reaching the elements; more scattering leads to longer, meandering paths through the bulk. The time delay of each ray is directly proportional to the length of the path it takes, as the bulk is assumed to have a constant index of refraction.

Figure 10.3 shows a comparison between the distribution of path lengths generated by both the path tracer and path sampler for various scattering parameters and an absorption parameter of $a = 0.01 \text{ m}^{-1}$. The solid and dashed lines are generated by the path sampler and path tracer respectively. The path tracer cast 600M rays, of which ~ 5000 landed on the detector element; generally the minimum required to provide a good constraint on the scattering parameter [49]. The path sampler took 200k samples, from which 40k were discarded as a burn-in period. The total CPU time of the path tracer was 70 minutes,

while the path sampler took 0.5 seconds. The path sampler produces a comparable distribution with significantly less run time.

10.4.2 Comparison of angular distributions

In this case, the angular distribution of arrival directions will be compared to a path tracer. The angular scattering distribution, σ , was chosen to be a more realistic mixture of a simplified Liu and Henyey-Greenstein distribution, as presented in section 4.3.1:

$$\sigma(\cos \psi) = f_{\text{SL}} p_{\text{SL}}(\cos \psi) + (1 - f_{\text{SL}}) p_{\text{HG}}(\cos \psi), \quad (10.21)$$

with $g = 0.95$, and $f_{\text{SL}} = 0.45$. The emission distribution remained the same as in section 10.4.1.

To simulate an environment where the scattering length is not isotropically uniform, the scattering length was defined as a smoothy varying hyperbolic tan function:

$$b(\vec{x}) = b_0 + (b_1 - b_0) \left(1 + \frac{1}{2} \tanh(\vec{x}_x) \right), \quad (10.22)$$

where b_1 was an adjustable value, and $b_0 = 0.01 \text{ m}^{-1}$.

In order to exacerbate the angular effects, the final vertex was constrained to a sphere with a radius of 10 meters. The conditional detection probability was also updated to better reflect the angular efficiency of a DOM:

$$\rho(\vec{f}_{n-1}) = \exp(3 \cos \omega - \ln \cosh(2 \cos \omega + 0.7) - 1), \quad (10.23)$$

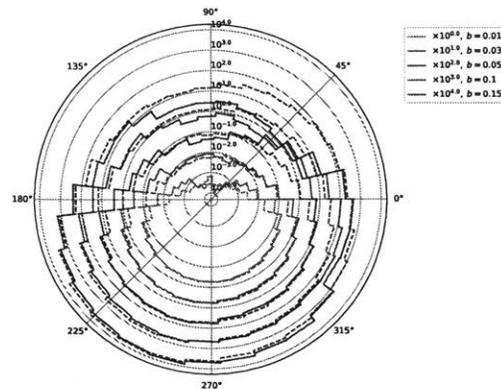
$$\cos \omega = \hat{\rho} \cdot \frac{\vec{f}_{n-1} - \vec{\eta}}{|\vec{f}_{n-1} - \vec{\eta}|} \quad (10.24)$$

with a κ of 3, a mean direction of $\hat{\rho} = (0, -1, 0)$, and a end point of $\vec{\eta} = (0, 0, 120)$ meters.

As the rearmost arrival angles at suppressed in probability by ~ 4 orders of magnitude, a more efficient sampler is required to explore this tail of the path distribution. A Riemann manifold Langevin Monte Carlo algorithm was implemented, which makes use of the second order derivatives of the likelihood to better match proposals to the path probability distribution. The second order derivatives were calculated using a reverse mode automatic differentiation algorithm called ‘‘edge-pushing’’ which specialises in calculations of the hessian of a function [121, 167]. The curvature metric of Riemann manifold Langevin Monte Carlo is required to be a positive definite matrix, but the Hessian is indefinite in general. Methods exist for finding a ‘nearby’ positive definite matrix such as the `softabs`

Figure 10.4:

The distribution of arrival angles for various values of b_1 in equation 10.22. The solid and dashed lines are generated by the path sampler and path tracer respectively. Each pair of distributions are offset by the stated value for visual presentation.



function [41] and the modified Cholesky decomposition [95]; however, best performance was found when just the main diagonal of the hessian was used, with all off diagonal elements removed.

Figure 10.4 shows a comparison between the angular arrival distribution of paths generated by both the path tracer and path sampler for various values of b_1 and an absorption parameter of $a = 0.01 \text{ m}^{-1}$. The solid and dashed lines are generated by the path sampler and path tracer respectively. The path tracer cast 600M rays, of which $\sim 1\text{M}$ landed on the detector element, owing to its much greater size in this test. The path sampler took 1M samples, from which 50k were discarded as a burn-in period. The path sampler produces a comparable distribution for all scattering parameters except $b_1 = 0.15 \text{ m}^{-1}$ where the rearmost tail of the distribution is under-sampled.

10.5 Discussion

In the synthetic test, a performance improvement of 8,000 times was observed on a CPU implementation. While this is a highly encouraging result, some caution must be exercised in predicting the performance increase when applied to light simulation in an experiment.

Path tracing is performed most efficiently on a GPU, where improvements of up to 100 times [59] are possible as it is a highly parallelisable algorithm. In contrast, an MCMC is an intrinsically serial algorithm. Nevertheless, multiple chains could be run in parallel; although, the utilisation of a GPU will likely be less efficient compared to the path tracer.

Experiments are usually composed of multiple hundreds of detector elements. A path tracer can check for intersection of the path with all of these elements, effectively running all possible endpoints

in parallel⁷⁷. In contrast, the path sampler only samples from one endpoint by definition; with that in mind, it will naturally spend most time sampling from endpoints where the probability is highest.

The efficiency of the path tracer is highly dependent on the position of the endpoints in relation to the emission distribution. If the endpoints are off-axis, the path tracer will suffer greatly due to the highly forward peaked scattering distributions. In contrast, the efficiency of the path sampler is only weakly dependent on the position of the endpoints. Moving the endpoints further away, or off-axis, simply requires the addition of more scattering vertices.

10.6 Summary

The simulation of light inside of a large neutrino detector can be formulated as a path integral, where the constraint that light must land on a light sensitive element can be naturally expressed. An implementation which samples paths from this integral was developed for the CPU, and compared against a path tracer. The path sampler reproduces the timing distribution for light traversing a distance of 120 meters in a medium with various scattering parameters. A performance improvement of $\sim 8,000$ times was observed, when compared to the CPU based path tracer. The path sampler also reproduces the angular arrival direction distribution in an anisotropic scattering environment, accurately reproducing the results of the path sampler for all but the tails of the $b_1 = 0.15 \text{ m}^{-1}$ distribution.

Path sampling can be extended to other kinds of experiments. In smaller neutrino detectors, light may be reflected or refracted from surfaces. These processes can be handled by combining the formulation presented here with those already developed in Metropolis Light Transport for reflection and refraction.

Particles other than photons can also be simulated. General particle physics simulations use path tracing to propagate particles such as neutrons or muons through matter, and could benefit from path sampling if the final state is constrained. An example of a constrained final state in biophysics is the diffusion of molecules to receptor sites. This technique shows promise for simulation of any particles where the initial and final states of the path are highly constrained.

⁷⁷ Although there will still be an additional computational cost incurred for checking the intersection with multiple detector elements, this can be partially alleviated with search algorithms.

Conclusion

These many threads of research are bound by one unifying principle: to find limiting factors in our analysis of physical phenomena, and address them with novel statistical and computational techniques. They span the entire domain of data analysis in particle and astrophysics; from meta-analysis of many experimental results, to analysis of a single data set, to reconstruction tools for the processing of raw data, to the simulation of a detector itself.

At the highest level, many disparate data sets from experiments around the world are combined into a global fit for sterile neutrinos in chapter 3. Some of these data sets have shown anomalous neutrino oscillation, while others have not — potential evidence for a fourth, sterile, neutrino. The global fit presented in Conrad et al. [79] was rewritten and the SciBooNE-MiniBooNE joint disappearance analysis results were included. In addition, by using a Markov chain Monte-Carlo sampler, Bayesian credible regions could be shown for the first time in the history of this global fit effort. Practical computation of a likelihood function that involves so many experiments requires the resources of a computer cluster, and a system for distributing the workload opportunistically was developed to this end. The new global fit to a $3 + 1$ model shows two 90% CL allowed regions, at $\Delta m_{41}^2 \sim 1.7 \text{ eV}^2$ and $\Delta m_{41}^2 \sim 1 \text{ eV}^2$. A $3 + 2$ model does not show significant improvement over the $3 + 1$ model. Tension between data sets in the fit remain, as muon flavour disappearance has yet to be observed. Inclusion of the recent IceCube sterile neutrino analysis in the global fit does not change the best fit point; however, the $\Delta m_{41}^2 \sim 1 \text{ eV}^2$ region is no longer favoured at 90% CL. The IceCube analysis is also sensitive to tau-sterile flavour mixing, allowing the complete $3 + 1$ mixing matrix to be constrained for the first time.

All scientific results are derived from analyses, and chapter 7 presents a search for point sources of astrophysical neutrinos with the IceCube experiment. Although IceCube has measured a flux of astrophysical neutrinos, to date no source has been identified. Previous IceCube point source searches tested a model of a single hot-spot. This analysis brought Non-Poissonian Template Fitting to bear on the problem: an advanced statistical tool for modelling an entire population of point sources. The NPTF method was adapted for use on IceCube, and simulations showed that it can

reconstruct the properties of an injected population. No evidence for populations of sources was observed, and limits on population models can be made for the first time using astrophysical neutrino data.

Analyses rely on ancillary calculations for topics outside of their scope, such as background processes that have no relation to the model under consideration. A calculation of atmospheric neutrino fluxes is given in chapter 5 to aid analyses for which atmospheric neutrinos are either a signal or a background. While atmospheric neutrino fluxes have been calculated for decades, computational constraints have prevented a detailed study on the effect of global atmospheric variations. In this calculation, the MCEq matrix cascade equation solver is used to efficiently compute the high energy neutrino flux. The flux was computed over an array of times and locations spanning a year and the surface of the Earth. In addition, various models for the cosmic ray spectrum and hadronic interaction model were considered. At each location, the local atmospheric properties were derived from the AIRS instrument on the Aqua climate science satellite, and were used to generate the atmospheric flux that would reach IceCube. The publicly released results form the most in-depth study on systematic effects for atmospheric fluxes, and were used in the IceCube sterile neutrino search.

Analyses are also reliant on reconstruction tools in order to recover physical observables from raw data. The increasing complexity of neutrino detectors and demands placed on analyses necessitates a new approach. In chapter 9, convolutional neural networks are used to develop a reconstruction algorithm to classify tau neutrino events. Tau neutrinos, while predicted to be present in IceCube data, have yet to be positively identified. The convolutional neural network approach provides a similar sensitivity to existing methods, but with half the background event rate.

At the lowest level, a fundamentally new method of simulating the propagation of light through bulk materials is demonstrated in chapter 10. By posing the problem as a path integral, the initial and final states of the light can be constrained. This is unlike current path tracing techniques, where many simulated photons must be discarded due to not meeting the final state requirements. Inspired by the Metropolis light transport algorithm used in the CGI industry, this path sampler uses Markov chain Monte-Carlo algorithms to draw samples from the path integrand. The path sampler agrees with a reference path tracer to within statistical uncertainty when comparing the timing and angular distributions of photons arriving at a detector element. In the case of the timing distribution, the path sampler is estimated to be faster by a factor of 8,000 compared to the path tracer in medium with relatively low amounts of scattering.

List of contributions

- The sterile neutrino global fit to short base-line data presented in sections 3.3 and 3.4 is entirely my own work, based on – but otherwise entirely overhauled – the previous global fit analysis of Conrad et al. [79], and advised by Prof. Janet Conrad and Prof. Mike Shaevitz. This result was published as Collin et al. [76].
- The sterile neutrino global fit including IceCube data presented in section 3.5 is the work of myself and Dr. Carlos Argüelles, advised by Prof. Janet Conrad and Prof. Mike Shaevitz. This result was published as Collin et al. [75].
- The atmospheric flux calculation presented in chapter 5 is entirely my own work. The fluxes have been made publicly available in Collin [77].
- The search for point sources of astrophysical neutrinos presented in chapter 7 is the work of myself, Nick Rodd, and Jimmy DeLaunay, advised by Dr. Ben Safdi.
- The application of convolutional neural networks to classify tau neutrino events presented in chapter 9 is entirely my own work.
- The demonstration of a path integral approach to simulating light propagation presented in chapter 10 is entirely my own work.

Bibliography

- [1] AMS Collaboration et al. “First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV”. In: *Physical Review Letters* 110.14 (Apr. 3, 2013), p. 141102. doi: 10.1103/PhysRevLett.110.141102.
- [2] M. G. Aartsen and others. “Searches for Sterile Neutrinos with the IceCube Detector”. In: (2016). arXiv: 1605.01990 [hep-ex].
- [3] M. G. Aartsen et al. “IceCube-Gen2: A Vision for the Future of Neutrino Astronomy in Antarctica”. In: *PoS FRAPWS2016* (Nov. 6, 2017), p. 004.
- [4] M. G. Aartsen et al. “Search for Astrophysical Tau Neutrinos in Three Years of IceCube Data”. In: *Phys.Rev.* D93 (Jan. 12, 2016), p. 022001. doi: 10.1103/PhysRevD.93.022001.
- [5] Martin Abadi et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”. In: (2015). Software available from tensorflow.org.
- [6] Kevork N. Abazajian. “Telling Three from Four Neutrinos with Cosmology”. In: *Astropart. Phys.* 19 (2003), pp. 303–312. doi: 10.1016/S0927-6505(02)00204-9. arXiv: astro-ph/0205238 [astro-ph].
- [7] Kevork N. Abazajian and Manoj Kaplinghat. “Neutrino Physics from the Cosmic Microwave Background and Large-Scale Structure”. In: *Ann Rev. Nuc. Part. Phys.* 66 (2016), pp. 401–420. doi: 10.1146/annurev-nucl-102014-021908.
- [8] Kevork Abazajian et al. “Cosmological Lepton Asymmetry, Primordial Nucleosynthesis, and Sterile Neutrinos”. In: *Phys. Rev.* D72 (2005), p. 063004. doi: 10.1103/PhysRevD.72.063004. arXiv: astro-ph/0410175 [astro-ph].

- [9] J. N. Abdurashitov et al. “Measurement of the Solar Neutrino Capture Rate with Gallium Metal. III. Results for the 2002\char21{}2007 Data-Taking Period”. In: *Physical Review C* 80.1 (July 30, 2009), p. 015807. doi: 10.1103/PhysRevC.80.015807.
- [10] B. Achkar et al. “Search for Neutrino Oscillations at 15, 40 and 95 Meters from a Nuclear Power Reactor at Bugey”. In: *Nuclear Physics B* 434.3 (Jan. 30, 1995), pp. 503–532. issn: 0550-3213. doi: 10.1016/0550-3213(94)00513-E.
- [11] Ackermann M. et al. “Optical Properties of Deep Glacial Ice at the South Pole”. In: *Journal of Geophysical Research: Atmospheres* 111.D13 (July 8, 2006). issn: 0148-0227. doi: 10.1029/2005JD006687.
- [12] P. Adamson and others. “Active to Sterile Neutrino Mixing Limits from Neutral-Current Interactions in MINOS”. In: *Phys. Rev. Lett.* 107 (2011), p. 011802. doi: 10.1103/PhysRevLett.107.011802. arXiv: 1104.3922 [hep-ex].
- [13] P. Adamson and others. “First Measurement of ν_{μ} and N Events in an Off-Axis Horn-Focused Neutrino Beam”. In: *Phys. Rev. Lett.* 102 (2009), p. 211801. doi: 10.1103/PhysRevLett.102.211801. arXiv: 0809.2447 [hep-ex].
- [14] O. Adriani et al. “PAMELA Measurements of Cosmic-Ray Proton and Helium Spectra”. In: *Science* 332.6025 (Apr. 1, 2011), pp. 69–72. issn: 0036-8075, 1095-9203. doi: 10.1126/science.1199172. arXiv: 1103.4055.
- [15] M. Ageron et al. “ANTARES: The First Undersea Neutrino Telescope”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 656.1 (Nov. 11, 2011), pp. 11–38. issn: 0168-9002. doi: 10.1016/j.nima.2011.06.103.
- [16] A. A. Aguilar-Arevalo and others. “A Search for Electron Neutrino Appearance at the $\Delta M^2 \sim 1\text{eV}^2$ Scale”. In: *Phys. Rev. Lett.* 98 (2007), p. 231801. doi: 10.1103/PhysRevLett.98.231801. arXiv: 0704.1500 [hep-ex].
- [17] A. A. Aguilar-Arevalo and others. “Event Excess in the MiniBooNE Search for $\nu_{\mu} \rightarrow \nu_e$ Oscillations”. In: *Phys. Rev. Lett.* 105 (2010), p. 181801. doi: 10.1103/PhysRevLett.105.181801. arXiv: 1007.1150 [hep-ex].

- [18] A. A. Aguilar-Arevalo and others. “Improved Search for $\nu_{\mu} \rightarrow N$ Oscillations in the MiniBooNE Experiment”. In: *Phys. Rev. Lett.* 110 (2013), p. 161801. doi: 10.1103/PhysRevLett.110.161801. arXiv: 1207.4809 [hep-ex].
- [19] A. A. Aguilar-Arevalo and others. “Unexplained Excess of Electron-Like Events From a 1-GeV Neutrino Beam”. In: *Phys. Rev. Lett.* 102 (2009), p. 101802. doi: 10.1103/PhysRevLett.102.101802. arXiv: 0812.2243 [hep-ex].
- [20] A.A. Aguilar-Arevalo. In: *Private communication* (2015).
- [21] A. Aguilar et al. “Evidence for Neutrino Oscillations from the Observation of Electron Anti-Neutrinos in a Muon Anti-Neutrino Beam”. In: *Physical Review D* 64.11 (Nov. 13, 2001). issn: 0556-2821, 1089-4918. doi: 10.1103/PhysRevD.64.112007. arXiv: hep-ex/0104049.
- [22] Eun-Joo Ahn et al. “Cosmic Ray Event Generator Sibyll 2.1”. In: *Physical Review D* 80.9 (Nov. 4, 2009). issn: 1550-7998, 1550-2368. doi: 10.1103/PhysRevD.80.094003. arXiv: 0906.4113.
- [23] J Ahrens et al. “Sensitivity of the IceCube Detector to Astrophysical Sources of High Energy Muon Neutrinos”. In: *Astroparticle Physics* 20.5 (Feb. 1, 2004), pp. 507–532. issn: 0927-6505. doi: 10.1016/j.astropartphys.2003.09.003.
- [24] E. N. Alekseev et al. “Possible Detection of a Neutrino Signal on 23 February 1987 at the Baksan Underground Scintillation Telescope of the Institute of Nuclear Research”. In: *JETP Lett.* 45 (1987), pp. 589–592.
- [25] Edoardo Amaldi. “From the Discovery of the Neutron to the Discovery of Nuclear Fission”. In: *Physics Reports* 111.1 (Sept. 1, 1984), pp. 1–331. issn: 0370-1573. doi: 10.1016/0370-1573(84)90214-X.
- [26] M. Antonello and others. “A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam”. In: (2015). arXiv: 1503.01520 [physics.ins-det].
- [27] David Applebaum. *Probability and Information: An Integrated Approach*. 2nd ed. New York, NY, USA: Cambridge University Press, 2008. isbn: 978-0-521-72788-4.

- [28] Maria Archidiacono et al. “Pseudoscalar—sterile Neutrino Interactions: Reconciling the Cosmos with Neutrino Oscillations”. In: *JCAP* 1608.08 (2016), p. 067. doi: 10.1088/1475-7516/2016/08/067. arXiv: 1606.07673 [astro-ph.CO].
- [29] Carlos A. Argüelles Delgado. “New Physics with Atmospheric Neutrinos”. University of Wisconsin - Madison, 2015.
- [30] Carlos Alberto Argüelles Delgado, Jordi Salvado, and Christopher N. Weaver. “A Simple Quantum Integro-Differential Solver (SQuIDS)”. In: (2014). arXiv: 1412.3832 [hep-ph].
- [31] Carlos Alberto Argüelles Delgado, Jordi Salvado, and Christopher N. Weaver. “ ν -SQuIDS”. In: ().
- [32] Carlos A. Argüelles et al. “High-Energy Behavior of Photon, Neutrino, and Proton Cross Sections”. In: *Phys. Rev. D* 92.7 (2015), p. 074040. doi: 10.1103/PhysRevD.92.074040. arXiv: 1504.06639 [hep-ph].
- [33] Carlos A. Argüelles, Teppei Katori, and Jordi Salvado. “New Physics in Astrophysical Neutrino Flavor”. In: *Physical Review Letters* 115.16 (Oct. 15, 2015). ISSN: 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.115.161303. arXiv: 1506.02043.
- [34] P. Astier et al. “Search for $N_{\mu} \rightarrow \nu_e$ Oscillations in the NOMAD Experiment”. In: *Physics Letters B* 570.1 (Sept. 18, 2003), pp. 19–31. ISSN: 0370-2693. doi: 10.1016/j.physletb.2003.07.029.
- [35] *Astroparticle Physics Roadmap - Phase I*. Astroparticle Physics European Coordination, p. 145.
- [36] C. Athanassopoulos et al. “The Liquid Scintillator Neutrino Detector and LAMPF Neutrino Source”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 388.1-2 (Mar. 1997), pp. 149–172. ISSN: 01689002. doi: 10.1016/S0168-9002(96)01155-2. arXiv: nucl-ex/9605002.
- [37] H. H. Aumann et al. “AIRS/AMSU/HSB on the Aqua Mission: Design, Science Objectives, Data Products, and Processing Systems”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2 (Feb. 2003), pp. 253–264. ISSN: 0196-2892. doi: 10.1109/TGRS.2002.808356.

- [38] David Balduzzi et al. “The Shattered Gradients Problem: If Resnets Are the Answer, Then What Is the Question?” In: (Feb. 27, 2017). arXiv: 1702.08591 [cs, stat].
- [39] Gianpaolo Bellini and L. Ludhova. *Neutrino Physics and Astrophysics*. IOS Press, 2012. 407 pp. ISBN: 978-1-61499-172-4.
- [40] Luis Bento and Zurab Berezhiani. “Blocking Active Sterile Neutrino Oscillations in the Early Universe with a Majoron Field”. In: *Phys. Rev. D* 64 (2001), p. 115015. DOI: 10.1103/PhysRevD.64.115015. arXiv: hep-ph/0108064 [hep-ph].
- [41] M. J. Betancourt. “A General Metric for Riemannian Manifold Hamiltonian Monte Carlo”. In: 8085 (2013), pp. 327–334. DOI: 10.1007/978-3-642-40020-9_35. arXiv: 1212.4693 [physics, stat].
- [42] Peter J. Bickel and Kjell A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. 2nd ed. Prentice Hall, 2001. 582 pp. ISBN: 978-0-13-850363-5.
- [43] Pasquale Blasi. “The Origin of Galactic Cosmic Rays”. In: *The Astronomy and Astrophysics Review* 21.1 (Nov. 2013). ISSN: 0935-4956, 1432-0754. DOI: 10.1007/s00159-013-0070-7. arXiv: 1311.7346.
- [44] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. New York: Springer-Verlag, 2002. ISBN: 978-0-387-95364-9.
- [45] A. M. Bykov et al. “Pulsar Wind Nebulae with Bow Shocks: Non-Thermal Radiation and Cosmic Ray Leptons”. In: *Space Science Reviews* 207.1-4 (July 2017), pp. 235–290. ISSN: 0038-6308, 1572-9672. DOI: 10.1007/s11214-017-0371-7. arXiv: 1705.00950.
- [46] Bob Carpenter et al. “The Stan Math Library: Reverse-Mode Automatic Differentiation in C++”. In: (Sept. 23, 2015). arXiv: 1509.07164 [cs].
- [47] George Casella and Roger L. Berger. *Statistical Inference*. Thomson Learning, 2002. 692 pp. ISBN: 978-0-534-24312-8.
- [48] G. Cheng and others. “Dual Baseline Search for Muon Antineutrino Disappearance at $0.1 \text{ eV}^2 < \Delta m^2 < 100 \text{ eV}^2$ ”. In: *Phys. Rev. D* 86 (2012), p. 052009. DOI: 10.1103/PhysRevD.86.052009. arXiv: 1208.0322.

- [49] Dmitry Chirkin. private communication. 2017.
- [50] Yi-Zen Chu and Marco Cirelli. “Sterile Neutrinos, Lepton Asymmetries, Primordial Elements: How Much of Each?” In: *Phys. Rev. D* 74 (2006), p. 085015. doi: 10.1103/PhysRevD.74.085015. arXiv: astro-ph/0608206 [astro-ph].
- [51] J. Clay. “Penetrating Radiation”. In: *Proceedings of the Section of Sciences, Koninklijke Akademie van Wetenschappen te Amsterdam* 30 (1927), pp. 1115–1127.
- [52] Stefan Coenders. “High-Energy Cosmic Ray Accelerators: Searches with IceCube Neutrinos”. Technische Universität München, 2016.
- [53] A. Weinstein for the VERITAS Collaboration. “Pulsar Wind Nebulae and Cosmic Rays: A Bedtime Story”. In: *Nuclear Physics B - Proceedings Supplements* 256-257 (Nov. 2014), pp. 136–148. issn: 09205632. doi: 10.1016/j.nuclphysbps.2014.10.017. arXiv: 1411.2532.
- [54] IceCube Collaboration. *IC86 Sterile Neutrino Data Release*. June 24, 2016. URL: <http://icecube.wisc.edu/science/data/IC86-sterile-neutrino>.
- [55] IceCube Collaboration et al. “All-Sky Search for Time-Integrated Neutrino Emission from Astrophysical Sources with 7 Years of IceCube Data”. In: *The Astrophysical Journal* 835.2 (Jan. 24, 2017), p. 151. issn: 1538-4357. doi: 10.3847/1538-4357/835/2/151. arXiv: 1609.04981.
- [56] IceCube Collaboration et al. “Anisotropy in Cosmic-Ray Arrival Directions in the Southern Hemisphere with Six Years of Data from the IceCube Detector”. In: *The Astrophysical Journal* 826.2 (Aug. 2, 2016), p. 220. issn: 1538-4357. doi: 10.3847/0004-637X/826/2/220. arXiv: 1603.01227.
- [57] IceCube Collaboration et al. “Flavor Ratio of Astrophysical Neutrinos above 35 TeV in IceCube”. In: *Physical Review Letters* 114.17 (Apr. 28, 2015). issn: 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.114.171102. arXiv: 1502.03376.
- [58] IceCube Collaboration et al. “IceTop: The Surface Component of IceCube”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 700 (Feb. 2013), pp. 188–220. issn: 01689002. doi: 10.1016/j.nima.2012.10.067. arXiv: 1207.6326.

- [59] IceCube Collaboration et al. “Measurement of South Pole Ice Transparency with the IceCube LED Calibration System”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 711 (May 2013), pp. 73–89. ISSN: 01689002. DOI: 10.1016/j.nima.2013.01.054. arXiv: 1301.5361.
- [60] IceCube Collaboration et al. “Observation and Characterization of a Cosmic Muon Neutrino Flux from the Northern Hemisphere Using Six Years of IceCube Data”. In: *The Astrophysical Journal* 833.1 (Dec. 1, 2016), p. 3. ISSN: 1538-4357. DOI: 10.3847/0004-637X/833/1/3. arXiv: 1607.08006.
- [61] IceCube Collaboration et al. “Search for Time-Independent Neutrino Emission from Astrophysical Sources with 3 Years of IceCube Data”. In: *The Astrophysical Journal* 779.2 (Dec. 3, 2013), p. 132. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/779/2/132. arXiv: 1307.6669.
- [62] IceCube Collaboration et al. “Searches for Extended and Point-like Neutrino Sources with Four Years of IceCube Data”. In: *The Astrophysical Journal* 796.2 (Nov. 12, 2014), p. 109. ISSN: 1538-4357. DOI: 10.1088/0004-637X/796/2/109. arXiv: 1406.6757.
- [63] IceCube Collaboration et al. “The IceCube Neutrino Observatory - Contributions to ICRC 2017 Part II: Properties of the Atmospheric and Astrophysical Neutrino Flux”. In: (Oct. 3, 2017). arXiv: 1710.01191 [astro-ph].
- [64] IceCube Collaboration et al. “The IceCube Neutrino Observatory: Instrumentation and Online Systems”. In: *Journal of Instrumentation* 12.03 (Mar. 14, 2017), P03012–P03012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/03/P03012. arXiv: 1612.05093.
- [65] IceCube Collaboration et al. “The IceCube Neutrino Observatory Part VI: Ice Properties, Reconstruction and Future Developments”. In: (Sept. 26, 2013). arXiv: 1309.7010 [astro-ph].
- [66] MiniBooNE Collaboration. “A Search for Muon Neutrino and Antineutrino Disappearance in MiniBooNE”. In: *Physical Review Letters* 103.6 (Aug. 4, 2009). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.103.061802. arXiv: 0903.2465.

- [67] The ALEPH Collaboration et al. “Precision Electroweak Measurements on the Z Resonance”. In: *Physics Reports* 427.5-6 (May 2006), pp. 257–454. ISSN: 03701573. DOI: 10.1016/j.physrep.2005.12.006. arXiv: hep-ex/0509008.
- [68] The AMANDA Collaboration and J. Ahrens. “Muon Track Reconstruction and Data Selection Techniques in AMANDA”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 524.1-3 (May 2004), pp. 169–194. ISSN: 01689002. DOI: 10.1016/j.nima.2004.01.065. arXiv: astro-ph/0407044.
- [69] The IceCube Collaboration. “The IceCube Data Acquisition System: Signal Capture, Digitization, and Timestamping”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 601.3 (Apr. 2009), pp. 294–316. ISSN: 01689002. DOI: 10.1016/j.nima.2009.01.001. arXiv: 0810.4930.
- [70] The IceCube Collaboration and R. Abbasi. “Time-Integrated Searches for Point-like Sources of Neutrinos with the 40-String IceCube Detector”. In: *The Astrophysical Journal* 732.1 (May 1, 2011), p. 18. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/732/1/18. arXiv: 1012.2137.
- [71] The IceCube Collaboration et al. “Calibration and Characterization of the IceCube Photomultiplier Tube”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 618.1-3 (June 2010), pp. 139–152. ISSN: 01689002. DOI: 10.1016/j.nima.2010.03.102. arXiv: 1002.2442.
- [72] The IceCube Collaboration et al. “Measurement of the Multi-TeV Neutrino Interaction Cross-Section with IceCube Using Earth Absorption”. In: *Nature* 551.7682 (Nov. 2017), pp. 596–600. ISSN: 1476-4687. DOI: 10.1038/nature24459.
- [73] The Pierre Auger Collaboration et al. “Multi-Resolution Anisotropy Studies of Ultrahigh-Energy Cosmic Rays Detected at the Pierre Auger Observatory”. In: *Journal of Cosmology and Astroparticle Physics* 2017.06 (June 13, 2017), pp. 026–026. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2017/06/026. arXiv: 1611.06812.

- [74] The Super-Kamiokande Collaboration and Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Physical Review Letters* 81.8 (Aug. 24, 1998), pp. 1562–1567. issn: 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003.
- [75] G. H. Collin et al. “First Constraints on the Complete Neutrino Mixing Matrix with a Sterile Neutrino”. In: *Physical Review Letters* 117.22 (Nov. 23, 2016). issn: 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.117.221801. arXiv: 1607.00011.
- [76] G. H. Collin et al. “Sterile Neutrino Fits to Short Baseline Data”. In: *Nuclear Physics B* 908 (July 2016), pp. 354–365. issn: 05503213. doi: 10.1016/j.nuclphysb.2016.02.024. arXiv: 1602.00671.
- [77] Gabriel Collin. *An Estimation of Systematics for Up-Going Atmospheric Muon Neutrino Flux at the South Pole*. Aug. 11, 2015.
- [78] J. M. Conrad and M. H. Shaevitz. “Limits on electron neutrino disappearance from the KARMEN and LSND ν_e -carbon cross section data”. In: *Physical Review D* 85.1 (Jan. 23, 2012), p. 013017. doi: 10.1103/PhysRevD.85.013017.
- [79] J. M. Conrad et al. “Sterile Neutrino Fits to Short Baseline Neutrino Oscillation Measurements”. In: *Adv. High Energy Phys.* 2013 (2013), p. 163897. doi: 10.1155/2013/163897. arXiv: 1207.4765 [hep-ex].
- [80] Amanda Cooper-Sarkar, Philipp Mertsch, and Subir Sarkar. “The High Energy Neutrino Cross-Section in the Standard Model and Its Uncertainty”. In: *JHEP* 08.042 (June 2011). eprint: 1106.3723.
- [81] C. L. Cowan et al. “Detection of the Free Neutrino: A Confirmation”. In: *Science* 124.3212 (July 20, 1956), pp. 103–104. issn: 0036-8075, 1095-9203. doi: 10.1126/science.124.3212.103. pmid: 17796274.
- [82] G. Cybenko. “Approximation by Superpositions of a Sigmoidal Function”. In: *Mathematics of Control, Signals and Systems* 2.4 (Dec. 1, 1989), pp. 303–314. issn: 0932-4194, 1435-568X. doi: 10.1007/BF02551274.

- [83] G. Danby et al. “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos”. In: *Physical Review Letters* 9.1 (July 1, 1962), pp. 36–44. doi: 10.1103/PhysRevLett.9.36.
- [84] Basudeb Dasgupta and Joachim Kopp. “Cosmologically Safe eV-Scale Sterile Neutrinos and Improved Dark Matter Structure”. In: *Phys. Rev. Lett.* 112.3 (2014), p. 031803. doi: 10.1103/PhysRevLett.112.031803. arXiv: 1310.6337 [hep-ph].
- [85] Tansu Daylan, Stephen K. N. Portillo, and Douglas P. Finkbeiner. “Inference of Unresolved Point Sources At High Galactic Latitudes Using Probabilistic Catalogs”. In: *The Astrophysical Journal* 839.1 (Apr. 6, 2017), p. 4. issn: 1538-4357. doi: 10.3847/1538-4357/aa679e. arXiv: 1607.04637.
- [86] *Distributed Messaging - Zeromq*. URL: <http://zeromq.org/> (visited on 05/07/2018).
- [87] D. A. Dwyer and T. J. Langford. “Spectral Structure of Electron Antineutrinos from Nuclear Reactors”. In: *Physical Review Letters* 114.1 (Jan. 7, 2015), p. 012502. doi: 10.1103/PhysRevLett.114.012502.
- [88] F. Dydak et al. “A Search for $\bar{\nu}_\mu$ Oscillations in the Δm^2 Range 0.3–90 eV²”. In: *Physics Letters B* 134.3 (Jan. 12, 1984), pp. 281–286. issn: 0370-2693. doi: 10.1016/0370-2693(84)90688-9.
- [89] A. M. Dziewonski and D. L. Anderson. “Preliminary Reference Earth Model”. In: *Physics of the Earth and Planetary Interiors* 25 (June 1981), pp. 297–356. doi: 10.1016/0031-9201(81)90046-7.
- [90] David J. Earl and Michael W. Deem. “Parallel Tempering: Theory, Applications, and New Perspectives”. In: *Physical Chemistry Chemical Physics* 7.23 (Nov. 16, 2005), pp. 3910–3916. issn: 1463-9084. doi: 10.1039/B509983H.
- [91] Ralph Engel et al. “Charm Production in SIBYLL”. In: (Feb. 23, 2015). arXiv: 1502.06353 [hep-ph].
- [92] Arman Esmaili and Alexei Yu. Smirnov. “Restricting the LSND and MiniBooNE Sterile Neutrinos with the IceCube Atmospheric Neutrino Data”. In: *JHEP* 12 (2013), p. 014. doi: 10.1007/JHEP12(2013)014. arXiv: 1307.6824 [hep-ph].

- [93] Amand Faessler et al. “Can One Measure the Cosmic Neutrino Background?” In: *International Journal of Modern Physics E* 26 (01n02 Jan. 2017), p. 1740008. ISSN: 0218-3013, 1793-6608. DOI: 10.1142/S0218301317400080. arXiv: 1602.03347.
- [94] Y. Fan and S. A. Sisson. “Reversible Jump Markov Chain Monte Carlo”. In: (Jan. 12, 2010). arXiv: 1001.2055 [stat].
- [95] Haw-ren Fang and Dianne P. O’Leary. “Modified Cholesky Algorithms: A Catalog with New Approaches”. In: *Mathematical Programming* 115.2 (Oct. 1, 2008), pp. 319–349. ISSN: 0025-5610, 1436-4646. DOI: 10.1007/s10107-007-0177-6.
- [96] Anatoli Fedynitch, Julia Becker Tjus, and Paolo Desiati. “Influence of Hadronic Interaction Models and the Cosmic Ray Spectrum on the High Energy Atmospheric Muon and Neutrino Flux”. In: *Physical Review D* 86.11 (Dec. 19, 2012). ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.86.114024. arXiv: 1206.6710.
- [97] Anatoli Fedynitch et al. “Calculation of Conventional and Prompt Lepton Fluxes at Very High Energy”. In: (Mar. 2, 2015). arXiv: 1503.00544 [astro-ph, physics:hep-ph].
- [98] G. Feinberg. “Decays of the μ Meson in the Intermediate-Meson Theory”. In: *Physical Review* 110.6 (June 15, 1958), pp. 1482–1483. DOI: 10.1103/PhysRev.110.1482.
- [99] ENRICO Fermi. “On the Origin of the Cosmic Radiation”. In: *Physical Review* 75.8 (Apr. 15, 1949), pp. 1169–1174. DOI: 10.1103/PhysRev.75.1169.
- [100] F. Feroz, M. P. Hobson, and M. Bridges. “MultiNest: An Efficient and Robust Bayesian Inference Tool for Cosmology and Particle Physics”. In: *Monthly Notices of the Royal Astronomical Society* 398.4 (Oct. 1, 2009), pp. 1601–1614. ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437.
- [101] F. Feroz et al. “Importance Nested Sampling and the MultiNest Algorithm”. In: (June 10, 2013). arXiv: 1306.2144 [astro-ph, physics:physics, stat].
- [102] F. Feroz et al. “Importance Nested Sampling and the MultiNest Algorithm”. In: (June 10, 2013). arXiv: 1306.2144 [astro-ph, physics:physics, stat].

- [103] Farhan Feroz and M. P. Hobson. “Multimodal Nested Sampling: An Efficient and Robust Alternative to MCMC Methods for Astronomical Data Analysis”. In: *Monthly Notices of the Royal Astronomical Society* 384.2 (Jan. 18, 2008), pp. 449–463. ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2007.12353.x. arXiv: 0704.3704.
- [104] Robert Foot and R. R. Volkas. “Reconciling Sterile Neutrinos with Big Bang Nucleosynthesis”. In: *Phys. Rev. Lett.* 75 (1995), p. 4350. DOI: 10.1103/PhysRevLett.75.4350. arXiv: hep-ph/9508275 [hep-ph].
- [105] Daniel Foreman-Mackey et al. “Emcee: The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125.925 (Mar. 2013), pp. 306–312. ISSN: 00046280, 15383873. DOI: 10.1086/670067. arXiv: 1202.3665.
- [106] J. A. Formaggio and G. P. Zeller. “From eV to EeV: Neutrino Cross Sections Across Energy Scales”. In: *Reviews of Modern Physics* 84.3 (Sept. 24, 2012), pp. 1307–1341. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.84.1307. arXiv: 1305.7513.
- [107] Nial Friel and Jason Wyse. “Estimating the Evidence – a Review”. In: (Nov. 8, 2011). arXiv: 1111.1957 [stat].
- [108] Edgar Gabriel et al. “Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation”. In: *Proceedings, 11th European PVM/MPI Users’ Group Meeting*. Budapest, Hungary, Sept. 2004, pp. 97–104.
- [109] T. K. Gaisser and M. Honda. “Flux of Atmospheric Neutrinos”. In: *Annual Review of Nuclear and Particle Science* 52.1 (Dec. 2002), pp. 153–199. ISSN: 0163-8998, 1545-4134. DOI: 10.1146/annurev.nucl.52.050102.090645. arXiv: hep-ph/0203272.
- [110] Thomas K. Gaisser. “Spectrum of Cosmic-Ray Nucleons and the Atmospheric Muon Charge Ratio”. In: *Astroparticle Physics* 35.12 (July 2012), pp. 801–806. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2012.02.010. arXiv: 1111.6675.
- [111] Thomas K. Gaisser, Todor Stanev, and Serap Tilav. “Cosmic Ray Energy Spectrum from Measurements of Air Showers”. In: (Mar. 14, 2013). arXiv: 1303.3565 [astro-ph].

- [112] Graciela Gelmini, Sergio Palomares-Ruiz, and Silvia Pascoli. “Low Reheating Temperature and the Visible Sterile Neutrino”. In: *Phys. Rev. Lett.* 93 (2004), p. 081302. doi: 10.1103/PhysRevLett.93.081302. arXiv: astro-ph/0403323 [astro-ph].
- [113] Vitali & I. Lazarevich Ginzburg and Serge & I. Ivanovich Syrovatski & I. *The Origin of Cosmic Rays, By V.L. Ginzburg and S.I. Syrovatskii. Translated by H.S.H. Massey and Edited by D. Ter Harr.* Macmillan, 1964. 426 pp.
- [114] Carlo Giunti and Chung W. Kim. *Fundamentals of Neutrino Physics and Astrophysics.* 2007.
- [115] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. “Deep Sparse Rectifier Neural Networks”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics.* Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. June 14, 2011, pp. 315–323.
- [116] M. C. Gonzalez-Garcia, F. Halzen, and M. Maltoni. “Physics Reach of High-Energy and High-Statistics Icecube Atmospheric Neutrino Data”. In: *Phys. Rev. D* 71 (2005), p. 093010. doi: 10.1103/PhysRevD.71.093010. arXiv: hep-ph/0502223 [hep-ph].
- [117] M. C. Gonzalez-Garcia et al. “Global Fit to Three Neutrino Mixing: Critical Look at Present Precision”. In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). issn: 1029-8479. doi: 10.1007/JHEP12(2012)123. arXiv: 1209.3023.
- [118] M.C. Gonzalez-Garcia, Michele Maltoni, and Thomas Schwetz. “Updated Fit to Three Neutrino Mixing: Status of Leptonic CP Violation”. In: *Journal of High Energy Physics* 2014.11, 52 (2014). doi: 10.1007/JHEP11(2014)052.
- [119] Jonathan Goodman and Jonathan Weare. “Ensemble Samplers with Affine Invariance”. In: *Communications in Applied Mathematics and Computational Science* 5.1 (Jan. 31, 2010), pp. 65–80. issn: 2157-5452, 1559-3940. doi: 10.2140/camcos.2010.5.65.
- [120] K. M. Gorski et al. “HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere”. In: *The Astrophysical Journal* 622.2 (Apr. 2005), pp. 759–771. issn: 0004-637X, 1538-4357. doi: 10.1086/427976. arXiv: astro-ph/0409513.

- [121] R. M. Gower and M. P. Mello. “A New Framework for the Computation of Hessians”. In: *Optimization Methods and Software* 27.2 (Apr. 1, 2012), pp. 251–273. ISSN: 1055-6788. DOI: 10.1080/10556788.2011.580098.
- [122] Peter J. Green. “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination”. In: *Biometrika* 82.4 (Dec. 1, 1995), pp. 711–732. ISSN: 0006-3444. DOI: 10.1093/biomet/82.4.711.
- [123] Richard H. R. Hahnloser et al. “Digital Selection and Analogue Amplification Coexist in a Cortex-Inspired Silicon Circuit”. In: *Nature* 405.6789 (June 2000), pp. 947–951. ISSN: 1476-4687. DOI: 10.1038/35016072.
- [124] Francis Halzen. “High-Energy Neutrino Astrophysics”. In: *Nature Physics* 13.3 (Mar. 2017), pp. 232–238. ISSN: 1745-2481. DOI: 10.1038/nphys3816.
- [125] Francis Halzen, Ali Kheirandish, and Viviana Niro. “Prospects for Detecting Galactic Sources of Cosmic Neutrinos with IceCube: An Update”. In: *Astroparticle Physics* 86 (Jan. 1, 2017), pp. 46–56. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2016.11.004.
- [126] Jan Hamann et al. “Sterile Neutrinos with eV Masses in Cosmology: How Disfavoured Exactly?” In: *JCAP* 1109 (2011), p. 034. DOI: 10.1088/1475-7516/2011/09/034. arXiv: 1108.4136 [astro-ph.CO].
- [127] Steen Hannestad, Rasmus Sloth Hansen, and Thomas Tram. “How Self-Interactions Can Reconcile Sterile Neutrinos with Cosmology”. In: *Phys. Rev. Lett.* 112.3 (2014), p. 031802. DOI: 10.1103/PhysRevLett.112.031802. arXiv: 1310.5926 [astro-ph.CO].
- [128] J. C. Hardy et al. “The Essential Decay of Pandemonium: A Demonstration of Errors in Complex Beta-Decay Schemes”. In: *Physics Letters B* 71.2 (Nov. 21, 1977), pp. 307–310. ISSN: 0370-2693. DOI: 10.1016/0370-2693(77)90223-4.
- [129] W. K. Hastings. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. ISSN: 0006-3444. DOI: 10.2307/2334940. JSTOR: 2334940.
- [130] W. Haxton. “THE SOLAR NEUTRINO PROBLEM”. In: *Annual Review of Astronomy and Astrophysics* 33.1 (Sept. 1995), pp. 459–503. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev.aa.33.090195.002331. arXiv: hep-ph/9503430.

- [131] A. C. Hayes et al. “Systematic Uncertainties in the Analysis of the Reactor Neutrino Anomaly”. In: *Physical Review Letters* 112.20 (May 20, 2014). issn: 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.112.202501. arXiv: 1309.4146.
- [132] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: (Dec. 10, 2015). arXiv: 1512.03385 [cs].
- [133] Kaiming He et al. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”. In: *CoRR* abs/1502.01852 (2015).
- [134] Donald Olding Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, Jan. 1, 1949. 368 pp. isbn: 978-0-471-36727-7.
- [135] D. Heck et al. “CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers”. In: (1998).
- [136] Hedin A. E. “Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere”. In: *Journal of Geophysical Research: Space Physics* 96.A2 (Sept. 20, 2012), pp. 1159–1172. issn: 0148-0227. doi: 10.1029/90JA02125.
- [137] L. G. Henyey and J. L. Greenstein. “Diffuse Radiation in the Galaxy”. In: *Astrophys.J.* 93 (1941), pp. 70–83. doi: 10.1086/144246.
- [138] V. F. Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten”. In: *Physikalische Zeitschrift* 13 (1912), pp. 1084–1091.
- [139] David E Hill. “Targeted Jumps by Salticid Spiders (Araneae, Salticidae, Phidippus)”. en. In: *Peckhamia Epublications* (2006), p. 28.
- [140] K. Hirata et al. “Observation of a Neutrino Burst from the Supernova SN1987A”. In: *Physical Review Letters* 58.14 (Apr. 6, 1987), pp. 1490–1493. doi: 10.1103/PhysRevLett.58.1490.
- [141] Chiu Man Ho and Robert J. Scherrer. “Sterile Neutrinos and Light Dark Matter Save Each Other”. In: *Phys. Rev. D* 87.6 (2013), p. 065016. doi: 10.1103/PhysRevD.87.065016. arXiv: 1212.1689 [hep-ph].
- [142] J. R. Hoerandel. “On the Knee in the Energy Spectrum of Cosmic Rays”. In: *Astroparticle Physics* 19.2 (May 2003), pp. 193–220. issn: 09276505. doi: 10.1016/S0927-6505(02)00198-6. arXiv: astro-ph/0210453.

- [143] M. Honda et al. “Calculation of the Flux of Atmospheric Neutrinos”. In: *Physical Review D* 52.9 (Nov. 1, 1995), pp. 4985–5005. ISSN: 0556-2821. DOI: 10.1103/PhysRevD.52.4985. arXiv: hep-ph/9503439.
- [144] M. Honda et al. “New Calculation of the Atmospheric Neutrino Flux in a Three-Dimensional Scheme”. In: *Phys. Rev. D* 70 (4 Aug. 2004), p. 043008. DOI: 10.1103/PhysRevD.70.043008.
- [145] Kurt Hornik. “Approximation Capabilities of Multilayer Feedforward Networks”. In: *Neural Networks* 4.2 (Jan. 1, 1991), pp. 251–257. ISSN: 0893-6080. DOI: 10.1016/0893-6080(91)90009-T.
- [146] Patrick Huber. “The 5 MeV Bump - a Nuclear Whodunit Mystery”. In: *Physical Review Letters* 118.4 (Jan. 27, 2017). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.118.042502. arXiv: 1609.03910.
- [147] Christina M. Ignarra. “Sterile Neutrino Searches in Mini-BooNE and MicroBooNE”. MIT, Cambridge, 2014.
- [148] H.-Th Janka. “Neutrino Emission from Supernovae”. In: (Feb. 28, 2017). arXiv: 1702.08713 [astro-ph, physics:hep-ph, physics:nucl-th].
- [149] E. T. Jaynes. “Information Theory and Statistical Mechanics”. In: *BRANDEIS UNIVERSITY SUMMER INSTITUTE. LECTURES IN THEORETICAL PHYSICS*. Ed. by W. Ford. W. A. Benjamin, Inc, pp. 181–218.
- [150] E. T. Jaynes. “Prior Probabilities”. In: *IEEE Transactions on Systems Science and Cybernetics* 4.3 (Sept. 1968), pp. 227–241. ISSN: 0536-1567. DOI: 10.1109/TSSC.1968.300117.
- [151] F Joliot and I Curie. “Un Nouveau Type de Radioactivite”. In: *C. R. Acad. Sci. Paris, , , 559 (1934)* 198 (1934), p. 254.
- [152] B. J. P. Jones. “Dynamical Pion Collapse and the Coherence of Conventional Neutrino Beams”. In: *Physical Review D* 91.5 (Mar. 4, 2015). ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.91.053002. arXiv: 1412.2264.
- [153] Benjamin J. P. Jones. “Sterile Neutrinos in Cold Climates”. MIT, 2015.
- [154] Benjamin J. P. Jones. “Sterile Neutrinos in Cold Climates”. PhD thesis. MIT, 2015. DOI: 10.2172/1221354.

- [155] KARMEN Collaboration et al. “Upper limits for neutrino oscillations $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ from muon decay at rest”. In: *Physical Review D* 65.11 (June 5, 2002), p. 112001. doi: 10.1103/PhysRevD.65.112001.
- [156] F. Kaether et al. “Reanalysis of the GALLEX Solar Neutrino Flux and Source Experiments”. In: *Physics Letters B* 685.1 (Feb. 2010), pp. 47–54. ISSN: 03702693. doi: 10.1016/j.physletb.2010.01.030. arXiv: 1001.2731.
- [157] Lewis D. Kaplan. “Inference of Atmospheric Structure from Remote Radiation Measurements*”. In: *JOSA* 49.10 (Oct. 1, 1959), pp. 1004–1007. doi: 10.1364/JOSA.49.001004.
- [158] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: (Dec. 22, 2014). arXiv: 1412.6980 [cs].
- [159] S. Kleinfelder. “Gigahertz Waveform Sampling and Digitization Circuit Design and Implementation”. In: *IEEE Transactions on Nuclear Science* 50.4 (Aug. 2003), pp. 955–962. ISSN: 0018-9499. doi: 10.1109/TNS.2003.815137.
- [160] K. Kodama et al. “Observation of Tau Neutrino Interactions”. In: *Physics Letters B* 504.3 (Apr. 12, 2001), pp. 218–224. ISSN: 0370-2693. doi: 10.1016/S0370-2693(01)00307-0.
- [161] Sacha E. Kopp. “Accelerator Neutrino Beams”. In: *Physics Reports* 439.3 (Feb. 2007), pp. 101–159. ISSN: 03701573. doi: 10.1016/j.physrep.2006.11.004. arXiv: physics/0609129.
- [162] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS’12. USA: Curran Associates Inc., 2012, pp. 1097–1105.
- [163] Yann LeCun et al. “Handwritten Digit Recognition with a Back-Propagation Network”. In: *Advances in Neural Information Processing Systems 2*. Ed. by D. S. Touretzky. Morgan-Kaufmann, 1990, pp. 396–404.
- [164] Samuel K. Lee et al. “Evidence for Unresolved Gamma-Ray Point Sources in the Inner Galaxy”. In: *Physical Review Letters* 116.5 (Feb. 4, 2016). ISSN: 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.116.051103. arXiv: 1506.05124.

- [165] Martin Lemoine, Kumiko Kotera, and Jérôme Pétri. “On Ultra-High Energy Cosmic Ray Acceleration at the Termination Shock of Young Pulsar Winds”. In: *Journal of Cosmology and Astroparticle Physics* 2015.07 (July 13, 2015), pp. 016–016. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2015/07/016. arXiv: 1409.0159.
- [166] Shirley Weishi Li and John F. Beacom. “Spallation Backgrounds in Super-Kamiokande Are Made in Muon-Induced Showers”. In: *Phys.Rev. D91* (May 5, 2015), p. 105005. DOI: 10.1103/PhysRevD.91.105005.
- [167] Tzu-Mao Li. *HAD*.
- [168] Manfred Lindner, Werner Rodejohann, and Xun-Jie Xu. “Sterile Neutrinos in the Light of IceCube”. In: (2015). arXiv: 1510.00666 [hep-ph].
- [169] Seppo Linnainmaa. “Taylor Expansion of the Accumulated Rounding Error”. In: *BIT Numerical Mathematics* 16.2 (June 1, 1976), pp. 146–160. ISSN: 0006-3835, 1572-9125. DOI: 10.1007/BF01931367.
- [170] Seppo Linnainmaa. “The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors”. Master’s Thesis. Univ. Helsinki, 1970.
- [171] Pingyu Liu. “A New Phase Function Approximating to Mie Scattering for Radiative Transport Equations”. In: *Physics in Medicine & Biology* 39.6 (1994), p. 1025. ISSN: 0031-9155. DOI: 10.1088/0031-9155/39/6/008.
- [172] X.-Q. Lu et al. *A Proposal to Search for Neutrino Oscillations with High Sensitivity in the Appearance Channels $\nu_\mu \rightarrow \nu_e$ and Bar $\nu_\mu \rightarrow$ Bar ν_e* . LA-11842-P. Los Alamos National Lab., 1990.
- [173] Cecilia Lunardini and Soebur Razzaque. “High Energy Neutrinos from the Fermi Bubbles”. In: *Physical Review Letters* 108.22 (June 1, 2012). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.108.221102. arXiv: 1112.4799.
- [174] MINOS Collaboration et al. “Observation of Muon Neutrino Disappearance with the MINOS Detectors in the NuMI Neutrino Beam”. In: *Physical Review Letters* 97.19 (Nov. 8, 2006), p. 191801. DOI: 10.1103/PhysRevLett.97.191801.

- [175] MINOS Collaboration et al. “Study of Muon Neutrino Disappearance Using the Fermilab Main Injector Neutrino Beam”. In: *Physical Review D* 77.7 (Apr. 4, 2008), p. 072002. doi: 10.1103/PhysRevD.77.072002.
- [176] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. New York, NY, USA: Cambridge University Press, 2002. ISBN: 978-0-521-64298-9.
- [177] Faisal Mahmood, Richard Chen, and Nicholas J. Durr. “Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training”. In: (Nov. 17, 2017). arXiv: 1711.06606 [cs].
- [178] K. B. M. Mahn and others. “Dual Baseline Search for Muon Neutrino Disappearance at $0.5 \text{ eV}^2 < \Delta m^2 < 40 \text{ eV}^2$ ”. In: *Phys. Rev. D* 85 (2012), p. 032007. doi: 10.1103/PhysRevD.85.032007. arXiv: 1106.5685 [hep-ex].
- [179] Dmitry Malyshev and David W. Hogg. “Statistics of Gamma-Ray Point Sources below the Fermi Detection Limit”. In: *The Astrophysical Journal* 738.2 (Sept. 10, 2011), p. 181. ISSN: 0004-637X, 1538-4357. doi: 10.1088/0004-637X/738/2/181. arXiv: 1104.0010.
- [180] Karl Mannheim. “Possible Production of High-Energy Gamma Rays from Proton Acceleration in the Extragalactic Radio Source Markarian 501”. In: *Science* 279.5351 (Jan. 30, 1998), pp. 684–686. ISSN: 00368075, 10959203. doi: 10.1126/science.279.5351.684. arXiv: astro-ph/9803241.
- [181] M. A. Markov. “On High Energy Neutrino Physics”. In: (1960), pp. 578–581.
- [182] Warren S. McCulloch and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1, 1943), pp. 115–133. ISSN: 0007-4985, 1522-9602. doi: 10.1007/BF02478259.
- [183] Kevin McFarland. “Neutrino Interactions”. In: (Apr. 24, 2008). arXiv: 0804.3899 [hep-ex].
- [184] G. Mention et al. “Reactor Antineutrino Anomaly”. In: *Physical Review D* 83.7 (Apr. 29, 2011), p. 073006. doi: 10.1103/PhysRevD.83.073006.
- [185] *MessagePack*. URL: <https://msgpack.org/> (visited on 05/07/2018).

- [186] Marvin Minsky, Seymour A. Papert, and Léon Bottou. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Sept. 15, 2017. 317 pp. ISBN: 978-0-262-53477-2.
- [187] Siddharth Mishra-Sharma, Nicholas L. Rodd, and Benjamin R. Safdi. “NPTFit: A Code Package for Non-Poissonian Template Fitting”. In: *The Astronomical Journal* 153.6 (May 19, 2017), p. 253. ISSN: 1538-3881. DOI: 10.3847/1538-3881/aa6d5f. arXiv: 1612.03173.
- [188] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN: 978-0-262-01802-9.
- [189] Jerzy Neyman and Egon Sharpe Pearson. “IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses”. In: *Phil. Trans. R. Soc. Lond. A* 231.694-706 (Feb. 16, 1933), pp. 289–337. ISSN: 0264-3952, 2053-9258. DOI: 10.1098/rsta.1933.0009.
- [190] V. A. Novikov et al. “Theory of Z Boson Decays”. In: *Reports on Progress in Physics* 62.9 (Sept. 1, 1999), pp. 1275–1332. ISSN: 0034-4885, 1361-6633. DOI: 10.1088/0034-4885/62/9/201. arXiv: hep-ph/9906465.
- [191] Kyoung-Su Oh and Keechul Jung. “GPU Implementation of Neural Networks”. In: *Pattern Recognition* 37.6 (June 1, 2004), pp. 1311–1314. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2004.01.013.
- [192] Yutaka Ohira, Shota Kisaka, and Ryo Yamazaki. “Pulsar Wind Nebulae inside Supernova Remnants as Cosmic-Ray PeVatrons”. In: (Feb. 20, 2017). arXiv: 1702.05866 [astro-ph].
- [193] Edward T. Olsen. *AIRS/AMSU/HSB Version 6 Data Release User Guide*. Nov. 27, 2017.
- [194] Sergey Ostapchenko. “Monte Carlo Treatment of Hadronic Interactions in Enhanced Pomeron Scheme: I. QGSJET-II Model”. In: *Physical Review D* 83.1 (Jan. 25, 2011). ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.83.014018. arXiv: 1010.1869.
- [195] Irene Di Palma, Dafne Guetta, and Elena Amato. “Revised Predictions of Neutrino Fluxes from Pulsar Wind Nebulae”. In: *The Astrophysical Journal* 836.2 (2017), p. 159. ISSN: 0004-637X. DOI: 10.3847/1538-4357/836/2/159.
- [196] Stephen Parke and Mark Ross-Lonergan. In: (2016). private communication.

- [197] Stephen Parke and Mark Ross-Lonegan. “Unitarity and the Three Flavour Neutrino Mixing Matrix”. In: (2015). arXiv: 1508.05095 [hep-ph].
- [198] C. Patrignani et al. “Review of Particle Physics”. In: *Chin.Phys.* C40 (Oct. 3, 2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.
- [199] Wolfgang Pauli. *Pauli Letter Collection : Letter to Lise Meitner*. Oct. 2, 2000. URL: <http://cds.cern.ch/record/83282> (visited on 03/30/2018).
- [200] Maxim Perelstein. “Introduction to Collider Physics”. In: (Mar. 2011), pp. 421–486. DOI: 10.1142/9789814327183_0008. arXiv: 1002.0274 [hep-ph].
- [201] Picone J. M. et al. “NRLMSISE-00 Empirical Model of the Atmosphere: Statistical Comparisons and Scientific Issues”. In: *Journal of Geophysical Research: Space Physics* 107.A12 (Dec. 24, 2002), SIA 15–1. ISSN: 0148-0227. DOI: 10.1029/2002JA009430.
- [202] Andrzej Pokraka and Andrzej Czarnecki. “Positronium Decay into a Photon and Neutrinos”. In: *Physical Review D* 94.11 (Dec. 21, 2016), p. 113012. DOI: 10.1103/PhysRevD.94.113012. arXiv: 1611.07645.
- [203] Simon Premože, Michael Ashikhmin, and Peter Shirley. “Path Integration for Light Transport in Volumes”. In: *Proceedings of the 14th Eurographics Workshop on Rendering*. EGRW '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003, pp. 52–63. ISBN: 978-3-905673-03-6.
- [204] Simon Premože et al. “Practical Rendering of Multiple Scattering Effects in Participating Media”. In: *Proceedings of the Fifteenth Eurographics Conference on Rendering Techniques*. EGSR'04. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2004, pp. 363–374. ISBN: 978-3-905673-12-8. DOI: 10.2312/EGWR/EGSR04/363-374.
- [205] Rajat Raina, Anand Madhavan, and Andrew Y. Ng. “Large-Scale Deep Unsupervised Learning Using Graphics Processors”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09. New York, NY, USA: ACM, 2009, pp. 873–880. ISBN: 978-1-60558-516-1. DOI: 10.1145/1553374.1553486.
- [206] J. Ranft. “Dual Parton Model at Cosmic Ray Energies”. In: *Physical Review D* 51.1 (Jan. 1, 1995), pp. 64–84. DOI: 10.1103/PhysRevD.51.64.

- [207] Sylvia Richardson and Peter J. Green. “On Bayesian Analysis of Mixtures with an Unknown Number of Components (with Discussion)”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59.4 (Jan. 1, 1997), pp. 731–792. ISSN: 1467-9868. DOI: 10.1111/1467-9868.00095.
- [208] David Rolnick and Max Tegmark. “The Power of Deeper Networks for Expressing Natural Functions”. In: (May 15, 2017). arXiv: 1705.05502 [cs, stat].
- [209] Jorge C. Romão and João P. Silva. “A Resource for Signs and Feynman Diagrams of the Standard Model”. In: *International Journal of Modern Physics A* 27.26 (Oct. 17, 2012), p. 1230025. ISSN: 0217-751X. DOI: 10.1142/S0217751X12300256. arXiv: 1209.6213.
- [210] F. Rosenblatt. *The Perceptron, a Perceiving and Recognizing Automaton Project Para*. Cornell Aeronautical Laboratory, 1957. book.
- [211] Bruno Rossi. “On the Magnetic Deflection of Cosmic Rays”. In: *Physical Review* 36.3 (Aug. 1, 1930), pp. 606–606. DOI: 10.1103/PhysRev.36.606.
- [212] Sebastian Ruder. “An Overview of Gradient Descent Optimization Algorithms”. In: (Sept. 15, 2016). arXiv: 1609.04747 [cs].
- [213] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning Representations by Back-Propagating Errors”. In: *Nature* 323.6088 (Oct. 1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0.
- [214] Krzysztof P. Rykaczewski. “Viewpoint: Conquering Nuclear Pandemonium”. In: *Physics* 3 (Nov. 8, 2010).
- [215] Andrew M. Saxe, James L. McClelland, and Surya Ganguli. “Exact Solutions to the Nonlinear Dynamics of Learning in Deep Linear Neural Networks”. In: (Dec. 20, 2013). arXiv: 1312.6120 [cond-mat, q-bio, stat].
- [216] David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. “Maps of Dust IR Emission for Use in Estimation of Reddening and CMBR Foregrounds”. In: *The Astrophysical Journal* 500.2 (June 20, 1998), pp. 525–553. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/305772. arXiv: astro-ph/9710327.
- [217] G. Sebestyen. “Review of ‘Learning Machines’ (Nilsson, Nils J.; 1965)”. In: *IEEE Transactions on Information Theory* 12.3 (July 1966), pp. 407–407. ISSN: 0018-9448. DOI: 10.1109/TIT.1966.1053912.

- [218] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 Physics and Manual”. In: *Journal of High Energy Physics* 2006.05 (May 9, 2006), pp. 026–026. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175.
- [219] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Computer Physics Communications* 178.11 (June 2008), pp. 852–867. ISSN: 00104655. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820.
- [220] John Skilling. “Nested Sampling for General Bayesian Computation”. In: *Bayesian Analysis* 1.4 (Dec. 2006), pp. 833–859. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/06-BA127.
- [221] Christian Spiering. “Towards High-Energy Neutrino Astronomy. A Historical Review”. In: *The European Physical Journal H* 37.3 (Aug. 2012), pp. 515–565. ISSN: 2102-6459, 2102-6467. DOI: 10.1140/epjh/e2012-30014-2. arXiv: 1207.4952.
- [222] I. E. Stockdale et al. “Limits on Muon-Neutrino Oscillations in the Mass Range $30 < \Delta m^2 < 1000 \text{ eV}^2 / \text{c}^4$ ”. In: *Physical Review Letters* 52.16 (Apr. 16, 1984), pp. 1384–1388. DOI: 10.1103/PhysRevLett.52.1384.
- [223] Meng Su, Tracy R. Slatyer, and Douglas P. Finkbeiner. “Giant Gamma-Ray Bubbles from Fermi-LAT: AGN Activity or Bipolar Galactic Wind?” In: *The Astrophysical Journal* 724.2 (Dec. 1, 2010), pp. 1044–1082. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/724/2/1044. arXiv: 1005.5480.
- [224] Tobin David C. et al. “Atmospheric Radiation Measurement Site Atmospheric State Best Estimates for Atmospheric Infrared Sounder Temperature and Water Vapor Retrieval Validation”. In: *Journal of Geophysical Research: Atmospheres* 111.D9 (Mar. 16, 2006). ISSN: 0148-0227. DOI: 10.1029/2005JD006103.
- [225] *U.S. Standard Atmosphere*. 1976.
- [226] Eric Veach and Leonidas J. Guibas. “Metropolis Light Transport”. In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '97. New York, NY, USA: ACM

- Press/Addison-Wesley Publishing Co., 1997, pp. 65–76. ISBN: 978-0-89791-896-1. DOI: 10.1145/258734.258775.
- [227] Yu. M. Volin and G. M. Ostrovskii. “Automatic Computation of Derivatives with the Use of the Multilevel Differentiating Technique—1. Algorithmic Basis”. In: *Computers & Mathematics with Applications* 11.11 (Nov. 1, 1985), pp. 1099–1114. ISSN: 0898-1221. DOI: 10.1016/0898-1221(85)90188-9.
- [228] Christopher Weaver. “Evidence for Astrophysical Muon Neutrinos from the Northern Sky”. University of Wisconsin–Madison, 2015.
- [229] B. Widrow and M.E. Hoff Jr. “Adaptive Switching Circuits”. In: *IRE WESCON Convention Record 4* (Aug. 1960), pp. 96–104.
- [230] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (Mar. 1938), pp. 60–62. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177732360.
- [231] Katherine Woodruff and the MicroBooNE Collaboration. “Automated Proton Track Identification in MicroBooNE Using Gradient Boosted Decision Trees”. In: (Oct. 2, 2017). arXiv: 1710.00898 [hep-ex, physics:physics].
- [232] Donglian Xu. “Search for Astrophysical Tau Neutrinos in Three Years of IceCube Data”. Alabama U., 2015.
- [233] Neha Yadav, Anupam Yadav, and Manoj Kumar. *An Introduction to Neural Network Methods for Differential Equations*. SpringerBriefs in Computational Intelligence. Springer Netherlands, 2015. ISBN: 978-94-017-9815-0.
- [234] GT Zatsepin and VA Kuz'min. “Neutrino Production in the Atmosphere”. In: *Soviet Physics JETP* 14.6 (1962).
- [235] V. I. Zatsepin and N. V. Sokolskaya. “Three Component Model of Cosmic Ray Spectra from 10 GeV to 100 PeV”. In: *Astronomy & Astrophysics* 458.1 (Oct. 2006), pp. 1–5. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361:20065108. arXiv: astro-ph/0601475.
- [236] J. Zurada. *Introduction to Artificial Neural Systems*. St. Paul, MN, USA: West Publishing Co., 1992. ISBN: 978-0-314-93391-1.

- [237] The Fermi-LAT collaboration et al. “Detection of the Characteristic Pion-Decay Signature in Supernova Remnants”. In: *Science* 339.6121 (Feb. 15, 2013), pp. 807–811. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1231160. arXiv: 1302.3307.

Index

- 3 + N model, 66
 - oscillation probability, 66
- 3 + 1 model, 65
 - oscillation probability, 65
- 3 + 2 model, 67
- absorption, 91, 165
 - length, 91, 166
- accuracy, machine learning, 141
- activation function, 140
- active galactic nuclei, 114
- adjoint, 144
- after-pulse, 90, 151
- AI winter, 136
- AIRS, 105
- alternative hypothesis, 45
- AMANDA, 86
- Analog Transient Waveform Digitizer, 90
- angular emission distribution, 169
- angular scattering distribution, 91, 166, 169
- anode, 90
- appearance, 62
- artificial neural network, 135
- astrophysical beam dump, 115
- atmosphere, Earth, 103
 - balloon measurements, 105
 - satellite measurements, 105
- atmospheric neutrino
 - conventional, 100
 - prompt, 100
- atmospheric shower, 100, 112
 - decay coefficients, 102
 - decay length, 102
 - interaction coefficients, 102
 - interaction length, 102
- automatic differentiation, 143

- edge-pushing, 173
- forward, 143
- reverse, 144

- backpropagation, 146
- baryon resonance, 33
- base-line
 - long, 62
 - short, 61
- batches, machine learning, 156
- Bayes' factor, 49, 129
- Bayes' theorem, 48
- Beer's law, 91
- best fit, 44
- beta decay, 24, 29
 - inverse, 30
- bias term, machine learning, 140
- birth/death move, 170
- bispherical coordinates, 170
- boost factor, 89, 101
- broken power-law, 121
- Bugey, 64

- cascade, 88, 150
- channel
 - s, 23
 - t, 22
- channel, machine learning, 154
- charged current, 23
- Cherenkov radiation, 37
- χ^2 , 45
 - saturated Poisson, 70
 - statistic, 70
- classification, 137, 141
- confidence region, 46
 - from samples, 72
- convolutional neural network, 136, 152, 153
- cosmic neutrino background, 29
- cosmic ray, 112
 - acceleration, 114
 - ankle, 113
 - knee, 113
- cosmic rays, 111
- coverage, 46
- credible region, 49

- from samples, 71
- highest posterior density, 49
- cross-correlation, 153
- cross-entropy, 42
- crossing symmetry, 30
- decay
 - muon, 32
 - pion, 32
- decay at rest, 34
- deep inelastic scattering, 35
- deep learning, 136, 141
- deep neural network, 141
- DeepCore, 87
- degrees of freedom, 45
- $\Delta\chi^2$, 46
- differential source count function, 119
- Digital Optical Module, 89
 - photo-multiplier tube, 90
- Digital Optical Module (DOM), 87
- digitiser, 90
- disappearance, 63
- double pulse, 150
- down-going, 123
- DUMAND, 85
- dynode, 90
- effective area, 126
- electron capture, 30
- entropy
 - information, 42
 - continuous, 43
 - statistical, 43
- epoch, machine learning, 162
- exposure, *see* effective area
- feature, machine learning, 153
- Fermi acceleration, 114
- Fermi bubble, 116
- flasher LED, 93, 165
- flavour, 24
- flavour ratio, astrophysical, 115, 117
- fractional flux distribution, 127
- fully connected neural network, 152
- GALLEX, 64

- generating function, 120
- Glashow resonance, 36
- global fit, sterile neutrino, 67
- gradient decent, 138, 142
- graphical processing unit, 137
- gyro-radius, 113

- healpix, 125
- Heney-Greenstein, 92, 173
- hidden layer, 140
- hole ice, 94
- Homestake, 31

- ice model, 92
- IceCube, 86, 116
 - angular resolution, 127
 - string, 87
- IceTop, 87
- importance sampling, 52
- infrared sounding, 105
- inverse CDF theorem, 52

- kernel, convolution, 153
- KL divergence, 43
 - continuous, 43

- late-pulse, 90, 151
- layer, neural network, 139
- learning
 - supervised, 136
- learning rate, 142
 - adaptive, 156
- likelihood
 - function, 44
 - normally distributed, 69
 - Poisson distributed, 69
 - ratio, 45, 46
- live time, 123
- looks linear initialisation, 160
- loss function, 138
 - cross-entropy, 142
- LSND, 62

- machine learning, 135
- marginal likelihood, 48
- Markov chain Monte-Carlo, 53

- acceptance probability, 54
- burn-in, 54
- detailed balance, 54
- ensemble, 55
- Metropolis-Hastings, 54
- parallelisation, 68
- replica exchange, 56
- temperature, 56
- mass splitting, 27
 - atmospheric, 64
 - solar, 64
- matrix cascade equation, 103
- matter effect, 77
- maximum likelihood, 44
- Metropolis light transport, 168
- Mie scattering, 92
- MiniBooNE, 62
- minimum ionising, 37
- mixing matrix
 - 3 + 1 model, 65
 - CKM, 24
 - PMNS, 25
- momentum, gradient decent, 156
- MSW, *see* matter effect
- multi-messenger astronomy, 111
- mutual information, 47

- nats, 42
- near detector, 63
- nested sampling, 58, 129
- neutral current, 23
- neutrino astronomy, 116
- neutrino capture, 31
- neutrino horn, 33
- neutrino oscillation
 - CP violation, 27
 - decoherence, 28
 - general formula, 27
 - maximum, 61
- neutrino sources
 - accelerator, 32, 33
 - astrophysical, 34, 115
 - atmospheric, 34, 124
 - geo, 31
 - isotope, 30, 64

- isotope (accelerator), 34
 - reactor, 30, 63
 - solar, 31
 - supernova, 31, 114
- neutrino-lepton scattering, 36
- non-parametric, 57
- Non-Poissonian Template Fitting, 119
 - likelihood, 122
- null hypothesis, 45
- numerical differentiation, 142

- OneWeight, 95
- optical depth, 169
 - absorption, 91, 166
 - scattering, 91, 166
- orthogonal initialisation, 161
- over-training, 146

- Pandel function, 96
- particle propagator, 95
- path integration, numerical, 168
- path tracer, 166
- path tracing, 95
- perceptron, 135
- phase function, 91
- photo-cathode, 90
- photo-electron, 90
- photo-multiplier tube, 90
- point sources, 18, 116
- point spread function, 127
- point-wise information, 43
- point-wise mutual information, 48
- pooling, machine learning, 155
- posterior, 48
- pre-pulse, 90, 151
- prediction function, 44
- prior, 48
- profiling, 47
- pulsar wind nebula, 114

- quasi-elastic scattering, 32

- reactor anomaly, 63
- reconstruction
 - cascade, 96
 - track, 96

rectified linear unit, 155
regularisation, 146
rejection sampling, 53
residual neural network, 160
reversible jump MCMC, 57, 168
 acceptance probability, 58
 transition function, 57
risk, 44

SAGE, 64

sample
 equally-weighted, 51
 re-sample, 52
 re-weighting, 52
 weighted, 50

scattering, 165
 length, 91, 166

scrambling, 124

short base-line approximation, 65

shower
 critical energy, 38
 electromagnetic, 38
 hadronic, 35

significance level, 45

simplified Liu, 92, 173

sky, northern and southern, 123

slant depth, 103

SPICE
 ³, 94
 Lea, 93
 Mie, 92

starburst galaxy, 114

sterile neutrino, 65

stochastic gradient decent, 156

strided convolution, 154

supernova remnant, 114

surprise, 44

symbolic differentiation, 142

tau decay, 39, 150

tau regeneration, 94

template, 121

test statistic, 45

testing set, 162

topology, event, 89

track, 87
training, 136
training set, 162

up-going, 123

virtual particle, 21

weight, machine learning, 140
 tying, 153
weights (machine learning), 135
Wilks' theorem, 45

Z decay width, 23