Automatic data processing for Baikal-GVD neutrino observatory

Bair Shaybonov on behalf of the Baikal-GVD collaboration

JOINT INSTITUTE FOR NUCLEAR RESEARCH

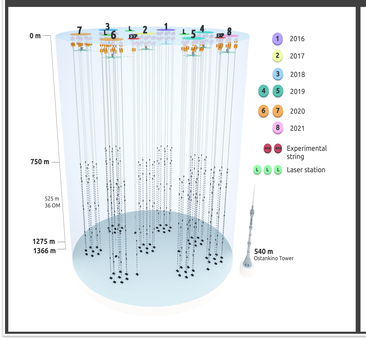
ONLINE ICRC 2021 THE ASTROPARTICLE PHYSICS CONFERENCE Berlin | Germany 37th International Cosmic Ray Conference 12–23 July 2021

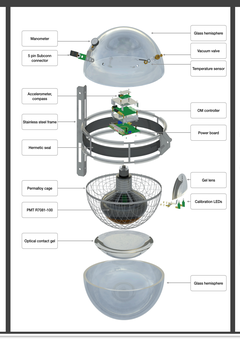
Outline

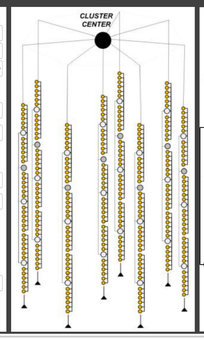
- Baikal-GVD (Gigaton Volume Detector) aims to investigate mainly the Galactic and extragalactic neutrino point sources in TeV – PeV energy range.
- The detector is a km³ - scale 3D-array of photo sensors located deep underwater and is currently under construction
- Flexible structure allows to rearrange the main building blocks , e.g. to optimize the neutrino detection energy threshold

Baikal-GVD Phase I Array:

- 2304 Optical Modules (OM) in GVD Phase I. OMs arranged in 8 Clusters with 8 Strings each
- Depth is 750 – 1275 m (91 m above the lakebed). 36 OMs at each string
- 15 m between OMs in a string
- 200-300 m between Clusters
- Each string consists of 3 sections - the main element of DAQ



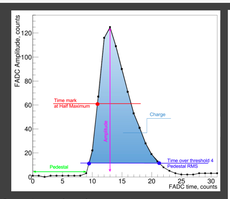


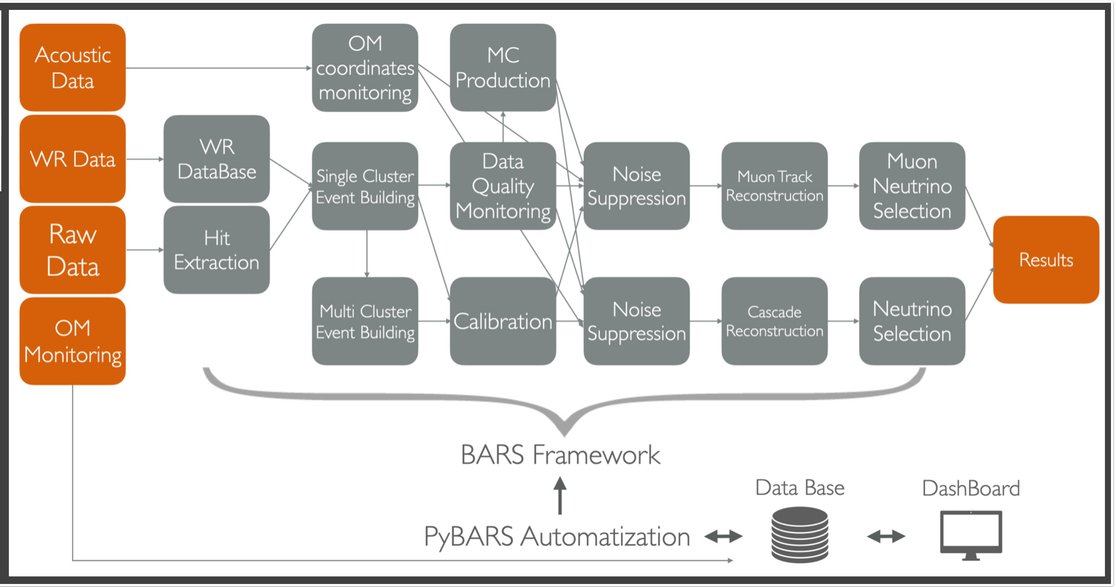


Data Processing Chain

Main parts:

- Extraction of PMT waveforms characteristics
  -> $T_i, A_i, Q_i, i = 1,..,N_{hit}^{section}$
- White-Rabbit synchronization data preparation
  -> + T_{WR}
- Single Cluster event building: combine hits from all CMs and timestamps from White-Rabbit and MSU Synchronisation systems
  -> $T_i, A_i, Q_i, i = 1,..,N_{hit}^{cluster} + header : T_{WR}, T_{sync}$
- Multi Cluster event building -> $T_i, A_i, Q_i, i = 1,..,N_{hit}^{telescope} + header : T_{WR}, T_{sync}$
- OM coordinates monitoring -> + $\vec{R}_i^{OM}(t)$
- Data quality monitoring:
  - validates the collected data, verifies the detector status
  - amplitude calibration constants
  - current lake noise level
  - trigger data quality monitoring
- Run-by-run MC simulation: $T_i^{MC}, A_i^{MC}, Q_i^{MC}, i = 1,..,N_{hit}^{cluster}$
- Time calibration applying $T_{shift}^i, i = 1,..,N_{hit}^{cluster}$
- Fast muon track reconstruction algorithm + specific noise hit suppression
- 2 versions of cascade reconstruction algorithms + specific noise hit suppression
- Quality upgoing reconstructed events selection -> Generation of alert



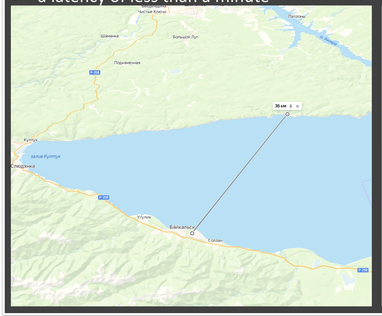


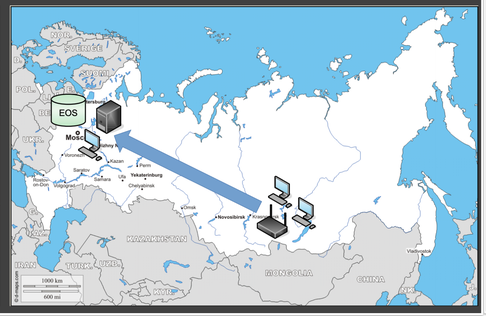
Raw Data:

- PMT digitised waveforms from Central Modules
- Acoustic positioning system data from acoustic modules,
- OM monitoring data from OM controllers,
- White-Rabbit synchronisation data

Raw data transfer:

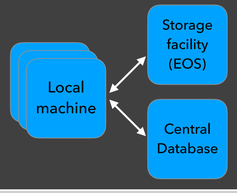
- transmission through the 300 Mbit/s radio link over the lake to Baikalsk town and then through the Internet
- storing in the central storage and processing facility in Joint Institute for Nuclear Research (JINR) in Dubna, Russia
- a latency of less than a minute

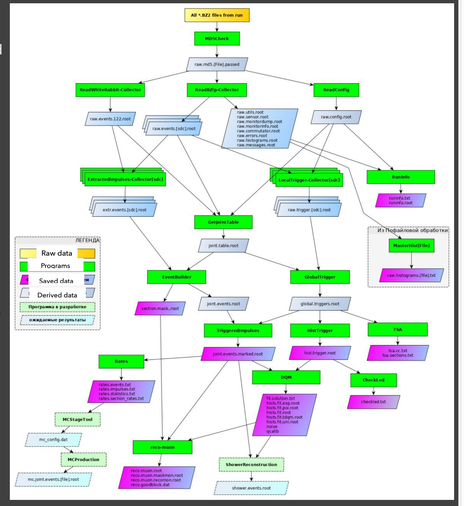




Workflow Management (PyBARS automatization):

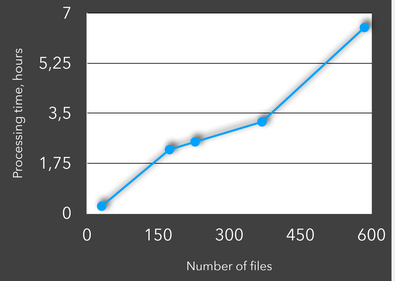
- Directed acyclic graph of data files and programs started from the raw data (yellow block)
- The system is started when a new run record has appeared in the Run Queue DB table which means that all run's raw files have come to the storage facility
- All raw data are copied from the storage facility to a local machine
- The programs (green blocks) take input data (blue or magenta or yellow blocks) and produce output data (blue or magenta blocks) one after another
- Some programs run in parallel in event based or over CMs manner
- If all programs are finished successfully selected data files (magenta blocks) are copied to the storage facility
- Processed data are stored with a specific version corresponding to the software version (git tag)
- Program execution status and run metadata are saved into the central database and displayed on the dashboard





Data Processing and Monitoring

- Raw data 2019 - 2020 were processed by the system in about 5 weeks (~200 CPU cores)
- Incoming data 2021 are being processed automatically
- It takes 2-7 hours to process a 24 hour cluster run depending on lake noise level



Summary

- Long-term R&D of automatic data processing system was finished by the first stable software version
- Collected data from 2019 - 2020 have been processed
- Incoming data 2021 are being processed automatically with a delay of several hours
- To generate alert signals with a delay of several tens of minutes we are going to move to per-file data processing

Cluster	Run	# files	Period, hour	# events, mln	Processing time, hour
6	316	**31**	3.3	0.76	**0.25**
5	281	**174**	23.8	5.4	**2.23**
6	313	**228**	22.3	5.0	**2.50**
2	286	**369**	17.5	7.0	**3.20**
2	297	**585**	24.2	11.0	**6.50**

References:

1. V.A. Allakhverdyan et al. (Baikal-GVD collaboration), Neutrino Telescope in Lake Baikal: Present and Nearest Future, these proceedings
2. V.A. Allakhverdyan et al. (Baikal-GVD collaboration), Monitoring of optical properties of deep lake water, these proceedings
3. A.D. Avrorin et al. (Baikal-GVD collaboration), The optical noise monitoring systems of the Lake Baikal environment for the Baikal-GVD telescope PoS-ICRC2019-875, doi:10.22323/1.358.0875
5. V.A. Allakhverdyan et al. (Baikal-GVD collaboration), The Baikal-GVD neutrino telescope: search for high-energy cascades, these proceedings
6. V.M. Aynutdinov et al. (Baikal Collaboration), The data acquisition system for Baikal-GVD, EPJ Web of Conferences, 116 (2016) 5004.
7. A.D. Avrorin et al. (Baikal-GVD collaboration), Data management and processing system for the Baikal-GVD telescope, PoS-ICRC2017-1046, doi:10.22323/1.301.1046
7. V.A. Allakhverdyan et al. (Baikal-GVD collaboration), Data Quality Monitoring system of the Baikal-GVD experiment, these proceedings
8. A.D. Avrorin et al. (Baikal-GVD collaboration), Baikal-GVD: Time Calibrations in 2016, PoS-ICRC2017-1036, doi:10.22323/1.301.1036
9. https://luigi.readthedocs.io/