Deep Underground Neutrino Experiment (DUNE)

Technical Proposal

Software and Computing Overview

May 9, 2018

DEEP UNDERGROUND
NEUTRINO EXPERIMENT

Chapter 1

Overview of Software and Computing

1.1 Overview

Offline computing for DUNE faces new and considerable challenges due to the large scale and diverse physics goals of the experiment. In particular, the advent of liquid argon (LAr) TPC's with exquisite resolution and sensitivity, combined with enormous physical volumes, creates challenges in acquiring and storing large data volumes and in analyzing and reducing them. The computing landscape is changing rapidly, with the traditional HEP architecture of individual cores running Linux being superseded by multi-core machines and GPU's. At the same time, algorithms for LAr reconstruction are still in their infancy and developing rapidly. As a result, we have reason to be optimistic about the future but we are not able to predict it accurately. The ProtoDUNE single and dual phase tests at CERN in the fall of 2018 will provide a wealth of data that will inform the future evolution of the DUNE computing models.

The DUNE offline computing challenges can be classified in several ways. We will start with the different detector/physics configurations that drive the large scale data storage and reconstruction. This discussion leans heavily on the data acquisition (DAQ) design described in Volume 2: The Single-Phase Far Detector and Volume 3: The Dual-Phase Far Detector of the DUNE Technical Proposal.

1.1.1 Detectors

The DUNE experiment will consist of four 17 kT far detector modules located at the Sanford Underground Research Facility, using either single or dual phase Liquid Argon TPC's, and an, as yet unspecified, near detector at Fermilab. The proposed full-size 17 kT modules for the far detectors will have an active volume 12m high, 14.5m wide and 58m long.

1.1.1.1 Single-phase estimates

Each single-phase (SP) module will consist of 150 alternating vertical cathode and anode planes spaced 3.5 m apart and operated at 180 kV for a 500 V/cm drift field. The anode planes are made up of anode plane assembly (APA)s which are 6.3 m tall by 2.3 m wide and have 2,560 readout channels each. Each channel is sampled with 12-bit precision every 500 nsec. For modules of this size, drift times in the liquid argon are of order 2.5 ms and raw data sizes before compression are of order 6 GB per module per 5.4 ms readout window. With no triggering and no zero suppression or compression, the raw data volume for four modules would be of order 145 exaB/year.

1.1.1.2 Dual-phase technology

For dual-phase, electrons drift the full height of the cryostat, emerge from the liquid and are collected - after gas amplification, on an grid of instrumented pads at the top of the detector. The WA105 3x1x1 m test of this technology ran successfully in the summer of 2017[2]. Each 17 kT module will have 153,600 channels. Drift time in the liquid argon is 7.5 ms. Given 20,000 samples in an 8 ms readout, the uncompressed event size is 4.2 GB (for 1 drift window). Due to gas amplification, the signal to noise ratio is quite high, allowing loss-less compression to be applied at the front-end with a compression factor of ten, bringing the event size/module to 0.42 GB. Recording the entire module drift window can be considered a pessimistic figure, since events are normally contained in smaller detector regions. A far detector module can be treated as 20 smaller detectors (with similar number of readout channels to the prototype currently being constructed at CERN), running in parallel, each one defining a Region of Interest (ROI). For beam or cosmic events it is possible to record only the interesting ROI(s) with the compressed size of a single ROI being 22 MB.

1.1.1.3 Beam coincident rates

Requiring coincidence with the Long-Baseline Neutrino Facility (LBNF) beam will reduce the effective live-time from the full 1.2-1.5 sec beam cycle to a 5.4 ms (8 ms for DP) readout window coincident with the 10 microsecond beam spill, leading to an uncompressed data rate for beam-coincident events of around 20 GB/sec for four 17 kT single-phase detector modules ($\sim$ 16 GB/sec for dual-phase), still too high to record permanently. Only a few thousand true beam interactions in the far detectors are expected each year. Compression and conservative triggering based on photon detectors and ionization should reduce the data rate from beam interactions by several orders of magnitude without sacrificing efficiency.

1.1.1.4 Near detector

The near detector configuration is not yet defined but we do have substantial experience from T2K and MicroBooNE at lower energies, and MINERvA at the DUNE beam energies on cosmic and beam interactions under similar conditions. We can expect that a near detector will experience $\sim$ 5-10 beam interactions/beam pulse and non-negligible rates of cosmic rays, spread over an area of a few square meters. MicroBooNE experience and ProtoDUNE simulations indicate compressed event sizes of 100-1000 MB, leading to yearly data volumes of 2-20 PB. Storing and disentangling this information will be challenging but comparable to the ProtoDUNE data expected in 2018.

1.1.2 Physics Challenges

DUNE physics will consist of several different processes with very different rates and event sizes.

1.1.2.1 Long-baseline neutrino oscillations

Neutrino oscillation measurements will require a near detector operating in a high rate environment and far detectors in which beam-coincident events are rare but in time with the beam spill and of sufficient energy to be readily recognizable. Studies discussed in the DAQ section of Technical Proposal Volumes 2 and 3 indicate that high efficiencies are achievable at an energy threshold of 10 MeV, leading to event rates for beam-initiated interactions of $\sim 6,400$/year and an uncompressed data volume of around 30 TB/year per 17 kT single-phase module.

Tables 1.1 and 1.2 summarize the event and data rates after appropriate filtering from the DAQ section of Volumes 2 and 3 of the Technical Proposal.

1.1.2.2 Processes not in synchronization with the beam spill

These include supernova physics, atmospheric neutrinos, proton decay, neutron conversion and solar neutrinos. These processes are generally at lower energy, making triggering more difficult, and asynchronous, thus requiring an internal or external trigger. In particular, supernovae signals will consist of a large number of low-energy interactions spread throughout the far detector volume over a time period of 1-30 seconds. Buffering and storing 10 seconds of data would require around 2000 readout windows, or around 50 TB per supernova readout. At a rate of one such event/month, this is 600 TB of uncompressed data per module/year.

1.1.2.3 Calibration

In addition to physics channels, continuous calibration of the detectors will be necessary. It is likely that, for the far detectors, calibration samples will dominate the data volume. Cosmic-ray muons and atmospheric neutrino interactions will provide a substantial sample for energy and position calibration. Dedicated runs with radioactive sources and laser calibration will also generate substantial and extremely valuable samples. Table 1.1 includes estimates for the single-phase far detector. Cosmic ray and atmospheric neutrino signals collected for calibration make

Table 1.1: Anticipated annual, uncompressed data rates for a single SP module from the far detector Technical Proposals. The rates for normal (non-SNB triggers) assume a readout window of 5.4 ms. In reality, lossless compression will be applied which is expected to provide as much as a $4\times$ reduction in data volume for each SP module.

Event Type	Data Volume PB/year	Assumptions
Beam interactions	0.03	800 beam and 800 rock muons; 10 MeV threshold in coincidence with beam time; include cosmics
Supernova candidates	0.5	30 seconds full readout, average once per month
Cosmics and atmospherics	10	10 MeV threshold
Radiologicals (^{39}Ar snd others.	$\leq$1	fake rate of $\leq$100 per year
Front-end calibration	0.2	Four calibration runs per year, 100 measurements per point
Radioactive source calibration	0.1	source rate $\leq$10 Hz; single fragment readout; lossless readout
Laser calibration	0.2	1×10^{6} total laser pulses, lossy readout
Random triggers	0.06	45 per day
Trigger primitives	$\leq$6	all three wire planes; 12 bits per primitive word; 4 primitive quantities; ^{39}Ar-dominated

rates-sp

Table 1.2: Anticipated annual, uncompressed data rates for one DP module. The rates for normal (non-SNB triggers) assume a readout window 7.5 ms. These numbers do not include lossless compression which is expected to provide as much as a $10\times$ reduction in data volume.

Event Type	Data Volume PB/year	Assumptions
Beam interactions (DP)	0.007	800 beam and 800 rock muons; this becomes 700 GB/year if just 2 ROIs/event are dumped on disk
Supernova candidates (DP)	0.06	10 seconds full readout, all ROIs are dumped on disk
Cosmics/atmospherics (DP)	2.33	This becomes 230 TB/year if two ROIs/event are dumped on disk
Radiologicals (^{39}Ar snd other).	$\leq$1	fake rate of $\leq$100 per year
Miscellaneous calibrations	0.5	similar to SP
Random triggers	0.02	45 per day
Trigger primitives	$\leq$6	similar to SP

rates-dp

up the bulk of the uncompressed SP data volume at $\sim$ 10 PB/year per 17 kT module and will dominate the rates from the far detectors.

1.1.2.4 Zero suppression

The data volumes discussed above are for un-zero-suppressed data. Efficient zero suppression mechanisms can substantially reduce the final data volume but previous experience in HEP indicates that signal processing must be done carefully and often happens well into data-taking when the data are well understood. Experience from MicroBooNE and the ProtoDUNE experiments will aid us in developing these algorithms but it is likely that they will be applied later in the processing chain for single-phase. No zero-suppression is planned for dual-phase.

The constrained environment at the Sanford Lab motivates a model where any further data reduction via zero-suppression is done downstream, either on the surface or after delivery to computing facilities at FNAL or elsewhere. This could be analogous to the HLT's used by LHC experiments. The relative optimization of data movement and processing location is an important consideration for the design of both the DAQ and offline computing.

1.1.3 Summary

In summary, uncompressed data volumes will be dominated by calibration for the far detectors ($\sim$10 PB/year/module SP or $\sim$ 3 PB/year/module DP) and by beam and cosmic ray interactions in the near detectors (2-20 PB/year). With four far detector (FD) modules but a conservative factor of four for compression a total compressed data volume of 12-30 PB/year is anticipated.

After discussion with the SP Trigger/DAQ group, we asked them to include as limits in their design a maximum data transfer rate from the far detectors to Fermilab of 100 Gbit/s, which is consistent with projected network bandwidths in the mid 2020's and a limit of 30 PB/year raw data stored to tape.

1.2 Building the computing model

-cmp-mdl

The DUNE computing model is a work in progress. We can expect that major advances will take place over the next year on several fronts, with data from ProtoDUNE and the full incorporation of lessons from MicroBooNE into LArSoft .

The overall model can be divided into several major parts: Infrastructure, Algorithms and Adaption for the future. These are in different stages of planning and completion. An overarching theme is evaluating and using community codes and resources wherever possible.

1.2.1 Infrastructure

This category includes the wealth of databases, catalogs, storage systems, compute farms and the software that drives them. HEP fortunately has already developed much of this technology and our plan is to adopt pre-existing systems wherever possible. As DUNE is a fully global experiment, integrating the resources of multiple institutions is both an opportunity and a logistical challenge.

We are currently planning to have the primary raw data repository at Fermilab, with derived samples and processing distributed among collaborating data centers. For ProtoDUNE, raw data will also be stored at CERN. Data processing is being designed to run on HEP grid resources with significant ongoing effort to containerize it so that we can make use of heterogenous resources worldwide.

1.2.1.1 Core HEP code infrastructure

We plan to use shared HEP infrastructure wherever possible. Notably the ROOT[3] and `geant4` [4, 5] frameworks. For event simulation, we plan to use and contribute to the broad range of available generators (GENIE [6], NuWro [7]...) shared with the worldwide neutrino community.

In addition, we are using the infrastructure developed for the LHC and the Intensity Frontier experiments at Fermilab, notably grid infrastructure, the *art* framework and the Sequential Access via Metadata (SAM) data catalog. The NOvA and MicroBooNE experiments are already using these tools for distributed computing and the ProtoDUNE data challenges are integrating CERN and Fermilab storage and CPU resources. We are now extending this integration to the institutions within the collaboration who have access to substantial storage and CPU resources.

1.2.2 Algorithms

This category includes the simulations, signal processing and reconstruction algorithms needed to reconstruct and understand our data. Algorithms are currently under development but are informed by existing general codes (for example GENIE and `geant4`) and the experience of other liquid argon experiments as encoded in the shared LArSoft project. Simulations are quite advanced but full understanding of reconstruction algorithms will need real data from ProtoDUNE.

1.2.2.1 External products

The image-like nature of TPC data allows us to make use of external machine-learning systems such as TensorFlow[8], Keras[9] and Caffe[10]. Many of these are being evaluated for pattern recognition. While they encapsulate a wealth of experience, they are also somewhat volatile as driven by external needs. We must have access to and preserve the underlying source codes in order to maintain reproducibility.

1.2.3 Adaptability

As the experiment will be expected to run at least two decades past the present we must be prepared for the inevitable and major shifts in the underlying technologies that will occur. The ability to keep operating over decades almost requires that we emphasize open source over proprietary technologies for most applications. We should also plan to be able to utilize and support a large range of compute architectures in order to fully utilize the resources available to the collaboration.

sw:computingTasks

Table 1.3 summarizes the responsibilities of the Software and Computing group and Reconstruction and Algorithms groups for both DUNE and ProtoDUNE.

Table 1.3: Computing Tasks - see the ProtoDUNE section for details on current status.

Task	Status
Code management	in place
Documentation and logging of DAQ and detector configurations	in design
Data movement	design rates achieved for short periods
Grid processing infrastructure	early version in use for data challenges
Data catalog	sam, in place
Beam instrumentation and databases	ifbeam, in test
Calibration and Alignment processing	needs development
Calibration and Alignment databases	needs development
Noise reduction	tested in simulation
Hit finding	tested in simulation
Pattern recognition algorithms	tested in simulation
Event simulation	use existing software
Analysis formats	no common format
Distribution of analysis samples to collaborators	needs development

ingTasks

1.2.4 Downstream activities

The previous sections have concentrated on movement and recording of raw data, as that is most time-critical and drives the primary data storage requirements. Basic simulation and reconstruction algorithms are in place, but other components, in particular physics analysis models, are in a much earlier stage of development.

1.2.4.1 Simulation

Our simulation efforts will build on the combined experience of multiple neutrino experiments and theorists for inputs. We already have a solid foundation of event and detector simulation codes thanks to prior work by the LArSoft and event generator teams. However, even with good software in place, detector simulation in detectors of this high resolution is highly CPU and memory intensive and we are actively following projects intended to exploit HPCs for more efficiency. As simulation is much less I/O and database intensive than raw data reconstruction, (due in part to our ability to trigger efficiently on signal), we can anticipate resource contributions to this effort being distributed across the collaboration and grid resources worldwide. Simulation sample sizes orders of magnitude larger than the number of beam events in the far detector will be reasonably easy to achieve while near detector samples would need to be prohibitively large to equal the millions of events that will be collected every year.

1.2.4.2 Reconstruction

We have working frameworks for large-scale reconstruction of simulated and real data in place thanks to the LArSoft effort. These, and the simulations, have been exercised in large scale data challenges. Optimization of algorithms awaits data from ProtoDUNE.

1.2.4.3 Data Analysis

The data analysis framework has not been defined yet. We are working to build a distributed model, where derived data samples are available locally and regionally, similar to the LHC experiments. Provision of samples of ProtoDUNE data and simulated samples for the Technical Design Report will help define the analysis models that are most useful to the collaboration. However, previous experience on the Tevatron experiments indicates that data analysis methods are often best designed by end-users rather than imposed by central software mandates.

1.3 Planning inputs

1.3.1 Running experiments

The Fermilab intensity frontier program experiments (MINOS[11], MINERvA[12], MicroBooNE[13] and NOvA[14] have developed substantial computing infrastructure for the storage, reconstruction and analysis of data on size scales of order 5% that of full DUNE and comparable to the ProtoDUNE experiments. While the LArTPC technology requires unique algorithms, the underlying compute systems, frameworks and database structures already exist and are being adapted for use on both ProtoDUNE and DUNE.

For algorithms, the MicroBooNE[15] experiment has been running since 2015 with a liquid Argon-TPC which shares many characteristics with the DUNE APA's. MicroBooNE has, over the past year, published studies of noise sources and signal processing [16, 17], novel pattern recognition strategies [18, 19] and calibration signatures such as Michel electrons and cosmic rays [20, 16]. DUNE shares both the LArSoft software framework and many expert collaborators with MicroBooNE and is taking direct advantage of their experience in developing simulations and reconstruction algorithms.

1.3.2 ProtoDUNE

The ProtoDUNE single and dual-phase experiments will run in the Fall of 2018. While the detectors themselves have only 4-5% of the channel count of the final far detectors, the higher beam rates (up to 100 Hz) and the presence of cosmic rays make the expected instantaneous data rates of 2.5 GB/sec from these detectors comparable to those from the full far detectors and similar to those expected for a near detector.

In addition, the entire suite of issues in transferring, cataloging, calibrating, reconstructing and analyzing these data are the same as for the full detectors and are driving the design and development of a substantial array of computing services necessary for both ProtoDUNE and DUNE.

Substantial progress is already being made on the infrastructure for computing, through a series of data challenges in late 2017 and early 2018. Development of reconstruction algorithms is currently restricted to simulation but is already informed by the experience with MicroBooNE data.

In summary, most of the important systems are already in place or are in development for full ProtoDUNE data analysis and should carry over to the full DUNE. We have indicated where infrastructure is in place in table 1.3.

1.3.2.1 Single-Phase prototype

The single-phase prototype (ProtoDUNE-SP) will utilize six prototype APAs with the full drift length envisioned for the final far detector. In the single-phase detector, the readout planes are immersed in the liquid Argon and no amplification occurs before the electronics. ProtoDUNE-SP is being constructed in the NP04 test beamline at CERN and should run with tagged beam for around 6 weeks in the Fall of 2018. In addition cosmic ray commissioning beforehand and cosmic running after the end of beam are anticipated. Table 1.4 shows the anticipated data rates and sizes.

Table 1.4: Parameters for the ProtoDUNE-SP run at CERN

Parameter	Value
Trigger rate	25 Hz
Spill duration	4.5 s
SPS Cycle	22.5 s
Readout time window	5 ms
# of APAs to be read out	6
Uncompressed single readout size (per trigger)	230.4 MB
Lossless compression factor	4
Instantaneous data rate (in-spill)	1440 MB/s
Average data rate	576 MB/s
3-Day buffer depth	300 TB
Planned total statistics of beam triggers in 42 beam days	18M
Planned overall storage size of beam events	1.0 PB
Requested storage envelope for ProtoDUNE-SP	5 PB at FNAL, 1.5 PB at CERN

ata_rate

1.3.3 Dual-Phase prototype

The Dual-Phase prototype will either run in the NP02 beamline in late Fall 2018, or run at high rate on cosmics soon thereafter. Given the most recent construction schedule for the Dual-Phase prototype is now likely that the collaboration will forgo beam data taking and focus on detector performance assessment with two charge-readout plane (CRP)s read out and cosmics only. ProtoDUNE-DP will then run with cosmics at a rate going from 20 to 100 Hz from late Fall 2018 to at least April 2019. During six months of operation, with 50% efficiency, ProtoDUNE-DP is expected to collect about 300 million cosmic triggers at various rates, corresponing to a total data volume of 2.4 PB.

Table 1.5: Parameters for a 6 month ProtoDUNE-DP cosmic run at CERN

Parameter	Value
Trigger rate	20-100 Hz
CRPs read out	2
Uncompressed single readout size (per trigger)	80 MB
Lossless compression factor	10
Maximum data rate	≤800 MB/s
Cosmic rays over a 6 month run	300 M
Requested cosmic storage envelope for ProtoDUNE-DP	2.4 PB

ata_rate

1.3.4 Data Challenges

Computing and software is performing a series of data challenges to ensure that systems will be ready when the detectors become fully operational in the summer of 2018. To date we have performed challenges using simulated single and double-phase data and real data from cold-box tests of single-phase electronics. We anticipate average rates of $\sim$ 600 MB/sec but have set our design criteria at 2.5 GB/sec for data movement from the experiments to CERN Tier-0 storage and from there to Fermilab.

In data challenge 1.5 in mid-January 2018, dummy data based on non-zero-suppressed simulated events were produced at EHN1 and successfully transferred via 10-50 parallel transfers to the EOS (EOS) disk systems in the CERN Tier-0 data center at a sustained rate of 2 GB/sec. Transfers to dCache/Enstore at Fermilab achieved rates of 500 MB/sec.

Data challenge 2.0 was performed in early April 2018 is still being analyzed but preliminary estimated rates of 4 GB/sec from Experiment Hall North One (EHN1) to the tier-0 were achieved over several days. Rates to Fermilab disk cache were 2 GB/sec. Movement from FNAL disk cache to tape was substantially slower due to configuration for a lower number of drives than needed and contention for mounts with other running experiments. Fermilab is in the process of upgrading their tape facilities but we may require additional offsite buffer space if data rates from the experiments exceed the $\sim$ 600 MB/sec expected.

A subsample of the data was used for data quality monitoring at CERN and the full sample was reconstructed automatically on the grid using resources at multiple sites, including CERN. Our overall conclusion from this test is that most components for data movement and automated processing are in place. Remaining issues are integration of beam information, detector configuration and calibrations into the main processing stream, and faster tape access.

1.3.5 Monte Carlo Challenges

The collaboration has performed multiple Monte Carlo challenges to create samples for physics studies for the Technical Proposal and in preparation for the Technical Design Report in early 2019. In the last major challenge, MCC10 in early 2018, 17M events, taking up 252 TB of space were generated and catalogued automatically using the central DUNE production framework in response to requests by the Physics groups.

1.3.6 Reconstruction tests

Reconstruction tests have been performed on simulated single-phase ProtoDUNE test beam interactions with cosmic rays and an electronic noise simulation based on MicroBooNE experience. Hit finding and shaping is found to take around 2 minutes/event with a 2 GB memory footprint, leading to a reduction in data size of a factor of four. Higher-level pattern recognition occupies

10-20 minutes/event with a 4-6 GB memory footprint. For real data, calibration, electric field non-uniformities and other factors will likely raise the CPU needs per event. We will learn this when real data starts to arrive in late summer.

1.4 Resource planning and prospects

The DUNE computing effort relies heavily on the human and hardware resources of multiple organizations, with the bulk of hardware resources at CERN, and national facilities worldwide. The DUNE computing organization serves as an interface between the collaboration and those organizations. Computational resources are currently being negotiated on a yearly basis, with additional resources available opportunistically. Human assistance is largely on a per-project basis, with substantial support when needed but very few personnel as yet permanently assigned to the DUNE or ProtoDUNE efforts. We are working with the laboratories and funding agencies to identify and solidify multi-year commitments of dedicated personnel and resources for ProtoDUNE and DUNE, analogous to, but smaller than, those assigned to the LHC experiments. In-kind contributions of computing resources and people can also be an alternative way for institutions to make substantial contributions to DUNE.

The ProtoDUNE efforts in 2018-2019 will exercise almost all computing aspects of DUNE, although at smaller scale. Much of the infrastructure needed for full DUNE, in particular databases, grid configurations and code management systems need to be fully operational for ProtoDUNE. We believe that the systems in place (and tested) will be adequate for that purpose.

However, ProtoDUNE represents only 4-5% of the final volume of the far detectors and the near detector technology is, as yet, unknown. At the same time, computing technology is evolving rapidly with increased need for flexibility and the ability to parallelize codes. Liquid Argon detectors, because of their reasonably simple geometry and image-like data, are already able to take advantage of parallelization and generic machine learning techniques. We have good common infrastructure such as the LArSoft suite and `geant4`, and will have an excellent testbed with the ProtoDUNE data, but our techniques will need substantial adaption to scale to full DUNE and to take full advantage of new architectures. This will be one of our major challenges, and opportunities for collaboration, over the next five years.

Glossary

art A software framework implementing an event based execution paradigm. 6

detector module The entire DUNE far detector is segmented into four modules, each with a nominal 10 kt fiducial mass. 14, 15

secondary DAQ buffer A secondary DAQ buffer holds a small subset of the full rate as selected by a trigger command. This buffer also marks the interface with the DUNE Offline. 14

DUNE Deep Underground Neutrino Experiment. 1, 3, 6, 7, 9, 12, 13

anode plane assembly (APA) One unit the SP detector containing the elements sensitive to activity in the LAr. It contains two faces each of three planes of wires, cold electronics and photo detection system.. 2

data acquisition (DAQ) The data acquisition system accepts data from the detector FE electronics, buffers it, performs a trigger decision, builds events from the selected data and delivers the result to the offline secondary DAQ buffer. 1

dual-phase (DP) Distinguishes one of the four 10 kt detector modules of the DUNE far detector by the fact that it operates using argon in both gas and liquid phases.. iii, 15

Experiment Hall North One (EHN1) Location at CERN of the ProtoDUNE experiments. 11

EOS (EOS) The XRootD based distributed file system developed by CERN. 11

far detector (FD) Refers to the detector or more generally the experimental site in or above the Homestake mine in Lead, SD. 5

liquid argon (LAr) The liquid phase of argon. 1

Long-Baseline Neutrino Facility (LBNF) The organizational entity responsible for developing the neutrino beam used by DUNE. 2

Sequential Access via Metadata (SAM) A data handling system to store and retrieve files and associated metadata, including a complete record of the processing which has used the files.

6

single-phase (SP) Distinguishes one of the four 10 kt detector modules of the DUNE far detector by the fact that it operates using argon in just its liquid phase.. iii, 2, 15

HPC High Performance Computing Facilities - generally computing facilities emphasizing parallel computing with aggregate power of more than a teraflop.. 7

LArSoft Liquid Argon Software (LArSoft), a shared base of physics software across Liquid Argon (LAr) Time Projection Chamber (TPC) experiments.. 6–8, 13

MicroBooNE The Liquid Argon TPC-based MicroBooNE neutrino oscillation experiment at Fermilab. 3, 5, 6, 9, 12

MINERvA The MINERvA neutrino cross sections experiment at Fermilab. 3, 9

NOvA The NOvA off-axis neutrino oscillation experiment at Fermilab. 6, 9

ProtoDUNE Two prototype detectors operated in a CERN beam test. One prototyping single-phase (SP) and the other dual-phase (DP) technology. i, 1, 3, 5–13

ROI region of interest. 2

trigger candidate Summary information derived from the full data stream and representing a contribution toward forming a trigger decision. 15

trigger command Information derived from one or more trigger candidates and which directs elements of the detector module to read out a portion of the data stream. 14, 15

trigger decision The process by which trigger candidates are converted into trigger commands. 14, 15

Acronyms

APA anode plane assembly. 2, 10

DAQ data acquisition. 1, 3, 5, 8, 14

DP dual-phase. iii, 4, 5, 10, 15

EHN1 Experiment Hall North One. 11

EOS EOS. 11

FD far detector. 5

LAr liquid argon. 1

LBNF Long-Baseline Neutrino Facility. 2

SAM Sequential Access via Metadata. 6

SNB supernova neutrino burst. 4

SP single-phase. iii, 2, 4, 5, 10, 15

References

[1] DOE Office of High Energy Physics, “Mission Need Statement for a Long-Baseline Neutrino Experiment (LBNE),” tech. rep., DOE, 2009. LBNE-doc-6259.

[2] **WA105** Collaboration, S. Murphy, “Status of the WA105-3x1x1 m^3 dual phase prototype,” tech. rep., 2017. `https://indico.fnal.gov/event/12345/session/1/contribution/5/material/slides/0.pdf`.

[3] R. Brun and F. Rademakers, “Root - an object oriented data analysis framework,” *Nucl. Inst. And Meth. in Phys. Res. A* **389** (1997) .

[4] S. Agostinelli *et al.*, “Geant4 - a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250 – 303. `http://www.sciencedirect.com/science/article/pii/S0168900203013688`.

[5] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[6] C. Andreopoulos *et al.*, “The GENIE Neutrino Monte Carlo Generator,” *Nucl. Instrum. Meth.* **A614** (2010) 87–104, `arXiv:0905.2517 [hep-ph]`.

[7] T. Golan, J. T. Sobczyk, and J. Zmuda, “NuWro: the Wroclaw Monte Carlo Generator of Neutrino Interactions,” *Nucl. Phys. Proc. Suppl.* **229-232** (2012) 499.

[8] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” *CoRR* **abs/1603.04467** (2016) , `arXiv:1603.04467`. `http://arxiv.org/abs/1603.04467`.

[9] F. Chollet, “Keras.” `https://github.com/fchollet/keras`, 2015.

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM ’14, pp. 675–678. ACM, New York, NY, USA, 2014. `http://doi.acm.org/10.1145/2647868.2654889`.

[11] **MINOS Collaboration** Collaboration, D. Michael *et al.*, “The magnetized steel and scintillator calorimeters of the MINOS experiment,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **596** no. 2, (2008) 190 – 228. `http://www.sciencedirect.com/science/article/pii/S0168900208011613`.

[12] **MINERvA collaboration** Collaboration, B. Osmanov, “MINERvA Detector: Description and Performance.” arXiv:1109.2855 [hep-ex], 2011.

[13] R. Acciarri *et al.*, “Design and construction of the microboone detector,” *Journal of Instrumentation* **12** no. 02, (2017) P02017. `http://stacks.iop.org/1748-0221/12/i=02/a=P02017`.

[14] **NOvA** Collaboration, P. Adamson *et al.*, “First measurement of muon-neutrino disappearance in NOvA,” *Phys. Rev.* **D93** no. 5, (2016) 051104, `arXiv:1601.05037 [hep-ex]`.

[15] **MicroBooNE** Collaboration, R. Acciarri *et al.*, “Design and Construction of the MicroBooNE Detector,” *JINST* **12** no. 02, (2017) P02017, `arXiv:1612.05824 [physics.ins-det]`.

[16] **MicroBooNE** Collaboration, R. Acciarri *et al.*, “Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC,” *JINST* **12** no. 08, (2017) P08003, `arXiv:1705.07341 [physics.ins-det]`.

[17] **MicroBooNE** Collaboration, C. Adams *et al.*, “Ionization Electron Signal Processing in Single Phase LArTPCs I. Algorithm Description and Quantitative Evaluation with MicroBooNE Simulation,” `arXiv:1802.08709 [physics.ins-det]`.

[18] **MicroBooNE** Collaboration, R. Acciarri *et al.*, “Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber,” *JINST* **12** no. 03, (2017) P03011, `arXiv:1611.05531 [physics.ins-det]`.

[19] **MicroBooNE** Collaboration, R. Acciarri *et al.*, “The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector,” *Eur. Phys. J.* **C78** no. 1, (2018) 82, `arXiv:1708.03135 [hep-ex]`.

[20] **MicroBooNE** Collaboration, R. Acciarri *et al.*, “Michel Electron Reconstruction Using Cosmic-Ray Data from the MicroBooNE LArTPC,” *JINST* **12** no. 09, (2017) P09014, `arXiv:1704.02927 [physics.ins-det]`.