CLAS12 Simulations on the OSG

Maurizio Ungaro

Jefferson Lab

OSG All-Hands Meeting

March 3, 2021

- Introduction: Jefferson Lab, CLAS12
- CLAS12 Simulations needs, and the old ways to run
- The OSG “Revolution”
- CLAS12 OSG Portal: web interface and backend
- CLAS12 Project Computing Resources on the OSG
- Summary

Jefferson Science Associates, LLC

Thomas Jefferson National Accelerator Facility

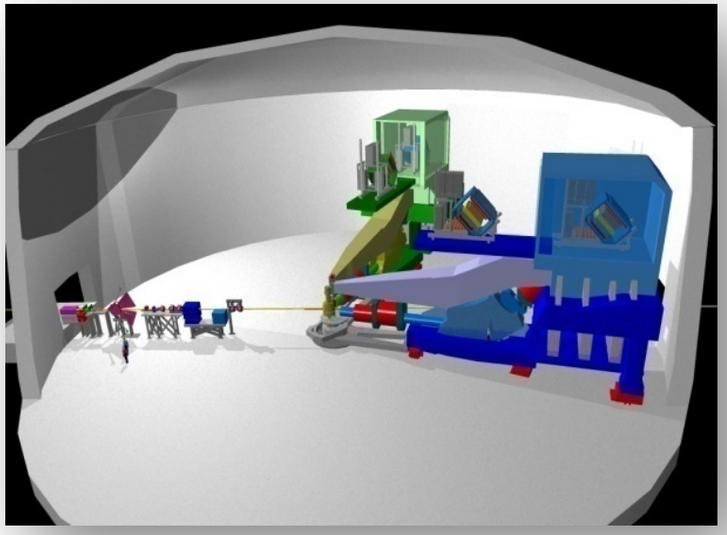
Jefferson Laboratory, Newport News, VA, USA

- Unique Superconducting electron-accelerating technology. Max energy of 6 GeV (up to 2012) and now 12 GeV
- Construction started in 1987
- Funded by the Department of Energy
- Started operation in 1997

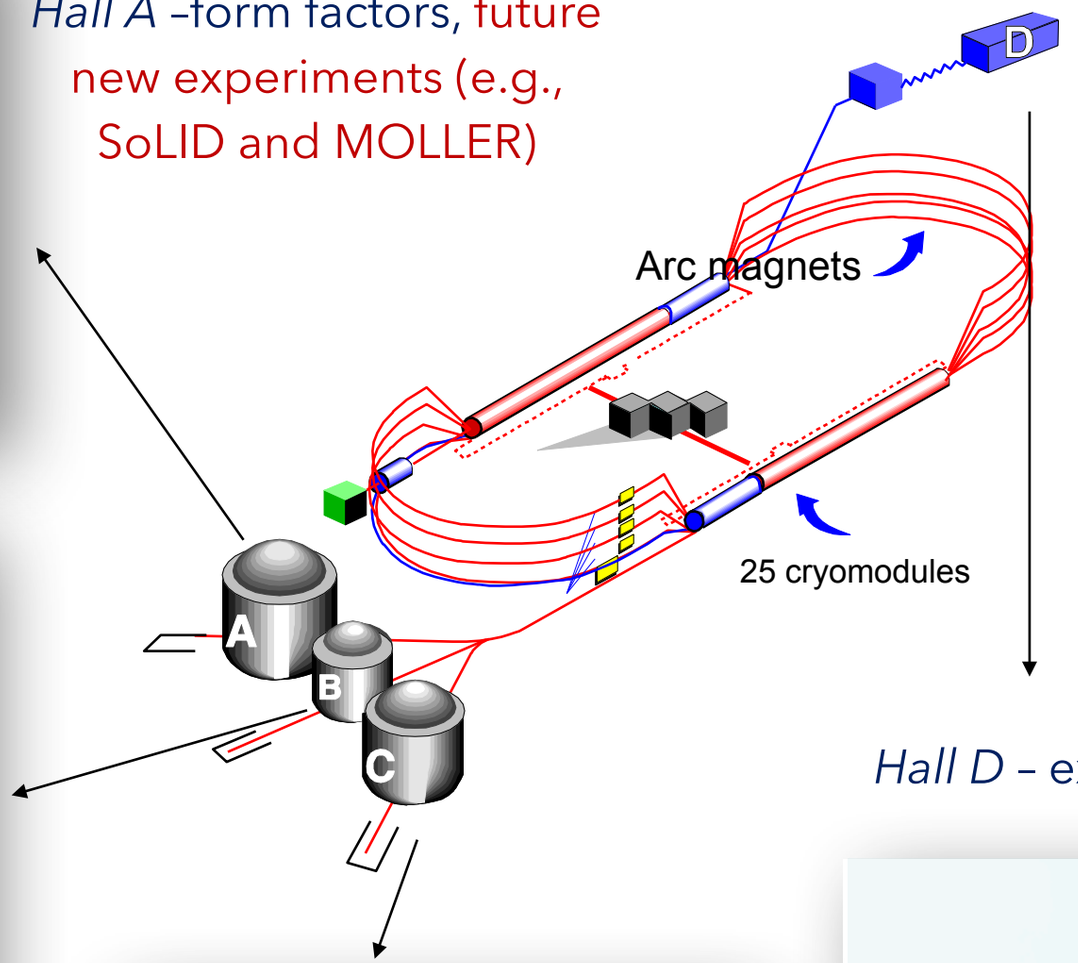
Approximately 700 people are employed at Jefferson Lab.

Approximately 1,300 scientists from around the world conduct experiments at Jefferson Lab.

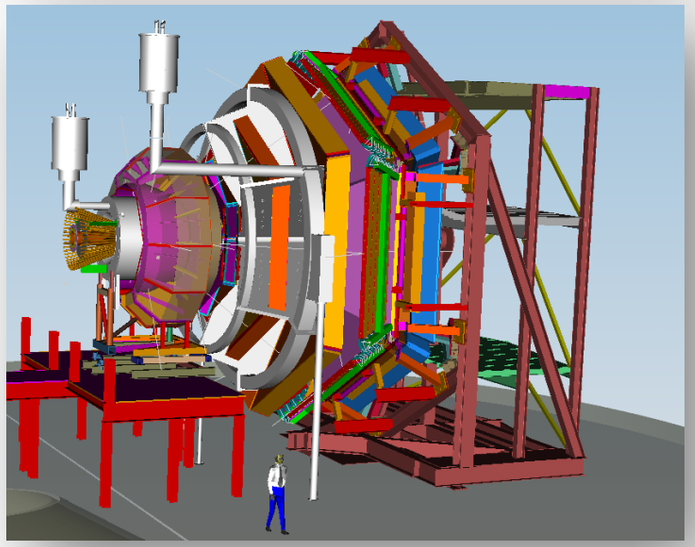
Continuous Electron Beam Accelerator Facility (CEBAF)



Hall A –form factors, future new experiments (e.g., SoLID and MOLLER)

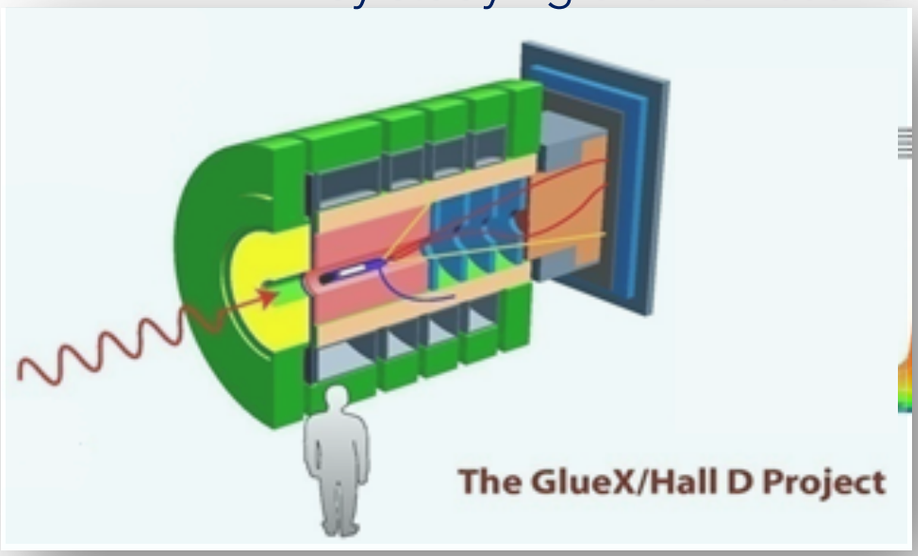


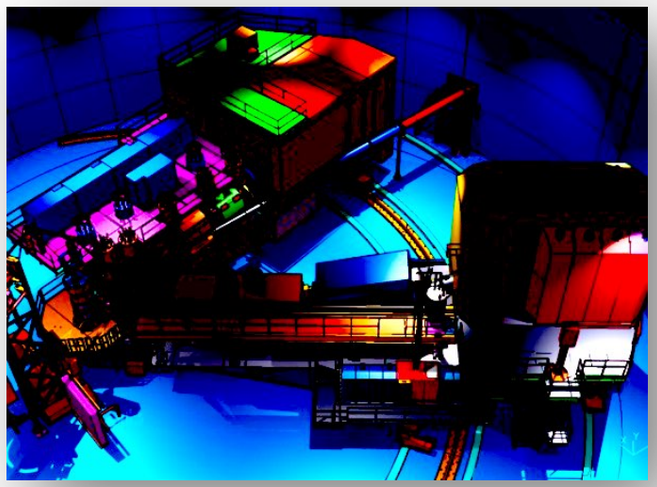
- Built 25 feet below the Earth's surface on the "Yorktown Formation"
- More than 2,200 magnets, from a few inches to three yards in size
- The electron beam travels around the 7/8-mile racetrack- shaped accelerator five times in about 22 millionths of a second at nearly the speed of light



Hall B – understanding nucleon structure via generalized parton distributions

Hall D – exploring origin of confinement by studying exotic mesons

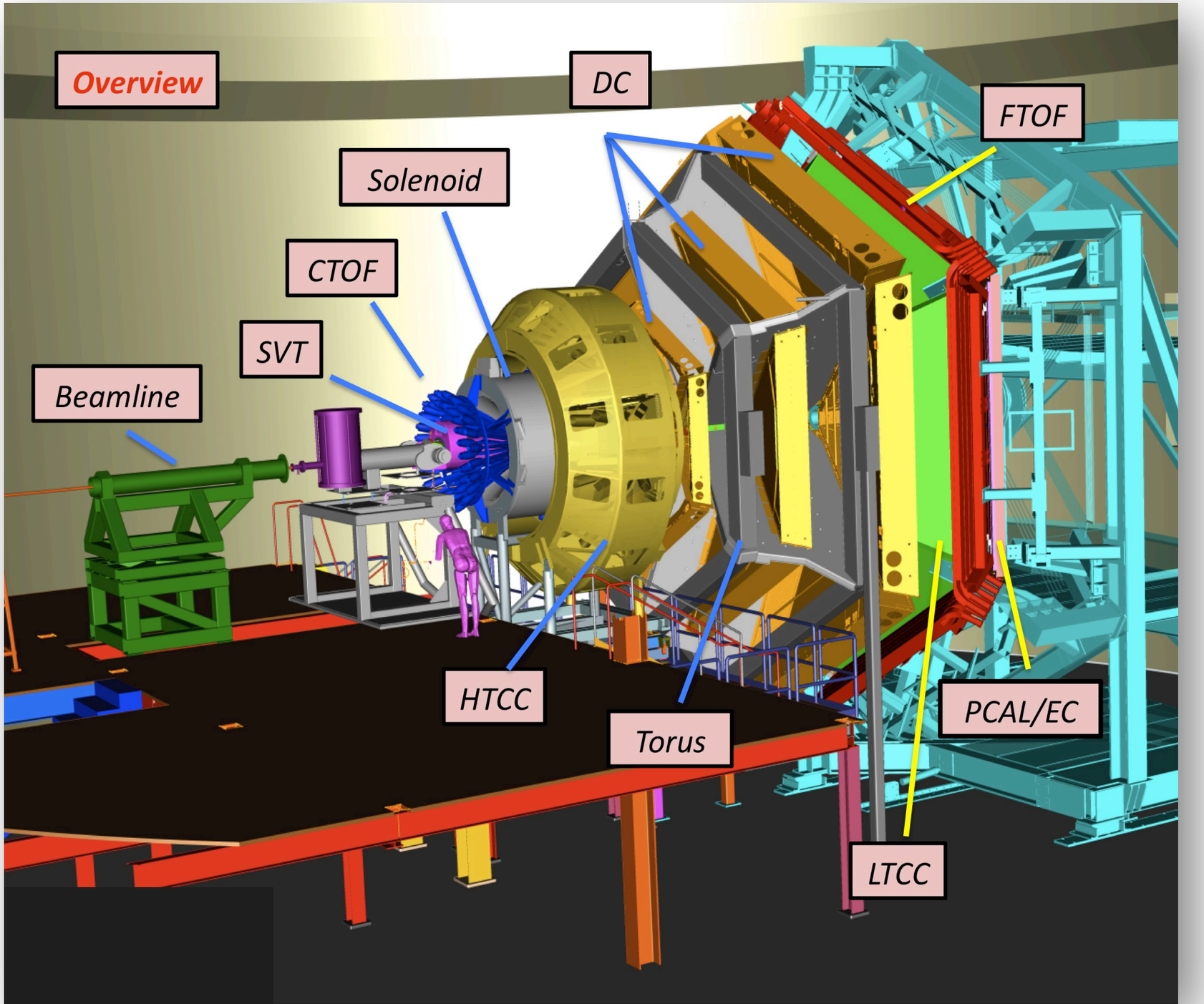


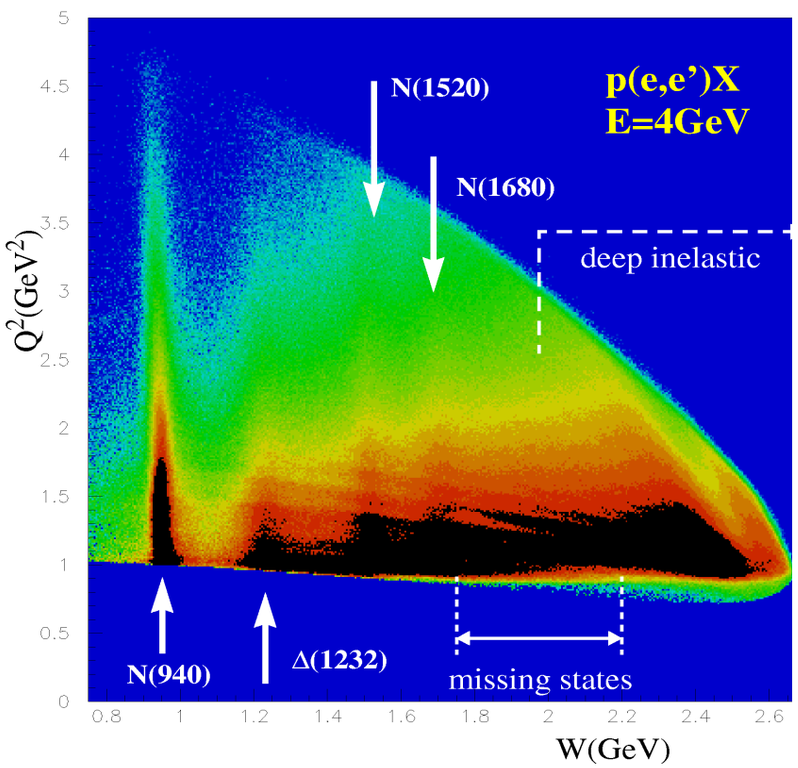


Hall C – precision determination of valence quark properties in nucleons and nuclei

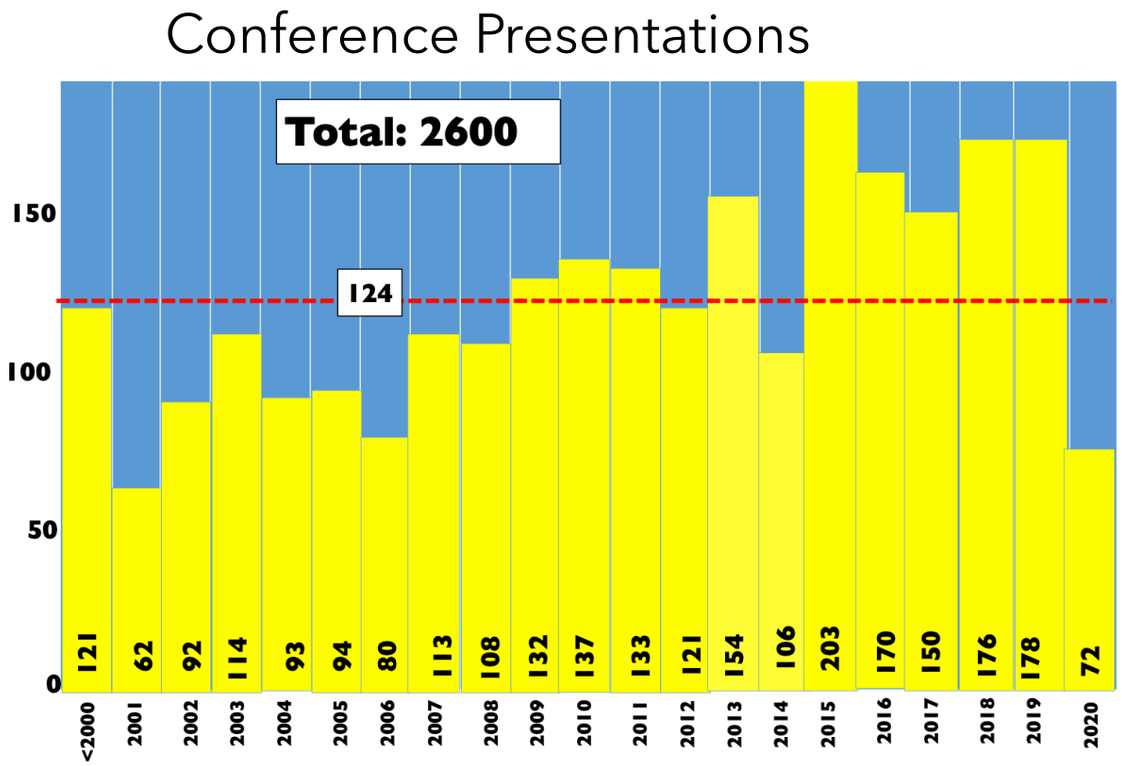
The CLAS12 Detector: Large Acceptance Spectrometer

- Several subsystem, tens of thousands of channels
- The Flash Analog to Digital Converter (FADC250) samples thousands of analog signals 250 million times per second. Each of the signals can be really fast: a few billionths of a second.
- One Jefferson Lab FADC channel can sample the detector data 250 million times / second, a bandwidth equivalent to 25,000 instruments.
- The electric signals in all the Hall-B detectors are equivalent to an orchestra of 100 million people.

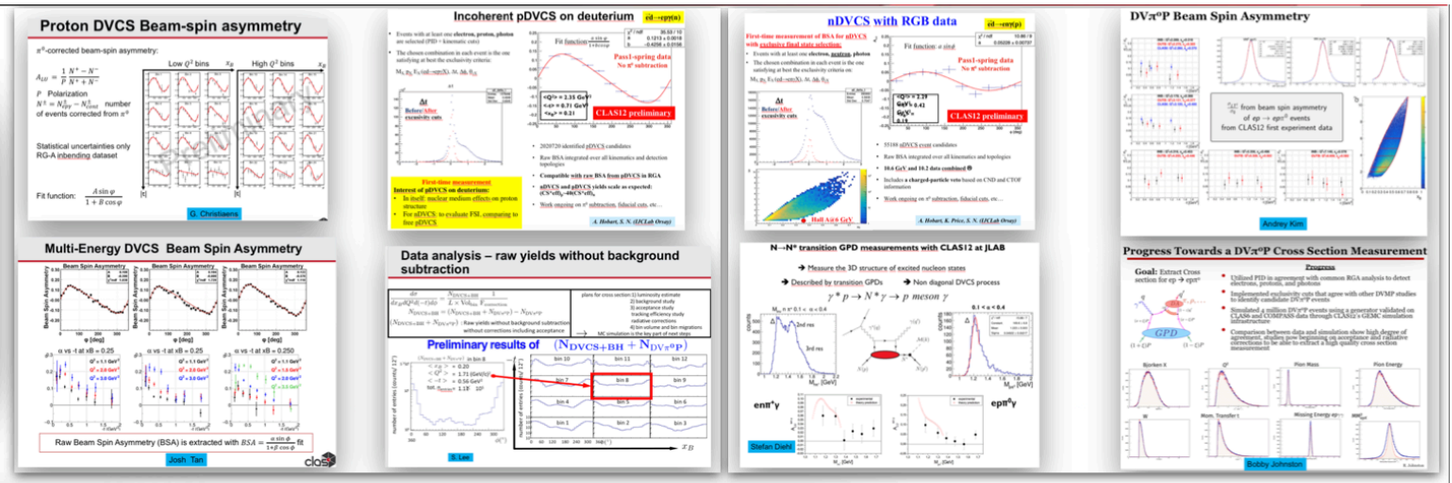




CLAS/CLAS12 Portfolio



- 230 physics papers in peer review journals (> 14,000 citations)
- 5 papers in Nature, 1 paper in Science
- 2600 conference talks (~1650 invited)



First CLAS12 publications: last month!

Mini-Symposium: Electromagnetic Form Factors of N*'s, Sessions DQ, EQ, and FQ, October 30, 2020

Goals: Facilitate joint efforts between experiment, phenomenology and theory on exploration of the spectrum and structure of the ground and excited states of the nucleons from the CLAS and CLAS12 data in order to get insight into strong interaction dynamics which underlie the baryon generation from quarks and gluons.

Organizers: K. Hicks, Ohio U., V.I. Mokeev, Jefferson Lab

Invited review talks:
1. Studies of Excited Nucleon Structure with CLAS and CLAS12
   Prof. K. Joo, University of Connecticut, USA
2. Ground and excited nucleon structure within continuum QCD approaches
   Prof. J. Segovia, Pablo de Olavide University, Seville, Spain

Novel direction:
1. Exploring the Emergence of Deformation Dominance in Nuclear Structure from Strong QCD
   Prof. J.P. Draayer, Louisiana State University, USA

and 18 contributed talks

The CLAS/CLAS12 experiments were designed to perform complementary measurements with different beam energies different targets and different combination of polarizations to study:

- protons and neutrons structure for both the ground and excited states, 3D imaging and mechanical structure of the nucleon with the core mission to understand the manner in which the constituents of protons are held together by the strong force and the emergence of the dominant part of hadron mass.

- quark confinement and the role of the glue in meson and baryon spectroscopy

- strong interaction in nuclei – evolution of quark hadronization, nuclear transparency of hadrons

… and many more:

- SIDID single pi+ BSA (S.Diehl)
- Di-hadron SIDIS (T.Hayward)
- SIDIS pion multiplicity (G.Angelini)
- BAND physics program (C.Fogler)
- BSA in resonance region (V.Klimenko, E.Isupov)
- Resonance electrocoupling (K.Neupane)
- Incluse cross section (N.Markov)
- RG-F (BONUS) report E.Christy)

Credit:L.Elouadrhiri

← Last APS/DNP CLAS12 Presentations

CLAS12 Simulations

facts sheet

- 13.5 B triggers of experimental data so far (2 years)
- We need x1-x10 simulated data
- Time to generate / simulate / reconstruct / skim 1 event ~ 2s
- 887 years on a single core for 1x nevents as data (just for this 2 years)

- Based on analyses workflow: CPU Time needed: 60 M CPU Hours / Year (this is a conservative estimate)

- Jefferson Lab Computing Farm is huge, but our CLAS12 slots are busy doing data reconstruction. We need offsite computing.

CLAS12 Simulations: the traditional way to run

Individual Experiments / Users have individual responsibility, for each analyses, to collect, modify, install for farm usage (including maintaining OS platforms), individual packages such as:

- generators (modified versions of PEPSI, Pythia, analysis specifics)
- CLHEP
- Geant4
- GEMC (Simulation Software)
- Calibration Databases
- Field Maps
- Background Data
- Reconstruction Software
- DST makers

Each package is collected with its own repository, version.

The collection is often undocumented: reproducibility/debugging nightmare

The farm flow submission/usage is done using custom scripts. Maintainability and debugging nightmare

The OSG Revolution: Part 1a

the docker/singularity happy river flow

- Containers are “natural” on the OSG.
- Tagged Container Images solve the software collection / simulation nightmares.
- Standard submissions scripts for all users mitigate submissions errors to basically zero.

CLAS12 Container Tag 4.4.1 (“production”)

> CCDB version: 1.07.00

> CLHEP version: 2.4.1.3

> GEANT4 version: 4.10.06.p02

> XERCESC version: 3.2.3

> EVIO version: 5.1

> MLIBRARY version: 1.4

> SCONS version: 1.9

> CLAS12Tag: version **4.4.1**

> CLAS12 Generators Collection: version 1.2

For all practical purposes ALL we need to do is modify a single dockerfile and everything is AUTOMAGICALLY ported onto a singularity image available to ALL OSG Nodes.

Incredible!

The OSG Revolution: Part 1b

tagged software and data through CVMFS

> Reconstruction Software Based on Java:

- **6.5.9 (rgb)**
- **6.5.6.1 (rga and rgk)**

> Calibration Software **Databases** tagged file 4.4.1

> Magnetic **Fields Maps**

> ROOT

> clas12-mcgen

- CVMFS is “natural” on the OSG
- We get all of this for free on all OSG nodes

To use:

modules environment files on /cvmfs/oasis.opensciencegrid.org/jlab/hallb

```
module load coatjava/6.5.9
module load jdk/1.8.0_31
module load root
module load mcgen
```

The OSG Revolution: Part 1c

XROOTD: Background Merging Files

Server: xroot://sci-xrootd.jlab.org//osgpool/hallb/

Files are organized under: clas12/backgroundfiles/

```
rga_fall2018
rga_spring2018
          tor-1.00_sol-1.00
                45nA_10604MeV/
                      10k/
                50nA_10604MeV/
                55nA_10604MeV/
          tor+1.00_sol-1.00
          tor+1.01_sol-1.00
rga_spring2019
rgb_spring2019
rgk_fall2018_FTOff
rgk_fall2018_FTOn
```

- tree path selectable by drop down menu on portal
- drop down many generated based on directory structure
- Currently we copy random file (among 100 each configuration, about 100MB) to the local node
- Plans to move to random-access direct reading of one file / configuration
- Random Access XROOTD reading would also allow us to run data reconstruction on the OSG

The OSG Revolution: Part 2a

and voilá: hundreds of cores for free

The process of adding an offsite resource:

- is straightforward
- done mostly by OSG Staff
- minimal requirements:
  - CVMFS, Singularity, Network Access

Easier Options

1. Make it “dedicated” to CLAS12
2. Make it “high priority”, but not limited to, CLAS12
3. Make it available to OSG in general

Steps:

- contact colleague for offsite resource
- advantage to both CLAS12 and collaborating institutions
- colleague and farm administrators in contact with OSG staff
- voilá: in a few days (weeks if bureaucracy is slow) we have resources
- magical, no overhead for us

Many thanks to Edgar Fajardo Hernandez for his work and patience

The OSG Revolution: Part 2

Technical Support

- Support and feedback always immediate
- Patience in explaining details
- Sincere support, OSG staff want us to “win”
- Volunteered solutions: Edgar suggested and made it happen SU university high priority usage for CLAS12
- OSG Staff support on local JLAB nodes (scosg16 and now scosg20)

The CLAS12 OSG Portal

CLAS12 Portal

Summary of current jobs

user	submission	total	done	run	idle
tyson	2	1002	20	980	2
mdefurne	1	10000	9715	285	0
jnewton	1	100	93	7	0
carman	4	20000	6848	7032	6120
sdiehl	2	240	230	10	0
total	10	31342	16906	8314	6122

Click to submit to OSG

Generator	LUND Files
- In-Container or gemc internal generator - Arbitrary number of jobs - Arbitrary number of events for each job (max 10,000)	- LUND files (.txt) from a web location or directory in /volatile - One job per LUND file

Details of current OSG Jobs

user	job id	submitted	total	done	run	idle	hold	osg id
tyson	1951	11/04 09:18	501	2	497	2	0	2778241
tyson	1952	11/04 09:44	501	18	483	0	0	2778242
mdefurne	1943	11/03 08:55	10000	9715	285	0	0	2778126
jnewton	1944	11/03 14:23	100	93	7	0	0	2778131
carman	1945	11/03 14:32	5000	4315	685	0	0	2778132
carman	1946	11/03 14:32	5000	2133	2862	5	0	2778133
carman	1947	11/03 14:33	5000	398	3485	1117	0	2778134
carman	1948	11/03 14:34	5000	2	0	4998	0	2778135
sdiehl	1949	11/03 15:51	120	119	1	0	0	2778136
sdiehl	1950	11/03 15:52	120	111	9	0	0	2778137

The CLAS12 OSG Portal

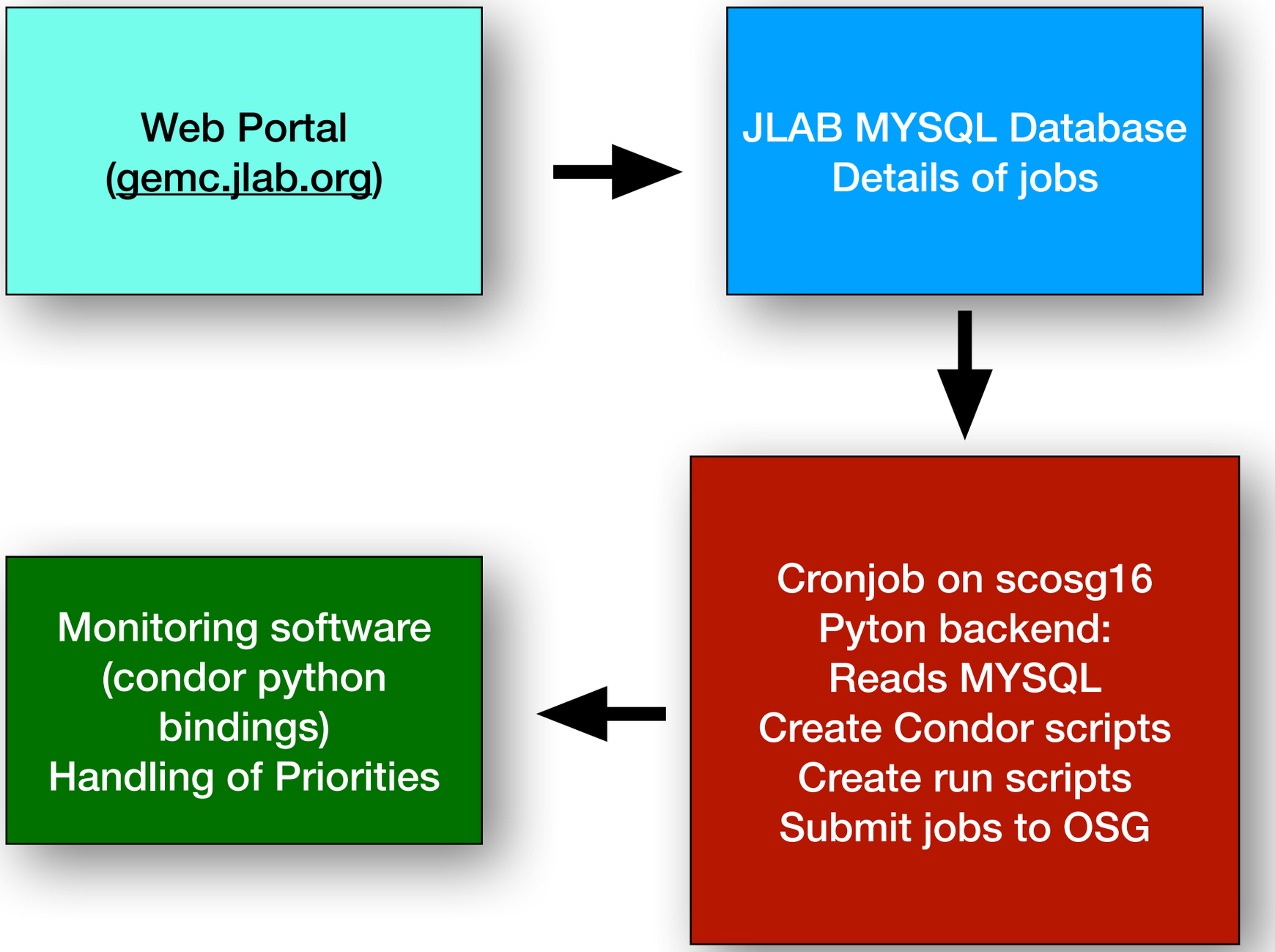
CLAS12 Portal

Configuration	rgb_spring2019	
Magnetic Fields	tor-1.00_sol-1.00	
Generator	clasdis	
Generator Options	--t 20 25	**clasdis options**
After selecting the generator, check the documentation and paste the needed options above. Notice: do no use the following options as they are automatically passed for you: --docker, output file name --trig options.		
Number of Events / Job	10000	
Number of Jobs	2000	
Total Number of Events	20 M	
Background Merging	50nA_10200MeV	
Output Options	☑ dst Warning: any of the choices below will enlarge the overall output size significantly. ☐ generator ☐ gemc ☐ gemc decoded ☐ reconstruction	
	Submit	

Submit complex CLAS12 simulations from anywhere (from phone!)

The CLAS12 OSG Portal

backend software



Offsite Dedicated Resources

OSG Name	Country/ Institution	Number of Nodes	MHours / Year
INFN-T1	Istituto Nazionale di Fisica Nucleare, Italy	500	4.4
UKI-SCOTGRID-GLASGOW	Glasgow University, Scotland	700	6.1
GRIF	Grille au service de la Recherche en Ile-de-France, France	500	4.4
MIT	Massachusetts Institute of Technology, USA	500	4.4
Upcoming: LAMAR	Lamar University, USA	1-200	1.7

Offsite Dedicated Resources

OSG Name	Country/ Institution	Number of Nodes	MHours / Year
INFN-T1	Istituto Nazionale di Fisica Nucleare, Italy	500	4.4
UKI-SCOTGRID-GLASGOW	Glasgow University, Scotland	700	6.1
GRIF	Grille au service de la Recherche en Ile-de-France, France	500	4.4
MIT	Massachusetts Institute of Technology, USA	500	4.4
Upcoming: LAMAR	Lamar University, USA	1-200	1.7

We need to make sure that the dedicate resources are fully utilized: their availability and expansion are problematic if they are “idle” most of the times

Current Status:

Testing Flags passed to OSG to prioritize offsite farms

Offsite High Priority and Opportunistic Resources

OSG Name	Country/Institution	Number of Nodes	MHours / Year
UCONN	University of Connecticut, USA	0-2,000	8.7
SU	Syracuse University	0-6,000	20
JLAB	Jefferson Lab, USA	0-1000	4.4
OSG	OSG Opportunistic	0-3000	10

Not dedicated but CLAS12 has high priority on these

CLAS12 Project Summary of Resources

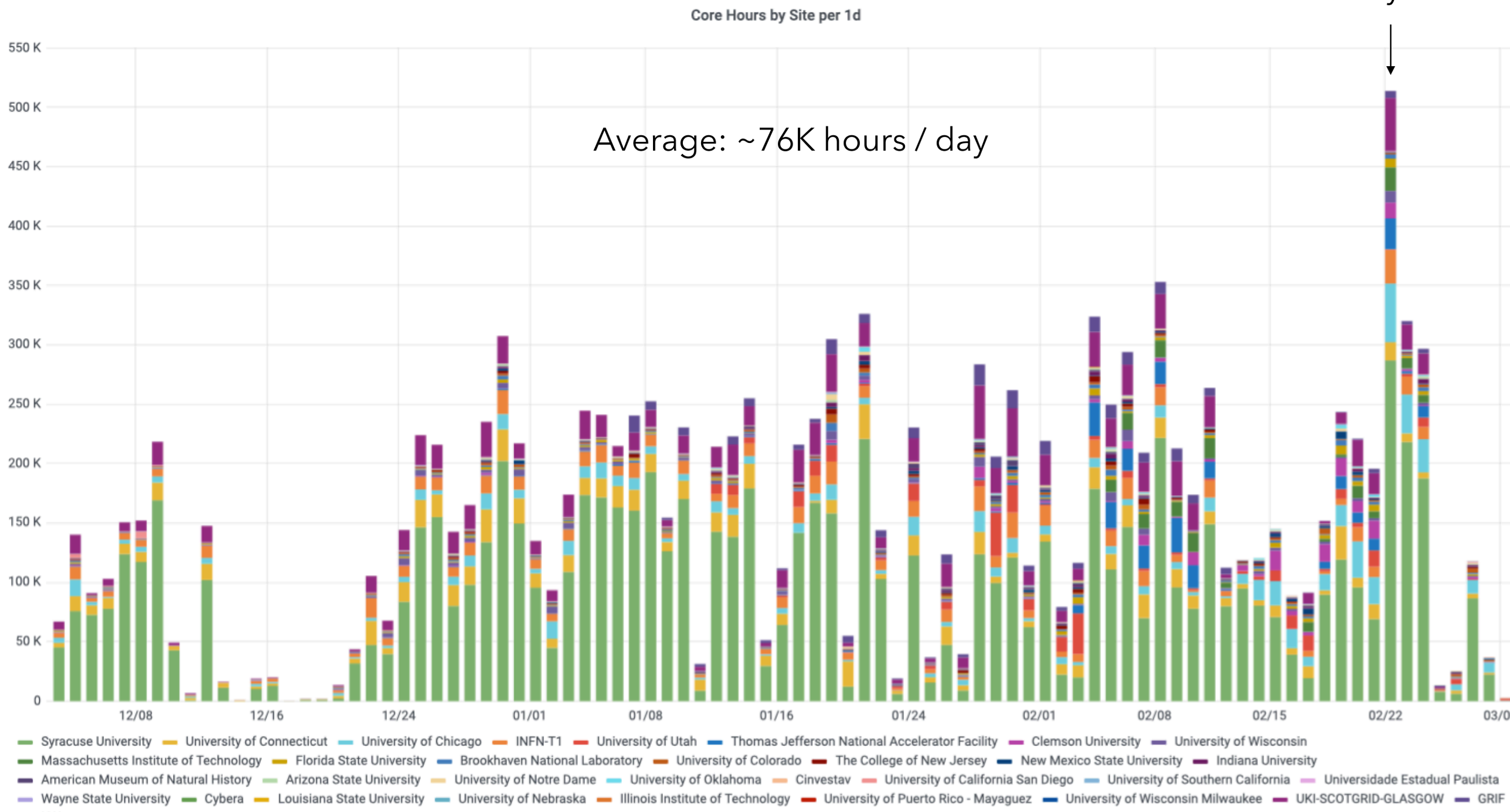
OSG Name	Number of Nodes	MHours / Year
Dedicated	2200	19
Priority	0-12,000	52
Opportunistic	0-3000	13
Total	2200-17200	84

CLAS12 Project Goal for Simulations:

60 M Hours

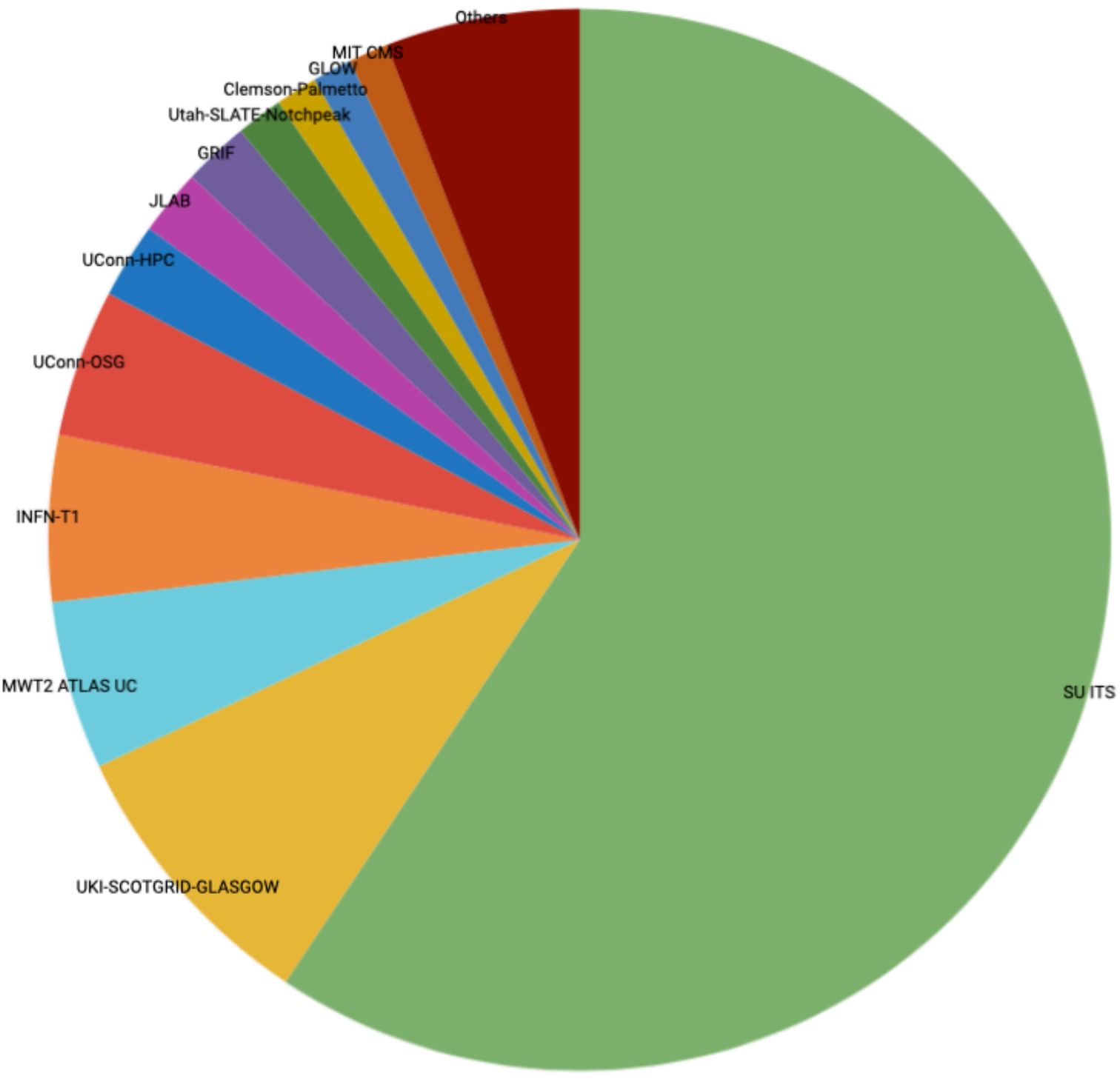
Last 90 days usage (March 2021)

best day: 520K



Average: ~76K hours / day

Last 90 days usage (March 2021)



2021-02-22 00:00:00

Syracuse University:	287 K
University of Connecticut:	15.3 K
University of Chicago:	49.4 K
INFN-T1:	28.3 K
University of Utah:	635
Thomas Jefferson National Accelerator Facility:	25.9 K
Clemson University:	13.3 K
University of Wisconsin:	9.85 K
Massachusetts Institute of Technology:	19.8 K
Florida State University:	7.33 K
Brookhaven National Laboratory:	3.59 K
University of Colorado:	1.17 K
The College of New Jersey:	0
New Mexico State University:	458
Indiana University:	664
American Museum of Natural History:	0
Arizona State University:	0
University of Notre Dame:	162
University of Oklahoma:	0
Cinvestav:	88.9
University of California San Diego:	0
University of Southern California:	35.8
Universidade Estadual Paulista:	197
Wayne State University:	0
Cybera:	76.2
Louisiana State University:	0
University of Nebraska:	0
Illinois Institute of Technology:	11.4
University of Puerto Rico - Mayaguez:	0
University of Wisconsin Milwaukee:	0
UKI-SCOTGRID-GLASGOW:	44.6 K
GRIF:	5.89 K

Summary

- The OSG has enabled CLAS12 Scientist and collaborators to run millions of hours of simulations in a straightforward, reproducible way
- Docker/Singularity/CVMFS usage is natural on the OSG
- Adding offsite resources is ridiculously easy
- Support is really great
- We could use more tools to monitor / diagnose problems when something goes wrong.

Thank you so much for your work, your passion and dedication