

Report of the Ad-Hoc Committee on Common Tools

Committee Members

R. De Vita (Chair)

D. Glazier

S. Kuhn

C. Smith

V. Ziegler

This report addresses the questions posed in the charge, based on information the committee members already had as CLAS collaborators, on information obtained from CLAS12 software experts and information collected from the spokespersons of the Run Group A experiments who are directly involved in the so-called “first experiment”.

Charge

1. What reaction channels should we focus on now for CLAS12 validation, calibration, and publication of the first papers? The identified reaction channels should allow defining and validating the procedure to extract novel physics observables, determining for example fiducial cuts, kinematic corrections, cross section normalization, etc.

2. Where should we draw the line for the tasks to perform with the Common Tools over the next fourteen months preceding the engineering run?

More specifically:

2.1 In the long term, what parts of the analysis procedure applied to the reconstructed data to extract physics observables from CLAS12 data should be standardized and become part of a “Common Toolset” for the collaboration to use?

2.2 How these common procedures should be integrated into software tools?

2.3 What should be the priorities in the development of Common Tools over the next fourteen months preceding the engineering run? This requires some estimate of the number of developers necessary and a realistic timeline.

3. What bottlenecks exist, hardware or software, that must be overcome before the start of the engineering run?

Definitions

CALIBRATION: collection of algorithms and procedures to determine the parameters that describe the detector response and are necessary to convert digitized signals recorded by the DAQ to energy, position and time.

EVENT RECONSTRUCTION: collection of algorithms and procedures to convert detector responses to particle 4-vectors, including PID, energy loss, vertex time (RF) corrections, helicity information, based on the best available knowledge of the detector calibration parameters, geometry and magnetic field.

EVENT SELECTION: collection of algorithms and procedures to select events from the event reconstruction output and skim the ones of interest for the analysis of a specific reaction, applying relevant kinematic corrections, e.g. momentum corrections, and cuts, e.g. fiducial cuts and quality checks selecting golden runs or files. The final output should be DST files with fully corrected 4-vectors for the final state of interest with the relevant luminosity information for absolute normalization.

SIGNAL SELECTION: collections of algorithms and procedures necessary to separate the reaction of interest from background events. These may also include refinement of the PID provided by the event reconstruction, kinematic fitting, background subtraction by for example Q-factor or sWeight or cuts based approaches including missing mass, neural networks and boosted decision trees.

PHYSICS ANALYSIS: collections of algorithms and procedures necessary to extract the physics observables of interest from the selected events. This includes χ^2 fits to histograms or event-by-event maximum likelihood approaches. The quantities extracted may be model independent, such as cross sections, asymmetries, or spin density matrix elements or directly apply some theoretical model such as amplitude analysis for meson photoproduction.

COMMON TOOLS: collections of software tools (not restricted to Java language software) that implement algorithms and procedures functional to specific tasks of the calibration, event reconstruction, event selection and physics analysis, independently of the specific experiment or physics reaction, and shared by the collaboration.

COATJAVA: collections of Java classes that permit functionalities of various tasks to be performed. They are organized in various packages in the JeffersonLab git code repository under the clas12rec main directory.

Charge #1

1. What reaction channels should we focus on now for CLAS12 validation, calibration, and publication of the first papers? The identified reaction channels should allow defining and validating the procedure to extract novel physics observables, determining for example fiducial cuts, kinematic corrections, cross section normalization, etc.

First reactions to be analyzed for validation of the detector calibration and reconstruction are inclusive electron scattering, single and double pion production. The measurement of the inclusive electron cross section will allow us to verify absolute normalization and efficiency evaluation. Single and double pion production would allow us to validate reconstruction of multi-particle events, test the selection of exclusive final states via missing mass and determine mass resolutions. Double pion production would permit to determine kinematic corrections.

These reactions could also give access to observables that could be the subject of a speedy publication such as the F2 structure function in poorly studied kinematics regions or N-pi beam asymmetries. The final choice will be made after completion of full simulation and reconstruction of the selected reactions and considering the scientific impact of the corresponding publication.

Charge #2, #2.1

2. Where should we draw the line for the tasks to perform with the Common Tools over the next fourteen months preceding the engineering run?

Common tools should provide at least the functionalities necessary to complete calibration, event reconstruction and event selection as defined in the previous section of this report.

More specifically:

2.1 In the long term, what parts of the analysis procedure applied to the reconstructed data to extract physics observables from CLAS12 data should be standardized and become part of a “Common Toolset” for the collaboration to use?

All tasks from calibration to event selection should be part of the Common Tools as stated above. In addition, in the long term, tasks that are part of the physics analysis, such as refined PID, kinematic fitting, background subtraction tools, could also become part of the Common Tools and be shared among collaborators. This extended level of standardization would ensure higher reliability and validation of the algorithms and related software packages resulting in a reduction of the analysis and review time.

Charge #2.2

2.2 How these common procedures should be integrated into software tools?

Given the level of development of COATJAVA, Common Tools for the first experiment, providing functionalities needed for calibration, event reconstruction and event selection, should be integrated in this package, possibly exploiting the work that was done for the “data mining” project.

The output of the COATJAVA event selection phase should be made available in evio, hipo and root formats. It should be possible to access, on request, lower level detector banks and/or fully reconstructed and corrected tracks with for example, 4-vectors, vertices, PID info, timing and covariance matrices.

While it may be difficult to converge, in the time frame between now and the first experiment, on particular scheme to implement tools for signal selection and physics analysis, strict guidelines and checks to ensure validity and encourage the use of common approaches should be enforced. As such we suggest any algorithm or technique should be pre-reviewed by an Analysis Review Group prior to any final Analysis Note to create a speedier final review process. Any individual analysis could then just be required to check the technique for their reaction through a simulated validation. Anyone implementing a new technique should be encouraged to make common software available to the rest of the collaboration, thereby building up our stock of innovative and reliable Common Tools.

Charge #2.3

2.3 What should be the priorities in the development of Common Tools over the next fourteen months preceding the engineering run? This requires some estimate of the number of developers necessary and a realistic timeline.

The first priority should be completing tools needed for calibration and event reconstruction. This should be done while maintaining a stable version of COATJAVA that allows users to become proficient with these tools.

The second priority should be developing tools necessary for the event selection phase, starting from fiducial cuts, kinematic corrections, event and reaction skimming. The development of such tools should be guided a new group of “wise experts” (see Action #4). These algorithms and methods should then be implemented in COATJAVA in collaboration with the software group. The manpower and timeline needed for the implementation was estimated by the software group to be 2 FTEs and one year.

Charge #3

3. What bottlenecks exist, hardware or software, that must be overcome before the start of the engineering run?

We have identified the following areas of possible improvements:

- increase the number of CLAS I 2 Collaboration members familiar with COATJAVA;*
- increase the manpower dedicated to software development;*
- strengthen the coordination of Run Group A leadership in the analysis organization and analysis tools development;*
- extend common tools to cover event selection.*

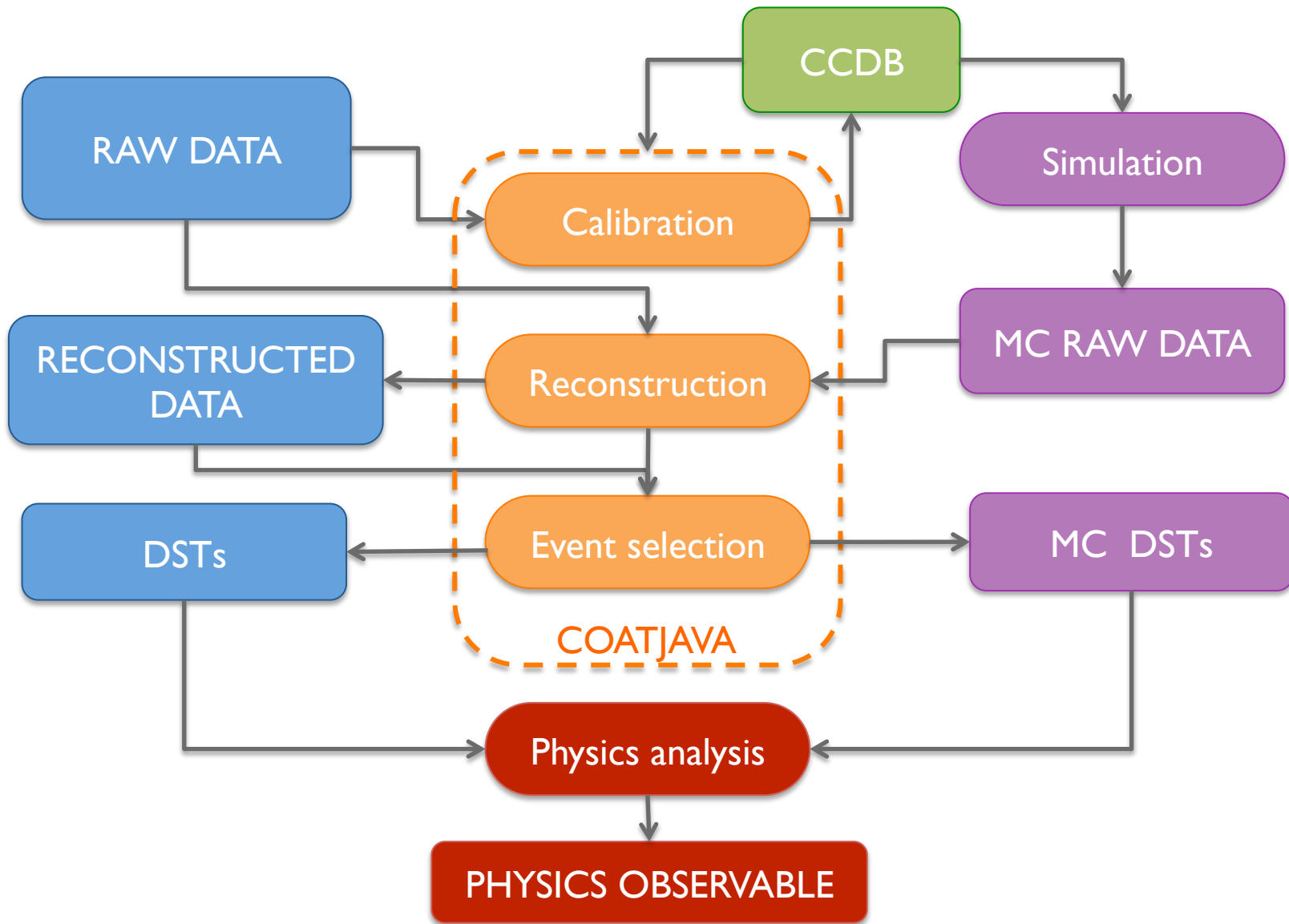
Actions

1. Complete the COATJAVA reconstruction and calibration tools. This must be done keeping future code developments backward compatible with the present version 3.0.
2. Extend COATJAVA to incorporate Event Selection, as per our definition.
3. Create a new group of “wise experts” (Analysis Group) who will guide the development of algorithms for momentum corrections, PID, background subtraction, fiducial cuts and other corrections, exploiting the expertise accumulated with analysis of CLAS data. Algorithms developed by this group should be reviewed by an analysis review committee (see 4). Upon approval they should be considered “standard”, requiring no further review when applied to specific analysis and only a short reference in future analysis notes.
4. Create a “First experiment analysis review committee” drawn from the relevant PWGs before the experiment even begins, and have that committee review each individual tool/algorithm/method used to arrive at Physics results, as they are being developed (by the RGA collaborators or the “experts” in 3). If this approach will prove to be effective for the First Experiment, it should be extended to any future CLAS12 experiment.

Actions

5. Set up a work plan defined by Run Group A to determine the optimal run conditions and be ready for data analysis prior the beginning of the first experiment. This includes:
 - setting up event generators,
 - running full simulation and reconstruction of a benchmark reaction for each Run Group A proposal,
 - becoming familiar with COATJAVA, GEMC and CED or other COMMON TOOLS that may become available,
 - determining the optimal detector, target, magnets, trigger configuration for data taking,
 - define a clear management structure, internal to the group, and distribute tasks of common interest with a clear timeline and milestones.
6. Utilize the CLAS12 software discussion forum for the exchange of information between software users; encourage the use of the clas12-software mailing list; organize software workshops during Collaboration meetings, and identify collaborators who will update documentation and tutorials on the software wiki.

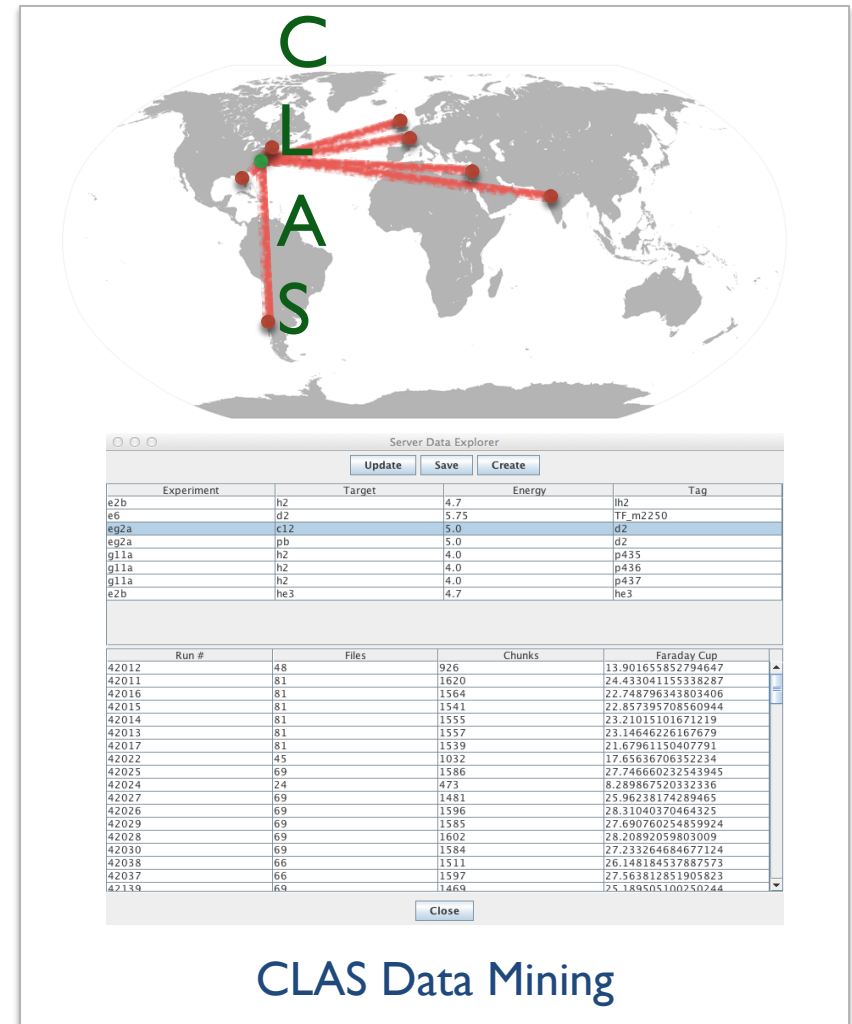
Data Analysis Scheme



Event selection tool

Collection of algorithms and procedures to select events from reconstruction output and provide DSTs for specific final states:

- Select golden runs and files
- Select events for specific final state
- Apply:
 - kinematic corrections
 - fiducial cuts
 - ...
- Output DST files with:
 - fully corrected 4-vectors for physics analysis
 - detector related info for refinement of PID and signal selection
 - luminosity and helicity related info
- Implement file tagging for easy data handling and distribution inspired by the CLAS data mining project



The image shows a world map with red lines connecting points labeled C, L, A, and S. Below the map is a screenshot of a 'Server Data Explorer' window. The window displays a table with columns for Experiment, Target, Energy, and Tag. Below this is another table with columns for Run #, Files, Chunks, and Faraday Cup. The 'CLAS Data Mining' logo is visible at the bottom of the screenshot.

Experiment	Target	Energy	Tag
e2b	h2	4.7	lh2
e6	d2	5.75	TF_m2250
eg2a	c12	5.0	d2
eg2a	pb	5.0	d2
g11a	h2	4.0	p435
g11a	h2	4.0	p436
g11a	h2	4.0	p437
e2b	he3	4.7	he3

Run #	Files	Chunks	Faraday Cup
42012	48	926	13.901655852794647
42011	81	1620	24.433041155338287
42016	81	1564	22.748796343803406
42015	81	1541	22.857395708560944
42014	81	1555	23.21015101671219
42013	81	1557	23.14646226167679
42017	81	1539	21.67961150407791
42022	45	1032	17.65636706352234
42025	69	1586	27.746660232543945
42024	24	473	8.28986752032336
42027	69	1481	25.96238174289465
42026	69	1596	28.31040370464325
42029	69	1585	27.690760254859924
42028	69	1602	28.20892059803009
42030	69	1584	27.23264684677124
42038	66	1511	26.148184537887573
42037	66	1597	27.563812851905823
42139	69	1469	25.189505100250244

Analysis Organization

