CLAS12 software status update

July 21, 2020

Outline

- Software organization
- Progress since last meeting:
  - Reconstruction
  - Common tools
  - Computing resources and tools
  - Simulations
- Documentation
- Ongoing and planned work

News from the software group

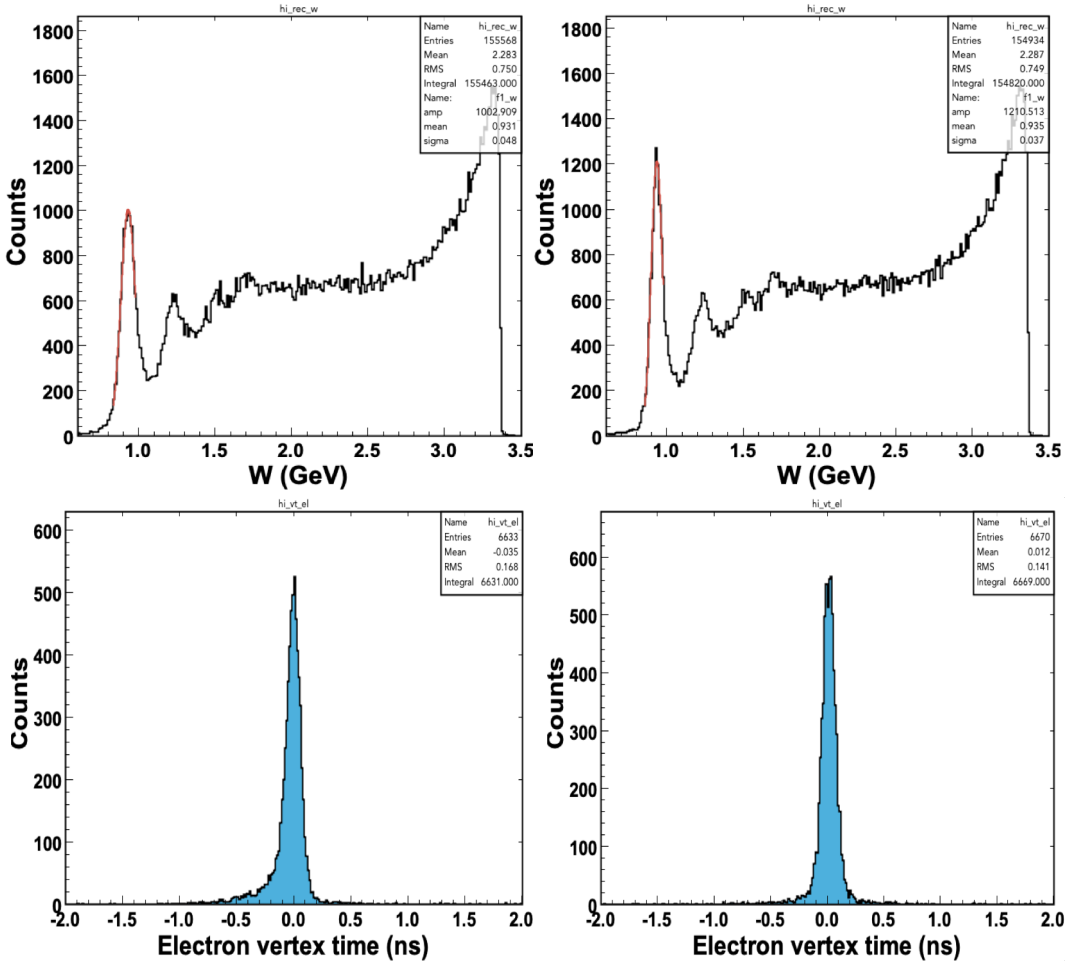
- Rafayel Paremuzyan joined the Hall B software group and will work both on offline and online software
- Roles of the Hall B software group reviewed:
  - Nathan Baltzell: Hall B software coordinator
  - Gagik Gavalian: architect
  - Veronique Ziegler: reconstruction algorithms
  - Maurizio Ungaro: simulations
  - Rafayel Paremuzyan: reconstruction and tools
- CLAS12 software coordinator:
  - work with Hall B coordinator and team to support the needs of the experiment
  - strengthen the role of liaison between the Collaboration and the software experts

Reconstruction progress

…since the last meeting:

- Finalization of software release for RG-A cooking (6.5.3, 6.5.6)
- Preparation of release for RG-B cooking (6.5.8):
  - (C)TOF clustering
  - CND-CTOF veto in EB
  - Updated BAND reconstruction
- RG-F support:
  - RTPC reconstruction
  - FMT reconstruction and alignment
- New run/detector: ALERT
- Ongoing:
  - CVT reconstruction restructuring
  - AI-based forward tracking
  - EB rerun from DSTs

6.3.1 (DNP cooking) | **6.5.3 ("Pass1" cooking)**



Reconstruction progress

…since the last meeting:

- Finalization of software release for RG-A cooking (6.5.3, 6.5.6)
- Preparation of release for RG-B cooking (6.5.8):
  - (C)TOF clustering
  - CND-CTOF veto in EB
  - Updated BAND reconstruction
- RG-F support:
  - RTPC reconstruction
  - FMT reconstruction and alignment
- New run/detector: ALERT
- Ongoing:
  - CVT reconstruction restructuring
  - AI-based forward tracking
  - EB rerun from DSTs

https://github.com/JeffersonLab/clas12-offline-software/releases

6.5.3
72578b8
Verified
Compare

COATJAVA release 6.5.3

Edit

raffaelladevita released this on Mar 6 · 86 commits to master since this release

Compatible with GEMC 4.3.2

- Reconstruction:
  - DC
    - remove bad tracks whose fits diverge and result in NaN values
  - EB
    - drop unnecessary layers from `REC::Traj` to reduce bank size
  - RICH
    - fix quadrant definition
    - fix CCDB electron table reading
    - fix `REC::ringCher` and `REC::hadCher` banks variables
    - fix ray-tracing matching resolution
    - fix bank list for schemas
- Other
  - updated kpp-plot jar

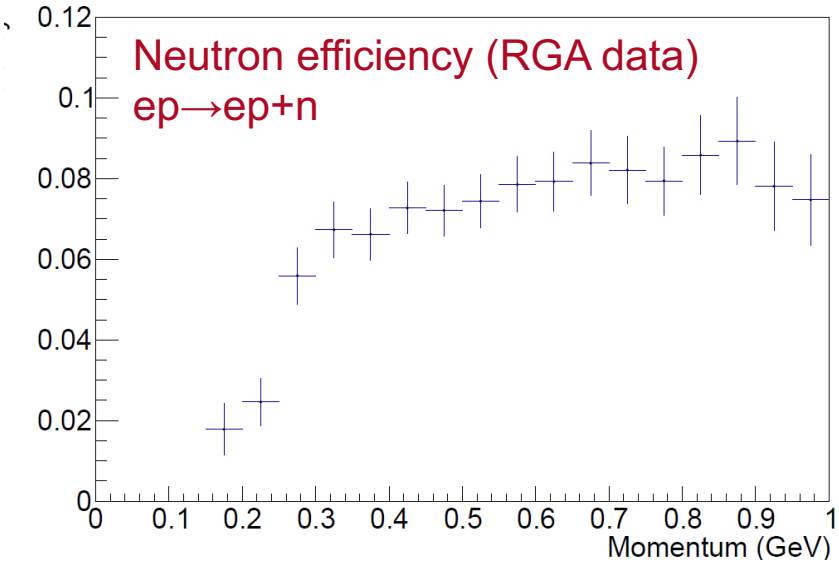
Commits since 6b.5.2
Commits since 6.3.1

Assets 3

CTOF-CND neutral identification

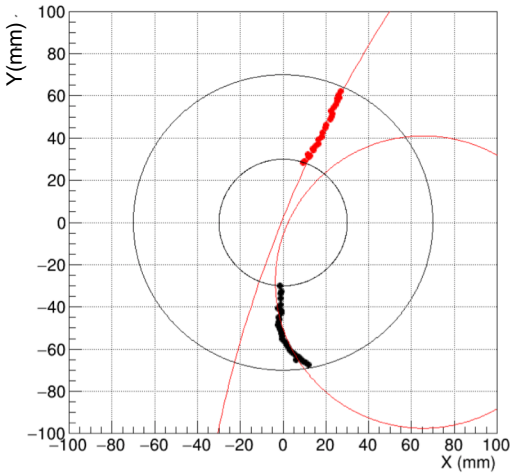
- Motivation:
  - Non-uniform acceptance of the CVT
  - Some of the « neutral » candidates reaching the CND are not neutral
- Requirements:
  - Vetoing charged particles in the CD using only CTOF and CND
  - Minimize contamination of charged particles
  - Minimize loss of neutrons
- Information used from CTOF and CND:
  - Energy deposition, hit multiplicity, layer multiplicity

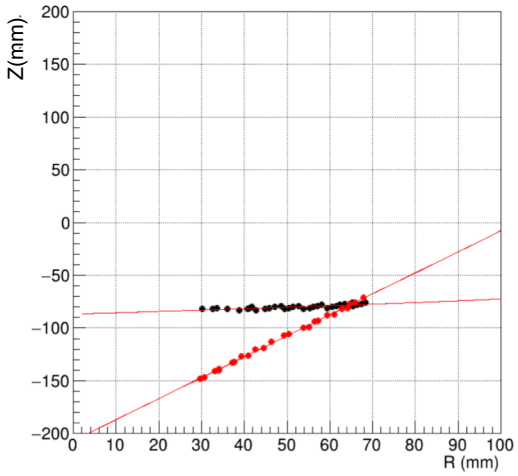
- First version of veto, based on single-particle simulations, implemented in the Event Builder
- Neutron detection efficiency for CND with CD slightly lower (~1%) than before
- Further optimization planned based on data analysis; use of neural networks being investigate

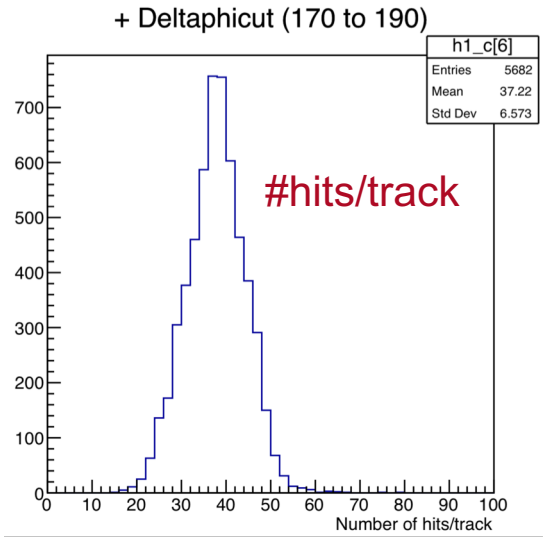


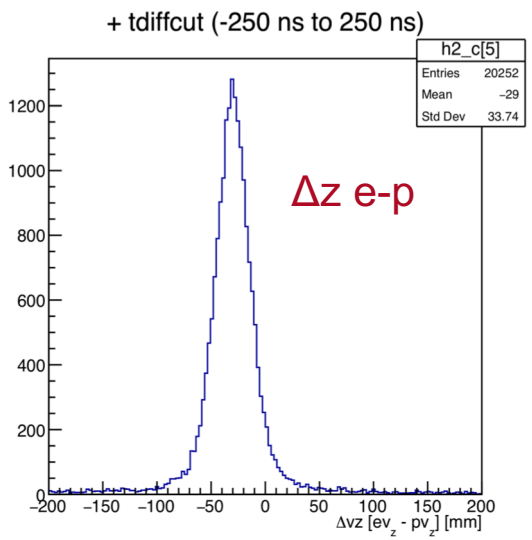
RTPC

- RTPC reconstruction implemented and being exercised on real data
- Ongoing work on:
  - experimenting with new r and phi parameterization
  - removing certain elements of the RTPC reconstruction to address tracking anomalies (broken tracks, short tracks
- Performance example on elastic electrons in coincidence with good proton tracks



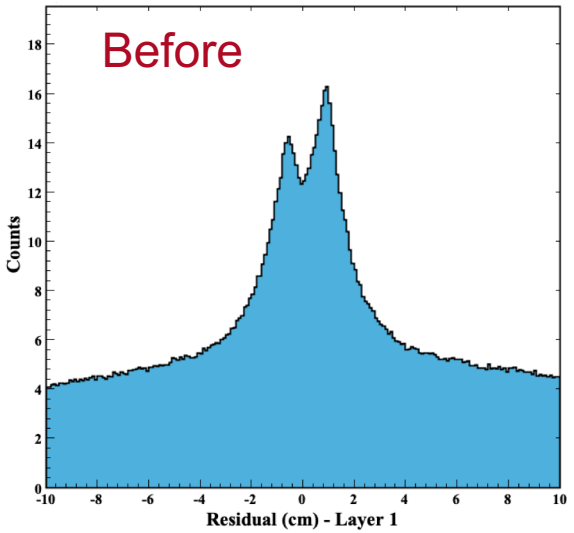


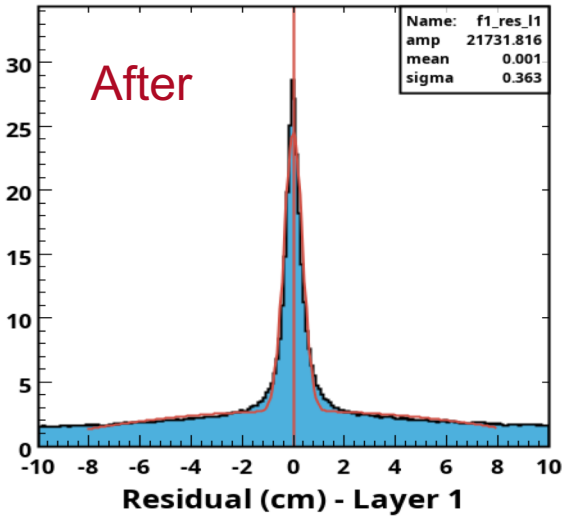




FMT alignment and reconstruction

- Alignment:
  - Performed using low luminosity data from RG-F
  - By heuristically selecting shifts, the residuals between a DC track and an FMT cluster are minimized
  - Shifts and rotations along/around 3 axes are applied: deltaX, deltaY, deltaZ, rotX, rotY, and rotZ
  - Results are currently being asserted, and are available in CCDB
- Reconstruction:
  - Implemented as a second pass, where DC tracks are matched to FMT clusters and refitted
  - Currently being modified to implement alignment information
  - Further studies and improvements to follow
- Dedicated effort of the UTFSM group: see Bruno's talk on Thursday

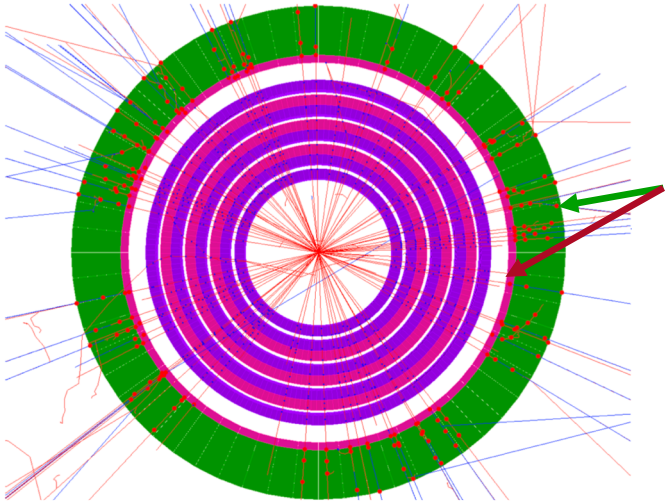


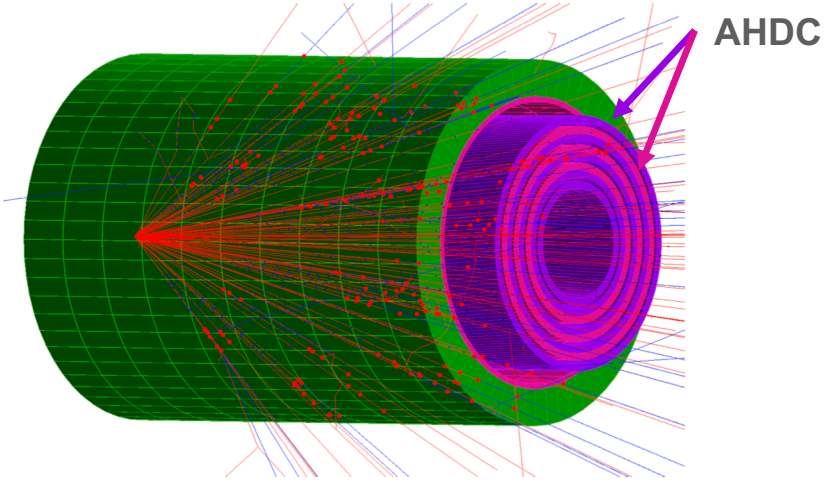


ALERT

Ongoing work on:

- Geometry:
  - Both ATOF and AHDC implemented in coatjava and transferred to gemc
- Simulation:
  - First version of detector digitization implemented
- Reconstruction:
  - Ongoing work to implement DC reconstruction starting from hits, clusters and crosses
  - Will use KF from tracking tool library
- Calibration:
  - Infrastructure based on coatjava tools in place now being populated starting from ATOF
- Work by ANL, Orsay, Temple with support from JLab





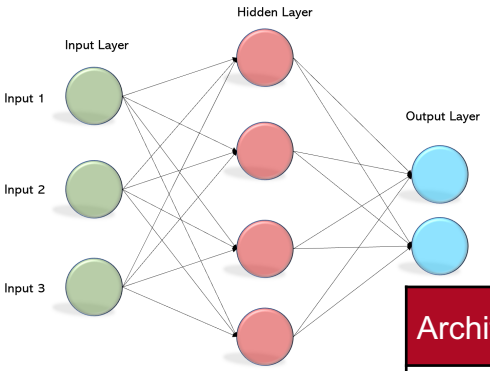
CVT restructuring and generic KF tool

- Stand-alone Kalman-Filter included in clas-tracking common tools package
  - KFitter, StateVec, MeasVec classes modified to remove ”built-in” geometry to propagate state vector to measurement sites, and compute projector value and matrix.
  - Surface class and surface Type enum to represent measurement surfaces and objects ➔ surfaces constructed to allow all translational and rotational degrees of freedom
    - Surfaces: planes & cylinders with measurement points, lines, strips
    - Strip object with centroid, position and uncertainty on position
    - Projector for strips and simple lines as DOCA
  - Implementation for CVT
    - Computation of pseudo-line representing cluster line in lab frame
    - CVT service creates surfaces ➔ passed to KF in initialization
  - Functionality to choose units (cm, mm)
  - Numeric estimate of covariance matrix
- KF implemented
- Now testing implementation for CVT
  - Dedicated test service
  - Compare efficiency and resolution with current service
  - Switch over to new service when validated
- Can be used for other detectors (ALERT)

AI tracking

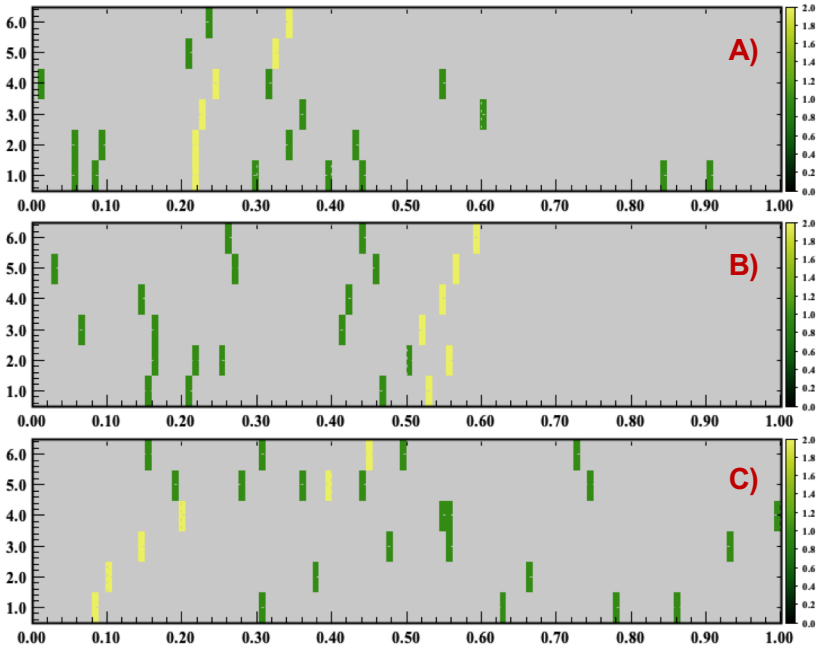
CLAS12 Tracking with Artificial Intelligence

- AI for CLAS12 Tracking:
  - Neural Network trained on cluster combinations.
  - Several Network Architectures are considered
    - Convolutional Neural Networks (CNN)
    - Multi-Layer Perceptron (MLP)
    - Extremely Randomized Trees (ERT)
  - Accuracy determined by Confusion Matrix
  - Multi-Layer Perceptron performed the best



Architecture	Accuracy	Inefficiency
MLP	**99.7%**	**0.3%**
CNN	95.6%	4.4%
ERT	98.5%	1.5%

AI Track reconstruction from cluster combinations

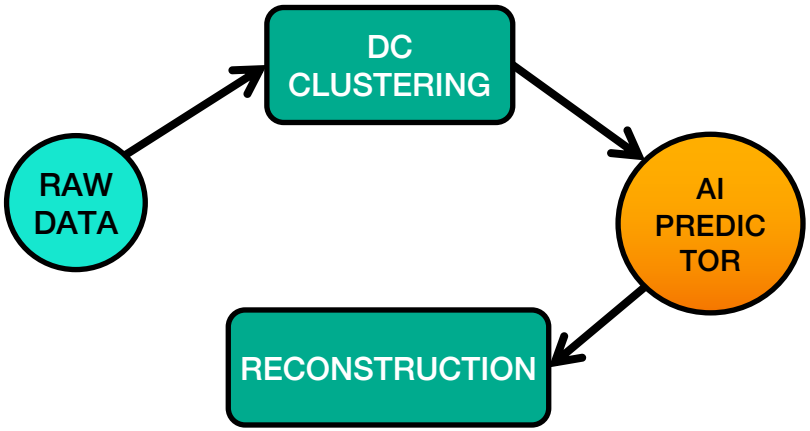


Number of combinations:
A) 2304, B) 2880, C) 7200

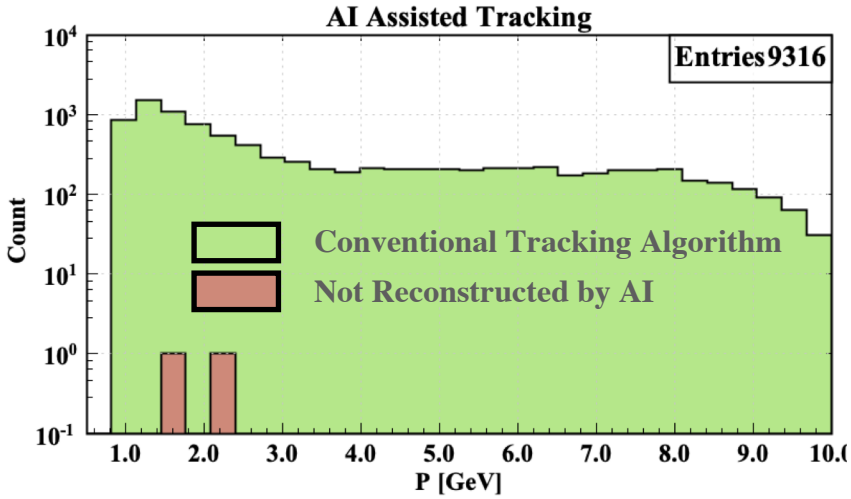
AI tracking

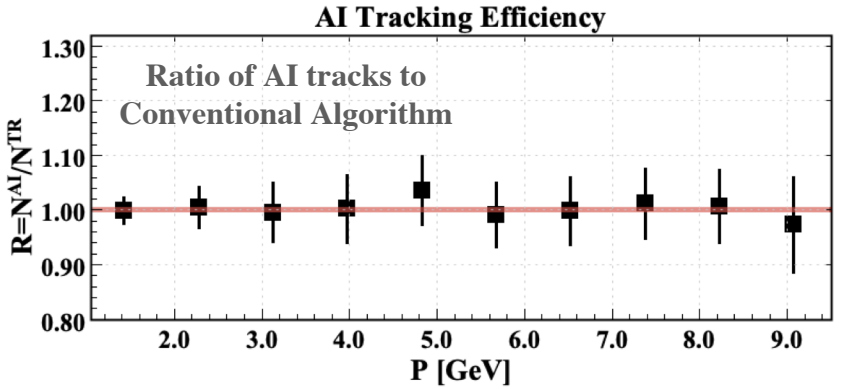
CLAS12 Tracking with Artificial Intelligence

- AI Tracking tools:
  - Training data extraction utility
  - Neural Network (MLP) to train network.
  - Utility to run track prediction algorithm on RAW data.



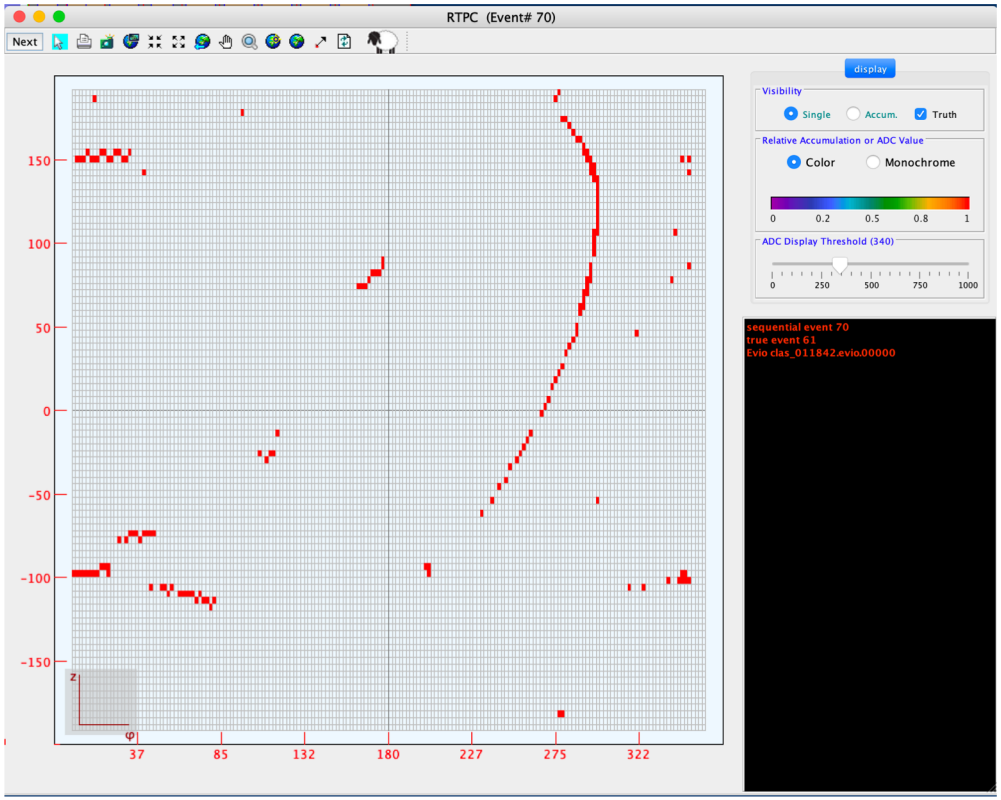
Reconstruction Speedup x6





CED updates

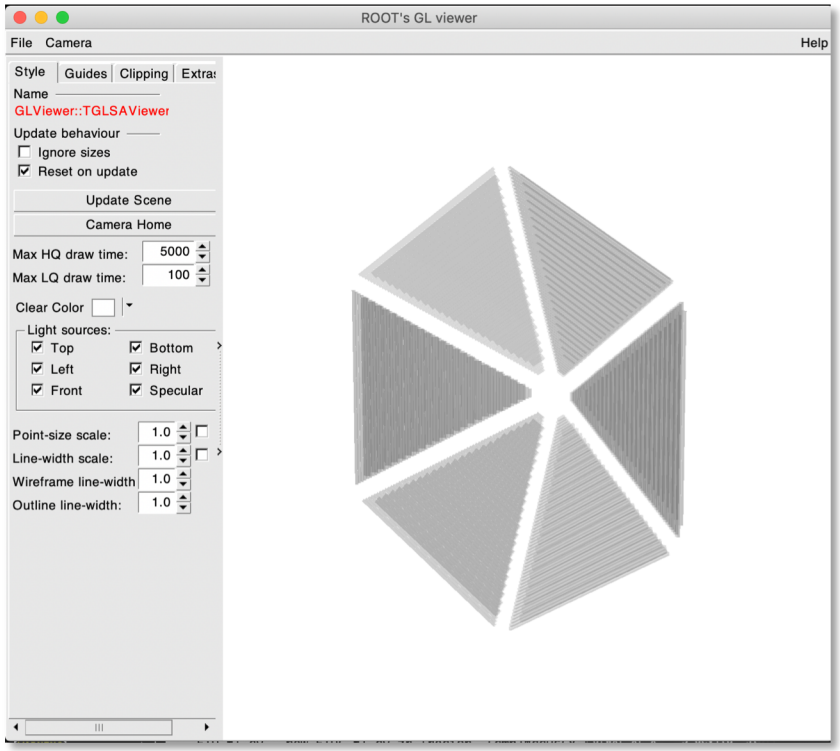
- Recently to include RG-F RTPC
- Other updates on:
  - "true" event number
  - better functionalities to navigate through events
  - display of AI tracking results
  - new 3D library
  - display data from REC::Calorimeter
- All included in CED 1.4.57



New geometry package

Requirements:

- Global service
  - loads all detector geometries
  - provides access to those geometries to users, e.g. reconstruction services
  - recreates geometries if run number changes (instead of using CCDB variations)
- Alignment support
  - uniform methods to apply shifts and rotations to detector elements
- Trajectory surface support
  - surfaces defined as part of the detector geometry and made accessible to user
- Support for the necessary GEANT4 volumes and their export to GEMC
- Support for importing STL volumes (e.g. CTOF)
- Support for querying line/track intersections with detectors
  - provide the detector elements intersected
  - provide intersection points (e.g. entrance and exit)
- Convenience methods for translation between global and local coordinates (e.g. for fiducials)
- Visualization capabilities

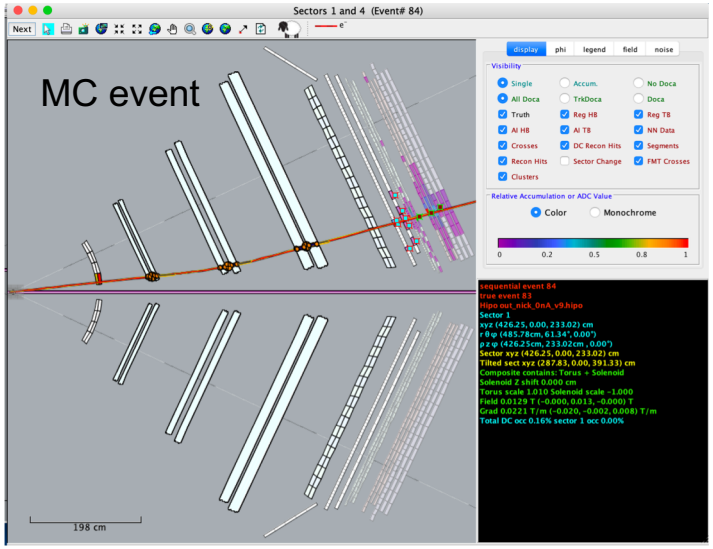


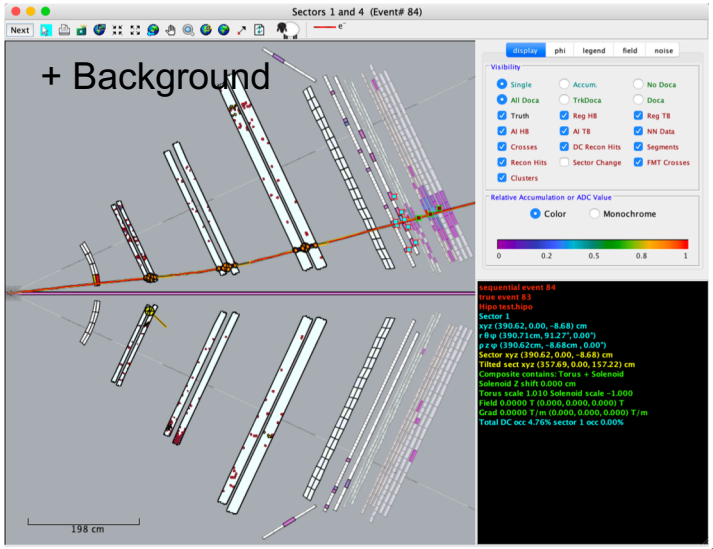
Status:

- General framework implemented
- FTOF geometry imported for testing
- Necessary refinements identified and currently in the works

Background merging

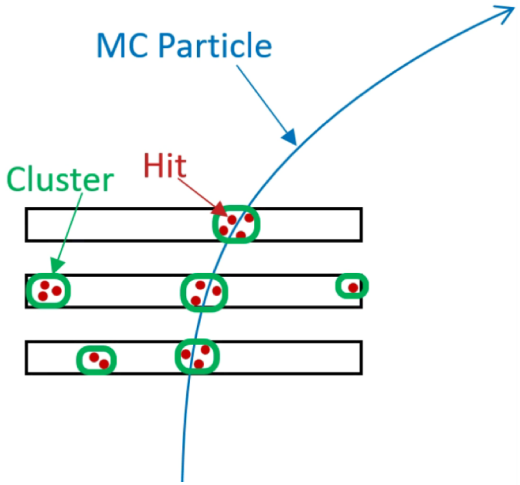
- Tools to filter and merge background from real events with real or simulated events included in coatjava
- Filter tool:
  - Selects events from a specific trigger bit applying a threshold on the beam current
  - Runs on hipo files
- Merging tool:
  - Merges raw banks (adc and tdc) of the primary event and the background event
  - Accounts for readout electronics behavior (multiple hit suppression, tdc jitter)
- Status:
  - Chain fully exercised on both data (low luminosity) and MC
  - Validation completed on data, in final stage for MC
  - Will be included in next release





See Stepan's talk for information on validation

Truth matching



- Core developed by F. Bossù
- The purpose is to rigorously connect MC particles with reconstructed particles
- The matching is based on the following fact:
  - For a given detector the order of hits in MC::True bank is the same as the order of ADC/TDC banks
- Output is the "MC::IsParticleMatched" bank

An example event

mcTindex :	1	2	3
recTindex :	-1	0	1
isInAcc :	0	1	1
nMCclusters :	11	37	37
fraction :	0.0000	0.9730	1.0000

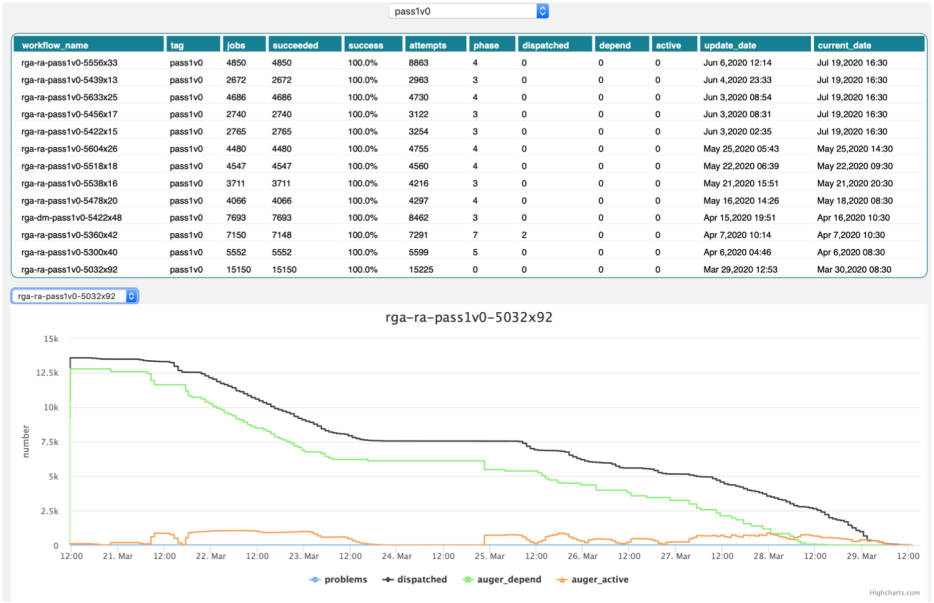
Index of MC particle
Index of Rec particle
If the trk is reconstructable (e.g. min 5 SLs for DC)
# of MC Clusters
Fraction of clusters used in Rec

- Currently works for charged particles only, being extended to neutrals
  - Needs mother information to be saved in gemc
  - Missing links to ECAL hits and clusters added
- Not yet compatible with background merging
- Available in the "iss540-mcMatching", expected to be merged to "development" soon

Data processing workflow

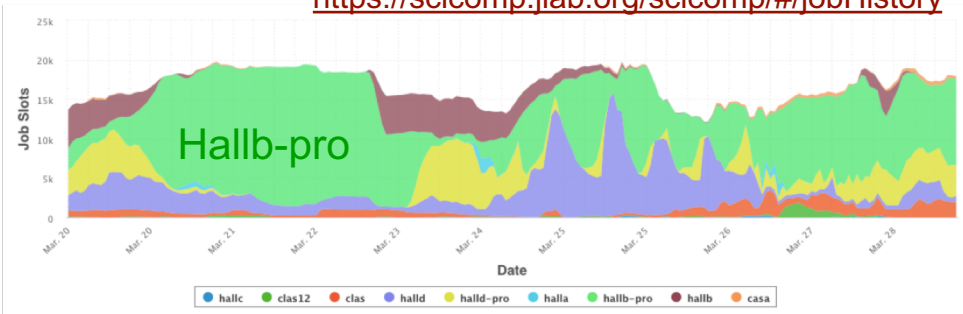
- Data processing for RG-A, B and K based on CLAS12 workflow tools:
  - Allows chefs to run decoding, reconstruction and skims on large data sets in a single process
  - Flexibility in selecting processing steps, reconstruction and skim configurations, file merging
  - Configurable resources (threads, memory, …) for individual job. Optimized configuration set by default
  - Relies on SciComp SWIF to implement job-job dependencies and automatic retries in case of jobs failures
- Exercised on about 2PB of data and more than 1M jobs with basically 100% success
- ”Cheffing” meeting every week to coordinate work

https://clas12mon.jlab.org/rga/status/processing/



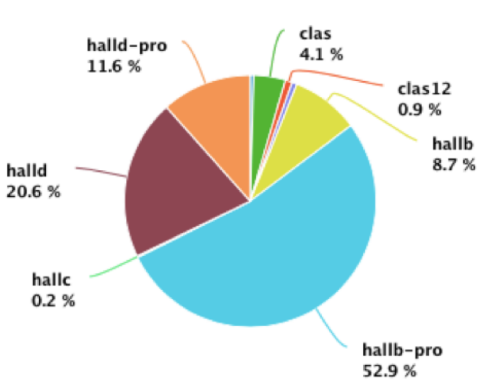
workflow_name	tag	jobs	succeeded	success	attempts	phase	dispatched	depend	active	update_date	current_date
rga-ra-pass1v0-5556x33	pass1v0	4850	4850	100.0%	8863	4	0	0	0	Jun 6,2020 12:14	Jul 19,2020 16:30
rga-ra-pass1v0-5439x13	pass1v0	2672	2672	100.0%	2963	3	0	0	0	Jun 4,2020 23:33	Jul 19,2020 16:30
rga-ra-pass1v0-5633x25	pass1v0	4686	4686	100.0%	4730	4	0	0	0	Jun 3,2020 08:54	Jul 19,2020 16:30
rga-ra-pass1v0-5456x17	pass1v0	2740	2740	100.0%	3122	3	0	0	0	Jun 3,2020 08:31	Jul 19,2020 16:30
rga-ra-pass1v0-5422x15	pass1v0	2765	2765	100.0%	3254	3	0	0	0	Jun 3,2020 02:35	Jul 19,2020 16:30
rga-ra-pass1v0-5604x26	pass1v0	4480	4480	100.0%	4755	4	0	0	0	May 25,2020 05:43	May 25,2020 14:30
rga-ra-pass1v0-5518x18	pass1v0	4547	4547	100.0%	4560	4	0	0	0	May 22,2020 06:39	May 22,2020 09:30
rga-ra-pass1v0-5538x16	pass1v0	3711	3711	100.0%	4216	3	0	0	0	May 21,2020 15:51	May 21,2020 20:30
rga-ra-pass1v0-5478x20	pass1v0	4066	4066	100.0%	4297	4	0	0	0	May 16,2020 14:26	May 18,2020 08:30
rga-dm-pass1v0-5422x48	pass1v0	7693	7693	100.0%	8462	3	0	0	0	Apr 15,2020 19:51	Apr 16,2020 10:30
rga-ra-pass1v0-5360x42	pass1v0	7150	7148	100.0%	7291	7	2	0	0	Apr 7,2020 10:14	Apr 7,2020 10:30
rga-ra-pass1v0-5300x40	pass1v0	5552	5552	100.0%	5599	5	0	0	0	Apr 6,2020 04:46	Apr 6,2020 08:30
rga-ra-pass1v0-5032x92	pass1v0	15150	15150	100.0%	15225	0	0	0	0	Mar 29,2020 12:53	Mar 30,2020 08:30

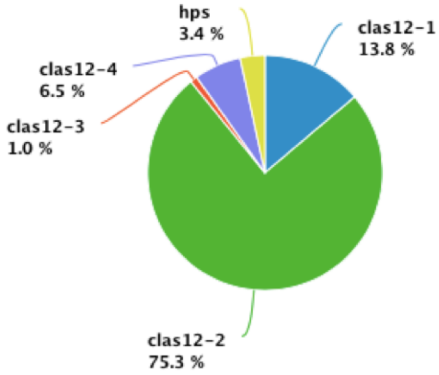
https://scicomp.jlab.org/scicomp/#/jobHistory

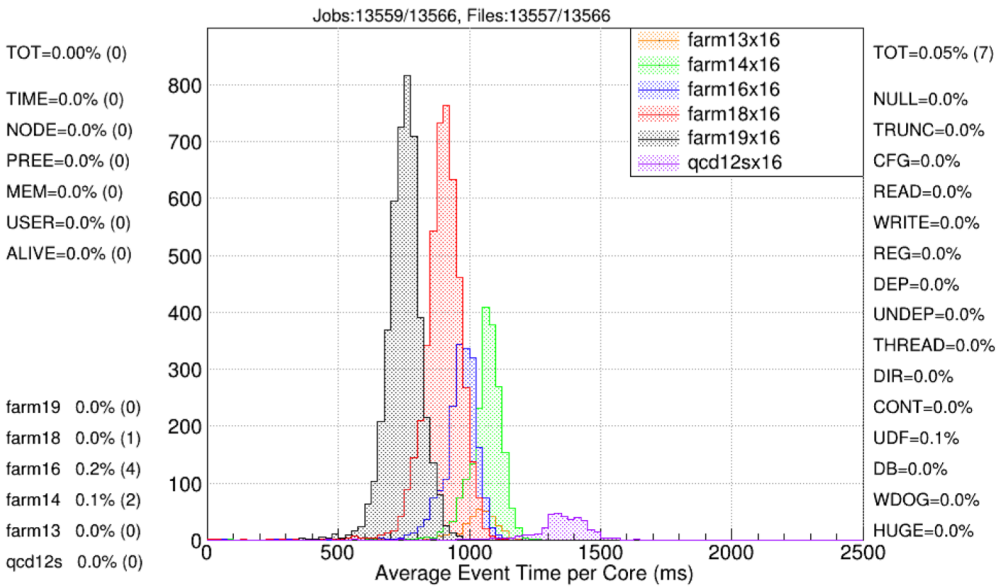


Data processing performance

- From the first week of RG-A processing:
  - 4.9 billion events @ 1.26 million thread/slot-hours (and ~700 million per day)
  - 930 ms per event, which matches expectation for the node-flavor distribution we received
  - Slurm reports we're running with CPU/Wall time ~90%
  - System/job failure rate low, <1%, all recovered successfully with automatic Swif retries
- Fair-share:
  - CLAS12 data processing capable of using the whole assigned fair-share (36% of the farm) or more if available







Computing resource estimates

	Events	Core Hours (M)		Tape (TB)		Disk (TB)				Ingress (TB)	Egress (TB)	
	(G)	Data	Simulation	Raw	Processed	Volatile	Cache	Work	SSD	Simulation	Simulation	DSTs
2021	52	48	20	1295	1500	290	650	150	200	120	760	1500
2022	57	26	40	1426	450	290	780	200	250	240	1500	435
2023	112	28	40	2022	500	325	780	250	250	240	1500	429
2024	130	23	60	3720	530	350	800	250	300	480	2300	507.5
2025	52	49	60	1296	1240	350	850	300	300	480	2300	933

- Estimates of computing resources for the next five years updated for recent Science&Technology review
- Based on:
  - Known data volumes for completed runs and projected data and trigger rate for upcoming runs
  - Current measurements of data processing speed for different data sets/luminosity
  - Fixed ratio between number of real and simulated events
  - DSTs and skims are kept on disk for recently processed runs
  - Ingress and egress rates accounts for transfer of skims to collaborating institutions and data transfers in support of offsite simulations

Skims and trains

- Ongoing efforts from Run Groups to optimize skims for physics analyses to increase reduction factor
- Skims should be tailored to specific final states to results in manageable data volumes
- Can be achieved using standard (event topology) or custom (additional cuts) wagons
- Implementation of custom wagons facilitated without unnecessary duplication by added shared functionalities (BeamTargetWagon)

Encouragement to all collaborators to invest in optimizing their skims!

RG-K train configuration

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		

GEMC development

- 4.4.0:
  - geant4 10.6 support ✅
  - conform all detectors to read RUNNO and DIGITIZATION_VARIATIONS in the digitization ✅
  - add time offsets for: LTCC ✅
  - add time offsets for: EC, PCAL ✅
  - HTCC mc_gain implementation ✅
  - conform all detectors to read DIGITIZATION_TIMESTAMP in the digitization ✅
  - dc, ec, htcc resolution matching (in progress)
  - gcards, yaml files from 4.4.0 and on in a dedicated "config" subdir ✅

Numbering scheme changes: hipo4 breaks backward compatibility. So this release is "major". Also, from now on we go to two numbers only.

- 5.0:
  - Hipo 4 output ➡ SOON

Dedicated "config" folder inside each release number with GCARDS and YAML files

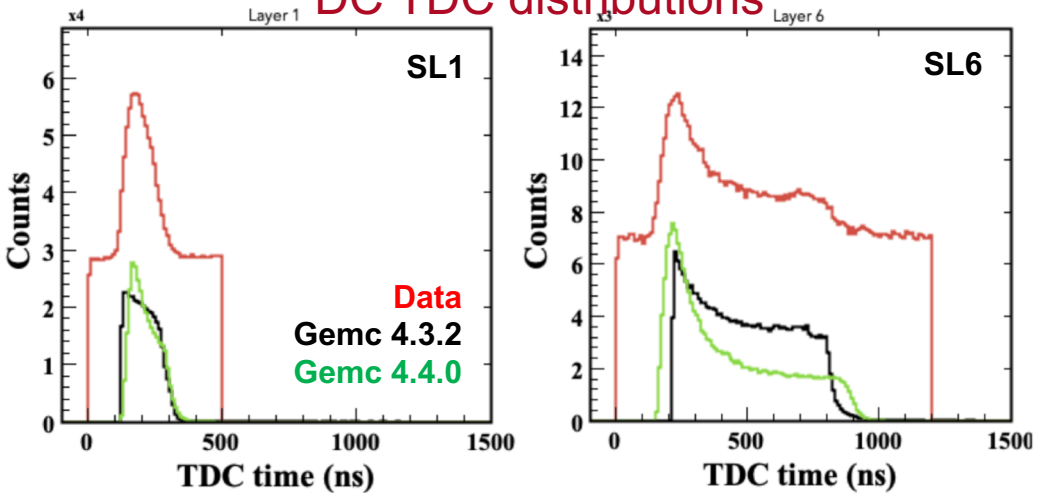
Example of DIGITIZATION_TIMESTAMP: "2020-02-19"

Visit the clas12tags repository at https://github.com/gemc/clas12Tags for plans, configurations, howtos

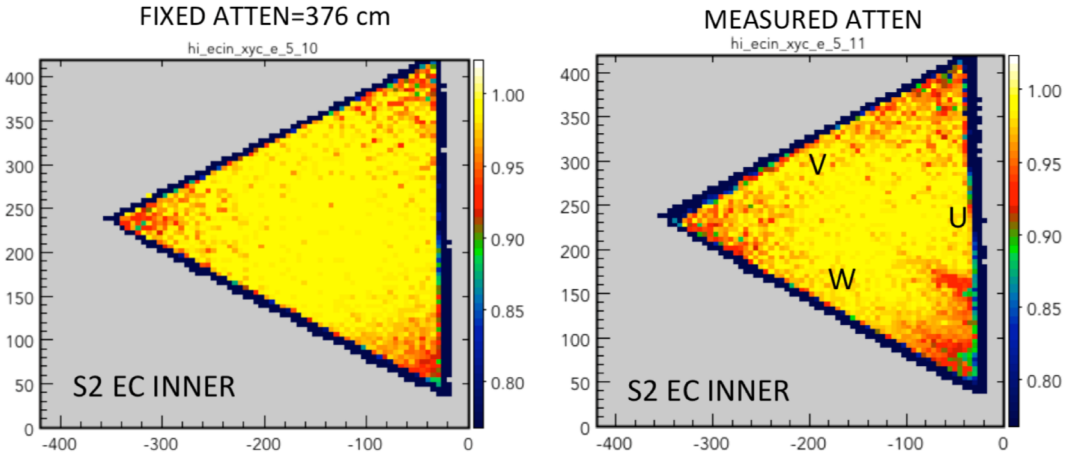
Detector response tuning

- Tuning of detector response to reproduce actual performance from data:
  - TOF resolution from calibration results
  - DC intrinsic efficiency and resolution
  - ECAL attenuation length and photoelectron statistics
  - HTCC response in terms of number of photoelectrons
- TOF resolution already available in 4.3.2, others coming with 4.4.0

DC TDC distributions



EC distributions



Event generators

Generators collected at https://github.com/JeffersonLab/clas12-mcgen/

- generators to be used in the docker container for offsite simulations are added as submodules to jeffersonlab/clas12-mcgen
- frozen version of each submodule in container
- requirements to access generators from OSG portal
- if you want to include a generator in the docker container, please visit the page for the requirements, and contact Nathan or Mauri

Additional Requirements

- An executable with the same name as the github repository name, installed at the top level dir
- If libraries are needed, they should be put inside a lib directory, at the top level dir
- The generator output file name must be the same name as the exectuable + ".dat". For example, the output of clasdis must be clasdis.dat
- To specify the number of events, the option "--trig" must be used
- If necessary, an environment variable (name in its README) where the executable will look for data
- The optional argument --docker will be added by default to all executable. This option can be ignored or used by the executable to set conditions to run on the OSG container

If you are the maintainer of a package and made changes that you want to include here, send emails to ungaro@jlab.org, baltzell@jlab.org (Mauri or Nathan) requesting the update.

List of Generators

name	summary description	maintainer	email	requirements met
clasdis	clas SIDIS MC based on PEPSI LUND MC	Harut Avakian	avakian@jlab.org	🔴
claspyth	SIDIS full event generator based on PYTHIA	Harut Avakian	avakian@jlab.org	🔴
dvcsgen	DVCS/pi0/eta generator based on GPD and PDF parameterizations	Harut Avakian	avakian@jlab.org	🔴
genKYandOnePion	KY, pi0P and pi+N	Valerii Klimenko	valerii@jlab.org	✅
inclusive-dis-rad	generates inclusive electron and optionally radiative photon using PDFs	Harut Avakian	avakian@jlab.org	🔴
tcsgen	Timelike Compton Scattering	Rafayel Paremuzyan	rafopar@jlab.org	🔴
jpsigen	J/Psi	Rafayel Paremuzyan	rafopar@jlab.org	✅

emails

ungaro@jlab.org, baltzell@jlab.org, avakian@jlab.org, valerii@jlab.org, rafopar@jlab.org

New docker container in preparation

clas12software:4.4.0

- **Installed Software:**
  - CCDB version: 1.07.00
  - CLHEP version: 2.4.1.3
  - GEANT4 version: 4.10.06.p02
  - QT using system installation
  - XERCESC version: 3.2.3
  - EVIO version: 5.1
  - MLIBRARY version: 1.4
  - SCONS version: 1.9
  - CLAS12 Tag: 4.4.0
- **CVMFS Software:**
  - Coatjava
  - Java

- xrootd access to background files @JLAB for merging
- cvmfs access to CLAS12 software, plus:
  - tagged CCDB SQL file
  - (magnetic field maps, not yet)
- conform all detectors to read DIGITIZATION_TIMESTAMP in the digitization
- self contained environment and configurations
- automatic (empty) tests

tag: **production** (currently 4.3.2)

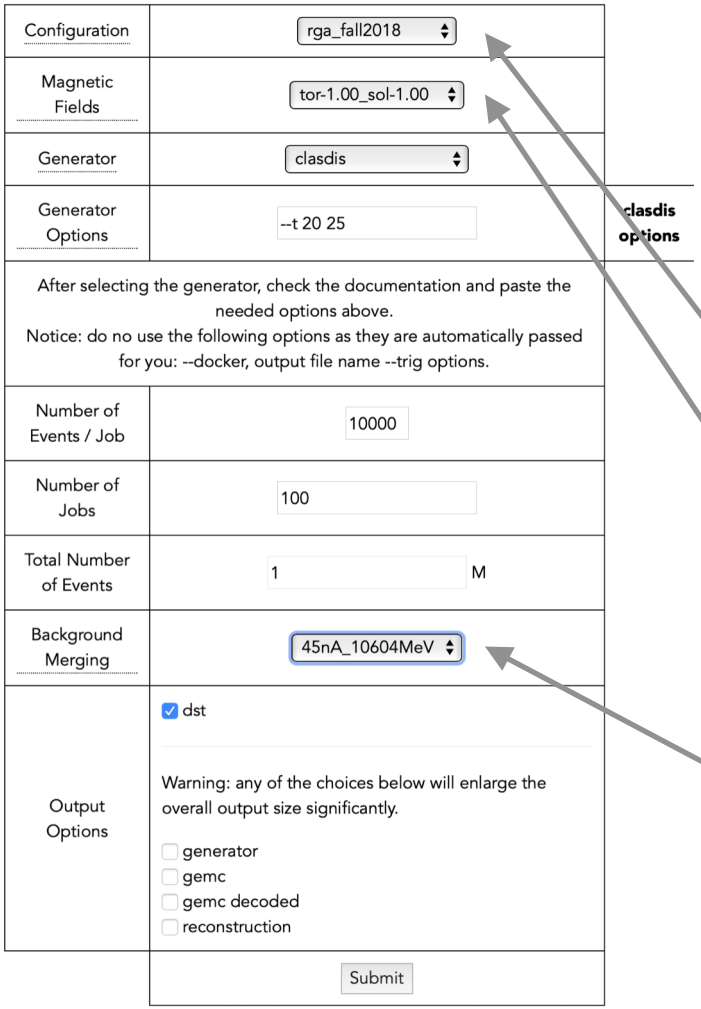
OSG Production Simulations through the portal

tag: **development** (currently 4.4.0)

For tests of the new software versions

OSG portal

https://gemc.jlab.org/web_interface/index.php



OSG Team: MIT Group Sangbaek Lee, Robert Johnston, Patrick Moran

Using MYSQL databases for PRODUCTION and TEST

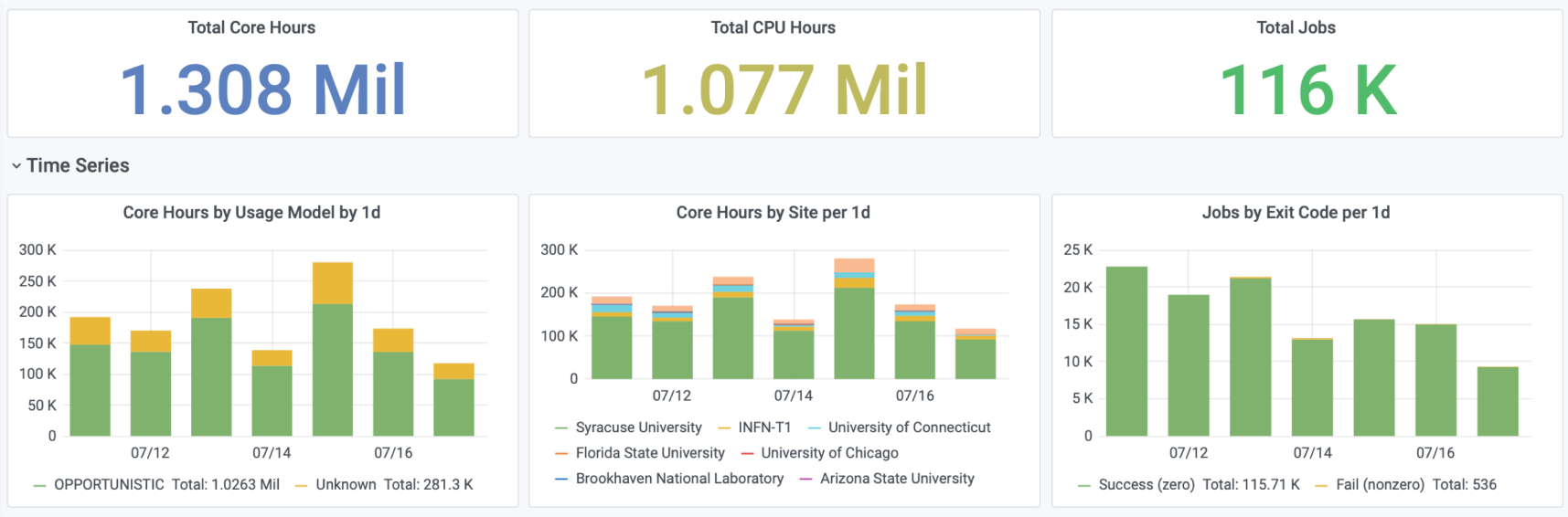
PHP, Javascript, MSYQL backend, GEMC server

Quantities based on actual experiments (ran or approved) configurations: JSON auto-generated.

- experiment configurations choices
- magnetic field choices
- background merging choices yes/no/choice

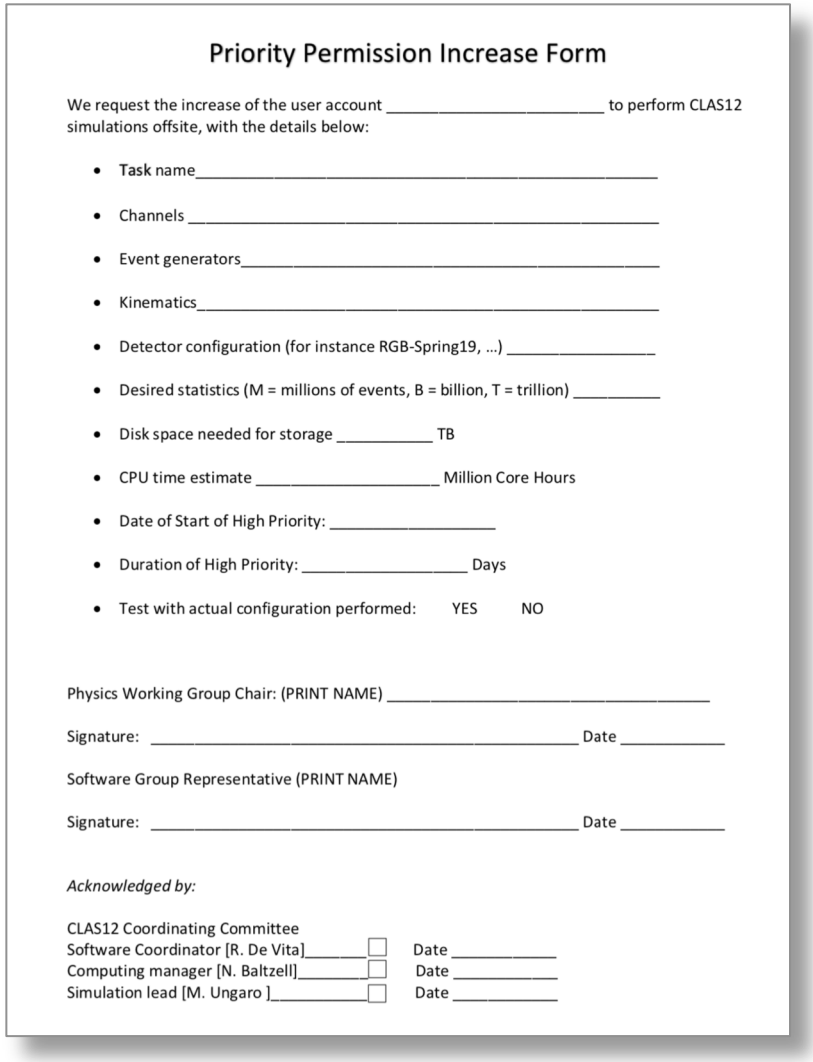
Offsite resources

- Offsite clusters accessible via OSG
  - Dedicated resources:
    - UK-GridPP: 400 cores guaranteed, 1000 next year
    - INFN-CNAF: 500 cores guaranteed
    - GRIF: test completed, request submitted
    - Others in the works
  - Semi-dedicated:
    - Uconn
    - Syracuse
  - Opportunistic



https://gracc.opensciencegrid.org/d/000000080/vo-summary?orgId=1&from=now-7d&to=now&var-interval=$__auto_interval_interval&var-vo=All&var-type=Payload&var-Filter=ProjectName%7C%3D%7CCLAS12

Priority request



Priority Permission Increase Form

We request the increase of the user account _____ to perform CLAS12 simulations offsite, with the details below:

- Task name_____
- Channels _____
- Event generators_____
- Kinematics_____
- Detector configuration (for instance RGB-Spring19, …) _____
- Desired statistics (M = millions of events, B = billion, T = trillion) _____
- Disk space needed for storage _____ TB
- CPU time estimate _____ Million Core Hours
- Date of Start of High Priority: _____
- Duration of High Priority: _____ Days
- Test with actual configuration performed: YES NO

Physics Working Group Chair: (PRINT NAME) _____

Signature: _____ Date _____

Software Group Representative (PRINT NAME)

Signature: _____ Date _____

Acknowledged by:

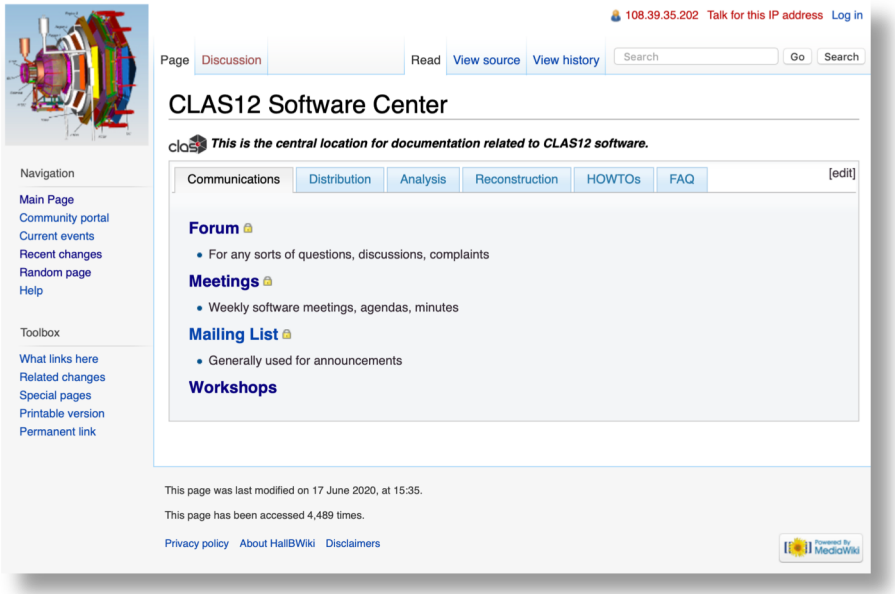
CLAS12 Coordinating Committee
Software Coordinator [R. De Vita]_____☐ Date _____
Computing manager [N. Baltzell]_____☐ Date _____
Simulation lead [M. Ungaro]_____☐ Date _____

- Procedure in place to request priority increase for large production
- Requires endorsement of relevant Physics Working Group and the Software Group
- Acknowledgment by CCC

- First usage for RG-A SIDIS simulations
  - 2B events in inbending configuration in about 35 days
  - New request to proceed with outbending

- See https://gemc.jlab.org/web_interface/about.html

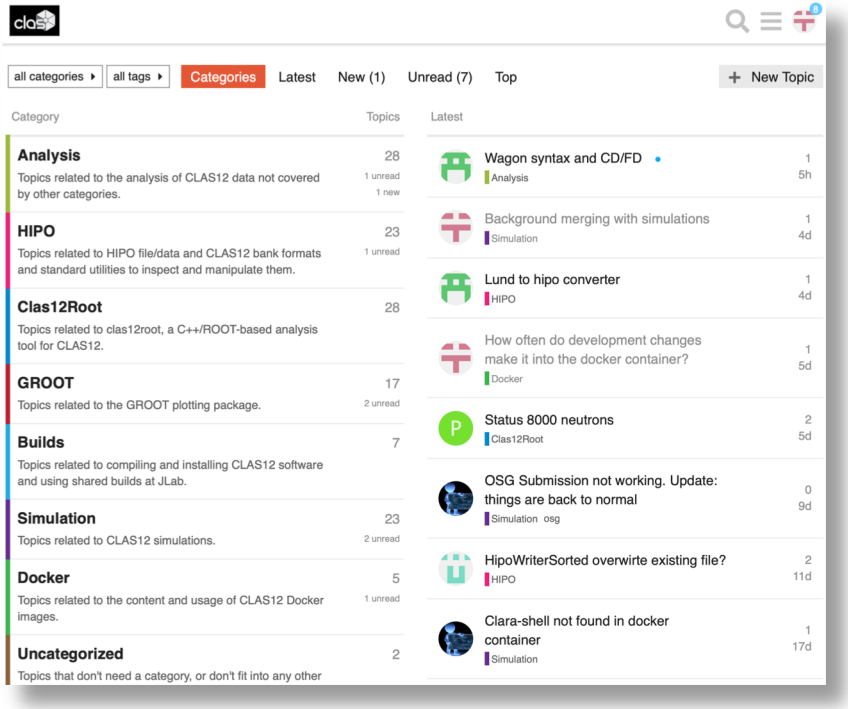
Documentation



- New HOWTOs tab:
  - How to run the full simulation chain single-threaded
  - How to test a custom wagon
  - How to interactively probe the magnetic field
  - How to mount and use the CLAS12 software releases on CVMFS
- Others to come based on user request

Centralized software wiki: https://clasweb.jlab.org/wiki/index.php/CLAS12_Software_Center

Discourse forum for software related questions and communications



Ongoing and planned work

- Ongoing work on:
  - CVT
  - AI tracking
  - EB rerun from DSTs
  - new reconstruction packages (RTPC, FMT, ALERT)
  - background merging
  - truth-matching
  - geometry restructuring
  - GEMC digitization
  - OSG performance and portal upgrade
  - CVMFS build deployment
- Longer term work plan developed by the Software Task Force will be presented in the near future