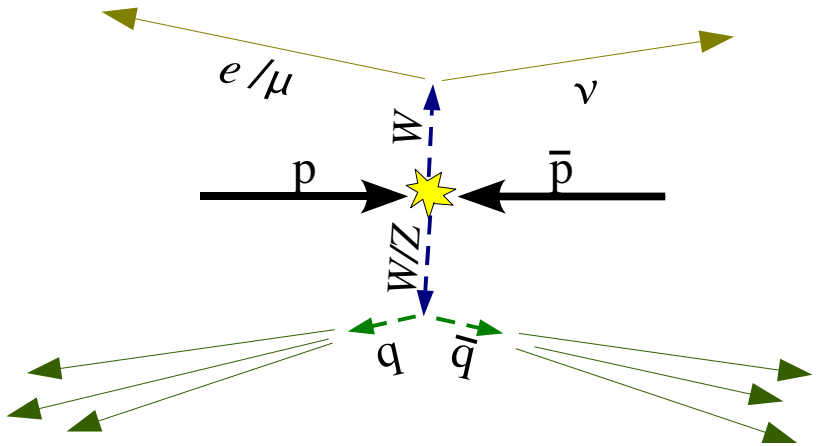
First Evidence for *WW* and *WZ* Diboson Production with Semi-leptonic Decays at a Hadron Collider

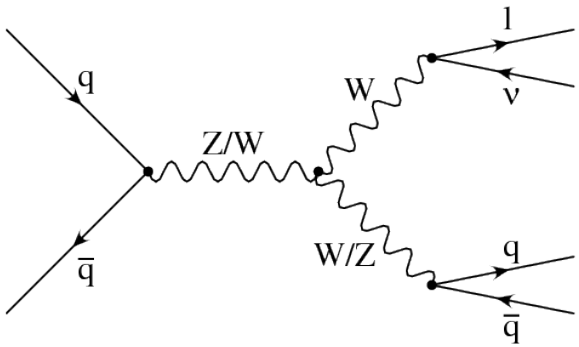


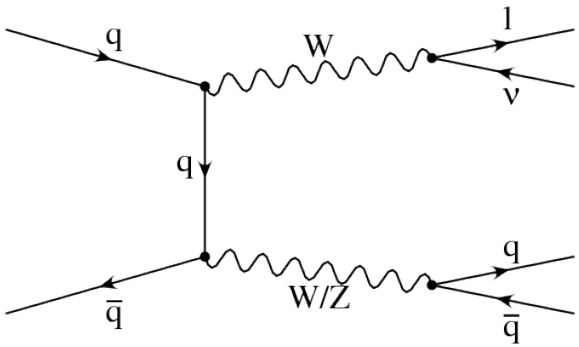
Joseph G. Haley
(D0 Collaboration)

April 14, 2009

Outline

- Motivation for Studying *WW/WZ→lνqq*
- The Apparatus
  - Accelerator Chain, D0 Detector
- Samples and Event Selection
  - Data, SM Simulation
- MC Corrections and Systematic Uncertainties
  - Efficiencies, Energies, MC Modeling
- Multivariate Discrimination
- The Measurement
  - Cross Section, Significance
- Summary
- Continuing Work



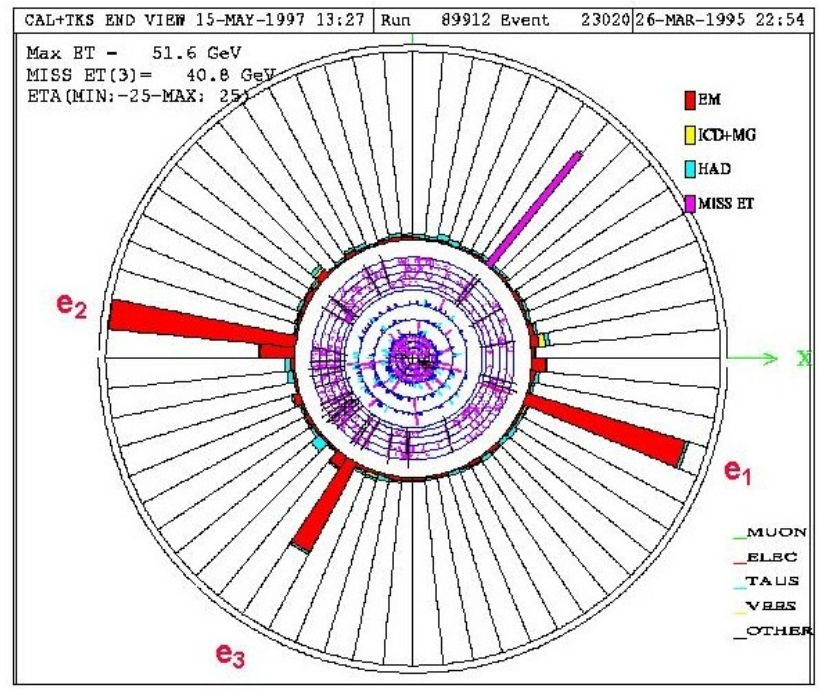


Motivation: Electroweak Studies

- Probe of the Electroweak sector of the SM at the Tevatron
- Deviations in production rate or event kinematics would indicate new physics
- Currently all diboson measurement are in fully leptonic final states
  - $WW \rightarrow l\nu l\nu$
    - DZero ~1.0/fb: $\sigma(WW) = 12.3 \pm 2.0$ pb
    - CDF ~0.83/fb: $\sigma(WW) = 13.6 \pm 3.1$ pb
    - SM prediction: $\sigma(WW) = 12.4 \pm 0.8$ pb
  - $WZ \rightarrow l\nu ll$
    - DZero ~1.0/fb: $\sigma(WZ) = 2.7\ ^{+1.7}_{-1.3}$ pb
    - CDF ~1.9/fb: $\sigma(WZ) = 4.3\ ^{+1.1}_{-1.1}$ pb
    - SM prediction: $\sigma(WZ) = 3.7 \pm 0.3$ pb

$M_T(e_2,\nu) = 74.7$ GeV/c^2

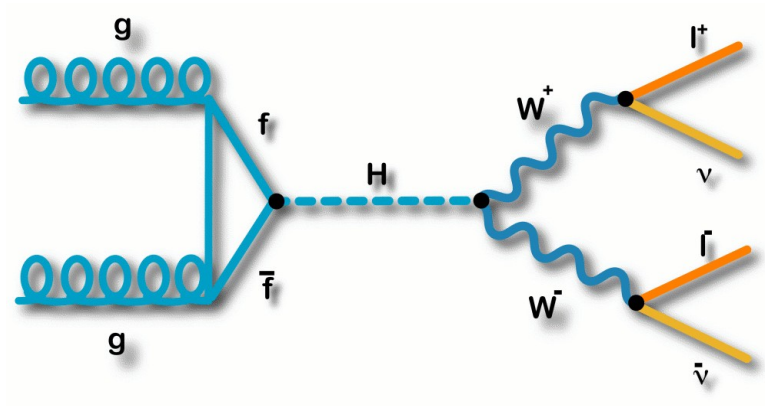
$M(e_1,e_3) = 93.6$ GeV/c^2

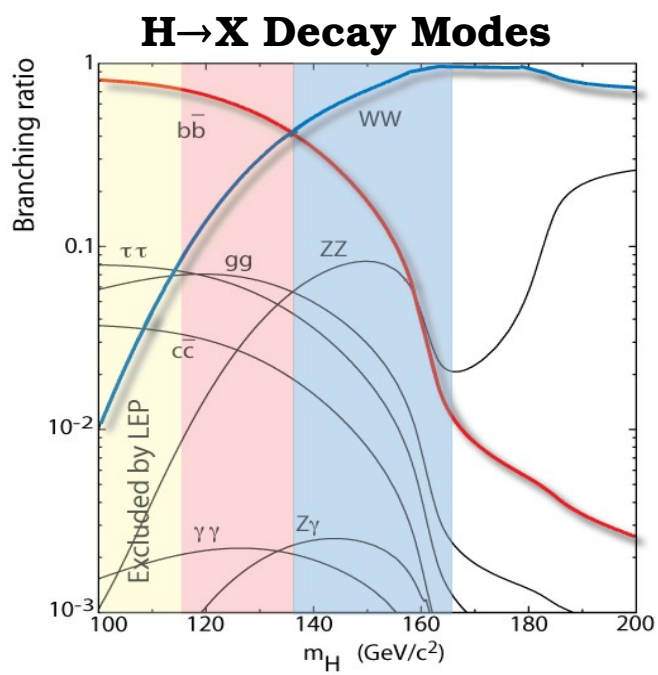


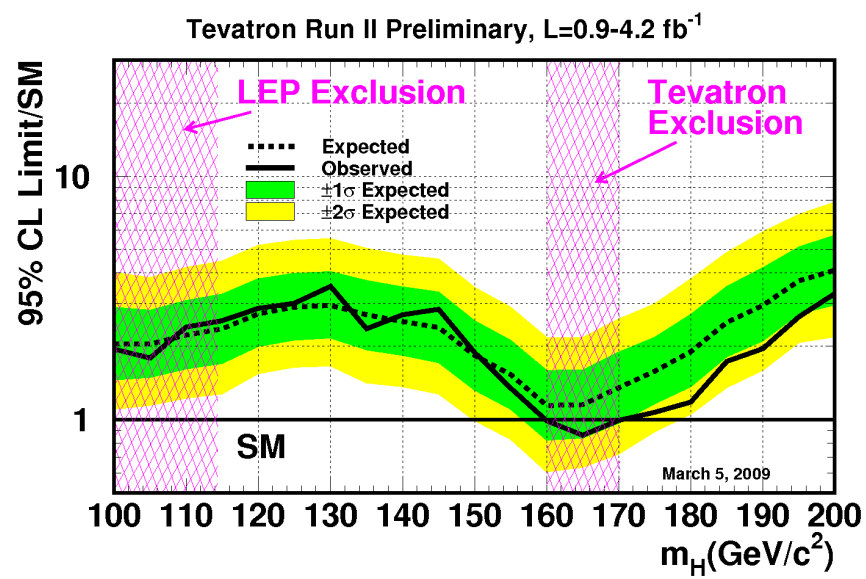
⇒ $WW/WZ \rightarrow l\nu qq$ complimentary to fully leptonic measurements

Motivation: Higgs Searches

- *H→WW* is the dominant decay mode for a high mass Higgs ($m_H > 135$ GeV/c^2)
  - Drives current exclusion limits
  - Direct diboson production is the single most important background
  - Important to understand this contribution

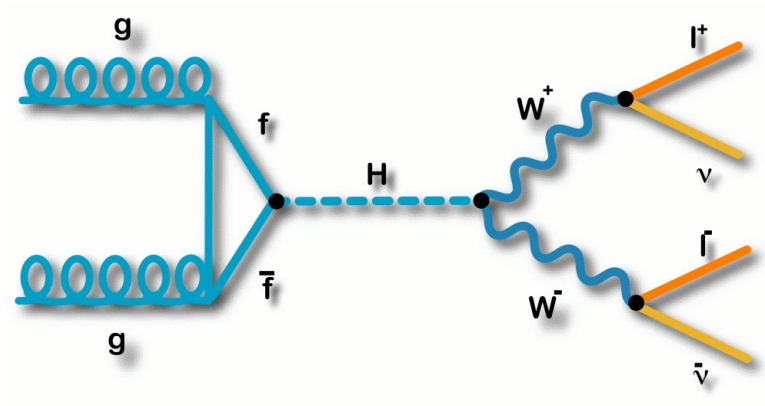




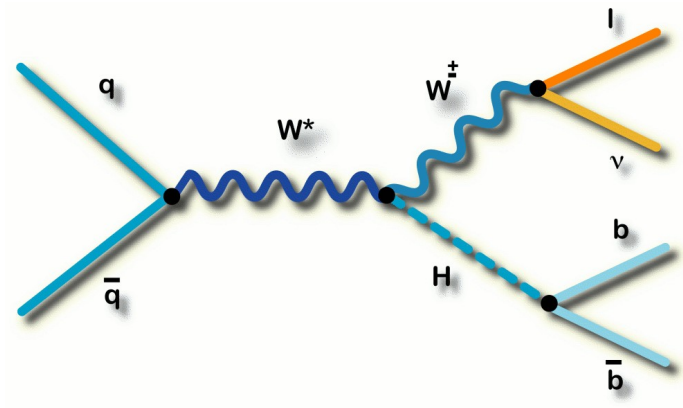


Motivation: Higgs Searches

- *H→WW* is the dominant decay mode for a high mass Higgs ($m_H > 135$ GeV/c^2)
  - Drives current exclusion limits
  - Direct diboson production is the single most important background
  - Important to understand this contribution



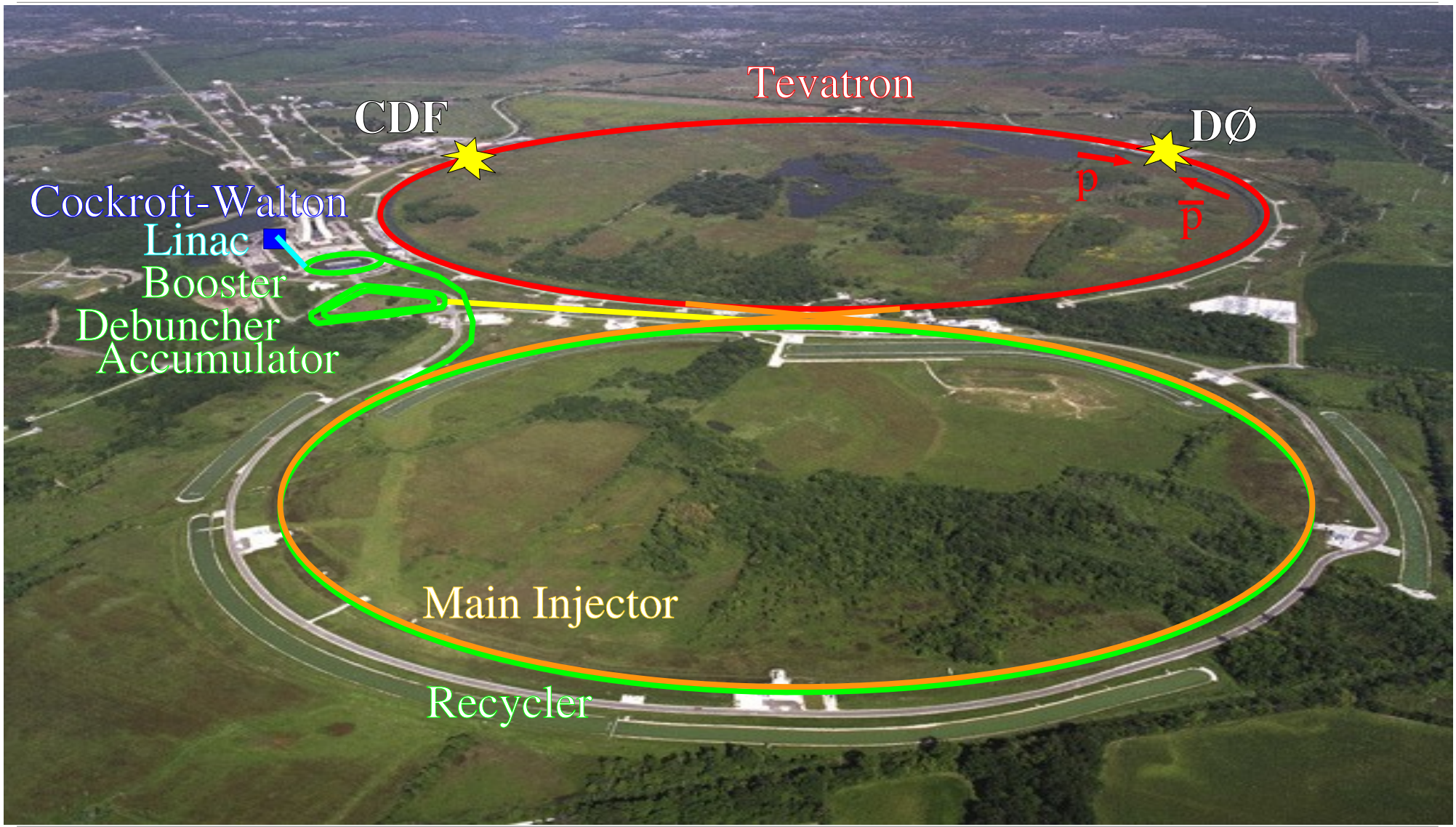
- *WH→lνbb* is a promising search mode for low mass Higgs ($m_H < 135$ GeV/c^2)
  - Similar topology/final state to *WW/WZ→lνqq*
  - Similar challenges
    - Small signal in a large background: *WW/WZ→lνqq* B/S ≈ 35, *WH→lνbb* B/S ≳ 90
  - Similar analysis techniques
    - Multivariate classifiers, statistical treatment



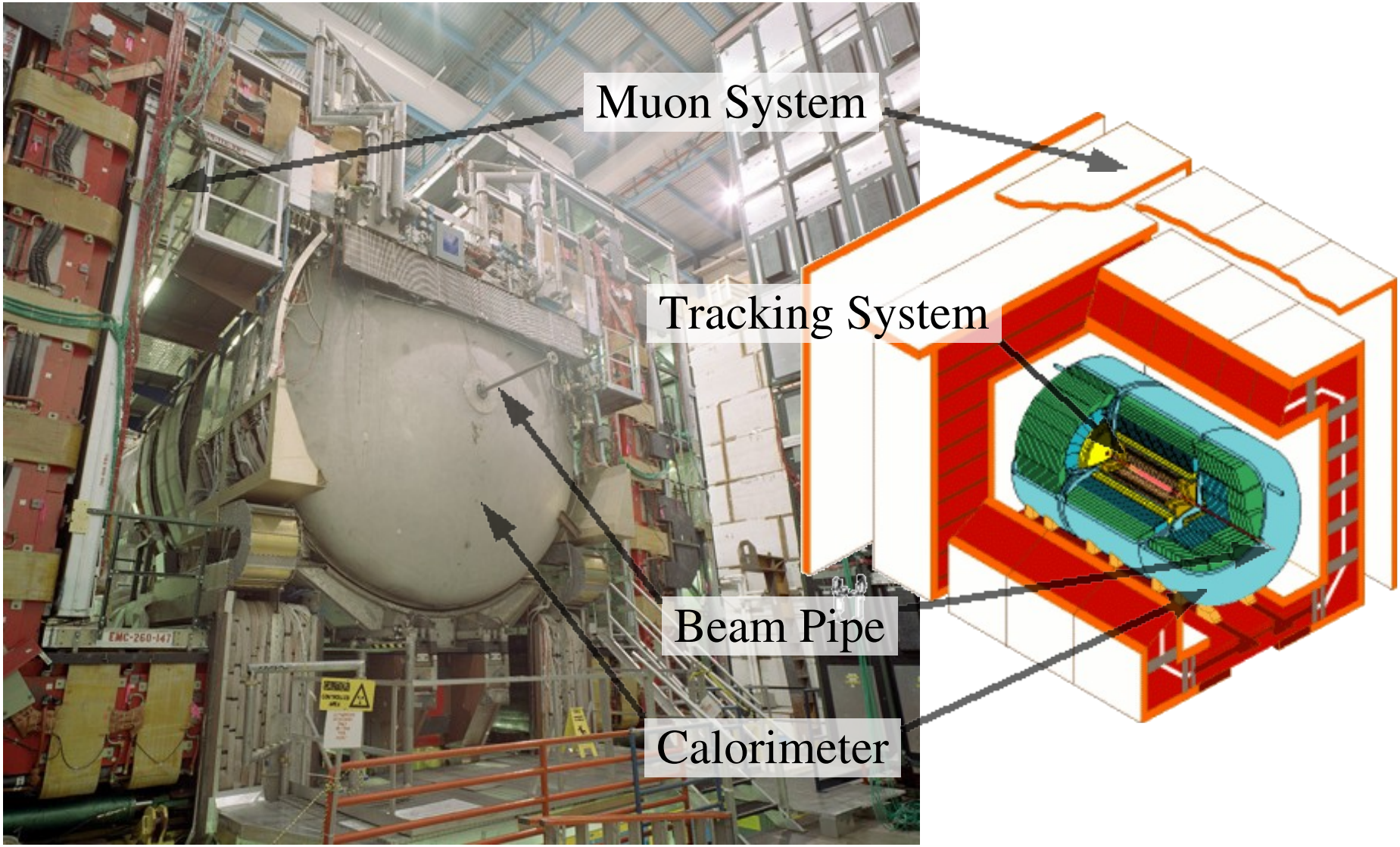
⇒ *WW/WZ→lνqq* is a proving ground for Higgs search

Apparatus

The Tevatron

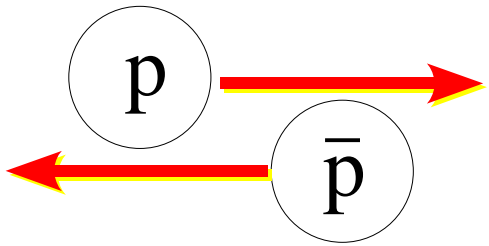


The D0 Detector



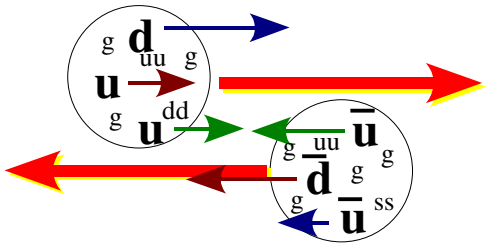
Hadron Collisions and Coordinate System

- Momentum of colliding protons and antiprotons divided among partons
  - Detector is not the zero-momentum-frame



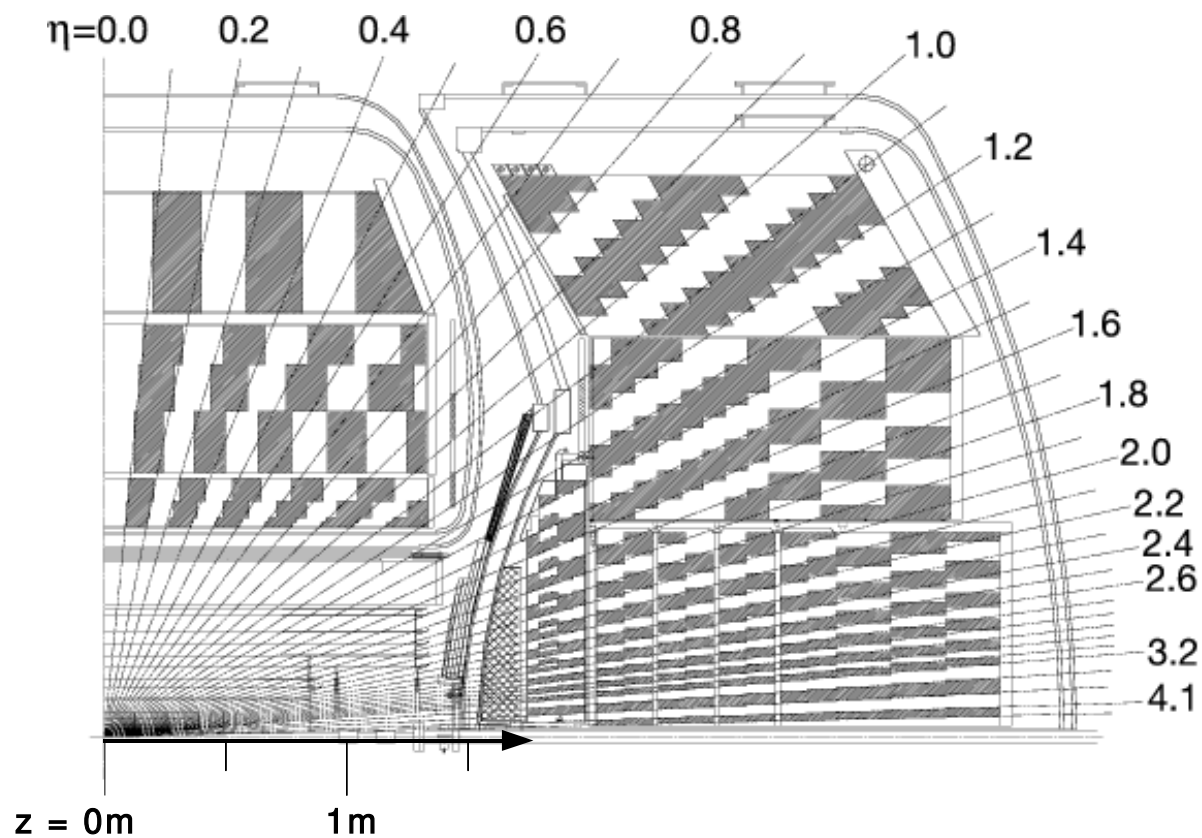
Hadron Collisions and Coordinate System

- Momentum of colliding protons and antiprotons divided among partons
  - Detector is not the zero-momentum-frame
    - Total transverse momentum is zero
      - $p_x = 0$, $p_y = 0$
    - Total longitudinal momentum is unknown
      - $p_z = ?$

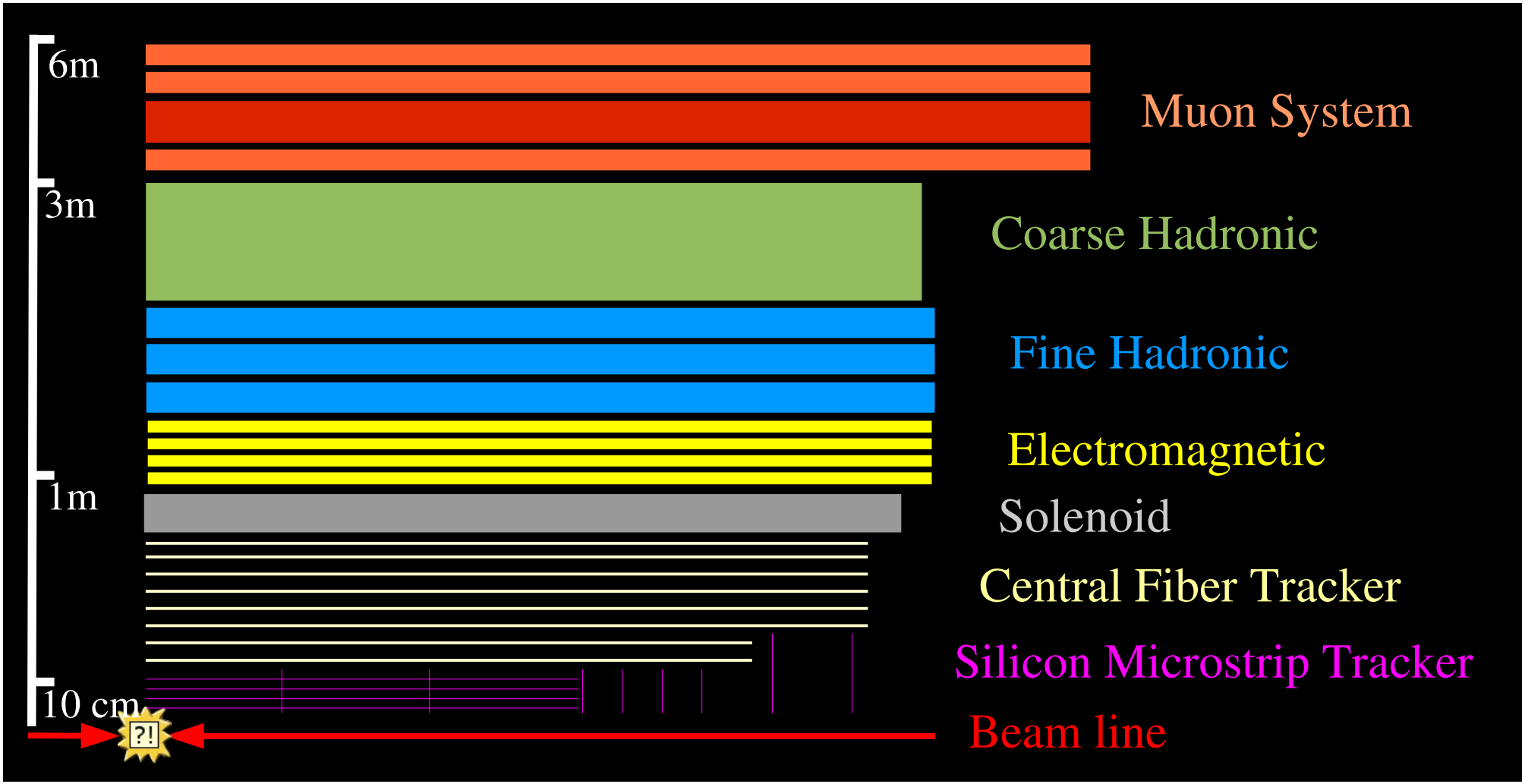


Hadron Collisions and Coordinate System

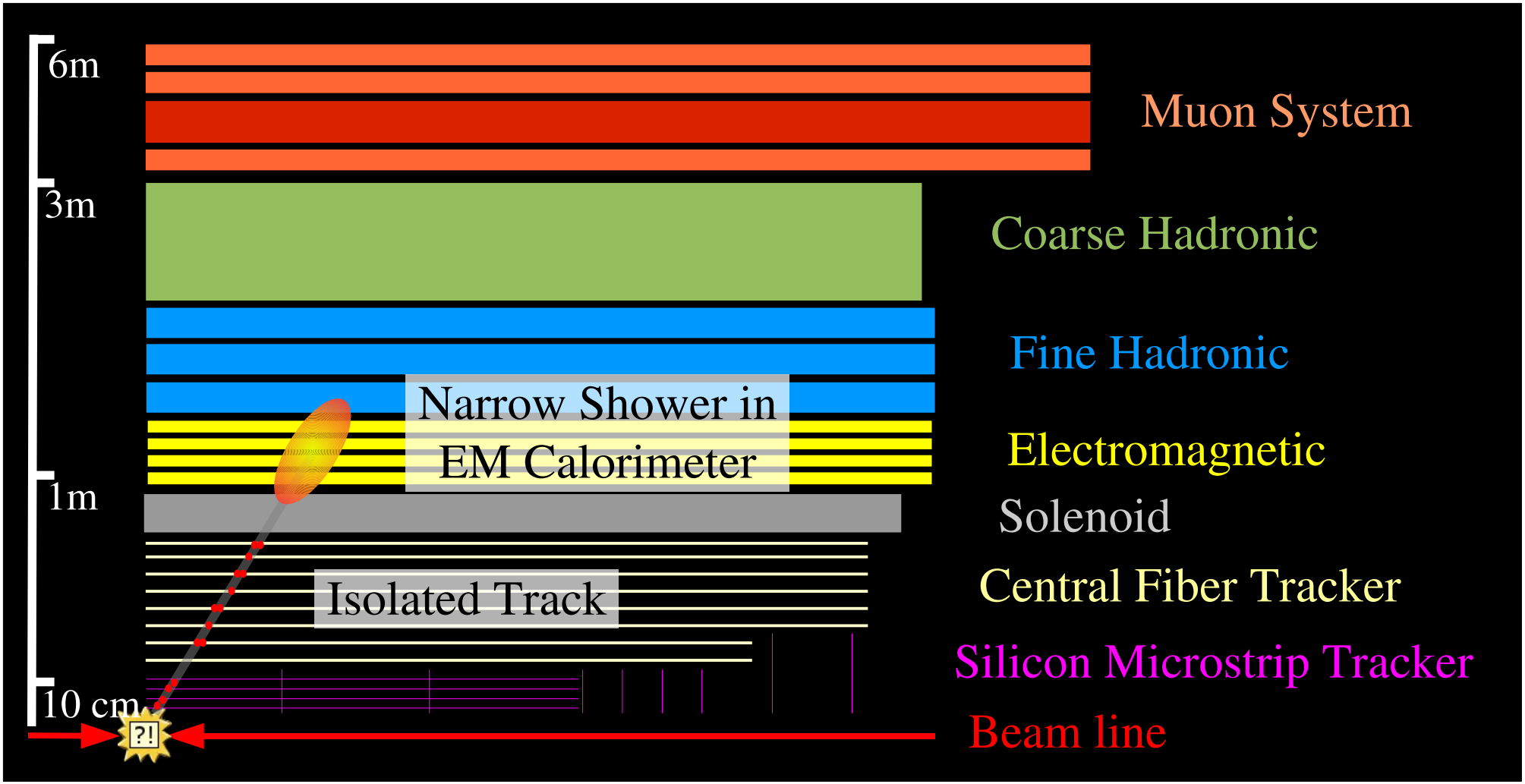
- Momentum of colliding protons and antiprotons divided among partons
  - Detector is not the zero-momentum-frame
    - Total transverse momentum is zero
      - $p_x = 0$, $p_y = 0$
    - Total longitudinal momentum is unknown
      - $p_z = ?$
  - Convenient to use pseudorapidity η instead of polar angle θ
    - $\eta \equiv - \ln[\tan(\theta/2)]$
    - Approaches Lorentz invariant rapidity for $E \gg m$



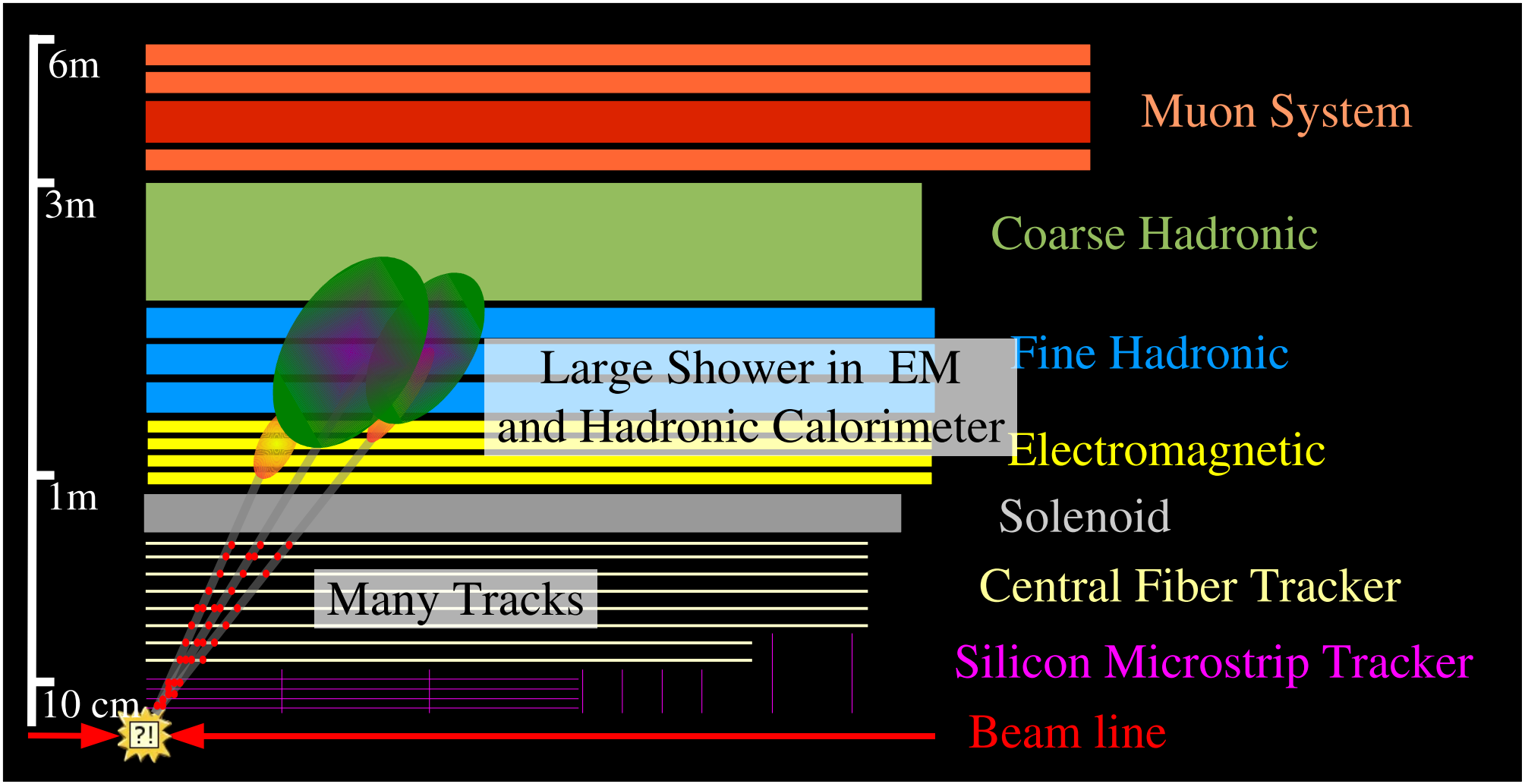
The D0 Detector: A Schematic View



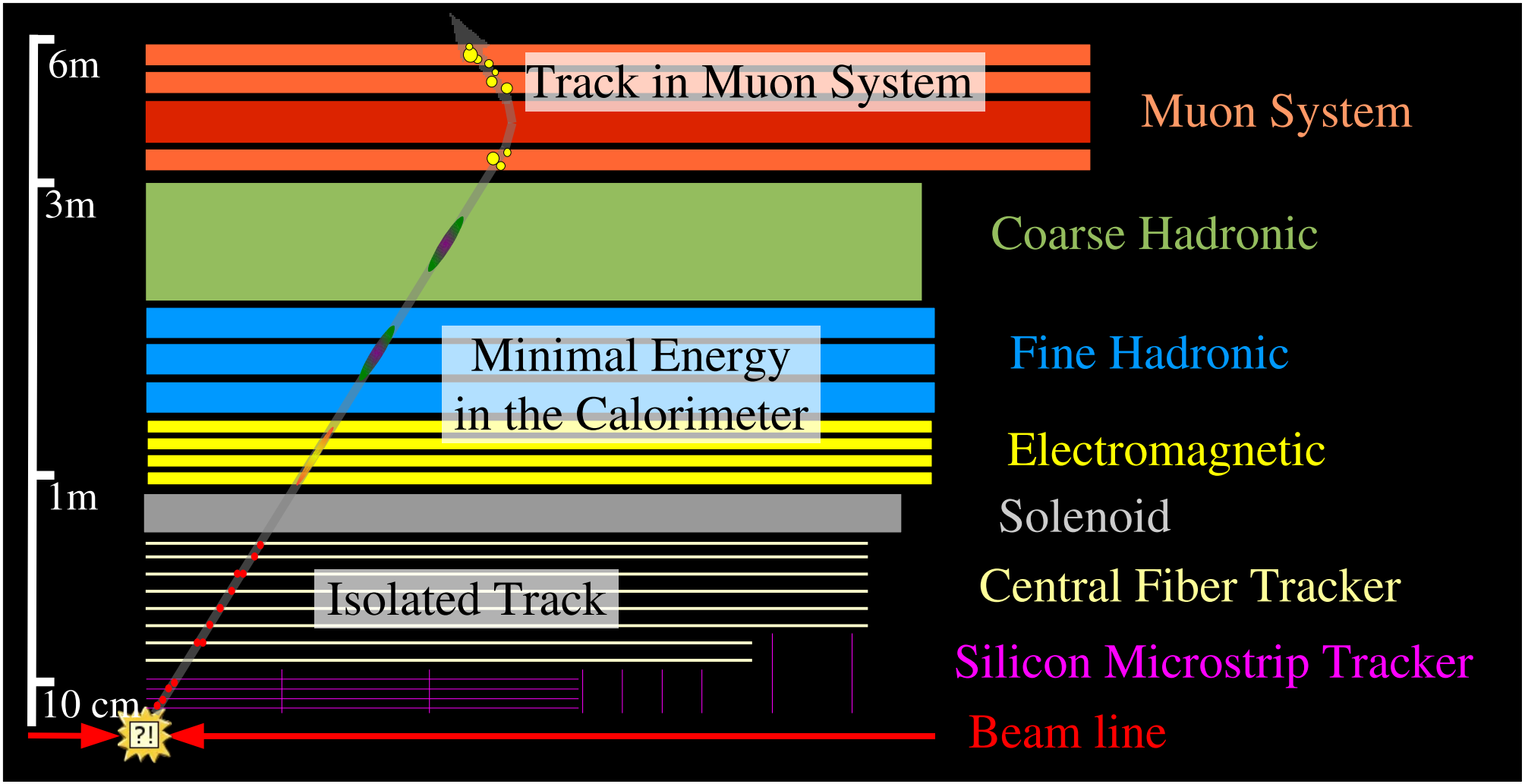
Electrons


6m
3m
1m
10 cm
Muon System
Coarse Hadronic
Fine Hadronic
Narrow Shower in
EM Calorimeter
Electromagnetic
Solenoid
Isolated Track
Central Fiber Tracker
Silicon Microstrip Tracker
?!
Beam line

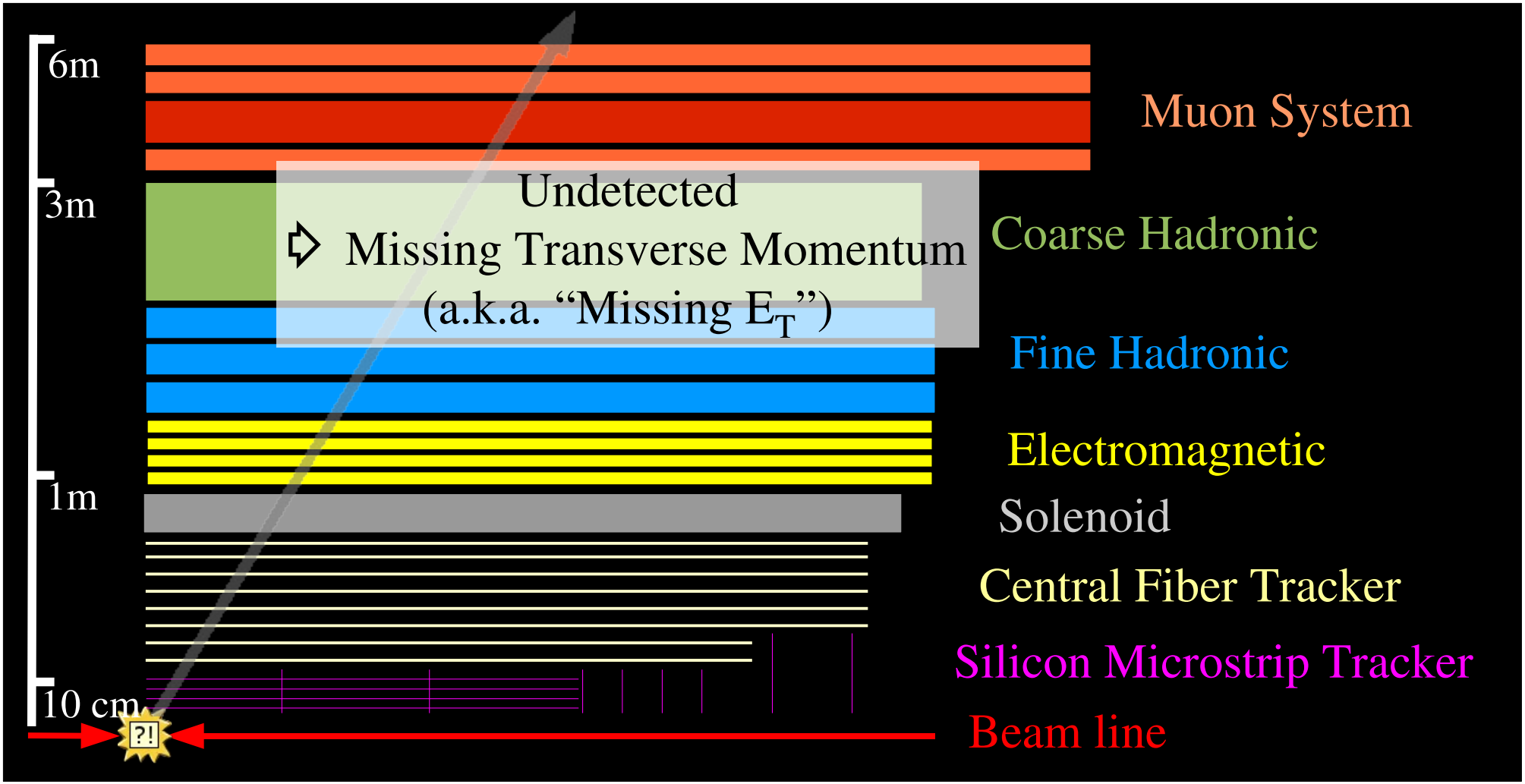
Jets


6m
3m
1m
10 cm
Muon System
Coarse Hadronic
Fine Hadronic
Large Shower in EM
and Hadronic Calorimeter
Electromagnetic
Solenoid
Central Fiber Tracker
Many Tracks
Silicon Microstrip Tracker
Beam line
?!

Muons


6m
3m
1m
10 cm
Track in Muon System
Muon System
Coarse Hadronic
Minimal Energy
in the Calorimeter
Fine Hadronic
Electromagnetic
Solenoid
Isolated Track
Central Fiber Tracker
Silicon Microstrip Tracker
?!
Beam line

Neutrinos


6m
3m
1m
10 cm
Muon System
Undetected
Missing Transverse Momentum
(a.k.a. "Missing E_T")
Coarse Hadronic
Fine Hadronic
Electromagnetic
Solenoid
Central Fiber Tracker
Silicon Microstrip Tracker
Beam line
?!

Trigger and DAQ

Detector

↓ 1.7 MHz

Level 1

↓ ~1 kHz

Level 2

↓ ~600 Hz

Level 3

↓ ~100 Hz

Tape/Storage

↓

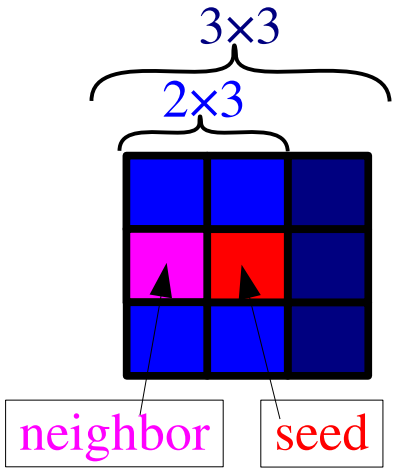
Analysis Selection

My Contributions to the Trigger

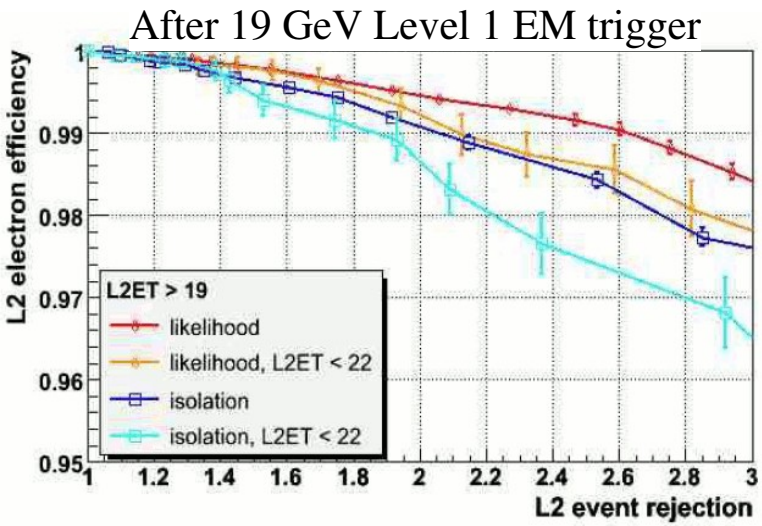
- Level 2 EM trigger upgrade for Run IIb
  - Needed to improve rejection to keep up with increased luminosity and improved Level 1 trigger
  - Constructed a Level 2 EM likelihood variable
    - Four inputs:

$$\frac{\text{EM } E_T \text{ seed}}{\text{TOT } E_T \text{ seed}} \qquad \frac{\text{EM } E_T\ 3\times3}{\text{TOT } E_T\ 3\times3}$$

$$\frac{\text{EM } E_T \text{ neighbor}}{\text{EM } E_T \text{ seed}} \qquad \frac{\text{EM } E_T \text{ seed}}{\text{EM } E_T\ 3\times3}$$



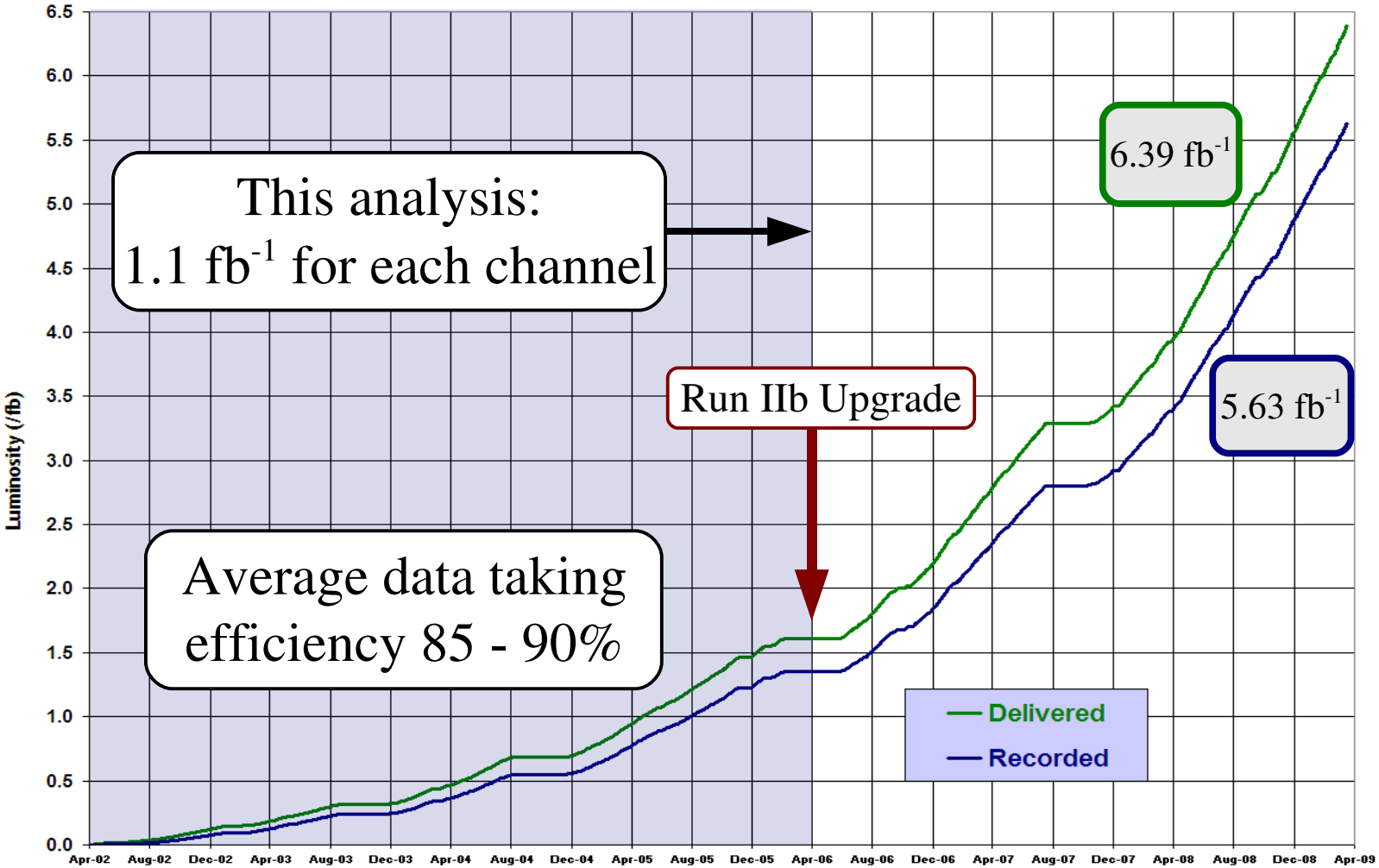
    - Factor of ~3 background rejection with signal efficiency near 99%
    - In some case, ~30% more rejection than Level 2 isolation variable



Samples and Event Selection

Data Sample

- Dataset: Run IIa of the Tevatron (2002 – 2006)



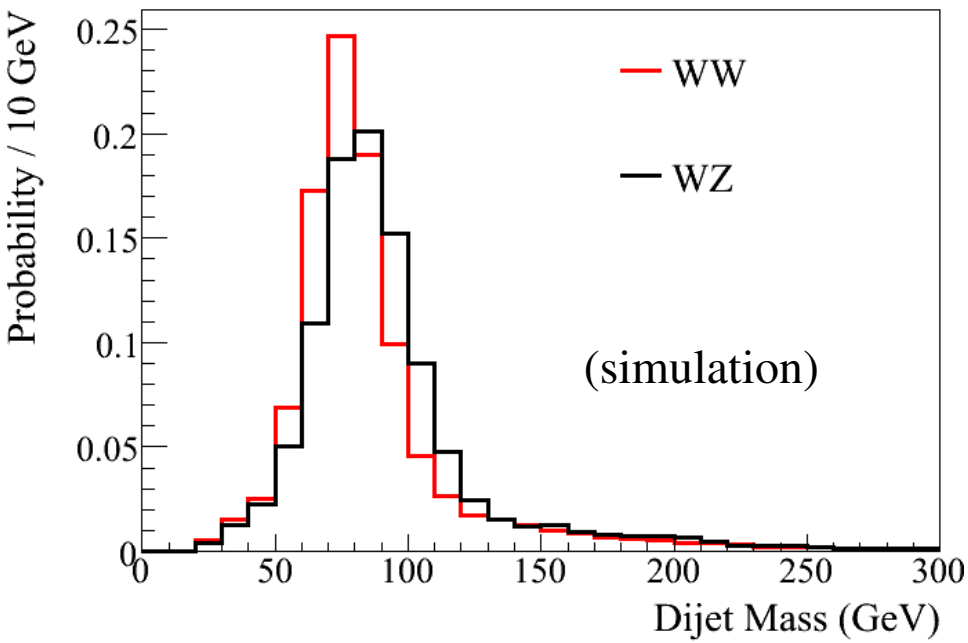
SM Expectation

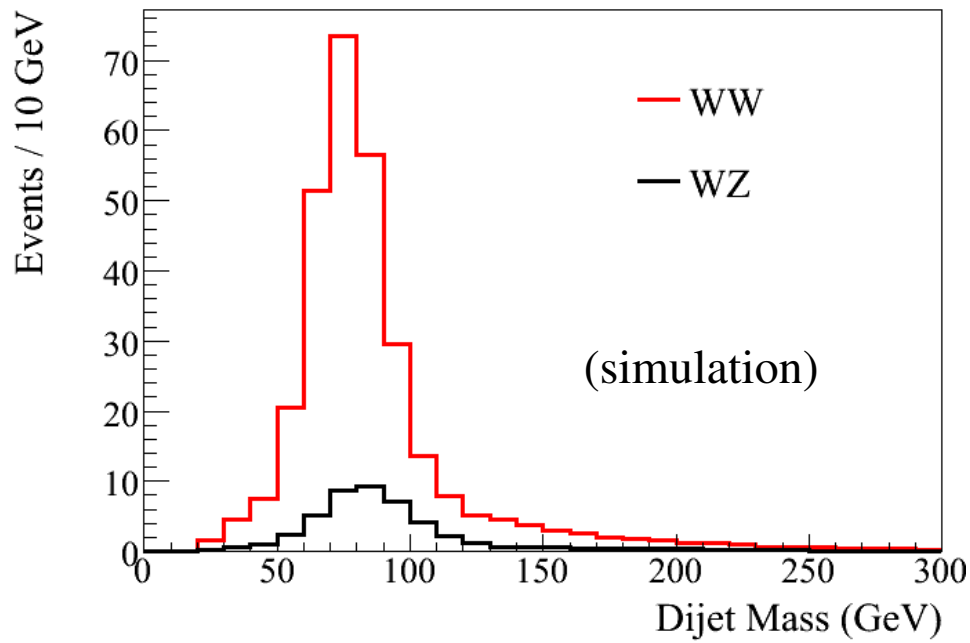
- Simulated Samples
  - Monte Carlo generators simulate the SM prediction for signal and backgrounds with real high p_T leptons
    - ALPGEN+Pythia: *W*+jet, *Z*+jets, *tt*+jets
    - COMPHEP+Pythia: Single-top
    - Pythia: Diboson Signal

- Multijet Estimation
  - Events without a real high p_T lepton
  - Probability for a jet to mimic a lepton is very small, but QCD processes have very large cross sections
  - Estimated from data using orthogonal selection
  - Corrected for contributions from samples modeled with MC

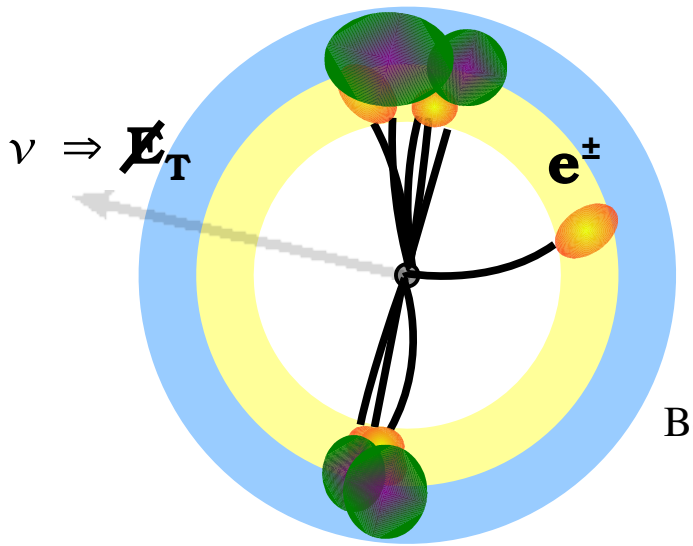
Note on WW vs. WZ

- *WW* and *WZ* treated as the same signal
  - Difference in mass of ~10 GeV/c^2
  - Dijet mass resolution ~ 20 GeV/c^2
- Relative contributions
  - $WW(WZ) \rightarrow l\nu qq$ $\sigma \times$BR $\cong$ 3.5(0.5) pb

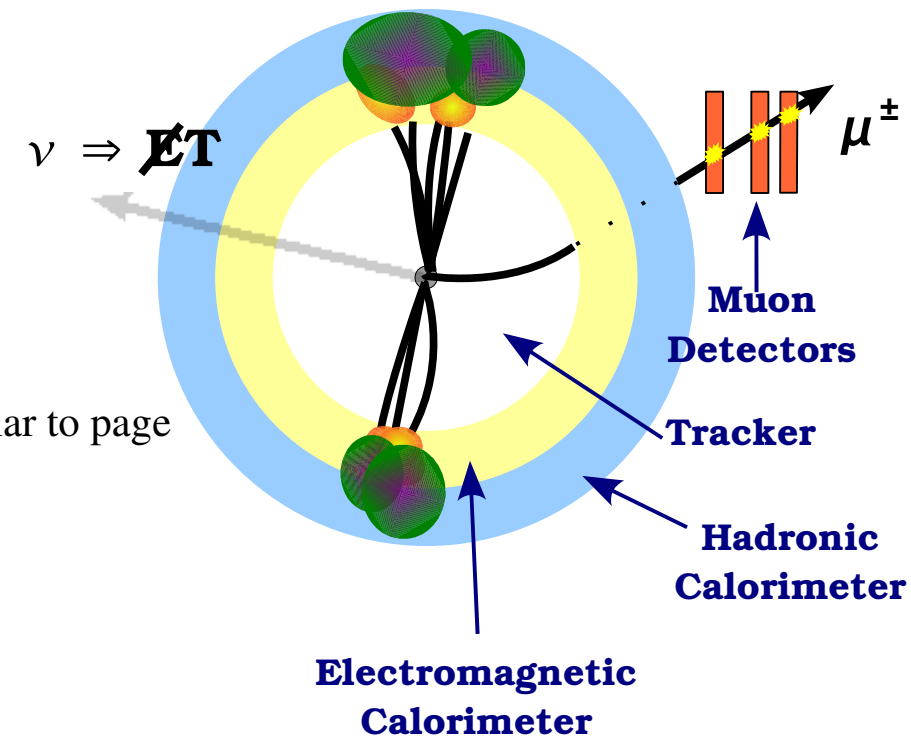




Event Selection



Beam perpendicular to page

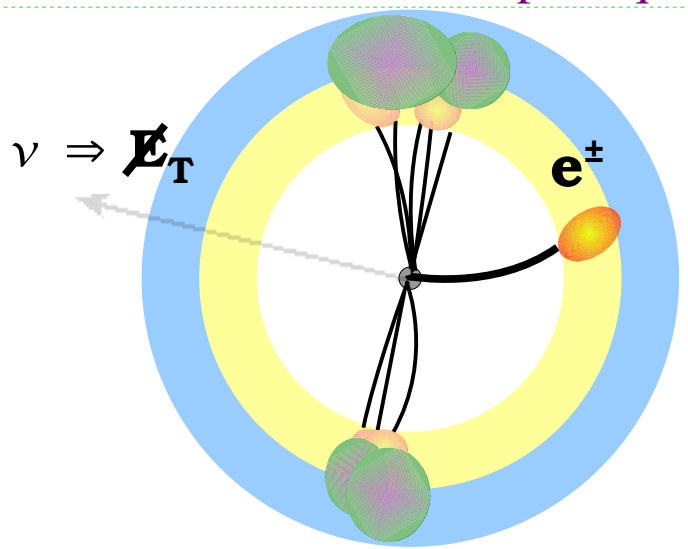


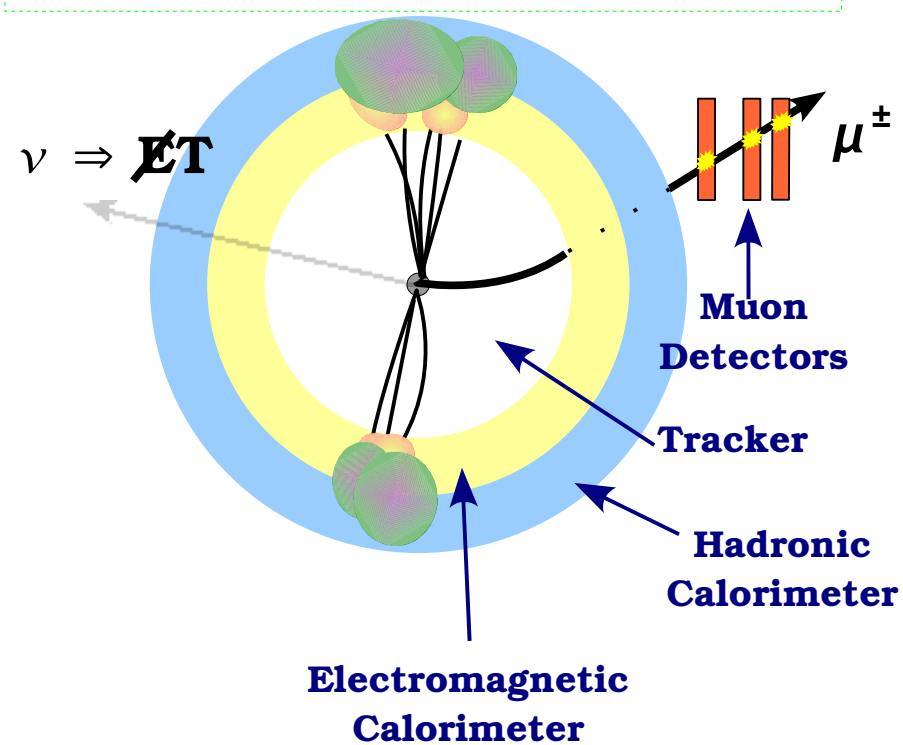
Event Selection

- $W \rightarrow l\nu$ Selection
  - Electron:
    - $p_T \geq 20$ GeV/c, $|\eta| < 1.1$
    - Isolated track and EM shower
    - Electron shower shape requirements

OR

  - Muon:
    - $p_T \geq 20$ GeV/c, $|\eta| < 2.0$
    - Hits in all three muons layers
    - Isolated in tracker and calorimeter



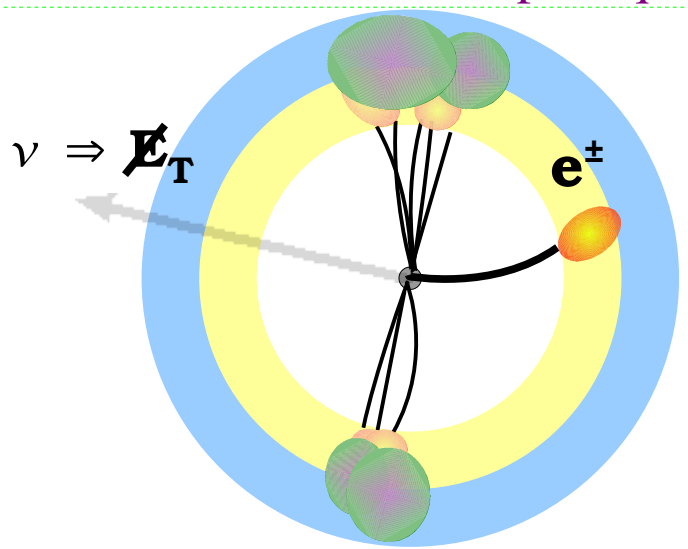


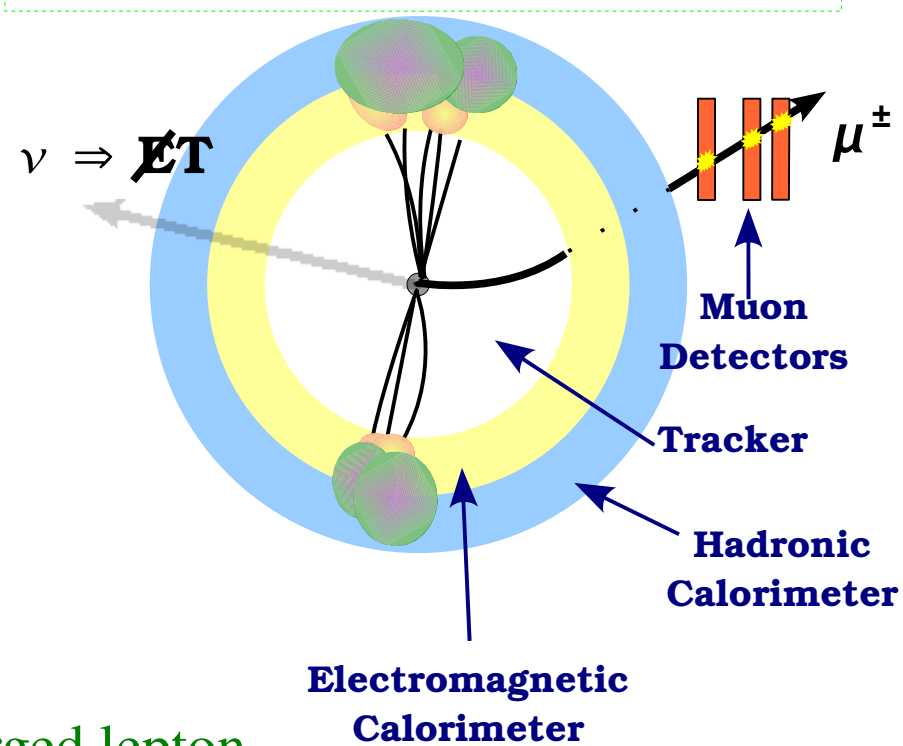
Event Selection

- $W \rightarrow l\nu$ Selection
  - Electron:
    - $p_T \geq 20$ GeV/c, $|\eta| < 1.1$
    - Isolated track and EM shower
    - Electron shower shape requirements

OR

  - Muon:
    - $p_T \geq 20$ GeV/c, $|\eta| < 2.0$
    - Hits in all three muons layers
    - Isolated in tracker and calorimeter



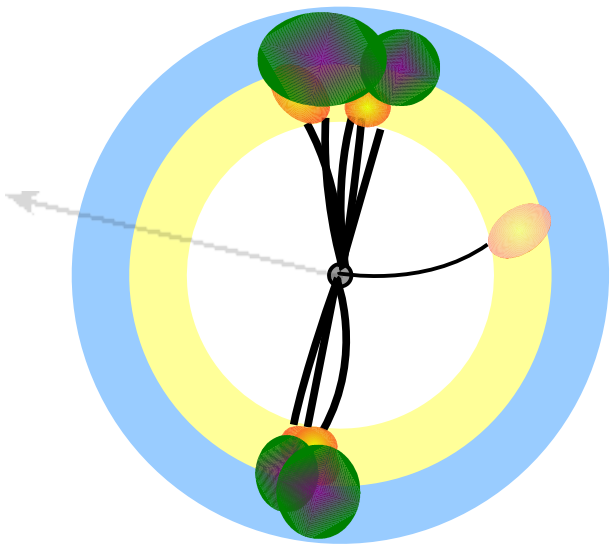


- Neutrino ⇒ Missing $E_T \geq 20$ GeV
- Transverse W mass ≥ 35 GeV/c^2
- Remove events with more than one charged lepton

Event Selection

- $W/Z \rightarrow qq$ Selection
  - At least two jets with radius $R = 0.5$ (cone algorithm)
  - Leading jet: $p_T \geq 30$ GeV/c, $|\eta| \leq 2.5$
  - Second jet: $p_T \geq 20$ GeV/c, $|\eta| \leq 2.5$

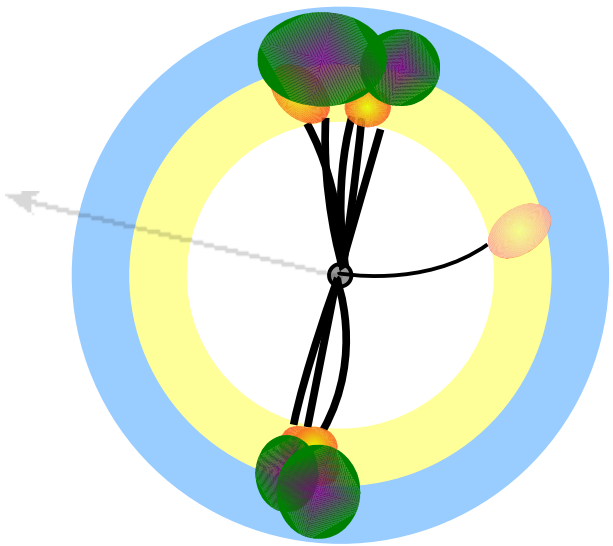
$$R \equiv \sqrt{(\Delta\phi + \Delta\eta)}$$



Event Selection

- *W/Z→qq* Selection
  - At least two jets with radius $R = 0.5$ (cone algorithm)
  - Leading jet: $p_T \geq 30$ GeV/c, $|\eta| \leq 2.5$
  - Second jet: $p_T \geq 20$ GeV/c, $|\eta| \leq 2.5$

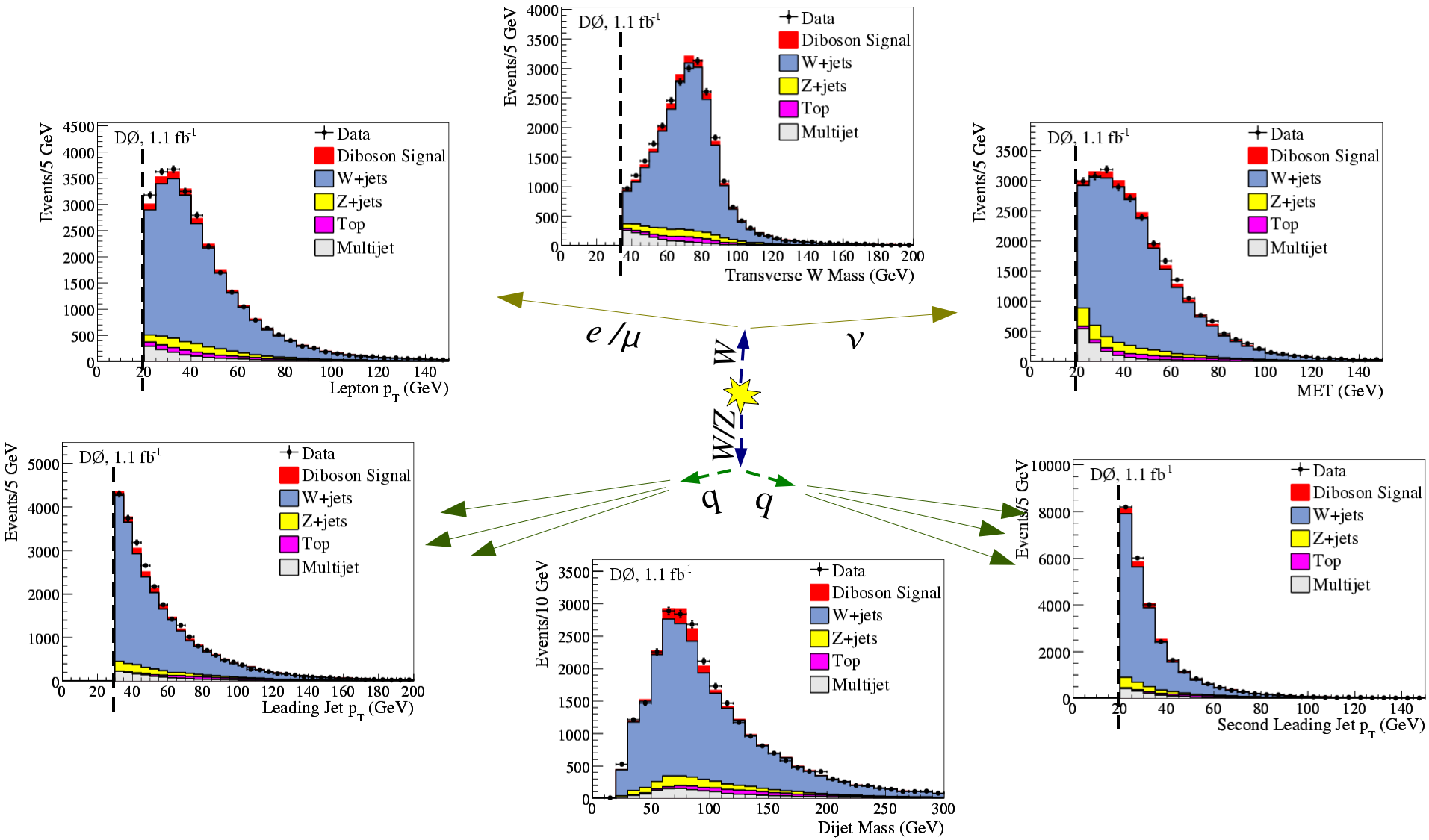
$$R \equiv \sqrt{(\Delta\phi + \Delta\eta)}$$



Selected Events

Source	$e\nu qq$	$\mu\nu qq$
Diboson	360.5 ± 2.3	427.3 ± 2.7
W+jets	10226 ± 76	12012 ± 88
Z+jets	408 ± 13	1239 ± 20
Top	463.3 ± 2.2	437.0 ± 2.2
Multijet	825 ± 11	327.0 ± 9.6
Total Predicted	12283 ± 78	14442 ± 91
Data	12473 ± 112	14392 ± 120
$S/\sqrt{S+B}$	3.25	3.56

Event Selection

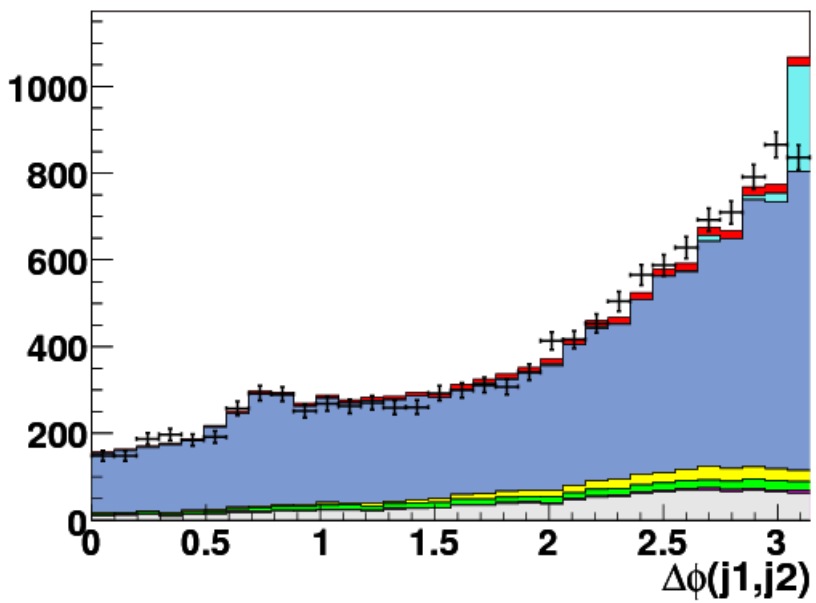


Monte Carlo Corrections and Systematic Uncertainties

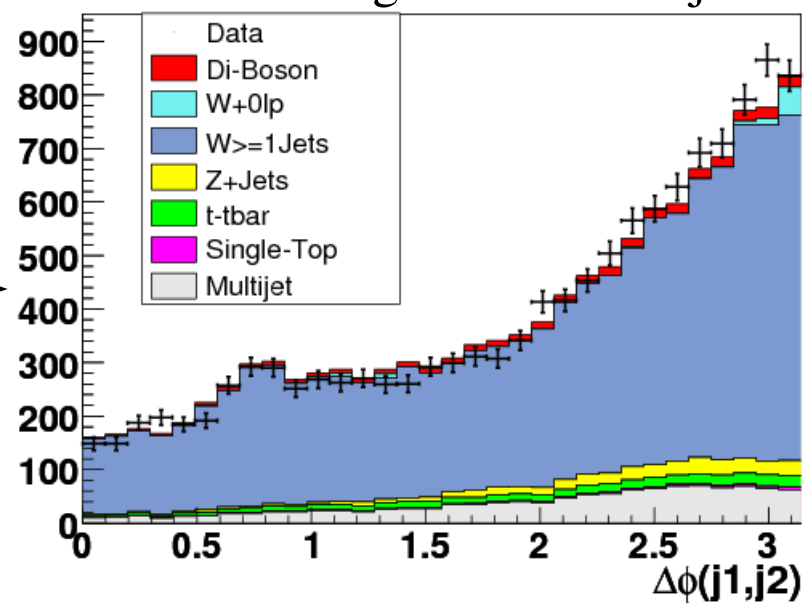
Modeling of Multiple Parton Interactions

- ALPGEN+Pythia MC generated with too many jets from multiple parton interactions (MPI)
  - Jets from more than one interaction in the proton-antiproton collision
    - Result in two jets that are back-to-back in ϕ
    - Mainly affects W+0 jet and W+1 jet events ⇒ too many pass the selection
  - χ^2 fit to data showed the number of selected MPI jets should be 80% less

Before correction

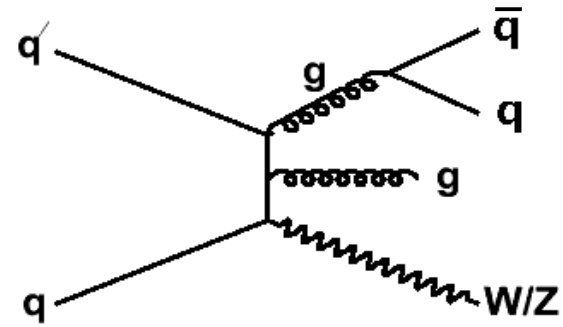


After removing 80% of MPI jets



ALPGEN W+jets Modeling

- Selected sample is dominated by *W*+jets background
  - *W*+jets is ~30 times larger than the signal!
  - Must accurately model *W*+jets at the percent level



- Potential problems with modeling of *W*+jets
  - Angles
    - Jet η and ΔR between jets is not modeled well by ALPGEN
    - To correct these differences we re-weight the MC to agree with data
  - Energies
    - ALPGEN does a fairly good job of modeling energies, however, energy distributions can be affected by ALPGEN generator and matching parameters
    - Studied the effects of these parameters to determine the uncertainty from these parameters
    - Determine best values for agreement with data
  - Cross section
    - The *W*+jets cross section is unconstrained when measuring diboson cross section

ALPGEN Modeling: Angular Corrections

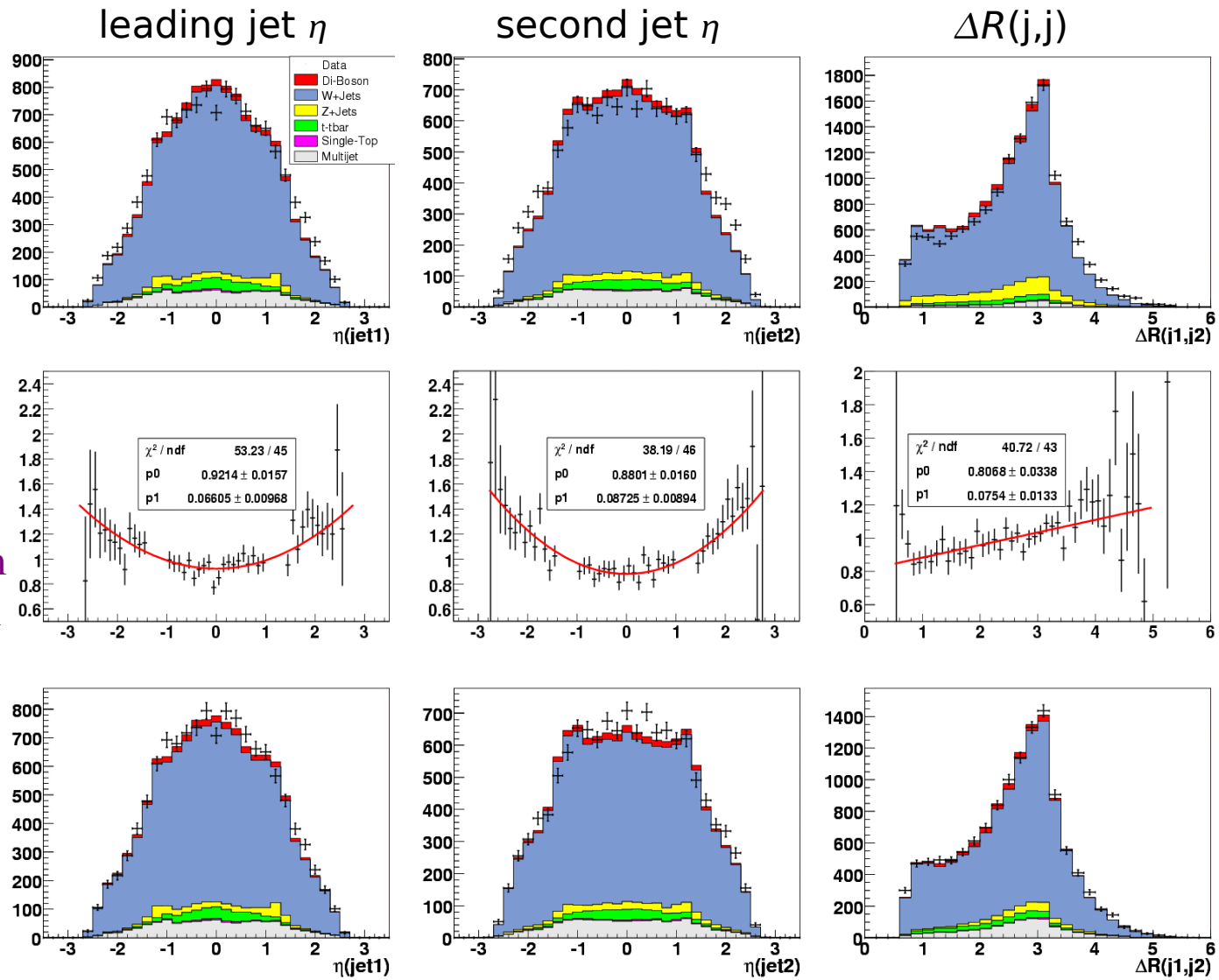
Poor agreement before:

Re-weight *W*+jets MC by a function fit to:

$$\frac{\text{Data} - (\text{non } W\text{+jets MC})}{W\text{+jets MC}}$$

(Include uncertainty on fit from varying the signal cross section from 0-2 × the SM prediction)

Improved agreement after:

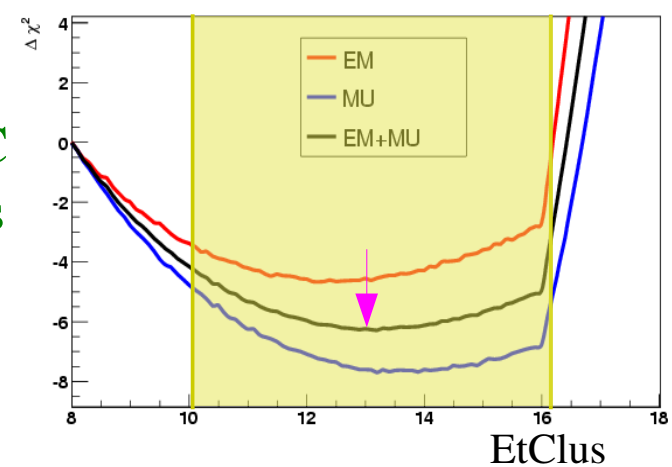


ALPGEN Modeling: Generator Parameters

- After correcting large angular discrepancies we can look at the smaller effects from ALPGEN generator and matching parameters
  - Qfac: Multiplies renormalization scale (Q-scale) by this factor
  - Ktfac: Multiplies parton kt scale (a.k.a. kperp or $k_\perp$) by this factor
  - RClus: Radius for clusters used in ALPGEN parton-jet multiplicity matching
  - **EtClus: E_T threshold for clusters used in parton-jet matching**
    - ALPGEN parton-jet matching required to keep from double counting events

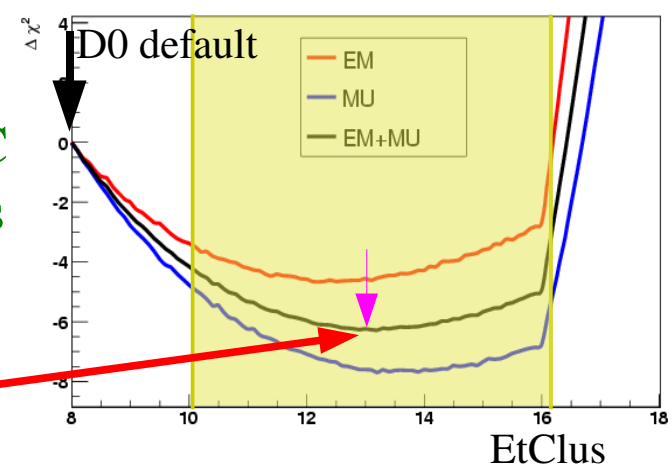
ALPGEN Modeling: Generator Parameters

- After correcting large angular discrepancies we can look at the smaller effects from ALPGEN generator and matching parameters
  - Qfac: Multiplies renormalization scale (Q-scale) by this factor
  - Ktfac: Multiplies parton kt scale (a.k.a. kperp or $k_\perp$) by this factor
  - RClus: Radius for clusters used in ALPGEN parton-jet multiplicity matching
  - **EtClus: E_T threshold for clusters used in parton-jet matching**
    - ALPGEN parton-jet matching required to keep from double counting events
- Remove signal events (55 GeV/c^2 < m_{jj} < 110 Gev/c^2)
- Use p_T of leading jet and recoiling $W \rightarrow l\nu$ to determine parameter values that best model data
  - Measure change in χ^2 between data and MC as the MC shape is morphed by varying these ALPGEN parameters
    - Minimum χ^2 tells the parameter value that best models data
    - $\Delta\chi^2 = 1$ tells the $\pm 1\sigma$ uncertainty range for the parameter



ALPGEN Modeling: Generator Parameters

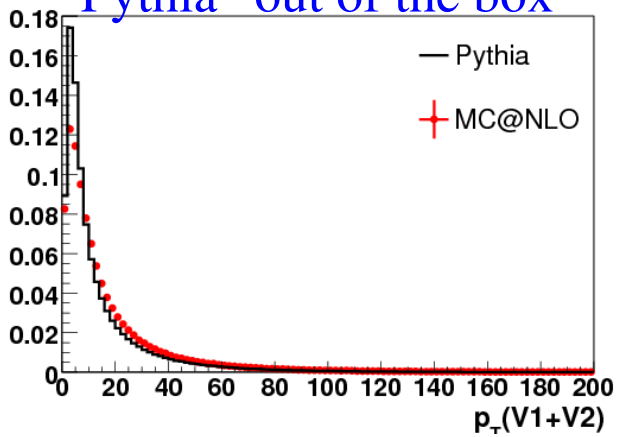
- After correcting large angular discrepancies we can look at the smaller effects from ALPGEN generator and matching parameters
  - Qfac: Multiplies renormalization scale (Q-scale) by this factor
  - Ktfac: Multiplies parton kt scale (a.k.a. kperp or $k_\perp$) by this factor
  - RClus: Radius for clusters used in ALPGEN parton-jet multiplicity matching
  - **EtClus: E_T threshold for clusters used in parton-jet matching**
    - ALPGEN parton-jet matching required to keep from double counting events
- Remove signal events (55 GeV/c^2 < m_{jj} < 110 Gev/c^2)
- Use p_T of leading jet and recoiling $W \rightarrow l\nu$ to determine parameter values that best model data
  - Measure change in χ^2 between data and MC as the MC shape is morphed by varying these ALPGEN parameters
    - Minimum χ^2 tells the parameter value that best models data
    - $\Delta\chi^2 = 1$ tells the $\pm 1\sigma$ uncertainty range for the parameter
- Data modeled best with **EtClus = 13.2 GeV**
  - ALPGEN authors recommend 13 GeV



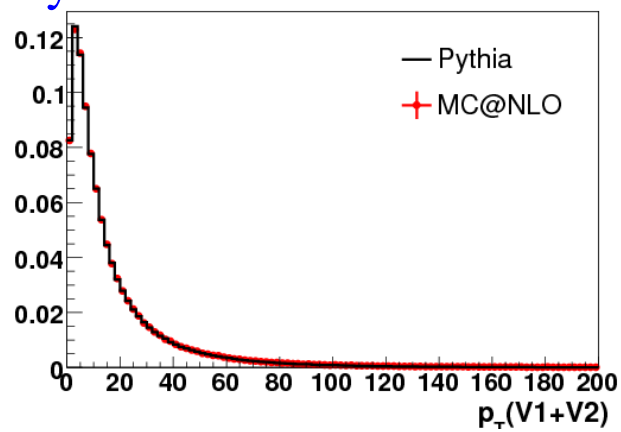
Diboson NLO Correction

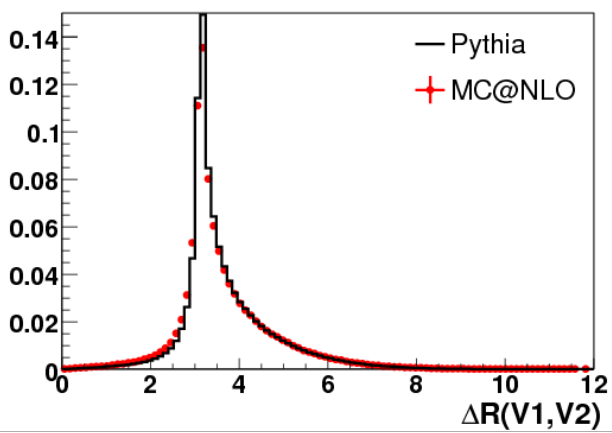
- Diboson signal generated at leading order
  - Use MC@NLO generator to correct signal MC to next-to-leading order
    - Generator level using a two dimensional re-weighting

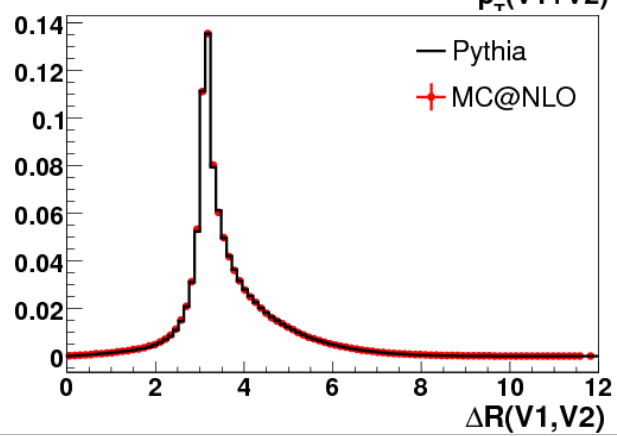
Pythia "out of the box"



Pythia after NLO correction







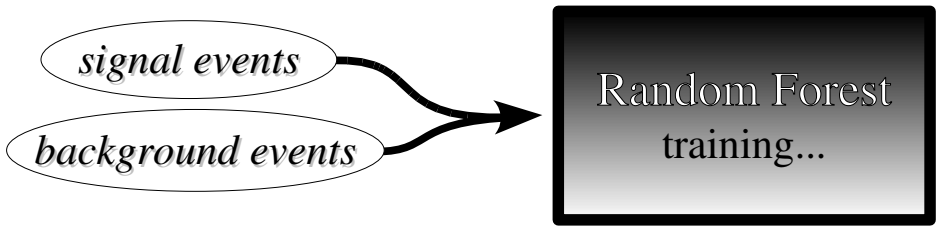
Other Corrections

- Efficiency, reconstruction, resolution differences between data and MC
  - Trigger Efficiencies
    - Data must pass the trigger to be selected ⇒ apply these efficiencies to the MC
  - Jet (Lepton) Energy Scale, Energy Resolution, and Identification
    - MC does not do a perfect job of modeling detector response to jets (and to a much lesser extent, leptons) ⇒ correct energies and apply data/MC scale factors
  - Instantaneous Luminosity Profile
    - MC overlayed with zero-bias data to account for other activity in the detector
    - Luminosity profile for zero-bias not the same as for the data ⇒ re-weight MC events to have the same profile
  - z Position of Primary Vertex
    - Slightly non-Gaussian in data ⇒ re-weight MC to match
  - Muon Track Efficiency
    - Poor MC modeling of muon track with low curvature significance ⇒ remove these events
- ⇒ In total, 28 independent systematic uncertainties

Multivariate Discrimination

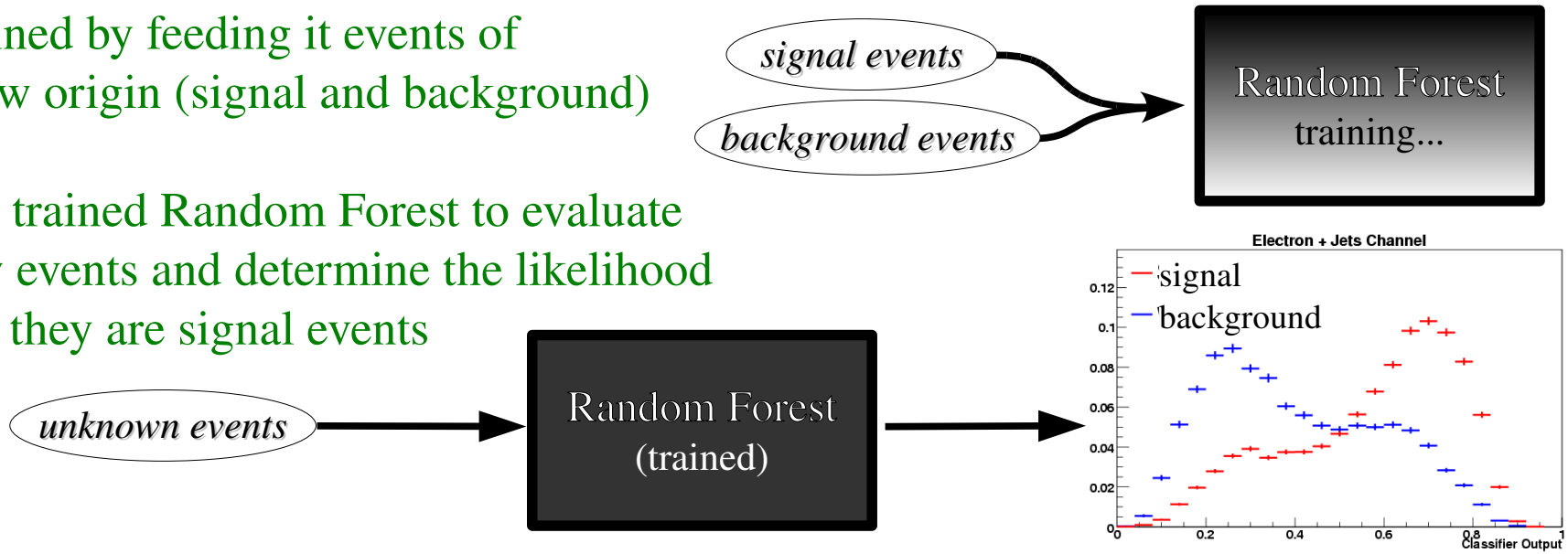
Multivariate Discrimination

- Use a multivariate classifier to improve separation of signal and background
  - Many options: Neural Networks, Decision Trees, Matrix Elements Method, etc.
  - Random Forest classifier powerful and robust
    - Similar technology as boosted decision trees used in single-top analyises
- From the outside (black box), the Random Forest works similar to other multivariate classifiers
  - Trained by feeding it events of know origin (signal and background)



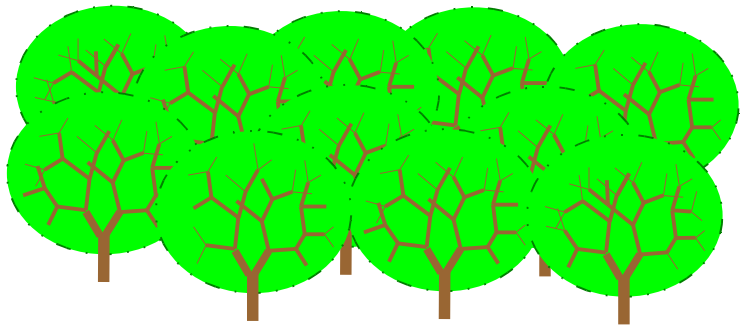
Multivariate Discrimination

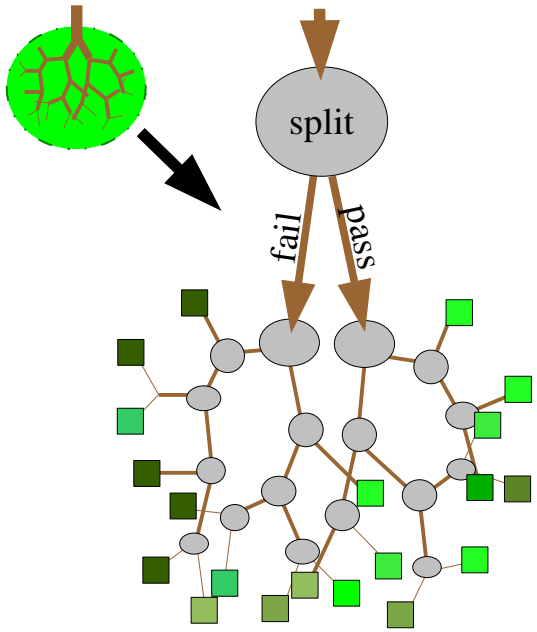
- Use a multivariate classifier to improve separation of signal and background
  - Many options: Neural Networks, Decision Trees, Matrix Elements Method, etc.
  - Random Forest classifier powerful and robust
    - Similar technology as boosted decision trees used in single-top analyses
- From the outside (black box), the Random Forest works similar to other multivariate classifiers
  - Trained by feeding it events of know origin (signal and background)
  - Use trained Random Forest to evaluate new events and determine the likelihood that they are signal events



Inside the Random Forest

A "forest" of many decision tree classifiers



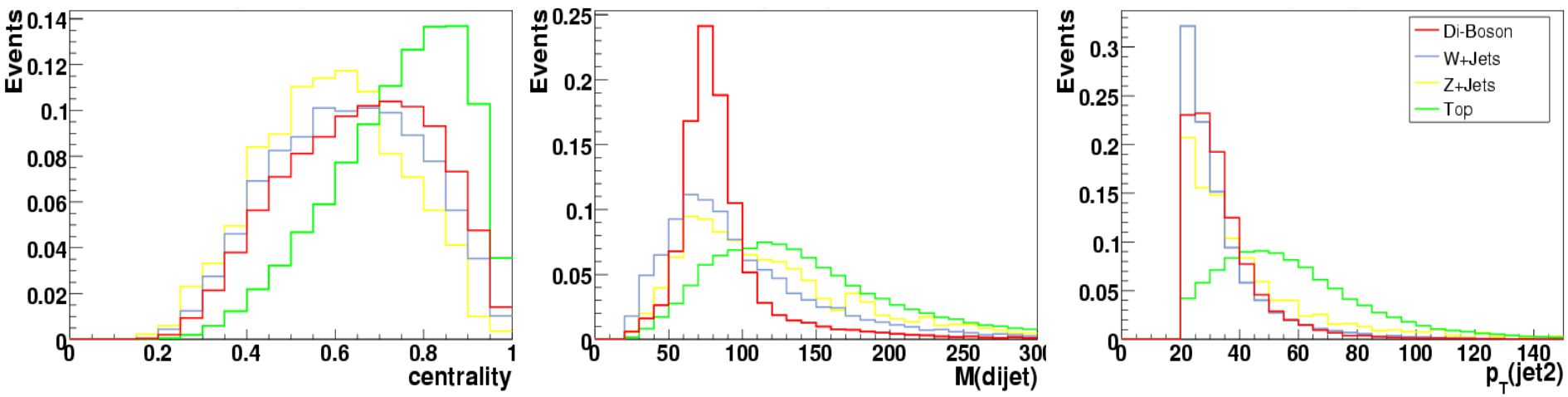


- Each tree is independent of the other trees
  - Each uses a random subset of input variables
    - Allows each tree to focus on a different subset of kinematics and correlations
  - Each is trained with a random subset of training events
    - Provides protection against over-training and high-weight events
- The Random Forest output averages the output from all the trees
  - Fluctuations and over-training that occur for a single decision tree are reduced because each tree will fluctuate differently

Multivariate Discrimination

- Choosing input variables to the Random Forest
  - Variables that are well modeled by the MC
    - χ^2 probability between data and MC greater than 5%
  - Variables that differ between signal and at least one background
    - However, well-modeled uninformative variables will not degrade performance

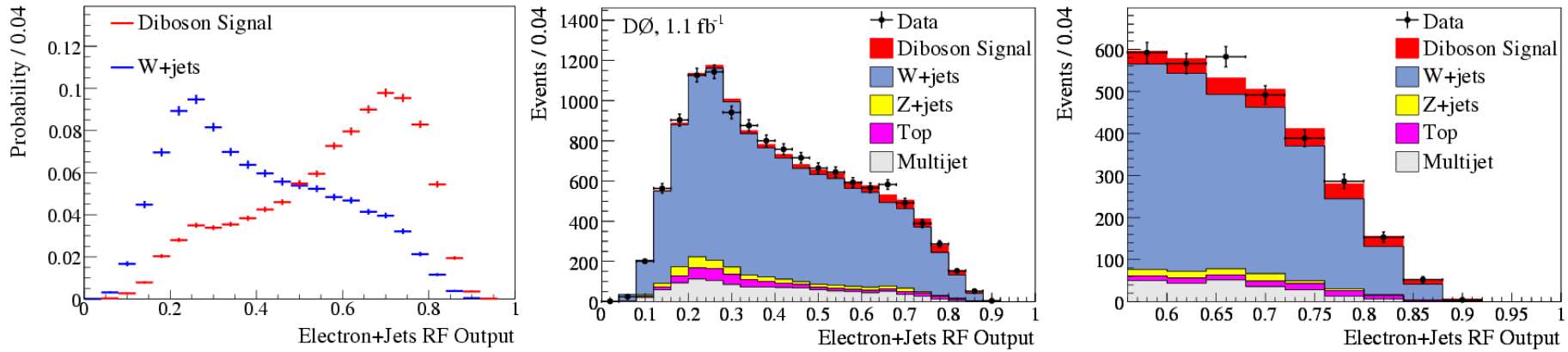
⇒ Came up with 13 input variables



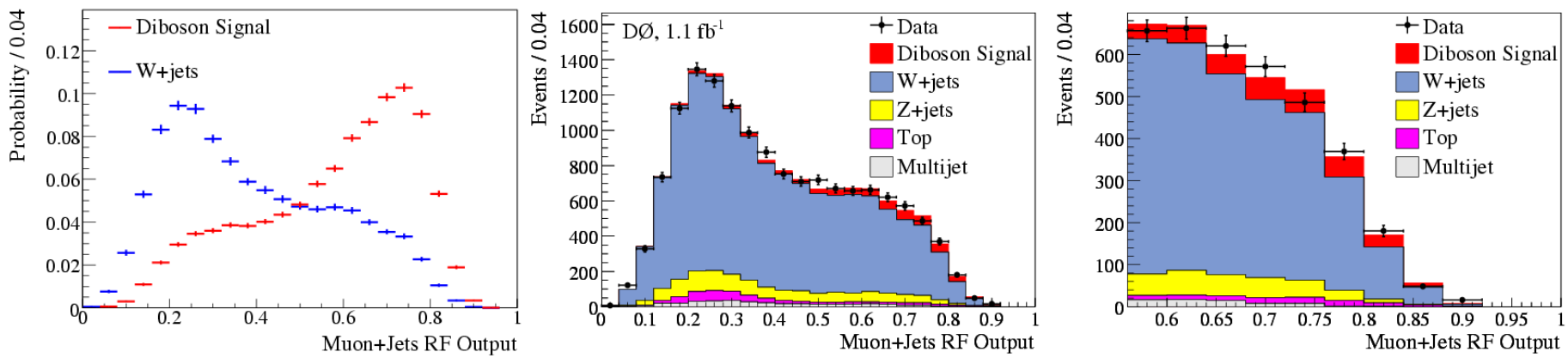
Multivariate Discrimination

- Output from the Random Forest
  - Discriminating variable with improved signal and background separation

$e\nu qq$ channel



$\mu\nu qq$ channel



The Measurement

Cross Section Measurement

- Perform "best fit" of distributions for signal and each background to the data
  - Predicted number of events, p_i, varied within the systematic uncertainties $\sigma_{k,i}$

$$p'_i = p_i \prod_k^{N_{\rm sys}} (1 + R_k \sigma_{k,i})$$

  - Minimizes the negative-log-likelihood between data and prediction

$$\mathrm{NLL} = -2 \ln \mathcal{L} = 2 \sum_i^{N_{\rm bins}} \left[p'_i - d_i - d_i \ln \left(\frac{p'_i}{d_i} \right) \right] + \sum_k^{N_{\rm sys}} (R_k)^2$$

Cross Section Measurement

- Perform "best fit" of distributions for signal and each background to the data
  - Predicted number of events, p_i, varied within the systematic uncertainties $\sigma_{k,i}$

$$p'_i = p_i \prod_k^{N_{\rm sys}} (1 + R_k \sigma_{k,i})$$

  - Minimizes the negative-log-likelihood between data and prediction

$$\mathrm{NLL} = -2 \ln \mathcal{L} = \boxed{2 \sum_i^{N_{\rm bins}} \left[p'_i - d_i - d_i \ln \left(\frac{p'_i}{d_i} \right) \right]} + \boxed{\sum_k^{N_{\rm sys}} (R_k)^2}$$

Ratio of Poisson likelihoods

Gaussian prior constraint for systematics

Cross Section Measurement

- Perform "best fit" of distributions for signal and each background to the data
  - Predicted number of events, p_i, varied within the systematic uncertainties $\sigma_{k,i}$

$$p'_i = p_i \prod_k^{N_{\rm sys}} (1 + R_k \sigma_{k,i})$$

  - Minimizes the negative-log-likelihood between data and prediction

$$\mathrm{NLL} = -2 \ln \mathcal{L} = \boxed{2 \sum_i^{N_{\rm bins}} \left[p'_i - d_i - d_i \ln \left(\frac{p'_i}{d_i} \right) \right]} + \boxed{\sum_k^{N_{\rm sys}} (R_k)^2}$$

Ratio of Poisson likelihoods

Gaussian prior constraint for systematics

  - Systematics "floated" by removing the $(R_k)^2$ term ⇒ free fit parameters
    - Float W+jets normalization ⇒ determined by sideband regions
    - Float signal cross section ⇒ fit the excess (over background) that is consistent with kinematics of WW and WZ production

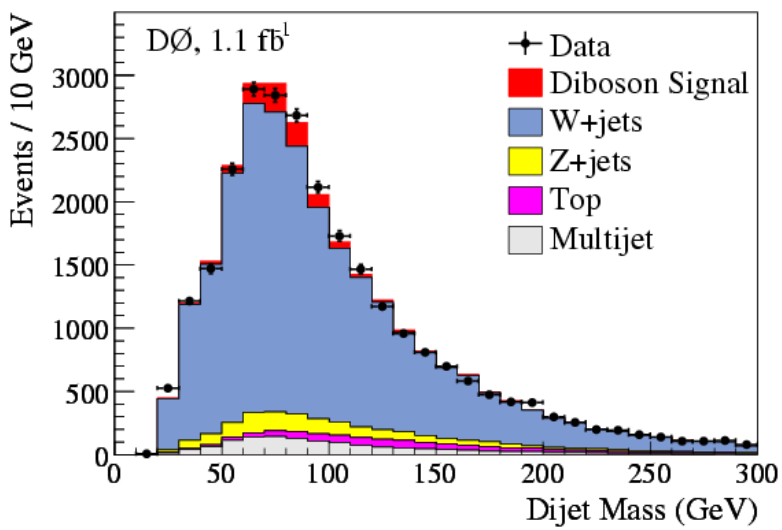
Cross Section Measurement

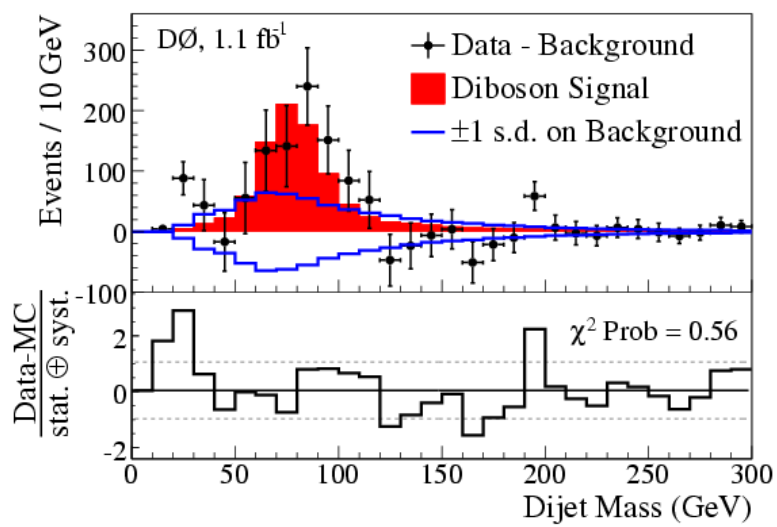
- Measured cross section using the **dijet mass**

  ⇒ $\sigma(WW+WZ)$ = **18.5 ± 2.8(stat) ± 4.9(syst) ± 1.1(lumi) pb**

(SM prediction: $\sigma(WW+WZ)$ = 16.1 ± 0.9 pb)

Dijet mass after best-fit using the dijet mass





Cross Section Measurement

- Measured cross section using the **dijet mass**

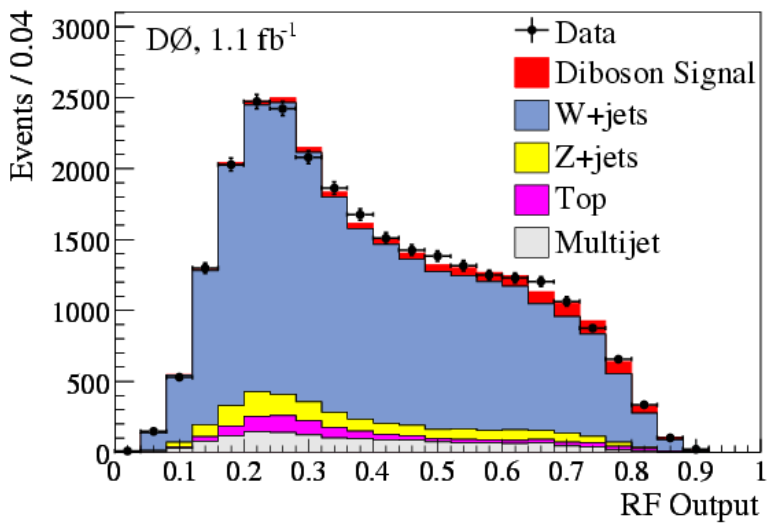
  ⇒ $\sigma(WW+WZ)$ = **18.5 ± 2.8(stat) ± 4.9(syst) ± 1.1(lumi) pb**

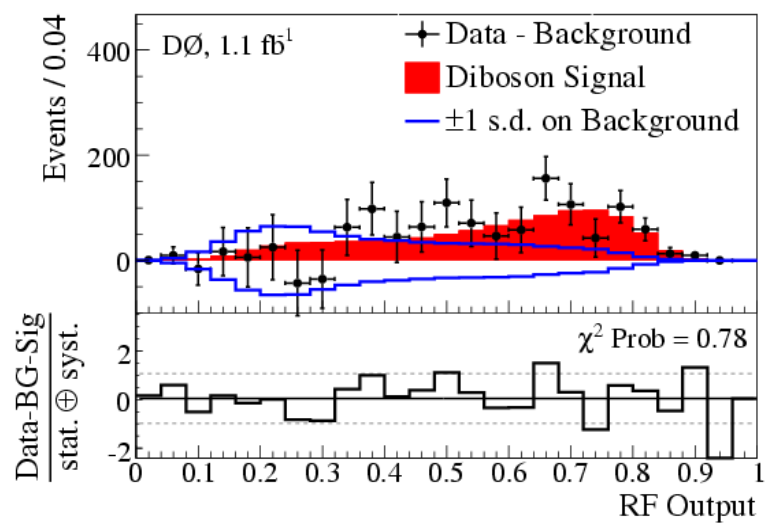
- Measured cross section using the **Random Forest**

  ⇒ $\sigma(WW+WZ)$ = **20.2 ± 2.5(stat) ± 3.6(syst) ± 1.2(lumi) pb**

  (SM prediction: $\sigma(WW+WZ)$ = 16.1 ± 0.9 pb)

Random Forest after best-fit using the Random Forest





Significance

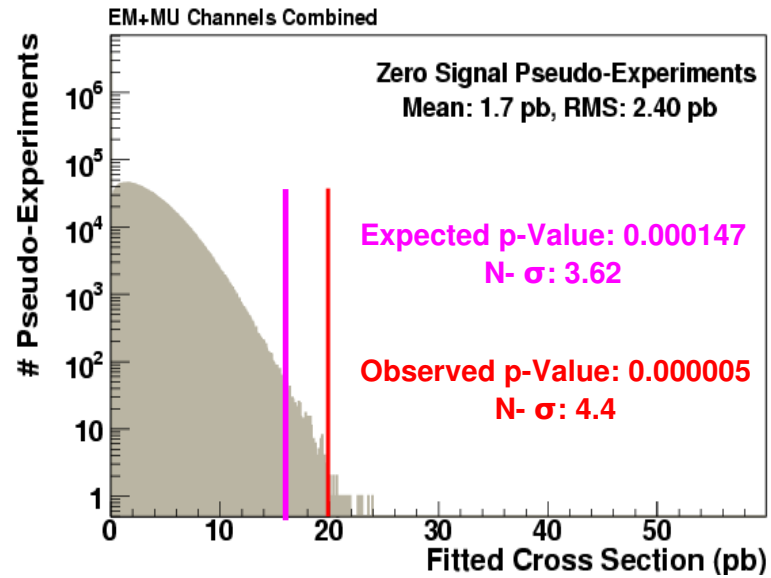
- Estimation of significance
- Frequentist approach: count fraction of outcomes for which the background-only scenario yields a measured cross section as large as we observe (or expect)
  - Generate pseudo-data for background-only scenario
    - randomly sample the systematics from their priors, then drawing Poisson trials for each bin
  - Fit the background-only pseudo-data outcomes in the same was as real data

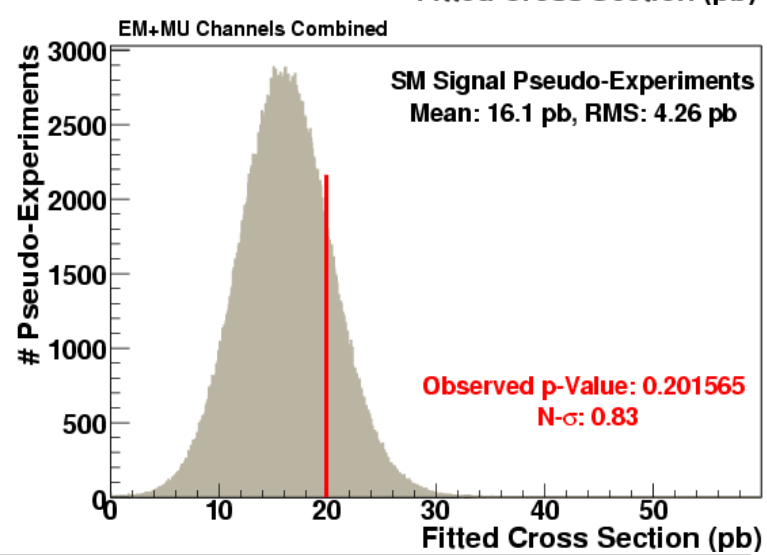
⇒ Expected significance: 3.6σ (p-value = $1.5 \cdot 10^{-4}$)

⇒ Observed significance: 4.4σ (p-value = $5.4 \cdot 10^{-6}$)

- Compatibility with the SM
- Same approach as calculating the significance, except with pseudo-data for the signal+background scenario

⇒ 0.83σ above SM prediction





Significance

- Estimation of significance
- Frequentist approach: count fraction of outcomes for which the background-only scenario yields a measured cross section as large as we observe (or expect)
  - Generate pseudo-data for background-only scenario
    - randomly sample the systematics from their priors, then drawing Poisson trials for each bin
  - Fit the background-only pseudo-data outcomes in the same was as real data

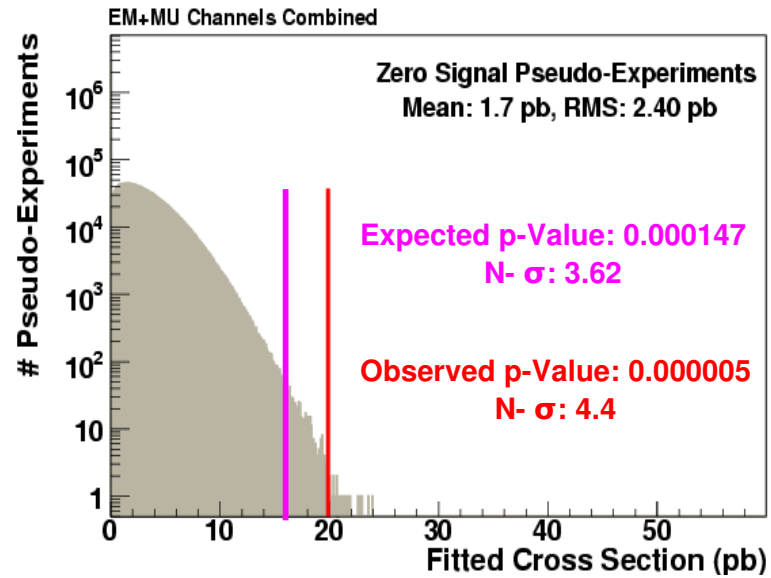
⇒ Expected significance: 3.6σ (p-value = $1.5 \cdot 10^{-4}$)

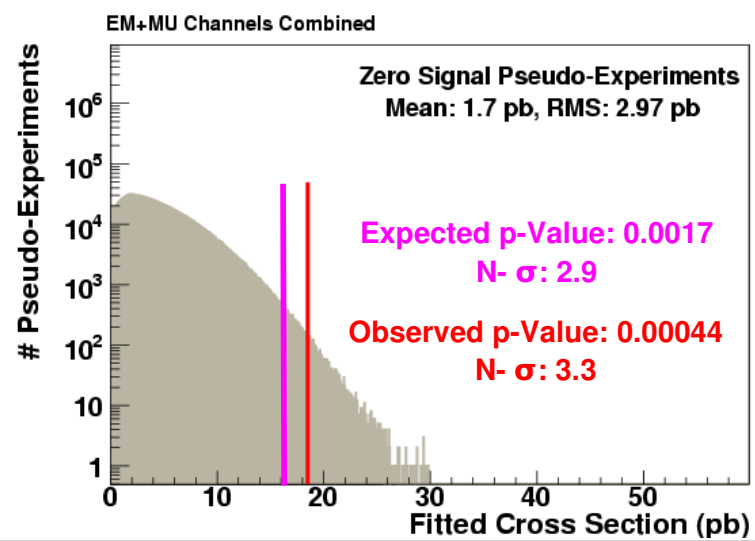
⇒ Observed significance: 4.4σ (p-value = $5.4 \cdot 10^{-6}$)

- Significance using only dijet mass:

⇒ Expected significance: 2.9σ (p-value = $1.7 \cdot 10^{-3}$)

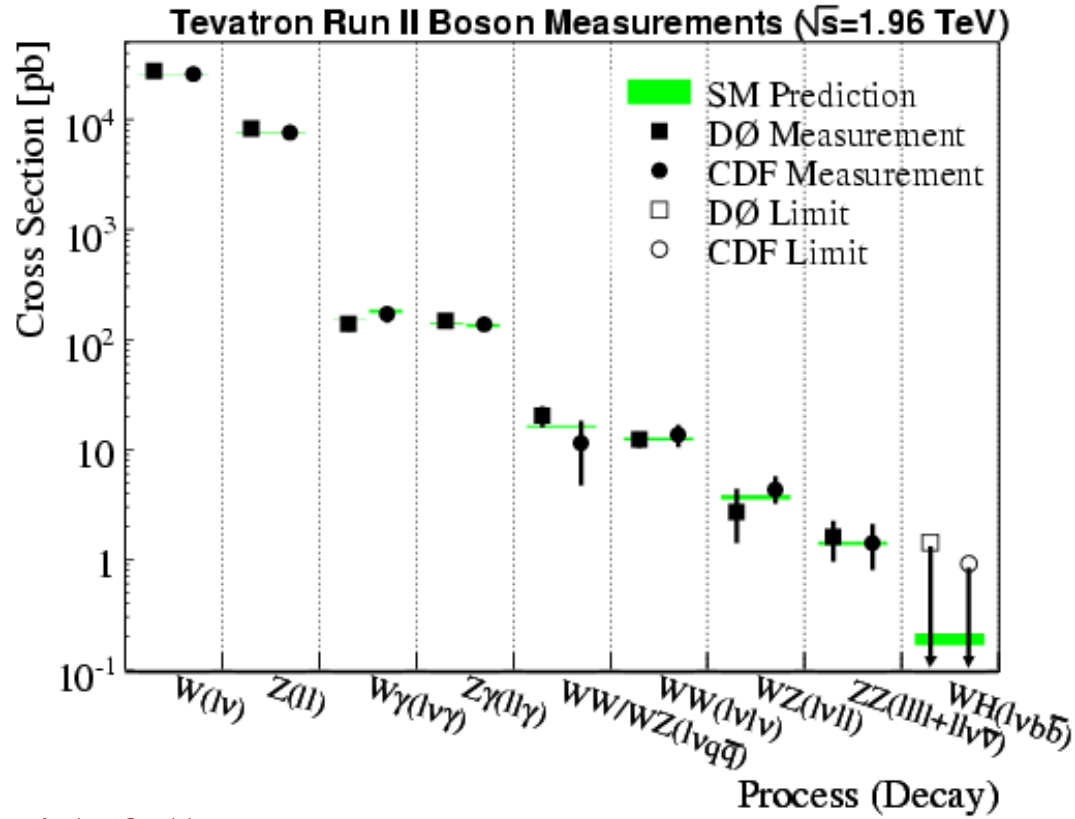
⇒ Observed significance: 3.3σ (p-value = $4.4 \cdot 10^{-4}$)





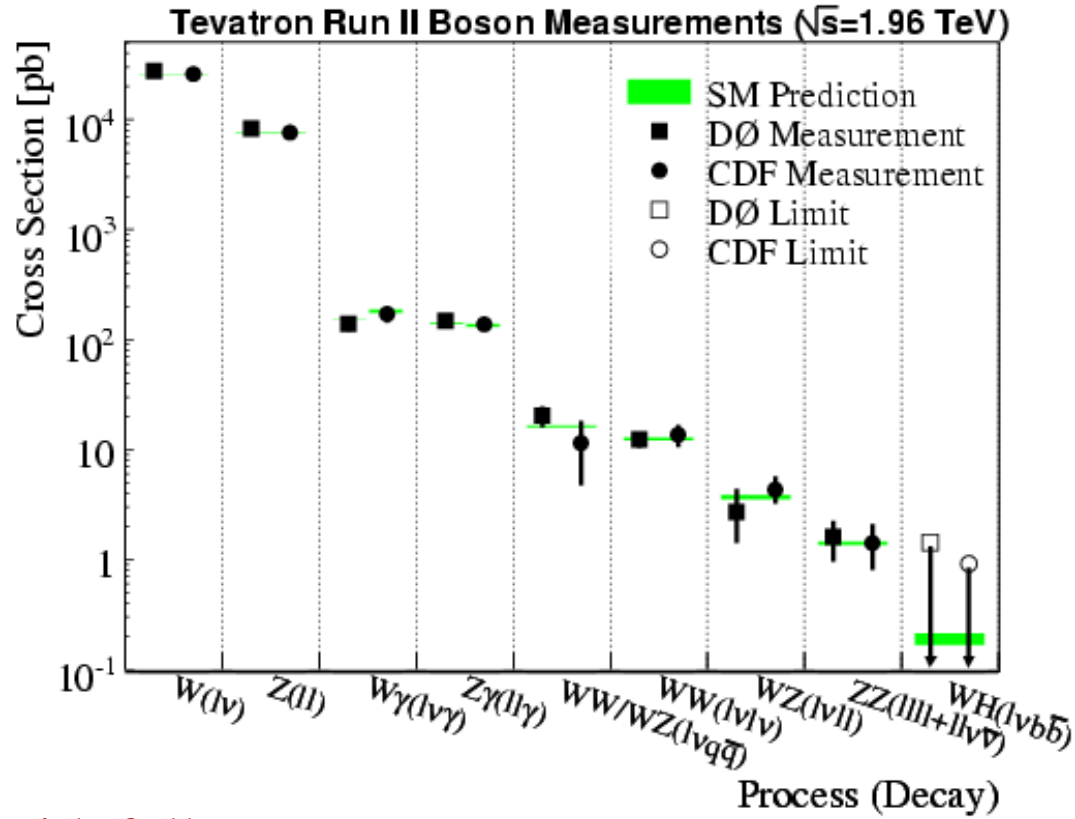
Summary

- First semi-leptonic diboson measurement at hadron collider
  - $\sigma(p\bar{p} \rightarrow WW+WZ) = 20.2 \pm 4.5$ pb
  - Observed significance: **4.4σ**
  - Accepted by PRL 3/25/09
- Only benchmark for low mass Tevatron Higgs analyses
  - First hadronic W mass outside of top quark events
- Direct constraint on dominant $H \rightarrow WW$ background
- Complimentary to and consistent with fully leptonic measurements
  - And more precise than previous leptonic measurements



Summary

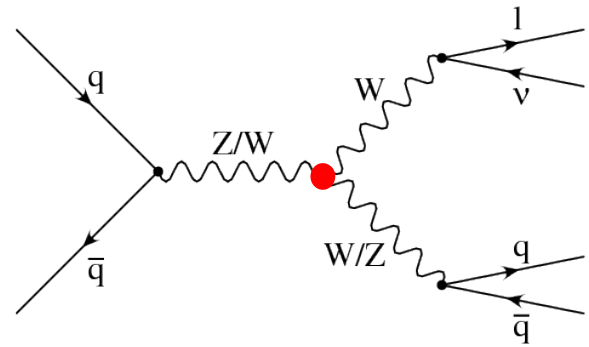
- First semi-leptonic diboson measurement at hadron collider
  - $\sigma(p\bar{p} \rightarrow WW+WZ) = 20.2 \pm 4.5$ pb
  - Observed significance: **4.4σ**
  - Accepted by PRL 3/25/09
- Only benchmark for low mass Tevatron Higgs analyses
  - First hadronic W mass outside of top quark events
- Direct constraint on dominant $H \rightarrow WW$ background
- Complimentary to and consistent with fully leptonic measurements
  - And more precise than previous leptonic measurements



...what's next?

Continuing Work

- Limits on Anomalous Couplings
  - Compare the dijet p_T (most sensitive variable) between data and the predictions for different anomalous couplings
    - Results soon!



- Updating measurement to full Run II dataset (>4 fb^{-1})
  - Should achieve >5σ significance ⇒ observation
  - Better (best Tevatron) limits on anomalous couplings
- Measure Z→bb Resonance
  - Apply b-tagging and measure contribution from only $WZ \to l\nu bb$
  - Also need to add $ZZ \to llbb$ channels
    - Expect sensitivity around 3σ
  - Analysis identical to $WH \to l\nu bb$ and $ZH \to llbb$, but 4 × more signal
    - Would be the first evidence of a bb resonance at the Tevatron

thank you

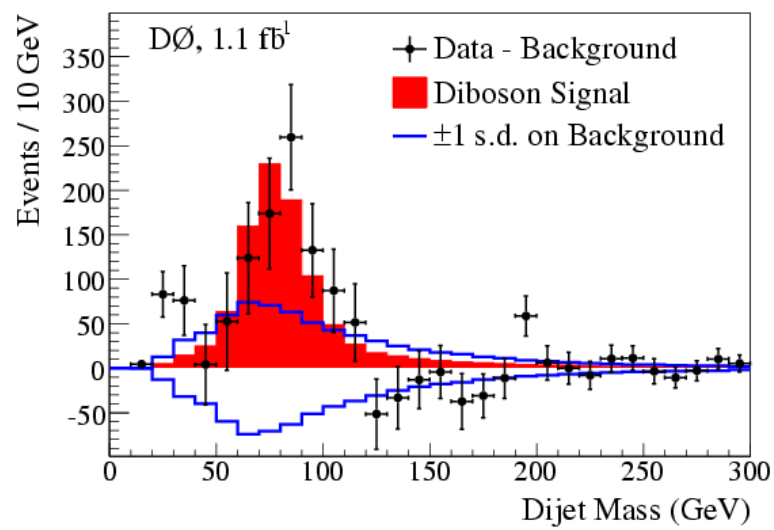
Backup Slides

Comparison of Measurements

- Measuring *WW*+*WZ* cross section
  - Using the **dijet mass**:
    $\sigma(WW+WZ) = 18.5 \pm 2.8(\text{stat}) \pm 4.9(\text{syst}) \pm 1.1(\text{lumi})$ pb
  - Using the **random forest**:
    $\sigma(WW+WZ) = 20.2 \pm 2.5(\text{stat}) \pm 3.6(\text{syst}) \pm 1.2(\text{lumi})$ pb

- Significance
  - Using the **dijet mass**:
    Expected = 2.9σ
    Observed = 3.3σ
  - Using the **random forest**:
    Expected = 3.6σ
    Observed = 4.4σ

Dijet mass distribution
after best-fit of Random Forest



Summary of Systematics

Percent change from nominal distribution for Randon Forest distribution.

Source of systematic uncertainty	Diboson	W+jets	Z+jets	Top	Multijet	Nature
Trigger efficiency, $e\nu q\bar{q}$	+2/-3	+2/-3	+2/-3	+2/-3		N
Trigger efficiency, $\mu\nu q\bar{q}$	+0/-5	+0/-5	+0/-5	+0/-5		S
Lepton identification	±4	±4	±4	±4		N
Jet identification	±1	±1	±1	$\pm <1$		S
Jet energy scale	±4	±9	±9	±4		S
Jet energy resolution	±3	±4	±4	±4		N
Cross section		$\pm20^a$	±6	±10		N
Heavy flavor jet contribution		±20				N
Diboson NLO correction	±10					S
PDF set	±1	±1	±1	±1		S
ALPGEN jet η and ΔR corrections		±1	±1			S
Renormalization and factorization scale		±3	±3			S
ALPGEN parton-jet matching parameters		±4	±4			S
Multijet normalization, $e\nu q\bar{q}$					±20	N
Multijet normalization, $\mu\nu q\bar{q}$					±30	N
Multijet shape, $e\nu q\bar{q}$					±6	S
Multijet shape, $\mu\nu q\bar{q}$					±10	S

Summary of Systematics

Contribution of each systematic to the total systematic uncertainty of 3.6 pb on measured cross section using the Random Forest classifier.

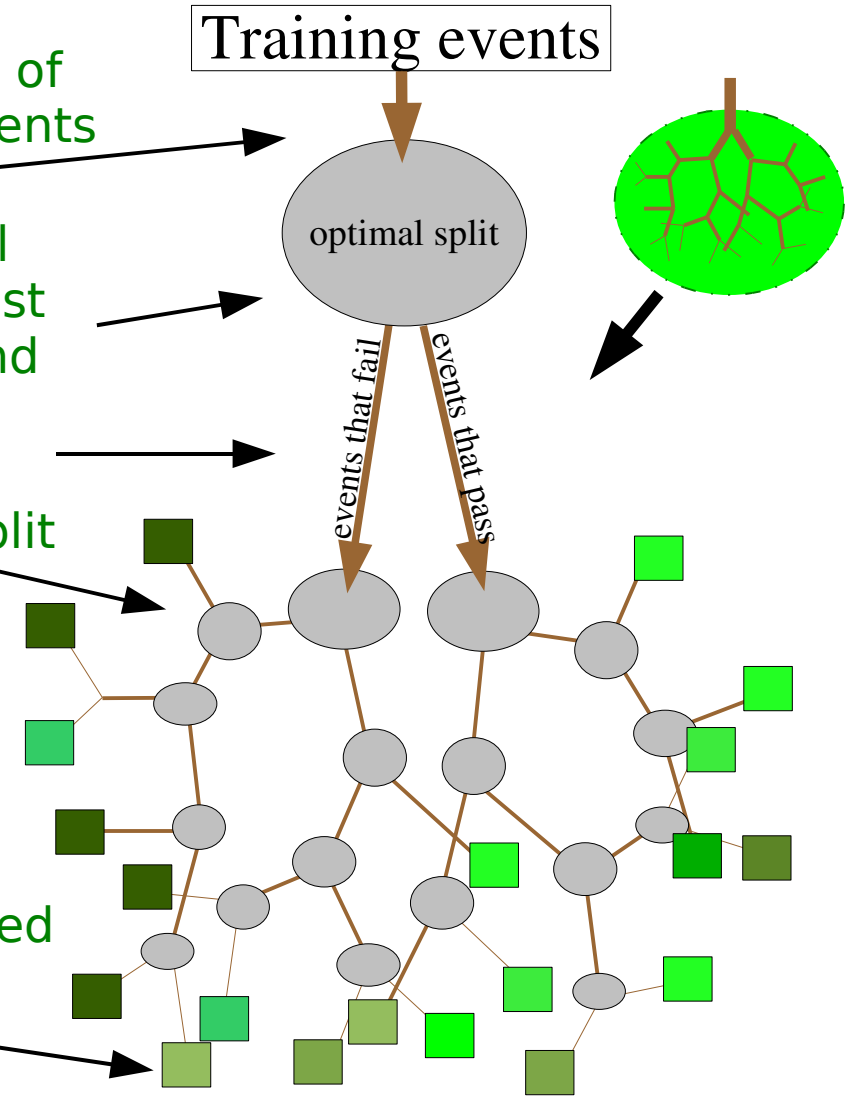
Source of systematic uncertainty	$\Delta\sigma$ [pb]
Trigger efficiency, $e\nu q\bar{q}$	< 0.1
Trigger efficiency, $\mu\nu q\bar{q}$	< 0.1
Lepton identification	< 0.1
Jet identification	0.3
Jet energy scale	1.9
Jet energy resolution	< 0.1
Cross section	1.1
Heavy flavor jet contribution	< 0.1
Diboson NLO correction	< 0.1
PDF set	0.2
ALPGEN jet η and ΔR corrections	< 0.1
Renormalization and factorization scale	0.9
ALPGEN parton-jet matching parameters	2.4
Multijet normalization, $e\nu q\bar{q}$	0.9
Multijet normalization, $\mu\nu q\bar{q}$	0.5
Multijet shape, $e\nu q\bar{q}$	< 0.1
Multijet shape, $\mu\nu q\bar{q}$	< 0.1

Making the Measurement

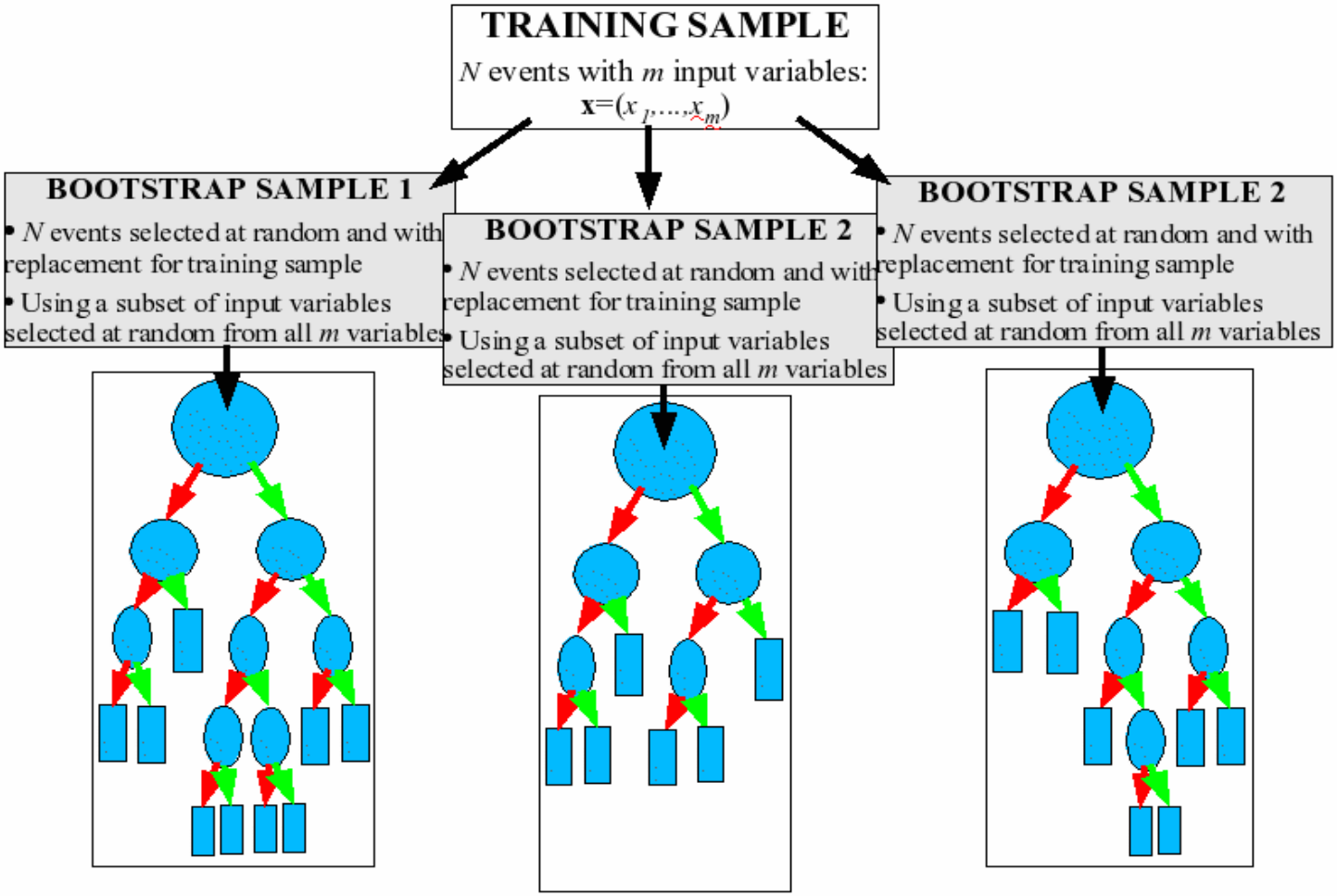
- Now that an adequate MC description of the data has been established
  - The random forest classifier should provide improved separation of signal events
    - Improved signal significance compared to individual variables, such as dijet mass
- The next step is to extract the signal
  - Cross section
    - Measure excess over background that is consistent with diboson production
  - Significance
    - Determine the probability the measured signal is really just a fluctuation in the background

Decision Tree Classifier

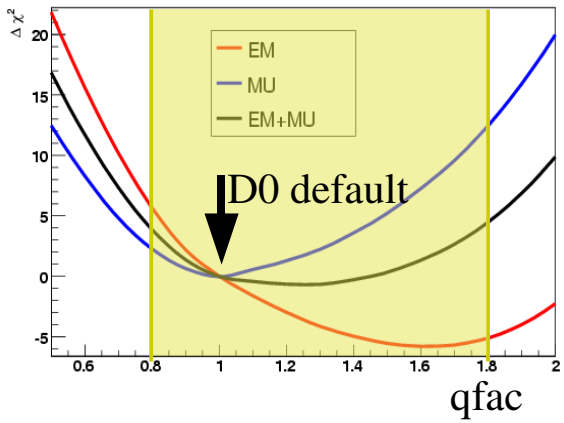
- Decision tree is trained/grown using a set of known signal and background training events ⇒ These events go into the root node
- Algorithm looks at all possible splits on all input variables and applies split giving best separation between signal and background (using Gini index)
- Events pass to one of two child nodes depending on whether they pass of fail split
- This process is repeated for each child node until
  - a node contains all or no signal events
  - the number of events per node is less than a some specified amount (must be optimized for each analysis)
- Output for an unknown event is determined by the signal purity of the terminal node that the event ends up in

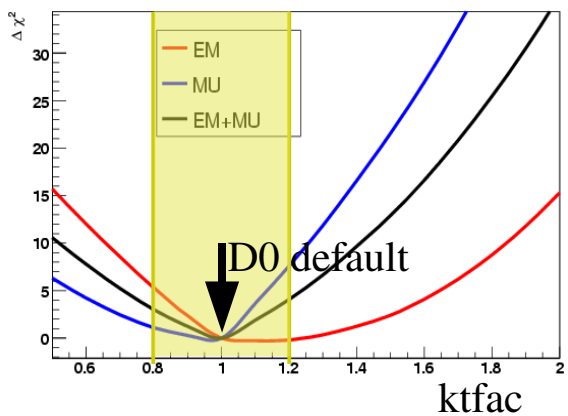


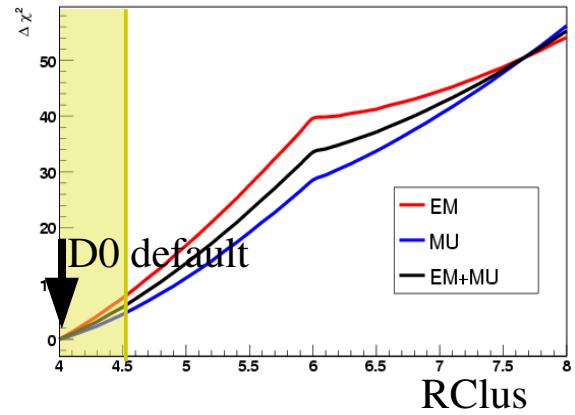
Random Forest Classifier



ALPGEN Modeling: Generator Parameters

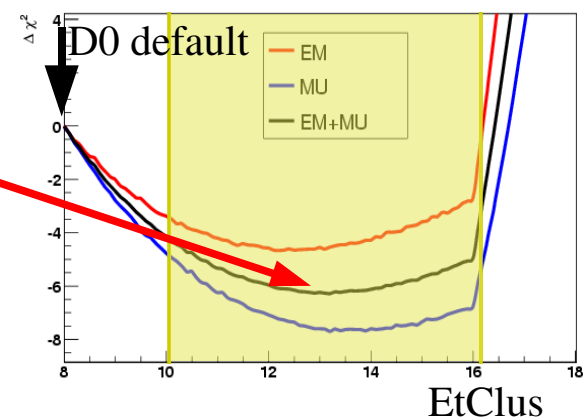




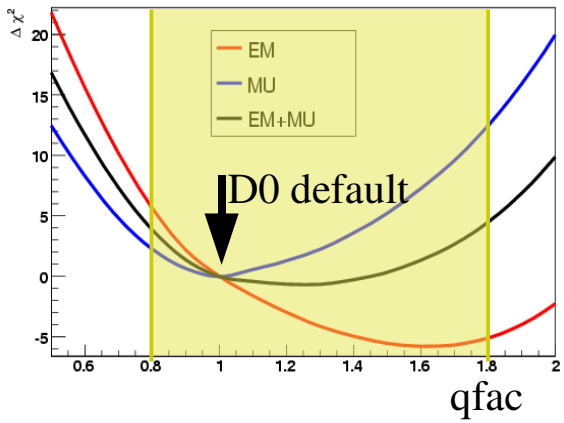


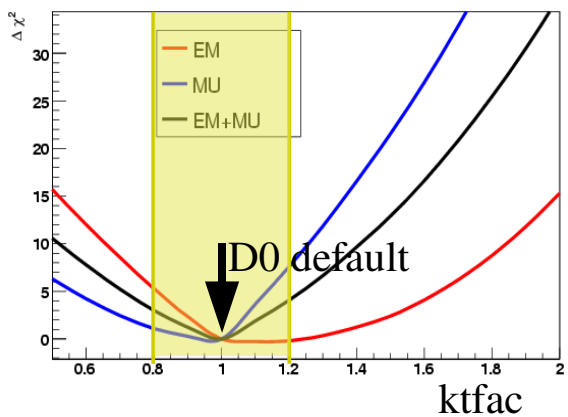
- Best values for qfac, ktfac, and RClus are consistent with default values used by D0
- EtClus models the data best for a value of 13.2 GeV
  - Value recommended by Alpgen authors is 13 GeV

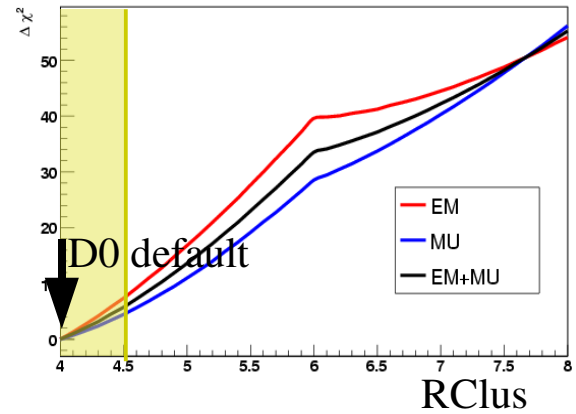
  ⇒ Re-weight Alpgen V+jets MC to EtClus = 13.2 GeV



ALPGEN Modeling: Generator Parameters

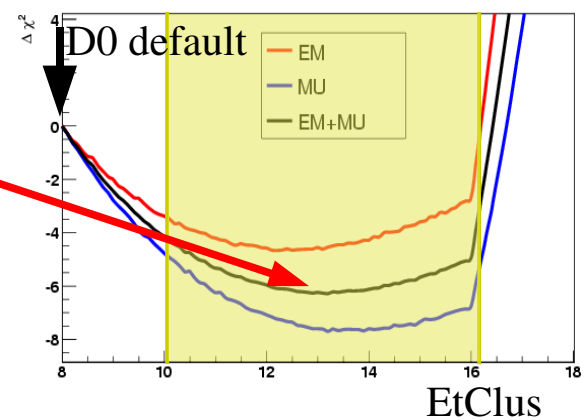




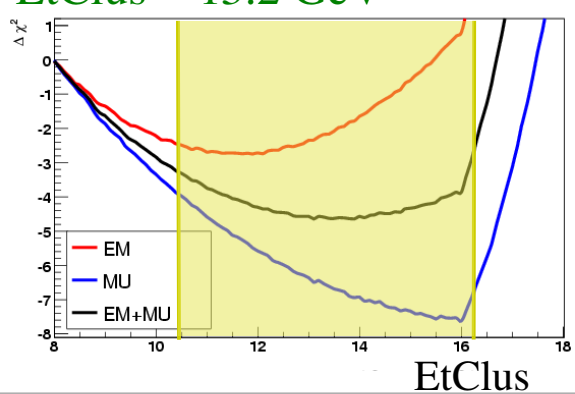


- Best values for qfac, ktfac, and RClus are consistent with default values used by D0
- EtClus models the data best for a value of 13.2 GeV
  - Value recommended by Alpgen authors is 13 GeV

  ⇒ Re-weight Alpgen V+jets MC to EtClus = 13.2 GeV

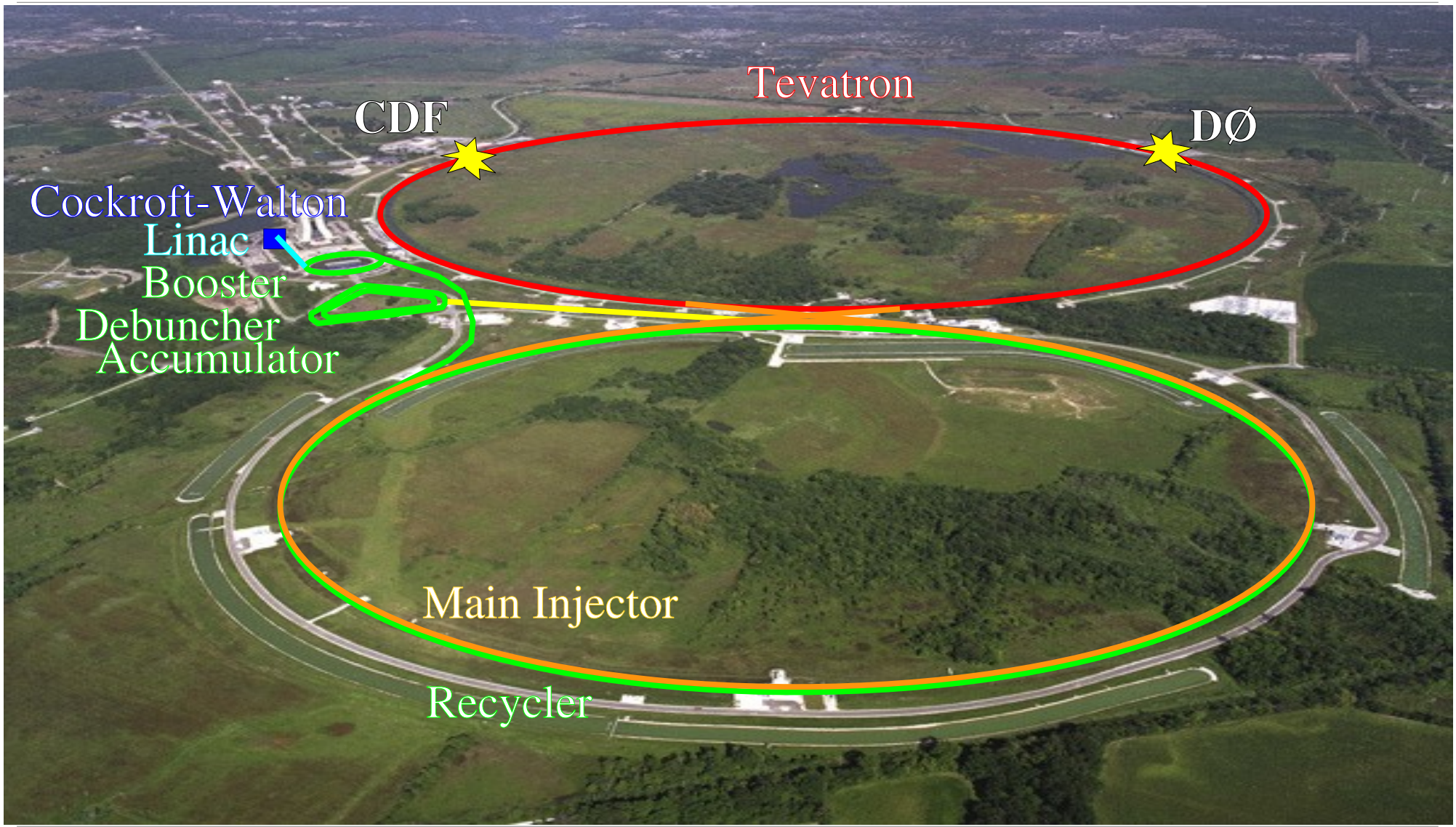
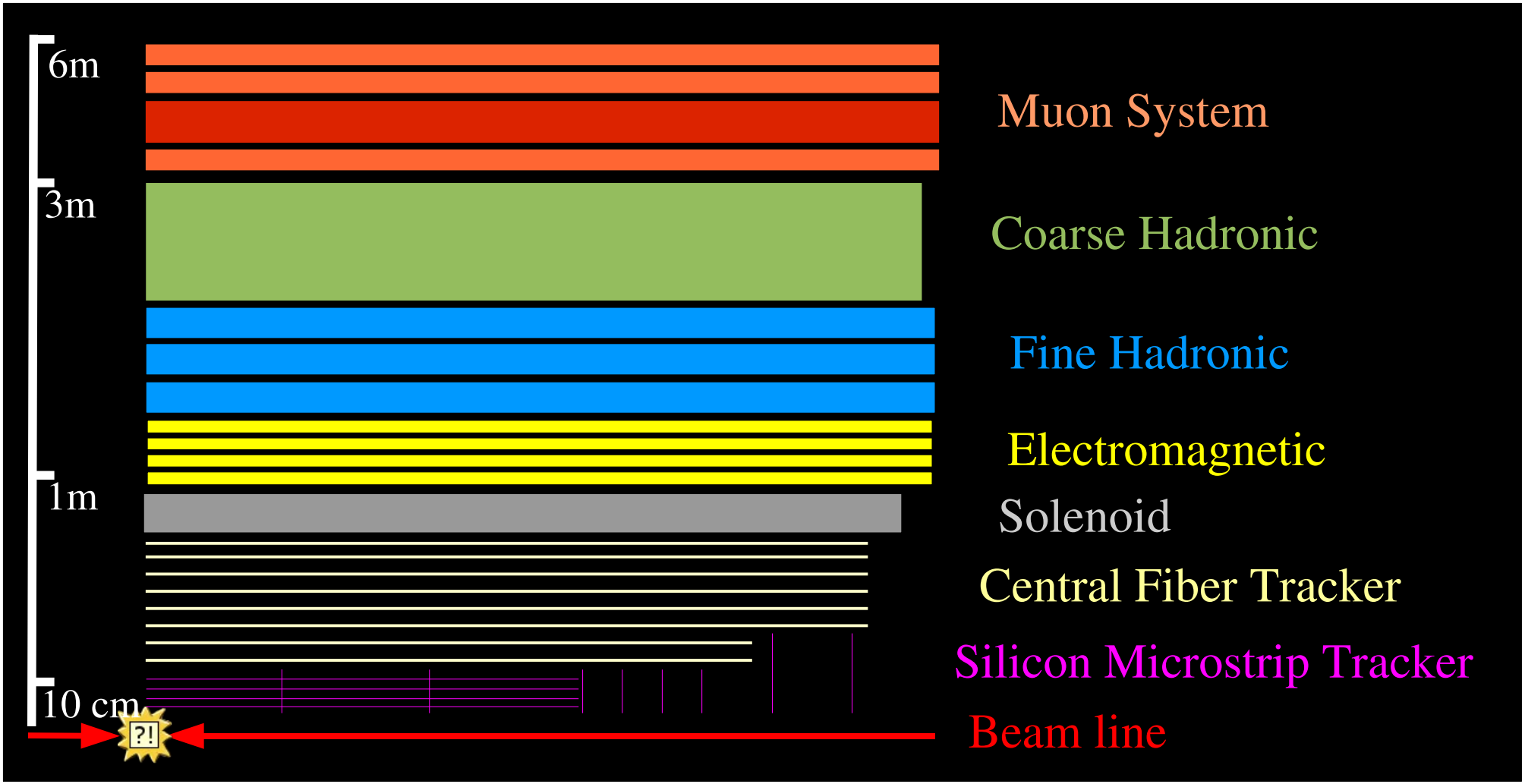


- Dijet mass shows similar behavior for EtClus (and other parameters)



The D0 Collaboration

The Accelerator Chain



Event Reconstruction