CLAS12 Calibration Task Force

Raffaella De Vita (Jefferson Lab)
and the Calibration Task Force

March 4, 2025

CLAS12 Calibrations Task Force

- Task Force appointed with the **objective** of making calibrations more efficient:
  - "calibrations": all the steps in between data taking and data processing for physics
  - "more efficient": (more) automatic, faster, using less resources, …

- **Charge**:
  - Perform a survey of the calibration process, i.e. procedure, execution time, tools, resources, most frequent issues/errors
  - Identify key items to reach the objectives
  - Define a work plan to address them
  - Execute the highest priority item(s)
  - Verify the impact
- **Members**:
  - N. Baltzell, D. Carman, R. De Vita (chair), C. Dilks, F. Hauenstein

Task Force Activity

- **Meetings**:
  - Kick-off meeting on November 1st, 2024
  - Meeting every other week since then

- **Focus**:
  - **Data Collection:**
    - Survey of the calibration process
    - Collected information on the most frequent issues/errors
    - Compiled list of potential upgrades for existing tools
  - **Identification of key areas and priorities:**
    - Procedural
    - Resource related
    - Tools related
  - **Implementation of first action items**
    - Calibration checklist
    - Reduction of data volumes
    - Unified calibration code repository

- **Documentation**
  - Collected material, presentations, meeting minutes in O365 shared drive

Survey of the calibration process

- Collect data on
  - Number of iterations
  - Number of processed runs and events
  - Expected/actual processing time on the JLab batch farm (full Hall-B production fairshare/actual fairshare at the given time
  - Total time including human time
  - Disk space usage

  for every calibration phase/step
- Focus on data sets with ongoing calibration for better accuracy via real-time data collection

Example of collected data

data set general information	name	**RG-X**
	number of triggers (B)	40
	number of settings	6
	data volume (TB)	1150
	number of runs	630
	raw event size (kB)	29
	decoded event size (kB)	6
	calib event size (kB)	17
	mon event size (kB)	40
	events in calibration runs (M)	100
	pass0 files/run	20

computing resources	CPU fairshare (M/day)	600
	startup overhead (days)	1
	disk quota (/volatile)	400

values in italics are preliminary
**includes data processing*

		Iterations	N. runs	N. of Events (B)	Time (days)*	Wall time (days)	Fairshare (days)	Disk (TB/iter.)
calibration step	FD alignment	3			15			
	CD alignment	0						
	beam offset	1		0	25	6	0	0
	RF	2		0	7	0	0	0
	FTOF calibration	2	45	4.5	30	20	9.50	52
	DC calibration	*3*	*15*	*1.5*			5.50	12
	Other detectors	*2*	*15*	*1.5*			4.50	17
	pass0s	5	630	4.38			12.30	40
	totals			**11.9**	**77.0**	**26.0**	**31.8**	**120.8**

Survey of the calibration process

Data across Run Groups are not always complete but lead to the same findings (not in order of relevance):

- Calibration protocols/algorithms keep being improved
  - Necessary to achieve the best quality or address new issues (for example, new DC alignment and calibration)
  - Significant overhead/delays during development
- Usage of computing resources is significant
  - Processed data amount up to 30% of total → can be weeks of processing time
  - Data volumes produced can be hundreds of TB, even if transient
- Human resources involved in the process are large
  - Many people and time
  - Delays unavoidable without (more) automation
- Non-negligible impact of accidental mistakes
  - Variety of examples and causes
  - Usually requires repeating the affected calibration step

New calibration checklist

- Outlines all the tasks of the calibration process
- Provide information on preparatory steps and interdependencies
- Developed in collaboration with CalCom
- Being tested with ongoing calibration

Calibration Checklist - v2

Step #	Category	Item	Contact Person	Prerequisites	Input Data	Outcomes	Validation	Documentation Source
1	Preparatory steps	Dataset archaeology	Daniel Carman			list of runs and known issues		RCDB, logbook, ...
2		Ready for Calibration review	Daniel Carman					[redacted]
3		Update geometry in CCDB	Analysis Coordinator			/geometry/shifts/* /geometry/*/alignment /geometry/rich/config		
4		Update target materials in CCDB	Analysis Coordinator			/geometry/materials/target		
5		Cable swaps	Analysis Coordinator + Subsystems			updated TTs and/or swap configuration		
6		TDC jitter	Analysis Coordinator + Subsystems			updated jitter tables		
7		RF config in CCDB	Raffaella De Vita			updated /calibration/eb/rf/config		
8		FTCAL leakage and angular correction tables	Raffaella De Vita			updated /calibration/ft/ftcal/ecorr, /calibration/ft/ftcal/thetacorr, /calibration/ft/ftcal/phicorr		
9		RICH reco tables	Marco Mirazita			/calibration/rich/reco_flags /calibration/rich/reco_parameters		
10		Pressure	Florian Hauenstein			updated /hall/weather/pressure		
11	Ancillary information	hwp	Analysis Coordinator			updated /runcontrol/hwp	physics timelines	
12		helicity	Analysis Coordinator			updated /runcontrol/helicity		
13		fcup and slm	Analysis Coordinator			updated /runcontrol/fcup and /runcontrol/slm		
14		Setup physics timelines	Christopher Dilks			Data-set specific settings		
15	Detector alignment	Raster calibration						
16		DC alignment	Raffaella De Vita	reasonable DC calibration, DC swaps, zeroed beam spot (raster+beam position)		updated /geometry/dc/alignment table in CCDB		https://github.com/JeffersonLab/clas12alignment/tree/master/dc
17		CVT alignment	Yuri Gotra	BMT timing cuts in CCDB		updated /geometry/cvt/svt/layeralignment /geometry/cvt/mvt/alignment		[redacted]
18		FD alignment	Cole Smith	#16		updated /geometry/ftof/alignment /geometry/ec/alignment /geometry/pcal/alignment		
19		FMT alignment (if necessary)	Yuri Gotra	DC alignment		updated /geometry/fmt/alignment		[redacted]
20		Beam offset		full pass0 with DC and CVT alignments	Pass0 recon files	updated CVT and beam position tables /geometry/beam/position /geometry/cvt/svt/position /geometry/cvt/mvt/position	beam spot analysis on new pass0	in preparation
21	Subsystem calibration - phase 1	RF calibration - remove global offsets	Raffaella De Vita	full pass0 after #19 and reasonable FTOF panel 1B calibration (sigma<100 ps)	Pass0 RF train	updated /calibration/eb/rf/on/eb/rf/offset	calibration timelines	[redacted]
22		FTOF calibration and HW status assessment	Daniel Carman	#21	Calibration train	updated FTOF calibration tables		[redacted]
23		DC time offset calibration and HW status	Florian Hauenstein	#22		updated /calibration/dc/time_corrections/T0Corrections	calibration timelines	
24		BAND time walk calibration	Florian Hauenstein		Laser data	updated /calibration/band/time_walk_amp_left /calibration/band/time_walk_amp_right	?	
25		RF calibration	Raffaella De Vita	#22	Pass0 RF train	updated /calibration/eb/rf/on/eb/rf/offset		[redacted]
26		Second FTOF calibration (if necessary)	Daniel Carman	#25	Calibration train	updated FTOF calibration tables	calibration timelines	[redacted]
27	Subsystem calibration - phase 2	DC time-to-distance calibration and TDC cuts	Florian Hauenstein	#26		updated /calibration/dc/time_to_distance/T2Dpressure /calibration/dc/time_corrections/TDCtimingcuts	calibration timelines	
28		BAND calibration and HW status	Florian Hauenstein			updated /calibration/band/*		
29		CND calibration and HW status	Silvia Niccolai		Calibration train	updated /calibration/cnd/*		See tutorial and algorithm documentation at https://github.com/JeffersonLab/clas12calibration-cnd
30		CTOF calibration and HW status	Daniel Carman		Calibration train	updated /calibration/ctof/*		[redacted]
31		ECAL calibration and HW status	Cole Smith			updated /calibration/ecal/*		
32		FT-Cal calibration and HW status	Raffaella De Vita		Calibration train	updated /calibration/ft/ftcal/*		[redacted]
33		FT-Hodo calibration and HW status	Nick Zachariou		Calibration train	updated /calibration/ft/fthodo/*		[redacted]
34		HTCC calibration and HW status	Izzy Illari		Calibration train	updated /calibration/htcc/*		[redacted]
35		LTCC calibration and HW status	Valerio Mascagna		Raw data	updated /calibration/ltcc/*		[redacted]
36		RICH timing calibration	Marco Mirazita		Calibration train	updated /calibration/rich/moduleX/time_offset /calibration/rich/moduleX/time_walk /calibration/rich/moduleX/mapmt_pixel		[redacted]
37	Subsystem calibration - phase 3	Additional subsystem calibrations	Subsystems	#26		updated /calibration/*/*	calibration timelines	
38		ECAL sampling fraction tables	Cole Smith	#31		updated /calibration/eb/electron_sf		
39		RICH Cherenkov angle calibration and HW status	Marco Mirazita	#36	RICH calibration train	updated /calibration/rich/moduleX/cherenkov_angle /calibration/rich/moduleX/status*		[redacted]
40	AI networks	Train DC AI track-finding networks	Analysis Coordinator	#27		new NN files	AI vs conventional studies with/without denoising and luminosity scan analysis	[redacted]
41		Test DC denoising network	Analysis Coordinator			decision on whether to use denoising		
42	Final check	Final calibration checks	Analysis Coordinator	All the above		readiness status	calibration and physics timelines (with new software release if necessary)	
43	PassX Readiness Review preparation	Data vs. MC efficiency study	Analysis Coordinator	All the above and a sample of background events		readiness status		
44		Physics trains definition	Analysis Coordinator	All the above				
45		Physics analyses	Run Group	#44				

Reduction of data volumes

- Data volumes produced for calibration and monitoring of a data set can amount to hundreds of TB
  - Transient data usually written to /volatile (400 TB CLAS12 quota)
  - Large volume due to reconstructed files with all necessary banks for calibration (~5xDSTs)
  - Can lead to early deletion when calibrating multiple data sets
  - Partial mitigation in the final detector specific calibration skims with event and bank filtering, but still 10s, up to 100s GB per typical calibration run (100 M events)
- Analysis of banks list and sizes in reconstructed files for calibration indicates:
  - Data required for a specific detector calibration is often spread in many banks that need to be included in the output
  - Often, only a fraction of those bank entries (e.g. hits, clusters, …) is used
  - Few detectors dominate the overall volume

- **Large reduction can be achieved by building calibration-specific banks with:**
  - **All the information needed**
  - **Only for the hits/clusters/… that are relevant**
  - **Only for the events of interest**

- Quantitative study based on first implementation:
  - New calibration banks designed with the help of detector experts
  - Tested on “major” contributors (DC, FTOF, RICH)
  - Reduction of corresponding volumes ranging from 5x (DC) to 30x (FTOF)
  - Overall reduction of reconstruction output of 3x
  - Benefit for calibration skims too

Calibration tools

- Currently, most detector calibrations are run interactively and/or require results inspection by the calibrator to assess convergence
- First list of upgrades necessary to minimize human intervention and move towards automation:
  - extend software functionalities
  - improve robustness of analysis algorithms
  - implement clear metrics for evaluating the results quality
- First step by collecting all calibration software in a single gitlab repository
  - facilitate implementation of new software functionalities
  - simplify code management and maintenance

Just at the beginning, more to come in the next months

Summary

- First phase focused on collection of data completed:
  - Several factors impact the calibration process efficiency, i.e. not a single culprit
  - Combined effort is needed to strengthen procedures, reduce usage of computing resources, and improve automation of calibration tools
- First action items identified and addressed
- Now moving the focus to calibration tools