

Evolution of the T2K-ND280 Computing Model

Thomas Lindner for ND280 Computing Group

E-mail: lindner@triumf.ca

Abstract. ND280 is the off-axis near detector for the T2K neutrino experiment. ND280 is a sophisticated, multiple sub-system detector designed to characterize the T2K neutrino beam and measure neutrino cross-sections. We have developed a complicated system for processing and simulating the ND280 data, using computing resources from North America, Europe and Japan. The first key challenge has been dealing with very different computing infrastructure in different continents; a second challenge has been dealing with the relatively large 1.5PB ND280 MC dataset. We will describe the software, data storage and data distribution solutions developed to meet these challenges.

1. Introduction

Neutrino oscillations were firmly established in the late 1990s with the observation by the Super-Kamiokande (SK) experiment that muon neutrinos produced by cosmic ray interactions in our atmosphere changed their flavor [1]. Measurements from the Sudbury Neutrino Observatory a few years later revealed that neutrino oscillations were responsible for the apparent deficit of electron neutrinos produced in the Sun [2]. More recently the T2K experiment [3, 4] and reactor experiments [5, 6, 7, 8] have made definitive measurements of the neutrino mixing angle, θ_{13} .

T2K is a long baseline neutrino oscillation, located in Japan. The T2K experiment involves creating a beam of almost pure muon neutrinos at the JPARC laboratory in Tokai in eastern Japan. The neutrinos are then measured 295 km away at the Super-Kamiokande detector. By comparing the type, energy and flux of the measured neutrinos at Super-Kamiokande to expectations we can extract the fundamental parameters that define neutrino oscillations.

However, to accurately predict the neutrino flux at the far detector we need precise measurements of the neutrino flux before any oscillations have occurred. This is the goal of the ND280 detector. ND280 is composed of a complicated set of sub-detectors located within the re-purposed UA1 magnet. The active parts of ND280 consist of Time Projection Chambers (TPCs) readout with microMegas and more dense scintillating bar detectors, readout by Multi Pixel Photon Counters (MPPCs). Figure 1 shows an example neutrino interaction occurring in ND280.

ND280 has developed a distributed computing model that involves the processing of data and production of Monte Carlo throughout North America and Europe. This distributed computing model uses elements of computing models from other large HEP experiments, in particular the WLCG. Key challenges of the ND280 computing model have been the heterogeneous nature of the computing resources available in different regions and the large data sets. In this note we will describe in details the ND280 software and computing models.



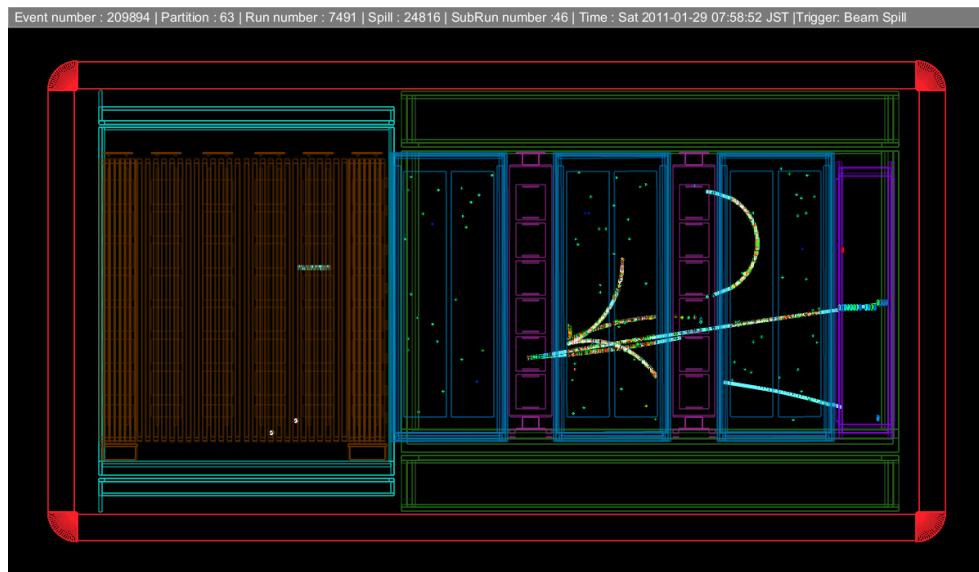


Figure 1. An example neutrino interaction occurring in one of the ND280 Fine Grained Detectors (FGDs); the charged particles from this interaction are imaged by the surrounding TPCs.

2. ND280 software

ND280 uses a fairly standard HEP software suite. The ND280 offline software suite consists of about 60 software packages totalling over 600,000 lines of original code. The design principles that have been followed include an emphasis on the use of standard particle physics software libraries and conventions where possible, structural modularity, good documentation and openness. Hence ROOT [9] is used for the underlying framework and data storage model, with Geant4 [10] as the basic simulation library. The CMT [11] software configuration management system provides the packaging mechanism, with CVS [12] used to store all versions of the source code files in a repository.

A representation of the geometry of the detectors is constructed in Geant4 code, and is converted to ROOT TGeoManager format and stored in version-controlled files. These are retrieved from a central repository to be used in the interpretation of raw data.

For Monte Carlo simulations, interfaces have been built between the neutrino beam simulation, the neutrino interaction generation packages, GENIE [13] and NEUT [14], and the ND280 software. The neutrino fluxes estimated from beam MC are passed through the detector geometries, and neutrino cross sections specific to the nuclei present in the geometries are used to generate interactions that are appropriate for the distribution of materials in the detector.

Geant4 is used to simulate the energy deposits from the final state particles that pass through the detector, and the response of the active detectors (scintillator bars, fibers, MPPCs and electronics, and TPC electron drift and electronics) is simulated through custom-written code in the elecSim package.

Individual subdetectors have dedicated packages designed to reconstruct event information internal to them. The Rec-Pack toolkit [15] is used as the framework for event reconstruction across the off-axis detector. It is an independent software package, but has been developed in close conjunction with the ND280 software to meet its needs. The full event information contained in the oaEvent format files is distilled by the oaAnalysis package into files based on trees which are built up from pure ROOT objects. An accompanying library of analysis tools helps end-users to process the summarized output using standardized C++ routines and

Python macros. An overall software control package allows for the fully automated running of the software, based on simple configuration files which list the inputs and processing steps.



Figure 2. Diagram of the geographic distribution of the ND280 data distribution and computing model.

3. ND280 MC Production and Data Processing

We organize our CPU efforts into large-scale productions, where a particular production uses a consistent set of software. We typically have one major production per year, since the calibration/reconstructions improvements typically take a while to develop.

There are a number of different types of ND280 production jobs, each with their different challenges:

- Data processing: the calibration and reconstruction of our real spill and calibration events.
- Monte Carlo (MC) spill simulation and reconstruction: these are simulations of neutrino interactions in the magnet. These are the principal simulation against which our neutrino data is compared. We typically aim to produce 10-30 times more MC statistics than data statistics. Consequently the MC spill simulations constitute our largest CPU and storage load.
- Monte Carlo simulations of backgrounds and calibration events: in particular, we have separate simulations for neutrino interactions that happen in the surrounding rocks and simulations of cosmic-ray events.

The 2014-2015 T2K-ND280 analyses used results from Production-6. In what follows we will describe the different modes of production work in Europe vs North America. Resources are provided in quite different ways in the two regions, significantly affecting the evolution of the ND280 computing model. The distribution of production work and data storage is shown graphically in Figure 2.

3.1. European Grid Processing

In Europe we use resources provided through LHC Computing Grid (LCG), largely provided by GridPP in the UK (with small amounts of additional computing from other countries like Spain). Our typical usage is as follows:

- We use 7 computing sites through the UK. The largest amount of CPU resource is provided by RAL-LCG2, which provides 50% of used CPU.

- Jobs are submitted through Workload Management Servers (WMS) [16] at RAL and Imperial College.
- The output from each job is copied to the local Storage Element (SE). Synchronization to the main RAL-LCG2 SE happens later, as explained in Section 4.2.

For the last production the European grid sites performed the spill MC simulation, which was the largest part of the CPU usage.

3.2. North American Batch Processing

The production work in North America uses three sites:

- Data processing and cosmic-ray MC on the Bugaboo cluster in Vancouver.
- Sand neutrino background simulation on the SCINET cluster in Toronto.
- GENIE spill MC on the CMS Tier-3 cluster in Colorado.

For these clusters we use traditional torque job submission. The output from these jobs is transferred back to the TRIUMF storage element.

3.3. Production Challenges

The ND280 production work has proven to be relatively manpower intensive, mostly because of the different resources provided in each region. Each different resource has required different scripts and expertise, leading to local experts that can only help with parts of the production work. In addition, there is no unified global view of the status of productions. We are unfortunately too small of a VO to develop our own software tools (which is what larger VOs do), but too large to not use a distributed computing model.

In addition, we have been hampered by our reliance on older, less supported tools. This is particularly the case for our European grid processing, where we use the LFC and WMS services which have been abandoned by the large LHC VOs.

Ongoing efforts to address these challenges will be presented in Section 6.

4. Data Management and Distribution

The data set for ND280 is a relatively large ≈ 1.5 PB of data. The majority of this data is the full set of reconstructed tracks and truth information from the MC spill simulation. Keeping this full set of information is important for our ability to efficiently re-process some portion of the data set. However, most end users only use the smaller set of analysis files.

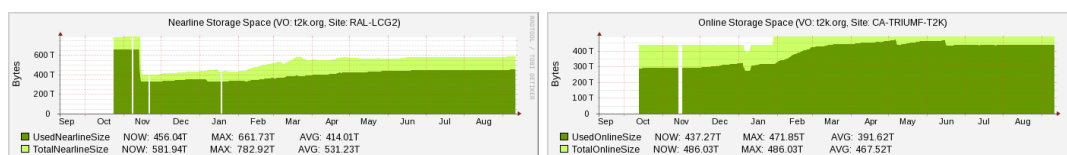


Figure 3. Evolution of the storage usage over time at main storage elements.

4.1. Storage Elements

The core of the data distribution system is a pair of large data storage; they are

- RAL-LCG2: a 700TB allocation at RAL GridPP-supported Tier-1.
- CA-TRIUMF-T2K: a 700TB storage element at TRIUMF, provided and maintained solely by T2K-Canada group.

All the output from the production work get saved at one or the other of these storage elements; all the files are then synchronized between the two, so that there is a copy of all the critical data in both continents. These main SEs provide the principal 'interface' between LCG-Europe and non-LCG-America; ie they are the unifying link in our heterogeneous computing system. Figure 3 shows the evolution of the SE usage over a production.

4.2. Management and Distribution

We use the LFC [17] providing a global view of the ND280 data. We augment this with a custom MySQL database which we use to store more detailed information about how each was produced.

Distribution of the files between the storage elements is mostly done using FTS3 [18]. There have been in the past some problems where FTS transfers occur, but the LFC registration fails. This can cause 'dark data'; files which sit on a storage element but aren't listed in the LFC, which therefore continue to waste storage space.

5. MySQL Databases

ND280 uses a set of MySQL databases for following purposes:

- ND280 Slow control information (temperatures, voltages, etc). Total size is 90GB.
- Calibration constants and data quality status. Total size is 7GB.
- File metadata catalogue, described in Section 4.2. Total size is 4GB.

Each databases has a different schema and a different set of C++ interface classes. In addition, the slow control and file catalogue databases have web interfaces. The slow control web interface is provided through the MIDAS DAQ framework [19]; this web interface has been critical for off-site experts to be able to quickly check the detector status.

The main databases are hosted at JPARC and TRIUMF, with some replication to Europe. Databases replication and loads are monitored with Zabbix, which provides clean web interfaces for checking database status.

6. Ganga and DIRAC

We have been working on integrating into our computing model some more modern frameworks. The main goal of this work has been to reduce the amount of time and experience needed by production personnel, by having a single interface for users to learn. Currently production work is a substantial burden for personnel.

The first of these new tools is Ganga [20]. Ganga is a python tool for managing job requirements, originally developed for ATLAS and LHCb. One benefit of Ganga is that it allows simpler transitions between test environments, like your desktop, and larger scale batch resources (by abstracting away the details that relate to where the job is being run).

We have started using Ganga to unify various types of ND280 computing task. This involved writing several Ganga plugins to wrap ND280 software and data sets. We have already transitioned over our main data processing and control sample creation to Ganga. The next step is to start switching over MC production tasks.

The second new tool is DIRAC [21]. There has been a long-standing plan to decommission the original glite WMS that we use in Europe to do job submission. In order to future-proof ourselves, we have been working with GridPP colleagues on testing DIRAC (Distributed Infrastructure with Remote Agent Control) as a replacement workload management system. DIRAC also offers a system for submitting jobs to different systems, as well as a way of deciding between different resources. DIRAC uses the more modern system of pilot jobs to more efficiently distribute work across different computing resources.

There have been some trouble with the DIRAC handling of LFC naming of existing T2K data-sets, but we have successfully submitted some jobs (see Figure 4). We are working towards larger scale tests.

JobID	Status	MinorStatus	ApplicationStatus	Site	JobName	LastUpdate [UTC]	LastSignOffLife	SubmissionTime	Owner
1742	Failed	Failed to perfor...	Unknown	ANY	Unknown	2014-10-02 16:55	2014-10-02 16:55	2014-10-02 16:55	j.perkin
1741	Failed	Input data not c...	Unknown	ANY	Unknown	2014-10-02 16:54	2014-10-02 16:54	2014-10-02 16:54	j.perkin
1740	Failed	Application Fini...	Unknown	LCG.UKI-LT2-Q...	Unknown	2014-10-02 17:00	2014-10-02 17:00	2014-10-02 16:54	j.perkin
1739	Failed	Application Fini...	Unknown	LCG.UKI-LT2-Q...	Unknown	2014-10-02 16:34	2014-10-02 16:34	2014-10-02 16:31	j.perkin
1738	Failed	Input data not c...	Unknown	ANY	Unknown	2014-10-02 16:29	2014-10-02 16:29	2014-10-02 16:29	j.perkin
1737	Failed	Input data not c...	Unknown	ANY	Unknown	2014-10-02 16:28	2014-10-02 16:28	2014-10-02 16:28	j.perkin
1736	Failed	Application Fini...	Unknown	LCG.Sheffield.uk	Unknown	2014-10-02 16:20	2014-10-02 16:20	2014-10-02 16:16	j.perkin
1735	Done	Execution Com...	Unknown	LCG.UKI-LT2-Q...	Simple_job	2014-10-02 14:55	2014-10-02 14:55	2014-10-02 13:49	j.perkin

Figure 4. DIRAC test instance at Imperial

7. Conclusion

The ND280 computing group has developed a distributed computing system to handle the data processing and simulation requirements of the experiment. One challenge is the rather different computing frameworks and support in different regions; we have partly solved this by having major data centers in North America and Europe, allowing different data access methods from each. Similarly, we have developed different set of productions scripts in the two continents. Work is ongoing with DIRAC and Ganga to try to unify these different script bases.

References

- [1] Y. Ashie et al. (Super-Kamiokande Collaboration), “Measurement of atmospheric neutrino oscillation parameters by Super-Kamiokande I”, Phys. Rev. D 71, 112005 (2005).
- [2] Q. R. Ahmad et al. (SNO Collaboration), “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”, Phys. Rev. Lett. 89, 011301 (2002).
- [3] K. Abe et al. (T2K Collaboration), “Evidence of Electron Neutrino Appearance in a Muon Neutrino Beam”, Phys. Rev. D 88, 032002 (2013), arXiv:1304.0841 [hep-ex].
- [4] K. Abe et al. (T2K Collaboration), “Observation of Electron Neutrino Appearance in a Muon Neutrino Beam”, Phys. Rev. Lett. 112, 061802 (2014), arXiv:1311.4750 [hep-ex].
- [5] F. P. An et al. (Daya Bay Collaboration), “Observation of Electron-Antineutrino Disappearance at Daya Bay”, Phys. Rev. Lett. 108, 171803 (2012), arXiv:1203.1669 [hep-ex].
- [6] F. P. An et al. (Daya Bay Collaboration), “Spectral Measurement of Electron Antineutrino Oscillation Amplitude and Frequency at Daya Bay”, Phys. Rev. Lett. 112, 061801 (2014), arXiv:1310.6732 [hep-ex].
- [7] J. K. Ahn et al. (RENO collaboration), “Observation of Reactor Electron Antineutrinos Disappearance in the RENO Experiment”, Phys. Rev. Lett. 108, 191802 (2012), arXiv:1204.0626 [hep-ex].
- [8] Y. Abe et al. (Double Chooz Collaboration), “Indication of Reactor e Disappearance in the Double Chooz Experiment”, Phys. Rev. Lett. 108, 131801 (2012).
- [9] R. Brun, F. Rademakers, “ROOT: An object oriented data analysis framework”, Nucl. Instrum. Meth. A389 (1997) 8186.

- [10] S. Agostinelli, et al., “Geant4a simulation toolkit”, Nucl. Instrum. Meth. A506 (2003) 250303.
- [11] C. Arnault (2000). “CMT: A software configuration management tool”, Prepared for International Conference on Computing in High-Energy Physics and Nuclear Physics (CHEP 2000), Padova, Italy, 7-11 Feb 2000.
- [12] B. Berliner, J. Polk, Concurrent Versions Systems (CVS), 2001. <http://www.cvshome.org>.
- [13] C. Andreopoulos, et al., “The GENIE neutrino Monte Carlo generator”, Nucl. Instrum. Meth. A614 (2010) 87104.
- [14] Y. Hayato, “A neutrino interaction simulation program library NEUT”, Acta Phys. Polon. B40 (2009) 24772489.
- [15] A. Cervera-Villanueva, J. J. Gomez-Cadenas, J. A. Hernando, “RecPack’ a reconstruction toolkit”, Nucl. Instrum. Meth. A534 (2004) 180183.
- [16] F Pacini, “EGEE Users Guide, WMS Service”, DATAMAT, 2005
- [17] EGEE, LCG File Catalogue, <https://svnweb.cern.ch/trac/lcgdm/wiki/Lfc>.
- [18] A A Ayllon et al., “FTS3: New Data Movement Service For WLCG”, 2014 J. Phys.: Conf. Ser. 513 032081
- [19] S. Ritt, P. Amaudruz, K. Olchanski, “MIDAS (Maximum Integration Data Acquisition System)”, <http://midas.triumf.ca>.
- [20] J.T.Moscicki et al., “Ganga: a tool for computational-task management and easy access to Grid resources; Computer Physics Communications”, Volume 180, Issue 11, (2009)
- [21] A Tsaregorodtsev et al., “DIRAC: a community grid solution”, 2008 J. Phys.: Conf. Ser. 119 062048