Development of the Double Cascade Reconstruction Techniques in the Baikal-GVD Neutrino Telescope

V.A. Allakhverdyan,[a] A.D. Avrorin,[b] A.V. Avrorin,[b] V.M. Aynutdinov,[b] R. Bannasch,[c] Z. Bardačová,[d] I.A. Belolaptikov,[a] I.V. Borina,[a] V.B. Brudanin,[a,1] N.M. Budnev,[e] V.Y. Dik,[a] G.V. Domogatsky,[b] A.A. Doroshenko,[b] R. Dvornický,[a,d] A.N. Dyachok,[e] Zh.-A.M. Dzhilkibaev,[b] E. Eckerová,[d,*] T.V. Elzhov,[a] L. Fajt,[f] S.V. Fialkovski,[g,1] A.R. Gafarov,[e] K.V. Golubkov,[b] N.S. Gorshkov,[a] T.I. Gress,[e] M.S. Katulin,[a] K.G. Kebkal,[c] O.G. Kebkal,[c] E.V. Khramov,[a] M.M. Kolbin,[a] K.V. Konischev,[a] K.A. Kopański,[h] A.V. Korobchenko,[a] A.P. Koshechkin,[b] V.A. Kozhin,[i] M.V. Kruglov,[a] M.K. Kryukov,[b] V.F. Kulepov,[g] Pa. Malecki,[h] Y.M. Malyshkin,[a] M.B. Milenin,[b] R.R. Mirgazov,[e] D.V. Naumov,[a] V. Nazari,[a] W. Noga,[h] D.P. Petukhov,[b] E.N. Pliskovsky,[a] M.I. Rozanov,[j] V.D. Rushay,[a] E.V. Ryabov,[e] G.B. Safronov,[b] B.A. Shaybonov,[a] M.D. Shelepov,[b] F. Šimkovic,[a,d,f] A.E. Sirenko,[a] A.V. Skurikhin,[i] A.G. Solovjev,[a] M.N. Sorokovikov,[a] I. Štekl,[f] A.P. Stromakov,[b] E.O. Sushenok,[a] O.V. Suvorova,[b] V.A. Tabolenko,[e] B.A. Tarashansky,[e] Y.V. Yablokova,[a] S.A. Yakovlev[c] and D.N. Zaborov[b]

[a]*Joint Institute for Nuclear Research, Dubna, Russia*

[b]*Institute for Nuclear Research, Russian Academy of Sciences, Moscow, Russia*

[c]*EvoLogics GmbH, Berlin, Germany*

[d]*Comenius University, Bratislava, Slovakia*

[e]*Irkutsk State University, Irkutsk, Russia*

[f]*Czech Technical University in Prague, Prague, Czech Republic*

[g]*Nizhny Novgorod State Technical University, Nizhny Novgorod, Russia*

[h]*Institute of Nuclear Physics of Polish Academy of Sciences (IFJ PAN), Kraków, Poland*

[i]*Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, Russia*

[j]*St. Petersburg State Marine Technical University, St.Petersburg, Russia*

E-mail: eliska.eckerova@fmph.uniba.sk

The Baikal-GVD is a neutrino telescope under construction in Lake Baikal. The main goal of the Baikal-GVD is to observe neutrinos via detecting the Cherenkov radiation of the secondary charged particles originating in the interactions of neutrinos. In 2021, the installation works concluded with 2 304 optical modules installed in the lake resulting in effective volume $\sim 0.4 \, \text{km}^3$. In this paper, the first steps in the development of double cascade reconstruction techniques are presented.

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

---

[1]Deceased.

[*]Presenter

1. Introduction

The Baikal Gigaton Volume Detector (Baikal-GVD) [1] is a cubic-kilometer scale neutrino telescope located in the deepest freshwater lake in the world – Lake Baikal installed approximately 3 - 4 km from shore at depths of $\sim$ 750 - 1275 m. This three dimensional array of photomultiplier tubes aims to detect the Cherenkov radiation emitted in water by products of neutrino interactions.

The basic and independently working unit of the Baikal-GVD detector is called cluster, which consists of 8 strings - one central and seven peripheral separated by about 60 m. The distance between the centers of two neighboring clusters is approximately 300 m (see Fig. 1a). On every string, 36 Optical Modules (OM) are installed (see Fig. 1b) with vertical spacing 15 m, resulting in 288 OMs for every cluster. The main component of the OM is a photomultiplier tube with hemispherical photocathode enclosed in pressure-resistant glass sphere with diameter of 42 cm.
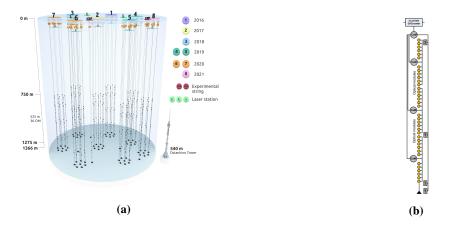

(a)

(b)

Figure 1: a) Design of the Baikal-GVD telescope in 2021. There are 8 clusters, 3 laser strings, and 2 experimental strings installed and operating. b) Illustration of the string of the Baikal-GVD telescope consisting of 36 OMs arranged in 3 sections.

One of the methods for astrophysical neutrino detection is an observation of ν_τ because the rate of ν_τ production in the atmosphere is almost negligible [2]. Therefore observed tau neutrino immediately confirms its astrophysical origin. In the charged current interaction of ν_τ, τ lepton is created in the hadronic cascade. If τ decays into electron or hadrons, the second cascade is produced. As a result, the double cascade signature is created.

In this paper, techniques for the reconstruction of double cascades are presented. The first method is a double pulse detection technique for the reconstruction of double cascade events with small distance between cascade vertices. The second method is based on the identification of two distinct cascades created in a single cluster and the third one combines single cluster single cascade reconstruction technique with multi-cluster events studies.

2. Search for Double Cascades with Double Pulse Detection Method

To be able to reconstruct double cascade events with very small distance between cascade vertices, the events have to be studied on the waveform level. Thus, double pulse detection method was developed.

To tag the potential double pulses, the differentiation method was selected. A potential double pulse is tagged if a sign of the first derivative changes from negative to positive (see Fig. 2a). Hence a creation of local minimum in the waveform is required.
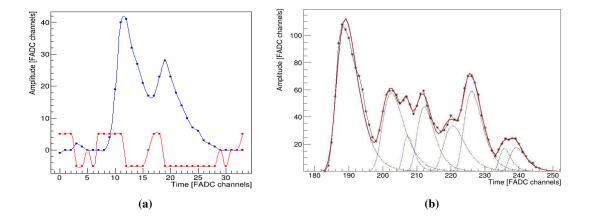

(a)   **(b)**

Figure 2: a) Illustration of the differentiation method. The waveform is displayed with the blue line. The red line corresponds to the sign of the first derivative multiplied by a factor of 5 for the sake of visualization. b) Multiple pulse fitting. The black line corresponds to experimental waveform. The fit with sum of the Gumbel functions is displayed with the red line. The black dotted peaks correspond to individual Gumbel functions.

To determine the minimal time delay between pulses needed for the formation of a local minimum with respect to the charges of both pulses, simulations of the Gumbel functions were studied. The analytical form of the Gumbel function is following:

$$f(x) = a e^{-((\frac{x-\mu}{\beta}) + e^{-(\frac{x-\mu}{\beta})})}, \tag{1}$$

where a is a normalization coefficient, μ is a position of maximum, and β is a width of the peak. This analysis showed that the minimal time delay between pulses required for formation of a local minimum is ~ 25 ns.

However, many potential double pulses tagged by the differentiation method are fake double pulses created by pedestal oscillations. To suppress them, a multivariate machine learning technique – Boosted Decision Trees (BDT) from ROOT TMVA package [3] was implemented. For BDT training, 12 parameters were used. Both datasets, the signal dataset (real double pulses) and background dataset (fake double pulses), contain attributes of approximately 3000 waveforms for training and testing of BDT.

The output from trained BDT – the distribution of BDT response value is shown in Fig. 3a. With a cut on BDT response value at -0.03, the signal efficiency is 99.6%, and the background efficiency is 1.4% (see Fig. 3b) [4].

This method can be used to identify not only the double pulses but any higher multiplicity of pulses (multiple pulses). Conclusively, individual pulses can be fit with the Gumbel function and their parameters can be determined. The demonstration of this procedure applied on experimental pulse is shown in Fig. 2b.
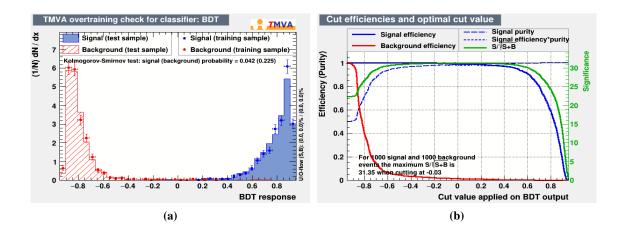
3

Figure 3: a) Distribution of BDT response value for background (red color) and signal (blue color). b) Distributions of signal efficiency and background efficiency with respect to the cut value applied on BDT output. [4]

3. Single Cluster Double Cascade Reconstruction Algorithm

The double cascade reconstruction algorithm consists of three main parts – signal/noise hits selection, division of pulses for separate reconstruction of two cascades, and two independent single cascade reconstructions. Input to this algorithm is a set of pulses recorded by photomultiplier tubes.

The first step is to separate signal and noise pulses. Pulse with the highest charge is chosen as a reference pulse. Other pulses are selected with respect to the reference pulse according to the causality criterion:

$$| t - t_i | < d_i/v + \delta t, \tag{2}$$

where t is detection time of the reference pulse, t_i is detection time of the pulse to be analyzed, d_i is distance between OMs of the two pulses, and δt is parameter that determine stringency of causality criterion. There are also additional criteria on neighboring OMs that detected hits. With $\delta t = 80$ ns, average efficiency of signal pulse selection is at the level of 80% and purity is 99%.

In the second step, the set of pulses from the first step is divided to two subsets for two independent single cascade reconstructions. The pulses are selected based on estimated positions and times of cascades.

Firstly, a set of five random pulses is chosen. These five pulses are used for cascade position and time estimation. The cascade position and time are estimated by solving a set of equations for distances between OMs that detected particular pulses and position of cascade vertex d_i:

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} = c/n(t - t_i), \tag{3}$$

where $x = (x, y, z, t)$ is space-time position of cascade vertex, and $x_i = (x_i, y_i, z_i, t_i)$ is space-time position of OM that recorded particular hit [5]. If these five pulses correspond to one cascade only, cascade position and time should be estimated accurately (see Fig. 4a). If this set is composed of pulses from two cascades, then the cascade position and time are estimated inaccurately (see Fig. 4b).
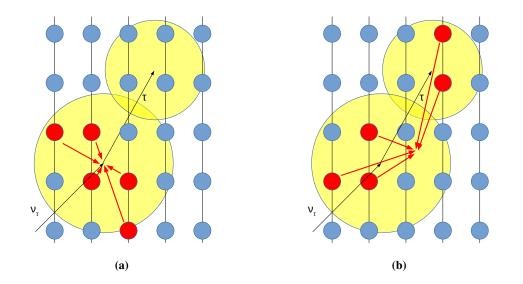
4

Figure 4: Illustration of cascade position estimation. a) Pulses for position estimation correspond to one cascade only ⇒ accurate position estimation. b) Pulses for position estimation correspond to two cascades ⇒ inaccurate position estimation.

Demonstration of this method on one event, for all combinations of five pulses, is shown in Fig. 5, two distinct peaks that correspond to vertices of two cascades are created. Accurately estimated position and time of cascade enable to select pulses that correspond to particular cascade. Pulses are selected with respect to the estimated position $\vec{R}$ and time T of cascade according to the criterion:

$$| T_i^{meas} - T_i^{exp}(\vec{R}, T) | \lesssim \delta T, \tag{4}$$

where T_i^{meas} is detection time of the pulse, T_i^{exp} is expected time of pulse detection, and δT determines strictness of the criterion. There are also additional criteria on the charge of studied pulse, and the distance between cascade position and OM that detected studied pulse. The estimation of position and time of the cascade is computationally extensive if all possible combinations of five pulses are taken into account. Therefore, if the chosen subset of pulses satisfies a set of certain conditions, this subset of pulses is chosen for single cascade reconstruction, and the selection of sets of five pulses is ceased.

The last step is to process two subsets of pulses with single cascade reconstruction [6]. An event has to have at least 6 hits on at least 3 strings (6/3 filter) to be reconstructed with single cascade reconstruction. For that reason both subsets of pulses for double cascade reconstruction have to pass this criterion. Subsequently, selected sets of pulses undergo single cascade reconstruction procedure. In the single cascade reconstruction, time and position of the cascade are estimated by minimization of χ^2:

$$\chi^2 = \frac{1}{N_{hit} - 4} \sum_{i=1}^{N_{hit}} \frac{(T_i^{meas} - T_i^{exp}(\vec{R}, T))^2}{\sigma_t^2}, \tag{5}$$

where N_{hits} is number of hit OMs, T_i^{meas} is experimentally measured time of pulse detected on i^{th}
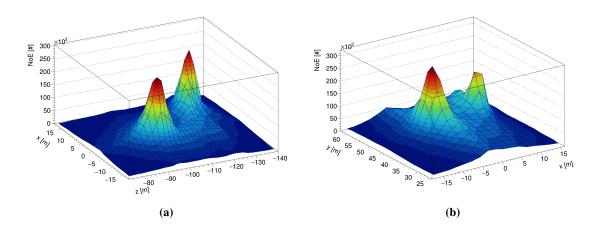
(a)   **(b)**

Figure 5: The results of position estimation procedure, considering all combinations of five pulses, for MC ν_τ event with ν_τ cascade vertex position [9, 47, -126], τ cascade vertex position [-3, 45, -100]. The MC positions are expressed in meters. a) Display of XZ plane. a) Display of XY plane.

OM, T_i^{exp} is theoretically expected time of pulse detection on i^{th} OM, given that cascade vertex time and position are T and $\vec{R}$, and σ_t is the uncertainty in time measurements. Energy and direction of a cascade are determined via minimization of likelihood function:

$$L = - \sum_{i=1}^{hitOM} log(P_i(q_i \mid Q_i)) - \sum_{i=1}^{unhitOM} log(P_i(q_i = 0 \mid Q_i)), \tag{6}$$

where P_i is the Poisson probability of detecting charge q_i on i^{th} OM while charge Q_i is expected.

The Monte Carlo (MC) simulations of ν_τ interaction were processed with aforementioned reconstruction algorithm [7]. Only double cascade events with distance between cascade vertices higher than 10 m were taken into account. The reconstruction efficiency of the double cascade events is defined as the ratio of events that passed certain criteria and total events simulated in MC. In Fig. 6a, the reconstruction efficiency after 6/3 filter is displayed. The reconstruction efficiency after whole double cascade reconstruction algorithm is shown in Fig. 6b (tolerance on mismatch between simulated and reconstructed position is 5 m for both cascades). An example of a reconstructed double cascade event is displayed in Fig. 7.

4. Multi-cluster Double Cascades Reconstruction

The multi-cluster double cascade identification technique combines single cascade reconstruction algorithm with multi-cluster events. A double cascade event is tagged if two single cascades were reconstructed separately in two different clusters satisfying time coincidence condition.

Approximately 87 000 cascades were reconstructed by means of single cascade reconstruction technique [7] in the experimental data collected in year 2019. The same dataset has been processed with multi-cluster double cascade identification algorithm. This method tagged one double cascade event which means that two single cascades were reconstructed within the coincidence time window. Reconstructed attributes of the two cascades are shown in Tab. 1.
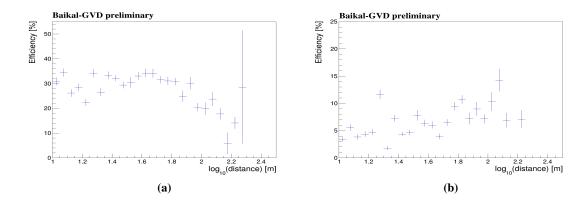
(a)

(b)

Figure 6: a) The dependence of reconstruction efficiency on MC simulated distance between cascades after 6/3 filter. b) The distribution of reconstruction efficiency with respect to MC simulated distance between cascades after double cascade reconstruction (tolerance on mismatch between simulated and reconstructed position is 5 m).
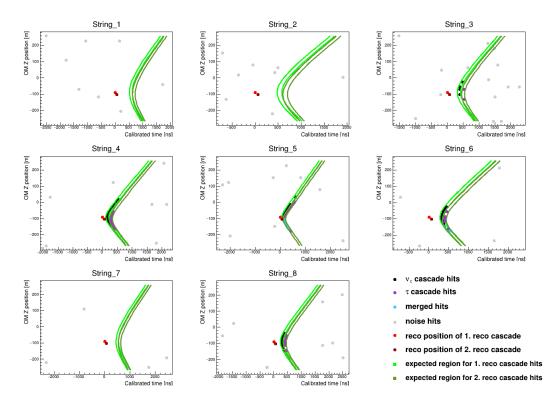

Figure 7: Reconstructed MC ν_τ event visualization. The dependence of time on the Z coordinate for all eight strings is shown. The MC simulated distance between cascades is 18.17 m. The mismatch between simulated and reconstructed position of ν_τ cascade is 1.84 m and 3.27 m for τ cascade.

According to the reconstructed positions of two cascades, the distance between cascades was determined as ~ 328.75 m. Taking into account reconstructed attributes of two cascades, the probability of ν_τ origin of this event is close to zero.

7

Table 1: Reconstructed parameters of cascades from double cascade multi-cluster event: zenith angle, azimuth angle, reconstructed energy, number of hits used in reconstruction, and likelihood.

	θ [rad]	ϕ [rad]	Energy [TeV]	nRecoHits	likelihood
cascade 1	2.20	3.85	8.06	19	0.92
cascade 2	2.32	3.66	4.72	17	1.08

5. Conclusion

The double pulse detection method based on machine learning technique BDT was developed. The signal efficiency gained is 99.6%, and the background efficiency is 1.4%. The usage of this method for multiple pulses identification and separation was studied. The first single cluster double cascade reconstruction algorithm has been developed. The individual steps of this algorithm were described. The first preliminary results were studied. The multi-cluster double cascade identification algorithm was applied on data collected in year 2019. One multi-cluster double cascade event was identified. According to the analysis of reconstructed parameters of two cascades, ν_τ origin of this event can be excluded.

6. Acknowledgments

References

[1] https://baikalgvd.jinr.ru/

[2] A. Palladino et al., *The importance of observing astrophysical tau neutrinos,* J. Cosmol. Astropart. P. 8 (2018).

[3] A. Hoecker et al., *TMVA 4 - Toolkit for Multivariate Data Analysis with ROOT, Users Guide,* ArXiv:physics/0703039 (2017).

[4] E. Eckerová, *Advanced Double Pulse Detection Techniques in the Baikal-GVD Neutrino Telescope,* Master's thesis, Comenius University, Slovakia (2020).

[5] B. Hartmann, *Reconstruction of Neutrino-Induced Hadronic and Electromagnetic Showers with the ANTARES Experiment,* PhD thesis, Univ. Erlangen (2006). arXiv:astro-ph/0606697

[6] Baikal-GVD Collaboration: A.D. Avrorin et al., *Search for cascade events with Baikal-GVD,* PoS-ICRC2019-0873, arXiv:1908.05430, doi:10.22323/1.358.0873

[7] O. Tange, *GNU Parallel - The Command-Line Power Tool*, The USENIX Magazine (2011), doi: 10.5281/zenodo.16303

PoS(ICRC2021)1167