

Initial Report of the DUNE Near Detector Task Force

J. Sinclair¹, L. Escudero², D. Cherdack³, T. Alion⁴, S. J. Brice⁴, L. Fields⁴, R. Hatcher⁴,
S. Lockwitz⁴, B. Rebel⁴, T. Yang⁴, D. Brailsford⁵, C. Andreopoulos⁶, G. Christodoulou⁶,
S. Dennis⁶, K. Mahn⁷, M. Bass⁸, and J. Martín-Albo⁸

¹University of Bern

²University of Cambridge

³Colorado State University

⁴Fermi National Accelerator Laboratory

⁵Lancaster University

⁶University of Liverpool

⁷Michigan State University

⁸Oxford University

Contents

| | | |
|-----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Charge | 1 |
| 1.2 | Personnel | 1 |
| 1.3 | Timeline | 1 |
| 2 | Overview | 2 |
| 3 | Neutrino Fluxes and Uncertainties | 3 |
| 4 | Neutrino Cross-Sections and Uncertainties | 5 |
| 5 | External Backgrounds | 9 |
| 5.1 | Cosmic Ray Induced Particles | 9 |
| 5.2 | Particles Induced by Neutrino Interaction in the Rock | 10 |
| 6 | Near Detector Samples | 11 |
| 6.1 | The Fine Grained Tracker | 11 |
| 6.1.1 | Simulation | 11 |
| 6.1.2 | Reconstruction | 13 |
| 6.2 | The High Pressure Gaseous Argon TPC | 14 |
| 6.2.1 | Simulation | 14 |
| 6.2.2 | Reconstruction | 15 |
| 6.3 | The Liquid Argon TPC | 16 |
| 6.3.1 | Simulation | 16 |
| 6.3.2 | Reconstruction | 17 |
| 7 | Near Detector Systematics | 17 |
| 8 | The VALOR Near Detector Fits: Constraining the Systematic Uncertainties | 18 |
| 8.1 | Parameter Estimation | 18 |
| 8.2 | Binning | 20 |
| 8.3 | Near-Detector Fitting Results | 21 |
| 9 | The Far Detector Samples | 22 |
| 9.1 | Simulation | 22 |
| 9.2 | Reconstruction | 23 |
| 10 | The Far Detector Fit | 23 |
| 10.1 | Overview | 23 |
| 10.2 | Systematic Uncertainties | 24 |
| 10.2.1 | Inputs from VALOR (flux and cross section) | 24 |
| 10.2.2 | Detector systematics | 24 |
| 10.3 | Fit Results | 24 |
| 10.3.1 | Results from the LOAF fits | 24 |
| 10.3.2 | The VALOR fits | 27 |
| 11 | The Figures of Merit | 27 |
| 11.1 | FOMs: The ability of the detectors to enhance the oscillation analyses | 27 |
| 11.2 | FOMs: The performance of the detectors in the beam | 28 |
| 11.3 | FOMs: The ability to do physics with the near detectors | 28 |
| 12 | Bibliography | 29 |

1 Introduction

1.1 Charge

The Near Detector Task Force of the DUNE collaboration started work in September of 2015 with the following charge from the collaboration spokespeople

- Develop GEANT4 simulations of the reference design near detector and possible alternatives
- Perform a full end-to-end simulation connecting the measurements in the near detector to the far detector systematics using, for example, the VALOR framework
- Evaluate the potential benefits of augmenting the reference design with ...
 - a LAr-TPC
 - the use of a High Pressure Gaseous TPC
- Produce a first report on their findings to the DUNE Technical Board by September 2016 and a final report by March 2017.

This document is the September 2016 Initial Report mandated by the last bullet of the charge.

1.2 Personnel

Responsibility for the work of the task force is divided into the following points of contact

| | |
|---|--|
| Leader and Deputies | Steve Brice, Daniel Cherdack, and Kendall Mahn |
| Infrastructure | Robert Hatcher |
| Flux | Laura Fields |
| Cross-Section Models and Systematics | Lorena Escudero |
| Fine Grained Tracker | Tyler Alion |
| Liquid Argon TPC | Sarah Lockwitz and James Sinclair |
| High Pressure Gaseous Argon TPC | Justo Martín-Albo |
| VALOR | Steve Dennis and Costas Andreopoulos |
| Far Detector | Tingjun Yang and Tyler Alion |
| Far Detector Fit | Daniel Cherdack |
| Figures of Merit | Brian Rebel |

1.3 Timeline

The work of the task force has been driven by a series of "Run Throughs", one every four months, where the complete processing chain is exercised from flux and cross-section calculation through near detector simulation and the constraining of systematics to the far detector simulation and final fitting. The processing chain is described in more detail in 2. The figure below shows the timeline for run throughs and how they match with collaboration meetings and the production of initial and final reports.

The 1st Run Through of the processing chain occurred in January 2016 was designed to exercise all links in the chain, but in a stripped down and corner cutting way. The idea was to make sure the chain works and to catch any fundamental issues as soon as possible. The 1st Run Through did not have physics content in its output. The 2nd Run Through occurred prior to the South Dakota collaboration meeting in April 2016 and upgraded the physics and had some physics-useful output. The 3rd Run Through brought in many more details of the physics with the aim of a much more defensible physics output and is the basis of this initial report. The 4th (and final) Run Through is scheduled for December 2016, just prior to the DUNE collaboration meeting at CERN. The final report of the task force will be the result of a period of validation and physics studies following the 4th Run Through.



Figure 1: The timeline of the Near Detector Task Force work

2 Overview

The processing chain starts with the beam Monte Carlo generating neutrino fluxes for both near and far locations and for all 4 relevant neutrino species (electron, muon, and their anti-particles). Along with the fluxes a flux error matrix is produced that encapsulates the effects and correlations of all known sources of neutrino flux uncertainty. The flux generation step is summarized by the orange region of the process flow diagram of Fig 2.

The yellow region of the diagram above shows how the GENIE generator takes near and far neutrino fluxes and produces simulated neutrino interactions for each of the 3 near detector technologies as well as for the far detector. The cross-section uncertainties and their effects on interaction probabilities and the outgoing particles from the struck nucleus also originate from GENIE.

There are three candidate near detector technologies being considered - Fine Grained Tracker (FGT), High Pressure Gaseous Argon TPC (HPTPC), and Liquid Argon TPC (LArTPC). Each of these has a dedicated GEANT4 simulation and reconstruction and is represented by the horizontal pink region in the process flow diagram of Fig 2. Each technology also has its own systematic uncertainties. Each simulation reads the GENIE events that have been generated for it, runs the particles through a GEANT4 simulation of the detector and a simulation of the detector response, and then a reconstruction of the final state particles and processes. All three simulations share an identical geometry description for the detector hall, shaft, and surrounding rock; cosmics and rock events are also be generated with this geometry and inserted into the separate detector simulations.

The pink region on the right of the process flow diagram above represents the VALOR step where the near detector simulated data and the systematic uncertainties are input to a minimization process that outputs constrained uncertainties.

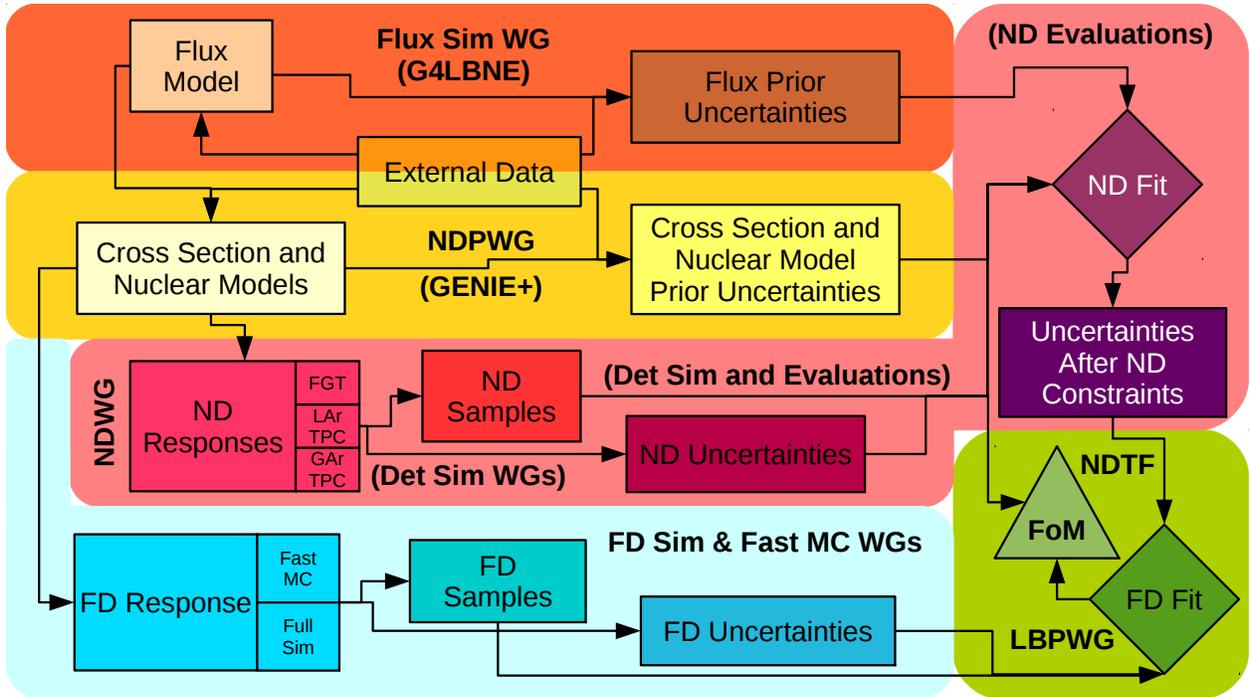


Figure 2: Overview of the processing chain

The cyan region of the process flow diagram describes the simulation of far detector data using fluxes derived from the same simulation that produced the near detector fluxes. The cross-section are also in common with those used in the near detector simulations. Far detector uncertainties and event samples are then fed into a final fit process.

The final fit extracts CP violation sensitivity from the far detector simulation and the constrained uncertainties output by VALOR. A number of figures of merit (FOMs) are extracted from this fit and combined with FOMs describing more basic near detector quantities like efficiencies, resolutions, and acceptances.

3 Neutrino Fluxes and Uncertainties

The neutrino flux simulation uses G4LBNF version v3r4p2 which is a configurable Geant4-based simulation of the DUNE beamline. For these fluxes, G4LBNF is configured to simulate the optimized beam described in the Beam Optimization Task Force Interim Report [1]. That beam includes a 200 cm long graphite target divided into fins that are each 13.4 mm wide by 20 mm high and separated by 0.2 mm. The primary proton beam momentum is 80 GeV, and has a RMS of 1.63 mm in both the x and y directions, and arrives in spills corresponding to $7.5e13$ protons per spill. Charged particles produced in the target are focused by three focusing horns. The target starts 8 cm downstream of the start of Horn 1 (the point commonly referred to as MCZERO), and the second (third) horn begins 2.63 (18.48) m downstream of MCZERO. The decay pipe is 194 meters long with a radius of 2 m. The hadron absorber and shielding around the target, horns and decay pipe are simulated but have little effect on the neutrino flux. Given current PIP II power estimates, we expect to receive 1.47×10^{21} protons on target per year at 80 GeV. The optimized and reference ν_μ fluxes are shown in Fig. 3.

The basic output of G4LBNF is an ntuple containing a branch for each neutrino produced in the beam simulation, formatted in the dk2nu common flux ntuple format. Output files corresponding to $1e8$ POT each in neutrino mode and antineutrino mode are output. A similar set of files corresponding to the reference design described in the 2015 DUNE CDR are also available.

The flux systematic uncertainties come in the form of a 208×208 covariance matrix, corresponding to two detector locations (near and far), two running modes (neutrino and antineutrino), and four neutrino species

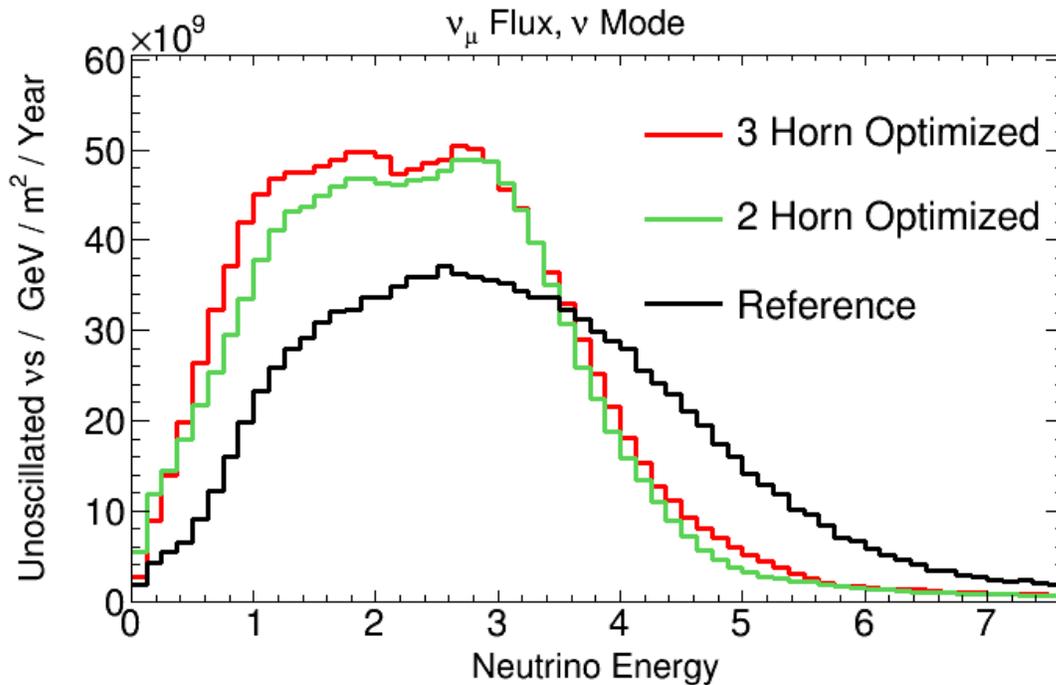


Figure 3: The optimized and reference ν_μ fluxes in forward horn current mode.

per mode (muon neutrinos, muon antineutrinos, electron neutrinos and electron antineutrinos). Nineteen energy bins are provided for the muon neutrino and muon antineutrinos fluxes, with bin edges at 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 12, 16, 20, 40, and 100 GeV. Seven bins are provided for the electron neutrino and electron antineutrino fluxes, with bin edges at 0, 2, 4, 6, 8, 10, 20, and 100 GeV. Matrices exist for both the optimized and reference beam options.

The covariance matrices are sums of two matrices which separately describe hadron production and beam alignment uncertainties. The hadron production uncertainties are estimated using the PPFX package developed by Leo Aliaga for MINERvA and extended to DUNE by Amit Bashyal. PPFX is documented in Leo Aliaga's thesis [2] while the expansion for Dune is described in a talk by Amit Bashyal [3]. The calculated hadron production uncertainties correspond to the flux at the center of the near and far detectors. For the far detector, this is an excellent approximation of uncertainties on flux over the entire detector, but this may not be the case for the near detector uncertainties. In particular, correlations between the near and far detector errors are likely overestimated as a result of this.

The alignment uncertainties are estimated by running simulations with each of the underlying uncertain parameters (horn currents, horn position, target positions, number of protons on target, etc) varied by one standard deviation from their nominal values. In some cases, we take the difference between the resulting flux and the nominal flux as the systematic uncertainty for that alignment parameter. In other cases, we run simulations for several other variations (for example two, four and six standard deviations) and fit the resulting fluxes to estimate the change in flux for one standard deviation. This work is described in more detail in two technical notes "LBNE Beam Alignment Tolerances and Systematic Uncertainties" [4] and <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1486> [5].

For all systematic uncertainties, underlying parameters are simultaneously changed for all neutrino fluxes, running modes and detector positions, so that correlations between the resulting uncertainties are properly calculated and propagated to the total covariance matrix. However, large statistical fluctuations are present in the case of the alignment parameter variations, which may cause under- or over- estimation of correlations.

The flux uncertainties over the 208 bins are shown in Fig. 4 and the full 208x208 correlation matrix is shown in Fig. 5

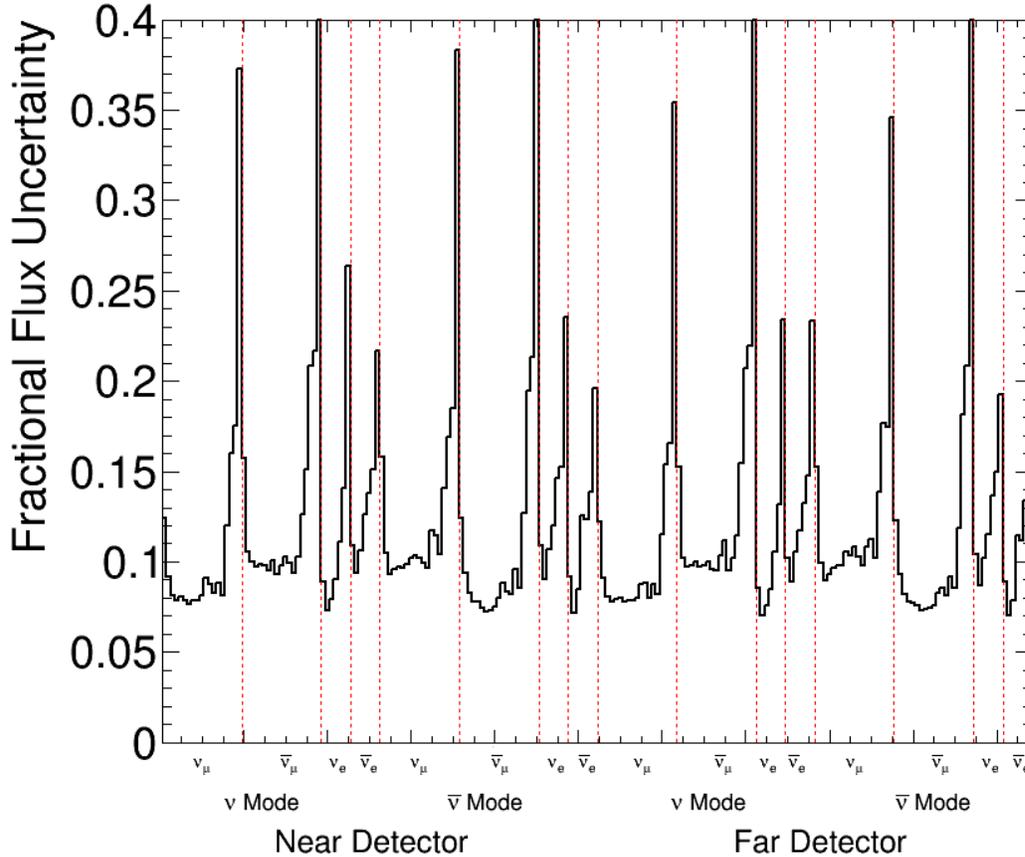


Figure 4: The *a priori* flux uncertainties. Details of the binning are given in the text of Sec. 3.

4 Neutrino Cross-Sections and Uncertainties

For the present analysis, a fairly comprehensive list of neutrino interaction systematics is implemented: we consider 43 neutrino interaction systematics parameterizing the uncertainties on a wide range of neutrino interaction modeling aspects relevant to DUNE, as listed below. For the DUNE oscillation sensitivity simulation and Near Detector optimization task, we want to give the VALOR fit sufficient freedom to vary the cross-section model and, also, ensure that it is the Near Detector data and not the priors that drive the DUNE systematics constraint. Therefore, appropriately conservative prior uncertainties are calculated by the VALOR group, using the GENIE event re-weighting tools. Multiple comparisons with external data are performed in collaboration with the GENIE group, in support of the systematic error assignments used in this analysis. Some of these data/MC comparisons are shown in the Appendix A of the VALOR DUNE technical note (available at DUNE-doc-1291 [6], and in preparation DUNE-doc-1712 [7]). Largely model-independent neutrino interaction systematics are used.

The following 43 neutrino interaction systematic parameters are included, defined in kinematical ranges chosen to ensure sufficient statistics in each bin:

0-2) 3 ν_μ CCQE systematics for the following true kinematic bins:

- $Q^2 < 0.2 \text{ GeV}^2$
- $0.2 \text{ GeV}^2 < Q^2 < 0.55 \text{ GeV}^2$
- $Q^2 > 0.55 \text{ GeV}^2$

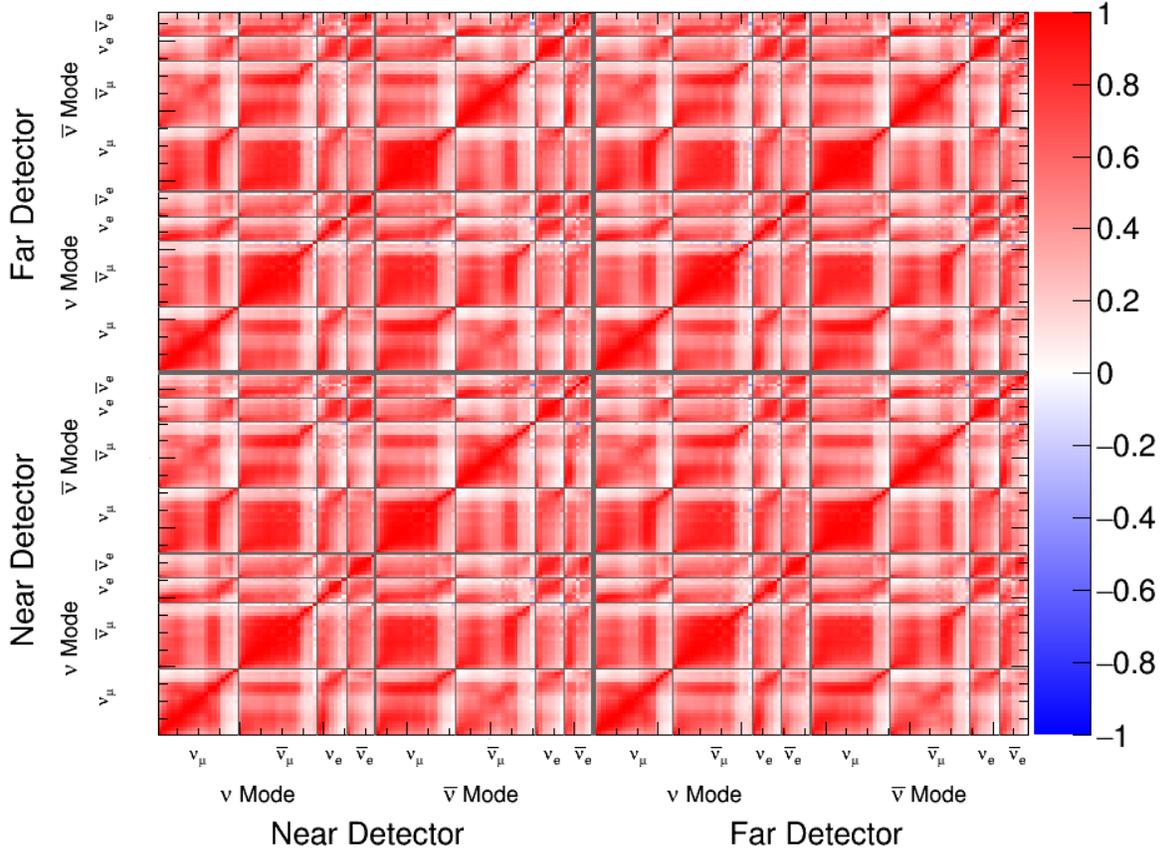


Figure 5: The *a priori* flux correlation matrix. Details of the binning are given in the text of Sec. 3.

- 3-5) $3 \bar{\nu}_\mu$ CCQE systematics for the following true kinematic bins:
- $Q^2 < 0.2 \text{ GeV}^2$
 - $0.2 \text{ GeV}^2 < Q^2 < 0.55 \text{ GeV}^2$
 - $Q^2 > 0.55 \text{ GeV}^2$
- 6) $1 \nu_\mu$ CC MEC systematic - 100% uncorrelated uncertainty added for now
- 7) $1 \bar{\nu}_\mu$ CC MEC systematic - 100% uncorrelated uncertainty added for now
- 8-10) $3 \nu_\mu$ CC $1\pi^\pm$ systematics for the following true kinematic bins:
- $Q^2 < 0.3 \text{ GeV}^2$
 - $0.3 \text{ GeV}^2 < Q^2 < 0.8 \text{ GeV}^2$
 - $Q^2 > 0.8 \text{ GeV}^2$
- 11-13) $3 \bar{\nu}_\mu$ CC $1\pi^\pm$ systematics for the following true kinematic bins:
- $Q^2 < 0.3 \text{ GeV}^2$
 - $0.3 \text{ GeV}^2 < Q^2 < 0.8 \text{ GeV}^2$
 - $Q^2 > 0.8 \text{ GeV}^2$

14-16) 3 ν_μ CC $1\pi^0$ systematics for the following true kinematic bins:

- $Q^2 < 0.35 \text{ GeV}^2$
- $0.35 \text{ GeV}^2 < Q^2 < 0.9 \text{ GeV}^2$
- $Q^2 > 0.9 \text{ GeV}^2$

17-19) 3 $\bar{\nu}_\mu$ CC $1\pi^0$ systematics for the following true kinematic bins:

- $Q^2 < 0.35 \text{ GeV}^2$
- $0.35 \text{ GeV}^2 < Q^2 < 0.9 \text{ GeV}^2$
- $Q^2 > 0.9 \text{ GeV}^2$

20) 1 ν_μ CC 2π systematic

21) 1 $\bar{\nu}_\mu$ CC 2π systematic

22-24) 3 ν_μ CC DIS systematics for the following true kinematic bins:

- $E_\nu < 7.5 \text{ GeV}$
- $7.5 \text{ GeV} < E_\nu < 15 \text{ GeV}$
- $E_\nu > 15 \text{ GeV}$

25-27) 3 $\bar{\nu}_\mu$ CC DIS systematics for the following true kinematic bins:

- $E_\nu < 7.5 \text{ GeV}$
- $7.5 \text{ GeV} < E_\nu < 15 \text{ GeV}$
- $E_\nu > 15 \text{ GeV}$

28) 1 ν_μ CC coherent systematic

29) 1 $\bar{\nu}_\mu$ CC coherent systematic

30) 1 ν_μ NC systematic

31) 1 $\bar{\nu}_\mu$ NC systematic

32) 1 ν_e/ν_μ cross-section ratio systematic

33-42) 10 final state re-interaction (FSI) systematics:

- 33) pion mean free path, controlling the pions re-interaction rate
- 34) nucleon mean free path, controlling the nucleon re-interaction rate
- 35) fraction of rescattered pions in charge exchange channels
- 36) fraction of rescattered pions in inelastic channels
- 37) fraction of rescattered pions in absorption channels
- 38) fraction of rescattered pions in pion production channels
- 39) fraction of rescattered nucleons in charge exchange channels
- 40) fraction of rescattered nucleons in inelastic channels
- 41) fraction of rescattered nucleons in absorption channels
- 42) fraction of rescattered nucleons in pion production channels

The last 10 systematics parameterize FSI effects. They are non-linear systematics and they are applied to every single multi-dimensional kinematic bin, of each one of the MC templates used for each one of the fit samples of the present analysis. The response of each individual bin to each of the 10 FSI systematics is pre-computed for a range of values of each systematic. Using the pre-computed values, the response of each bin is parameterized using cubic splines which are interrogated during the process of producing any single set of VALOR DUNE predictions. Separate response functions are calculated for each of the 3 detector options, using the outputs of the corresponding simulation chain. Although in the covariance matrix these FSI parameters are taken to be uncorrelated with other systematic parameters, correlations between all kinematical bins are taken into account.

The first 33 parameters listed describe the uncertainty in the cross-section for the corresponding process in the absence of FSI effects, and they are linear parameters applied in the appropriate kinematical range of the MC templates corresponding to the appropriate true reaction mode (more details are given in Appendix B of [6]). For them, a covariance matrix is computed using a sample of 100k ν_μ and 100k $\bar{\nu}_\mu$ events generated using GENIE v2.10.6 and the optimized neutrino flux histograms. The procedure itself is largely model agnostic, and translates the effect of several model-dependent parameters to an appropriately chosen set of primarily linear model-independent ones. This covariance matrix is constructed using GENIE re-weighting tools.

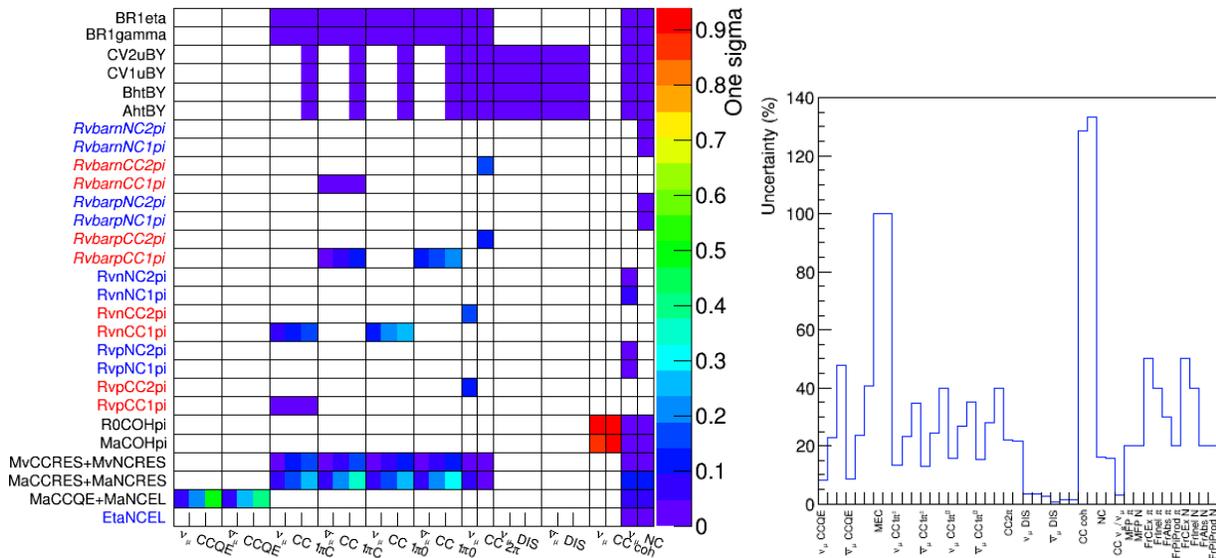


Figure 6: The *a priori* cross-section uncertainties. Left: effect of the 1σ variation of each GENIE parameter (in the Y axis) in each one of our model-independent parameters (in the X axis). Right: uncertainties in the model-independent parameters, calculated by adding in quadrature the effects of each GENIE parameter shown in the left plot.

A two-step process is used to obtain conservative prior errors:

1. Firstly, the effect of the 1σ variation of each GENIE model parameter is calculated by varying it independently of the others in order to avoid potential cancellations that may occur when several parameters affecting the same bins are tweaked at the same time. For each of the model-independent parameters listed above, the effects of the 1σ variation of each GENIE parameter are added in quadrature. This is calculated by varying the GENIE parameters with a $+1\sigma$ variation and a -1σ variation, and choosing the maximum total uncertainty. Figure 6 (left) shows the effect of the 1σ variation of each GENIE parameter (in the Y axis) in each one of our model-independent parameters (in the X axis); these effects are then added in quadrature to compute the uncertainties in the model-independent parameters, presented in 6 (right). Notice that the red(blue) labels in the Y axis in Fig. 6 (left) denote

parameters related to CC(NC) interactions, while black denotes both CC+NC interactions, and italic text is used when the labels refer to parameters applied only to antineutrino interactions.

- Secondly, the correlations between any of the 33 parameters listed above are calculated as well. To do this, all the GENIE parameters are varied simultaneously. These correlations are illustrated in Figure 7.

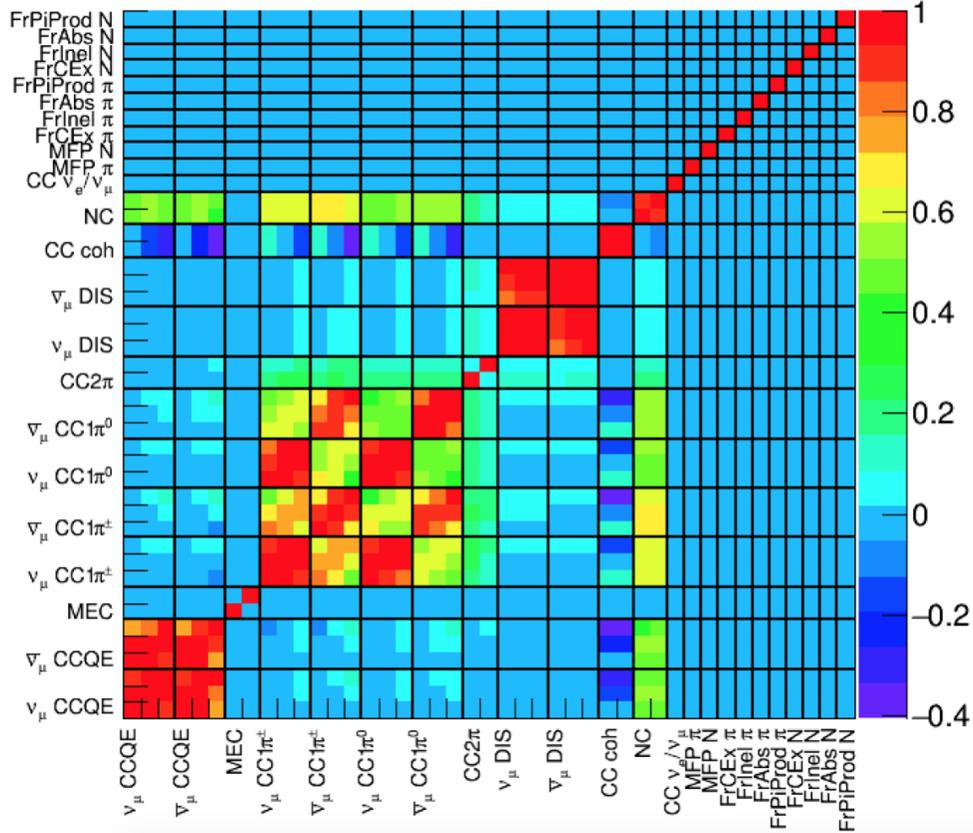


Figure 7: The *a priori* cross-section correlation matrix

By combining the results from both steps (uncertainties from the first one and correlations from the second one), the final covariance matrix is built for the neutrino interaction uncertainties, input to the VALOR fit.

5 External Backgrounds

The neutrino interaction samples described in the later sections ??, ??, and ?? have mixed in with them the appropriate amounts of cosmic ray induced particles and particles originating from neutrino interactions in the rock surrounding the DUNE near hall. The next two sections describe the creation of these two background samples.

5.1 Cosmic Ray Induced Particles

The cosmics simulation for the NDTF is based in LArSoft and relies on LArSoft modules that have been developed and used by other experiments to produce cosmics samples. The overall approach is to use a cosmic ray Monte Carlo generation library to generate muons and neutrons that are then propagated

through the LArND geometry using Geant4. The outputs are extracted at the end and made available as ntuples describing cosmogenic particles passing through the boundaries of the ND hall.

Throughout the simulation the LArND geometry is used because it is compatible with LArSoft and so allows the use of available LArSoft modules. In the end the particles are extracted at the boundary of the ND hall so the detector geometry is irrelevant to the goals of the overall cosmic simulation. The cosmic-ray shower library (CRY) is used via the CosmicsGen larsim module to generate muons and neutrons at the level of the ND. These particles are then moved back to the edge of the world volume using their trajectories. The output is then propagated through the geometry using the largeant module and the particles reaching the boundary of the ND hall are extracted into separate ntuples with a custom larsoft producer module. From the surface level to the top of the ND hall there is 53 m of material with the first 20 m being dirt (1.7 g/cm³) and the remaining 33 m being rock (2.43 g/cm³).

A sample of 10,000 5 ms blocks of time are simulated. The simulation predicts a muon rate of 2.7 Hz/m² at the top of the ND hall. This is in reasonable agreement with the measured rate in the MINOS hall, 0.8 Hz/m² given their differences in depth (53 m vs 93 m). In addition to muon fluxes the other particles coming out of Geant4 are also kept (neutrons, protons, pions, electrons, kaons) allowing them to be overlaid in the various ND simulations. Specific issues that could be addressed to generate more detailed/accurate fluxes in the future are:

- The geometry has values for the density of dirt (1.7 g/cm³) that differs from the value used in the MINOS simulations (2.29 g/cm³). These values should be updated to the MINOS values which will decrease the cosmogenic flux slightly.
- The world volume of the geometry should be expanded to increase the dirt/rock that is traversed by the more horizontal muon flux. This would reduce the horizontal flux which is expected to be an overestimate in its current configuration.
- CRY generates the cosmic flux expected at sea level, 2100 m, or 11,300 m. A possible improvement here could use a higher elevation with a modified geometry or a different generator (e.g. CORSIKA).
- CRY has wide bins in energy and zenith angle which lead to binning artifacts in the resulting flux. This is a relatively small effect, but could be addressed by moving to another generator.

5.2 Particles Induced by Neutrino Interaction in the Rock

The simulation of particles induced by neutrino interactions occurring in the rock surrounding the Near Detector (ND) cavern (hereafter referred to as rock events) is separated into three stages, all of which are outside of the ART (and LArSoft) framework. The ultimate output of the simulation is a set of particle states (position, momentum etc.) located on the inner side of the ND cavern walls which are matched to the rock events that created them. All steps of the rock event generation use the ND world geometry in gdml format.

The first step of the simulation is generation of the neutrino flux incident on the cavern and the surrounding rock. The simulation utilises a converter (generate_gsimple.sh from gsimple v2.8.6c) which converts "dk2nu" flux information to a set of neutrino rays ("gsimple" rays) that pass through a user defined 2D flux window. To ensure that the simulation is all encompassing, the flux window is chosen to be 760 x 540 m² in size, 275 m downstream of the ND cavern and oriented such that the beam direction vector is normal to the flux window. All chosen values are based on simple path length calculations for a 120 GeV muon (the simulation's highest available energy) just reaching the ND cavern. The DUNE reference beam design (ref_01 in dk2nu format) is used to produce 1,000 gsimple files, each containing 10,000 neutrino rays. Note the reference beam flux is used here rather than the optimized flux used everywhere else. This difference is expected to have negligible consequences. The POT per file is 4 x 10³.

The second step of the simulation is neutrino event generation in the surrounding rock. The gev_gen_fnal executable (GENIE v2.8.6c) is used, taking a randomized subset of the gsimple flux files as inputs. An adaptable fiducial volume (known as the flexible rock box) is used in the processing. It defines a varying fiducial volume around the ND cavern whose size is calculated based on the path length of a given neutrino interaction's final state particles. The flexible rock box helps to remove neutrino events in which the final state particles have no hope of reaching the ND cavern. The output of gev_gen_fnal is a set of single neutrino

interaction events with POT accounting and a recording of the relevant gsimple information used in the event processing. A total of 1950 files have been produced, each containing 10,000 GENIE interactions. The POT per output file is 1.5×10^{14} .

The final step of the simulation is the tracking of the GENIE final states. Geant4 (v4.10.1.p03), setup with the QGSP_BERT_HP physics list, is used to propagate the GENIE final states (labeled with `kIStStableFinalState`) up to their entrance into the ND cavern, where the tracking stops. The particle state information at this point (position, momentum and particle species) is repackaged into the original GENIE event record along with a recording of the original final state as the particles mother. Finally, the ND cavern entering particles are labeled such that they should be tracked (with `kIStStableFinalState`) whereas the original final state particles are labeled with some other GENIE label (`kIStUndefined`). This relabeling scheme means that any downstream tracking algorithm that takes a genie event record as an input will only track the particles from where they enter the ND cavern, whilst correctly ignoring the original GENIE final states. A total of 1911 have been successfully produced, each containing 1200 rock events. The POT per file is identical to that in step two (1.5×10^{14}). The expected number of rock events for a DUNE spill (7.5×10^{13} POT) is 600.

There are a few items to note:

- The calculation of the flux window dimensions are based on a 120 GeV muon originating from the corners of the flux window and just reaching the ND cavern. Such an event would be very forward going, meaning that the flux window could be shrunk in future iterations. The approach used here however is more inclusive of the physics and the trade off, which is minor in terms of processing time, is CPU efficiency.
- All path length calculations for the flux window assume a rock density of 1.7 g/cm³ whereas the density used in GENIEs flexible rock box fiducial volume is 2.5 g/cm³. The lower density for the former will have further unnecessarily inflated the size of the flux window.
- Any neutrinos produced in the tracking stage, including those originating from the GENIE vertex, are immediately killed.
- The QGSP_BERT_HP physics list for the tracking stage is chosen due to its wide use in HEP experiments and its ability to track neutrons more precisely than other physics lists.
- Any intermediaries produced in the tracking stage have not been recorded in the genie event record.

6 Near Detector Samples

For each of the three candidate near detector technologies a large sample of simulated events has been produced in both forward and reverse horn current modes. These samples are used to understand how well each ND technology can constrain the systematic errors relevant to a long baseline oscillation analysis. The samples are also used to enumerate some basic detector properties, efficiencies, acceptances, etc.

Because of the 18 month duration of the taskforce work it is not possible to develop a full and mature reconstruction for each of the 3 technologies. The question with each reconstruction then becomes how best to mimic the reconstruction situation 10 years from now when algorithms have been fully developed. One can use the current state of a reconstruction algorithm as a mimic or one can use past experience with similar technologies and algorithms to smear truth information in the way that best mimics how an algorithm will eventually perform. This latter approach is referred to as "cheating".

For all three near detector options the reconstruction is done by a mixture of proper reconstruction and cheating. The decision which approach to apply is made algorithm by algorithm with the goal of best mimicking the reconstruction situation 10 years from now.

6.1 The Fine Grained Tracker

6.1.1 Simulation

The full Geant 4 simulation of the FGT is kept in a Fermilab redmine repository, called `dunefgt`, which uses ART for data handling and NuTools for interfacing with GENIE and Geant4. The geometry description

includes the Straw Tube Tracker (STT) complete with several targets and many radiators, the ECAL, and the Muon ID system comprising an arrangement of many Resistive Plate Chambers (RPC). The geometry is shown Fig. 8, with RPC trays in purple, the magnet in green, ECAL modules in red, and the STT in the center.

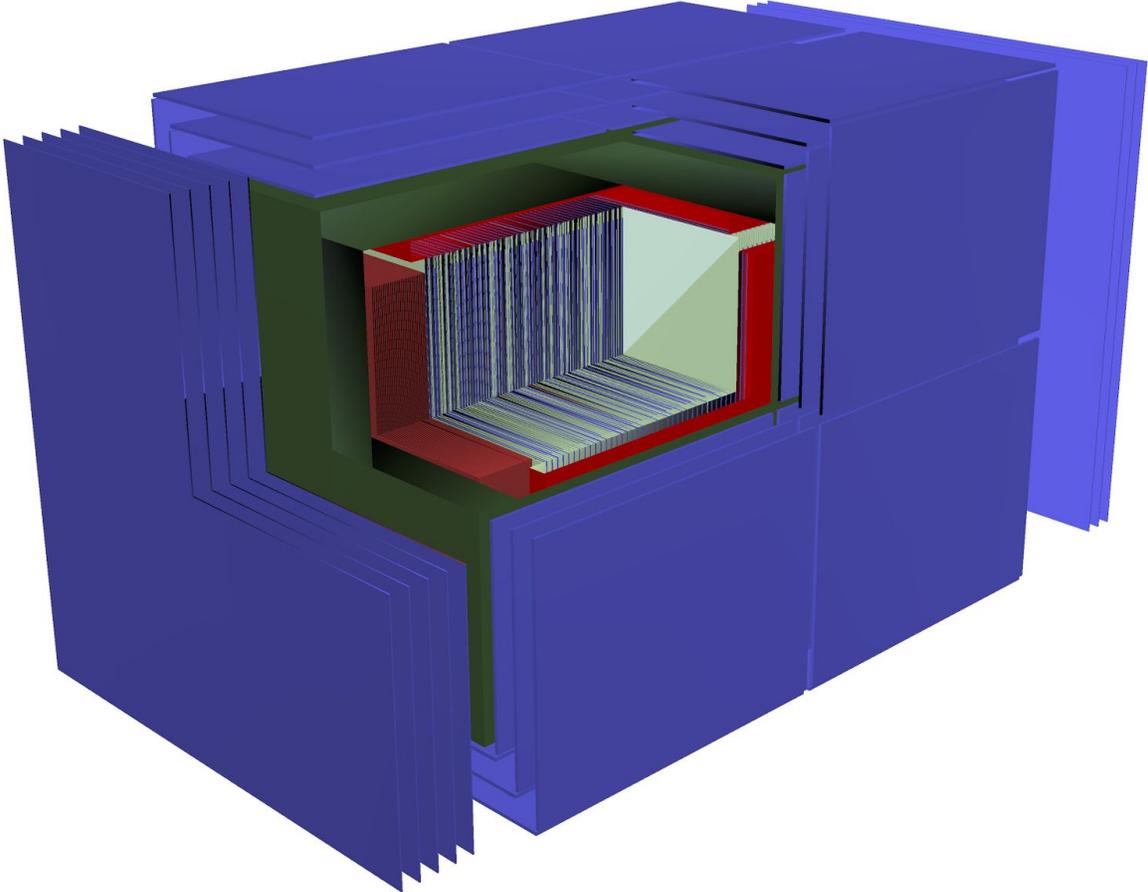


Figure 8: The Fine Grained Tracker geometry

The FGT simulation uses GENIE v2.10.10. Pileup is not yet simulated, though machinery to merge multiple interactions and background has been developed. Less than one interaction per spill is expected in the fiducial volume of the tracker, with most of the background coming from magnet and ECAL events. These generated events are tracked with Geant4 (currently v4.10.1_p03a), where user actions are defined to save the necessary information for each step in the most efficient way. This involves a Geometry interface aware of the world-coordinate placements and boundaries of all active sub-detectors, where each Geant4 step is passed to the appropriate active volume as quickly as possible and the minimum sufficient data is stored. Storing each step and simulated particle in a coherent and accessible way (C++ standard maps) takes a significant amount of virtual memory. Simulation of the detector electronics has been tabled for now.

About 1M neutrino and antineutrino events have been generated in a volume restricted to the STT. Once the proper POT per spill, as well as rock event and cosmic backgrounds, are implemented, new samples in the entire detector geometry will be generated for background rejection challenges. The Geant4 tracking and data-saving process dominates the virtual memory requirement for the entire Simulation and Reconstruction, as well as dominating run time and disk space of an event.

The most significant cut corner in the simulation, aside from the lack of pileup and background, is the lack of Transition Radiation (TR), which boosts the dEdx of electrons to help distinguish them from pions and muons.

6.1.2 Reconstruction

The desired neutrino energy, hadronic energy, and understanding of the event topology are each reconstructed using cheated and smeared tracks rather than a full reconstruction. A Gaussian smear is applied to account for the first-order detector effects on hit-level energy deposition, track direction, and momentum. The widths of the smearing are derived from functions fits to NOMAD data, which is expected to be quite conservative. Relative to the NOMAD detector, the FGT will have 6x (2x) the sampling transverse (parallel) to the beam. Furthermore, hit readout resolution will be less angle-dependent than NOMAD, due to the cylindrical readout of the straw tubes.

Being the root of the final resolutions on neutrino and hadronic energy, the momentum smear has the most significant effect. The best fit to fractional momentum resolution in NOMAD data involves a multiple scattering term and a measurement term:

$$\frac{\sigma_p}{p} = \frac{0.05}{\sqrt{L}} + \frac{\sigma_x}{0.3BL^2p} \sqrt{\frac{720}{N+4}}$$

The multiple scattering term depends only on the component of track length perpendicular to the magnetic field, L . The 5% is extracted from NOMAD and needs to be updated for the different FGT density. The measurement term depends on the magnetic field $B = 0.4\text{T}$, true particle momentum p , number of hits N , and spatial resolution of the hits σ_x , which is conservatively taken to be $200\mu\text{m}$ (given the ATLAS experience of $100\mu\text{m}$ resolution). Note that the component of any particles trajectory parallel to the magnetic field does not directly give any information toward its momentum, and therefor is not used. This parameterization captures this effect by blowing up as the trajectory becomes more parallel to the field, L going to zero. It is hoped that this happens infrequently and nothing is currently done to try to get a better momentum measurement by a different technique.

Shower energy in the ECAL is smeared based on MC studies, depending on containment and which part of the ECAL is hit. Contained showers in the downstream ECAL, having 3 samples per radiation length, are smeared by $6\%/\sqrt{E}$, and are otherwise smeared by 11% where there are 1.6 sample per radiation length. If an electron or positron leaves a STT track and an ECAL shower providing a better constraint on the energy than curvature and PID, that energy smear is used.

When a photon converts in the STT, its 4-momentum is simply the sum of the smeared electron/positron 4-momenta, and when it reaches the ECAL before converting, the energy is smeared in the ECAL way described above. More care with the direction needs to be taken in the future, as this is currently cheated and might impose a significant effect. When both photons from a π^0 decay are reconstructed in either of these ways, the π^0 is considered to be reconstructed, assuming that they will virtually always be associable. When one or no photon daughter is found, that π^0 's energy is lost, affecting both neutrino energy and topology classification. This may be too drastic and needs attention, since significant information may be regained from a lone but clear photon.

Once every track and shower is smeared, the reconstructed neutrino 4-momentum is simply the sum of the smeared 4-momenta of each primary track or π^0 , which has been smeared via sum of its daughters. It is assumed that, if a track has enough hits to be reconstructed, it can be identified as a primary track. In this way, missed π^0 's, neutrons, other neutral particles, and short-traveling charged tracks are missing from both final reconstructed energy and topology. Recovering information from neutron activity has been considered, but is too entangled with background effects to be well understood until further study.

Identifying the event topology is the most significantly cheated. The first step to reconstructing topology is finding the lepton, most commonly the muon. Any in-time track measured in the MuID detector is assumed to be associable with its corresponding STT track, given that it has one. Without background simulation, this is always the case, and with background, it may be safe to assume that STT tracks can be fit to and associated with their MuID track. When a muon has low enough energy, however, it can stop in the solenoid coils before reaching the MuID, faking a pion and potentially faking a NC event. This becomes quite geometry dependent since there is not steel to the sides, where lower-energy muons will be detected. In this case, muon vs. pion is left to other PID methods.

6.2 The High Pressure Gaseous Argon TPC

6.2.1 Simulation

The simulation of the pressurized gaseous argon TPC near detector (GArTPC-ND) is a stand-alone Geant4 application i.e. it does not depend on NuTools or the art processing framework kept at a public repository hosted on GitHub.

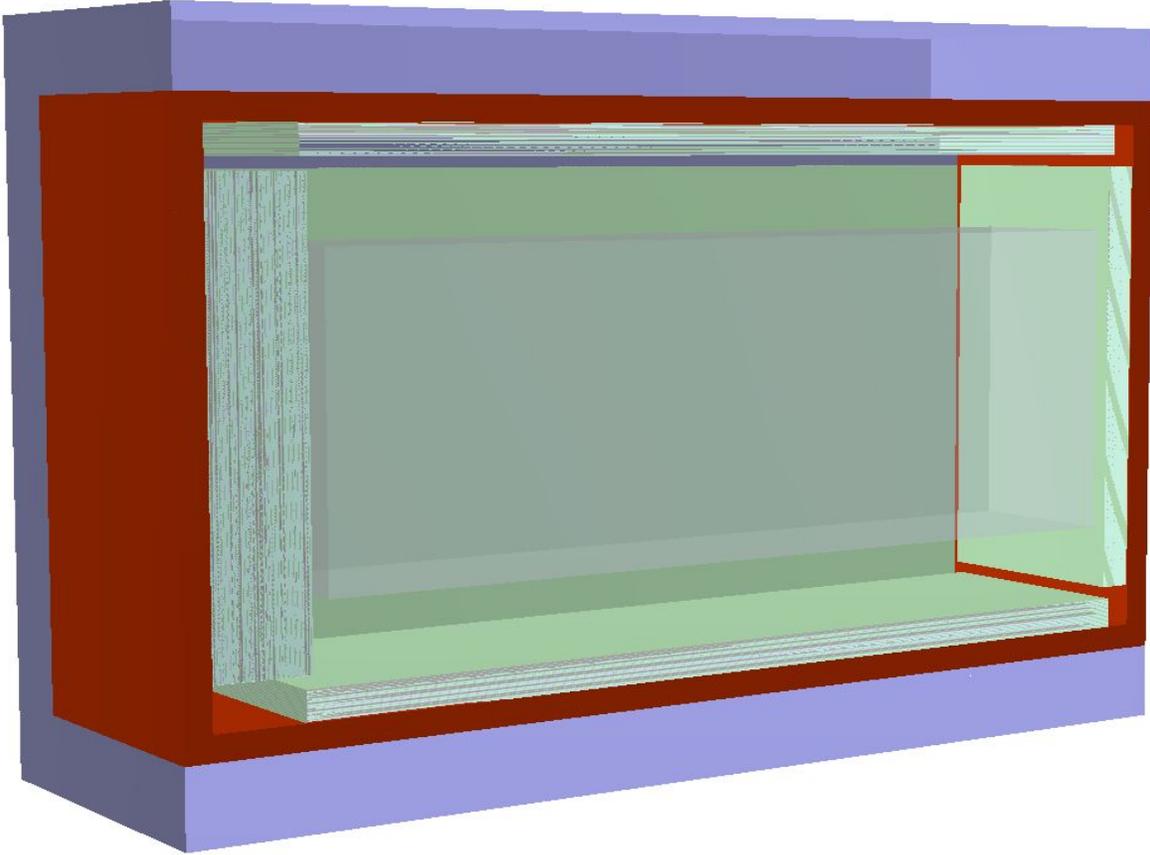


Figure 9: The High Pressure Gaseous Argon TPC geometry

The simulated detector geometry is shown in Fig. 9. The central element is a cylindrical pressure vessel, 6.5m long and 3.5m in diameter, made of a titanium alloy (UNS-R56323), that can hold about 1 tonne of argon at 10 bar. The pressure vessel barrel is 9.9 mm thick, whereas the endcaps are 17.5 mm thick. These thicknesses are estimated following the ASME code and assuming torispherical endcaps (not simulated yet). The pressure vessel houses a time projection chamber (TPC) of square cross section of $2.45 \times 2.45 \text{ m}^2$ with an electric drift field perpendicular to the vessels longitudinal axis and parallel to the floor. The pressure vessel is surrounded by sampling calorimeters (the so-called ECALs) of lead and plastic scintillator of design identical to those considered for the FGT-ND. The vessel and ECALs are enclosed within a solenoidal magnet, consisting of four copper coils and a return yoke made of iron, that creates a uniform magnetic field of 0.4 T parallel to the electric field of the TPC.

Each simulation event represents a spill of the neutrino beam ($\sim 7.5 \times 10^{13}$ POT). At the primary generation level the neutrino-rock interactions and the cosmic ray interactions described in Secs. ?? and ?? are mixed with neutrino detector interactions. On average, 144 neutrino interactions occur in the detector in each spill, with most of them ($> 85\%$) happening in the magnet and only 0.08% in the argon gas. Each simulation job starts with the production, using GENIE (v2.10.10 with MEC physics activated), of enough

neutrino detector interactions to fill 50 spills. Those interactions are then randomly mixed with interactions from cosmics and rock, and the chosen primary particles are propagated through the detector geometry down to zero range using Geant4 (currently v4.10.1.p03), leaving hits (position, time and associated energy deposits) at the various sensitive volumes of the near detector i.e. the TPC and the scintillator layers of the ECALs. A record of all the generated particles (including their genealogy) and their associated hits is stored in an output disk using ROOT. Processes involving the generation of primary signals, such as charge drift or scintillation light propagation, which are, in general, quite demanding computationally, are not simulated at the moment. Instead, the event reconstruction, described in the next section, makes use of the MC truth hits smeared to take into account the detector spatial and energy resolutions. In the 3rd Run Through, a total of 1 million of spills have been generated for each one of the beam operation modes.

6.2.2 Reconstruction

The event reconstruction used for the 3rd Run Through makes use of a combination of smeared simulation hits and cheated information.

The association of TPC hits to form tracks is currently being cheated, that is, TPC hits are always correctly assign to their MC true track. Therefore, the reconstruction starts with the selection of candidate vertices in the gas. Given that we only expect, on average, 0.15 argon interactions per spill, only one vertex is allowed per event at the moment. For vertices containing multiple tracks the selection is essentially pure. The selection of single-track vertices, however, poses a greater challenge due to the presence of background tracks originating from neutrino interactions outside the TPC, and it is still work in progress. For this reason, the legitimacy of any single-track vertex is cross-checked for now with the MC truth information. This results, probably, in an underestimation of the background for single-track channels. We are also assuming that vertices can be timed using the ECAL tracks/showers or with the primary scintillation signal from the TPC. The connection of TPC tracks and ECAL track/showers is currently cheated.

Mock reconstruction for momentum and angle in the TPC has been implemented following classic formulas for the prediction of resolution in trackers (see, for instance, R.L. Gluckstern, NIM 24 (1963) 381 or Blum and Rolandi, Particle detection with drift chambers):

$$\sigma(p_T)/p_T = \frac{\sigma_x p_T}{0.3BL^2} \sqrt{\frac{720}{N+4}} + \frac{0.05}{BL} \sqrt{\frac{1.43L}{X_0}}$$

$$\sigma_\theta = \frac{\sigma_X}{L} \sqrt{\frac{12(N-1)}{N(N+1)}} + \frac{0.015}{\sqrt{3}P} \sqrt{L/X_0}$$

where $p_T = p \sin\theta$ is the transverse momentum component (i.e. the momentum projection into a plane perpendicular to the magnetic field), σ_X is the point resolution, B is the magnetic field strength, L is the track length and N is the number of measurements in the track.

The dE/dx measurement, used for particle identification of the TPC tracks, is smeared according to the following empirical formula for argon:

$$\sigma(dE/dx) = 0.41N^{-0.43}(NP)^{-0.32}$$

Shower energy in the ECAL is smeared by 10% according to the MC studies done for the FGT. π^0 reconstruction is currently being cheated in the following way: if both photons resulting from the decay are detected, the π^0 is considered to be reconstructed, assuming therefore that they are always associable.

Once every track and shower is smeared, the reconstructed neutrino 4-momentum is simply the sum of the smeared 4-momentums of each primary track or π^0 , which has been smeared via sum of its daughters. It is assumed that, if a track has enough hits to be reconstructed, it can be identified as a primary track. In this way, missed π^0 s, neutrons, other neutral particles, and short-travelling charged tracks are missing from both final reconstructed energy and topology. Missing energy from neutrons is not being considered for the 3rd Run Through.

6.3 The Liquid Argon TPC

6.3.1 Simulation

The simulation of the LArND option uses a magnetised modular single-phase LArTPC based on ArgonCube (<http://cenf-argoncube.web.cern.ch/>).

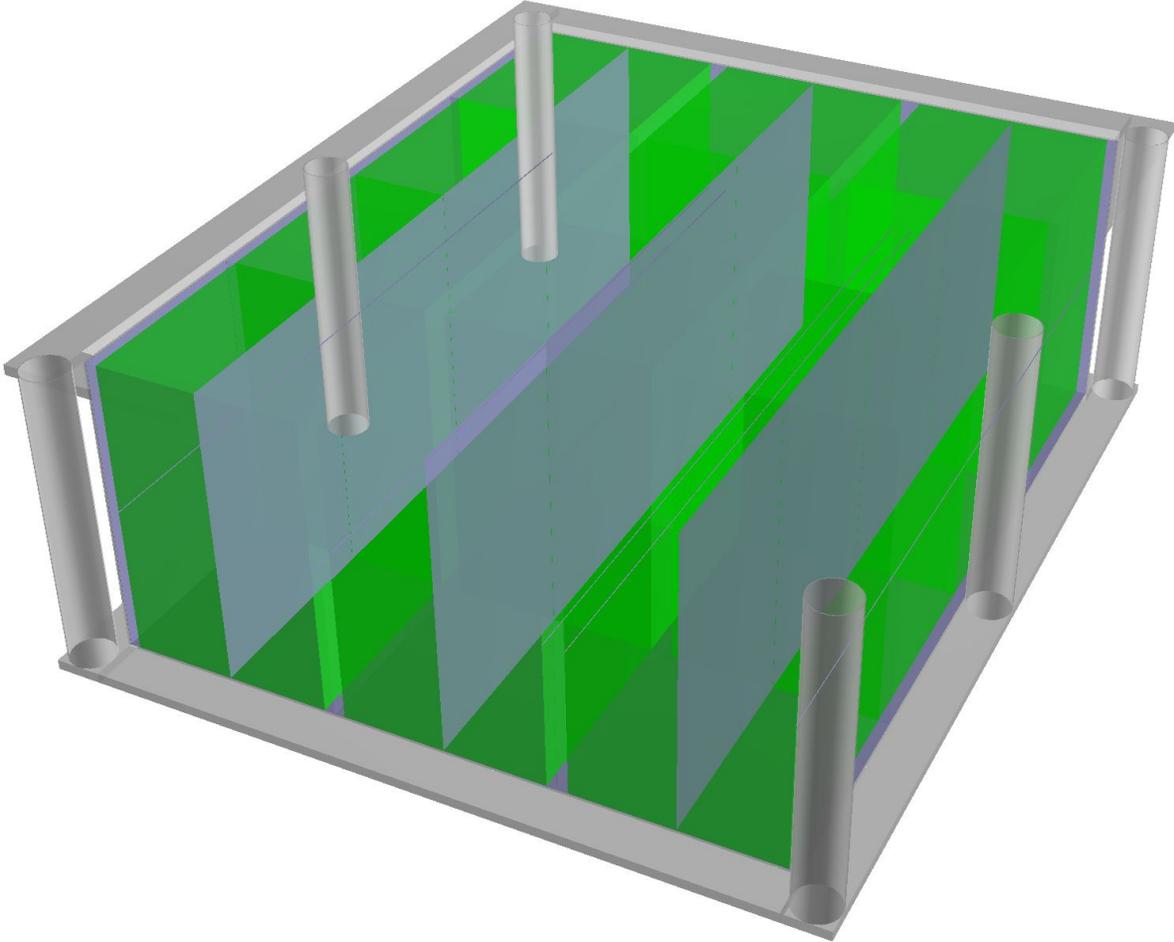


Figure 10: The Liquid Argon TPC geometry

Each module has a 2 m x 2 m footprint and the TPC volume is 3 m tall. The walls are ~ 2 mm thick with the field cages printed onto them. Dead space between modules is sub cm. Each module contains two symmetric TPCs, the cathode separating the TPCs runs along the centre of the module in the beam direction. With drift lengths of only 1 m, only a 100 kV field is intended and a with drift time of 0.7 ms is expected. For the LArND we are using 4 x 3 modules (6 m x 8 m), longest in beam direction. Containing 200 tonnes of active LAr.

A pixel-readout system, similar to MicroMegas, is used. The pixels provide live 3D readout, which means that the full reconstruction chain of LArSoft cannot be used. The pixel pitch used in the HW demonstrator is 2.86 mm. Simulating detector electronics, will be implemented once the pixel readout is better characterized.

The B-field of 1T is aligned vertically, perpendicular to the E-field and beam direction, to give greater separation of charged particles in the drift direction. Deflection in drift direction is desirable since we have greater resolution in timing. The magnet is based on a CERN design for magnetizing a single ICARUS module with a Helmholtz coil formed from two ATLAS toroids. This design requires no return yoke, therefore

the bulk of the material is from the Al stabilizer surrounding the NbTi cable, and the support structure. Only the bulk material has been simulated, the design will need optimizing.

Like the FGT, the LArND geometry is developed in a DUNE-group github repository (<https://github.com/DUNE/duneggd>). The geometry has been part of the tagged release of the LArSoft dunetpc repository (<https://cdcv.sfnal.gov/redmine/projects/dunetpc/repository>) since v06. The geometry is shown in Fig. 10, the module walls are shown in green, the grey planes within the modules are the cathodes. The bulk material of the magnet and its support structure are the grey components surrounding the modules.

Although the full LarSoft reconstruction chain cannot be used, event generation can be run through Larsoft. Each event represents a beam spill with 7.5×10^{13} POT. Based on tests, 20 neutrino interactions per spill are expected. After GENIE generation of the beam related interactions, additional interactions for rock and cosmic events are merged as discussed in Secs. 5.1 and 5.2.

For the 3rd Run Through 100k events have been generated for both neutrinos and antineutrinos. Rock and cosmic events have been merged into the neutrino events. Antineutrino events have only been merged with cosmics, rock events will be included once files are available. All samples include MEC interactions, and have been generated using only the active volume. Further samples will be generated using the entire geometry for background studies.

6.3.2 Reconstruction

Event reconstruction for the 3rd Run Through relies on MC truth information and smeared GEANT energy depositions. The most simplistic approach is taken where the neutrino energy smeared by a Gaussian with a sigma of 0.25.

The rest of this section describes the more advanced cheated reconstruction that is in development and will be implemented in the 4th Run Through.

The reconstruction will use extracted GEANT4 information; voxelising charge, and grouping particles based on Track ID. Smearing will be applied to the final products of this.

For tracks, the truth-based momentum value will be taken, and then smeared. The reconstruction of the momentum and angle from curvature and MCS, will use the Highland equation (V. L. Highland, Some Practical Remarks on Multiple Scattering, NIM 129 1975 104-120.) and methods described in the PDG sections 32.23 (Passage of Particles Through Matter Multiple scattering through small angles) and 33.12 (Detectors at Accelerators Measurement of particle momenta in a uniform magnetic field). The curvature will be fit with a half-reco algorithm, first fitting a curve to track points, then determining the expected resolution based on energy and the number of hits along the curve combined with the MCS contribution. PID will be based on tagging of wrong-sign muon decays, hard scatters, track length, curvature and end topology.

For showers PID will use gaps, dEdx, and the determination of one or two tracklets. A method for accounting for total energy and angle is still to be developed.

As there is no charge amplification at the pixels, their energy resolution is assumed to be the same as wires of similar pitch. MicroBooNE does not yet have published results for dE/dx, therefore the smearing applied to dE/dx will utilize the results of ArgoNeuT (R. Acciarri, et al, A study of electron recombination using highly ionizing particles in the ArgoNeuT Liquid Argon TPC 2013 JINST 8 P08005). This is overly conservative given ArgoNeuT's 4 mm wire pitch; A more accurate smearing needs to be determined.

Beyond fiducial volume and directionality cuts, only a basic method of rejecting background events will be used; if a track originating outside the LAr crosses a track from an event with the LAr, that event will be deemed unreconstructed. This method needs improving to include a cut on intersection distance to vertex.

7 Near Detector Systematics

Currently, simplistic and tentative detector acceptance parameters and uncertainties are used. More realistic detector uncertainties are not yet available from detailed studies using the actual simulation, reconstruction and event selection chain by each detector group. For this analysis a simple 10% value is assumed for the following 9 efficiencies:

- ν_μ CC 1-track QE enhanced sample efficiency
- ν_μ CC 2-track QE enhanced sample efficiency
- ν_μ CC 1π charged sample efficiency
- ν_μ CC 1π neutral sample efficiency
- ν_μ CC 1π neutral + 1π charged sample efficiency
- ν_μ CC other sample efficiency
- ν_e CC inclusive sample efficiency
- wrong sign ν_μ CC inclusive sample efficiency
- NC inclusive sample efficiency

8 The VALOR Near Detector Fits: Constraining the Systematic Uncertainties

The VALOR DUNE analysis is capable of fitting both the DUNE near and far detector data, independently or jointly. The near detector-only fit provides a constraint to the flux and cross-section systematic uncertainties, via the joint fit of neutrino flux and neutrino interaction parameters to Near Detector event samples.

The VALOR near detector fits use cuts on the simulated data to for 23 samples, for each of the two beam running modes (FHC, RHC) and for each of the 3 detector technologies. These samples are chosen to maximize the ability to constrain specific contributions to the total uncertainty.

The FHC samples are defined in Fig. 11. Equivalent samples with the appropriate change of sign are defined for RHC running.

8.1 Parameter Estimation

A near detector measures the event rates for specific final-state topologies (a set of states not entirely aligned with generator-level event classifications). To compare these measurements with theory, one needs to build an ND event rate model. This model is built from the convolution of flux Φ , interaction σ and detector (acceptance) models ϵ . These models have (parameterised) uncertainties $(f_{\Phi 1}, f_{\Phi 2}, \dots, f_{\Phi k}, f_{\sigma 1}, f_{\sigma 2}, \dots, f_{\sigma l}, f_{\epsilon 1}, f_{\epsilon 2}, \dots, f_{\epsilon m})$. In VALOR, a simultaneous fit of important (neutrino flux and neutrino interaction) systematic parameters to many ND samples is used to determine the values of these parameters, reduce their uncertainty, and obtain their correlations given the ND event rate constraint.

Measurements of a set of physics parameters $\vec{\theta} = (\theta_0, \theta_1, \dots, \theta_{m-1})$ in the presence of several systematic parameters $\vec{f} = (f_0, f_1, \dots, f_{N1})$ are obtained by comparing the observed and expected reconstructed kinematical distributions (in an N_r -dimensional space K_r) for a series of samples.

The expected number of events $n_{d;b;s}^{pred}(r; \vec{\theta}; \vec{f})$ in reconstructed bin r for each sample s , recorded in a detector d and exposed to a beam configuration b is thus calculated:

$$n_{d;b;s}^{pred}(r; \vec{\theta}; \vec{f}) = \sum_m \sum_t P_{d;b;m}(t; \vec{\theta}) \cdot R_{d;b;s;m}(r, t; \vec{f}) \cdot T_{d;b;s;m}(r, t) \quad (1)$$

where $P_{d;b;m}(t; \vec{\theta})$ encapsulates the effect of a physics hypothesis (e.g. neutrino oscillations in a 3-flavour framework), and $R_{d;b;s;m}(r, t; \vec{f})$ parameterizes the response of a template bin to systematic variations.

A binned likelihood-ratio method is typically used by VALOR. With the addition of binned data $n_{d;b;s}^{obs}(r)$, the following log-likelihood function is constructed:

$$-2 \ln \lambda_{d;b;s}(\vec{\theta}; \vec{f}) = 2 \sum_r \left\{ \left(n_{d;b;s}^{pred}(r; \vec{\theta}; \vec{f}) - n_{d;b;s}^{obs}(r) \right) + n_{d;b;s}^{obs}(r) \cdot \ln \frac{n_{d;b;s}^{obs}(r)}{n_{d;b;s}^{pred}(r; \vec{\theta}; \vec{f})} \right\} \quad (2)$$

- ν_μ CC
 1. 1-track 0π (μ^- only)
 2. 2-track 0π (μ^- + nucleon)
 3. N-track 0π (μ^- + (>1) nucleons)
 4. 3-track Δ -enhanced ($\mu^- + \pi^+ + p$, with $W_{reco} \approx 1.2$ GeV)
 5. $1\pi^\pm$ ($\mu^- + 1\pi^\pm + X$)
 6. $1\pi^0$ ($\mu^- + 1\pi^0 + X$)
 7. $1\pi^\pm + 1\pi^0$ ($\mu^- + 1\pi^\pm + 1\pi^0 + X$)
 8. Other
- Wrong-sign ν_μ CC
 9. 0π ($\mu^+ + X$)
 10. $1\pi^\pm$ ($\mu^+ + \pi^\pm + X$)
 11. $1\pi^0$ ($\mu^+ + \pi^0 + X$)
 12. Other
- ν_e CC
 13. 0π ($e^- + X$)
 14. $1\pi^\pm$ ($e^- + \pi^\pm + X$)
 15. $1\pi^0$ ($e^- + \pi^0 + X$)
 16. Other
- Wrong-sign ν_e CC
 17. Inclusive
- NC
 18. 0π (nucleon(s))
 19. $1\pi^\pm$ ($\pi^\pm + X$)
 20. $1\pi^0$ ($\pi^0 + X$)
 21. Other
- ν -e
 22. $\nu_e + e^-$ elastic
 23. Inverse muon decay $\bar{\nu}_e + e^- \rightarrow \mu^- + \bar{\nu}_\mu$ (including the annihilation channel $\nu_\mu + e^- \rightarrow \mu^- + \nu_e$).

Figure 11: The 23 samples defined for each of the two horn current mode and for each detector technology

The following quantity is constructed in the VALOR DUNE analysis by summing up the contributions from all fit samples:

$$-2 \ln \lambda(\vec{\theta}; \vec{f}) = -2 \sum_d \sum_b \sum_s \ln \lambda_{d;b;s}(\vec{\theta}; \vec{f}) - 2 \ln \lambda_{prior} \quad (3)$$

Most physics and systematic parameters in the VALOR fit come with prior constraints from external data. The following Gaussian penalty term is computed:

$$-2 \ln \lambda_{prior}(\vec{\theta}; \vec{f}) = \left\{ (\vec{\theta} - \vec{\theta}_0)^T \mathbf{C}_\theta^{-1} (\vec{\theta} - \vec{\theta}_0) + (\vec{f} - \vec{f}_0)^T \mathbf{C}_f^{-1} (\vec{f} - \vec{f}_0) \right\} \quad (4)$$

where C_θ is an $M \times M$ physics parameter covariance matrix, C_f is an $N \times N$ systematic parameter covariance matrix, $\vec{\theta}_0$ is a vector with the nominal values of the measured physics parameter, and \vec{f}_0 is a vector with the nominal values of the systematic parameters.

Best-fit values are obtained by maximizing:

$$\lambda(\vec{\theta}; \vec{f}) = \lambda_{DUNE}(\vec{\theta}; \vec{f}) \cdot \lambda_{prior}(\vec{\theta}; \vec{f}) \quad (5)$$

The advantage of the likelihood ratio method is that, in the large-sample limit, the quantity $-2 \ln \lambda(\vec{\theta}; \vec{f})$ has a χ^2 distribution and it can therefore be used as a goodness-of-fit test.

8.2 Binning

Within VALOR, a different reconstructed kinematical distribution can be fitted for each sample. For the 3rd Run Through analysis, we have chosen to fit 2-dimensional (E_{reco}, y_{reco}) distributions for all CC-like samples and 1-dimensional E_{vis} distributions for all NC-like samples, where E_{reco} is the reconstructed neutrino energy, y_{reco} is the reconstructed inelasticity, and E_{vis} is the reconstructed visible energy.

The following binning of reconstructed kinematical variables is used:

- 10 E_{reco} (E_{vis}) bins defined by the following bin edges:
(0, 1.0, 1.4, 2.0, 3.0, 4.0, 5.0, 7.0, 10.0, 15.0, 1000) GeV
- 8 y_{reco} bins for CC-like samples defined by the following bin edges:
(0, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1)
- 1 y_{reco} bins for NC-like samples.

Events are also binned according to their true GENIE reaction mode. The definition of these modes is influenced by the interaction parameter model being fitted. Sixteen reaction mode types are used, and in the near detector these are defined for each of ν_μ , $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$. The reaction mode types are defined by the process used to generate the event and the number of particles leaving the primary vertex (before FSI). The reaction mode types used for the current analysis are CC QE, CC MEC, CC $1\pi^\pm$, CC $1\pi^0$, CC $2\pi^\pm$, CC $2\pi^0$, CC $1\pi^\pm + 1\pi^0$, CC coherent, inverse muon decay, CC elastic scattering, CC other, NC $1\pi^\pm$, NC $1\pi^0$, NC coherent, NC elastic scattering, and NC other. For each of ν_e and $\bar{\nu}_e$, there is a single combined mode for CC and NC $\nu - e$ elastic scattering.

The templates are also binned in two true kinematic variables, E_{true} and Q^2 . The Q^2 binning varies based on the true mode to enable proper application of the binned interaction systematics. The truth binning scheme used has:

- 19 E_{true} bins defined by the following bin edges:
(0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0, 12.0, 16.0, 20.0, 40.0, 1000) GeV
- A number of Q_{true}^2 bins depending on the true reaction mode:
 - For neutrino and anti-neutrino CCQE, 3 Q_{true}^2 bins defined by the following bin edges:
(0, 0.20, 0.55, 1000) GeV^2
 - For neutrino and anti-neutrino CC $1\pi^\pm$, 3 Q_{true}^2 bins defined by the following bin edges:
(0, 0.30, 0.80, 1000) GeV^2
 - For neutrino and anti-neutrino CC $1\pi^0$, 3 Q_{true}^2 bins defined by the following bin edges:
(0, 0.35, 0.90, 1000) GeV^2
 - For all other reaction modes, 1 Q_{true}^2 bin.

8.3 Near-Detector Fitting Results

The primary result of the VALOR DUNE ND-only fit is a 147x147 parameter covariance matrix containing the 104 far-detector flux parameters and the 43 interaction parameters for each of the three near detector designs. To produce this matrix, all ND-only systematics are marginalized.

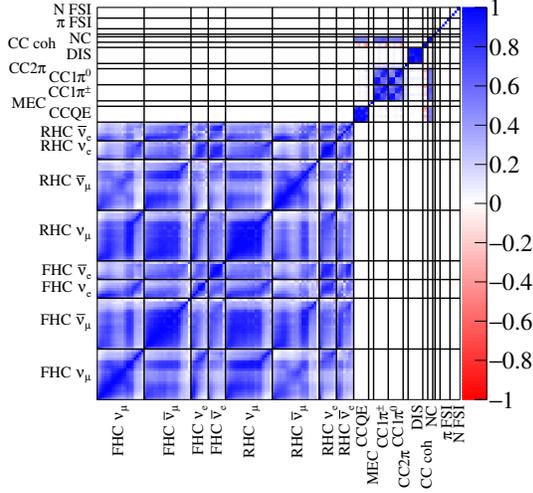


Figure 12: The prefit correlation matrix.

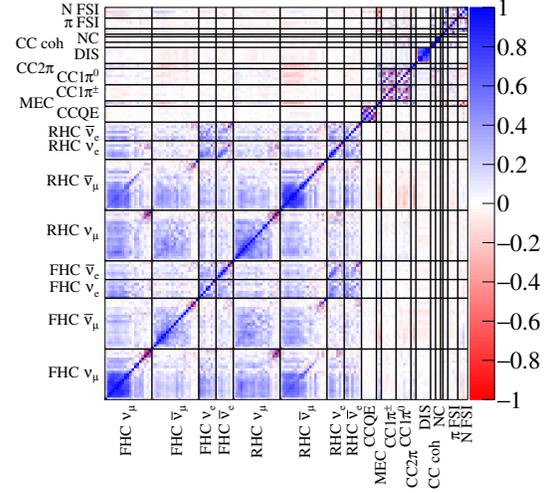


Figure 13: The postfit correlation matrix using the Fine-Grained Tracker.

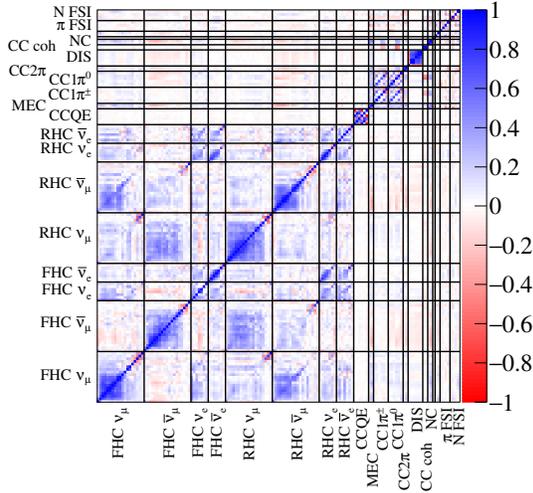


Figure 14: The postfit correlation matrix using the High Pressure Gaseous Argon TPC.

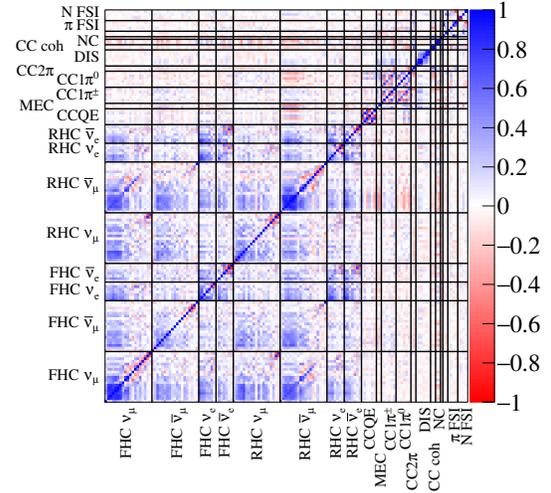


Figure 15: The postfit correlation matrix using the Liquid Argon TPC.

The fits shown in this section use a high exposure - 1.5×10^{22} POT for each of FHC and RHC beam, equivalent to 10.4 years running in each using the optimized beam design. As intended by the near detector designs, the fit shows a powerful ability to constrain the fitted parameters, and to break the correlations between them. The prefit correlation matrix can be seen in Fig. 12 while the postfit correlation matrices for the three detectors can be seen in Figs 13, 14 and 15.

The fitted single-parameter errors are shown for the prefit matrix and all three detectors in Fig 16. The overall fitted errors are highly constrained, likely a result of the relatively pure samples and overly naive

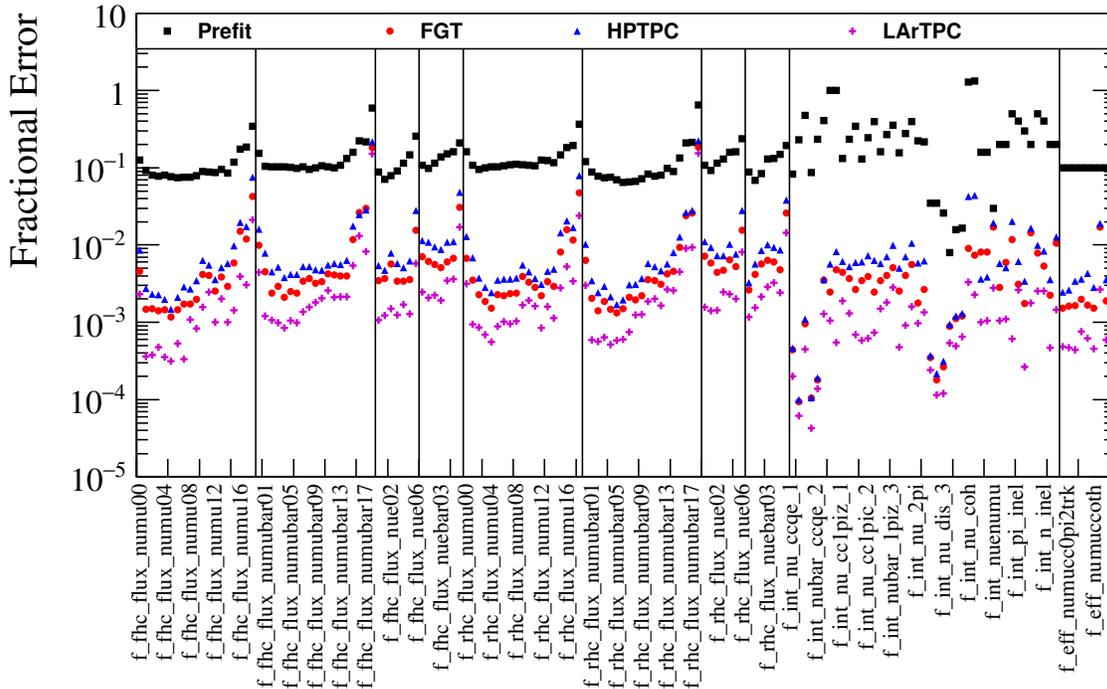


Figure 16: The uncertainties on each individual parameter. The flux parameters shown here are the fitted fluxes at the near detector.

uncertainty models. Work to produce a more realistic set of uncertainties and to evaluate the plausibility of the current set is underway.

9 The Far Detector Samples

9.1 Simulation

In the latest MC production far detector neutrino samples are simulated using GENIE version v2.10.10 with the optimized flux. Unoscillated samples are produced in both the neutrino and antineutrino modes as well as fully oscillated samples where all ν_{μ} 's in the beam are converted to either ν_{τ} 's or ν_{τ} 's. In order to save processing time, events are simulated using a smaller version of the full 10 kt far detector geometry. This geometry is 13.9 m long, 12 m high and 13.3 m wide, which consists of 12 APAs and 24 TPCs. The TPC wire spacing is 5 mm, the wire angle is 36 degrees.

The particles generated in the GENIE event generator step are passed to a GEANT v4.10.1_p03 based detector simulation. In this step, each primary particle from the generator and its decay or interaction daughter particles are tracked as they traverse liquid argon. The energy deposition is converted to ionization electrons and scintillation photons. Some electrons are recombined with the positive ions while the rest are drifted towards the wire planes. The number of electrons is further reduced by the existence of impurities in the liquid argon, which is commonly parameterized as an electron lifetime. The default electron lifetime is 3 ms in the simulation. The longitudinal diffusion smears the arrival time of the electrons at the wires and the transverse diffusion smears the electron location among neighboring wires.

The electrons on each wire are converted into raw wire signal (ADC vs Time) by convolution with the field response and electronics response. The field response on each wire plane is simulated with Garfield while the ASIC electronics response is simulated with the BNL SPICE simulation. Currently, the signal on each wire can only be produced from the ionization electrons going through the wire. Improvement upon this approximation is underway. This is important for the induction wire planes, where the induction signal depends on the local ionization charge distribution. By default, the ASIC gain is set to 14 mV/fC

and the shaping time is set to $2\mu\text{s}$. The noise level is set to 2.5 ADC RMS. In the current simulation, the electronic noise is assumed to be white, which is a uniform distribution in the frequency domain. The implementation of a more realistic electronics noise model is ongoing. More details on simulation can be found in DUNE-doc-1689.

9.2 Reconstruction

The reconstruction chain that produces track, shower and calorimetric information is fully automated. The current track reconstruction is described below and is efficient and robust. Other components of the reconstruction chain can be found in DUNE-doc-1689.

The first step in the reconstruction converts the raw signal from each wire to a standard Gaussian shape. This is achieved by passing the raw data through a calibrated deconvolution algorithm. Deconvolution removes the impact of the field and electronics responses from the measured signal to recover the number of ionized electrons.

The next step uses GausHitFinder, a hit-finding algorithm, that works by starting from deconvolved signals on wires and defines areas above threshold known as pulses. Once a pulse is found, an "n" Gaussian hypothesis is applied where "n" is defined by the number of peaks initially identified within the pulse. Based on the outcome of the fit an object known as a hit is formed and stored in the event.

Next the Line Cluster algorithm constructs two-dimensional line-like clusters using local information. The concept is to construct a short line-like seed cluster of proximate hits in an area of low hit density where hit proximity is a good indication that the hits are indeed associated with each other. Additional nearby hits are attached to the leading edge of the cluster if they are similar to the hits already attached to it. The conditions are that the impact parameter between a prospective hit and the cluster projection is similar to those previously added and the hit charge is similar as well. These conditions are moderated to include high charge hits that are produced by large dE/dx fluctuations and the rapid increase in dE/dx at the end of stopping tracks while rejecting large charge hits from delta-rays. Seed clusters are formed at one end of the hit collection so that crawling in only one direction is sufficient. Line Cluster uses disambiguated gausshits as input and produces a new set of refined hits.

The Projection Matching Algorithm (PMA) provides 3D reconstruction of individual particle trajectories (trajectory fit). Reconstructed 3D objects also provide basic physics quantities like particle directions and dE/dx evolution along the trajectories. PMA uses as its input the output from 2D pattern recognition: clusters of hits. For the purposes of the DUNE reconstruction chain the Line Cluster algorithm is used as input to PMA. The PMA builds and optimizes objects in 3D space (formed as polygonal lines with iteratively increased number of segments) by minimizing the cost function calculated simultaneously in all available 2D projections. The track can be reconstructed using clusters from two projections while the distance of hits to the track projection in the third plane is used to validate correct association of clusters.

10 The Far Detector Fit

10.1 Overview

The final oscillation fit uses the Long baseline Oscillations Analysis Fitter (LOAF). The LOAF fitter is built for speed and so rather than building spectra event by event, it uses prebuilt histograms and response function which encode the information needed to alter those histograms given a set of fit parameter values. Since oscillation probabilities are a function of true neutrino energy and species, template histograms of true neutrino energy broken out by species are required as inputs. Smearing functions are used to convert the oscillated true neutrino energy spectra to reconstructed neutrino energy spectra. Systematic fluctuations are applied to either the true energy spectrum before smearing (flux, cross section, nuclear models), or to the reconstructed energy spectra (reconstruction and efficiency uncertainties, and other detector effects) via response functions which provide the relevant spectral distortions induced as functions of systematic (fit) parameter changes. Parameter variations from nominal are used to determine penalty terms in the fit χ^2 .

The fit χ^2 assumes Poisson probability distributions for event counts, and Gaussian probability distributions for the priors used in penalty terms. In the case of correlated priors a covariance matrix is used to

| Detector | Fraction of δ_{CP} | Fraction of δ_{CP} |
|----------|----------------------------|----------------------------|
| | Above 3σ NH (IH) | Above 5σ NH (IH) |
| | [%] | [%] |
| LAr TPC | 75.3 (72.3) | 56.4 (51.5) |
| HPTPC | 73.3 (72.3) | 56.4 (48.5) |
| FGT | 74.0 (72.3) | 57.0 (48.5) |
| Prefit | 9.9 (5.9) | 0.0 (0.0) |

Table 1: The fraction of points in δ_{CP} where the fit to the Asimov data set is above 3 (5) σ .

determine the penalty term. The formulation is consistent with the one used in the VALOR ND fits. The χ^2 is minimized with the MIGRAD algorithm in MINUIT2 as implemented in ROOT.

As this is a sensitivity study Mock Data must be generated. There are two classes of Mock Data used. The first is the most probable data, often referred to as the Asimov Data Set. This is created by assuming that the true value of each parameter is the nominal value from the simulation. This includes both systematic parameters and the event counts (statistical variations). The second class of Mock Data are Toy MC Data Sets where the true values of each parameter are chosen at random based on the prior probability distributions. Again this applies for both systematic parameters and event counts (statistical variations). Asimov Data Set studies only require a single fit, however special care must be taken to ensure correlations between data sets are properly dealt with. Toy MC Data Sets require a series of fits, and the results are given by examining the ensemble of fit results.

10.2 Systematic Uncertainties

10.2.1 Inputs from VALOR (flux and cross section)

Flux and cross section uncertainties are parameterized in the same manner as in VALOR. For each systematic a response function is generated to propagate variations in underlying model parameters to the Far Detector (FD) spectra. The results of the VALOR ND fit for each detector option are propagated to the LOAF fitter via a covariance matrix that gives the covariance of the ensemble of fit parameters that describe the uncertainty on the predicted event rate at the FD. These covariance matrices (broken into correlation matrices and diagonal values) are shown in Figs. 13 to 16. Detector response systematics for the NDs have been marginalized, and are not propagated. The constraints on the fit parameters are enforced through a penalty term calculated by inverting the matrix, C , and multiplying by the vector of fit parameters, f , $penalty = f^T C^{-1} f$.

10.2.2 Detector systematics

The fits of the 3rd Run Through do not include any estimates of FD response uncertainties. These include energy scale, calibration, efficiency, etc. A plan to include these has been developed, however current reconstruction efforts are not advanced enough to provide reasonable estimates of LAr TPC uncertainties in the 10 - 20 year timescales.

10.3 Fit Results

10.3.1 Results from the LOAF fits

The results of the LOAF fits are shown in Tab. 1, and Figs. 17 and 18. The use of all 3 detectors represents significant improvement over the external constraints used to generate the prefit matrix. The LAr TPC provides slightly better sensitivity than the other two options, which have almost identical responses. However, given the quality of the simulation and reconstruction processes, these small sensitivity difference should not be taken seriously.

Before the final report the following LOAF development tasks should have been completed:

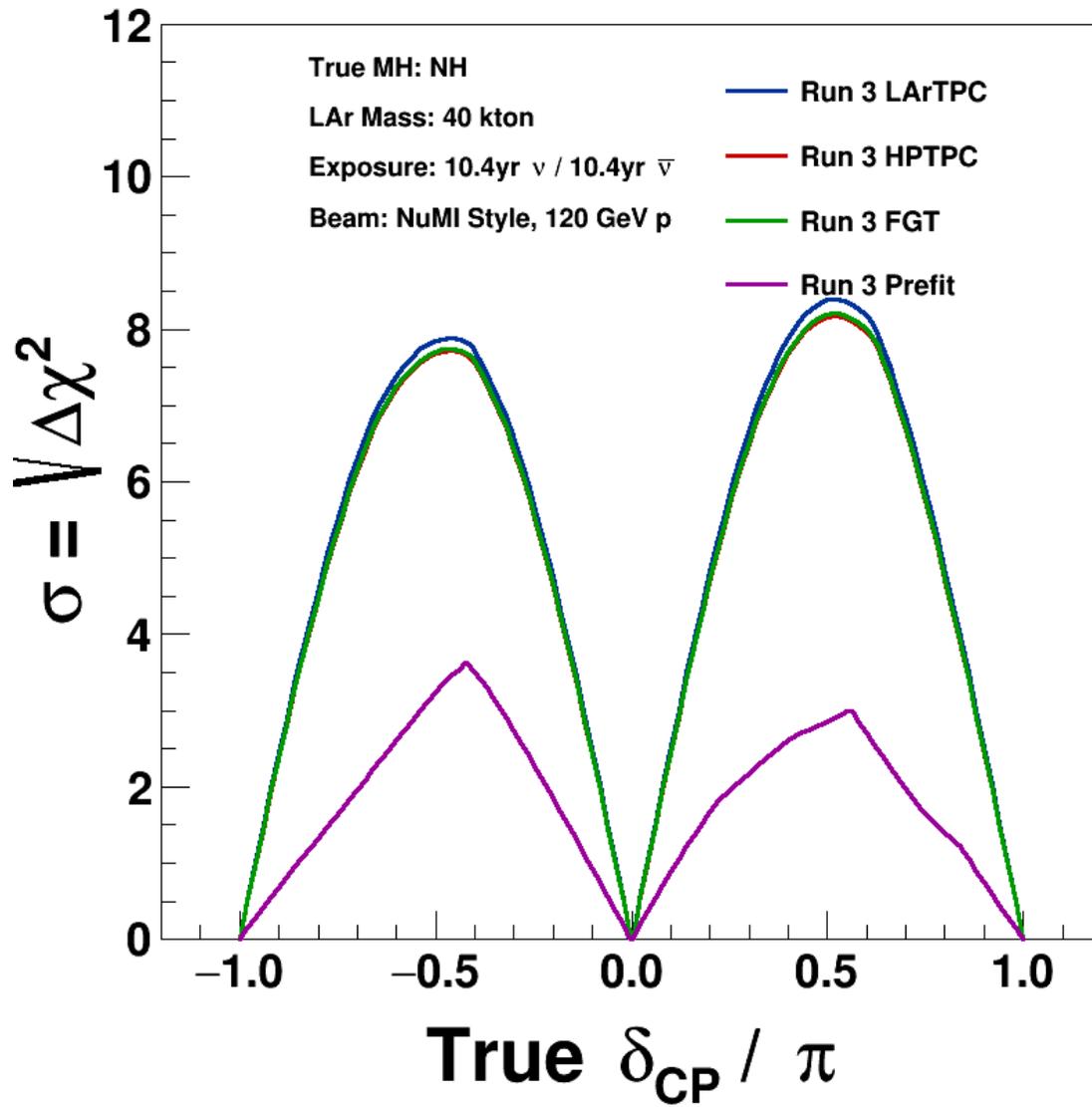


Figure 17: CPV Sensitivity for the NH

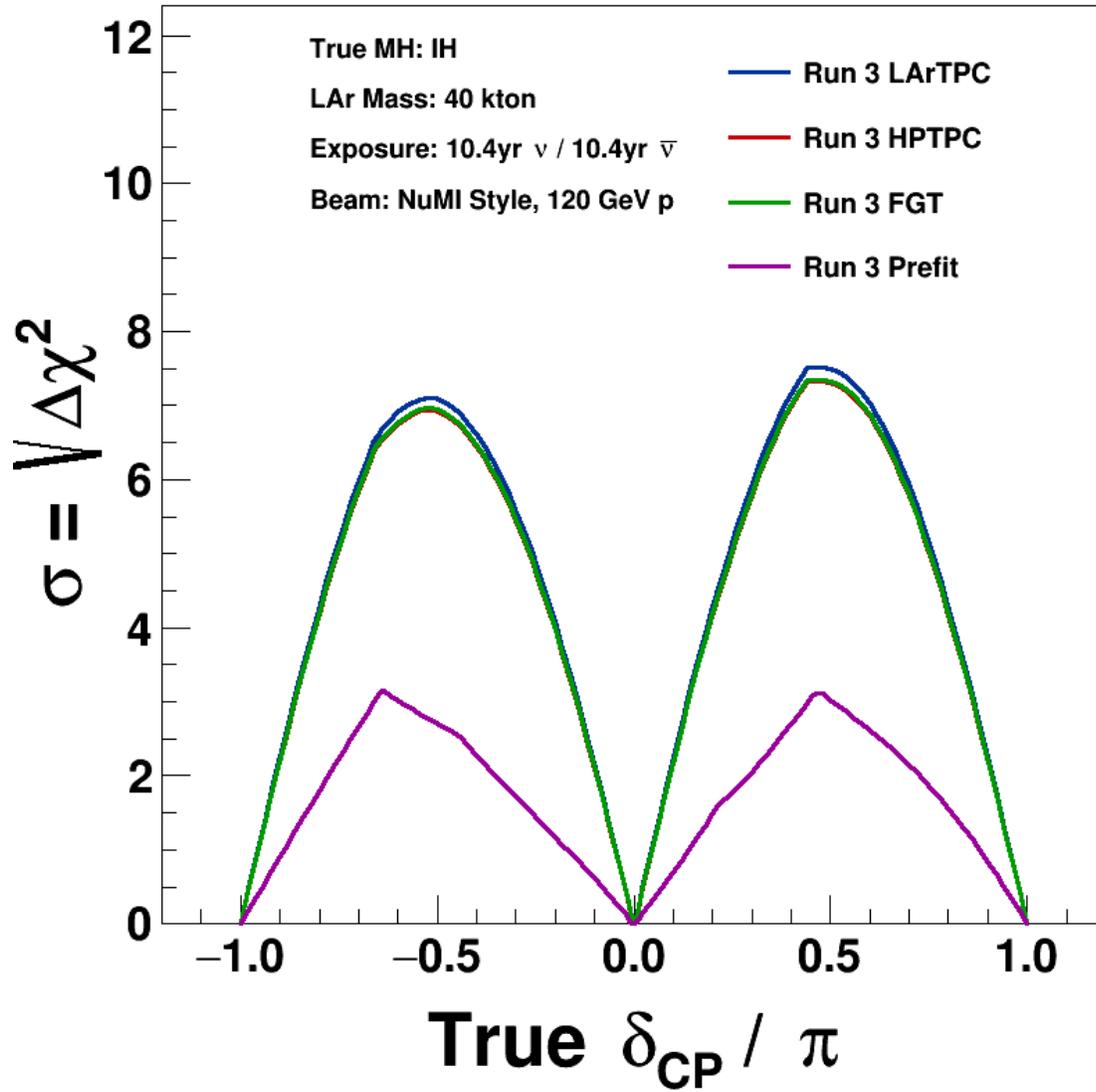


Figure 18: CPV Sensitivity for the IH

- Detector systematics
- Re-implement the MCMC
- Near/Far flux uncertainties
- New response function to reflect the run 3 VALOR parameterization of uncertainties.
- Diagonalize the covariance matrix and fit in minimal number of parameters
- Oscillation parameter prior correlations (can they be included?)
- Incorporate inputs from the full simulation rather than using FMC inputs

10.3.2 The VALOR fits

As mentioned in Sec. 8 a combined ND+FD fit is also under development using VALOR. This oscillation analysis is performed in a full 3-flavour framework including matter effects. Frameworks with sterile neutrinos (3+1, 3+2, 1+3+1) are also supported; more details about the calculation and validation of the oscillation probability can be found in [6] and [7]. Near and Far detector samples are fit jointly, and neutrino flux, interaction and detector systematics are marginalized during the oscillation fit. Results from this approach are not yet available, but will be included in the final report. A joint Near and Far detector oscillation analysis is an important piece as it avoids concerns regarding the accuracy of the very large covariance matrix produced by the fit to the Near detector data.

In the VALOR Joint Near and Far oscillation analysis, the expected number of events of the different samples in the Near and Far detector are calculated using Eq. 1 and fit jointly using a binned likelihood-ratio method constructed following Eq. 2. Several of the parameters in $\vec{\theta}$ and \vec{f} need to be eliminated, while the parameters of interest are measured. For example, all (flux, cross-section and detector) systematics in \vec{f} , as well as all squared-mass splittings and mixing angles in $\vec{\theta}$ are elicited in a 3-flavour analysis for measuring the CP-invariance violating phase δ_{CP} , and in the sensitivity studies presented.

Multiple studies, performing the joint Near and Far detector oscillation analysis using the 3 Near detector options and the 2 MC simulations of the far detector, are currently under development by the VALOR group.

11 The Figures of Merit

The potential options for figures of merit listed below are primarily based on experience from MINOS, NOvA and T2K. There are three broad categories of metrics one can use to evaluate the different near detector options and these are described in the next three sections. In each category are a set of Figures of Merit (FOMs). The intent is to provide figures or tables for each in the final report. In this initial report the FOMs are listed, but for only some of them have the corresponding figures or numbers been produced. Please note, some of the following figures show distributions for only the LArTPC and HPTPC options; the FGT distributions are not included in those figures because the source information stored for the FGT option in the 3rd Run Through is corrupted. The FGT distributions will be fixed for future iterations.

11.1 FOMs: The ability of the detectors to enhance the oscillation analyses

The primary figures of merit are:

- The sensitivity of discovering a non-zero CP violating phase:
See Figs. 17 and 18.
- The sensitivity for establishing the octant of the atmospheric mixing angle

11.2 FOMs: The performance of the detectors in the beam

The proposed figures of merit are

- The number of interactions in the detector per proton on target:
Fig. 19 shows the number of neutrino interactions normalized by 10^{15} POT.
- The pile-up in the detector due to the beam intensity
- The fraction of energy shared between neutrino interactions in the same beam spill
- The fraction of energy shared between cosmic rays and neutrino interactions

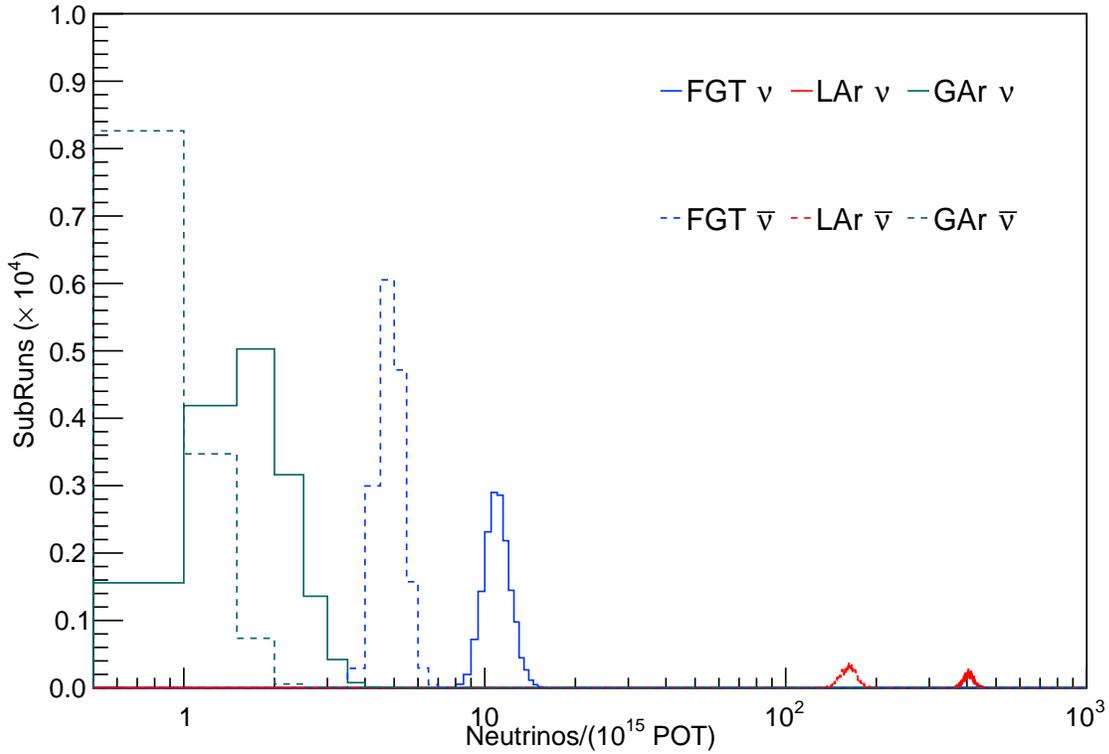


Figure 19: Average number of neutrino interactions per 10^{15} protons on target in each of the detector and beam configuration options.

11.3 FOMs: The ability to do physics with the near detectors

The quantities related to the detector performance are:

- The vertex positions and resolution:
Figure 20 shows the vertex positions in the detector coordinate systems.
- The energy resolution for EM showers, hadronic showers, minimum ionizing particles, and the total neutrino interaction energy resolution:
Figure 21 shows the true neutrino spectrum along with the reconstructed spectrum and energy resolution for the full interaction. The true neutrino spectrum for the LArTPC option is systematically lower than for the other options, an effect which is not currently understood.

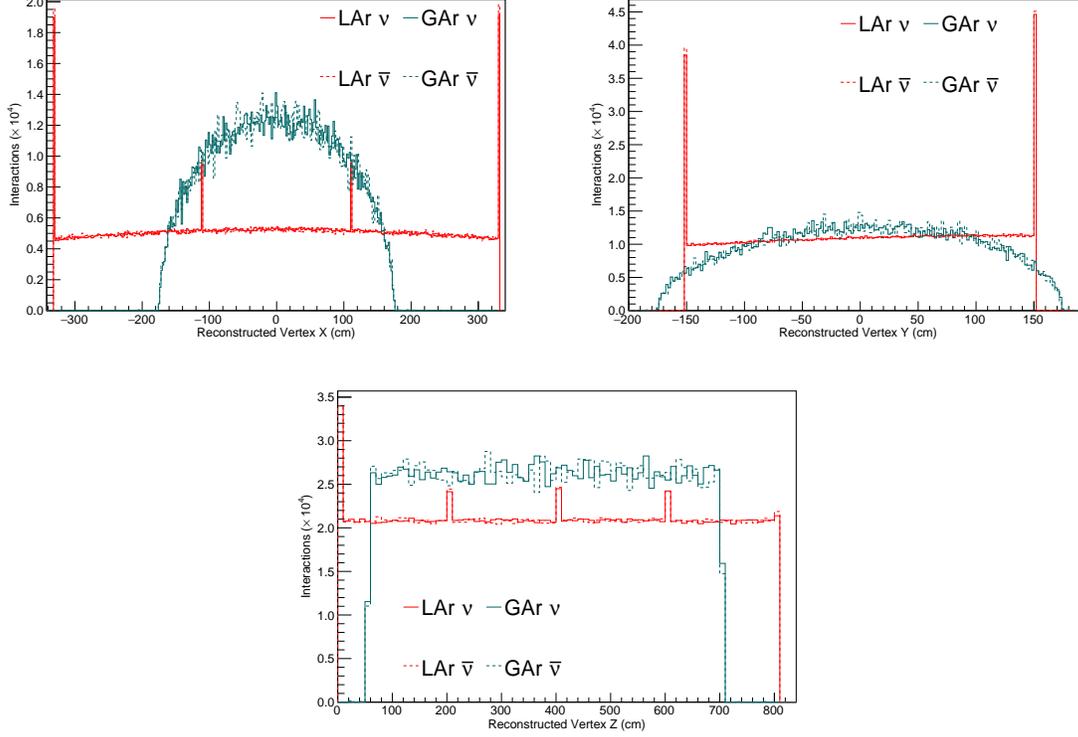


Figure 20: The vertex positions for the different Near Detector and beam configuration options. The X and Y positions are top left and right respectively, with the Z positions in the bottom plot.

- The acceptance of final state particles as a function of energy and direction
- The fraction of neutrino interactions on each species of nuclear target:
 Tab. 2 shows which nuclear targets are present in each sample as well as the fraction of interactions on each. The GArTPC option recorded interactions only in the volume of the TPC, which explains the fact that all interactions are on argon. The vast majority of the interactions in the FGT are on carbon.
- The fraction of energy contained in the detector as a function of the vertex distance from an edge of the detector
- The ability to distinguish different interaction types (NC, beam electron neutrino, muon neutrino/antineutrino) as a function of energy:
 Figure 22 shows the fraction of correctly identified interaction types as a function of energy for each of the detector and beam configurations. The fraction is unity for each of the configurations in this iteration because the full reconstruction is not available for the detector technologies and the truth information is used to make this figure.
- The energy thresholds for observing different particle species (proton, neutron, pion)

12 Bibliography

[1] Beam optimization task force interim report, <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1238>.

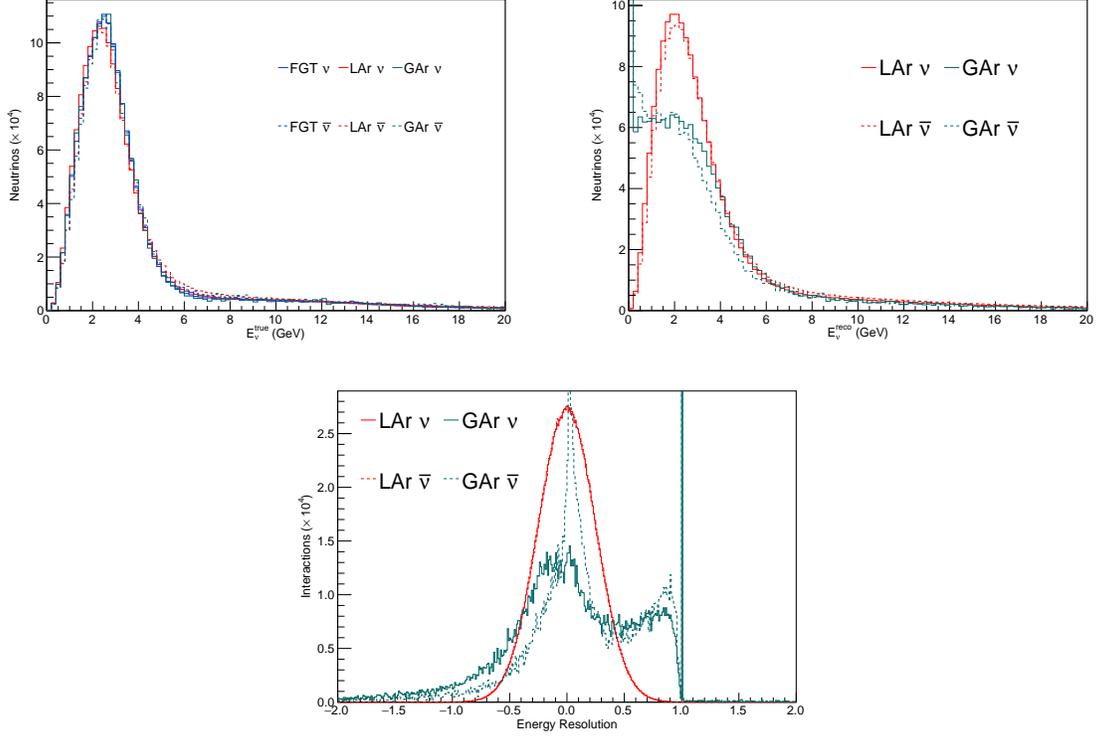


Figure 21: Energy spectra for selected interactions from the Monte Carlo simulated truth information (top left) and the reconstruction (top right) for the different detector and beam options. The energy resolution is also shown (bottom).

- [2] L. Aliaga, Neutrino Flux Prediction for the NuMI Beamline, PhD thesis, College of William and Mary, 2016, <http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2016-03.pdf>.
- [3] A. Bashyal, <https://indico.fnal.gov/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=12707>.
- [4] Lbne beam alignment tolerances and systematic uncertainties, <http://lbne2-docdb.fnal.gov:8080/cgi-bin/ShowDocument?docid=8410>.
- [5] Lbnf beam alignment parameters and dune neutrino flux uncertainties, <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1486>.
- [6] VALOR dune joint oscillation and systematics constraint fit (version 2016a), <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1291>.
- [7] VALOR dune joint oscillation and systematics constraint fit (version 2016b), <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1712>.

| Detector | A | Z | Fraction of Interactions |
|----------|----|----|--------------------------|
| FGT | 12 | 6 | 0.88 |
| | 14 | 7 | 0.02 |
| | 16 | 8 | 0.04 |
| | 40 | 18 | 0.03 |
| | 40 | 20 | 0.02 |
| | 56 | 26 | 0.01 |
| LArTPC | 12 | 6 | 0.00 |
| | 40 | 18 | 0.90 |
| | 52 | 24 | 0.02 |
| | 56 | 26 | 0.07 |
| | 59 | 28 | 0.01 |
| GArTPC | 40 | 18 | 1.00 |

Table 2: Nuclear targets from neutrino interactions in each of the candidate detector technologies. The GasTPC option only records interactions in the TPC volume which is why the fraction of interactions on argon is unity.

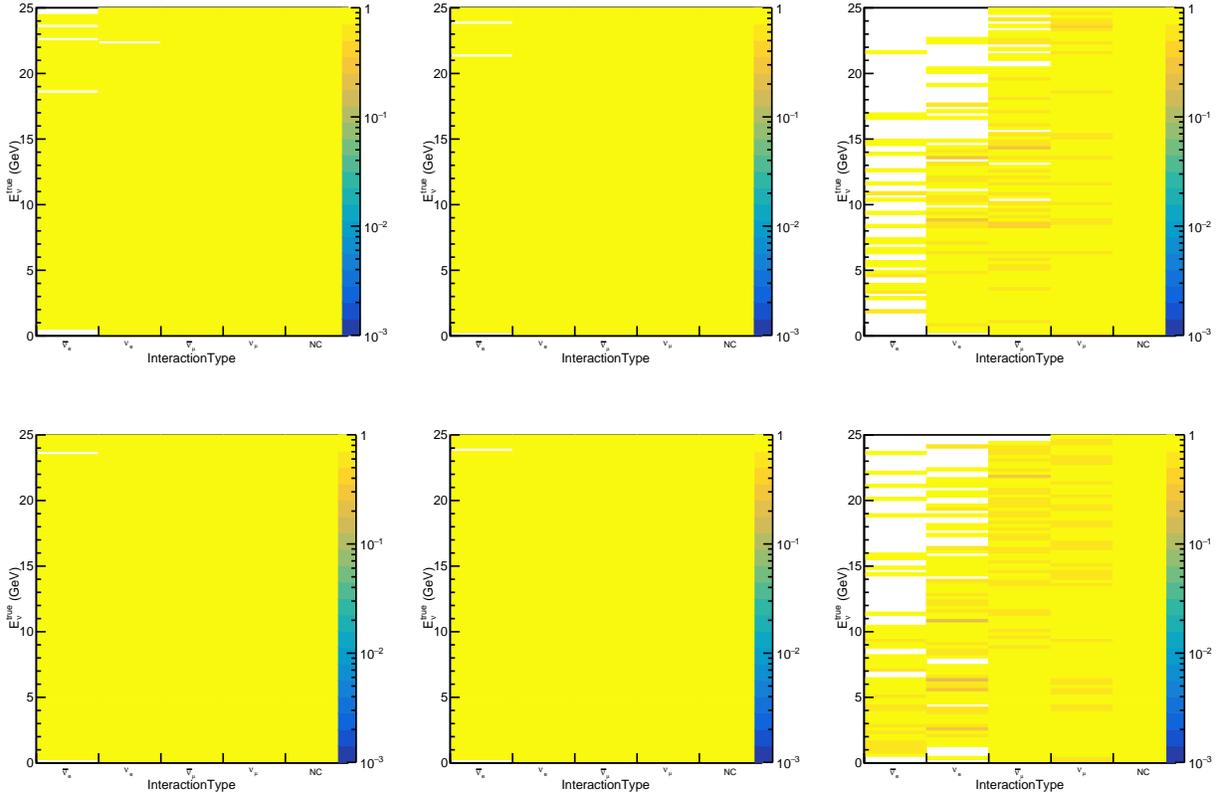


Figure 22: Fraction of correctly identified interaction types as a function of true neutrino energy for the FGT (left), LArTPC (center) and GArTPC (right). The top row is for the neutrino beam and the bottom row is for the anti-neutrino beam.