

Software Projects in Double Chooz

I made a large impact in almost all software aspects in Double Chooz. Here, I list only those where I lead the development effort from the start excluding others where I made an external contribution outside the core development team.

FileDB ... is a file tracking MySQL database infrastructure with user interface written in C++ and Python. It was used for all file access purposes including analysis, data production, and inter-cluster file transfer protocols.

DCProd ... is a data production framework written in Python with necessary MySQL database infrastructure for tracking grid job status. It was also integrated with FILEDB to track both input and output files via database. It supports a web browser interface to both monitor and operate production process.

OnlineDB ... is the base database infrastructure used for storing information related to data taking. APIs support both C++ and Python, and used by the run control GUI for recording run number and configuration, operation shifters to record comments and findings, and also by **DataMigro** which is a file transfer software to send data from the detector to the permanent storage in the off-site computing cluster.

DataMonitor ... is an automated data quality control software. It runs inspection algorithms developed by system experts to identify potential problems in data and it is operated online.

CommonTrunk ... is an integration software that runs both calibration and reconstruction algorithms to process both simulated and real detector data in the production chain.

Software Projects in MicroBooNE

My software development contributions to the MicroBooNE experiment have been instrumental, and it is widely spread from detector operations, calibrations, to event reconstruction and analysis. Here, I list software efforts which I founded and led the development group work except for DAQ for which I was involved as a major contributor.

DAQ ... Implemented interface APIs for communicating with readout electronics hardware, threaded trigger data readout stream, and electronics status readout which is monitored by experiment shifters via SLOWCONTROL monitoring software.

RunConfigDB ... SQL-based (POSTGRES) framework equipped with C++ and Python API to store and track data taking configuration for all systems, including both software and hardware, for all data taking runs in the history. It supports graph-like parameter set data structure. It optionally supports FHICL file format as an output (used by DAQ).

PUBS ... Python based data processing framework. It is a database state-machine that can be implemented with either POSTGRES or MYSQL back-end for executing tree-structured process units. For online data processing it is responsible for registering DAQ output file automatically into the database, decode binary file header to extract file meta-data, transfer data files to permanent storage, cleaning of data files and monitoring of online computing resources. Equipped with a robust failure handling scheme it has processed over 1 million files without any unanticipated mistake. It is also used for offline simulation and reconstruction data production for handling number of staged grid job submissions.

SWTrigger ... C++ based software framework for implementing software trigger logic. It provides base structures on which physicists can implement pure algorithm logic C++ code. It is currently used to apply trigger at DAQ level which reduces the data volume by a factor of 20 with negligible loss in signal, which is necessary for sustainable data taking.

LArLite ... C++ code development framework. It provides a set of scripts and build procedures to make C++ learning experience easy for beginners. The build supports Python CAPI binding by default which allows users to use compiled C++ from Python automatically. This tool is used by many students and post-docs to generate numbers of portable C++ software frameworks for data reconstruction and analysis purpose, and therefore it is the most productive invention of mine to date.

LArLite Analysis ... C++ framework that mimicked LARSOFT for data product representation and IO stream handling. Unlike LARSOFT, it supported wide variation of Linux and OS X distributions from the beginning. Setting up and compilation is much easier and faster than LARSOFT, and attracted number of analyzers. The process execution speed is also a lot faster than LARSOFT as it is a stripped-off version with simpler architecture, and it contributed to find problems in LARSOFT execution speed in early

days through a speed comparison report brought by myself. This resulted in immediate speed up in LARSOFT by a factor of roughly 3 orders of magnitude after fixing a problem which was executing unnecessary calculation unintentionally by almost all users in MicroBooNE.

OpticalDetectorSim ··· A software package that simulates the logic in the PMT read-out electronics as well as hardware PMT trigger logic. It is now a part of the experiment's software repository.

OpHitReco ··· A software framework for reconstructing optical signal in individual PMT waveform. It supports modular algorithm implementation. Algorithms are developed across several collaborators as well as some contributors from other LArTPC experiments. This is integrated into LARSOFT and used for outside MicroBooNE as well.

UBFlashFinder ··· A software framework like OPHITRECO but this operates algorithms that reconstruct an optical *flash*, a cluster reconstructed optical hit across PMTs and time. While a module to execute single hard-coded algorithm is implemented in LARSoft, we implemented a framework in the MicroBooNE code repository to allow flexible development by others. This is used for the default flash reconstruction in MicroBooNE.

OpT0Finder ··· A software framework equipped with numbers of algorithms to find the right pair of optical flash and reconstructed TPC interaction which is a collection of 3D geometrical objects (track and showers) . The latter could be anything but a typical case uses a neutrino interaction. This reconstruction step helps to reduce cosmic background as well as providing T0 to TPC reconstructed interactions.

ShowerReco3D ··· A software framework to reconstruct 3D shower objects given a cluster of 2D hits as inputs. It supports modular algorithm design and attracted dozens of physicists to co-develop. This is the default shower reconstruction routine in MicroBooNE to date.

MichelReco ··· A software framework to support 2D Michel electron reconstruction initiated by a summer student at Columbia. The study was successful and made a dramatic progress since then. Currently it serves for the first publishable physics analysis in the experiment, which is in the internal publication review.

GeoAlgo ··· A collection of 3D and 2D geometrical analysis algorithms, developed based on a fast collision detection analysis textbook used for computer game code development. This is widely used in physics analysis to assess correlation of 3D geometrical objects such as tracks and showers.

ERTool ··· A software framework for reconstructing particle correlations and stores in a graph data structure. It was used for the latest sensitivity study in the experiment where many algorithms were developed under the leadership of the oscillation analysis group.

It heavily depends on GEOALGO package for fast assessment of object correlation in 3D space.

LArOpenCV . . . A reconstruction software framework to process LArTPC data as an image format and support easy interfacing with open source image processing library called OPENCV. This was the first attempt of using OPENCV for event reconstruction and analysis in LArTPC community and attracted numbers of contributors to develop algorithms. This is currently used for 2D clustering stage of electromagnetic showers, and in particular π^0 mass peak analysis from real detector data. It is also under heady development for the on-going low energy ν_e oscillation signal reconstruction by the deep learning group.

LArCV . . . An analysis framework developed for processing LArTPC image data. Unlike LAROPENCV, this software is not reconstruction specific but for more generic use by supporting C++ data structures, IO interface, and data processing machinery. It is used to directly manipulate the image data with or without OPENCV, and also it plays a key role to interface with open source deep learning softwares including CAFFE by Berkeley Lab. and TENSORFLOW by Google.