Reply To:

Readiness Review Report for Pass-1 Processing of the CLAS12 RG-K Dataset

June 19 2020

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, comments, and recommendations. RG-K has done a great job with calibrations and understanding the dataset. The quality of the calibration for all detectors is adequate for the pass-1 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 pass-1 processing of their dataset. 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 datasets 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.

Reply:

Two reactions from on-going analysis of RG-K data may be put on fast track to get publication-quality results: (1) BSA from DVCS on proton target at 6.5 GeV and 7.5 GeV; (2) beam-recoil Λ polarization from $K^+\Lambda$ electroproduction on the proton.

1) BSA from DVCS on proton target

Using 9 runs at 7.5 GeV corresponding to 1.64 mC and 9 runs at 6.5 GeV corresponding to 2.27 mC (available from the currently processed data samples), we estimated the number of exclusive events available on tape for publication. We expect a total of about 200k exclusive DVCS events with the proton detected in the FD, and an additional 1220k with the proton detected in the CD. These represent a factor 3 improvement in statistics over the originally published 6 GeV BSA data, **and clearly represents a publishable dataset.**

The analysis of protons in the FD has already been demonstrated. For protons in the CD, the quality of the CVT alignment will not allow us to use momentum information. However, this information is not necessary to reconstruct the exclusive reaction kinematics. The CD protons without CVT tracks have information equivalent to the neutrons in RG-B. It is possible to reconstruct their momentum from the CTOF information, and to compare this with the momentum transfer calculated between the virtual photon and the real photon. This strategy was also used by the Hall A DVCS publications. We **will** produce simulations to further demonstrate this point. These simulations will require some radiative effects to be included.

Finally the beam asymmetry observable is not sensitive to the possible variations of the detection efficiency of the proton in the central region, since it cancels out in the evaluation of the difference over the sum of events with opposite helicities. These data will also form the basis for the PhD thesis of Joshua Artem Tan from KNU.

2) Beam-recoil Λ polarization from $K^+\Lambda$ electroproduction on the proton

Based on the same sample of 9 runs at 7.5 GeV and 9 runs at 6.5 GeV, detailed statistical projections have been completed for the K^+Y analysis at both beam energies for the entire fall 2018 RG-K beam time for both the $e'K^+$ and $e'K^+p$ event topologies (with the electron in the ECAL). See the RG-K KY analysis wikipage at:

https://clasweb.jlab.org/wiki/index.php/Run_Group_K#tab=KY_Analysis_Work

The relevant Q^2 range for these outbending torus polarity data spans from 0.3 to 6.0 GeV^2. The statistics from the available RG-K data is $\approx 3\times$ more than the CLAS e1f dataset in the same Q^2 range (1.4–3.8 GeV^2), which represent the dominant part of the available world data for these final states in the nucleon resonance region. In the low Q^2 range from 0.3-1.4 GeV^2, covered by the acceptance of CLAS12 in the RG-K data, the statistics is a factor of $13\times$ the published data from CLAS, even just selectin the topology with the K^+ in the FD.

The short-term plans are to focus on the extraction of the beam-recoil Λ polarization in the CLAS12 acceptance with the reaction products (K^+ and p) in the FD (the dominant reaction topology). This analysis will be extended next to the topology with the K^+ in the CD and the p in the FD (the second most important topology) after implementation of the momentum correction from the PCORR task force. For the polarization analysis the proton angle measurement is critical and will still be provided by the accurate measurement in the FD.

The momentum smearing of the electron is the dominant contribution in all topologies for the hadrons going forward or backward. As long as the lineshapes for the Λ and Σ^0 describe the data (after the Gaussian convolution procedure), and after the efficiency function for the CD is derived based on the current reconstruction, then the polarization analysis can be extended in the CM angle of the K^+.

The expectation is that this analysis can proceed toward publication within 6-8 months after the pass-1 data processing is complete.

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.

Reply:

Since the pass-1 Readiness Review, we have revisited the list of data skims that we want to have in place during the pass-1 processing. The current list of skims is detailed in our answer to review charge #4 below. As shown in Fig. 1, the present total size of the full set of skims that have been developed is now estimated to be about 5% of the DST size for 7.5 GeV runs and 6.5% for the 6.5 GeV runs.

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.

Reply:

We thank the Review Committee for the positive comment.

Charge #2: (a) Is data quality as a function of run number or time for the dataset that is proposed for pass-1 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.

Reply:

As soon as the reconstructed efficiencies are assessed and the beam background merging software for the data and MC are available, the comparison will be studied at both 6.5 GeV and 7.5 GeV beam energies.

Charge #3: (a) Are analysis plans for the dataset 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 FT-On configuration (7.5 GeV beam energy) and four skims for FT-Off configuration (6.5 GeV). The total volume of skims is $\simeq 80\%$ 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.

Recommendation:

None.

Reply:

The QA for high-level physics quantities was shown only for four fully cooked runs during the pass-1 Readiness Review. Timelines for the full data range have now been produced using the pass-0 outputs (for which 10 files from each run have been processed).
The timelines are now available at the link: https://clas12mon.jlab.org/dilks/rgk_pass0/timeline/ and on the RG-K pass-1 Readiness Review wikipage:
https://clasweb.jlab.org/wiki/index.php/Run_Group_K#tab=Pass1_Review_Documents
under the section "Additional Comments and Links". All of the quantities are stable for the entire RG-K run range, compatible with the available statistics.

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 dataset.

Comments:

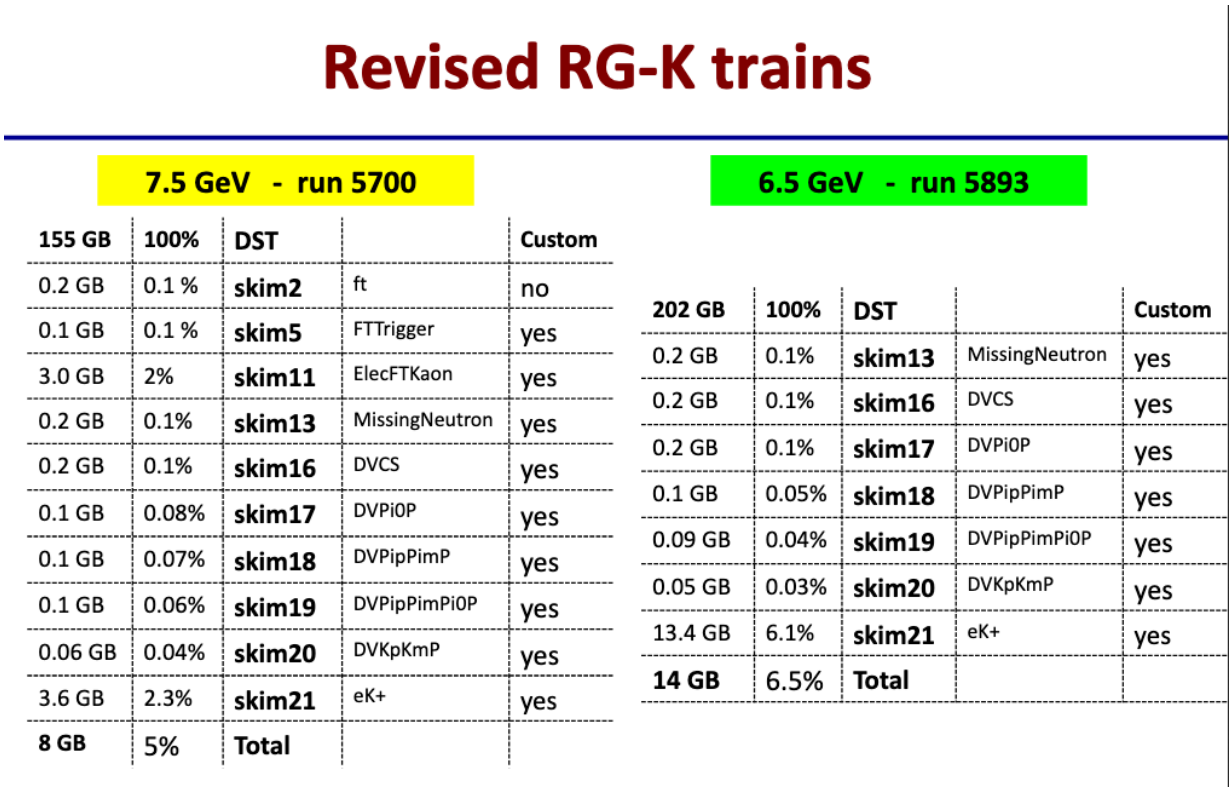
It will take about 1 month to complete the pass-1 processing and about 40TB of recon DSTs and about 32 TB of skimmed files for physics analysis will be produced.

Recommendation:

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 datasets 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).

Reply:

A new train scheme has been developed and tested in collaboration with the Software Group, producing a total fraction of skimmed output files that is about 5% of the DSTs for 7.5 GeV runs and 6.5% for 6.5 GeV runs. Custom wagons have been designed to isolate the event topologies and the kinematical region of interest for the ongoing data analyses.
The new train scheme is shown in Fig. 1 and is available on the RG-K pass-1 Readiness Review wikipage at the link: https://clasweb.jlab.org/wiki/images/1/10/RGK_trains.pdf.
Should additional skims become necessary, additional wagons can be defined and new trains can be run using the DST files from pass-1.



Revised RG-K trains

7.5 GeV - run 5700

155 GB	100%	DST		Custom
0.2 GB	0.1 %	skim2	ft	no
0.1 GB	0.1 %	skim5	FTTrigger	yes
3.0 GB	2%	skim11	ElecFTKaon	yes
0.2 GB	0.1%	skim13	MissingNeutron	yes
0.2 GB	0.1%	skim16	DVCS	yes
0.1 GB	0.08%	skim17	DVPi0P	yes
0.1 GB	0.07%	skim18	DVPipPimP	yes
0.1 GB	0.06%	skim19	DVPipPimPi0P	yes
0.06 GB	0.04%	skim20	DVKpKmP	yes
3.6 GB	2.3%	skim21	eK+	yes
8 GB	5%	Total		

6.5 GeV - run 5893

202 GB	100%	DST		Custom
0.2 GB	0.1%	skim13	MissingNeutron	yes
0.2 GB	0.1%	skim16	DVCS	yes
0.2 GB	0.1%	skim17	DVPi0P	yes
0.1 GB	0.05%	skim18	DVPipPimP	yes
0.09 GB	0.04%	skim19	DVPipPimPi0P	yes
0.05 GB	0.03%	skim20	DVKpKmP	yes
13.4 GB	6.1%	skim21	eK+	yes
14 GB	6.5%	Total		

Figure 1: New RG-K train scheme, including new defined custom wagons. The total data volume from the trains output is about 5% of the total DST data volume for 7.5 GeV runs and 6.5% for 6.5 GeV runs.

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.

Reply:

Full QA monitoring of normalized yields is now available for the full run range from the pass-0 file subset. The timelines are now available at the link: https://clas12mon.jlab.org/dilks/rgk_pass0/timeline/ and on the RG-K pass-1 Readiness Review wikipage: https://clasweb.jlab.org/wiki/index.php/Run_Group_K#tab=Pass1_Review_Documents

under the section "Additional Comments and Links". All of the quantities are stable for the entire RG-K run range, compatible with the available statistics.

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 pass-1 processing of the RG-K dataset.

Recommendations:

None.

Reply:

We thank the Review Committee foe their constructive comments and their support.