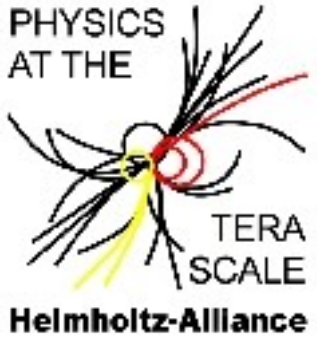
BAT

A Bayesian Analysis Toolkit



A. Caldwell[1], M. Corradi[2], G. D'Agostini[3],
A. Knue[4], D. Kollar[1], **K. Kröninger[4]**, A. Quadt[4]

[1]*Max-Planck Institut für Physik, München*
[2]*INFN, Bologna*
[3]*University of Rome La Sapienza*
[4]***II. Physikalisches Institut, Universität Göttingen***

Motivation

- **Aims of data analyses:**
  - Compare data and models
  - Judge validity of models
  - Parameter estimate

$$p(\lambda|\vec{x}) = \frac{p(\vec{x}|\lambda)\, p_0(\lambda)}{\int p(\vec{x}|\lambda)\, p_0(\lambda)\, d\lambda}$$

- **BAT: Software package for phrasing models and solving numerical problems related to Bayesian analyses**
  - C++ based framework (flexible, OO)
  - Sets of algorithms for num. Integration, optimization, etc.
  - Interfaces to ROOT, Minuit, CUBA + user defined

Flow diagram / Implementation

- **Definition of models (classes)**
  - Parameters
  - Conditional probabilities
  - Prior probabilities

$$p(\lambda|\vec{x}) = \frac{p(\vec{x}|\lambda)\, p_0(\lambda)}{\int p(\vec{x}|\lambda)\, p_0(\lambda)\, d\lambda}$$

- **Data set definition**
  - Read in from file
  - .txt, .root, histogram, user defined

- **Analysis**
  - (Normalize)
  - Optimize
  - **Marginalize**
  - **Validity check ("goodness-of-fit")**
  - **Error propagation (error band)**
  - Write output file
  - Graphical output

Tools and algorithms

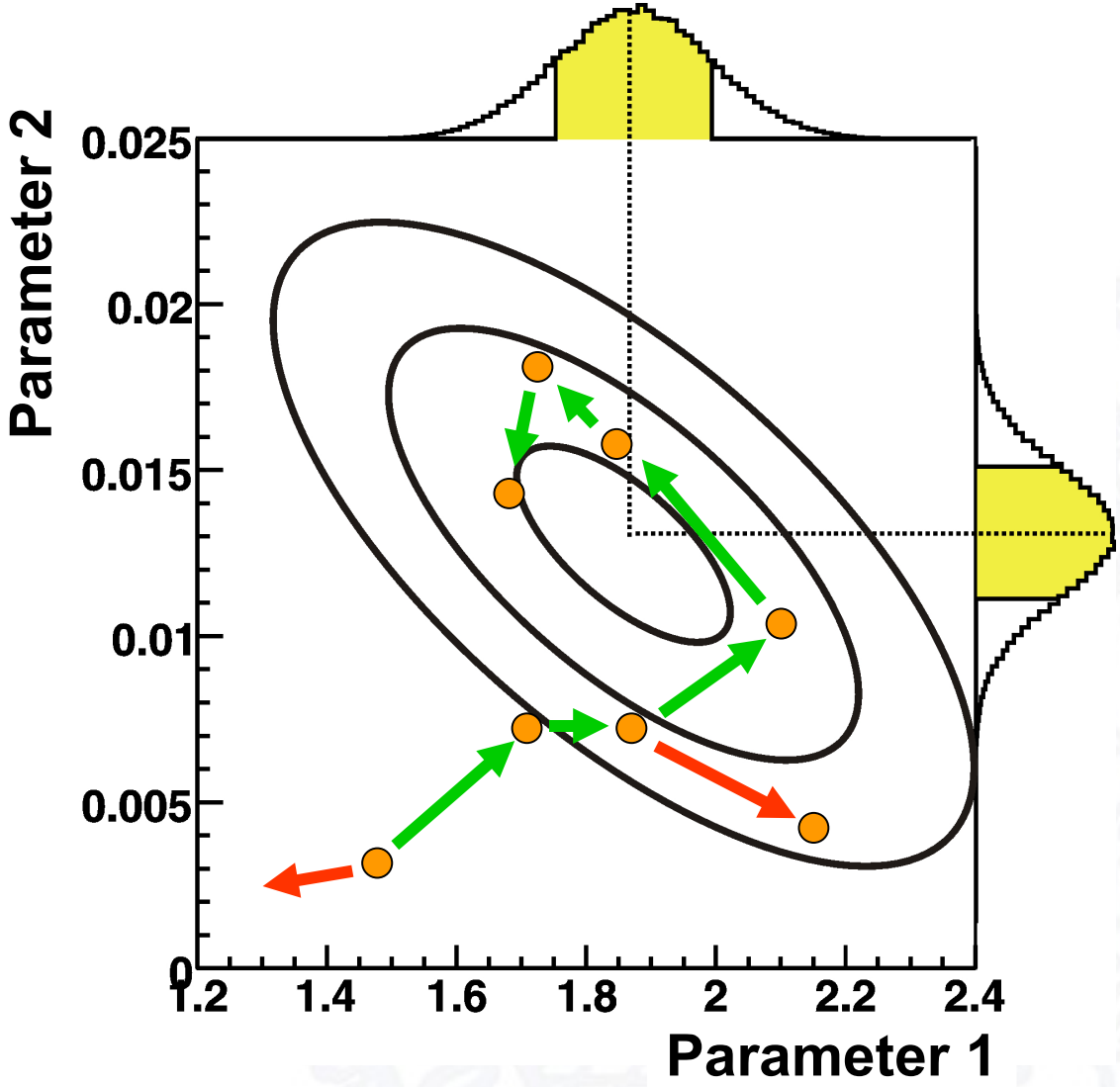
- **Integration**
  - Monte Carlo (sampled mean)
  - Importance sampling
  - Interface to CUBA (VEGAS, ...)
- **Marginalization**
  - MCMC (Metropolis)
- **Validation**
  - Ensemble tests and p-value
- **Optimization**
  - Monte Carlo (hit&miss)
  - Metropolis (quasi-simulated annealing)
  - Interface to Minuit
- **Error propagation**
  - Calculate any function of the parameters during a run

Large variety of numerical tools

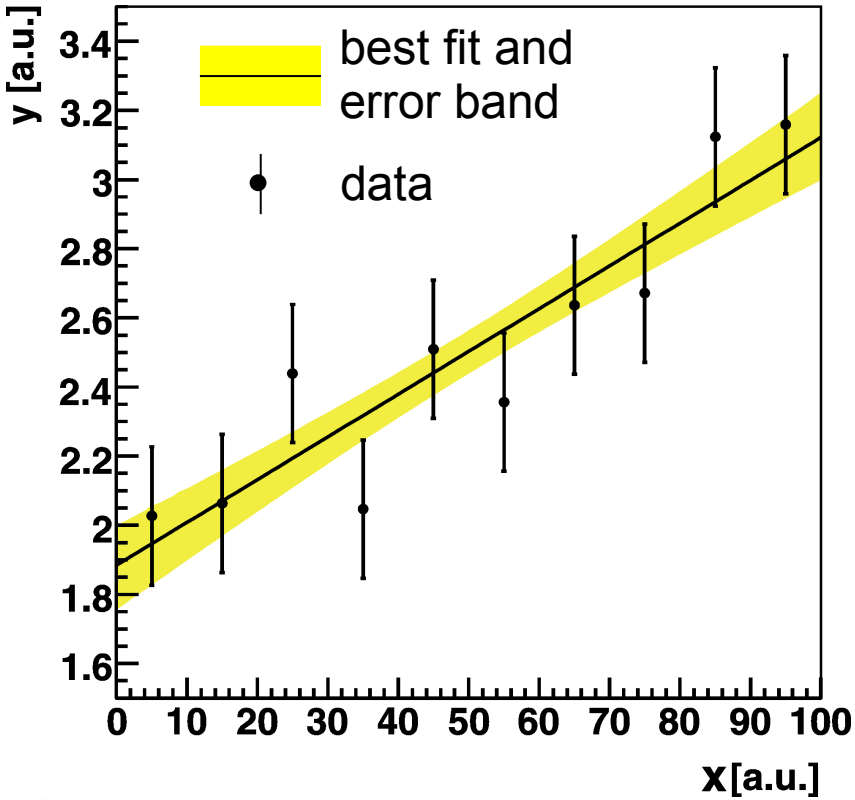
Key element: Markov Chain Monte Carlo (MCMC)

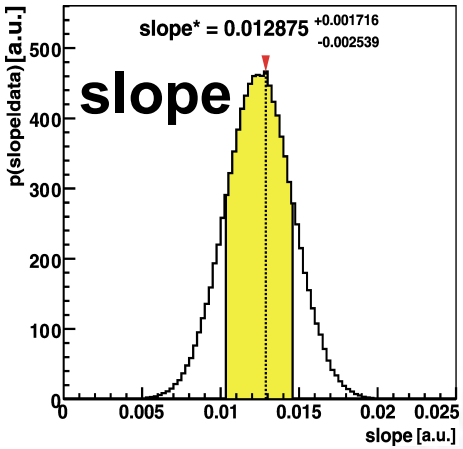
Markov Chain Monte Carlo

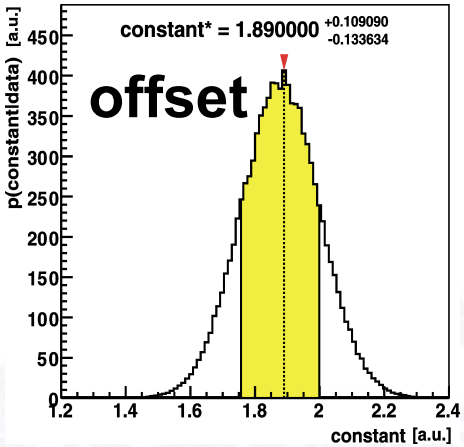
- Sample parameter space by migrating to regions of larger probability
- Converge towards underlying distribution
- Marginalize while walking
- Calculate any function of parameters while walking (error propagation)
- Find global maximum

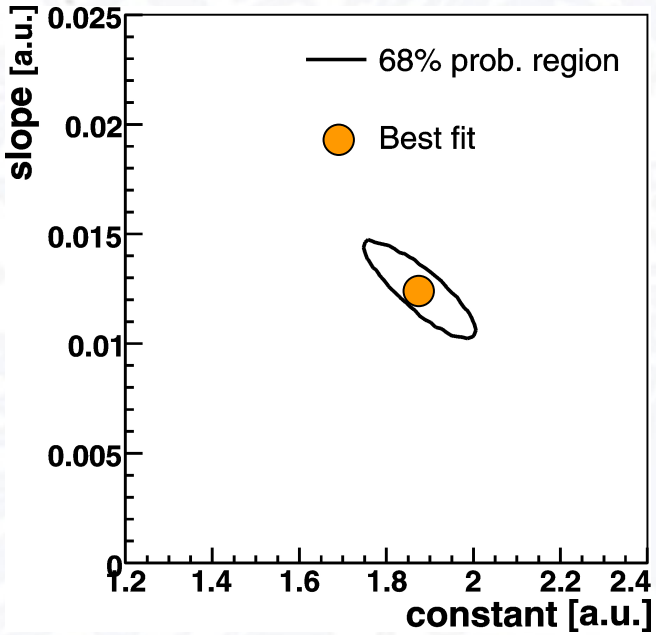


Example: Line Fit





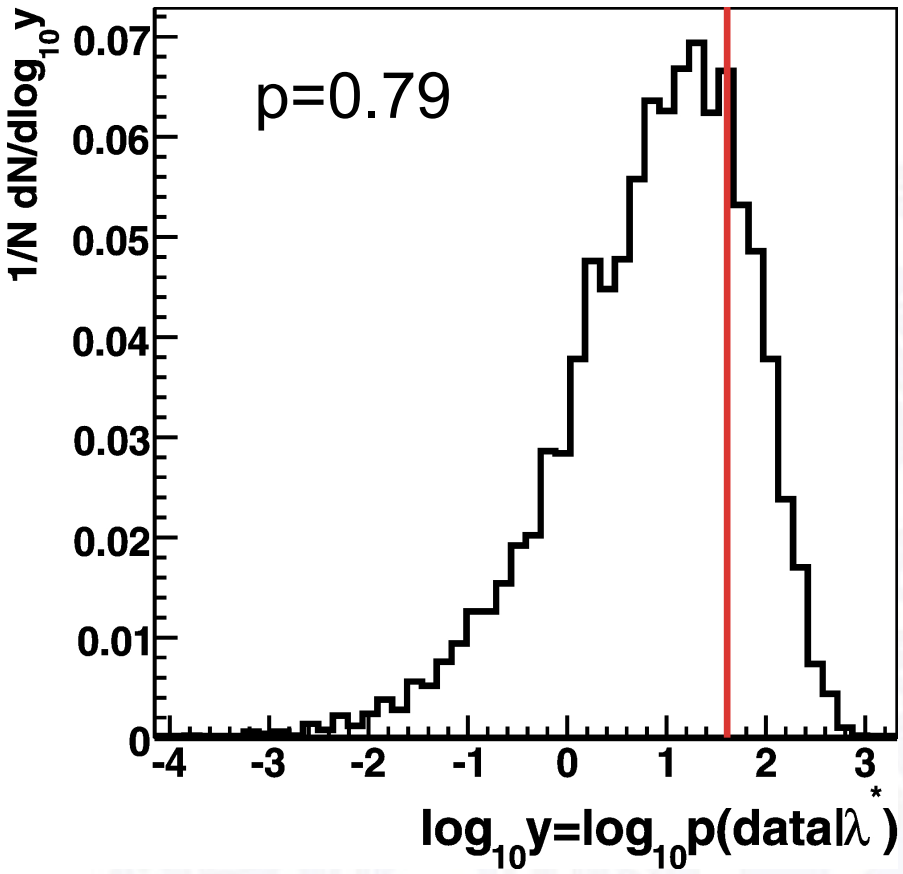




- Assume linear correlation between points
- Gaussian uncertainties
- Error band calculated during MCMC

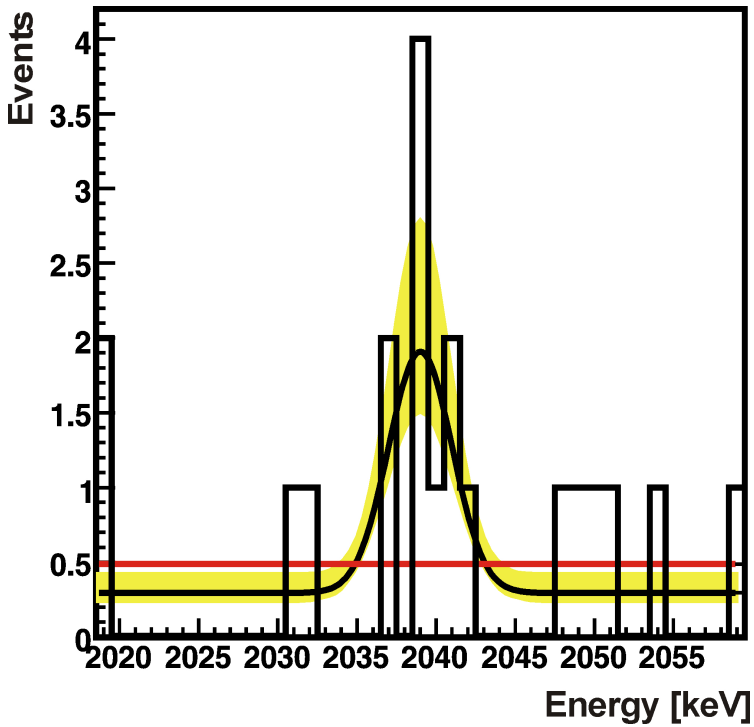
Example: Line Fit

- What is the probability to observe the data given your model and best fit parameters?
- Ensemble tests:
  - Create Monte Carlo data sets assuming best fit parameters
  - Calculate the likelihood for each ensemble
  - Calculate the **p-value**: integral of the frequency distribution to find a less likely result

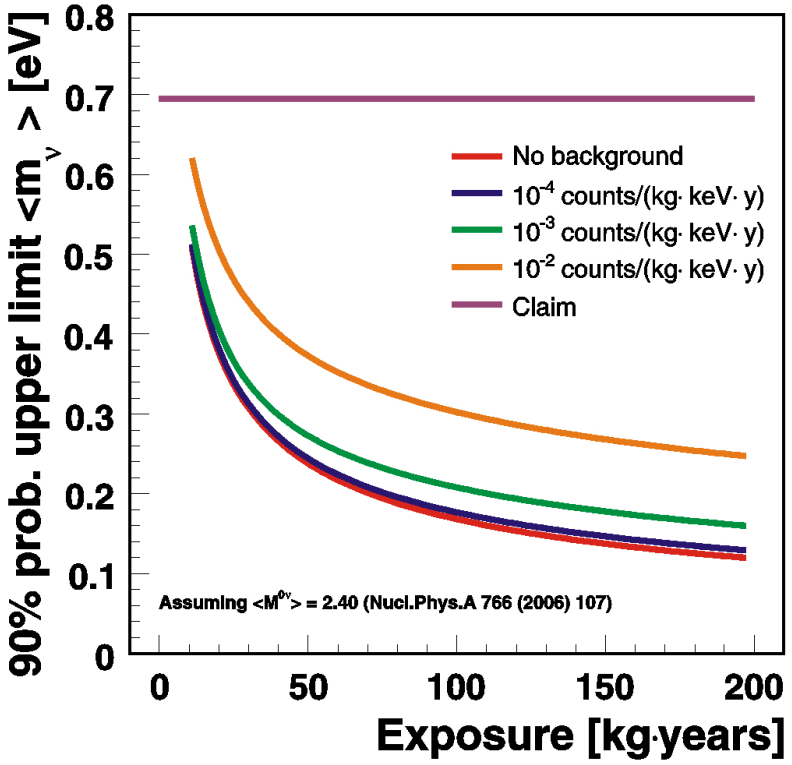


Example: 0νββ-decay @ GERDA

- Search for signal of 0νββ-decay at 2,039 keV
- Gaussian energy resolution assumed



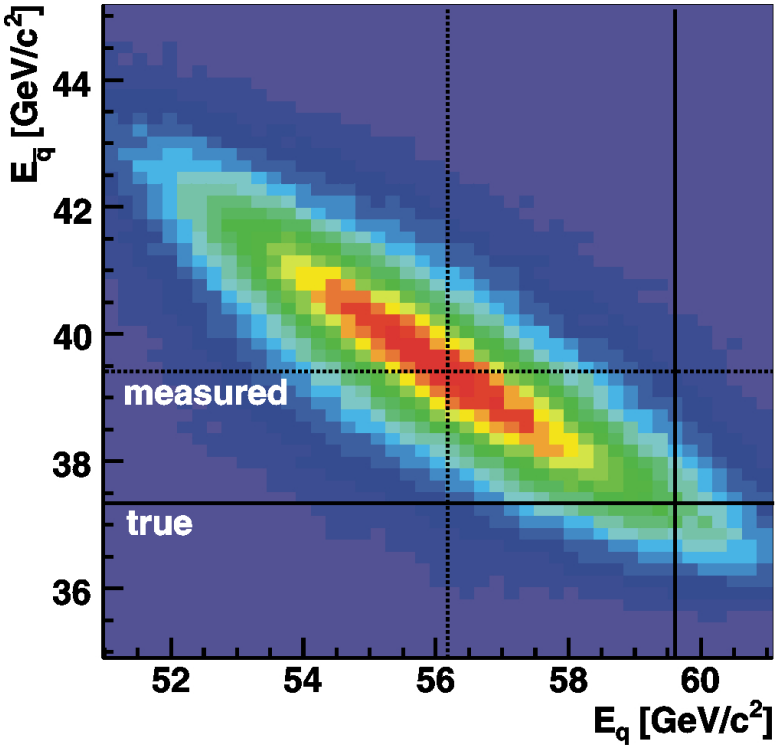
Energy spectrum in the region of interest. The model contains signal and flat background



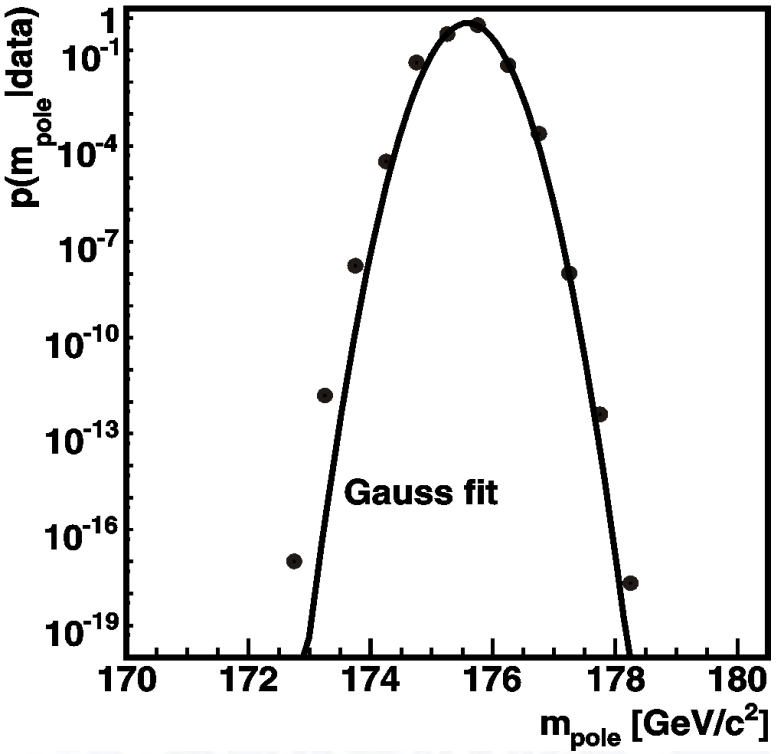
Upper limit on neutrino mass vs. exposure

Example: Top mass @ ATLAS

- Estimate top mass using the matrix element method (A. Knue)
- Integrate over 5 measured energies (4 jets + 1 lepton)



Correlation between two jets combining to a W



Probability of the top pole mass in an ensemble of 300 events

Summary / Outlook

- Toolkit with strong numerical tools **(MCMC)**
- A variety of possible applications in physics:
  - ZEUS (structure functions) (publication in preparation)
  - GERDA (sensitivity calculation)
  - ATLAS (top mass estimate) (see talk by A. Knue)
  - Cosmology, ...
- Publication on BAT in preparation
- **First release expected in May 2008**
- ROOTified version being worked on