DUNE Software Management Welcome and To-Do list

Tom Junk

DUNE Software Management Meeting

October 26, 2017

Many thanks to Ruth for suggestions!

Fermilab DUNE

The DUNE Software Management Group

- Official Roles:
  - Christoph Alt: dunetpc librarian
  - David Adams: dunetpc release manager
  - Tom Junk: ex DUNE software and computing coordinator (emeritus). Self-Charged with getting this group going.
- Man groups on DUNE and the ProtoDUNEs depend on software management and contribute to the thinking and deliverables
  - Brett Viren
  - Nektarios Benekos
  - Dorota Stefan
  - Maxim Potekhin
  - Karol Hennessy
  - Very large amounts of help from Lynn Garren and the art and LArSoft teams
  - Support and interest from the CERN Neutrino Platform
- Meeting time: Looks like Wednesdays at 9 or 10 AM CT are the best according to a doodle poll.
- This meeting's a one-off
- Mailing list? We don't have one. Should we?

CVMFS

- How-To use files in CVMFS, how to put them there:
  https://cdcvs.fnal.gov/redmine/projects/dune/wiki/Access_files_in_CVMFS
  which has a link to David Adams's how-to-make-releases page:
  https://dune.bnl.gov/wiki/DUNE_LAr_Software_Releases
- Responsibles (from the .k5login)
  - LArSoft-related DUNE code: David Adams, Tingjun Yang, Alex Himmel, Dorota Stefan, Tom Junk, Christoph Alt
  - Physics: Elizabeth Worcester, Evan Niner
  - Beam Simulations: Laura Fields
  - artDAQ components: Kurt Biery
  - Near Detector: Justo Albo-Simon, Brian Rebel
  - Management: Andrew Norman
  - Maybe need more ProtoDUNE responsibles?
  - BSM group may need a responsible

Some Jobs We Do

- Maintain the dunetpc repository
  - tag releases. As needed and when LArSoft releases
  - install releases: CVMFS, /grid/fermiapp (which is going away), scisoft.fnal.gov, CERN
  - Look after dependencies and version numbers.
  - Coordinate with maintainers of products DUNE supplies that dunetpc depends on. e.g. lbne_raw_data and duneutil
- Deferred maintenance tasks https://en.wikipedia.org/wiki/Deferred_maintenance
  - Clean up unused code (DUNE has high turnover, but people also come back)
  - Split into separate repositories (faster builds. We may move to a model where we can check out and build part of a repository in the future)
  - Modernize some code – GetValidHandle instead of GetByLabel for instance

Coordinate with the LArSoft team

- Transmit information/news/requests in both directions
  - Attend bi-weekly LArSoft coordination meetings
  - Monthly LArSoft Steering Group meetings (less regular now)
  - Bi-Monthly LArSoft Offline Leads meetings

LArSoft Coordination Duties

- Frequently the LArSoft team requires sign-off from the experiments when making breaking changes or simply updating to newer versions of important packages, like ROOT, GEANT, and GENIE.
- Tom's been the go-to person for that, but now that the reco teams report to the Physics Coordinators, Tom refers signoff questions to Ryan and Elizabeth, and also asks the reco coordinators.
- The LArSoft team also regularly prioritizes their work with input from the experiments. Tom attends these meetings, but needs feedback from people working on reco and analysis to make the most of this opportunity.
- Robert Sulej – part of the Fermilab SCD reco group: frequent interface with LArSoft and *art* developers and managers.

art Stakeholders' Meetings

- Tom has been the DUNE representative at the *art* stakeholders' meetings.
- Weekly at 3:00 PM CT on Thursdays
- artists also sometimes ask for permission from experiments before making breaking changes.
- Much of the discussion is driven by experiments needing new features or requesting different behaviors of *art* to make their work easier.
- Much of DUNE's issues come from memory and CPU needs
  - Multithreading – more cores using same memory: possible to improve efficiency of use of cores if we go this route. Thought needed!
  - DUNE events are quite large. Split them up?

Helping the LArSoft Team Help Us

- The LArSoft team recruits new effort – either new hires, or borrowing time from members of other groups.
  - Two new people onboard LArSoft:
    - Soon Yung Jun, borrowed from the FNAL GEANT group. Interested in profiling. "New" to liquid argon, though he has given tutorials at LArSoft and *art* meetings.
    - Guilherme Lima: Working on vectorization improvements. SIMD
  - Core team plus part-timers: Erica Snider, Gianluca Petrillo (getting a job at SLAC), Saba Sehrish, Giuseppe Cerati, Hans Wenzel
- They frequently need real live use cases to test their ideas out or to see what is needed by us. They seek our advice and suggestions for how our work can be done better or more easily.

DAQ Interface

- 3x1x1 Data
  - Import software by Kevin Fusshoeller, Christoph Alt and Andrea Scarpelli checked in to dunetpc and run on a sample of 3x1x1 data
  - Makes raw::RawDigits
  - Checked with gallery scripts – looks good!
- ProtoDUNE-DP Support needed moving forwards

DAQ Interface

- Tom attends ProtoDUNE-SP DAQ meetings when he can, and has interacted closely with JJ Russell (SLAC, RCE firmware provider), and Frank Filthaut (NIKHEF, FELIX provider) regarding data unpacking software that must run offline.
- 35-ton data format was awkward, and code was not optimized.
- ProtoDUNE-SP data will come in at least two formats – compressed and uncompressed.
- JJ wrote unpacker code that comes in three types:
  - "generic" – runs anywhere compiled code runs
  - "AVX" – requires a CPU with AVX instructions
  - "AVX2" – requires a CPU with AVX2 instructions
- JJ provided makefiles and utilities to build all three versions

DAQ Interface

- AVX2 version only builds on SL7 machines -- the compiler we distribute with LArSoft is new enough, but the assembler is not distributed. The system assembler is used, and on SLF6 it is quite old.
- Versions: (SL6 or SL7)x(prof or debug)x(generic,AVX or AVX2)
- Distinguished by flavor (SL6 or SL7) and qualifier (everything else) in UPS
- Tom has written a script to go through all the build options and make a UPS-friendly directory with the products and sources. Need to ups declare it and test it, and install it. v0_0_1 is JJ's version number.
- How to set up? Default should be generic, but allow user to pick the better one if possible.
- JJ supplied a script that queries /proc/cpuinfo and sets PATH and LD_LIBRARY_PATH appropriatley. Almost what we want.
- Lynn Garren is collecting a group of people to address this. This is not the only use case.

DAQ Interface To-Do

- Work with Frank Filthaut to provide similar unpackers.
- Need a module in dunetpc to call the unpacking methods and create raw::RawDigits. Memory should be allocated in the dunetpc module, filled by the unpacker, and std::Move'd to the raw::RawDigit data product.
- Jingbo Wang already has used JJ's initial unpacker to test online monitoring code.

Beam Instrumentation Interface

- This is being handled by the databases and the beam instrumentation group – Jonathan Paley is our database leader as well as a beam instrumentation expert.
- Caroline Zhang working on beam instrumentation simulation
- Igor Mandrichenko and others working on DB access, synchronization (IFBEAM)
- DRA group is looking for effort here.
- ProtoDUNE-DP approach – same DB. Need some responsibles. Timeline?

Software Repositories

- Redmine for all LArSoft-based code
- github is a favorite for many DUNE efforts
  - gegede
  - edep-sim
  - Near Detector work (some of it, others in Redmine)
  - DUNE's port of CAFANA
  - JJ's overlay classes for ProtoDUNE work
  - Many other uses, including our CDR (and TDR).

LArSoft Simulation Project

- Hans Wenzel has been working to separate the LArG4 step into two:
  1. GEANT4 part – produces G4 steps
  2. LArSoft part – LArVoxelReadout and downstream.
     - recombination
     - lifetime
     - diffusion
     - space charge
     - putting true charge on wires: output: sim::SimChannel data products

LArSoft Simulation Project Impacts

- Writing out the G4 steps allows for subsequent simulation processing to be plugged in more easily, as well as upstream parts.
  - WireCell simulation has a 3D drift and wire induction signal model
  - FLUKA can replace G4 upstream. S. Bordoni started this but now works on pion cross sections. May resurrect this if the G4 step interface becomes real
- Bill Seligman asked if we needed NEST moving forwards or if we would cry if it went away
  - NEST is a detailed recombination model by Matthew Szydagis and collaborators that takes pains to fit existing data and model a statistical anticorrelation between photons and drift electrons.
  - Our "separate" recombination model calculates the expected numbers of photons and electrons and throws independent random numbers.
  - The problem: NEST uses G4 routines. If NEST is rewritten not to do that, it can fit in with Hans and Bill's project.
  - We don't use NEST but have an interest in alternative models for systematic studies of the detector.

Continuous Integration

- Each push to dunetpc triggers a CI build and test (actually a waiting period to collect grouped commits).
- Documentation and links: https://cdcvs.fnal.gov/redmine/projects/lar-ci/wiki/Wiki
- Vito Di Benedetto has been our absolute hero on this
  - Watches experiments' CI test results
  - Looks for commits that are correlated with broken builds or failed tests
  - A script compares data product counts and sizes from a reference run to the curren test. Vito looks at the output of the script, locates a plausible commit, and alerts the relevant experiment
  - Tingjun and I get the DUNE e-mails and investigate what went wrong, prodding the person who committed problematic code or files to fix it, or, if the change is intentional, request Vito to update the reference files.
  - This is a lot of work for Vito. Erica has asked the experiments to start doing this work for themselves.

Software and Computing Tutorials

- Typically run as a "satellite" meeting around collaboration meetings
- Eileen Berman has been organizing them
- Many presentations from the FIFE, *art* and LArSoft teams
- We also like to give DUNE users some hands-on experience with code
  - homework exercises
  - working session time
- SCD says "we don't teach people how to program" but DUNE may need to fill that need for some collaborators. At least to give some practical advice on how programming work is done (work flows and asking critical questions).
- Next tutorial: DUNE Physics Week Nov. 14-17
- We have a request from J. Stock to provide a tutorial on how to make a CI test in dunetpc and larsoft
- Would like to add specific examples that work at CERN.

Support for Distributed Offline Computing

- duneutil contains production tools
  - DUNE-specific components for project.py
  - MCC XML files
  - 35-ton data-handling tools
- Supporting distributed computing at CERN has a software component: Nektarios's work