2019 Joint HSF/OSG/WLCG Workshop - HOW19
18-22 March 2019

Computing activities for the JLab 12 GeV science program

R. De Vita
INFN – Genova

Outline

- Jefferson Lab landscape and mission
- Scientific computing at Jlab
- Software and computing for the 12 GeV program
  - Experiments requirements and approaches
  - Use of onsite and offsite resources
  - Future challenges
- Supporting activities
  - Advanced computing initiatives
  - Grand Challenge
- Summary

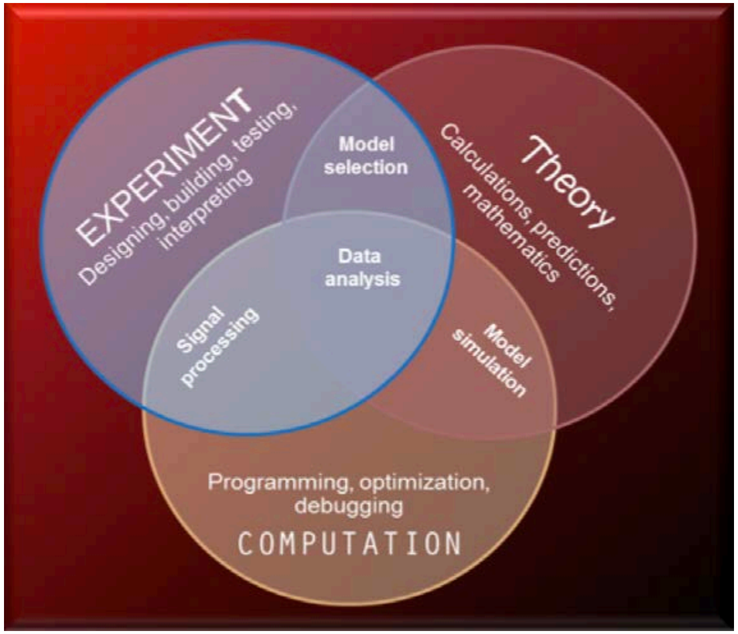
Jefferson Lab landscape

- Multi-hall nuclear physics user facility hosting the Continuous Beam Electron Accelerator Facility (CEBAF)
  - 6 GeV era: 173 Experiments completed over 17 years
  - 1500 users
  - >500 PhDs awarded
  - Tight coupling between Theory and Experiment
  - LQCD major driver for the computing program
- CEBAF upgrade to 12 GeV
  - Upgrade of the existing experimental Halls and construction of the new Hall D, with two new major experiments
  - Simultaneous running of the four Halls since 2017
  - Over 10 year scientific program is already planned with 32 weeks of operation/year
  - Increased demand on computing resources for the realization of the 12 GeV science program
- In the future:
  - New experimental facilities (Moller, Solid)

Jefferson Lab Agenda

MISSION		We support the DOE Office of Science and serve the Nuclear Physics User Community as a world-leading center for fundamental nuclear science and associated technologies		
SCIENCE & TECHNOLOGY	STRATEGIC OUTCOMES	Enable scientific discoveries by the Nuclear Physics User Community through our unique, world leading facilities and capabilities	Plan for future facilities and capabilities to realize the long-term scientific goals in Nuclear Physics research	Provide technology solutions that support the NP community, the larger DOE mission and societal needs
	MAJOR INITIATIVES	1 Operate CEBAF accelerator and experimental facilities to execute the FY19 experimental nuclear physics program 2 Prepare CEBAF accelerator and experimental equipment for future 3-5 year experimental physics program 3 Perform R&D to enable enhanced performance and future new capabilities for CEBAF and experimental halls 4 Perform theoretical research in support of the CEBAF 12 GeV program 5 Perform theoretical and experimental research in support of the broader NP research community 6 Provide software and computational resources for theoretical and experimental nuclear physics research	1 Continue to develop the MOLLER and SoLID initiatives 2 Perform Accelerator R&D towards an Electron Ion Collider 3 Perform Detector R&D towards an Electron Ion Collider 4 Pre-project design and planning for an Electron Ion Collider 5 Engage with the EIC user community and further develop the anticipated scientific program for a future Electron Ion Collider 6 Develop and expand expertise in Scientific Computation and Data Science	1 Plan and execute projects to construct equipment for facilities (DOE and possibly others) beyond Jefferson Lab. 2 Perform R&D to enable other future (non-CEBAF, non-EIC) accelerator capabilities and enhance the reputation of JLab in SRF and large-scale cryogenics 3 Develop and promote technology transfer activities that align with the lab's research portfolio, further DOE missions, and promote national and economic security.
OPERATIONS	STRATEGIC OUTCOMES	Provide, protect, and improve the human, physical and information resources that enable world class science		
	MAJOR INITIATIVES	1 Business Process Streamlining 2 Cyber Operations Laboratory 3 Facilities Engineering and Reliability Enhancement	4 Safety and Quality Processes 5 Internal and External Communications 6 Talent Management	

Scientific computing at JLab



Simultaneous support of theory and experiments:

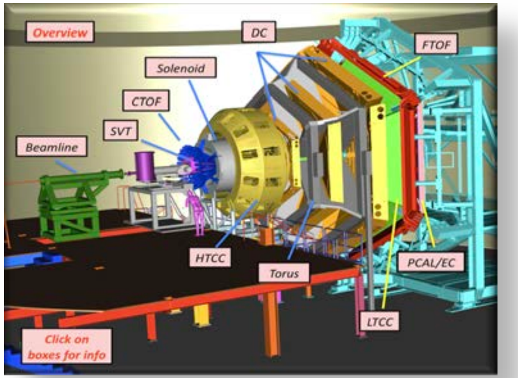
- Integration of experimental and theoretical analysis
- Crucial for the success of the 12 GeV program

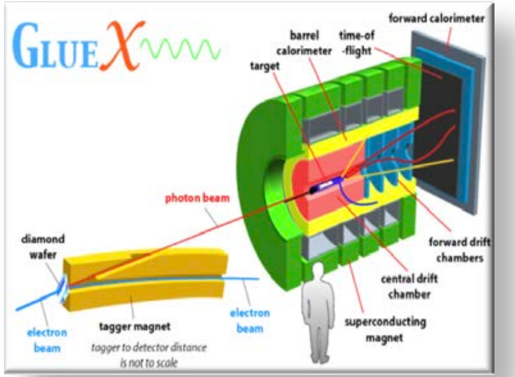
- Physics computation –Large Scale Parallel Computing
  - Lattice QCD: JLAB known for Science; software; hardware
  - Accelerator simulation
- Experimental computing is coordinated by Physics Division
  - Physics Division Staff often play lead role
  - Relatively small efforts: in the few FTEs
- Scientific Computing group in IT supports
  - Hardware: Disk; Tape; Compute clusters;
  - Runs key services: the batch system
  - Provides in-house tools: e.g. monitoring and workflow
  - Provides technical support and expertise, especially for parallel computing
- IT provides cyber security, networking and desktops

Experimental Halls

- 12 GeV program currently underway in all four Halls
- Diverse needs and challenges depending on rates and event sizes
- Computing models developed and improved based on experiment needs and experience from ongoing program

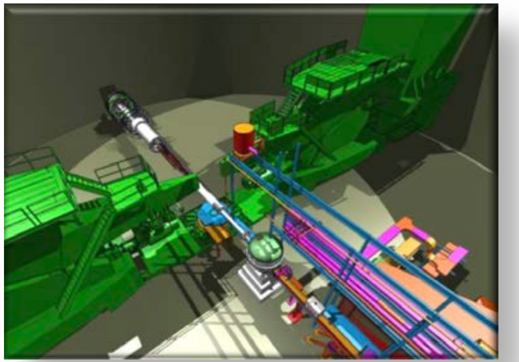
Hall B – nucleon structure via generalized parton distributions and transverse momentum distributions

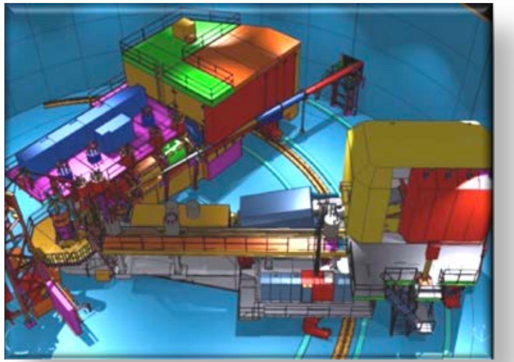




Hall D - exploring origin of confinement by studying exotic mesons

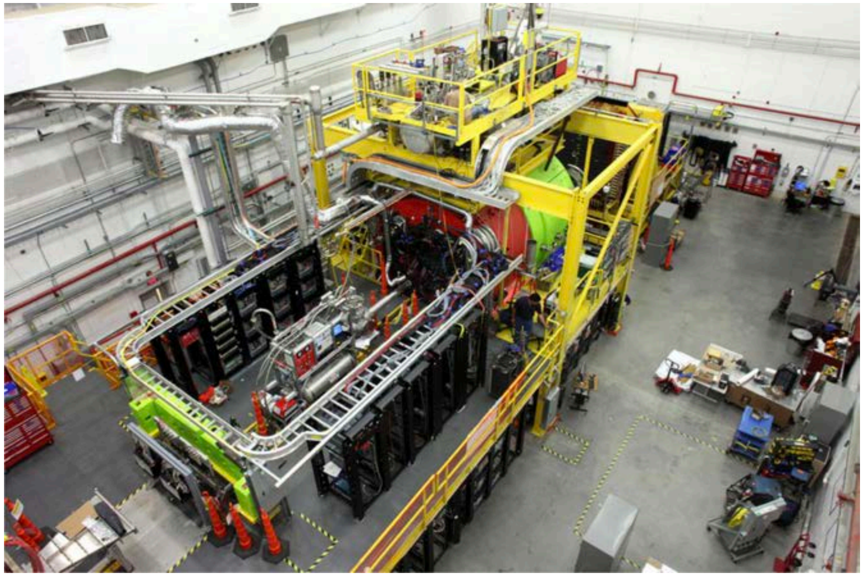
Hall A – short range correlations, form factors, hyper-nuclear physics, future new experiments (e.g., SoLID and MOLLER)





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

Hall D/GlueX

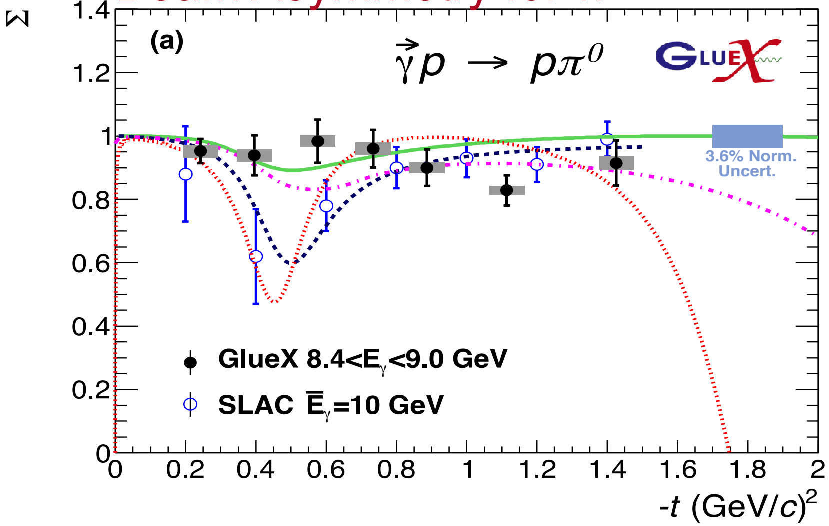


Exploring the origin of quark-gluon confinement by studying meson photo-production and searching for exotics

- Large acceptance, hermetic multi-detector spectrometer
- Reconstruct exclusive photoproduction final states
- Perform Partial-Wave-Analysis to extract individual meson resonances

- Commissioning started in Fall 2014
- Physics started in Spring 2017, GlueX-I (low luminosity) has completed data taking
- GlueX-II (high-luminosity) starts in Fall 2019 and at least 5 years of running

Beam Asymmetry for π^0



Phys. Rev. C95, 042201 (2017)

Hall D/GlueX

JANA:

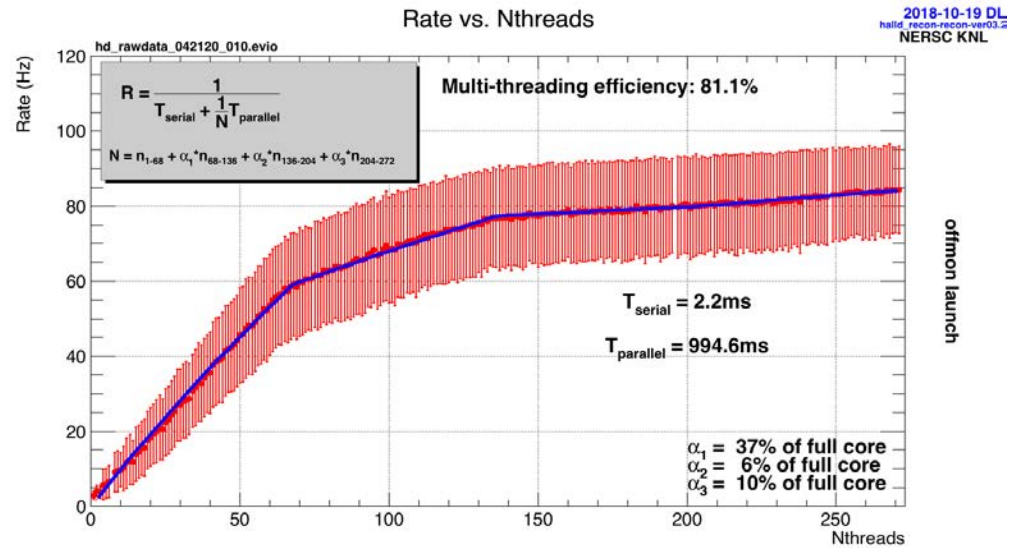
- Multithreaded, factory based, plugin driven C++ framework for reconstruction and analysis

AmpTools:

- C++ libraries for Partial Wave Analysis (PWA), i.e. unbinned maximum likelihood fits to data using user-provided sets of interfering amplitudes
- Multi-core, multi-machine support, GPU-enabled

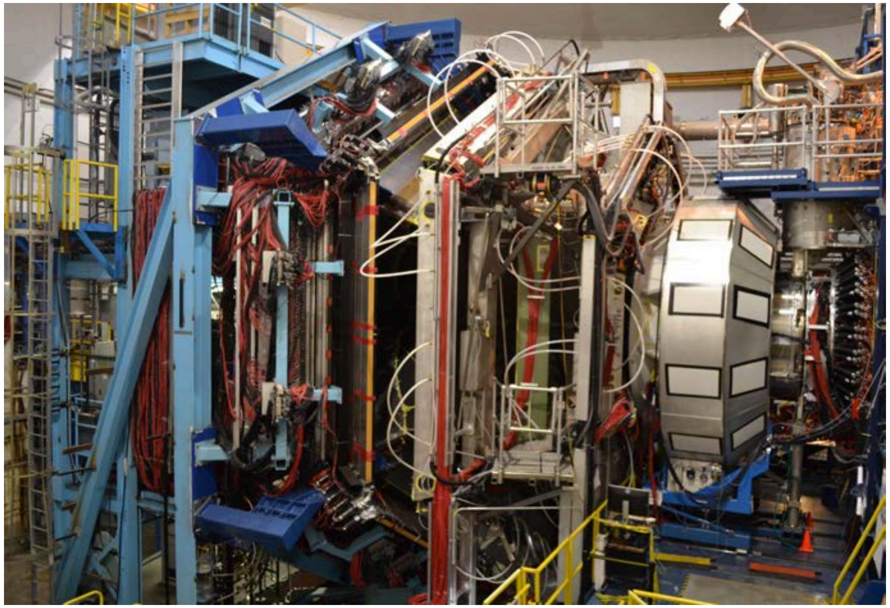
Data format:

- EVIO and REST data formats for raw and reconstructed (DST) data formats



	Low Intensity	High Intensity
Beam	2.4 x 10^7 γ/s	5 x 10^7 γ/s
Trigger	42 kHz	90 kHz
Front End	0.5 GB/s	1.2 GB/s
Disk	0.5 GB/s	600 MB/s
Tape	4.2 PB/yr	5.8 PB/yr

Hall B/CLAS12

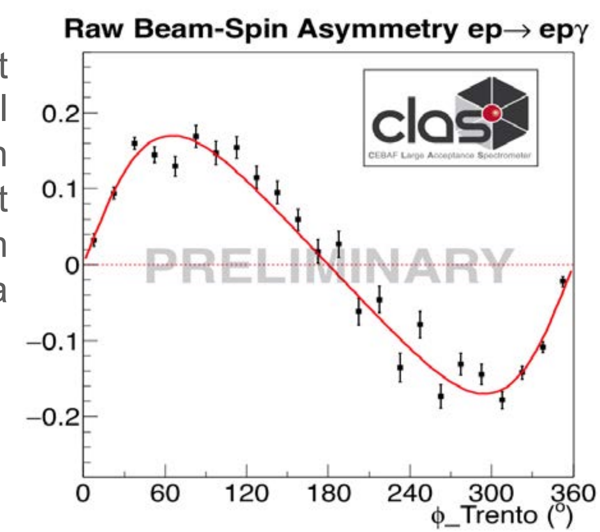


Understanding nucleon and hadron structure via electro-production of inclusive, semi-inclusive and exclusive final states

- Large acceptance spectrometer based on two superconducting magnets
- 16 sub-detectors, >100k readout channels
- Large coverage for charged and neutral particles

- Commissioning started in 2017
- Physics data taking started in Spring 2018:
  - First run on hydrogen target in 2018 (13 parallel physics proposals)
  - Currently running on deuterium

First look at Deeply virtual Compton scattering at CLAS12 from Spring 18 data



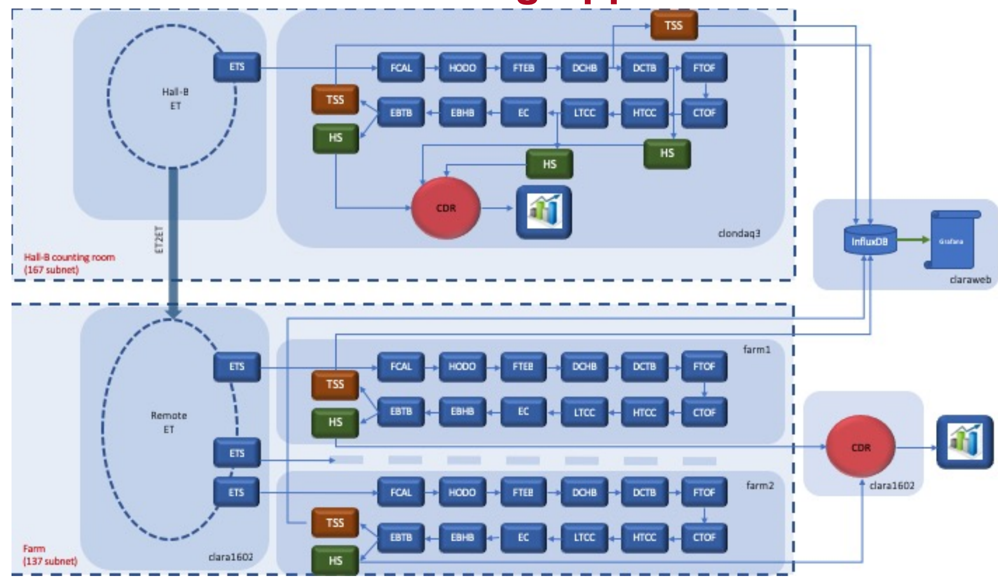
Hall B/CLAS12

Trigger:
- Highly selective, multi component FPGA-based (majority of recorded events are retained for physics analysis)

Offline software:
- Java based toolset (I/O, geometry, calibration, analysis, …) and reconstruction packages
- **CLAS**12 **R**econstruction and **A**nalysis Framework (CLARA) glues together isolated, independent micro-services with reactive resource allocation and multithreading capability
- Geant4 Monte Carlo (GEMC)

Data format:
- EVIO and HIPO data formats for raw and reconstructed (DST) data formats

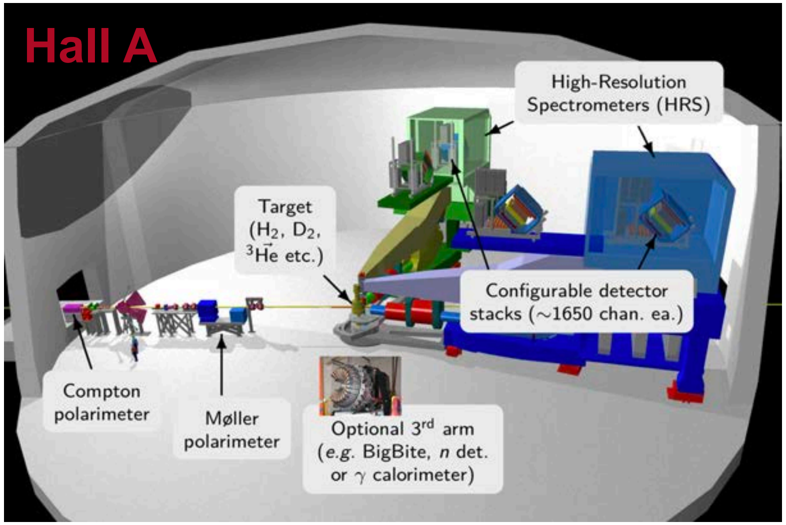
CLAS12 Data Processing Application

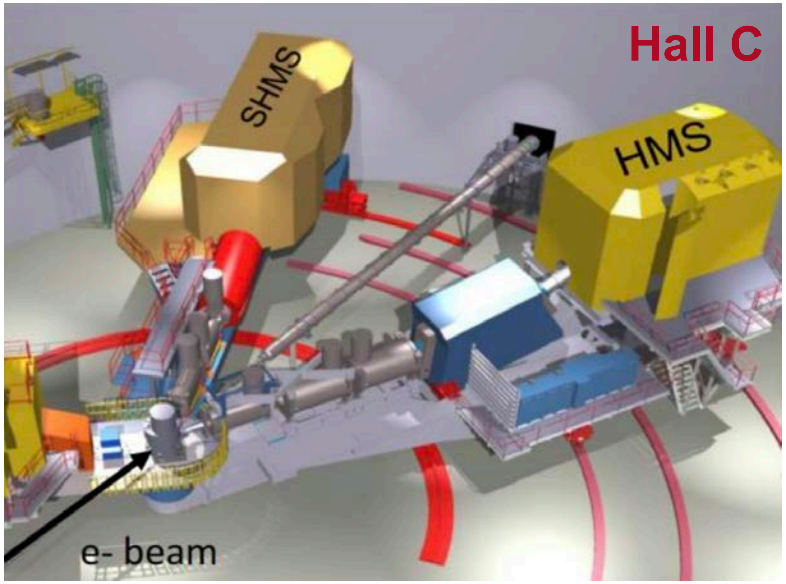


	Run Group A (LH2 target)
Luminosity	0.7 x 10^{35} cm^{-2} s^{-1}
Trigger	13 kHz
Front End	360 MB/s
Tape	1.7 PB/yr

See V. Ziegler's talk on Wednesday

Hall A&C





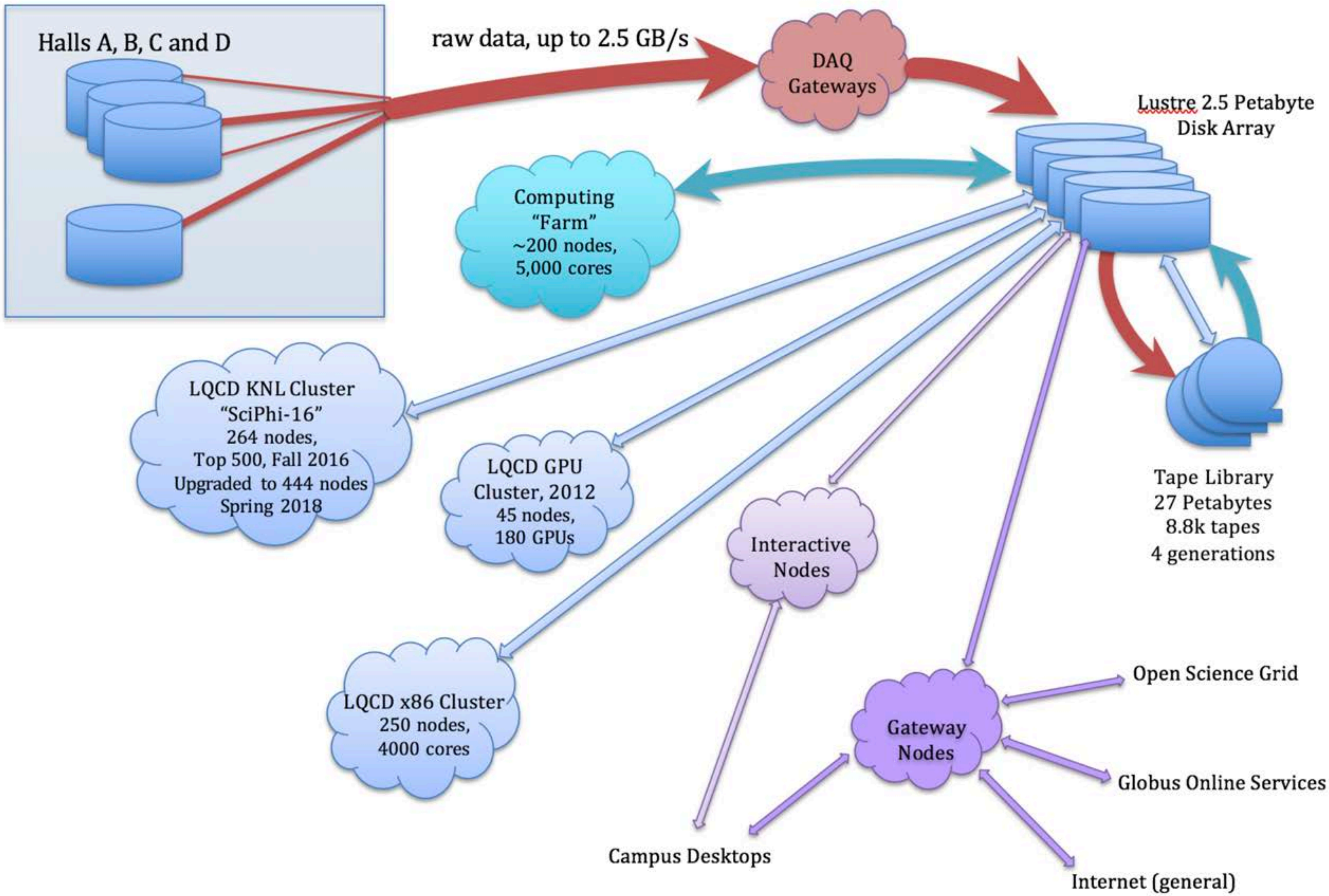
Precision measurements on nucleon structure, form factors, … , and BSM physics:

- High resolution magnetic spectrometers
- Dedicated, experiment-dependent equipment and configuration
- Space for large installation
- Future facilities (Moller, SOLID)

Software and computing:

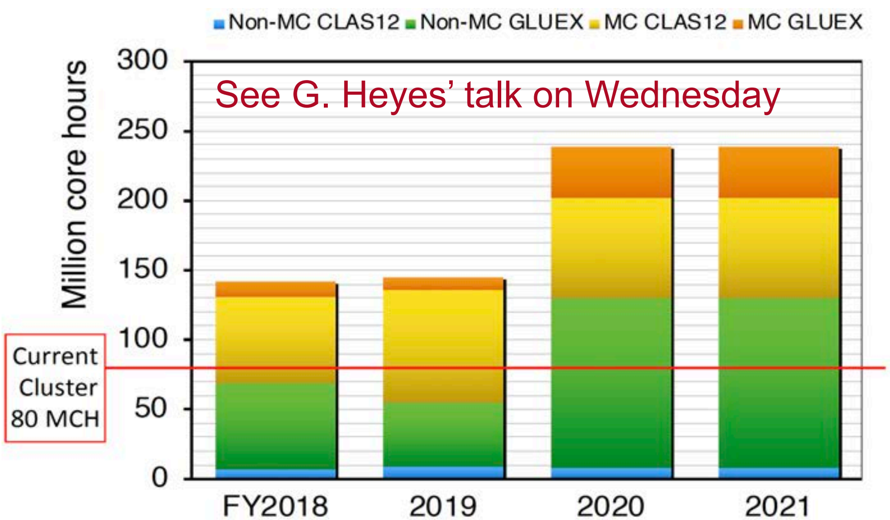
- Relatively small event size and rate, will grow with planned upgrades
- Flexible, plugin-based C++ reconstruction, calibration and analysis framework
  - Highly modular and run-time configurable
  - Large application libraries
  - User-friendly to support large user community and diverse physics goals

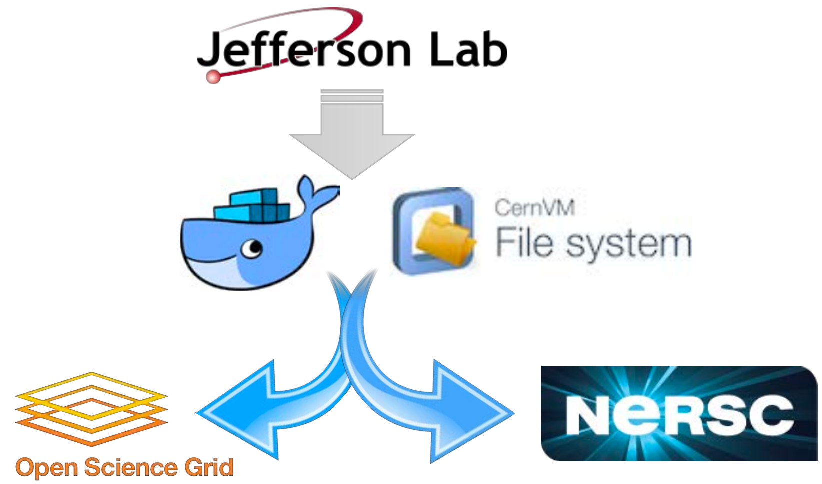
Scientific computing at JLab



Exploitation of onsite & offsite resources

- Onsite computing resources adequate for supporting small scale experiments and large fraction of GlueX and CLAS12 needs
- Offsite resources exploited to satisfy total request:
  - OSG via collaborating institutions
  - NERSC allocation equivalent to 50% of demand in 2019
  - Others…



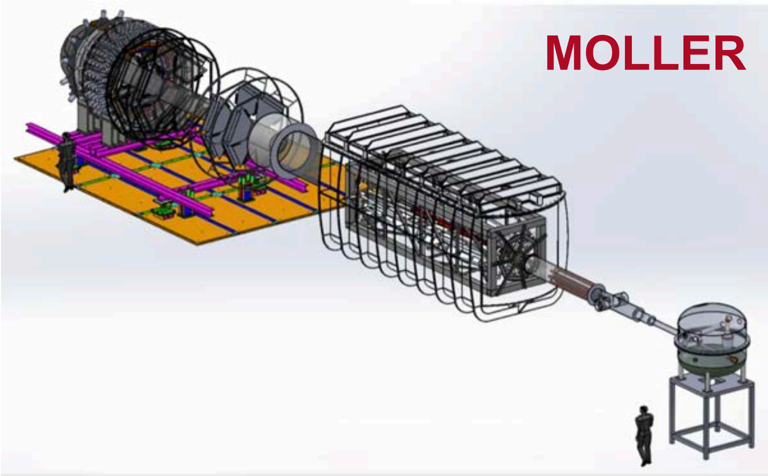


Deployment approach:

- Docker container (converted to Singularity and Shifter)
- CVMFS share
  - Experiment software builds
  - 3rd party software
  - Calibration Constants (CCDB SQLite file)
  - Resource files (field and material maps)
- Full exploitation by GlueX, CLAS12 gearing up

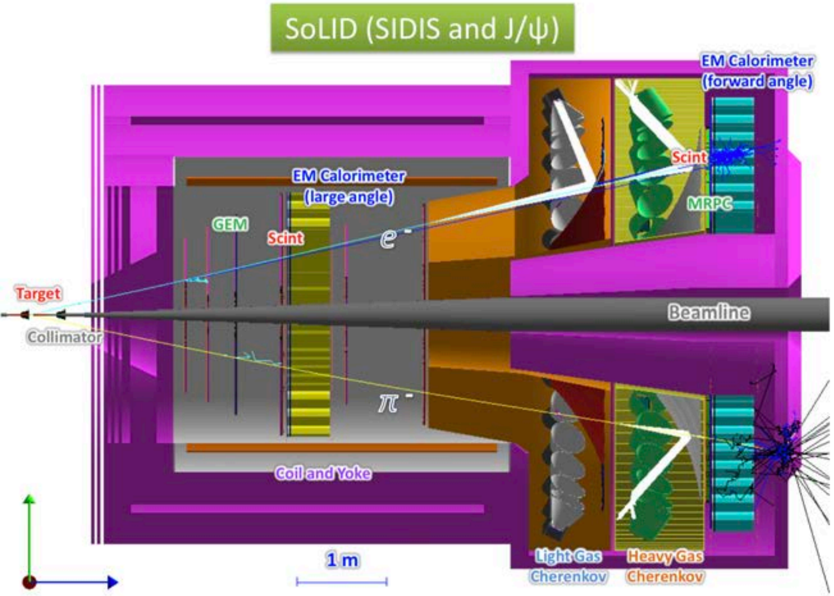
See R. Jones' talk on Wednesday

Future challenges



Standard model test via parity violating Moller scattering at 11 GeV

- CD-0 approved, Dec. 2016
- Need ~3x10^{18} scattered electrons and aims to reach 10^{-9} precision
- ~118 MByte/sec – 425 GB/hour – 4PB total
- Real time analysis for prompt feedback and control of systematics



SoLID – Solenoidal Large Intensity Device

- Multi-configuration 2π forward detector for SIDIS and PVDIS
- CLEO Solenoid, GEM (165K channels), Gas Cherenkov, Shower, MRPC, Scintillator
- 15-100 kHz
- 3-6 GB/s
- 100-200 PB per experiment

… and going beyond the 12 GeV program, EIC see M. Diefenthaler's talk

Advanced computing

Computation is crucial to all aspects of our NP Program

Goal: Develop computing and computation for the success of the 12 GeV Physics Program that transitions toward the EIC era with computational science as a pillar of Femtography

- **Initiative 1**: Integrated Start to End Experimental Computing Model for 12 GeV Physics Program and future EIC
  - Modern computational and data science techniques and hardware technologies provide an opportunity to modernize the experimental computing paradigm
  - Proposed 'Streaming Grand Challenge' organizes 8 ongoing tactical initiatives from DAQ through data analysis towards this integrated model
- **Initiative 2**: Develop computational and data science methodology and infrastructure to realize the scientific goals of Nuclear Femtography.
- **Initiative 3**: Apply Machine Learning for accelerator modeling/control

- Invited talks on streaming at ASCR workshops
- Invited talk on Streaming Data at the DoE booth at Supercomputing 18
- Instigated streaming readout consortium
- Participation in NSAC subcommittee on QIS
- Co-organized Virginia Symposium on Imaging & Visualization in Science
- Host HOW2019

Machine learning

- Growing interest in ML & AI applications to experimental and theoretical physics problems
- Ongoing efforts in:
  - Detector calibration and monitoring
  - Event reconstruction, e.g. tracking and PID
  - Accelerator physics
  - Theoretical analysis of experimental data
  - …
- Lab-wide initiatives to expand expertise and develop synergies between interested groups:
  - Roundtables
  - Seminars
  - …

Streaming Grand-Challenge

Evolve from the current experiment model

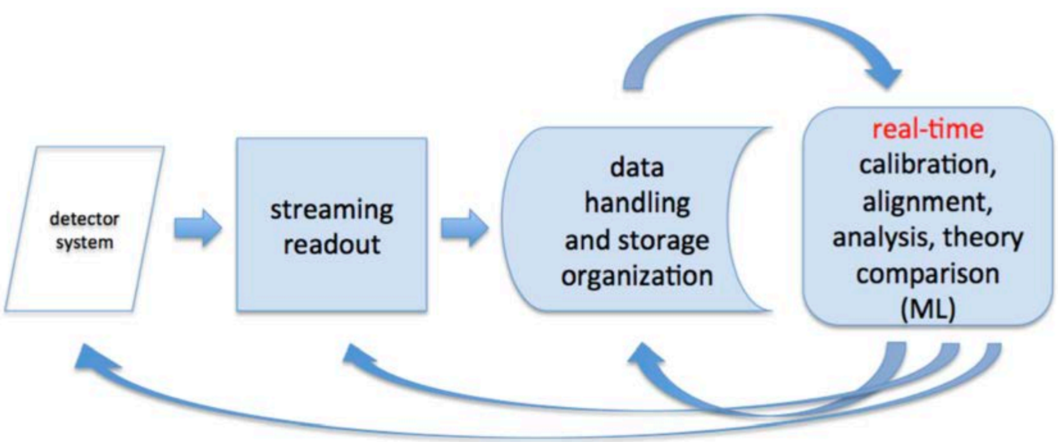
- *triggered DAQ, raw data file, multi-step offline processing, high-level physics analysis and theoretical interpretation*

To an “integrated” model removing the distinction between offline and online computing

- *“Streaming readout” where detectors are read out continuously*
- *Continuous data quality control and calibration via integration of machine learning technologies*
- *Local, task based high performance computing and offsite distributed bulk data processing offsite*
- *Modern, and forward looking, statistical and computer science methods*

“Streaming Grand Challenge” organizes tactical initiatives from DAQ through data analysis towards this integrated model as a proof of concept:

- Workshop and conferences
- Streaming-readout consortium
- Funding of specific projects (LDRD – Laboratory Directed Research Developments)



Summary

- Rich physics program at 12 GeV poses significant software and computing challenges
- Scientific computing model with exploitation of onsite and offsite resources for a cost-effective support of diverse experimental needs
- Bottom-up development approach supported via advanced computing initiatives
- Investment on emerging technologies such as streaming readout, machine learning, … to support the 12 GeV science and build for the future