

Measurement of neutron production in atmospheric neutrino interactions at Super-Kamiokande

S. Han²⁸, K. Abe,^{1,49} S. Abe,^{1,49} Y. Asaoka,^{1,49} C. Bronner,¹ M. Harada,¹ Y. Hayato,^{1,49} K. Hiraide,^{1,49} K. Hosokawa,¹ K. Ieki,^{1,49} M. Ikeda,^{1,49} J. Kameda,^{1,49} Y. Kanemura,¹ R. Kaneshima,¹ Y. Kashiwagi,¹ Y. Kataoka,^{1,49} S. Miki,¹ S. Mine,^{1,7} M. Miura,^{1,49} S. Moriyama,^{1,49} M. Nakahata,^{1,49} S. Nakayama,^{1,49} Y. Noguchi,¹ G. Pronost,¹ K. Sato,¹ K. Okamoto,¹ H. Sekiya,^{1,49} H. Shiba,¹ K. Shimizu,¹ R. Shinoda,¹ M. Shiozawa,^{1,49} Y. Sonoda,¹ Y. Suzuki,¹ A. Takeda,^{1,49} Y. Takemoto,^{1,49} A. Takenaka,¹ H. Tanaka,^{1,49} T. Yano,¹ T. Kajita,^{2,49,23} K. Okumura,^{2,49} R. Nishijima,² T. Tashiro,² T. Tomiya,² X. Wang,² S. Yoshida,² P. Fernandez,³ L. Labarga,³ N. Ospina,³ D. Samudio,³ B. Zaldivar,³ B. W. Pointon,^{6,53} C. Yanagisawa,^{4,35} E. Kearns,^{5,49} J. L. Raaf,⁵ L. Wan,⁵ T. Wester,⁵ J. Bian,⁷ B. Cortez,⁷ N. J. Griskevich,⁷ M. B. Smy,^{7,49} H. W. Sobel,^{7,49} V. Takhistov,^{7,25} A. Yankelevich,⁷ J. Hill,⁸ M. C. Jang,⁹ S. H. Lee,⁹ D. H. Moon,⁹ R. G. Park,⁹ B. S. Yang,⁹ B. Bodur,¹⁰ K. Scholberg,^{10,49} C. W. Walter,^{10,49} A. Beauchêne,¹¹ O. Drapier,¹¹ A. Giampaolo,¹¹ A. Ershova,¹¹ Th. A. Mueller,¹¹ A. D. Santos,¹¹ P. Paganini,¹¹ C. Quach,¹¹ R. Rogly,¹¹ T. Nakamura,¹² J. S. Jang,¹³ L. N. Machado,¹⁴ F. P. Soler,¹⁴ J. G. Learned,¹⁵ K. Choi,¹⁶ N. Iovine,¹⁶ S. Cao,¹⁷ L. H. V. Anthony,¹⁸ D. Martin,¹⁸ N. W. Prouse,¹⁸ M. Scott,¹⁸ Y. Uchida,¹⁸ A. A. Sztuc,¹⁸ V. Berardi,¹⁹ N. F. Calabria,¹⁹ M. G. Catanesi,¹⁹ E. Radicioni,¹⁹ A. Langella,²⁰ G. De Rosa,²⁰ G. Collazuol,²¹ M. Feltre,²¹ F. Jacob,²¹ M. Mattiazzi,²¹ L. Ludovici,²² M. Gonin,²³ L. Périssé,²³ B. Quilain,²³ C. Fujisawa,²⁴ S. Horiuchi,²⁴ A. Kawabata,²⁴ M. Kobayashi,²⁴ Y. M. Liu,²⁴ Y. Maekawa,²⁴ Y. Nishimura,²⁴ R. Okazaki,²⁴ R. Akutsu,²⁵ M. Friend,²⁵ T. Hasegawa,²⁵ T. Ishida,²⁵ T. Kobayashi,²⁵ M. Jakkapu,²⁵ T. Matsubara,²⁵ T. Nakadaira,²⁵ K. Nakamura,^{25,49} Y. Oyama,²⁵ A. Portocarrero Yrey,²⁵ K. Sakashita,²⁵ T. Sekiguchi,²⁵ T. Tsukamoto,²⁵ N. Bhuiyan,²⁶ G. T. Burton,²⁶ F. Di Lodovico,²⁶ J. Gao,²⁶ A. Goldsack,²⁶ T. Katori,²⁶ J. Migenda,²⁶ R. M. Ramsden,²⁶ Z. Xie,²⁶ S. Zsoldos,^{26,49} T. Sone,²⁷ A. T. Suzuki,²⁷ Y. Takagi,²⁷ Y. Takeuchi,^{27,49} S. Wada,²⁷ H. Zhong,²⁷ J. Feng,²⁸ L. Feng,²⁸ J. Hikida,²⁸ J. R. Hu,²⁸ Z. Hu,²⁸ M. Kawaue,²⁸ T. Kikawa,²⁸ M. Mori,²⁸ T. Nakaya,^{28,49} T. V. Ngoc,²⁸ R. A. Wendell,^{28,49} K. Yasutome,²⁸ S. J. Jenkins,²⁹ N. McCauley,²⁹ A. Tarrant,²⁹ P. Mehta,²⁹ M. Fani,³⁰ M. J. Wilking,³⁰ Y. Fukuda,³¹ Y. Itow,^{32,33} H. Menjo,³² Y. Yoshioka,³² K. Ninomiya,³² J. Lagoda,³⁴ S. M. Lakshmi,³⁴ M. Mandal,³⁴ P. Mijakowski,³⁴ Y. S. Prabhu,³⁴ J. Zalipska,³⁴ M. Jia,³⁵ J. Jiang,³⁵ C. K. Jung,³⁵ W. Shi,³⁵ K. Hamaguchi,³⁶ Y. Hino,³⁶ H. Ishino,³⁶ Y. Koshio,^{36,49} F. Nakanishi,³⁶ S. Sakai,³⁶ T. Tada,³⁶ T. Tano,³⁶ T. Ishizuka,³⁷ G. Barr,³⁸ D. Barrow,³⁸ L. Cook,^{38,49} S. Samani,³⁸ D. Wark,^{38,44} A. Holin,³⁹ F. Nova,³⁹ S. Jung,⁴⁰ J. Y. Yang,⁴⁰ J. Yoo,⁴⁰ J. E. P. Fannon,⁴¹ L. Kneale,⁴¹ M. Malek,⁴¹ J. M. McElwee,⁴¹ T. Peacock,⁴¹ P. Stowell,⁴¹ M. D. Thiesse,⁴¹ L. F. Thompson,⁴¹ S. T. Wilson,⁴¹ H. Okazawa,⁴² S. M. Lakshmi,⁴³ S. B. Kim,⁴⁵ E. Kwon,⁴⁵ M. W. Lee,⁴⁵ J. W. Seo,⁴⁵ I. Yu,⁴⁵ A. K. Ichikawa,⁴⁶ K. D. Nakamura,⁴⁶ S. Tairafune,⁴⁶ K. Nishijima,⁴⁷ A. Eguchi,⁴⁸ S. Goto,⁴⁸ S. Kodama,⁴⁸ Y. Mizuno,⁴⁸ T. Muro,⁴⁸ K. Nakagiri,⁴⁸ Y. Nakajima,^{48,49} S. Shima,⁴⁸ N. Taniuchi,⁴⁸ E. Watanabe,⁴⁸ M. Yokoyama,^{48,49} P. de Perio,⁴⁹ S. Fujita,⁴⁹ C. Jesús-Valls,⁴⁹ K. Martens,⁴⁹ Ll. Marti,⁴⁹ K. M. Tsui,⁴⁹ M. R. Vagins,^{49,7} J. Xia,⁴⁹ M. Kuze,⁵⁰ S. Izumiyama,⁵⁰ R. Matsumoto,⁵⁰ K. Terada,⁵⁰ R. Asaka,⁵¹ M. Ishitsuka,⁵¹ H. Ito,⁵¹ Y. Ommura,⁵¹ N. Shigeta,⁵¹ M. Shinoki,⁵¹ M. Sugo,⁵¹ M. Wako,⁵¹ K. Yamauchi,⁵¹ T. Yoshida,⁵¹ Y. Nakano,⁵² F. Cormier,⁵³ R. Gaur,⁵³ V. Gousy-Leblanc,^{53,*} M. Hartz,⁵³ A. Konaka,⁵³ X. Li,⁵³ B. R. Smithers,⁵³ S. Chen,⁵⁴ Y. Wu,⁵⁴ B. D. Xu,⁵⁴ A. Q. Zhang,⁵⁴ B. Zhang,⁵⁴ M. Girgus,⁵⁴ P. Govindaraj,⁵⁴ M. Posiadala-Zezula,⁵⁵ S. B. Boyd,⁵⁶ R. Edwards,⁵⁶ D. Hadley,⁵⁶ M. Nicholson,⁵⁶ M. O'Flaherty,⁵⁶ B. Richards,⁵⁶ A. Ali,^{57,53} B. Jamieson,⁵⁷ S. Amanai,⁵⁸ D. Hamaguchi,⁵⁸ A. Minamino,⁵⁸ Y. Sasaki,⁵⁸ R. Shibayama,⁵⁸ R. Shimamura,⁵⁸ and S. Suzuki⁵⁸

(Super-Kamiokande Collaboration)

¹Kamioka Observatory, Institute for Cosmic Ray Research, University of Tokyo,
Kamioka, Gifu 506-1205, Japan

²Research Center for Cosmic Neutrinos, Institute for Cosmic Ray Research, University of Tokyo,
Kashiwa, Chiba 277-8582, Japan

³Department of Theoretical Physics, University Autònoma Madrid, 28049 Madrid, Spain

⁴Science Department, Borough of Manhattan Community College / City University of New York,
New York, New York, 1007, USA

⁵Department of Physics, Boston University, Boston, Massachusetts 02215, USA

⁶Department of Physics, British Columbia Institute of Technology, Burnaby,
British Columbia V5G 3H2, Canada

⁷Department of Physics and Astronomy, University of California, Irvine,
Irvine, California 92697-4575, USA

⁸Department of Physics, California State University, Dominguez Hills, Carson, California 90747, USA

- ⁹*Institute for Universe and Elementary Particles, Chonnam National University, Gwangju 61186, Korea*
- ¹⁰*Department of Physics, Duke University, Durham, North Carolina 27708, USA*
- ¹¹*Ecole Polytechnique, IN2P3-CNRS, Laboratoire Leprince-Ringuet, F-91120 Palaiseau, France*
- ¹²*Department of Physics, Gifu University, Gifu, Gifu 501-1193, Japan*
- ¹³*GIST College, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea*
- ¹⁴*School of Physics and Astronomy, University of Glasgow, Glasgow, Scotland, G12 8QQ, United Kingdom*
- ¹⁵*Department of Physics and Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA*
- ¹⁶*Center for Underground Physics, Institute for Basic Science (IBS), Daejeon, 34126, Korea*
- ¹⁷*Institute For Interdisciplinary Research in Science and Education, ICISE, Quy Nhon, 55121, Vietnam*
- ¹⁸*Department of Physics, Imperial College London, London, SW7 2AZ, United Kingdom*
- ¹⁹*Dipartimento Interuniversitario di Fisica, INFN Sezione di Bari and Università e Politecnico di Bari, I-70125, Bari, Italy*
- ²⁰*Dipartimento di Fisica, INFN Sezione di Napoli and Università di Napoli, I-80126, Napoli, Italy*
- ²¹*Dipartimento di Fisica, INFN Sezione di Padova and Università di Padova, I-35131, Padova, Italy*
- ²²*INFN Sezione di Roma and Università di Roma “La Sapienza,” I-00185, Roma, Italy*
- ²³*ILANCE, CNRS-University of Tokyo International Research Laboratory, Kashiwa, Chiba 277-8582, Japan*
- ²⁴*Department of Physics, Keio University, Yokohama, Kanagawa, 223-8522, Japan*
- ²⁵*High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan*
- ²⁶*Department of Physics, King’s College London, London, WC2R 2LS, United Kingdom*
- ²⁷*Department of Physics, Kobe University, Kobe, Hyogo 657-8501, Japan*
- ²⁸*Department of Physics, Kyoto University, Kyoto, Kyoto 606-8502, Japan*
- ²⁹*Department of Physics, University of Liverpool, Liverpool, L69 7ZE, United Kingdom*
- ³⁰*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*
- ³¹*Department of Physics, Miyagi University of Education, Sendai, Miyagi 980-0845, Japan*
- ³²*Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi 464-8602, Japan*
- ³³*Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Nagoya, Aichi 464-8602, Japan*
- ³⁴*National Centre For Nuclear Research, 02-093 Warsaw, Poland*
- ³⁵*Department of Physics and Astronomy, State University of New York at Stony Brook, New York 11794-3800, USA*
- ³⁶*Department of Physics, Okayama University, Okayama, Okayama 700-8530, Japan*
- ³⁷*Media Communication Center, Osaka Electro-Communication University, Neyagawa, Osaka, 572-8530, Japan*
- ³⁸*Department of Physics, Oxford University, Oxford, OX1 3PU, United Kingdom*
- ³⁹*Rutherford Appleton Laboratory, Harwell, Oxford, OX11 0QX, United Kingdom*
- ⁴⁰*Department of Physics and Astronomy, Seoul National University, Seoul 151-742, Korea*
- ⁴¹*Department of Physics and Astronomy, University of Sheffield, S3 7RH, Sheffield, United Kingdom*
- ⁴²*Department of Informatics in Social Welfare, Shizuoka University of Welfare, Yaizu, Shizuoka, 425-8611, Japan*
- ⁴³*August Chelkowski Institute of Physics, University of Silesia in Katowice, 75 Pułku Piechoty 1, 41-500 Chorzów, Poland*
- ⁴⁴*STFC, Rutherford Appleton Laboratory, Harwell Oxford, and Daresbury Laboratory, Warrington, OX11 0QX, United Kingdom*
- ⁴⁵*Department of Physics, Sungkyunkwan University, Suwon 440-746, Korea*
- ⁴⁶*Department of Physics, Faculty of Science, Tohoku University, Sendai, Miyagi, 980-8578, Japan*
- ⁴⁷*Department of Physics, Tokai University, Hiratsuka, Kanagawa 259-1292, Japan*
- ⁴⁸*Department of Physics, University of Tokyo, Bunkyo, Tokyo 113-0033, Japan*
- ⁴⁹*Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
- ⁵⁰*Department of Physics, Institute of Science Tokyo, Meguro, Tokyo 152-8551, Japan*
- ⁵¹*Department of Physics, Faculty of Science and Technology, Tokyo University of Science, Noda, Chiba 278-8510, Japan*
- ⁵²*Faculty of Science, University of Toyama, Toyama City, Toyama 930-8555, Japan*
- ⁵³*TRIUMF, 4004 Wesbrook Mall, Vancouver, British Columbia V6T2A3, Canada*
- ⁵⁴*Department of Engineering Physics, Tsinghua University, Beijing, 100084, China*
- ⁵⁵*Faculty of Physics, University of Warsaw, Warsaw, 02-093, Poland*

⁵⁶*Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom*⁵⁷*Department of Physics, University of Winnipeg, MB R3J 3L8, Canada*⁵⁸*Department of Physics, Yokohama National University, Yokohama, Kanagawa, 240-8501, Japan*

(Received 15 May 2025; accepted 23 June 2025; published 7 July 2025)

We present measurements of total neutron production from atmospheric neutrino interactions in water, analyzed as a function of electron-equivalent visible energy over a range of 30 MeV to 10 GeV. These results are based on 4,270 days of data collected by Super-Kamiokande, including 564 days with 0.011 wt% gadolinium added to enhance neutron detection. Neutron signal selection is based on a neural network trained on simulation, with its performance validated using an Am/Be neutron point source. The measurements are compared to predictions from neutrino event generators combined with various hadron-nucleus interaction models, which include an intranuclear cascade model and a nuclear deexcitation model. We observe significant variations in the predictions depending on the choice of hadron-nucleus interaction model. We discuss key factors that contribute to describing our data, such as in-medium effects in the intranuclear cascade and the accuracy of statistical evaporation modeling.

DOI: [10.1103/4d71-d69k](https://doi.org/10.1103/4d71-d69k)

I. INTRODUCTION

A large fraction of neutrino experiments rely on nuclear targets, yet significant uncertainties remain regarding the influence of nucleon correlations on interaction cross sections and particle kinematics. Modeling these “nuclear effects” is particularly critical for GeV-scale neutrino experiments aiming to measure neutrino oscillation parameters, including CP violation and mass ordering.

Outgoing hadrons serve as valuable probes of these effects, with recent advancements in neutrino detectors enabling precise measurements of hadron multiplicities and kinematics. For instance, proton measurements in neutrino-argon interactions have revealed discrepancies between observed and predicted kinematic distributions [1]. However, detecting neutrons in tracking detectors is challenging due to the limited detection efficiency from the small (n, p) reaction cross section [2,3]. This issue is crucial, as inaccuracies in estimating the “missing energy” carried by neutrons can bias key measurements, such as the Dirac CP phase [4].

Neutrons with kinetic energies of a few MeV or lower tend to thermalize and can be detected via the radiative neutron capture (n, γ) reactions with well-defined time-scales and energy signatures, enabling clean signal selection with virtually no energy threshold. Historically, this made neutrons effective tags for antineutrino charged-

current (CC) interactions (e.g., $\bar{\nu}_e p \rightarrow e^+ n$) compared to neutrino interactions (e.g., $\nu_e n \rightarrow e^- p$) [5]. Neutron tagging remains relevant today, for instance, in atmospheric neutrino oscillation analyses, where it enhances sensitivity to both neutrino mass ordering and CP violation, by preventing the cancellation of opposite-sign effects in neutrino and antineutrino oscillation probabilities [6]. It also helps suppress atmospheric neutrino backgrounds in searches for rare events, such as proton decay (e.g., $p \rightarrow e^+ \pi^0$ [7]), which in many cases is not expected to produce neutrons, or inverse beta decay ($\bar{\nu}_e p \rightarrow e^+ n$) induced by supernova $\bar{\nu}_e$, which emits only one neutron.

Accurate prediction of detectable neutrons is essential and requires well-constrained uncertainties. The modeling approach commonly adopted by GeV-scale neutrino experiments is illustrated in Fig. 1. Neutrino event generators sample outgoing hadrons from the initial neutrino interaction, either at the nucleon or quark level, and

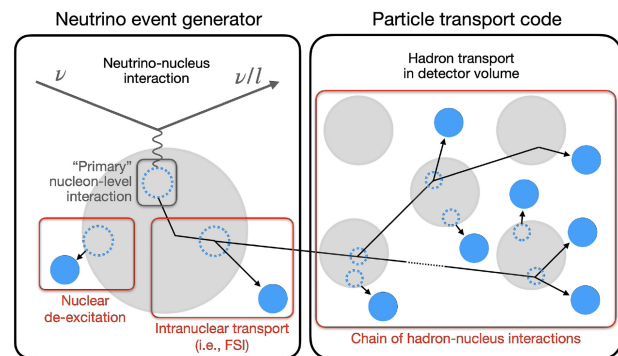


FIG. 1. Schematic illustration of nucleon production mechanisms in a typical GeV-scale neutrino interaction simulation. Black arrows indicate nucleon trajectories; solid blue circles represent detectable nucleons ejected from nuclei, while dashed blue circles denote resulting nucleon holes.

*Also at University of Victoria, Department of Physics and Astronomy, PO Box 1700 STN CSC, Victoria, British Columbia V8W 2Y2, Canada.

subsequently model intranuclear hadron transport (often referred to as final-state interactions, or FSI), followed by nuclear deexcitation. Similarly, particle transport codes such as Geant4 [8–10] simulate hadron transport and nuclear deexcitation to describe downstream hadron interactions within the detector.

Accurately modeling secondary neutron production is particularly important. For hadron transport in the $O(0.1\text{--}1)$ GeV energy range, intranuclear cascade (INC) models [11] are commonly used. Subsequent nuclear deexcitation involves an evaporation process that releases neutrons with kinetic energies of a few MeV and also contributes significantly to total neutron production. Variations in how these models account for nuclear effects often lead to significant discrepancies in predictions [12]. Several studies have measured neutron production from atmospheric or artificial neutrino interactions using water (T2K [13]), heavy water (SNO [14]), and hydrocarbon (MINERvA [2,15], KamLAND [16]) as target materials. Several of these studies [2,13,15] reported deficits in observed neutron signals compared to predictions from neutrino event generators, with the discrepancies often attributed to the inaccuracy of hadron transport models.

In this paper, we present a measurement of neutron production following atmospheric neutrino interactions in water. Using data collected by the Super-Kamiokande (SK) detector from 2008 to 2022, we evaluated the average multiplicity of (n, γ) reactions (“neutron captures”) as a function of electron-equivalent visible energy, a calorimetric proxy for neutrino momentum transfer. The results were compared with predictions from various models relevant to secondary neutron production. This study focuses on an event sample with visible energy greater than 30 MeV, distinct from our previous neutron measurement [17] targeting neutral-current quasielastic (NCQE) events with visible energy below 30 MeV.

This paper is structured as follows. Section II provides a brief overview of the SK detector. The selection process for atmospheric neutrino events and neutron signals, along with the estimation of selection performance, are detailed in Secs. III and IV. Section V outlines the methodology for determining the average (n, γ) multiplicity per visible energy bin and the associated systematic uncertainties. Section VI introduces the interaction models used for generating predictions. Finally, Secs. VII and VIII present a comparison of observations with predictions and discuss the implications of the results.

II. THE SUPER-KAMIOKANDE DETECTOR

Super-Kamiokande (SK) [18] is an underground water Cherenkov detector located in Gifu, Japan. It consists of two optically separated, concentric cylindrical volumes: the inner detector (ID) containing 32.5 ktons of water and equipped with 11,129 inward-facing photomultiplier tubes (PMTs) and the outer detector (OD) serving as a cosmic-ray veto. The

detector registers a PMT signal with a pulse height greater than 0.25 photoelectron-equivalent charge as a “hit.” If the number of ID or OD PMT hits within a 200-ns sliding time window ($N_{200\text{-ns}}$) exceeds a given threshold, an event trigger is issued. The details of the detector can be found in [18,19].

Charged particles, namely electrons and muons produced by charged-current neutrino interactions, are identified through Cherenkov radiation. The radiation is projected onto the PMTs as a characteristic ring pattern that depends on the particle type and energy. This ring pattern serves as the basis for particle reconstruction. Neutrons are indirectly identified via Compton-scattered electrons resulting from (n, γ) reactions. In pure water, most occur on ^1H , emitting a single 2.2 MeV γ -ray. With the recent addition of gadolinium (Gd), a large fraction of neutrons are expected to be captured by Gd isotopes, resulting in a total γ -radiated energy of around 8 MeV.

The $O(1)$ MeV signal identification performance is significantly influenced by variations in detector characteristics. Parameters such as individual PMT gain, timing properties, quantum efficiency, and optical absorption and scattering in water are continuously monitored using cosmic-ray muons and light sources [19]. Additionally, the uncertainty in Cherenkov ring energy reconstruction (described in Sec. III B) is evaluated over a wide energy range using naturally occurring particles, including cosmic-ray muons, Michel electrons, and neutral pions produced in neutral-current (NC) atmospheric neutrino interactions in water. Figure 2 illustrates the agreement between data and simulation in energy reconstruction for the Gd-loaded SK-VI phase, which is mostly within 2% across the $O(10\text{--}10^4)$ MeV range and consistent with the pure water phase results reported in the recent atmospheric neutrino oscillation analysis [6].

The detector has operated through eight different phases. Neutron detection began with the fourth phase SK-IV, following the electronics upgrade [20] that allowed extended event recording up to 535 μs after certain ID triggers. This has enabled analysis of delayed neutron captures that occur with a timescale of $O(10\text{--}100)$ μs following an atmospheric neutrino interaction. Between SK-IV and SK-V, in 2018, the detector underwent refurbishment, during which malfunctioning PMTs were replaced. The later phases, SK-VI, SK-VII, and SK-VIII involved the dissolution of $\text{Gd}_2(\text{SO}_4)_3$ into the water volume to enhance neutron detection efficiency [21,22]. Table I summarizes the relevant operational conditions.

III. ATMOSPHERIC NEUTRINO EVENTS

A. Event selection

We followed a typical selection process for atmospheric neutrino interactions that are fully contained within the ID, similar to previous studies conducted at SK [6]. All events were required to pass the ID trigger with the threshold

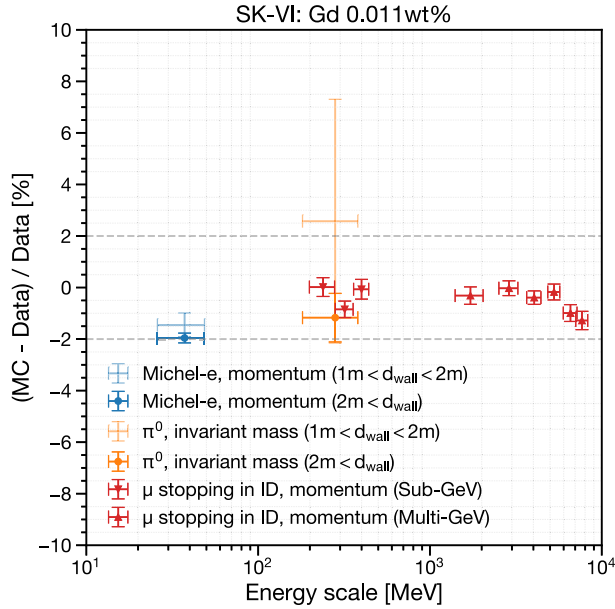


FIG. 2. Comparison of energy reconstruction performance between Monte Carlo (MC) simulation and observed data for naturally occurring particles in the Gd-loaded SK-VI phase. The variable d_{wall} represents the distance (in meters) from the reconstructed vertex to the nearest ID photodetector wall. The fiducial volume for this study is defined as $d_{\text{wall}} > 1$ m.

$N_{200-\text{ns}} \geq 58$ PMT hits—roughly corresponding to a 10 MeV electron—followed by the extended event window of 535 μs for neutron detection. Background events from cosmic-ray muons, radioactivity, and neutrino interactions with exiting particles were reduced using OD veto and ID charge cuts.

Selected events were reconstructed as described in Sec. III B. To further reject low-energy backgrounds, we required that the reconstructed vertex be more than 1 m away from the ID tank wall (defining the fiducial volume with 27.2 kton of water) and that the visible energy be larger than 30 MeV. The remaining background contamination, mainly due to cosmic-ray muons stopping in the ID and PMT discharges, was estimated to be below 0.2%, based on visual inspection [7].

B. Reconstruction of prompt Cherenkov rings

Figure 3 shows a simulated PMT hit time distribution of a typical $\bar{\nu}_\mu$ charged-current (CC) quasielastic (QE) event followed by a Michel electron and a $\text{Gd}(n, \gamma)$ reaction, along with the corresponding event displays. For the “prompt” radiation due to charged particles (namely, electrons and muons) produced via the primary neutrino interaction, we followed the Cherenkov ring reconstruction process [24] as applied in previous SK analyses [6,7].

The visible energy of an event is defined as the sum of the reconstructed kinetic energies of all Cherenkov rings, assuming each ring originates from an electron. This calorimetric measure serves as a reliable proxy for neutrino momentum transfer, as demonstrated by the positive correlation shown in Fig. 4.

C. Baseline simulation setup

The simulation of atmospheric neutrino events involves a convolution of the atmospheric neutrino flux, neutrino event generator, and detector simulation.

To determine the incoming neutrino kinematics and event rate, we used atmospheric flux calculations for ν_e , ν_μ , $\bar{\nu}_e$, and $\bar{\nu}_\mu$ at the detector site, assuming no oscillations, as provided by Honda *et al.* [28] for $[10^2, 10^7]$ MeV range. Neutrinos with energies below 100 MeV, as well as the ν_τ and $\bar{\nu}_\tau$ interactions, were neglected as their overall presence in data is expected to be small.

Neutrino interactions in water were modeled using neutrino event generators [29], which compute the cross sections for each interaction channel and sample outgoing particle kinematics. NEUT 5.4.0 [30] was used as the baseline event generator. It covers major interaction channels, including QE scattering with single (1p1h) or double (2p2h) nucleon knockout, single pion production due to Δ resonance (RES), and multiple pion production via deep inelastic scattering (DIS). The models used are consistent with the latest SK atmospheric neutrino oscillation analysis [6]. The transport of hadronic final states within the target nucleus (FSI) was modeled separately

TABLE I. SK operational phases and neutron-related characteristics. SK-IV, V, VI data were used in this analysis.

Phase	Years	Livetime	Gd concentration ^a	Expected (n, γ) fraction ^b		(n, γ) time constant ^c
		[days]	[wt%]	$H(n, \gamma)$ [%]	$\text{Gd}(n, \gamma)$ [%]	[μs]
SK-I-III	1996–2008	2805.9	...	> 99.9	...	No data
SK-IV	2008–2018	3244.4	...	> 99.9	...	204.8 ± 19.8
SK-V	2019–2020	461.0	...	> 99.9	...	199.8 ± 10.2
SK-VI	2020–2022	564.4 ^d	0.0110 ± 0.0001 [21]	56.1 ± 1.5	43.9 ∓ 1.5	116.2 ± 12.3
SK-VII-VIII	2022–present	...	0.0332 ± 0.0002 [22]	29.7 ± 0.7	70.3 ∓ 0.7	61.8 ± 0.1 [22]

^aBased on the amount of dissolved Gd.

^bBased on the evaluated thermal (n, γ) reaction cross sections and uncertainties of ENDF/B-VII.1 [23].

^cWeighted mean of all Am/Be neutron source measurements, explained in Sec. IV B.

^dExcludes earlier runs which showed signs of nonuniform Gd concentration, i.e., varying time constant by position.

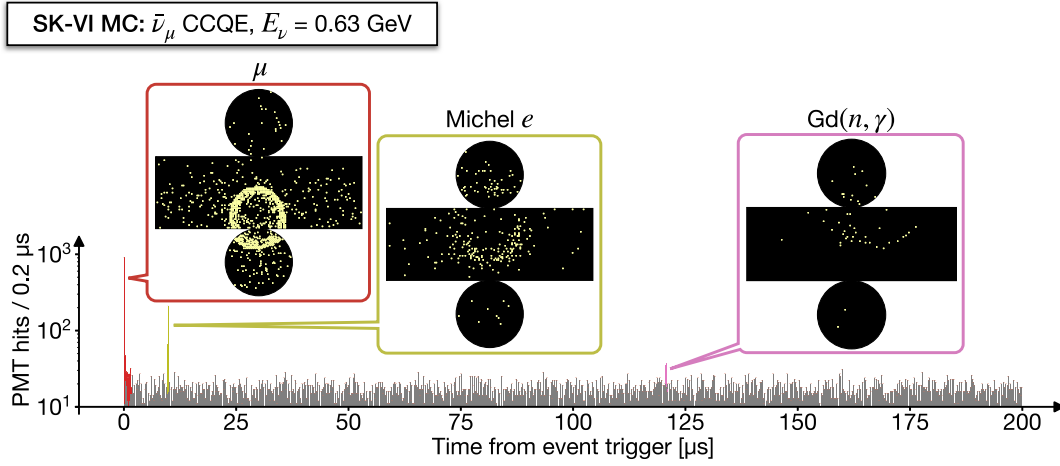


FIG. 3. A PMT hit time distribution for a typical $\bar{\nu}_\mu$ charged-current (CC) quasielastic (QE) interaction with a muon and a neutron as final states. The zero is set at the event trigger. The event displays feature the “prompt” muon signal (red) and the two types of “delayed” coincident signals—Michel electrons from muon decay (olive) and neutron captures on Gd (pink). The gray bars represent randomly recorded background PMT hits. This event was simulated in the SK-VI configuration.

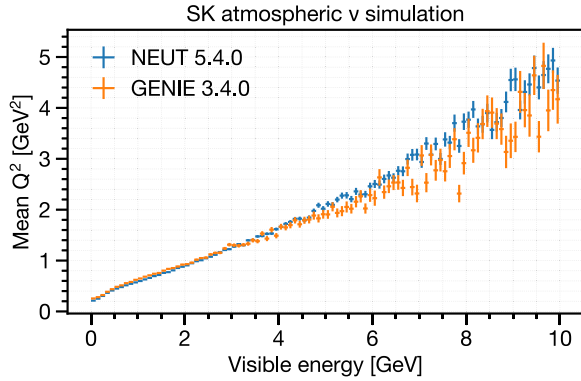


FIG. 4. Average squared momentum transfer (Q^2) per visible energy bin for simulated atmospheric neutrino events at SK, compared between two different neutrino event generators. NEUT 5.4.0 (blue) uses the baseline setup described in Sec. III C, while GENIE 3.4.0 [25–27] uses the “hN” setup as described in Sec. VI.

from the primary neutrino interaction. The transport of low-momentum (<500 MeV/c) pions was based on Salcedo *et al.* [31], with parameters tuned to fit external pion-nucleus scattering data [32]. For higher energy nucleons and pions, FSIs were modeled using the INC approach, which includes elastic scattering and single/double pion production. The cross sections for reactions with free target nucleons were sourced from Bertini [33] for nucleon projectiles and from external pion scattering data for high-momentum pion projectiles [34]. The deexcitation of an oxygen target was modeled using tabulated occupation probabilities for nucleon energy states [35] and branching ratios for knockout of each state [36].

The subsequent particle transport in water and detector responses were simulated using GCALEOR [37] coupled with GEANT 3.21 [38]. Propagation of nucleons and charged pions below 10 GeV was simulated based on the Bertini cascade model [39]. For cross sections, GCALEOR uses ENDF/B-VI [40] for low-energy (below 20 MeV, or 195 MeV/c in momentum) neutrons, and Bertini-based tabulation [39] for the remaining nucleons and charged pions. The transport of low-momentum pions was separately modeled using the NEUT pion FSI routine to ensure consistency between NEUT and GEANT.

For the transport of low-energy neutrons in the Gd-loaded SK-VI phase, we imported reaction cross sections from the GEANT 4.10.5.p01 [8–10] NeutronHP model [41] based on ENDF/B-VII.1 [23], replacing ENDF/B-VI for neutron energies below 20 MeV. Additionally, we modeled the γ -cascade resulting from neutron captures on $^{155/157}\text{Gd}$ using the ANNRI-Gd model [42,43].

The characteristics of individual PMTs and the optical parameters in water were adjusted to align with calibration data obtained from light sources and through-going cosmic-ray muons. To accurately account for detector noise, randomly recorded PMT hits were included as background, represented as gray bars in Fig. 3.

Five hundred years of atmospheric neutrino events were simulated for each SK phase and processed as described in Secs. III B and III A. The events in the final sample were weighted based on the standard three-flavor oscillation probability in matter [44], using the oscillation parameters fitted with reactor constraints in the previous SK analysis [45] and the PREM model [46] for Earth’s matter density. Corrections for atmospheric neutrino flux accounting for solar activity were also applied.

Figure 5 shows the visible energy distributions of the simulated event sample. Single-ring events exhibit a higher

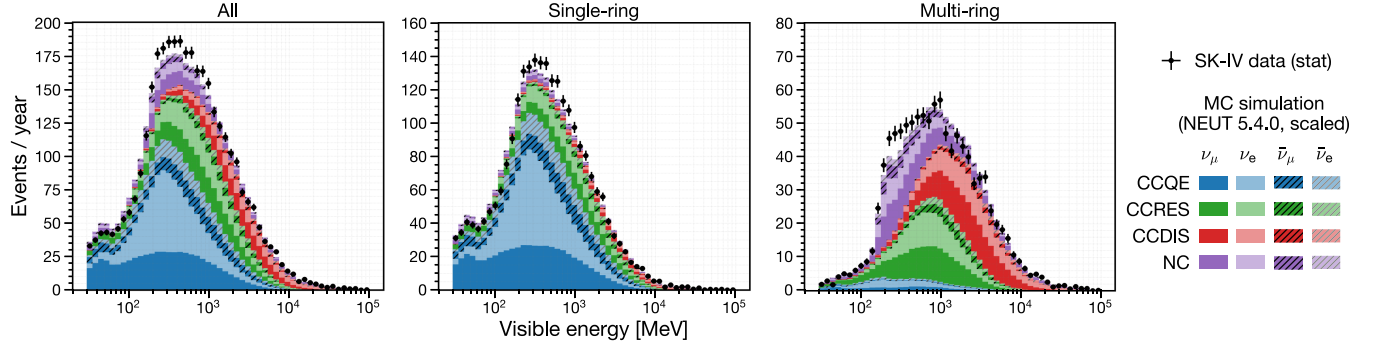


FIG. 5. Observed and simulated visible energy distributions for selected atmospheric neutrino interactions, for the longest SK-IV phase: all events (left), events with one reconstructed Cherenkov ring (middle), and events with two or more rings (right). CCQE events dominate the single-ring sample, while CCRES and CCDIS events dominate the multiring sample. NC events in the multiring sample are mostly RES or DIS interactions. The simulation setup is detailed in Sec. III C. Simulated distributions are scaled to match the SK-IV data to account for the uncertainty in absolute neutrino flux normalization.

fraction of QE interactions and a smaller fraction of DIS compared to multiring events. For the same visible energy, multiring events are expected to produce more secondary neutrons than single-ring events, due to a higher fraction of DIS interactions.

IV. NEUTRON SIGNAL SELECTION

Neutrons from atmospheric neutrino interactions in the SK ID are mostly captured by ^1H or $^{155/157}\text{Gd}$ within $O(100)$ μs , with capture ratios summarized in Table I. The resulting $O(1)$ MeV γ -rays scatter electrons, producing faint Cherenkov rings on the ID wall (hereafter referred to as the “neutron signal”; see Fig. 3).

Since fast neutrons, γ -rays, and scattered electrons travel only a few tens of centimeters, the associated Cherenkov photons can be reliably treated as originating from a single vertex. This enables time-of-flight (ToF) corrections to suppress random PMT hits from dark noise. The vertex can either be assumed *a priori*—e.g., using the reconstructed lepton vertex—for better resolution and signal efficiency at the cost of potential bias, or reconstructed independently from the lower energy neutron signal (Appendix A), yielding lower resolution and efficiency but greater robustness to bias and kinematic uncertainties. Residual Michel electrons from muon decays are reduced using timing and energy cuts.

A. Signal selection algorithm

The signal selection algorithm is based on Ref. [47], and consists of candidate search and classification stages.

In the candidate search stage, PMT hit times are corrected for photon ToF from an *a priori* vertex given by the reconstructed lepton vertex. A 14-ns sliding time window is applied, with thresholds of 5 hits for SK-IV/V (pure water) and 7 for SK-VI (Gd-loaded), optimized for signal efficiency [48]. Overlapping candidates within 50 ns are resolved by selecting the one with the most PMT hits. In SK-IV/V, the search window was set to $[18, 534]$ μs from

the event trigger, following earlier analyses [6,7,47] designed to avoid PMT afterpulses (<15 μs) and Michel electrons. For SK-VI, it was extended to $[3, 534]$ μs to capture more of the faster neutron captures with Gd, aided by improved candidate classification.

In the candidate classification stage, we extract features of each candidate and use a neural network to classify each candidate into signal and noise based on input features. These features characterize the signal energy, the background hit level, timing spread assuming the vertex, correlation between the input vertex and the hit PMT positions, correlation to the known properties of PMT noise, and angular correlation among hit PMTs relative to the Cherenkov cone opening angle.

The major changes from the original algorithm [47] include a simplified algorithm, a reduced set of features, and a heuristically tuned neural network architecture. These modifications aim to reduce performance bias between the data and the simulation that is used to train the neural networks.

Here, we provide the definition and unit of each feature used for the classification of signal candidates, along with their expected distributions as shown in Fig. 6:

- (i) **NHits**
The number of selected PMT hits within the 14-ns sliding time window.
- (ii) **NResHits**
The number of PMT hits within $[-100, +100]$ ns from the center of the 14-ns sliding time window, minus NHits.
- (iii) **TRMS [ns]**
The root mean square (RMS) of the ToF-corrected time distribution of the selected PMT hits.
- (iv) **FitGoodness**
The normalized likelihood of the ToF-corrected time distribution of the selected PMT hits, given the *assumed* signal vertex and the Gaussian PMT timing resolution of 5 ns.

(v) DWall [cm]

The distance from the *assumed* signal vertex to the nearest tank wall.

(vi) DWallMeanDir [cm]

The shorter of the radial and vertical distances from the *assumed* signal vertex to the tank wall, calculated along the average direction of the unit vectors connecting the vertex to each hit PMT.

(vii) BurstRatio

The ratio of the selected PMTs with a preceding hit within 10 μ s, which are likely caused by scintillation within the irradiated PMT glass.

(viii) DarkLikelihood

The normalized log likelihood ratio based on measured individual PMT dark rates, given by

$$\text{DarkLikelihood} = \sigma \left(\log \prod_{i=1}^{\text{NHits}} \frac{r_i}{\langle r \rangle} \right) \quad (1)$$

where σ represents the sigmoid function, r_i is the dark rate of the i^{th} PMT, and $\langle r \rangle$ is the average dark rate of all ID PMTs.

(ix) OpeningAngleStdev [deg]

The standard deviation of the opening angles of cones formed by every possible combination of three hit PMTs and the *assumed* signal vertex.

(x) Beta(k), $k \in \{1, 2, 3, 4, 5\}$

$$\text{Beta}(k) = \frac{2}{\text{NHits}(\text{NHits} - 1)} \sum_{i \neq j} P_k(\cos \theta_{ij}) \quad (2)$$

where P_k is the k^{th} Legendre polynomial and θ_{ij} is the opening angle between the assumed signal vertex and the i^{th} and j^{th} hit PMTs.

For training the neural network, we utilized the features and labels of first-stage candidates from a particle-gun simulation of thermal neutrons, with vertices randomized within the ID. Each first-stage candidate, as shown in Figs. 6 and 7, was labeled as signal if it was triggered within 50 ns of the simulated (n, γ) reaction, and otherwise as background. Of the training dataset, 80% was used for updating the neural network weights, while the remaining 20% was reserved for validation.

We implemented a feed-forward fully connected neural network using Keras 2.6.0 [49]. The network consisted of an input layer with 14 features, followed by three dense layers, each comprising 128 ReLU-activated nodes with a 50% dropout rate, and a single sigmoid output node. Weights and biases were initialized following He *et al.* [50] and optimized by minimizing the binary cross-entropy

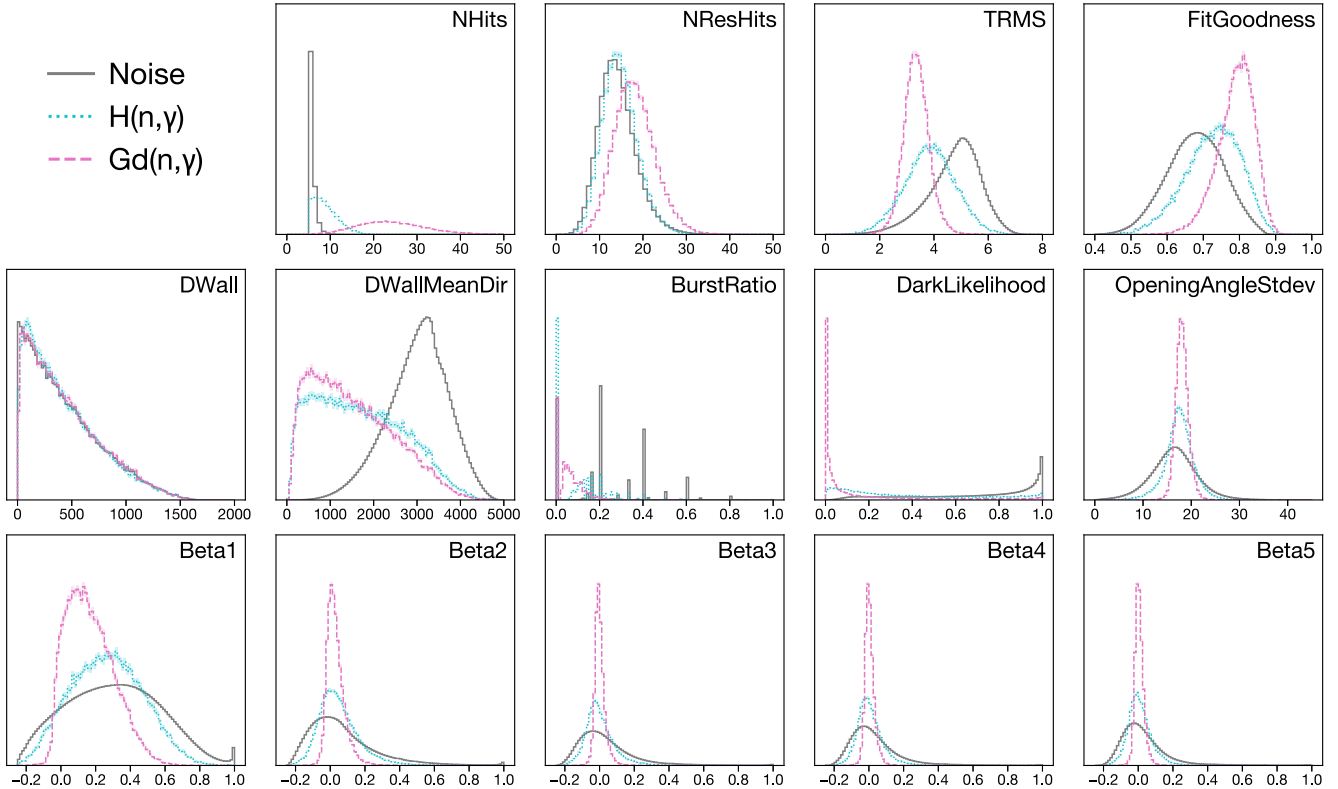


FIG. 6. The features (area-normalized) of neutron signals and noise from the thermal neutron MC simulation used for training the neural network for SK-VI phase.

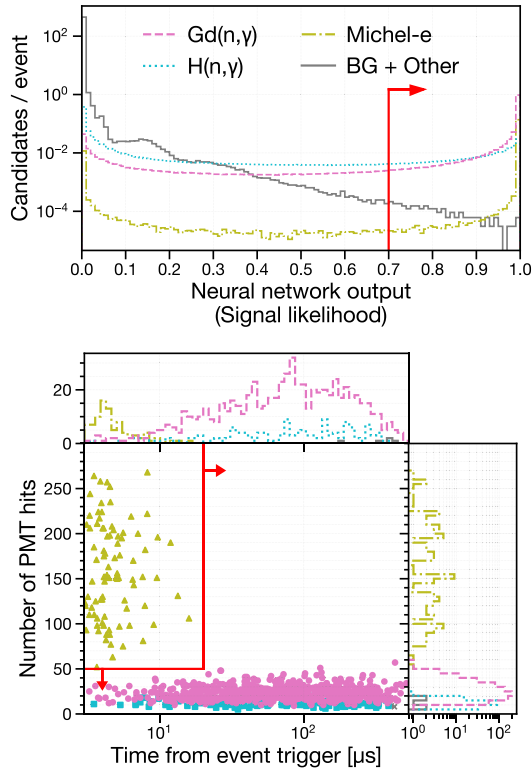


FIG. 7. The neural network response to neutron signals and backgrounds in the SK-VI atmospheric neutrino simulation (top), and the time versus energy distribution of signal candidates passing the neural network selection (bottom). Red arrows indicate the corresponding cut points.

loss iteratively on minibatches of size 2,048 using the Adam optimizer [51]. The initial learning rate was set to 0.0001. Training was stopped when signal efficiency on the validation set showed no improvement for 5 consecutive epochs. A neural network was trained for each SK phase: SK-IV, SK-V, and SK-VI.

Candidates with a neural network output greater than 0.7 were classified as signals, while those with a large number of PMT hits ($N_{\text{Hits}} > 50$) and occurring earlier than the typical neutron capture timescale ($< 20 \mu\text{s}$) were identified as Michel electrons and excluded. The effectiveness of this Michel electron rejection is illustrated in Fig. 7. When applied to cosmic-ray muons decaying within the ID, the selection achieved an efficiency of $98.4 \pm 1.3\%$ and a purity of $98.7 \pm 0.5\%$.

B. Signal selection performance on calibration data

An Am/Be neutron source with a measured total intensity of 236.8 ± 5.0 neutrons/s [52] was used to obtain calibration data for estimating the signal detection performance. The first excited state of the alpha-absorbed ^9Be , with a roughly 60% branching ratio, emits a fast neutron and a 4.44 MeV γ -ray simultaneously. This source was encapsulated with Bismuth Germanate ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$, BGO)

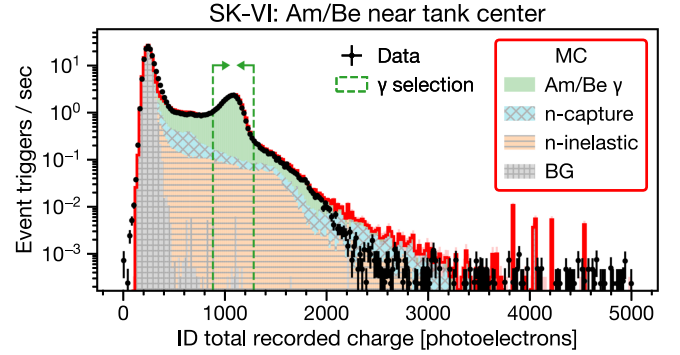


FIG. 8. Distribution of recorded charge within the time window $[-0.5, 1.0] \mu\text{s}$ relative to event triggers, measured with an Am/Be neutron source positioned near the ID tank center. The black points represent SK-VI data, while the red line shows the simulated prediction. The green dashed arrows indicate the selection window for 4.44 MeV γ -ray-induced scintillation. Roughly 95% of the selected events are attributed to γ 's from the Am/Be source ("Am/Be γ "), with the remaining 5% arising from ambient neutron captures ("n-capture") and neutrons inelastically producing charged particles in the scintillator ("n-inelastic"). The contribution from background events ("BG") is negligible. The pink shades indicate MC statistical errors.

crystals so that the 4.44 MeV γ 's can induce scintillation. The setup was deployed in various positions within the ID, and events were recorded for 30 minutes to 1 hour. Events with a trigger charge yield corresponding to the 4.44 MeV γ -ray scintillation were regarded as the single neutron control sample.

The observed light yield distribution was compared with dedicated simulation, as shown in Fig. 8. The simulation accounts for continuous source activity and pile-up, by reorganizing the simulated detector response to Am/Be neutron emission on a single global time axis, based on the measured total neutron intensity and the estimated branching ratios to each excited state of alpha-absorbed ^9Be . As shown in Fig. 8, this simulation accurately models event triggers due to ambient neutron captures and neutron inelastic interactions within scintillator crystal. The contamination of such unwanted event triggers in the 4.44 MeV γ -ray event selection was estimated to be at a few percent level.

Within the selected events in the single neutron control sample, signal candidates were obtained following the algorithm described in Sec. IV A, with the *assumed* signal vertex set at the source position.

Figure 9 shows an example time distribution of the selected signal candidates. Such distributions of the time t were fitted with a function f of the form

$$f(t) = A(1 - e^{-t/\tau_{\text{thermal}}})e^{-t/\tau_{\text{capture}}} + B \quad (3)$$

where the normalization constant A , the background constant B , the neutron thermalization timescale τ_{thermal}

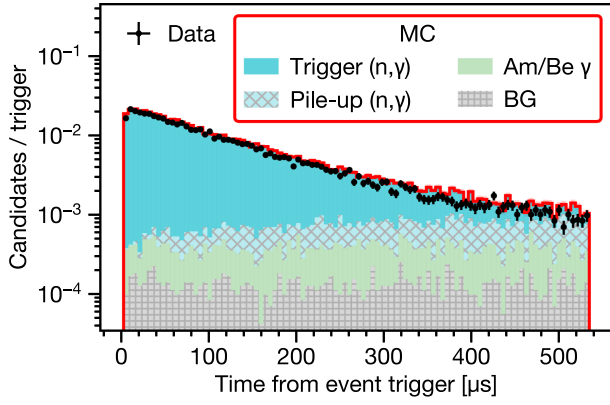


FIG. 9. Exponential decrease of the selected neutron signal candidates as a function of the time from the selected event triggers with the Am/Be neutron source positioned near the ID tank center, in the SK-VI phase. The label “Trigger (n, γ)” indicates captures of neutrons produced within 350 ns from the event trigger, while the label “Pile-up (n, γ)” indicates captures of piled-up neutrons without such correlation to the event trigger. The pink shades indicate MC statistical errors.

(set to 0 for SK-IV/V assuming a velocity-independent neutron capture rate), and the neutron capture time constant τ_{capture} are free parameters. The measured τ_{thermal} in SK-VI was $4.71 \pm 0.04 \mu\text{s}$. The signal efficiency was evaluated as the number of identified signals per selected event triggers, corrected by the constant background term B . Figure 10 shows the estimated neutron detection efficiencies for various source positions in the ID.

The major sources of systematic uncertainty are summarized in Table II. In the pure water phases (SK-IV and V), the dominant source of uncertainty lies in the potential bias caused by the calibration setup, such as the unwanted event triggers or time correlation of false positives to event triggers, often leading to an overestimation of the background constant B . The size of this uncertainty was

TABLE II. Major sources of systematic uncertainty in the neutron detection efficiency estimated with the Am/Be neutron source.

Source	SK-IV	SK-V	SK-VI
Am/Be neutron characterization	0.5%	0.9%	0.5%
Detector response	2.2%	3.3%	1.2%
Bias due to calibration setup	6.9%	4.6%	1.1%
Gd(n, γ) fraction	2.1%
Gd(n, γ) γ emission model	2.6%
Total	7.3%	5.7%	3.8%

conservatively estimated by comparing the true and estimated signal efficiencies from the simulations and quantifying the fluctuation within each SK phase. In the Gd-loaded phase (SK-VI), the dominant source of uncertainty is in the fraction and the γ emission model of the Gd(n, γ) reaction. The size of these uncertainties is estimated based on the evaluated thermal neutron capture cross section uncertainties in ENDF/B-VII.1, as well as variations in the estimated signal efficiency when using an alternative photon strength function to describe the Gd continuum γ -cascade in the ANNRI-Gd model [42].

The discrepancy between observed and predicted signal efficiencies in SK-VI (Fig. 10) likely arises from an overestimation of the Gd(n, γ) fraction in Geant4 [53,54]. For example, the Gd(n, γ) fraction estimated using SK-VI Am/Be data was $(44 \pm 3)\%$ [53], consistent with the $(43.9 \pm 1.5)\%$ evaluated with thermal neutron capture cross sections in ENDF/B-VII.1 [55], yet both were lower than the 52% predicted by GEANT 4.10.5.p01 NeutronHP using the same cross sections. This problem with Geant4 is attributed to its treatment of hydrogen as free rather than bound in water molecules, leading to an underestimation of the competing $^1\text{H}(n, \gamma)$ reaction [54].

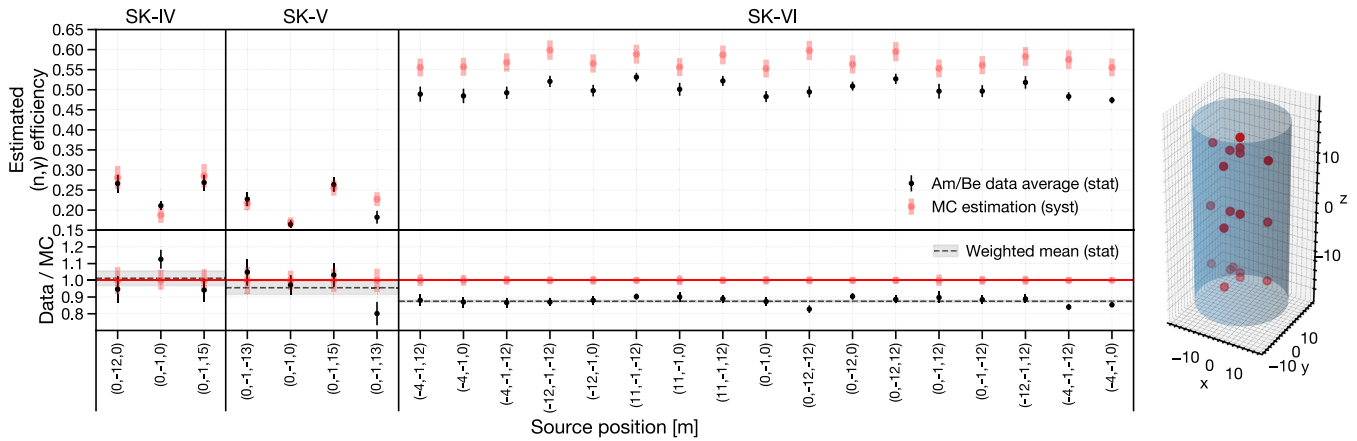


FIG. 10. Estimated (n, γ) signal selection efficiencies for each calibration position within the ID tank (red markers in the schematic on the right). For SK-VI positions with multiple measurements taken on different dates, the reported efficiency is the average of those measurements. Systematic uncertainties considered in the MC simulation are summarized in Table II.

To address this discrepancy, we applied a correction based on the weighted mean of Am/Be data-to-MC efficiency ratios across all source positions. Signal efficiencies in the pure water phases with limited calibration data were further corrected to ensure consistency in the overall neutron counts with the well-calibrated Gd-loaded phase, and the final correction factors were 0.90 ± 0.12 (SK-IV), 0.94 ± 0.04 (SK-V), and 0.88 ± 0.01 (SK-VI) (see Appendix A for details).

The measured τ_{capture} in Eq. (3) were 200.4 ± 3.7 μs (pure water, SK-IV/V) and 116.9 ± 0.3 μs (Gd-loaded, SK-VI), consistent with ENDF/B-VII.1 predictions of 204.7 ± 5.3 μs and 114.9 ± 2.5 μs .

V. (n, γ) MULTIPLICITY ESTIMATION

The average multiplicity of total (n, γ) reactions is computed as the average of the expected number of (n, γ) reactions estimated on an event-by-event basis, as follows:

$$\langle N \rangle = \left\langle \frac{N_i^{\text{detected}} - N_i^{\text{BG}}}{\epsilon_i} \right\rangle \quad (4)$$

Here, N_i^{detected} is the count of detected signals, N_i^{BG} is the estimated number of false positives, and ϵ_i is the estimated signal detection efficiency of the i th event.

Accurate estimation of N_i^{BG} and ϵ_i is crucial. While Am/Be calibration data provides a basis for these estimates, additional factors in atmospheric neutrino events—such as neutrino vertex reconstruction accuracy and larger neutron kinetic energy—may significantly impact the signal selection performance. To better account for these effects in the calculation of N_i^{BG} and ϵ_i , we trained generalized additive models (GAMs) [56] on the baseline atmospheric neutrino simulations. Using pyGAM 0.9.0 [57], a total of six linear GAMs were constructed across the three SK phases, for two output metrics: signal selection efficiency and false positive rate.

Using the baseline MC simulation generated with NEUT 5.4.0, GAMs were fitted to the simulated distributions of each metric, averaged within bins in a five-dimensional feature space. The features consist of reconstructed neutrino event variables, including visible energy, Cherenkov ring multiplicity, the particle type of the most energetic ring, and the radial and vertical displacements of the neutrino interaction vertex. No assumptions were made regarding feature correlations, and appropriate smoothing was applied to mitigate overfitting. The 1σ prediction interval was determined based on the MC statistical uncertainties within each bin.

To evaluate uncertainty in its performance, trained GAMs were deliberately tested on unseen simulation data produced with NEUT 5.1.4, which predicts roughly 10% lower (n, γ) multiplicity overall. Figure 11 shows the performance of the trained GAM in estimating the true

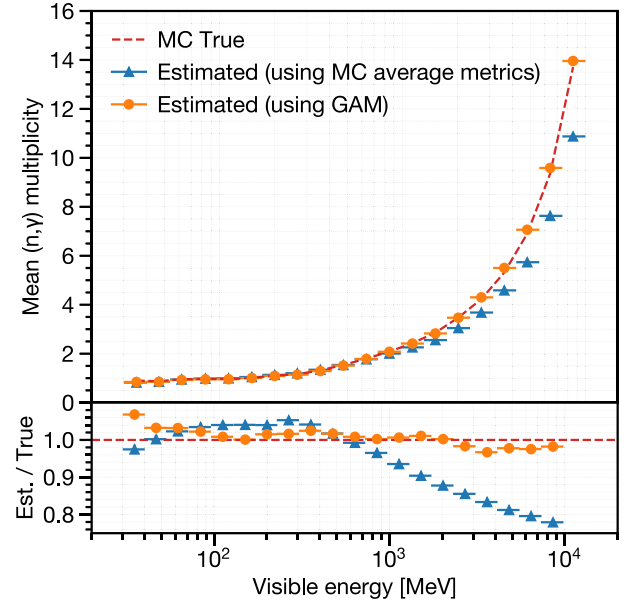


FIG. 11. The true (dashed lines) and estimated average (n, γ) multiplicity (circle markers: using GAM, triangle markers: using the overall average signal efficiency and false positive rate obtained from the given simulation) as a function of visible energy, for the test simulations of atmospheric neutrino events produced with NEUT 5.1.4.

average signal multiplicity per visible energy bin. By reconstructing N_i^{BG} and ϵ_i on an event-by-event basis, the GAM helps reduce potential biases in the results, particularly in the multi-GeV bins.

The following major systematic uncertainties affecting signal counting were evaluated on a bin-by-bin basis for each operational phase and data subsample:

- (1) Overall signal efficiency scale
This includes calibration uncertainties (Sec. IV B) and phase-dependent variations (to be explained in Appendix A).
- (2) Signal selection performance modeling (GAM)
This is quantified as the difference between true and estimated signal multiplicities in simulations. The ratio is also used to correct the estimated distribution.
- (3) Neutrino event reconstruction
This assumes 2% visible energy resolution (Fig. 2), with ring-counting errors accounted for single- and multiring events.

Figure 12 shows fractional uncertainties per visible energy bin for the full data sample. The largest uncertainty lies in the overall signal efficiency scale and statistics.

VI. TESTED MODELS

To compare with data, we generated predictions for the average (n, γ) multiplicity as a function of neutrino event

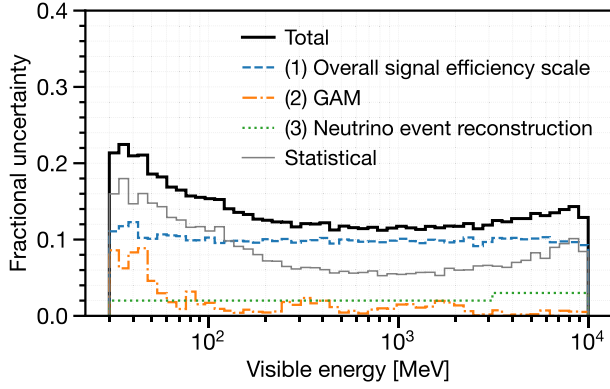


FIG. 12. Fractional uncertainties assigned to the estimated average (n, γ) multiplicity per bin for the full data sample.

visible energy using various neutrino event generators and hadron-nucleus interaction models. Six generator configurations were tested: NEUT 5.4.0, NEUT 5.6.3, GENIE hA, GENIE hN, GENIE BERT, and GENIE INCL. The GENIE setups use GENIE 3.4.0 with G18a10x0211b physics tunes [27], where $x \in \{a, b, c, d\}$ corresponds to the FSI model: INTRANUKE/hA [26] (hA), INTRANUKE/hN [26] (hN), the Geant4 Bertini cascade model [58] (BERT), and the Liège INC model [59] (INCL).

NEUT 5.4.0 follows the setup described in Sec. III C, while NEUT 5.6.3 includes a modified nuclear binding energy, removing roughly 10% of QE interactions in which the struck nucleon falls below the revised threshold. The GENIE setups share QE and single-pion production models with NEUT but differs in FSI and hadronization. Except for GENIE hA, all FSI models use the full INC approach. Also, the Liège INC model in GENIE is coupled with ABLA07 [60] for nuclear deexcitation.

Secondary hadron-nucleus interaction models were tested using SK detector simulations with GEANT 3.21 and GEANT 4.10.5.p01, in six configurations: SK-IV/V default, SK-VI default, G3 GCALOR, G4 BERT, G4 BERT_PC, and G4 INCL_PC. The SK-VI setup differs from SK-IV/V in its use of the Geant4 NeutronHP and Bertini cascade models for neutron tracking below 20 MeV and μ^- captures, respectively. G3 GCALOR relies entirely on GCALOR within GEANT 3.21, while the baseline setups use the NEUT pion FSI routine. Geant4-based models adopt ENDF/B-VII.1 for low-energy neutron transport, with G4 BERT and G4 BERT_PC using the Bertini cascade, and G4 INCL_PC using the Liège INC model. Configurations with the PC suffix employ the Geant4 precompound model [61] for nuclear deexcitation, while G4 BERT uses a simpler native model [58]. GEANT-3-based models rely on Bertini-based tabulation [39] for hadron-nucleus cross sections (excluding low-energy neutrons), whereas Geant4-based models use Glauber parametrization [62].

Figure 13 compares outgoing neutron momentum distributions predicted by the six neutrino event generator

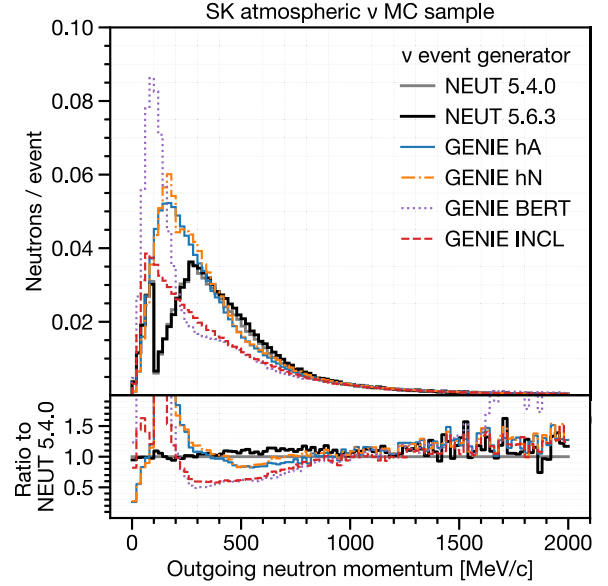


FIG. 13. Comparison of neutrino event generator options with different FSI models: Outgoing neutron momentum distribution per selected atmospheric neutrino event.

options, highlighting notable differences below 1 GeV/c, which are further discussed in Sec. VIII.

Figure 14 shows the average signal multiplicities as a function of projectile neutron momentum for the six secondary hadron-nucleus interaction models, with variations of up to 40%.

For each neutrino event generator, a full MC simulation equivalent to 50 years of atmospheric neutrino exposure

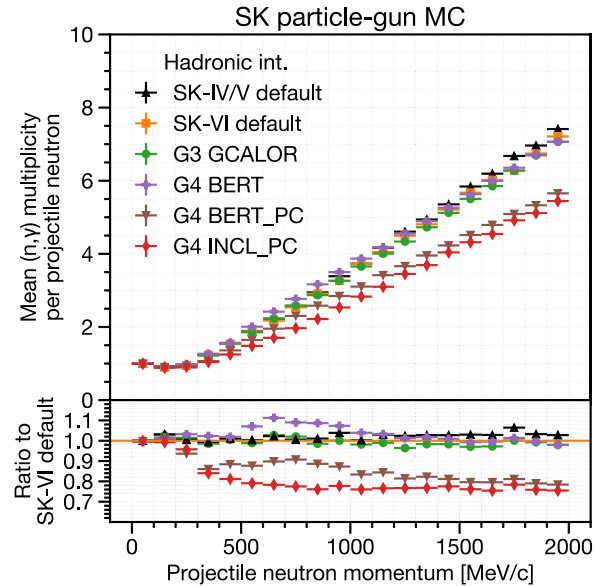


FIG. 14. Comparison of secondary hadron-nucleus interaction models: Average (n, γ) multiplicity per projectile neutron momentum bin, based on simulations using a single-neutron particle-gun setup in water.

was generated using the baseline detector model (SK-VI default). To account for variations in secondary hadron-nucleus interactions without additional full simulations, we precomputed average signal multiplicities for neutron-producing projectiles (n , p , π^\pm , and μ^-) in water up to 10 GeV/c, based on particle-gun MC simulations.

Predictions were made by convolving projectile momentum distributions with momentum-to-multiplicity tables. For each event generator, ratios relative to the baseline detector simulation model (SK-VI default) were used to scale the signal multiplicities obtained from the full MC simulation. In configurations using either NEUT 5.4.0 or NEUT 5.6.3 with SK-IV/V default, the discrepancy between this prediction method and the 500-year exposure full MC simulation was within 5% across all visible energy range and data subsamples.

Figures 15 and 16 show the model predictions. In visible energy range above a few hundred MeV, the average (n, γ) multiplicities are expected to increase linearly with visible energy. Differences across neutrino event generator configurations are more pronounced at visible energies below 1 GeV, while variability among secondary hadron-nucleus interaction models remains relatively constant. GENIE INCL predicts fewer neutrons than the other FSI models, as already suggested by Fig. 13. NEUT 5.4.0 predicts fewer neutrons than NEUT 5.6.3, due to a larger QE fraction. Additionally, the Geant4 Bertini cascade model with the precompound model (G4 BERT_PC) predicts fewer neutrons than when using the native deexcitation model (G4 BERT).

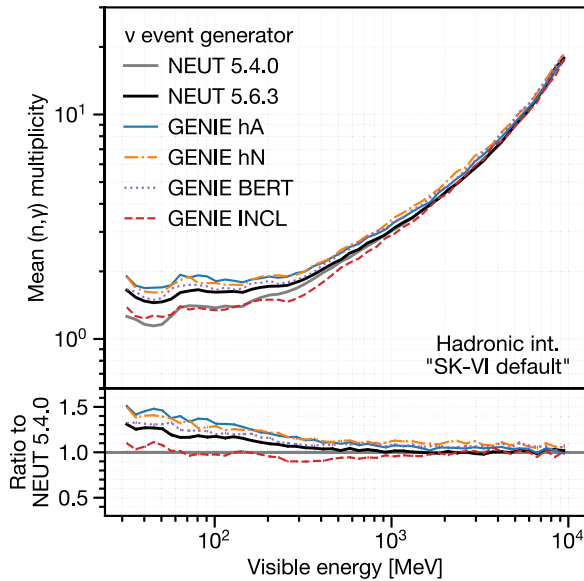


FIG. 15. Comparison of the predicted average (n, γ) multiplicity predicted by the neutrino event generator options paired with SK-VI default.

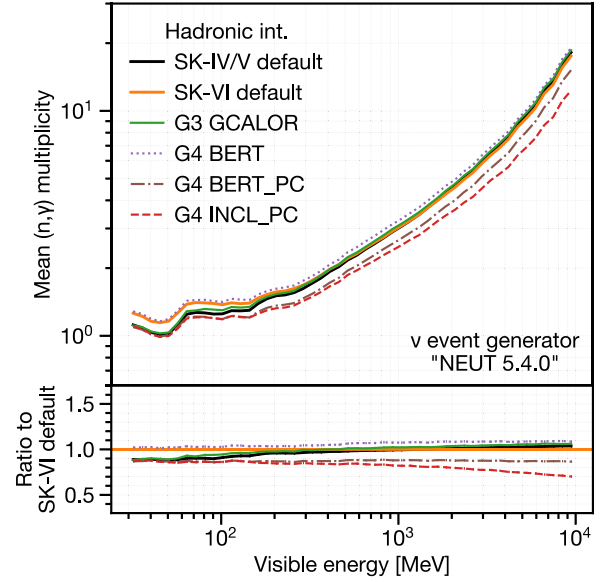


FIG. 16. Comparison of the predicted average (n, γ) multiplicity predicted by the secondary hadron-nucleus interaction models paired with NEUT 5.4.0.

VII. RESULTS

Table III summarizes the number of atmospheric neutrino events and detected neutron signals in the final data sample. The overall average (n, γ) multiplicity per neutrino event, $\langle N \rangle_{\text{overall}}$, is estimated using Eq. (4) for all selected atmospheric neutrino events without binning.

The combined data was compared with the baseline simulation and neutron production estimates in water from SNO [14], as shown in Fig. 17. For visible energies above 200 MeV—where outgoing lepton Cherenkov rings are well reconstructed—a linear relationship between visible energy and the average (n, γ) multiplicity was observed, as expected. While the data aligned well with the SNO estimate, it was 10%–30% lower than the baseline predictions in the sub-GeV energy range. A similar deficit was reported by SNO in heavy water [14], where their data was compared to GENIE 2.10.2 (hA FSI model) coupled with the Geant4 Bertini cascade model.

Predictions from various model combinations, as described in Sec. VI, were compared with the combined data across different subsamples, as shown in Fig. 18. The bottom panel of Fig. 18 contrasts the data with predictions from selected FSI models (GENIE hN and INCL) and secondary interaction models (G4 BERT, G4 BERT_PC, and G4 INCL_PC). The observed deficit in the [0.1, 0.4] GeV range was primarily found in the single-ring sample, which is expected to be dominated by CCQE interactions (see Figure 5). This deficit was only reproduced when using models that predict fewer neutrons than the baseline for both FSI and secondary interactions, such as GENIE INCL and G4 BERT_PC.

TABLE III. Summary of the number of atmospheric neutrino events (“ ν events”) and detected neutron signals (“ n signals”) in the final sample. $\langle N \rangle_{\text{overall}}$ is the unbinned application of Equation (4), followed by signal efficiency scale corrections with the factors given in Sec. IV B. These factors are specifically chosen to ensure consistency of $\langle N \rangle_{\text{overall}}$ across the three SK phases (see Appendix A). Errors for the observed $\langle N \rangle_{\text{overall}}$ are listed as statistical first, followed by systematic uncertainty. Other errors are statistical only. The bottom two rows show the expected $\langle N \rangle_{\text{overall}}$ and the true overall (n, γ) multiplicity extracted from the baseline full MC simulations (NEUT 5.4.0 with SK-IV/V default for SK-IV/V, and SK-VI default for SK-VI).

	SK-IV	SK-V	SK-VI
ν events	29,942	4,231	5,203
Events/day	9.23 ± 0.05	9.18 ± 0.14	9.22 ± 0.13
n signals	15,705	2,035	5,752
n signals/event	0.525 ± 0.004	0.481 ± 0.011	1.106 ± 0.015
Observed $\langle N \rangle_{\text{overall}}$	$2.49 \pm 0.03 \pm 0.23$	$2.49 \pm 0.10 \pm 0.11$	$2.49 \pm 0.06 \pm 0.05$
Expected $\langle N \rangle_{\text{overall}}$	2.83	2.84	2.85
True (n, γ) multiplicity	2.85	2.85	2.86

In the multiring sample at a few-GeV energies, there was significant variation in the predicted “slope” of the average (n, γ) multiplicity as a function of visible energy. In particular, G4 INCL_PC predicted a smaller slope compared to G4 BERT and G4 BERT_PC.

The goodness of fit between model predictions and data for the entire visible energy range of [0.03, 10] GeV was quantified using the following chi-square (χ^2) definition, considering only the two most dominant sources of uncertainty: statistical uncertainty and signal efficiency scale uncertainty.

$$\chi^2 = \sum_i^{N_{\text{bins}}} \frac{(sO_i - E_i)^2}{\sigma_{\text{stat},i}^2} \quad (5)$$

Here, i is the bin number, O_i and E_i represent the observed and expected average (n, γ) multiplicity, respectively, and $\sigma_{\text{stat},i}$ is the statistical uncertainty in the i -th bin. The total number of bins, N_{bins} , is 20 per subsample for both single-ring and multiring events, resulting in 40 bins in total. The binning was optimized to ensure a sufficient number of events per bin so that the average (n, γ) multiplicity follows a normal distribution. The normalization scale s was allowed to float within [0, 2]. Two χ^2 calculations were performed: (1) a normalization-free χ^2 , where s was minimized without constraint, and (2) a constrained χ^2 , where s was minimized with a penalty term $(s - 1)^2 / \sigma_{\text{scale}}^2$ and an additional constraint $|s - 1| < \sigma_{\text{scale}}$, with σ_{scale} set to 10%.

Table IV compares the reduced χ^2 values across different model combinations. Among the neutrino event generator options, GENIE INCL showed the lowest χ^2 for both free and constrained s . Among the secondary interaction models, G4 BERT_PC provided the best agreement with data under the 10% normalization constraint. The combination of these two models yielded the best overall agreement under the same constraint. Other secondary interaction models such as SK-IV/V default showed reasonable agreement when coupled with GENIE INCL or when normalization was allowed to float. In contrast, G4 INCL_PC showed higher unconstrained χ^2 values, due to the shallower slope prediction for the multiring events.

Model performance, separated into contributions from target nucleus FSI and secondary interactions, is further visualized using effective metrics in Appendix B.

VIII. DISCUSSION

Overall, our data favor models that predict relatively lower neutron production for single-ring sub-GeV events

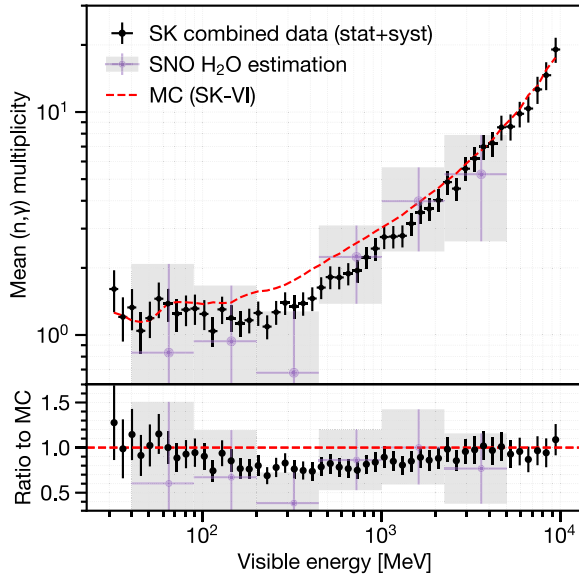


FIG. 17. Average (n, γ) multiplicity in atmospheric neutrino events as a function of their visible energy. Black data points represent combined SK data with statistical and systematic uncertainties. Purple crosses with shades are estimates for pure water based on SNO measurements [14] using a D_2O target. The red dashed line represents the true average (n, γ) multiplicity from the baseline SK-IV MC simulation.

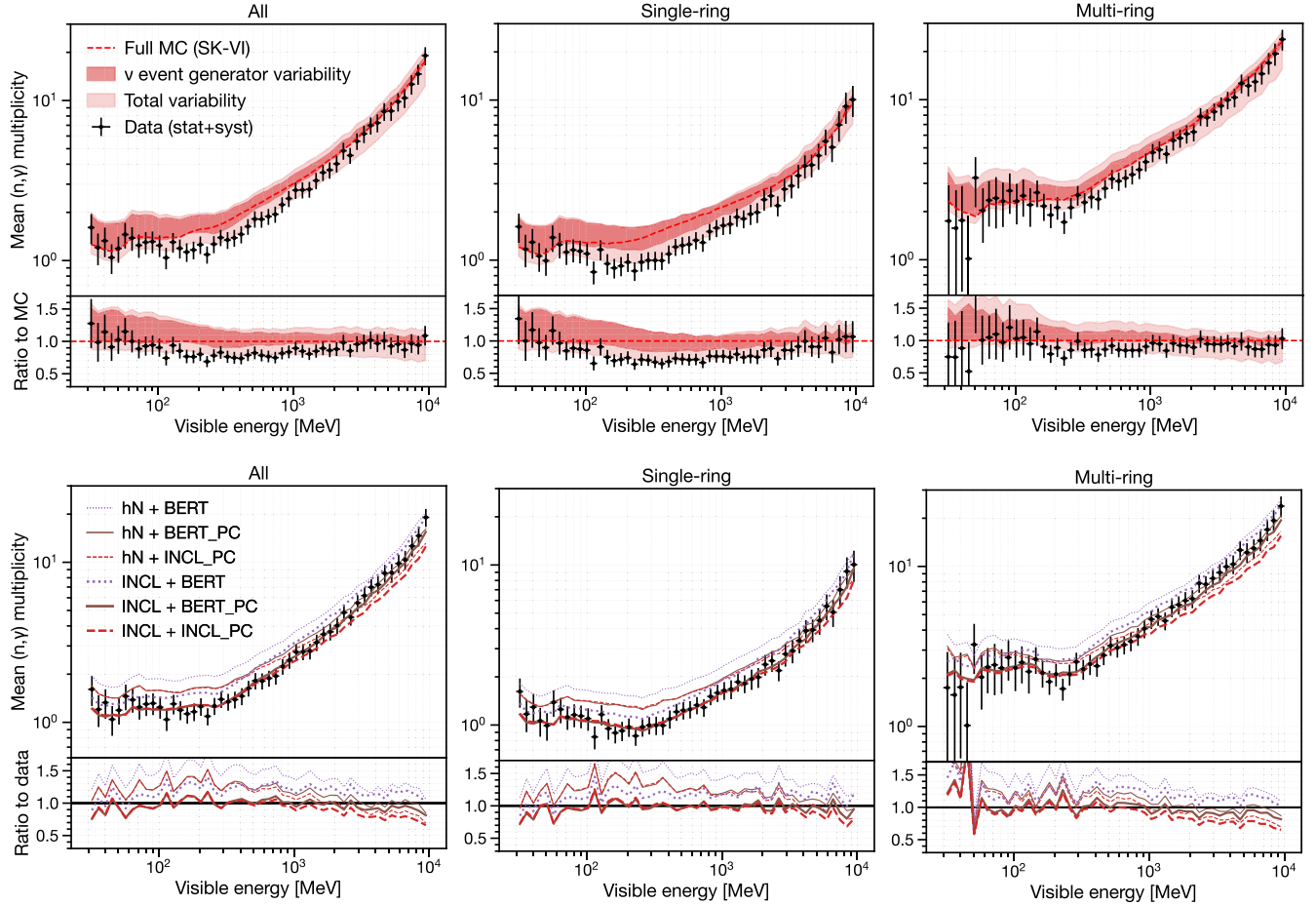


FIG. 18. Comparison of various model predictions with the combined data, showing the average (n, γ) multiplicity as a function of atmospheric neutrino event visible energy, for all events (left), single-ring events (middle), and multiring events (right). The black data points represent the combined data, including both statistical and systematic uncertainties. In the top panel, the combined data are compared to the prediction intervals of the models. Thicker shaded regions represent the range of predictions obtained by varying only the neutrino event generator options while keeping the secondary hadron-nucleus interaction model fixed to the baseline (SK-VI default). Lighter shaded regions indicate the broader range of predictions from all model combinations. The bottom panel shows corresponding plots for specific model choices that are based on GENIE and Geant4, including the one (GENIE INCL + G4 BERT_PC) that shows good agreement with the data (see Table IV).

and higher neutron production for multiring multi-GeV events. In this section, we examine key features of these models that influence neutron production predictions, focusing on Fig. 13, which illustrates differences in neutron momentum distributions across the tested FSI models. By qualitatively comparing the features of these models, we discuss potential factors contributing to the observed discrepancies in predictions.

A. Nuclear deexcitation

Nuclear deexcitation models are crucial for predicting total neutron production, as they govern low-energy neutron emission per nuclear interaction. This is evident when comparing G4 BERT and G4 BERT_PC, which differ only in their deexcitation treatment. For instance, the inaccuracy of the Geant4 Bertini cascade model prediction is indicated

by the distinct large neutron peak below 250 MeV/c in GENIE BERT (Fig. 13), compared to the other models such as ABLA07 in GENIE INCL or the native deexcitation model in NEUT. (The GENIE hA and hN options omit nuclear deexcitation entirely.) Coupling the Geant4 precompound model with the Geant4 Bertini cascade model significantly reduces neutron production across all visible energy ranges, leading to much better agreement with data (Fig. 18).

In the Geant4 Bertini cascade model, neutron emission follows the Weisskopf statistical evaporation model [63]. A key input determining neutron emission probability is the inverse reaction (neutron absorption) cross section. The inaccuracy of the Bertini model and its variants in predicting neutron evaporation may stem from its use of the simplified cross section parametrization by Dostrovsky [64]. In comparison, the Geant4 precompound model uses

TABLE IV. Reduced χ^2 [see Eq. (5)] and p -values for each model combination. The signal efficiency scale s was either freely minimized or constrained to 10%. The fit has 38 fixed degrees of freedom. p -values assume Gaussian bin errors without bin correlations.

	ν generator	Secondary interaction	Minimizing s		Reduced χ^2		p -value	
			Free s	Constrained s	Free s	Constrained s	Free s	Constrained s
1	NEUT 5.4.0	SK-IV/V default	0.75	0.90	2.76	11.66	$<10^{-4}$	$<10^{-4}$
2	NEUT 5.4.0	SK-VI default	0.74	0.90	3.64	14.38	$<10^{-4}$	$<10^{-4}$
3	NEUT 5.4.0	G3 GCALOR	0.73	0.90	2.14	13.47	$<10^{-4}$	$<10^{-4}$
4	NEUT 5.4.0	G4 BERT	0.70	0.90	2.14	20.76	$<10^{-4}$	$<10^{-4}$
5	NEUT 5.4.0	G4 BERT_PC	0.84	0.90	2.73	3.68	$<10^{-4}$	$<10^{-4}$
6	NEUT 5.4.0	G4 INCL_PC	0.89	0.90	5.08	5.11	$<10^{-4}$	$<10^{-4}$
7	NEUT 5.6.3	SK-IV/V default	0.72	0.90	2.92	15.94	$<10^{-4}$	$<10^{-4}$
8	NEUT 5.6.3	SK-VI default	0.71	0.90	4.40	19.97	$<10^{-4}$	$<10^{-4}$
9	NEUT 5.6.3	G3 GCALOR	0.71	0.90	2.24	18.30	$<10^{-4}$	$<10^{-4}$
10	NEUT 5.6.3	G4 BERT	0.67	0.90	2.53	27.88	$<10^{-4}$	$<10^{-4}$
11	NEUT 5.6.3	G4 BERT_PC	0.81	0.90	3.38	5.87	$<10^{-4}$	$<10^{-4}$
12	NEUT 5.6.3	G4 INCL_PC	0.86	0.90	6.38	6.86	$<10^{-4}$	$<10^{-4}$
13	GENIE hA	SK-IV/V default	0.73	0.90	3.17	14.83	$<10^{-4}$	$<10^{-4}$
14	GENIE hA	SK-VI default	0.67	0.90	6.71	30.43	$<10^{-4}$	$<10^{-4}$
15	GENIE hA	G3 GCALOR	0.71	0.90	2.47	17.69	$<10^{-4}$	$<10^{-4}$
16	GENIE hA	G4 BERT	0.65	0.90	3.73	36.46	$<10^{-4}$	$<10^{-4}$
17	GENIE hA	G4 BERT_PC	0.77	0.90	5.29	10.91	$<10^{-4}$	$<10^{-4}$
18	GENIE hA	G4 INCL_PC	0.80	0.90	9.64	12.93	$<10^{-4}$	$<10^{-4}$
19	GENIE hN	SK-IV/V default	0.71	0.90	1.75	16.50	0.0030	$<10^{-4}$
20	GENIE hN	SK-VI default	0.66	0.90	4.32	31.29	$<10^{-4}$	$<10^{-4}$
21	GENIE hN	G3 GCALOR	0.69	0.90	1.48	20.77	0.028	$<10^{-4}$
22	GENIE hN	G4 BERT	0.63	0.90	2.29	40.19	$<10^{-4}$	$<10^{-4}$
23	GENIE hN	G4 BERT_PC	0.76	0.90	3.13	10.10	$<10^{-4}$	$<10^{-4}$
24	GENIE hN	G4 INCL_PC	0.79	0.90	6.48	10.34	$<10^{-4}$	$<10^{-4}$
25	GENIE BERT	SK-IV/V default	0.75	0.90	1.22	9.79	0.17	$<10^{-4}$
26	GENIE BERT	SK-VI default	0.70	0.90	3.31	21.35	$<10^{-4}$	$<10^{-4}$
27	GENIE BERT	G3 GCALOR	0.73	0.90	1.21	13.69	0.17	$<10^{-4}$
28	GENIE BERT	G4 BERT	0.67	0.90	1.79	27.70	0.0020	$<10^{-4}$
29	GENIE BERT	G4 BERT_PC	0.80	0.90	2.56	6.06	$<10^{-4}$	$<10^{-4}$
30	GENIE BERT	G4 INCL_PC	0.83	0.90	5.60	7.12	$<10^{-4}$	$<10^{-4}$
31	GENIE INCL	SK-IV/V default	0.87	0.90	0.84	1.21	0.74	0.18
32	GENIE INCL	SK-VI default	0.80	0.90	1.33	4.59	0.082	$<10^{-4}$
33	GENIE INCL	G3 GCALOR	0.83	0.90	1.36	2.75	0.067	$<10^{-4}$
34	GENIE INCL	G4 BERT	0.76	0.90	0.87	8.55	0.71	$<10^{-4}$
35	GENIE INCL	G4 BERT_PC	0.92	0.92	0.95	0.96	0.56	0.53
36	GENIE INCL	G4 INCL_PC	0.97	0.97	2.74	2.74	$<10^{-4}$	$<10^{-4}$

a more refined parametrization based on the nuclear optical model potential [65], fitted to a broader dataset [66], resulting in lower evaporation rates particularly for low-energy neutrons. Additionally, the Geant4 precompound model features tuned level density parameters and allows Fermi breakup in nuclei up to mass 16 [61], including oxygen, further reducing isolated neutrons that contribute to (n, γ) signals.

B. Nuclear in-medium effects

A key distinction among the FSI cascade models is their treatment of Pauli blocking, which prevents nucleons from occupying the same quantum state, as dictated by the Pauli

exclusion principle. For instance, NEUT 5.4.0, NEUT 5.6.3 and GENIE BERT implement “strict” Pauli blocking, which forbids all nucleon scattering below the pre-defined oxygen Fermi momentum around 225 MeV/c, modeling the nucleus as a degenerate Fermi gas. This approach results in a characteristic dip in the outgoing neutron momentum distributions for NEUT 5.4.0 and NEUT 5.6.3 within the corresponding energy range, as shown in Fig. 13. The Liège INC model applies a probabilistic approach for nucleon collisions throughout the cascade (except for the initial collision), accounting for nucleon holes in the surrounding phase space volume [67,68]. This results in smoother neutron distributions near

the Fermi momentum, as seen in Fig. 13. In contrast, the GENIE hA and hN models omit Pauli blocking entirely, leading to significantly higher predicted neutron production than NEUT 5.4.0, NEUT 5.6.3, and GENIE INCL.

In the 0.2–1 GeV/c range, the number of neutrons predicted by the Liège INC and Geant4 Bertini cascade models is lower than that from the GENIE hA, hN, and NEUT nucleon FSI models, as illustrated in Fig. 13. One contributing factor may be the inclusion of cluster formation—such as deuterons, tritons, and α particles—among cascade products, which reduces the number of isolated neutrons contributing to the observable signal. Reference [69] also identifies cluster formation as a major mechanism behind the reduction of outgoing protons in the Liège INC model. Additional in-medium effects, considered either explicitly or implicitly in the two models—such as short-range nucleon repulsion and local nuclear density reduction after each collision—further reduce the number of nucleon-nucleon collisions during the cascade.

C. π^\pm production

Accurate modeling of pion-nucleus interactions is crucial for predicting neutron production in multiring events with energetic pions (see Figs. 24 and 25 in Appendix C). The difference in predicted neutron production in the few-GeV multiring sample between the Liège INC model (G4 INCL_PC) and the Geant4 Bertini cascade model (G4 BERT_PC), as observed in Fig. 18, can be attributed to differences in pion production cross sections for pion projectiles. Our results are consistent with those of Ref. [70], which report that the Liège INC model tends to predict lower inclusive π^\pm production cross sections on light nuclei (mass number $A < 20$) compared to the Geant4 Bertini cascade model. In addition, the pion potential within the nucleus differs between the two models. In the Liège INC model, the pion potential is generally deeper [71] than the constant 7 MeV potential used in the Geant4

Bertini cascade model [58], which impacts pion transparency. For instance, Ref. [72] demonstrated that the Liège INC model shows lower pion transparency for pions with momenta below 300 MeV/c compared to the GENIE hA, hN, or NEUT pion FSI models. Also, using the NEUT pion FSI routine for low-energy π^\pm transport in water slightly improves agreement with the data (e.g., SK-IV/V default vs G3 GCALOR).

D. Other considerations

Neutrons produced from μ^- captures contribute 5%–10% of the average (n, γ) multiplicity for visible energies below 300 MeV, as shown in Fig. 16 (e.g., SK-VI default, which includes μ^- capture simulation, vs SK-IV/V default, which does not).

Our model predictions do not account for interactions of ν_τ and $\bar{\nu}_\tau$ from oscillations, nor neutrinos with energies below 100 MeV. However, their contributions are expected to be smaller than our uncertainty budget in the upper and lower ends of the visible energy range, where their effects become relevant, respectively.

Tables V and VI summarize the components of each tested model and their agreement with the data. The differences in model components indicate that accurately describing the data requires both moderate evaporation of low-energy neutrons (as modeled in NEUT, ABLA07, and Geant4 precompound) and suppression of medium-energy neutrons in the [0.3, 1] GeV/c range via in-medium effects such as Pauli blocking, cascade step correlations, and cluster formation. Among the tested neutrino event generator configurations, only GENIE INCL satisfies both conditions as seen in Fig. 13. For secondary hadron-nucleus interaction models, G4 BERT_PC and G4 INCL_PC meet these requirements, with G4 BERT_PC performing better at higher energies due to its relatively larger π^\pm production. These features make GENIE INCL and G4 BERT_PC the most successful model combinations

TABLE V. Summary of interaction models for final-state hadrons from neutrino-nucleus interactions used in the neutrino event generators, including their component choices and the resulting minimum reduced χ^2 values when compared to data (with a 10% constraint on the signal efficiency scale). Footnotes highlight notable features of specific models or components.

Model	Hadron transport	Nuclear deexcitation	Minimum χ^2 (coupled model)
NEUT 5.4.0	GCALOR Bertini ^a	Refs. [35,36]	3.68 (G4 BERT_PC)
NEUT 5.6.3 ^b	GCALOR Bertini ^a	Refs. [35,36]	5.87 (G4 BERT_PC)
GENIE hA	INTRANUKE/hA ^c	...	10.91 (G4 BERT_PC)
GENIE hN	INTRANUKE/hN ^c	...	10.10 (G4 BERT_PC)
GENIE BERT	Geant4 Bertini ^d	Geant4 Bertini ^c	6.06 (G4 BERT_PC)
GENIE INCL	Geant4 INCL++ ^{ef}	ABLA07	0.96 (G4 BERT_PC)

^a π^\pm inelastic scattering below 500 MeV/c are modeled based on Refs. [31,32].

^bPredicts a lower fraction of QE interactions compared to NEUT 5.4.0, due to a modified treatment of nuclear binding energy.

^cPauli blocking is not included.

^dIncludes cluster formation and nucleon correlations that effectively reduce collisions, in contrast to GCALOR Bertini.

^ePredicts larger neutron evaporation compared to the NEUT model or ABLA07.

^fIncludes a probabilistic treatment of Pauli blocking and predicts smaller π -induced π^\pm production than Geant4 Bertini.

TABLE VI. Summary of secondary hadron–nucleus interaction models used in the detector simulator, including their component choices and the resulting minimum reduced χ^2 values when compared to data (with a 10% constraint on the signal efficiency scale). Footnotes highlight notable features of specific models or components.

Model	Hadron transport	Nuclear deexcitation	μ^- capture	Minimum χ^2 (coupled model)
SK-IV/V default	GCALOR Bertini ^a	GCALOR Bertini	...	1.21 (GENIE INCL)
SK-VI default	GCALOR Bertini ^a	GCALOR Bertini	Geant4 Bertini	4.59 (GENIE INCL)
G3 GCALOR	GCALOR Bertini	GCALOR Bertini	...	2.75 (GENIE INCL)
G4 BERT	Geant4 Bertini ^b	Geant4 Bertini	Geant4 Bertini	8.55 (GENIE INCL)
G4 BERT_PC	Geant4 Bertini ^b	Geant4 precompound ^c	Geant4 Bertini	0.96 (GENIE INCL)
G4 INCL_PC	Geant4 INCL++ ^{bd}	Geant4 Precompound ^c	Geant4 Bertini	2.74 (GENIE INCL)

^a π^\pm inelastic scattering below 500 MeV/c are modeled based on Refs. [31,32].

^bIncludes cluster formation and nucleon correlations that effectively reduce collisions, in contrast to GCALOR Bertini.

^cPredicts smaller neutron evaporation compared to Geant4 Bertini.

^dIncludes a probabilistic treatment of Pauli blocking and predicts smaller π -induced π^\pm production than Geant4 Bertini.

for reproducing the data. Additionally, the native deexcitation routine in the Geant4 Bertini cascade model significantly overpredicts neutron evaporation; thus, caution is advised when using it for neutron multiplicity predictions.

IX. CONCLUSIONS AND PROSPECTS

Accurately modeling neutron production in neutrino interactions is essential for characterizing incoming neutrinos, which is crucial for advancing precision measurements of neutrino oscillation parameters and rare event searches involving neutron tagging. Recent studies have suggested potential inaccuracies in the modeling of secondary hadron-nucleus interactions.

This paper reports a measurement of total neutron production following atmospheric neutrino interactions within the water volume of the Super-Kamiokande (SK) detector. Atmospheric neutrino events were binned by their electron-equivalent “visible energy,” a semicalorimetric proxy for neutrino momentum transfer, in the range of [0.03, 10] GeV. The dominant systematic uncertainty was the roughly 10% uncertainty in the overall signal efficiency scale. The average neutron capture multiplicity in each visible energy bin was compared against predictions from various combinations of neutrino event generators and secondary hadron-nucleus interaction models.

Our data provides strong discriminative power for evaluating these models, with predictions varying by up to 50%. Two key observations were made: a reduction in neutron production for sub-GeV single-ring events, where secondary hadron interactions are minimal, and a nearly linear increase in neutron production with visible energy, especially in the multi-GeV region. The first observation highlights the need for moderate neutron evaporation and consideration of nuclear in-medium effects. The second is more sensitive to pion production in cascade models, which become important at higher energies.

This study highlights the crucial role of hadron-nucleus interaction models in accurately predicting total neutron production from neutrino interactions. The observed

discrepancy with the evaporation model [64] commonly used in Bertini-based cascade models aligns with previous studies, including neutron measurements with NCQE selections [17] at SK and secondary γ -ray measurements using Germanium detectors with a neutron beam on a water target [73–75]. Our data also supports the selection of specific models, such as GENIE INCL and G4 BERT_PC, which reduce the uncertainty in total neutron production from atmospheric neutrino events to around 10%. These models may help explain the deficits in total neutron capture signals reported by T2K [13] and SNO [14], and offer insights into the low-energy neutron signal deficits observed in MINERvA [2]. Lastly, this work will improve the use of neutron tagging for identifying atmospheric antineutrino events at SK, enhancing sensitivity to neutrino mass ordering and searches for proton decays and the diffuse supernova neutrino background.

The data and model predictions to produce Fig. 18 and Table IV are openly available [76].

ACKNOWLEDGMENTS

We gratefully acknowledge the cooperation of the Kamioka Mining and Smelting Company. The Super-Kamiokande experiment has been built and operated from funding by the Japanese Ministry of Education, Culture, Sports, Science and Technology; the U.S. Department of Energy; and the U.S. National Science Foundation. Some of us have been supported by funds from the National Research Foundation of Korea (Grants No. NRF-2009-0083526, No. NRF-2022R1A5A1030700, and No. NRF-2202R1A3B1078756) funded by the Ministry of Science, Information and Communication Technology (ICT); the Institute for Basic Science (IBS-R016-Y2); and the Ministry of Education (Grants No. 2018R1D1A1B07049158, No. 2021R1I1A1A01042256, No. 2021R1I1A1A01059559, and No. RS-2024-00442775); the Japan Society for the Promotion of Science; the National Natural Science Foundation of China under Grant No. 12375100; the Spanish Ministry of Science, Universities and Innovation

(Grant No. PID2021-124050NB-C31); the Natural Sciences and Engineering Research Council (NSERC) of Canada; the Scinet and Westgrid consortia of Compute Canada; the National Science Centre (Grants No. UMO-2018/30/E/ST2/00441 and No. UMO-2022/46/E/ST2/00336) and the Ministry of Science and Higher Education (2023/WK/04), Poland; the Science and Technology Facilities Council (STFC) and Grid for Particle Physics (GridPP), UK; the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 754496; H2020-MSCA-RISE-2018 JENNIFER2 Grant Agreement No. 822070, H2020-MSCA-RISE-2019 SK2HK Grant Agreement No. 872549; and European Union’s Next Generation EU/PRTR Grant No. CA3/RSUE2021-00559; the National Institute for Nuclear Physics (INFN), Italy.

DATA AVAILABILITY

The data that support the findings of this article are openly available [76].

APPENDIX A: $\langle N \rangle_{\text{overall}}$ CONSISTENCY ACROSS SK PHASES

When applying a signal efficiency correction based on the weighted mean of Am/Be data-to-MC efficiency ratios—as shown in Fig. 10 and described in Sec. IV B—the efficiency-corrected estimate of total neutron production $\langle N \rangle_{\text{overall}}$ was $2.21 \pm 0.03(\text{stat}) \pm 0.11(\text{syst})$ in SK-IV, lower than the values observed in the other two phases: $2.46 \pm 0.10 \pm 0.11$ for SK-V and $2.50 \pm 0.06 \pm 0.05$ for SK-VI. This discrepancy may be due to unaccounted systematic uncertainties in SK-IV, which had the longest

data-taking period (nearly 10 years) but only a limited amount of calibration data to constrain the entire duration. In contrast, SK-VI—with a much shorter runtime (about 2 years)—periodically took calibration data across various positions within the detector volume. Thus, the signal efficiency scale for SK-VI is considered more reliable than that of SK-IV.

To test the robustness of the signal efficiency correction derived from the Am/Be calibration data and simulation, we applied an alternative neutron signal detection algorithm to both the calibration and atmospheric neutrino datasets in SK-VI. This algorithm uses a likelihood-based vertex fitter [77] to reconstruct the $\text{Gd}(n, \gamma)$ vertex independently of the neutrino interaction vertex, making it more resilient to uncertainties in outgoing neutron kinematics that affect the detection efficiency in the baseline algorithm described in Sec. IV A. Using this reference algorithm, we obtained $\langle N \rangle_{\text{overall}} = 2.49 \pm 0.06 \pm 0.05$ for SK-VI, consistent with the baseline result.

Relying on the signal efficiency scale obtained for SK-VI, we adjust the correction factors for all phases such that $\langle N \rangle_{\text{overall}}$ aligns with the SK-VI reference value of 2.49. The difference between the calibration-based and phase-consistency-based correction factors is treated as an independent systematic uncertainty in the signal efficiency scale and is included in the uncertainty budget shown in Fig. 12. This yields a final $\langle N \rangle_{\text{overall}}$ of $2.49 \pm 0.03(\text{stat}) \pm 0.26(\text{syst})$, corresponding to a total systematic uncertainty of approximately 10%. Figure 19 illustrates the impact of the two signal efficiency correction approaches: one based on Am/Be neutron source calibration and the other based on phase consistency.

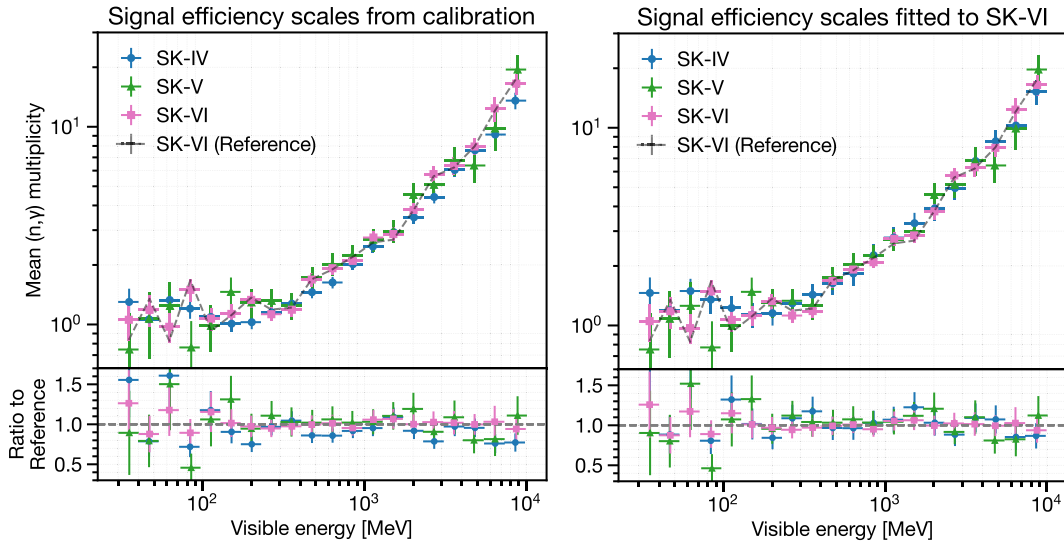


FIG. 19. Average (n, γ) multiplicities plotted against neutrino event visible energy across SK phases. The left plot uses signal efficiency scales from Am/Be calibration (Sec. IV B), while the right plot adjusts SK-IV and SK-V scales to match $\langle N \rangle_{\text{total}}$ of SK-VI. Error bars include statistical and systematic uncertainties, with the right plot also accounting for scale differences. The dashed “SK-VI (Reference)” line shows SK-VI data using a neutron-energy-independent reference algorithm.

APPENDIX B: EFFECTIVE METRICS FOR MODEL EVALUATION

To separately evaluate the contributions from FSI within the target nucleus and downstream secondary interactions, we define two effective metrics. The first metric is the low-energy (low-E) multiplicity as the average (n, γ) multiplicity in the $[0.1, 0.3]$ GeV visible energy range, where the contribution of low-energy nucleon FSI is significant (see Fig. 23 in Appendix C). Events below 0.1 GeV are excluded due to large systematic uncertainties. The second metric is the slope of the linear increase in average (n, γ) multiplicity as a function of visible energy, obtained by performing a linear fit over the range $[0.3, 10]$ GeV, as shown in Fig. 20.

Figure 21 compares the measured slopes and low-energy multiplicities across different SK phases and event topologies. For the slope fits shown in the figure, the binning

scheme was adjusted to ensure sufficient statistics (typically $\gtrsim 30$ neutrino events per bin). The measured values are consistent across SK phases within the assigned uncertainties. Both metrics are higher for multiring events than for single-ring events, as expected due to the higher fraction of deep inelastic scattering (DIS) interactions in the multiring sample (see Fig. 5). While the observed slopes agree with baseline model predictions, the measured low-energy multiplicities were lower than expected.

Figure 22 compares the measured slopes and low-energy multiplicities with model predictions. The predictions show three distinct groups of slopes: the G4 INCL_PC model gives the smallest slopes, G4 BERT_PC predicts intermediate slopes, and other variants of the Bertini cascade model produce the largest slopes. The slope observed in the single-ring data better matches the G4 INCL_PC predictions, while the slope in the multiring data is closer to the G4 BERT_PC predictions. This distinction is most evident

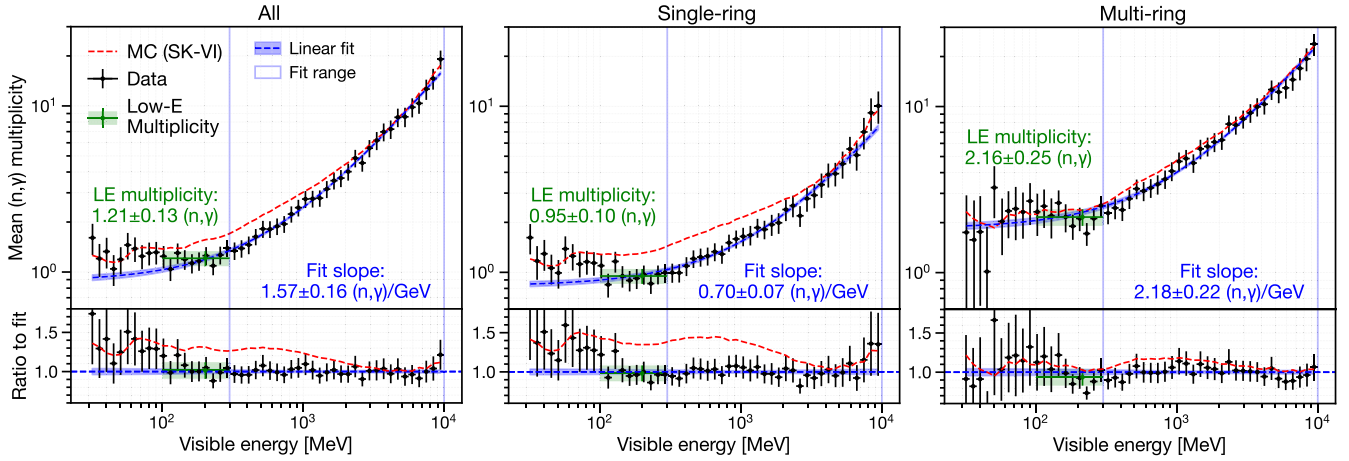


FIG. 20. Average (n, γ) multiplicity as a function of visible energy for neutrino events: all events (left), single-ring events (center), and multiring events (right). Red dashed lines show predictions from the full MC simulation in the SK-VI baseline setup. Blue dashed lines represent fitted linear functions with 1σ prediction intervals shaded in blue, with the fit covering the energy range $[0.3, 10]$ GeV. Green points represent low-energy multiplicity in the $[0.1, 0.3]$ GeV range. Error bars include both statistical and systematic uncertainties.

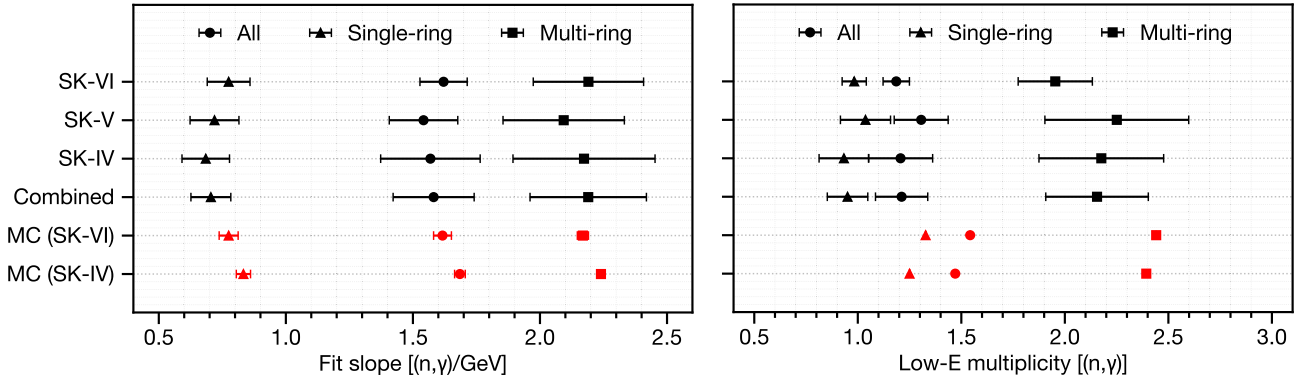


FIG. 21. Measured slopes (left) and low-energy multiplicities (right) across different SK phases and event types. Data error bars (black) represent both statistical and systematic uncertainties, while MC error bars in the left panel indicate statistical uncertainties. In the right panel, MC predictions correspond to the “true” average (n, γ) multiplicity.

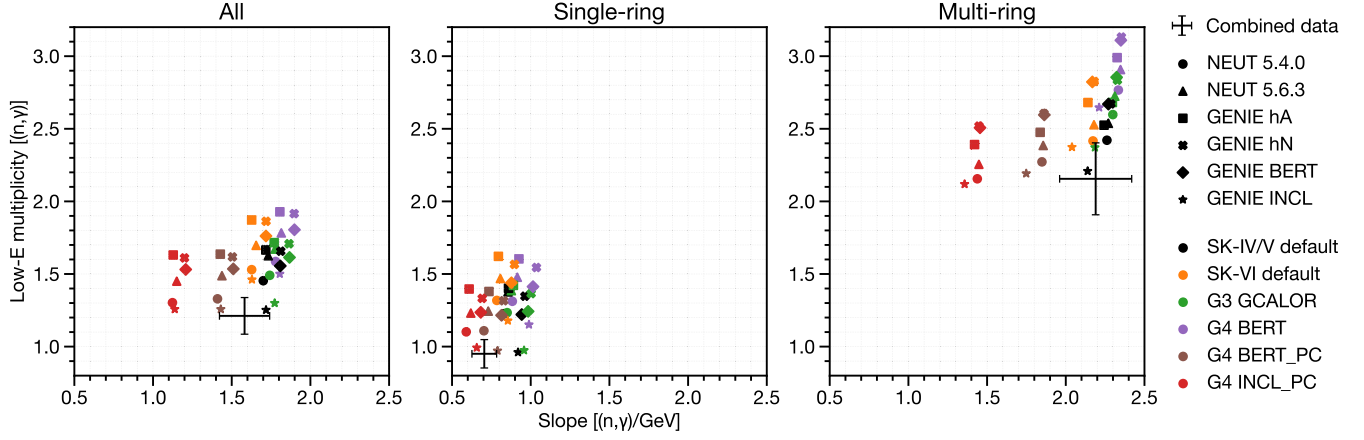


FIG. 22. Scatter plots of fitted slopes and low-energy multiplicities in model predictions and data (black crosses, error bars include both statistical and systematic uncertainties), shown for all events (left), single-ring events (middle), and multiring events (right). Shapes represent neutrino event generator options used, while colors indicate the secondary hadron-nucleus interaction model options.

in the bottom panel of Fig. 18, where G4 INCL_PC matches well with sub-GeV single-ring data, and G4 BERT_PC is a better fit for multi-GeV multiring data. For low-energy multiplicities, NEUT 5.4.0, NEUT 5.6.3, and GENIE INCL are preferred, as they predict lower neutron production. In contrast, models such as GENIE hA, hN, and BERT overestimate low-energy multiplicities, producing values significantly higher than the estimated 1σ uncertainty. These trends in model predictions and agreement with data remained robust under variations in the visible energy ranges used to define both metrics.

APPENDIX C: SOURCES OF NEUTRON PRODUCTION

Here, we clarify the sources of neutron production in our simulation. Figure 23 shows the baseline SK-IV neutron production as a function of visible energy. The number of outgoing neutrons from neutrino-nucleus interactions remains nearly constant across visible energies. At lower energies, the outgoing neutrons have small momenta, with each neutron corresponding closely to a single (n, γ) reaction. At higher energies, larger hadron momentum results in a linear increase in secondary neutron production. This indicates that, for the two metrics defined in Appendix B, the low-energy multiplicity is sensitive to nucleon transport modeling, while the slope is primarily determined by the secondary hadron-nucleus interaction model.

Figure 24 shows the contributions of nucleons and pions to the (n, γ) multiplicity as a function of visible energy, for different secondary hadron-nucleus interaction models. A general trend across all models is that, in the single-ring sample, most (n, γ) reactions are induced by outgoing nucleons from neutrino interactions, while in the multiring

sample, outgoing pions dominate in the few-GeV range. Figure 25 further breaks down the contributions to the (n, γ) multiplicity by outgoing particle momentum, for two FSI models (GENIE INCL and GENIE hN) and two secondary interaction models (G4 INCL_PC and G4 BERT_PC). In single-ring atmospheric neutrino events with visible energy in the range $[0.1, 0.3]$ GeV, the total (n, γ) multiplicity is primarily driven by neutrons below 1 GeV/c (from neutrino-nucleus interactions), with negligible variation across secondary interaction models. However, in multiring events with visible energies above 2 GeV, the total contribution from outgoing pions at all momenta becomes comparable to that from neutrons.

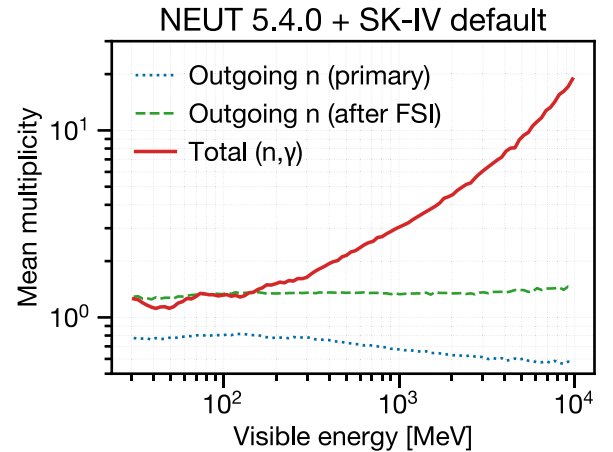


FIG. 23. Mean multiplicity of outgoing neutrons from the primary neutrino interaction (blue, dotted) and subsequent FSIs (green, dashed) and resulting total (n, γ) reactions (red, solid), in the baseline SK-IV simulation setup.

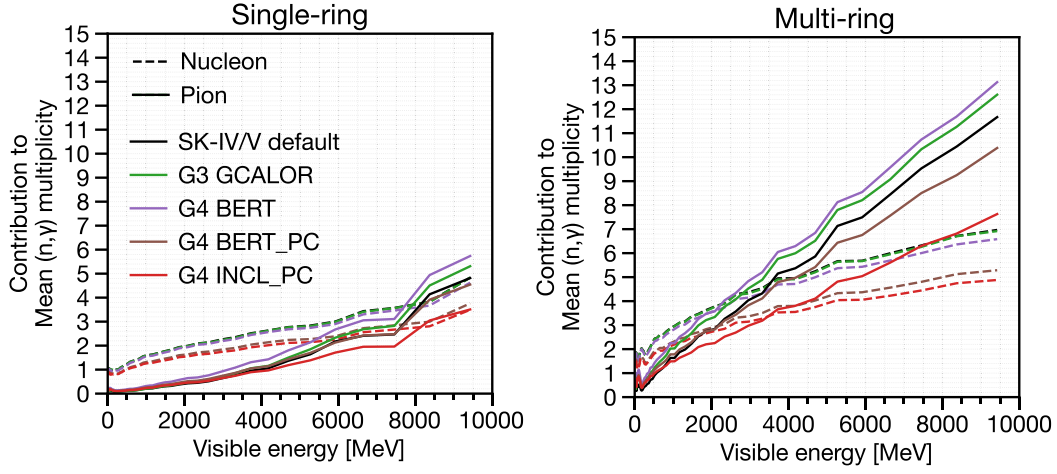


FIG. 24. Average contribution of outgoing nucleons (dashed lines) and pions (solid lines) to the observed (n, γ) multiplicities as a function of visible energy, shown separately for single-ring events (left) and multiring events (right). The line colors represent different secondary hadron-nucleus interaction models used in the predictions.

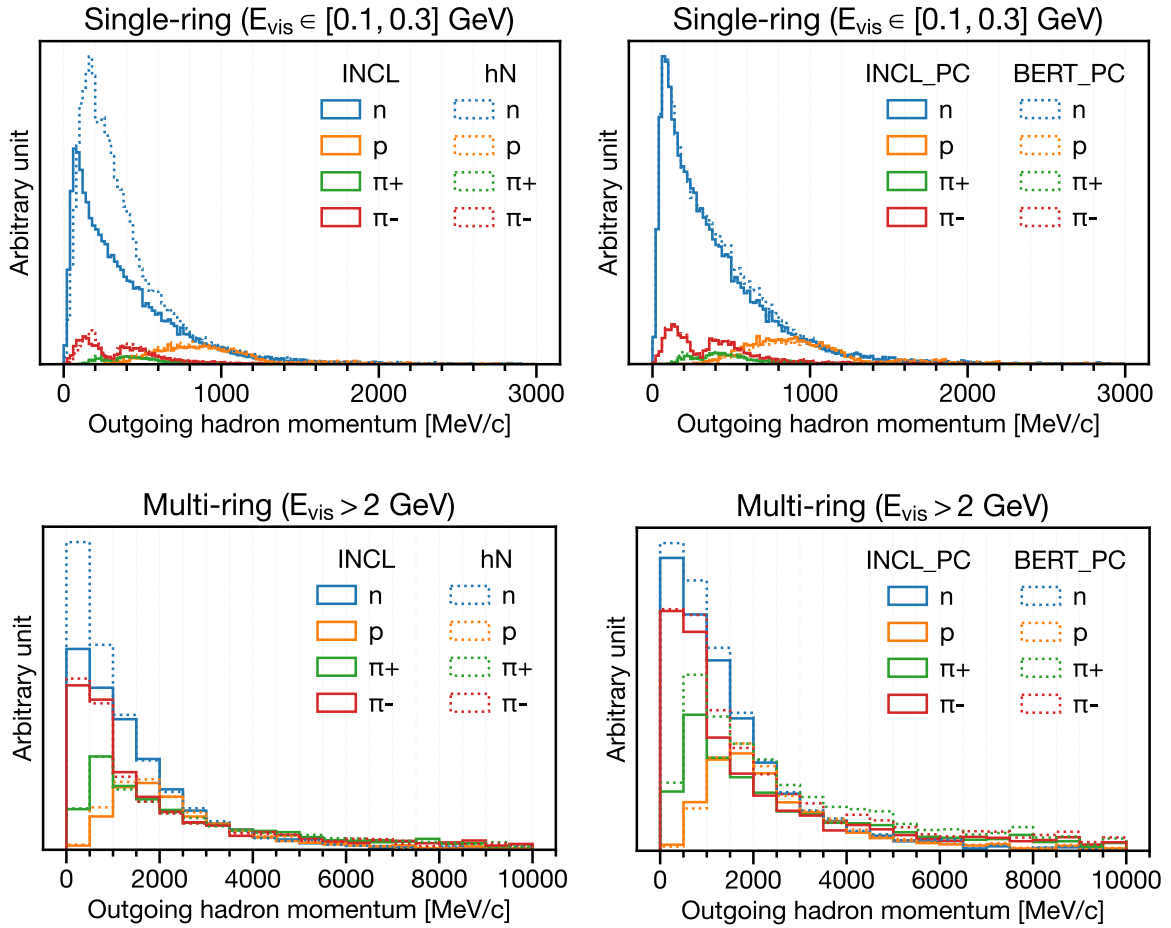


FIG. 25. Momentum distributions of outgoing hadrons predicted by neutrino event generators, weighted by the average (n, γ) multiplicity per projectile hadron momentum bin as predicted by secondary hadron-nucleus interaction models. The area under each histogram bin represents the average contribution of a specific hadron at a given momentum to the observable (n, γ) multiplicity. The top panels show predictions for single-ring events with visible energy (E_{vis}) in the range $[0.1, 0.3]$ GeV, while the bottom panels correspond to multiring events with visible energy above 2 GeV. The left plots compare two different cascade models for final-state interactions (GENIE INCL and GENIE hN) coupled with the secondary hadron-nucleus interaction model G4 INCL_PC, whereas the right plots compare two secondary interaction models (G4 INCL_PC and G4 BERT_PC) coupled with the neutrino event generator option GENIE hN.

- [1] P. Abratenko *et al.* (MicroBooNE Collaboration), *Phys. Rev. D* **110**, 013006 (2024).
- [2] M. Elkins *et al.* (MINERvA Collaboration), *Phys. Rev. D* **100**, 052002 (2019).
- [3] P. Abratenko *et al.* (MicroBooNE Collaboration), *Eur. Phys. J. C* **84**, 1052 (2024).
- [4] A. M. Ankowski, P. Coloma, P. Huber, C. Mariani, and E. Vagnoni, *Phys. Rev. D* **92**, 091301 (2015).
- [5] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, *Science* **124**, 103 (1956).
- [6] T. Wester *et al.* (Super-Kamiokande Collaboration), *Phys. Rev. D* **109**, 072014 (2024).
- [7] A. Takenaka *et al.* (Super-Kamiokande Collaboration), *Phys. Rev. D* **102**, 112011 (2020).
- [8] S. Agostinelli *et al.* (Geant4 Collaboration), *Nucl. Instrum. Methods Phys. Res., Sect. A* **506**, 250 (2003).
- [9] J. Allison *et al.*, *IEEE Trans. Nucl. Sci.* **53**, 270 (2006).
- [10] J. Allison *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **835**, 186 (2016).
- [11] J. Cugnon, *Nucl. Phys.* **A387**, 191 (1982).
- [12] J.-C. David *et al.*, *Prog. Nucl. Sci. Technol.* **2**, 942 (2011).
- [13] R. Akutsu, Ph.D. thesis, The University of Tokyo, 2019.
- [14] B. Aharmim *et al.* (SNO Collaboration), *Phys. Rev. D* **99**, 112007 (2019).
- [15] A. Olivier *et al.* (MINERvA Collaboration), *Phys. Rev. D* **108**, 112010 (2023).
- [16] S. Abe *et al.* (KamLAND Collaboration), *Phys. Rev. D* **107**, 072006 (2023).
- [17] S. Sakai *et al.* (Super-Kamiokande Collaboration), *Phys. Rev. D* **109**, L011101 (2024).
- [18] S. Fukuda *et al.* (Super-Kamiokande Collaboration), *Nucl. Instrum. Methods Phys. Res., Sect. A* **501**, 418 (2003).
- [19] K. Abe *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **737**, 253 (2014).
- [20] S. Yamada *et al.*, *IEEE Trans. Nucl. Sci.* **57**, 428 (2010).
- [21] K. Abe *et al.* (Super-Kamiokande Collaboration), *Nucl. Instrum. Methods Phys. Res., Sect. A* **1027**, 166248 (2022).
- [22] K. Abe *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1065**, 169480 (2024).
- [23] M. B. Chadwick *et al.*, *Nucl. Data Sheets* **112**, 2887 (2011).
- [24] M. Shiozawa, *Nucl. Instrum. Methods Phys. Res., Sect. A* **433**, 240 (1999).
- [25] C. Andreopoulos *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **614**, 87 (2010).
- [26] C. Andreopoulos *et al.*, [arXiv:1510.05494](https://arxiv.org/abs/1510.05494).
- [27] J. Tena-Vidal *et al.* (GENIE Collaboration), *Phys. Rev. D* **104**, 072009 (2021).
- [28] M. Honda, T. Kajita, K. Kasahara, and S. Midorikawa, *Phys. Rev. D* **83**, 123001 (2011).
- [29] R. L. Workman *et al.* (Particle Data Group), *Prog. Theor. Exp. Phys.* **083C01**, 729 (2022).
- [30] Y. Hayato and L. Pickering, *Eur. Phys. J. Special Topics* **230**, 4469 (2021).
- [31] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio, *Nucl. Phys.* **A484**, 557 (1988).
- [32] E. S. P. Guerra *et al.*, *Phys. Rev. D* **99**, 052007 (2019).
- [33] H. W. Bertini, *Phys. Rev.* **131**, 1801 (1963).
- [34] P. de Perio, NEUT pion FSI, in *AIP Conf. Proc.*, edited by S. K. Singh, J. G. Morfin, M. Sakuda, and K. D. Purohit (AIP, Melville, NY, 2011).
- [35] A. M. Ankowski, O. Benhar, T. Mori, R. Yamaguchi, and M. Sakuda, *Phys. Rev. Lett.* **108**, 052505 (2012).
- [36] H. Ejiri, *Phys. Rev. C* **48**, 1442 (1993).
- [37] C. Zeitnitz and T. A. Gabriel, *Nucl. Instrum. Methods Phys. Res., Sect. A* **349**, 106 (1994).
- [38] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanzarini, GEANT3, CERN Technical Report No. CERN-DD-EE-84-1, 1987).
- [39] H. W. Bertini, *Phys. Rev.* **188**, 1711 (1969).
- [40] P. F. Rose, ENDF-201: ENDF/B-VI Summary Documentation, BNL Technical Report No. BNL-NCS-17541, 1991.
- [41] E. Mendoza and D. Cano-Ott, Update of the evaluated neutron cross section libraries for the Geant4 code, IAEA Technical Report No. INDC(NDS)-0758, 2018.
- [42] K. Hagiwara *et al.*, *Prog. Theor. Exp. Phys.* **2019**, 023D01 (2019).
- [43] T. Tanaka *et al.*, *Prog. Theor. Exp. Phys.* **2020**, 043D02 (2020).
- [44] V. Barger, K. Whisnant, S. Pakvasa, and R. J. N. Phillips, *Phys. Rev. D* **22**, 2718 (1980).
- [45] M. Jiang *et al.* (Super-Kamiokande Collaboration), *Prog. Theor. Exp. Phys.* **2019**, 053F01 (2019).
- [46] A. M. Dziewonski and D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981).
- [47] K. Abe *et al.* (Super-Kamiokande Collaboration), *J. Instrum.* **17**, P10029 (2022).
- [48] S. Han, Master's thesis, The University of Tokyo, 2021.
- [49] F. Chollet *et al.*, Keras, <https://keras.io> (2015).
- [50] K. He, X. Zhang, S. Ren, and J. Sun, *2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile* (IEEE, New York, 2015).
- [51] D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [52] H. Ito, K. Wada, T. Yano, Y. Hino, Y. Ommura, M. Harada, A. Minamino, and M. Ishitsuka, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1057**, 168701 (2023).
- [53] S. Han, Ph.D. thesis, The University of Tokyo, 2024.
- [54] Y. Hino *et al.*, *Prog. Theor. Exp. Phys.* **2025**, 013C01 (2024).
- [55] B. Pritychenko and S. F. Mughabghab, *Nucl. Data Sheets* **113**, 3120 (2012).
- [56] T. J. Hastie and R. J. Tibshirani, *Stat. Sci.* **1**, 297 (1986).
- [57] D. Servén, C. Brummitt, and H. Abedi, *dswah/pyGAM: v0.8.0* (2018).
- [58] D. Wright and M. Kelsey, *Nucl. Instrum. Methods Phys. Res., Sect. A* **804**, 175 (2015).
- [59] D. Mancusi, A. Boudard, J. Cugnon, J.-C. David, P. Kaitaniemi, and S. Leray, *Phys. Rev. C* **90**, 054602 (2014).
- [60] A. Kelic, M. V. Ricciardi, and K.-H. Schmidt, [arXiv:0906.4193](https://arxiv.org/abs/0906.4193).
- [61] J. Quesada Molina, V. Ivanchenko, A. Ivanchenko, M. Cortés-Giraldo, G. Folger, A. Howard, and D. Wright, *Prog. Nucl. Sci. Technol.* **2**, 936 (2011).
- [62] V. M. Grichine, *Eur. Phys. J. C* **62**, 399 (2009).
- [63] V. Weisskopf, *Phys. Rev.* **52**, 295 (1937).
- [64] I. Dostrovsky, Z. Fraenkel, and G. Friedlander, *Phys. Rev.* **116**, 683 (1959).
- [65] A. Chatterjee, K. H. N. Murthy, and S. K. Gupta, *Pramana* **16**, 391 (1981).
- [66] C. Kalbach, PRECO-2000, Exciton model Preequilibrium code system with direct reactions, NEA Technical Report No. PSR-0226, 2002.

- [67] A. Boudard and J. Cugnon, INCL4—The Liège INC Model for high-energy hadron-nucleus reactions: A short description of the INCL4.2 and INCL4.4 versions, IAEA Technical Report, INDC(NDS)-0530, 2008.
- [68] A. Boudard, J. Cugnon, J.-C. David, S. Leray, and D. Mancusi, *Phys. Rev. C* **87**, 014606 (2013).
- [69] A. Ershova *et al.*, *Phys. Rev. D* **106**, 032009 (2022).
- [70] D. Mancusi *et al.*, *Eur. Phys. J. A* **53**, 80 (2017).
- [71] T. Aoust and J. Cugnon, *Phys. Rev. C* **74**, 064607 (2006).
- [72] S. Dytman, Y. Hayato, R. Raboanary, J. T. Sobczyk, J. Tena-Vidal, and N. Vololoniaina, *Phys. Rev. D* **104**, 053006 (2021).
- [73] Y. Ashida *et al.*, *Phys. Rev. C* **109**, 014620 (2024).
- [74] T. Tano *et al.*, *Prog. Theor. Exp. Phys.* **2024**, 113D01 (2024).
- [75] Y. Hino, Y. Ashida, T. Tano, and Y. Koshio, [arXiv: 2506.14388](https://arxiv.org/abs/2506.14388).
- [76] Super-Kamiokande Collaboration, [10.5281/zenodo.15392411](https://zenodo.org/record/15392411) (2025).
- [77] M. Smy, Low energy event reconstruction and selection in Super-Kamiokande-III, in *Proceeding of the 30th Int. Cosmic Ray Conf.*, edited by R. Caballero, J.C. D’Olivo, G. Medina-Tanco, L. Nellen, F.A. Sánchez, and J.F. Valdés-Galicia (UNAM, Mexico City, Mexico, 2008).