Lessons from the Tevatron

Gustaaf Brooijmans

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

YETI 09

Outline

- Trigger
- Ramp-up
  - Not transposable, but lessons nevertheless
- Data Quality
- Data Analysis
  - Logistics
  - Techniques
- Nevertheless

Tevatron

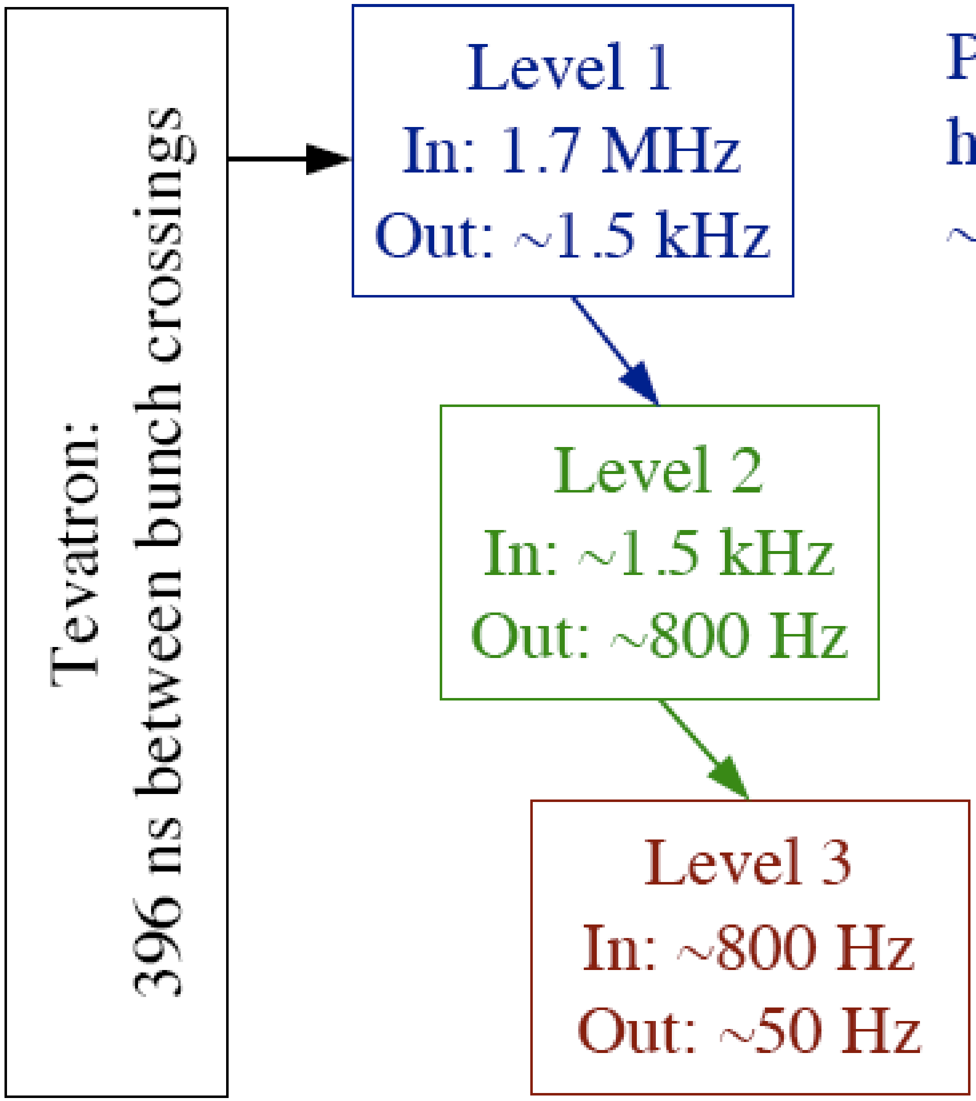
- proton-antiproton (quantum numbers of the vacuum)
- ~2 TeV center-of-mass
- Particle bunches cross at 2.5 (1.7) MHz, 3-10 interactions/crossing now
- Over 5 fb^{-1} recorded by each experiment

LHC

- proton-proton (well, gluon-gluon mainly)
- 14 (?) TeV
- Bunch crossings at 40 MHz, ~20 interactions/crossing
- ~30 times Tevatron luminosity
- Start soon

Trigger

The DØ Trigger System



Pipelined (but not deadtimeless), hardware only, coarse readout, ~4.5 μs latency

Hardware/Software mix, L1 inputs, ~100 μs latency

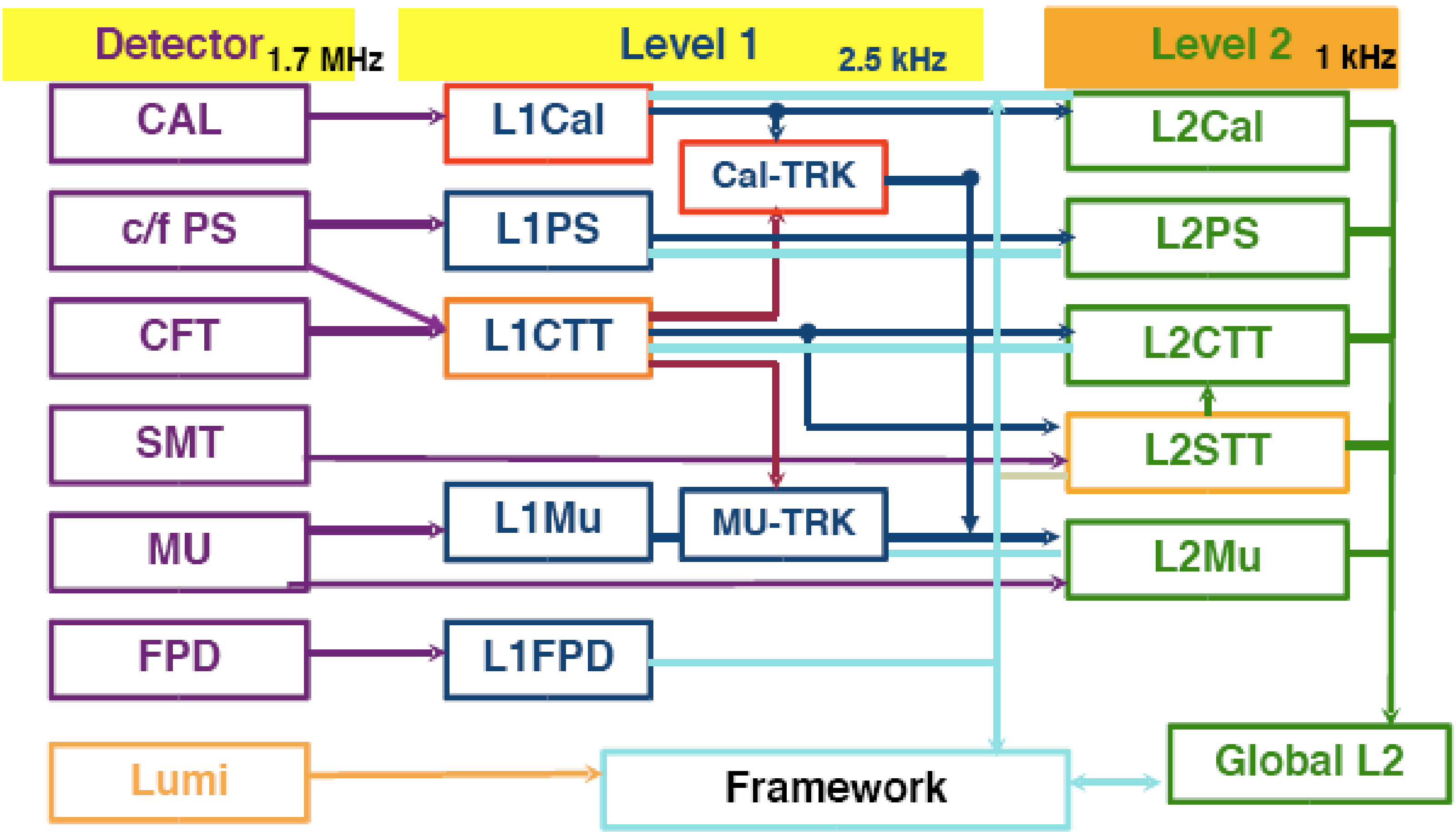
CPU farm, access to full event information, ~200 ms/event

50 Hz = few million events/day

Modulo rates, similar to LHC!

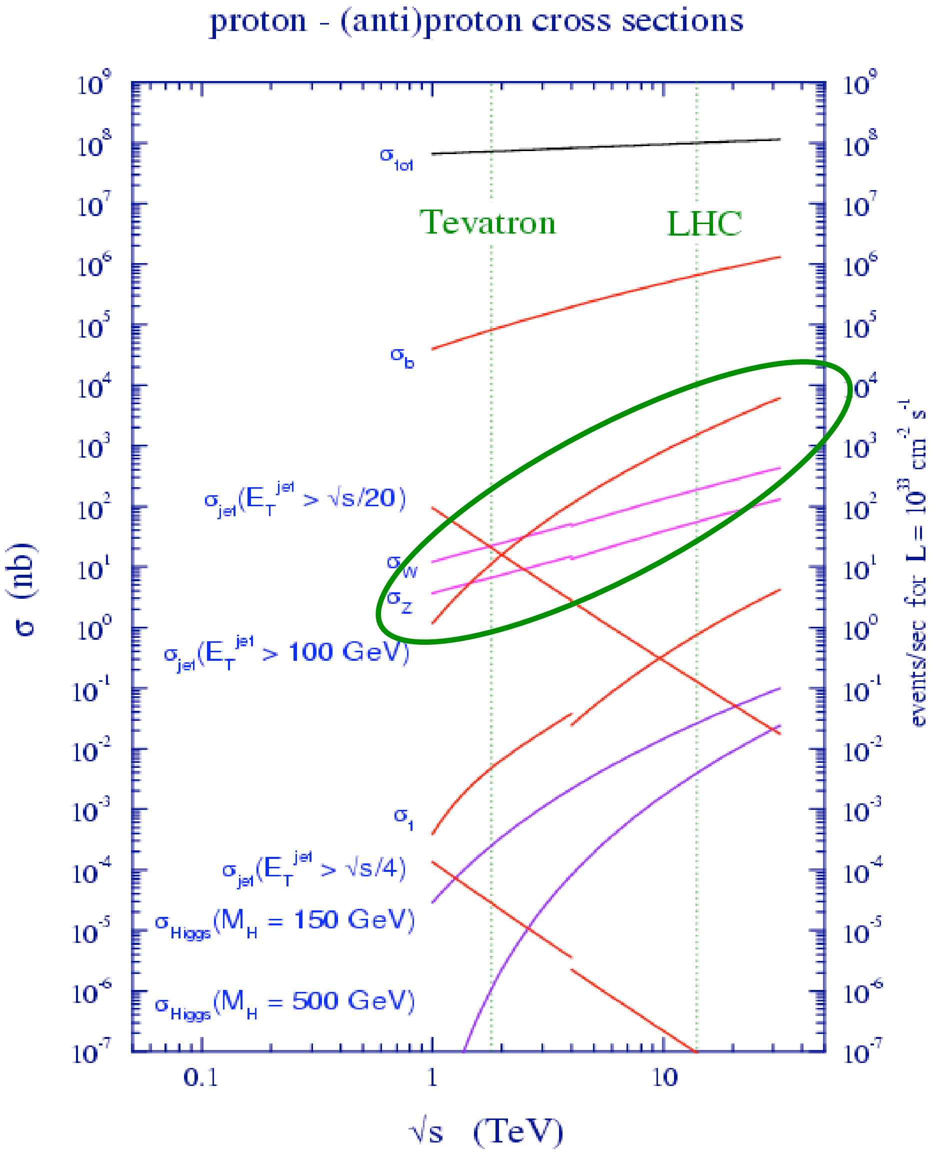
Technical Differences

- Tevatron experiments have lower L1 bandwidth
  - Requires more complex L1 systems (better rejection)
  - But LHC detectors have better resolution



Tevatron vs LHC

- LHC input rate ~20x larger
  - But “interesting” cross-sections ~10x larger
  - And output rate only ~4x larger
  - At design luminosity, have to prescale leptonic W…
- We will need to do better at the trigger level
  - (The Tevatron is a Mega-W machine…)



Physics and the Trigger

- "Basic" physics analysis:
  - Select "loose" and "tight" samples
  - Use loose to help determine tight sample composition
  - Use various distributions in tight sample to search for signal/measure properties
- 200 Hz/40 MHz = 5×10^{-6} → trigger rejects 99.9995% of the events
  - The first 99.9995% of physics analysis, i.e. the first "loose selection" is done in the trigger
    - ➡ Multiple preselection stages with different resolutions lead to complicated biases

Trigger List/Menu

- Maximize physics yield within allowed rates
- Combination of many (i.e. hundreds) of different "triggers" (in ATLAS these are called "trigger chains")
  - E.g. in DØ have O(100) single electron triggers with different, somewhat correlated requirements, combination optimizes signal yield/rate-to-tape
  - Dependent on instantaneous luminosity delivered by accelerator (controlled on fine scales using prescales)
  - Complex interplay between triggers: "overlaps"
- ~6 months to develop a new menu for higher luminosities

Trigger List Development

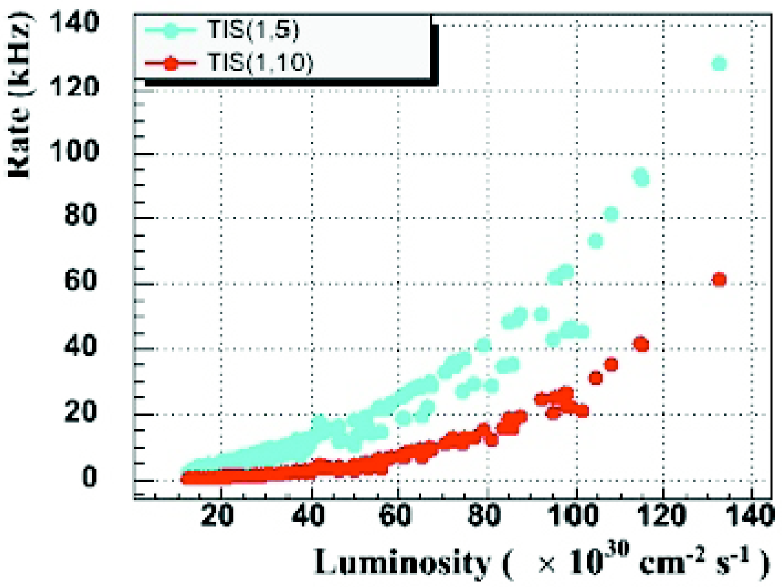
- Complex:
  - Optimize efficiency within a certain rate budget
    - Implies being able to estimate rates
  - Many signatures in multi-purpose experiments
    - How do you "prioritize" physics?
  - Enormous flexibility, especially at higher levels
- Currently:
  - DØ has about 600 triggers, including monitoring
  - CDF uses about 180

Rate Estimates

- Rates are very sensitive to events that are not normally recorded
  - Ideally, would like ~10 seconds of unbiased accelerator data
    - Not practical: at LHC: 40 MHz x 10s / 200 Hz = 2 10^6s, or 1-2 months of exclusive data taking
  - Take "enhanced bias" data: use lowest thresholds for each of the Level 1 objects, apply prescales at EF (but still run algorithms in pass-through mode)
    - Still need a lot of bandwidth, but no need to reconstruct
      - Could try to take in parallel with normal data but....

Rate Projections

- Can't take all data at low luminosity, when bandwidth available
    - Even at high lumi, typically designing menu for even higher luminosities
- Many trigger objects have non-linear rates due to increased occupancy. Two options:
    - Fit the rate vs lumi curve
        - Done online at CDF
    - Re-weight events as a function of the number of primary vertices



Physics: Efficiency Estimates

- Trigger objects from simulation useful tool for initial efficiency estimates
  - MC usually does a fair job at reproducing p_T distribution of signals
    - Ok, maybe not for jets in W/Z+jets, but the jets shouldn't be crucial in your trigger strategy there
  - OTOH, MC is usually not so good at reproducing variables that depend on occupancy, like isolation, "hadronic veto", missing E_T
    - Often, these involve the absence of signal
- Trigger simulation also needed for verification!

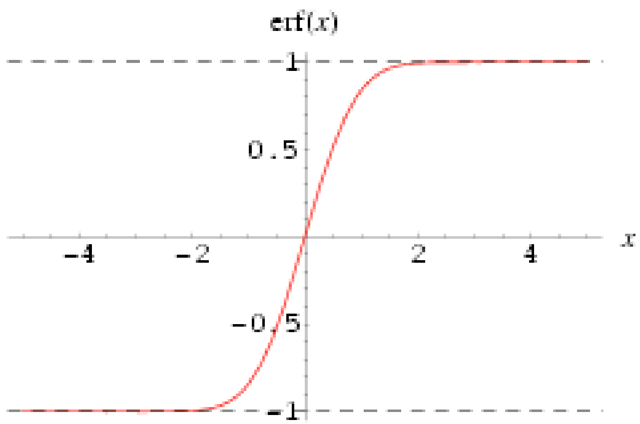
Efficiency Measurement

- Trigger efficiency w.r.t. what?
  - Absolute? Not necessarily useful, can be difficult
  - Offline reconstruction? Yes, but can be a moving target
- Really need to determine trigger inefficiency
  - I.e., which events *didn't* you get?
    - Monitoring triggers have lower thresholds/are less tight, but heavily correlated (same object → same acceptance, etc.)
    - Orthogonal triggers: exploit diverse trigger menu, e.g. use events that passed muon triggers for jet efficiency
      - Logistics! Now need to look at muon-triggered events!

Functional Form

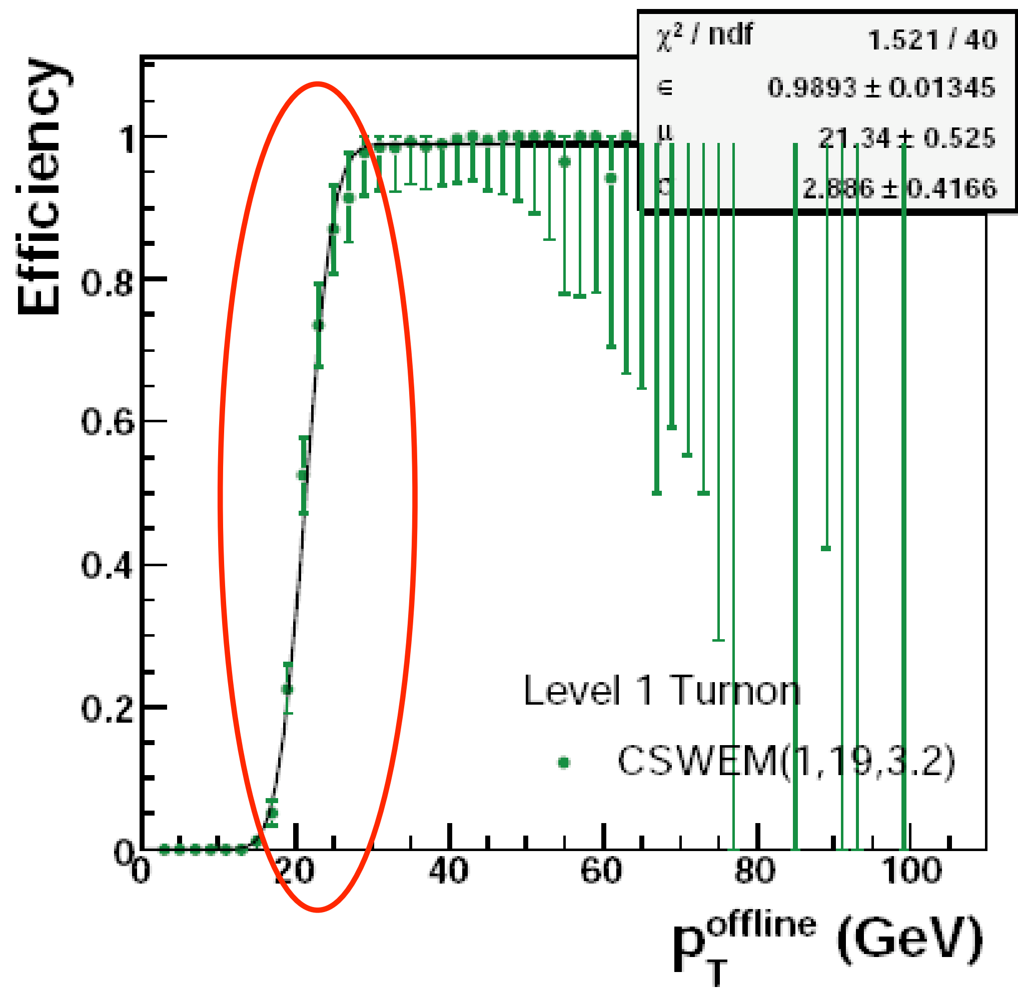
- At perfect resolution, trigger efficiency vs a certain parameter is a step function
- But detectors aren't perfect, so step is convoluted with - usually - a gaussian
- Integral of gaussian is the error function:

$$\mathrm{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt.$$



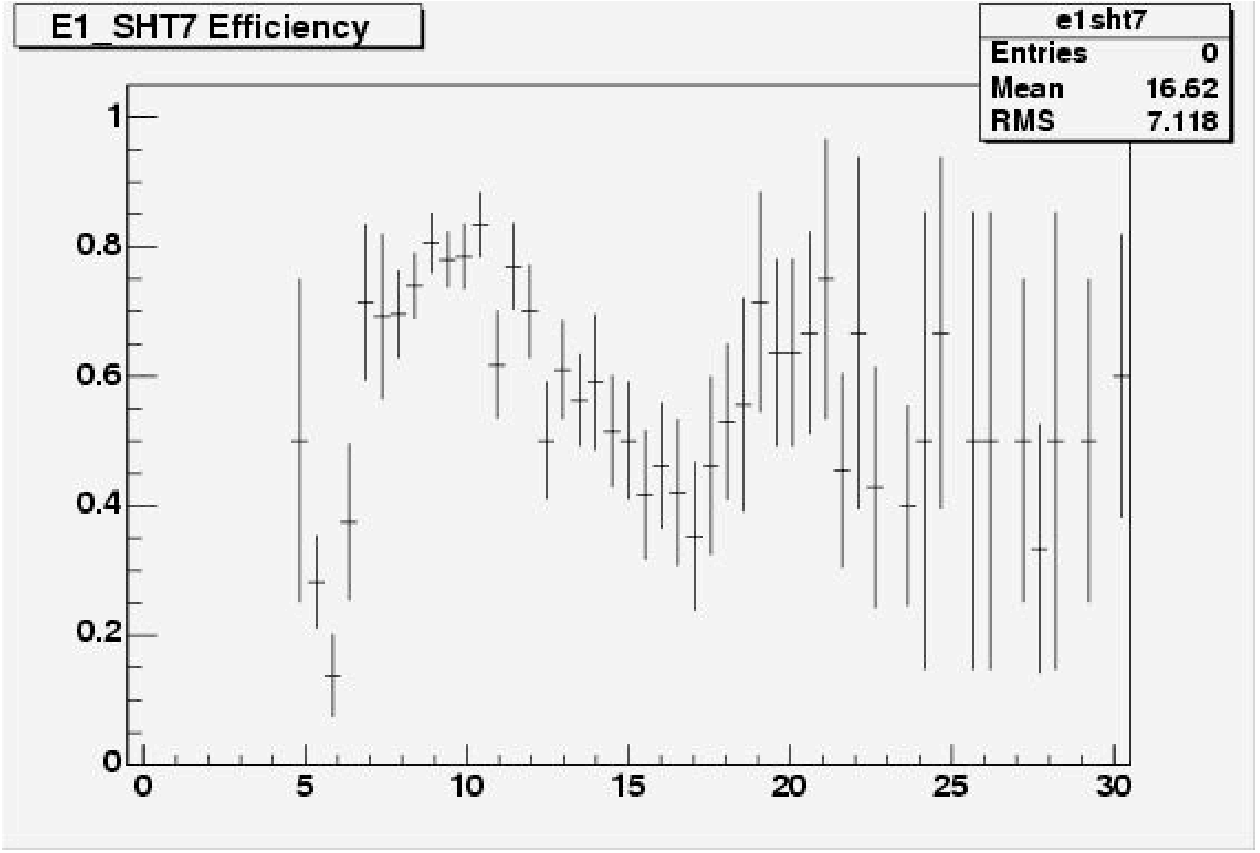
In Practice

- Most used is efficiency vs p_T
  - Plot is "turn-on" curve
  - Turn-on "point" is where efficiency reaches plateau (or sometimes midpoint)
  - Many analyses use only data in plateau region (severe systematics below)
  - To get rate, convolute with exponentially dropping QCD spectrum
    - Most events are at low end!



Example: Electrons

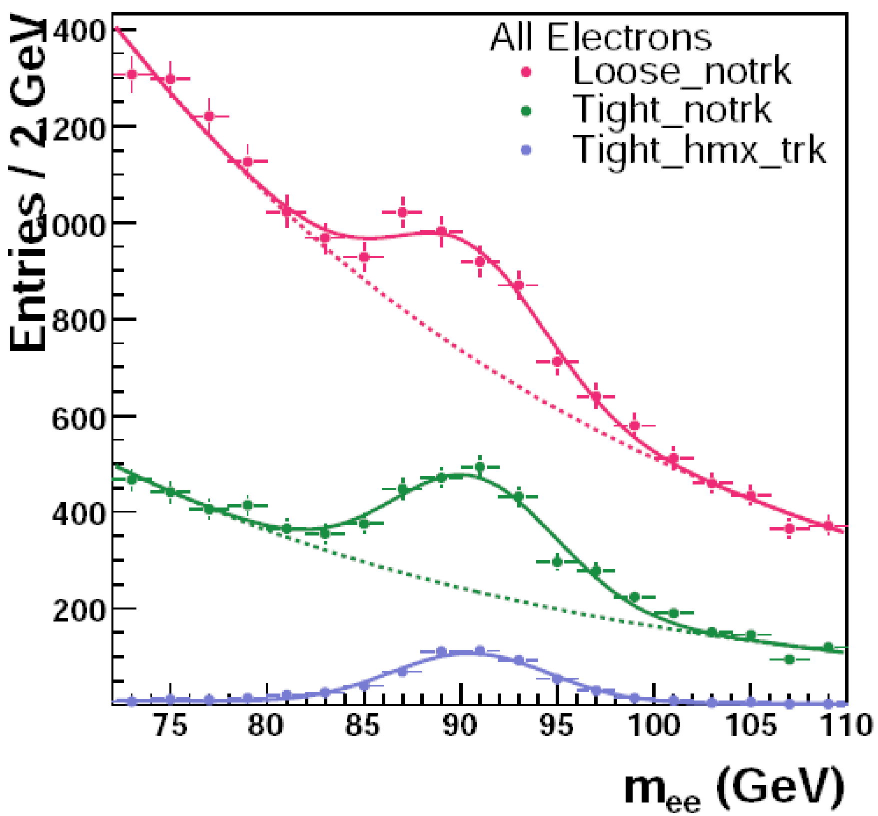
- Interested in efficiency for electrons
  - But most medium p^T objects that satisfy good calorimetric criteria (EM fraction, isolation, shower shape) are jets

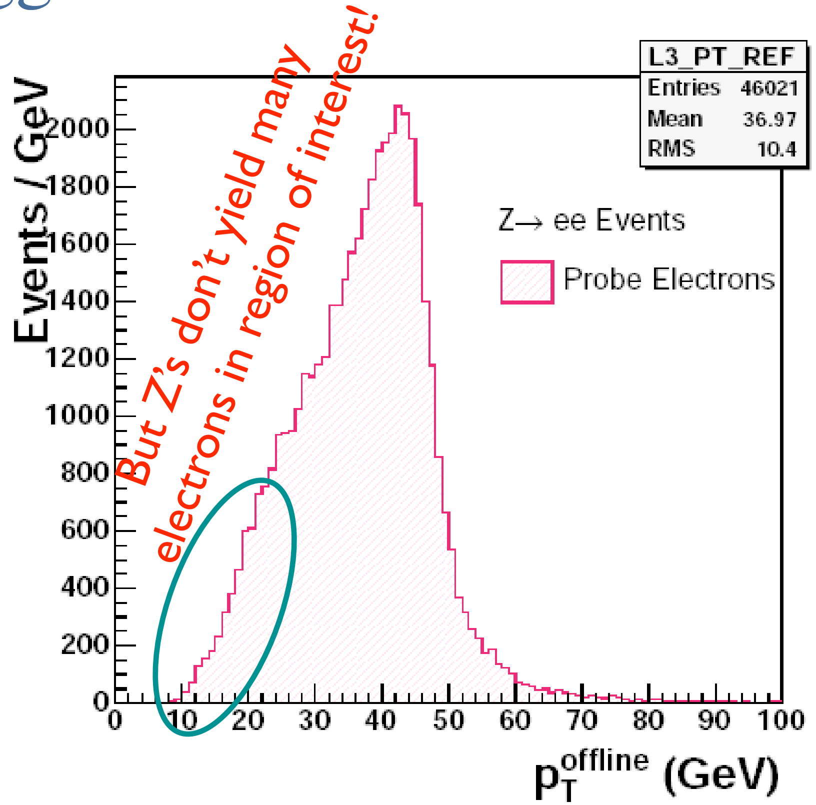


In this case, no track requirement made, which would help, but purity would still be a problem

Tag-and-Probe

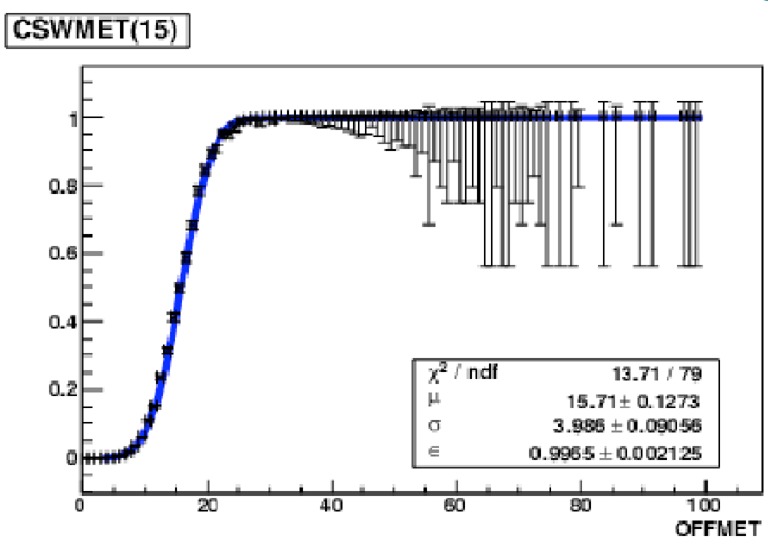
- Z's are a good source of **two** real electrons
  - Select Z's from events with two good electrons with Z mass
  - Tag = matched to electron from single electron trigger, check how often second electron fires trigger as well



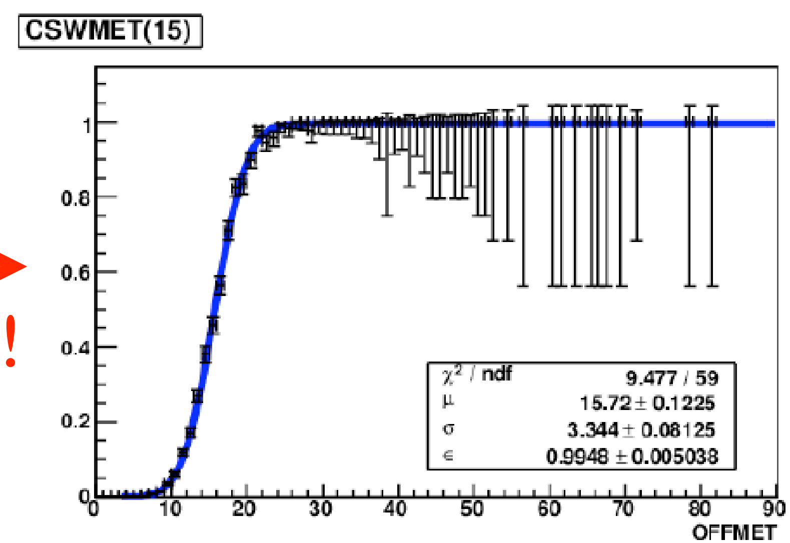


Impact of Calibration

- Calibration sharpens turn-on curve
  - Substantially reduces “garbage” events
- But... rates can change substantially
  - After all, bulk of events are at low end
  - Depending on main source of events, rates can go up!
    - Then need to readjust thresholds → new trigger menu!



Rates +~20%!



Calibration @ Trigger

- Example: L1 calorimeter calibration constants. In principle, a simple problem:
  - Determine gains by comparing with offline
    - Of course, that requires offline to be "calibrated"
  - Determine pedestals from "noise runs":
    - So, pedestal = "number of ADC counts without signal"
    - What about pile-up? Underlying event? Offline, we correct for these
    - At the trigger level, in principle would prefer to factor pileup into pedestal... but then they depend on luminosity!

Ramp-Up

From the Tevatron to the LHC

- You will often hear
  - “It took CDF & DØ one year to go from first collisions to physics in Run II” (implying the situation will be similar for the LHC)
  - This is misleading
    - Run II was considered an “upgrade” (even though > 80% of electronics, both trackers, etc. were replaced)
    - First collisions happened in 2001
      - Main software development effort started in 1998
      - When detectors rolled in (4/2001), large fraction of readout electronics was missing

- However, we did learn things that will likely impact LHC experiments as well:
  - The trigger is the one system where individual subdetectors can/will have a large impact on each other
    - Pathological behavior that doesn't affect one system *will* bring down another
    - You cannot use teststands and testbeams to come close to emulating the real system
  - The trigger is the nervous system of the experiment: it's very complicated, relatively fragile, and bad behavior can be very debilitating. It's also often how you discover problems

Lessons

- You can’t fully debug the trigger and readout until the downstream system can take the full rate
  - No matter how sophisticated the simulated triggers in the lab, the real system will find a pattern that leads to problems (race conditions, bad buffer management,...)
  - Corollary: if you increase the rates in steps, you need to verify the data integrity at each step (in addition to finding and fixing crashes/hangs,...)
    - E.g. are all parts of the calorimeter reading out the same event?
  - Increasing the rates in steps is, of course, a typical commissioning strategy

- You can never have enough diagnostic tools
  - But they need to be easy to adapt (i.e. clear code) so you can react to a new situation
  - Very important
    - Online: data flow GUIs
    - Offline and online: Capability of examine data and status registers (in hex) at all stages (i.e. the software to do it, AND the expertise)
- Things that are “impossible” happen regularly
  - “My hardware can’t do that!” ... 2 days pass ... “If you do this simultaneously with that, then... but I didn’t think it could be done”

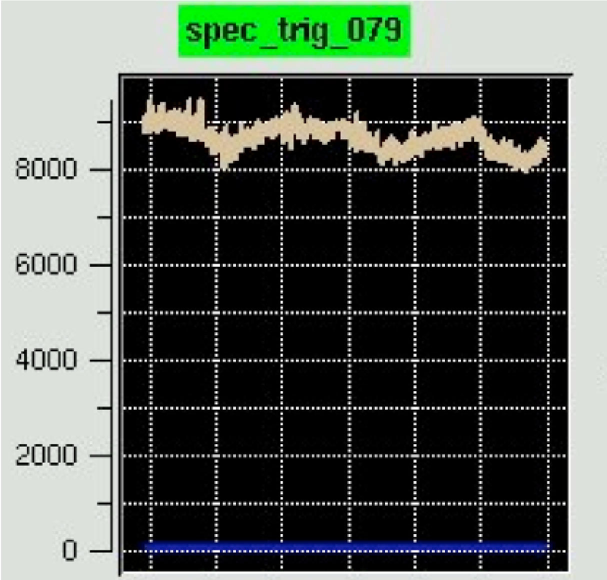
- Expertise, expertise, expertise
  - Experiments rely on dedicated individuals, many “very young” (students & postdocs), who move on to physics analysis at some point
  - Logbooks are not an appropriate repository of expertise
  - Difficult problem, approaches through institutional commitments and similar techniques
  - Becoming an expert at something rather technical is an important part of every (experimental) physicist’s development
    - It will be recognized, and it will be an asset in landing your next job!

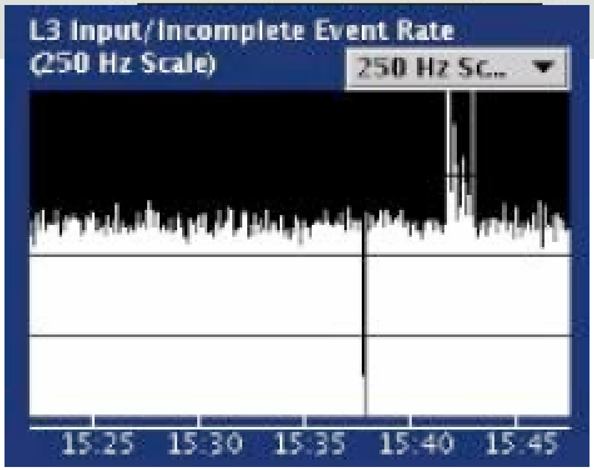
Data Quality

Online Data Quality

- Due to the complexity of the detectors, it is remarkably easy to take *bad* data
- Basic data quality monitoring:
  - Event display (don't underestimate!)
  - Occupancy plots
    - Geometrical, but also timing, number of trigger objects, etc
  - Full reconstruction
    - Can only be done for a small sample (+ with larger delay "express stream")
  - Smarter: calorimeter occupancy for events with large MET, ..

- Trigger is the first line of defense (system usually just breaks down because of corrupt data, hot cells, ...)
  - But: system is also built to throttle itself ("busy")...
- Some problems subtle
  - Rate oscillations
  - Transient effects
  - Bunch-crossing dependence
  - ➡ Maximize # of x-checks, monitors

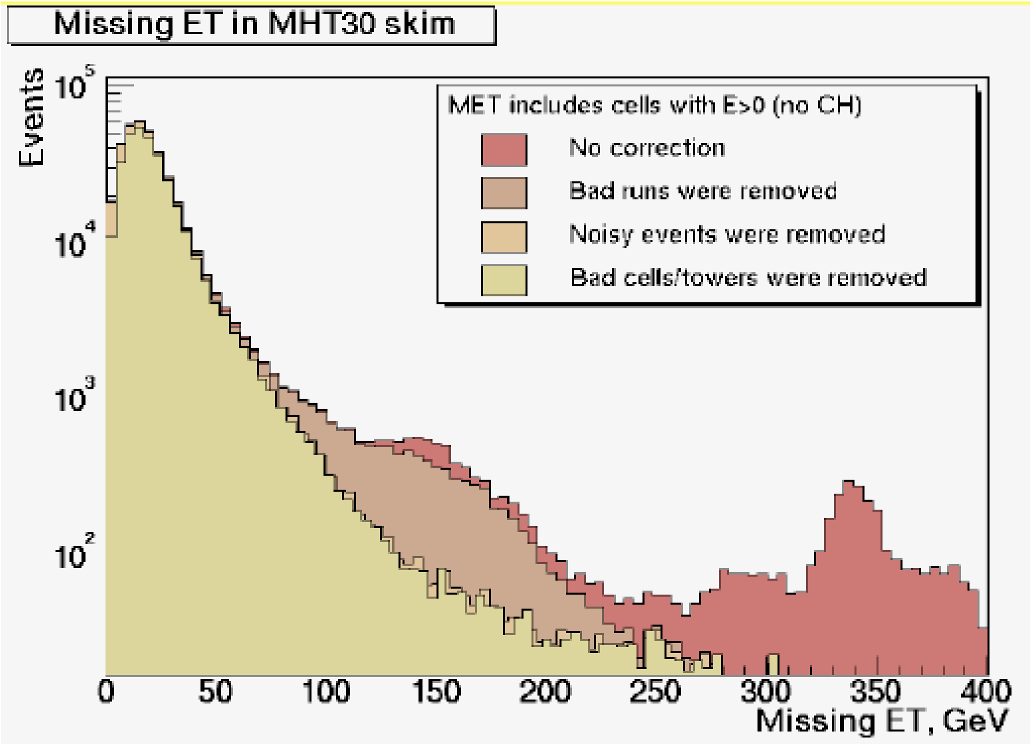




Offline

- Anything that happens at < 1% rate is almost impossible to detect online
  - As long as you don't know what to look for
    - Keep track of TGV schedules, TV programs, multitude of cron jobs, people welding in 500 m radius, ...
  - Continuous feedback from analysis is a necessity
    - Really subtle stuff can take years to find
- Doesn't mean you can't take good data starting on day 2
  - But detailed understanding takes lots of time & effort

- Analyses using MET are particularly sensitive
  - Requires the full calorimeter to behave, and calorimeter is generally the most sensitive subdetector (analog, ~16 bits)
  - Easy: basic DQ (missing board, etc.)
  - Hard: low frequency
  - Can't spot a 10^{-5} Hz (once a day) effect online or in first pass DQ
    - But can be biggest part of dataset after cuts!
    - Everytime dataset x5, find new source of rare noise...



DØ Calorimeter DQ

See L. Zivkovic talk on MET for ATLAS

- Online with alarms for shifters
- Offline: 3 steps
  - Per run (~4 hour block)
  - Per luminosity block (60 seconds)
  - Per event (known noise patterns)
    - Patterns have names: "noon noise", "ring of fire", "purple haze"
- "Spanish fan" was discovered in monojet analysis
  - First event was a lone event in the tail
  - Subsequent analysis of larger dataset showed a handful of events with the *same jet*.....

Data Analysis I
Logistics

Datasets

- Tevatron experiments now run at ~100 Hz with high efficiency
  - Around 10^9 events/year (a Tevatron year is longer than an LHC year)
    - 250 TB of raw data
  - LHC experiments will have similar event yields
    - 5x larger events (ATLAS) → > 1 pB of data/year
  - To this, you need to add MC (similar volume), calibration data, etc.
  - All of this in various formats (raw, cell-level reconstructed, analysis-level reconstructed)

Formats

- "Reco Full" contains the full info, i.e. individual cell energies, hits, etc. to allow re-running of high level reconstruction algorithms
- "Reco Ana" contains 4-vectors for jets, leptons, ...
- There is a strong tendency to migrate as much info as possible to "Reco Ana" as well
    - Which gets bigger and bigger
    - But this was found to be necessary at the Tevatron!

Streaming

- To reduce volume of data to be studied by an individual, data is "streamed" based on final state objects (e.g. 2+ muons, 2+ electrons/photons, etc.)
  - Streaming can be
    - Inclusive: events that satisfy multiple stream conditions are sent to each of those streams
    - Exclusive: events go to a single stream, according to preset priorities
  - Streaming can be done
    - Online: based on objects identified by the trigger (pre-reco)
    - Offline: based on reconstructed objects (post-reco)

- Option cons:
  - Inclusive: Need to set up streams to minimize overlaps; duplicate events!
  - Exclusive: Need to be very careful about priorities
  - Online:
    - Events that failed your trigger (crucial for trigger efficiency) are not in the same stream!
  - Offline:
    - When re-reconstructing, some objects may not pass object selection cuts anymore: events can migrate to a different stream
    - Can't preferentially (re)process a given stream

“Fixing”

- Complex detector → reconstruction program is a complex assembly of algorithms
  - Lots of debugging during development
  - Validation on small samples
- Nevertheless, some problems (from detector or reconstruction) are only found during detailed analysis
  - Can sometimes be “fixed” on analysis format
  - But need to make sure corrections are only applied once, and correctly → best done centrally

Skimming

- Physics analysis needs
  - "Loose" and "tight" samples:
    - Loose is used to x-check understanding of sample composition and backgrounds
    - Tight is the "physics sample"
- "Skim" = from the relevant stream, select subset satisfying "loose" constraints (and "fix" on the fly)
  - Tempting to make this very loose so it can be used by many
    - Many to find problems, common solutions, ...
  - But then it becomes very large ....

Choices

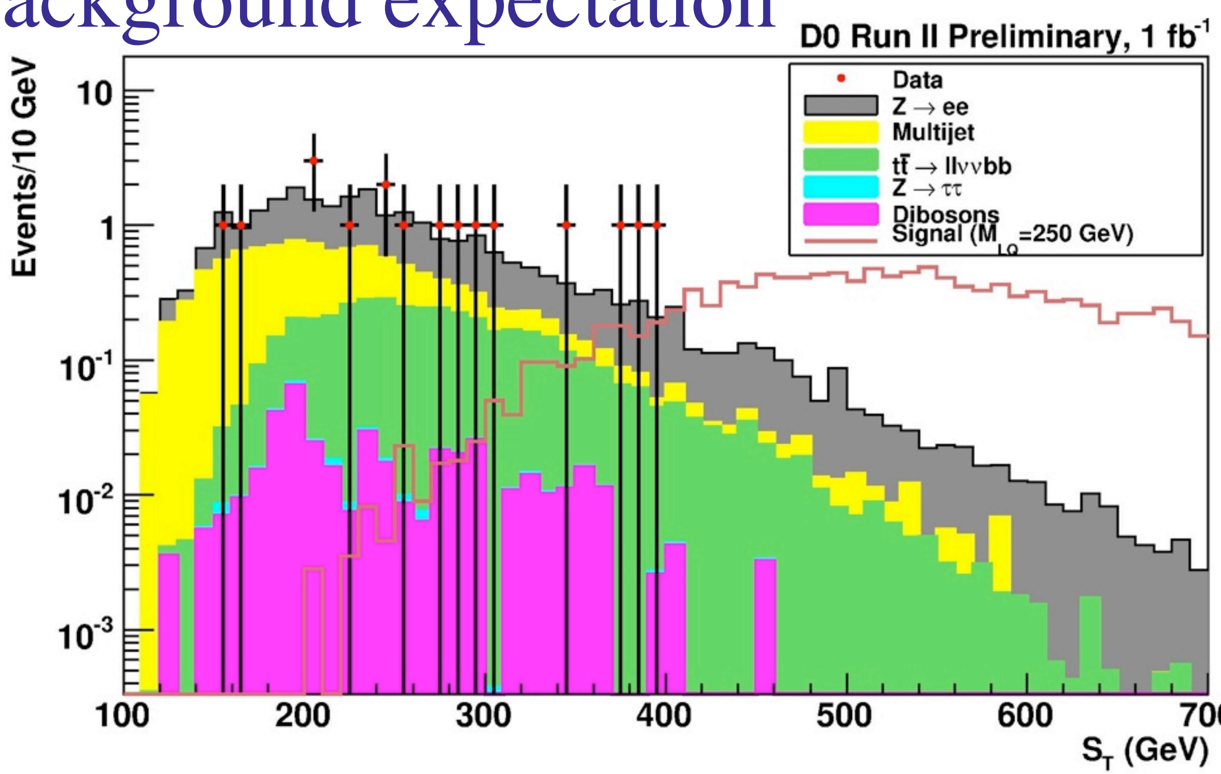
- At DØ:
  - Inclusive, offline streaming
  - Central skimming (includes fixing):
    - 2EMhighpt, 2MUhighpt, 3JET, EMinclusive, MUinclusive, EMMU, Higgs, JPSI, NP (MET), QCD, TAUTRIG, TOPJETTRIG, ZBMB
  - Analyzers apply next layer of skimming themselves
    - Resulting sample fits in << 1 TB, then ntuple fits on laptop
- At ATLAS:
  - Inclusive, online streaming
  - Skimming to address trigger efficiency sample

Data Analysis II
Techniques

Counting Experiments

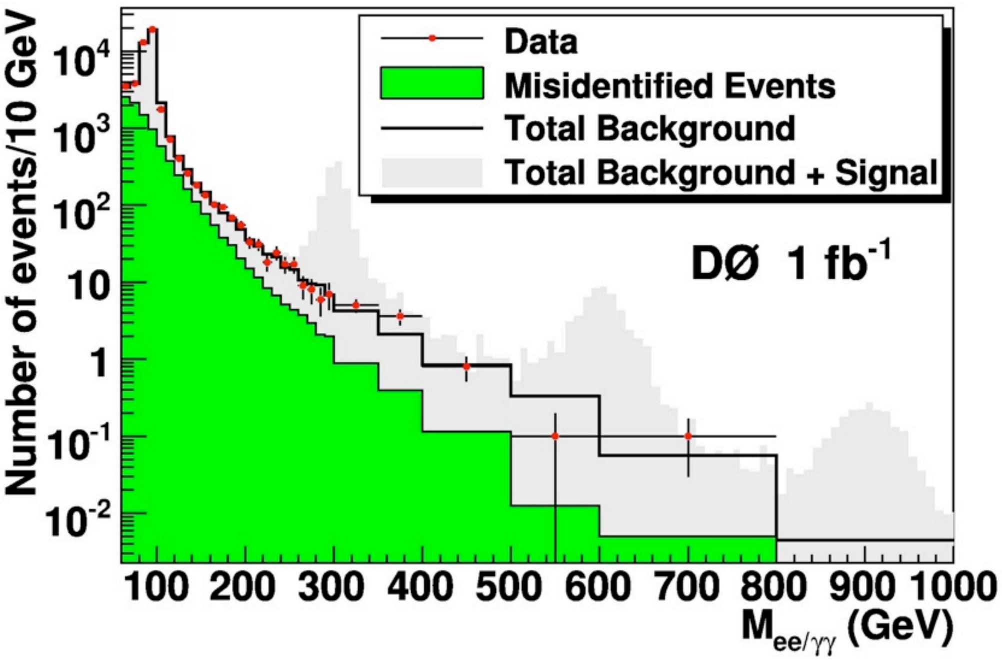
- Traditional analysis method:
  - Convince yourself that you understand the backgrounds
    - Loose sample etc.
- Count events that pass certain cuts, compare with (usually poisson) background expectation

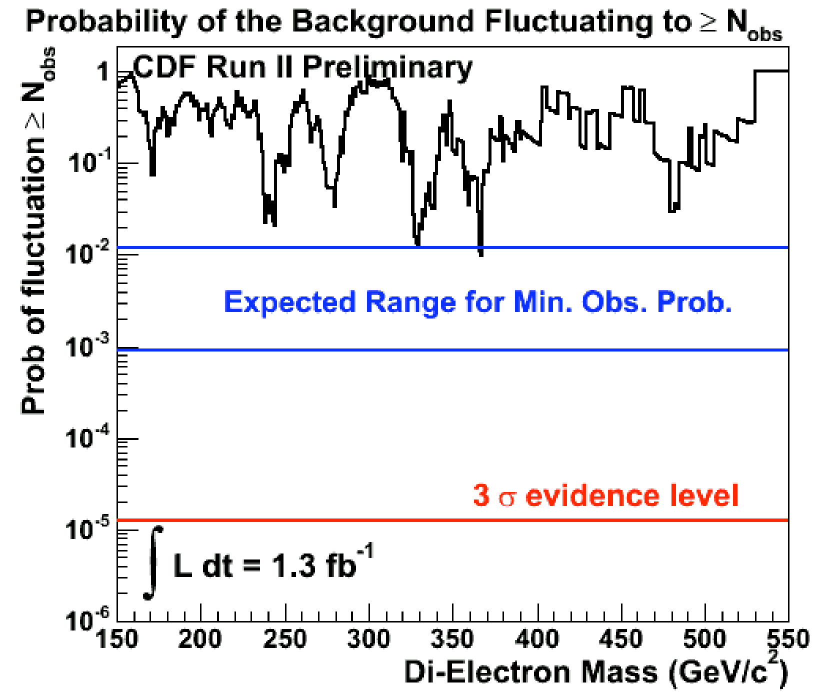
Search for LQs:
Require S_T > x GeV
so that background
expectation = 0 events



Counting II

- Alternatively, if you're looking for an excess populating a specific region, count there
  - E.g. for Z'→ee count events in a sliding mass window:



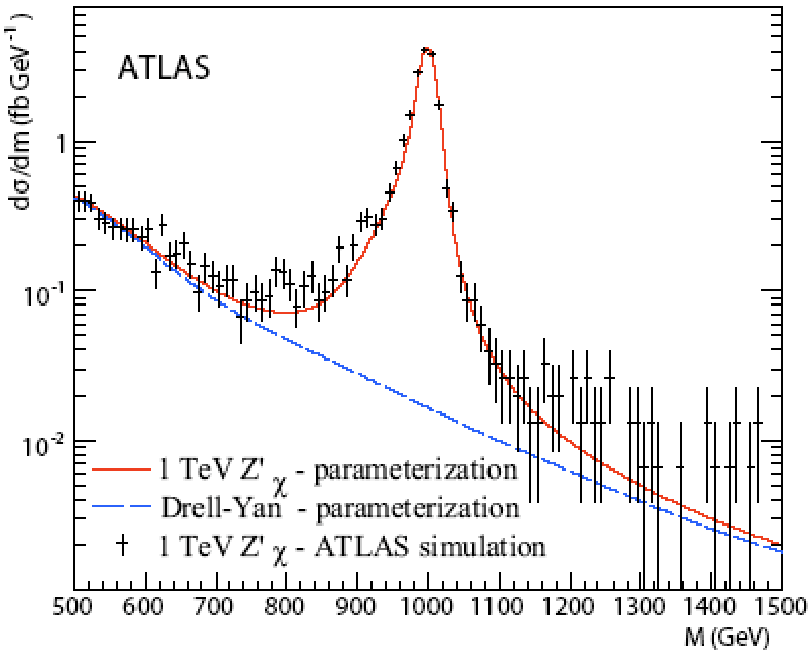


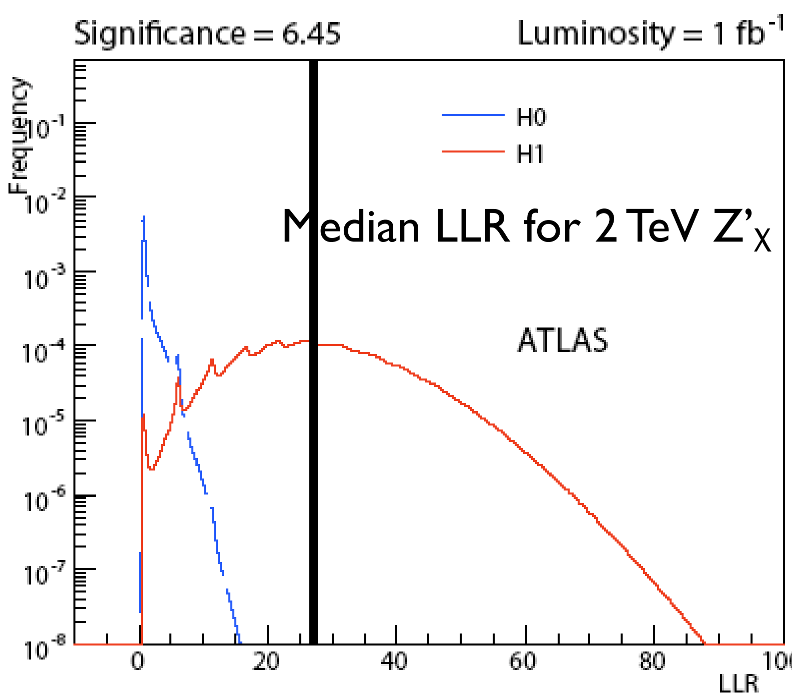
Beware of "Look Elsewhere" effect, since look in many mass windows!

Shapes!

- But a lot more information is available in shapes
  - E.g. a Z' is more than just an excess of events in a mass window - it has a Breit-Wigner (⊕ gaussian) shape!
  - Need parametrized signal & background shapes
  - Compute likelihood/χ^2 of data to B and S+B hypos

Only really useful if background is non-negligible!

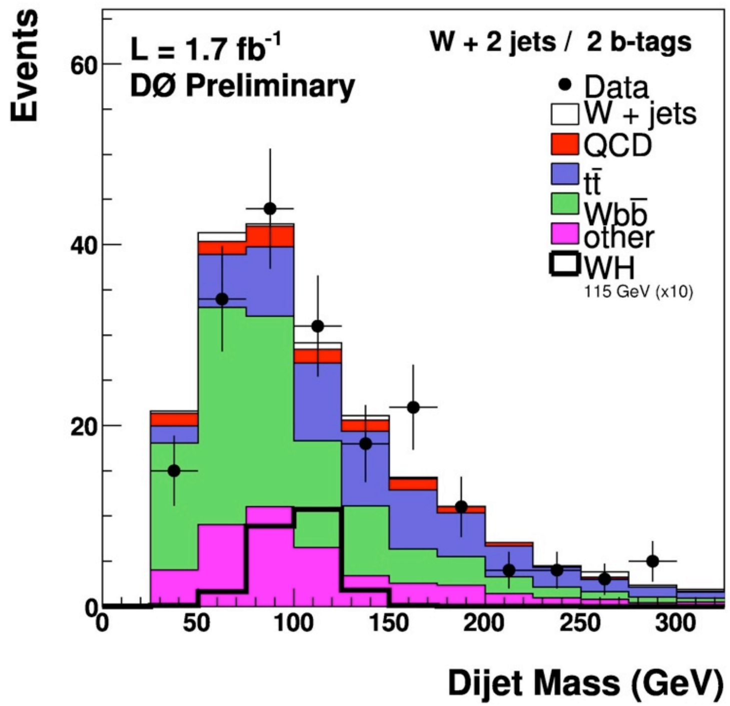


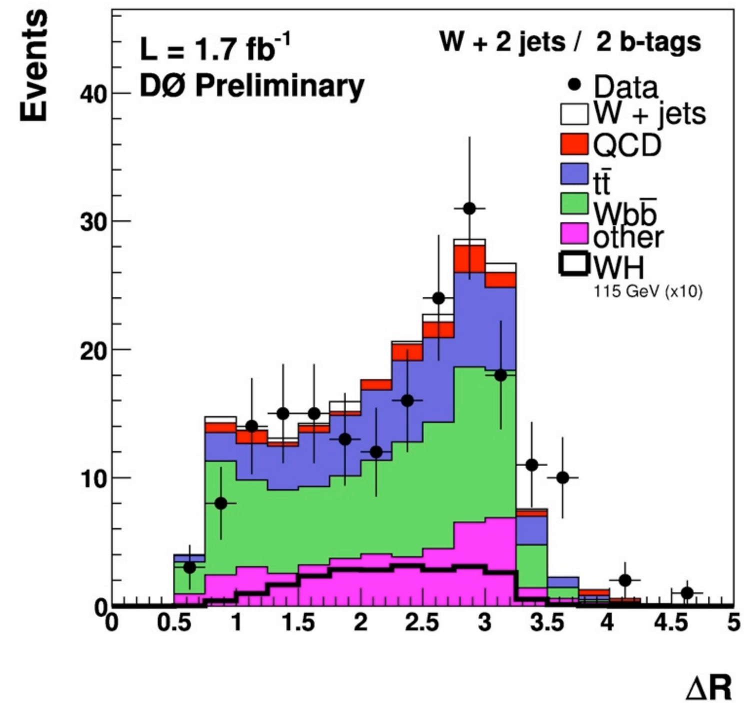


Multivariate Tools

- Most signals are not as clear as Z’
  - E.g. pp → WH → ℓνbb: signal is much much smaller than dominant Wbb background (and tt etc.)
  - Find a way to exploit (correlated) differences in many distributions

Nice overview in application to single top:
arXiv:/0803.0739 [hep-ex]
Published in Phys. Rev. D 78, 012005 (2008)





Matrix Element Analyses

- Currently yield the most precise measurement of the top quark mass, also
  - Major contribution to the evidence for single top
  - Big contribution in Higgs searches
- Basically unbinned maximum likelihood fits
  - Event-by-event measured uncertainties
    - More weight for more signal-like event
    - Determine event's "signal probability":

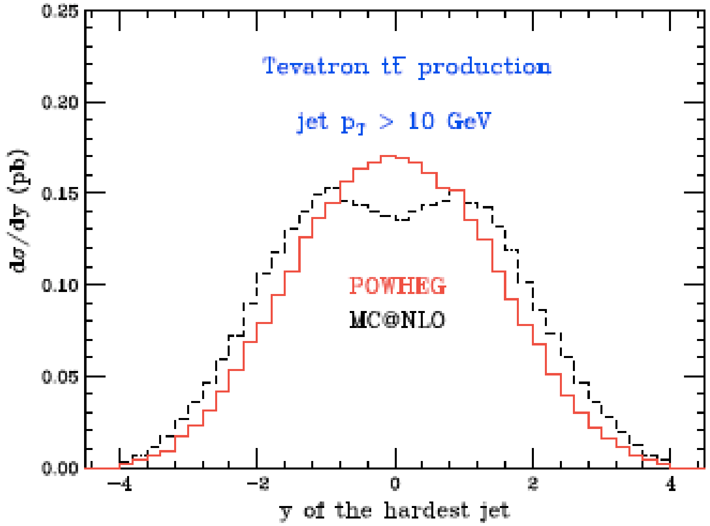
"Transfer functions": generated → measured momenta

b-tag prob

matrix element

$$\sum_{\text{perm}} w_i \int_{q_1,q_2,y} \sum_{\text{flavors}} dq_1 dq_2 f(q_1) f(q_2) \frac{(2\pi)^4 \left|\mathcal{M}(q\bar{q} \to t\bar{t} \to y)\right|^2}{2 q_1 q_2 s} d\Phi_6 W(x, y; JES)$$

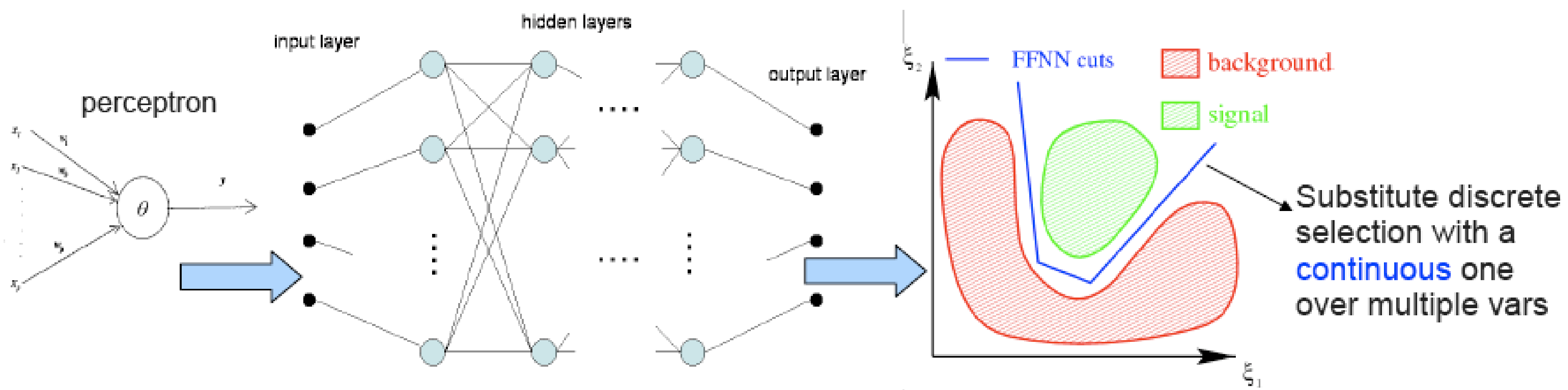
- Caveats:
  - LO matrix elements:
    - Require exact number of jets
    - Evaluation of NLO systematic not so easy
  - Recent development: replace madevent with MCFM
    - Done in Higgs searches, where likelihood output is injected in neural net
    - Increases Higgs sensitivity by ~1.3 (equiv to 1.7 x more data...)
  - Of course....



Very recently determined to be a problem with phase space coverage in Herwig (Nason)

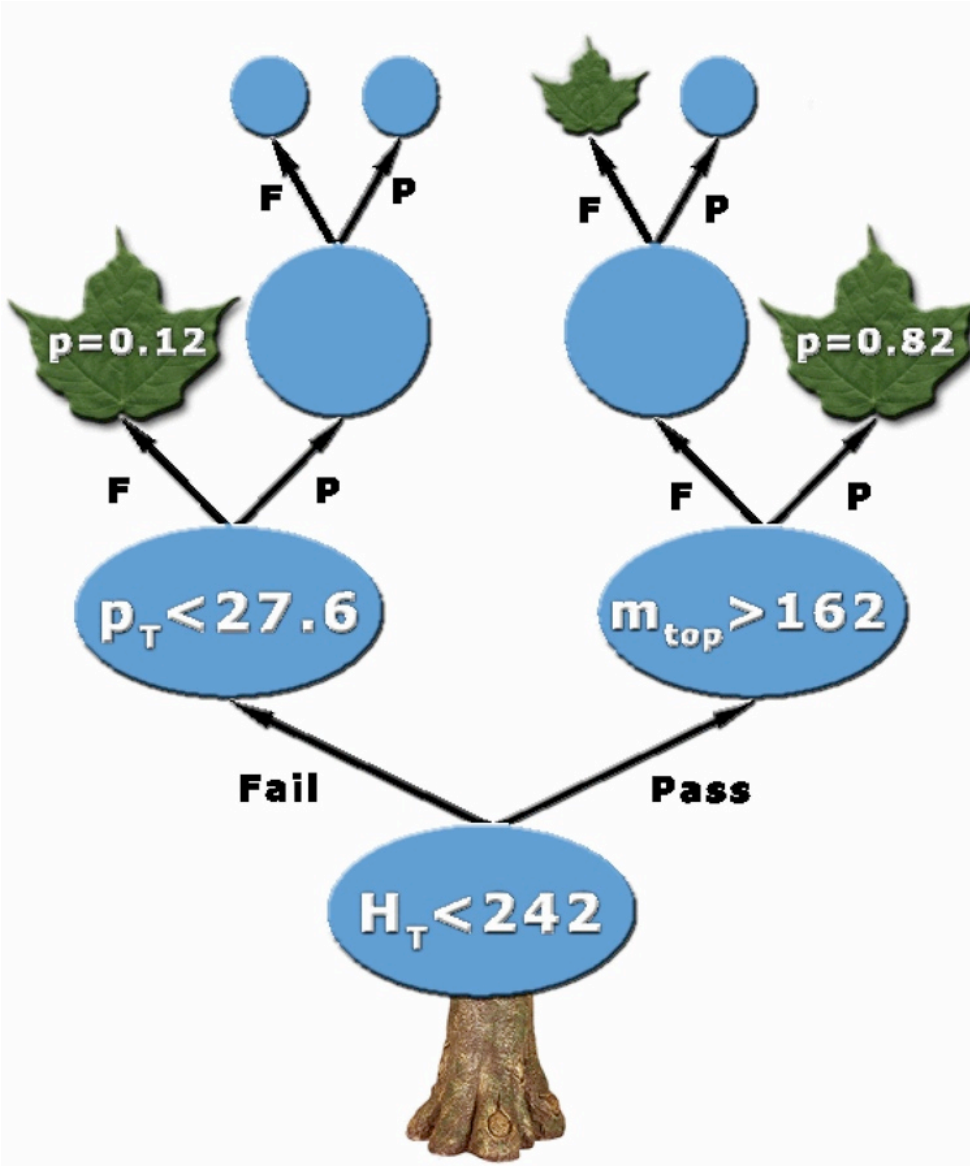
Neural Net

- So the matrix element approach is good, but not perfect
  - Squeeze out more sensitivity by using it as an input to a neural net, along with various kinematic distributions sensitive to the signal (yes there is some “double use”)
  - Neural net: computing system aimed at approximating a given mapping from a subset D of $\mathbf{R}^n$ (input variables) into **[0,1]** on the basis of known examples

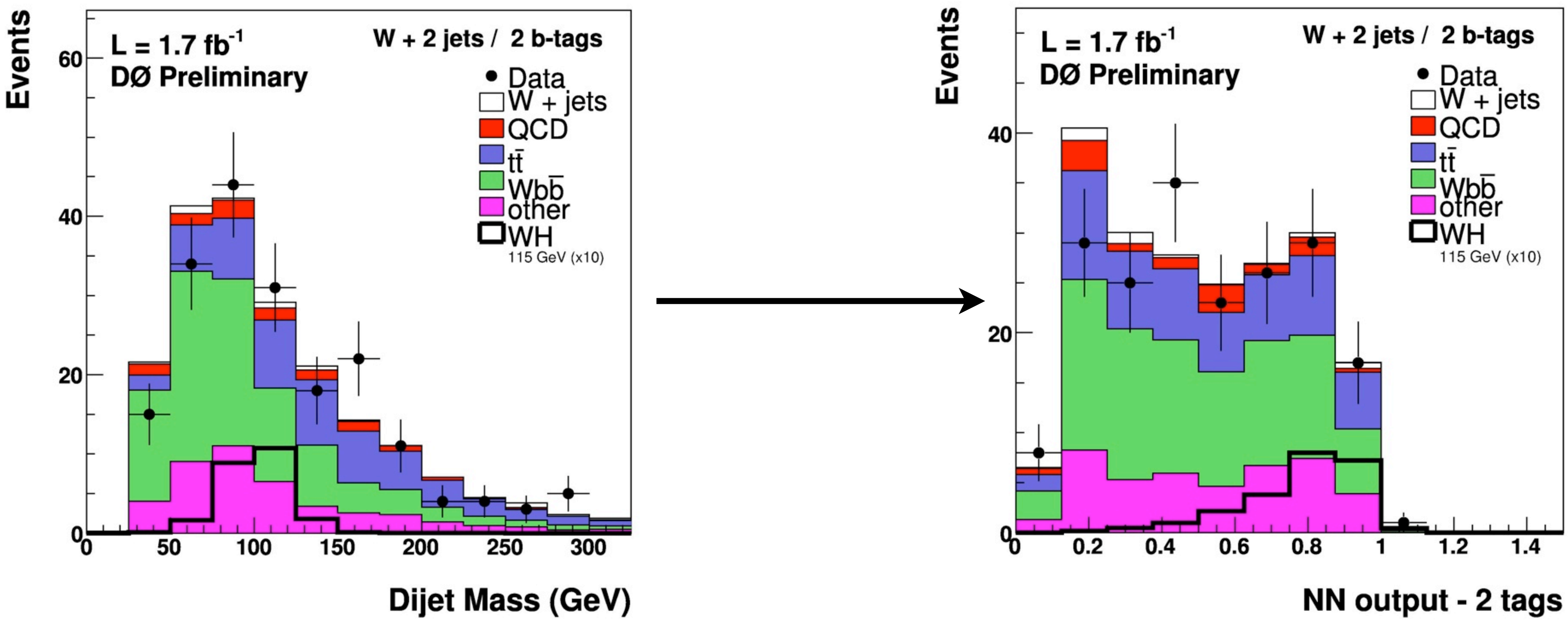


Boosted Decision Tree

- Sequence of “square cuts”
  - Events are not rejected
  - All events end up in leaves with certain signal purity
- Train to optimize cut values at each stage
- Boosting = look at misclassified events, give these extra weight and re-train
  - Force to work harder to separate signal-like events



Neural Net/BDT Output

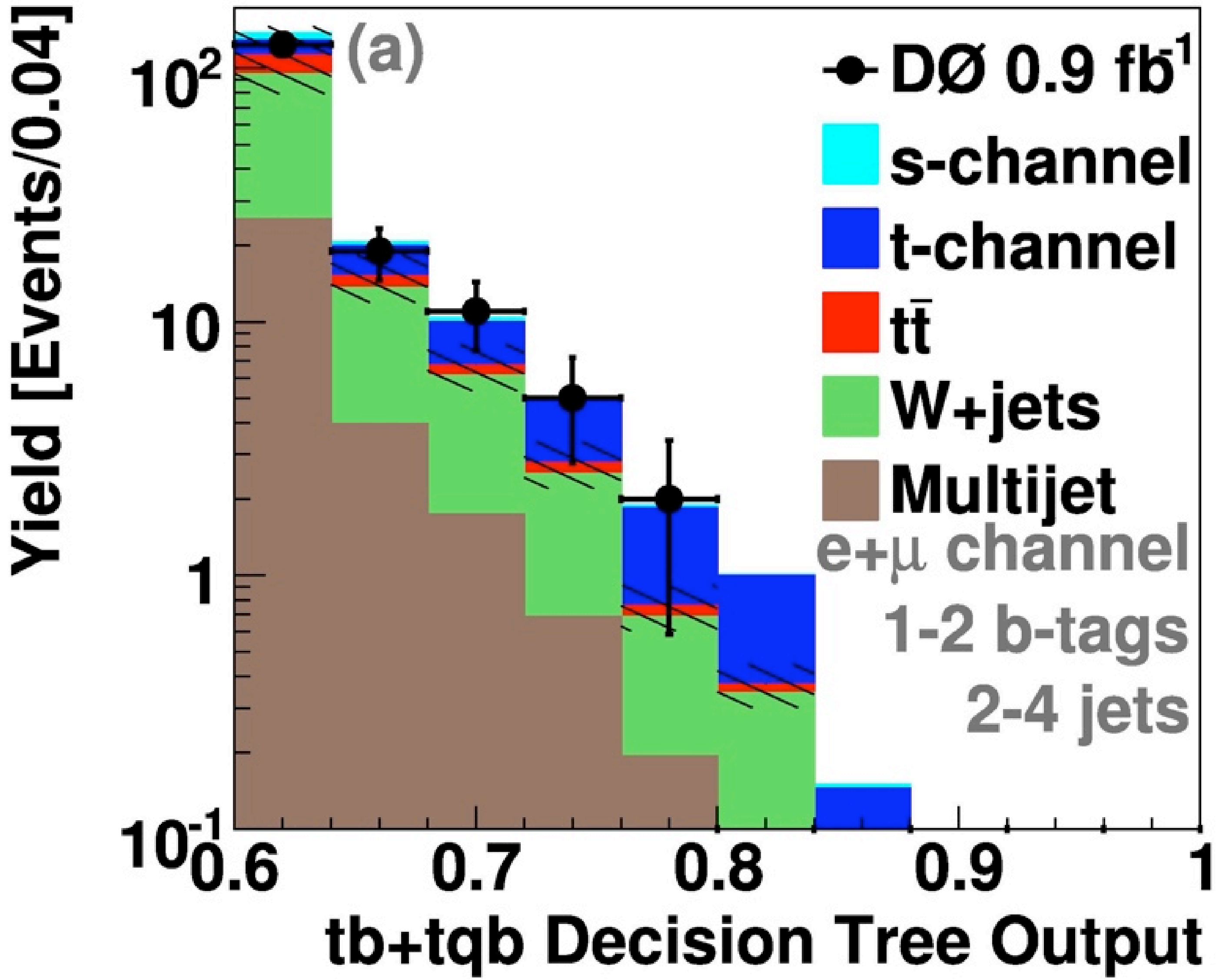


- Could just require NN/BDT > x and count events
  - Can do better: use shape of the distribution (i.e. **all** bins)
  - And then use the correlation between bins to constrain the systematics (systematics “profiling”)...

Systematics

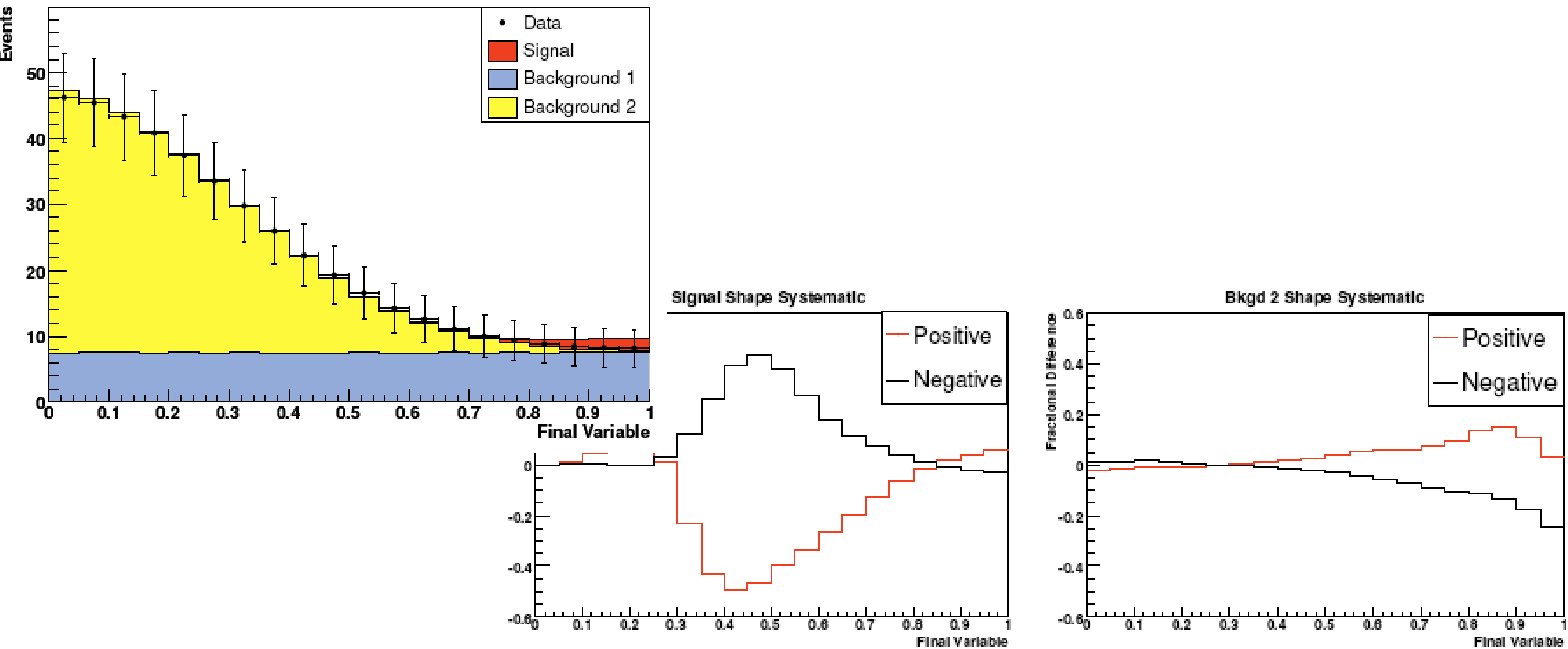
- Systematic uncertainties reduce the sensitivity of any analysis
  - In statistical terms, they are "nuisance parameters"
- Typically, you try to evaluate the impact of systematics by propagating through the full analysis chain. E.g.
  - You measure the jet energy scale in some (independent) way
  - You repeat your analysis shifting all jets up in scale by 1σ
  - You repeat your analysis shifting all jets down in scale by 1σ
  - You get two "new" NN/BDT/... output distributions yielding a systematic uncertainty "band"

- In this plot, hatches = 1σ uncertainty on background determination



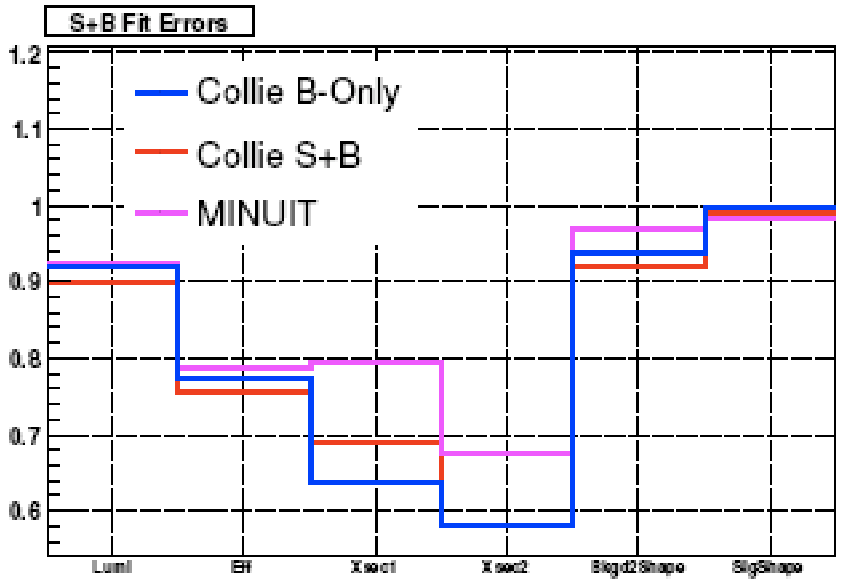
Systematics Profiling

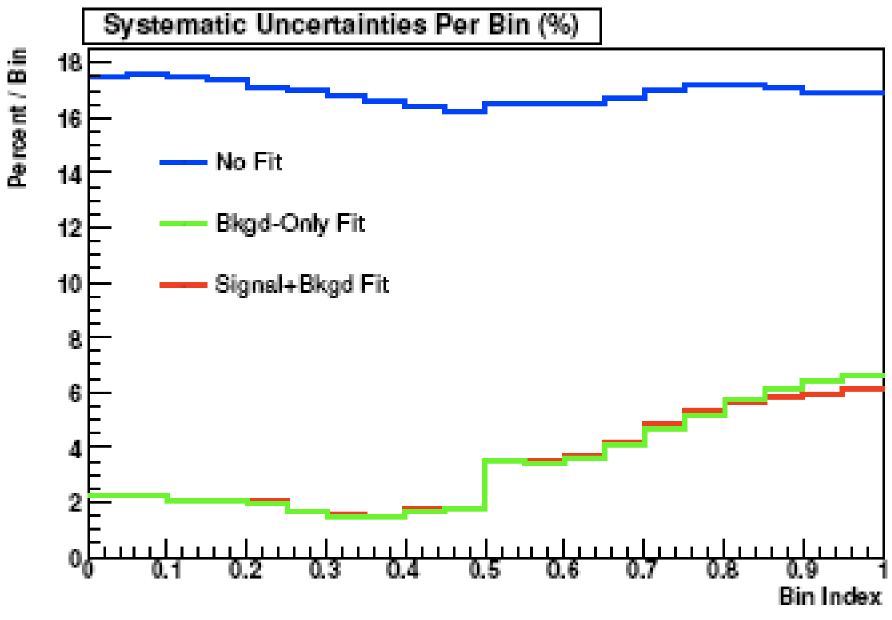
- Nuisance parameters tend to be correlated, but not 100%, among backgrounds
  - Can affect rates, shapes, or both (in any distribution), and often asymmetric and non-gaussian



- 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 you can maximize likelihood ratio as a function of nuisance parameters → constraint them
  - I.e. use full shape of distributions 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)

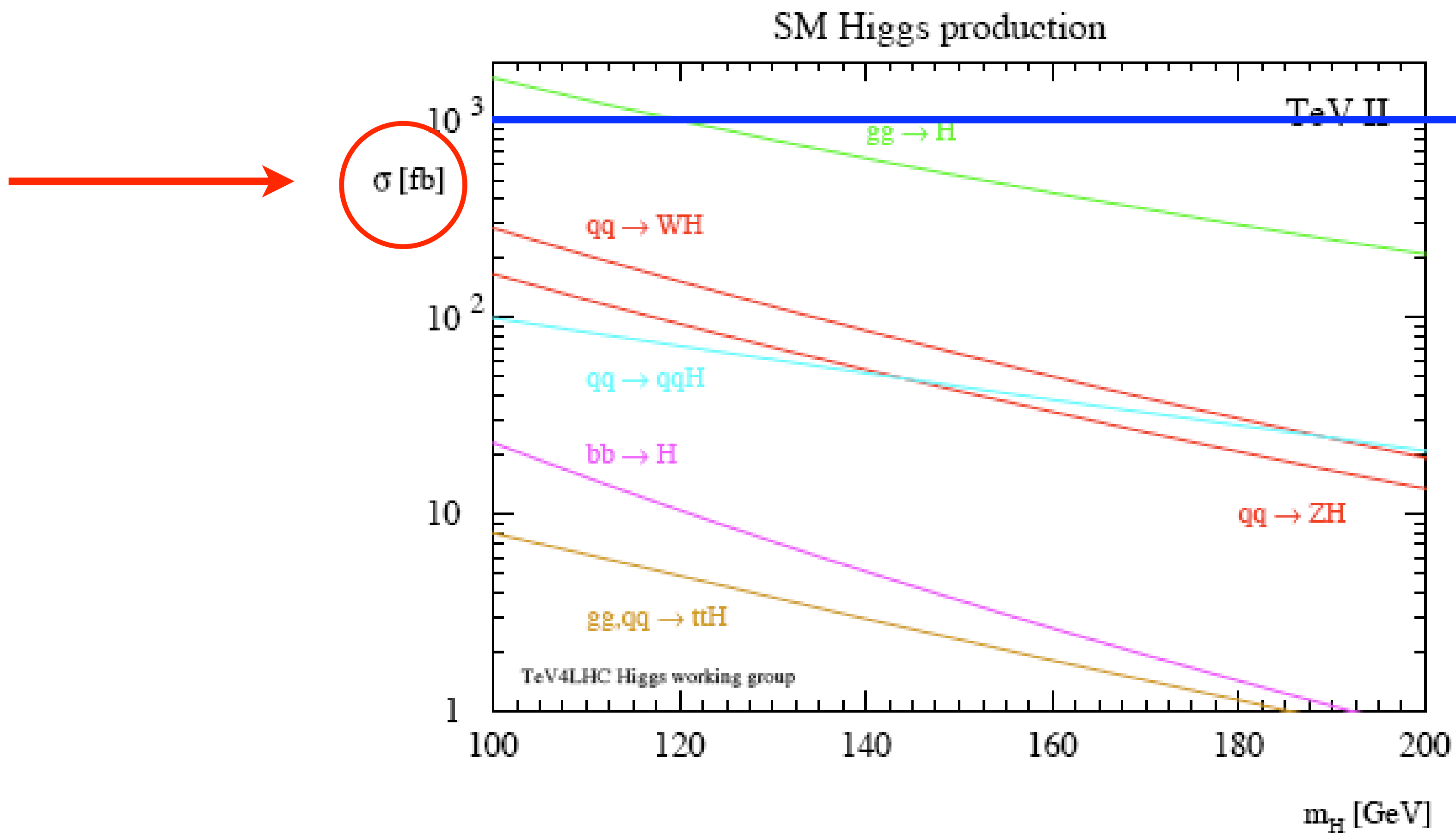




Dessert

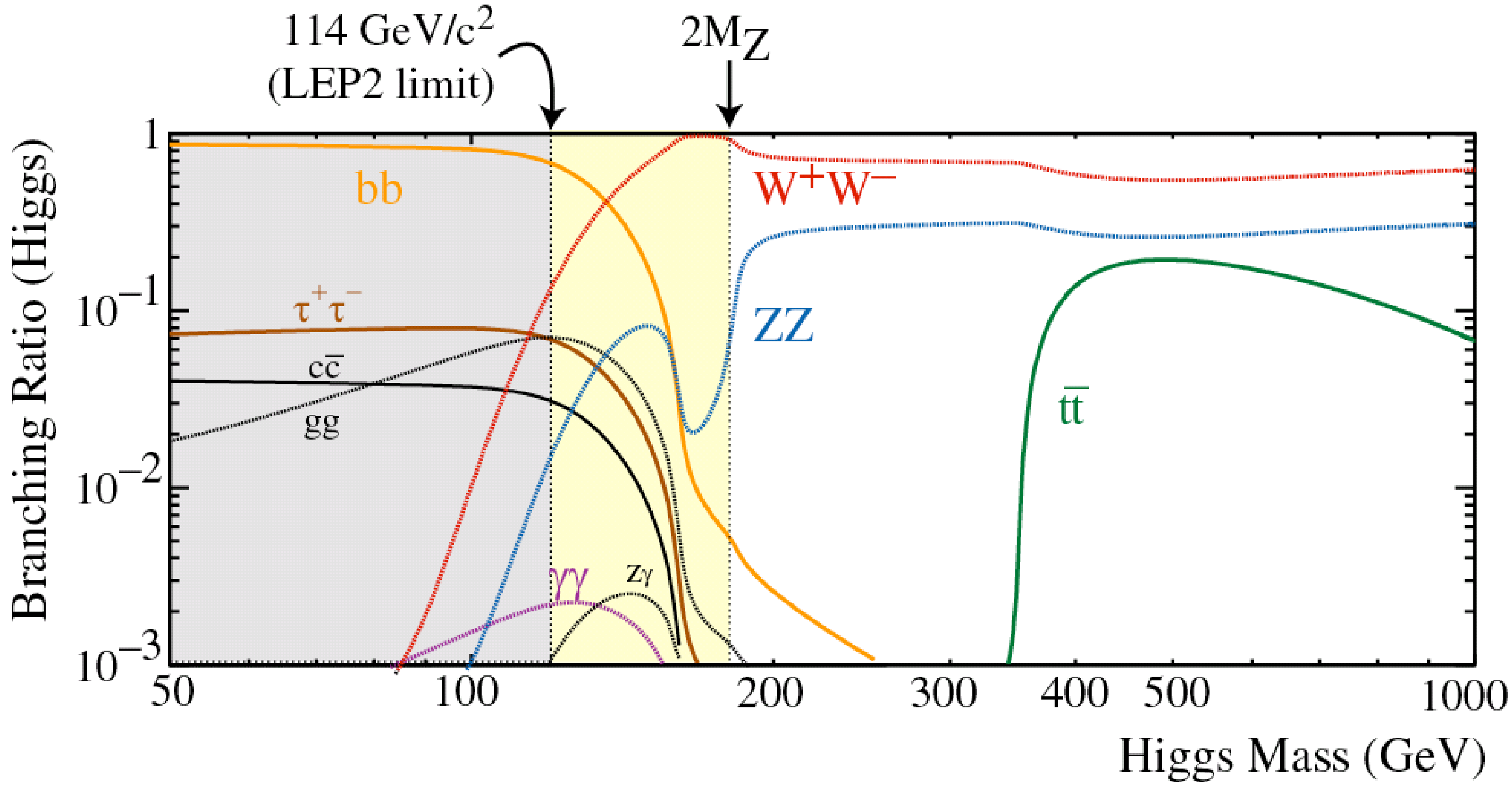
Producing Higgses at the Tevatron

- We currently have 5+ fb^{-1} of data on tape



5000 events (x many efficiency factors)

Higgs Decay



Low Mass
$H \to bb$

High Mass
$H \to WW$

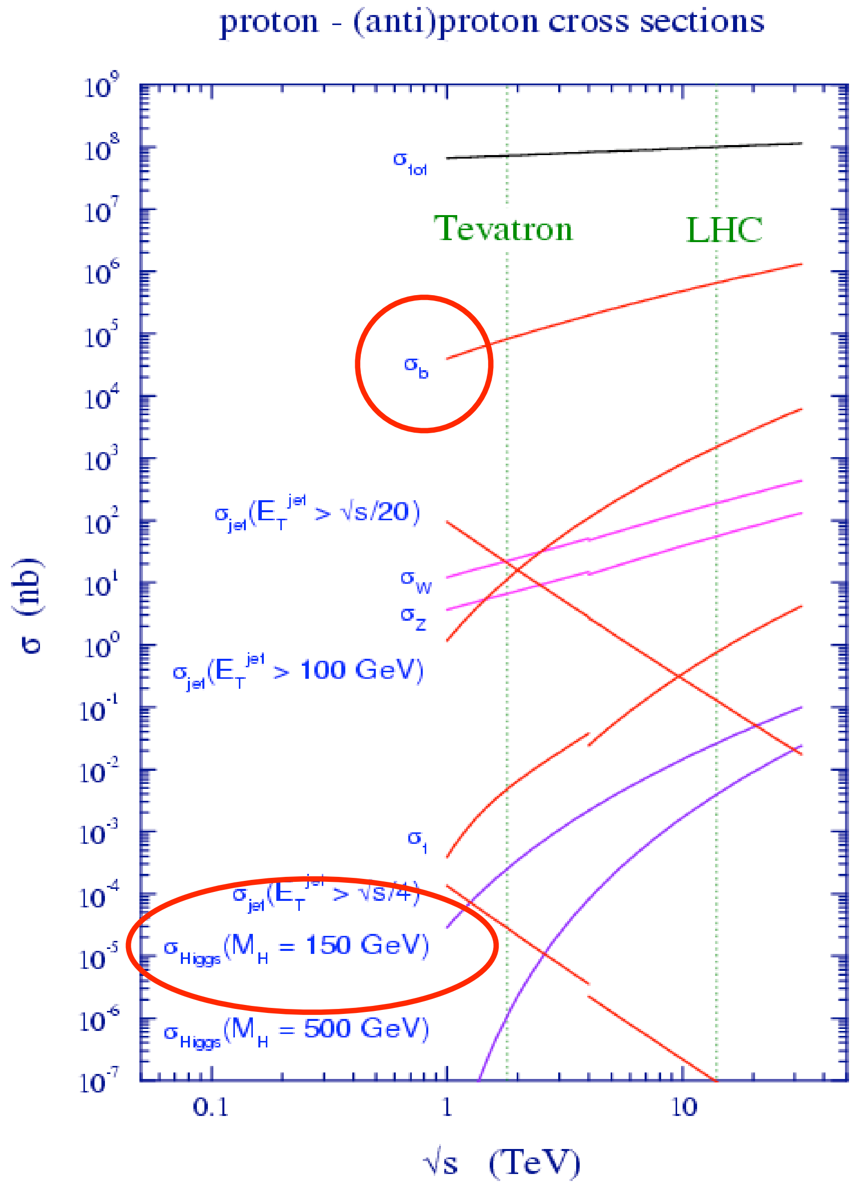
Search Channels

- Hadron collider
  - bb production ~9 orders of magnitude larger than H
  - gg → H → bb swamped

➡ At low mass look for pp → WH or ZH → W/Z bb (so down to 500 events)

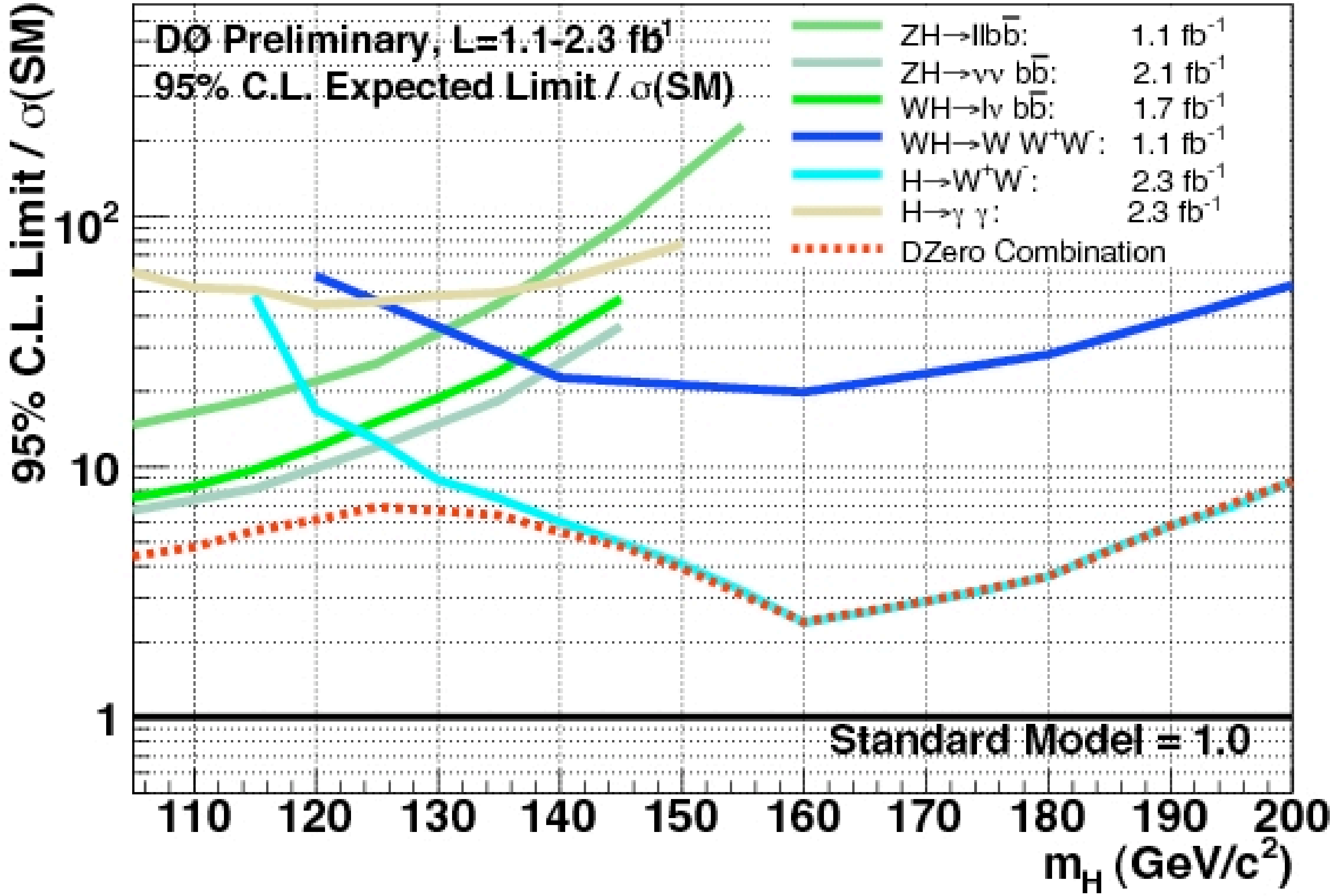
  - With leptonic W, Z decay (→ ~100)

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

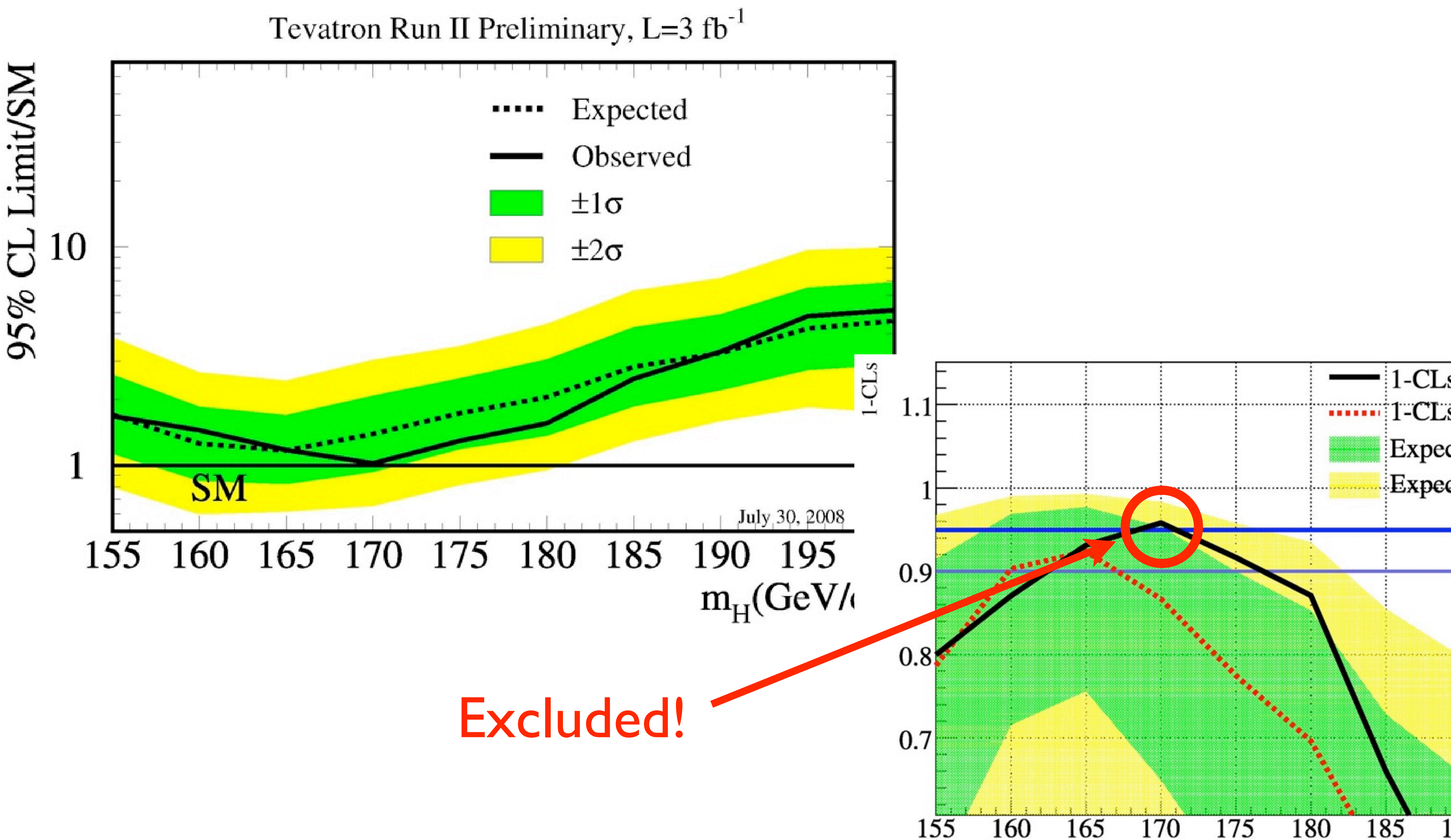


Many Channels

- For each optimize trigger, work on DQ, handle logistics, measure all biases, multivariate tools, systematics, ...
- Then combine...



- ... and with the work from our 600 buddies at CDF



Concluding

- Experimental physics is hard, and each measurement represents a huge amount of work
  - Cannot do justice to this in this short talk
  - E.g. calorimeter data quality requires manual inspection of all deviations
    - Is this known? New? How can it be fixed?
  - At each step, including MC generation, there are such difficulties
    - They all take time, but solving the problems is very rewarding
- And the physics result at the end is worth it!

➡ "Don't sweat the small stuff!"