

CLAS12 software session

November 13, 2018



Latest developments...

Very intense 4-month period:

- Major improvement of reconstruction speed (x3)
- Important upgrades to CLARA to improve performances when using full nodes
- Several fixes in algorithms and geometries leading to significantly better reconstruction performances
- Increased flexibility in reconstruction configuration
- Important upgrades to handle new bit-packed fADC and MM raw data
- Implementation of analysis trains
- Development of new simulation software distribution for easy-deployment on offsite computing resources
- Analysis tools weekly meetings started

Planned vs. Achieved

From the July meeting presentation:

During the next 6 months:

- Support to preparations for the fall-winter data taking
- Improvement/completion of reconstruction software
- Tuning/speed-up/portability of simulation software
- Development of analysis framework and tools
- Support to the First Experiment data processing (calibration, reconstruction, analysis, ...)
- Documentation

Planned vs. Achieved

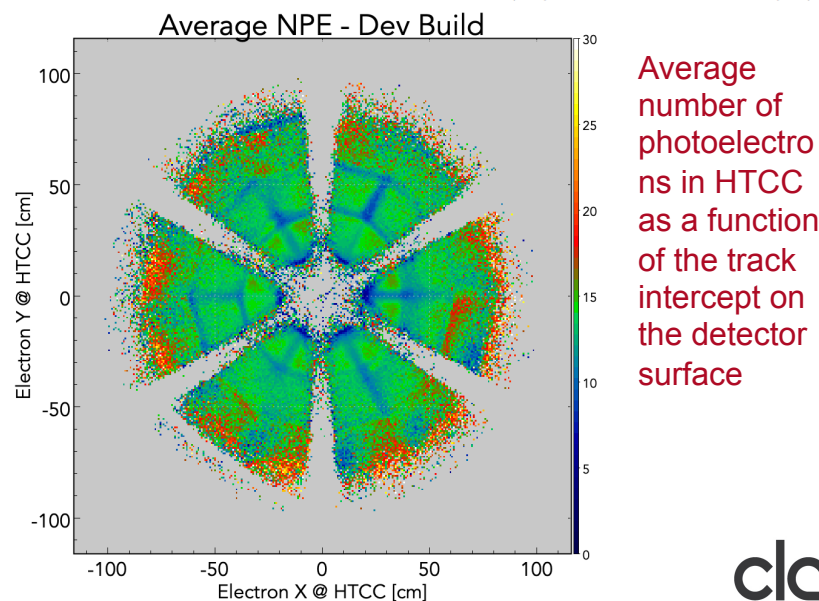
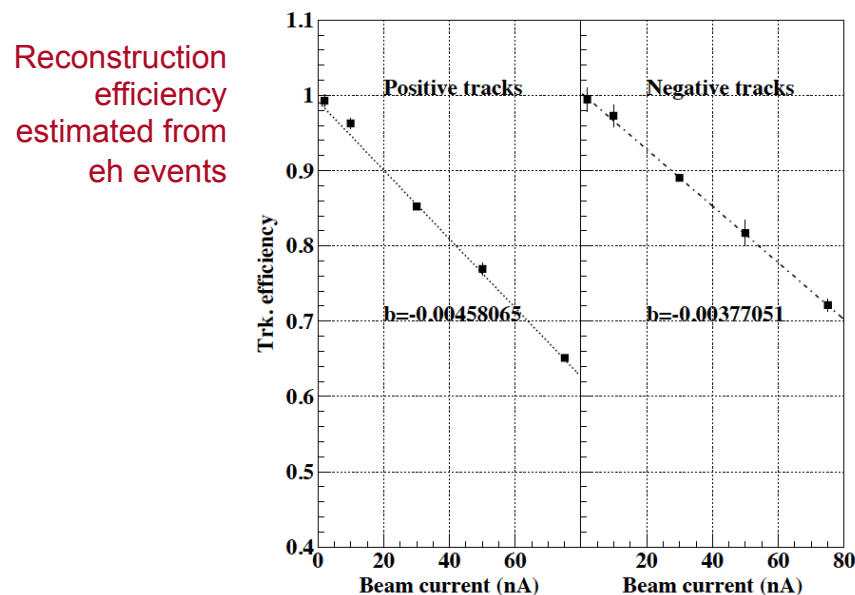
From the July meeting presentation:

During the next 6 months:

- Support to preparations for the fall-winter data taking
✓ Road dictionary, bit-packing, ...
- Improvement/completion of reconstruction software
~ Certainly improved but not yet complete
- Tuning/speed-up/portability of simulation software
✓ Presentation today and demonstration on Friday
- Development of analysis framework and tools
✓ Trains runs successfully for DNP analysis, tools development continues
- Support to the First Experiment data processing (calibration, reconstruction, analysis, ...)
✓ Ongoing
- Documentation
✗ Still inadequate

Improving reconstruction

- Each of the reconstruction packages has been debugged and tested but we can reasonably expect to have still a lot of room for improvements:
 - more efficient or performing algorithms
 - more robust, fast and maintainable software implementations
 - ...and keep searching for and getting rid of bugs
- How?
 - Need input from ongoing analyses of both simulated and real data to understand what the possible problems are and what the priorities should be
- When? Now!
 - Needed to define work plan for next months in reconstruction development to be ready for pass1 processing
 - Need quantitative information from analyzers and availability to repeat the studies when the source of the problem is found and fixed



Documentation

- Still behind in updating/ completing documentation
- List of links from the CLAS12 page recently cleaned up to remove obsolete information but still to be better organized
- More coatjava/reconstruction documentation coming
- New forum at:
<https://clas12.discoursehosting.net>
- Help from Collaborators:
 - Use the forum to document issues and solution
 - Get involve in writing documentation

The image displays four overlapping screenshots of CLAS12 documentation resources:

- Top Left:** A 3D CAD model of the CLAS12 detector components, including the Torus, HTCC, SVT, and CTOP. A label 'Baseline Equipment' points to a list of components.
- Top Right:** A list of links under the heading 'CLAS12 Databases' and 'CLAS12 Reviews'. The databases include Calibration Constant Database - CCDB, Run Control Database - RCDB, and CLAS Equipment Inventory Database. The reviews include Torus Cooledown Review, Ancillary Equipment Review, Torus Power-up Review, CLAS12 Equipment Review, Solenoid Cool-down and Power-up Review, and Ready for Science Review.
- Bottom Left:** A screenshot of the Discourse forum at <https://clas12.discoursehosting.net>. It shows a table of topics with columns for Topic, Category, Users, Replies, and Activity.
- Bottom Right:** A list of links under the heading 'CLAS12 Documentation' and 'CLAS12 Upgrades'. The documentation includes CLAS12 Wiki, CLAS12 Operations Documentation, and First Experiment Document Database. The upgrades include Luminosity Increase and Drift Chamber Readout Boards (DCRB).

Topic	Category	Users	Replies	Activity
F1D integral for convolution fit	Data Plotting	F	5	12d
Bos2hipo issues	Offline Software	G	1	20d
GROOT: Logarithmic Scale Issue	Data Plotting	G	1	26d
Histogram titles in groot	Data Plotting	T	5	29d
GEMC on OSX 10.13.6	Simulations	M	10	Sep 28
NaN photons in DSTs	Offline Software	B	18	Sep 26
This is a complete sentence		B	4	Sep 21
Test 15 char minimum	Site Feedback	G	3	Sep 20
Welcome to Discourse		G	0	Sep 19

Computing resources

- New estimates of computing resources to process CLAS12 data based on improved reconstruction processing speed:
 - Reconstruction of 2018-2020 data expected to take approximately 14 months on our current farm allocation, not including tape access or inefficiencies
 - Factor two contingency to be “used well” (currently we have a 9 months backlog)
- Disk space currently needs to be expanded to support processing of multiple data sets:
 - Estimated data volumes from 2018-2020:
 - DSTs: 340 TB (to cache/tape)
 - Skims: 78 TB (2 versions, each <10% of DSTs)
 - Calibration and data monitoring: 150 TB assuming two data sets actively being calibrated in parallel
 - File staging: 100 TB
 - Users: 50 TB
 - Total work area needed: 330 TB with respect to 130 TB currently available
 - Request for increase presented to SciComp
- Estimates being refined in view of JLAB 12 GeV software and computing review scheduled for November 27-28 2018

Computing resources for processing on 2018/2019/2020 CLAS12 data

Batch farm usage

Assumptions

Data processing will involve primarily farm18 (87 nodes, 80 job slots/node), farm16 (46 nodes, 72 job slots/node), farm14 (83 nodes, 48 job slots/node). The total number of nodes and job slots is 216 and 14256, respectively. With a fair-share of 45%, CLAS12 can utilize about 92 nodes and 6415 job slots.

Estimate for RG-A-Spring run

The following estimate assumes the data set consists of 500k raw files (1 PB) with ~45k events each and uses one file from run 4013 as example.

- **Decoding:** the time to decode one file on a clara node is 12 min. This is likely less than when running on the farm occupying a full node. Assuming a factor two larger time, the estimate for the decoding of the full data set is 8400 core*days that would be less than two days with the fair share indicated above. This does not account for the time to retrieve files from tape, which will be largely dominant. Estimating this time for the Spring data will require more studies, since the raw data files are distributed over a large number of tapes: this is partially due to the way the files were initially stored on tape but has been magnified with the migration from LTO6 to LTO8 tapes. As an example, the 2078 files of run 4013 are currently spread over 9 tape volumes. Presently, the option that is being considered to speed-up the decoding process is to process files tape-by-tape instead of run-by-run, possibly staging the output on disk to save the hipo files from the same run on tape in sequential order.
- **Reconstruction:** according to the latest benchmark, the reconstruction rate per node, computed as the average on farm14, farm16 and farm18 nodes weighted by the corresponding node number, is of the order of 92 Hz. Assuming the fair share indicated above, this results in an overall reconstruction rate of about 9 kHz. At this rate, reconstruction of the spring data (22.5 billion events) would require about 30 days. This does not include tape access, failures etc. More precise numbers will be obtained with the planned tests on Slurm: for now, assuming a safety factor of 2, we could estimate a total of 60 days for the reconstruction of the RG-A Spring data.

Estimates for subsequent runs (RG-A-Fall, RG-K, RG-B, RG-A-2019, RG-F)

Based on the beam time allocation for 2018 and 2020, we can expect to accumulate the following statistics:

- RG-A-Fall (59 days at 13 kHz trigger rate and 50% efficiency = 33 billion events)
- RG-K (18 days at 13 kHz trigger rate and 50% efficiency = 10 billion events)

https://www.jlab.org/Hall-B/secure/claschair/software/CLAS12_computing_2018-2020.pdf

Software review

- JLAB 12 GeV software and computing review scheduled for November 27-28 2018
- Hall B/CLAS12 scheduled to give two talks:
 - Overview and Progress (Stepan): schedule for next 5 years, path to publication, computing requirements, etc...
 - Deep Dive (Raffaella, Graham): software status, software organization, recent successes, etc...

Work plan for next months

- Detailed list of software task being reviewed on a weekly basis
- Main tasks for next months:
 - Upgrade to HIPO4
 - Continue reconstruction performance assessment and improvement
 - Implement geometry manager for coherent and efficient loading of all CLAS12 detector geometries
 - Reconstruction monitoring and online reconstruction
 - Extend functionalities of analysis trains and tools
 - Continue support for running simulation offsite
 - Improve data processing workflow
 - Continue support to ongoing or future data taking

Implement unit tests for all packages	N. Harrison	Medium	
Extend validation tests	N. Harrison	Medium	

Alignment and Geometry:	Lead	Priority	Completion
Redo SVT geometry		Medium	Fall 2018
Extend current geometry services adding the ability to move and rotate detectors	Andrey/Brandon	High	September 30

Completed
In progress
To be started
New
Late

DAQ and trigger:	Lead	Priority	Completion
Tracking trigger	Sergey/Ben	High	July 31
Generate roads dictionary	Veronique	High	August 10
Develop tracking trigger monitoring		High	August 15
Geometry matching	Sergey/Ben	High	July 31
Writing of full trigger information including level 2	Sergey	Medium	August
Compressed-bit-packed IADC output	Sergey	High	July 31
Cam TDC INL		Medium	August
Bit-packed FMT output	Saclay/Sergey	Medium	August
Switch to large evio files	Sergey	Medium	Waiting on evio6
Write hipo instead of evio	Gagik/Sergey	Low	

Simulations:	Lead	Priority	Completion
Set default field map to new 4/24 map	Mauri	done	June 1st
Rotate solenoid field	Mauri	done	June 20th
Check fast mc	Zhiwen	done	
Implement capability of handling non symmetric field map	Mauri	Low	
Production cut test	Mauri/VRG-A	High	Mid July
Speed up: production cuts for active and passive material, goal should be 500 ms/event	Mauri	High	Fall 2018
Support for alternative, run dependent, geometries	Mauri	Medium	Fall 2018
Trigger simulations		Medium	2019
Hipo output	Mauri	Low	
Support porting GENIE to outside farms	Mauri	High	Continuous

Reconstruction:	Lead	Priority	Completion
Use event start time in TBT	Veronique/Latif	done	June 30
CVT: suppress fake/duplicate tracks	Francisco	done	June 8
Magnetic field map handling	Dave	done	June 8
Complete FMT and use information in tracking	Veronique	done	June 15
Swimming optimization	Dave	High	June 23
Tracking speed-up	Veronique	Medium	Fall 2018
Algorithmic improvements to tracking	Veronique	Low	
Roads	Veronique/Dave	Medium	
Validation of CDC distortions due to gravity	Veronique/Raffaella	Medium	Fall 2018
FTOP create TB and HB banks	Veronique	Medium	Oct. 31
ECAL fix v-v views and moments	Gagik	Medium	Oct. 31
ECAL index from strip to cluster			
Other existing packages: TOF	Raffaella/Veronique	Medium	ongoing
Finalize LTCC and CND	Mauri/Rong/Nathan	Medium	ongoing
Missing packages: FTT, RICH		Low	
Geometry manager	Gagik	High	Nov. 30
Reconstruction monitoring histograms	Gagik	Medium	Fall 2018
Service to handle true information	Marouen/Francesco	Low	
Code cleanup and optimization	Bruno/Vardan	Medium	ongoing

completion

Completion
End of July
End of July
End of July
Fall 2018

Completion
August

Completion
August 1
August 15
August 15

Completion

Completion

Session agenda

14:00 - 18:20

CLAS Collaboration Meeting - Plenary: CLAS12 Software

Convener: Raffaella De Vita (INFN - Genova)



Location: F113 - Bluejeans connection: <https://bluejeans.com/183745339>

14:00 **Introduction and latest news 15'**

Speaker: Raffaella De Vita (INFN - Genova)

14:15 **CLAS12 reconstruction: code speed-up and latest improvements 25'**

Speaker: Veronique Ziegler (Jefferson Lab)

Material: **Slides**  

14:40 **Event Builder and DSTs update 25'**

Speaker: Nathan Baltzell (Jefferson Lab)

15:05 **Code benchmarking and CLARA updates 20'**

Speaker: Dr. Vardan Gyurjyan (Jefferson Lab)

15:30 **Coffee break 30'**

16:00 **Running simulations offsite: docker containers and CLAS12 distribution 25'**

Speaker: Dr. Maurizio Ungaro (Jefferson Lab)

16:25 **CLAS12 MC at MIT 15'**

Speaker: Richard Milner (MIT)

16:40 **Opportunities for tier-2 computing at Argonne 15'**

Speaker: Dr. Sylvester Joosten (Temple University)

16:55 **Common Tools update and Analysis Tools development 30'**

Speaker: Gagik Gavalian (Jefferson Lab)

17:25 **Analysis trains: status and performances 20'**

Speaker: William Phelps (The George Washington University)

17:45 **MesonEx analysis tools 20'**

Speaker: Dr. Derek Glazier (University of Glasgow)

+Workshop on Friday!