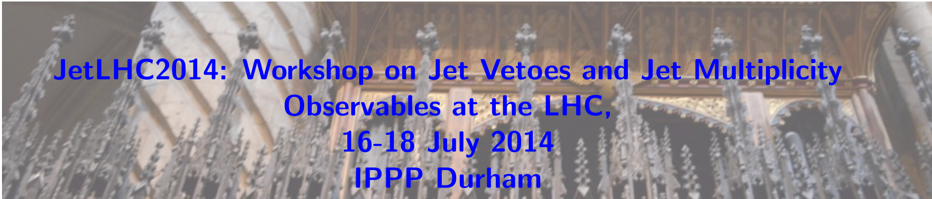
Top Quark Production with N-jets and with jet vetoes at CMS

Carmen Diez Pardos
for the CMS collaboration

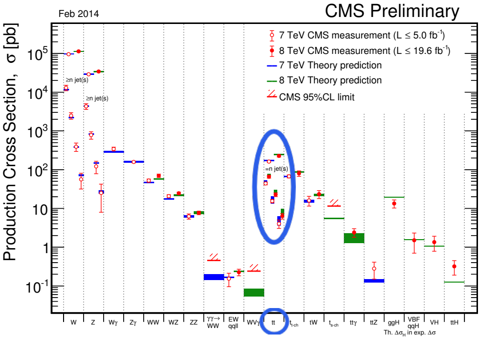
DESY



JetLHC2014: Workshop on Jet Vetoes and Jet Multiplicity Observables at the LHC,
16-18 July 2014
IPPP Durham

Outline

1. Introduction
2. Normalised differential $\sigma_{t\bar{t}}$ as a function of NJets
3. $\sigma_{t\bar{t}}$ as a function of N additional jets
4. Kinematic distributions for the leading p_T additional jets
5. Veto on additional jets
6. Summary

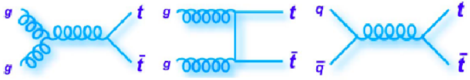


Motivation

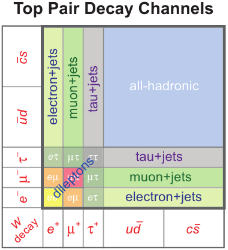
- At LHC energies, more than half of $t\bar{t}$ events are produced with additional hard jets (not coming from the $t\bar{t}$ decay).
- Precise understanding of these events is important:
  - Test higher-order QCD predictions
  - Anomalous production of $t\bar{t}$(+jets) could reveal new physics
  - $t\bar{t}$+jets is a background for many searches and for $t\bar{t}H$
- In general, sizeable uncertainty from QCD radiation for many top quark analysis
  - Theory predictions and models need to be tuned and tested with new measurements
- Large samples of $t\bar{t}$ events provide a great opportunity to study the details of the production mechanisms
  - Potential of constraining QCD radiation at the scale of the top quark mass

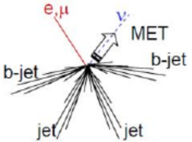
Top quark production and decay

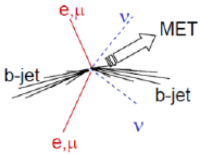
$t\bar{t}$ production mainly by gluon fusion at LHC (∼80% at 7-8 TeV)

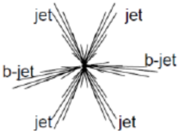


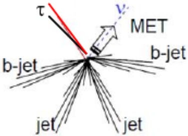
W decay defines final state:











Semileptonic [e/μ]: BR∼30% and manageable BG (ie. W+jets)

Dileptonic [e/μ]: BR∼5% and small BG (ie. DY+jets)

All-jets: BR∼46% but largest BG (ie. QCD multijet)

τ+jets: BR∼15%

Introduction

Present studies of $t\bar{t}$+jets with 7 TeV data, both in the dilepton and the lepton+jets channels, and 8 TeV data in the dilepton channels.

- Measurements performed:
  - ◇ Differential cross-section measurement in the dilepton and l+jets channels as a function of jet multiplicity
  - ◇ l+jets: $t\bar{t}$ production as a function of the additional jet multiplicity.
  - ◇ dilepton: $t\bar{t}$ production with a veto on additional jet activity.

  $\longrightarrow$ All measurements in visible phase space, corrected to particle level.

- 7 TeV results: arXiv:1404.3171, submitted to EPJC (lepton+jets, dilepton channels)
- 8 TeV results: CMS-PAS-TOP-12-041 (dilepton channel)

Generator setups for $t\bar{t}$ at CMS

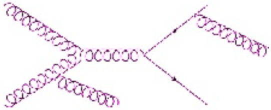
Matrix element	Shower & Hadronization	PDF	Tune
MadGraph v5	Pythia 6	cteq6l	Z2 (7 TeV) Z2* (8 TeV)
Powheg	Pythia 6	cteq6m (7 TeV) CT10 (8 TeV)	Z2 (7 TeV) Z2* (8 TeV)
MC@NLO v3.4	Herwig 6 + Jimmy	cteq6m	default tune

$m_t = 172.5$ GeV

- Matrix Element + Parton Shower generators
  - Better description of high multiplicities
  - Initial and final state radiation (ISR/FSR) modelling via ME from assumed Q^2 variation
  - Matching procedure to remove double counting between partons produced by ME and PS
- Next to Leading Order generators
  - More accurate in normalisation
  - Smaller uncertainty on Q^2

Radiative corrections

- The Q^2 scale variations address two aspects:
  - renormalisation and factorisation scale (ME)
  - amount of ISR/FSR



- For each event, Q^2 is defined as:
  - MadGraph: $Q^2 = m_t^2 + \sum p_T^2$
  - Powheg/MC@NLO: $Q^2 = m_t^2$
- Parton showering:
  - p_T-ordered evolution scale of ISR/FSR
  - shares Q^2 factor α_s with ME (starting scale changes with ΔQ^2)
- MadGraph(+Pythia) is the default MC
  - tree-level diagrams for hard radiation and interferences (up to 3 final-state partons for $t\bar{t}$)
  - parton showering for soft and collinear region (with Pythia 6.42X)
  - matching with ktMLM, thresholds varied by a factor 0.5 and 2.0 (nominal = 20 GeV)

$\rightarrow$ Uncertainty on radiation covered by variations of Q^2 and ME-PS matching

Event selection

Dilepton

1. Dilepton triggers
2. At least two isolated leptons ($p_T > 20$ GeV, $|\eta| < 2.4$), opposite sign
3. ≥ 2 jets (anti-kT, R<0.5) with $p_T > 30$ GeV, $|\eta| < 2.4$
4. QCD veto: $M_{ll} >$ 12-20 GeV
5. In *ee*, $\mu\mu$: Z veto ($76 < M_{ll} < 106$ GeV), $E_T^{miss} >$30-40 GeV
6. At least one b-tagged jet

Lepton+jets

1. Single muon or electron+jets trigger
2. One isolated lepton ($p_T > 30$ GeV, *e*: $|\eta| < 2.4$, μ: $|\eta| < 2.1$)
3. ≥ 3 jets (anti-kT, R<0.5) with $p_T > 35$ GeV, $|\eta| < 2.4$ (≥ 4 jets with $p_T > 30$ GeV - additional jet measurement)
4. Loose lepton veto in both channels
5. At least two b-tagged jets

Selection and analysis strategy

Dilepton channels

- Event reconstruction: Kinematic reconstruction of the $t\bar{t}$ system
- Background estimation:
    - Z/γ^*+jets estimated from data
    - Other BGs (single top, dibosons, etc) estimated from MC

Lepton+jets channels

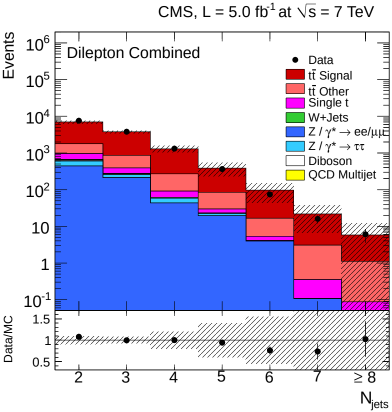
- Background estimation
    - W+jets estimated from data
    - QCD: data driven
    - Single top, Z/γ^*+jets and diboson are from MC

$\rightarrow$ Signal: $t\bar{t}$ MadGraph+Pythia (normalised to NNLO+NNLL)

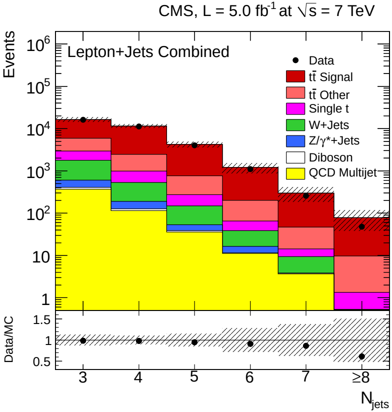
Control Plots: Reconstructed jet multiplicity

7 TeV, dilepton: p_T >30 GeV, l+jets: p_T >35 GeV

Dilepton



l+jets



Good description of data within uncertainties. (Similar description with 8 TeV data)

Normalised differential cross sections

$$\boxed{\frac{1}{\sigma_{t\bar{t}}}\frac{d\sigma^i}{dN_j} = \frac{1}{\sigma_{t\bar{t}}}\frac{x^i}{\Delta_X^i\, L}}$$

- x^i number of events after background subtraction, corrected for detector efficiencies, acceptances and migration to particle level (regularised unfolding).
- $\sigma_{t\bar{t}}$ inclusive $t\bar{t}$ cross section in the same phase space (visible).
- Δ_X: bin width (=1)

Measurement done in the visible phase space:

- p_T^l >20(30) GeV, $|\eta^\mu|$ <2.4 (2.1) dilepton (l+jets), $|\eta^e|$ <2.4
- p_T^{jet} >30(35) GeV dilepton (l+jets), $|\eta^{jet}|$ <2.4, jets required $\Delta R(j,l)$ >0.4, b-jets identified by B-hadron

Comparing results to predictions from:

- Different generators (POWHEG+Pythia, MC@NLO+Herwig)
- MadGraph+Pythia with varied Q^2 scale, matching threshold

Systematic uncertainties

- Sources considered:
  - Jet energy scale and resolution
  - Background estimate
  - Model uncertainties: Q^2 scale (using samples with 2*Q, 0.5*Q), jet-parton matching threshold (threshold halved/doubled), hadronisation model, the color reconnection modelling and PDF
  - Other sources: luminosity, PU, b-tagging, lepton Id/Iso, trigger
- Uncertainties determined individually for each bin.
- Normalised cross sections: bin-to-bin correlated uncertainties cancel (luminosity, flat SF, etc.), only shape uncertainties contribute (shape variation for b-tag, BG, scale, hadronisation etc.)

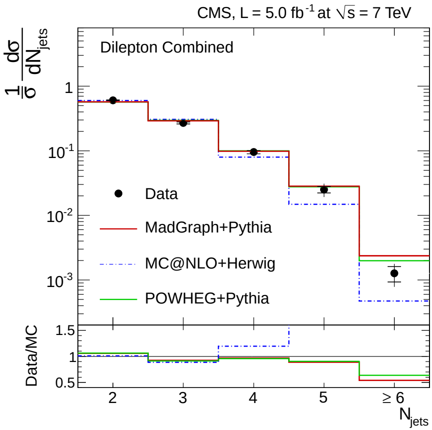
  Total syst. uncertainty varies from 3-6% in the low multiplicity bins to 20-30% for the highest multiplicities.

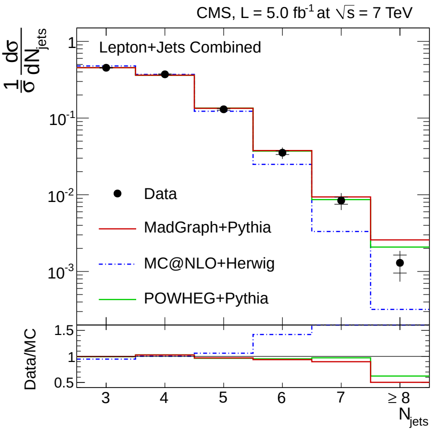
  Most important sources: JES and Q^2 scale and ME/PS Matching

Normalised diff. $\sigma_{t\bar{t}}$ as a function of NJets

Results 7 TeV: dilepton and l+jets channels

- MadGraph+Pythia, POWHEG+Pythia provide a reasonable description
- MC@NLO+Herwig doesn't describe large jet multiplicities



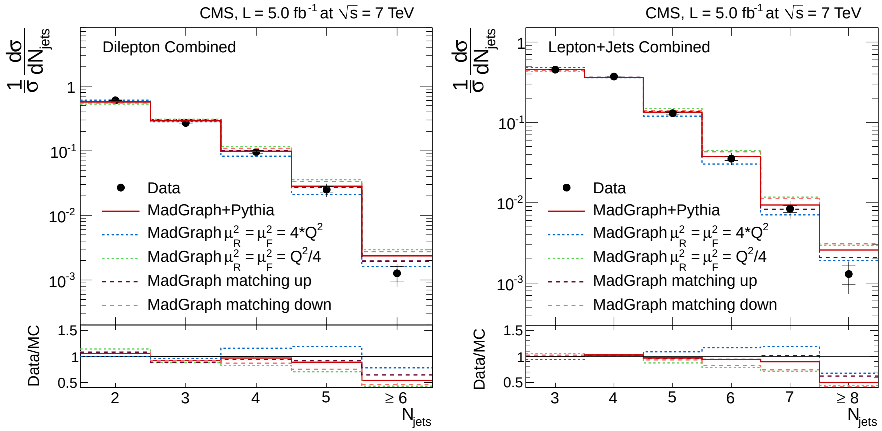


Results are consistent among channels (dilepton, l+jets)

Normalised diff. $\sigma_{t\bar{t}}$ as a function of NJets

Results 7 TeV: dilepton and l+jets channels

- Comparison with MadGraph, varying Q^2 scale and jet/parton matching threshold up and down



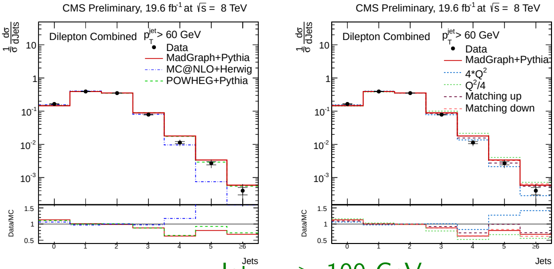
◇ Choice of lower scale gives slightly worse description of the data.

Normalised diff. $\sigma_{t\bar{t}}$ as a function of NJets

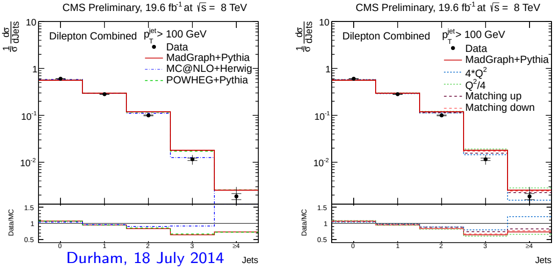
Results 8 TeV: dilepton for jets $p_T > 60$ GeV, $p_T > 100$ GeV

- Larger data samples allow to measure higher p_T
  - MadGraph+Pythia, POWHEG+Pythia provide a reasonable description, MC@NLO+Herwig describes data better for high jet-p_T.
  - Choice of lower Q^2 scale gives slightly worse description of the data.
  - Results consistent with 7 TeV measurement.

Jet $p_T > 60$ GeV



Jet $p_T > 100$ GeV

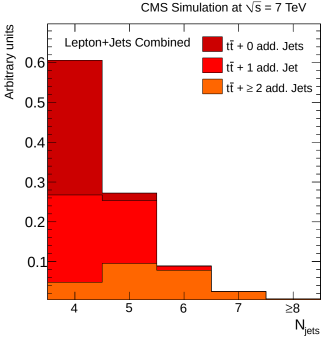


Cross section as a function of the additional jet number

7 TeV: Lepton+jets channel

- Categorise every $t\bar{t}$ MC event as a function of the number of generated jets NOT matching any of top decay products (b quarks, light quarks and prompt lepton)
- Jets with $\Delta R > 0.5$: additional radiated jets

  $\rightarrow$ classification of events in $t\bar{t}+0,1,\geq 2$ additional jets
- Extracting rates of these $t\bar{t}$ classes from data via a template fit
- Selection: similar to previous section, but selection events with at least 4 jets with $p_T >$ 30 GeV.



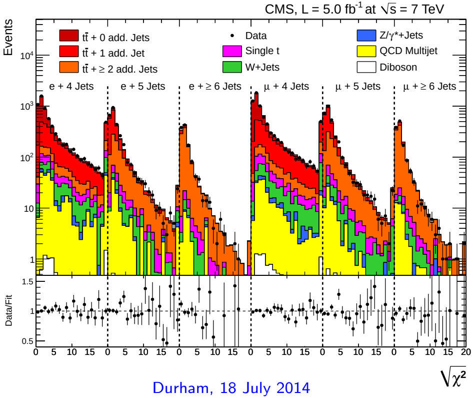
Template fit results

- Simultaneous fit to data in three jet multiplicity bins in e/μ+jets channels
- Templates built from smallest χ of any jet combination (using b-tag info)

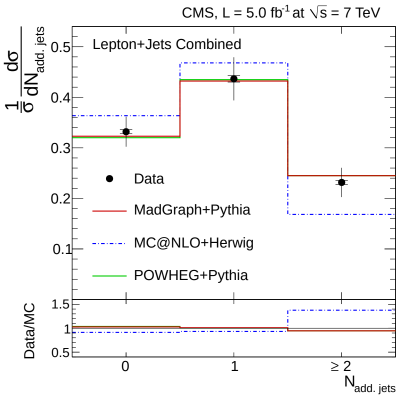
$$\chi = \sqrt{\left(\frac{m_{Whad}^{rec} - m_{Whad}^{true}}{\sigma_{Whad}}\right)^2 + \left(\frac{m_{thad}^{rec} - m_{thad}^{true}}{\sigma_{thad}}\right)^2 + \left(\frac{m_{tlep}^{rec} - m_{tlep}^{true}}{\sigma_{tlep}}\right)^2}$$

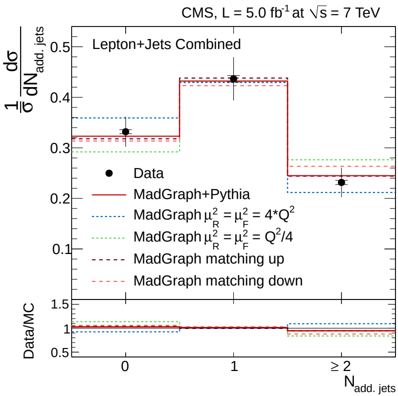
from full event reconstruction

- χ lower if all jets from the $t\bar{t}$ decay are reconstructed



Diff. cross section: additional jet multiplicity





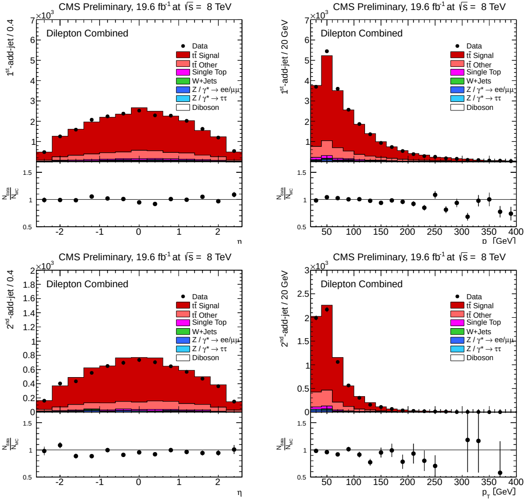
- Systematic uncertainties evaluated with pseudoexperiments
- Best agreement with MadGraph+Pythia and POWHEG+Pythia, MC@NLO+Herwig shows discrepancies

Kinematic distributions for the leading p_T additional jets

- Additional jets defined as the jets in visible PS NOT selected by the kinematic reconstruction

1^{st} leading add. jet

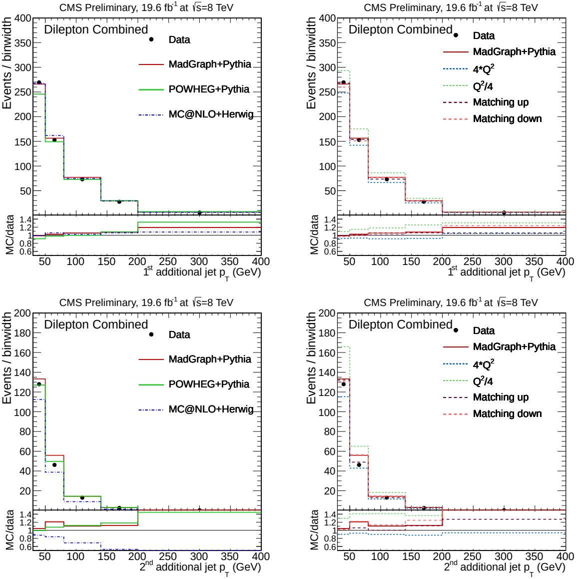
2^{nd} leading add. jet



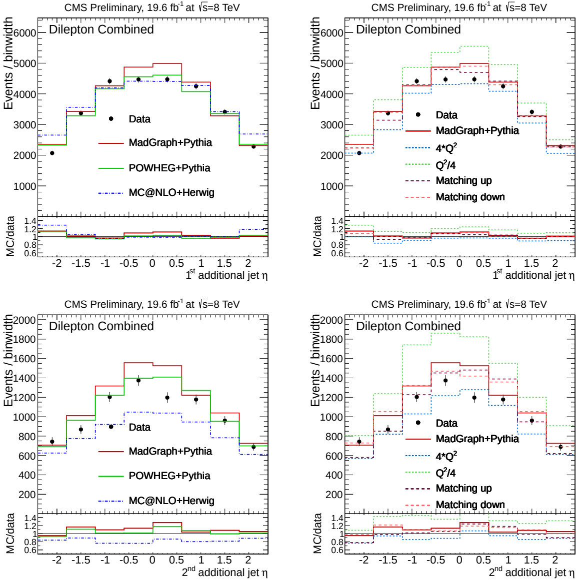
Good agreement between data and MC.

p_T of 1^{st} and 2^{nd} leading additional jets (8 TeV)

- Comparison with various theory predictions
- Distributions at reconstructed level (no unfolding applied!), background is subtracted using MC predictions
- All predictions normalised to in situ measured cross-section (with MadGraph)



η of 1^{st} and 2^{nd} leading additional jets (8 TeV)



- MadGraph+Pythia, Powheg+Pythia provide a reasonable description
- MC@NLO+Herwig fails to describe 2^{nd} add. jet distributions
- Lower scale (Q^2) sample predicts lower yields than data

$t\bar{t}$ production with a veto on additional jet activity

gap fraction - dilepton channel

- Jet activity arising from quark and gluon radiation produced with the $t\bar{t}$ system is quantified with a jet veto

$$f(p_T) = \frac{N(p_T)}{N_{total}} \tag{1}$$

$N(p_T)$ are the events without 1 (2) additional jets with p_T above a threshold (sensitive to the (2^{nd}) leading-p_T emission).

- The veto can be extended by defining it as

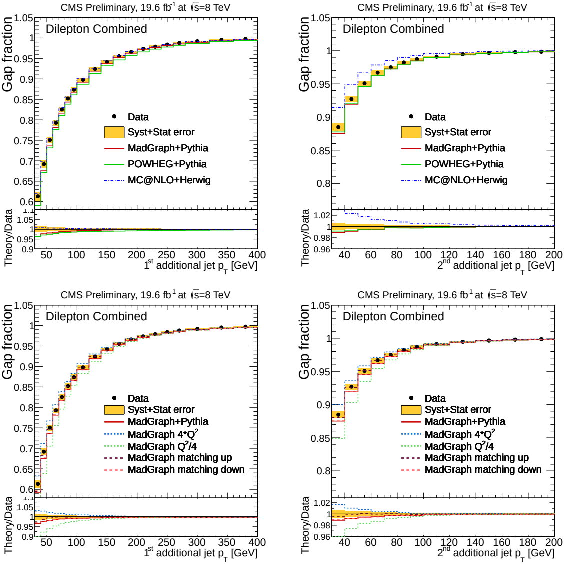
$$f(H_T) = \frac{N(H_T)}{N_{total}} \tag{2}$$

$N(H_T)$ is the number of events in which the scalar sum of the p_T of the additional jets is less than a certain threshold, $H_T = \sum p_T^{add.jets}$, (sensitive to all hard emission accompanying $t\bar{t}$ system).

Gap fraction in data corrected for detector effects to particle level using MadGraph: for each value of the threshold the ratio of the "true" and "reconstructed" simulated gap fraction is computed and applied to data.

Gap fraction (1^{st} add. jet, 2^{nd} add. jet p_T)

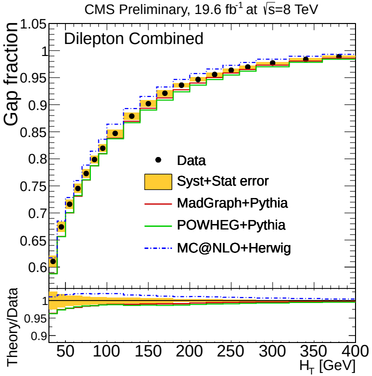
Full rapidity range

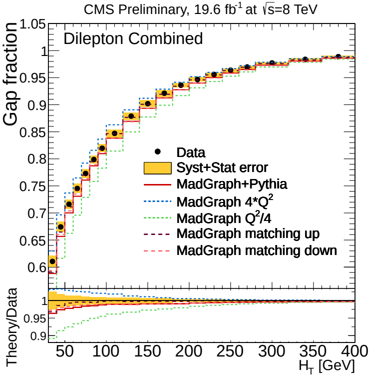


- MadGraph+Pythia, Powheg+Pythia provide a reasonable description for both jet-p_T threshold.
- Gap fraction as a function of 1^{st} add. jet p_T better described with MC@NLO+Herwig.
- MadGraph with decreased Q^2 scale predicts lower gap fraction values than the measured ones.

Gap fraction as a function of H_T

$H_T = \sum p_T^{add.jets}$

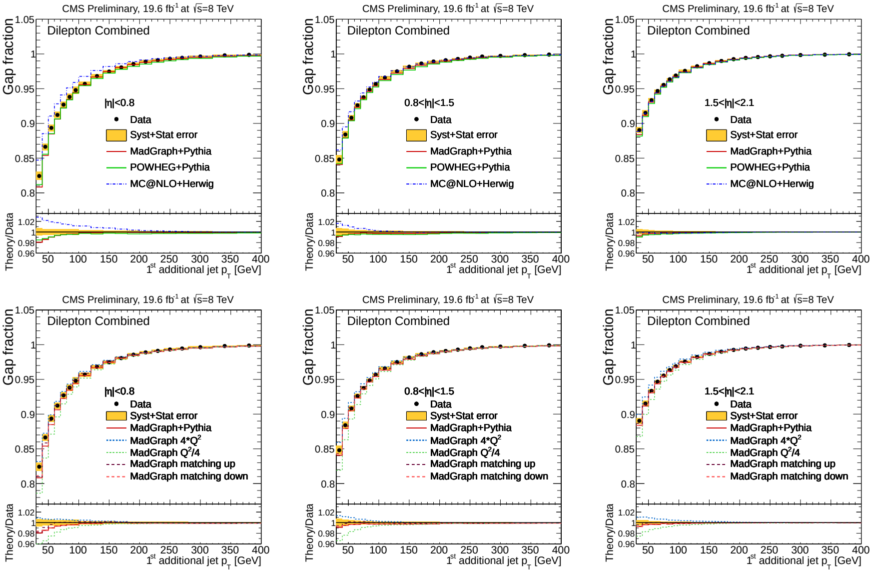




- Reasonable agreement between MadGraph+Pythia, Powheg+Pythia and data.
- Gap fraction slightly better described by MadGraph with increased matching threshold.

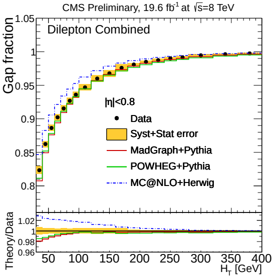
Gap fraction: p_T first additional jet

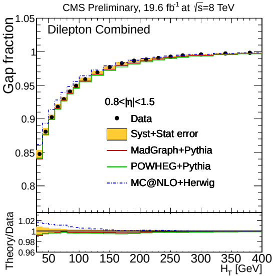
Different rapidity ranges

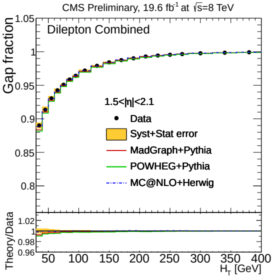


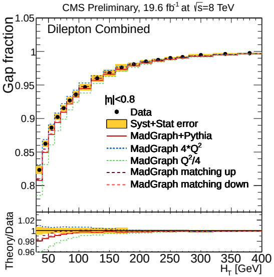
Gap fraction as a function of H_T

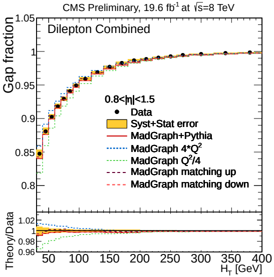
Different rapidity ranges

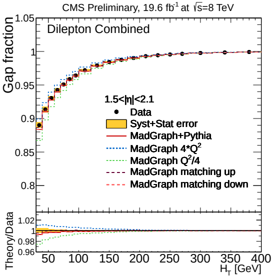












Summary and Outlook

- Presented $t\bar{t}$ production with additional jet activity in the l+jets and in the dilepton channels. (All measurements in visible phase space, corrected to particle level.)
  - ◇ Normalised $t\bar{t}$ production cross section as a function of jet multiplicity and additional jet multiplicity
  - ◇ Kinematics of the additional jets
  - ◇ Gap fraction
- Compared to different MCs and parameter variations from Madgraph.
- In general, good agreement between data-MC observed with different predictons.
- Consistent result among channels and measurements.
- Often, experimental precision smaller than spread due to parameter variation.
- Working towards comparisons with other predictions (Powheg+Herwig) and NLO+Parton Showering multileg generators like aMC@NLO.

BACK UP

Top quarks

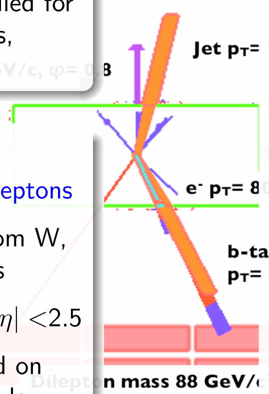
MET

- Typical cut range 20-40 GeV, not applied for all analysis (lep+jets, dilepton)

Leptons

- Up to two high p_T leptons
- Isolated, high p_T from W, soft leptons in b-jets

  With p_T >20 GeV $|\eta|$ <2.5
- Trigger largely based on leptons (Single/double (isolated) lepton)



Jets

- Two to six high p_T jets (up to 2 b-tags, might use τ-tagging)
- Jets defined with anti-kT algorithm with cone ΔR=0.5
- p_T >20-30 GeV $|\eta|$ <2.5
- Trigger on l+jets signatures

b-tagging

- Uses secondary vertices and/or IP information
- Efficiencies and fake rates are calibrated by using data

Selection and analysis strategy

- Full 2011 data (7 TeV, 5.0±0.1 fb^{-1}) and 2012 data (8 TeV, 19.7±0.1 fb^{-1})

◇ Signal: $t\bar{t}$ MadGraph+Pythia (normalised to NNLO+NNLL)

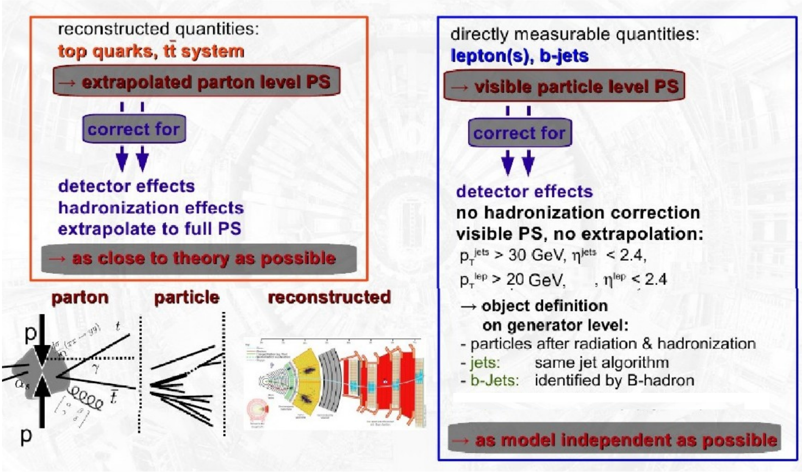
  - For comparison:
    POWHEG+Pythia and MC@NLO+Herwig
    MadGraph+Pythia with varied hadronisation/renormalization scale (Q^2) and jet-parton matching threshold.

◇ Backgrounds:

  - Z/γ^*+jets: MadGraph+Pythia (dominant BG ee, $\mu\mu$ channels)
  - W+jets: MadGraph+Pythia (dominant BG l+jets)
  - Single top (s-, tW-channel): POWHEG+Pythia
  - Diboson (WW,WZ,ZZ): Pythia
  - QCD: Pythia (data driven for l+jets)

Phase space definitions

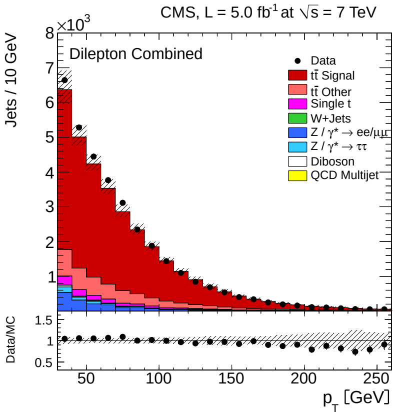
- Measurements are presented at particle level in the visible phase space.
  Additionally GenJets must fullfill: ΔR(genJet,selected leptons)>0.4



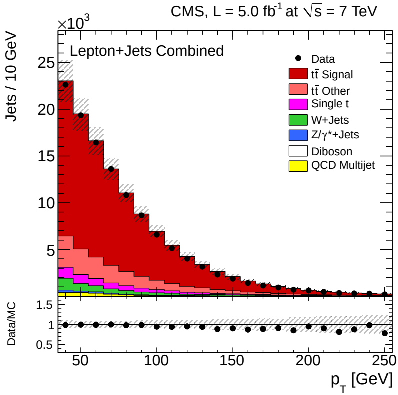
Control Plots: Jet p_T

Dilepton: p_T >30 GeV, l+jets: p_T >35 GeV

Dilepton



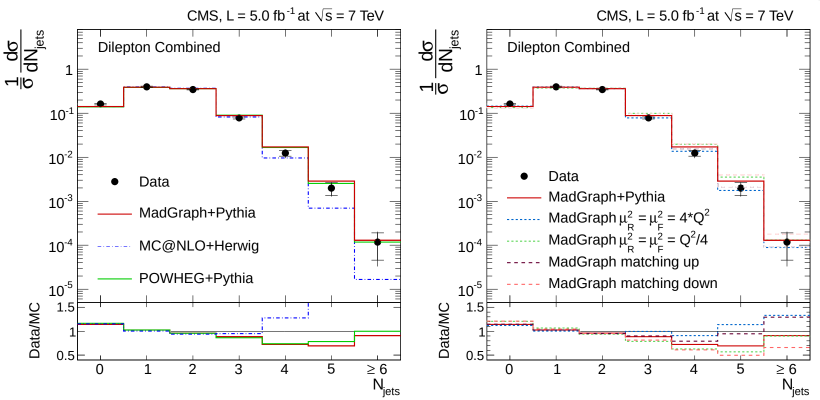
l+jets



Good description of data within uncertainties.

Results: dilepton channel

Jet $p_T > 60$ GeV

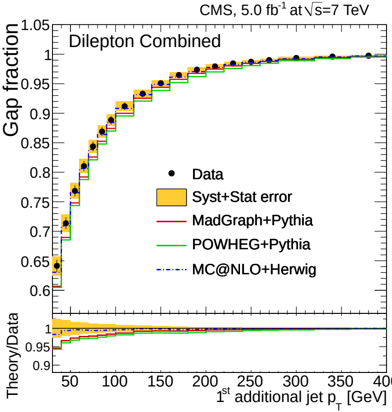


- MadGraph+Pythia, POWHEG+PYTHIA provide a reasonable description
- MC@NLO+HERWIG doesn't generate large jet multiplicities.
- Choice of lower scale gives slightly worse description of the data.

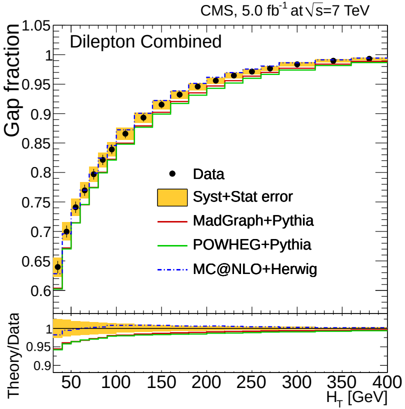
Gap fraction: results 7 TeV

Gap fraction in data corrected for detector effects to particle level using MadGraph

p_T



H_T

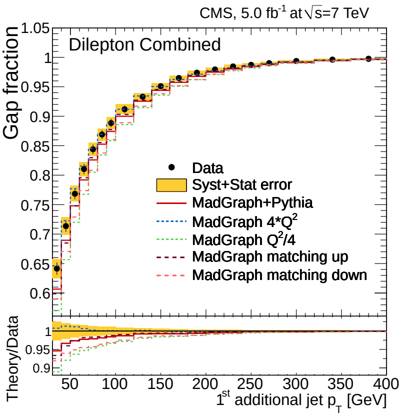


Gap fraction better described by MC@NLO+HERWIG

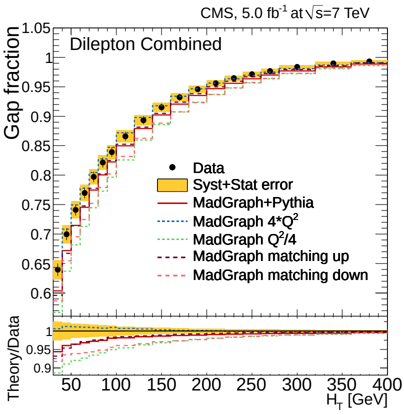
Dominant systematic uncertainties: JES uncertainty, BG contamination

Gap fraction: results 7 TeV

p_T



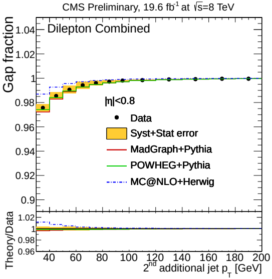
H_T

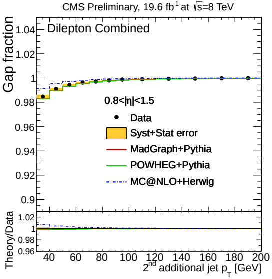


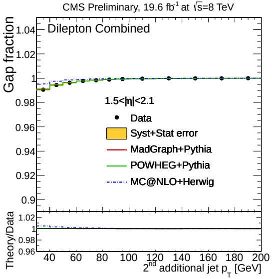
Decreasing the Q^2 scale, jet-parton matching threshold worsens the agreement with data

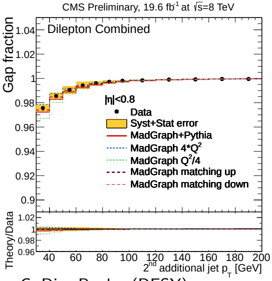
p_T second additional jet

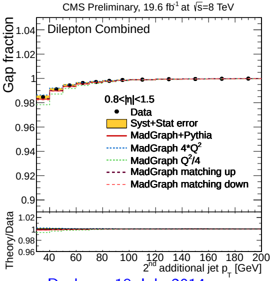
- Gap fraction for different rapidity ranges.

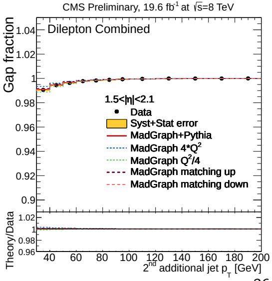










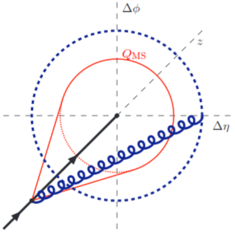


Matching scale (MadGraph only)

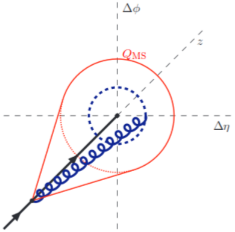
MadGraph uses

- tree-level diagrams for hard radiation & interferences (up to 3 final state partons)
- parton-showering for soft & collinear region (with Pythia 6.42x)
- matching via ktMLM (k_T instead of cone algorithm)

The idea:



ME emission



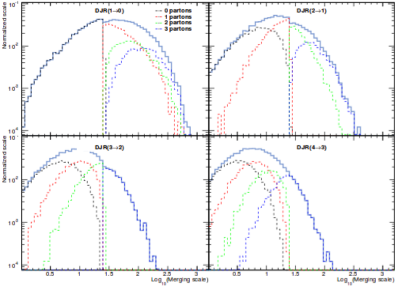
PS emission

(for illustrative purposes only)

Matching scale (MadGraph only)

MadGraph uses

- tree-level diagrams for hard radiation & interferences (up to 3 final state partons)
- parton-showering for soft & collinear region (with Pythia 6.42x)
- matching via ktMLM (k_T instead of cone algorithm)



- study smoothness N-jet vs N+1-jet transitions (y-axis) as function of scale
- matching thresholds (xqcut) drives the optimal matching scale (qcut)

- default xqcut = 20 (qcut = 40) for $t\bar{t}$+jets
- thresholds are varied by factor 0.5 to 2.0

Martijn Gosselink (Universität Hamburg) Signal modelling uncertainties in CMS September 18, 2012 6 / 11