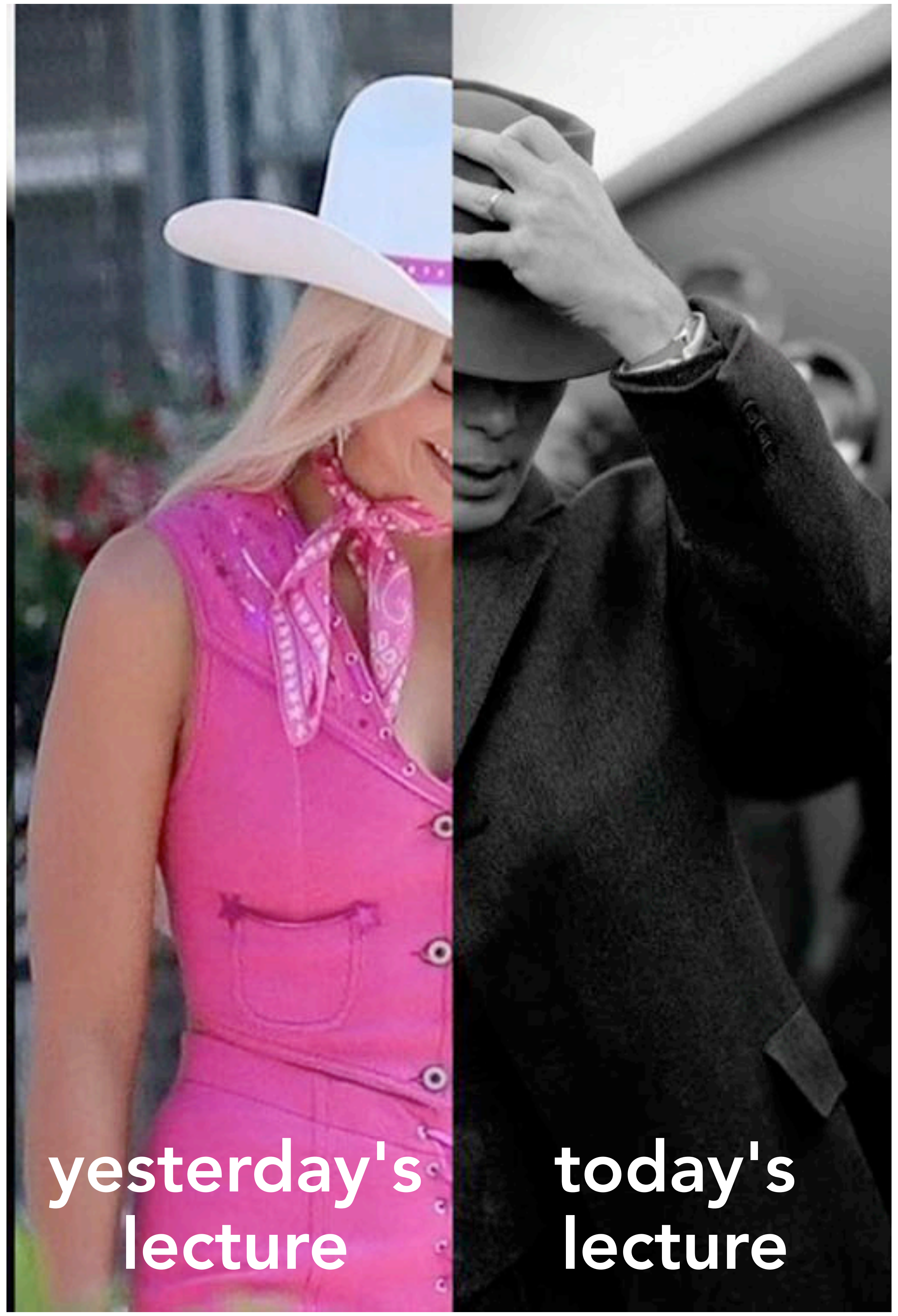
Let's talk about fits

Vitalii



9 November 2023

Setting your expectations right

- **This lecture is**
  - **Not a lecture.** I will be mostly live coding in front of you and dive into practical scenarios relevant for (LHCb-biased) flavour physics.
  - **Not a course in statistics.** To make me happy, ask questions about physics instead :)
    I did also not plan to talk about limit setting.
  - **Not a polished course.** It is the first time I try to teach this. But thank you for inviting! Let's keep in interactive so that I survive 7 hours of talking.
  - I will only cover RooFit (I am an old grumpy person). Many of you use alternative packages for fitting, but they (should) have similar internal logic, so what you learn today can be applied elsewhere.

- Some of you are probably familiar with 95% of what I say (= are better experts than me)
  - My hope is that everyone learns something (remaining 5%)
- Please ask questions at any time. (Or I will ask them to you.)
  - My most likely answer will be "I don't know". But remember – we have real experts among you.

Warm-up: Q1

- What is the fundamental difference between a typical plot from particle physics and a typical plot from plasma physics?

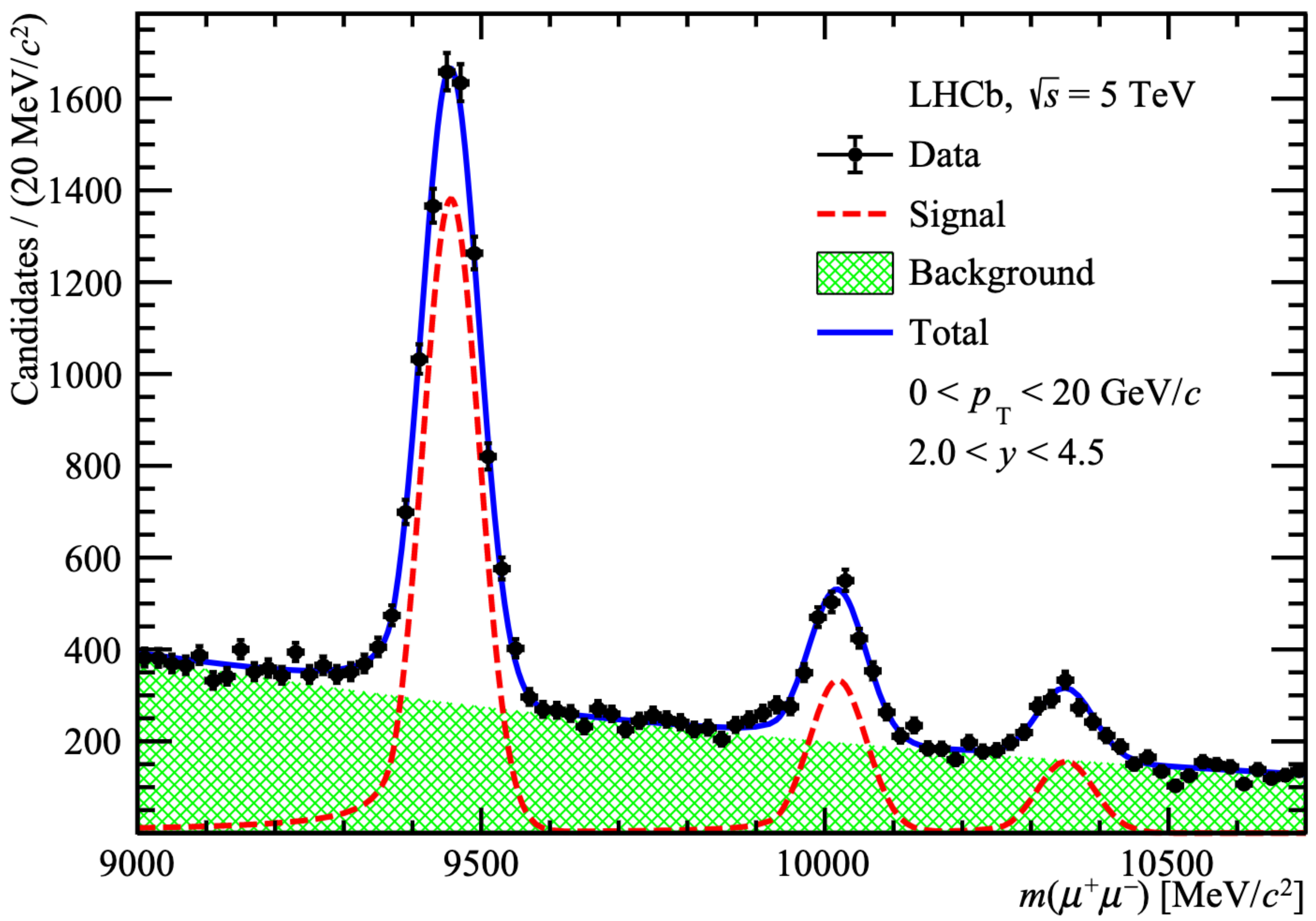


Figure 1: Invariant mass distribution of Υ candidates in the kinematic range $p_T \in [0, 20]$ GeV/c and $y \in [2.0, 4.5]$. The result of the fit with three Crystal Ball functions for the signal plus an exponential function for the background is also shown.

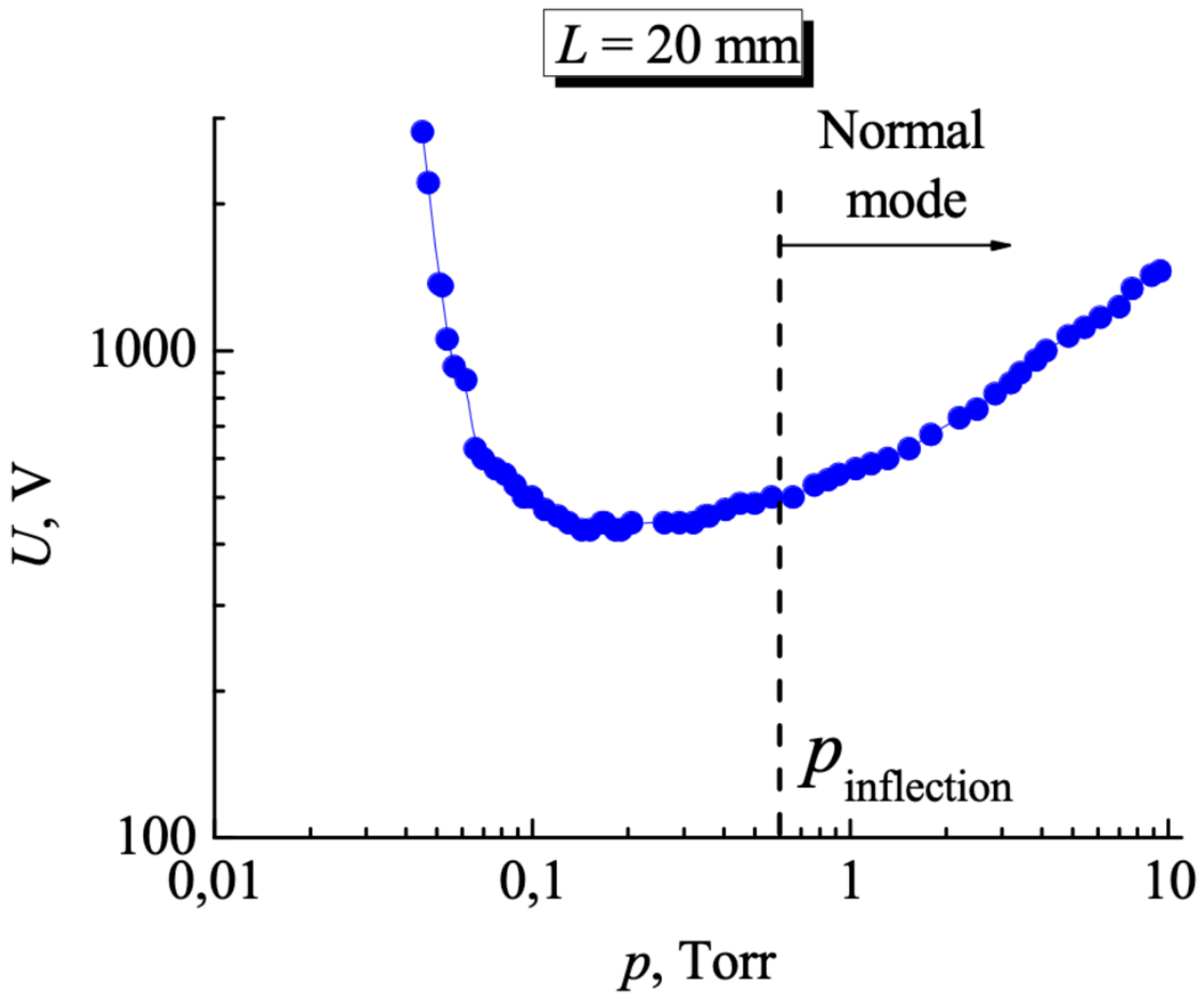
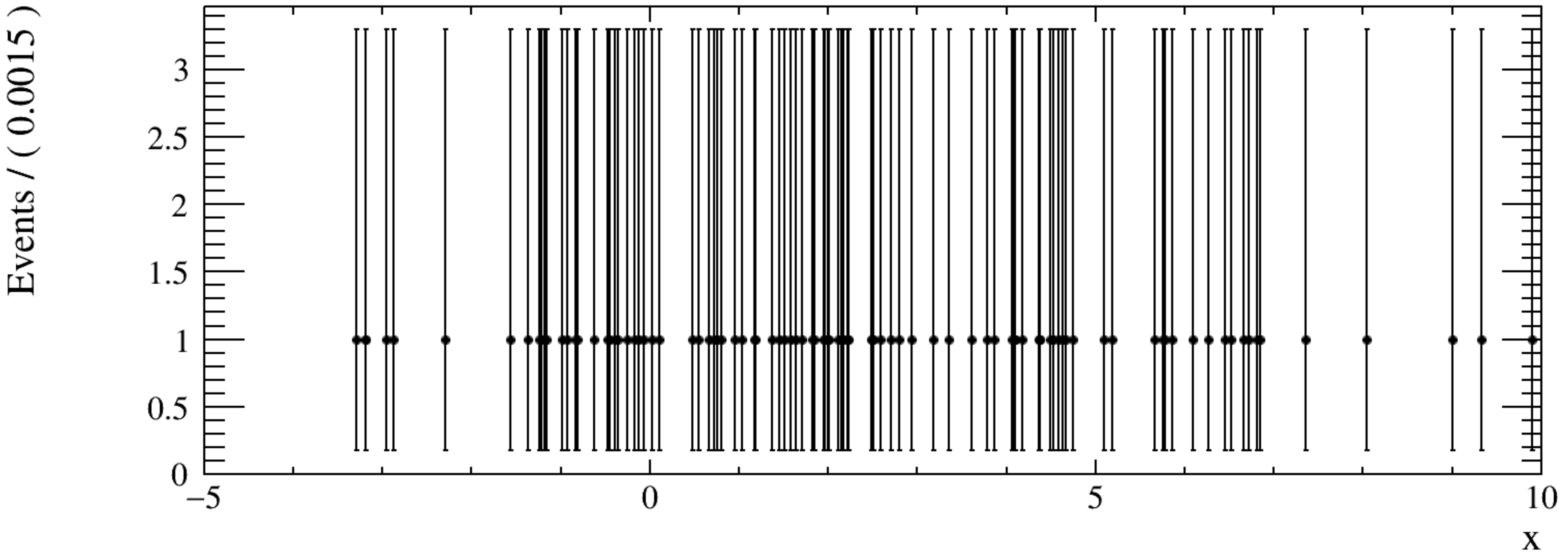
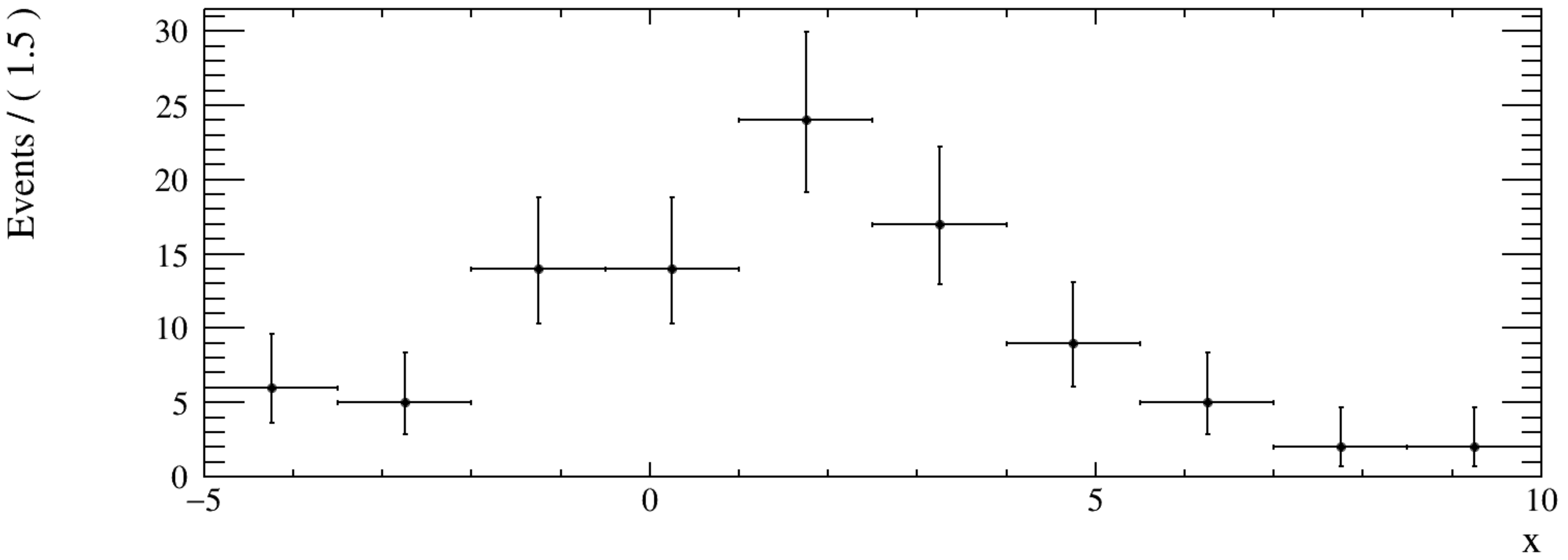


Fig. 4. DC breakdown curve in N_2O for the inter-electrode gap of 20 mm.

Warm-up: intermezzo

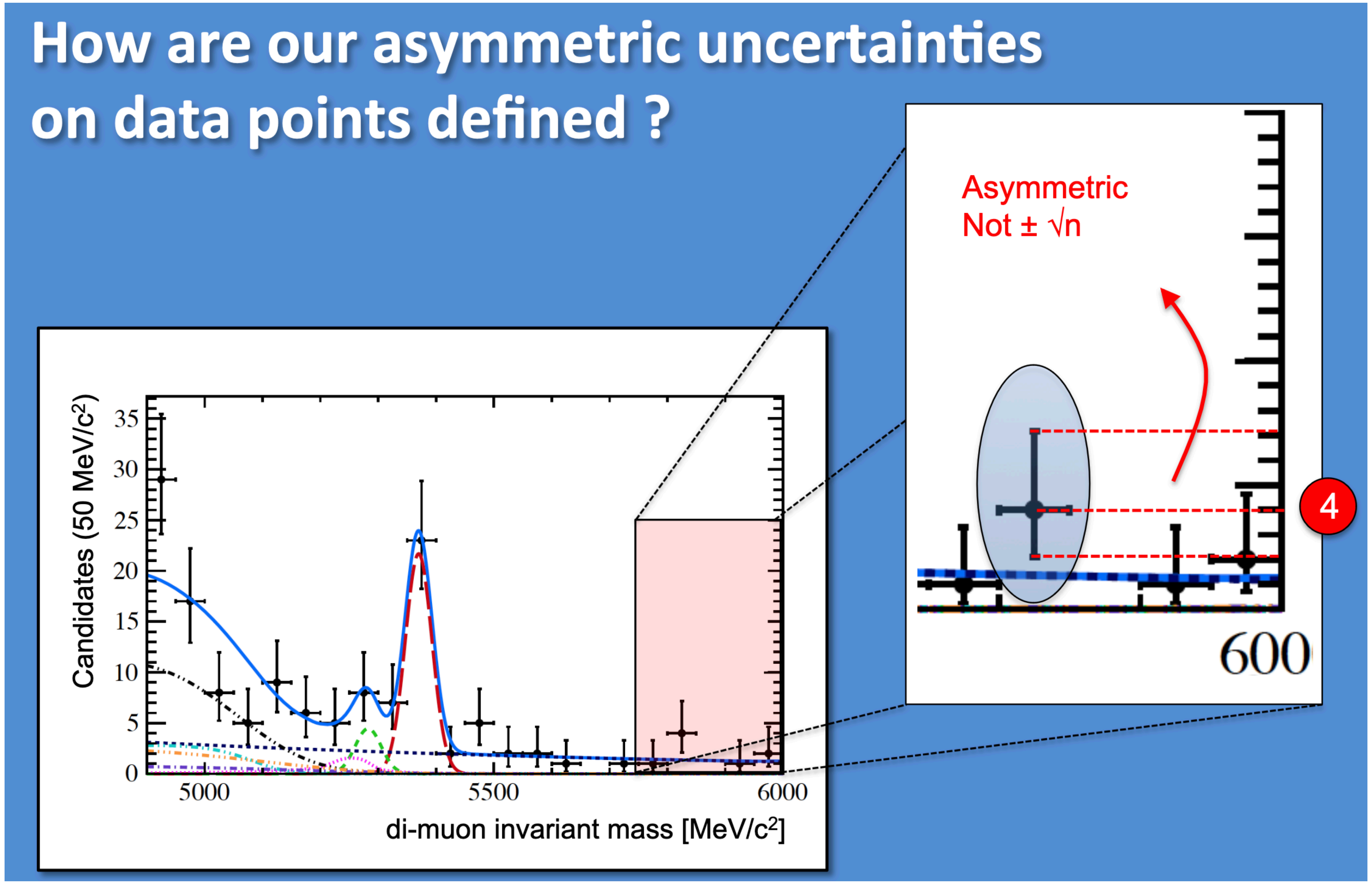
- Same histogram, different binnings





Warm-up: Q2

Based on https://www.nikhef.nl/~ivov/Statistics/PoissonError/2017_05_15_PoissonError_LHCb_IvovanVulpen.pdf



Warm-up: Q2

Based on https://www.nikhef.nl/~ivov/Statistics/PoissonError/2017_05_15_PoissonError_LHCb_IvovanVulpen.pdf
See also code in https://www.nikhef.nl/~ivov/Statistics/PoissonError/PoissonError.C (Ioption=1)

Poisson distribution: probability to observe n events when λ is expected:

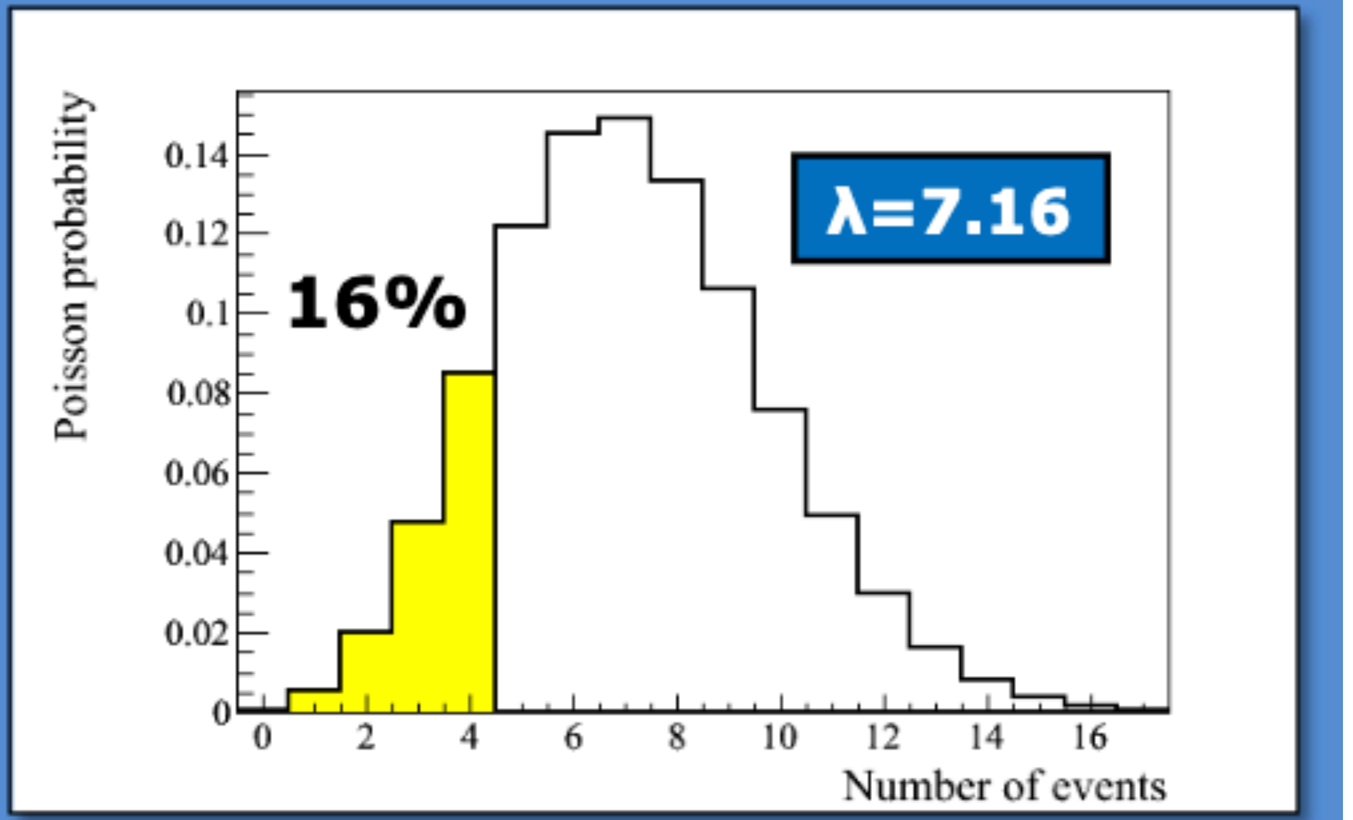
$$P(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

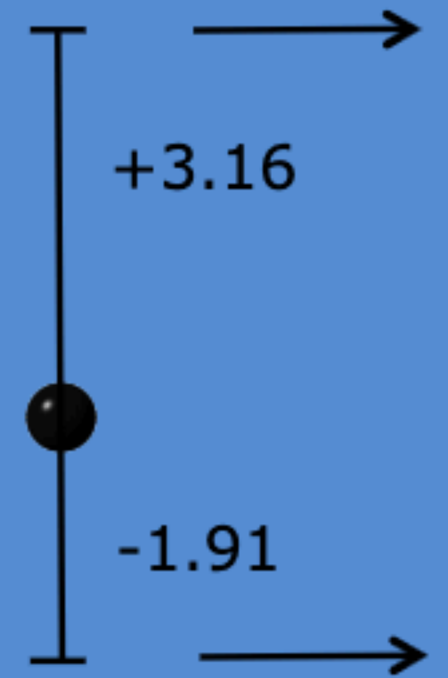
In the RooFit code:
look for RooHist / RooHistError

Find values of λ that are on border of being compatible with observed #events

If λ > 7.16 then probability to observe 4 events **(or less)** <16%

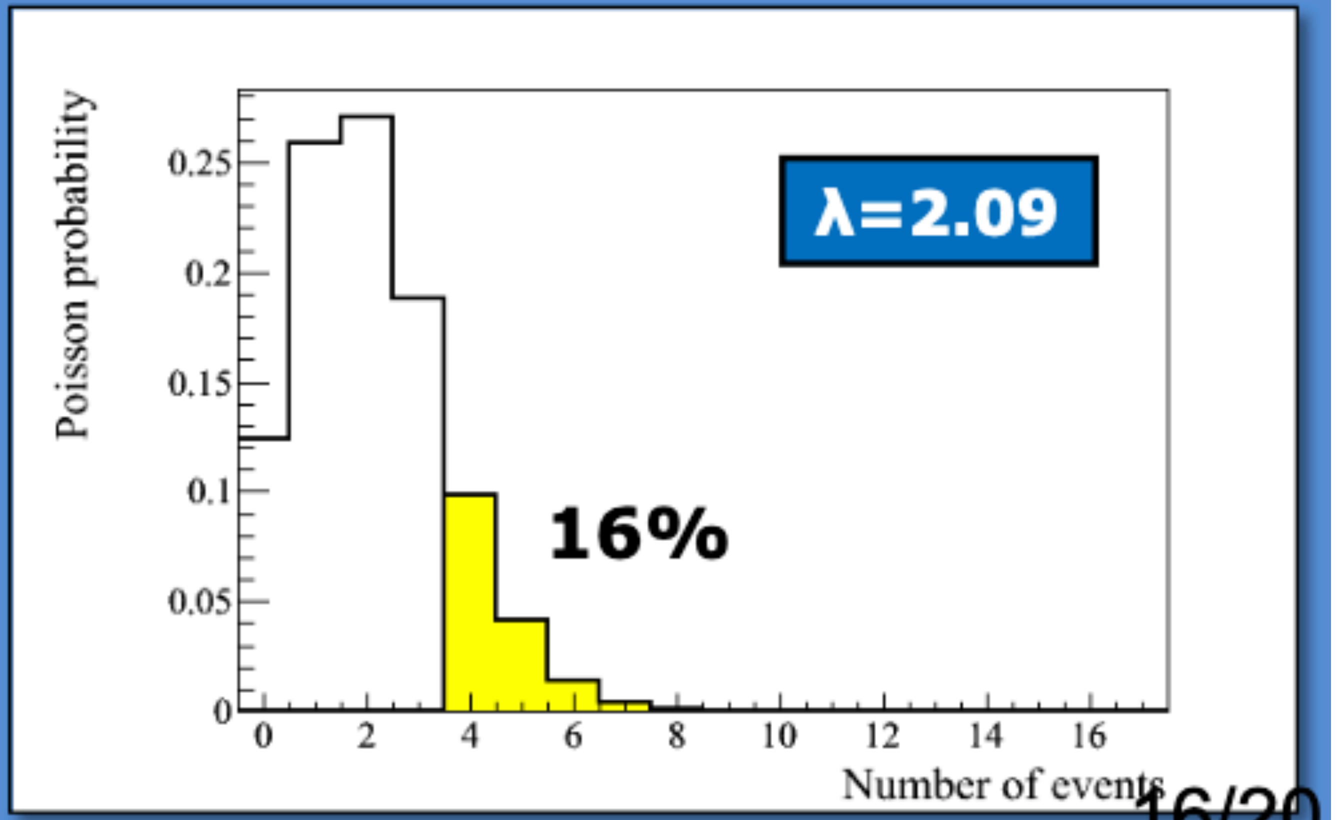
Note: also uses 'data you didn't observe', i.e. a bit like definition of significance





smallest λ (>n) for which $P(n \le n_{obs}|\lambda) \le 0.159$

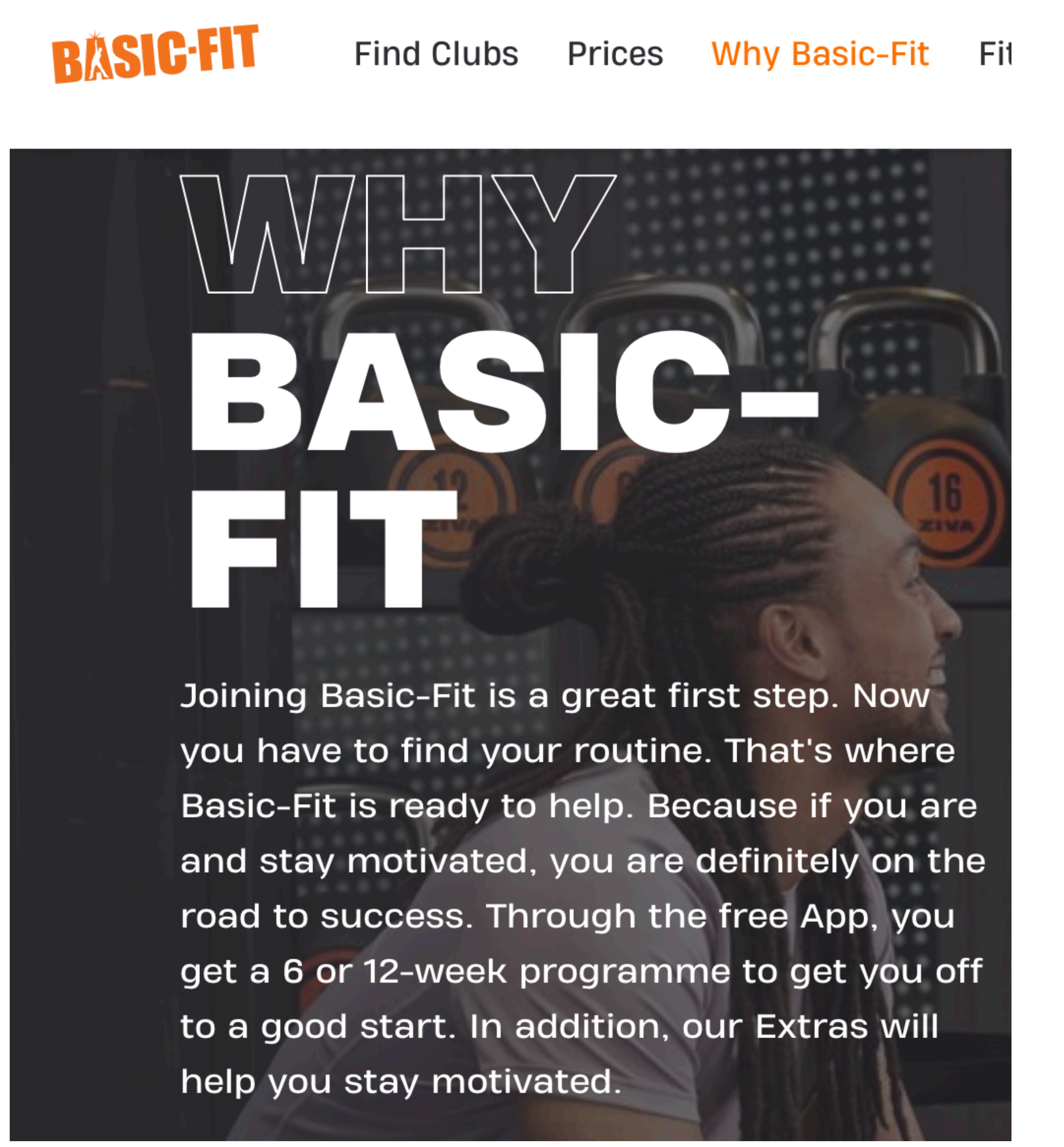
largest λ (<n) for which $P(n \ge n_{obs}|\lambda) \le 0.159$

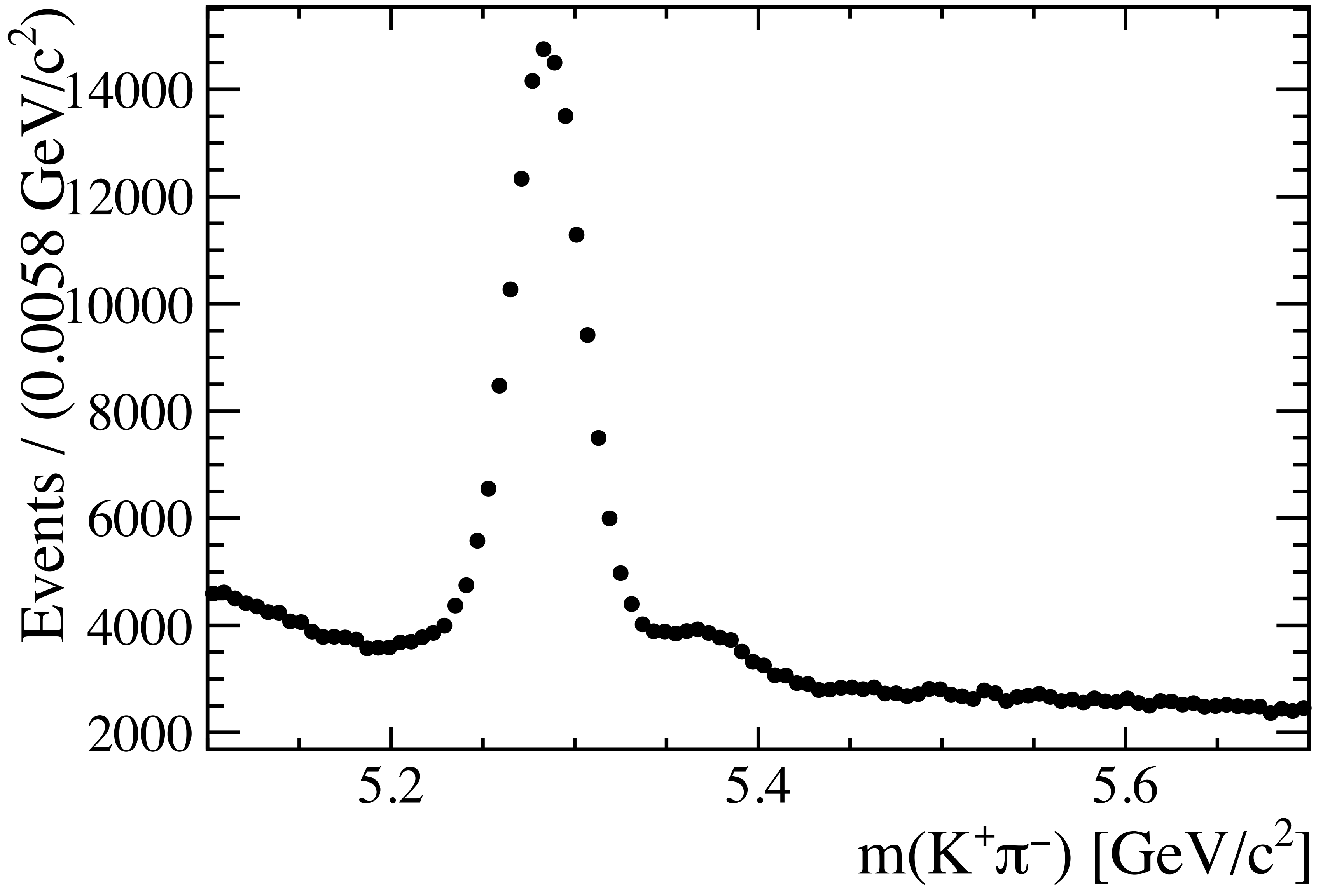


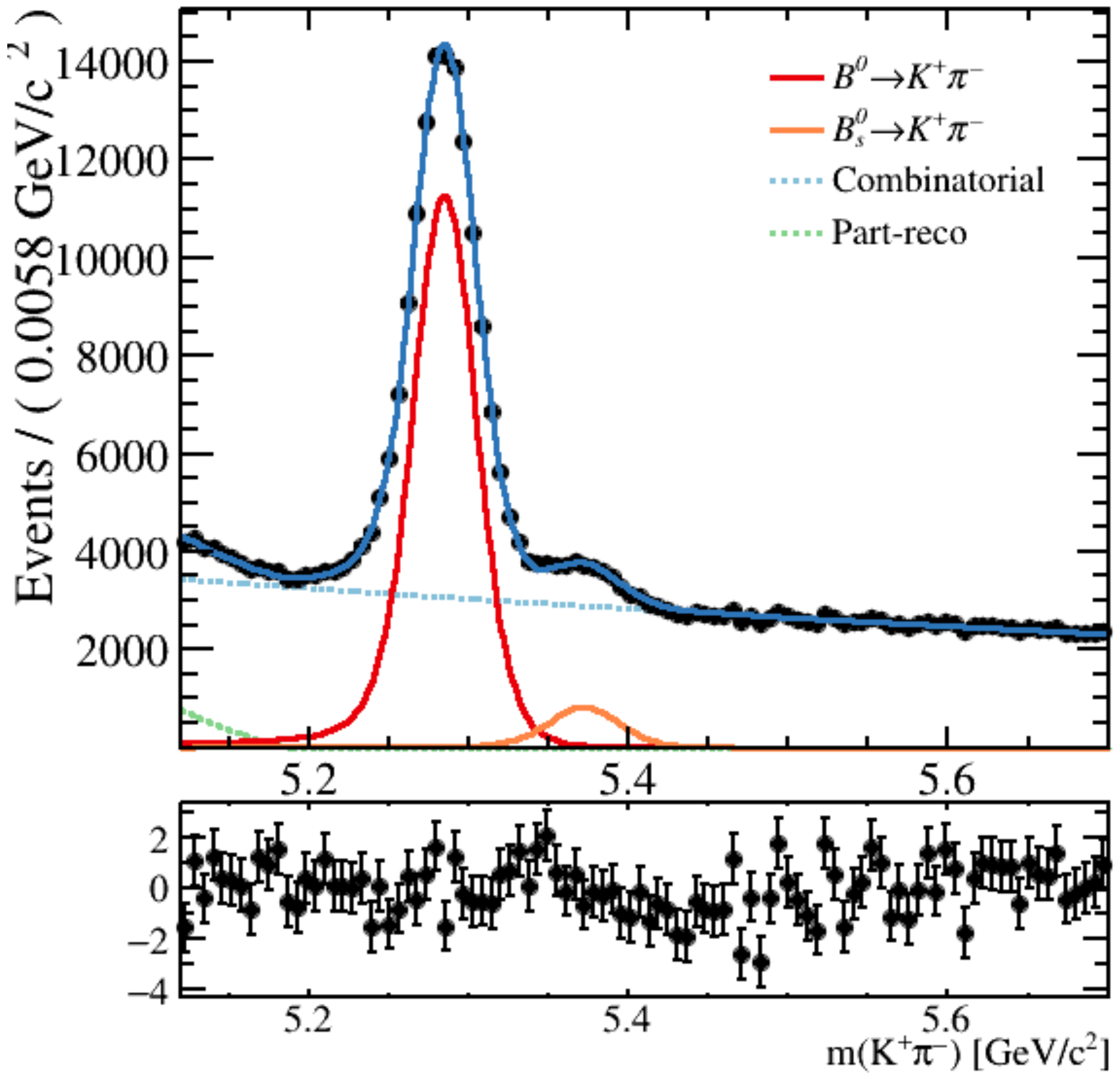
Note: 1-0.159-0.159 = 0.682
(1 sigma interval)

Fit basics: why?

- Choose a discrimination variable (signal/bkg separation)
  - Typically invariant mass, why?
- Create a model that describes the data, extract parameters of interest (e.g. signal yield)
- Estimate uncertainties on extracted parameters







Model = collection of Probability Density Functions (PDF): normalisation allows to interpret their fractions as yields. Challenge: calculate normalisation for nontrivial PDFs.

Fit basics: max likelihood

https://www.physi.uni-heidelberg.de/~reygers/lectures/2017/smipp/stat_methods_ss2017_05_parameter_estimation.pdf

Likelihood Function

Suppose we have a measurement of n independent values

$$\vec{x} = (x_1, x_2, \ldots, x_n)$$

drawn from the distribution

$$f(x; \vec{\theta}), \quad \vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$$

The joint pdf for the observed values $\vec{x}$ is given by:

$$L(\vec{x}; \vec{\theta}) = \prod_{i=1}^{n} f(x_i; \vec{\theta})$$ "likelihood function"

We consider $\vec{x}$ as constant. The *maximum likelihood estimate* (MLE) of the parameters are the values $\hat{\vec{\theta}}$ for which $L(\vec{x}; \vec{\theta})$ has a global maximum.

In other words, we ask the question:

"For which parameters do the observed data have the highest probability?"

Maximum Likelihood Example 1: Exponential Decay

Consider exponential pdf: $f(t; \tau) = \frac{1}{\tau} e^{-t/\tau}$

Independent measurements drawn from this distribution: $t_1, t_2, \ldots, t_n$

Likelihood function: $L(\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau}$

$L(\tau)$ is maximum when $\ln L(\tau)$ is maximum:

$$\ln L(\tau) = \sum_{i=1}^{n} \ln f(t_i; \tau) = \sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

Find maximum:

$$\frac{\partial \ln L(\tau)}{\partial \tau} = 0 \quad \rightsquigarrow \quad \sum_{i=1}^{n} \left(-\frac{1}{\tau} + \frac{t_i}{\tau^2} \right) = 0 \quad \rightsquigarrow \quad \hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

Fit basics: chi2

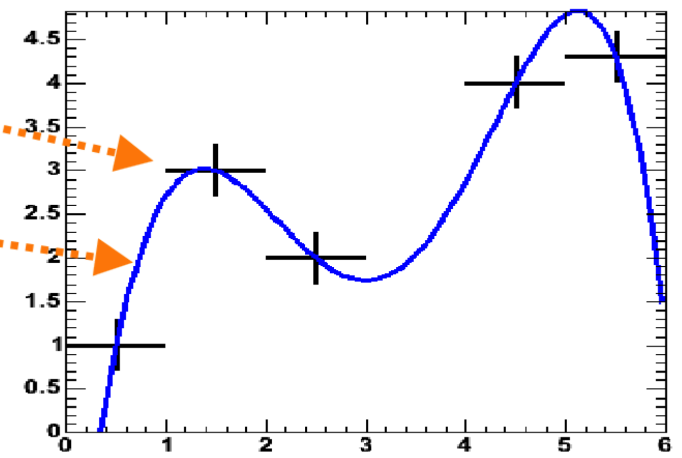
Works well in Gaussian regime
Less so in low-stat case

https://www.physik.hu-berlin.de/de/gk1504/block-courses/autumn-2010/program_and_talks/Verkerke_part3/

A well known estimator – the χ^2 fit

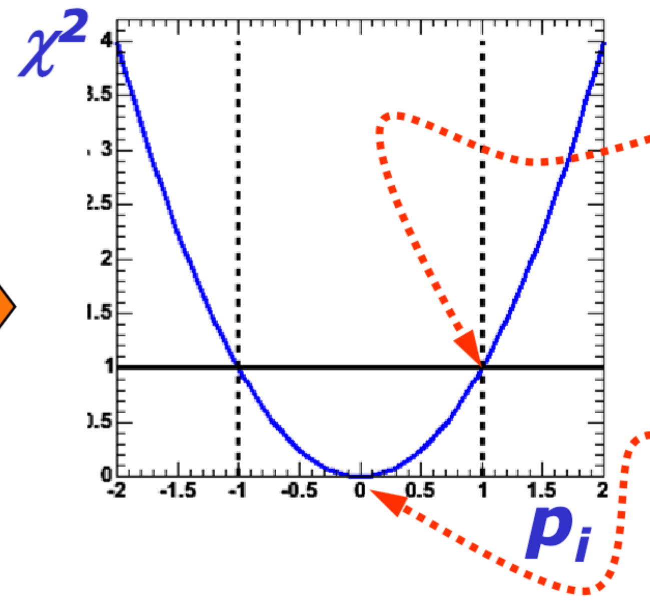
- Given a set of points $\{(\vec{x}_i, y_i, \sigma_i)\}$ and a function $f(\vec{x},\vec{p})$ define the χ^2

$$\chi^2(\vec{p}) = \sum_i \frac{(y_i - f(\vec{x};\vec{p}))^2}{\sigma_y^2}$$

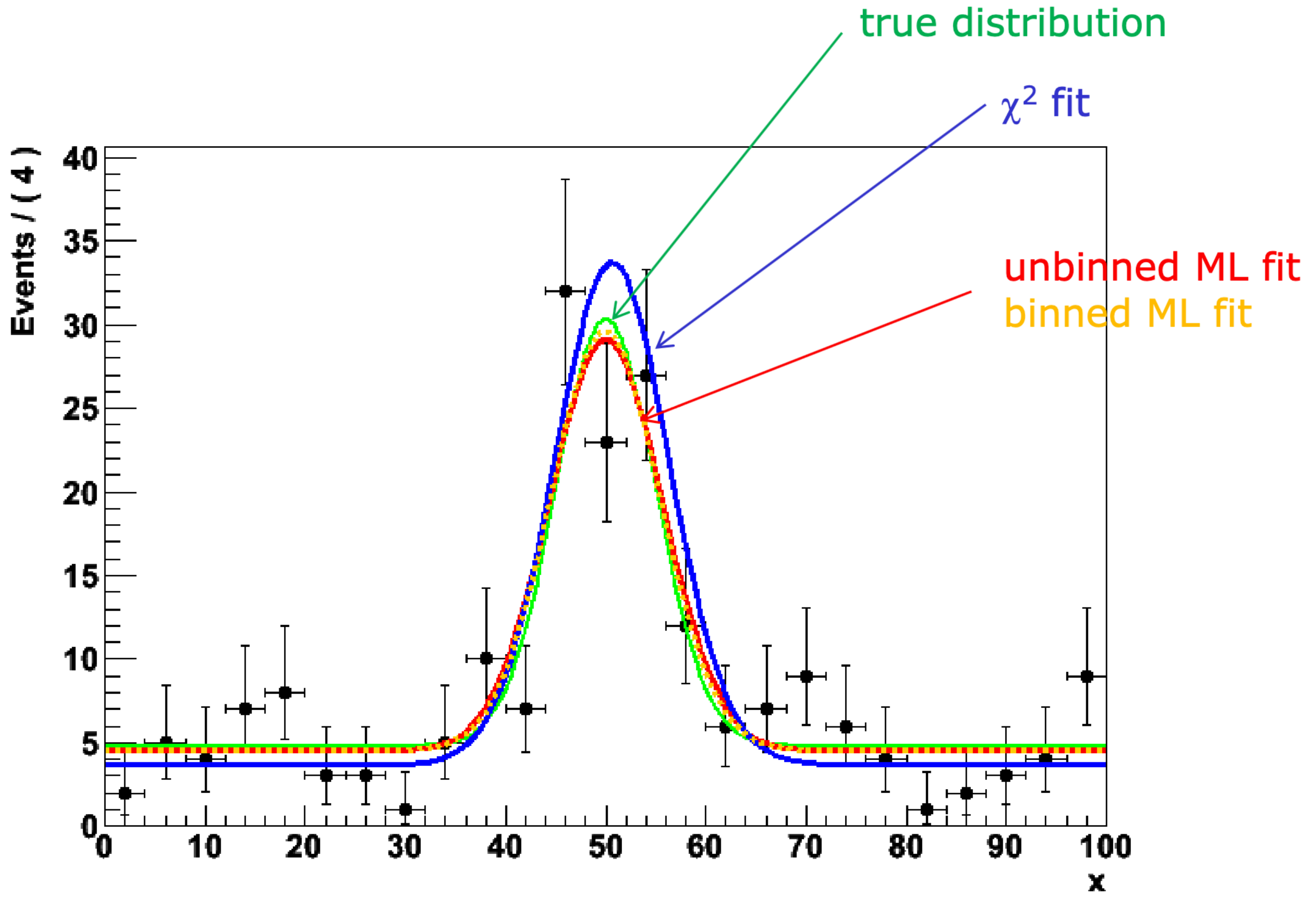


- Estimate parameters by minimizing the $\chi^2(p)$ with respect to all parameters p_i
  - *In practice, look for*

$$\frac{d\chi^2(p_i)}{dp_i} = 0$$



Example with many low statistics bins



Maximum Likelihood or χ^2 – What should you use?

- χ^2 fit is fastest, easiest
  - Works fine at high statistics
  - Gives absolute goodness-of-fit indication
  - Make (incorrect) Gaussian error assumption on low statistics bins
  - Has bias proportional to 1/N
  - Misses information with feature size < bin size
- Full Maximum Likelihood estimators most robust
  - No Gaussian assumption made at low statistics
  - No information lost due to binning
  - Gives best error of all methods (especially at low statistics)
  - No intrinsic goodness-of-fit measure, i.e. no way to tell if 'best' is actually 'pretty bad'
  - Has bias proportional to 1/N
  - Can be computationally expensive for large N
- Binned Maximum Likelihood in between $-\ln L(p)_{\text{binned}} = \sum_{\text{bins}} n_{\text{bin}} \ln F(\vec{x}_{\text{bin-center}};\vec{p})$
  - Much faster than full Maximum Likihood
  - Correct Poisson treatment of low statistics bins
  - Misses information with feature size < bin size
  - Has bias proportional to 1/N

Wouter Verkerke, UCSB

Minuit

A brief description of MINUIT functionality

https://web2.ba.infn.it/~pompili/teaching/data_analysis_lab/Verkerke-RooFit-part2.pdf

Watch for correlated parameters, find ways to avoid them

E.g.: instead of using two correlated parameters, take the 1st one and the ratio of the two.

Fix some of them (iteratively) or constrain allowed ranges / starting values

- MIGRAD
  - Find function minimum. Calculates function gradient, follow to (local) minimum, recalculate gradient, iterate until minimum found
    - To see what MIGRAD does, it is very instructive to do RooMinuit::setVerbose(1). It will print a line for each step through parameter space
  - Number of function calls required depends greatly on number of floating parameters, distance from function minimum and shape of function

Beware of local minima: starting values might matter

- HESSE
  - Calculation of error matrix from 2nd derivatives at minimum
  - Gives symmetric error. Valid in assumption that likelihood is (locally parabolic)

$$\hat{\sigma}(p)^2 = \hat{V}(p) = \left(\frac{d^2 \ln L}{d^2 p} \right)^{-1}$$

  - Requires roughly N^2 likelihood evaluations (with N = number of floating parameters)

Good approximation for large number of events

Wouter Verkerke, NIKHEF

- MINOS
  - Calculate errors by explicit finding points (or contour for >1D) where Δ-log(L)=0.5
  - Reported errors can be asymmetric
  - Can be very expensive in with large number of floating parameters

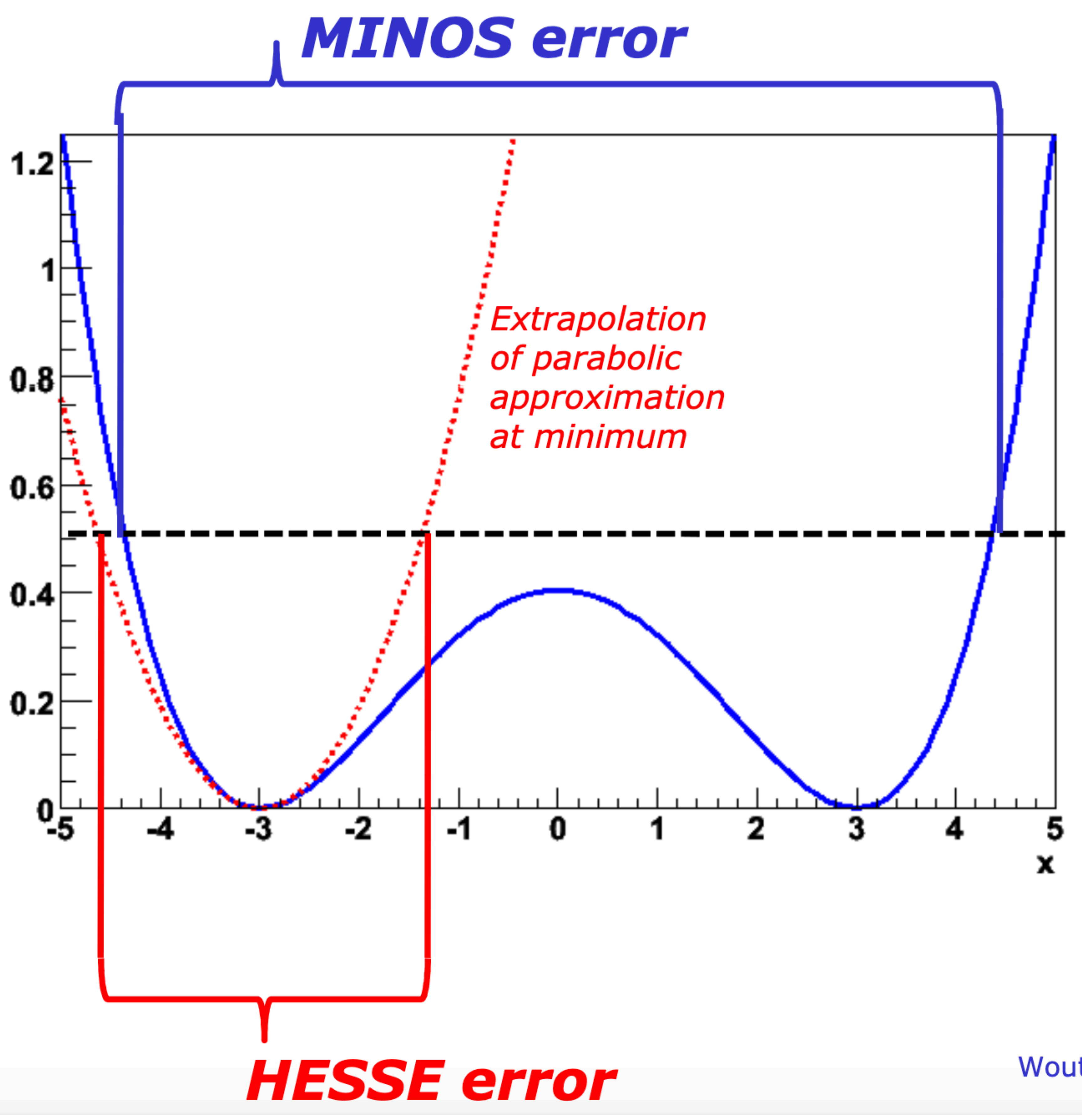
Useful in low-stats case

Minuit

https://web2.ba.infn.it/~pompili/teaching/data_analysis_lab/Verkerke-RooFit-part2.pdf

Illustration of difference between HESSE and MINOS errors

- 'Pathological' example likelihood with multiple minima and non-parabolic behavior



Wouter Verkerke, NIKHEF

Mass peaks: Q1

- Why are our mass peaks asymmetric?

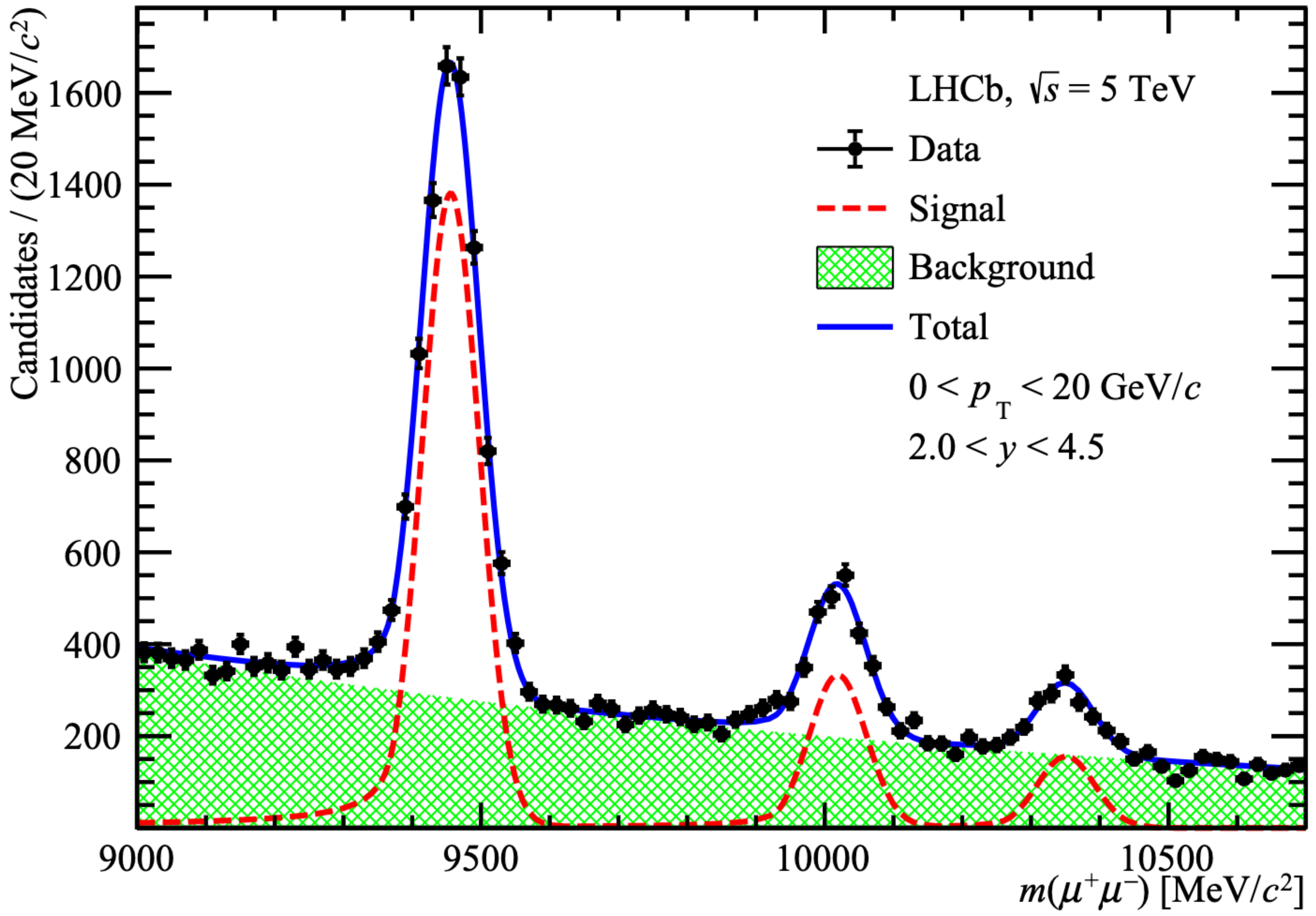


Figure 1: Invariant mass distribution of Υ candidates in the kinematic range $p_T \in [0, 20]$ GeV/c and $y \in [2.0, 4.5]$. The result of the fit with three Crystal Ball functions for the signal plus an exponential function for the background is also shown.

Mass peaks: Q1

- Why are our mass peaks asymmetric?

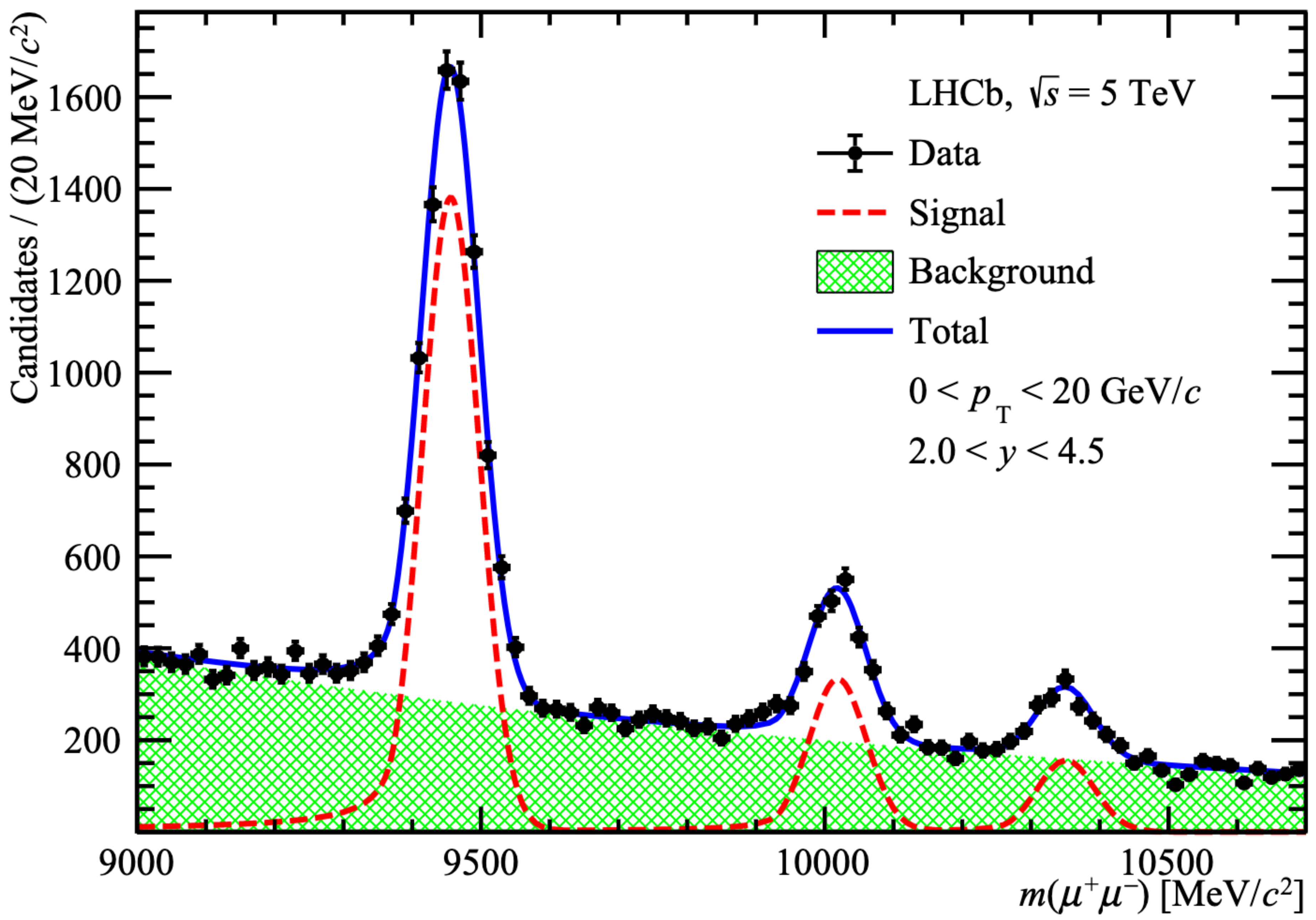
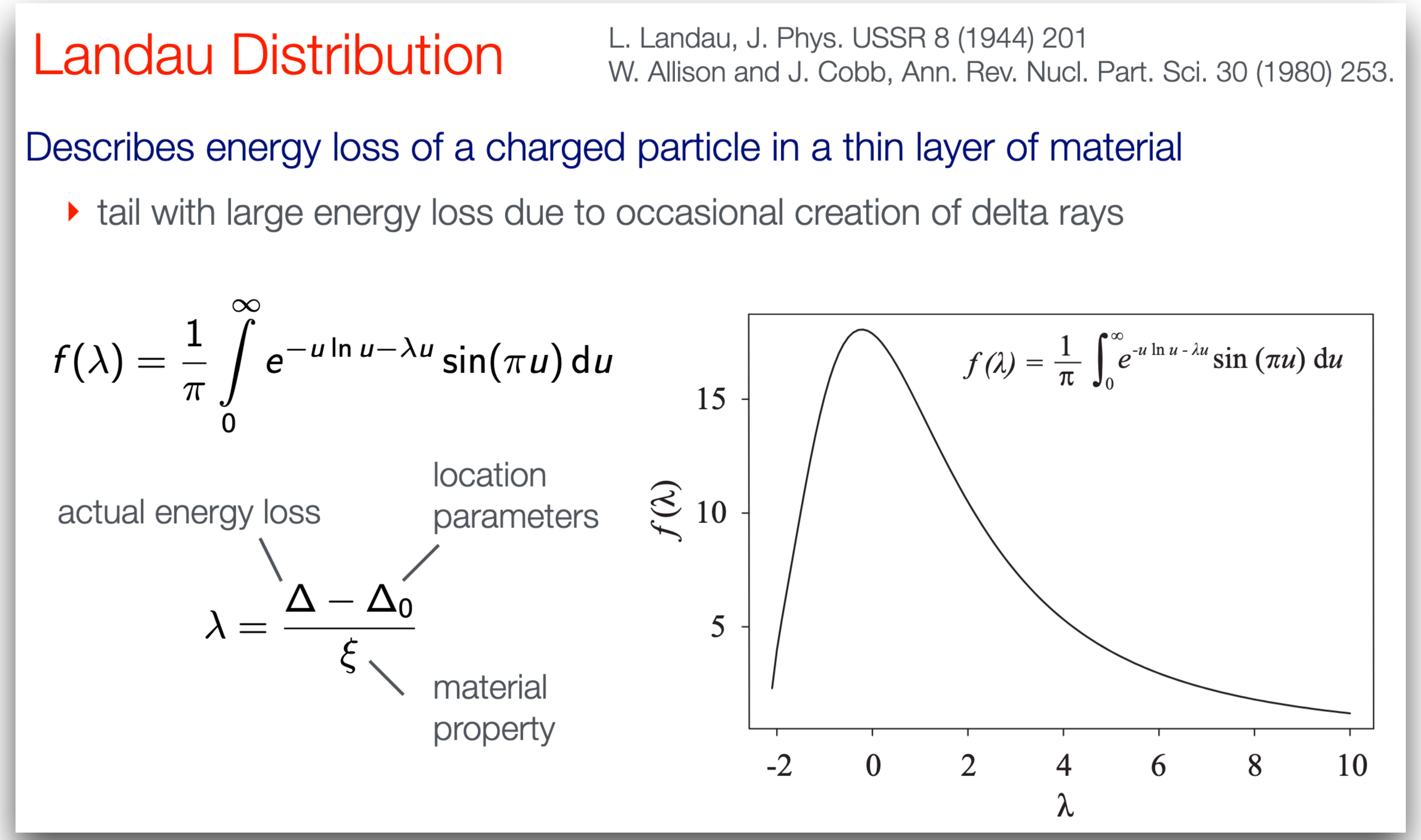
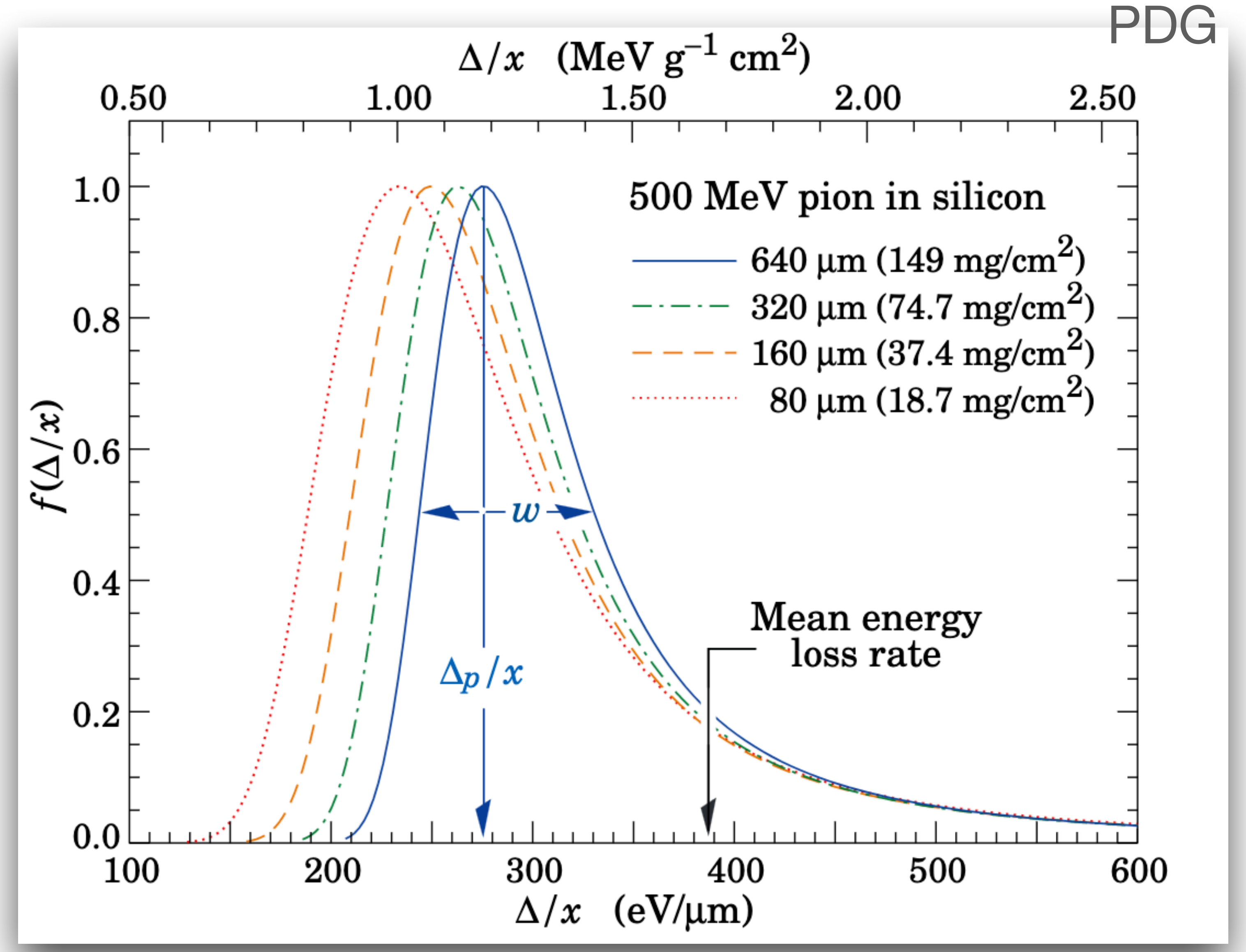


Figure 1: Invariant mass distribution of Υ candidates in the kinematic range $p_T \in [0, 20]$ GeV/c and $y \in [2.0, 4.5]$. The result of the fit with three Crystal Ball functions for the signal plus an exponential function for the background is also shown.

https://www.physi.uni-heidelberg.de/~reygers/lectures/2017/smipp/stat_methods_ss2017_02_distributions.pdf



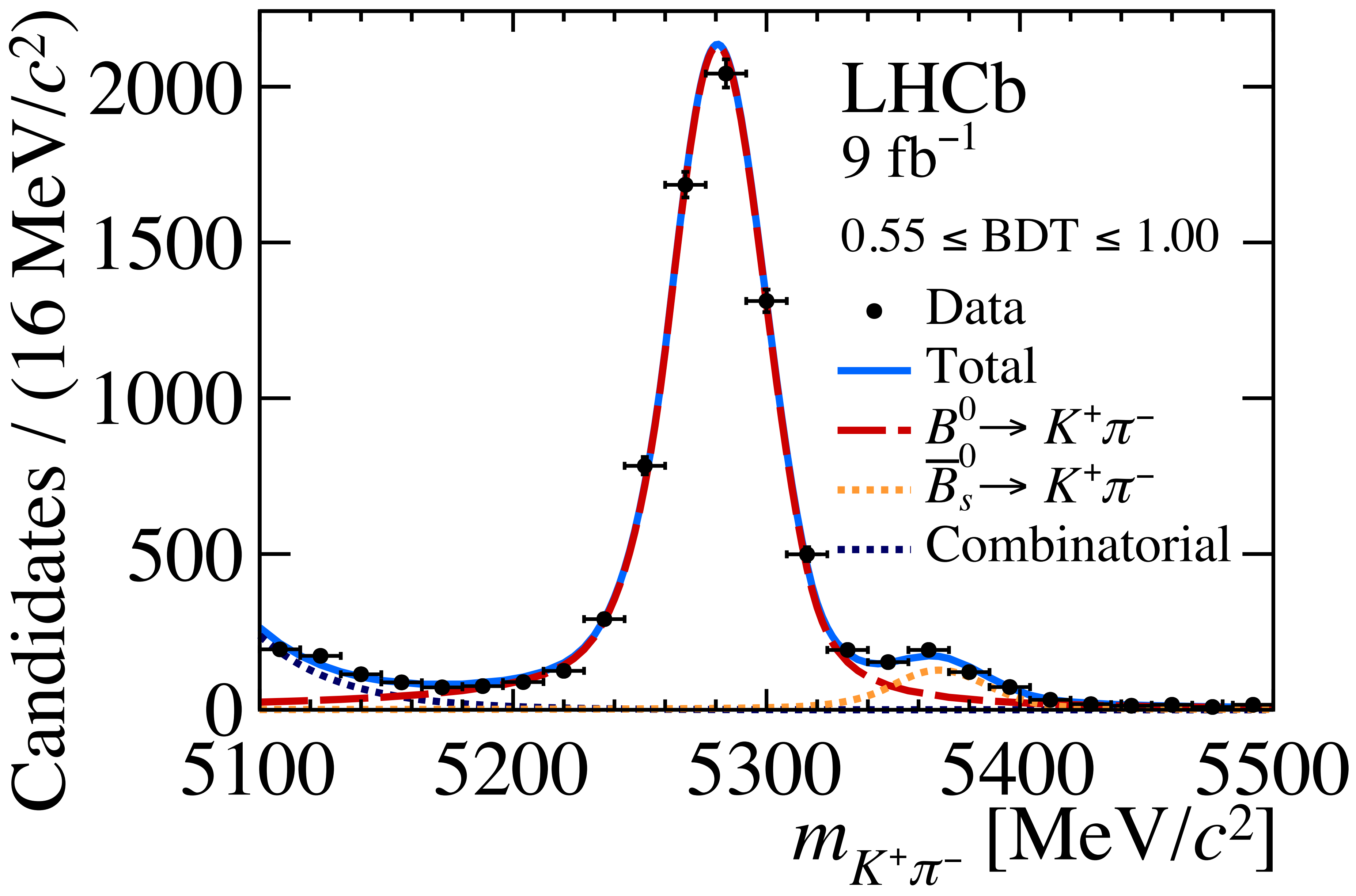
PDG



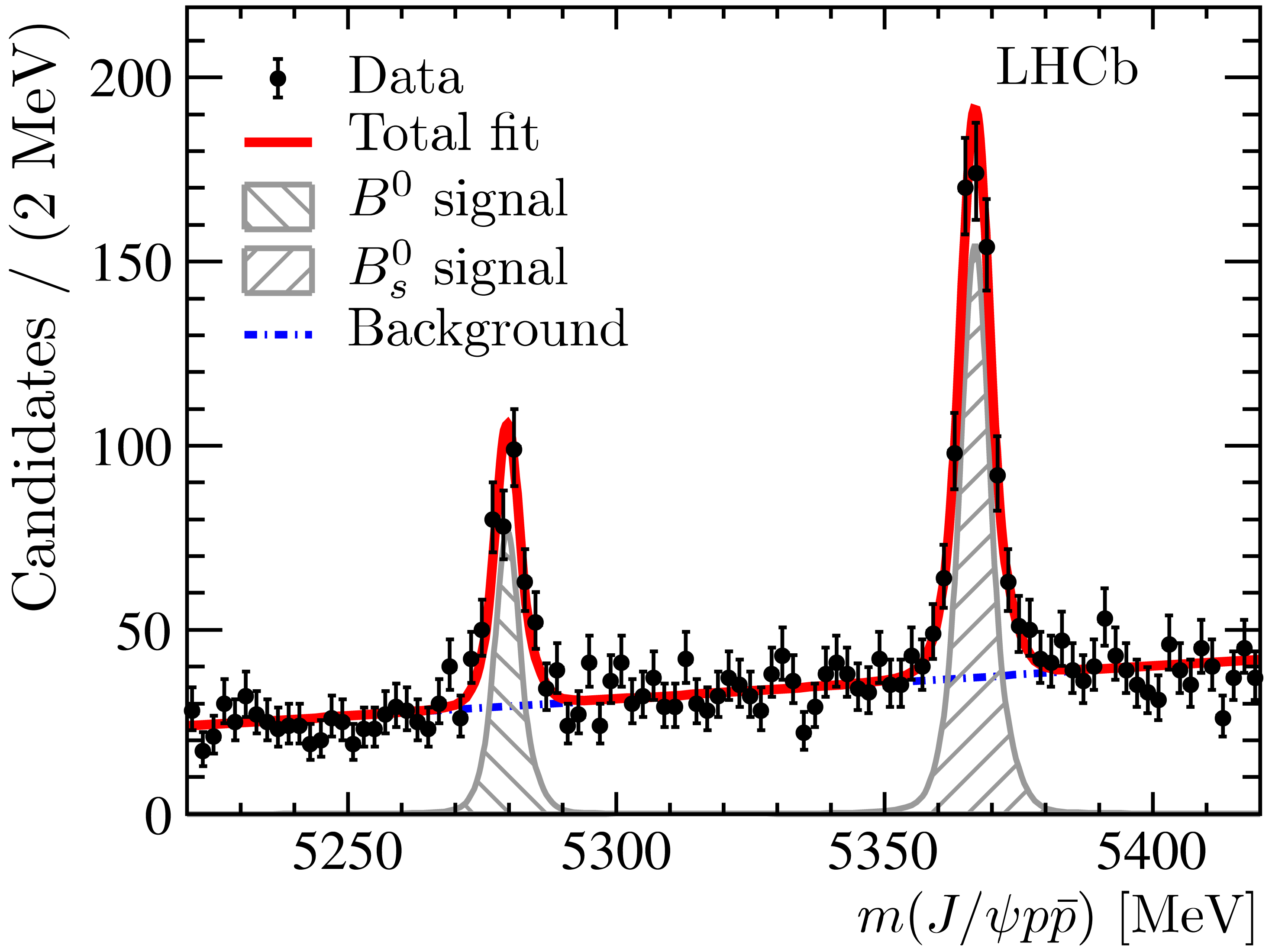
Add multiple scattering (not Gaussian either), hit resolution...

Mass peaks: Q2

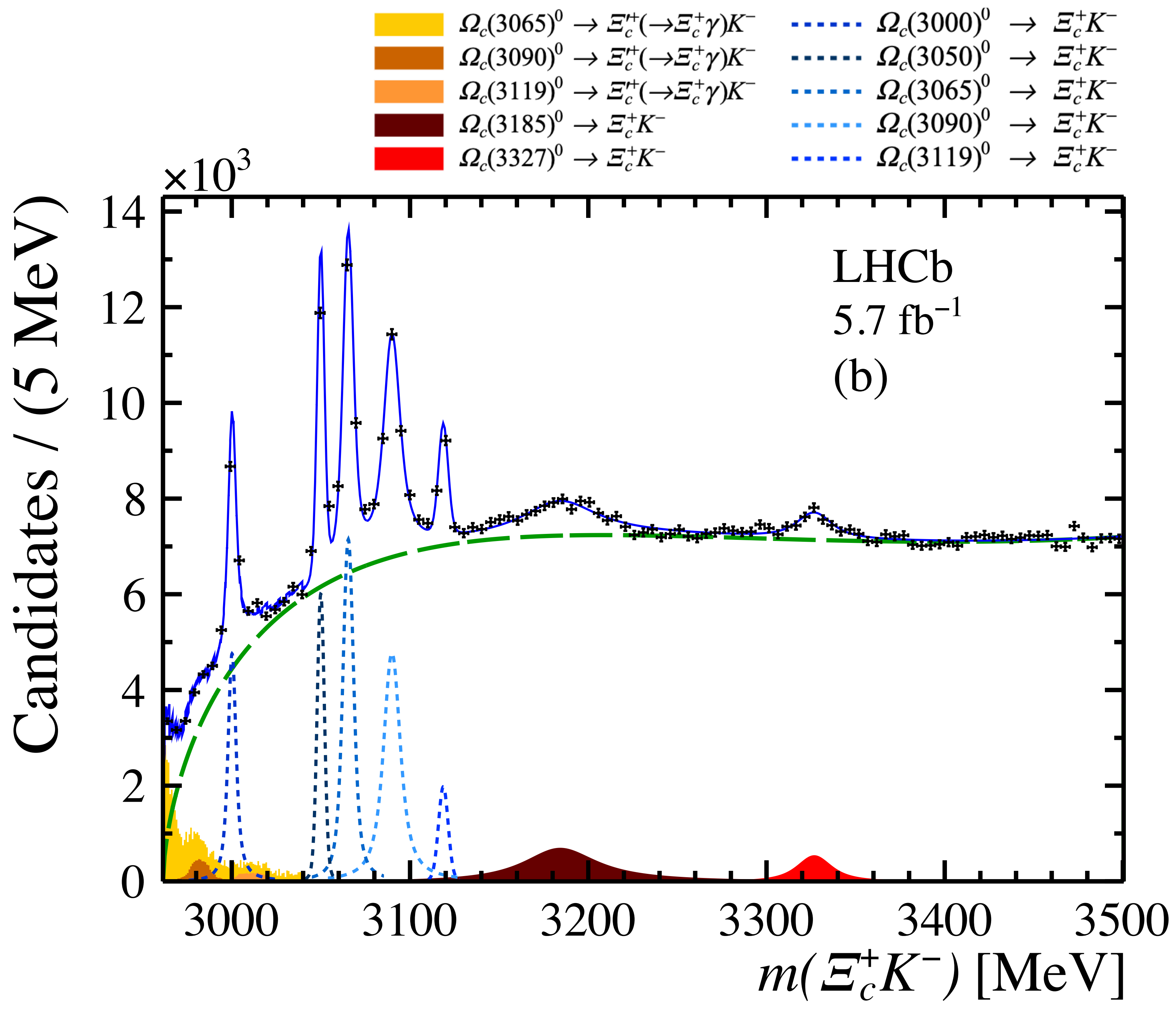
- Why do these mass peaks have different widths?



$B^0 \to K^+\pi^-$ $B_s^0 \to K^+\pi^-$



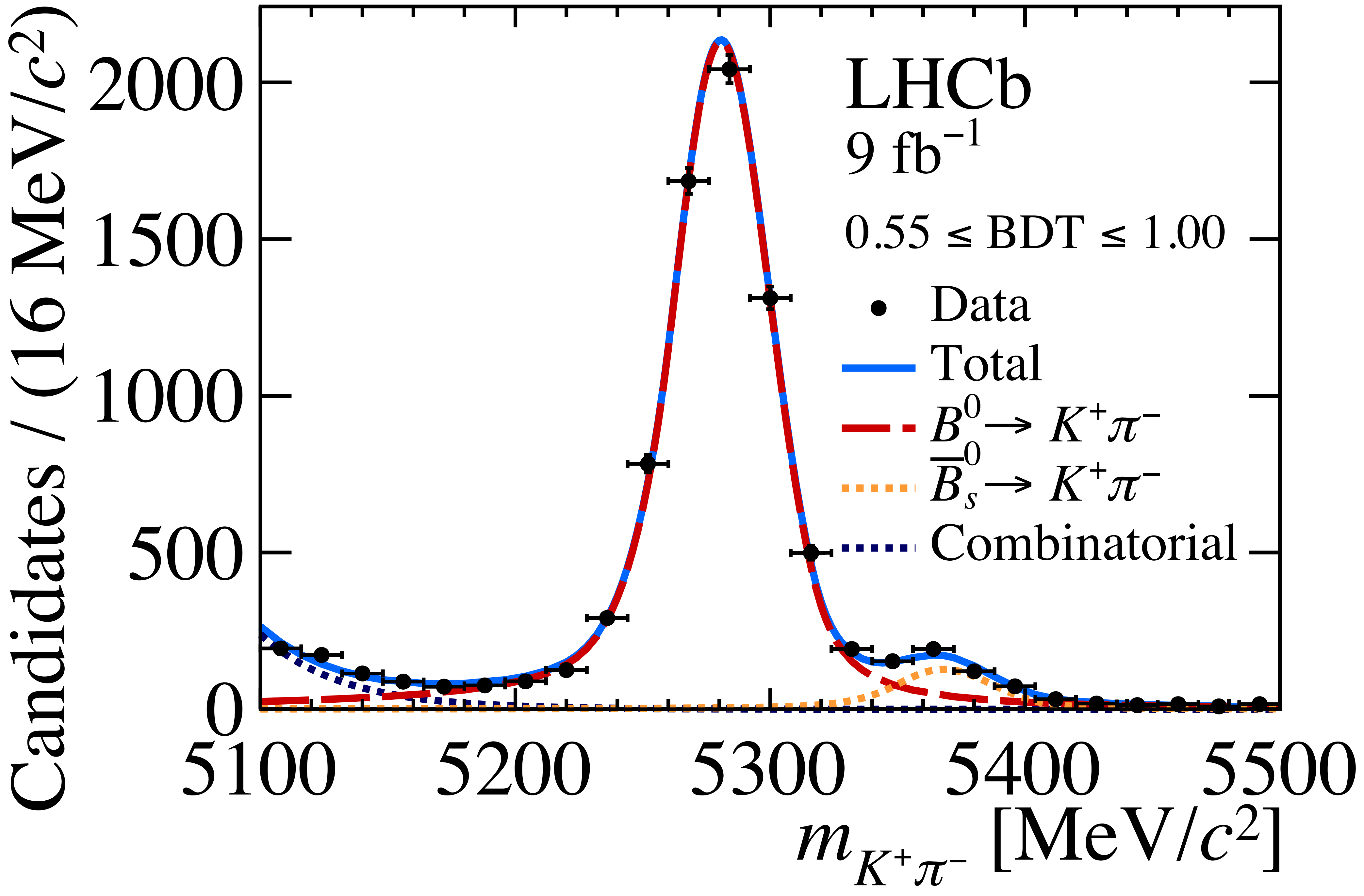
$B^0 \to J/\psi p\bar{p}$ $B_s^0 \to J/\psi p\bar{p}$

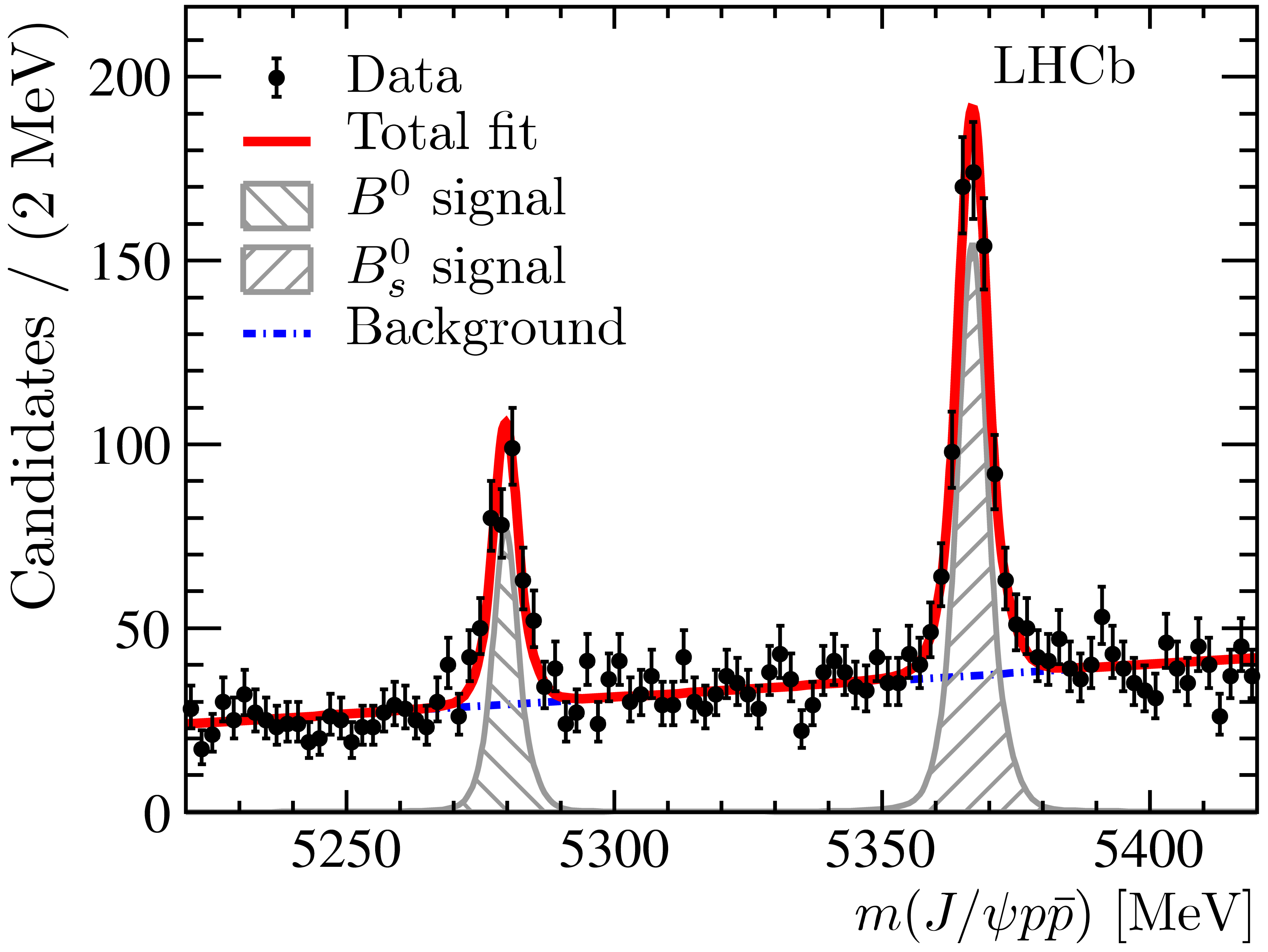


various $\Omega_c^* \to \Xi_c^+ K^-$

Mass peaks: Q2b

- Why do these B-meson mass peaks have different resolutions?
- Which final state is better for mass measurements?





Hint: in tracking, $p(\mathrm{GeV}) \sim 0.3B(\mathrm{T})R(\mathrm{m})$

$$\frac{\delta p}{p} \sim \frac{\delta R}{R} \sim \frac{R}{L^2}\delta y \sim \frac{p}{BL^2}\delta y$$

(exact details differ slightly between "forward" and "cylindrical" detectors)

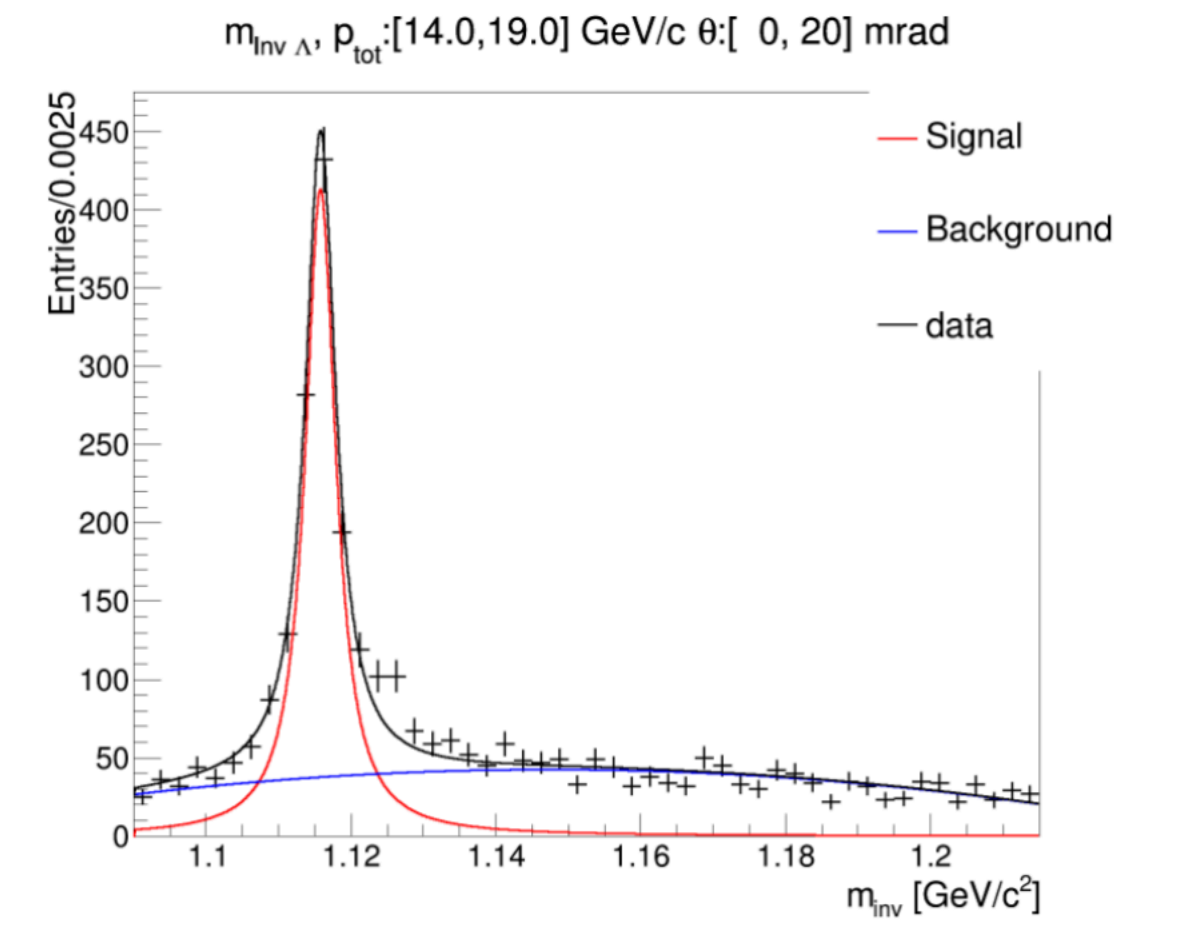
Second effect: multiple scattering (~const vs momentum, so relatively more important at low momentum)

Mass peaks: Q3

- What is wrong with this procedure?

Measurements of K^0_S, Λ and $\bar{\Lambda}$ production in 120 GeV/c p + C interactions

The NA61/SHINE Collaboration



The signal model used is a Lorentzian:

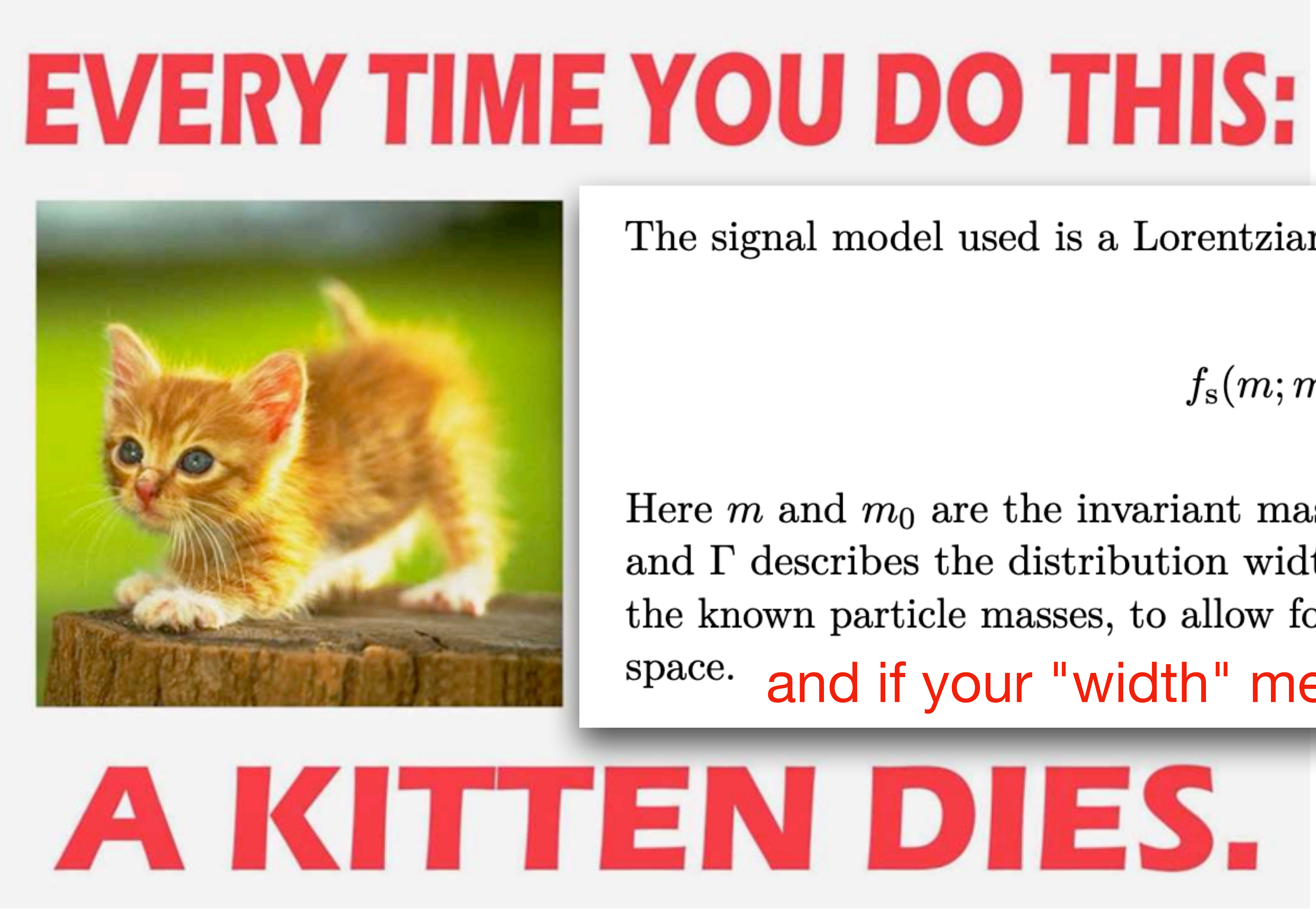
$$f_s(m; m_0, \Gamma) = \frac{1}{\pi\Gamma} \frac{\Gamma^2}{(m - m_0)^2 + \Gamma^2}. \quad (4)$$

Here m and m_0 are the invariant mass and offset from the accepted best-fit value, respectively, and Γ describes the distribution width. Central invariant masses were allowed to deviate from the known particle masses, to allow for momentum mis-reconstruction in certain regions of phase space.

Mass peaks: Q3

- What is wrong with this procedure?

Measurements of K^0_S, Λ and $\bar{\Lambda}$ production in 120 GeV/c p + C interactions



The signal model used is a Lorentzian:

$$f_s(m; m_0, \Gamma) = \frac{1}{\pi\Gamma} \frac{\Gamma^2}{(m - m_0)^2 + \Gamma^2}. \tag{4}$$

Here m and m_0 are the invariant mass and offset from the accepted best-fit value, respectively, and Γ describes the distribution width. Central invariant masses were allowed to deviate from the known particle masses, to allow for momentum mis-reconstruction in certain regions of phase space.

and if your "width" means "resolution"


FE
ENCE
OK CLASS, I KNOW WE'VE COVERED A **NEW** TOPIC TODAY...
ARE THERE ANY QUESTIONS?
©AHEARN
zoodraws