

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

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

(The Super-Kamiokande Collaboration)

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

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

<sup>3</sup>Department of Theoretical Physics, University Autónoma Madrid, 28049 Madrid, Spain

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

<sup>5</sup>Department of Physics, Boston University, Boston, MA 02215, USA

<sup>6</sup>Department of Physics, British Columbia Institute of Technology, Burnaby, BC, V5G 3H2, Canada

<sup>7</sup>Department of Physics and Astronomy, University of California, Irvine, Irvine, CA 92697-4575, USA

<sup>8</sup>Department of Physics, California State University, Dominguez Hills, Carson, CA 90747, USA

<sup>9</sup>Institute for Universe and Elementary Particles, Chonnam National University, Gwangju 61186, Korea

<sup>10</sup>Department of Physics, Duke University, Durham NC 27708, USA

<sup>11</sup>Ecole Polytechnique, IN2P3-CNRS, Laboratoire Leprince-Ringuet, F-91120 Palaiseau, France

<sup>12</sup>Department of Physics, Gifu University, Gifu, Gifu 501-1193, Japan

<sup>13</sup>GIST College, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea

<sup>14</sup>School of Physics and Astronomy, University of Glasgow, Glasgow, Scotland, G12 8QQ, United Kingdom

<sup>15</sup>Department of Physics and Astronomy, University of Hawaii, Honolulu, HI 96822, USA

<sup>16</sup>Center for Underground Physics, Institute for Basic Science (IBS), Daejeon, 34126, Korea

<sup>17</sup>Institute For Interdisciplinary Research in Science and Education, ICISE, Quy Nhon, 55121, Vietnam

<sup>18</sup>Department of Physics, Imperial College London , London, SW7 2AZ, United Kingdom

<sup>19</sup>Dipartimento Interuniversitario di Fisica, INFN Sezione di Bari and Università e Politecnico di Bari, I-70125, Bari, Italy

<sup>20</sup>Dipartimento di Fisica, INFN Sezione di Napoli and Università di Napoli, I-80126, Napoli, Italy

<sup>21</sup>Dipartimento di Fisica, INFN Sezione di Padova and Università di Padova, I-35131, Padova, Italy

<sup>22</sup>INFN Sezione di Roma and Università di Roma "La Sapienza", I-00185, Roma, Italy

<sup>23</sup>ILANCE, CNRS - University of Tokyo International Research Laboratory, Kashiwa, Chiba 277-8582, Japan

<sup>24</sup>Department of Physics, Keio University, Yokohama, Kanagawa, 223-8522, Japan

<sup>25</sup>High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan

<sup>26</sup>Department of Physics, King's College London, London, WC2R 2LS, UK

<sup>27</sup>Department of Physics, Kobe University, Kobe, Hyogo 657-8501, Japan

<sup>28</sup>Department of Physics, Kyoto University, Kyoto, Kyoto 606-8502, Japan

<sup>29</sup>Department of Physics, University of Liverpool, Liverpool, L69 7ZE, United Kingdom

<sup>30</sup>School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA

<sup>31</sup>Department of Physics, Miyagi University of Education, Sendai, Miyagi 980-0845, Japan

<sup>32</sup>Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi 464-8602, Japan

<sup>33</sup>Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Nagoya, Aichi 464-8602, Japan

<sup>34</sup>National Centre For Nuclear Research, 02-093 Warsaw, Poland

<sup>35</sup>Department of Physics and Astronomy, State University of New York at Stony Brook, NY 11794-3800, USA

<sup>36</sup>Department of Physics, Okayama University, Okayama, Okayama 700-8530, Japan

<sup>37</sup>Media Communication Center, Osaka Electro-Communication University, Neyagawa, Osaka, 572-8530, Japan

<sup>38</sup>Department of Physics, Oxford University, Oxford, OX1 3PU, United Kingdom

<sup>39</sup>Rutherford Appleton Laboratory, Harwell, Oxford, OX11 0QX, UK

<sup>40</sup>Department of Physics and Astronomy, Seoul National University, Seoul 151-742, Korea

<sup>41</sup>Department of Physics and Astronomy, University of Sheffield, S3 7RH, Sheffield, United Kingdom

<sup>42</sup>Department of Informatics in Social Welfare, Shizuoka University of Welfare, Yaizu, Shizuoka, 425-8611, Japan

<sup>43</sup>August Cichkowski Institute of Physics, University of Silesia in Katowice, 75 Putku Piechoty 1, 41-500 Chorzów, Poland

<sup>44</sup>STFC, Rutherford Appleton Laboratory, Harwell Oxford, and

Daresbury Laboratory, Warrington, OX11 0QX, United Kingdom

<sup>45</sup>Department of Physics, Sungkyunkwan University, Suwon 440-746, Korea

<sup>46</sup>Department of Physics, Faculty of Science, Tohoku University, Sendai, Miyagi, 980-8578, Japan

<sup>47</sup>Department of Physics, Tokai University, Hiratsuka, Kanagawa 259-1292, Japan

<sup>48</sup>Department of Physics, University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

<sup>49</sup>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

<sup>50</sup>Department of Physics, Institute of Science Tokyo, Meguro, Tokyo 152-8551, Japan

<sup>51</sup>Department of Physics, Faculty of Science and Technology, Tokyo University of Science, Noda, Chiba 278-8510, Japan

<sup>52</sup>Faculty of Science, University of Toyama, Toyama City, Toyama 930-8555, Japan

<sup>53</sup>TRIUMF, 4004 Wesbrook Mall, Vancouver, BC, V6T2A3, Canada

<sup>54</sup>Department of Engineering Physics, Tsinghua University, Beijing, 100084, China

<sup>55</sup>Faculty of Physics, University of Warsaw, Warsaw, 02-093, Poland

<sup>56</sup>Department of Physics, University of Warwick, Coventry, CV4 7AL, UK

<sup>57</sup>Department of Physics, University of Winnipeg, MB R3J 3L8, Canada

<sup>58</sup>Department of Physics, Yokohama National University, Yokohama, Kanagawa, 240-8501, Japan

(Dated: January 23, 2025)

We present measurements of neutron production from atmospheric neutrino interactions in water, analyzed as a function of the 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. The measurements are compared to predictions from neutrino event generators combined with multiple hadron-nucleus interaction models, which describe intranuclear hadron transport and subsequent nuclear de-excitation. Significant variations were observed in the predictions depending on the choice of hadron-nucleus interaction models. Among the models tested, our data favors models that predict relatively fewer secondary neutrons. We also discuss the model-specific features that contribute to these variations.

teraction 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, and advancements in neutrino detectors have enabled precise measurements of hadron multiplicities and kinematics. For instance, measurements of proton production in neutrino-argon interactions have revealed several discrepancies between observed and predicted kinematic distributions [1].

Detecting neutrons is challenging since they do not leave ionization tracks. Tracking detectors identify neutrons via displaced proton tracks, with detection efficiency limited by the small  $(n, p)$  reaction cross section [2]. This issue is critical with GeV-scale neutrino interactions, where neutrons often carry substantial energy. Misestimating this “missing energy” can bias key measurements, such as the Dirac CP phase [3].

Neutrons can also be detected via the radiative neutron capture  $(n, \gamma)$  reaction, which has no practical energy threshold. Neutrons with kinetic energies of a few MeV or lower are likely to thermalize, undergoing  $(n, \gamma)$  reactions with well-defined timescales and energy signatures that enable clean signal selection. This characteristic historically made neutrons effective tags for antineutrino charged-current (CC) interactions (e.g.,  $\bar{\nu}_\mu p \rightarrow \mu^+ n$ ) compared to neutrino interactions (e.g.,  $\nu_\mu n \rightarrow \mu^- p$ ). 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 [4]. Additionally, neutron tagging helps suppress atmospheric neutrino backgrounds for rare events, such as proton decay (e.g.,  $p \rightarrow e^+ \pi^0$  [5]), 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 typically produces only one neutron. Accurately predicting the number of detectable neutrons from neutrino interactions is crucial for these analyses and requires well-constrained uncertainties in interaction models.

The modeling approach commonly adopted by GeV-scale neutrino experiments is illustrated in Figure 1. Neutrino event generators simulate outgoing hadrons from the initial neutrino interaction, either at the nucleon or quark level, and subsequently model intranuclear hadron transport (often referred to as final-state interactions, or FSI), followed by nuclear de-excitation. Similarly, particle transport codes such as Geant4 [6–8] simulate hadron transport and nuclear de-excitation to describe downstream hadron-nucleus interactions within the detector.

Accurate prediction of the observable neutron multiplicity requires careful consideration of secondary neutron production. For hadron transport in the  $O(0.1\text{--}1)$  GeV energy range, semi-classical Intranuclear Cascade (INC) models [9] are commonly used. These models approximate the process as a series of binary collisions be-

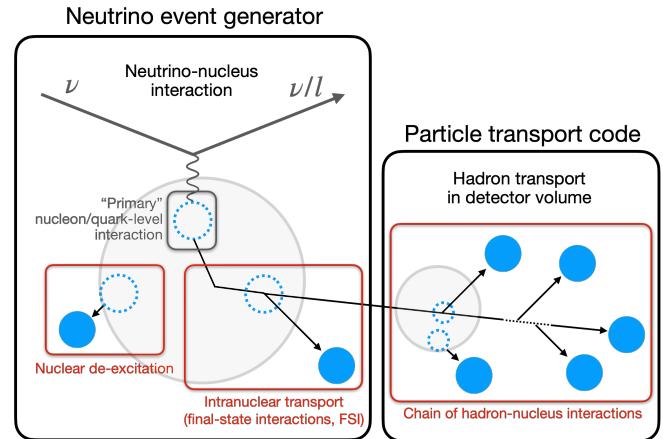


FIG. 1. Schematic of sources of nucleon production in a typical GeV-scale neutrino event simulation setup. Dashed blue circles represent nucleon holes, black arrows represent nucleon trajectories, and solid blue circles represent detectable nucleons ejected from nuclei.

tween a hadron projectile and a quasi-free nucleon within the target nucleus. Subsequent nuclear de-excitation, modeled as a statistical process that releases neutrons with kinetic energies of a few MeV, also contribute significantly to neutron production at lower energies. Notably, variations in how these models account for nuclear effects often lead to significant discrepancies in predictions [10]. Several studies have measured neutron production from atmospheric or artificial neutrino interactions using water (T2K [11]), heavy water (SNO [12]), and hydrocarbon (MINERvA [13, 14]) as target materials. These studies 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 from events fully contained within the Super-Kamiokande detector from 2008 to 2022, we evaluated the average multiplicity of  $(n, \gamma)$  reactions (or “neutron captures”) as a function of electron-equivalent visible energy, a calorimetric proxy for neutrino momentum transfer. These observations were compared with predictions generated using various models relevant to secondary neutron production.

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 Sections III and IV. Section V outlines the methodology for determining the average  $(n, \gamma)$  multiplicity per visible energy bin and the associated systematic uncertainties. Section VI introduces the interaction models used for generating predictions. Finally, Sections VII and VIII present a comparison of observations with predictions and discuss the implications of the results.

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

Phase	Dates	Livetime [days]	Gd concentration <sup>a</sup> [wt%]	H( $n, \gamma$ ) [%]	Expected ( $n, \gamma$ ) ratio <sup>b</sup> H( $n, \gamma$ ) [%]	( $n, \gamma$ ) time constant <sup>c</sup> [μs]
SK-I-III	1996-2008	2805.9	-	>99.9	-	No data
SK-IV	2008-2018	3244.4	-	>99.9	-	204.8 ± 9.8
SK-V	2019-2020	461.0	-	>99.9	-	199.8 ± 10.2
SK-VI	2020-2022	564.4 <sup>d</sup>	0.0110 ± 0.0001 [16]	56.1 ± 1.5	43.9 ± 1.5	116.2 ± 2.3
SK-VII-VIII	2022-present	-	0.0332 ± 0.0002 [17]	29.7 ± 0.7	70.3 ± 0.7	61.8 ± 0.1 [17]

<sup>a</sup> Based on the amount of dissolved Gd.

<sup>b</sup> Based on the evaluated thermal ( $n, \gamma$ ) reaction cross sections and uncertainties of ENDF/B-VII.1 [15].

<sup>c</sup> Weighted mean of all Am/Be neutron source measurements, explained in Section IV B.

<sup>d</sup> Excludes earlier runs which showed signs of non-uniform Gd concentration, i.e., varying time constant by position.

## 209 II. THE SUPER-KAMIOKANDE DETECTOR

210 Super-Kamiokande (SK) [18] is an underground water  
 211 Cherenkov detector located in Gifu, Japan. It consists of  
 212 two optically separated, concentric cylindrical volumes:  
 213 the inner detector (ID) containing 32.5 ktons of water  
 214 and equipped with 11,129 inward-facing photomultiplier  
 215 tubes (PMTs) and the outer detector serving as a cosmic-  
 216 ray veto. The detector registers a PMT signal with over  
 217 0.25 photoelectron-equivalent charge as a “hit”. If the  
 218 number of ID or OD PMT hits within a 200-ns sliding  
 219 time window ( $N_{200\text{-ns}}$ ) exceeds a given threshold, an  
 220 event trigger is issued. The details of the detector can be  
 221 found in [18, 19].

222 Charged particles, namely electrons and muons pro-  
 223 duced by charged-current neutrino interactions, are iden-  
 224 tified through Cherenkov radiation. The radiation is pro-  
 225 jected onto the PMTs as a characteristic ring pattern  
 226 that depends on the particle type and energy. This ring  
 227 pattern serves as the basis for particle reconstruction.  
 228 Neutrons are indirectly identified via Compton-scattered  
 229 electrons resulting from ( $n, \gamma$ ) reactions. In pure water,  
 230 most occur on  $^1\text{H}$ , emitting a single 2.2 MeV gamma-  
 231 ray. With the recent addition of gadolinium (Gd), the  
 232 majority of neutrons are expected to be captured by Gd  
 233 isotopes, resulting in a total gamma-radiated energy of  
 234 around 8 MeV.

235 The  $O(1)$  MeV signal identification performance is sig-  
 236 nificantly influenced by variations in detector character-  
 237 istics. Parameters such as individual PMT gain, tim-  
 238 ing properties, quantum efficiency, and optical absorption  
 239 and scattering in water are continuously monitored using  
 240 cosmic-ray muons and light sources [19]. Additionally,  
 241 the uncertainty in Cherenkov ring energy reconstruc-  
 242 tion (described in Section III B) is evaluated over a wide en-  
 243 ergy range using naturally occurring particles, including  
 244 cosmic-ray muons, Michel electrons, and neutral pions  
 245 produced in neutral-current (NC) atmospheric neutrino  
 246 interactions in water. Figure 2 illustrates the agreement  
 247 between data and simulation in energy reconstruction for  
 248 the Gd-loaded SK-VI phase, which is mostly within 2%  
 249 across the  $O(10\text{-}10^4)$  MeV range and consistent with the  
 250 pure water phase results reported in [4].

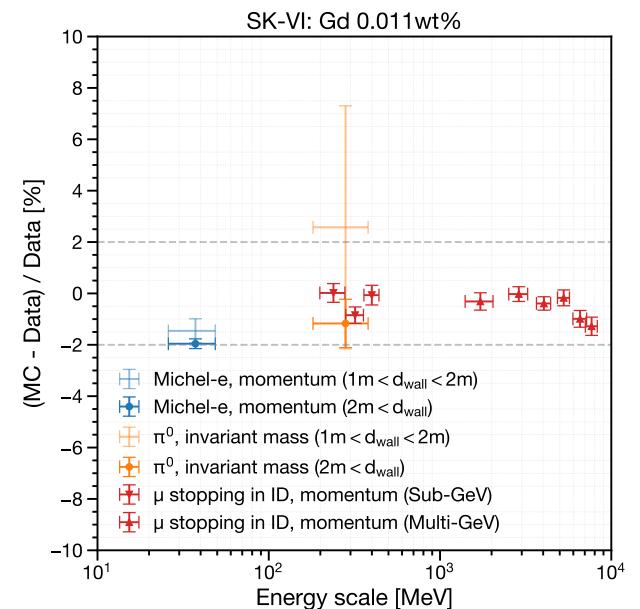


FIG. 2. Comparison of energy reconstruction performance between Monte Carlo (MC) simulation and observed data for naturally occurring particles in the Gd-loaded SK-VI phase. The variable  $d_{\text{wall}}$  represents the distance (in meters) from the reconstructed vertex to the nearest ID photodetector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

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 time scale 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 [16, 17]. Table I summarizes the relevant operational conditions.

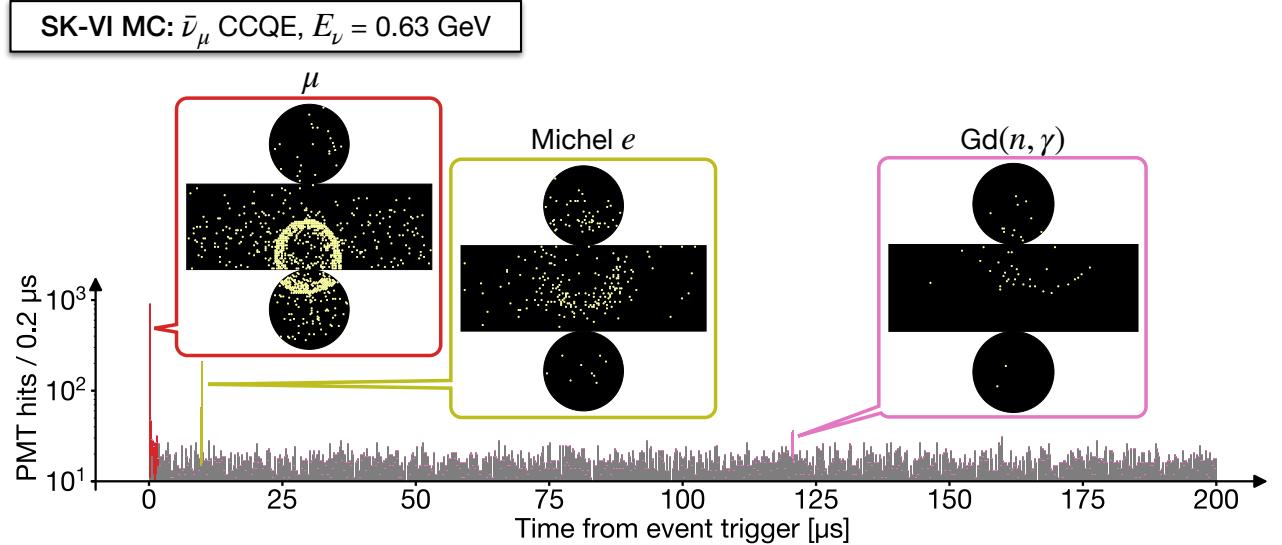


FIG. 3. A PMT hit time distribution for a typical  $\bar{\nu}_\mu$  charged-current (CC) quasi-elastic (QE) interaction with a muon and a neutron as final states. The zero is set at the event trigger. The event displays feature the “prompt” muon signal (red) and the two types of “delayed” coincident 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.

### 265 III. ATMOSPHERIC NEUTRINO EVENTS

#### 266 A. Data reduction

267 We followed a typical selection process for atmospheric  
268 neutrino interactions that are fully contained within the  
269 ID, similar to previous studies conducted at SK [4]. All  
270 events were required to pass the ID trigger with the  
271 threshold  $N_{200\text{-ns}} \geq 58$  PMT hits—roughly correspond-  
272 ing to a 10 MeV electron—followed by the extended 535  
273  $\mu\text{s}$  event window for neutron detection. Background  
274 events arising from cosmic-ray muons, low-energy ra-  
275 dioactivity, and neutrino events with particles exiting the  
276 tank were reduced through OD veto and ID charge cuts.  
277

278 Selected events were reconstructed as described in Sec-  
279 tion III B. To further reject low-energy backgrounds, we  
280 required that the reconstructed vertex be more than 1 m  
281 away from the ID tank wall (defining the fiducial volume  
282 with 23.2 kton of water) and that the visible energy be  
283 larger than 30 MeV. The remaining background contam-  
284 ination, mainly due to corner-clipping cosmic-ray muons  
285 and PMT discharges, was estimated to be below 0.2%,  
286 based on visual inspection [5].

#### 286 B. Reconstruction of prompt Cherenkov rings

287 Figure 3 shows a PMT hit time distribution of a typi-  
288 cal  $\bar{\nu}_\mu$  charged-current (CC) quasi-elastic (QE) event fol-  
289 lowed by a Michel electron and a  $Gd(n, \gamma)$  reaction, along  
290 with the corresponding event displays. For the “prompt”  
291 radiation due to charged particles (namely, electrons and  
292

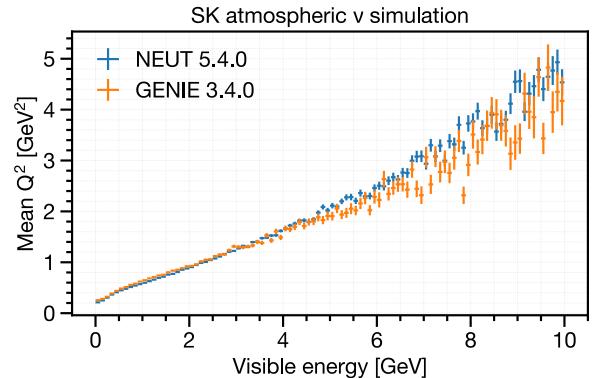


FIG. 4. Average squared momentum transfer ( $Q^2$ ) per visible energy bin for simulated atmospheric neutrino events at SK, compared between two different neutrino event genera-  
tors. NEUT 5.4.0 (blue) uses the baseline setup described in  
Section III C, while GENIE 3.4.0 [21–23] uses the “hN” setup  
as described in Section VI.

292 muons) produced via the primary neutrino interaction,  
293 we followed the typical Cherenkov ring reconstruction  
294 process [24] as applied in previous SK analyses [4, 5].

295 The visible energy of an event is defined as the sum of  
296 the reconstructed kinetic energies of all Cherenkov rings,  
297 assuming each ring originates from an electron. This  
298 calorimetric measure serves as a reliable proxy for neu-  
299 trino momentum transfer, as demonstrated by the pos-  
300 itive correlation shown in Figure 4.

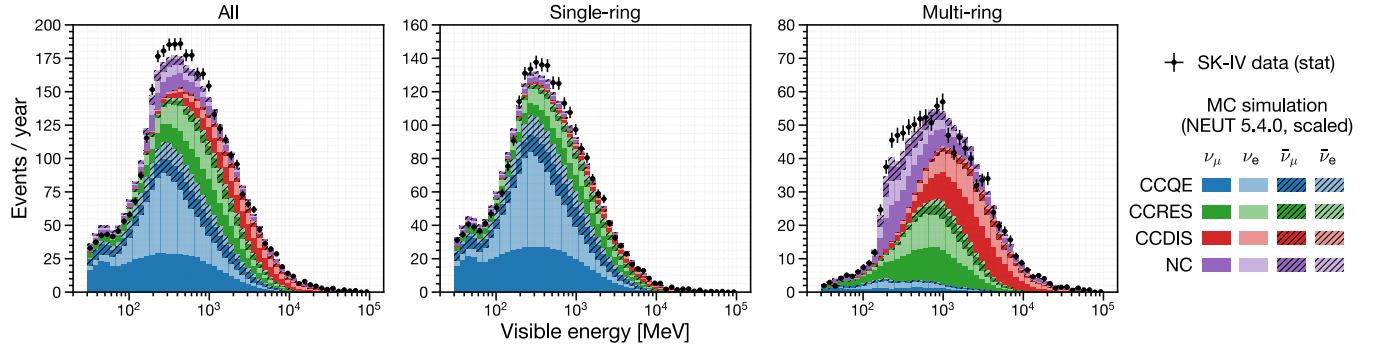


FIG. 5. Visible energy distributions for observed and simulated atmospheric neutrino interactions in water, fully contained within the fiducial volume: all events (left), events with one reconstructed Cherenkov ring (middle), and events with two or more rings (right). CCQE events dominate the single-ring sample, while RES and DIS events dominate the multi-ring sample. The simulation setup is detailed in Section III C. Simulated distributions are scaled to match the SK-IV data to account for neutrino flux uncertainties.

### 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  $\nu_e$ ,  $\nu_\mu$ ,  $\bar{\nu}_e$ , and  $\bar{\nu}_\mu$  at the detector site, assuming no oscillations, as provided by Honda et al. [25] for neutrino energies ranging from 100 MeV to 10 TeV. Interactions of  $\nu_\tau$  and  $\bar{\nu}_\tau$  were not simulated.

Neutrino interactions in water were modeled using neutrino event generators [26], which compute the cross sections for each interaction channel and sample outgoing particle kinematics. NEUT 5.4.0 [27] 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  $\Delta$  resonance (RES), and multiple pion production via deep inelastic scattering (DIS). The models used are consistent with the latest SK atmospheric neutrino oscillation analysis [4].

The transport of hadronic final states within the target nucleus (FSI) was modeled separately from the primary neutrino interaction. The transport of low-momentum ( $< 500$  MeV/c) pions was based on Salcedo et al. [28], with parameters tuned to fit external pion-nucleus scattering data [29]. 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 [30] for nucleon projectiles and from external pion scattering data for high-momentum pion projectiles [31]. The de-excitation of an oxygen target was modeled using tabulated occupation probabilities for nucleon energy states [32] and branching ratios for knockout of each state [33].

The subsequent particle transport in water and detector responses were simulated using GCALOR [34] cou-

pled with Geant3.21 [35]. Hadron propagation above 10 GeV was simulated using FLUKA [36], while those below 10 GeV was simulated based on the Bertini cascade model [37]. For low-energy neutrons with kinetic energy below 20 MeV, GCALOR calls MICAP [38], using ENDF/B-V [39] cross sections. The transport of low-momentum pions was separately modeled using the NEUT pion FSI routine to ensure consistency between NEUT and Geant3.

For the transport of low-energy neutrons in the Gd-loaded SK-VI phase, we imported reaction cross sections from the Geant4.10.03.p01 [6–8] NeutronHP database [40] based on ENDF/B-VII.1 [15], replacing MICAP for neutron energies below 20 MeV. Additionally, we modeled the gamma-cascade resulting from neutron captures on  $^{155}/^{157}\text{Gd}$  using the ANNRI-Gd model [41, 42].

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 Figure 3.

500 years of atmospheric neutrino events were simulated for each SK phase and processed as described in Sections III B and III A. The events in the final sample were weighted based on the standard three-flavor oscillation probability in matter [43], using the oscillation parameters fitted with reactor constraints in the previous SK analysis [44] and the PREM model [45] 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, demonstrating reasonable agreement with the observations. Single-ring events exhibit a higher fraction of QE interactions and a smaller fraction of DIS compared to multi-ring events. For the same visible energy, multi-ring events are expected to produce more secondary neutrons than single-ring events, due to a higher fraction of DIS interactions.

378 **IV. NEUTRON SIGNAL SELECTION** 432

379 Neutrons produced by atmospheric neutrino interactions in the SK ID are primarily captured by  $^1\text{H}$  or 434  
 380  $^{155}/^{157}\text{Gd}$  isotopes within approximately  $O(100)$   $\mu\text{s}$ , with 435  
 381 expected capture ratios summarized in Table I. Electrons 436 that are scattered by the  $O(1)$  MeV gamma-rays generate 437 Cherenkov photons that form faint rings on the ID tank wall 438 (hereafter referred to as “neutron signal”), as shown in Figure 3. 439

387 Fast neutrons, along with the subsequent gamma-rays 440 and scattered electrons, typically travel only a few tens 441 of centimeters in water. Thus, signal photons can be approximated as originating from a single vertex. While 442 reconstructing this vertex is challenging due to the limited 443 number of PMT hits, *assuming* it to be near another known 444 vertex, such as the reconstructed neutrino interaction 445 vertex, helps identify signals from random PMT coincidences 446 caused by dark current. Remaining Michel electrons from muon decays can be effectively suppressed 447 through cuts based on signal timing and energy. 448

450 **A. Signal selection algorithm** 451

399 The signal selection algorithm is based on Ref. [46], 452 and consists of two stages. 453

401 In the first candidate search stage, we initially subtract 455 the expected photon time-of-flight (ToF) from the individual 456 PMT hit times for a given *assumed* signal vertex, 457 which was provided by the reconstructed neutrino event 458 vertex. Then, we slide a time window of 14 ns width 459 on the ToF-corrected PMT hit times to trigger on the 460 number of included PMT hits. The threshold was 5 for 461 pure water phases SK-IV and V, and 7 for the Gd-loaded 462 phase SK-VI. The search span for each event was [18, 534] 463  $\mu\text{s}$  from the event trigger for SK-IV and V, and [3, 534] 464  $\mu\text{s}$  for SK-VI with a shorter neutron capture time. For 465 overlapping candidates within 50 ns, only the candidate 466 with the largest number of PMT hits is selected. 467

468 In the second candidate classification stage, we extract 469 features of each candidate and use a neural network to 470 classify each candidate into signal and noise based on 471 input features. These features characterize the signal 472 energy, the background hit level, timing spread assuming 473 the vertex, correlation between the input vertex and the 474 hit PMT positions, correlation to the known properties 475 of PMT noise, and angular correlation among hit PMTs 476 relative to the Cherenkov cone opening angle. 477

478 The major changes from the original algorithm [46] include 479 a simplified algorithm, a reduced set of features, 480 and a heuristically tuned neural network architecture. 481 These modifications aim to reduce performance bias between 482 the data and the simulation that is used to train 483 the neural networks. 484

485 Here, we provide the definition and unit of each feature 486 used for the classification of signal candidates, along with 487 their expected distributions as shown in Figure 6. 488

- **NHits**

The number of selected PMT hits within the 14-ns sliding time window.

- **NResHits**

The number of PMT hits within [-100,+100] ns from the center of the 14-ns sliding time window, subtracted from NHits.

- **TRMS** [ns]

The root mean square (RMS) of the ToF-corrected time distribution of the selected PMT hits.

- **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.

- **DWall** [cm]

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

- **DWallMeanDir** [cm]

The shorter of the radial and vertical distances from the *assumed* signal vertex to the tank wall, weighted by the mean of the unit vectors connecting the vertex to each hit PMT.

- **BurstRatio**

The ratio of the selected PMTs with a preceding hit within 10  $\mu\text{s}$ , which are likely caused by scintillation within the irradiated PMT glass.

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

- **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.

- **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  $k^{\text{th}}$  Legendre polynomial and  $\theta_{ij}$  is the opening angle between the assumed signal vertex and the  $i^{\text{th}}$  and  $j^{\text{th}}$  hit PMTs.

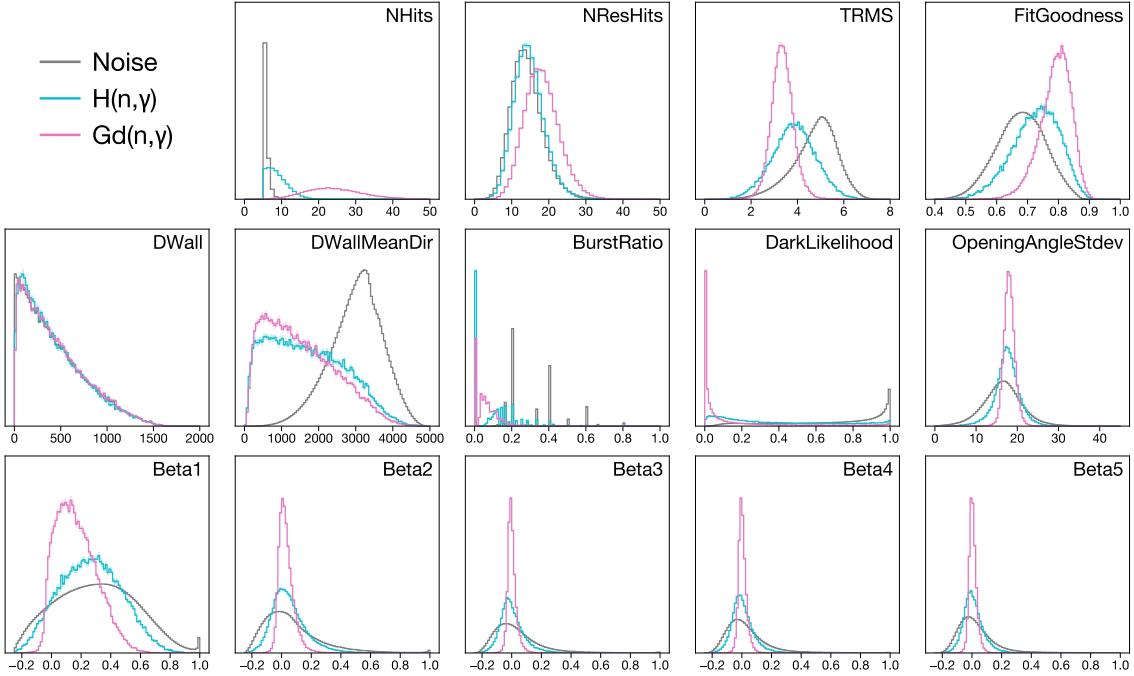


FIG. 6. The features (area-normalized) of neutron signals and noise from the thermal neutron MC simulation used for training the neural network for SK-VI phase.

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  $(n, \gamma)$  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 [47]. 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. [48] and optimized by minimizing the binary cross-entropy loss iteratively on minibatches of size 2,048 using the Adam optimizer [49]. 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 ( $\text{NHits} > 50$ ) 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 Figure 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  $98.7 \pm 0.5\%$ .

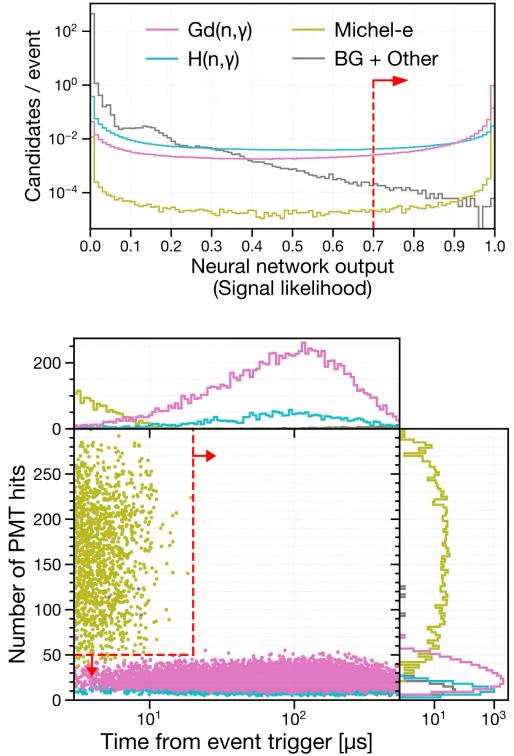


FIG. 7. The neural network response to neutron signals and backgrounds in the SK-VI atmospheric neutrino simulation (top), and the time versus energy distribution of signal candidates passing the neural network selection (bottom). Red arrows indicate the corresponding cut points.

## 506 B. Signal selection performance on calibration data

507 An Am/Be neutron source with a measured total intensity  
 508 of  $236.8 \pm 5.0$  neutrons/s [50] was used to obtain cal-  
 509 ibration data for estimating the signal detection perfor-  
 510 mance. The first-excited state of the alpha-absorbed  ${}^9\text{Be}$ ,  
 511 with a roughly 60% branching ratio, emits a fast neutron  
 512 and a 4.44 MeV gamma-ray simultaneously. This source  
 513 was encapsulated with Bismuth Germanate ( $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ ,  
 514 BGO) crystals so that the 4.44 MeV gamma-rays can  
 515 induce scintillation. The setup was deployed in various  
 516 positions within the ID, and events were recorded for 30  
 517 minutes to 1 hour. Events with a trigger charge yield  
 518 corresponding to the 4.44 MeV gamma-ray scintillation  
 519 were regarded as the single neutron control sample.

520 The observed light yield distribution was compared  
 521 with dedicated simulation, as shown in Figure 8. The  
 522 simulation accounts for continuous source activity and  
 523 pile-up, by reorganizing the simulated detector response  
 524 to Am/Be neutron emission on a single global time axis,  
 525 based on the measured total neutron intensity and the es-  
 526 timated branching ratios to each excited state of alpha-  
 527 absorbed  ${}^9\text{Be}$ . As shown in Figure 8, this simulation  
 528 accurately models event triggers due to ambient neutron  
 529 captures and neutron inelastic interactions within scintil-  
 530 lator crystal. The contamination of such unwanted event  
 531 triggers in the 4.44 MeV gamma-ray event selection was  
 532 estimated to be at a few percent level.

533 Within the selected events in the single neutron con-  
 534 trol sample, signal candidates were obtained following the  
 535 algorithm described in Section IV A, with the *assumed*  
 536 signal vertex set at the source position.

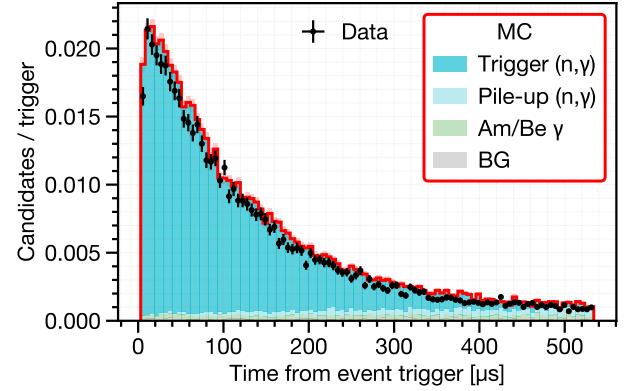


FIG. 9. Exponential decrease of the selected neutron signal candidates as a function of the time from the selected event triggers with the Am/Be neutron source positioned near the ID tank center, in the SK-VI phase. The label “Trigger ( $n, \gamma$ )” indicates captures of neutrons that are produced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

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 time scale  $\tau_{\text{thermal}}$ , and the neutron capture time constant  $\tau_{\text{capture}}$  are free parameters. 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 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  $\gamma$  emission model of the  $Gd(n, \gamma)$  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 gamma cascade in the ANNRI-Gd model [41].

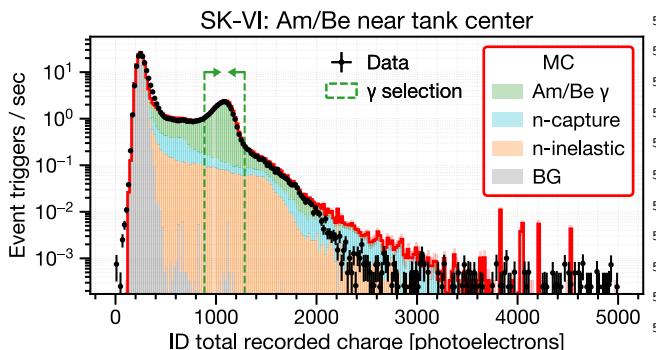


FIG. 8. Distribution of recorded charge within the time window  $[-0.5, 1.0]$   $\mu\text{s}$  relative to event triggers, measured with an Am/Be neutron source positioned near the ID tank center. The black points represent SK-VI data, while the red line shows the simulated prediction. The green dashed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

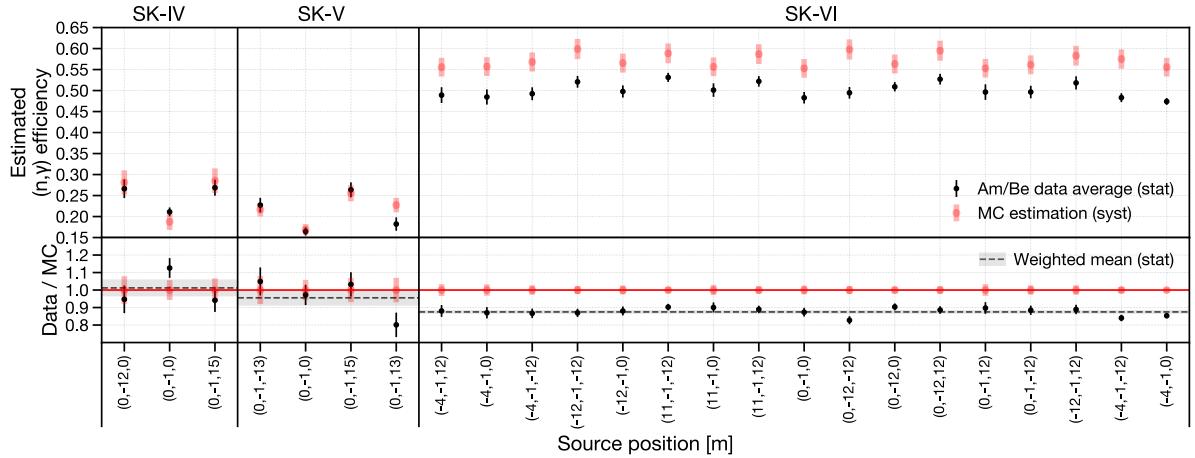


FIG. 10. Estimated  $(n, \gamma)$  signal selection efficiency for each calibration position within the tank. For positions in SK-VI where there are multiple measurements at different dates, the estimated efficiencies were averaged. The systematic uncertainties considered for MC simulation are summarized in Table II.

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( $n, \gamma$ ) fraction	-	-	2.1%
Gd( $n, \gamma$ ) $\gamma$ emission model	-	-	2.6%
Total	7.3%	5.7%	3.8%

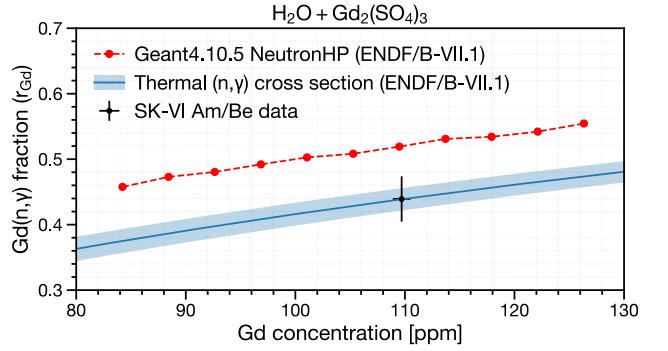


FIG. 11. Predicted Gd( $n, \gamma$ ) fraction in Gd-loaded water as a function of Gd concentration, compared with SK-VI Am/Be data (black). The red dots represent predictions by MC simulation using the NeutronHP model in Geant4.10.05.p01 with neutron cross sections from ENDF/B-VII.1, while the blue line and shades represent evaluation of Equation 4 based on the evaluated thermal  $(n, \gamma)$  cross sections and uncertainties in ENDF/B-VII.1.

The noticeable difference between the observed and predicted signal efficiencies in SK-VI is thought to stem from the overestimation of the Gd( $n, \gamma$ ) fraction in our MC simulation setup. Figure 11 compares the MC-simulated Gd( $n, \gamma$ ) fraction  $r_{\text{Gd}}$  with the analytically evaluated fraction, assuming completely thermalized neutrons:

$$r_{\text{Gd}} \approx 1 - r_{\text{H}} \approx 1 - \frac{n_{\text{H}} g_{\text{H}}(T) \sigma_{\text{H}}(v_{\text{thermal}})}{\sum_i n_i g_i(T) \sigma_i(v_{\text{thermal}})} \quad (4)$$

Here, for the  $i^{\text{th}}$  isotope,  $n_i$  is the number density,  $g_i(T)$  is the Westcott  $g$ -factor for temperature  $T$ , and  $\sigma_i(v_{\text{thermal}})$  is the neutron capture cross section evaluated at thermal neutron speed  $v_{\text{thermal}} = 2200$  m/s.

The value estimated with SK-VI Am/Be data agrees well with the analytical evaluation based on  $g_i$  and  $\sigma_i$  from ENDF/B-VII.1 [51], while both differ from the MC simulation using Geant4.10.05.p01 NeutronHP and ENDF/B-VII.1. The reason seems to be that the NeutronHP model considers the thermal motion of hydrogen as free rather than bound in a water molecule [52], thus underestimating the fraction of  ${}^1\text{H}(n, \gamma)$  that competes with the Gd( $n, \gamma$ ) reaction.

To account for such a difference between signal efficiencies evaluated with Am/Be data and MC simulation, we used the weighted mean of the ratios of the two evaluated at all Am/Be source positions as the correction factor. The obtained efficiency correction factor was  $1.01 \pm 0.04$  for SK-IV,  $0.96 \pm 0.04$  for SK-V, and  $0.88 \pm 0.01$  for SK-VI, including both statistical and systematic errors.

The average neutron capture time constants were measured as  $200.4 \pm 3.7$   $\mu\text{s}$  for the pure water phase (SK-IV, V) and  $116.9 \pm 0.3$   $\mu\text{s}$  for the Gd-loaded phase (SK-VI). These results are consistent with ENDF/B-VII.1 predictions of  $204.7 \pm 5.3$   $\mu\text{s}$  and  $114.9 \pm 2.5$   $\mu\text{s}$ , where the errors are derived from evaluated cross section uncertainties. Relative to this, the Geant4 prediction for the Gd-loaded phase was  $112.4$   $\mu\text{s}$ .

## 604 V. $(n, \gamma)$ MULTIPLICITY ESTIMATION

605 The average multiplicity of  $(n, \gamma)$  reactions, or neutrino captures, is computed as the average of the expected number of  $(n, \gamma)$  reactions estimated on an event-by-event basis, as follows:

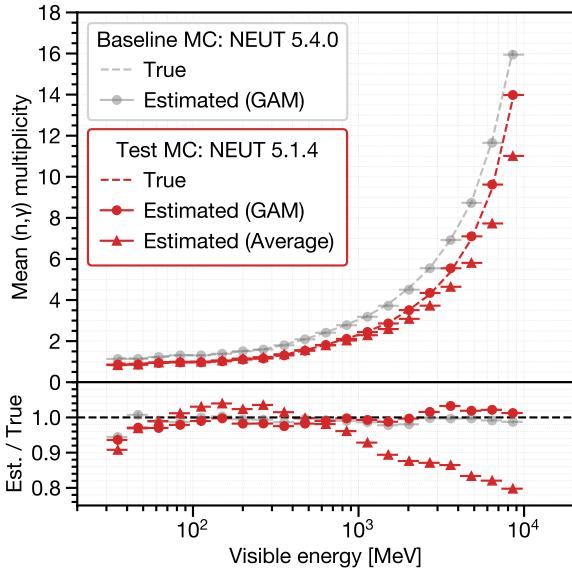
609

$$636 \langle N \rangle = \left\langle \frac{N_i^{\text{detected}} - N_i^{\text{BG}}}{\epsilon_i} \right\rangle \quad 637 \quad (5) \quad 638$$

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

640 Accurate estimation of  $N_i^{\text{BG}}$  and  $\epsilon_i$  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 performance of signal selection. To better account for these effects in the calculation of  $N_i^{\text{BG}}$  and  $\epsilon_i$ , we trained Generalized Additive Models (GAMs) [53] on default atmospheric neutrino event simulations. A total of six GAMs were constructed across the three SK phases, for two output metrics: signal efficiency and false positive rate, using the LinearGAM class in pyGAM 0.9.0 [54].

641 The input features for the analysis included five reconstructed neutrino event variables: visible energy, Cherenkov ring multiplicity, the particle type of the most energetic ring, and the radial and vertical displacements



642 FIG. 12. Comparison of the true (dashed lines) and estimated (circle markers: using GAM, triangle markers: using the overall average signal efficiency and false positive rate) average  $(n, \gamma)$  multiplicity as a function of visible energy, for both baseline and test simulations of atmospheric neutrino events.

643 GAMs were fitted against the distributions of the average simulated output metric per bin in the five-dimensional feature space. No assumptions were made regarding feature correlations, and appropriate smoothing was applied to prevent overfitting. The  $1\sigma$  prediction interval was derived based on the statistical uncertainties within each bin.

644 GAMs were trained on the baseline MC simulation, as described in Section III C, and tested on a simulation generated with NEUT 5.1.4, which predicts roughly 10% lower  $(n, \gamma)$  multiplicity overall. Figure 12 shows the performance of the trained GAM in estimating the true average signal multiplicity per visible energy bin. Compared to using a single overall average metric for each of  $N_i^{\text{BG}}$  and  $\epsilon_i$ , the GAM performs better in reproducing the true distribution, especially for multi-GeV bins.

645 The following major systematic uncertainties affecting signal counting were evaluated on a bin-by-bin basis:

### (1) Overall signal efficiency scale

646 This includes uncertainties in calibrated efficiency correction factors, as well as neutron momentum and detector modeling, which are accounted for by comparing the results across different SK phases as described in Section VII.

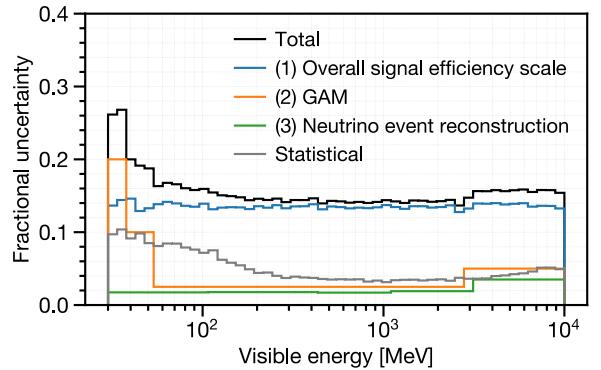
### (2) Signal selection performance modeling (GAM)

647 This refers to potential errors in the regression model used to describe the signal selection performance for atmospheric neutrino events. The uncertainty is quantified as the difference between the true and estimated signal multiplicity in a test simulation setup, as shown in Figure 12.

### (3) Neutrino event reconstruction

648 This assumes a uniform 2% visible energy resolution, as described in Sec. II and Figure 2.

649 Figure 13 illustrates the distribution of fractional uncertainties assigned to each visible energy bin. The most dominant factor is the overall signal efficiency scale.



650 FIG. 13. Fractional uncertainties assigned to the estimated average  $(n, \gamma)$  multiplicity on a bin-by-bin basis.

## 666 VI. TESTED INTERACTION MODELS

667 To compare with the data, we generated predictions  
 668 for the average  $(n, \gamma)$  multiplicity as a function of visible  
 669 neutrino energy using various configurations of neutrino  
 670 event generators and secondary hadron-nucleus inter-  
 671 action models. Six setups of neutrino event generators were  
 672 tested: NEUT 5.4.0, NEUT 5.6.3, and GENIE 3.4.0 with  
 673 G\_18a\_10x\_02\_11b physics tunes [23], where  $x \in a, b, c, d$   
 674 represents FSI models: INTRANUKE/hA, hN [22], the  
 675 Geant4 Bertini cascade model (“BERT”) [55], and the  
 676 Liège INC model (“INCL”) [56].

677 NEUT 5.4.0 was configured as described in Section  
 678 III C, while NEUT 5.6.3 used a modified nuclear bind-  
 679 ing energy, slightly reducing the fraction of CCQE inter-  
 680 actions. GENIE 3.4.0 used interaction models for QE  
 681 and single-pion production similar to NEUT but featured  
 682 distinct FSI and hadronization models. All FSI models,  
 683 except GENIE “hA,” are full INC models based on free  
 684 nucleon cross sections. The Liège INC model was paired  
 685 with ABLA07 [57] for nuclear de-excitation. Figure 14 il-  
 686 lustrates the outgoing neutron momentum distributions  
 687 for these setups, highlighting significant differences be-  
 688 low 1 GeV/c. These variations and their implications  
 689 are discussed in Section VIII.

690 Secondary hadron-nucleus interaction models were  
 691 tested using SK detector simulations based on Geant3.21  
 692 and Geant4.10.05.p01. Six configurations were evaluated:  
 693 “SK-IV/V default,” “SK-VI default,” “G3 GCALOR,”  
 694 “G4 BERT,” “G4 BERT\_PC,” and “G4 INCL\_PC.” The  
 695 SK-IV/V and SK-VI defaults differ mainly in their treat-  
 696 ment of low-energy neutron transport below 20 MeV,  
 697 using MICAP (based on ENDF/B-V) and NeutronHP  
 698 (based on ENDF/B-VII.1), respectively. The “G3  
 699 GCALOR” setup fully relies on Geant3.21 GCALOR,  
 700 differing from the baseline setups that use the NEUT  
 701 pion FSI routine. The Geant4-based configurations (“G4  
 702 BERT,” “G4 BERT\_PC,” and “G4 INCL\_PC”) use the  
 703 NeutronHP model for the low-energy neutron transport.  
 704 For the other hadron transport, “G4 BERT” and “G4  
 705 BERT\_PC” employ the Bertini cascade model, while “G4  
 706 INCL\_PC” uses the Liège INC model. Configurations  
 707 with the “PC” suffix employ the Geant4 Precompound  
 708 model [58] for nuclear de-excitation, whereas “G4 BERT”  
 709 uses a simpler native model [55]. Figure 15 compares  
 710 the average signal multiplicity as a function of projectile  
 711 neutron momentum across these configurations, reveal-  
 712 ing variability in model predictions up to 40% for neutron  
 713 momentum over a few hundred MeV/c. 723

724 For each neutrino event generator, a full MC simula-  
 725 tion equivalent to 50 years of atmospheric neutrino obser-  
 726 vation was generated using the baseline detector model  
 727 (“SK-VI default”). To account for variations in sec-  
 728 onary interactions without additional full simulations,  
 729 we tabulated average signal multiplicities for projectile  
 730 hadrons ( $n, p, \pi^\pm$ , and  $\mu^-$ ) in water up to 10 GeV/c, 730  
 731 based on particle-gun MC simulations. Figure 15 shows  
 732 an example for neutrons. 732

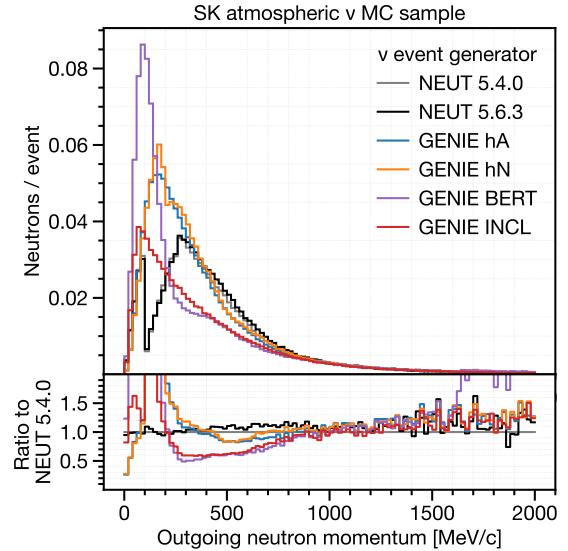


FIG. 14. Model comparison: Outgoing neutron momentum distribution per atmospheric neutrino event in the sample.

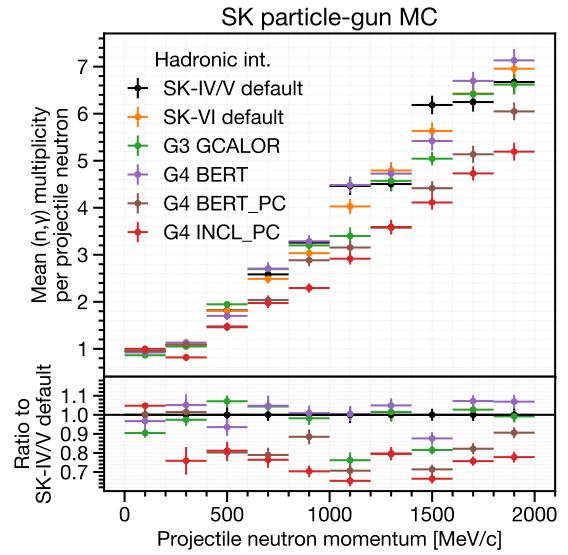


FIG. 15. Model comparison: Average  $(n, \gamma)$  multiplicity per projectile neutron momentum bin, based on simulations using a single-neutron particle-gun setup in water.

Predictions were made by convolving hadron 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. Figure 16 illustrates the sources of variability in these predictions. Differences across neutrino event generator configurations are more pronounced at visible energies below 1 GeV, while variability among secondary hadron-nucleus interaction models remains rela-

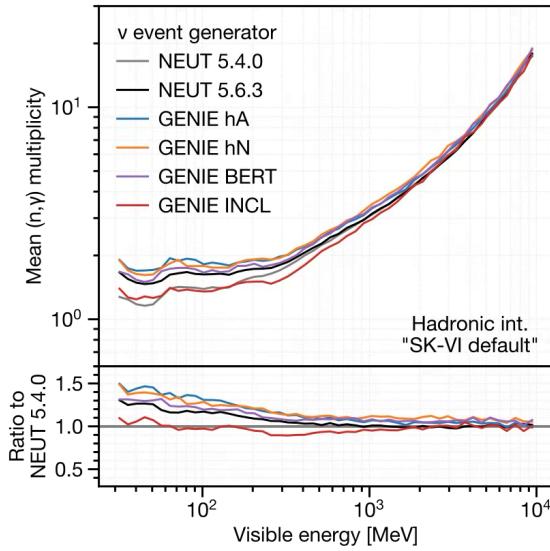
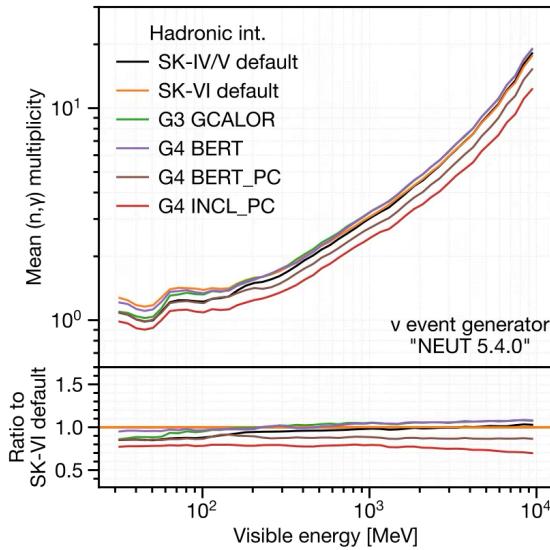


FIG. 16. Comparison of the neutrino event generator options (top) and secondary hadron-nucleus interaction models (bottom) on the predicted average  $(n, \gamma)$  multiplicity. In the top figure, each neutrino event generator option is paired with the baseline secondary interaction model (“SK-VI default”). In the bottom figure, each model is paired with the baseline neutrino event generator (“NEUT 5.4.0”).

tively constant, with a slight increase at higher energies. The Liège INC model predicts fewer neutrons than the other FSI models in NEUT and GENIE, as well as the Bertini cascade model and its variants. NEUT 5.4.0 predicts fewer neutrons than NEUT 5.6.3, due to a larger QE fraction. Additionally, the Geant4 Precompound model predicts fewer neutrons than the nuclear de-excitation model native to the Bertini cascade model.



## VII. RESULTS

Table III summarizes the number of atmospheric neutrino events (“ $\nu$  events”) and detected neutron signals (“ $n$  signals”) in the final data sample. The differences in  $\langle N \rangle_{\text{total}}$  are primarily due to uncertainty in the signal efficiency scale.

TABLE III. Summary of atmospheric neutrino events and detected neutron signals in the final data sample.  $\langle N \rangle_{\text{total}}$  denotes the average  $(n, \gamma)$  multiplicity and its systematic uncertainty. Other errors shown are purely statistical.

	SK-IV	SK-V	SK-VI
$\nu$ events	29,942	4,231	5,203
Events/day	$9.23 \pm 0.05$	$9.18 \pm 0.14$	$9.22 \pm 0.13$
$n$ signals	15,705	2,035	5,752
$n$ signals/event	$0.525 \pm 0.004$	$0.481 \pm 0.011$	$1.106 \pm 0.015$
$\langle N \rangle_{\text{total}}$	$2.04 \pm 0.34$	$2.33 \pm 0.27$	$2.36 \pm 0.21$

Figure 17 compares the estimated average  $(n, \gamma)$  multiplicity per visible energy bin across SK phases. To account for uncertainties in outgoing neutron kinematics that affect signal detection efficiency, an independent algorithm using a likelihood fitter [59] was applied to SK-VI data to reconstruct the  $Gd(n, \gamma)$  vertex, providing a reference for the results in Table III. The scale differences between SK phases and the reference are included as systematic uncertainties in the signal efficiency scale, as shown in Table III and Figure 13.

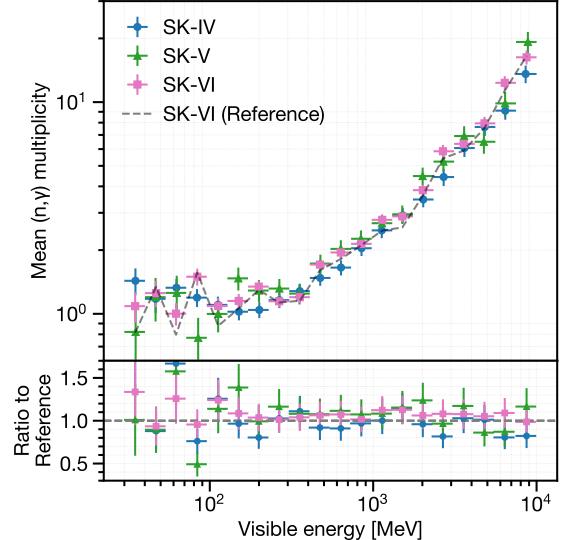


FIG. 17. Comparison of average  $(n, \gamma)$  multiplicity across SK phases, with error bars representing both statistical and systematic uncertainties. The dashed line labeled “SK-VI (Reference)” shows the result from SK-VI data using a reference algorithm whose performance is largely independent of neutron kinetic energy.

The combined data was compared with the SK-VI MC simulation and neutron production estimates in water from SNO [12], as shown in Figure 18. While the combined data matched well with the SNO estimate, it was 20-40% lower than the SK-VI simulation results across different energy ranges. For neutrino events with visible energy greater than a few hundred MeV, where the Cherenkov rings of outgoing leptons are accurately reconstructed, a linear relationship was observed between the visible energy and the average  $(n, \gamma)$  multiplicity.

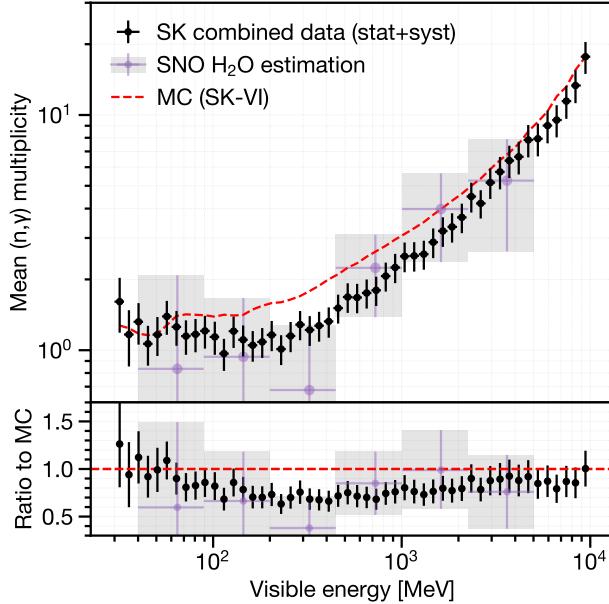


FIG. 18. Average  $(n, \gamma)$  multiplicity estimated for atmospheric neutrino interactions versus visible energy. The black dots and error bars represent combined SK data, including statistical and systematic uncertainties. Purple crosses with shading indicate estimates of neutron production in water derived from SNO measurements using a D<sub>2</sub>O target volume [12]. The red dashed line represents the true average  $(n, \gamma)$  multiplicity per visible energy bin obtained from SK-IV MC simulation.

Figure 19 shows the sources of neutron production in the baseline SK-IV simulation as a function of visible energy, offering insight into the origins of neutron production across different neutrino energy regions. The number of outgoing neutrons from neutrino-nucleus interactions is expected to be nearly constant regardless of visible energy. In events with lower visible energies, the momentum of the outgoing neutrons is small, so each outgoing neutron corresponds almost directly to a  $(n, \gamma)$  reaction. However, at higher energies, the outgoing hadrons gain more energy, leading to an increase in secondary neutron production. This suggests that the  $(n, \gamma)$  multiplicity in lower visible energies is sensitive to the modeling of low-energy nucleon transport within the nucleus, while the linear increase is primarily determined by the choice of secondary hadron-nucleus interaction model.

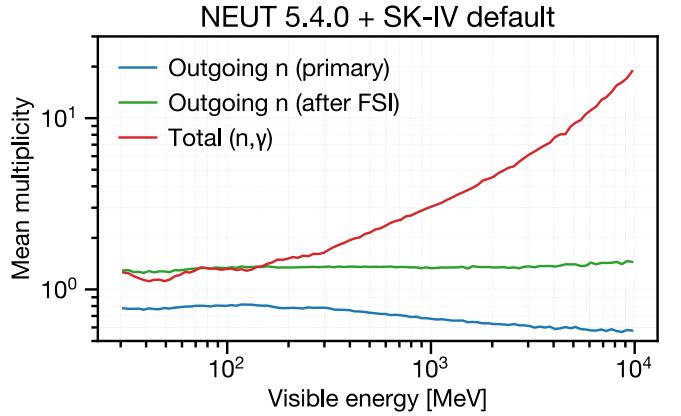


FIG. 19. Mean multiplicity of outgoing neutrons from the primary neutrino interaction (blue) and subsequent FSIs (green) and resulting total  $(n, \gamma)$  reactions (red), in the baseline SK-IV simulation setup.

### A. Effective metrics for model evaluation

The “slope” of the linear increase in the average  $(n, \gamma)$  multiplicity as a function of visible energy was determined through a linear fit over the energy range [0.3, 10] GeV, as shown in Figure 20. Only the vertical statistical uncertainties of the data points were considered in the fit, since the dominant systematic uncertainty arises from the signal efficiency scale. This scale is assumed to be fully correlated across bins, and does not affect the fitted slope.

At lower visible energies, the effects of hadronization in DIS events become negligible, and sensitivity to low-energy nucleon FSI modeling becomes significant. To quantify neutron production in low energy neutrino events, the “low-energy (low-E) multiplicity” was defined as the average  $(n, \gamma)$  multiplicity observed in [0.1, 0.3] GeV visible energy range. Events below 0.1 GeV were excluded due to large systematic uncertainties.

Figure 21 compares the measured slopes and low-energy multiplicities across different SK phases and event types. Both metrics were higher for multi-ring events compared to single-ring events, consistent with MC simulations. This difference is due to multi-ring events having a higher fraction of DIS events, as shown in Figure 5. While only statistical errors were considered, some tension was observed in the slopes between the pure water phases (SK-IV and SK-V) and the Gd-loaded phase (SK-VI). The slopes in SK-VI were slightly higher and closer to MC predictions, while the MC predictions showed consistency between pure water and Gd-loaded phases. Both the slopes and low-E multiplicities from the combined data were lower than the baseline predictions.

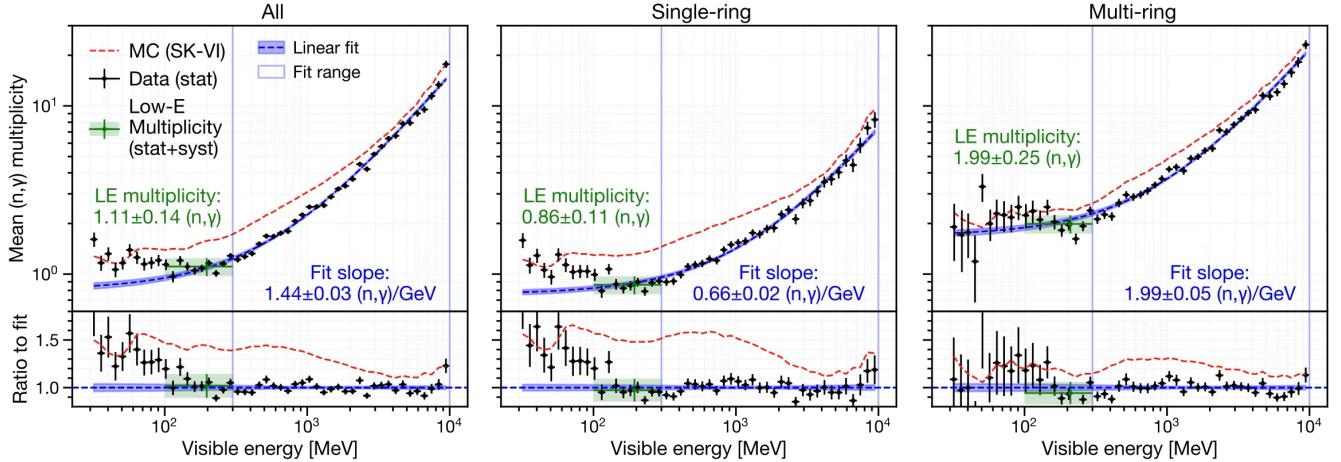


FIG. 20. The average  $(n, \gamma)$  multiplicity as a function of the visible energy of neutrino events observed in the combined data, shown for all events (left), single-ring events (middle), and multi-ring events (right). The red dashed lines denote the predictions from the full MC simulation in the SK-VI baseline setup. The blue dashed lines represent the fitted linear functions, with the blue-shaded regions indicating the  $1\sigma$  prediction intervals. The fit was performed over the energy range [0.3, 10] GeV, with data points weighted by the inverse of their statistical errors. The green data points with error bars represent the low-energy multiplicity, defined as the average  $(n, \gamma)$  multiplicity in [0.1, 0.3] GeV range.

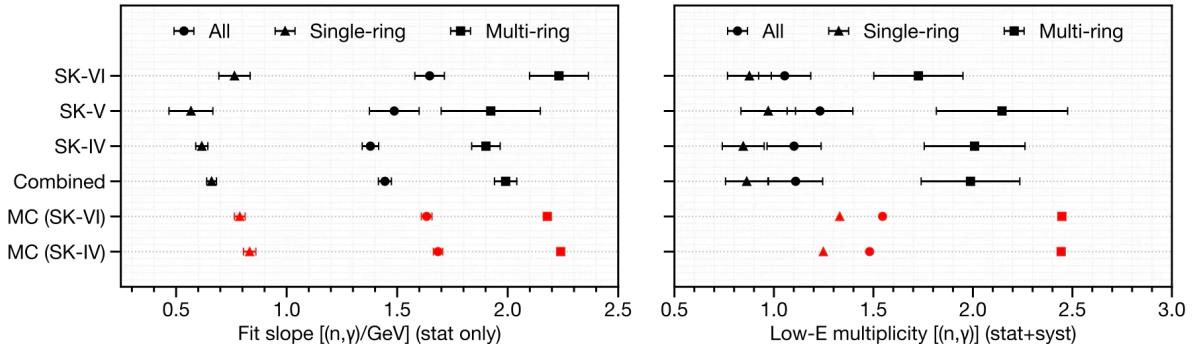


FIG. 21. The measured slopes (left) and low-energy multiplicities (right) across different SK phases and event types. The low-energy multiplicities are defined as the average  $(n, \gamma)$  multiplicity observed within the visible energy range of [0.1, 0.3] GeV. In the left panel, the error bars represent statistical uncertainties only, whereas in the right panel, the error bars account for the full systematic uncertainty, including the signal efficiency scale error. The MC predictions (red) in the right panel correspond to the “true” average  $(n, \gamma)$  multiplicity, rather than the efficiency-corrected values.

## B. Comparison with model predictions

815 Predictions from various model combinations (ex-  
816 plained in Section VI) were compared with the combined  
817 data, as shown in Figure 22. A notable feature in the  
818 data is a valley in the [0.1, 0.3] GeV region, which is  
819 less pronounced in the baseline simulation. The largest  
820 discrepancy between the data and predictions occurs in  
821 this region. This valley is only reproduced when us-  
822 ing models that predict fewer neutrons than the base-  
823 line for both FSI and secondary interaction models, such  
824 as “GENIE INCL” and “G4 BERT\_PC.” The bottom  
825 panel of Figure 22 highlights this, by comparing the  
826 data with predictions from specific FSI models (“GENIE  
827 hN,” which predicts more neutrons, and “GENIE INCL,”  
828 which predicts fewer neutrons) and secondary interac-  
829 tion models (“G4 BERT,” which predicts more neutrons,  
830 and “G4 BERT\_PC” and “G4 INCL\_PC,” which predict  
831 fewer neutrons). Among all tested combinations, “GE-  
832 NIKE INCL” and “G4 BERT\_PC” provide the best agree-  
833 ment with data, showing the lowest chi-squared statistic  
834 when considering statistical errors only.

835 Figure 23 compares the measured slopes and low-  
836 energy multiplicities with model predictions. The pre-  
837 dictions show three distinct groups of slopes: the  
838 “G4 INCL\_PC” model gives the smallest slopes, “G4  
839 BERT\_PC” predicts intermediate slopes, and other vari-  
840 ants of the Bertini cascade model produce the largest

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

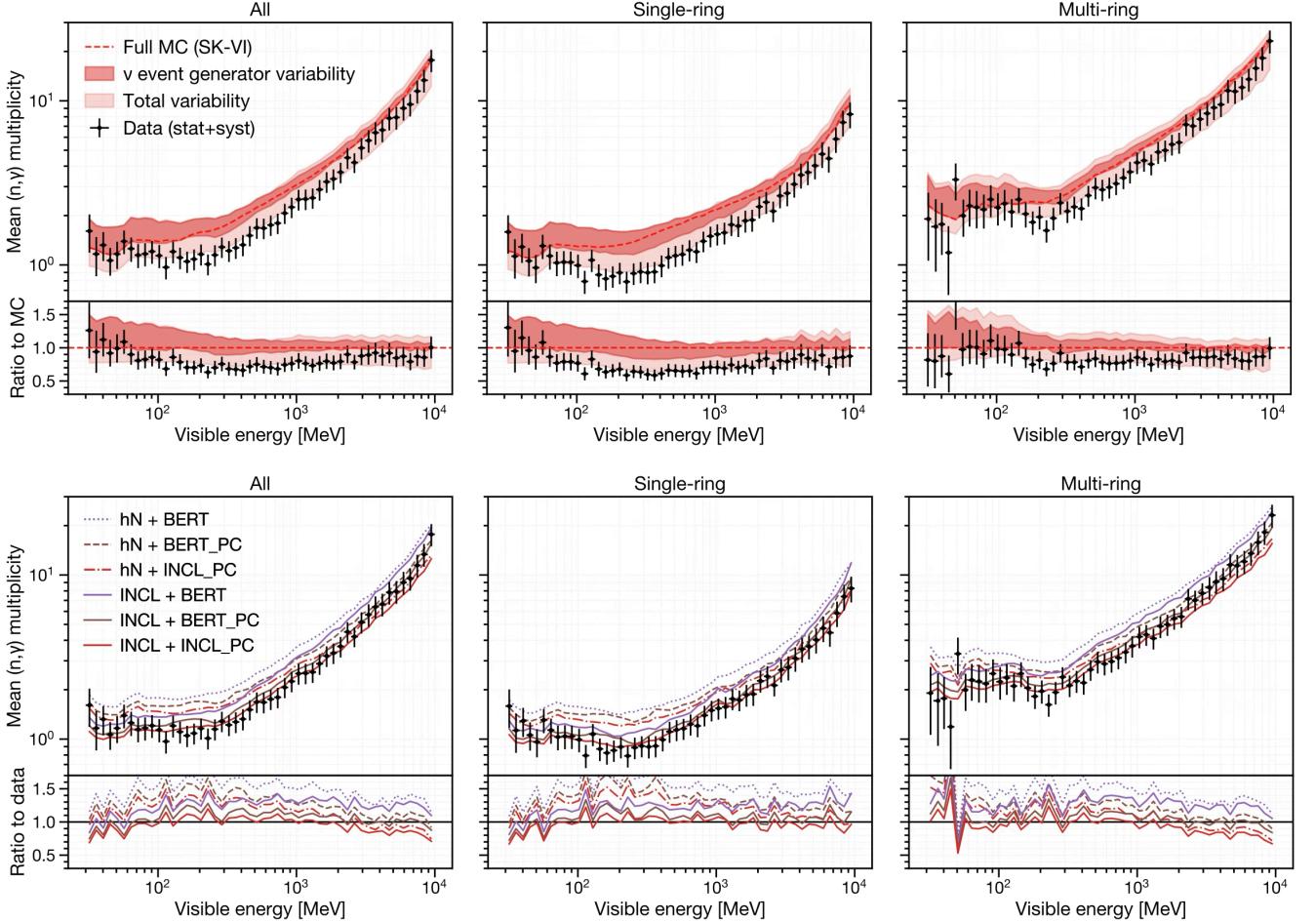


FIG. 22. Comparison of various model predictions with the combined data, showing the average  $(n, \gamma)$  multiplicity as a function of atmospheric neutrino event visible energy for all events (left), single-ring events (middle), and multi-ring events (right). The black dots with error bars represent 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 (“INCL + INCL\_PC”) that shows the overall best agreement with the combined data.

slopes. The slope observed in the single-ring data better<sup>855</sup>  
 matches the “G4 INCL\_PC” predictions, while the slope  
 in the multi-ring data is closer to the “G4 BERT\_PC”  
 predictions. This distinction is most evident in the bot-  
 tom panel of Figure 22, where “G4 INCL\_PC” matches  
 well with sub-GeV single-ring data, and “G4 BERT\_PC”  
 is a better fit for multi-GeV multi-ring data.  
 842 slopes. The slope observed in the single-ring data better<sup>855</sup>  
 843 matches the “G4 INCL\_PC” predictions, while the slope  
 844 in the multi-ring data is closer to the “G4 BERT\_PC”  
 845 predictions. This distinction is most evident in the bot-  
 846 tom panel of Figure 22, where “G4 INCL\_PC” matches  
 847 well with sub-GeV single-ring data, and “G4 BERT\_PC”  
 848 is a better fit for multi-GeV multi-ring data.

For low-energy multiplicities, the two versions of NEUT and ”GENIE INCL” are preferred, as they pre-<sup>860</sup>  
 850 dict lower neutron production. In contrast, models<sup>861</sup>  
 851 such as GENIE’s “hA,” “hN,” and “BERT” overestimate<sup>862</sup>  
 852 low-energy multiplicities, producing values significantly<sup>863</sup>  
 853 higher than the uncertainty in signal efficiency.<sup>864</sup>

## VIII. DISCUSSION

Overall, our data favor models that predict relatively lower neutron production. In this section, we examine key features of these models that affect neutron production, with reference to Figure 14 that highlights differences in nucleon momentum distributions across various neutrino event generators. By qualitatively comparing the variations among the INC models and nuclear de-excitation models used in these generator options, we identify the sources of the observed differences in their predictions.<sup>865</sup>

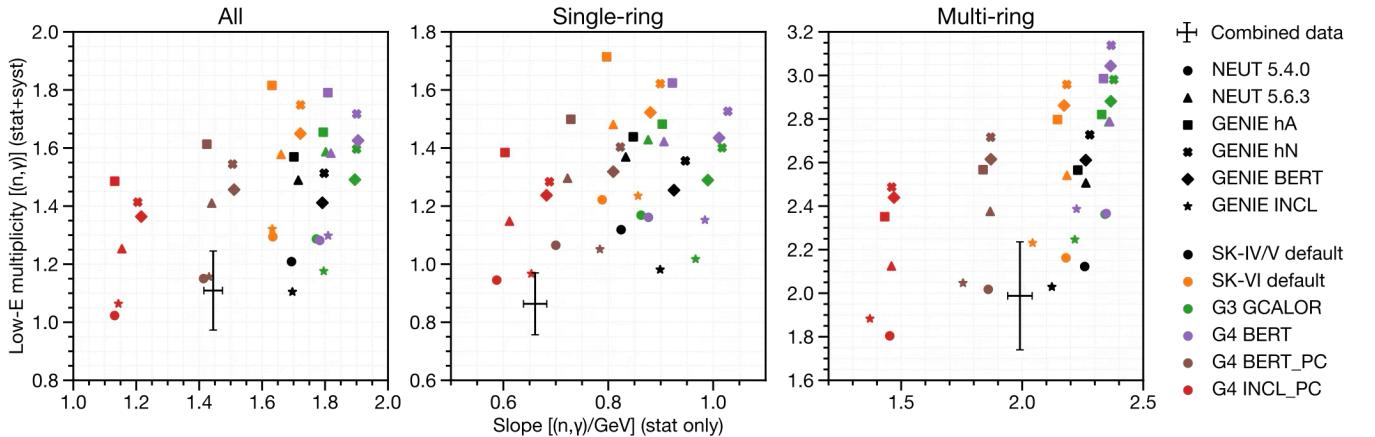


FIG. 23. Scatter plots of fitted slopes and low-energy multiplicities in model predictions and data (black crosses, error bars include systematic uncertainty in low-energy multiplicity and statistical uncertainty in slope), shown for all events (left), single-ring events (middle), and multi-ring events (right). Shapes represent neutrino event generator options used, while colors indicate the secondary hadron-nucleus interaction model used.

866

### A. Nuclear de-excitation

Nuclear de-excitation models play a critical role in predicting the total  $(n, \gamma)$  multiplicity per secondary interaction, as they govern the production of very-low-energy neutrons, which directly contribute to  $(n, \gamma)$  reactions. One notable feature in Figure 14 is the significant peak in the “GENIE BERT” prediction for neutron momenta below 250 MeV/c. This peak arises primarily from the nuclear de-excitation model in the Geant4 Bertini cascade [55], which is based on the Weisskopf statistical evaporation model [60], as implemented by [61]. This model relies on crude parameterizations of nucleon absorption (inverse reaction) cross sections, developed before the availability of extensive experimental data. In contrast, the more recent Geant4 Precompound model [58] provides smaller, data-driven estimates for neutron emission probabilities.<sup>1</sup> Similarly, the ABLA07 model, used in the “GENIE INCL” option, predicts neutron emission rates comparable to those of the Geant4 Precompound model and the NEUT oxygen de-excitation model. Conversely, the “GENIE hA” and “hN” options omit nuclear de-excitation altogether.

888

### B. Pauli blocking

Another key difference among the models lies in their treatment of “Pauli blocking,” a quantum mechanical effect that prohibits nucleons from occupying the same quantum state, as dictated by the Pauli exclusion principle. For instance, NEUT and the “GENIE BERT”

894

options implement strict Pauli blocking, which forbids all nucleon scattering below the oxygen Fermi momentum of 225 MeV/c, modeling the nucleus as a degenerate Fermi gas. This strict treatment manifests in NEUT as a characteristic dip in the corresponding energy range, as illustrated in Figure 14. By contrast, the Liège INC model applies a probabilistic approach to Pauli blocking during the cascade, accounting for nucleon holes created by prior cascade steps. This results in smoother neutron distributions near the Fermi momentum. In comparison, the “GENIE hA” and “hN” models omit Pauli blocking entirely, leading to significantly higher predicted neutron production than either NEUT or “GENIE INCL.”

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

915

916

917

918

919

920

921

922

### C. Other considerations

The Liège INC model and the Geant4 Bertini cascade model, which apply medium corrections such as nucleon-nucleon repulsion and interaction correlations across cascade steps, predict lower neutron production at intermediate energies (250–1000 MeV/c) compared to the other options shown in Figure 14. This energy difference among outgoing hadrons directly affects the number of secondary interactions. In contrast, variations in pion-nucleus interaction models (e.g., NEUT’s low-energy pion FSI model in “SK-IV/V default” and the Bertini cascade model in “G3 GCALOR”) or low-energy neutron reaction cross-section datasets (e.g., ENDF/B-V cross sections in the S“K-IV/V default” model and ENDF/B-VII.1 in the “SK-VI default” model) appear to have relatively minor effects on neutron production predictions.

<sup>1</sup> The predictions of this model are not shown in Figure 14, as it is only applied in the secondary hadron-nucleus interaction options.

923

## IX. SUMMARY

978

## ACKNOWLEDGMENTS

924 Accurately modeling neutron production in neutrino 979  
 925 interactions is essential for characterizing incoming neu- 980  
 926 trinos, which is crucial for advancing precision measure- 981  
 927 ments of neutrino oscillation parameters and rare event 982  
 928 searches involving neutron tagging. Recent studies have 983  
 929 suggested potential inaccuracies in the modeling of sec- 984  
 930 ondary hadron-nucleus interactions. 985

931 This paper reports a measurement of total neutron 987  
 932 production following atmospheric neutrino interactions 988  
 933 within the water volume of the Super-Kamiokande (SK) 989  
 934 detector. In SK, neutrons with kinetic energies below a 990  
 935 few MeV are typically captured by surrounding nuclei, 991  
 936 such as  $^1\text{H}$  or  $^{155}/^{157}\text{Gd}$ , releasing 2–8 MeV of energy 992  
 937 through gamma-ray emissions. These radiative neutron 993  
 938 captures ( $(n, \gamma)$  reactions) were identified using a detec- 994  
 939 tion algorithm that combines a simple low-energy trigger 995  
 940 with a neural network binary classifier, calibrated with 996  
 941 an Am/Be neutron source. Event-by-event neutron cap- 997  
 942 ture multiplicity was estimated using a multivariate non- 998  
 943 linear regression technique. Atmospheric neutrino events 999  
 944 were binned by their electron-equivalent “visible energy,”<sup>1000</sup>  
 945 a semi-calorimetric proxy for neutrino momentum trans-<sup>1001</sup>  
 946 fer, in the range of [0.03, 10] GeV. The average neutron<sup>1002</sup>  
 947 capture multiplicity in each visible energy bin was com-<sup>1003</sup>  
 948 compared against predictions from various combinations<sup>1004</sup>  
 949 of neutrino event generators and secondary hadron-nucleus<sup>1005</sup>  
 950 interaction models.<sup>1006</sup>

1007

951 The extensive dataset provides strong discriminative<sup>1008</sup>  
 952 power for evaluating hadron-nucleus interaction models.<sup>1009</sup>  
 953 Variability in model predictions across the energy range<sup>1010</sup>  
 954 was found to be as large as 50%. Notably, the variabil-<sup>1011</sup>  
 955 ity in FSI model predictions was more pronounced in the<sup>1012</sup>  
 956 sub-GeV range, while variability from secondary hadron-  
 957 nucleus interaction models dominated in the multi-GeV  
 958 range. The observed data, which showed fewer neutron  
 959 signals than predicted by the baseline simulation, could  
 960 only be explained by using FSI and secondary interaction  
 961 models that predict relatively lower neutron produc-  
 962 tion. For example, the Liège INC model for intranuclear  
 963 hadron transport and the Geant4 Precompound model  
 964 for nuclear de-excitation produced lower neutron predic-  
 965 tions that aligned more closely with the observations. In  
 966 contrast, the widely used Geant4 Bertini cascade model,  
 967 when coupled with its native nuclear de-excitation model,  
 968 significantly overestimated neutron production beyond  
 969 the estimated uncertainties. Qualitative comparisons  
 970 suggest that the major sources of variation in neutron  
 971 production predictions stem from the modeling of sta-  
 972 tistical neutron emission in nuclear de-excitation, Pauli  
 973 blocking, and the inclusion of nucleon-nucleon corre-  
 974 lations beyond the semi-classical treatment in INC models.  
 975 This study highlights the critical role of hadron-nucleus  
 976 interaction models in predicting total neutron production  
 977 from neutrino interactions.

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 (NRF-2009-0083526, NRF-2022R1A5A1030700, 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 (2018R1D1A1B07049158, 2021R1I1A1A01042256, 2021R1I1A1A01059559, RS-2024-00442775); the Japan Society for the Promotion of Science; the National Natural Science Foundation of China under Grants No.12375100; the Spanish Ministry of Science, Universities and Innovation (grant 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 (UMO-2018/30/E/ST2/00441 and 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 CA3/RSUE2021-00559; the National Institute for Nuclear Physics (INFN), Italy.

- [1] P. Abratenko *et al.* (MicroBooNE collaboration), Phys. Rev. D **110**, 013006 (2024).
- [2] P. Abratenko *et al.* (MicroBooNE collaboration), The European Physical Journal C **84**, 1052 (2024).
- [3] A. M. Ankowski, P. Coloma, P. Huber, C. Mariani, and E. Vagnoni, Phys. Rev. D **92**, 091301 (2015).
- [4] T. Wester *et al.* (Super-Kamiokande collaboration), Phys. Rev. D **109**, 072014 (2024).
- [5] A. Takenaka *et al.* (Super-Kamiokande collaboration), Phys. Rev. D **102**, 112011 (2020).
- [6] S. Agostinelli *et al.* (Geant4 collaboration), Nucl. Instrum. Methods A **506**, 250 (2003).
- [7] J. Allison *et al.*, IEEE Transactions on Nuclear Sciences **53**, 270 (2006).
- [8] J. Allison *et al.*, Nucl. Instrum. Methods A **835**, 186 (2016).
- [9] J. Cugnon, Nucl. Phys. A **387**, 191 (1982).
- [10] J.-C. David, D. Filges, F. Gallmeier, M. Khandaker, A. Konobeyev, S. Leray, G. Mank, A. Mengoni, R. Michel, N. Otsuka, and Y. Yariv, Progress In Nuclear Science and Technology **2**, 942 (2011).
- [11] R. Akutsu, A study of neutrons associated with neutrino and antineutrino interactions on the water target at the T2K far detector, Ph.D. thesis, The Univ. of Tokyo (2019).
- [12] B. Aharmim *et al.* (SNO collaboration), Phys. Rev. D **99**, 112007 (2019).
- [13] M. Elkins *et al.* (MINERvA collaboration), Phys. Rev. D **100**, 052002 (2019).
- [14] A. Olivier *et al.* (MINERvA collaboration), Phys. Rev. D **108**, 112010 (2023).
- [15] M. B. Chadwick *et al.*, Nuclear Data Sheets **112**, 2887 (2011), special Issue on ENDF/B-VII.1 Library.
- [16] K. Abe *et al.* (Super-Kamiokande collaboration), Nucl. Instrum. Methods A **1027**, 166248 (2022).
- [17] K. Abe *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1065**, 169480 (2024).
- [18] S. Fukuda *et al.* (Super-Kamiokande collaboration), Nucl. Instrum. Methods A **501**, 418 (2003).
- [19] K. Abe *et al.*, Nucl. Instrum. Methods A **737**, 253–272 (2014).
- [20] S. Yamada, K. Awai, Y. Hayato, K. Kaneyuki, Y. Kouzuma, S. Nakayama, H. Nishino, K. Okumura, Y. Obayashi, Y. Shimizu, *et al.*, IEEE Trans. Nucl. Sci. **57**, 428 (2010).
- [21] C. Andreopoulos *et al.*, Nucl. Instrum. Methods A **614**, 87 (2010), arXiv:0905.2517 [hep-ph].
- [22] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher, G. Perdue, and J. Yarba, The GENIE Neutrino Monte Carlo Generator: Physics and User Manual (2015), arXiv:1510.05494 [hep-ph].
- [23] J. Tena-Vidal *et al.* (GENIE), Phys. Rev. D **104**, 072009 (2021), arXiv:2104.09179 [hep-ph].
- [24] M. Shiozawa, Nucl. Instrum. Methods A **433**, 240 (1999).
- [25] M. Honda, T. Kajita, K. Kasahara, and S. Midorikawa, Phys. Rev. D **83**, 123001 (2011).
- [26] R. L. Workman *et al.* (Particle Data Group), PTEP **083C01**, 729 (2022).
- [27] Y. Hayato and L. Pickering, Eur. Phys. J Spec. Top. **230**, 4469–4481 (2021).
- [28] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio, Nucl. Phys. A **484**, 557 (1988).
- [29] E. S. P. Guerra *et al.*, Phys. Rev. D **99**, 052007 (2019).
- [30] H. W. Bertini, Phys. Rev. **131**, 1801 (1963).
- [31] P. de Perio, AIP Conf. Proc. **1405**, 223 (2011), arXiv:1405.3973 [nucl-ex].
- [32] A. M. Ankowski, O. Benhar, T. Mori, R. Yamaguchi, and M. Sakuda, Phys. Rev. Lett. **108**, 052505 (2012).
- [33] H. Ejiri, Phys. Rev. C **48**, 1442 (1993).
- [34] C. Zeitnitz and T. A. Gabriel, Nucl. Instrum. Meth. A **349**, 106 (1994).
- [35] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, GEANT3, Tech. Rep. (1987).
- [36] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, FLUKA: A multi-particle transport code (Program version 2005), Tech. Rep. (2005).
- [37] H. W. Bertini, Phys. Rev. **188**, 1711 (1969).
- [38] T. A. Gabriel, J. O. Johnson, and J. Brau, MICAP: A program for low energy neutron, ion and gamma-ray transport and one of its applications in calorimeter design, Tech. Rep. (United States, 1987) cONF-8708186-7.
- [39] C. S. E. W. Group, ENDF/B summary documentation, Tech. Rep. BNL-NCS-1754 (ENDF-201) (National Nuclear Data Center, Brookhaven National Laboratory, Upton, NY, USA, 1979).
- [40] E. Mendoza and D. Cano-Ott, Update of the Evaluated Neutron Cross Section Libraries for the Geant4 Code (see also INDC(NDS)-0612), Tech. Rep. (International Atomic Energy Agency (IAEA), 2018) iINDC(NDS)-0758.
- [41] K. Hagiwara *et al.*, Prog. Theor. Exp. Phys. **2019**, 023D01 (2019), <https://academic.oup.com/ptep/article-pdf/2019/2/023D01/27970473/ptz002.pdf>.
- [42] T. Tanaka *et al.*, Prog. Theor. Exp. Phys. **2020**, 043D02 (2020), <https://academic.oup.com/ptep/article-pdf/2020/4/043D02/33040537/ptaa015.pdf>.
- [43] V. Barger, K. Whisnant, S. Pakvasa, and R. J. N. Phillips, Phys. Rev. D **22**, 2718 (1980).
- [44] M. Jiang *et al.* (Super-Kamiokande collaboration), Progress of Theoretical and Experimental Physics **2019**, 053F01 (2019), <https://academic.oup.com/ptep/article-pdf/2019/5/053F01/28638877/ptz015.pdf>.
- [45] A. M. Dziewonski and D. L. Anderson, Physics of the Earth and Planetary Interiors **25**, 297 (1981).
- [46] K. Abe *et al.* (Super-Kamiokande collaboration), J. Instrum. **17** (10), P10029.
- [47] F. Chollet *et al.*, Keras, <https://keras.io> (2015).
- [48] K. He, X. Zhang, S. Ren, and J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification (2015), arXiv:1502.01852 [cs.CV].
- [49] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2017), arXiv:1412.6980 [cs.LG].
- [50] H. Ito, K. Wada, T. Yano, Y. Hino, Y. Ommura, M. Harada, A. Minamino, and M. Ishitsuka, Nucl. Instrum. Methods A **1057**, 168701 (2023).
- [51] B. Pritychenko and S. F. Mughabghab, Nuclear Data Sheets **113**, 3120 (2012).
- [52] Y. Hino, K. Abe, R. Asaka, S. Han, M. Harada, M. Ishitsuka, H. Ito, S. Izumiya, Y. Kanemura, Y. Koshio, F. Nakanishi, H. Sekiya, and T. Yano, Modification on

- 1135 thermal motion in geant4 for neutron capture simulation in gadolinium loaded water (2024), arXiv:2412.04186 [hep-ex].  
 1136  
 1137  
 1138 [53] T. Hastie and R. Tibshirani, *Generalized Additive Models*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Taylor & Francis, 1990).  
 1139  
 1140 [54] D. Servén, C. Brummitt, and H. Abedi, dswh/pygam v0.8.0 (2018).  
 1141  
 1142 [55] D. Wright and M. Kelsey, Nucl. Instrum. Methods A **804**, 175 (2015).  
 1143  
 1144 [56] S. Leray, D. Mancusi, P. Kaitaniemi, J. C. David, A. Boudard, B. Braunn, and J. Cugnon, Journal of Physics: Conference Series **420**, 012065 (2013).  
 1145  
 1146 [57] A. Kelic, M. V. Ricciardi, and K.-H. Schmidt, Abla07 - towards a complete description of the decay channels of a nuclear system from spontaneous fission to multifragmentation (2009), arXiv:0906.4193 [nucl-th].  
 1147  
 1148 [58] J. Quesada Molina *et al.*, Progress in Nuclear Science and Technology **2**, 936 (2011).  
 1149  
 1150 [59] M. Smy, in *International Cosmic Ray Conference*, International Cosmic Ray Conference, Vol. 5 (2008) pp. 1279–1282.  
 1151  
 1152 [60] V. Weisskopf, Phys. Rev. **52**, 295 (1937).  
 1153  
 1154 [61] I. Dostrovsky, Z. Fraenkel, and G. Friedlander, Phys. Rev. **116**, 683 (1959).  
 1155  
 1156  
 1157  
 1158  
 1159