LArSoft Work Plan for 2018

by: Erica Snider and Katherine Lato
Last updated 3/8/17

Introduction

This document records the LArSoft work plan for 2018. Progress on specific items going forward are recorded in redmine issues with updates to this document given quarterly at Steering Group meetings.

Erica and Katherine discussed 2018 priorities with the Offline leads starting in October of 2017 and through November. The experiments detailed their plans for the next year, the implied requirements for LArSoft, and how LArSoft could help, as well as what the experiments might be able to contribute to LArSoft code. Some items were raised by LArSoft, not a particular experiment.

Based on those discussions, LArSoft proposed short-term and long-term priorities outlined below at a Steering Meeting in December 2017. Note, when defining work, we include items that can be done by members of the collaboration. There is a cost associated with making things workable for other experiments, but the benefit is that other experiments develop software that is usable by all experiments. The more this happens, the more all experiments benefit.

Short-term priorities; happening concurrently

These are happening concurrently. The order does not imply priority.

1. **Investigate having a new Event Display framework common to all experiments**. Some important features:
    a. Provides a simple, intuitive means to navigate through large events and multi-TPC detectors.
    b. Ability to zoom and pan with a data density that matches the screen resolution.

c. Provides basic interfaces to art for controlling runs and events, the geometry service, run conditions services. Re-running reconstruction workflows and displaying the result should be possible.

This work does not cover development of drawing routines. It is expected that the core event display users within the experiments will provide this code. Arrangements for this follow-on work will be part of this project.

Tracking this item is via https://cdcvs.fnal.gov/redmine/issues/19034
Resources: ?? + pre-arranged people from the experiments

3/7/18 status:

Requirements have been gathered and ranked. These are being reviewed by the community at LArSoft Coordination meetings. The plan for this investigation is documented and tracked in LArSoft issue 19034, and is summarized below:

i) Develop a set of requirements, ranked according to the following categories: essential, strongly-recommended, desired, and possibly-useful.
ii) Get community input on the ranking and final list of requirements at two LArSoft Coordination meetings to ensure that people can give opinions.
iii) Establish criteria to judge between the technical options, and review these with the community.
iv) Identify a list of possible technologies such as Root, Venu, ParaView, QT and other potential technologies and how applicable they are.
v) Compare technologies. Ideally, write a toy display that demonstrates the requirements, or sufficient subset of critical requirements. Code it up on each platform
vi) Evaluate based on established criteria in task 19037
vii) Take the results to SCD management for review, along with an initial deployment proposal and a request for funding.

2. **Pixel-detectors within LArSoft**. The simulation moves to LArG4 phase 2. The reconstruction piece has two possible scenarios.
    a. Pattern recognition algorithms shared between pixel and SP/DP LArTPCs. This requires agreeing on a common data format to represent 3D data, whether obtained directly from a pixel detector, or inferred from 2D views in SP/DP detectors. There will likely be implications for the geometry interface, which will then need to support both conventional and pixel-based detectors. Some study will be needed to determine whether these constraints can be met. If code is shared, then support for the development of those algorithms will also be needed.

b. No sharing of pattern recognition algorithms between pixel and conventional SP/DP LArTPCs. The code bases in this case diverge, so no further work is required aside from the normal support for data product, service and algorithm development.

Resources: ??

2/19/18 status: Need agreement on the data representation with the pixel community which we were hoping to get at the 3D reconstruction workshop. Since that hasn't been scheduled, to move ahead, we propose using the Space Point Associated Charge in the LArSoft repository. There may be additional requirements, and LArSoft is willing to consider them as they arise. We'd like to agree on the information that needs to be carried around. It may not matter if we agree on the truth part.

Next step: Erica will schedule a meeting with Alan Bross (and his compatriots), Erica Snider, Tracy Usher, Brett Viren, Jonathan Asaadi, Jen Raaf, Chris Backhouse, Daniel Dwyer, Kazu Terao, Xin Qian

3. **Use of SIMD vectorization to optimize LArSoft algorithms**. Introduce vectorized types to redesign some existing LArSoft data structures and modify existing algorithms, with the goal of exploiting vectorization opportunities to significantly improve performance of existing LArSoft algorithms. This project involves profiling real use case scenarios, redesigning data structures and applying changes to the most performance-critical algorithms, documenting performance comparisons and other studies, preparing reports and writing documentation as needed. Addresses in part the 2017 long-term priority on concurrency.

   https://cdcvs.fnal.gov/redmine/issues/17920
   Resources: Giulherme Lima

   2/19/18 status: Big loops over waveforms are candidates for this work: DetSim (SimWire), WireCell, CalWire. Guilherme consulted with Brett about WireCell. Brett noted that updating the fftw library to a new release recently produced a factor of 4 improvement in CPU performance. Guilherme surmised that this meant that they are using a vectorized fftw, so is not a candidate. DetSim and WireCell, must choose one or the other until the LArG4 refactoring is done.

   SimWire shows up in Soon's profiling results as a 60% consumer of CPU, verifying this as a good target. Goal at this point is still only to demonstrate on some piece of code that this strategy can have big benefits on performance in places.

4. **Optimization / profiling work**. Profile production LArSoft workflows to identify problem areas and potential solutions. This work includes:
   a. setting up an application to run LArSoft

b. identifying a major production workflow that will be the target of the profiling work
c. preliminary profiling of LArSoft applications, looking for opportunities for optimization if provided
d. promoting ease of use for profiling tools. --

https://cdcvs.fnal.gov/redmine/issues/17921
Resources: Soon Jun

2/19/18 update: LArSoft team has examined profiling results for ProtoDUNE, 35t and FD production workflow chains compiled by Soon Jun Yun. Noted that TensorFlow was the dominant consumer of resources in reconstruction phase, with few other large targets. At the 2/14/18 Offline Leads meeting we asked the experiment to provide a person to review the results in more detail with LArSoft team and identify actionable bottlenecks.

3/7/18 update: Thomas Junk will look at the profiling results (both memory and CPU) available at https://g4cpt.fnal.gov/LArSoft.html

From a LArSoft perspective, this work is now almost done. Soon Jun is providing help in easing the use of profiling tools.

5. **Error handling policy for LArSoft.** The policy should also prescribe what common conditions constitute an "error" versus a "warning", etc. An education campaign will then be needed to disseminate this information.

   https://cdcvs.fnal.gov/redmine/issues/12778
   Resources: Katherine Lato, Erica Snider

   3/7/18 - Work on this item was deferred to the second quarter of 2018.

6. **Introduce support for global wire coordinates**. The requirements for this work are documented in https://cdcvs.fnal.gov/redmine/issues/11522. At present, DUNE has implemented an ad hoc solution to provide global wire coordinate functionality. This project will provide a native LArSoft solution. Addresses in part the 2017 long-term priority of the TPC / locality interface and optimization in the Geometry service.

   Resources: Leigh Whitehead

   3/6/18 status: WIth the departure of Robert Sulej, Leigh Whitehead becomes the new point of contact. He has been working on CNN imagine ID, one of the two types of reconstruction/event selection algorithms that can benefit from this. The version he was using for the CVN isn't quite complete and generic. He looked for one that Robert Sulej had mentioned, and found the current version of the code. It works in the wire channel / TDC space as required for the image production, but might be useful for standard

reconstruction too if it is included at the early steps and continued through the clustering and tracking.

LArSoft goal is to have this "generic" version use the existing LArSoft infrastructure, and then infer a useful interface to make code like that easier to write.

7. **Architecture-dependent libraries**. Support is needed to allow coexisting builds that include / exclude, support for different computing backends under a given OS flavor.
    a. An immediate goal will be to allow the standard setup procedure to support a generic library as well as one built with avx2, or other vectorization technologies
    b. More generally, there are a number of areas that might benefit from this feature, such as allowing use of GPU backends when available, or selecting the optimal SIMD instruction set for a given computing node.The types of backend support required needs to be studied, and appropriate tools adapted or developed to allow the required setup procedures.
    c. Figure out if running LArSoft on NERSC is coupled to this.

   Resources: ??

   3/7/18: Since we are waiting on Spack for this, an issue has not yet been created.

8. **LArG4 re-factoring work.**
    - Phase 1: Continuing with existing plan to deliver equivalent functionality to the current code, but re-structured so as to de-couple various models from LArG4 and, where possible, from Geant4, and to expose the result of energy deposition by Geant4, and more generally by upstream particle interaction simulation.
    - Phase 2: Abstract the anode simulation model, which is required to handle conventional single-phase, dual-phase, and pixel based detectors within the common framework.

   https://cdcvs.fnal.gov/redmine/issues/14454

   Resources: Hans Wenzel and William Seligman for phase 1.

2/21/18 There is a May 1 deadline from MicroBooNE for incorporating this code into MCC9, their next major MC challenge / production processing run. Hans expects that the first stage of work (which meets MicroBooNE's needs) will be completed on that time-scale, though there appears to be some risk in meeting the deadline. Bill Seligman, who is working on the experiment-side component, is ready to test the downstream changes. The re-factoring is currently blocking Wirecell integration testing, so MicroBooNE is proceeding with a temporary backup plan that takes considerable shortcuts in achieving the re-factoring so that testing can take place.

3/7/18 update - Hans Wenzel gave a LArG4 re-factoring status report at the Feb. 27 LArSoft Coordination Meeting. The presentation highlighted the relationship with the Geant4 Collaboration, and discussed various new features of the GDML configuration, such as defining step limits for specific volumes. The G4 collaboration has implemented our requests for a new interface at the energy deposition level, eliminated performance bottlenecks, and improved cross sections e.g. for kaons. An output data translation layer is still required. No time estimates or task lists for completion were provided, though Hans informally mentioned that the changes related to Geant4 will be completed on the timescale of a few days to a week. The downstream code is ready for testing, though experiments will need to update detector-specific GDML files in order to participate.

9. **SPACK - new build system for art and LArSoft** (deferred from 2017).
   Purpose:
   a. To migrate to a standard set of build tools that have broad community support
   b. To address portability and configurability issues raised by experiments with the current build system
   c. To allow continued full support of Mac OSX as a LArSoft development platform within the context of Apple's System Integrity Protection system

   http://cdcvs.fnal.gov/redmine/issues/15313
   Resources: Chris Green

10. **Re-architecture of art services in LArSoft to ensure thread safety**.
    Gianluca Petrillo, Erica Snider and Jason Stock met with art team leader Kyle Knoepfel in December 2017 to continue the discussion of art service use cases within LArSoft in the context of thread safety. The discussion concluded with an agreement on how to re-architect a certain class of services (such as those that tag event-by-event transients in various electronics channels, and the calculation of event trigger times to be used in the global clock service) such that they are stateless, and therefore inherently thread safe, while still providing the capability of on-the-fly updates. The LArSoft team will work on an execution plan for this project.

    https://cdcvs.fnal.gov/redmine/issues/19288
    Resources: Erica Snider, ??

    2/26/18 update: Need to know what LArSoft has to do to be ready. Have had meetings to discuss the models with the art team and LArSoft.
    3/7/18 - Have identified a person to work on this, and will develop a plan over the next couple of weeks. The plan will be presented at a LArSoft Coordination Meeting in April

11. **Support for transitioning code to art-independent repositories**, making them available to run in external frameworks such as gallery, in order to meet the growing demand to run LArSoft code in art-independent contexts.
    2/19/18 status:
    - LArSoft has a written policy for managing this code.
    - Repositories for the code have been provided.
        - https://cdcvs.fnal.gov/redmine/issues/18302
    - LArSoft has to move code to these repositories. Gianluca is doing this after we move to art 2.10.
        - https://cdcvs.fnal.gov/redmine/issues/17194
        - https://cdcvs.fnal.gov/redmine/issues/18281
        - https://cdcvs.fnal.gov/redmine/issues/17179
    - Once done, there will be a presentation at a Coordination Meeting.

    Resources: Gianluca Petrillo, Lynn Garren, Erica Snider

Longer-term priorities

We welcome input from the Steering Group to help shape the long-term priorities of LArSoft.

A) Concurrency -
- Purpose: To use multi-threading to address memory usage issues and provide flexibility in resource utilization - working with art on this, starting with understanding thread safety issues in the LArSoft use cases for art services.
- To introduce vectorization to those components of the code where speed improvements could be obtained, and to make use of currently unutilized resources - see short-term item above
- Resources: LArSoft team and (proposed) experiment effort

B) Architectural changes to define algorithm interfaces for layered algorithms.
- Purpose: Provide a well-defined set of interfaces for layering algorithms within a framework that allows for run-time configurability below the level of art modules, e.g., the art "tool" currently under development
- The primary target for this change are the algorithms used to generate, process and deconvolve raw signals from the detector
- A major thrust is to provide a set of tools that will allow algorithms that are currently detector dependent to be generalized and incorporated into the core LArSoft code
- Resources: The LArSoft team and (proposed) experiment effort.
- DUNE has already contributed effort to a similar project

C) TPC topology / locality in the Geometry service.
- Purpose: To provide a mechanism to efficiently determine the TPCs adjacent to a given TPC, as would be needed to follow tracks and showers that cross TPC boundaries
- Resources: The LArSoft team and (proposed) experiment effort
- http://cdcvs.fnal.gov/redmine/issues/9818 -- requested by DUNE

- 11/15/17 - asking DUNE for priority call on this.

Other topics under consideration to work on include the list of accepted, but not assigned, redmine issues.
These can be found under accepted redmine issues.