Experience with a Physicist-Developed, Collaborative, Large-Scale Software Project for Discovering the Nature of Neutrinos

Erica L. Snider
Fermi National Accelerator Laboratory[1]
P.O. Box 500, Batavia IL 60510
erica@fnal.gov

Abstract— **We will discuss the features of new liquid argon detectors to study the properties of neutrinos, which may hold the answer to a number of fundamental questions about the nature of the universe. We will describe our experience in developing and maintaining a large scale collaborative software suite to enable analysis of data from these detectors, despite vast difference in scale and configuration.**

AUDIENCE: [Software Architecture and Design][Software Development and Deployment][Software Engineering Foundations]

[Beginner Technical Talk]

I. Introduction

The properties of neutrinos [1], sub-atomic particles that scarcely interact with matter, may hold the answer to a number of puzzles in modern particle physics, such as why the universe consists of matter [2]. Due to their highly elusive nature, however, they have proven to be extremely difficult to study in detail. A new neutrino detector technology, the liquid argon time projection chamber (LArTPC) [3], promises to change this. When placed in intense accelerator-driven neutrino beams, such as the NuMI [4] and LBNF [5] beams at Fermilab, LArTPCs will usher in a new era of precision measurement of neutrino properties that may finally provide answers to some of these long-standing questions.

The various LArTPC-based experiments that are now in operation or under development [6],[7], however, face a number of challenges in exploiting this data. Unlike previous generations of particle detectors, these instruments provide richly detailed digital images of particle interactions with excellent spatial and charge resolution. Developing the software needed to extract neutrino properties at the desired accuracy and precision from this new type of data will require innovative algorithms, and represents in itself a significant research and development project. In an effort to overcome these challenges with the limited effort available, the experiments have elected to exploit some of the general features of LArTPCs to embark on a unique undertaking in high-energy phyiscs, in which the experiments jointly develop and share the needed software solutions.

The LArSoft Collaboration [9] is the group of neutrino experiments representing well over 1000 neutrino physicists from around the world, working with laboratories and other software projects, that develop and use this set of shared software algorithms for the simulation, reconstruction and analysis of data from LArTPCs.

In this presentation, we will discuss the features of LArTPCs that enable the use of common software across detectors of vastly different scale. We will then describe some of the details of the LArSoft collaboration and the particular challenges posed by collaborating on software that until this project has been wholly owned by individual experiments. We conclude by discussing and how the collaboration overcomes the software engineering and organizational challenges in order to realize the benefits that shared software solutions offer for these critical pieces of software. The presentation will be given from the perspective of a physicist serving as technical lead of the collaboration.

II. General Features of LArTPCs

The LArTPC consists of a volume of liquid argon with a uniform electric field applied. Charged particles created in neutrino-argon interactions traverse the liquid and produce trails of ionization and scintillation light along their trajectories. The electric field drives the deposited ionization electrons at a known velocity to one end of the TPC where they induce signals on two or more planes of parallel readout wires or strips. Each readout plane is perpendicular to the electric field and the drift direction of the electrons. The combined, digitized signals on the strips in a given plane yield a 2-D image of the deposited ionization projected onto a plane perpendicular to the strips. When combined with the views from other readout planes, it is possible to construct a 3-D image of the particle trajectories and associated charge deposition. The scintillation light is (typically) observed using photo-detectors arrayed around the drift volume, and is used to estimate the time of the interaction. When subtracted from the arrival time of the ionization, the exact position of the 3D image can be inferred.

This basic description applies to every LArTPC for every neutrino experiment that uses this technology, and forms the foundation for shared software. The size of individual TPCs, the operating parameters, the number and configuration of readout strips and planes within the TPCs, the number of TPCs within an experiment, etc., although vastly different

[1] Operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.

among the various experiments,[2] can be abstracted into interfaces to otherwise generic classes. Similarly, the data representations, from the digitized waveforms from the TPC and photo-detectors, to reconstructed objects such as tracks and electro-magnetic showers, through to the end products of the reconstruction — a complete description of the neutrino interaction in terms of the final state particles produced — all lend themselves to common definitions, interpretations and interfaces. Algorithms can then be fashioned to use these common interfaces to construct work-flows that extract features from the data based solely on the physics of charged particle interactions within liquid argon, which is the same across all detectors. Most importantly, by using the provided generic interfaces, all of the algorithms can be written such that they are detector independent. The direct, code-level leveraging of expertise and experience across experiments that results is one of the primary benefits that accrues to collaborating experiments. For small, effort limited experiments such as LArIAT [7], the interoperability of algorithms provided by LArSoft is an enabling technology.

III. LArSoft: the Collaborative Solution

At present, there are five neutrino experiments participating in the LArSoft collaboration, with another engaged in on-boarding activities. Each of the experiments contributes to the shared toolkit according to the needs of their own experiment. By adhering to an agreed upon set of design principles and practices, these algorithms can function as designed on any LArTPC within the project with only changes in run-time configuration. At present, the 17 shared repositories contain about 260k lines of C++, with another 200k lines across experiment-specific repositories. More than 100 authors from over 25 institutions have contributed to the core software suite.

To provide for the common needs of the collaboration and its members, a dedicated group of physicists, software engineers, developers and others comprise a core project team that works under management oversight by the experiments. Major elements of the core projects mission include provisioning and supporting the LArSoft framework, software architecture and design; performing release management and testing functions; coordinating code integration and planning across experiments; and providing software engineering expertise to the LArSoft community.

IV. Outcomes: Addressing the Challenges of a Collaborative Solution to Proprietary Software

In a typical high-energy physics experiment, the participating physicists develop custom simulation, reconstruction and analysis software. With physics results and the reputation of the experiment at stake, the content of and access to this software is usually tightly controlled. To thrive within this culture, the collaboration must at once provide positive value added to each of the participating experiments, while also allowing each experiment to pursue independent physics agendas, schedules and priorities, and to handle detector-specific customizations, all without compromising the confidentiality of experimental data and results. In short, each experiment must feel as though it has a custom solution that is attained with only a fraction of the effort otherwise needed to produce it.

Communication is the key: Developing code within such an environment requires excellent communication and coordination across experiments. The core LArSoft project facilitates this exchange in a number of ways. The primary meas is via a regular "coordination meeting." Members from the various experiments attend this highly technical forum to propose, discuss and coordinate code and policy changes, release plans and other issues relevant to the collaboration.

The project team also gathers requirements on a regular basis via a special series of meetings with experiment software coordinators, and more broadly from the community at dedicated workshops. Discussion of major technical issues and initiatives and short term project work occur in regularly scheduled meetings with experiment software coordinators.

The overall direction of the collaboration, and the priorities and goals of the core project are set or approved in quarterly meetings with the experiment spokespeople.

Our developers are physicists! Most of the physicist-developers have little or no prior programming experience, let alone software design or engineering training. To help improve the quality of the code, software engineers and other experts regularly contribute to special projects for the benefit of the community via the core project. These individuals also participate in targeted code reviews conducted in collaboration with code authors. Finally, the core project team provides consulting services, and runs tutorials and workshops with collaboration members and software experts in an effort to enhance the skills of contributing physicists, improve the compliance of the code they produce with the LArSoft design principles, and generally promote a culture of problem solving.

Continuous integration testing: The broad and rapid pace of development, coupled with an aggressive integration release schedule and constant pressure to minimize time to production, demands a robust testing regime to ensure that the head of the main development branch always works, and that changes made by one experiment do not have adverse effects on another. LArSoft uses a Fermilab-supported continuous integration system built on top of the Jenkins open-source automation system [8]. The system provides a number of features and tools to facilitate construction of integration test workflows, analysis of program output, and tracking of historical CPU and memory usage.

With every commit to the central repositories, the system automatically runs a rapid cycling test suite for all experiments. With these tests, core project and experiment software coordinators can quickly determine whether changes had

[2] To give a flavor for the range of variation in the detectors accommodated by LArSoft, the smallest experiment uses a single TPC less than a meter cubed with only two readout planes. The largest, when the first stage is completed, will have over 100 TPCs, each on the scale of 10 meters with three readout planes, where some of the planes share readout channels with adjacent TPCs.

unintended effects. The system routinely finds several such defects per week before they go into releases.

Developers define and configure tests via simple local configuration files. The system also allows developers to run the entire test suite against their changes prior to being merged with the main development branch.

Framework flexibility: The architecture of LArSoft is layered such that data structures and algorithm code are uncoupled from the underlying event processing framework that is bundled with the code. This feature not only simplifies unit testing, but also allows experiments to develop and run components of LArSoft in an environment of their choice, with seamless integration back into the core repositories. The flexibility gained through this design allows developers access to a broad range of development environments, should the native system not suite their tastes.

Document, document, document: Finally, the core project provides dedicated effort to create and maintain documentation and learning materials for collaboration members. LArSoft.org [9] provides a one-stop starting point for new users, with links to guides, information on the overall conceptual framework, basic workflows and features of LArSoft, and LArSoft policies, design principles, etc. The site also hosts high-level descriptions of algorithms submitted by individual developers. More detailed guidelines for developers, code submission procedures, release notes and the like are maintained on a separate wiki [10]. All of this content is regularly reviewed and updated. LArSoft makes use of the Doxygen [11] system to disseminate low-level documentation of code interfaces and implementations.

In all interactions and tasks, the core project seeks to accommodate the particular needs of the individual experiments and developers, crafting balanced solutions and adjudicating differences as needed in order to maintain and enhance the code sharing regime, the quality of the code and the development environment, and the overall strength of the collaboration. In return, the collaborating experiments remain highly engaged in LArSoft at all levels, from providing new ideas and requirements, to contributing new code, or using and improving existing code shared by other experiments. The collaboration is a vibrant community, with many plans and ideas for future work.

PARTICIPATION STATEMENT

If accepted, Erica Snider will attend the conference and make this presentation.

BIO

Erica Snider received her Ph.D. in high energy physics from University of Chicago, and has been a staff physicist in the Scientific Computing Division of the Fermi National Accelerator since 1998. As a member of the CDF experiment, she studied multi-particle production, top and bottom quarks, and searched for the higgs boson and evidence for physics beyond the standard model in proton-anti-proton collisions. She moved to the study of neutrino interactions at the MINERvA experiment in 2010, and more recently at the MicroBooNE and DUNE experiments. Beginning early in her graduate work, she has acquired extensive experience within software and computing for high energy physics, from writing reconstruction software, coordinating software development as leader of reconstruction groups, to serving as coordinator of software and computing infrastructure for CDF. Beyond experiments, she served as co-chair of the Open Science Grid Council in 2011 and 2012, and in ISO20k-based IT service management roles for Fermilab, and holds an ITIL service management certification. She has served as technical lead of the LArSoft project since 2013.

REFERENCES

[1] See for instance, R. Oerter, *The Theory of Almost Everything: The Standard Mode, the Usung Triumph of Modern Physics*, Penguin Group, ISBN 0-13-236678-9 (2006)

[2] M Fukugita and T Yanagida, Phys. Lett. B **174**, 45 (1986).

[3] C. Rubbia, *Preprint* CERN-EP-INT-77-08 (1977).

[4] K. Anderson *et al.*, *The NuMI Facility Technical Design Report*, FERMILAB-DESIGN-1998-01 (1998); P. Adamson *et al.*, Nucl. Instrum. Meth. A **806**, 279 (2016).

[5] C. anderson *et al.*, "The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab", JINST **7**, P10019 (2012). See also `http://t962.fnal.gov`; J.Strait *et al.*, "Long-BAseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Volume 3: Long-Baseline Neutrino Facility for DUNE", FERMILAB-DESIGN-2016-03, arXiv:1601.06823 [physics.ins-det]

[6] R. Accarri *et al.*, 2016, Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report Vol. 1: The LBNF and DUNE Projects, arXiv:1601.05471. See also `http://www.dunescience.org/`; R. Accarri *et al.*, JINST **12**, no. 02, P02017 (2017). See also `http://www-microboone.fnal.gov`; See R. Accarri *et al.*, "A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam", arXiv:1503.01520 (2015). See also `http://sbn-nd.fnal.gov.`

[7] J. Paley *et al.*, "LArIAT: Liquid Argon In A Testbeam", arXiv:1406.5560 (2014). See also `http://lariat.fnal.gov.`

[8] See `https://jenkins.io|.`

[9] R. Pordes and E. Snider, *Proceedings of Science (ICHEP2016)*, 182 (2016). See also `http://larsoft.org.`

[10] `https://cdcvs.fnal.gov/redmine/projects/larsoft/wiki`

[11] `http://www.stack.nl/ dimitri/doxygen/`