Experimental Particle Physics PHYS6011

Performing an analysis

Lecture 5

Data Analysis

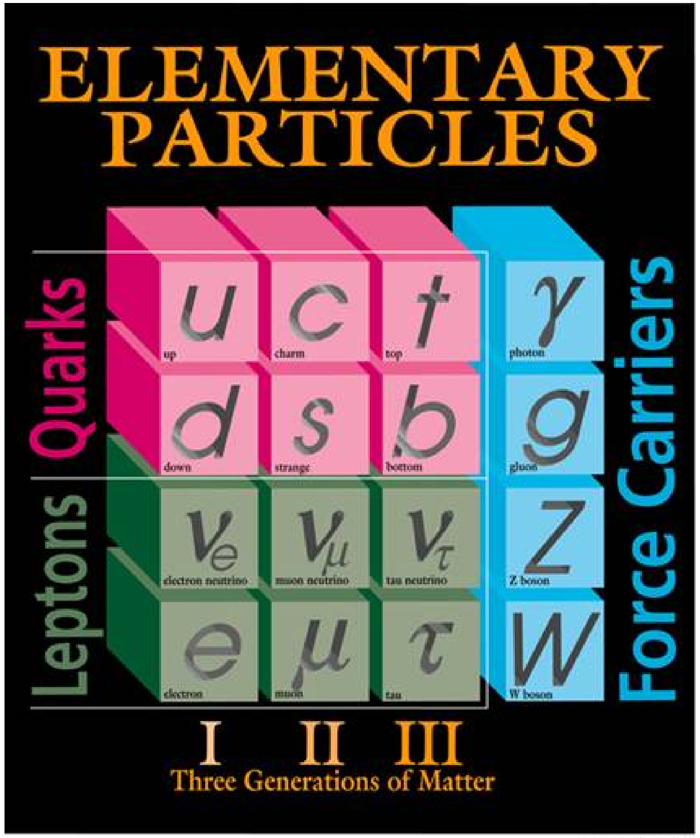
Extract physics from data

- Measure a quantity
- Search for new particles

1. Basic concepts
2. Monte Carlo methods
3. Signal
4. Backgrounds
5. Errors
6. Statistics

What do we measure?

In principle:



Fermilab 95-759

But in reality:

LIGHT UNFLAVORED MESONS
$(S = C = B = 0)$

For $I = 1$ (π, b, ρ, a): $u\overline{d}$, $(u\overline{u} - d\overline{d})/\sqrt{2}$, $d\overline{u}$;
for $I = 0$ (η, η', h, h', ω, ϕ, f, f'): $c_1(u\overline{u} + d\overline{d}) + c_2(s\overline{s})$

Quark content

$\pi^\pm$ $\quad I^G(J^P) = 1^-(0^-)$

Spin

Mass

Lifetime

Mass $m = 139.57018 \pm 0.00035$ MeV (S = 1.2)
Mean life $\tau = (2.6033 \pm 0.0005) \times 10^{-8}$ s (S = 1.2)
$c\tau = 7.8045$ m

$\pi^\pm \to \ell^\pm \nu \gamma$ form factors [a]

$F_V = 0.017 \pm 0.008$
$F_A = 0.0115 \pm 0.0005$ (S = 1.2)
$R = 0.059^{+0.009}_{-0.008}$

Form factor

π^- modes are charge conjugates of the modes below.

For decay limits to particles which are not established, see the appropriate Search sections (Massive Neutrino Peak Search Test, A^0 (axion), and Other Light Boson (X^0) Searches, etc.).

Decay Modes

Branching Fraction

π^+ DECAY MODES			Fraction (Γ_i/Γ)	Confidence level	p (MeV/c)
$\mu^+\nu_\mu$		[b]	(99.98770 ± 0.00004) %		30
$\mu^+\nu_\mu\gamma$		[c]	$(2.00 \pm 0.25) \times 10^{-4}$		30
$e^+\nu_e$		[b]	$(1.230 \pm 0.004) \times 10^{-4}$		70
$e^+\nu_e\gamma$		[c]	$(1.61 \pm 0.23) \times 10^{-7}$		70
$e^+\nu_e\pi^0$			$(1.036 \pm 0.006) \times 10^{-8}$		4
$e^+\nu_e e^+ e^-$			$(3.2 \pm 0.5) \times 10^{-9}$		70
$e^+\nu_e\nu\overline{\nu}$			$< 5 \times 10^{-6}$	90%	70
Lepton Family number (LF) or Lepton number (L) violating modes					
$\mu^+\overline{\nu}_e$	L	[d]	$< 1.5 \times 10^{-3}$	90%	30
$\mu^+\nu_e$	LF	[d]	$< 8.0 \times 10^{-3}$	90%	30
$\mu^- e^+ e^+ \nu$	LF		$< 1.6 \times 10^{-6}$	90%	30

Particle Properties

- Properties
  - Mass
    - Measure momentum and energy: $E^2 = p^2 + m^2$
  - Mass width → Lifetime
    - Measure momentum and energy or:
    - How many particles exist after t seconds
  - Branching Fraction
    - Reconstruct the decays and see how many there are.
  - Charge
    - Direction in a magnetic field
  - Spin
    - Angular distribution of decays
  - Structure e.g. Proton/Neutron/Nucleus
    - Scatter particles of the proton and look at distribution

Data Flow

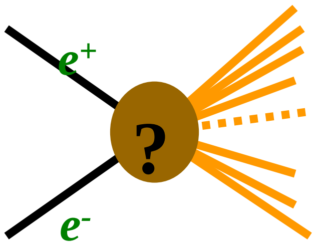
High Signal:Background

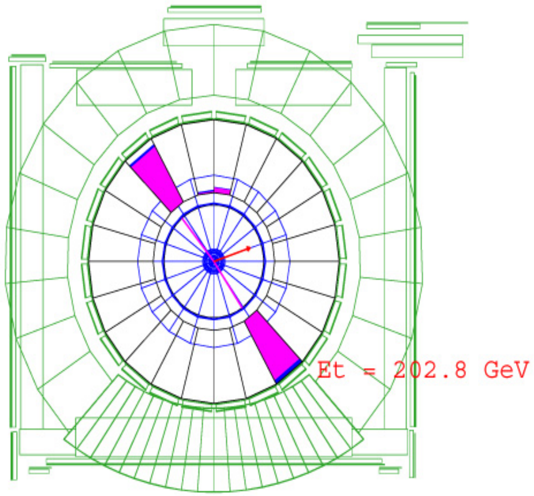
Low Signal:Background

Trigger and DAQ

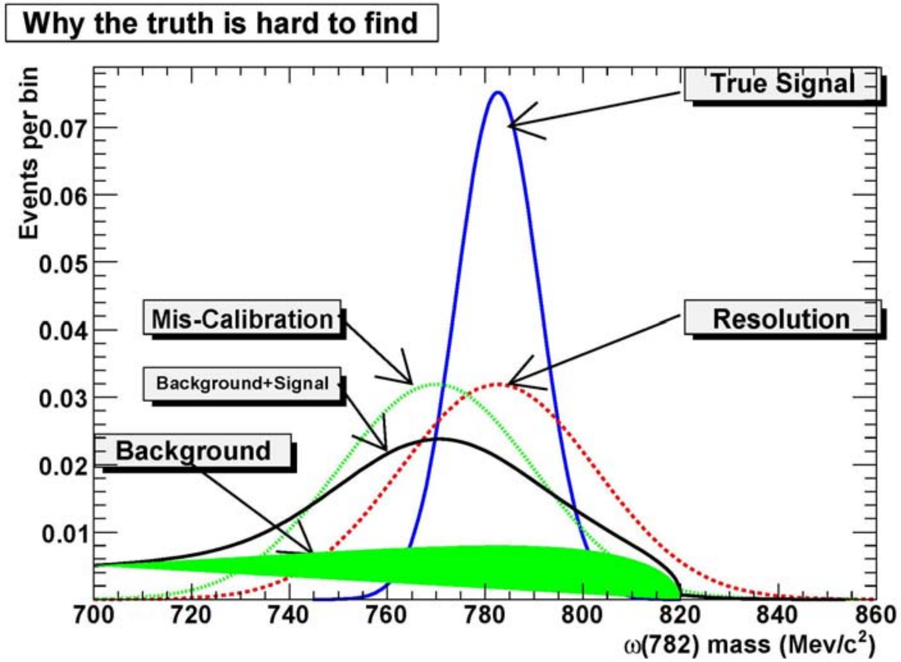
Event selection

Collision → Detector → Trigger and DAQ → Storage → Event selection → Analysis





Elements of Analysis



Not only *Data* but...

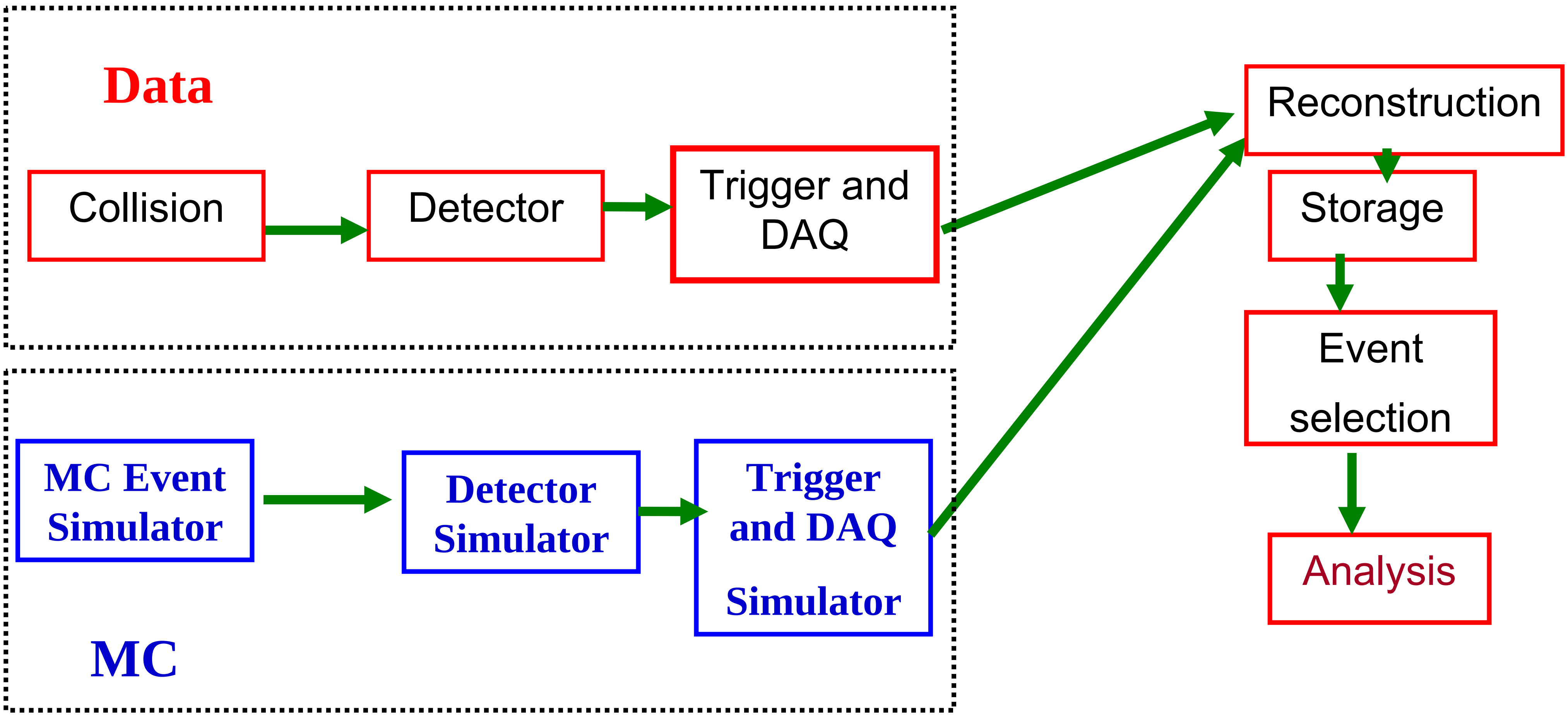
- Detector response to signal
- Background estimates
- Errors
  - statistical
  - systematic
- How to solve?
  - Try and evaluate from data
  - Sometimes need more...
    - Monte Carlo

Monte Carlo

1. Generate artificial data
2. Simulate detector response
3. Analyse simulated data as if it were real
   - Response to known input can be calculated
   - *Also used in detector design*

- Computer intensive
- Must be carefully tuned and checked

Data and Monte Carlo

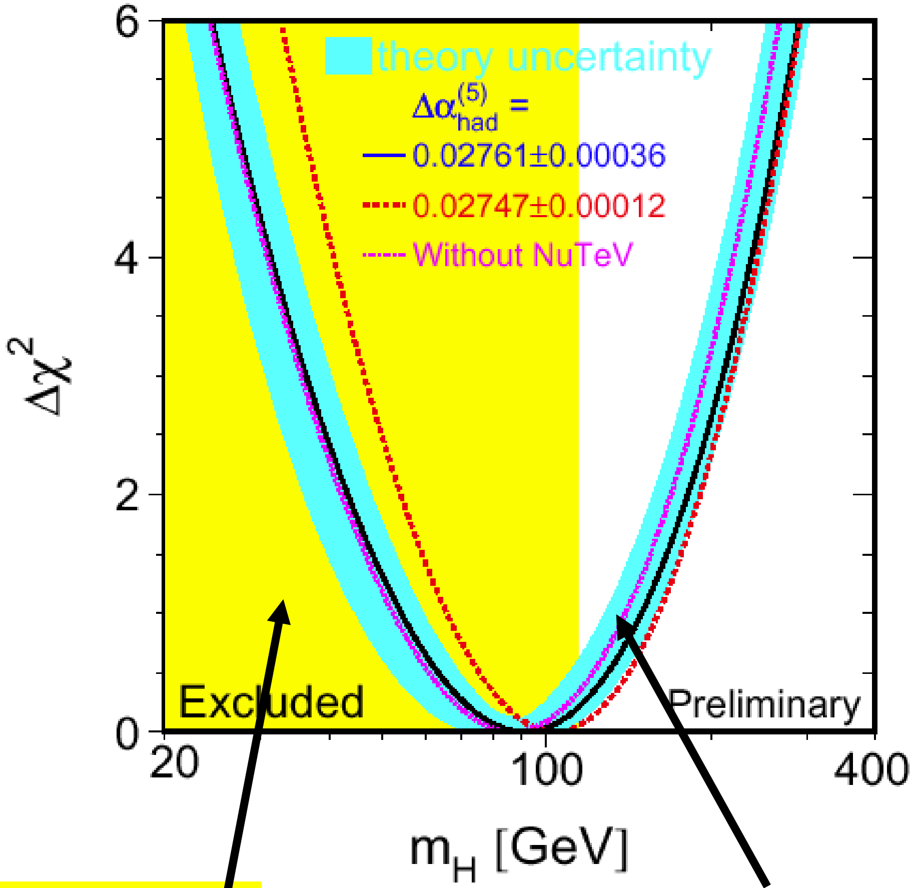


Search for the Higgs Boson

- Missing piece of Standard Model
- Standard Model Higgs theory well understood:
  - Mass is only free parameter
  - Clear predictions to test
- Most "New Physics" models have something equivalent to a Higgs boson ("MSSM Higgs", "little Higgs", etc…).
- Could be more than one type of Higgs boson
- Current limit M_H>115 GeV (LEP)

- Particle masses are generated by interactions with the scalar (Higgs) field.
- Couplings are fixed by the masses.
- Once M_H is known everything is predicted.
- So by measuring the coupling of the Higgs to particles of known mass we can test theory.

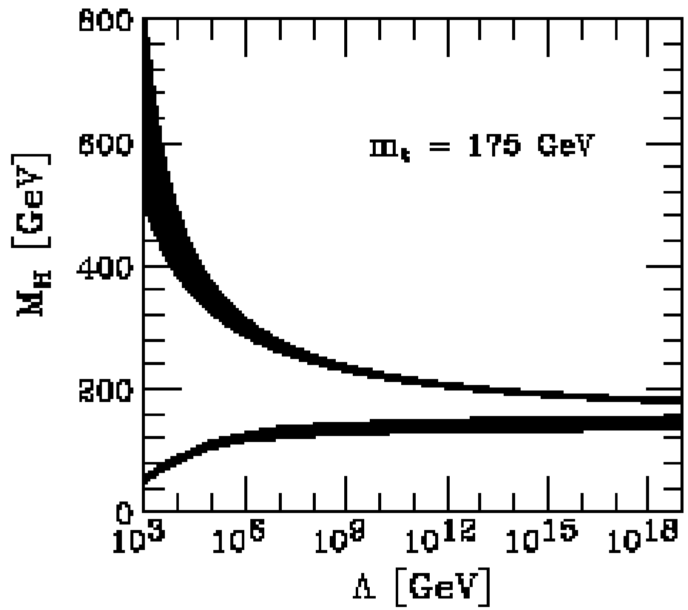
Search for the Higgs Boson



Direct searches
$M_H > 114.4$ GeV
@ 95% C.L.

Electroweak fit
$M_H < 219$ GeV
@ 95% C.L.

$$M_H^2 \leq \frac{8\pi^2 v^2}{3\log\frac{\Lambda_{QCD}^2}{v^2}} \qquad \upsilon^2 = 246 \text{ GeV}$$



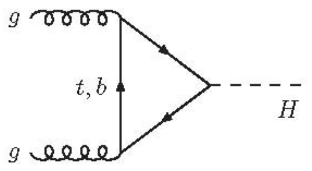
If no new physics up to Planck scale (~10^{19}GeV quantum gravity significant) small mass range for Higgs: $130 < M_H < 190$ GeV

Higgs Production

First: *understand signal*

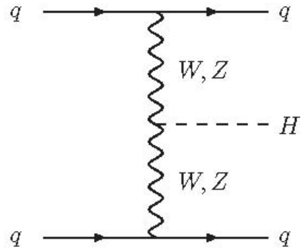
How is the Higgs produced?

gg fusion

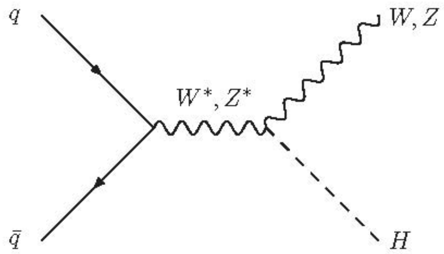


(a)

WW/ZZ fusion

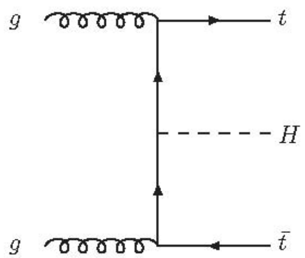


(b)



(c)

Associated WH, ZH



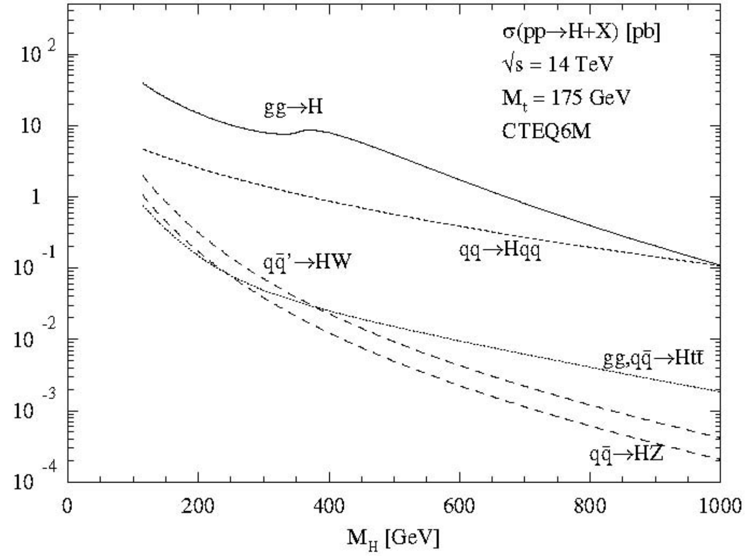
(d)

Associated ttH

Gluon fusion most promising

How often is it produced?

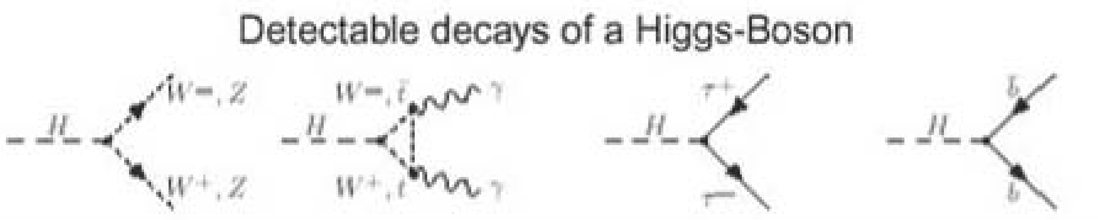
σ (pb)

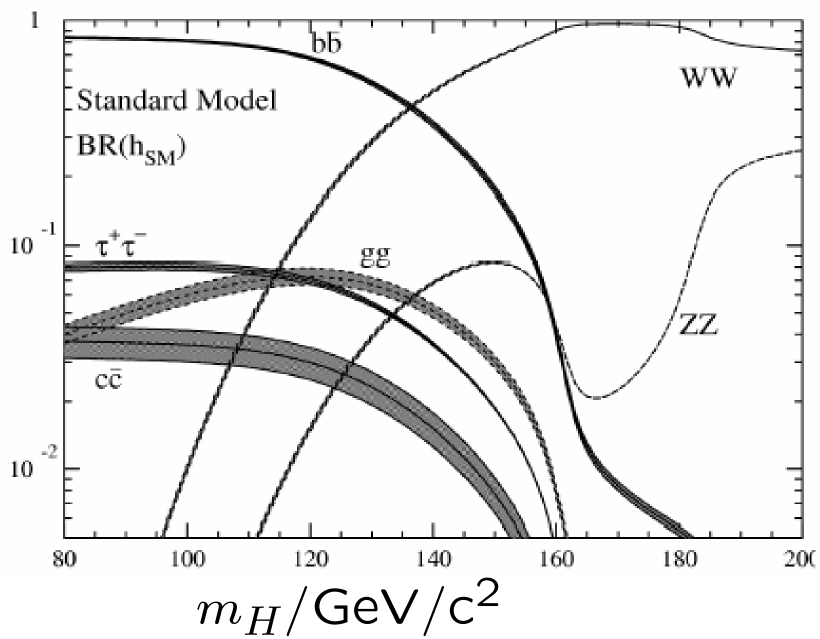


Process	Events/s	Events/year
$W \to e\nu$	40	$4 \cdot 10^8$
$Z \to ee$	4	$4 \cdot 10^7$
$t\bar{t}$	1.6	$1.6 \cdot 10^7$
$b\bar{b}$	10^6	10^{13}
$\tilde{g}\tilde{g}$ (m = 1 TeV)	0.002	$2 \cdot 10^4$
Higgs (m= 120 GeV)	0.08	$8 \cdot 10^5$
Higgs (m= 120 GeV)	0.08	$8 \cdot 10^5$
Higgs (m= 800 GeV)	0.001	10^4
QCD jets $p_T > 200$ GeV	10^2	10^9

$L = 2\times10^{33}\,cm^{-2}s^{-1}$

Higgs Decay



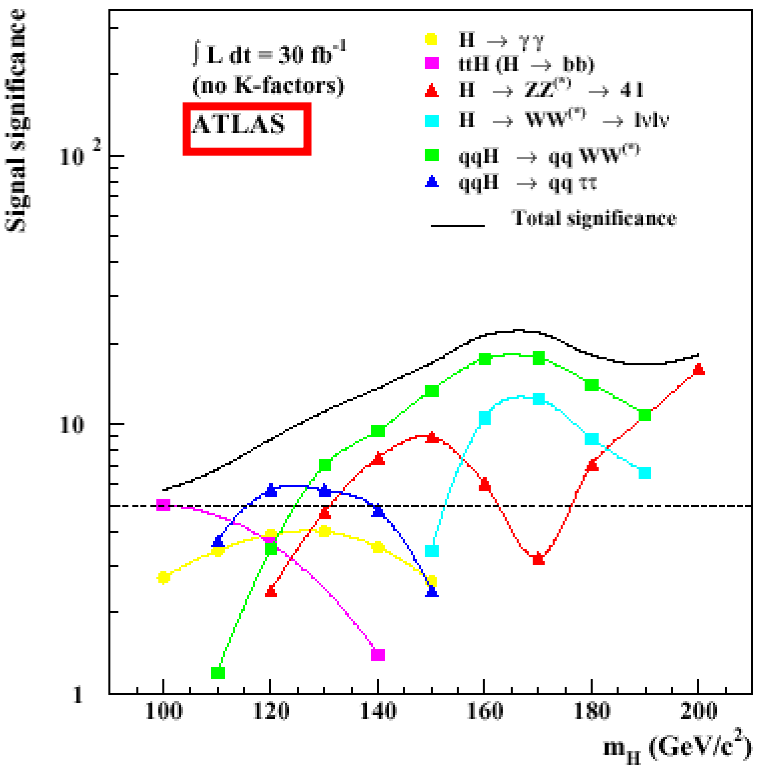


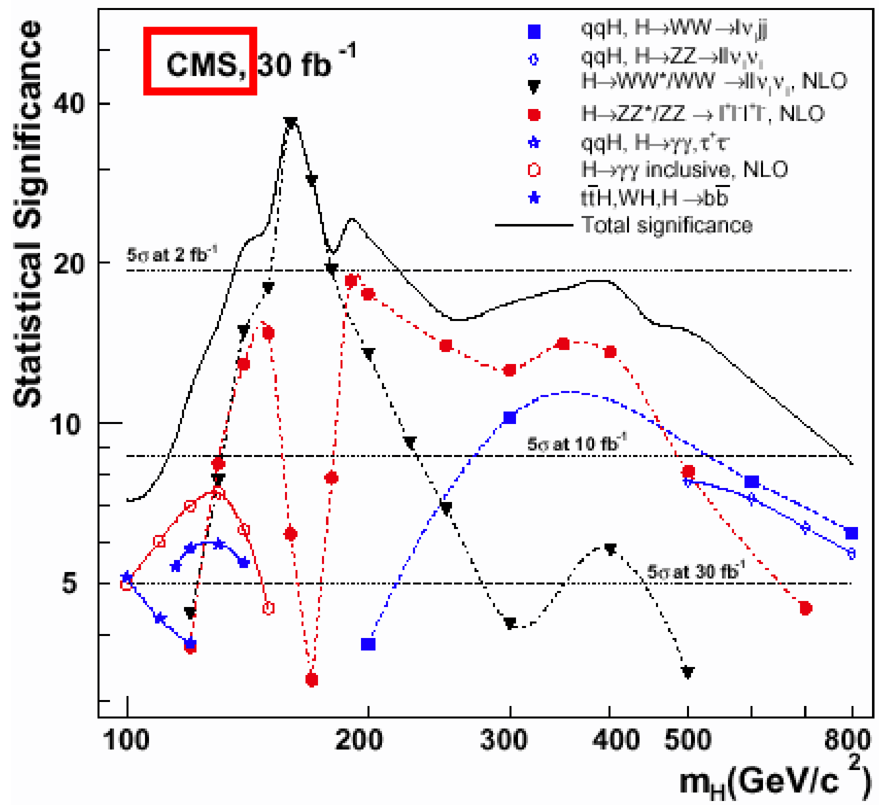
Branching Fraction: If produce 10^8 Higgs and measure only 20 decays H→gg with an efficiency of 0.00025% then Branching Fraction:

$$BF(Higgs \rightarrow gg) = \frac{N_{decays}}{N_H * \eta} = \frac{20}{10^8 * 2.5\times 10^{-6}} = 0.08$$

- Which decay to look at?
- Depends on Higgs Mass
  - M_H <150 GeV
    - gg → H→γγ
    - H→ZZ* →4l
    - gg → HW,Htt: H→bb
    - H→WW* →2l 2ν
    - qq→qqH : H →γγ, WW*, $\tau^+\tau^-$
  - M_H <500 GeV
    - H→ZZ →4l
  - M_H >500 GeV
    - H→ZZ,WW → jets

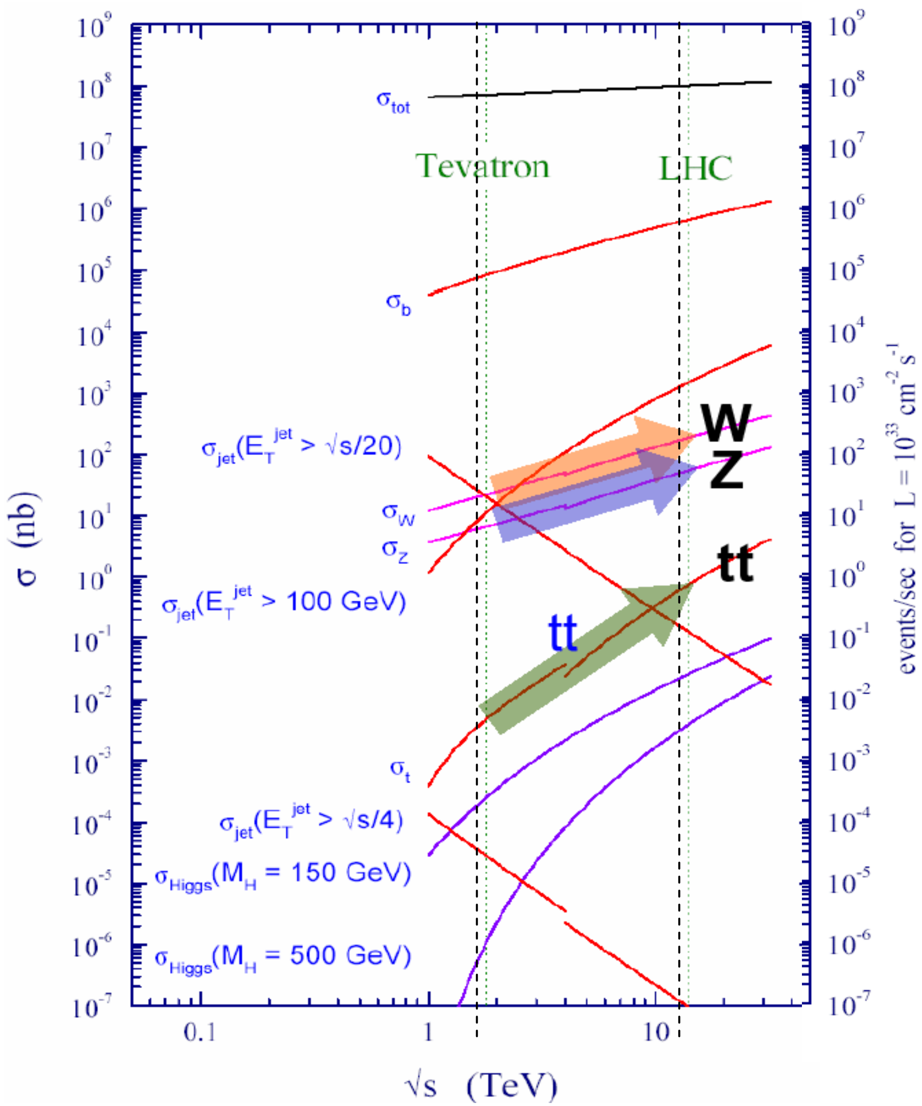
Best Modes to look at





Compare to list on previous slide

Backgrounds - Tevatron to the LHC



Huge stats for Standard Model signals. Rates @ $10^{33} cm^{-2} s^{-1}$

~10^9 events/10 fb^{-1} W (200 Hz)

~10^8 events/10 fb^{-1} Z (50 Hz)

~10^7 events/10 fb^{-1} tt (1 Hz)

(10 fb^{-1} = 1 year of LHC running at low luminosity $10^{33} cm^{-2} s^{-1}$, hence by ~end 2009)

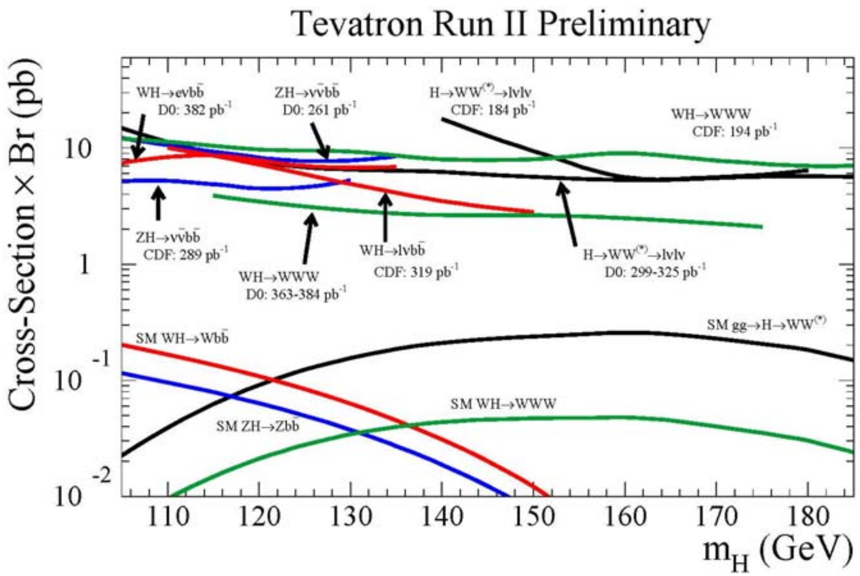
Background is anything with signature similar to signal

- *W+X* (*X* can be *W*, *Z* or just 2 QCD jets)
- *ZZ→ qql+l-* (one lepton not identified)
- $\tau^+\tau^-$
- *b*-tags can be real, charm or fakes

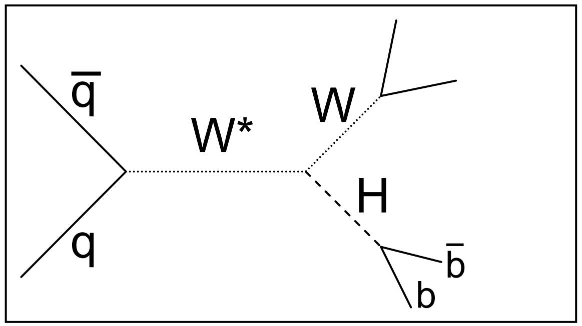
Current Results - Tevatron

- At 120 GeV $H \to bb$ dominates
- Signature $gg \to H \to bb$:
  - 2 jets
  - One or two *b*-tags
- Swamped by dijet production
  - *bb* ~ μb
  - *qq* ~ mb *(fake b-tag rate small but not zero)*
- Have to use *W/Z+H* channel (Associated Production)

Tevatron/CDF - Associated Production



qq→WH with *H→bb*



W→qq 70%

- final state *qqbb*
- Four jet backgrounds still too large

$W \rightarrow e\nu_e$ 10% $W \rightarrow \mu\nu_\mu$ 10%

- Final state *lvbb*
  - One electron or muon
  - Missing transverse momentum
  - Two jets
  - One or two *b*-tags
- Easy to select in trigger and offline

$\sigma \times Br \approx 0.02$ pb

Efficiency at the Tevatron/CDF

- Nature provides 20 fb of *WH→lvbb* events – *a handful per year*
- How many pass CDF trigger and analysis selection?
  - *Cleanly identified electron or muon in acceptance*
  - *Two jets*
  - *At least one b-tag*
  - *Large missing momentum*
  - *None overlapping*

  - Run thousands of MC events
  - Efficiency
  - Observe 2 per fb^{-1} per year

$$\epsilon = \frac{N_{selected}}{N_{generated}} \approx 10\%$$

How do we report this result?

Statistical

- Mostly counting events (data or MC)
- Poisson distribution: $\sigma = \sqrt{\mu} \approx \sqrt{N}$
  - *NB fractional error ~ $1/\sqrt{N}$*
- Efficiency follows binomial distribution: $\sigma_\epsilon = \sqrt{\epsilon(1-\epsilon)/N}$

Systematic

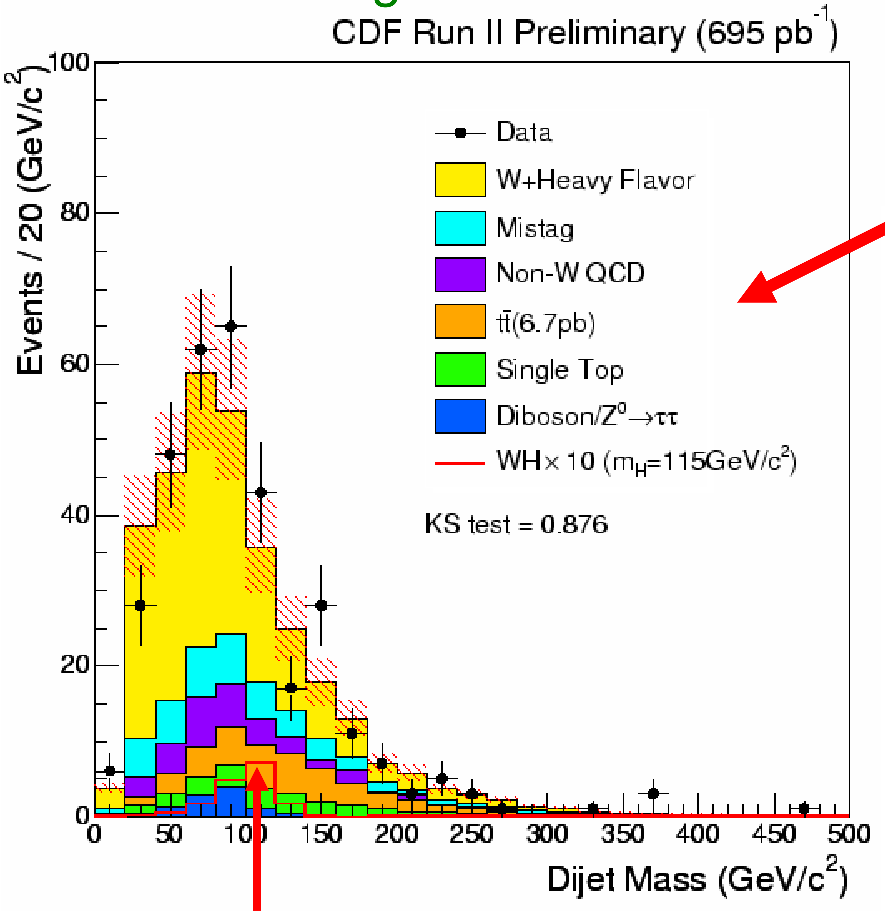
- Anything not completely understood may affect result
  - Detector performance, background rates, Monte Carlo modeling…
- Estimate range of parameter
- Vary in Monte Carlo

Significance

- In a given amount of data we expect:
  - N_B background events
  - Statistical error on background ≈ $\sqrt{N_B}$
  - Systematic error on background = σ_{sys}
  - Add errors in quadrature to get σ_{TOT}
- Observe $N(>N_B)$ events in data. Could be:
  - random fluctuation in $N_B \pm \sigma_{TOT}$ background events
  - N_B background events & N_S signal events
- Significance $S = N_S/\sigma_{TOT}$
  - S = 3: probability of fluctuation $\sim 10^{-3}$ – interesting...
  - S = 5: probability of fluctuation $\sim 10^{-5}$ – discovery!!

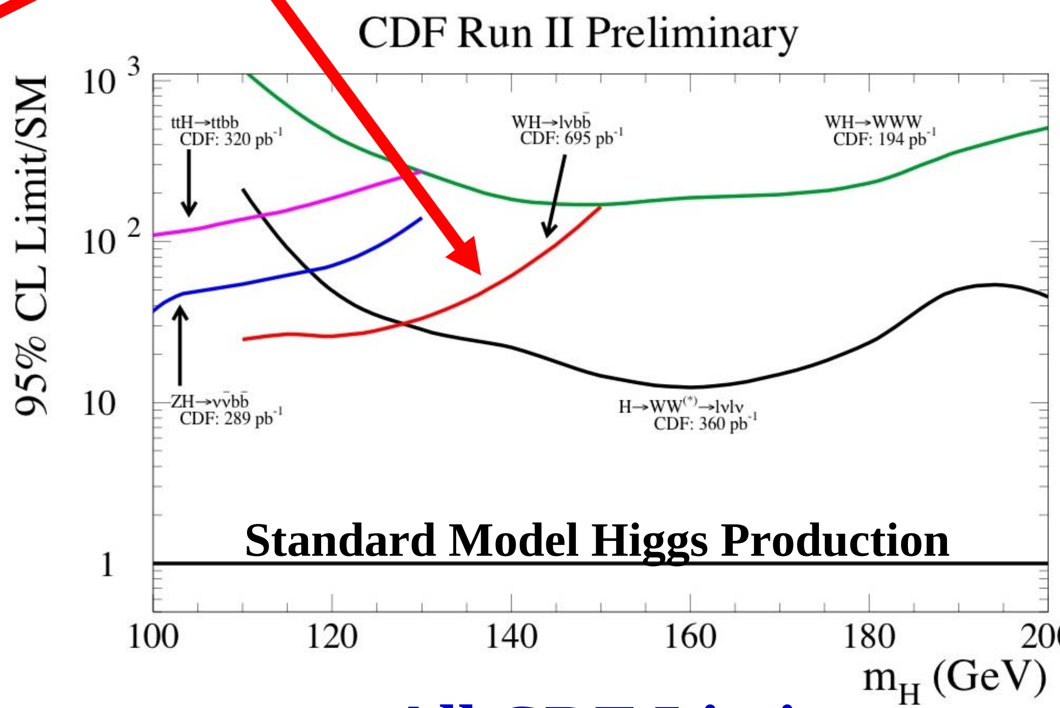
Latest CDF Higgs Results

- Data and background as function of *bb* mass



Expected signal × 10

$WH \to l\nu b\bar{b}$

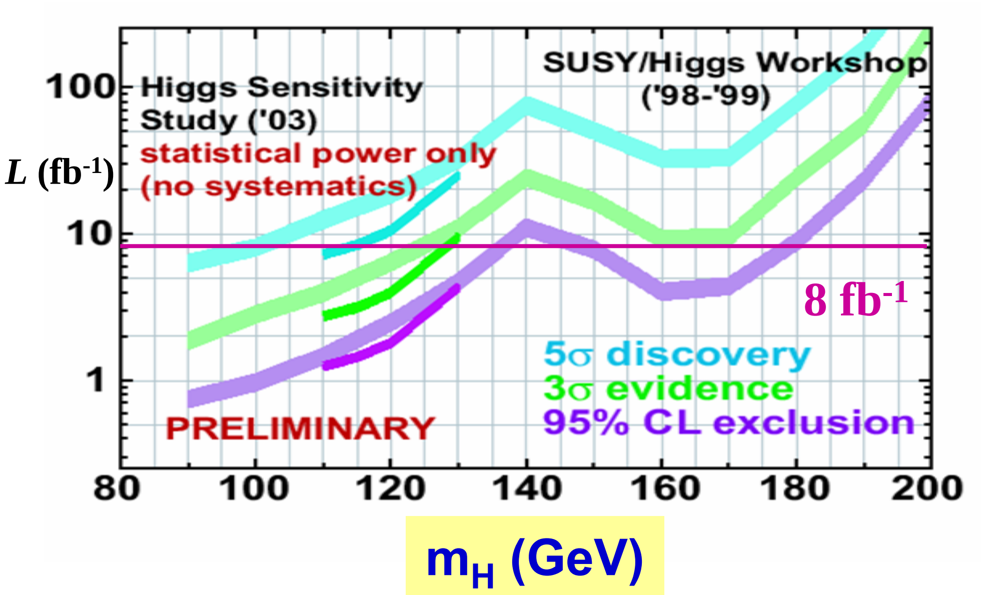


All CDF Limits

Predicted Sensitivity – Tevatron v LHC

Tevatron

- CDF expects a maximum of 8 fb^{-1} by 2009
  - 15-20 signal events
  - 2000 background
  - 8 years of running



LHC