An efficient hit finding algorithm for Baikal-GVD muon reconstruction

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: avrorin@inr.ru, bairsh@yandex.ru

The Baikal-GVD is a large scale neutrino telescope being constructed in Lake Baikal. The majority of signal detected by the telescope are noise hits, caused primarily by the luminescence of the Baikal water. Separating noise hits from the hits produced by Cherenkov light emitted from the muon track is a challenging part of the muon event reconstruction. We present an algorithm that utilizes a known directional hit causality criterion to contruct a graph of hits and then use a clique-based technique to select the subset of signal hits. The algorithm was tested on realistic detector Monte-Carlo simulation for a wide range of muon energies and has proved to select a pure sample of PMT hits from Cherenkov photons while retaining above 90% of original signal.

---

[1]Deceased.

[*]Presenter

1. Introduction

Baikal-GVD [1, 2] is a cubic kilometer scale underwater neutrino telescope currently under construction in the southern basin of Lake Baikal. The telescope operates by detecting Cherenkov radiation from neutrino induced secondary particles in the Baikal water with an array of PMTs submerged between the depths of 750 and 1275 meters. The PMTs are organized into clusters, with 288 PMTs in each cluster. In its present configuration Baikal-GVD consists of 8 clusters and is currently taking data.

Most of the PMT hits are produced by noise rather than Cherenkov photons. While there are several sources of noise in GVD, including PMT dark current and electronic noise, they are dominated by 1 p.e. signal from the Baikal water luminiscence [3]. The PMT noise counting rates can vary between 10 and 200 kHz and, as can be seen in [4], exhibit heavy variation with time and elevation. Suppressing noise hits while preserving signal is of particular importance for the muon track reconstruction where a charge deposition on a single PMT can often be comparable to the noise.

We present an approach for selecting signal hits based on the ScanFit technique introduced in [5]. Similarly to ScanFit, for each event we assume a single muon track model and cover the sky with the lattice of possible track origin points corresponding to a considered track direction. A set of hits is selected for each direction using a directional causality test (called 1-D clustering in [5]) and the set that produces the optimal hit time based fit for a single track hypothesis is considred to be the signal, providing both noise suppression and a track prefit for reconstruction.

Our approaches to hit selection for a particular direction and picking the best direction, however, are novel. For each considered direction we are using a maximal clique search algorithm to find subsets of hits where each pair of hits passes a directional causality test. The hits in the largest or second-largest subset with an optimal signal track fit are then selected as the signal. The direction of the winning fit can also be used as a track prefit for track reconstruction. Due to the computational cost of clique search for multiple directions we also apply a prefilter to remove easily identifiable noise hits while keeping as much of the signal as possible.

This approach operates on the assumption that if we treat the directional causality condition as a binary test for a pair of hits, each pair of signal hits should pass, while a noise hit is unlikely to pass it in a pair with every signal hit in the event. This prevents rejecting otherwise isolated hits, increasing hit selection efficiency and, as the number of signal hits increases, the number of causality tests a noise hit must pass grows as well, increasing hit selection purity. For an atmospheric neutrino sample and low noise conditions, the resulting hit selection has an average hit-wise purity and efficiency of 95%.

2. Directional causality

In this work we heavily use a directional causality criterion from [5]. For a predefined track direction and a pair of causaly related hits i and j, the following conditions must be true:

$$|\Delta r_{i,j}| < R$$
$$\Delta z_{i,j} - kR - w \leq c\Delta t_{i,j} \leq \Delta z_{i,j} + kR + w$$

Here, $\Delta r_{i,j}$ is the difference between hit PMTs distances to the muon track, $\Delta z_{i,j}$ is the distance between hit PMTs along the muon track, R is the distance between hit PMTs in the plane orthogonal to the track, $\Delta t_{i,j}$ is time difference between hits, w is a tunable parameter used to compensate for hit timing uncertainty and k is a constant depending only on the refraction index of the observed medium. Because the track direction is predefined, the time window imposed by this criterion on a pair of hits is narrow and noise hits are less likely to fall into it.

Note that the directional causality relies only track direction, which makes the scanning part of the primary algorithm possible.

3. Algorithm description

The hit selection algorithm is performed in three steps: initial noise rejection, scanning the sky and selecting hit subsets for each direction, and selecting the optimal direction and a subset of hits. Each step is described in a subsection below.

3.1 Prefilter

At this stage a set of hits is preselected for processing. Processing time in the following stages increases quadratically with the number of hits and linearly with the number of considered directions, so at this stage we want to remove easily identifiable noise. This is accomplished in three steps. First, we construct a causality graph of the event - a graph where each hit is represented by a vertex and two vertices share an edge if the corresponding hits satisfy causality conditions as described in [6]. The causality parameters are relaxed to avoid rejecting signal hits.

Next, we use the Bron-Kerbosch algorithm [7] to find the maximal cliques in the causality graph - the largest subsets of vertices where all vertices are connected to each other. This way, each clique corresponds to a subset of hits where each pair of hits satisfies the causality conditions. Once the maximal cliques are found, all the hits that are not in the largest clique are rejected. If there are multiple largest cliques of equal size, the hits from all of these cliques are preserved.

Finally, the events with fewer then 6 hits on 3 strings are discarded. This primarily filters purely noise induced events. For an atmospheric neutrino sample and low noise conditions, average prefilter hit-wise efficiency is 98% and average hit-wise purity is 69%.

3.2 Scan

A set of fixed directions in the sky - a scan grid - is considered. In this work we use a rectangular grid in spherical coordinates as a scan grid. Each direction in the scan grid is a muon origin point and is associated with a muon track parametrized by the tuple $(\theta, \phi, X_0, Y_0, T_0)$, where θ and ϕ are polar and azimuth angle of the track, X_0 and Y_0 are coordinates of the track projection to the plane orthogonal to the track direction and T_0 is the time when the track crosses this plane. For each track θ and ϕ are fixed, while X_0, Y_0 and T_0 are free parameters.

For each direction in the scan grid, similarly to the prefilter step, we construct a causality graph from the event hits. This time, however, we are using directional causality to establish whether two vertices corresponding to a pair of hits should share an edge. Because each direction in the scan grid is associated with a different track model, each direction will have its own causality graph.
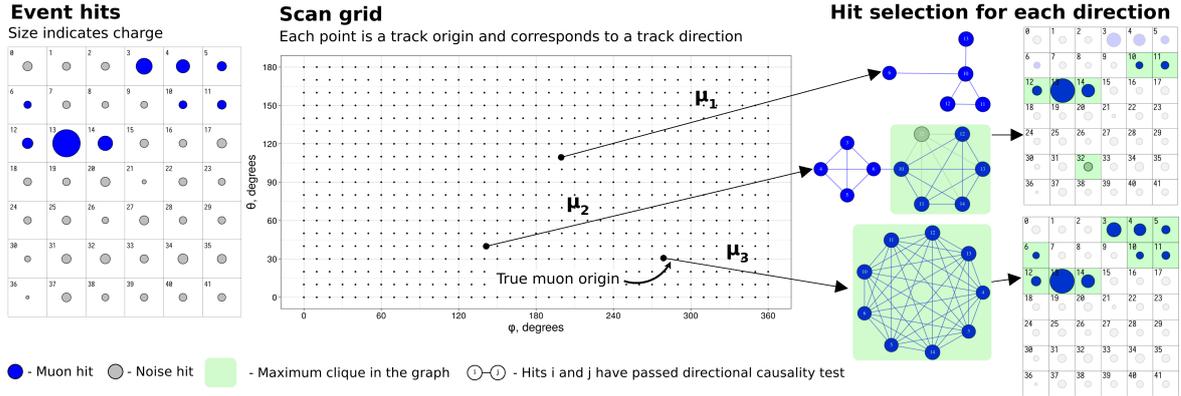
Figure 1: Example of the scanning step for an upgoing muon event.

Next, we use Bron-Kerbosch search to find maximal cliques for all the graphs considered at the previous step. Here, minimum clique size is a tunable parameter and directions without maximal cliques above minimum clique size are rejected from further consideration. Minimum clique size is set to the minimum number of hits required for reconstruction and typically equals 8 hits. For each direction we then pick the largest cliques in the graph associated with it, called the maximum cliques. A direction may have multiple maximum cliques, but they all should be of the same size. All cliques beside the maximum cliques are rejected from further consideration.

In this approach we are using the Bron-Kerbosch search that produces all maximal cliques in the graph rather then a faster equivalent that would only search for maximum cliques. The reasoning behind this decision is that the number of maximal cliques associated with a direction can inform the assumptions about topology of the event (such as the number of muons and showers) that may be useful for further steps in the muon track reconstruction pipeline.

At the end of this step, each scan grid direction should be associated with a muon track, a causality graph and a set of cliques of equal size. The example of the scanning step for 3 considered track directions of an upgoing muon event with the minimal clique size set to 6 is presented on Figure 1. Three scan grid directions are considered. μ_1 produces a causality graph with the maximum clique of size 3 and is ignored. μ_2 produces a graph with two maximal cliques and only the largest one is considered. μ_3 is close to the true track direction and produces a graph with only one clique that contains all signal hits. Finally, two cliques pass to the clique selection stage: a single clique from μ_3 and the largest clique from μ_2.

3.3 Clique selection

Once the maximum cliques for each direction in the scan grid are found, we determine the largest clique size for the event N_{max} and reject all cliques with sizes below $N_{max} - 1$. This step typically rejects the majority of directions in the grid, as they would no longer have a single clique associated with them. For each of the remaining cliques, we fit the associated tracks on the space of X_0, Y_0 and T_0 to a single track model. This is accomplished with a hit time based M-estimator,

(a) Atmospheric neutrino sample.

(b) High energy (E^{-2}) neutrino sample.

(c) Multicluster atmospheric neutrino sample
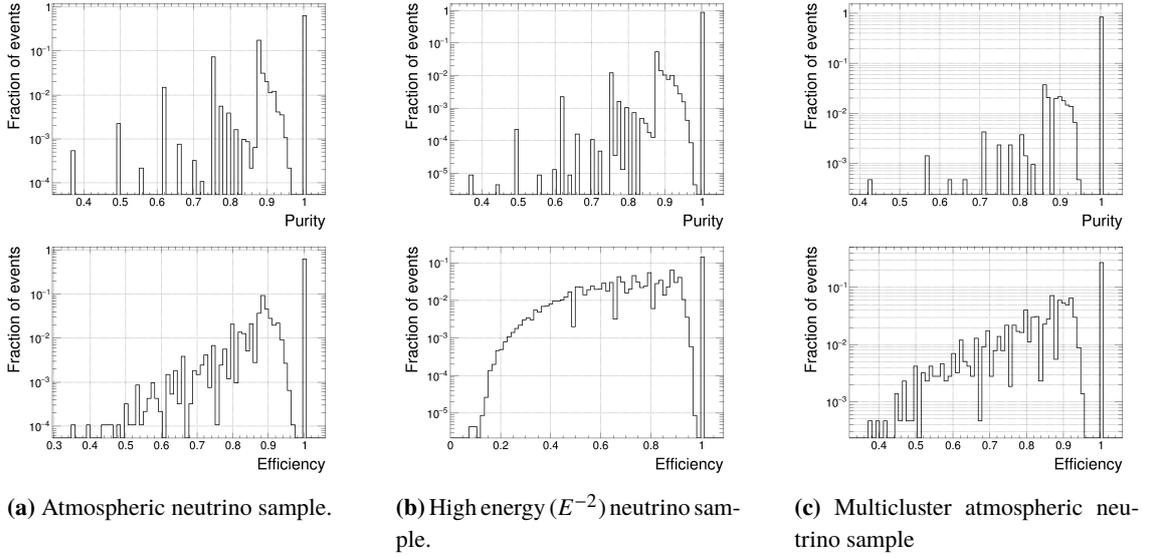
Figure 2: Hit-wise purity and efficiency distributions for various samples.

to mitigate effect of the outliers:

$$M = \frac{1}{N-2} \sum_i^N ln(1 + \frac{(T_0 + t_i^{th}(X_0, Y_0) - t_i)^2}{2\sigma^2})$$

If the M-estimator minimization does not converge, the corresponding clique is rejected.

Following this step, each remaining clique is associated with a fully-defined track and its fit quality represented by the M-estimator score. Finally, we select the clique with the best M-estimator value. The hits of this clique are considered to be the signal hits of the event and the associated track can be used as a prefit for a more comprehensive reconstruction technique.

4. Performance and results

The performance of the presented appoach was studied on three samples. The first sample consists of single cluster atmospheric muon neutrinos, simulated with Bartol flux and the low threshold at 20 GeV. The second sample is equivalent to the first one, except it was simulated with E^{-2} spectrum. Finally, the third sample is equivalent to the first one, however it only contains multicluster events. All samples correspond to a low noise period in season 2019 characterized by PMT noise counting rates of ∼ 20 kHz for all elevations. The samples have been processed with a rectangular scan grid with a step of 3 degrees in both θ and ϕ, minimum clique size of 8 hits, and a directional causality window set to 20 ns.

The key estimated hit selection quality characerics for each sample in this performance study are purity and efficiency. The purity of hit selection for an event (also known as precision, or positive predictive value) is defined as the ratio of the number of correctly identified signal hits to the total number of hits identified as a signal. The efficiency of hit selection for an event (also known as recall, or true positive rate) is defined as the ratio of the number of correctly identified

signal hits to the total number of signal hits in the event. The distributions of purity and efficiency for the processed samples are presented on Figure 2.

The average values of purity and efficiency for each sample are presented in Table 1. Note that compared to the hit selection procedure used in [6], the efficiency for atmospheric neutrino events has almost doubled. Lower efficiency for the high energy and multicluster samples is explained by the increased fraction of the delayed cascade hits due to the higher energies of observed particles. The directional causality uses the model of a single muon and so some cascade hits are excluded from cliques, decreasing overall efficiency of hit selection. Importantly, cascade hits that fit the track model within the timing uncertainty can be included in cliques along with the muon hits. However, if we restrict the samples to events with at least 8 track hits, efficiency improves.

Table 1: Hit selection quality characterics for various samples

Sample	Average purity	Average efficiency
Atmospheric neutrino sample (hit selection from [6])	0.99	0.5
Atmospheric neutrino sample	0.95	0.95
High energy (E^{-2}) neutrino sample	0.98	0.74
High energy (E^{-2}) neutrino sample (at least 8 track hits)	0.98	0.89
Multicluster atmospheric neutrino sample	0.98	0.87
Multicluster atmospheric neutrino sample (at least 8 track hits)	0.98	0.9

The examples of a scan for two events are presented on Figure 3. Note that all the scan grid directions that passed clique selection are below the horizon for the upgoing event and above the horizon for the downgoing event. As can be seen on Figure 4, this remains the case for most of the vertices in the case of upgoing muons from atmospheric neutrinos and downgoing bundles, allowing for improved rejection of the atmospheric muon background in further analysis.

5. Conclusion

We have developed an efficient hit selection technique for suppressing noise hits in Baikal-GVD based on the approach presented in [5]. Our approach alters key steps in the previous

(a) Upgoing muon from atmospheric neutrino

(b) Downgoing atmospheric muon bundle
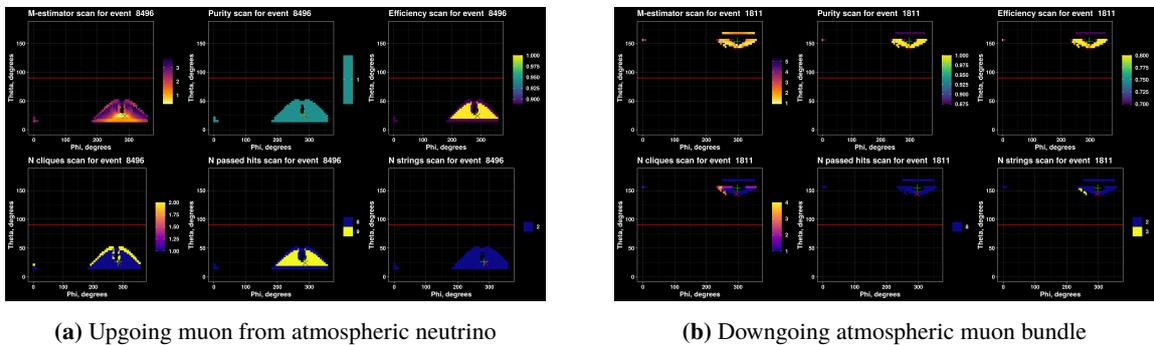
Figure 3: Event scan examples. Each pixel corresponds to a scanned direction that passed the clique selection. Green cross corresponds to the true event origin, red cross to the selected direction.
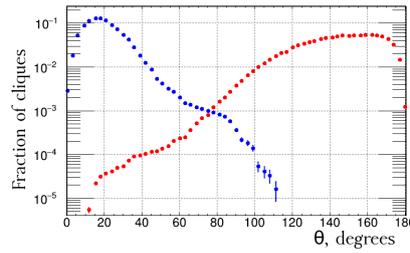
Figure 4: Distribution of cliques over the polar angle for an atmospheric neutrino (blue) and atmospheric muon bundle (red) samples.

work by considering pairwise causal relationships between PMT hits as an undirected graph and assuming that signal hits should form fully connected subgraphs (cliques) in it. The new approach has demonstrated a 95% purity and 95% efficiency for an atmospheric neutrino sample, showing almost two-fold inrease in efficiency compared to the hit selection algorithm previously used in Baikal-GVD. The algorithm has shown comparable results for multicluster and high energy neutrino samples (see Table 1). Besides noise suppression, this approach yields useful information about event configuration which can be used further along the processing pipeline. In Baikal-GVD this approach is used in the latest version of the muon reconstruction, as described in [8].

6. Acknowledgements

References

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

[2] V. A. Allakhverdyan et al. Neutrino telescope in lake Baikal: Present and nearest future. PoS *ICRC2021 (these proceedings) 002*.

[3] A. D. Avrorin et al. The optical noise monitoring systems of the Lake Baikal environment for the Baikal-GVD telescope. *PoS*, ICRC2019:875, 2021.

[4] V. A. Allakhverdyan et al. The Baikal-GVD neutrino telescope as an instrument for studying Baikal water luminescence. PoS *ICRC2021 (these proceedings) 400*.

[5] Ronald Bruijn et al. *The Antares Neutrino Telescope: Performance studies and analysis of first data*. Universiteit van Amsterdam, 2008.

[6] V. A. Allakhverdyan et al. Measuring muon tracks in Baikal-GVD using a fast reconstruction algorithm. *arXiv preprint arXiv:2106.06288*, 2021.

[7] Coen Bron and Joep Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

[8] V. A. Allakhverdyan et al. Performance of the muon track reconstruction with the Baikal-GVD neutrino telescope. PoS *ICRC2021 (these proceedings) 1080.*