



Open Science Grid

Catching up with GlueX

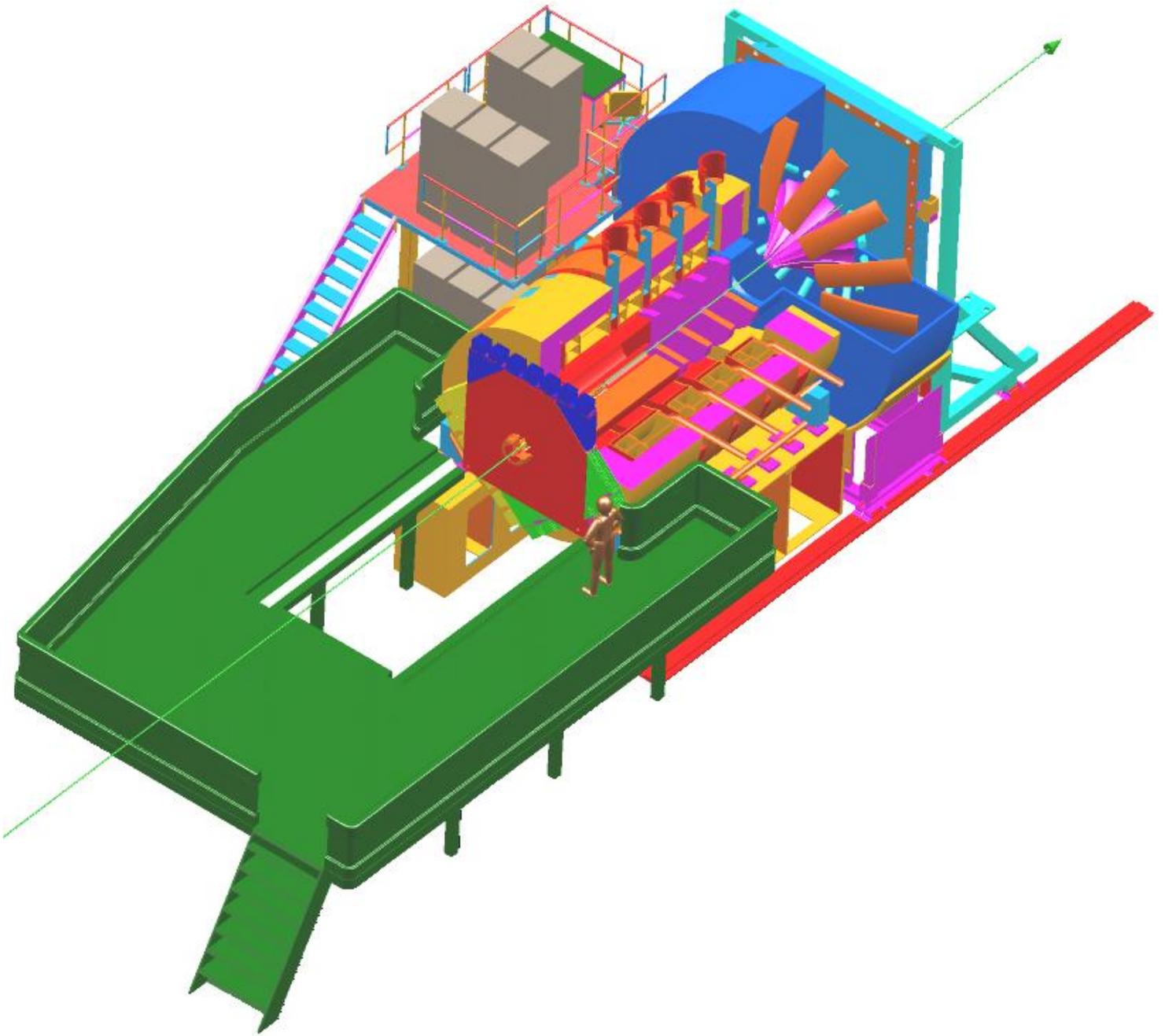


In spring 2014 the GlueX collaboration, a group pursuing physics using photon beams at the Thomas Jefferson National Accelerator Facility (Jefferson Lab) in Virginia, carried out their second data challenge on the Open Science Grid (OSG), and by October they were undertaking their first “commissioning” run. This winter seemed an appropriate time, therefore, to follow up on our February 2013 interview with Richard Jones, an associate professor of physics at the University of Connecticut and a member of the GlueX collaboration.

Image: Professor Richard Jones

In 2013, Jones had just run a data challenge on the OSG to simulate experimental data for GlueX. Jones says the goal of this experiment is to understand the strong forces that bind quarks and gluons together inside nuclei by directly exciting them using linearly polarized photons directed onto a liquid hydrogen target. The excitations are revealed in the pattern of new particles that are emitted from the target when the excitations decay. Detecting these decay products and analyzing their patterns to infer the types of excitations that produced them requires a sophisticated detector, together with a data analysis infrastructure capable of handling very large volumes of data.

We caught up with Jones this winter to ask what they have been learning. Jones reports that GlueX has progressed beyond the construction phase to actually running the beam and detector, its first “commissioning” run having just taken place in October-December 2014. Besides working the bugs out of the new beam and detector systems, commissioning also provides the first opportunity to get real data to compare with the simulations.



(<http://www.opensciencegrid.org/wp-content/uploads/2015/01/GlueX.jpeg.jpg>)

Figure 1: Thomas Jefferson National Accelerator Facility (Jefferson Lab) in Virginia. Image used with permission from Jefferson Lab.

With the first real GlueX data now in hand, he says it's too early to tell whether they are getting what they expected. "The rough answer is yes," Jones said. "We're seeing trackers showing particles coming out of the target. We're seeing the photon beam going where we expect. Now we'll look at the data and get into the details of understanding exactly what it is we are seeing."

This past spring (2014), the group ran a second data challenge on the OSG, similar to the first simulations. "It was like a rerun of the first one but with important improvements based on what we learned in the first challenge," Jones said. "We also made specific upgrades to the physical realism based on the recommendations of our software-readiness review panel. We added new features like noise in the detector and beam background interactions in the target. These were not in the first challenge because they are computationally expensive."

The first data challenge was to prove the software stack could run in production mode. “Production mode has to run smoothly,” notes Jones, “because once the experimental data starts rolling in, we want to be able to analyze them in real time so we can answer the initial scientific questions quickly and direct the future experimental program based on what the experiment is telling us.” The second data challenge served to fine-tune the software. The first run uncovered obstacles that they overcame in the second, having learned a lot about running on the OSG. “Putting together the lessons learned during round one, we achieved a more realistic workflow, and we included those improvements in the simulation software that evolved during the construction phase of the detector,” he says. “We achieved more realism in the trackers, put that in the simulation, and we gained a number of specific improvements in the reconstruction libraries.” Those improvements are now in production.

The new features slow the process of getting data back. To run under full background conditions like noise in the detector requires about five times more processing time as it does to run with no background. Jones says they decided to split the OSG data challenge into three pieces: one at full background intensity, one at one-fifth background, and one at zero background. “Most of the jobs we ran were at the one-fifth background rate because we plan to run at reduced intensity during this phase of the experiment,” Jones says.

“We will run at reduced intensity during GlueX phase one, which covers the first two years. The commissioning phase helps us understand how to proceed. Commissioning the photon beam is not yet at 12 GeV (gigaelectronvolt), which is where we’re headed in 2015. So far we’re running unpolarized at 10 GeV. That’s not really phase one yet because we still have to achieve beam operations with linear polarization.”

For the second data challenge in April 2014, the GlueX team used 3.8 million wall clock hours over four weeks on the OSG, and an average of 5,700 cores simultaneously — with a peak of 11,000 cores running simultaneously. Most of that was opportunistic, with about 600 cores at the University of Connecticut. Northwestern University contributed another few hundred cores and storage for the output files, making the files available at more than one grid-accessible location.

Jones said it was important to know they could do it with several different sites. The Jefferson Lab computer center devoted a large part of their production cluster to executing this data challenge on the OSG, with the Lattice Quantum Chromodynamics (QCD) cluster contributing so they could reach their testing goals. Jones says there is still some skepticism that a grid can be used as a stable platform for production. This data challenge showed they could produce in a heterogeneous environment.

“We also proved that the environment, the middleware, has made it possible to bridge all the heterogeneities in a real production environment, making it possible to manage the differences from site to site,” Jones added. “The OSG produced 80% of the results in the first challenge, but that was without significant dedication of Jefferson Lab resources. For the second challenge, Jefferson Lab resources were mandated by our software review committee, and they came through for us with flying colors.” In the end, the OSG produced about 60% of the results from the second data challenge, with most of the remaining 40% coming from Jefferson Lab.

Jones says the spring 2014 data challenge also broadened key infrastructure partnerships. “We were able to strengthen ties with a University of Connecticut campus in another part of the state (Farmington). They have significant IT resources. We are working on overcoming the barriers to collaboration across the campuses. We received a National Science Foundation grant to build a network infrastructure to facilitate low-latency, high-throughput transfer of scientific data between campuses and over Internet2. The grant has a component that goes to the Health Center on the Farmington campus. I’m anxious to see how OSG storage works using this improved throughput.”

Read more about GlueX (<http://www.gluex.org/Gluex/Home.html>).

~ Greg Moore and Sarah Engel

Funded by National Science Foundation (<http://nsf.gov>) and Department of Energy
(<http://science.energy.gov/>)

© 2013 Open Science Grid | Privacy Policy (<https://twiki.grid.iu.edu/bin/view/Operations/IUPrivacyPolicy>)