Muon reconstruction and optimal event classification in AMANDA

Gary C. Hill

University of Wisconsin- Madison

for the AMANDA Collaboration

INTERNATIONAL WORKSHOP ON

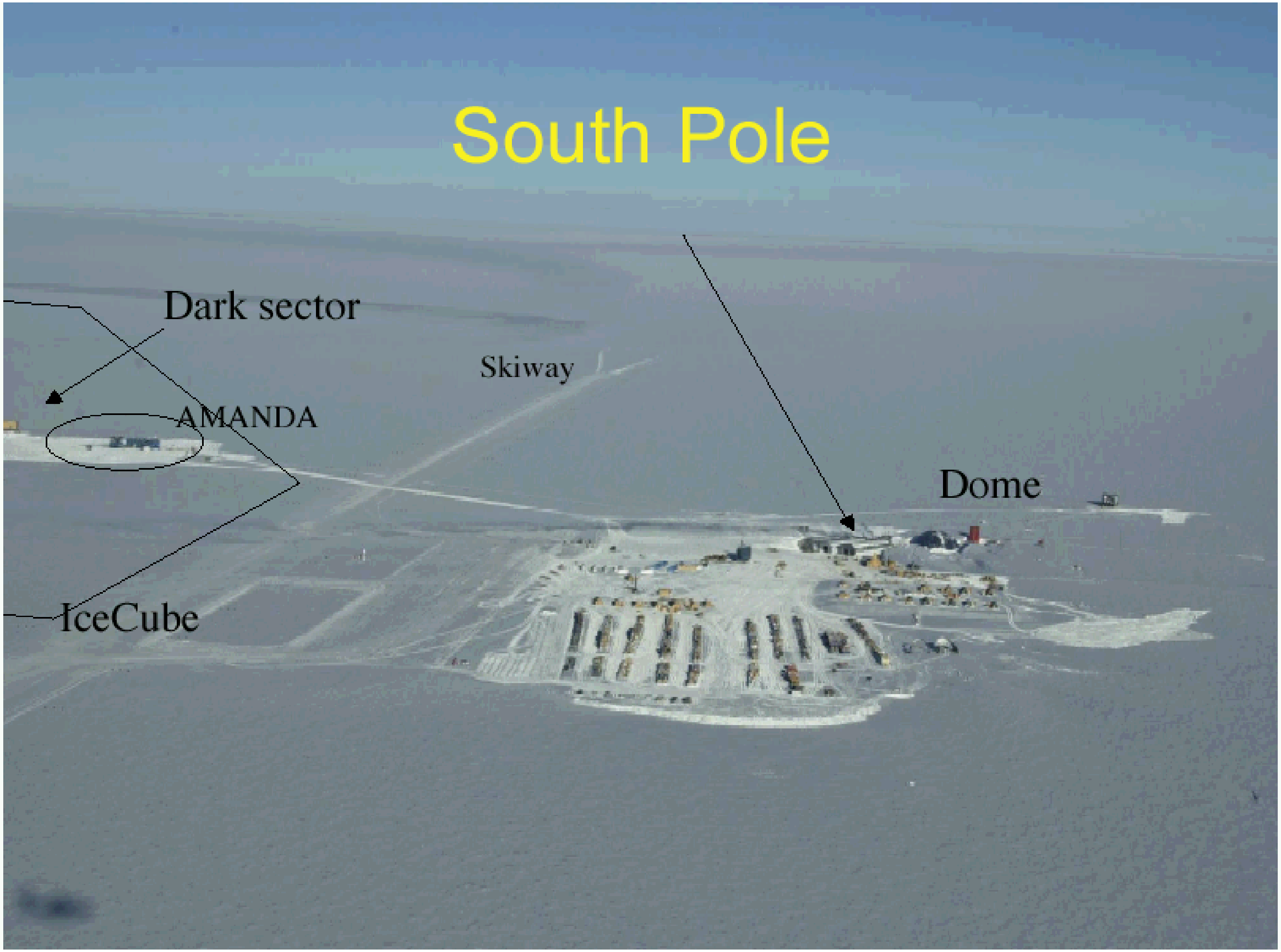
Ultra High Energy Neutrino Telescopes

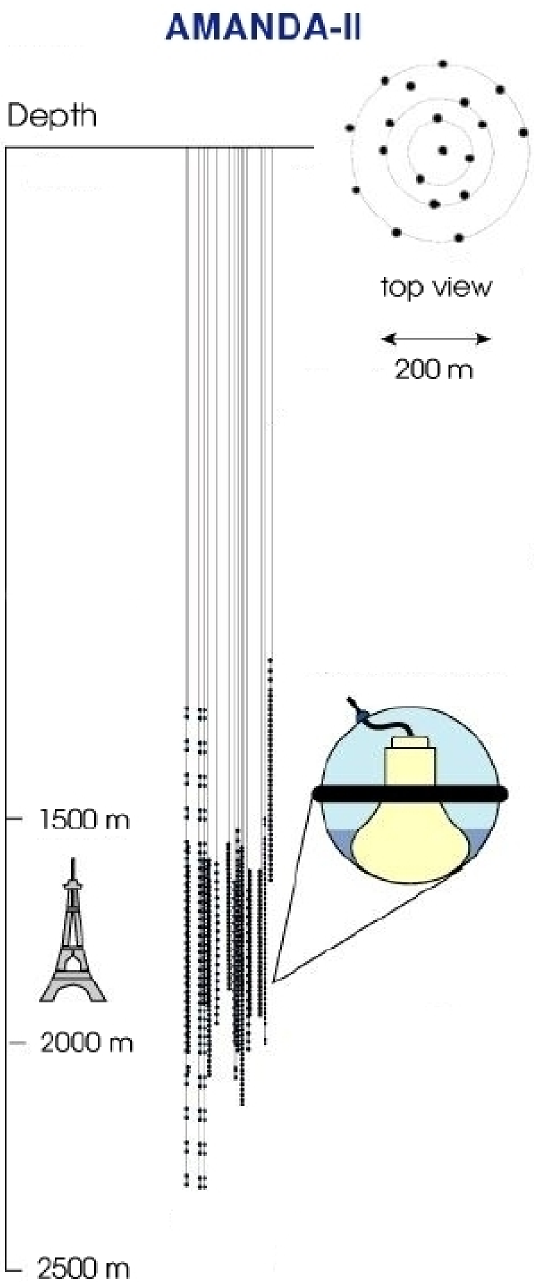
Chiba University, Chiba, Japan July 29, 2003

Outline

- Maximum likelihood reconstruction in the presence of light scattering in ice
- Optimal event classification - our job of reconstruction isn't done until we've assigned the most probable origin (background, signal) to an observed event
  - Zenith weighted "Bayesian" event reconstruction
  - Optimal classification with modern machine learning methods
- Try to give a unifying theme to the problem of reconstruction and event classification

AMANDA-II Location



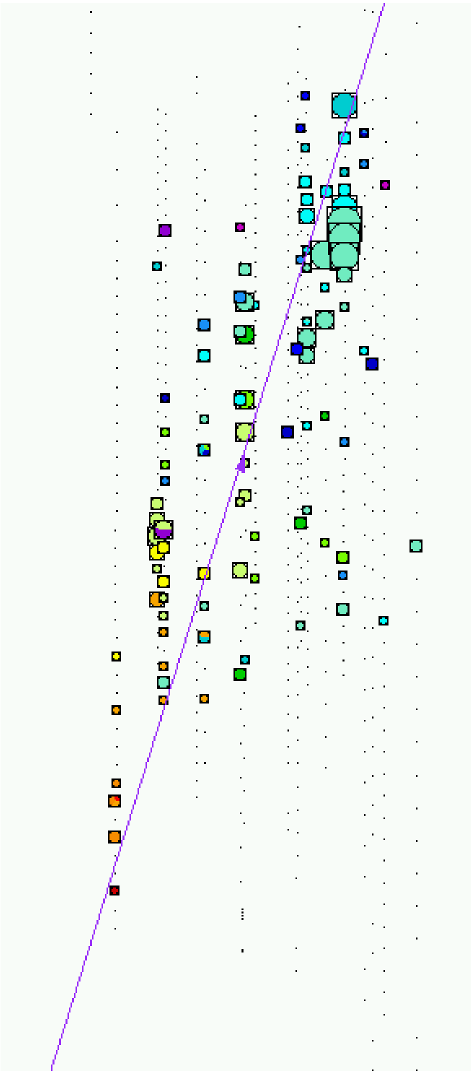


AMANDA-II Experiment

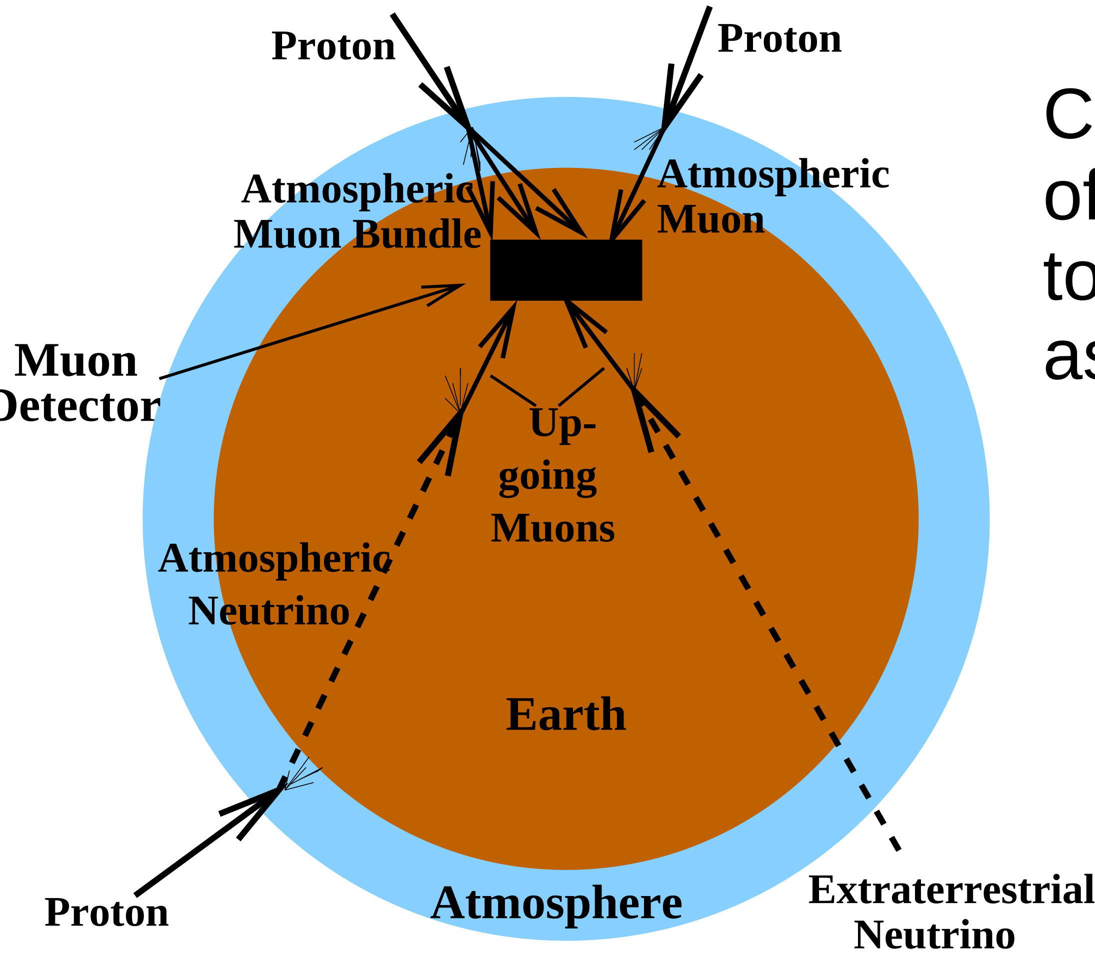
- 19 strings
- 677 Optical Modules (OM)
- 200 meters diameter
- 500 meters tall

- completed in 1999
- 1997-99 AMANDA-B10
  - 10 strings, 300 OM

Experimental ν_μ event



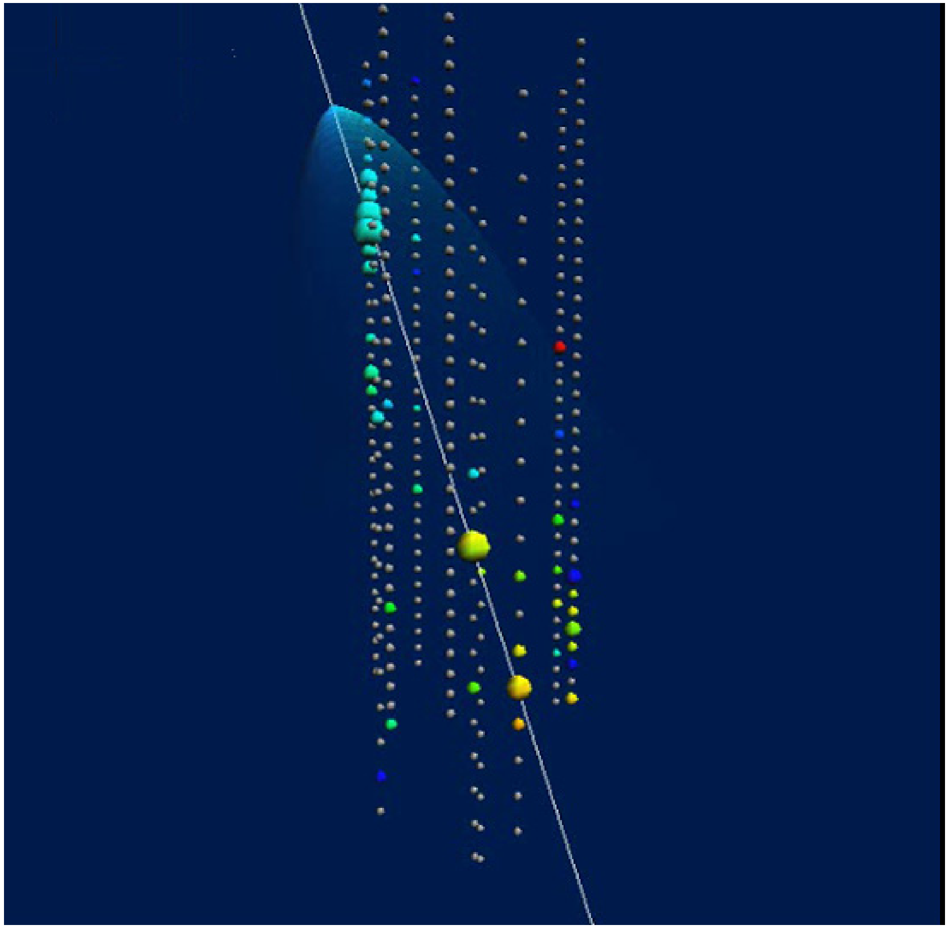
Potential (muon) event origins?



Consider three types of hypothetical origin to which we will try to assign an event

- Downgoing muons
- Upgoing atmospheric neutrinos
- Upgoing extraterrestrial neutrinos

Reconstruction principle



- Cherenkov photons are detected by PMTs
- tracks are reconstructed by maximum likelihood method of photon arrival times

Muon track reconstruction

- Cherenkov photons from the muons are recorded by the array optical modules
- each module records photon arrival times and amplitudes
- an event E is described by a vector of times and amplitudes of all the hits :

$$E \equiv \{t_1, ...t_n; \rho_1, ..., \rho_n\}$$

- Wish to fit a track hypothesis :

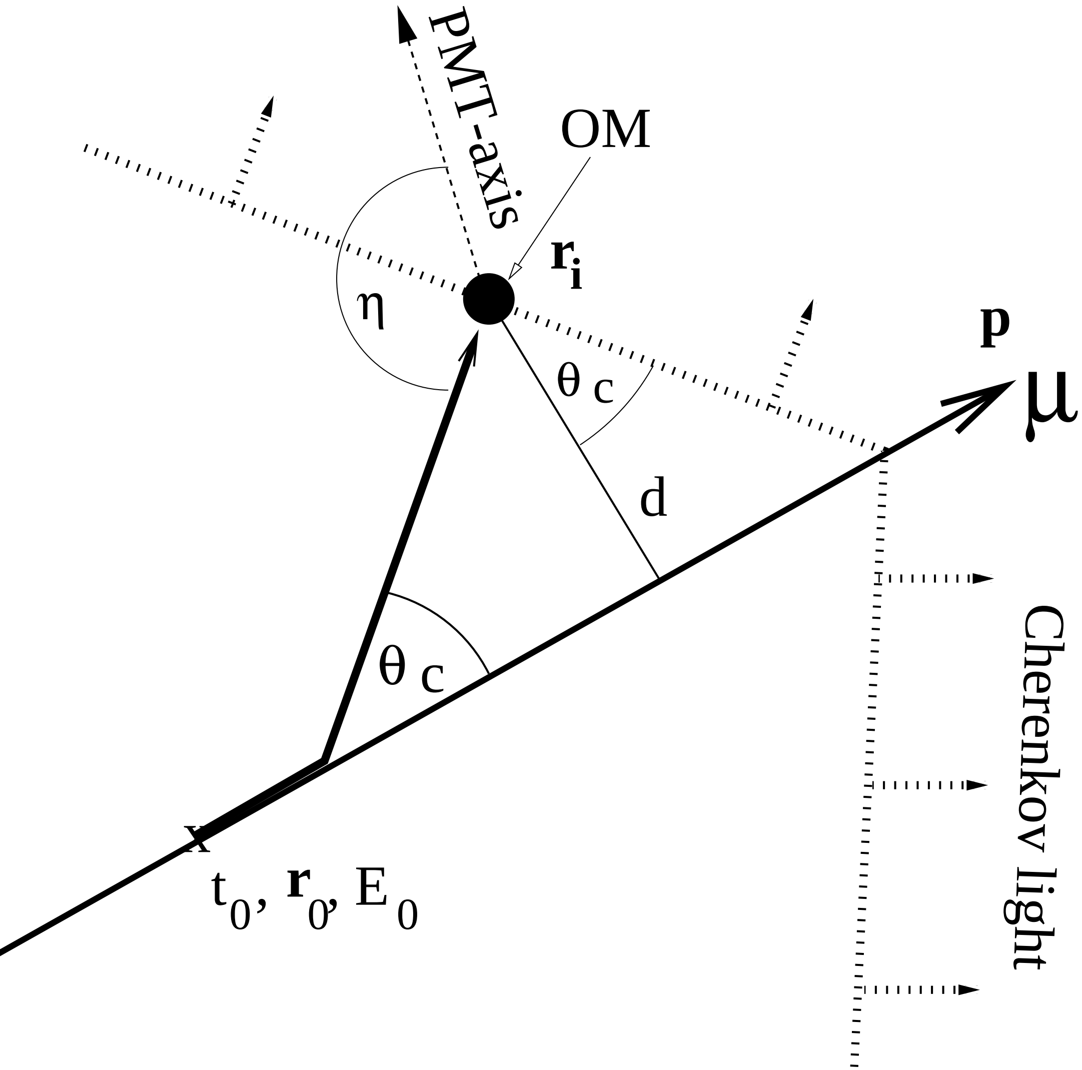
$$H \equiv \{x, y, z, \theta, \phi\}$$

Fitting a muon track hypothesis to the event information

- To connect the event $E \equiv \{t_1, ...t_n; \rho_1, ..., \rho_n\}$ and the track hypothesis $H \equiv \{x, y, z, \theta, \phi\}$ we need the likelihood function

$$\mathcal{L}(E \equiv \{t_1, ...t_n; \rho_1, ..., \rho_n\} \mid H \equiv \{x, y, z, \theta, \phi\})$$

Muon track Cherenkov cone geometry



- Given a track hypothesis we can calculate the expected photon arrival times from an unscattered Cherenkov cone

Likelihood reconstruction in the absence of scattering

- Expected photon arrival times derived from Cherenkov geometry smeared with Gaussian PMT jitter
- Straightforward form of $p(\mathrm{times} \mid \mathrm{track})$

$$\mathcal{L} = \Pi_{\mathrm{OMs}}\, p(\mathrm{time}_{\mathrm{OM}} \mid \mathrm{track})$$

- Essentially χ^2 fit
- This method insufficient in ice with scattering
- Need to use a likelihood with full photon propagation information

Detemining the PMT time residuals

- Time residual is the delay in photon arrival time after the expected "direct" Cherenkov arrival time
- Full photon propagation simulation (e.g. PTD (Albrecht Karle), Photonics (Ped Miocinović)) used to tabulate residuals as a function of possible muon tracks
- These tables can be used as the reconstruction likelihood

Analytic form - the Pandel function

- Dirk Pandel (diploma student at DESY-Zeuthen in mid-90's) solved the propagation equations of light in the presence of absorption and scattering and found an analytic form for the time residuals

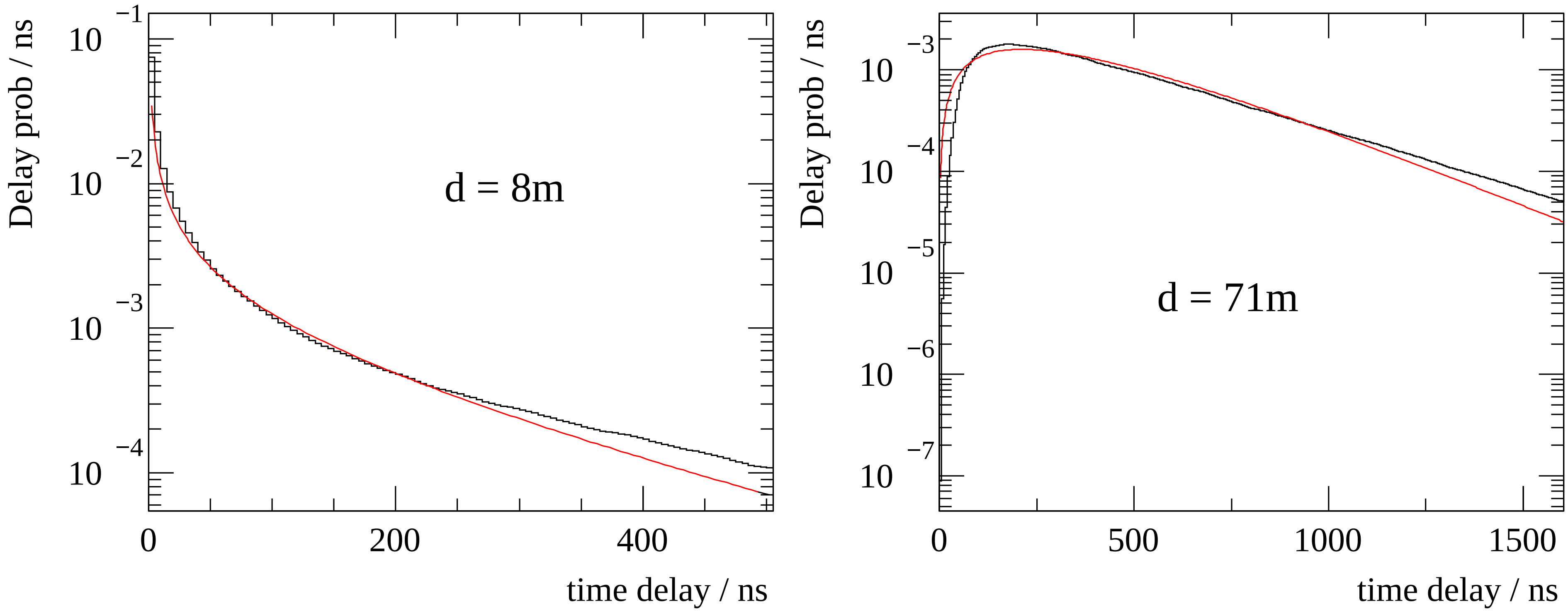
$$p(t_{\mathrm{res}}) \equiv \frac{1}{N(d)} \frac{\tau^{-(d/\lambda)} \cdot t_{\mathrm{res}}^{(d/\lambda - 1)}}{\Gamma(d/\lambda)} \cdot e^{-\left(t_{\mathrm{res}} \cdot \left(\frac{1}{\tau} + \frac{c_{\mathrm{medium}}}{\lambda_a}\right) + \frac{d}{\lambda_a}\right)}$$

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

Fitting the Pandel function free parameters

- The free parameters are fitted to make the Pandel form of the residual distributions match the full photon simulation
- Gives an analytic form that can be used in the reconstruction algorithm
- Need to add PMT jitter - old method was a simple patching of a Gaussian with the Pandel
- Recently an analytic form of the convolution of the Pandel with a Gaussian was found (George Japaridze)

Time residuals - full simulation and Pandel fit



Pandel fit (red) to photon tables (black)

Reconstructing an event!

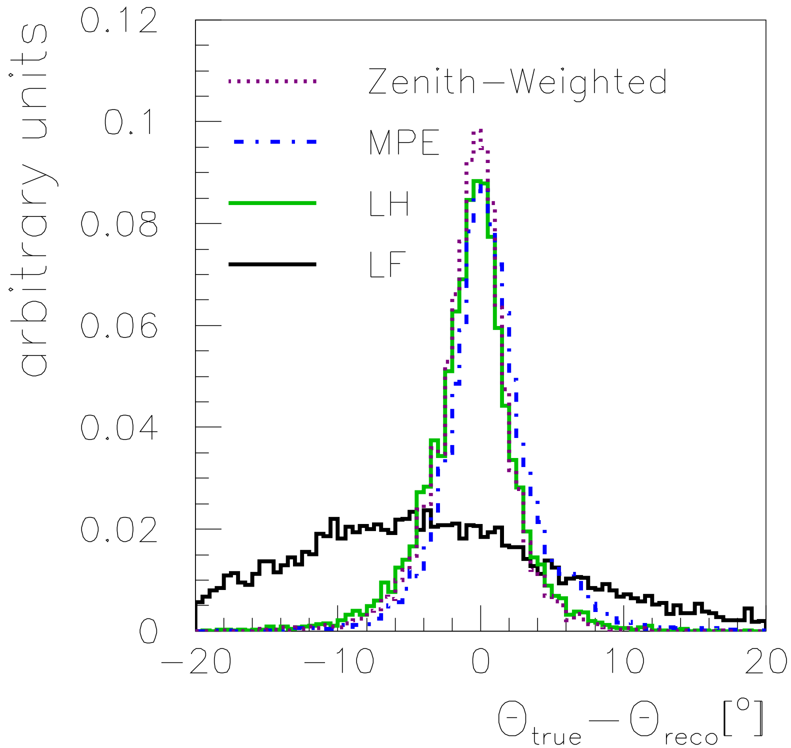
- Our likelihood function

$$\mathcal{L}(E \equiv \{t_1, ...t_n; \rho_1, ..., \rho_n\} \mid H \equiv \{x, y, z, \theta, \phi\})$$

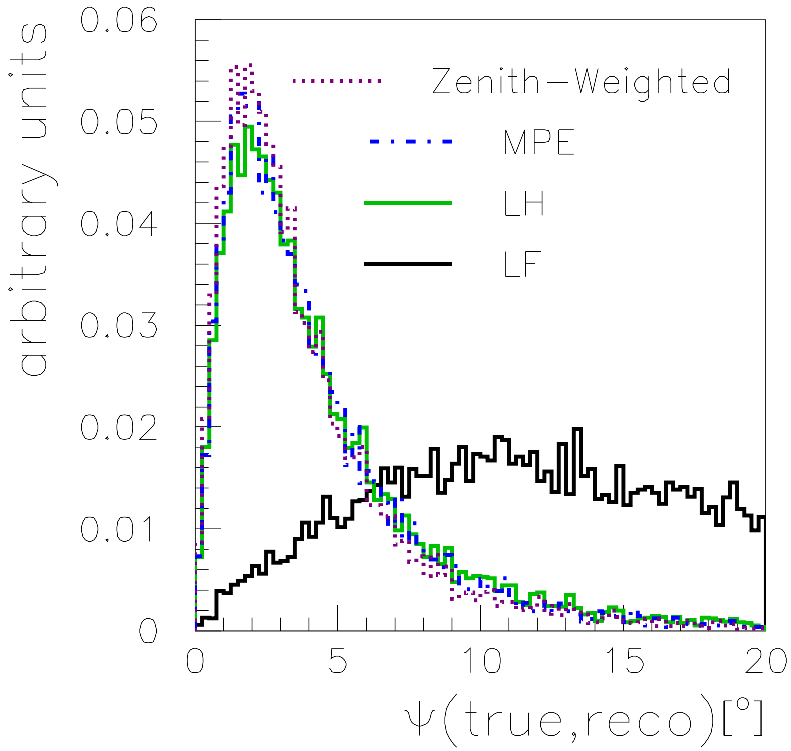
is given by the track geometry and the time residual function (tabulated photon simulation or Pandel function)

- use a minimisation algorithm (Nelder-Mead simplex, Powell's gradient descent, Minuit) to fit the track parameters

Reconstruction performance



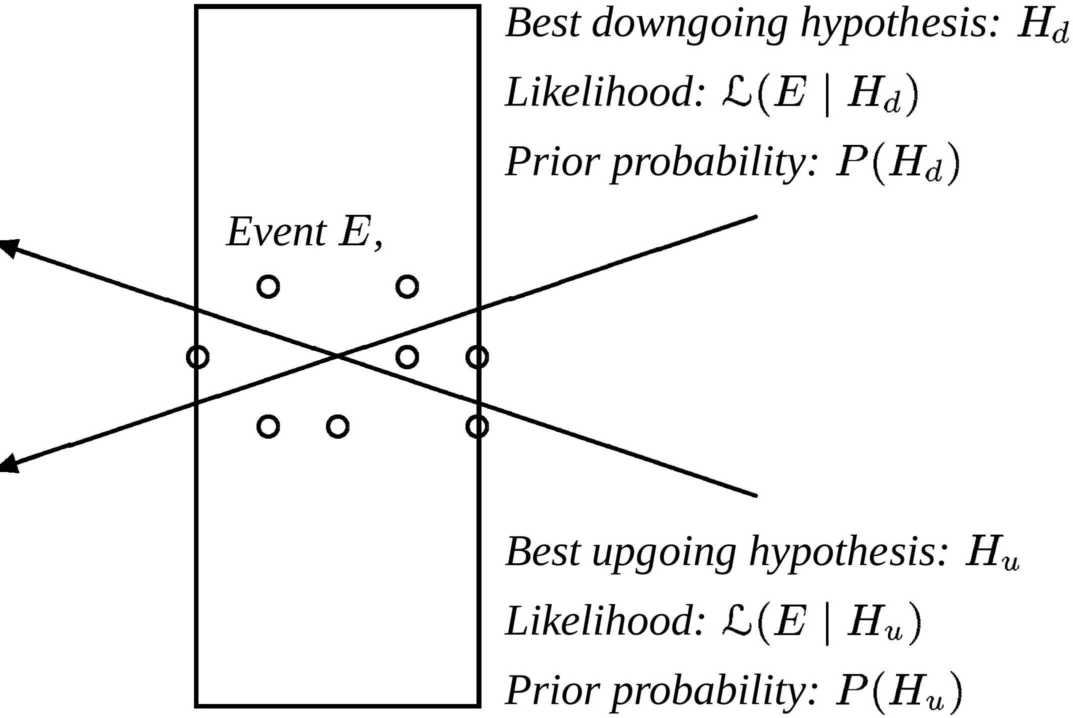
zenith angle



space angle

- Pandel function and photon tables yield similar results

What is the most probable origin of an observed event?



What if $\mathcal{L}(E \mid H_u)$ is only slightly better than $\mathcal{L}(E \mid H_d)$?

- Should we still choose H_u over H_d?
- We know that $P(H_d) > P(H_u)$ i.e. more downgoing muons passing through the detector
- Also strong zenith dependence of $P(H_d)$
- how is this accounted for?

Joint conditional probability distribution

Most probable downgoing muon hypothesis is the one that maximises joint probability distribution

$$P(E \mid H_d)\ P(H_d) \quad \equiv \quad \mathcal{L}(E \mid \mu_d)\ \Phi(\mu_d)$$

where $P(H_d) \equiv \Phi(\mu_d)$, the flux of downgoing muons in the vicinity of the detector.
Most probable upgoing hypothesis : maximise

$$P(E \mid H_u)\ P(H_u) \quad \equiv \quad \mathcal{L}(E \mid \mu_u)\ \Phi(\mu_u)$$

where $P(H_u) \equiv \Phi(\mu_u)$, the flux of upgoing muons in the vicinity of the detector (taken as uniform).

Zenith weighted reconstruction in practice

- Treat the downgoing muon prior as a simple function of the zenith angle (polynomial fit to simulated muon flux at the detector)
- For each event, find the maximised downgoing and upgoing likelihoods, then take the ratio.
- Use this ratio as a cut parameter, optimised on simulated downgoing and upgoing events
- Rejection of mis-reconstructed atmospheric muons improved by a couple of orders of magnitude over conventional “all hypotheses are equal” method
- Cuts are simplified (in principle, this is the only cut we need)

Bayesian statistics interpretation

The probabilities of observing an event E due to up and downward muons are found by integration over the likelihood and priors

$$P_d(E) = \int_{H_d} \mathcal{L}(E \mid H_d)\, P(H_d)\, \mathrm{d}H_d$$

$$P_u(E) = \int_{H_u} \mathcal{L}(E \mid H_u)\, P(H_u)\, \mathrm{d}H_u$$

The ratio $P_d(E)/P_u(E)$ is known as the Bayes' discriminant and is the statistically most powerful separator of classes of hypotheses

Are we evaluating the discriminant?

We have approximated the Bayes' discriminant ratio of integrals by the ratio of the maximum values of the integrands :

$$\frac{\int_{H_d} \mathcal{L}(E \mid H_d)\, P(H_d)\, \mathrm{d}H_d}{\int_{H_u} \mathcal{L}(E \mid H_u)\, P(H_u)\, \mathrm{d}H_u} \simeq \frac{\mathcal{L}(E \mid \hat{H}_d)\, P(\hat{H}_d)}{\mathcal{L}(E \mid \hat{H}_u)\, P(\hat{H}_u)}$$

Don't most physicists reject Bayesian inference?

- Absolutely yes when used incorrectly!
- Classic example is in upper limit calculations where uniform priors are used to represent subjective "degree-of-belief" about an unknown physical quantity (e.g. the rate of a Poisson process λ, or the mass of a particle m)
- After measuring x, an inference on m is made from $P(m \mid x) \propto P(x \mid m)P(m)$
- Usually take $P(m)$ to be uniform in some interval
- However $P(m)$ uniform does not yield same inference as taking $P(m^2)$ uniform and both choices of "metric" (m or m^2) are equally valid

What about our "Bayesian" reconstruction?

- Acid test - Advanced Statistics in Particle Physics Workshop, Durham, 2002, Ty DeYoung with an audience of the staunchest Bayesians and anti-Bayesians
- Bayesians naturally said the technique was fine....

Bob Cousins for the frequentists....

- Bayes theorem applies to all types of probability - both subjective degree of belief (e.g. "I think the mass of the Higgs is uniform in the interval 80-200 GeV") and to classical relative frequency probabilities ("the distribution of cosmic rays arriving at earth is uniform and follows a power law energy spectrum")
- Our muon flux “prior” is a relative frequency probability - it’s very easy to define $P(\mu_d)$ - the muon flux is well measured, theoretically calculated and understood - not subjective at all
- More explictly : the procedure is “Bayesian” only in that Bayes’ theorem was used – definitely not “subjective” Bayesian!

NEVOD experiment developed this technique independently

- When presenting this work in 2001 in Hamburg, during discussion time A.A. Petruhkin from the NEVOD experiment explained how they did exactly the same thing...
- ... and where able to separate an atmospheric neutrino candidate from a 10^{10} to one background of atmospheric muons in a tiny ($6 \times 6 \times 7.5\mathrm{m}^3$) surface detector!

Modern machine learning classification

- Machine learning - feed a routine a bunch of labelled signal and background, build a model of the Bayes' posterior for future classification of new data
- Neural networks are an example
- Modern methods - Support Vector Machines, Penalised Likelihood methods (Reproducing Kernel Hilbert Space methods)

Penalised likelihood method

- Build a model of the Bayes' discriminant using weighted sums of basis functions and regularisation methods to control the smoothness of the solution
- Currently building a model of atmospheric and isotropic E^{-2} neutrinos for our diffuse limit analysis (work in collaboration with UW Statistics)

Conclusions

- Reconstruction of muon tracks in a scattering medium has been successful
- Methods of optimally classifying events as signal and background have been implemented (zenith weighted reconstruction) and are under development (Penalised Likelihood Estimate model building)
- These provide a unifying framework for the reconstruction and classification problem