Comparisons and challenges of modern neutrino-scattering experiments

M. Buizza Avanzini [1], M. Betancourt,[2] D. Cherdack,[3] M. Del Tutto [2,4] S. Dytman [5] A. P. Furmanski,[6,7]
S. Gardiner,[2] Y. Hayato [8] L. Koch [9] K. Mahn [10] A. Mastbaum [11] B. Messerly,[5,7] C. Riccio [12,13]
D. Ruterbories [14] J. Sobczyk,[15] C. Wilkinson,[16] and C. Wret [14]

[1]*Laboratoire Leprince-Ringuet, CNRS, Ecole Polytechnique,
Institut Polytechnique de Paris, 91128 Palaiseau, France*
[2]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[3]*University of Houston, Houston, Texas 77204, USA*
[4]*Harvard University, Cambridge, Massachusetts 02138, USA*
[5]*University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA*
[6]*University of Manchester, Manchester M13 9PL, United Kingdom*
[7]*University of Minnesota, Minneapolis, Minnesota 55455, USA*
[8]*Kamioka Observatory, ICRR, The University of Tokyo, Hida, Gifu 506-1205, Japan*
[9]*University of Oxford, Denys Wilkinson Building, Oxford OX1 3RH, United Kingdom*
[10]*Michigan State University, East Lansing, Michigan 48823, USA*
[11]*Rutgers University, Piscataway, New Jersey 08854, USA*
[12]*INFN Sezione di Napoli and Università di Napoli, Napoli 80126, Italy*
[13]*State University of New York at Stony Brook, Stony Brook, New York 11794, USA*
[14]*University of Rochester, Rochester, New York 14627, USA*
[15]*University of Wrocław, Wrocław 50-204, Poland*
[16]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

A set of comparisons among neutrino interaction experiments [MiniBooNE, MINERvA, Tokai-to-Kamioka (T2K), and MicroBooNE] is presented. This gives a broad view of the field of neutrino-nucleus interactions. The emphasis is on charged-current inclusive, quasielastic-like, and pion production experiments. Measurements are compared in new ways. Comparisons of recent data with available event generator codes are made more comprehensively than is regularly found in most previous publications. Generator studies show sensitivities for experimental model dependence. Efficiencies calculated with different generators are presented in a novel way. A comparison of different forward-folding techniques is also presented.

I. INTRODUCTION

Neutrino oscillations are a rich subject which sees a variety of long baseline neutrino oscillation experiments in the 0.1–10-GeV range currently running [1–3], studying the parameters of the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [4,5] and other effects beyond the Standard Model [6–8]. Future experiments [9,10] aim to be dominated by systematic uncertainties and will depend critically on the details of modeling of neutrino-nucleus interactions. These experiments aim to reconstruct the neutrino energy of a particular flavor using the products of the neutrinos

charged-current interaction, at a specific distance from the neutrino source. Although most running experiments employ the use of a near detector to constrain some aspect of the neutrino flux and interaction uncertainties, the study of neutrino interactions is important to understand the effects which cause the neutrino energy and flavor to be poorly reconstructed, as these do not always cancel in a near-far extrapolation. An experimental and theoretical review can be found in Ref. [11].

There have been a variety of reviews of neutrino interactions [11] with the main goal of describing the physics content. Many types of interaction are possible, all denoted by the principal interaction, i.e., what the neutrino does microscopically. The dominant interactions are quasielastic (QE, where the neutrino interacts with a single nucleon and only a single nucleon is emitted), multinucleon processes (usually referred to as $2p2h$ because the interaction is predominantly with 2 nucleons), resonant (RES, where the struck nucleon is excited to one of the broad

nucleon resonances which decay to a nucleon and various mesons), and deep inelastic scattering (DIS, where the neutrino interacts with either a nucleon or predominantly a quark and a variety of mesons are emitted). Neutrinos can interact via charged current (CC) or neutral current (NC). For CC interactions, a charged lepton is emitted and its flavor provides a way to infer the neutrino flavor.

Many neutrino cross-section measurements have moved away from publishing data according to interaction mode—which requires heavily model-dependent corrections to be made—and now publish data of a more objective nature. Emphasis is now on topological content of the final state observed in the detector. For example, CCQE cross-section measurements have generally been replaced with measurements that specify a final state of one muon and no pions. These measurements are referred to as CC0π or CCQE-like. Thus cross-section measurements tend to have a mix of contributions to their signal; for instance, CC0π without any restriction on the number of nucleons will see contributions from CCQE, two-particle, two-hole (2p2h), and pion production processes where the pion is absorbed. CC1π measurements have different meanings depending on the neutrino energy. At low energies ($E_\nu < 1.5$ GeV), the $\Delta(1232)$ resonance dominates. However, higher-energy neutrinos can excite a variety of higher-mass resonances or interact directly with the quarks (DIS).

Neutrino beams are wideband, i.e., the flux distribution width is a large fraction of the peak energy. In addition, the beam contains neutrinos of different flavors. All these properties must be measured or calculated by each experiment. Although most neutrino detectors have good acceptance over a broad range of kinematics, there are still kinematic holes, e.g., very low-energy hadrons, and neutral particles are often hard to detect. These aspects make the experiments difficult and Monte Carlo (MC) calculations are often used to fill the holes. Hence, the established method for interpreting the results is to generate events and make predictions for interaction generators using the same signal definition, neutrino flux, and target as the experiment, and comparing the calculations to data. We follow this method throughout this article.

The Tensions workshop series attempts to examine how measurements are defined/carried out and make comparisons with a variety of Monte Carlo calculations. The first Tensions (2016) workshop [12] discussed difficulties in comparing results with different signal definitions, published comparisons of data against a variety of models, and discussed their model dependence. All cross-section experiments were represented, and generator experts were present. General issues of experiments and modeling were discussed from different points of view, and adjustments to the methods were suggested to the experiments.

One of the interesting studies in the first Tensions workshop was pion production where signal definitions are especially complicated. Each experiment chooses

analysis methods and signal definition to best utilize their detector and their view of what measurements are needed. The result is that comparing measurements is a difficult job that must be done via generators which are capable of reproducing a wide range of signal definitions. Event generator codes were able to accurately simulate the signal definitions of the MiniBooNE [13] and first MINERvA results [14] and as a result compare the same models to each dataset. One of the main conclusions of Tensions 2016 was that the two datasets were largely incompatible with our models as a result of comparison with a variety of calculations.

This article comes from the second Tensions workshop which was held in the summer of 2019 at University of Pittsburgh, USA, as well as further investigation which followed this workshop. For this article, we discuss new measurements from Tokai-to-Kamioka (T2K), MINERvA, and MicroBooNE, which have been published with significant improvements in methodology. We use three recent versions of GENIE [15], one recent version of NuWro [16], and one recent version of NEUT [17]. To compare the generators to each other and to experiments' cross-section data, we use NUISANCE [18].

The emphasis is on CC-inclusive (Sec. VI), CC0π (Sec. VII), and CC1π (Sec. VIII) measurements. We make comparisons of the event generators against recent cross-section data, and present novel comparisons beyond model-to-data comparisons. In the last few years, collaboration between theorists and generator authors has significantly increased. The result is improved models and the plots contained in this paper show the results. We compare similar measurements in the same kinematics, and pseudoefficiencies from generator predictions were compared with actual efficiencies for each class of measurement to look for model dependence. We give a retrospective on the MiniBooNE results and their relevance to modern measurements is provided (Sec. IX and the Appendix). In Sec. X we compare generator predictions for quantities that are of great interest to experiments, such as neutral particle energy content and the dependence of efficiency on detection threshold. Finally, some important observations on forward-folding techniques are made (Sec. XI).

The workshop attendees were experts from each of the experiments' cross-section programs, and generator experts were present throughout, all of whom are contributing authors to this article.

II. EVENT GENERATOR OVERVIEW

This section briefly introduces the generator models that are compared to data later in this work. Generator models are essential to experiments, providing a means by which to estimate efficiencies and background contributions, develop selection cuts and corrections, and assess systematic uncertainties due to interaction modeling. Understanding interaction-model dependence

TABLE I.   QE-like models implemented in each generator.

Generator	Nuclear model	QE model	M_A^{QE} (GeV)	2p2h NN correlations	Long-range NN correlations
GENIE v3 G18_02a	RFG [23]	Llewellyn-Smith [24]	0.99	Empirical [29]	None
GENIE v3 G18_10a	LFG	Nieves [25]	1.05	Nieves [26,28]	RPA [25]
GENIE v3 G18_10b	LFG	Nieves [25]	1.05	Nieves [26,28]	RPA [25]
NuWro 19.02	LFG	Llewellyn-Smith [24]	1.03	Nieves [26,28]	RPA [33]
NEUT v5.4.0.1	LFG	Nieves [25]	1.05	Nieves [26,28]	RPA [26]

in neutrino-oscillation experiments is a topic of particular interest [9,19,20]. The generators used in existing accelerator experiments, GENIE [15] and NEUT [17,21], have been developed primarily within the experimental community, drawing on published theoretical work or direct collaborations with theorists to develop models. While in many cases models are similar with respect to the underlying theory, differences in implementation and parameter choices can lead to important differences in generator predictions; this is evident in the comparisons shown in later sections. The use of generators for interpretation of experimental data introduces additional challenges. Complete coverage of phase space is necessary for use in a full detector simulation, typically leading to inconsistency in regions which are poorly understood or where models overlap. Mechanisms for assessment of systematic uncertainties are also necessary and challenging; these must incorporate all relevant degrees of freedom within models while remaining computationally tractable. An additional generator, NuWro [16], is frequently used as a benchmark. With rapid integration of improved theoretical models and a more consistent handling of certain interaction modes, NuWro has provided both a point of comparison and an avenue for developing new models which are later integrated into GENIE and NEUT. Detailed discussion of the models implemented in each of these generators is provided in the following subsections. First, the general structure and common components are introduced.

In all generators considered here, neutrino-nucleus interactions are modeled as a two-step process in an impulse approximation. Here, the first step is the primary interaction, where interactions occur on individual bound and moving nucleons. The second step is final-state interactions (FSI), where interactions of particles from the primary vertex with the residual nucleus are considered. A major difficulty in the interaction modeling is to consistently describe the nucleus where different nuclear model choices are important. For quasielastic scattering, meson exchange current or two-particle two-hole interactions where the neutrino scattering is with a correlated nucleon pair, and resonance production, the nucleus is modeled as an ensemble of nucleons. In DIS, nucleon substructure in the form of quarks becomes most relevant. Finally, in coherent neutrino-nucleus scattering (COH), the nucleus is essentially a single composite object. These three

classes of models tend to be developed separately, and must be merged into a consistent picture within the generators. The primary focus of this work is to study processes where the neutrino interacts with one or two nucleons at a time: QE, MEC or 2p2h, and RES.

Model choices for QE-like nucleon-level interactions are summarized in Table I. For modeling of the initial-state bound nucleon momentum distributions, two types of models are typically used. The relativistic Fermi gas (RFG) model is the traditional approach, with implementations of the Smith-Moniz [22] and Bodek-Ritchie [23] versions available in generators. The more recent local Fermi gas (LFG) model provides a more realistic distribution based on the position-dependent local nucleon density. For describing quasielastic scatters, generators have historically used the RFG and Llewellyn-Smith [24] model; in GENIE v3.00.06 (v3), the G18_02a model set uses these models, to provide a point of comparison to these choices. In current versions, generators have shifted to use the LFG and Valencia group's self-consistent QE model. This model by Nieves *et al.* [25,26] includes long-range nucleon-nucleon random-phase approximation (RPA) correlations and Coulomb effects for the outgoing charged lepton on single-nucleon [one-particle, one-hole ($1p1h$) or true QE] and multinucleon ($2p2h$) interactions. These effects can modify interactions significantly at energies near or below 1 GeV, so these models are more applicable to these lower neutrino energies while predictions are identical to Llewellyn-Smith for higher energies. For multinucleon interactions, the Valencia $2p2h$ model [27] has been widely used, particularly since Gran and Sanchez [28] studied its features and application in collaboration with the theory authors. It is included as a distinct interaction channel which explicitly incorporates additional nucleons in the final state. Broader applicability is gained by suppressing events for which $q_3 > 1.2$ GeV/c, where q_3 is the magnitude of the three-momentum transfer. An alternative model, based on an empirical enhancement of the total cross section in a region of energy-momentum transfer space between QE and RES due to multinucleon interactions [29], is used in GENIE v3 G18_02a which is similar to code commonly used in experiments discussed here. Final-state nucleons from $2p2h$ processes are generally distributed via phase space [30]. Consistency in the isospin decomposition of emitted nucleons is an ongoing

TABLE II.   Models for pion production in the $\Delta(1232)$ resonance region. R-S refers to Rein-Sehgal [34], B-Y to Bodek-Yang [37], B-S to Berger-Sehgal [36], and S-O to Salcedo-Oset [38].

Generator	Resonance model	M_A^{RES} (GeV)	Nonresonant model	Form factor	π FSI model
GENIE v3 G18_02a	B-S	1.23	Scaled B-Y [39]	Ref. [40]	Empirical [15]
GENIE v3 G18_10a	B-S	1.23	Scaled B-Y	Ref. [40]	Empirical [15]
GENIE v3 G18_10b	B-S	1.23	Scaled B-Y	Ref. [40]	S-O
NuWro 19.02	Δ-only [40]	0.94	Scaled B-Y	Ref. [40]	S-O
NEUT v5.4.0.1	B-S	0.95	Scaled B-Y	Ref. [40]	S-O

problem. While some advanced theoretical models [31] have explored interference between one- and two-body currents, the implementation of these in event generators is just beginning to be explored [32].

The models for production of pions in the $\Delta(1232)P_{33}$ resonance are summarized in Table II. At the core of most generator resonance models is the Rein-Sehgal [34] (R-S) model. The R-S model uses a nonrelativistic quark model [35] to derive helicity amplitudes to produce resonances, and then describes the subsequent decay of those resonances. Berger and Sehgal [36] updated the R-S model to include effects due to lepton mass.

The resonance parameters such as masses, decay widths, and form factors have changed significantly as the data improved since the development of the R-S model in 1981. All generator groups have implemented these updates, and in certain models have incorporated updated tuning to neutrino scattering data. There are several ways to describe nonresonant pion production [11]. Strength can come from the tail of DIS processes, referred to here as a scaled Bodek-Yang [39] (B-Y) model, or via low-order diagrams [26]. The scaled B-Y choice uses a factor that decreases model strength to achieve agreement with data, and necessarily includes both resonant and nonresonant contributions.

For the kinematic region for values of W (invariant mass) greater than 1.4 GeV, model choices are given in Table III. This is the kinematic region that is critical to successful interpretation of DUNE [9] data. Berger and Sehgal [36] provide models for all highly rated nucleon resonances [41] as well as coherent pion production [42], both of which are updated versions of the corresponding Rein-Sehgal [34,43] models. Unlike Δ production, these resonances have weaker excitation strength from neutrino interactions and are poorly known as a result. The relevant couplings have not been fit to modern electron-nucleon data and the decay distributions are considered to be isotropic. A major challenge is finding the optimal way to describe the soft, or shallow, inelastic scattering (SIS) kinematic region, which covers values of W between 1.4 and 2.0 GeV. There exists no hard boundary and theoretical guidance is meager. As a result, modeling at the boundary between RES and DIS is empirical and event generator groups have adopted different strategies, using both RES and DIS models to describe the data in this region. All generators have a transition between resonance- and DIS-dominated

kinematic regimes, though they differ in the location and treatment of this transition. GENIE, for example, has a sharp boundary at $W \sim 1.9$ GeV (depends on model) while NuWro linearly interpolates over a range just above the Δ peak. Lacking true nonresonant models, generator codes use scaled versions of the B-Y model.

For $W > 2.2$ GeV, true DIS processes are dominant and the strategy of using PYTHIA [44] is optimal. At low values of W, resonance decays to baryon and meson states are reasonably well understood and are the appropriate description. However, treatments that extend DIS models into the SIS region at lower W rely on empirical models [37,46] for hadronization processes.

Final-state interactions occurring during the propagation of produced particles through the remnant nucleus are described using intranuclear cascade models based on free hadron-nucleon cross sections. In models simulating the full intranuclear cascade (NEUT, NuWro, and GENIE G18_10b), nuclear medium modifications are added in a local-density approximation. These have been derived by Salcedo-Oset [38] for pions and by Pandharipande-Pieper [47] for nucleons. The GENIE G18_02a and G18_10a configurations have a data-driven model [15] which has partial inclusion of medium dependence effects.

A. GENIE overview

The GENIE [15] generator evolved from NEUGEN, the primary event generator for the MINOS experiment [48]. The current major version, GENIE v3, was released in October 2018. GENIE includes a variety of model sets, which are user selectable via configuration files. The work described here uses three GENIE configurations: G18_02a, G18_10a, and G18_10b. These alphanumeric codes represent the choice of GENIE physics models: the G18_02a model set is an updated version of the historical default model from GENIE v2, and the G18_10a and G18_10b configurations have newer models for almost all processes relevant here [49]. For all G18 model sets, the last character defines the FSI treatment used. The letter "a" refers to the hA effective cascade model and "b" to the hN full cascade model, each described below. The official labels for the GENIE configurations studied here all include an additional suffix _02_11a, which is omitted for simplicity. This suffix indicates that several model parameters were tuned to

TABLE III.   Models for meson production in the kinematic region with resonances of mass larger than $\Delta(1232)$ resonance.

Generator	Resonance model	DIS model	RES/DIS boundary	Coherent model
GENIE v3 G18_02a	B-S	B-Y [39]/PYTHIA [44]	1.93	B-S [42]
GENIE v3 G18_10a	B-S	B-Y/PYTHIA [44]	1.93	B-S [42]
GENIE v3 G18_10b	B-S	B-Y/PYTHIA [44]	1.93	B-S
NuWro 19.02	None	B-Y/PYTHIA [44,45]	1.3–1.6	B-S
NEUT v5.4.0.1	B-S	B-Y with Custom [21] (W < 2.0) and PYTHIA (W > 2.0)	1.4–2.0	B-S

neutrino scattering data on hydrogen and deuterium targets [50].

The G18_10 model sets use a LFG nuclear model and the Nieves models for CCQE [26] and $2p2h$ [26,28]. The $2p2h$ implementation in GENIE is fully described in Ref. [51]. The G18_02a model set provides an updated version of models used within the community for many years: a Bodek-Ritchie [23] RFG nuclear model including a high-momentum tail due to short-range nucleon-nucleon correlations, the Llewellyn-Smith model [24] for primary CCQE processes, and an empirical $2p2h$ model based on fits to MiniBooNE data [29]. All model sets use a dipole axial form factor and BBBA07 vector form factors [52], and for nuclear targets they apply Pauli blocking requiring that the momentum of the outgoing nucleon exceeds the Fermi momentum k_F for the nucleus in question.

Although all GENIE resonance models are based on the Rein-Sehgal treatment [34], a variety of changes have been implemented, e.g., regular updates for new resonance masses and widths. For all versions, the effect of the lepton masses on the allowed region of phase space is taken into account. G18_10a/b fully include Berger-Sehgal lepton-mass corrections [53] and the pion-pole diagram [36]. In the G18_02a model set, the axial and vector form factors are the modified dipole forms as in the Berger-Sehgal model. In GENIE v3 G18_10a/b, the Δ form factors have been updated from fits to MiniBooNE data [54]. While the $\Delta \rightarrow \pi$ decay is isotropic for the G18_02a model set, G18_10a/b use the angular distribution from Rein-Sehgal [34] which was fit to Argonne National Laboratory (ANL) data [55]. All GENIE models neglect interference between resonances. The nonresonant contribution to pion production comes from scaled versions of the Bodek-Yang [39] model, with hadronization described by the custom Andreopoulos, Gallagher, Kehayias, and Yang (AGKY) model [37,46] and PYTHIA [44].

DIS processes are handled in GENIE with a mix of Bodek-Yang [56], a special fragmentation model [15], and PYTHIA [44]. The transition between resonance and DIS processes comes at a cutoff value of W which is part of the single-nucleon fit [50]; the value for model sets used here is 1.93 GeV. However, the Bodek-Yang model is valid for all energies above the πN threshold and is used (scaled to neutrino-hydrogen and neutrino-deuterium scattering data) for nonresonant processes below the cutoff value.

GENIE has a unique FSI model [15,57] called hA which uses a single interaction to approximate the multiple steps

in traditional cascade models. This has been tuned to hadron-nucleus scattering data for a wide range of nuclei and energies. This model is denoted with an a in the configuration name and is used in G18_02a and G18_10a. A multistep cascade model called hN, which includes medium corrections for pions [38] and nucleons [47], is denoted with a b and used in G18_10b. Both models use SAID hadron-nucleon fits to data [58] in calculations of mean-free path and various angular distributions.

B. NEUT overview

The NEUT Monte Carlo generator has been developed for Super-Kamiokande, T2K, and other experiments, and simulates neutrino-nucleus interactions from $\sim$100 MeV to $\sim$100 GeV. Simulations shown in this paper were performed using NEUT version 5.4.0.1.

The QE and pion production models are as described in Tables I–III. NEUT takes into account inter-resonance interference, consistently using the Rein-Sehgal [34] model. Multipion production events are generated with the custom code that assumes the Koba-Nielsen-Olsen scaling [59] and the measured multiplicity of pions as a function of W. DIS events are generated with PYTHIA5/JETSET [44]. For the multipion and DIS channels, the GRV98 parton distribution functions, including Bodek-Yang corrections [60], are used. $2p2h$ events are simulated with the Valencia model [26], and coherent pion production events are simulated with the Berger-Sehgal model [42]. To reflect the idea of the formation zone, hadrons and mesons produced by interactions other than the (quasi-)elastic scatterings or coherent scattering have their production positions shifted toward the direction of the outgoing particles.

For pion FSI, the mean-free paths (MFP) of absorption and inelastic scattering are calculated with the model developed by Oset et al. [61] below 400 MeV/c. Above 500 MeV/c, the MFP are extracted from pion-nucleon scattering data. In the transition region in between, the fraction of low-energy model decreases linearly from 1 at 400 MeV/c to 0 at 500 MeV/c. The normalizations of the mean-free paths were tuned using pion-nucleus scattering data [62]. The kinematics are determined using the results of a phase-shift analysis with the medium correction suggested by Seki et al. [63]. If a pion is absorbed in the nucleus, multiple nucleons are emitted. Kaon and eta

FSI is handled similarly to pions by the cascade model, with the MFP deduced using kaon-nucleon or eta-nucleon scattering data. The nucleon FSI model is based on the work by Bertini *et al.* [64]. Finally, nuclear deexcitation is considered in the case of an oxygen target, and includes production of additional low-energy γ or nucleons following the neutrino interaction.

C. NuWro overview

The NuWro [16] generator has become an important "sandbox" for other generators since its inception at University of Wrocław around 2005, introducing new theoretical models which are used for testing before being adopted by NEUT and GENIE. It covers a neutrino energy range from $\sim$100 MeV to $\sim$100 GeV. For neutrino scattering on a free nucleon, NuWro includes contributions from the three different regions discussed above.

Quasielastic interactions are described using the Llewellyn-Smith [24] model with BBBA05 [65] vector and dipole axial vector form factors. For nuclear effects, NuWro offers many options: global and local Fermi gas, hole spectral function [66], effective spectral function [67], and a density- and nucleon momentum-dependent potential [68]. In the case of LFG, long-range correlations calculated with an RPA technique can be included [33]. For the simulations used in this paper, the LFG model including RPA effects has been used.

Resonance production is described with a model optimized for the Δ-resonance peak region. The Δ resonance is explicitly included, with nucleon-Δ form factors taken from Ref. [40] with parameters obtained as a simultaneous fit to ANL [55] and Brookhaven National Laboratory (BNL) [69] of the DIS contribution, extrapolated down to the pion threshold. This is added incoherently to the Δ contribution as described in Ref. [45]. In the region $W \in (1.3, 1.6)$ GeV, NuWro employs a linear interpolation between the described model and the DIS pion production cross sections. For nuclear target reactions the Δ self-energy is included in an approximation based on Ref. [70]. The effect of the finite Δ lifetime is included as described in Ref. [16]. The angular distribution of pions resulting from Δ decays is described based on values of density matrix elements informed by ANL and BNL experimental studies [55,69].

In the DIS region, NuWro uses the Bodek-Yang prescription [71]. Hadronic final states are generated using PYTHIA [44] fragmentation routines [44], with modifications described in Ref. [72]. NuWro performance is optimized to reproduce charged hadron multiplicities reported in Ref. [73]. For DIS events, formation zone effects are included [16].

Simulation of $2p2h$ events can be done using a variety of models for the overall contribution to the cross section and distribution of final-state leptons. For the charged-current reaction, the default is the Valencia model [26,28]. Other options include a transverse enhancement model (which can be applied to neutral current reactions) [74], the Marteau-Martini model [75–77], and the SuSAv2 model [78]. Simulations in this work use the Valencia $2p2h$ model. For the hadronic part, NuWro uses a model proposed in Ref. [30]. The basic assumption is that the distribution of outgoing nucleons in the hadronic center of mass frame is uniform. However, this assumption can be relaxed with a suitable parameter. Implementation of the Valencia and SuSAv2 $2p2h$ models is done with five tabulated nuclear response functions for carbon and oxygen. Extrapolation to heavier targets is done using methods similar to those proposed in Ref. [51].

A custom cascade model for pions and nucleons propagating through the nucleus with a realistic density profile [16,79] is employed. The key inputs are microscopic pion-nucleon and nucleon-nucleon in-medium cross sections, with Pauli-blocking effects implemented locally. In the Δ region, pion-nucleon cross sections are described with the Salcedo-Oset [80] model. At larger pion energies, free pion-nucleon total cross sections are taken from experimental data and differential cross sections are provided by the SAID model [58]. For nucleon-nucleon elastic interactions, in-medium modifications are taken from Ref. [47], while for inelastic reactions the model from Ref. [81] is adopted. Short-range correlation effects are included by reducing the nuclear density near every hadron-nucleon interaction point [82,83] and introducing a compensating factor at larger distances [79].

III. CROSS-SECTION EXTRACTION METHODS

In recent years, there have been significant developments in our understanding of the potential for bias in neutrino cross-section extraction techniques. The bias of particular concern is towards the input Monte Carlo used when developing an analysis and extracting the cross section; the assumptions made about the channel to be measured can affect the analysis. The bias can enter in a number of ways, including signal definitions and strategies for handling backgrounds. The data utilized in this work have been produced over a number of years, as these techniques have evolved, and as such the results do not consistently use the latest techniques or follow the measurement strategies that each experiment would use today. In this work, we consider all results as they were published or released, without systematically evaluating the techniques used to extract them. However, we briefly comment here regarding the potential issues in the extraction of neutrino cross-section measurements, as these issues are critically important and their treatment will determine the long-term utility of measurements from currently operating experiments.

Crucial areas of potential model dependence are (1) the choice of signal definition and selection variables; (2) the treatment of unmeasured phase space and efficiency corrections; and (3) unfolding methods and techniques [84].

(1) Extracting cross-section measurements as a function of variables other than those which can be directly measured in the detector, or which are a convolution of them, requires Monte Carlo corrections to translate to them from the measured variables, which necessarily introduces some model dependence. More subtly, selection cuts which are as a function of variables which are not accessible in the detector also require similar model-dependent corrections. Typically safe variables are kinematic variables of final-state particles.

(2) A related issue is the treatment of unmeasured regions of phase space and efficiency corrections for the final-state particles. If a selection excludes certain regions of phase space (e.g., low proton momenta), then a naive efficiency correction would simply add the missing strength based on the Monte Carlo prediction. A now widespread technique to mitigate this problem is to explicitly remove such regions of low or no efficiency from the signal definition. More subtle issues relate to regions of rapidly changing efficiency in binned cross-section measurements, where the model may be implicitly relied upon when integrating the efficiency across each bin. Despite increasing awareness of issues relating to efficiency corrections across the field, this remains a very challenging problem with no easy solution, particularly for signal definitions which allow event topologies with many final-state particles. Reconstruction algorithms tend to perform badly for very high-multiplicity events, and correlations between particles in the final state may have a significant impact on the efficiency, in ways that are very challenging to capture in an analysis.

(3) Unfolding is a general term for removing the smearing of resolution due to properties of a measuring device used for a measurement [85–87]. It describes the process for producing a result as a function of a true variable (e.g., the true muon momentum) from a reconstructed variable (e.g., the reconstructed muon momentum). A common issue when unfolding is that statistical fluctuations in reconstructed space can cause large fluctuations between bins in true space. Various methods exist for regularizing, or smoothing, unfolded results, by preferring results in which the result fits some prior expectation, which necessarily adds some bias [85–87]. The challenge is to tune the strength of the regularization to balance the bias in the result with the variance in each bin (known as the bias-variance trade-off). The most popular unfolding method used in the field is D'Agostini unfolding [88,89], which can be characterized as an algorithm for maximum-likelihood estimation with early stopping [86,87]. In the D'Agostini method, the input Monte Carlo is used as a reference for regularizing the result. Each iteration of the algorithm reduces the strength of the regularization, and the size of the bias, by allowing the bin-to-bin variance to increase. The main problem for the D'Agostini method is that the stopping criterion is generally set by Monte Carlo studies of the potential bias. If the simulations used for those studies are substantially different from data, it is likely that the result is strongly affected by the bias towards the input Monte Carlo. Similar issues exist for other unfolding and regularization methods. One solution would be to present regularized and unregularized results or to avoid regularization altogether, which in the D'Agostini case would mean iterating until convergence. Unregularized results can sometimes look unphysical, but are statistically correct. It is also possible to avoid unfolding altogether, and smear the model to match the data, rather than unsmear the data to match the model [90], as is discussed in Sec. XI. Even in this case, there remain complications when resolutions and efficiencies depend on more than the variable being measured; experiments must ensure the smearing functions are as complete as possible.

The extent to which any of the above issues can bias a result depends on many aspects of the detector design and analysis methodology, so cannot be assessed outside the collaboration reporting the result, and therefore cannot be reliably quantified for historical purposes. As a consequence, studies here are only based on Monte Carlo studies and are then necessarily incomplete. This can indicate sensitivity, but not fully assess the consequences.

Adopting methods that minimize the model dependence of neutrino cross-section results will help to ensure their continued reliability when the models currently used in neutrino interaction simulation packages become obsolete. Heavily model-dependent results can only be judged in the context of the models used to extract them, and will have a limited utility as a result.

Furthermore, one always needs to take care not to overinterpret fluctuations in plots of differential cross-section measurements, especially for unfolded results. The unfolding procedure causes correlations in the data points' uncertainties, making it difficult to impossible to judge how well a given prediction fits the data by just looking at the plots. Instead, it is necessary to calculate measures for the goodness of fit that take the correlations into account and then judge the model-data agreement based on these. The most commonly used measure for this purpose is the Mahalanobis distance, which is often called "the chi square" in the field of physics. This is largely objective but interpreting a χ^2 value is not always easy as the effect of correlations is not shown in the plot and the value can be dominated by single bins which are not easily visible. Comparing the generator predictions without any
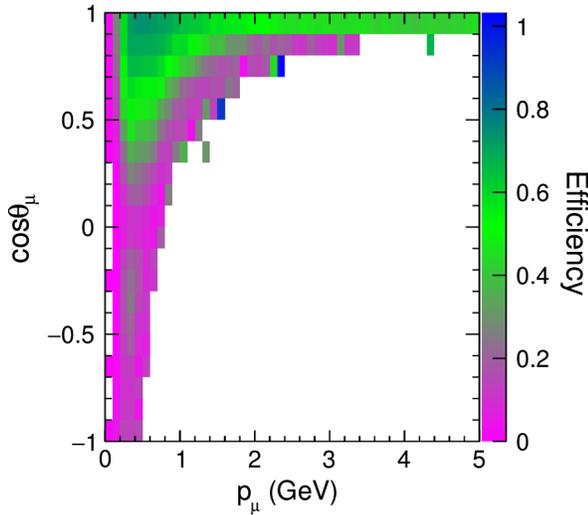
FIG. 1. T2K selection efficiency for the ν_μ–hydrocarbon CC0π measurement presented as a function of p_μ–$\cos\theta_\mu$, corresponding to the analysis presented in Ref. [91].

bin-to-bin correlations directly to one another is a more straightforward exercise.

IV. EFFICIENCIES

Determination of efficiencies is a very important part of every experiment. Although they should be based on data as much as possible, Monte Carlo is often used to supplement or replace data. To study one aspect of the resultant model dependence, selection efficiencies were provided by participating experiments for the measurements discussed in this work. These make it possible to see potential sources of difference between measurements. For example, although two experiments may measure the same process,

they might be sensitive to different regions of phase space. As well as the providing efficiencies in terms of measured variables presented by each analysis (e.g., p_μ–$\cos\theta_\mu$), they were also provided in terms of various true kinematic quantities of interest, for example true q_0–q_3, or Q^2–W. There are well-documented problems with making measurements in these variables, but they allow the qualitative comparison of experiments in terms of the variables of interest from a theoretical point of view.

Efficiencies were typically provided in two-dimensional phase space, with all other degrees of freedom implicitly integrated out. For example, Fig. 1 shows the selection efficiency as a function of p_μ–$\cos\theta_\mu$ for T2K's ν_μ–hydrocarbon CC0π measurement from Ref. [91]. For ease of presentation, in this paper we compare single-dimensional efficiencies with generator predictions, for different measurements. In order to collapse a two-dimensional efficiency as provided by the experiments to the single-dimensional efficiencies shown here, it is necessary to go through an intermediate, model-dependent step. This can be understood by considering Fig. 1 and asking what the efficiency is for a p_μ bin. Clearly it depends on the distribution of events within that bin in $\cos\theta_\mu$, as the efficiency also varies as a function of $\cos\theta_\mu$. Therefore we have to multiply the efficiency in two-dimensions by the predicted rate given the relevant flux, incident neutrino species, and target material, and then collapse that distribution onto the axis of interest, before dividing by the total number of simulated events in that bin. This *pseudoefficiency* is in principle dependent on the model used to transform from two to one dimensions. The resulting efficiencies as a function of p_μ when following this procedure from the p_μ–$\cos\theta_\mu$ in Fig. 1 are shown in Fig. 2 for all generators used in this work. The differences
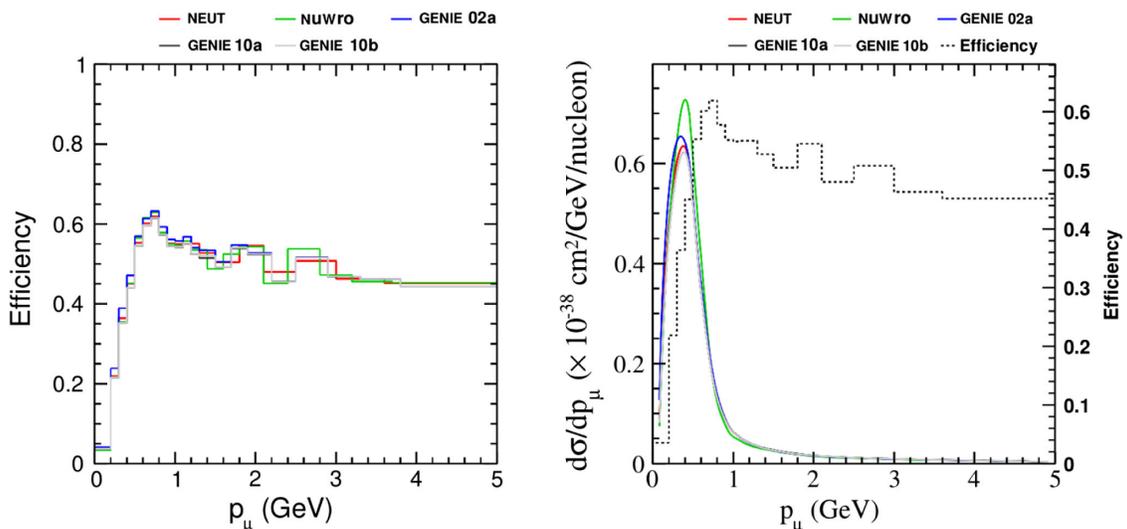
FIG. 2. Left: A comparison of the one-dimensional selection efficiencies when different generator models are used to collapse from two to one dimension. Right: Various generator predictions are shown and compared to the NEUT selection efficiency.
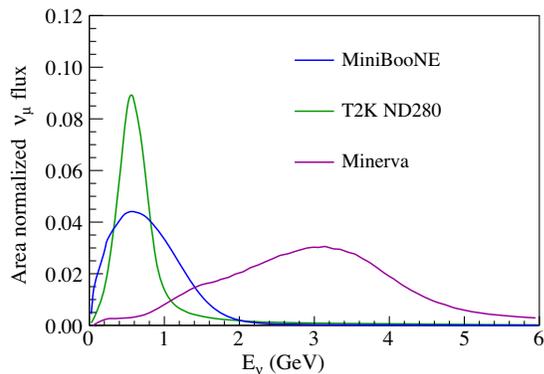
FIG. 3. Area-normalized neutrino flux distributions for the T2K, MINERνA low-energy run, and MiniBooNE experiments. Only the muon neutrino component is considered; the wrong-sign background contributions are not presented. The T2K flux is that at the off-axis near detector, ND280. The MicroBooNE flux shape is almost identical to the MiniBooNE flux shape.

between the one-dimensional efficiencies produced using different model assumptions are small in this case, but are present. Figure 2 also shows an example of the sort of efficiency plots found later in this work, where different generator predictions of the cross section are compared with the efficiency. In this figure, and all others in the work, the NEUT generator prediction is used to calculate the overlaid efficiency.

We note that this procedure, integrating over a model prediction for other kinematic variables, is explicitly done by experiments when they present results as a function of a single variable, and has the potential to introduce severe model dependence in some cases.

V. EXPERIMENT OVERVIEW

A. Neutrino fluxes

Neutrino interaction measurements are intrinsically produced averaged over an incoming neutrino energy spectrum. Figure 3 shows the neutrino flux spectra for the three experiments considered in this paper. The different distributions of neutrino energies impact the interpretation of measurements and their relationships to one another. All beams start with a primary proton beam which impinges on a target to produce pions and kaons. These secondary particles are focused using magnetic horns and subsequently decay into neutrinos. Table IV shows the proton energy, target material, decay pipe length, and resultant

peak neutrino energy for the three neutrino beams used by the running experiment discussed in this work. It should be noted that the T2K experiment uses the off-axis strategy, which results in a lower and narrower energy spectrum than the on-axis flux in the same beam.

B. The T2K experiment

The T2K experiment has detectors both on- and off-axis. The data considered in this paper are from the off-axis near detector, ND280 [92], which sits 2.5° from the beam center. The off-axis angle provides a narrow-band neutrino beam peaked at $E_\nu \sim 600$ MeV with a suppressed high-energy tail. FLUKA is used to model the proton-graphite target interaction and GEANT3/GCALOR is used to propagate the particles through the horns and decay volume. The simulation uses proton beam monitor measurements as inputs, and the modeling of hadronic interactions in the target is constrained using thin target hadron production measurements from the NA61/SHINE experiment at ~30 GeV [93]. Downstream of the decay volume and absorbers, there are two muon monitors (MUMON) [94,95] monitoring the muon direction for muons with $p_\mu > 5$ GeV. The neutrino rate and direction is monitored by the INGRID detector, which sits near ND280 but is centered on the neutrino beam, spanning ±5 m in both dimensions [96].

ND280 is composed of several subdetector systems, all enclosed in a 0.2-T magnetic field. The Pi-Zero Detector (P0D) [97] is composed of orthogonal scintillator tracking planes interleaved with refillable water layers, and sheets of brass. The two fine-grained detectors (FGDs) [98] are composed of orthogonal scintillator tracking planes (CH), one of which also contains alternating planes filled with water (FGD2). The FGD's scintillator bars are composed of 86.1% carbon, 7.4% hydrogen, 3.7% oxygen, 1.7% titanium, 1% silicon, and 0.1% nitrogen by mass. Interleaved between each of the three subdetector modules (P0D, FGD1, FGD2), as well as downstream of FGD2, are gaseous argon time-projection chambers (TPCs) [99] which measure track characteristics at high resolution, providing sign selection and momentum measurements of tracks. All of the subdetector modules are surrounded by electromagnetic calorimeters (ECals) [100]. The ECals surround the tracker and consist of 13 modules made up of plastic scintillator bars alternating with lead sheets. Finally, surrounding the ECals, burrowed in slats in the magnet, is the side muon range detector [101], used to tag

TABLE IV. Various parameters that control the neutrino beam energy distributions for the beams considered in this paper.

Experiment	Proton energy	Target material	Decay pipe length	Peak energy
T2K	30 GeV	Graphite	90 m	0.6 GeV
MINERνA	120 GeV	Graphite	500 m	3 GeV
Mini/Micro-BooNE	8 GeV	Beryllium	50 m	0.6 GeV

escaping particles and particles entering from outside the detector, e.g., cosmic muons.

The neutrino event generator NEUT [17] is used to simulate neutrino interactions in the detector; more details are reported in Sec. II B and GEANT4 version 4.9.4 [102] is used to simulate the detector response and passage of particles through materials.

C. The MINERvA experiment

MINERvA is located on axis in the NuMI beamline at Fermilab. The on-axis beam peaks at 3 GeV and contains 95% ν_μ, with the remainder consisting of $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$ [103]. The data presented here are from the Low Energy run; the MINERvA collaboration has started to produce several cross-section measurements with the NuMI Medium Energy flux [104,105]. The neutrino beam is simulated with GEANT4 9.2.p03 [102], and constrained with thin-target hadron production measurements and an *in situ* neutrino electron scattering constraint [106].

The MINERvA detector uses plastic scintillator bars with a triangular cross section, arranged in three each 60° from the other two. The MINERvA detector [107] is segmented longitudinally into several regions: nuclear targets, the scintillator tracker, and downstream electromagnetic and hadronic calorimeters. The nuclear target region contains five solid passive targets of carbon (C), iron (Fe), and lead (Pb), separated from each other by four or eight scintillator planes for vertex and particle reconstruction. Targets 1, 2, and 3 contain distinct segments of Fe and Pb planes that are 2.6 cm thick; target 3 also has a C segment which is 7.6 cm thick, and target 5 has Fe and Pb segments which are 1.3 cm thick. The tracker is made solely of scintillator planes; the fiducial volume contains 106 planes. The target mass of the fiducial volume is a mix of carbon in 88.51%, hydrogen in 8.18%, oxygen in 2.5%, titanium in 0.47%, chlorine in 0.2%, aluminum in 0.07%, and silicon in 0.07%. The MINOS Near Detector is 2 m downstream of the MINERvA detector and serves as a magnetized muon spectrometer [108].

The neutrino event generator GENIE2.12.6 with some additions is used to simulate neutrino interactions in the detector. $2p2h$ interactions and long-range correlations estimated using the RPA from Valencia model are included. The interactions and decays of particles produced in the neutrino interactions of the final-state particles that exit the nucleus are simulated by GEANT4 9.4.2 [102].

D. The MiniBooNE experiment

MiniBooNE used neutrinos from the Booster Neutrino Beam. The beam has an average energy of 800 MeV, and is 93.6% ν_μ with 5.9% (0.5%) contamination of $\bar{\nu}_\mu(\nu_e, \overline{\nu_e})$. The beam simulation was tuned to external hadron production measurements from HARP experiment [109].

The detector is composed of 800 tons of mineral oil (CH_2) that serves as both the target for neutrino interactions

and the medium in which charged particles produced in neutrino interactions radiate Cherenkov and scintillation photons. The photons are detected on an array of 1520 photomultipliers, and the resulting spatial and temporal patterns of light are used to identify and reconstruct the interactions. For particles above Cherenkov threshold, the scintillation light is a minor component; however, the scintillation light is important for interactions that do not produce any particles above threshold. Scintillation light provides only position and energy information, whereas Cherenkov light additionally provides direction information.

The NUANCE event generator was used to simulate neutrino interactions and GEANT3-based program to simulate the response of the detector to neutrino interactions. The NUANCE generator main components include: a relativistic Fermi gas model for CCQE and NC elastic scattering, a baryonic resonance model for CC/NC single-pion production model, a deep inelastic scattering model, and a final-state interaction model to simulate reinteraction of final-state hadrons in nuclear medium. The simulation did not include long-range correlation such as RPA or $2p2h$ nuclear effects.

E. The MicroBooNE experiment

MicroBooNE sits in the booster neutrino beam (BNB) upstream of MiniBooNE. As a smaller detector than MiniBooNE it subtends a smaller angle relative to the neutrino beam direction, and as such has a slightly different energy spectrum, but the difference is very small. MicroBooNE uses the same flux simulation chain and data constraints as MiniBooNE.

The MicroBooNE detector is a Liquid Argon Time Projection Chamber with 85 tons of active mass. In the MicroBooNE detector, charged particles leave trails of ionization electrons as they traverse the argon and also create prompt ultraviolet scintillation photons. The electrons drift in an electric field to one side of the detector, where they are detected by a series of sensing wires on three separate planes. The scintillation photons are instead detected by 32 photomultipliers. The liquid argon acts both as target material and as detector for charged particles.

For the analyses covered here, neutrino interactions are simulated using the GENIE v2.12.2 version with the addition of the empirical MEC, while cosmogenic particles (which constitute a significant background in many MicroBooNE analyses) are simulated with CORSIKA [110]. Particles are then propagated by GEANT4, while the simulation of the MicroBooNE detector is performed in the LARSOFT framework [111].

VI. INCLUSIVE INTERACTIONS

A. Introduction

Inclusive interactions include all the neutrino interactions with the nucleus, without any particular requirement on the

number or type of final-state particles. Inclusive measurements are important because they often allow selecting a large sample of neutrino interactions with high efficiency and purity, they are mildly sensitive to hadron uncertainties, and finally, allow testing multiple contributing processes at once. In addition, some calculations require integration over the hadronic final state and are available only for the inclusive cross sections. In this section, a comparison between charged-current inclusive measurements from T2K, MicroBooNE, and MINERvA is presented. The measurements from T2K and MicroBooNE use similar energy neutrino fluxes (see Sec. V A) and the same observables from the muons—$\cos\theta_\mu$ and p_μ. The measurement from MINERvA uses a neutrino flux with a higher mean of 3 GeV, and is made with different observables—both muon and hadronic information is measured to test more information about the models. Additionally, the signal definition is different so MINERvA sees different contributions from interaction channels, such as CCQE and CC DIS. MINERvA's charged-current inclusive measurement reports events with low three-momentum transfer, $|q_3| < 0.8$ GeV, which does not include the multi-π and DIS events, with more details in Sec. VI D. Meanwhile, T2K and MicroBooNE include all the events in the inclusive sample. The next sections outline details about the T2K, MicroBooNE, and MINERvA measurements.

B. T2K results

The T2K collaboration produced a muon-neutrino CC-inclusive double-differential cross section on a carbon target using a beam of muon neutrinos with a peak energy of 0.6 GeV [112]. The cross section was extracted as a function of the unfolded muon momentum p_μ and $\cos\theta_\mu$, where θ_μ is the angle between the muon and the average incoming neutrino direction.

The main selection consists of four samples of ν_μ charged-current interactions inside an $\sim 1 - m^3$ FV in FGD1, based on the angle of the muon with respect to the detector axis: forward going (FWD), backward going (BWD), high-angle forward going (HAFWD), and high-angle backward going (HABWD). The aim of the selection is to find events with at least a muon in the final state. Depending on the angle of the muon inside the detector, different selection criteria are used. Forward-going and backward-scattered muons are identified by the energy deposited in the gaseous argon TPCs and their track curvature in the magnetic field, i.e., their momentum. Muons that are scattered close to perpendicular to the neutrino direction do not cross a TPC, and are instead identified in the electromagnetic calorimeters by a multivariate discriminator which separates muon and pion tracks from showering particles. Additionally, the Side Muon Range Detector is used to tag forward-going high-angle muons, as well as veto seemingly backward-going cosmic muon background events. The composition of the selected

TABLE V. Relative composition (%) of the charged-current signal in the four samples of the T2K measurement according to NEUT. [112].

	FWD	BWD	HAFWD	HABWD
QE	44.7	82.0	67.3	83.2
$2p2h$	7.5	5.5	7.2	5.5
RES	25.4	8.6	17.6	8.0
DIS	19.9	3.8	7.2	3.4
COH	2.5	0.0	0.7	0.0

signal events depends on the sample and is shown in Table V. This is due to a differing efficiency under the two model assumptions, especially at low-momentum, forward-going bins in muon kinematics. In this region, there is an especially strong contribution of DIS with a large fraction of the energy in the hadronic system. Since the muon kinematics should not differ much within a single bin, the deviations probably come from differences in the distributions of hadronic particles, which suffer from large modeling uncertainties.

Migration of events between the kinematic bins is handled by an unfolding procedure (see [112] for more details) producing a spectrum of events in "true" kinematic variables. The number of background events is constrained by the selection of dedicated control regions, and fitted to the data in a simultaneous fit. That means both the signal and background event rates are determined together and correlations and migrations between samples are handled naturally. The resulting signal event distributions are then scaled by a bin-dependent efficiency correction to account for detection and reconstruction inefficiencies, and converted into a flux-integrated cross section using the known neutrino flux profile and number of target nuclei.

To judge the model dependence of the unfolding procedure, the result is extracted with two different models as the nominal assumptions. Despite best efforts to make the unfolding procedure as model independent as possible, the extracted cross sections differ slightly but noticeably between the two. This is due to a differing efficiency under the two model assumptions, especially at low-momentum, forward-going bins in muon kinematics. Since the muon kinematics should not differ much within a single bin, these deviations probably come from differences in the distributions of hadronic particles. Although they are integrated over for the inclusive cross section, they still affect the overall reconstruction efficiency. This underlines the need to understand and investigate the efficiency performance of an analysis for "hidden" variables that are not included in the signal definition.

C. MicroBooNE results

The MicroBooNE collaboration produced a muon-neutrino CC-inclusive double-differential cross section using a

beam of muon neutrinos with a mean energy of 0.8 GeV [113]. The cross section was extracted as a function of the reconstructed muon momentum, p_μ^{reco}, and the muon direction, $\cos\theta_\mu^{\text{reco}}$, where θ is the angle between the muon and the beamline.

The inclusive sample of ν_μ charged-current interactions is selected [114,115] inside a 44-t fiducial volume (FV), requiring one muon with or without the presence of other particles in the final state. The muon can be either contained inside, or can exit the detector, and can have any direction. The muon momentum is calculated using multiple Coulomb scattering, by fitting an argon-tuned Highland formula along the candidate muon trajectory [116]. This method is equally applicable to muons that are fully contained, and those that exit, but suffers from relatively poor (10–20%) resolution. No angular or energy cut is applied. The deposited charge per unit length (dQ/dx) is used to discriminate muons from protons. Several algorithms ensure the quality of the fitted track by limiting the allowed spatial dispersion of the reconstructed hits with respect to the track hypothesis. Since the MicroBooNE detector is on the Earth's surface and takes several milliseconds to read out data, cosmic rays are the dominant background for an inclusive muon neutrino analysis. A series of algorithms is used to identify these background event, by looking at tracks that traverse the detector from top to bottom, that do not match with the light activity arriving in time with the neutrino beam, and by looking at the Bragg peak and Michel electrons to identify stopping muons, which overall reduce the cosmic rate by more than 3 orders of magnitude.

The analysis follows a so-called forward-folding technique and the measurement is presented in terms of reconstructed variables instead of true, unfolded, ones. More details on the forward-folding method and its limits of applicability are given later in Sec. XI. The analysis reported in the following uses data collected between February and July 2016, and corresponds to 1.6×10^{20} protons on target.

D. MINERvA results

MINERvA reported the first inclusive charged-current double-differential cross section as a function of three-momentum transfer and available energy. The three-momentum transfer is obtained using the four-momentum transfer Q^2 and the energy transfer q_0:

$$q_3 = \sqrt{Q^2 + q_0^2}, \qquad (1)$$

where the Q^2 is obtained using the energy of the neutrino, muon angle, and momentum ($Q^2 = 2E_\nu(E_\mu - p_\mu \cos\theta_\mu) - M_\mu$). The muon momentum is calculated by using the ionization energy loss for a muon traversing the material in the MINERvA detector in

conjunction with the momentum reconstructed from MINOS experiment. The neutrino energy is reconstructed using ($E_\nu = E_\mu + q_0$). The energy transfer, q_0 is estimated by summing the visible hadronic energy and applying model-dependent corrections for unobserved neutrons and nucleon removal energy [117]. A new variable called available energy (E_{avail}) was defined to unfold and report the cross section. This variable is close to the true energy transfer, but does not include energy of the neutrons (because they leave very small energy in the detector), or other forms of missing energy (nuclear recoil, binding energy, etc.). The resolution of E_{avail} varies from 55 to 38%. The true E_{avail} is defined as

$$E_{\text{avail}} = \sum T_p + \sum T_{\pi^\pm} + \sum E_{\text{particles},} \qquad (2)$$

where $\sum T_p$ is the proton kinetic energy, $\sum T_{\pi^\pm}$ is the pion kinetic energy

$$\sum E_{\text{particles}}$$
$$= \sum E_{K^\pm} + \sum E_{e^\pm} + \sum E_{\pi^0} + \sum E_{\gamma,} \qquad (3)$$

and $\sum E_{\text{particles}}$ is the total energy of other particles except neutrons. In the reconstruction E_{avail} is estimated using the calorimetric sum of the visible energy not associated with the muon.

The inclusive sample of ν_μ charged-current interactions is selected using events in MINERvA's 5.3 ton active-tracker FV; the sample includes muon tracks that are matched to a track in the MINOS detector and $\theta_\mu < 20°$ and $p_\mu > 1.5$ GeV. The signal definition is charged-current ν_μ with 2 GeV $< E_\nu < 6$ GeV in the true neutrino energy, $p_\mu > 1.5$ GeV and $\theta_\mu < 20°$. The measurement is reported for low three-momentum transfer ($q_3 < 0.8$ GeV). Selection on neutrino energy might introduce model dependence; a better signal definition should avoid cuts on neutrino energy or any other observable with model dependence.

An unfolding procedure [88] with four iterations was applied in two dimensions to translate the data from reconstructed quantities to true (E_{avail}, q_3). GENIE2.8.4 was used to correct for the acceptance of the FV, the efficiency of the MINOS muon match, and the subtraction of small (3%) neutral-current and μ^+ backgrounds [117].

E. Comparisons of event generator predictions

To explore in more detail the contributions from each experiment, Tables VI–VIII show the event generator predictions broken down by true interaction channels, including the signal definition cuts for each measurement. Concerning the QE predictions of the generators at the three experiments, we see by far the lowest prediction from G18_02a at MicroBooNE, which then is largest prediction

TABLE VI.   Event generator cross-section predictions for the different true interaction channels for T2K charged-current inclusive cross section ($\times 10^{-38}$ cm^2/nucleon).

Generator	QE	$2p2h$	$1\pi^+1p$	$1\pi^0$	$1\pi^+1n$	$N\pi$	1η	DIS	$1\pi^{+,0} + N\pi + 1\eta + DIS$
GENIE v3 18_02a	2.80	0.82	1.04	0.31	0.32	0.53	0.03	0.42	2.65
GENIE v3 18_10a	2.79	0.55	1.12	0.33	0.34	0.57	0.03	0.46	2.84
NuWro 19.02	3.19	0.55	1.16	0.28	0.23	0.12	0.03	0.79	2.61
NEUT v5.4.0.1	2.84	0.56	1.19	0.36	0.37	0.56	0.07	0.50	3.05

TABLE VII.   Event generator cross-section predictions for the different true interaction channels for MicroBooNE charged-current inclusive cross section ($\times 10^{-38}$ cm^2/nucleon).

Generator	QE	$2p2h$	$1\pi^+1p$	$1\pi^0$	$1\pi^+1n$	$N\pi$	1η	DIS	$1\pi^{+,0} + N\pi + 1\eta + DIS$
GENIE v3 18_02a	2.88	0.75	1.03	0.40	0.39	0.31	0.03	0.05	2.21
GENIE v3 18_10a	3.30	0.73	1.10	0.43	0.41	0.34	0.03	0.05	2.36
NuWro 19.02	3.42	0.58	1.07	0.38	0.30	0.14	0.00	0.23	2.13
NEUT v5.4.0.1	3.33	0.61	1.14	0.48	0.49	0.42	0.08	0.06	2.68

TABLE VIII.   Event generators predictions for the different true interaction channels for MINERvA charged-current inclusive cross section ($\times 10^{-38}$/nucleon).

Generator	QE	$2p2h$	$1\pi^+1p$	$1\pi^0$	$1\pi^+1n$	$N\pi$	1η	DIS	$1\pi^{+,0} + N\pi + 1\eta + DIS$
GENIE v3 18_02a	7.99	1.62	1.88	0.33	0.43	0.05	0.00	0.00	2.69
GENIE v3 18_10a	6.76	1.49	1.93	0.33	0.44	0.05	0.00	0.00	2.75
NuWro 19.02	7.85	1.49	2.60	0.39	0.46	0.05	0.00	0.00	3.51
NEUT v5.4.0.1	7.29	1.40	2.30	0.48	0.81	0.07	0.02	0.00	3.68

at MINERvA and T2K has similar prediction for G18_02a and G18_10a. The G18_10a prediction roughly agrees with NEUT and NuWro at MicroBooNE, but is 10% lower at MINERvA energies, whereas NEUT and NuWro seem to scale similarly to each other. At T2K energies, NuWro is 13% higher compared with NEUT and both GENIE versions. This is interesting because many similar choices have been made according to Table I, making implementation important for understanding these differences. Looking at the $2p2h$ prediction, NEUT, NuWro, and G18_10a all utilize the Valencia $2p2h$ model, but we see up to 20% difference. This difference could come from binding-energy implementation and final-state interaction modeling. GENIE G18_02a uses Empirical MEC for the $2p2h$ model, yet still produces similar $2p2h$ predictions for MicroBooNE and T2K, but smaller prediction for MINERvA. This is likely due to both models being tuned to MiniBooNE CCQE-like data, but handling the neutrino energy scaling differently. This suggests that different datasets could be used to constrain the energy dependence of the models. In T2K, $2p2h$ predictions are similar for each generator except for G18_02a, which is 27% higher because it uses a different $2p2h$ model (Empirical) than the others and different nuclear model (RFG).

Focusing on the MicroBooNE and T2K breakdown, the single-π production, multi-π production, single-η production, and DIS contributions are all different. Some of this is due to nomenclature of defining interaction modes, e.g., summing the resonant, multi-π, η, and DIS contributions is required to get directly comparable contributions from the generators. There are differences in strategy; for instance, NuWro favors the CC1$\pi + 1$p for its tuning whereas NEUT tries to tune to all the data and inflates uncertainties accordingly. GENIE [50] also does a fit to all available data with different choices. NuWro, GENIE, and NEUT are tuned to ANL and BNL bubble chamber data [40,118]. All have similar predictions for MicroBooNE where $\Delta(1232)$ excitation dominates due to the neutrino energy range. When scaling up the same interaction mode to MINERvA energies, the predictions separate with NuWro largest; NEUT is somewhat smaller and GENIE smaller yet. This again displays the different choices for energy dependence of different interaction modes. Although the choices allowed given the large uncertainties in the single-nucleon data are important, nuclear modeling can also cause differences. This is especially important for higher-energy long-baseline neutrino oscillation experiments such as NOvA and DUNE.

F. Comparisons of generators with T2K, MicroBooNE, and MINERvA inclusive data

This paper reports comparisons of the T2K, MicroBooNE, and MINERvA's double-differential cross-section measurements with the different event generator predictions in Figs. 4–6. T2K and MicroBooNE have similar kinematic variables and use similar fluxes, but different targets. MINERvA's target is the same as T2K's, but the measurement variables, acceptance, and fluxes are different from T2K and MicroBooNE. Therefore, the comparisons between these experiments is not straightforward. Importantly, while T2K and MicroBooNE exclusively measure the muon in these data, MINERvA's measurement uses kinematic variables that contain the muon and hadronic information from the event, which is a further test of the models used to reconstruct the neutrino energy in oscillation experiments, noting that the NOvA and DUNE neutrino experiments both use the full kinematics of the event to reconstruct the neutrino energy at the cost of model-dependent corrections.

Figure 4 shows the inclusive T2K result compared to some model predictions and Fig. 5 shows the same model comparisons with the MicroBooNE data.

Based on the χ^2 values, the models describe the T2K data poorly. The best $\chi^2/\mathrm{dof} = 105/71$ is given by the NEUT event generator, while NuWro shows the worst agreement among the compared generators, with a $\chi^2/\mathrm{dof} = 201/71$. Most of the difference between NuWro and the other generators seems to be located as a lower prediction in the high-momentum, very forward-going data points ($\cos\theta > 0.92$ and $p_\mu > 1$ GeV), as well as a higher prediction in the peak of the cross section at around $p_\mu = 500$ MeV/c.

In MicroBooNE's muon-neutrino double-differential cross-section measurement, about half of the events are quasielastic processes, and the remaining half are $2p2h$ and resonance processes, with a small contribution from deep inelastic scattering. The data likewise are described poorly by the generators especially in the forward-going region where there is the largest tension between the data and the generators. Part of the disagreement with the MC prediction could be from the $2p2h$ description which uses equal
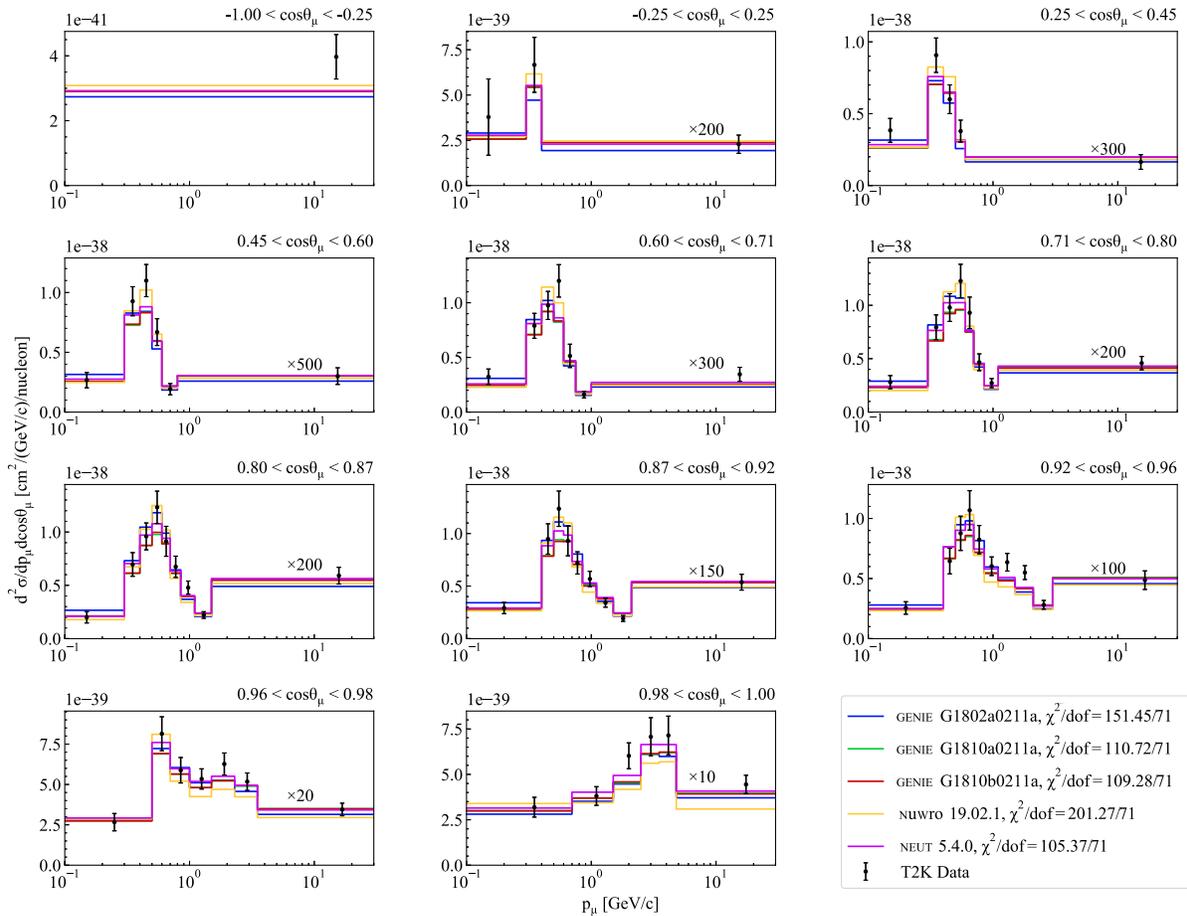
FIG. 4. T2K double-differential cross-section $\mathrm{d}^2\sigma/(\mathrm{d}p_\mu^{\mathrm{reco}}\mathrm{d}\cos\theta_\mu^{\mathrm{reco}})$ in nine regions of $\cos\theta_\mu^{\mathrm{reco}}$ is compared to NUWRO, NEUT, and three versions of GENIE.
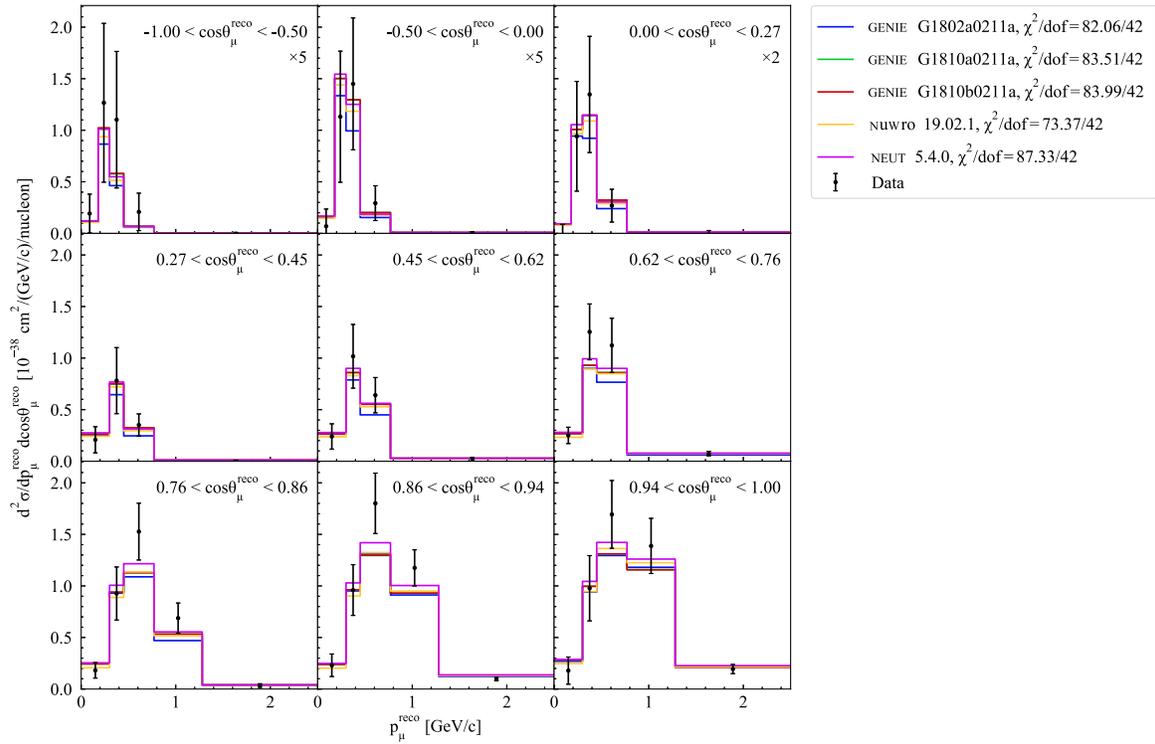
FIG. 5. MicroBooNE double-differential cross-section $d^2\sigma/(dp_\mu^{reco}d\cos\theta_\mu^{reco})$ in nine regions of $\cos\theta_\mu^{reco}$ is compared to NuWro, and three versions of GENIE.

number of protons and neutrons for argon. NuWro and NEUT has the worst score of $\chi^2/dof = 87/42$. The difference is much smaller than in the T2K case, though. Also, as the difference between NuWro and the other generator predictions is less pronounced, it is more difficult to say what kinematic region is actually causing the difference. Based

on separate χ^2 analysis for each measurement, the generators have equivalent ability to describe general characteristics at neutrino energies of ~1 GeV in carbon and argon.

In MINERvA's double-differential cross-section measurements, the region at low available energy below
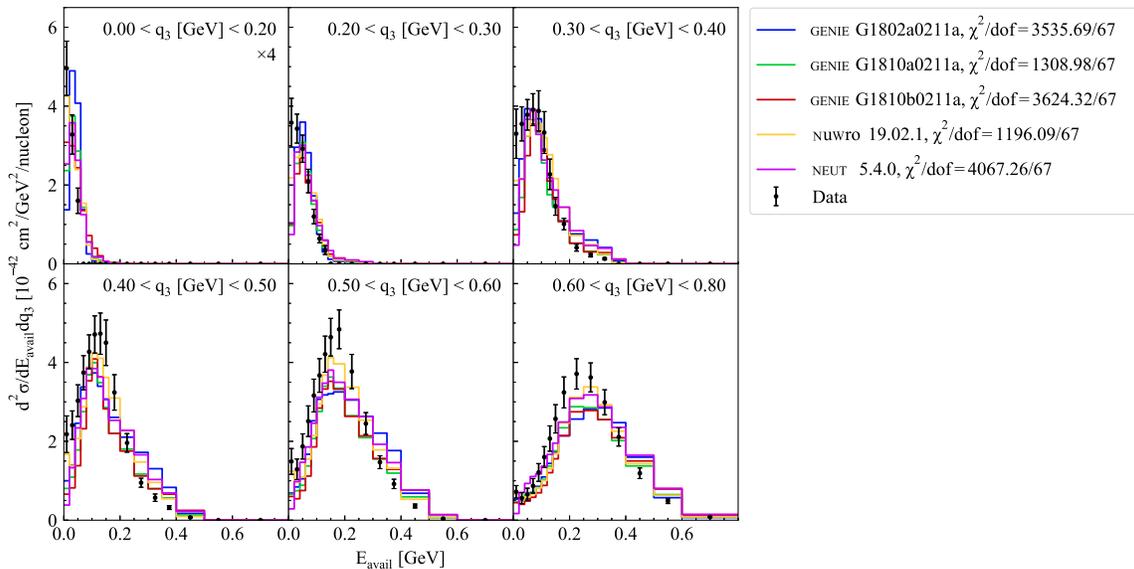
FIG. 6. MINERvA double-differential cross-section $d^2\sigma/dE_{avail}dq_3$ in six regions of q_3 is compared to NuWro, and three versions of GENIE.

0.15 GeV is dominated by QE processes, the region at high available energy above 0.2 GeV is dominated by delta resonance events, and events in the intermediate region contains $2p2h$ contributions. Discrepancies between data and the different generator predictions are visible. None of the generators correctly predict the first bin of available energy for the momentum transfer region 0 GeV/c < q_3 < 0.4 GeV/c—dominated by QE events. In the region between the QE and RES processes, where $2p2h$ events are expected, all generators underestimate the data. The best prediction is from the NuWro event generator with a total $\chi^2/\mathrm{dof} = 1196$ and the GENIE $G18\_10a$ with a $\chi^2/\mathrm{dof} = 1308$, both simulations contain the same nuclear model (LFG), the same $2p2h$ model from Valencia, and different long-range correlations (RPA) models. However, the χ^2/dof are unusually large values, most likely due to the strong correlations in the uncertainties among the data points. Overall, none of the event generators predict the inclusive data well; data are underpredicted in different regions of QE, MEC, and RES. The main sources of disagreement is for low values of available energy and the middle region where the MEC and RES events are located. Because the MINERvA data also use a hadronic quantity as the independent variable, final-state interaction model effects are present in these distributions. The lowest E_{avail} bins pick up a contribution from QE events when a p → n process leads to only neutrons in the final state, and similarly following pion absorption. In the antineutrino version of the analysis, the opposite happens for QE, but the same happens for the component that started with pion before FSI [119]. The follow-up measurement from MINERvA [120] quantifies this effect and it is accounted for in the uncertainty estimates.

In the publication [117], the data were compared with GENIE2.8.4 with reduced pion production. In that publication a discrepancy was reported, specifically in the region between QE and delta. In addition, the data were compared to a simulation that contained RPA [25,121] and $2p2h$ contributions from the Valencia group. The simulation with the addition of RPA showed good agreement at the lowest E_{avail} bins. The new version of GENIE has different predictions at the lowest E_{avail} bins due to a different treatment of the binding energy of the protons, new nuclear model, and the adoption of Valencia QE and $2p2h$; see more details in Sec. II. The new versions of GENIE are consistent with predictions of other generators at the lowest E_{avail} bins. However, all event generator underestimate the data in the lowest E_{avail} bins.

G. Comparisons of efficiency generator predictions

Figures 7 through 9 show the event selection efficiency as a function of the muon angle (T2K and MicroBooNE), muon momentum (T2K and MicroBooNE), hadronic energy (MicroBooNE and MINERvA), and three-momentum transfer (MINERvA). For MicroBooNE and
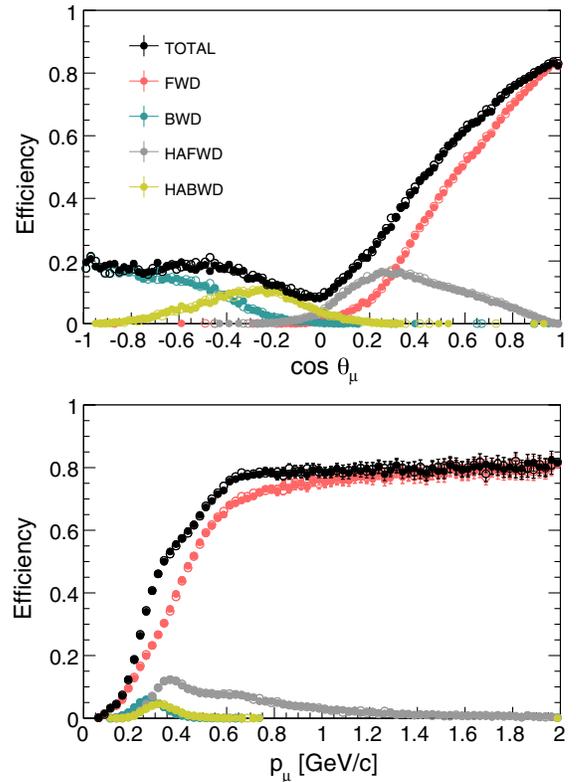
FIG. 7. T2K efficiency for the GENIE (empty dots) and NEUT (filled dots) generators as function of the muon momentum (top) and the angle (bottom) [112]. The different colors show the percentage of true signal events being reconstructed in the four different selection samples, split by muon direction (forward, backward, high-angle forward, and high-angle backward). While the efficiency seems to be identical for both generators in these projections, the 2D efficiency map presented in the original paper actually shows some significant differences in the low-momentum, forward-going bins.

MINERvA, the plots also show the cross section predicted by the generators studied in this paper as a function of the same variables.

The T2K measurement is sensitive to muon momenta above ∼200 MeV/c, with a flat efficiency above 600 MeV/c. This is shown in Fig. 7. The efficiency for backscattered muons is lower than for forward-going muons, mostly due to those muons also being of lower momentum. Muons produced at high angles are reconstructed without the help of the TPCs and can travel along the length of the FGD scintillator bars, making them harder to reconstruct. It is important to note that the 2D efficiency shows model dependence in some of the measurement bins. In low-momentum, forward-going bins the efficiencies evaluated with the NEUT and GENIE event generators differ significantly. This is caused by event properties that are not part of the analysis binning, but which influence the event reconstruction. In this case, it seems to be caused by the different handling of DIS events by NEUT and GENIE. The
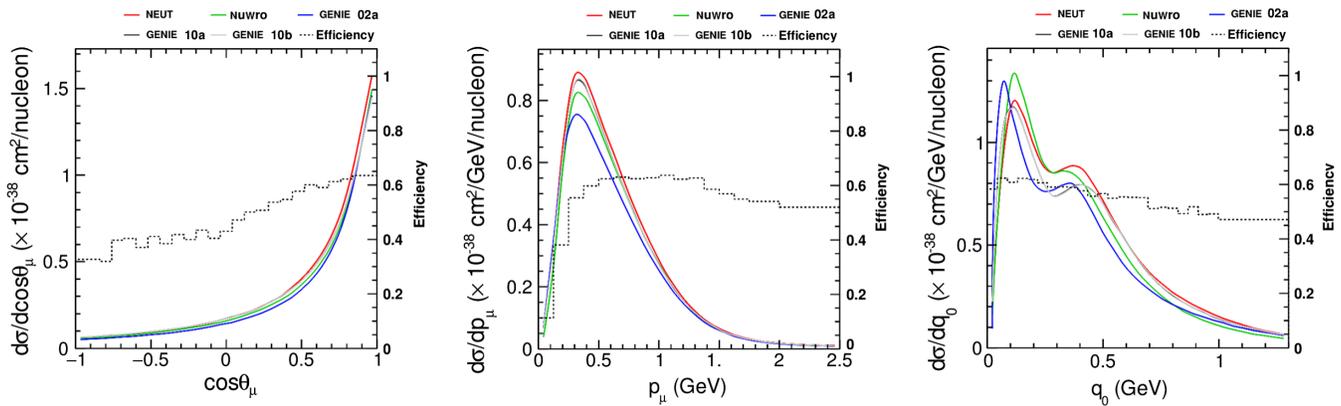
FIG. 8.   MicroBooNE efficiency (dotted line) and cross-section predictions (solid line) for the different generators as function of the muon angle (left), muon momentum (middle), and the hadronic energy q_0 (right).

difference is not visible in the 1D projections of the efficiencies as shown in Fig. 7. It is very important to check the efficiencies not only for dependence on the single measured variables, but also on their multidimensional combinations, as well as any other implicit assumptions about nuisance event property distributions.

In MicroBooNE, the efficiency in muon-momentum increases from 0 to 0.5 GeV/c due to the effect of detector and reconstruction thresholds; it is more constant above 0.5 GeV/c. Figure 8 shows that the biggest tension between different cross-section models happens just before 0.5 GeV, right at the place where the efficiency changes rapidly. On the other hand, the efficiency in q_0 is quite constant with a slight decrease moving towards higher q_0 values. At higher q_0 the events become more complicated to reconstruct and hence more difficult to select. The different models shown in the figure present a quite different behavior as a function of q_0. Since the efficiency depends strongly on other event properties (like the muon momentum), it must be assumed that the overall efficiency difference between the considered models is stronger than the flat efficiency in q_0 suggests.

In MINERvA, the efficiency as a function of the muon angle, muon momentum, three-momentum transfer, and the hadronic energy, q_0, are shown in Fig. 9. The efficiency as a function of angle only covers $\theta_\mu < 20°$, the muon tracks which exit the downstream end of MINERvA are matched to tracks in the MINOS near detector, and the efficiency as a function of muon momentum ranges from 1.5 GeV to higher momentum values up to 6 GeV compared with MicroBooNE and T2K, which covers momentum up to 2.5 GeV. The efficiency at low three-momentum transfer is 70% and decreases for values of $q_3 > 0.3$ GeV to 50%, where the models show different predictions. Event generators show different predictions in normalization, for example at $q_3 = 1$ GeV NEUT and GENIE v3 G18_02a event generators differ by 20%. In the region $0.4 < q_3 < 1$ GeV, the event generators have different predictions in shape and

normalization. The efficiency as a function of q_0 is 60% for values $q_0 > 0.2$ GeV and efficiency is higher at low $q_0 < 0.2$ GeV. Model predictions are different in shape and normalization for all values of the hadronic energy.

H. Discussion

The q_0 efficiency for MicroBooNE and MINERvA, shown in Figs. 8 and 9, is similar for the two experiments, about 80% efficiency which slightly decreases with increasing q_0. Having a constant efficiency for the MINERvA measurement is extremely important giving the fast-changing cross sections and also the differences among the models shown in the same figures.

The MicroBooNE cross section, shown in Fig. 5, shows a poor agreement with the different predictions, with tension especially visible in the forward-going bins, where the MC shows some deficits compared to the data in some momentum bins. From the χ^2, calculated with the full covariance matrix, NuWro appears to give the best prediction, though the other generators are comparable.

The comparisons of MINERvA's inclusive cross section with different event generator predictions in Fig. 6 show disagreement in different regions of q_3 and available energy. A deficit at low available energy is observed from all generators and strength is missing for the regions with MEC and RES, around 0.4 GeV $< q_3 < 0.8$ GeV.

Comparisons of T2K and MicroBooNE with MINERvA cannot be directly made. The inclusive charged-current measurements were done as a function of different variables and with different neutrino energy spectra. A common outcome from the comparisons for MicroBooNE and MINERvA inclusive charged-current measurements is that NuWro provides better predictions for some but not all kinematics, although no generator is able to successfully describe the MINERvA data at low available energy. Conversely, for the T2K measurement NEUT has better χ^2/dof compared with NuWro. This apparent tension is hard to resolve. The GENIE G18_10a and G18_10b predictions
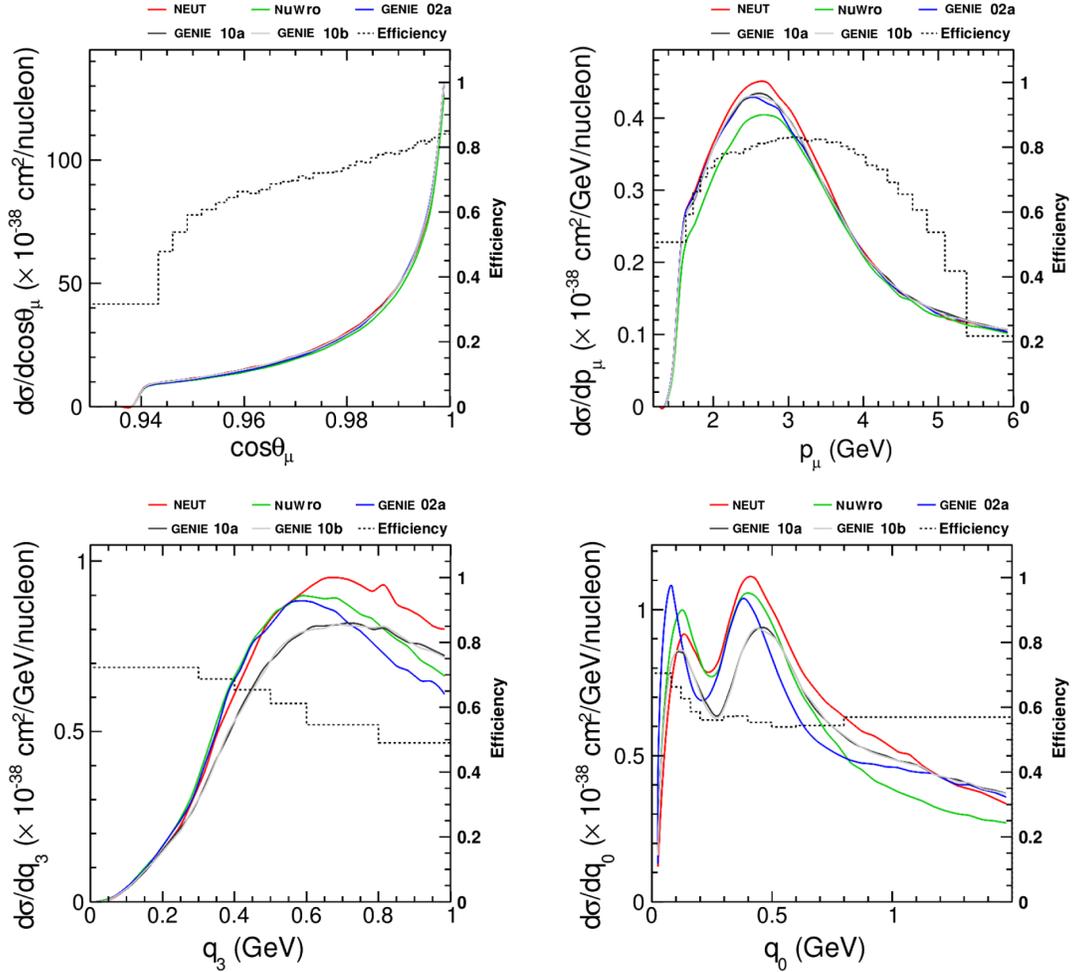
FIG. 9.   MINERvA efficiency (dotted line) and cross-section predictions (solid line) for the different generators as function of muon angle (left top), muon momentum (right top), three-momentum transfer (left bottom), and the hadronic energy q_0 (right bottom).

have χ^2 values between NEUT and NuWro for MicroBooNE, but have the best χ^2 values for T2K. It is interesting that for MINERvA G18_10b has χ^2 twice as large as G18_10a even though the only difference is choice of FSI model. This shows an interesting sensitivity to FSI in these data. Although GENIE G18_02a has older models than G18_10a and G18_10b, the χ^2 values are not significantly worse overall.

VII. CC-MESONLESS INTERACTIONS

A. Introduction

Quasielastic-like, also referred to as CC-0π, refers to a topological classification of neutrino-nucleus interactions where the resulting particles exiting the nucleus contain only nucleons and no mesons. This interaction is a critical process providing a dominant channel for neutrino oscillation experiments operating in the few-GeV region [1–3,6,117]. Appropriate estimators of neutrino energy and the ability to simulate these kinematics is of the utmost importance to this experimental program.

The CC-0π topology is mainly composed of CCQE events, where indeed only one nucleon is expected to exit the interaction vertex. Recent measurements have shown the importance of the $2p2h$ process, although the details remain uncertain. Other effects due to nucleon-nucleon correlations in the nucleus, like RPA, are now included in simulations, but also remain uncertain. Finally, FSI can impact the CC-0π channel, by altering the nucleon final-state kinematics or by reabsorbing final-state pions before they exit the nucleus.

In this section we will review several CC mesonless measurements from T2K and MINERvA. We will start in Sec. VII B and Sec. VII C with a comparison between the cross-section measurements in muon kinematics from both experiments and the predictions from generators described in this paper, using samples which include protons below tracking thresholds. With the same samples, in Sec. VII D and Sec. VII E we will focus on a more direct comparison between T2K and MINERvA results, trying to select a region of the phase space common to both experiments.

A comparison of the $q_0 - q_3$ phase space for the two experiments is also discussed.

In order to focus more on the ability of current generators to describe nuclear effects, in Sec. VII F we will review measurements of a number of variables, known as Transverse Kinematic Imbalance variables, which have specific sensitivity to a variety of nuclear effects.

Measurement of QE-like interactions in Q_{QE}^2, using muon kinematics, from MINERvA is also compared to MC predictions in Sec. VII G, while in Sec. VII H generator predictions for different targets (CH, Fe, and Pb) will be compared to MINERvA results in Q_{QE}^2 variable, using proton kinematics. We will finally discuss the presented comparisons in Sec. VII I.

B. T2K results and comparisons with generators

The T2K collaboration has published two CC-0π measurements combining different targets and flavors, and using the data taken at the off-axis near detector (ND280): the first simultaneous extraction of the ν_μ and $\bar{\nu}_\mu$ CC-0π cross sections on hydrocarbon employing the data taken with a neutrino and antineutrino beam [91] and the first simultaneous extraction of the ν_μ CC-0π cross sections over oxygen and carbon [122]. The cross sections have been extracted as function of the muon momentum and $\cos\theta$ without any phase-space restriction.

A simultaneous measurement has many advantages. The knowledge of the correlation between the measured cross sections allows further information (cross-section ratio, asymmetry, sum, difference) to be obtained through a proper combination, often reducing common systematics uncertainties. In addition, a joint measurement further allows a less model-dependent background subtraction, as is the case for the combined neutrino-antineutrino CC-0π cross section, where the neutrino background in antineutrino beam is relatively large. By fitting ν and $\bar{\nu}$ samples at the same time, it is possible to simultaneously extract both cross sections, thus disentangling the neutrino and antineutrino contributions, without needing a bare background subtraction. A similar approach is exploited for the combined carbon-oxygen cross-section measurement. In this case, the two cross sections are extracted using the interactions occurring in the ND280 FGDs. The first FGD (FGD1) is completely made of plastic scintillator bars, while in the second one (FGD2) the scintillator bars are interleaved with water targets [1]. The sample of CC-0π interactions on carbon is a background for the oxygen measurement, since the water modules are passive and all the interactions are reconstructed in the scintillator layers. Based on the starting position of the muon track, it is possible to construct carbon- and oxygen-enhanced samples to be used in a simultaneous fit, thus allowing the oxygen and carbon components to be extracted at the same time.

In the following, we will only consider the ν_μ CC-0π cross sections on carbon and hydrocarbon, in order to allow a more direct comparison with MINERvA ν_μ CC-0π cross section described in Sec. VII C, which uses the same target material.

The two T2K measurements use the same event selection for the ν_μ CC-0π interactions; two control samples are also employed to constrain the background in the signal sample, mainly made up of interactions from CC resonant pion production and CC deep inelastic scattering. Events are selected exploiting the particle identification capabilities of the FGDs and the TPCs, and the timing between sub-detectors to distinguish between forward- and backward-going (with respect to beam direction) tracks. Five different signal samples with at least a negatively charged muon entering in TPC or fully contained in FGD, and, eventually, one or more protons, have been selected. For the control samples, events with one negatively charged muon and one positively charged pion (CC-$1\pi^+$) or more than one (CC-Other) entering in the TPC are selected.

The two measurements also share the same cross-section extraction method. An extended binned likelihood fit is used to extract the true number of CC-0π events in bins of muon kinematics (momentum and $\cos\theta$) that are subsequently corrected by the signal efficiency, the integrated flux, the number of targets, and the bin width. Uncertainties are taken into account by adding a penalty term to the likelihood and are then propagated when estimating the cross section.

The differences between the two measurements are related to

(i) the combined cross section: in one case, the complementary measurement is the $\bar{\nu}_\mu$ CC-0π on CH and in the other case is the ν_μ CC-0π on O;

(ii) the target: in one case the detector target is the FGD1 and the cross section is extracted per hydrocarbon nucleons, while in the second case both FGD1 and FGD2 are used and the cross section is extracted per carbon nucleon[1];

(iii) the binning: for the measurement on CH the $p_\mu - \cos\theta_\mu$ binning is finer than for the measurement on C. However the latter better matches the MINERvA phase space.

Due to the number of common points between the two measurements, results on carbon or hydrocarbon should in principle give similar information. However, since they are partial results of more complex and different analyses, we decided to report both.

---

[1]It should be underlined that when quoting the CH cross section, the full detector mass of the FGD1 is considered and it thus includes also small percentage of non-CH elements (as detailed in Sec. V B). On the other hand, when quoting the cross section on carbon, the contribution from all the other elements is removed.
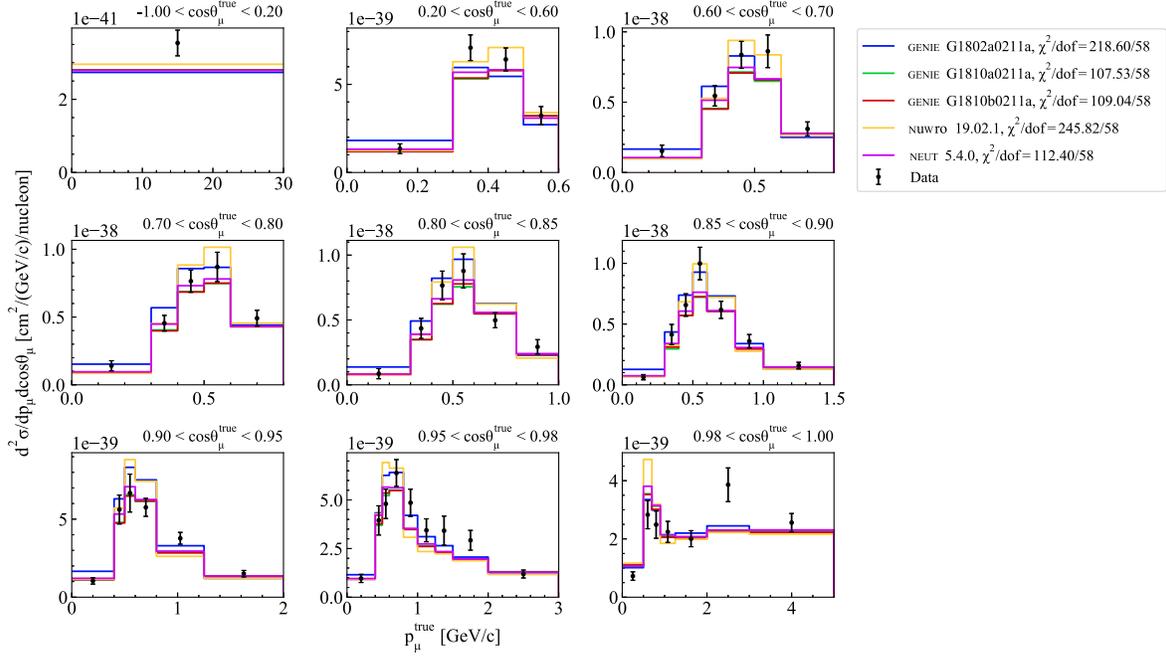
FIG. 10.   Measured T2K ν_μ CC-0π double-differential cross sections on hydrocarbon in bins of true muon kinematics. The results are compared to GENIE v3 G18_02a (blue), G18_10a (green), and G18_10b (red), NuWro 19.02.1 (orange), and NEUT 5.4.0 (violet). The last bin in momentum is not displayed for readability.

Figures 10 and 11 show the measured ν_μ CC-0π double-differential cross sections on hydrocarbon and carbon, respectively, in bins of true muon kinematics, compared with the MC predictions. Although the two measurements share several data samples and the cross-section extraction method, some differences can be noticed when compared to generator predictions.

In the legends of Figs. 10 and 11, we report the χ^2 obtained using a reduced covariance matrix for the neutrino-only and the carbon-only part of the measurements. Although the two results show similar preferences, it

should be noticed that χ^2 values are in general smaller for the CC-0π measurement on carbon than on hydrocarbon. This can be partially explained by the fact that the two measurements use different binning, different statistics, and are subset of two more complex measurements, one including antineutrino and the other one including oxygen. Looking at the χ^2, carbon data seem to clearly prefer NEUT, GENIE G18_10a and G18_10b. We should notice that the only difference between GENIE G18_10a and G18_10b is related to the pion FSI model, and CC0π measurements in muon kinematics are not very sensitive to these model
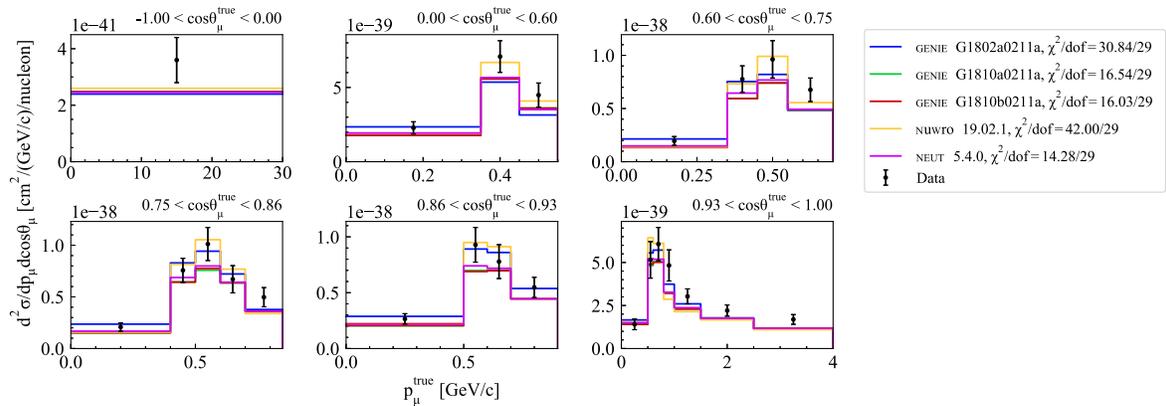
FIG. 11.   Unregularized T2K ν_μ CC-0π double-differential cross sections on carbon in bins of true muon kinematics. The results are compared to GENIE v3 G18_02a (blue), G18_10a (green), and G18_10b (red), NuWro 19.02.1 (orange), and NEUT 5.4.0 (violet). The last bin in momentum is not displayed for readability.
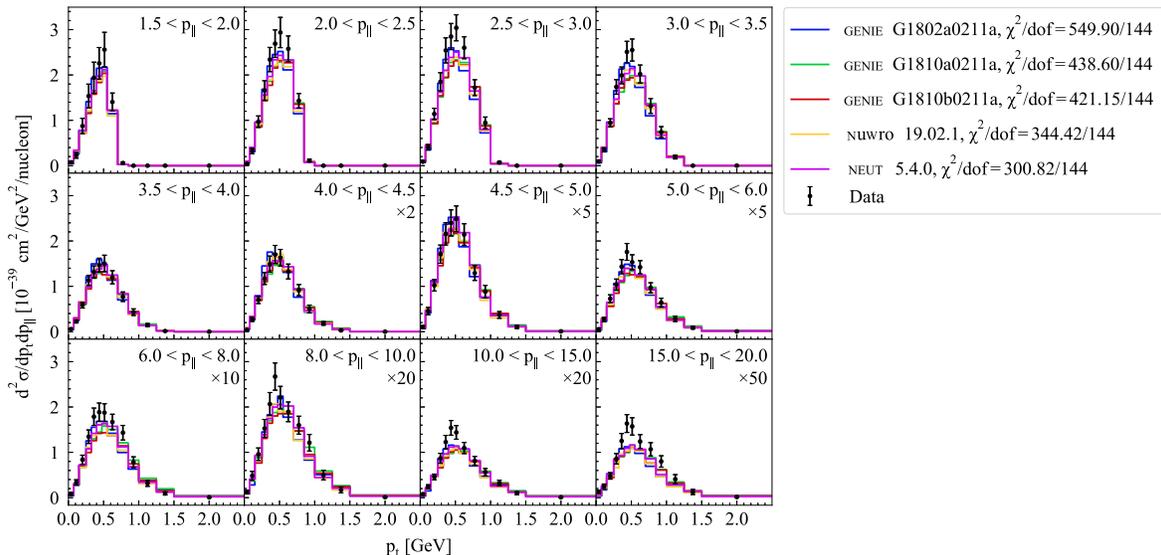
FIG. 12.   MINERvA CCQE-like cross-section measurement on hydrocarbon compared to various models in regions of $p_\parallel$ [GeV].

differences. On the other hand, the measurement on hydrocarbon is clearly overestimated in the most forward bin for momenta below 1 GeV. This can be due to incorrect $1p1h$ predictions in the region of small energy transfer to the nucleus, where the treatment of various nuclear effects, like RPA, is not well understood.

C. MINERvA result and comparison with generators

The MINERvA collaboration produced a CCQE-like double-differential cross-section, shown in Fig. 12, using a beam of primarily muon neutrinos at a mean energy of 3.5 GeV. The measurement uses as observables the transverse (p_T) and longitudinal ($p_\parallel$) muon momentum [123]. This variable combination was chosen because at high neutrino energy, as is the case for MINERvA, p_T^2 is correlated to Q_{QE}^2 and $p_\parallel$ is correlated to the neutrino energy.

A selection of CCQE-like events employs a combination of selection criteria using particle identification to remove different subsamples of backgrounds. A Michel electron tagging algorithm is used to identify late in time electrons from pion decay near the interaction vertex, and all track end points. An isolated cluster algorithm is used to count the multiplicity of showers in the interaction. In addition, a 500-MeV restriction on the visible energy not associated with tracked particles is imposed to remove the DIS and neutral resonant pion production with large pion energies.

The selected sample requires no Michel electrons and no more than one isolated cluster. Three control samples are populated using a combination of the interactions failing these cuts. A charged pion-dominated sample is constructed by requiring a single Michel electron and no more than one isolated cluster. A multipion-dominated sample is constructed by requiring more than one Michel electron

and more than two isolated clusters. The third control region is a mix of single charged, neutral, or multipion and is constructed by selecting interactions with no Michel electrons and two isolated showers.

The cross section is extracted using a data-constrained background-subtracted sample which is unfolded, using four iterations, using D'Agostini unfolding via the ROOUNFOLD package [124]. The sample is efficiency corrected using the GENIE prediction. The result is then corrected for the number of nucleons in the FV and the integral of the NuMI flux between 0 and 100 GeV.

Each prediction's χ^2 values using the full covariance matrix are reported in the legend of Fig. 12. We note that none of the generators is able to well reproduce the data because the χ^2 values are all larger than 300 for 144 bins. The smallest χ^2 values are shown for NEUT and NuWro, while GENIE G18_02a shows the largest disagreement with data. The region where models struggle the most in reproducing data is in the highest $p_\parallel$ region. Unlike what was shown in Sec. VII B, NuWro and NEUT predictions are very similar at higher beam energies. In general, predictions from generators using LFG as nuclear models are similar and show differences, as expected, from G18_02a.

D. T2K/MINERvA phase-space comparison

In this section we compare true experimental efficiencies with projected calculations of 2D cross sections calculated by the generators (see Sec. IV for details), specifically for the T2K and MINERvA measurements described in Sec. VII B and in Sec. VII C, respectively. We remind here that, although both T2K and MINERvA performed 2D measurements and thus provided two-dimensional efficiency maps, for display purposes 2D efficiencies were projected in each dimension and compared with generator predictions.
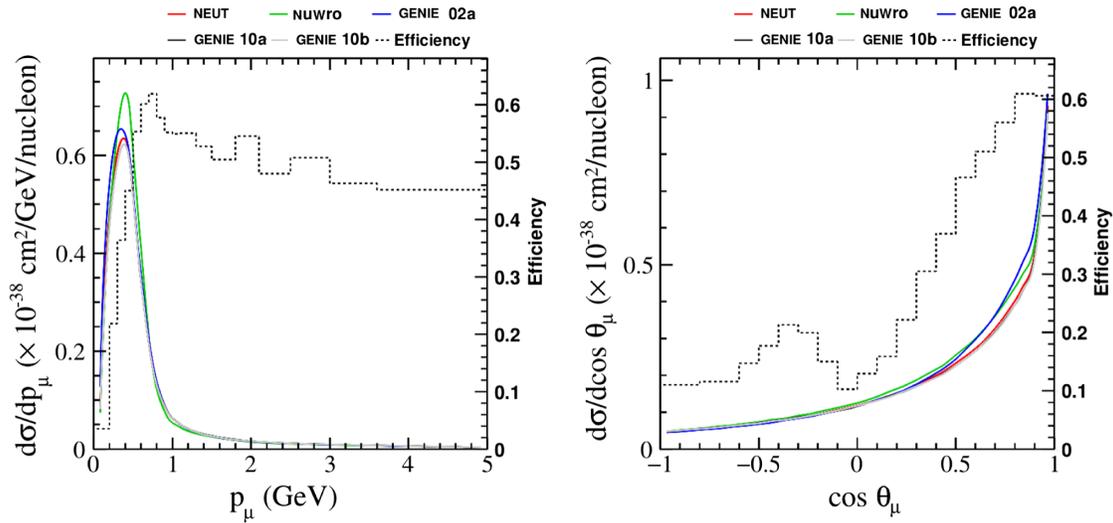
FIG. 13.   Efficiency for the T2K CC0π measurement (dotted line using the right vertical scale) and cross-section predictions (solid line using the left vertical scale) for the different generators employed in this paper as function of true muon momentum (left) and cosine of the muon scattering angle (right).

The efficiencies of the T2K CC-0π selected sample as a function of true muon momentum and cosine of the scattering angle are shown in Fig. 13 (dotted line). On the same plot, the cross sections (solid lines) predicted by the different generators studied in this paper are compared. The dip for $\cos\theta_\mu = 0$ is a result of the intrinsic inefficiency of the detector to track particles perpendicular to the beam direction. One goal is to examine model dependence in situations where the efficiency is rapidly changing. The efficiency is rapidly rising for muon momentum values up to about 0.5 GeV. Although the cross-section predictions

are also changing rapidly in that region, no significant disagreements among models are seen, excluding NuWro, which predicts a higher cross section.

The efficiencies of the MINERvA CCQE-like sample as a function of transverse and longitudinal momentum are shown in Fig. 14, as a dotted line. The cross-section predictions from the generators used in this paper are also shown. Also in this case, all the generators show a similar behavior, although we can notice that G18_02a predicts a slightly higher cross section in p_T. The projected efficiency ranges from 40% at large p_T to 70% at small p_T and ranges
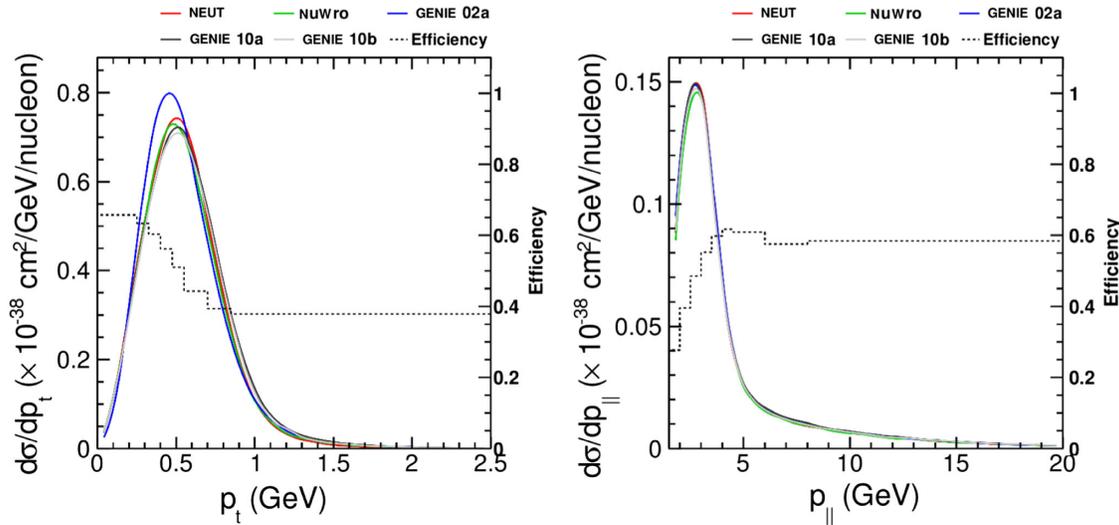
FIG. 14.   Projected efficiency in p_T (left) and $p_\parallel$ (right) for the MINERvA CCQE-like measurement. The dotted line is the efficiency using the right vertical scale while the colored lines are cross-section predictions using the left vertical scale for the generators used in this paper.
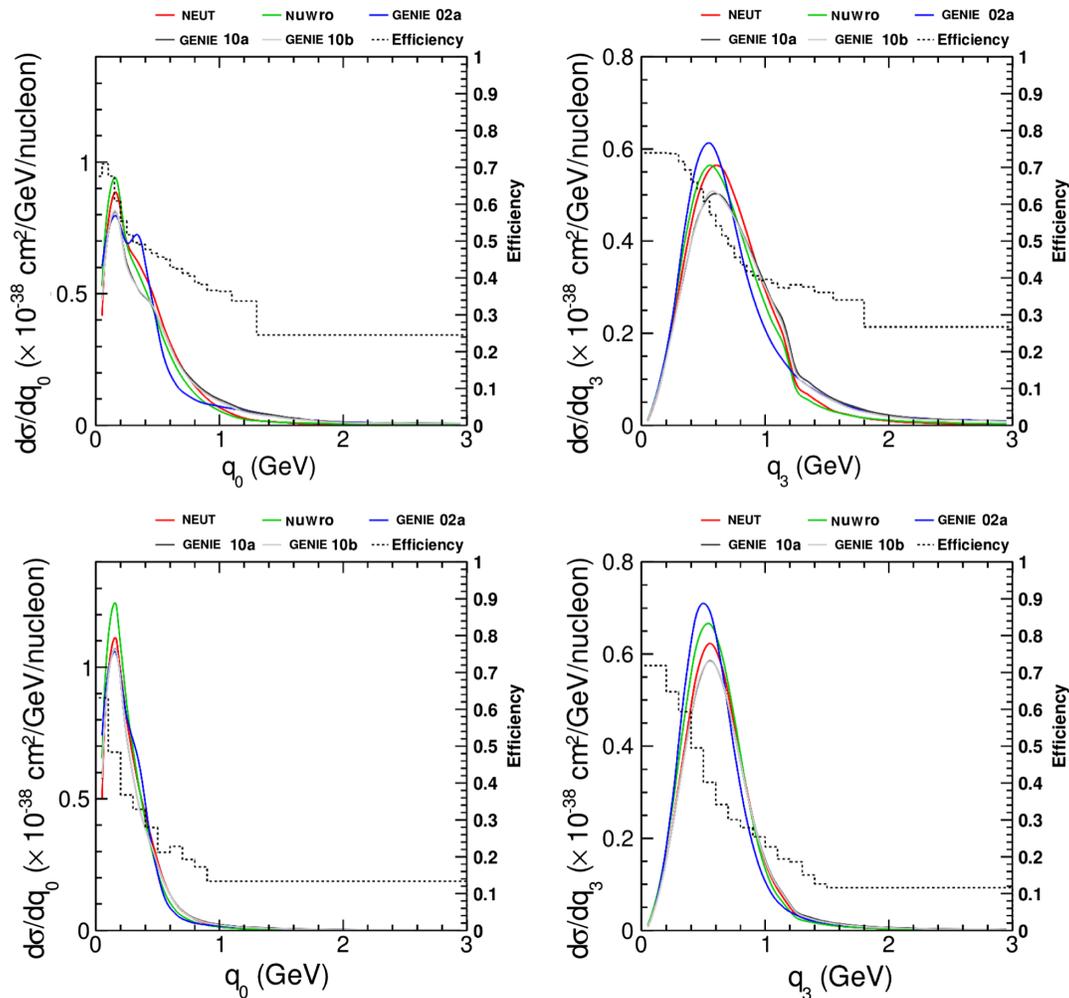
FIG. 15.   Comparison of the projected efficiencies and predicted cross sections for q_0 (left column) and q_3 (right column) for MINERvA (top) and T2K (bottom).

from 30% at small $p_{||}$ to a plateau of about 60% for $p_{||}$ greater than 4 GeV. The decrease in efficiency at low $p_{||}$ is due to the requirement that the muon is reconstructed in MINOS, which puts a threshold of about 2 GeV depending on the vertex location in the scintillator tracker. The efficiency dropoff at large p_T is due to the same MINERvA -MINOS track requirement.

In addition to the efficiencies described above, in order to facilitate a direct comparison between T2K and MINERvA results, the $q_0 - q_3$ projected phase spaces and relative efficiencies are shown in Fig. 15, using the method described in Sec. IV. Due to the higher beam energy, the accessible phase space in both q_0 and q_3 is wider in MINERvA than in T2K. For these plots, the same kinematic ranges were used. MINERvA and T2K have efficiencies which are large at low values of q_0 and q_3, about 70% for MINERvA and 65–70% for T2K, which decrease with increasing q_0 and q_3. The lowest efficiency MINERvA has is about 35%, while for T2K it is about 15%.

E. Comparisons between T2K and MINERvA data sets

In the upper panel of Fig. 16 the MINERvA and T2K ν_μ CC-0π neutrino cross-section measurements are compared with the MC samples studied in this paper and described in Secs. II A to II C. Even if MINERvA reports the cross section as function of $p_{||}$ and p_T, given the restricted phase space in which the cross section has been measured, such variables can be associated with p_μ and $\cos\theta_\mu$, respectively. Although MINERvA has a higher beam energy and therefore larger range of kinematics among the final-state particles, the muon must be at less than 20°. The phase space of the T2K measurements is restricted for the purposes of this comparison to match MINERvA phase space ($p_\mu > 1.5$ GeV and $\theta_\mu < 20°$). For T2K both measurements over hydrocarbon and carbon are included, and two different sets of phase-space cuts, are compared:

(i) In one case, the full muon momentum phase space is exploited, while the cosine of the muon-scattering angle is required to be greater than 0.94 for the cross
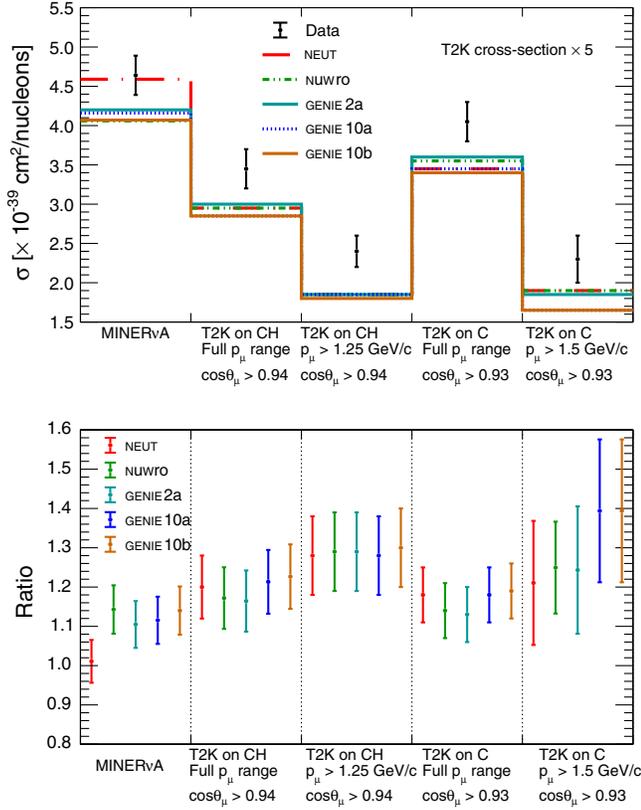
FIG. 16. Top figure: MINERvA and T2K CC0π cross-section measurements compared with the MC employed in this paper. The phase space of the T2K measurements is restricted to match MINERvA and the obtained values of the cross section are multiplied by a factor of 5 for display purposes. Bottom figure: Ratio between MINERvA and T2K CC0π cross-section measurements and the MC.

section on hydrocarbon and 0.93 for the cross section on carbon; those cuts correspond to require θ_μ smaller than about 20°.

(ii) In the other case, the momentum is restricted to be greater than 1.25 GeV/c for hydrocarbon and 1.5 GeV/c for carbon, while the cut on the muon $\cos\theta$ remains the same.

With respect to the binning used for the T2K measurement on hydrocarbon, the binning used for the measurement on carbon allows to better match the MINERvA p_μ phase space. The values plotted in Fig. 16 represent the cross section per nucleon assuming pure CH or C targets, respectively.

In order to mitigate the energy dependence due to the different neutrino fluxes at which MINERvA and T2K detectors are exposed, and thus to allow a clearer comparison between the two experiments, the ratio between the measured cross sections and the MC predictions has been computed and is shown in the lower panel of Fig. 16. In the case of MINERvA, the ratio between the measured cross section and NEUT is consistent with one, while T2K

measures a cross section higher than what was predicted by NEUT. The other ratios are compatible across the different predictions and both experiments show that the MC underestimate the data. It is important to stress that for T2K this effect is especially true for high-momentum bins, while for low momentum the MC slightly overestimate the data, as shown in the last angular bins in Figs. 10 and 11 and evident when comparing the full and limited p_μ range ratios in Fig. 16. Indeed, when using the full momentum phase space for this forward region, the MC underestimation is less visible.

In summary, most generators seem to systematically underestimate the data with one exception: NEUT is in excellent agreement with the MINERvA data. Concerning the T2K restricted phase space, this trend is confirmed, even if the low statistics in this region results in larger error bars.

F. Transverse kinematic imbalance variables and comparisons with generators

As explained in Secs. VII B and VII C, both T2K and MINERvA include events with outgoing protons detected in their CC-0π selections. By using the subsamples where one or two protons are reconstructed, T2K and MINERvA have measured the CC-0πNp cross section as a function of transverse kinematic imbalance (TKI) variables, as reported in [125,126], respectively. In this section, published results have been compared with MC predictions studied in this work.

TKI variables quantify the imbalance between the outgoing lepton and proton kinematics in the plane transverse to the incoming neutrino, and are thus able to offer a probe of nuclear effects [126,127]. They are defined [128] as follows:

$$\delta p_T = |\delta \mathbf{p}_T| = |\mathbf{p}_T^\mu + \mathbf{p}_T^p|, \tag{4}$$

$$\delta\alpha_T = \arccos\left(-\frac{\mathbf{p}_T^\mu \cdot \delta\mathbf{p}_T}{p_T^\mu \delta p_T}\right), \tag{5}$$

$$\delta\phi_T = \arccos\left(-\frac{\mathbf{p}_T^\mu \cdot \mathbf{p}_T^p}{p_T^\mu p_T^p}\right) \tag{6}$$

where $\mathbf{p}_T^\mu$ and $\mathbf{p}_T^p$ are the momentum of the outgoing muon and the highest momentum proton, respectively, projected on the plane transverse to the incoming neutrino. In the case of an interaction on a free nucleon, δp_T and $\delta\phi_T$ are expected to be zero (while $\delta\alpha_T$ is undefined) and any difference from zero is an indication of nuclear effects. In particular, δp_T is most sensitive to the nuclear structure, specifically the momentum distribution of the struck nucleon, while $\delta\alpha_T$ is most sensitive to FSI.

Both T2K and MINERvA measure cross sections as a function of TKI variables over a restricted phase space

TABLE IX. Signal phase-space restrictions for the T2K and MINERvA results in TKI variables.

Analysis	p_p	$\cos\theta_p$	p_μ	$\cos\theta_\mu$
T2K	0.45–1.0 GeV	>0.4	>250 MeV	>−0.6
MINERvA	0.45–1.2 GeV	>0.342	1.5–10 GeV	>0.940

where their detectors are sensitive, as summarized in Table IX.

In the left panels of Fig. 17, T2K TKI variables are shown and compared with generator predictions. All calculations use the LFG nuclear model except G18_02a, which uses

RFG. Considering the cross section as a function of δp_T, models using LFG generally have better agreement with the data, while G18_02a has the worst agreement. It is surprising that NuWro has poor agreement with the data, although it also uses LFG as nuclear model. The high-momentum tail where $2p2h$ and FSI processes are more relevant, is well described by all models.

Concerning $\delta\alpha_T$, most of the generator predictions have similar shapes, and show a rise at high angles, as is expected in the presence of FSI effects. While G18_02a shows the smallest rise at high $\delta\alpha_T$, NuWro has an almost flat shape in this variable. Data have a slight preference for G18_02a. Finally, G18_02a and NuWro
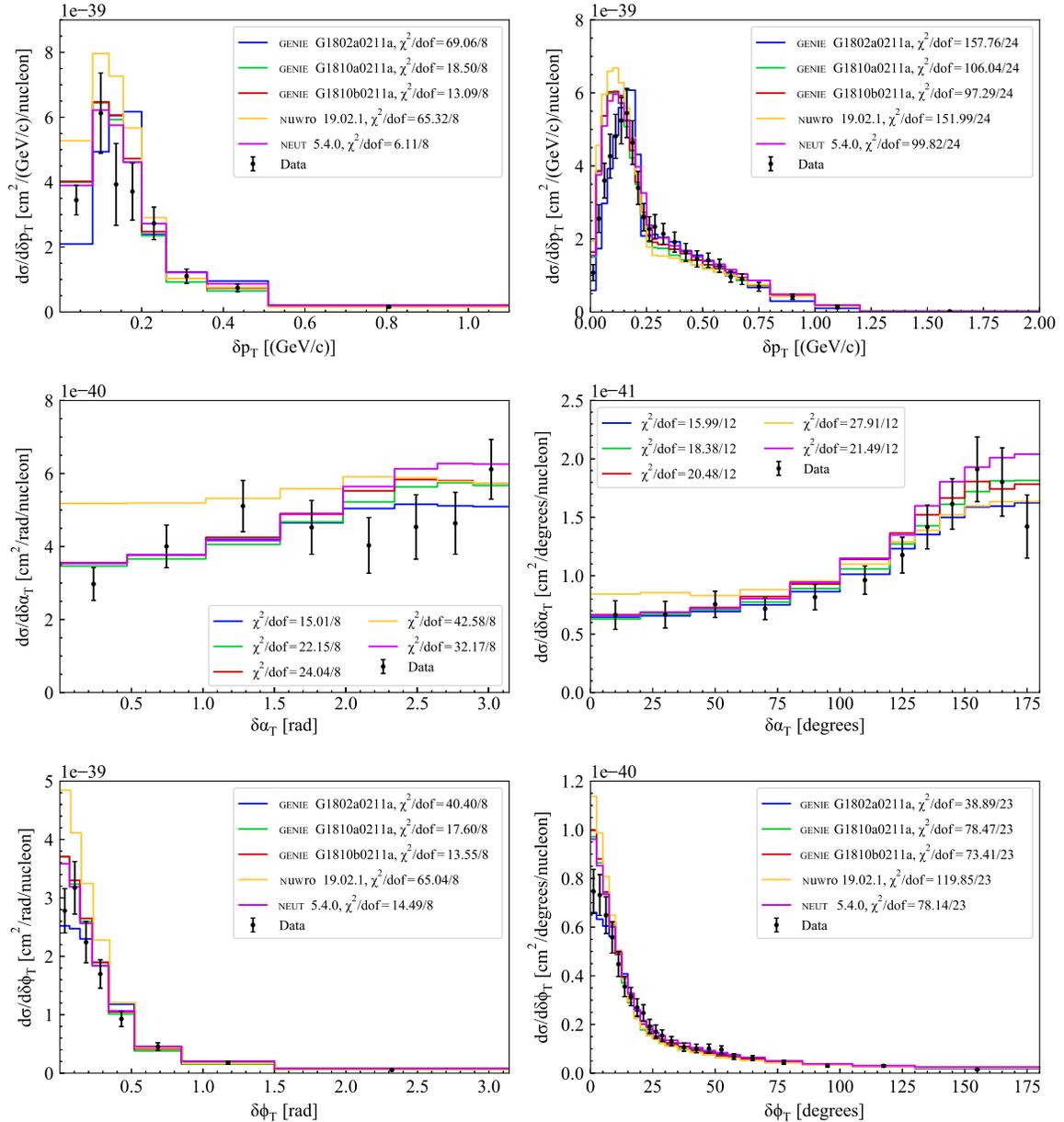
FIG. 17. T2K (left) and MINERvA (right) CC0π TKI cross-section measurements on hydrocarbon compared to various models. Generator names, suppressed in the middle row, follow the same colors as above and below.
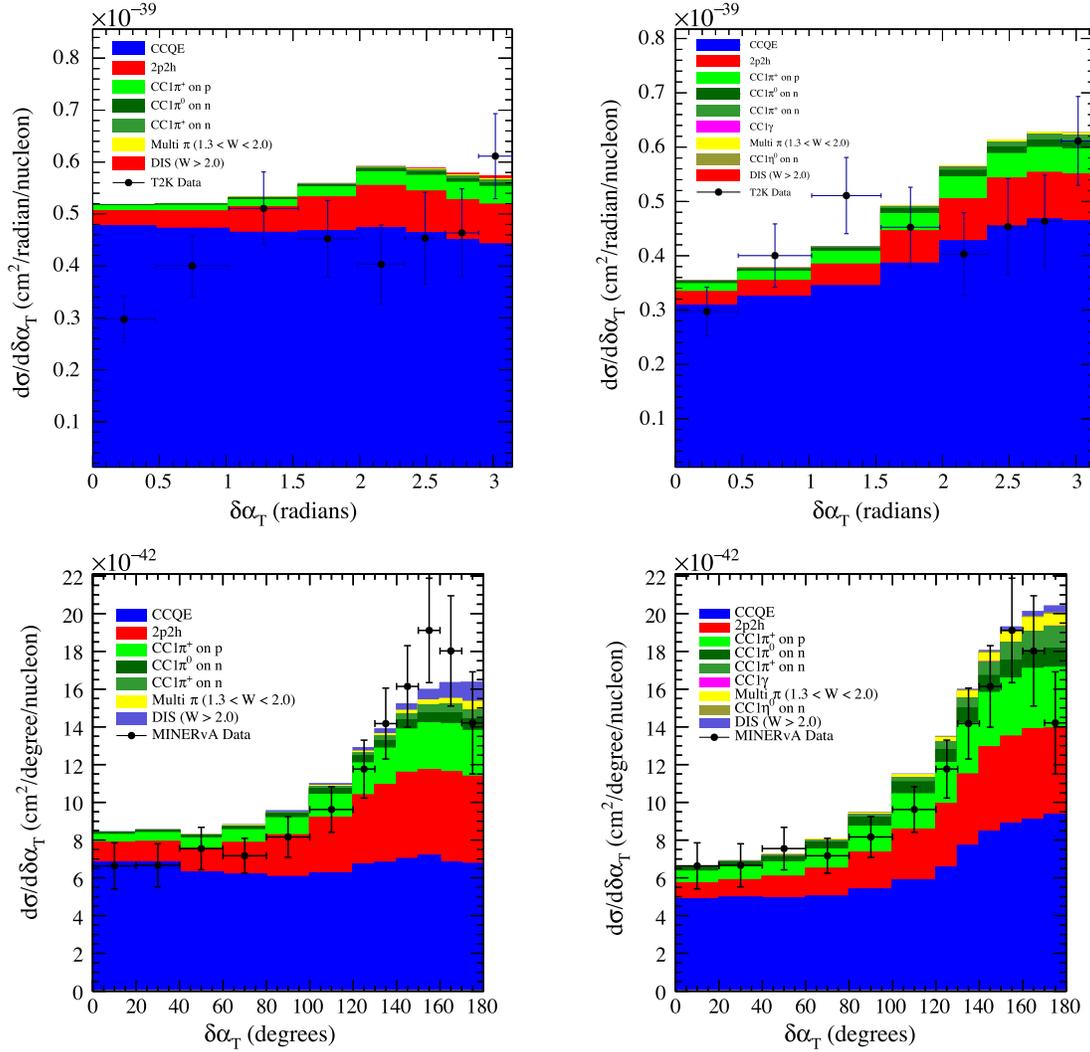
FIG. 18. T2K CC0π (top) and MINERvA (bottom) TKI cross-section measurements on hydrocarbon compared to NuWro (left) and NEUT (right) predictions, split by neutrino interactions.

show a different shape with respect to the other generators also in $\delta\phi_T$.

Although none of the generators correctly reproduces all the data in the three variables, the largest disagreement is shown by NuWro and G18_02a, which have χ^2/number of degrees of freedom (ndf) values between 5 and 8 and between 2 and 8, respectively. On the other hand, NEUT and GENIE with LFG show similar and lower χ^2/ndf values, between 0.8 and 3, depending on the variable; NEUT and both G18_10a and G18_10b are fairly good in reproducing δp_T and $\delta\phi_T$ distributions. Focusing on G18_10, we do not see a particular preference for either of the two FSI models used.

MINERvA TKI results are compared to the generator predictions in the right panels of Fig. 17. Comparisons indicate varying degrees of data-MC agreement. In particular, generator predictions for δp_T show different peak positions from one another, indicating a difference in the description of the nucleus; G18_02a shows the largest

disagreement with data (χ^2/ndf $\simeq 6.5$). In general, predictions from generators that use LFG as nuclear model show similar shapes in δp_T, although χ^2 values are smaller for NEUT and GENIE (χ^2/ndf $\simeq 4$) with respect to NuWro (χ^2/ndf $\simeq 6$). Overall, none of considered generators is able to well reproduce the δp_T distribution.

Concerning $\delta\alpha_T$, all the generators show similar predictions and thus have a similar fairly good agreement with the data (χ^2/ndf $\simeq 1.5$), with the only exception of NuWro that shows a disagreement (χ^2/ndf $\simeq 2.3$) and a flatter distribution as already observed for T2K. The variable $\delta\alpha_T$ in principle is sensitive to FSI effects. MINERvA data have a slight preference for the G18_10a.

Concerning $\delta\phi_T$, almost none of the generator shows agreement with MINERvA data, although G18_02a predictions shows the smaller χ^2 value (χ^2/ndf $\simeq 1.7$), while NuWro the largest (χ^2/ndf $\simeq 5$); χ^2 values for NEUT and G18_10a /b are similar (χ^2/ndf $\simeq 3$).

To better understand the difference between NuWro and others generators that use LFG in the prediction of $\delta\alpha_T$, we report in Fig. 18 the NuWro and NEUT TKI distributions split by neutrino interactions for T2K and MINERvA.

For T2K, by looking at first plots of Fig. 18, it is evident that the remarkable difference between the predicted $\delta\alpha_T$ shapes is due to CCQE interactions: NuWro predicts an almost flat shape for CCQE, while NEUT shows a dip in CCQE at low $\delta\alpha_T$. Contributions from $2p2h$ and CC1π are instead largely similar. We can also notice that the normalization difference visible in the three variables is still due to different predictions of CCQE, that has a higher cross section in NuWro than in NEUT.

For MINERvA (see Fig. 18), the differences between NuWro and NEUT in TKI variable prediction are less relevant, but are again due to the different contribution from CCQE interactions, that also in this case have a flat distribution in $\delta\alpha_T$ and a higher cross section.

In general, none of the generators investigated here are able to correctly reproduce the data in the three variables; in particular, both T2K and MINERvA data disfavor NuWro predictions in all the three TKI variables. Generators have difficulties in reproducing the MINERvA δp_T distribution, while the agreement is better with T2K data, especially for NEUT and GENIE LFG predictions. Conversely, generators seem to well reproduce MINERvA $\delta\alpha_T$ distribution, while they have more difficulties with the T2K $\delta\alpha_T$ distribution. The dominant difference between the NuWro prediction and others in the $\delta\alpha_T$ variable is due to differences in the pre-FSI spectrum of protons resulting from quasielastic interactions. Finally, concerning $\delta\phi_T$, T2K data seems to favor NEUT and GENIE LFG calculations, while this is not evident for MINERvA data.

G. Comparison of generators using lepton-derived four-momentum transfer

Although recent measurements tend to report cross sections as a function of variables related to detectable muon observables, in this and following sections we focus on CCQE-like measurements that have been performed as a function of the momentum transfer, Q^2.

An estimator of the four-momentum transfer can be derived for QE-like events under the assumption of a stationary target, as shown in Eqs. (7) and (8):

$$E_{\nu,\mathrm{QE}} = \frac{M_p^2 - (M_n - E_b)^2 - M_\mu^2 + 2(M_n - E_b)E_\mu}{2(M_n - E_b - E_\mu + P_\mu \cos\theta_\mu)}, \quad (7)$$

$$Q_{\mathrm{QE}}^2 = 2E_{\nu,\mathrm{QE}}(E_\mu - P_\mu \cos\theta_\mu) - M_\mu^2, \quad (8)$$

where M_p and M_n are the masses of proton and nucleon, respectively, E_b is the binding energy, E_μ is the energy of the outgoing muon, and P_μ and $\cos(\theta_\mu)$ are the muon momentum and direction. The MINERvA experiment measures QE-like interactions in the scintillator tracker
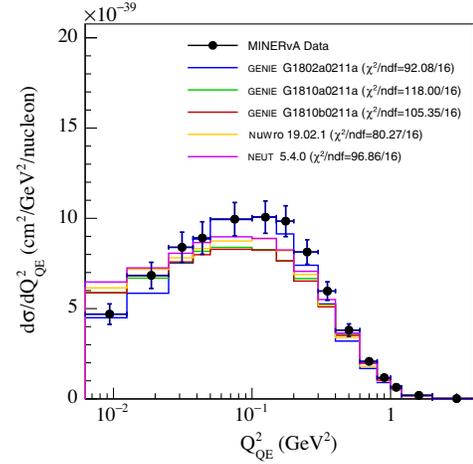
FIG. 19.   MINERvA CCQE-like Q_{QE}^2 cross-section measurement on hydrocarbon compared to various models.

region between 0 and 4 GeV^2/c^2 using an on-axis beam with an average energy of 3.5 GeV [123]. Figure 19 shows a comparison of this result with a set of predictions from NuWro, NEUT, and GENIE.

The NuWro prediction shows the lower χ^2, while G18_02a and G18_10b show the highest disagreement with data. However, χ^2/ndf values are between 5 and 7, thus indicating that none of the generators properly describes the data.

H. Comparisons of generators with data from different targets (CH, Fe, and Pb)

The MINERvA experiment measured QE-like interactions on different targets using the same neutrino beam [129]. Although this result was not published recently, it has a unique standing as it spans various nuclei. The signal is defined as an event with one muon, no pions, and at least one proton with momentum greater than 450 MeV/c exiting the nucleus. The measurements have been performed using the four-momentum transfer (Q_p^2) reconstructed using the proton kinetic energy (T_p). Under the assumption of CCQE scattering from a neutrino at rest, the Q_p^2 value is reconstructed using

$$Q_p^2 = (M_n - \epsilon_B)^2 - M_p^2 + 2(M_n - \epsilon_B)(T_p + M_p - M_n + \epsilon_B),$$
$$(9)$$

where $M_{n,p}$ is the appropriate nucleon mass, and ϵ_B is the effective binding energy of 34 MeV/c^2.

Figure 20 shows comparisons of data in different materials (CH, Fe, and Pb) with the NuWro, NEUT, and GENIE predictions.

Since proton kinematics in the final state are very sensitive to final-state interactions, these measurements test the atomic mass (A) dependence of models. For the CH
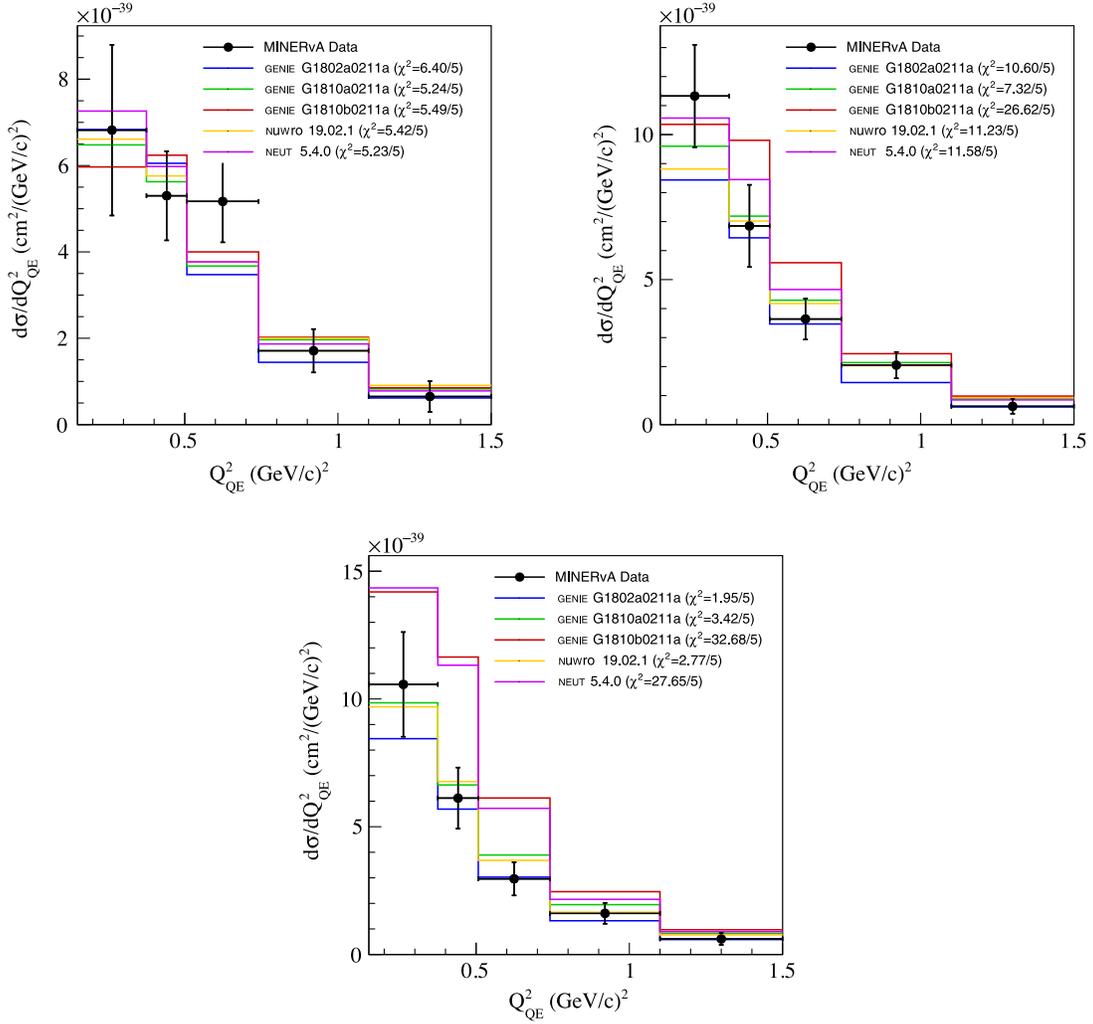
FIG. 20. Differential cross sections as a function of Q_p^2 for CH (top left), Fe (top right), and Pb (bottom) compared with NuWro, NEUT, and three versions of GENIE.

target measurement, all generators show similarly good agreement with data and have a χ^2/ndf of about 1.0, although no tuning was done for any of the models. The heavier targets provide a more interesting test of models, since we expect larger effects from FSI. The spread of distributions predicted by generators and the spread of corresponding χ^2 both increase significantly for these nuclei. For Fe the best prediction comes from G18_10a and for Pb the best predictions are from G18_02a, G18_10a, and NuWro. NEUT5.4.0 and G18_10b have large disagreement with the Pb data with χ^2 per data point of about 5. The most significant deviations are at low Q_p^2, where nuclear effects tend to be more important. Perhaps the most interesting comparison is between G18_10a and G18_10b because only the FSI model changes between them. Although G18_10a describes the A dependence well, G18_10b (as well as NEUT) grows faster at low Q_p^2 than the data.

I. Discussion

A number of interesting comparisons against three results from T2K [91,122,125] and MINERvA [123,126,129] have been shown in this section. All use ν_μ beams and all use C or CH targets except Ref. [129].

Recent cross-sections measurements are presented as a function of muon kinematics, in either $p_\mu - \cos\theta_\mu$ (T2K, Figs. 10 and 11) or $p_T - p_{||}$ (MINERvA, Fig. 12), that are variables directly measured, thus ensuring a minimization of model dependence. In order to help the comparison between the two mentioned experiments, in Fig. 15, model predictions for $q_0 - q_3$ are also reported, together with the corresponding detection efficiencies.

When comparing data and generator predictions, it is interesting to note that the older and less sophisticated model (G18_02a) is the worst model overall in describing the data but is not significantly worse. Also, NuWro seems unable to reproduce T2K data well, while it behaves

slightly better for MINERvA. GENIE G18_02a and G18_10b show similar behaviors and are very close to NEUT predictions in the case of T2K, while they slightly differ when looking at MINERvA results.

It should be stressed that χ^2 presented in this section are obtained using the covariance matrices as provided by the two collaborations. This means that normalization and shape errors are accounted at the same time, thus implying significant correlations between bins, due for instance to the flux normalization uncertainty. Therefore, the χ^2 calculation could be affected by Peelle's Pertinent Puzzle [130], although according to T2K publications [91,122], where also shape-only χ^2 are provided, the conclusions about data-MC agreement do not change.

A comparison of T2K and MINERvA data in a similar phase-space region in muon kinematics was made. To take out the known energy dependence (larger cross section for the higher beam energies of MINERvA), the ratio data/Monte Carlo are also shown. In general, in this particular phase-space region, MC predictions underestimate the data. The only exception is represented by NEUT that matches very well the MINERvA results. For all the other generators, the underestimation of the data seems to be equivalent in T2K and MINERvA, thus suggesting that the energy dependence of the models is approximately correct for the CC-0π interaction.

The TKI variables (Fig. 17) provide a more detailed way to explore nuclear and FSI models. δp_T distributions, which are more sensitive to the nuclear model, disfavor G18_02a for both T2K and MINERvA. The T2K preference is clearly for LFG models as implemented in GENIE and NEUT, while NuWro is disfavored. Concerning FSI models, that can be in principle tested by looking at $\delta\alpha_T$ distributions, T2K shows no particular preference for G18_10a or G18_10b while MINERvA has a slight preference for G18_10a; all χ^2 values are similar, except for NuWro that is the most disfavored.

Two different MINERvA analyses measuring Q^2 have been considered: one where the Q^2 variable is estimated starting from muon kinematics and the other one where the proton in the final state is used to evaluate the Q^2; this second, and older, analysis used data of neutrino interactions on three different nuclei (see Fig. 20). It has been reexamined because of its uniqueness as a way to probe the A dependence. Agreement between data and MC is best for the lightest target (CH) that does not show a particular preference for one generator. NuWro and G18_10a are also able to nicely describe cross section on heavier nuclei, while NEUT and G18_02a show a more rapid increase in the low-Q^2 region and a significantly higher normalization than the Pb data. Both Fe and Pb data seem to prefer G18_10a over G18_10b, thus preferring the data-driven FSI model.

VIII. PION PRODUCTION INTERACTIONS

Pion production data are crucial for neutrino oscillation experiments at higher beam energies such as NOvA and DUNE, where final states with one or several pions are dominant. Models for neutrino-induced single-pion production in NEUT and GENIE are based on modifications to the Rein-Sehgal model, whereas NuWro uses a Δ-only model. These models are generally tuned to low statistics bubble chamber data, where the neutrino scatters on deuterium and/or hydrogen. Effective models for the nuclear medium are added, but none of the models contain effects on the Δ production amplitude shown to be important in descriptions of pion-nucleus scattering data [131]. In addition, the effects of long-range nucleon–nucleon (NN) correlations are unknown. NN correlations and the nucleon density distributions [66] are known to be important for CC-mesonless interactions (see Sec. VII) but have not been applied to pion production interactions. Another major challenge in using neutrino-*nucleus* scattering data to tune the underlying nucleon production model is the presence of pion FSI. Because of the difficulties in describing hadron propagation through nuclei with a quantum-mechanical model, generators use intranuclear cascade models. These are often semi-classical models with corrections accounting for nuclear effects via some effective approach. As described in Table II, GENIE v3, NEUT, and NuWro all have implementations of the nuclear effects in the Salcedo-Oset [38] cascade model, although the details differ slightly. GENIE additionally offers an effective "single-step" cascade model with its hA model which is tuned to hadron-nucleus data. FSI models can additionally be informed by pion-nucleus scattering data [38,132], although the relationship between pion-nucleus scattering and pions being produced in medium and propagating out is nontrivial [57,79].

In this section, we investigate recent charged pion production publications from T2K and MINERvA which serves as an update on the previous TENSIONS2016 work [12], where tensions between the MiniBooNE CC1π^+ result [13] and the first MINERvA CC1$\pi^\pm$ result [14] were investigated.

A. T2K results

The most recent CC1π^+ analysis from T2K [133] used data corresponding to 1.51×10^{21} POT and used FGD1, a plastic scintillator detector (C_8H_8), as its target. The measurement is dominated by single-pion production via a resonant interaction, most prominently the $\Delta(1232)$ resonance due to T2K's beam energy. Coherent pion production and soft inelastic interactions also make small contributions to the single-pion topology, with very few deep inelastic events. The analysis used NEUT5.1.4.2 as its neutrino interaction simulation (Table II) during the development of the analysis.

1. Selection

The analysis selects events with a forward-going muon and a charged pion in the final state. The pion is tagged either by the presence of a pionlike track in a TPC, that shares a common vertex in FGD1 with the muon track, or by a time-delayed Michel electron in FGD1 implying a pion with momentum below tracking threshold. Protonlike tracks were allowed but not required. The selection provides a pure sample with minimal muon-pion confusion, primarily due to the charge as determined by track curvature from the magnetic field in the TPC, with good momentum and angle resolution. The selection limits the ranges in track momentum and angle, since muon tracks must have enough momentum to escape FGD1 and they must have relatively forward angles with respect to the beam direction to enter the TPC. The signal definition includes restrictions on both muon and pion kinematics to reflect this.

2. Variable definitions

The differential cross section is analyzed in seven distributions: the two-dimensional $d^2\sigma/dp_\mu d\cos\theta_\mu$, and the one-dimensional $d\sigma/dQ^2_{\rm rec}$, $d\sigma/dp_\pi$, $d\sigma/d\theta_\pi$, $d\sigma/d\theta_{\pi\mu}$, $d\sigma/d\phi_{\rm Adler}$, and $d\sigma/d\cos\theta_{\rm Adler}$.

Importantly, the "theory variables" of $Q^2_{\rm rec}$, $\cos\theta_{\rm Adler}$, and $\phi_{\rm Adler}$ used in the measurement do not correspond to the usual theory variables which require exact knowledge about the neutrino energy and direction, which would involve model-dependent corrections to the observed event. Instead they are proxy variables, derived using the lepton and pion kinematics only, assuming that there were three outgoing particles. The variables require both a muon and a pion track, and $Q^2_{\rm rec}$ is reconstructed as $Q^2_{\rm rec} = -q^2 = -(p_\nu - p_\mu)^2 = -(m_\mu^2 - 2E_\nu^{\rm rec}(E_\mu - |\vec{p}_\mu|\cos\theta_\mu))$, where the neutrino energy is *reconstructed* using detector observables to reduce model dependence:

$$E_\nu^{\rm rec} = \frac{m_p^2 - (m_p - E_b - E_\mu - E_\pi)^2 + |\vec{p}_\mu + \vec{p}_\pi|^2}{2(m_p - E_b - E_\mu - E_\pi + \vec{d}_\nu \cdot (\vec{p}_\mu + \vec{p}_\pi))}, \quad (10)$$

with $E_b = 25$ MeV and $\vec{d}_\nu$ being the average predicted neutrino direction. Although this means $Q^2_{\rm rec} \neq Q^2_{\rm true}$ for the T2K result, it was preferred in the publication as it avoids model dependence, and still maps well to $Q^2_{\rm true}$.

The Adler angles are defined in Fig. 21 which, when combined with Eq. (10) to estimate $E_\nu^{\rm rec}$, provide a more model-independent proxy variable for the usual Adler angles, which are defined on the nucleon level in the resonance rest frame. Hence the Adler angle measurements produced by T2K do not perfectly correspond to the more common Δ decay angles, since producing them involves significant dependence on the nuclear model [134]. Furthermore, the publication uses the Adler angles defined by the neutrino and muon kinematics instead of the more
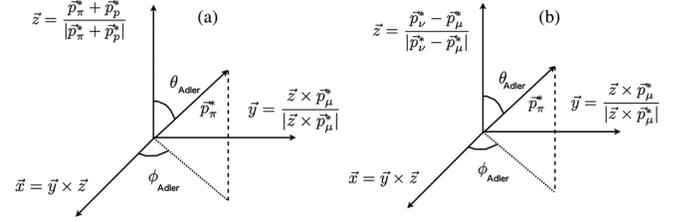
FIG. 21.   Definition of the Adler angles in the resonant rest frame, using (a) pion and nucleon kinematics and (b) neutrino and muon kinematics. Reproduced from Ref. [133].

commonplace pion and nucleon kinematics, since the ability of ND280 to reconstruct the outgoing nucleon is relatively poor. The Adler angles can be used to infer the strength of interference between resonant and nonresonant pion production, which polarizes the resonance [134].

3. Signal definitions

The cross-section measurements' signal definitions are summarized in Table X, and differ slightly from each other due to the different efficiencies in the variables, and the kinematics that the variables require. For instance, the $d^2\sigma/dp_\mu d\cos\theta_\mu$ distribution only requires a measurement of the muon and a tag for the pion; hence, T2K can tag the pion via a delayed Michel electron in addition to the usual track criteria, so the restrictions on p_π and $\cos\theta_\pi$ are removed. The statistical power of the measurement is therefore the highest out of all the variables in the publication.

The cuts are always on the outgoing observed particle kinematics according to truth (MC), and not on auxiliary variables, such as the hadronic mass, W.

4. Cross-section extraction

The measurement uses three sideband samples looking for a right-signed muon, further enriched in $CC0\pi 1p$, $CC2\pi^+$, and $CCNe^\pm 0\pi^+$ events to constrain pion/proton confusion, missing charged pions, and neutral pion backgrounds, respectively. The event rate extractions were performed separately for each differential cross section, and the number of events for overlapping regions of phase

TABLE X.   Phase-space cuts on particle kinematics included in the signal definition in the various T2K CC$1\pi^+$ measurements.

Measurement	$\cos\theta_\mu$	p_μ (GeV/c)	$\cos\theta_\pi$	p_π (GeV/c)
$d^2\sigma/dp_\mu d\cos\theta_{\mu,\nu}$	>0.0			
$d\sigma/dp_\pi$	>0.2	>0.2	>0.2	
$d\sigma/d\theta_{\pi,\nu}$	>0.2	>0.2	>0.0	>0.2
$d\sigma/dQ^2_{\rm rec}$				
$d\sigma/d\theta_{\pi,\mu}$				
$d\sigma/d\phi_{\rm Adler}$			>0.2	
$d\sigma/d\cos\theta_{\rm Adler}$				

FIG. 22. Double-differential cross section as a function of p_μ and $\cos\theta_\mu$ from T2K, showing the predictions of various neutrino interaction generators.

space were not required to match. However, all results were found to be consistent with the model within statistical and systematic uncertainties.

After background subtraction, the number of events in each bin of the true kinematic variable was estimated from the data with a single iteration of the D'Agostini [88] unfolding procedure, and was repeated for each of the aforementioned kinematic variables. The single-iteration unfolding was chosen by studying full NEUT and GENIE productions on ND280 with different alternate parameter sets, balancing the statistical error size with the unfolding robustness [135]. However, the use of a single iteration of D'Agostini, and the features of the efficiency function (see Sec. VIII C), opens potential for bias toward the input signal MC in this analysis.

5. Results

The differential cross section in muon momentum in the four muon angular bins is shown in Fig. 22. The overall variation in predictions among the generators is relatively small, and they are generally good against data. NuWro has the largest χ^2, coming predominantly from the two highest-momentum bins in the $0.80 < \cos\theta_\mu < 0.90$ region. NEUT has the lowest χ^2, and a notably different prediction in the 0.4–1.2-GeV bin; however, no particular region drives the lower χ^2. Little variations is seen between the various GENIE models, which use different form factors and have a different FSI treatment. The six single-differential measurements shown in Figs. 23–25 provide more detailed information, and exhibit larger differences between the generators.
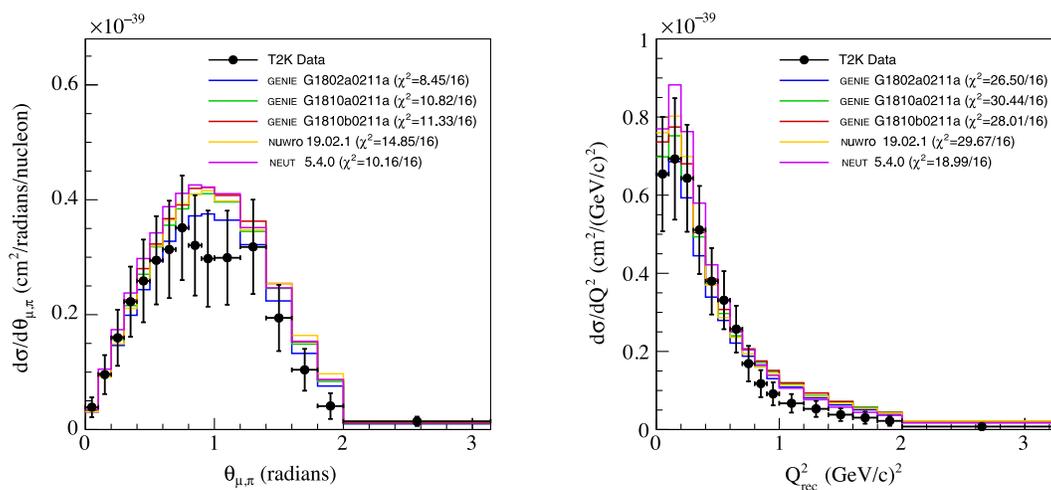
FIG. 23. Single-differential cross section in $\theta_{\mu,\pi}$ and Q^2_{rec} from T2K, showing the predictions of various neutrino interaction generators. The Q^2_{rec} definition does not correspond to the fundamental interaction vertex Q^2_{true}, but is instead reconstructed from the visible particles observed in the detector after they exit the nucleus.
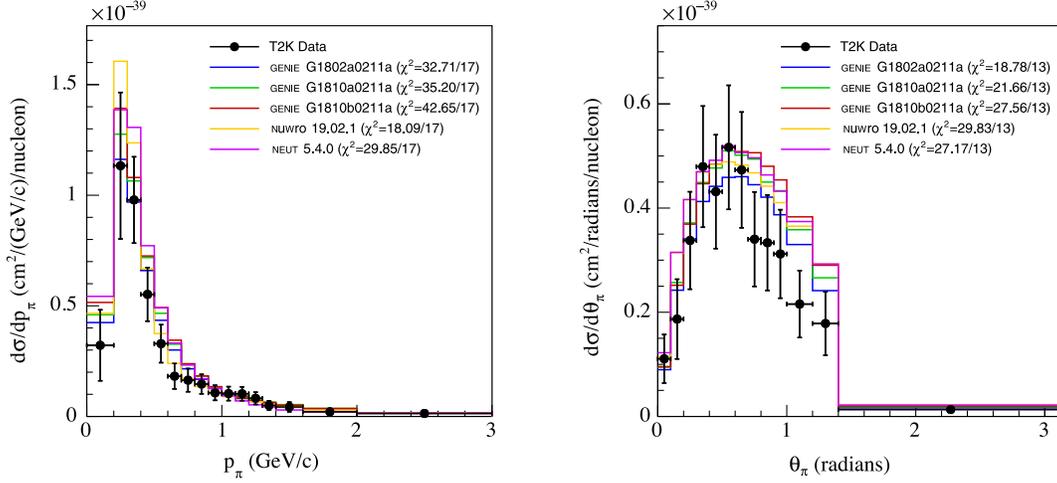
FIG. 24. Single-differential cross section in p_π and $\theta_{\pi,\nu}$ from T2K, showing the predictions of various neutrino interaction generators.

For the Q^2_{rec} distribution in Fig. 23, we note that all generators exhibit a turnover behavior at low Q^2_{rec}, also observed in the data. The publication [133] used older versions of GENIE (2.8.4) and NEUT (5.1.4.2), and saw an overprediction at low Q^2_{rec}, which does not appear with our recent generator versions. NEUT shows the best overall agreement, although it consistently overpredicts at $Q^2_{\text{rec}} < 0.5$ GeV^2, possible due to the strong bin-by-bin correlations in the data. Looking only at the shape of the distribution, GENIE underpredicts the low- and overpredicts the high-Q^2_{rec} region, whereas NEUT and NuWro do better at low Q^2_{rec}, but also overpredict the high-Q^2_{rec} region, although to a lesser degree. NuWro and the various GENIE versions receive largest $\Delta\chi^2 \sim 4$ contributions from the bins around $Q^2_{\text{rec}} = 1$–1.5 GeV/c^2 and all generators do poorly in the last bin, with $\Delta\chi^2 \sim 2$–3.

The muon-pion opening angle distribution in Fig. 23 is sensitive to a variety of effects related to the initial state and the resonance decay. The generators all have good χ^2/N_{bins}, even though the predictions are notably different, due to the weak statistical power of the measurement. Generally, NEUT has the largest prediction, and G18_10a, G18_10b, and NuWro all predict similar distributions, with G18_02a differing from the other GENIE versions mostly by normalization.

Figure 24 shows the pion momentum distribution, and all generators tend to be higher than the data for $p_\pi < 0.8$ GeV/c. A peak at 0.2–0.3 GeV/c is seen which is shaped from the fundamental nucleon interaction, Δ-resonance propagation in the nucleus, and pion FSI. Although the generators predict the same peak, NuWro has a more pronounced peak which appears to overpredict the data considerably. This is also true without enforcing the T2K p_μ, $\cos\theta_\mu$ and $\cos\theta_\pi$ cuts, and we conclude it is a genuine feature of the generator model choice. However, NuWro also has the best χ^2 by a factor of 2, coming

predominantly from the $0.5 < p_\pi < 1$ GeV/c region, where the other generators consistently overpredict. All GENIE versions have the worst χ^2, coming from the last pion momentum bin, which contributes 9 units of χ^2, compared to less than 1 unit of χ^2 for NEUT and NuWro. NuWro's dominant bin contribution is 1.3–1.4 GeV/c with $\Delta\chi^2 = 4$, and NEUT's is 0.6–0.7 GeV/c with $\Delta\chi^2 = 5$.

The changes to the GENIE pion FSI model are most pronounced around the pion momentum peak, although this region does not particularly affect the χ^2 against the data. Comparing the shapes, all three GENIE versions produce similar predictions. We theorize that p_π is more sensitive to the details of the resonance decay in the nuclear medium and secondarily to pion FSI, so the larger deviations in p_π compared to $p_\mu - \cos\theta_\mu$ and Q^2_{rec}, likely come from these treatments.

Conversely, the situation is reversed for the distribution in the lab pion angle, θ_π, shown in Fig. 24. Here GENIE agrees best with the data, with NEUT, NuWro, and G18_10b having almost double the χ^2 of G18_02a, indicating that no simple picture is possible. G18_02a has the best χ^2 due to capturing the aggressive dropoff in cross section at higher pion angles. We also note that the pion FSI model in GENIE has barely any effect until $\theta_\pi > 0.6$ rad. All generators predict somewhat larger cross sections than data at larger pion angles ($\theta_\pi > 0.8$ rad), which is also the region where we observe largest difference between the generators. At lower angles, NEUT is the outlier, predicting a larger cross section than the other generators and the data.

The Adler angles are traditionally considered as the angles of the outgoing pion in the rest frame of the decaying resonance, before final-state interactions. As previously discussed, the T2K measurement's Adler angles instead concern a proxy variable of the true Adler angles, using only the outgoing muon and pion kinematics to form the variable, avoiding corrections for nuclear effects such as
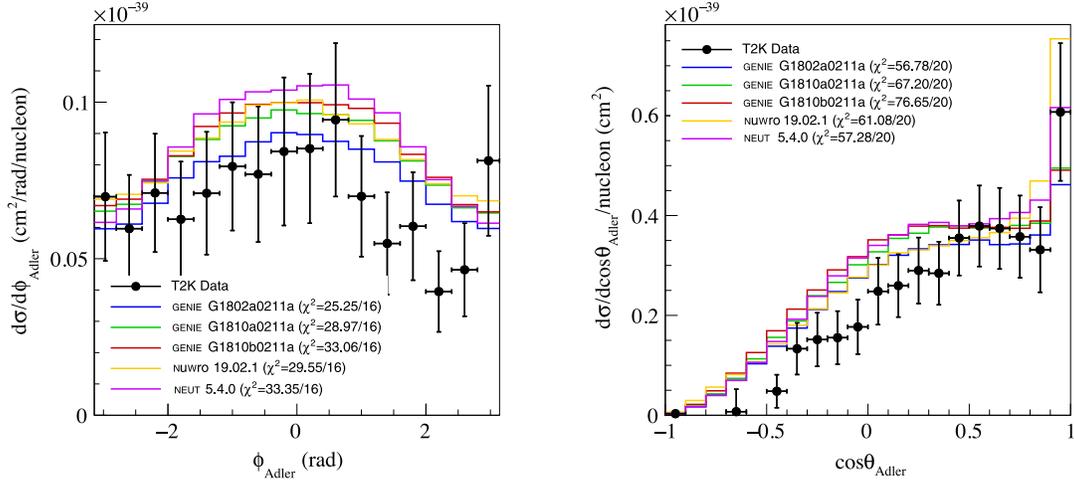
FIG. 25.   Single-differential cross section in the Adler angles ϕ_{Adler} and $\cos\theta_{\text{Adler}}$ from T2K, showing the predictions of various neutrino interaction generators. The Adler angles do not correspond to the fundamental interaction vertex Adler angles, but are instead reconstructed from the visible particles observed in the detector after they exit the nucleus.

FSI. As such, they do not perfectly map to the traditional Adler angles, but also do not contain the strong model dependence that deriving the true Adler angles would. The true Adler angle variables are the decay angles of the resonance in its rest frame which is the output of electro-production experiments. These decay angles were measured in bubble chamber experiments [55] and are the source of the event generators' decay models; GENIE, NEUT, and NuWro all use parametrizations of the $\Delta(1232) \to \pi N$ system based on neutrino H/D data. The Adler angles may also have some sensitivity to resonance/nonresonance interference [134], which the generators also implement differently: GENIE ignores interference, NuWro only models a single noninterfering $\Delta(1232)$ resonance, and NEUT has 17 interfering resonances and one noninterfering, nonresonant $I_{1/2}$-only background. A variety of models have been used to refit the H/D data [55]. Although the new results are not significantly different than the original work, the applicability is uncertain because the new models are different than the models used to extract the original weak form-factor data.

The Adler angles for the T2K data in Fig. 25 are very poorly predicted by all generators, although the predictions differ. In ϕ_{Adler} NEUT shows the most peaked distribution and has the worst χ^2, whereas the other generators show similar behavior up to a normalization. Although the G18_02a prediction has the lowest cross section and the best χ^2 with $\chi^2/N_{\text{bins}} \sim 1.5$, the shape is the most asymmetric. All three GENIE versions have the same decay implementation, and the small differences show the effect of other aspects of the dynamics. The effect of pion FSI between G18_10a and G18_10b appears mostly flat up to a normalization. In the $\cos\theta_{\text{Adler}}$ distribution, we note a consistent overestimation in $\cos\theta_{\text{Adler}} < 0.5$ for all generators. The χ^2 are about 3 per bin for $\cos\theta_{\text{Adler}}$; the

problems are with both shape and magnitude. Comparing the generators, we note significantly different behavior in the most forward $\cos\theta_{\text{Adler}}$ angles, where NuWro rapidly rises, with almost 50% larger cross section than the G18_02a. NEUT also rises rapidly, but to a lesser extent, possibly a reflection of the similar nonisotropic Δ decay the two generators have in common.

A summary of the χ^2 contributions for all variables is provided in Table XI. Overall all the generators appear to be able to describe the data reasonably well in muon kinematics and $\theta_{\pi,\mu}$. The Q^2_{rec} distribution is moderately well described, with all generators having $p_{\chi^2} > 0.01$ in common. NuWro is notably best at describing the pion momentum distribution, G18_02a at describing the pion and pion-muon angular distributions, and NEUT at describing the muon kinematic distributions and Q^2_{rec}. A general trend is that the generators all agree better with data in muon kinematic variables than pion kinematic variables.

TABLE XI. χ^2 contributions for the generators against T2K CC1π^+ data, over a total of 114 bins. The χ^2 values for individual distributions are calculated using the provided covariance matrices. The summed χ^2 for each generator does not account for correlations between measurements in different variables, and so do not represent a proper global χ^2.

Measurement	N_{bins}	G18_02a	G18_10a	G18_10b	NuWro	NEUT
$p_\mu - \cos\theta_\mu$	16	16.58	18.94	19.03	24.17	11.81
p_π	17	32.71	35.20	42.65	18.09	29.85
$\cos\theta_\pi$	13	18.78	21.66	27.56	29.83	27.17
$\theta_{\mu,\pi}$	16	8.45	10.82	11.33	14.85	10.16
Q^2_{rec}	16	26.50	30.44	28.01	29.67	18.99
ϕ_{Adler}	16	25.25	28.97	33.06	29.55	33.35
$\cos\theta_{\text{Adler}}$	20	56.78	67.20	76.65	61.08	57.28
All	114	185.05	213.23	238.29	207.24	188.61

The worst agreement is seen for the variables derived from a number of measured quantities, Q^2_{rec} and the Adler angles. Amongst the GENIE flavors, G18_02a performs best in every distribution. The impact of pion FSI in G18_10a vs G18_10b has very little effect on the predictions for $p_\mu - \cos\theta_\mu$, $Q^2_{\text{rec,}}$ and $\theta_{\mu,\pi}$. For the pion kinematic variables and Adler angles, G18_10a is consistently preferred, although the overall agreement with data in those variables is still not good.

B. MINERvA results

MINERvA has published pion production measurements on CH with varying final states—ν_μ CC1π^+ [14], $\bar{\nu}_\mu$CC1π^0 [136], ν_μCC1π^0 [137], and $\bar{\nu}_\mu$CC1π^- [138]. For this paper, we have selected ν_μ CC1$\pi^\pm$ data to allow optimal comparison with T2K [133] and MiniBooNE [13]. The analysis has been updated since Ref. [14] to use the improved signal definitions of Ref. [136], and is available as a public data release [139]. The updated results for this measurement include an improved flux estimate [103] and a modified signal definition. As with the CC-mesonless measurement discussed in Sec. VII C, the measurement uses a muon neutrino beam with mean energy of 3.5 GeV, and it selects events in the MINERvA central tracker, composed mainly of CH.

1. Signal definition

The analysis focuses on $\pi^\pm$ production in the Δ (P$_{33}$(1232)) resonance region which is the most prominent baryon resonance for neutrino interactions. It is the successor to Ref. [14] with some significant improvements that were developed for Ref. [136] including a new flux calculation [103] and an updated signal definition. It has no signal definition restriction corresponding to the acceptance limitations for pion kinetic energy, pion angle, muon momentum, or muon angle.[2]

There are kinematic limits on the signal in terms of quantities that are not directly observable in the detector. The true neutrino energy is required to be between 1.5 and 10 GeV. Additionally, a key kinematic constraint in the analysis is on the invariant hadronic mass, W. The purpose is to suppress higher mass resonances and allow comparison with measurements and calculations that emphasize $\Delta(1232)$ resonance. In the original version of the analysis [14], a signal constraint was placed on the true, generator-invariant hadronic mass, requiring $W < 1.4$ GeV. In the new version (method in Ref. [136] and data in Ref. [139]), the signal constraint was placed on the event's W_{exp} calculated with true muon kinematics and true E_ν, as

---

[2]The original analysis [14] did include a version with a muon angle constraint in its Appendix, but this constraint was removed for the updated analysis, to better align with MINERvA's later pion measurements, specifically Ref. [136].

defined in Eq. (13). Equations (11) and (12) show the way the scalar and invariant transferred energy are calculated. The equation for W_{exp} has the significant assumption that the struck nucleon is at rest.

$$\nu = E_\nu - E_\mu, \tag{11}$$

$$Q^2 = 2E_\nu(E_\mu - |\vec{p}_\mu|\cos\theta_{\mu\nu}) - m_\mu^2, \tag{12}$$

$$W^2_{\text{exp}} = m_p^2 - Q^2 + 2m_p\nu. \tag{13}$$

The signal definition requires a single charged pion ($\pi^\pm$), though the Michel requirement selects >99%π^+. The CC1π^- contribution to the CC1$\pi^\pm$ cross section is calculated to be less than 2% at MINERvA energies. Any additional baryons and mesons (including π^0) are allowed.

2. Selection

This analysis uses the selection of Ref. [136] as applied to the signal definition defined above. It selects events with a muon track and charged-pion track that share a common vertex. The muon track must exit the back of MINERvA and enter MINOS, effectively restricting its momentum to about $p_\mu > 1.5$ GeV/c and angle to $\theta_\mu < 20°$. All π^+ in this analysis are identified by a Michel electron in MINERvA, and are required to produce a track with an energy deposition signature consistent with a charged pion. This imposes limitations on the kinetic energy of the pion, T_π: there is a $T_\pi > 35$ MeV tracking threshold in MINERvA, and pions with $T_\pi > 350$ MeV typically exit the back of MINERvA, and therefore cannot be fully reconstructed. Due to MINERvA 's planar design, pion-tracking efficiency drops to zero between $80° < \theta < 110°$. Since the p_μ, θ_μ, T_π, θ_π restrictions exist in the selection but not in the signal definition, the model is relied upon in these regions at the efficiency correction stage of the cross-section calculation which necessarily implies model dependence.

The reconstructed energy transfer, ν^{reco}, referred to as E_{recoil} and E_{hadronic} in the publications, is estimated by summing the visible energy not associated with the muon, and applying corrections determined from simulation that correct for energy missing in passive material, energy escaping the detector, energy below threshold, and energy lost to the nucleus (i.e., binding energy). The reconstructed neutrino energy, E_ν^{reco}, is calculated as $E_\mu^{\text{reco}} + \nu^{\text{reco}}$. The reconstructed Q^2 and W_{exp} are calculated as in Eq. (12) and Eq. (13), respectively, but using reconstructed muon kinematics and the reconstructed energy transfer.

3. Cross-section extraction

The signal selection efficiency is approximately 8%, with the largest losses coming from the Michel selection. The high-W_{exp} background is constrained to sidebands as
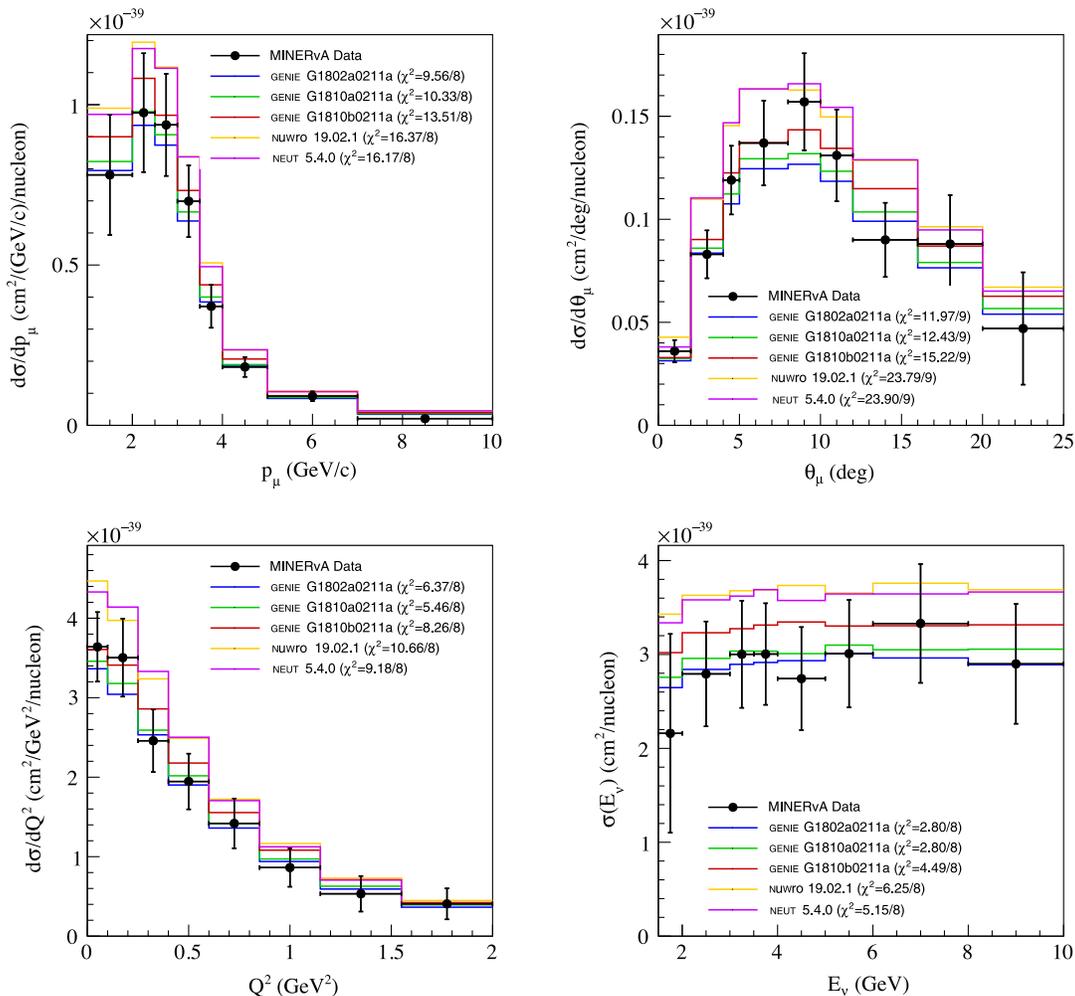
FIG. 26.   MINERvA CC1π^+ cross-section measurements and model comparisons in muon variables (top) and Q^2 and E_ν (bottom).

in Ref. [14] with the new signal definition. The cross section is extracted using D'Agostini unfolding [88] with four iterations. The number of iterations was determined by examining MC simulations and the value chosen had the best balance between correctly unfolding and problems with statistics. The second major improvement in the updated result is an improved neutrino flux [103] which resulted in a increase in cross section which was flat across pion kinetic energy. The redefined signal definition resulted in a decrease in cross section largely independent of pion energy.

4. Results

One-dimensional differential cross sections were measured in muon and pion kinematic variables, four-momentum transfer squared, Q^2, and neutrino energy, E_ν. Since all quantities were unfolded from the reconstructed quantities to the true quantities, theoretical calculations can be compared directly with them (acknowledging that this unfolding is model dependent as it corrects for nuclear

effects that can impact the relationship between E_ν^{true} and E_ν^{reco}). Results are shown in Figs. 26 and 27 with comparisons to the same calculations as for the previous T2K section. In the previous TENSIONS2016 paper [12], the generators were 20–30% above the MINERvA data [14]. Since then, the data have been reanalyzed (Sec. VIII B 1) and the GENIE calculations have been improved with better fits to the ν_μ–D$_2$ data [140,141] and GENIE, NuWro and NEUT have also made improvements to their FSI models.

The lepton variables p_μ and θ_μ can be used to gauge the overall cross-section magnitude. By eye we see NEUT and NuWro have very similar predictions, and consistently overpredict the data below 5 GeV, and GENIE predicts the data better. However, when accounting for the strong bin-by-bin correlations, we see very large χ^2 contributions from a single bin: the $7 < p_\mu < 10$ GeV/c bin, contributing 9 units of χ^2 for NEUT, G18_10b and NuWro, and 6 units for the other GENIE versions. Once this bin is excluded, NEUT and NuWro go from $\chi^2/N_{\text{bins}} \sim 2$ to a very agreeable $\chi^2/N_{\text{bins}} \sim 1$, and we see no definitively superior
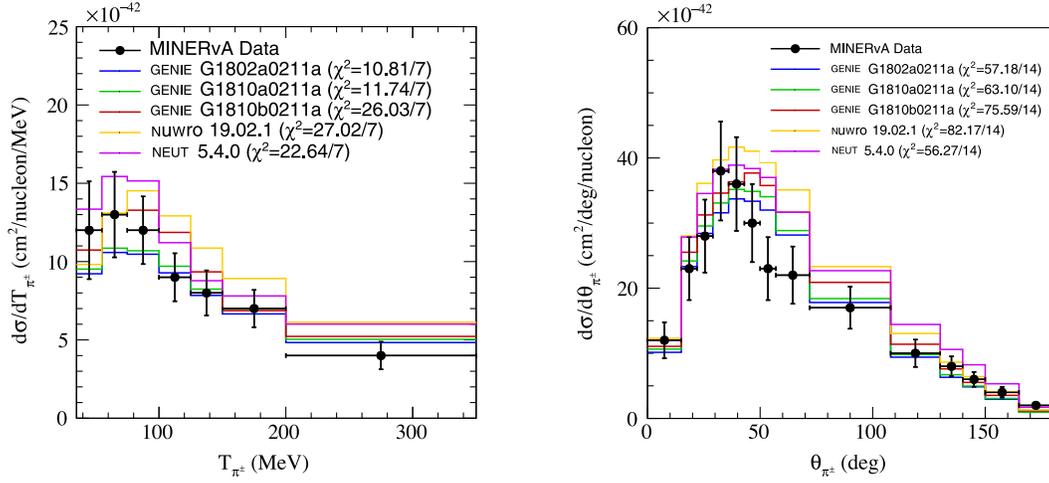
FIG. 27.   MINERvA CC1π^+ cross-section measurements and model comparisons in pion kinetic energy and angle.

prediction. When comparing the shapes of the generators in p_μ, they are remarkably similar, and the total predictions differ almost exclusively in normalisation. For the θ_μ distribution, the predominant χ^2 contribution is the $12° < \theta_\mu < 16°$ bin, which contributes $\chi^2 \sim 10$ for NEUT and NuWro, and 7 units for GENIE. Again, if this bin is excluded all generators perform well in regard to the χ^2/N_{bins}.

Q^2 can probe nuclear structure, and E_ν separates the interaction modes relatively well. However, the Q^2 and E_ν distributions have significantly more model dependence than the purely kinematic distributions observed in the detector. This is because they require corrections for initial and final-state physics, coming from the input generator model. Comparing the predictions in Fig. 26, the generators perform adequately in regard to the χ^2, although we see relatively large differences between the predictions; at low Q^2 G18_02a predicts a 30% smaller cross section than NuWro. NuWro and NEUT perform very similarly, likely due to similar form factors and tuning to data. They are somewhat higher than the data with comparable χ^2, coming from the limited power of the data. The CC1$\pi^\pm$ cross section does not indicate a missing low-Q^2 dropoff and calculations appear to follow the data well, in common with CC1π^- [138] production, but in contrast to MINERvA's measurements of CC1π^0 [137]. This is discussed more in comparison to T2K data in Sec. VIII C.

The pion variables, shown in Fig. 27, are more sensitive to resonance decay processes and FSI effects. We observe very similar behavior between G18_10a and G18_02a, with a significantly lower cross section than G18_10b NEUT and NuWro, and with a significantly better χ^2 in T_π; largely observed in the previous muon-based variables. Although the peak below $T_\pi < 100$ MeV is underpredicted, the higher region drives the χ^2, with the last bin contributing 6.5 units of χ^2 for G18_02a, and 16 units for NEUT. NuWro however receives its largest χ^2 penalty in the 55–76 MeV

and 100–125 MeV bins, with 6.5 and 7.5 units, respectively. Comparing G18_10a and G18_10b, which differ only in FSI model, the empirical model produces an acceptable χ^2 for the kinetic energy distribution, whereas the cascade-based G18_10b does not, and shares more features with NEUT and NuWro in prediction. Interestingly, the very similar predictions of NEUT and NuWro in the lepton variables in Fig. 26 are different in the pion variables; NEUT overpredicts almost the entire range but does well in the shape, whereas NuWro peaks at a higher T_π than data, underpredicts the lowest bin, and overpredicts the remaining distribution at higher T_π, with a different shape from the other generators.

In θ_π all generators produce an unsatisfactory χ^2, with more than 4 units per bin on average. Similar to the T2K measurement, the rising shape of the peak is well modeled, but the dropoff at higher angles is overpredicted. There is a large impact from the GENIE FSI model choice, and we see differences between NEUT and NuWro not observed in the muon kinematics. We note that the high-angle region is where the model dependence of the data may be the strongest, and note in the region of $\theta_\pi \sim 100°$ that the data are almost perfectly reproduced by G18_02a and G18_10a predictions.

It is clear from the χ^2 table in Table XII, that G18_02a generally exhibits the best agreement with MINERvA data, with the exception of the Q^2 and θ_π distributions. We note G18_02a is most similar to the generator used in the analysis (GENIE2.8.6). Comparing G18_10a and G18_10b—which differ only by their pion FSI model—the empirical model used in the former consistently performs better, often similar to G18_02a.

NuWro generally has the worst χ^2, and is very similar to NEUT in muon variables, similarities which disappear when pion variables are concerned. We note that the Q^2 and E_ν cross-section results holds very little statistical power, which is likely a result of large model uncertainties that

TABLE XII. χ^2 contributions for the generators against MIN-ERvA CC1π^+ data, over a total of 54 bins. The χ^2 values for individual distributions are calculated using the provided covariance matrices. The summed χ^2 for each generators does not account for correlations between measurements in different variables, and so do not represent a proper global χ^2.

Measurement	N_{bins}	G18_02a	G18_10a	G18_10b	NuWro	NEUT
p_μ	8	9.56	10.33	13.51	16.37	16.17
θ_μ	9	11.97	12.43	15.22	23.79	23.90
T_π	7	10.81	11.74	26.03	27.02	22.64
θ_π	14	57.18	63.10	75.59	82.17	56.27
Q^2	8	6.37	5.46	8.26	10.66	9.18
E_ν	8	2.80	2.80	4.49	6.25	5.15
All	54	98.69	105.86	143.10	166.26	133.31

come from correcting and unfolding to these variables in data. The θ_π predictions presents the largest χ^2/N_{bins} with no generator performing well. This has also been seen in previous studies of these data [142].

C. T2K and MINERvA pion production phase-space comparisons

1. Cross-section comparisons

Direct comparison of the T2K and MINERvA cross-section data is worthwhile because the targets are the same and the signal definitions are similar. However, it is challenging given the differences in the neutrino fluxes: MINERvA has a peak energy of 3.5 GeV, whereas the peak T2K energy is at 600 MeV, shown in Fig. 3. T2K's ND280 detector is at an off-axis angle of 2.5° which suppresses the high-energy tail as compared with the on-axis MINERvA detector. In addition, the signal definitions and cuts (e.g., pion kinetic energy) tend to make overlap less common. MINERvA's cut on $W_{\text{exp}} < 1.4$ GeV ensures that the cross section is dominated by the $\Delta(1232)$ as in T2K. However, the strength of the resonant and nonresonant contributions

still depend on E_ν, W, and Q^2, and so will see different strengths at the two experiments.

The small overlap region where a direct comparison is possible between MINERvA and T2K is for events with $1 < p_\mu < 2$ GeV/c. MINERvA's muon measurements are limited to a range in muon angle between roughly $0.9 < \cos\theta_\mu < 1.0$ where T2K also has coverage, and both cover a nearly full angular range for pions. Thus, the range where the two datasets can be compared is limited to the first bin of p_μ for MINERvA (Fig. 26), and the two highest-momentum bins for T2K (lower right panel of Fig. 22). In this limited region, the two results are largely consistent with each other, and the generator-data comparisons are consistent for the two, which is reassuring. However, we note that the $(q_0, \vec{q})$ phase space after the signal definition cuts are included for T2K and MINERvA are notably different, even with the $W < 1.4$-GeV cut. Hence, the agreement is more coincidence than sharing identical physics processes.

Although the selections for T2K and MINERvA occupy different ranges of T_π and p_π, there is some overlap in the pion phase space. Firstly, $100 < T_\pi < 350$ MeV measured by MINERvA roughly corresponds to the $0.2 < p_\pi < 0.5$ GeV/c range measured by T2K. Secondly, both measurements are dominated by the $\Delta(1232)$, albeit at different neutrino energies, meaning the pions come from a similar interaction process at both experiments. The GENIE G18_02a, NuWro, and NEUT predictions for the single-pion production measurements are broken down by mode in Figs. 28 and 29. Restricting the range to the aforementioned T_π and p_π, it is clear that an overprediction in one experiment means an overprediction in the other; GENIE almost perfectly follows the data in the MINERvA measurement, and does very well in the equivalent range for T2K. NuWro and NEUT's overestimates are present for both experiments, and follow a similar shape. However, NuWro and NEUT have an almost identical CC1π^+1p contribution at MINERvA, but NuWro has a visibly larger prediction in
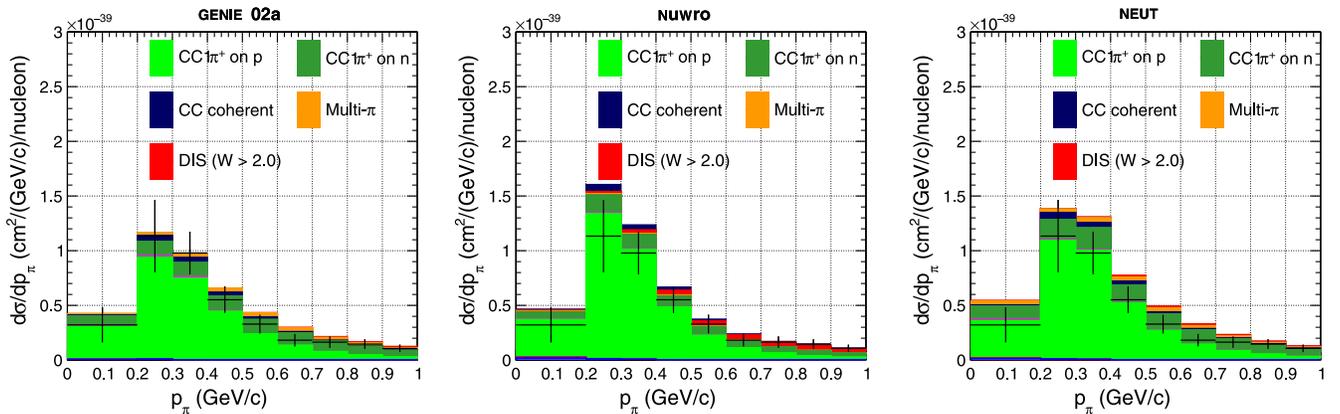
FIG. 28.   T2K CC1π^+ cross section in p_π for G18_02a, NuWro, and NEUT, broken down by true interaction mode, shown up to $p_\pi = 1$ GeV/c.
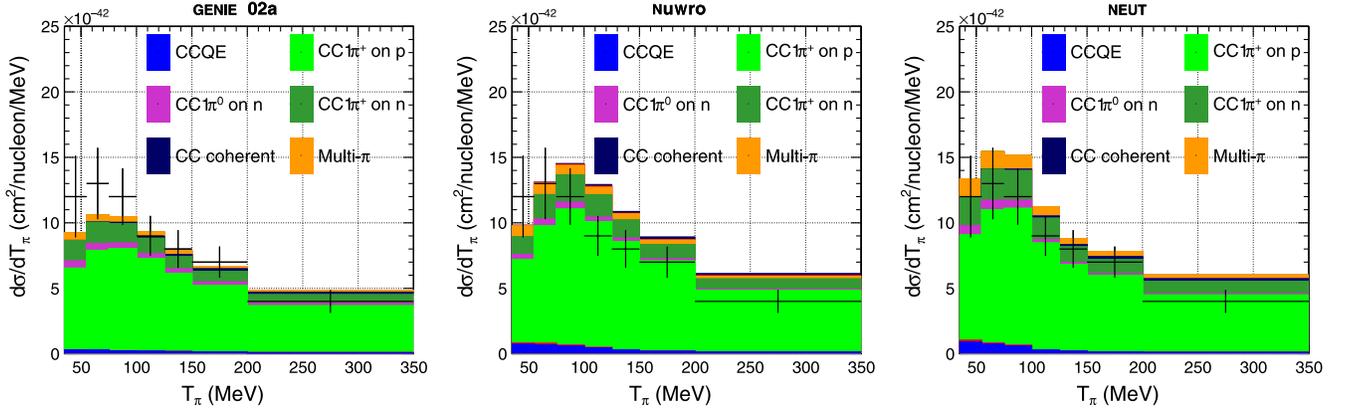
FIG. 29.   MINERvA CC1π^+ cross section in T_π for G18_02a, NuWro, and NEUT, broken down by true interaction mode.
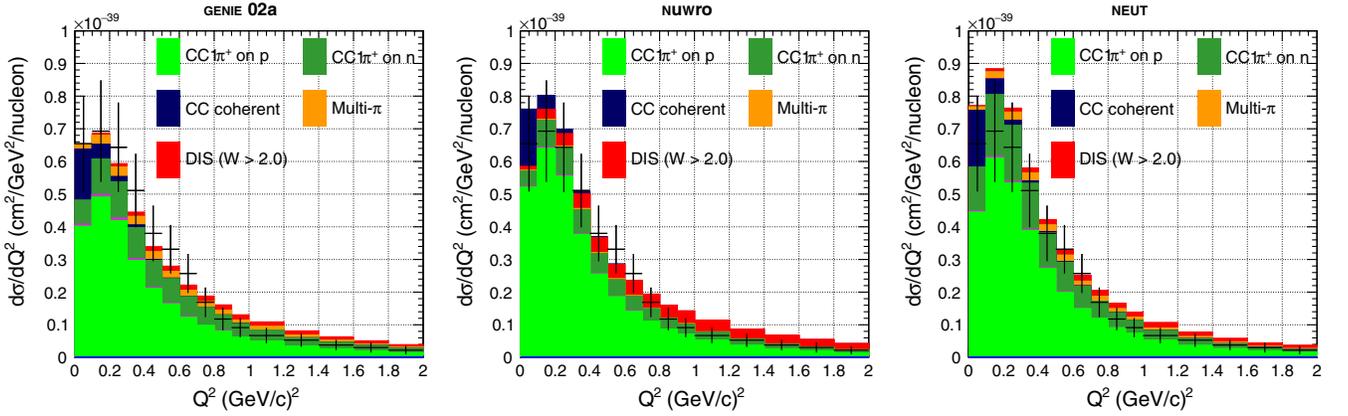
FIG. 30.   T2K CC1π^+ cross section in Q^2_{rec} for G18_02a, NuWro, and NEUT, broken down by true interaction mode.

the peak of the T2K distribution. There is a lack of CCQE events at low T_π for GENIE, whereas it makes up ∼10% for NuWro and NEUT in the first two bins of T_π; this is likely due to the nucleon FSI model present in the generators. We also note very few CC coherent events in the MINERvA distribution, whereas they are present in T2K. Inspecting the coherent cross section, it rises rapidly for MINERvA as

T_π increases, which implies that the low-Q^2 region of MINERvA may contain model dependence, as the pions from coherent events will mostly be from pions above MINERvA 's detection ability. We explore this further in Sec. VIII C 2.

In Figs. 30 and 31, we show the breakdown of the predictions for Q^2_{rec} and Q^2 according to true interaction
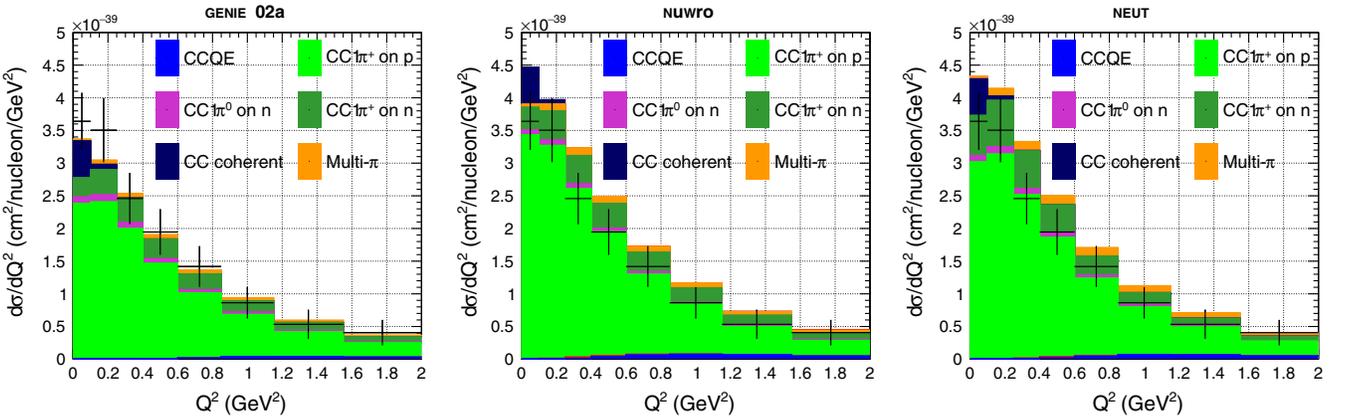
FIG. 31.   MINERvA CC1$\pi^{\pm}$ cross section in Q^2 for G18_02a, NuWro, and NEUT, broken down by true interaction mode.

mode for the T2K and MINERvA measurements, respectively. Separate panels are provided for G18_02a, NuWro, and NEUT. As expected from single-nucleon cross sections, the composition is dominated by $CC1\pi^+$ on a struck proton target via resonances, with smaller contributions from $CC1\pi^+$ on a struck neutron target via resonances, and barely visible contribution from $CC1\pi^0$ via resonances, where the π^0 produces a π^+ in the final state via FSI. All generator codes include both nonresonant pion production and DIS processes. Although each code has different definitions of each process (see Sec. II), the sum is contained in the multi-π category. There is no ambiguity for $W >$ 2 GeV where true DIS models are most appropriate. We see approximately similar amounts of CC coherent at low Q^2 for all generators, which all use the Berger-Sehgal model. Importantly, the $1\pi^+$ via resonance contributions are suppressed at lower Q^2 for all the generators. Interestingly, NuWro and NEUT have similar $CC1\pi^+1p$ contributions, whereas GENIE's prediction is significantly lower, even though the generators are tuned to similar bubble chamber data. This highlights the need for consistent high-statistics neutrino-nucleon data. The strength of the $CC1\pi^+1n$ channel in the peak region is also different, and varies by a factor of 2 ranging from NuWro (smallest) to GENIE to NEUT. Coherent and DIS contributions are larger for the T2K signal definition than MINERvA, partially due to the $W_{rec} < 1.4$ GeV signal definition used by MINERvA which cuts ∼50% of coherent events below $Q^2 < 0.2$ GeV^2. The ratio of $CC1\pi^+$ from proton to $CC1\pi^+$ from neutron is very similar at both low- and higher-energy experiments.

Low-Q^2 suppression in single-pion production has seen some recent discussion in the community; NOvA introduced a low-Q^2 suppression to better match their charged-current resonant enhanced selections when using GENIE v2 [143], as did MINERvA and NUISANCE collaborators when tuning GENIE v2 to their CC1, $N\pi^{+,0}$ data [142]. When updating to GENIE v3, NOvA no longer requires such a suppression [144], and neither does T2K when using their updated single-pion model in NEUT. Our findings support these choices. All generators now include lepton mass effects in the Berger-Sehgal resonance model, and have updated the Δ form factors. It seems that these improvements to the nucleon-level interactions have improved the low-Q^2 discrepancies which were previously attributed to "nuclear effects." Additionally, all generators now also use an updated Berger-Sehgal coherent model, which MINERvA has found to better predict charged-current coherent data [145].

2. Model dependence via signal definition and signal selection differences

Here we investigate how each of the selection cuts applied to the true particles generated by G18_02a affects MINERvA's $CC1\pi^\pm$ $d\sigma/dQ^2$ cross section with
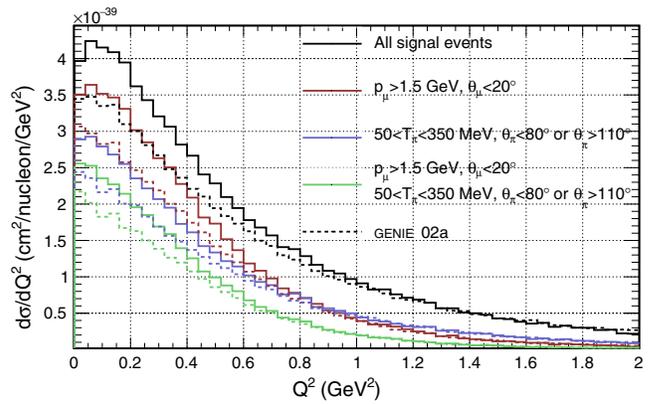
FIG. 32. MINERvA $CC1\pi^\pm$ cross section in Q^2 for NEUT 5.4.0.1 (solid) and G18_02a (dashed), looking at the effect of selection cuts on the generated events.

$W_{exp} < 1.4$ GeV and 1.5 GeV $< E_\nu^{true} < 10$ GeV. The selection requires a forward-going (back-exiting) muon that is well measured in MINOS, effectively limiting $p_\mu > 1.5$ GeV and $\theta_\mu < 20°$. It also requires the pion to be contained within MINERvA and be well reconstructed, limiting the pion to $50 < T_\pi < 350$ MeV and $\theta_\pi < 80°$ or $\theta_\pi > 110°$. The cross section is reported for all events with a single pion in the final state, although only single-pion events with specific muon and pion kinematics were actually observed. The simulation was used to correct the data for the unmeasured region of phase space, at the cost of introducing a dependence on the simulation used to make the correction. Technically, the cuts should be applied on the reconstructed candidates' kinematics present in the full experiment simulation, but this was not available. Interestingly, a small coherent sample passes the signal definition in Fig. 31. However, these events are gone once the selection cuts are applied (see Sec. VIII D). When inspecting interaction mode contributions to the cross section, there are no charged-current coherent events in G18_02a passing both the muon and pion kinematic cuts. This can be seen in the T_π by-mode contribution in Fig. (29) which shows no CC coherent contribution, whereas the Q^2 distribution in Fig. 31 shows a large contribution at low Q^2. Hence MINERvA's phase-space cuts in the signal selection removes the possibility to actually observe coherent events in the detector with these selection cuts, and the low-Q^2 coherent contribution enters purely through the underlying simulation.

This is investigated in more detail in Fig. 32, where the cut sequence and signal definition is applied using GENIE (dashed lines) and NEUT (solid lines). GENIE cuts events at all Q^2 but each cut has a different effect on the shape. With NEUT as the base model instead, the impact of the cuts changes which events are observed in MINERvA, especially when $Q^2 < 0.4$ GeV^2 where the physics content is not well known and the difference between GENIE and NEUT
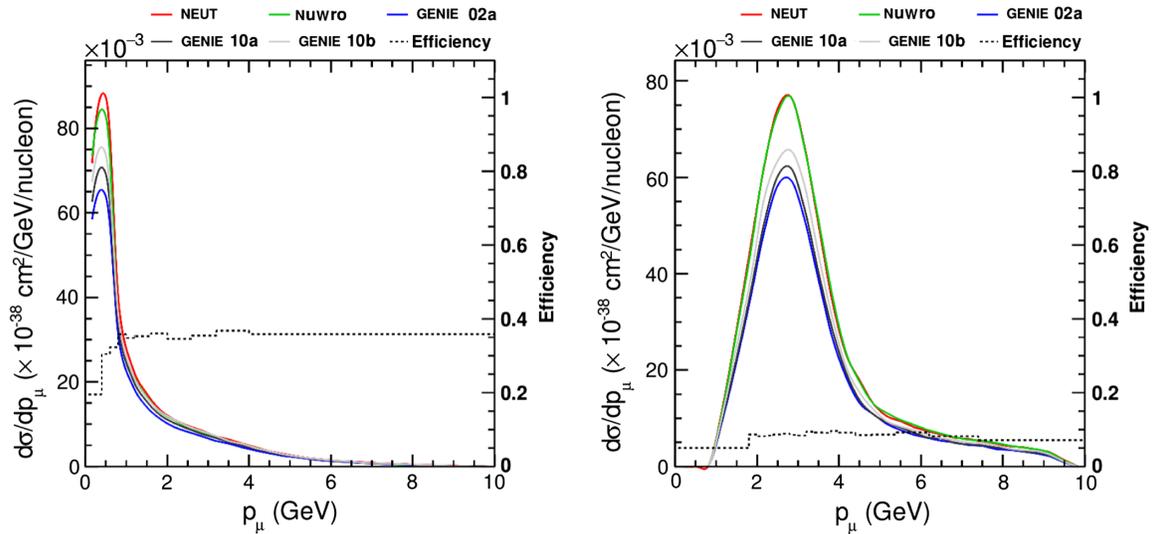
FIG. 33.   Muon momentum distributions for T2K (left) and MINERvA (right). The dotted line is the efficiency for each experiment using the right-axis scale and the solid lines are generator predictions for this variable using the left-axis scale.

is 25%. This implies that the extracted data may have been different if MINERvA had used NEUT as the model for the analysis, and therefore implies some degree of model bias. At the same time, different choices in nuclear model and binding energy can be involved. How significantly this affects the results is impossible to accurately quantify in this exercise.

3. Efficiency comparisons

This section shows the T2K CC1π^+ and MINERvA CC1$\pi^\pm$ efficiencies, calculated using NEUT v5.3.6 and GENIE v2.6.2, respectively, as a function of p_μ, $\cos\theta_\mu$, T_π, and $\cos\theta_\pi$, using the method described in Sec. IV to show the efficiencies as a function of a single variable.

The efficiency as a function of muon momentum is shown in Fig. 33 for T2K and MINERvA, overlaid with the respective model predictions from NEUT, NuWro, and GENIE. T2K has a flat muon momentum efficiency of ∼35% above 1 GeV, dropping to ∼30% at 200 MeV, and falling further to ∼20% at lower momentum. This was the basis for a cut on the muon momentum to avoid model dependence when extracting the result. Still, the majority of T2K events fall in the region of changing efficiency where models disagree at the 15% level. This could lead to model dependence in the efficiency correction. We note that T2K does not extract a result solely as a function of the muon momentum and this effect is likely to be largely mitigated by extracting the cross section in $p_\mu - \cos\theta_\mu$. In contrast to T2K, MINERvA 's muon momentum efficiency is relatively flat at approximately 9% across a broad momentum range but vanishes below 1.5 GeV. Below ∼2 GeV, the various generator models agree very closely. Here, the efficiency drops below 7%, averaged across a single bin. At the peak of the

distribution, 30% differences are observed and the agreement worsens with increasing momentum. The shape of the efficiency is largely caused by the MINOS track-matching requirements.

The efficiency as a function of muon angle is shown in Fig. 34. Coverage in this variable is restricted for both T2K and MINERvA. However, the efficiency is largest at forward angles where the cross section is largest. MINERvA 's MINOS track-matching requirements limits angular acceptance to less than $\cos\theta_\mu \sim 0.93$ (20°), although the signal definition is not restricted to this region. The range for T2K extends much further, up to $\cos\theta_\mu \sim 0.26$ (75°) for measurements that integrate over muon momenta (all forward angles are considered for the muon double-differential cross section). MINERvA falls off steadily from 12 to 5% over that range, while T2K falls off from 37 to 10%. In both cases the bulk of the events are forward going, somewhat mitigating the effects of the efficiency decline on measurements that integrate over muon angle. T2K published a two-dimensional distribution in $p_\mu - \theta_\mu$, in part to decrease the model dependence shown here.

The efficiency as a function of pion kinetic energy is shown in Fig. 35. T2K and MINERvA have reasonable efficiency down to low T_π. The acceptance region extends up to high values of T_π containing the full range of the cross-section model predictions, although the MINERvA efficiency falls with higher T_π whereas T2K's increases due to the two methods of tagging pions present in the analysis.

The variations in efficiency over the peak regions in T_π, coupled with the model disagreement in those regions, are troubling for measurements that integrate over the full pion phase space. For T2K, the efficiency drops from 25% down to 20%, and then rebounds up to 33% in the
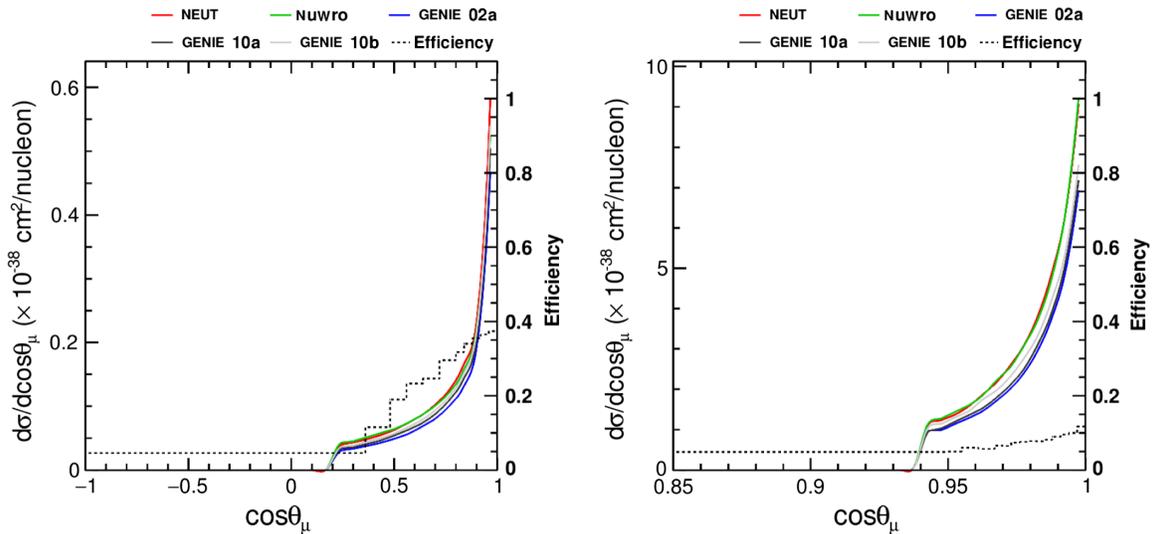
FIG. 34.   Muon angle for T2K (left) and MINERvA (right). Experimental efficiency is shown as a black dotted line using the right vertical scale and various generator predictions are shown using the left vertical axis.

$0 < T_\pi < 0.6$ GeV region where the vast majority of the cross section lies. The cross section drops rapidly and generator predictions vary here by up to 25%. The efficiency uncertainties are further complicated by nature of the event selection, which includes Michel tagged pions at the lower pion kinetic energies and tracked pions at higher energies. The details of the efficiency function shape over this region are sensitive to systematic errors in either of these two selections. However, the Michel tagged pion sample is only used in the muon and pion T_π measurements: all other measurements include a $p_\pi > 200$ MeV/c ($T_\pi > 104$ MeV) phase-space cut which intentionally eliminates most of the Michel sample's contributions. The issues with the increasing tracked-pion efficiency through 600 MeV is still problematic.

MINERvA suffers from similar issues, with the efficiency rising steadily from 6 to 14% in the region below the predicted cross-section peak near $T_\pi = 100$ MeV. At the peak the generator models differ in their predictions by up to 30%. Bins are large with respect to the change in efficiency on the high-T_π side of the peak as well, where the cross-section prediction continues to change quickly, and large model difference persists. In summary, there is potential for large model dependence across the full range of T_π, especially since no signal-definition cut is placed on the variable.

The efficiency as a function of the cosine of the pion angle is shown in Fig. 36. T2K's efficiency is flat at ~20% for backward and perpendicular pions and rises gradually to 35% for forward pions. MINERvA has essentially no acceptance of perpendicular pions because of the scintillator strip orientation. Nevertheless, the MINERvA CC1$\pi^\pm$ analysis shown here [14] covers the full range of pion angles. This may be partially mitigated by the smoothness
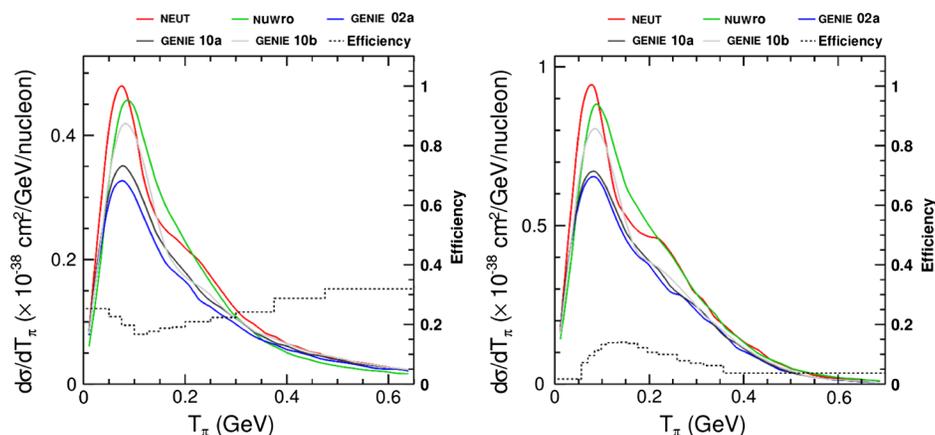
FIG. 35.   Pion kinetic energy for T2K (left) and MINERvA (right). Experimental efficiency is shown as a black dotted line using the right vertical scale and various generator prediction are shown using the left vertical axis.
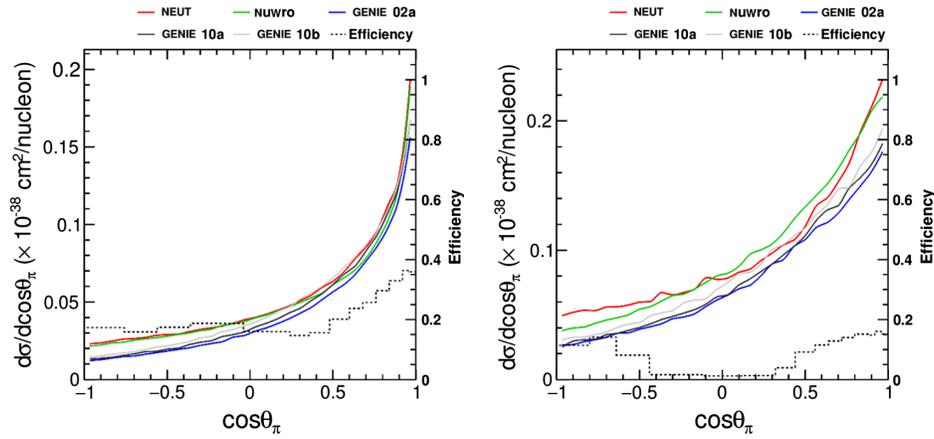
FIG. 36. $\cos\theta_\pi$ for T2K (left) and MINERvA (right). Experimental efficiency is shown as a black dotted line using the right vertical scale and various generator predictions are shown using the left vertical axis.

of the cross-section behavior in this region, and the relative agreement between the generators as shown, but it clearly risks model dependence, which is impossible to accurately assess after the analysis. In particular, the MINERvA θ_π bins which correspond to this low-efficiency region in the published cross-section measurement are not informative.

D. Discussion and summary

For this publication, the recent CC1π^+ data from T2K [133] and the updated CC1$\pi^\pm$ MINERvA results [139] were considered. Both measurements studied ν_μ-CH interactions and report inclusive distributions of muon and pion angle and kinetic energy/momentum and other quantities of interest. These measurements are at different neutrino energies and give the opportunity to examine energy dependence of the interaction.

One of the primary sources of tension at the 2016 workshop [12] came from comparison of the pion production results of MiniBooNE [13] and MINERvA [14]. The results had somewhat different signal definition (CC1π^+ vs CC1$\pi^\pm$, E_ν, and W cuts) and were at different beam energies. It was also noted that the lack of covariance matrices in the MiniBooNE data release severely complicated interpretation of the results. Although no direct comparison between datasets was possible due to differences in signal definition, there was clearly a more rapid evolution of the cross section in neutrino energy for the calculations than was observed in the data. Later calculations at the time [146] were unable to resolve the differences.

The first publication of pion production results from T2K [133] provides significant new information with a more modern treatment of uncertainties, at a similar neutrino energy to MiniBooNE. Attempts at comparing MINERvA and T2K measurements run into similar issues as MINERvA and MiniBooNE comparisons. Each measurement seeks a result that is objective and reproducible.

However, detector technology, geometry, and selection sculpt the particle acceptances differently, imparting potential model dependence to the data. Although the process of minimizing bias from detector effects or model dependence is still underway, an attempt was made to compare data at the same kinematics to decrease the differences. For the small overlapping phase space of MINERvA and T2K, we find good agreement, with NuWro and NEUT often describing T2K better, and GENIE performing better at MINERvA energies. We also note that the generators used to extract the cross section at each experiment is often the one with the best description of the data. This may be coincidence or come from biases towards the underlying MC program used in analysis.

In TENSIONS2016, an instance of model dependence related to the signal definition limitation on W in the original MINERvA analysis signal definition was identified and improved for the data release [139] result studied here. However, no additional changes in the phase-space restrictions on particle kinematics in the signal definition were made. Separately, all the models changed due to more complete fitting to neutrino H/D data [50,147]. The main effect is to make the change with beam energy smaller and all the models are now in better agreement with MINERvA data. Agreement of generators with both datasets is reasonable for pion kinematics for the quoted uncertainties. On the other hand, all generators have similar systematic problems describing the Adler angle variables presented by T2K.

The G18_02a, NuWro, and NEUT predictions for the single-pion production measurements are broken down by mode for T2K and MINERvA Q^2 and pion momentum/kinetic energy signal definitions in Figs. 30 and 29. Although T2K data extend to higher pion energies, both peak at the same energy within the resolutions provided as both are dominated by Δ production processes. This peak is produced by a balance of pion production and FSI processes. Thus, both effects must be considered.

Generator predictions have small-to-moderate deviations from the data based on the χ^2 values in Tables XI and XII. It is notable that NuWro has the lowest χ^2 for T2K and the largest for MINERvA. The deviations are not large and all calculations seem to have general agreement with the data. The calculations tend to have better agreement for lepton variables. For variables involving pion kinematics the differences between generators grows and the agreement with data is worse, with the largest data–MC differences being observed in the Adler angle distributions. This is also largely reflected in the comparisons to MINERvA data, where the $d\sigma/d\theta_\pi$ distribution is poorly described by all generators, followed by the $d\sigma/dp_\pi$ distribution.

It is difficult to make a quantitative assessment of the agreement of data and calculations with regard to energy dependence. The calculations have similar agreement with both T2K and MINERvA data according to χ^2. In addition, the average cross section of both datasets for overlapping kinematics (Sec. VIII C) have a ratio very similar to that of the calculations. The generators are also very similar in lepton kinematics, and differ mostly in hadron kinematics. These observations all give confidence that the energy dependence of the calculations matches that of the data within the accuracy of the data.

Beyond comparisons with calculations, a more detailed examination of the T2K and MINERvA analyses was made. Both measurements have restricted muon angle ranges, much more so for MINERvA. MINERvA also has a significant hole in the pion angle acceptance around 90°, and for higher-momentum pions. Low-efficiency regions are excluded from the T2K analysis, but corrected for using the nominal generator in the MINERvA analysis. The MINERvA data for $\theta_\pi \sim 90°$ and disagreements with simulations should be discounted. The MINERvA CC1$\pi^\pm$ measurement produces a full phase-space cross-section measurement in which the selection for observables $p_\mu, \cos\theta_\mu, T_\pi, \theta_\pi$ are selected to better match the detector's reconstruction capabilities. Model dependence from these choices is demonstrated. Although the analysis contains systematic errors to account for these effects, future analyzers should minimize the model dependence.

Efficiency studies uncover problems in regions where both the efficiency changes rapidly and model predictions from different generators vary. One example is for MINERvA $T_\pi \lesssim 100$ MeV, which applies to the lowest bin in Fig. 27. The other case is $p_\mu \lesssim 1$ GeV/c for T2K, which applies to the lowest two bins in Fig. 22, although the importance is less clear when the data are presented as a two-dimensional distribution. T2K's $p_\mu \cos\theta_\mu$ measurement uses both TPC and Michel tagged pions, leading to a complex T_π efficiency shape in a region where the cross-section predictions vary. However, this is not the case for the other T2K distributions, which require $p_\pi > 200$ MeV/c since the efficiency for TPC-only pions falls quickly at low momentum. As a result, the $p_\mu \cos\theta_\mu$ and other distributions have different model dependence in the low-momentum region.

The importance of the choice of signal definition and analysis methods was discussed in Sec. III. The pion production data available are all from early measurements and many of the cautions listed there apply to these data. The T2K analysis used a single iteration of d'Agostini unfolding and the MINERvA analysis used a signal definition with derived values. Both introduce model dependence that could be avoided. Both collaborations have more advanced analyses in progress, but were not available for this work.

A useful strategy is to use a multigenerator approach. The cross section is either extracted separately with different generators [148], or a bias test using a different generator is performed (e.g., Ref. [114]). An example would be to measure the efficiency with two generators making different choices or varying models within a specific generator to their extremes. A simpler alternative is to use an alternate generator or model set as fake data. However, these methods are necessarily limited to model dependencies based on the models implemented in generators.

The theoretical treatment of the pion production process is also important. It is complicated and clearly the current models implemented in generators are unable to explain all of the data, with pion kinematic variables a particular weakness. It is important to stress that these studies are only for $\Delta(1232)$ dominated interactions, and will be increasingly complicated at DUNE energies, due to higher resonances and soft inelastic scattering contributions, and a more complex ^{40}Ar target, amongst others. This motivates the need to improve our understanding by including better models in the event generators. The description of the basic pion production process with nucleon targets is limited by the quality of data [50,147]. Better accounting for existing electron inelastic scattering from the nucleon [149] would be a significant advance because vector resonance form-factor improvements since Ref. [34] have not been included in the event generators. Furthermore, much improved data for pion electroproduction using nuclear targets from the e4nu experiment [150,151] at JLab are expected. This will be important for testing pion production and FSI with high-quality data. Additionally, none of the common event generators include nuclear medium corrections to the Δ production operator, but various models are available [131].

IX. MINIBOONE DATASETS

MiniBooNE produced a series of key measurements: CCQE and CCQE-like [152], CC1π^+ [13], CC1π^0 [153], NCπ^0 [154], and NC elastic (NCEL) [155]. These datasets have had a long-lasting impact on the field of neutrino cross-section physics, as they pioneered important approaches in measuring neutrino cross sections with

intense, modern beams. MiniBooNE developed the presentation of double-differential results with high-statistics datasets, and a broad signal definition (only based on final-state properties which gives reduced model dependence). Additionally, they set standards regarding the information made available to describe and support each analysis in data releases that have now become a regular part of a modern neutrino cross-section publication. The results spawned extensive theoretical development of models to explain the data, and experimental efforts to assess uncertainties on models. However, limitations of MiniBooNE data have been uncovered with time. In some of the early MiniBooNE cross-section results, the off-diagonal correlations were not reported due to concerns about the influence of finite MC statistics in the generation of the detector model uncertainties. For example, the lack of a complete correlation error matrix for some of the measurements (CCQE-like, $CC1\pi^+$) prevents the ability to compare data to updated models in a statistically valid way. This is exacerbated by the unfolding method used, which leads to strongly regularized and therefore correlated values between bins. In another example, flux and background model improvements developed after the measurements were completed made some of the earlier analyses more difficult to interpret. Finally, the measurements themselves are being superseded by other experiments (e.g., T2K, MicroBooNE) at a similar neutrino energy regime. One major appeal of the MiniBooNE results was the uniformity of acceptance of the detector, but improvements to T2K reconstruction and detectors will overcome this limitation. In addition, the statistical precision of MINERvA has surpassed or is approaching that of MiniBooNE.

It is clear that the MiniBooNE NCEL and π^0 ($NC\pi^0$) production measurements are of great value because of large statistics, excellent efficiency uniformity, and up-to-date error analysis. NCEL is a forward-folded result, and they are published for neutrino and antineutrino enhanced running, each with a full covariance matrix. Both channels are important and cover reactions that are otherwise very poorly known. They should be employed with confidence for comparison with calculations and can be used with the same confidence as most other modern datasets. Since these results are still unique, we have chosen not to include them in the present work where the emphasis is on comparing different datasets.

Datasets such as CCQE-like and $CC1\pi^+$ should be used with care as the estimated uncertainties did not include the full correlation information. As these data are still of high quality, we urge the collaboration to reanalyze the events using more modern methods. For now, any comparison of calculation with theory should include at least one other data set if one of these MiniBooNE measurements is used; the Appendix includes plots of the MiniBooNE data which can be qualitatively

compared to the other datasets discussed in earlier sections of this document.

X. GENERATOR STUDIES

The opportunity to compare generator outputs from the samples produced for the data comparisons led to the following sections. These are studies that have applicability to the field as a whole.

A. Fraction of energy transfer imparted to neutral final-state particles

In the challenging task of neutrino energy estimation, high-quality modeling of the production of neutral particles is critical due to the relative difficulty of reconstructing these particles in a detector [9,20]. Large detector volumes are needed for good neutral particle reconstruction; unfortunately, this is not true for existing detectors and Monte Carlo programs are required to estimate neutral energy contributions. Significant differences in neutral particle production modeling among the predictions of GENIE, NEUT, and NuWro can exist and they are examined briefly in this section.

To facilitate comparisons between the generator models, we define the observable F to be the fraction of the total energy contained in the final-state hadrons (equal to the leptonic energy transfer q_0) which is associated with the production of neutral final-state particles:

$$F \equiv \frac{1}{q_0} \sum_j (E_j - \delta_j^B m_j). \tag{14}$$

Here the sum runs over all neutral particles in the final state, and E_j is the total energy of the jth neutral particle. The second term is subtracted to avoid counting the neutron mass m_n for baryons in the final state: the symbol δ_j^B is 1 if the jth particle is a baryon and zero otherwise. For definiteness, the following particle species (and their antiparticles) are included in the sum in Eq. (14): n, π^0, γ, D^0, Λ^0, Σ^0, K_L^0, K_S^0, η, and ω. The unstable members of the list decay into both charged and neutral particles. Deexcitation γ rays are not simulated (and thus do not contribute to F) for the results shown in this subsection.

The mean value of the neutral energy-transfer fraction F is shown as a function of neutrino energy in the upper left panel of Fig. 37. The simulated events used in the calculation, which are common to all panels of the figure, are for inclusive charged-current ν_μ scattering on ^{40}Ar at energies most relevant for MicroBooNE (up to $E_\nu = 5$ GeV). The upper right panel shows the generator predictions for the same distribution when at least one final-state neutral particle is required to be present. The bottom two panels repeat the same calculations using a modified definition of F in which only final-state neutrons contribute
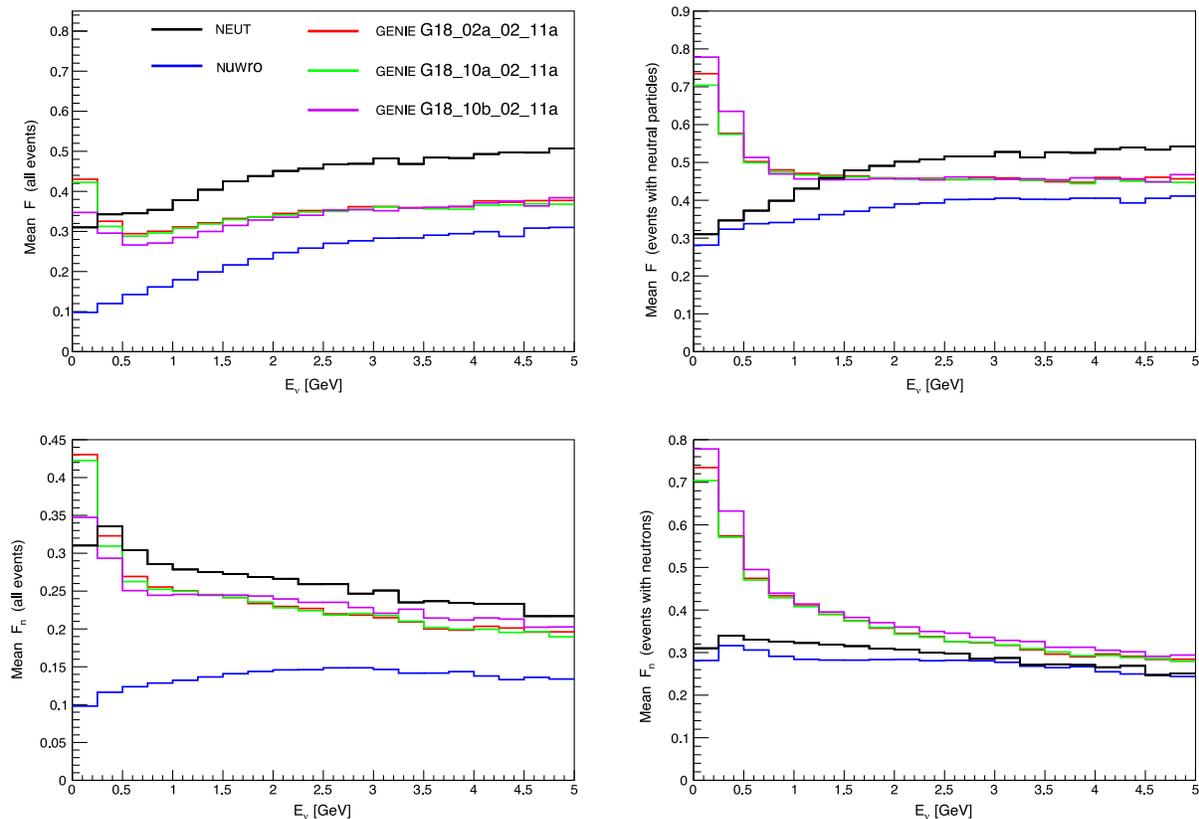
FIG. 37. Mean fraction of the leptonic energy transfer imparted to final-state neutral particle species. All panels show distributions calculated for charged-current ν_μ interactions on ^{40}Ar. Top left: Predictions including all neutral particles and all events in the sample. Top right: Predictions including all neutral particles for events containing at least one final-state neutral particle. Bottom left: Predictions including final-state neutrons only and all events in the sample. Bottom right: Predictions including final-state neutrons only for events containing at least one final-state neutron.

to the sum. This observable, referred to as F_n, is equal to the sum of final-state neutron kinetic energies divided by the energy transfer q_0.

Variations in the GENIE model set have a significant impact on these distributions only at low neutrino energies, and typically only when comparing the two FSI models available in v3.0.6: hA2018 (red and green histograms) and hN2018 (purple histogram). Disagreements between GENIE and the other two generators are more pronounced, with the spread in the predicted mean values of F reaching roughly a factor of 2 for an inclusive sample (upper left panel). For events containing at least one neutral particle (right-hand panels), a bifurcation between GENIE and NuWro/NEUT at low energies highlights an important physics difference: in both FSI models used by GENIE, an approximate treatment of compound nuclear decay (but not γ-ray emission) is included which leads to an enhancement of low-energy neutron emission. The similar shapes seen in multiple panels of Fig. 37 for the NEUT and NuWro distributions can be partially attributed to their similar approaches to FSI modeling (see Sec. II).

B. Energy threshold dependence of proton and pion event yields

Exclusive measurements with charged hadrons in the final state provide important insights into the physics of both the neutrino scattering and the nuclear environment. Most available data consider protons [114,129] and charged pions [14,133], both detectable through their ionization signatures. Each experiment makes choices consistent with their particle detection capabilities, including imposing a threshold for efficient tracking. MINERvA uses scintillator technology, with a threshold of ∼100 MeV (50 MeV) kinetic energy for protons (pions). T2K uses a combination of scintillator and TPC detectors, achieving a similar threshold to MINERvA. Liquid argon TPC detectors offer the promise of lower thresholds, with the 47-MeV proton threshold reported by MicroBooNE [114] the lowest available in a fully automated reconstruction at this time.

Generator models differ in their final-state hadron kinematics, leading to differences in efficiency for a fixed threshold. To assess the impact of these effects among the generators used in this work, we determine the fraction of events above threshold as a function of the threshold value
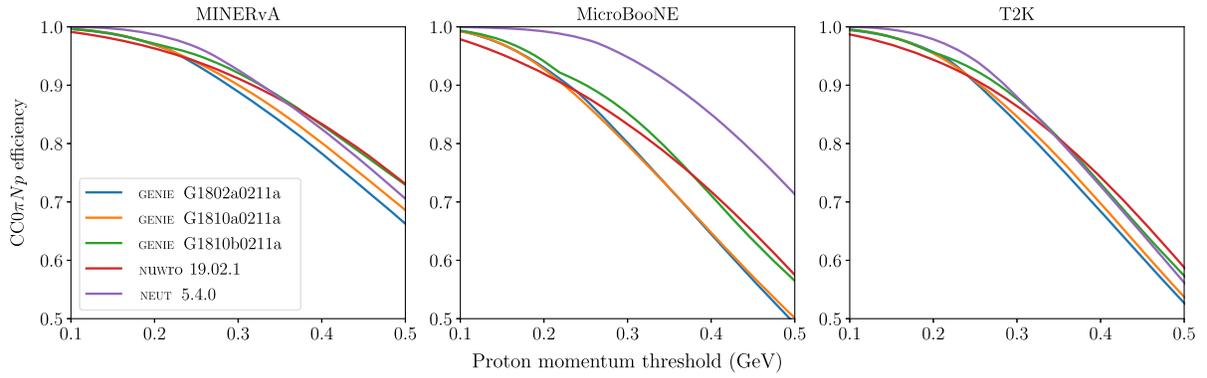
FIG. 38.   The fraction of true $CC0\pi Np$ events above threshold as a function of the true proton momentum threshold for different generators.

for specific final-state topologies in T2K, MINERvA, and MicroBooNE, using the appropriate flux and target material in each case.

Figure 38 shows the effect of threshold on efficiency for protons detected for $CC0\pi$ final states. With several potential sources of low-energy protons that are challenging to calculate, including varied assumptions in FSI models, this is an especially interesting issue. With lower-energy neutrinos, MicroBooNE and T2K are expected to be particularly sensitive to the threshold. While the codes generally show the same trends, some large discrepancies are evident. NEUT results are substantially different from the others due to its suppression of low-energy protons. At 300 MeV/c proton momentum (47 MeV kinetic energy) in MicroBooNE, the fraction of protons seen for NEUT is 13% larger than the average of the others. G18_02a and G18_10a, which use an effective cascade FSI model, have about 5% smaller fraction than NuWro or G18_10b, which use a full cascade model. Experiments must carefully consider the effects of threshold choice.

Figure 39 shows the effect of threshold for pions detected in $CC0\pi^{0}N\pi^{\pm}$ final states. It is notable that the largest deviations for protons are at low energy where the yield is

growing rapidly. For positively charged pions, the cross section is decreasing close to threshold and the largest deviations visible here come at larger energies. Since pions have lower ionization yields than protons at a given momentum, all experimental thresholds in use are in the region where there is very little dependence on generator. Effects of threshold are therefore relatively less important for charged pions than protons.

XI. METHODOLOGICAL LIMITATIONS OF THE RECENT FORWARD-FOLDING CROSS-SECTION RESULTS PRESENTED BY THE MICROBOONE EXPERIMENT

Measured observables are always convoluted with detector effects. As described in Sec. III, often, neutrino cross sections are presented after a deconvolution of these effects. Such measurements are usually called "unfolded.". Unfolded measurements correct for smearing and other detector effects in order to get to the true underlying distributions. If the observed neutrino interaction events in a certain observable bin i is N_i, the cross section in true bin j is usually calculated as
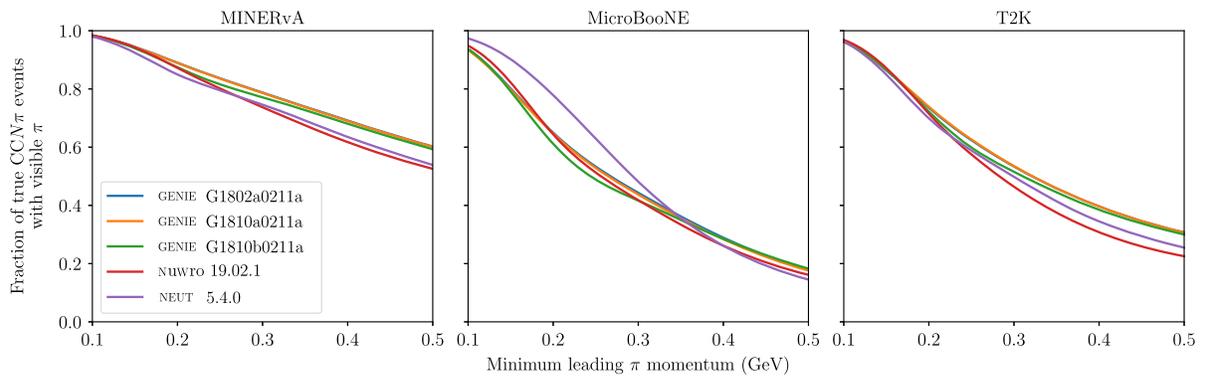
FIG. 39.   The efficiency for true $CCN\pi^{\pm}$ events as a function of the true leading $\pi^{\pm}$ momentum threshold, as compared among different generators.
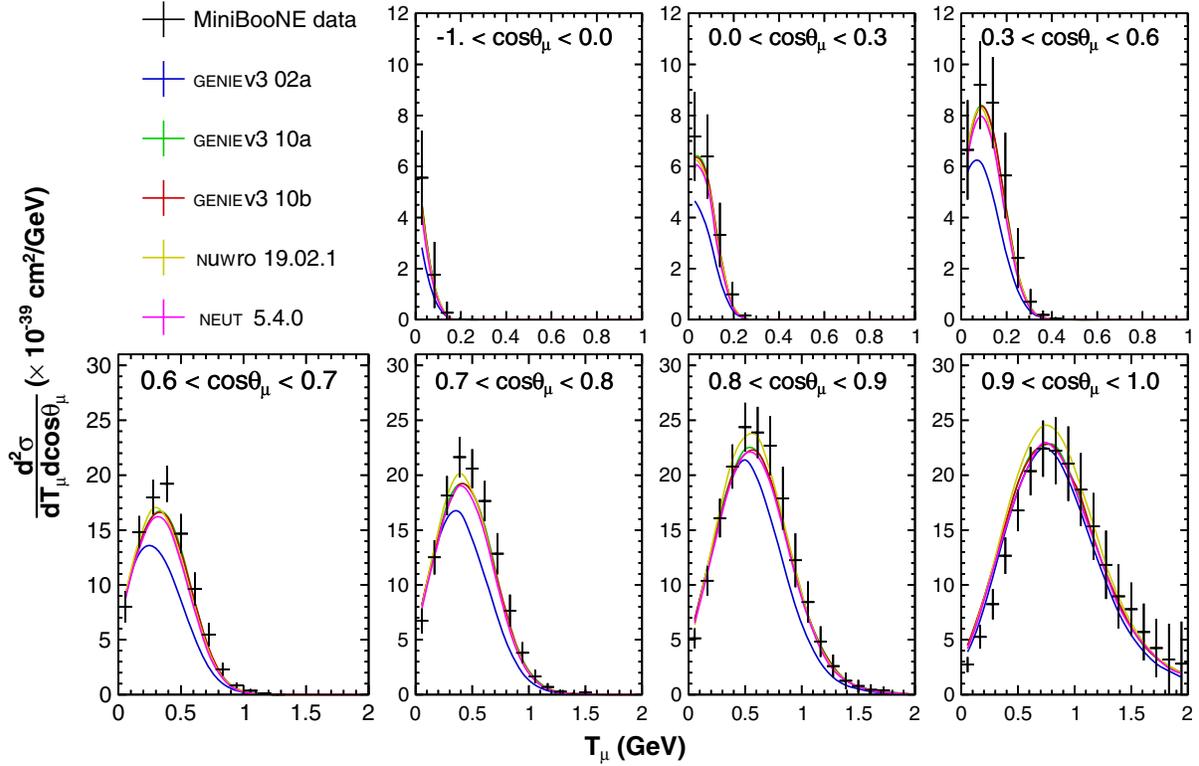
FIG. 40. Comparison of MiniBooNE CCQE-like T_μ, $\cos\theta_\mu$ data [152] with GENIE, NEUT, and NuWro calculations. Data are determined from the CCQE sample assuming true CCQE kinematics. Cross sections displayed come from the MiniBooNE data release [156] for Ref. [152] by adding the pion absorption correction back into the published CCQE values. Each bin in the array is a θ_μ angular distribution for the range in μ^- kinetic energy in the legend.The shown uncertainties are the shape-only uncertainties. There is an additional 10.7% normalization uncertainty.

$$\left(\frac{d\sigma}{dx}\right)_j = \sum_i \frac{U_{ji}(N_i - B_i)}{\Phi T \epsilon_j (\Delta x)_j}, \qquad (15)$$

where B_i is the number of background events in the observable bin, Φ is the total flux, T number of target nucleons, ϵ_j is the efficiency in truth bin j, Δx its width (or area), and U_{ji} is the unsmearing matrix. Unfolded measurements can be easily compared with theory predictions, and with unfolded results from other experiments. The difficult part is to calculate the unsmearing matrix, and there exist different approaches to do it.

Unfortunately, unfolding is an "ill-posed problem" in which small statistical fluctuations can lead to large variations in the unfolded spectrum [with generally strongly (anti)correlated unfolded data points]. Also, they usually make some sort of assumption about the linearity of the statistical uncertainties (i.e., things are treated like normally distributed errors). Another way of presenting cross-section results is by using the "forward-folding" approach, in which the data are not deconvolved, but instead are compared with "smeared" (i.e., forward-folded) theoretical predictions. Forward-folding measurements are currently being investigated with increased interest as they sidestep some issues of the unfolding methods and can thus

be preferable from a statistical point of view (see, e.g., [90]). To make these measurements comparable to other predictions, the method to smear the theoretical prediction has to be provided alongside with the measurement.

The MicroBooNE experiment has recently used the following method for their double-differential muon-neutrino charged-current inclusive cross-section measurement on argon [113] and their measurement with N protons and no pions [114]. In the MicroBooNE method, a background-subtracted, efficiency-corrected cross-section measurement in reconstructed space is extracted from the measured data:

$$\sigma_i = \left(\frac{d^2\sigma}{dp_\mu d\cos\theta_\mu}\right)_i = \frac{N_i - B_i}{\tilde{\epsilon}_i T \Phi_{\nu_\mu} (\Delta p_\mu \Delta \cos\theta_\mu)_i}, \qquad (16)$$

where N_i and B_i are the number of selected data events and the expected number of background events in reconstructed bin i, respectively. $(\Delta p_\mu \cdot \Delta\cos\theta_\mu)_i$ is ith bin area. T and Φ_{ν_μ} are the number of target nucleons and the integrated muon-neutrino flux, respectively. $\tilde{\epsilon}_i$ is the average event selection efficiency:

$$\tilde{\epsilon}_i = \frac{\sum_j S_{ij} N_j^{\text{sel}}}{\sum_j S_{ij} N_j^{\text{gen}}}, \tag{17}$$

where N_j^{sel} and N_j^{gen} are the number of selected and generated events in truth bin j. S_{ij} is the smearing matrix describing the migration of events between kinematic bins, $S_{ij} = P(\text{measured in bin } i | \text{generated in bin } j)$.

A covariance matrix E describes the statistical and systematic uncertainties of the result. The systematic part is generated by varying the input parameters of the simulation according to their respective uncertainties. These variations can affect every single variable in Eqs. (16) and (17) except for the number of selected events, the bin area, and the number of target nucleons.

To compare some model cross-section predictions σ^{model} with the data, they are folded through the *nominal* smearing matrix S. Then, the Mahalanobis distance d (i.e., chi square) between the predicted cross section and the measured data points is calculated using the covariance matrix:

$$d^2 = \sum_{i,k} \left[\left(\sigma_i - \sum_j S_{ij} \sigma_j^{\text{model}} \right) E_{ik}^{-1} \left(\sigma_k - \sum_l S_{kl} \sigma_l^{\text{model}} \right) \right]. \tag{18}$$

If the measured data are a random variation of the model prediction, d^2 should be χ^2 distributed.

The idea behind providing the covariance matrix E is that new generator predictions can be tested against the data using Eq. (18). This requires all systematic uncertainties to be encoded in the covariance matrix. Unfortunately, it seems like this does not hold true in the general case. If the systematic uncertainties on the smearing itself are not negligible, the formalism described in (16) may underestimate these systematics.

This can be easily shown when considering the extreme case of completely flat efficiencies (the ideal case in most scenarios). When the selection efficiency ϵ is flat over all truth bins j, Eq. (17) becomes a constant expression:

$$\tilde{\epsilon}_i = \frac{\sum_j S_{ij} N_j^{\text{sel}}}{\sum_j S_{ij} N_j^{\text{gen}}} = \frac{\sum_j S_{ij} N_j^{\text{gen}} \epsilon}{\sum_j S_{ij} N_j^{\text{gen}}} = \epsilon. \tag{19}$$

No amount of variation in the smearing matrix will change this number, so in this case, the contribution of the smearing uncertainty to the covariance matrix is exactly 0. For the measurement described in Ref. [113], performance studies done by MicroBooNE [157] show that the effect of the smearing uncertainty is small compared to all other systematic uncertainties. The interpretation of model comparisons using the covariance matrix E are thus "correct," i.e., the partially ignored uncertainties have only a small effect.

This shows though that this method is not necessarily generalizable to an experiment in which the uncertainty on S plays a bigger role. To fully include the smearing uncertainty in the model comparisons, it must be taken into account when folding the truth-space model prediction to the reconstructed space. One possibility of doing this is by using a set of response matrices, varied according to the detector uncertainties, as shown in Ref. [158].

XII. CONCLUSION

The field of neutrino interaction physics has expanded significantly in recent years, with numerous new measurements and calculations, and improved techniques. Importantly, analysis improvements now facilitate clearer connections among experimental results, and between those results and generator predictions. One notable shift is that nearly all experiments now use signal definitions expressed in terms of direct kinematic observables, e.g., final-state lepton momentum and angle, and often make kinematic restrictions to better match experimental capabilities. This alignment to the topology of experimental measurements has helped to decrease the level of model dependence in extracted cross sections. However, no direct comparison between results of different experiments is possible at this time. Detailing these issues is the primary purpose of this paper.

Comparison of the results here with those in the first Tensions 2016 paper [12] shows the evolution. At that time, both MiniBooNE and MINERvA used broad signal definitions and corrected for kinematic regions with low efficiency. This has the advantage of being closer to theoretical results. Since then, the dangers of model dependence have become much more apparent. Nevertheless, some of the measurements here (particularly the MINERvA 1π measurements) are done in the earlier style. We use them as test cases showing the results of choices made in analysis.

Event generators continue to play a key role in comparing theory with measurements. Although model implementations necessarily lag behind most recent theory advances and certain assumptions are required to produce a composite model with complete phase-space coverage, generator simulations can readily reproduce a broad array of signal definitions due to the availability of all variables in the calculation. Recent efforts have led to dramatic enhancements in the event generators, including a wider array of available models, inclusion of collective nuclear effects and processes involving nucleon-nucleon correlations, and improved tuning to data. These improvements are a result of a growing body of theoretical work, and strong collaboration among theorists and generator developers working toward implementation. We note that while generators tend to make similar advances, e.g., Valencia models [26] for QE and $2p2h$ interactions and Salcedo-Oset [38] for pion FSI, full calculations to match

experimental conditions involve many elements, and we find that agreement among calculations is good but not exact. The studies presented here include up-to-date configurations of the GENIE, NuWro, and NEUT generators, as well as a GENIE configuration using older model sets, as a point of comparison to earlier work. Notably, most of the more modern calculations were not in the original publications of the experimental measurements, and are presented here for the first time.

The Tensions workshops have attempted to assess the overall status in the field of neutrino interactions in recent years. The first Tensions publication [12] was based on results available in 2016, focusing primarily on the final results from MiniBooNE and early T2K and MINERvA results. Comparisons made with generators available at the time revealed discrepancies among the experimental data, and highlighted the role of model dependence in cross-section measurements. The results presented in this paper summarize the work conducted and initiated during a second workshop in 2019. In addition to an updated perspective including more recent and more detailed measurements alongside new and substantially improved event generators, we have presented new, detailed cross-experiment comparisons, studied key model differences among generators, and explored the limitations of forward-folding methods. The comparisons in this document include a novel comparison of each interaction type in terms of the model dependence of the efficiency.

Since the results are primarily for CH targets, the A dependence cannot be studied in a detailed way. However, the CH data from T2K ($\langle E_\nu \rangle \sim 1$ GeV) and MINERvA ($\langle E_\nu \rangle \sim 3.5$ GeV) allow examination of the energy dependence in a limited way. In some cases, bins with identical final state kinematics in each experiment were examined. The following subsection summarizes those findings.

A. Comparisons by interaction type

First, we consider recent measurements of charged-current inclusive cross sections from T2K, MicroBooNE, and MINERvA, which provide a general view of neutrino interactions. The measurements show clearly that the modeling of the lepton kinematics works much better than the modeling of the hadronic part of the interactions. The T2K and MicroBooNE measurements using lepton kinematics yield χ^2 per degree of freedom of order 2, while MINERvA's measurement including hadronic information is of order 10. The large χ^2 values are likely an effect from strong correlations, and must be the subject of a future work because it is a complicated issue. Although clear differences exist in the relative performance of the generators among experiments, the poor fits overall make it difficult to draw definitive conclusions on particular sources of disagreement; all p values are well below 0.01.

The different inclusive datasets seem to give contrasting conclusions when compared with various generators. For

T2K, NEUT does a better job describing these data, and NuWro predicts a smaller cross section in the forward region of the lepton compared to the other generators there. It also consistently predicts a larger peak at intermediate muon momenta, while these differences are not vis NuWro. For MicroBooNE, GENIE predictions have similar values of χ^2 for all three models, while the value from NEUT is the highest. All generators seem to have equal understanding of A dependence in general. For MINERvA measurements, NuWro gives the lowest χ^2 compared with other predictions. However, all event generators show disagreement at low available energy and the peak of the distributions for the regions between $2p2h$ and RES.

Concerning the ν_μ CC-0π cross section, comparisons have been made between T2K and MINERvA results and the generator predictions employed for this work. While it is difficult to draw an unambiguous conclusion, it is possible to underline some considerations. As described in Sec. VII D, T2K and MINERvA cover different $q_0 - q_3$ regions and in order to directly compare results from the two experiments it is necessary to carefully select a common phase space. As reported in Sec. VII E, the integrated ν_μ CC-0π cross sections in a phase space common to T2K and MINERvA ($p_\mu > 1.5$ GeV/c and $\theta_\mu < 20°$) are systematically higher than what is predicted by generators under study (see Fig. 16). The only exception is for NEUT that seems to perfectly reproduce the MINERvA integrated cross section. The other generator results tend to be $\sim 20\%$ below the data, indicating that the energy dependence is better reproduced than the individual measurements.

For T2K, the systematic MC underestimation of the cross section is true also when considering the full muon momentum phase space in the same angular region. However, the detailed two-dimensional measurements shown in Figs. 10–12 reveal specific behaviors of the generators depending on the kinematic bins considered, as well as different degrees of agreement/disagreement with the data. In general, T2K carbon data seem to prefer NEUT and GENIE using LFG, while MINERvA data do not show agreement with considered generators. Because of the different $q_0 - q_3$ accessible to T2K and MINERvA, the energy dependence of cross section is difficult to determine. Both experiments indicate an underprediction in the forward lepton direction, for T2K is for high-momentum (>1 GeV) bins, but the additional pion contribution via FSI at higher beam energies obscures the robustness of the predicted energy dependence.

In a similar way, the different T2K and MINERvA TKI variables prefer different generators as shown in Fig. 17. Both T2K and MINERvA data show poor χ^2 agreement with respect to NuWro. Generators suffer in reproducing the low δp_T region for MINERvA, suggesting limits in the prediction of the initial-state nucleons. On the other side, MINERvA $\delta \alpha_T$ is well reproduced by the generators other

than NuWro, suggesting a more correct treatment of FSI at energies relevant for MINERvA. T2K δp_T shows lower χ^2 values when compared to GENIE and NEUT with LFG, while GENIE with the G18_02a tuning better reproduces $\delta\alpha_T$, which can be interpreted as a preference for GENIE 's traditional effective cascade FSI model. Future measurements with additional statistics as well as further comparisons between T2K and MINERvA data will certainly help in clarifying this complex picture. Finally, generator comparisons with MINERvA measurements in Q^2_{QE} (see Sec. VII H) underline the relevance of a correct FSI treatment for nuclei with increasing A. The comparisons suggest a preference for GENIE's traditional effective cascade FSI model (tuning G18_10a and G18_02a).

The pion production interaction remains a central concern in the field at the same time as its importance grows through the needs of DUNE [7]. It was the source of many issues with signal definition and data compatibility in TENSIONS2016 [12]. Although MINERvA has published a variety of measurements, the $1\pi^\pm$ data are of particular interest here since they can be compared to T2K. While T2K has added measurements and MINERvA has updated and expanded upon their first measurements, data remain insufficient to fully constrain models due to large uncertainties and limitations in data coverage. At the same time, improvements in modeling this interaction remain slow to implement.

The focus in this work is on a detailed comparison of the recent T2K CC1π^+ [133] and the updated MINERvA CC1$\pi^\pm$ [139] (which is heavily dominated by π^+) cross sections. Both T2K and MINERvA measurements stress inclusive measurements; this makes them sensitive to underlying efficiency evaluations. However, generator predictions show that both are largely sensitive to π^+ production from bound protons. The efficiency study (Sec. VIII C 3) show holes in both T2K and MINERvA acceptance and the likelihood of resulting increased model dependence. The overlap in kinematics is small but encouraging. Within the estimated uncertainties, the measurements are in agreement, indicating that the energy dependence is handled correctly (in direct contrast to the MiniBooNE-MINERvA comparisons of TENSIONS2016 [12]). This provides a more detailed test than straight comparison with theory.

Both T2K and MINERvA present numerous distributions that probe different parts of a complicated interaction. Both qualitative and quantitative comparisons are described; see Tables XI and XII for details. The muon distributions (Figs. 22 and 26) are largely insensitive to the details of FSI and show features of the treatment of the initial state. All generators do a reasonable job of describing these inclusive distributions with χ^2 typically 1–2/bin. Pion observables such as kinetic energy/momentum and polar angle (Figs. 24 and 27) are more sensitive to the production mechanism and FSI. Features of the data in

these observables are described poorly in general and values of χ^2 are typically larger than 2 per bin. This is especially true for MINERvA θ_π, but it should be noted that MINERvA θ_π shows some issues with model dependence in our studies. The pion momentum distribution appears most sensitive to modeling. Although the magnitude of various calculations is close to correct, the values of χ^2 are large, indicating a problem with shape. The most direct test of FSI is a comparison of GENIE hA and hN models with all other features held constant. The more empirical hA model describes the data better than the more theoretical hN model.

Finally, the Adler angle distributions (angles of the decay pion in the parent resonance rest frame) for the T2K measurement show details of the $\Delta(1232)$ decay. This required a very detailed analysis by the experiment and shows the largest discrepancies between data and generator predictions. Many issues could be involved. The nuclear modeling in the generators could be wrong. There is incompatibility between older analyses for H/D targets [55] and the newer data involving both nuclear corrections and the treatment of resonance/nonresonance components.

There is an interesting trade-off between MINERvA and T2K 1π^+ measurements. Both have the goal of providing accurate data that stand on their own with minimal model dependence and allow theory calculations to be compared directly with the data. T2K puts more emphasis on decreasing the model dependence by restricting the phase space of the final-state particles. MINERvA does this also, but to a smaller extent. For example, MINERvA uses a $W < 1.4$ GeV constraint in the signal definition which allows more direct studies of the Δ resonance but creates significant potential for model dependence because events from the higher resonances become background.

MiniBooNE published a large body of neutrino cross-section data about 10 years ago. Although it has excellent statistics and broad kinematic coverage, some of those results have significant model dependence and incomplete systematic uncertainty analysis as compared to the more modern treatments. In particular, the lack of published correlation matrices for some MiniBooNE datasets presents a major difficulty. Quantitative model comparisons based on goodness of fit to these data should not be attempted unless great care is taken to address the statistical implications of the missing bin-to-bin correlations [159]. We feel that a reanalysis of this important dataset should be undertaken using modern methods. Provisional qualitative conclusions based on model comparisons to the MiniBooNE CC0π and CC1π measurements should be confirmed with more rigorous testing against modern measurements. The MiniBooNE NC elastic and NC1π^0 results are not subject to the same caveats: both provide high-quality measurements with full covariance matrices of otherwise poorly known cross sections. Models may be

compared to the NCEL and NC1π^0 data with the same level of confidence as typical modern measurements.

Only qualitative analysis of the MiniBooNE results was attempted in the Appendix. All modern calculations are in good visual agreement with the MiniBooNE CC0π data considered in this publication with the most notable disagreement at very low Q^2. For CC1π, conclusions are harder to ascertain because the calculations are not in agreement with each other. The most notable problem is the significant normalization disagreement for NEUT, NuWro, and the older GENIE version with respect to the CC1π^+ MiniBooNE data. Better agreement of calculations with the MiniBooNE CC1π^0 data may indicate an issue with CC1π^+ data.

A series of separate generator studies compared features of the models independent of comparison with data. The first considered the dependence of yield as a function of momentum threshold, and showed the sensitivity to choices the experiments make. Both proton and pion momentum yields vary significantly among the generators at the values typically chosen, indicating significant model sensitivity. The second study was of the neutral energy content which is an important contribution of the generators to all experiments. The variations among the generators are significant, both for the total neutral energy and for that contained in neutrons. Any experimental measurements in variables of this type are difficult but will be extremely valuable for constraining this energy which varies widely among generators.

The community continues to investigate and improve statistical methods to present cross-section measurements in mathematically correct and useful ways. Disentangling the effects on the detector response from model-dependent nuisance parameters and from the actual parameters of interest remains a challenge. The work within this group explored limitations of the forward-folding implementation employed by MicroBooNE in its first cross-section results [113]. In this case, the method breaks down when the uncertainties of the detector smearing become large enough to have a notable effect on the result. Other forward-folding strategies without this limitation exist, but have so far not been used in a published cross-section measurement, and introduce other challenges and complexities. Statistical methods that have been used by other experiments, or even by the same experiment in the past, might not be sufficient for future experiments and ever more complicated analyses. Additional effort from experiments to explore these techniques is essential [158].

B. Looking forward

We conclude with a general outlook on the field. Progress since TENSIONS2016 [12] is impressive. The number and quality of experimental results for neutrino-nucleus cross sections has increased significantly and is still improving. As a result, our studies expanded into inclusive interactions. There have been a large number of mesonless (i.e., pionless at low neutrino energies such in T2K) quasielastic scattering results, both for neutrinos and antineutrinos. The statistics of MINERvA results are approaching and in some cases exceeding that of MiniBooNE (the previous standard) and have the ability to differentiate calculations better. The additional detection of protons in the final state has allowed more focus on true QE interactions. For pion production, MINERvA has published a variety of results. The body of pion production results has expanded in the last few years with MINERvA results for all pion charges. However, measurements are still limited in statistics and have potential problems with model dependence of various sources. Of course, the pion production measurements remain challenging because of the difficulties identifying and measuring the energies of pions in tracking detectors or calorimeters.

The measurements covered in this work are largely for CH targets because of the preponderance of scintillator target/detectors at the time. More recent results from MINERvA stress high atomic weight nuclei and MicroBooNE provides a liquid argon (LAr) target/detector. Heavier nuclear targets and more accurate measurements with hadrons are clearly the frontier. Recent MicroBooNE results [114,160] have a proton tracking threshold of 47 MeV and liquid argon detectors should be able to go lower. This is especially interesting because the yield of protons and neutrons is expected to rise at low energy due to compound nucleus processes. DUNE [7] will focus on LAr target/detectors. Pion detection remains largely unexplored for LAr experiments because of the low fraction of events with a Michel electron. The T2K near-detector upgrade [161] will emphasize back-angle particle detection and oxygen targets in preparation for upcoming measurements and Hyper-Kamiokande. These are all very positive steps forward toward improved measurements and this document will hopefully aid that development.

ACKNOWLEDGMENTS

APPENDIX: ADDITIONAL MINIBOONE QUALITATIVE COMPARISONS

Given the concerns raised in Sec. IX we provide comparisons to MiniBooNE data in a limited way for this paper. We note specifically that a lack of correlations may lead to incorrect conclusions being drawn from the data when a naive bin-by-bin comparison is made. But, there is no established procedure to overcome this shortcoming for the currently available data. Only qualitative comparisons are provided, and a full interpretation should rely on other datasets presented (e.g., T2K, MINERvA, MicroBooNE).

Here we discuss for this specific study CCQE-like and CC1π^+ results. The first measurement, CCQE-like, included only the final-state muon in the signal definition. The CC1π^+ measurement introduced a novel way to identify positively charged pions in a Cerenkov detector through their inelastic scattering signature. Comparisons of the CCQE-like Q^2 distribution, the CC1π^+ pion kinetic energy, and the CC1π^0 pion momentum are shown in Figs. 41, 42, and 43, respectively.

While NEUT, G18_10a, and G18_10b are in qualitative agreement with the CCQE-like Q^2 distribution, GENIE G18_02 has problems in describing the shape properly. GENIE G18_02 is significantly low at $Q^2 > 0.3$ GeV^2 and all calculations are above the data point at the lowest Q^2. The corresponding plot for the MINERvA CCmeson-less measurement is shown in Fig 19 for a higher-energy neutrino beam. As a result, the Q^2 range is larger. In addition, the statistical advantage at very low Q^2 is clear.

FIG. 42. Comparison of MiniBooNE T_{π^+} data [13] with GENIE, NEUT, and NuWro calculations.

The agreement of the calculations with the MINERvA data is improved, but the problems at very low Q^2 persist. A broader view is obtained with Fig. 40 where the full range of $p_\mu - \cos\theta_\mu$ can be seen. This has the advantage of separating the components of the Q^2 calculation within the same sample. All modern calculations are in good qualitative agreement with these data with the only problems seen for the older GENIE version at back angles and low momentum.

One of the central issues in TENSIONS2016 [12] was the discrepancy in the T_π normalization and shape between MINERvA [14] and MiniBooNE [13] data. As discussed in Ref. [162], the normalization in the calculation is largely driven by the elementary $\nu_\mu p \to \mu^- \pi^+ p$ cross section and that makes a disagreement in normalization hard to understand from physics considerations.

The calculations for CC1π^+ in Fig. 42 have a wide variation in magnitude and shape. Figure 27 shows the

FIG. 41. Comparison of MiniBooNE Q^2_{QE} data with GENIE, NEUT, and NuWro calculations. The shown uncertainties are the shape-only uncertainties. There is an additional 10.7% normalization uncertainty.
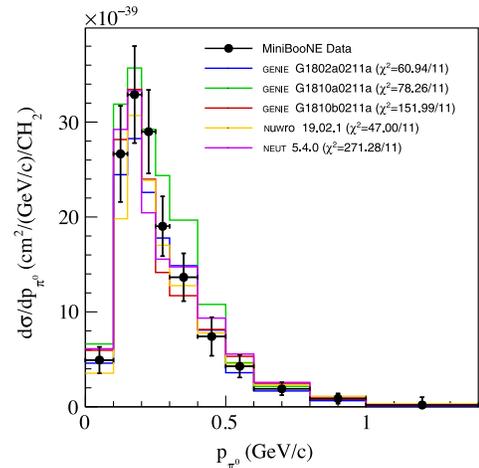
FIG. 43. Comparison of MiniBooNE CC1π^0 p_{π^0} data [153] with GENIE, NEUT, and NuWro calculations.

corresponding result for MINERvA (similar target and higher energy) and Fig. 23 for T2K (similar energy and target). It is interesting to note that NuWro and NEUT tend to overpredict the T2K data but underpredict the MiniBooNE data. The same tendency is seen with the GENIE calculations, especially G18_10a. Since T2K and MiniBooNE have very similar target and energy distribution, the conclusion from this comparison is that the two experiments are inconsistent. A proper error treatment would be required to make a more quantitative assessment. On the other hand, NEUT and NuWro tend to overpredict MINERvA data and the GENIE calculations tend to be smaller than the data. In contrast to the first comparisons [12], GENIE v3 G18_10a calculation is in agreement with both MiniBooNE and MINERvA. Additional information is provided by the MiniBooNE CC1π^0 measurement [153]. The π^0 momentum distribution is shown in Fig. 43; here, all calculations are in good qualitative agreement with the data, implying differences in treatment of π^+ and π^0 in some calculations or in the data. In detail, the χ^2/N_{bins} values are all large.

[1] K. Abe *et al.* (T2K Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **659,** 106 (2011).

[2] D. S. Ayres *et al.* (NOvA Collaboration), 10.2172/935497.

[3] M. Antonello *et al.* (MicroBooNE, LAr1-ND, ICARUS-WA104 Collaborations), arXiv:1503.01520.

[4] Z. Maki, M. Nakagawa, and S. Sakata, Prog. Theor. Phys. **28,** 870 (1962).

[5] S. M. Bilenky and B. Pontecorvo, Phys. Rep. **41,** 225 (1978).

[6] B. Abi *et al.* (DUNE Collaboration), Eur. Phys. J. C **80,** 978 (2020).

[7] A. Abed Abud *et al.*, Instruments **5,** 31 (2021).

[8] B. Abi *et al.* (DUNE Collaboration), Eur. Phys. J. C **81,** 322 (2021).

[9] B. Abi *et al.* (DUNE Collaboration), arXiv:2002.03005.

[10] K. Abe *et al.*, arXiv:1109.3262.

[11] L. Alvarez-Ruso *et al.* (NuSTEC Collaboration), Prog. Part. Nucl. Phys. **100,** 1 (2018).

[12] M. Betancourt *et al.*, Phys. Rep. **773–774,** 1 (2018).

[13] A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **83,** 052007 (2011).

[14] B. Eberly *et al.* (MINERvA Collaboration), Phys. Rev. D **92,** 092008 (2015).

[15] C. Andreopoulos *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **614,** 87 (2010).

[16] T. Golan, C. Juszczak, and J. T. Sobczyk, Phys. Rev. C **86,** 015505 (2012).

[17] Y. Hayato, Acta Phys. Pol. B **40,** 2477 (2009).

[18] P. Stowell *et al.*, J. Instrum. **12,** P01016 (2017).

[19] K. Abe *et al.* (T2K Collaboration), Phys. Rev. Lett. **121,** 171802 (2018).

[20] M. A. Acero *et al.* (NOvA Collaboration), Phys. Rev. Lett. **123,** 151803 (2019).

[21] Y. Hayato and L. Pickering, Eur. Phys. J. Special Topics **230,** 4469 (2021).

[22] R. A. Smith and E. J. Moniz, Nucl. Phys. **B43,** 605 (1972); **B101,** 547(E) (1975).

[23] A. Bodek and J. L. Ritchie, Phys. Rev. D **23,** 1070 (1981).

[24] C. H. Llewellyn Smith, Phys. Rep. **3,** 261 (1972).

[25] J. Nieves, J. E. Amaro, and M. Valverde, Phys. Rev. C **70,** 055503 (2004); **72,** 019902(E) (2005).

[26] J. Nieves, I. R. Simo, and M. J. V. Vacas, Phys. Rev. C **83,** 045501 (2011).

[27] J. Nieves *et al.*, Nucl. Part. Phys. Proc. **273–275,** 1830 (2016).

[28] R. Gran, J. Nieves, F. Sanchez, and M. J. Vicente Vacas, Phys. Rev. D **88,** 113007 (2013).

[29] T. Katori, AIP Conf. Proc. **1663,** 030001 (2015).

[30] J. T. Sobczyk, Phys. Rev. C **86,** 015504 (2012).

[31] A. Lovato, S. Gandolfi, J. Carlson, S. C. Pieper, and R. Schiavilla, Phys. Rev. Lett. **112,** 182502 (2014).

[32] J. L. Barrow, S. Gardiner, S. Pastore, M. Betancourt, and J. Carlson, Phys. Rev. D **103,** 052001 (2021).

[33] K. M. Graczyk and J. T. Sobczyk, Eur. Phys. J. C **31,** 177 (2003).

[34] D. Rein and L. M. Sehgal, Ann. Phys. (N.Y.) **133,** 79 (1981).

[35] R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3,** 2706 (1971).

[36] C. Berger and L. Sehgal, Phys. Rev. D **76,** 113004 (2007).

[37] T. Yang, C. Andreopoulos, H. Gallagher, K. Hoffmann, and P. Kehayias, Eur. Phys. J. C **63,** 1 (2009).

[38] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio, Nucl. Phys. **A484,** 557 (1988).

[39] A. Bodek and U. Yang, AIP Conf. Proc. **670,** 110 (2003).

[40] K. M. Graczyk, D. Kielczewska, P. Przewlocki, and J. T. Sobczyk, Phys. Rev. D **80,** 093001 (2009).

[41] P. A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020,** 083C01 (2020).

[42] C. Berger and L. M. Sehgal, Phys. Rev. D **79,** 053003 (2009).

[43] D. Rein and L. M. Sehgal, Nucl. Phys. **B223,** 29 (1983).

[44] T. Sjostrand, S. Mrenna, and P. Z. Skands, J. High Energy Phys. 05 (2006) 026.

[45] C. Juszczak, J. A. Nowak, and J. T. Sobczyk, Nucl. Phys. B, Proc. Suppl. **159,** 211 (2006).

[46] J. Tena-Vidal, C. Andreopoulos, C. Barry, S. Dennis, S. Dytman, H. Gallagher, S. Gardiner, W. Giele, R. Hatcher, O. Hen, I. D. Kakorin, K. S. Kuzmin, A. Meregaglia, V. A. Naumov, A. Papadopoulou, M. Roda, V. Syrotenko, and J. Wolcott (GENIE Collaboration), Phys. Rev. D **105,** 012009 (2022).

[47] V. R. Pandharipande and S. C. Pieper, Phys. Rev. C **45,** 791 (1992).

[48] H. Gallagher, Nucl. Phys. B, Proc. Suppl. **112,** 188 (2002).

[49] L. Alvarez-Ruso *et al.* (GENIE Collaboration), Eur. Phys. J. Special Topics **230**, 4449 (2021).

[50] J. Tena-Vidal *et al.* (GENIE Collaboration), arXiv:2104 .09179.

[51] J. Schwehr, D. Cherdack, and R. Gran, arXiv:1601.02038.

[52] A. Bodek, S. Avvakumov, R. Bradford, and H. S. Budd, Eur. Phys. J. C **53**, 349 (2008).

[53] D. Rein and L. M. Sehgal, Phys. Lett. B **657**, 207 (2007).

[54] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **83**, 052007 (2011).

[55] G. M. Radecky *et al.*, Phys. Rev. D **25**, 1161 (1982); **26**, 3297(E) (1982).

[56] A. Bodek and U. K. Yang, AIP Conf. Proc. **670**, 110 (2003).

[57] S. Dytman, Y. Hayato, R. Raboanary, J. T. Sobczyk, J. Tena-Vidal, and N. Vololoniaina, Phys. Rev. D **104**, 053006 (2021).

[58] Said partial-wave analysis facility, http://gwdac.phys.gwu .edu/.

[59] Z. Koba, H. B. Nielsen, and P. Olesen, Nucl. Phys. **B43**, 125 (1972).

[60] A. Bodek and U.-k. Yang, AIP Conf. Proc. **792**, 257 (2005).

[61] E. Oset and L. Salcedo, Nucl. Phys. **A468**, 631 (1987).

[62] E. S. Pinzon Guerra *et al.* (DUET Collaboration), Phys. Rev. C **95**, 045203 (2017).

[63] R. Seki and K. Masutani, Phys. Rev. C **27**, 2799 (1983).

[64] H. W. Bertini, Phys. Rev. C **6**, 631 (1972).

[65] R. Bradford, A. Bodek, H. Budd, and J. Arrington, Nucl. Phys. B, Proc. Suppl. **159**, 127 (2006).

[66] O. Benhar, A. Fabrocini, S. Fantoni, and I. Sick, Nucl. Phys. **A579**, 493 (1994).

[67] A. M. Ankowski and J. T. Sobczyk, Phys. Rev. C **74**, 054316 (2006).

[68] C. Juszczak, J. A. Nowak, and J. T. Sobczyk, Eur. Phys. J. C **39**, 195 (2005).

[69] S. J. Barish *et al.*, Phys. Rev. D **19**, 2521 (1979).

[70] J. T. Sobczyk and J. Zmuda, Phys. Rev. C **87**, 065503 (2013).

[71] A. Bodek and U. K. Yang, Nucl. Phys. B, Proc. Suppl. **112**, 70 (2002).

[72] J. A. Nowak, Phys. Scr. **T127**, 70 (2006).

[73] J. A. Nowak and J. T. Sobczyk, Acta Phys. Pol. B **37**, 2371 (2006).

[74] A. Bodek, H. Budd, and M. Christy, Eur. Phys. J. C **71**, 1726 (2011).

[75] J. Marteau, J. Delorme, and M. Ericson, Nucl. Instrum. Methods Phys. Res., Sect. A **451**, 76 (2000).

[76] J. T. Sobczyk, arXiv:nucl-th/0307047.

[77] M. Martini, M. Ericson, G. Chanfray, and J. Marteau, Phys. Rev. C **80**, 065501 (2009).

[78] G. Megias, J. Amaro, M. Barbaro, J. Caballero, T. Donnelly, and I. Ruiz Simo, Phys. Rev. D **94**, 093004 (2016).

[79] K. Niewczas and J. T. Sobczyk, Phys. Rev. C **100**, 015505 (2019).

[80] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio, Nucl. Phys. **A484**, 557 (1988).

[81] D. Klakow, G. Welke, and W. Bauer, Phys. Rev. C **48**, 1982 (1993).

[82] Vmc calculations of two-nucleon densities, https://www .phy.anl.gov/theory/research/density2/.

[83] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Rev. Mod. Phys. **87**, 1067 (2015).

[84] K. Mahn, C. Marshall, and C. Wilkinson, Annu. Rev. Nucl. Part. Sci. **68**, 105 (2018).

[85] *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, edited by H. B. Prosper and L. Lyons (CERN, Geneva, 2011).

[86] M. Kuusela, Statistical issues in unfolding methods for high energy physics, Master's thesis, Aalto University, 2012.

[87] M. J. Kuusela, Uncertainty quantification in unfolding elementary particle spectra at the Large Hadron Collider, Ph.D. thesis, SB, Lausanne, 2016.

[88] G. D'Agostini, Nucl. Instrum. Methods Phys. Res., Sect. A **362**, 487 (1995).

[89] G. D'Agostini, arXiv:1010.0632.

[90] R. D. Cousins *et al.*, arXiv:1607.07038.

[91] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **101**, 112001 (2020).

[92] K. Abe *et al.* (T2K, UK Collaborations), Nucl. Instrum. Methods Phys. Res., Sect. A **659**, 106 (2011).

[93] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **87**, 012001 (2013); **87**, 019902(A) (2013).

[94] K. Suzuki, Nucl. Phys. B, Proc. Suppl. **229–232**, 453 (2012), neutrino 2010.

[95] K. Matsuoka *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **624**, 591 (2010).

[96] K. Abe *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **694**, 211 (2012).

[97] S. Assylbekov *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **686**, 48 (2012).

[98] P.-A. Amaudruz *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **696**, 1 (2012).

[99] N. Abgrall *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **637**, 25 (2011).

[100] D. Allan *et al.*, J. Instrum. **8**, P10019 (2013).

[101] S. Aoki *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **698**, 135 (2013).

[102] J. Allison *et al.*, IEEE Trans. Nucl. Sci. **53**, 270 (2006).

[103] L. Aliaga *et al.* (MINERvA Collaboration), Phys. Rev. D **94**, 092005 (2016); **95**, 039903(E) (2017).

[104] E. L. Valencia *et al.* (MINERvA Collaboration), Phys. Rev. D **100**, 092001 (2019).

[105] D. Ruterbories *et al.* (MINERvA Collaboration), Phys. Rev. D **104**, 092007 (2021).

[106] J. Park *et al.* (MINERvA Collaboration), Phys. Rev. D **93**, 112007 (2016).

[107] L. Aliaga *et al.* (MINERvA Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **743**, 130 (2014).

[108] D. Michael *et al.* (MINOS Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **596**, 190 (2008).

[109] M. G. Catanesi *et al.*, Eur. Phys. J. C **52**, 29 (2007).

[110] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, and T. Thouw, Forschungszentrum Karlsruhe Report FZKA, Report No. FZKA-6019, 1998.

[111] R. Pordes and E. Snider, Proc. Sci., ICHEP2016 (**2016**) 182.

[112] K. Abe, J. Amey, C. Andreopoulos, L. Anthony, M. Antonova, S. Aoki, A. Ariga, Y. Ashida, Y. Azuma, S. Ban *et al.*, Phys. Rev. D **98**, 012004 (2018).

[113] P. Abratenko *et al.*, Phys. Rev. Lett. **123**, 131801 (2019).

[114] P. Abratenko *et al.* (MicroBooNE Collaboration), Phys. Rev. D **102**, 112013 (2020).

[115] MicroBooNE Collaboration, Report No. MICROBOONE-NOTE-1045-PUB, 2018, https://microboone.fnal.gov/public-notes/.

[116] P. Abratenko *et al.*, J. Instrum. **12**, P10010 (2017).

[117] P. Rodrigues *et al.* (MINERvA Collaboration), Phys. Rev. Lett. **116**, 071802 (2016); **121**, 209902(E) (2018).

[118] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **103**, 112008 (2021).

[119] R. Gran, M. Betancourt *et al.* (MINERvA Collaboration), Phys. Rev. Lett. **120**, 221805 (2018).

[120] M. V. Ascencio *et al.* (MINERvA Collaboration), arXiv:2110.13372.

[121] R. Gran, arXiv:1705.02932.

[122] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **101**, 112004 (2020).

[123] D. Ruterbories *et al.* (MINERvA Collaboration), Phys. Rev. D **99**, 012004 (2019).

[124] T. Adye, in *PHYSTAT 2011* (CERN, Geneva, 2011), pp. 313–318,

[125] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **98**, 032003 (2018).

[126] X. Lu *et al.* (MINERvA Collaboration), Phys. Rev. Lett. **121**, 022504 (2018).

[127] T. Cai *et al.* (MINERvA Collaboration), Phys. Rev. D **101**, 092001 (2020).

[128] X. G. Lu, L. Pickering, S. Dolan, G. Barr, D. Coplowe, Y. Uchida, D. Wark, M. O. Wascko, A. Weber, and T. Yuan, Phys. Rev. C **94**, 015503 (2016).

[129] M. Betancourt *et al.* (MINERvA Collaboration), Phys. Rev. Lett. **119**, 082001 (2017).

[130] K. Hanson, T. Kawano, P. Talou, AIP Conf. Proc. **769**, 304 (2005).

[131] R. Freedman, G. Miller, and E. Henley, Nucl. Phys. **A389**, 457 (1982).

[132] E. Pinzon Guerra *et al.*, Phys. Rev. D **99**, 052007 (2019).

[133] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **101**, 012007 (2020).

[134] F. Sánchez, Phys. Rev. D **93**, 093015 (2016).

[135] R. Castillo Fernández, Measurement of the muon neutrino charged current interactions and the muon neutrino single pion cross section on CH using the T2K near detector, Ph.D. thesis, School Barcelona, Autonoma U., 2015.

[136] C. L. McGivern *et al.* (MINERvA Collaboration), Phys. Rev. D **94**, 052005 (2016).

[137] O. Altinok *et al.* (MINERvA Collaboration), Phys. Rev. D **96**, 072003 (2017).

[138] T. Le *et al.* (MINERvA Collaboration), Phys. Rev. D **100**, 052008 (2019).

[139] MINERvA Collaboration, http://minerva.fnal.gov/data-release-page/.

[140] J. Tena-Vidal *et al.* (GENIE Collaboration), Phys. Rev. D **104**, 072009 (2021).

[141] C. Wilkinson *et al.*, Phys. Rev. D **93**, 072010 (2016).

[142] P. Stowell *et al.* (MINERvA Collaboration), Phys. Rev. D **100**, 072005 (2019).

[143] M. A. Acero *et al.* (NOvA, R. Group Collaborations), Eur. Phys. J. C **80**, 1119 (2020).

[144] A. Himmel, New oscillation results from the nova experiment (2020), 10.5281/zenodo.4142045.

[145] A. Mislivec *et al.* (MINERvA Collaboration), Phys. Rev. D **97**, 032014 (2018).

[146] J. T. Sobczyk and J. Żmuda, Phys. Rev. C **87**, 065503 (2013).

[147] P. Rodrigues, C. Wilkinson, and K. McFarland, Eur. Phys. J. C **76**, 474 (2016).

[148] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **93**, 112012 (2016).

[149] L. Tiator, D. Drechsel, S. S. Kamalov, and M. Vanderhaeghen, Eur. Phys. J. Special Topics **198**, 141 (2011).

[150] A. Papadopoulou *et al.* (Electrons for Neutrinos Collaboration), Phys. Rev. D **103**, 113003 (2021).

[151] M. Khachatryan *et al.* (CLAS, e4v Collaborations), Nature (London) **599**, 565 (2021).

[152] A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **81**, 092005 (2010).

[153] A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **83**, 052009 (2011).

[154] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **81**, 013005 (2010).

[155] A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **82**, 092005 (2010).

[156] R. Tayloe, Miniboone ccqe data release, https://rtayloe.pages.iu.edu/MB/data-releases/ccqe/.

[157] MicroBooNE Collaboration, Report No. MICROBOONE-NOTE-1049-PUB, 2018, 10.2172/1573058.

[158] L. Koch, J. Instrum. **14**, P09013 (2019).

[159] L. Koch, Phys. Rev. D **103**, 113008 (2021).

[160] P. Abratenko *et al.* (MicroBooNE Collaboration), Phys. Rev. Lett. **125**, 201803 (2020).

[161] D. Sgalaberna (T2K Collaboration), Proc. Sci., ICHEP2020 (**2021**) 175.

[162] J. T. Sobczyk and J. Zmuda, Phys. Rev. C **91**, 045501 (2015).