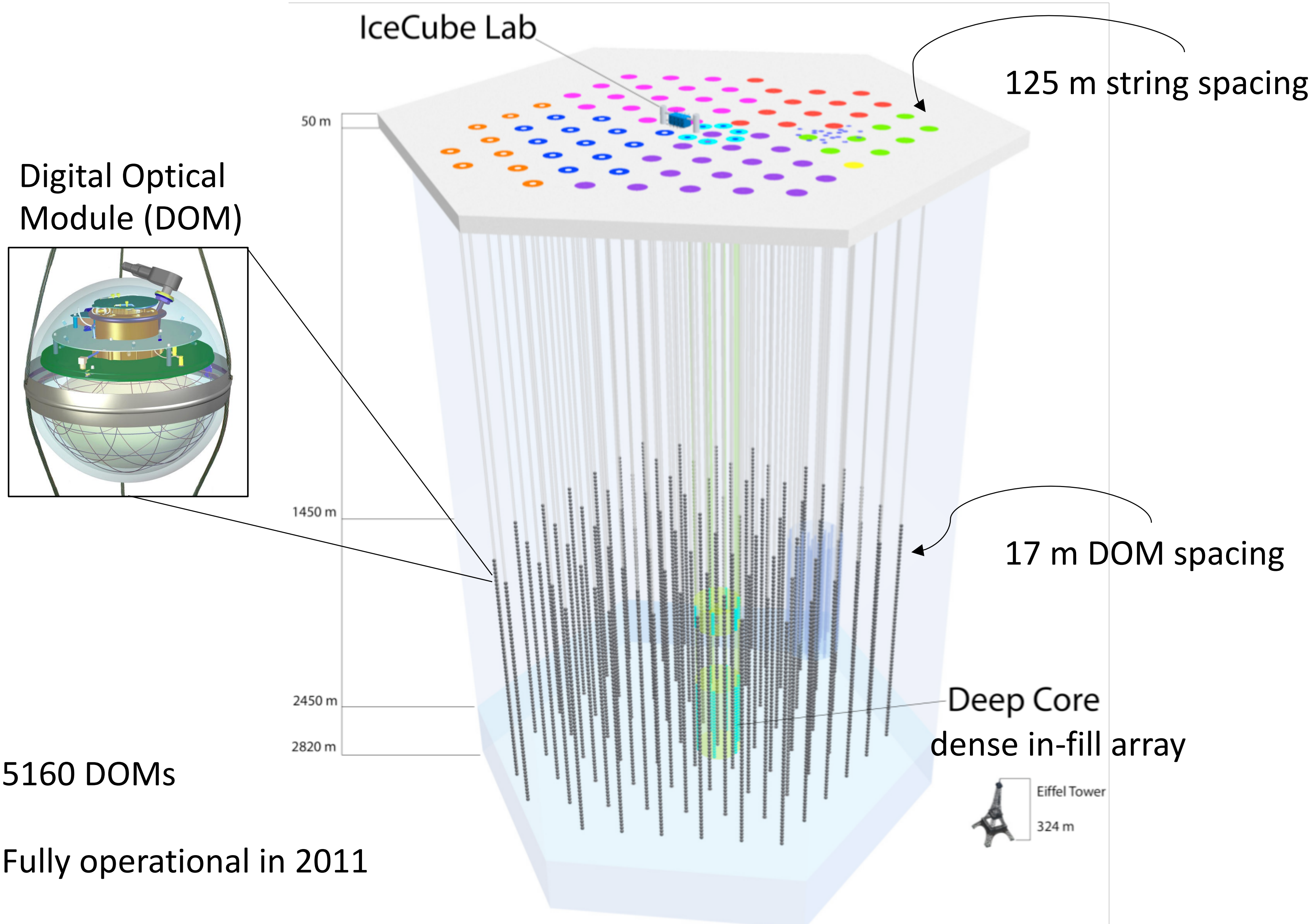
IceCube event reconstruction

Tianlu Yuan

IceCube Bootcamp

June 10, 2021

IceCube



5160 DOMs

Fully operational in 2011

Detection principals

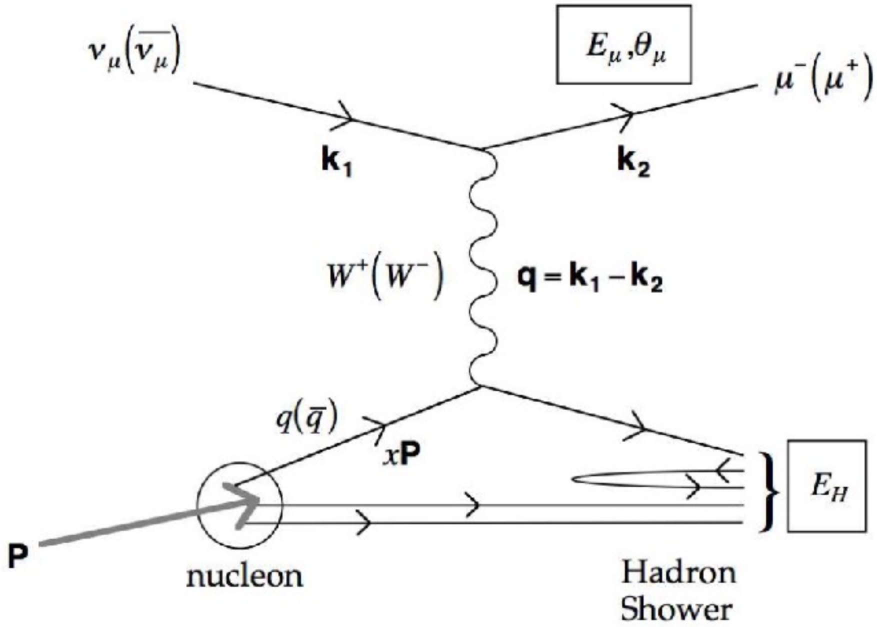
Neutrino interacts via weak force with targets in ice

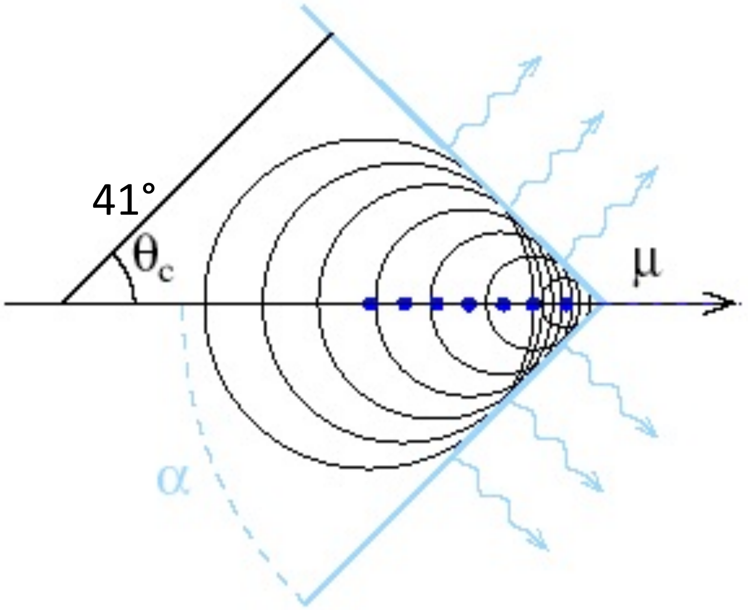
- At IceCube energies, primarily deep-inelastic scattering (DIS) off nucleons

Nucleon breaks apart; outgoing particles may be charged

Charged particles emit **Cherenkov radiation** detectable by PMTs

Rev. Mod. Phys. 84, 1307





Cherenkov radiation

Occurs when a charged particle travels faster than light-in-medium

Constructive interference of EM-field to form a plane wave

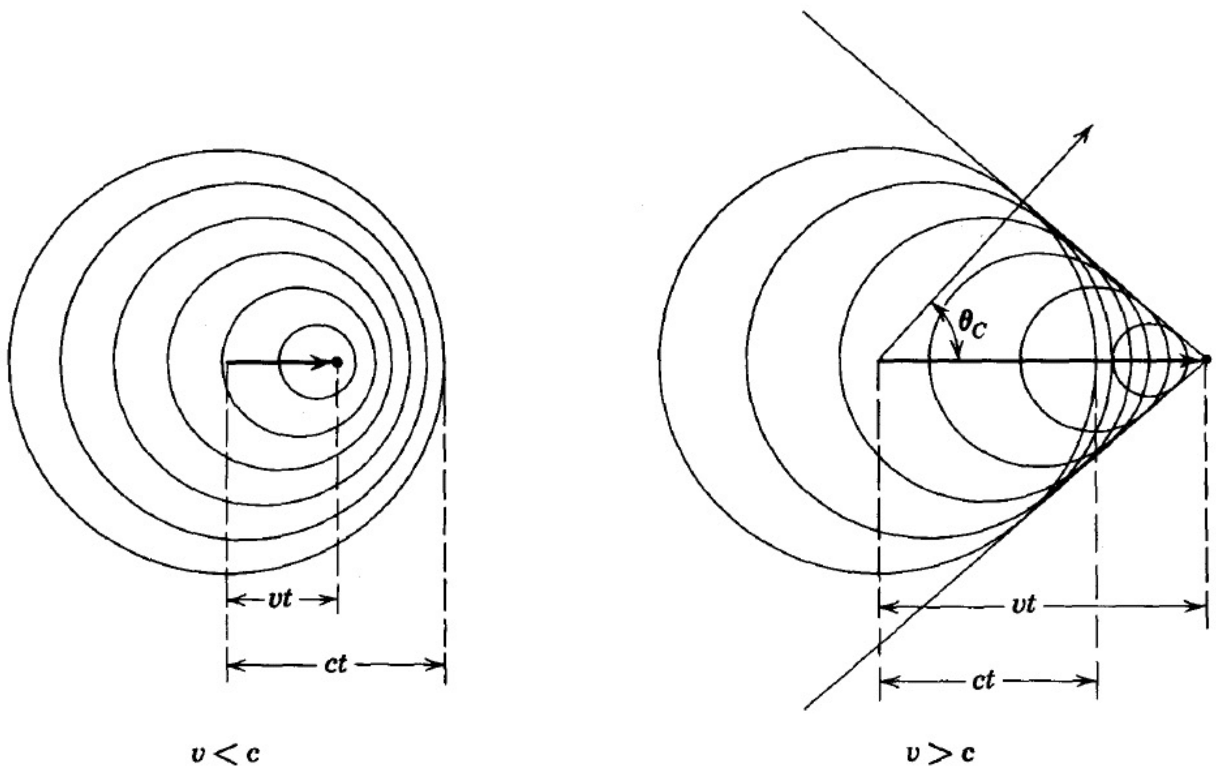
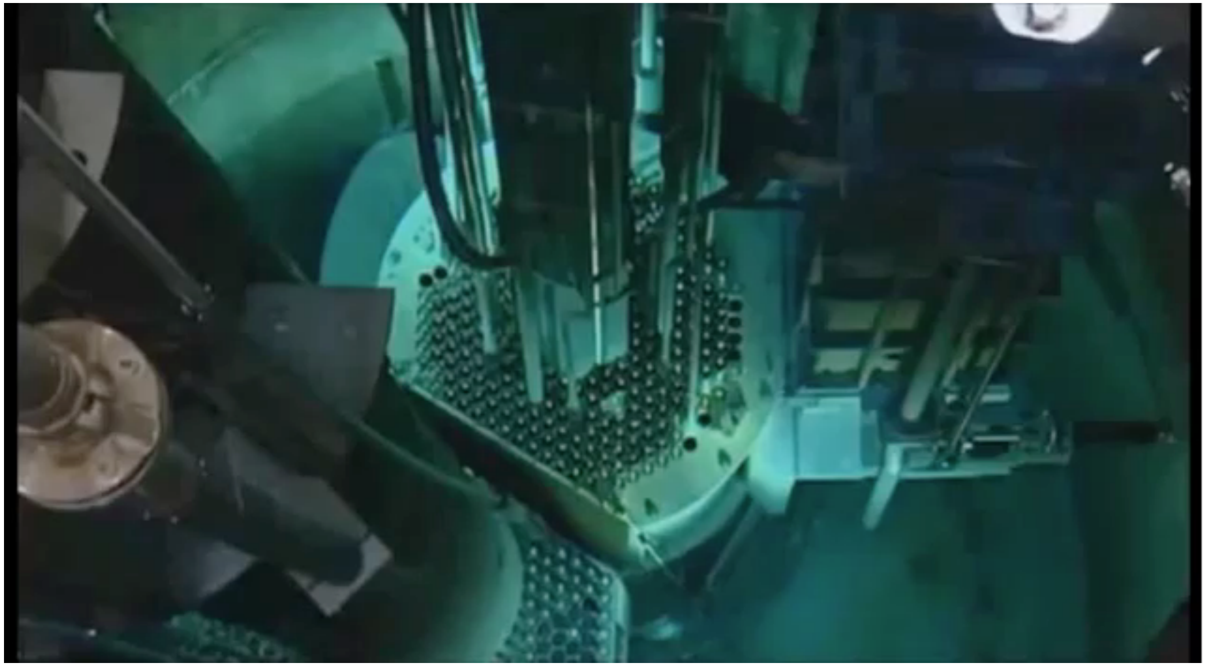
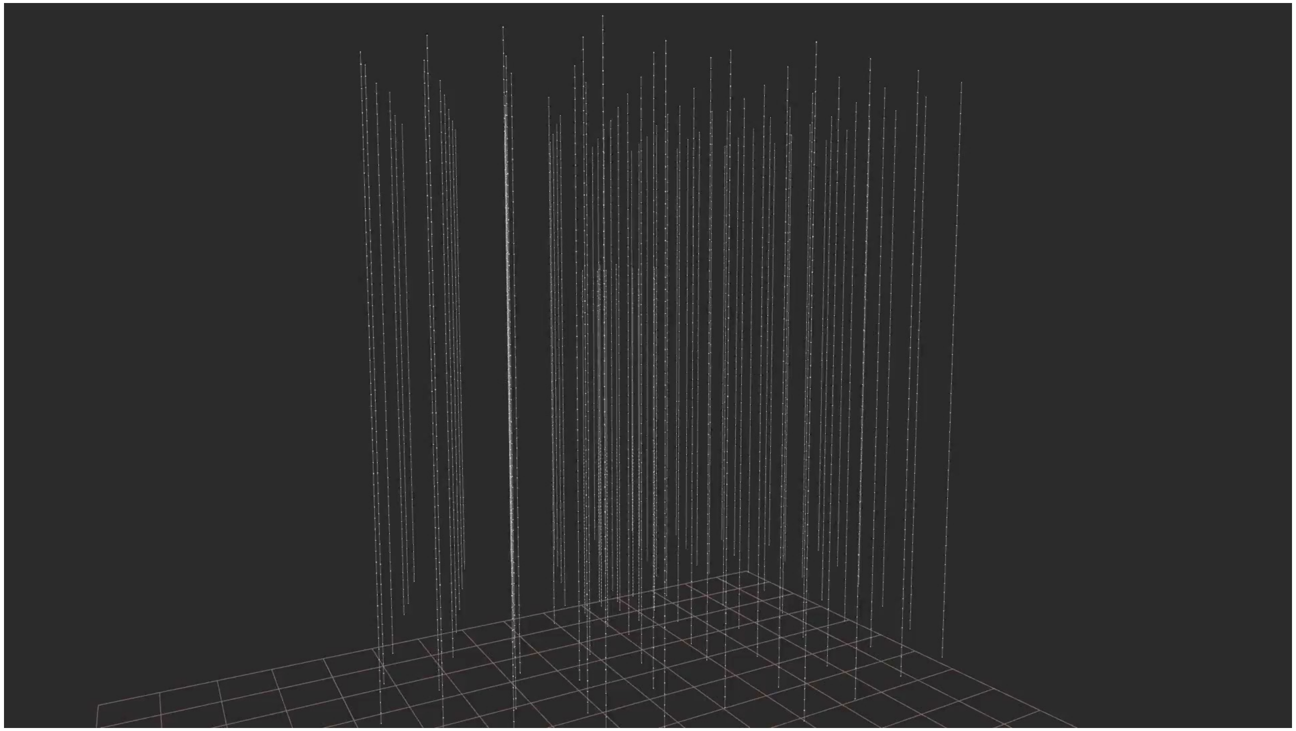


Fig. 14.14 Cherenkov radiation. Spherical wavelets of fields of a particle traveling less than, and greater than, the velocity of light in the medium. For $v > c$, an electromagnetic "shock" wave appears, moving in the direction given by the Cherenkov angle θ_C.

Cherenkov radiation in water

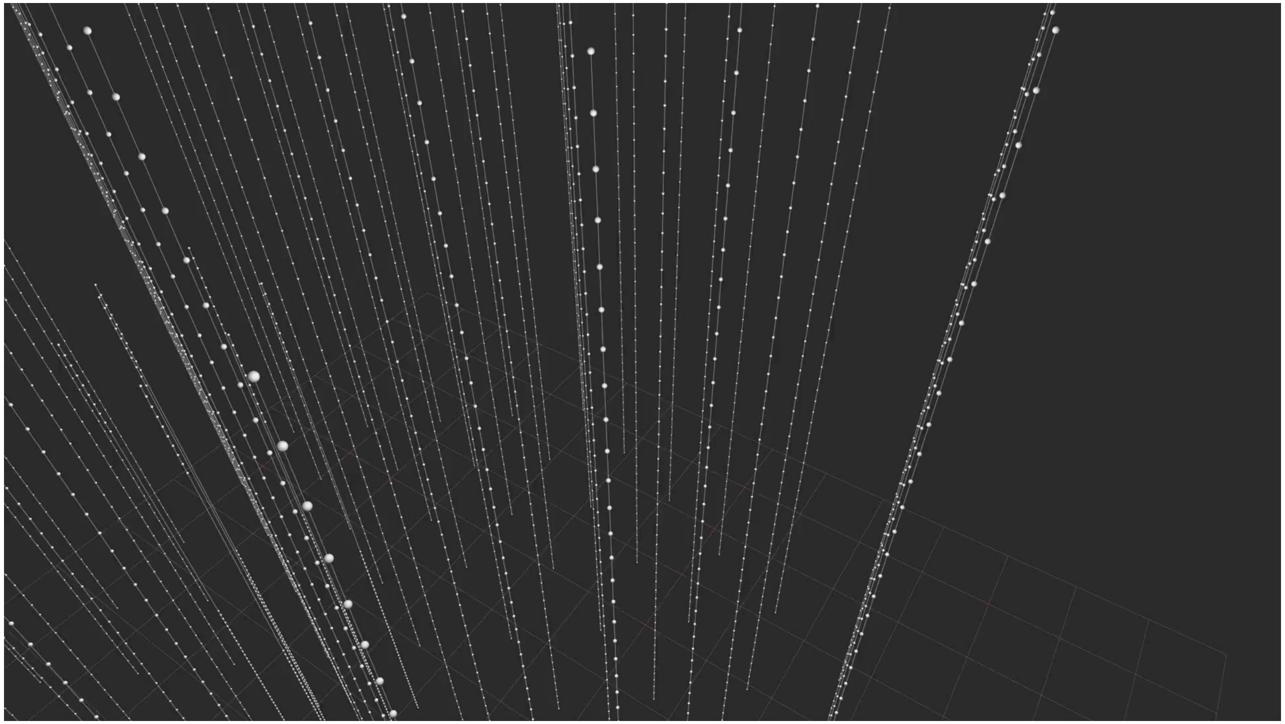


What do neutrinos look like in IceCube?



muons: long paths in the detector → **track**

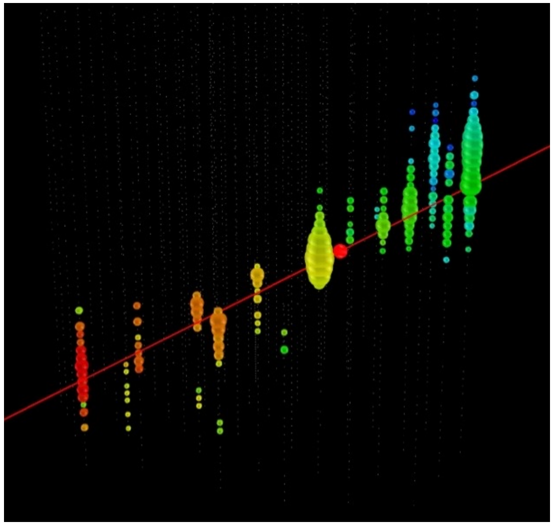
What do neutrinos look like in IceCube?



electrons/hadrons: shower of light → **cascade**

What IceCube actually sees (high-energy)

CC muon neutrino

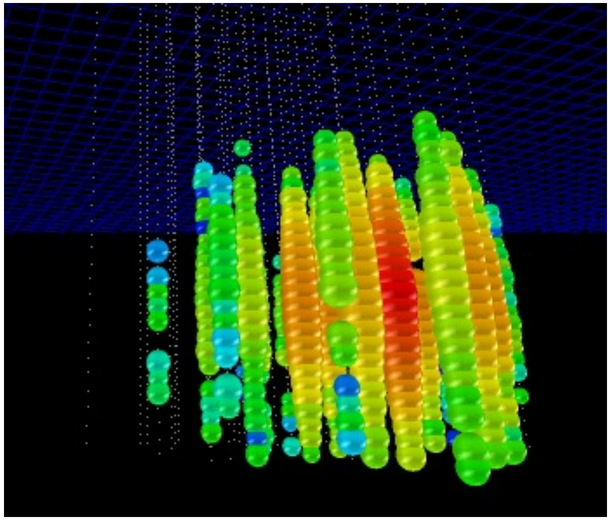


$\nu_\mu + N \rightarrow \mu + X$

track (data)

angular resolution ~ 0.5°
energy resolution ~ x2

NC or CC electron neutrino

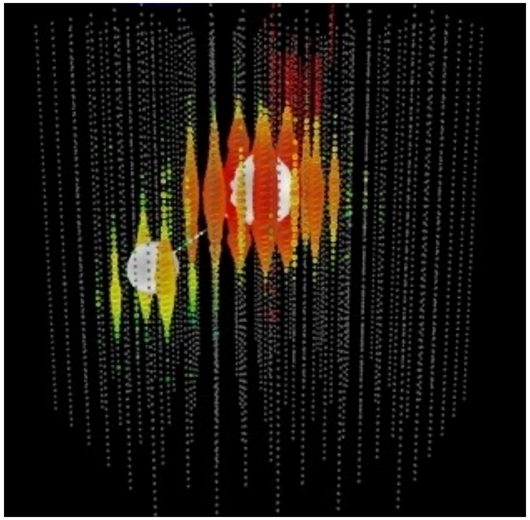


$\nu_e + N \rightarrow e + X$
$\nu_x + N \rightarrow \nu_x + X$

shower (data)

angular resolution ~ 10°
energy resolution ~ 15%

CC tau neutrino



$\nu_\tau + N \rightarrow \tau + X$

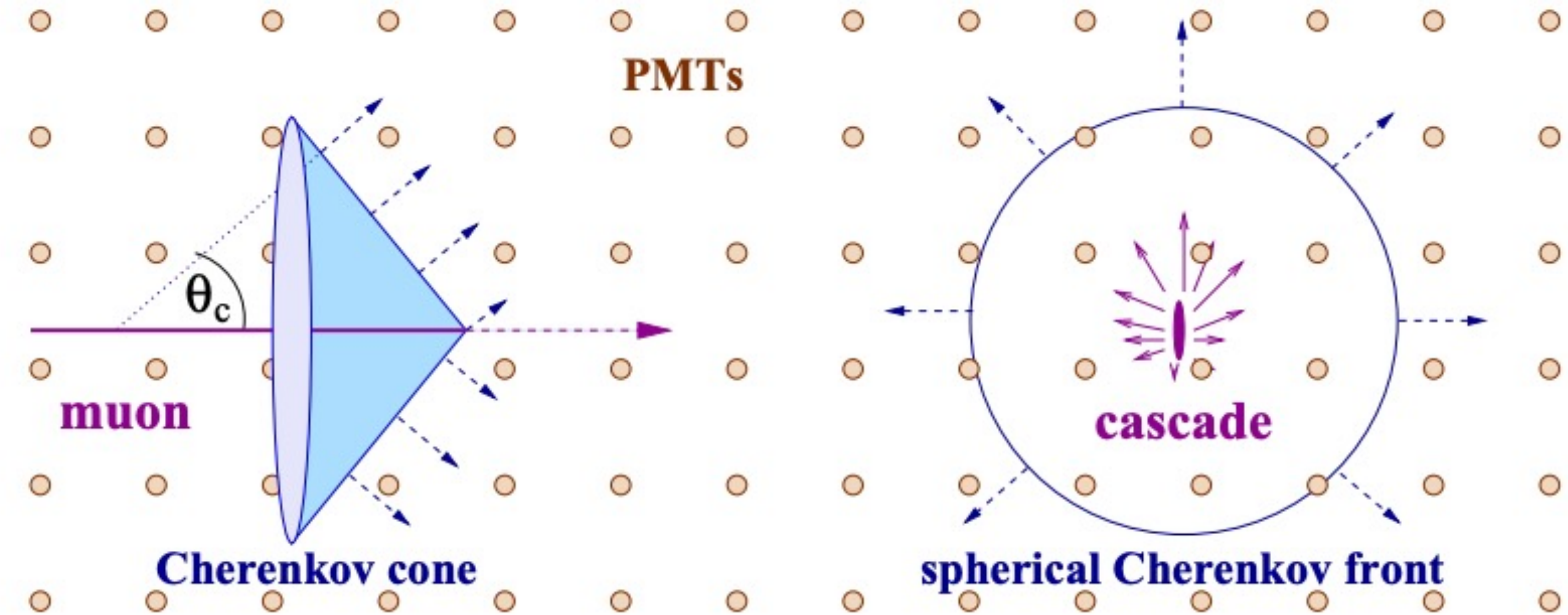
"double-bang"
(simulation)

~2 expected in 6 years

Tracks vs cascades

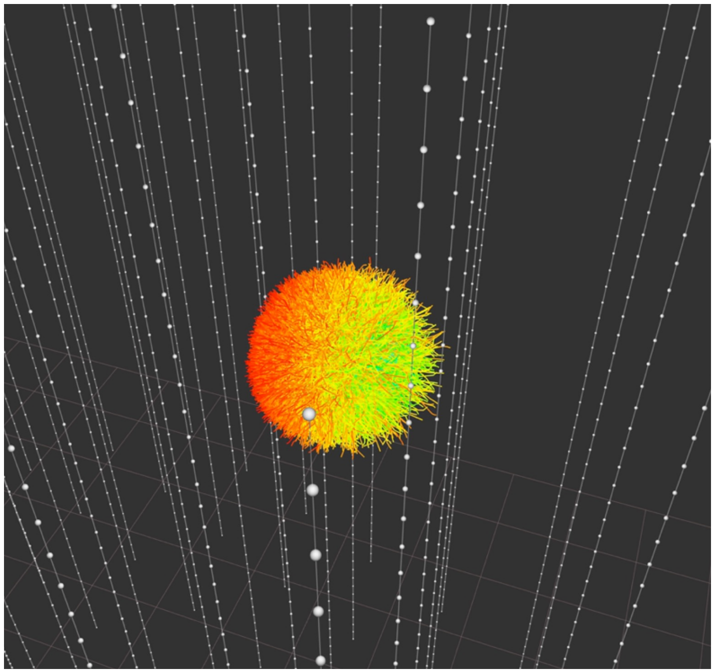
Tracks can travel large distance ~ first photons on Cherenkov cone

Cascades travel relatively short distance ~ diffuse photons w. spherical front



Event reconstruction

Emitted

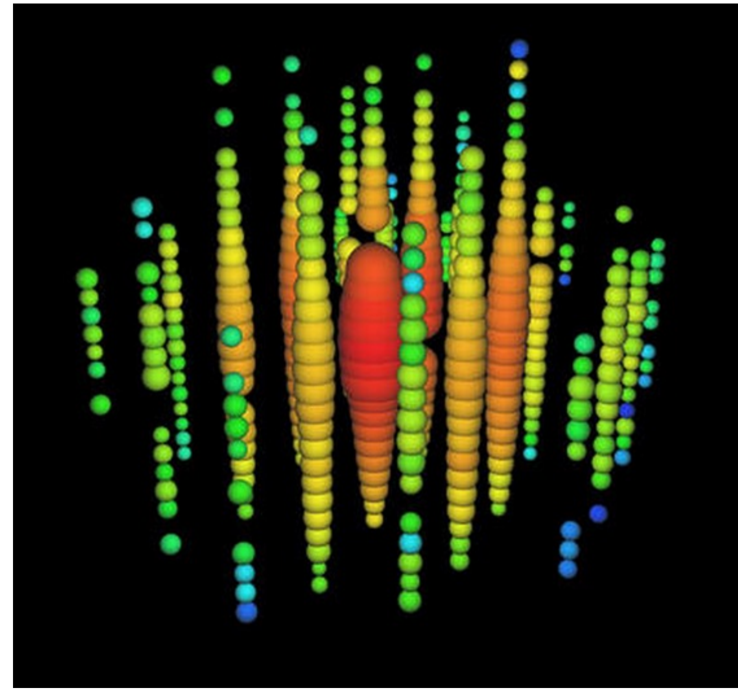


Asymmetry in photon emission helps with directional reconstruction

Information loss

$$\nu_e + N \rightarrow e + X$$
$$\nu_l + N \rightarrow \nu_l + X$$

Detected



Physics model (emission, absorption, scattering)

$\boldsymbol{\Theta} = (x, y, z, \theta, \phi, E, t)$

$X = q_{DOM}\,(t)$

Maximize $\mathcal{L}(\boldsymbol{\Theta}|\mathrm{X}_{\mathrm{Data}})$ or train ML algorithm

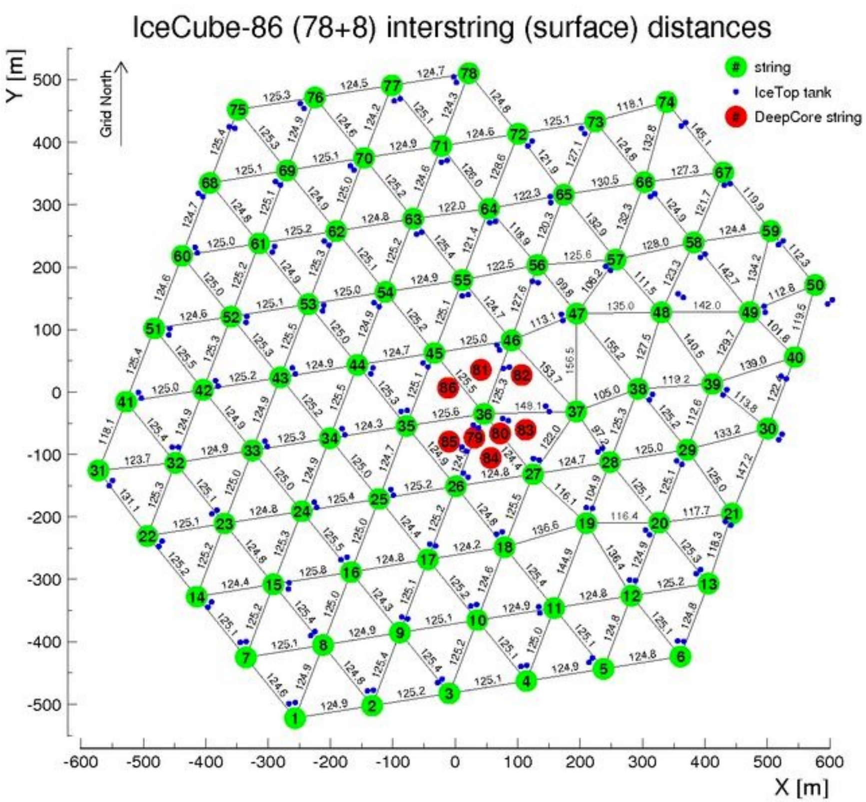
Physics parameters and IceCube coordinates

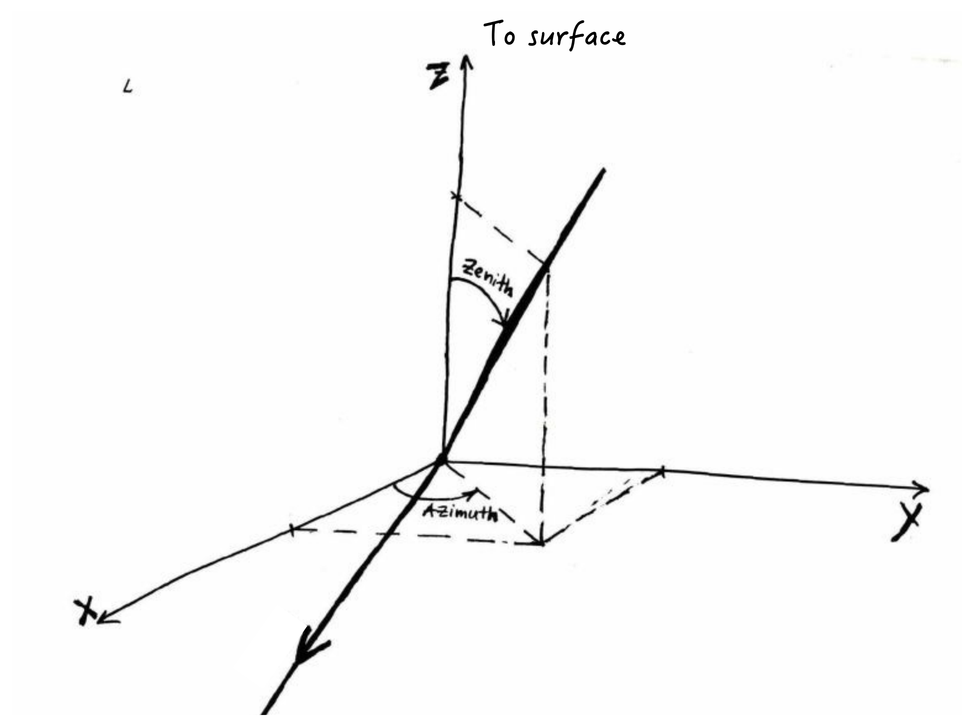
$$\mathbf{\Theta} = (x, y, z, \theta, \phi, E, t)$$

Detector coordinate system centered in middle of detector

(θ, ϕ) = (zenith, azimuth) corresponds to *arrival* direction

Usually, (θ, ϕ, E) are the physics parameters we're most interested in

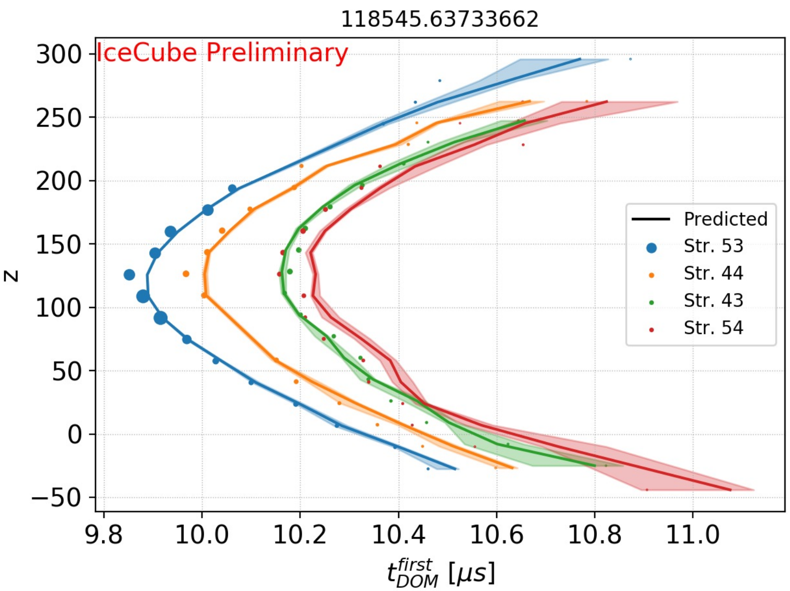


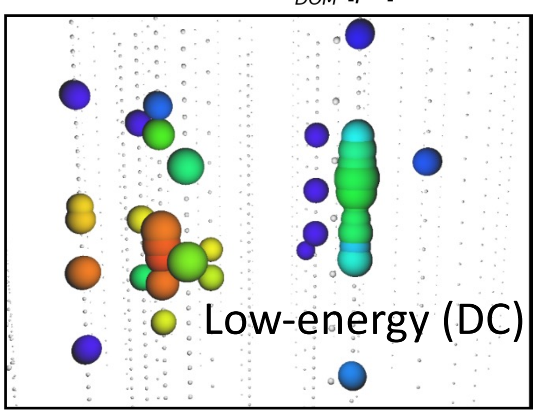


Approaches for reconstruction

Tracks

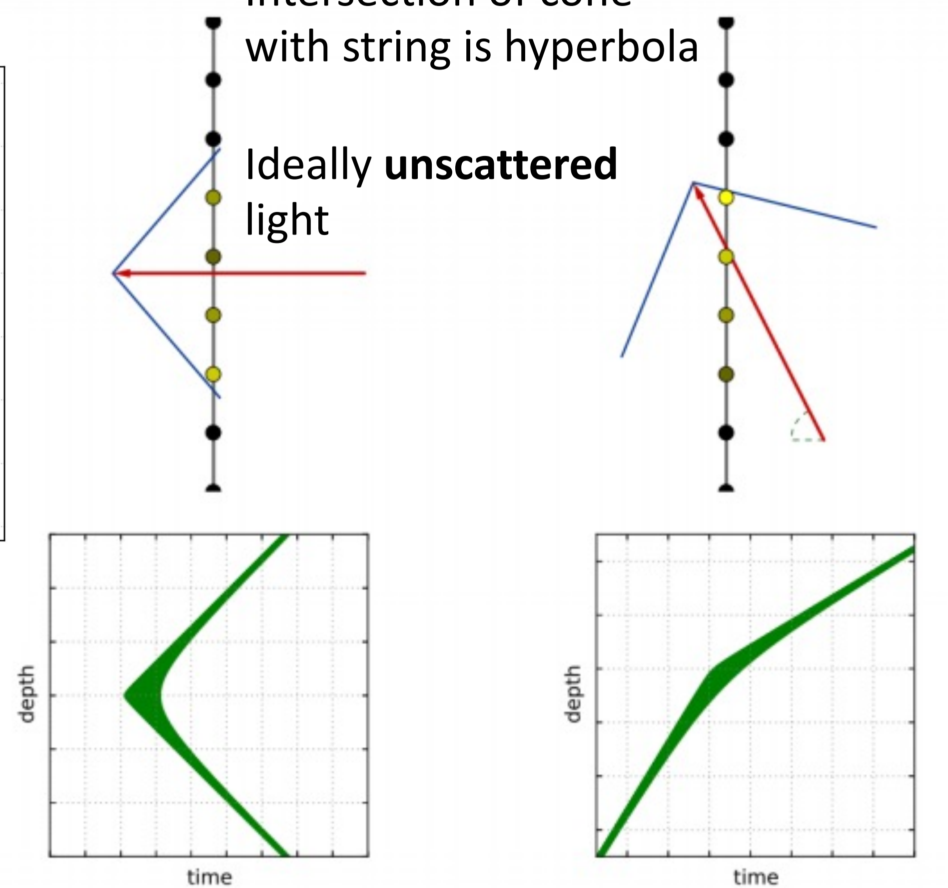
- Use **first-hit times** for *directional* reconstruction (SANTA, SplineReco)





Intersection of cone with string is hyperbola

Ideally **unscattered** light

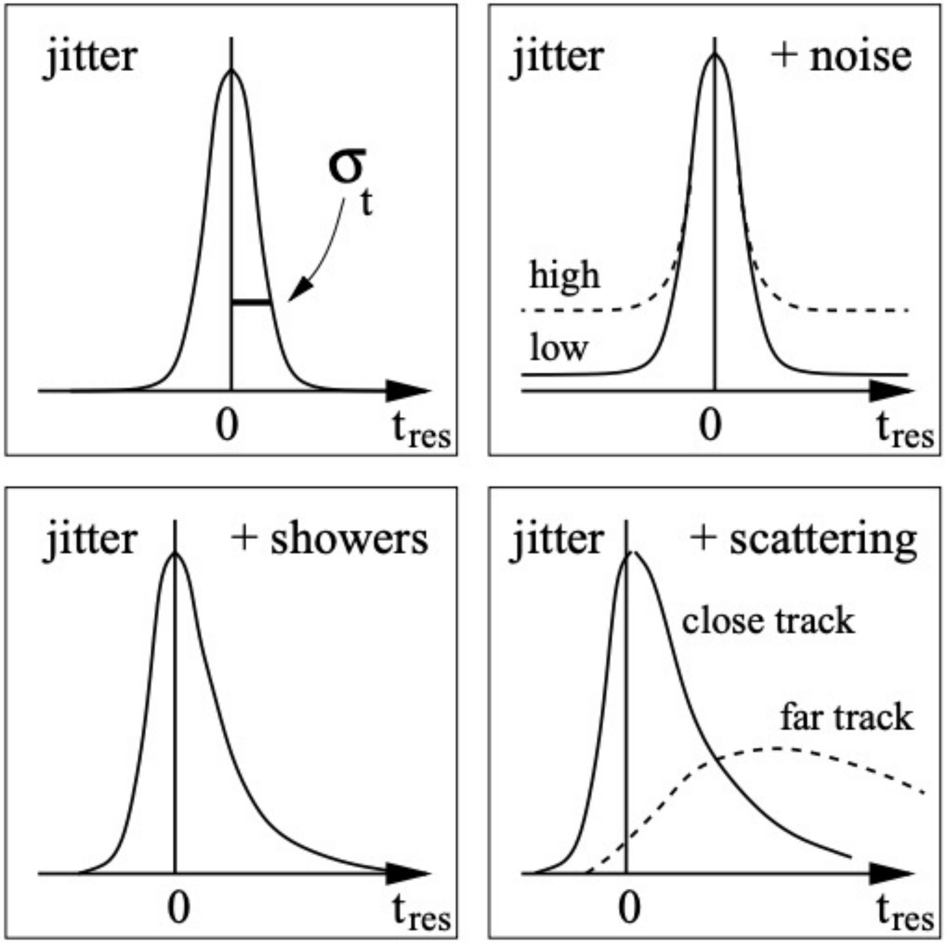


Juan Pablo Yanez

First-arrival time pdfs

PMT jitter (Transit time spread) due to spread in initial energies/momenta of photoelectrons

arxiv:0407044



Additional effects due to:

- noise
- additional cascades along track
- scattering

Original analytic parameterization

- “Pandel function” (gamma dist.)

$$p(t_{\rm res}) \equiv \frac{1}{N(d)} \frac{\tau^{-(d/\lambda)} \cdot t_{\rm res}^{(d/\lambda - 1)}}{\Gamma(d/\lambda)} \cdot e^{-\left(t_{\rm res} \cdot \left(\frac{1}{\tau} + \frac{c_{\rm medium}}{\lambda_a}\right) + \frac{d}{\lambda_a}\right)},$$

$$N(d) = e^{-d/\lambda_a} \cdot \left(1 + \frac{\tau \cdot c_{\rm medium}}{\lambda_a}\right)^{-d/\lambda},$$

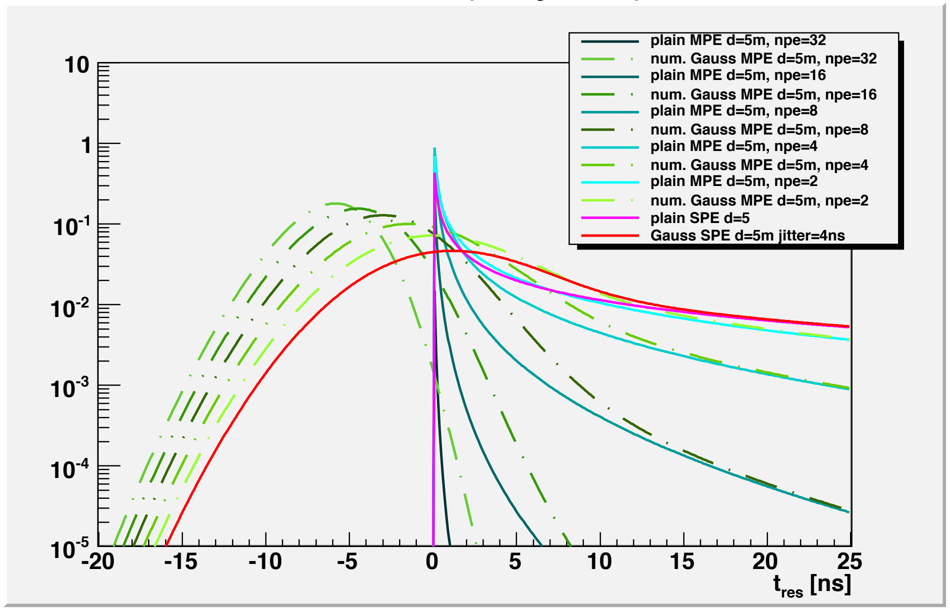
- Now based on splines

MPE Pandel likelihood

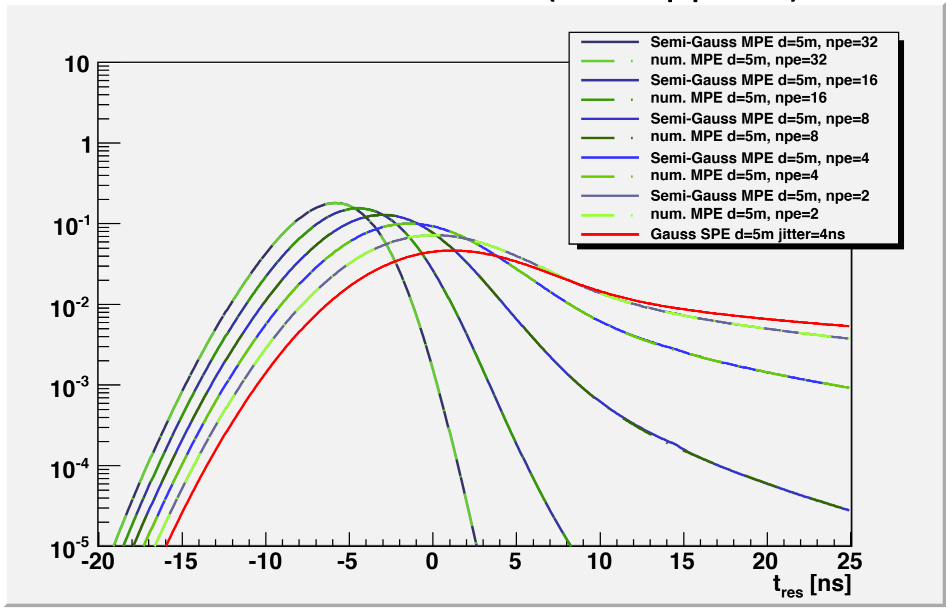
Pandel function cannot cope with negative time residuals so need to convolute with Gaussian

- https://user-web.icecube.wisc.edu/~boersma/PandelUpdates/MPEplots/

Plain MPE (no jitter)



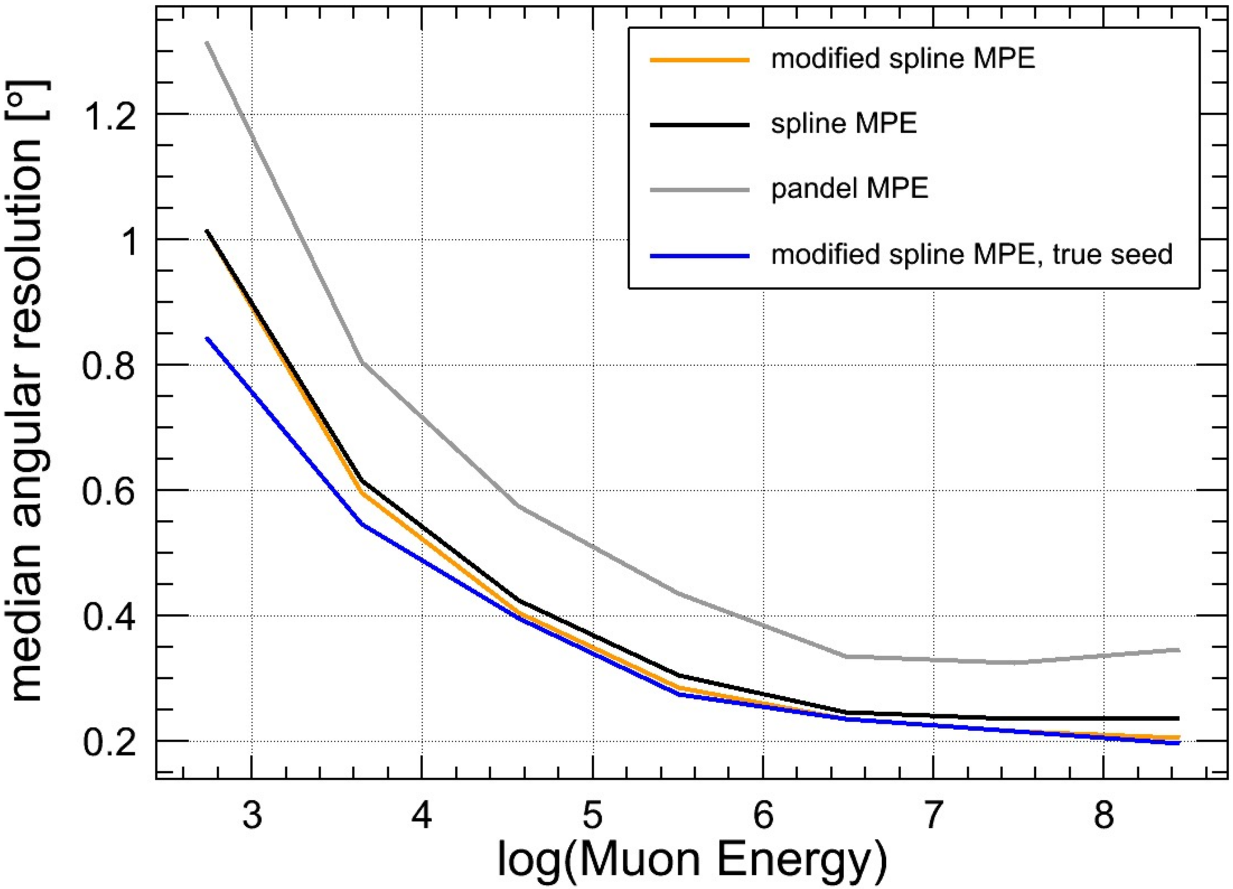
Gauss convoluted (fast-approx.)



SplineReco Resolutions

Improvements were made by moving to (photo)spline tables based on simulation (c.f. K. Schatto thesis)

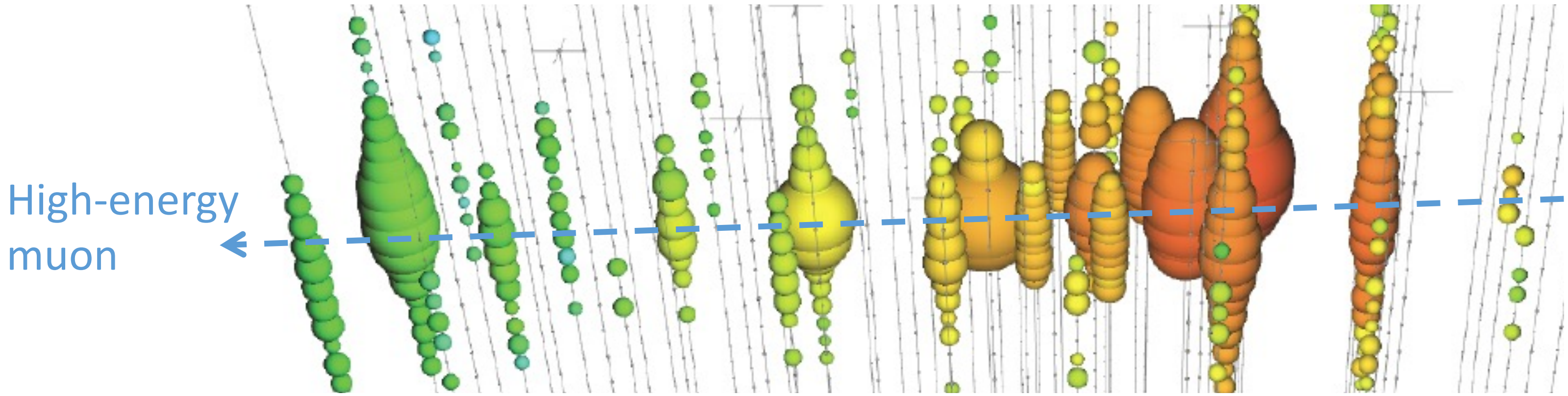
Fast 1D Gaussian convolution using IIR approximation



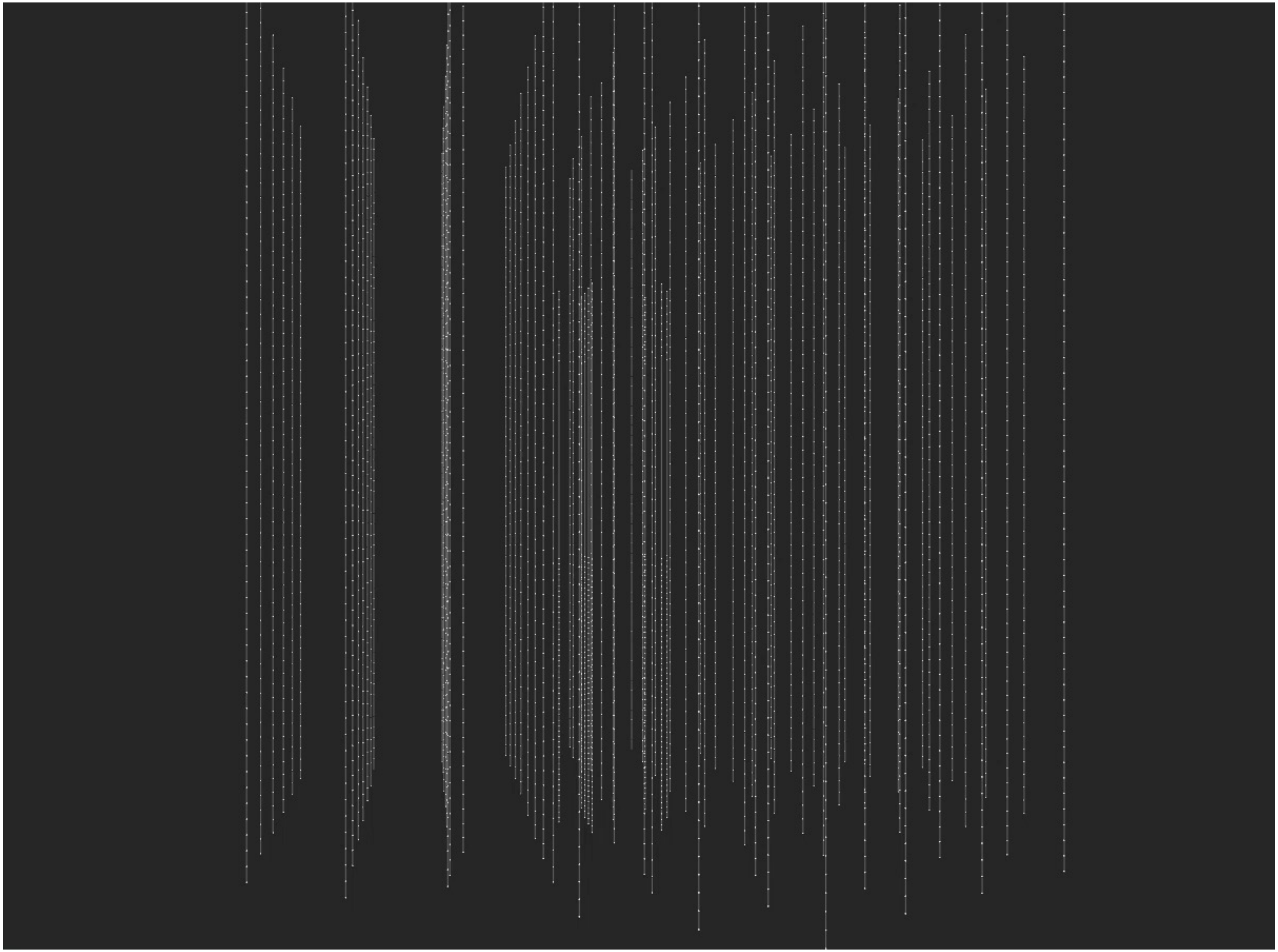
Approaches for reconstruction

Tracks

- Use **first-hit times** for *directional* reconstruction (SANTA, SplineReco)
- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, MuEx, Millipede, DirectFit)
  - Millipede works for **high-energy tracks** by breaking it up into multiple cascades along the track due to muon stochastic energy losses



Example reco for a data event



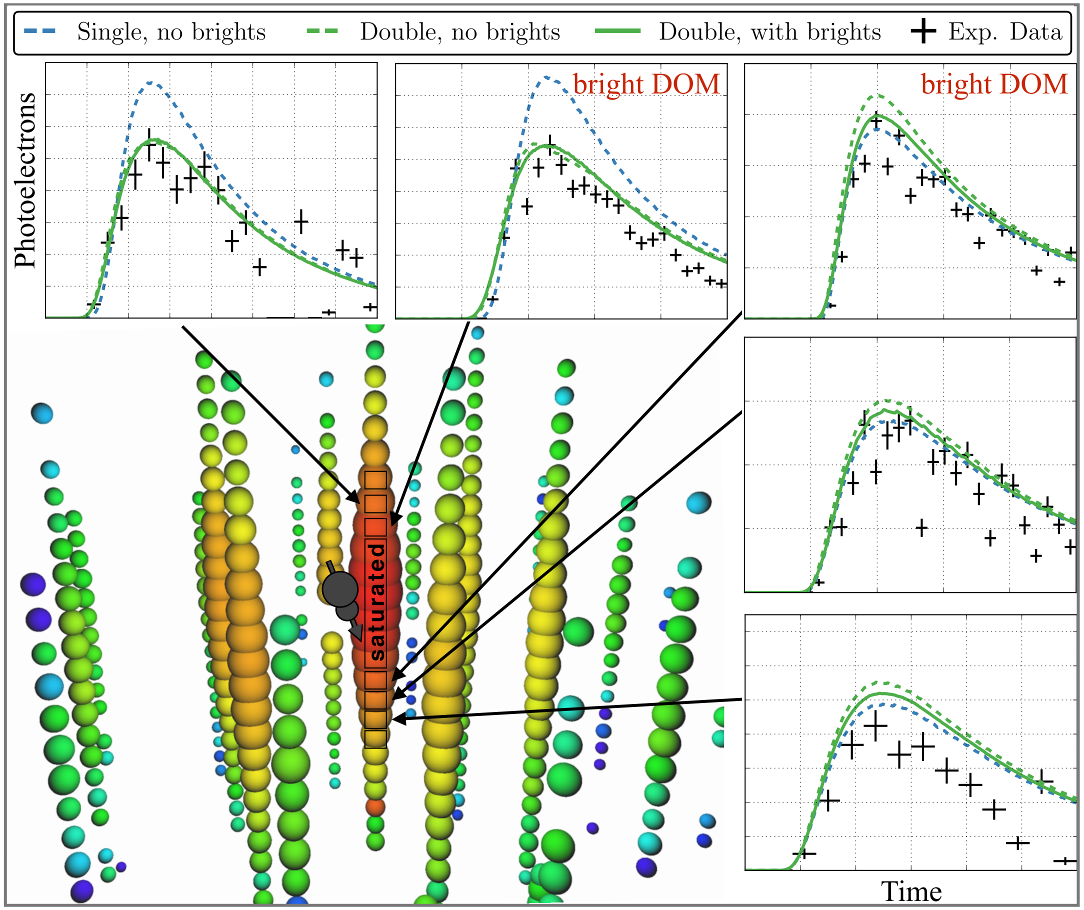
Two approaches to **full-waveform** reconstruction

Tabulated photon yields

- Pros: Fast runtime; gradients
- Cons: Limited ice-models

Direct photon propagation

- Pros: Any ice-model can be used
- Cons: Statistical errors from both data and MC; slow



IC collaboration, 1311.4767
D. Chirkin, arXiv:1304.0735

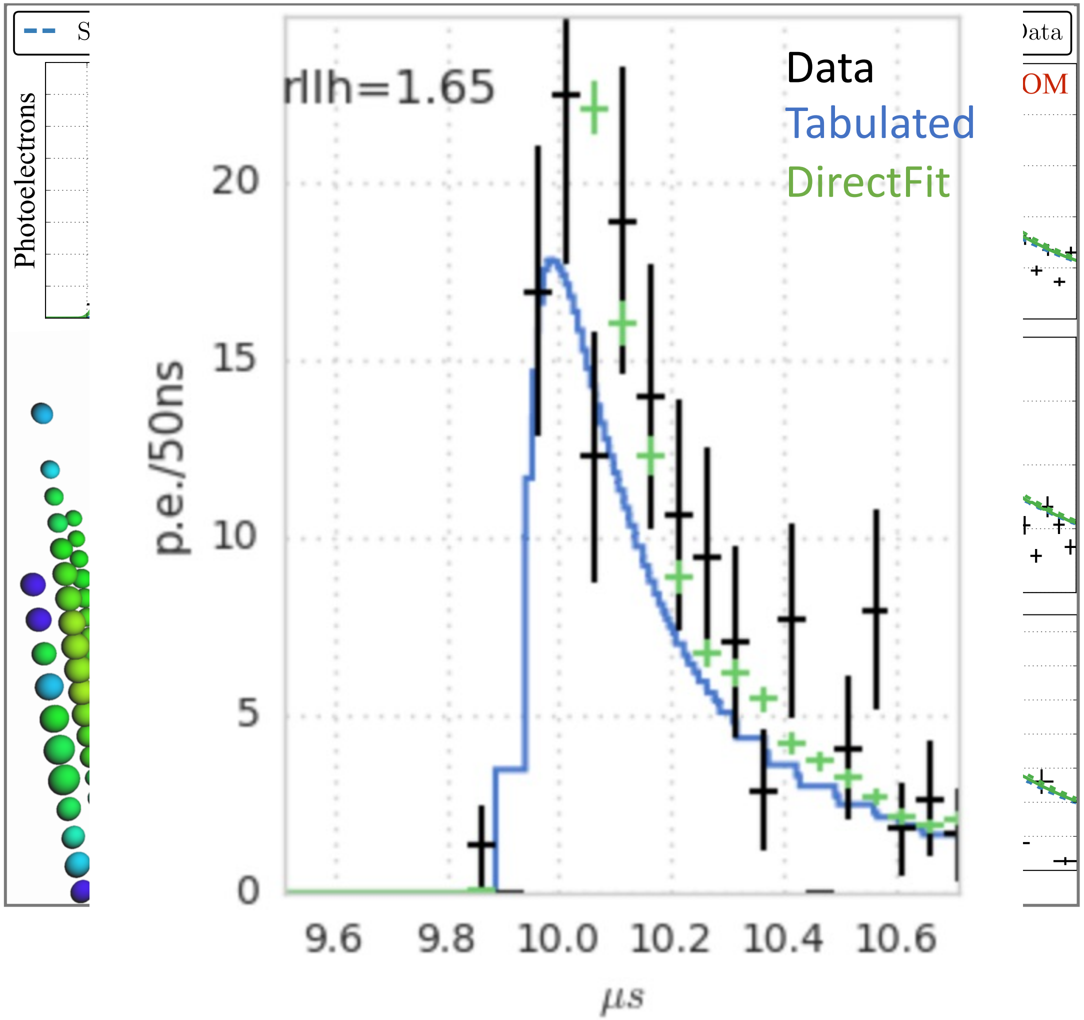
Two approaches to **full-waveform** reconstruction

Tabulated photon yields

- Pros: Fast runtime; gradients
- Cons: Limited ice-models

Direct photon propagation

- Pros: Any ice-model can be used
- Cons: Statistical errors from both data and MC; slow



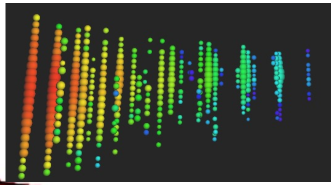
IC collaboration, 1311.4767
D. Chirkin, arXiv:1304.0735

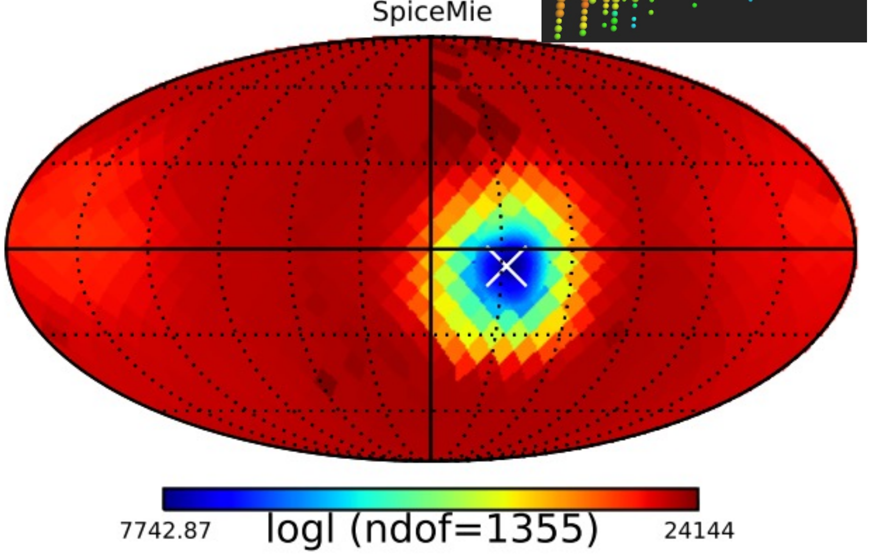
Minimization approaches

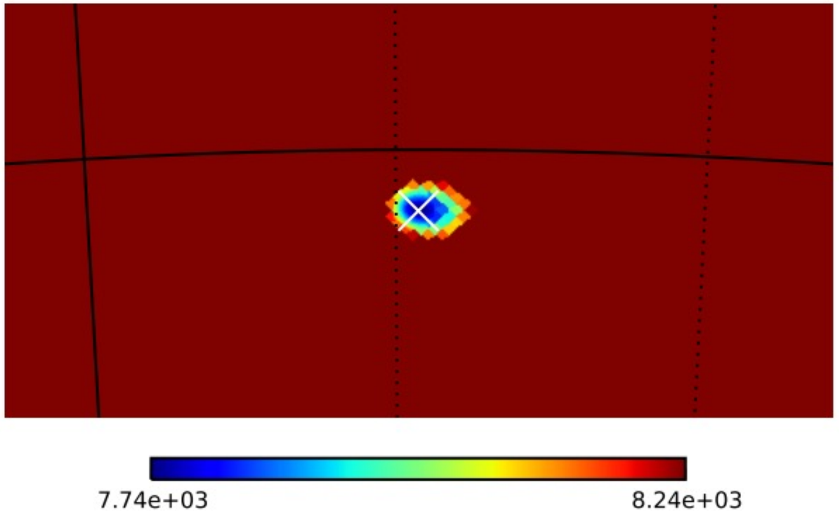
Given a likelihood $\mathcal{L}(\mathbf{\Theta}|\mathrm{X}_{\mathrm{Data}})$ as a function of $\mathbf{\Theta} = (x, y, z, \theta, \phi, E, t)$, want to find $\mathbf{\Theta_0}$ that minimizes the negative likelihood

- Millipede uses **photon tables** which allows for iterative gradient descent

- DirectFit reruns **photon simulation** which is more computationally intensive

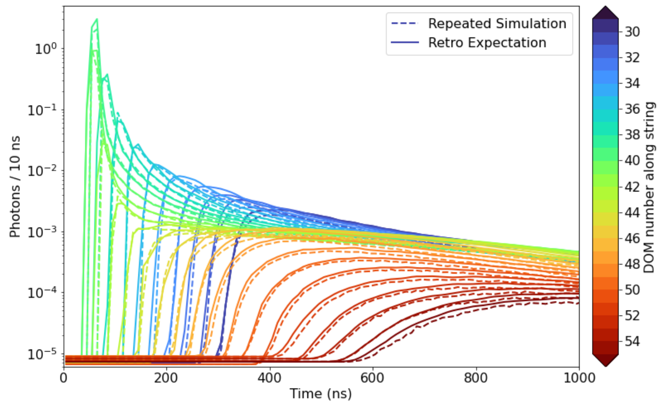


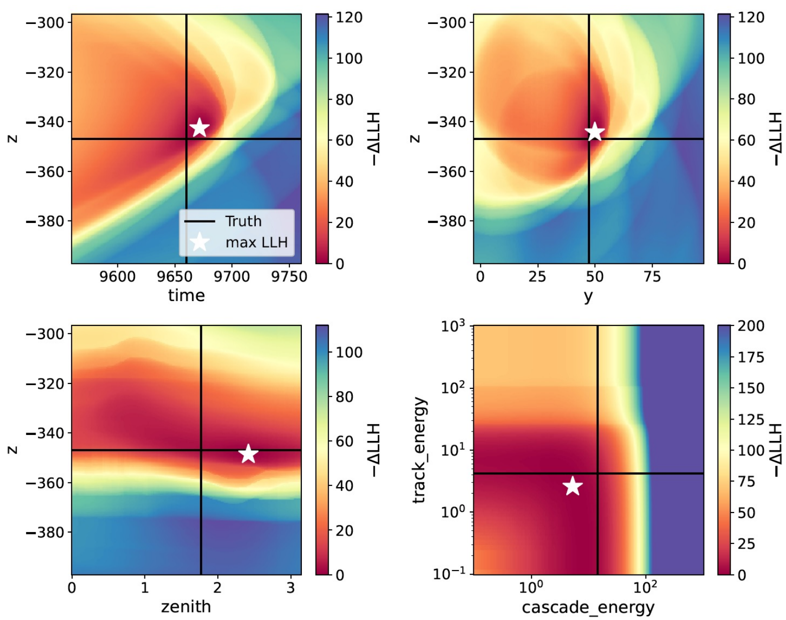




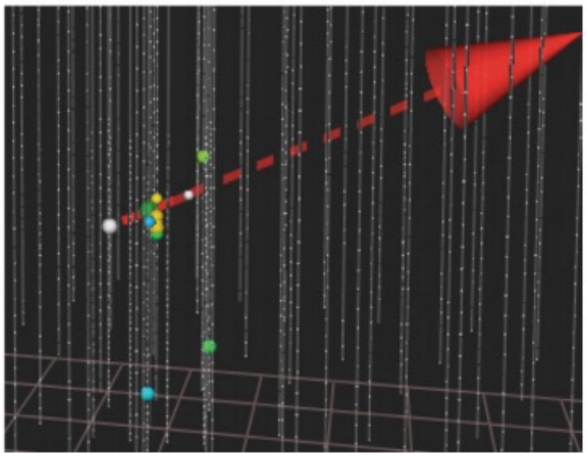
Low-energy reco: RetroReco and DirectReco

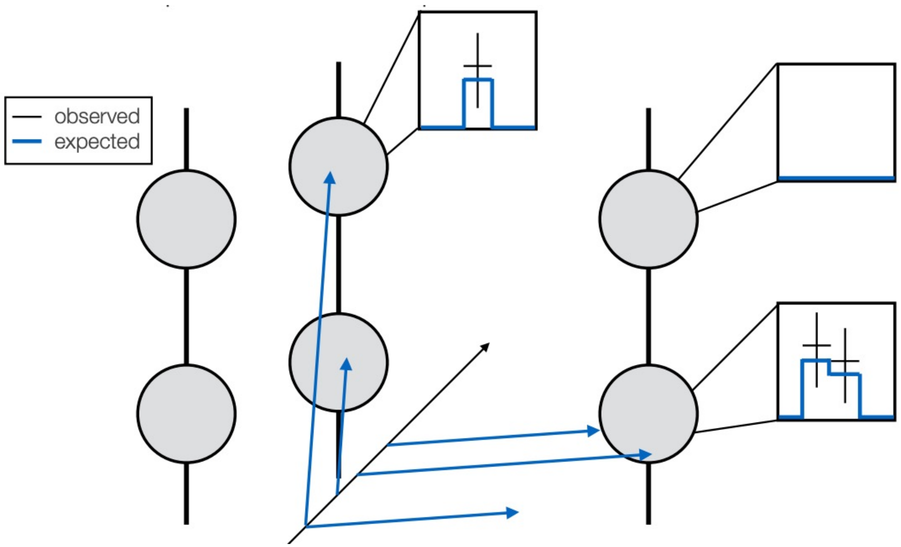
RetroReco: emit photons from DOM and track → then build retro tables





DirectReco: like DirectFit but for lower energies





Approaches for reconstruction

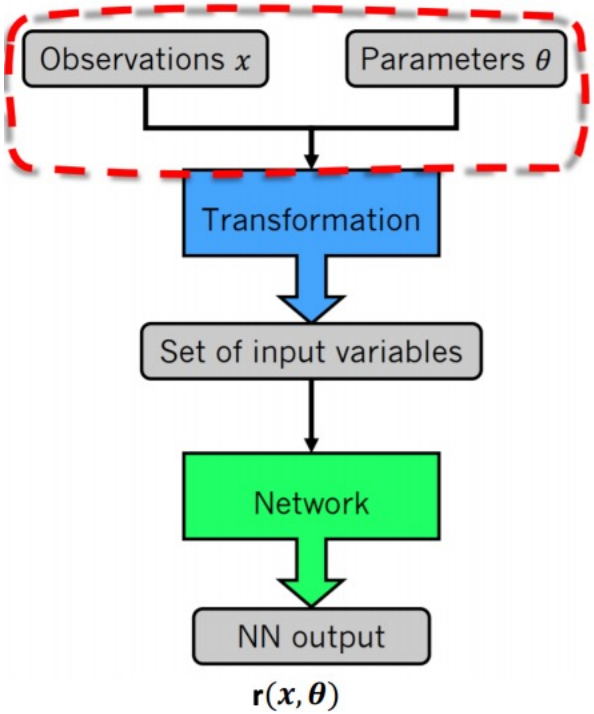
Tracks

- Use **first-hit times** for *directional* reconstruction (SANTA, SplineReco)
- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, MuEx, Millipede, DirectFit)
  - Millipede works for **high-energy tracks** by breaking it up into multiple cascades along the track due to muon stochastic energy losses
- ML+LLH approaches (CascadeGenerator)
- Likelihood-free inference (FreeDOM)
- Energy reco (TruncatedEnergy)

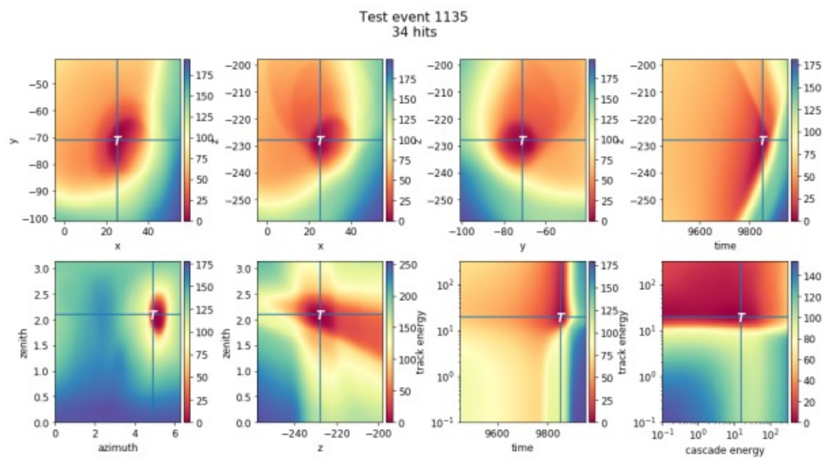
FreeDOM

Likelihood-free inference using NN

Train a binary classifier that can be converted back into a likelihood



- We replace $\frac{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x})}{p(\boldsymbol{x})}$ with the output of our neural network, $\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\theta})$
  - $\boldsymbol{r}$ is a ratio estimator; approximates the likelihood-to-evidence ratio
- $\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\theta})$ can be used anywhere you'd typically use a likelihood function
- Evaluating $\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\theta})$ is very fast (tens of microseconds)



T: true parameters

eight hypothesis parameters:
x, y, z, t, azimuth, zenith, cascade energy, track energy

parameters not being scanned are set to their truth values

Approaches for reconstruction

Tracks

- Use **first-hit times** for *directional* reconstruction (SANTA, SplineReco)
- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, MuEx, Millipede, DirectFit)
  - Millipede works for **high-energy tracks** by breaking it up into multiple cascades along the track due to muon stochastic energy losses
- ML+LLH approaches (CascadeGenerator)
- Likelihood-free inference (FreeDOM)
- Energy reco (TruncatedEnergy)

Cascades

- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, Monopod, DirectFit)

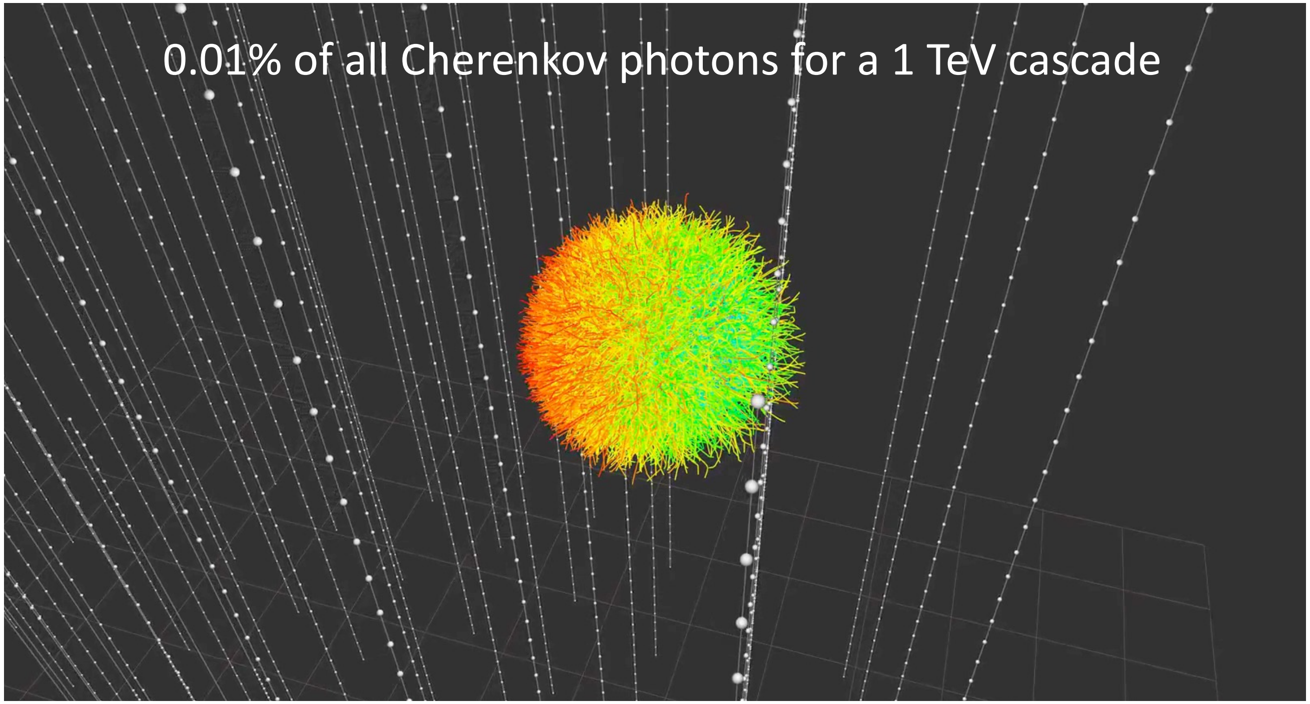
Challenges in cascade reconstruction

Large distances between DOMs means not many detected photons

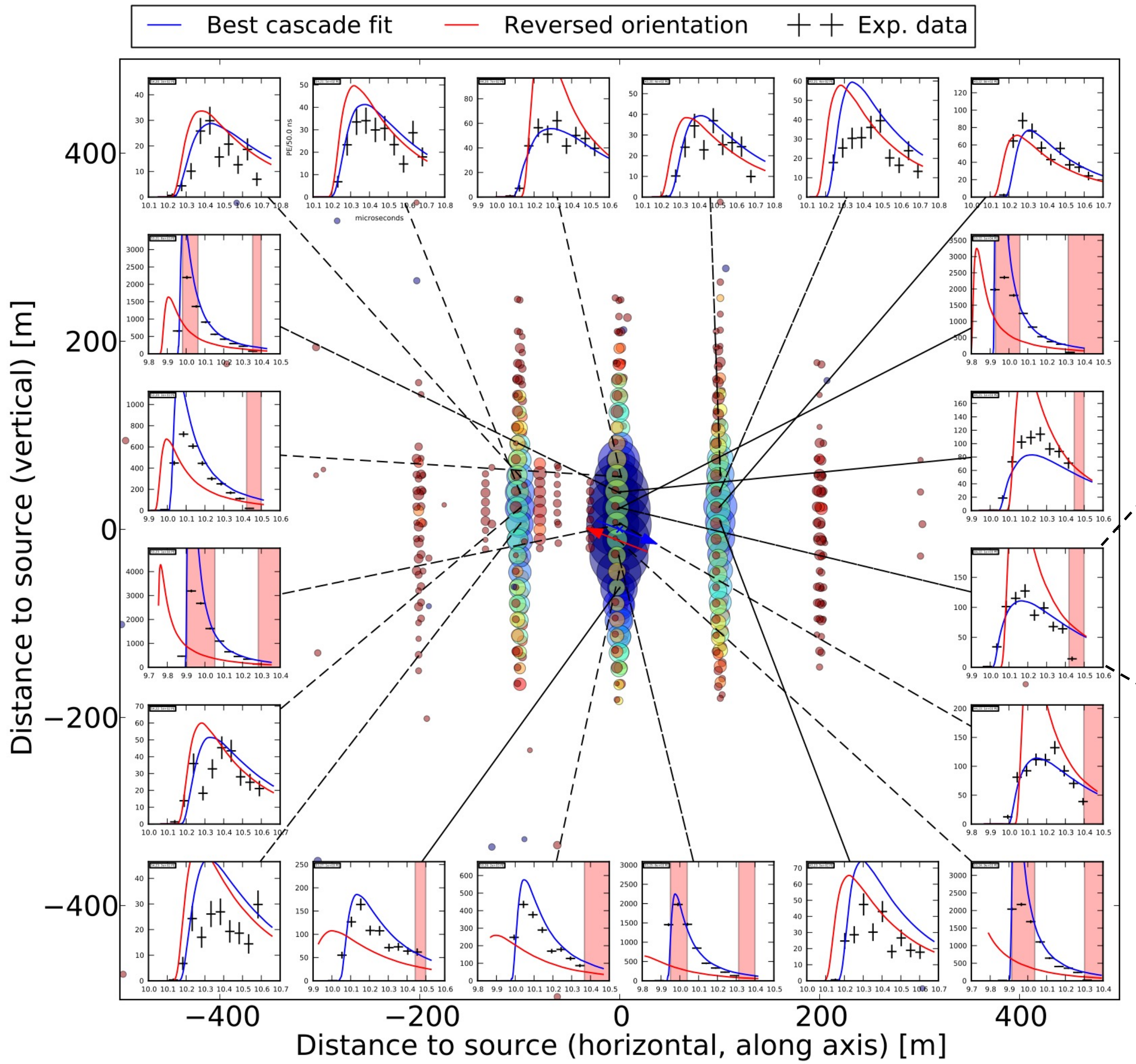
Small asymmetry means high dependence on ice modeling

Sheer number of photons difficult to simulate

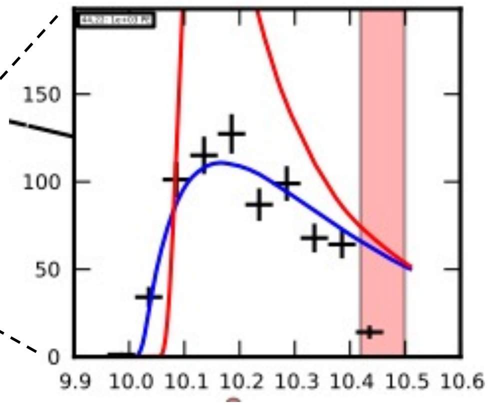
1. Tabulate photon yields for a single ice model (Millipede/Monopod)
   - Fast, less flexible, table generation time-consuming
2. Directly propagate all photons for any ice model (DirectFit)
   - Slow, more flexible



Cascade orientation from full-waveform



Differences between best-fit and reversed-orientation from Monopod



Time-windows where PMT saturates or calibration failed are shaded

Minimization approaches

Given a likelihood $\mathcal{L}(\mathbf{\Theta}|\mathrm{X}_{\mathrm{Data}})$ as a function of $\mathbf{\Theta} = (x, y, z, \theta, \phi, E, t)$, want to find $\mathbf{\Theta_0}$ that minimizes the negative likelihood

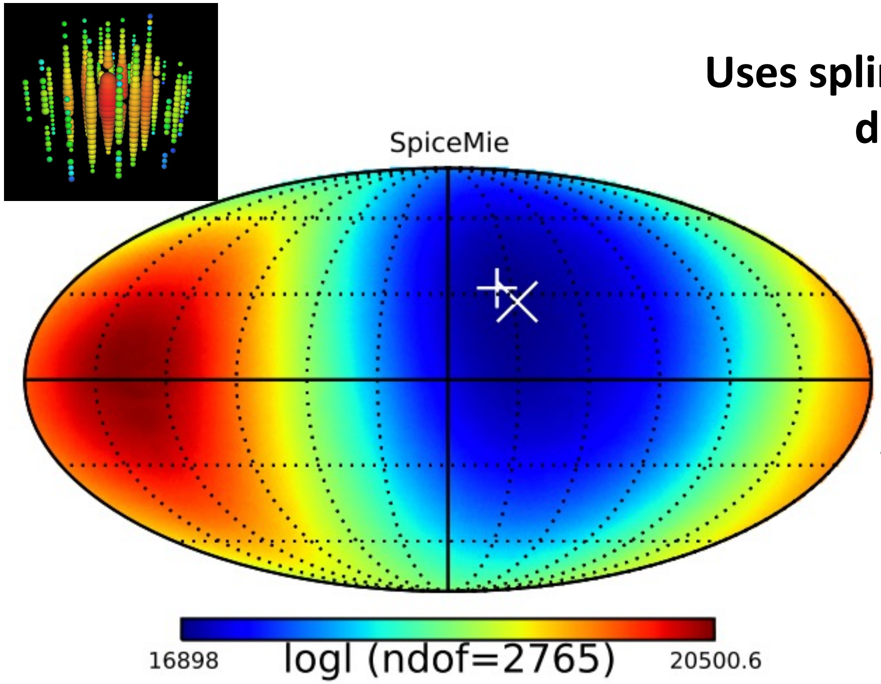
Need to explore 7D space which is challenging

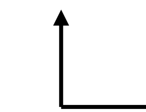
- Monopod uses **photon tables** which allows for iterative gradient descent
  - Doesn't always find the global minimum

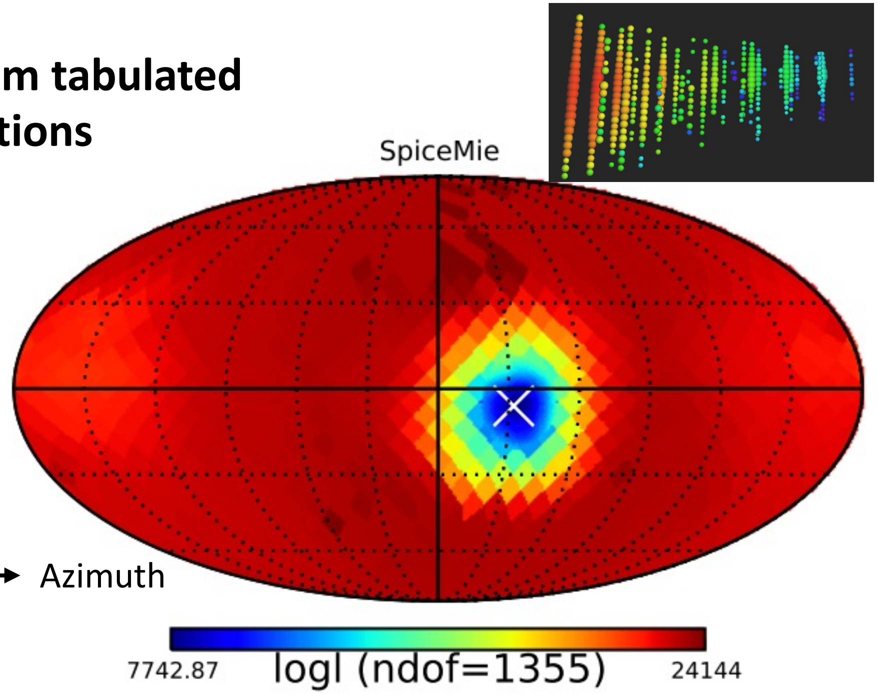
- Can also **brute force** all possible directions (θ, ϕ) to reduce the minimization to only 5 dimensions

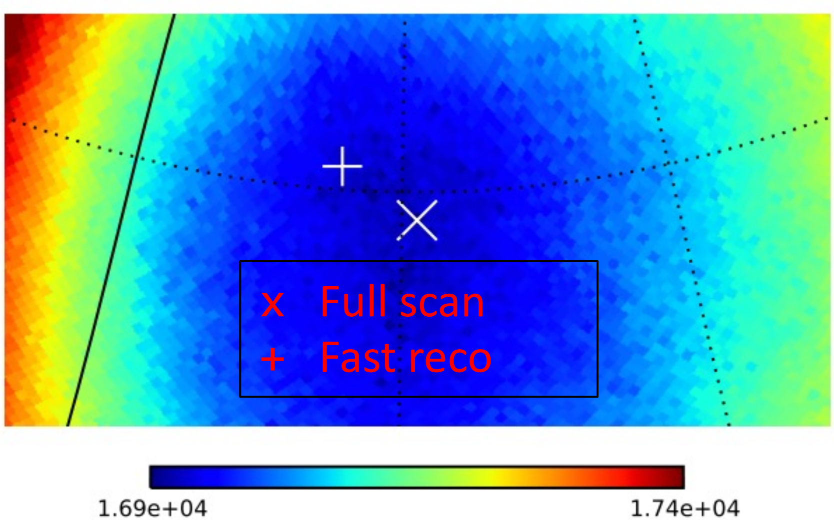
Cascade vs track skymap

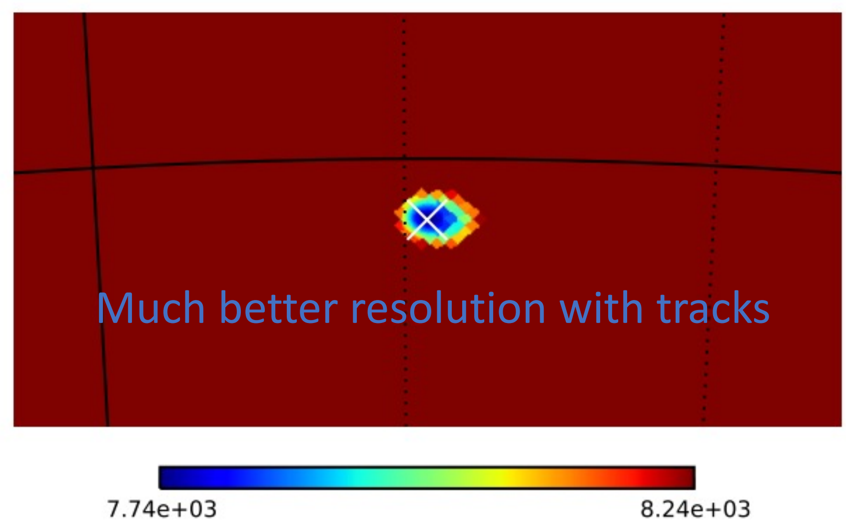
Uses splines from tabulated distributions







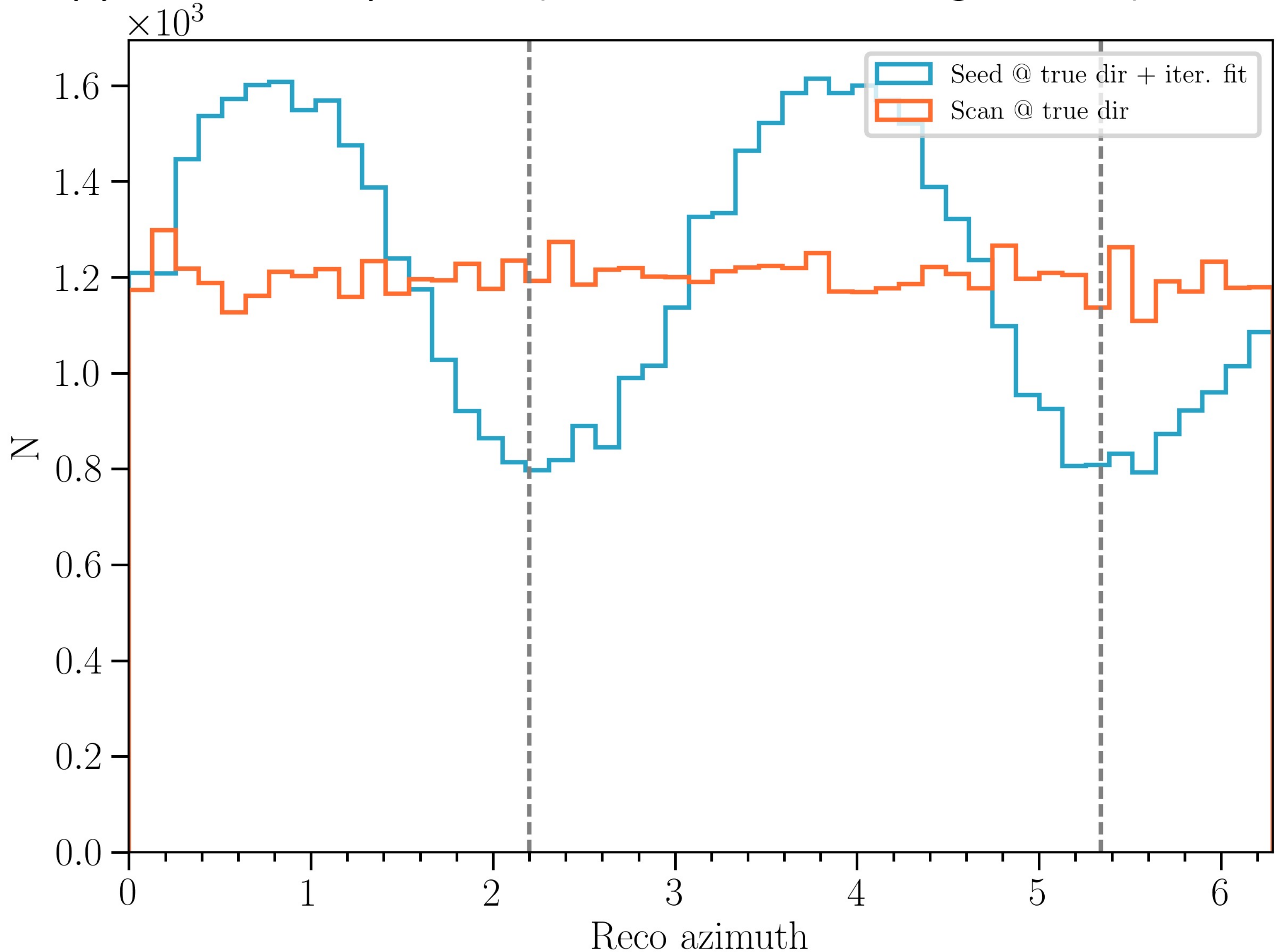




Caveats with tables-based reco of cascades

Spice 3.2 simulation reco'd with Spice Lea tables + effective distance

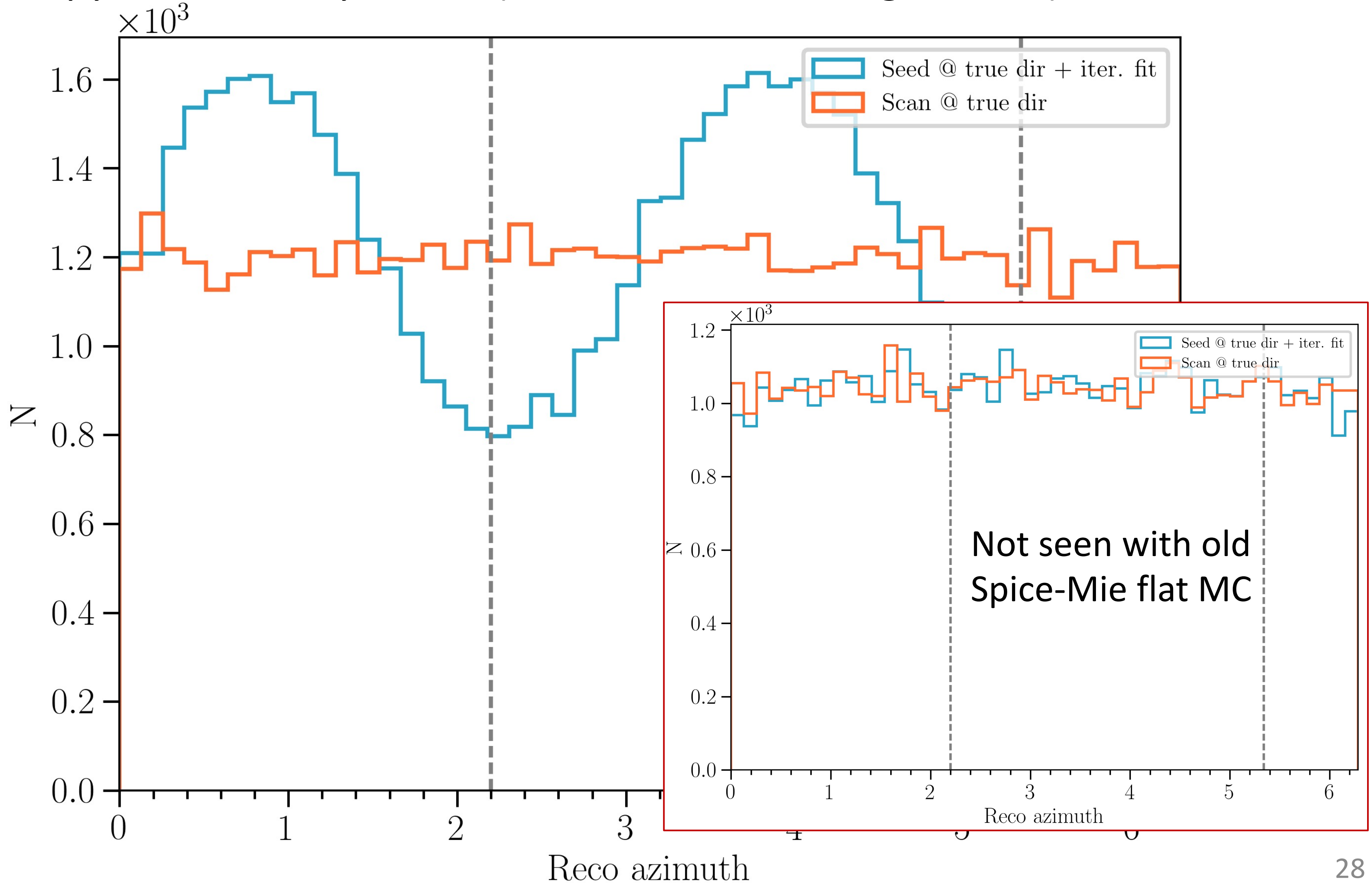
Anisotropy effect clearly visible (also seen when using Lea MC)



Caveats with tables-based reco of cascades

Spice 3.2 simulation reco'd with Spice Lea tables + effective distance

Anisotropy effect clearly visible (also seen when using Lea MC)



Minimization approaches

Given a likelihood $\mathcal{L}(\mathbf{\Theta}|\mathrm{X}_{\mathrm{Data}})$ as a function of $\mathbf{\Theta} = (x, y, z, \theta, \phi, E, t)$, want to find $\mathbf{\Theta_0}$ that minimizes the negative-likelihood

Need to explore 7D space which is challenging

- Millipede/Monopod uses **photon tables** which allows for iterative gradient descent
  - Doesn't always find the global minimum

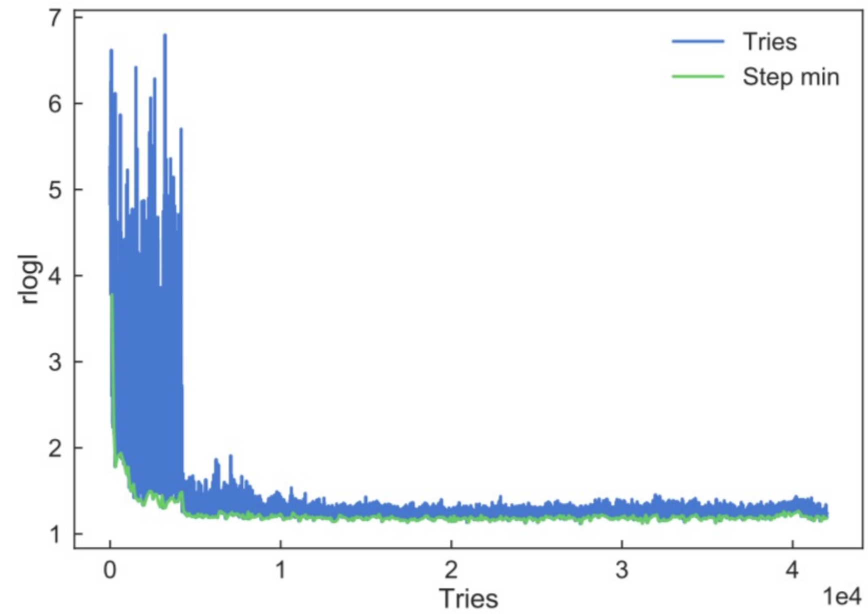
- Can also **brute force** all possible directions (θ, ϕ) to reduce the minimization to only 5 dimensions

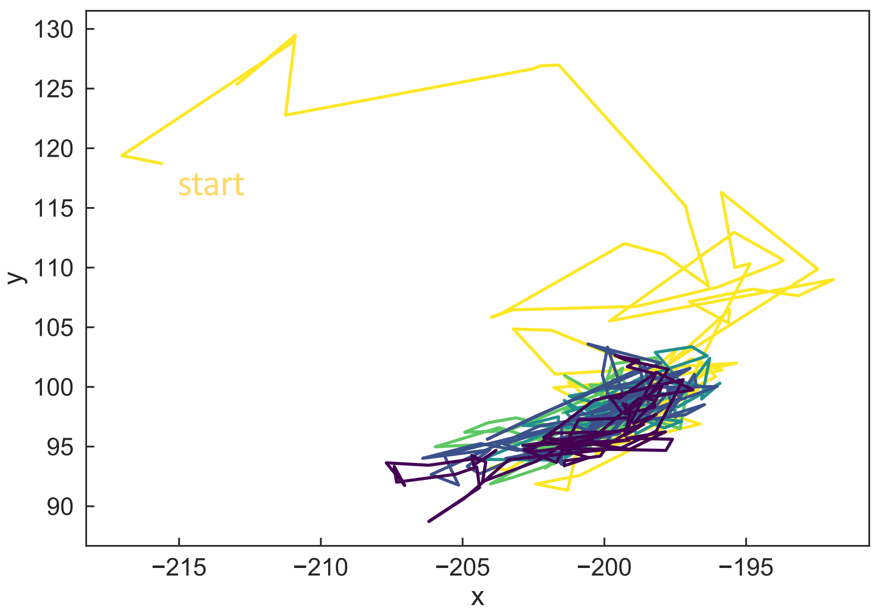
- DirectFit attempts to find minimum using **localized random search**, randomly sampling points in (x, y, z, θ, ϕ) within a "search radius" that is refined iteratively

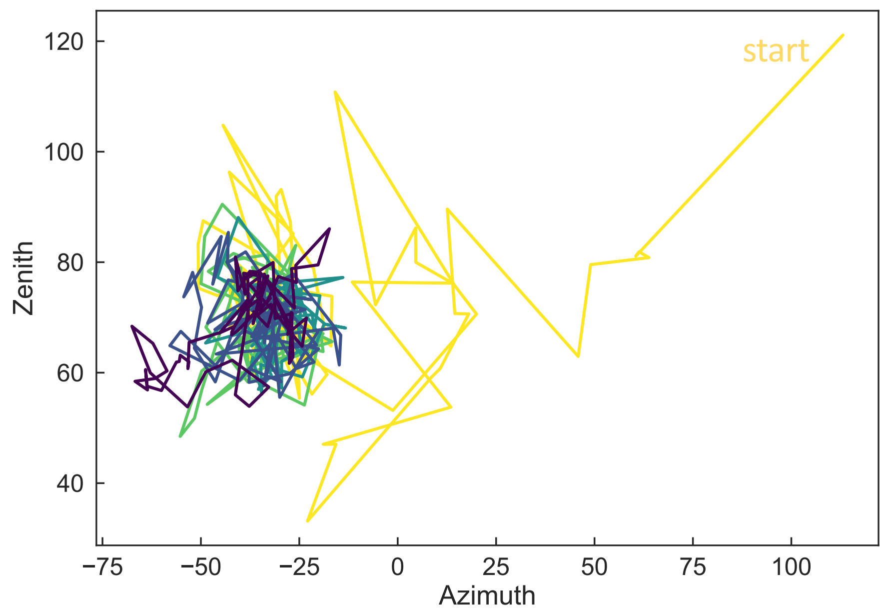
DirectFit minimization

Likelihood gradually improves from start to finish

Following this, MCMC approach to sample from posterior pdf

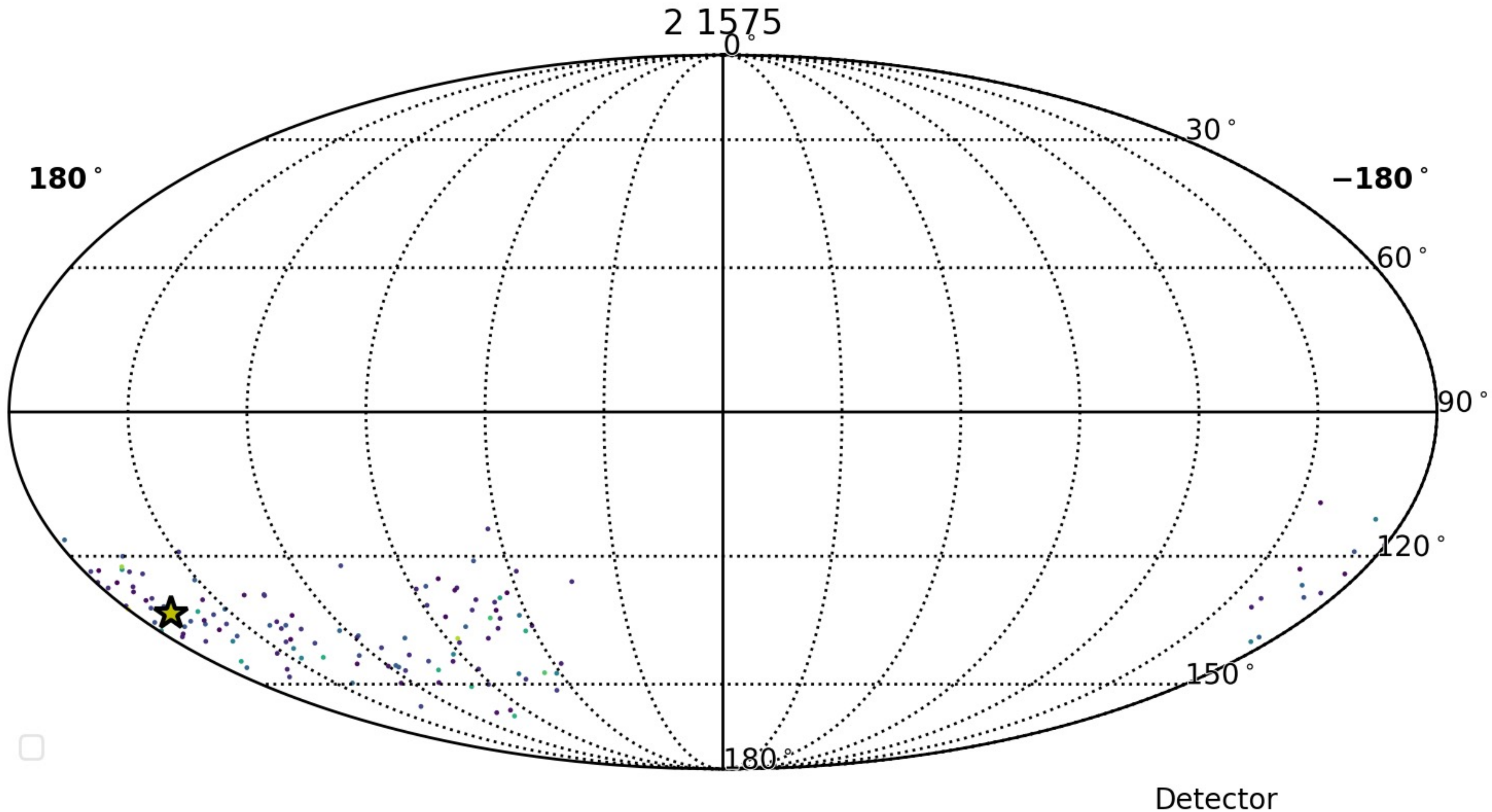






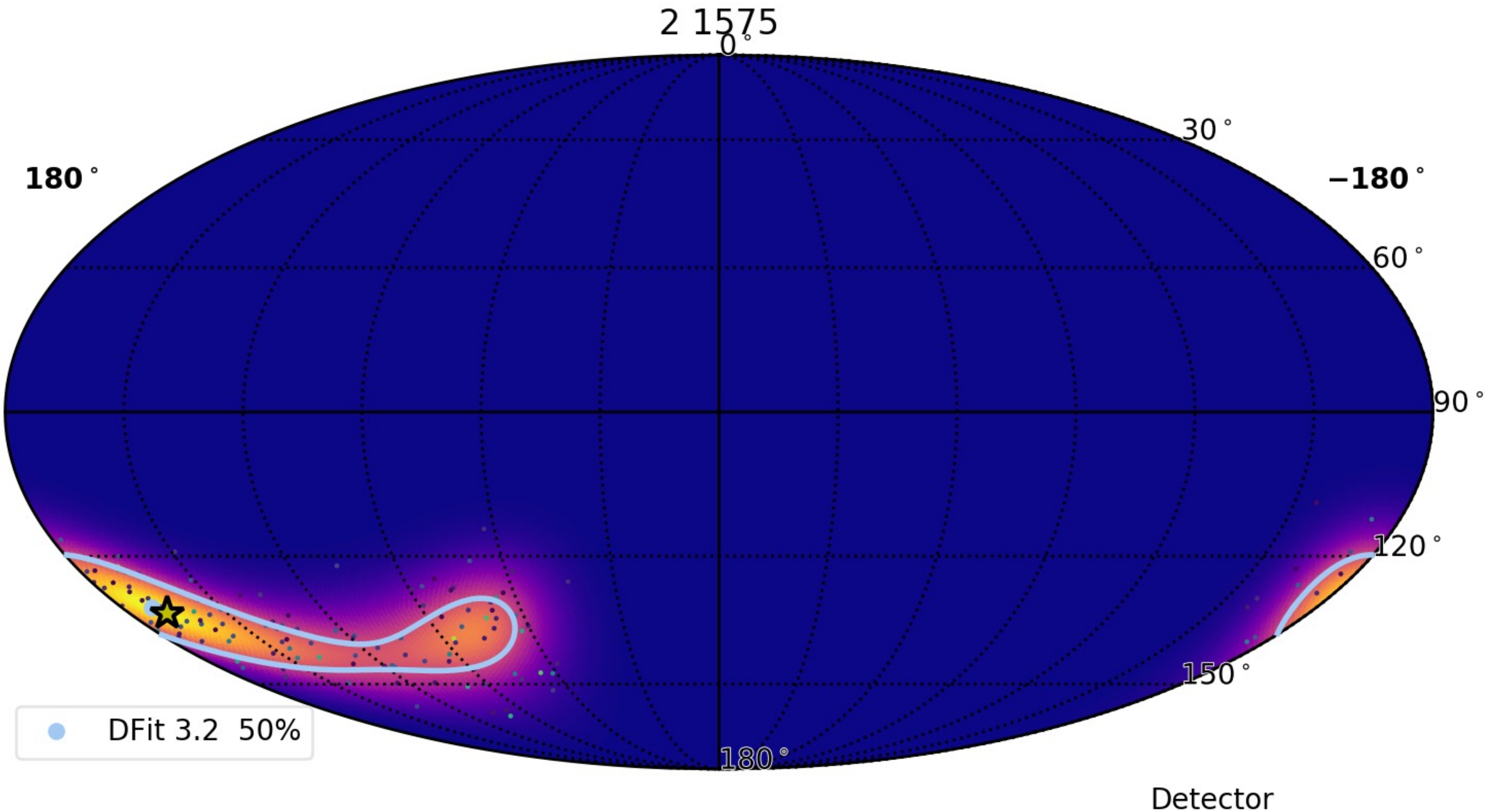
DirectFit with directional PDFs

ABC outputs points on unit sphere (simulated event)



DirectFit with directional PDFs

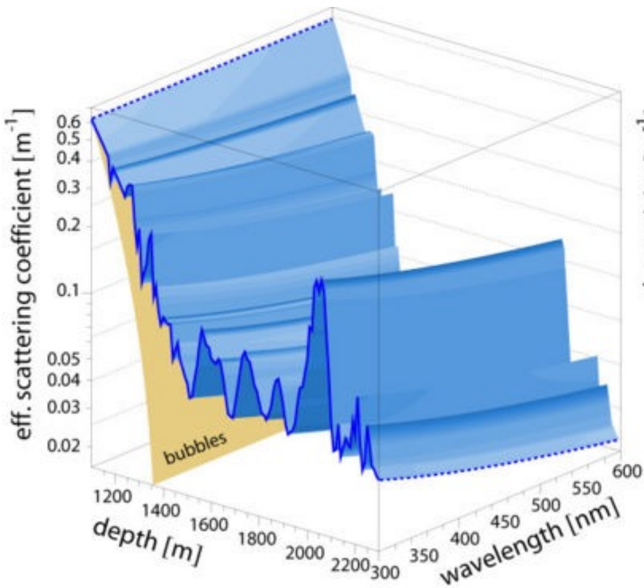
ABC outputs points on unit sphere (simulated event)



Can then fit a PDF on a sphere to those points

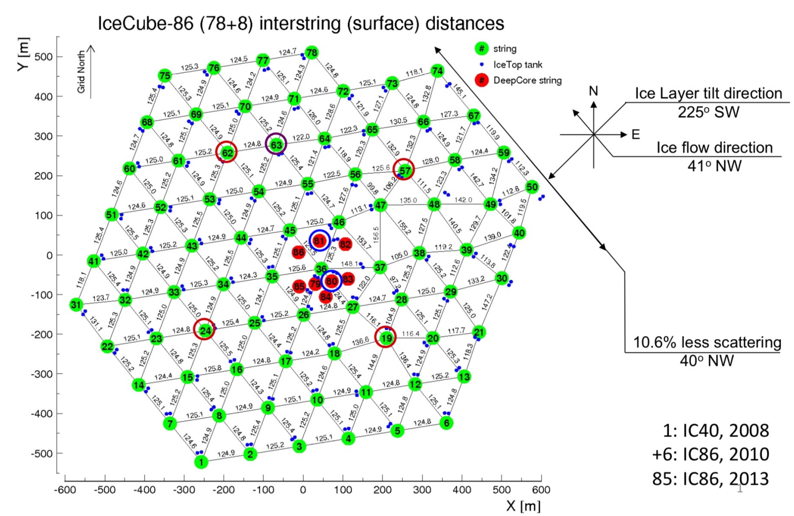
Ice modeling is important!

Bulk ice described by scattering and absorption coefficients as a function of depth → these have been refined over time



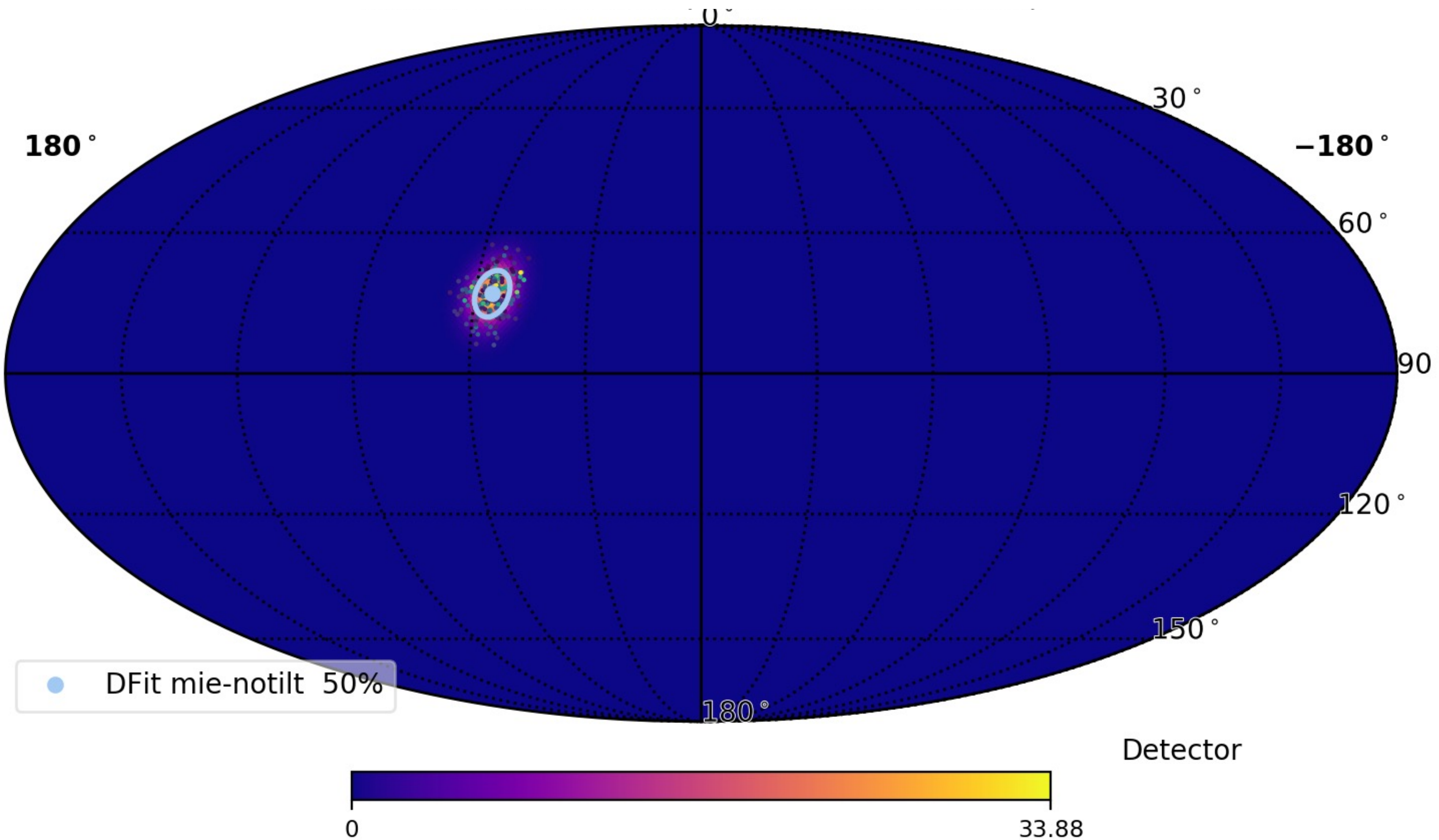
Ice layers were found to be tilted [arXiv:1301.5361]

Ice was also discovered to be anisotropic [ICRC 2013, 0580]



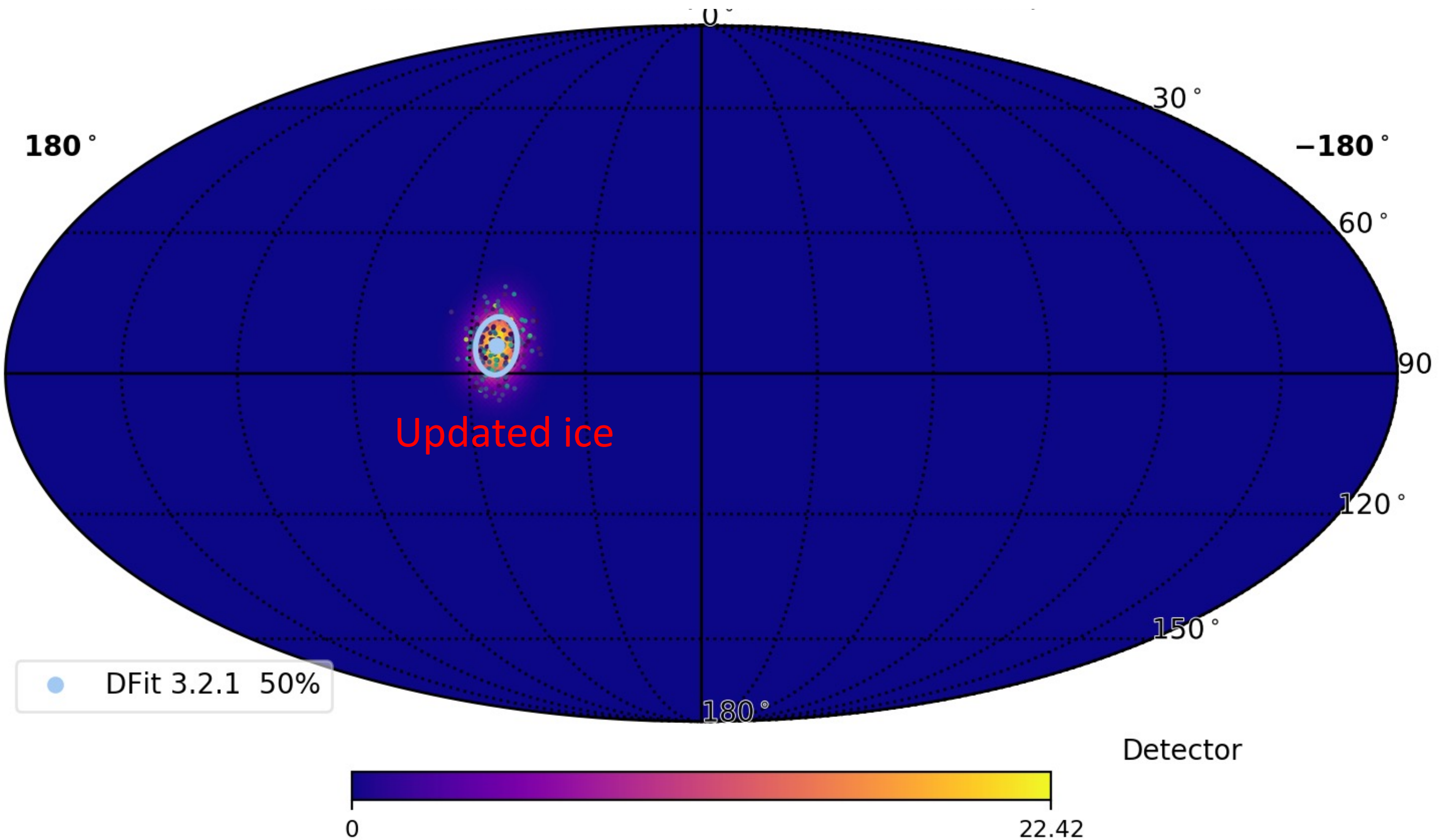
Directional bias due to different ice models

Ice affects cascade reconstruction



Directional bias due to different ice models

Ice affects cascade reconstruction

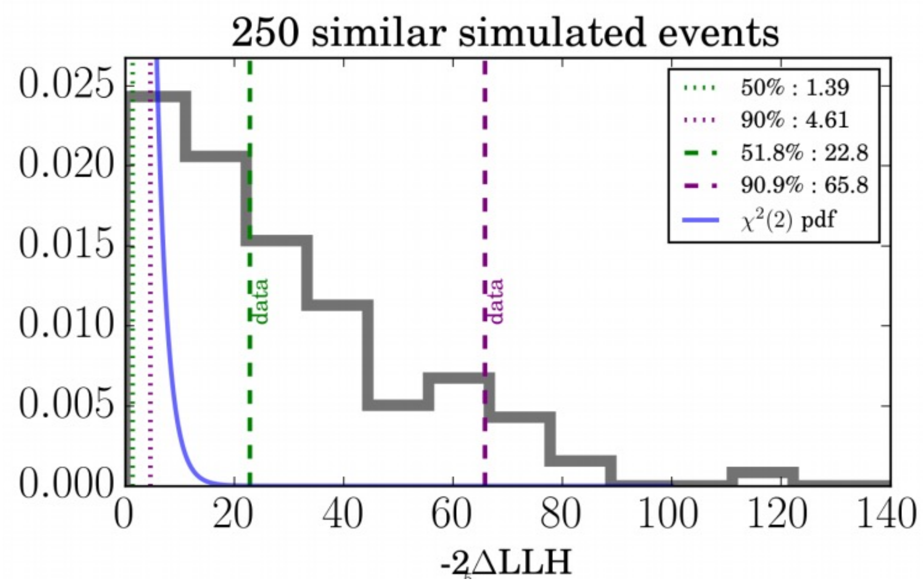


Uncertainty estimation

Ice uncertainties affect reconstructed directions

Directional uncertainties important for point-source searches

With Millipede/Monopod full-sky scan, can draw a contour at some value of Δllh derived from resimulations with different ice models



With DirectFit, can reconstruct with different ice-models and combine into larger contour

Approaches for reconstruction

Tracks

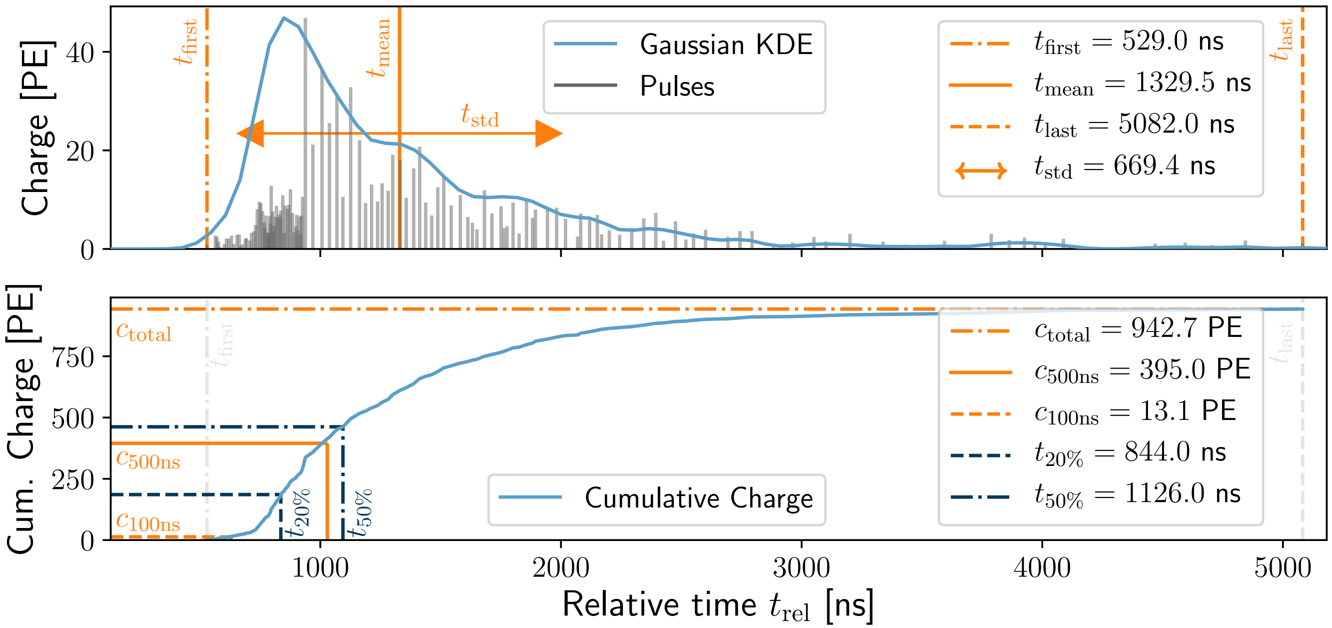
- Use **first-hit times** for *directional* reconstruction (SANTA, SplineReco)
- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, MuEx, Millipede, DirectFit)
    - Millipede works for **high-energy tracks** by breaking it up into multiple cascades along the track due to muon stochastic energy losses
- ML+LLH approaches (CascadeGenerator)
- Likelihood-free inference (FreeDOM)
- Energy reco (TruncatedEnergy)

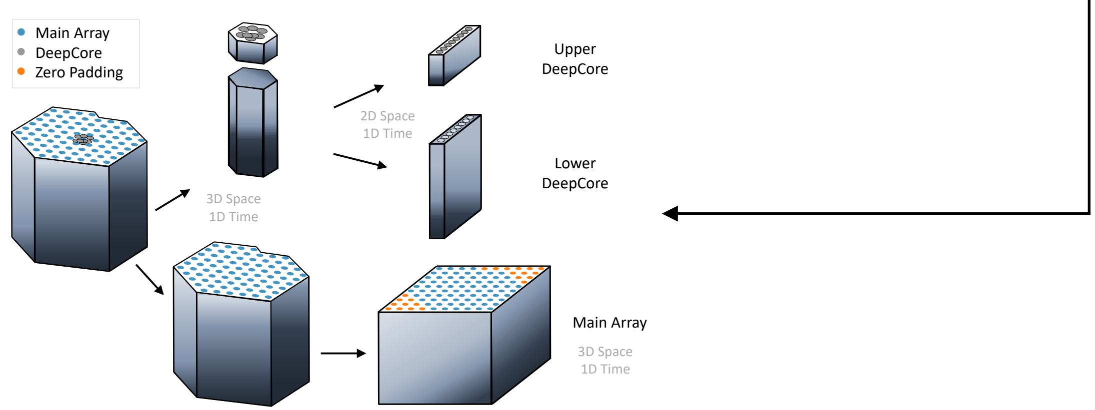
Cascades

- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, Monopod, DirectFit)
- ML (FLERCNN, DNN)
- ML+LLH approaches (CascadeGenerator)
- Likelihood-free inference (FreeDOM)

DNN

Input pulseseries features into CNN



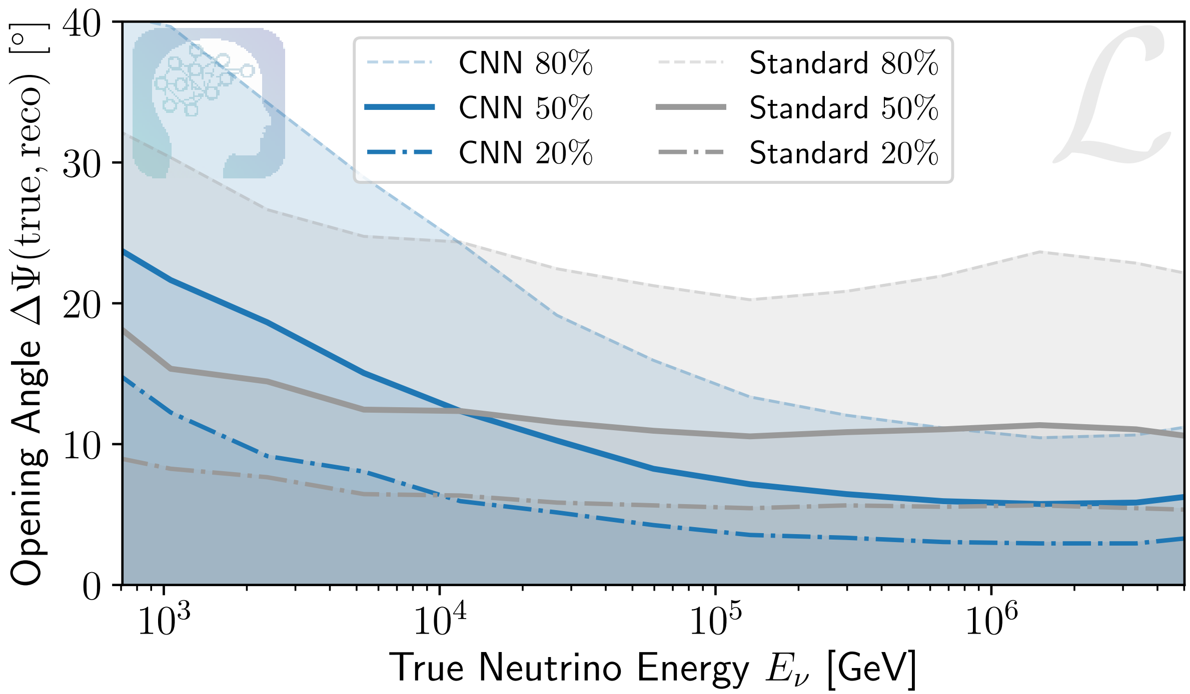


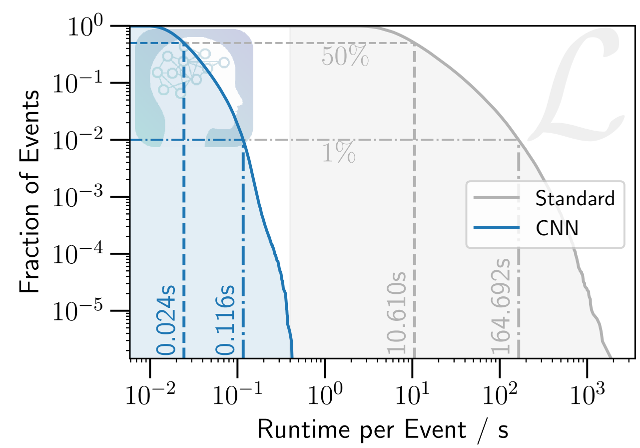
Comparison to Monopod

DNN trained on newer ice models

Monopod relies on tables built for older ice models (SpiceMie, effective SpiceLea)

Better resolution above ~10 TeV

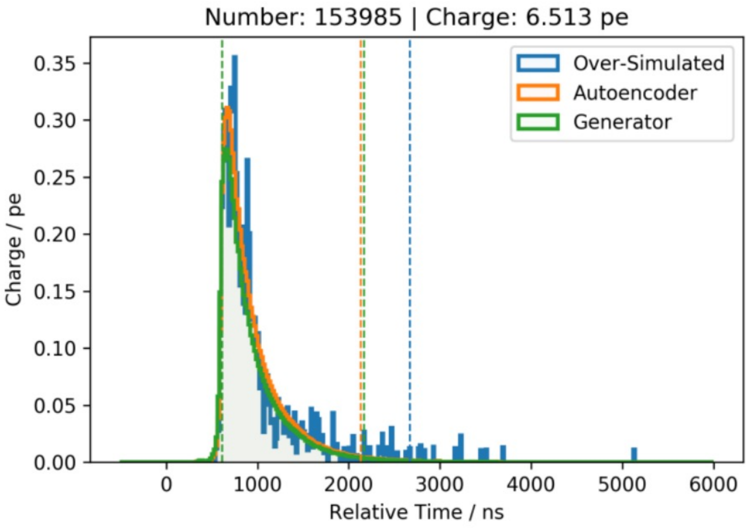


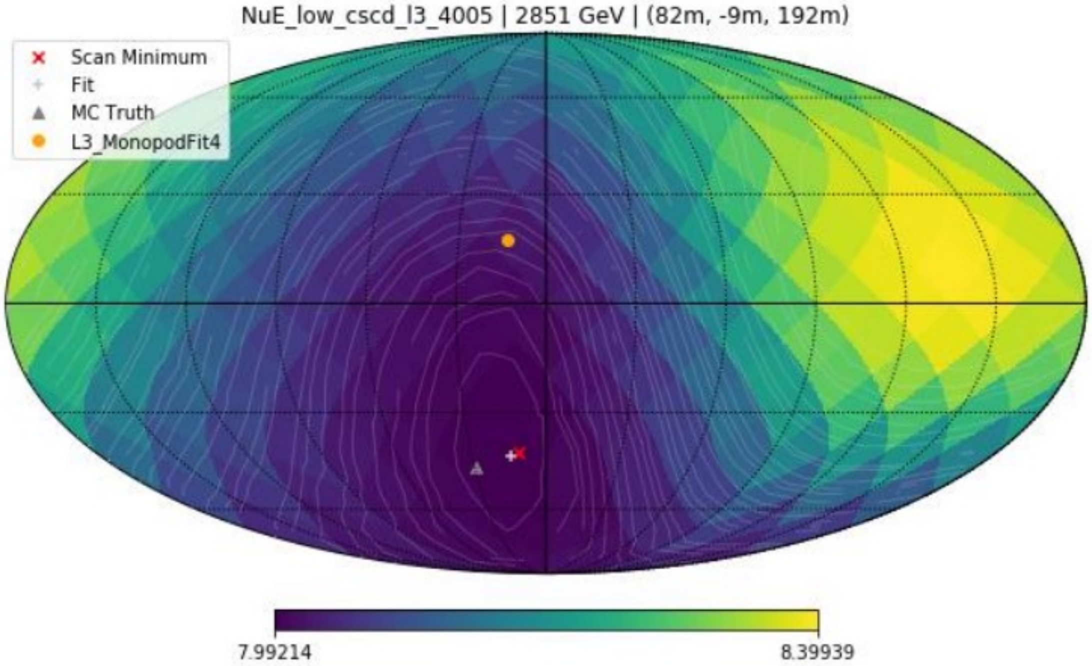


CascadeGenerator

DNN outputs most probable direction and uncertainty estimators

CascadeGenerator outputs waveforms that can be used to construct a traditional likelihood





Approaches for reconstruction

Tracks

- Use **first-hit times** for *directional* reconstruction (SANTA, SplineReco)
- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, MuEx, Millipede, DirectFit)
  - Millipede works for **high-energy tracks** by breaking it up into multiple cascades along the track due to muon stochastic energy losses
- ML+LLH approaches (CascadeGenerator)
- Likelihood-free inference (FreeDOM)
- Energy reco (TruncatedEnergy)

Cascades

- Use **full-waveform** information by fitting predicted light yields to what is actually seen (RetroReco, DirectReco, Monopod, DirectFit)
- ML (FLERCNN, DNN)
- ML+LLH approaches (CascadeGenerator)
- Likelihood-free inference (FreeDOM)

Track + Cascade

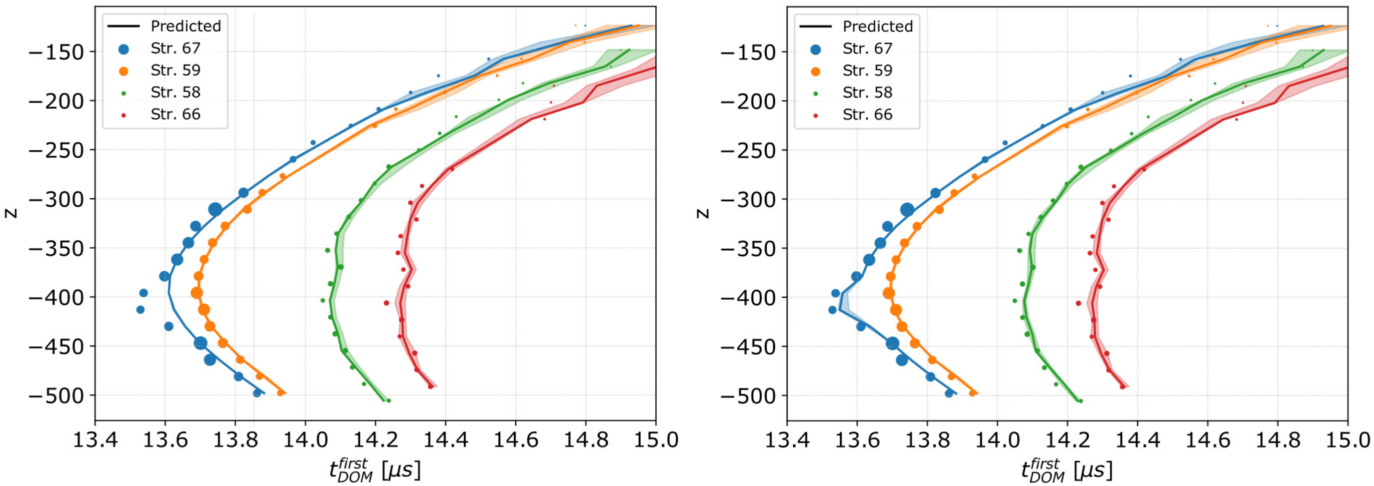
- PeV hadronic showers with early muons (Lollipop)

Combining tracks and cascades

Nature **591,** 220–224 (2021)

Hadronic showers at PeV energies may be accompanied by muons

- Outrun shower Cherenkov wavefront

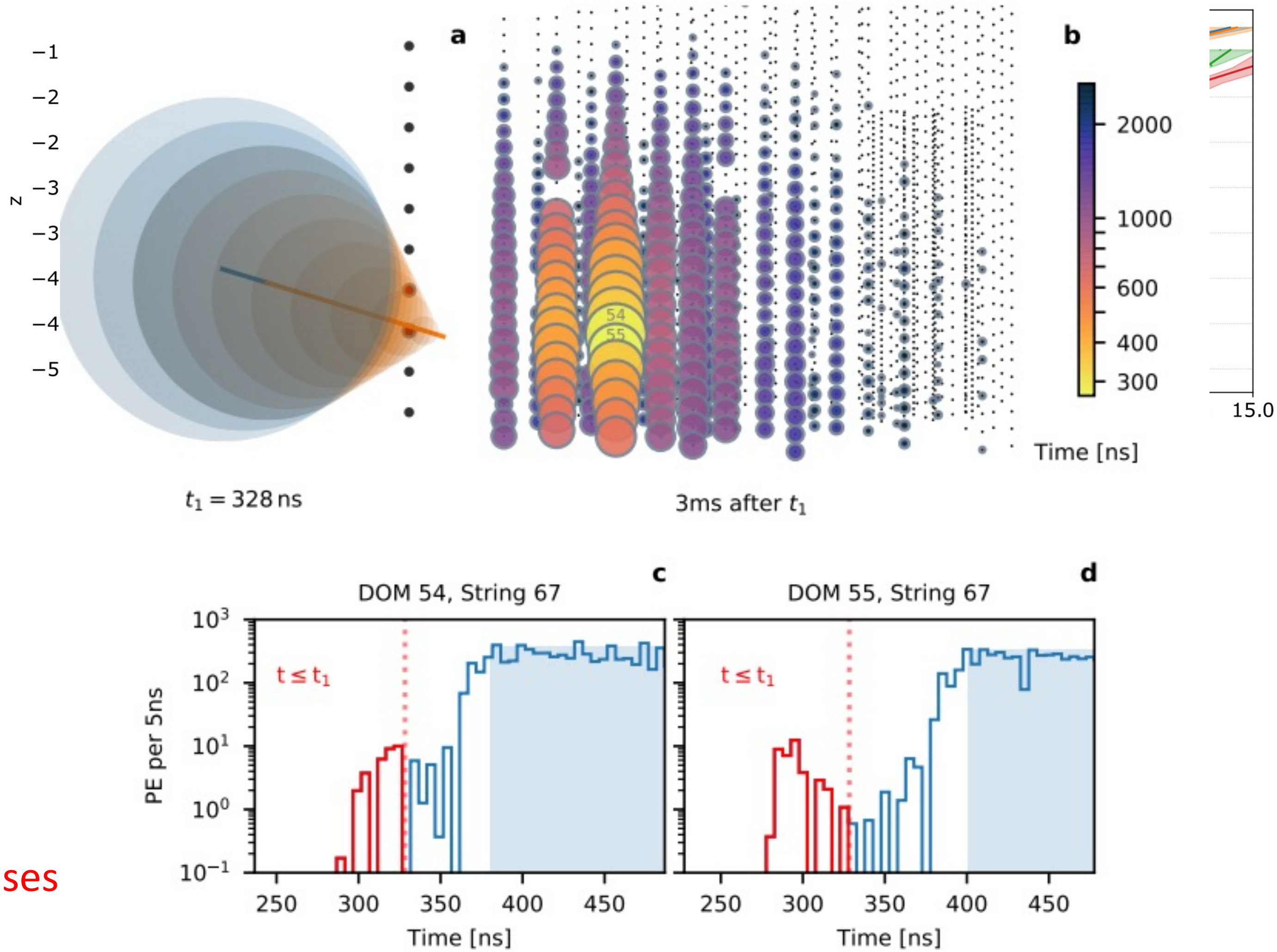


Combining tracks and cascades

Nature **591,** 220–224 (2021)

Hadronic showers at PeV energies may be accompanied by muons

- Outrun shower Cherenkov wavefront



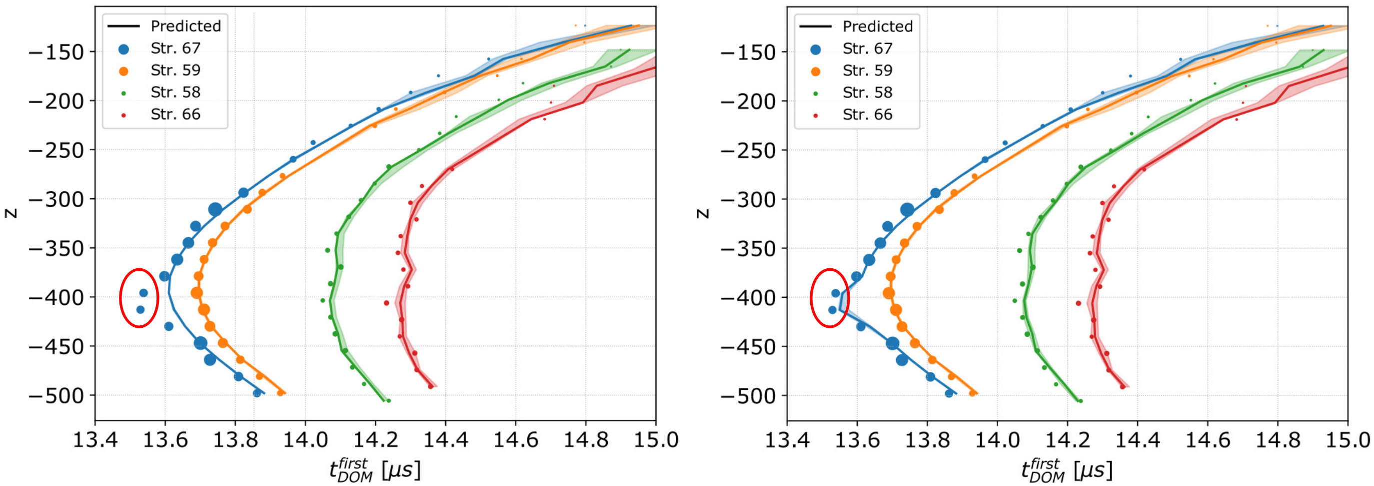
Early pulses

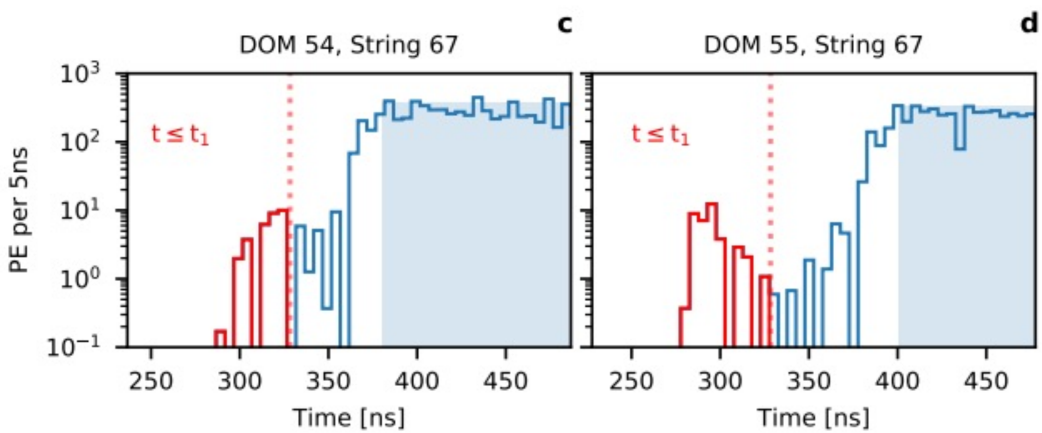
Combining tracks and cascades

Nature **591,** 220–224 (2021)

Hadronic showers at PeV energies may be accompanied by muons

- Outrun shower Cherenkov wavefront

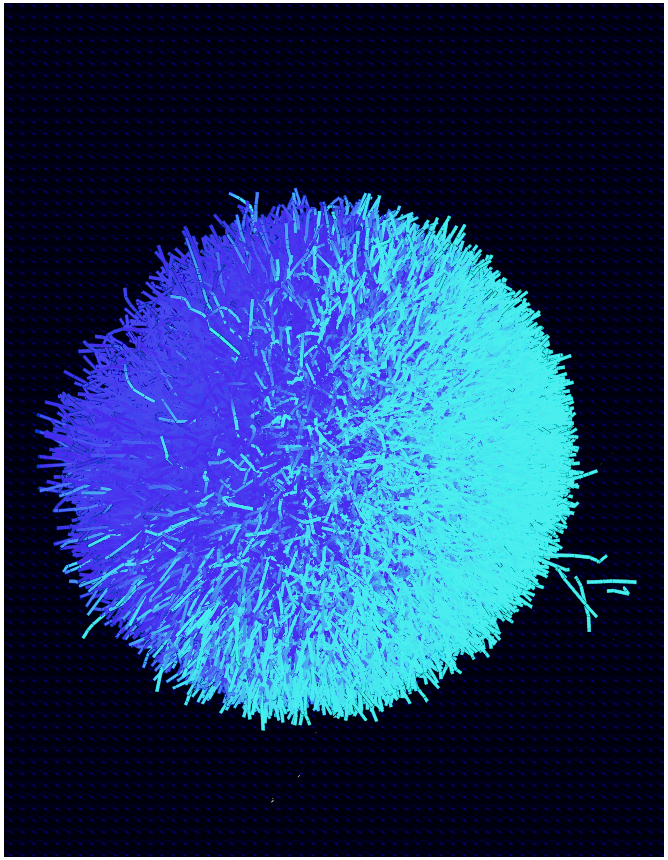




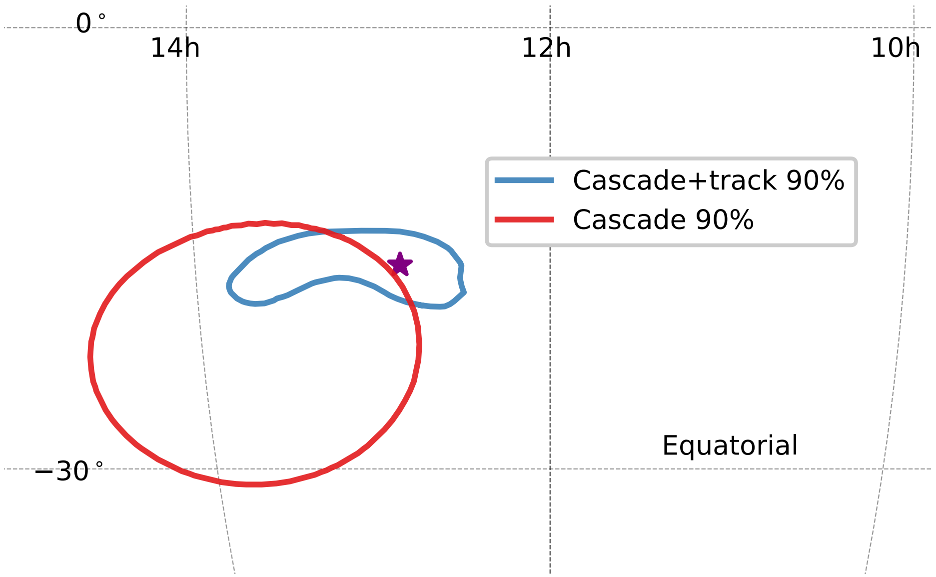
Early pulses

Improvements in directional reconstruction

Cascade reco → reco vertex/direction/energy → Track reco w. vertex prior



Improvements in directions possible!



Summary

Reconstruction in IceCube is often a challenge

Many algorithms exist, separable into high-energy/low-energy and track/shower

- Ice modeling is most important for cascades

Traditionally LLH-based approaches; recently a lot of ML/hybrid developments

Each has pros and cons ~ymmv

New approaches always welcome!

References

SANTA: https://doi.org/10.1016/j.astropartphys.2011.01.003

SplineReco: https://docs.icecube.aq/icetray/main/projects/spline-reco/index.html

RetroReco: https://github.com/icecube/retro

DirectReco: https://indico.cern.ch/event/593812/contributions/2499791/attachments/1468178/2270620/snowicki_IC_directreco_CAPtalk2017.pdf

MuEx: https://docs.icecube.aq/icetray/main/projects/mue/muex.html

TruncatedEnergy: https://docs.icecube.aq/icetray/main/projects/truncated_energy/index.html

Millipede: https://docs.icecube.aq/icetray/main/projects/millipede/index.html

DirectFit: http://icecube.wisc.edu/~dima/work/WISC/papers/2013_ICRC/dir/icrc2013-0581.pdf

FLERCNN: https://github.com/jessimic/LowEnergyNeuralNetwork

DNN: https://icecube.wisc.edu/~mhuennefeld/docs/dnn_reco/html/pages/about.html

CascadeGenerator: https://events.icecube.wisc.edu/event/115/contributions/5977/attachments/5029/5566/2019_09_18_Tokyo_cgenerator.pdf

FreeDOM: https://events.icecube.wisc.edu/event/125/contributions/7228/attachments/5679/6634/fienberg_freeDOM_plenary.pdf

Backups

Local effects

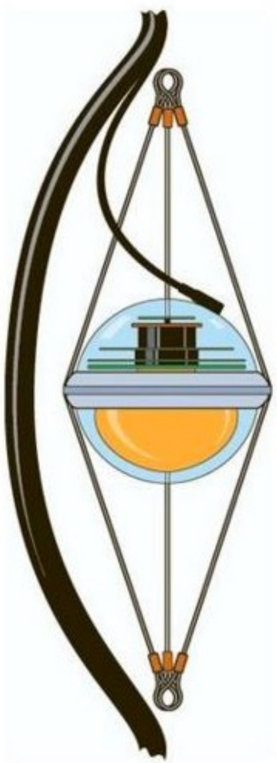
Hole-ice

- Refrozen central column with high scattering

DOM orientation

- Thick, support cable may impede direct photons if vertex is nearby
- A few DOMs may not be perfectly horizontal

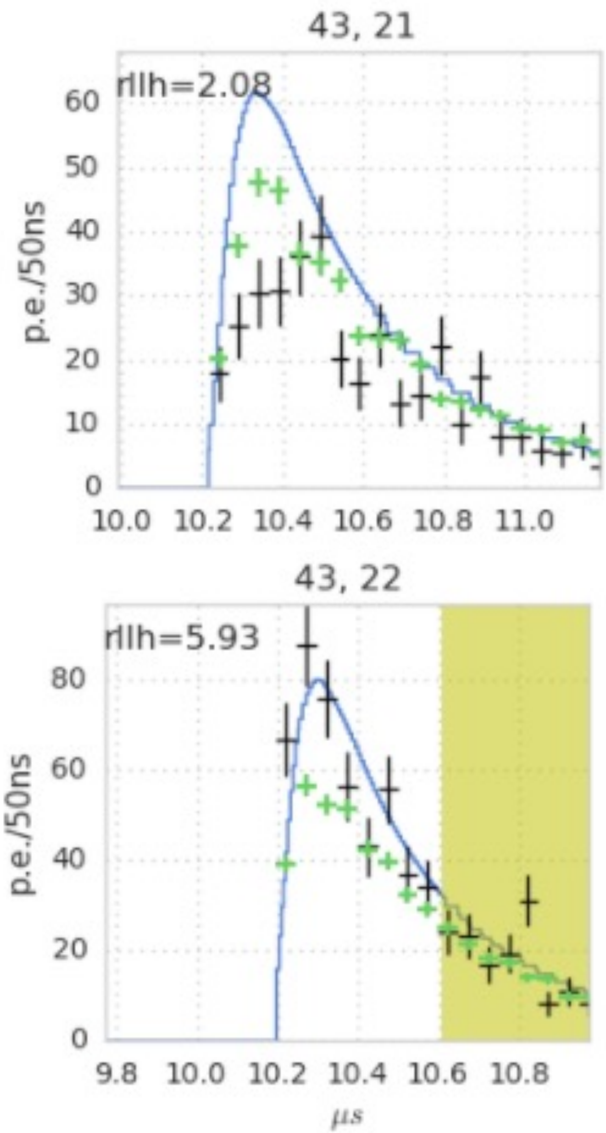
Looking up the string



Local effects: DOM orientation and cable position

Bert data
Direct photon MC
Effective photon MC

Without local effects



With local effects