

# The GlueX Collaboration Meeting

October 4-6, 2012

Jefferson Lab

Curtis Meyer

## The Review

- Was held on June 7-8, 2012 at Jefferson Lab.
- Reviewed all four halls as well as online and the computer center.
- Mark Ito, David Lawrence and Curtis Meyer made presentations.
- Overall the review was very positive, but a LOT of work!

# Presentations

- Morning session ---
  - Overview of the Experiment – Curtis
  - Overview of the Offline System – Mark
- Afternoon parallel with Hall B ---
  - Details of the Offline System - David

“...special thanks to Hall D for early and comprehensive documentation.”

# General Recommendations

- **Presentations in future reviews should cover end user utilization of and experience with the software in more detail.**
- **Talks from end users on usage experience with the software and analysis infrastructure would be beneficial.**

# Hall D Observations



- The framework is written in C++ and is decomposed into a series of components. The event processing framework does event level parallelism, which is the appropriate scope for their problem domain.
- They do a nightly build of the software, which signals experts in the case of problems. In addition, they have twice weekly regression tests using known MC samples in order to find performance problems in new code.
- Calibration and alignment software are in a rather advanced state. Nonetheless completing the development of the calibration software is estimated to be the largest remaining offline software effort to complete in terms of FTE-years.
- They have implemented a run based calibration system, CCDB, which can use ASCII or MySQL back ends. The system keeps a full history and makes it easy to create a new calibration 'era' by cloning an existing era and modifying only the calibrations one is interested in. It also includes an easy to use python shell interface.
- HDDS as the single source of detector geometry description for MC and reconstruction looks very promising. This guarantees that reconstruction and simulation are in synchronization. In addition it should aid in the migration from Geant3 to Geant4.
- Innovative use is made of GPUs for Partial Wave Analysis.
- They are evaluating the use of the visualization library built by CLAS12 in order to do event visualization.

# Findings

- Workflow tools for handling bulk processing on the batch farm are only in conceptual stage.
- The detector simulation is based on Geant3.
- The collaboration has explored using grid resources; this is worthwhile so long as required manpower levels are low, to improve flexibility and capability in analysis.
- The JANA framework is very stable. They report they haven't changed the code for a year.
- Their JANA framework is not specific to GlueX, it could be adopted by others. It hasn't been thus far.

# Recommendations

- **A series of scale tests ramping up using JLab's LQCD farm should be planned and conducted.**
- **The data volume and processing scale of GlueX is substantial but plans for data management and workload management systems supporting the operational scale were not made clear. They should be carefully developed.**
- **Consider ROOT (with it's schema evolution capabilities) as a possible alternative for the HDDM DST format.**
- **To ensure a smooth transition from development and deployment to operations, ... , an explicitly planned program of data challenges, directed both at exercising the performance of the full analysis chain and at exercising the scaling behavior and effectiveness of the computing model at scales progressively closer to operating scale, is recommended. We heard more explicit plans from Hall D..... This data challenge program should be underway now, and should not await the full completion of the offline software.**

# Response of GlueX

- We have set up our “data challenge” working group to move us towards progressively larger data challenges.
  - Jakes work with 3pi in 2011 ~10 hours of beam time
  - Next data challenge should try to increase this by about a factor of 10.
  - Ultimately reach ~1month of simulated data.
- The first issue we encountered was software performance.
  - Tremendous push within the collaboration to understand the problems with tracking. We are now in much better shape than we were.
  - Substantial work has been put into photon reconstruction in both the BCAL and the FCAL.
  - Major push on a more global handling of particle identification has been implemented.
  - Our software is in substantially better shape now than it was in one year ago!



# Response of GlueX

- The limiting issue we encountered was data storage.
  - Richard has developed the REST format for DSTs which is smaller than what we had estimated.
  - Paul Mattione has developed an analysis package and has been stressing the REST format.
  - We are “close” to having the real online format for the experiment.
  - We are now trying to refine the size estimates of our raw data.
- Can we process data on the JLab Farms?
  - What happens when we submit ~1000 jobs to the farm?
  - Can we track the results?
  - Can we recover from errors?
  - How does the software perform?
  - What tools do we need to manage this?

# Response of GlueX

- Generate a very large set of PYTHIA background events on the GRID and make available in REST format.
  - Many of our analyses efforts are now limited by the amount of PYTHIA background we have.
  - We in principle have access to a very large amount of core time on the grid. Again, we are limited by data storage.
  - Once we are happy with REST, we are ready to roll on this.
- Generate a very large set of PYTHIA background events at Jefferson Lab.
  - This effort is aimed more at understanding the issues of processing raw data at Jefferson Lab.
  - We are currently building up to run this and have requested tape and disk resources.
  - We are waiting for the Raw Data.
  - We still need to define what will get kept in this challenge.

# Calibrations

- We identified calibrations as the largest hole in our offline effort.
- We also identified that the natural time for this to move forward is after detectors construction is complete.
- We are moving into that phase now, and need to start moving in the direction of calibrations for GlueX.

# Summary

- The software review was a very useful exercise that made us really look carefully at what we had, what we needed and how we could get to where we needed to be.
- It is fair to say that we are in good shape in this regard, but that does not by any means say we are ready for data.
- We have taken the appropriate response to the review, and I think that we have made substantial progress since June.
- I hope that we will not be going through another review like this anytime soon!