# CLAS12 Calibration Task Force

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

July 8, 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

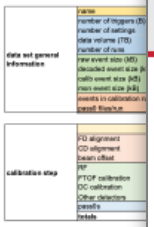
- **Documentation**
  - Collected material, presentations, meeting minutes in O365 shared drive

# 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*



March 4, 2025

# Survey of the calibration process

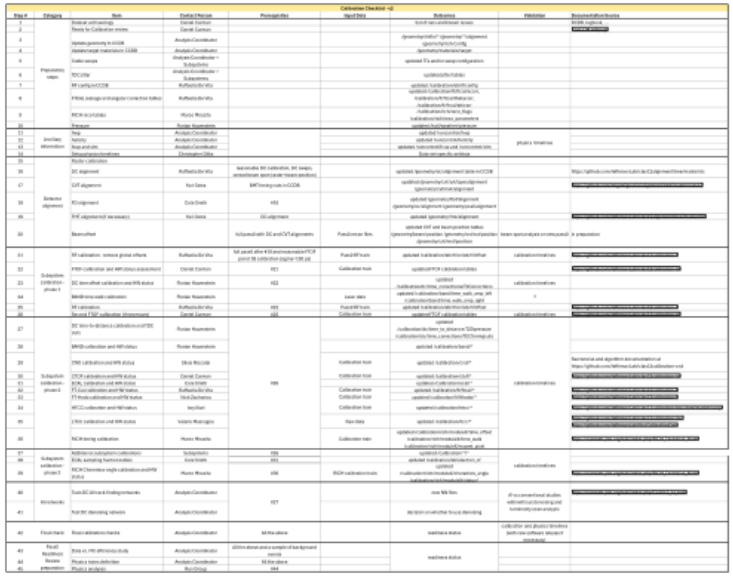
Data across Run Groups ar… findings (not in order of rele…

- Calibration protocols/alg…
  - Necessary to achieve t… example, new DC align…
  - Significant overhead/d…
- Usage of computing reso…
  - Processed data amoun… processing time
  - Data volumes produce…
- Human resources involv…
  - Many people and time
  - Delays unavoidable wit…
- Non-negligible impact of…
  - Variety of examples an…
  - Usually requires repeat…

March 4, 2025

# 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



March 4, 2025 6 CLAS Collaboration Meeting

# 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

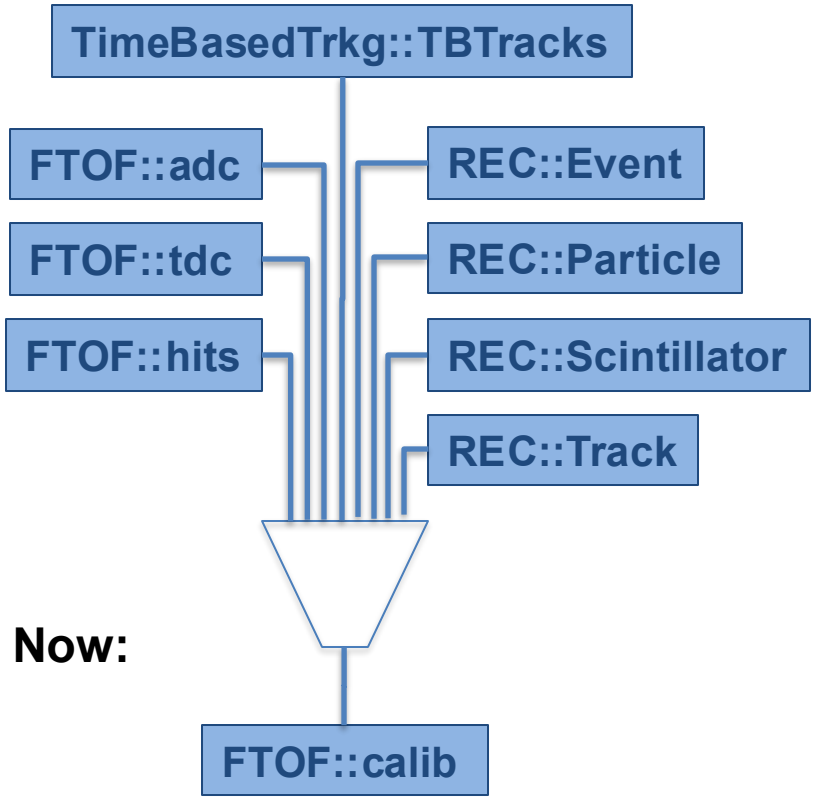
- **Large reduction can be achieved by building calibration-specific banks with:**
  - **All the information needed**
  - **Only for the hits/clusters/… that are relevant**
  - **Only for the events of interest**

# Reduction of data volumes

Currently being deployed:

- New calibration banks designed with the help of detector experts for CTOF, DC, FTOF, RICH
- Dedicated service implemented to create the new banks during reconstruction, filtering on event topology and detector hit/cluster
- Calibration codes updated to use the new banks
- Reduction of corresponding volumes ranging from 5x (DC) to 20x (FTOF)
- Overall reduction of reconstruction output of 3x
- Benefit for calibration skims too, e.g. , FTOF skim 10x smaller and 10x faster to process

**Before:**



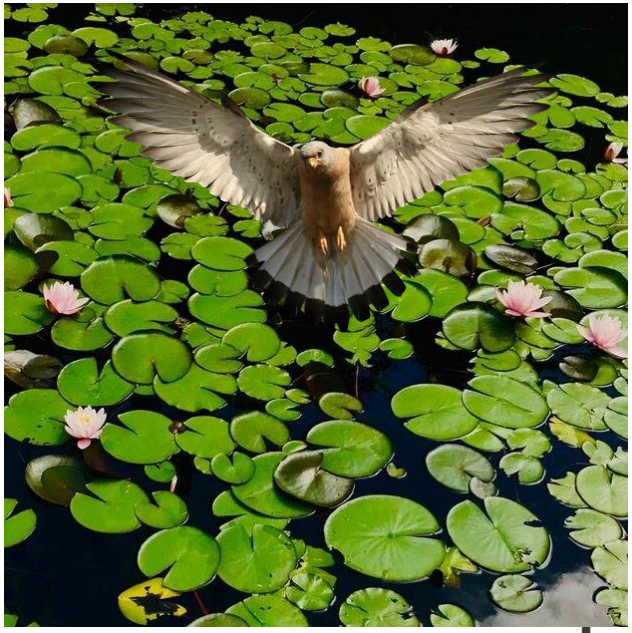

**Now:**

FTOF::calib

# Calibration code reorganization

## Motivation

- We have several calibration suites: repositories that contain calibration software for each detector
  - This is *great* for modularity
    - One detector → one (or more) software suite maintainers
    - Maintainer(s) → full control over their own calibration software
  - This is not so great for:
    - Generalizing, e.g., the build system
    - Standardizing / sharing common things
    - Automated testing
    - Versioning and releasing → ~10 repos means ~10 releases

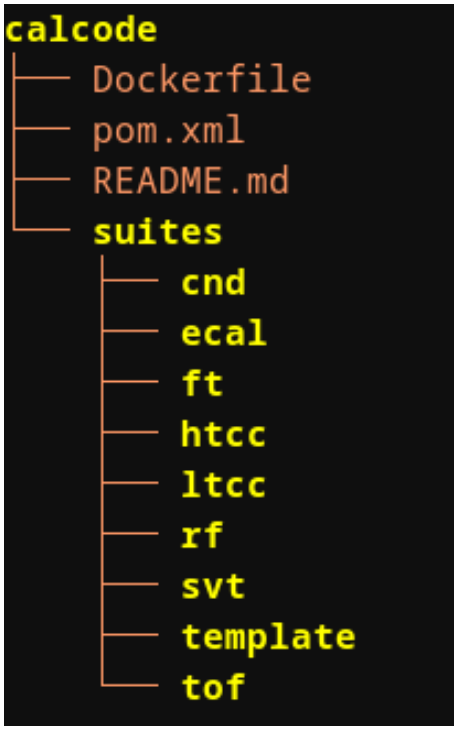



# From Multi-repo to Mono-repo

New Repository: **https://code.jlab.org/hallb/clas12/calibration/calcode**

- Multi-repo: one repository per suite
- Mono-repo: one repository (“repo”), each detector suite in a subdirectory
  - New repo is hosted at code.gitlab.org
    - Log in with your JLab account
    - The new name is “**CALCODE**”
- Many calibration-suite repos have been imported, with git history preserved
  - For example, this allows one to answer the question “where did this line of code come from?”
- Developers should make branches for their work, following standard ‘git’ workflows (branches, merge requests, etc.)

**Calcode Layout**



# CALCODE Status**

| Suite | Status* | Maintainer(s) | Original Repository | Notes |
|---|---|---|---|---|
| AHDC | ❓ | Michael Paolone | | Experiment ongoing |
| ATOF | ❓ | Michael Paolone | | Experiment ongoing |
| BAND | 🕘 | Florian Hauenstein | Link | |
| CND | ✅ | Silvia Niccolai | Link | |
| DC | 🕘 | Veronique Ziegler | Link | |
| ECAL | ✅ | Cole Smith | Link | Work continues on `calcode` fork |
| FT | ✅ | Raffaella De Vita, Nicholas Zachariou | Link | |
| HTCC | ✅ | Tatsuhiro Ishige | Link | |
| LTCC | ✅ | Valerio Mascagna, Maurizio Ungaro | Link | |
| RF | ✅ | Raffaella De Vita | Link | |
| RICH | 🚧 | Marco Mirazita | | Importing to `calcode` on `RICH` branch |
| RTPC | ❌ | Mohammad Hattawy | | Experiment done; standalone suite preferred for now |
| SVT | ✅ | Yuri Gotra | Link | |
| TOF | ✅ | Daniel Carman, Aron Kripko | Link | |

**Status symbols:

- ✅ = actively maintained in `calcode`
- 🚧 = importing to `calcode` is in progress
- 🕘 = not yet ready for `calcode` import
- ❌ = standalone suite preferred for now
- ❓ = unknown

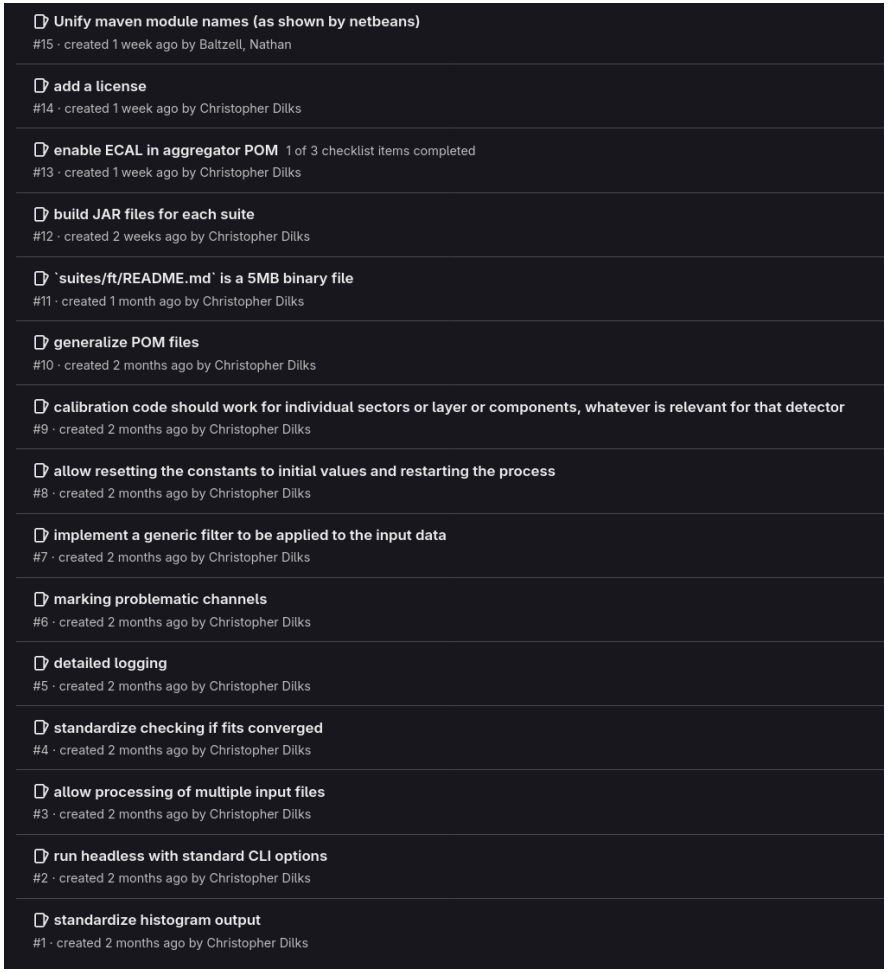
Maintainer names get highlighted automatically when they make an account (just log in)

**For the most up-to-date status, see the main README at: https://code.jlab.org/hallb/clas12/calibration/calcode

# Next steps

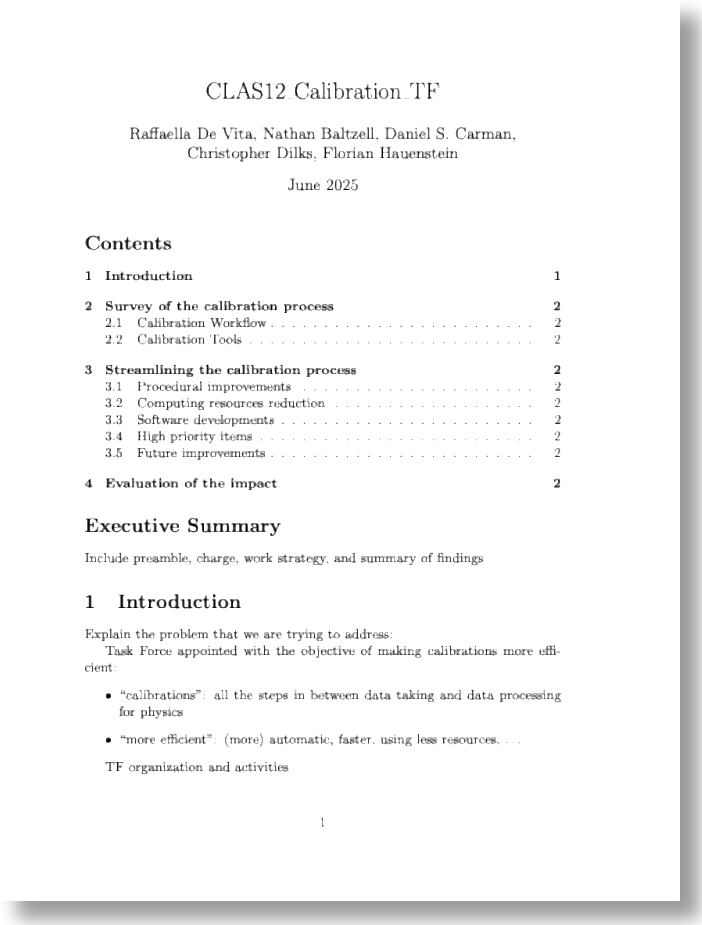
- Using “Issues” to track
  - things to do
  - problems
  - Requests
- Main focus on upgrades necessary to minimize human intervention and move towards automation:
  - extend software functionalities
  - improve robustness of analysis algorithms
  - implement clear metrics for evaluating the results quality
- Proof of principle with “automation” of RF and FTCal (the “simple” ones), TOF as ultimate goal
- Work in progress…

https://code.jlab.org/hallb/clas12/calibration/calcode/-/issues



# 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:
  - Calibration checklist
  - Data volume reduction
  - New calibration code repository
- Now moving the focus to “automating” calibration tools
- Task Force report drafting started



CLAS12 Calibration TF

Raffaella De Vita, Nathan Baltzell, Daniel S. Carman, Christopher Dilks, Florian Hauenstein

June 2025

Contents



Executive Summary

Include preamble, charge, work strategy, and summary of findings

1 Introduction

Explain the problem that we are trying to address:
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, . . .

TF organization and activities

1