



A novel approach to neutrino–hydrogen measurements

H. Duyang, B. Guo, S. R. Mishra, R. Petti^a

Department of Physics and Astronomy, University of South Carolina, Columbia, SC 29208, USA

Received: 17 August 2024 / Accepted: 27 October 2024
© The Author(s) 2024

Abstract The limited statistics of the available (anti)neutrino–hydrogen (H) interactions has been a longstanding impediment for high-energy neutrino physics. We discuss a practical way to achieve accurate (anti)neutrino–hydrogen measurements, addressing the principal limitations of earlier experiments. Interactions on hydrogen are extracted by subtracting measurements on thin dedicated graphite (pure C) and polypropylene (CH₂) targets within a highly segmented low-density detector. A kinematic selection is used to increase the purities to 80–95% before subtraction. A statistics of $\mathcal{O}(10^6)$ can be realistically achieved in modern neutrino beams for the various $\nu(\bar{\nu})$ -H event topologies. The availability of such samples would allow a precise determination of neutrino and antineutrino fluxes, as well as to directly constrain nuclear effects from a comparison with corresponding measurements on heavy materials within the same detector. The (anti)neutrino fluxes and the nuclear smearing are typically the leading sources of systematic uncertainties in long-baseline oscillation experiments. (Anti)neutrino–hydrogen interactions also provide an ideal tool for a wide range of precision tests of fundamental interactions.

1 Introduction

Neutrino experiments typically use massive nuclear targets, which are particularly relevant in long-baseline oscillation experiments given the reduced flux at the far sites [1, 2]. An understanding of the structure and interactions of hadrons within the nuclear targets is, therefore, crucial to achieve accurate measurements of neutrino interactions. The existing uncertainties in the modeling of the nuclear effects, which modify the neutrino cross sections, as well as the final state interactions within the nucleus are insufficient for the precisions required by next-generation neutrino experiments [3].

Modern (anti)neutrino beams can deliver intense fluxes allowing the use of high-resolution detectors with a relatively modest fiducial mass of a few tons to obtain a more accurate reconstruction of (anti)neutrino interactions. However, to exploit the physics potential of such developments for precision measurements of fundamental interactions and searches for new physics beyond the Standard Model, neutrino detectors have to address a number of issues. In addition to a high experimental resolution and a large acceptance, they must achieve an accurate calibration of the energy scales, as well as a control of the configuration, chemical composition, and mass of the neutrino targets comparable to electron scattering experiments [4]. The fact that the energy of the projectile (anti)neutrino is unknown on an event-by-event basis still represents an intrinsic limitation, as neutrino detectors have to infer the (anti)neutrino energy from the reconstructed final state particles emerging from the nucleus, which are affected by a substantial nuclear smearing and related systematic uncertainties. The difficulties above are illustrated by the many outstanding discrepancies among different experiments, and with existing theoretical models [3]. We could therefore argue that the availability of a hydrogen target—the only hadron target of known energy—is necessary to go beyond the precision level of existing neutrino scattering experiments [4, 5].

The available data from $\nu(\bar{\nu})$ -H interactions were collected by the early bubble chamber experiments ANL 12-foot [6], BNL 7-foot [7], FNAL E31 [8] and E45 [9], CERN WA21 [10], and WA24 [11]. In spite of the excellent experimental resolution of those measurements, the overall statistics is limited to a total of about 16k ν -H and 9k $\bar{\nu}$ -H CC interactions. Since then safety requirements related to the underground operation and practical considerations have prevented new measurements of high-statistics $\nu(\bar{\nu})$ -H samples.

In this paper, we discuss a novel approach for precision measurements of $\nu(\bar{\nu})$ -H charged current (CC) interactions, which is both safe and inexpensive to implement [4, 5, 12]. Interactions on hydrogen are extracted by subtracting measurements on dedicated graphite (pure C) targets from those on CH₂ plastic targets, integrated within a low-density high-resolution detector. The concept appears to be a viable and realistic alternative to liquid H₂ detectors.

^a e-mail: Roberto.Petti@cern.ch (corresponding author)

This paper is organized as follows. In Sect. 2, we discuss the detection technique and the basic concept of the subtraction between CH_2 and C targets. In Sect. 3, we present a detailed kinematic selection of $\nu(\bar{\nu})$ -H interactions from the CH_2 plastic target for various event topologies, and in Sect. 4, we discuss our results.

2 Detection technique

A detector technology designed to offer a control of the configuration, chemical composition, and mass of the neutrino targets similar to electron scattering experiments is a Straw Tube Tracker (STT), in which the targets are physically separated from the actual tracking system [13]. The target mass is distributed within a relatively large volume ($\sim 40 \text{ m}^3$) with the average density similar to that of liquid deuterium $\rho \leq 0.17 \text{ g/cm}^3$ and all dimensions comparable to one radiation length, to achieve an accurate reconstruction of the four-momenta of the visible final state particles, as well as of the event kinematics in a plane transverse to the beam direction. A large number (70–100) of thin planes—each typically 1–2% of radiation length X_0 —of various passive materials with comparable thickness are alternated and dispersed throughout active layers—made of four straw planes—of negligible mass in order to guarantee the same acceptance to final state particles produced in (anti)neutrino interactions. The STT allows to minimize the thickness of individual active layers and to approximate the ideal case of a pure target detector, as the targets constitute about 97% of the total mass [4, 5, 12]. The lightness of the tracking straws and the chemical purity of the targets, together with the physical spacing among the individual target planes, make the vertex resolution ($\ll 1 \text{ mm}$) less critical in associating (anti)neutrino interactions to the correct target material. Each target plane can be removed or replaced with different materials during data taking, providing a flexible target configuration.

The detector considered here is based upon a central STT inserted in a 0.6-T magnetic field for the measurement of charged-particle momenta, and surrounded by a 4π electromagnetic calorimeter (ECAL) [14] for the detection of photons and neutral hadrons including neutrons. The base tracking technology is provided by low-mass straws similar to the ones used in many modern experiments for precision physics or the search for rare processes [15–20].

In this paper, we will focus on the “solid” hydrogen technique, in which $\nu(\bar{\nu})$ interactions on free protons are obtained by subtracting measurements on dedicated graphite (C) targets from those on polypropylene (CH_2) targets [4, 5]. The latter are based on one of the plastic materials with the highest hydrogen content (14.4% by mass) and can be easily manufactured in thin foils of arbitrary size. Both C and CH_2 targets must have comparable thickness in terms of radiation and nuclear interaction lengths, and are alternated throughout the entire tracking volume to ensure that they result in a difference between their detector acceptances for final state particles within 10^{-3} [12]. Given the chemical purity achievable for the thin passive targets ($\sim 100\%$) and the accuracy in associating the interactions to each target, the normalization of the H signal and the C background is based upon the relative abundances in the CH_2 compound. The technique is conceived to be model-independent, as the data from the graphite targets automatically include all types of processes, as well as detector effects, relevant for the selection of interactions on H. Furthermore, it can be safely implemented to obtain a relatively large (~ 0.7 ton with 5 tons of CH_2) fiducial mass of hydrogen. In the following, we will present detailed studies of the corresponding event selection with realistic assumptions for the detector smearing and the physics modeling (Sect. 3.1).

3 Selection of $\nu(\bar{\nu})$ -H interactions

The subtraction technique described in Sect. 2 can be used to select any inclusive and exclusive process in both CC and neutral current (NC) $\nu(\bar{\nu})$ interactions on free protons [5]. For CC interactions, we can improve the signal/background ratio in the selection of H interactions by exploiting the event kinematics. Since the H target is at rest, CC events are expected to be perfectly balanced in a plane transverse to the beam direction (up to the tiny beam divergence) and the muon and hadron vectors are back-to-back in the same plane. Instead, events from nuclear targets are affected by the smearing with the energy–momentum distribution of bound nucleons (Fermi motion and binding), the off-shell modifications, meson exchange currents, and nuclear shadowing [3, 21–23], as well as by final state interactions (FSI) [3]. These nuclear effects result in a significant missing transverse momentum and a smearing of the transverse plane kinematics. The use of transverse plane variables and event kinematics to select various neutral current (NC) and CC (anti)neutrino topologies was pioneered by the NOMAD experiment [24–27]. The analysis described in this paper is largely based upon the variables and techniques developed and validated with NOMAD data [25–27].

3.1 Analysis framework

We simulate (anti)neutrino interactions on CH_2 , H, and C targets with three different event generators: NuWro [28], GiBUU [29], and GENIE [30]. While we use NuWro as our default generator, we compare our results with both GiBUU and GENIE to check the sensitivity of our analysis to the details of the input modeling. These generators are based upon different assumptions and modeling of nuclear effects and final state interactions, with GiBUU using a conceptually different approach from heavy ion physics, based upon the Boltzmann-Uehling-Uhlenbeck equation for the particle propagation through the nuclear medium. For a detailed review

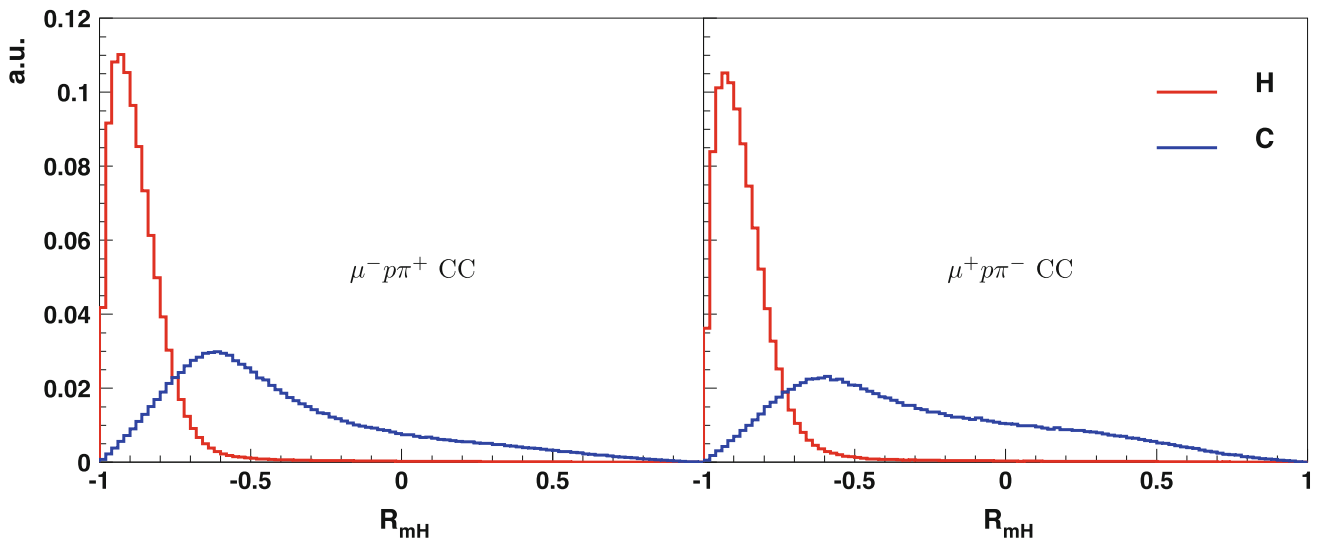


Fig. 1 Comparison of the reconstructed p_T asymmetry $R_{mH} = (p_T^m - p_T^H)/(p_T^m + p_T^H)$ in $\nu_\mu p \rightarrow \mu^- p \pi^+$ and $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$ processes on H (signal) and in the corresponding background topologies from the C nucleus. Results for both neutrino (left panel) and antineutrino (right panel) interactions are shown. All distributions are normalized to unit area

and comparison of the generators used, we refer to Ref. [31]. The studies of nuclear FSI by the MINER ν A and T2K experiments [32, 33] have found an unphysical excess of hA elastic scattering processes in the FSI simulated by GENIE, which is in disagreement with (anti)neutrino data. We follow the corresponding prescriptions by MINER ν A and T2K and discard such elastic hA FSI processes in GENIE. We generate inclusive CC interactions including all processes available in the event generators—quasi-elastic (QE), $\Delta(1232)$ and higher resonances (RES), nonresonant processes and deep inelastic scattering (DIS)—with default settings and the input (anti)neutrino spectra expected at the future Long-Baseline Neutrino Facility (LBNF) [1, 34, 35].

We use the GEANT4 [36] simulation package to evaluate detector effects and apply to all final state particles a parameterized reconstruction smearing consistent with the NOMAD data [24, 37]. The detector smearing has also been independently checked with the FLUKA [38] simulation program. The acceptance for individual final state particles (p , n , π^\pm , π^0 , μ) takes into account the detector geometry, the event topology, and the material traversed by the particles and is folded into the analysis. We emphasize that the STT detector has been explicitly designed to offer the same acceptance for particles produced in both the CH_2 and graphite targets, as discussed in Sect. 2.

For charged tracks, the average momentum resolution is about 5% and the angular resolution about 2 mrad. The reconstruction of protons is typically worse due to the shorter track length, with average momentum resolution of 6.5% for H events and 8.4% for C events. The difference between H and C events is related to the different momentum and angular distributions introduced by nuclear effects. The vertex resolution can vary from $100 \mu\text{m}$ to about $600 \mu\text{m}$ for multi-track events, depending upon the geometry and the event topology [24, 37]. As discussed in Sect. 2, this parameter is not critical in STT, as even for events with a single reconstructed charged track (Sect. 3.2.2) for which the vertex cannot be reconstructed, the uncertainty in associating the event to the correct target material is given by the ratio between the thickness of the straw walls ($< 20 \mu\text{m}$) and the thickness of a single target layer, typically below 0.5%. Particle identification is provided by various methods including dE/dx , range, and transition radiation in STT, as well as the energy deposition, timing, and track segments in the surrounding ECAL and muon identifier.

We analyze ν_μ and $\bar{\nu}_\mu$ CC interactions originated from the CH_2 and graphite targets described in Sect. 2. We determine the momentum vectors of charged particles from the track curvature in the B field, while for neutral particles we use either the energy deposited in the various sub-detectors or, whenever available, the secondary charged tracks originated by the neutral particles in STT. The momentum vector of the total hadron system, \vec{p}_H , is obtained from the sum of the momenta of all the reconstructed final state hadrons [25].

In the following section, we discuss a unified approach to the kinematic selection of all the event topologies available in ν_μ -H and $\bar{\nu}_\mu$ -H CC interactions. For more details about the analysis technique used, we refer to Ref. [25].

3.2 Kinematic analysis

3.2.1 Selection of $\nu_\mu H \rightarrow \mu^- p \pi^+$ and $\bar{\nu}_\mu H \rightarrow \mu^+ p \pi^-$

In order to illustrate the potential of the proposed technique, we start from an analysis of the cleanest topologies $\nu_\mu p \rightarrow \mu^- p \pi^+$ and $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$, mainly originating from resonance production. Detailed GEANT4 simulations indicate that the average proton

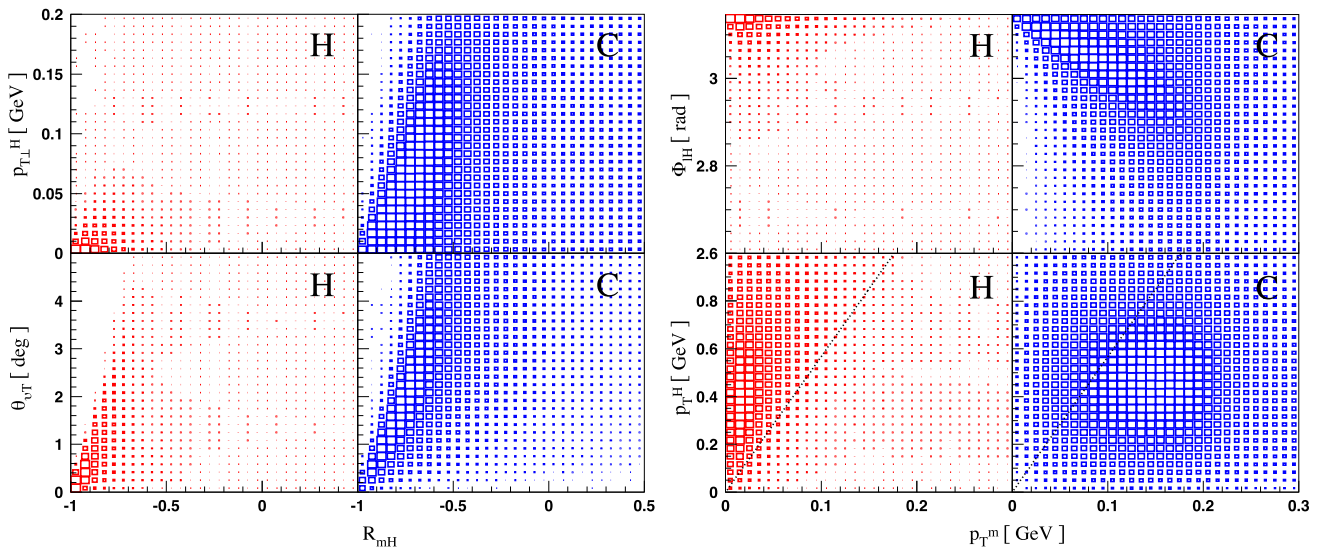


Fig. 2 Some of the correlations between kinematic variables used to construct $\ln \lambda^H$ and $\ln \lambda_4^H$ for H signal (red color) and C background (blue color) exclusive $\mu^- p\pi^+$ CC topologies. The dotted line on the bottom right plots corresponds to a cut $R_{mH} < -0.7$ (Fig. 1)

reconstruction efficiency¹ is about 93% for H events and 67% in C events, since in this latter case nuclear effects result, on average, in smaller kinetic energies and larger angles for the protons. As discussed in Sect. 3.1, the final state particles from these processes on H can be accurately reconstructed in the detector described in Sect. 2, thus resulting in an excellent measurement of all the relevant kinematic variables.

The most discriminant kinematic variable to separate H interactions from the ones originated in nuclear targets is found to be $R_{mH} \equiv (p_T^m - p_T^H)/(p_T^m + p_T^H)$, the asymmetry between the missing transverse momentum, p_T^m , and the transverse momentum of the hadron vector, p_T^H . For H interactions, p_T^m is consistent with zero up to reconstruction effects and hence we expect $R_{mH} \sim -1$. Instead, if the interactions occur inside a nuclear target we expect, on average, a substantial p_T^m together with smaller values of p_T^H , due to the nuclear smearing. Furthermore, the missing transverse momentum is mainly generated inside the hadron system and it is expected to be correlated with the latter. All these nuclear effects result in much larger values of R_{mH} . As shown in Fig. 1, this variable can be efficiently used to separate H interactions, as well as to probe various aspects of the nuclear smearing. Another useful variable is the magnitude of the component of the hadron transverse momentum perpendicular to the transverse momentum of the lepton, $p_{T\perp}^H$. In H interactions, the transverse momenta of the lepton and of the hadron system are back-to-back, thus resulting in a sharp peak around zero in $p_{T\perp}^H$. Interactions from nuclear targets have a much broader distribution originating from the nuclear smearing. The use of this variable to study H interactions within composite targets was suggested in Ref. [39]. Since $p_{T\perp}^H$ is selecting topologies in which the transverse momenta of the lepton and the hadron system are back-to-back, the effect of this variable is similar to the use of the angle between those transverse vectors, Φ_{lH} . Used as a single variable in the H selection, $p_{T\perp}^H$ has less discriminating power than both R_{mH} and the other kinematic variables discussed in this section. However, it provides information complementary to R_{mH} , so that the combined effect of both these variables improves the overall selection efficiency.

We can improve the selection of H interactions by using multivariate techniques exploiting the complete event kinematics [25–27]. Assuming the two momentum vectors of the lepton and hadron system, we have in total three transverse and two longitudinal degrees of freedom in the event selection, due to the invariance for an arbitrary rotation in the transverse plane. Since we want to separate the same type of CC events with and without nuclear effects, we can further assume that the overall reconstructed energy spectra are similar, thus somewhat reducing the rejection power of one of the longitudinal variables compared to the transverse ones. As a result, we can define a complete kinematic set as three transverse plus one longitudinal variables. We select the angle between the total visible momentum vector and the incident neutrino direction (z axis), θ_{vT} , as the variable including longitudinal information. This variable is expected to be close to zero in H interactions, up to the tiny beam divergence, while it is much larger in interactions originated from nuclear targets.

We use likelihood functions incorporating multi-dimensional correlations among kinematic variables. A study of the kinematic selection suggests the following function using only three-dimensional correlations:

$$\mathcal{L}^H \equiv \left[R_{mH}, p_{T\perp}^H, \theta_{vT}, p_T^m, \Phi_{lH} \right] \quad (1)$$

¹ To be reconstructed, the proton track emerging from a vertex must have at least four hits in the bending plane to allow a momentum determination. We do not make any attempt to reconstruct events failing such a requirement from the energy deposition of the available hits.

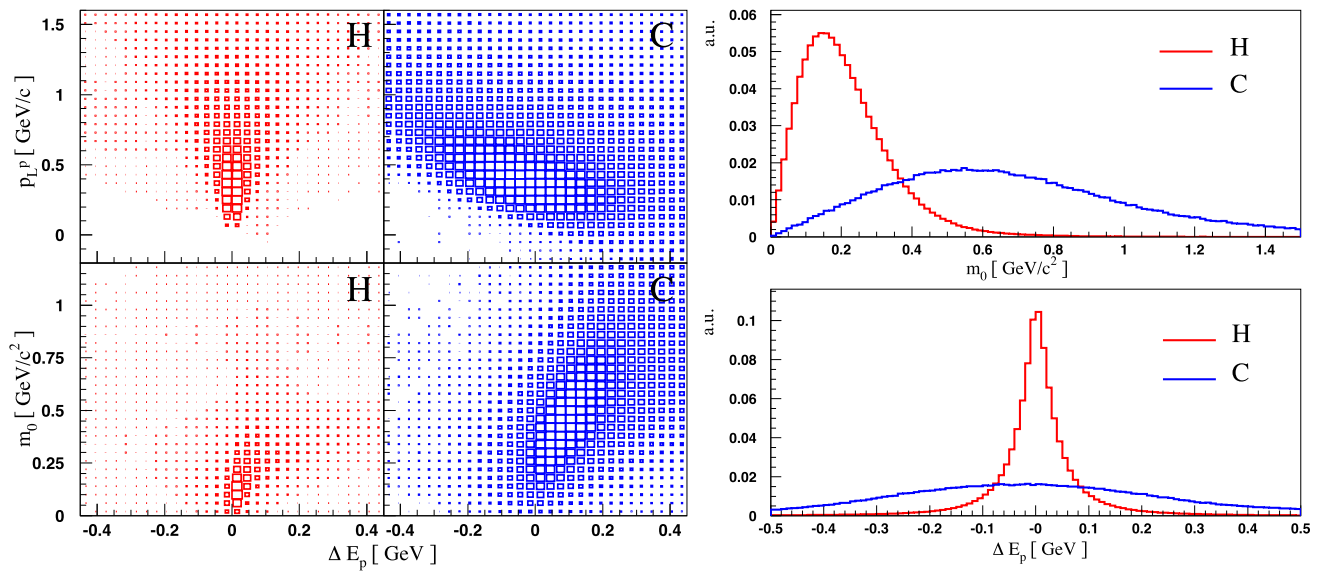


Fig. 3 Some of the kinematic variables and correlations used to construct $\ln \lambda_{\text{IN}}^H$ for H signal (red color) and C background (blue color) exclusive $\mu^- p \pi^+$ CC topologies. All histograms in the right plots are normalized to unit area

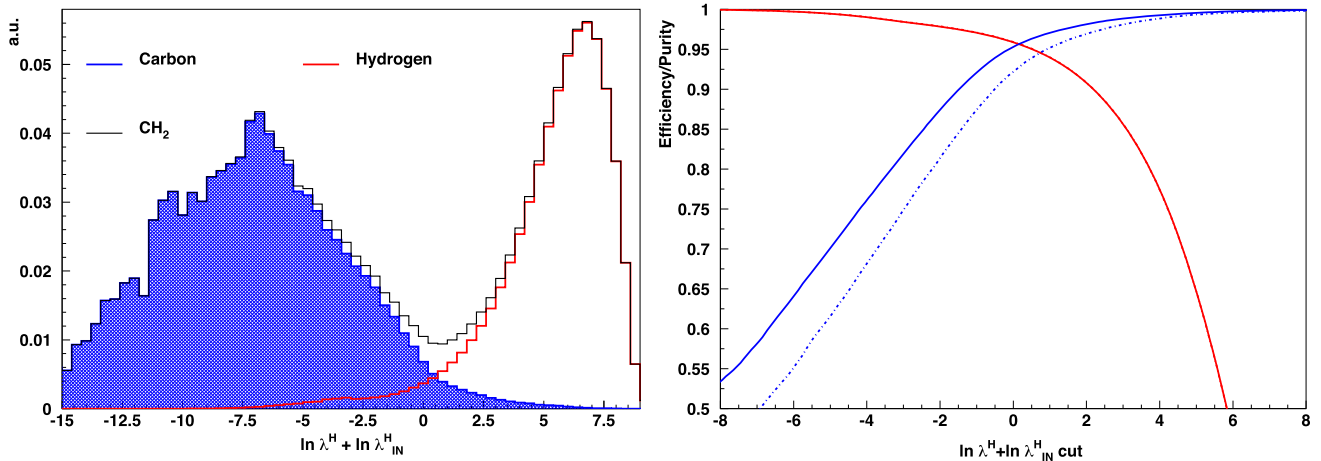


Fig. 4 Left plot: Distributions of $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ for the H signal, the C background, and the CH_2 plastic (sum) for the exclusive $\mu^- p \pi^+$ CC topologies. The multiple peaks result from the binning used to build \mathcal{L}^H and $\mathcal{L}_{\text{IN}}^H$. The C distribution is normalized to unit area while the H distribution is normalized to the expected relative abundance in CH_2 . Right plot: Efficiency (red color) and purity (blue color) as a function of the cut on $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ for the kinematic selection of the exclusive processes $\nu_\mu p \rightarrow \mu^- p \pi^+$ (solid lines) and $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$ (dashed-dotted lines) on hydrogen from the CH_2 target. The efficiency curves are common to both channels

where the square brackets denote correlations (Fig. 2). A function strictly based upon a complete set of kinematic variables is the four-dimensional:

$$\mathcal{L}_4^H = [p_T^H, p_T^m, \Phi_{lH}, \theta_{vT}] \quad (2)$$

which incorporates the R_{mH} variable through its underlying correlation $[p_T^H, p_T^m]$ (Fig. 2). Although the use of the \mathcal{L}_4^H function requires a larger statistics to achieve a sensible binning, we used it for our kinematic selection and obtained results similar to \mathcal{L}^H . We build the \mathcal{L}^H and \mathcal{L}_4^H probability density functions (pdf) for the two test hypotheses of H interactions (signal) and C interactions (background). The individual pdf are properly smoothed and are built with samples independent from the test ones to avoid large statistical biases. As it is common practice, the logarithm of the final likelihood ratio between signal and background hypotheses, $\ln \lambda^H$ or $\ln \lambda_4^H$, is used as discriminant [25]. We note that in the actual measurements we will directly use interactions from the graphite targets to build the pdf for the C hypothesis, eliminating thus any dependence from the nuclear modeling.

We can also exploit the information related to the individual particles inside the hadron system. For $\mu^\mp p \pi^\pm$ topologies, we have a total of three additional degrees of freedom, since the total hadron momentum vector is constrained by the global event kinematics in \mathcal{L}^H and \mathcal{L}_4^H . We select one such variable as the difference ΔE_p between the energy of the neutrino calculated from the μ and

Table 1 Efficiency ε and purity for the kinematic selection of H interactions from the CH₂ plastic target using the likelihood ratio $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ or $\ln \lambda_4^H + \ln \lambda_{\text{IN}}^H$. For the μ^+n QE topologies $\ln \lambda_{\text{QE}}^H$ is used instead. The cuts applied for each channel are chosen to maximize the sensitivity defined as $S/\sqrt{S+B}$, where S is the H signal and B the C background. The CC inclusive samples are obtained from the combination of the corresponding exclusive channels

| Process | $\nu\mu$ -H CC | | | | $\bar{\nu}\mu$ -H CC | | | | | |
|--------------------|-----------------------|-------------------------|------------------------------|------------------|-----------------------|--------------------|---------------|-------------------------|--------------------------|------------------|
| | $\mu^- p\pi^+$ (%) | $\mu^- p\pi^+ X$ (%) | $\mu^- n\pi^+\pi^+ X$ (%) | Inclusive (%) | $\mu^+ p\pi^-$ (%) | $\mu^+ n\pi^0$ (%) | $\mu^+ n$ (%) | $\mu^+ p\pi^- X$ (%) | $\mu^+ n\pi\pi X$ (%) | Inclusive (%) |
| Eff. ε | 96 | 89 | 75 | 93 | 94 | 84 | 75 | 85 | 82 | 80 |
| Purity | 95 | 93 | 70 | 93 | 95 | 84 | 80 | 94 | 84 | 84 |

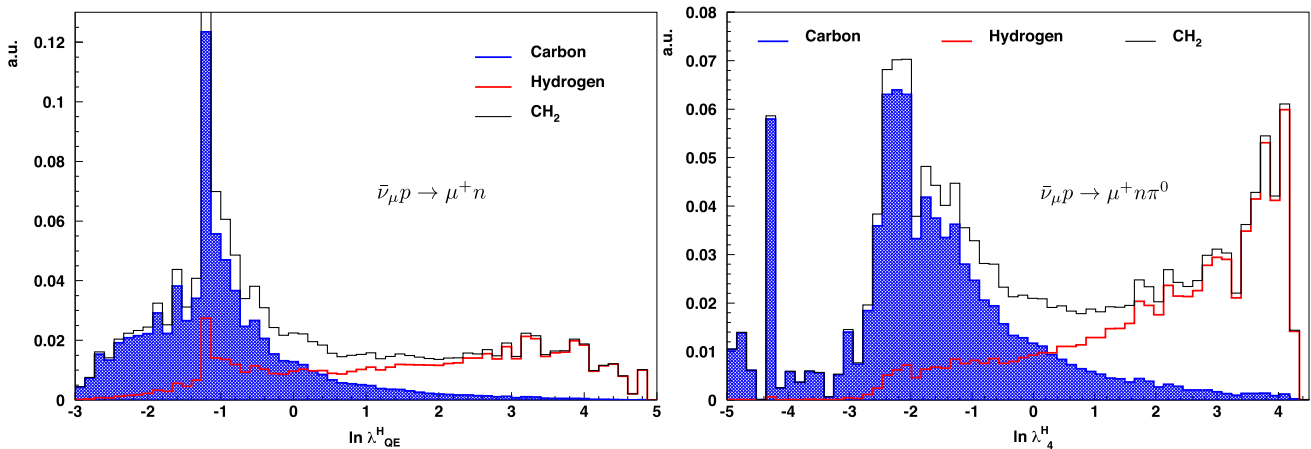


Fig. 5 Left plot: Distributions of $\ln \lambda_{\text{QE}}^H$ for the H signal, the C background, and the CH₂ plastic (sum) for the selection of $\bar{\nu}_\mu p \rightarrow \mu^+ n$ QE events. Right plot: Distributions of $\ln \lambda_4^H$ for the selection of $\bar{\nu}_\mu p \rightarrow \mu^+ n\pi^0$ events. The multiple peaks are the effect of the binning used to build the likelihood functions. The C distributions are normalized to unit area, while the H distributions are normalized to the expected relative abundance in CH₂. See the text for details

π momenta by applying energy–momentum conservation and the one reconstructed from the sum of the momentum vector of all three particles $\mu^\mp p\pi^\pm$:

$$\Delta E_p = \frac{m_\mu^2 - m_{\pi^\pm}^2 + 2M_p(E_\mu + E_{\pi^\pm}) - 2\vec{p}_\mu \cdot \vec{p}_{\pi^\pm}}{2(M_p - E_\mu - E_{\pi^\pm} + |\vec{p}_\mu| \cos \theta_\mu + |\vec{p}_{\pi^\pm}| \cos \theta_{\pi^\pm})} - |\vec{p}_\mu + \vec{p}_{\pi^\pm} + \vec{p}_p| \quad (3)$$

where M_p , M_n , m_μ , m_{π^\pm} are the masses of the proton, neutron, muon, and pion, respectively, $p_{\mu(\pi^\pm)}$, $\vec{p}_{\mu(\pi^\pm)}$, $E_{\mu(\pi^\pm)}$ and $\theta_{\mu(\pi^\pm)}$ are the four-momentum, momentum vector, energy, and angle of the outgoing muon (pion), and \vec{p}_p is the proton momentum vector. The variable ΔE_p is close to zero up to reconstruction effects in hydrogen, while it is significantly larger in carbon events, due to the nuclear smearing. Another useful variable is the invariant mass of the reconstructed neutrino, calculated as $m_0^2 = (p_\mu + p_{\pi^\pm} + p_p - p_N)^2$ where p_p and p_N are the four-momenta of the outgoing proton and of the target proton assumed at rest, respectively. We use the following likelihood function using information from the internal $p\pi$ structure:

$$\mathcal{L}_{\text{IN}}^H = [\Delta E_p, p_L^p, m_0] \quad (4)$$

where p_L^p is the longitudinal component of the proton momentum vector along the beam direction. Figure 3 shows the main variables and correlations included in $\mathcal{L}_{\text{IN}}^H$. Since $\mathcal{L}_{\text{IN}}^H$ is essentially independent from \mathcal{L}^H and \mathcal{L}_4^H we multiply the corresponding density functions and use the sum $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ or $\ln \lambda_4^H + \ln \lambda_{\text{IN}}^H$ as the final discriminant for our analysis.

The distributions of $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ for the H signal and the C background in $\mu^- p\pi^+$ topologies are shown in Fig. 4 (left plot). The corresponding purity and efficiency achievable as a function of the $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ cut are given in the right plot of Fig. 4, for both the $\nu_\mu p \rightarrow \mu^- p\pi^+$ and $\bar{\nu}_\mu p \rightarrow \mu^+ p\pi^-$ samples. Both the efficiency and the purity appear relatively uniform as a function of the neutrino energy. Table 1 summarizes the results obtained by applying the cut on $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ maximizing the sensitivity $S/\sqrt{S+B}$, where S is the number of events for H signal and B for the C background. The fact that the maximum sensitivity corresponds to regions with high purity for the selected H signal indicates that the kinematic selection is optimal. An advantage of this approach is that the likelihood function allows to evaluate, on an event-by-event basis, the probability that a given (anti)neutrino interaction originated from either the hydrogen or the carbon nucleus. Furthermore, it provides a better control of the selection procedure with respect to a simple cut-based analysis, by allowing an easier variation of the efficiency/purity and by offering relatively clean control samples.

3.2.2 Selection of $\bar{\nu}_\mu H \rightarrow \mu^+ n$

An important exclusive process available in $\bar{\nu}_\mu$ CC interactions is the quasi-elastic $\bar{\nu}_\mu p \rightarrow \mu^+ n$ on hydrogen. The reconstruction of this topology is more complex because of the presence of a neutron in the final state and a single charged track. We performed detailed GEANT4 simulations of the detector to study the reconstruction efficiency of the neutrons. The results were independently checked using FLUKA simulations. About 25% of the neutrons produced in $\bar{\nu}_\mu p \rightarrow \mu^+ n$ QE events on H interact inside the STT and are detected from the corresponding hits in the straws. An additional fraction of about 55% of the neutrons are detected in the ECAL [40] surrounding the tracker, thus allowing a combined detection efficiency of 80.5% for H events. Conversely, the combined detection efficiency for neutrons originated from the C background events is only about 64.8%, since they are typically affected by nuclear effects, resulting, on average, in a smaller kinetic energy. For events with a single charged track, the resolution on the position of the primary vertex is worse than for multi-track events, and is essentially defined by the thickness of a single CH₂ (or C) target plane by noting the absence of straw tube hits preceding the presumed target. However, events can still be efficiently associated to the correct target material, due to the lightness of the tracking straws and the purity of the target itself. The corresponding uncertainty is given by the ratio between the thickness of the straw walls ($< 20 \mu\text{m}$) and the thickness of a single CH₂ target, resulting in an efficiency $> 99\%$ (Sect. 3.1). From the position of the first track hit and that of the neutron interaction within the detector, we can reconstruct the neutron line of flight. Assuming that the target proton is at rest, we calculate the energy of the incoming antineutrino as:

$$E_\nu = \frac{M_n^2 - m_\mu^2 + 2M_p E_\mu - M_p^2}{2(M_p - E_\mu + |\vec{p}_\mu| \cos \theta_\mu)} \quad (5)$$

with the same notations as in Eq. (3). The energy of the neutron is $E_n = E_\nu + M_p - E_\mu$ and the momentum vector of the neutron is obtained from the measured neutron direction and the calculated energy E_n ². We note that Eq. (5) is correct only for interactions on hydrogen and not for the ones on carbon, due to nuclear effects. We use a realistic smearing on the measured direction of the neutron obtained from a detailed FLUKA simulation of the detector and including the non-Gaussian tails, mainly related to elastic scattering of the neutrons before interacting. We can then reconstruct the complete event kinematics for $\bar{\nu}_\mu p \rightarrow \mu^+ n$ interactions on hydrogen and apply the same kinematic selection described in Sect. 3.2. The resolution achievable on many of the kinematic variables is dominated by the reconstruction of the neutron direction rather than that of the μ^+ momentum vector, given the excellent momentum and angular resolutions of the STT.

An additional background source to consider for the selection of the process $\bar{\nu}_\mu p \rightarrow \mu^+ n$ on H is given by the uncorrelated neutrons originated from (anti)neutrino interactions occurring in the large amount of materials surrounding the STT, including the ECAL, the magnet elements, and the external rocks. We estimate this background by randomly overlaying a genuine $\bar{\nu}_\mu p \rightarrow \mu^+ n$ event on H with a neutron extracted from either a different CH₂ event or from an event in the surrounding materials within the same beam spill. We then reconstruct the neutron momentum using the measured direction of the random neutron and the calculated energy E_n , as described above. We require a small time difference between the interaction point of the random neutron and the primary vertex, $\Delta t \equiv t_n - t_{\text{vtx}}$. We also require that the “measured” time-of-flight Δt is consistent with the expected one using the calculated neutron energy, with small values of the variable $\Delta t' \equiv \Delta t - d/(\beta_n c)$. We further reject the random neutrons which are closely correlated—in time and space—to a detected activity from an external (anti)neutrino interaction detected either in STT or in the surrounding ECAL. The combined effect of these timing and topological cuts rejects most of the random neutrons by retaining 90.4% of H events.

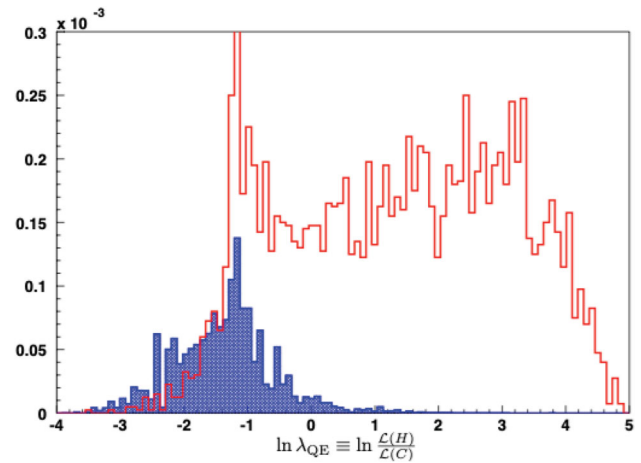
We have less kinematic information available for the selection of $\bar{\nu}_\mu p \rightarrow \mu^+ n$ QE on H with respect to other channels. We use the optimal function \mathcal{L}_4^H from Eq. (2) as the basis of our kinematic analysis. The likelihood function $\mathcal{L}_{\text{IN}}^H$, as defined in Eq. (4), is not strictly applicable, since the hadron system is composed of a single neutron without internal structure. However, we can still exploit one additional longitudinal degree of freedom with respect to Eq. (4). In analogy of Eq. (3), we calculate $\Delta E_n = E_\nu - |\vec{p}_\mu + \vec{p}_n|$, where E_ν is given by Eq. (5) and \vec{p}_n is the neutron momentum calculated as described above. We use the following likelihood function for the selection of $\bar{\nu}_\mu p \rightarrow \mu^+ n$ QE on H:

$$\mathcal{L}_{\text{QE}}^H = \left[|\vec{p}_T^H|, p_T^m, \Phi_{lH}, \theta_{vT}, \Delta E_n, p_L^n \right] \quad (6)$$

The variable p_L^n represents the longitudinal component of the neutron momentum vector along the beam direction. We note that, although the two-dimensional correlation $[\Delta E_n, p_L^n]$ appears similar to the corresponding one in Eq. (4), it has a different role in the QE channel, as a part of the global event kinematics. We therefore include it into a single likelihood function, since it is expected to be correlated with the other kinematic variables. Figure 5 shows the distributions of $\ln \lambda_{\text{QE}}^H$ for the H signal and the C background. Table 1 summarizes the efficiency and purity for a cut on $\ln \lambda_{\text{QE}}^H$ maximizing the sensitivity of the analysis. Since the random neutron background is uncorrelated with the primary vertex from the $\bar{\nu}_\mu p \rightarrow \mu^+ n$ interaction on H, the corresponding kinematic distributions are inconsistent with those expected from a genuine H event. Figure 6 shows the distribution of $\ln \lambda_{\text{QE}}^H$ for the random neutrons passing the timing and topological cuts described above. We do not recalculate the likelihood functions using

² It is possible to calculate the complete momentum vector of the neutron from energy–momentum conservation. However, the direct use of transverse plane kinematics would void the kinematic selection against the C background.

Fig. 6 Distribution of $\ln \lambda_{QE}^H$ for the random neutron background and for the $\bar{\nu}_\mu p \rightarrow \mu^+ n$ events on H. The likelihood functions calculated from the C and H hypotheses are used, as in Fig. 5. The multiple peaks are the effect of the binning used to build the likelihood functions. Timing and topological cuts are applied as described in the text



the random neutron events and we keep the same $\ln \lambda_{QE}^H$ cuts as in the H/C separation, resulting in no additional efficiency loss for the H signal. We note that the efficient kinematic selection could allow looser timing and topological cuts against random neutrons.

3.2.3 Selection of $\bar{\nu}_\mu H \rightarrow \mu^+ n \pi^0$

Another exclusive process available only in $\bar{\nu}_\mu$ CC interactions is $\bar{\nu}_\mu p \rightarrow \mu^+ n \pi^0$ on hydrogen, mainly originated from resonance production. The kinematic selection of this channel is similar to the one discussed in Sect. 3.2.1 for the complementary process $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$, with the exception of a few specific features related to the reconstruction of the neutron and the π^0 . The relative fractions of these two channels are about comparable in $\bar{\nu}_\mu$ CC interactions on hydrogen. Detailed GEANT4 simulations indicate that the average neutron reconstruction efficiency is about 85% for H events and 76.5% in C using both the STT and the surrounding ECAL (Sect. 3.2.2).

For the reconstruction of the π^0 , we distinguish the case in which γ s are converted into e^+e^- pairs within the STT volume and the one in which γ s are detected only in the surrounding ECAL. In the former case, we use the momentum and angular smearing for e^\pm tracks in STT, while in the latter case, we apply the corresponding smearing obtained from a detailed ECAL simulation. On average, about 50% of all π^0 have at least one converted γ within the STT volume, providing a more accurate reconstruction of the π^0 direction. We calculate the energy of the neutrons interacting in the detector as:

$$E_n = \frac{M_n^2 - m_\mu^2 - m_{\pi^0}^2 + 2M_p(E_\mu + E_{\pi^0}) - 2p_\mu \cdot p_{\pi^0} - M_p^2}{2(M_p - E_\mu - E_{\pi^0} + |\vec{p}_\mu| \cos \theta_\mu + |\vec{p}_{\pi^0}| \cos \theta_{\pi^0})} + M_p - E_\mu - E_{\pi^0} \quad (7)$$

where p_{π^0} , \vec{p}_{π^0} , E_{π^0} and θ_{π^0} are the four-momentum, momentum vector, energy and angle of the outgoing π^0 , respectively. The first term on the r.h.s. of Eq. (7) represents the neutrino energy calculated from the muon and π^0 momenta using energy–momentum conservation. We note that Eq. (7) is correct only for interactions on hydrogen since it assumes a target proton at rest. We reconstruct the neutron momentum vector \vec{p}_n following the same procedure described in Sect. 3.2.2 for the QE process, combining the measured (smeared) direction of the neutron with the energy E_n calculated from Eq. (7).

We use the likelihood function \mathcal{L}_4^H from Eq. (2) to describe the global event kinematics. In analogy to the $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$ case, it is possible to exploit the additional information related to the individual particles within the hadron system with the function $\mathcal{L}_{IN}^H = [\Delta E_{\pi^0}, p_L^{\pi^0}, m_0]$, where $\Delta E_{\pi^0} = E_\nu - |\vec{p}_\mu + \vec{p}_{\pi^0} + \vec{p}_n|$, E_ν is the neutrino energy calculated from the muon and neutron momenta using energy–momentum conservation, and $p_L^{\pi^0}$ is the longitudinal component of the momentum vector of the π^0 . The use of ΔE_{π^0} is preferable with respect to the equivalent quantity for the neutron, ΔE_n , since this latter is partially biased by the calculation of the neutron energy from Eq. (7). Since the average reconstruction smearing for the neutron and π^0 is larger than for most charged particles, the $\ln \lambda_{IN}^H$ improves only marginally the selection of $\bar{\nu}_\mu p \rightarrow \mu^+ n \pi^0$ on H and can be dropped for this channel. Figure 5 shows the distributions of $\ln \lambda_4^H$ for the H signal and the C background. Table 1 summarizes the efficiency and purity for a cut on $\ln \lambda_4^H$ maximizing the sensitivity of the analysis. Similar results are obtained with $\ln \lambda^H$.

3.2.4 Selection of $\nu_\mu H \rightarrow \mu^- p \pi^+ X$ and $\bar{\nu}_\mu H \rightarrow \mu^+ p \pi^- X$

In this section and in the following one, we consider the collective selection of all the inelastic topologies produced in $\nu_\mu p$ and $\bar{\nu}_\mu p$ CC interactions on H and different from the $\mu^- p \pi^+$, $\mu^+ p \pi^-$, and $\mu^+ n \pi^0$ topologies discussed in Sects. 3.2.1 and 3.2.3. These samples are dominated by the deep inelastic scattering (DIS): about 63% of the $\mu^- \pi^+ X$ events and 7% of the $\mu^- p \pi^+$ sample have $W > 1.8$ GeV with the default LBNF beam spectra [35].

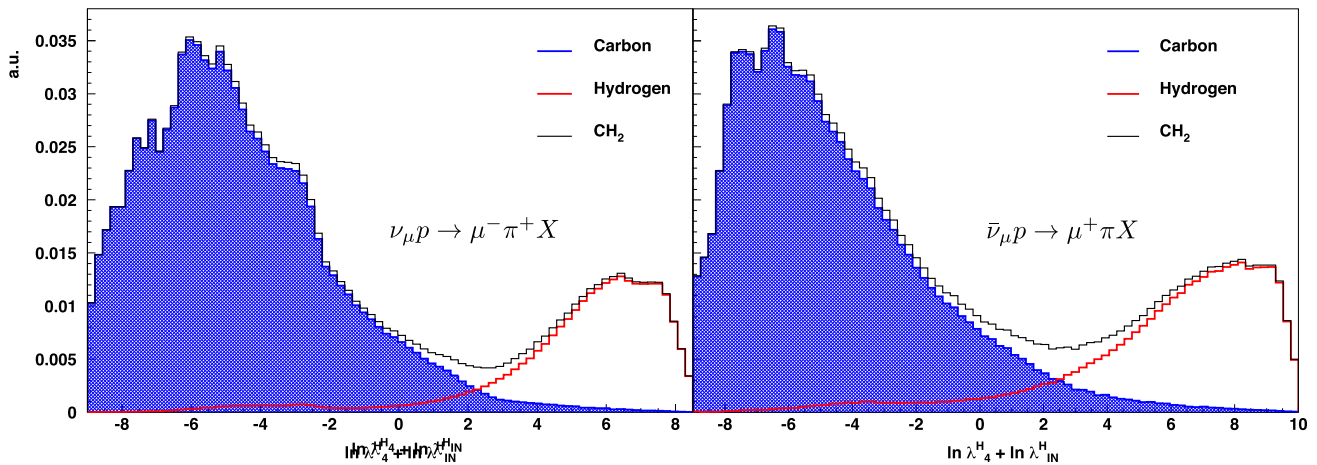


Fig. 7 Distributions of $\ln \lambda_4^H + \ln \lambda_{IN}^H$ for the H signal, the C background, and the CH_2 plastic (sum) for $\nu_\mu p \rightarrow \mu^- \pi^+ X$ (left plot) and $\bar{\nu}_\mu p \rightarrow \mu^+ \pi X$ (right plot), including the sum of the subsamples with a proton and a neutron. The multiple peaks are the effect of the binning used to build \mathcal{L}_4^H and \mathcal{L}_{IN}^H . The C distributions are normalized to unit area, while the H distributions are normalized to the expected relative abundance in CH_2 . See the text for details

The inelastic $\nu_\mu p \rightarrow \mu^- \pi^+ X$ and $\bar{\nu}_\mu p \rightarrow \mu^+ \pi X$ samples on H are characterized, in general, by a higher multiplicity and a worse reconstruction of some of the events, compared to the $\mu^- p \pi^+$, $\mu^+ p \pi^-$, and $\mu^+ n \pi^0$ topologies. The larger detector smearing directly affects the reconstruction of the kinematic variables, thus somewhat reducing their discriminating power. However, these effects are even larger for the DIS interactions originated in nuclear targets, primarily because of the final state interactions. We can therefore still achieve an adequate separation of the interactions on hydrogen from the carbon background, which largely dominates the statistics of the $\mu^- \pi^+ X$ and $\mu^+ \pi X$ samples from the CH_2 target.

Since we want to select interactions on protons, we require that the total charge measured at the primary vertex is $C = +1$. This cut rejects a large fraction of the interactions on neutrons in background events and has an efficiency of about 23%(33%) for $\nu_\mu(\bar{\nu}_\mu)$ CC interactions on C. We also require the presence of at least one π^+ in $\nu_\mu p$ and one π^- or $\bar{\nu}_\mu p$ events to select inelastic interactions.

In $\nu_\mu p$ and $\bar{\nu}_\mu p$ CC interactions on H, we expect a single nucleon—either a proton or a neutron—in the final state, while for interactions on C nuclear effects including final state interactions can result in higher multiplicities. We perform a separate analysis of the two subsamples with a proton and a neutron. In this section, we discuss the selection of the subsamples with a reconstructed proton. Detailed GEANT4 simulations indicate that the average proton reconstruction efficiency is about 96% for H events and 74% in C events.

The kinematic selection of $\nu_\mu p \rightarrow \mu^- p \pi^+ X$ and $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^- X$ on H can exploit additional degrees of freedom with respect to the $\mu^- p \pi^+$, $\mu^+ p \pi^-$, and $\mu^+ n \pi^0$ topologies, due to the higher number of particles in the hadron system. We use the likelihood function \mathcal{L}_4^H from Eq. (2) to describe the global event kinematics. We then tag the single particle within the hadron system potentially affected by the largest nuclear effects by maximizing the magnitude $|\Delta E_{h_i}|$, defined as:

$$\Delta E_{h_i} = \frac{m_{h_i}^2 - m_\mu^2 - (W')^2 + 2M_p(E_\mu + E_{H'}) - 2\vec{p}_\mu \cdot \vec{p}_{H'} - M_p^2}{2(M_p - E_\mu - E_{H'} + |\vec{p}_\mu| \cos \theta_\mu + |\vec{p}_{H'}| \cos \theta_{H'})} - |\vec{p}_\mu + \sum_k \vec{p}_{h_k}| \quad (8)$$

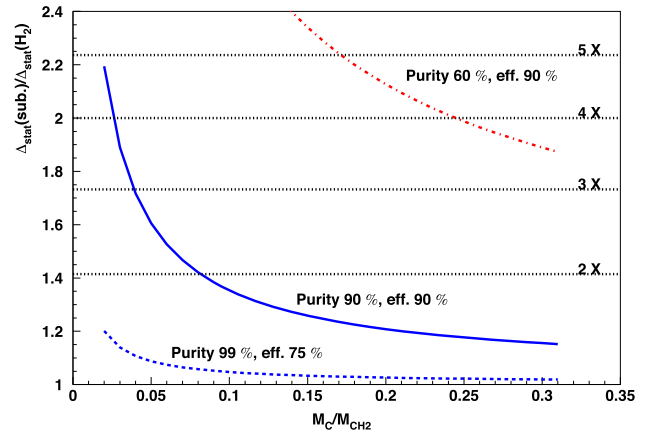
which is calculated for each hadron particle h_i of mass m_{h_i} and momentum vector \vec{p}_{h_i} . The first term on the r.h.s. of Eq.(8) is similar to the equivalent one in Eq. (7), with the π^0 replaced by a reduced hadron system H' excluding the single particle h_i being considered. The quantity $(W')^2 = E_{H'}^2 - |\vec{p}_{H'}|^2$ is the invariant mass of the reduced hadron system, and $E_{H'} = \sum_{k \neq i} E_{h_k}$ and $\vec{p}_{H'} = \sum_{k \neq i} \vec{p}_{h_k}$ are the corresponding energy and momentum vector. We select the hadron particle maximizing the magnitude $|\Delta E_{h_i}|$ and use the corresponding value $\Delta E_{h_i}^{\max}$ (with sign) and the longitudinal momentum of such a particle, $p_L^{h_i}$, as input for the likelihood function³ based upon information internal to the hadron system:

$$\mathcal{L}_{IN}^H = [\Delta E_{h_i}^{\max}, p_L^{h_i}, m_0,] \quad (9)$$

where we use similar notations as in Eq. (4). Figure 7 shows the distributions of $\ln \lambda_4^H + \ln \lambda_{IN}^H$ for the H signal and the C background in $\nu_\mu p \rightarrow \mu^- \pi^+ X$ and $\bar{\nu}_\mu p \rightarrow \mu^+ \pi^- X$ topologies. Table 1 summarizes the efficiency and purity in the selection of both $\nu_\mu p \rightarrow \mu^- p \pi^+ X$ and $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^- X$ processes on H with a cut on $\ln \lambda_4^H + \ln \lambda_{IN}^H$ maximizing the sensitivity of the analysis. Similar results are obtained with $\ln \lambda^H + \ln \lambda_{IN}^H$.

³ In principle, the same approach can be used for $\mu^- p \pi^+$, $\mu^+ p \pi^-$, and $\mu^+ n \pi^0$ topologies as well.

Fig. 8 Ratio between the relative statistical uncertainty Δ_{stat} on the H sample obtained after the C background subtraction and the corresponding ideal one from a pure H_2 sample of the same size, as a function of the ratio between the fiducial masses of the graphite and CH_2 targets, M_C/M_{CH_2} . The curves for different values of the H purities and efficiencies are shown to illustrate the impact of the kinematic selection. See text for details



3.2.5 Selection of $\nu_\mu H \rightarrow \mu^- n \pi^+ \pi^+ X$ and $\bar{\nu}_\mu H \rightarrow \mu^+ n \pi \pi X$

In this section, we discuss the selection of the subsamples of the inelastic $\nu_\mu p$ and $\bar{\nu}_\mu p$ CC interactions on H with a detected neutron in the final state and different from the $\mu^+ n \pi^0$ topologies (Sect. 3.2.3). These samples are complementary with respect to the similar ones with a reconstructed proton in the final state described in Sect. 3.2.4. Detailed GEANT4 simulations indicate that the average neutron reconstruction efficiency is about 87.4% for H events and 80.1% in C events using both the STT and the surrounding ECAL (Sect. 3.2.2). For $\nu_\mu p$ CC on H, the presence of a neutron in the final state is tagged by requiring that no proton and two π^+ are present. These criteria correctly tag the neutron in 99.8% of the H events with a neutron, including events in which the neutron is not detected. Events with a neutron in $\bar{\nu}_\mu p$ CC on H are tagged with an efficiency of 88% by requiring that no proton and a number of pions ($\pi^- + \pi^+ + \pi^0$) ≥ 2 with total charge equal to zero are present in the final state. The analysis follows closely the one described in Sect. 3.2.4 for the subsamples with a proton, the main difference being the treatment of the neutron. We calculate the neutron energy using energy–momentum conservation and Eq. (7) with the π^0 replaced by a reduced hadron system H' excluding the neutron:

$$E_n = \frac{M_n^2 - m_\mu^2 - (W')^2 + 2M_p(E_\mu + E_{H'}) - 2p_\mu \cdot p_{H'} - M_p^2}{2(M_p - E_\mu - E_{H'} + |\vec{p}_\mu| \cos \theta_\mu + |\vec{p}_{H'}| \cos \theta_{H'})} + M_p - E_\mu - E_{H'} \quad (10)$$

where we use the same notations as in Eq. (8) with $h_i \equiv n$. We reconstruct the neutron momentum vector \vec{p}_n following the same procedure described in Sect. 3.2.2 for the QE process, combining the measured (smeared) direction of the neutron with the energy E_n calculated from Eq. (10). For events with more than one neutron detected the calculation above is not applicable and we ignore the neutrons.

We use \mathcal{L}_4^H from Eq.(2) to describe the global event kinematics and $\mathcal{L}_{\text{IN}}^H$ from Eq.(9) for the information related to the individual particles inside the hadron system. Since the angular smearing for the detected neutrons (Sect. 3.2.2) is typically larger than for other particles, we use the track with the largest angle with respect to the beam direction to calculate $\Delta E_{h_i}^{\text{max}}$ in Eq. (9), rather than explicitly maximizing $|\Delta E_{h_i}|$. Figure 7 shows the distributions of $\ln \lambda_4^H + \ln \lambda_{\text{IN}}^H$ for the H signal and the C background in $\nu_\mu p \rightarrow \mu^- \pi^+ X$ and $\bar{\nu}_\mu p \rightarrow \mu^+ \pi X$ topologies. Table 1 summarizes the efficiency and purity in the selection of both $\nu_\mu p \rightarrow \mu^- n \pi^+ \pi^+ X$ and $\bar{\nu}_\mu p \rightarrow \mu^+ n \pi \pi X$ processes on H with a cut on $\ln \lambda_4^H + \ln \lambda_{\text{IN}}^H$ maximizing the sensitivity of the analysis. Similar results are obtained with $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$.

3.2.6 Selection of $\nu_\mu H$ and $\bar{\nu}_\mu H$ CC inclusive

In the previous sections, we optimized the selection of the various exclusive topologies available in $\nu_\mu p$ and $\bar{\nu}_\mu p$ CC interactions on H by maximizing independently the corresponding sensitivities. The results summarized in Table 1 are characterized by varying efficiencies and purities across different channels. For measurements requiring the inclusive CC samples, we can combine the individual exclusive topologies with their corresponding relative fractions in $\nu_\mu p$ and $\bar{\nu}_\mu p$ CC interactions on H. The average efficiency and purity of the resulting inclusive CC samples on H are listed in Table 1.

3.3 Achievable statistics

In the following, we will assume an overall fiducial mass of 5 tons for the CH_2 targets, corresponding to a “solid” hydrogen mass of about 700 kg. This value is realistically achievable with the detector technology discussed in Sect. 2 and a relatively compact

Table 2 Number of events expected in the selection of all the various processes on H with the default low-energy (anti)neutrino beams available at the LBNF [35], assuming a fixed exposure of 5.5×10^{21} POT, a detector located at 574 m from the source and a fiducial target mass of 5 tons of CH_2 . The first two columns (CH_2 and H targets) refer to the initial statistics, while the last three include all selection cuts described in Sect. 3 and Table 1). For the CH_2 and C targets, the numbers refer to the given final state topologies originated from either p or n interactions. The fifth column shows the total residual C background to be subtracted from the corresponding CH_2 selected samples. We use a ratio $M_C/M_{C/\text{CH}_2} = 0.12$ to measure the C backgrounds from the graphite targets. See text for details

| CC process | CH_2 target | H target | CH_2 selected | C bkgnd | H selected |
|---|----------------------|-----------|------------------------|---------|------------|
| $\nu_\mu p \rightarrow \mu^- p \pi^+$ | 5,615,000 | 2,453,000 | 2,305,000 | 115,000 | 2,190,000 |
| $\nu_\mu p \rightarrow \mu^- p \pi^+ X$ | 11,444,000 | 955,000 | 877,000 | 61,000 | 816,000 |
| $\nu_\mu p \rightarrow \mu^- n \pi^+ \pi^+ X$ | 3,533,000 | 183,000 | 158,000 | 48,000 | 110,000 |
| ν_μ CC inclusive | 34,900,000 | 3,591,000 | 3,340,000 | 224,000 | 3,116,000 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ n$ | 4,450,000 | 1,688,000 | 1,274,000 | 255,000 | 1,019,000 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$ | 827,000 | 372,000 | 342,000 | 17,000 | 325,000 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ n \pi^0$ | 791,000 | 366,000 | 295,000 | 48,000 | 247,000 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^- X$ | 2,270,000 | 176,000 | 153,000 | 9,000 | 44,000 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ n \pi \pi X$ | 2,324,000 | 280,000 | 220,000 | 35,000 | 185,000 |
| $\bar{\nu}_\mu$ CC inclusive | 13,000,000 | 2,882,000 | 2,284,000 | 364,000 | 1,920,000 |

tracking volume around 40 m^3 , depending upon the specific configuration of the main STT parameters. The distributions of the generic kinematic variables $\vec{x} \equiv (x_1, x_2, \dots, x_n)$ in $\nu(\bar{\nu})$ -H CC interactions are obtained as:

$$N_H(\vec{x}) \equiv N_{\text{CH}_2}(\vec{x}) - \frac{M_{C/\text{CH}_2}}{M_C} N_C(\vec{x}) \quad (11)$$

where N_{CH_2} and N_C are the numbers of events selected from the polypropylene and graphite targets, respectively. The interactions from this latter are normalized by the ratio between the total fiducial masses of C within the graphite and CH_2 targets, $M_{C/\text{CH}_2}/M_C$. The subtraction in Eq. (11) is performed after all the selection cuts, including the kinematic analysis described in Sect. 3 and resulting in the purities and efficiencies summarized in Table 1. Practical considerations require the graphite targets to be smaller than the actual mass of C inside the CH_2 plastic, thus resulting in a statistical penalty associated with the subtraction procedure. Figure 8 illustrates how the total statistical uncertainty on N_H from Eq. (11) compares to the ideal one expected from a pure H_2 sample equivalent to the statistics of H interactions within CH_2 . For a given efficiency, the purity of the H samples achievable by the kinematic selection can drastically reduce the overall statistical uncertainty from the subtraction technique. Our analysis suggests that a fiducial mass for the graphite targets around 600 kg corresponding to $M_C/M_{\text{CH}_2} \sim 0.12$ provides a reasonable compromise with a statistical penalty around 30%. We note that this statistical penalty can be further reduced by analytically smoothing the measured distributions from the graphite target and/or by using a tighter kinematic selection, as illustrated in Fig. 8.

We consider the beam spectra expected in the LBNF project [35] and assume a fixed exposure of 5.5×10^{21} protons on target (POT) for both the low-energy neutrino and antineutrino beams, achievable in about two years with the default beam power of 1.2 MW. Table 2 summarizes the total number of events expected for the various topologies and targets. An interesting option available at LBNF is a high-energy beam optimized to detect the ν_τ appearance from neutrino oscillations in the far detector, which would result in an increase by a factor 2.4 of the $\nu(\bar{\nu})$ -H rates, combined with a much harder spectrum. It is conceivable to have a dedicated two year run with such a high-energy beam after the completion of the nominal data taking. By that time the planned upgrades of the beam intensity to a nominal power of 2.4 MW would further roughly double the available POT.

4 Discussion

4.1 Systematic uncertainties

The kinematic analysis described in Sect. 3.2 allows the identification of all the various $\nu(\bar{\nu})$ -H CC topologies within the CH_2 target in STT with little residual backgrounds (5–20%) from interactions on the carbon nucleus. The kinematic selection can reduce the statistical uncertainty from the background subtraction procedure (Sect. 3.3) and the impact of systematic uncertainties on the modeling of nuclear effects in carbon [3]. These latter are further suppressed by a model-independent background subtraction using the data obtained from the dedicated graphite target [4, 5], which also provides a pure background sample to build the likelihood functions used in the selection. The detector technology discussed in Sect. 2 allows the integration of a large number CH_2 and C targets, which are configured as thin (1–2% X_0) passive layers with the same equivalent thickness in terms of radiation and nuclear

Table 3 Comparison of the efficiency and purity for the kinematic selection of H interactions from the CH₂ target (Sect. 3.2) with the NuWro [28], GiBUU [29], and GENIE [30] event generators. The same cuts on the likelihood ratios as in Table 1 are used. In all cases, the likelihood functions are built from the default NuWro generator and are not recalculated when different generators are tested. See text for details

| Process | Selection | NuWro | | GiBUU | | GENIE | |
|---|---|----------------|------------|----------------|------------|----------------|------------|
| | | Efficiency (%) | Purity (%) | Efficiency (%) | Purity (%) | Efficiency (%) | Purity (%) |
| $\nu_\mu p \rightarrow \mu^- p \pi^+$ | $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ | 96 | 95 | 96 | 85 | 96 | 96 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ p \pi^-$ | $\ln \lambda^H + \ln \lambda_{\text{IN}}^H$ | 94 | 95 | 94 | 87 | 94 | 98 |
| $\bar{\nu}_\mu p \rightarrow \mu^+ p n$ | $\ln \lambda_{\text{QE}}^H$ | 75 | 80 | 75 | 88 | 75 | 94 |

interaction lengths, and are alternated throughout the detector volume. Detailed detector simulations indicate that in this way the acceptance difference between targets can be kept within 10^{-3} for all particles [12]. The data from the graphite target automatically include all types of interactions, as well as reconstruction effects, relevant for our analysis.

The kinematic selection of ν_μ -H and $\bar{\nu}_\mu$ -H CC interactions described in Sect. 3.2 relies upon the fact that the target proton is at rest and on energy–momentum conservation, rather than on the specific kinematics of the interactions with the free nucleon. Uncertainties on these latter, including form factors and structure functions, affect only marginally the resulting efficiencies through the corresponding kinematic dependence for individual exclusive topologies. Furthermore, uncertainties related to the structure of the free nucleon would be common to both signal and backgrounds, largely canceling in a selection based upon the differences introduced by nuclear effects. We note that the relevant proton form factors and structure functions can be directly determined in a model-independent way from the measured Q^2 and x distributions, using the procedure described in details in Ref. [41].

Although the “solid” hydrogen technique is conceived to be model-independent, in the absence of actual data from STT the efficiencies and purities listed in Table 1 are sensitive to the details of the interaction modeling implemented in the simulations. In order to estimate the impact of such effects, we repeat the event selection described in Sect. 3.2 with three event generators: NuWro [28], GiBUU [29], and GENIE [30]. These generators use rather different assumptions for the modeling of (anti)neutrino–nucleus interactions—including both initial and final state nuclear effects—as outlined in Ref. [31]. To this end, we do not recalculate the likelihood functions but rather use the ones obtained from the default NuWro generator throughout. Given the differences among generators [31, 41], this assumption can help to understand the impact of potential discrepancies between data and simulations on the H selection. As shown in Table 3, our kinematic selection of H interactions from the CH₂ targets is relatively stable across the three event generators. We emphasize that this test is only meant to estimate a possible outer envelope for the numbers in Table 1, since the actual technique is entirely data-driven and all backgrounds and efficiencies will be directly determined from the measured interactions. Furthermore, we will also use data from the graphite targets to build the likelihood functions independently of simulations.

Reconstruction effects on the four-momenta of the final state particles can in principle degrade the kinematic selection. For this reason in our studies, we used a realistic detector smearing and checked its consistency with GEANT4 and FLUKA simulations (Sect. 3.1). Furthermore, we validated the effects of the detector acceptance, smearing, and track reconstruction with NOMAD data [24], although the NOMAD detector had worse acceptance and granularity than the STT. We note that similar kinematic selections were successfully demonstrated by NOMAD⁴ in more severe background conditions (rejections up to 10^5) in various published analyses [25–27], as well as in single track measurements of $\bar{\nu}_\mu$ QE and inverse muon decay [42].

The momentum scale of charged particles can be calibrated with the mass peak of the large samples of reconstructed $K^0 \rightarrow \pi^+ \pi^-$ decays [41]. In our study, we assume the same energy scale uncertainty of 0.2% achieved by the NOMAD experiment using this technique [43]. We note that the STT at LBNF would provide 25 times higher granularity than NOMAD and about 40 times higher K^0 statistics [41]. Similarly, the proton identification and reconstruction efficiency can be accurately calibrated with the large samples of $\Lambda \rightarrow p \pi^-$ decays available [41]. Both Λ and K^0 decays can be used to constrain the systematic uncertainty on the reconstruction of the track angles.

The use of a likelihood function in the kinematic analysis (Sect. 3.2) provides a simple way to vary the purity and efficiency of the selected samples (Fig. 4) to validate the background subtraction and the selection efficiencies through appropriate control samples [25].

4.2 Physics measurements

The availability of high-statistics samples of $\nu(\bar{\nu})$ -H CC interactions would be extremely relevant for neutrino scattering physics, as well as for long-baseline oscillation experiments. In this section, we briefly outline some of the physics measurements [4, 5, 44].

⁴ The higher neutrino energy in NOMAD implies higher backgrounds and more difficult kinematic selections for low multiplicity processes like RES and QE compared to our case study. Furthermore, the lowest usable energy in NOMAD was about 5 GeV, which is reasonably close to the spectra expected at LBNF.

The limited knowledge of the (anti)neutrino flux has always been a major limitation for accelerator-based neutrino experiments. The exclusive $\nu_\mu H \rightarrow \mu^- p \pi^+$ and $\bar{\nu}_\mu H \rightarrow \mu^+ n$ processes with small energy transfer ν offer an excellent tool to measure the relative (anti)neutrino flux as a function of energy with little hadronic uncertainties. The use of the samples described in Sect. 3.3 allows a determination of the ν_μ and $\bar{\nu}_\mu$ relative flux to a precision better than 1% [41] in conventional wide-band beams, which is not achievable with other known techniques using nuclear targets. The $\bar{\nu}_\mu H \rightarrow \mu^+ n$ interactions at small momentum transfer Q also provide an accurate measurement of the absolute $\bar{\nu}_\mu$ flux, since the corresponding cross section in the limit $Q \rightarrow 0$ is known to high accuracy from neutron β decay [41, 45].

A comparison of $\nu(\bar{\nu})$ -H CC interactions with the corresponding ones from nuclear targets within the same detector provides a direct measurement of nuclear effects [3, 21–23], which typically introduce a substantial smearing of the observed interactions. This study can be performed with both inclusive CC events and with various exclusive topologies. Constraining the nuclear smearing from initial and final state interactions is required to reduce the systematic uncertainties in the unfolding of data collected from nuclear targets. To this end, the combined use of ν -H and $\bar{\nu}$ -H CC interactions provides a control sample free from nuclear effects to calibrate the neutrino energy scale in CC interactions [5].

The unique combination of nuclear and “solid” hydrogen targets would enable a broad program of precision measurements and searches for new physics [4, 5] complementary with the ones planned in the collider [46], fixed target [47], and nuclear physics communities. An example is given by the Adler sum rule [48], which is based upon current algebra and was tested only by BEBC [49] with a few thousand events. Similarly, by exploiting the isospin symmetry $F_2^{\nu n} = F_2^{\bar{\nu} p}$, we can obtain a direct determination of the free neutron structure functions, as well as a measurement of the large x behavior of the d/u quark ratio [50, 51]. These measurements can also be used for precision tests of the isospin (charge) symmetry [5] and would help to elucidate the flavor structure of the nucleon [52]. Furthermore, using a combination of both isoscalar and non-isoscalar nuclear targets can provide valuable insights on the physics mechanisms responsible of the nuclear modifications of the nucleon properties [21, 23, 53].

Acknowledgements The authors express their gratitude to the DUNE experiment for the use of some detector simulation tools. We thank L. Camilleri, R. Ent, X. Lu, and X. Qian for fruitful discussions. We thank F. Ferraro, L. Di Noto, P. Sala, and M. Torti for providing the reconstruction smearing of n and π^0 with the FLUKA simulation package.

Funding Open access funding provided by the Carolinas Consortium.

Data Availability Statement No data associated in the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. B. Abi *et al.* (DUNE), (2020), [arXiv:2002.03005](https://arxiv.org/abs/2002.03005) [hep-ex]
2. K. Abe *et al.* (Hyper-Kamiokande), (2018), [arXiv:1805.04163](https://arxiv.org/abs/1805.04163) [physics.ins-det]
3. L. Alvarez-Ruso *et al.*, Prog. Part. Nucl. Phys. **100**, 1 (2018). <https://doi.org/10.1016/j.pnpnp.2018.01.006>. [arXiv:1706.03621](https://arxiv.org/abs/1706.03621) [hep-ph]
4. R. Petti, in *Proceedings, 27th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2019): Torino, Italy, April 8-12, 2019* (2019) PoS (DIS2019) 235, [arXiv:1910.05995](https://arxiv.org/abs/1910.05995) [hep-ex]
5. R. Petti, Phys. Lett. B **834**, 137469 (2022). <https://doi.org/10.1016/j.physletb.2022.137469>. [arXiv:2205.10396](https://arxiv.org/abs/2205.10396) [hep-ph]
6. S.J. Barish *et al.*, Phys. Rev. D **16**, 3103 (1977). <https://doi.org/10.1103/PhysRevD.16.3103>
7. G. Fanourakis, L.K. Resvanis, G. Grammatikakis, P. Tsilimigras, A. Vayaki, U. Camerini, W.F. Fry, R.J. Loveless, J.H. Mapp, D.D. Reeder, Phys. Rev. D **21**, 562 (1980). <https://doi.org/10.1103/PhysRevD.21.562>
8. M. Derrick *et al.*, Phys. Rev. D **25**, 624 (1982). <https://doi.org/10.1103/PhysRevD.25.624>
9. J. Bell *et al.*, Phys. Rev. D **19**, 1 (1979). <https://doi.org/10.1103/PhysRevD.19.1>
10. G. T. Jones *et al.* (WA21), Z. Phys. C **46**, 25 (1990) <https://doi.org/10.1007/BF02440830>
11. M.A. Parker *et al.*, BEBC TST Neutrino. Nucl. Phys. B **232**, 1 (1984). [https://doi.org/10.1016/0550-3213\(84\)90358-4](https://doi.org/10.1016/0550-3213(84)90358-4)
12. R. Petti, (2023), [arXiv:2301.04744](https://arxiv.org/abs/2301.04744) [hep-ex]
13. R. Petti, T2K-ND280 Workshop, Rome, 6-8 December (2004)
14. M. Adinolfi *et al.*, Nucl. Instrum. Meth. A **482**, 364 (2002). [https://doi.org/10.1016/S0168-9002\(01\)01502-9](https://doi.org/10.1016/S0168-9002(01)01502-9)
15. A. Sergi, *Proceedings, 2nd International Conference on Technology and Instrumentation in Particle Physics 2011: Chicago, Illinois, 9-14 Jun 2011*, Phys. Procedia **37**, 530 (2012). <https://doi.org/10.1016/j.phpro.2012.03.713>
16. H. Nishiguchi *et al.*, *Proceedings, 14th Vienna Conference on Instrumentation (VCI 2016): Vienna, Austria, February 15-19, 2016*, Nucl. Instrum. Meth. A **845**, 269 (2017). <https://doi.org/10.1016/j.nima.2016.06.082>
17. M. Lee, Nuclear and Particle Physics Proceedings **273-275**, 2530 (2016), 37th International Conference on High Energy Physics (ICHEP) <https://doi.org/10.1016/j.nuclphysbps.2015.09.448>
18. M. Anelli *et al.* (SHiP), (2015), [arXiv:1504.04956](https://arxiv.org/abs/1504.04956) [physics.ins-det]
19. E. Abat *et al.*, JINST **3**, P10003 (2008). <https://doi.org/10.1088/1748-0221/3/10/P10003>
20. W. Erni *et al.*, PANDA. Eur. Phys. J. A **49**, 25 (2013). <https://doi.org/10.1140/epja/i2013-13025-8>. [arXiv:1205.5441](https://arxiv.org/abs/1205.5441) [physics.ins-det]
21. S.A. Kulagin, R. Petti, Nucl. Phys. A **765**, 126 (2006). <https://doi.org/10.1016/j.nuclphysa.2005.10.011>. [arXiv:hep-ph/0412425](https://arxiv.org/abs/hep-ph/0412425) [hep-ph]

22. S.A. Kulagin, R. Petti, Phys. Rev. D **76**, 094023 (2007). <https://doi.org/10.1103/PhysRevD.76.094023>. arXiv:hep-ph/0703033 [hep-ph]
23. S.A. Kulagin, R. Petti, Phys. Rev. C **90**, 045204 (2014). <https://doi.org/10.1103/PhysRevC.90.045204>. arXiv:1405.2529 [hep-ph]
24. J. Altegoer et al., NOMAD. Nucl. Instrum. Meth. **A404**, 96 (1998). [https://doi.org/10.1016/S0168-9002\(97\)01079-6](https://doi.org/10.1016/S0168-9002(97)01079-6)
25. P. Astier et al., NOMAD. Nucl. Phys. B **611**, 3 (2001). [https://doi.org/10.1016/S0550-3213\(01\)00339-X](https://doi.org/10.1016/S0550-3213(01)00339-X). arXiv:hep-ex/0106102 [hep-ex]
26. D. Naumov et al., NOMAD. Nucl. Phys. B **700**, 51 (2004). <https://doi.org/10.1016/j.nuclphysb.2004.09.013>. arXiv:hep-ex/0409037 [hep-ex]
27. P. Astier et al., NOMAD. Phys. Lett. B **483**, 387 (2000). [https://doi.org/10.1016/S0370-2693\(00\)00611-0](https://doi.org/10.1016/S0370-2693(00)00611-0)
28. C. Juszczak, J. A. Nowak, J. T. Sobczyk, *NuInt05, proceedings of the 4th International Workshop on Neutrino-Nucleus Interactions in the Few-GeV Region, Okayama, Japan, 26-29 September 2005*, Nucl. Phys. Proc. Suppl. **159**, 211 (2006), [211(2005)], arXiv:hep-ph/0512365 [hep-ph] <https://doi.org/10.1016/j.nuclphysbps.2006.08.069>
29. O. Buss, T. Gaitanos, K. Gallmeister, H. van Hees, M. Kaskulov, O. Lalakulich, A.B. Larionov, T. Leitner, J. Weil, U. Mosel, Phys. Rept. **512**, 1 (2012). <https://doi.org/10.1016/j.physrep.2011.12.001>. arXiv:1106.1344 [hep-ph]
30. C. Andreopoulos et al., Nucl. Instrum. Meth. **A614**, 87 (2010). <https://doi.org/10.1016/j.nima.2009.12.009>. arXiv:0905.2517 [hep-ph]
31. U. Mosel, arXiv:1904.11506 [hep-ex] (2019)
32. X.G. Lu et al., MINERvA. Phys. Rev. Lett. **121**, 022504 (2018). <https://doi.org/10.1103/PhysRevLett.121.022504>. arXiv:1805.05486 [hep-ex]
33. K. Abe et al., T2K. Phys. Rev. D **98**, 032003 (2018). <https://doi.org/10.1103/PhysRevD.98.032003>. arXiv:1802.05078 [hep-ex]
34. J. Rout, S. Roy, M. Masud, M. Bishai, P. Mehta, Phys. Rev. D **102**, 116018 (2020). <https://doi.org/10.1103/PhysRevD.102.116018>. arXiv:2009.05061 [hep-ph]
35. LBNF Fluxes, <https://glauco.crc.nd.edu/DUNEFluxes/> (2019)
36. S. Agostinelli et al., GEANT4. Nucl. Instrum. Meth. **A506**, 250 (2003). [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8)
37. M. Anfreville et al., Nucl. Instrum. Meth. **A481**, 339 (2002). [https://doi.org/10.1016/S0168-9002\(01\)01371-7](https://doi.org/10.1016/S0168-9002(01)01371-7). arXiv:hep-ex/0104012
38. G. Battistoni et al., Annals of Nuclear Energy **82**, 10 (2015)
39. X.G. Lu, D. Coploue, R. Shah, G. Barr, D. Wark, A. Weber, Phys. Rev. D **92**, 051302 (2015). <https://doi.org/10.1103/PhysRevD.92.051302>. arXiv:1507.00967 [hep-ex]
40. M. Anelli et al., KLOE. Nucl. Instrum. Meth. **A598**, 244 (2009). <https://doi.org/10.1016/j.nima.2008.08.120>
41. H. Duiyang, B. Guo, S.R. Mishra, R. Petti, Phys. Lett. B **795**, 424 (2019). <https://doi.org/10.1016/j.physletb.2019.06.003>. arXiv:1902.09480 [hep-ph]
42. V. Lyubushkin et al., NOMAD. Eur. Phys. J. C **63**, 355 (2009). <https://doi.org/10.1140/epjc/s10052-009-1113-0>. arXiv:0812.4543 [hep-ex]
43. Q. Wu et al., NOMAD. Phys. Lett. B **660**, 19 (2008). <https://doi.org/10.1016/j.physletb.2007.12.027>. arXiv:0711.1183 [hep-ex]
44. R. Petti, Workshop on Near Detector Physics at Neutrino Experiments, CERN, 18-22 June 2018
45. R. Petti, R.J. Hill, O. Tomalak, Phys. Rev. D **109**, L051301 (2024). <https://doi.org/10.1103/PhysRevD.109.L051301>. arXiv:2309.02509 [hep-ph]
46. R. Abdul Khalek et al., Nucl. Phys. A **1026**, 122447 (2022). <https://doi.org/10.1016/j.nuclphysa.2022.122447>. arXiv:2103.05419 [physics.ins-det]
47. J. Dudek et al., Eur. Phys. J. A **48**, 187 (2012). <https://doi.org/10.1140/epja/i2012-12187-1>. arXiv:1208.1244 [hep-ex]
48. S.L. Adler, Phys. Rev. **135**, B963 (1964). <https://doi.org/10.1103/PhysRev.135.B963>
49. D. Allasia et al., Z. Phys. C **28**, 321 (1985). <https://doi.org/10.1007/BF01413595>
50. S.I. Alekhin, S.A. Kulagin, R. Petti, Phys. Rev. D **96**, 054005 (2017). <https://doi.org/10.1103/PhysRevD.96.054005>. arXiv:1704.00204 [nucl-th]
51. S.I. Alekhin, S.A. Kulagin, R. Petti, Phys. Rev. D **105**, 114037 (2022). <https://doi.org/10.1103/PhysRevD.105.114037>. arXiv:2203.07333 [hep-ph]
52. S. Alekhin, J. Blümlein, S. Kulagin, S.-O. Moch, R. Petti, in *26th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2018) Port Island, Kobe, Japan, April 16-20, 2018* (2018) arXiv:1808.06871 [hep-ph]
53. S.I. Alekhin, S.A. Kulagin, R. Petti, Phys. Rev. D **107**, L051506 (2023). <https://doi.org/10.1103/PhysRevD.107.L051506>. arXiv:2211.09514 [hep-ph]