SK comments for draft v1.4

M. Miura

Please add the following authors to the author list.
M. Fan\`{i} Minnesota
D. Samudio UAM
R. Shinoda Kamioka
A. Kawabata Keio
T. Sone Kobe
S. Wada Kobe
K. Hamaguchi Okayama
R. Nishijima RCCN
S. Kodama UT
M. Sugo TUS
M. Wako TUS
Y. Sasaki YNU
D. Hamaguchi YNU
J. Hikida Kyoto
Added all.

Y. Ashida

[General]
- PRD style suggests "$gamma (= γ)" not "gamma".
Done.

[Abstract]
- L108: You tried not only the hadron-nucleus SI but also the variation of neutrino FSI as well, right?
Yes, here "hadron-nucleus interaction model" refers to "intranuclear cascade model + de-excitation model" used in both SI and FSI.

[Section 1]
- L152 which "typically": Is it not always, as IBD on free protons leaves only one neutron? Or what you mean here is oxygen as a target?
Removed "typically." Originally I was thinking of additional neutrons produced via secondary interactions in water. But I guess in this sentence things get clearer by simply removing "typically."

- You could mention the NCQE neutron study at SK (Sakai et al.) somewhere in the intro part, and better to define your signal clearly different from NCQE.
Added: *This study focuses on an event sample with visible energy greater than 30 MeV, distinct from our previous neutron measurement \cite{sakai} targeting neutral-current quasi-elastic (NCQE) events with visible energy below 30 MeV.*

- Figure 1: Why do you have black arrows arising from "holes" causing the later interactions? Not nucleons themselves?
Revised: 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*.
Nucleons that are sitting in the nucleus are not drawn for simplicity. I have clarified that these holes are resulting from the ejected nucleons.

[Section 2]
- How large is the oxygen capture fraction?
O(0.01)% level. Already suggested by >99.9% H capture fraction shown in Table I.

- L268: It is better to mention live times of each operational period here.
It's shown in Table I so I'll just leave as is for now.

- Would you want to explain why there is a phase separate between VII and VIII, i.e., coil off trouble?
Not in this paper. I'd prefer to leave it to the other people.

[Section 3]
- A "Data reduction": Is this ATMPD specific? I would prefer a better naming, e.g., pre-selection.
Changed to "Event selection"

- L288: Please clarify the hit distribution in Figure 3 is made by simulation.
Added "simulated"

- Figure 3: Can you add an event display for noise-only? It would help readers to understand how different signal looks from noise.

For now, leave as is. I understand it can help readers catch the signal-noise difference. I hope readers can imagine randomly positioned hit PMTs from the caption.

- L309: The flux input is down to 100 MeV while your visible energy is down to 30 MeV. Would it be better to mention that the flux below 100 MeV contributes only few to the selected events?

Added: *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.

Also added in Discussion, Sec VIII D: 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.

- As far as I read, you are missing description about the abbreviations of CC and NC.

Actually there are; in L141 (CC) and L246 (NC)

- L347: Do you have a proper reference?

For now, leave as is. There is one T2K technote describing neut pion FSI usage in skdetsim, but I don't think there is a public literature describing it.

- L371 ", demonstrating reasonable agreement with the observations": I am not sure if you can state this as I see mismatch between data and MC.

Removed the description.

- You may want to explain what kind of NC events dominate the multi-ring sample?

Added: "NC events in the multi-ring sample are mostly RES or DIS interactions" in Figure 5 caption.

- Figure 5: Why do you show the result from SK-IV here? The other parts always show SK-VI. I guess this is because of the statistics... but if so, it is better to mention this.

Added in caption: Observed and simulated visible energy distributions for selected atmospheric neutrino interactions, *for the longest SK-IV phase*

- Figure 5 "fully contained": This was not defined in the main text?

Described in L269-270.

[Section 4]

- Table 1: "Dates" --> "Years"

Done.

- Figure 9: I can't see the pink shade... maybe you can do the log scale in y?

For now, leave as is. I understand it is very small, and I'm afraid plotting it in log scale wouldn't make much difference.

- L541: no indent

Done.

- Figure 10: You may need a tank schematic so that readers can easily understand where?

Added tank schematic and revised Figure 10 caption: Estimated (n,γ) signal selection efficiencies for each calibration position within the tank *(red markers in the schematic on the right).*

- L579: Ref. [52] is now published on PTEP.

Done.

- L587: Do you need this scaling for SK-IV and V? If so, why do I not see such gap between data and MC for SK-IV and V in Figure 10?
Moved “see appendix A for details” to the end of the paragraph.
You’re right, Am/Be estimated efficiency comparison between data and MC for SK-IV and V look fine and this Am/Be calibration has little practical impact on this study as we’re mostly calibrating our efficiency using SK6 Am/Be data as described in Appendix A.
So if we apply no scaling for SK-IV, which seems ok by Am/Be calib results shown in Figure 10, then it results in 10% smaller neutron multiplicity compared to SK-VI. (SK-V agrees with SK-VI without scaling) We don’t know why, and the only possible/reasonable explanation at this stage is to blame the lack of calibration data and that there may be unaccounted source of uncertainty in SK-IV. Indeed, SK-IV with 10-year span has only two calibration runs, one in 2008 and the other in 2016, at only three source positions. Also, data structures (e.g., triggers, AFT length) seem to be somewhat different between the two, with little documentation. Of course with SK-VI we have lots of periodic calibration data once in every 2 weeks or a month or so, and the data span the whole tank volume.
So we apply scaling based on relaive difference in the estimated neutron multiplicity in atmospheric neutrino data, instead of Am/Be. For SK-VI (for both nominal and reference algorithms) and SK-V, this scaling is consistent with Am/Be.

[Section 5]
- Just from my curiosity, can you please explain a bit more what you are doing with GAM? I was not able to understand well this part...
So GAM is a regression model, whose internal parameters can be trained/fitted to estimate neutron tagging efficiency and false positive rate, given some high-level reconstructed variables of a neutrino event. I’ll brieflly explain why we need this.
The performance of the neutron tagging algorithm is quite dependent on neutrino event properties, such as neutrino energy and interaction vertex. This is because our neutron tagging algorithm is dependent on resolution of neutron capture vertex and the vertex position itself. For example, high energy neutrino events tend to exhibit smaller neutron tagging efficiency because a) the initial neutrino vertex reconstruction with APFit that we use as an input to neutron tagging doesn’t work well with multi-ring, and b) neutrons themselves tend to travel further from the initial vertex estimated with APFit. If we does not consider such an event-by-event difference in neutron tagging performance correctly, our final results will be biased as seen in Figure 11.
For a given neutrino event in our data, in principle we can look up MC events with similar reconstructed properties and check average neutron tagging efficiency and false positive rate. GAM is basically a lookup table trained on baseline MC. Plus, it gives uncertainty estimates based on the training MC.
Some extra uncertainties not considered in the training MC can be divided into two parts; a) neutron capture vertex reconstruction algorithm and b) outgoing neutron kinematics and interaction model. To evaluate their impact, we have the “reference” algorithm that uses BONSAI to directly reconstruct neutron capture vertex instead of using APFit neutrino vertex as a proxy. This algorithm is also much less dependent on neutron KE, and shows very similar result as the nominal APFit-based algorithm. Small difference (within stat errors) between the two algorithms are taken as a systematic uncertainty in the overall signal efficiency scale, as described in Appendix A. Additionally we have tested NEUT 5.1.4 with slightly different neutron KE spectrum with 10% lower neutron counts as shown in Figure 11, and GAM seems to perform well.

[Section 6]
- L659-660: Because you have "a" and "b" in G_18a_10x_02_11b, it is a bit confusing that you show {a, b, c, d}.
I’ll just leave as is for now, as this is a problem with GENIE’s naming…

- Figure 14: What is the main cause for the difference between SK-IV/V and SK-VI defaults?
As written in Sec VI: The SK-IV/V and SK-VI setups differ mainly in neutron reaction cross sections below 20 MeV, using ENDF/B-VI and ENDF/B-VII.1, respectively.

- L703-707: How did you treat position and direction of outgoing particles? Did you fire particle gun always at the tank center?
At tank center, at random direction. SK ID is large enough to contain all primary hadrons and secondary neutrons.

- L725-726: You can mention Figure 13 when explaining GENIE INCL.
Done.
- L727: How much more QE does exist from 5.4.0 than 5.6.3?
Added in L666: \texttt{NEUT 5.4.0} follows the setup described in Section~\ref{sec:defsimsetup}, while \texttt{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.*

[Section 7]

- Figure 17 legend label: SK "combined" means SK-IV, V, and VI?
Yes.

- L775: no indent
Done.t

- Table 3: Please explicitly show the setup of "baseline full MC simulation".
Added (\texttt{NEUT 5.4.0} with \texttt{SK-IV/V default} for SK-IV/V, and \texttt{SK-VI default} for SK-VI) in the caption.

- Table 4: Please have the caption at the top of table.
Done.

[Section 8]
- Mention Appendix B and its appeal in the main text.
Added at the end of Results (Sec VII): Model performance, separated into contributions from target nucleus FSI and secondary interactions, is further visualized using effective metrics in Appendix \ref{appendix:slopeintercept}.

- Also mention that the model discriminative power of the metrics are largely robust to the fit/measurement Evis range.
Added: These trends in model predictions and agreement with data remained robust under variations in the visible energy ranges used to define both metrics.

- Mention that Fig 22 visually separates FSI and SI impacts of model predictions.
Added in the beginning of Appendix B: *To separately evaluate the contributions from FSI within the target nucleus and downstream secondary interactions*, we define two effective metrics.

[Section 9]
- L956: Can you please have the following references in addition to Ref. [72]?
  * https://journals.aps.org/prc/abstract/10.1103/PhysRevC.109.014620
  * https://academic.oup.com/ptep/article/2024/11/113D01/7845878?login=true
Done.