Ad Hoc Review Response for the g14/CLAS paper
"Beam Target Helicity Asymmetry E in $K^0\Lambda$ and $K^0\Sigma^0$ Photoproduction on the Neutron"
dated March 7, 2018

CLAS Ad Hoc Committee: William Briscoe, Marco Mirazita, Gabriel Niculescu
Respondents: Reinhard Schumacher and the g14 Run Group

Reply dated: April 5, 2018
Updated Draft of paper dated: April 5, 2018

March 27 2018 comments by Marco Mirazita

I have read the paper, I don't have found any major issue. My main comment would be that I've found the section III rather complicated to follow for a non-expert on the BDT analysis technique, as I am. But probably this is the best one can do given the limited space of a paper. Beside that, I list here my comments.

Reply: Yes, the description of the BDT method is most difficult part of the paper to read, especially if one has not encountered it before. This is the first CLAS analysis publication that makes use of the method, so we think it is important to outline what we did in enough detail.

1) Page 2, Section II, line 7 (the line numbering is missing here) Add ref. 25 together to ref. 24

Reply: Thank you for catching that. Fixed.

2) Pag 6, line 6 after Fig. 4 "...true the K0Y..." -> "...true K0Y ..."

Reply: Done.

3) Figures 7, 8, 9 Each one of the model is shown with two curves corresponding to the minimum and maximum W values. The result is that at some particular cos(theta) values the two curves intersect each other, as if at those points there would be no variation at all in the whole W range, but I hardly believe that. I would suggest to show instead one curve averaged within the measured W range, with a band for the calculation uncertainty.

Reply: In our analysis group (g14) we have settled on the way of representing the variation in the model calculations after quite some discussion. Our trouble is that the W bins are very wide, so the models can vary a lot between one edge of the bin and the other. Plotting a "middle" curve could be done but we did not like that since it does not capture the (sometimes large and rapid) change in the model across a bin. When the curves cross it does not mean the model is unchanging across the whole bin, of course! But this, we think, will not confuse the reader. We have already used this way of showing the range of variation of the models in our previous publication, Ref 44: D. Ho et al. (CLAS), "Beam-Target Helicity Asymmetry for $\vec{\gamma}\vec{n} \rightarrow \pi^- p$ in the N^* Resonance Region," Phys. Rev. Lett. 118, 242002 (2017). Therefore, we plan to stay consistent with showing the boundary curves as given in the draft paper.

4) Page 9, lines 377-386 and Figure 9 I don't like Figure 9 too much. The left plot is a repetition of Fig. 7, the right plot is just a comparison between proton and neutron models. If the goal is just to say that neutron models must be improved, this is already suggested by Fig. 7 and 8 (although the statistics of the experiment is not allowing any definite statement). On the other hand, it could be interesting to compare the present neutron data with proton ones and eventually comment on what one can infer from the comparison.

Reply: We admit that we are trying, with Fig. 9, to make the most out of a rather modest experimental result. The value in keeping this figure is that it shows the reader very directly, by eye, that the predictions for proton and neutron observables are quite different, which we cannot say convincingly with words alone. The Bn-Ga curves in the right-hand plot are also new from our co-authors, and we want to highlight their recent work in this area. Thus, we plan to keep this figure in the paper.

March 25 comments by Gabriel Niculescu

Review for: Beam-Target Helicity Asymmetry E in K0Lambda and K0Sigma0 photoproduction on the neutron

I. Introduction:
The authors attempt a very challenging analysis involving strange, neutral particles in the final state. The rarity of this process (at least compared with the non-strange backgrounds), combined with the need of extracting the polarization observable E makes for a statistically-challenged environment. The small (~80 MeV) mass difference between the two hyperons studied provides an extra degree of complexity due to the substantial overlap between the Lambda and Sigma0 states. The article tries to make the most out of this situation, relying to a large extent on the Boosted Decision Tree (btd) technique for event selection and reaction channel separation, in the process introducing the Hall B/CLAS community to this useful, though not new (earliest references go back to the early-mind 90s. Usage in high energy/particle physics dates back at least a decade) classification tool.

Reply: Thank you for recognizing the challenge that we faced in extracting this information!

II. General suggestion/Comments.
This reviewer finds the article generally well written and relatively easy to follow. Given the modest statistical significance of the result the conclusions are not (nor they can be, one thinks) very forceful, though in the context of PWA analyses every bit of data helps, especially in previously unexplored/less explored channels (such as the ones reported here) - and this point is well established in the text. Below are a few specific comments which, in the view of this reviewer, should clarify/strengthen the paper and, if implemented, they will hopefully increase the chances of this article being accepted/published.

Reply: Again, thanks for the comments.

III. Suggestion/Comments.
1. Title: The paper promises "beam-target ... on the neutron" - this is strictly not true. Asymmetry we're extracting is off of a bound (quasi-free) neutron. There are quite a few theorists that will argue that nuclear effects need to be accounted for in all cases. If this paper get a journal reviewer with that train of thought we might need to alter the title and/or argue in the narrative why we think these effects can be neglected.

Reply: You are correct that we have theorist who are very picky about this issue. One of the pickiest is Igor Strakovsky, and you will note that he is already on the author list. We investigated this issue in connection with our previous publication, where we used the high statistics channel (pi- p) for which we could study the effect of the recoil momentum cut on the results. That is, Ref 44: D. Ho et al. (CLAS), "Beam-Target Helicity Asymmetry for $\vec{\gamma}\vec{n} \rightarrow \pi^- p$ in the N^* Resonance Region," Phys. Rev. Lett. 118, 242002 (2017). summarizes this issue and we discussed a lot about it in the analysis note . The "narrative" discussion you recommend is already in the draft paper. See lines 293 to 318. We make an explicit correction, as discussed there, for the polarization state of the neutron inside the deuteron.

2. Line 118: "Background events..." - here, as well as in several other places in the paper one wishes the authors will be more forthcoming with actual numbers. As it stands now the paper is sprinkled with statements that seem to be quantitative, but they are not so: "... were small in number compared to signal events.". As far as this reviewer can tell there is only one place where actual numbers are reported, Table I which gives the overall result and its uncertainties.

Reply: Some examples of places where we do have more quantitative results include: lines 280, 291, 303, 318. To your specific objection about the statement at line 118, we have modified the text to refer instead to the section of the paper where background subtraction is treated fully.

3. There is a whole paragraph between lines 118 and 119 that is not numbered.

Reply: (Our version of RevTex macros does not number lines in paragraphs that have actual math equations, sorry.)

4. Line 128: This reviewer is not sure how to interpret the phrase that starts here. Do the authors literally mean that they carried out the statistically rich gamma d \to pi- p (X) reaction, obtained the yields through the three methods listed and concluded that the btd method is superior because it provided "larger event yields" - One fails to see how this is a valid argument as written. One would hope that when carrying out an analysis one strives to get as close as possible to the RIGHT event yield, be it LARGE or SMALL. Just because a technique provides a larger number of events for a particular channel it does not necessarily mean it is a better technique compared to one that provides a smaller number of events. If one were to follow the logic implied in the text the following situation will occur: Assume that there are (but of course we do not know it!) 100 counts in a data sample (buried in a lot background, etc.).

Technique A "finds" 150 "good events" while technique B "finds" only 95 counts. According to the paper the 150 figure should be used in subsequent analysis, resulting in a result that is way off the mark. Perhaps there is more information in the analysis paper/PhD dissertation that can make this point clearer, hopefully without much addition to the text.

Reply: You have indeed found a place where the text is less than optimally clear. Yes we DID do the full analysis of the pi- p channel and compared three independent ("parallel") analyses to verify that the BDT method preserves the most SIGNAL events. All three methods gave consistent results for the E asymmetry we are measuring. The more signal events we keep, the better the precision of the result, of course. The BDT method resulted in the smallest error bars of the three methods because the statistics were highest: it minimized the unnecessary loss of good events.

The key thing is that the three analysis methods led to consistent results, with an advantage of the BDT method being that the event count was highest. This was discussed in our previous paper (cited above), and presented in enormous detail in the corresponding analysis note.

If we were measuring, say, a cross section instead of an asymmetry, then the logic would be: the BDT method preserves for us the most "signal" events, but the associated acceptance would **also** be calculated to be bigger, and so the measured cross section should be unchanged from what it would be using a less sophisticated method.

We are measuring an asymmetry, for which no Monte Carlo calculation was required, but the logic is the same: if we have a method that preserves more true signal events, we can use it.   We have changed the text of the paper near line 132 to make this more clear.

5. Data analysis section (as a whole):
a. As the text defines and refers frequently the 3 BDTs used, perhaps it would be a good idea to label this in some way (maybe BDT1... the way it is done further down with the W bins. This should shorten the text and streamline the explanation.

Reply: We changed the text starting at (old) line 178 to make it clearer that a three-step process was used.   You are right to point out that we did not make it clear to the read what was coming in the discussion.   We think the revised text sets the reader up to not be confused by the sequence of BDT's used in this analysis.

b. As alluded to at line 192 - maybe a graphical representation of the BTDs would help guide the reader. This could in principle be done either in addition or instead of some of the descriptions associated with the BDT implementation and usage.

Reply: The long paragraph between lines 188 and 213 is Dao's compact description of the BDT method.   We debated adding a figure to make it more clear to a neophyte.   Given the modest results we have to show, it did not seem like a good idea to lengthen the paper with more description of this step.   We wanted to have one paragraph of introduction to the method, however, since without such a paragraph many readers would be clueless.   If anything, we would shorten the paragraph a bit.   Mainly, we want to point the reader to the relevant references. We have shortened it (see below), hoping that it comes across at a reasonable level for most readers.

c. Line 211 (and the whole section as well). The authors rightfully point out that BDT usage is a decade-old (if not older) technique in our field. While, as remarked in the Introduction, this is a very welcomed (in the reviewer's opinion) addition to the CLAS software pool, the latter part of the statement (which explicitly mentions CLAS) should probably be left out. One hopes that based on the dissertation and the analysis paper the authors get all the internal CLAS credit they deserve for this.

Reply: Yes, we are bragging a bit here.   You are probably correct. We have removed this sentence.

d. Following up on c. - one can probably shave off some of the "technical/implementation" details relating for the BTDs (for example the phrase starting at 195 can be shortened/omitted) - after all, we argue that this is an established technique, therefore the reader can be pointed to earlier work in the field that made use of the technique.

Reply: Yes, the discussion is too general here. Sentence removed.

6. Line 232: See comment 2. Can we be more quantitative about this remaining background?

Reply: Changed the wording to point the reader to the next section where this is handled.

7. Line 269: See comment 2. For any classification machine learning technique, such as the one used in this analysis, one can build the so-called "confusion matrix" that gives a complete picture of the rate at which events are mis-classified. This reviewer thinks that such an addition to the paper will go a long way in substantiating the overall analysis reported in the paper and give a more solid footing to the numbers in Table I.

Reply: We think that Figures 5 and 6 tell the story already in enough detail to convince the reader of our results. The discussion about the performance of the BDT's with respect to sample purity, confusion, and the specific values of the BDT cut parameters that were selected are in Dao's thesis. We think it would be overkill to go into that much detail in this paper. We made no changes to the text in this case.

8. Line 253: Very nice recovery and explanation for not believing the dip shown in the top part of figure 4. Presumably the degree of the polynomial was varied and the difference in backgrounds for the various choices was taken into account in the final systematic tally, right?

Reply: Yes. There is a long section in Dao's thesis on the systematic checks. The degree of the polynomial was less of an issue than the value of the second BDT selection cut. Varying this cut was tested to estimate the range of correction factors for this issue. Table 5.20 in his thesis shows the results if you care to deep-dive into it. The systematic variations are smaller by far than the final statistical uncertainties. It's tricky with low-statistics data because varying the cut tends to lead to over-estimation of the systematic uncertainty since the statistical uncertainty in involved at the same time.

9. A lot of the results rely on color plots. This was presumably done on purpose and the authors are confident that the target publication will not demand other ways of distinguishing between the various distributions/curves shown.

Reply: The real results in Figs 7 to 9 are carefully prepared to use both color and line-type. Those plots are fine. Figs 5 and 6 could be objected to, but I doubt they will be. If the journal insists, we could mention the direction of the hatch marks in the panels to distinguish the histograms.

10. Figures 5 and 6: Bin size changes between these figures. One understands that poor statistics dictate the size of the binning for Fig.6 - The argument these figures support will be equally as valid if Fig.5 is rebinned to match Fig.6. Presumably there is/are strong argument(s) in the dissertation/analysis paper that can be brought to bear (if need be in the journal review process) if one asks about the credibility of training the BDT on a simulation that seems to have a different resolution than the data. Same goes for the presence/absence of noise in data vs. simulation. Besides Fig. 5, which is very nice, one would argue for the inclusion of a figure equivalent to Fig.6, except from the simulation.

Reply: It's notable that you noticed this! I asked Dao about it and he answered that Fig. 6 was rebinned more coarsely in order to make the histograms tolerably smooth given the statistics we have. Unfortunately, there is no possibility of remaking any of these figures; too much time has gone by.

11. Figures 7-9:
00: there is no label on the Y axis for these plots.

Reply: Figures remade, for this and subsequent reasons.

a. There is a substantial change in the quality of the plots going from 1-6 vs 7-9. One can distinguish (maybe?) the change from a ROOT/C++ output to a F77/PAW output. Depending on the target publication, the journal might insist in uniformity in the way the figures look. As it stands now this is not the case; the labels and axes of figures 7-9 are smaller fonts than their 1-6 counterparts. If the authors do not have access to the numbers associated with the theoretical predictions directly (for example if using some of the online model facilities) then a possible workaround would be to:
- digitize the theory curves off of a png, gif, etc. (there are several stand-alone as well as online programs that can do this). Yes, it can be tedious, but it can be done.
- redo the figure(s) in ROOT, ensuring uniformity.

Reply: Figures remake, but using the same plot package as before.

b. Not sure if the W labeling ought to be "(GeV)" or "(in GeV)" or "2.02 GeV < W < 2.34 GeV". Also, are we 100% sure the intervals were open ended at both ends?

Reply: We use "(GeV)" to save space.

c. Depending on the target journal, the authors might have to explicitly state the size of the W bins in the caption.

Reply: Good idea. We did this.

d. A significant amount of space can be saved if one were to not include the vertical axis for the RHS panel - thus giving more room to show the results. As they stand now Figs 7-9 are very "tall and skinny".

Reply: Figures remade to take care of this, in part.

e. Figures are pretty busy. Is it really necessary to have grid lines at y=0,+/-1 ?

Reply: Yes, some of us really like putting some markers at the physical boundaries.

f. For each of the 3(2) models two sets of curves are shown, one for each of the (arguably) large W bins used in this analysis. The reader is left for his/her-self to figure out which curve corresponds to the lower/upper W value. A possible suggestion would be to use different line thickness to distinguish between the two.

Reply: It hardly matters, we think, since the data and the models are both so spread out.   The modelers will surely know which way the trends go. The more casual reader does not care or need to know.

g. Many of the theory curves seem to "cross" when evaluated at the two W bin edges. See for example the Gatchina and SAID models in the higher W bin on Fig7., etc. Not clear if the KaonMAID predictions cross in both panels of Fig. 7 (or the lower panel of Fig.9) - they might they might not (just come very close/touching, then parting ways.).

Reply: The curves definitely can cross, just according to whatever the models give. Again, the modelers will know the trends and the more casual reader will not really care or need to know, we believe.

h. Furthermore, quite a few of the model predictions seem to change the sign of their second derivatives between the two W bin ends. See for example KaonMaid in the W2 bin of Fig.8 for cos theta in the -0.3 – 0 range.

Reply: That's perfectly acceptable behavior for the models: asymmetries can do things like that.

i. From g. and h. one can conclude that unless we have absolute proof (and this reviewer doubts that we do) that indeed the bin edges are extremum (minima, maxima, saddle) points for all three models shown, for all channels - (this would be a REALLY interesting coincidence!) then showing the model evaluated at the bin edges is of little use, because presumably the model will have (at least one) point between the two W limits of a bin where it will go even higher (or lower) than the two curves shown. So all conclusions that claim, as in line 418 or about, (slightly) better or worse agreement with a model or another should take this fact into account.

Reply: The data implicitly average over the range between the extrema, of course. If there is a higher or lower value in between (which is entirely possible), then the data average over those extremes. Our goal is to stake out the boundaries of the averaging range by plotting the extrema of the model curves. We think this is OK since it is the method we used in the previously-cited PRL on the (pi- p) reaction. Of course here the energy bins are wider, but the idea is the same.

j. Picking up on i.: - the CLAS results shown represent, to the best of our knowledge, the cross-section averaged value for the observable E. This is not the case for the model predictions. As the authors rightfully point out (line 373...) there cannot be a directly comparable theoretical prediction is one lacks the corresponding (differential) cross-section such that an equivalent averaging can be done for the model as nature has done for the experimental data.

Reply: Yes, same logic as stated above.

k. Given i. and j. and the still ongoing "hunt" for "missing resonances" one would have hoped that one of the more forceful/forward looking conclusion of the paper would be for more (a lot more!) data in the strangeness sector, possibly using the CLAS12 spectrometer (or other JLab setups?).

Reply: Our last paragraph is about as strong a statement as we care to make about these results.

IV. Conclusion.

The authors extend the ongoing quest for missing resonances by providing new results in the form of beam-target polarization observables in the neutral, open strangeness sector. The analysis takes full advantage of the powerful BDT technique for particle/channel identification. Though modest in their statistical precision, this data will be very useful for the several theory groups active in the field. In the view of this reviewer the paper would benefit from a more explicit, quantitative of the uncertainties associated with the various analysis steps. These should be fairly straightforward to implement as suggested above. The latter figures in the paper might require refurbishing - one can proceed with them as they are but it is a very good chance they might delay publication if they are found inconsistent, fonts undersized, etc. These figures might also benefit if steps aimed at clarifying the meaning of various curves is made more explicit. See comments j. and k. about the physics conclusions and (possible) future work.

Reply: Thank you for your thoughtful comments.   Please have a look whether the revised draft of the paper is OK.