Hall D Computing

Alexander Austregesilo
Thomas Jefferson National Accelerator Facility

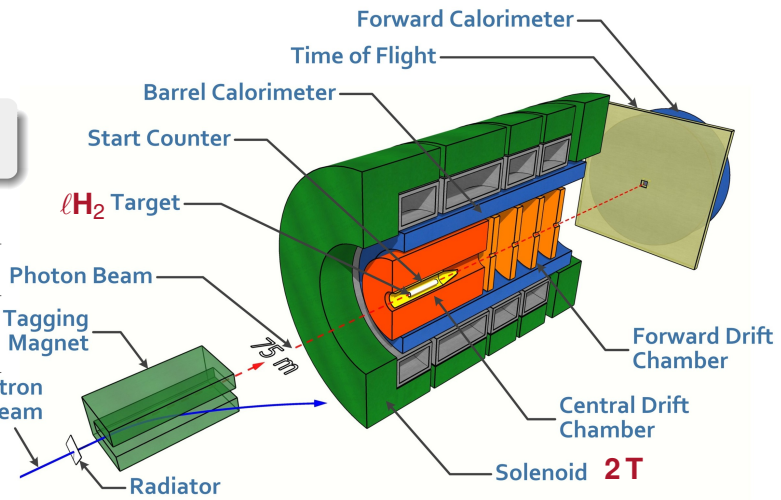
JLab Software and Computing Workshop
May 19th, 2023

GlueX Detector

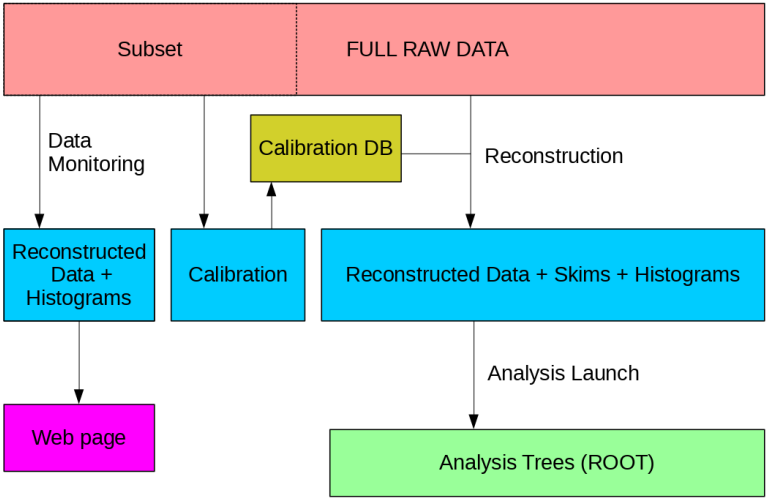
Jefferson Lab
Thomas Jefferson National Accelerator Facility

Light quark meson spectroscopy
with **nearly complete coverage**

Timeline	Events ($\times 10^9$)	Status
2017	47	published
2018	223	analyzed
2020	326	analyzed
2023	153	calibrating
2024	500	scheduled
GlueX-III	?	

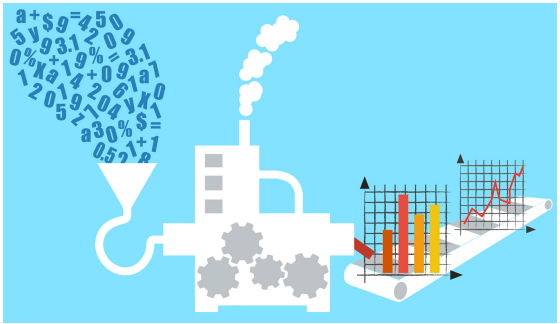


GlueX Processing Overview



Outline

Software and Infrastructure



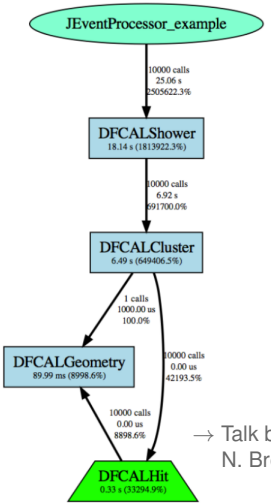
- C++ framework used for monitoring, calibration, reconstruction and first step of analysis
- Multi-threaded to optimize processing rate on modern multi-core machines

Event Processor

- Drives the program, one instance shared between all threads
- Usually implemented in a plugin
  → Library loaded at runtime, no recompiling / relinking

Factory Based

- All data objects inherit from JObject
- Created on-demand for each event
- One instance for each thread
- Tagging allows for multiple versions of algorithms:
  e.g. DBCALShower_factory_KLOE



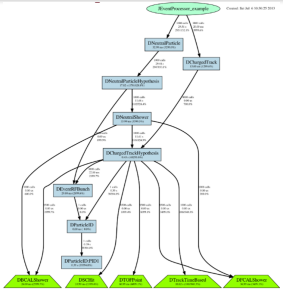
→ Talk by N. Brei

Event Reconstruction

Jefferson Lab
Thomas Jefferson National Accelerator Facility

REST Production

- Primary executable: hd_root (JApplication)
- Beam Photon
- Calorimeter Shower
- Time-based tracks
- Particle ID Hits
- Saved in REST format



Automated Reconstruction Test

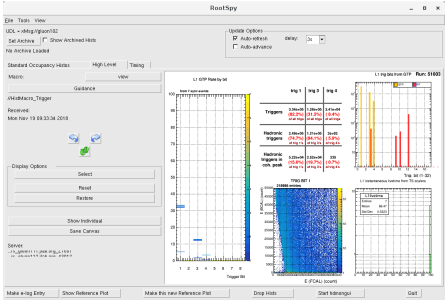
- Reconstruction software under active development, managed on github
- Software stack automatically built every midnight and after every pull request
- Automatically reconstruct same EVIO file for each run period every 3 days
- ROOT file and histograms available online and sent per email

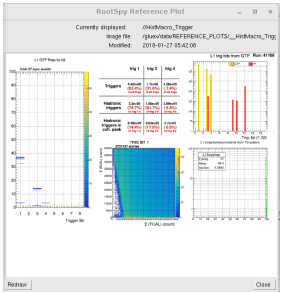
⇒ **Early problem detection**

Online Monitoring and Calibration

RootSpy

- C++ software package based on ROOT
- Incoming data stream analyzed on multiple nodes, monitor histograms during filling
- Communication over network via xMsg
- Reference plots for shift crew





→ D. Lawrence

`www.jlab.org/RootSpy`

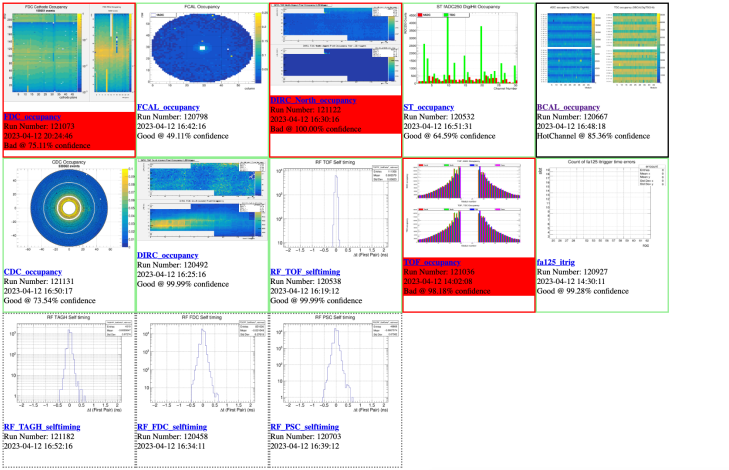
Real Time Data Quality Monitoring

Hydra

Jefferson Lab
Thomas Jefferson National Accelerator Facility

GlueX 121073 Last Updated: 44778.74 second(s) ago HYDRA

showing 13 / 13 frames



TPC_occupancy
Run Number: 121073
2023-04-12 20:24:46
Bad @ 75.11% confidence

FCAL_occupancy
Run Number: 120798
2023-04-12 16:42:16
Good @ 49.11% confidence

BSC_NoHit_Occupancy
Run Number: 121122
2023-04-12 16:30:16
Bad @ 100.00% confidence

ST_occupancy
Run Number: 120532
2023-04-12 16:51:31
Good @ 64.59% confidence

BCAL_occupancy
Run Number: 120667
2023-04-12 16:48:18
HotChannel @ 85.36% confidence

CDC_occupancy
Run Number: 121131
2023-04-12 16:50:17
Good @ 73.54% confidence

DIRC_occupancy
Run Number: 120492
2023-04-12 16:25:16
Good @ 99.99% confidence

RF_TOF_selftiming
Run Number: 120538
2023-04-12 16:19:12
Good @ 99.99% confidence

TOF_occupancy
Run Number: 121036
2023-04-12 14:02:08
Bad @ 99.18% confidence

fa125_ttrig
Run Number: 120927
2023-04-12 14:30:11
Good @ 99.28% confidence

RF_TAGH_selftiming
Run Number: 121182
2023-04-12 16:52:16

RF_FDC_selftiming
Run Number: 120458
2023-04-12 16:34:11

RF_PSC_selftiming
Run Number: 120703
2023-04-12 16:39:12

Computer Vision

- Training of neural network model with existing monitoring plots
- AI classifies incoming data quickly and consistently
- Website alarms shift crew

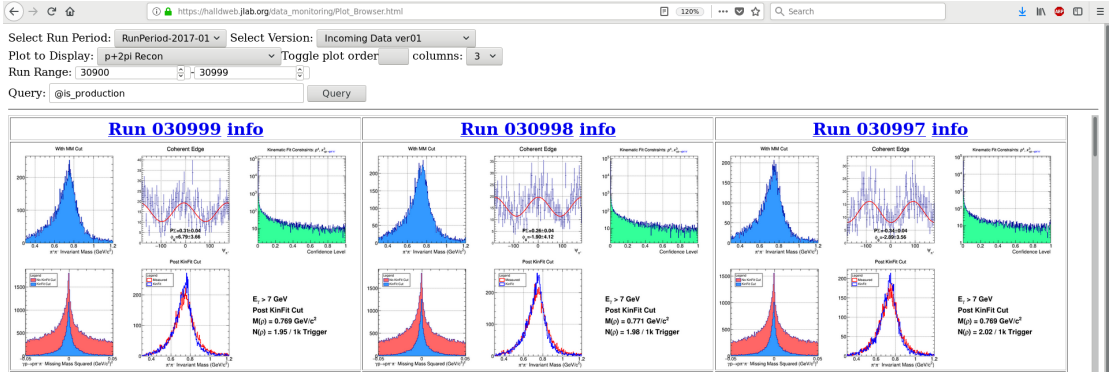
Browser-Based Monitoring

Jefferson Lab
Thomas Jefferson National Accelerator Facility

- Automatic processing of first five files of each run
- Low- and high-level histograms displayed on webpage
- Same set of histograms are produced for each iteration of processing

Browser-Based Monitoring

- Automatic processing of first five files of each run
- Low- and high-level histograms displayed on webpage
- Same set of histograms are produced for each iteration of processing



Goal: optimize calibration workflows for timely data production

Limitation: speed of track Reconstruction

Online Calibrations

- Automatically perform well-understood, run-dependent calibrations
  Examples: timing calibrations, drift chamber gains
- Results deployed to calibration database, logbook

Modified Tables

/PHOTON_BEAM/RF/time_offset

Row,Column	Old Value	New Value
[0] TAGH	-43.970043	-11.814913
[0] PSC	-52.329247	51.894828

/CDC/base_time_offset

Row,Column	Old Value	New Value
[0] CDC_BASE_TIME_OFFSET	-221.043	-367.3075718945941

Calibration

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Goal: optimize calibration workflows for timely data production

Limitation: speed of track Reconstruction

Online Calibrations

- Automatically perform well-understood, run-dependent calibrations
  Examples: timing calibrations, drift chamber gains
- Results deployed to calibration database, logbook

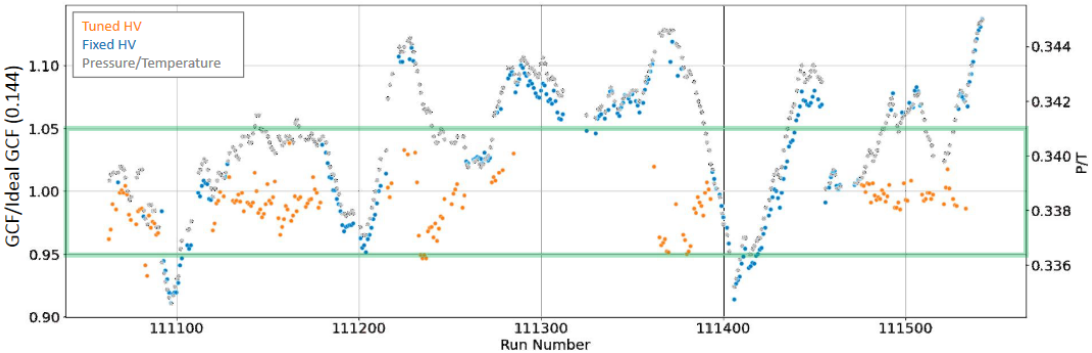
Offline Calibrations

- Run after online calibrations complete and data is copied out of the counting house
- $\approx$10% of data reconstructed for more complicated calibration processes
  Example: pi0 for calorimeter calibrations
- Iterative procedures worked into monitoring workflow
- Generate additional data streams for special trigger types, calibration tasks

AI-Driven Experiment Control

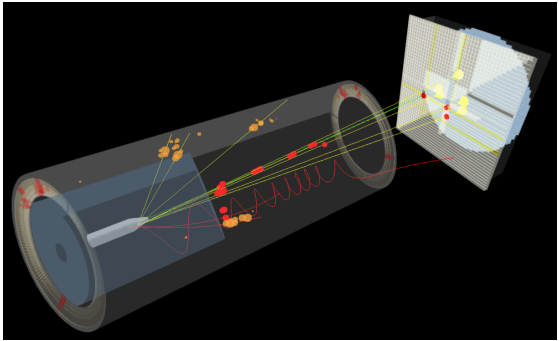
AIEC

Jefferson Lab
Thomas Jefferson National Accelerator Facility



- Gains for Central Drift Chamber have large variation, detailed calibration necessary
- Idea: tune detector parameters to counteract environmental changes
- Use ML to determine high voltage and calibration constants
- In development for other parts of detector, e.g. π^0 calibration for FCal

Event Reconstruction



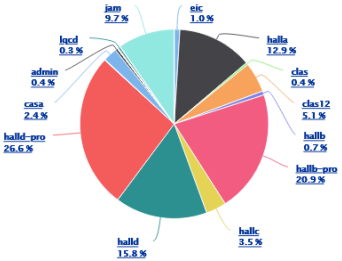
(click on the image for interactive 4D event display)

Reconstruction at Jefferson Lab

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Local Computing

- ≈70M CPUh at JLab in 2022, similar in previous years
- Scientific Computing farm is constantly growing



Local Computing

- ≈70M CPUh at JLab in 2022, similar in previous years
- Scientific Computing farm is constantly growing
- Total usage by Hall D of Jefferson Lab farm close to 50% (central monitoring, production, analysis + users)

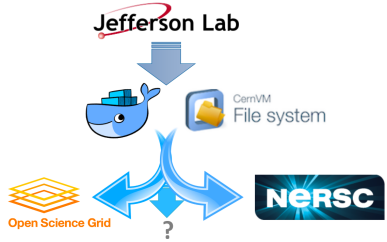
⇒ **Necessity to expand to alternative resources**

Offsite Data Production

Jefferson Lab
Thomas Jefferson National Accelerator Facility

Prerequisites

- Docker (Singularity, Shifter) container
- CernVM File System:
  - GlueX software builds and dependencies
  - Calibration constants and resources
- Data transfer with Globus
- Allocations at NERSC, PSC, BigRed (IU)

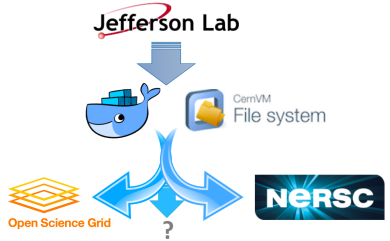


Offsite Data Production

Jefferson Lab
Thomas Jefferson National Accelerator Facility

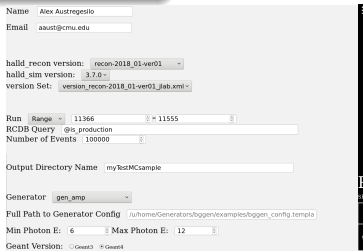
Prerequisites

- Docker (Singularity, Shifter) container
- CernVM File System:
  - GlueX software builds and dependencies
  - Calibration constants and resources
- Data transfer with Globus
- Allocations at NERSC, PSC, BigRed (IU)



MC Simulation @ Open Science Grid

- Request through web form
- Presets for different run periods
- Automatic testing and submission



Central Data Analysis

Analysis Software

Jefferson Lab
Thomas Jefferson National Accelerator Facility

GlueX Analysis Library

- Creates particle combinations out of charged tracks and showers (from REST)
- Common selection criteria for exclusivity and PID
- Supports displaced vertices and missing particles
- Performs kinematic fit: Vertex and P4
- Output saved in common ROOT tree format
- Measures to reduce memory footprint:
  - Objects managed by pools, shared among threads
  - Reuse particle combinations between channels, e.g. $\bar{K}^0$

Example:

$\gamma p \to K^*(892)^0 \bar{K}^0(p)$
$K^*(892)^0 \to K^+\pi^-$
$\bar{K}^0 \to \pi^+\pi^-$

Need: 2 q+'s, 2 q-'s
Measure: 3 q+'s, 2 q-'s
Test each q+ as K+ (3x), π+ (2x)
Test each q- as each π-: 2x
Total Combinations: 12

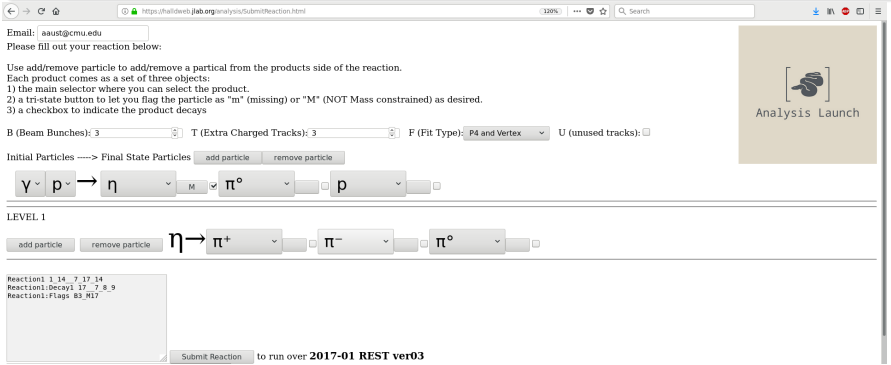
ReactionFilter Plugin

- One plugin, that performs analysis and generates trees for arbitrary reactions
- Steered by human readable configuration file
  Example: `1_14__8_9_14` for $\gamma p \to \pi^+\pi^- p$
- Special flags to select missing or unconstrained particles, type of kinematic fit, etc.

Analysis Launch

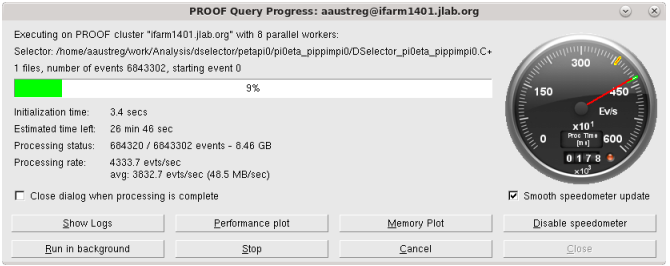
Jefferson Lab
Thomas Jefferson National Accelerator Facility

- Massive parallel campaigns with up to 100 channels
- Efficiently use computing resources and tape access for REST files
- Results in common format, standardized name, accessible for the whole collaboration
- Regular reruns, web form to add channels



GlueX ROOT Analysis

Jefferson Lab
Thomas Jefferson National Accelerator Facility



DSelector

- Inherits from ROOT TSelector: read TTrees, multi-threaded analysis with PROOF-Lite
- Provides C++ interface to TTree data: DParticleCombo, DChargedTrackHypothesis, ...
- Scripts generate reaction-tree specific code with examples
- Common histogram actions, cut actions and other utilities (e.g. coordinate trafo)
- Optional output: TTree with selected events (same format or flat)

Summary

- Production of data, event skims, and MC centrally managed by collaboration
  Do physics, not data production!
- Web forms to request event skims and MC samples
  User does not need to know technical details (to get it right)!
- Workshops and tutorials held regularly, videos and presentations available online
  Low threshold for students and new collaborators!