Top and top-pair mass measurements at CMS

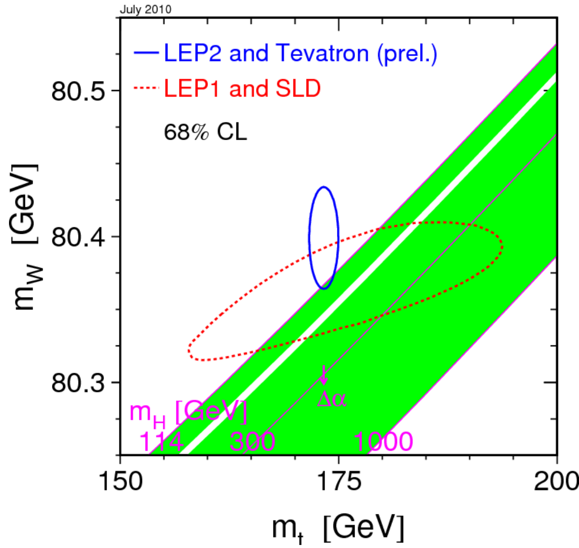
Roberto Chierici
(CNRS/IPN Lyon)

Outline

- Introduction
- The top mass in the di-lepton channel (TOP-10-006)
  - Experimental challenges at the LHC
- The top-pair mass in the semi-leptonic channels
  - Standard reconstruction techniques (TOP-10-007)
  - Tools for boosted top reconstruction (JME-10-013)
- Outlook

Introduction

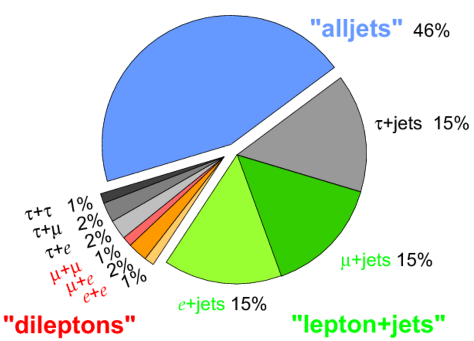
- Top physics is one of the main pillars of the physics program at the LHC
  - ➢ The top quark is intriguing
    - o The heaviest fundamental particle known
    - o The only quark not hadronizing
  - ➢ The top mass is a fundamental parameter of the Standard Model (SM)
    - o Whose precise knowledge allows to constrain the model itself and predict the Higgs boson mass (in the frame of the SM)



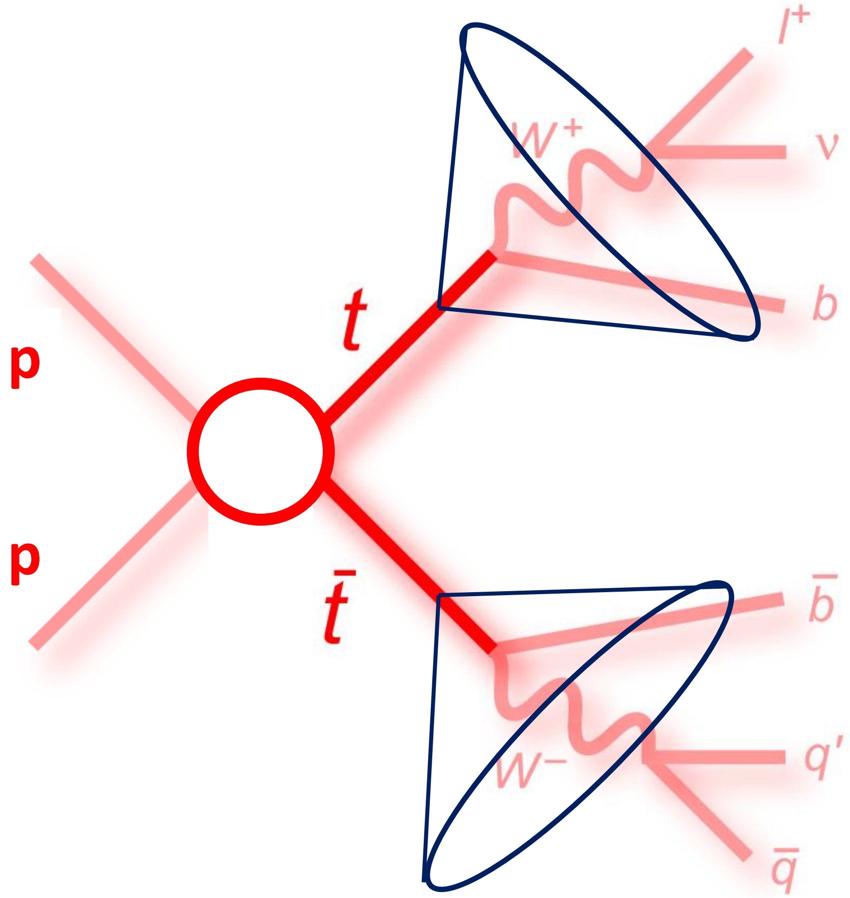
- The top quark represents a potential portal to physics beyond the SM
  - ➢ "Strongly" coupled to the Higgs/EWSB sector
  - ➢ Many models predict favorable couplings to the third family

- Top physics needs a full understanding of all detector components
  - ➢ Muons, electrons, MET, jets
  - ➢ The hadronic component of top-pair events is particularly crucial for the full event reconstruction
    - o Jet energy scale and resolution
    - o Jet pairing
    - o Heavy flavour tagging
    - o Jet substructure reconstruction

Top Pair Branching Fractions



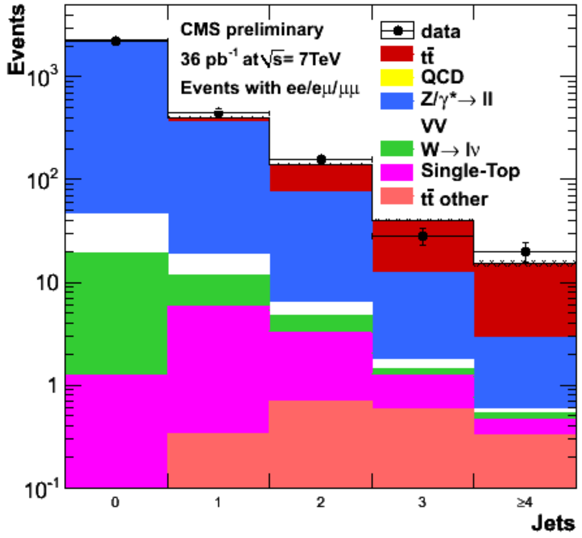
The top mass

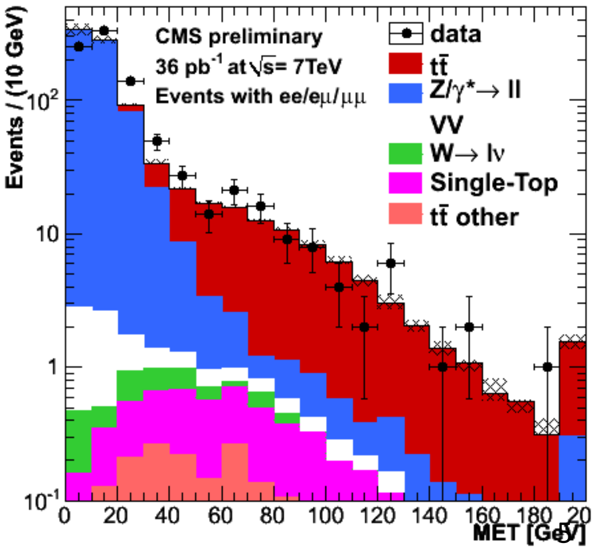


Top mass: event selection

- CMS performed the first measurement of the top mass outside Tevatron
  - ➢ Di-lepton channel (low cross-section, but more background free)

- Event selection is straightforward
  - ➢ 2 isolated, prompt, opposite charge leptons with p_T>20GeV/c and $|\eta|$<2.5
  - ➢ For same flavour leptons, $|m(\ell\ell)-m_Z|$<15GeV/c²
  - ➢ Two or more jets with p_T>30GeV/c and $|\eta|$<2.5
    - o b-tagging is not used for selection, but used for ranking the jets which enter the mass reconstruction
  - ➢ MET>30(20) GeV for the ee/μμ (eμ) channels

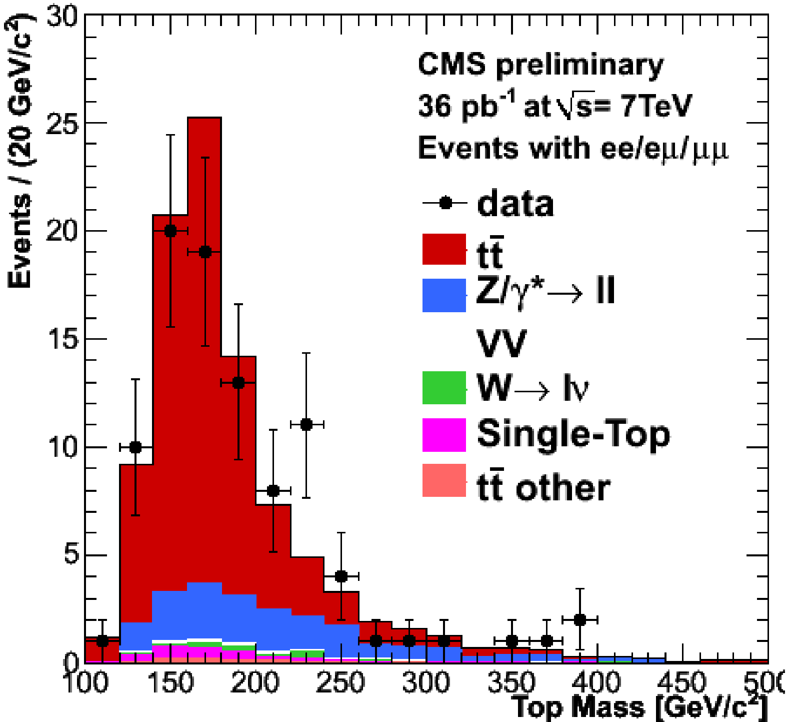


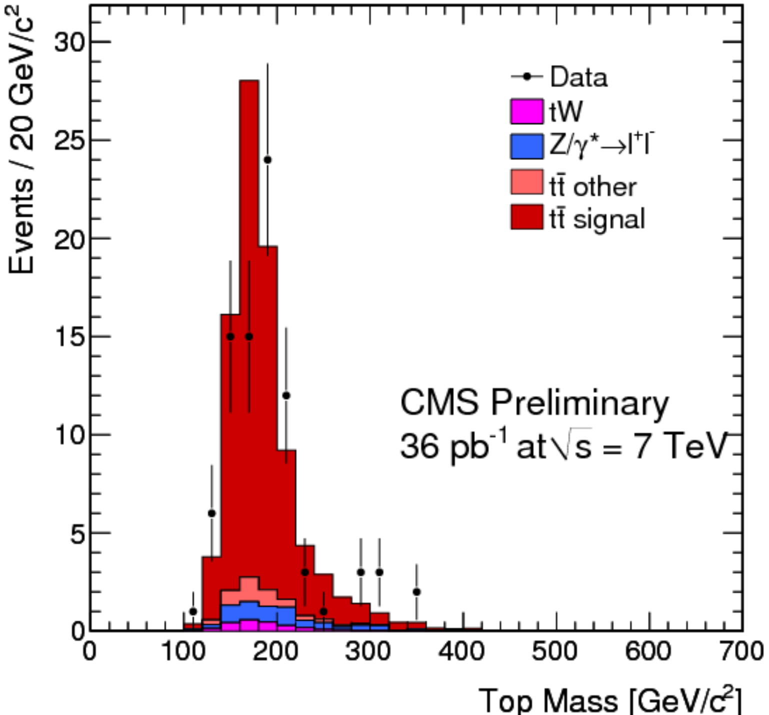


Selection cut	Data	Total expected	$t\bar{t}$ signal	Total background
		pre-tagged sample		
≥2 isolated leptons	27257	28934 ± 49	158.8 ± 0.9	28775 ± 49
opposite sign	26779	28545 ± 42	157.3 ± 0.9	28388 ± 42
Z/quarkonia-veto	2878	2873 ± 27	139.3 ± 0.8	2734 ± 27
≥2 jets	204	193 ± 2	103.1 ± 0.7	90 ± 2
$\not\!\!E_T$	102	$108.5 \pm 0.9\,^{+3}_{-2}$	$92.1 \pm 0.7\,^{+2}_{-1}$	$16.3 \pm 0.7\,^{+1}_{-1}$

Top mass: event reconstruction

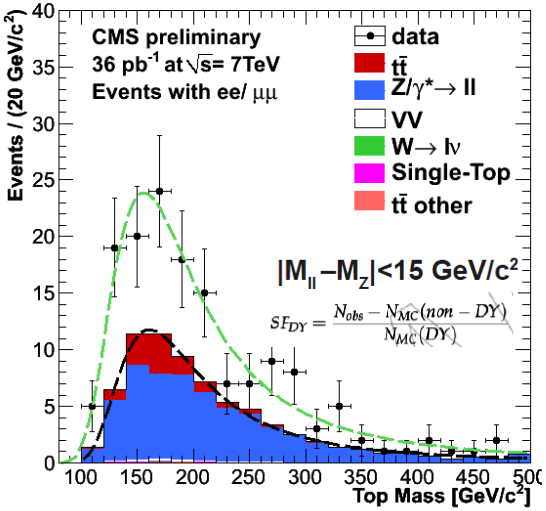
- Two partially independent methods solving the event equation
  - ➢ Impose equality of “top” masses and the W mass constraints
  - ➢ Do it for every lepton-jet combination in the event
    - o Favour b-tagged jets
  - ➢ Iterate for several mass hypotheses, keep the highest weight solution



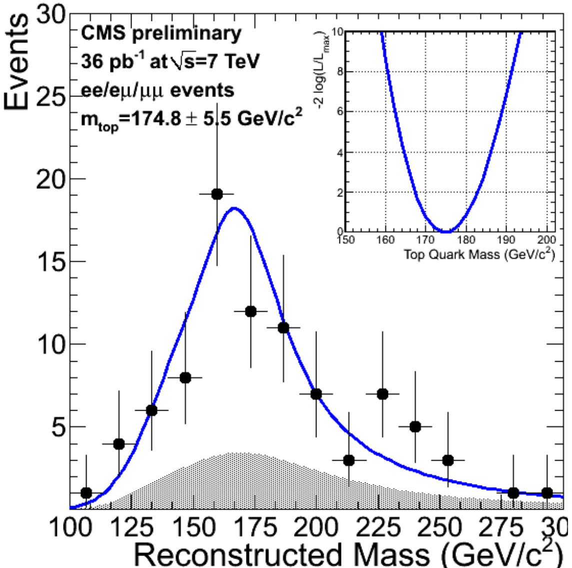


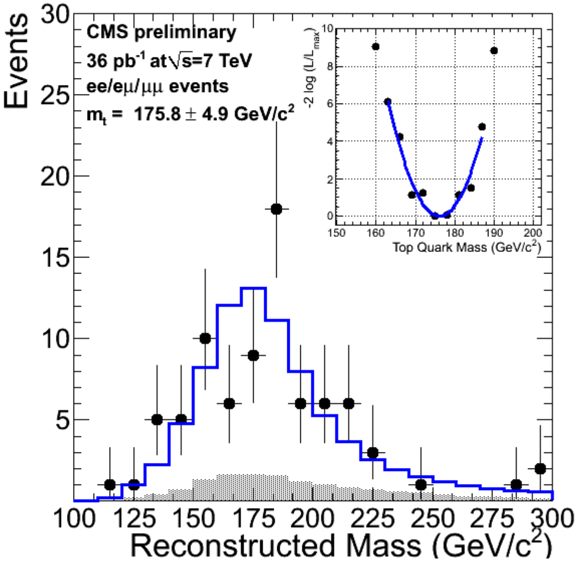
Top mass: template fitting

- Use template fitting to extract m_t from the mass distributions
  - ➢ Signal is taken from MC predictions at different m_t
  - ➢ Background parametrized with MC and data
    - ○ Single top, tt, W+jets, di-boson from MC
    - ○ Z+jets in di-leptons from data
      - – Scale factor from mass distributions inside the Z peak



- Apply the likelihood fit to the data
  - ➢ Combine 0, 1, >=2 btag in the fit
  - ➢ Methods crosschecked to be linear in m_t and with small bias (corrected for)
  - ➢ The analyses provide compatible results

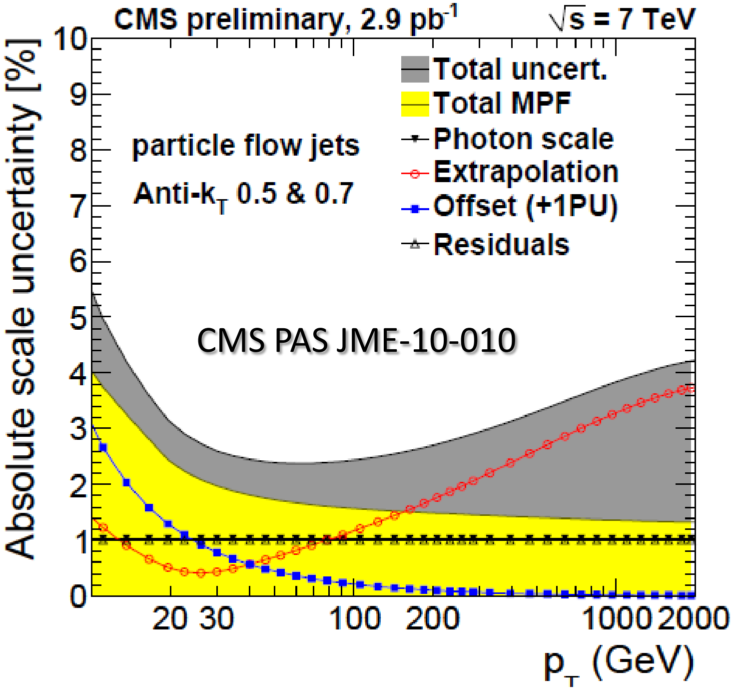




Top mass: results

- JES is the most relevant systematic error
  - Flavour specific uncertainties accounted for

- MC modelling also accounted for
  - Radiation, PS-ME matching thresholds
  - Different generators, UE tunes, Pile-up

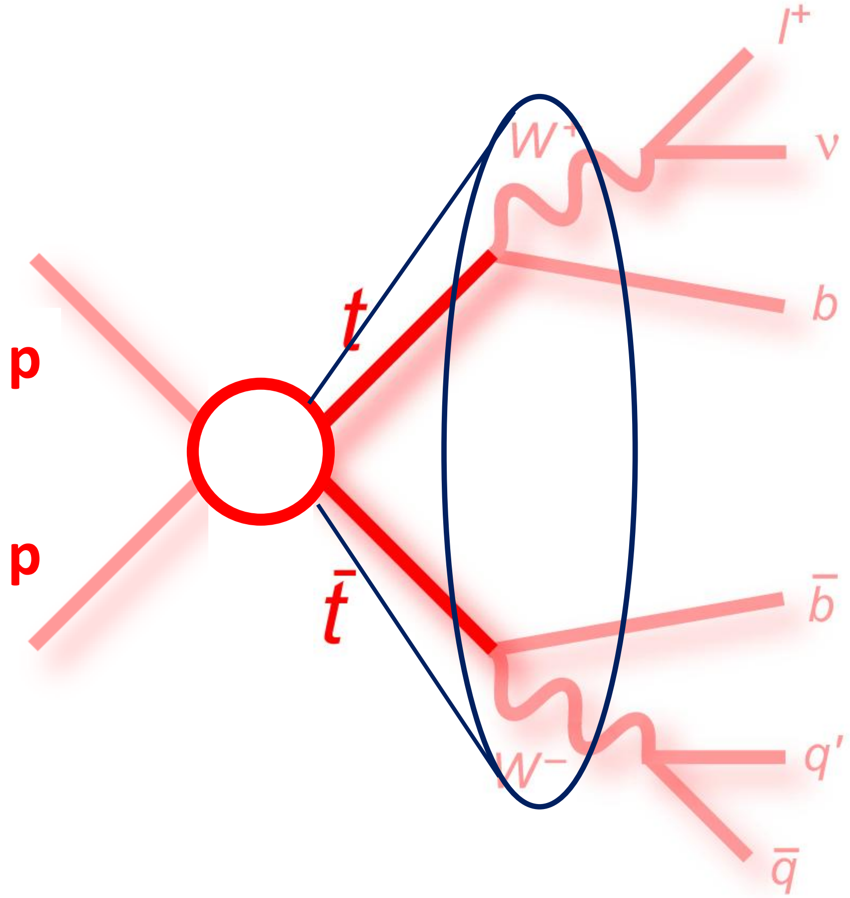


Source	KINb	AMWT
jet energy scale	+3.1/-3.7	3.0
b-jet energy scale	+2.2/-2.5	2.5
Underlying event	1.2	1.5
Pileup	0.9	1.1
Jet-parton matching	0.7	0.7
Factorization scale	0.7	0.6
Fit calibration	0.5	0.1
MC generator	0.9	0.2
Parton density functions	0.4	0.6
b-tagging	0.3	0.5

- Analyses are combined by using BLUE
  - Statistical correlation determined via pseudo experiments to be 0.57
  - Statistical and systematic errors already of the same size

Method	Measured m_{top} (in GeV/c^2)	Weight
AMWT	$175.8 \pm 4.9(stat) \pm 4.5(syst)$	0.65
KINb	$174.8 \pm 5.5(stat)^{+4.5}_{-5.0}(syst)$	0.35
combined	$175.5 \pm 4.6(stat) \pm 4.6(syst)$	χ^2/dof=0.040 (p-value=0.84)

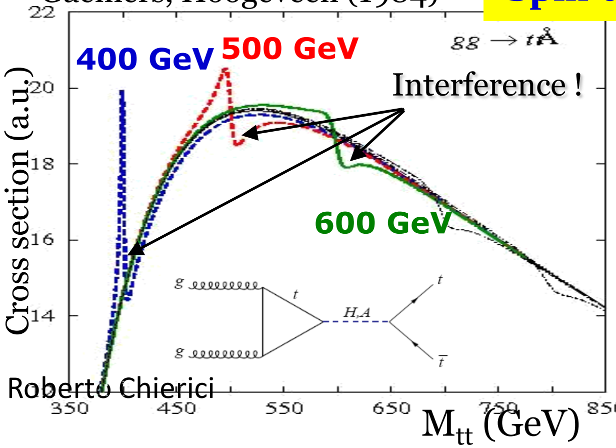
The top-pair mass



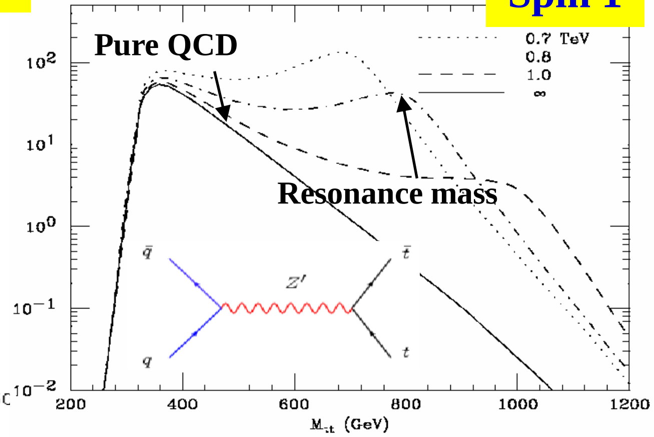
Why is that interesting?

- Studying m(tt) is particularly important in many respects:
  - ➢ As a measure within the SM
    - o top pair kinematics
    - o indirect probe of the top mass
  - ➢ Undiscovered heavy s-channel resonances can decay to a pair of top quarks
    - o MSSM Higgs (spin 0) (H/A, if mH,mA>2mt, BR(H/A→tt)≈1 for tanβ≈1)
    - o Technicolor, strong EW SB, Topcolor (spin 1)
    - o KK excitations (spin 2)
  - ➢ Distortions in the top pair mass distributions are predicted by other models
    - o Associated production of invisible scalars, SUSY, …

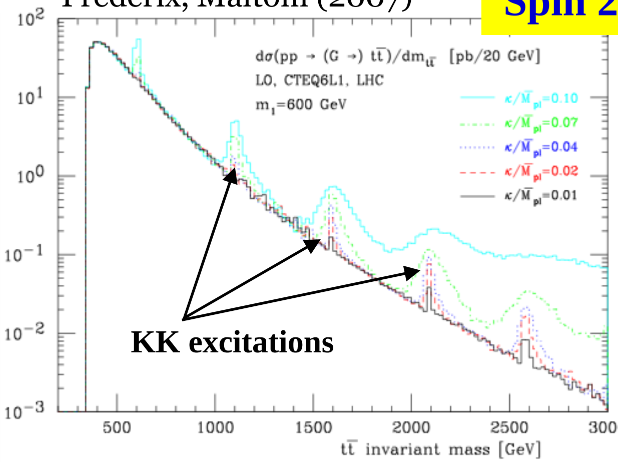
Gaemers, Hoogeveen (1984) **Spin 0**



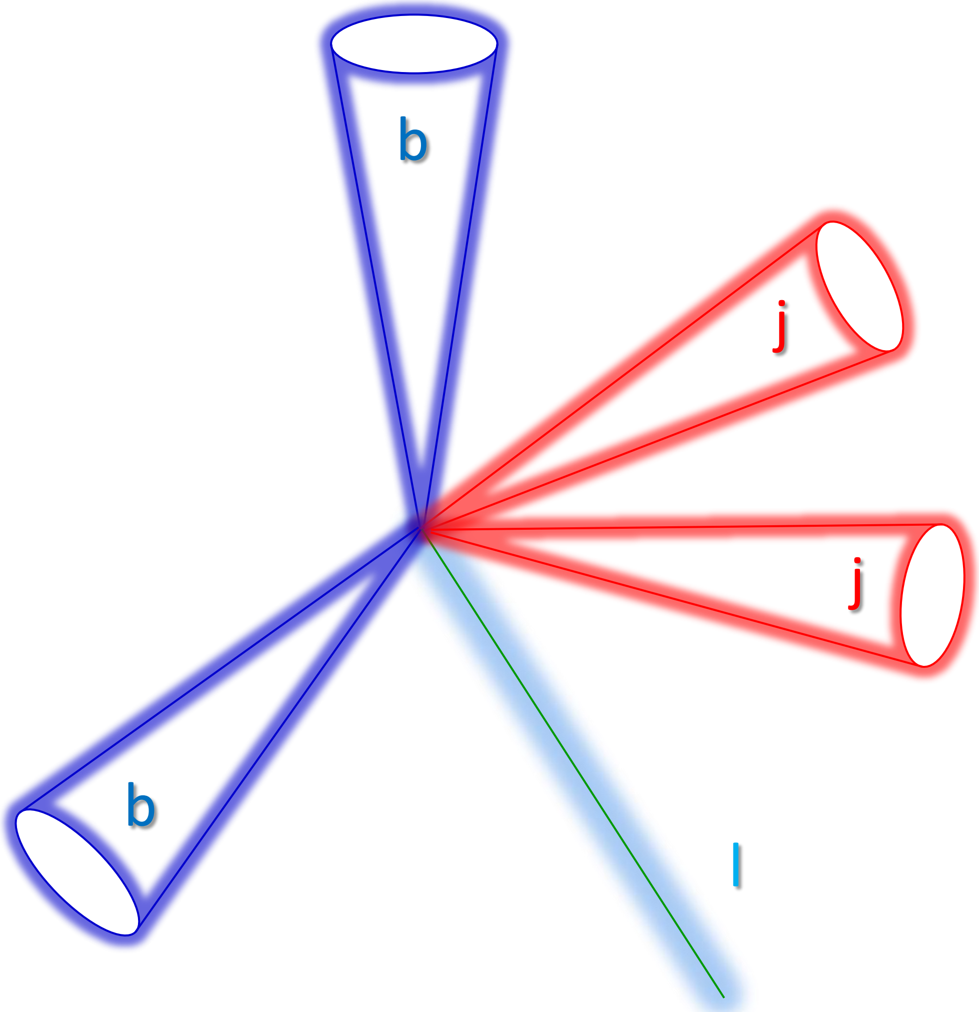
Hill, Parke (1994) **Spin 1**



Frederix, Maltoni (2007) **Spin 2**



Standard reconstruction -moderate top boost-



Event selection and yields

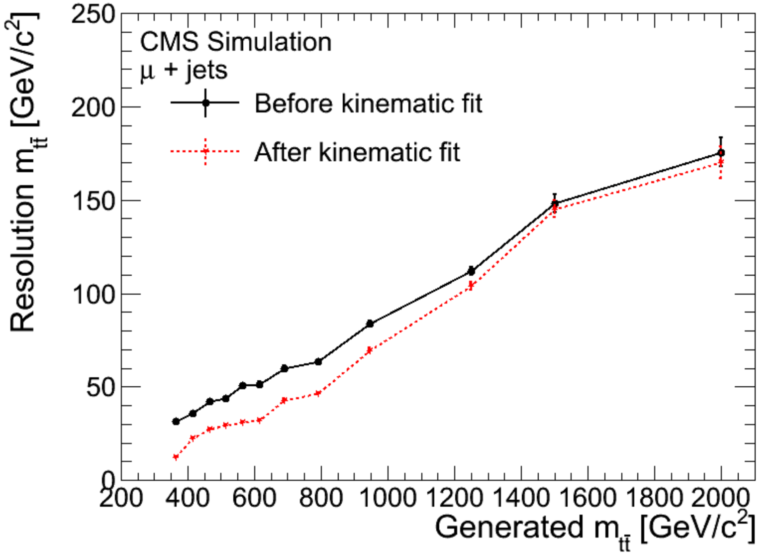
- Focus on semi-leptonic events with an electron or a muon
  - Single muon/electron trigger and good primary vertex
  - One isolated lepton in the acceptance with p_T>20GeV/c (30 GeV/c for electrons), veto on a second isolated lepton
  - At least three (four) jets with p_T>70/50/30(/30)GeV/c and $|\eta|<2.4$
  - MET>20 GeV

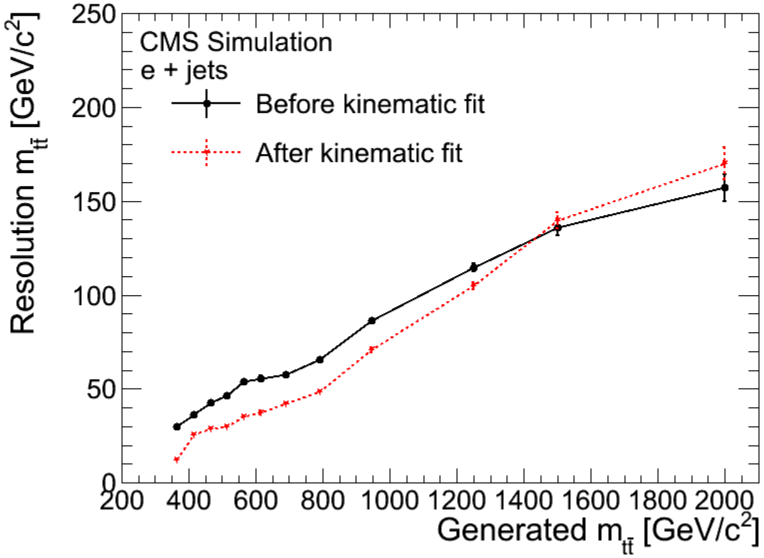
- Yields in agreement with the expectations
  - Divided into jet+btag (SV) bin multiplicities
  - All then taken from MC, with exception for QCD, taken from data estimates

Yields	$t\bar{t}$	W/Z+LF	W/Z+HF	Single-top	QCD	Data	Sum BG
μ 3j1t	96.9 ± 0.6	7.9 ± 0.2	28.6 ± 1.1	11.6 ± 0.1	8.2 ± 8.2	142 ± 11.9	153.2 ± 8.3
μ 4j0t	40.4 ± 0.5	62.8 ± 2.2	25.0 ± 1.0	2.5 ± 0.1	4.5 ± 4.5	107 ± 10.3	135.1 ± 5.1
μ 4j1t	84.8 ± 0.6	3.8 ± 0.1	12.5 ± 0.7	4.2 ± 0.1	5.1 ± 5.1	112 ± 10.6	110.5 ± 5.2
μ 4j2t	51.6 ± 0.4	0.1 ± 0.0	2.4 ± 0.2	2.0 ± 0.0	1.0 ± 1.0	58 ± 7.6	57.1 ± 1.1
e 3j1t	80.3 ± 0.6	5.4 ± 0.1	22.8 ± 1.0	8.5 ± 0.1	9.4 ± 9.4	114 ± 10.7	126.4 ± 9.5
e 4j0t	31.8 ± 0.4	47.0 ± 1.9	19.1 ± 0.9	1.9 ± 0.0	10.8 ± 10.8	106 ± 10.3	110.4 ± 11.0
e 4j1t	66.7 ± 0.5	2.8 ± 0.1	9.0 ± 0.6	3.2 ± 0.1	3.0 ± 3.0	80 ± 8.9	84.7 ± 3.1
e 4j2t	40.9 ± 0.4	0.1 ± 0.0	2.1 ± 0.2	1.5 ± 0.0	0.1 ± 0.1	50 ± 7.1	44.6 ± 0.5

Event reconstruction

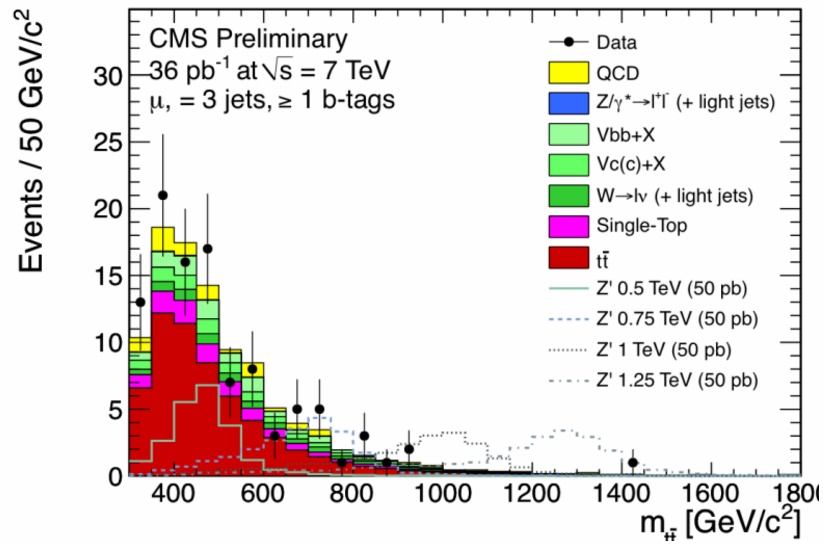
- Reconstruct neutrino p_Z component by using the W mass constraint
- Associate jets to form the two top systems
  - Use a χ^2 method with information from the hadronic and leptonic reconstructed masses, the p_T of the tt system, the H_T of the event
- Four jet events: apply a full kinematic fit
  - Exploit known top and W masses
  - Use jet resolutions from simulation
- Improve the resolution and linearity on m(tt) in most of the interesting range

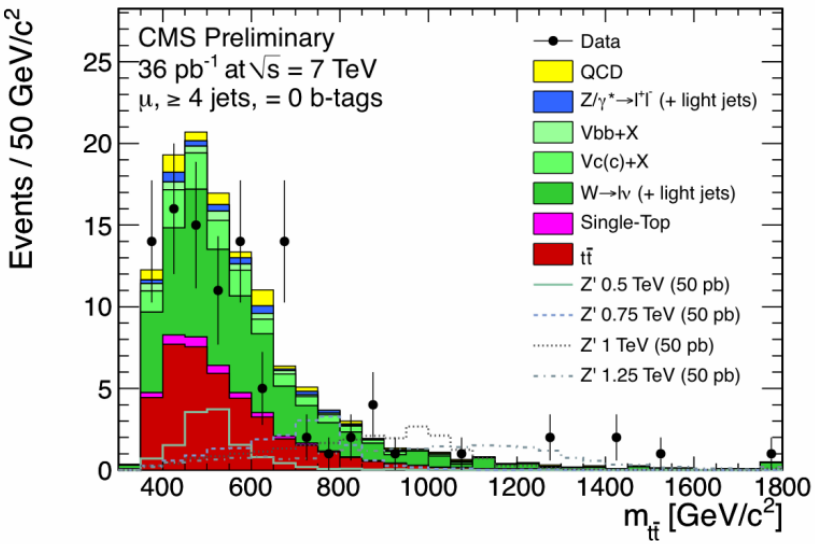


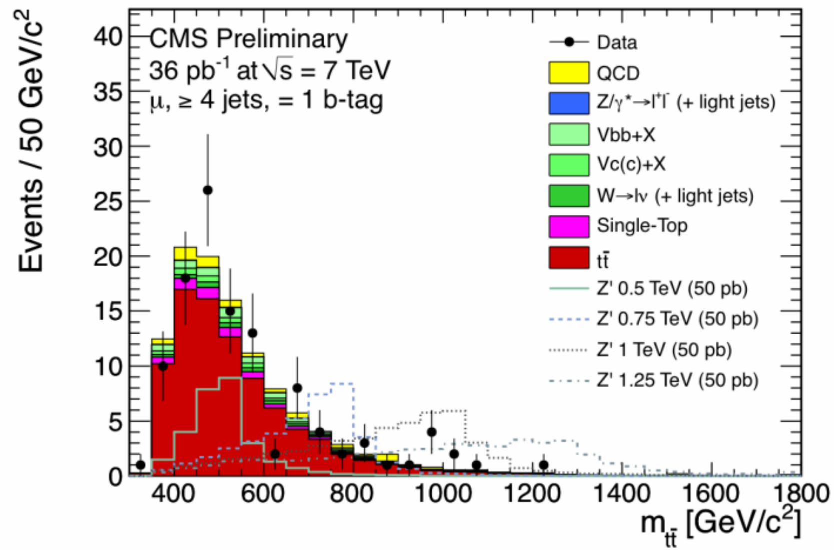


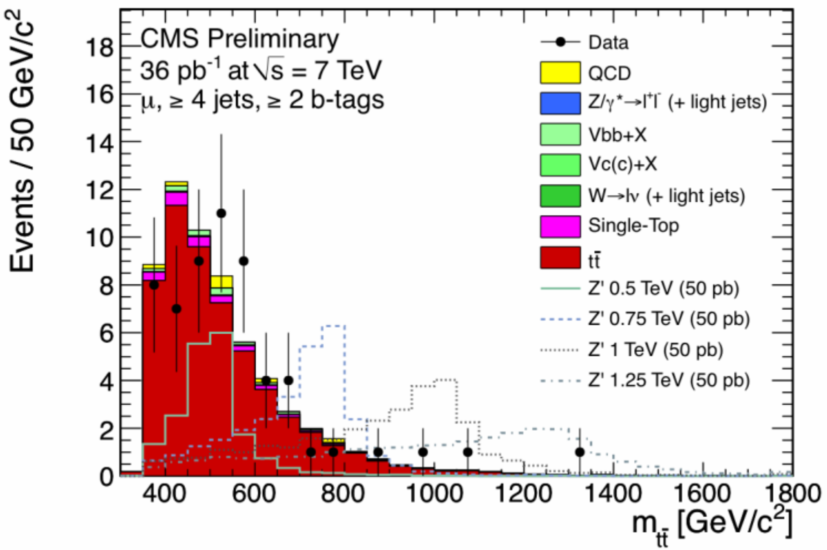
Muon+jets m(tt) distributions

- Data is superimposed to MC expectations and Z' signals in arbitrary normalization



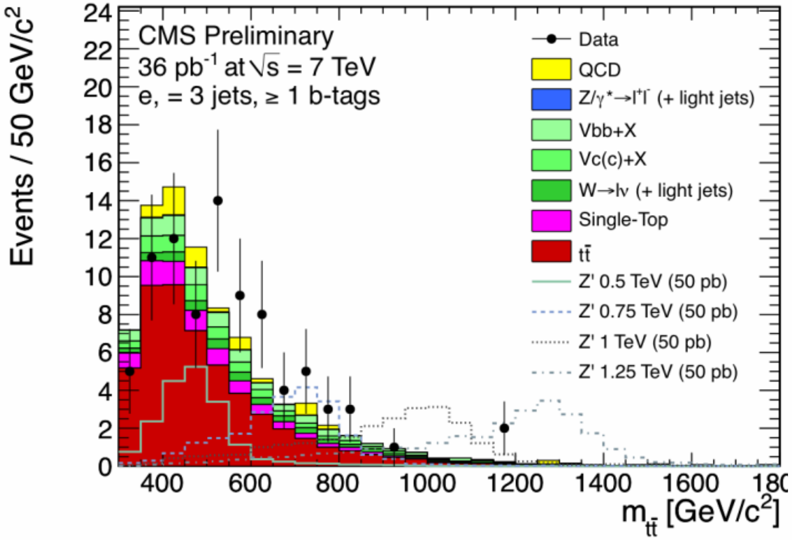


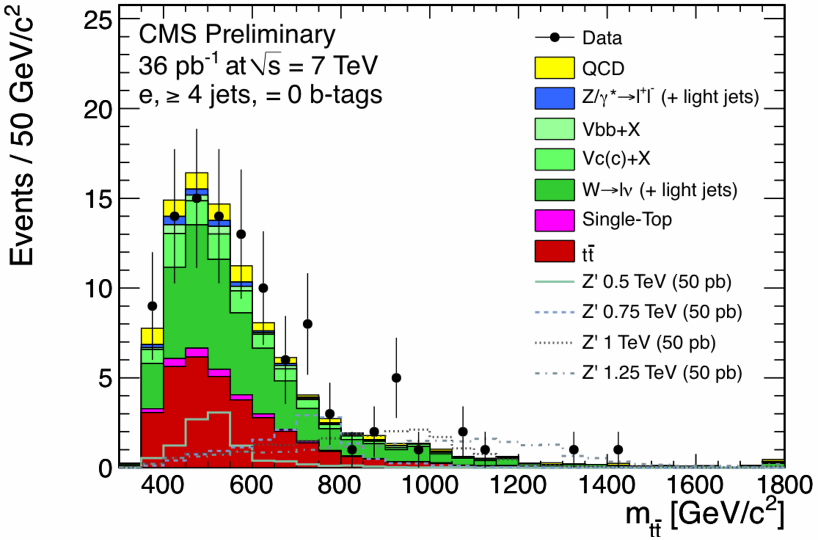


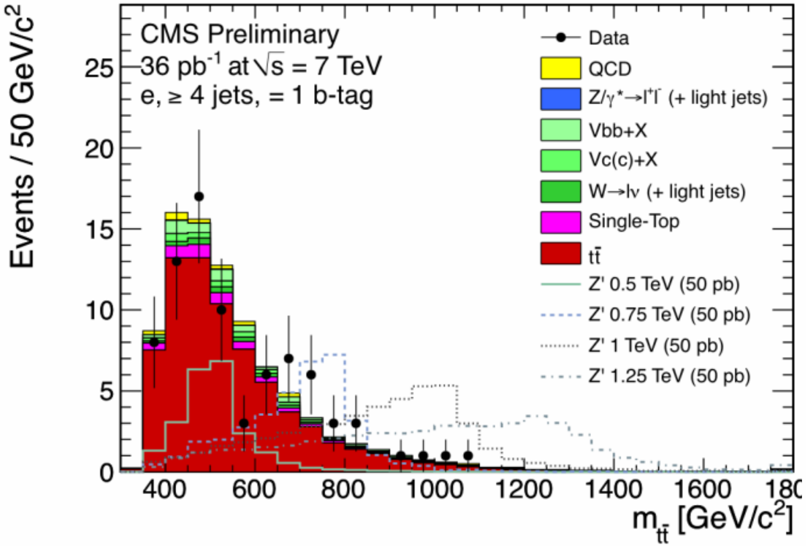


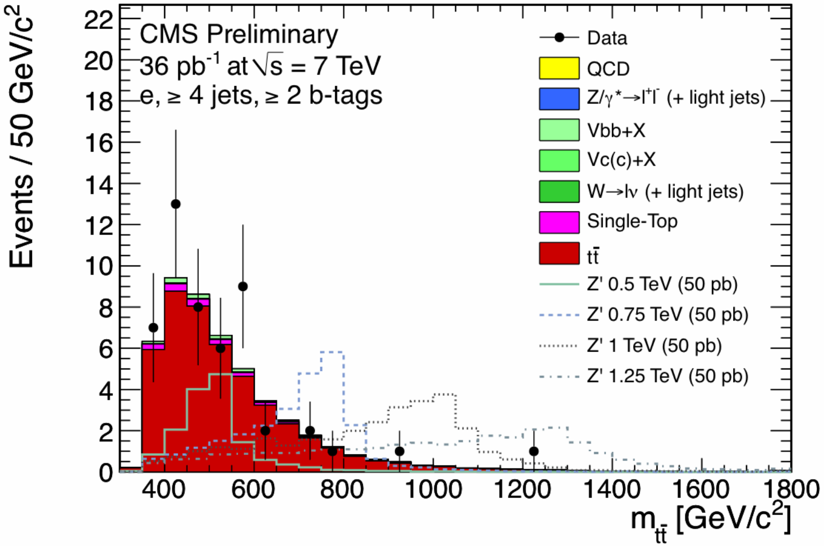
Electron+jets m(tt) distributions

- Data is superimposed to MC expectations and Z' signals in arbitrary normalization









Statistical treatment

- Mass distributions modeled by templates for signal and background

$$\begin{aligned} n_k(m_{t\bar{t}}, \vec{\sigma}^r, \vec{\sigma}^s) &= N_k^{signal}(\vec{\sigma}^r, \vec{\sigma}^s) \cdot \mathrm{pdf}^{signal}(m_{t\bar{t}}, \vec{\sigma}^s) \\ &+ \sum_i N_{ki}^{background}(\vec{\sigma}^r, \vec{\sigma}^s) \cdot \mathrm{pdf}^{background}(m_{t\bar{t}}, \vec{\sigma}^s), \end{aligned}$$

shape-changing nuisances

rate-changing nuisances

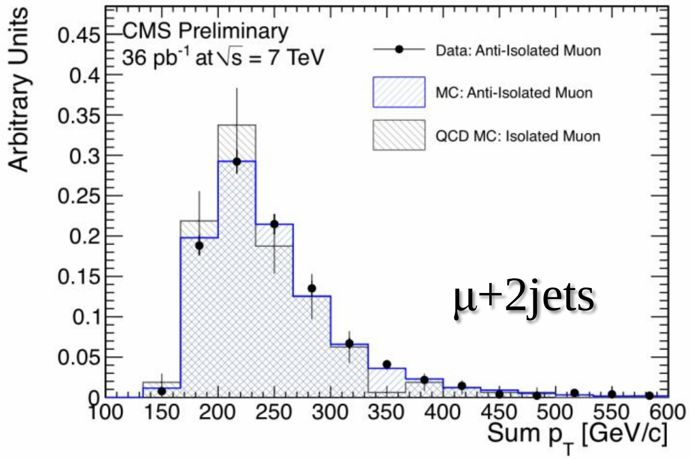
- Fully bayesian approach, all uncertainties are included as nuisance parameters modifying the templates in rates and shapes

- ➢ Gaussian or log-normal priors
- ➢ Shape changing nuisances extrapolated bin-by-bin by fitting the variation as a function of the systematic source with cubic functions
- ➢ Full marginalization over nuisances is granted via a numerical integration using Markov chains MC
- ➢ Background rates and shapes taken from both MC and data
  - ➢ All reference rates and shapes, except for QCD, taken from MC
  - ➢ QCD constrained with data

Uncertainty	Variation	Type
Luminosity	4%	rate
Electron efficiency (trigger + ID + isolation)	5%	rate
Muon efficiency (trigger + ID + isolation)	5%	rate
$t\bar{t}$ cross section	20%	rate
Single top cross section	30%	rate
W+jets cross section	50%	rate
Ratio Drell-Yan to W cross section	30%	rate
Ratio W/Z+HF to σ(W)	100%	rate
Muon QCD yield	100%	rate
Electron QCD yield	100%	rate
Jet energy scale	p_T,η dependent	shape
Jet energy resolution	10%	shape
Unclustered energy	10%	shape
b tagging efficiency (b jets)	15%	shape
b tagging efficiency (c jets)	30%	shape
Q^2 scale for W and Drell-Yan events		shape
$t\bar{t}$ modelling		shape
Q^2 scale for $t\bar{t}$ events		shape
Amount of ISR/FSR for $t\bar{t}$ events		shape
Matching scale for $t\bar{t}$ events		shape

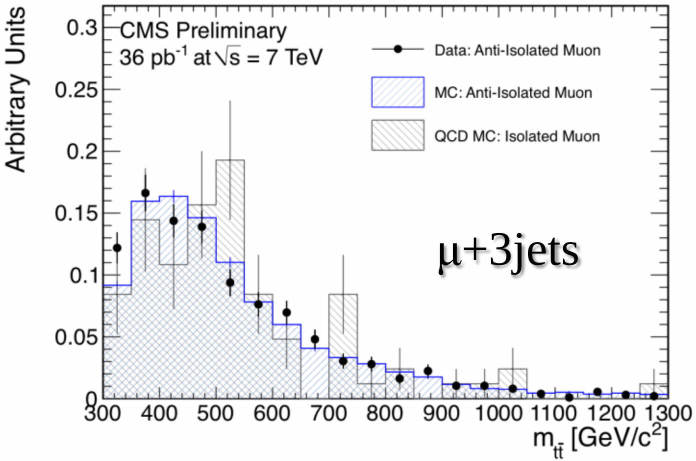
Determining the QCD component

- Use control regions enriched in QCD to determine shape and difference in rate with respect to the MC predictions
  - ➢ Electron: fit to the relative isolation of the lepton extrapolated to signal region
  - ➢ Muon: matrix method using the lepton relative isolation and the (x,y) distance to the primary vertex



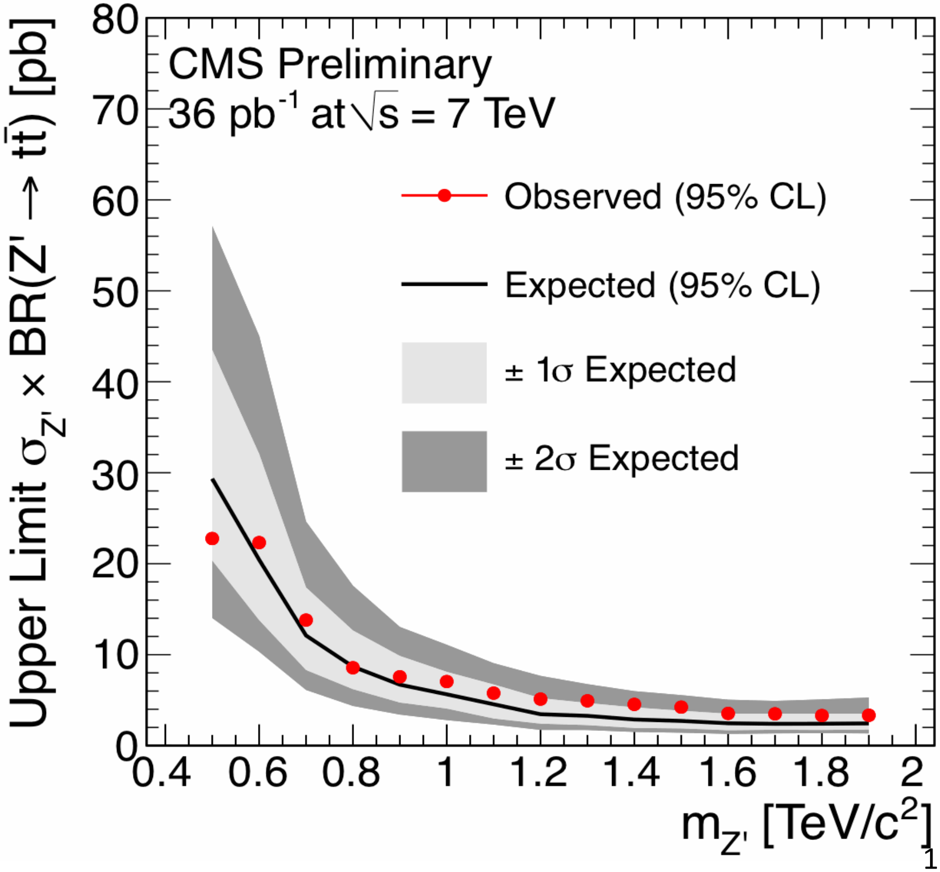
- Factor ~two underestimation of the QCD component by the MC

- Shapes are studied in the control regions
  - ➢ Verify data and MC shape agree in the control regions (dominated by QCD)
  - ➢ Verify m(tt) is not correlated with the definition of the control region
  - ➢ Use the shape of data in the control region to describe QCD in the signal region

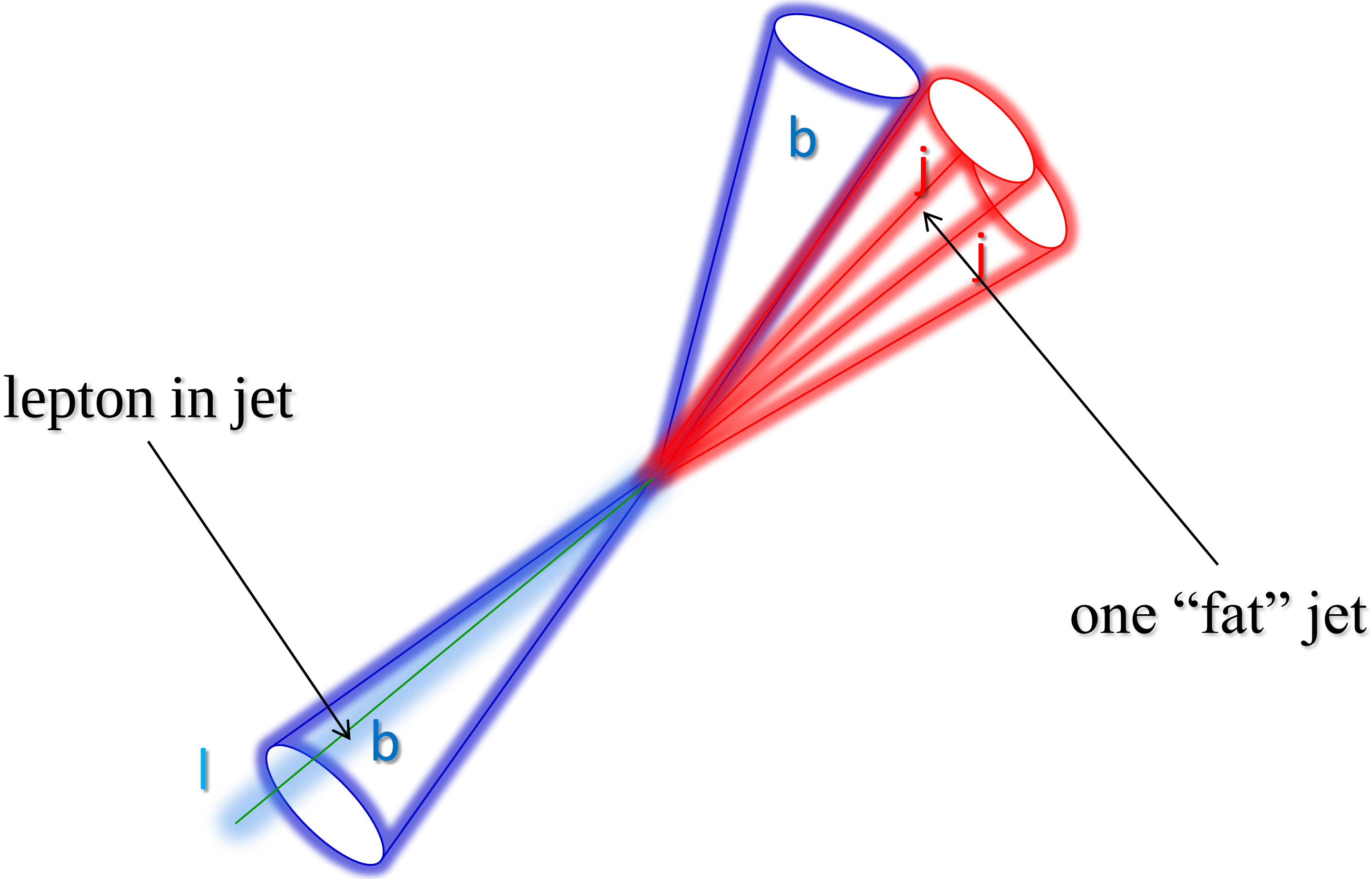


Results

- No observed excess of events in the mass range in reach
- Bayesian integration over nuisances to derive 95% upper limits
  - Use all data collected in 2010, corresponding to 36/pb
- Limits presented in terms of the production cross-section x BR of a Z’
  - Narrow width hypothesis (Γ/m<10%)
  - Expected and observed limits are in good agreement
  - No observed significant discrepancy with respect to the SM expectations
  - Exclusion possible for models predicting cross sections of about 10 pb for masses about 1TeV



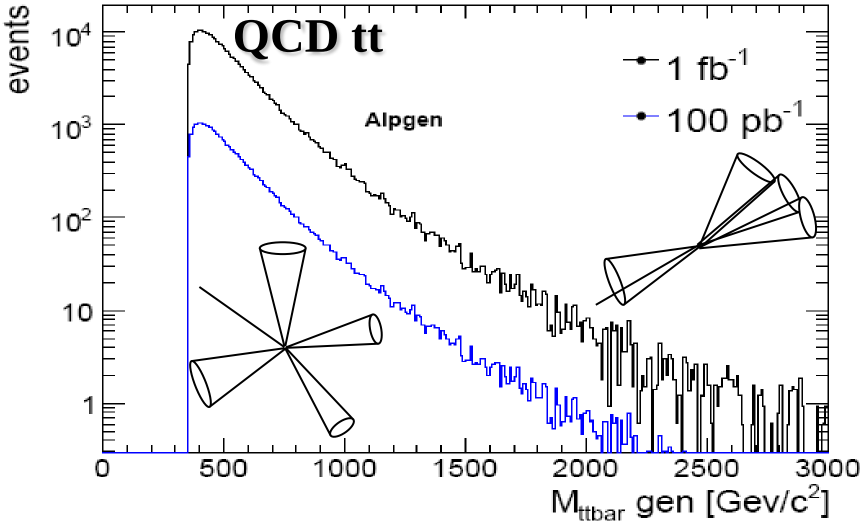
Boosted tops !



Boosted tops

- CMS prepares to cover all portions of the phase space for top-pair production
  - adapt to final state configuration that are very different from threshold top-pair production

- The event kinematics drastically change as a function of m(tt)
  - Leptons progressively loose their isolation
  - Individual jet reconstruction becomes an issue because of (partial) jet merging

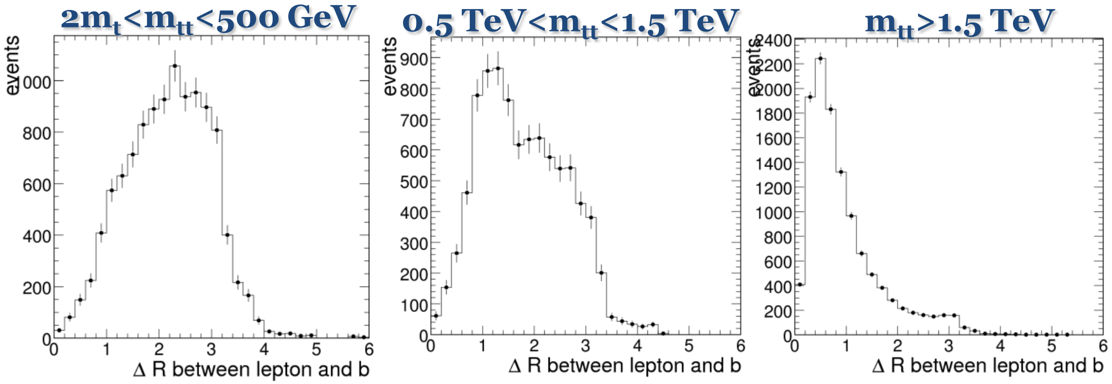


- This imposes stringent requests to:
  - Event triggering
  - Event reconstruction

- Study a top-tagging algorithm
  - Identify jet substructures
  - Based on:

    Kaplan et al: Phys. Rev. Lett. 101, 142001

    Ellis et al: Phys Rev. n D80 (2009)


$2m_t<m_{tt}<500$ GeV

0.5 TeV$<m_{tt}<1.5$ TeV

$m_{tt}>1.5$ TeV

Top tagging in a nutshell

- Cambridge-Aachen (C-A) is a κ_T like algorithm:
  - ➢ Finds the min d_{min} of $\{d_{ij}, d_{iB}\}$ for all pairs. If d_{min} is a d_{ij} merge the pair, if it is a d_{iB} then final jet
  - ➢ κ_T algo: n=2. Anti-κ_T: n=-2. C-A: n=0 (no κ_T weighing)

κ_T of particle i w.r.t. beam axis

κ_T of particle i w.r.t. particle j

$$d_{ij} = \min(k_{\mathrm{T},i}^{n}, k_{\mathrm{T},j}^{n}) \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{i\mathrm{B}} = k_{\mathrm{T},i}^{n}$$

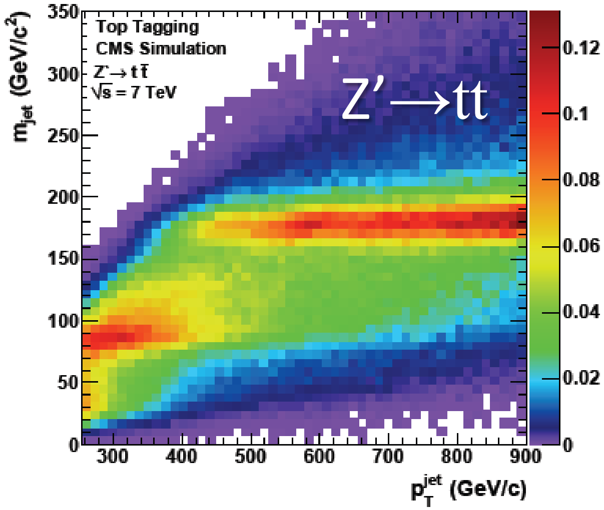
- The top tagging C-A (modified) algorithm (for boosted tops):
  - ➢ 1. Cluster all with R=0.8, consider only hard jets with p_T>250 GeV, |y|<2.5
  - ➢ 2. Decomposition of the jets into pieces:
    - o find iteratively 2 "parent jets" with more than 5% the energy of the jet
    - o if only 2 jets are found in a), find 2 "grandparents" with the same procedure. Decomposition successful if at least one of the parent has two subjets.
  - ➢ 3. Kinematic standard conditions for the decomposed jets
    - o Mass m_{jet} of the 4vector sum of the towers of the initial jet between 100 and 250 GeV.
    - o The min invariant mass m_{min} of all the subjet pairs is required to be larger than 50 GeV.

- The jet pruning algorithm
  - ➢ At each step of the C-A clustering sequence cuts are imposed to avoid too soft and too large-angle pairings
    - o Reject soft radiation at large angle

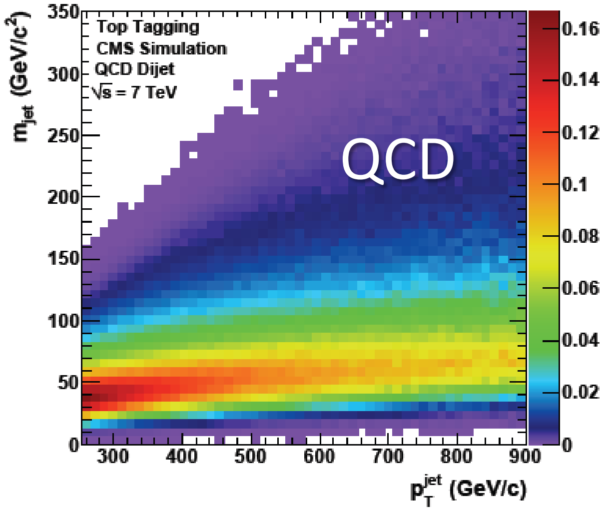
$$\begin{array}{rcl} z_{ij} \equiv min(p_{T,i}, p_{T,j})/p_{T,p} & < & z_{cut} \\ \Delta R_{ij} & > & D_{cut} \end{array}$$

m_{jet} and m_{min}

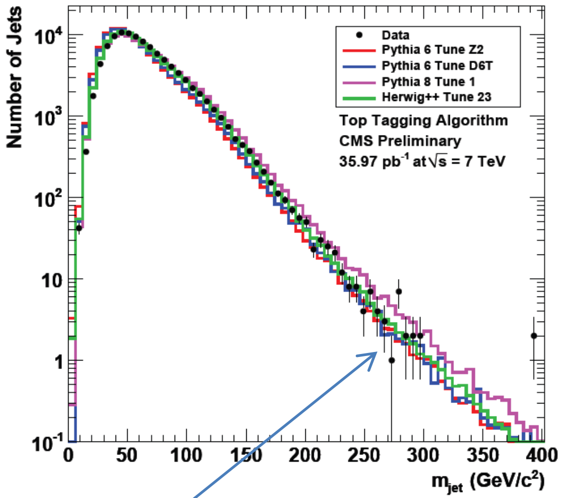
- Main variables for the top tagging already well understood in QCD data



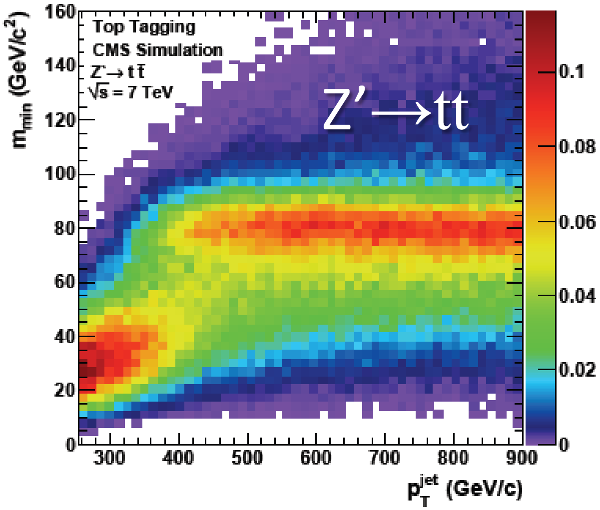
(a)

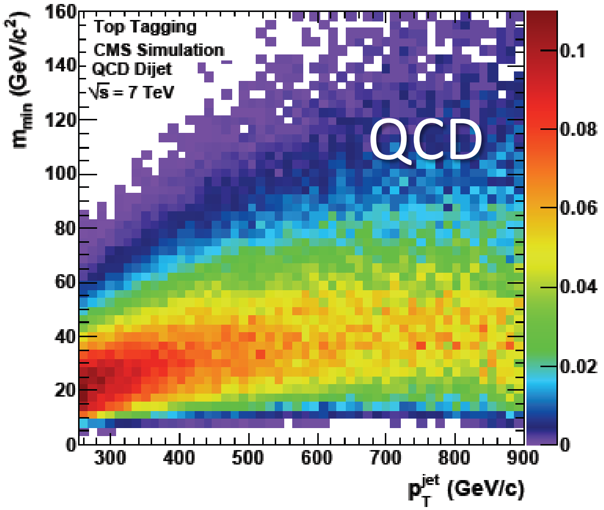


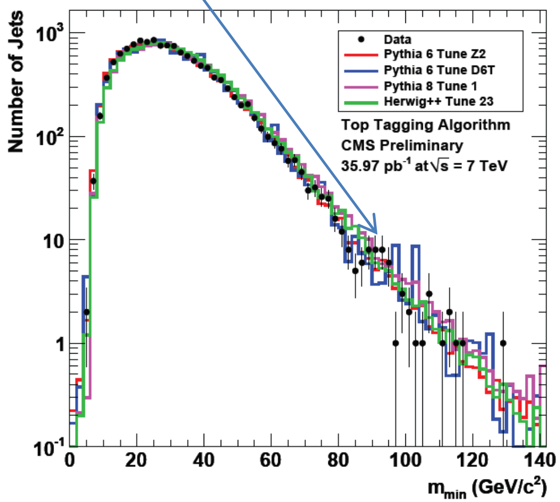
(b)



Sensitivity to tunes !

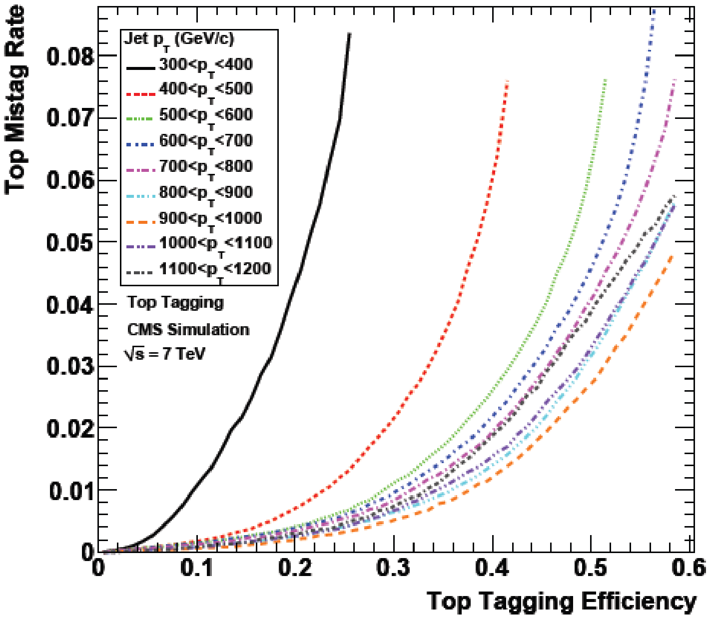


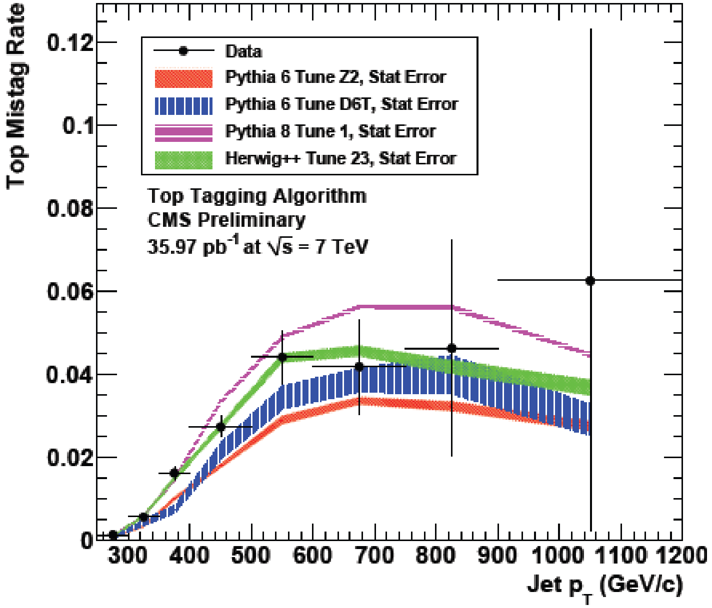




Mistag rate from data

- The algorithm’s parameters can be varied to optimize its performance
  - ➢ Plot results in a 2D way showing top efficiency versus fake rate
    - ○ allow a better and more direct comparison between different algorithm
    - ○ allow a choice of a working point of the algorithm
  - ➢ Top efficiency defined w.r.t. the jets matching a top jet decaying hadronically
  - ➢ Mistag rate defined w.r.t. QCD jets
  - ➢ Optimization done by minimizing the mistag rate at a certain efficiency by changing the cuts on m_{jet} and m_{min}



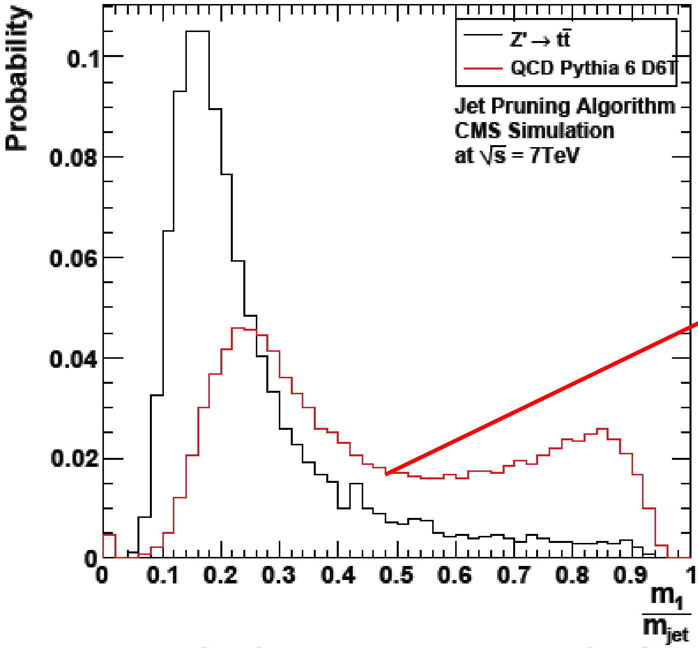


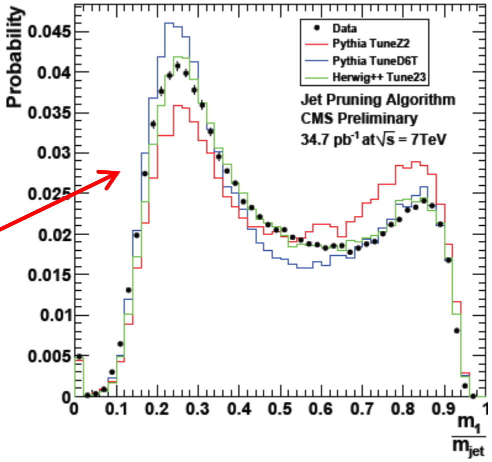
Mistag rate can be directly determined from data by selecting anti-tag and probe in QCD events

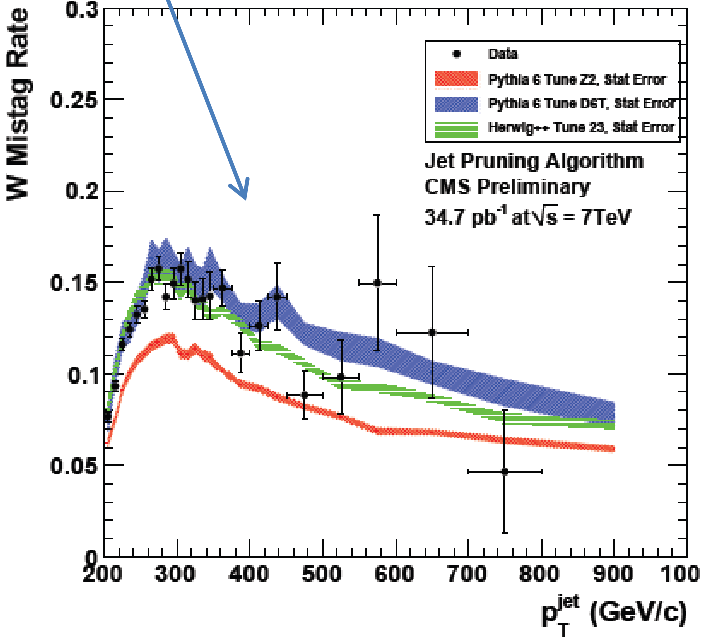
- Use di-jet topologies
- Good agreement with predictions from simulation

W tagging

- A jet pruning algorithm with the requirement of a two-jet substructure
  - With $60\ \mathrm{GeV}/c^2 < m_{jet} < 100\ \mathrm{GeV}/c^2$
- Extra conditions applied to the “mass drop” and the “p_T asymmetry”
  - Mass drop: $m_{highest\ pT}/m_{jet}$ (typically required to be <0.4)
  - p_T asymmetry: $\min(p_{T1}^2, p_{T2}^2)\Delta R_{12}^2/m_{jet}^2$ ensures a minimum energy to the less energetic subjet
- Mistag rate can already be measured with QCD data (tag and probe in di-jet)
  - Extremely good accord between data and MC expectations







Summary and outlook

- The CMS program for exploiting top events for precision physics or for beyond the Standard Model searches is well under way

- Jet reconstruction, tagging, pairing, substructures are key aspects for a complete analysis of top-pair production at the LHC
  - ➢ Top physics at CMS addresses all of them in the top mass and top-pair mass analyses

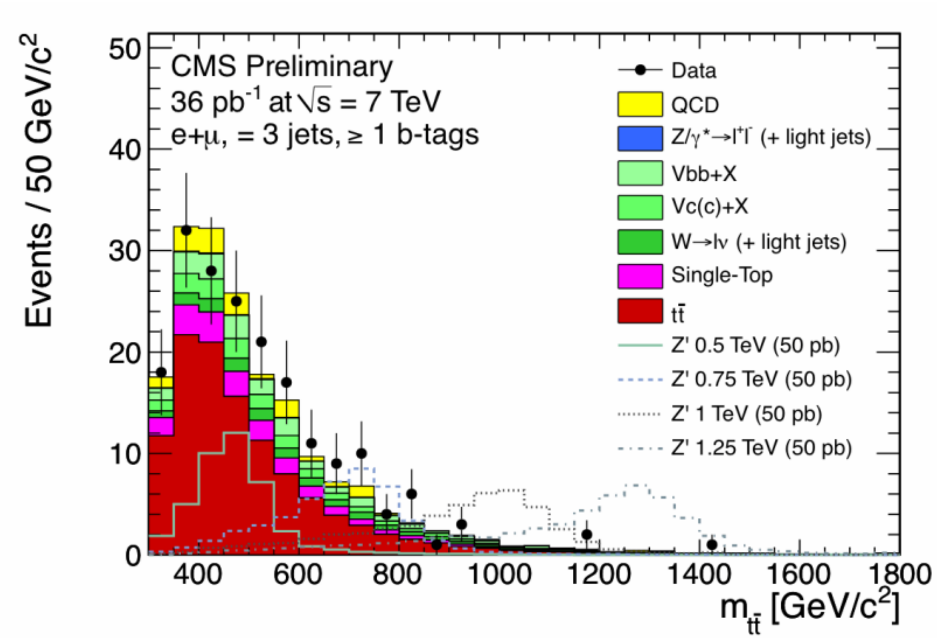
- Analyses with these reconstruction techniques are applied to 2010 data
  - ➢ Top mass determined in the leptonic channel
    - o Semileptonic channels under way
  - ➢ Top-pair mass distribution in the semi-leptonic channels
    - o Analysis with standard reconstruction shows no presence of new physics so far
    - o Boosted top (and W) tagging techniques are validated with QCD data and ready to be exploited

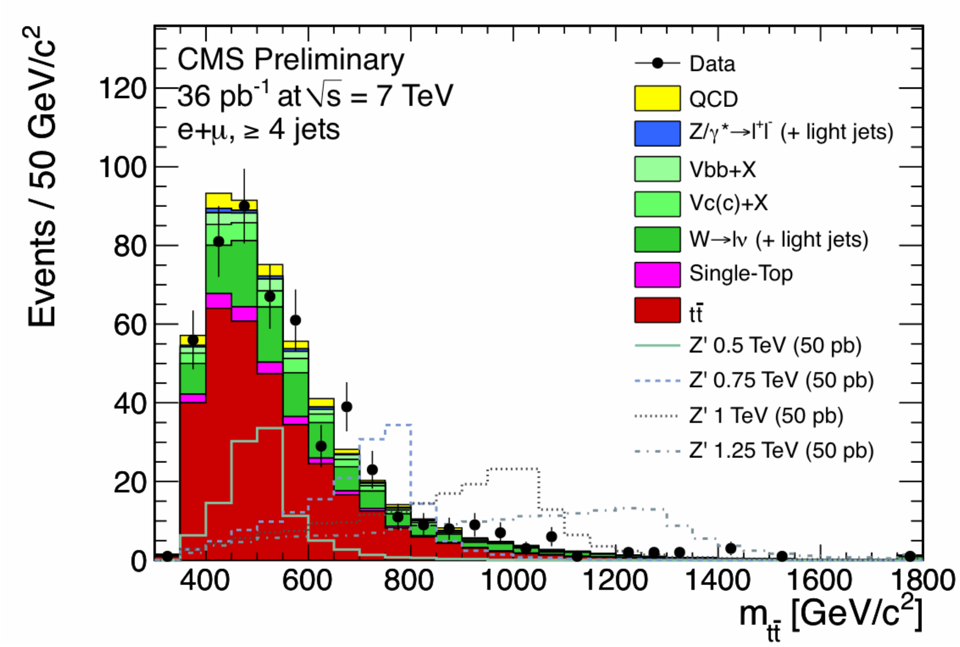
- Excellent performance of detector and simulation so far
  - ➢ CMS is ready for the increase in statistics in 2011

Backup

M(tt) distributions

- Electron and muon channels added up together
  - ➢ Distributions to be taken with grain of salt: limits are not derivable from these





Crosscheck analysis for m(tt)

- A simpler analysis is used for crosschecking the m(tt) result
  - ➢ Different (looser) lepton isolation to grant high efficiency in moderately boosted top configurations
  - ➢ Jet b-tagging is used as a selection criteria (no optimal use of all information)
  - ➢ Background directly derived from data by definition of “side-bands” regions obtained by inverting the b-tagging condition

- Limits in accord with the reference analysis. Minor expected sensitivity