Computing for Experimental Nuclear Physics at Jefferson Lab.

Graham Heyes

Data Acquisition Support Group lead

ENP computing coordinator

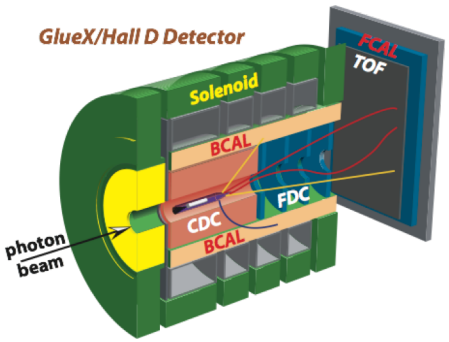
CLAS12 software group co-lead

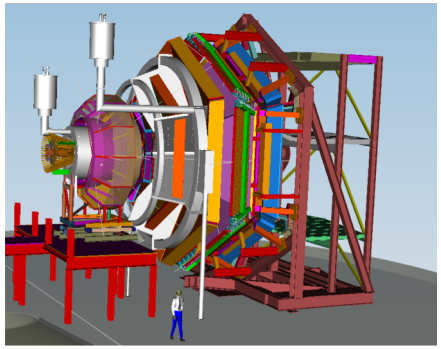
Jefferson Lab

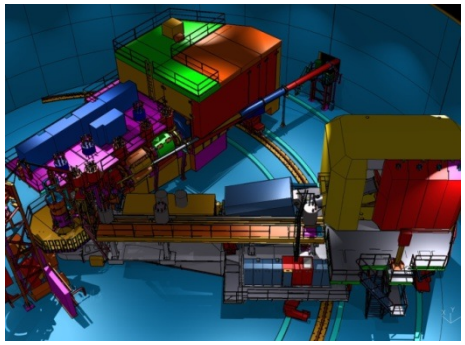
Outline

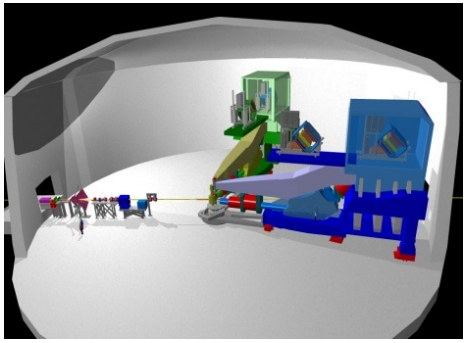
- Brief introduction to the four experiment halls, the detectors, experiments and how things are organized.
- The various computing challenges and requirements that the physics program presents.
- The past approach to computing and lessons learned.
- Where we are now.
- Where we would like to be – near term.
- How are we going there.
- Where we may go in the future.
- A slide for people who like a summary.

Detector capabilities









Hall D	Hall B	Hall C	Hall A
excellent hermeticity	luminosity 10^{35}	energy reach	custom installations
polarized photons	hermeticity	precision	
E_γ~8.5-9 GeV	11 GeV beamline		
10^8 photons/s	target flexibility		
good momentum/angle resolution		excellent momentum resolution	
high multiplicity reconstruction		luminosity up to 10^{38}	
particle ID			

Jefferson Lab

12 GeV Approved Experiments by Physics Topics

Topic	Hall A	Hall B	Hall C	Hall D	Other	Total
The Hadron spectra as probes of QCD (GluEx and heavy baryon and meson spectroscopy)		1		2		3
The transverse structure of the hadrons (Elastic and transition Form Factors)	4	3	2	1		10
The longitudinal structure of the hadrons (Unpolarized and polarized parton distribution functions)	2	2	6			10
The 3D structure of the hadrons (Generalized Parton Distributions and Transverse Momentum Distributions)	5	10	4			19
Hadrons and cold nuclear matter (Medium modification of the nucleons, quark hadronization, N-N correlations, hypernuclear spectroscopy, few-body experiments)	4	2	6		1	13
Low-energy tests of the Standard Model and Fundamental Symmetries	2			1	1	4
Total	**17**	**18**	**18**	**4**	**2**	**59**

12 GeV Approved Exp. by PAC Days

Topic	Hall A	Hall B	Hall C	Hall D	Other	Total
Hadron spectra as probes of QCD	0	219	11	540	0	**770**
Transverse structure of the hadrons	150.5	85	110	25	0	**370.5**
Longitudinal structure of the hadrons	65	230	211	0	0	**506**
3D structure of the hadrons	409	872	197	0	0	**1478**
Hadrons and cold nuclear matter	220	275	205	0	14	**714**
Low-energy tests of the Standard Model and Fundamental Symmetries	547	180	0	79	60	**866**
Total Days	**1392**	**1861**	**734**	**644**	**74**	**4704.5**
Total Days - (includes MOLLER)	**917.5**	**1861**	**734**	**644**	**28**	**4184.5**
Total Approved Run Group Days (includes MOLLER)	**917.5**	**1026**	**691**	**444**	**28**	**3106.5**
Total Days Completed	**165.5**	**77**	**94.5**	**120**	**0**	**457**
Total Days Remaining	**752**	**949**	**596.5**	**324**	**28**	**2650**

- **1 PAC day ~ 2 calendar days.**
- **Backlog: 7 to 9 yrs depending on accelerator availability per year.**
  - Run multiple halls In parallel ~300 to ~370 PAC days per year.

Some numbers

- The halls vary in the scale and longevity of experiments.
  - Halls A and C
    - High turnaround of short lived individual experiments, weeks or months.
    - Typically high luminosity but simple, and therefore small events.
    - Data rates < 100 MB/s at 5kB/event.
  - Hall B
    - General purpose detector – group experiments in “run groups” that share a common dataset, typically few months to a year each.
    - Data rates ~300 MB/s (13 kHz event rate, 22 kB/event).
  - Hall D
    - GLUEX represents a single experiment.
    - Low and high intensity phases running for years each.
    - Up to 1.5 GB/s data rate (90 kHz event rate, 17 kB/event) in high intensity phase.

Computing requirements

- Four main components to the computing workflow :
  - **Calibration**, calculate parameters used to convert raw data from detectors into derived values like momenta, energies, particle ID etc.
    - Done using a sample of the dataset ~5%.
  - **Reconstruction**, apply the calibration to the raw data and reconstruct the individual events.
    - Applied to all of the data, maybe more than once.
  - **Analysis**, take the output of reconstruction and extract the physics measurements.
    - Applied to a varying fraction of the reconstruction output depending on physics of interest.
  - **Simulation**, use MC algorithms from standard packages such as GEANT to generate a simulated dataset.
    - Has a very small input but the output scales with the size of the dataset taken by the experiment.
- The computing resource requirements scale linearly with the number of events in the dataset.
- When calculating requirements it is useful to group calibration, reconstruction and analysis as non-MC workload since they require access to the data from the experiment

Compute requirements process

- Each hall is encouraged to “own” their computing requirements.
- Up to last year the computing requirements were, for the most part, calculated based on expected rates and assumptions about workflow.
- The calculations involved are rather simple and can be done with a spreadsheet or script. For any component of the workflow :

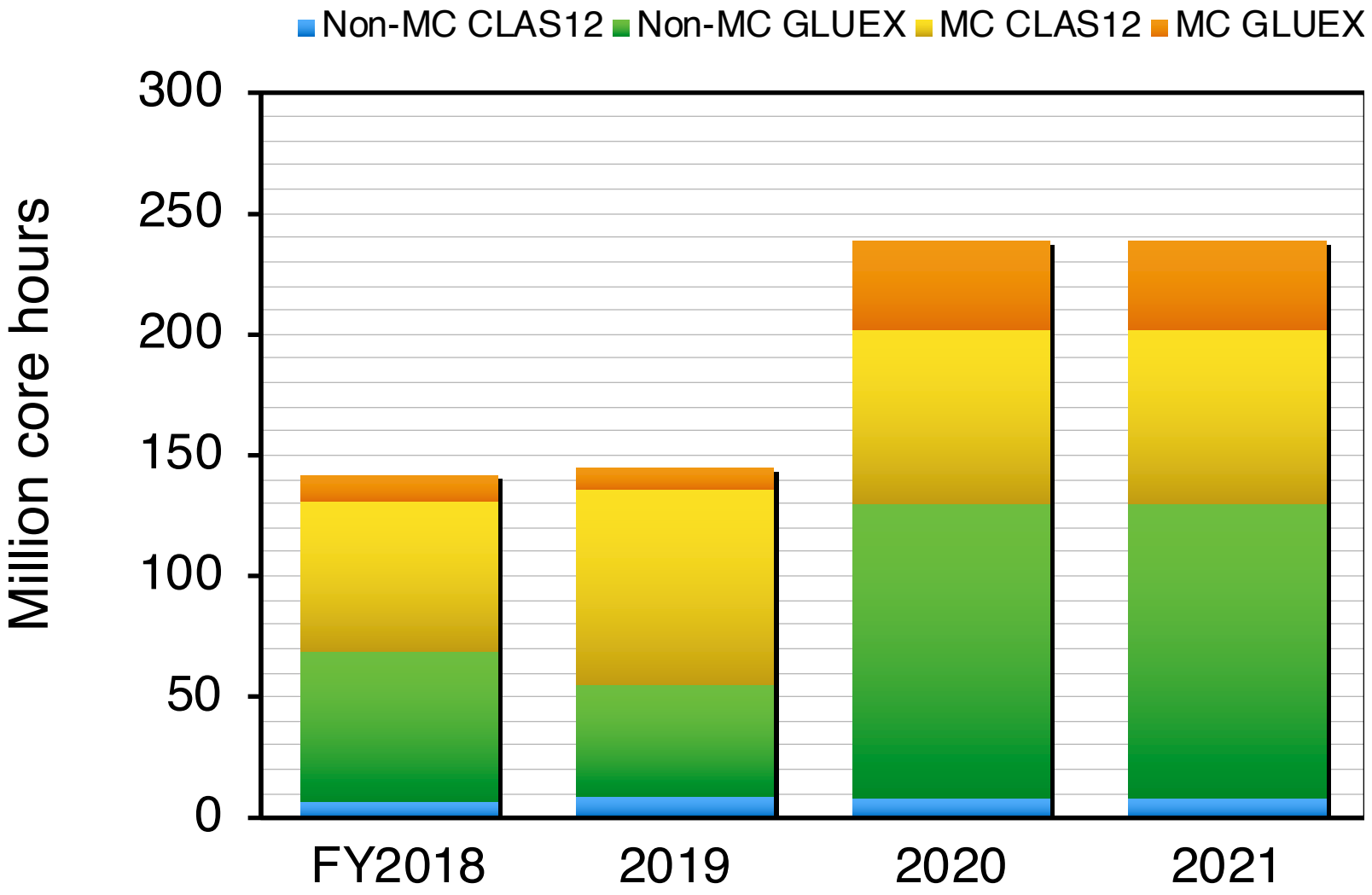
  Computing load = (compute time per event) * (events in total dataset) * (scaling factor)

- The “art” is in knowing what the scaling factors are.

- By last year GLUEX had significant experience of real workflow patterns and how their offline software and DAQ behave.
  - They now have a script that calculates computing requirements based on a set of observables.

- CLAS12 are not yet this advanced but the requirements calculations are now based on real experience from recent post-commissioning runs.

Summary in chart form



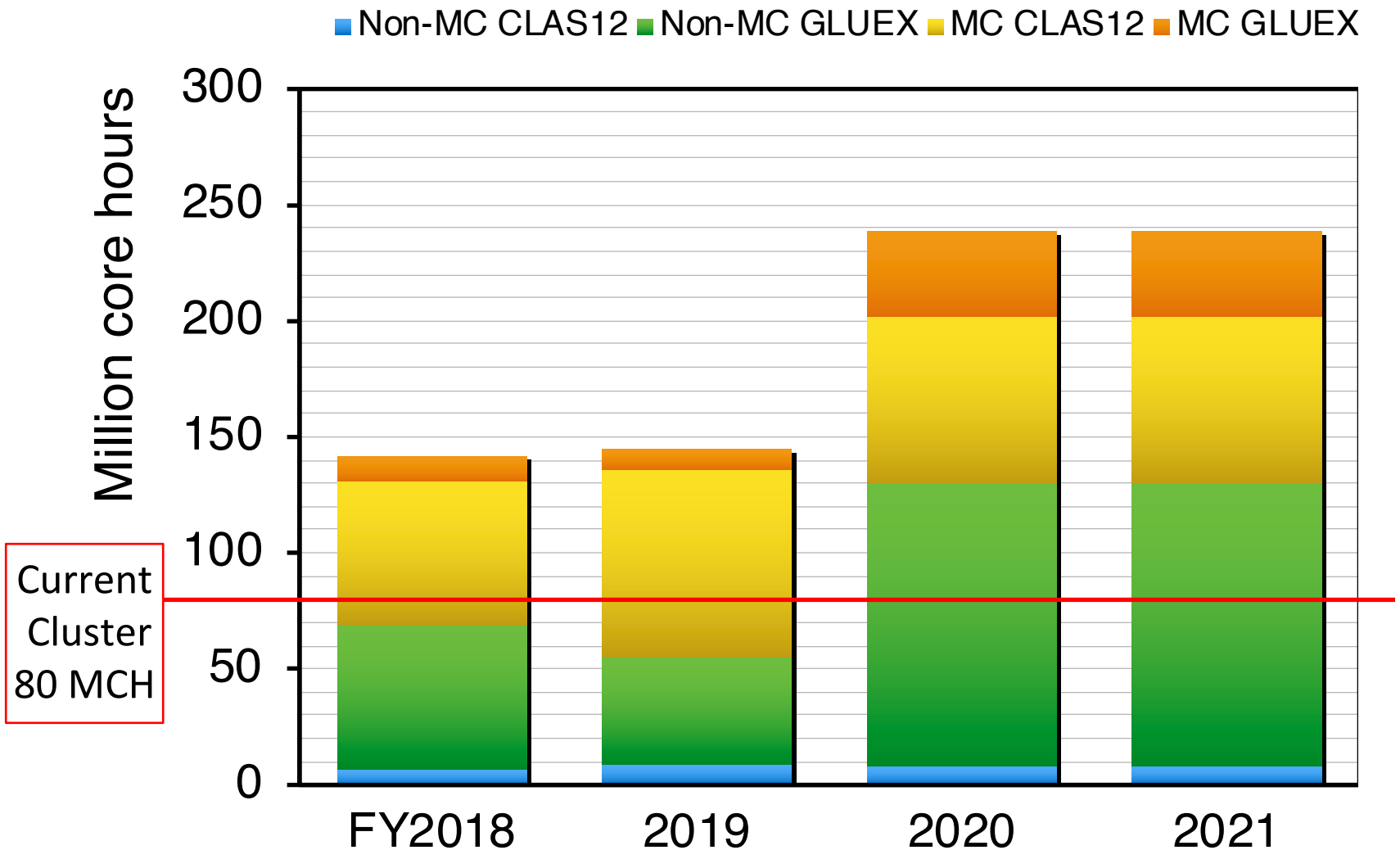
Jefferson Lab

Provisioning for the workload

- The approach at JLab during the 6 GeV program was:
  - A local compute cluster large enough to perform the entire workflow.
  - All the raw and processed data stored at JLab in a tape library.
  - Enough disk to keep the whole thing humming along.

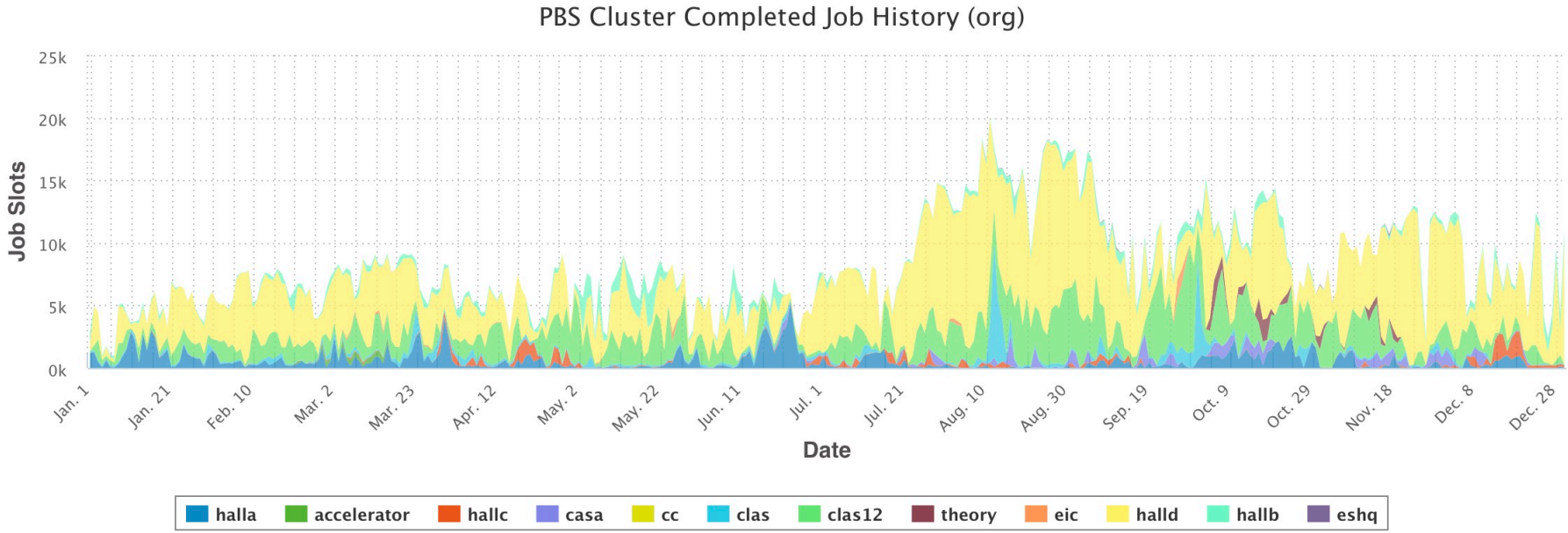
- We continued that into the 12 GeV era and currently have:
  - Compute : 80M core hours per year (Broadwell equivalent)
    - 5000 cores, mix of Skylake, Broadwell, Haswell
  - Disk : 1 PB ENP, of Lustre, ZFS with14 GB/s capacity
    - Soon to be almost doubled.
    - Contains a mix of data from tape, scratch, work
  - Tape: 200 PB library capacity if all LTO-8.
    - but LTO migration expected before that happens
    - 5 GB/s across LTO-5 / 6 / 8 generations today
  - Network
    - 40gE DAQ Halls to tape stage disks; IB (FDR, QDR mix) to data movers
    - **10gE offsite to ESNet**

Where we need to be



Jefferson Lab

Why not just expand?



- A combination of the workflow and number of small experiments make the load on the cluster bursty.
  - Careful calibration leads to a delay before data can be processed.
  - Hard to be efficient with many small experiments.
  - Don't want to wait 12 months for the reconstruction to complete.
    - Wasteful to complete a year's work in 3 months and sit idle for 9.

Jefferson Lab

Lessons learned

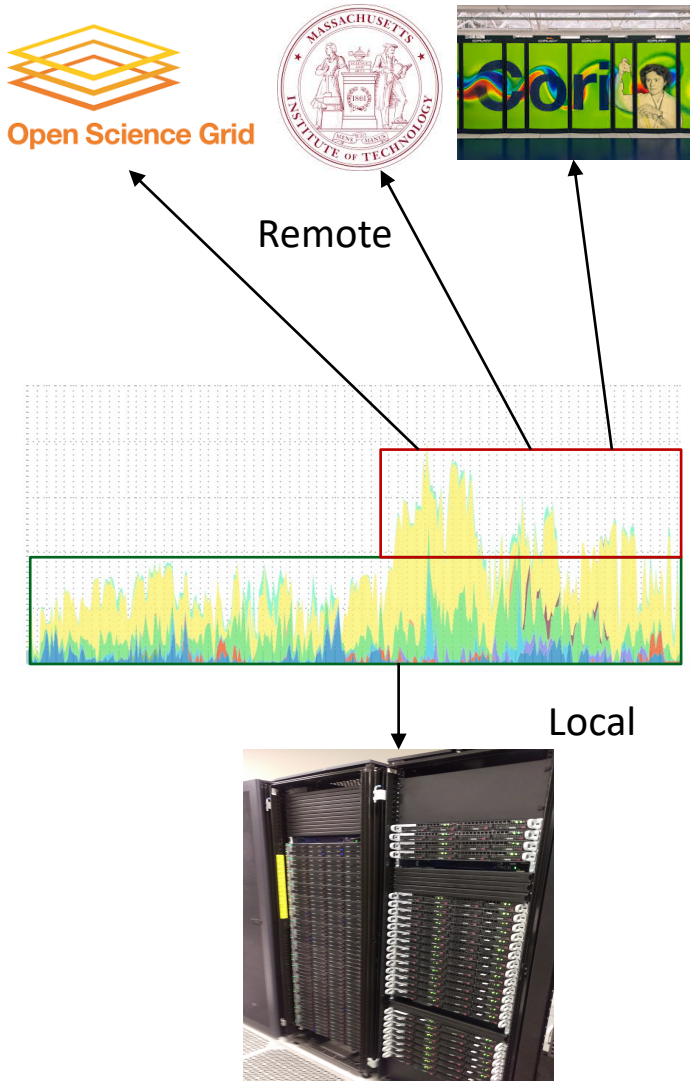
- A combination of science workflow and the number of independent experiments makes the load bursty.

- Calculation of computing requirements looks simple but depends on scaling factors that vary from experiment to experiment and are hard to guess up front.
  - GLUEX has the best estimates but they are also close to being a single experiment.

- Budgets and staffing levels make it hard to argue that we provision for the peaks.
  - In the 6 GeV era the background “hum” of users submitting small jobs could fill in the valleys, there just isn’t enough of that at the moment.

Share the load

- Better to provision locally to fit the usual daily load and use offsite resources for the peaks.
- Going forward the computing model is :
  - Provision to do locally as much of the processing tied to the raw datasets as possible.
    - Doing all would require locally 130 MCH (or less as code improves).
    - Offsite is limited by 10 Gb/s WAN connection.
  - Plan to do offsite as much of the simulation as possible .
    - Total 109 MCH.
  - Plan to expand the WAN capacity to support doing more offsite in the future.
- Offsite compute resources will be a mix of :
  - Locally at collaborating labs/universities (not via OSG).
  - Remote via OSG (includes collaborators via OSG).
  - Remote at supercomputer centers such as NERSC.



Where we are now

- GLUEX has been using OSG for some time now (see next talk).
  - Per year resources :
    - UConn - 10 MCH
    - FSU - 5 MCH (so far, more on the horizon)
    - Northwestern - 2 MCH
    - Regina - 2 MCH (so far, maybe more can be found)
    - Indiana - 4 MCH
    - Florida International - 2 MCH
    - Opportunistic OSG – 10 MCH (rough estimate, from experience so far)
  - **Total anticipated : 35-50 MCH – close to requirement**

- NERSC for processing raw data (their preference).
  - GlueX can utilize up to 30 MCH per month with existing 10 Gbps WAN. Depends on allocation, 2018 got ~16-45 Mhr*

- CLAS12 started data taking later but is following GLUEX's lead.

- So far the use of OSG is not well integrated and relies on "volunteer administrators."

* CORI-II is KNL which our code runs more slowly on than CORI-I which has Haswell processors similar to the ones we have locally.

Where we want to be

- We should enable our users to submit “jobs” at JLab in a uniform way independent of where they run.
  - Choose where to run or let the system decide based on load and priority?
  - Properly manage data and work flows automatically. *
- We should enable our collaborating institutions to easily make resources available for use by the collaboration.
  - Turnkey setup.
  - Meet needs of funding agencies by managing resources. Some sites are restricted from making resources “freely” available via OSG.
- We should at least pay back any opportunistic resources that we take advantage of.
  - Suggestion from the 2018 computing review.

* Automatically because we have a relatively small scientific computing staff. We are not able to do a lot of manual interaction.

How are we planning to get there?

- JLab Scientific Computing have developed a tool, SWIF, to manage the typical NP data processing workflow.
  - Originally designed for use with the local batch system.
  - SWIF 2 has been designed to allow incorporation of offsite resources into SWIF workflows.
  - SWIF 2 does NOT yet support OSG.

- The two main JLab software frameworks, CLARA for Hall-B and JANA for GLUEX, are both designed bottom up to support multi threaded parallel processing of NP data.
  - Both frameworks are used in a mode where “jobs” request whole nodes and well defined memory footprints. These are not small single thread jobs.
  - We are also gaining experience in code optimization that uses thread affinity and knowledge of node architecture.

- SWIF would have to be able to request sole access to nodes of a known architecture via OSG.

Were going to need some tools

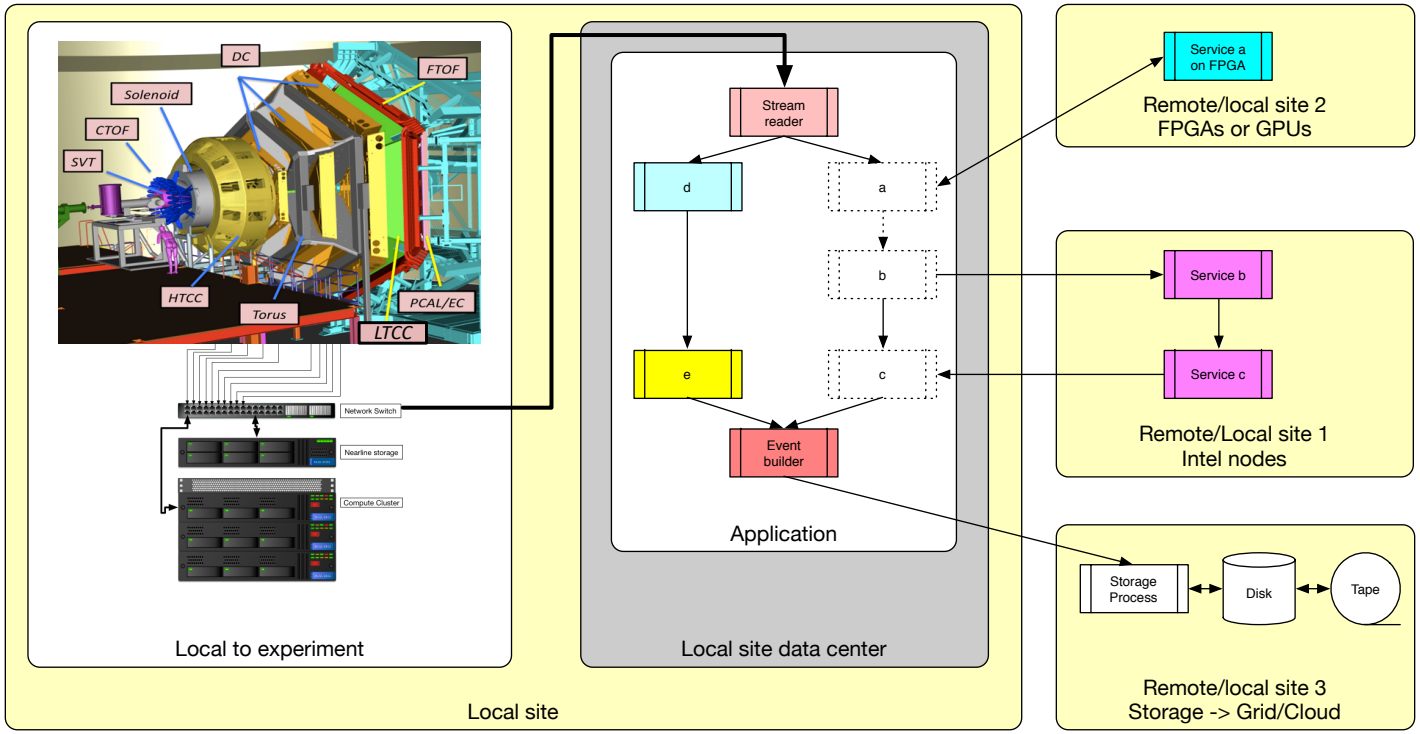
- Since we are still relatively new to this we are learning what is and isn't available.
- We are containerizing our software and learning how to streamline that.

- GLUEX have made excellent use of CernVM-FS as a read-only repository for executables, libraries and configurations.

- We need to get together with our community to firm up the ideas on what an ideal system would look like then with the broader communities, such as OSG and ASCR, to see what fits.

Future concepts

- Currently we run an entire application on OSG or at NERSC.
  - Is there a better way?
- We could reimagine applications as nets of services processing streams of data objects.
  - Standardized application building blocks, one data type in, another out.
  - Route data to services running on appropriate hardware. Parts of algorithm run on Intel, others on FPGA or GPU.
  - Integrate Machine Learning hardware and software.
- We have begun a ”Grand Challenge” project that will investigate this type of computing model.



Jefferson Lab

Summary

- For the first 20+ years of the labs history the bulk of the compute was done locally.
- The scale of 12 GeV era experiments and collaborations have lead to a model where some or most of the compute has to use offsite resources.
  - Fortunately these resources exist, the question is how to take advantage of them?
- The “holy grail” is an completely integrated system where the user can get on with the science and the system figures out the best place(s) to run the workflow.
- To get there we have a lot of ideas but need help and cooperation to convert them into reality.

Experimental studies at low Q^2 of the spin structure of the nucleon at Jefferson Lab

A. Deur[*]
Thomas Jefferson National Accelerator Facility
E-mail: deurpam@jlab.org

We summarize the experimental program of Jefferson Lab that studies the nucleon spin structure at low Q^2. This program completes the precise experimental mapping of the nucleon spin structure functions $g_1(\nu, Q^2)$ and $g_2(\nu, Q^2)$ and their moments started at SLAC, CERN and DESY at high Q^2, and continued at Jefferson Lab at intermediate Q^2. The results presented cover the domain where Chiral Effective Field Theory (χEFT) should describe the strong interaction. They provide a comprehensive set of benchmark measurements for χEFT. The preliminary conclusion is that nucleon spin structure data are still challenging for χEFT in spite of the notable improvements in these calculations.

Jefferson Lab