REFEREE D COMMENTS

Alicia's Note: The following are the comments from the fourth referee of our manuscript from PRL. Of the first three referees, two supported this work's publication in PRL, and one suggested that it should be published in Phys. Rev. D. The fourth referee was more critical. I believe their comments reflect two general trends. First, the expectations set in the introduction of the paper did not match the conclusions made in the discussion. To address this issue, we have made substantial changes to the introduction, title, and abstract. Second, certain aspects of the experimental setup were not clear to the referee. Clarifications have been added, and correspond to the bold text in the current draft of the manuscript.

The manuscript re-submitted for publication in PRL deals with a measurement of beam-helicity induced azimuthal modulations in exclusive electroproduction of positive pions off the proton. In certain kinematic regimes and under certain assumptions they can be related to generalized parton distributions (GPDs) and as such to a three-dimensional picture of the proton. The latter is clearly a topic of high interest with several world-wide activities, past, ongoing, and future (including the physics program at the future Electron-Ion Collider to be built at Brookhaven National Lab). Other phenomenological approaches are also available that might be more relevant in different kinematic regimes (than where GPDs are expected to yield a good description). Leaving for the moment technical aspects about the measurement presented aside, is the expected impact, interest, and level of innovation sufficient to meet the requirements for publication in PRL?

A) First of all, I must admit that while I was initially looking forward to reading the manuscript, I was somewhat disappointed about the level of sloppiness and of being too technical that has survived internal and the ongoing external review. To give some examples:

The abstracts starts with "The cross section ratio $\sigma_{LT'}/\sigma_0$ was extracted from the beam-spin asymmetry ALU" without further specifying what these cross sections are? While not wrong and possibly fine for a specialized journal, it does not seem to target the general PRL readership, and this style continues (the symbols are used multiple times in the abstract, the kinematic ranges are provide by just the symbols of the kinematic variables, and in fact, the significance of the prime on the T does not seem to be elucidated).

Fig. 1b) the pi+ cannot be simply produced at the vertex but involves another soft part not even appropriately discussed anywhere in the manuscript, casting doubt about the seriousness of the entire physics discussion (at least for the GPD description).

It is stated that the DEMP reactions can be described by three kinematic variables, only to then have four for ALU in equation (1). [Admittedly, phi is not a Lorentz invariant but a relevant quantity to describe DEMP.]

B) Impact & Interest: as already mentioned, the 3D structure of the nucleon is of high interest and several programs were/are/will be dedicated to study GPDs. However, that physics motivation comes way too late and remains way too vague to be convincing. Moreover, the Regge approach does not provide such a direct 3d picture, at least to my knowledge. Even the GPD picture is somewhat hampered here as rigorous only for longitudinal photons. The GK model involves additional assumptions.

The main problem for me is that at the end, the final result does not seem to provide insight into the 3d structure. The conclusions are weak except that none of the models seems to work but if the paper is dedicated to the "Comparison of beam-spin asymmetries ALU. with [several model] expectations", one would hope for more physics insights, at least to spur the interest of the PRL readership.

C) Innovation: While there are certainly new aspects (enlarged kinematic region, different experimental approach, etc.), the overall level of innovation does not convince me to be substantial enough to grant publication in PRL. The mere comparison to models can not be considered an overly innovative step forward, if the models are not further scrutinized and varied to get further physical insights. Also, the Q2 dependence is looked at using actually more data points from previous experiments than from this work.

Again, little indicates a leap step forward as to recommend publication in PRL, neither concerning a much increased data set nor a fundamental understanding of modes or even the nature of proton structure.

With these considerations, I cannot full-heartedly endorse a publication in PRL, but rather recommend a more specialized journal like PRD, after above concerns about the actual manuscript and further details listed below are addressed.

Further detailed comments:

YCK1 provides the closest description both visually and from chi2; however, it does not manage to catch all the pronounced features of the data, especially when going to lower xB. Moreover, the chi2 is too large to talk about a satisfactory description. None of the other models even get close to describe the data in their entirety. The statement “Quantitatively, YCK1 has the lowest average chi2/NDF, followed by VR [34].” is almost misleading, as it might suggest that both do well, but the p-value, especially for the VR model must be dismissive. [Disclaimer: I did not check it myself.]

The Q2 dependence is tested in two (x,t) bins with three or four data points not even spanning a factor two in range with still very large uncertainties. The statement “YCK1 better predicts the Q2-dependence at x =0.4 than x =0.25” is to be taken with some caution, as the actual Q2 dependence appears flat, better in agreement with the Q2 dependence predicted by YCK2. By contrast, YCK1 predicts an increase in the range considered. That the overall magnitude is off, especially for YCK2, is a different point.

p2: while the definitions of Q2 and t are Lorentz invariant, there are not in the cases of W and xB -¿ specify the (target rest) frame or even better use a Lorentz-invariant definition.

p2: “can be treated as a twist-3 effect in this approach [6]” - probably better to write ”has been treated as a twist-3 effect in this approach [6]” as it is not clear (at least to my knowledge) whether indeed this can be done in a rigorous way.

Fig. 1: suggest to reverse order of the two diagrams (and text in caption) as to follow flow of discussion in the main text.

p2: “can be used to probe fundamental quantities such as the still unknown tensor charge” - how viable is this actually as one needs to evaluate the complete integral of the GPD?

Eq. (2): parentheses in ALU definition probably not needed.

Eq. (2): a priori, the event yields should be luminosity normalized. There is an indication later on that this is fulfilled by the way polarized beams are provided, but this is not clear at the moment the equation gets introduced. AND p3: “subtracted from charge normalized event counts.” - here the N’s are not any longer just event yields, but normalized ones. Again, the charge-normalization might not be obvious to somebody that measures luminosity differently (or not at all as being a theorist). It makes more sense to already in Eq. (2) indicate a luminosity normalization, use that terminology, but also explain that this is done at looking at the total beam charge delivered.

p3: “with a charge asymmetry of about 0.1%” - this probably refers to the way luminosity is measured at Hall C and is not about a physical beam-charge asymmetry. This should be made clearer.

p3: “Two aluminum foils placed 10 cm apart were used” - I guess they were used as a separate “target” setup and during data taking you switched between these setups? Discussion might profit from some additional details. How much data were taken with the Al foils?

“SHMS, momentum acceptance delta p/p from -10 to +20%” was not entirely clear to me. “delta p/p” is often used to refer to resolution. Here, it is the range around some central momentum value? Which one? The “11 GeV/c” mentioned in the line above? But why 11 GeV – that’s larger than the beam momentum. And if the electron carries about 7 GeV +/- 8% than the pion should be much less energetic. How are these numbers to be understood?

Fig. 2: what is the right tail of the Mx distribution? This is important as it influences the way of how it gets treated. It is not so obvious that these cannot be K-Lambda events with the wrong mass assigned to the particles. Also, can QED radiative corrections contribute to the tail?

p3: “they cancel in the calculation of ALU” is not a general statement. If not evaluated fully differential, it requires a certain dependence of the asymmetry efficiencies on the variables integrated over.

p4 “Using the over-determined p(e, e’ p) reaction” - do you mean “p(e, e’)p”? Moreover, it’s better to explain why p(e, e’)p is used and how, e.g., can you use the exactly same setup to detect the proton or do you have to change something?”

p4 “An error-weighted average was then taken” - why not simply use all the data points separately? That should not harm the fit. By contrast, by forming the average, you might hide (average out) discrepancies between data.

p4 “for central kinematics of Q2 =3 GeV2 , xB =0.25” - are these really central values or just more or less central and used here mainly as ”labels”. AND Fig. 5 the x and Q2 labels could be misleading as the values seem a bit arbitrary. They are not even the average values of these quantities. AND - Fig. 5 (d): the supplemental material suggests that this panel is at x around 0.21 and not 0.25 as stated in the figure.

p4 “was determined using Eqn. 1” - how stable is the fit when including parameters in the denominator, which is often a problem? Have you tried different starting values of the parameters?

Fig. 5 a, b) the first data point is usually many sigma away from zero where the behavior of the curves suggests that this is already a range outside the possible t range (e.g., below tmin)? Or do the curves continue but becoming negative? Also, why not bin in -t’, subtracting the minimum t that is (x,Q2) dependent?

p5 “Q2-dependence at fixed xB and t.” - strictly speaking, xB and t are not really fixed. Moreover, would be nicer to spell out the values of $< xB >$ and $< -t >$.