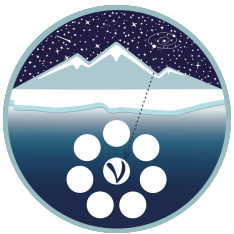
Review of software for automatic processing and analysis of data from the Baikal-GVD neutrino observatory

A.G. Solovjev
on behalf of the Baikal-GVD Collaboration

Joint Institute for Nuclear Research, Dubna, Russia, 141980

Parallel computational technologies, April 2-4, 2024



Outline

Introduction
- Method for neutrino detection
- Baikal-GVD
- Computing system
- Data to be processed

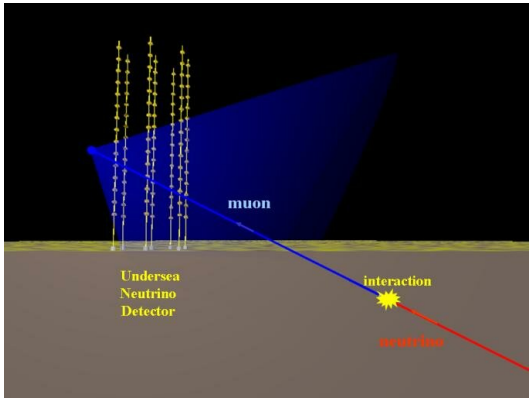
Core of processing software
- Overview of programs in use

Workflow management system
- Luigi
- Per-file workflow
- Per-run workflow
- Multicluster workflow

Conclusion

Introduction

Method for neutrino detection



M. A. Markov, 1960

recording Cherenkov radiation from secondary muons and/or high-energy showers produced by neutrinos during their interactions with matter in transparent natural media

$$\nu_\mu + n \rightarrow p + \mu^-$$
$$\nu_\mu + p \rightarrow n + \mu^+$$

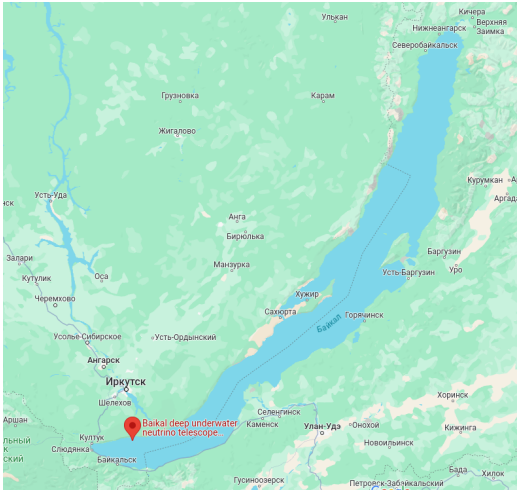
Introduction

Baikal-GVD (Gigaton Volume Detector)

Location Lake Baikal, ≈4 km away from shore

Advantages

- the depth (1366 m)
- flat lakebed
- water transparency
- railway infrastructure
- possibility of deployment right away from the ice during late winter and early spring



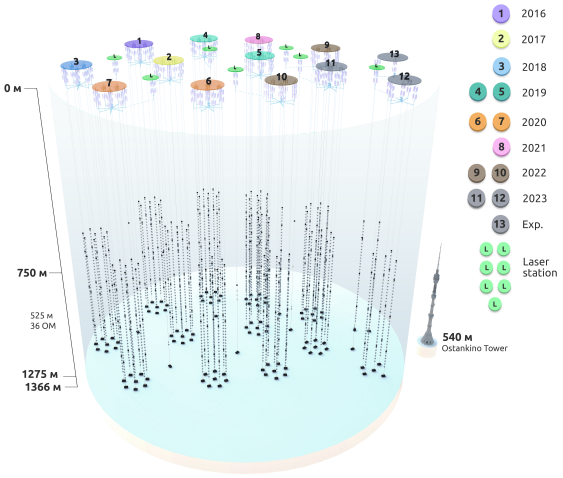
Introduction

Baikal-GVD (Gigaton Volume Detector)

Basic elements are optical modules (OMs).

The OMs are mounted to the strings, with 36 OMs per string:

- ≈15 m spacing
- depth 750 — 1275 m



The strings are grouped in clusters, with 8 strings per cluster:

- diameter ≈120 m
- 250-300 m from each other
- 13 clusters
- effective volume $> 0.5 km^3$

Each cluster functions as a standalone detector.

Computing system

The virtual machines (VMs) have been deployed in the JINR cloud infrastructure.

- 28-32 CPU cores
- 235-480 GB of memory

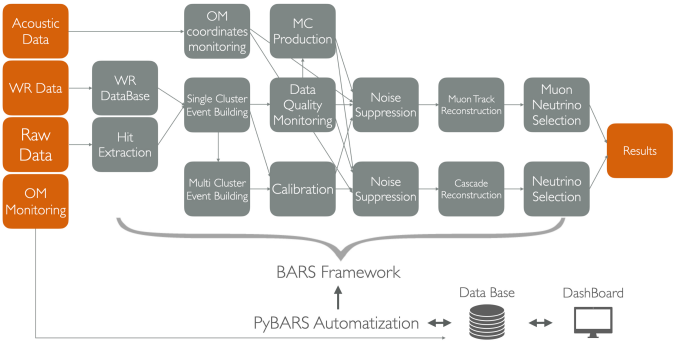
The number of VMs = the number of clusters.

Each VM has its own copy of the processing software.

Introduction

Data to be processed

The data processing chain:



A cluster produces the raw data for typical 24 hours run. The number of files per run ≈100.

Core of processing software

A set of C++ programs based on

BARS framework

(Baikal Analysis and Reconstruction Software)

Provides the facilities to build processing pipelines

forms a directed acyclic graph, where the output of one program is the input to others.

An unified command line interface with the following common arguments:

- `--season` — sets season (4 digits, like 2019)
- `--cluster` — sets cluster number (from 1)
- `--run` — sets run number (from 1)
- `--file` — sets file number in a run (optional)
- `--promise` — prints list of the output files

Specifying these arguments uniquely determines the input and the output for each program.

Core of processing software

Overview of programs in use

md5sum	Each raw file is first checked for integrity
readconfig	For the first run file, the detector configuration is read
stage1	Reads the specified raw file, extracts hit parameters (time and charge) from waveforms or evaluates local trigger condition
read-wr-data	Prepares WR timestamps from the White-Rabbit database
getjointtable	Combines section data into cluster-wide events
eventbuilder	Combines scattered information into single cluster-wide events
globaltrigger	Combines local trigger information into single cluster trigger
tsa	Analyzes global tigger data
bexport-exp	Applies calibrations to cluster-wide events
rates	The program evaluates the rate of counting channel noise
dqm	Monitors the status of the detector and collected data
reco-muon	Reconstructs muon track events
reco-cascade	Reconstructs cascade events

Workflow management system

PyBARS - a set of Python scripts based on

Luigi package

Developed by Spotify

`https://github.com/spotify/luigi`

arranges the programs into a processing graph and ensures their execution in the pre-defined order, taking into account the dependencies between them.

Workflow management system

Luigi

Fundamental building blocks of Luigi:

Task

Tasks are where the execution takes place
Each C++ processing program has its own Luigi wrapper as a Task
Tasks depend on each other and on their own output targets

Target

The Target class corresponds to a file on a disk
The Task is considered as done if its output files exist
(checkup uses the `--promise` argument of underlying program)

Parameter

The task parameters depend on the data being processed
(passed to the processing program as arguments `--season`, `--cluster`, `--run`, `--file`)

Workflow management system

Two processing systems

Multi messenger astronomy

GW170817

To make the Baikal-GVD a contributor, software must provide:

- obtaining the high-level event data necessary for physics analyses
- issuing alerts

Fast Processing

operates on raw files as they arrive and does not include information about data quality. It reconstructs event coordinates with a lower precision. However, it allows for a reduction in the time delay to 3–5 min between detecting a potential neutrino event and generating a preliminary notice about the candidate.

Offline processing

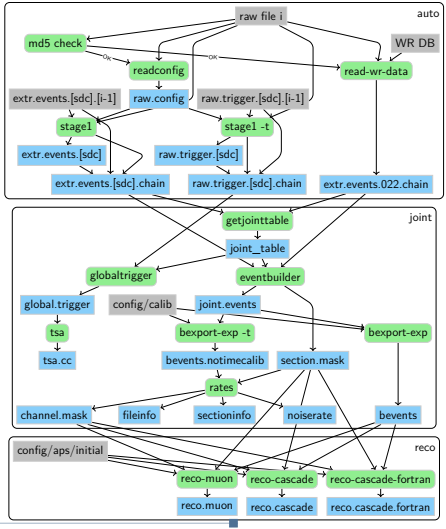
operates on completed data taking runs. It uses the data quality information and provides more accurate events reconstruction than the Fast Processing. However it requires more time - 24 hours for the run to complete and 2–4 hours for processing.

Workflow management system

Per-file workflow

Stacked-Parallel form

Independent operations are located on the same level (tier)



The daemon monitors the appearance of the data file. It arrives every 3-10 minutes.

When a new file arrives, a new instance of the graph is created. The processing takes 2-3 minutes.

The first stage is necessary for further per-run processing. The next two are preliminary stages.

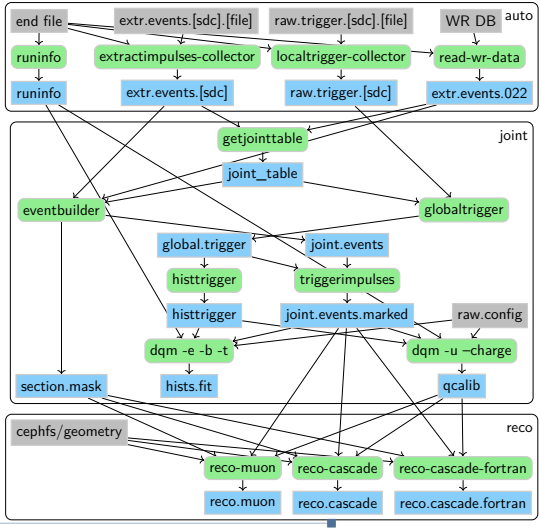
Workflow management system

Per-run workflow

The daemon monitors the appearance of the 'end' file.

When all run files have been processed as they arrive, the processing takes about an hour.

When it is necessary to perform per-file processing, processing the run takes up to 5 hours.



Workflow management system

Multicluster workflow

High energetic muon track can trigger more than one cluster.
Such multicluster (MCL) events are of particular interest for physics analyses.

There is a separate machine to which others send certain data from one cluster.

Finding time periods where all clusters are active:

Once MCL events are found, they are reconstructed.
Then the PyBARS package broadcast a message to email and telegram channel.

Conclusion

The Baikal-GVD data processing system has been developed.

Its advantages are simplicity, modularity and parallelism.

The modular architecture of the system makes it easy to modify or add components without violating the integrity.

The parallelism of the system consists of several levels.

1. Processing of individual clusters runs in parallel on different virtual machines. Another virtual machine is then used to combine single-cluster events into telescope-wide events.
2. Single-cluster processing is performed in two sequential workflows:
   - *Fast Processing*: each file is processed as it arrives. A potential neutrino event can be detected reconstructed and analyzed within a 3–5 min delay from the event detection.
   - *Offline processing*: the entire run is processed as it finishes. The results of per-file workflow are used. The comprehensive information about the event is produced. This yields more accurate results than the Fast Processing, albeit after a more substantial delay (2–4 hours after the run completion).

   In both workflows, some tasks are executed in parallel.

Conclusion

The current state of the processing system has already provided astrophysical results:

- Diffuse neutrino flux measurements with the Baikal-GVD neutrino telescope.
  Phys. Rev. D, 107(4):042005, 2023.
- High-energy neutrino-induced cascade from the direction of the flaring radio blazar TXS 0506 + 056 observed by Baikal-GVD in 2021.
  Monthly Notices of the Royal Astronomical Society, 527(3):8784–8792, 11 2023.
- Search for directional associations between baikal gigaton volume detector neutrino-induced cascades and high-energy astrophysical sources.
  Monthly Notices of the Royal Astronomical Society, 526(1):942–951, 09 2023.

The fast processing subsystem allows quickly respond to astrophysical events:

- Baikal-gvd observation of a high-energy neutrino candidate event from the blazar pks 0735+17 at the day of the icecube-211208a neutrino alert from the same direction.
  https://www.astronomerstelegram.org/?read=15112.