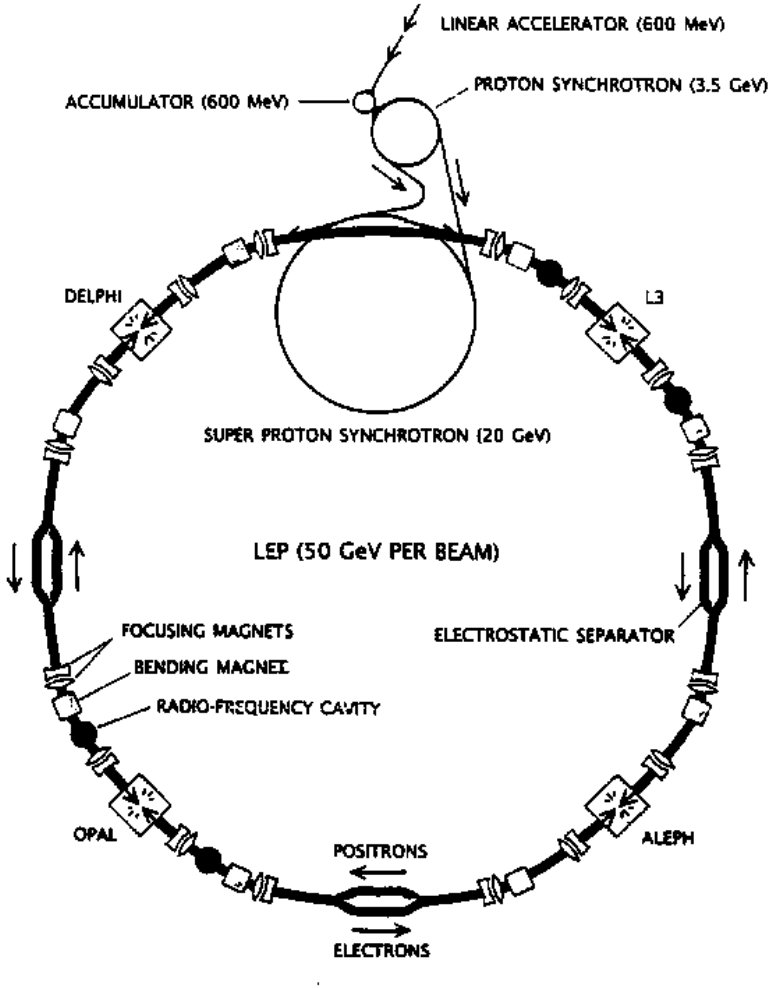
Collider Experience

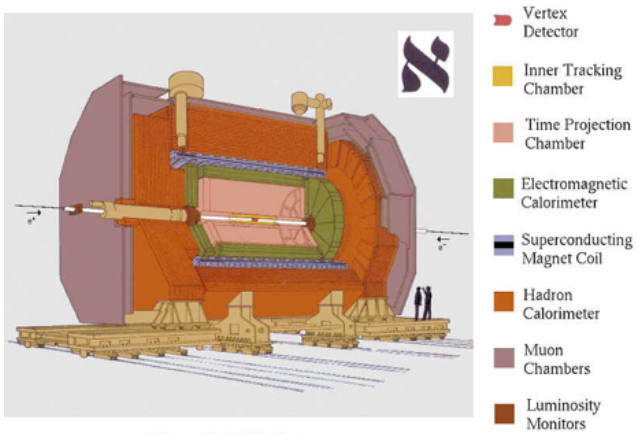
Thomas R. Junk

Fermilab

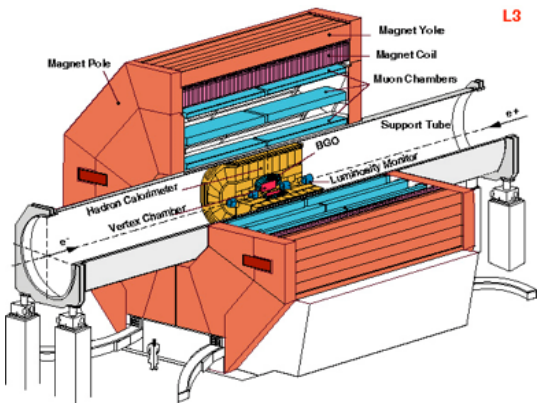
Phystat-Nu 2016, Sep. 20, 2016

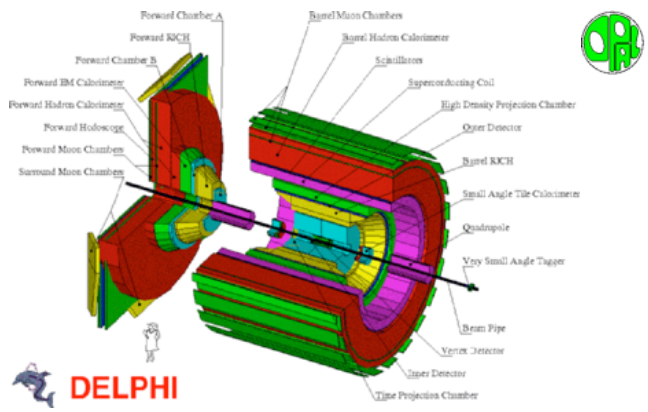
LEP: e^+e^- at up to 209 GeV. Four Detectors

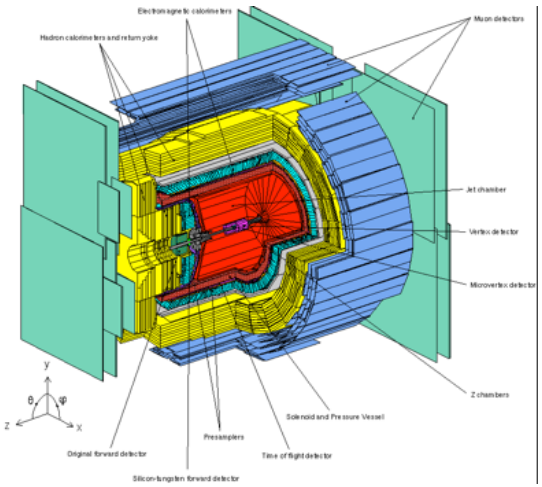




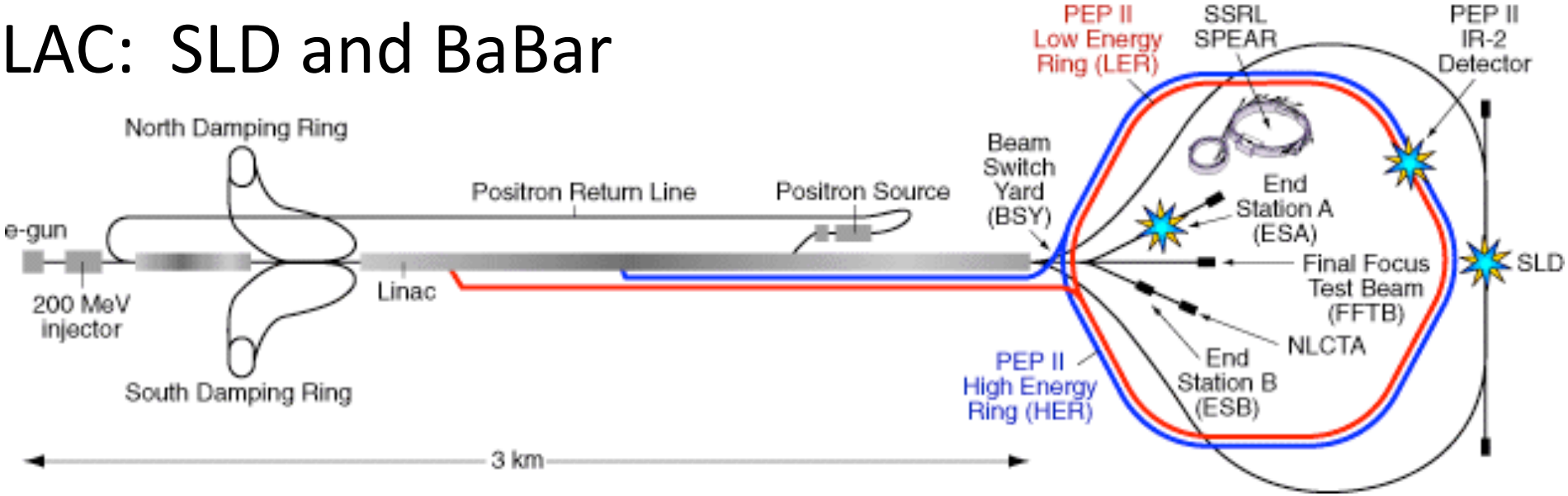
The ALEPH Detector

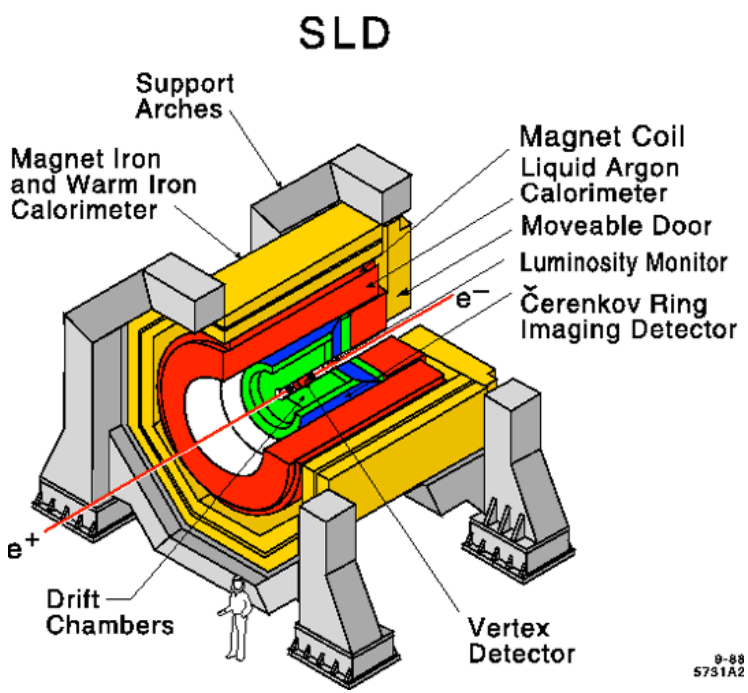




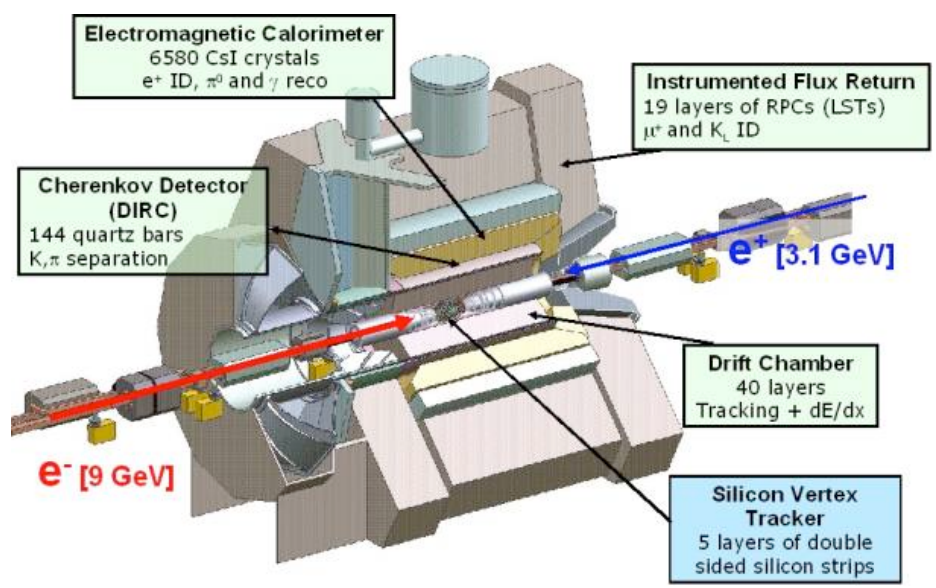


SLAC: SLD and BaBar

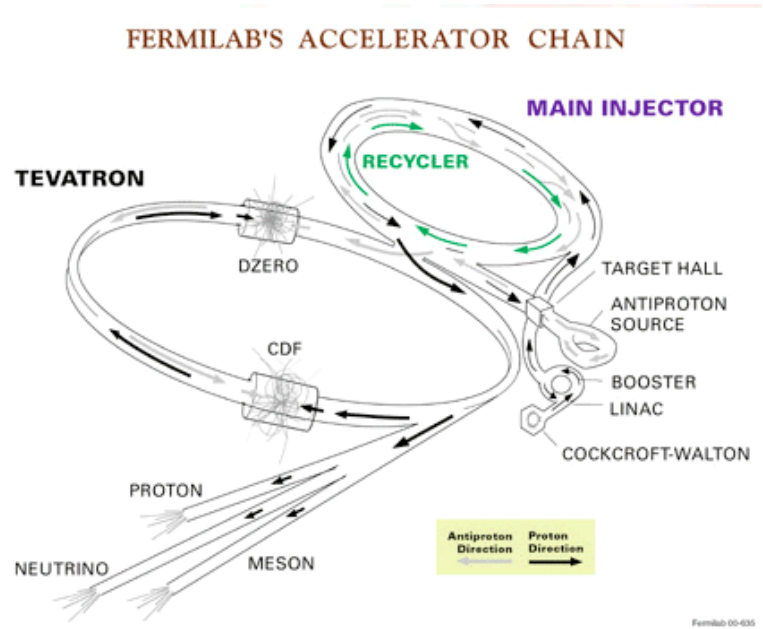




BaBar



The Tevatron and Associated Accelerators



Cockroft-Walton

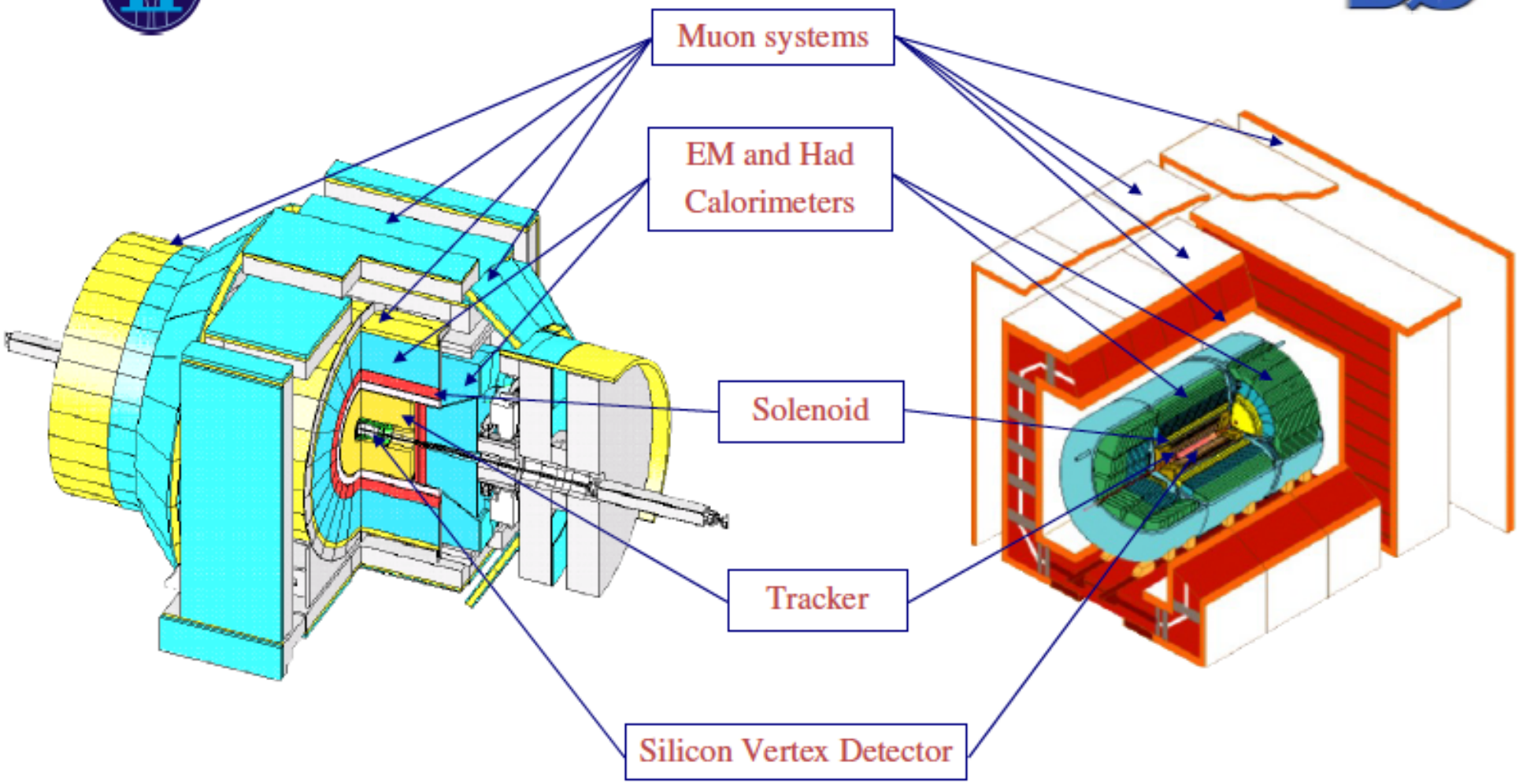
Linac

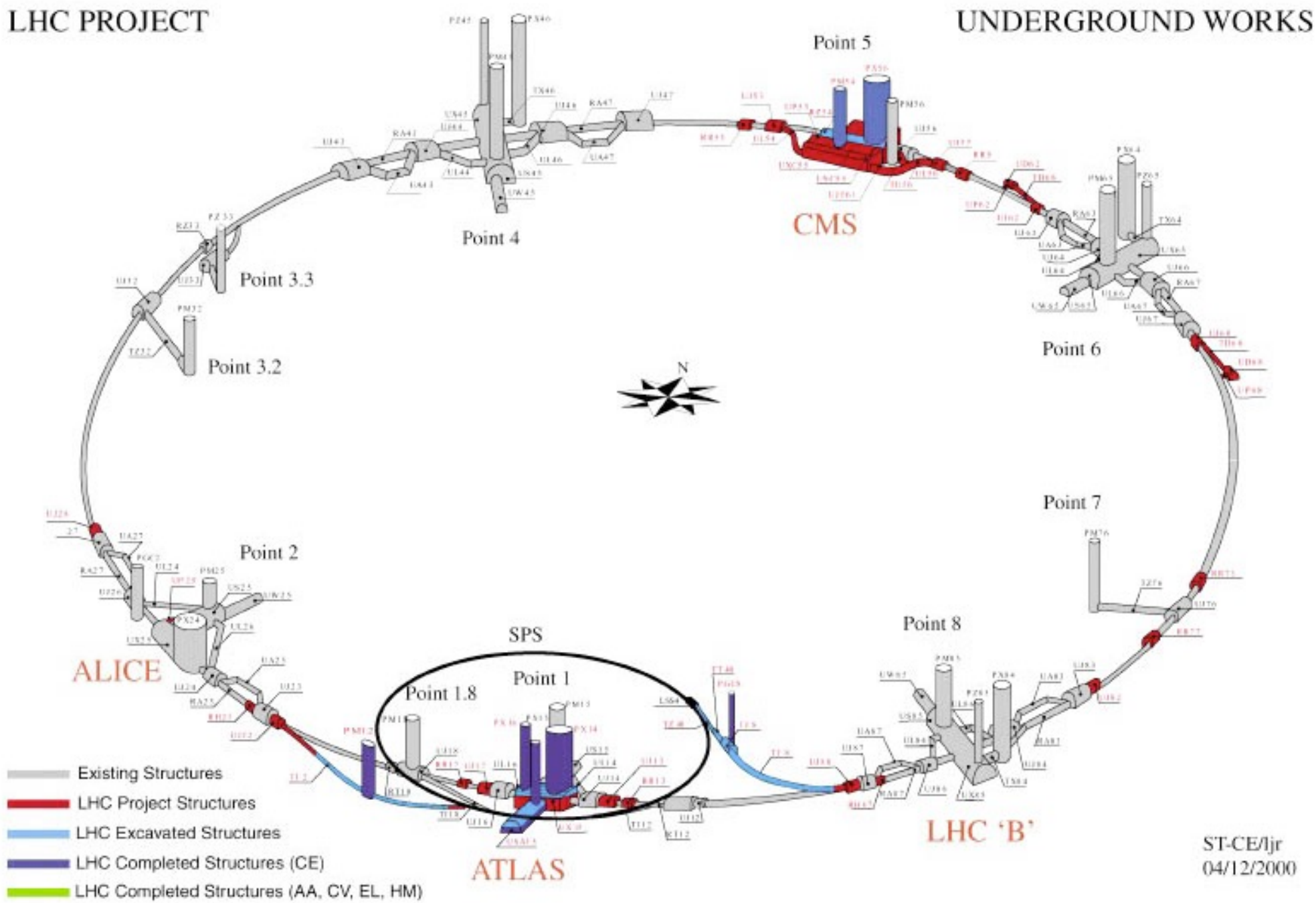
Booster

pbar debuncher and accumulator

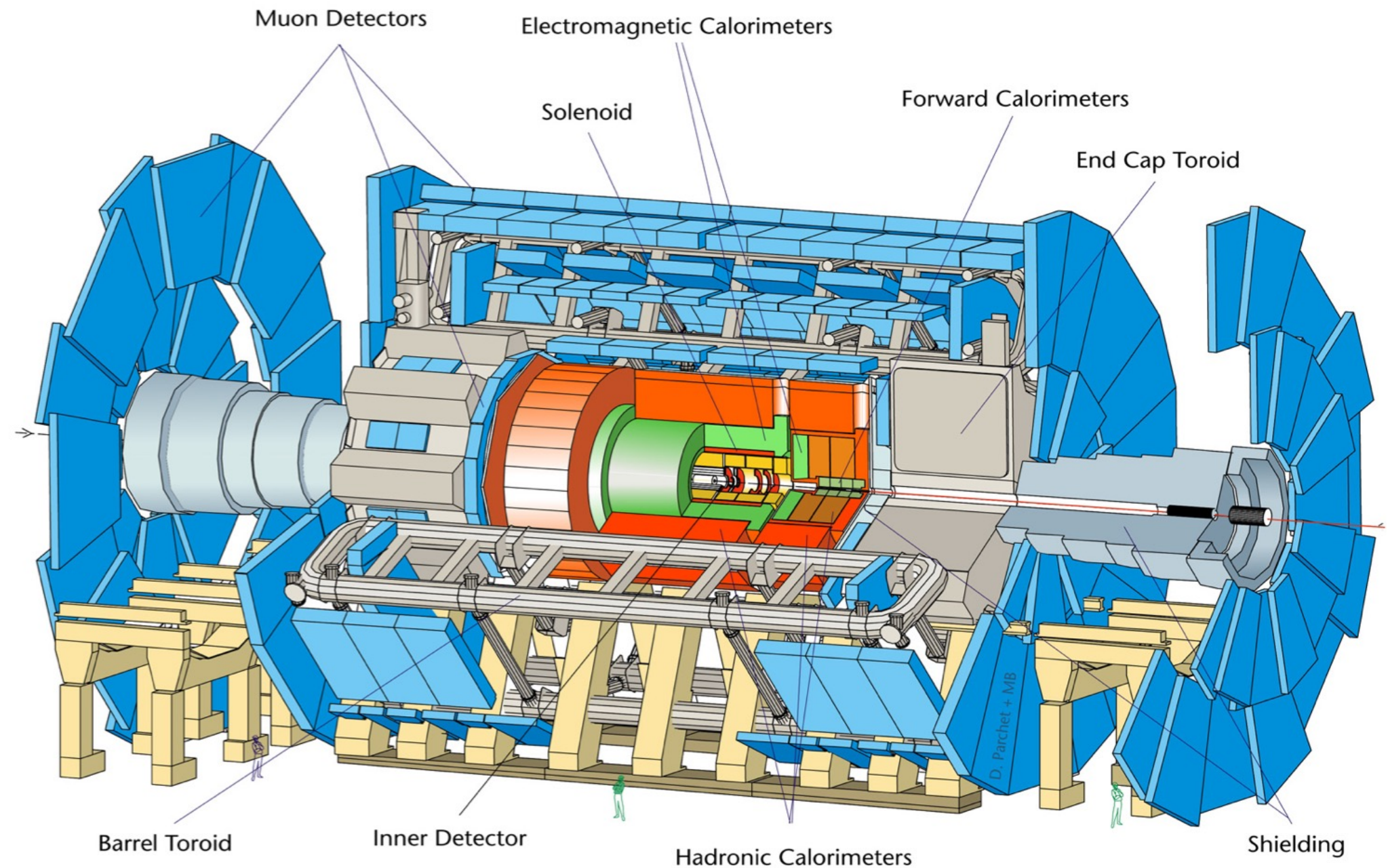
ppbar collisions at 1.96 TeV
Luminosity up to 400E30 $cm^{-2}s^{-1}$
A very good week: ~80 pb^{-1}

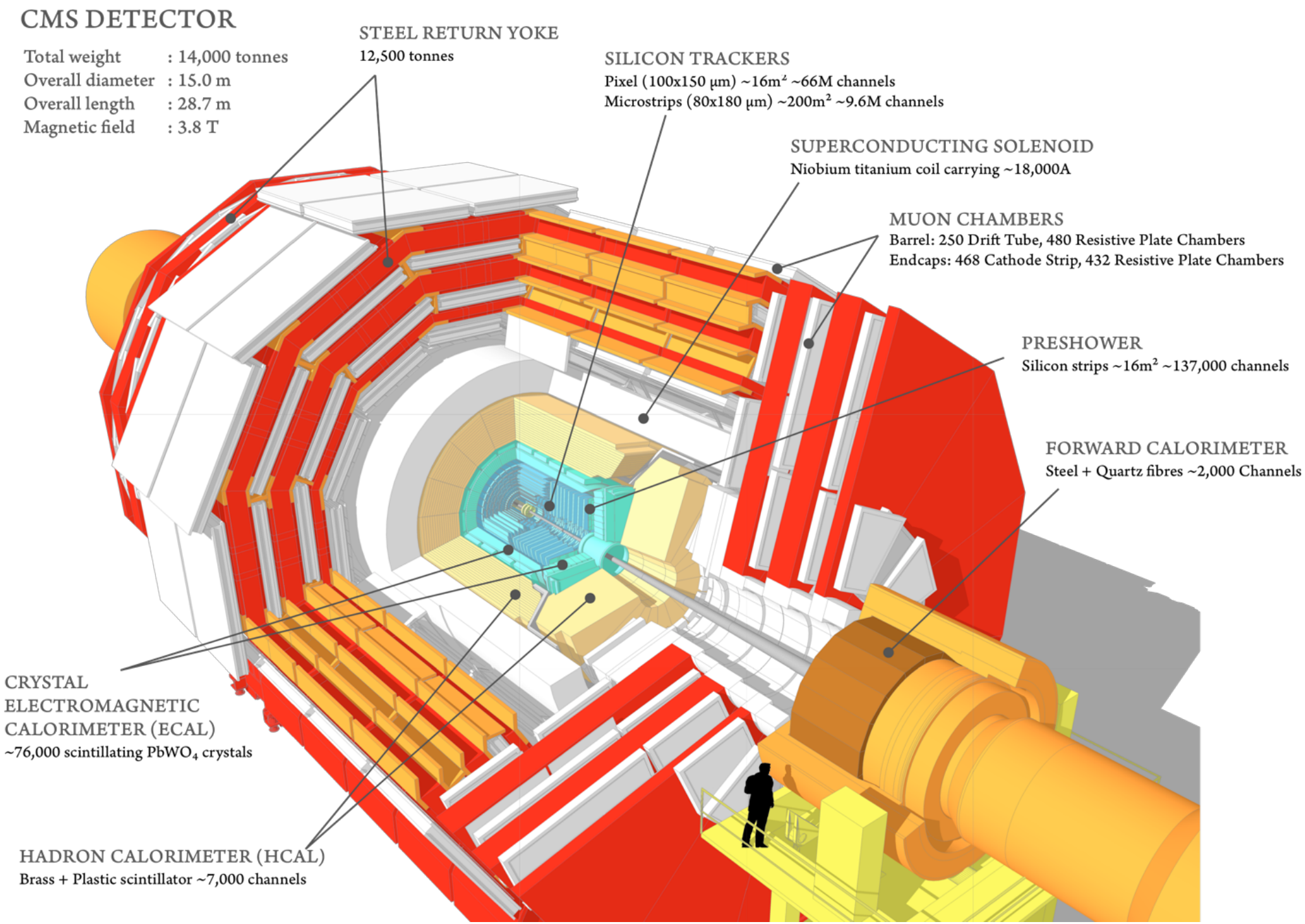
The CDF and D0 Detectors




LHC PROJECT
UNDERGROUND WORKS
Point 5
CMS
Point 4
Point 3.3
Point 3.2
Point 6
N
Point 7
Point 2
ALICE
SPS
Point 1
Point 1.8
Point 8
LHC 'B'
ATLAS
Existing Structures
LHC Project Structures
LHC Excavated Structures
LHC Completed Structures (CE)
LHC Completed Structures (AA, CV, EL, HM)
ST-CE/ljr
04/12/2000

The ATLAS Detector at the LHC




CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO4 crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

CDF Run II Trigger System

Bunch Crossing Rate: ~1.7 MHz

Level 1 trigger ~15 KHz

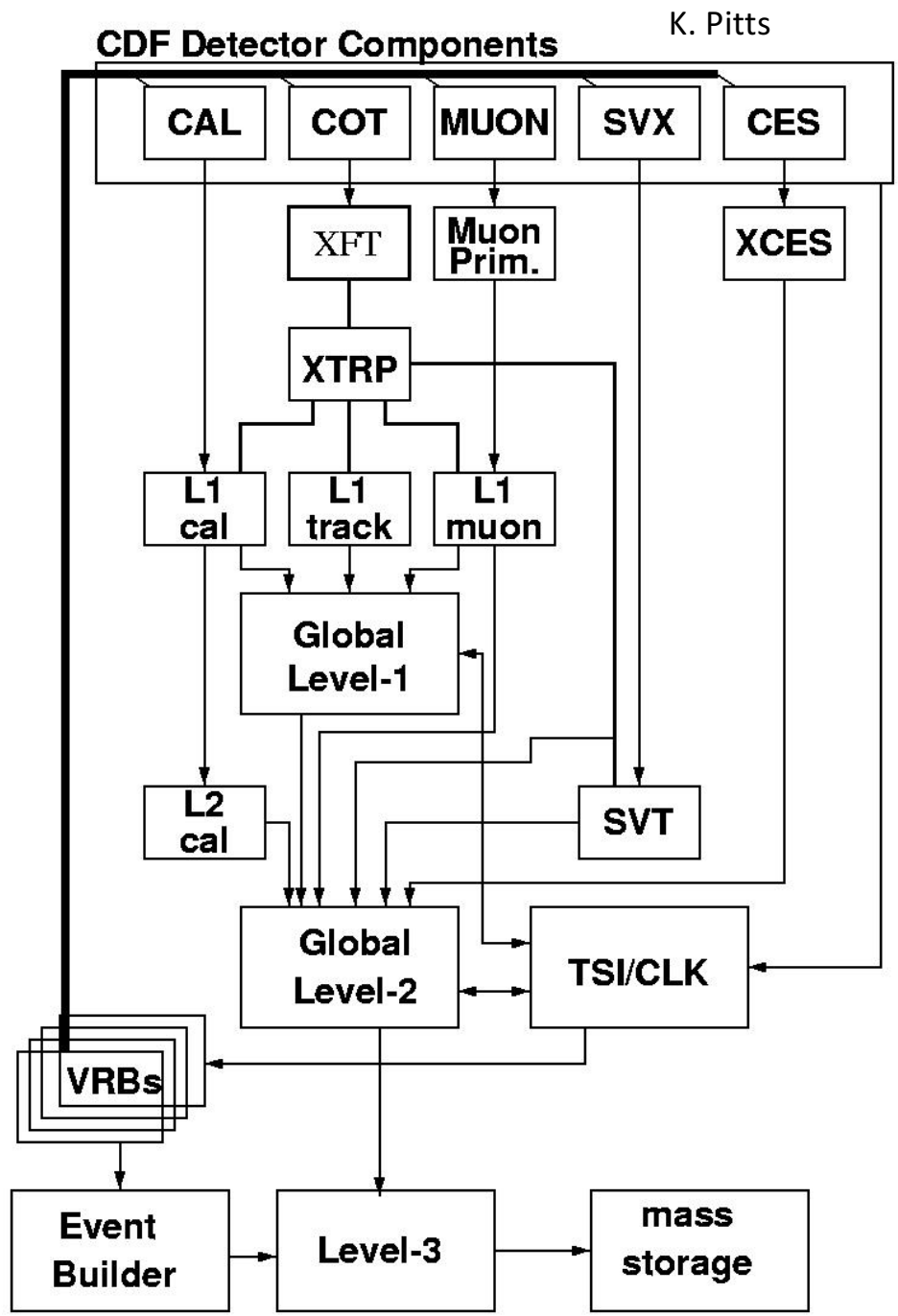
- tracking
- calorimeter: jets & electrons
- muons

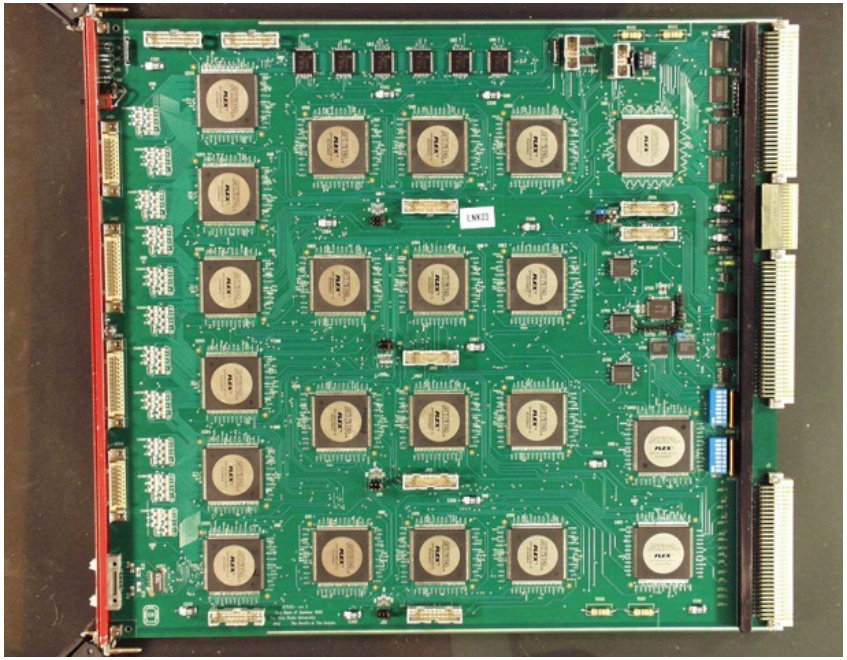
Level 2 trigger ~800 Hz

- L1 information (tracks, e, μ)
- calorimeter shower max
- silicon information
- algorithms run in L2 processor

Level 3 trigger ~200 Hz to tape

- full detector readout
- event building
- "offline" processing

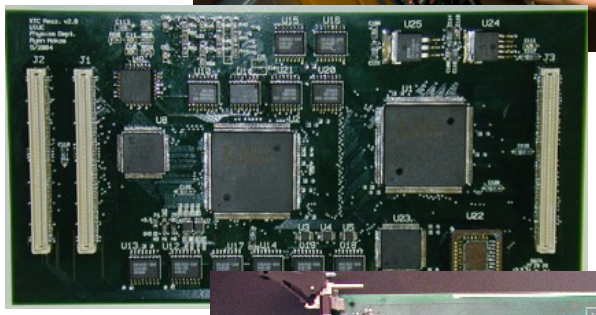




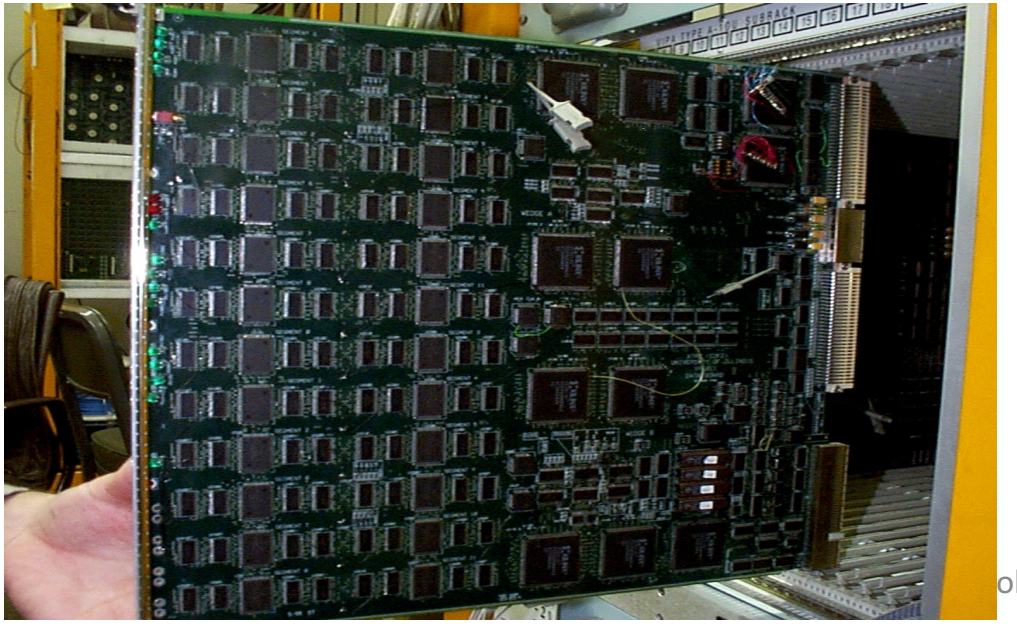
CDF
Trigger
Electronics

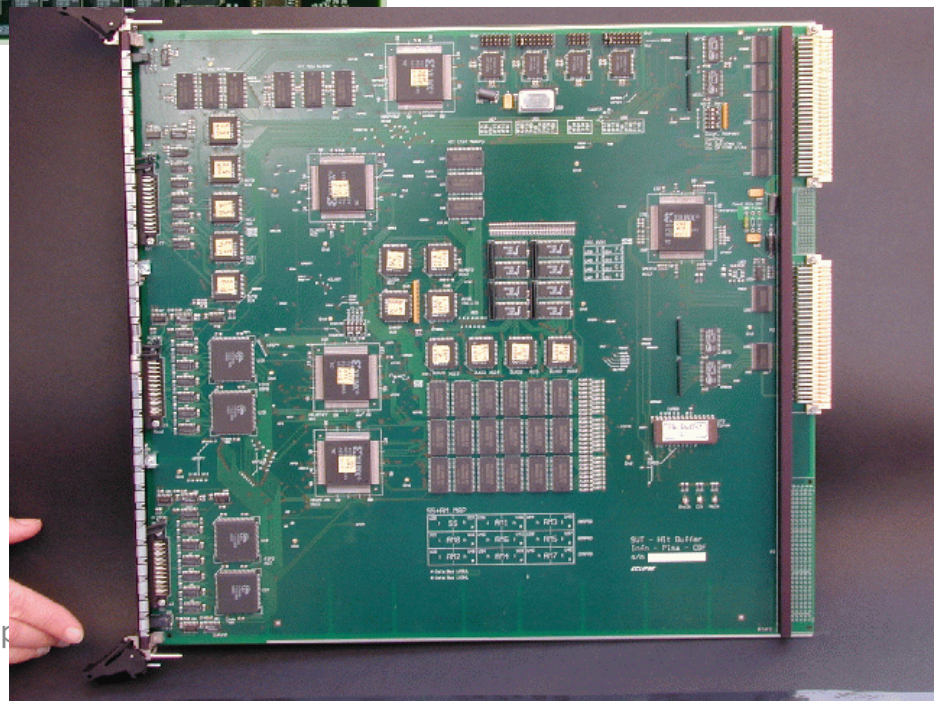
K. Pitts

Hundreds of application- specific boards working together, passing and processing data on the nanosecond timescale…

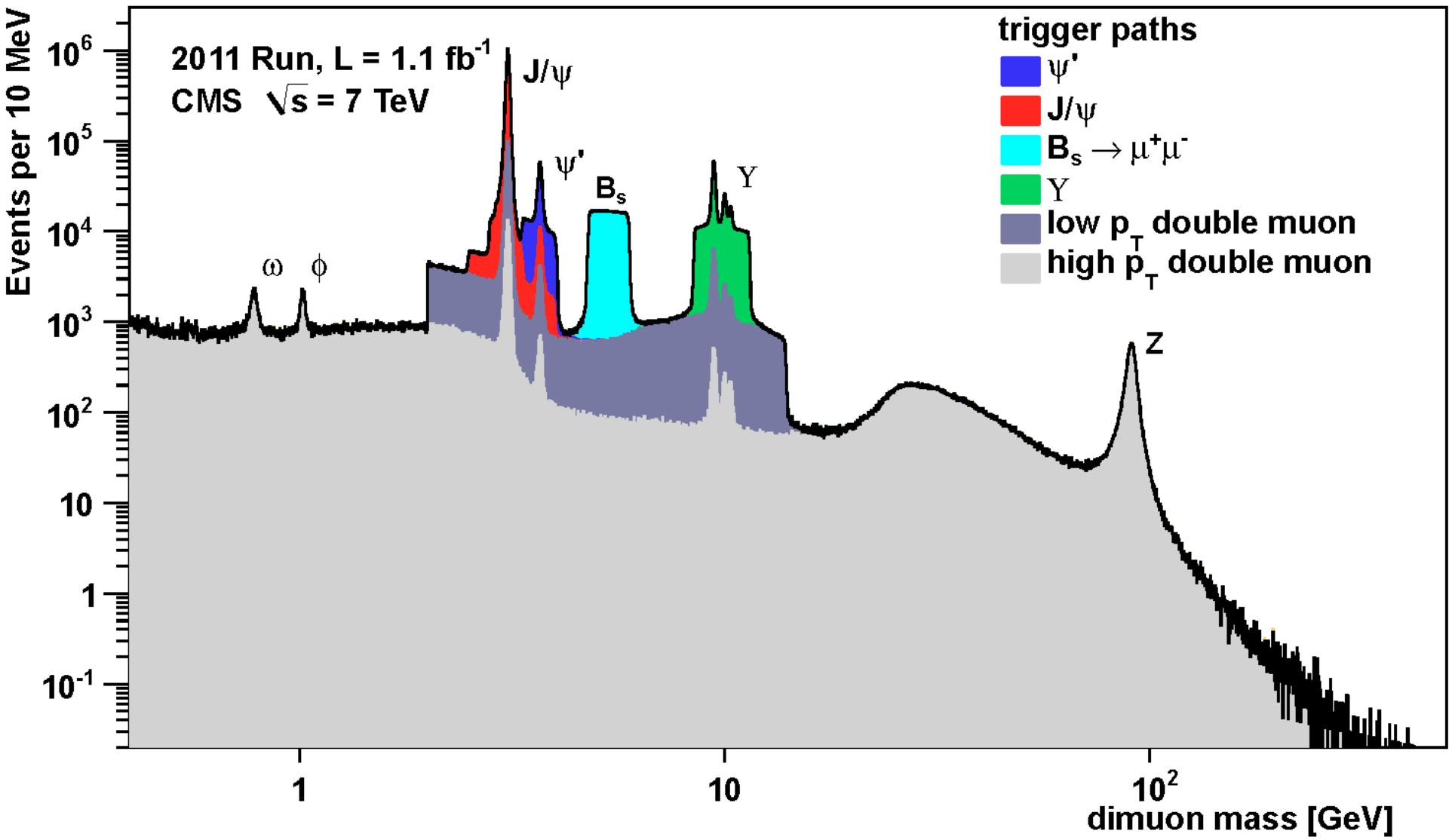


System continually upgraded to deal with increasing luminosity/pileup





Dimuon Mass Spectrum in a Small Initial Data Sample



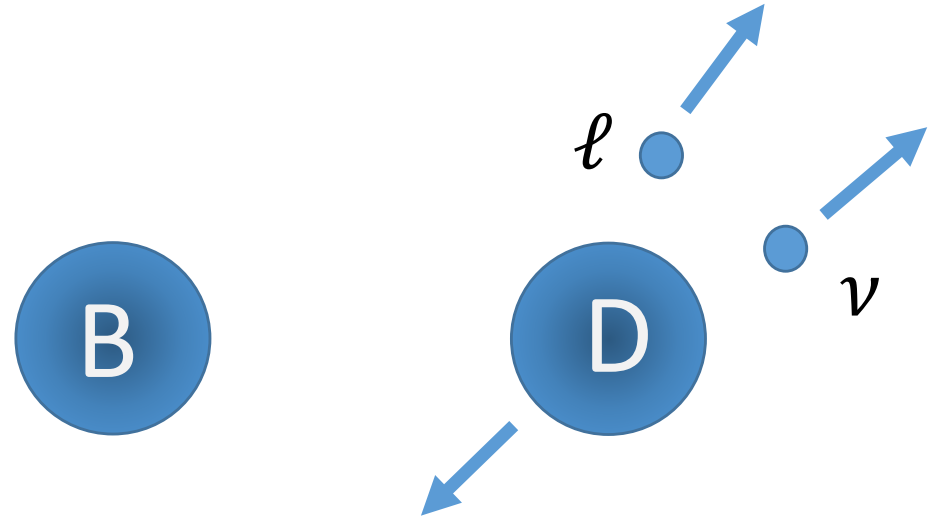
Triggers have different acceptances for different mass ranges
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsMUO#Full_invariant_mass_spectrum_of

Measuring $|V_{cb}|$ in Simple Events

$$B \rightarrow D \ell \nu$$

A *b* quark is very heavy, and it has a lighter quark and other stuff (gluons, sea quarks) bound to it.

A charm quark (in the D meson) is also fairly heavy but lighter than the *b* quark.



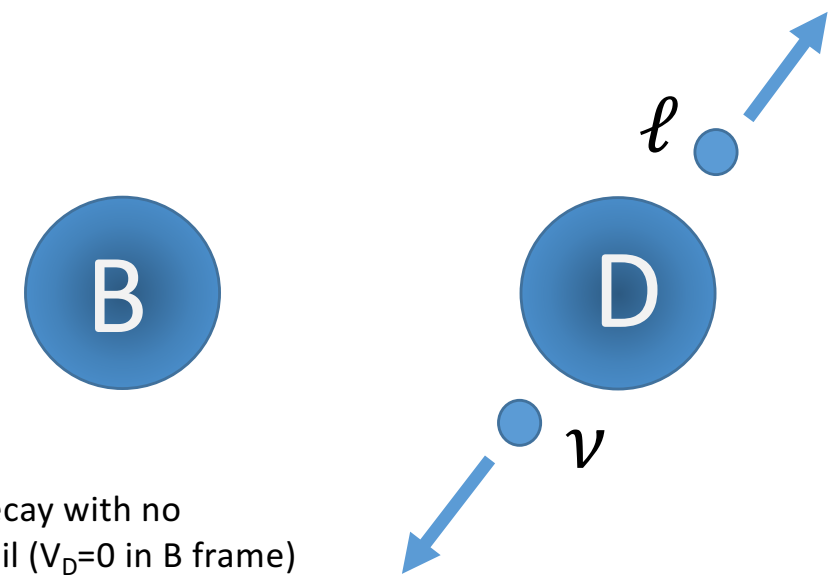
A typical decay with lots of recoil

In a recoil-less decay, the "other stuff" is minimally disturbed.

Easier to calculate the rate of this, but the closer to recoil-free you get, the fewer events there are.

$$q^2 = (P_\ell + P_\nu)^2$$

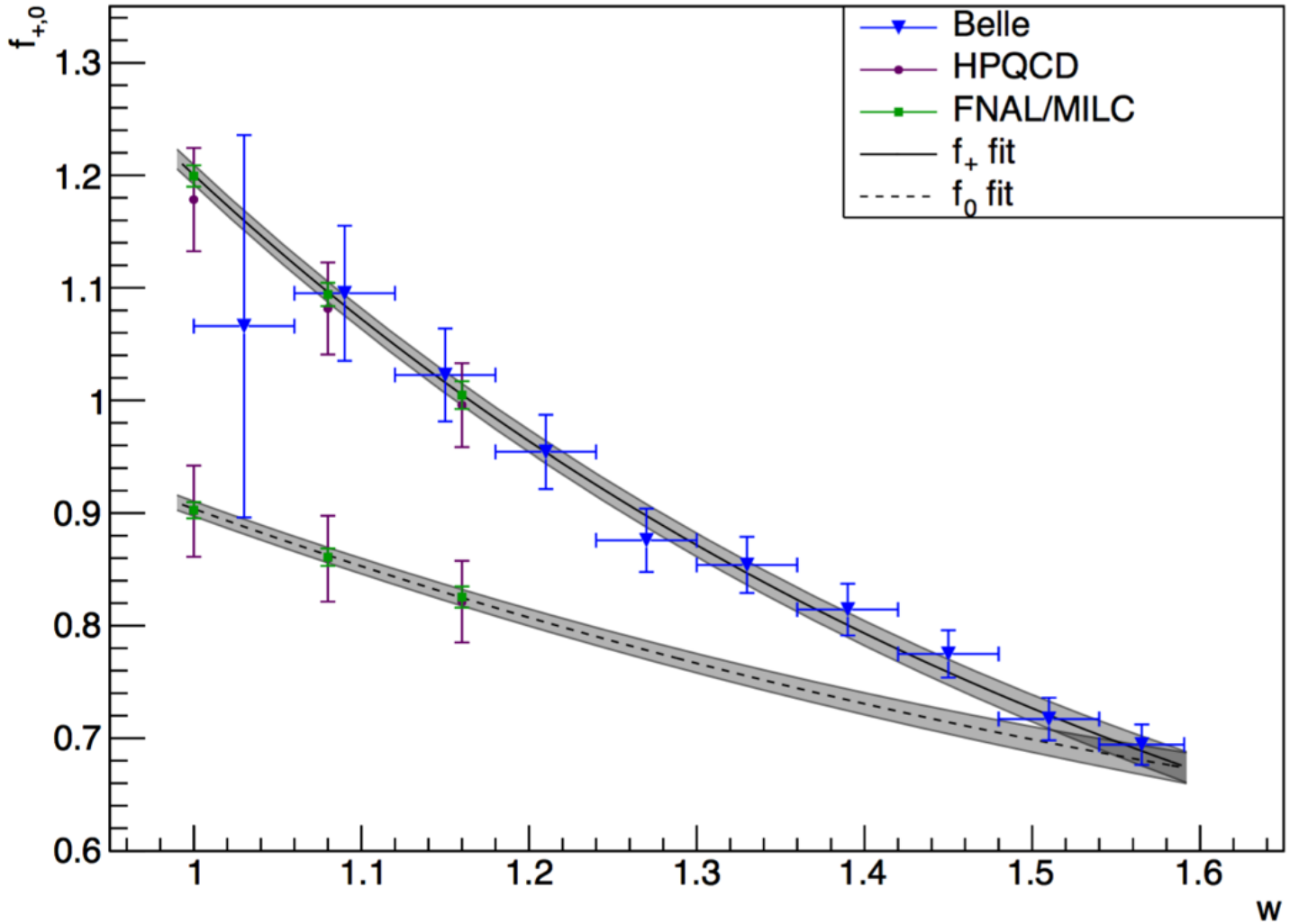
$$w = V_B \cdot V_D = \frac{m_B^2 + m_D^2 - q^2}{2 m_B m_D}$$



A decay with no recoil (V_D=0 in B frame)

Extrapolating to Zero Recoil

Belle Collab., Phys. Rev. D **93**, 032006 (2016)



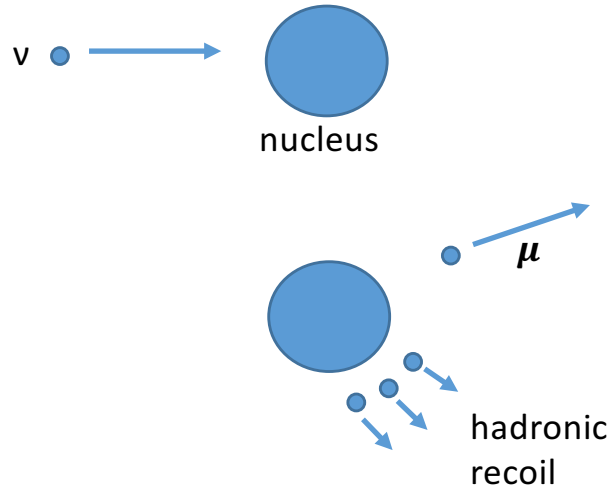
w=1 means no recoil

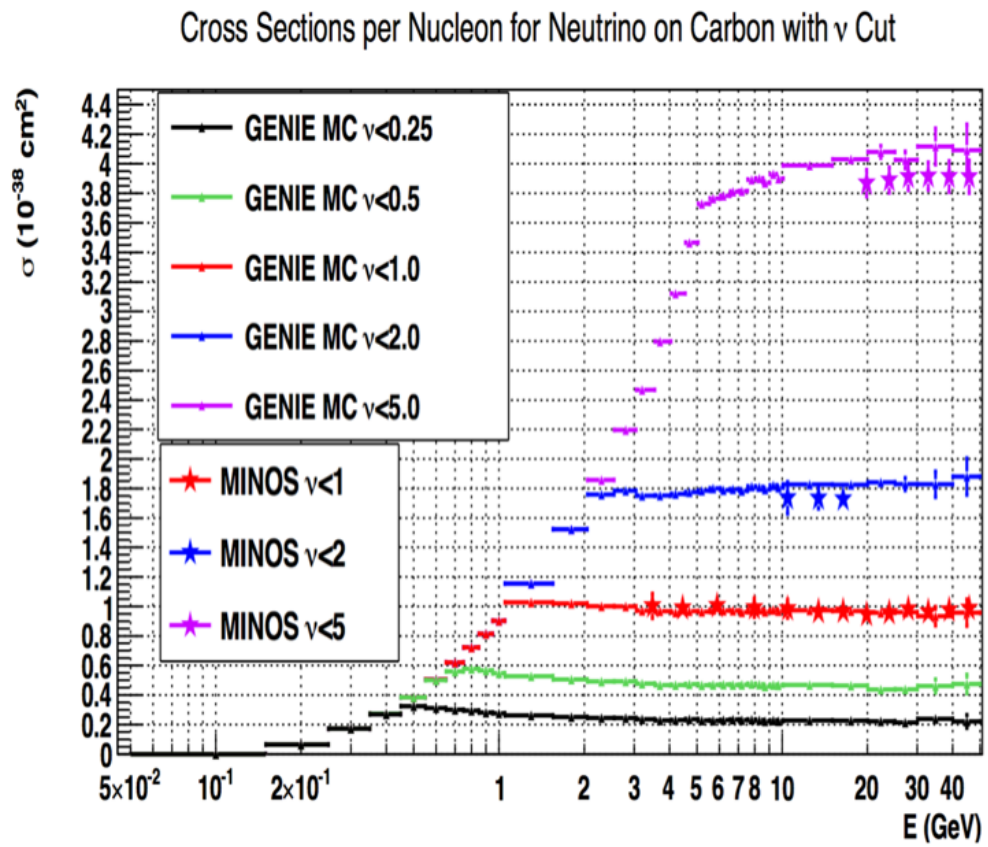
$$d\Gamma \propto G_F^2 |V_{cb}|^2 \left| L_\mu \langle D|\bar{c}\gamma^\mu b|B\rangle \right|^2$$

$$\langle D|\bar{c}\gamma^\mu b|B\rangle = f_+(q^2)\left[(P_B+P_D)^\mu - \frac{m_B^2-m_D^2}{q^2}q^\mu\right] + f_0(q^2)\frac{m_B^2-m_D^2}{q^2}q^\mu$$

FIG. 8. Form factors of the decay $B \to D\ell\nu_\ell$ and result of the combined fit to experimental and lattice QCD (FNAL/MILC and HPQCD) data. The BGL series (Eq. (8)) is truncated after the cubic term. The points with error bars are Belle and LQCD data. The solid curve is the f_+ form factor and the dashed curve represents f_0. The shaded areas around these curves indicate the uncertainty in the coefficients of the BGL expansion.

Similarity to the "low-ν Method"





Here, $\nu = E_{hadronic}$ in the lab frame

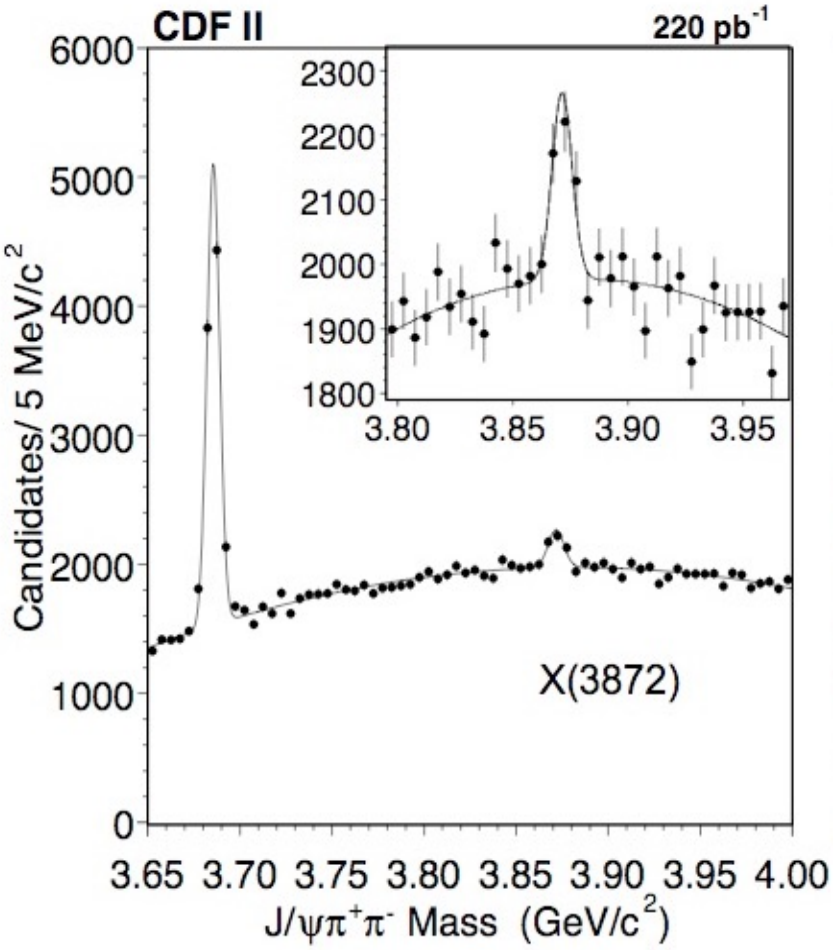
Little breakup of nucleus, not so complicated.

Problem is, fewer and fewer collision events the lower the cut on ν

CCFR/NUTeV: W. Seligman Ph.D. Thesis (1997),
S. Mishra, in Proceedings of the Workshop on Hadron Structure Functions and Parton Distributions, ed. D. Geesman *et al.*
A. Bodek, U Sarica, D. Naples, and L. Ren EPJC **72**, 1973 (2012).

“On-Off” Example

Select events with J/ψ(→ll) $\pi^+\pi^-$ candidates. Lots of nonresonant background which is poorly understood *a priori*, but there’s a *lot* of it.



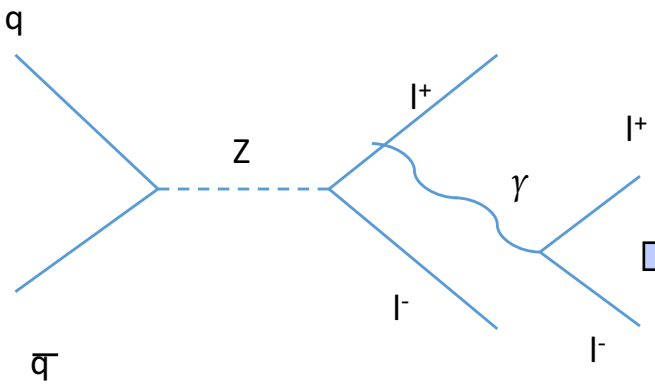
Typical strategy:
Fit the background outside of the signal peak, and interpolate the background under the signal to subtract it off.

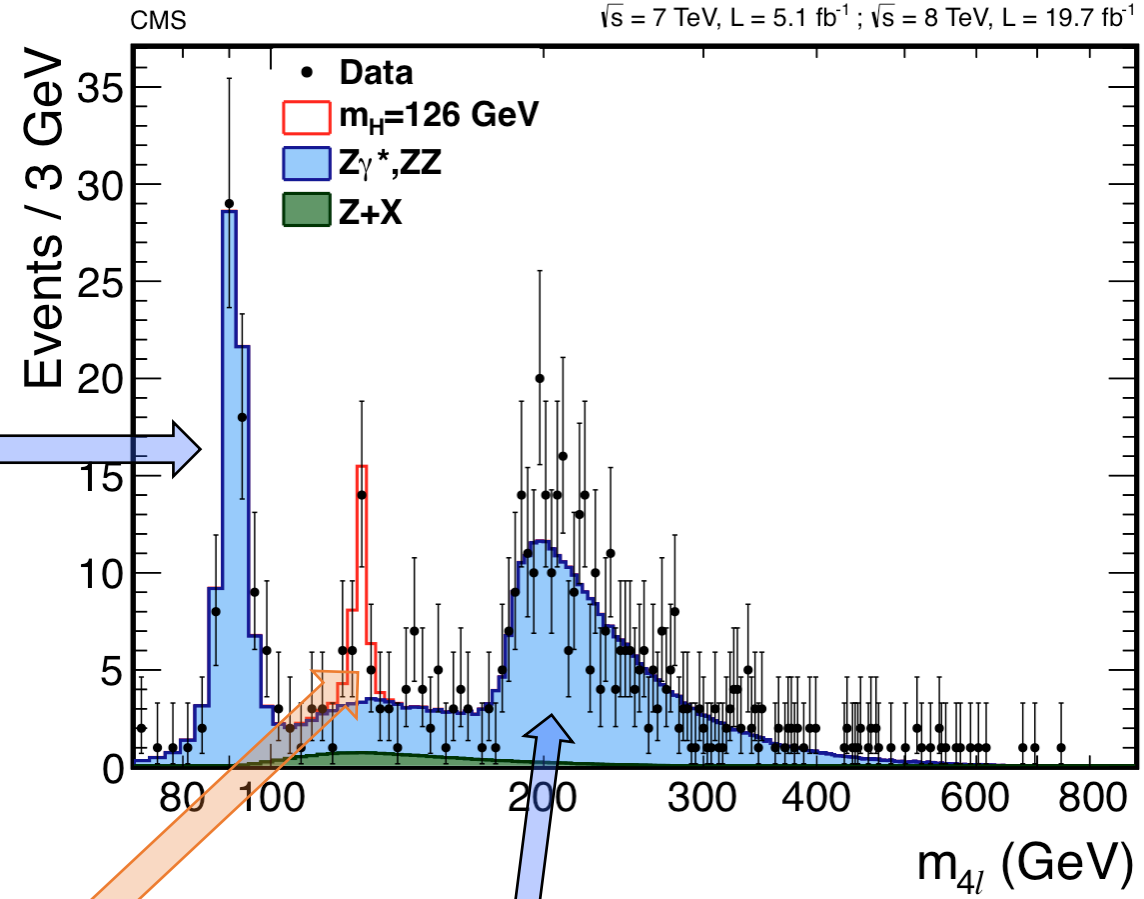
The ratio of events in the sidebands to the background prediction under the signal is called **τ**

Guess a shape that fits the backgrounds, and fit it with a signal.

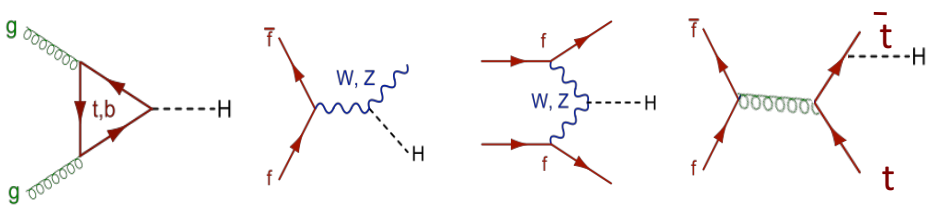
The M(4L) Distribution

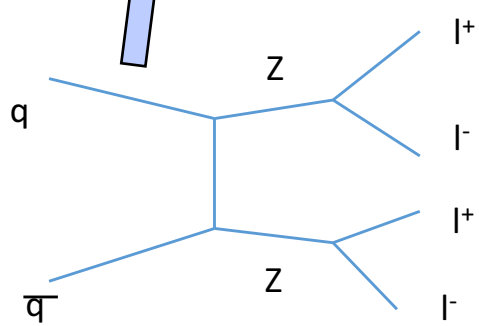
Summed plot: eeee, eeμμ, μμμμ



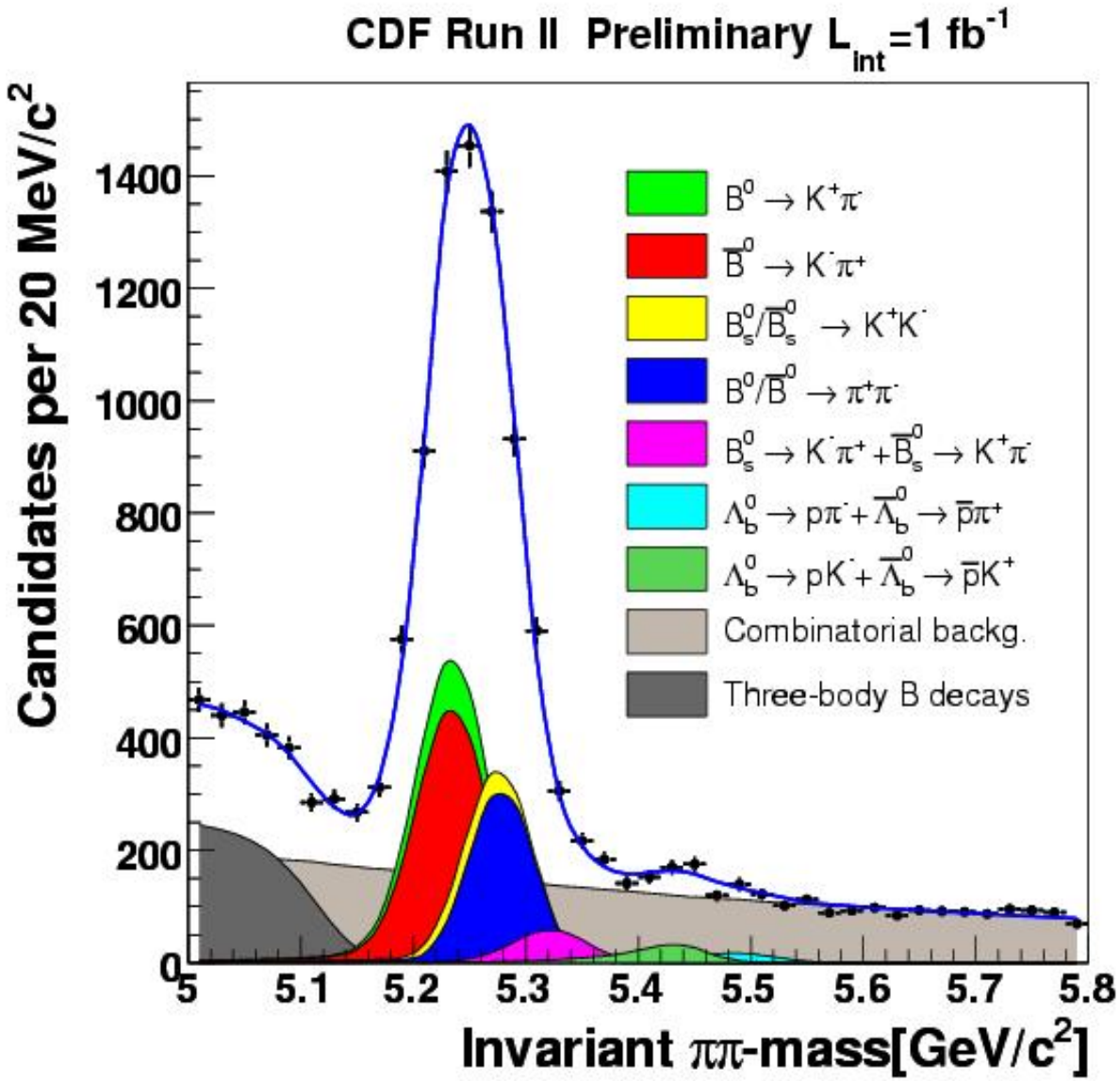


I'm not too sure what the error bars on the data points correspond to.





Several "on's", Several "off's"



Weak Sidebands

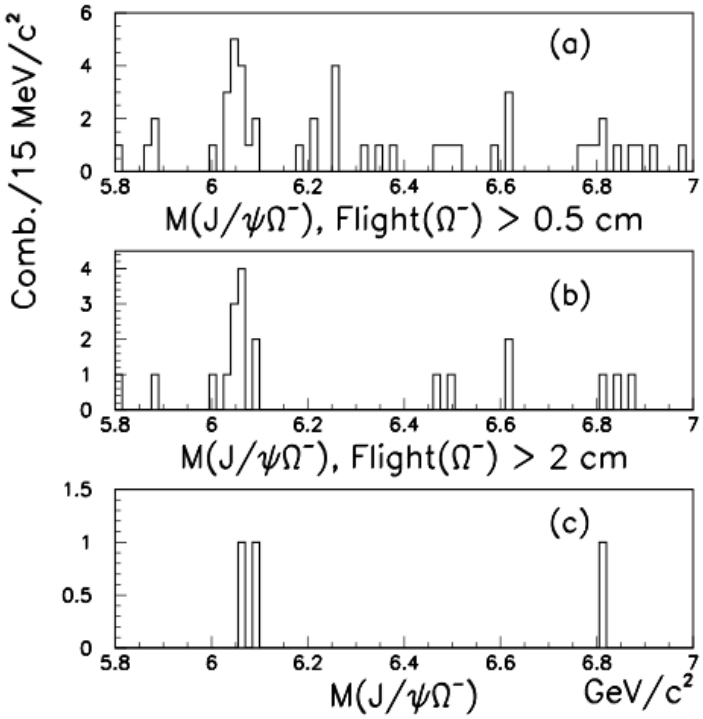
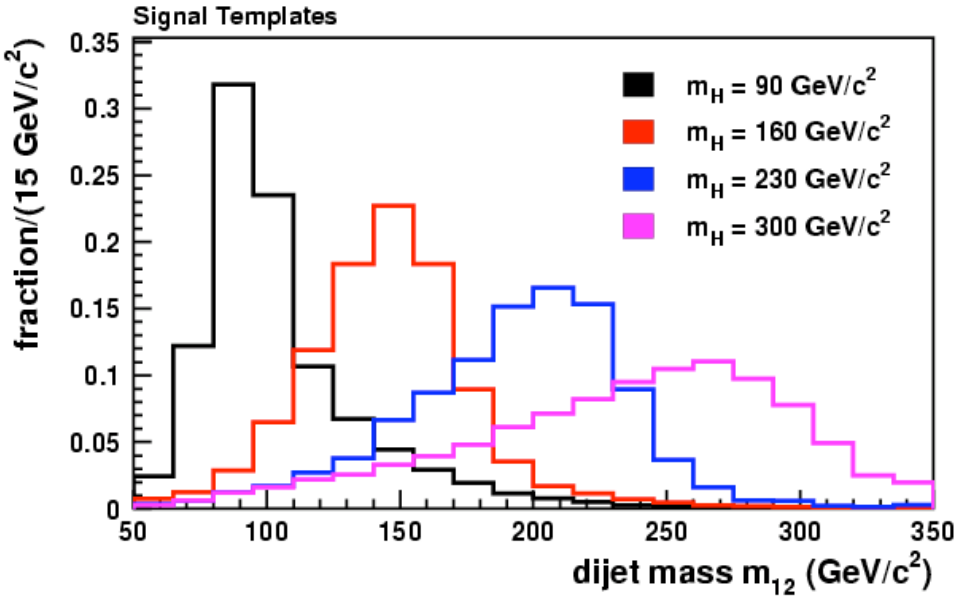


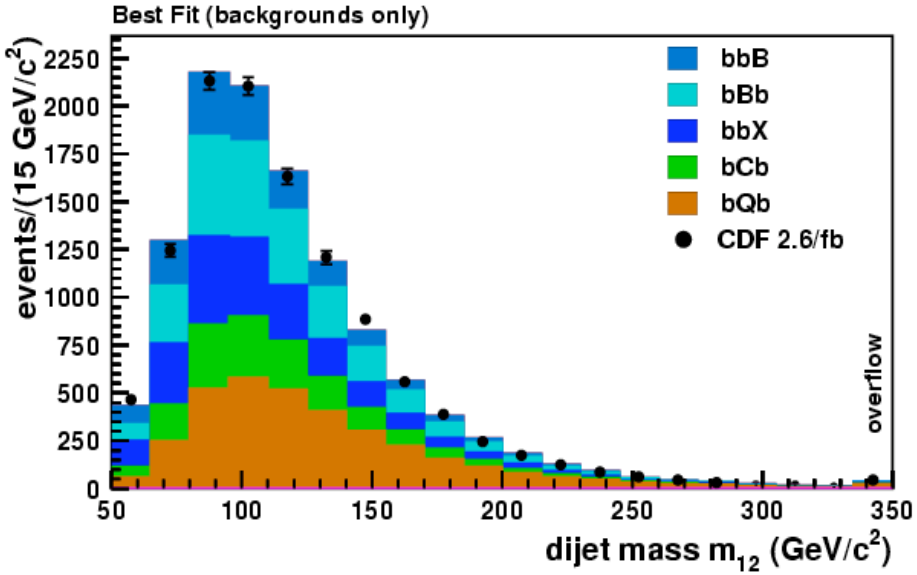
FIG. 8: (a,b) The invariant mass distribution of $J/\psi\,\Omega^-$ combinations for candidates where the transverse flight requirement of the Ω^- is greater than 0.5 cm and 2.0 cm. (c) The invariant mass distribution of $J/\psi\,\Omega^-$ combinations for candidates with at least one SVXII measurement on the Ω^- track. All other selection requirements are as in Fig. 5(c).

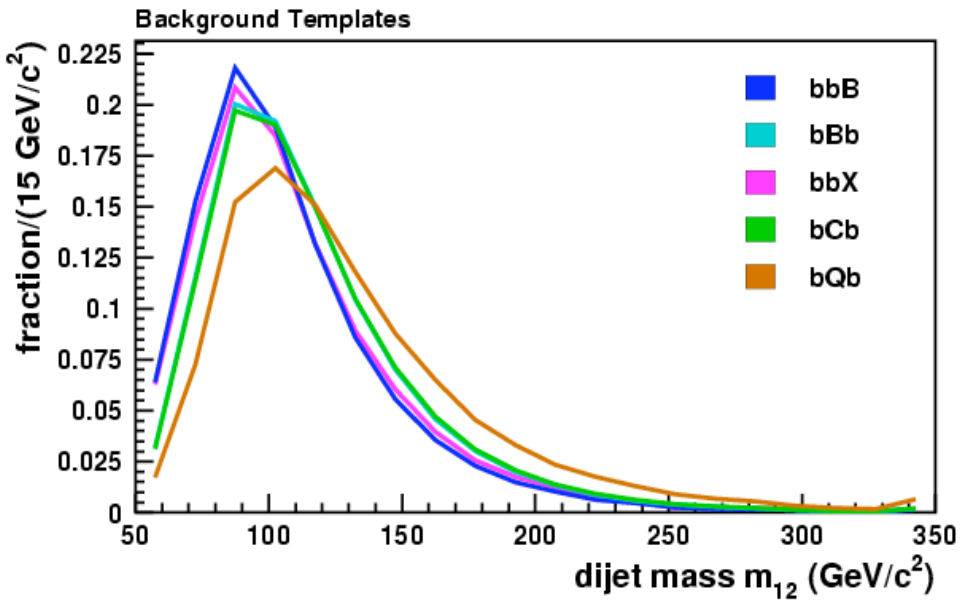
CDF's Ω_b observation paper:

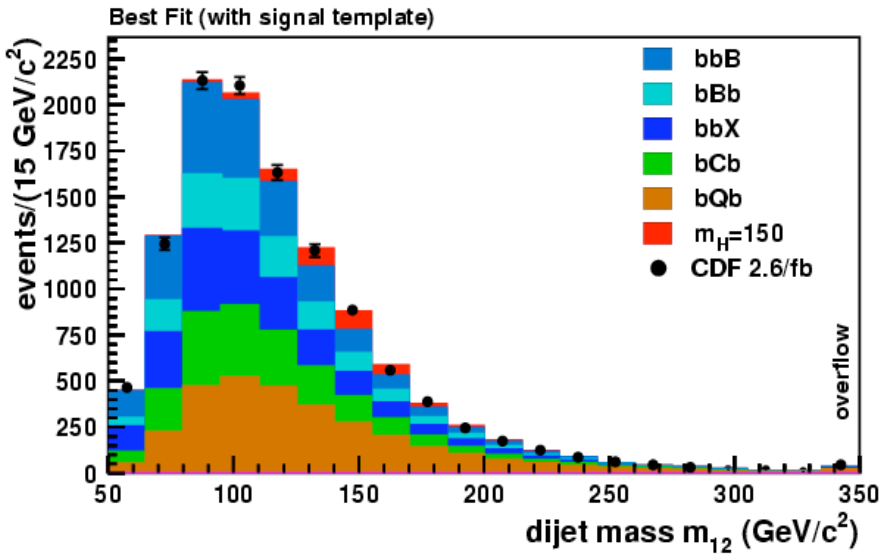
Phys.Rev. D80 (2009) 072003

Another Weak Sideband Constraint Example that Looks Like a Strong Sideband Constraint





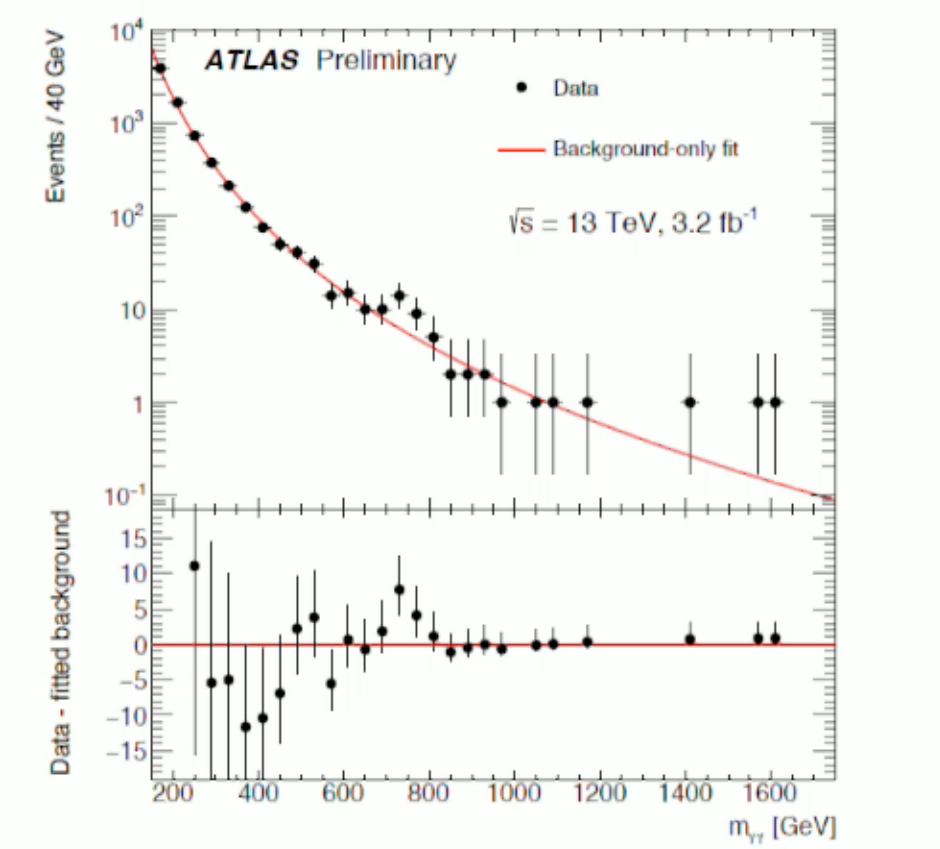


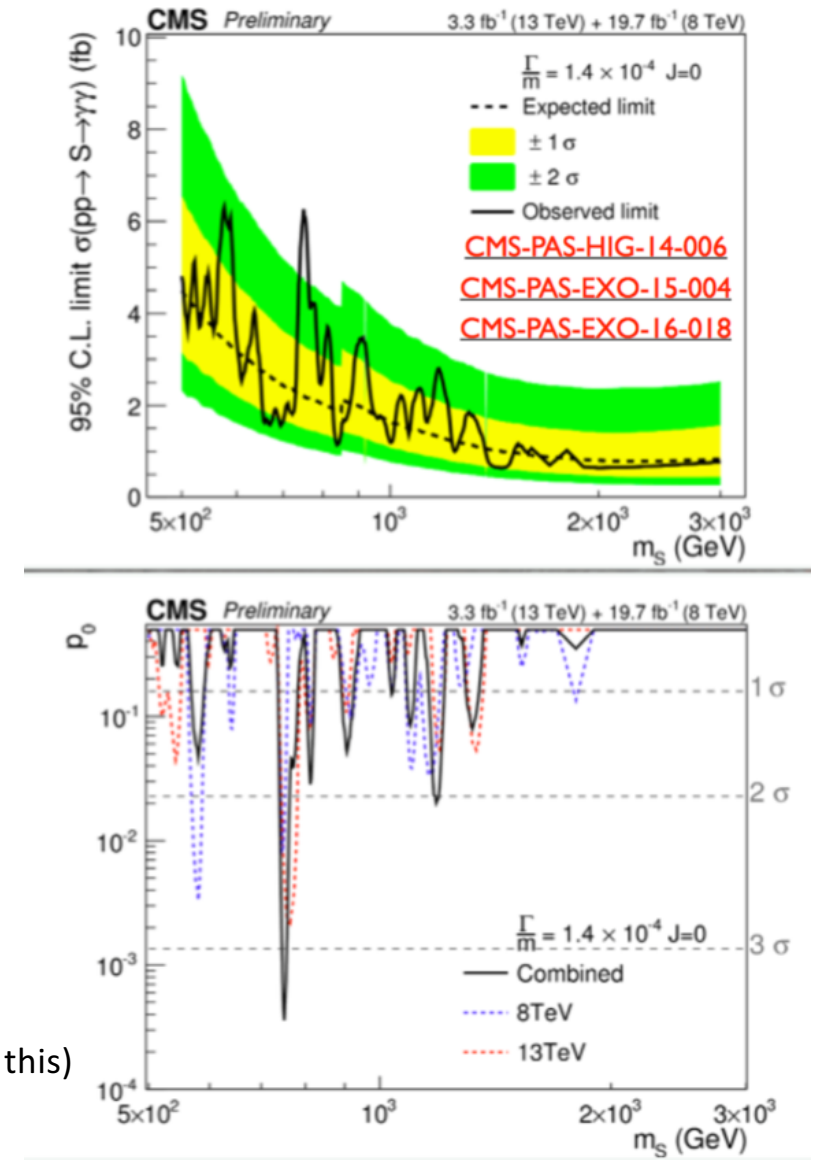


Phys.Rev. D85 (2012) 032005
e-Print: arXiv:1106.4782 [hep-ex]

Diphoton Excess at 750 GeV

The distribution that caught everyone's attention in 2015:





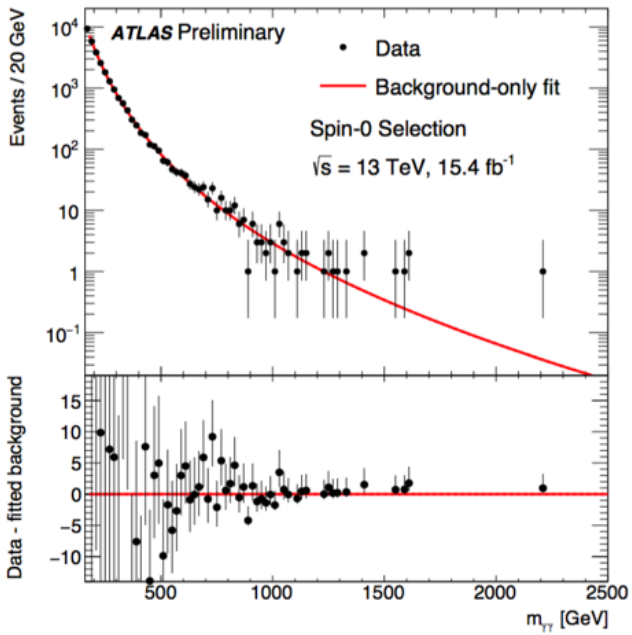
This is exactly what a new particle on the edge of detectability would look like! (A Banff Challenge 2 problem looked a lot like this)

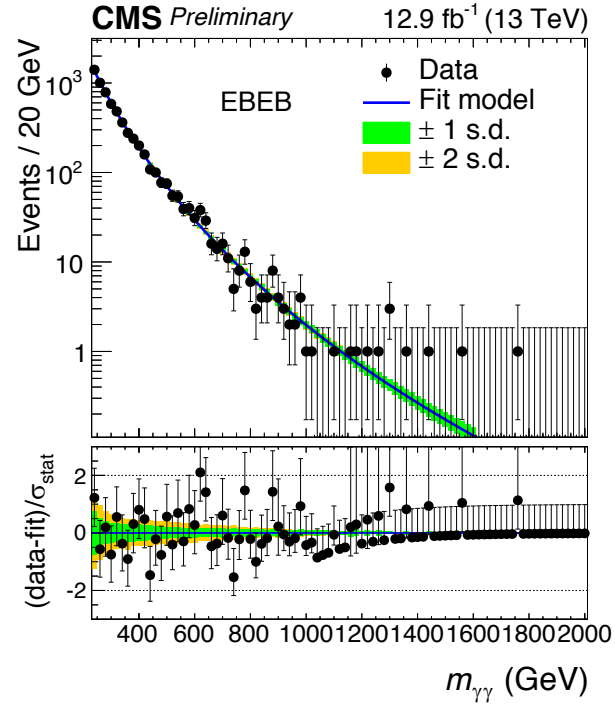
ATLAS-CONF-2015-089

Local significance ~3.5σ, global: ~2σ

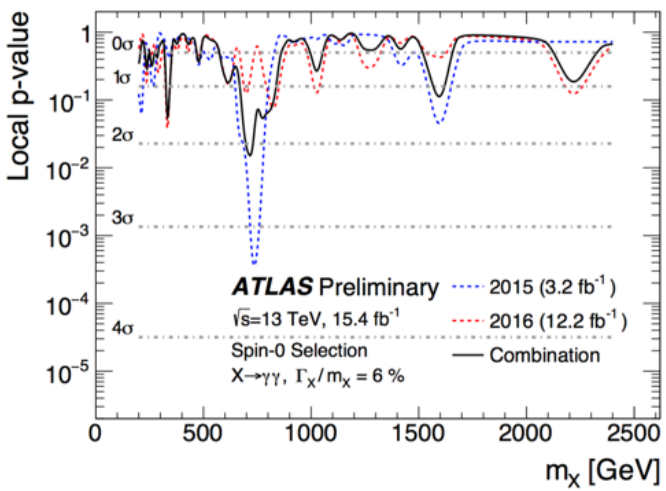
CMS saw an excess too!

But it did not get confirmed with new data

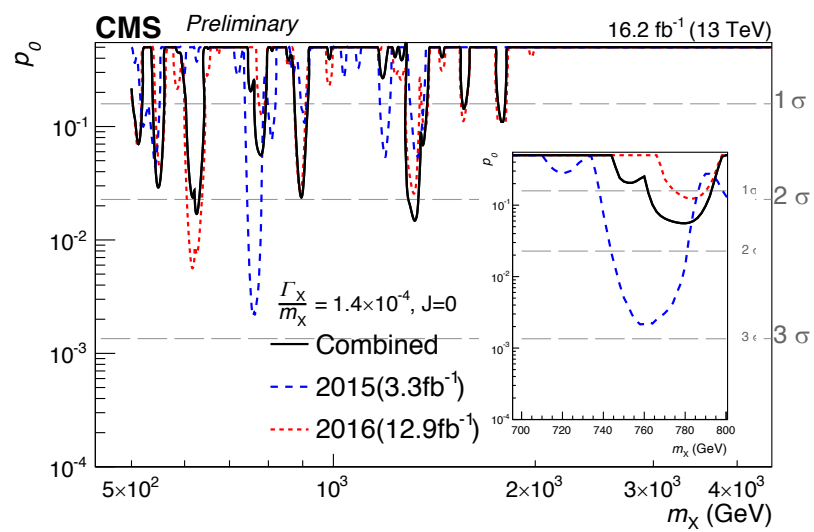




Results presented at ICHEP 2016



ATLAS-CONF-2016-059



CMS-PAS-EXO-16-027

An Approximate LEE Correction for Peak Hunting

E. Gross and O. Vitells, **Eur.Phys.J. C70 (2010) 525-530**.

Approximate formula applies to bump hunts on a smooth background.

Requires a few fully simulated pseudoexperiments with complete p-value calculations over the region of interest. Count up-crossings of a threshold. Extrapolates to higher thresholds assuming large-sample behavior. Specifically, that the LR test statistic has a chisquared distribution.

An interesting feature – specific to bump hunts but may be more general:

As the expected significance goes up, so does the LEE correction

This makes lots of sense: LEE depends on the number of separate models that can be tested. As we collect more data, we can measure the position of the peak more precisely.

So we can tell more peaks apart from each other, even with the same reconstruction resolution.

But: Combine a poor resolution low s/b search with a high resolution high s/b but very tiny s and very tiny b search – may not get the right answer.

An earlier bump in $m_{\gamma\gamma}$

S. C. C. Ting, DPF '92
Conf.Proc. C921110 (1992) 53-108

$e^+e^- \rightarrow \ell\ell\gamma\gamma$ at LEP 1

Possible to use Gross and Vitells' LEE? Some parts of the histogram have very few expected events and are thus highly non-Gaussian. Other bins have more expected events.

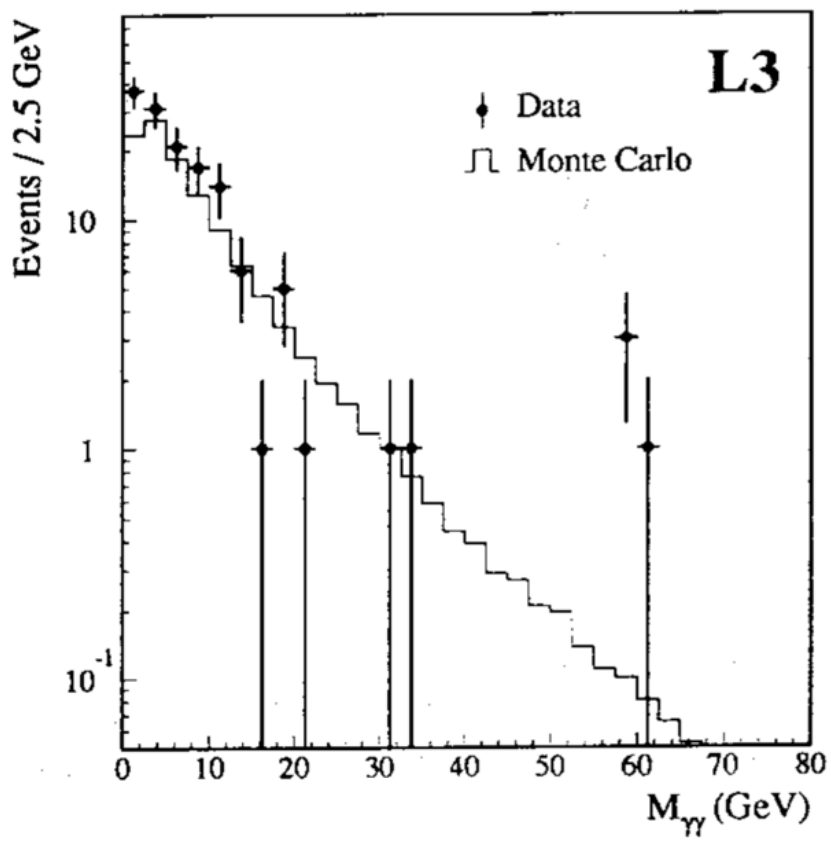


Figure 54: The measured two photon invariant mass spectrum from L3 compared with the QED Monte Carlo expectations.

DELPHI also observed two such events with $M_{\gamma\gamma}$ close to 60 GeV. The event pictures are shown in Figure 55. The invariant mass of the photon pair is $M_{\gamma\gamma} = 58.4\pm$

Where is "Elsewhere?"

A collider collaboration is typically very large; >1000 Ph.D. students. ATLAS+CMS is another factor of two. (Four LEP collaborations, Two Tevatron collaborations).

Many ongoing analyses for new physics. The chance of seeing a bump somewhere is large. What is the LEE?

Do we have to correct our previously published p-values for a larger LEE when we add new analyses to our portfolio?

How about the physicist who goes to the library and hand-picks all the largest excesses? What is LEE then?

"Consensus" at the Banff 2010 Statistics Workshop: LEE should correct only for those models that are tested within a single published analysis. Usually one paper covers one analysis, but review papers summarizing many analyses do not have to put in additional correction factors.

For the Winter 2012 Higgs search analyses, we had several LEE's computed, depending on the mass range defined to be elsewhere.

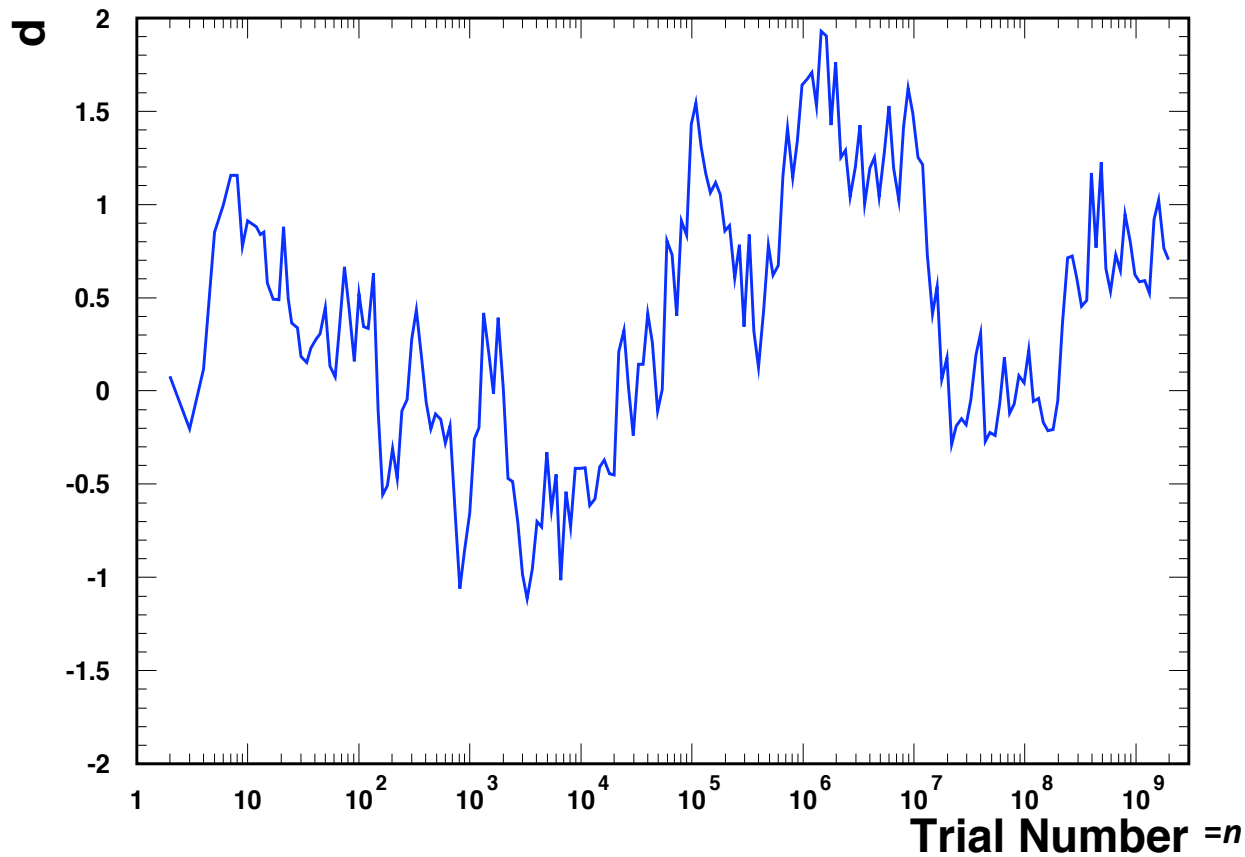
Caveat lector.

Look ElseWHEN

Running averages converge on correct answer, but the deviations in units of the expected uncertainty have a random walk in the logarithm of the number of trials

$$d_n = \frac{\sum_{k=1}^{n} r_k / n}{1/\sqrt{n}}$$

The r_k are IID numbers drawn from a unit Gaussian.

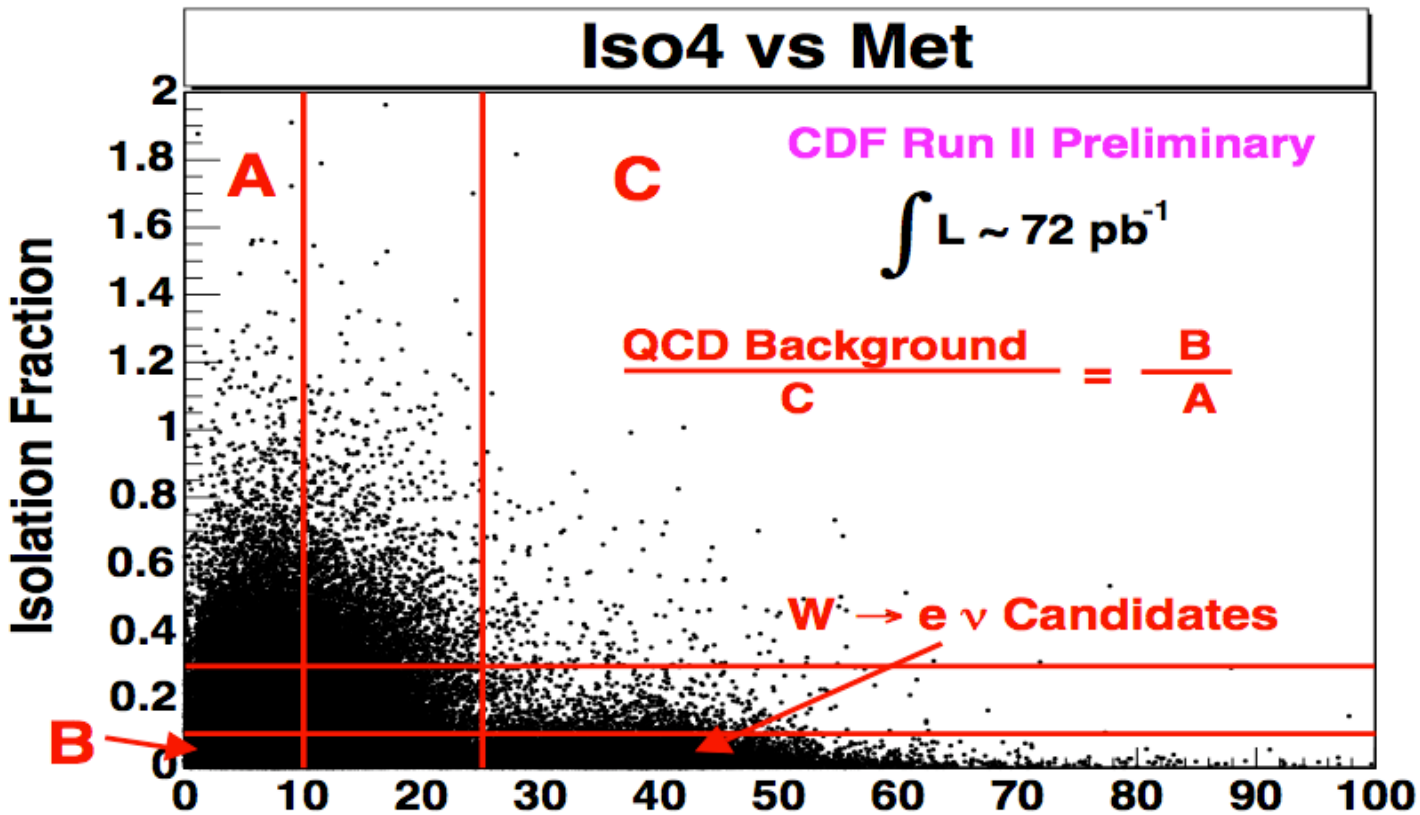


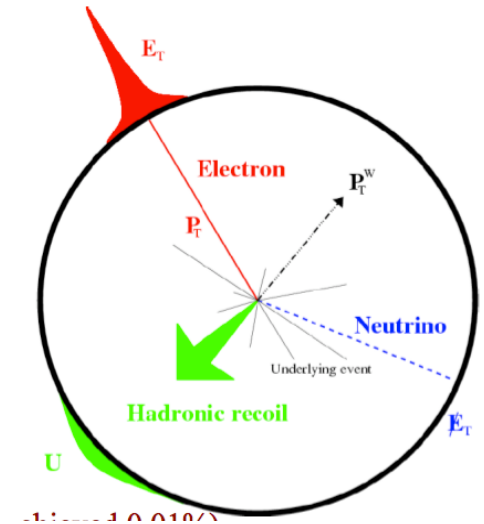
It's possible to cherry-pick a dataset with a maximum deviation. **"Sampling to a foregone conclusion"**

Stopping Rule: In HEP, we (almost always!) take data until our money is gone. We produce results for the major conferences along the way. Some will coincidentally stop when the fluctuations are biggest. We take the most recent/largest data sample result and ignore
(or should!) results performed on smaller data sets. p-values still distributed uniformly from 0 to 1. A recipe for generating "effects that go away"

"ABCD" Methods

CDF's W Cross Section Measurement





Isolation fraction=

Energy in a cone of radius 0.4 around lepton candidate not including the lepton candidate / Energy of lepton candidate

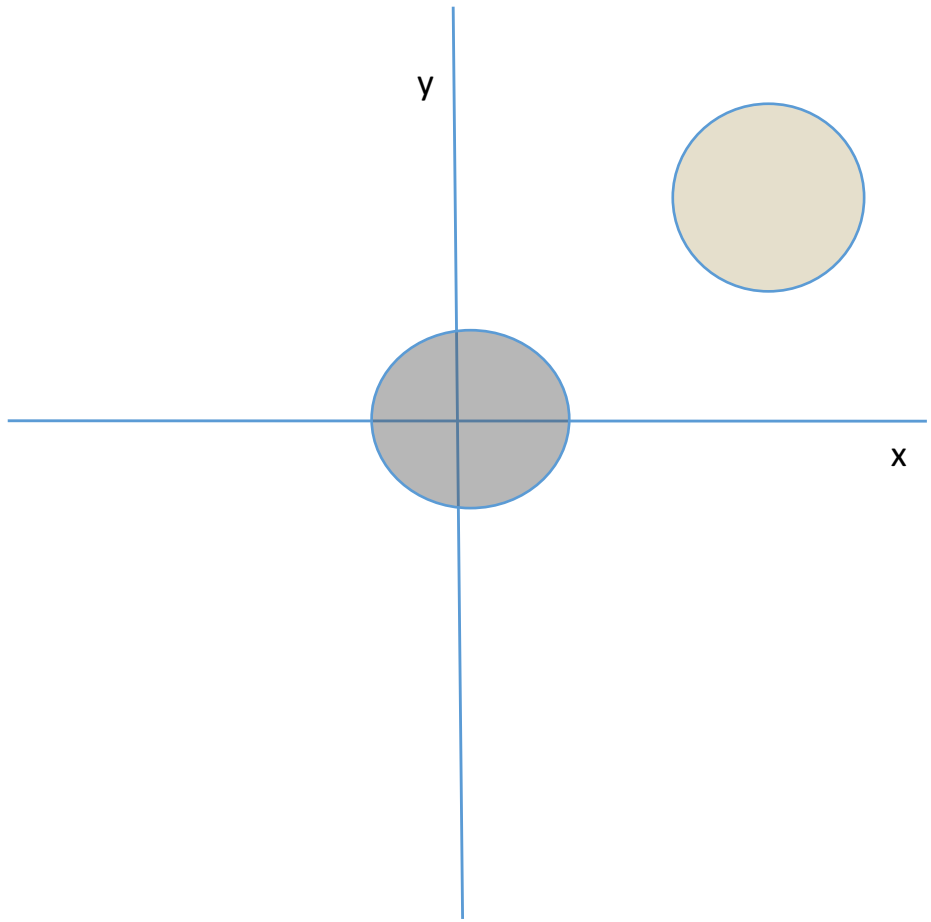
Want QCD contribution to the "D" region where signal is selected.

Assumes: MET and ISO are uncorrelated sample by sample

Signal contribution to A,B, and C are small and subtractable

ABCD methods are really just on-off methods where **τ is measured using data samples**

The Sum of Uncorrelated 2D Distributions may be Correlated



Knowledge of one variable helps identify which sample the event came from and thus helps predict the other variable's value even if the individual samples have no covariance.

Examples of ABCD Methods

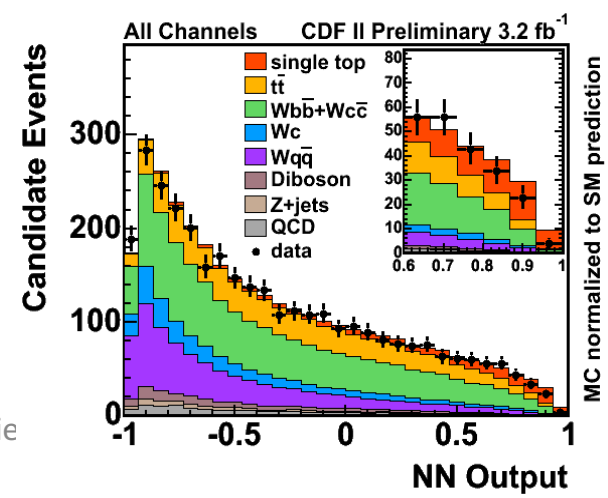
- Sideband calibration of background under a peak. ("what if the background peaks also where the signal peaks?")
- **The on-off problem with τ=A/C**. Very frequently samples A and C are in MC simulations, where we can be sure not to contaminate the background estimations wtih signal. Example: Using the MC to estimate acceptance for a cut for background, to be scaled with a data control sample.

But we pay the price of unknown MC mismodeling.

Uncorrelated variable assumption == assumption that τ is the same in the data and the MC. (check modeling of shape of distribution in the MC)

Equivalent of previous problem: Even if the background shapes are well modeled by the MC, if there are multiple background processes which contribute, they can have different fractional contributions, distorting the total shapes.

- Fitting an MVA shape to the data. Low-score MC = A, High-Score MC = C
  Low-score data = B, High-score Data=D.



An Example from ICHEP 2016, CMS H→ττ Analysis

Higgs at 13 TeV

New!

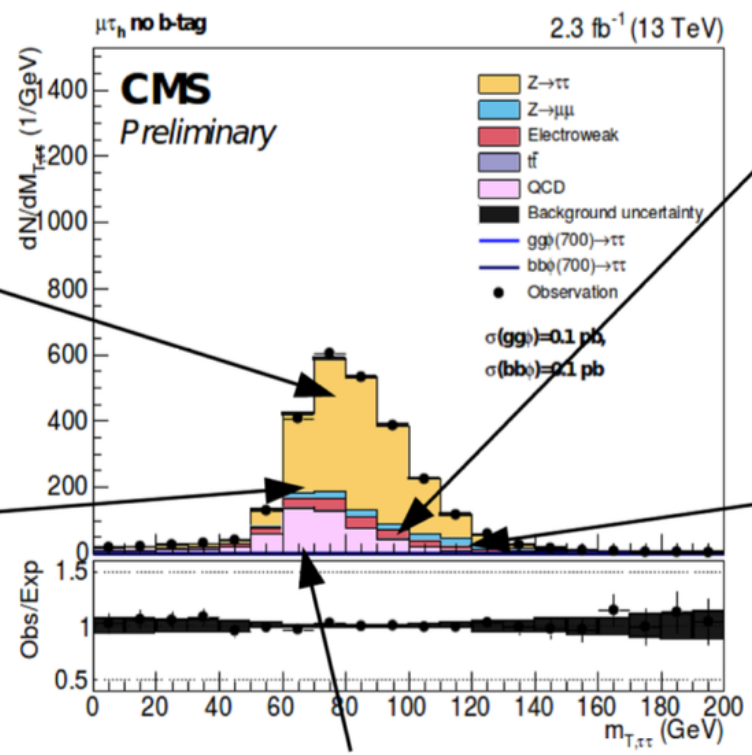
CMS HIG-16-006-PAS

- Analysis Background Methods
  - Utilize both MC based background and data driven backgrounds

$Z \rightarrow \tau\tau$:
- MC based, data driven corrections applied

$Z \rightarrow ll$:
- MC based

W+jets:
- $\mu\tau_h$ & $e\tau_h$ uses data driven scale factor
- $\tau_h\tau_h$ & $e\mu$: MC based



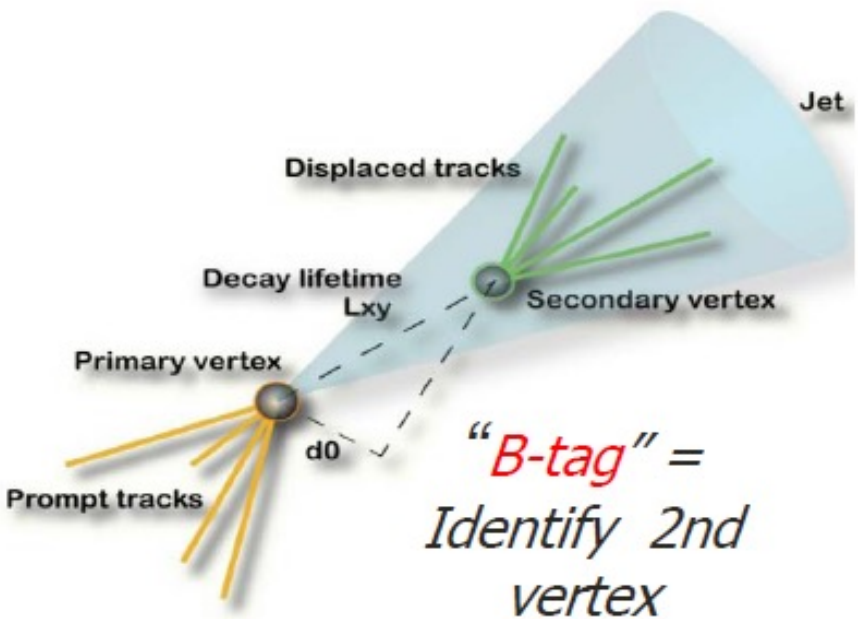
ttbar:
- MC based, validated in data control region

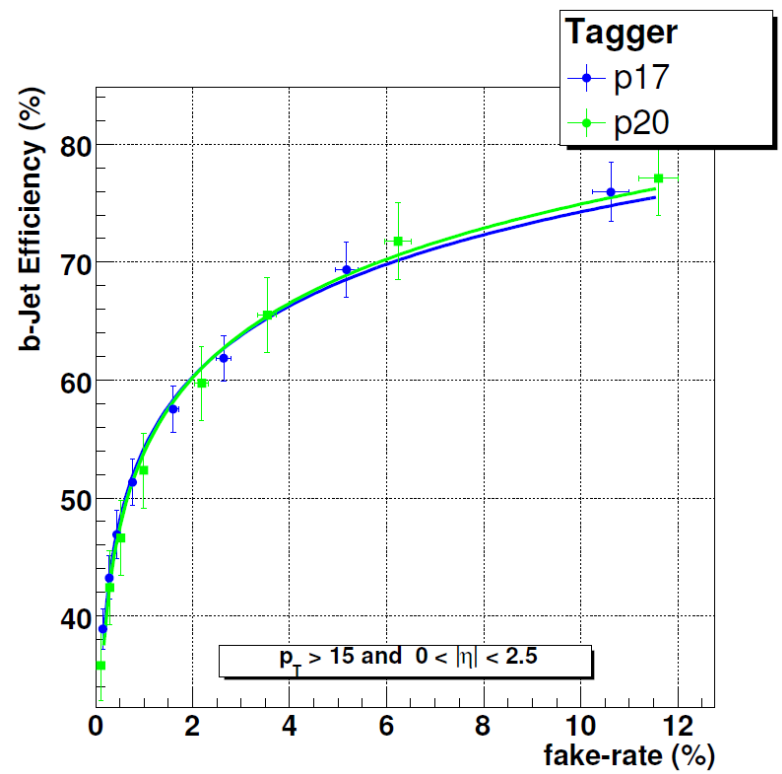
QCD: fully data driven estimation
- τ_h, $e\mu$ & $e\tau_h$ - shape from SS with OS/SS scale factor
- $\tau_h\tau_h$ - shape from relaxed loose isolation OS

Pooja Saxena

Identifying B hadrons at Colliders

- *b* decays with finite lifetime, travels millimeters before decay
  - Displaced vertex where b decays
  - Neural Network with jet and track variables, impact parameter
  - Typically 50-70% efficient for 0.5-1.5% fake rate
- Similar methods used by MARK II, ALEPH, DELPHI, L3, OPAL, SLD, CDF, D0, ATLAS, CMS





Slide from K. Herner

b-Tag Operating Points

DELPHI's 2000 re-calibration of detector parameters (TPC drift velocity and vertex detector alignment), reclassified a very high-weight Higgs boson candidate because it fell on the other side of a chisquared cut for the displaced vertex.

http://delphiwww.cern.ch/pubxx/delsec/talks/lepc/LEPCnov2000/lepc.ps.gz

Want to avoid high-weight events (that is, in bins with high s/b) next to cuts!

Would like a continuous b-tag discriminant variable, used as input to other MVA techniques.

But: Practical demands creep in.

Calibration of efficiency scale factors and mistag matrices required cut-based methods for practicality. One could generalize them to calibrate shapes, but data in control samples may not be enough.

Data Categorization

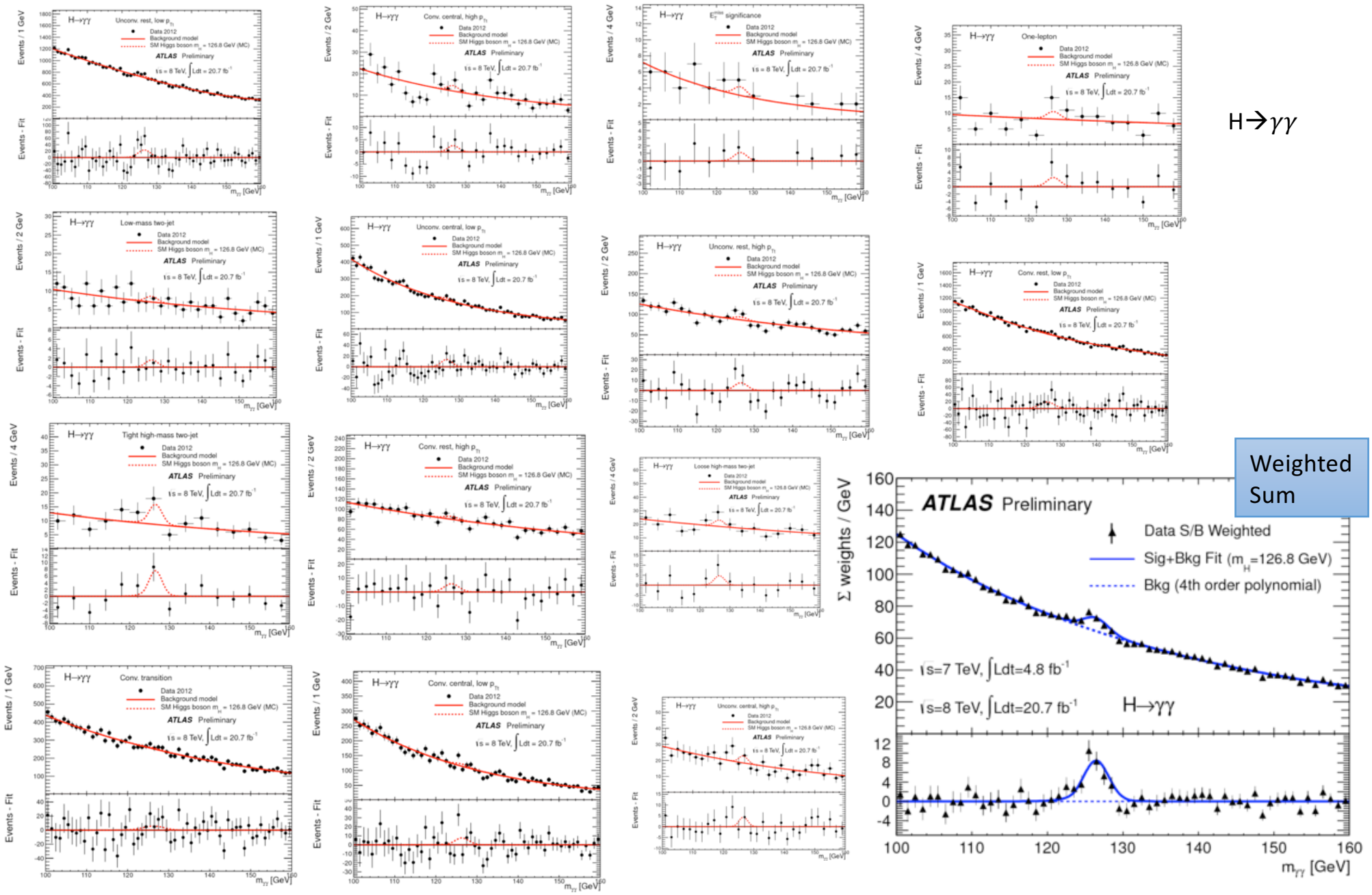
It is often useful to divide the data into subsamples with different characteristics

number of jets
number of b-tags
missing energy or no
leptons reconstructed or not
momentum of candidate particle
converted photons or no
looser selection cuts on leptons
barrel vs. endcap objects (better resolution or worse)
...

Even when searching for a single particle, like the Higgs boson. Different signal production mechanisms and different backgrounds will dominate in each sample.

Categorization of events to Improve Sensitivity

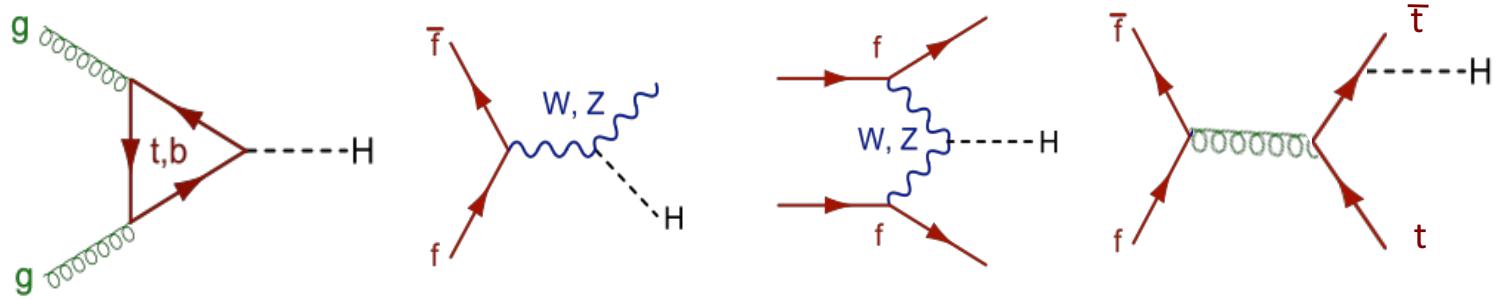
ATLAS-CONF-2013-012

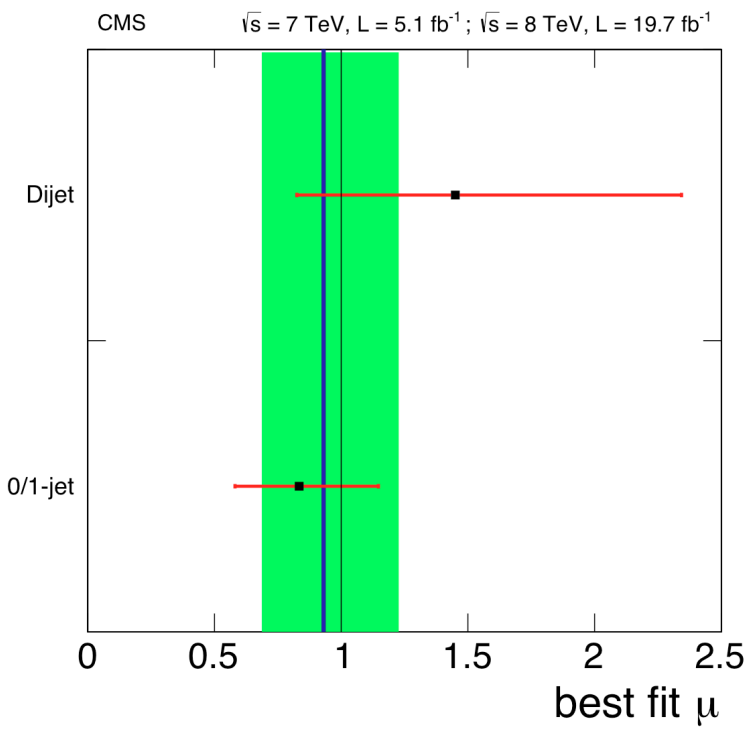


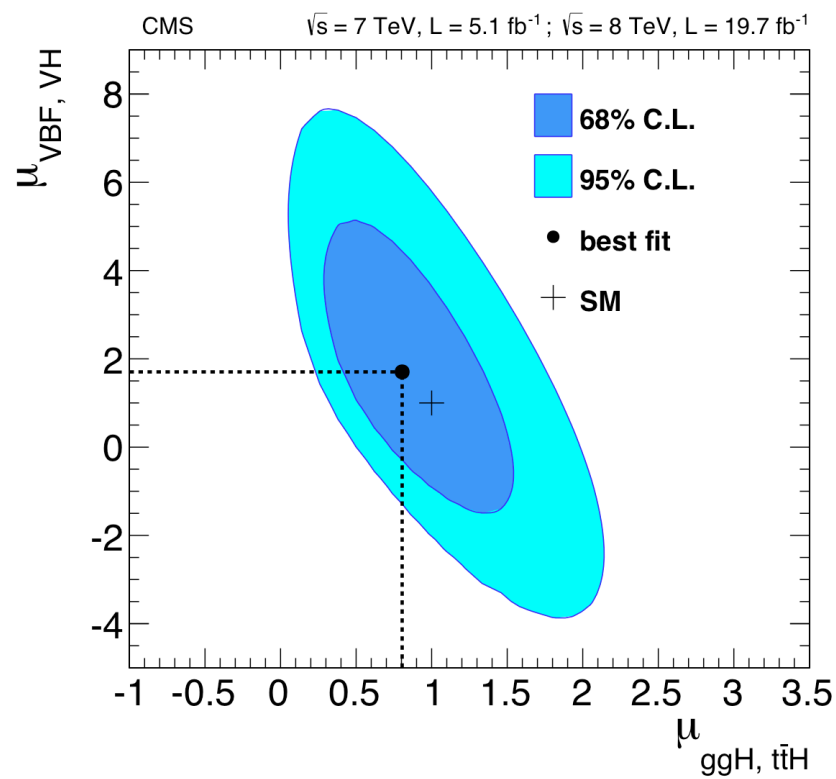
H→$\gamma\gamma$

Divide and Conquer has a Side Benefit – More Physics

Different analysis, but same idea.







Profile Likelihood used – fit for parameters of interest and nuisance parameters

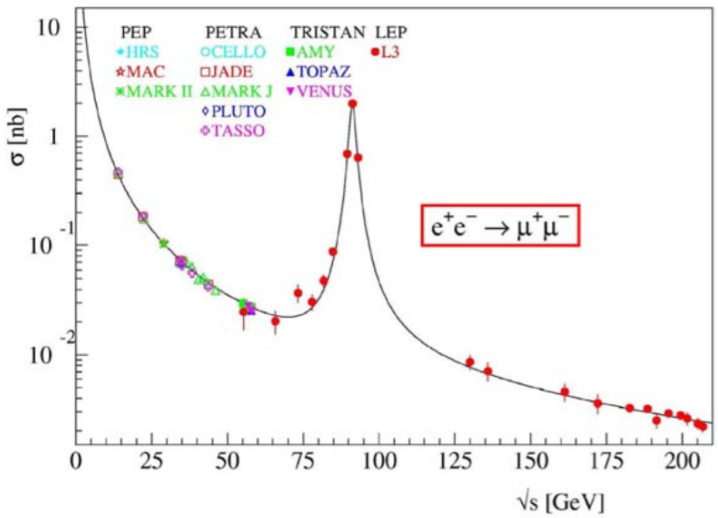
Several Analyses on the Same Data

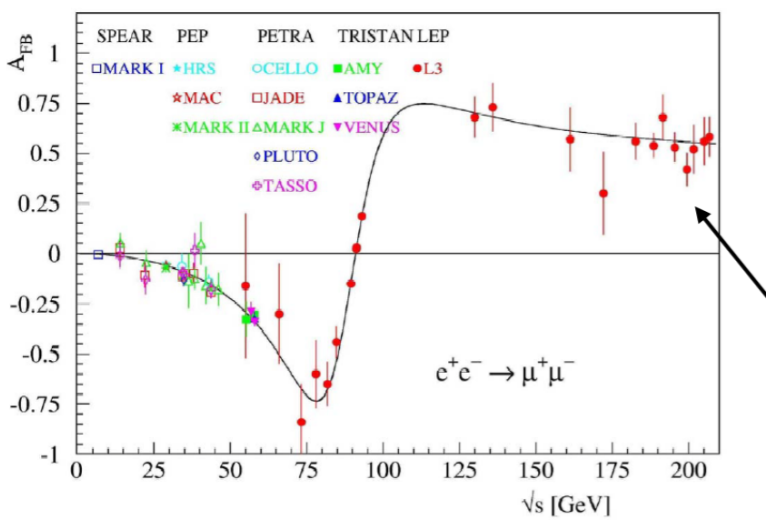
- Different groups are interested in the same search/measurement using the same data.
- May have slightly different selection requirements (Jet energies, lepton types, missing Et, etc).
- Usually have different choices of MVA or even training strategies for the same MVA
- Always will give different results!

- What to do?
  - Pick one and publish it – criterion: best sensitivity. Median expected limit, median expected discovery sensitivity, median expected measurement uncertainty. How to pick it if the result is 2D? Need a 1D figure of merit.
  - Can check consistency with pseudoexperiments. A p-value using Δ(measurement) as a test statistic. What's the chance of running two analyses on the same data and getting a result as discrepant as what we got?
  - Combine MVA's into a super-MVA
    - Keeps everyone happy and involved
    - Usually helps sensitivity
    - Requires coordination and alignment of each event in data and MC
    - Easiest when overlap in data samples is 100%. Otherwise have to break sample up into shared and non-shared subsets and analyze them separately
- What not to do: Pick the one with the "best" observed result. (LEE!)

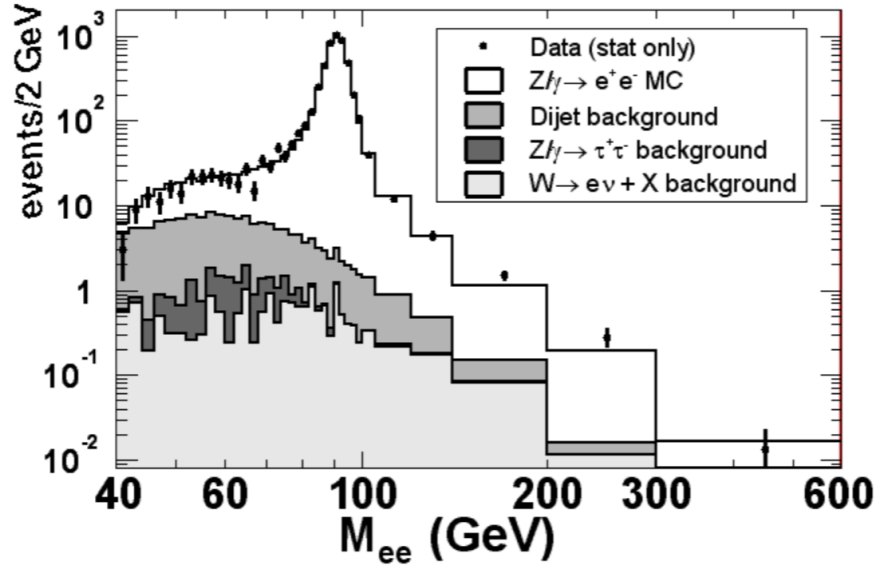
Forward-Backward Asymmetry of Leptons at a Hadron Collider and Unfolding

What it **should** look like – from lepton colliders



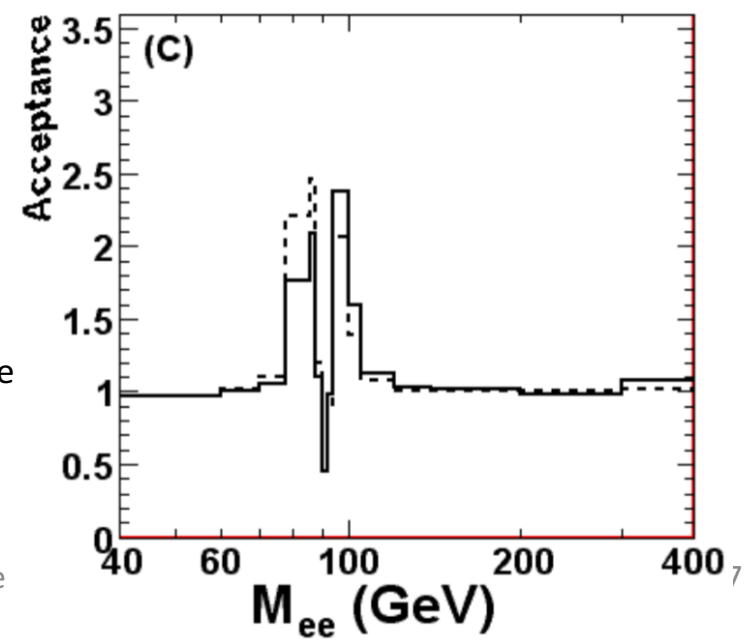


Big Z peak at the Tevatron too!



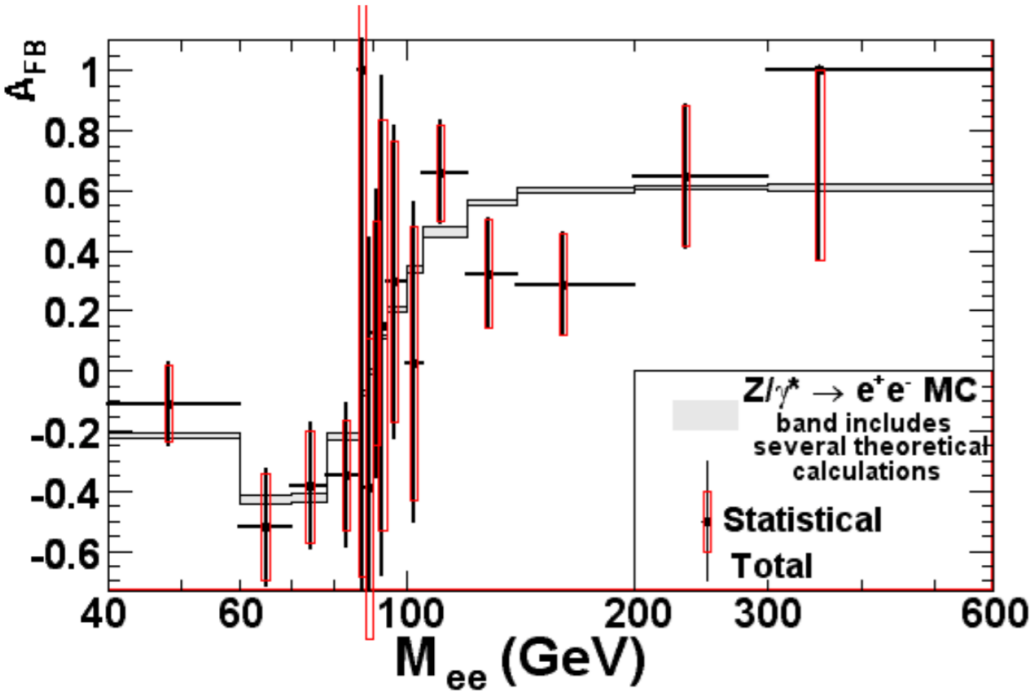
solid line:
forward

dashed line
backward



Three Attempts: Unfolding, Regularizing, and Injecting Model Shape

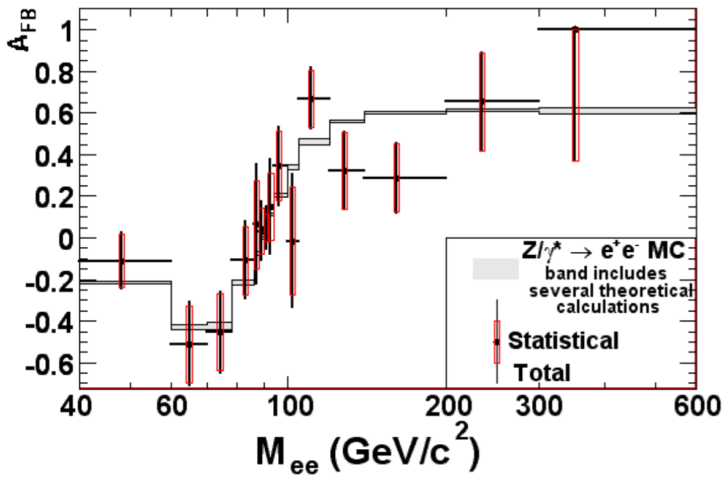
Invert a transfer matrix (four, actually, FF,FB, BF, BB) plot asymmetry



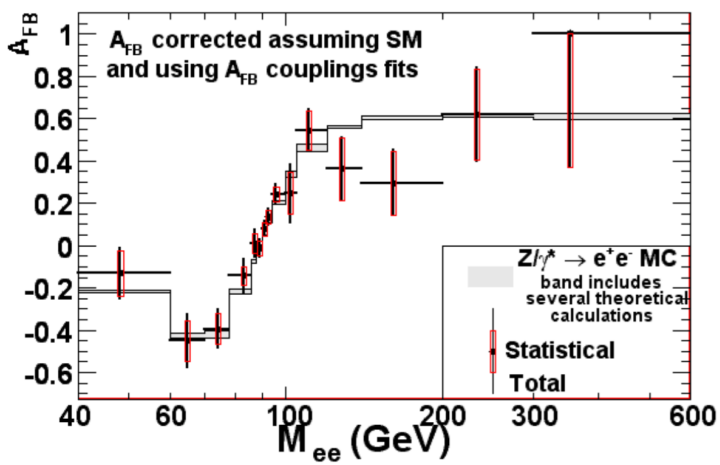
Z bosons spill over to nearby bins, obscuring their asymmetry.

Add Tikhonov Regularlization

- Result #1:



Result #2 (SM with arbitrary Z couplings):



Prior-Predictive Handling of Systematic Uncertainty

- Analyzers at the Tevatron and LEP typically used the prior-predictive prescription.
  - Examples: top quark discovery, single top observation, Higgs boson limits and p-values.

- Sometimes the supremum p-value was used, especially if the effect was very clear anyway and analyzers wanted to be conservative. Some examples in B physics

- In the prior-predictive prescription, systematics are varied for each pseudo-dataset, and then data are randomly drawn from the systematically varied predictions.

- It's a Mixed Bayesian-Frequentist method.

- Usual citation: R. Cousins and V. Highland, Nucl. Instrum. Meth. A320, 331 (1992).

LHC Higgs Prescription

ATLAS and CMS prescription for Higgs results combination (2011)
CMS-NOTE-2011-005 ; ATL-PHYS-PUB-2011-11
https://cds.cern.ch/record/1379837?ln=en

advocate the posterior-predictive method:

5. Generate toy Monte Carlo pseudo-data to construct *pdf*s $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\rm obs})$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\rm obs})$ assuming a signal with strength μ in the *signal+background* hypothesis and for the *background-only* hypothesis ($\mu = 0$). These distributions are shown in Fig. 1. Note, that for the purposes of generating a *pseudo-dataset*, the nuisance parameters are fixed to the values $\hat{\theta}_\mu^{\rm obs}$ or $\hat{\theta}_0^{\rm obs}$ obtained by fitting the observed data, but are allowed to float in fits needed to evaluate the test statistic. This way, in which the nuisance parameters are fixed to their maximum likelihood estimates, has good coverage properties [12].

Relationship to Pulls

http://www-cdf.fnal.gov/physics/statistics/notes/cdf5776_pulls.ps.gz
By Luc Demortier and Louis Lyons

CDF Note 5776, "Everything you always wanted to know about pulls" says in Section 5 that the method of generating toy experiments in the prior predictive method is wrong, at least when computing pulls for Gaussian constrained fits.

The pull distributions back it up! Fluctuating the constraint in the fit but not the values of the parameters in the toy experiments produces better-looking pulls.

In practice it rarely matters much.

Prior vs. Posterior Predictive Uncertainties

- Posterior predictive uncertainties – hard to predict what the performance of an experiment is going to be before the data come in. The same problem occurs with the prior-predictive method.

- We only analyze the data once, and we would like to simulate all possible data. Is it worthwhile also to simulate all possible versions of Nature (nuisance parameter values) that our analyses could be confronted with?

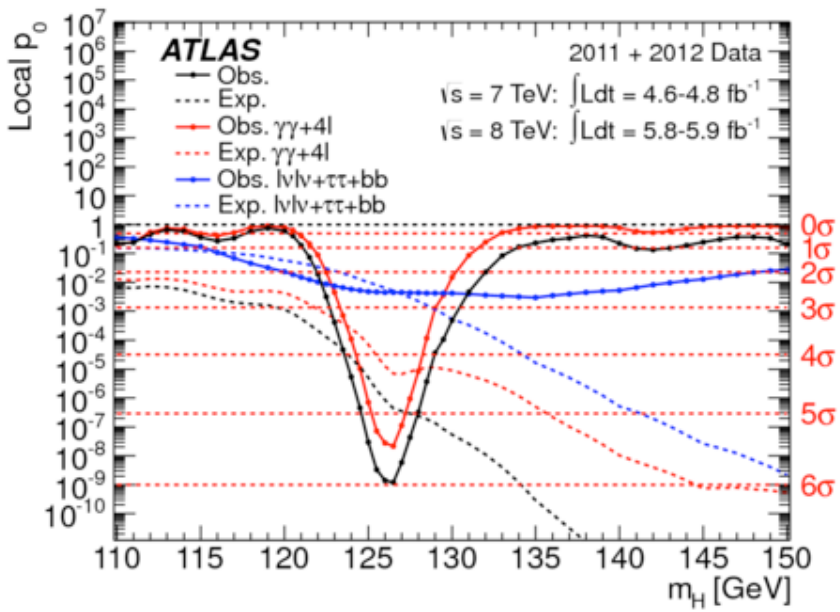
- LHC prescription splits outcomes in the distribution of the test statistic. For a counting experiment, the distribution is not just a set of delta functions at the observed counts in the signal sample, but each one is refined by what the control samples' outcomes are.

- Bob Cousins found this long ago -- you can improve the sensitivity of a search just by adding in some barely-relevant information, as long as it serves to break the discreteness of a low-rate counting experiment. This is a good thing!

- Can't use just any test statistic with LHC method, like the event count in the signal region. Must use a test statistic that ranks outcomes including the auxiliary outcomes (even when they are fictitious, as in theory uncertainties).

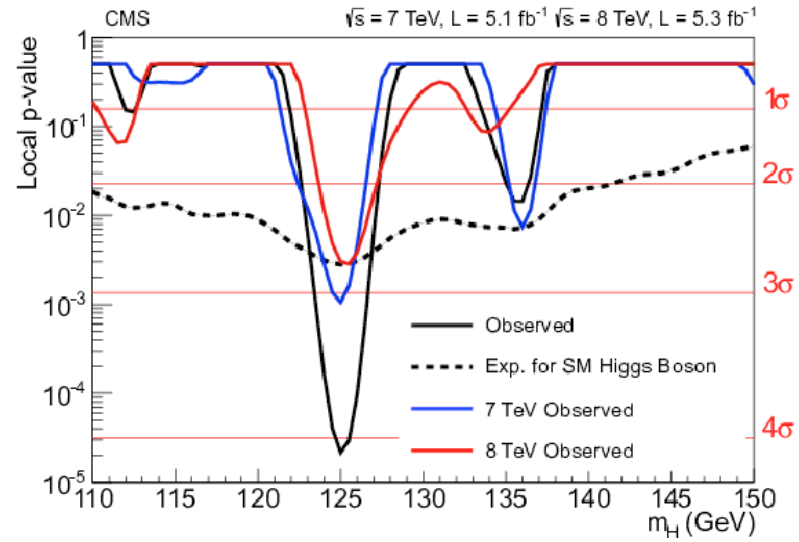
Sensitivity is reported to be slightly better where actual particles (or upward fluctuations) are.

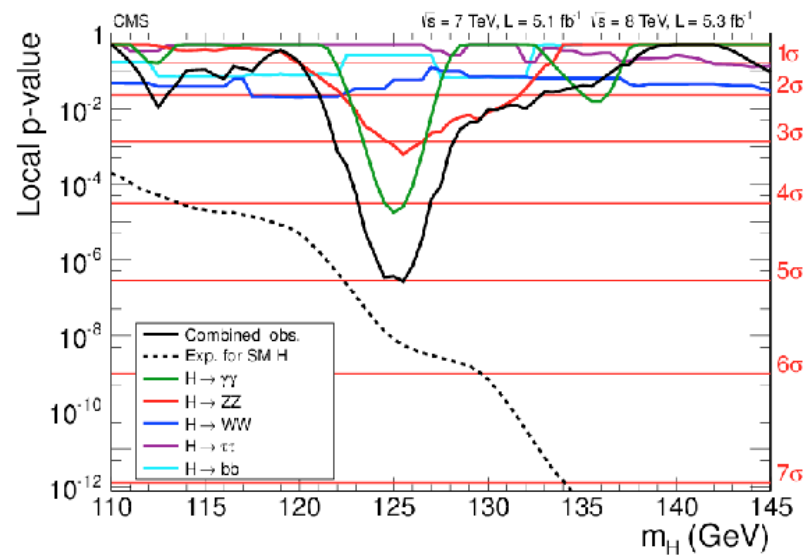
Reason – the best-fit values of the acceptances are higher when there is real signal to fit (or a fluctuation).



ATLAS Collaboration,

Phys. Lett. B 716 (2012) 1-29





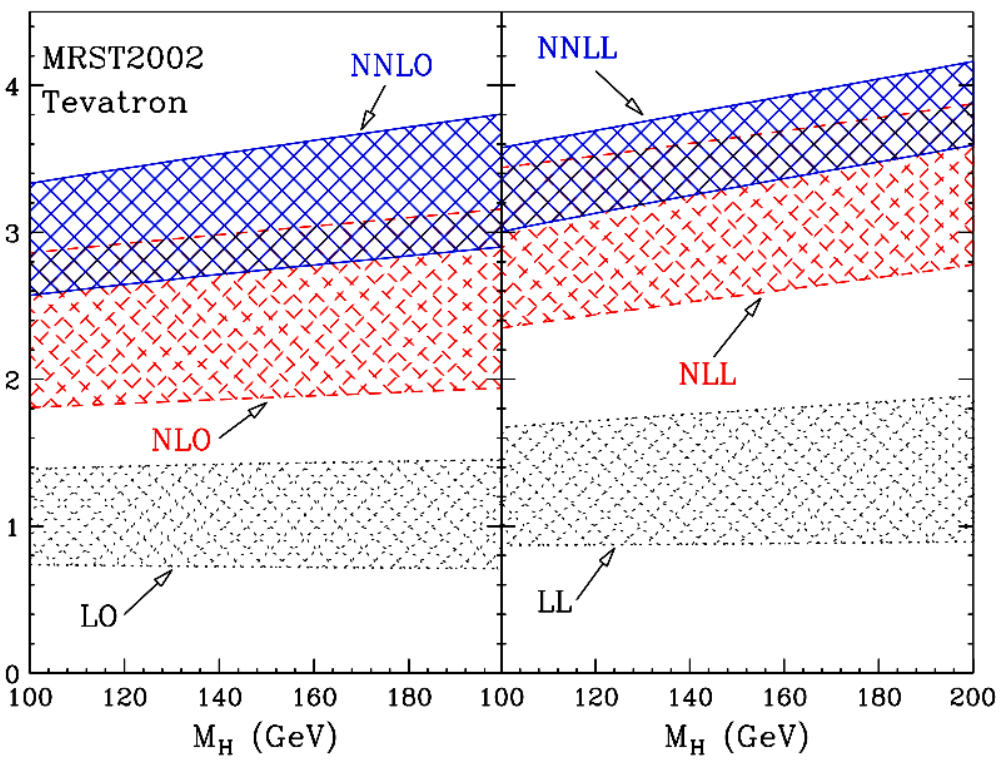
CMS Collaboration,
Phys. Lett. B 716 (2012) 30

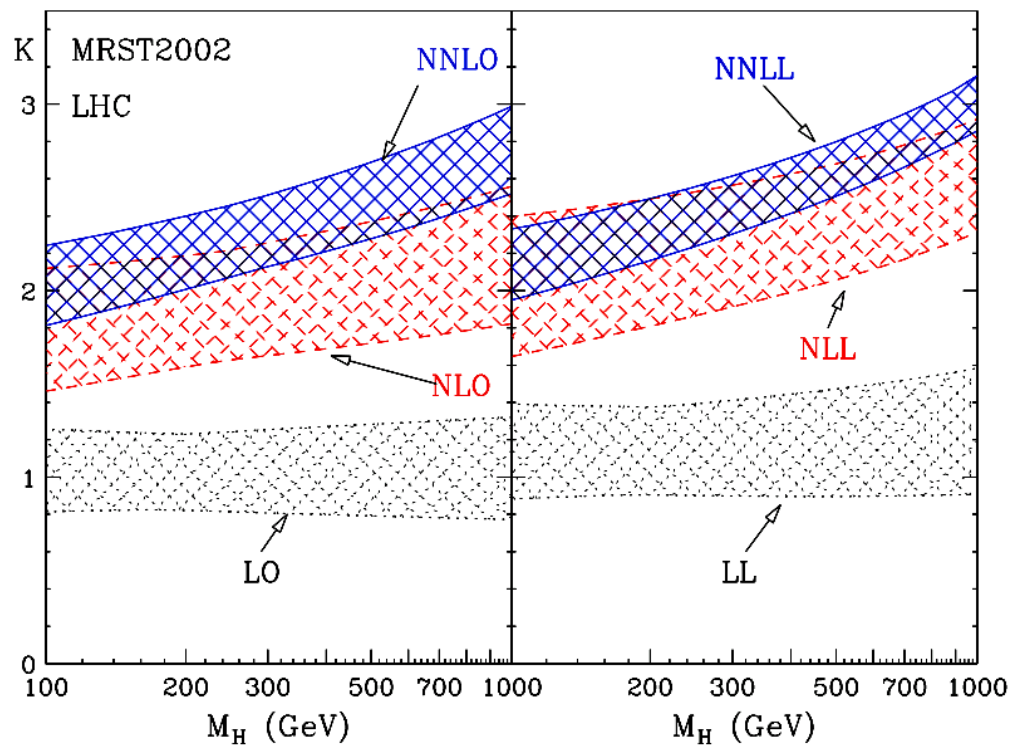
Theoretical Uncertainties

- Some push and pull here between theorists and experimentalists, especially in high-profile analyses, like Higgs boson work.
- Theorists often think of the uncertainties they provide as upper bounds and not $\pm 1\sigma$ uncertainties.
- Understandable when the uncertainty comes from a bound on a series sum. But this is often not the case (examples next slide).
- Theory uncertainties usually from factorization and renormalization scale, but theorists also use PDF uncertainties and call them theoretical.

Predictions of the cross section for gg→H

- Leading order is one-loop, so NLO involves more pieces -- gluon radiation from gluons, and more loops.
- NLO corrections -- ~80% (almost double the cross section)!
- NNLO QCD corrections -- An additional 40% on top of that! Residual uncertainty ~10%. Catani, de Florian, Grazzini, Nason JHEP **0307**, 028 (2003) hep-ph/0306211





NLO,NNLO bands: $0.5m_H<\mu_F,\mu_R<2M_H$. Bands on LO unreliable.

Special Handling of Theory Uncertainties

It has been proposed to handle theory uncertainties (THU) using a maximum-bound method while allowing prior-predictive methods for "parametric uncertainties" (PU).

S. Dittmaier et al., "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables" arXiv:1101.0593

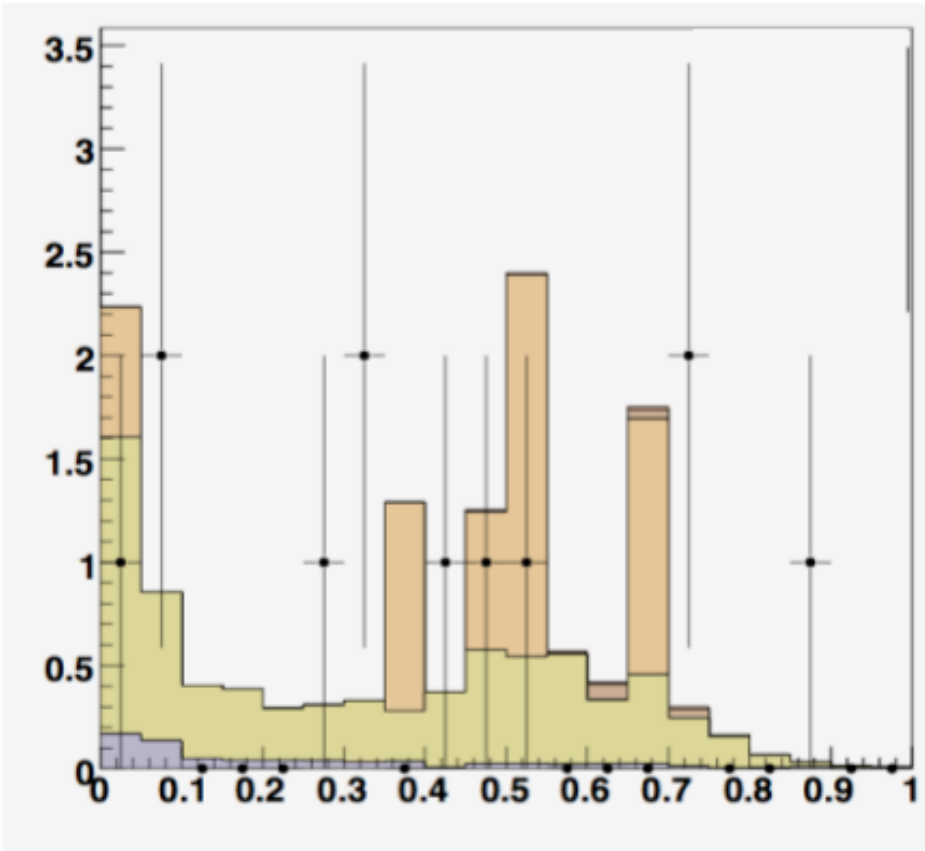
Sec. 12 proposes to add THU linearly, and PU in quadrature. The reasoning is suspect -- even though the text says Bayesian interpretation is possible, the choice of a flat prior inside the uncertainty region implies linear addition of uncertainty.

Usual prior-predictive methods are equivalent to adding uncertainties in quadrature.

One can draw THU parameter values from uniform priors, but adding them linearly implies correlation

A Pitfall -- Not Enough MC (data) To Make Adequate Predictions

An Extreme Example (names removed)



Cousins, Tucker and Linnemann tell us prior predictive p-values undercover with 0±0 events are predicted in a control sample.

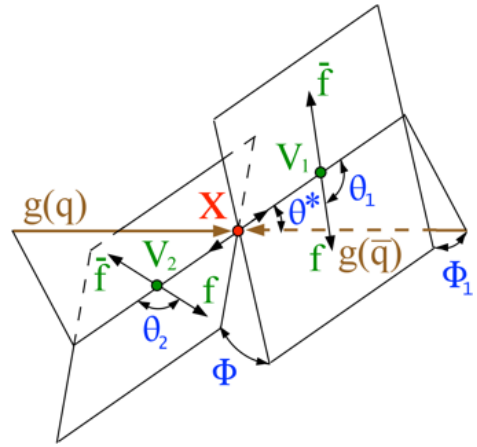
CTL Propose a flat prior in true rate, use joint LF in control and signal samples. Problem is, the mean expected event rate in the control sample is n_{obs}+1 in control sample. Fine binning → bias in background prediction.

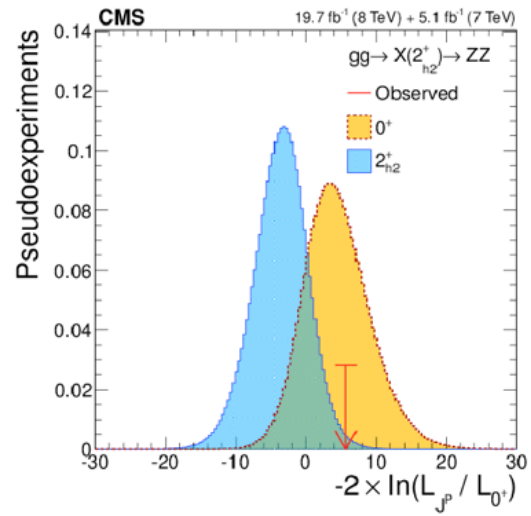
Questions: What's the shape we are trying to estimate? What is the uncertainty on that shape?

Overcovers for discovery, undercovers for limits?

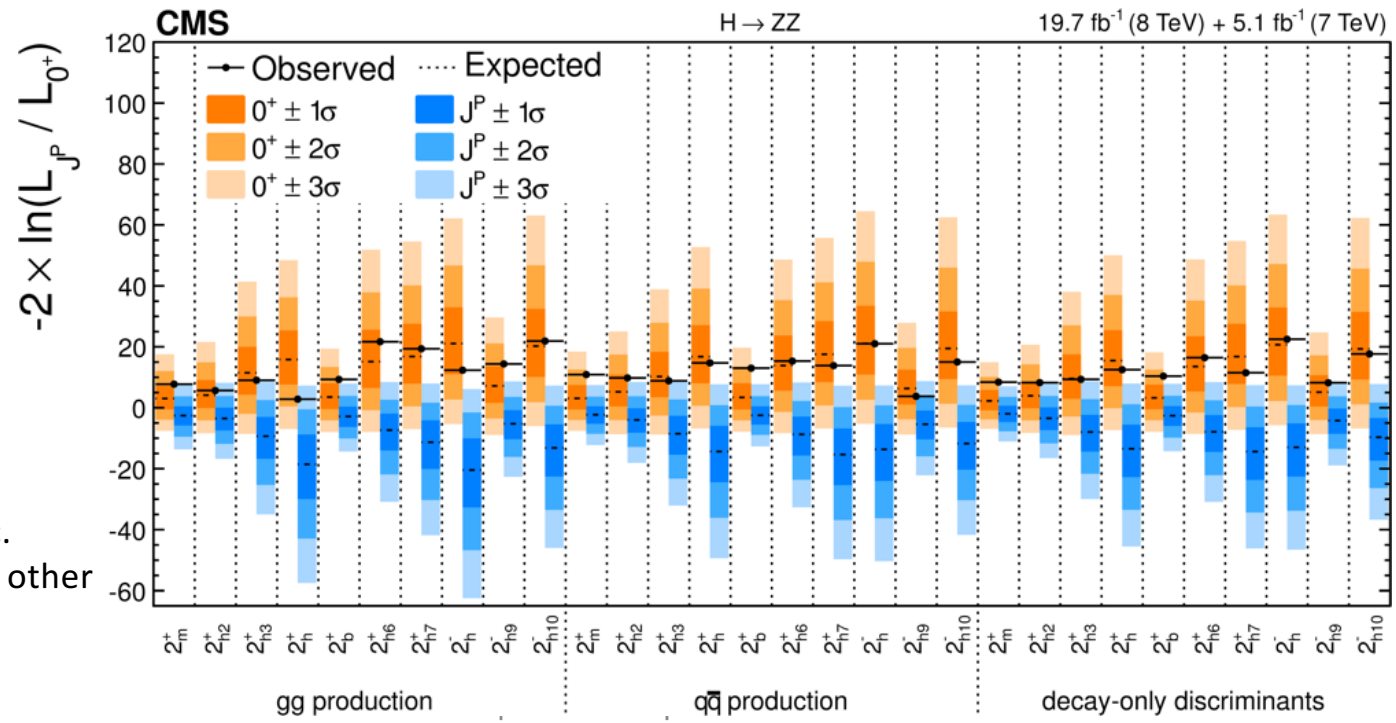
Examples of Two-Hypothesis Tests

Higgs Sin and Parity with H→ZZ and WW and $\gamma\gamma$ decays





CMS Collab., Phys. Rev. D 92 (2015) 012004



Clear Null Hypothesis: $J^P=0^+$. Others are exotic. Some even excluded by other considerations (like 1^+)

Sensitivity improves over time

Faster than 1/sqrt(exposure)

Analysis techniques improve!

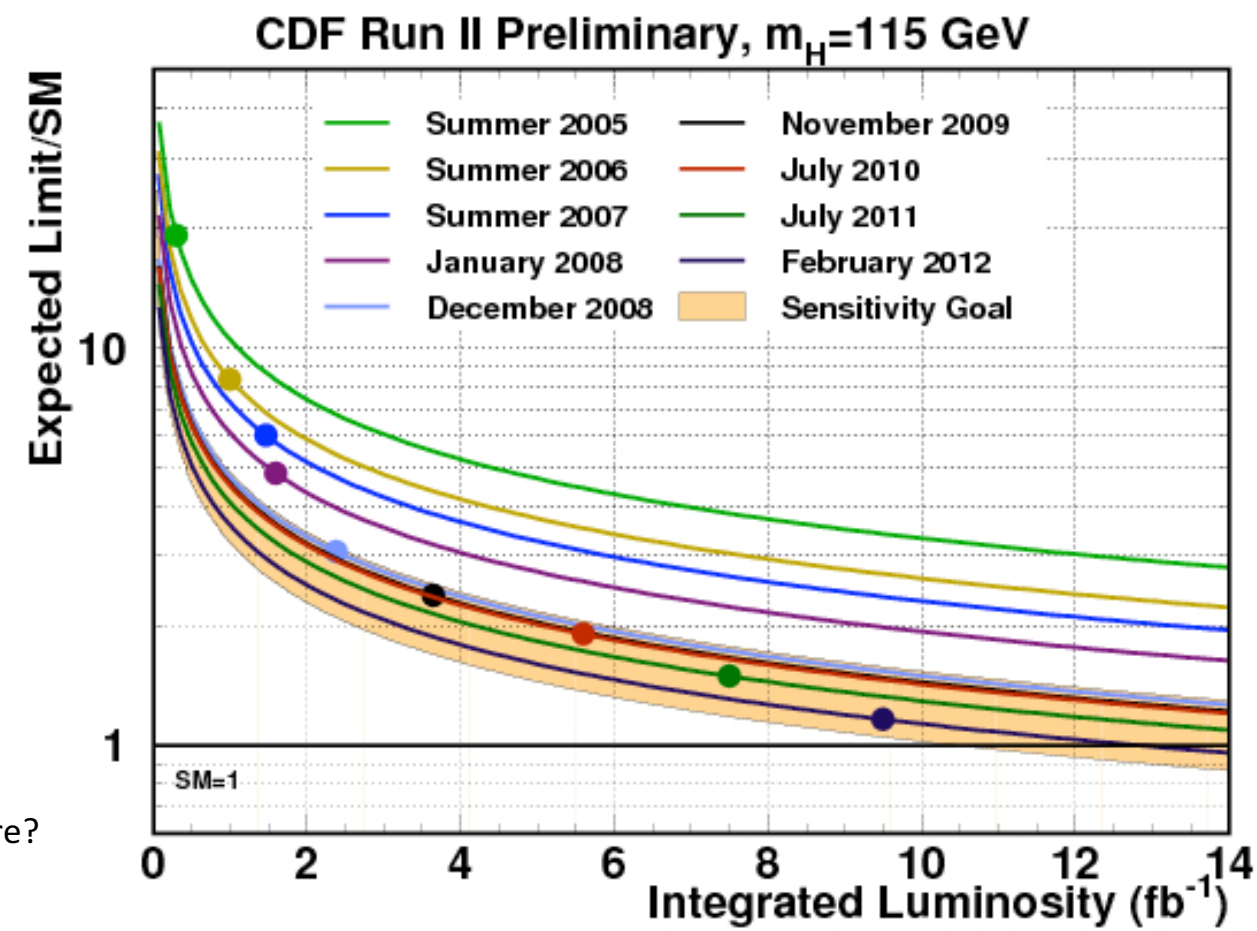
Control samples also acquire more data.

More channels are added, and existing ones divided into pieces.

Data constraints on nuisance parameters become more powerful than MC predictions

A challenge to predict how well an experiment will do before it is built!

How smart will we be decades in the future?



Extras

Matrix-Element Discriminants

- Calculate probability density of an event resulting from a given process

Phase space factor: Integrate over unknown or poorly measured quantities

Parton distribution functions

$$P(p_l^\mu, p_{j1}^\mu, p_{j2}^\mu) = \frac{1}{\sigma} \int d\rho_{j1} d\rho_{j2} dp_\nu^z \sum_{comb} \phi_4 \, | M(p_i^\mu) |^2 \, \frac{f(q_1) f(q_2)}{| q_1 || q_2 |} W_{jet}(E_{jet}, E_{part})$$

Inputs: lepton and jet 4-vectors – no other information needed!

Matrix element: Different for each process. Leading order, obtained from MadGraph

Transfer functions: Account for detector effects in measurement of jet energy

- The input variables are the same for all matrix elements – adding a new matrix element requires more calculation but does not use any different information from the data

Matrix-Element Discriminants

In principle, nothing performs better than these.

If processes cannot be separated because they contribute to the final state in the same way, this is all there is.

BUT:

- Four-vectors are imperfectly measured. Transfer functions are also imperfect.

- Only the modeling needs systematics; construction of the discriminant does not incur additional systematics, so even if the discriminant is imperfect or naive, it's okay – just an optimization question.

- Matrix elements are usually leading-order only.

- Particles are sometimes not reconstructed at all, even when they should be

- Some processes do not have well defined matrix elements – like data-derived fakes.

- Non-kinematic information is important, too, such as b-tags (help reduce combinatorics)

- Not clear whether integrating over all possibilities or just picking the best one is the most optimal for the purposes we set out for (more on this later).

How Likely is it?

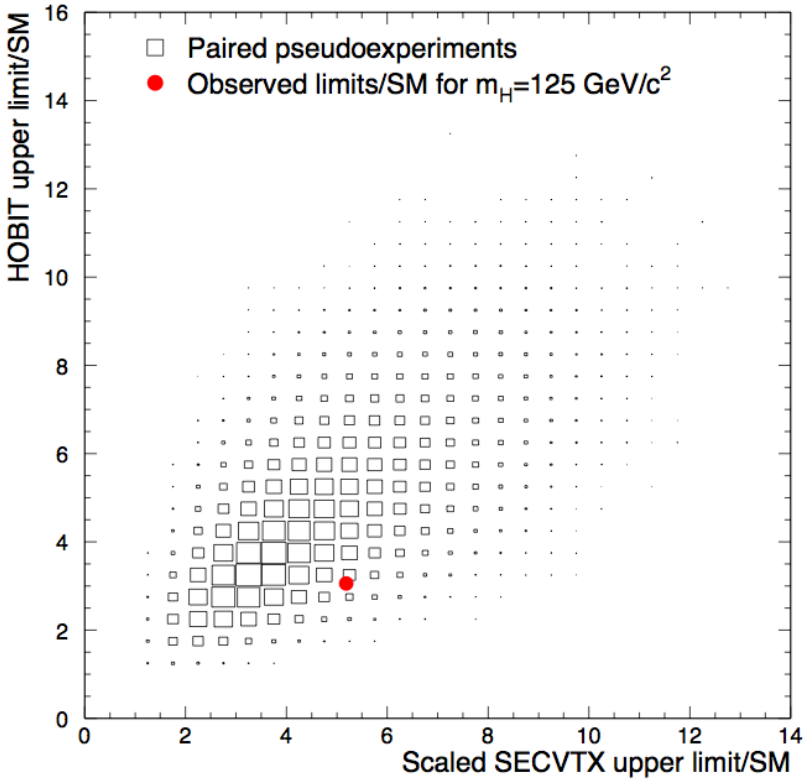
Compare CDF METbb HOBIT analysis with the re-done SECVTX+JP analysis

Paired pseudoexperiments:

- HOBIT channels TT, TL, 1T
- SECVTX channels SS, SJ, 1S
- Use expected overlaps in the $NN_{SIG} < 0.8$ and $NN_{SIG} > 0.8$ regions to generate independent samples of events in 15 categories:

TTSS, TTSJ, TT1S, TTnone
TLSS, TLSJ, TL1S, TLnone
1TSS, 1TSJ, 1T1S, 1Tnone
noneSS, noneSJ, none1S

Calculate limit for each pair of pseudoexperiments & compare



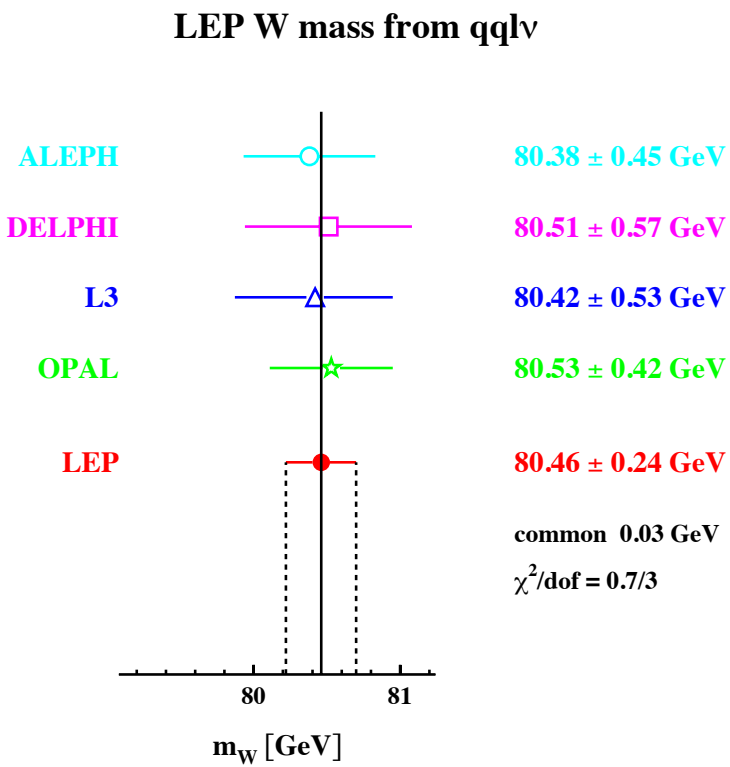
P-value for |Limit difference| to be as big as observed: 7%
Highly correlated over m_H range:
P-value for all limits to be low:
3-5%

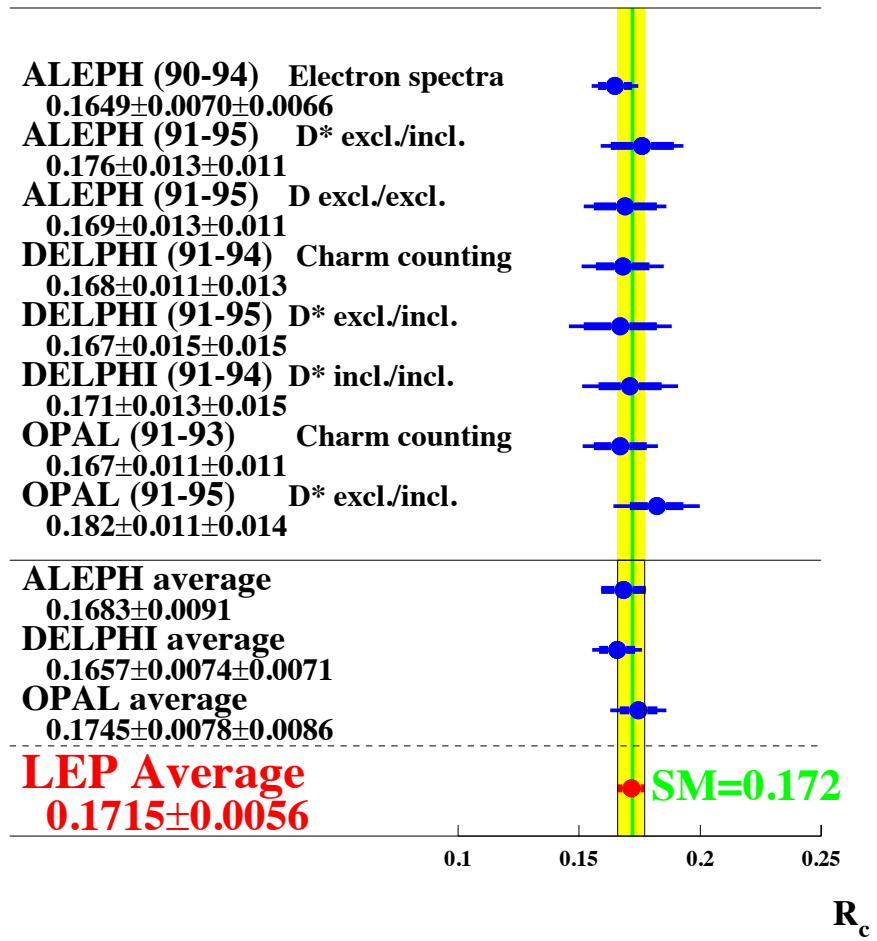
http://www-cdf.fnal.gov/~trj/wc_trj_cdfhiggs_18jan_pub.pdf

CDF Collab, Phys.Rev. D **87** (2013) 052008

Some LEP Plots from lepewwg.cern.ch (1997 and 1996) with very small chisquare sums

(cherry-picked, I confess)





A Dilemma – Can't we test just *one* model?

Something experimentalists come up with from time to time:

- Make distributions of every conceivable reconstructed quantity
- Compare data with Standard Model Predictions
- Use to test whether the Standard Model can be excluded
- Example: CDF's Global Search for New Physics Phys.Rev. D **79** (2009) 011101

The case *for* doing this:

- We might miss something big and obvious in the data if we didn't
- Searches that are motivated by specific new physics models may point us away from actual new physics.

More potential for discovery if you look in more places.

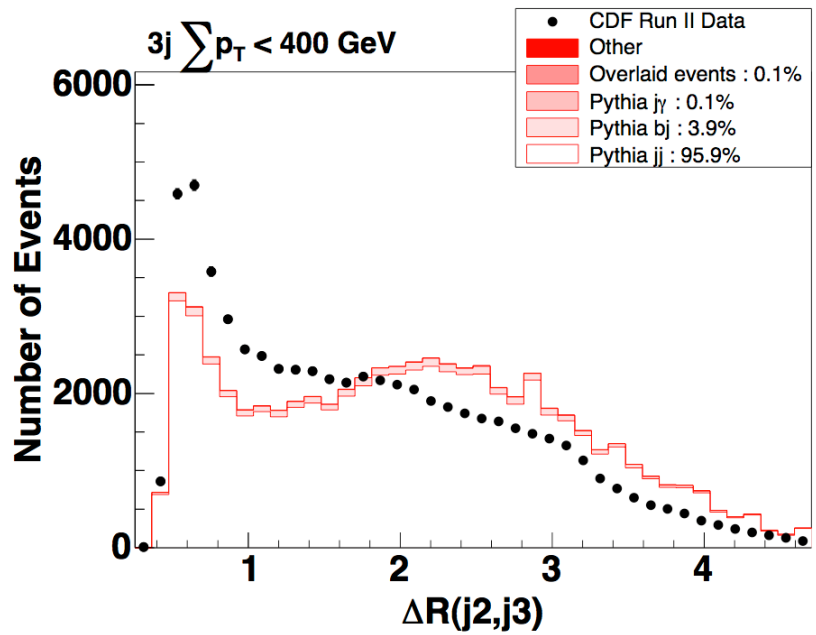
Testing Just One Model – Difficulties in Interpretation

- Look-Elsewhere Effect can be very large. But at least it is well defined.

- More worrisome is what to do when systematic flaws in the modeling are discovered.

Example: angular separation between the two least energetic jets in three-jet events.

Not taken as a sign of new physics, but rather as an indication of either generator (Pythia) or detector simulation (CDF's GEANT simulation) mismodeling. Or an issue with modeling trigger biases. Each of these is a responsibility of a different group of people.



Phys.Rev. D79 (2009) 011101

Testing Just One Model – Difficulties in Interpretation

- Need a definition of what counts as "interesting" and what's not. Already, using triggered events at a high-energy collider is a motivation for seeking highly-energetic processes, or signatures of massive new particles previously inaccessible.

- Analyzers chose to make ΣP_T distributions for all topologies and investigate the high ends, seeking discrepancies.

We just lost some generality! Some new physics may now escape detection.

But we now have alternate hypotheses – no longer are we just testing the SM (really our clumsy Monte Carlo representation of it).

Boxed into a corner trying to test just one model

- Of course our MC is wrong (that's what systematic uncertainty is for)
- Of course the SM is incomplete (but is it enough to describe our data?)

But without specifying an alternative hypothesis, we cannot exclude the null hypothesis ("maybe it's a fluctuation. Maybe it's mismodeling.")

An Ambitious Goal...

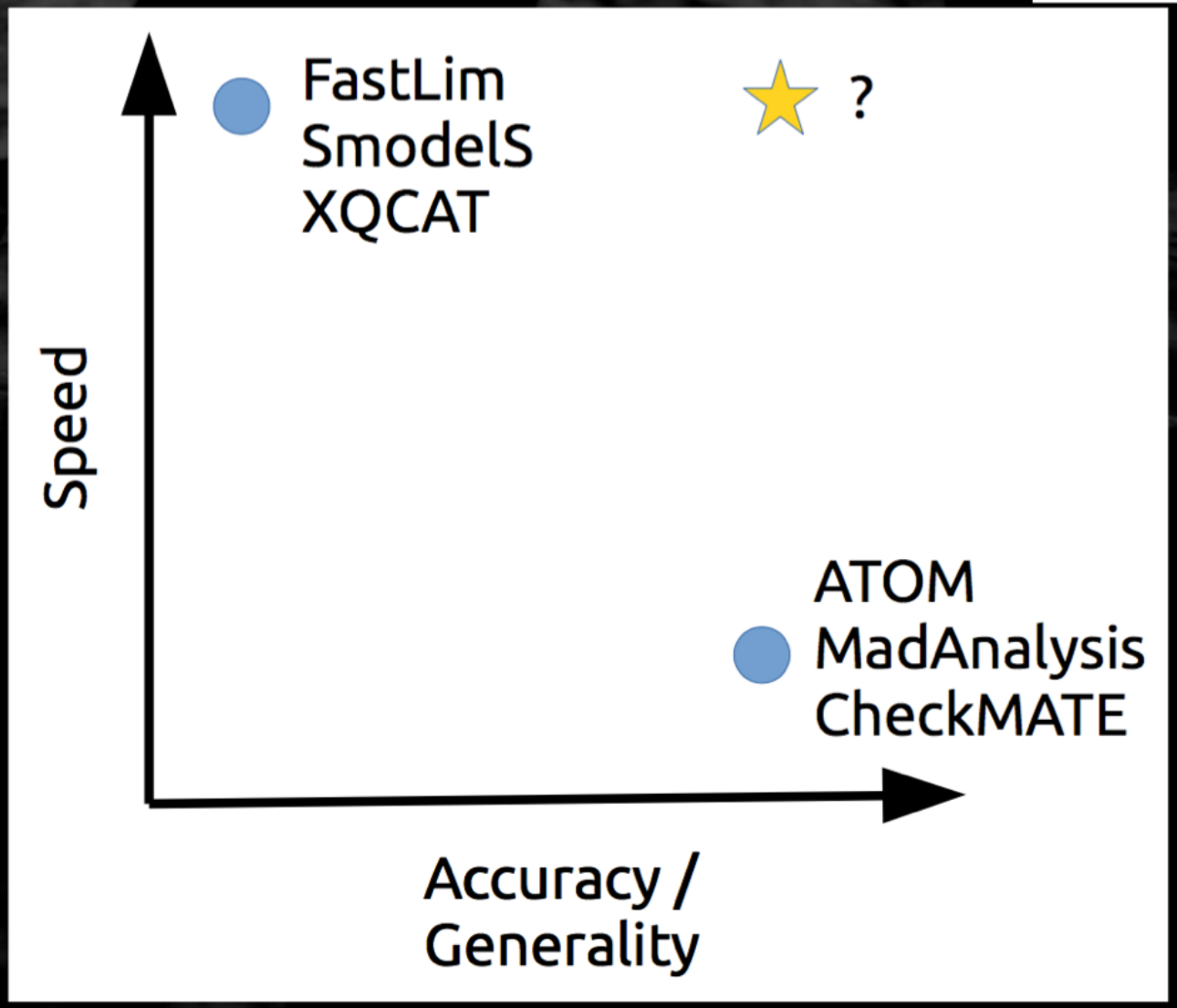
What do we want?

Write the Lagrangian down

Press "Enter"

Computer fits the model parameters to all relevant observables

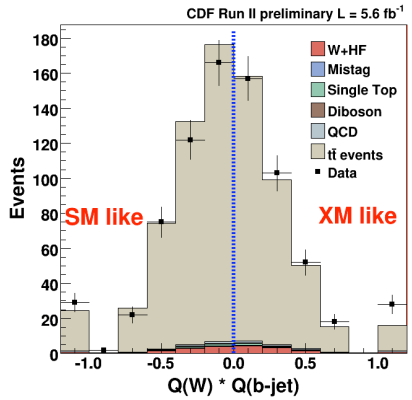
Can we do better?

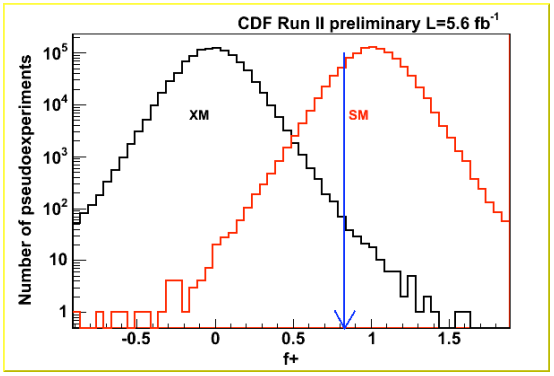


Another Classical Two-Hypothesis Test: What is the Charge on the Top Quark?

$$t \rightarrow W^+ b$$
$$\text{or} \quad t \rightarrow W^+ \bar{b}$$

H_0: q(top) = 2*e*/3
H_1: q(top) = 4*e*/3.





These are the only allowed possiblities assuming t→Wb (Wbbar) proceeds.

See: CDF Collab., **Phys.Rev.Lett. 105 (2010) 101801.** Even here we introduced no-decision regions to keep with our 95% CL exclusion and 3σ, 5σ conventions

Author Lists

Version 2.5 FERMILAB-PUB-10-xxx-E

Combination of Tevatron searches for the standard model Higgs boson in the W^+W^- decay mode

(The CDF and D0 Collaborations)

[110] Rutgers University, Piscataway, New Jersey 08855, USA
[111] Princeton University, Princeton, New Jersey 08544, USA
[112] University of New Mexico, Albuquerque, New Mexico 87131, USA
[113] State University of New York, Buffalo, New York 14260, USA
[114] Columbia University, New York, New York 10027, USA
[115] The Rockefeller University, New York, New York 10021, USA
[116] University of Rochester, Rochester, New York 14627, USA
[117] State University of New York, Stony Brook, New York 11794, USA
[118] Brookhaven National Laboratory, Upton, New York 11973, USA
[119] Duke University, Durham, North Carolina 27708, USA
[120] The Ohio State University, Columbus, Ohio 43210, USA
[121] Langston University, Langston, Oklahoma 73050, USA
[122] University of Oklahoma, Norman, Oklahoma 73019, USA
[123] Oklahoma State University, Stillwater, Oklahoma 74078, USA
[124] University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
[125] Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
[126] University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA
[127] Brown University, Providence, Rhode Island 02912, USA
[128] University of Texas, Arlington, Texas 76019, USA
[129] Texas A&M University, College Station, Texas 77843, USA
[130] Southern Methodist University, Dallas, Texas 75275, USA
[131] Rice University, Houston, Texas 77005, USA
[132] Baylor University, Waco, Texas 76798, USA
[133] University of Virginia, Charlottesville, Virginia 22901, USA
[134] University of Washington, Seattle, Washington 98195, USA
[135] University of Wisconsin, Madison, Wisconsin 53706, USA
(Dated: January 11, 2010)

A CDF+D0 Author list circa 2010., with 1042 authors.
About ¼ the size of just one LHC author list!

https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ListAuthors

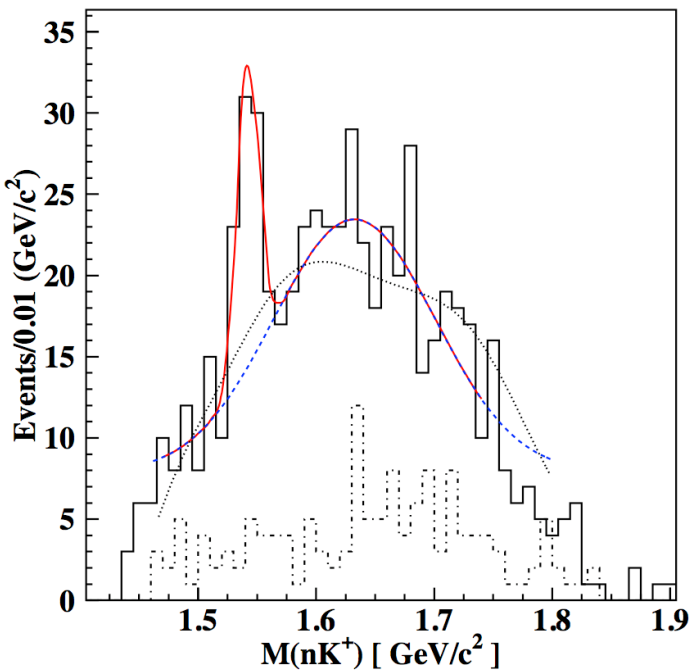
Mixtures with different weights

MC samples are often weighted (matrix elements)

Negative weights!

Example plot

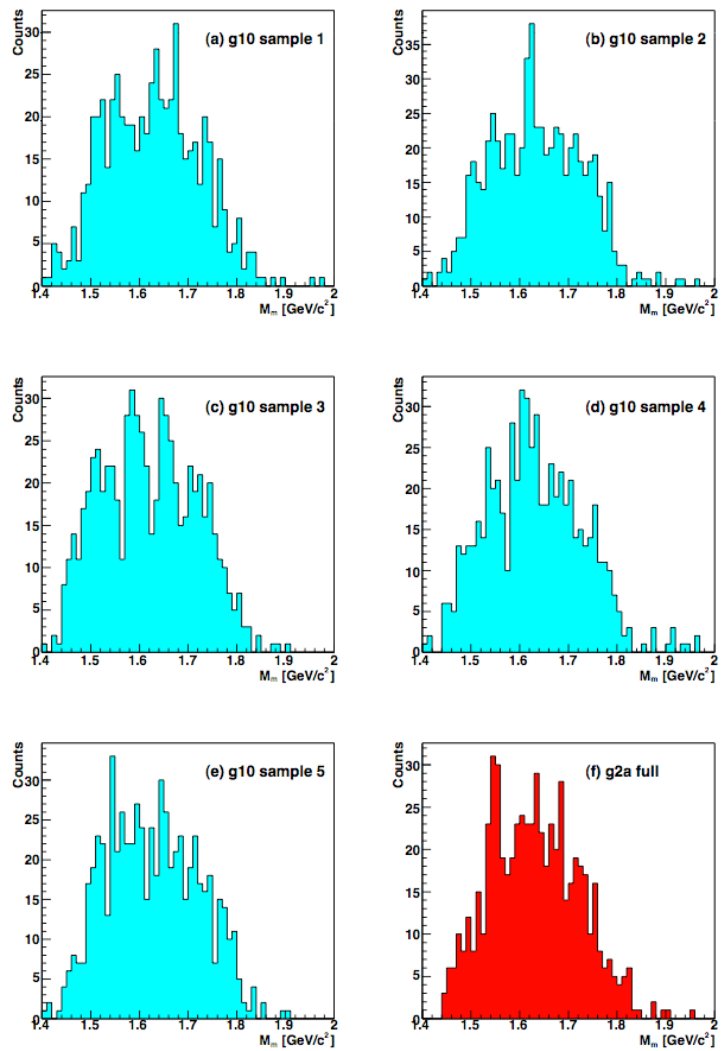
A Cautionary Tale – The Pentaquark "Discoveries"



CLAS Collab., **Phys.Rev.Lett. 91 (2003) 252001**

Significance = 5.2 ± 0.6 σ

Watch out for the background function parameterization!

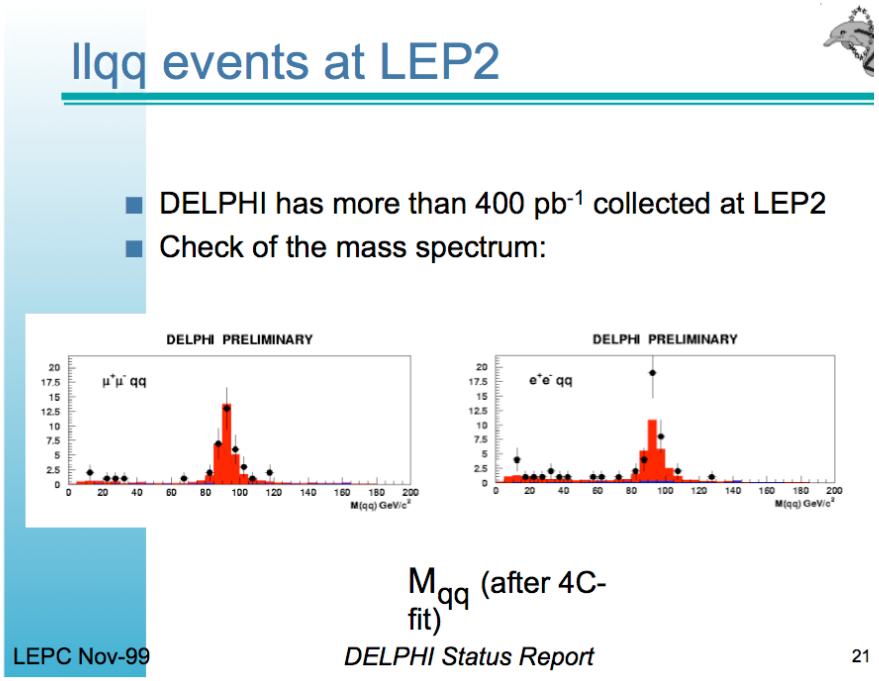


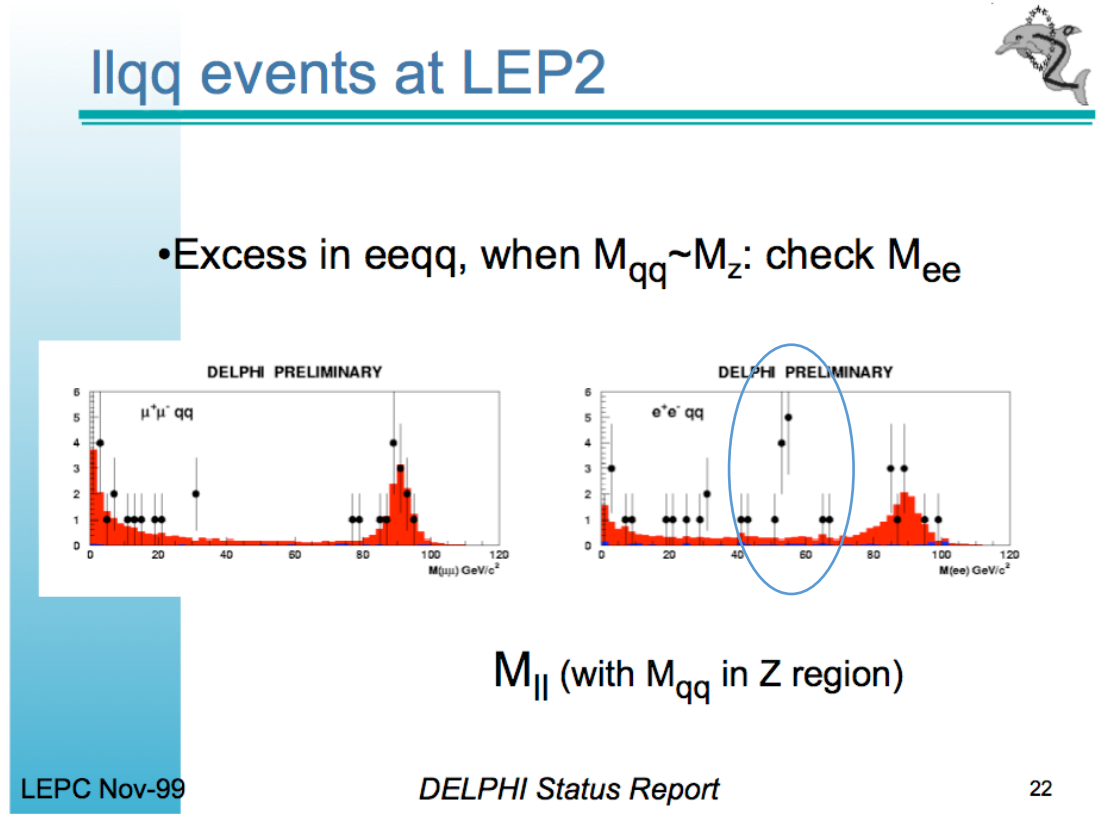
Five times the data sample
CLAS Collab., **Phys.Rev.Lett. 100 (2008) 052001**

n.b. the Bayesian analysis in this paper is flawed –
see the criticism by R. Cousins, **Phys.Rev.Lett. 101 (2008) 029101**

Another Bump That Went Away

A preliminary set of distributions shown at a LEPC presentation





Benefit of having four LEP experiments – at the very least, there's more data.
This one was handled very well – cross checked carefully.

But, they shared models – Monte Carlo programs, and theoretical calculations.

The Literature is Full of Bumps that Went Away

See Sheldon Stone, “Pathological Science”, **hep-ph/0010295**

My personal favorite is the “Split A_2 resonance”

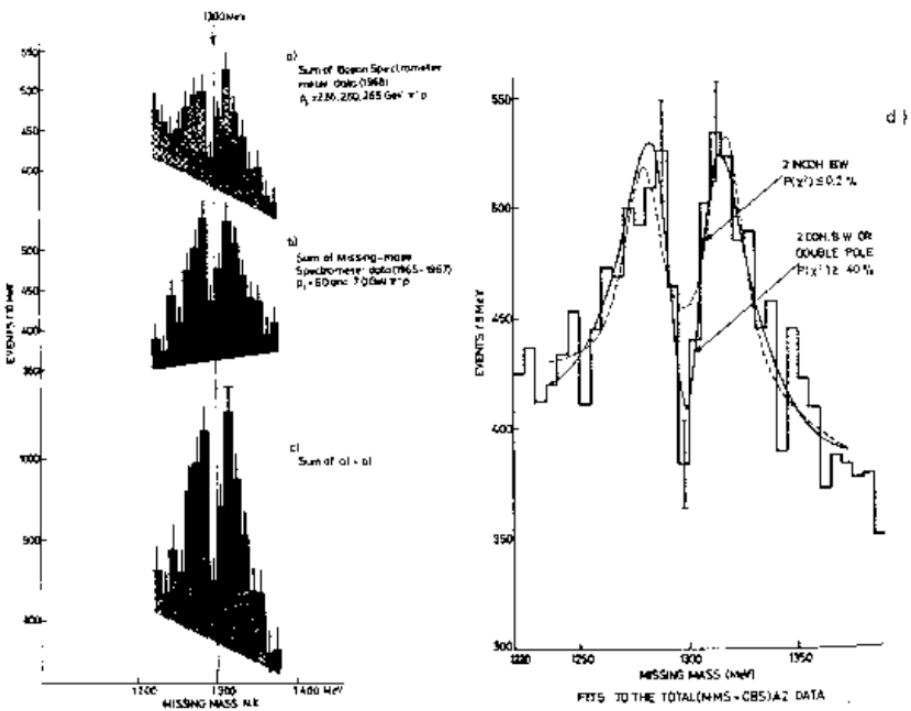


Figure 3: (a-c) Evidence for A_2 splitting in $\pi^- p \to pX^-$ collisions in the two CERN experiments, (d) same as (c) in 5 MeV bins fit to two hypotheses.

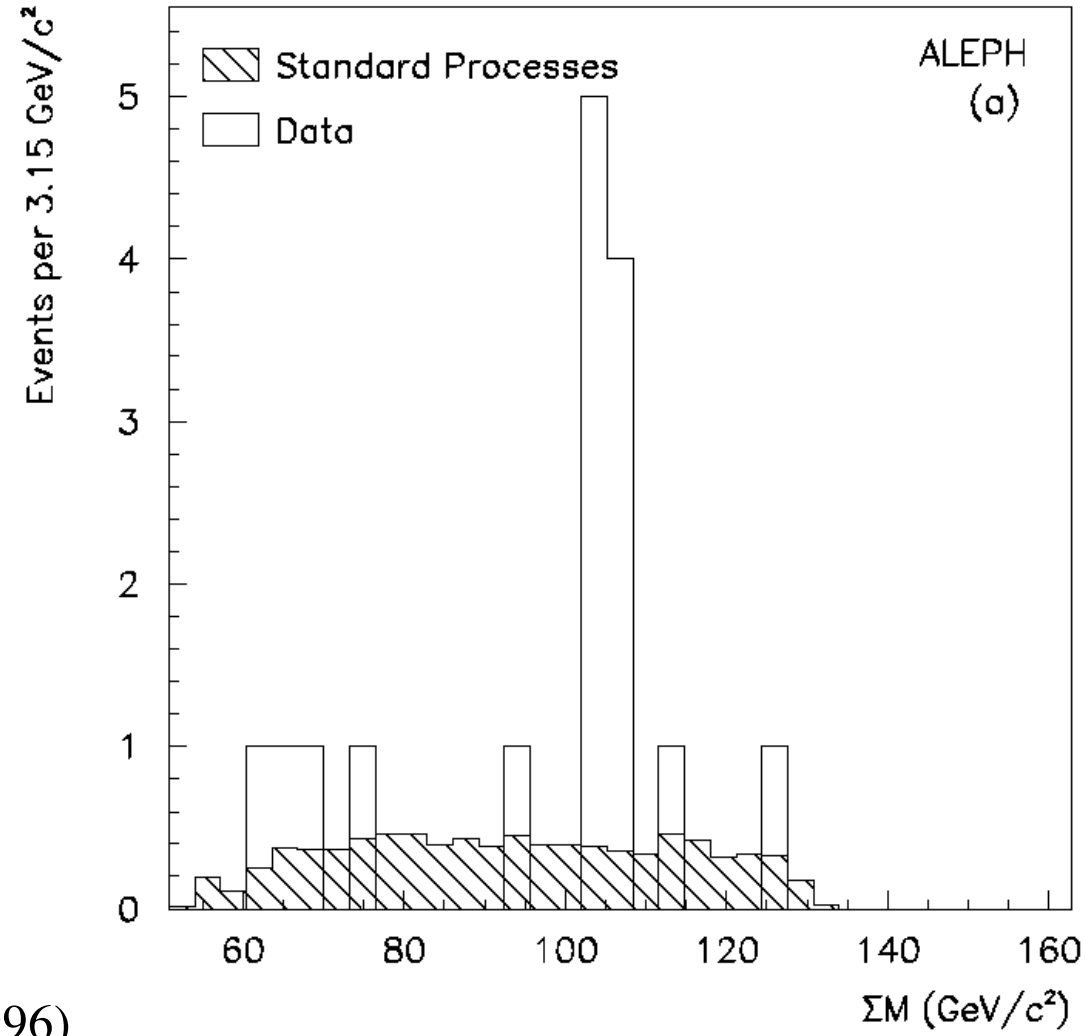
Text from Sheldon’s article:

How did this happen? I have heard several possible explanations. In the MMS experiment, I was told that they adjusted the beam energy so the dip always lined up! Another possibility was revealed in a conversation I had with Schübelin, one of the CBS physicists. He said: “The dip was a clear feature. Whenever we didn’t see the dip during a run we checked the apparatus and always found something wrong.” I then asked him if they checked the apparatus when they did see the dip, and he didn’t answer.

What about the other experiments that did see the dip? Well there were several experiments that didn’t see it. Most people who didn’t see it had less statistics or poorer resolution than the CERN experiments, so they just kept quiet. Those that had a small fluctuation toward a dip worked on it until it was publishable; they looked at different decay modes or t intervals, etc. (This is my guess.)

At Least ALEPH Explained what They Did

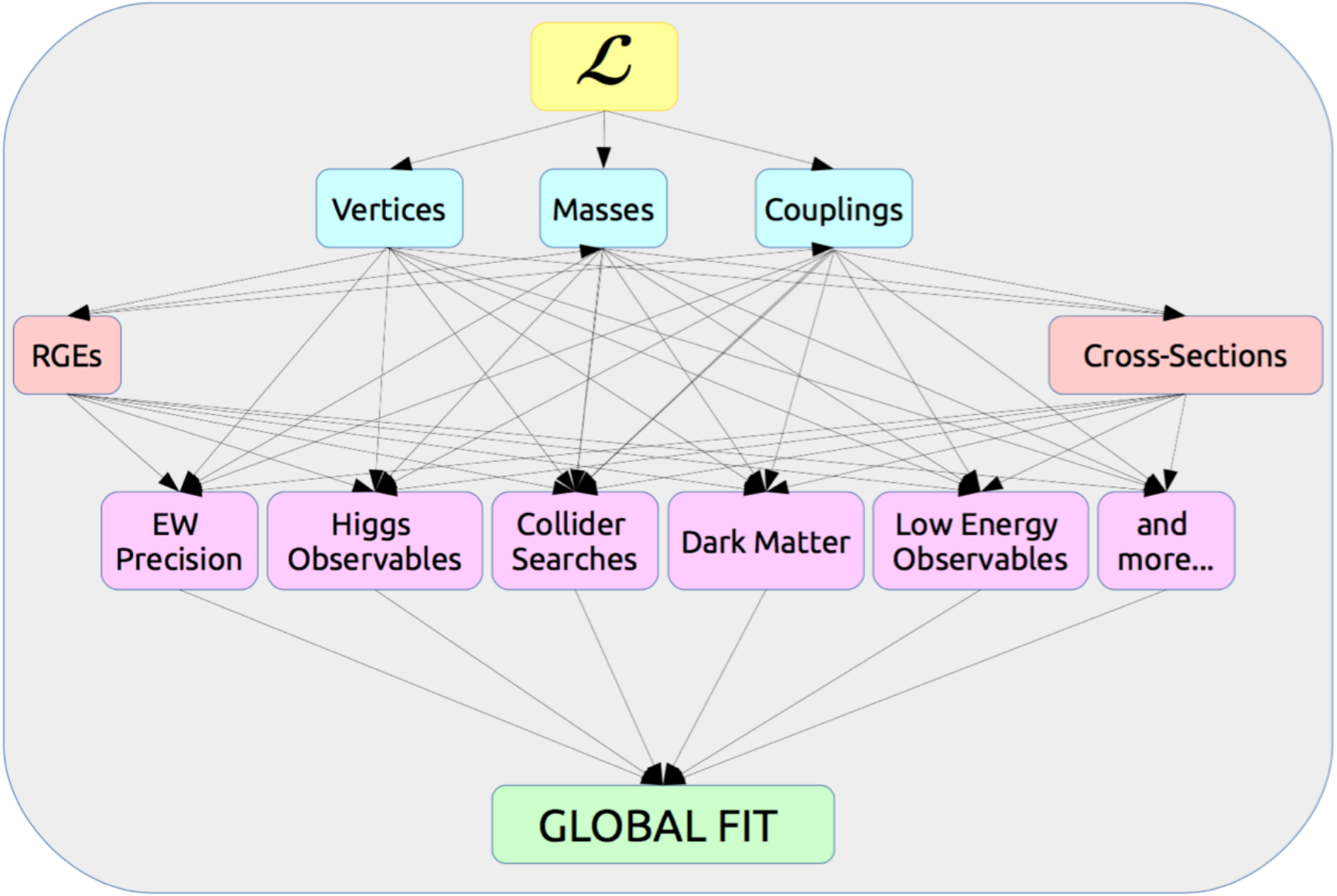
"the width of the bins is designed to correspond to twice the expected resolution ... and their origin is deliberately chosen to maximize the number of events found in any two consecutive bins"



Dijet mass sum in $e^+e^- \rightarrow jjjj$

ALEPH Collaboration, Z. Phys. C71, 179 (1996)

The CheckMATE Plan – limit input to Collider Searches for now

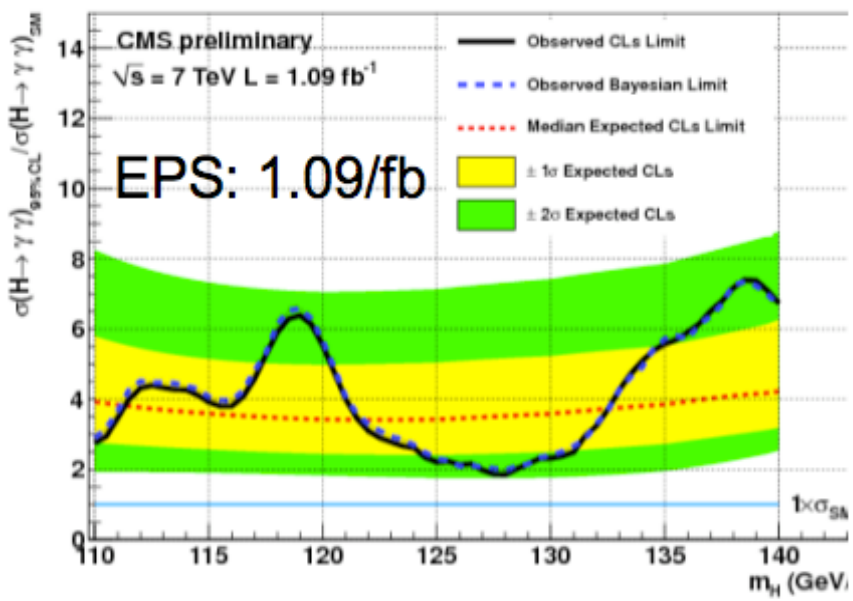


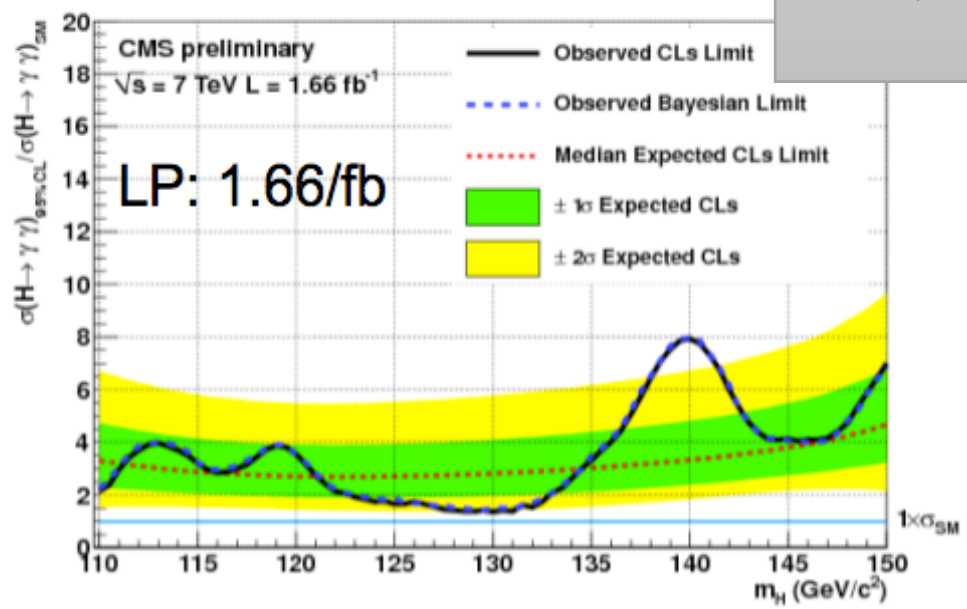
It wasn't always so clear...

Show ATLAS animation

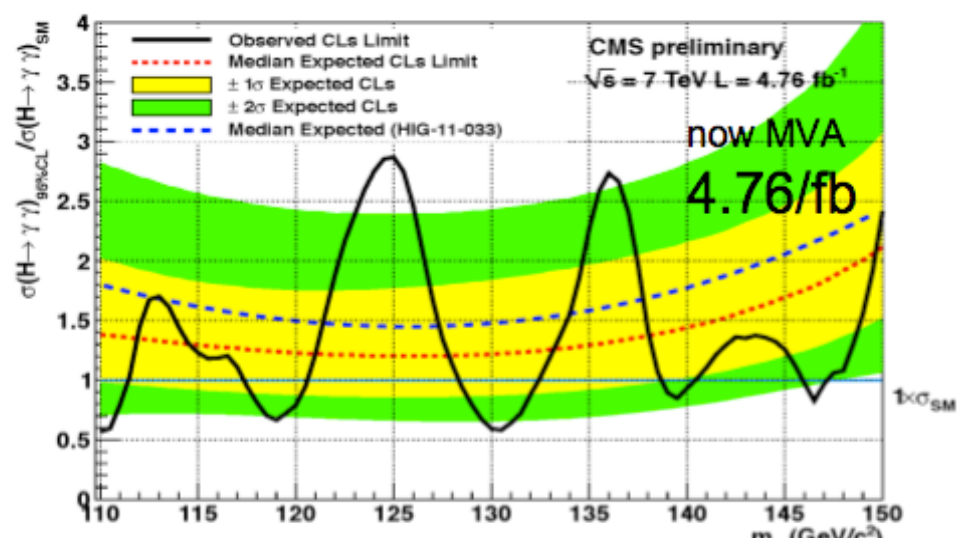
C. Paus, Implications Workshop, Mar. 27, 2012

CMS History: $H \rightarrow \gamma\gamma$





- EPS (1.09/fb) LP (1.66/fb) Dec 19 (4.76/fb)
- 'peaks' come and go
- we are getting into interesting territory, and peaks can also stay



Anastasiou, Dixon, Melnikov, and Petriello, **Phys.Rev. D69 (2004) 094008**

arXiv: hep-ph/0312266

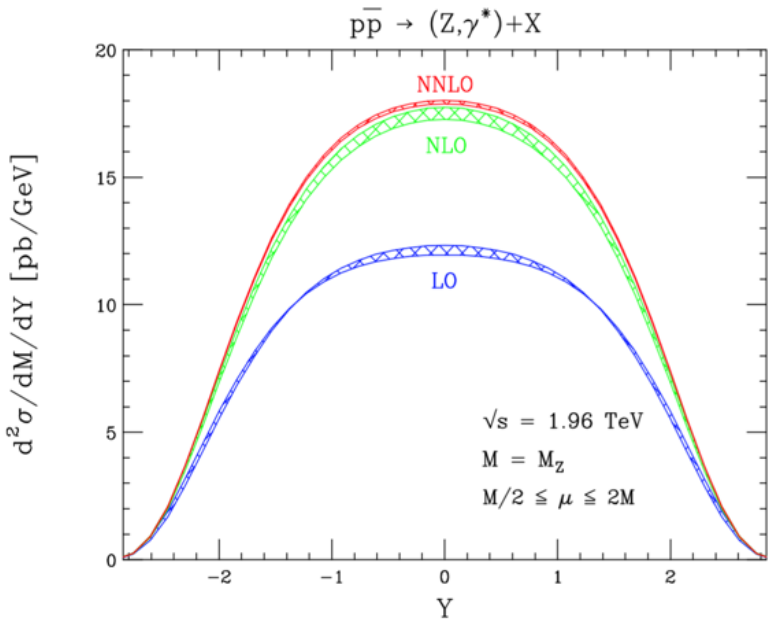


Figure 5: The CMS rapidity distribution of an on-shell Z boson at Run II of the Tevatron. The LO, NLO, and NNLO results have been included. The bands indicate the variation of the renormalization and factorization scales in the range $M_Z/2 \leq \mu \leq 2M_Z$.

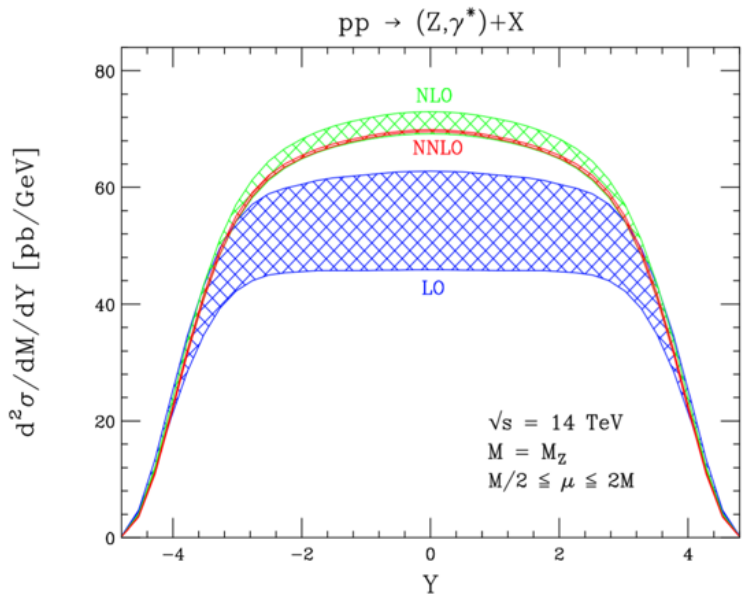


Figure 3: The CMS rapidity distribution of an on-shell Z boson at the LHC. The LO, NLO, and NNLO results have been included. The bands indicate the variation of the renormalization and factorization scales in the range $M_Z/2 \leq \mu \leq 2M_Z$.

Optimizing Histogram Binning

Two competing effects:

1) Separation of events into classes with different s/b improves the sensitivity of a search or a measurement. Adding events in categories with low s/b to events in categories with higher s/b dilutes information and reduces sensitivity.

→ Pushes towards more bins

2) Insufficient Monte Carlo can cause some bins to be empty, or nearly so. This only has to be true for one high-weight contribution.

Need reliable predictions of signals and backgrounds in each bin

→ Pushes towards fewer bins

Note: It doesn't matter that there are bins with zero data events – there's always a Poisson probability for observing zero.

The problem is inadequate prediction. Zero background expectation and nonzero signal expectation is a discovery!

Overbinning = Overlearning

A Common pitfall – Choosing selection criteria after seeing the data.
"Drawing small boxes around individual data events"

The same thing can happen with Monte Carlo Predictions –

Limiting case – each event in signal and background MC gets its own bin.
→Fake Perfect separation of signal and background!.

Statistical tools shouldn't give a different answer if bins are shuffled/sorted.

Try sorting by s/b. And collect bins with similar s/b together. Can get arbitrarily good performance from an analysis just by overbinning it.

Note: Empty data bins are okay – just empty prediction is a problem. It is our job however to properly assign s/b to data events that we did get (and all possible ones).