SoLID Computing

1 Description of Research

1.1 Overview and Context

SoLID (**So**lenoidal **L**arge **I**ntensity **D**evice) is a large-acceptance spectrometer and detector system under development for Jefferson Lab's Hall A, designed to handle very high rates. This system will be the base equipment for a continued high-impact physics program in the 12 GeV era at Jefferson Lab that requires both high luminosity and large acceptance and will facilitate the study of some aspects of the QCD spin structure of the nucleon with unprecedented precision.

1.2 Research Objectives for the Next Decade

Five A-rated experiments have been approved for SoLID to date. Three will investigate spin-dependent Semi-Inclusive Deep Inelastic Scattering (SIDIS) from transversely and longitudinally polarized ^{3}He and transversely polarized ^{2}H, mapping the Collins, Sivers and "pretzelosity" asymmetries and transverse momentum-dependent parton distributions (TMDs) of the nucleon, which will ultimately offer insight into quark orbital motion and spin-orbit correlations. The fourth experiment will study parity-violating deep-inelastic scattering (PVDIS) with proton and deuterium targets to carry out a measurement of $\sin^2\theta_W$ at intermediate Q^2. The fifth will focus on J/ Ψ production near threshold, a measurement sensitive to non-perturbative aspects of QCD in the charmonium sector.

2 Computational and Data Strategies

2.1 Approach

For the next several years, SoLID will perform extensive physics and detector simulations to refine the physics cases, estimate backgrounds, optimize the experimental design and develop the offline reconstruction software. Data taking may start as early as 2023. A high-level software trigger (HLT) is foreseen for SoLID to perform real-time data reduction before writing raw data to tape. The total raw data volume of all approved experiments (after the HLT) is expected to be of order 15 PB. Offline calibration and reconstruction will consist of multiple processing passes over significant parts of these data sets. Physicist users will be provided with DST and ntuple-type ROOT files for final physics analysis.

2.2 Codes and Algorithms

At present, simulations for SoLID are being carried out using GEMC, a Geant4-base simulation package that was originally written for CLAS12. An integrated simulation and reconstruction framework for SoLID is under development. We are currently evaluating Fermilab's C++-based *art* framework for this purpose. Reconstruction is expected to require only standard NP/HEP calorimeter clustering and helical track finding and fitting algorithms, such as Kalman filters, not unlike the ones employed by GlueX. Although *art* is thread-safe, it does not at present support multithreaded or distributed computing. We anticipate these capabilities to become available, or to develop them in-house, by the time SoLID runs.

3 Current and Future HPC Needs

The following table provides preliminary, very rough estimates of the SoLID computing requirements, based on typical algorithm performance and current (2015 vintage) CPUs.

Anticipated SoLID Computing Needs	
Computational Hours	5×10^7 (simulations) 2.5×10^7 (calibrations & reconstruction)
Parallelism	Event-level
Memory	$\leq$ 1 GB/core
Scratch Data and I/O	$\geq$ 250 TB (production output) $\geq$ 100 TB (user files) $\geq$ 300 MB/s disk read/write (large files) $\geq$ 300 MB/s LAN networking $\geq$ 1 TB/hour tape read/write
Long-term and Shared Online Data	10 TB (?)
Archival Data Storage	30 PB (raw + raw duplicates) 500 TB (simulations) 5 PB (production results)
Workflows	Processing is expected to be feasible in-house, assuming the availability of a batch farm with $\geq$ 5000 nodes. DAQ: DAQ $\rightarrow$ online disks $\rightarrow$ HLT $\rightarrow$ stage disk $\rightarrow$ Tape Production: Tape $\rightarrow$ cache disk $\rightarrow$ event distribution $\rightarrow$ worker nodes in data center (reconstruction & user analyzer tasks) $\rightarrow$ event collection $\rightarrow$ production/user files $\rightarrow$ tape Analysis: User files $\rightarrow$ ROOT $\rightarrow$ visualization, physics extraction
Many-Core and/or GPU Readiness	None at present Use of Geant4 multithreading in 1-2 years Parallel/distributed framework in 3-4 years GPU usage not planned