Getting the LHC Physics Out

TASI 2006 - Lecture 2 - June 30, 2006
John Conway

Outline

A. The First Inverse Femtobarn
B. The Statistics of Discovery
C. Discovering the Higgs
D. Discovering SUSY
E. PGS and the LHC Olympics

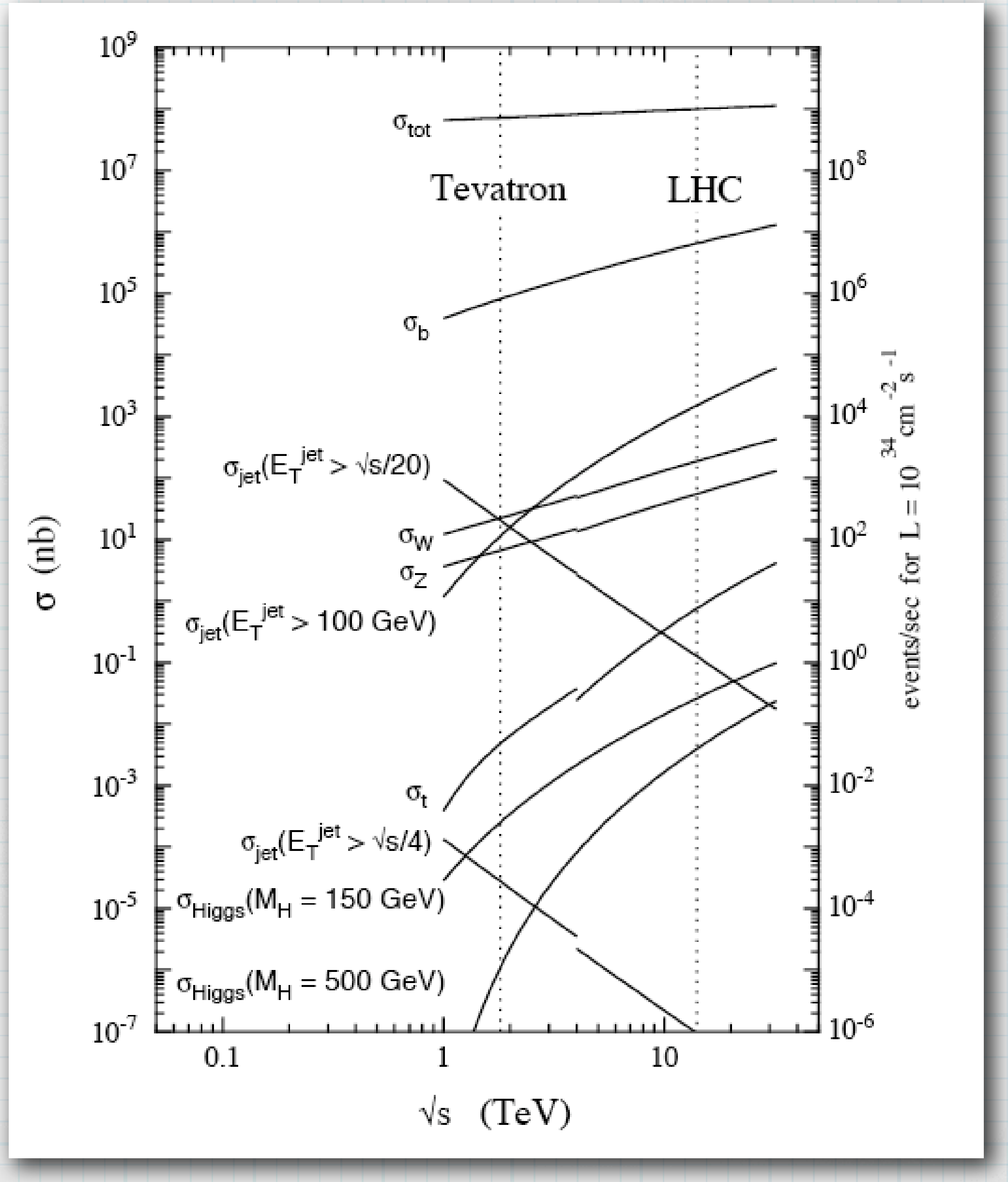
You give 100% for the first half of the game, and if that’s not enough, in the second half you give what’s left.

- Yogi Berra

The First Inverse Femtobarn

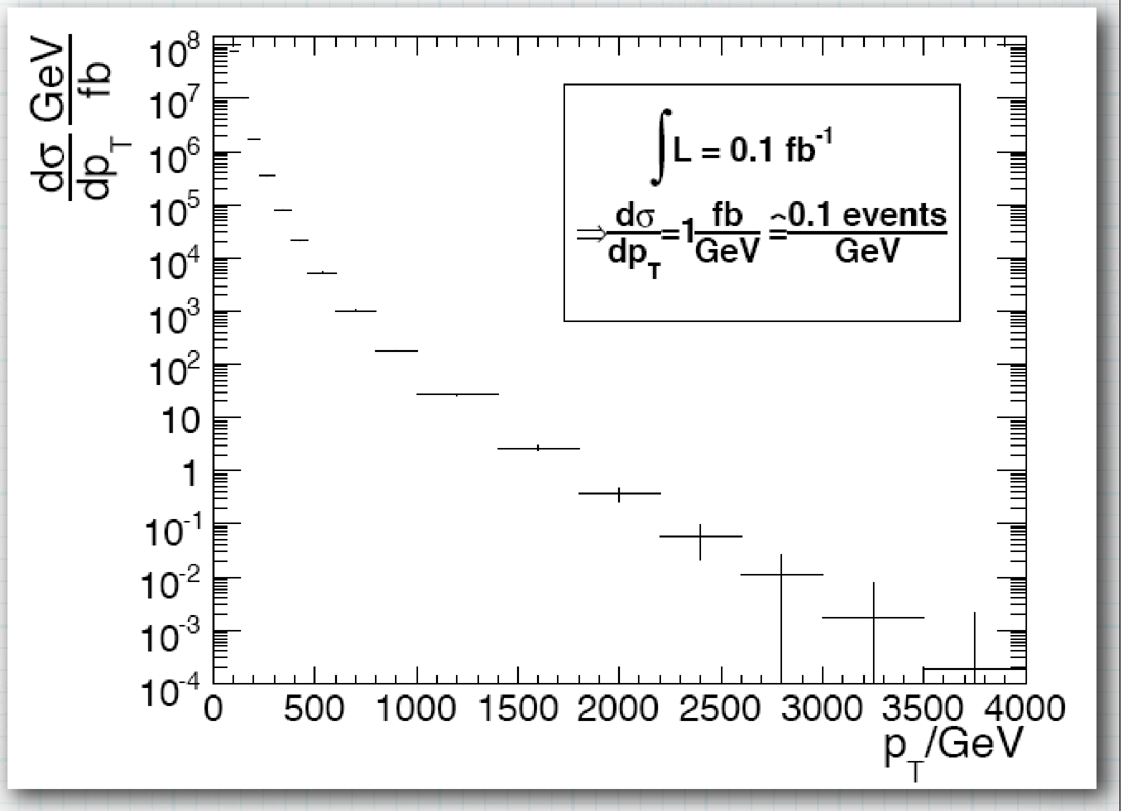
- 2008 - the first real physics running!
- low $\mathcal{L}$ at first...
- dijet cross section is enormous
- gamma + jet is huge
- $\sigma(Z) \sim 60$ nb
  $B(Z \to \ell\ell) = 3.6\%$

$N(Z \to \ell\ell) = 2.2 \times 10^{6}/\mathrm{fb}^{-1}$



Dijets

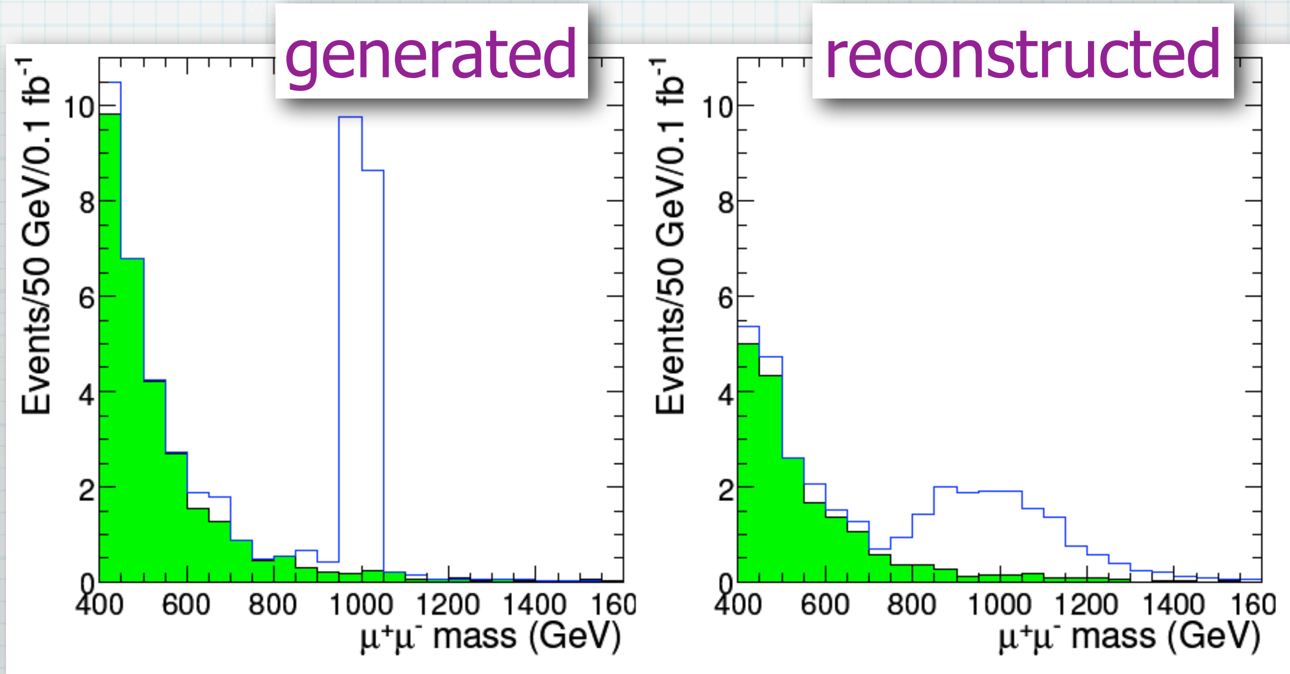
- expect hundreds of > 1 TeV events even with just $0.1\ \mathrm{fb}^{-1}$
- can use these to calibrate jet reconstruction methods
- need to tune simulation to match jet propoerties
- good place to look for new physics in the early days!



Z′s and W′s

- Z is actually Z/γ* from Drell Yan $q\bar{q}$ annihilation
- get large Z peak plus continuum out to very high invariant masses
- excellent place to look for new physics early on...if it’s something striking

Z′ signal in CMS with 0.1 fb^{-1}



Calibration

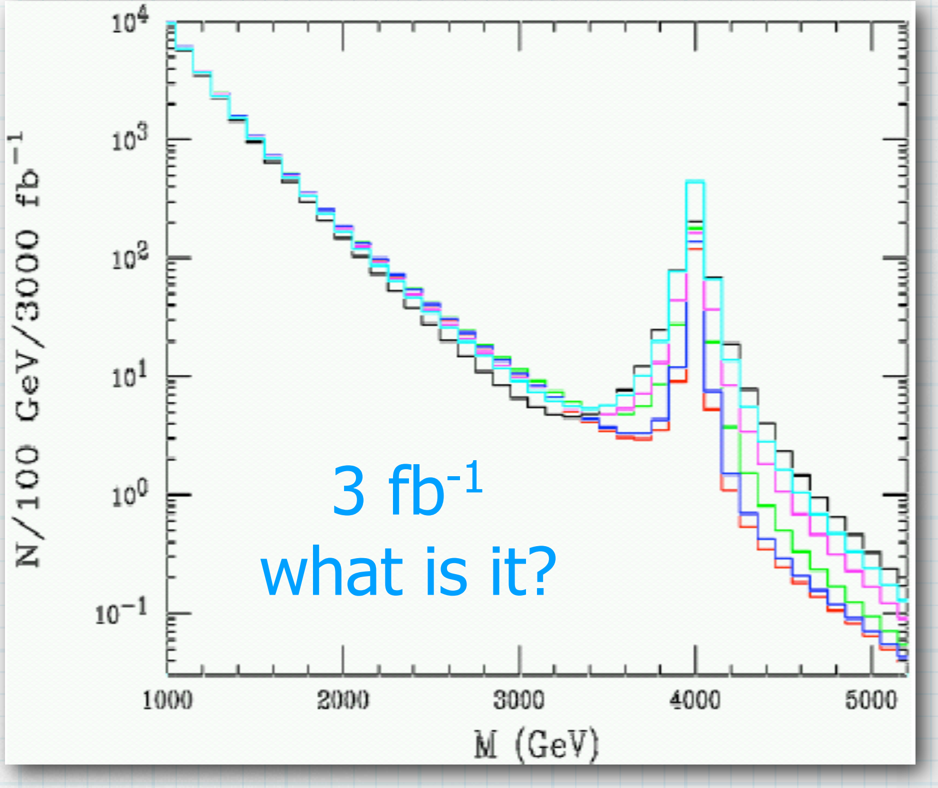
- early work: align tracking system using copious tracks from multijet events (20k parameters!)
- early work: do basic tower-to-tower calibrations of calorimeter cells using jets, electrons, etc.
- use $Z \rightarrow \mu\mu$ events to calibrate tracking/magnetic field, etc.
- use $Z \rightarrow ee$ events to calibrate e.m. section of calorimeter
- use jet-jet and γ-jet balancing to calibrate whole calorimeter
- tune missing energy (p_T) corrections

Calibration

- the main work of 2008 will be to perform this sort of calibration!
- we need to “re-discover” the Standard Model
    - dijet spectra
    - jet multiplicity
    - Z, W
    - ttbar, diboson, ...
- it is difficult to imagine making discoveries except for the most striking phenomena such as a huge resonance in $\ell\ell$ or dijets....still...??

Possible Early Discovery?

- suppose the Randall-Sundrum extra dimensions picture is true, and the effective Planck mass is low
- or perhaps a heavy Z′ ?
- might observe huge resonances in dijet spectra
- would be good to catalog all the possible early discovery scenarios, yes?



3 fb^{-1}
what is it?

The Statistics of Discovery

The challenge:

Take several analyses which yield multivariate spectra of observed quantities, with multiple imperfectly understood background sources, and combine them together to obtain a quantitative measure of the level of significance for the presence of some (possibly unknown) new signal process, or, in the absence of a signal, determine quantitative exclusion bounds on the possibility of its existence, in one or more dimensions.

Whew!

The Statistics of Discovery

- The field has labored for many years to refine the statistical techniques for
  - signal-background separation (selection)
  - combining results
  - incorporating systematic uncertainty
  - interpreting the physics

- PHYSTAT series of workshops: emerging sub-specialty within experimental particle physics

Bayesianism/Frequentism

- Frequentist: only talk about the probability of the observation given the (physics) parameter:

$$P(x|\alpha)$$

- Bayesian: only talk about the probability that the parameter has some value, given the observation:

$$P(\alpha|x)$$

"Bayesians address the question everyone is interested in, by using assumptions no-one believes, and frequentists use impeccable logic to deal with an issue of no interest to anyone" - L. Lyons, SCMA IV, June 2006

Quoting Results

- We tend to quote three types of results:
  1. Limit (exclusion bound)
  2. Discovery (significance)
  3. Measurement (value with uncertainty)
- The field has not yet come to a set of standards for calculating all three types of results with a uniform method.
- Physicists tend toward frequentism, but think like Bayesians more than they realize!

The Simplest Limit

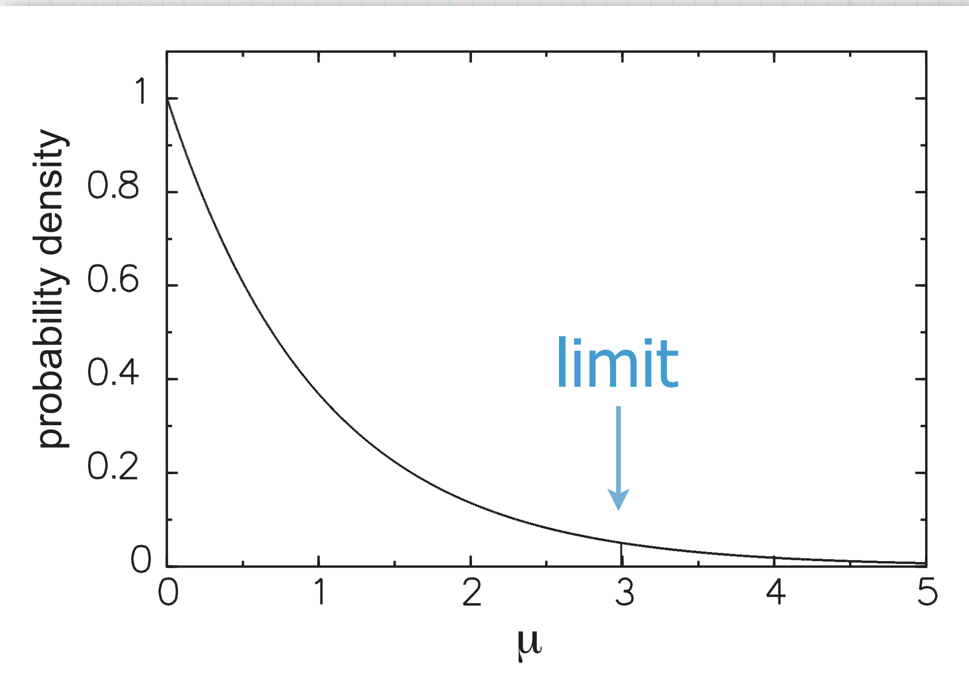
- suppose we search for a new particle X
- we have a striking signature: no background!
- we run the experiment and observe zero events
- at what level can we say “X does not exist” ?
- the problem: X can exist, but there is a possibility that we don’t see it due to statistical fluctuation

Poisson distribution:

$$P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!} \Rightarrow P(0|\mu) = e^{-\mu}$$

The Simplest Limit

- Frequentist says "there is a 5% chance that, if the true value of μ is 2.996, that we would have observed 0 events, therefore with 95% confidence we can say that the value of μ must be 2.996 or less" and from that obtain a cross section limit, given the luminosity and acceptance.
- Bayesian says "probability density for the true value of μ looks like this" and therefore 0<μ<2.996 with 95% probability; obtain interval for cross section as in frequentist case.
- same result!



Bayes' Theorem

- Rev. Bayes was trying to prove the existence of God mathematically...late 1700's
- Bayes' Theorem reads

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- for us, with n events given μ expected,

$$P(\mu|n) = \frac{P(n|\mu)P(\mu)}{P(n)}$$

"prior" prob. of μ

"get some n"

Two Fun Examples

- “99% Accurate AIDS Test”

- particle beam type identifier

Coverage

- if we say “we exclude a new particle X with 95% confidence” then what is 95% of what?
- incorrect: “there is less than a 5% chance the particle X exists”
- correct: “if particle X exists, there is less than a 5% chance that we would have observed what we did”
- coverage: the fraction of cases where X does exist and we falsely exclude it

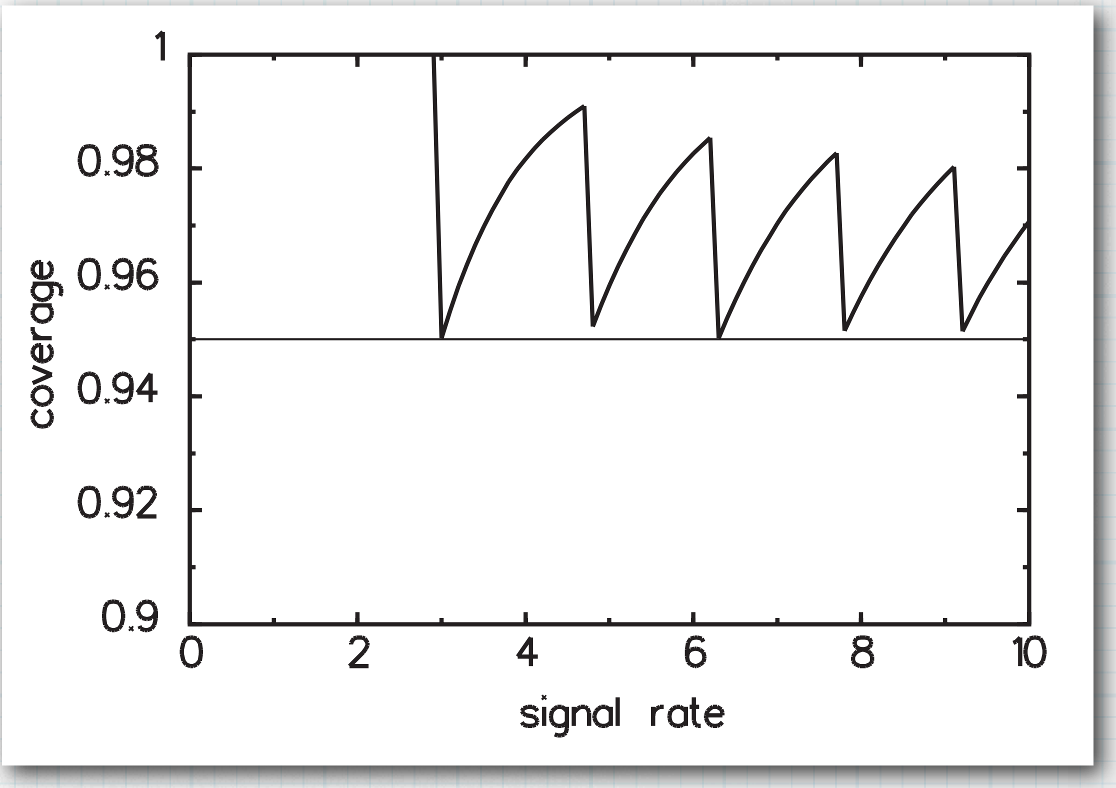
Example of Coverage

- let's consider a range of possible (true, unknown) rates for the number of events we expect from new particle X, from 0 to 10
- for each true rate we consider all possible experimental outcomes (n observed events)
- given n observed events, a frequentist will exclude at 95% CL any μ such that would give us n or fewer events with 5% or less probability:

$$0.05 < \sum_{k=0}^{n} \frac{\mu^k e^{-\mu}}{k!}$$

“Dinosaur” Plot of Coverage

- here we plot the fraction of cases where the 95% CL interval contains the true signal rate



- how do we explain the strange jumps in this plot?
- method “overcovers” for any value of the signal rate
- overcoverage is desirable (want 95% to mean at least 95%)

Significance

- basically we answer the question “what were the chances of that happening if there isn’t something new going on”
- probability of the null hypothesis: a “P-value”
- example: see 20 events, expect 3.2±1.0 from background only
- equate to gaussian probability to quote how many “sigma” the discovery is
- we usually demand 5σ ($P \sim 3 \times 10^{-5}$, or 1 in 33k)
- but what about systematics?

Likelihood

- Bayesians and frequentists use likelihoods
- a likelihood is a quantity proportional to the probability for something to occur
- suppose we have an observed spectrum:

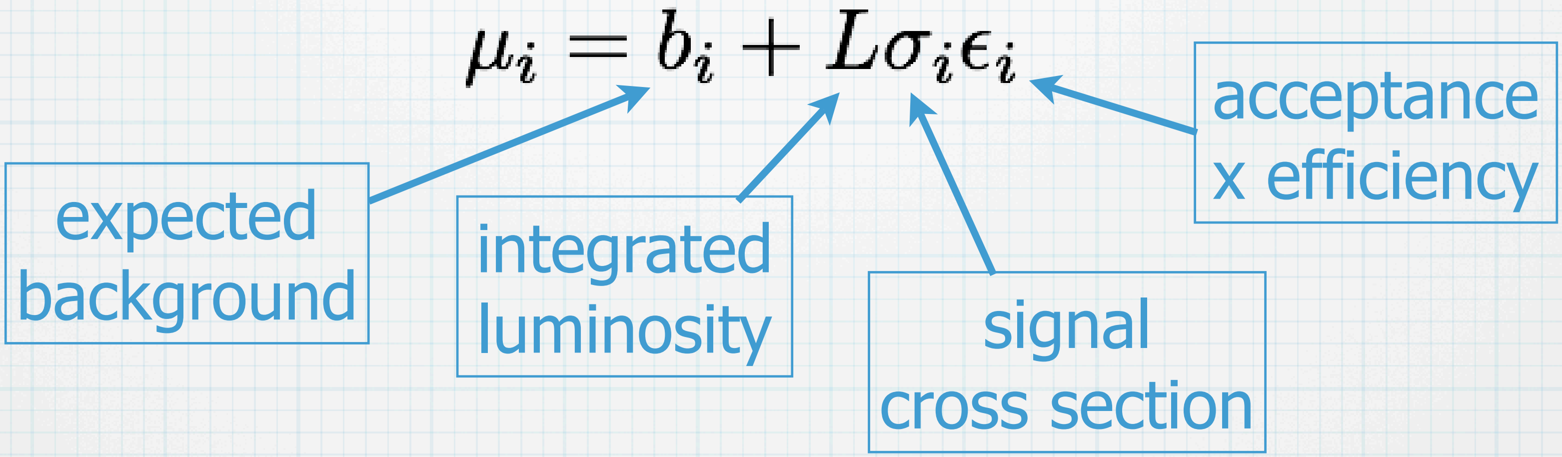
  obs: $\{y_i, i=1,\ldots,n_{bin}\}$ exp: $\{\mu_i, i=1,\ldots,n_{bin}\}$

- likelihood is joint probability of all the observations, which are treated as independent:

$$\mathcal{L} = \prod_{i=1}^{n_{bin}} P(y_i|\mu_i) = \prod_{i=1}^{n_{bin}} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}$$

Likelihood

- the μ_i are in general a sum of background(s) plus any new signal:



$$\mu_i = b_i + L\sigma_i\epsilon_i$$

expected background

integrated luminosity

signal cross section

acceptance x efficiency

- none of these things are known with certainty
- (for a frequentist this is a big problem, since they can only make statements about the observation, given the true value)

Likelihoods

- normalization of likelihood
- log likelihood and chi square
- minimizing likelihoods

Systematic Uncertainties

- how do we handle systematic uncertainties on efficiency (signal rate) or background?
- in general, presence of uncertainty should make limits worse (less stringent), discovery significances less so, and uncertainties on measurements larger
- not all methods guarantee this, surprisingly!
- Frequentist: there is only one true value for the background, or efficiency, and we do not know it
- Bayesian: we treat unknown parameters as uninteresting "nuisance" parameters
- nearly all modern treatments of systematics: Bayesian

Nuisance Parameters

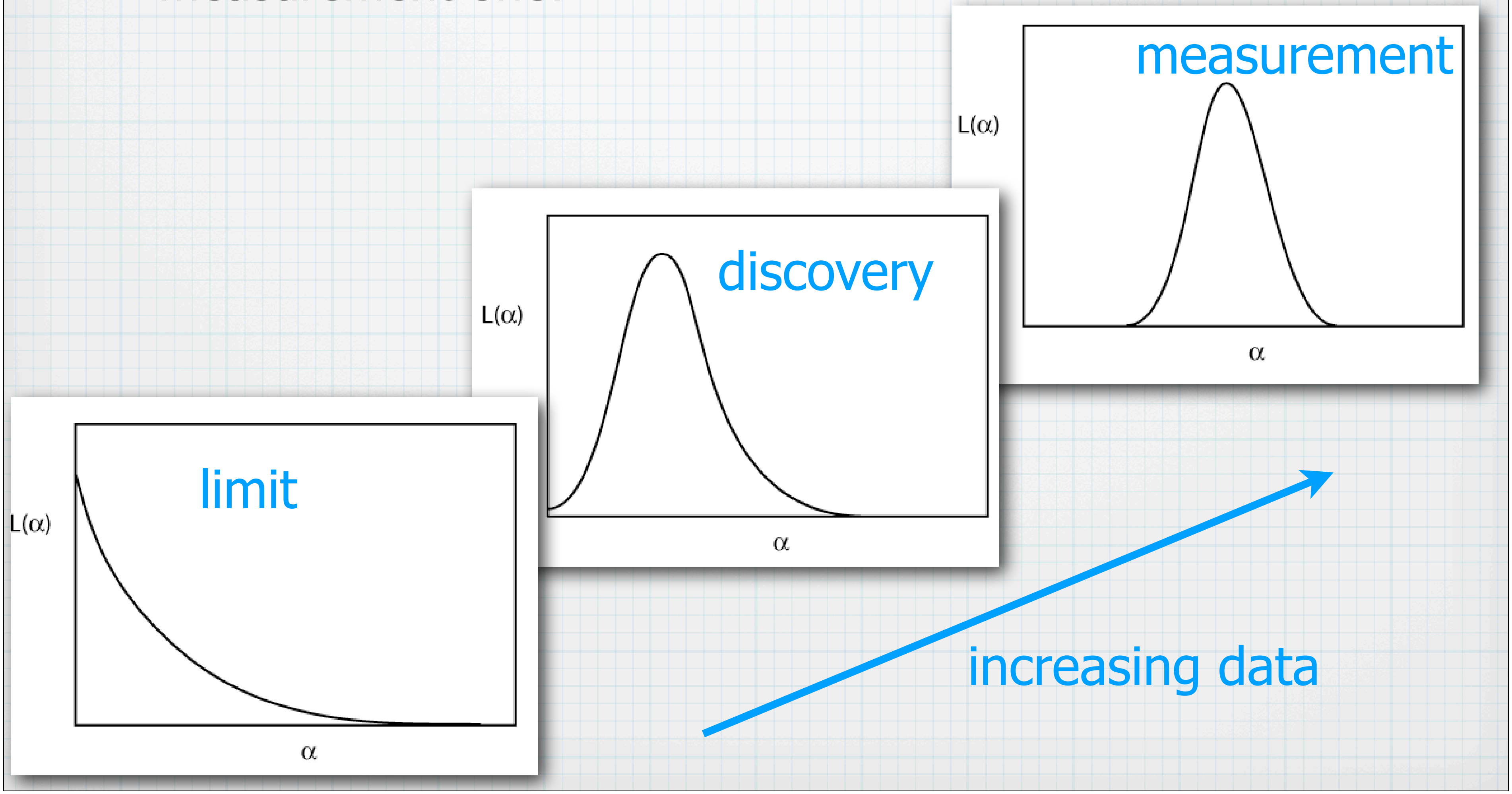
- we most often treat nuisance parameters by marginalization; we integrate them away
- example: single channel with unknown background but perfectly known signal rate gives likelihood

$$\mathcal{L}(n|s,b,\sigma_b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \int_0^\infty \frac{(s+b')^n e^{-(s+b')}}{n!} e^{-(b-b')^2/2\sigma_b^2} db'$$

- straightforward application of Bayes' Theorem!
- often called "smearing the likelihood" or "convoluting likelihood with gaussian"
- nearly equivalent to maximizing L w.r.t. nuisance parameters

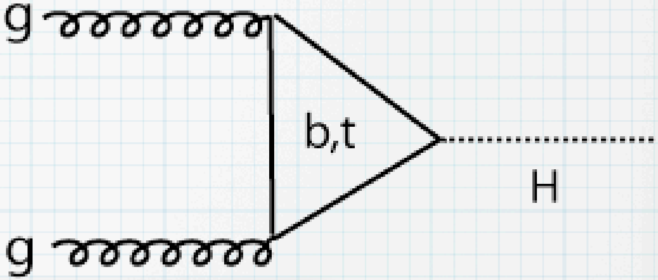
Limits→Discoveries→Measurements

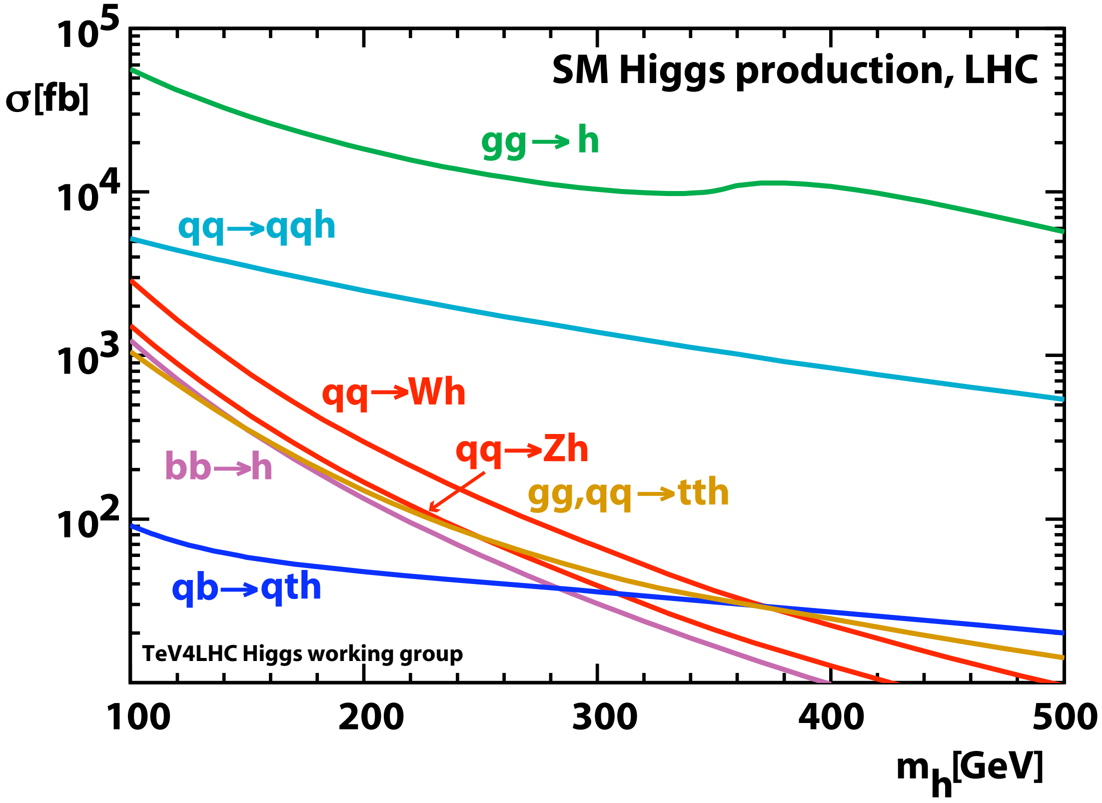
- as we gather more data, the likelihood evolves from an exponentially falling limit-type one to a gaussian-type measurement one:



Discovering the Higgs

- most copious source of SM Higgs bosons at the LHC: gluon fusion



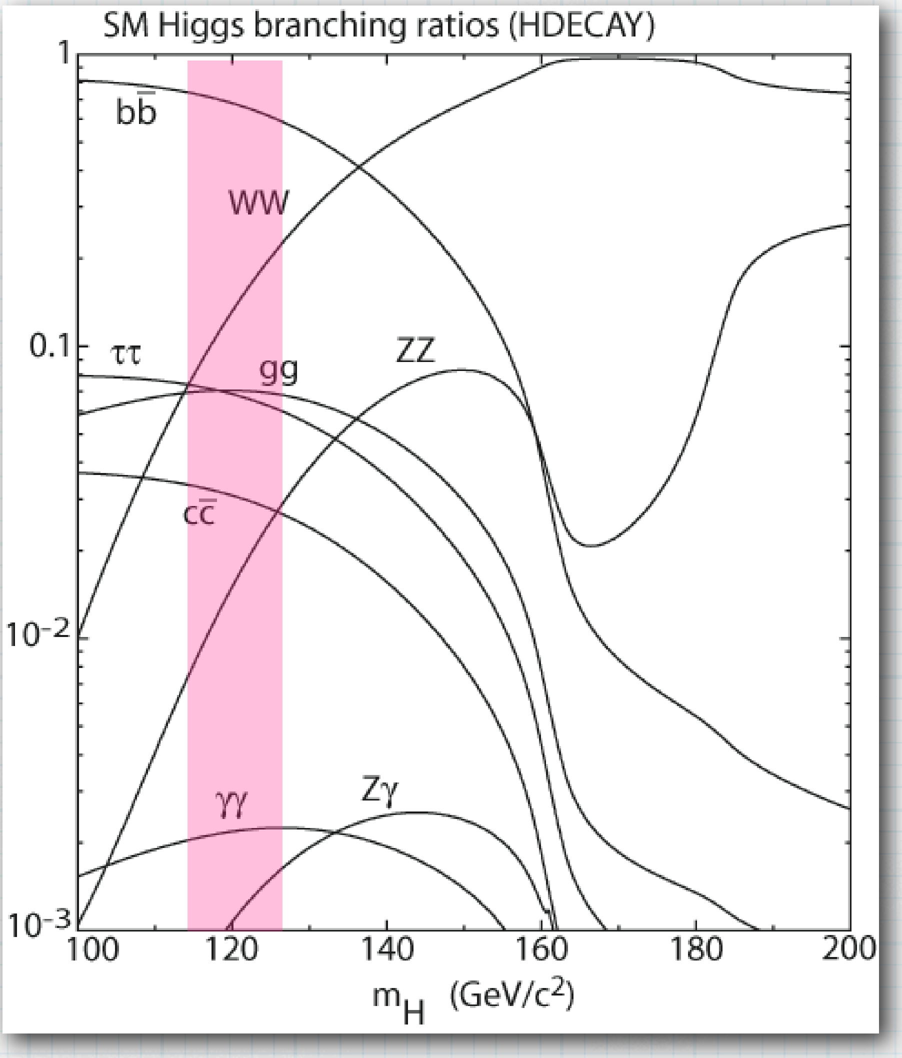


SM Higgs Branching Ratios

- you might not have thought γγ was so important:

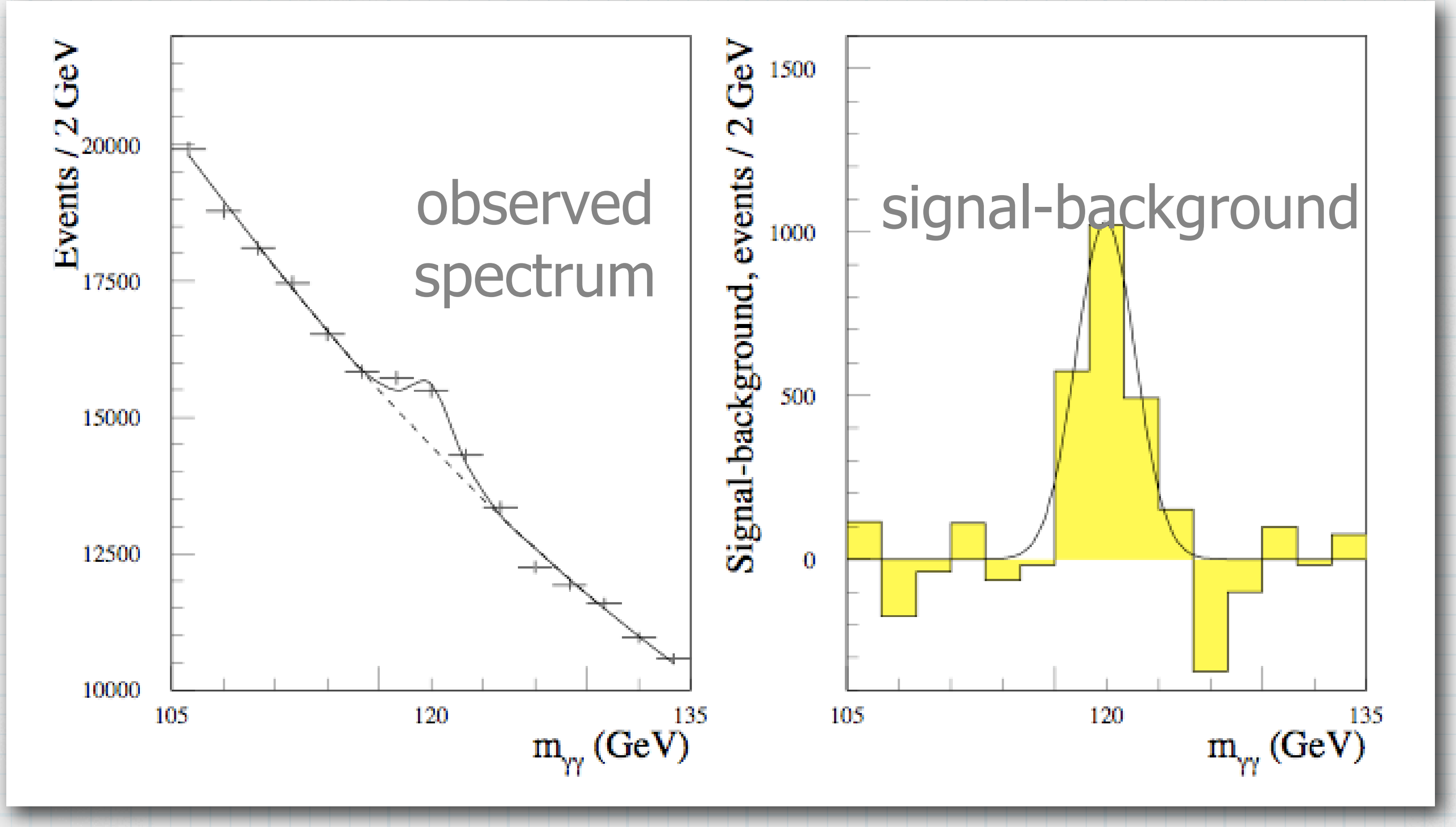
  B(H→γγ) ~ 2 x 10^{-3}

- bb mode nearly impossible given huge dijet rate
- gg, cc even worse
- what about ττ ? (Z→ττ)
- lots of random combinatoric background for γγ mode (π^0's)
- e.m. cal resolution!



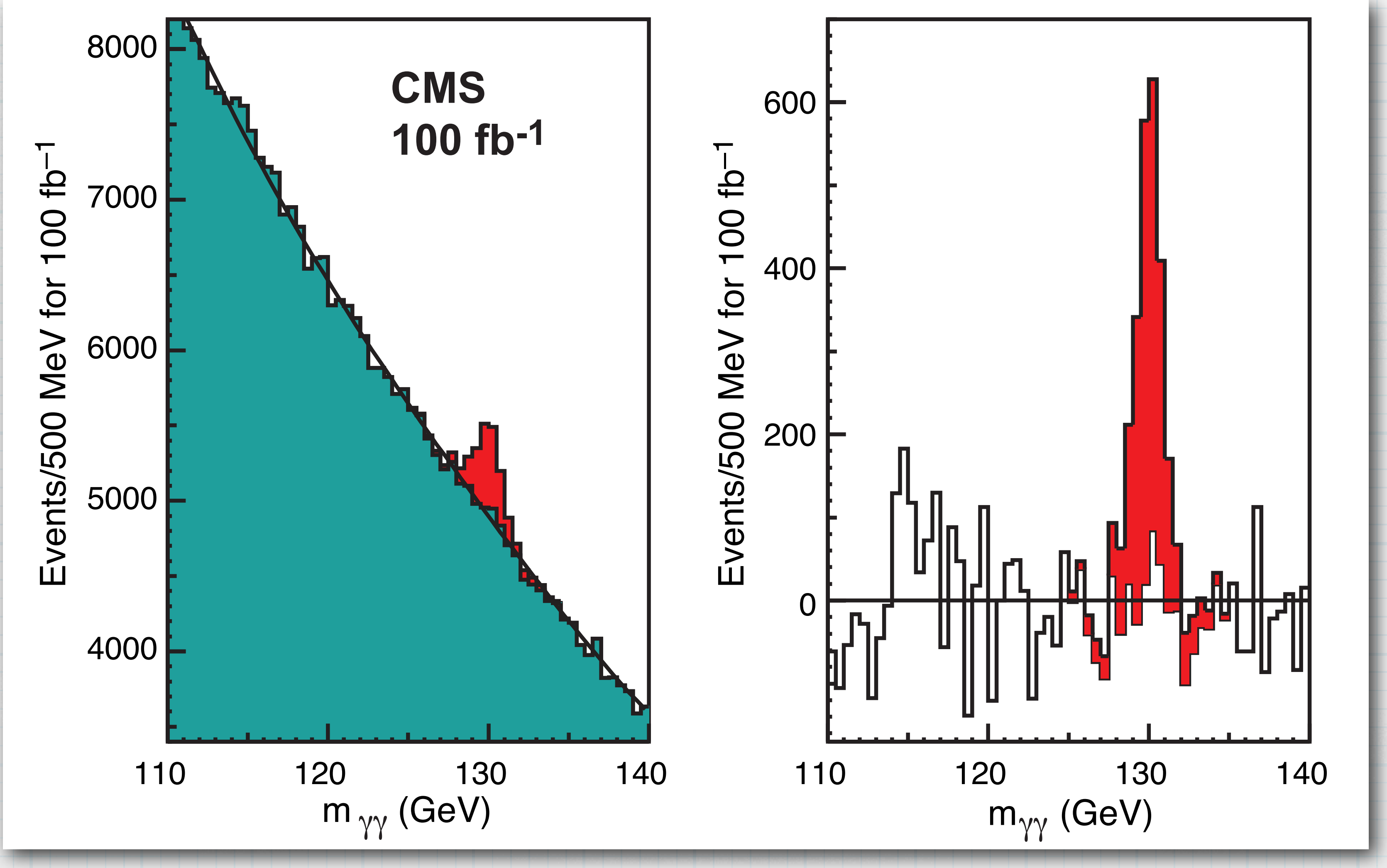
SM H→γγ

- ATLAS analysis: expected result after 100 fb^{-1}



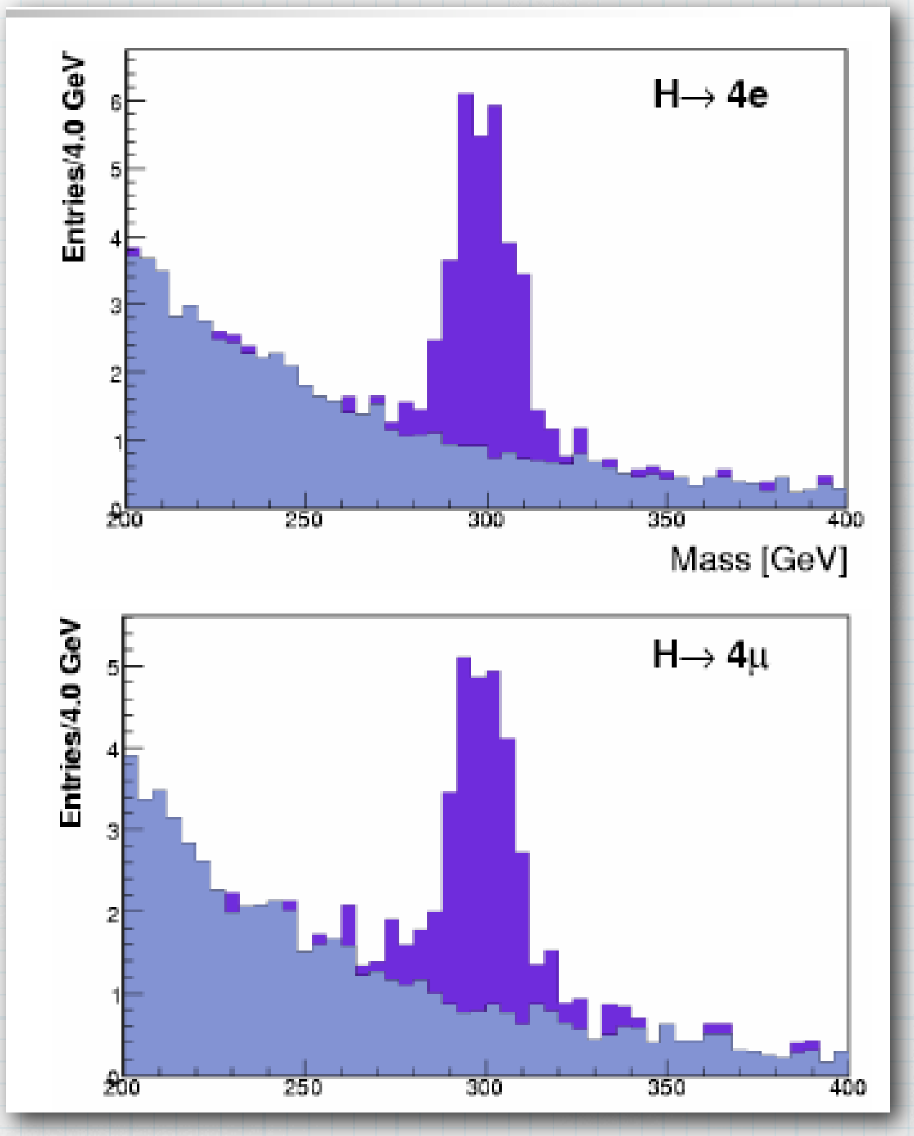
SM H→γγ

- CMS analysis: does better than ATLAS...

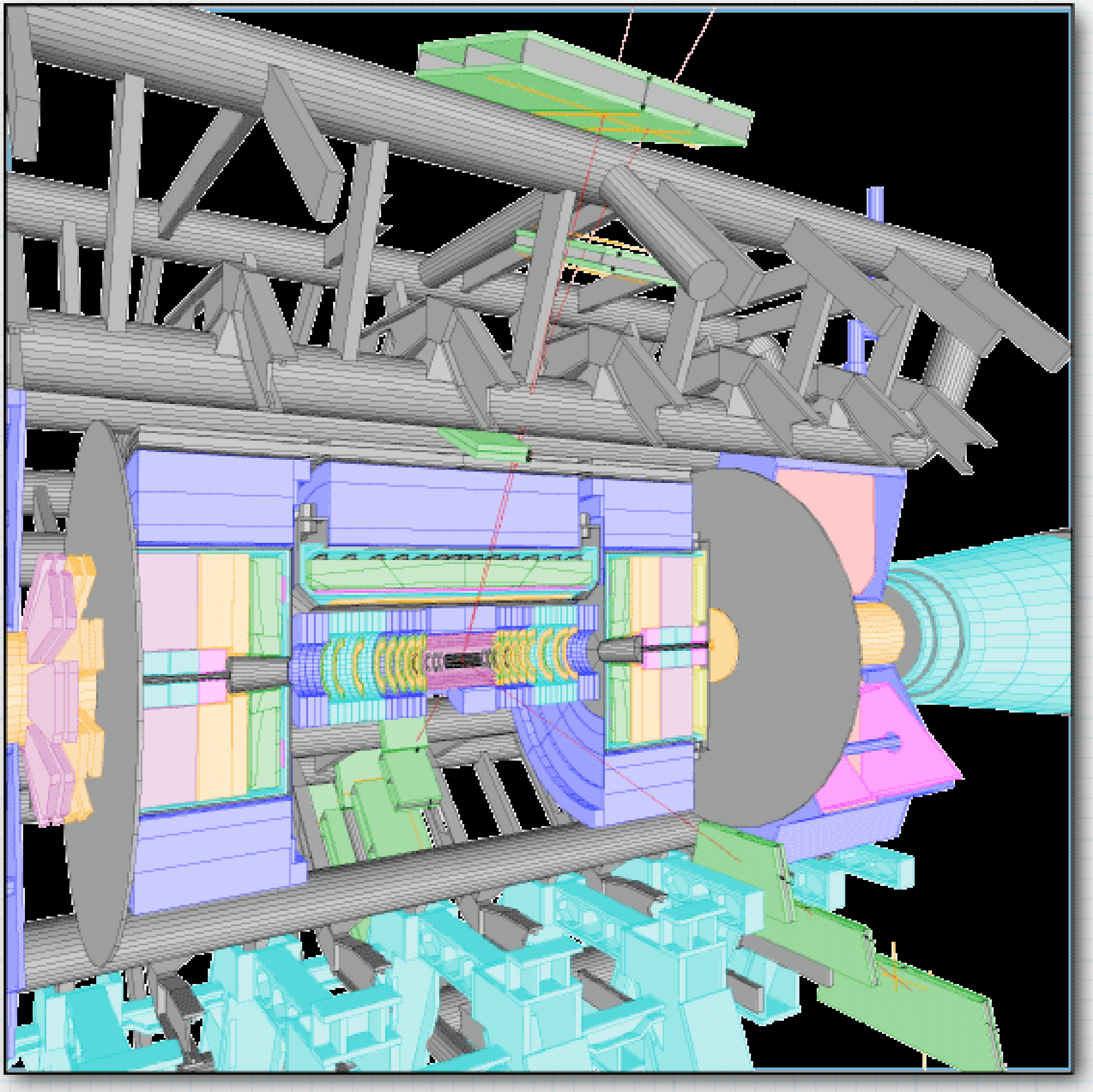


Heavy H→ZZ→4ℓ

- super-clean golden mode for higgs discovery at the LHC ... if the Higgs is very massive
- BR ~ 10^{-3}
- but get about 10^3 Higgs events per day!



ATLAS H→4ℓ Event



Tau: the new b ?

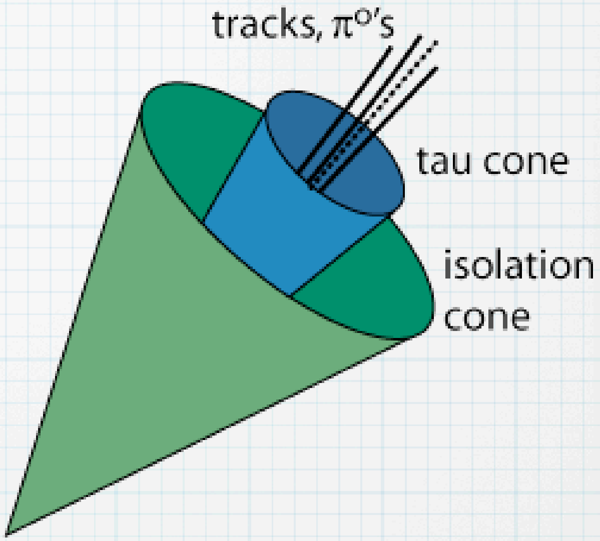
- tau decays weakly
- always get a ν_τ
- 35% of the time we get $e\nu\nu$ or $\mu\nu\nu$
- 65% of the time we get one or three charged hadrons
- can reconstruct taus and distinguish from hadronic jets

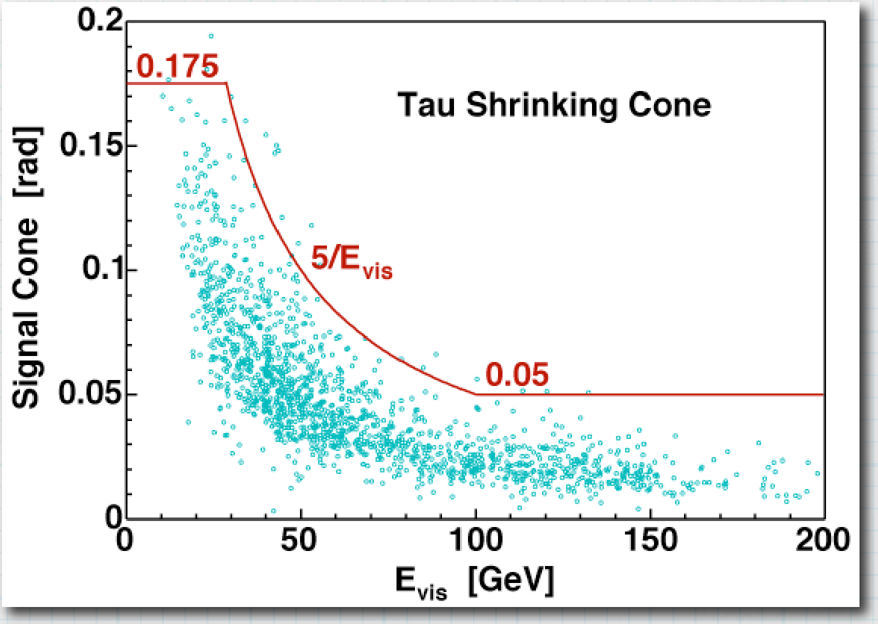
$e^- \nu\nu$	17.8%
$\mu^- \nu\nu$	17.4%
$h^- \nu$	49%
$\pi^- \nu$	11%
$K^- \nu$	0.7%
$\rho^- \nu$	25.4%
$h^+ h^- h^- \nu$	15%

tau is the most massive lepton...must tell us about the Higgs

Tau Identification

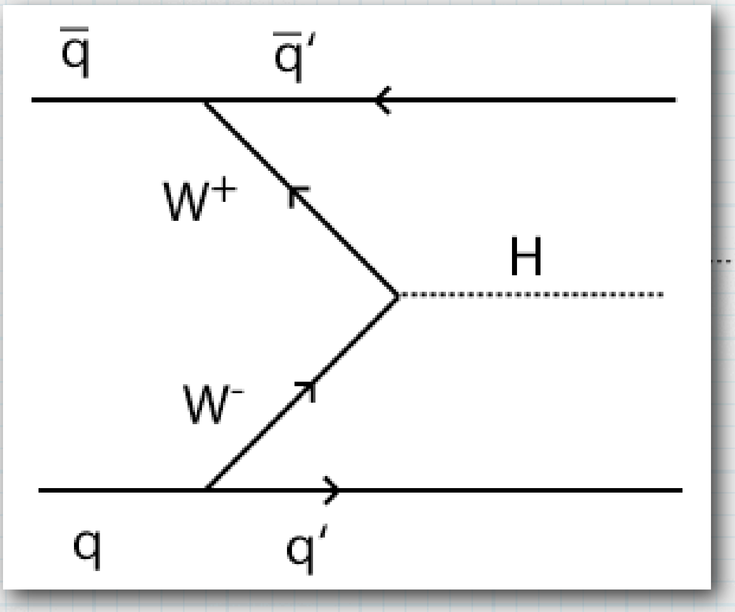
- at the Tevatron we have developed sophisticated tau identification
- use tracks to form cone
- define tau region and isolation region
- demand tracks and pi0’s lie in cone
- big challenge: hadronic jets fluctuate to look like taus

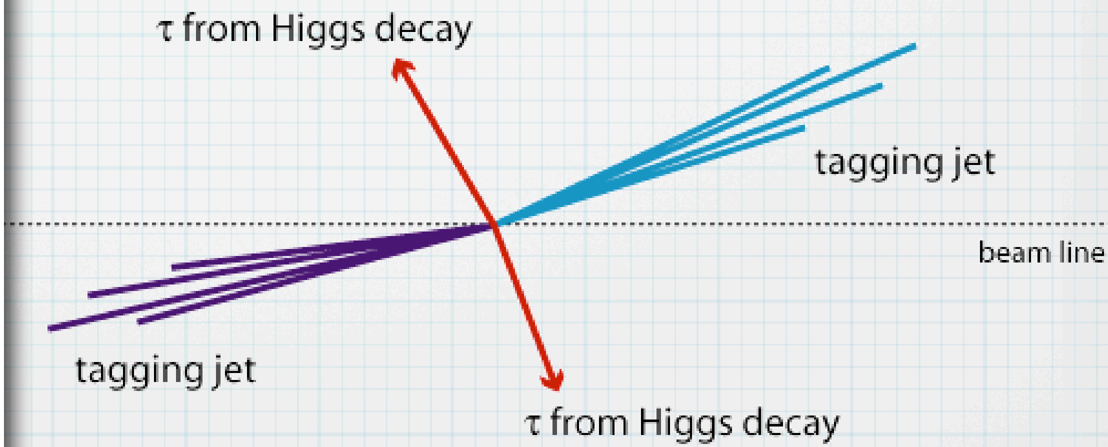




SM Higgs with Taus

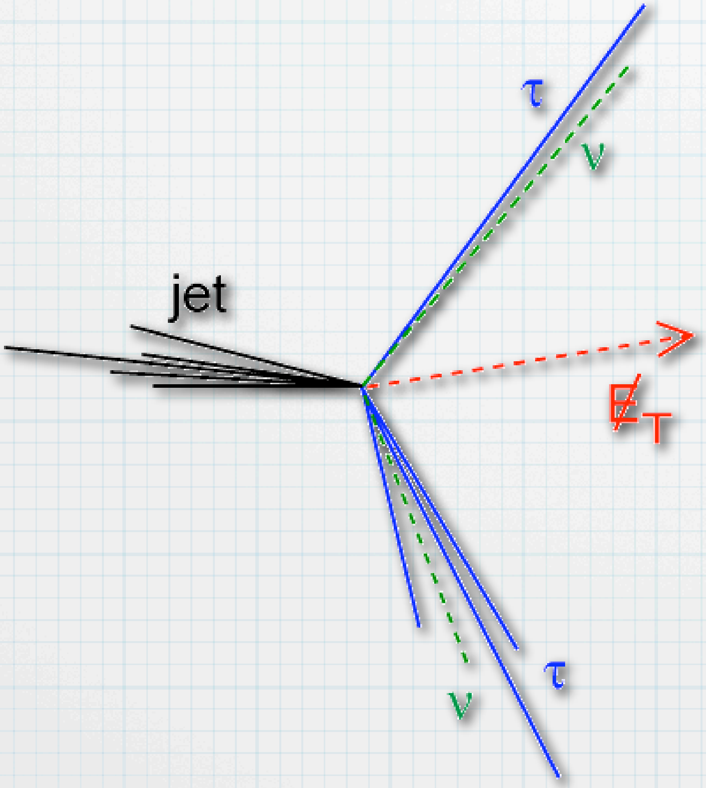
- vector boson fusion can produce Higgs, plus two forward “tagging” jets
- main background: VBF to Z
- need to trigger: probably require one e or μ

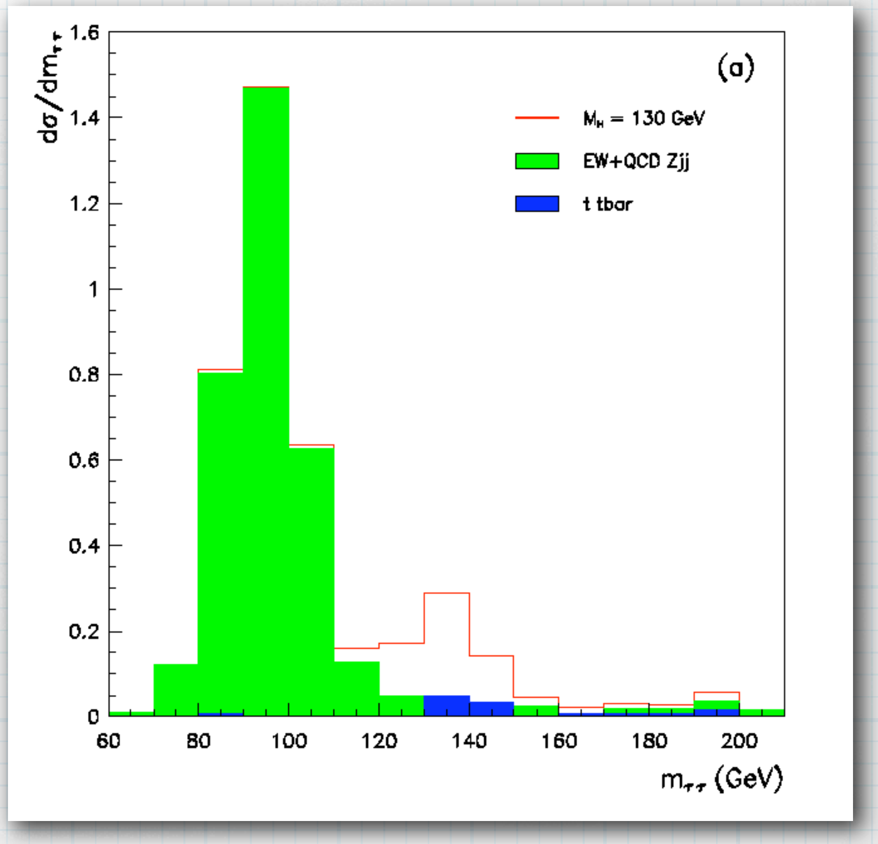




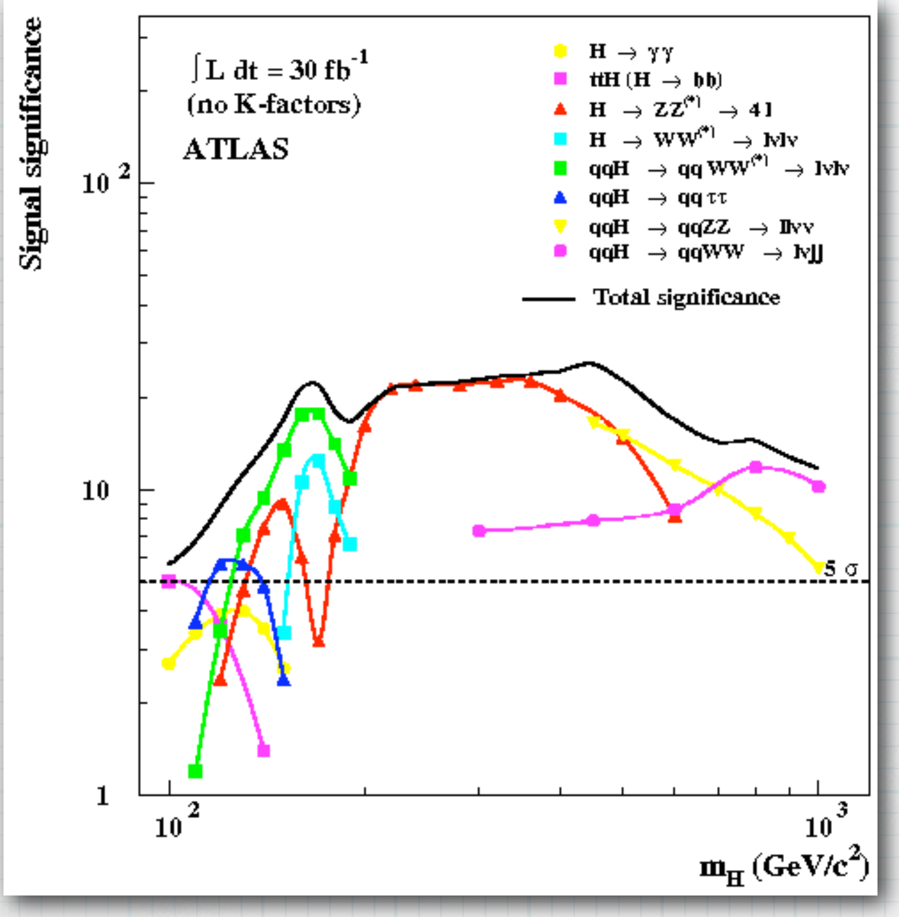
Tau Pair Mass

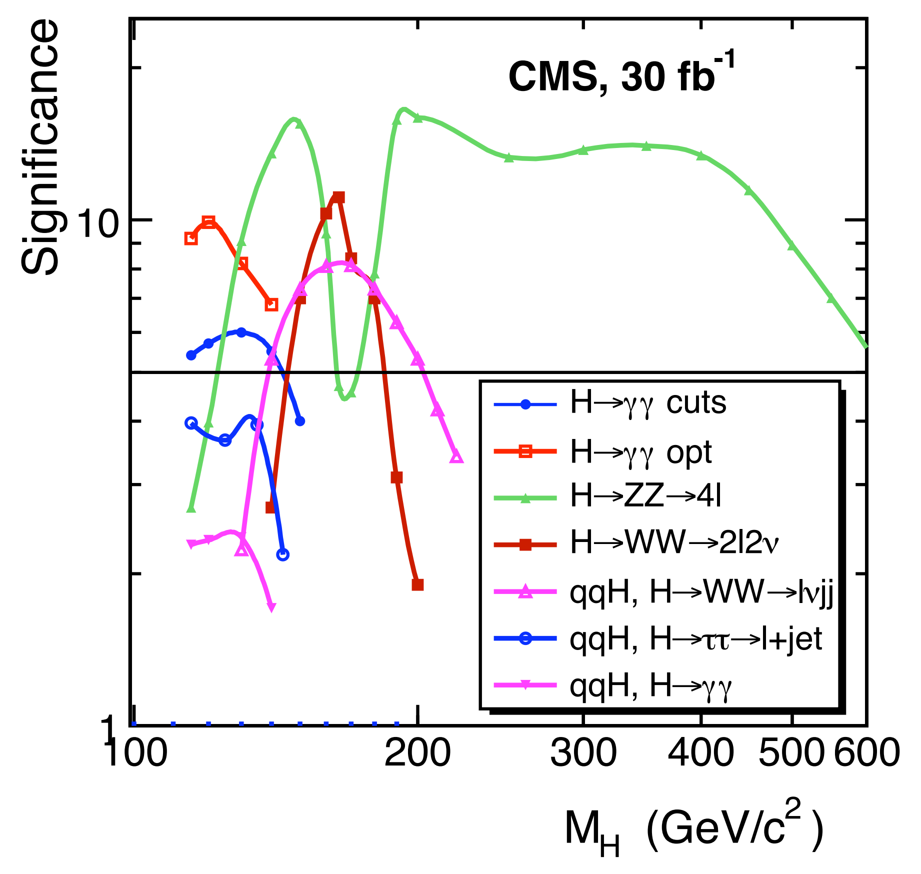
- need to know neutrino energies - assume they point along the tau visible decay direction and use missing E_T to calculate





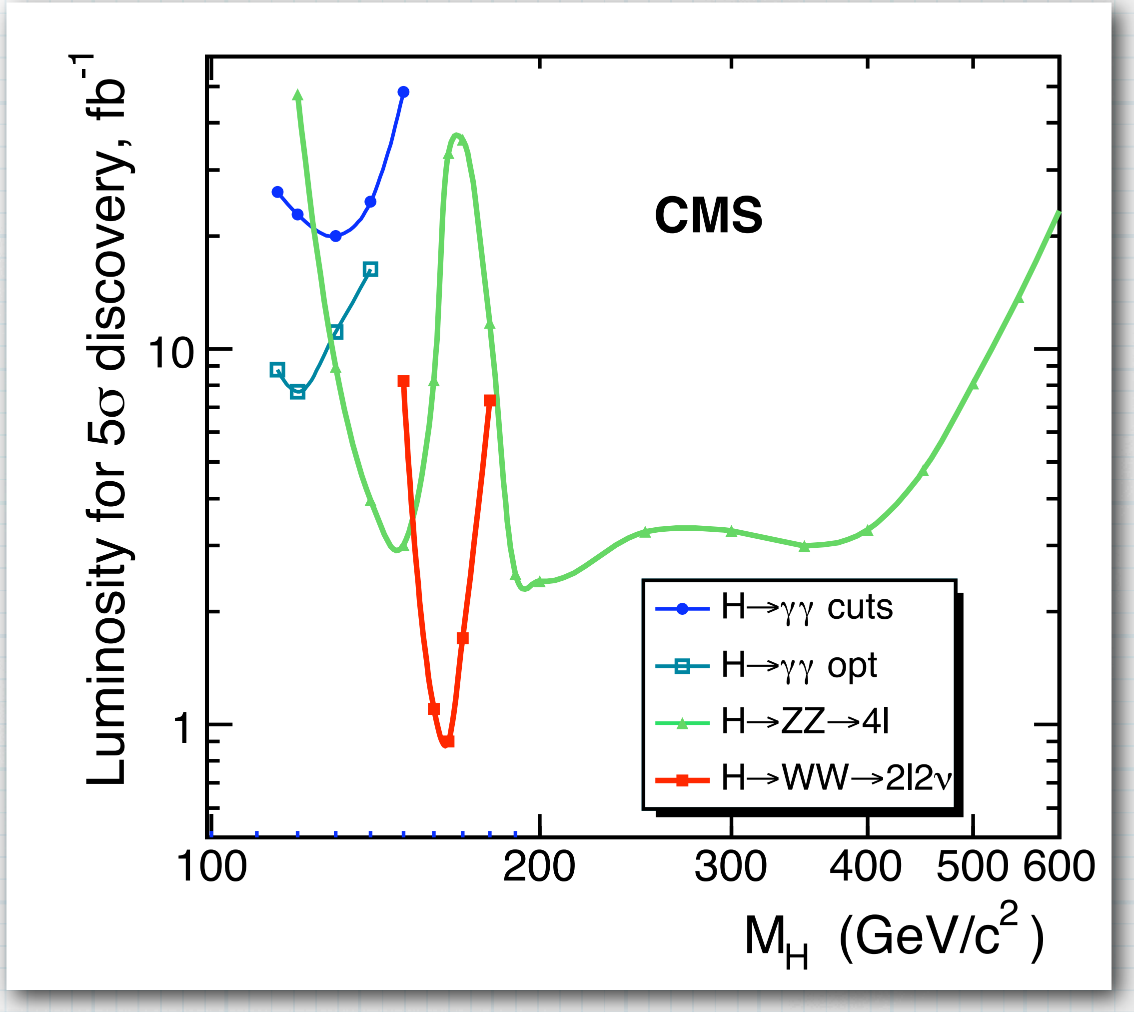
SM Higgs Reach





SM Higgs Reach

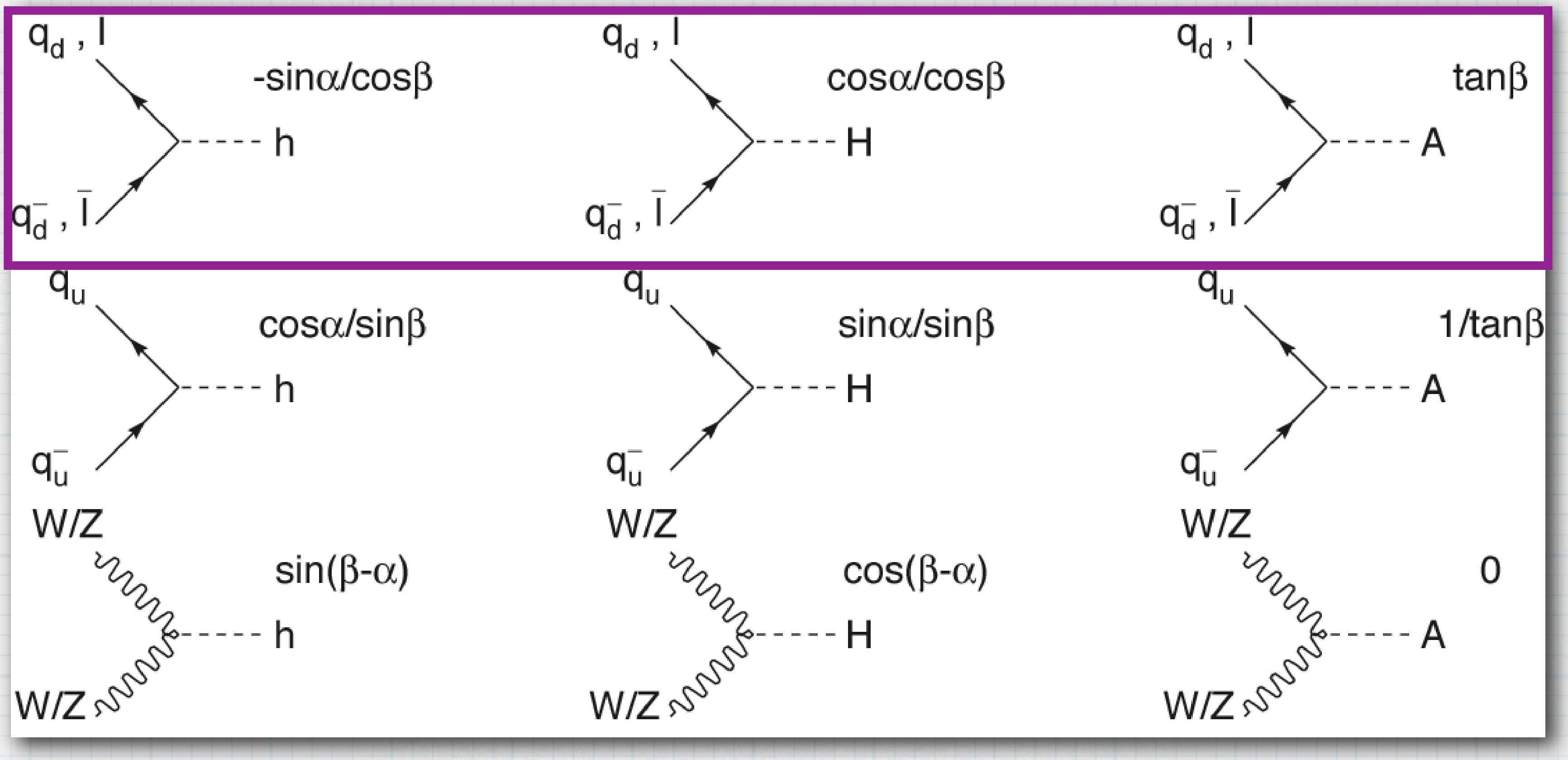
- “when will we discover the Higgs?”
- 2009, most likely...
- if it’s a SM Higgs!



What if we find it?

- even if we find an SM-like Higgs, our work has just begun
- how can we know if it’s the Higgs of the SM, or the one from SUSY
- precision measurements of the Higgs couplings will take time, and are limited by the messy pp environment
- if nature is kind we will get SUSY particles directly, or discover a heavy SUSY-like Higgs
- but nature may not be kind...

SUSY Higgs Search

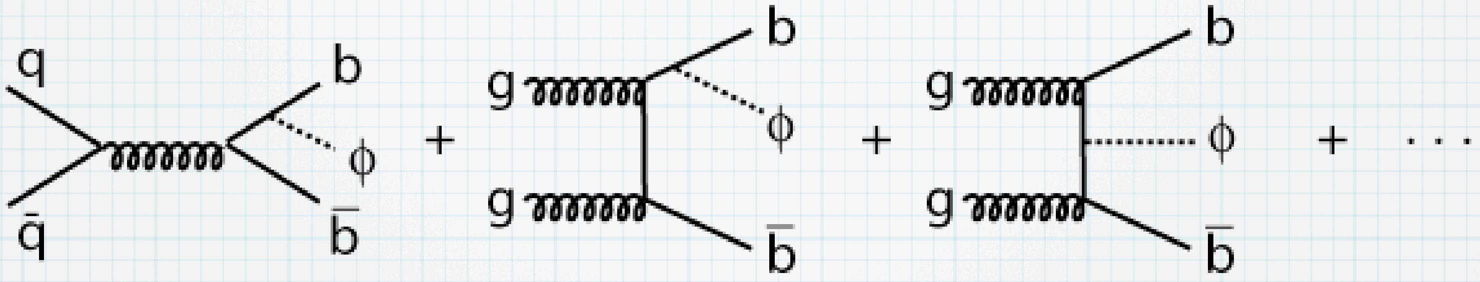


Higgs production enhanced by a factor of $\tan^2\beta$

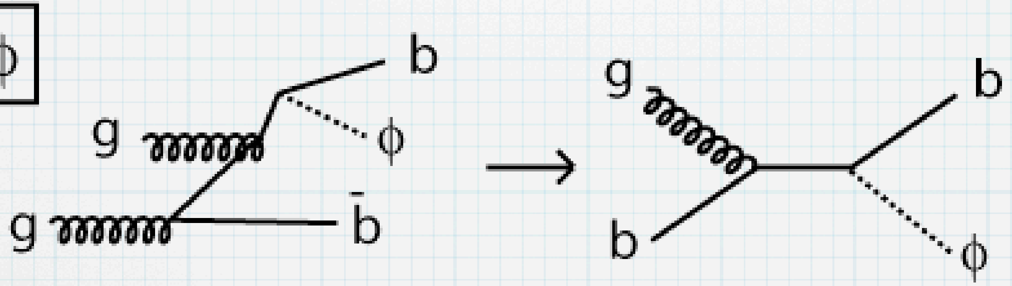
Taus play the dominant role here!

MSSM Higgs + b(b)

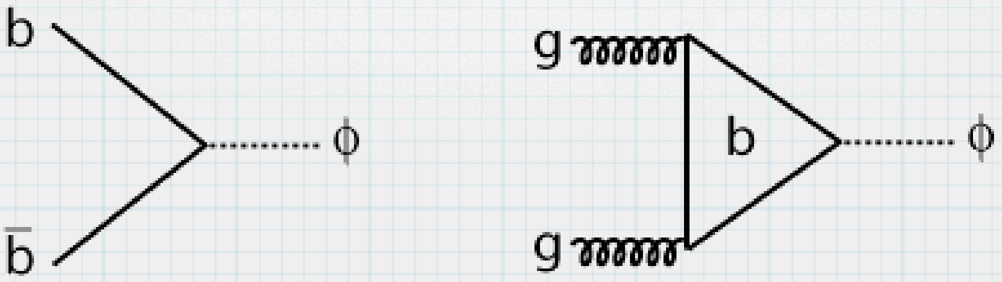
$\boxed{b\bar{b}\phi}$



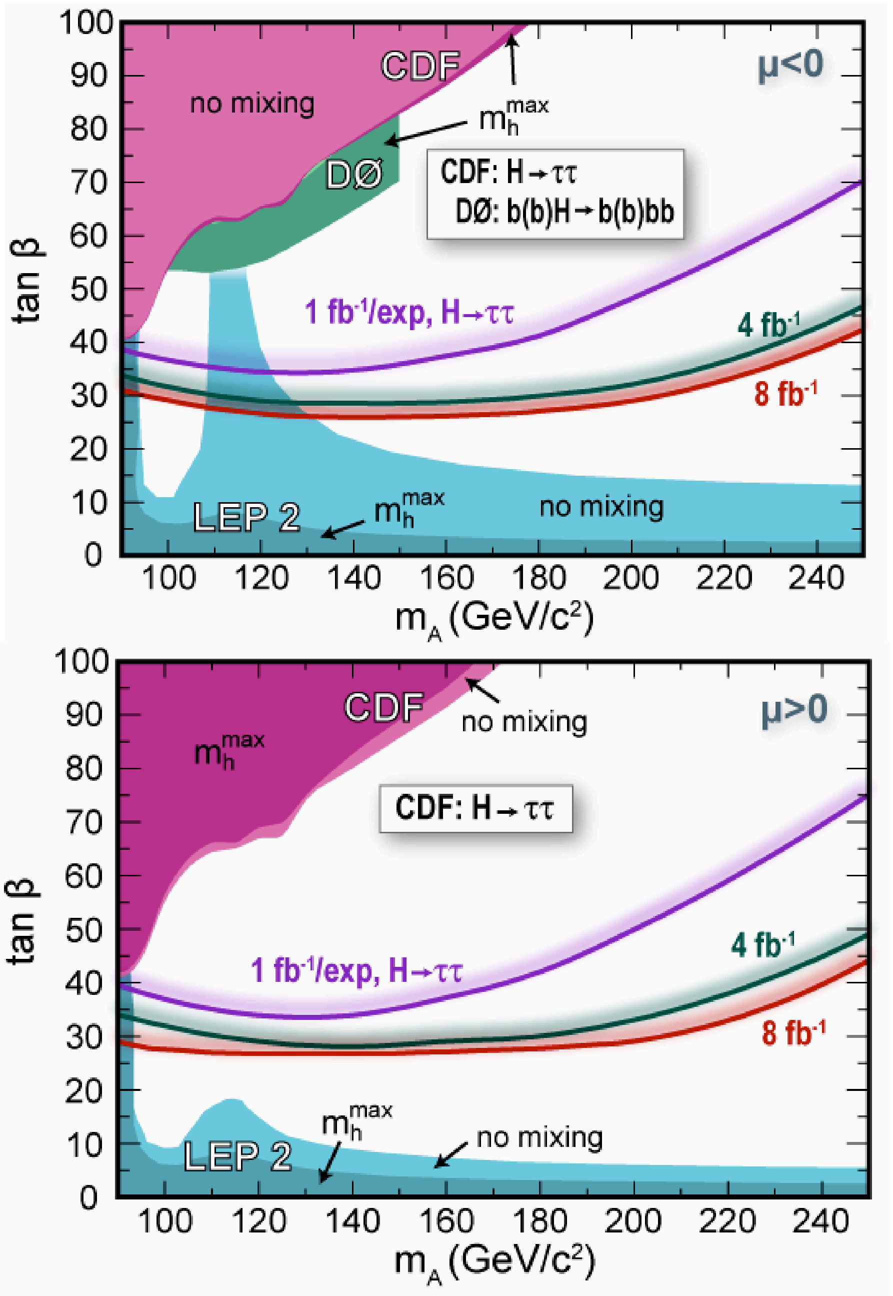
$\boxed{b\phi}$



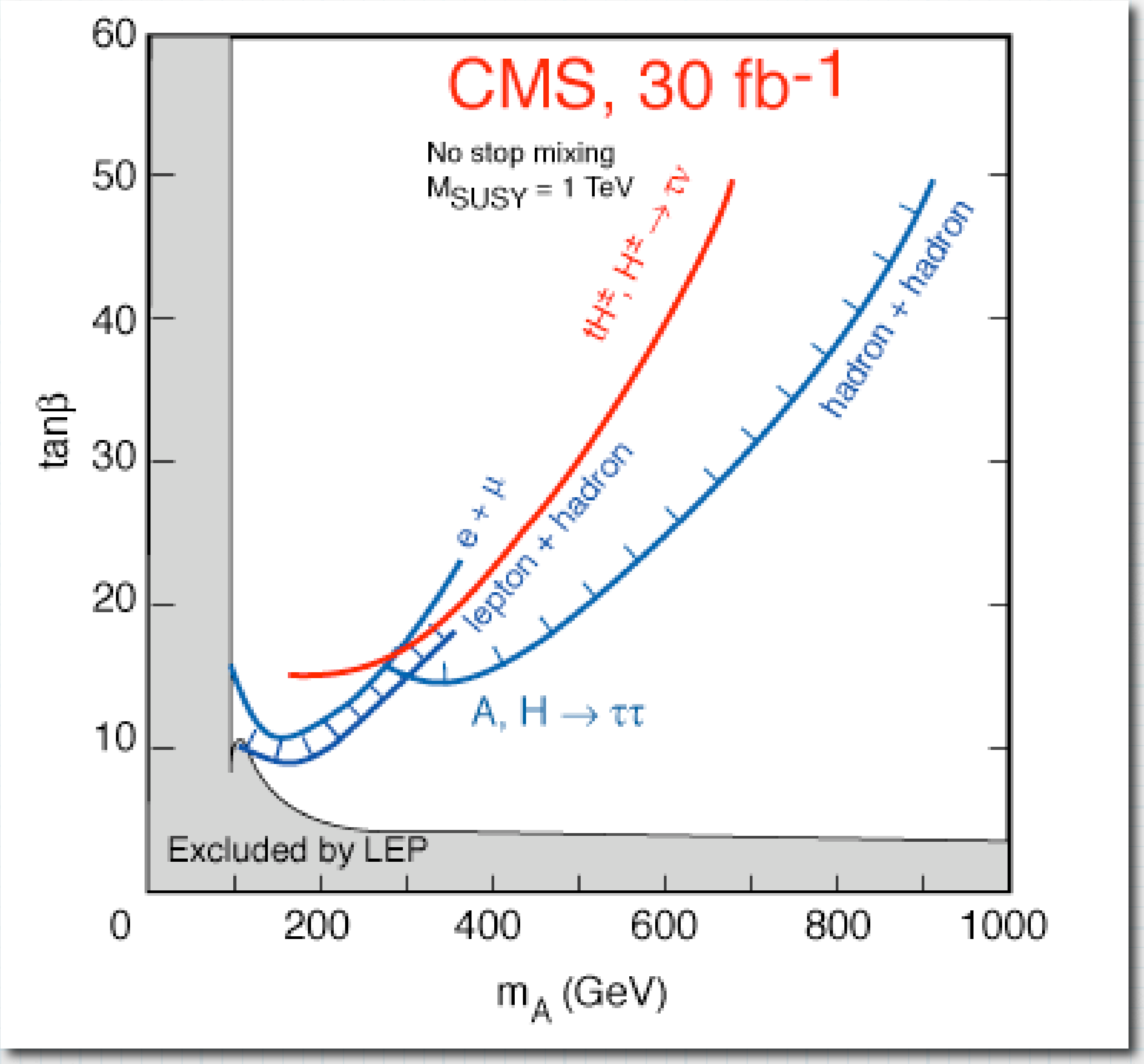
$\boxed{\phi}$



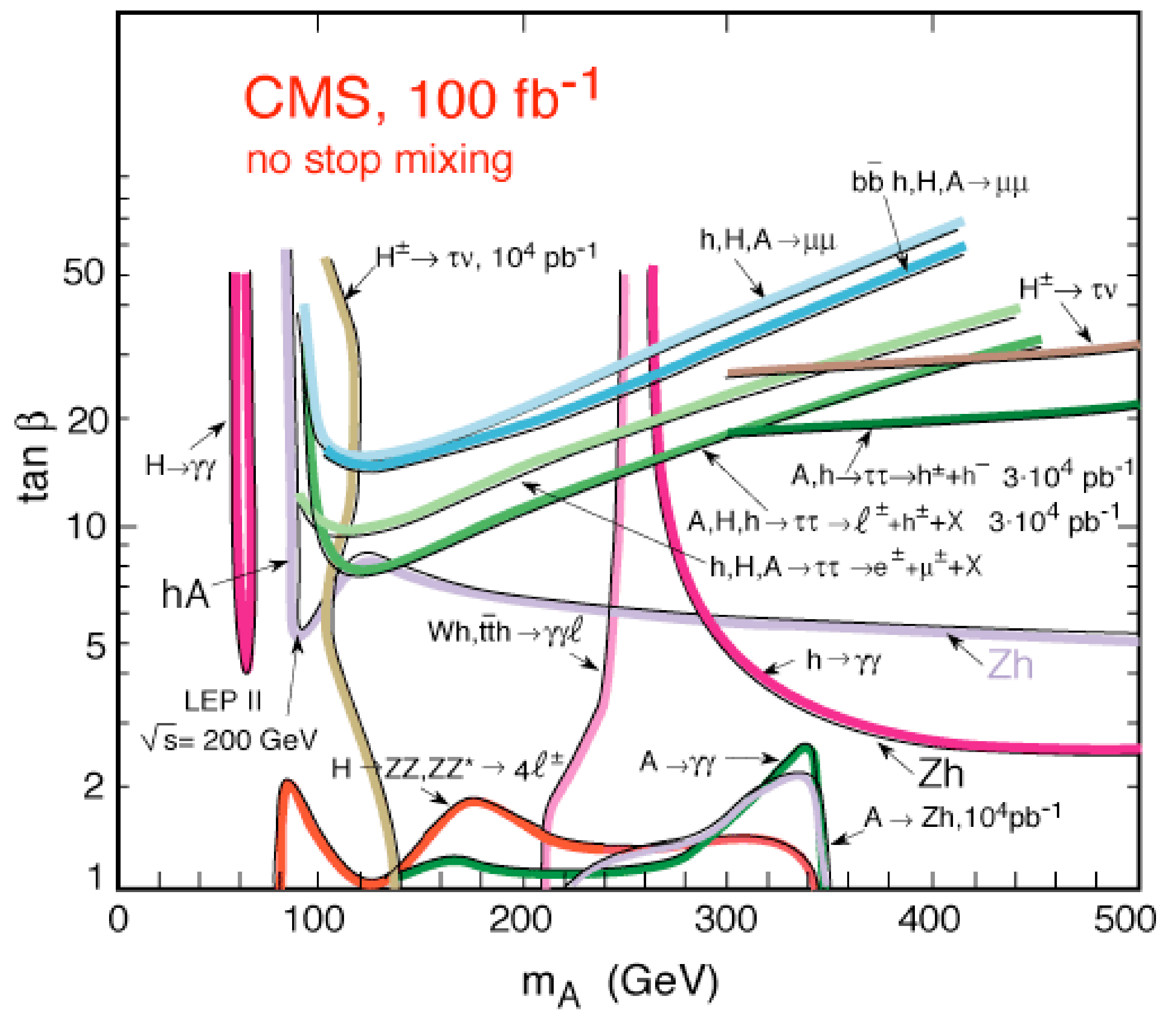
- Tevatron: use both the tau pair and b pair decay modes of the Higgs
- set exclusion bounds in m(A) versus tanβ
- with some luck, Tevatron can see the MSSM Higgs before LHC!
- but if m(A) is large it’s up to the LHC



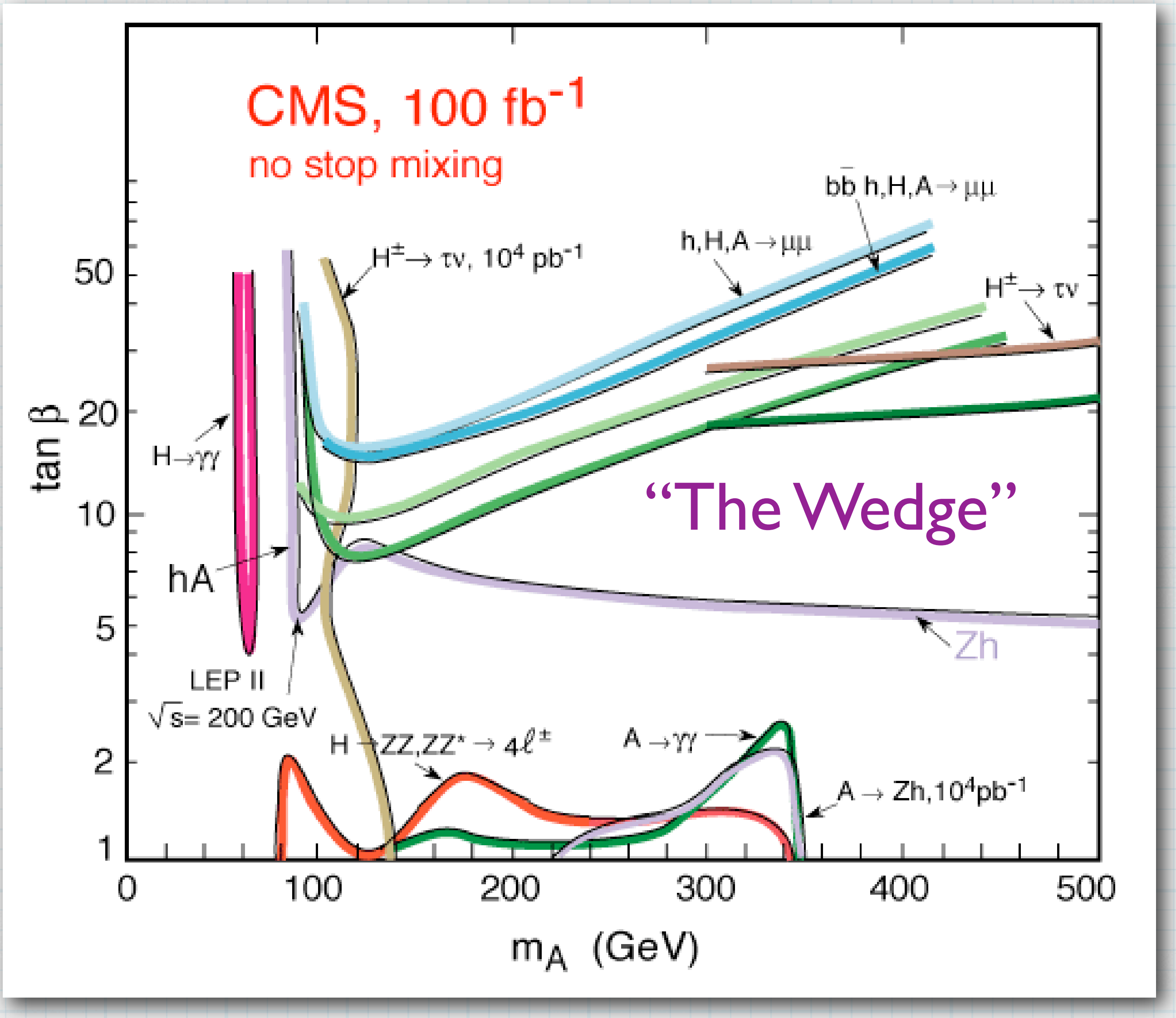
CMS MSSM Higgs (tau)



MSSM Higgs Sensitivity



MSSM Higgs Sensitivity



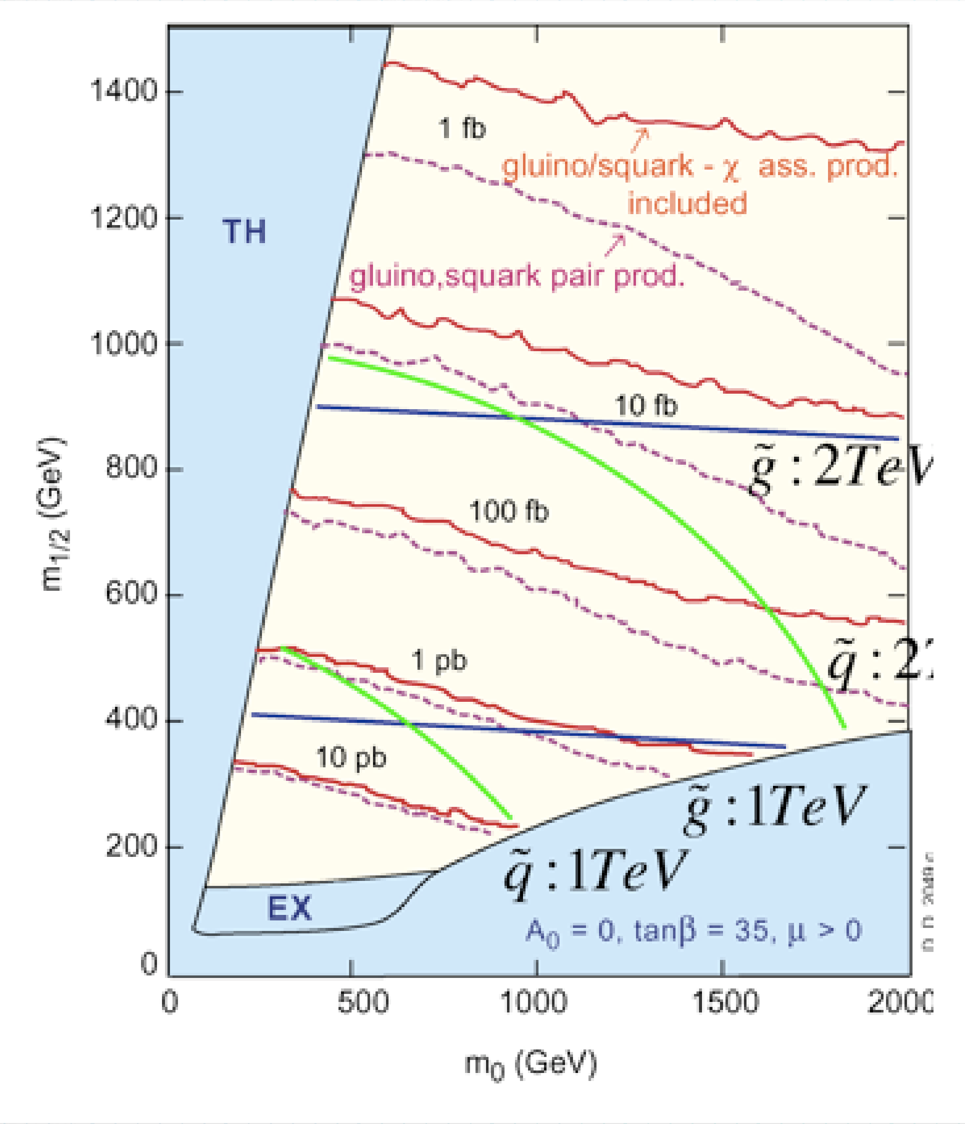
Discovering SUSY Particles

- clearly it would be awesome to discover SUSY particle directly at the LHC
- most promising: strongly produced ones!
- squark/gluino production leads to multijet final states with large missing energy
- some SUSY cascades have high-p_T leptons too
- can play kinematic tricks to tease out some signals

Challenge of SUSY

- classic signature: missing energy from neutralino
- but what if LSP is not neutralino? (sneutrino is okay...)
- SUSY decays of Higgs?
- Higgs decays of SUSY particles?
- signals/signatures depend utterly on completely unknown many-parameter mass hierarchy

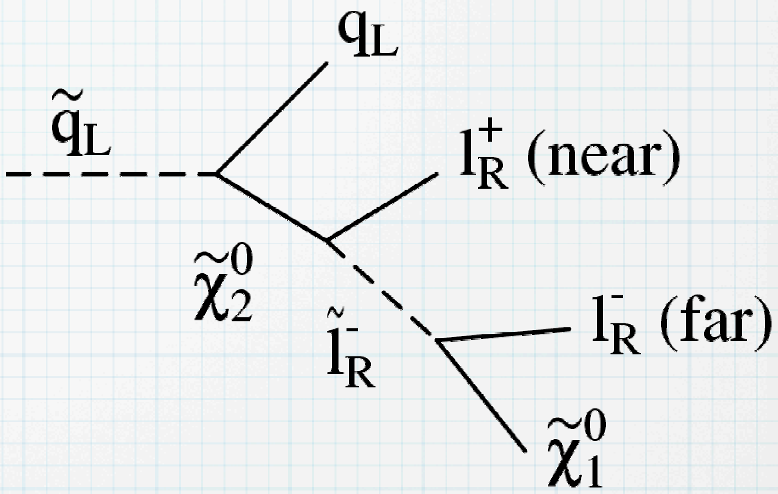
Squark/Gluino Search

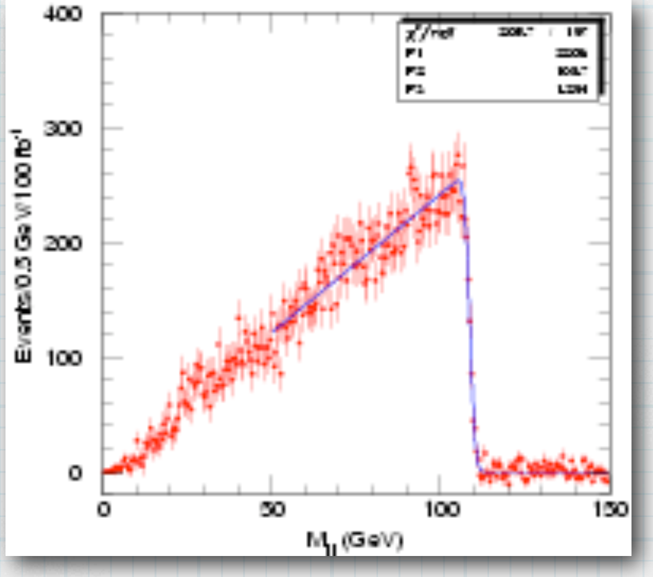


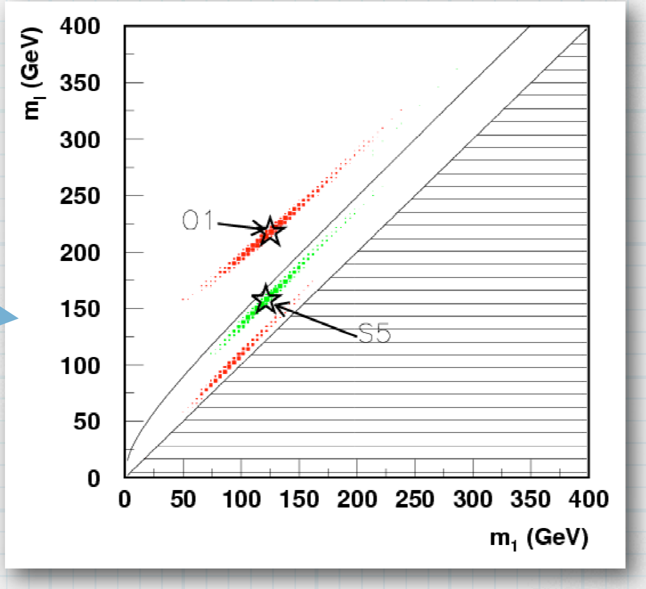
- assume LSP is neutralino
- look for multijets plus missing p_T
- estimate that we can become sensitive with the very first data

Mass Differences: A Hope?

- squark cascade decay gives two leptons
- study lepton mass,p_T distributions

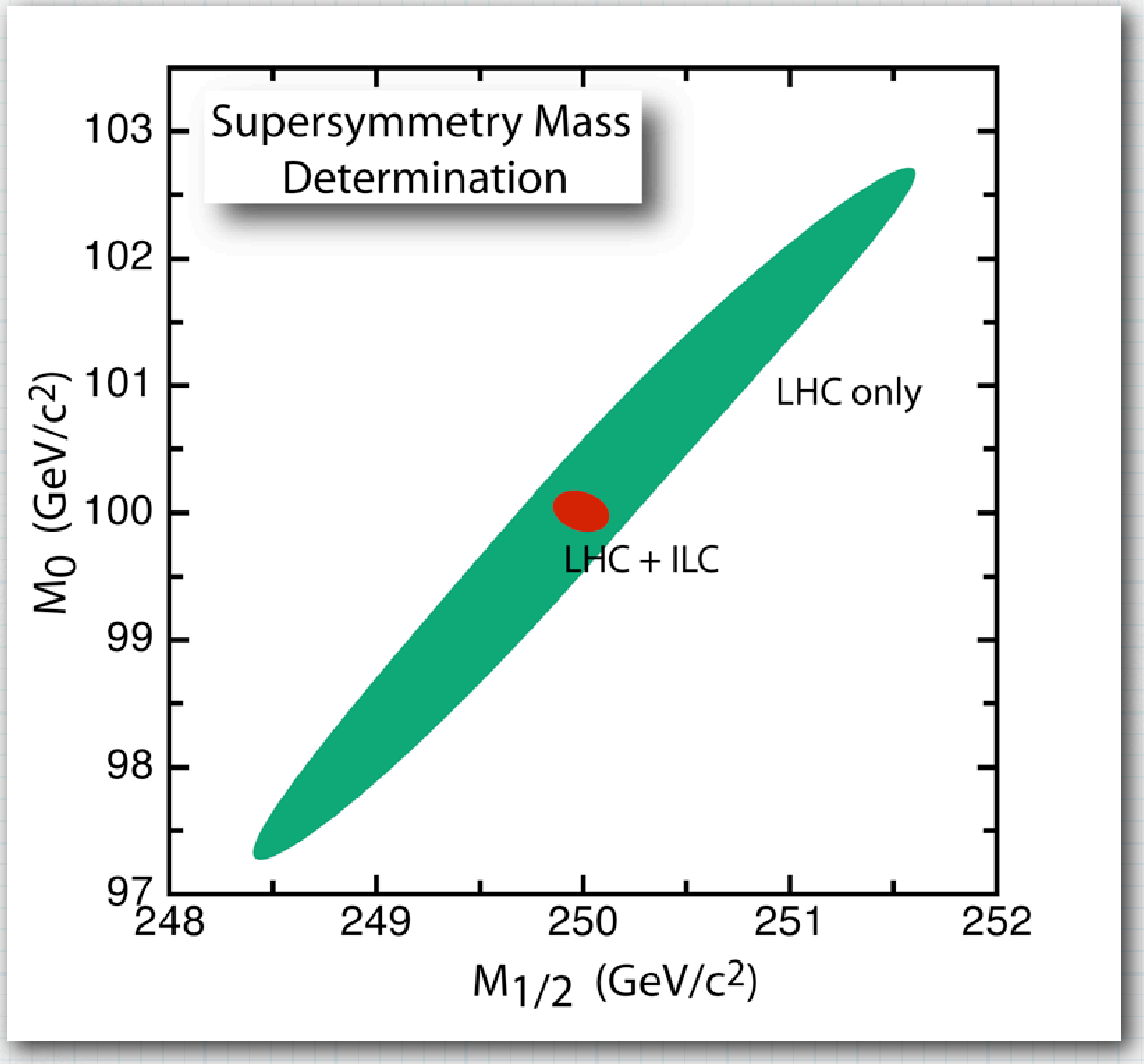






compare with models

We Need the ILC...



PGS and the LHC Olympics

- March 1998: kickoff of the Tevatron Run 2 SUSY/Higgs Workshop
- no Run 2 CDF/D0 simulations available then
- developed “SHW” simulation as average of CDF/D0
- published SHW Higgs report: hep-ph/0010338
- still a reliable resource for Tevatron Higgs reach!
- SHW -> PGS for Snowmass 2001
- used for VLHC, LHC, LC, Tevatron comparisons especially by theorists

PGS - Pretty Good Simulation

- written in Fortran, several thousand lines
- contributions from many people, maintained and extended by me
- major revision in 2006: PGS 4
- driving force: LHC Olympics

Flow of PGS

event generation (PYTHIA, HERWIG, ...)

STDHEP common blocks

event simulation, object reconstruction

user analysis

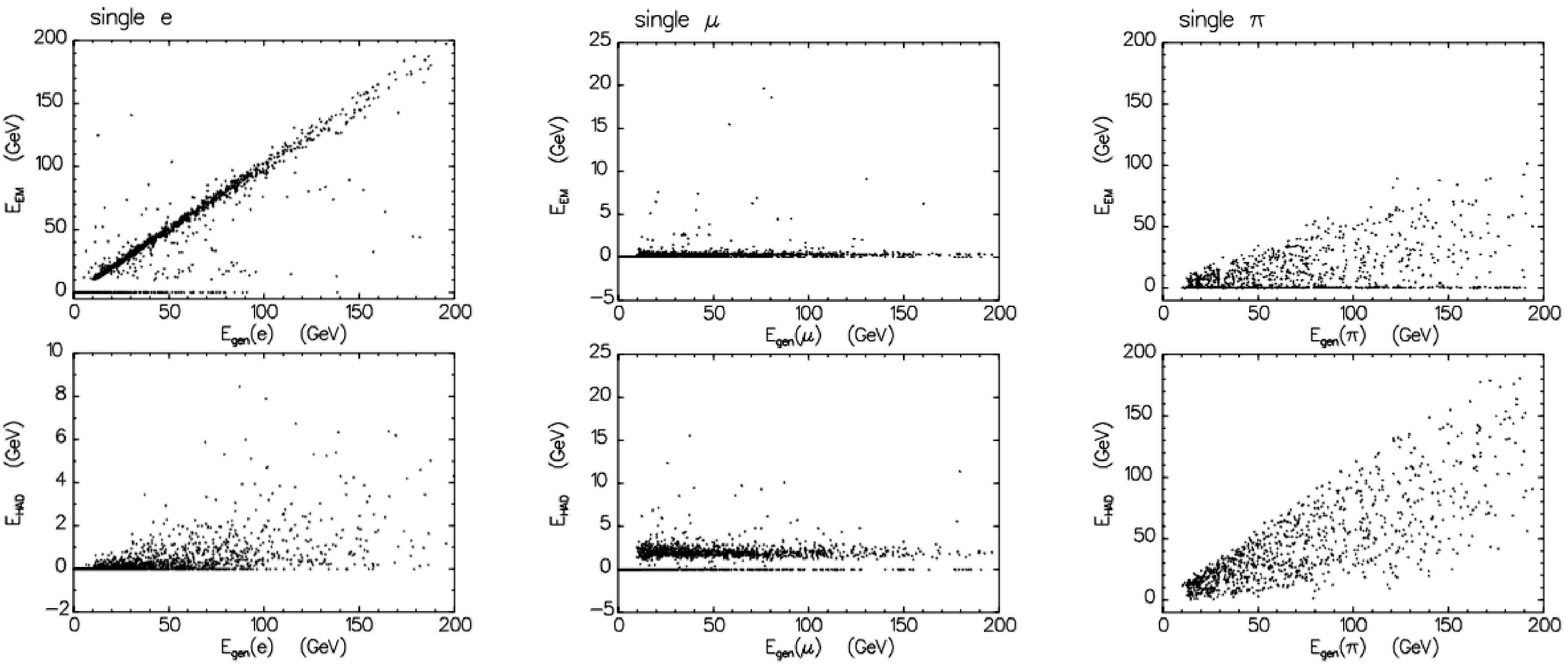
user output

PGS Paradigm

- final state stable particles:
- charged tracks (e, mu, pi, K, ...)
- calorimeter deposits
- take "raw" calorimeter and track information and reconstruct PGS "physics objects":
- gamma, e, mu, tau, jet (b tag), MET
- also have realistic trigger-type objects
- user configurable detector parameters

PGS Simulation

- a charged particle makes a track with some resolution, and some efficiency
- particles deposit energy in an em and had tower:



PGS Parameters

```
LHC                   ! parameter set name
320                   ! eta cells in calorimeter
200                   ! phi cells in calorimeter
0.0314159             ! eta width of calorimeter cells  |eta| < 5
0.0314159             ! phi width of calorimeter cells
0.01                  ! electromagnetic calorimeter resolution  const
0.2                   ! electromagnetic calorimeter resolution * sqrt(E)
0.8                   ! hadronic calolrimeter resolution * sqrt(E)
0.2                   ! MET resolution
0.01                  ! calorimeter cell edge crack fraction
5.0                   ! calorimeter trigger cluster finding seed threshold (GeV)
1.0                   ! calorimeter trigger cluster finding shoulder threshold
0.5                   ! calorimeter kt cluster finder cone size (delta R)
2.0                   ! outer radius of tracker (m)
4.0                   ! magnetic field (T)
0.000013              ! sagitta resolution (m)
0.98                  ! track finding efficiency
1.00                  ! minimum track pt (GeV/c)
3.0                   ! tracking eta coverage
3.0                   ! e/gamma eta coverage
2.4                   ! muon eta coverage
2.0                   ! tau eta coverage
```

What PGS Does Not Do

- no vertex smearing (r or z)
- no multiple interaction
- no secondary particle production (nuclear, brems)
- no gamma to pair conversion
- no multiple scattering (resolution is simulated)
- no lateral shower development (one particle puts its energy in only one calorimeter tower)
- no magnetic field track curvature effects
- no muon brem taken into account!

PGS 4

- k_T jet finding algorithm now the standard
- standardized object ID
- remove dependency on external libraries
- added support for HERWIG, ALPGEN, ...
- more realistic b-tagging
- improved calorimeter resolution
- numerous (but mercifully minor) bug fixes
- new web site

http://www.physics.ucdavis.edu/~conway/research/software/pgs/pgs4-olympics.htm

What is PGS Good For?

- can get very good idea of geometric and kinematic acceptance for various physics processes
- tends to be a bit optimistic - can be used as a quick feasibility test for an analysis
- can get a reasonable idea of gamma, e, mu, tau, jet reconstruction rates

- in short, it's a Pretty Good Simulation, not a great one
- <u>very</u> idealized version of LHC detectors/analyses

LHC Olympics

- original idea from Gordy Kane and Steve Mrenna
- two workshops so far at CERN (July 2005, Feb 2006)
- 3rd LHC Olympics Workshop: 24-25 August, Kavli Institute for Theoretical Physics, UC Santa Barbara
- basic idea: generate "blind" samples of LHC-like data with certain new physics signals embedded
- Olympians will analyze these blind samples to see if they can discover the nature of the new physics

You are welcome to join the Olympics and see what it is like to explore the new energy frontier!

LHC Olympics

main Olympics web site:

http://physics.princeton.edu/~verlinde/research/lhco/

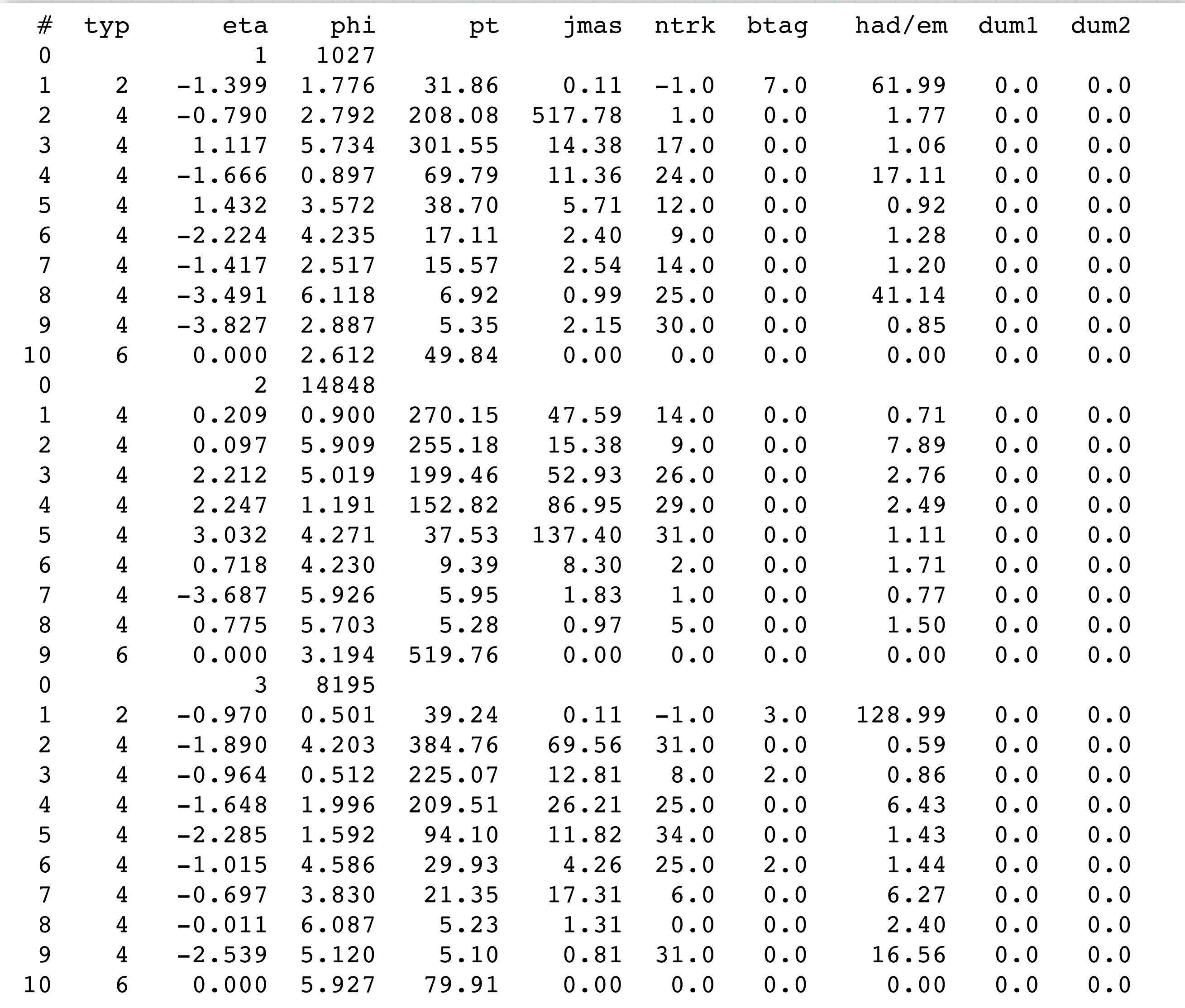
two black-box blinded samples available, one challenging, one relatively easy

can analyze the data with your own program, or use the provided Mathematica-based Chameleon package

can make your own PGS background or signal samples using the "olympics" executable (Linux, OSX, Cygwin!)

black box data and code available at:
http://physics.princeton.edu/~verlinde/research/lhco/BB/

LHC Olympics Data



```
  #  typ      eta    phi       pt     jmas  ntrk  btag   had/em  dum1  dum2
  0             1   1027
  1    2   -1.399  1.776    31.86     0.11  -1.0   7.0    61.99   0.0   0.0
  2    4   -0.790  2.792   208.08   517.78   1.0   0.0     1.77   0.0   0.0
  3    4    1.117  5.734   301.55    14.38  17.0   0.0     1.06   0.0   0.0
  4    4   -1.666  0.897    69.79    11.36  24.0   0.0    17.11   0.0   0.0
  5    4    1.432  3.572    38.70     5.71  12.0   0.0     0.92   0.0   0.0
  6    4   -2.224  4.235    17.11     2.40   9.0   0.0     1.28   0.0   0.0
  7    4   -1.417  2.517    15.57     2.54  14.0   0.0     1.20   0.0   0.0
  8    4   -3.491  6.118     6.92     0.99  25.0   0.0    41.14   0.0   0.0
  9    4   -3.827  2.887     5.35     2.15  30.0   0.0     0.85   0.0   0.0
 10    6    0.000  2.612    49.84     0.00   0.0   0.0     0.00   0.0   0.0
  0             2  14848
  1    4    0.209  0.900   270.15    47.59  14.0   0.0     0.71   0.0   0.0
  2    4    0.097  5.909   255.18    15.38   9.0   0.0     7.89   0.0   0.0
  3    4    2.212  5.019   199.46    52.93  26.0   0.0     2.76   0.0   0.0
  4    4    2.247  1.191   152.82    86.95  29.0   0.0     2.49   0.0   0.0
  5    4    3.032  4.271    37.53   137.40  31.0   0.0     1.11   0.0   0.0
  6    4    0.718  4.230     9.39     8.30   2.0   0.0     1.71   0.0   0.0
  7    4   -3.687  5.926     5.95     1.83   1.0   0.0     0.77   0.0   0.0
  8    4    0.775  5.703     5.28     0.97   5.0   0.0     1.50   0.0   0.0
  9    6    0.000  3.194   519.76     0.00   0.0   0.0     0.00   0.0   0.0
  0             3   8195
  1    2   -0.970  0.501    39.24     0.11  -1.0   3.0   128.99   0.0   0.0
  2    4   -1.890  4.203   384.76    69.56  31.0   0.0     0.59   0.0   0.0
  3    4   -0.964  0.512   225.07    12.81   8.0   2.0     0.86   0.0   0.0
  4    4   -1.648  1.996   209.51    26.21  25.0   0.0     6.43   0.0   0.0
  5    4   -2.285  1.592    94.10    11.82  34.0   0.0     1.43   0.0   0.0
  6    4   -1.015  4.586    29.93     4.26  25.0   2.0     1.44   0.0   0.0
  7    4   -0.697  3.830    21.35    17.31   6.0   0.0     6.27   0.0   0.0
  8    4   -0.011  6.087     5.23     1.31   0.0   0.0     2.40   0.0   0.0
  9    4   -2.539  5.120     5.10     0.81  31.0   0.0    16.56   0.0   0.0
 10    6    0.000  5.927    79.91     0.00   0.0   0.0     0.00   0.0   0.0
```