Snowmass2021 - Letter of Interest

New Computing Model for Experiments Utilizing Large Scale Liquid Argon TPCs

Topical Group(s):

☑ (CompF01) Experimental Algorithm Parallelization

☑ (CompF04) Storage and processing resource access

Contact Information:

Michael Kirby (Fermilab) [kirby@fnal.gov]

Authors: Michael Kirby, PhD (Fermilab), Kenneth Herner, PhD (Fermilab)

Abstract:

During the coming decade, several large scale liquid argon TPC based experiments will be commissioned and start taking data. From the experiences of ArgoNeuT, MicroBooNE, and ICARUS, LAr TPC detectors are capable of the precision necessary for the next generation of neutrino oscillation measurements. At the same time, it has become apparent that future LAr TPC experiments will have both considerable computing resource needs, and that the data volume and processing workflows are considerably different than collider-based experiment computing models. This letter lays out some of the unique computing challenges for large-scale LAr TPCs and areas of development that can help address those challenges.

Unique Challenges for LAr TPC Computing Models

The computing model for large-scale LAr TPCs must address three main challenges:

- data volume and I/O versus reconstruction processing time
- parallelization of signal processing for efficient use of accelerators
- rapid processing of supernova candidate events for multi-messenger observations

The data volume recorded from the single drift window of a 10 kT scale LAr TPC can easily reach 10 GB per trigger for typical detector configurations with annual data volume estimates reaching as high as 40 PB [1]. The processing time for a trigger based upon experiences at MicroBooNE and ProtoDUNE would suggest that the ratio of I/O to processing time for offline reconstruction would be 1.5 to 2 orders of magnitude greater than for collider experiments such as ATLAS and CMS. The need for access to low-level, pixelated data in training and inference of machine learning algorithms also complicates the data volume reduction within the experiment computing model. In order to address the possible mismatch in resources, development is needed for data management and workflow management infrastructure to ensure timely processing of data recorded.

One of the excellent opportunities for utilizing accelerators and leadership class facilities within the computing model of LAr TPCs is parallel processing of the raw signals from the induction and collection planes of the TPCs. With typical TPCs designed with anode planes with thousands of identical wires (either induction or collection), the processing of raw data signals is an ideal parallel computation. Of course, utilizing these facilities has significant hurdles with limited direct network access to cores, unique architectures, and provisioning that is optimized differently than the late-binding strategy of the OSG and WLCG. Significant effort is needed so that large-scale LAr TPC experiments can provision resources on Leadership Class Facilities (LCFs), co-locate data near resources, access conditions and calibration information needed for reconstruction, and have workloads that will efficiently run on novel architectures.

Supernova candidates pose a unique problem for data acquisition and reconstruction within LAr TPCs. A classic core-collapse supernova 10 kpc away would be expected to yield around 3,000 charged-current electron neutrino interactions in a 40 kT LAr TPC detector [2]. LAr TPC's fine-grained tracking should allow significant pointing power with the most optimistic scenario of four modules and high electron neutrino fraction yielding pointing resolutions of less than 5 degrees [3]. The ability to produce a reasonably fast pointing signal would be extremely valuable to optical astronomers doing followup, especially if a supernova was in a region where dust masks the primary optical signal. For example, a compressed supernova readout of a 40 kT LAr TPC will be of order 184 TB in size and take a minimum of 4 hrs to transfer over a 100 Gbps network, and then take of order 130,000 CPU-hrs for signal processing at present speeds. If processing takes the same time as transfer, a peak of 30,000 CPU cores would be needed to accomplish the task of 5° pointing accuracy within 8 total hours. The development of rapid computational resource provisioning software, and networking to support that provisioning, need significant effort but would enable important the observation of supernova core-collapse and have significant scientific impact.

In summary, there are significant development needs for the unique computing challenges of large-scale LAr TPCs that will need to be addressed in the coming decade. Without dedicated effort to ensure that these challenges are met, the processing and analysis of data from these detectors will be inefficient and overburden the current infrastructure.

References

[1] B. Abi *et al.* [DUNE Collaboration], "The DUNE Far Detector Technical Design Report Volume I: Introduction to DUNE", arXiv:2002.02967 [physics.ins-det] (2020).

[2] B. Abi *et al.* [DUNE Collaboration], "Supernova Burst Detection with the Deep Underground Neutrino Experiment", arXiv:2008.06647 [hep-ex] (2020).

[3] A.J. Roeth, "Supernova Pointing Resolution of DUNE", https://indico.cern.ch/event/868940/contributions/3813598/attachments/2081577/3496427/Point_Res_ICHEP_2020_07_AJRoeth.pdf