SoLID Software & Simulation Framework

Ole Hansen

Jefferson Lab

SoLID Collaboration Meeting
May 14, 2015

Outcome of Director's Review: Simulations & Software

- **Finding:** "The results from the present, simplified track reconstruction effort do not reveal evident showstoppers, but also do not demonstrate measurement capability."
- **Recommendation:** "End-to-end simulations with realistic subsystem responses and material budgets, and complete track finding and reconstruction should be developed."
- **Recommendation:** "The development of a simulation framework with realistic reconstruction and analysis should be pursued with high priority and increased resources."
- **Recommendation:** "The collaboration is strongly encouraged to develop an end-to-end realistic simulation and reconstruction to further optimize cost and physics reach and derive clear performance requirements for the individual subdetectors."
- **Recommendation:** "Having functional simulation and reconstruction routines as soon as possible should be a high priority in the software effort. Such software will pay off many times over in experimental design and avoiding pitfalls."
- **Recommendation:** "Acceptances, efficiencies, and systematic uncertainties should be simulated for each of the core measurements."

Outcome of Director's Review: Simulations & Software

- **Finding:** "The results from the present, simplified track reconstruction effort do not reveal evident showstoppers, but also do not demonstrate measurement capability."
- **Recommendation:** "End-to-end simulations with realistic subsystem responses and material budgets, and complete track finding and reconstruction should be developed."
- **Recommendation:** "The development of a simulation framework with realistic reconstruction and analysis should be pursued with high priority and increased resources."
- **Recommendation:** "The collaboration is strongly encouraged to develop an end-to-end realistic simulation and reconstruction to further optimize cost and physics reach and derive clear performance requirements for the individual subdetectors."
- **Recommendation:** "Having functional simulation and reconstruction routines as soon as possible should be a high priority in the software effort. Such software will pay off many times over in experimental design and avoiding pitfalls."
- **Recommendation:** "Acceptances, efficiencies, and systematic uncertainties should be simulated for each of the core measurements."

Interpretation

- **Main message:** Simulations are very important at this stage, and we need to give them their due attention. Therefore:
  - Build a simulation & reconstruction framework
  - Make simulations an essential part of the framework
  - Simulate your experiment(s) properly
  - Put more resources into these efforts
- "...realistic subsystem responses and material budgets..."
  - Include all subsystems in the simulation
  - Include support structures and inactive material (don't we already?)
  - Do full digitization
- "...complete track finding and reconstruction..."
  - Reconstruct curve tracks
  - Reconstruct momentum, vertex
  - Extract relevant physics quantities from tracks (e.g. Q^2, x_{Bj})

Comments

- Note that we are not being asked to *demonstrate* track reconstruction "measurement capability".
- Items left unspecific
  - Which costs to optimize?
  - Which detector performance requirements to derive?
  - What constitues demonstrating track measurement capability?
- "Having functional simulation and reconstruction routines as soon as possible should be a high priority . . . Such software will pay off many times over . . . "
  - Rapid development and longevity of software are contradictory priorities
  - My take: Prioritize the development of software now, so that meaningful simulation results will be available early enough still to inform design decisions.

Outcome of Director's Review: Simulations & Software II

- **Observation:** "The manpower currently associated with software for SoLID is estimated to be 6 FTE-years. Numbers from both Hall-B/CLAS-12 and Hall-D/GlueX are in the range of 30 to 50 FTE-years."
- **Finding:** "Consultation with appropriate people from the other halls would be useful to get a more accurate estimate of software needs, including manpower."
- **Recommendation:** "Compare the resource levels you have assumed in some key areas (particularly in software, . . .) to make sure the estimates align with other similar projects or there is a good reason they do not."

- **Finding:** "Early exploration of the tools available at Jefferson Lab that can handle the data at the expected scale of SoLID will be crucial in minimizing false starts in software development."
- **Recommendation:** "Closer communication with the other JLab experiments and the JLab computing center is strongly encouraged."

Interpretation II

- Manpower needs for software development seem severely underestimated
  - But: What constitues a "software" FTE for the other Halls?
  - Hall-D Computing Document states that a total of "47 FTE-years are available to work on software projects and developing physics analyses."
  - Are things like DAQ firmware development, slow controls, system administration, workflow tools development counted? All these are within the scope of the Hall-D computing document.
- Learn (and adopt solutions) from the other halls as much as possible
  - Resource needs
  - "Tools", presumably for batch job and tape library management
  - Otherwise thin on specifics, but this may also include approaches to simulation and reconstruction software etc.

Addressing the Recommendations

- Software group will address the 4 Recommendations regarding developing an "end-to-end simulation and reconstruction framework". A design document will be available by the end of the summer.
- Software group will address the Finding about demonstrating track measurement capability, if necessary. ETA 1 year.
- Simulation group should address the Recommendation regarding simulating acceptances etc. for each of the core measurements. ETA?
- Software group will address the Recommendation to estimate software needs in consultation with other halls. ETA 2-3 months.
  Note: Already without such consultations, the available SoLID software manpower does appear insufficient at this time.
- Both the software and the DAQ groups should follow the Recommendation to communicate closely with other JLab experiments and the JLab computing center regarding data handling. ETA 2-3 months. This should yield:
  - Estimate of required computing resources
  - Plans for calibration and production analyses
  - Specifications for the level-3 trigger farm
  - Performance expectations for simulation and reconstuction software (time per event etc.)

Software Planning In Progress

- Formed working group w/ Simulation and Reconstruction subgroups
- Weekly meetings
- Considering Hall A/B/D & Phenix frameworks for ideas
- Developing specifications (=features wanted and how to get them)
- Computing document being drafted (`JeffersonLab/SoLID-docs-softspec` on GitHub)
  - Hall-B document: `http://www.jlab.org/Hall-B/clas12-software-nov13.pdf`
  - Hall-D document Gluex-doc-2350: `http://argus.phys.uregina.ca/cgi-bin/private/DocDB/ShowDocument?docid=2350`
- To-do/task list being developed

Software Design Specifications — Key Points

- Simulation/General
  - Simulation engine: Geant4
  - Simulation framework under discussion: `GEMC`, `GEMC-SoLID`, ...
  - Database access convenience, consistency between simulation and reconstruction, esp. geometry
- Reconstruction
  - Reconstruction software to be based on C++ & ROOT
  - DST file format: ROOT (unless anyone objects)
  - Reconstruction should support its own DST output format as input format (for multi-stage analysis)
  - User interface: ROOT (CINT/Cling) interpreter?
    (alternatives: command line/shell scripts, Python)

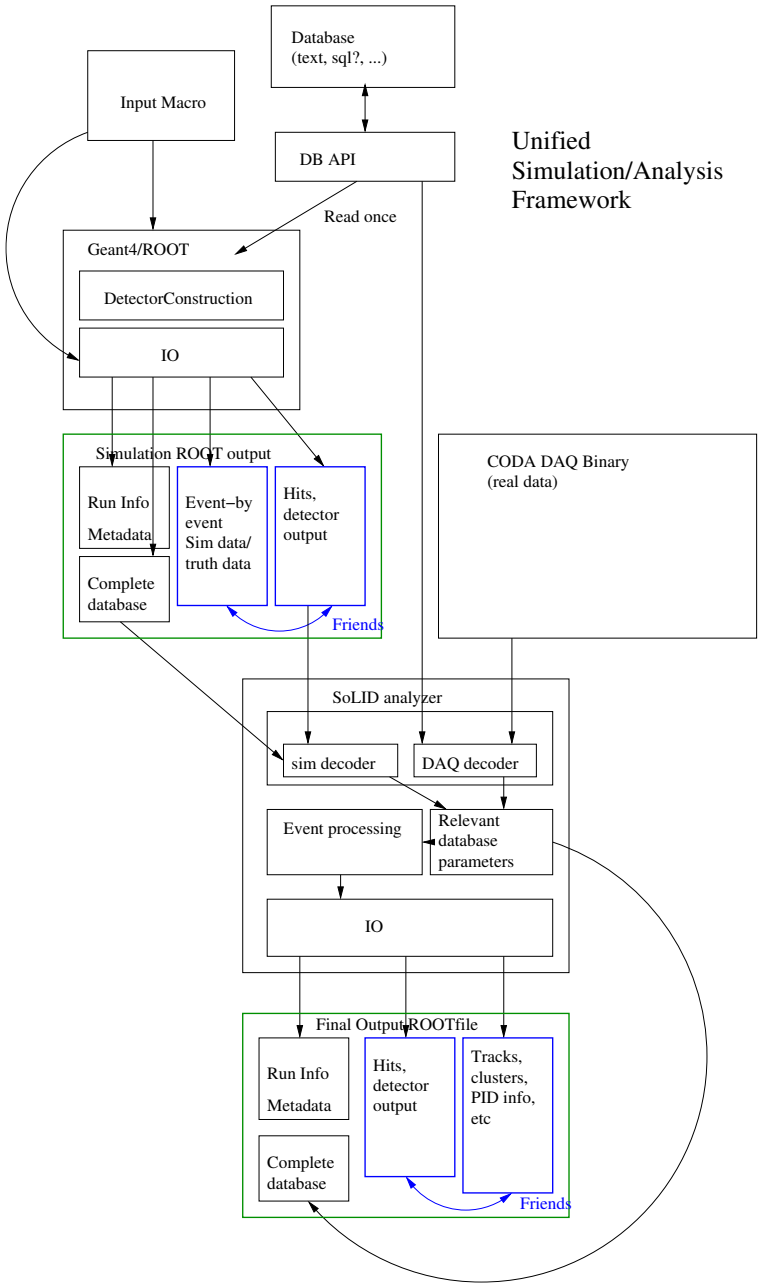
Unified simulation and analysis database API

General considerations

- Single common database abstraction
- Database holds "core parameters", esp. for geometry information
- Each component expands these parameters to a suitable internal representation
- Store sets of core parameters (for relevant run numbers) as objects in output files

Database requirements

- Should be easy to set up a local database
- DB should contain change history
- DB should support "variations" and local overrides of parameters
- Hall B's CCDB seems to fit the bill



Framework Comparison (very preliminary)

Feature	Hall A Podd	Hall B CLARA	Hall D JANA/DANA	Phenix Fun4All
Language	C++	Java & C++	C++	C++
Base Package	ROOT	–	–	ROOT
Raw data format	EVIO	EVIO	EVIO	(non-EVIO)
DST format	ROOT	EVIO	REST (HDDM)	ROOT(?)
Configurable Output	yes		no	no
Database	Text	CCDB	CCDB, XML	
User Interface	CINT	Groovy	command line	CINT
Plugins	yes	yes(?)	yes	yes
Multi-threaded	soon	yes	yes	no(?)
Distributed	no	yes	no	yes(?)
Multi-stage analysis	no		yes	yes
Recalibration support	no	no	no	yes
Sim truth data API	yes		yes	yes
Event display	no	yes	yes	

Framework Selection

- Continue study of available frameworks, esp. Hall B's
- If appropriate, invite presentations from other Halls' experts
- Probably will not find any single available framework optimal
  - Pick one (least work now, but limitations may cost time later)
  - Build a new one, based on an existing one and using the best ideas of the others (more work, we will get what we want, but bugs may cost time later)
  - How much time can we afford to spend?

Tracking Studies

- Continuing tracking studies (within Hall A Analyzer framework)
- Weizhi porting progressive tracking code to Hall A analyzer digitization framework, will test on PVDIS data
- Need to do SIDIS digitization & analysis. Requires implementation of SIDIS geometry in both digitization and reconstruction code, which will be time-consuming