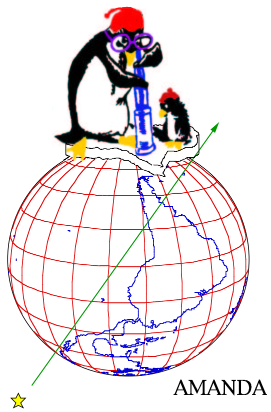
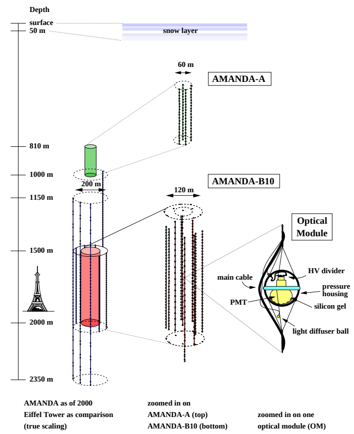
Antarctic Muon And Neutrino Detector Array (AMANDA) on Millennium

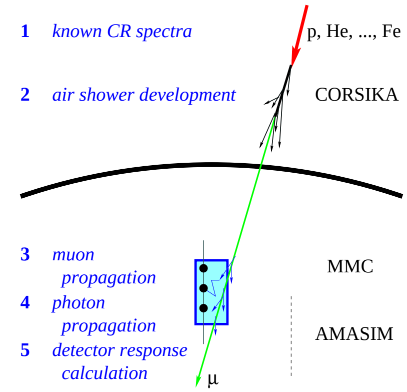


Dmitry Chirkin
chirkin@physics.berkeley.edu

University of California at Berkeley, USA



AMANDA is a detector located in the ice 1 km away from the geographical South Pole. Its main goal is to look for the extra-galactic neutrinos. The majority of the recorded events, however, come from muons and neutrinos created in the Earth's atmosphere. This *background* must be carefully studied in order to find ways to remove it (annoyance) and to employ it for detector calibration (useful).



To simulate background one needs to run an air shower generator (CORSIKA), muon propagator (MMC) and then propagate photons and evaluate detector response (AMASIM).

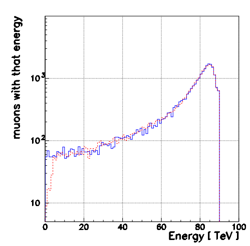
Calculation times and output sizes after each stage for a typical run with 10^8 primaries are summarized below:

step	CPU (.5 GHz) time (sec)	output size (Mb)
CORSIKA	$2.4 \cdot 10^6$	80
MMC	$5 \cdot 10^3$	1200
AMASIM	$30 \cdot 10^3$	150

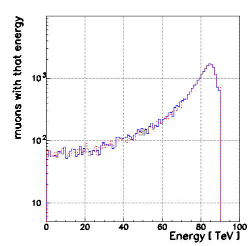
Since generation at the CORSIKA stage takes such a long time, each CORSIKA file is put together from 100 smaller files (each containing 10^6 primaries). On the millennium cluster (with ~ 200 nodes available) several 10^8-files can be produced per day. Each such file corresponds to about 2 minutes of detector lifetime. To get any kind of meaningful statistics one needs at least a couple of hours, for some types of analysis days or even months of simulated lifetime. To save computational time each CORSIKA file can be oversampled, e.g. used with different random generator seed several times (usually 10-100) in the rest of the Monte Carlo chain. With this in mind **it takes about one full day of millennium computation for each hour of simulated data**.

MMC optimizations

Millennium cluster played significant role in developing and optimizing Muon Monte Carlo (MMC). The program was developed with code clarity and uncompromising precision as main goals. Execution time was only a secondary issue and originally integral parameterizations were not implemented as the program was intended to be run on large computer clusters. Now there are many levels of parameterizations and the fully optimized (parametrized) program is as fast or even faster than other muon propagation projects. Below is the example of how comparing parametrized and non-parametrized (exact) versions of the program helped catch some flaws in parameterizations schemes. **It took a few minutes to propagate 10^5 muons on a single node parametrized and one day on the millennium cluster for the exact version**.

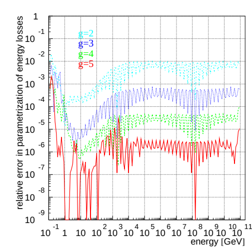


version 0.08.4

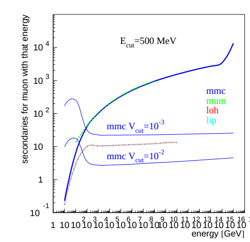


version 0.08.5

MMC optimizations



Parametrization errors

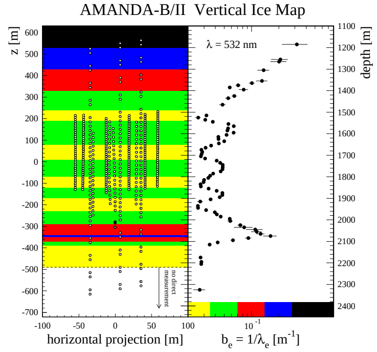


Number or secondaries

MMC employs a very customizable parametrization scheme. It uses both polynomial and rational function interpolations spanned over varying number of points (5 by default). The algorithm is very stable and allows calculations to be performed in an energy range even bigger than physically interesting without losing precision (over 15 orders of magnitude in energy). Other muon propagation codes have limited energy resolution and smaller precision in only part of MMC's region of applicability.

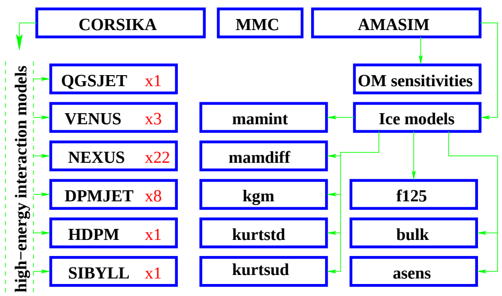
Simulating Light in Ice

AMANDA-B/II Vertical Ice Map



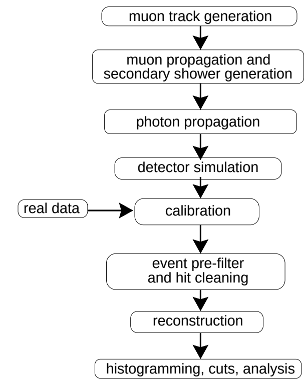
- Light emitted by muons in ice can travel over 100 m
- Due to inhomogeneous optical properties of ice, all emission/detection pairs have to be considered by light tracking simulation
- Depending on requested accuracy, CPU usage can vary from 3000 - 9000 CPU-seconds (0.7 GHz) per 10^6 photons, resulting in minimum of $\sim$ **625 CPU-days** for production quality tables.

Different Monte Carlo settings



Many different settings for the AMANDA Monte Carlo must be used in order to estimate systematic errors. Because of the lack of the accelerator data in the region of energies and scattering angles that occur in air showers of interest, different high-energy interaction models have to extrapolate from available data and may lead to different results. Some of these models are up to **22 times slower** than the used default (QGSJET). Since it is very difficult to determine the exact ice conditions at the depth of the detector, different ice models are used. Eight shown are the most commonly used.

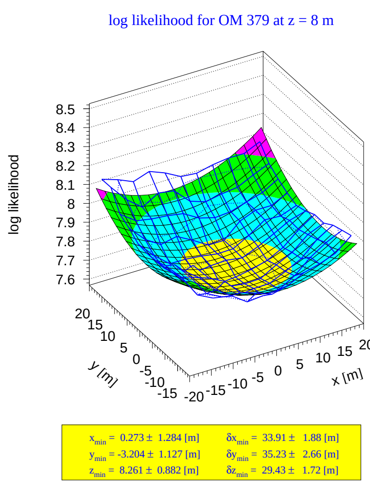
Complete data analysis



The Monte Carlo production discussed above is only part of the analysis. Calibration, cleaning and reconstruction take about as much time as detector simulation. These tasks must also be performed on the real AMANDA data. For one year (2001) processing done in Madison **1.7 TBytes** were required to store raw data and another **500 GBytes** for filtered data. The processing time was 2250 CPU-days = **6.2 CPU-years**.

Other projects: Ice Flow with AMANDA

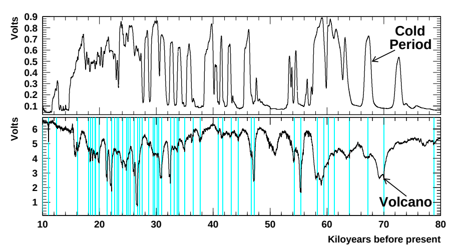
D. Chirkin, Price Group



Downgoing muon data can be used to reconstruct positions of the lowest optical modules to observe possible shear ice flow over the years. This will require full reconstruction of at least one day per season (10-20 seasons) for the lowest 24 modules.

Paleoclimatology with optical borehole devices

R. Bay, Price Group



Substantial computational resources needed for:

- Monte Carlo simulations of instrument performance (~ 100 CPU days)
- Data parsing, processing, analysis
- Monte Carlo calculations of correlations

Conclusions

1. All of the discussed projects require a lot of long-term disk space (some up to several TBytes) and significant amounts of local short-term disk cache (several GBytes per node).
2. Most projects are Monte Carlo simulations or data processing, i.e. they can be broken into many parts: Monte Carlo can be started on different nodes with different random number generator seed, and data can be processed one file at a time (several hours of the detector lifetime worth).
3. Large long-term storage capacity would allow to produce months of background Monte Carlo and to process future AMANDA data on millennium.