Electron Beam Energy Reconstruction for Neutrino Oscillation Measurements

M. Khachatryan,[1] A. Papadopolou,[2] A. Ashkenazi,[2,*] F. Hauenstein,[1,2] A. Nambrath,[2] A. Hrnjic,[2] L.B. Weinstein,[1] O. Hen,[2] E. Piasetzky,[3] M. Betancourt,[4] S. Dytman,[5] P. Coloma,[7] and K. Mahn[6]

(The CLAS and $e4\nu$ Collaborations)

[1] *Old Dominion University, Norfolk, Virginia 23529*
[2] *Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3] *School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel*
[4] *Fermi National Accelerator Laboratory, Batavia, IL*
[5] *University of Pittsburgh, Pittsburgh, PA*
[6] *Michigan State University, East Lansing, MI*
[7] *Instituto de Fisica Corpuscular, Universitat de Valencia and CSIC, Paterna, E-46980 Valencia, Spain and Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*

Neutrinos exist in one of three types or "flavors" (ν_e, ν_μ or ν_τ) which oscillate from one to another when propagating through space. This phenomena is one of the few that cannot be described using the Standard Model of particle physics [1]. Thus, its experimental study can provide new insight into the nature of our universe [2]. Neutrinos oscillate as a function of their propagation distance divided by their energy (L/E). Therefore experiments extract oscillation parameters by measuring their energy distribution at different locations. As accelerator-based oscillation experiments cannot directly measure E, their interpretation relies heavily on phenomenological models of neutrino-nucleus interactions to infer E. Here we exploit the similarity of electron- and neutrino-nucleus interactions, and use electron scattering data with known beam energies to test energy reconstruction methods and interaction models. We find that even in simple interactions where no pions are detected, only a small fraction of events reconstruct to the correct incident energy. While widely-used interaction models reproduce kinematical distributions such as transverse momentum well, they describe the reconstructed energy distribution only qualitatively. Applying our measured and modeled incident-energy reconstruction results to the Deep Underground Neutrino Experiment (DUNE, USA) [3] neutrino flux we find that the discrepancies between the data and the model can bias the reconstruction of neutrino oscillation parameters. This shows the need to improve current models to meet the requirements of next-generation, high-precision experiments such as Hyper-Kamiokande (HK, Japan) [4] and DUNE.

The three types of neutrinos are described in two different bases: flavor and mass. The weak nuclear interaction of neutrinos with other particles is described using flavor (ν_e, ν_μ, and ν_τ) while their propagation is described using mass. Each flavor state is a linear combination of the three mass states (ν_1, ν_2, and ν_3) [1].

In the simpler case of two neutrino flavors the oscillation probability from ν_μ to ν_e is given by [5]

$$P_{\nu_\mu \to \nu_e}(E, L) \approx \sin^2(2\theta_{12}) \sin^2\left(\frac{\Delta m^2_{12} L}{4E}\right), \quad (1)$$

where $\Delta m^2_{12} = m^2_{\nu_1} - m^2_{\nu_2}$ is the neutrino mass difference squared that determines the oscillation wavelength as a function of L/E and θ_{12} is the neutrino mixing angle that determines the oscillation amplitude. CP symmetry violation in the leptonic sector would add a phase (δ_{CP}) to the oscillation with an opposite sign for neutrinos and anti-neutrinos ($\bar{\nu}$) [6, 7].

Accelerator-based measurements produce beams that predominantly contain either ν_μ or $\bar{\nu}_\mu$. At a distance L from the neutrino production point some ν_μ will oscillate to ν_e, resulting in fluxes of

$$\begin{aligned} \Phi_e(E, L) &\propto P_{\nu_\mu \to \nu_e}(E, L)\, \Phi_\mu(E, 0), \\ \Phi_\mu(E, L) &\propto \left[1 - P_{\nu_\mu \to \nu_e}(E, L)\right] \Phi_\mu(E, 0), \end{aligned} \quad (2)$$

where the proportionality constant depends on the experiment geometry.

$\nu_\mu \to \nu_e$ oscillations are thus observed by measuring the neutrino fluxes $\Phi_e(E, L)$ and $\Phi_\mu(E, L)$ as a function of energy or distance. The three-flavor oscillation equations are similar but include additional terms, see Methods for details.

Experimentally, the neutrino flux is extracted from the measured neutrino interaction rate with atomic nuclei in neutrino detectors. This interaction rate is given by:

$$N_e(E_{rec}, L) \propto \sum_i \int \Phi_e(E, L) \sigma_i(E) f_{\sigma_i}(E, E_{rec}) dE, \quad (3)$$

where $\sigma_i(E)$ is the neutrino interaction cross section for process i (e.g. quasi-elastic scattering, resonance production, etc.), E_{rec} is the neutrino energy reconstructed from the angles and momenta of the particles measured in the detector.

* Contact Author adishka@mit.edu

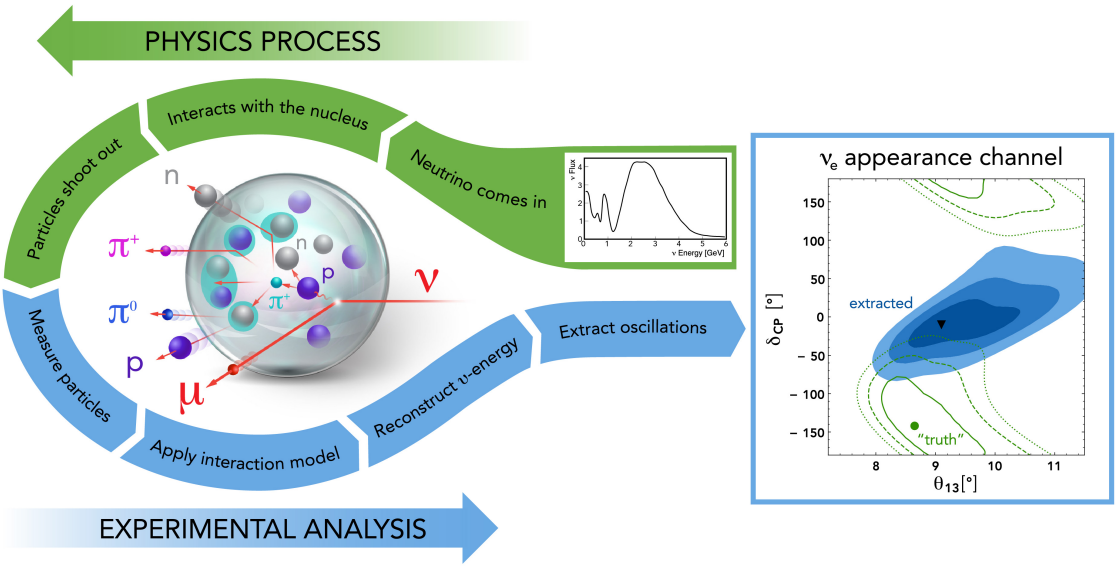

Fig. 1: **Neutrino oscillations and energy spectra measurements** | (Left) Neutrino energy spectra reconstruction depends on our ability to model the interaction of neutrinos with atomic nuclei and the propagation of particles through the atomic nucleus. This flow chart shows the process, starting with the oscillated far-detector ν_e incident-energy spectrum and differentiating the physical neutrino interactions (green arrows) from the experimental analysis (blue arrows). (Right) simulated extracted neutrino oscillation parameters (θ_{13} vs. δ_{CP}) in the $\nu_\mu \to \nu_e$ appearance channel for three years of running with DUNE. The green dot (labelled "truth") signifies the input values of the parameters. The solid, dashed and dotted green lines show the 1, 2 and 3σ confidence regions obtained from simulating the neutrino interactions with the CLAS *data*-derived smearing matrices and fitting them with the same matrices. The solid blue areas (labelled "extracted") show the resulting confidence regions when fitting using the *model*-derived smearing matrices instead. The black triangle shows the best fit parameters.

$f_{\sigma_i}(E, E_{rec})$ is a smearing matrix relating the real (E) and reconstructed (E_{rec}) neutrino energies. E_{rec} differs from E due to both experimental effects (e.g. detector resolutions, inefficiencies, backgrounds) and nuclear interaction effects (e.g. nucleon motion, meson currents, nucleon reinteraction). While experimental effects are generally understood and can be minimized using improved detectors, nuclear effects are irreducible and must be accounted for using theoretical models, typically implemented in neutrino event generators.

The precision to which oscillation parameters can be determined experimentally therefore depends on our ability to extract $\Phi_\alpha(E, L)$ from $N_\alpha(E_{rec}, L)$, see Fig. 1. This is largely determined by the accuracy of the theoretical models used to calculate $\sigma_i(E)$ and $f_{\sigma_i}(E, E_{rec})$. Current oscillation experiments report significant systematic uncertainties due to these interaction models [7–10] and simulations show that energy reconstruction errors can lead to significant biases in extracting δ_{CP} at DUNE [11]. There is a robust theoretical effort to improve these models [12–14].

Because there are no mono-energetic high-energy neutrino beams, these models cannot be tested for individual neutrino energies. Instead, experiments tune models of $\sigma_i(E)$ and $f_{\sigma_i}(E, E_{rec})$ to reproduce their near-detector data, where the unoscillated flux $\Phi(E, 0)$ is well known from hadronic calculations [15, 16].

To improve on this, experiments also typically select simpler charged-current event "topologies", such as events with one detected muon, one proton, and zero pions ($1p0\pi$) that are hopefully dominated by well-understood quasi-elastic (QE) scattering (i.e., by scattering of the neutrino from a single moving nucleon).

While highly informative, such integrated constraints are insufficient to ensure that the models are correct for each value of E. Therefore, even if the models are tuned to reproduce the near-detector data, there is no guarantee that they are suitable for analyzing far-detector data, where the neutrino flux can be very different due to oscillations.

Here we report the first measurement of $f_{\sigma_i}(E, E_{rec})$ for mono-energetic electron-nucleus scattering, and use it to test interaction models used by neutrino oscillation analyses. Both types of leptons, e and ν, interact similarly with nuclei. Both particles interact with nuclei via a vector current, while neutrinos have an additional axial-vector current. The nuclear ground state is the same in both cases and many of the nuclear reaction effects are similar. Therefore, any model of neutrino interactions (vector+axial-vector) should also be able to reproduce electron (vector) interactions. The data presented here can therefore test and constrain neutrino-nucleus interac-

tion models to be used in analysis of neutrino oscillation measurements.

We found that $1p0\pi$ events, which were expected to be predominantly quasi-elastic, reconstruct to the correct energy less than 50% of the time, and that the model overestimates the amount of mis-reconstructed events due to non-QE processes by $\sim 10-20\%$. To estimate the possible effect of this data-model difference, we analyzed a DUNE far-detector flux simulation [11, 17] with $f_{\sigma_i}(E, E_{rec})$ taken from our data and from the model. We found non-negligible differences in the extracted neutrino oscillation parameters. This highlights a significant shortcoming in our current understanding of neutrino interactions which, if not corrected, could limit the exploitation of the full potential of next-generation, high-precision oscillation experiments, namely DUNE and HK.

I. ELECTRON DATA SELECTION

The experiment measured electron scattering from ^{4}He, ^{12}C, and ^{56}Fe nuclei at beam energies of 1.159, 2.257 and 4.453 GeV, detecting the scattered electron and knocked out particles over a wide range of angles and momenta in the CEBAF Large Acceptance Spectrometer (CLAS) [18] at the Thomas Jefferson National Accelerator Facility (Jefferson Lab). We detected electrons with energy $E_e \geq 0.4, 0.55$ and 1.1 GeV for $E_{beam} = 1.159, 2.257$ and 4.453 GeV respectively and angles $15^\circ \leq \theta_e \leq 45^\circ$, hadrons with momenta above 150 to 300 MeV/c and $10-20^\circ \leq \theta_h \leq 140^\circ$, and photons with energy $E_\gamma \geq 300$ MeV. These hadron detection thresholds are similar to those of neutrino detectors [19], however neutrino detectors have full angular coverage and lower lepton energy thresholds. See Methods for details.

The incident energies used here span the range of typical accelerator-based neutrino beam energies (Extended Data Fig. 1). The carbon data are relevant for scintillator-based experiments such as MINERνA and NOνA [9] and similar to the oxygen in water-based Čerenkov detectors such as Super-Kamiokande [7, 8] and Hyper-Kamiokande [20]. The iron is similar to the argon in the liquid argon time projection chambers of MicroBoone [21], the Fermilab short-baseline oscillation program [22] and DUNE [23]. Many nuclear interaction processes are mass dependent, so it is important to measure a range of target nuclei.

We selected events with one electron and zero pions or photons from π^0 decay above threshold. We did this to maximize the contribution of well-understood events where the incident lepton scattered quasi-elastically from a single nucleon in the nucleus, as is done in many neutrino oscillation analyses [1, 24].

Electrons, unlike neutrinos, radiate bremsstrahlung photons in the electric field of the nucleus. We vetoed events where the photons from scattered-electron radiation were detected in CLAS. We did not correct our data for electron radiation, but instead added electron radiation to the simulations described below. See Methods for details.

We subtracted from our data contributions from events where unwanted pions or photons were produced but not detected due to the incomplete CLAS angular acceptance ($\approx 50\%$). We used events with a detected unwanted particle (e.g., pion or extra proton), and for each event we constructed a "simulation" where we rotated the unwanted-particle momentum around the (known) momentum transfer direction many times to determine the probability P of detecting similar events. We then applied a weight $W = (1-P)/P$ to account for the undetected similar events, and subtracted it from our data set. See Methods for details.

This produced an $(e,e')_{0\pi}$ data set where events included any number of detected or undetected protons and neutrons as well as charged pions and photons below the CLAS detection threshold. We also separately examined the subset of events with exactly one detected proton, i.e. $(e,e'p)_{1p0\pi}$, subtracting contributions from events with additional undetected protons above threshold.

Due to the difference in mass of the exchanged vector bosons ($W^\pm$ for charged-current ν interactions and γ for electrons), the electron-nucleon cross section is much more forward peaked than the neutrino cross section. We accounted for that by weighting each event by Q^4, to account for the different electron- and neutrino-nucleon cross sections. See Methods for details.

We considered several different sources of systematic uncertainties, including the angular dependence of the pion-production cross section (for the undetected-pion subtraction), the effects of fiducial cuts on undetected particle subtraction, and photon identification cuts. The overall point-to-point systematic uncertainty from these effects on the subtracted spectrum was about 2% at 1.159 and 2.257 GeV, and about 5% at 4.453 GeV. See Methods for details.

II. FROM NEUTRINO TO ELECTRON SCATTTERING

We compared our mono-energetic electron data to predictions of the GENIE [25] simulation, which is used by most neutrino experiments in the USA and has an electron-scattering version (e-GENIE) that was recently updated to be consistent with the neutrino version. We used a GENIE "tune" that reproduces measured neutrino [26] and electron inclusive cross sections, see Extended Data Fig. 2 (top). GENIE includes quasi-elastic lepton scattering (QE), interactions of the lepton with a meson exchanged between two nucleons (meson exchange currents or MEC, often referred to "$2p2h$"), resonance production in nuclei (RES) and "deep inelastic scatterin'g" (DIS, which also includes all non-resonant meson production), as well as rescattering (final state interactions) of

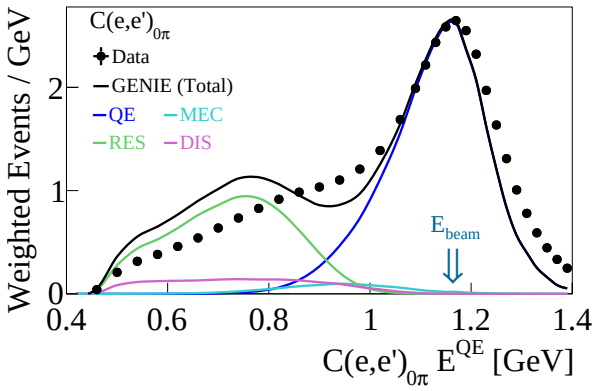

Fig. 2: **Quasi-Elastic Reconstructed Energy** | The number of weighted 1.159 GeV C$(e,e')_{0\pi}$ events per GeV plotted as a function of the reconstructed energy E_{QE} for data (black points) and GENIE (black curve). The colored lines show the contribution by different processes to the GENIE simulation: QE (blue), MEC (cyan), RES (green) and DIS (magenta). The GENIE results are normalized to the same integral as the data. Statistical uncertainties are smaller than the data points. Systematic uncertainties are not shown.

the outgoing hadrons.

We generated events using e-GENIE, propagated the events through CLAS acceptance maps to determine which particles were detected, and smeared the momenta of these particles based on the known CLAS resolution. We then analyzed the resulting simulated events using the same code as the data (including the Q^4 weighting and the subtraction for undetected particles) and compared the two. See Methods for details.

Electron radiation was added to the simulation based on the formalism of Ref. [27] and verified against 4 GeV H(e,e') elastic scattering data. Radiation was negligible for photons with more than 10 MeV. See Extended Data Fig. 3.

To demonstrate the relevance of our electron study to neutrino interactions, Extended Data Fig. 2 (bottom) compares the predictions of e-GENIE and ν-GENIE for the events described above. For this study we turned off electron radiation and weighted e-GENIE events by Q^4. The resulting event distributions are very similar for e-GENIE and ν-GENIE.

III. INCIDENT ENERGY RECONSTRUCTION

There are two general approaches for reconstructing the incident neutrino energy, based on the particle detection capabilities of the neutrino detector.

Water Čerenkov detectors only measure charged leptons and pions. If the neutrino scattered quasi-elastically (QE) from a stationary nucleon in the nucleus, its energy can be reconstructed from the measured lepton as:

$$E_{QE} = \frac{2M_N\epsilon + 2M_N E_l - m_l^2}{2(M_N - E_l + k_l\cos\theta_l)}, \tag{4}$$

where $\epsilon \approx 20$ MeV is the average nucleon separation energy, M_N is the nucleon mass, and $(m_l, E_l, k_l, \theta_l)$ are the scattered lepton mass, energy, momentum, and angle. Nucleon motion will doppler-broaden E_{QE}.

Figure 2 shows the E_{QE} distribution for 1.159 GeV C$(e,e')_{0\pi}$ events, which are most relevant for T2K and HK. We observe a broad peak centered at the real beam energy with a large tail extending to lower energies. The peak is doppler-broadened by the motion of the nucleons in the nucleus. The tail is caused by non-quasi-elastic reactions that pass the $(e,e')_{0\pi}$ selection. The tail is cut off at the lowest energies by the CLAS minimum detected electron energy threshold of 0.4 GeV.

The e-GENIE peak is too narrow, with a Gaussian width of $\sigma = 76$ MeV, compared to 89 MeV for the data. This is due to inexact modeling of the nuclear ground state momentum distribution. The tail dips below the data at around 0.9 GeV, and is larger than the data at lower reconstructed energies. This energy reconstruction discrepancy is observed despite the overall good reproduction of the differential inclusive cross-sections (Extended Data Fig. 2) and the transverse missing momentum (discussed below), which are two of the main observables to which current experiments tune models in order to trust their energy reconstruction predictions.

Tracking detectors measure all charged particles above their detection thresholds. The "calorimetric" incident neutrino energy is then the sum of all the detected particle energies:

$$E_{cal} = \sum E_i + \epsilon, \tag{5}$$

where E_i are the detected nucleon kinetic energies and the lepton and meson total energies and ϵ is the average total removal energy for the detected particles.

Figure 3 shows the E_{cal} distribution for 1.159, 2.257 and 4.453 GeV C$(e,e'p)_{1p0\pi}$ events and 2.257 and 4.453 GeV Fe$(e,e'p)_{1p0\pi}$ events. All spectra show a sharp peak at the real beam energy, followed by a large tail at lower energies. For carbon only about 25–50% of the events reconstruct to within 5% of the real beam energy, see Extended Data Table 1. For iron this fraction is only about 25%, highlighting the crucial need to well model the low-energy tail of these distributions. e-GENIE reproduces the fraction of events in the peak to within 10–20% of the fraction, overestimating the fraction at 1.159 GeV and underestimating it at 2.257 and 4.453 GeV. The peak underestimate at 2.257 GeV could be due to an overestimate of π and proton multiplicities, see Extended Data Fig. 4, or to an overestimate of neutrons. e-GENIE also predicts the peak position to be about 20 MeV too low at all three energies.

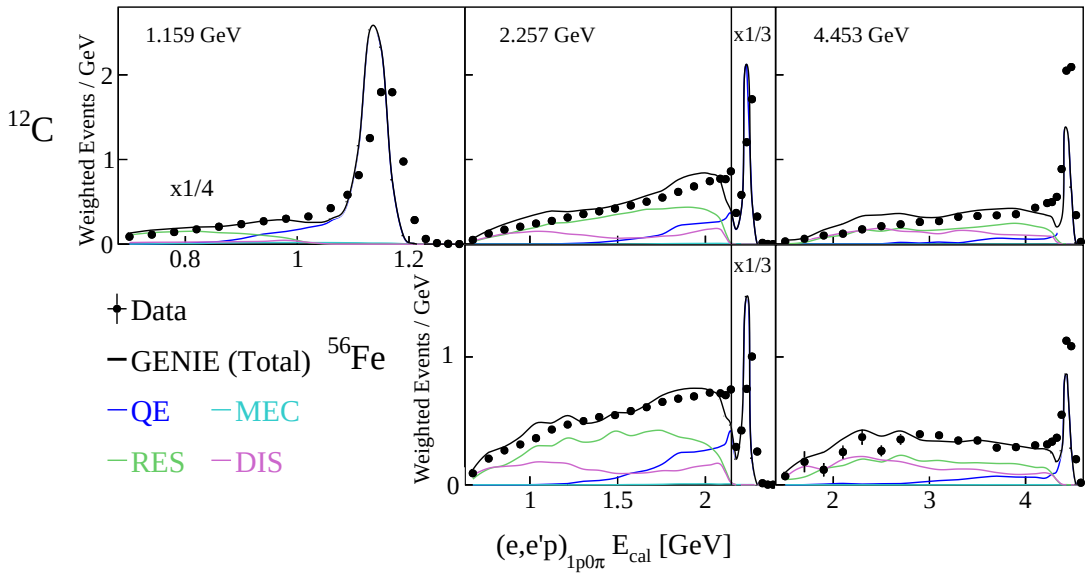

Fig. 3: **Calorimetric Reconstructed Energy** | The number of weighted A$(e,e'p)_{1p0\pi}$ events per GeV plotted as a function of the reconstructed calorimetric energy E_{cal} for data (black points) and GENIE (black curve). Different panels show results for different beam energies and target nuclei combinations: (top row) Carbon target with (left to right) 1.159, 2.257, and 4.453 GeV incident beam and (bottom) Iron target with (left) 2.257 and (right) 4.453 GeV incident beam. All curves have been scaled by a factor of 1/4 for the 1.159 GeV spectrum. Colored lines show the contribution by different processes to the GENIE simulation: QE (blue), MEC (cyan), RES (green) and DIS (magenta). The GENIE results are normalized to the same integral as the data in each panel. Error bars include statistical uncertainties at the 68% (1σ) level. Error bars are not shown when they are smaller than the size of the data point. Systematic uncertainties of 2% for 1.159 and 2.257 GeV and 4% for 4.453 GeV data are not shown.

e-GENIE underpredicts the low energy tail slightly for the 1.159 GeV data, but overpredicts it for higher energies. The tail seems to be dominated by resonance production that did not result in the production of charged particles above detection threshold. At higher energy, contributions from DIS processes also become significant. MEC contributions are very small because the empirical MEC model contribution decreases rapidly with Q^2. However, increased MEC contributions would increase the fraction of events in the tail and therefore decrease the discrepancy between GENIE and data at 1.159 GeV and increase it at 2.257 and 4.453 GeV.

Extended Data Fig. 6 shows that there are similar discrepancies between data and e-GENIE even at the QE peak (i.e., at $0.8 \leq x_B \leq 1.2$, where $x_B = Q^2/(2m_N\omega)$). This cut was done using knowledge of the true beam energy, which is not possible in neutrino experiments.

While the $(e,e')_{0\pi}$ quasi-elastic reconstruction of Eq. 4 gives a much broader peak at the true beam energy than the calorimetric energy E_{cal} due to the effects of nucleon motion (see Extended Data Fig. 5), it has the same tail of lower energy events for the same $(e,e'p)_{1p0\pi}$ data set. The two energy reconstruction methods agree remarkably well within their respective resolutions and therefore consistency between the two methods does not indicate accuracy, see Extended Data Fig. 7.

IV. TRANSVERSE VARIABLES AND MODEL TUNING

Neutrino experiments that use tracking detectors can accurately measure the transverse missing momentum of the detected particles (using the known incident neutrino direction),

$$P_T = |\vec{P}_T^{\,e'} + \vec{P}_T^{\,p}|, \tag{6}$$

where $\vec{P}_T^{\,e'}$ and $\vec{P}_T^{\,p}$ are the three-momenta of the detected lepton and proton perpendicular to the direction of the incident lepton, respectively. Purely quasi-elastic events, where the lepton scattered from a bound moving proton, will have small P_T, consistent with the motion of the struck nucleon. Events with small P_T should thus reconstruct to the correct incident energy. Non-quasi-elastic events, where neutral or sub-detection-threshold charged particles were produced, will have larger P_T and will not reconstruct to the correct incident energy. P_T is thus an ideal observable for tuning reaction models to ensure they correctly account for non-QE processes.

The P_T distribution for 2.257 GeV C$(e,e'p)_{1p0\pi}$ events is shown in Fig. 4 (and the other energies are shown in Extended Data Fig. 8). Both data and e-GENIE peak at relatively low momenta, as expected, but both have a

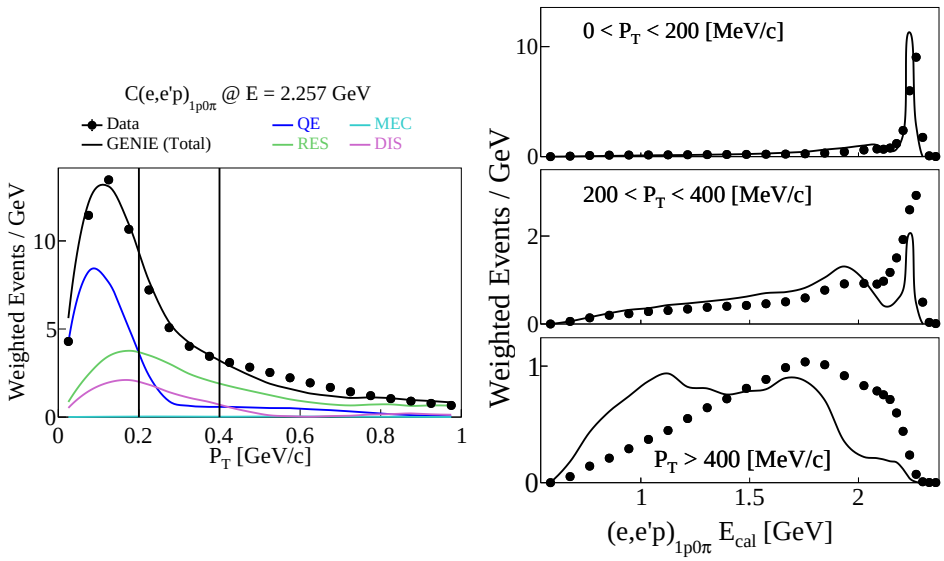

Fig. 4: **Reconstructed energies and perpendicular momenta** | (Left) the missing transverse momentum, P_T for 2.257 GeV C$(e,e'p)_{1p0\pi}$ events for data (black points) and GENIE (black histogram). The vertical lines at 200 MeV/c and at 400 MeV/c separate the three bins in P_T. The colored lines show the different components of the GENIE simulation, QE (blue), MEC (cyan), RES (green) and DIS (magneta). (Right) The calorimetric energy E_{cal} for different bins in P_T: (top) $P_T < 200$ MeV/c, (middle) 200 MeV/$c \leq P_T \leq 400$ MeV/c and (bottom) $P_T > 400$ MeV/c . The GENIE results are normalized to the same integral as the data in the left plot. Each bin has been scaled by the bin width.

large tail extending between 0.3 and 1 GeV/c and containing almost half of the measured events. The high-P_T tail is predominantly due to resonance production that did not result in an additional pion or nucleon above the detection threshold. e-GENIE reproduces the data remarkably well, suggesting adequate reaction modeling, including the contribution of non-QE processes such as resonance production.

Despite this excellent agreement, the reconstructed energy distributions, which are not testable in neutrino measurements, are not well reproduced by e-GENIE at high-P_T (see Fig. 4). Events with $P_T < 200$ MeV/c almost all reconstruct to the correct incident energy and are well described by e-GENIE. However, events with $P_T \geq 400$ MeV/c do not reconstruct to the correct energy and are poorly reproduced by e-GENIE.

This disagreement, which is not seen in the P_T distribution, indicates that including high-P_T data in oscillation analyses could bias the extracted parameters. As high-P_T data accounts for $25-50\%$ of the measured events, cutting it out would significantly increase the statistical uncertainties of the measurements. Thus, care must be taken to improve the models implemented in GENIE, so that they can reproduce the high-P_T data.

V. IMPACT ON NEUTRINO OSCILLATION PARAMETERS

To test the impact of the data-e-GENIE energy reconstruction differences on the extraction of oscillation parameters, we created two $f_{\sigma_i}(E, E_{rec})$ smearing matrixes, one for data and one for e-GENIE. The matrixes use as input the energy feed-down spectra, $(E_{cal} - E_{true})/E_{true}$ (see Extended Data Fig. 5), and linearly interpolate and extrapolate over $0.5 \leq E \leq 5$ GeV.

We then simulated the measured oscillated ν_e appearance spectrum at the DUNE far-detector for 1250 events using nominal neutrino oscillation parameters (green dot in Fig. 1) and the data-based smearing matrix. We fit the simulated spectra (see Eq. 7) using the same smearing matrix and using the e-GENIE-based smearing matrix. See Methods for details.

As expected, when consistently using the data-based smearing matrix the best fit parameters are consistent with the input values. However, when fitting using the e-GENIE-based smearing matrix the best fit value of δ_{CP} differed by more than 3σ from its input value (see Fig. 1). While it may be possible to shift some of the δ_{CP} bias to other parameters by simultaneously fitting the oscillated ν_μ spectrum, these other parameters are increasingly well-constrained by reactor and other precision measurements. See Methods for details.

While DUNE is overall similar to CLAS, several differences between the CLAS and DUNE data still re-

main. (1) DUNE covers smaller lepton scattering angles than the CLAS minimum angle of 15°. This excludes the DUNE low-Q^2 region from our analysis. However, this region is known to be poorly described by current models [26, 28]. (2) The CLAS data included only $1p0\pi$ events, while the DUNE data will include all interactions, of which only about 25% are $1p0\pi$. However, since $1p0\pi$ events primarily come from the relatively well understood quasi-elastic scattering, these should have the best agreement between data and e-GENIE. (3) The carbon target used for electron scattering should be better understood than the argon in the DUNE detectors. This comparison is less relevant for T2K due to its lower $\pi^{\pm}$ detection thresholds and lower incident neutrino energies [7].

VI. SUMMARY

In conclusion, we have used Jefferson Lab CLAS electron-nucleus scattering data with known incident energies to perform the first test of our ability to reconstruct incident neutrino energies from measured neutrino-nucleus collisions. Understanding the incident energy reconstruction is a crucial step in the neutrino-oscillation experiment analysis chain.

Most of the $1p0\pi$ events do not reconstruct to the correct incident energy. The interaction model describes the size but not the exact shape of the low-energy tail for quasi-elastic energy reconstruction, despite reproducing both differential inclusive electron-scattering cross sections and transverse missing momenta. The same interaction model moderately mis-estimates the low-energy tail for calorimetric energy reconstruction.

We found that smearing the incident neutrino energy spectrum with CLAS data and then analyzing that with e-GENIE led to significantly different neutrino oscillation parameters. This shows the importance of having the model (e.g., GENIE) describe reality as accurately as possible.

Combining the neutrino energy reconstruction studies presented here with the standard near-detector neutrino data analyses could significantly reduce the systematic modeling uncertainties of next generation oscillation experiments. Future experiments with the improved CLAS12 spectrometer will extend these measurements to more nuclei, much smaller scattering angles, and to a wider range of momentum transfers.

[1] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98**, 030001 (2018).
[2] R. Mohapatra *et al.*, Rept. Prog. Phys. **70**, 1757 (2007), arXiv:hep-ph/0510213.
[3] B. Abi *et al.* (DUNE), (2018), arXiv:1807.10334 [physics.ins-det].
[4] K. Abe *et al.* (Hyper-Kamiokande), (2018), arXiv:1805.04163 [physics.ins-det].
[5] M. Freund, Phys. Rev. D **64**, 053003 (2001).
[6] M. Fukugita and T. Yanagida, Phys. Lett. B **174**, 45 (1986).
[7] K. Abe *et al.* (T2K), Nature **580**, 339 (2020).
[8] K. Abe *et al.* (T2K), Phys. Rev. Lett. **121**, 171802 (2018), arXiv:1807.07891 [hep-ex].
[9] L. Aliaga *et al.*, Nucl. Instrum. Methods **A743**, 130 (2014).
[10] L. Alvarez-Ruso *et al.*, Prog. Part. Nucl. Phys. **100**, 1 (2018), arXiv:1706.03621 [hep-ph].
[11] A. M. Ankowski, P. Coloma, P. Huber, C. Mariani, and E. Vagnoni, Phys. Rev. D **92**, 091301 (2015), arXiv:1507.08561 [hep-ph].
[12] N. Rocco, Frontiers in Physics **8**, 116 (2020).
[13] S. Dolan, G. D. Megias, and S. Bolognesi, Phys. Rev. D **101**, 033003 (2020).
[14] N. Rocco, A. Lovato, and O. Benhar, Phys. Rev. Lett. **116**, 192501 (2016).
[15] K. K. Maan (NOvA), PoS **ICHEP2016**, 931 (2016).
[16] L. Haegel (T2K), in *18th International Workshop on Neutrino Factories and Future Neutrino Facilities Search* (2017) arXiv:1701.02559 [hep-ex].
[17] P. Coloma and P. Huber, Phys. Rev. Lett. **111**, 221802 (2013), arXiv:1307.1243 [hep-ph].
[18] B. A. Mecking *et al.*, Nucl. Instrum. Meth. **A503**, 513 (2003).
[19] M. Betancourt *et al.* (MINERvA), Phys. Rev. Lett. **119**, 082001 (2017), arXiv:1705.03791 [hep-ex].
[20] K. Abe *et al.* (Hyper-Kamiokande), (2018), arXiv:1805.04163 [physics.ins-det].
[21] R. Acciarri *et al.* (MicroBooNE), JINST **12**, P02017 (2017), arXiv:1612.05824 [physics.ins-det].
[22] M. Antonello *et al.* (MicroBooNE, LAr1-ND, ICARUS-WA104), (2015), arXiv:1503.01520 [physics.ins-det].
[23] R. Acciarri *et al.* (DUNE), (2015), arXiv:1512.06148 [physics.ins-det].
[24] T. Katori and M. Martini, J. Phys. G **45**, 013001 (2018), arXiv:1611.07770 [hep-ph].
[25] C. Andreopoulos, A. Bell, D. Bhattacharya, F. Cavanna, J. Dobson, S. Dytman, H. Gallagher, P. Guzowski, R. Hatcher, P. Kehayias, A. Meregaglia, D. Naples, G. Pearce, A. Rubbia, M. Whalley, and T. Yang, Nucl. Inst. and Meth. A **614**, 87 (2010).
[26] P. Abratenko *et al.* (MicroBooNE), Phys. Rev. Lett. **123**, 131801 (2019), arXiv:1905.09694 [hep-ex].
[27] L. W. Mo and Y.-S. Tsai, Rev. Mod. Phys. **41**, 205 (1969).
[28] P. Abratenko *et al.* (MicroBooNE), (2020), arXiv:2006.00108 [hep-ex].
[29] A. Cervera, A. Donini, M. Gavela, J. G. Cádenas], P. Hernández, O. Mena, and S. Rigolin, Nuclear Physics B **579**, 17 (2000).
[30] A. Cervera, A. Donini, M. Gavela, J. G. Cádenas, P. Hernández, O. Mena, and S. Rigolin, Nuclear Physics B **593**, 731 (2001).
[31] O. Hen *et al.*, Science **346**, 614 (2014), arXiv:1412.0138 [nucl-ex].
[32] D. Rein and L. M. Sehgal, Annals of Physics **133**, 79

(1981).
[33] R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971).
[34] T. Katori, "Meson exchange current (mec) models in neutrino interaction generators," (2013), arXiv:1304.6014 [nucl-th].
[35] S. Dytman and A. Meyer, AIP Conference Proceedings **1405**, 213 (2011).
[36] R. Merenyi, W. A. Mann, T. Kafka, W. Leeson, B. Saitta, J. Schneps, M. Derrick, and B. Musgrave, Phys. Rev. D **45**, 743 (1992).
[37] S. G. Mashnik, A. J. Sierk, K. K. Gudima, and M. I. Baznat, J. Phys. Conf. Ser. **41**, 340 (2006).
[38] N. Markov *et al.* (The CLAS Collaboration), Phys. Rev. C **101**, 015208 (2020).
[39] B. Abi *et al.* (DUNE), (2020), arXiv:2002.03005 [hep-ex].
[40] I. Esteban, M. Gonzalez-Garcia, A. Hernandez-Cabezudo, M. Maltoni, and T. Schwetz, JHEP **01**, 106 (2019), arXiv:1811.05487 [hep-ph].
[41] P. Huber, M. Lindner, and W. Winter, Comput. Phys. Commun. **167**, 195 (2005), arXiv:hep-ph/0407333.
[42] P. Huber, J. Kopp, M. Lindner, M. Rolinec, and W. Winter, Comput. Phys. Commun. **177**, 432 (2007), arXiv:hep-ph/0701187.
[43] P. Coloma, P. Huber, J. Kopp, and W. Winter, Phys. Rev. D **87**, 033004 (2013), arXiv:1209.5973 [hep-ph].
[44] H. Dai *et al.* (The Jefferson Lab Hall A Collaboration), Phys. Rev. C **99**, 054608 (2019).

Acknowledgments We acknowledge the efforts of the staff of the Accelerator and Physics Divisions at Jefferson Lab that made this experiment possible. We thank Luke Pickering for useful discussions. The analysis presented here was carried out as part of the Jefferson Lab Hall B Data-Mining project supported by the U.S. Department of Energy (DOE). The research was supported also by DOE, the National Science Foundation, the Israel Science Foundation, the Chilean Comisión Nacional de Investigación Científica y Tecnológica, the French Centre National de la Recherche Scientifique and Commissariat a l'Energie Atomique, the French-American Cultural Exchange, the Italian Istituto Nazionale di Fisica Nucleare, the National Research Foundation of Korea, and the UK's Science and Technology Facilities Council. PC acknowledges support from project PROMETEO/2019/083. This project has been supported by the European Union Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 674896 (Elusives, H2020-MSCA-ITN- 2015-674896). Jefferson Science Associates operates the Thomas Jefferson National Accelerator Facility for the DOE, Office of Science, Office of Nuclear Physics under contract DE-AC05-06OR23177. The raw data from this experiment are archived in Jefferson Lab's mass storage silo.

Author Contributions The CEBAF Large Acceptance Spectrometer was designed and constructed by the CLAS Collaboration and Jefferson Lab. Data acquisition, processing and calibration, Monte Carlo simulations of the detector and data analyses were performed by a large number of CLAS Collaboration members, who also discussed and approved the scientific results. The analysis presented here was performed by M.K., A.P., and A.A. with guidance from A.A., F.H., O.H., E.P., and L.B.W., and was reviewed by the CLAS Collaboration. S.D., M.B., and K.M. provided expertise on neutrino scattering. S.D. helped guide development of *e*-GENIE. P.C. performed a simulation of the DUNE sensitivity to the oscillation parameters, and determined the impact of our results on the fit.

Author Information Reprint and permission information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to O.H. (hen@mit.edu).

Author List Full CLAS Collaboration. To be added after opt in period.

Methods

Three-flavor neutrino oscillation.

Eq. 7 gives the neutrino oscillation probability for the simplified case of only two types of neutrino. The full three-flavor probability for $\nu_\mu \to \nu_e$ oscillation (in vacuum) is given by [5, 29, 30]

$$P_{\nu_\mu\to\nu_e}(E,L) \approx A\sin^2\frac{\Delta m^2_{13}L}{4E} - B\cos\left(\frac{\Delta m^2_{13}L}{4E}+\delta_{CP}\right)\sin\frac{\Delta m^2_{13}L}{4E}, \quad (7)$$

where $\Delta m^2_{13} = m^2_{\nu_1} - m^2_{\nu_3}$ is the neutrino mass difference squared that determines the oscillation wavelength as a function of L/E and δ_{CP} is a phase that might break charge-parity (CP) symmetry. The coefficients A and B depend primarily on the neutrino oscillation mixing angles, $A = \sin^2\theta_{23}\sin^2 2\theta_{13}$ and $B = -\frac{\sin 2\theta_{12}\sin 2\theta_{23}}{2\sin\theta_{13}}\sin\frac{\Delta m^2_{21}L}{4E}\sin^2 2\theta_{13}$. The different flavor neutrinos (labelled ν_e, ν_μ and ν_τ) are linear combinations of the different mass neutrinos labelled $1, 2, 3$.

Experimental setup and particle identification: CLAS used a toroidal magnetic field with six sectors of drift chambers, scintillation counters, Čerenkov counters and electromagnetic calorimeters to identify electrons, pions, protons, and photons, and to reconstruct their trajectories [18].

We used the e2a data, which was measured in 1999. We measured the momentum and charge of the outgoing charged particles from their measured positions in the drift chambers and the curvature of their trajectories in the magnetic field. We identified electrons by requiring that the track originated in the target, produced a time-correlated signal in the Čerenkov counter, and deposited enough energy in the electromagnetic calorimeter. We identified charged pions and protons by requiring that the track originated in the target and that the measured time of flight agreed (to within ±three times the standard deviation of the detector resolution) with that calculated from the particle's momentum and assumed mass. We identified photons by requiring a signal in the electromagnetic calorimeter which implied a velocity greater than about $0.96c$ (see Ref. [31] for details).

We detected protons with momenta $p_p \geq 300$ MeV/c and angles $\theta_p \geq 10°$, charged pions with momenta $p_\pi \geq 150$ MeV/c and angles $\theta_{\pi+} \geq 10°$ and $\theta_{\pi-} \geq 22°$, and photons with energy $E_\gamma \geq 300$ MeV and $8 \leq \theta_\gamma \leq 45°$. We applied separate fiducial cuts for electrons, π^-, positive particles, and photons, to select momentum-dependent regions of CLAS where the detection efficiency was constant and close to one.

The beam energy equaled the injector energy plus the pass number times the linac energy. The three-pass beam energy was measured using the Hall A arc measurement to be 3.3547 ± 0.0002 (stat) ± 0.0005 (syst) GeV and the injector energy was 0.061 ± 0.0024 GeV. This gave a central linac energy of 1.0979 GeV and Hall B one-, two-, and four-pass beam energies of $1.159 \pm 0.0015, 2.257 \pm 0.001$, and 4.453 ± 0.001 GeV, respectively.

The electron momentum in each sector was scaled to give the correct missing mass for the ^{3}He$(e,e'pp)n$ reaction. These correction factors were all less than 1%.

Nucleon removal energies: The average nucleon removal energy, ϵ, used in reconstructing the incident energies in Eqs. 4 and 5, was determined from the data. We used $\epsilon = M_A - M_{A-1} - m_p + \Delta\epsilon$, where $M_A - M_{A-1} - m_p$ is the difference in the binding energies for knocking a proton out of nucleus A. We adjusted $\Delta\epsilon$ so that the peaks in the E_{cal} spectrum for low-P_T events reconstructed to the correct beam energy. We found $\Delta\epsilon = 5$ and 11 MeV for ^{12}C and ^{56}Fe, respectively, which are consistent with average excitation energies from single-nucleon knockout from nuclei.

Subtraction of undetected pions and photons: Because the CLAS geometrical coverage is incomplete, we needed to subtract for undetected pions and photons to achieve a true 0π event sample. We assumed that the photons came from either radiation by the outgoing electron approximately parallel to its motion or from π^0 decay. We identified the radiated photons by requiring that they be detected within about 30° of the scattered electron and removed them from the data set.

We determined the undetected pion or photon contribution from the events with detected pions or photons. We assumed that the pion-production cross section was independent of $\phi_{q\pi}$, the angle between the electron-scattering plane (the plane containing the incident and scattered electrons and the virtual photon) and the hadron plane (the plane containing the virtual photon and pion). For each detected $(e,e'\pi)$ event, we rotated the pion around the momentum transfer direction $\vec{q}$ randomly many times. For each rotation we determined if the particle would have been detected. We did this by checking to see if the particle was within the fiducial region of the detector. If it was, we used acceptance maps to determine the probability that it would have been detected. The particle acceptance is then $A_\pi = N_{det}/N_{rot}$, where N_{rot} is the number of rotations and N_{det} is the number of times the pion would have been detected. Then the corresponding number of undetected $(e,e'\pi)$ events for that detected $(e,e'\pi)$ event is $(N_{rot} - N_{det})/N_{det}$. We used that as a weight to subtract for the undetected pion events. We did this separately for π^+, π^- and photons.

In order to subtract the undetected $(e,e'p\pi)$ and $(e,e'p\gamma)$ events to get the $(e,e'p)_{0\pi}$ sample, we rotated the proton and pion (or photon) together around $\vec{q}$ and determined the number of detected proton-only events N^p_{det} and the number of detected proton and pion events $N^{p\pi}_{det}$. We used $(N^p_{det} - N^{p\pi}_{det})/N^{p\pi}_{det}$ as a weight to subtract for the undetected $(e,e'p\pi)$ and $(e,e'p\gamma)$ events. We also subtracted the $(e,e'p)$ event sample for extra protons in the same way to get a true $0\pi 1p$ sample. The proton and

pion multiplicity plots are shown in Extended Data Fig. 4.

We also accounted for the effects of, for example, events with two detected pions (or photons). These events could appear as either 0π events, which we subtracted from our event sample, or as 1π events which we subtracted from the 1π sample that we used to subtract from our event sample as described above (i.e., this subtraction from the subtracted events increased the yield). For these events, we calculated the probability that neither pion was detected and subtracted those events from our (e, e') or $(e, e'p)$ samples. We then calculated the probability that only one of the pions was detected and subtracted those events from the $(e, e'\pi)$ or $(e, e'p\pi)$ event samples, which then reduced the subtraction from the (e, e') or $(e, e'p)$ samples.

We considered event multiplicities up to four pions and photons (total) for (e, e') and up to four protons, pions and photons (total) for $(e, e'p)$, where the subtraction converged. The effects of the subtraction (and its convergence) can be seen in Extended Data Fig. 9. The number of events with an undetected $\pi^{\pm}$ or photon is about equal to the number of events with a detected $\pi^{\pm}$ or photon, consistent with the $\approx 50\%$ CLAS geometrical acceptance. The effect of including two $\pi^{\pm}$ or photon events is much less than that of the one $\pi^{\pm}$ or photon events and the effect of including three $\pi^{\pm}$ or photon events is negligible.

GENIE Simulations: We generated events with the electron-scattering version of GENIE (e-GENIE), one of the standard neutrino event generators. e-GENIE has been significantly modified recently to fix known issues and to use reaction mechanisms as close to those of ν-GENIE as possible (version v3.00.06) and, on top of that, we included the effects of electron bremsstrahlung [27]. We used CLAS acceptance maps to determine the probability that each particle was detected and smeared the momenta of the particles with an effective CLAS resolution (we used electrons and proton momentum resolutions of 0.5% and 1%, respectively, for the 2.257 and 4.453 GeV data and 1.5% and 3% for the 1.159 GeV data which was taken with a lower torus magnetic field). We then analyzed the events in the same way as the data.

We used e-GENIE (tune G18_10a_02_11a) with the Local Fermi Gas model; the Rosenbluth cross section for electron-nucleon quasi-elastic scattering; the Berger-Sehgal model [32] of electron-nucleon resonance production, which includes cross sections of 16 resonances calculated in the Feynman-Kislinger-Ravndal (FKR) model [33], without interference between them; the Dytman model [34] of MEC (Meson Exchange currents or $2p2h$ currents), which describes it as a Gaussian distribution located between the quasi-elastic and Δ peaks; and an intranuclear cascade (INC) model of outgoing nucleon final state interactions and hadronization using the Intranuke package [35, 36] with hA2018, an empirical data-driven method, using the cross section of pions and nucleons with nuclei as a function of energy up to 1.2 GeV and the CEM03 [37] calculation normalized to low-energy data for higher energies.

For ν-Genie, we used the same tune (G18_10a_02_11a) with the Local Fermi Gas model; the Nieves cross section for quasi-elastic scattering; the Berger-Sehgal model [32] of resonance production; the Nieves model [34] of MEC; and an intranuclear cascade (INC) model of outgoing nucleon final state interactions and hadronization using the Intranuke package [35, 36] with the same two options as in e-GENIE, hA and hN.

This version of e-GENIE and ν-GENIE described electron- (see Extended Data Fig. 2) and charged current neutrino-scattering cross sections [26] well.

Systematic Uncertainties We considered three major sources of point-to-point systematic uncertainty. When we rotated events containing pions around the momentum transfer vector, we assumed that the cross section did not depend on $\phi_{q\pi}$. We tested the $\phi_{q\pi}$ independence of the pion-production cross section by weighting the subtraction using the measured $\phi_{q\pi}$-dependent $H(e, e'p\pi)$ cross sections of Ref. [38]. This changed the subtracted spectra by about 1%.

Our subtraction of events with undetected pions depends on the CLAS acceptance for such particles. The final spectrum should be independent of the CLAS pion acceptance. We estimated the effect of varying the CLAS acceptance on the undetected particle subtraction by comparing the results using the nominal fiducial cuts and using fiducial cuts with the ϕ acceptance in each CLAS sector reduced by 6° or about 10–20%. This changed the resulting subtracted spectra by about 1% at 1.159 and 2.257 GeV and by 4% at 4.453 GeV.

We also varied the photon identification cuts. We identified photons as neutral particle hits in the calorimeter with a velocity greater than two standard deviations (3σ at 1.159 GeV) below the mean of the photon velocity peak (at $v = c$). We varied this limit by $\pm 0.25\sigma$. This gave an uncertainty in the resulting subtracted spectra of 0.1%, 0.5% and 2% at 1.159, 2.257 and 4.453 GeV, respectively.

We did not consider normalization uncertainties because we normalized the e-GENIE results to the data.

Numerical fit to the oscillation parameters Here, we perform a simplified analysis to demonstrate that the choice of energy response function is critical to the correct extraction of oscillation parameters.

The DUNE experiment will have a kton-scale liquid Argon (LAr) detector, located at approximately 1300 km from the neutrino source. In our fit we considered only the $\nu_\mu \to \nu_e$ appearance channel, and an exposure of 168 kt MW yr. This corresponds to 3.5 years of data taking for a 40 kton detector and a 1.2 MW beam. The beam configuration used corresponds to an optimized design for the DUNE TDR [39] for 120 GeV protons and 1.1×10^{21} protons on target per year. The detection effi-

ciencies and cross sections used also correspond to those used in the DUNE TDR.

This channel has a signal contribution that is mostly ν_e appearance, but there are some $\bar{\nu}_e$ (13 out of 1250 events). The most relevant backgrounds come from events due to ν_e in the initial beam (about 140 events), plus about 60 neutral-current (NC) events that are misidentified as charged-current (CC) events. Thus, for the considered exposure we obtain approximately 1250 signal events, and about 205 background events, after efficiencies (for $\delta_{CP} = 0$, normal mass ordering, and the rest of the oscillation parameters set at the best-fit values for NuFIT 4.1 [40]). The fit has been carried out using GLoBES [41, 42], with the new systematics implementation as in Ref. [43].

In practice, a full DUNE simulation should include data from ν_μ disappearance and combine that with the ν_e appearance data, as the two channels provide complementary information. In our analysis, ν_μ disappearance was not included since the smearing matrices do not reach high enough energies to reproduce the full feeddown effect to lower energies for this channel. Including the ν_μ disappearance results might allow shifting the parameter bias between δ_{CP} and other parameters, such as Δm^2_{31} and θ_{23}. However, these other parameters are well constrained by precision measurements including reactors and other long-baseline experiments. Therefore, shifting significant bias to these other parameters would lead to inconsistencies with world data. In addition, the size of the bias will depend on the true values of the oscillation parameters.

The smearing of the NC background has been done using the same matrices as in Ref. [11]. For the CC events, on the other hand, we used either the smearing matrices derived from the CLAS data or the e-GENIE smearing matrices, see below for additional details. In all cases, the simulated data is binned in 100 MeV bins, between 0.5 and 5 GeV. The effect of the matter potential is fully accounted for in the analysis, using a constant matter density of 2.848 g/cm^3.

The 1250 DUNE events include all event topologies at all outgoing lepton angles. The CLAS and GENIE events used to smear and unsmear the incident energy spectrum include only $1p0\pi$ events for the CLAS acceptance. However, GENIE should describe these QE-like events more accurately than other, more complicated channels.

The simulated data is computed using input values of the oscillation parameters in accordance with the NuFIT 4.1 results from a global fit to present neutrino oscillation data [40]. During the fitting procedure, the solar parameters are kept fixed for simplicity, since their impact on the fit is expected to be minimal. On the other hand, minimization is performed on $\sin^2\theta_{23}$ and Δm^2_{13} according to current uncertainties from the NuFIT 4.1 results, for normal mass ordering. Minimization is also performed over nuisance parameters associated with signal and background systematics (2% for the signal and 5% for the background), using the pull method.

The incident ν_e flux $\Phi_e(E, L)$ smeared with $f_{\sigma_i}(E, E_{rec})$ derived from e-GENIE is shifted to lower energies by about 100 MeV relative to the incident flux as smeared by $f_{\sigma_i}(E, E_{rec})$ derived from CLAS data.

The green lines in Fig. 1 show the confidence regions for simulating and fitting using the CLAS smearing matrices; the true input values for θ_{13} and δ_{CP} are indicated by the green dot. The shaded regions show the confidence regions for simulating the spectra using the CLAS smearing matrices, with the same true input values for the oscillation parameters. However, in this case the data has been fitted using the e-GENIE smearing matrices. Shifting the GENIE $f_{\sigma_i}(E, E_{rec})$ by about 20 MeV so that the energy reconstruction peaks were centered at the beam energy did not significantly affect the results. The best-fit point in the this case (indicated by the black triangle) has $\chi^2 = 10.9$. We used 50 bins in neutrino energy, and fit $\theta_{13}, \delta_{CP}, \theta_{23}$ and Δm^2_{13}, which gives $\chi^2/dof = 0.24$.

Extended Data

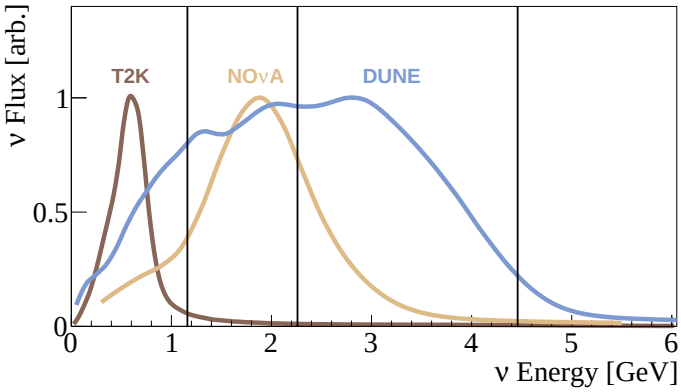

Fig. Extended Data Fig. 1: The energy distribution of different neutrino beams before oscillation. The vertical lines show the three electron beam energies of this measurement.

Table Extended Data Table 1: Fraction of $(e, e'p)_{1p0\pi}$ events reconstructed within 5% of beam energy and the Data/GENIE ratio (D/G).

		1.159 GeV		2.257 GeV		4.453 GeV	
		E_{QE}	E_{cal}	E_{QE}	E_{cal}	E_{QE}	E_{cal}
^{4}He	Data	-	-	17	40	32	37
	GENIE	-	-	20	37	26	28
	Data/GENIE	-	-	0.85	0.93	1.23	1.32
^{12}C	Data	25	51	18	35	31	27
	GENIE	23	63	21	32	26	24
	Data/GENIE	1.09	0.81	0.86	1.09	1.19	1.13
^{56}Fe	Data	-	-	17	23	22	26
	GENIE	-	-	20	24	27	23
	Data/GENIE	-	-	0.85	0.96	0.81	1.13

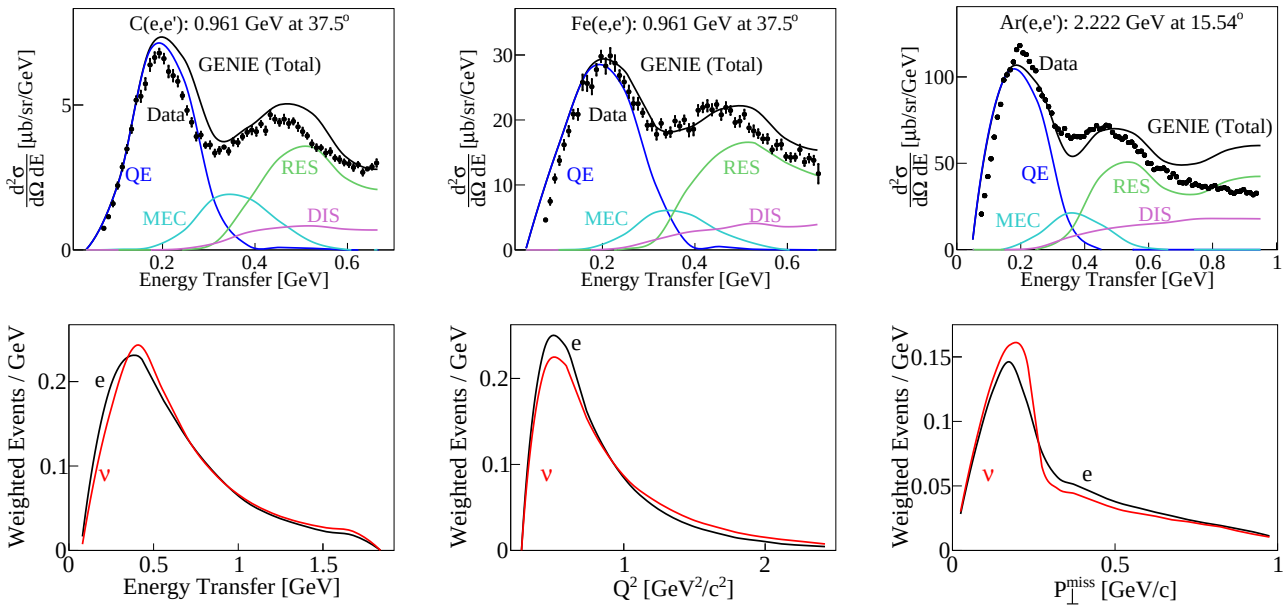

Fig. Extended Data Fig. 2: (Top) 0.961 GeV inclusive electron scattering cross section at 37.5° plotted versus ω for data (points), e-GENIE (black line), and different reaction mechanism components of e-GENIE (quasi-elastic (QE), meson exchange currents (MEC), resonance production (RES), and deep inelastic scattering (DIS)) for (Left) C(e,e') and (Middle) Fe(e,e'). (Right) The same for 2.22 GeV Ar(e,e') data at 15.54° [44]. (Bottom) Comparison of generated $(e,e'p)_{1p0\pi}$ event distributions for e-GENIE (black) and ν-GENIE (red) for 2.257 GeV leptons incident on ^{56}Fe for $Q^2 \geq 0.2$ GeV^2. The plots show the number of events as a function of (Left) energy transfer, (Middle) Q^2, and (Right) P_T. The e-GENIE events are weighted by Q^4 and the plots have been area normalized (i.e., normalized to the same total integral).

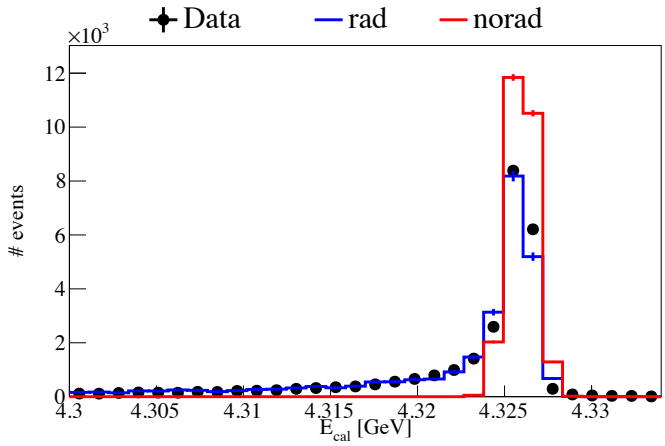

Fig. Extended Data Fig. 3: The number of events plotted as a function of the calorimetric energy for 4 GeV H$(e,e'p)$ for data (black points) overlayed on top of the e-GENIE prediction without (blue) and with (black) electron radiation.

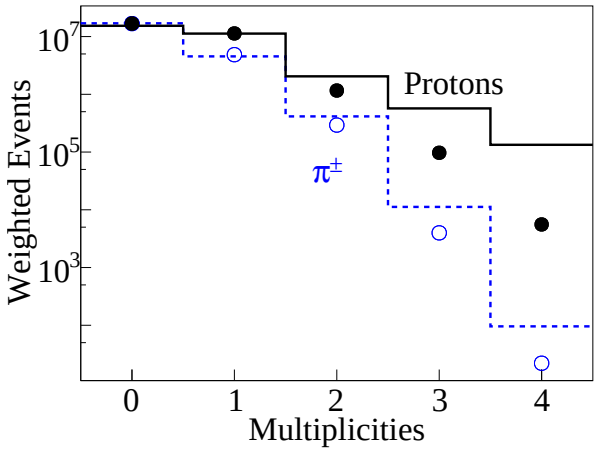

Fig. Extended Data Fig. 4: The proton (black) and charged pion (blue) multiplicities for data (points) and GENIE (lines) for 2.257 GeV carbon.

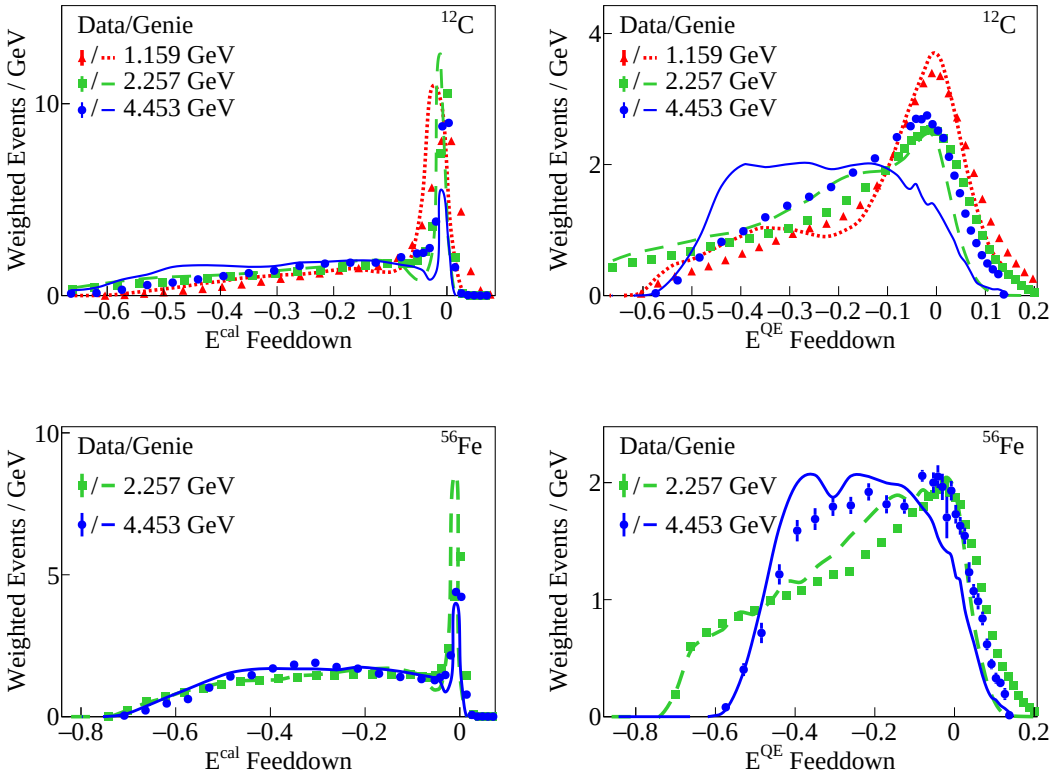

Fig. Extended Data Fig. 5: Fractional energy feed-down $(E_{rec} - E_{true})/E_{true}$ for data (points) and GENIE (histograms) for 1.159 GeV (red triangles and dotted lines), 2.257 GeV (green squares and dashed lines) and 4.453 GeV (blue dots and solid lines) for (upper left) C E^{cal}, (upper right) C E^{QE}, (lower left) Fe E^{cal}, and (lower right) Fe E^{QE}. The plots are area normalized and each bin has been scaled by the bin width.

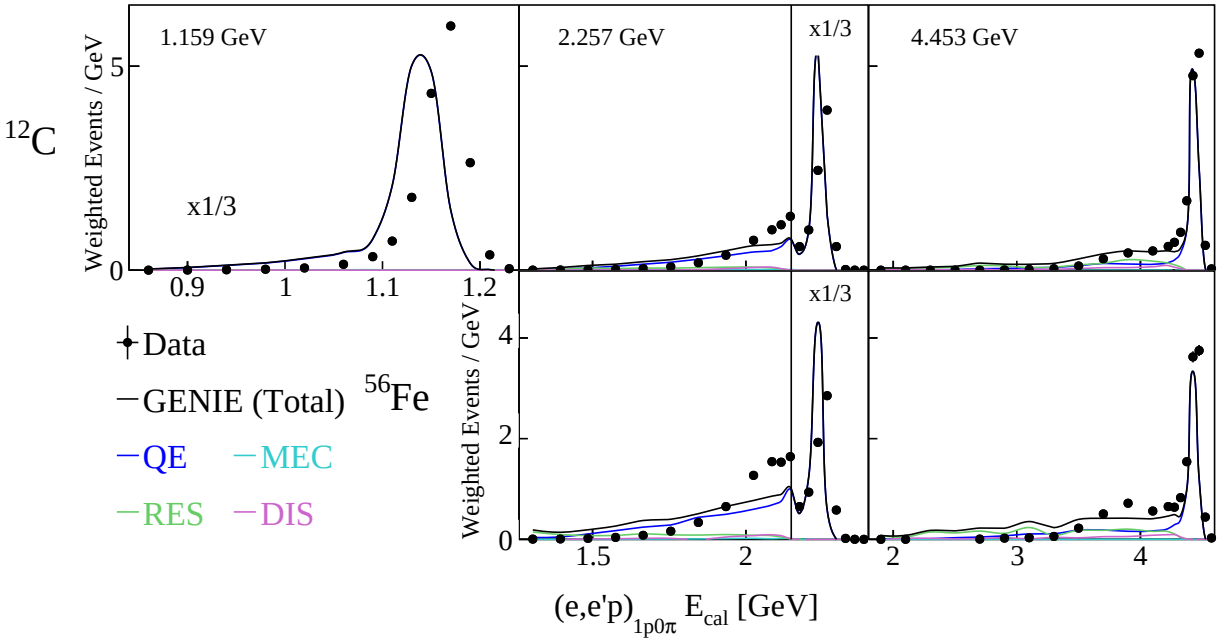

Fig. Extended Data Fig. 6: **Calorimetric Reconstructed Energy for QE events** | The number of weighted A$(e,e'p)_{1p0\pi}$ events per GeV plotted as a function of the reconstructed calorimetric energy E_{cal} for data (black points) and GENIE (black line) cut on $0.8 \leq x_B \leq 1.2$. Different panels show results for different beam energies and target nuclei combinations: Carbon target with 1.159 (Left), 2.257 (Middle) and 4.453 (Right) GeV incident beam and iron target with 2.257 (Middle) and 4.453 (Right) GeV incident beam. Colored lines show the contribution by different processes to the GENIE simulation: QE (blue), MEC (cyan), RES (green) and DIS (magenta). The GENIE results are normalized to the same integral as the data in each panel. Error bars include statistical and systematical uncertainties at the 68% (1σ) level. Error bars are not shown when they are smaller than the size of the data point.

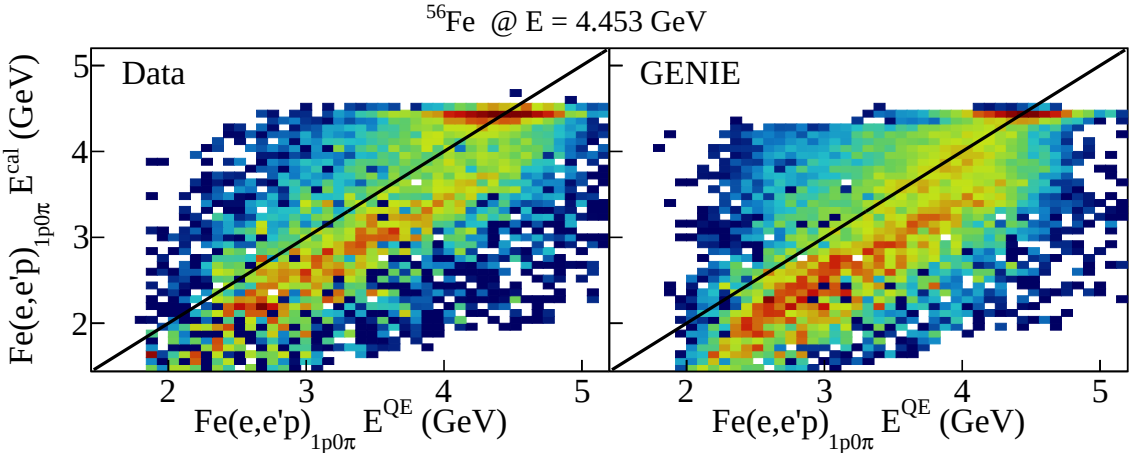

Fig. Extended Data Fig. 7: (a) E_{cal} vs. E_{QE} for 4.453 GeV Fe$(e,e'p)$ events, (left) data and (right) GENIE.

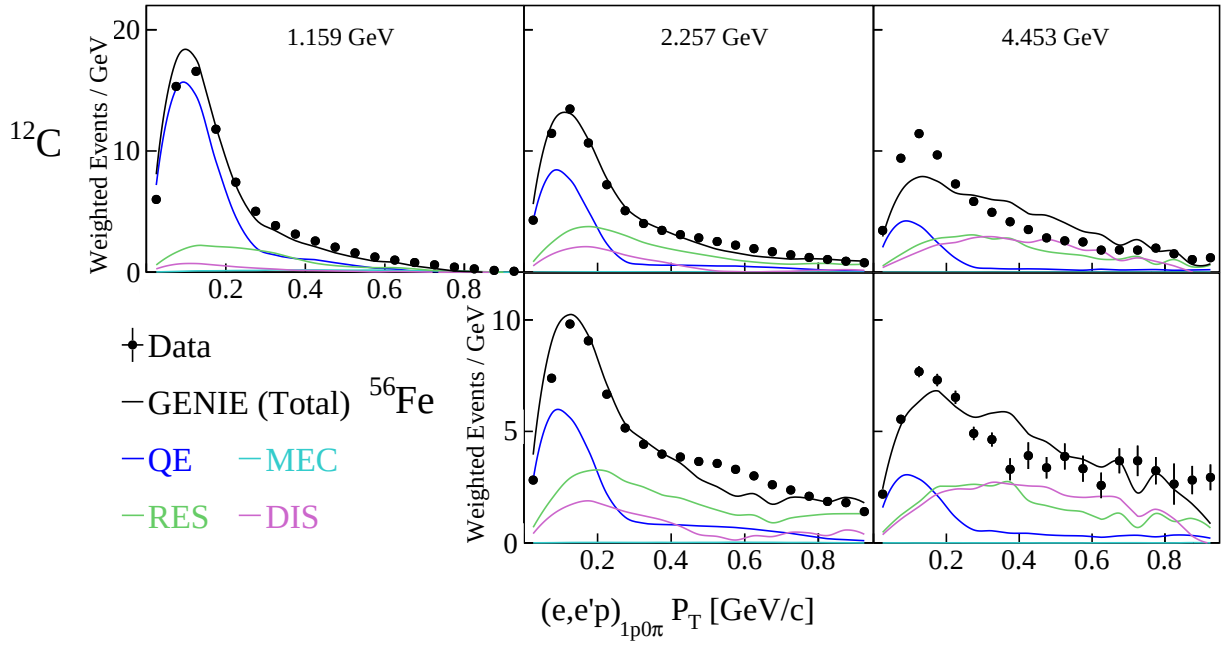

Fig. Extended Data Fig. 8: Transverse missing momentum for ^{12}C and ^{56}Fe for all available energies. The plots are area normalized and each bin has been scaled by the bin width.

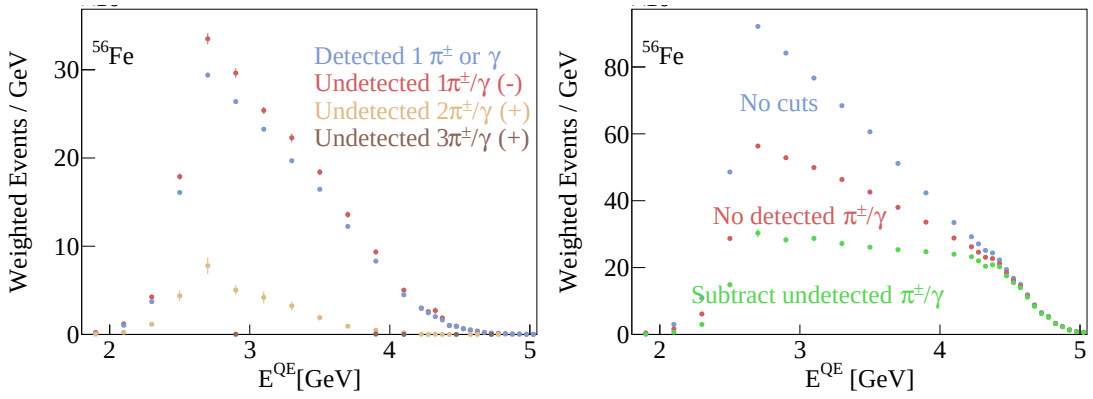

Fig. Extended Data Fig. 9: The effect of undetected pion subtraction. The number of weighted events as a function of reconstructed energy E_{QE} for 4.453 GeV Fe(e,e') events for (Left) events with a detected $\pi^{\pm}$ or photon (blue), events with one (red), two (light brown) or three (dark brown) undetected $\pi^{\pm}$ or photons and (Right) all $(e,e'X)$ events with detected or undetected $\pi^{\pm}$ or photon (blue), (e,e') events with no detected $\pi^{\pm}$ or photon (red), and (e,e') events after subtraction for undetected $\pi^{\pm}$ or photon (green). The uncertainties are statistical only and are shown at the 1σ or 68% confidence level, except for when they are smaller than the size of the data-point in which case they are not visible by eye.