The generative model of the surface detector signal in the Pierre Auger Experiment

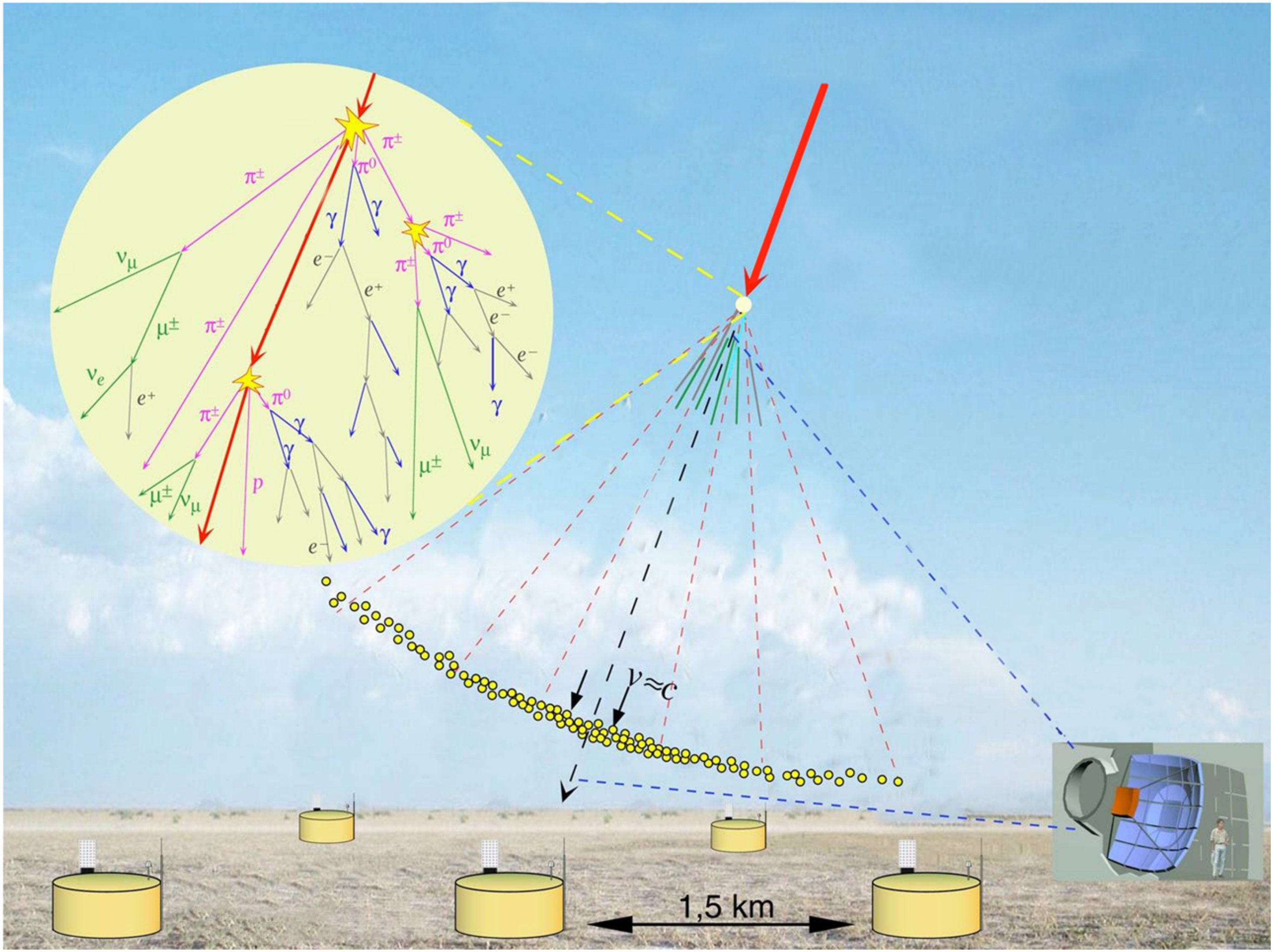
Balázs Kégl

LAL, CNRS/Université Paris Sud

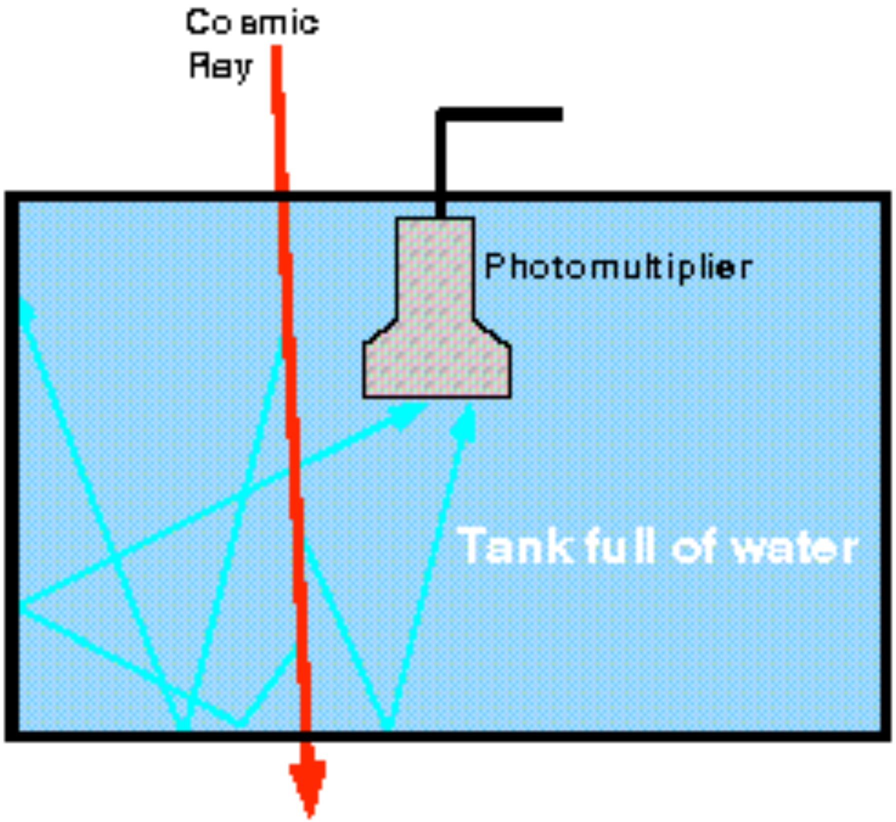
SIMINOLE meeting

Telecom ParisTech, October 25, 2010

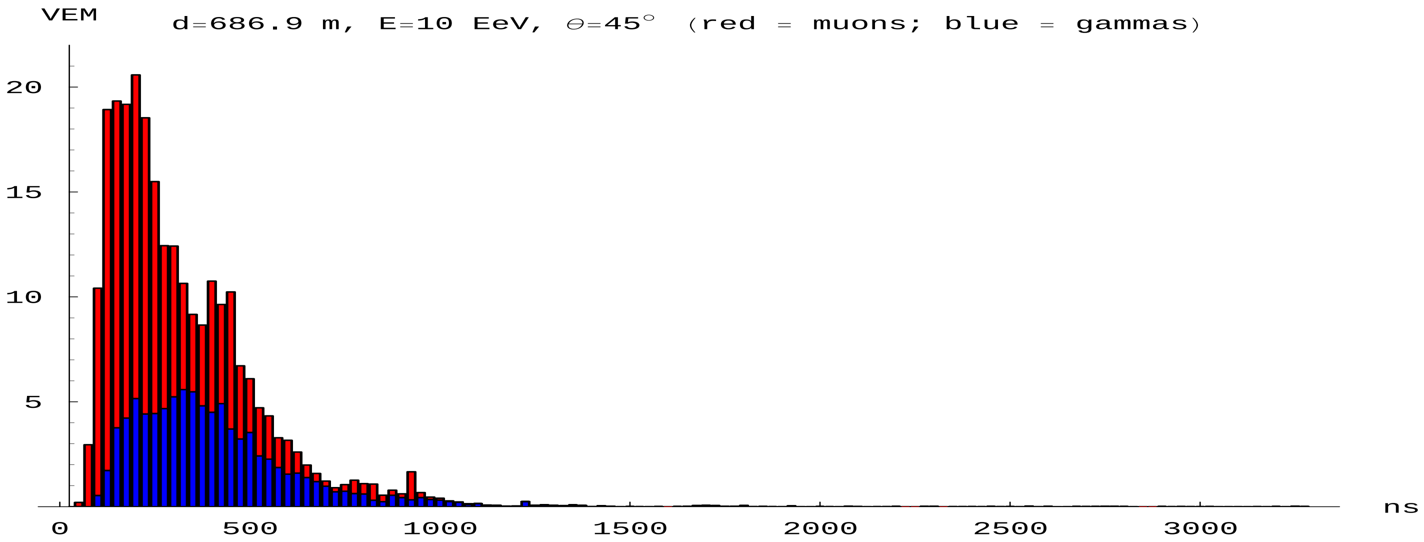
The shower disk

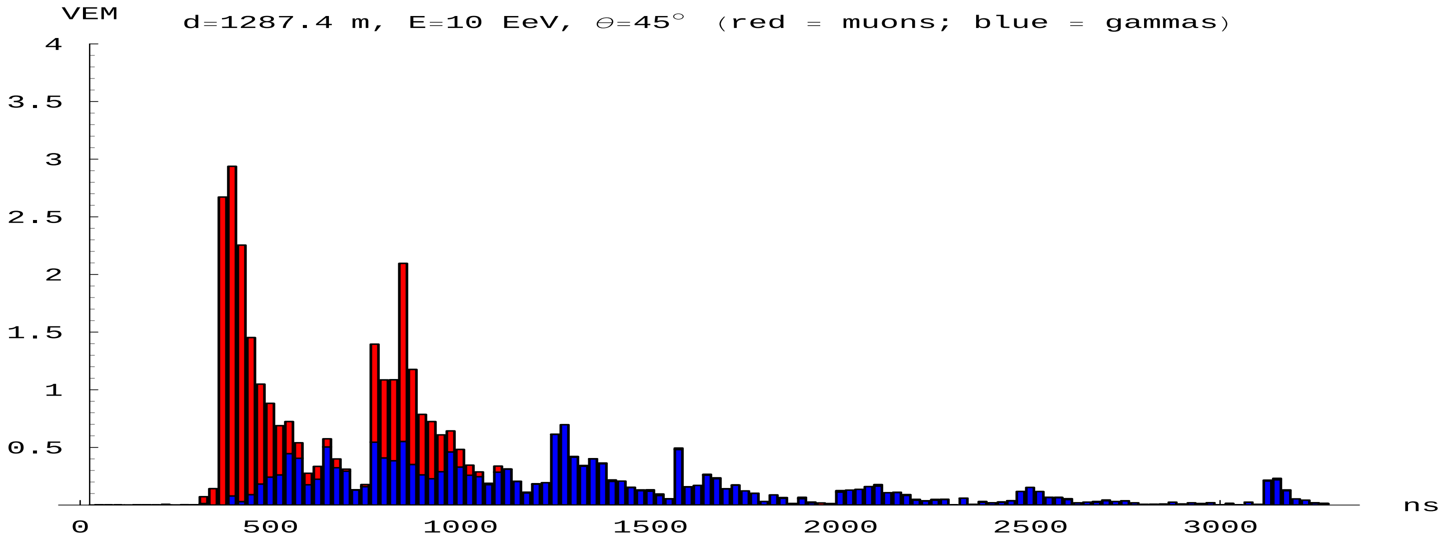


The Cherenkov light



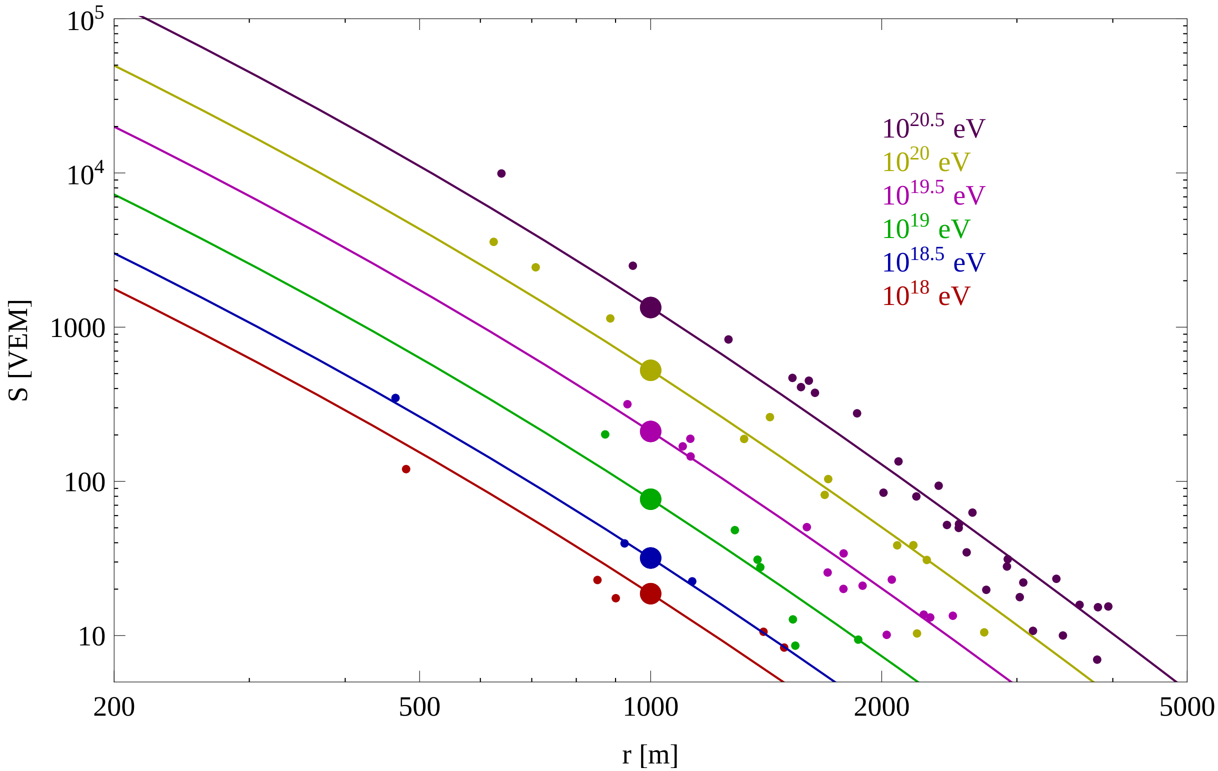
The (simulated) tank signal





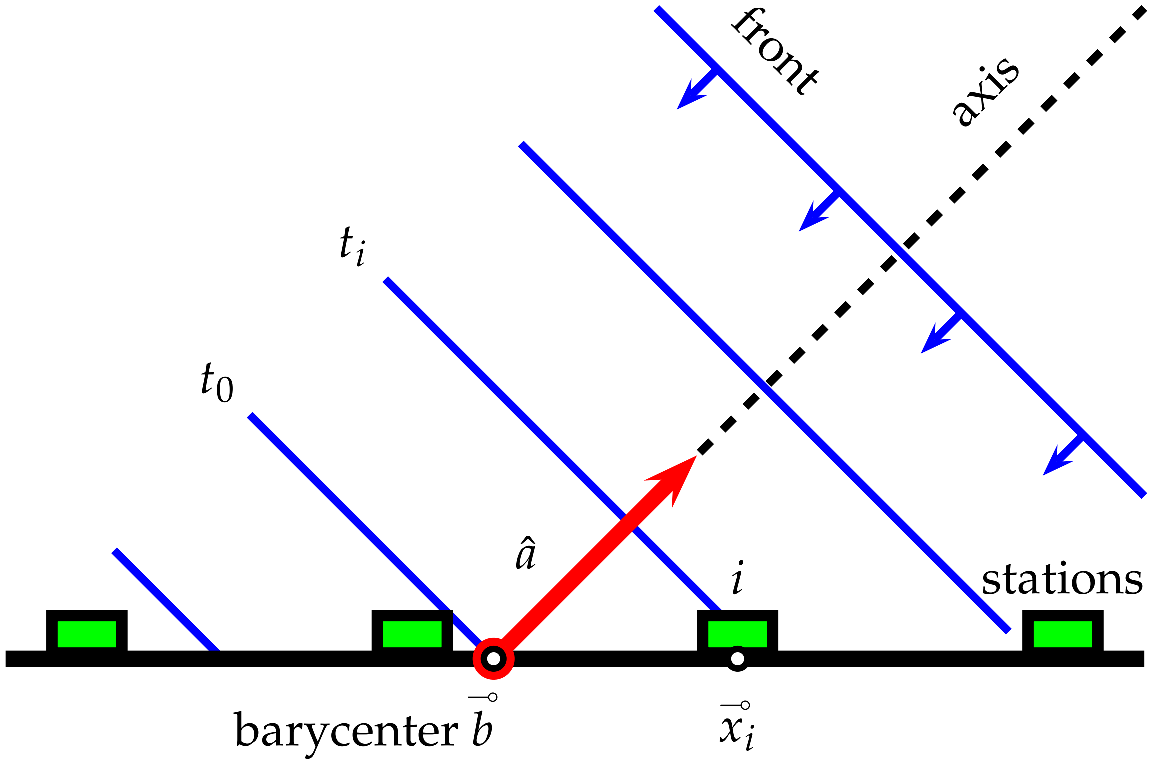
Energy reconstruction

- The primary energy is related to the tank signal integrals



Angle reconstruction

- The angles (zenith, azimuth) are related to the start times of the tank signals

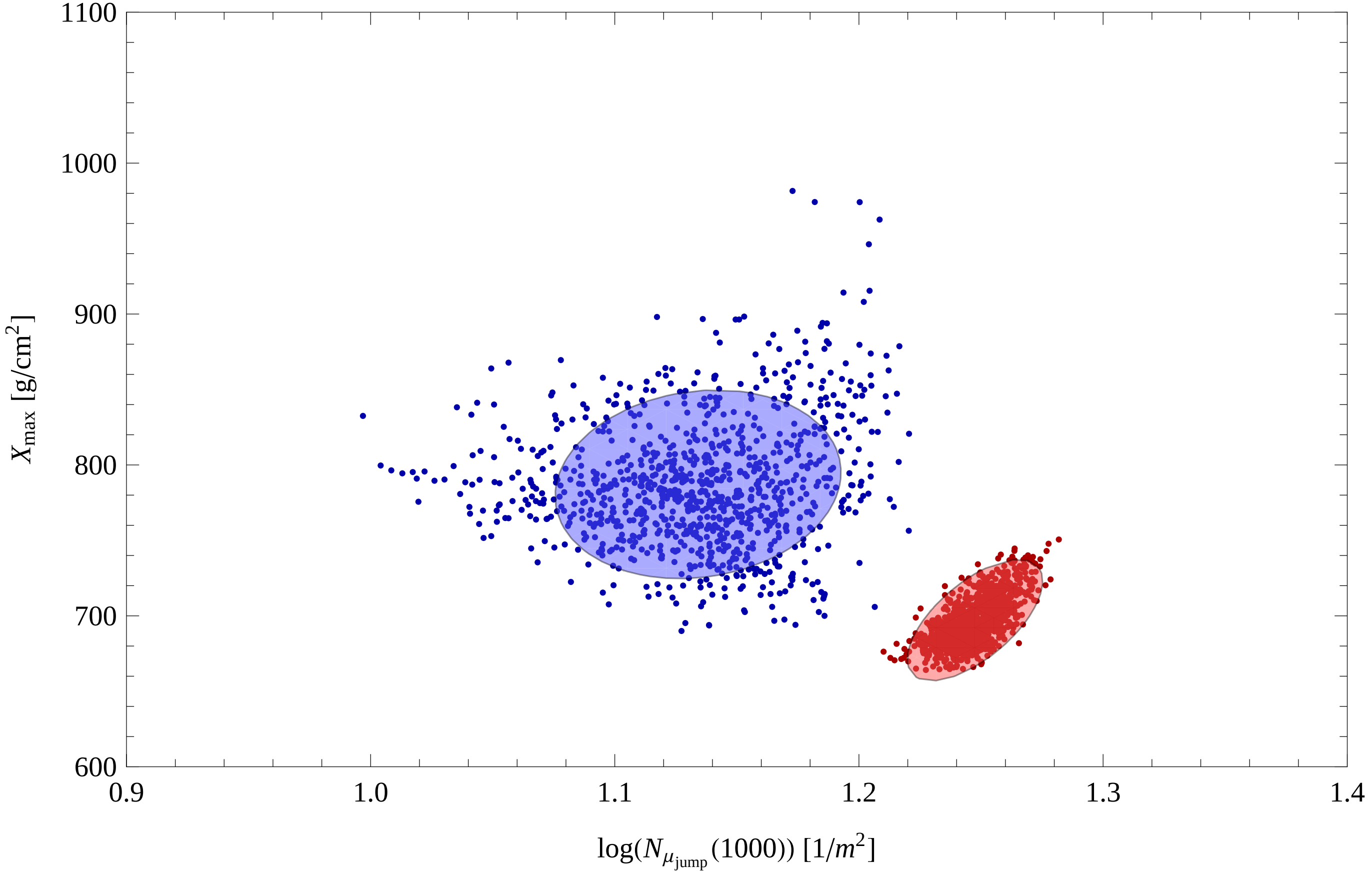


Primary type reconstruction

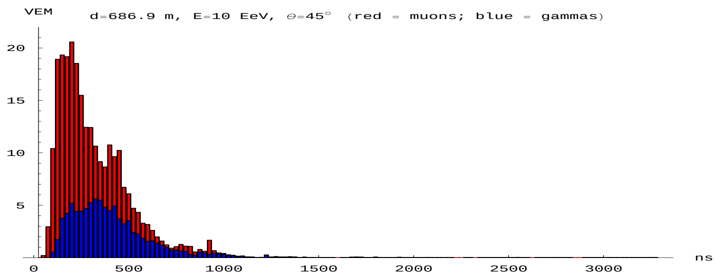
- Candidate nuclei: proton, iron, in between (CNO, Helium)
- Two discriminants:
  - the heavier the nucleus, the shallower is the shower $\rightarrow X_{\max}$, asymmetry, risetime
  - the heavier the nucleus, the higher is the number of muons

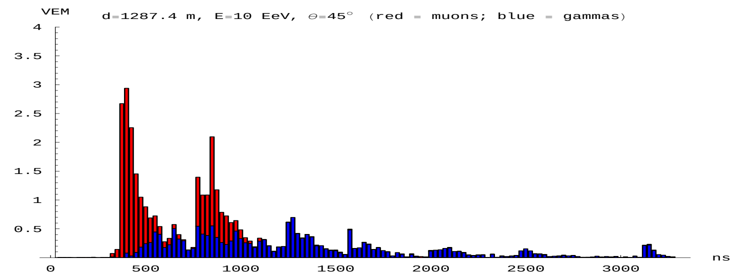
Primary type reconstruction

- proton, iron; zenith angle $\theta = 50^\circ$; energy $E = 10^{19}\,\mathrm{eV}$; no measurement error; bias in simulation



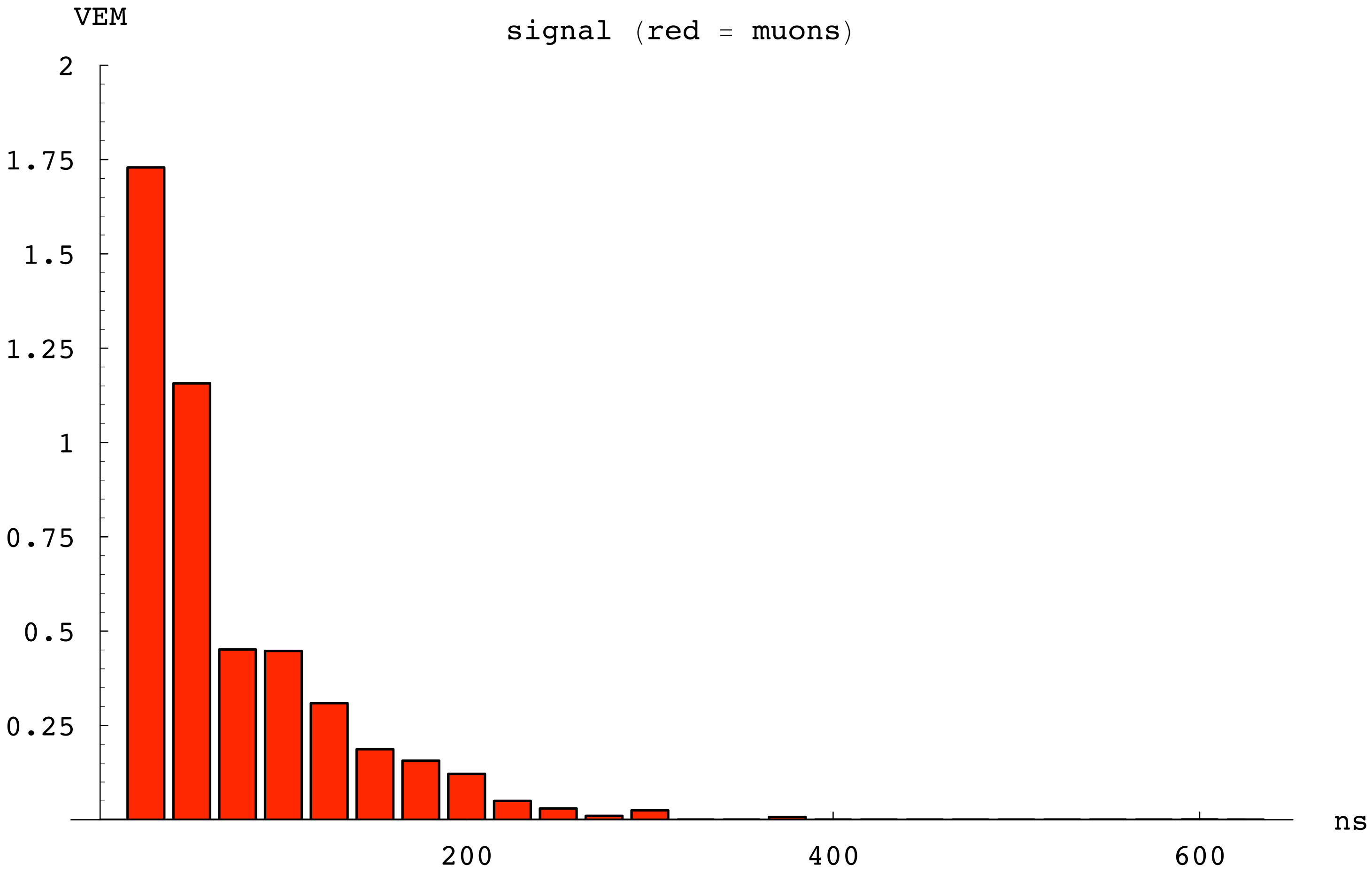
The surface detector signal



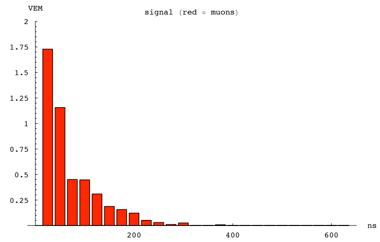


- Energy of the primary is related to the signal integral
- Close to the core only the general shape of the shower "disk" is visible
- Further the signal is more elongated and smaller, so individual particles become "visible"

The signal of one muon



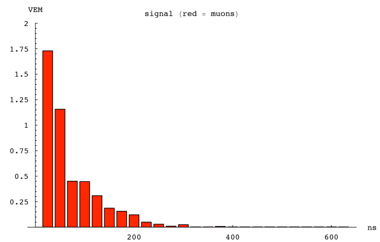
The model of the muonic signal $p(\mathbf{x}|t_\mu, \theta)$



$$p(\mathbf{x}|t_\mu, \theta)$$

- t_μ: arrival time
- θ: zenith angle

The model of the muonic signal $p(\mathbf{x}|t_\mu,\theta)$



- Conditioning on the expected number of photoelectrons per bin $\bar{\mathbf{n}} = (\bar{n}_1,\ldots,\bar{n}_N)$:

$$p(\mathbf{x}|t_\mu,\theta) = \int_{\mathbf{n}\in\mathbb{R}^{+N}} p(\mathbf{x},\bar{\mathbf{n}}|t_\mu,\theta) = \int_{\mathbf{n}\in\mathbb{R}^{+N}} p(\mathbf{x}|\theta,t_\mu,\bar{\mathbf{n}})\,p(\bar{\mathbf{n}}|\theta,t_\mu)$$

- Given $\bar{\mathbf{n}}$, $\mathbf{x}$ does not depend on θ and t_μ:

$$p(\mathbf{x}|t_\mu,\theta) = \int_{\mathbf{n}\in\mathbb{R}^{+N}} p(\mathbf{x}|\theta,t_\mu,\bar{\mathbf{n}})\,p(\bar{\mathbf{n}}|\theta,t_\mu) = \int_{\bar{\mathbf{n}}\in\mathbb{R}^{+N}} p(\mathbf{x}|\bar{\mathbf{n}})\,p(\bar{\mathbf{n}}|\theta,t_\mu)$$

The model of the binwise signal $p(\mathbf{x}|\bar{\mathbf{n}})$

- Binwise independence:

$$p(\mathbf{x}|\bar{\mathbf{n}}) = \prod_{i=1}^{N} p(x_i|\bar{n}_i)$$

- Conditioning on the actual number of photoelectrons per bin $\mathbf{n} = (n_1, \ldots, n_N)$:

$$p(x_i \mid \bar{n}_i) = \sum_{n_i=0}^{\infty} p(x_i, n_i \mid \bar{n}_i) = \sum_{n_i=0}^{\infty} p(x_i \mid n_i, \bar{n}_i) p(n_i \mid \bar{n}_i)$$

- Given n_i, x_i does not depend on $\bar{n}_i$:

$$p(x_i \mid \bar{n}_i) = \sum_{n_i=0}^{\infty} p(x_i \mid n_i) p(n_i \mid \bar{n}_i)$$

The model of the binwise signal $p(x_i|n_i)$

- Conditioning on the noiseless signal per bin $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_N)$:

$$p(x_i \mid n_i) = \int_{\bar{x}_i \in \mathbb{R}^+} p(x_i, \bar{x}_i \mid n_i) = \int_{\bar{x}_i \in \mathbb{R}^+} p(x_i \mid \bar{x}_i, n_i) p(\bar{x}_i \mid n_i)$$

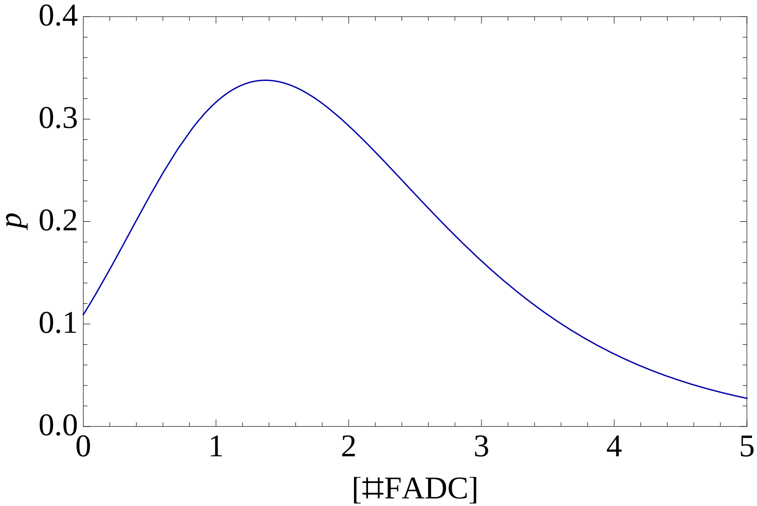
- Given $\bar{x}_i$, x_i does not depend on n_i:

$$p(x_i \mid n_i) = \int_{\bar{x}_i \in \mathbb{R}^+} p(x_i \mid \bar{x}_i) p(\bar{x}_i \mid n_i)$$

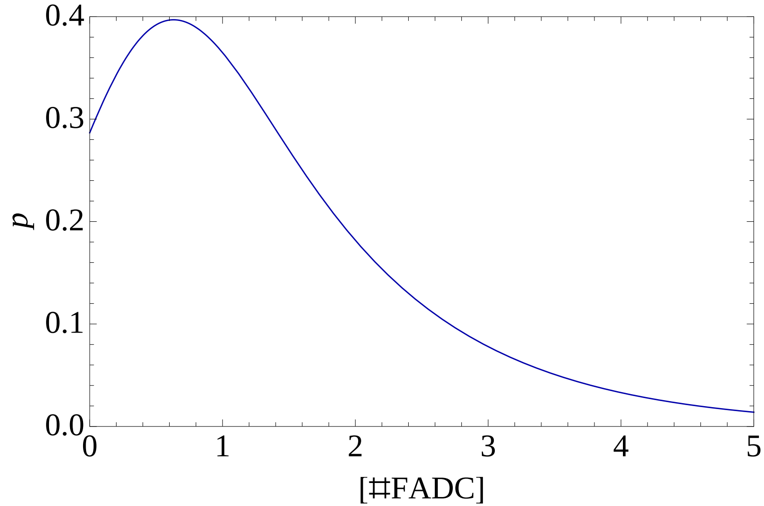
The model of the noiseless signal $p(\bar{x}_i \mid n_i)$ and the noise $p(x_i|\bar{x}_i)$

- $p(\bar{x}_i \mid n_i) = \Gamma_{k,\mu/k}(\bar{x}_i)^{*n_i} = \Gamma_{n_i k,\mu/k}(\bar{x}_i)$
    - μ: gain
    - k: shape
- $p(x_i \mid \bar{x}_i) = \mathcal{N}_{b+\bar{x},\sigma}(x_i)$
    - b: baseline (pedestal)
    - σ: noise standard deviation

$p(x_i \mid n_i = 1)$



(a) $k = 2.5$, $\mu = 1.9$, $\sigma = 0.6$



(b) $k = 1.05$, $\mu = 1.2$, $\sigma = 0.6$

The model of the binwise signal $p(\mathbf{x}|\bar{\mathbf{n}})$

- The model of the number of photoelectrons:

$$p(n_i \mid \bar{n}_i) = \mathrm{Poi}_{\bar{n}_i}(n_i)$$

- Putting together:

$$\begin{aligned} p(x_i \mid \bar{n}_i) &= \sum_{n_i=0}^{\infty} \mathrm{Poi}_{\bar{n}_i}(n_i) \int_{\bar{x}_i \in \mathbb{R}^+} \Gamma_{n_i k, \mu/k}(\bar{x}_i) \mathcal{N}_{b+\bar{x}_i,\sigma}(x_i) \\ &= \int_{\bar{x}_i \in \mathbb{R}^+} \mathcal{N}_{b,\sigma}(x_i - \bar{x}_i) \underbrace{\sum_{n_i=0}^{\infty} \mathrm{Poi}_{\bar{n}_i}(n_i) \Gamma_{n_i k, \mu/k}(\bar{x}_i)}_{\text{compound Poisson}} \end{aligned}$$

One last step: discretization

- x_i is an integer FADC count

$$p(x_i \mid \bar{n}_i) = \sum_{n_i=0}^{\infty} \mathrm{Poi}_{\bar{n}_i}(n_i) \int_{\bar{x}_i \in \mathbb{R}^+} \Gamma_{n_i k, \mu/k}(\bar{x}_i) \int_{x'_i \in [x_i - 0.5, x_i + 0.5]} \mathcal{N}_{b+\bar{x}_i, \sigma}(x'_i)$$

  - Could be approximated, especially for large signals x_i
  - Or explicitly include nuisance variables x_i, $\bar{x}_i$, and n_i in MCMC

- Not yet modeled:
  - Average of three PMs

The model of the average muonic signal $p(\bar{\mathbf{n}}|t_\mu, \theta)$

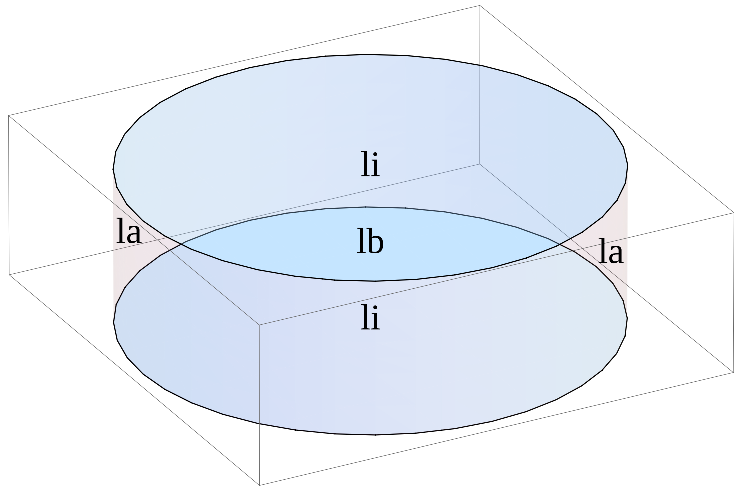
- Conditioning on the tracklength L_μ and the energy E_μ of the muon:

$$p(\bar{\mathbf{n}}|t_\mu, \theta) = \int_{L_\mu, E_\mu} p(\bar{\mathbf{n}}|t_\mu, \theta, L_\mu, E_\mu) p(L_\mu, E_\mu|t_\mu, \theta)$$

- L_μ depends only on θ; $\bar{\mathbf{n}}$ is independent of θ given L_μ and E_μ; L_μ is independent of E_μ:

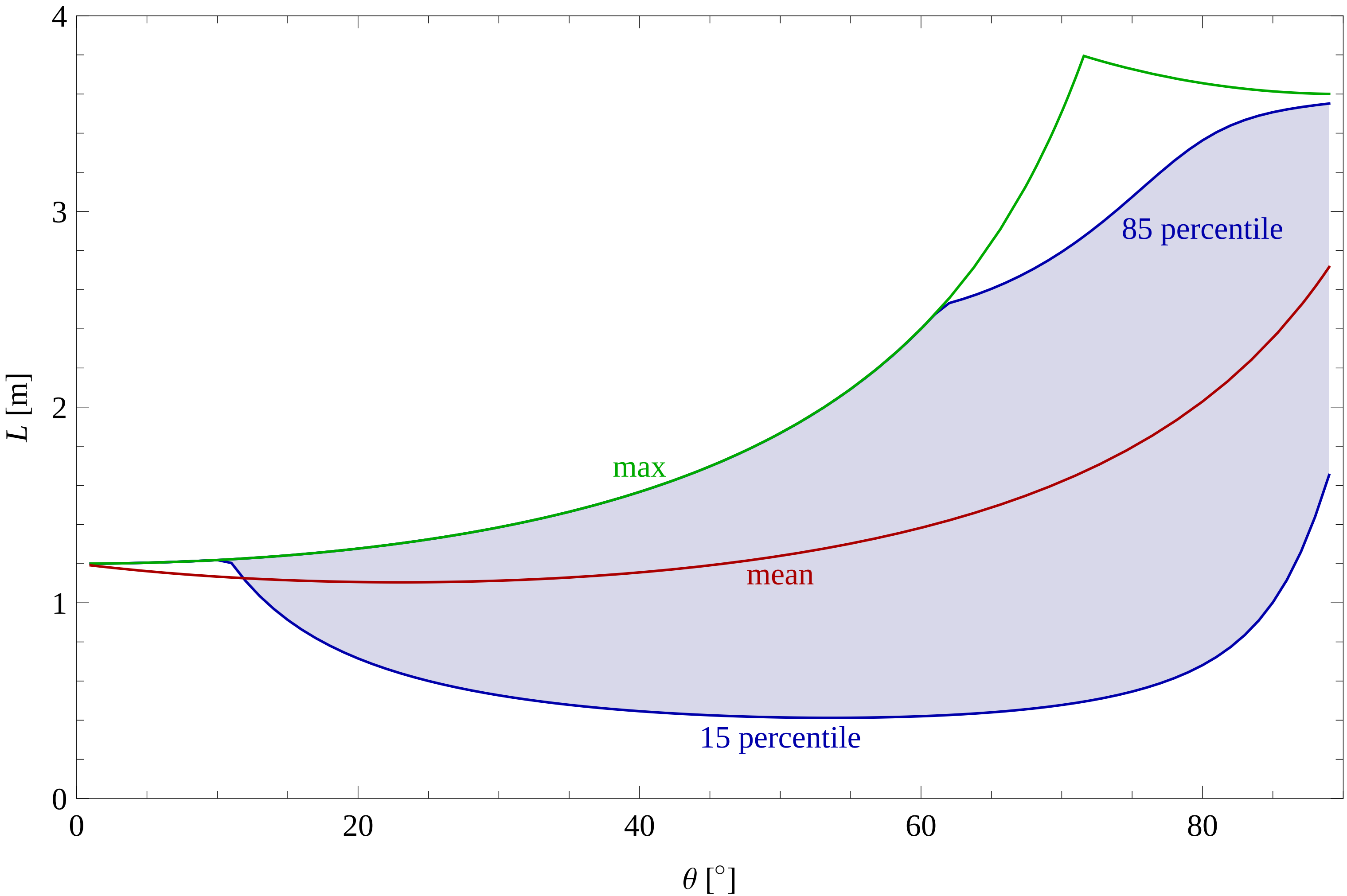
$$p(\bar{\mathbf{n}}|t_\mu, \theta) = \int_{L_\mu, E_\mu} p(\bar{\mathbf{n}}|t_\mu, L_\mu, E_\mu) p(L_\mu|\theta) p(E_\mu|t_\mu, \theta)$$

The model of the tracklength distribution $p(L_\mu|\theta)$



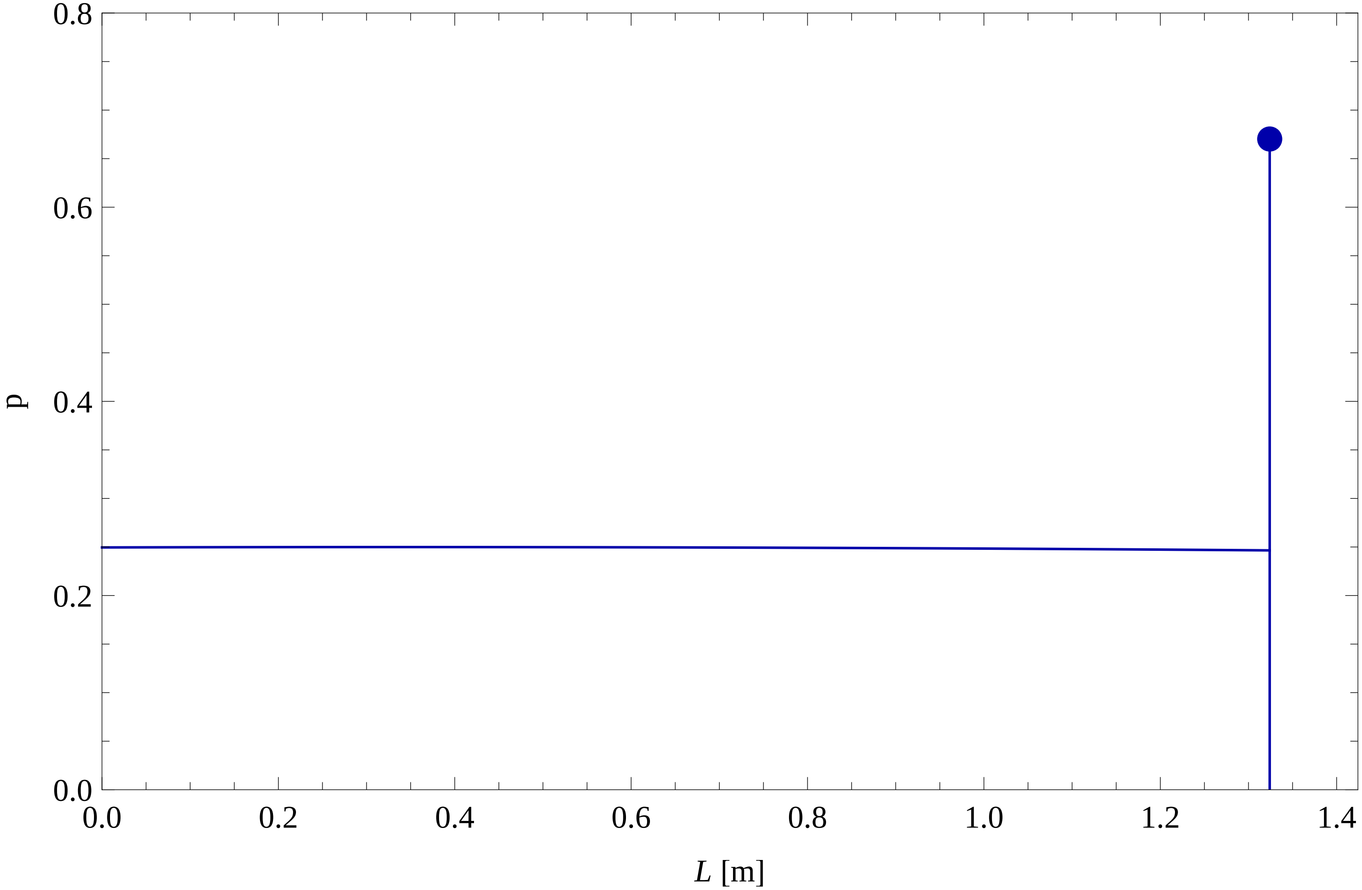
$$
\begin{aligned}
p(L_\mu|\theta) &= p_{\mathrm{lb}}(L_\mu|\theta) + p_{\mathrm{li}}(L_\mu|\theta) + p_{\mathrm{la}}(L_\mu|\theta) \\
p_{\mathrm{lb}}(L_\mu|\theta) &= \frac{1}{\sigma(\theta)}\delta_{s_{\max}(\theta)}\left(2R^2 \arccos\left(\frac{h\tan\theta}{2R}\right)\cos\theta - \frac{1}{2}h\sin\theta\sqrt{4R^2 - h^2\tan^2\theta}\right) \\
p_{\mathrm{li}}(L_\mu|\theta) &= \frac{1}{\sigma(\theta)}\sin(2\theta)\sqrt{4R^2 - (s\sin\theta)^2} \\
p_{\mathrm{la}}(L_\mu|\theta) &= \frac{1}{\sigma(\theta)}\frac{s\,(h - s\cos\theta)\sin^3\theta}{\sqrt{4R^2 - (s\sin\theta)^2}}
\end{aligned}
$$

The model of the tracklength distribution $p(L_\mu|\theta)$



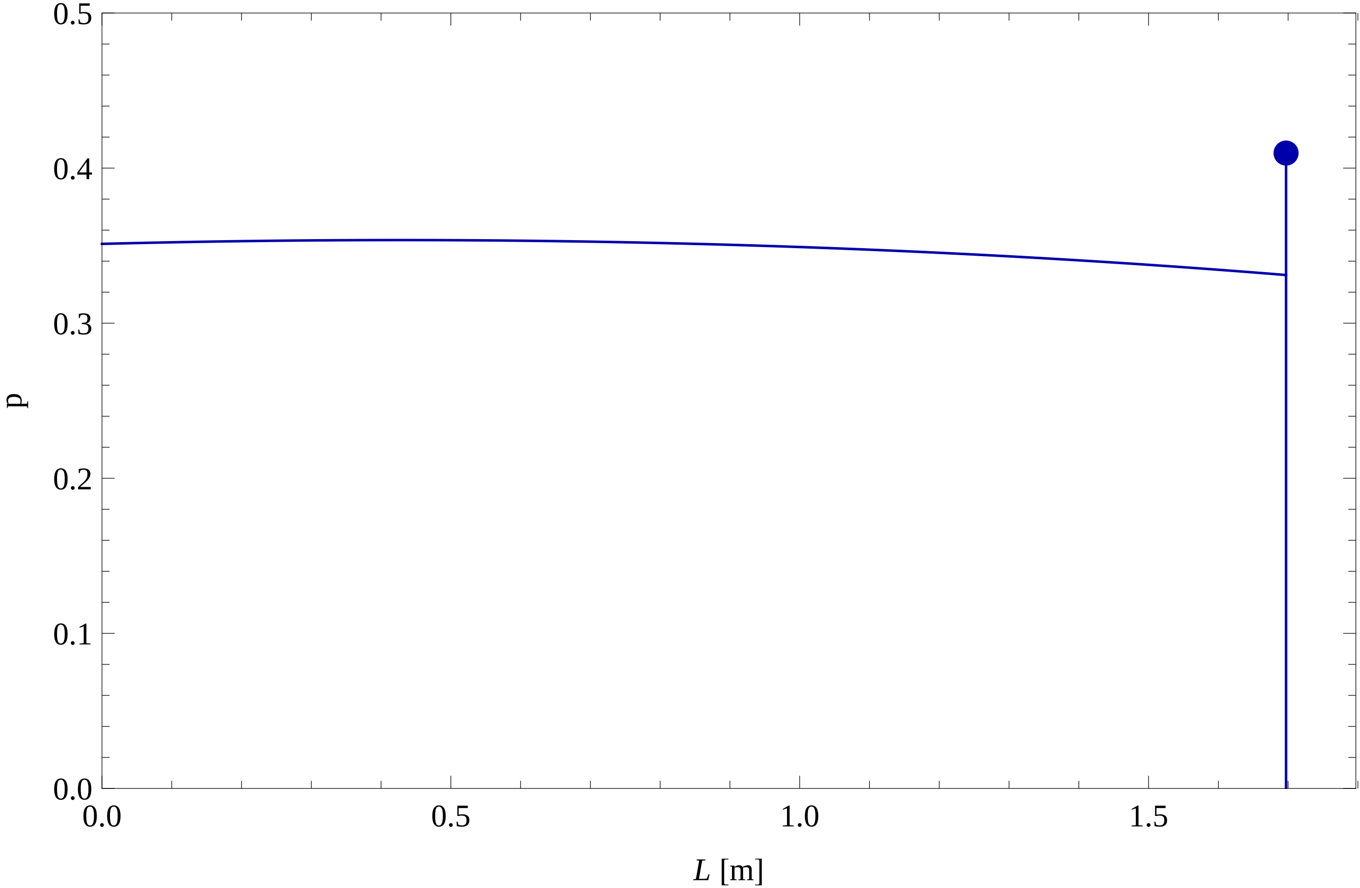
The model of the tracklength distribution $p(L_\mu|\theta)$

- $\theta = 20^\circ$



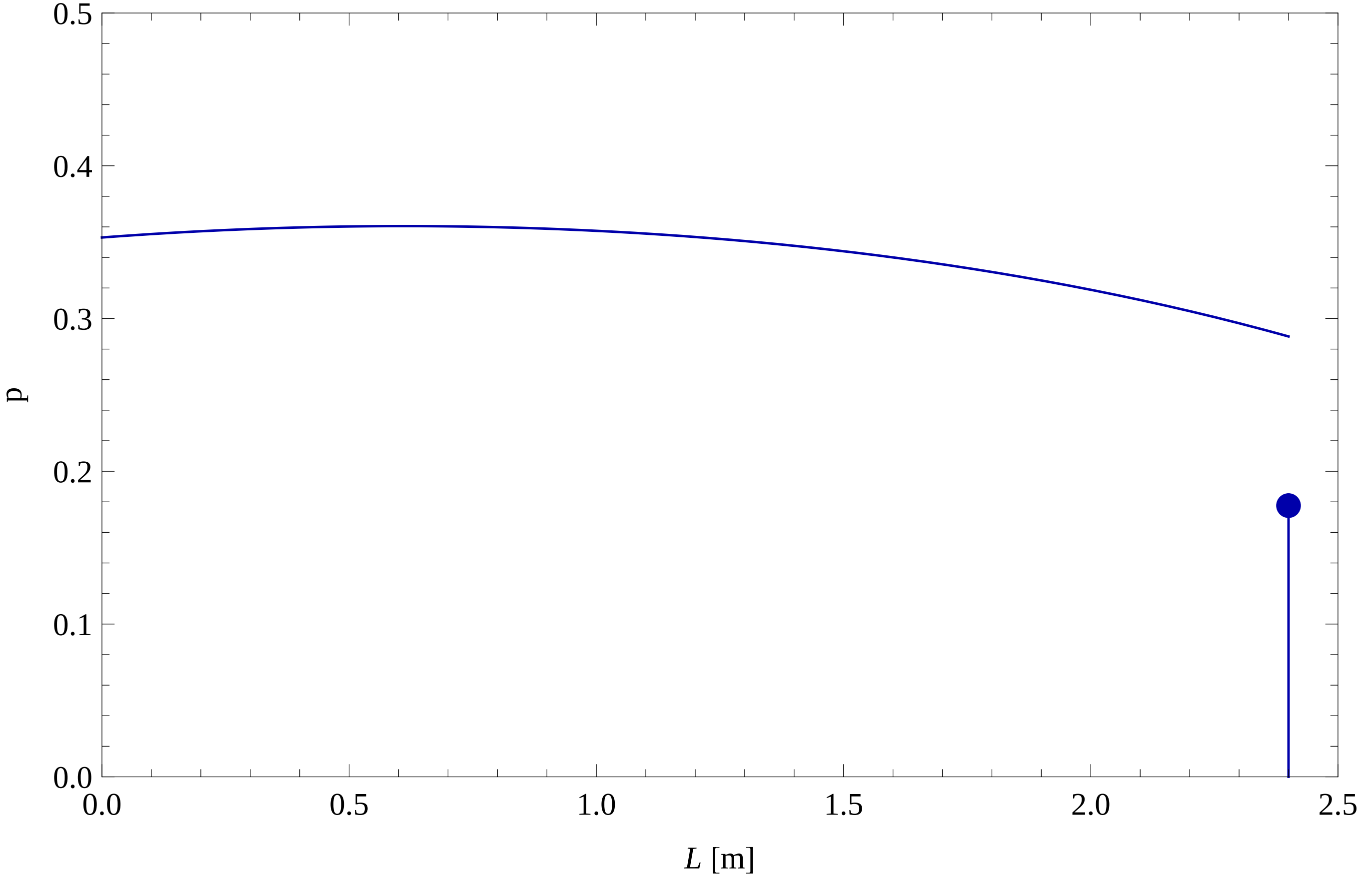
The model of the tracklength distribution $p(L_\mu|\theta)$

- $\theta = 45^\circ$



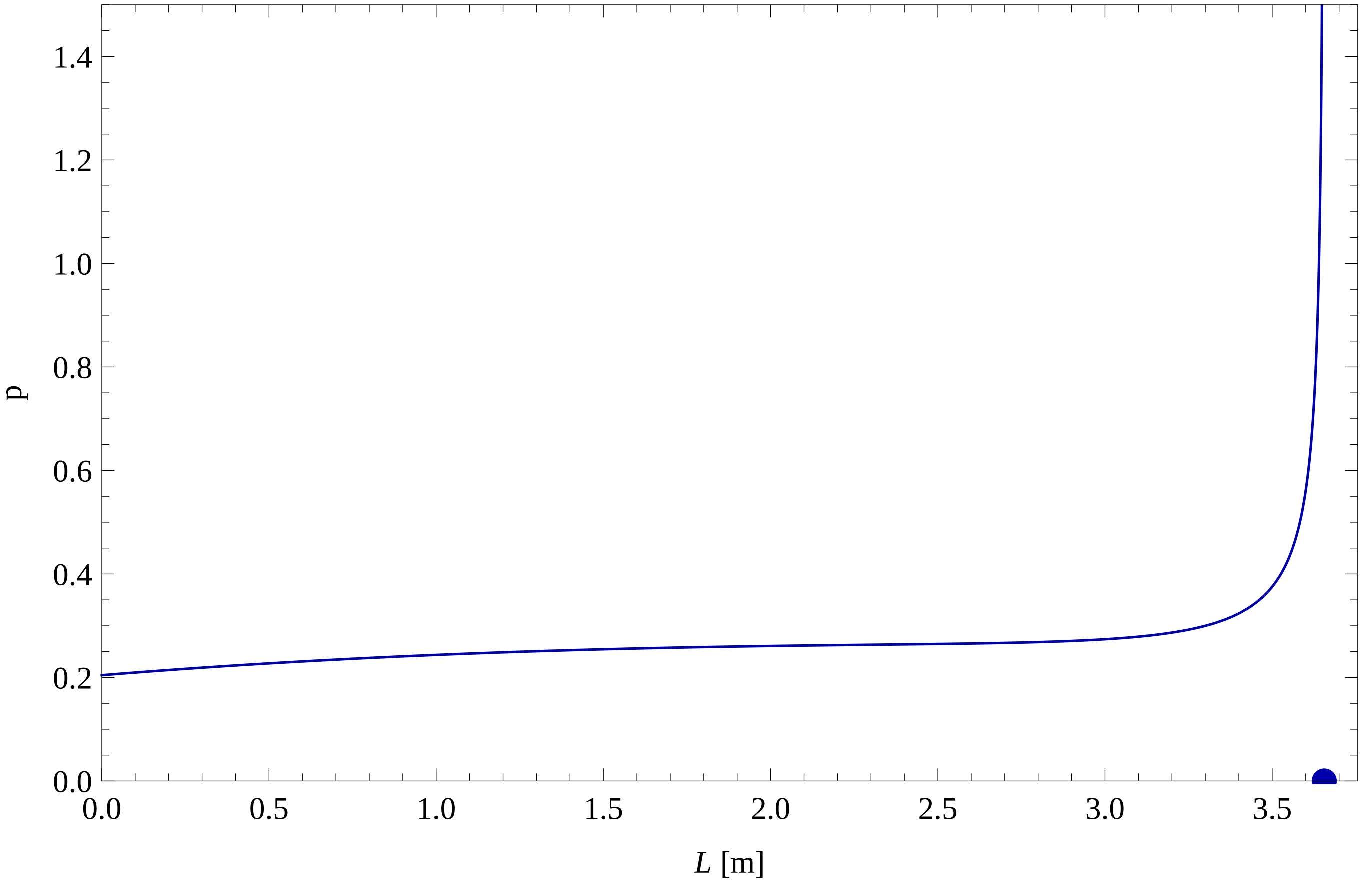
The model of the tracklength distribution $p(L_\mu|\theta)$

- $\theta = 60^\circ$



The model of the tracklength distribution $p(L_\mu|\theta)$

- $\theta = 80^\circ$



The model of the energy distribution $p(E_\mu|t_\mu, \theta)$

- See Karim's talk
  - In fact: $p(E_\mu|t_\mu, \theta, r, \text{etc.})$
  - Also: $p(t_\mu|\theta, r, \text{etc.})$

The model of the average number of PEs $p(\bar{\mathbf{n}}|t_\mu, L_\mu, E_\mu)$

- Conditioning on the energy-dependent signal amplitude ϕ:

$$p(\bar{\mathbf{n}}|t_\mu, L_\mu, E_\mu) = \int_\phi p(\bar{\mathbf{n}}|\phi, t_\mu, L_\mu, E_\mu)\, p(\phi|t_\mu, L_\mu, E_\mu)$$

- ϕ depends only on E_μ; $\bar{\mathbf{n}}$ is independent of E_μ given ϕ:

$$p(\bar{\mathbf{n}}|t_\mu, L_\mu, E_\mu) = \int_\phi p(\bar{\mathbf{n}}|\phi, t_\mu, L_\mu)\, p(\phi|E_\mu)$$

The model of the time profile $p(\bar{\mathbf{n}}|\phi, t_\mu, L_\mu)$

- $p(\bar{\mathbf{n}}|\phi, t_\mu, L_\mu) = \delta_{\bar{\mathbf{n}}(\phi, t_\mu, L_\mu)}$

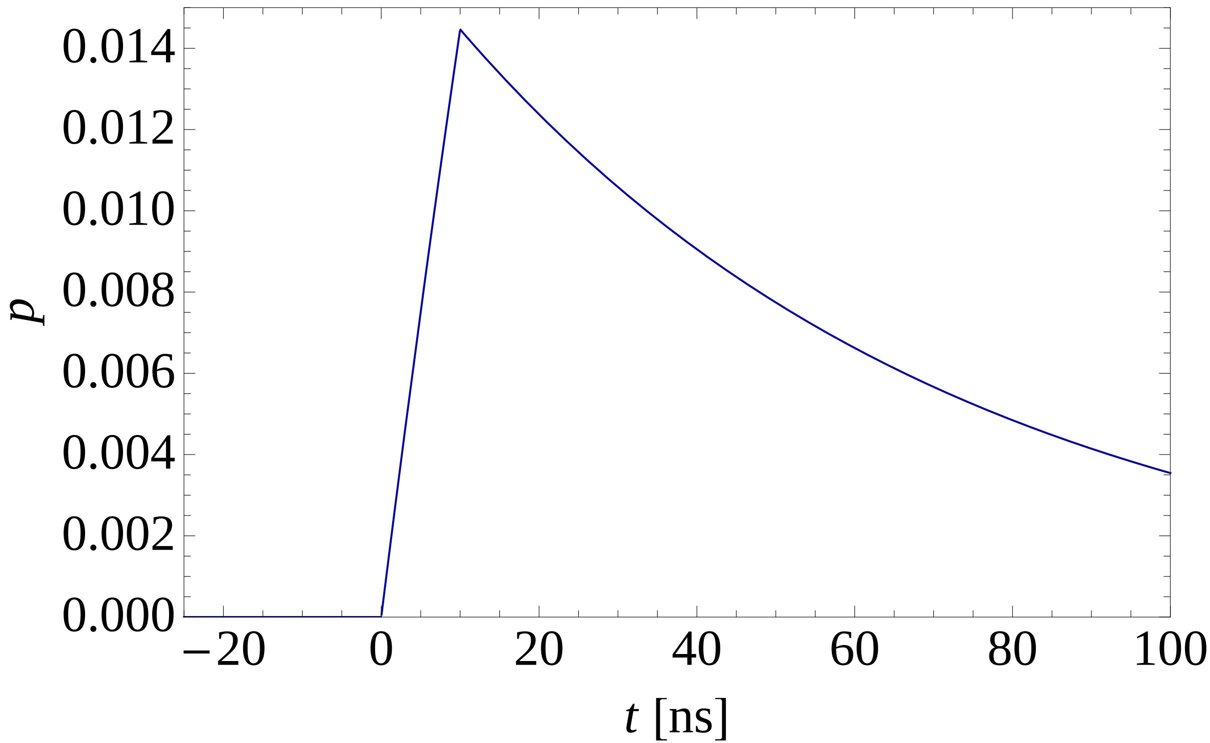
$$\bar{n}_i(\phi, t_\mu, L_\mu) = \phi L_\mu \nu \int_{t \in [t_{i-1}, t_i]} p_{\tau, t_d}(t - t_\mu)$$

- $p_{\tau, t_d}(t) = \frac{1}{t_d} \cdot \begin{cases} 0 & \text{if } t < 0, \\ 1 - \exp(-t/\tau) & \text{if } 0 \leq t < t_d, \\ \exp\big(-(t - t_d)/\tau\big) - \exp(-t/\tau) & \text{if } t_d \leq t. \end{cases}$

- ν is the total number of photoelectrons emitted by an “average” muon on $1\,\mathrm{m}$ tracklength

The model of the time profile $p(\bar{\mathbf{n}}|\phi, t_\mu)$

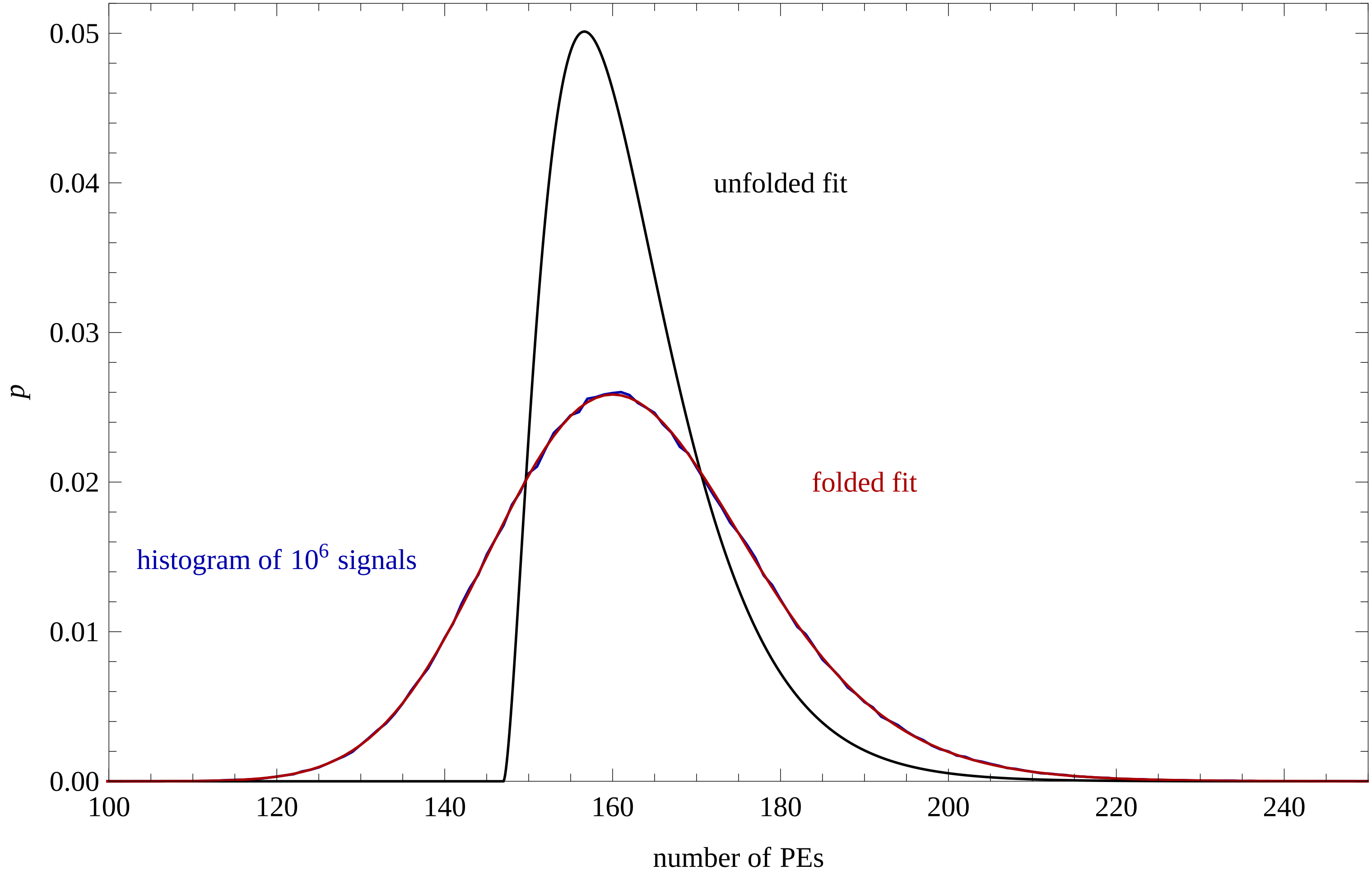
- $p_{\tau,t_d}(t)$ with $\tau = 60\,\mathrm{ns}$ and $t_d = 4\,\mathrm{ns}$



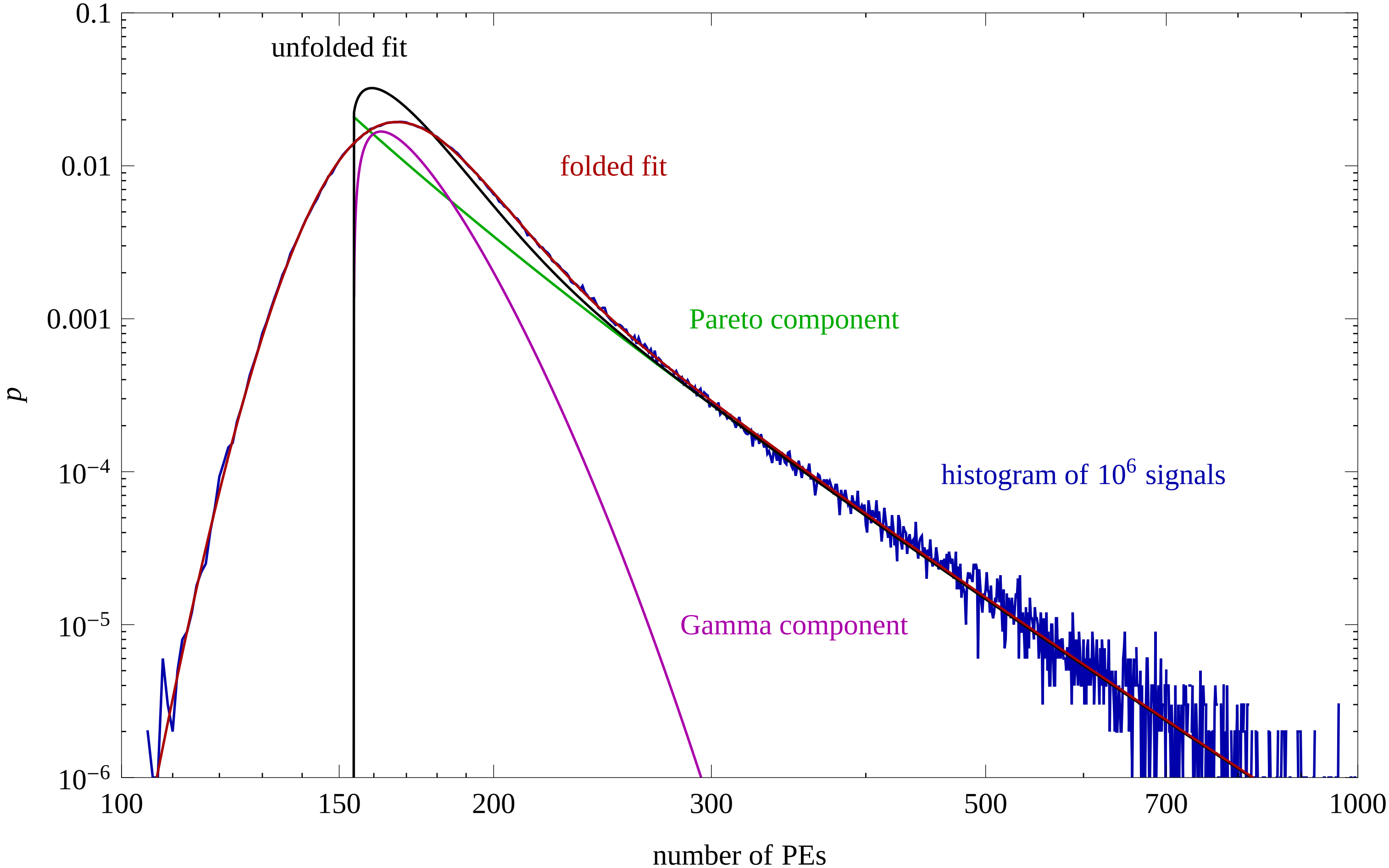
The model of the energy-dependent signal amplitude $p(\phi|E_\mu)$

- $\mathbb{E}\{p(\phi)\} \approx 1$, in fact $p(\phi) \approx \delta_1(\phi)$
- A little bit better: $p(\phi) \approx \mathcal{N}_{1,0.1}(\phi)$
- Otherwise it gets really complicated

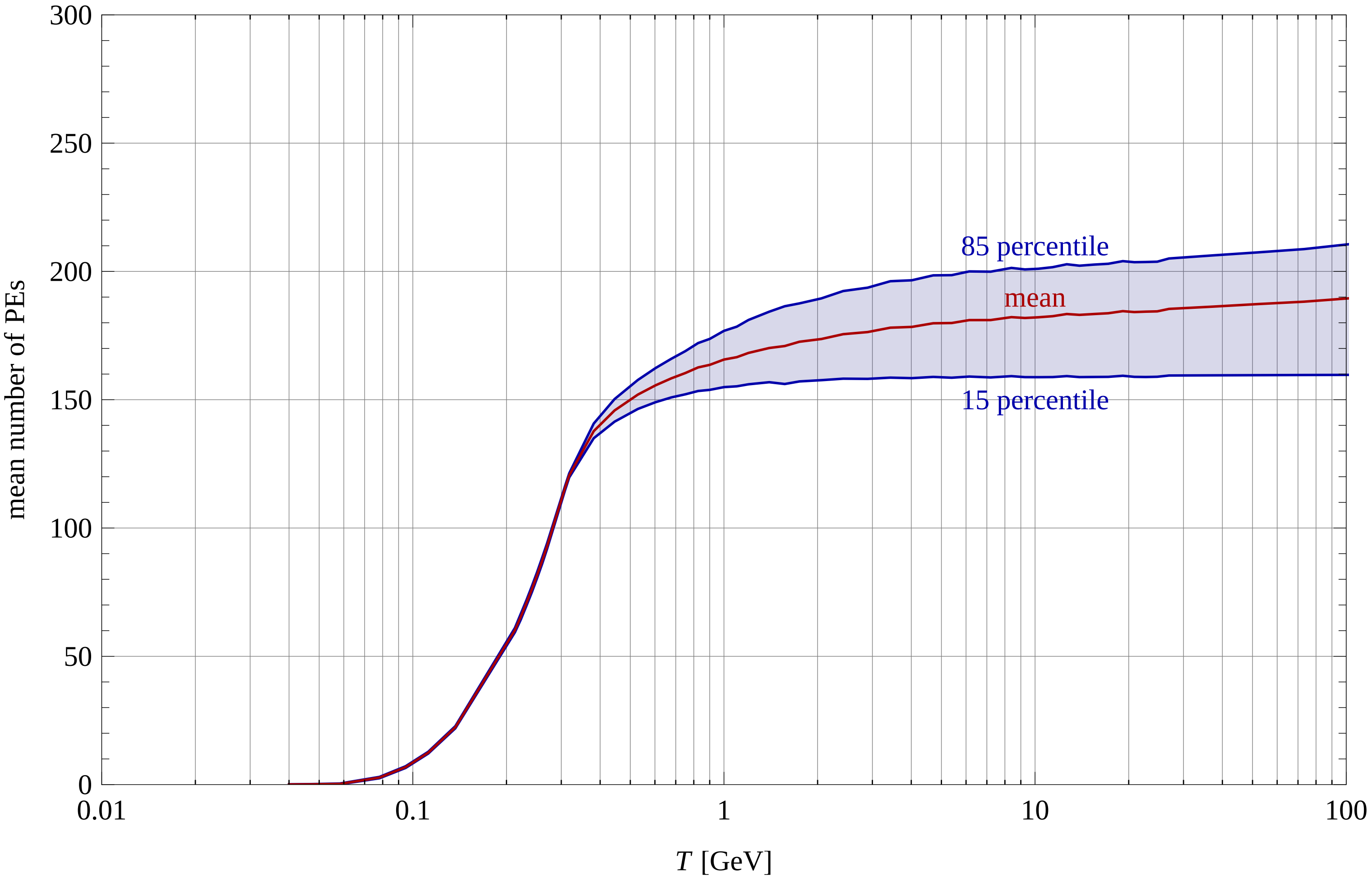
$T = 0.8\,\mathrm{GeV}$: Gamma



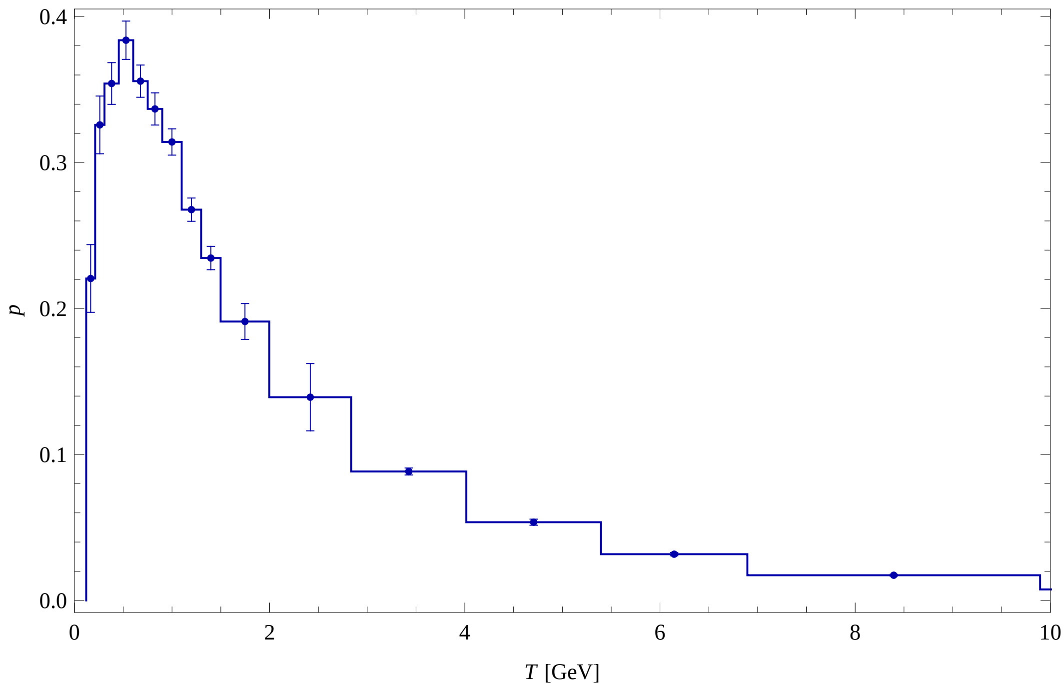
$T = 8\,\mathrm{GeV}$: Gamma + heavy tail



Spread of number of PEs vs. muon kinetic energy

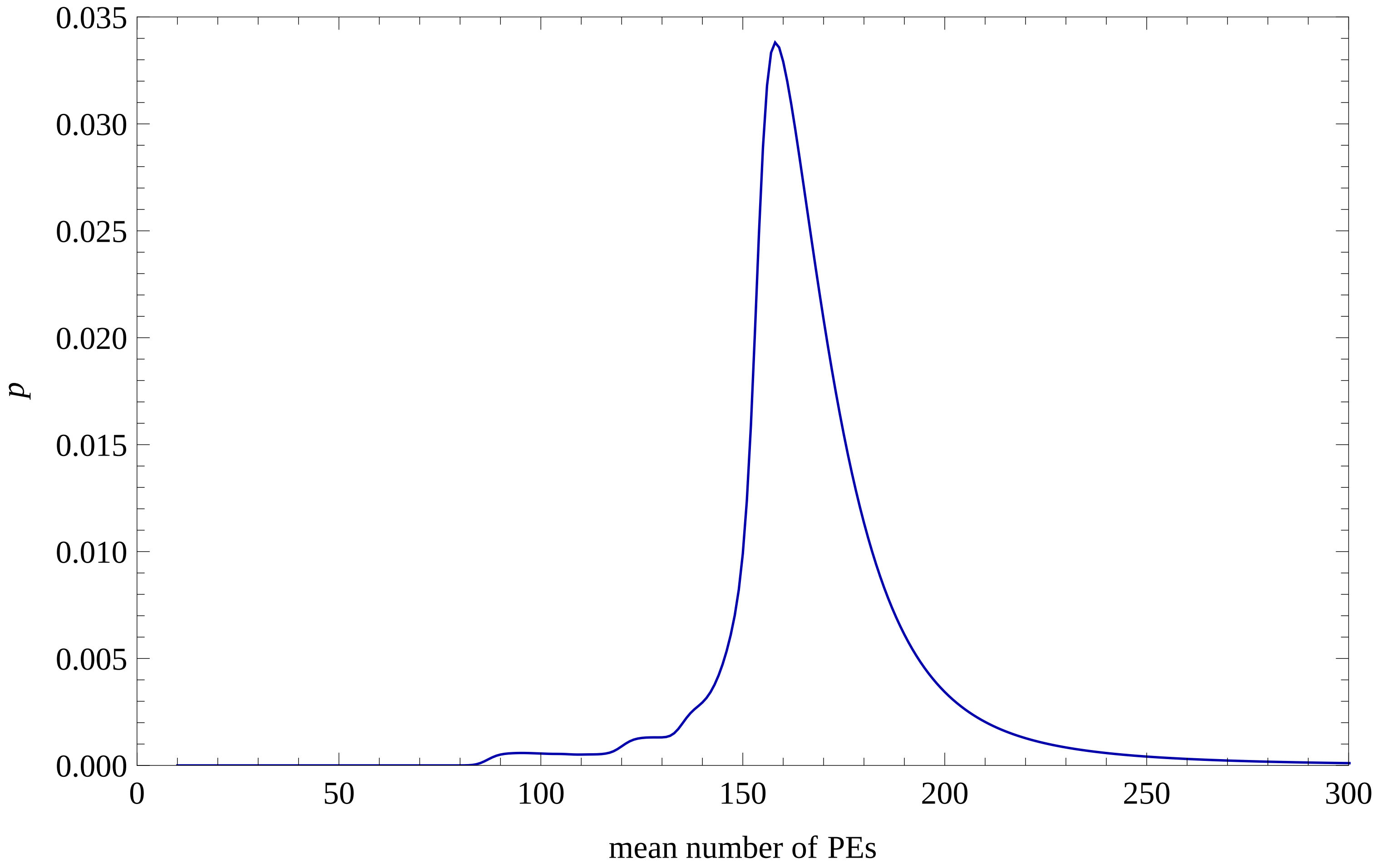


Atmospheric muon energy spectrum (CAPRICE98)

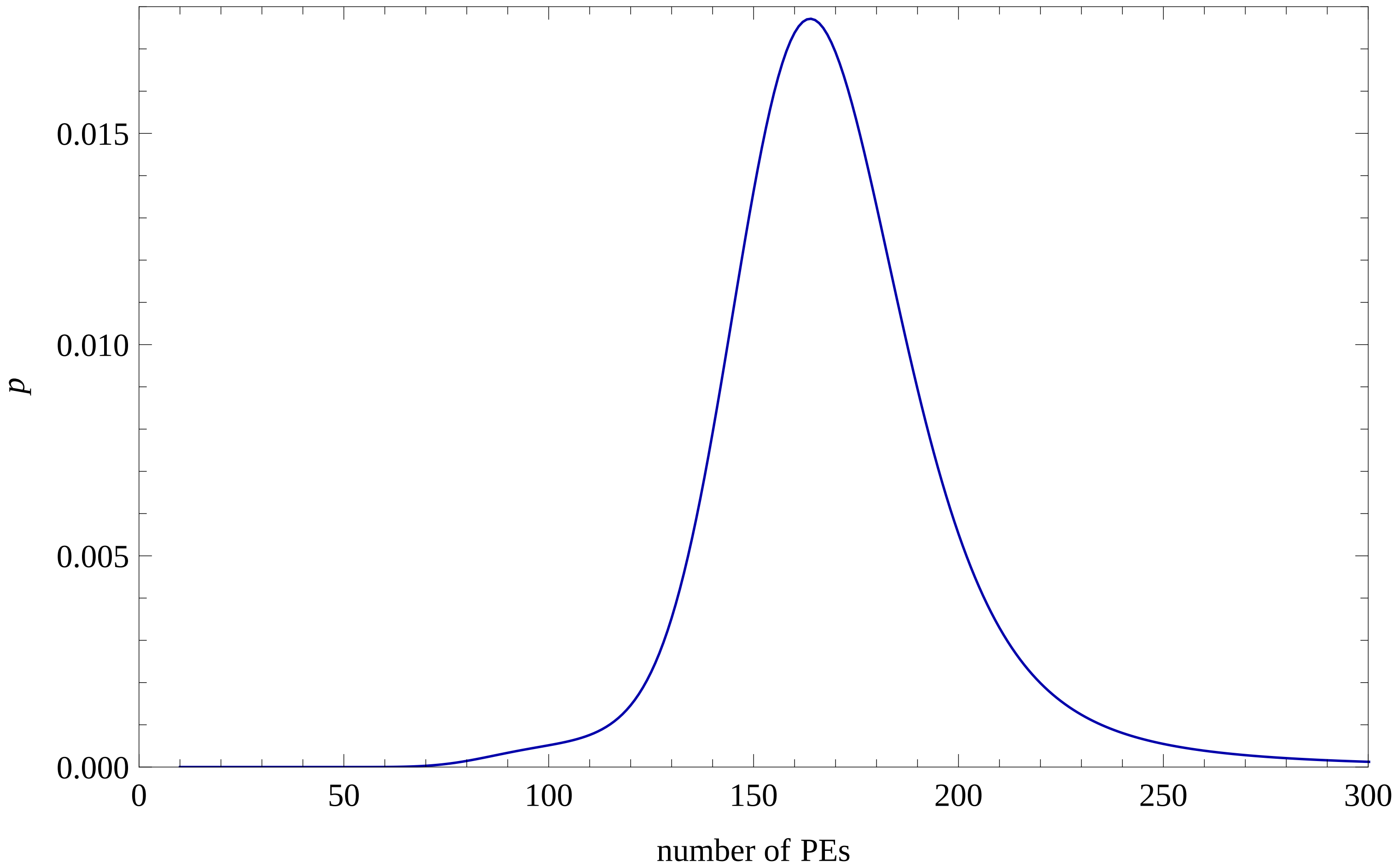


- altitude: $1230\,\mathrm{m} \leftrightarrow 885\,\mathrm{g/cm^2}$
- latitude: North 34°
- zenith angle $\theta \in [0^\circ, 20^\circ]$, average $\bar{\theta} = 9^\circ$

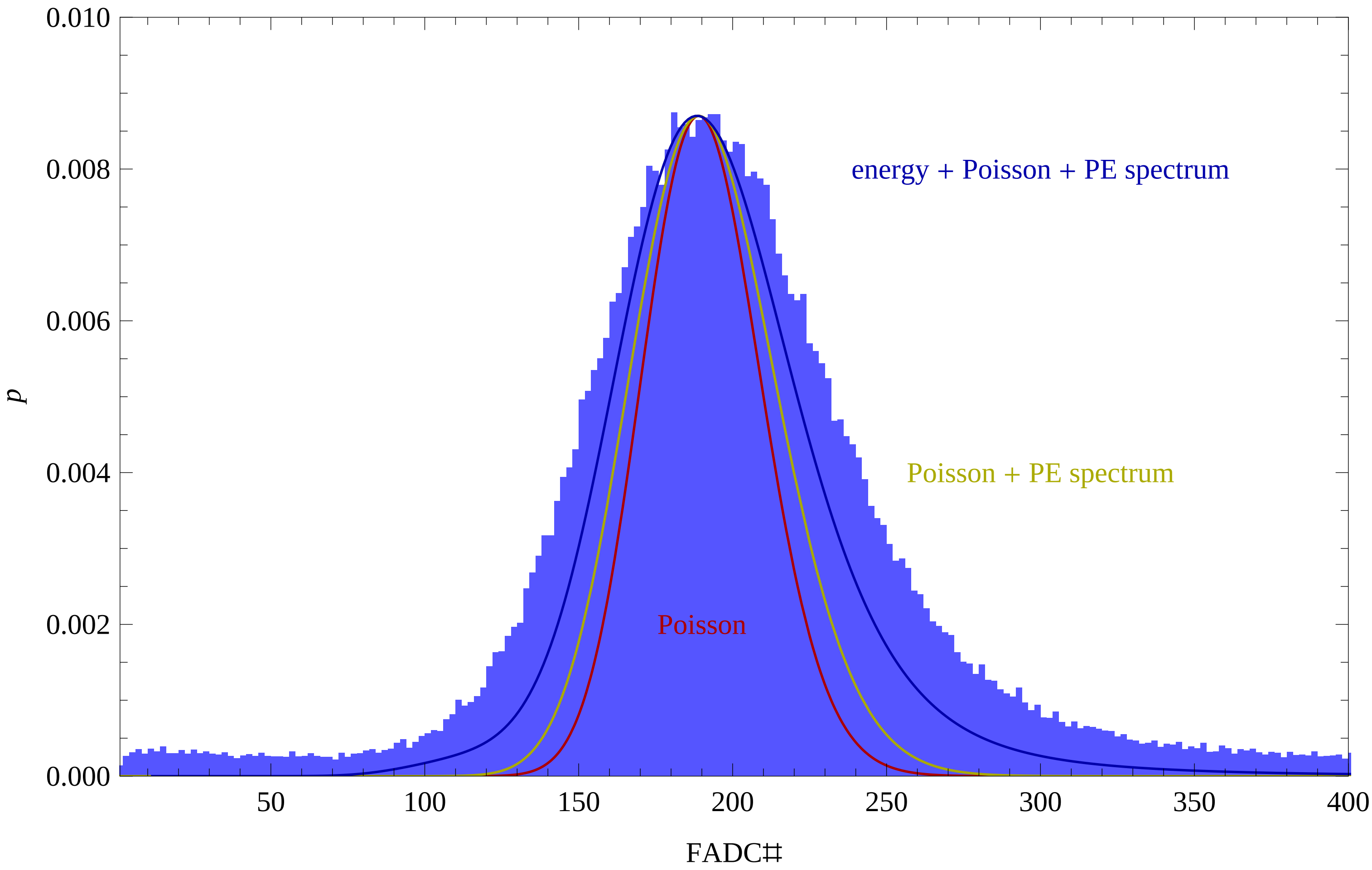
Atmospheric muon PE spectrum (CAPRICE98)



Atmospheric muon PE spectrum (CAPRICE98)



Partly explains the missing variance



Missing elements

- Direct light (refinement of $p_{\tau,t_d}(t)$)
- Time profiles $p(t_\mu)$, $p(t_\gamma)$ (connected to muon production)
- Energy distributions $p(E_\mu)$, $p(E_\gamma)$ (connected to muon production)
- Photon tank response $p(\phi_\gamma|E_\gamma)$
- Priors $p(N_\mu|\ldots)$, $p(N_\gamma|\ldots)$