Report

Readiness Review for pass1 processing of the HallB/CLAS12 RG-K data set

Monday, June 8, 2020

Review committee:
Nathan Baltzell
Marco Mirazita
Cole Smith
Stepan Stepanyan (chair)
Larry Weinstein

The review to assess the readiness of RG-K to process the data set acquired during the December 2018 run took place on June 6, 2020, remotely over bluejeans. For the agenda and presentations please refer to the review page:
https://clasweb.jlab.org/wiki/index.php/Run_Group_K#tab=Pass1_Review_Documents

The review committee thanks the RG-K team for preparing the presentations, providing ancillary information, and for patiently answering our questions during the review meeting. Below are our answers to the Review Charge questions, and some comments and recommendations. RG-K has done a great job with calibrations and understanding the data set. The quality of the calibration for all detectors is adequate for the Pass1 data processing. The only recommendation in this review report is about the physics skims (see below).

General Remarks (also for CCC and the CLAS12 software group):

A. This is the third run group that came forward for the Pass1 processing of their data sets. It has been our understanding that the goal of this large-scale data processing at this stage is to provide enough statistics for the first publication(s) of physics results from CLAS12 in FY20. Since certain aspects of the CLAS12 software, alignment, and calibrations are not yet final, e.g. CVT alignment and the tracking algorithm for the central tracker, these data sets will be re-processed in the future after everything is ready. With that said, RG-A and B had clear plans and demonstrated efforts to publish analyses this year without CVT information. It was not clear from the RG-K presentations which, if any, reactions can be put on fast track analysis to get publication-quality results in a half-year from now. Moreover, all of the presented reaction channels can benefit significantly from proper alignment and software for CVT, which may be ready in a few months.
B. The importance of small skims have been discussed many times by the software group. We all accept that the small size physics skim helps run analyses faster and preserves common resources. RG-K is not the first group to get a recommendation for production

of a large size skims. It will be helpful for all RGs to have clear guidelines from the CCC and the software group on skim sizes, skim frequencies and storage in line with the scheme for “trains/wagons” adopted for CLAS12 in the 2018 software review.

Review Charge:

Charge #1: *Is the quality of detector calibration and alignment adequate to achieve the performance specifications foreseen for CLAS12 or achievable at the current time, given the “state of the art” calibration, alignment and reconstruction algorithms?*

Yes – The current algorithms and methods used for calibration, alignment, and the reconstruction of data are adequate to achieve the CLAS12 performance specifications as has been demonstrated for RG-A & B and now RG-K.

Comments:

Achieved performance of the CLAS12 detector with RG-K data (at both energies) are well within specifications.

Recommendation:

None.

Charge #2: *(a) Is data quality as a function of run number or time for the data set that is proposed for pass1 cooking stable and understood? (b) Is reconstruction efficiency consistent with expectations and reproducible by appropriate MC simulations?*

(a) Yes – timelines for most of the parameters have been shown and discussed. Most of the monitored parameters are within expected tolerances, close to what have been achieved with RG-A & B.

(b) No – but, there is an ongoing work in the collaboration to understand the reconstruction efficiencies, develop and validate beam background merging software for the data and MC. It is expected to have everything ready in about a month.

Comments:

Timelines showed consistent and stable performance of detector components. Calibrations are as good, in some cases even better, than RG-A & B.

Recommendation:

None.

Charge #3: *(a) Are analysis plans for the data set developed at adequate levels? (b) Is the list of planned skims defined and tested running analysis trains on preliminary data? (c) Are preliminary analysis results for the main reaction channels and observable available and consistent with expectations? (d) Is all ancillary information (helicity, Faraday Cup, …) available and understood?*

(a) Yes – There are several ongoing analyses but plans for the leading candidates for the first publication of RG-K have not been discussed.

(b) Yes – The skim list is in place. Eight skims are proposed for FTOn configuration (7.5 GeV beam energy) and four skims for FTOff configuration (6.5 GeV). The total volume of skims is ~80% of full DST output.

(c) Yes – The preliminary analysis results for few physics reactions are encouraging.

(d) No – Beam spin asymmetries for DVCS and DVMP (π^0, η) and beam-recoil Λ transferred polarization have been shown, but not beam charge normalized yields.

Comments:

Full QA for high level physics quantities (yields of different final states, beam spin asymmetries …) should be demonstrated, see below.

Recommendations:

None.

Charge #4: *Are data processing tools that will be used adequate for the proposed processing task? Is the data management plan (staging area, tape destination, directory structure, logs, …) defined and appropriate given the available resources? Is the estimate of resources needed to complete the task sound?*

Yes – Data processing tools, the data management plan, and the required resources are adequate for processing of this data set.

Comment:

It will take about 1 month to complete the pass1 processing and about 40 TB of recon DSTs and about 32 TB of skimmed files for physics analysis will be produced.

Recommendations:

The total amount of skimmed information is 80% of the DSTs. The fraction of data in some of skims is more than 1/3rd of DSTs. We find this to be an inefficient way of doing analysis and using unnecessary disk space. Note, RG-A&B having much larger data sets have a much smaller footprint in terms of disk space for physics analysis. We recommend that the run group should determine proper physics skims, using more custom “wagons”, and follow the CLAS12 software group and CCC guidelines (see above).

Charge #5: *What are the plans for monitoring the quality of the cooking output and identify/correct failures?*

Monitoring tools used by RG-A & B will be deployed to perform QA of the processed data. It has been tested for a few runs but more runs must be included to verify the quality of processed data.

Comment:

The monitoring of normalized yields and beam spin asymmetries have been shown (web portal of the monDB) only for four runs.

Recommendations:

None.

Charge #6: *Is the manpower adequate for the proposed data processing?*

Yes – The available manpower is adequate for completing the processing and monitoring of the pass1 processing of the RG-K data set.

Recommendations:

None.