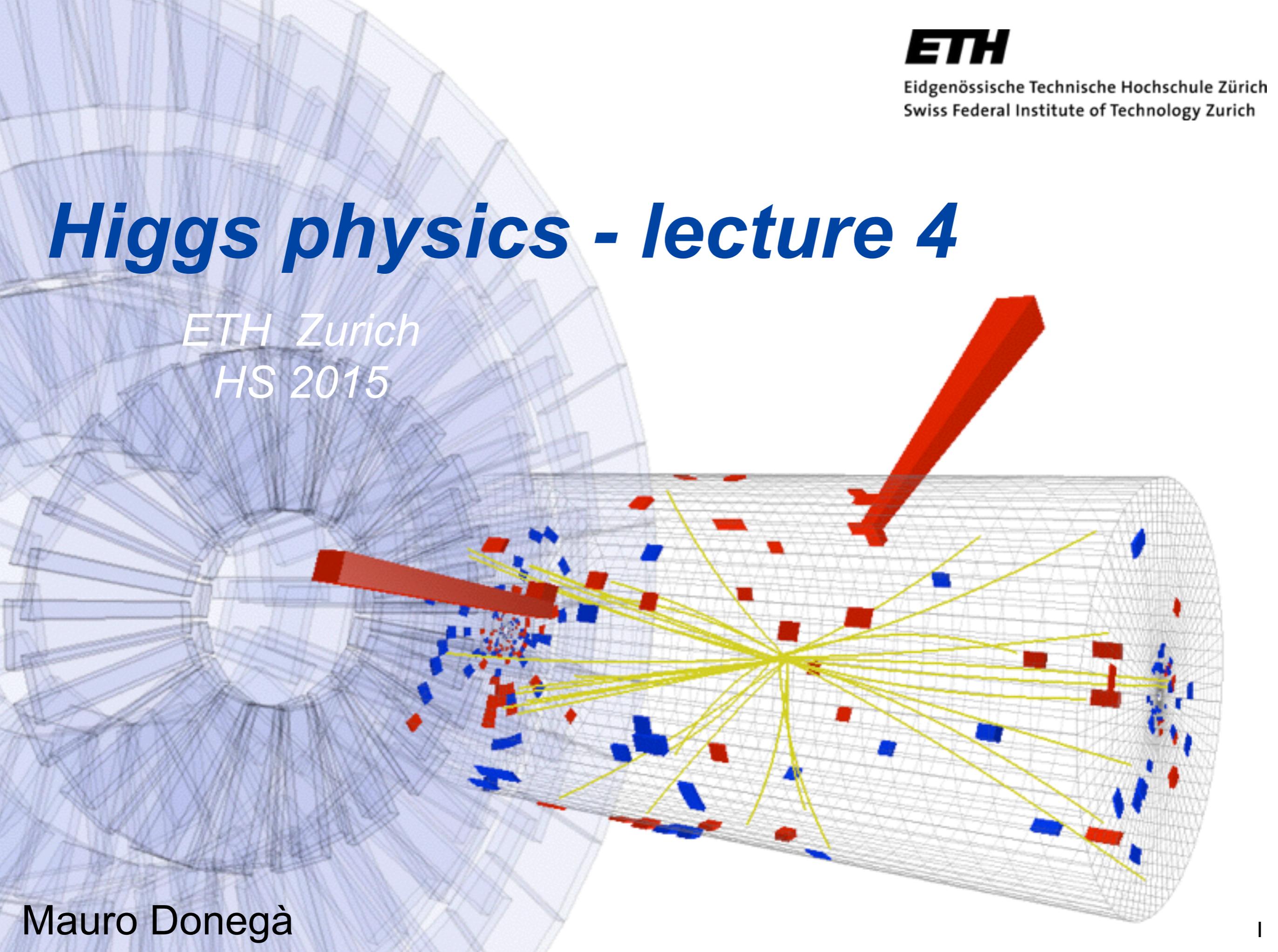


Higgs physics - lecture 4

*ETH Zurich
HS 2015*



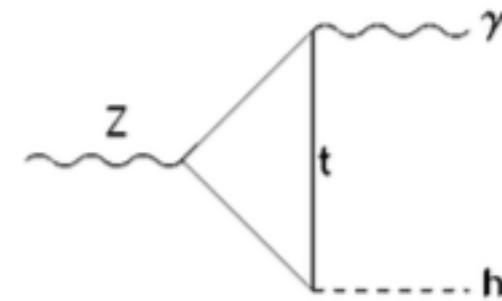
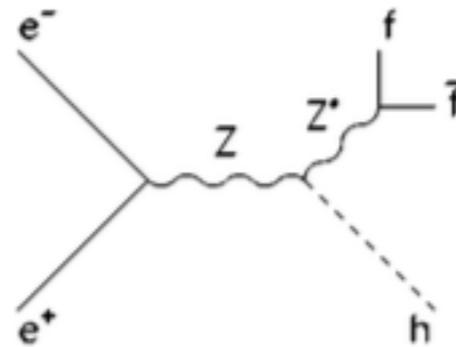
Outline

- ▶ 1
 - Introduction
 - Accelerators
 - Detectors
 - EW constraints
- ▶ 2
 - Search at LEP1 / LEP 2
 - Statistics: likelihood and hypothesis testing
- ▶ 3
 - TeVatron
 - Multivariate Analysis Techniques
 - Results
- ▶ 4
 - LHC
 - channels overview
- ▶ 5
 - dissect one analysis
 - Higgs combination
- ▶ 6
 - Extras
 - differential distributions
 - off shell
 - Beyond Standard Model
 - pseudo-observables / EFT

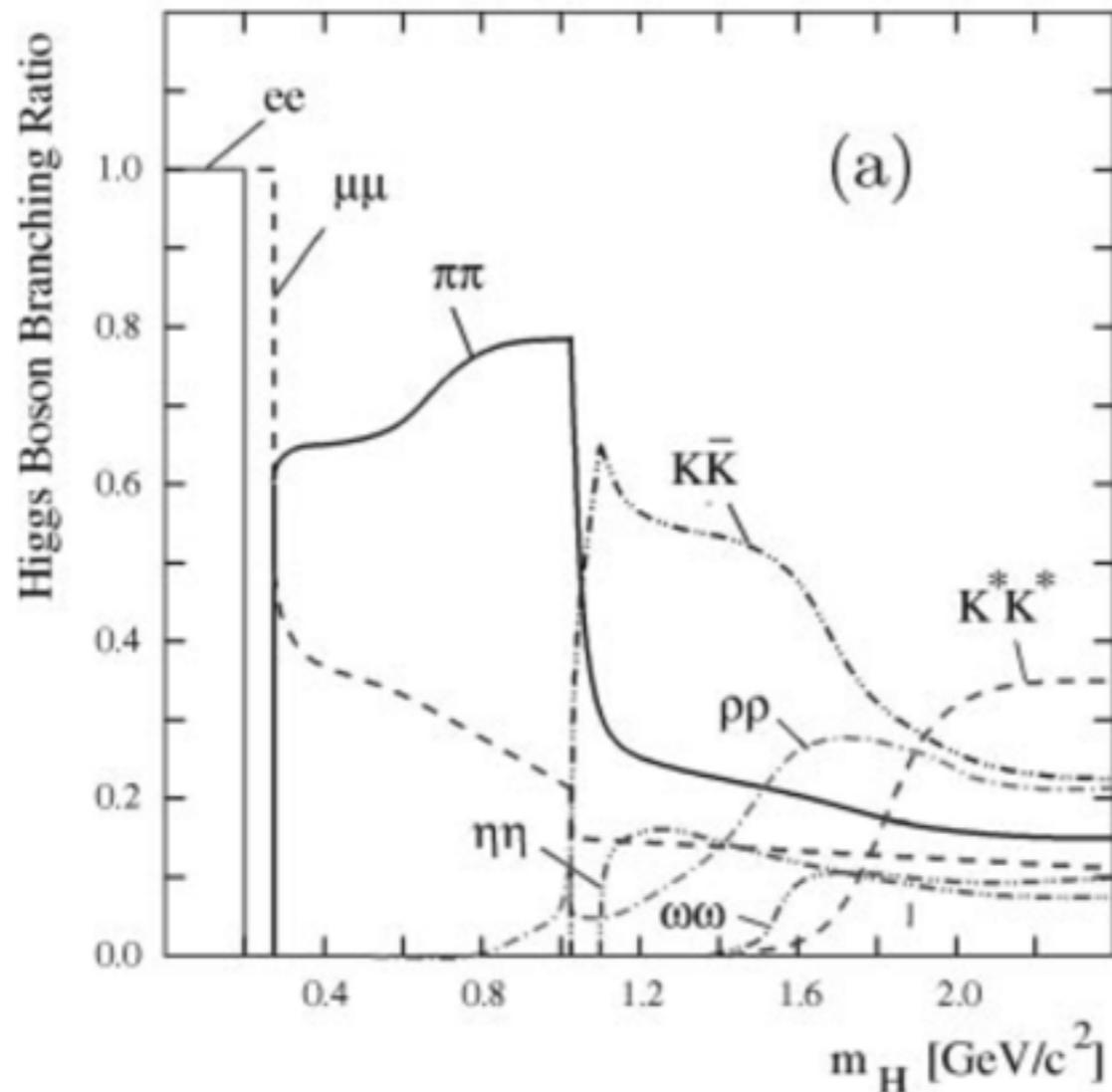
Higgs searches at LEP

LEP 1 searches (1989-1995)

Production mechanisms :



(much smaller cross section
larger backgrounds $ee \rightarrow qq\gamma$)



$2 < m_H < 20 \text{ GeV}$

look for heaviest fermion kinematically allowed

$m_H > 20 \text{ GeV}$

Higgsstrahlung: $Z \rightarrow \nu\nu$ or $Z \rightarrow e^+e^-$ ($\mu^+\mu^-$)

$H \rightarrow b\bar{b}$

10 signal events expected in $13 \cdot 10^6$ events
(all other channels swamped by $Z \rightarrow qq$ bkg)

LEP 1 limit : $m_H < 65.6 \text{ GeV @ 95\% CL}$

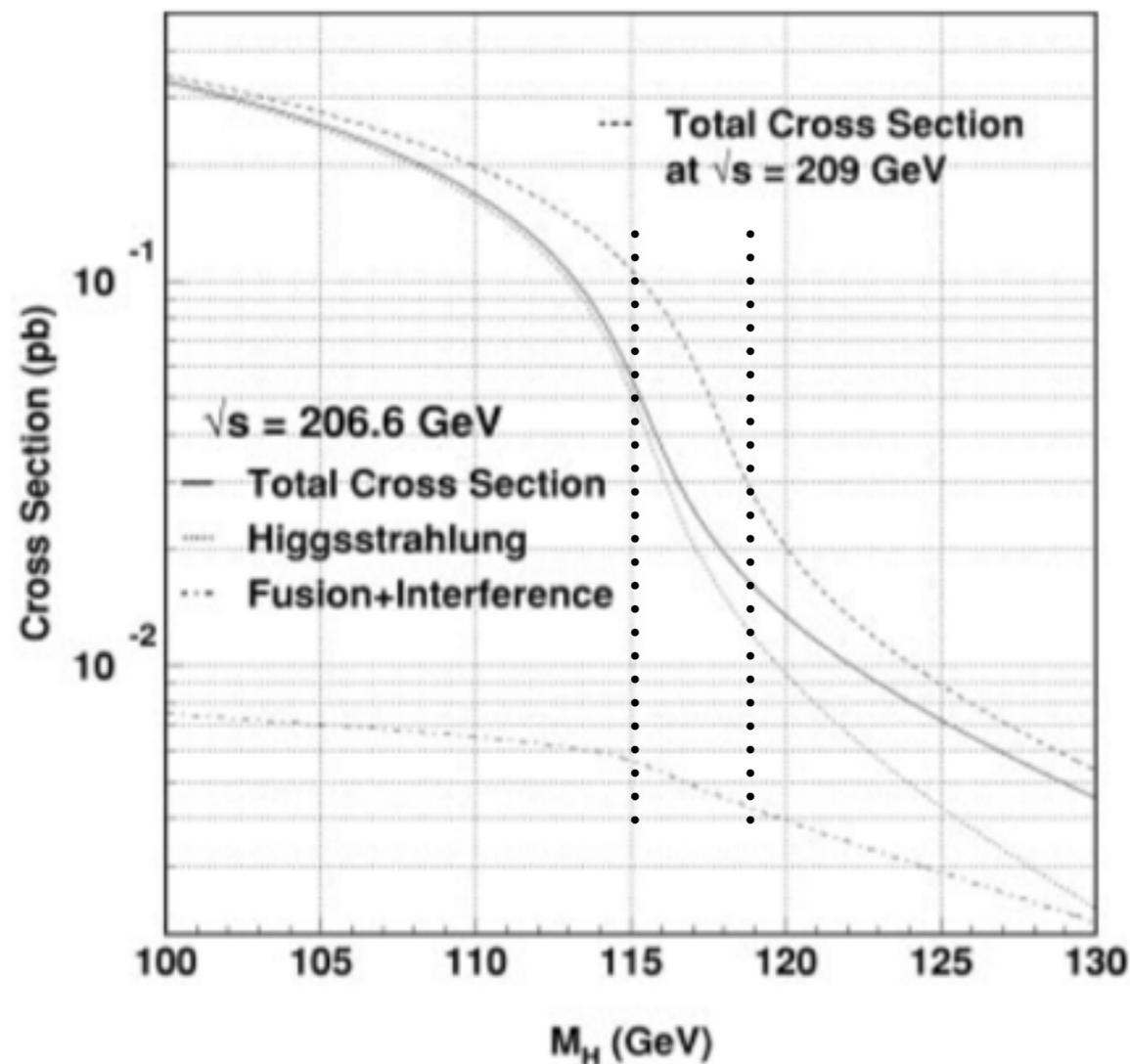
What does it happen below the e^+e^- threshold ?
(not shown in this plot)

LEP 2 (1995-2000)

Threshold effect : $m_{\text{threshold}} = \sqrt{s} - m_Z$

Confront this with an hadron collider

⇒ the LEP Higgs sensitivity depends dramatically on \sqrt{s}



Machine design highest energy:

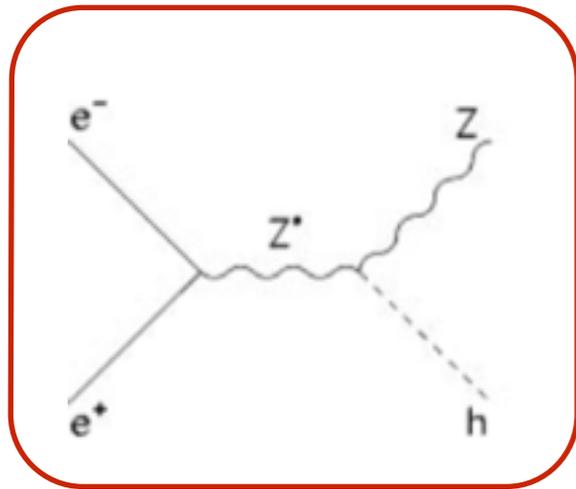
RF: 6 MV/m, $\sqrt{s} = 192$ GeV

mH sensitivity up to ~ 100 GeV

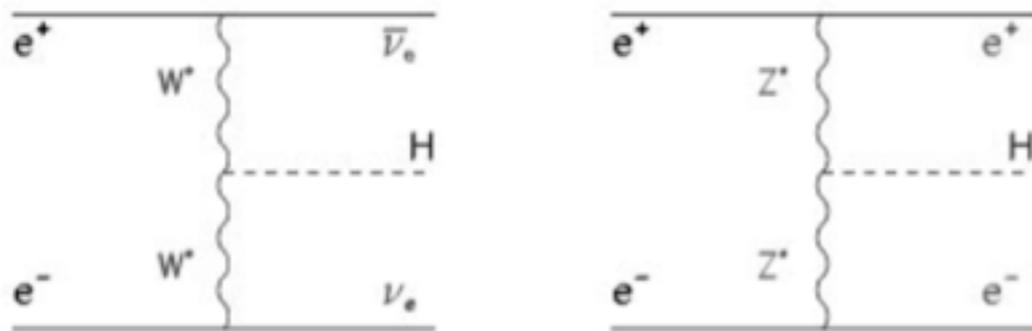
Upgrade and be smart to surpass the design capabilities:

- upgrade cryogenics: RF up to 7.5 MV/m (increased stability)
 $\sqrt{s} = 204$ GeV ; $mH < 112$ GeV
- run with one klystron margin (average trip frequency reduced below 1h thanks to improved stability)
 $\sqrt{s} = 205.5$ GeV ; $mH < 113$ GeV
- reduce 350MHz RF by 100Hz (different orbit effectively more bending from quadrupoles)
 $\sqrt{s} = 206.9$ GeV ; $mH < 113.6$ GeV
- Unused orbit correctors used as dipole
 $\sqrt{s} = 207.3$ GeV ; $mH < 113.85$ GeV
- reinstall 8 old Cu cavities from LEP 1
 $\sqrt{s} = 207.7$ GeV ; $mH < 114$ GeV
- miniramps (tradeoff between energy/stability/fill time)
 $\sqrt{s} = \sim 209$ GeV ; $mH \sim 115$ GeV

Production and decay

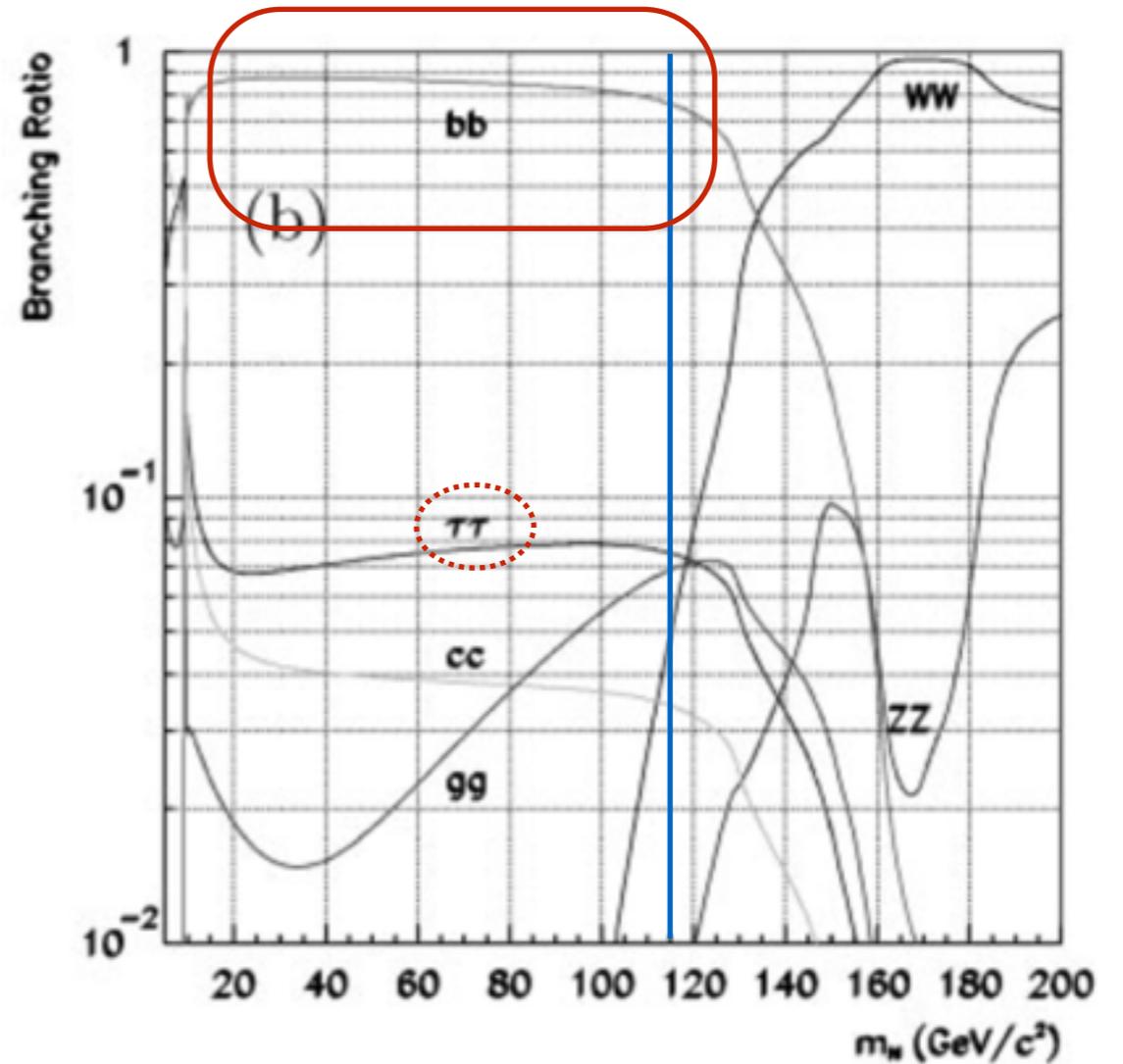


Higgsstrahlung



Vector Boson Fusion (VBF)

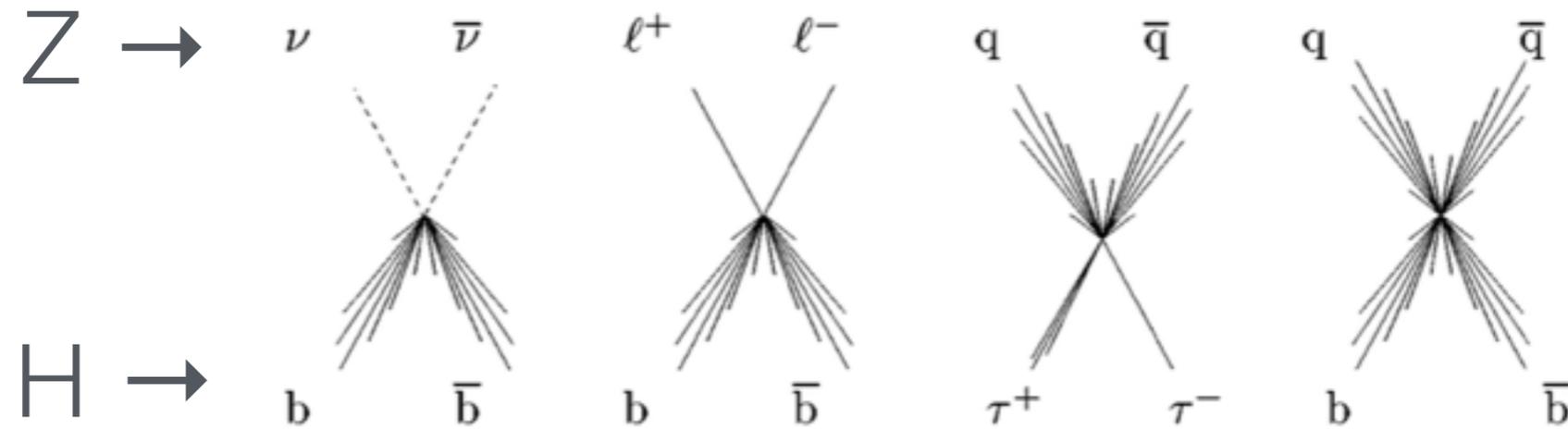
(and their interference
 $Z \rightarrow e^+e^-$, $Z \rightarrow \nu\nu$)



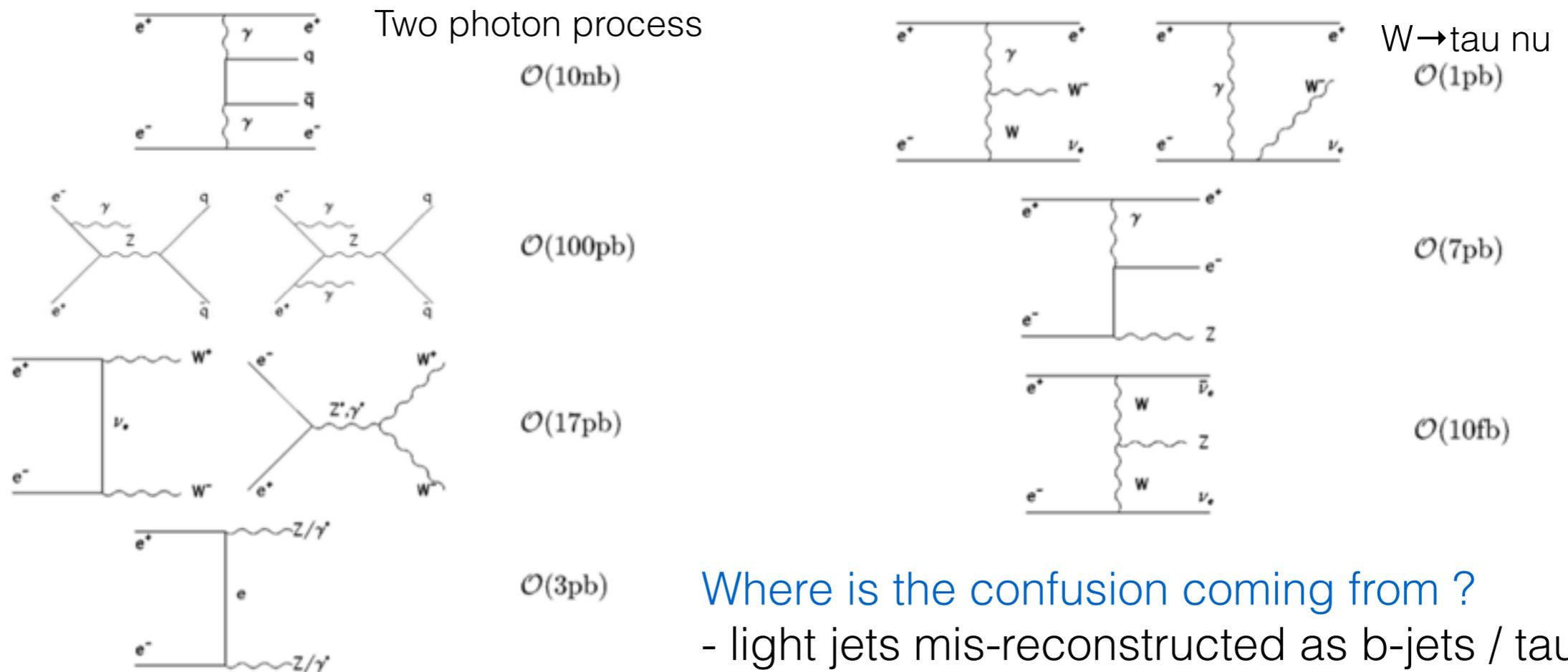
Compare this with an updated plot

Signal topologies / Backgrounds

Signal:



Background:



Where is the confusion coming from ?

- light jets mis-reconstructed as b-jets / taus
- non reconstructed particles giving ME

Analyses

SM backgrounds at LEP were well modelled (generators) and simulated (detectors).

Often use Monte Carlo to model the backgrounds !

Calibrate the detectors at the Z peak

Confront this situation with hadron colliders

Measurements development: get started by studying SM rare processes WW/ZZ production before attacking the Higgs

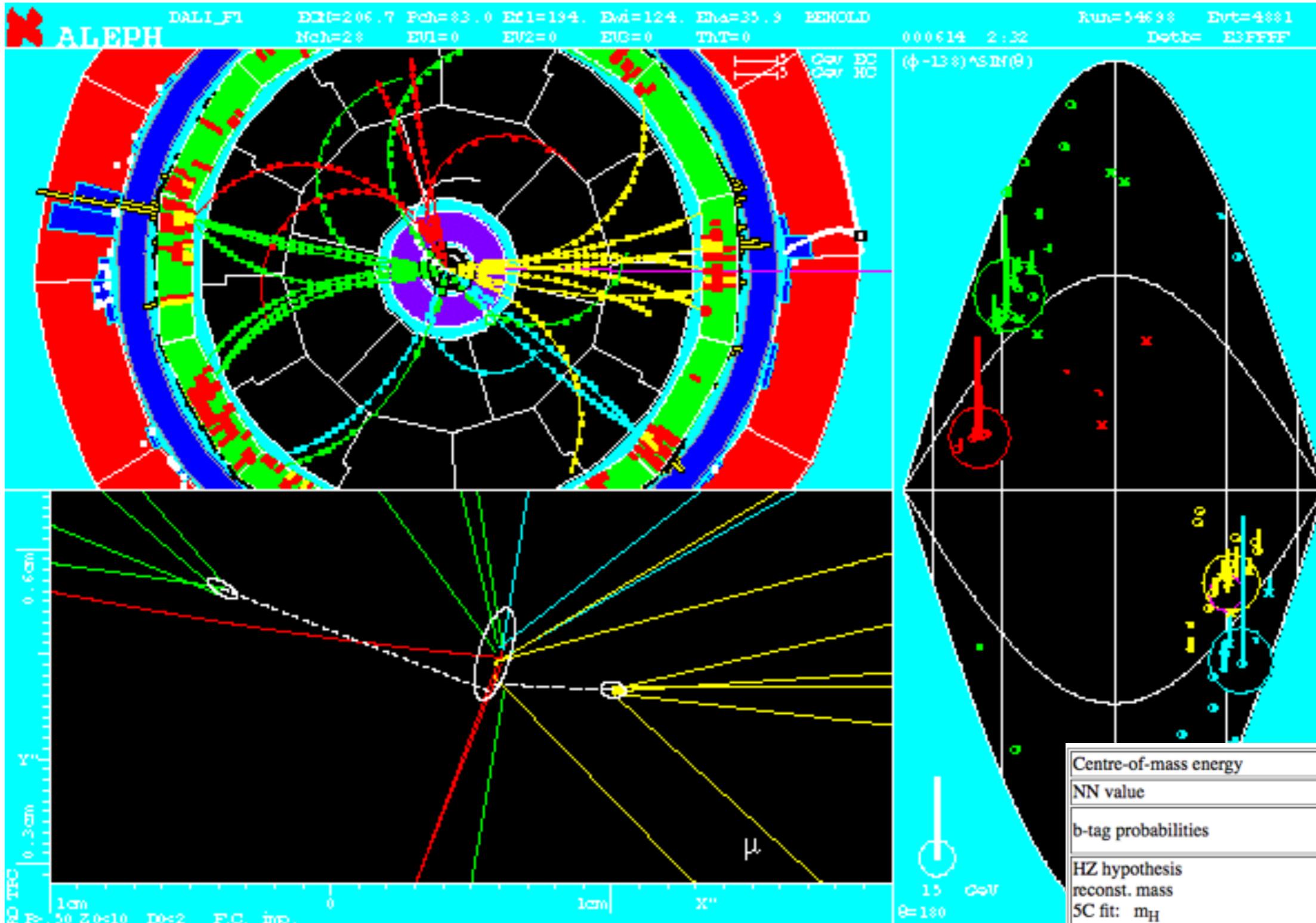
4 jets channel:

the most sensitive at LEP2 (at LEP 1 swamped by the $Z \rightarrow qq$ bkg):

- at kinematic threshold Z and H are produced at rest: 4 jets in one plane
 - main backgrounds $e^+e^- \rightarrow ZZ$, $e^+e^- \rightarrow WW$, $e^+e^- \rightarrow qq$
- $Z \rightarrow bb$: 4b case high purity but pairing ambiguities
 - main background ZZ
- typical mass resolution ~ 3 GeV

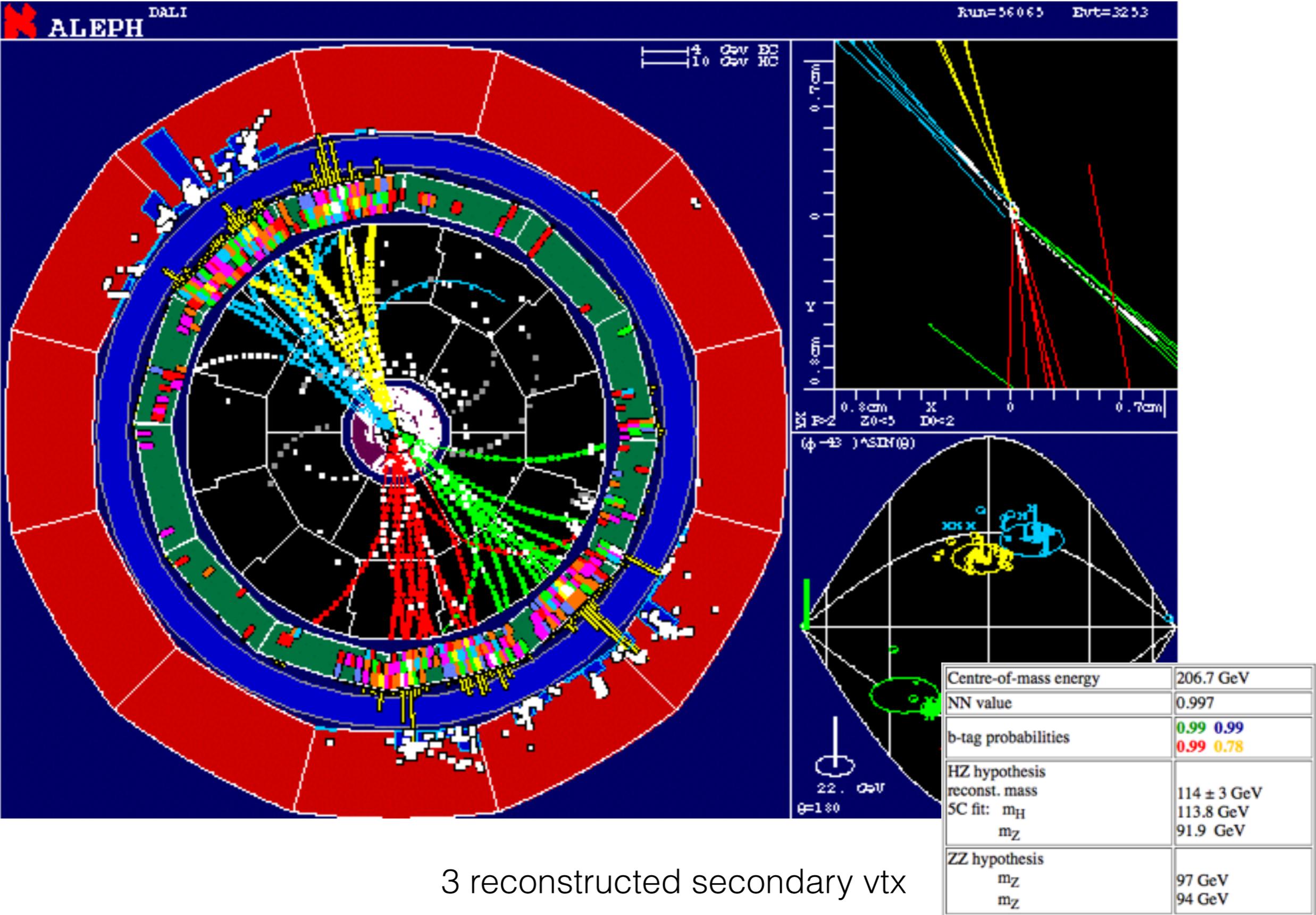
Channels sensitivity is different at LEP and at the hadron colliders: why ?
what happens to the 4 jets channel ?

qqbb



Centre-of-mass energy	206.7 GeV
NN value	0.996
b-tag probabilities	0.99 0.99 0.14 0.01
HZ hypothesis reconst. mass	114 ± 3 GeV
5C fit: m _H	112.4 GeV
m _Z	93.3 GeV
ZZ hypothesis	
m _Z	102 GeV
m _Z	91.7 GeV

bbbb



3 reconstructed secondary vtx

Analyses

Missing Energy: Why at an hadron collider we use the missing *transverse* energy ?

- $H \rightarrow bb$ and $Z \rightarrow \nu\nu$
- Main background : ZZ (irreducible)

- typical mass resolution as in the 4jets (!) ~ 3 GeV

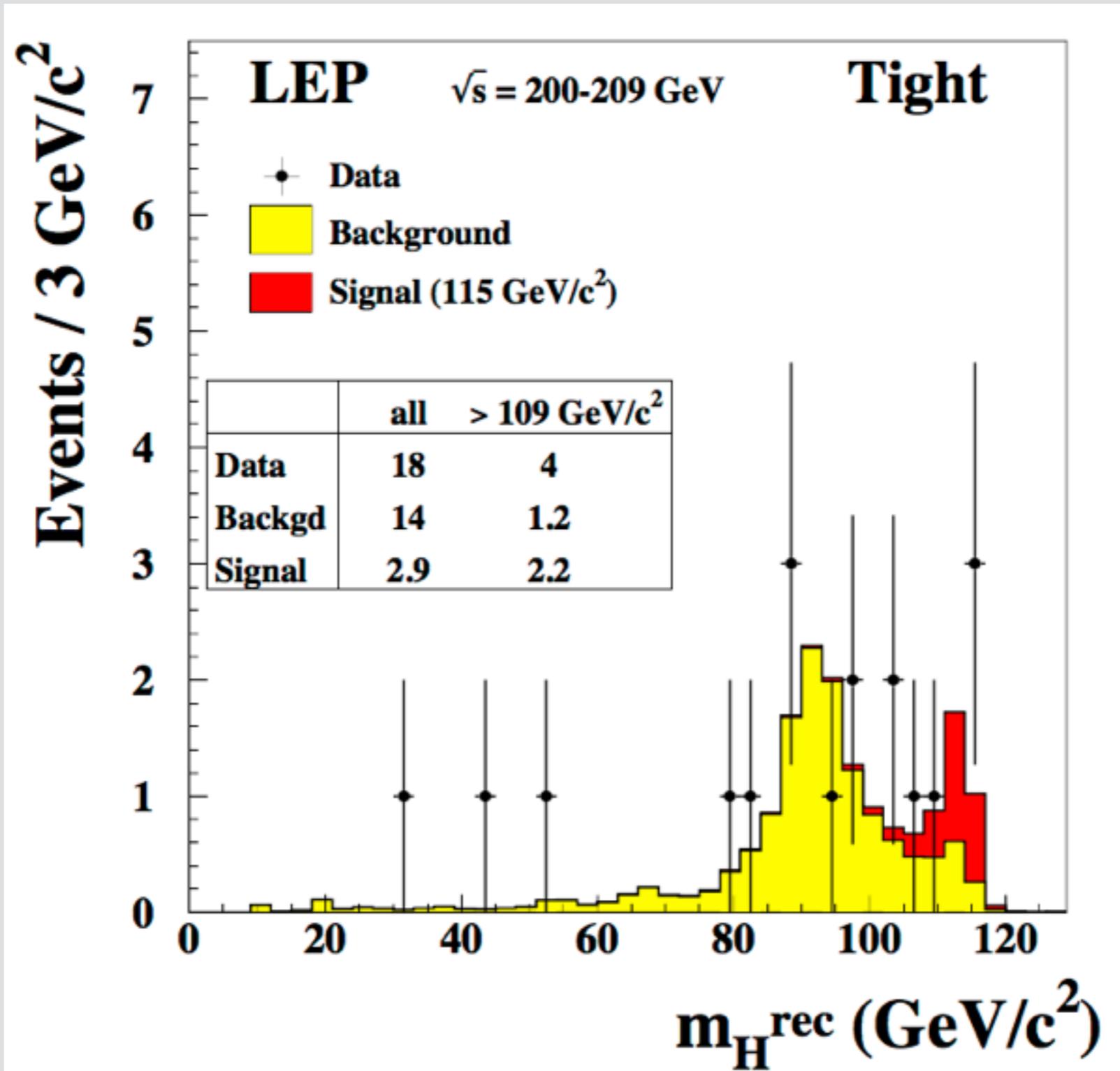
l^+l^- channel:

- $H \rightarrow bb$ and $Z \rightarrow l^+l^-$
- very small branching ratio (3% $Z \rightarrow ll$)
- Main background : ZZ (irreducible)

$\tau^+\tau^-$ channel:

- $H \rightarrow bb$ and $Z \rightarrow \tau^+\tau^-$ (neutrino in the final state)
- very small branching ratio (3% $Z \rightarrow ll$)
- Main background : ZZ (irreducible) and $Z \rightarrow bb$ (mis-reconstructed as τ)

Statistical inference at LEP



Use data to take decisions

Formulate an **hypothesis** (precisely), collect data, test the data against the hypothesis then **accept or reject**.

The way the hypothesis are defined is “reversed”, i.e. you always check that a hypothesis is NOT consistent with data.

In statistics/physics one cannot meaningfully accept a hypothesis: one can ONLY reject them.

Definitions:

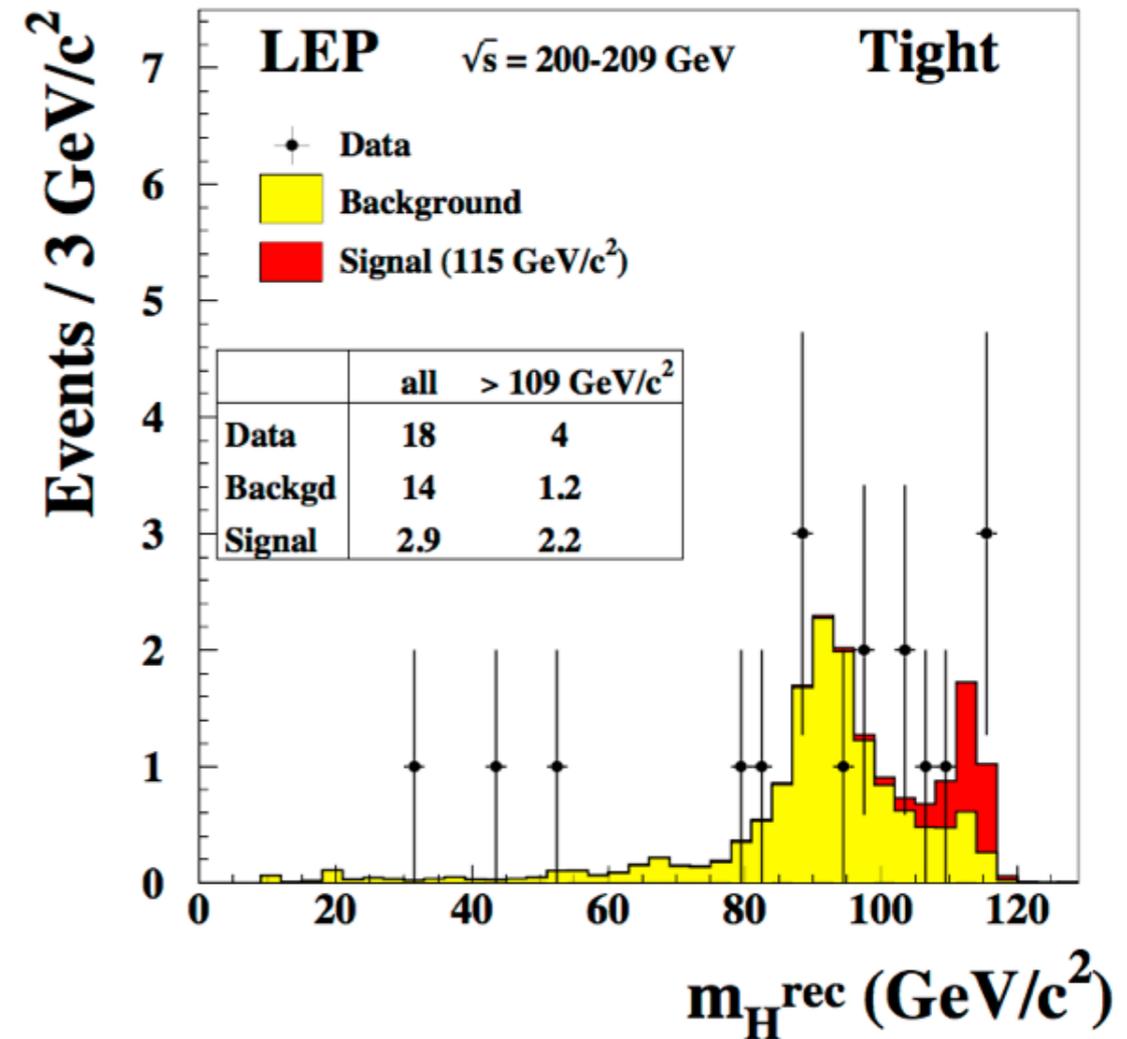
H_0 : null hypothesis defined to be the hypothesis under consideration.

H_1 : Alternative hypothesis

Typically H_0 is the **background only hypothesis** while H_1 adds the presence of some signal.

Simple hp: the expected PDF of the random variable (data) is completely fixed/specified

Composite hp: not all parameters are fixed, but they lie within a range



Test statistics

To quantify the agreement between the observed data and a given hypothesis one constructs a function of the measured data (x) and the given hypothesis H

$$\text{test statistics} := t(x|H_p)$$

The test statistics can be one-dimensional $t(x|H_p)$ or multidimensional $t(x|H_p)$.

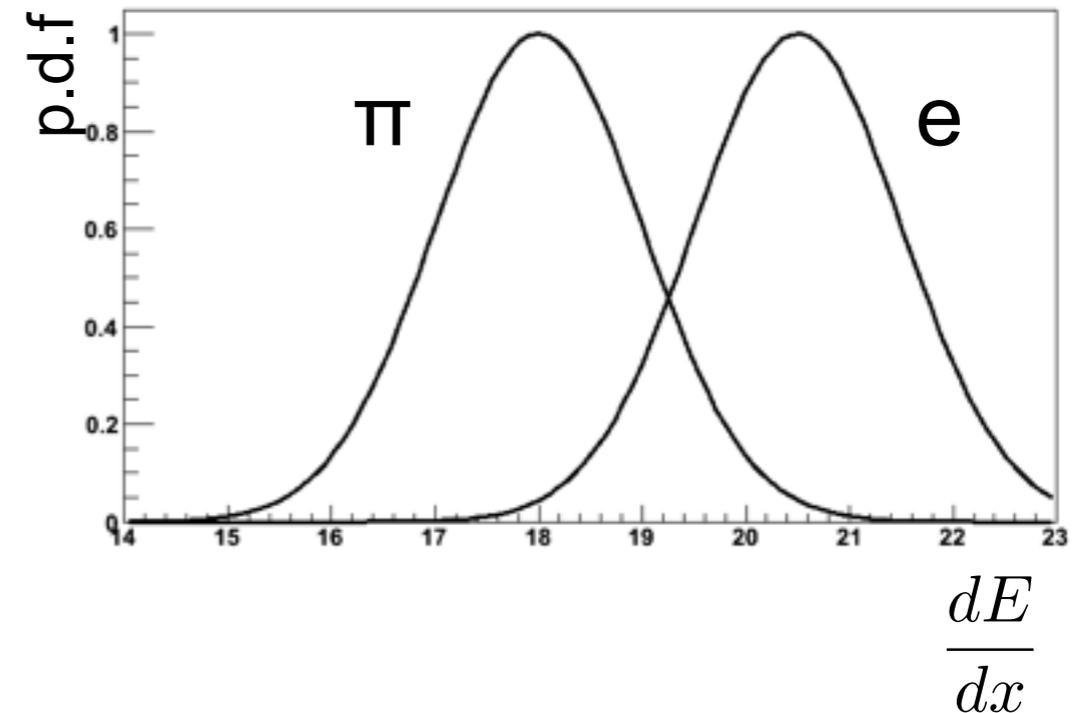
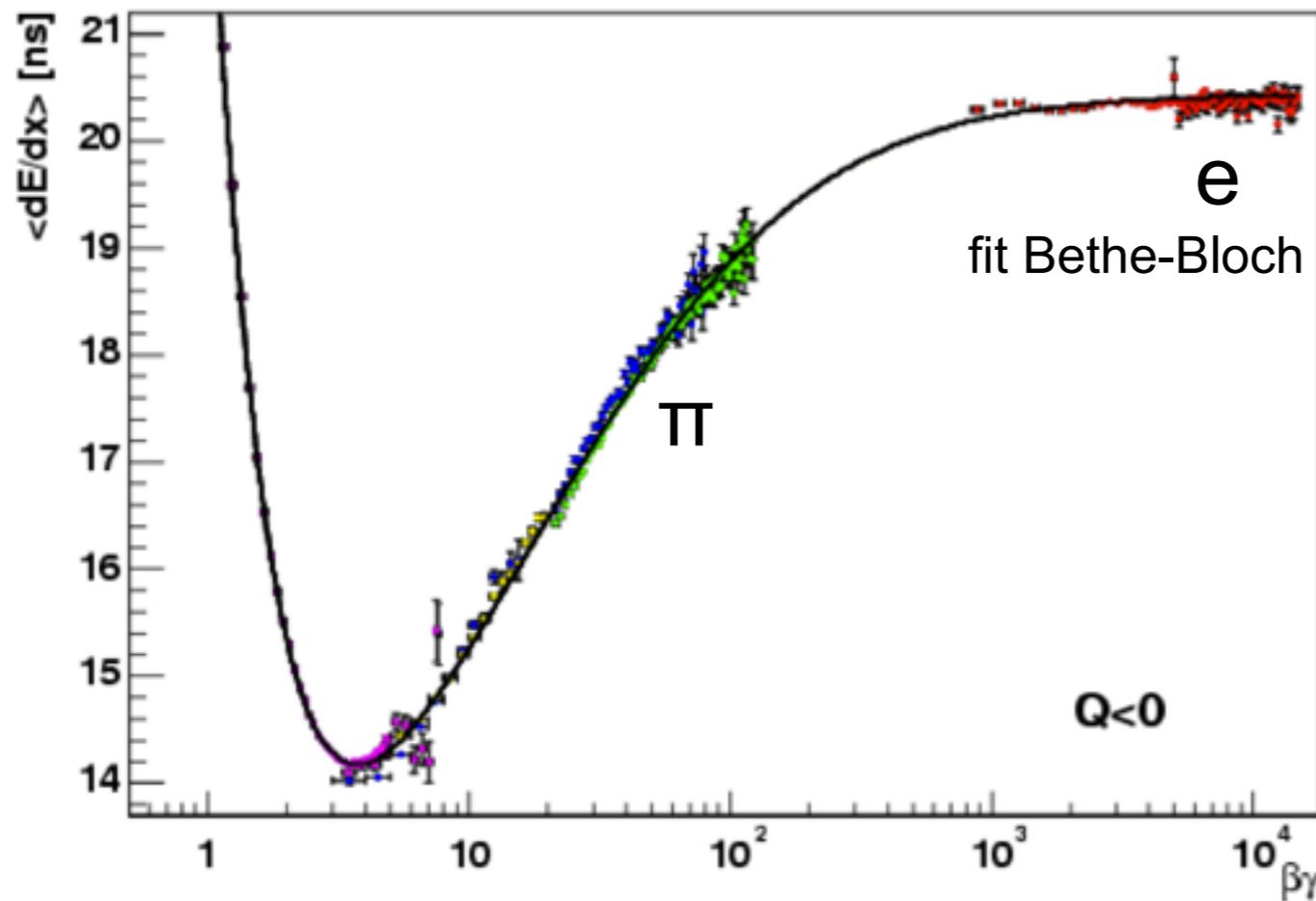
Typical test statistics:

- number of events
- an observable ($dE/dx, \dots$)
- a function of the observables (invariant mass from a 4 vector)
- a likelihood
- a ratio of likelihoods
- ...

The choice of the test statistic $t(x|H_p)$ depends on the particular case,
there is no general rule !

Different test statistics will give different “results”: **PHYSICS judgement is important !**

Example: electron/ π separation



Here we built the test statistics from DATA using selected labeled candidates

Test statistics: dE/dx by a charged particle in a ionization chamber.
The ionization is a statistical process

Null hypothesis:
$$P(x|H_0) = P\left(\frac{dE}{dx} | e^\pm\right)$$

Alternative hypothesis:
$$P(x|H_1) = P\left(\frac{dE}{dx} | \pi^\pm\right)$$

pdf from TOY experiments

In general we don't have a labelled sample of signal and background samples from data. Sometimes we can use data in "signal-free" control regions to build the pdf for the background, but often (and by definition in case of searches) we don't have a clean sample of "background-free" signal to build its pdf.

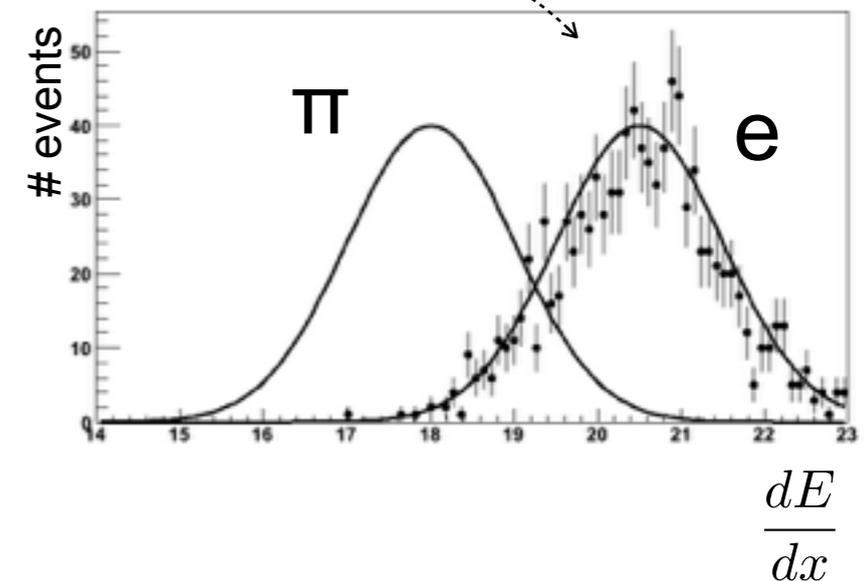
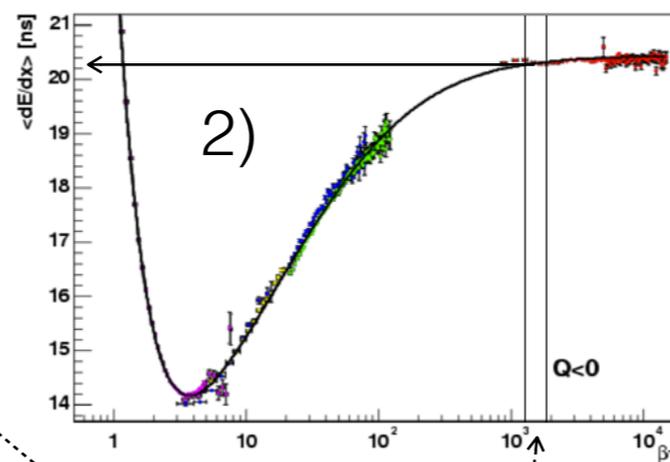
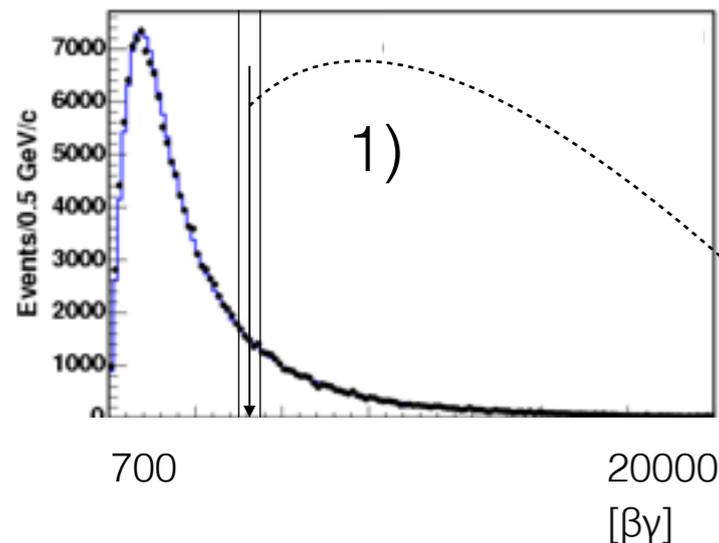
Use Monte Carlo "toy samples"

(toy means you don't run the full generator+detector simulation.

You generate pseudo-data sampling some high level distribution)

Example: build the pdf for the electron sample

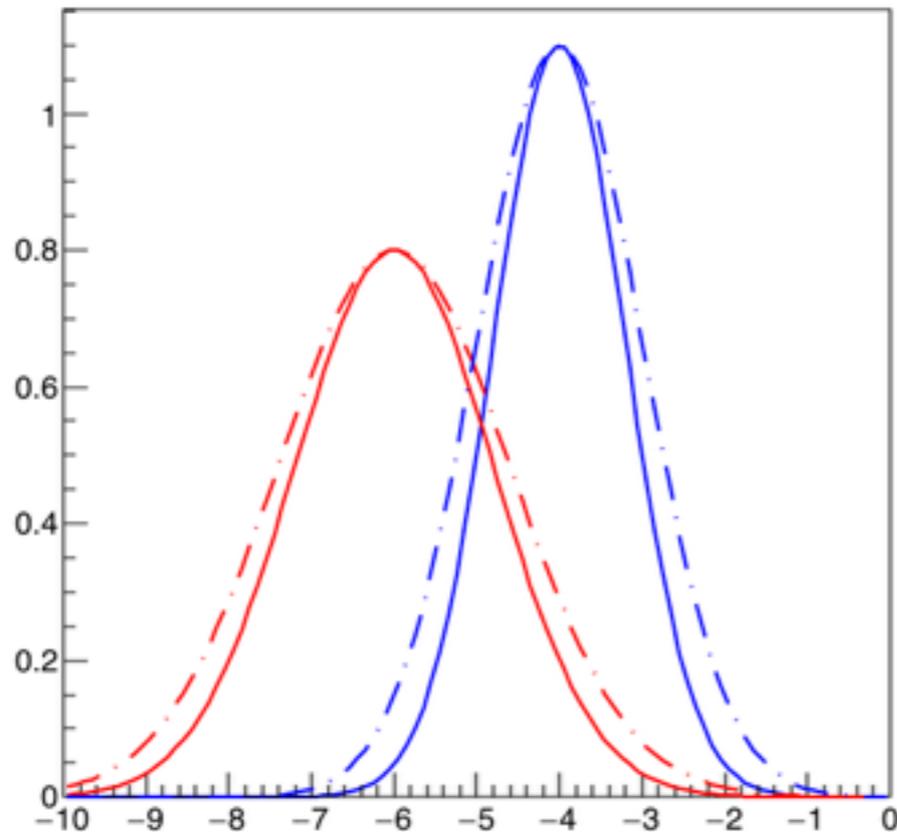
- 1) Sample the momentum distribution of the electrons (e.g. hit / miss)
- 2) Read out the dE/dx corresponding to $\beta\gamma$



Nuisances

What if signal (or background) descriptions are **NOT known exactly**.

In the previous example suppose the dE/dx is known with an uncertainty



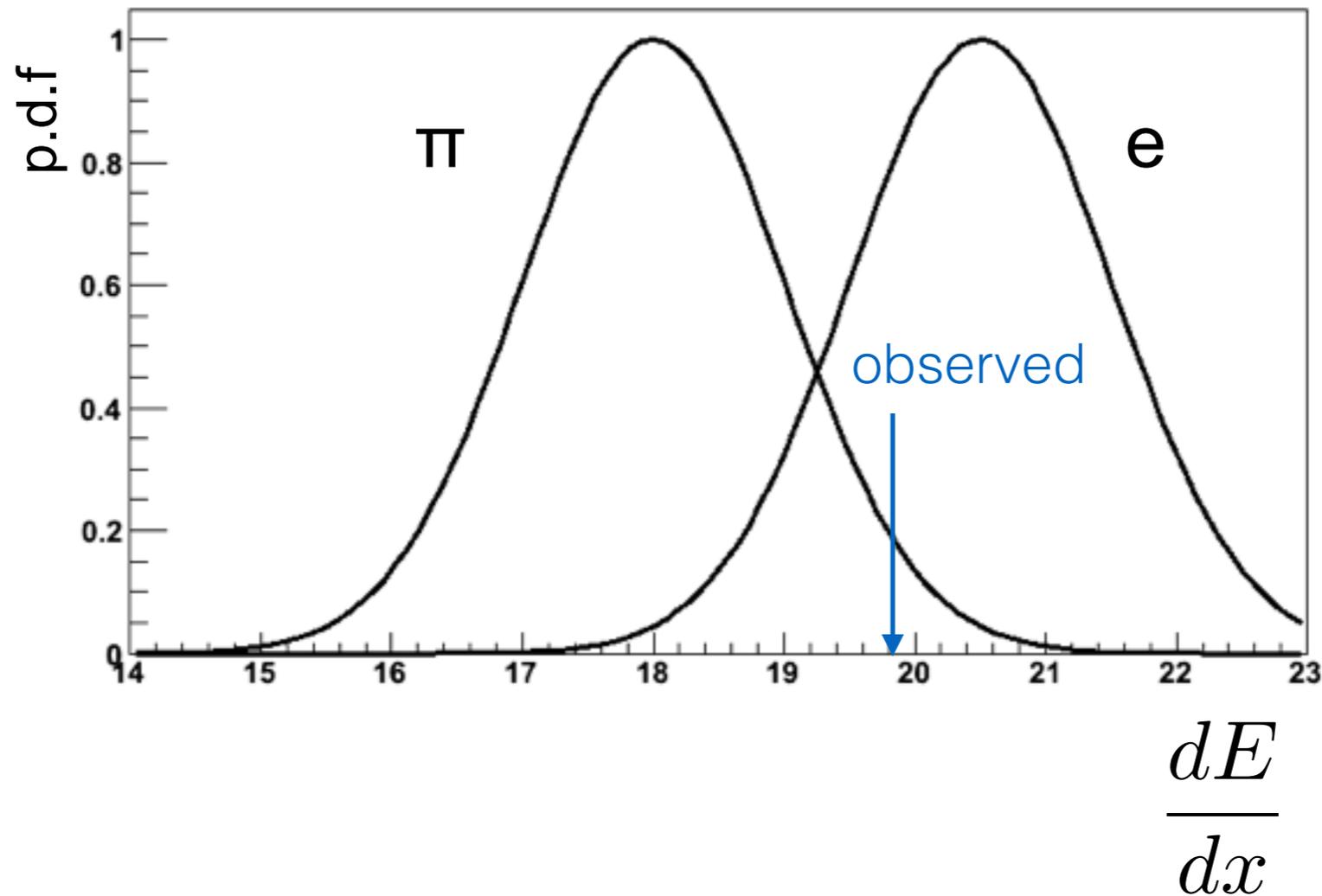
Both $f(Q|s+b)$ and $f(Q|b)$ get **broader** because of the uncertainty on b .

Intuitively: your discrimination power is reduced

Such a systematic in the model is named: **NUISANCE** (\Rightarrow composite hypothesis)

Signal or background ?

Use the expected [distributions of the test statistics](#) to decide if a candidate is signal or background.



The name of the game will be to use the test statistics to [quantitatively say if your data contains signal](#)

Error types

Different ways of mistakenly interpret the data:

Type I: reject a true hypothesis (loss or false negative)

Type II: accept a false hypothesis (contamination or false positive)

eg: “Law court” The accused proclaims himself as innocent (H_0).

Type I: he’s really innocent and the jury rejects the hypothesis and convict him

Type II: he’s really guilty (and a liar) and the court accept the hypothesis and let him off.

	H_0 is true Truly not guilty	H_1 is true Truly guilty
Accept Null Hypothesis Acquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

twiki

eg: “Bump hunting” You analyze a mass spectrum. The hypothesis is bkg only (H_0).

Type I : there’s really no resonance, you reject the H_0 and you publish rubbish

Type II: there is a real resonance, you accept H_0 and you miss the Nobel

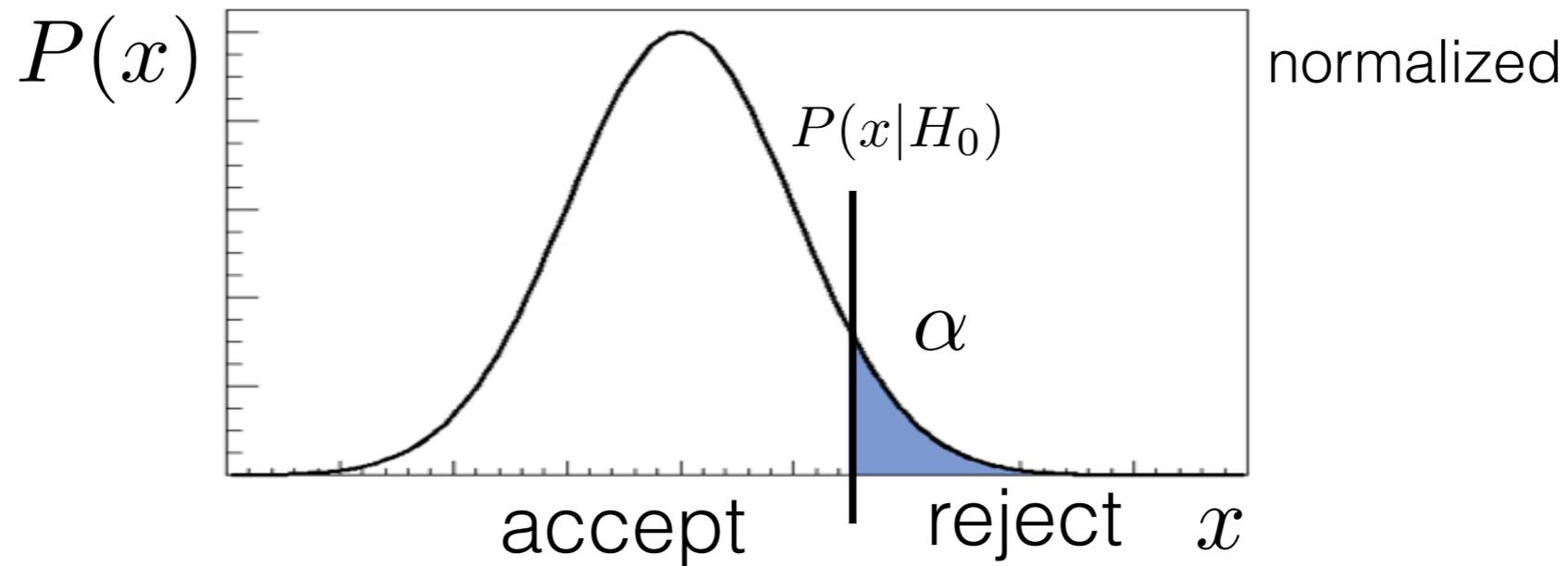
Test statistics properties

1) Significance (or Size)

Type I errors can be controlled pretty well:

Suppose you have a test statistics x (the data itself) and the null hypothesis $H_0: P(x|H_0)$

Partition the range of x in 2 regions. Define: acceptance / rejection



The probability for a type I error is the integral of $P(x|H_0)$ over the rejection region. This is called **significance of the test**

$$\alpha = \int_{t_{cut}}^{\infty} P(x|H_0) dx$$

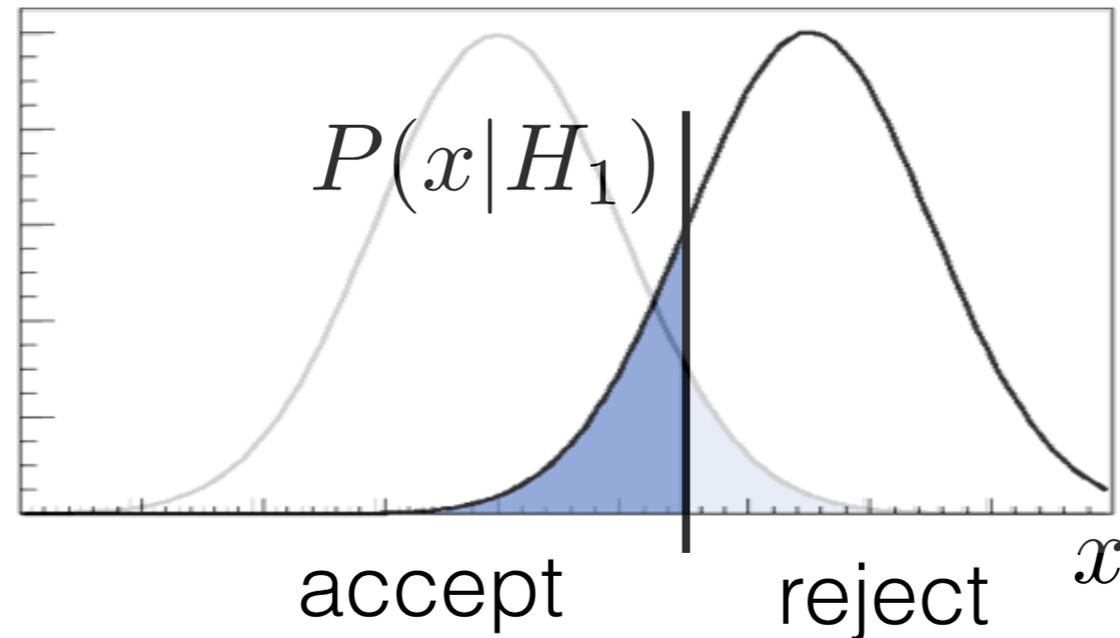
Typical values of α are 5%, 1%.

A test has a **significant level of $1-\alpha$** if the integrated probability to reject a true hypothesis is less or equal to α

Test statistics properties

2) Power

Suppose you have an alternative simple hypothesis H_1 and $P(x|H_1)$ is known. $P(x)$



$$\beta = \int_{-\infty}^{t_{cut}} P(x|H_1) dx$$

$1 - \beta$ is called the **power** of the test

A **good test** is the one with both α and β small,
i.e. high significance and high power

(i.e. H_0 and H_1 very different; large separation)

Likelihood ratio

Remember: the “best” test is the one that makes both α and β as small as possible. Such a test can be found if-and-only-if the hypothesis and the alternative are simple.

Neyman-Pearson lemma: choice of the acceptance region

the acceptance region giving the highest power (i.e. the highest purity) for a given significance level α (or efficiency $1-\alpha$) is the region of the space such that

$$\frac{g(\mathbf{t}|H_0)}{g(\mathbf{t}|H_1)} > c.$$

c is determined by the desired efficiency.

This is equivalent to a one dimensional statistics given by: $r = \frac{g(\mathbf{t}|H_0)}{g(\mathbf{t}|H_1)}$
the “likelihood ratio”

In practice: **it is usually difficult to determine c** , because one needs to know the complete joint pdfs of x : $g(\mathbf{t}|H_0)$ and $g(\mathbf{t}|H_1)$.

In general it's challenging to write a pdf with the proper correlations !

In practice: easier to use test statistics based on multivariate methods - MVA.

(We'll come back to this in the next lectures)

Statistical model

Take as an example variable the reconstructed mass of the Higgs candidate. The variable is **binned** (finite resolution).

For each bin we know the expected number of events from Signal and Background.

The probability to observe a number of events n with ν expected is given by:

$$P(n, \nu) = \frac{\nu^n}{n!} e^{-\nu} \quad (\text{Poisson})$$

The test statistics was chosen to be the **likelihood ratio**: $Q = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}$

$$\mathcal{L}_{s+b} = \frac{(s+b)^n}{n!} e^{-(s+b)} \cdot \prod_{j=1}^n \frac{sS(x_j) + bB(x_j)}{s+b}$$

$$\mathcal{L}_b = \frac{b^n}{n!} e^{-b} \cdot \prod_{j=1}^n B(x_j)$$

where:

s = # expected sig events, **function of m_H**

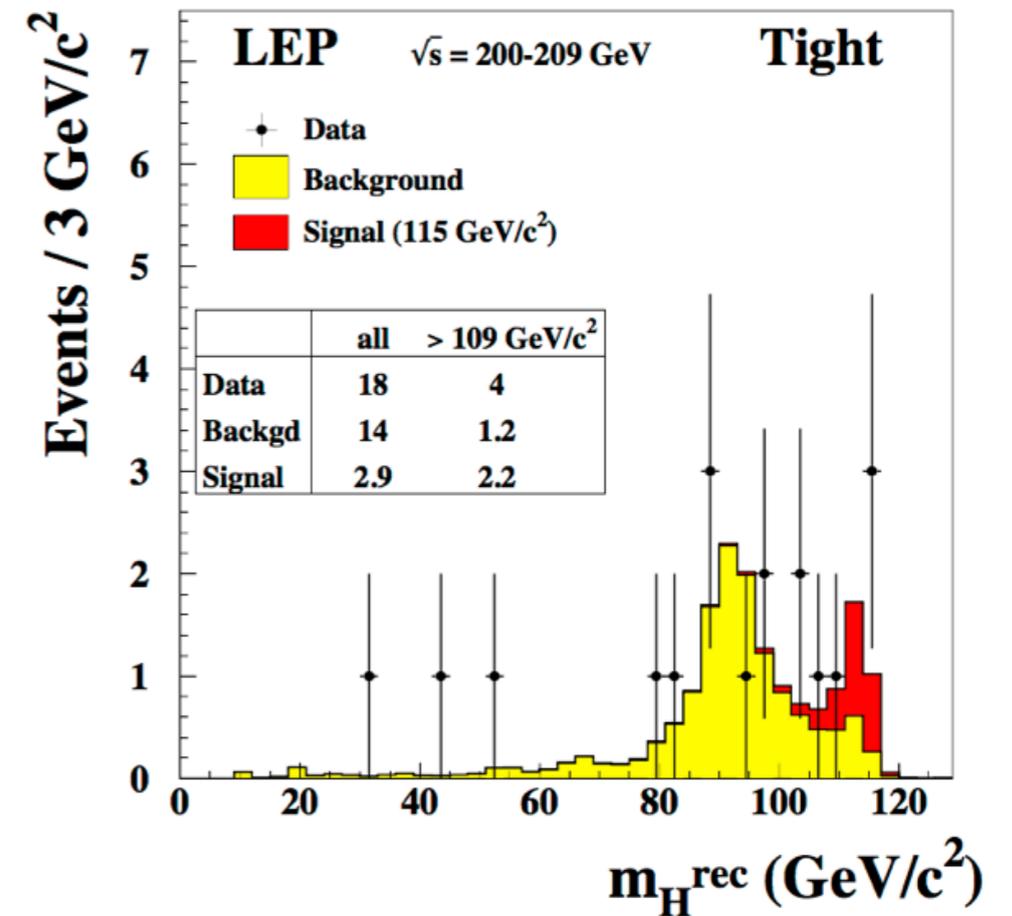
b = # expected bkg events

n = # observed events

x_j = value of the discriminating var j

$S(x_j)$ = signal pdf for the vars x , **function of m_H**

$B(x_j)$ = bkg pdf for the vars x



Statistical model

Generalization to N measurements (k runs over the measurements)

$$\mathcal{L}(\eta) = \prod_{k=1}^N \frac{\exp[-(\eta s_k(m_H) + b_k)] (\eta s_k(m_H) + b_k)^{n_k}}{n_k!} \times \prod_{j=1}^{n_k} \frac{\eta s_k(m_H) S_k(\vec{x}_{jk}; m_H) + b_k B_k(\vec{x}_{jk})}{\eta s_k(m_H) + b_k}$$

where: $\eta = 1$ gives \mathcal{L}_{s+b}

$\eta = 0$ gives \mathcal{L}_b

k runs over the N channels (different decay, different data periods, etc...)

s_k = # expected sig events

b_k = # expected bkg events

n_k = # observed events in channel k

x_{jk} = value of the discriminating var j in channel k

$S(x_{jk})$ = signal pdf for the vars x in channel k

$B(x_{jk})$ = bkg pdf for the vars x in channel k

$$Q = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}$$

$$q = -2 \ln Q(m_H) = 2 \sum_{k=1}^N \left[s_k(m_H) - \sum_{j=1}^{n_k} \ln \left(1 + \frac{s_k(m_H) S_k(\vec{x}_{jk}; m_H)}{b_k B_k(\vec{x}_{jk})} \right) \right]$$

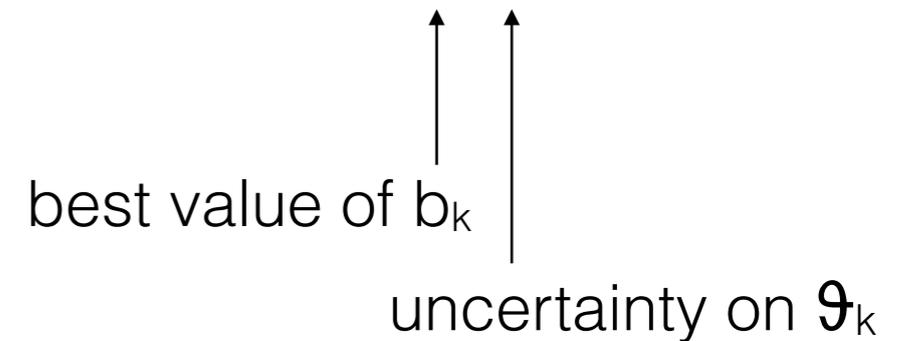
Each event contributes with a weight to the test statistics

To avoid numerical precision issues in treating very small numbers (we're multiplying several small probabilities, i.e. numbers $0 \leq p \leq 1$) we usually work with the [logarithm of Q](#)

Systematic uncertainties

Systematics uncertainties are included through nuisance parameters.

e.g. the background is known with an uncertainty, so: $b_k \rightarrow b_k \cdot f(\theta_k, \sigma_k)$



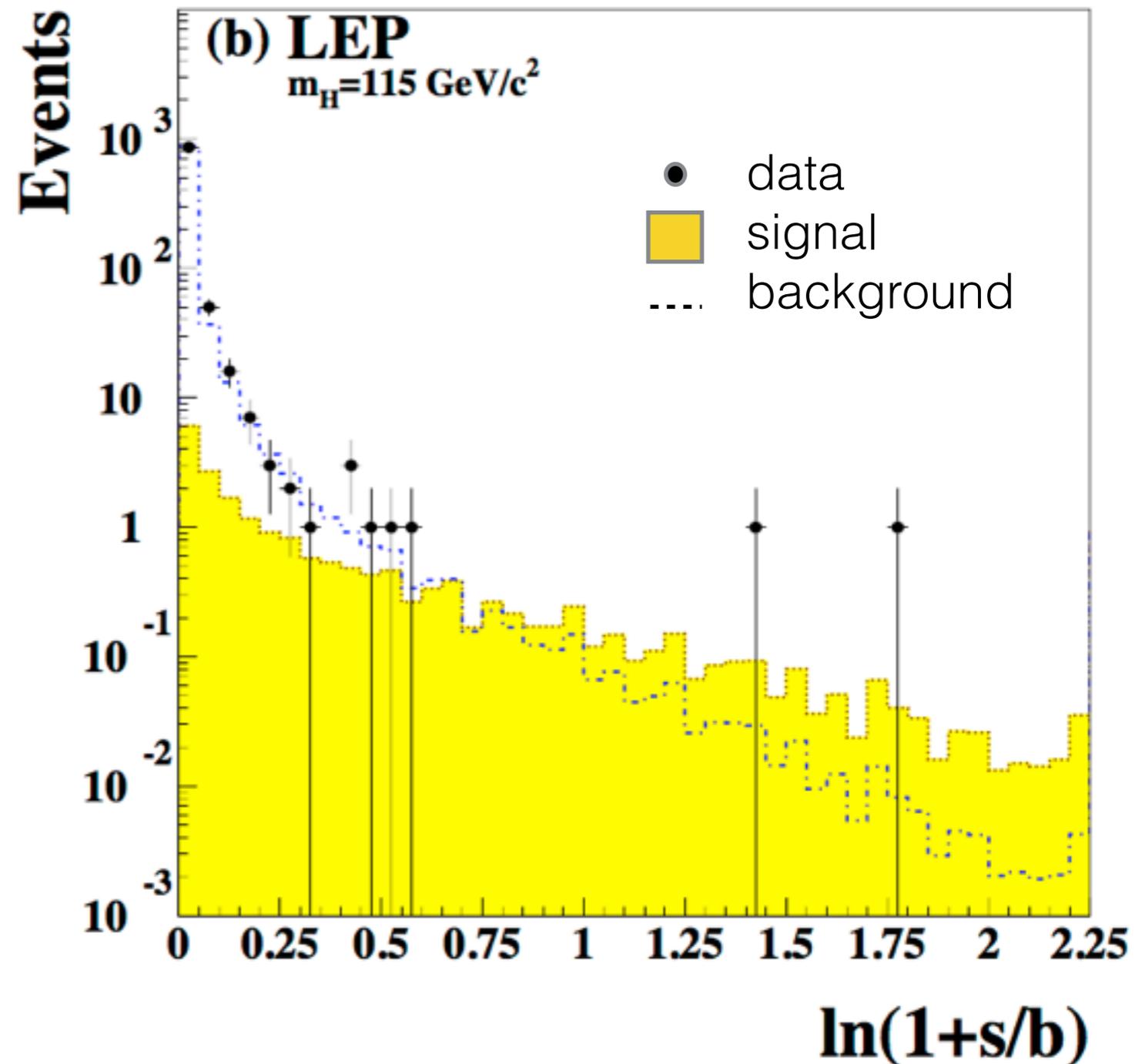
The function f represents a **constraint** on the parameter b_k

Each term of the likelihood affected by a systematic uncertainty gets multiplied by a constraint term (the functional form of the constraint depends on the variable and it can be a gaussian, log-norm, etc...)

LEP results

$$-2 \ln Q(m_H) = 2 \sum_{k=1}^N \left[s_k(m_H) - \sum_{j=1}^{n_k} \ln \left(1 + \frac{s_k(m_H) S_k(\vec{x}_{jk}; m_H)}{b_k B_k(\vec{x}_{jk})} \right) \right]$$

Because the separation between signal and background depends on logarithmically on S/B, the plot of $\log(1+s/b)$ shows the important region for signal search.



-2 ln Q for different hypotheses

$$q = -2 \ln Q = 2 \sum_{k=1}^N \left(s_k - \sum_{j=1}^{n_k} \ln \left(1 + \frac{s_k S_{kj}}{b_k B_{kj}} \right) \right)$$

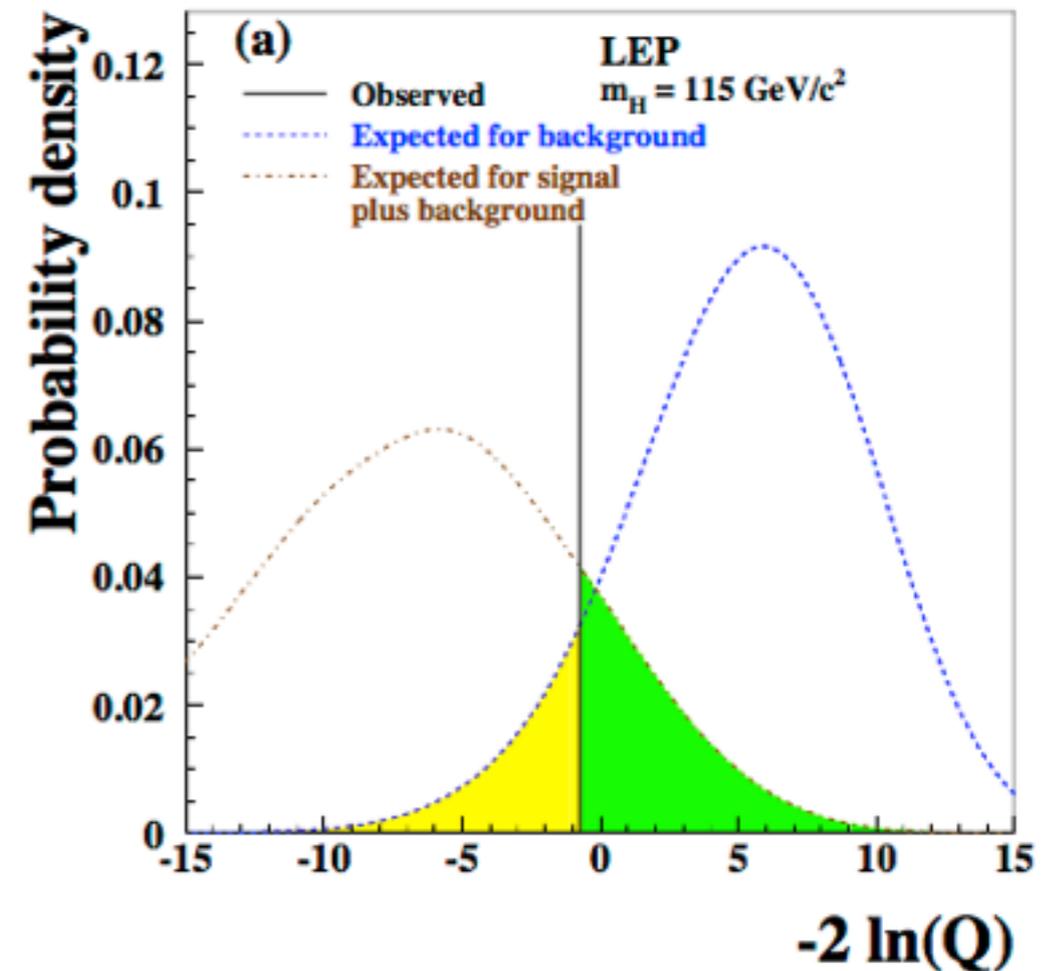
The “observed” is the value of q computed on data

“s+b” hp: $Q > 1 \Rightarrow q < 0$

“b” hp: $Q < 1 \Rightarrow q > 0$

Notes:

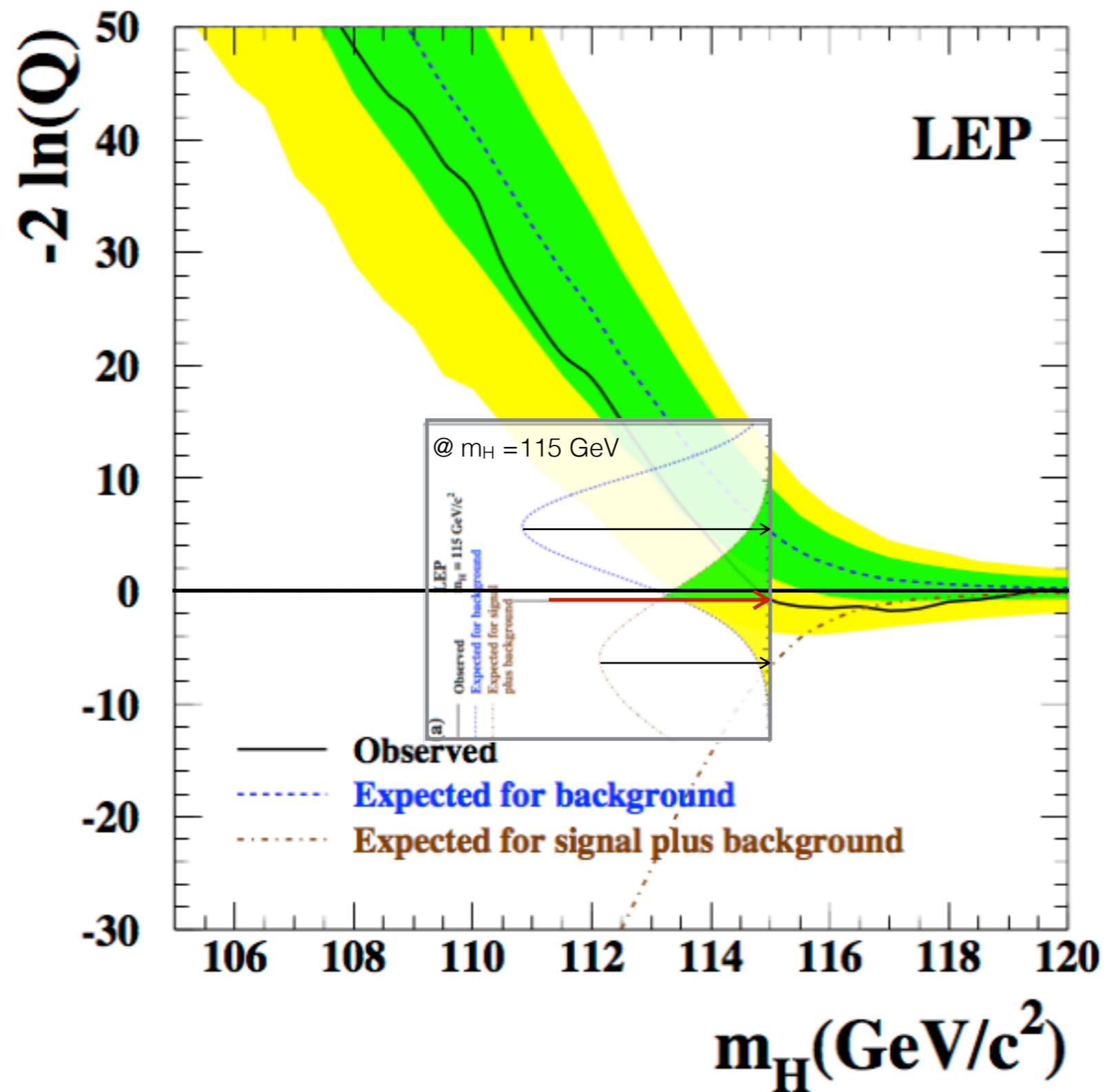
- q depends on the test mass
- q on data is computed with the nuisance at their best value



LEP results

“s+b” hp: $Q > 1 \Rightarrow q < 0$

“b” hp: $Q < 1 \Rightarrow q > 0$



ADLO results

