The Road to Discovery

Experimental Search Methods

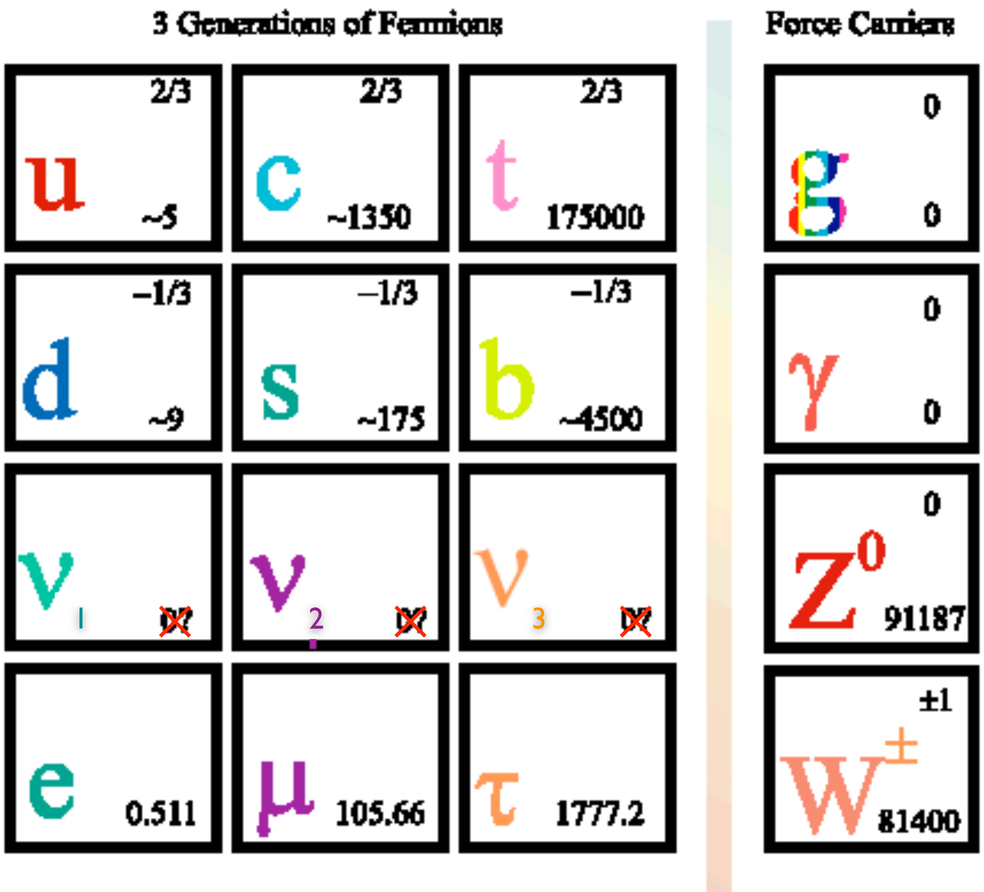
Gustaaf Brooijmans

Columbia University
In the City of New York

Goals

- For theorists (and experimenters):
  - What is it these experimenters *do*?
    - (Except for using fancy equipment to build fancy detectors...)
  - How come it takes them forever to release a result?
  - Why can't they just give us the 4-vectors as they record the data?
  - What can('t) they look for?
- For experimenters:
  - Why does it take so many layers of review to get a result out?   My analysis seems pretty simple at first sight....

HEP in 2011



Masses are in MeV

CKM elements:

Observable	Central ± 1 σ
$\|V_{ud}\|$	0.97430 [+0.00019 -0.00019]
$\|V_{us}\|$	0.22521 [+0.00082 -0.00082]
$\|V_{ub}\|$	0.00350 [+0.00015 -0.00014]
$\|V_{cb}\|$	0.04117 [+0.00038 -0.00115]
$\|V_{ud}\|$ (meas. not in the fit)	0.97444 [+0.00028 -0.00028]
$\|V_{us}\|$ (meas. not in the fit)	0.2257 [+0.0011 -0.0011]
$\|V_{ub}\|$ (meas. not in the fit)	0.00350 [+0.00015 -0.00016]
$\|V_{cb}\|$ (meas. not in the fit)	0.04399 [+0.00069 -0.00397]
$\|V_{cd}\|$	0.22508 [+0.00082 -0.00082]
$\|V_{cs}\|$	0.97347 [+0.00019 -0.00019]
$\|V_{td}\|$	0.00859 [+0.00027 -0.00029]
$\|V_{ts}\|$	0.04041 [+0.00038 -0.00115]
$\|V_{tb}\|$	0.999146 [+0.000047 -0.000016]

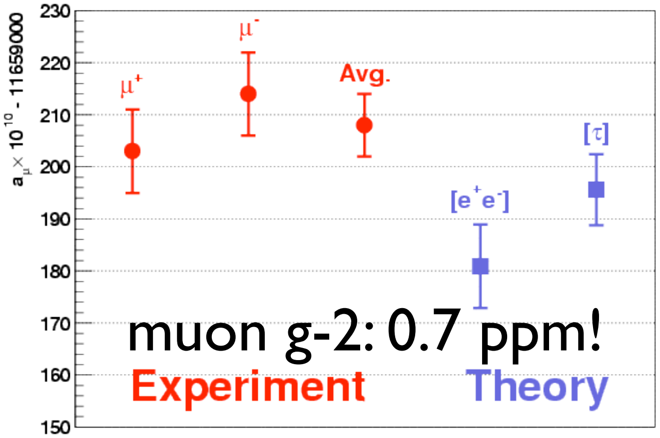
In Words

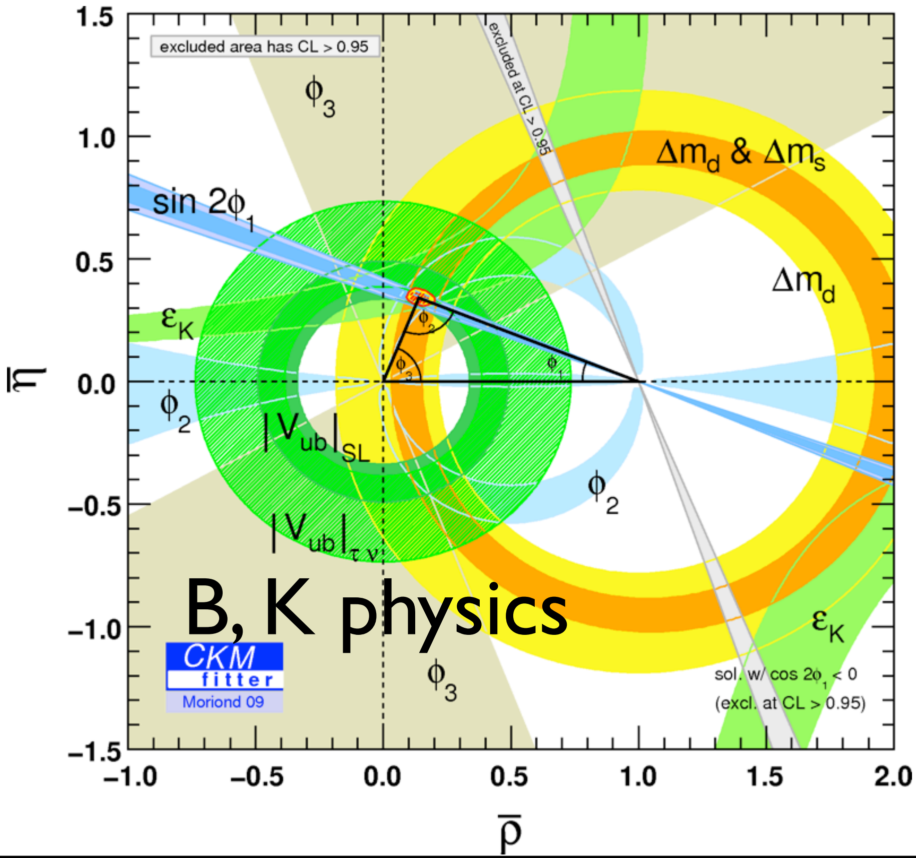
- Matter is built of spin 1/2 particles that interact by exchanging 3 different kinds of spin 1 particles corresponding to 3 different (gauge) interactions
- There appear to be 3 generations of matter particles
- The 4 different matter particles in each generation carry different combinations of quantized charges characterizing their couplings to the interaction bosons
- The matter fermions and the weak bosons have “mass”
- Gravitation is presumably mediated by spin 2 gravitons
- Gravitation is extremely weak for typical particle masses
- There appear to be 3 macroscopic dimensions

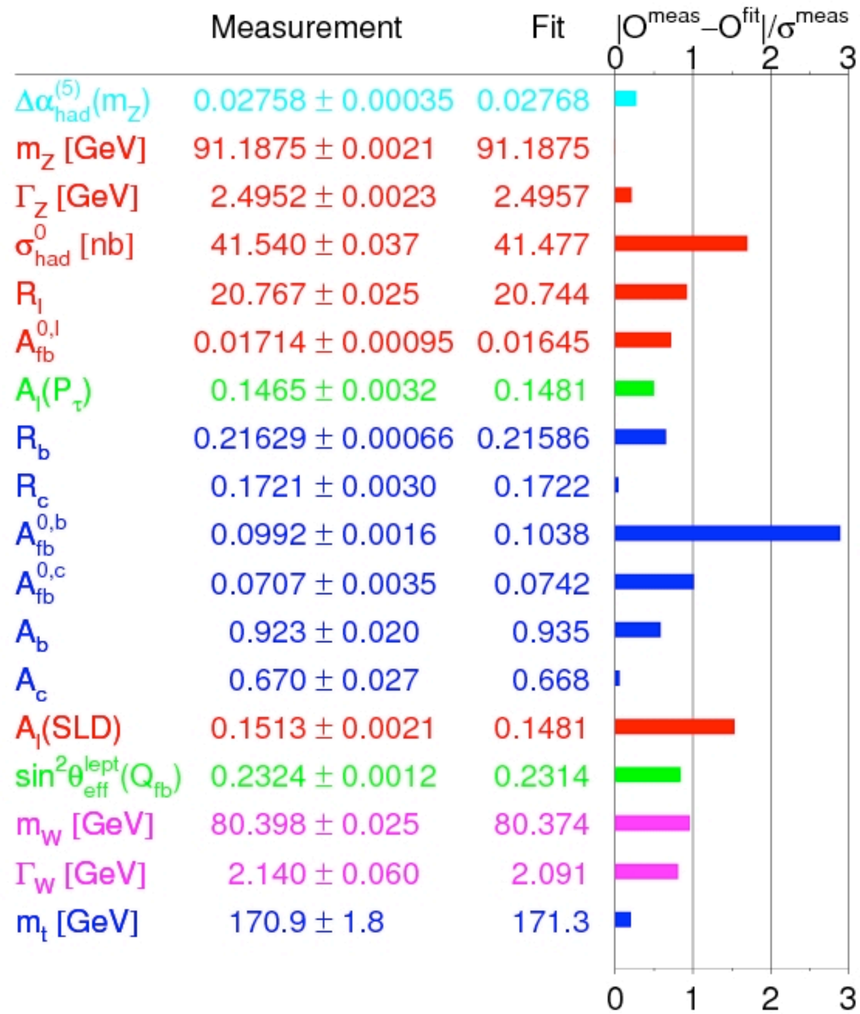
About the Standard Model

- It's a theory of interactions:
  - Properties of fermions are inputs
  - Properties of interaction bosons in terms of couplings, propagations, masses are linked:
    - Measuring a few allows us to predict the rest, then measure and compare with expectation
- It's remarkably successful:
  - Predictions verified to be correct at sometimes incredible levels of precision
  - After ~30 years, still no serious cracks

Precision Results







	Measurement	Fit	$\|O^{meas}-O^{fit}\|/\sigma^{meas}$
$\Delta\alpha_{had}^{(5)}(m_Z)$	0.02758 ± 0.00035	0.02768	
m_Z [GeV]	91.1875 ± 0.0021	91.1875	
Γ_Z [GeV]	2.4952 ± 0.0023	2.4957	
σ_{had}^0 [nb]	41.540 ± 0.037	41.477	
R_l	20.767 ± 0.025	20.744	
$A_{fb}^{0,l}$	0.01714 ± 0.00095	0.01645	
$A_l(P_\tau)$	0.1465 ± 0.0032	0.1481	
R_b	0.21629 ± 0.00066	0.21586	
R_c	0.1721 ± 0.0030	0.1722	
$A_{fb}^{0,b}$	0.0992 ± 0.0016	0.1038	
$A_{fb}^{0,c}$	0.0707 ± 0.0035	0.0742	
A_b	0.923 ± 0.020	0.935	
A_c	0.670 ± 0.027	0.668	
A_l(SLD)	0.1513 ± 0.0021	0.1481	
$\sin^2\theta_{eff}^{lept}(Q_{fb})$	0.2324 ± 0.0012	0.2314	
m_W [GeV]	80.398 ± 0.025	80.374	
Γ_W [GeV]	2.140 ± 0.060	2.091	
m_t [GeV]	170.9 ± 1.8	171.3	

LEP, SLD & Tevatron

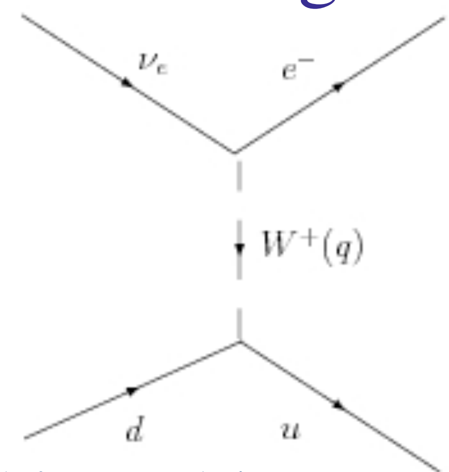
Many Fundamental Questions

- What exactly *is* spin? Or color? Or electric charge? Why are they quantified?
- Are there only 3 generations? If so, why?
- Why are there e.g. no neutral, colored fermions?
- What is mass? Why are particles so light?
- Is there a link between particle and nucleon masses?
- How does all of this reconcile with gravitation? How many space-time dimensions are there really?
- ...

The Plot

Vector Boson Scattering

- There is in fact one known problem with the Standard Model:
  - If we collide W's and/or Z's (not so easy...), the scattering cross-section grows with the center of mass energy, and gets out of control at about 1.7 TeV
- This is similar to "low" energy neutrino scattering:
  - If $q^2 << (M_W)^2$, looks like a "contact interaction", and cross-section grows with center of mass energy
  - But when $q^2 \approx (M_W)^2$, W-boson propagation becomes visible, and "cures" this problem

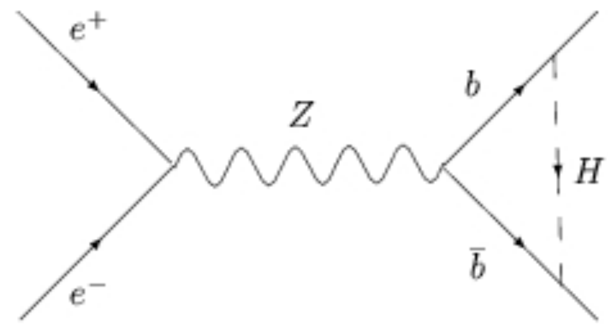


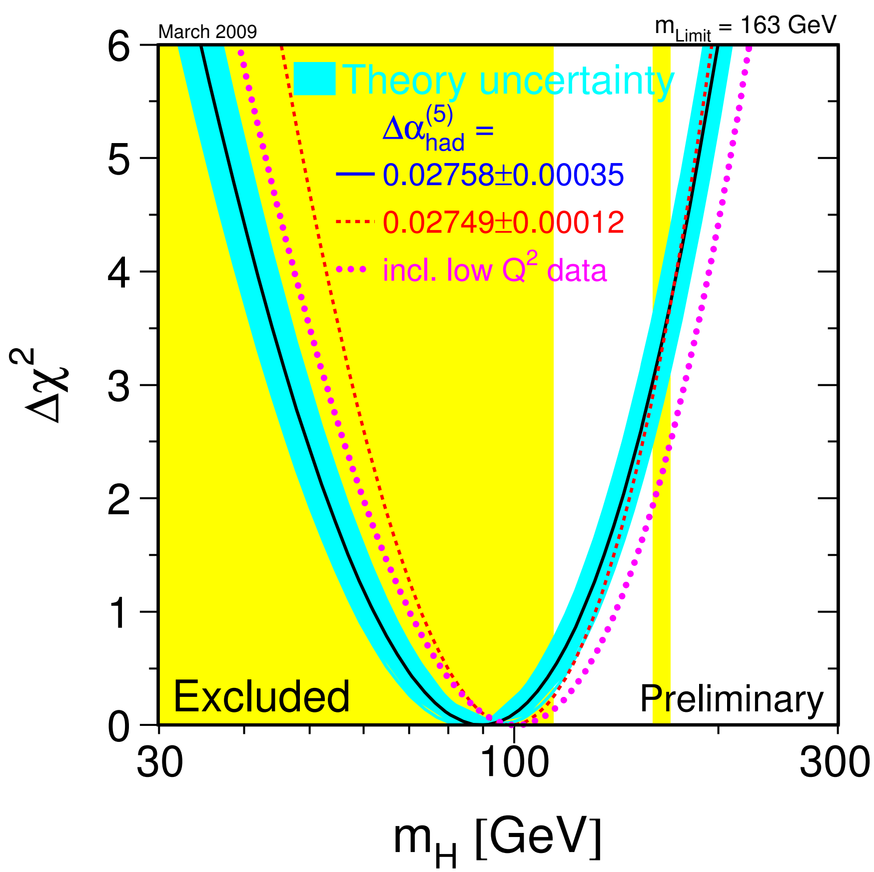
The Higgs Boson

- One way to solve this, is to introduce a massive, spinless particle (of mass < ~1 TeV)
  - Couplings to W and Z are fixed, quantum numbers are known...
  - .... to be those of the vacuum
  - Its mass is unknown, and its couplings to the fermions are unknown.... well, maybe
    - Fermions can acquire mass by coupling to this Higgs boson, so their couplings could be proportional to their masses. This is called the "Standard Model Higgs"

Precision Measurements

- In fact, we can say something about the standard model Higgs mass
  - If the fermions get their masses from the Higgs, we know all couplings and can infer the Higgs mass from precision measurements
  - Result is very sensitive to measured top quark, W boson masses
    - Really wants a "light" Higgs boson



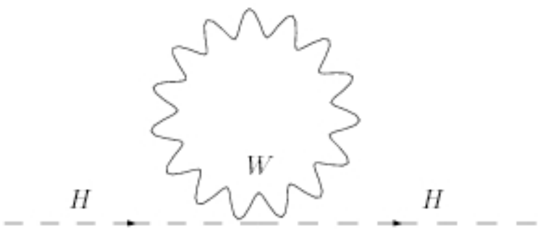


Higgs Drawbacks

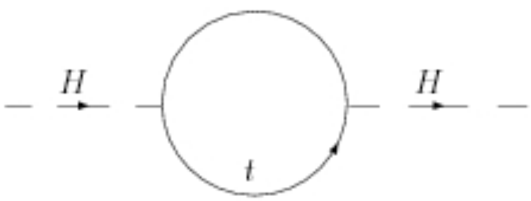
- In principle, with the addition of a Higgs boson around 150 GeV particle physics could be "complete"
  - Like Mendeleev's table for chemistry
- But by itself, the Higgs is very unsatisfactory:
  - Why are the couplings to the fermions what they are?
    - Dumb luck (aka landscape)?
  - What is the link to gravity?
  - Why does the Higgs break the symmetry?
  - Why are there 3....?

The Plot Thickens

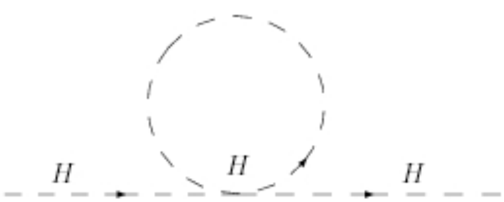
Higgs Mass



$\rightarrow \frac{1}{16\pi^2} g^2 E^2$



$\rightarrow \frac{3}{16\pi^2} y_t^2 E^2$



$\rightarrow \frac{1}{16\pi^2} \lambda E^2$

- Higgs, in fact, also acquires mass from coupling to W's, fermions, and itself!
  - These "mass terms" are quadratically divergent
  - Drive mass to limit of validity of the theory
- So we expect the Higgs mass to be close to the scale where new physics comes in....

Unravelling the Mystery

Hunting for Answers

- Get more information
  - Measure particles and their interactions in detail
    - Precision measurements
  - Observe new particles or interactions
    - Search in new areas in “phase space”
- Find the underlying pattern(s)
  - Hypothesize, build models
    - Internally consistent? Consistent with data?
    - Suggestions on where to look

Experiment

Theory

Where to Start?

- BSM physics **must** couple to SM (weakly?), but is it
  - "SM-like"?
    - Does it have new massive particles decaying to electrons, muons, quarks,...?
  - Quasi "SM-like"?
    - Same but includes some new long-lived particles in the decay chain... (dark matter candidate)
  - No new "particles" in reach
    - Hidden or too heavy or.... don't exist
  - Are there new interactions?

Dark Matter??

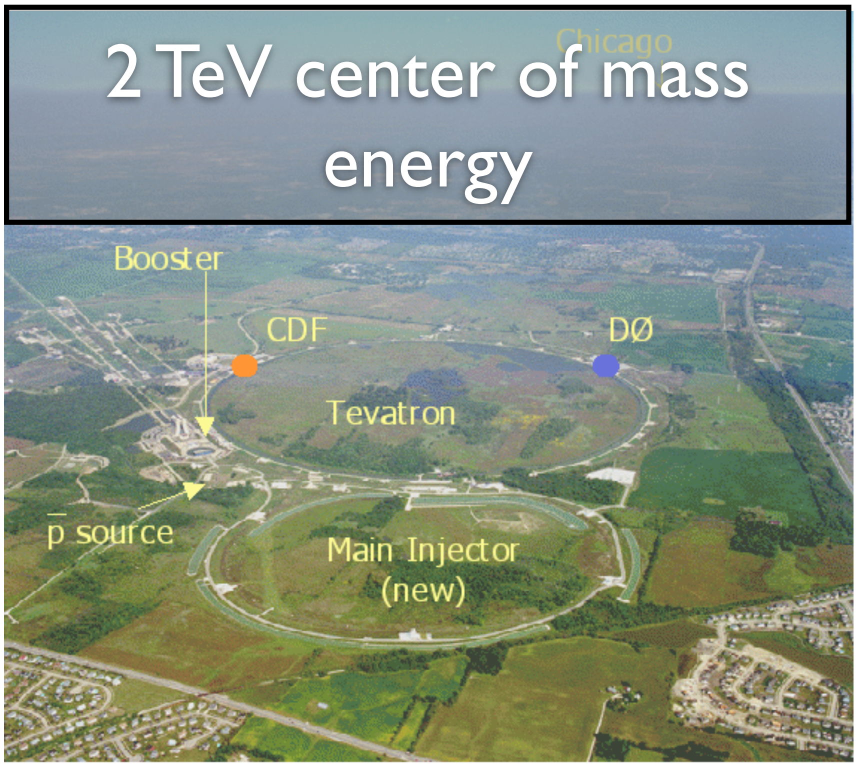
So....

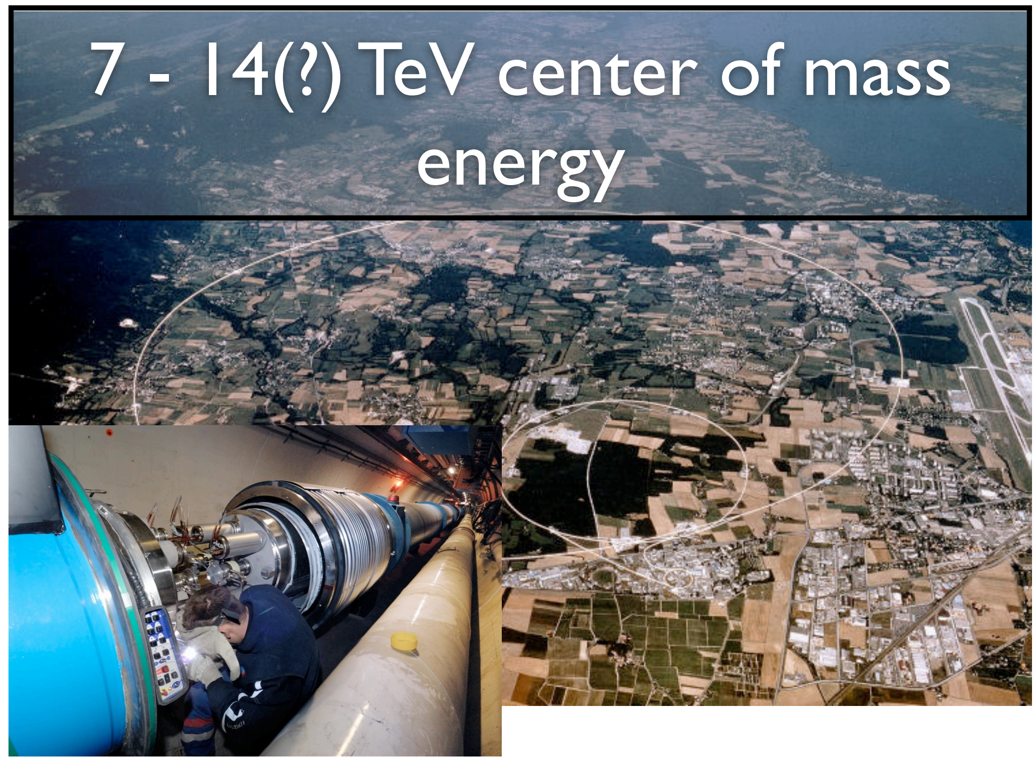
- Go look where the SM breaks down (high energy)...
- ... or for subtle anomalies
- Assume new physics manifestations lead to anomalous production of SM particles
  - Resonant or not (and maybe in loops only)
  - Short-lived or less so
- Rely on guidance from models to some extent
  - What are implications of known constraints? What signatures are "allowed"?
  - Some scenarios do require new approach

The Tools

Colliders

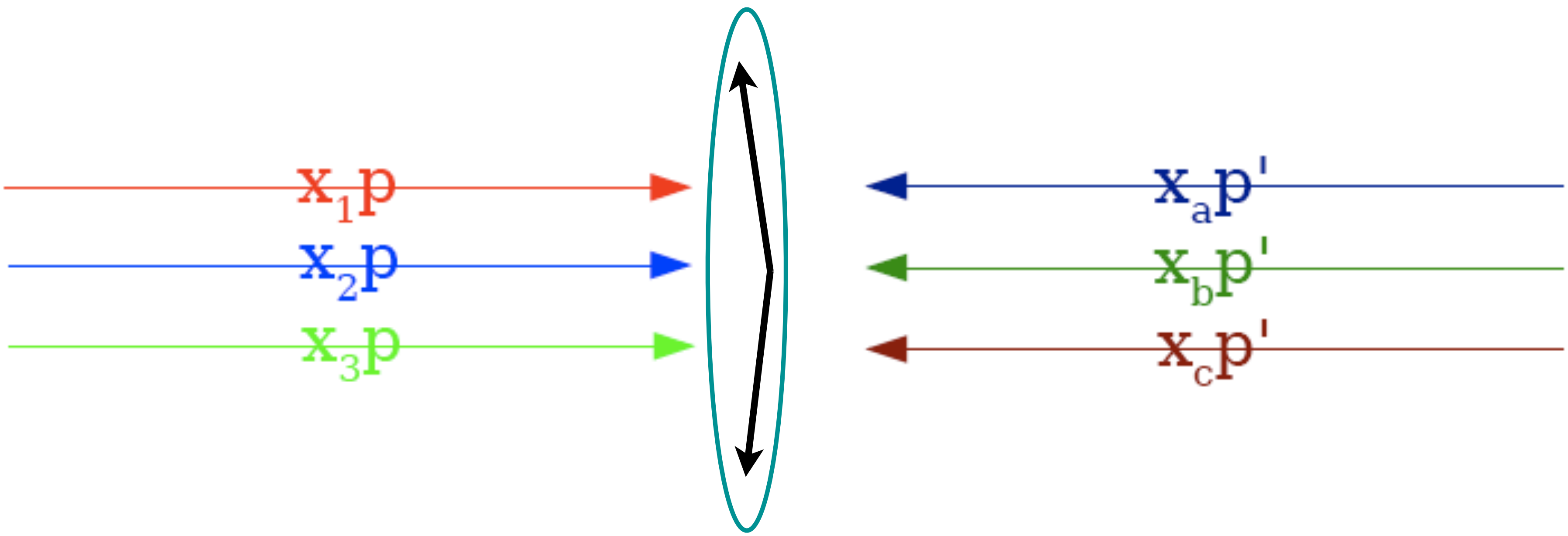
- Currently, hadron colliders:
  - High energy implies probing of short distances, and production of other, massive particles





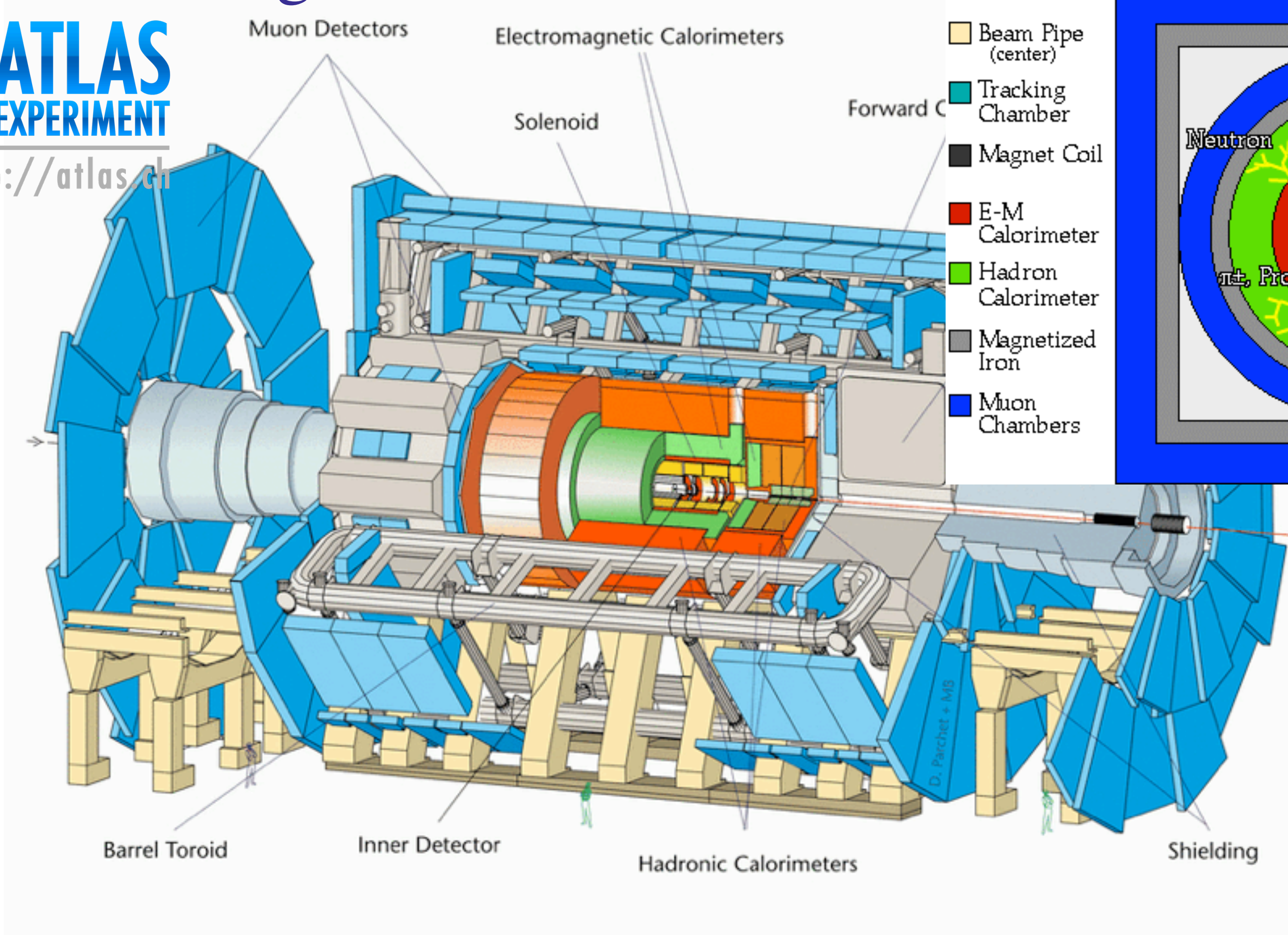
Hadron Colliders

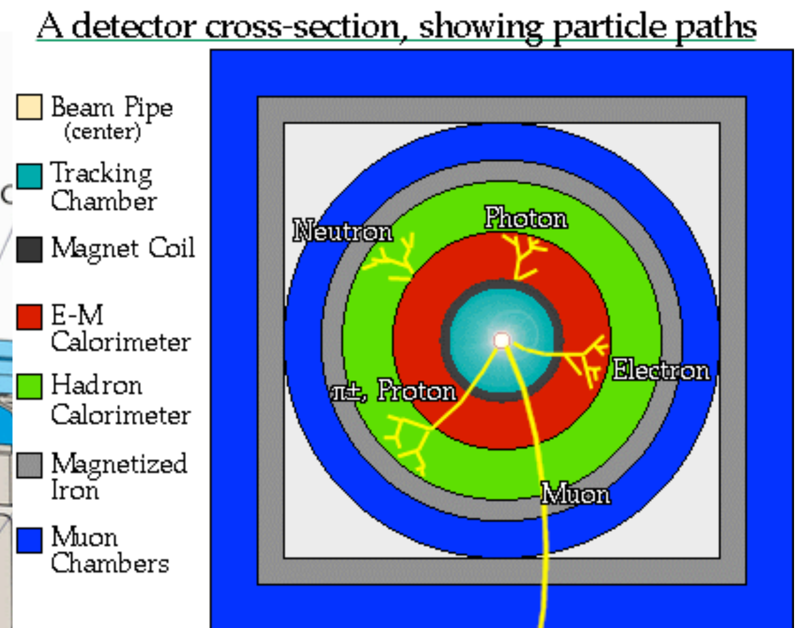
- Incoming longitudinal momentum not known:
  - "Hard interaction" is between one of the quarks and/or gluons from each proton, other quarks/gluons are "spectators"
- Longitudinal boost "flattens" event to a pancake
- We usually work in the plane transverse to the beam



ATLAS

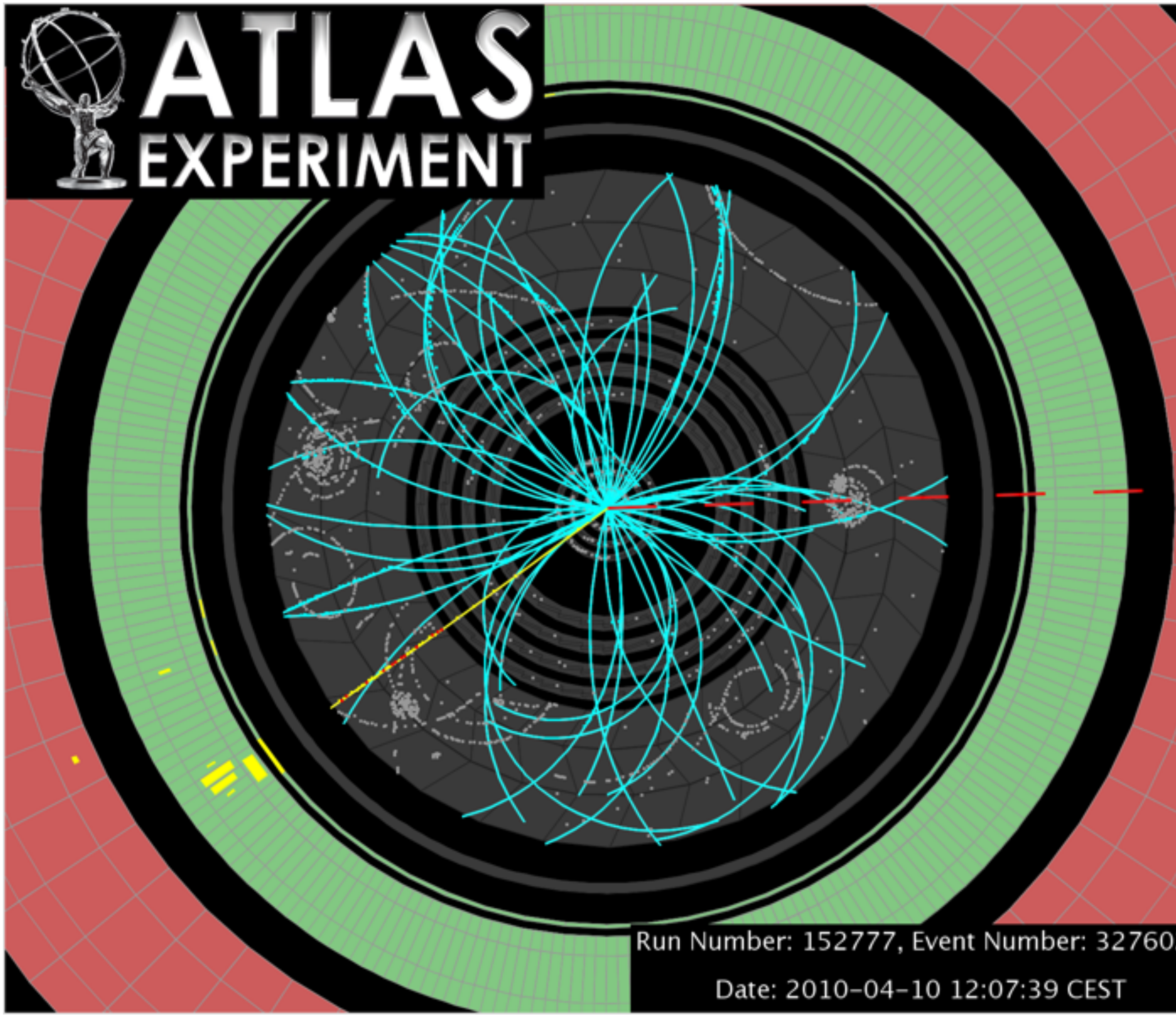
- Make best possible measurement of all particles coming out of collisions

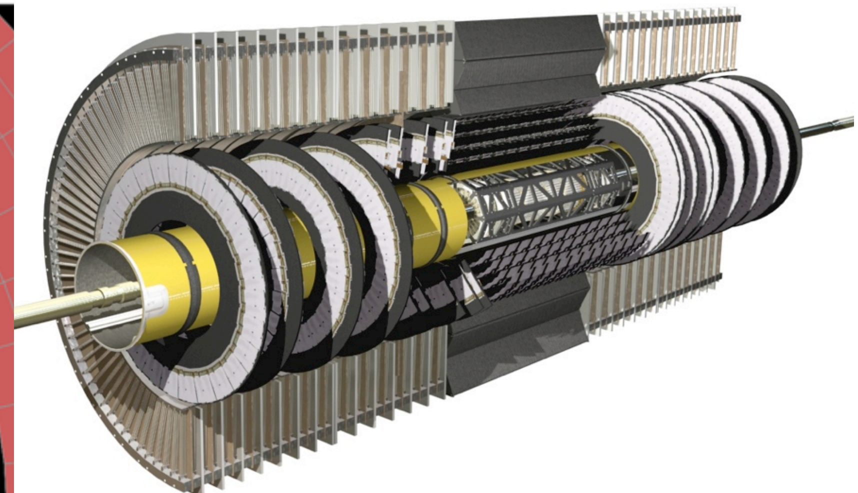


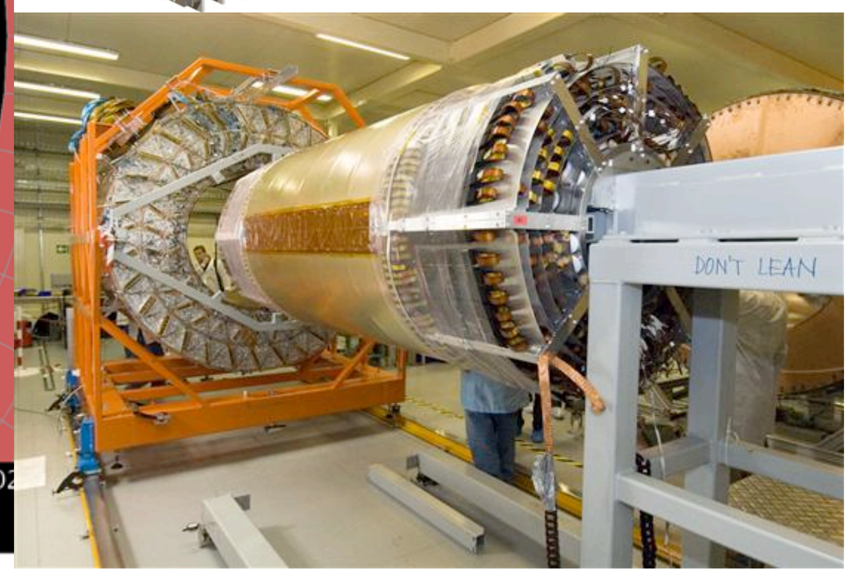


Tracking

- Combination of pixels, silicon strips (“SCT”) and straw tube transition radiation tracker (TRT)

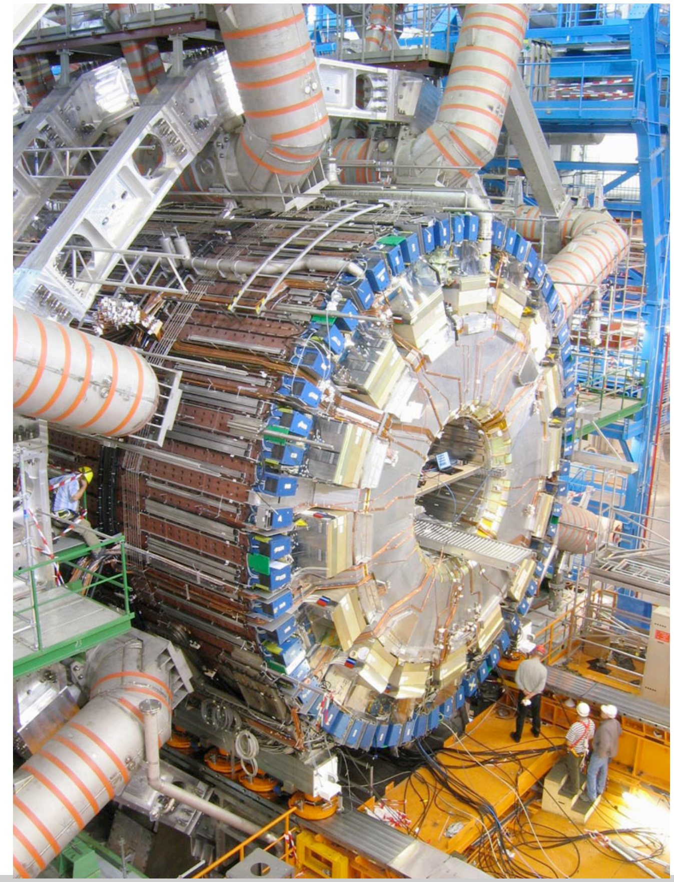


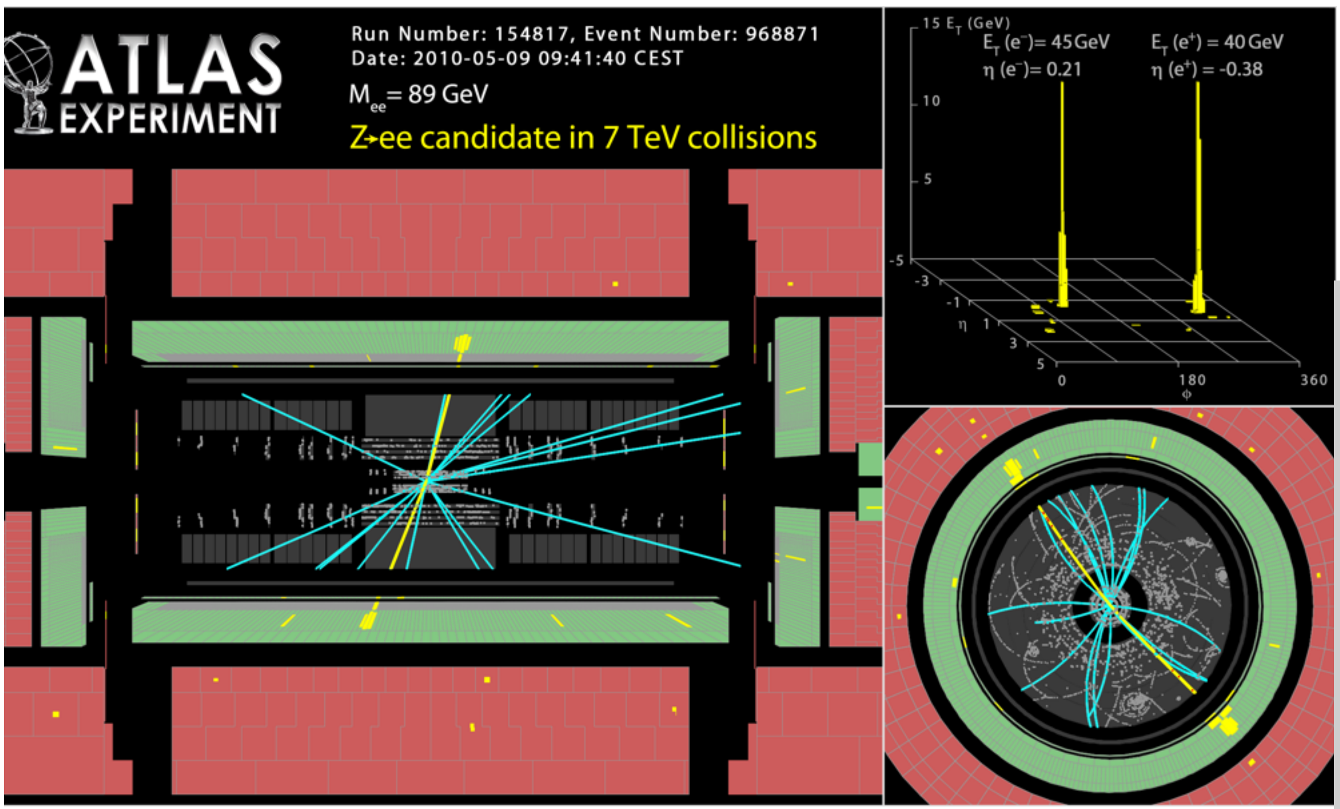


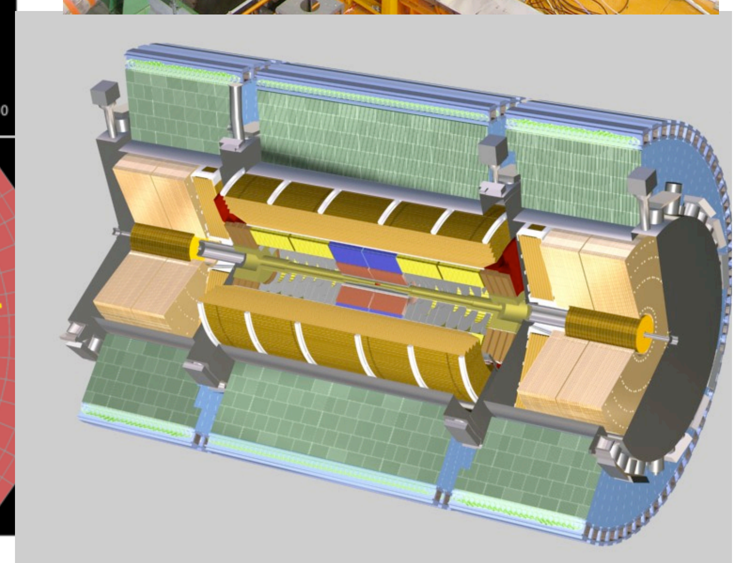


Calorimetry

- Liquid Argon & Pb accordion (EM & forward)
- Scintillator & Pb (hadronic)

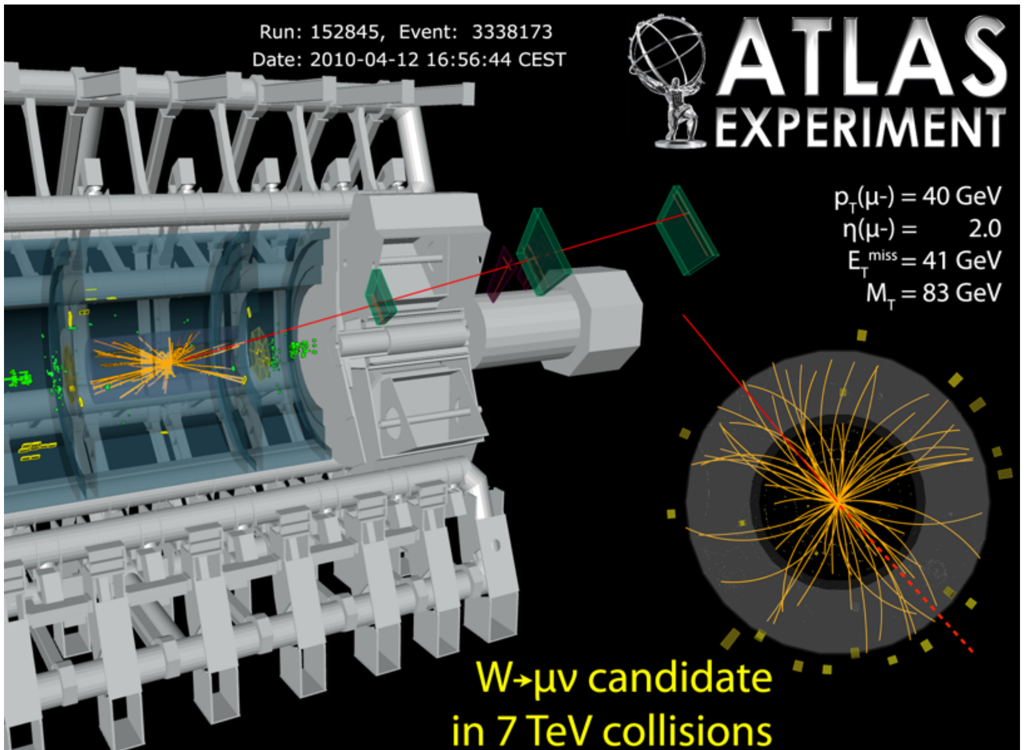


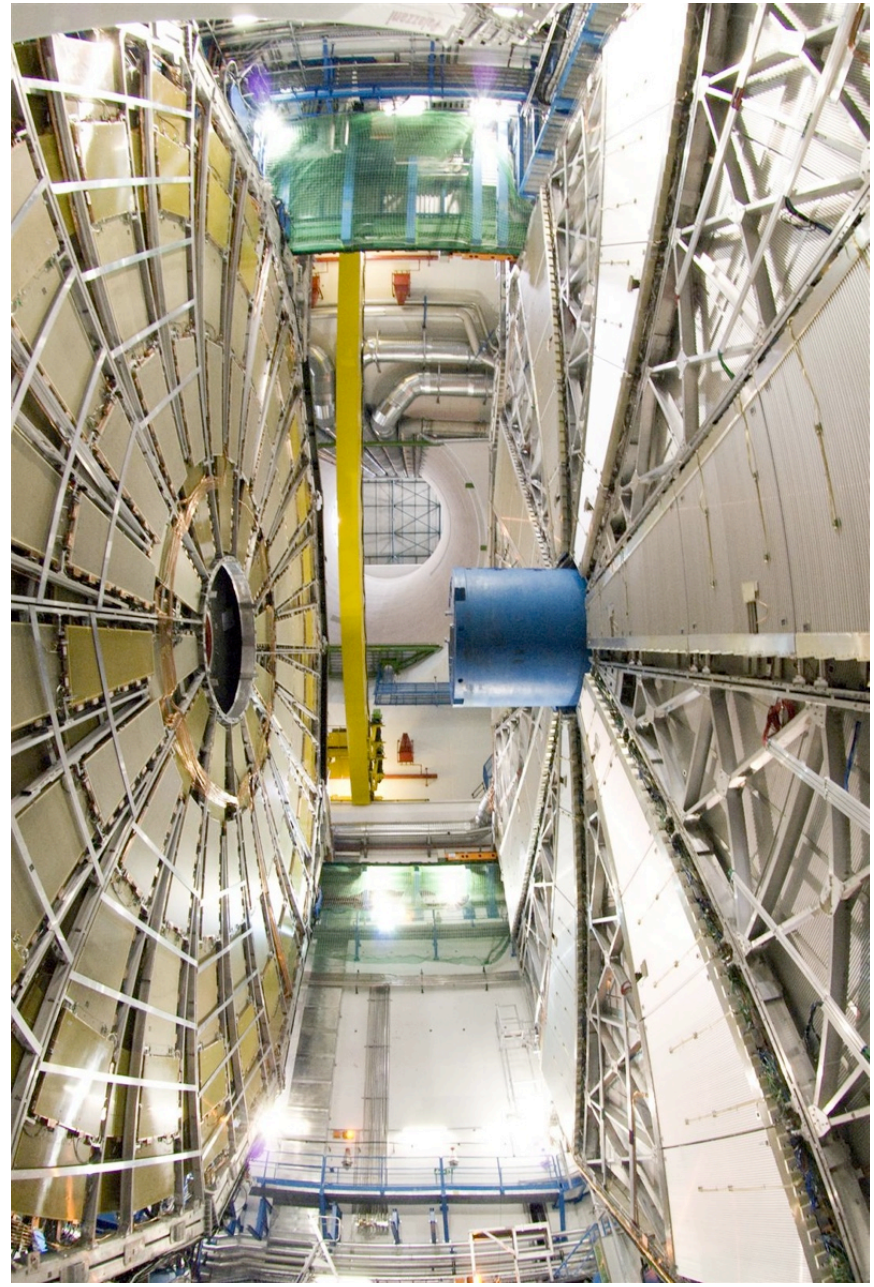




Muons

- Air-core toroids
  - Makes ATLAS big

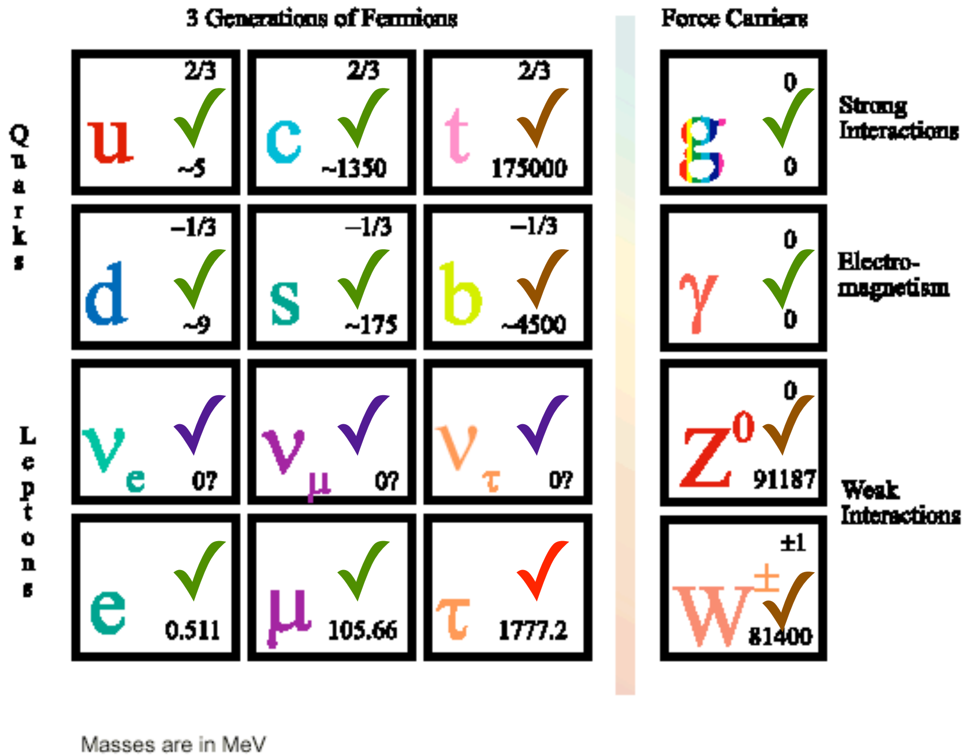




Neutrinos*

*(100% acceptance)

Detecting Particles



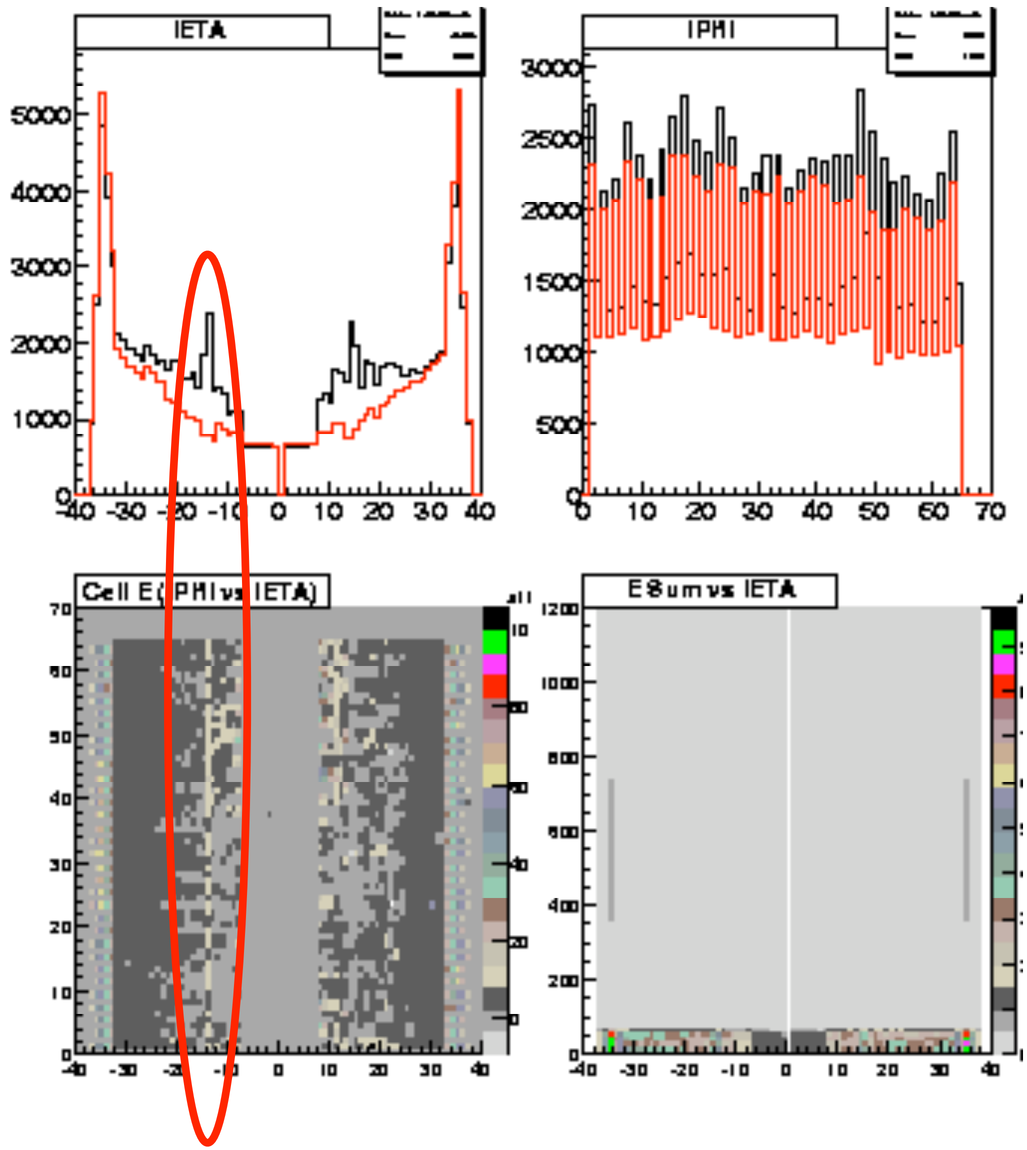
✓: Detect with high efficiency

✓: Detect by missing transverse energy

✓: Detect through decays: t→Wb, W/Z → leptons

Anecdotes From the Field (I)

- DØ's "ring of fire"
  - Noise in a few "eta rings"
  - Occurred on rare occasions
- Originally thought to be a ground fault in HV distribution
- Found to be concurrent with welding in building
  - Finally traced to liquid Argon purity & temperature monitoring



The Work

Steps in a Physics Search

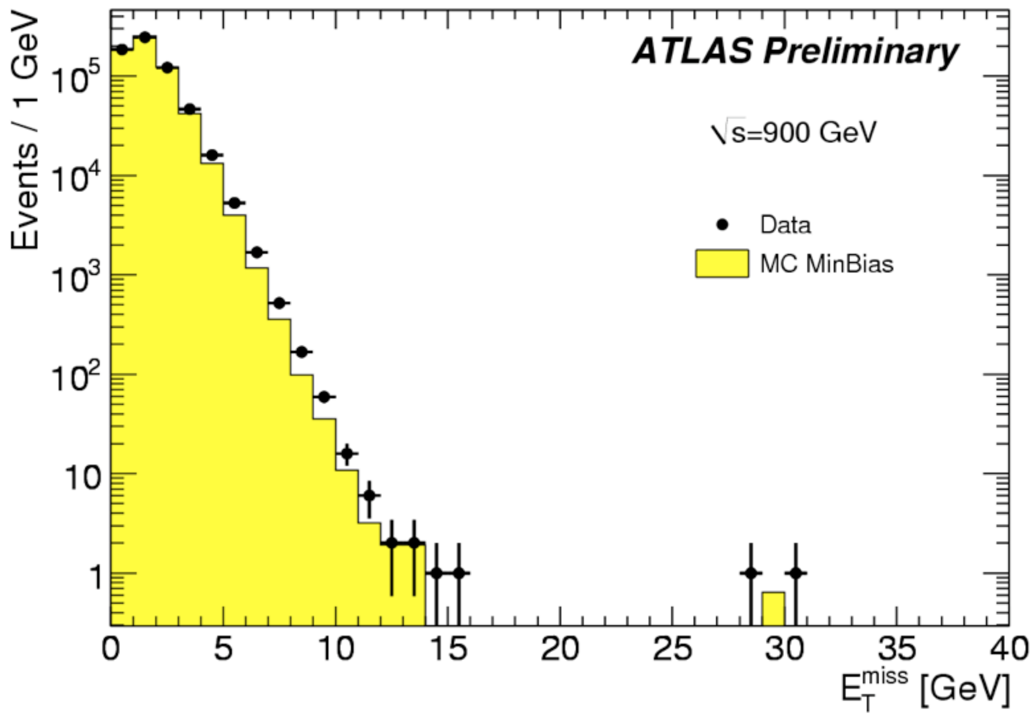
- What is the final state? $\Rightarrow$ "Preselection"
  - Sufficiently loose to be signal-poor
    - Prove you understand the detector response, physics processes contributing
  - But sufficiently tight to have a manageable data volume
    - ATLAS/CMS write 200-400 Hz $\times$ 1+ MB/event = 200-400+ MB/s
    - "4-vectors" is not enough, need some amount of detector info
    - In practice, often have preselected sample for frequent analysis, + looser sample for e.g. multijet background with rare passes
- Note that data volume ~ running time, not $\int\mathcal{L}$

Steps (II)

- Determine preselected sample's composition
  - MC and data to understand each contribution
    - QCD multijet background to leptons often extracted from data: rejection factor ~10^{-4}, difficult for simulation to be that accurate
    - MC for most other processes, with corrections from data, since generators are LO or NLO
  - Also need to correct MC for real-life data conditions
    - Different alignment, small fraction of dead channels etc.
  - As statistics increase, more difficult, since mis-modelings not hidden by large statistical uncertainties anymore

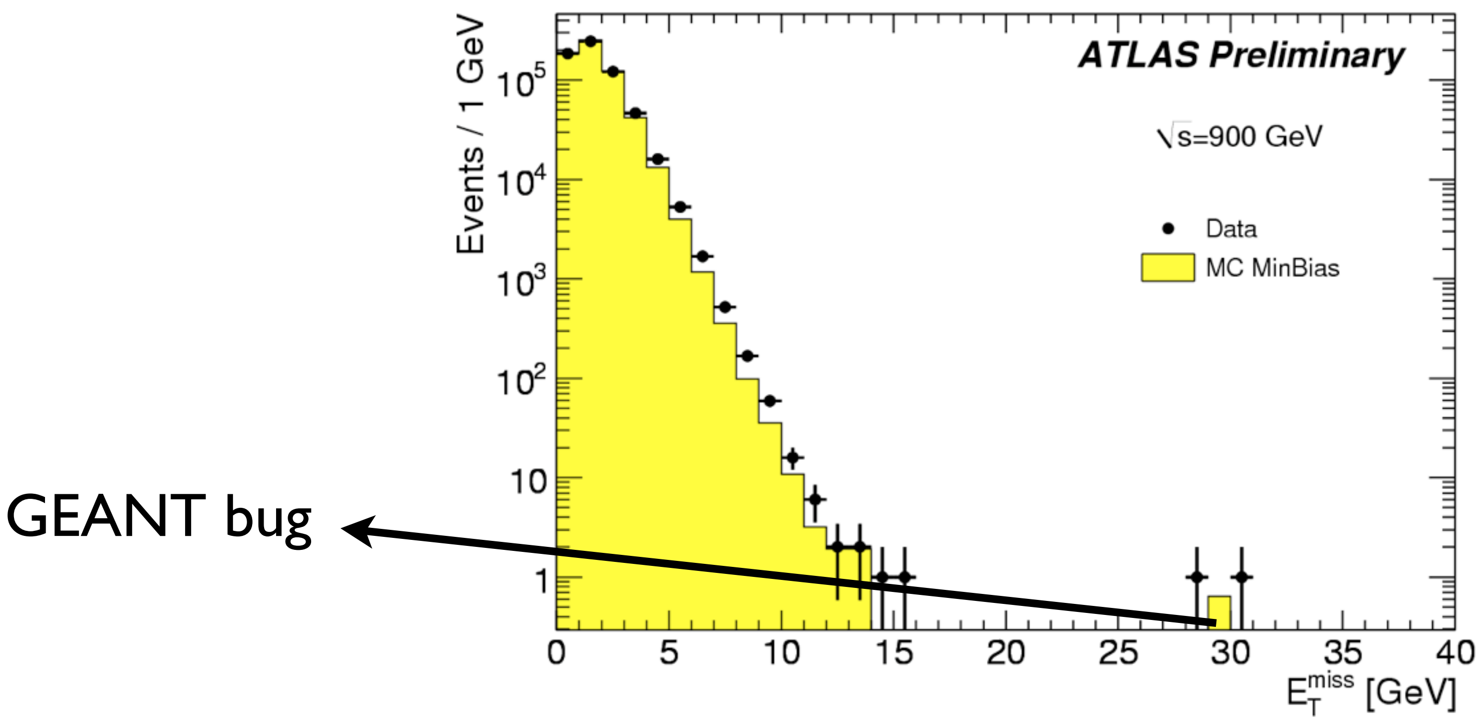
Anecdotes From the Field (II)

- Everybody wants experimenters to produce results fast
  - Lots of pressure in the early days of LHC...
    - Only jets, background composition "easy"



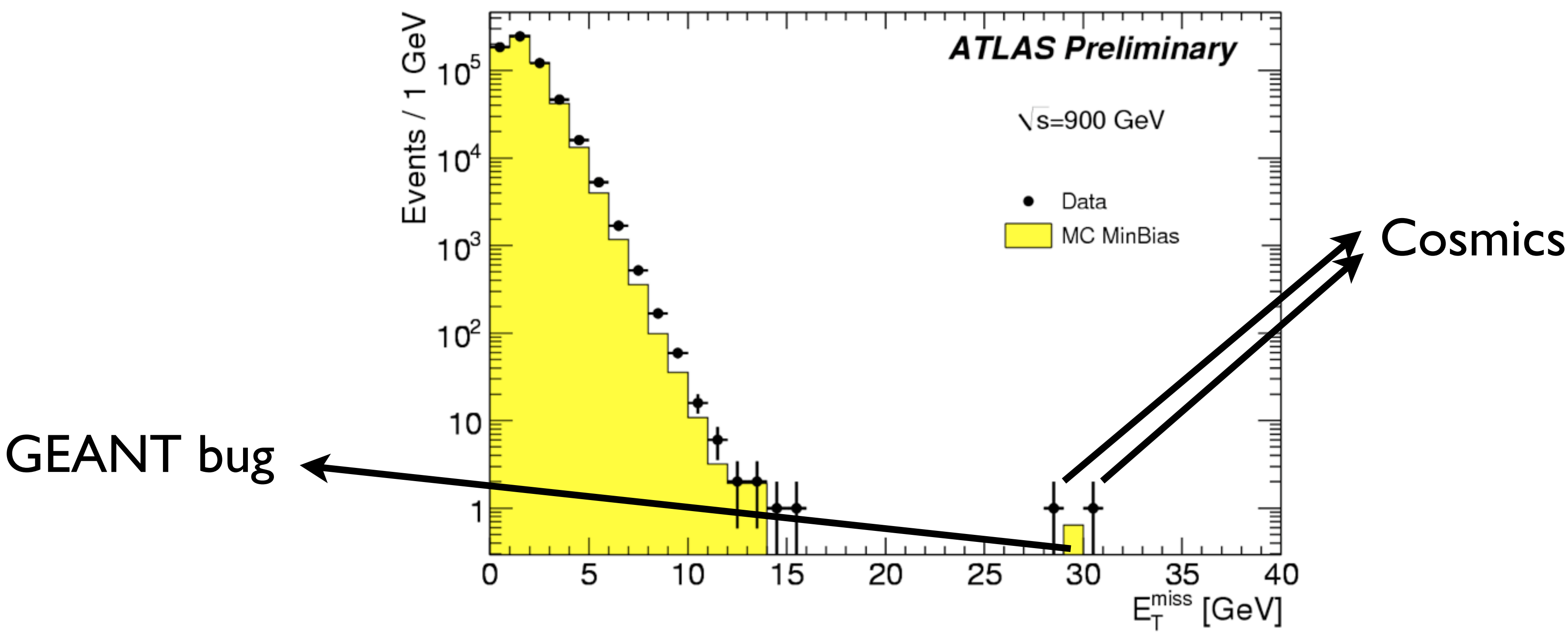
Anecdotes From the Field (II)

- Everybody wants experimenters to produce results fast
  - Lots of pressure in the early days of LHC...



Anecdotes From the Field (II)

- Everybody wants experimenters to produce results fast
  - Lots of pressure in the early days of LHC...

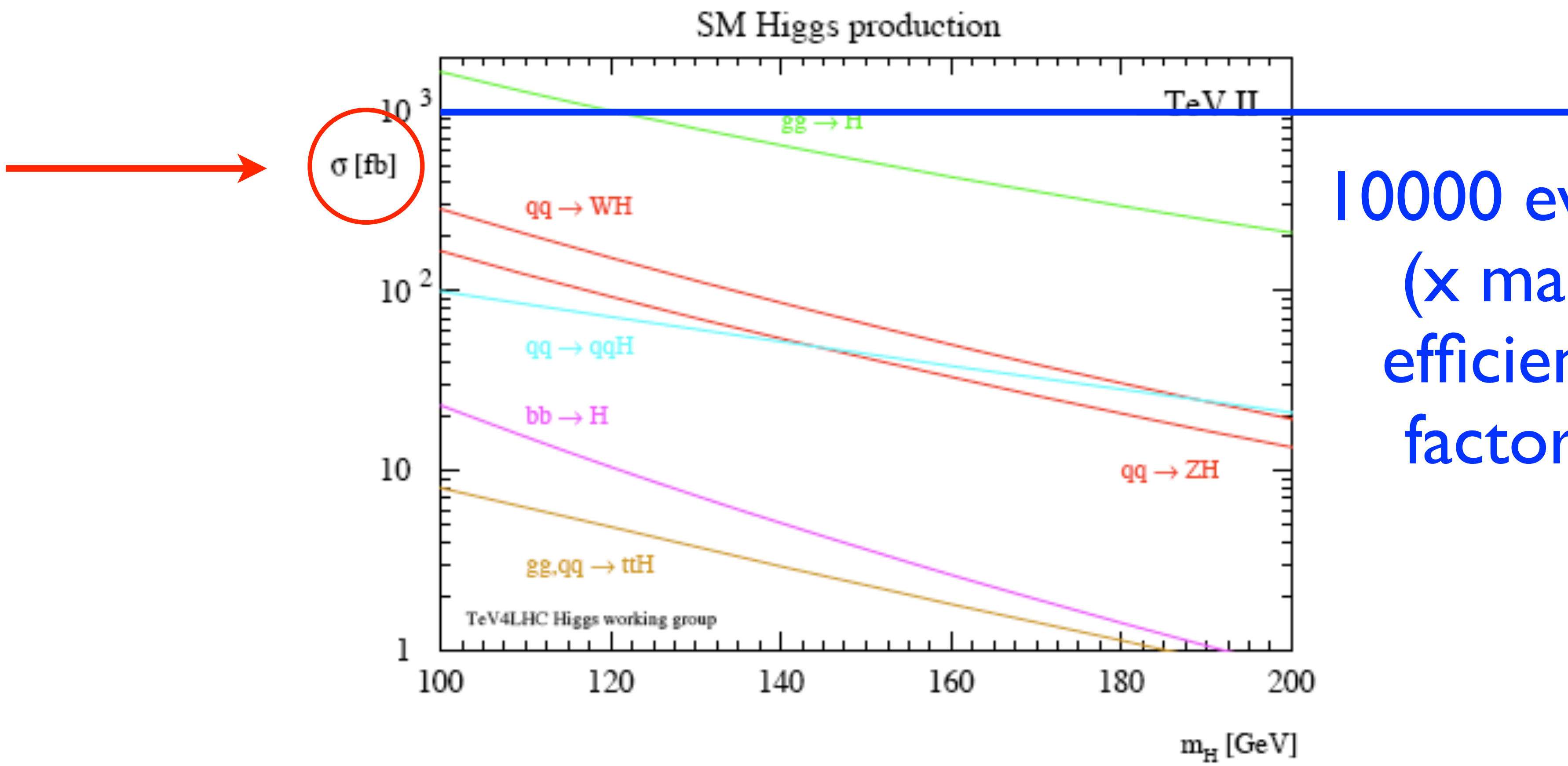


  - Sometimes, it's better to take the appropriate time to investigate

A Challenging Search: The Higgs Hunt at the Tevatron

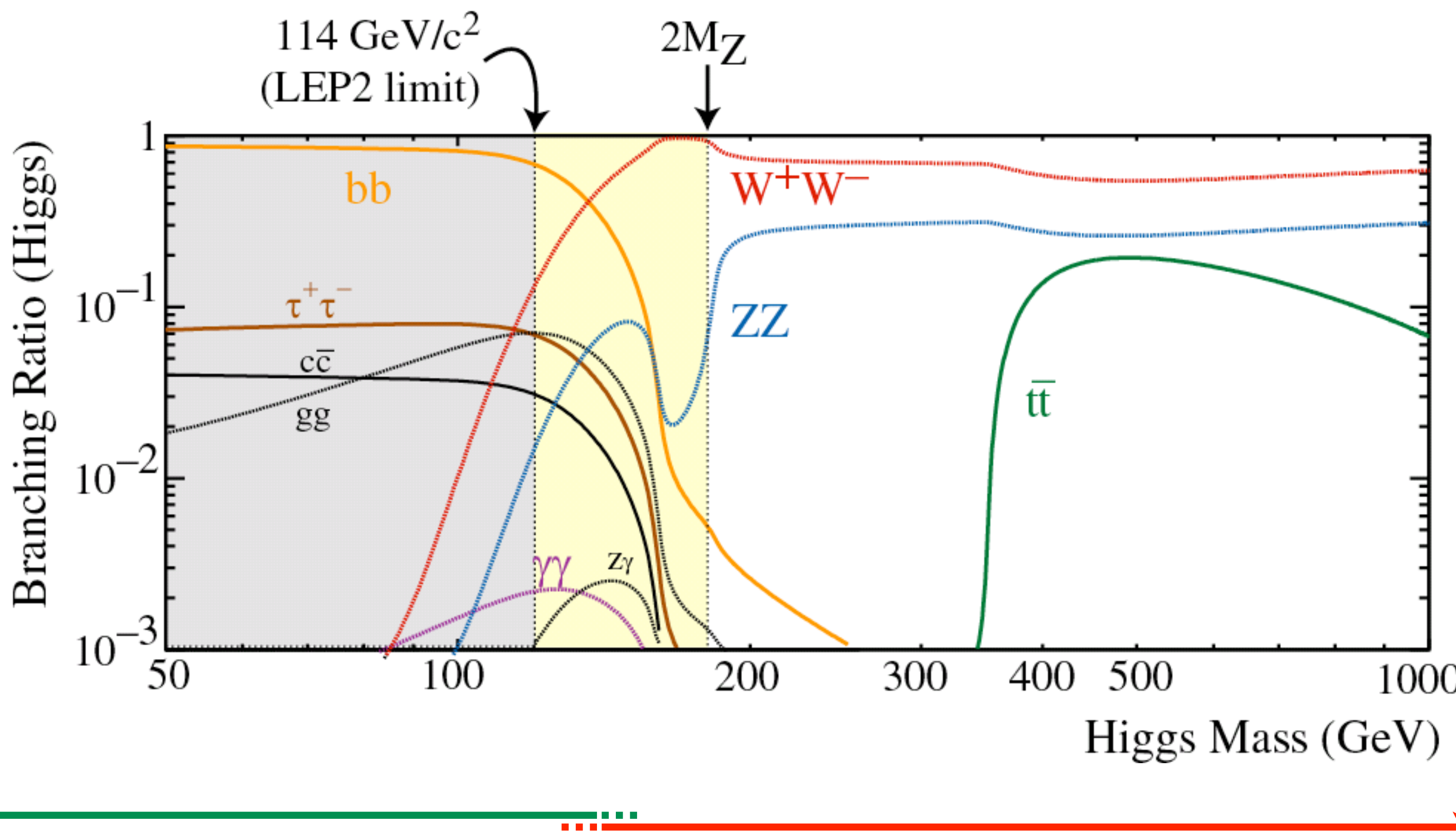
Producing Higgses

- Tevatron experiments currently have ~10 fb^{-1} of data on tape
  - (Data taking efficiency is ~90%)



10000 events (x many efficiency factors)

Higgs Decay



Low Mass

$H \to b\bar{b}$

High Mass

$H \to WW$

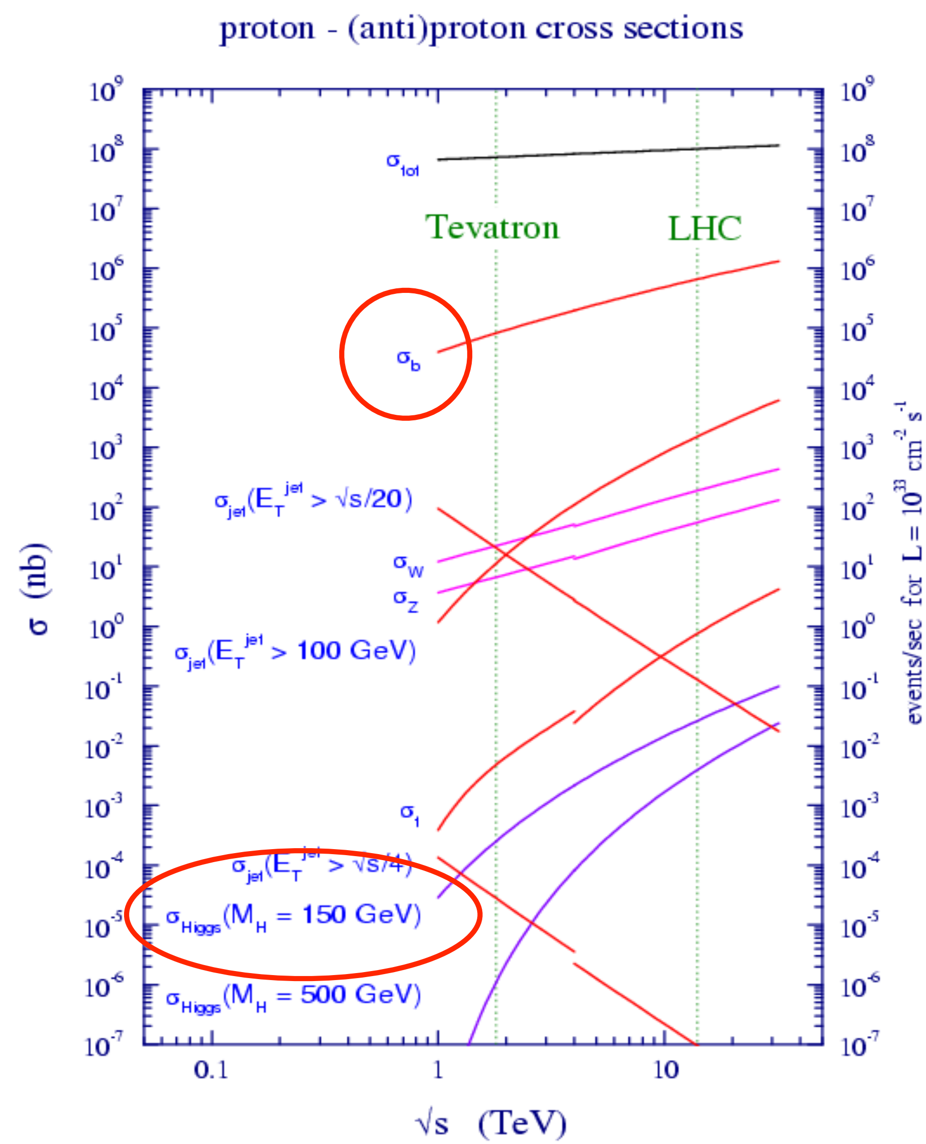
Search Channels

- Hadron colliders
  - $b\bar{b}$ production ~9 orders of magnitude larger than H
  - $gg \to H \to b\bar{b}$ swamped

➡ At low mass look for $pp \to WH$ or $ZH \to W/Z\ b\bar{b}$

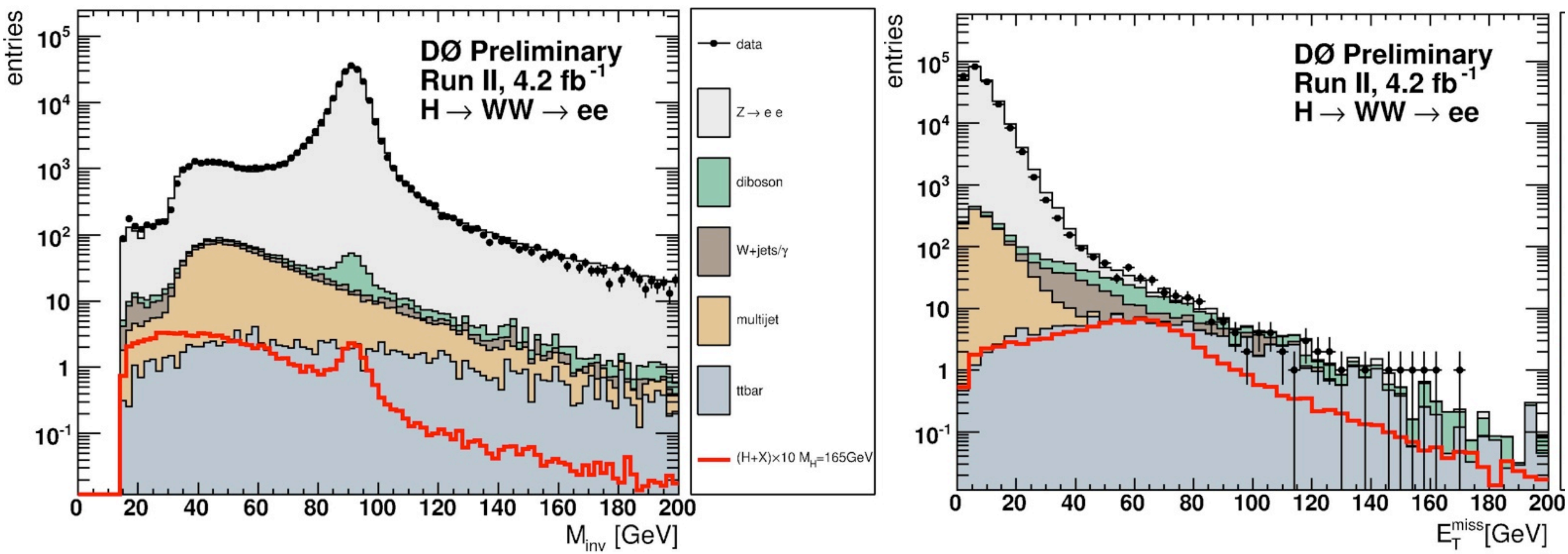
  - With leptonic W, Z decay, so # of events /~50!

- At high mass, $gg \to H \to WW$ accessible if at least one W decays leptonically



Dilepton + MET

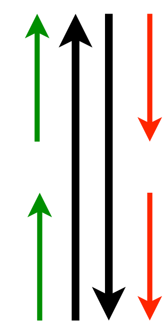
- "Golden" channel:
  - Main background $Z \rightarrow \ell\ell$ also a great reference signal
  - "Easy" to suppress using MET, angle between leptons, ...



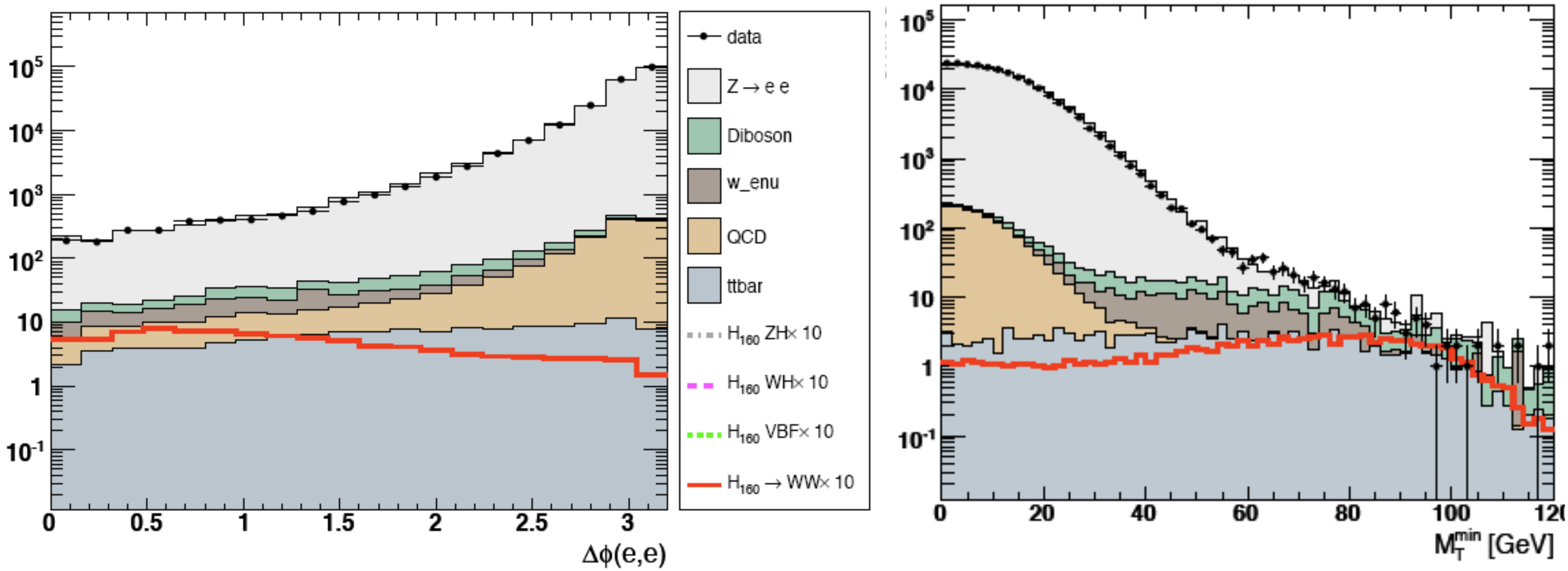
Angles

- In $Z \rightarrow \ell\ell$ (and dijets faking leptons), leptons preferentially emitted back-to-back
  - In Higgs decays, W^+W^- spins back-to back, so charged leptons in similar direction! (One LH, other RH)

Spins

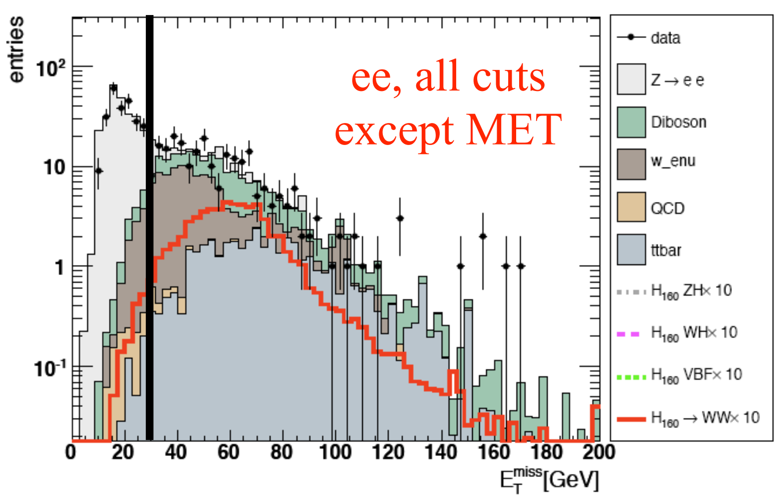


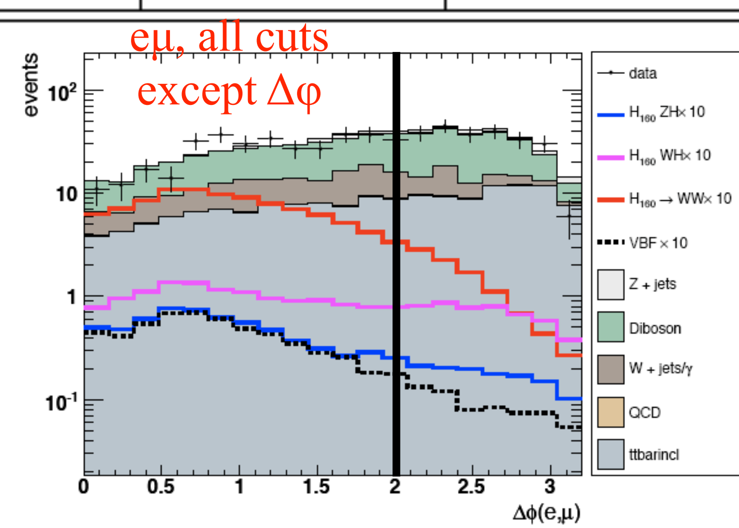
  - In Z, smallest transverse mass tends to be small



Preselection

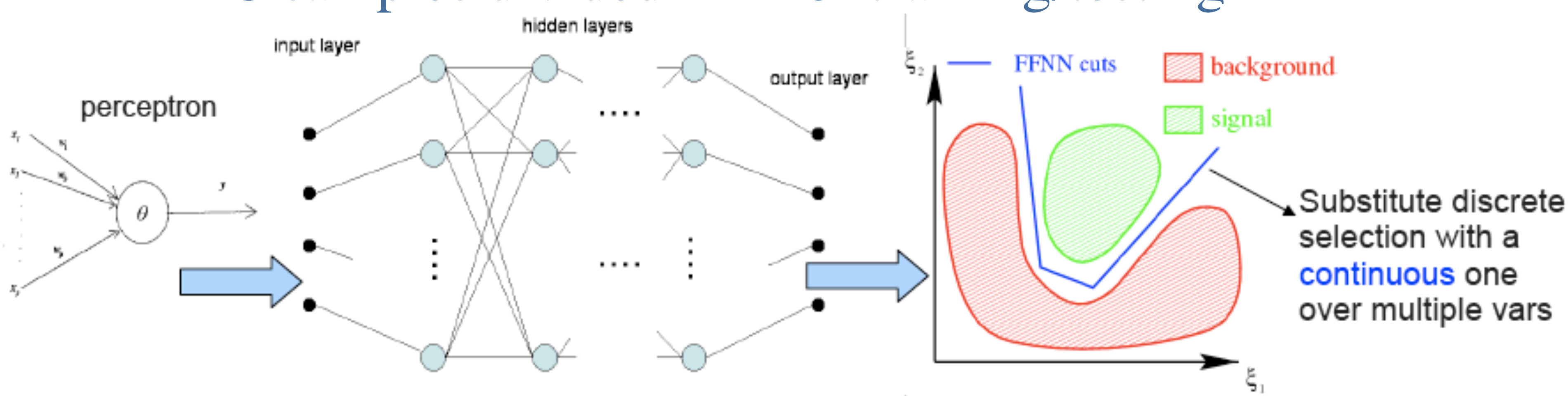
	Final state	$e\mu$	ee	$\mu\mu$
Cut 0	Pre-selection	lepton ID, leptons with opposite charge and $p_T^\mu > 10$ GeV and $p_T^e > 15$ GeV invariant mass $M_{\ell\ell} > 15$ GeV $\mu\mu$: $n_{\rm jet} < 2$ for $p_{\rm T}^{\rm jet} > 15$ GeV, $\Delta\mathcal{R}(\mu, {\rm jet}) > 0.1$ and $p_T^\mu > 15$ GeV for the leading μ		
Cut 1	Missing Transverse Energy $\not\!\!E_T$ (GeV)	> 20	> 20	
Cut 2	$\not\!\!E_T^{\rm Scaled}$	> 6	> 6	
Cut 3	M_T^{min} $(\ell, \not\!\!E_T)$ (GeV)	> 20	> 30	
Cut 4	$p_{\rm T}^{\mu\mu}$ (GeV) for $n_{\rm jet} = 0$ $\not\!\!E_T$ (GeV) for $n_{\rm jet} = 1$			> 20 > 20
Cut 5	$\Delta\phi(\ell, \ell)$	< 2.0	< 2.0	< 2.5





Multivariate Tools

- After preselection, S/B not good (~1/30, 1/50, 1/1000 in eμ, ee and μμ final states)
- Use multivariate tools to exploit correlations between observables for S ↔ B discrimination
- In the dilepton + MET ($H \to WW \to \ell\nu\ell\nu$), use neural nets
    - MC samples divided in 2 for training/testing

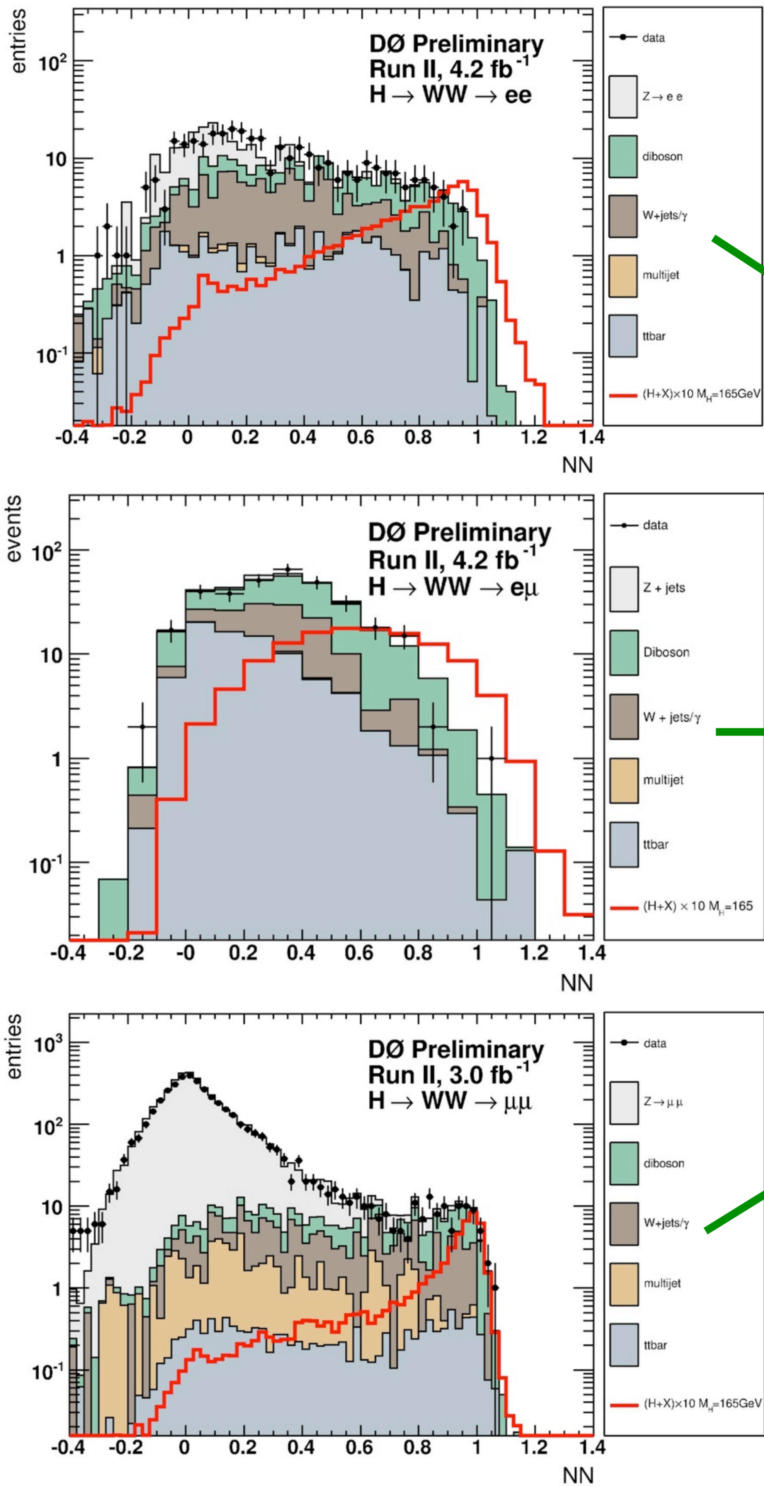


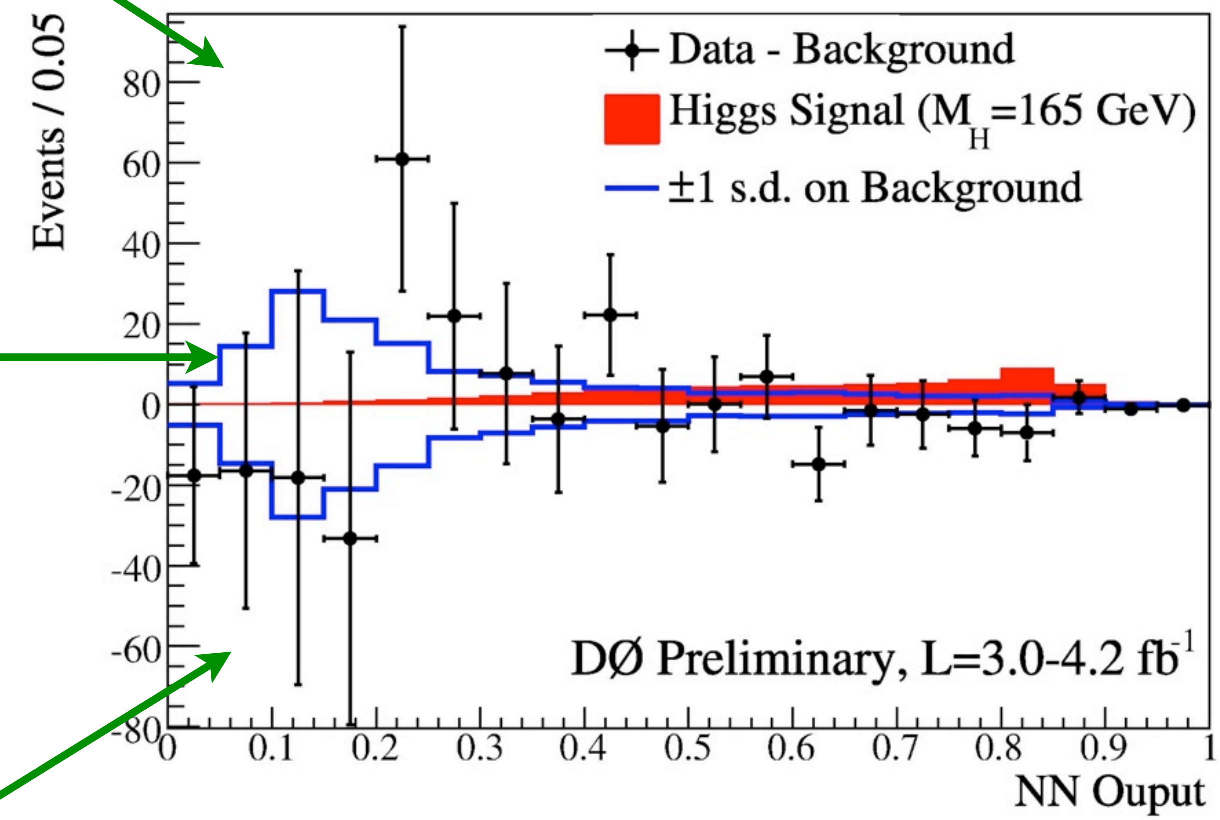
Variables

Only accept variables that are well-modeled!

NN Analysis Variables	
p_T of leading lepton	$p_T(\ell_1)$
p_T of trailing lepton	$p_T(\ell_2)$
Minimum of both lepton qualities	$\min(q_{\ell 1}, q_{\ell 2})$
Vector sum of the transverse momenta of the leptons:	$p_T(\ell_1) + p_T(\ell_2)$
Scalar sum of the transverse momenta of the jets:	$H_T = \sum_i \lvert p_T(\text{jet}_i) \rvert$
Invariant mass of both leptons	$M_{\text{inv}}(\ell_1, \ell_2)$
Minimal transverse mass of one lepton and $\not\!\!E_T$	M_T^{min}
Missing transverse energy	$\not\!\!E_T$
Scalar transverse energy	E_T^{scalar}
Azimuthal angle between selected leptons	$\Delta\phi(\ell_1, \ell_2)$
Solid angle between selected leptons ($e\mu$ only)	$\Delta\Theta(\ell_1, \ell_2)$
ΔR between selected leptons ($e\mu$ only)	$\Delta R(\ell_1, \ell_2)$
Azimuthal angle between leading lepton and $\not\!\!E_T$	$\Delta\phi(\not\!\!E_T, \ell_1)$
Azimuthal angle between trailing lepton and $\not\!\!E_T$	$\Delta\phi(\not\!\!E_T, \ell_2)$

NN Outputs


DØ Preliminary
Run II, 4.2 fb^{-1}
H → WW → ee
entries
NN
data
Z → e e
diboson
W+jets/γ
multijet
ttbar
(H+X)×10 M$_H$=165GeV
DØ Preliminary
Run II, 4.2 fb^{-1}
H → WW → eμ
events
NN
data
Z + jets
Diboson
W + jets/γ
multijet
ttbar
(H+X) × 10 M$_H$=165
DØ Preliminary
Run II, 3.0 fb^{-1}
H → WW → μμ
entries
NN
data
Z → μ μ
diboson
W+jets/γ
multijet
ttbar
(H+X)×10 M$_H$=165GeV

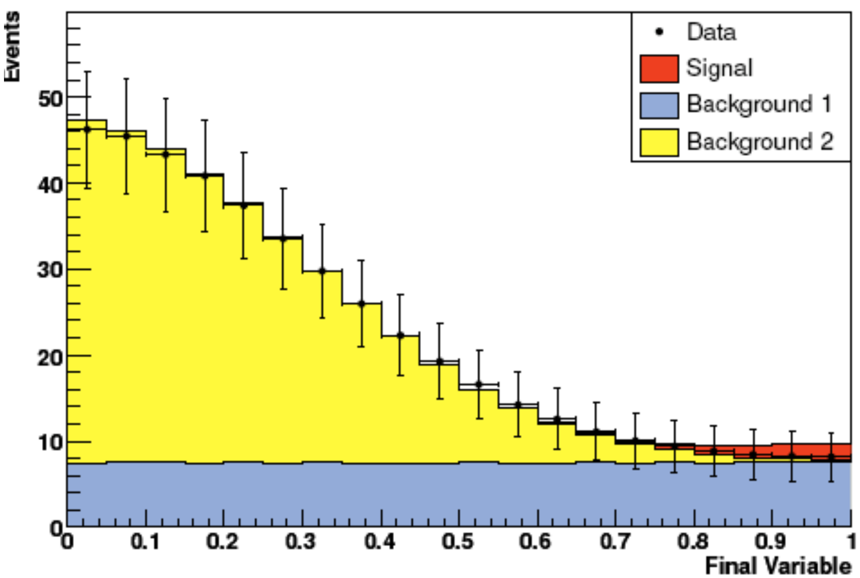

Data - Background
Higgs Signal (M$_H$=165 GeV)
±1 s.d. on Background
Events / 0.05
DØ Preliminary, L=3.0-4.2 fb^{-1}
NN Ouput

Systematics Profiling

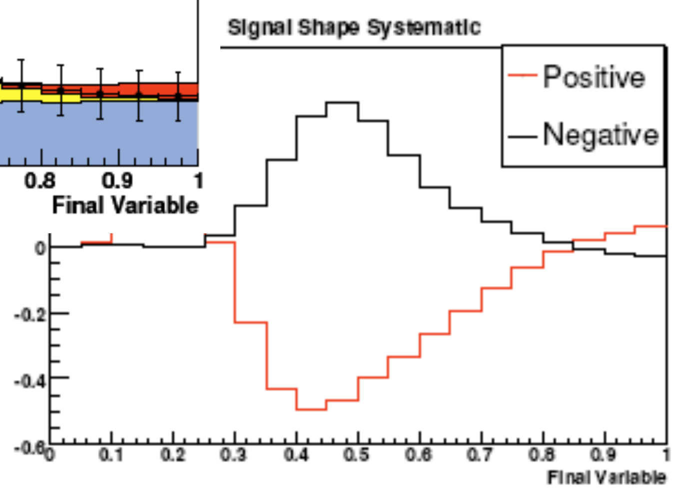
- Systematic uncertainties are propagated through the full analysis chain to the NN output distribution
  - E.g. we repeat the analysis with jet energy scale shifted up & down by 1σ
  - Some systematic uncertainties affect shape (jet reconstruction efficiency, energy scale and resolution, boson p^T distributions), others only normalization (lepton reconstruction efficiencies and momentum calibration, modeling of multijet background, theoretical cross-sections and luminosity)
  - Systematic uncertainties are treated as nuisance parameters

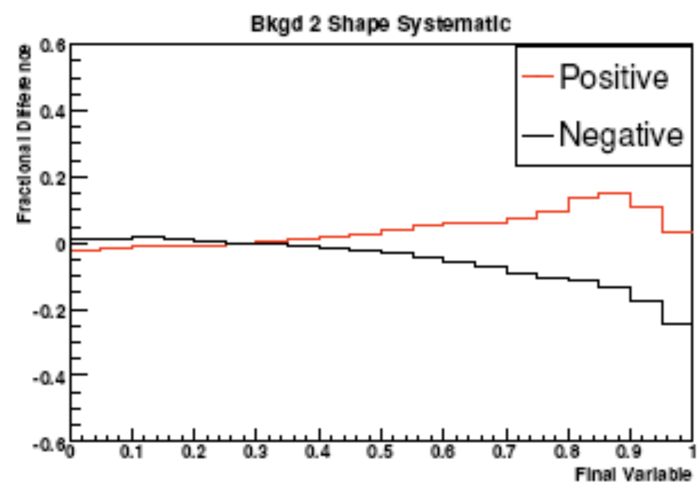
Systematics Profiling

- Nuisance parameters tend to be correlated, but not 100%, among backgrounds
  - Can affect rates, shapes, or both (in any distribution), and often asymmetric and non-gaussian



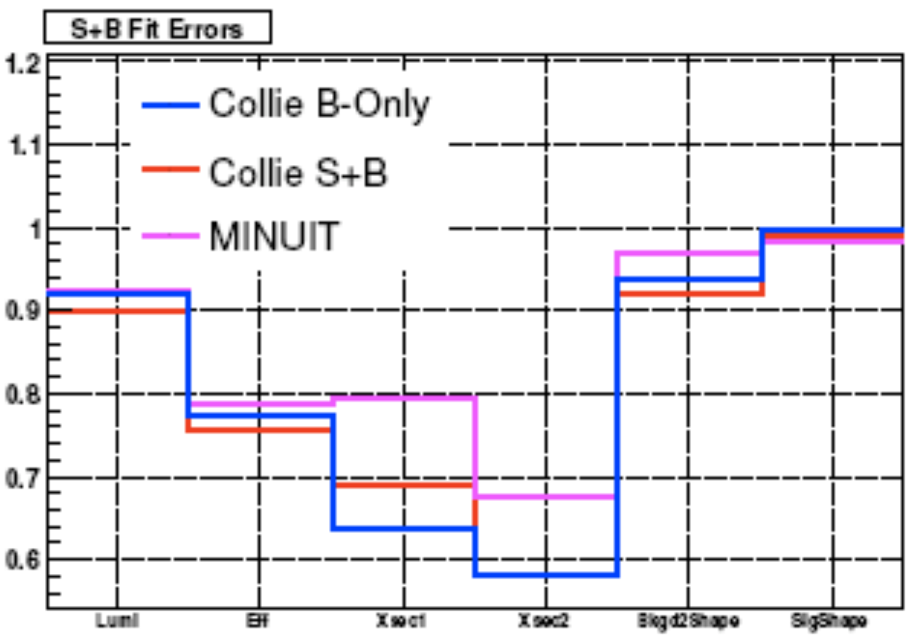
Toy Example

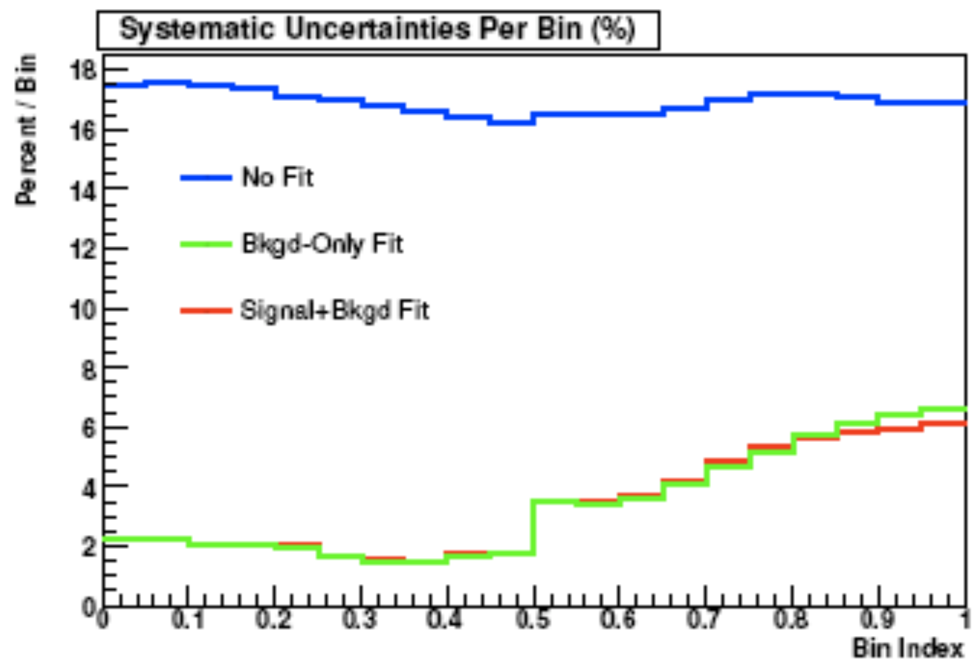




- Can generate pseudo-experiments (events in bins according to poisson), then for each experiment vary nuisance parameters
    - Variations in background (& S+B) prediction
        - Compare results to data using log-likelihood ratio
- So we can maximize likelihood ratio as a function of nuisance parameters → constraint them
    - I.e. use full shape of distribution(s) to see which background uncertainties are over/underestimated
        - Of course limited to size of statistical fluctuations
    - Can remove bins with large S/B if needed
        - Mostly important if uncertainties lead to similar shape distortions

- Test example:
  - Data constructed to disagree with background-only hypothesis (wrong estimates for background uncertainties)
  - But to agree with background-only better than signal+ background
    - Improvement quite spectacular (but by construction)

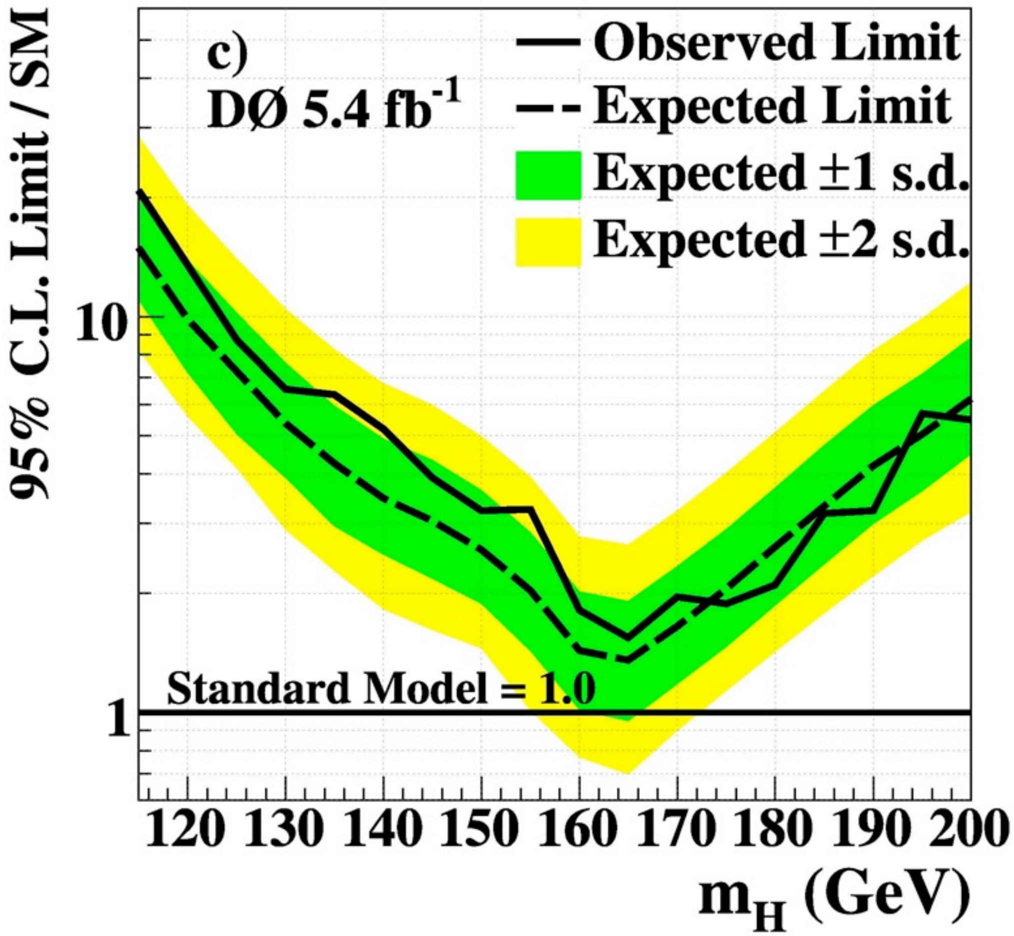




Dilepton + MET Result

- Present result as a 95% C.L. limit in units of the SM Higgs production x-section

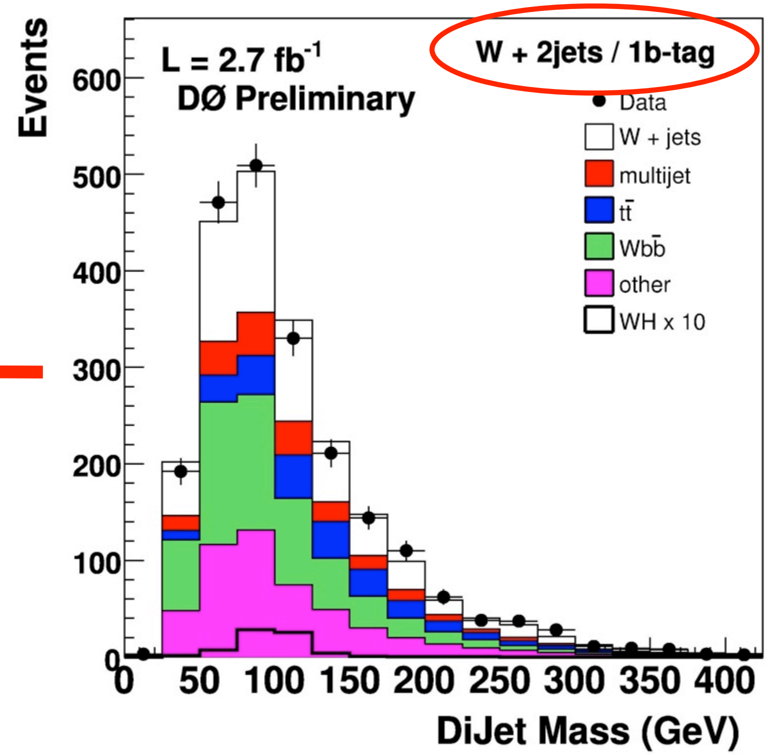
Phys. Rev. Lett. 104, 061804 (2010)



Wjj and the Higgs

- The final state consisting of W + 2 jets is critical
  - At low mass (WH, H→bb), they're b-jets with $m_{bb} = m_H$
  - At high mass (H→WW), $m_{jj} = m_W$, $m_{WW} = m_H$
- But dijet mass resolution is so-so:

And lots more background!



Sample Composition

- After preselection, low S/B allows to verify shapes of dominant backgrounds
  - For WH, first before *b*-tagging, then with 1 tag
- Determining the sample's composition
  - I.e. which processes contribute, and how

Increasing difficulty ↓

- Diboson from MC simulation (usually small, + "trust" MC)
- Top from simulation (relatively small @ Tevatron)
- Z+jets from data & MC ("easy" to get a clean sample, correct MC)
- QCD multijet from data (no choice)
- W + jets from MC, but ....

Generators Used

- We use four kinds of Monte Carlo generators
  - "Calculators" (often NNLO) do not actually generate events, they just calculate some (limited) distributions, like W p^T
  - Traditional $2 \rightarrow 2$ generators: LO, e.g. $q\bar{q} \rightarrow WZ$
    - Include parton shower, i.e. QCD radiation, and hadronization to jets
  - "Matrix Element" $2 \rightarrow n$ ($n < 9$): LO, e.g. $q\bar{q} \rightarrow e\nu jjjj$
    - Necessary to generate events with multiple hard jets
    - Require matching to parton shower to avoid double counting
  - NLOwPS $2 \rightarrow 2$ generators: include NLO corrections
    - I.e. in a sense they are $2 \rightarrow 2$ & 3 <u>with</u> virtual corrections

Correction Factors

- Of course, the ME's are LO, so "K-factors" needed
  - Different ones for heavy flavor etc..... convention to avoid confusion....

• K-factor is purely theoretical, and denotes a (N)NLO/LO ratio of cross sections;

• K'-factor is also theoretical, and denotes a (N)NLO/LL ratio of cross sections.
According to Steve, ALPGEN cross sections are Leading Log;

• S-factor is empirical, and comes on top of K or K' to bring MC in agreement with data. MC should be initially normalized to luminosity, and all correction (a.k.a. scale) factors should be applied (trigger, ID…);

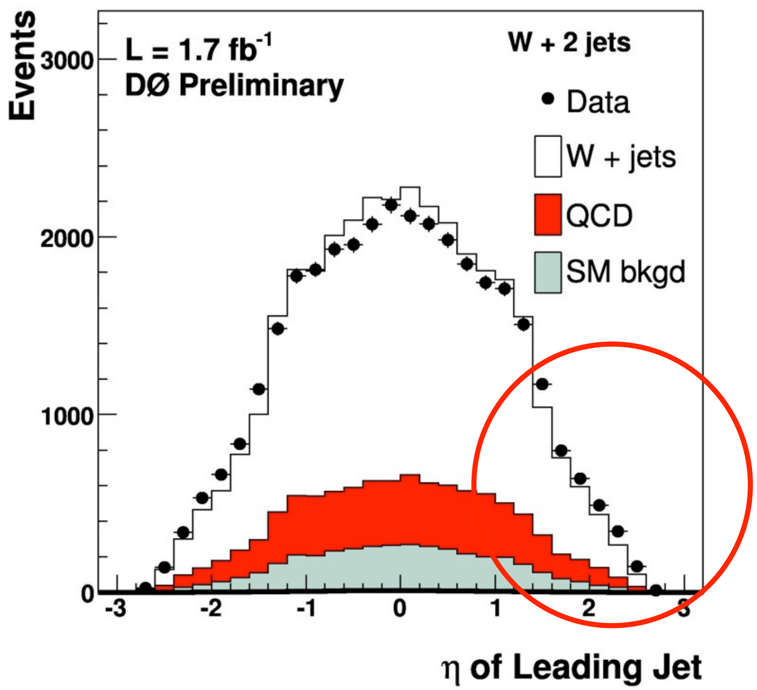
• HF-factor is, in principle, theoretical, but in practice only theory inspired.
It tells you by how much heavy flavor production should be increased, on top of K or K', and possibly S;

• S_HF-factor is empirical, and comes on top of K or K', S, and HF, to bring MC in agreement with data, after b-tagging.

In addition to W/Z p_T reweighting

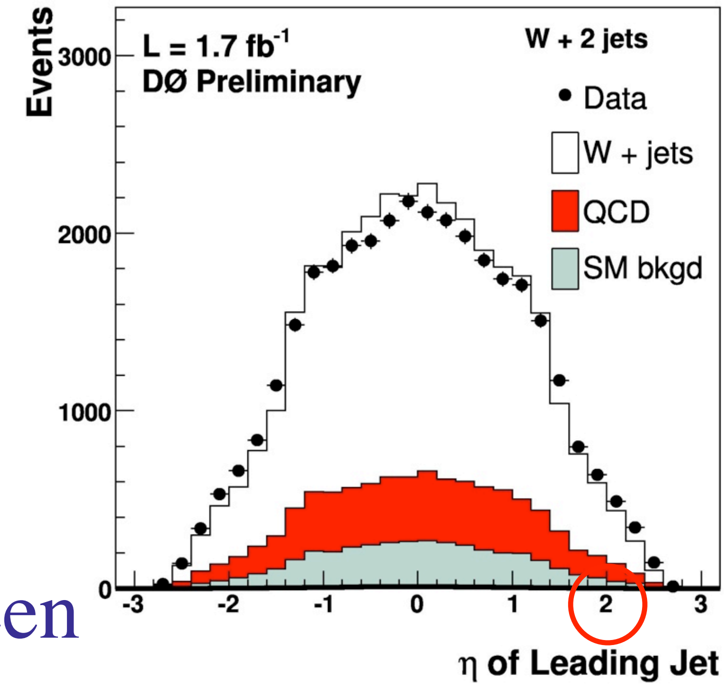
Anecdotes From the Field (III)

- Pile-up events ("minimum bias") do produce jets
  - At high $\mathscr{L}$, require that tracks pointing to jets originate from same vertex as lepton
  - High η excess disappeared!
  - Eta-dependence of jet-vertex match turns out to have shape very very similar to excess
  - After correcting for this, excess is back....

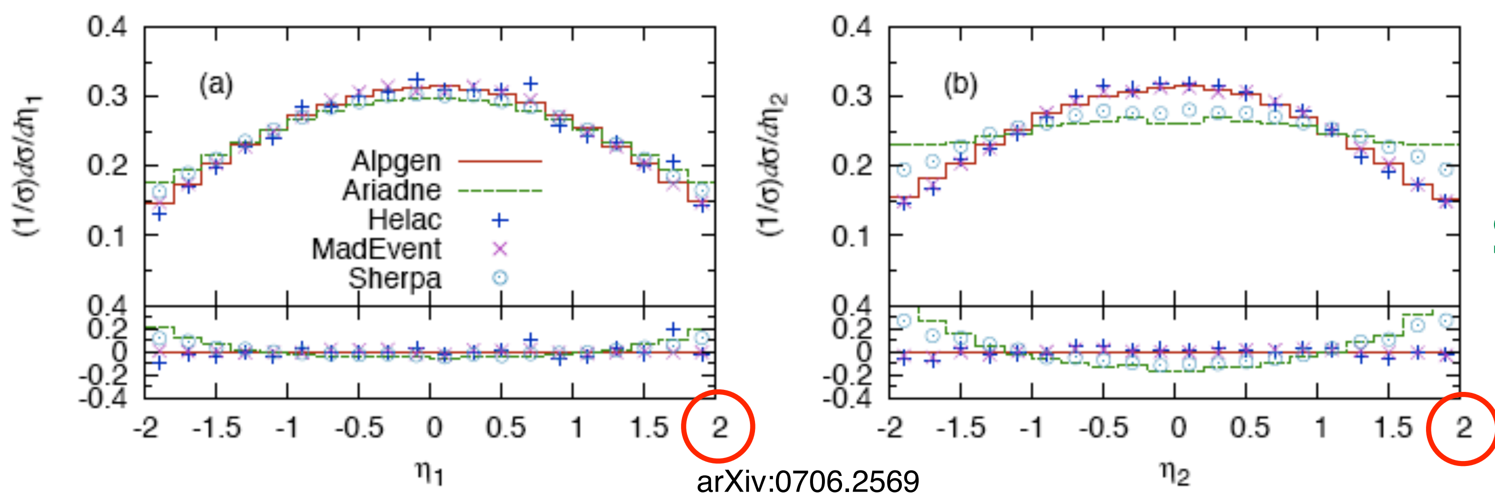


So...

- After all K/K'/S/HF-factors and boson p_T reweighing:



- Similar angular differences between generators: reweigh alpgen to sherpa



arXiv:0706.2569

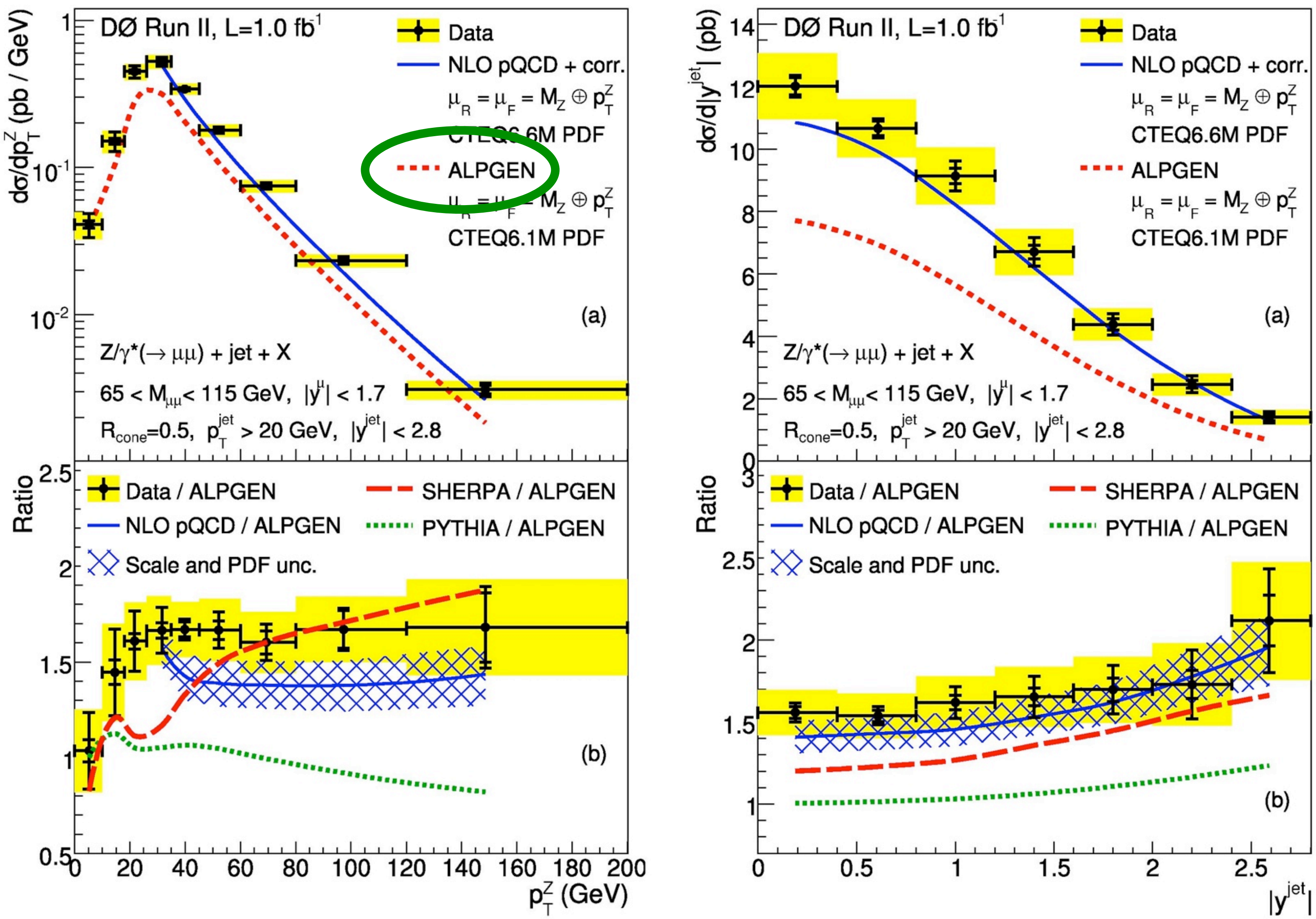
Alpgen, MadEvent, Helac with MLM, Sherpa and Ariadne with CKKW matching

Z (→ℓℓ) + jets

- Can get a clean sample, check if our simulation reproduces the data

⇒ yes, with ~expected deviations

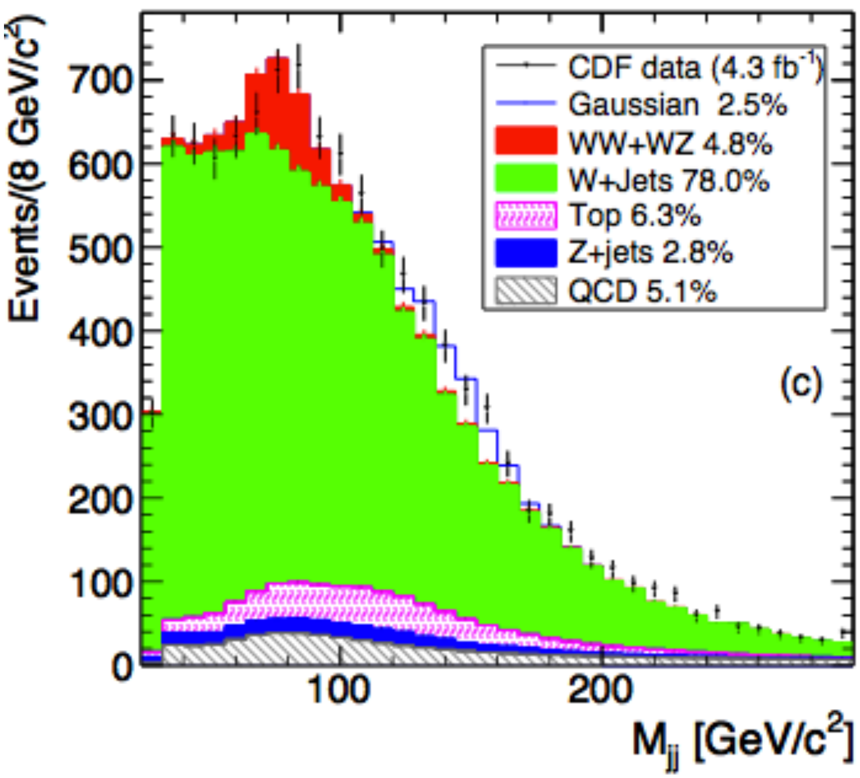
Need reweighing of MC

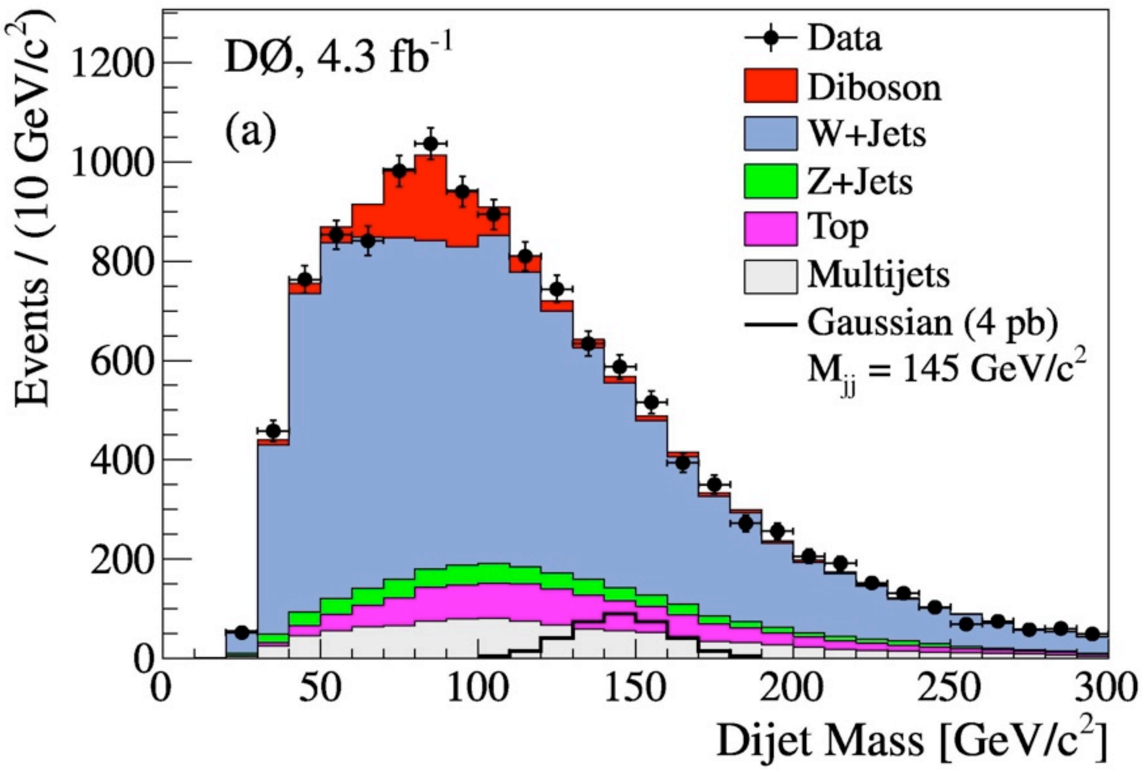


Anecdotes From the Field (III)

- CDF searched for WW/WZ in $\ell\nu jj$

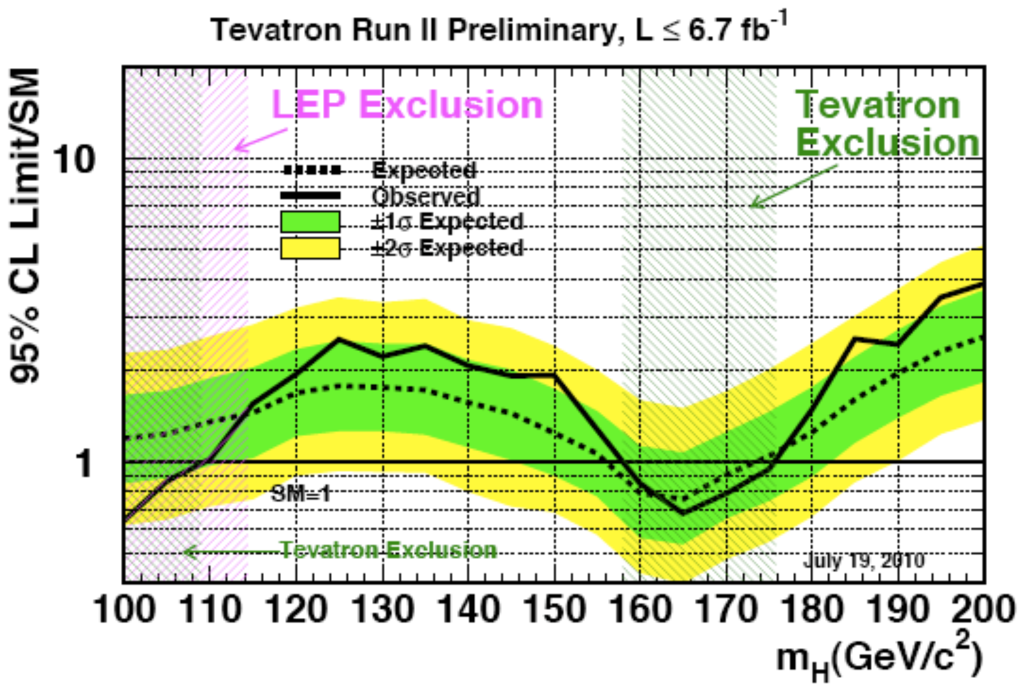
Phys.Rev.Lett.106:171801

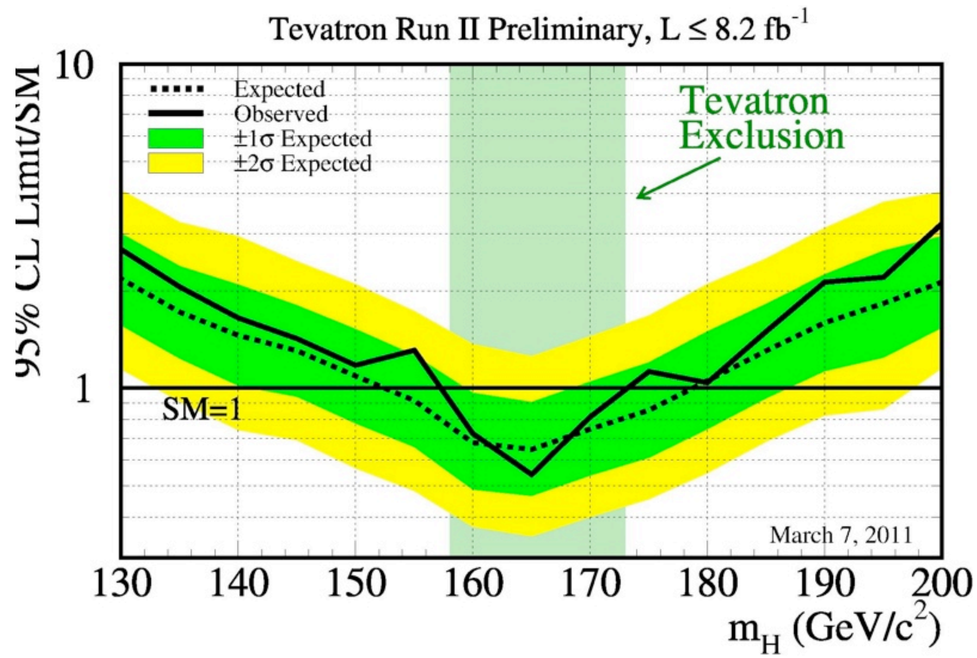




- The background here is not SM, it is uncorrected alpgen!!
  - But this is not the issue.....

All Channels, CDF + DØ





- Low mass a struggle!

129 Channels!!

So, Physic Analysis

- Start from:
  - "*How well* do we understand data *and* the SM?"
    - How confident are we in corrections we apply?
- Given that:
  - Which measurements can we make?
  - What do we need to do to improve our understanding?

Complementary measurements!

- Balance the work!
  - Early, low background searches
  - Detailed understanding/verification of SM predictions