Comments from the paper committee

Title

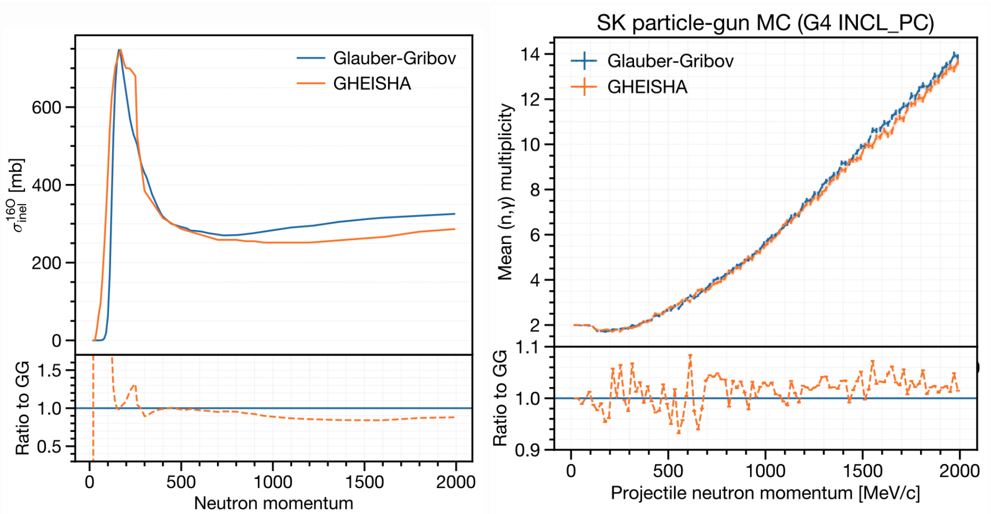
- (Name) L#: comment

Abstract

- (Name) L#: comment

General

- (Name) L#: comment
- Benda: the line number of the right column can be put to the right of the text. e.g. \usepackage[switch]{lineno}
    - SH: It looks like it conflicts with revtex4-2. (For example, see here) For now, I'll keep the current scheme.
- RA: Geant4 uses a common hadron-nucleus inelastic cross-section dataset, but the draft does not seem to address its uncertainty. Given the absence of relevant data in the energy region for oxygen, the uncertainty may be significant.
    - SH: Although we have not explicitly evaluated the impact of hadron-nucleus reaction cross sections, we have used two different cross-section datasets: one from Geant3.21/GCALOR, which employs Bertini's tabulation based on his own cascade model, and the other from Geant4, which uses the Glauber-Gribov+Barashenkov (BGG) parametrization. These datasets, derived from two distinct methods, should provide some coverage of the uncertainties in inelastic interaction cross sections.
    - Comparing the predictions of G3 GCALOR and G4 Bertini—both of which use similar INC and de-excitation models but different cross-section datasets—reveals that they are nearly identical at high energies and differ by at most 10% below 100 MeV visible energy, as shown in Figures 15 or 16. This suggests that uncertainties in cross section choices are likely smaller than those arising from the selection of INC and de-excitation models.
    - I tested the "G4 INCL_PC" model with two different cross-section parameterizations available in Geant4. The left figure compares the reaction cross section on 16O, while the right figure corresponds to Figure 15 in paper draft v1.0. Although the lack of experimental data for 16O is concerning, if these parameterizations are reasonable for other elements, this suggests that the uncertainty in the reaction cross section impacts neutron multiplicity at only the few-percent level.



    - I have clarified that Geant3 uses Bertini's tabulation for hadronic cross sections, while Geant4 employs its own BGG parametrization. Additionally, in Section VIII C, I state that while this study does not explicitly analyze the impact of cross-section uncertainties, the comparison between G3 GCALOR and G4 Bertini suggests that their effect is likely smaller than that of the chosen INC and de-excitation models.

- YH: It would be useful to analyze which particles, at what momenta, contribute most to the total neutron prediction. (Especially in terms of INCL underprediction in multi-ring events) This will help us decide which model is appropriate for each particle and momentum range.
    - SH: I have added the predicted hadron momentum distributions weighted by the mean (n,γ) multiplicity predicted by SI models, at the end of the text. This represents the average contribution of hadrons at a given momentum to the observable signals. INCL tends to underpredict neutron production from outgoing pions above 500 MeV/c. Jean-Christophe David, one of the INCL developers, pointed out that this coincides with the momentum range where the multiple-pion production channel opens. I have included this discussion in the text, but further investigation is needed.

1. Introduction

- (Name) L#: comment
- Benda 140–142: add a reference for the historical neutron tag.
    - SH: Cited 1956 Cowan-Reines paper.
- Benda Fig 1: The left schematic is a zoomed-in part of the left one. The schematic could be improved to make such a relation more obvious.
    - SH: Enlarged the right part of Fig. 1.

2. The Super-Kamiokande detector

- (Name) L#: comment
- Benda Table 1: 6 column, why minus-plus sign is used?
  - SH: This is because the Gd capture fraction uncertainty is anti-correlated with H capture fraction uncertainty (shown with ± sign)
- Benda L217: please confirm whether it is 0.25 charge or pulse height.
  - SH: Yes, this is indeed pulse height. The line is edited:
    → The detector registers a PMT signal with <u>a pulse height greater than</u> 0.25 photoelectron-equivalent charge as a “hit”.

3. Atmospheric neutrino events

- (Name) L#: comment

4. Neutron signal selection

- (Name) L#: comment
- BX: L#584-588: Can the impact of the hydrogen thermal motion modeling error in Geant4 be observed through the neutron capture time distribution in pure water for SK4/5?
  - SH: We have lack of Am/Be calibration data so the errors on SK4/5 neutron capture time (shown in Table I) is not small enough to validate this error with hydrogen-only data.
- Benda: Figure 9 caption, 350 ns got splitted into 2 lines. Needs a fix.
  - SH: Fixed.

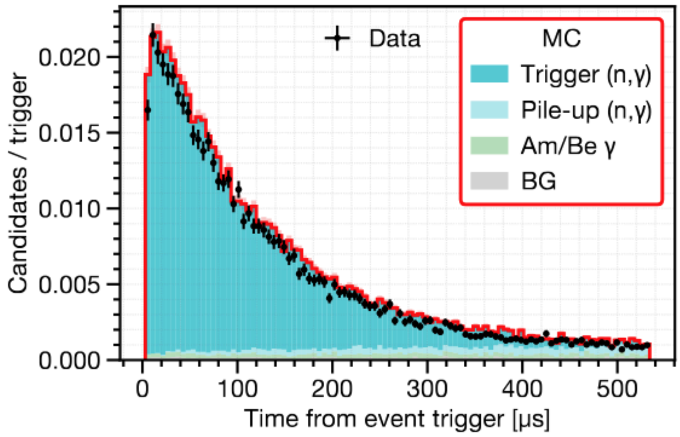


FIG. 9. Exponential decrease of the selected neutron signal candidates as a function of the time from the selected event triggers with the Am/Be neutron source positioned near the ID tank center, in the SK-VI phase. The label "Trigger (n, γ)" indicates captures of neutrons produced within 350 ns from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

5. (n,γ) multiplicity estimation

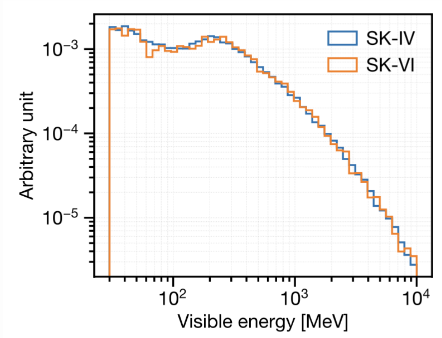
- (Name) L#: comment
- Benda L683: “hA,” <- likely a typo.
  - SH: This is not a typo. “hA” is a model name used in GENIE neutrino event generator.
  - Benda: Why is there a comma attached to it? e.g. L854 in draft 1.1. Likewise in Figure 24, “G4 INCL_PC,” ends with a comma while “GENIE INCL.” ends with a period. Maybe you meant to interchange the comma and quotation marks?
    - SH: I guess this is the punctuation rule (see here) According to this rule “G4 INCL_PC”, is wrong, but “G4 INCL_PC,” is correct.
    - SH: Regardless, I changed the qutoation marks with \texttt{}, to improve readability and bypass punctuation issues.

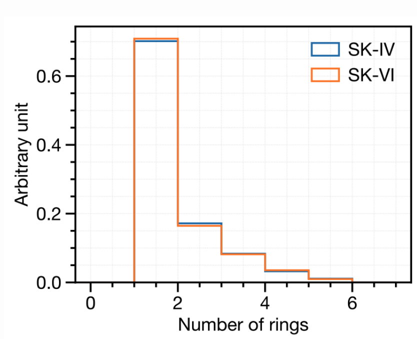
6. Tested interaction models

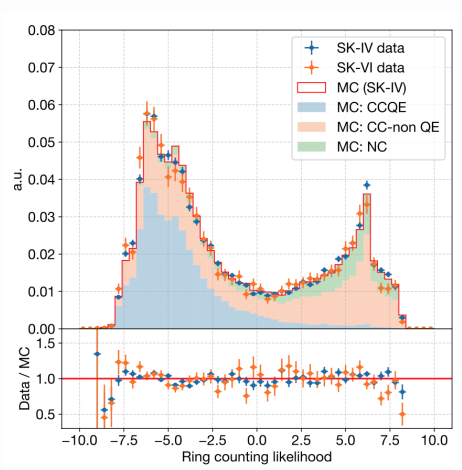
- (Name) L#: comment

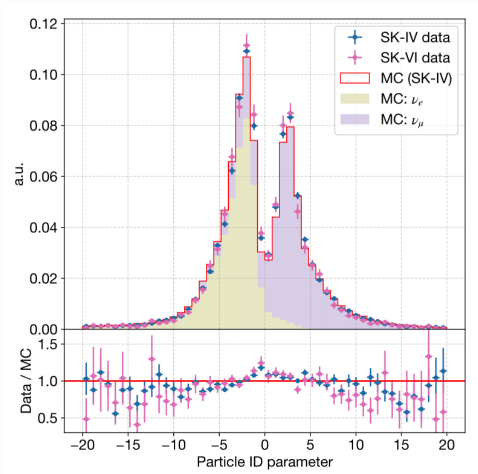
7. Results

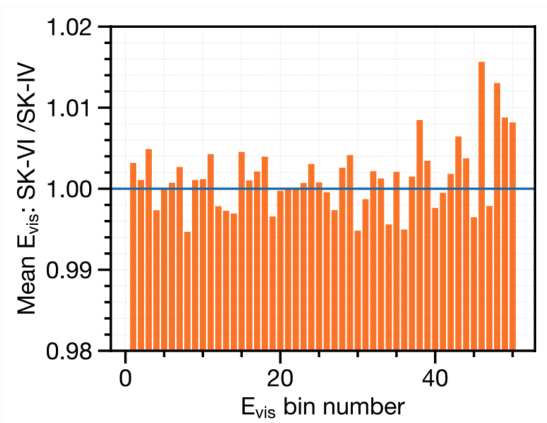
- (Name) L#: comment
- RA Fig. 21: Regarding the difference in SK4/5 and SK6 fitted slopes, differences in prompt lepton selection efficiencies between SK4/5 and SK6 might explain the discrepancy.
  - SH: PID, NRing, Evis x errors seem consistent between SK4 and SK6.











  - SH: Slope fit errors were dominated by y errors so that adding Evis x errors had no impact on the fit result.
  - SH: I was calculating statistical errors wrong. It was changed from:

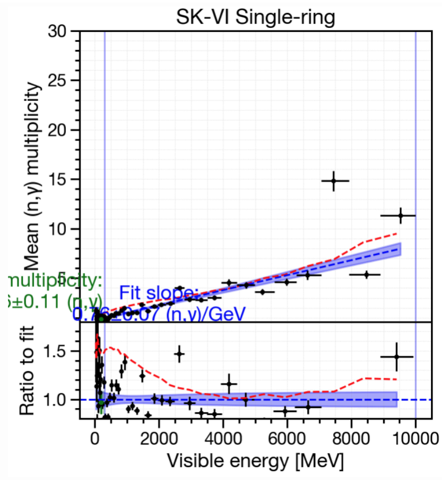
$$\frac{\sqrt{\langle N_{\text{detected}} \rangle (1 - \langle \epsilon \rangle)}}{\sqrt{N_{\text{event}}} \langle \epsilon \rangle}$$

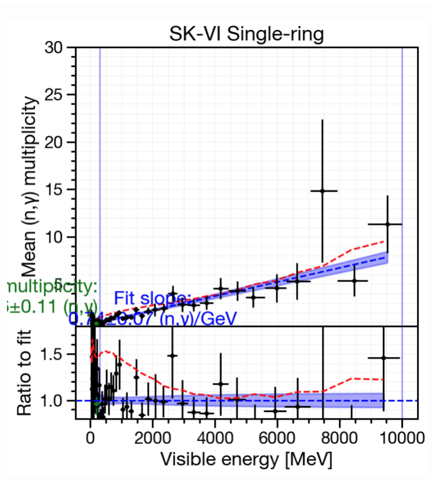
which assumes binomial variance with "constant" number of trials (i.e., number of true neutrons), to:

$$\frac{\hat{\sigma}_{N_{\text{detected}}}}{\sqrt{N_{\text{event}}}}$$

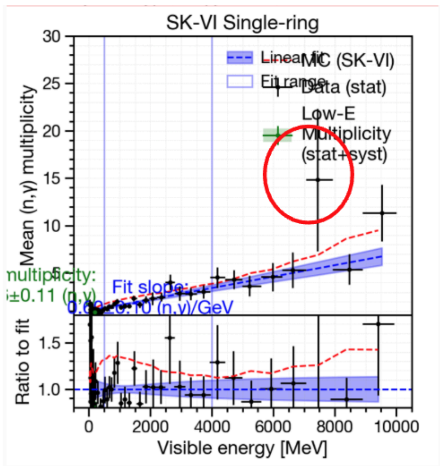
which roughly accounts for the latent variance of the number of true neutrons. $\hat{\sigma}$ stands for the unbiased sample standard deviation.

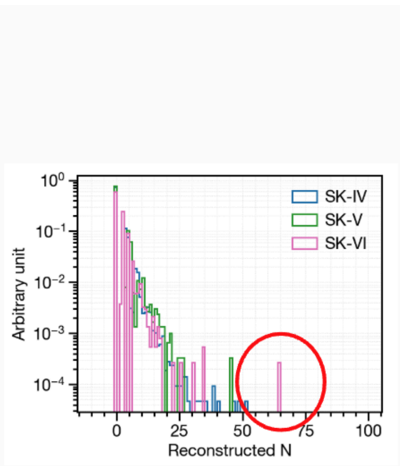
The plots below show SK6 single-ring slope fit result before (left) and after (right) the stat error fix. The fitted slope has changed from 0.76±0.07 to 0.74±0.07.



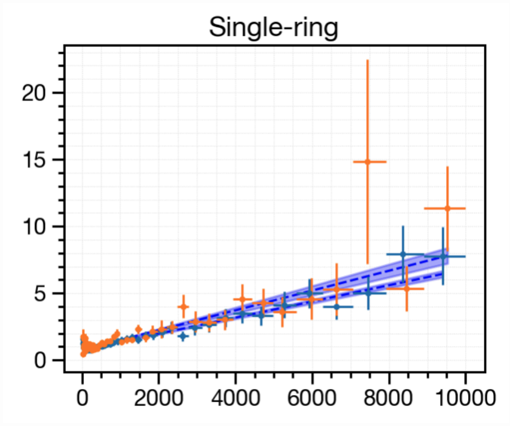


As seen in the plots above, the slope is affected by outliers. For example, in the SK6 single-ring sample, a single event significantly increases the slope. This event may be a RES/DIS interaction (mis)identified as single-ring.

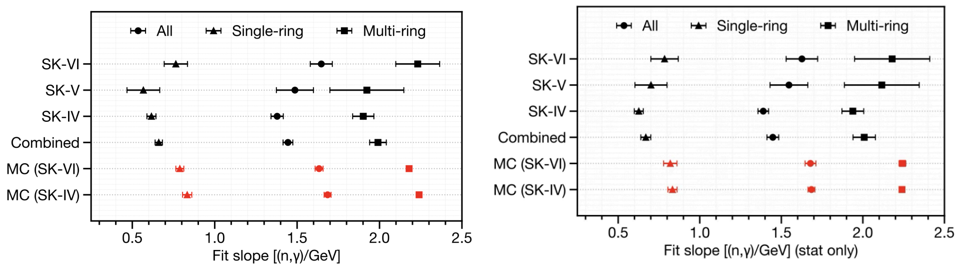




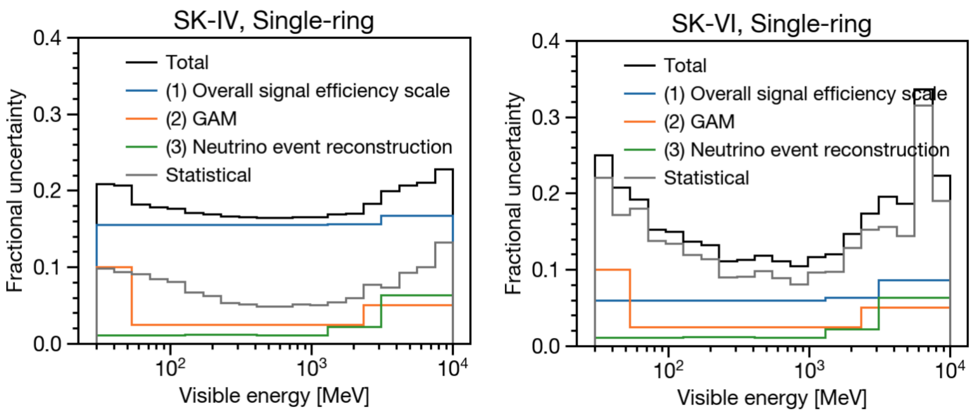
Comparing the single-ring results from SK4 (blue) and SK6 (orange) on linear scales, the overall trend appears consistent. However, the fitted slopes differ by 1–2σ, possibly due to outliers present in both datasets. It is difficult to determine whether this difference is a systematic effect.



To ensure each bin has enough number of events and mitigate outlier effects, the number of bins were reduced from 50 to 20 for the slope fit for each SK phase data. The resulting change in the phase consistency plot (Fig 21) is from left to right:



After applying the statistical error correction, SK6 uncertainty is primarily driven by statistics, unlike SK4.



All data plots and fit results will be updated with the statistical error correction. While the errors have increased slightly due to larger statistical uncertainties, the overall impact on the output figures is negligibly small.

8. Discussion

- (Name) L#: comment
- RA: L#873-882: The Fermi breakup process should be called more often (even for a 100 MeV KE neutron projectile) compared to the statistical evaporation process in question. Additionally, level density parameters significantly affect nucleon emission and should be worth mentioning in the paper.
    - SH: Replaced the explanation with the following:
      This peak arises primarily from the nuclear de-excitation model in the Geant4 Bertini cascade \cite{g4_bertini}. *A key characteristic of this model is its restriction on Fermi breakup, which is disallowed for nuclear targets with mass numbers greater than 11 \cite{g4_bertini}, including oxygen (and also carbon, which is particularly relevant to past MINERvA studies \cite{mv_neutron_2019,mv_neutron_2023}). The more recent Geant4 Precompound model \cite{g4preco} allows Fermi breakup for mass numbers up to 16, including oxygen. This additional fragmentation mechanism is expected to reduce the number of ``single'' neutrons contributing to (n,γ) signals. Further differences between these models include variations in nucleon absorption (inverse reaction) cross sections and level density parameters \cite{g4preco}, which may impact single-neutron production in the statistical evaporation process.*

9. Summary

- (Name) L#: comment