BAIKAL-GVD

JOINT INSTITUTE FOR NUCLEAR RESEARCH JINR

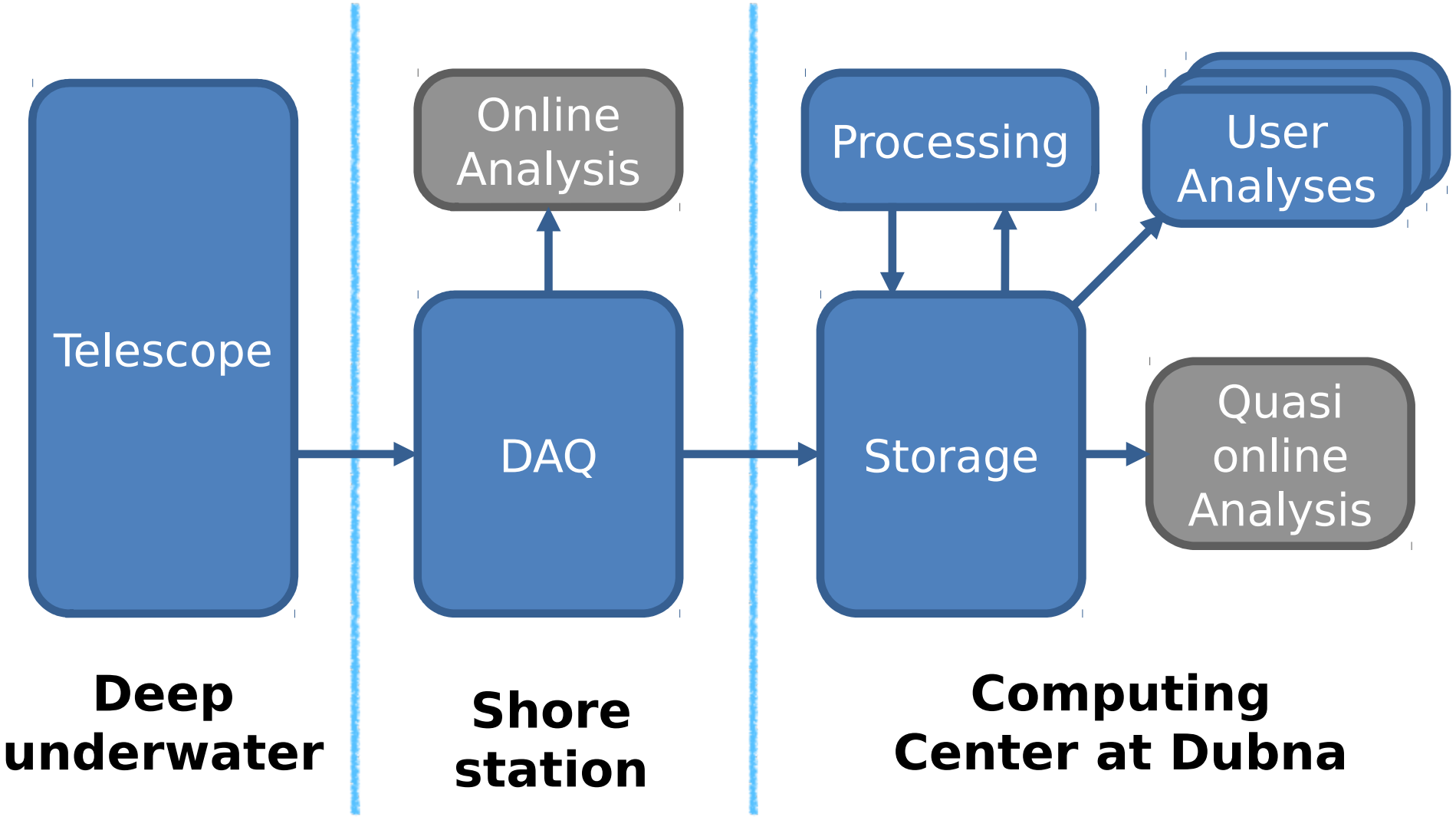
Data processing and quality monitoring of GVD

E. Khramov and B. Shaybonov (JINR)
on behalf of the GVD Collaboration

VLVnT – 2018, Dubna, Russia

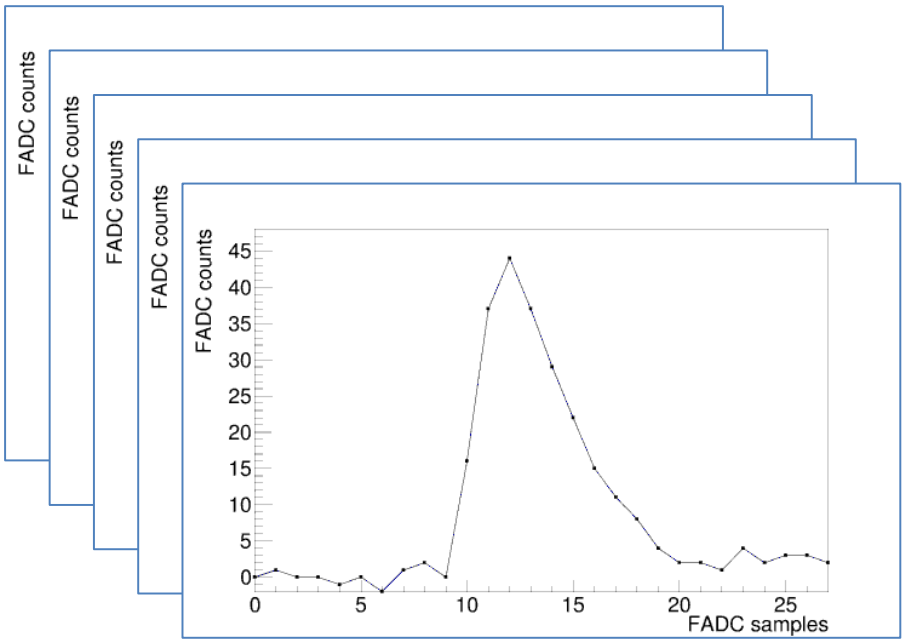
3 October 2018,

Overview


Telescope
Online Analysis
DAQ
Processing
Storage
User Analyses
Quasi online Analysis
Deep underwater
Shore station
Computing Center at Dubna

Telescope

- Offshore Trigger:
  - pair of adjacent OM in 100 ns window with High (~4.5 p.e.) and Low (~1.8 p.e.) thresholds (event rate 50 - 200 events/sec/cluster)
- Fiber Optic Line from Cluster Center to the shore
- Full FADC waveform data (12 bit, 5 ns)

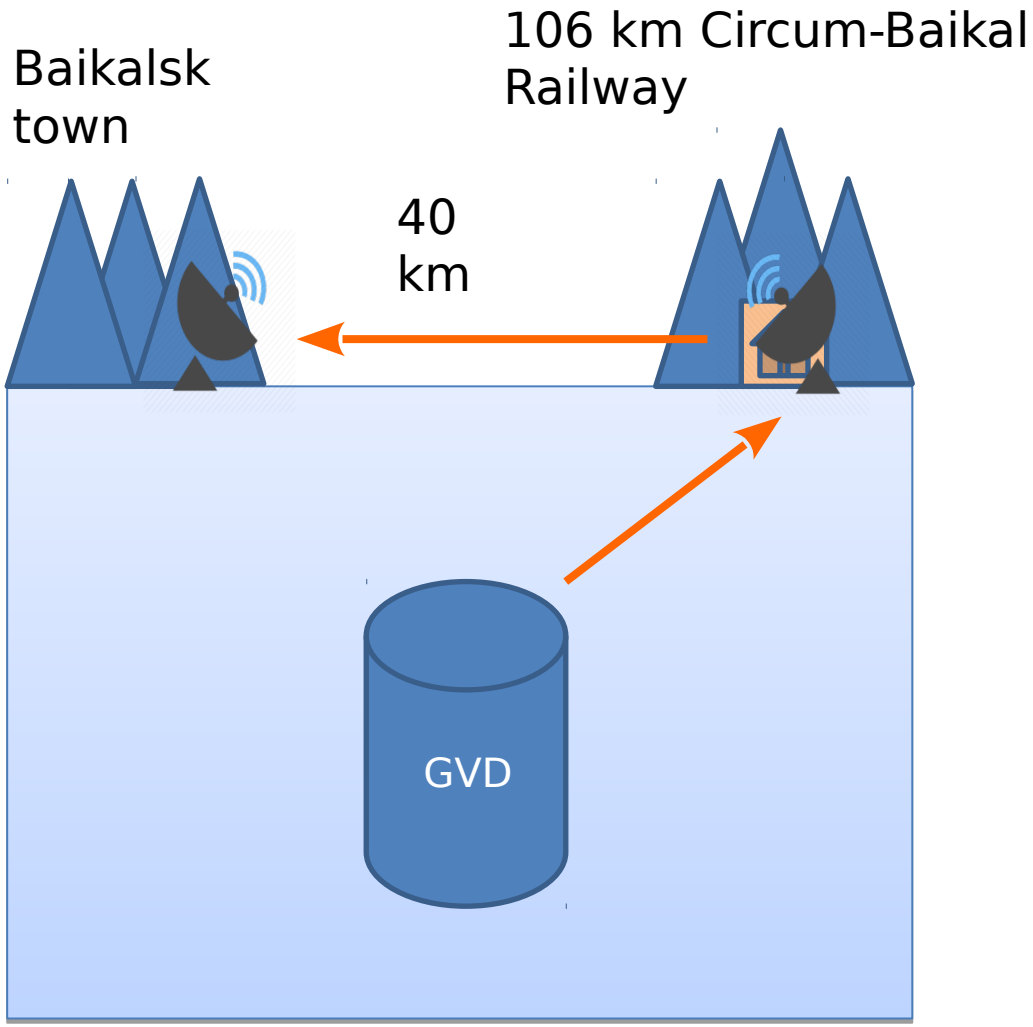


DAQ

- TCP-based network
- Output is archived in raw custom-made format
- Save output archive each 6 min (~36K events)
- 10 Gb/cluster/day archived data (compressed by factor of 4)

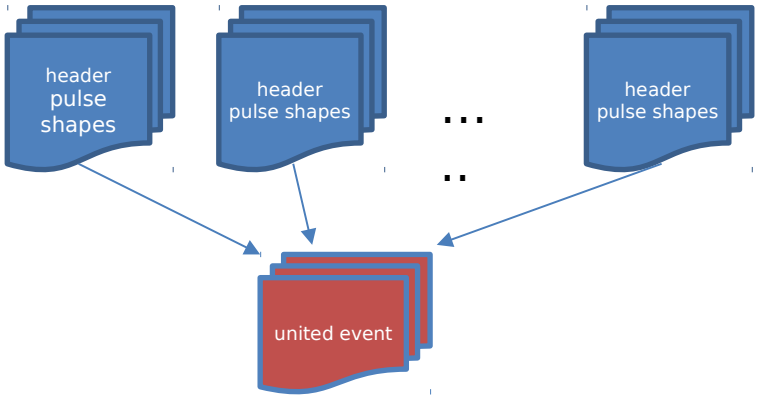
Storage

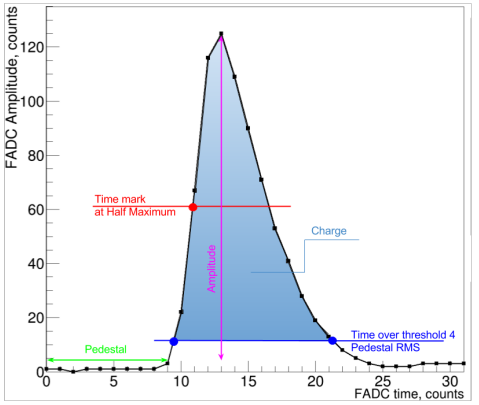
- Radio channel to another shore (40 km, 5 Mb/s)
- Raw files are copied to Dubna Computing Center as soon as it appears
- Stored on Dubna Cloud system (hard drives)

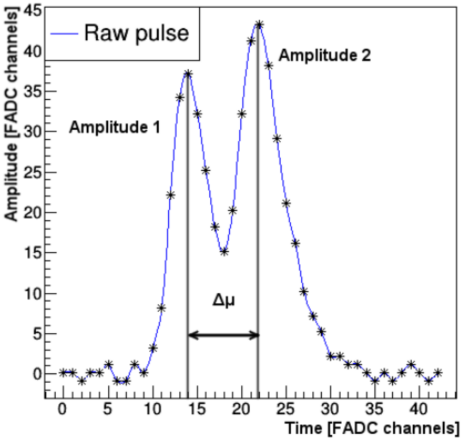


Processing

- Perform processing using Software Framework BARS
- Read raw files and convert to ROOT format
- Main jobs are scheduled, managing by MySQL and python scripts:
  - Event Builder
  - Pulse Shape Analysis
  - etc.
- Output in ROOT trees



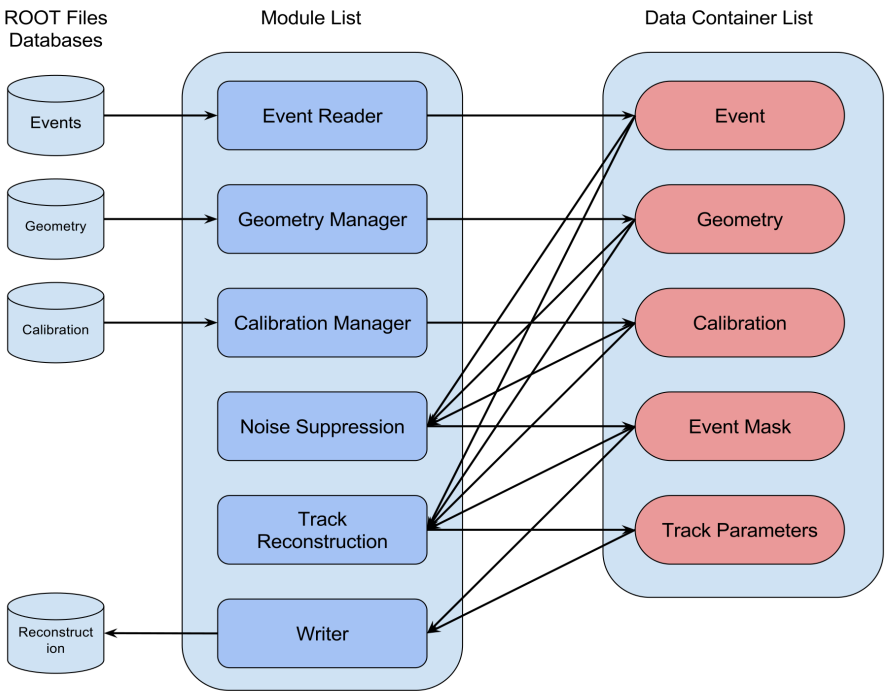




Analysis

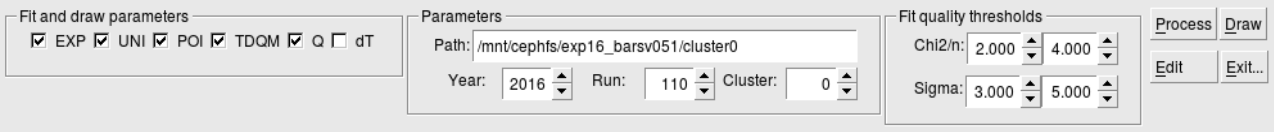
- User Analyses and Processing Jobs are performed in Baikal Analysis and Reconstruction Software (BARS)
- Written in C++, based on ROOT package, core from MAGIC framework MARS
- Link Modules to each other
- Data are encapsulated in Containers
- Support changing telescope geometry, calibrations, background noise via config files
- Easy to merge code from different groups into a single processing scheme

General scheme



Data quality monitoring

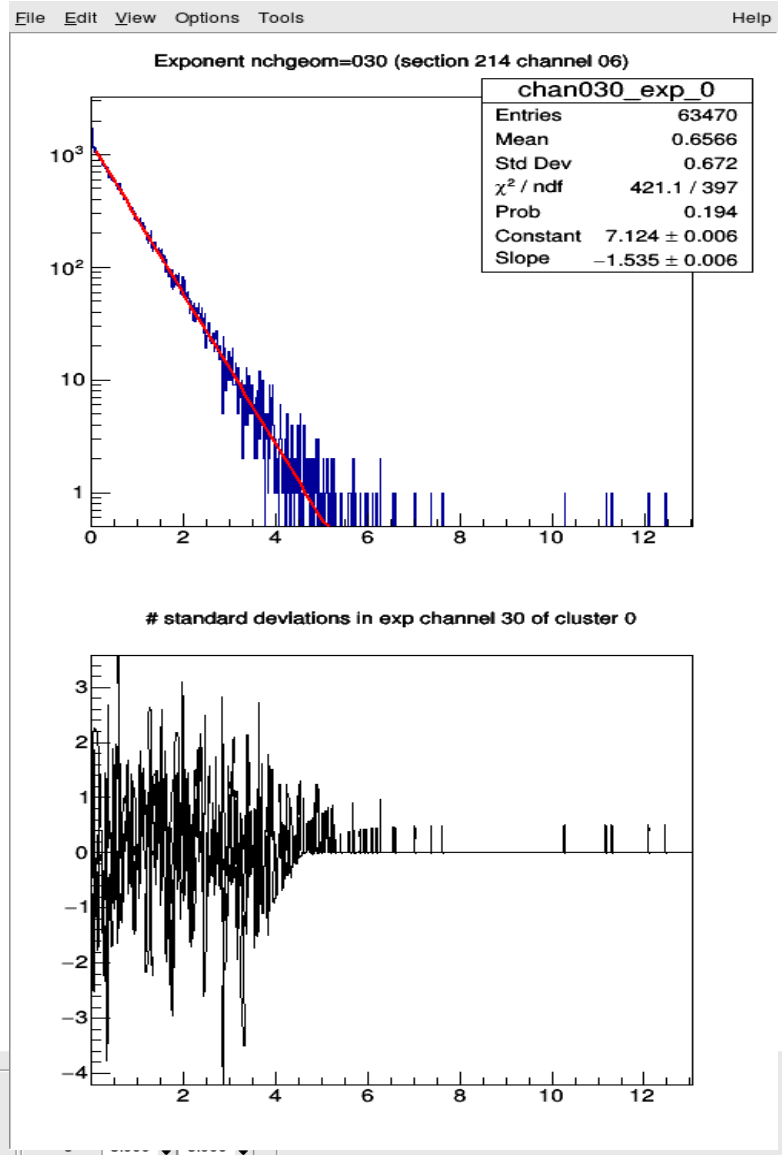
- ➔ Time difference between two neighbor events
- ➔ Events rate
- ➔ Average numbers of events per given time interval
- ➔ Triggers quality monitoring
- ➔ Charge distribution analyses:
  - 1 p.e. → amplitude calibration
  - High and low trigger thresholds
  - Full range analysis wrt baseline distributions
  - Sensitivity-wise monitoring



Data quality monitoring

time difference between events

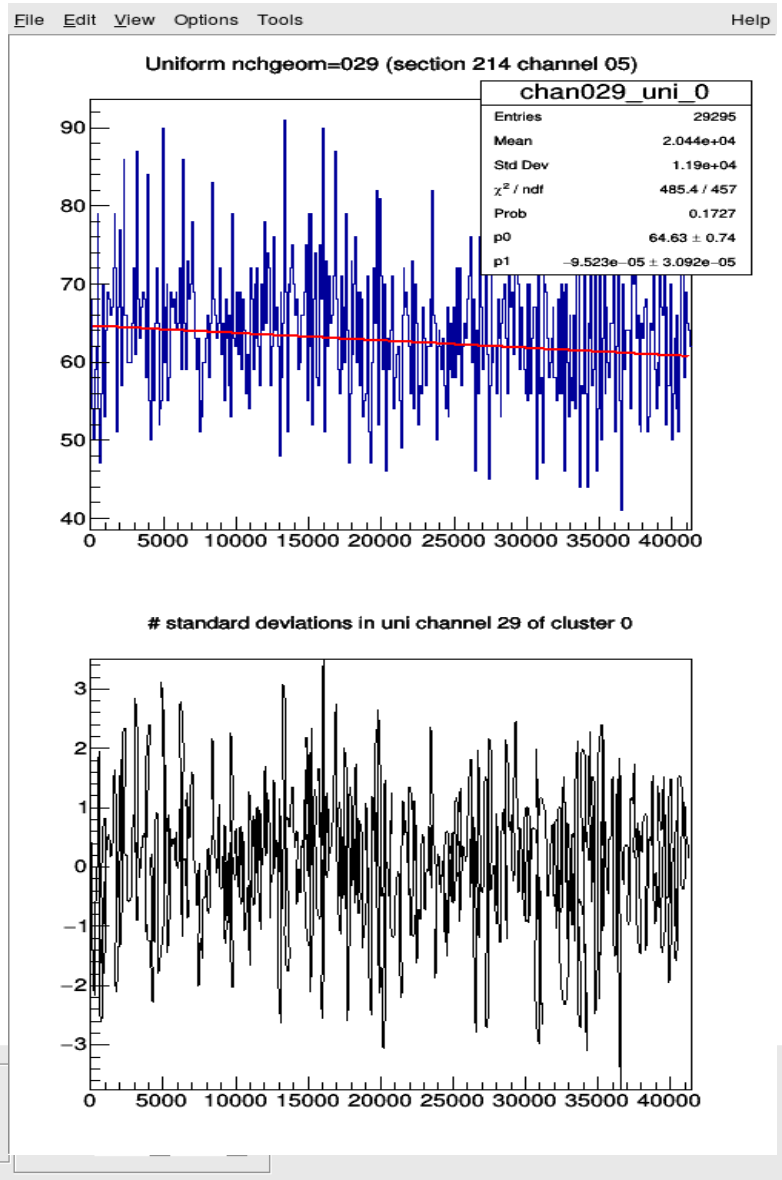
- Monitored at the levels of OM, section and cluster
- Expected to have exponential distribution
- Along with fit quality the fraction of bins with medium/high deviation from the fit function is the subject of solution



Data quality monitoring

events rate

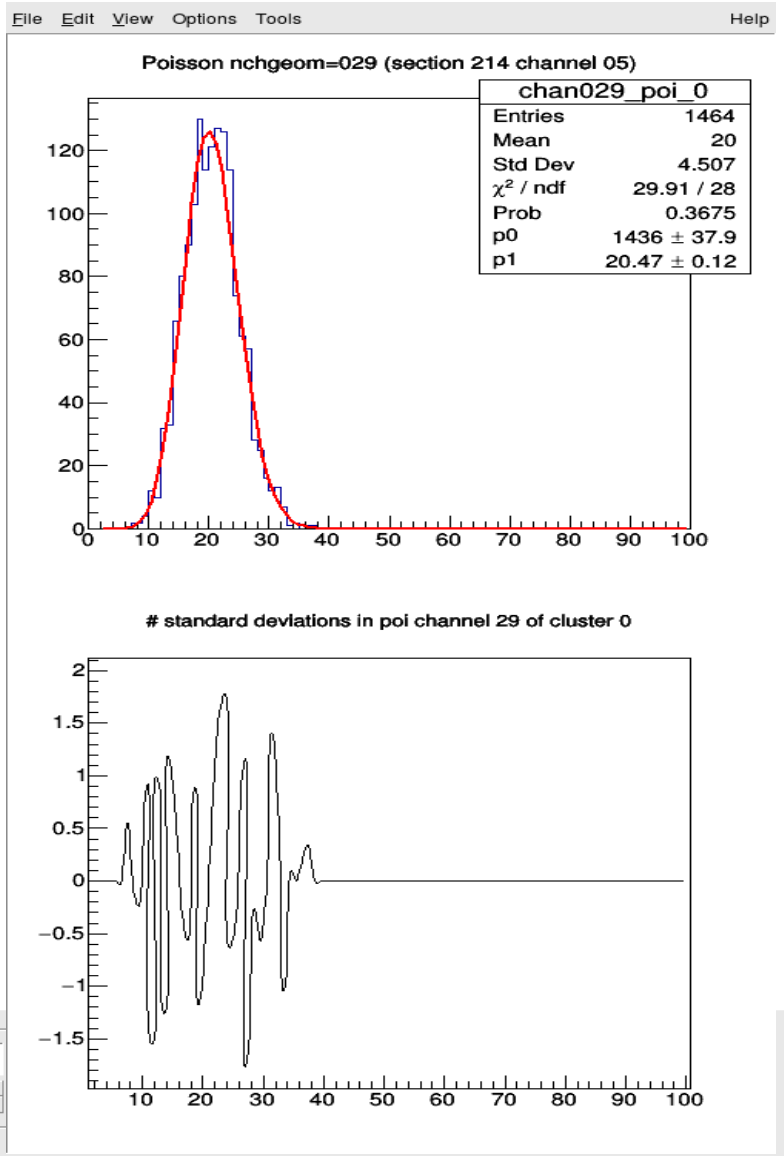
- Monitored at the levels of OM, section and cluster
- Expected to have linear distribution
- Along with fit quality the fraction of bins with medium/high deviation from the fit function is the subject of solution



Data quality monitoring

average numbers of events per given time interval

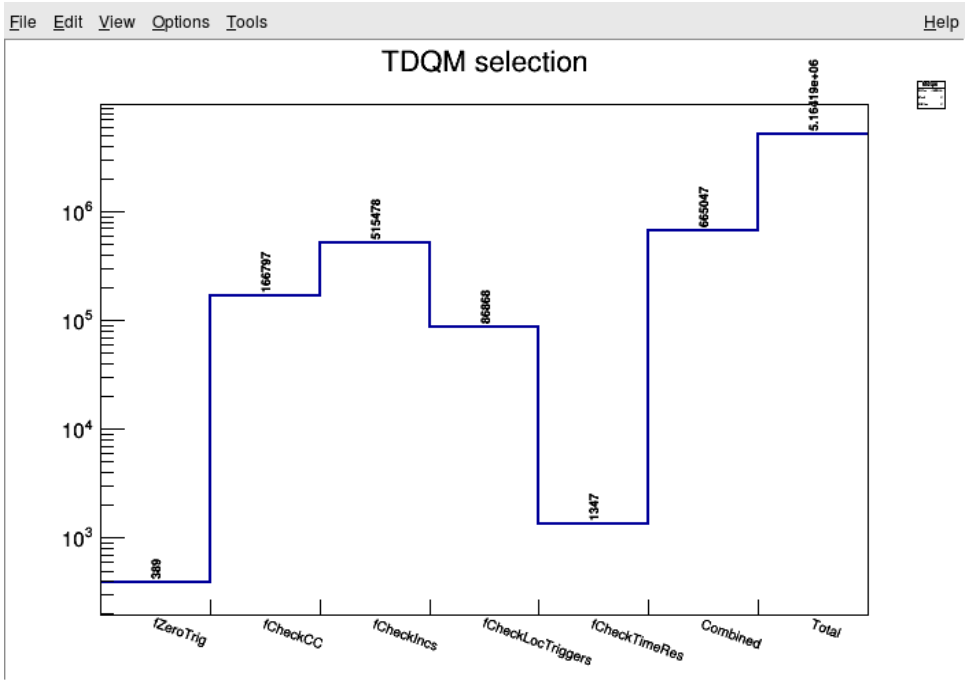
- Monitored at the levels of OM, section and cluster
- Expected to have Poisson distribution
- Along with fit quality the fraction of bins with medium/high deviation from the fit function is the subject of solution

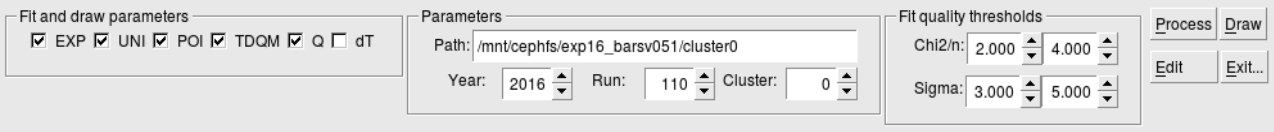


Data quality monitoring

trigger quality monitoring

- Monitored at the level of cluster
- Cross check if number of local/global request/acknowledgement increments are properly corresponds to each other
- Trigger efficiency should be higher that 90%

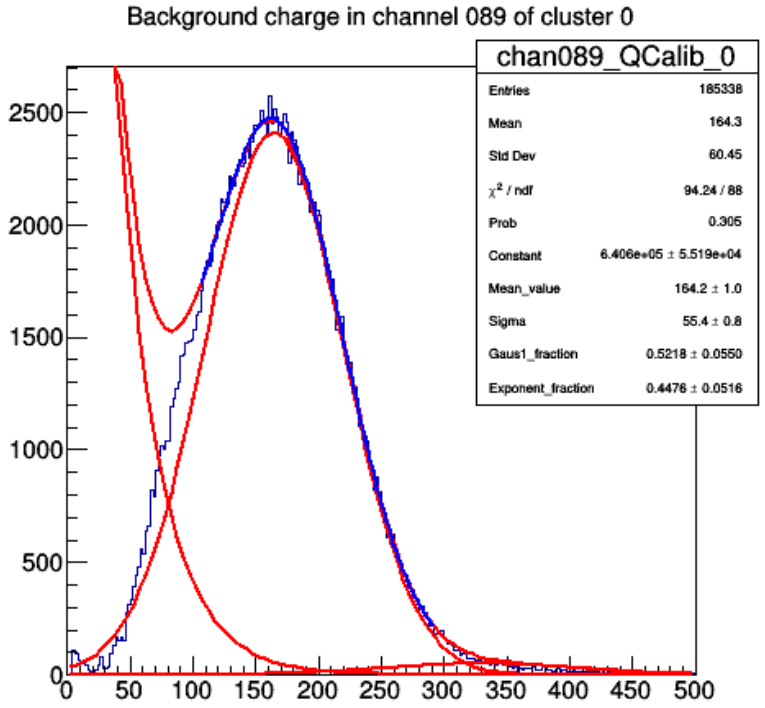


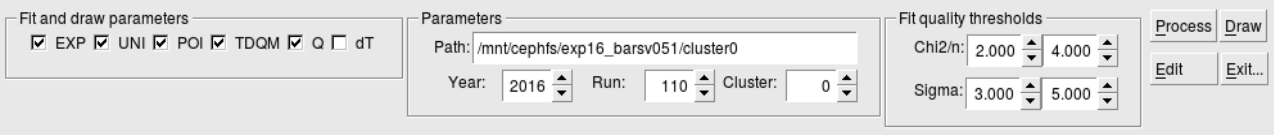


Data quality monitoring

Charge distribution analysis: amplitude calibration

- Monitored at the level of OM
- Expected to have exponential dark currents, 1 and 2 p.e. Gaussian impacts to the distribution
- Only fit quality in the given range is the subject of solution
- Mean values of the 1 p.e. Gaussian is extracted

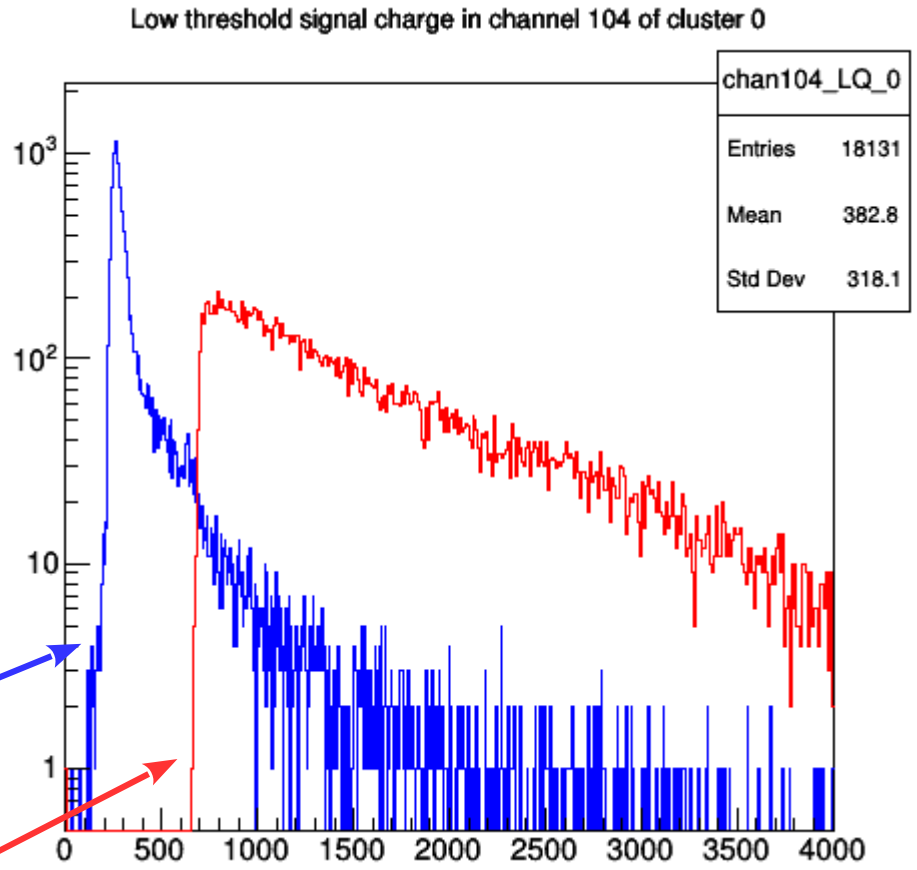




Data quality monitoring

Charge distribution analysis: high and low trigger thresholds

- Monitored at the level of OM
- Expected to have stair form
- No subject for solution
- Extract high and low threshold values



Low trigger threshold

High trigger threshold

Fit and draw parameters
EXP UNI POI TDQM Q dT

Parameters
Path: /mnt/cephfs/exp16_barsv051/cluster0
Year: 2016 Run: 110 Cluster: 0

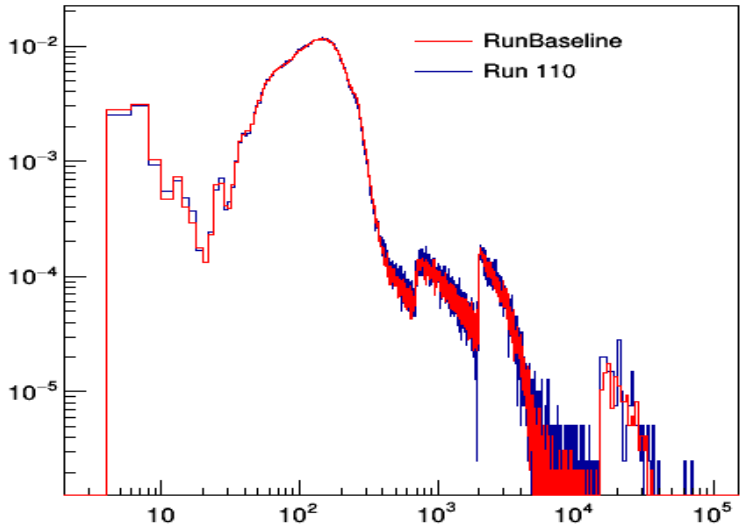
Fit quality thresholds
Chi2/n: 2.000 4.000
Sigma: 3.000 5.000

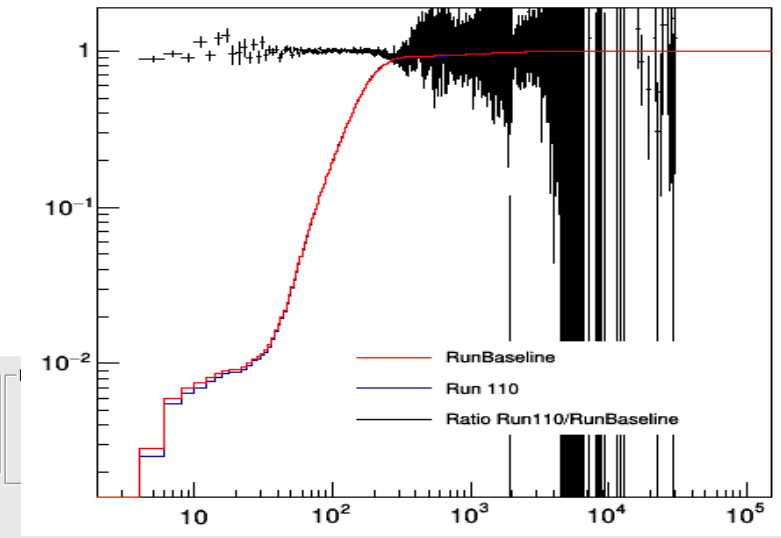
Process Draw
Edit Exit...

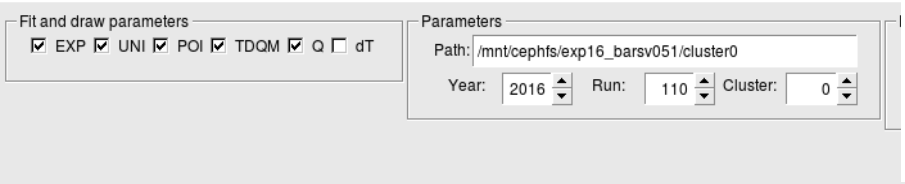
Data quality monitoring

Charge distribution analysis: range analysis wrt baseline distributions

- Monitored at the level of OM
- Non-equidistant binning
- Fit by pattern distribution of given OM from baseline run
- Only fit quality in the given range is the subject of solution



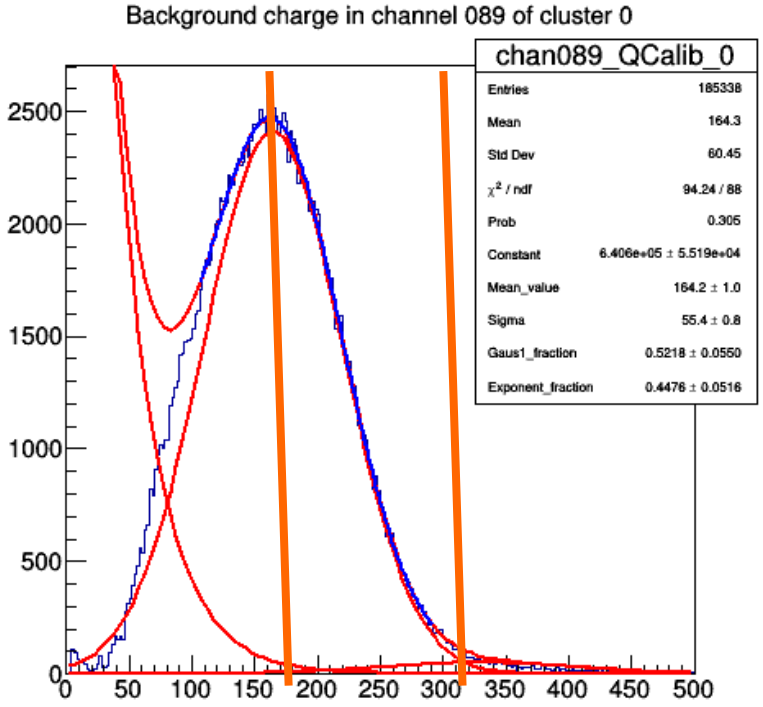


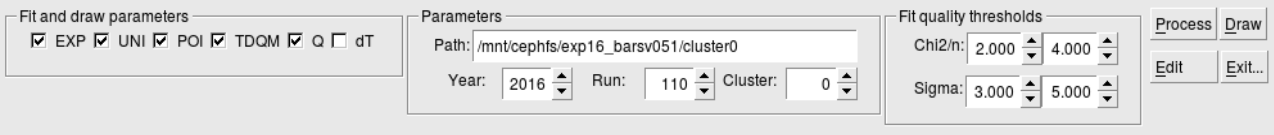


Data quality monitoring

Charge distribution analysis: sensitivity-wise monitoring

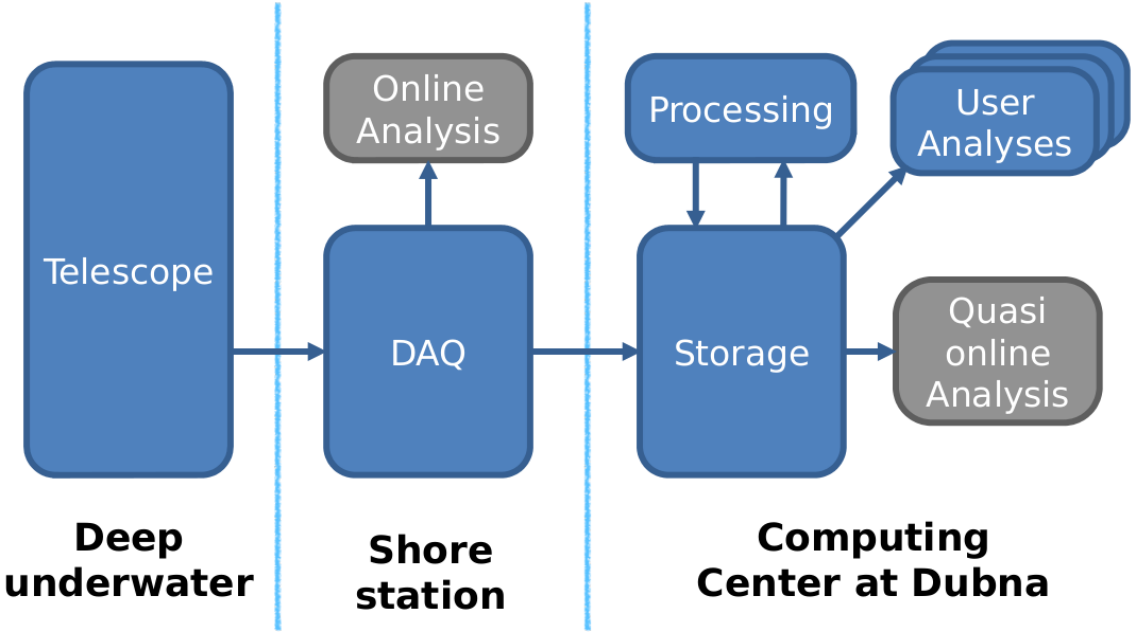
- Monitored at the level of OM
- Integral of 1 p.e. distribution in [1-2] p.e. range
- Estimate deviation of the given OM from average over layer
- This deviation is the subject of solution





Online Analysis

- Real time data stream that is available through TCP socket on the shore
- Latest raw data file (6 min of exposition) that is available in Dubna CC after few minutes



are under development

Thank you!

Backup slides

DQM decision making

What color of each cell will be?

Decision chain: channel → section → string → cluster

1) For EXP, UNI and POI check each bin for deviation in addition to χ^2/NDF check
   a) $N_{NormalBins}/N_{TotalBins}<25\%$ && $N_{BadBins}/N_{TotalBins}<1\%$ → **GOOD CELL → Masked as GREEN**
   b) $N_{NormalBins}/N_{TotalBins}<50\%$ && $N_{BadBins}/N_{TotalBins}<5\%$ → **NORMAL CELL → Masked as YELLOW**
   c) Any other case → **BAD CELL → Masked as RED**

2) For **channels** fit goodness of Charge and dT is checked as well

3) For **sections** check of numbers good/bad channels in addition to checks above:
   a) $N_{NormalChannels}<2$ && No and bad channel → **GOOD SECTION → Masked as GREEN**
   b) $N_{NormalChannels}<6$ && $N_{BadChannels}<2$ → **NORMAL SECTION → Masked as YELLOW**
   c) Any other case → **BAD SECTION → Masked as RED**

4) For **strings** check only number of good/bad sections in it:
   a) $N_{NormalSections}<2$ && No and bad section → **GOOD STRING → Masked as GREEN**
   b) $N_{NormalSections}<3$ && $N_{BadSectionss}<2$ → **NORMAL STRING → Masked as YELLOW**
   c) Any other case → **BAD STRING → Masked as RED**

DQM decision making

What color of each cell will be?

Decision chain: channel → section → string → cluster

1) For EXP, UNI and POI check each bin for deviation in addition to χ^2/NDF check
   a) $N_{NormalBins}/N_{TotalBins}<25\%$ && $N_{BadBins}/N_{TotalBins}<1\%$ → **GOOD CELL → Masked as GREEN**
   b) $N_{NormalBins}/N_{TotalBins}<50\%$ && $N_{BadBins}/N_{TotalBins}<5\%$ → **NORMAL CELL → Masked as YELLOW**
   c) Any other case → **BAD CELL → Masked as RED**

5) For **cluster** check of numbers good/bad strings in addition to checks above:
   a) $N_{NormalStrings}<2$ && No and bad string → **GOOD CLUSTER → Masked as GREEN**
   b) $N_{NormalStrings}<5$ && $N_{BadSections}<2$ → **NORMAL CLUSTER → Masked as YELLOW**
   c) Any other case → **BAD CLUSTER → Masked as RED**
   d) Fraction of events that DO NOT PASS global trigger quality criteria (TDQM) < 1% → **GOOD CLUSTER → Masked as GREEN**
   c) Fraction of events that DO NOT PASS global trigger quality criteria (TDQM) < 10% → **NORMAL CLUSTER → Masked as YELLOW**
   d) Any other case → **BAD CLUSTER → Masked as RED**