Jets produced in association with W-bosons in CMS at the LHC

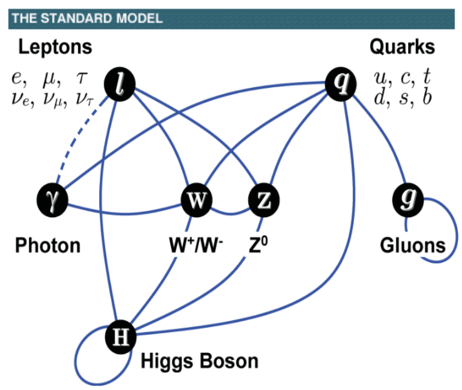
Kira Grogg
UW-Madison
Ph.D. Defense
20 July 2011

Outline

- Introduction
  - Standard Model
  - Importance of W+jets
- Experiment
  - Large Hadron Collider
  - Compact Muon Solenoid
    - Tracker
    - Calorimeters
    - Trigger
- Monte Carlo Simulation
- Reconstruction
  - Electrons, E_T^{miss}, jets
- W+jets analysis
  - Samples
  - Selection
  - Efficiency
  - Data-MC comparisons
  - Signal Extraction
  - Unfolding
- Results
- Summary/Outlook

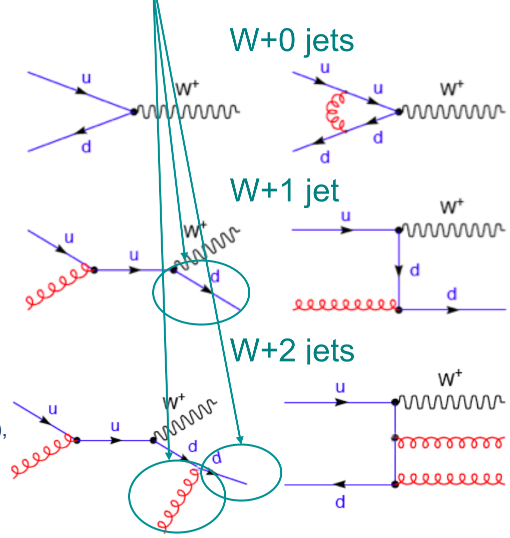
The Standard Model

- Fundamental particles:
  - Fermions (matter)
    - Electron, muon, tau, corresponding neutrinos
    - up, down, charm, strange, top, bottom quarks
      - Combine into hadrons
  - Bosons (force carriers)
    - Photon (EM)
    - W, Z (EW)
    - Gluon (Strong)
  - Higgs? (source of EWK symmetry breaking and mass)



Importance of W + Jets as Signal

- Measure of Electroweak Interaction at much higher energies
- Test of perturbative QCD calculations
  - Verification of theoretical cross-section and parton distribution functions (PDFs)
- Goal: measure the rate of events with jets and a W boson decaying to electron and neutrino
  - Inclusive rate of n jets (i.e., ≥ n jet), not corrected for acceptance
  - Starting with ratio measurements where systematics uncertainties partially cancel



Perturbative QCD (pQCD) and NLO

- QCD involves the strong force
  - Difficult to calculate cross sections exactly
  - Strong coupling α_s increases with distance
- pQCD is possible at high momentum transfer (large Q^2) and small distances ➔ α_s is small $\alpha_s(Q^2) \propto 1/\ln(Q^2/\Lambda^2_{QCD})$
  - Q^2 is large for W+jets events
  - Can use perturbation and expand calculation in different orders of α_s
    - $A = A_0 + \alpha_s{}^1 A_1 + \alpha_s{}^2 A_2 + \alpha_s{}^3 A_3 + \ldots$

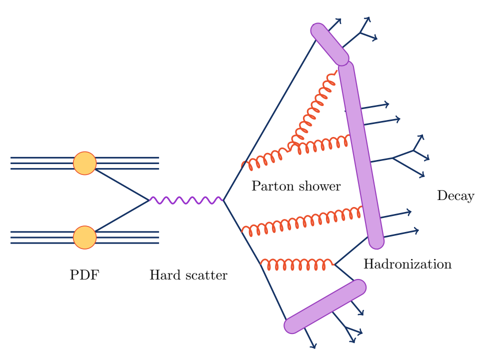
      Leading order (LO) ($\alpha_s{}^1 A_1$)

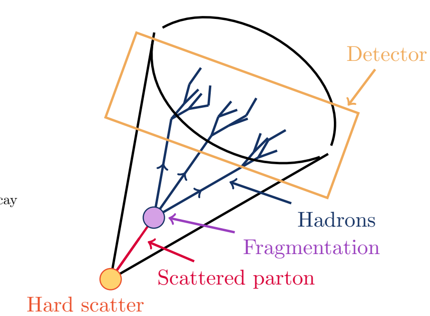
      Next-to-Leading order (NLO) ($\alpha_s{}^2 A_2$)

      Next-to-next-to-Leading order (NNLO) ($\alpha_s{}^3 A_3$)
    - α_s (Q=M_W=80 GeV) ~ 0.1 → possible to expand perturbatively
    - α_s (Q=1 GeV) ~ 0.62 → perturbative series is not as effective
    - α_s (Q≈Λ_{QCD}) ~ very large → need different, non-pQCD, method

Jets and Non-pQCD

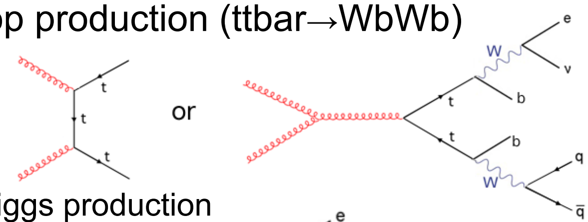
- Non-pQCD is needed for parton showers (creation of jets)
  - Large distances and small energies make pQCD impossible
  - Use previous experimental measurements to model
- Partons (quarks and gluons) radiate more partons, which hadronize and decay to form a jet



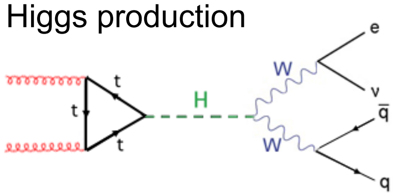


W+jets: Background for Top, Higgs, New Particles

- top production (ttbar→WbWb)



- Higgs production

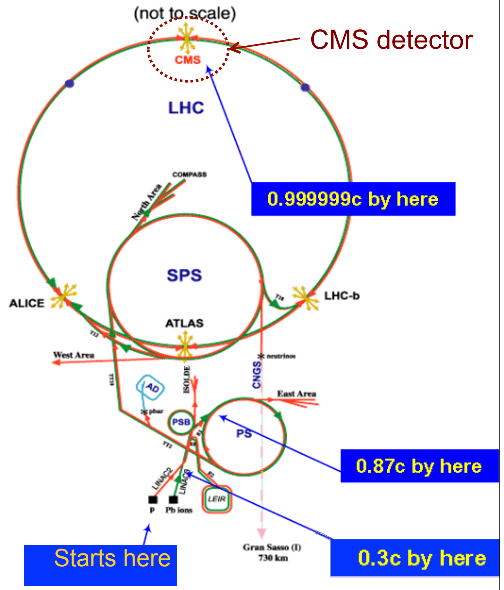


- WW production
- W', Z' decay into the W+jet-jet final state
  - Z' →WW→eνjj

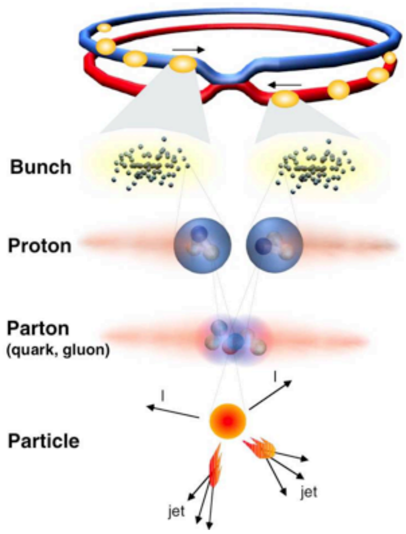
The Large Hadron Collider

- 7 TeV proton-proton collider
  - 3.5 TeV per beam
  - Design: 14 TeV
- 4T magnets
  - Design: 8T
- Circumference of 27 km
- Luminosity of 10^{32} $cm^{-2}s^{-1}$
  - Design: 10^{34} $cm^{-2}s^{-1}$
- The acceleration process
  - Linac2, produces 50 MeV protons
  - Proton Synchrotron Booster (PSB) increases energy to 1.4 GeV, Proton Synchrotron (PS) increases energy to 24 GeV
  - Super Proton Synchrotron (SPS) increases energy up to 450 GeV

CERN Accelerators
(not to scale)



Proton-Proton interaction at the LHC



Proton-Proton	**2835 bunch/beam**	(368 bunch 1st yr)
Protons/bunch	**10^{11}**	
Beam energy	3.5 TeV (3.5×10^{12} eV)	
Luminosity	**10^{34} cm^{-2} s^{-1}**	

2010: ~2×10^{32} cm^{-2}s^{-1}, Now: 1.2×10^{33} cm^{-2}s^{-1}

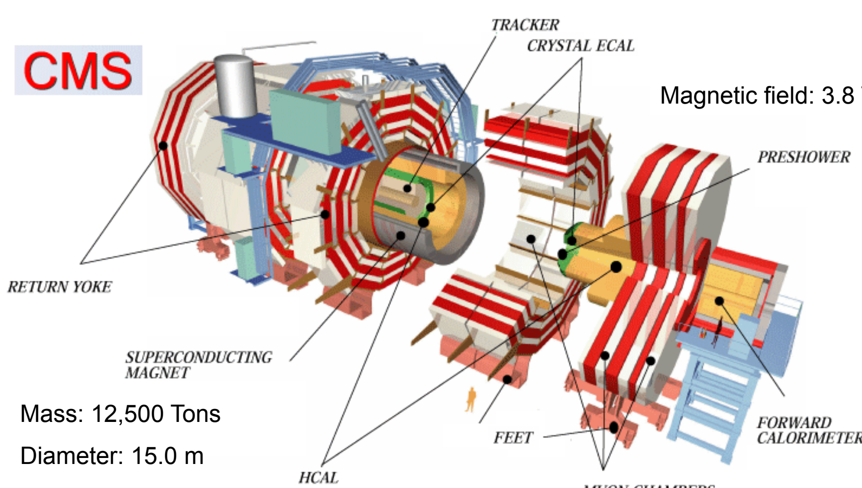
Crossing rate	**40 MHz**
Collisions ≈	**10^7 - 10^9 Hz**

Luminosity L = particle flux/time

Interaction rate: $\frac{dN}{dt} = L\sigma$

Cross section σ = "effective" area of interacting particles

Compact Muon Solenoid (CMS)



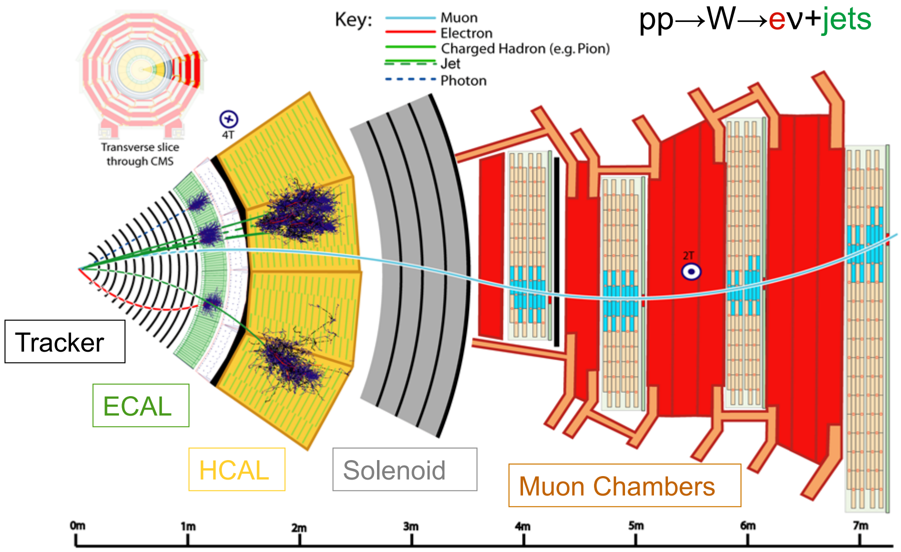
Magnetic field: 3.8 T

Mass: 12,500 Tons

Diameter: 15.0 m

Length: 21.5 m

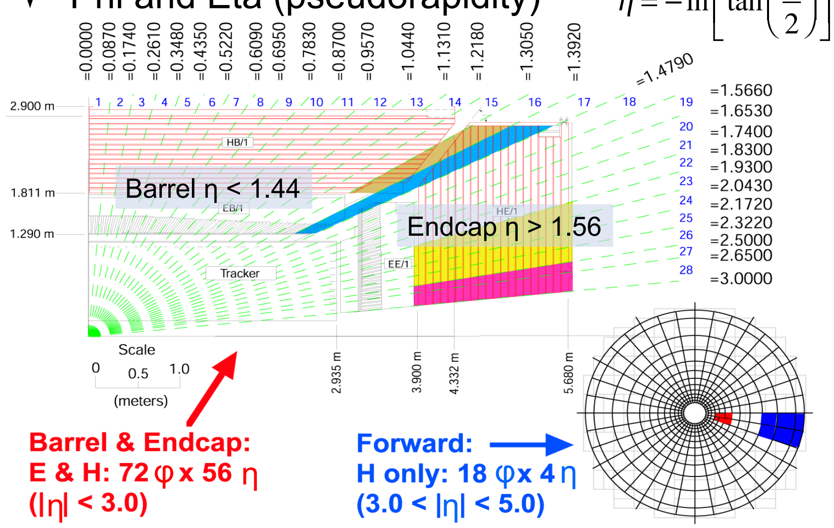
CMS Detector Parts



CMS Geometry

- Phi and Eta (pseudorapidity)

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

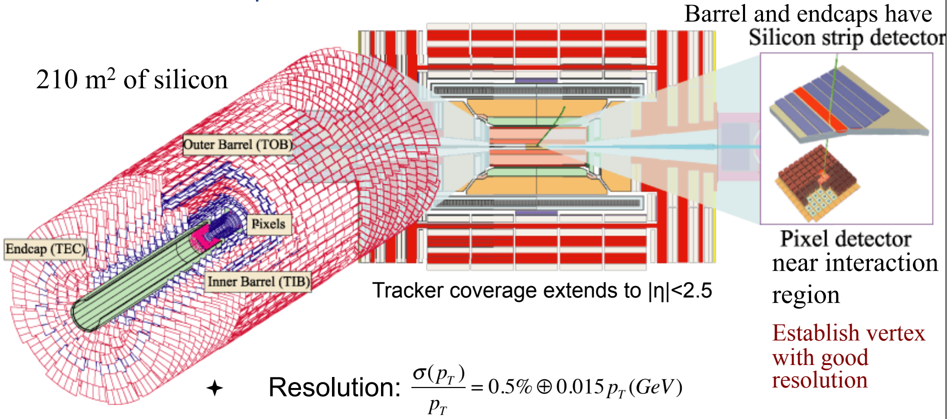


Barrel & Endcap:
E & H: 72 φ x 56 η
(|η| < 3.0)

Forward:
H only: 18 φ x 4 η
(3.0 < |η| < 5.0)

Tracker

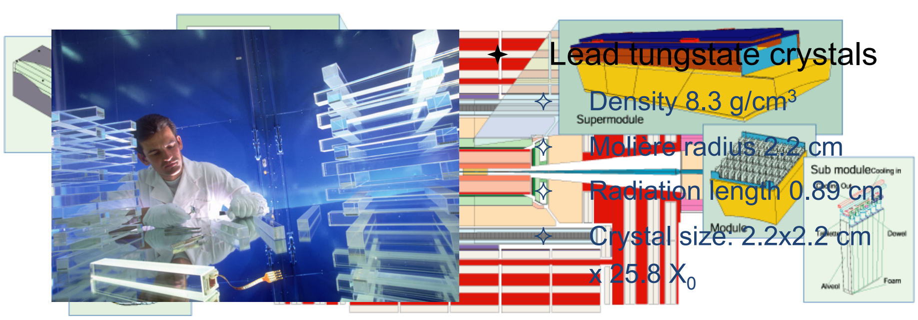
- Measures path and transverse momentum (p_T) of charged objects
  - Will help ID electrons from W decays, measure the p_T, and eliminate photons



- Resolution: $\frac{\sigma(p_T)}{p_T} = 0.5\% \oplus 0.015\, p_T(GeV)$

Electromagnetic Calorimeter

- Measures e/γ energy within |η| < 3 using 76,000 lead tungstate ($PbWO_4$) crystals
  - Will measure energy of electron from W decay

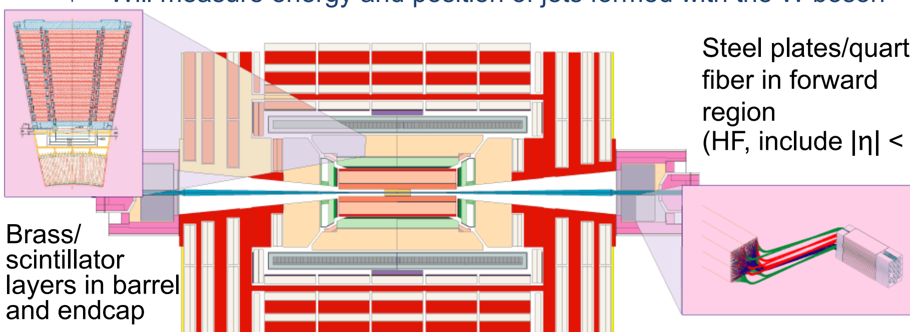


- Lead tungstate crystals
  - Density 8.3 g/cm³
  - Moliere radius 2.2 cm
  - Radiation length 0.89 cm
  - Crystal size: 2.2x2.2 cm x 25.8 X_0

- Resolution: $\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{41.5MeV}{E}\right)^2 + (0.3\%)^2$

Hadron Calorimeter

- Measures shower energy and location
  - Sampling calorimeter
  - Will measure energy and position of jets formed with the W boson



Steel plates/quartz fiber in forward region (HF, include |η| < 5)

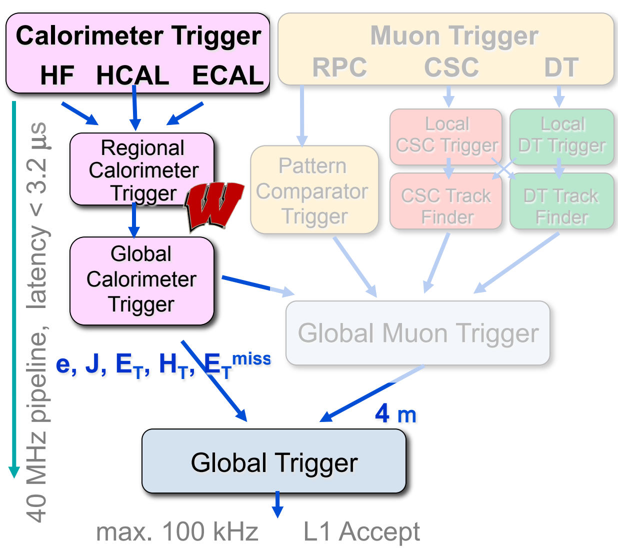
Brass/ scintillator layers in barrel and endcap (|η|<3)

Single particle resolution

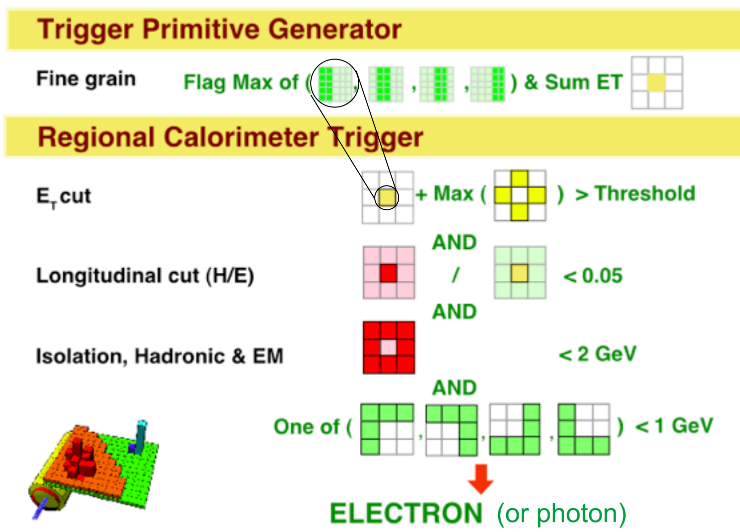
- Barrel resolution: $\left(\frac{\sigma}{E}\right)^2 = \left(\frac{90\%}{\sqrt{E}}\right)^2 + (4.5\%)^2$ HF resolution: $\left(\frac{\sigma}{E}\right)^2 = \left(\frac{198\%}{\sqrt{E}}\right)^2 + (9.0\%)^2$

Level 1 Trigger

- 0.5 GHz frequency (~ 25 ns bunch crossings * 2.2 interactions) , not all of the 0.2 MB events can be retained
- L1 trigger electronics select 50-100 kHz of interesting events
- Triggers
  - **Electron**/photon
    - 5 or 8 GeV
    - ~100% efficient
  - Jets
  - Missing E_T
  - Muon



L1 Electron Trigger



High Level Trigger

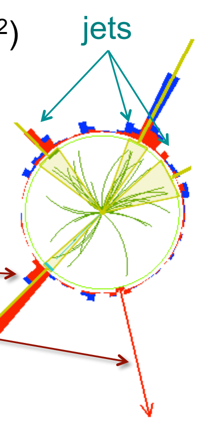
- Software trigger
  - Multi-processor farm
  - Reduces Level-1 rate from 100kHz to 300 Hz
  - Processes events every 40 ms (compared to L1 in 3.2 μs)
- Electron HLT
  - Start from L1 electron/photon seed (E_T = 5 or 8 GeV)
  - Energy deposit in ECAL
    - H/E < 0.15
  - Track reconstruction
  - Match ECAL and track information
  - Required either 15 or 17 GeV electron
    - Additional selection applied as the luminosity increased

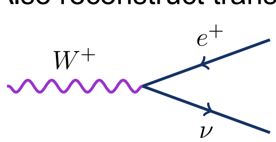
Analysis Outline

- Characteristics of W+jets
  - Electron & neutrino
  - Jets
- Previous W+jets studies at CDF and D0
  - Jet multiplicity
  - Jet transverse energy
- Simulation
  - Samples
    - Monte Carlo
    - Data
- Selection
  - Variable plots and cuts
- Efficiency
  - Tag & probe and MC
- Data-MC comparisons
- Signal Extraction
  - Fits
- Unfolding
  - Jet multiplicity
- Results
  - Cross section ratios

W+jets characteristics

- Cross section of W→eν ~10 nb (10^{-32} cm^2)
  - Measurable soon after LHC start up
  - event rate ≈ 3×10^5 events / 36 pb^{-1}
    - First year instant luminosity: 2.1 x 10^{32} cm^{-2}s^{-1}
- Reconstruct using "particle flow" (PF) technique
  - Electron
  - E_T^{miss} (from neutrino)
  - N jets
- Also reconstruct transverse W mass





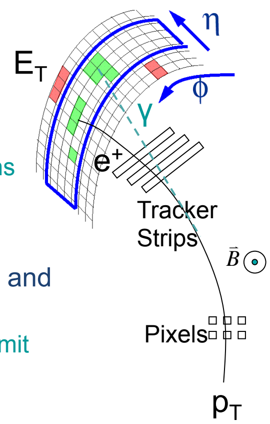
$$m_T = \sqrt{2 p_T^{(e)} p_T^{(\nu)} (1 - \cos\Delta\phi)}$$

Electron Reconstruction

- Electron reconstruction
  - E_T > 20 GeV for an EM cluster
    - $|\eta_{cluster}| < 1.44$ for barrel electrons
    - $1.56 < |\eta_{cluster}| < 2.5$ for endcap electrons
    - Wider in φ to include bremsstrahlung photons
  - Small energy deposit in HCAL
    - $E_{Had}/E_{Em} < 0.15$
  - Tracks reconstructed from hits in the pixels and strips
    - Accounts for changing radius as electrons emit bremsstrahlung photons
  - ECAL clusters matched to track, within

$$\Delta r = \sqrt{\Delta\varphi^2 + \Delta\eta^2} \leq 0.15$$

  - Isolated: no nearby energy or other tracks



Particle Flow Algorithm

- Collects information from all sub-detectors
  - Tracker, ECAL, HCAL, muon system
- Clusters of information are formed in each sub-detector and then linked to clusters from other sub-detectors
  - e.g., track is reconstructed and then link to an ECAL deposit
  - Links are based on particle compatibility between calorimeter deposits and track momentum
- All activity (above a noise threshold) is included as part of a PFlow particle
  - Electron, photon, muon, charged hadron, or neutral hadron
- Particles can then be formed into composite objects such as jets

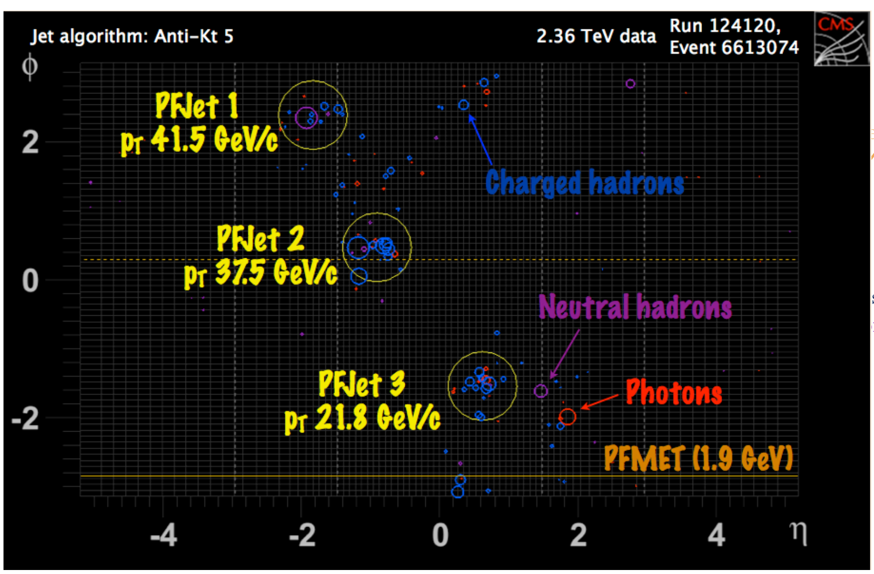
Missing Transverse Energy

- Missing Transverse Energy
  - Neutrino only 'detectable' from missing energy
    - Only interacts weakly
    - Constructed from opposite of sum of transverse momentum of all particles, *i*, reconstructed with the PFlow algorithm

$$E_T^{miss} = -\sum_i \left(E_x^i \hat{\mathbf{x}} + E_y^i \hat{\mathbf{y}} \right)$$

    - Because the initial transverse momentum of the collision is zero, so should the final
  - Expect about 40 GeV of E_T^{miss}
    - Shares the 80 GeV W boson mass with the electron

Jets Reconstruction



B-tagging jets

- Major difference between W+jets events and top quark events is the distribution of jets from b-quarks
  - Top events necessarily have a b-jet from $t \to Wb$ decay
- B-hadrons leave a distinctive pattern in the detector that can be used to distinguish them from other jets
  - B-hadrons travel a measureable distance in the tracker before decaying into lighter particles
  - Create a discriminator, based on a displaced vertex, for which b-jets are more likely to have a higher values than other jet "flavors"
    - Cut on a value and calculate the efficiency and purity at that value
- Jets are tagged as b-quarks with about 63% efficiency and a 2.7% mistag rate using the chosen algorithm and cut
  - Calculated from MC, validated on data

Tevatron (D0) W+jets

Phys. Rev. D 77, 011108 (2008)

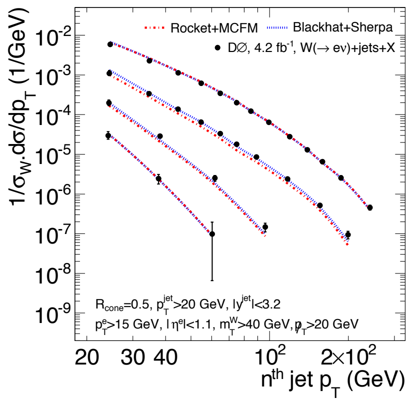
- Tevatron info:
  - p -p_{bar} collisions
  - $\sqrt{s}$ = 1.96 TeV
- Backgrounds to W+jets at Tevatron:
  - Leptonic
    - Top
    - $W \to \tau\nu$
    - $Z \to e^+e^-$
  - Multi-jet
    - QCD
    - Y+jets

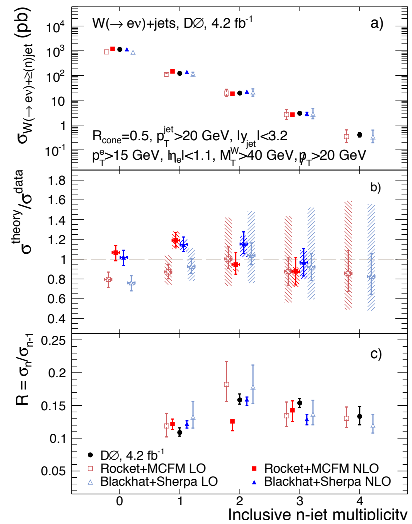
- Measurement at D0
  - L = 4.2 fb^{-1}
  - Select events with electron $E_T > 15$ GeV and $|\eta| < 1.1$; $E_T^{miss} > 20$ GeV; $M_T > 40$ GeV
  - N jets, found using
    - $\Delta R = 0.5$ cone algorithm
    - $|\eta| < 3.2$
    - $E_T > 20$ GeV for counting

D0 W+jets Results

hep-ph/1106.1457v1

- Good agreement seen between data and MCs in σ by jet p_T and σ by jet multiplicity





Simulation (Monte Carlo) for CMS

- W + jets simulated with MadGraph
  - Fixed order matrix element calculations of cross sections
  - Generates multi-parton processes in hadronic collisions.
- Subsequent generator level simulation with Pythia6 Tune Z2
  - Creates underlying event
  - Generates event hadronization, parton shower, and initial and final state radiation (IFSR)
- Detector simulated using GEANT4
  - Toolkit for the simulation of the passage of particles through matter

Hard scattering MadGraph → Hadronization, showers, IFSR PYTHIA → Detector simulation GEANT4 → Reconstruction of event CMSSW

Data/Monte Carlo (MC) Samples for CMS

- Data collected from June through October 2010
  - Only included declared “good” runs
  - Total of 36.1 ± 1.4 pb^{-1}

Run Range	Trigger Name
136033 - 137028	HLT_Photon10_L1R
138564 - 140401	HLT_Photon15_Cleaned_L1R
141956 - 144114	HLT_Ele15_SW_CaloEleId_L1R
146428 - 147116	HLT_Ele17_SW_CaloEleId_L1R
147196 - 148058	HLT_Ele17_SW_TightEleId_L1R_v1
148822 - 149063	HLT_Ele17_SW_TighterEleIdIsol_L1R_v2
149181 - 149442	HLT_Ele17_SW_TighterEleIdIsol_L1R_v3

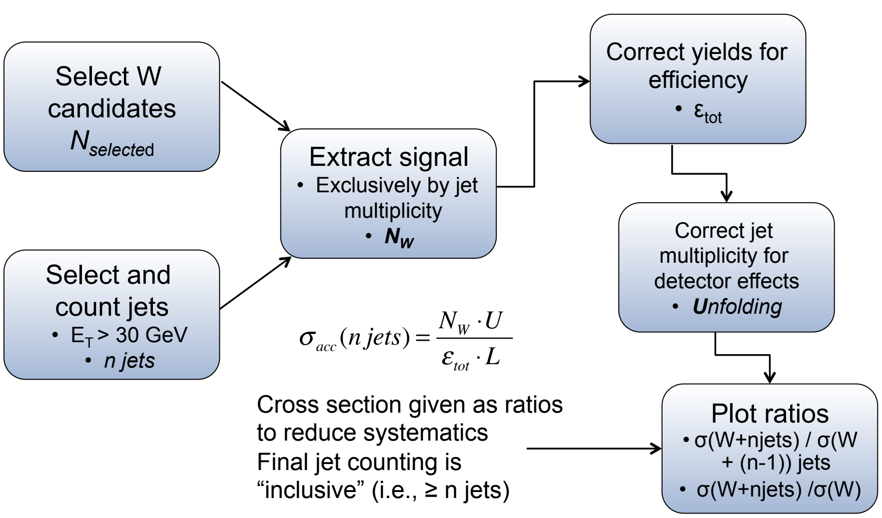
Process	Generator	Cross sec. (pb)	
W+jets	MadGraph	31314	NNLO
Z+jets (M_{ll} > 50 GeV)	MadGraph	3048	NNLO
Ttbar	MadGraph	157	NLO
QCD (20 < p_T < 170 GeV)	Pythia	~10^6	LO
Y+jet (15 < p_T < 80 GeV)	Pythia	~10^4-10^6	LO

- MC samples listed in table
  - Madgraph TuneZ2 is default
  - Pythia and Madgraph TuneD6T used for systematic studies

NNLO cross section calculations done with “Fully Exclusive W and Z production” (FEWZ) OR Monte Carlo for FeMtobarn processes (MCFM) simulation code (EWK and top respectively)

July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 29

Analysis Flow



$$\sigma_{acc}(n\,jets) = \frac{N_W \cdot U}{\varepsilon_{tot} \cdot L}$$

Cross section given as ratios to reduce systematics
Final jet counting is “inclusive” (i.e., ≥ n jets)

July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 30

Event Selection

After HLT: 15,041,836 events

- Electron Selection
  - Acceptance
    - p_T > 20 GeV
    - $|\eta|$ < 2.5
      - exclu. 1.4442 < $|\eta|$ < 1.566

After acceptance: 6,823,434 events

  - Identification
  - Conversion rejection
  - Isolation
    - relative to p_T

After electron selection: 328,701 events

- No other electrons forming Z mass with 1st
  - ! (60 < m_{ll} < 120 GeV)
- No muons with p_T > 15 GeV
- HLT object match
- M_T > 20 GeV
  - From electron and PFlow Missing E_T
  - Necessary for data-driven fitting

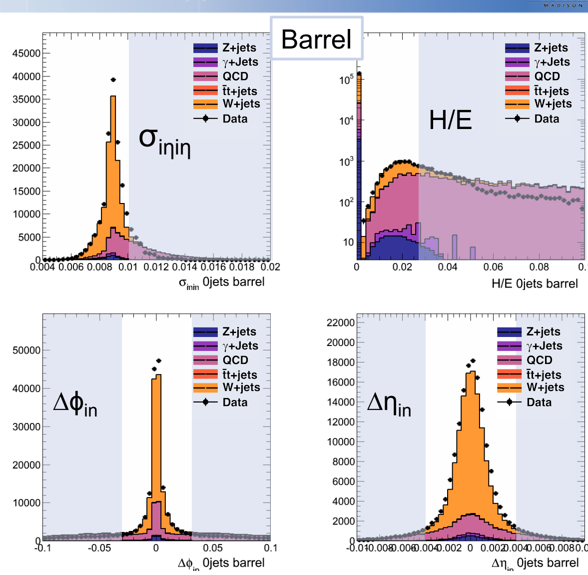
$$m_T = \sqrt{2 p_T^{(e)} p_T^{(\upsilon)} (1 - \cos\Delta\phi)}$$

After full selection: 219,815 events

- Next slides: ID, conv. rej. and isolation variables with all cuts applied but for the variable shown, with shaded area for rejected region
  - Need some selection applied to be compatible with QCD Monte Carlo and HLT paths used in data and MC

Electron Identification: $\sigma_{i\eta i\eta}$, H/E, $\Delta\phi_{in}$ & $\Delta\eta_{in}$

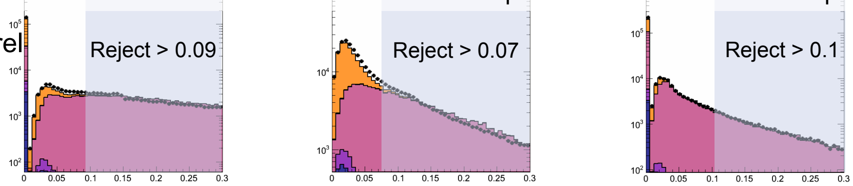
- $\sigma_{i\eta i\eta}$: Width of EM cluster
  - Reject
    - > 0.01 (barrel)
    - > 0.03 (endcap)
- H/E: Hadronic activity
  - Reject
    - > 0.040 (barrel)
    - > 0.025 (endcap)
- $\Delta\phi_{in}$ ($\Delta\eta_{in}$): Spread from track to supercluster
  - Reject $\Delta\phi_{in}$
    - > 0.03 (barrel)
    - > 0.02 (endcap)
  - Reject $\Delta\eta_{in}$
    - > 0.004 (barrel)
    - > 0.005 (endcap)



Electron selection: Isolation

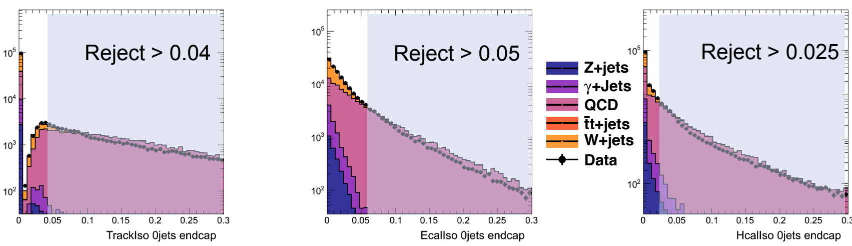
Track	ECAL	HCAL
Sum of p_T of tracks around electron track	Sum of ECAL deposits around electron deposit	Sum of HCAL deposits around electron deposit

Barrel



Isolation removes a large portion of the QCD background

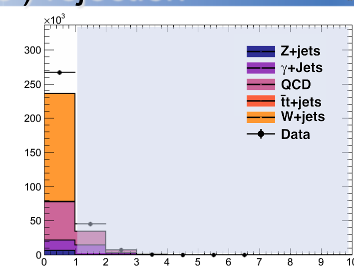
Endcap

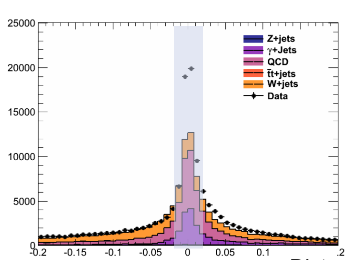


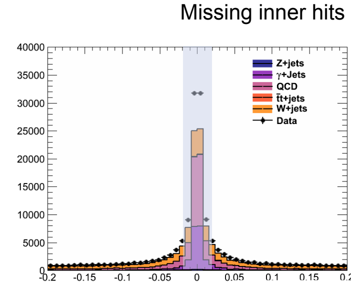
July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 33

Electron selection: Conversion ($\gamma \rightarrow e^+e^-$) rejection

- Missing Inner Hits
  - No missing inner hits between vertex and first hit of reconstructed electron track
- Dist
  - Distance of closest approach of "partner" track
- ΔCot(θ)
  - Difference in polar angle between track and "partner" track
- Reject if Missing hits OR (Dist < 0.02 && ΔCot(θ) < 0.02)







July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 34

Electron Selection: Summary

- Table at the right shows a summary of the values used for the identification, conversion rejection, and isolation variables

After acceptance: 6,823,434 events

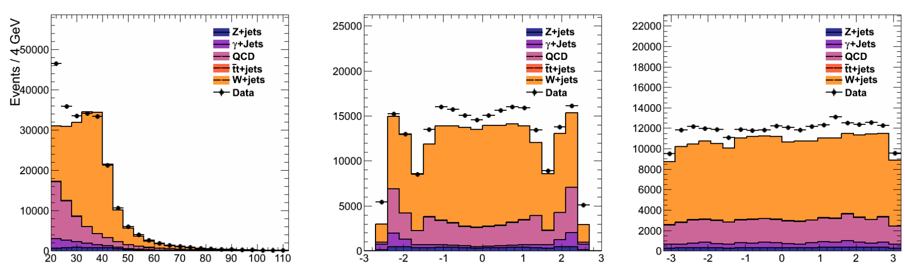
After ID Cuts: 1,205,840 events

After Isolation Cuts: 514,511 events

After conversion rejection: 328,701 events

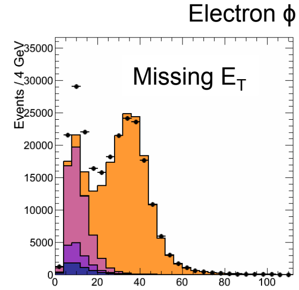
	Barrel	Endcap
Identification		
$\sigma_{i\eta i\eta}$	0.01	0.03
$\Delta\phi_{in}$	0.03	0.02
$\Delta\eta_{in}$	0.004	0.005
H/E	0.04	0.025
Isolation		
Track iso	0.09	0.04
Ecal iso	0.07	0.05
Hcal iso	0.10	0.025
Conversion rejection		
Missing hits	0 OR	
Dist	(0.02 AND	
$\Delta\cot(\theta)$	0.02)	

Electron variables and Missing E_T



Electron p_T Electron $|\eta|$ Electron ϕ

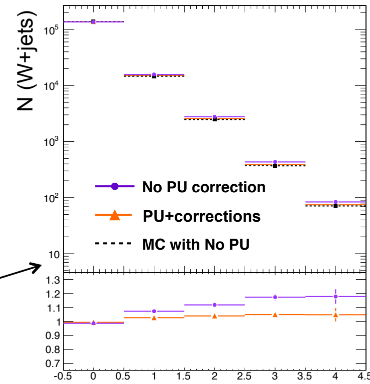
- MC is scaled to cross-section x 36.1 pb^{-1}
- QCD scale is underestimated in Monte Carlo, so data dominates
  - Signal and background yields will be fit to extract the signal without relying on the QCD scaling
- Electrons in data more central in η than in Monte Carlo



Jet Selection

- Particle Flow Jets
  - Corrected for pile-up and non-uniformity in η and E_T
  - $E_T > 30$ GeV
    - ★ Removes jets from underlying event
    - ★ Smaller pile-up corrections needed
  - $|\eta| < 2.4$ (within tracker acceptance)
  - Loose identification requirements
    - ★ Remove noise, assure true particles
  - If selected electron is within $\Delta R < 0.5$, remove jet
  - Effect of pile-up on jet multiplicity
    - ★ Pile-up comes from additional proton interactions in a bunch
    - ★ Adds energy to jets and needs to be removed

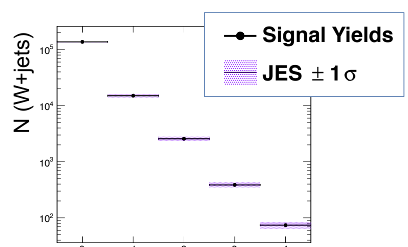
Pile-up (PU) and corrections study on jet multiplicity

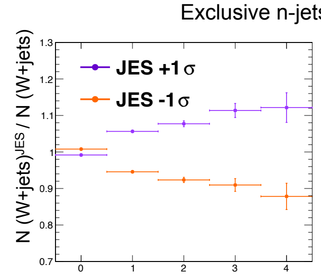


Effect of JES uncertainty on n-jets

- Jet energy scale (JES) uncertainty
  - Add in quadrature: Energy corrections + Pile-up + Flavor
    - ★ Jet energy corrections (JEC) dependent on eta and p_T (~3%)
    - ★ Pile-up dependent on jet p_T (~1.2 % for 30 GeV jet)
    - ★ Flavor (b-jets) ~ 2-3%
  - Additional PU uncertainties on njets: (0.5, 2, 4, 5, 5)%

njets	+ 1 σ (%)	- 1 σ (%)
= 0	1.02	1.06
= 1	6.2	6.5
= 2	9.0	9.0
= 3	10.6	12.9
≥ 4	13.1	14.4

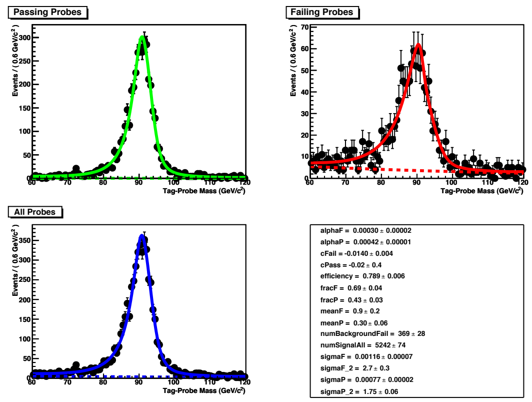




Selection efficiency:
Tag and Probe for data-driven efficiency

- We reconstruct Z events which have two good electrons. One of them is "tagged" to select events, and the efficiency of measuring the other is "probed"
  - Three steps: $\varepsilon_{T\&P} = \varepsilon_{reconstruction} \times \varepsilon_{selection} \times \varepsilon_{trigger}$
- Example fits to the passing and failing probes for the WP80 selection, $\varepsilon_{selection}$

$$\varepsilon = \frac{2N_{TT} + N_{TP}}{2N_{TT} + N_{TP} + N_{TF}}$$



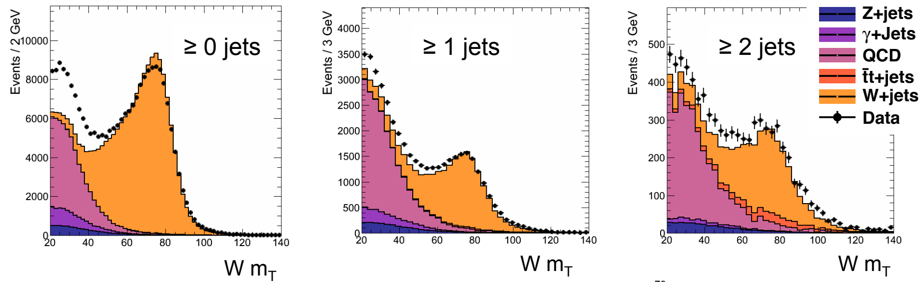
See full T&P explanation

Selection Efficiency:
Full Event Selection

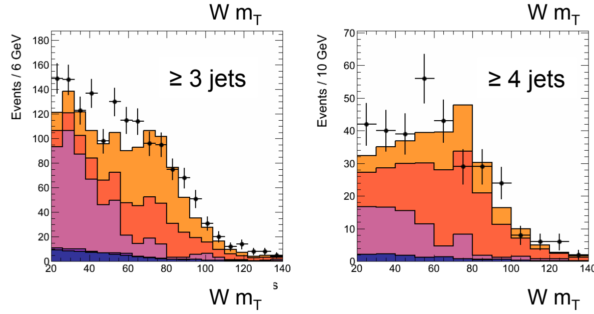
- Measure efficiency using tag-and-probe strategy on Z+jets data and MC samples
  - Electron selection efficiency found as a function of jet multiplicity
  - Use jet E_T > 15 GeV to increase statistics
- Tag-and-probe results combined with the full W+jets MC selection for final selection efficiency
  - W+jets MC efficiency: full selection / generator electrons in acceptance
    - Acceptance: generator electron $p_T > 20$ GeV, $\eta < 2.5$ (not in gap)
  - $\varepsilon_{Total} = MC_W$ * T&P data / T&P MC

Efficiency	0 jets	1 jets	2 jets	3 jets	≥ 4 jets
MC_W (full selection)	0.694	0.646	0.595	0.540	0.486
T&P data	0.752	0.743	0.722	0.735	0.693
T&P MC	0.732	0.733	0.729	0.720	0.710
ε_{Total} = MC * T&P data / T&P MC	**0.713**	**0.655**	**0.589**	**0.551**	**0.474**

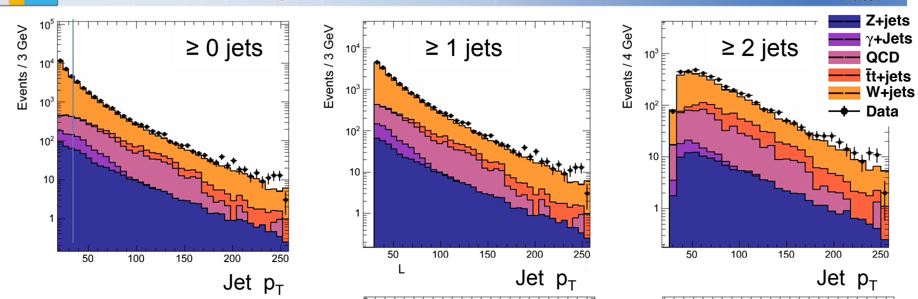
Data-MC comparisons of event variables:
W Transverse Mass



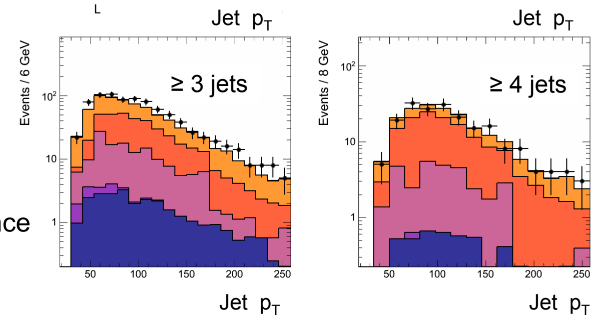
- M_T > 20 GeV
- MC is scaled to 36.1 pb^{-1}
- Calculation of QCD sample known to be underestimated
  - Signal extraction does not rely on MC



Data-MC comparisons of event variables:
Leading Jet Transverse Momentum

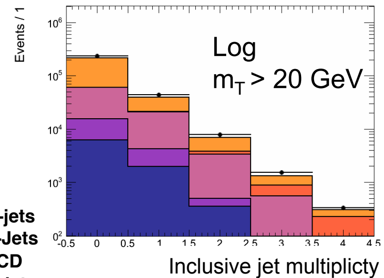


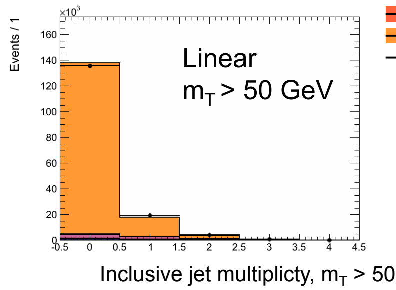
- Log scale plots
- MC is scaled to 36.1 pb^{-1}
- M_T > 50 GeV to enhance signal

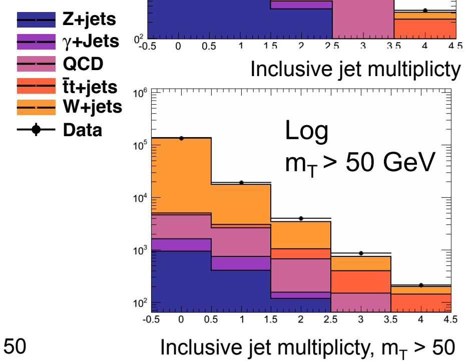


Data-MC comparisons of event variables: Jet Multiplicity

- MC is scaled to 36.1 pb^{-1}
- QCD sample cross section known to be underestimated
  - Signal extraction use to determine signal and background cross sections







Signal Extraction: Strategy

- Use functional fits to W m_T to distinguish signal from majority of backgrounds
  - Probability distribution function (PDF)
    - Parameterized on MC
- Use fit to number of b-tagged jets to distinguish signal from top
  - Top quark decays to W, so it also peaks in M_T
  - Method validated on data, no reliance on MC cross sections
- Perform 2D fits of M_T x $n_{btagged}$ for each exclusive jet multiplicity
- Species:
  - Signal (W+jets)
  - Top (ttbar, single top)
    - Divided into three subspecies based on number of b-jet (0, 1, ≥ 2)
  - Others (QCD, Z, W→τν, γjets)
    - Model based on a background enriched sample in data

Signal Extraction: Fitting method to m_T

- Fit W m_T distribution with a cruijff function
  - Mean and resolution can then be floated to be compatible data
  - The "cruijff" function is a modified Gaussian with left and right tails

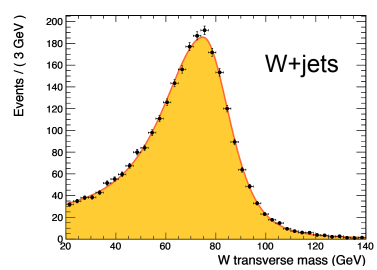
$$f(x;m,\sigma_L,\sigma_R,\alpha_L,\alpha_R) = N_s \cdot e^{-\frac{(x-m)^2}{2\sigma^2+\alpha(x-m)^2}}$$

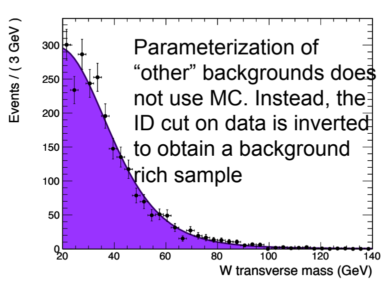
where $\sigma = \sigma_L(\sigma_R)$ for $x < m (x > m)$ and $\alpha = \alpha_L(\alpha_R)$ for $x < m (x > m)$.

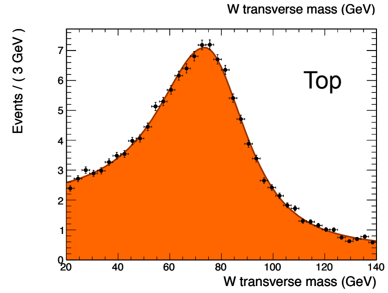
  - Cruijff accounts for the irregular tails – m_T has a jacobian peak
  - Two cruijffs used for 0-1 jets
    - Accounts for kinematic effects of electron p_T > 20 GeV
- The function is fit to the MC for each species, and then the three are combined and fit to data
  - Yields of each are floated
    - ttbar and W yields separated using n_bjets (next slide)
  - Mean and resolutions of signal are floated (for 0, 1 & 2 jets)
  - Mean for signal (3 & 4 jets) is floated
  - Top parameters are set to MC values, parameters are floated for "others"

Signal Extraction: Example M_T Cruijff Fits to MC

- Fit to MC m_T for initial parameterization
- Njets == 2
- Points are MC
- Histograms are the probability distribution function (PDF) fit to the MC







Signal Extraction:
Fitting method to n_b-jets

- Number of b-tagged jets distribution is different between W and top events
  - Use probability distribution function (PDF) to describe (depends on number of jets, number of b-flavored jets (n_{bj}), mistag rate and tag rate (from data-driven study)

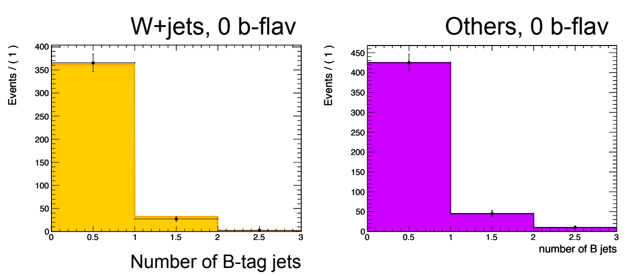
$$P(n_j^{tagged}|n_j, n_{bj}, \epsilon_{nob}, \epsilon_b) =$$

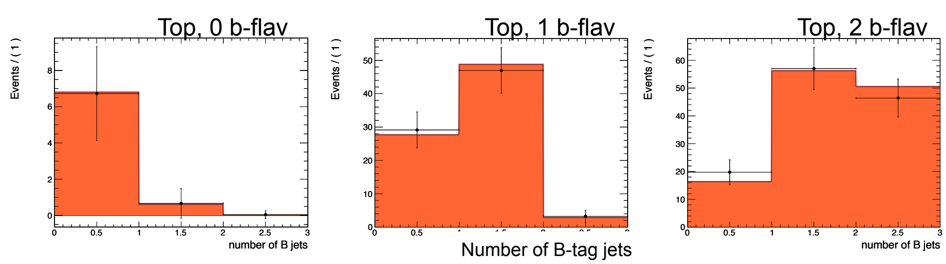
$$\begin{cases} (1-\epsilon_{nob})^{n_j-n_{bj}} \cdot (1-\epsilon_b)^{n_{bj}} & n_j^{tagged}=0 \\ (1-\epsilon_{nob})^{n_j-n_{bj}-1} \cdot \epsilon_{nob} \cdot (n_j - n_{bj}) \cdot (1-\epsilon_b)^{n_{bj}} + \\ \quad (1-\epsilon_{nob})^{n_j-n_{bj}} \cdot (1-\epsilon_b)^{n_{bj}-1} \cdot (\epsilon_b) \cdot n_{bj} & n_j^{tagged}=1 \\ 1 - P(0) - P(1) & n_j^{tagged} \geq 2 \end{cases}$$

  - n_b = number of b-tagged jets
  - n_{bj} = number of jets in acceptance that are b-flavored (true)
  - ε_{nob} = mistag rate
    - 2.42 ± 0.03 (stat) ± 0.5 (syst)% from MC and validated on data
  - ε_b = tag rate
    - 63 ± 6.3% from MC and validated on data

Signal Extraction:
Example of number of B-tagged

- Number of b-tagged jets in MC
- Points are MC
- Histograms are PDF
- Njets == 3
- PDF describes MC well

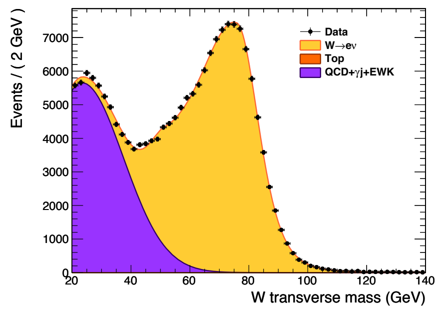




Signal Extraction: Fit to m_T, == 0 jets

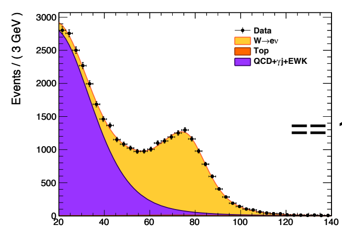
- Fit to transverse mass for events with no jets E_T > 30 GeV
- W+Jets PDF in yellow
- Ttbar PDF in orange
- QCD + γjets + Z+jets + W→τν PDF in purple
- Signal Yield: 131376 ± 423
  - efficiency corrected: 184258
- Cruijff fits model the data well

30 GeV jets

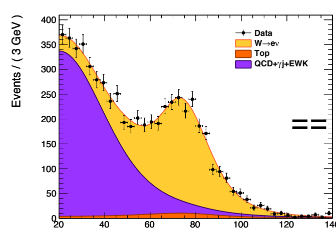


Signal Extraction: Fit to m_T for 1 and 2 jet events

- W+Jets PDF in yellow
- Ttbar PDF in orange
- QCD + γjets + Z+jets + W→τν PDF in purple
- Signal yields:
  - 15476 ± 189 for 1 jet
    - Efficiency corrected: 23627
  - 2730 ± 82 for 2 jets
    - Efficiency corrected: 4634
- Crujiff fits model data well



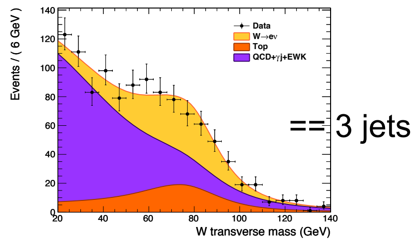
== 1 jet



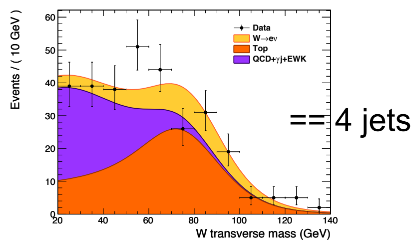
== 2 jets

Signal Extraction:
Fit to m_T for 3 and 4 jet events

- W+Jets PDF in yellow
- Ttbar PDF in orange
- QCD + γjets + Z+jets + W→τν PDF in purple

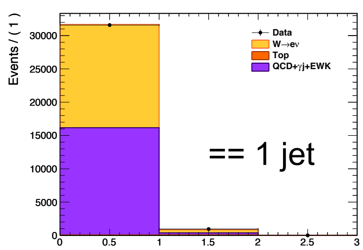


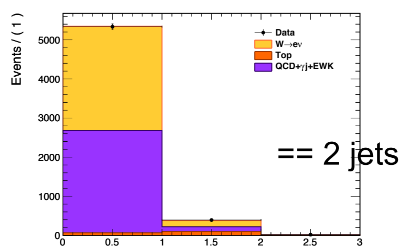
- Signal yields:
  - 362 ± 38 for 3 jet
    - Efficiency corrected: 657
  - 60.1 ± 17.8 for 4 jets
    - Efficiency corrected: 127



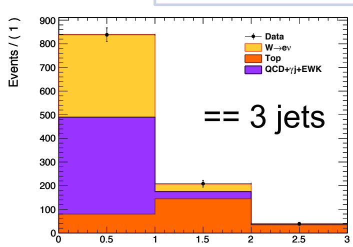
- Low statistics and high ttbar make the 4 jet bin difficult to fit

Signal Extraction:
Fit to number of b-tag jets

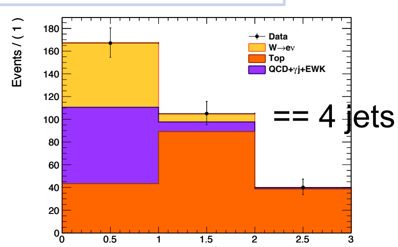




PDFs fit well to the b-tag jet distributions in data



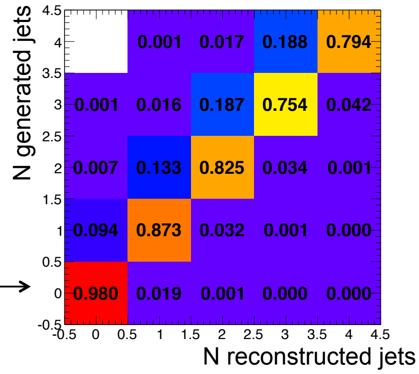
Number of b-tagged jets



Number of b-tagged jets

Unfolding the Jet Multiplicity

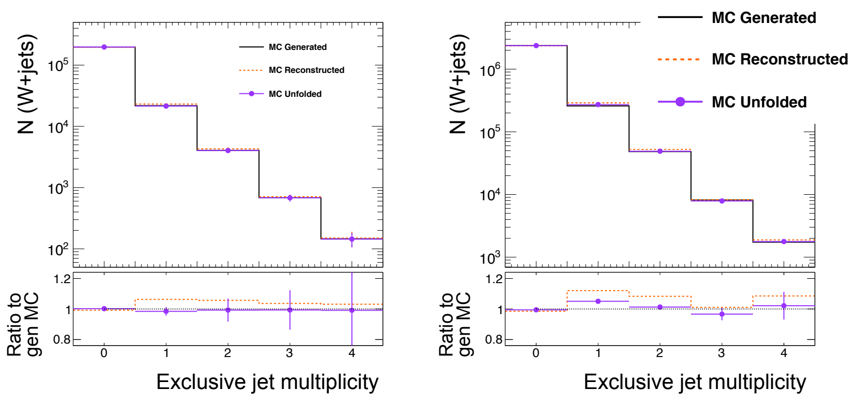
- An "unfolding" technique is used to correct the jet multiplicity distribution that is smeared from detector effects
- Unfolding "unsmears" the distribution based on the relationship between MC reconstructed and generated jets
  - A migration matrix M_{ij} is used to describe the n-jet migrations between measured (reconstructed) and true (generated) jets
  - $R_i = M_{ij} T_j$
    - In principle, invert the matrix to recover the true distribution (but slightly more complicated)
- Use the Singular Value Decomposition (SVD) method
  - Regularizes to prevent fluctuations
  - Gives the best results on MC validation compared to other methods



- Migration matrix from MadGraph TuneZ2 w/pile-up+corrections
- Only acceptance cuts are applied
  - Will match with data corrected for eff

Unfolding jet multiplicity: Closure test

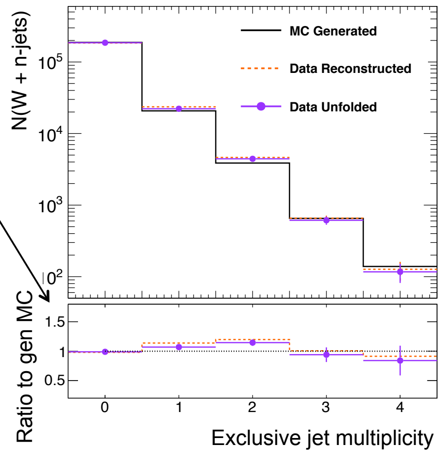
- Closure shown below:
  - Unfolding MadGraph TuneZ2 with matrix from MadGraph TuneZ2 (left)
    - Data sized sample, full selection + efficiency corrections
  - Unfolding MadGraph TuneD6T with matrix from MadGraph TuneZ2 (right)
- SVD regularization term k = 5 gives most realistic errors



Unfolding jet multiplicity: Data Yields

- Unfolding done on data for exclusive jet multiplicity
- Data has been corrected for selection efficiency
- Ratio is comparison of pre-unfolded and post-unfolded data to the generated N-jets distribution from MadGraph TuneZ2
- Systematic uncertainty in unfolding
  - Unfold with different methods
    - Different tune (Z2 vs D6T), generator (MadGraph vs Pythia), or algorithm (SVD vs Bayes)

------- Madgraph Z2 Generated
- - - - Data Reconstructed
---•--- Data Unfolded



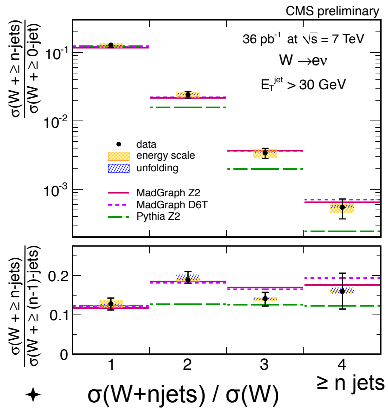
Sources of Systematic Uncertainty (%)

- Jet energy scale
  - Jet energy corrections
    - dependent on η and p_T (~3%)
    - Pile-up (~1.2 % for 30 GeV jet)
    - Flavor set to 2-3%
- Missing E_T
  - ± 10% on MET_x & MET_y
  - Affects M_T > 20 GeV cut
- Efficiency
  - From Tag and Probe and MC counting
- Fit
  - B-tag variables uncertainties
  - QCD modeling
  - Fixed parameters in m_T fit

Njets	0	1	2	3	4
JES +1σ JES -1σ	1.02 1.06	6.2 6.5	9.0 9.0	10.6 12.9	13.1 14.4
Missing E_T	0.1	0.3	0.5	0.5	1.4
Efficiency	0.5	0.3	0.8	1.4	2.7
Fit	0.1	0.8	1.26	4.16	8.95
Total + **-**	**1.14** **1.18**	**6.27** **6.56**	**9.14** **9.14**	**11.5** **13.6**	**16.2** **17.2**

- Unfolding uncertainty estimated by unfolding with different methods and comparing to the nominal
  - Not included in table above but is included in final results

Final Cross Section Ratios and uncertainties



- σ(W+njets) / σ(W)
  - No uncertainty in luminosity
  - Reduces event selection uncertainty
- σ(W+njets) / σ(W+(n-1)) jets
  - Reduces JES uncertainty
- Signal extraction, efficiency corrections and unfolding are performed on exclusive n-jet bins (i.e., n=0, n=1, n=2, n=3, n≥4)
  - Statistical + uncorrelated systematics are black error bars
    - Lepton efficiency, fit
  - Central values shifted by correlated systematics, orange band
    - Jet counting
  - Unfold with different methods, blue band
    - Different tune (Z2 vs D6T), generator (MadGraph vs Pythia), or algorithm (SVD vs Bayes)
- Good agreement between data and MadGraph MC

July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 57

Conclusions / Outlook

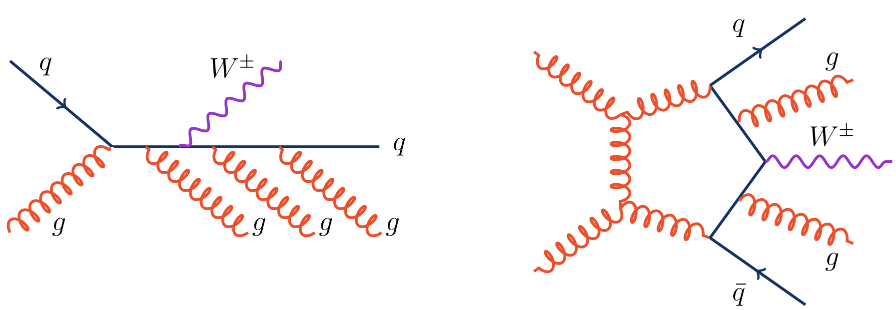
- Presented results for the W + jets cross section by jet multiplicity using **36 pb^{-1}** of data
  - Jet E_T threshold of 30 GeV
  - Extensive use of data-driven methods for efficiency and signal extraction
- The results are in agreement with MadGraph Monte Carlo predictions
  - Specific matrix element generator such as MadGraph is necessary for modeling events with > 1 jets
    - Generators without multiple final state partons, such a pythia, do not model W+jets data well
  - MadGraph will prove useful in new physics searches
- Outlook
  - Higher statistics in the future (1 fb^{-1} in 2011 already) will mean more precise measurement
    - Absolute cross sections
    - Unfolded cross section as a function of jet E_T
  - Starting point for new physics searches

July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 58

Backup

W + 4 jet examples

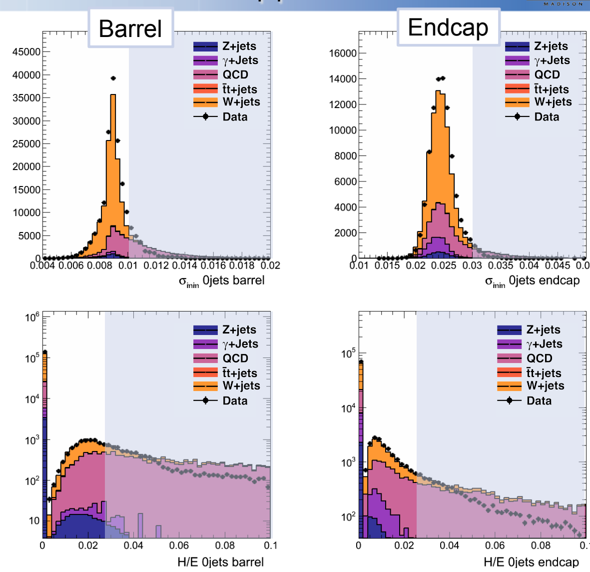
Two of the 498 possible W + 4 jet Feynman diagrams



Electron Identification: $\sigma_{i\eta i\eta}$ and H/E

- $\sigma_{i\eta i\eta}$
- Shape variable, measures width of EM cluster in η
- Reject
  - > 0.01 (barrel)
  - > 0.03 (endcap)

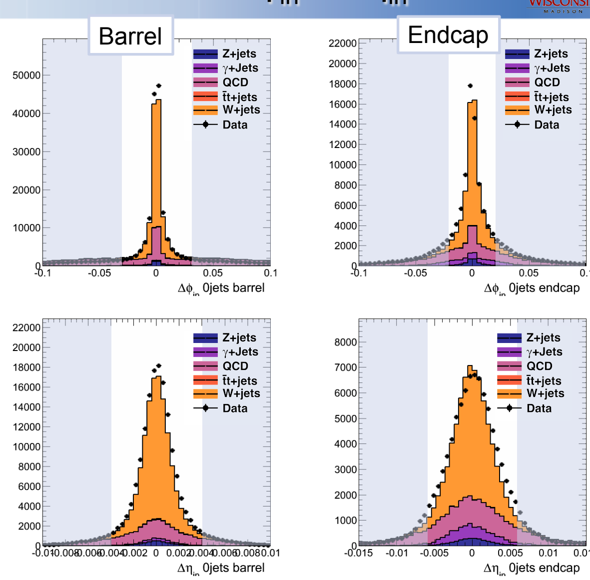
- H/E
- Measures hadronic activity in the calorimeter
- Reject
  - > 0.025 (barrel)
  - > 0.025 (endcap)



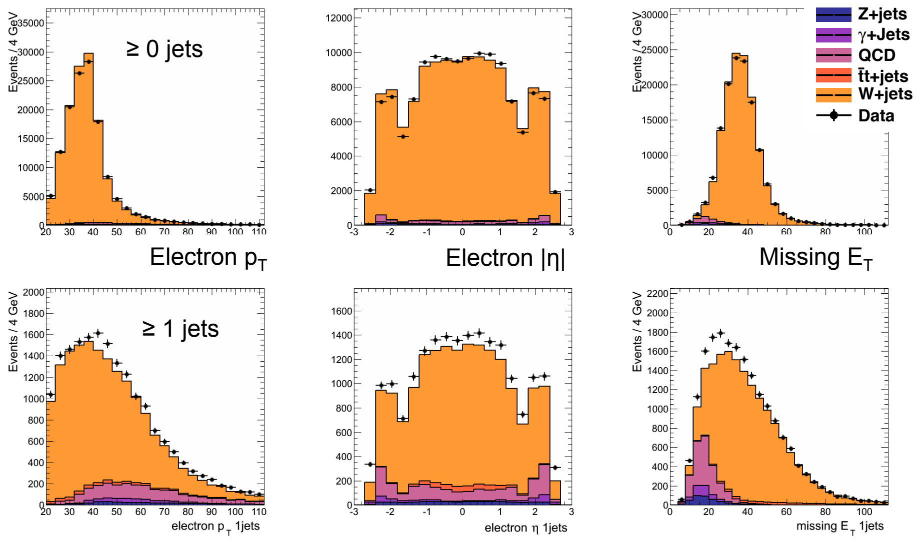
Electron Identification: $\Delta\phi_{in}$ & $\Delta\eta_{in}$

- $\Delta\phi_{in}$
- Spread in of electron ϕ from gsf track and from supercluster position
- Reject
  - > 0.03 (barrel)
  - > 0.02 (endcap)

- $\Delta\eta_{in}$
- Spread in of electron η from gsf track and from supercluster position
- Reject
  - > 0.004 (barrel)
  - > 0.005 (endcap)



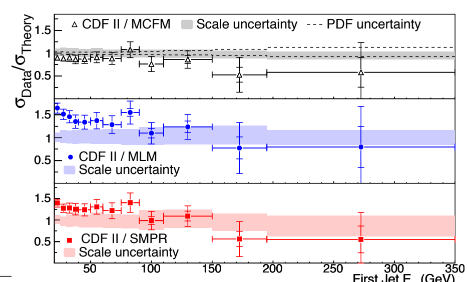
Electron variables and MET after selection
$m_T > 50$

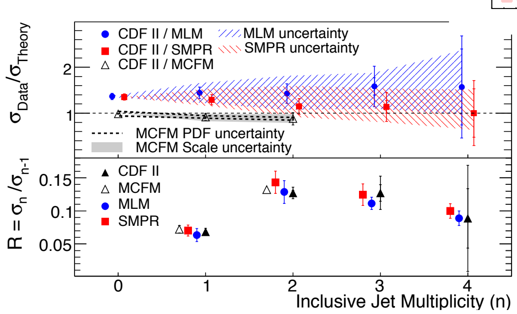


July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 63

CDF W + n jets, Jet E_T

- Differential cross section by jet transverse energy
  - Ratio of data to three different MCs
- Reasonably well described by MC samples -- after PDF tuning





- Inclusive jet multiplicity
  - Ratio of data to three MC simulations
  - Ratio of $\sigma(n)/\sigma(n-1)$ jets
- Data is well described by the NLO MC.

July 20, 2011 Kira Grogg, U. of Wisconsin -- Madison 64

Selection Efficiency: Tag & Probe

- Use data-driven "Tag-and-probe" method as part of the efficiency calculation
  - Start from Z/γ^* + jets data sample (very little background)
    - Two electrons forming an invariant mass, $60 < m_{ee} < 120$ GeV
  - One electron, the "tag", passes full selection (reduces background)
  - Second "probe" electron is divided into two samples
    - Passing the desired requirement
      - i.e., reconstruction, WP80, or HLT
    - Failing the same requirement
  - Fits are performed on the passing and failing samples to extract the number of Z electrons from the remaining background
  - Efficiency is the number of probes passing the current requirement relative to the total number of probes, e.g., $\varepsilon_{trigger} = N_{trig} / N_{WP80}$
    - $\varepsilon_{T\&P} = \varepsilon_{reconstruction} \times \varepsilon_{selection} \times \varepsilon_{trigger}$

See T&P fits

Breit-Wigner and Crystal Ball functions

Functions used in T&P fitting:

Crystal-Ball

Gaussian with power-law low-end tail

$$f(x;\alpha,n,\bar{x},\sigma) = N \cdot \begin{cases} \exp(-\frac{(x-\bar{x})^2}{2\sigma^2}), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leqslant -\alpha \end{cases}$$

where

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right),$$

$$B = \frac{n}{|\alpha|} - |\alpha|,$$

Breit-Wigner

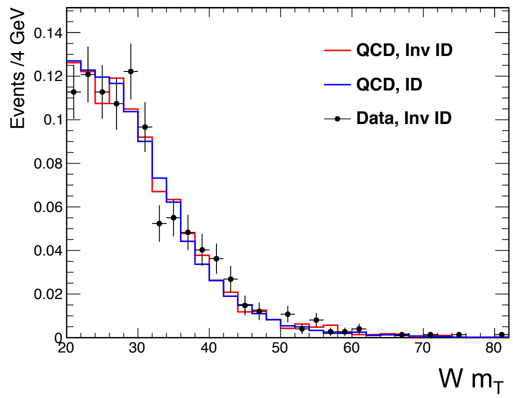
$$P(x) = \frac{\gamma}{\pi(\gamma^2 + x^2)}$$

Check on QCD m_T shape with ID inversion

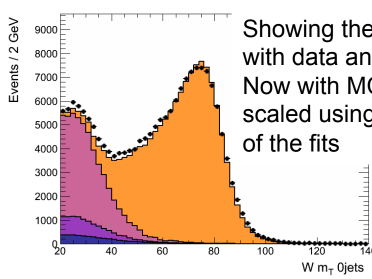
Cuts applied:
Isolation
H/E
Inverted $\Delta\phi$ and $\Delta\eta$

Isolation and H/E correlated with MET so use same cuts

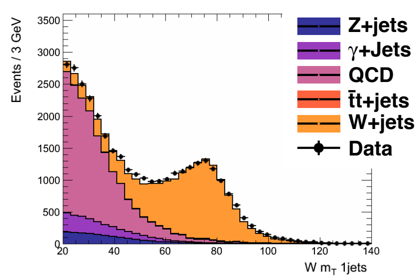
$\Delta\phi$ and $\Delta\eta$ have least correlation with MET

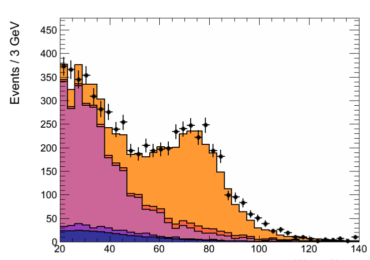


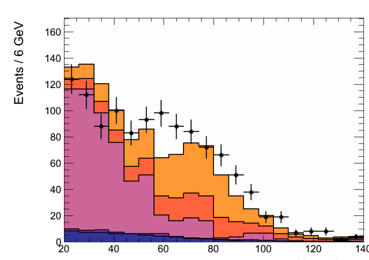
W Transverse Mass – Scaled to Fit Results



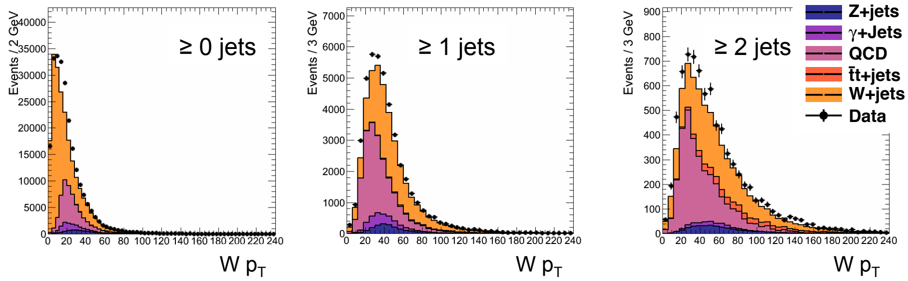
Showing the W m_T with data and MC
Now with MC scaled using results of the fits



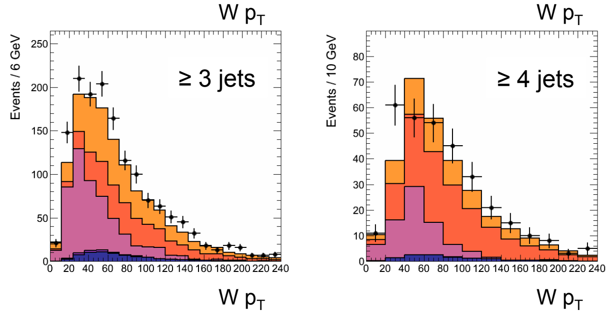




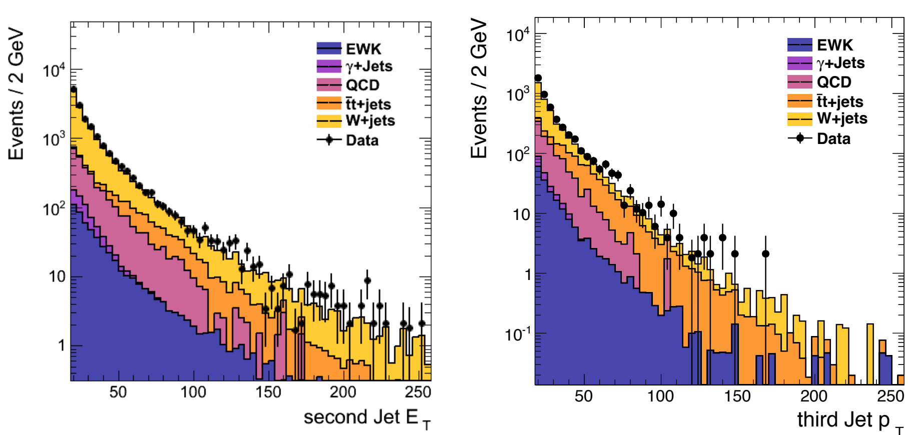
Data-MC comparisons of event variables: W Transverse Momentum



- MC is scaled to 36.1 pb^{-1}
- QCD known to be underestimated
  - Signal extraction does not rely on MC

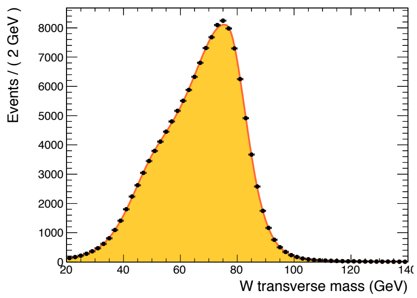


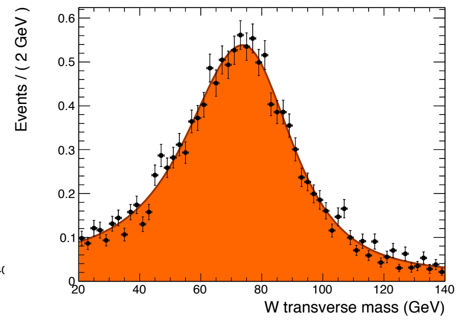
2nd and 3rd jet E_T



0 jet MC distributions

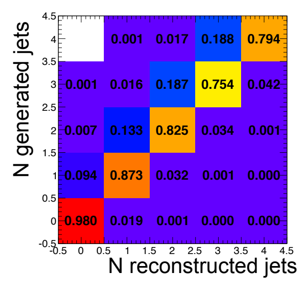
W+jets | Top

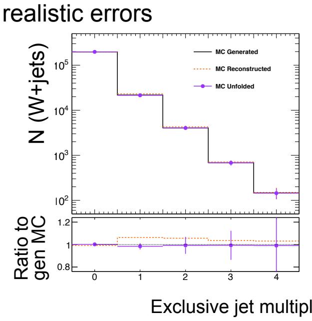


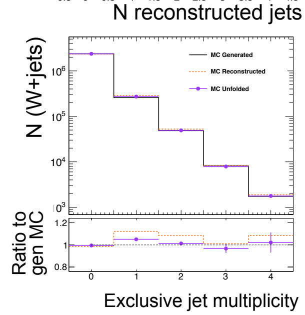


Unfolding jet multiplicity: Closure test

- Migration matrix made using MadGraph TuneZ2 w/pile-up+corrections →
- Closure shown below:
  - Unfolding MadGraph TuneZ2 matrix from with MadGraph TuneZ2 (left)
  - Unfolding MadGraph TuneZ2 with matrix from MadGraph TuneD6T (right)
- SVD regularization term k = 5 gives most realistic errors







Results: N events for = n jets

PF jet $p_T > 30$ GeV							
					Unfolding systematic deviation		
n jets	N_{obs}	ϵ_{tot}	N_{effcor}	N_{unf}	SVD - Bayes	MC generator	MC tune
0	131376 ± 423	0.713 ± 0.0049	184258 ± 1399	185946 ± 1525	4.0	697.0	-26.0
1	15476 ± 189	0.655 ± 0.00624	23627 ± 366	22198 ± 473	-7.2	-926.8	-84.9
2	2730 ± 81.6	0.589 ± 0.0115	4635 ± 165	4433 ± 217	7.6	208.1	90.4
3	362 ± 38.1	0.551 ± 0.0269	657 ± 76	613 ± 81	-6.2	14.7	9.1
4	60 ± 17.8	0.474 ± 0.0421	127 ± 39	117 ± 35	0.4	-2.3	10.1

Table 8.1: N_{obs} are the results from the signal extraction, N_{effcor} are the results after correcting for electron efficiency, ϵ_{tot}, and N_{unf} are the results after unfolding, all with with exclusive jet counting. The last three columns represent the deviation from the nominal unfolding results when changing the algorithm, the MC generator, and the MC tune, respectively.

Results: σ(W+≥njets)

PF jet $p_T > 30$ GeV								
n jets	σ	stat	stat+sys	JES syst error		Unfolding systematic deviation		
	in acceptance			(±)		SVD - Bayes	MC generator	MC tune
≥ 0 jets	5909	33.4	44.7	2.50	2.92	-0.04	-0.26	-0.04
≥ 1 jets	758	12.8	14.6	60.0	62.7	-0.15	-19.6	0.68
≥ 2 jets	143	5.92	6.49	14.2	14.6	0.05	6.11	3.04
≥ 3 jets	20.2	2.30	2.44	2.36	2.88	-0.16	0.34	0.53
≥ 4 jets	3.23	0.91	0.97	0.44	0.51	0.01	-0.06	0.28

Table 8.2: Results for cross section $\sigma\,(\geq n\,\text{jets})$ within the acceptance with inclusive jet counting. Sources of uncertainty shown are statistical, statistical + uncorrelated systematics (fit and efficiency), correlated systematics (jet energy scale, JES), and deviations when using different unfolding methods (algorithm, generator, and tune). There is also an overall 4% uncertainty for the luminosity.

Results: σ(W+≥njets)/σ(W)

PF jet $p_T > 30$ GeV								
n jets	σ ratio in acceptance	stat	stat+sys	JES syst error ($\pm$)		Unfolding systematic deviation		
						SVD - Bayes	MC generator	MC tune
≥ 1 / ≥ 0 jets	0.128	0.002	0.00234	0.0101	0.0106	-2.47e-05	-0.00331	0.000117
≥ 2 / ≥ 0 jets	0.0242	0.000987	0.00109	0.00239	0.00246	8.33e-06	0.00103	0.000514
≥ 3 / ≥ 0 jets	0.00342	0.000388	0.000413	0.000397	0.000486	-2.75e-05	5.83e-05	9.02e-05
≥ 4 / ≥ 0 jets	0.000547	0.000155	0.000164	7.35e-05	8.63e-05	1.73e-06	-1.08e-05	4.75e-05

Table 8.3: Results for cross section ratio $\sigma(W+\geq n \text{ jets})/\sigma(W)$ within the acceptance with inclusive jet counting. Sources of uncertainty shown are statistical, statistical + uncorrelated systematics (fit and efficiency), correlated systematics (jet energy scale, JES), and deviations when using different unfolding methods (algorithm, generator, and tune).

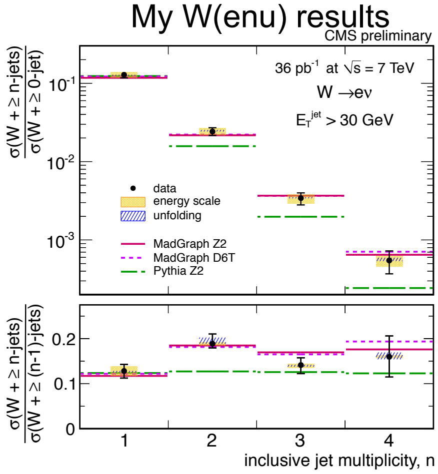
Results: σ(W+≥njets)/σ(W+≥(n−1)jets)

PF jet $p_T > 30$ GeV								
n jets	σ ratio in acceptance	stat	stat+sys	JES syst error ($\pm$)		Unfolding systematic deviation		
						SVD - Bayes	MC generator	MC tune
≥ 1 / ≥ 0 jets	0.128	0.002	0.00234	0.0101	0.0106	-2.47e-05	-0.00331	0.000117
≥ 2 / ≥ 1 jets	0.189	0.00694	0.00767	0.00351	0.004	0.000101	0.0133	0.00383
≥ 3 / ≥ 2 jets	0.141	0.0148	0.0158	0.00223	0.00636	-0.00118	-0.00349	0.000708
≥ 4 / ≥ 3 jets	0.16	0.0415	0.044	0.0026	0.00292	0.0018	-0.00577	0.00941

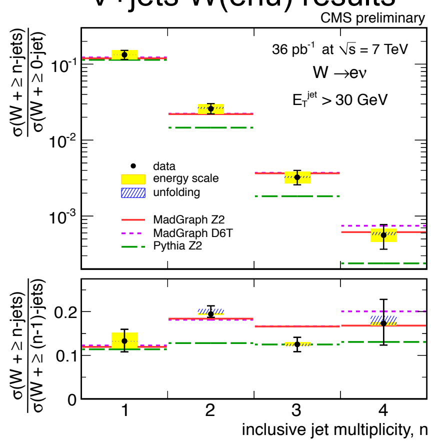
Table 8.4: Results for cross section ratio $\sigma(W+\geq n \text{ jets})/\sigma(W+\geq (n-1) \text{ jets})$ within the acceptance with inclusive jet counting. Sources of uncertainty shown are statistical, statistical + uncorrelated systematics (fit and efficiency), correlated systematics (jet energy scale, JES), and deviations when using different unfolding methods (algorithm, generator, and tune).

Comparison with V+jets

My W(enu) results



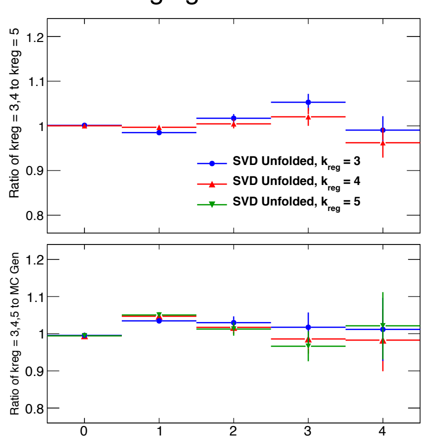
V+jets W(enu) results

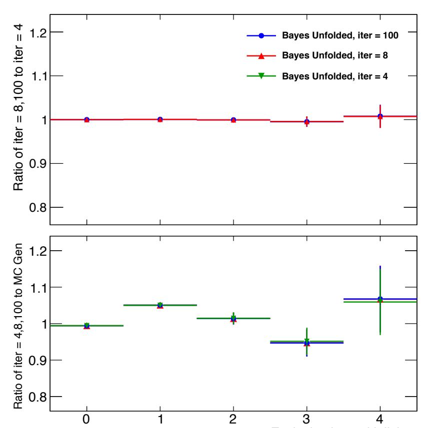


Results consistent with those from V+jets paper

Unfolding MC with different Regularization terms

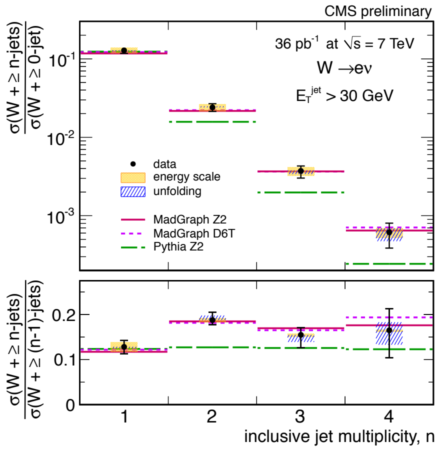
Small difference seen when changing SVD regularization term
Changing the iterations for Bayes has almost no effect



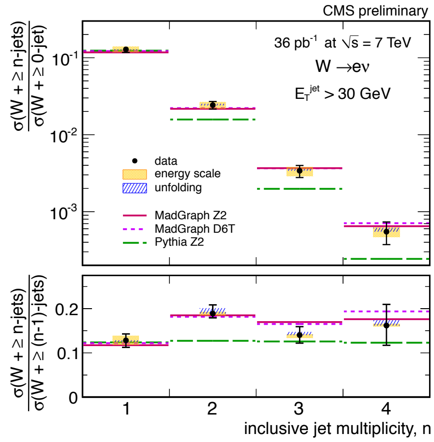


Data unfolded with alternate methods

Data unfolded with SVD k_{reg} = 4

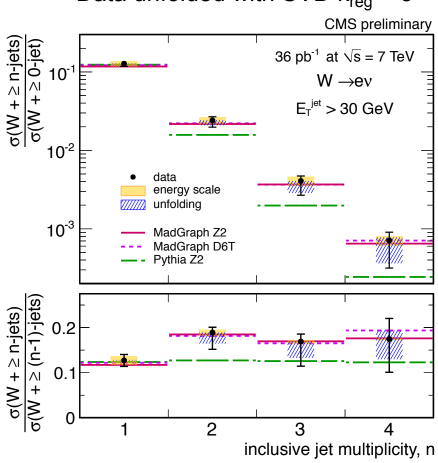


Data unfolded with Bayes iter = 4



Data unfolded with alternate methods 2

Data unfolded with SVD k_{reg} = 3



Data not unfolded