Anatomy of large-volume neutrino telescopes

Grigory Safronov (INR RAS, JINR)

17th Baksan school on particle astrophysics
April 3-11 2025, Terskol

Lecture 3

Event reconstruction and selection

Event reconstruction and analysis

Events are read-out and built
Detector is positioned and calibrated

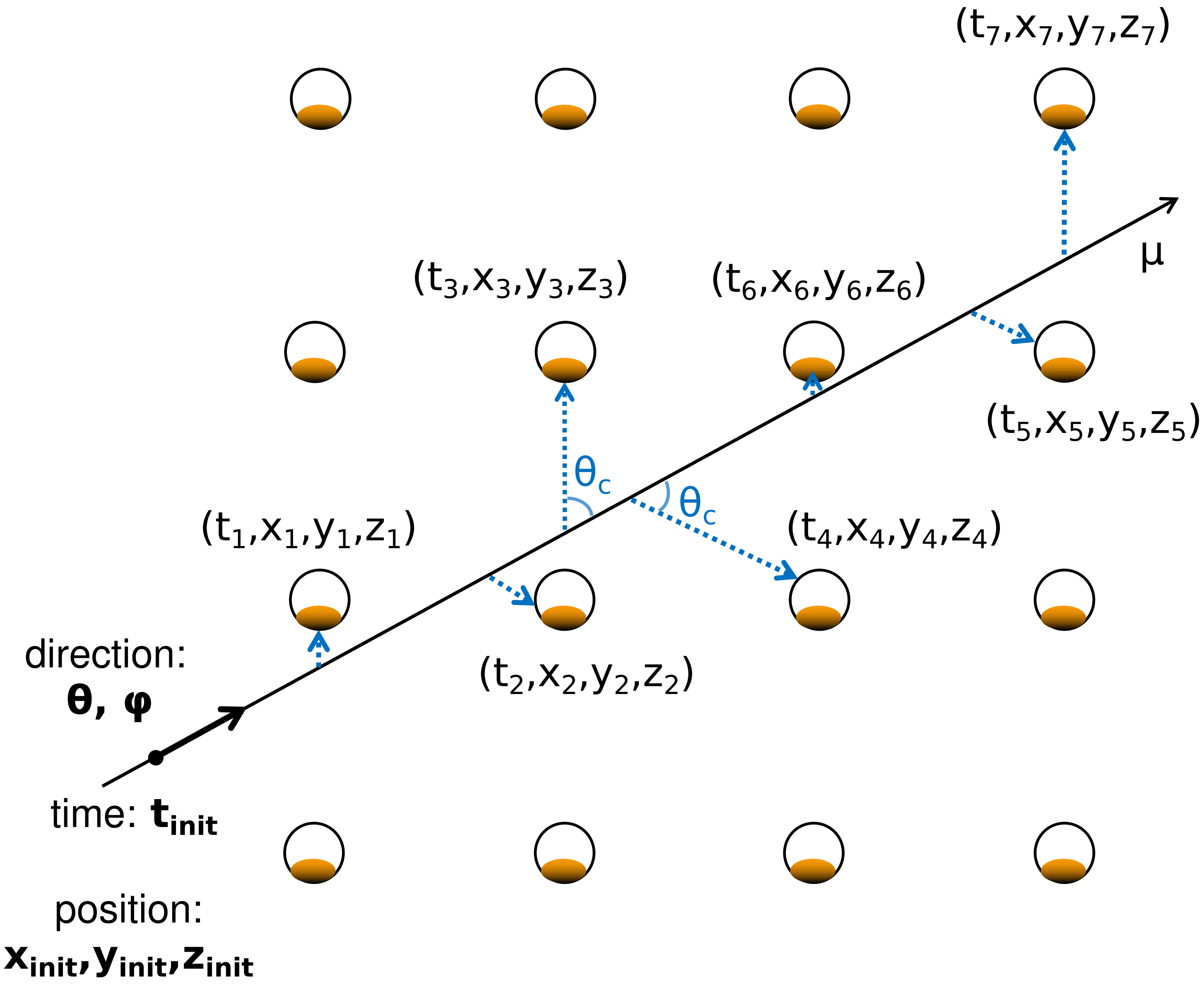
We can associate OM pulses (**hits**) in different sections of the detector with some event hypothesis

Focus on tracks here

Simplest MIP track hypothesis has 5 free parameters

Time of each hit can be unambigously calculated using track hypothesis

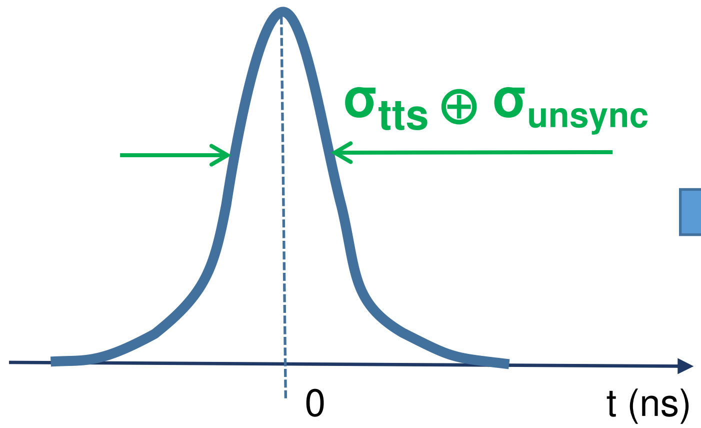
Fit the data to the track hypothesis



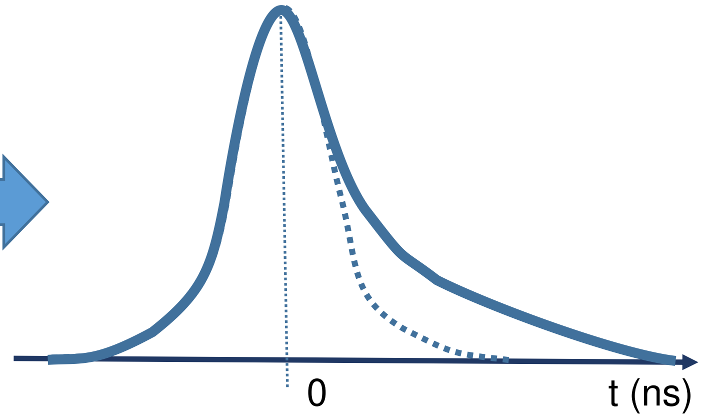
Track models

Different track models can be used depending on importance of effects

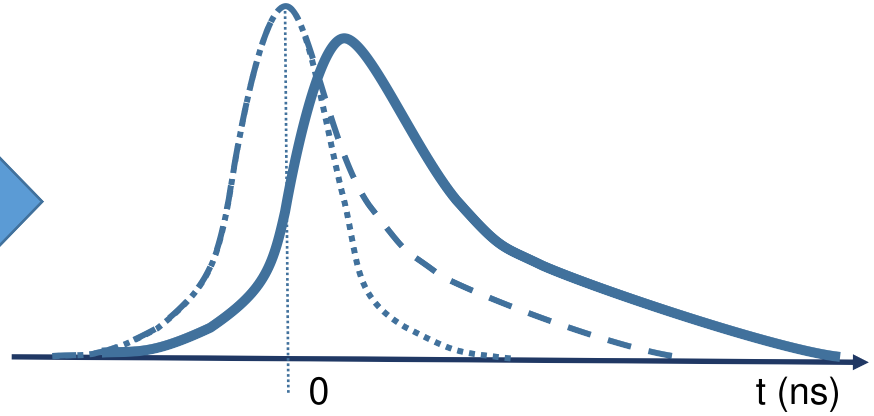
MIP muon



MIP muon + showers

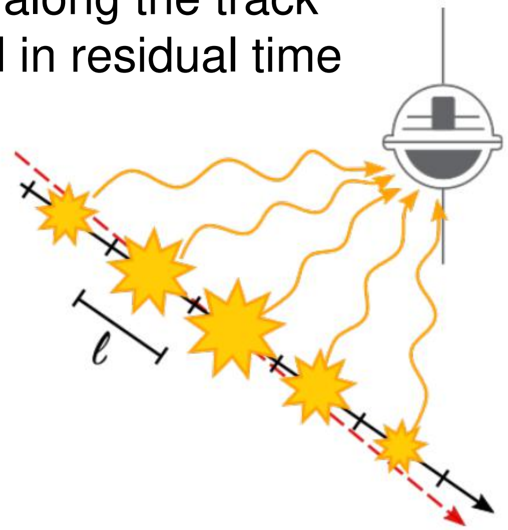


MIP muon + showers + scattering



Simplest model:
Time chi2 fit with sigma defined by the time measurement precision

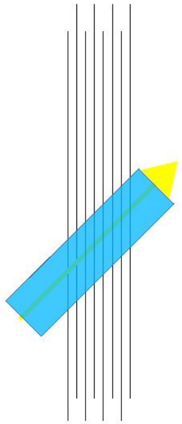
Showers along the track cause tail in residual time



Showers and scattering require complex likelihood functions derived from MC

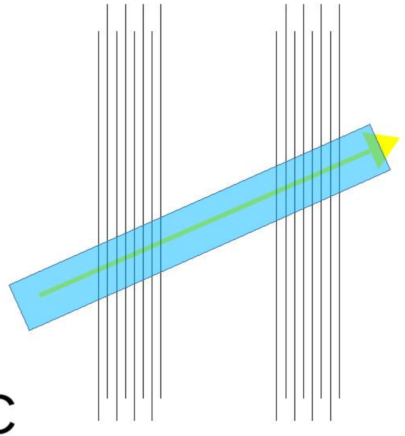
Baikal-GVD analysis pipelines

Single-cluster tracks



- Low energy threshold
- Optimal sensitivity to nearly vertical tracks
- 90% of recorded track events

Multi-cluster tracks

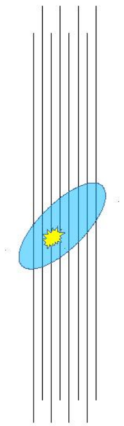


- Moderately low energy threshold
- Optimal sensitivity to inclined tracks
- 10% of recorded track events

ν_μ CC

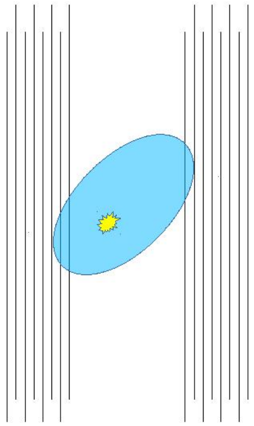
NC, ν_e ν_τ CC

Single-cluster cascades



- High energy threshold
- Good energy resolution
- Relatively rare events

Multi-cluster cascades



- Very high energy threshold
- Excellent energy resolution
- Very rare events

Track reconstruction in Baikal-GVD

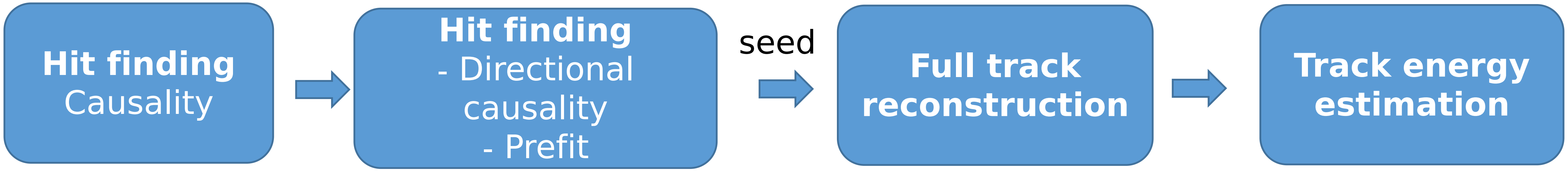
Due to high noise levels and substantial absorption the most challeging task is to collect weak Cerenkov signal among noise hits

- Crucial for efficient detection at ~TeV energies

For the track reconstruction simplest track model is used

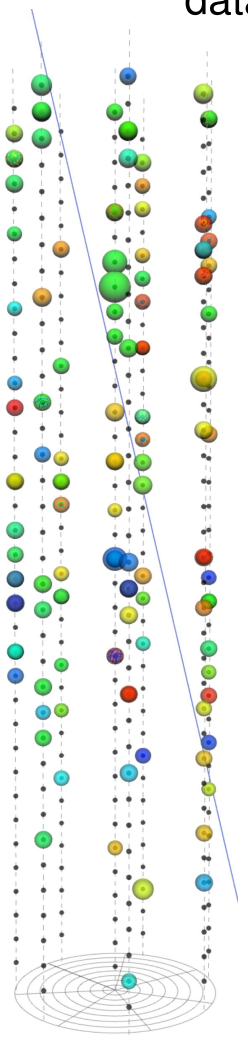
- Neglect scattering and showering

Track reconstruction chain in in Baikal-GVD

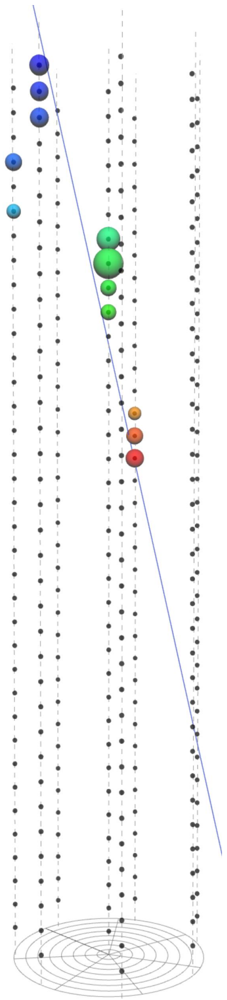


Track reconstruction in Baikal-GVD

track-like event,
data 2019



full event with noise



event after the hit selection

From previuous lecture, rough estimate of response to 10 TeV muon

distance (m)	light field (γ/cm²)	photoelectron yield	non-zero p.e. prob. (%)
10	0.226	18.3	~100
22	0.014	1.13	67.7
30	0.0027	0.22	19.7

Response at lower energies is sparse, low-charge and fluctuating

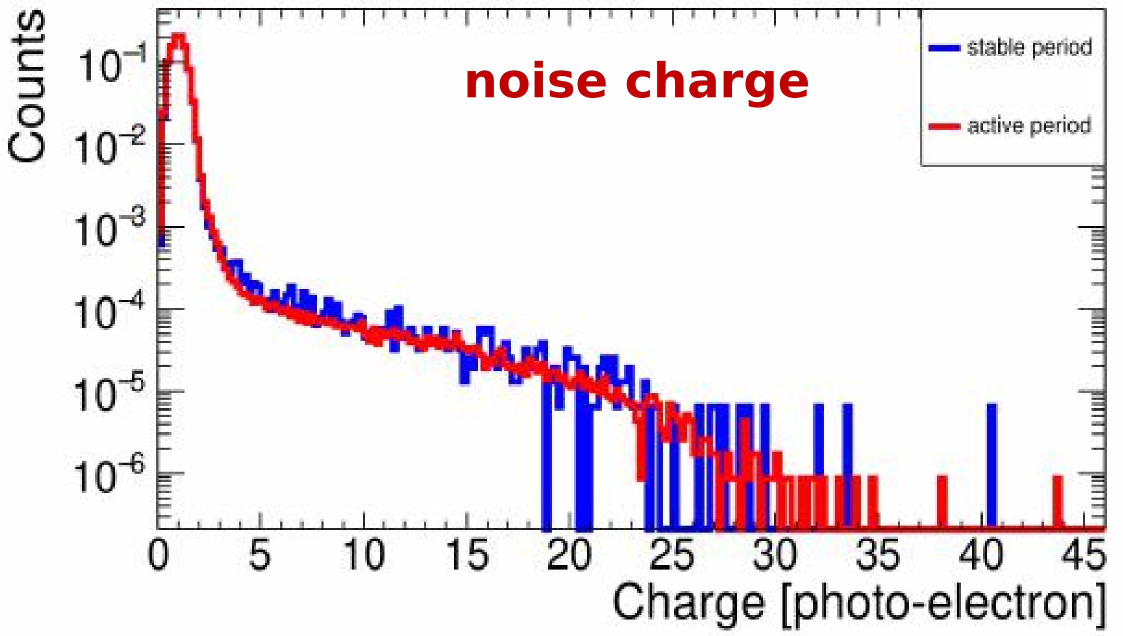
Suppress noise hits at the minimum cost for Cerenkov hits

- Scanning algorithm **[PoS(ICRC2021)1063]**
- Multicluster algorithm
- Deep learning -based algorithms **[I. Kharuk et al 2023 JINST 18 P09026]**

Hit finding: causality criterium

The noise charge distribution is at the level of 1 p.e.

- Simplest hit selection technique: cut on the hit charge
- Baikal-GVD cascade analysis: q > 1.5 p.e.



Charge cut would affect the efficiency of low-energy (~ 1TeV) track reconstruction

Signal hits are correlated in time while noise hits are not

Causality criterium: $\Delta t_{ij} < \Delta R_{ij} \frac{c}{n} + \delta$

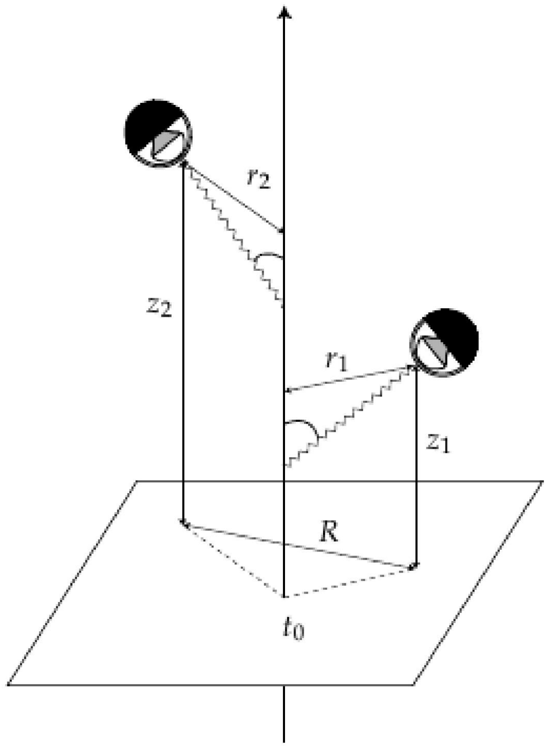
Where δ ~ 10 ns, accounts for time measurement precision

Event purity of ~70-80% is achieved keeping efficiency at >99%

- For atmospheric neutrino spectrum, E_{median} ~ 500 GeV

For good reconstruction precision we need to further purify signal hits

Hit finding: directional causality



$$t_j = t_0 + \frac{1}{c}(z_j + kr_j)$$

$$c(t_j - t_i) - (z_j - z_i) = k(r_j - r_i)$$

$$|r_j - r_i| < R$$

R.Bruijin, 2008, ANTARES

More strict causality condition for predefined track direction

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

Perform scan on (θ,φ) - find the set of largest cliques of causally-connected hits

- Graph theory algorithms, e.g. Bron-Kerbosch clique search algorithm

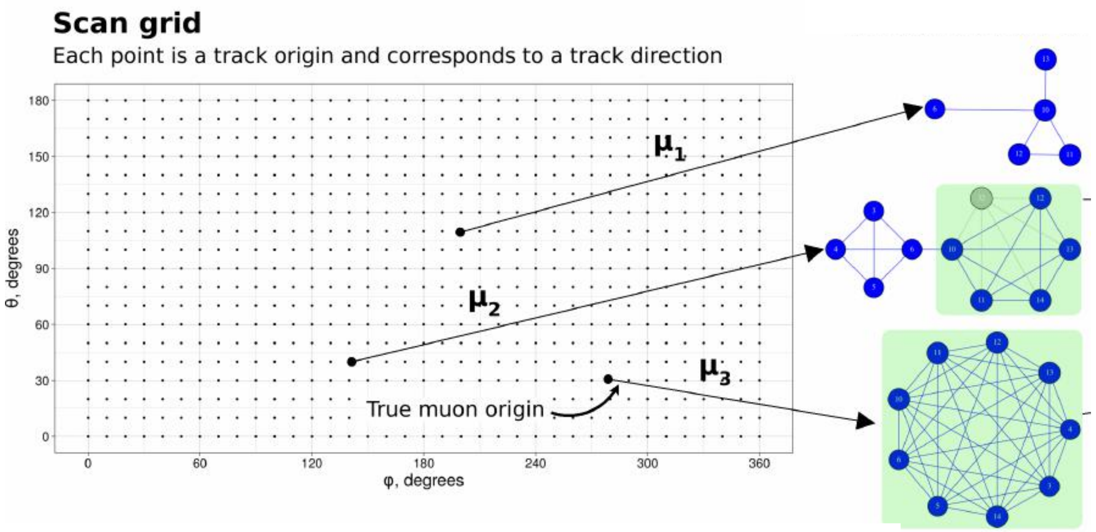
Few largest hit cliques are fit with fixed direction to find optimal track position

For atmospheric neutrino spectrum (E_{median} ~ 500 GeV):

- Purity ~ 95%
- Hit selection efficiency ~95%

Few best hit collections are passed to full-scale muon reconstruction

- At least 8 hits on 2 strings
- Track direction and position are used as seeds



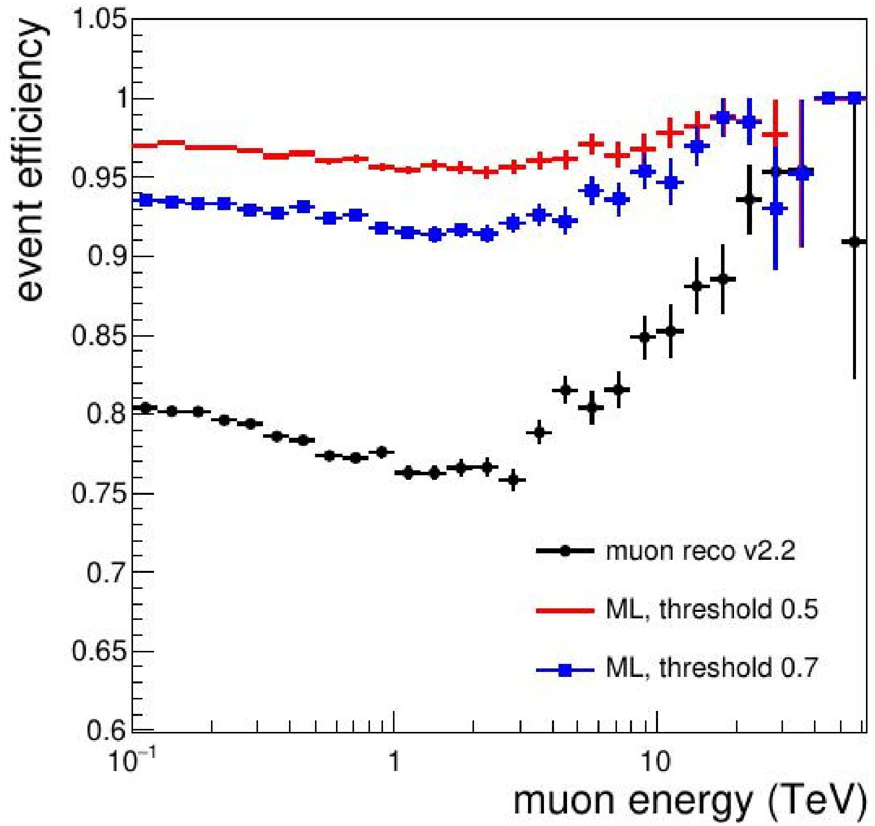
Hit finding with neural networks

Noise suppression algorithm based on deep learning: **I. Kharuk et al 2023 JINST 18 P09026**

Neural networks allow to impove detection efficiency at low energies, when we look for smaller number of low-charge hits hidden in noise

Application of ML in other areas of event reconstruction is being developed

Efficiency of the detection of event with at least 8 signal hits on 2 strings



Full track reconstruction at Baikal-GVD

Full track reconstruction

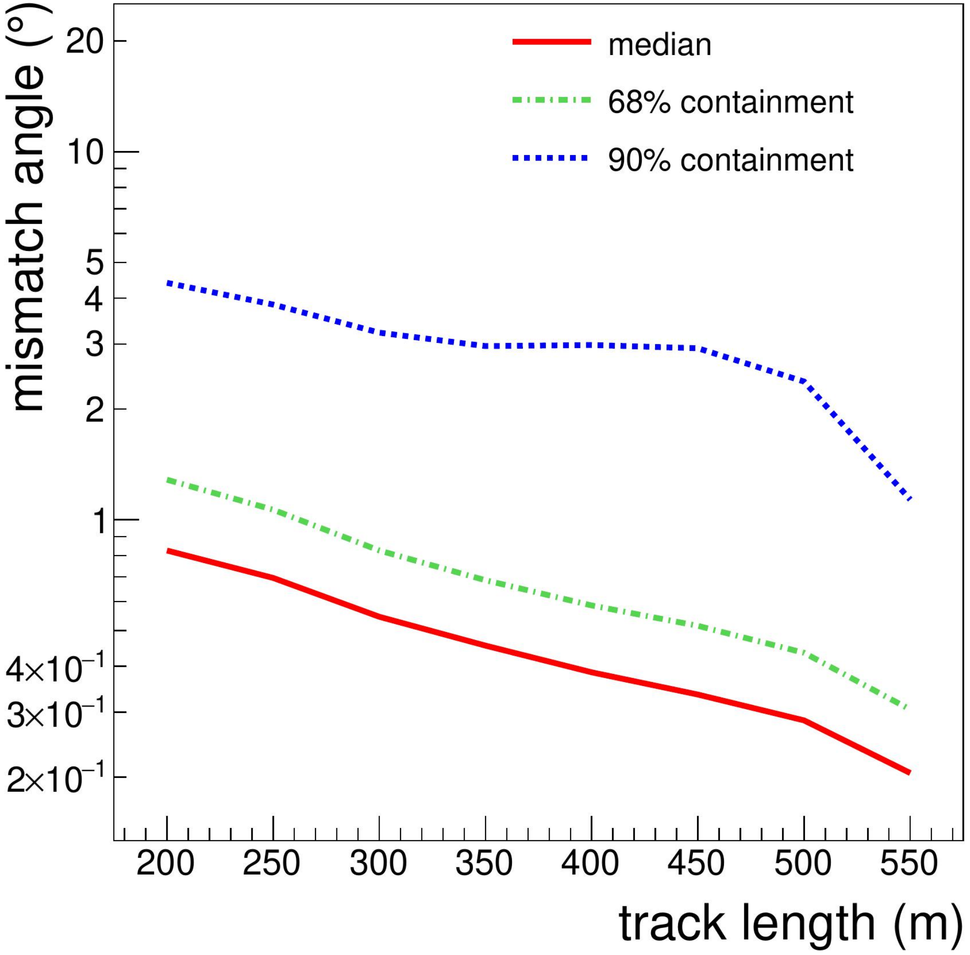
- For each seed direction few iterations are performed with gradual hit collection optimisation
- Minimisation of the loss function with time and charge parts

$$Q = \sum \left(\frac{(t_i - t_i^{th})^2}{\sigma^2} + 0.3 \frac{(N_{hits} - 6)}{q_{sum}} \frac{a_0 q_i \sqrt{d_1^2 + d_i^2}}{\sqrt{a_0^2 + q_i^2}} \right)$$

time χ^2 in assumption of MIP muon model

~Q*R penalty

Full track reconstruction at Baikal-GVD



Full track reconstruction

- For each seed direction few iterations are performed with gradual hit collection optimisation
- Minimisation of the loss function with time and charge parts

$$Q = \sum \left(\frac{(t_i - t_i^{th})^2}{\sigma^2} + 0.3 \frac{(N_{hits} - 6)}{q_{sum}} \frac{a_0 q_i \sqrt{d_1^2 + d_i^2}}{\sqrt{a_0^2 + q_i^2}} \right)$$

time χ^2 in assumption of MIP muon model

~Q*R penalty

Angular resolution

- For tracks with length ~ 500m median resolution is 0.2-0.3°
- Median resolution for short tracks ~150m: ~ 1°

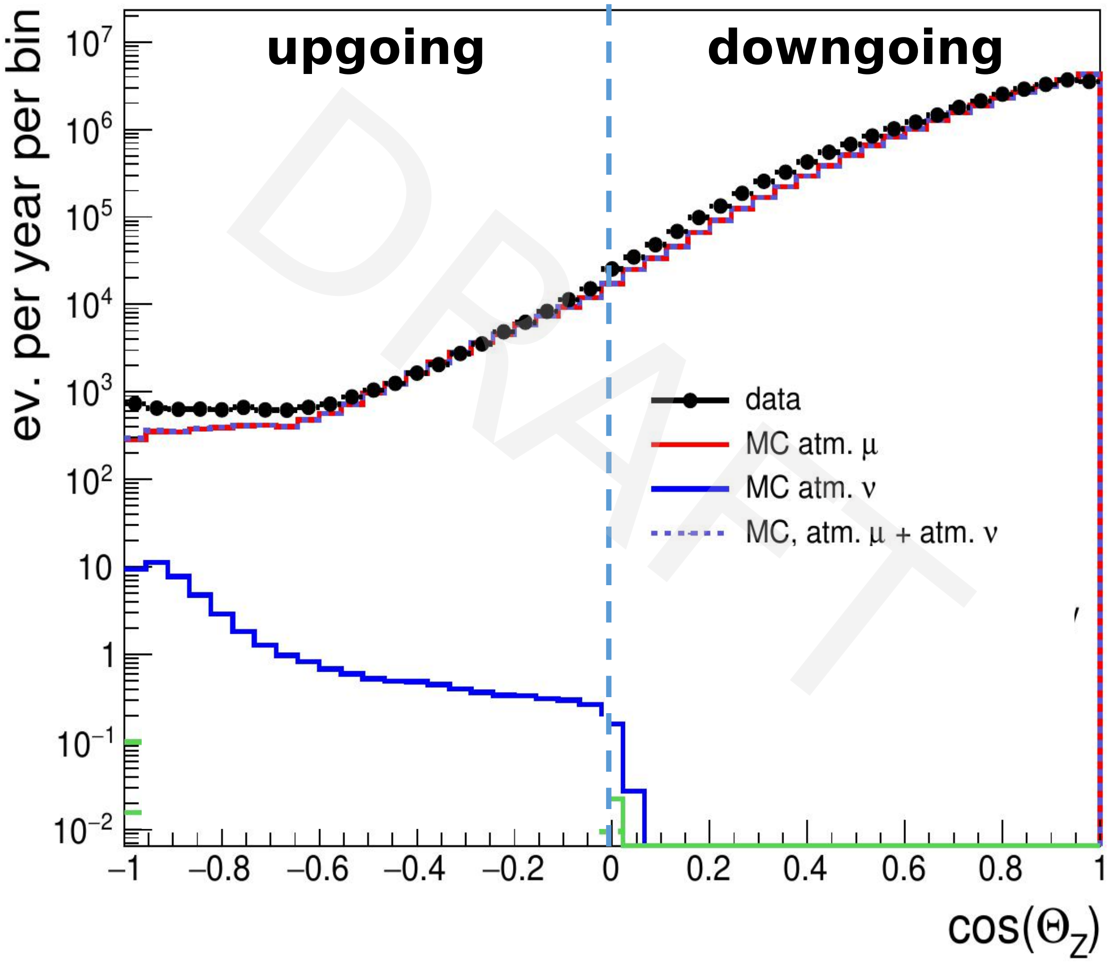
Neutrino event selection

Reconstructed tracks are dominated by atmospheric muon background

On average ~3 events per second

Background rate in the region of upgoing events up to ~10000 times larger than neutrino rate

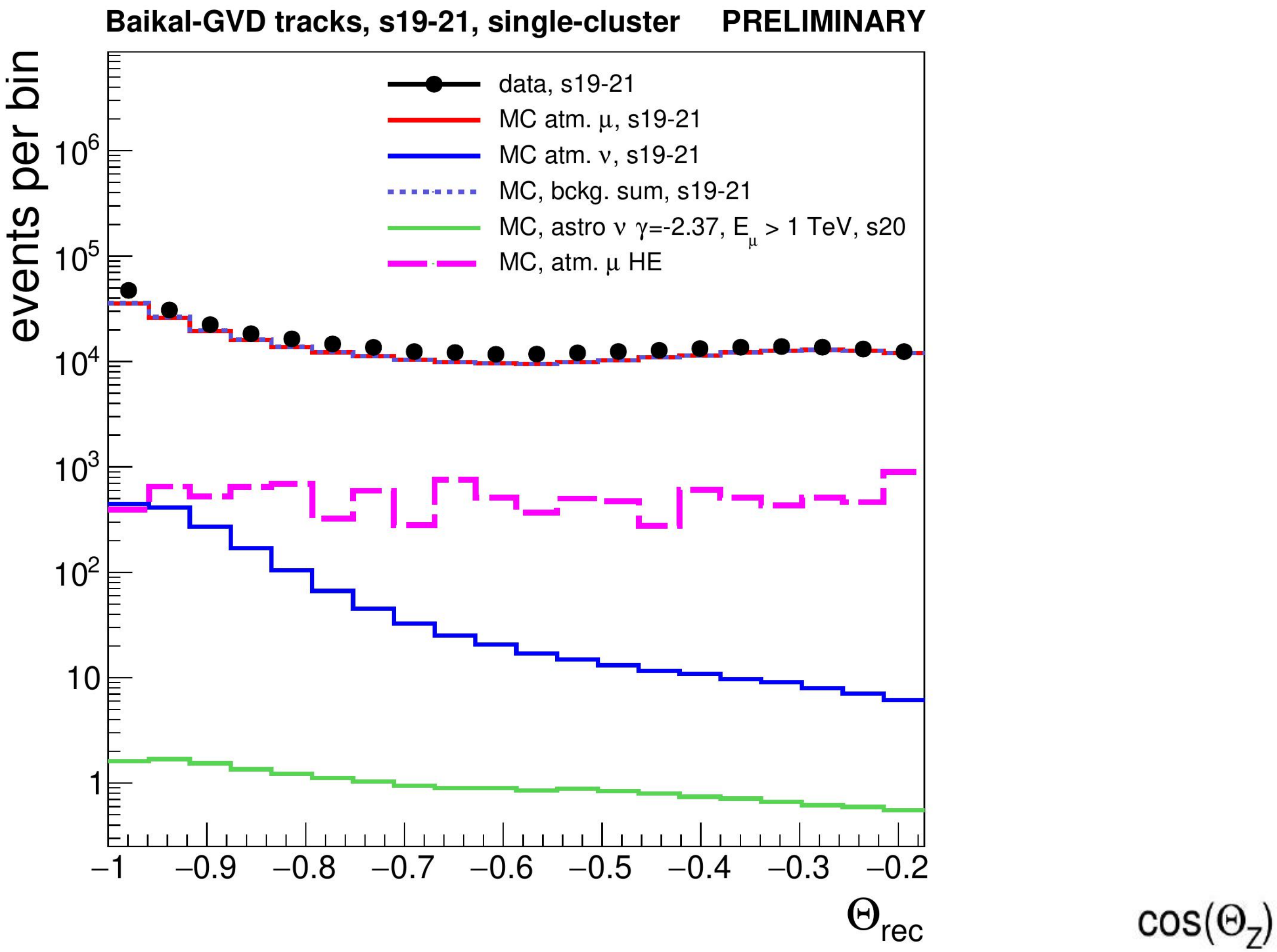
Zenith angle of reconstructed tracks



Track-like event selection

The most convinient region for single-cluster track analysis: upgoing tracks θ > 100°

Upgoing events (θ>100°)
before neutrino selection



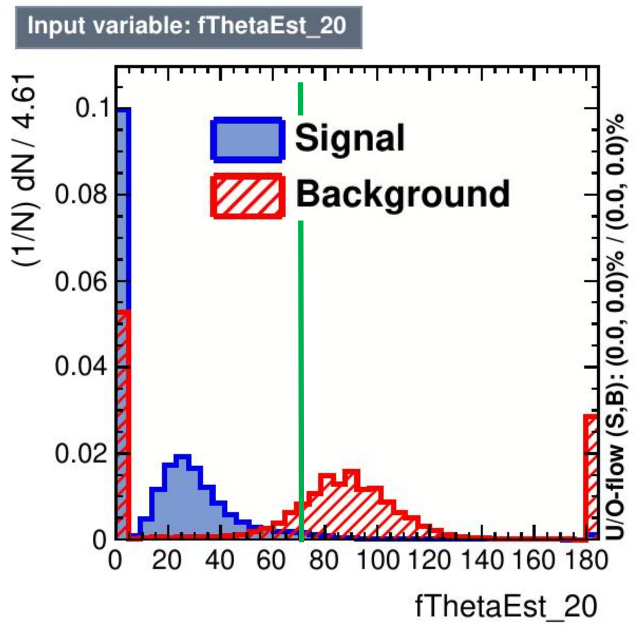
Track-like event selection

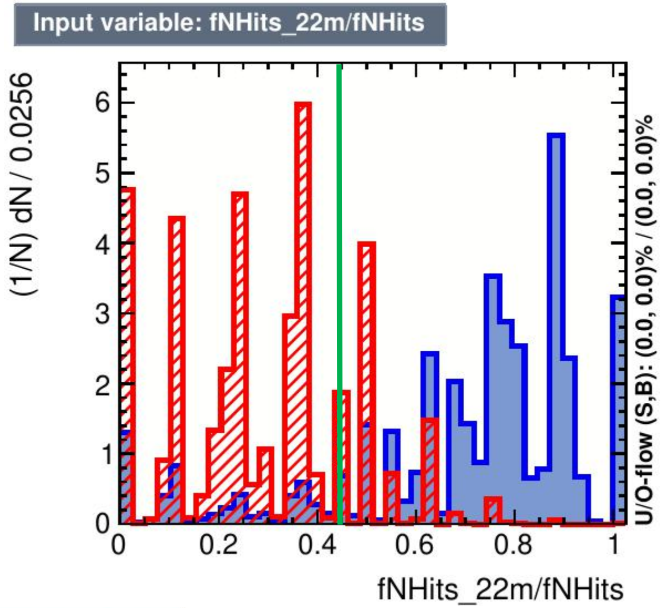
Rejection of badly reconstructed tracks is performed with cuts on various quality variables

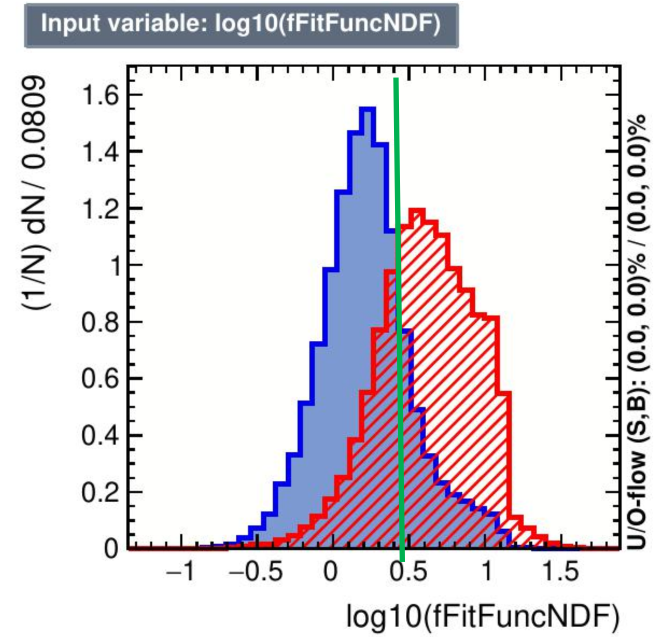
One needs to suppress background 10^3 time larger than signal

Often it is not possible to design cuts on limited set of variables with enough rejection power and high signal efficiency

In this case Boosted Decision Trees (BDT) are used





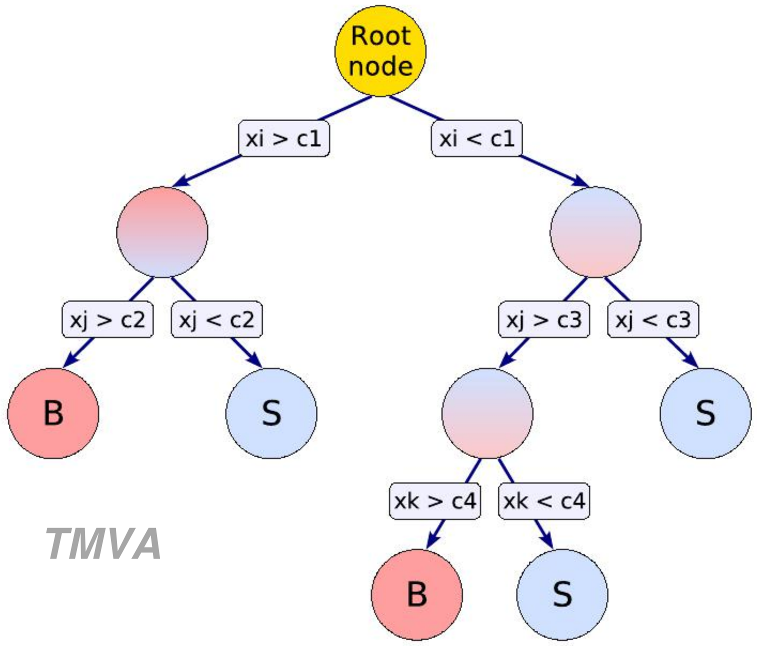


Boosted decision trees (BDT)

Active usage in HEP since mid-2000s
- Icreased signal/background (S/B) separation wrt usual cuts analysis
- Very good out-of-the box performance

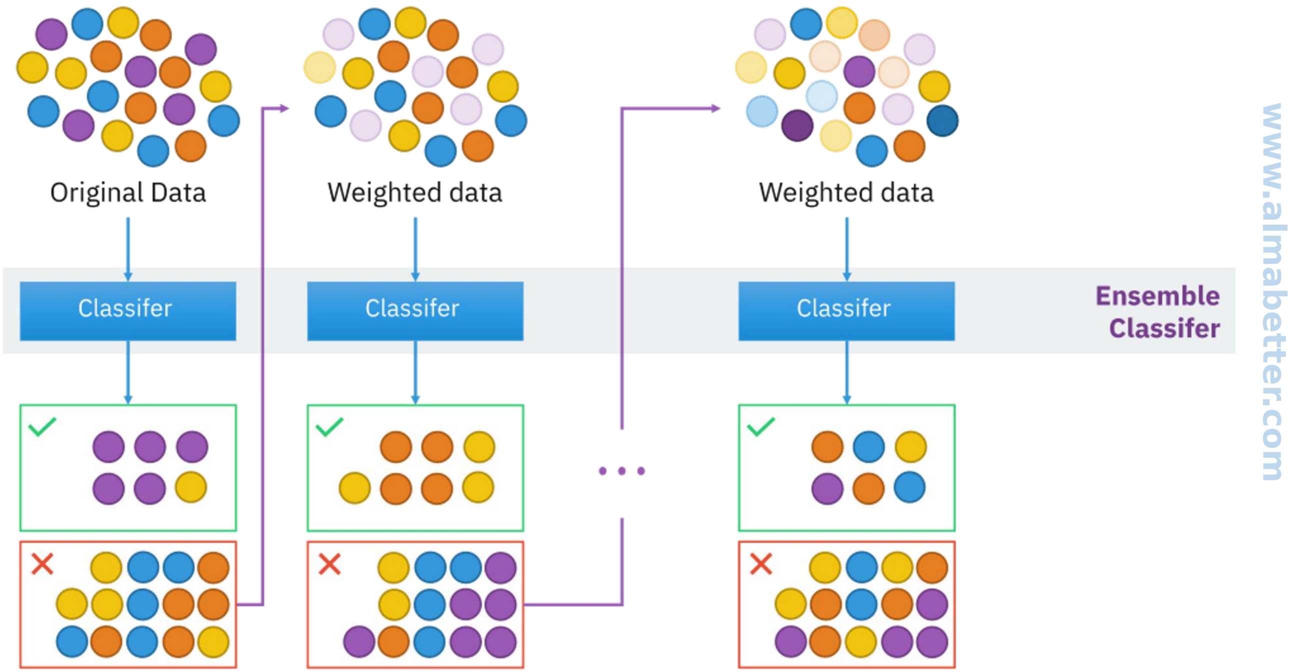
Elementary decision tree

Variable x* at each step is chosen to give the best S/B separation power



Decision tree response:
-1 for bckg.; +1 for noise

Boosting: build sequental set of weighted trees trained on weighted event samples
- Events missclassified on previous step gain increased weight
- BDT response (score): normalised sum of weighted resposes



Suggested reading on BDTs in HEP: **Yann Cadou, arXiv:2206.09645**

Track-like event selection

Classification with BDTs with 20 weakly correlated variables

- Variables of reconstruction quality
- Variables of event topology

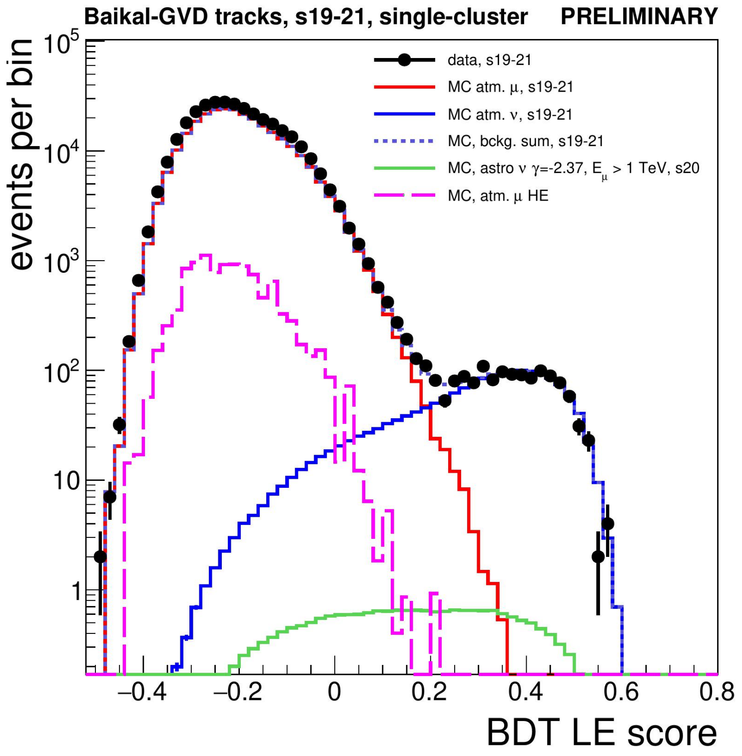
Two classifiers

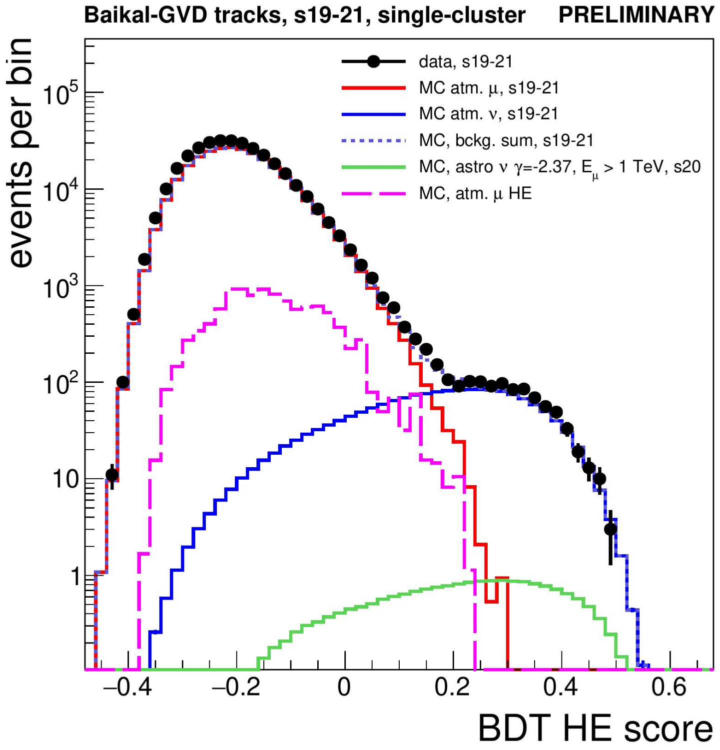
- Low-energy BDT (BDT_LE), E_μ<10 TeV, atmospheric neutrino spectrum
- High-energy BDT (BDT_HE): E_μ>10 TeV, astrophysical neutrino spectrum ν ~E^{-2}

Backgrounds used in training

- CORSIKA muon bundles natural spectrum
- CORSIKA muon bundles, leading μ E_μ > 100 TeV

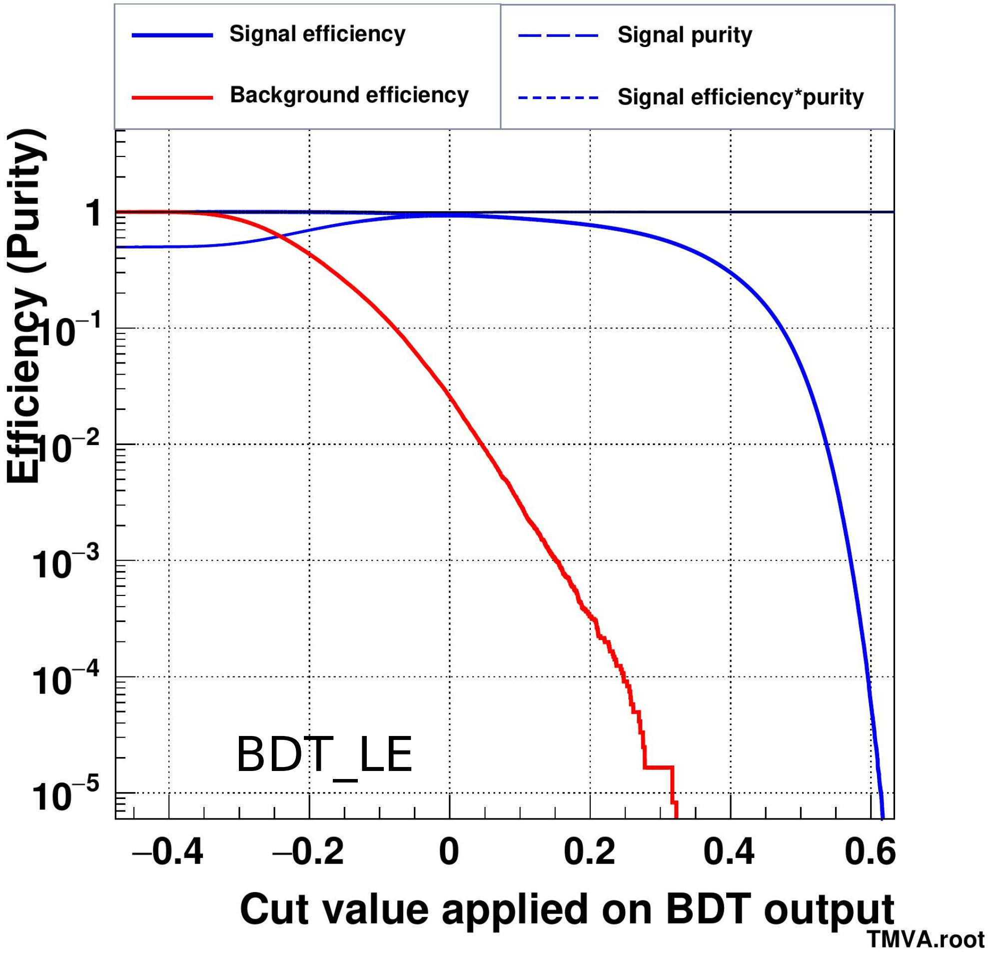
Important: data and MC match each other, that justifies the application of BDT





Track-like event selection

BDT signal and background efficiencies



Cut at 0.25:
Background reduction by the factor 10^4,
signal efficiency: ~70%

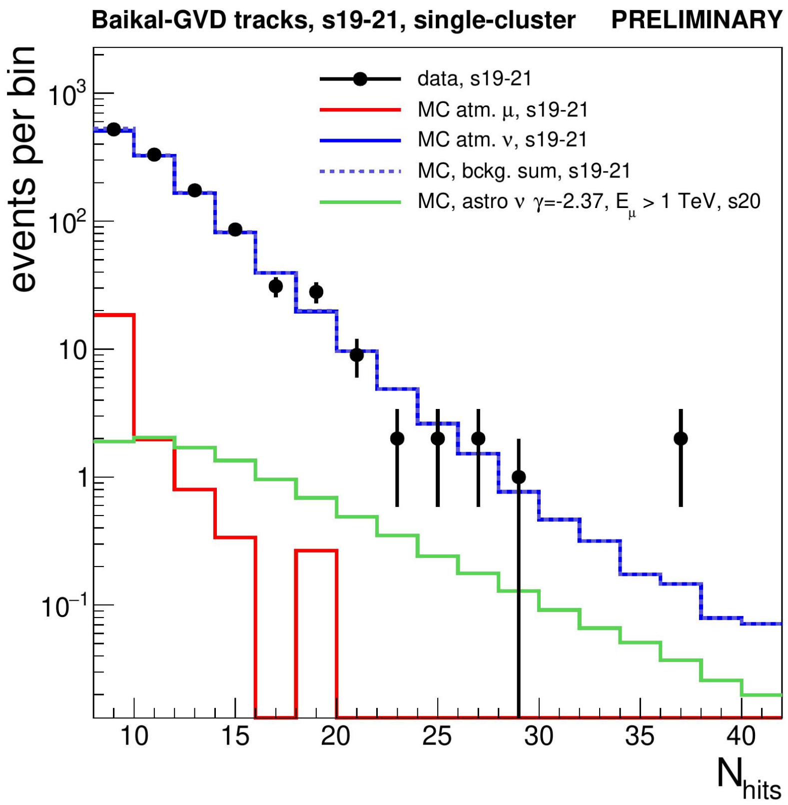
Neutrino candidates

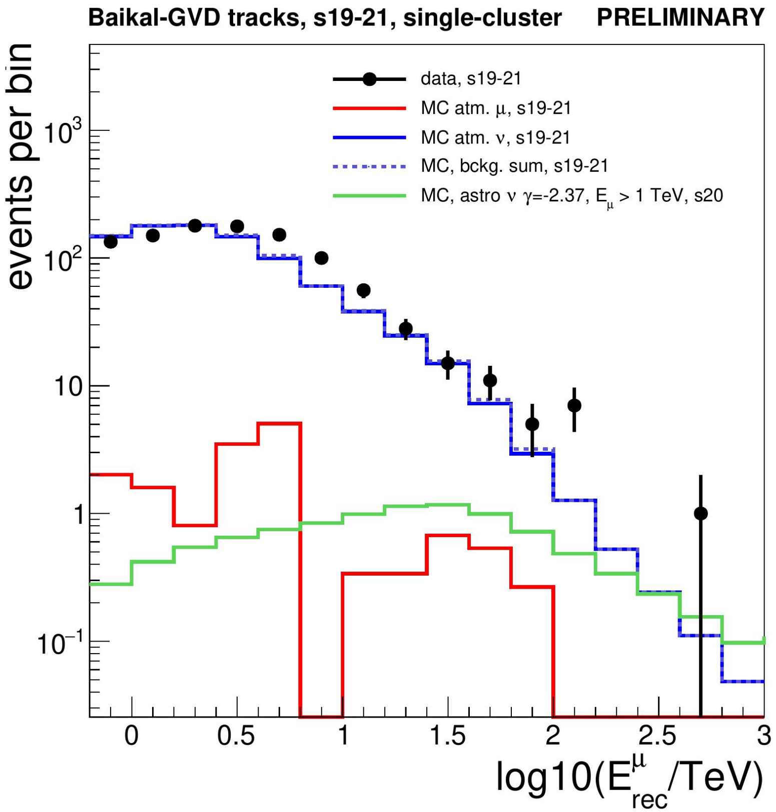
An illustration of cleaned track-like neutrino candidate sample

- A cut on BDT: BDT_LE>0.25|| BDT_HE>0.25
- Cut is not optimized, it could change depending on analysis: time-integrated PS search or time-dependent search, energy threshold etc..

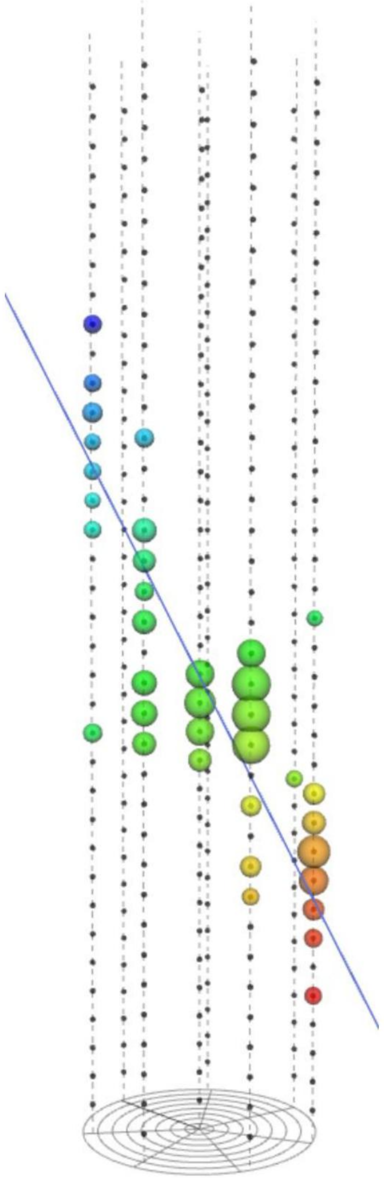
Sample of track-like events selected in data-taking seasons 19-21

- 14.37 years of taking data in single-cluster configuration
- ~Half of processed data





Some interesting events s19-21



Season 2019, December
Cluster 3

N_{hits}	36
E^{μ}_{rec}	62.1 ТэВ
θ_{rec}	153.1°
L_{track}	332.4 м

Angular precision:

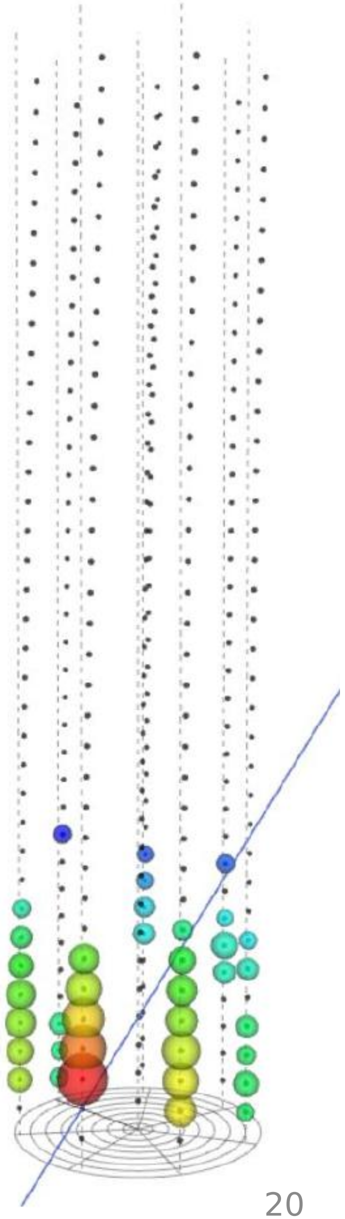
50%:	0.5°
68%:	0.7°
90%:	1.0°

Season 2020, September
Cluster 5

N_{hits}	37
E^{μ}_{rec}	107.2 ТэВ
θ_{rec}	116.7°
L_{track}	140.1 м

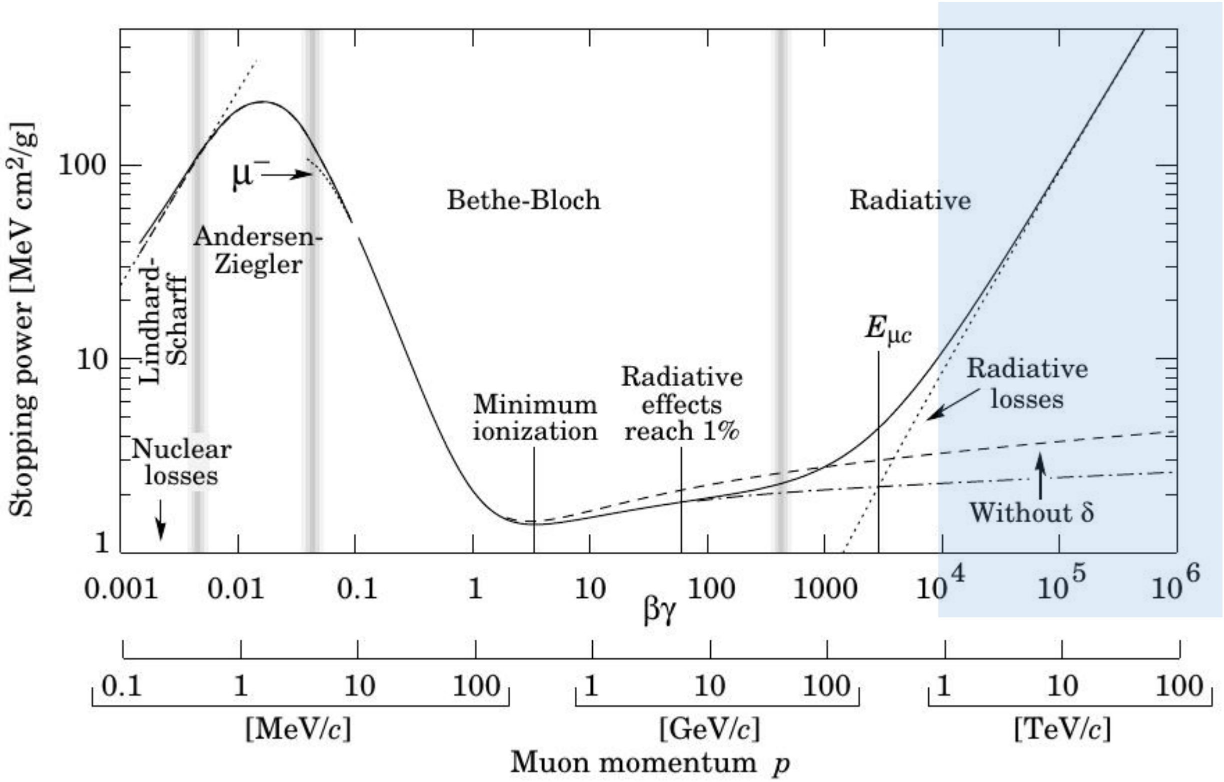
Angular precision:

50%:	0.7°
68%:	1.0°
90%:	1.5°



Muon energy reconstruction

PDG



Only part of the muon trajectory is observed in the detector

Muon energy can be estimated using energy losses (or light deposition) along the track

- 200-300% precision

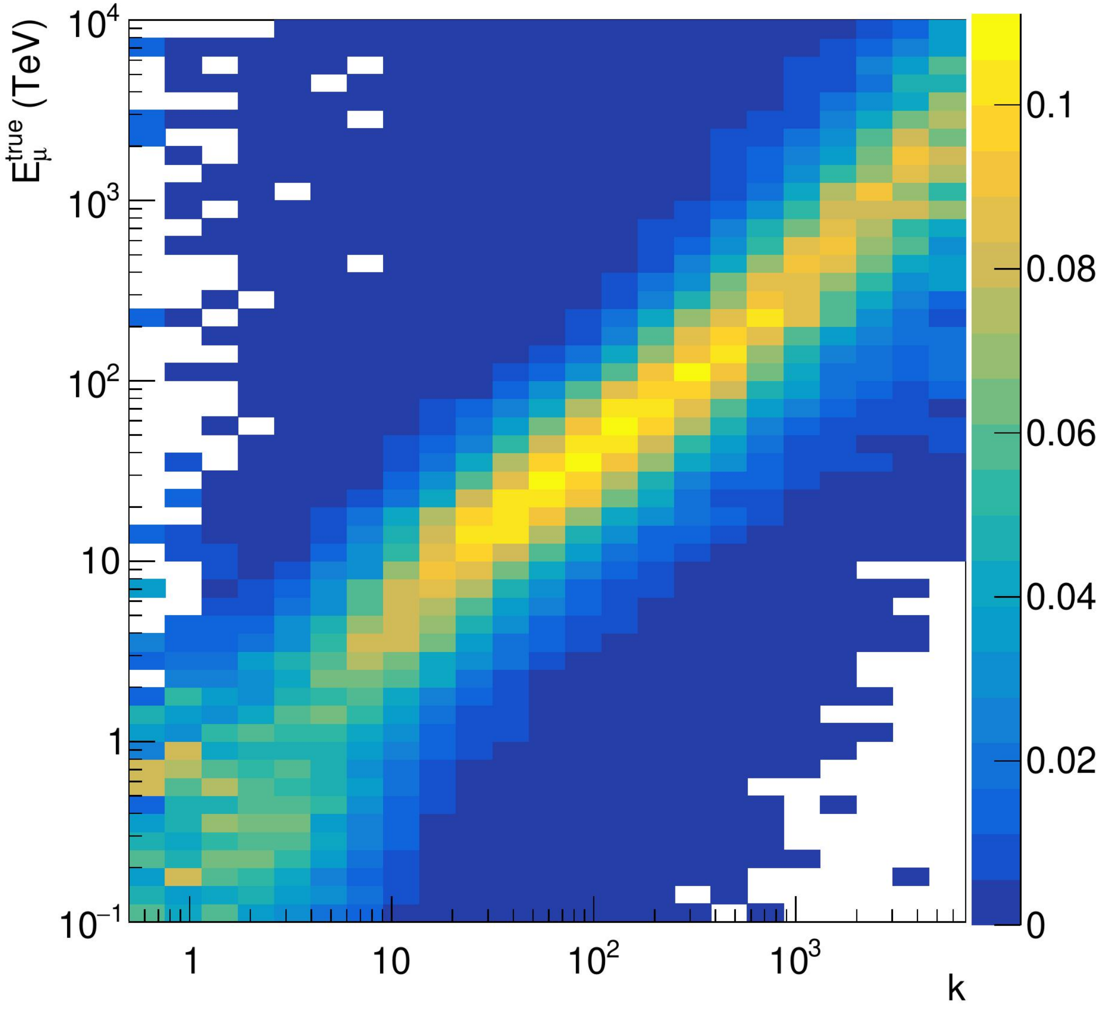
~ linear dependency of dE/dX on energy for E > 1 TeV

Muon energy estimation

Energy reconstruction is optimised for tracks passing the neutrino selection criteria

"Median" muon energy estimator:

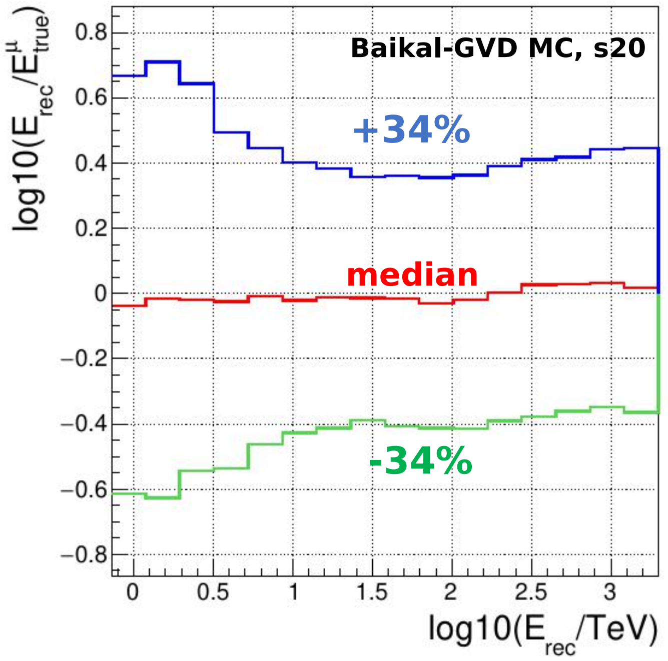
- Consider hits within 40m from the track
- Estimate number of photons emitted by the muon based on the hit charge
- Take median of these estimates along the track

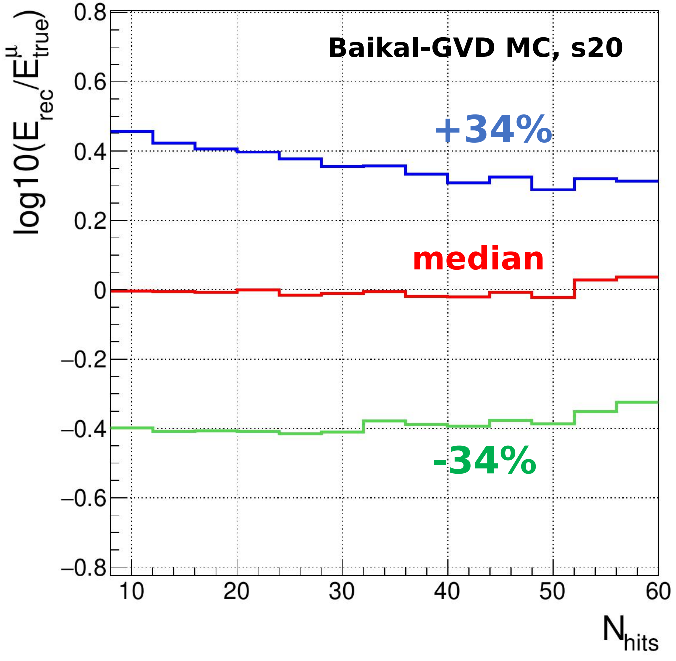


Muon energy estimation

68% confidence interval for E^{μ}_{true} :

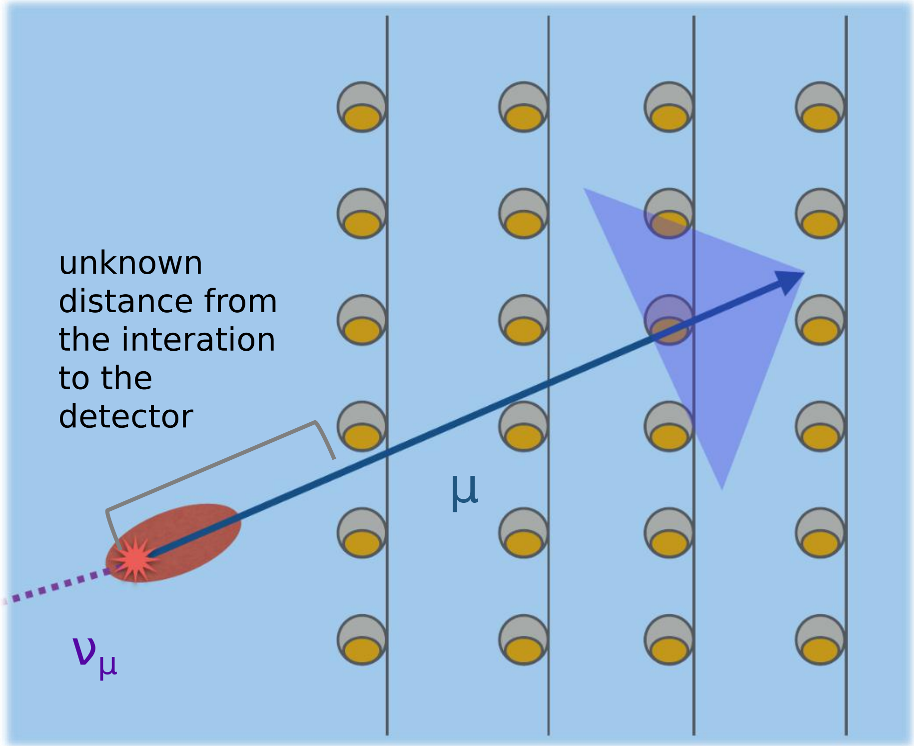
- ~ **(E_{rec}/2.5, 2.5×E_{rec})** in the range 10<E_{rec}<1000 (TeV)
- Resolution improves with N_{hits} increase





But what about the neutrino energy?

Neutrino energy reconstruction

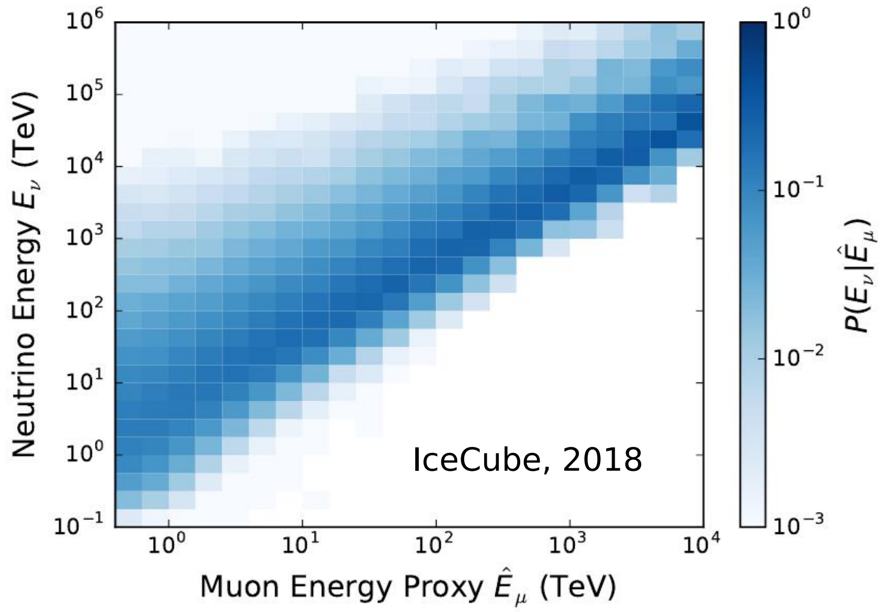


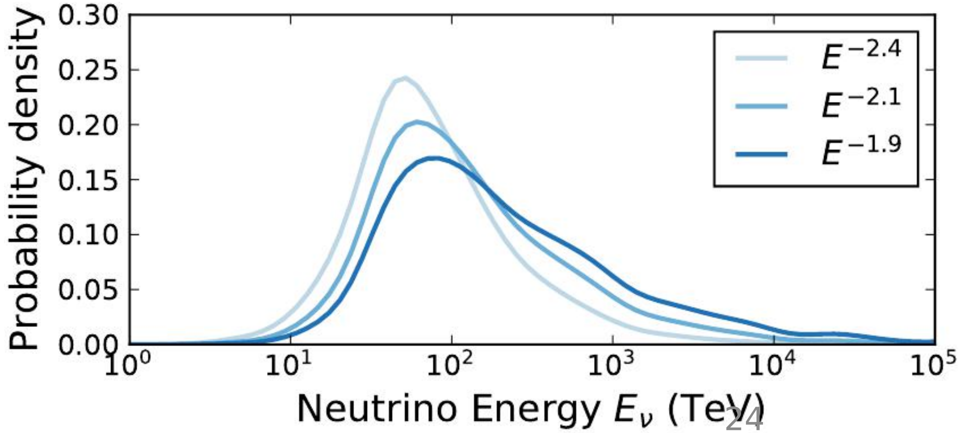
Energy losses before muon has entered the detector are not known

Derived from MC simulation:

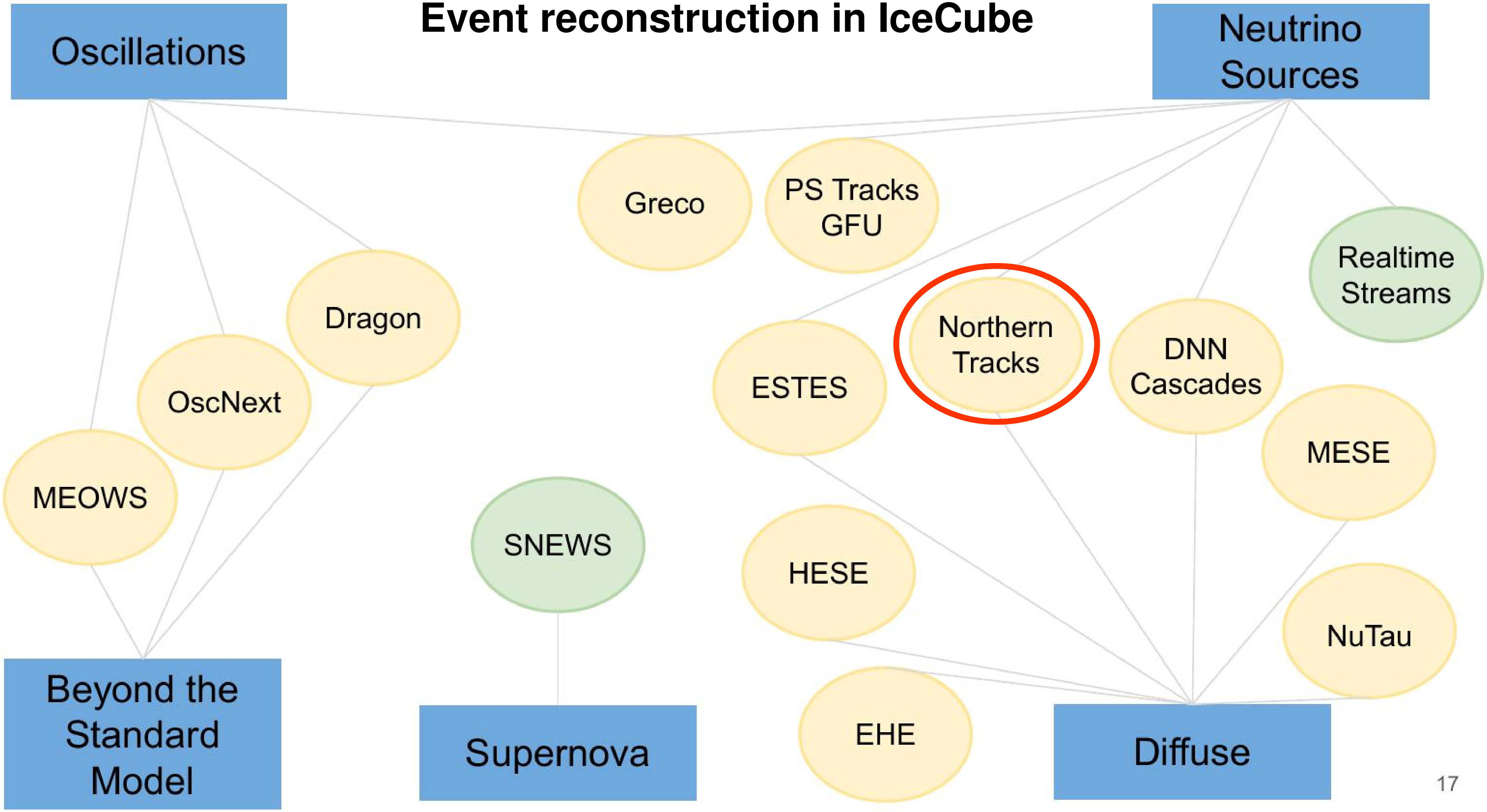
Lower bound on neutrino energy is available

Neutrino most probable energy is often quoted





Event reconstruction in IceCube



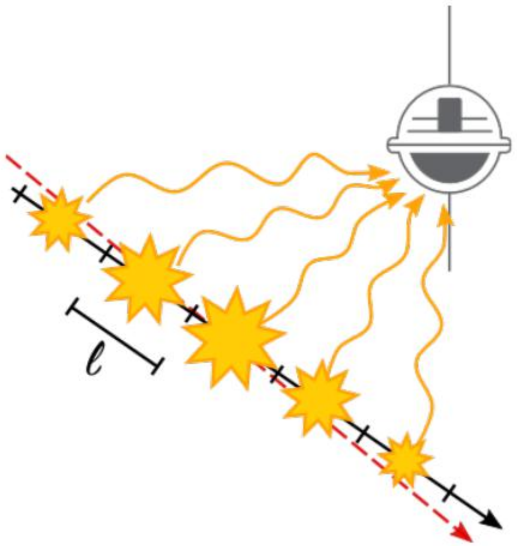
Track event reconstruction in IceCube

Likelihood for all hits at the channel:

$$L(x, y, z, t, \theta, \phi) = \prod_{j=1}^{N_{\mathrm{DOM}}} \prod_{i=1}^{N_{\mathrm{hit}}} [p_j(t_i)]^{q_i}$$

Likelihood for the first photon:

$$L_{1\mathrm{st}}(x, y, z, t, \theta, \phi) = \prod_{j=1}^{N_{DOM}} p_{j,1\mathrm{st}}(t_1)$$



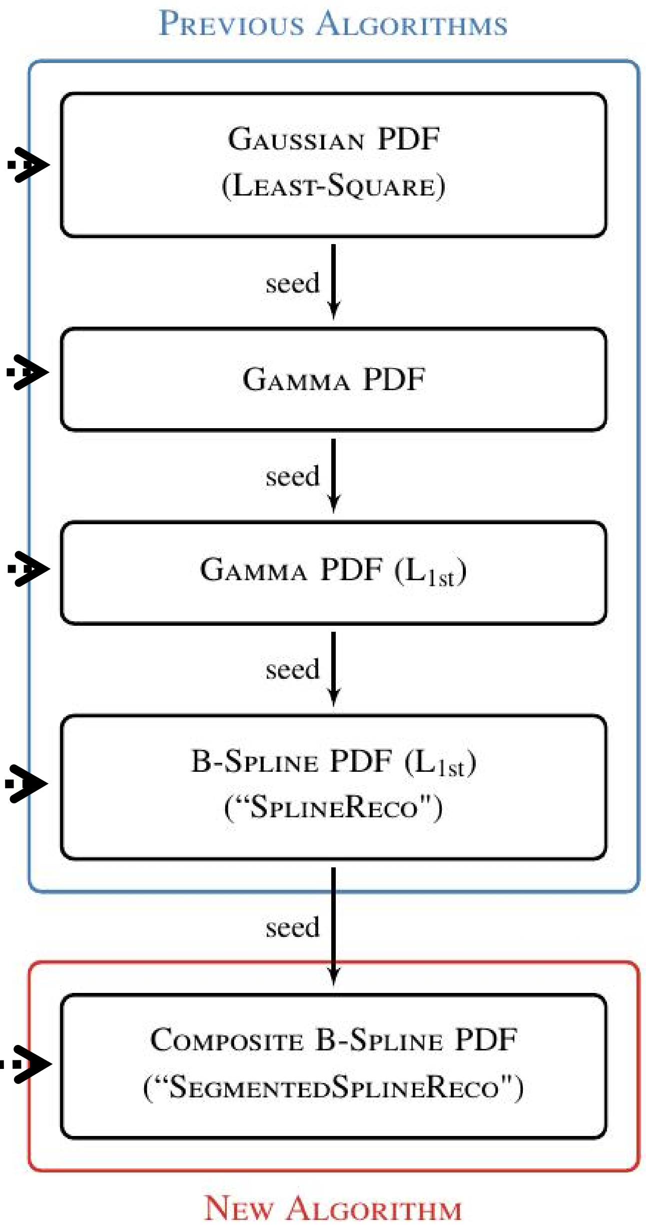
"Plane wave" approximation

Full MIP likelihood with homogenous ice model

First photon MIP likelihood with homogenous ice model

First photon MIP likelihood with realistic ice model

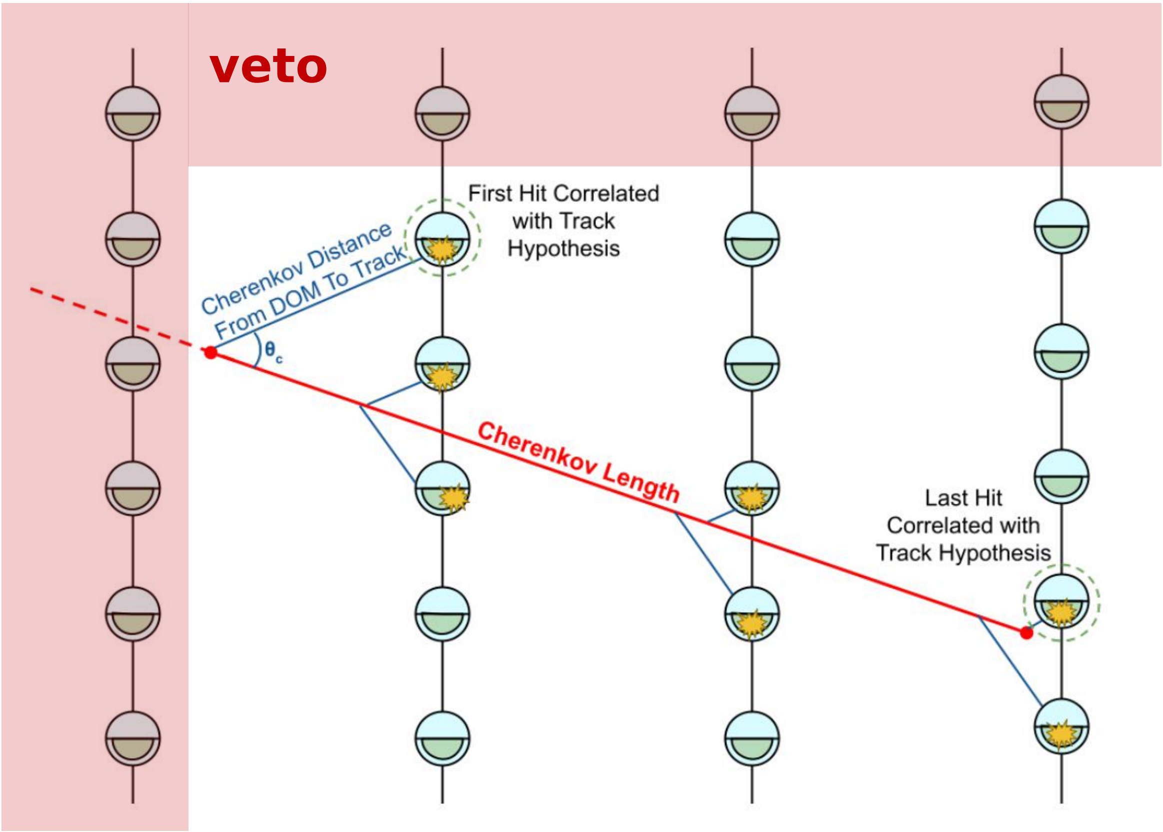
Likelihood with realistic ice model accounting for EM showers at stochastic energy losses



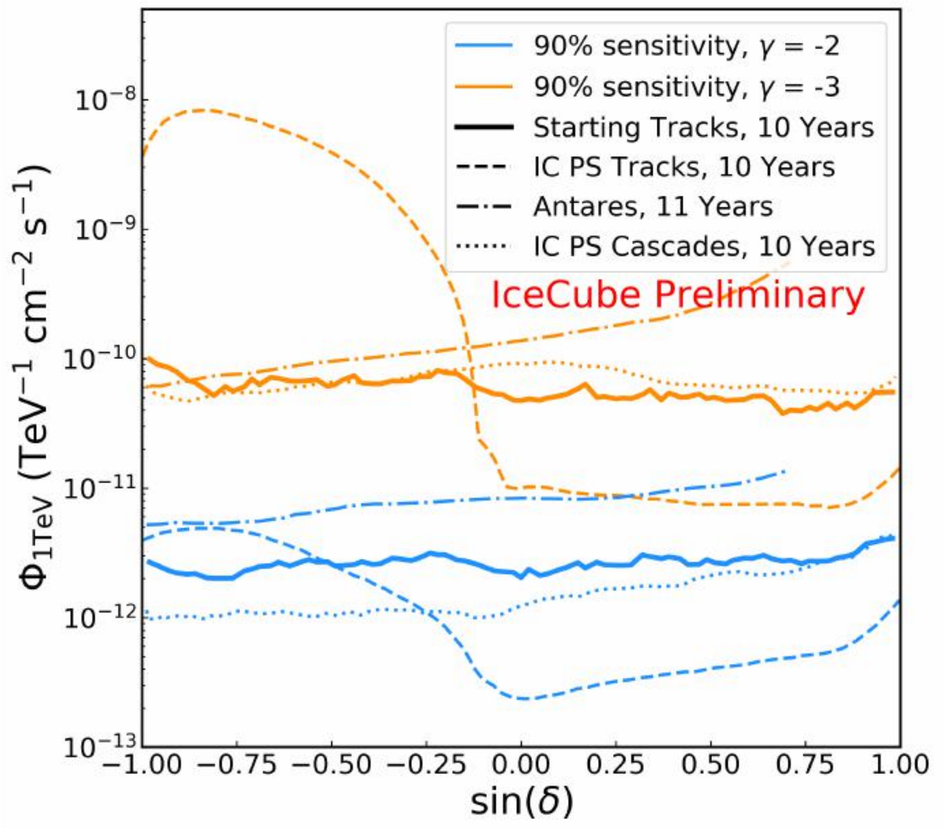
JINST 16 P08034

Starting tracks in IceCube

Starting tracks allow to study downgoing numu neutrino flux
- Worse angular resolution
- Higher energy threshold



Sensitivity compared to through-going tracks

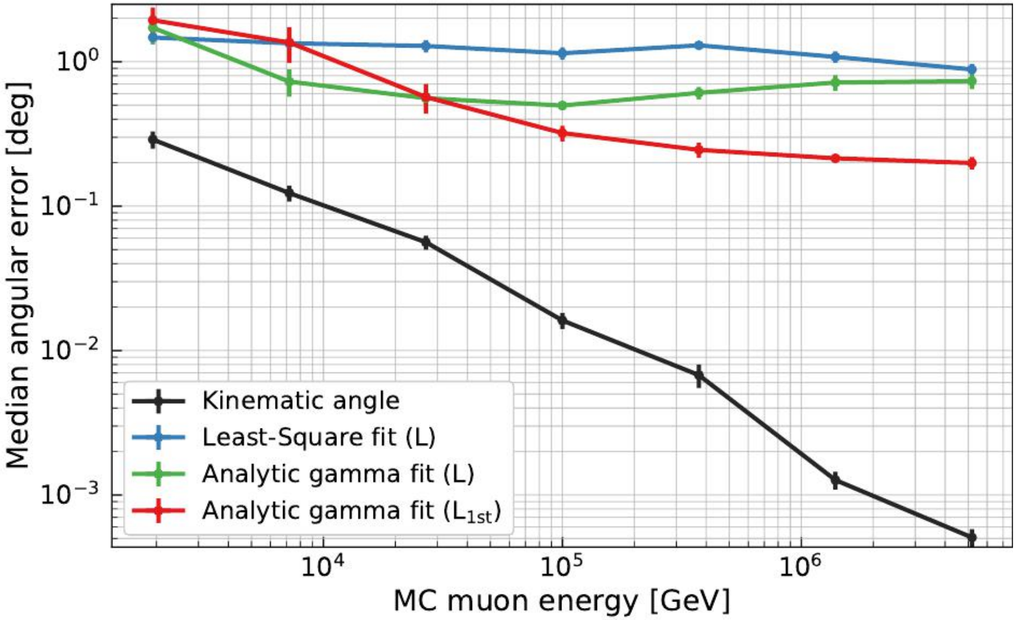


Sarah Mancina, IC thesis, 2022

Track events reconstruction in IceCube

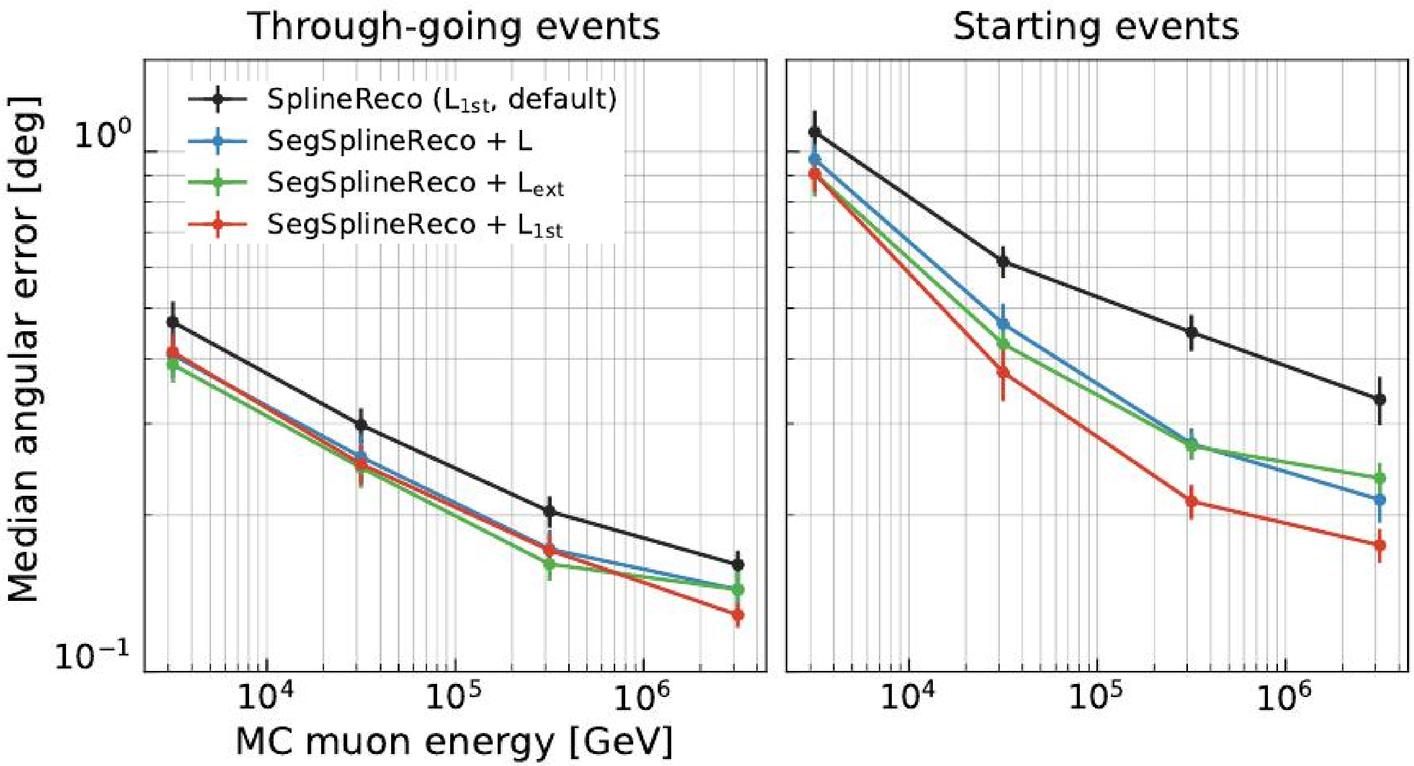
Angular resolutions for different stages of track-like event reconstruction

L > 700m



JINST 16 P08034

L > 700m **L > 400m**



IceCube ice properties

Feature of IceCube: complex ice light propagation properties
Very small absorption and enormous light scattering

Effective scattering length is few times shorter than absorption length

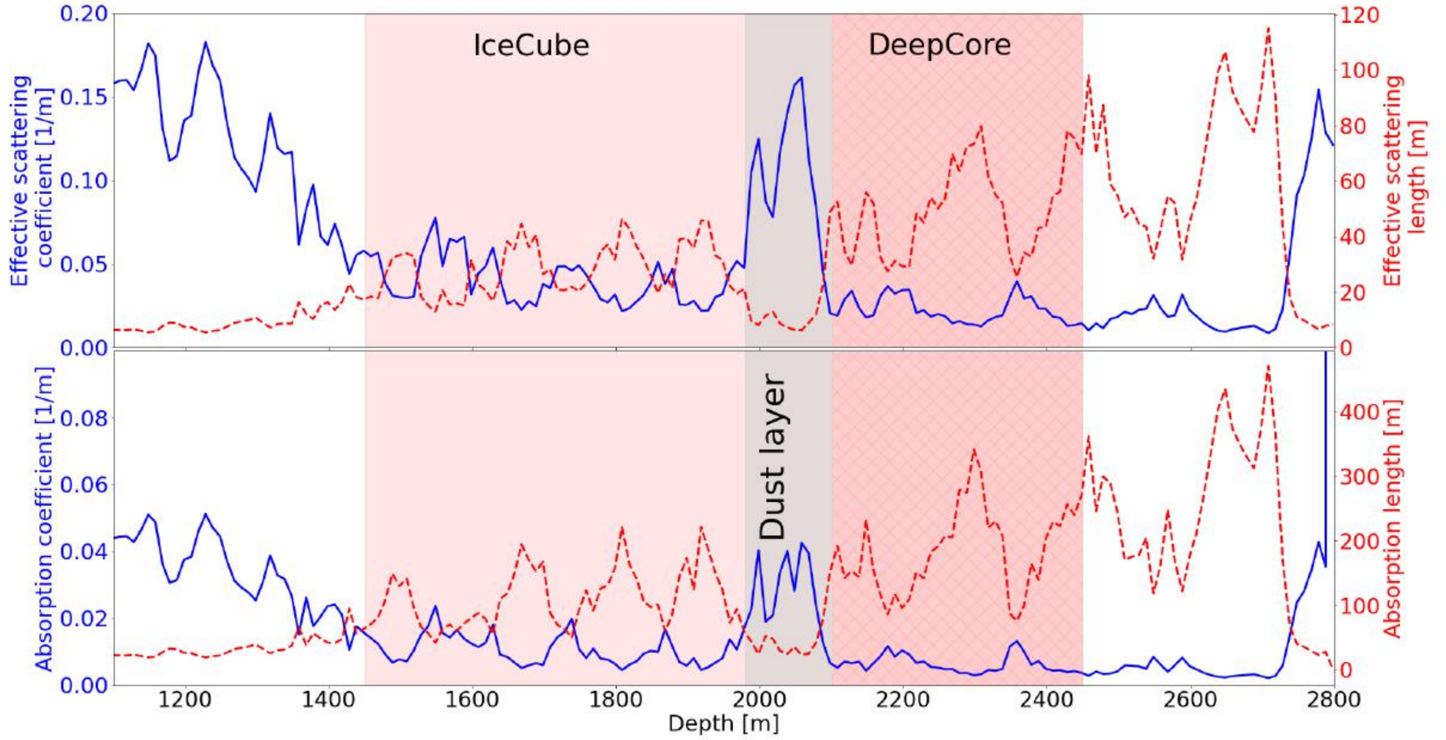
Large dependence on depth of both scattering and absorption

“Dust layer” with both large absorption and scattering

Light propagation anisotropy

Layers with stable ice properties are not at constant depth

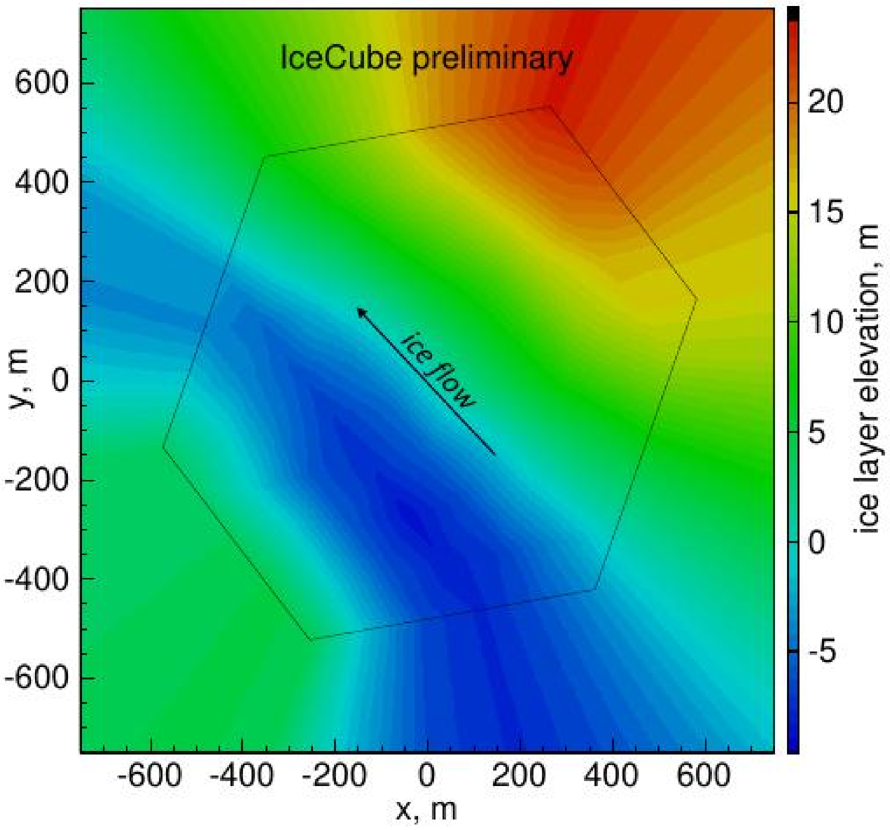
Scattering on bubbles in drill holes



IC, The Cryosphere Discussions 2022 (2022) 1–48

IceCube ice properties

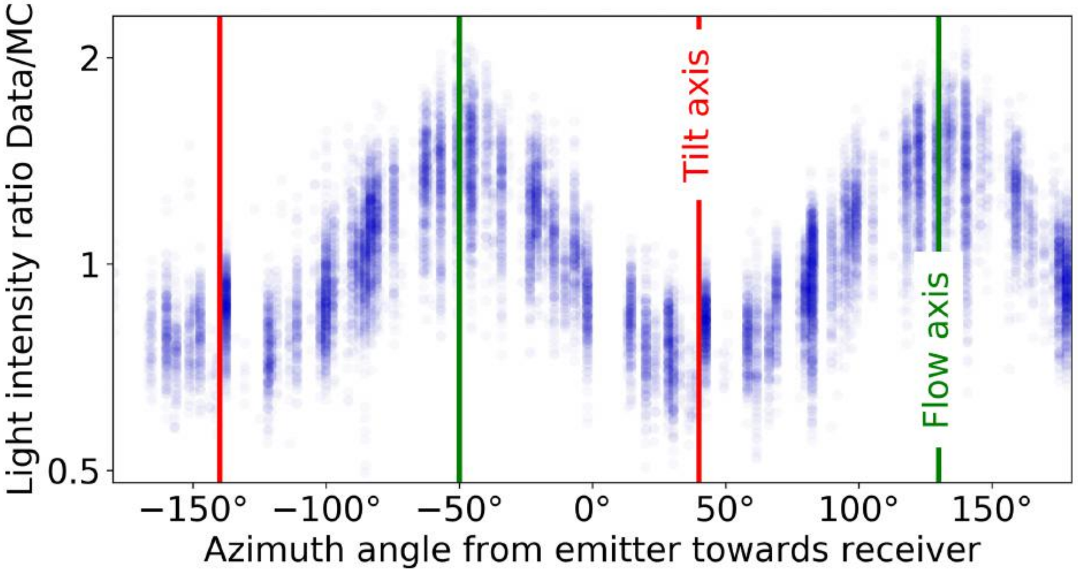
Antarctic Ice Shield flows at
10m/year which defines “flow axis”

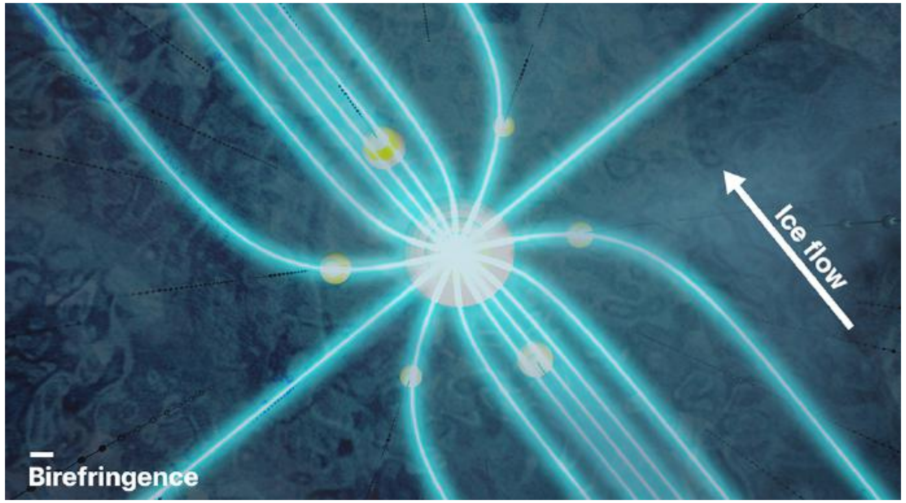


Layer undulations: Ice layers with constant ice properties change their depth

PoS(ICRC2023)975

Light tends to be deflected towards flow axis

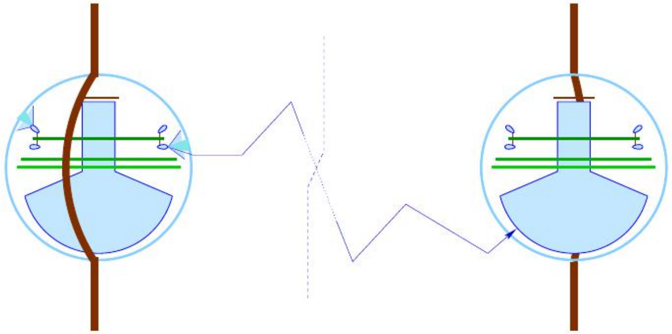


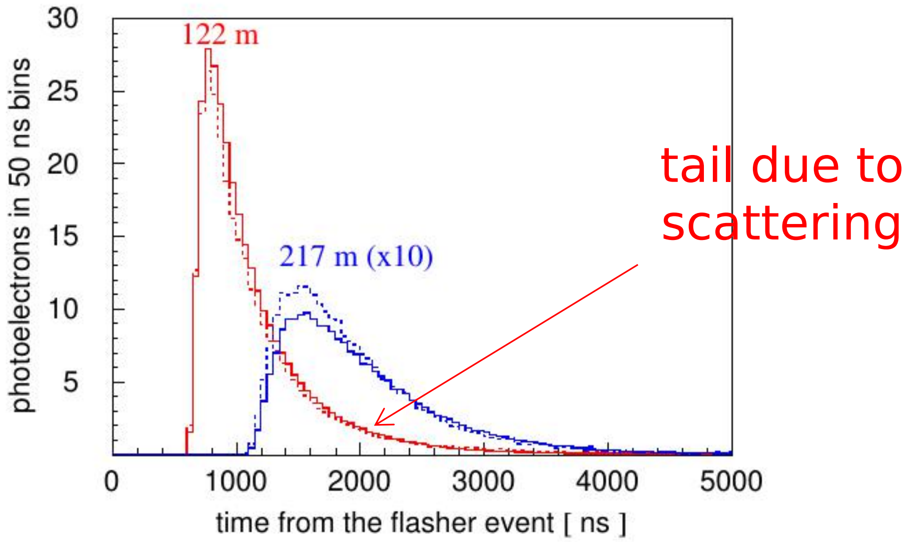


IC, The Cryosphere 18 (2024) 1

IceCube ice properties

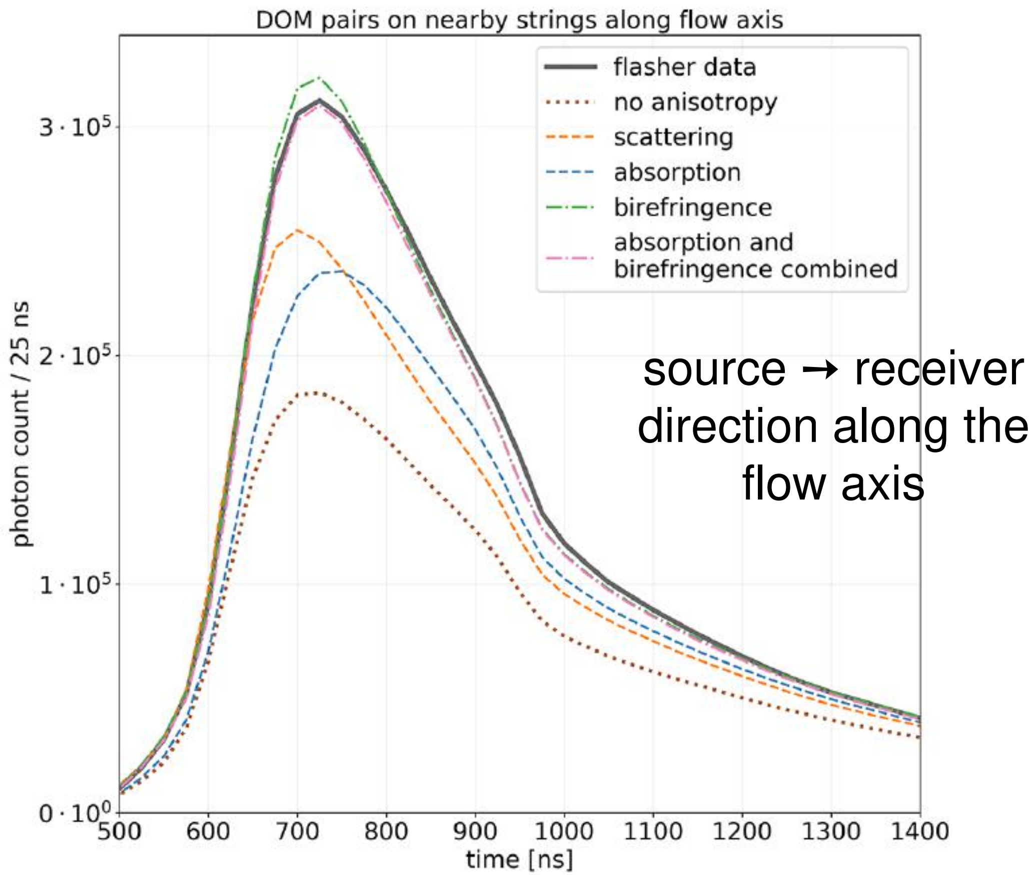
Ice models are established by fits to flasher data





NIM A, Volume 711, 21 May 2013, Pages 73-89

Results of fits of models incorporating different effects to flasher data



IC, The Cryosphere 18 (2024) 1

IceCube ice properties

Steady progress in ice description models over the years

AMANDA ice models:			model error
bulk, f125, mam, mamint, stdkurt, sudkurt, kgm, …			
millennium (published 2006) → AHA (2007)			55%
IceCube ice models:			
WHAM	(2011)		42%
SPICE 1	(2009)		29%
SPICE 2, 2+, 2x, 2y	(2010)	added ice layer tilt	
SPICE Mie	(2011)	fit to scattering function	29%
SPICE Lea	(2012)	fit to scattering anisotropy	20%
SPICE (Munich)	(2013)	7-string, LED unfolding	17%
$SPICE^3$ (CUBE)	(2014)	llh fixes, DOM sensitivity fits	11%
SPICE 3.0	(2015)	improved RDE, ang. sens. fits	10%
SPICE 3.1, 3.2	(2016)	85-string, correlated model fit	<10%
SPICE HD, 3.2.2	(2017)	direct HI and DOM sens., cable, DOM tilt	
SPICE EMRM	(2018)	absorption-based anisotropy	*single*
SPICE BFR	(2020)	birefringence-based anisotropy	*LEDs*

D.Chirkin, M.Rongen
IceCube Polar Science workshop, 2021

IceCube ice properties

Event selections are updated incorporating more data, new reconstructions and improved in ice models

HESE events:
diffuse flux discovery (2013)

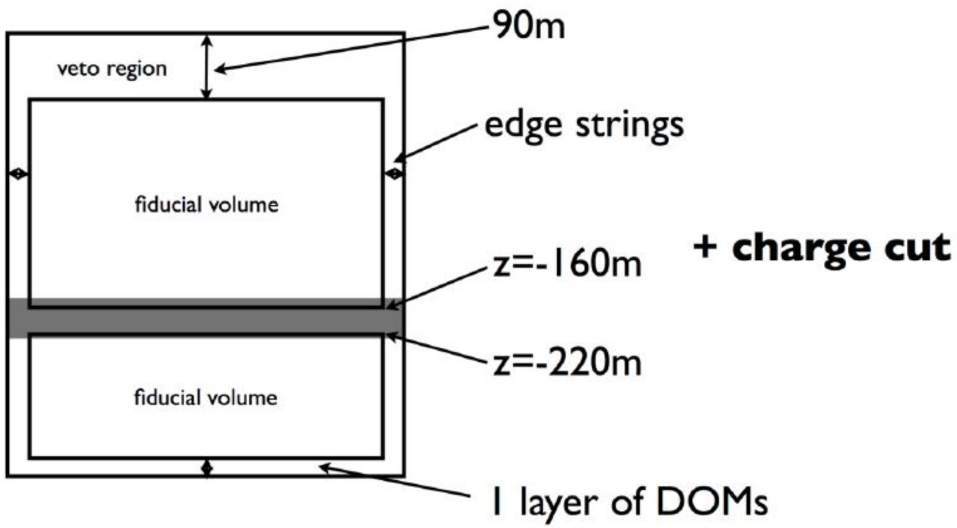
An update of High-Energy Starting Events (HESE) event sample **PoS(ICRC2023)1030**

- Added 4.5 years of data to **[Phys. Rev. D 104, 022002]**
- Ice model with birefringence
- Ice layer undulations
- Reconstruction by re-simulation

102 events before, 62 new events found

Re-simulation: event reconstruction by multiple event simulation finiding event parameters fitted to data

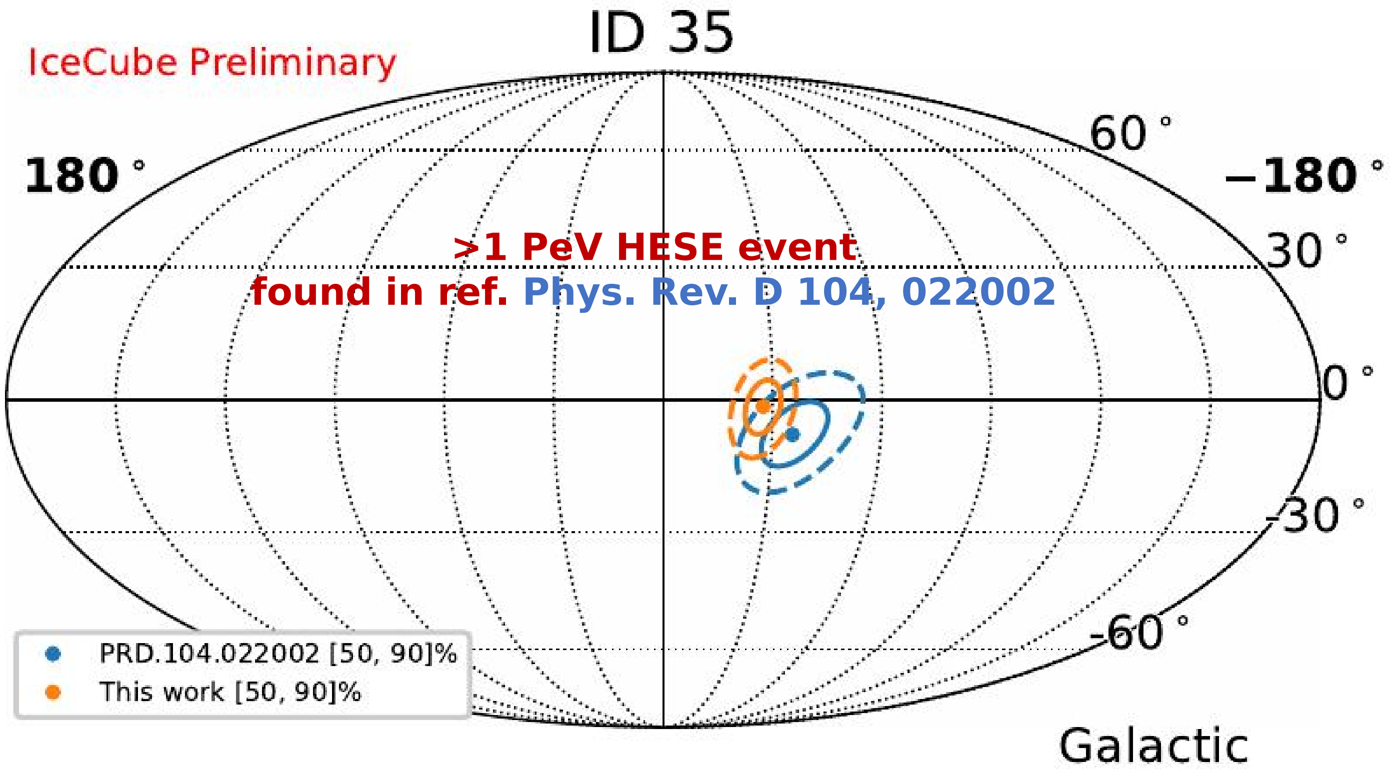
HESE event definition



charge cut: $Q_{tot} > 6000$ p.e.

IceCube ice properties

Directions of previously found HESE events were updated
Considerable change in direction of some events



Impact on the direction reconstruction: Main contribution due to improved ice modelling **[PoS(ICRC2023)1030]**

An update has entered IceCube 12 year data release: **https://doi.org/10.7910/DVN/PZNO2T**

Systematic uncertainties

Uncertainties in the knowledge of the detector and theoretical calculations affect the measurement results

Main sources of uncertainty in neutrino telescope

- Detection medium properties: Absorption and scattering measurement uncertainties, ice properties
- Sensitivity of optical modules: In-situ optical module sensitivity, module rotations, sedimentations, drill holes (IC)
- Theoretical uncertainties in background fluxes: atmospheric neutrino flux

Typically considered systematic uncertainties, e.g. Baikal-GVD diffuse flux **[Phys.Rev.D 107 (2023) 4, 042005]**

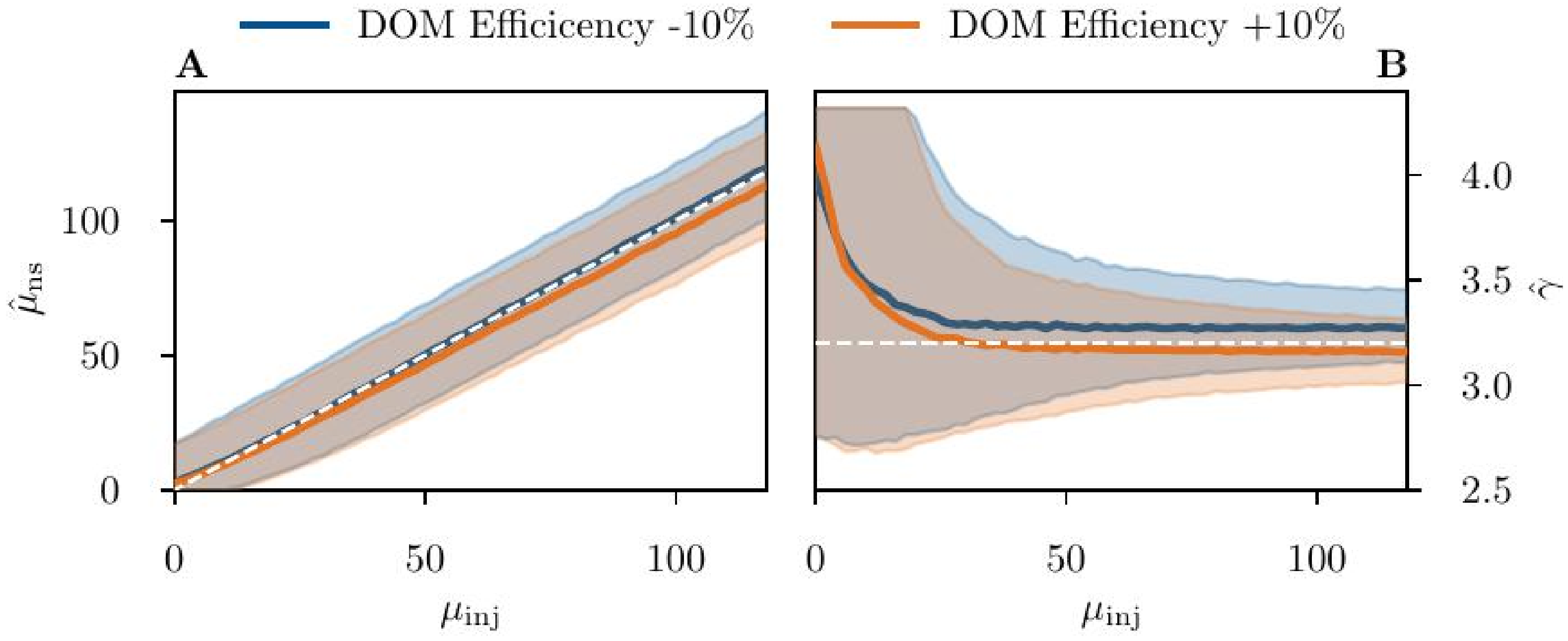
- Light absorption +- 5%
- Optical module sensitivity +-10%
- Theoretical: Atmospheric neutrino flux +-15%

Systematic uncertainties

IceCube ν emission from NGC1068 paper, PS tracks
Science 378, 538 (2022)

Considered uncertainties

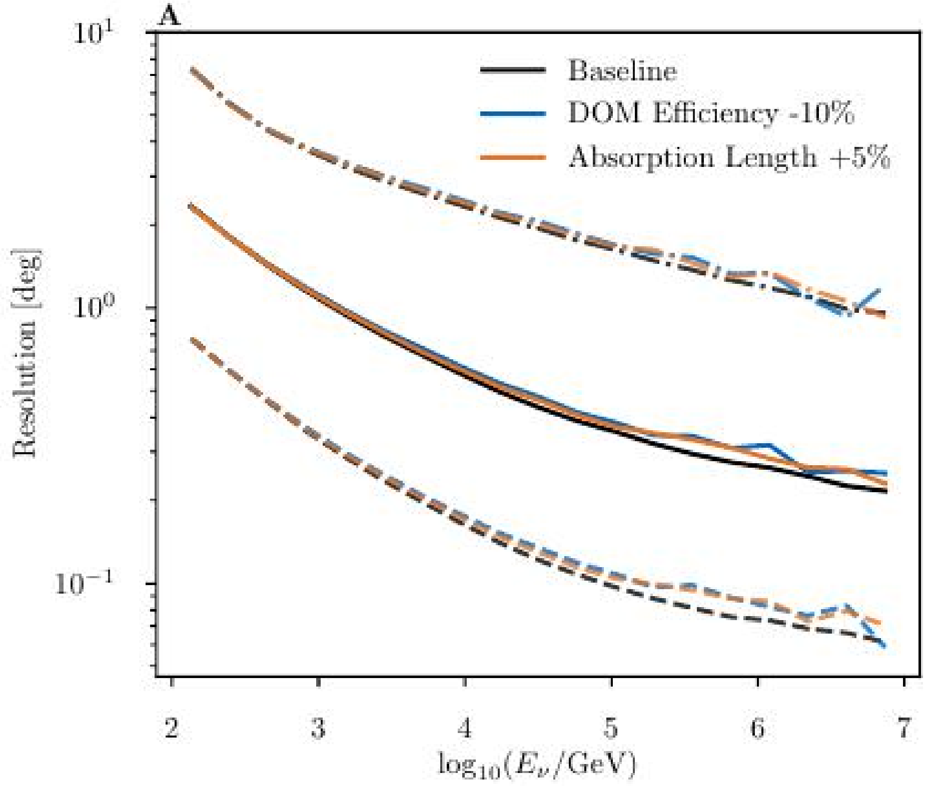
- Absorption length ±5%
- Scattering length ±5%
- DOM efficiency ±10%
- Angular acceptance due to drill hole: shadowing of head-on photon direction



The largest impact from

- Absorption length +5%
- DOM efficiency -10%

Up to 10% effect on resolution



BACKUP

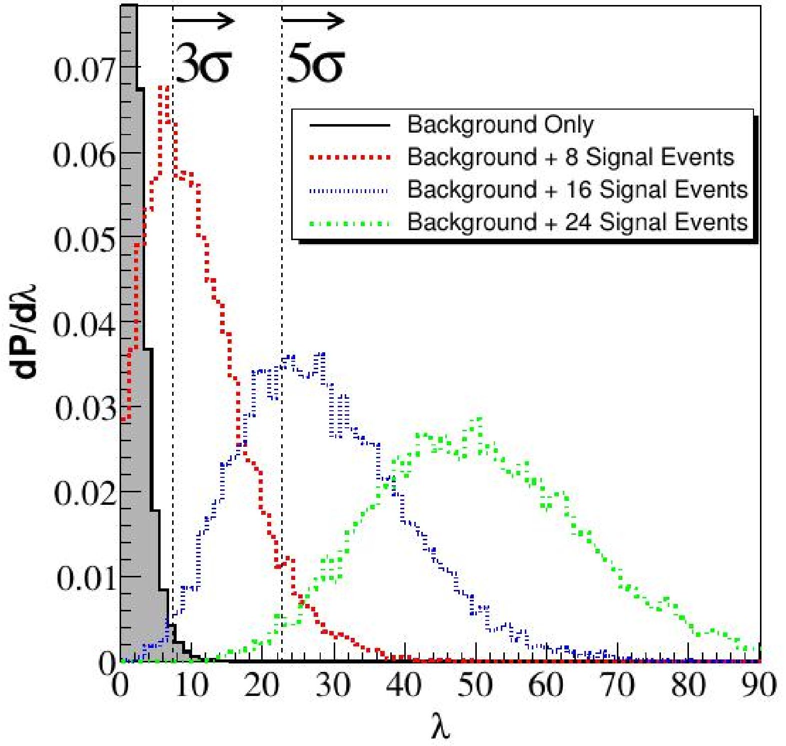
Point source search

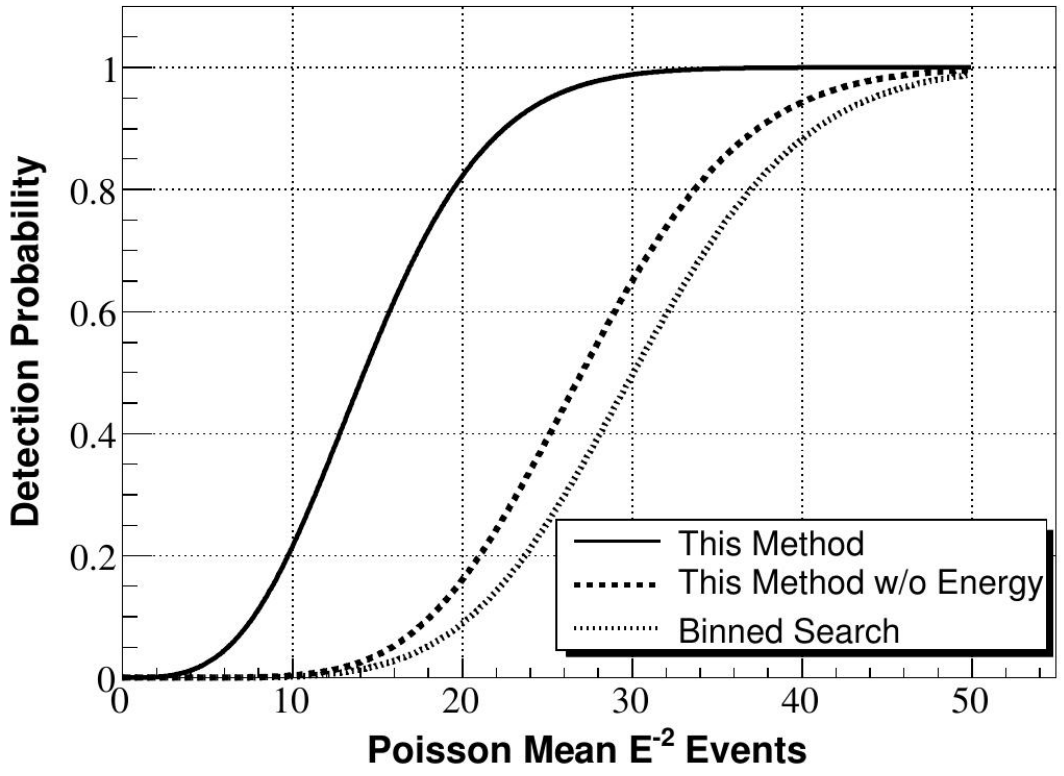
$$\mathcal{L}(\vec{x}_s, n_s, \gamma) = \prod_N \left(\frac{n_s}{N} \mathcal{S}_i + (1 - \frac{n_s}{N}) \mathcal{B}_i \right)$$

$$\mathcal{S}_i(\vec{x}_i, \vec{x}_s, E_i, \gamma) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\vec{x}_i - \vec{x}_s|^2}{2\sigma^2}} P(E_i|\gamma)$$

$$\mathcal{B}_i = P(\vec{x}_i, E_i | \phi_{atm} + \phi_{mu} + \phi_{diffuse})$$

$$\lambda = -2 \cdot sign(\hat{n}_s) \cdot log \left[\frac{\mathcal{L}(\vec{x}_s, 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma})} \right]$$





Point source search

Widely used point source unbinned likelihood search approach **[J. Braun et. al.,2008]**

Fix potential source direction in equatorial coordinates: $\vec{x}_s$

Assume the symmetric detector angular resolution σ

Choose some (RA,dec) region around the source, e.g. 4 σ

Suppose we have a data sample and N events have entered the (RA,dec) region

Evaluate the likelihood function over all data events

$$\mathcal{L}(\vec{x}_s, n_s, \gamma) = \prod_N \left(\frac{n_s}{N}\mathcal{S}_i + (1 - \frac{n_s}{N})\mathcal{B}_i \right)$$

Where S_i and B_i are the signal and background probability density functions (PDF)
n_s is the number of signal events, free parameter

Point source search

PDF for signal: $$\mathcal{S}_i(\vec{x}_i, \vec{x}_s, E_i, \gamma) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\vec{x}_i - \vec{x}_s|^2}{2\sigma^2}} P(E_i|\gamma)$$

data event coordinates

spatial term

spectral term
E_i data event energy estimate, **γ is the source spectrum power**

PDF for background: $$\mathcal{B}_i = P(\vec{x}_i, E_i|\phi_{atm} + \phi_{mu} + \phi_{diffuse})$$

Likelihood is maximised with free n_s and γ

- Depending on analysis γ can also be fixed to some value or set of values

The test statistic of the form $$\lambda = -2 \cdot sign(\hat{n}_s) \cdot log\left[\frac{\mathcal{L}(\vec{x}_s, 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma})}\right]$$ is used for hypothesis testing

The power of good angular resolution

- The better the angular resolution - the smaller is the (RA, dec) region, the less background enters N
- The test statistic gains larger value for optimal parameters, thus better discovery potential

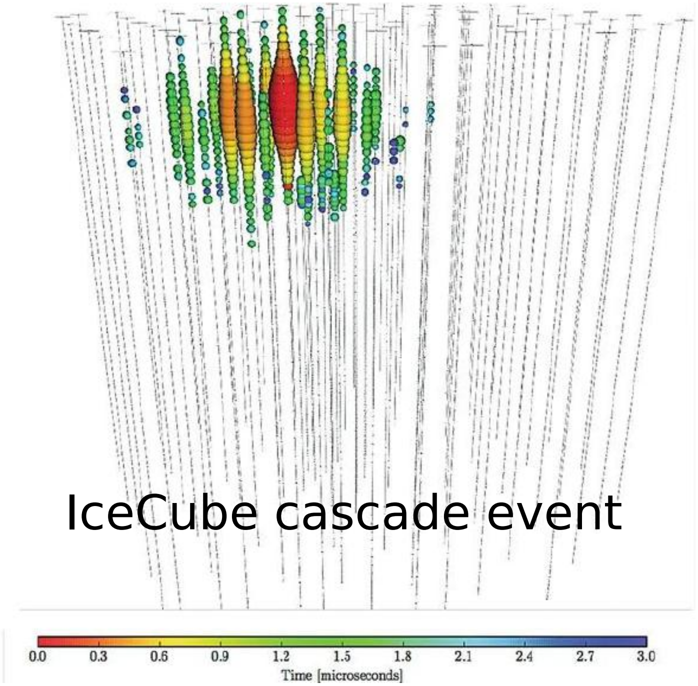
Energy reconstruction

Charge deposited in PMTs is used for cascade or muon energy reconstruction

Cascades:
- Calorimeter-type energy measurement
- Full energy deposition of cascade can be reconstructed

Precision depends on cascade location wrt the detector
- Worse for partially-contanied cascades

Neutrino energy resolution: 10-30%



IceCube cascade event

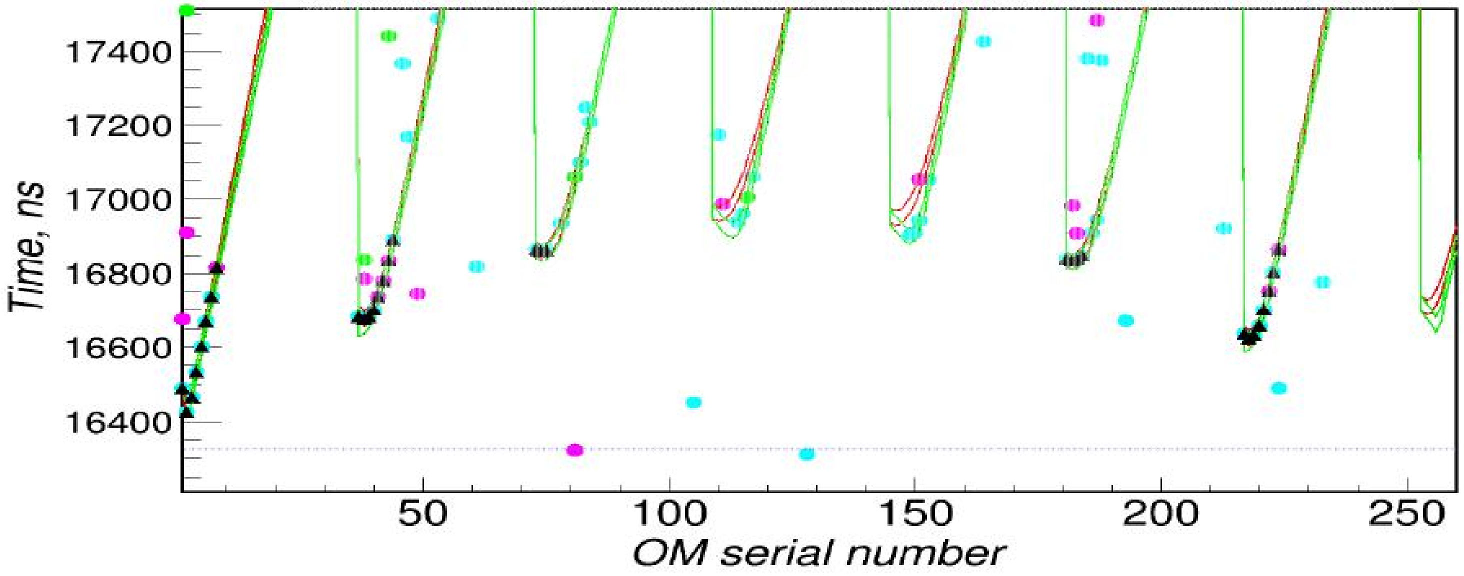
Event GVD200906

Zhan Djilkibaev, Rubakov 70 conference, February 2025

Track reconstruction

Cascade energy – 85 TeV; Zenith angle – 117°

OM hit arrival time

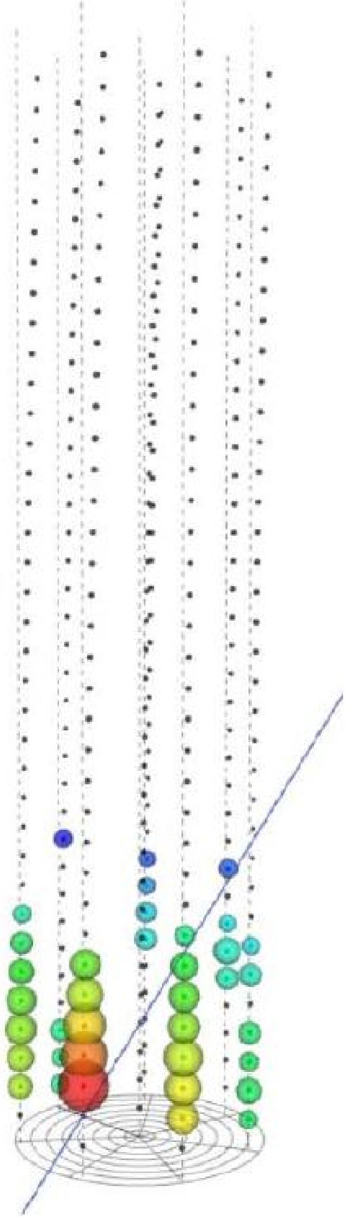


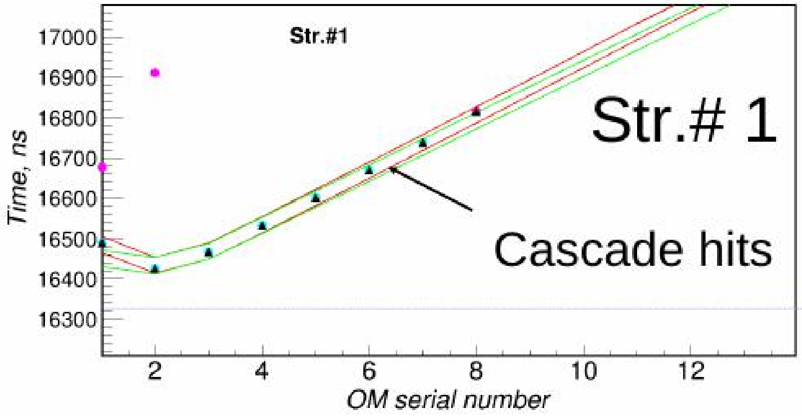
Season 2020, September
Cluster 5

N_{hits}	37
E^{μ}_{rec}	107.2 ТэВ
θ_{rec}	116.7°
L_{track}	140.1 м

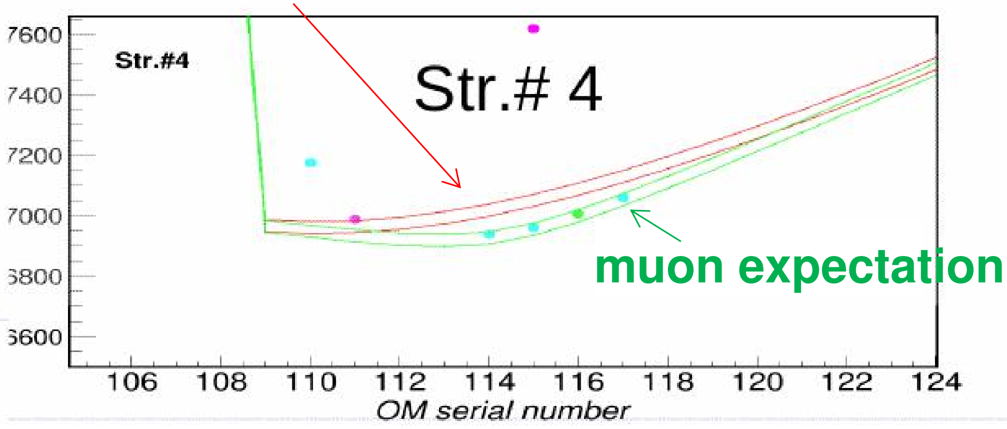
Angular precision:

50%:	0.7°
68%:	1.0°
90%:	1.5°

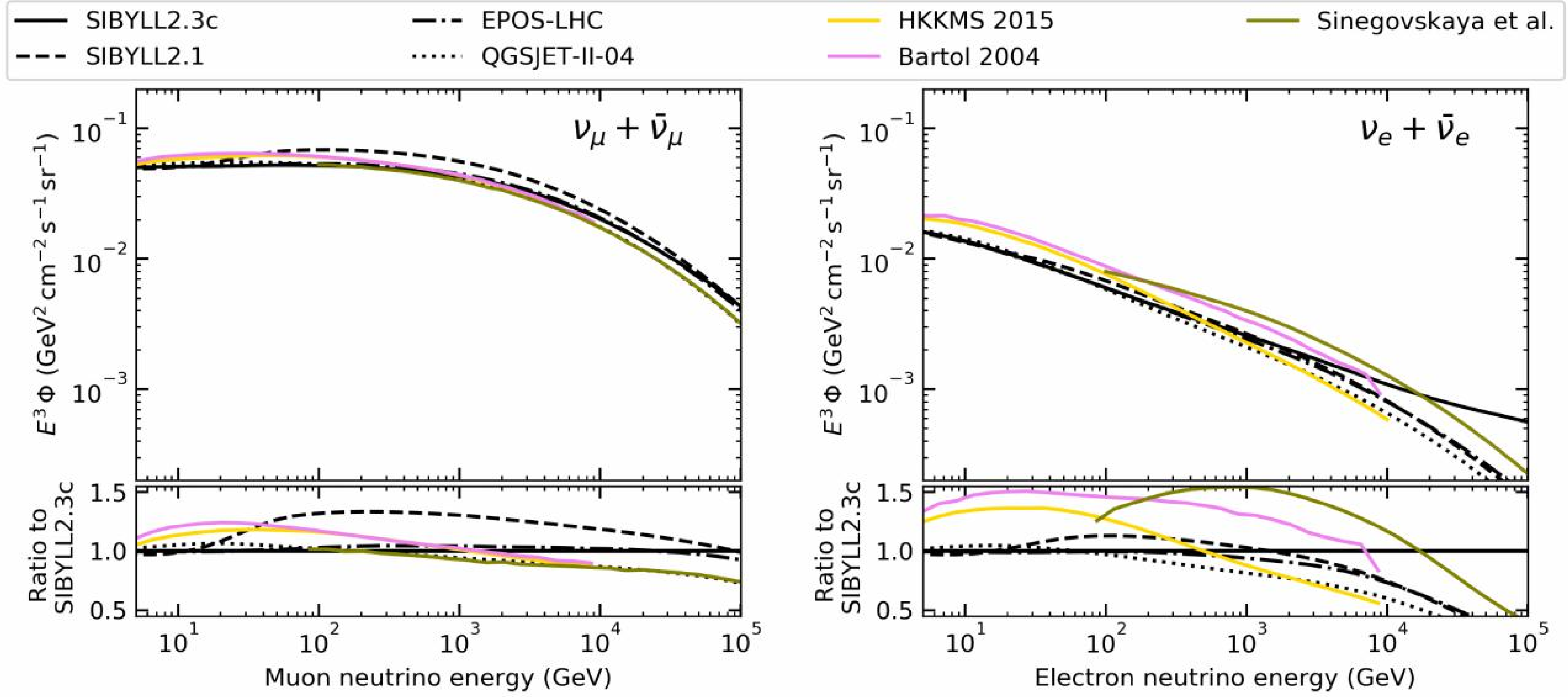




point source expectation



HKKMS: M. Honda et al., PRD 92 (2015)
Bartol: G. Barr et al., PRD 70 (2004)
Sinegovskaya et al. PRD 91 (2015)
MCEq: AF, R. Engel in prep.



A. Fedynich et al. Phys. Rev. D 100, 103018 (2019)

Event reconstruction

Detector response for Baikal-GVD