Low exposure long-baseline neutrino oscillation sensitivity of the DUNE experiment

Abed Abud, A.; DUNE Collaboration; Gupta, V.

DOI
10.1103/PhysRevD.105.072006

Publication date
2022

Document Version
Final published version

Published in
Physical Review D. Particles, Fields, Gravitation, and Cosmology

License
CC BY

Citation for published version (APA):
Low exposure long-baseline neutrino oscillation sensitivity of the DUNE experiment

A. Abed Abud et al.*
(DUNE Collaboration)

(Received 8 September 2021; accepted 14 March 2022; published 25 April 2022)

The Deep Underground Neutrino Experiment (DUNE) will produce world-leading neutrino oscillation measurements over the lifetime of the experiment. In this work, we explore DUNE’s sensitivity to observe charge-parity violation (CPV) in the neutrino sector, and to resolve the mass ordering, for exposures of up to 100 kiloton-megawatt-years (kt-MW-CY), where calendar years include an assumption of 57% accelerator uptime based on past accelerator performance at Fermilab. The analysis includes detailed uncertainties on the flux prediction, the neutrino interaction model, and detector effects. We demonstrate that DUNE will be able to unambiguously resolve the neutrino mass ordering at a 4σ (5σ) level with a 66 (100) kt-MW-CY far detector exposure, and has the ability to make strong statements at significantly shorter exposures depending on the true value of other oscillation parameters, with a median sensitivity of 3σ for almost all true δCP values after only 24 kt-MW-CY. We also show that DUNE has the potential to make a robust measurement of CPV at a 3σ level with a 100 kt-MW-CY exposure for the maximally CP-violating values δCP = ±π/2. Additionally, the dependence of DUNE’s sensitivity on the exposure taken in neutrino-enhanced and antineutrino-enhanced running is discussed. An equal fraction of exposure taken in each beam mode is found to be close to optimal when considered over the entire space of interest.

DOI: 10.1103/PhysRevD.105.072006

I. INTRODUCTION

The Deep Underground Neutrino Experiment (DUNE) [1] is a next-generation long-baseline neutrino oscillation experiment which will utilize high-intensity νμ and ¯νμ beams with peak neutrino energies of ≈2.5 GeV over a 1285 km baseline to carry out a detailed study of neutrino mixing. Some of DUNE’s key scientific goals are the definitive determination of the neutrino mass ordering, the definitive observation of charge-parity symmetry violation (CPV) for more than 50% of possible true values of the charge-parity violating phase, δCP, and the precise measurement of other three-neutrino oscillation parameters. These measurements will help guide theory in understanding if there are new symmetries in the neutrino sector and whether there is a relationship between the generational structure of quarks and leptons [2]. Observation of CPV in neutrinos would be an important step in understanding the origin of the baryon asymmetry of the Universe [3,4]. DUNE has a rich physics program beyond the three-neutrino oscillation accelerator neutrino program described here. These include beyond standard model searches [5], supernova neutrino detection [6], and solar neutrino detection [7]. Additional physics possibilities with DUNE are discussed in Refs. [8,9].

Neutrino oscillation experiments have so far measured five of the three-neutrino mixing parameters [10–12]: the three mixing angles θ12, θ23, and θ13, and the two squared-mass differences Δm21 and |Δm23|, where Δm2j = m2j − m2j is the difference between the squares of the neutrino masses. The neutrino mass ordering (the sign of Δm23) is not currently known, though recent results show a weak preference for the normal ordering (NO), where Δm23 > 0, over the inverted ordering (IO) [13–15]. The value of δCP is not well known, though neutrino oscillation data are beginning to provide some information on its value [13,16].

The oscillation probability of νμ → νe through matter in the standard three-flavor model and a constant matter density approximation can be written as [17]

\[
P(\bar{\nu}_\mu \to \nu_e) \approx \sin^2\theta_{23}\sin^2\theta_{13}\frac{\sin^2(\Delta_{31} - aL)}{(\Delta_{31} - aL)^2}\Delta_{31}^2
\]

\[+ \sin 2\theta_{23}\sin 2\theta_{13}\sin 2\theta_{12}\frac{\sin(\Delta_{31} - aL)}{(\Delta_{31} - aL)}\Delta_{31}
\]

\[\times \frac{\sin(aL)}{(aL)}\Delta_{21}\cos(\Delta_{31} \pm \delta_{CP})
\]

\[+ \cos^2\theta_{23}\sin^2\theta_{12}\frac{\sin^2(aL)}{(aL)^2}\Delta_{21}^2,
\]

1

*Corresponding author.
cwilkinson@lbl.gov

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP3.
where
\[ a = \pm \frac{G_F N_e}{\sqrt{2}} \approx \pm \frac{1}{3500 \text{ km}} \left( \frac{\rho}{3.0 \text{ g/cm}^3} \right), \]

\[ aL \approx 0.367, \quad G_F \text{ is the Fermi constant, } N_e \text{ is the number density of electrons in Earth’s crust, } \Delta_{ij} = 1.267 \Delta m^2_{ij} L / E_\nu, \Delta m^2_{ij} \text{ is in eV}^2, \quad L \text{ is the baseline in kilometers, and } E_\nu \text{ is the neutrino energy in GeV. Both } \delta_{CP} \text{ and } a \text{ terms are positive (negative) for } \nu_\mu \rightarrow \nu_e (\bar{\nu}_\mu \rightarrow \bar{\nu}_e) \text{ oscillations}. \]

The matter effect asymmetry encapsulated in the \( a \) terms arises from the presence of electrons and absence of positrons in Earth’s crust [18,19]. In the analysis described here, the oscillation probabilities are calculated exactly [20].

DUNE has published sensitivity estimates [21] to CPV and the neutrino mass ordering, as well as other oscillation parameters, for large exposures which show the ultimate sensitivity of the experiment. Sophisticated studies with a detailed treatment of systematic uncertainties were carried out only at large exposures. In this work, DUNE’s sensitivity at low exposures is explored further, with a detailed systematics treatment, including an investigation into how the run plan may be optimized to enhance sensitivity to CPV and/or mass ordering. It is shown that DUNE will produce world-leading results at relatively short exposures, which highlights the need for a high-performance near detector complex from the beginning of the experiment.

The DUNE far detector (FD) will ultimately consist of four modules, each with a 17 kt total mass. The neutrino beamline has an initial design intensity of 1.2 MW, with a planned upgrade to 2.4 MW. We assume a combined yearly Fermilab accelerator and neutrino beam-line uptime of 57% [8] and define kiloton-megawatt-calendar years (kt-MW-CY) as the exposure that would be collected by DUNE per kiloton of FD mass per megawatt of beam power per calendar year of nominal running (not per year of 100% uptime). This is for ease of interpretation, so the reader can easily translate the given exposures into expected years of DUNE operation.\(^1\) As the FD deployment schedule and beam power scenarios are both subject to change, the results shown in this work are consistently given in terms of exposure in units of kt-MW-CY, which is agnostic to the exact staging scenario, but can easily be expressed in terms of experiment years for any desired scenario. For example, with two FD modules, assuming a fiducial mass of 10 kt and a beam intensity of 1.2 MW, exposure would accumulate at a rate of 24 kt-MW-CY per calendar year, although a ramp-up in beam power is expected before reaching the design intensity in early running. In this work, the single-phase horizontal drift technology is assumed for all FD modules (see Sec. II D), which is a necessary simplification, but alternative technologies which may have slightly different performance are under investigation for some FD modules.

The analysis framework is described in Sec. II, including a description of the flux, neutrino interaction and detector models and associated uncertainties. A study of the dependence of the sensitivity to CPV and mass ordering on the fraction of data collected in neutrino-enhanced or antineutrino-enhanced running is given in Sec. III. A detailed study of the CPV and mass ordering sensitivities at low exposures are described in Secs. IV and V, respectively. Finally, conclusions are presented in Sec. VI.

II. ANALYSIS FRAMEWORK

This work uses the flux, neutrino interaction and detector model described in detail in Ref. [21], implemented in the CAFAna framework [22]. This section provides an overview of the key analysis features. Further details on all aspects can be found in Ref. [21].

A. Neutrino flux

DUNE will operate with two different beam modes, which depend on the polarity of the electromagnetic horns used to focus secondary particles produced after protons from the primary beamline interact in the target. Forward horn current (FHC) corresponds to neutrino-enhanced running, and reverse horn current (RHC) corresponds to antineutrino-enhanced running. In both FHC and RHC there are significant contributions from neutrinos with energies between 0.5 and 6 GeV, with a flux peak at \( \approx 2.5 \text{ GeV} \). The neutrino flux prediction is generated with G4LBNF [8,23], using the LBNF optimized beam design [8]. Flux uncertainties are due to the production rates and kinematic distributions of hadrons produced in the target and the parameters of the beamline, such as horn currents and horn and target positioning (“focusing uncertainties”) [8]. They are evaluated using current measurements of hadron production and estimates of alignment tolerances, giving flux uncertainties of approximately 8% at the first oscillation maximum, which are highly correlated across energy bins and neutrino flavors. A flux covariance as a function of neutrino energy, beam mode, detector, and neutrino species is generated with a “toy throw” approach, which is built using variations (“throws”) of the systematics propagated through the full beamline simulation. To reduce the number of parameters used in the fit, the covariance matrix is diagonalized, and each principal component is treated as an uncorrelated nuisance parameter. Only the first \( \approx 30 \) principal components (out of 108) were found to have a significant effect in the analysis and were included. The shapes of the unoscillated fluxes at the ND and FD are

---

\(^1\)In previous publications, DUNE has referred to the same quantity as kiloton-megawatt-years (kt-MW-yr). Here we use “calendar years” to make clear that we are referring to elapsed time rather than live time, at the request of the referee. We regret any confusion this change may cause.
similar, and the differences between them are understood at the percent level.

B. Neutrino interaction model

The interaction model used is based on GENIE v2.12.10 [24,25], although the combination of models used is much closer to some of the physics tunes available with GENIE v3.0.06, including a number of uncertainties beyond those provided by either GENIE version. These are motivated by data, although the available (anti)neutrino data taken on argon targets are sparse, leading to an uncertainty model that relies in a number of places on light target (mostly hydrocarbon) data. Variations in the cross sections are implemented either using GENIE reweighting parameters or with ad hoc weights of events designed to parametrize uncertainties or cross-section corrections currently not implemented within GENIE. The latter were developed using alternative generators or GENIE configurations, or custom weightings using the NUISANCE package [26]. Further details about the uncertainties used can be found in Ref. [21] (Sec. III).

The nuclear model which describes the initial state of nucleons in the nucleus is the Bodek-Ritchie global Fermi gas model [27], which includes empirical modifications to the nucleon momentum distribution to account for short-range correlation effects. The quasielastic model uses the Llewellyn-Smith formalism [28] with a simple dipole axial form factor, and BBBA05 vector form factors [29]. Nuclear screening effects and uncertainties are included based on the T2K 2017/8 parametrization [30] of the Valencia group’s [31,32] random phase approximation model. The Valencia model of the multinucleon, 2p2h, contribution to the cross section [31,32] is used, as described in Ref. [33]. Both MINERvA [34] and NOvA [35] have shown that this model underpredicts observed event rates on carbon at relevant neutrino energies for DUNE. Modifications to the model are constructed to produce agreement with MINERvA CC-inclusive data [34], which are used in the analysis to introduce additional uncertainties on the 2p2h contribution, with energy-dependent uncertainties, and extra freedom between neutrinos and antineutrinos. Uncertainties are added on scaling the 2p2h prediction from carbon to argon on electron-scattering measurements of short-range correlated pairs taken on multiple targets [36], separately for neutrinos and antineutrinos. GENIE uses a modified version of the Rein-Sehgal (R-S) model for pion production [37]. A data-driven modification to the GENIE model is included based on reanalyzed neutrino-deuterium bubble chamber data [38,39]. The deep inelastic scattering (DIS) model implemented in GENIE uses the Bodek-Yang parametrization [40] and GRV98 parton distribution functions [41]. Hadronization is described by the Andreopoulos-Kehayias-Gallagher-Yang model [42], which uses the Koba-Nielsen-Olesen scaling model [43] for invariant masses \( W \leq 2.3 \) GeV and PYTHIA6 [44] for invariant masses \( W \geq 3 \) GeV, with a smooth transition between the two models for intermediate invariant masses. Additional uncertainties developed by the NOvA Collaboration [45] to describe their resonance to DIS transition region data are also included. The final state interaction model and uncertainties available in GENIE are retained [46–48].

The cross sections include terms proportional to the lepton mass, which are significant at low energies where quasielastic processes dominate. Some of the form factors in these terms have significant uncertainties in the nuclear environment. Separate (and anticorrelated) uncertainties on the cross-section ratio \( \sigma_\nu / \sigma_\bar{\nu} \) for neutrinos and antineutrinos are adopted from Ref. [49]. Additionally, some \( \nu_e \) charged-current (CC) interactions occur at four-momentum transfers where \( \nu_\mu \) CC interactions are kinematically forbidden and so cannot be constrained by \( \nu_\mu \) cross-section measurements. To reflect this, a 100% uncertainty is applied in the phase space present for \( \nu_\mu \) but absent for \( \nu_\bar{\nu} \).

Systematic effects beyond what can be obtained by shifting parameters of the GENIE model can be studied by fitting simulated data produced with entirely distinct interaction generators. In general, the fit is not able to exactly reproduce the alternate simulated data, resulting in biases to extracted oscillation parameters. Demonstrating that these biases are small is an important test of the robustness of the systematic uncertainties. Such a study has been performed using simulated data based on the NuWro generator [50]. Controlling these effects is critical for the long-term physics goals of DUNE, and Ref. [9] describes how near detector measurements can be used to mitigate the discrepancy observed between the reference model and the alternate simulated data. For the early physics milestones presented in this paper, the sensitivity is limited by far detector statistics, and the impact of additional uncertainties to cover these out-of-model effects is expected to be small, although they are not explicitly included. A detailed treatment of these issues can be expected in future works.

C. Near detector simulation and reconstruction

The near detector (ND) hall will be located 574 m downstream of the proton target and \( \approx 60 \) m underground. The reference design for the DUNE ND system is fully described in Ref. [9] and consists of a liquid argon (LAr) time projection chamber (TPC) referred to as ND-LAr, a magnetized high-pressure gaseous argon TPC (ND-GAr), and an on-axis beam monitor (SAND). Additionally, ND-LAr and ND-GAr are designed to move perpendicular to the beam axis in order to take data at various off-axis angles (the DUNE-PRISM technique). ND-LAr is a modular detector based on the ArgonCube design [51–53], with a total active LAr volume of 105 m³ (a LAr mass of 147 tons). ND-GAr is implemented in this analysis as a cylindrical TPC filled with a 90/10 mixture of argon and CH₄ at 10 bar, surrounded by a granular, high-performance electromagnetic calorimeter
Neutrino interactions are simulated in the active volume of ND-LAr. The propagation of neutrino interaction products through the ND-LAr and ND-GAr detector volumes is simulated using a GEANT4-based program [54]. As pattern recognition and reconstruction software has not yet been fully developed for the ND, this analysis uses a parameterized reconstruction based on the GEANT4 simulated energy deposits in active detector volumes.

Only CC-inclusive interactions originating in the LAr are considered in this analysis, with a fiducial volume (FV) which excludes 50 cm from the sides and upstream edge and 150 cm from the downstream edge of the active region, containing a total fiducial mass of 50 t. Most muons with kinetic energies greater than 1 GeV exit ND-LAr. Energetic forward-going muons pass into ND-GAr, where their momentum and charge are reconstructed by curvature. Muon energy is reconstructed by range for tracks that stop in the LAr, and the charge cannot be determined event by event. Events with muons that exit the LAr active volume and do not match to a track in ND-GAr are rejected, as the muon momentum is not well reconstructed. For FHC beam running, the wrong-sign background is small and the charge is assumed to be negative for all LAr-contained muons. For RHC beam running, a Michel electron is required at the end of these stopped tracks to suppress the wrong-sign $\mu^-$ by a factor of 4.

All generated muons and charged pions are evaluated as potential muon candidates. Tracks are classified as muons if their length is at least 1 m, and their mean energy deposit is less than 3 MeV/cm. The minimum length requirement imposes an effective threshold on the true muon kinetic energy of about 200 MeV but greatly suppresses potential neutral current (NC) backgrounds with low-energy, non-interacting charged pions. Charged-current events are required to have exactly one muon candidate, and if the charge is reconstructed by curvature, it must be of the appropriate sign. Hadronic energy in the ND is reconstructed by summing all charge deposits in the LAr active volume that are not associated with the muon. To remove events where the hadronic energy is badly reconstructed due to charged particles exiting the detector, a veto region is defined as the outer 30 cm of the active volume on all sides, and events with more than 30 MeV total energy deposited in the veto region are rejected. Only a fraction of neutron kinetic energy is typically observed (24% on average with large fluctuations), resulting in poor energy reconstruction of events with energetic neutrons. The reconstructed neutrino energy $E_{\nu}^{rec} = E_{\mu}^{rec} + E_{had}^{rec}$ is the sum of the reconstructed hadronic energy $E_{had}^{rec}$ and the reconstructed muon energy $E_{\mu}^{rec}$. The reconstructed inelasticity $y_{rec} = 1 - E_{\mu}^{rec}/E_{\nu}^{rec}$ is the fraction of the neutrino energy that is carried by hadrons.

**D. Far detector simulation and reconstruction**

The DUNE FD design consists of four separate LArTPC detector modules, each with a total LAr mass of 17 kt, installed $\approx$ 1.5 km underground at the Sanford Underground Research Facility (SURF) [55]. The technologies to be deployed for the four modules and their order of construction are still under investigation, so in this analysis, only the single-phase design with a horizontal drift [56] is used. In this design, signals from drift electrons in the 13.3 $\times$ 12.0 $\times$ 57.5 m$^3$ active volume are read out by $\approx$ 5 mm spaced wires in anode readout planes. Scintillation light produced at the time of the neutrino interaction is detected and used to reconstruct the start time of the electron drift. We have developed a full simulation chain, which generates neutrino events in a GEANT4 model of the FD geometry and simulates the electronics readout. We have developed a reconstruction package to calculate efficiencies and reconstructed neutrino energy estimators for the four CC-inclusive FD samples used in the analysis ($\nu_\mu$-like FHC, $\nu_e$-like FHC, $\bar{\nu}_\mu$-like RHC and $\bar{\nu}_e$-like RHC).

The electronics response to the ionization electrons and scintillation light is simulated in the wire planes and photon detectors, respectively. Algorithms are applied to remove the impact of the LArTPC electric field and electronics response from the raw detector signal to identify hits and to cluster hits that may be grouped together due to proximity in time and space. Clusters from different wire planes are matched to form high-level objects such as tracks and showers using the Pandora toolkit [57,58]. Event classification is carried out through image recognition techniques using a convolutional neural network [59] which classifies events as $\bar{\nu}_\mu$-CC, $\bar{\nu}_e$-CC, $\nu_\tau$-CC, and NC. The $\bar{\nu}_e$ and $\bar{\nu}_\mu$ efficiencies in both beam modes exceed 90% in the flux peak.

The neutrino energy for $\bar{\nu}_\mu$-CC ($\bar{\nu}_e$-CC) events is estimated by the sum of the energy of the longest reconstructed track (highest energy reconstructed electromagnetic shower) and the hadronic energy. For both event types, the hadronic energy is estimated from the charge of reconstructed hits that are not in the primary track or shower, and corrections are applied to each hit charge for recombination and electron lifetime effects. For $\bar{\nu}_e$-CC events, the energy of the longest track is estimated by range if the track is contained or by multiple Coulomb scattering if it is exiting. For 0.5–4 GeV neutrino energies, the observed neutrino energy resolution is 15%–20%. The muon energy resolution is 4% for contained tracks and 18%
for exiting tracks. The electron energy resolution is approximately $4\% \oplus 9\% / \sqrt{E}$. The hadronic energy resolution is 34\%.

### E. Detector systematics

Detector effects impact the event selection efficiency as well as the reconstruction of neutrino energy and inelasticity (the variables used in the oscillation fits). The main sources of detector systematic uncertainties are limitations of the expected calibration and modeling of particles in the detector. Important differences between the ND and FD LArTPC design, size, detector environment, and calibration strategy lead to uncertainties that do not fully correlate between the two detectors. The degree of correlation is under active study, but in this analysis they are treated as being completely uncorrelated. Detailed simulations of detector effects are under development. In this analysis, uncertainties on the energy scale, energy resolution, particle responses, and detector acceptance are included to encapsulate these effects. The absolute scale uncertainties shift the reconstructed energy distributions, while the resolution uncertainties narrow or broaden them.

An uncertainty on the overall energy scale is included in the analysis presented here, as well as particle energy scale and resolution uncertainties that are separate and uncorrelated between four particle classes: muons, charged hadrons, neutrons, and electromagnetic showers. In the ND, muons reconstructed by range in LAr and by curvature in the ND-GAr are treated separately and assigned uncorrelated uncertainties. For each class of particle, uncertainties on the energy scale are introduced as a function of the reconstructed particle energy $E$, with a constant term, a term proportional to $\sqrt{E}$, and a term proportional to $1 / \sqrt{E}$. A 10\% uncertainty on the energy resolution is also included and treated as uncorrelated between the four particle classes. The parameters produce a shift to the kinematic variables in an event, as opposed to simply assigning a weight to each simulated event. The scale of the uncertainties is motivated by what has been achieved in recent experiments, including calorimetric-based approaches (NOvA, MINERvA) and LArTPCs (LArIAT, MicroBooNE, ArgoNeut).

In addition to impacting energy reconstruction, the E-field model also affects the definition of the FD FV, which is sensitive to electron drift. An additional 1\% uncertainty is therefore included on the total fiducial mass, which is conservatively treated as uncorrelated between the $\bar{\nu}_\mu$ and $\bar{\nu}_e$ samples due to the potential distortion caused by large electromagnetic showers in the electron sample.

The FD is sufficiently large that acceptance is not expected to vary significantly as a function of event kinematics. However, the ND acceptance does vary as a function of both muon and hadronic kinematics due to various containment criteria. Uncertainties are evaluated on the muon and hadron acceptance at the ND based on the change in the acceptance as a function of muon kinematics and true hadronic energy.

### F. Sensitivity methods

Systematics are implemented in the analysis using one-dimensional response functions for each analysis bin, and oscillation weights are calculated exactly, in fine (50 MeV) bins of true neutrino energy. For a given set of inputs—flux, oscillation parameters, cross sections, detector energy response matrices, and detector efficiency—an expected event rate can be produced. Minimization is performed using the MINUIT [60] package.

Oscillation sensitivities are obtained by simultaneously fitting the $\nu_\mu$-like FHC, $\nu_e$-like FHC, $\bar{\nu}_\mu$-like RHC and $\bar{\nu}_e$-like RHC FD spectra along with the $\nu_\mu$ FHC and $\bar{\nu}_\mu$ RHC samples from the ND. In the studies, all oscillation parameters shown in Table I are allowed to vary. Gaussian penalty terms (taken from Table I) are applied to $\theta_{12}$, $\Delta m^2_{31}$, and the matter density $\rho$ of Earth along the DUNE baseline [61]. Some studies presented in this work include a Gaussian penalty term on $\theta_{13}$ (also taken from NuFIT 4.0, given in Table I), which is precisely measured by experiments sensitive to reactor antineutrino disappearance [62–64]. The remaining parameters, $\sin^2 \theta_{23}$, $\Delta m^2_{32}$, and $\delta_{CP}$ are allowed to vary freely, with no penalty terms. The penalty terms are treated as uncorrelated with each other and uncorrelated with other parameters.

Flux, cross-section, and FD detector parameters are allowed to vary in the fit but are constrained by a penalty term corresponding to the prior uncertainty. ND detector uncertainties are included via a covariance matrix based on the shape difference between ND prediction and the “data”

#### TABLE I. Central value and relative uncertainty of neutrino oscillation parameters from a global fit [65,66] to neutrino oscillation data. The matter density is taken from Ref. [61]. Because the probability distributions are somewhat non-Gaussian (particularly for $\theta_{23}$), the relative uncertainty is computed using $1 / 6$ of the $3\sigma$ allowed range from the fit, rather than $1 / 2$ of the $1\sigma$ range. For $\theta_{23}$, $\theta_{13}$, and $\Delta m^2_{31}$, the best-fit values and uncertainties depend on whether NO or IO is assumed. The best fit for $\delta_{CP}$ is used as a test point in the analysis, but no uncertainty is assigned.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Central value</th>
<th>Relative uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{12}$</td>
<td>0.5903</td>
<td>2.3%</td>
</tr>
<tr>
<td>$\theta_{23}$ (NO)</td>
<td>0.866</td>
<td>4.1%</td>
</tr>
<tr>
<td>$\theta_{23}$ (IO)</td>
<td>0.869</td>
<td>4.0%</td>
</tr>
<tr>
<td>$\theta_{13}$ (NO)</td>
<td>0.150</td>
<td>1.5%</td>
</tr>
<tr>
<td>$\theta_{13}$ (IO)</td>
<td>0.151</td>
<td>1.5%</td>
</tr>
<tr>
<td>$\Delta m^2_{21}$</td>
<td>$7.39 \times 10^{-5}$ eV$^2$</td>
<td>2.8%</td>
</tr>
<tr>
<td>$\Delta m^2_{31}$ (NO)</td>
<td>$2.451 \times 10^{-3}$ eV$^2$</td>
<td>1.3%</td>
</tr>
<tr>
<td>$\Delta m^2_{31}$ (IO)</td>
<td>$-2.512 \times 10^{-3}$ eV$^2$</td>
<td>1.3%</td>
</tr>
<tr>
<td>$\rho$</td>
<td>2.848 g cm$^{-3}$</td>
<td>2%</td>
</tr>
<tr>
<td>$\delta_{CP}$ (NO)</td>
<td>$-2.53$ (rad)</td>
<td>...</td>
</tr>
<tr>
<td>$\delta_{CP}$ (IO)</td>
<td>$-1.33$ (rad)</td>
<td>...</td>
</tr>
</tbody>
</table>
(which come from the simulation in this sensitivity study).

The covariance matrix is constructed with a throwing technique. For each “throw,” all ND energy scale, resolution, and acceptance parameters are simultaneously thrown according to their respective uncertainties, and the modified prediction is produced by varying the relevant quantities away from the nominal prediction according to their respective uncertainties, and the parameters are simultaneously thrown according to their nominal value (see Table I) except the parameters of interest, which are set to a test point. Then a fit is carried out in which all parameters can vary, constrained by their prior uncertainty where applicable. For the smallest exposures investigated with an Asimov study in this work, all samples have at least 100 events, satisfying the Gaussian approximation inherent in the Asimov method. Toy throwing studies are performed (in Secs. IV and V) in which an ensemble of systematic, oscillation parameter and statistical throws are made. Systematic throws are made according to their prior Gaussian uncertainties, oscillation parameters are randomly chosen as described in Table II, and Poisson fluctuations are then applied to all analysis bins, based on the mean event count for each bin after the systematic adjustments have been applied. For each throw in the ensemble, the test statistic is minimized, and the best-fit value of all parameters is determined. The expected resolution for parameters of interest are then determined from the spread in the distribution of their postfit values.

Asimov studies are computationally efficient and, for Gaussian parameters and uncertainties, give a good sense of the median sensitivity of an experiment. Toy throwing studies are computationally expensive, fully explore the parameter space, and make fewer assumptions about the behavior of parameters and uncertainties.

### G. Near and far detector samples and statistics

In this work, the sensitivity as a function of FD exposure is explored and results are reported in terms of kt-MW-CY, which does not assume any specific FD or beam intensity staging scenario. However, the ND used in this analysis (ND-LAr with a downstream muon spectrometer) is assumed not to be staged, and as such the ND sample size corresponding to a particular FD exposure will vary based on the staging scenario. The nominal staging scenario from Ref. [21] is therefore retained for the purpose of normalizing the ND samples at each FD exposure. In that scenario, a 7 year exposure corresponds to 336 kt-MW-CY at the FD and 480 t-MW-CY at the ND, summed over both beam modes. The ND statistics used in this analysis are scaled assuming this ratio throughout, using the same fraction of exposure in each beam mode as used at the FD. The ND samples used in this analysis are relatively quickly systematics limited in both beam modes, and so these approximations are unlikely to have a significant impact on the results.

The oscillation analysis presented here includes two CC-inclusive samples originating in the ND-LAr FV, an FHC $\nu_\mu$ and an RHC $\bar{\nu}_\mu$ sample. These samples are both binned in two dimensions, as a function of reconstructed neutrino
FIG. 1. ND samples in both FHC and RHC, shown in the reconstructed neutrino energy and reconstructed inelasticity binning ($y_{\text{rec}}$) used in the analysis, for a 105 t-MW-CY exposure (equivalent to a 100 kt-MW-CY exposure at the FD) with all relevant exposure assumptions including 57% accelerator uptime as described in the text, with an equal split between FHC and RHC. The size of the systematic uncertainty bands from all of the flux, cross-section and ND detector systematics used in the analysis are shown, as well as the postfit uncertainty bands obtained by performing an Asimov fit to the ND data. NC backgrounds and wrong-sign contributions to the total event rate are also shown. Statistical uncertainties are too small to be visible on this plot scale. (a) FHC (b) RHC.
energy and inelasticity, \( y_{\text{rec}} = 1 - E_{\mu}^{\text{rec}} / E_{\nu}^{\text{rec}} \). The sample distributions for both FHC and RHC are shown in Fig. 1 for an exposure of 105 t-MW-CY, corresponding to 100 kt-MW-CY at the far detector with the assumptions stated above. The size of the systematic uncertainty bands from all of the flux, cross-section and ND detector systematics used in the analysis and described above are shown, as well as the postfit uncertainty bands obtained by performing an Asimov fit to the ND data. It is clear that, even after a relatively small exposure of 105 t-MW-CY, the ND samples are very high statistics and are systematics limited in the binning used in the analysis. Backgrounds in the \((\bar{\nu}_\mu)^{\text{CC}}\) samples are also shown in Fig. 1. NC backgrounds are predominantly from NC \( \pi^\pm \) production where the pion leaves a long track and does not shower. Wrong-sign contamination in the beam is a background where the charge of the muon is not reconstructed, which particularly affects low reconstructed neutrino energies in RHC. The wrong-sign background is also larger at high reconstructed inelasticity \( y_{\text{rec}} \) due to the kinematics of neutrino and antineutrino scattering.

The expected FD FHC \( \nu_e \) and RHC \( \bar{\nu}_e \) samples are shown in Fig. 2 for a 100 kt-MW-CY total FD exposure, split equally between FHC and RHC beam modes. The systematic uncertainty bands with and without the ND constraint applied are shown, as well as the background contributions. There are contributions from both \( \nu_e \) and \( \bar{\nu}_e \) in both beam modes. The NC, intrinsic beam \( \bar{\nu}_e \) and wrong flavor contamination is also shown; the largest background comes from the intrinsic \( \bar{\nu}_e \) beam contribution in both modes. After a 50 kt-MW-CY exposure in FHC, the \( \nu_e \) sample statistical uncertainty is close to the systematic uncertainty before the ND constraint, although it is still clearly statistics limited when the ND constraint is applied. The \( \bar{\nu}_e \) sample is still strongly statistics limited after 50 kt-MW-CY exposure in RHC. The difference is largely due to the difference in the \( \nu_e \) and \( \bar{\nu}_e \) cross sections.

The expected FD FHC \( \nu_\mu \) and RHC \( \bar{\nu}_\mu \) samples are shown in Fig. 3 for a 100 kt-MW-CY total FD exposure, split equally between FHC and RHC beam modes. The systematic uncertainty bands with and without the ND constraint applied are shown, as well as the background contributions. Although the wrong-sign \( \nu_\mu \) contribution to the RHC \( \bar{\nu}_\mu \) sample is shown separately, it still provides useful information for constraining the oscillation parameters and is included in the analysis. The statistics are much higher than in Fig. 2; the statistical uncertainty on the \( \nu_\mu \) FHC sample is smaller than the systematic uncertainty band for some regions of phase space, even after the ND constraint is applied, although the statistical uncertainty is larger than the constrained systematic uncertainty in the “dip” region, around 2.5 GeV, which is likely to have the most impact on the analysis. The statistical uncertainty on the \( \bar{\nu}_\mu \) RHC sample is larger, again due to the smaller \( \bar{\nu}_\mu \) (than \( \nu_\mu \)) cross section and lower fluxes in RHC running. The statistical uncertainty around the 2.5 GeV dip region is significantly larger than the systematic uncertainty band, although, as for the FHC \( \nu_\mu \) sample, the statistical uncertainty is smaller than the systematics for some regions of the parameter space.
the dependence of the median CPV and mass ordering significances are studied, for different fractions of time spent in each beam mode, using the full analysis framework described in Sec. II.

Figure 4 shows DUNE’s Asimov sensitivity to CPV for a total 100 kt-MW-CY exposure, with different fractions of FHC and RHC running, at the NuFIT 4.0 best-fit value in both NO and IO (see Table I), shown with and without a penalty on $\theta_{13}$ applied. For each point tested, all oscillation and nuisance parameters are allowed to vary, and three fits are carried out, two where $\delta_{\text{CP}}$ is set to the CP-conserving values $\delta_{\text{CP}} = 0$ and $\delta_{\text{CP}} = \pm \pi$, the minimum of which is the CP-conserving best-fit value, and another where $\delta_{\text{CP}}$ is allowed to vary. The difference in the best-fit $\chi^2$ values is calculated:

$$\Delta \chi^2_{\text{CPV}} = \min \left\{ \chi^2_{\delta_{\text{CP}}=0}, \chi^2_{\delta_{\text{CP}}=\pm \pi} \right\} - \chi^2_{\text{CPV}},$$

and the square root of the difference is used as the figure of merit on the $y$ axis in Fig. 4. There are some caveats associated with this figure of merit, which are discussed in Sec. IV. A 100 kt-MW-CY exposure is shown as it was identified in Ref. [21] as the exposure at which DUNE’s median CPV significance exceeds $3\sigma$ at $\delta_{\text{CP}} = \pm \pi/2$, an important milestone in DUNE’s physics program (with equal beam mode running).

Figure 4 shows that, when the reactor constraint on $\theta_{13}$ is included, the sensitivity to CPV can be increased in some regions of $\delta_{\text{CP}}$ parameter space with more FHC than RHC running. However, this degrades the sensitivity in other regions, most notably for $\delta_{\text{CP}} > 0$ regardless of the true mass ordering. This is due to a degeneracy between $\delta_{\text{CP}}$ and the octant of $\sin^2 2\theta_{13}$ because both parameters impact the rate of $\nu_e$ appearance. The degeneracy is resolved by including antineutrino data; the octant of $\sin^2 2\theta_{13}$ affects the rate of $\bar{\nu}_e$ and $\bar{\nu}_e$ appearance in the same way, but the effect of $\delta_{\text{CP}}$ is reversed for antineutrinos.

For regions of phase space where the octant degeneracy does not affect the result (e.g., $\sin^2 \theta_{23} \approx 0.5$), there is no degradation in the sensitivity, and enhanced FHC running increases the sensitivity for all values of $\delta_{\text{CP}}$. Increasing the fraction of RHC decreases the sensitivity for the entire $\delta_{\text{CP}}$ range when the reactor $\theta_{13}$ constraint is included, relative to equal beam mode running. This is due to the lower statistics of the $\bar{\nu}_e$ sample (see Fig. 2) because of the reduced antineutrino flux and cross section. For short exposures, DUNE will not have a competitive independent measurement of $\theta_{13}$, so the main analysis will include the reactor $\theta_{13}$ constraint. Nonetheless, it is instructive to look at the results without the penalty applied. In this case, the sensitivity is severely degraded (as expected) for 100% running in either beam mode.

Figure 5 shows DUNE’s Asimov sensitivity to the mass ordering for a total 24 kt-MW-CY far detector exposure,
with different fractions of FHC and RHC running, and the same four true oscillation parameter sets. A 24 kt-MW-CY exposure is used in Fig. 5 as it is around the exposure at which DUNE’s median mass ordering significance exceeds 5σ for some values of δCP [21]. For each point tested, all oscillation and nuisance parameters are allowed to vary, and two fits are carried out, one using each ordering. The difference in the best-fit χ² values is calculated:

$$\Delta \chi^2_{MO} = \chi^2_{IO} - \chi^2_{NO},$$

and the square root of the difference is used as the figure of merit on the y axis in Fig. 5. There are some caveats associated with this figure of merit, which are discussed in Sec. V.

It is clear from Fig. 5 that the mass ordering sensitivity has a strong dependence on the fraction of running in each beam mode. As in the CPV case, the effect is very different with and without the reactor θ_{13} constraint included. If the true ordering is normal and the reactor θ_{13} penalty is applied, the sensitivity increases significantly with increasing FHC running, with a full 1σ increase in the sensitivity between equal beam running and 100% FHC for most

FIG. 4. The Asimov CPV sensitivity as a function of the true value of δCP, for a total exposure of 100 kt-MW-CY with different fractions of FHC and RHC running, with and without a θ_{13} penalty applied in the fit. Results are shown for both true normal and inverted ordering, with the true oscillation parameter values set to the NuFit 4.0 best-fit point in each ordering (see Table I). All exposures include an assumption of 57% accelerator uptime as described in the text. (a) NO, with θ_{13}-penalty (b) IO, with θ_{13}-penalty (c) NO, no θ_{13}-penalty (d) IO, no θ_{13}-penalty.
values of $\delta_{CP}$. Conversely, if the ordering is inverted, 100% FHC running would degrade the sensitivity by $\geq 1\sigma$ for all values of $\delta_{CP}$ at the NuFIT 4.0 best-fit point. Overall, the sensitivity to the inverted ordering is improved by a more equal split between the beam modes. It is clear that 100% RHC running gives poor sensitivity for all values tested.

Without the reactor $\theta_{13}$ constraint, the greatest sensitivity is obtained with close to an equal split of FHC and RHC running, and the sensitivity is significantly reduced with 100% FHC running. This is because of a degeneracy between the effect of $\theta_{13}$ and the mass ordering on the rate of $\nu_e$ appearance in FHC mode. If the mass ordering is normal, the $\nu_e$ rate in FHC will be enhanced; without the reactor constraint, this excess can be accommodated by increasing the value of $\theta_{13}$.

For comparison, Fig. 6 shows the Asimov CPV and mass ordering sensitivities, with and without the reactor $\theta_{13}$ constraint included, for true normal ordering only, for a large exposure of 336 kt-MW-CY, with different fractions of FHC and RHC running. At large exposures, running with strongly enhanced FHC no longer improves the sensitivity over equal beam mode running, with or without the $\theta_{13}$ penalty applied, for either CPV or mass ordering determination. This can be understood because the enhancement to the statistics that enhanced FHC brings is no longer as important to the sensitivity, and DUNE is
able to place a constraint on the value of $\theta_{13}$ with its own data.

Overall, the sensitivity to CPV and the mass ordering is dependent on the division of running time between FHC and RHC, but a choice that increases the sensitivity in some region of parameter space can severely decrease the sensitivity in other regions. If there is strong reason to favor, for example, normal over inverted ordering when DUNE starts to take data, Fig. 5 shows that this could be more rapidly verified by running with more FHC data than RHC data, as the reactor $\theta_{13}$ constraint will be used in the main low exposure analysis. However, if this choice is wrong, this might cause DUNE to take longer to reach the same significance. Clearly this is an important consideration which should be revisited shortly before DUNE begins to collect data. Similarly, the CPV sensitivity shown in Fig. 4 might be optimized if there is a strong reason to favor gaining sensitivity for $\delta_{CP} > 0$ or $\delta_{CP} < 0$, at a cost of reducing the sensitivity to CPV if $\delta_{CP}$ has the other sign. But, it is clear from Figs. 4 and 5 that equal running in FHC and RHC gives a close to optimal sensitivity across all of the parameter space and as such is a reasonable a priori choice of run plan for studies of the DUNE sensitivity. Additionally, it is clear from Fig. 6 that the improvement in
the sensitivity with unequal beam running is a feature at low exposures, but not at high exposures, particularly because at high exposures when DUNE is able to constrain all the oscillation parameters with precision [21], there is a stronger motivation to run a DUNE-only analysis, without relying on the reactor $\theta_{13}$ measurement.

IV. CP VIOLATION SENSITIVITY

In this section, CPV sensitivity results are presented. For simplicity, only true NO will be shown unless explicitly stated. In all cases, a joint ND + FD fit is performed, and a $\theta_{13}$ penalty is always applied to incorporate the reactor measurement, as described in Sec. II. An equal split between FHC and RHC running is assumed based on the results obtained in Sec. III. Asimov sensitivities, as shown in Sec. III, are instructive but do not give information on how the expected sensitivity may vary with statistical or systematic uncertainties or for variations in the other oscillation parameters of interest.

Figure 7 shows the significance with which CPV ($\delta_{CP} \neq \{0, \pm \pi\}$) can be observed for both NO and IO, for exposures of 66 and 100 kt-MW-CY. The sensitivity metric used is the square root of the difference between the best-fit $\chi^2$ values obtained for a CP-conserving fit and one where $\delta_{CP}$ is allowed to vary, as shown in Eq. (3), which is calculated for each throw of the systematic, other oscillation parameters and statistics.

The sensitivity shown in Fig. 7 has a characteristic double peak structure because the significance of a CPV measurement decreases around CP-conserving values. The systematic and statistical variations mean that all throws have $\Delta \chi^2_{CPV} \geq 0$, and therefore neither the median significance nor the band showing the central 68% of throws reach exactly 0 at CP-conserving values. This is entirely expected, it simply means that random variations in the data will cause us to obtain a $1\sigma$ measurement of CPV $\approx 32\%$ of the time for CP-conserving values. Median significances are slightly higher for IO than for NO, and by exposures of 100 kt-MW-CY, the median significance exceeds $3\sigma$ for the maximal CP-violating values of $\pm \pi/2$. This presentation of the CPV sensitivity was followed in Ref. [21] and is very informative at high exposures. Around CP-conserving values ($\delta_{CP} = \{0, \pm \pi\}$), the distribution of the sensitivity metric $\sqrt{\Delta \chi^2_{CPV}}$ is non-Gaussian for all exposures. Additionally, at lower exposures, as shown in Fig. 7, the distribution of $\sqrt{\Delta \chi^2_{CPV}}$ around maximally CP-violating values of $\delta_{CP} = \pm \pi/2$ is increasingly non-Gaussian, making the spread in sensitivity harder to interpret with this presentation.

The CPV significance in Fig. 7 (and previously in Ref. [21]) is calculated using constant $\Delta \chi^2_{CPV}$ critical values, where $\Delta \chi^2_{CPV} \leq 1, 4, 9$ corresponds to a significance of $1, 2$ and $3\sigma$ for one degree of freedom. This assumption holds when Wilks’ theorem can be applied [69] but can lead to incorrect coverage where it cannot. It is known to break down for low-statistics samples, around physical boundaries, in the case of cyclic parameters, and where there are
significant degeneracies. It is likely that a constant \( \Delta \chi^2 \) treatment will break down for \( \delta_{\text{CP}} \), where all of these issues apply, as has indeed been shown by the T2K Collaboration [13].

The Feldman-Cousins method [70] is a brute force numerical method to calculate confidence intervals with correct coverage. A large number of toy experiments are produced, where the parameter(s) of interest (here \( \delta_{\text{CP}} \)) is set to a desired true value, all other systematic and oscillation parameters are thrown, as described in Sec. II, and a statistical throw is made, for the two ND samples and four FD sample used in the analysis. Then two fits are performed, one where the parameter(s) of interest are fixed to the true value and another where the test statistic is minimized with respect to the parameter(s) of interest. In both fits, all other parameters are allowed to vary. For each throw, the profile likelihood ratio \( \Delta \chi^2_{\text{FC}} \) is calculated using the minimum \( \chi^2 \) values for those two fits, as in Eq. (5):

\[
\Delta \chi^2_{\text{FC}} = \chi^2(\theta_{\text{true}}) - \min_{\theta} \chi^2(\theta). \tag{5}
\]

The distribution of these throws is used to calculate the \( \Delta \chi^2_{\text{FC}} \) value that gives the desired coverage, with the appropriate fraction of toys above or below the calculated value. These are labeled critical values and are denoted \( \Delta \chi^2_c \). A distribution of \( \Delta \chi^2_{\text{FC}} \) values is shown in Fig. 8 for an example ND + FD analysis with a 100 kt-MW-CY exposure at the far detector, equal FHC and RHC run fractions, and the reactor \( \theta_{13} \) constraint applied. In Fig. 8, the \( \Delta \chi^2_c \) values corresponding to 68.27% (1\( \sigma \)), 90%, 95.45% (2\( \sigma \)) and 99.73% (3\( \sigma \)) of the throws are indicated. The \( \Delta \chi^2_c \) values were only calculated up to the 3\( \sigma \) level due to the very large number of throws required for higher confidence levels.

An uncertainty on the value of \( \Delta \chi^2_c \) obtained from the toy throw distribution (e.g., Fig. 8), is obtained using a bootstrap rethrowing method [71]. The empirical probability density function (PDF) obtained from the throws is treated as the true PDF, and \( B \) independent samples of size \( n \) are drawn from it, where \( n \) is the total number of throws used to build the empirical PDF. Each throw can be drawn multiple times in this method, so the ensemble of throws is different in each sample. Then, the standard deviation \( \sigma_b \), on the \( \Delta \chi^2_c \) values of interest, \( \theta \), are calculated for each of the \( B \) samples using

\[
\sigma_b = \sqrt{\frac{1}{B-1} \sum_{i=0}^{B} (\theta^*_i - \bar{\theta}^*)^2}, \tag{6}
\]

where \( \theta^*_i \) denotes the calculated \( \Delta \chi^2 \) value of interest for each of the samples and \( \bar{\theta}^* \) is their average value.

Figure 9 shows the evolution of the \( \Delta \chi^2_c \) values as a function of exposure for \( \delta_{\text{CP}} = 0 \), the relevant value for

**FIG. 8.** Distribution of \( \Delta \chi^2_{\text{FC}} \) values, calculated using Eq. (5), for a large number of throws with true \( \delta_{\text{CP}} = 0 \), and a 100 kt-MW-CY exposure. The \( \Delta \chi^2_c \) values (vertical lines) obtained using the Feldman-Cousins method show the \( \Delta \chi^2_{\text{FC}} \) value below which 68.27% (1\( \sigma \)), 90%, 95.45% (2\( \sigma \)) and 99.73% (3\( \sigma \)) of the throws reside, with the calculated values given in the legend. The number of throws used is also given. All exposures include an assumption of 57% accelerator uptime as described in the text.

**FIG. 9.** The \( \Delta \chi^2_c \) values corresponding to 68.27% (1\( \sigma \)), 90%, 95.45% (2\( \sigma \)) and 99.73% (3\( \sigma \)) of throws, shown for true \( \delta_{\text{CP}} = 0 \), as a function of exposure. A linear x-axis scale is used to highlight the stability of \( \Delta \chi^2_c \) values for large exposures. The uncertainty on the \( \Delta \chi^2_c \) values is obtained using Eq. (6) and is indicated as the shaded line. To guide the eye, horizontal dashed lines are included which indicate the 1\( \sigma \), 90%, 2\( \sigma \) and 3\( \sigma \) \( \Delta \chi^2_c \) values assumed using the constant-\( \Delta \chi^2 \) method, with one degree of freedom. The distribution of throws used produced to calculate the \( \Delta \chi^2_c \) values shown are given in Fig. 19. All exposures include an assumption of 57% accelerator uptime as described in the text.
CPV sensitivity, for an ND + FD analysis with equal FHC and RHC running and the reactor $\theta_{13}$ constraint applied. For all significance levels tested, the $\Delta\chi^2$ rise quickly as a function of exposure and stabilize at values slightly higher than those suggested by the constant $\Delta\chi^2$ method by exposures of $\approx 100$ kt-MW-CY. The initial rise in the $\Delta\chi^2$ values is due to the low statistics at those exposures. Overall, this implies that the CPV significance is slightly weaker than what would be inferred from $\sigma = \sqrt{\Delta\chi^2_{CP}}$, as used for example in Fig. 7. Crucially, there is no constant increase in the $\Delta\chi^2$ values over time as has been reported by the T2K Collaboration [13]. Details on the number of toy throws used at each point of Fig. 9 are given in the Appendix, and the toy throw distributions from which the $\Delta\chi^2$ values and their uncertainties were calculated are shown in Fig. 19.

As $\delta_{CP}$ is a cyclical parameter, with physical boundaries at $\pm \pi$, it is interesting to see how the $\Delta\chi^2$ values evolve as a function of it. Figure 10 shows the $\Delta\chi^2$ as a function of true $\delta_{CP}$, for an ND + FD analysis with equal FHC and RHC running including the reactor $\theta_{13}$ constraint, for both 100 and 336 kt-MW-CY exposures. There is a noticeable, although not large, depression in the $\Delta\chi^2$ values at $\delta_{CP} = \pm \pi/2$ for all significance levels considered. This effect is larger at the lower, 100 kt-MW-CY, exposure and is larger at higher significance levels. It is also clear from Fig. 10 that the $\Delta\chi^2$ behavior is very similar at $\delta_{CP} = \pm \pi/2$ as at $\delta_{CP} = 0$. Although the $\Delta\chi^2$ values are relevant for CPV sensitivity, this evolution of the $\Delta\chi^2$ values with $\delta_{CP}$ will be important for estimating DUNE’s $\delta_{CP}$ resolution. Details on the number of toy throws used at each point of Fig. 10 are given in the Appendix, and the toy throw distributions used to calculate the $\Delta\chi^2$ values and uncertainties are shown for the 100 kt-MW-CY (336 kt-MW-CY) test points in Fig. 20 (Fig. 21).

DUNE’s CPV sensitivity is calculated using Eq. (3) from an ensemble of throws of all systematic, other oscillation parameters and statistics. Figure 11 shows the fraction of throws for which DUNE would observe a CPV significance above a discrete threshold, as a function of the true value of $\delta_{CP}$, for 1–3$\sigma$ significances and for a variety of exposures. The shaded histograms show the complete treatment including FC, while the dashed histograms show the constant-$\Delta\chi^2$ treatment, to show the deviation from Wilks’ theorem, and to facilitate comparison at higher significances where the FC treatment becomes computationally prohibitive.

The point at which the median significance (50% of throws) passes different significance thresholds can be easily read from the figures and can be compared with those shown in Fig. 7. The same double peak structure seen in Fig. 7 can be observed. The median significance for measuring CPV exceeds 3$\sigma$ after $\approx 100$ kt-MW-CY at $\delta_{CP} = \pm \pi/2$, but a significant fraction of throws exceed 3$\sigma$ at shorter exposures for those values. For a 336 kt-MW-CY exposure, the fraction of throws for which the significance is less than 3$\sigma$ at maximal values of $\delta_{CP}$ is very small. In general, the effect of the Feldman-Cousins correction is to reduce the fraction of toy throws that cross each significance threshold (with respect to the constant-$\Delta\chi^2$ result), by a maximum of $\approx 10\%$, but the exact fraction changes as a function of true $\delta_{CP}$ value and exposure. An exception to this general trend is the 3$\sigma$ behavior at 24 kt-MW-CY, the lowest exposure shown, where the significance increases. This is due to fall...
FIG. 11. Fraction of throws for which significance of DUNE’s CP-violation test ($\delta_{CP} \neq 0, \pm \pi$) exceeds 1–3$\sigma$, calculated using both the FC (shaded histograms) and constant-$\Delta \chi^2$ (dashed lines) methods, as a function of the true value of $\delta_{CP}$. Shown for NO, for a number of different exposures. The number of throws used to make each figure is also shown. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) 24 kt-MW-CY (b) 66 kt-MW-CY (c) 100 kt-MW-CY (d) 150 kt-MW-CY (e) 197 kt-MW-CY (f) 336 kt-MW-CY.

FIG. 12. Fraction of throws for which the significance of DUNE’s CP-violation test ($\delta_{CP} \neq 0, \pm \pi$) exceeds 1–3$\sigma$, for $\delta_{CP} = -\pi$ and for 50% of $\delta_{CP}$ values, calculated with the FC (solid lines) and constant-$\Delta \chi^2$ (dashed lines) methods, as a function of exposure. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) $\delta_{CP} = -\pi/2$ (b) 50% of $\delta_{CP}$ values.
FIG. 13. Fraction of throws for which the significance of DUNE’s CP-violation test ($\delta_{CP} \neq \{0, \pm \pi\}$) exceeds 1–5$\sigma$, as a function of the true value of $\delta_{CP}$. Shown for NO, for a number of different exposures. The number of throws used to make each figure is also shown. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) 24 kt-MW-CY (b) 66 kt-MW-CY (c) 100 kt-MW-CY (d) 150 kt-MW-CY (e) 197 kt-MW-CY (f) 336 kt-MW-CY.

FIG. 14. Fraction of throws for which the significance of DUNE’s CP-violation test ($\delta_{CP} \neq \{0, \pm \pi\}$) exceeds 1–5$\sigma$, both assuming $\delta_{CP} = -\pi/2$, and for 50% of $\delta_{CP}$ values, shown as a function of exposure, for NO. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) $\delta_{CP} = -\pi/2$ (b) 50% of $\delta_{CP}$ values.
in the $3\sigma$ $\Delta\chi^2$ value toward the lowest exposures observed in Fig. 9. The number of throws carried out at each exposure is indicated on each plot. The number of throws decreases as a function of exposure because fixed computing resources were used for each configuration, and the time for the ensemble of fits carried out for each throw to complete increases slightly with exposure. The final 336 kt-MW-CY exposure has more throws because it was generated for the analysis presented in Ref. [21], where more than one projection was considered—requiring more throws to sample the space.

Figure 12 shows the fraction of throws which exceed different significance thresholds at the maximal $\delta_{CP} = -\pi/2$ and for $50\%$ of $\delta_{CP}$ values as a function of exposure, with and without FC corrections, for $1-3\sigma$ significance values. Figure 12 was produced using the same throws used for Fig. 11, with additional points from higher exposures used in Ref. [21], but not shown in Fig. 11 (646 and 1104 kt-MW-CY). After $\approx200$ kt-MW-CY, the median significance (including FC correction) for $50\%$ of the $\delta_{CP}$ range is greater than $3\sigma$. It is clear from Fig. 12 that the effect of the FC correction is not large, and $\approx10\%$ longer exposures are required for the median expected significance to cross each threshold than without correction, at both $\delta_{CP} = -\pi/2$ and for the $50\%$ range of $\delta_{CP}$ values.

Calculating $\Delta\chi^2$ values above $3\sigma$ using the FC method is challenging due to the large number of throws to explore the tails of the $\Delta\chi^2_{FC}$ distribution and prohibitive computational cost. In Fig. 13, the fraction of throws that exceed $1-5\sigma$ significance calculated only with the constant-$\Delta\chi^2$ method is shown in order to explore DUNE’s sensitivity at higher significance levels. All the caveats described above relating to the constant-$\Delta\chi^2$ method still apply. Figure 13 shows that, although the median significance to CPV does not exceed $5\sigma$ for $\delta_{CP} = -\pi/2$ until $\approx336$ kt-MW-CY, there are significant fractions of throws at lower exposures which reach $5\sigma$ significance. Figure 14 shows the fraction of throws which exceed different significance thresholds at the maximal CP-violating value of $\delta_{CP} = -\pi/2$, and for $50\%$ of all $\delta_{CP}$ values, as a function of exposure. By $\approx200$ kt-MW-CY, where the median significance for $50\%$ of the $\delta_{CP}$ range is greater than $3\sigma$, the sensitivity at $\delta_{CP} = -\pi/2$ exceeds $4\sigma$.

V. NEUTRINO MASS ORDERING SENSITIVITY

In this section, the toy-throwing approach described in Sec. II is used to explore the neutrino mass ordering sensitivity as a function of exposure in detail. In all cases, a joint ND $+$ FD fit is performed, and the reactor $\theta_{13}$ constraint is always applied, as described in Sec. II. An equal split between FHC and RHC running is assumed based on the results obtained in Sec. III.

Figure 15 shows the significance with which the neutrino mass ordering can be determined for both true NO and IO, for exposures of 66 and 100 kt-MW-CY. The sensitivity metric used is the square root of the difference between the best-fit $\chi^2$ value obtained using each ordering, as shown in Eq. (4), which is calculated for each throw of the

FIG. 15. Significance of the DUNE determination of the neutrino mass ordering, as a function of the true value of $\delta_{CP}$, for 66 kt-MW-CY (blue) and 100 kt-MW-CY (orange) exposures. The width of the transparent bands cover $68\%$ of fits in which random throws are used to simulate systematic, oscillation parameter and statistical variations, with independent fits performed for each throw constrained by prior uncertainties. The solid lines show the median significance. All exposures include an assumption of $57\%$ accelerator uptime as described in the text.
The characteristic shape of the MH sensitivity in Fig. 15 results from near degeneracy between matter and CPV effects that occurs near $\delta_{CP} = \frac{\pi}{2}$ ($\delta_{CP} = -\frac{\pi}{2}$) for true normal (inverted) ordering. Dedicated studies have shown that special attention must be paid to the statistical interpretation of neutrino mass ordering sensitivities [72–74] because the $\Delta \chi^2_{MO}$ metric does not follow the expected chi-square distribution for one degree of freedom, so the interpretation of the $\sqrt{\Delta \chi^2_{MO}}$ as the sensitivity is complicated.

Given the complications with the interpretation of significance for mass ordering determination, it is instructive to look at the distribution of the test statistic [Eq. (4)], which gives more information than the 68% central band and median throw shown in Fig. 15. Figure 16 shows the distribution of $\Delta \chi^2_{MO}$ obtained for a large ensemble of throws, for both true normal and inverted orderings, for a number of different exposures. There is a uniform distribution of true $\delta_{CP}$ used in the throws at each exposure. The change in shape at higher exposures in Fig. 16 is due to the degeneracy between $\delta_{CP}$ and the effect of the mass ordering, and as might be expected from Fig. 15, the separation between hierarchies is greater for some true values of $\delta_{CP}$ than others. This additional structure starts to become obvious from a $\approx 66$ kt-MW-CY exposure, at which point the CPV sensitivity is not very strong (see Sec. IV). For all exposures, the shape of the throw distribution is highly non-Gaussian, which makes it difficult to apply simple corrections to the sensitivity of the sort described in Ref. [74]. As a result alternatives to $\sqrt{\Delta \chi^2_{MO}}$ as a sensitivity metric are not explored, as the full information is given in Fig. 16. Figure 16 also indicates the probability for the test statistic $\Delta \chi^2_{MO}$ to be less (more) than zero from the toy throws for true normal (inverted) orderings at each exposure.

**FIG. 16.** The distribution of $\Delta \chi^2_{MO} = \chi^2_{IO} - \chi^2_{NO}$ values shown for both true normal (red) and true inverted (blue) hierarchies built using random throws of the systematic parameters, the oscillation parameters and with statistical variations. In each case, the $\chi^2$ values are separately minimized with respect to all variable parameters before calculating the test statistic. The fraction of throws for which the value of $\Delta \chi^2_{MO}$ is greater than (less than) 0 is also given for inverted (normal) hierarchies. For each ordering and exposure, approximately 100 000 throws were used. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) 6 kt-MW-CY (b) 12 kt-MW-CY (c) 24 kt-MW-CY (d) 66 kt-MW-CY (e) 100 kt-MW-CY (f) 336 kt-MW-CY.
exposure. This information is summarized in Fig. 17. This marks the proportion of toys which appear more like the incorrect ordering than the true ordering for the toy and gives a sense of the ambiguity between the hierarchies, although it is not easily converted to a single number sensitivity. It is clear from Figs. 16 and 17 that DUNE is sensitive to the mass ordering even from very low ($\approx 12$ kt-MW-CY) exposures, with a small probability for preferring the incorrect ordering. By exposures of 66 kt-MW-CY, the overlap between the orderings is very small with $\approx 1\%$ of toy throws which appear more like the incorrect ordering than the true ordering.

Figure 18 shows an alternative way to present the result of the throws as a function of $\delta_{CP}$, which is complementary to Fig. 15. The fraction of throws for which the simple figure of merit $[\text{the square root of Eq.}(4)]$ exceeds different confidence levels are shown, for $1–5\sigma$ significances, and a variety of exposures, all for true NO. The same throws are used as in Fig. 16. Despite the caveats regarding the interpretation of $\Delta \chi^2_{MO}$ as units of $\sigma$, the general trend is clear and provides more information about the expected exposure.

![Graph showing probability for preferring the wrong neutrino mass ordering as a function of exposure, shown for both true NO and IO. All exposures include an assumption of 57% accelerator uptime as described in the text.](image)

**FIG. 17.** The probability for preferring the wrong neutrino mass ordering as a function of exposure, shown for both true NO and IO. All exposures include an assumption of 57% accelerator uptime as described in the text.

![Graphs showing fraction of throws for which the DUNE sensitivity to the mass ordering exceeds 1–5\sigma significance, as a function of the true value of $\delta_{CP}$. Shown for NO, for a number of different exposures. The number of throws used to make each figure is also shown. All exposures include an assumption of 57% accelerator uptime as described in the text.](image)

**FIG. 18.** Fraction of throws for which the DUNE sensitivity to the mass ordering exceeds 1–5\sigma significance, as a function of the true value of $\delta_{CP}$. Shown for NO, for a number of different exposures. The number of throws used to make each figure is also shown. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) 6 kt-MW-CY (b) 12 kt-MW-CY (c) 24 kt-MW-CY (d) 66 kt-MW-CY (e) 100 kt-MW-CY (f) 336 kt-MW-CY.
DUNE sensitivity at low exposures. As with Figs. 11 and 13, the point at which the median significance (50% of throws) passes different significance thresholds can be easily read from the figures and can be compared with those shown in Fig. 15. The same general shape as a function of $\delta_{\text{CP}}$ as was observed in Fig. 15 can be seen. The general trend would be very similar in IO, reflected in the line $\delta_{\text{CP}} = 0$, although a slightly longer exposure is required to reach the same sensitivity. The median significance for $\delta_{\text{CP}} = -\pi/2$ exceeds $5\sigma$ for 24 kt-MW-CY, at which point the fraction of throws for which the significance is $3\sigma$ or smaller is only $\approx 2\%$. By 66 kt-MW-CY, 100% of the throws exceed $5\sigma$ at $\delta_{\text{CP}} = -\pi/2$. By 100 kt-MW-CY exposures, the median significance approaches $5\sigma$ for all true values of $\delta_{\text{CP}}$. At long exposures of 336 kt-MW-CY, almost 100% of the throws exceed $5\sigma$ for all values of $\delta_{\text{CP}}$.

VI. CONCLUSION

In this work a detailed exploration of DUNE’s sensitivity to CPV and the mass ordering at low exposures has been presented. The analysis uses the same framework, flux, cross section and detector models and selections as were used in Ref. [21], which showed the ultimate DUNE sensitivity to CPV, the neutrino mass ordering and other oscillation parameters, with large statistics samples after long exposures.

The effect of operating with different run plans, involving different ratios of FHC and RHC beam modes, on the mass ordering and CPV Asimov sensitivities was explored. It was found that, for low exposures, the sensitivity to both CPV and the mass ordering can be increased for certain regions of parameter space, but at a cost to the sensitivity in other regions. This sensitivity increase is in part produced by leveraging the strong $\theta_{13}$ constraint available from reactor experiments. If there is a strong reason to favor the exploration of a given region of parameter space when DUNE begins to take data, this issue should be revisited. However, with no strong motivation to focus on a given ordering or region of $\delta_{\text{CP}}$ parameter space, equal FHC and RHC beam running provides a close to optimal sensitivity across all of the parameter space, so was used for the subsequent detailed sensitivity studies. The increase in sensitivity for unequal beam running is also a feature of low exposure running and degrades the sensitivity almost uniformly across the parameter space investigated for large exposures, with and without a $\theta_{13}$ constraint applied.

The studies presented here demonstrate that a full treatment of DUNE’s sensitivity at low exposures supports the conclusions made in Refs. [8,21] using Asimov studies. In particular, the median CPV sensitivity is $\approx 3\sigma$ for $\delta_{\text{CP}} = \pm \pi/2$ after approximately a 100 kt-MW-CY FD exposure. Variations in the expected sensitivity around the median value were also explored. Additionally, it was shown that the CPV sensitivity is not significantly degraded when Feldman-Cousins corrections are included, leading to $\approx 10\%$ longer exposures to reach a given significance level. Crucially, it was found that after an initial low exposure rise, the Feldman-Cousins $\Delta \chi^2_{\text{CP}}$ do not change as a function of exposure, unlike the rise with exposure which has been observed by the T2K experiment [13].

It has also been shown that strong statements on the mass ordering can be expected with very short exposures of $\approx 12$ kt-MW-CY, which supports the results shown in Refs. [8,21] with a more complete treatment of the systematic uncertainty.

Although the analysis used here makes no assumptions about the FD staging scenario, and results are given as a function of exposure only, the results are dependent on having a performant ND complex from the start of the experiment. In particular, the low exposures necessary to make world-leading statements about the mass ordering can only be given with confidence with ND samples included in the fit.

ACKNOWLEDGMENTS

This document was prepared by the DUNE Collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359. This work was supported by CNPq, FAPERJ, FAPEG and FAPESP, Brazil; CFI, Institute of Particle Physics (IPP) and NSERC, Canada; CERN; MSMT, Czech Republic; ERDF, H2020-EU and MScA, European Union; CNRS/IN2P3 and CEA, France; INFN, Italy; FCT, Portugal; NRF, South Korea; Comunidad de Madrid (CAM), Fundación “La Caixa,” Junta de Andalucía-FEDER, and MICINN, Spain; SERI and SNSF, Switzerland; TÜBITAK, Turkey; The Royal Society and UKRI/STFC, United Kingdom; DOE and NSF, United States of America. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

APPENDIX: FELDMAN-COUSINS THROW DISTRIBUTIONS

The distribution of throws used to calculate the $\Delta \chi^2_{\text{CP}}$ values for Fig. 9 for nine different exposures with $\delta_{\text{CP}} = 0$ are shown in Fig. 19; for Fig. 10(a) for nine different values of $\delta_{\text{CP}}$ with an exposure of 100 kt-MW-CY in Fig. 20; and for Fig. 10(b) for nine different values of $\delta_{\text{CP}}$ with an exposure of 336 kt-MW-CY in Fig. 21. For each distribution shown in Figs. 19–21, the calculated $\Delta \chi^2_{\text{CP}}$ values corresponding to for $68.27\%$ (1$\sigma$), $90\%$, $95.45\%$ (2$\sigma$) and $99.73\%$ (3$\sigma$) of the throws are given and indicated with a vertical line. The number of throws used is also given. The $\Delta \chi^2_{\text{CP}}$ values were only calculated up to the $3\sigma$ level due to the very large number of throws required for higher confidence levels.
FIG. 19. Distribution of $\Delta \chi^2$ values, calculated using Eq. (5), for a large number of throws with true $\delta_{CP} = 0$, for a variety of exposures. The $\Delta \chi^2$ values (vertical lines) obtained using the Feldman-Cousins method show the $\Delta \chi^2_{FC}$ value below which 68.27% ($1\sigma$), 90%, 95.45% ($2\sigma$) and 99.73% ($3\sigma$) of throws reside, with the calculated values given in the legend. The number of throws used is also given. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) 24 kt-MW-CY (b) 66 kt-MW-CY (c) 100 kt-MW-CY (d) 150 kt-MW-CY (e) 197 kt-MW-CY (f) 336 kt-MW-CY (g) 500 kt-MW-CY (h) 646 kt-MW-CY (i) 936 kt-MW-CY.
FIG. 20. Distribution of $\Delta \chi^2$ values, calculated using Eq. (5), for a large number of throws for nine different values of true $\delta_{\text{CP}}$, for a 100 kt-MW-CY exposure. The $\Delta \chi^2$ values (vertical lines) obtained using the Feldman-Cousins method show the $\Delta \chi^2_{\text{FC}}$ value below which 68.27% ($1\sigma$), 90%, 95.45% ($2\sigma$) and 99.73% ($3\sigma$) of throws reside, with the calculated values given in the legend. The number of throws used is also given. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) $\delta_{\text{CP}}/\pi = -1$ (b) $\delta_{\text{CP}}/\pi = -0.75$ (c) $\delta_{\text{CP}}/\pi = -0.5$ (d) $\delta_{\text{CP}}/\pi = -0.25$ (e) $\delta_{\text{CP}}/\pi = 0$ (f) $\delta_{\text{CP}}/\pi = 0.25$ (g) $\delta_{\text{CP}}/\pi = 0.5$ (h) $\delta_{\text{CP}}/\pi = 0.75$ (i) $\delta_{\text{CP}}/\pi = 1$. 

LOW EXPOSURE LONG-BASELINE NEUTRINO OSCILLATION ... PHYS. REV. D 105, 072006 (2022)
FIG. 21. Distribution of $\Delta \chi^2$ values, calculated using Eq. (5), for a large number of throws for nine different values of true $\delta_{CP}$, for a 336 kt-MW-CY exposure. The $\Delta \chi^2$ values (vertical lines) obtained using the Feldman-Cousins method show the $\Delta \chi^2_{FC}$ value below which 68.27% ($1\sigma$), 90%, 95.45% ($2\sigma$) and 99.73% ($3\sigma$) of throws reside, with the calculated values given in the legend. The number of throws used is also given. All exposures include an assumption of 57% accelerator uptime as described in the text. (a) $\delta_{CP}/\pi = -1$ (b) $\delta_{CP}/\pi = -0.75$ (c) $\delta_{CP}/\pi = -0.5$ (d) $\delta_{CP}/\pi = -0.25$ (e) $\delta_{CP}/\pi = 0$ (f) $\delta_{CP}/\pi = 0.25$ (g) $\delta_{CP}/\pi = 0.5$ (h) $\delta_{CP}/\pi = 0.75$ (i) $\delta_{CP}/\pi = 1$. 

A. ABED ABUD et al. PHYS. REV. D 105, 072006 (2022)

(DUNE Collaboration)

1Abilene Christian University, Abilene, Texas 79601, USA
2University of Albany, SUNY, Albany, New York 12222, USA
3University of Amsterdam, NL-1098 XG Amsterdam, Netherlands
4Antalya Bilim University, 07190 Dışemelalt/Antalya, Turkey
5University of Antananarivo, Antananarivo 101, Madagascar
6Universidad Antonio Nariño, Bogotá, Colombia
7Argonne National Laboratory, Argonne, Illinois 60439, USA
8University of Arizona, Tucson, Arizona 85721, USA
9Universidad Nacional de Asunción, San Lorenzo, Paraguay
10University of Athens, Zografou GR 157 84, Greece
11Universidad del Atlántico, Barranquilla, Atlántico, Colombia
12Augustana University, Sioux Falls, South Dakota 57197, USA
13Banaras Hindu University, Varanasi—221 005, India
14University of Basel, CH-4056 Basel, Switzerland
15University of Bern, CH-3012 Bern, Switzerland
16Beykent University, Istanbul, Turkey
17University of Birmingham, Birmingham B15 2TT, United Kingdom
18Università del Bologna, 40127 Bologna, Italy
19Boston University, Boston, Massachusetts 02215, USA
20University of Bristol, Bristol BS8 1TL, United Kingdom
21Brookhaven National Laboratory, Upton, New York 11973, USA
22University of Bucharest, Bucharest, Romania
23Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ 22290-180, Brazil
24IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
25CERN, The European Organization for Nuclear Research, 1211 Meyrin, Switzerland
26CIEMAT, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, E-28040 Madrid, Spain
27Central University of South Bihar, Gaya, 824236, India
28University of California Berkeley, Berkeley, California 94720, USA
29University of California Davis, Davis, California 95616, USA
30University of California Irvine, Irvine, California 92697, USA
31University of California Los Angeles, Los Angeles, California 90095, USA
32University of California Riverside, Riverside California 92521, USA
33University of California Santa Barbara, Santa Barbara, California 93106 USA
34California Institute of Technology, Pasadena, California 91125, USA
35University of Cambridge, Cambridge CB3 0HE, United Kingdom
36Universidade Estadual de Campinas, Campinas—SP, 13083-970, Brazil
37Università di Catania, 2—95131 Catania, Italy
38Universidad Católica del Norte, Antofagasta, Chile