Search for $t\bar{t}$ resonances in semileptonic final state using 19.6 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV with the CMS detector

GDR Terascale – Montpellier

Sébastien Brochet[1]

[1]Institut de Physique Nucléaire de Lyon – Université Claude Bernard Lyon 1

May 15, 2013

Based on public CMS-PAS-B2G-12-006

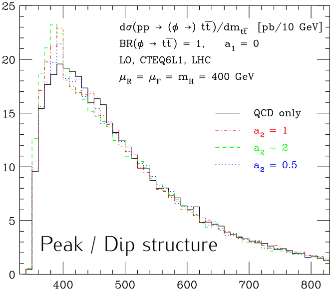
Contents

Physics motivation

- Numerous extensions to the standard model predict gauge interactions with enhanced couplings to the top quark.
  - Massive Higgs boson
  - Extra-dimension (Randall-Sundrum models for example)
  - Massive color singlet Z′ bosons
  - Colorons or axigluons
  - ...
- These new models can have very different signatures, like a bump or a dip in a mass spectrum, forward-backward asymmetry of top pair production, ...

Physics motivation

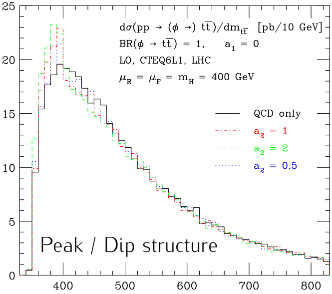
- Numerous extensions to the standard model predict gauge interactions with enhanced couplings to the top quark.
  - Massive Higgs boson
  - Extra-dimension (Randall-Sundrum models for example)
  - Massive color singlet Z′ bosons
  - Colorons or axigluons
  - ...
- These new models can have very different signatures, like a bump or a dip in a mass spectrum, forward-backward asymmetry of top pair production, ...



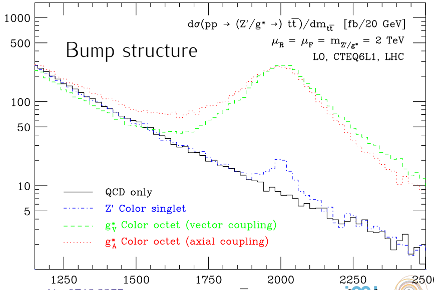
arXiv:0712.2355

Physics motivation

- Numerous extensions to the standard model predict gauge interactions with enhanced couplings to the top quark.
  - Massive Higgs boson
  - Extra-dimension (Randall-Sundrum models for example)
  - Massive color singlet Z′ bosons
  - Colorons or axigluons
  - ...
- These new models can have very different signatures, like a bump or a dip in a mass spectrum, forward-backward asymmetry of top pair production, ...



arXiv:0712.2355



arXiv:0712.2355

Physics motivation

- In this analysis, we perform a "bump hunt" in the $t\bar{t}$ mass spectrum, considering two particular models:

Physics motivation

- In this analysis, we perform a "bump hunt" in the $t\bar{t}$ mass spectrum, considering two particular models:
  - A **generic** spin 1 Z′ boson*, with the same fermion coupling as the SM Z boson. In this model, the width and cross-section are **independent** of the mass. We consider two kinds of resonances: *narrow*, with $\Gamma_{Z'}/m_{Z'} = 1.2\%$; and *wide*, with $\Gamma_{Z'}/m_{Z'} = 10\%$. This is a benchmark scenario since the Tevatron.

*Harris et al.

Physics motivation

- In this analysis, we perform a "bump hunt" in the $t\bar{t}$ mass spectrum, considering two particular models:
  - A **generic** spin 1 Z′ boson*, with the same fermion coupling as the SM Z boson. In this model, the width and cross-section are **independent** of the mass. We consider two kinds of resonances: *narrow*, with $\Gamma_{Z'}/m_{Z'} = 1.2\%$; and *wide*, with $\Gamma_{Z'}/m_{Z'} = 10\%$. This is a benchmark scenario since the Tevatron.
  - The Kaluza-Klein partner of SM gluon†, in the Randall-Sundrum 1 model. In this model, the width and cross-section are **fixed** by the mass.

*Harris et al.
†Agashe et al.

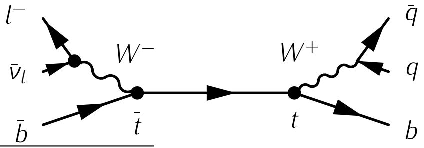
Physics motivation

- In this analysis, we perform a “bump hunt” in the $t\bar{t}$ mass spectrum, considering two particular models:
  - A **generic** spin 1 Z′ boson*, with the same fermion coupling as the SM Z boson. In this model, the width and cross-section are **independent** of the mass. We consider two kinds of resonances: *narrow*, with $\Gamma_{Z'}/m_{Z'} = 1.2\%$; and *wide*, with $\Gamma_{Z'}/m_{Z'} = 10\%$. This is a benchmark scenario since the Tevatron.
  - The Kaluza-Klein partner of SM gluon†, in the Randall-Sundrum 1 model. In this model, the width and cross-section are **fixed** by the mass.
- Signal has been generated for different mass points: 0.5, 0.75, 1, 1.5, 2, and 3 TeV.

*Harris et al.
†Agashe et al.

Physics motivation

- In this analysis, we perform a "bump hunt" in the $t\bar{t}$ mass spectrum, considering two particular models:
  - A **generic** spin 1 Z′ boson*, with the same fermion coupling as the SM Z boson. In this model, the width and cross-section are **independent** of the mass. We consider two kinds of resonances: *narrow*, with $\Gamma_{Z'}/m_{Z'} = 1.2\%$; and *wide*, with $\Gamma_{Z'}/m_{Z'} = 10\%$. This is a benchmark scenario since the Tevatron.
  - The Kaluza-Klein partner of SM gluon†, in the Randall-Sundrum 1 model. In this model, the width and cross-section are **fixed** by the mass.
- Signal has been generated for different mass points: 0.5, 0.75, 1, 1.5, 2, and 3 TeV.
- A model independent search for heavy resonances in $t\bar{t}$ spectrum is performed, where one top decays to $bl\nu$, and the other to bjj (semileptonic channel)



*Harris et al.
†Agashe et al.

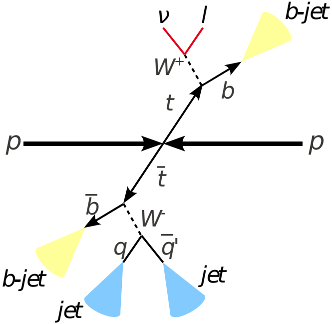
Analysis strategy

- Search performed over a wide mass range, with different event topologies

Analysis strategy

- Search performed over a wide mass range, with different event topologies

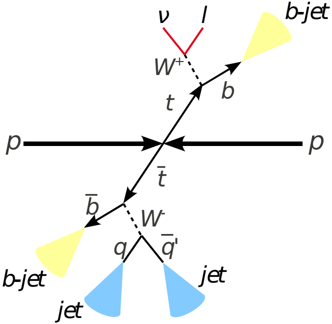
low mass (< 1 TeV) – SM like



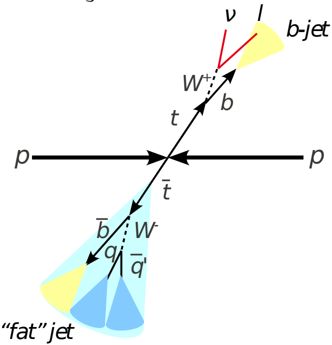
Analysis strategy

- Search performed over a wide mass range, with different event topologies

low mass (< 1 TeV) – SM like



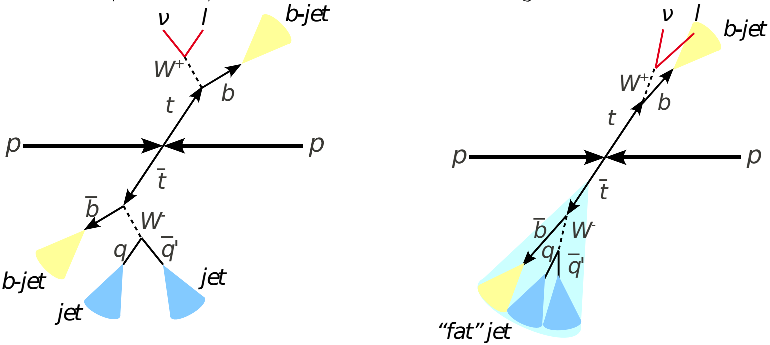
high mass, boosted



Analysis strategy

- Search performed over a wide mass range, with different event topologies

low mass (< 1 TeV) – SM like high mass, boosted



- Two different analyses:
  - *Low mass* analysis, using standard top quark reconstruction techniques
  - *High mass* analysis, using techniques optimized for boosted top quark reconstruction

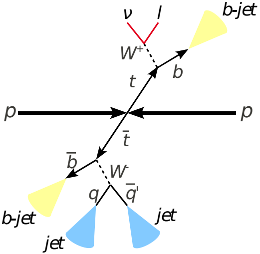
What's next

- Because each analysis has a similar structure, I'll present for each
  - The selection
  - The reconstruction of $t\bar{t}$ events
  - The number of signal events extraction, with a fit on data
    - For the *low mass* analysis, data driven approach for background modelization
    - For the *high mass* analysis, mainly MC based background modelization

Low mass analysis

Selection

- Selection optimized for non-boosted top pairs:
  - One isolated lepton + 3 jets trigger
  - $\not{E}_t > 20$ GeV
  - At least 4 jets with $p_t > 30$ GeV.
  - Exactly one isolated lepton
  - At least 1 b-tagged jet



- Event categorized in four categories, based on lepton flavor and number of b-tagged jets

muons	$\otimes$	1 b-tagged jet
electrons		≥ 2 b-tagged jets

- Typical efficiency of ~30%.

Reconstruction of $t\bar{t}$ events

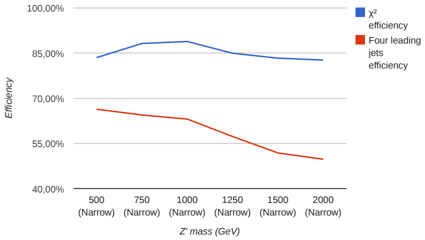
- The lepton and the $\not{E}_t$ (interpreted as the p_t of the neutrino) are assigned to the leptonic leg of $t\bar{t}$ event.
- An hadronic collision produces a lot of jets. Therefore, among the selected jets, we need to correctly choose the ones coming from $t\bar{t}$ decay. For that, we use a χ^2 sorting algorithm.

χ^2 sorting algorithm

$$\chi^2 = \chi^2_{m_{top}^{lept}} + \chi^2_{m_{top}^{hadr}} + \chi^2_{m_W^{hadr}} + \chi^2_{p_t^{t\bar{t}} \ system}$$

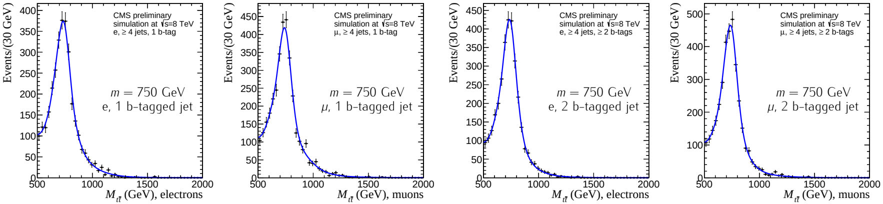
where $\chi^2_x = (x_{\text{meas}} - x_{\text{MC}})^2 \ / \ \sigma^2_{\text{MC}}$

- Iterates over all jets combinations and chooses the solution with the smallest χ^2



Number of signal events extraction

- Simultaneous extended maximum likelihood fit to the mass distribution, with a fit range from 550 GeV to 2 TeV.
- Likelihood build with two components: background and signal.
- Signal parametrization derived from MC using a kernel estimation.

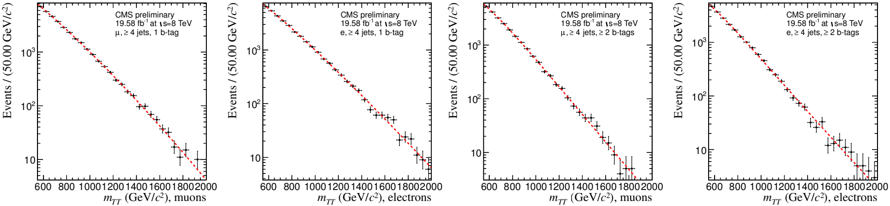


Number of signal events extraction

- A data driven method is used to parametrize the background functional form.

$$\frac{d\sigma}{dm_{t\bar{t}}} = \frac{\left(1 - m/\sqrt{s}\right)^{c_1}}{\left(m/\sqrt{s}\right)^{c_2 + c_3 \ln m/\sqrt{s}}}$$

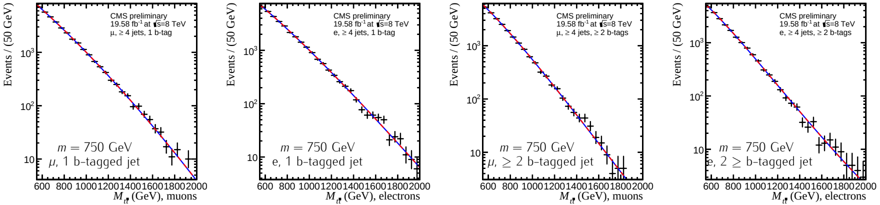
- This PDF is inspired by the parton density function of protons:



$$\chi^2 \ / \ \mathrm{NDF} = 1.05$$

Number of signal events extraction

- The signal PDF is fixed during the fit, and background parameters are left floating.



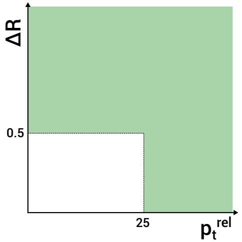
- No evidence of signal is found, so we can proceed to set limits.

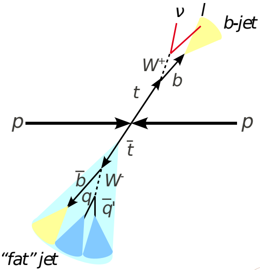
High mass analysis

Selection

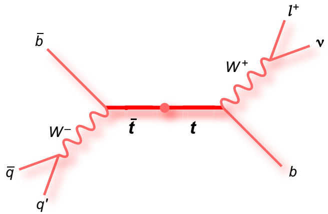
- Selection optimized for boosted top pairs:
  - Single lepton trigger
  - At least 2 jets with $p_t > 150/50$ GeV
  - Exactly one lepton, without isolation requirement
  - 2D cut: $\Delta R(l, \text{closest jet}) > 0.5$ or $p_t^{\text{rel}}(l, \text{closest jet}) > 25$ GeV
  - $\not{E}_t > 50$ GeV and $H_T^{\text{lep}} > 150$ GeV
  - Triangular cut: removes event where $\not{E}_t$ is parallel to the electron.
- Event categorized in four categories, based on lepton flavor and number of b-tagged jets

muons	$\otimes$	0 b-tagged jet
electrons		≥ 1 b-tagged jets





Reconstruction of $t\bar{t}$ events



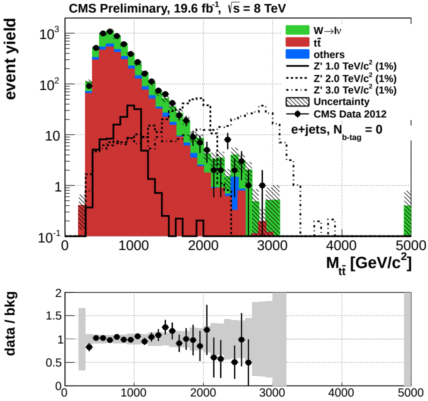
- A list of reconstruction hypothesis is created
  - Neutrino momentum from $\not{E}_t$ and W boson mass constraint
  - A jet is assigned to the leptonic leg
  - One or more jets are allowed in the hadronic top leg
- Select the hypothesis with minimal χ^2. Drop event if minimal $\chi^2 > 10$.

$$\chi^2 = \chi^2_{m_{top}^{lept}} + \chi^2_{m_{top}^{hadr}}$$

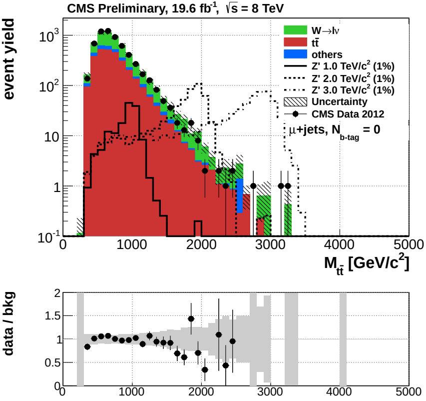
Number of signal events extraction

- Both signal and background shapes are taken from MC.
- Binned likelihood fit of the invariant mass of the reconstructed $t\bar{t}$ system :
- Combined likelihood for the 4 channels (electron/muon, without/with b-tagging)

$m_{t\bar{t}}$ distribution for $N_{\text{b-tag}} = 0$



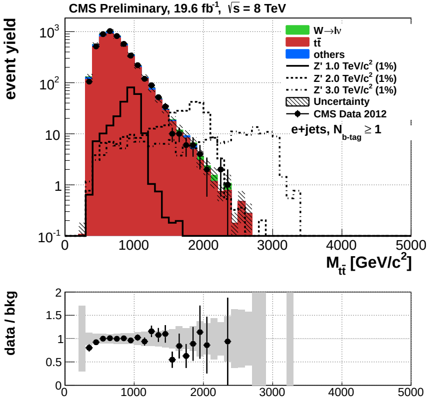
electron channel



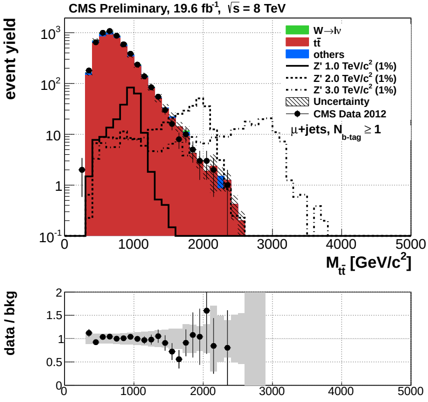
muon channel

- Backgrounds: ~50% from W+jets, ~50% from SM $t\bar{t}$ production.
- Signal should be visible experimentally for high mass point at 1 pb.

$m_{t\bar{t}}$ distribution for $N_{\text{b-tag}} \geq 1$



electron channel



muon channel

- Backgrounds dominantly SM $t\bar{t}$ production.

Combination

Systematic uncertainties

- Log-normal distributions are used as prior for all the systematic uncertainties.
- Some systematic uncertainties are common to both analysis, but some are very specific, like uncertainties related to the fit procedure or the MC modelization.

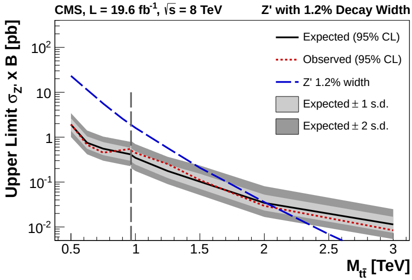
Systematic uncertainty	*low mass* analysis		*high mass* analysis	
	signal	background	signal	background
event pileup	×		×	×
luminosity	×		×	×
lepton ID and trigger	×		×	×
jet energy scale and resolution	×		×	×
signal probability density function	×			
background probability density function		×		
background cross section				×
parton distribution functions	×		×	×
background modeling				×

Methodology

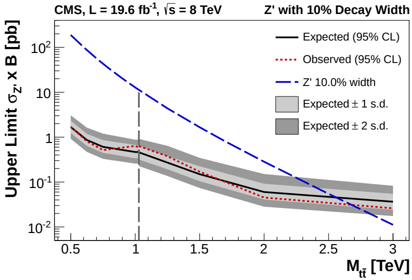
- Limits are produced for topcolor Z′ and for Kaluza-Klein gluons, for each analysis, using a Bayesian statistical method to extract the 95% C.L. upper limits.
- Selections between *low mass* and *high mass* analysis are not orthogonal: a simple combination is not possible.
- Instead, the limits obtained from the two analyses are combined by transitioning from the *low mass* to the *high mass* results based on the sensitivity of the expected limit.

Limits – Topcolor Z'

$\Gamma_{Z'}/m_{Z'} = 1.2\%$

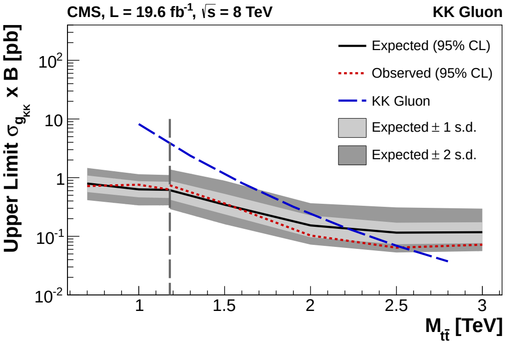


$\Gamma_{Z'}/m_{Z'} = 10\%$



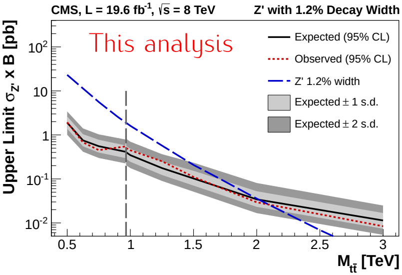
- Masses below 2.1 TeV are excluded for narrow Z'.
- Masses below 2.7 TeV are excluded for wide Z'.
- Model independent: **limits can be interpreted with other models**.

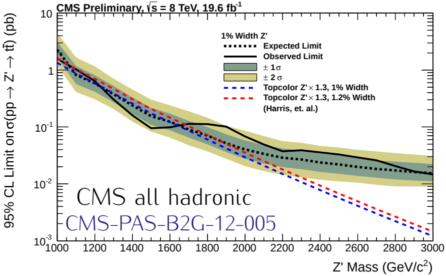
Limits – Kaluza-Klein gluons

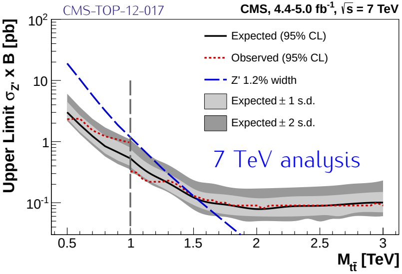


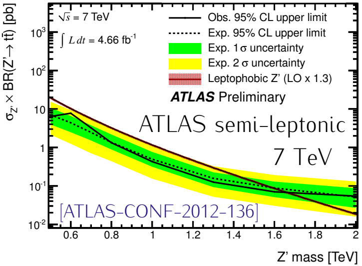
- Masses below 2.5 TeV are excluded for Kaluza-Klein gluons.

Comparison with other results









Summary

- Search for resonant $t\bar{t}$ production using 19.6 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV with the CMS detector has been presented.
- No evidence of new physics
- Topcolor Z′ and Kaluza-Klein gluons can be excluded up to:

Model	narrow Z′ (1%)	wide Z′ (10%)	Kaluza-Klein gluons
Exclusion	2.1 TeV	2.7 TeV	2.5 TeV

- **Best limits** up to date on these models
- On going combination with the all hadronic analysis

Backup

Statistical analysis – choice of background function

- A data driven method is used to parametrize the background functional form.
- Several (five) functional forms are considered. A χ^2 like test using the saturated algorithm on data allows us to select only three PDFs from the five:

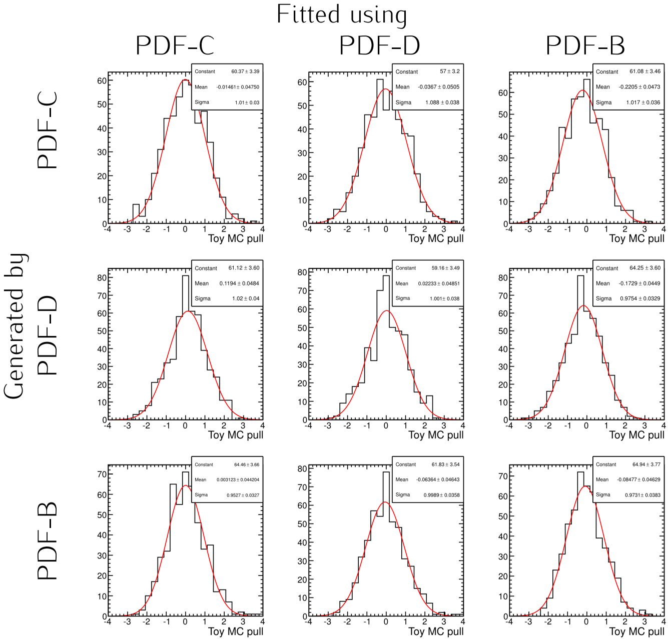
$$\frac{d\sigma}{dm_{t\bar{t}}} = \frac{1}{1 + \exp \frac{m/\sqrt{s} - c_1}{c_2}} \qquad \text{(PDF-B)}$$

$$= \frac{\left(1 - m/\sqrt{s} + c_3 \left(m/\sqrt{s}\right)^2\right)^{c_1}}{m^{c_2}} \qquad \text{(PDF-C)}$$

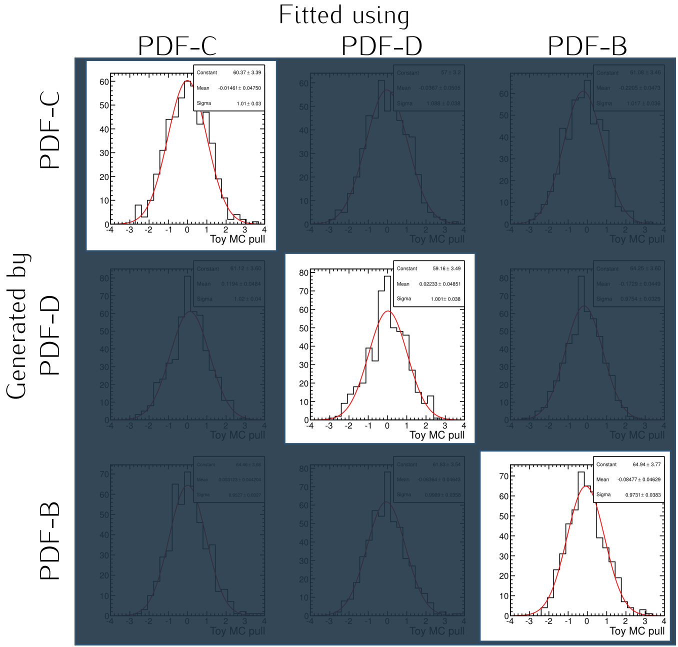
$$= \frac{\left(1 - m/\sqrt{s}\right)^{c_1}}{\left(m/\sqrt{s}\right)^{c_2 + c_3 \ln m/\sqrt{s}}} \qquad \text{(PDF-D)}$$

- For each of these functions, a set of 500 pseudo-experiments is generated, with a number of signal events equivalent to a 2σ fluctuation.
- Each pseudo-experiment is then fitted with the three possible functions. For each mass point, we obtain a set of 9 pull distributions for the number of signal events.

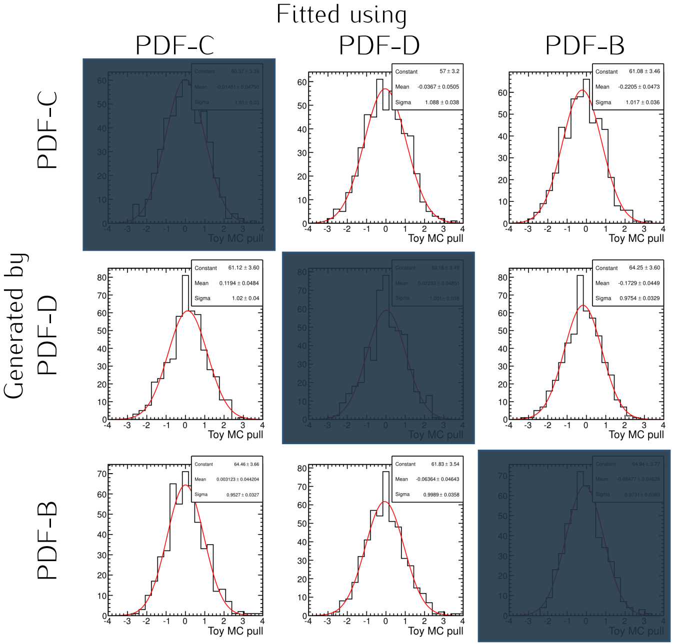
Statistical analysis – pull matrix for $m_{Z'} = 750$ GeV



Statistical analysis – pull matrix for $m_{Z'} = 750$ GeV



Statistical analysis – pull matrix for $m_{Z'} = 750$ GeV



Statistical analysis – Background bias estimation

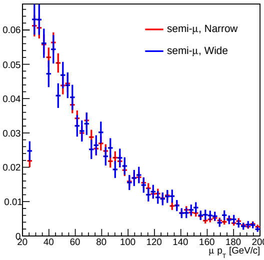
- We look at the mean of the pull distribution of number of signal events (which is consistent with 0 if the fit is unbiased)
- On diagonal, pull distributions show no sign of bias
- In order to have only one estimator by function and by mass point, we sum the toy fitted by one PDF and generated by the others (i.e., we exclude the set of toys fitted by one function and fitted by itself).
- Mean values (with errors) are summarized in the table below, for each mass point and each function

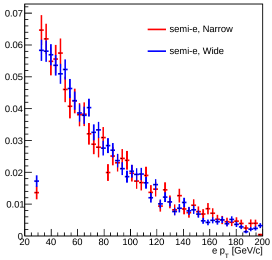
$m_{Z'}$ [GeV/c^2]	PDF-C	PDF-B	PDF-D
500	0.061 ± 0.034	-0.029 ± 0.032	-0.045 ± 0.033
750	-0.12 ± 0.03	-0.046 ± 0.032	-0.032 ± 0.032
1000	$\mathbf{0.32} \pm 0.03$	0.088 ± 0.036	0.059 ± 0.032
1250	-0.17 ± 0.03	$\mathbf{0.37} \pm 0.03$	0.090 ± 0.037
1500	-0.43 ± 0.03	0.046 ± 0.034	$\mathbf{0.21} \pm 0.04$

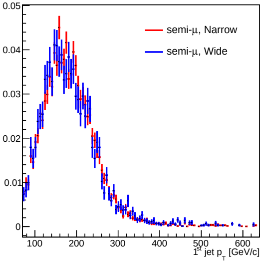
Table 1: Mean values (with errors) from Gaussian fit to pull distributions obtained, for each fitting function, on pseudo-experiments generated with different functions.

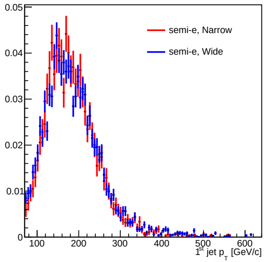
PDF-D minimizes the bias on the number of signal events. For this reason, PDF-D is used as our background function

Narrow / wide resonances comparison ($m_{Z'} = 750\text{GeV}$)

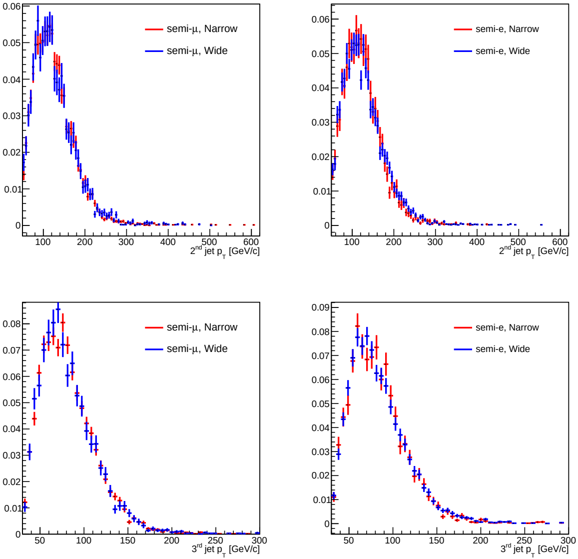








Narrow / wide resonances comparison ($m_{Z'} = 750\text{GeV}$)



χ^2 terms study

Definitions

- An event is said:

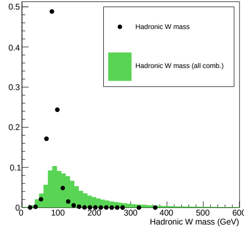
  associable if we can find a reco-object matching the parton for the four jets of the $t\bar{t}$ system (hadronic b, leptonic b, two light jets from W)

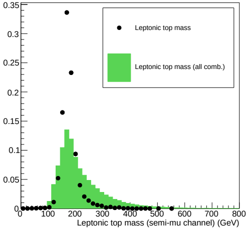
  matched if it's associable, and if the χ^2 algorithm correctly choose the reco-objects (position is not important)

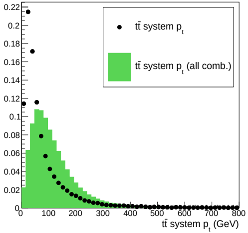
  matched and well placed if it's matched and the reco-objects are correctly positioned.

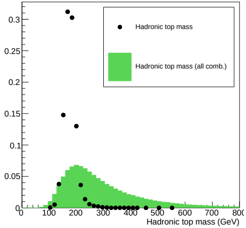
χ^2 terms study

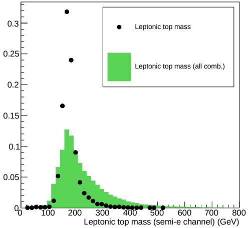
- Using only matched events and well placed from $t\bar{t}$ sample, we compute optimal parameters for the χ^2



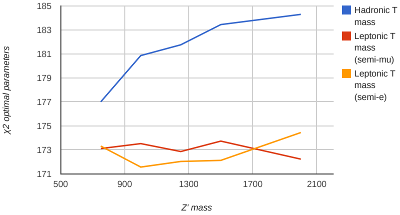


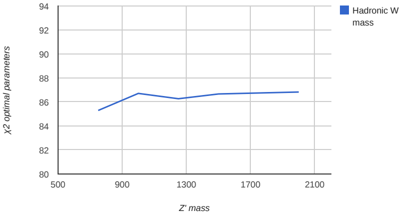






Evolution of χ^2 parameters versus Z' mass

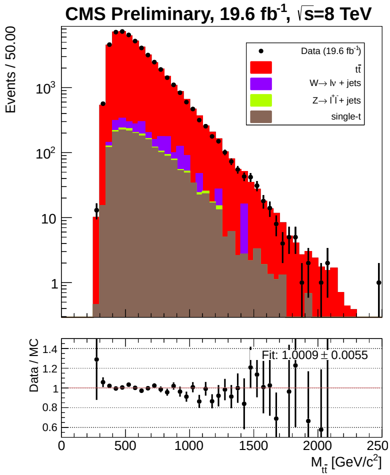




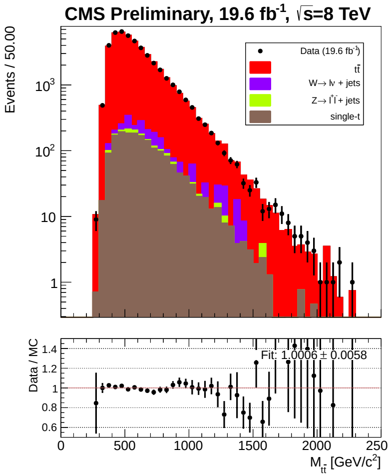
Data / MC comparison

- Data / MC comparison for $m_{t\bar{t}}$ distribution is shown below. (only for sanity checks)

Muons, at least 2 b-tagged jets



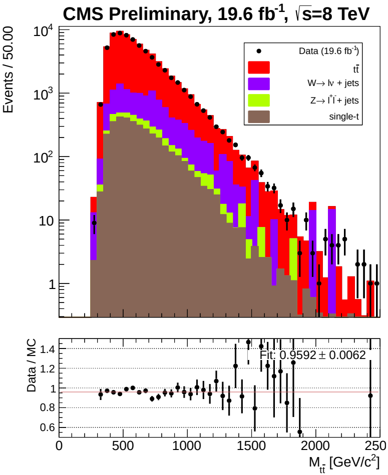
Electrons, at least 2 b-tagged jets



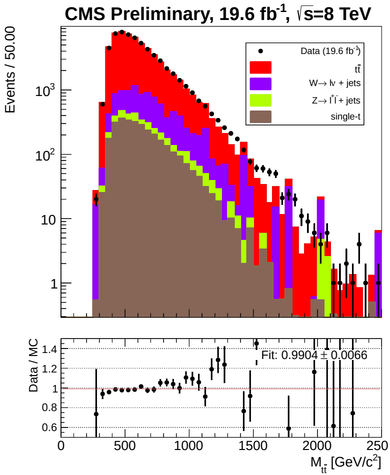
Data / MC comparison

- Data / MC comparison for $m_{t\bar{t}}$ distribution is shown below. (only for sanity checks)

Muons, 1 b-tagged jet



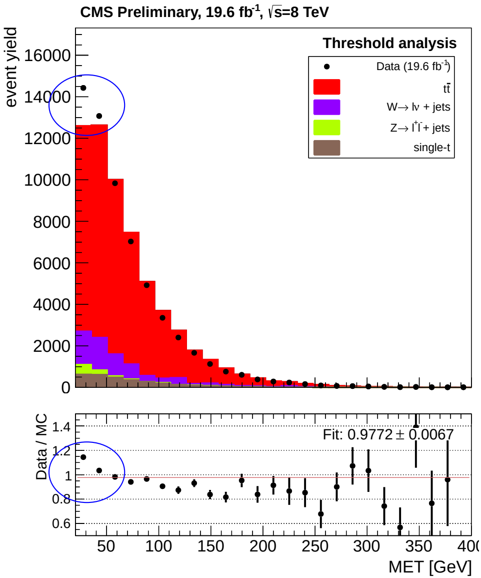
Electrons, 1 b-tagged jet



Data/MC disagreement in semi-e, 1 b-tag category

- We see a slight disagreement between data and MC, especially in semi-electronic, 1 b-tagged jet category
- This disagreement has multiple explanations:
  1. Only statistical uncertainties are taken into account, and no systematic uncertainties.
  2. QCD (multijet) sample is missing. Next slide shows more in detail the QCD contamination (disagreement between data and MC at low $\not{E}_t$)
  3. When cutting at $\not{E}_t > 80$ GeV (removing QCD contamination), data and MC are in agreement.

QCD contamination



- We show data / MC comparison for $\not\!\!E_t$ for the semi-electronic channel, 1 b-tagged jet.
- The disagreement at low $\not\!\!E_t$ is a clear sign of missing QCD.
- When cutting at $\not\!\!E_t > 80$ GeV, we find:
    - $N_{\text{bkg}} = 16682 \pm 136$
    - $N_{\text{data}} = 16634$

High mass analysis selection

- Triangular cut:

$$-\frac{1.5\not{E}_t}{75} + 1.5 < \Delta\varphi\left\{(e \text{ or } j), \not{E}_t\right\} < \frac{1.5\not{E}_t}{75} + 1.5$$