

Precision measurements in the top quark sector with the CMS experiment at the LHC

Pedro Ferreira da Silva, *for the CMS collaboration*

CERN, Genève, Switzerland

E-mail: psilva@cern.ch

Abstract. A review of the measurements of the $t\bar{t}$ production cross section (inclusive and differential) performed by the CMS experiment at the LHC is given. The results are used to constrain the magnitude of most of the relevant systematics affecting precision measurements in the top quark sector. In addition from the determination of the top quark pair production cross section it is possible to determine the strong coupling constant as $\alpha_S(m_Z) = 0.1178^{+0.0046}_{-0.0040}$ in agreement with the current world average. After the combination of several channels and techniques explored to measure the top quark mass in data, a precise determination of this quantity is made at CMS: $m_t = 173.4 \pm 0.4_{\text{stat}} \pm 0.9_{\text{syst}}$ GeV. The measurement of the difference $\Delta m_t = m_t - m_{\bar{t}}$ is also presented.

1. Introduction

Amongst the quark family the top quark has revealed experimentally exceptional characteristics. A mass above the electroweak (EWK) symmetry breaking scale ($m_t = 173.18 \pm 0.94$ GeV), a small width which compels it to decay before it can fragment and bound into a hadronic state ($\Gamma_t = 2.0^{+0.7}_{-0.6}$ GeV) and an almost exclusive decay mode into a real W boson and a b quark ($|V_{tb}| = 0.999145^{+0.000021}_{-0.000046}$) [1]. These properties give to this particle a unique role in the Standard Model (SM) of interactions - it contributes fundamentally to the radiative corrections of some of the fundamental parameters of the SM such as the Higgs boson mass. The precise determination of the top quark properties is therefore crucial to the success of understanding the SM and its limitations in particular the origin of mechanism which originates EWK symmetry breaking.

At the Large Hadron Collider (LHC) the top quark is expected to be produced abundantly: at the end of the 2012 run it is expected that $> 4 \cdot 10^6$ top quark pairs have been produced. The experiments located at the LHC have a unique opportunity to analyse this high statistics top quark sample and to study its properties with improved accuracy with respect to the Tevatron experiments. In this manuscript we highlight precision measurements performed by the CMS experiment [2] of the top quark pair production cross section. The study of the inclusive and differential cross section has been used to derive the strong coupling constant (α_S) and to characterise accurately not only the QCD $t\bar{t}$ production but also its production environment. With both an accurate description of the $t\bar{t}$ production and an improved knowledge of the performance and calibration of CMS, precision measurements of the top quark mass were made.



2. Top quark pair production

At the LHC top/anti-top pairs are mostly produced through gluon-gluon fusion ($\approx 90\%$). At CMS the measurement of the $t\bar{t}$ cross section has been carried out in all the different final states which result from the combinatorics of the W boson decays produced after the dominant $t \rightarrow Wb$ decay. The $t\bar{t}$ decay channels comprise therefore fully-hadronic (46%), lepton+jets (45%) and dileptonic (9%) final states. One crucial aspect in the measurement of the top cross section is modelling of the signal which is briefly summarised next.

2.1. $t\bar{t}$ modelling

The modelling of the $t\bar{t}$ signal at CMS is based on MADGRAPH [3] which generates $t\bar{t}$ +up to 3 additional partons at Born level. The generated matrix-element-based events are then used as input to PYTHIA [4] which takes care of performing the parton showering and simulating the proton remnants and the decays of unstable particles. TAUOLA [5] is specifically used for the decays of τ leptons. The matrix-element to parton-shower (ME-PS) matching is done using the so-called k_T - MLM algorithm. In the generation two parameters are crucial: the ME-PS matching threshold and the factorisation and normalisation scales defined as a single parameter $Q^2 = \mu_f = \mu_r = m_t^2 + \sum p_T^2(\text{parton})$. Besides characterising the scale of the hard interaction, the Q^2 -scale is also shared with the α_S -based evolution scale for ISR/FSR in PYTHIA.

2.2. Inclusive $t\bar{t}$ cross section

Using the 7 TeV proton-proton collision data the most precise measurement is obtained in the dilepton channel [6]. Even if not being the most favoured in terms of branching ratio it has high purity and it is less affected by some of the instrumental uncertainties such as jet energy scale or knowledge of the trigger/selection efficiency owing to a smaller jet activity with respect to the lepton+jets or fully hadronic channels. The selection of this channel requires at least two opposite sign charged leptons (e or μ) with a transverse momentum - $p_T > 20$ GeV - and centrally produced in the detector - $|\eta| < 2.5$ (e) or 2.1 (μ). The lepton candidates are required to be isolated in such a way that the p_T flux of the particle candidates reconstructed with the particle flow algorithm [7] in a cone of radius $R = 0.3$ built around the momentum of the lepton does not exceed 20% (17%) of the muon's (electron's) p_T . Two jets, clustered with the anti- k_T algorithm with a cone $R=0.5$, with $p_T > 30$ GeV and $|\eta| < 2.4$ are furthermore required. b -tagging information may be used in the analysis given the expectations for two b -jets in the final state. In the same flavour channels (ee or $\mu\mu$) it is also imposed that the total balance of the p_T of all particle candidates of the event (missing transverse momentum) is $E_T^{\text{miss}} > 30$ GeV. This last requirement is useful to reject further the contamination of Drell-Yan events (DY). In the same flavour channels dilepton candidates with an invariant mass compatible with the Z boson are rejected for the measurement of the cross section but counted with the purpose of estimating the DY contamination in the signal region. In the $e\mu$ channel the residual $Z/\gamma^* \rightarrow \tau\tau$ contamination is estimated from a fit to the dilepton invariant mass distribution. The contribution of events with fake lepton candidates (multijets and $W \rightarrow \ell\nu$) is estimated using a data-driven probability of selecting a fake or non-isolated lepton candidate and the control sideband sample where the lepton candidates are rejected by the isolation criteria. After event selection the purity of the sample is expected to be high ($\approx 77\%/86\%$ for the same/opposite flavour channels).

The cross section is extracted with a profile likelihood ratio method where the expected event yields for signal and background processes are parametrized taking into account the effect of the systematic uncertainties stemming from both instrumental and theoretical sources. The fit is performed to a total of 3×10 categories corresponding to the permutations of the dilepton channels, jet multiplicities and number of observed b -tagged jets (see Fig. 1, *left*). The three channels, although statistically independent, have common systematic uncertainties which are assumed to be fully correlated across the channels. The measurement yields $\sigma_{t\bar{t}}(7 \text{ TeV}) =$

$161.9 \pm 2.5_{\text{stat}} \pm 5.1_{\text{syst}} \pm 3.6_{\text{lumi}}$ pb in good agreement with the available approximate next-to-leading order (approx. NLO) calculations [8, 9, 10]. The systematic uncertainty is dominated by the theoretical uncertainty on the contribution of single top (tW) and by experimental uncertainties related to the jet energy scale, lepton efficiencies or the current knowledge on $BR(W \rightarrow l\nu)$. The result is furthermore cross checked with a simple cut and count analysis which requires at least one of the jets to be b -tagged. The same cut and count analysis has been performed at 8 TeV [11] and measures $\sigma_{t\bar{t}}(8 \text{ TeV}) = 227 \pm 3_{\text{stat}} \pm 11_{\text{syst}} \pm 10_{\text{lumi}}$ pb which allows to extract the ratio of cross sections $R_{8/7} = 1.41 \pm 0.1$ in agreement with the expectations from theory (see Fig. 1, *right*).

The cross section extracted from other channels are observed to be in agreement with the ones quoted above but have larger total uncertainty. The 4.1% (6.7%) relative uncertainty attained experimentally at 7 (8) TeV surpasses the current theoretical calculations. Work is ongoing towards the full next-to-next-to-leading order (NNLO) QCD calculation of the inclusive $t\bar{t}$ cross section, which is not yet available for all production channels.

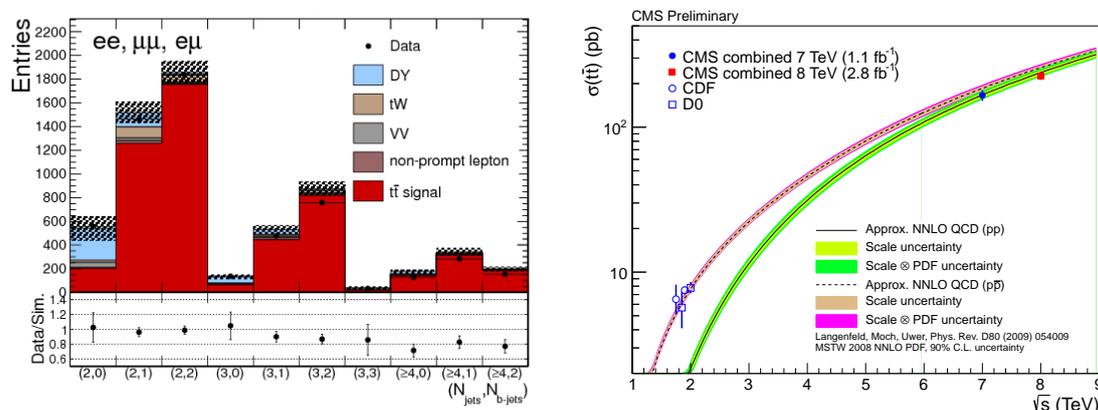


Figure 1. *Left:* Number of events selected for the three combined dilepton channels, as a function of the number of jets and b -tagged jets. The ratios of data to the sum of the $t\bar{t}$ and background predictions are given at the bottom. *Right:* Top pair cross section as function of centre-of-mass energy.

The $t\bar{t}$ cross section measurement can be re-interpreted in order to extract α_S or m_t once the dependency of $\sigma_{t\bar{t}}$ as a function of these quantities is established. Given the strong dependency of m_t on α_S both parameters cannot be extracted simultaneously. Different PDFs and approx. NLO generators have been used to derive the dependency of $\sigma_{t\bar{t}}$ on α_S (see Fig. 2, *left*). α_S is then extracted from a likelihood which convolutes the probability distribution of the experimental measurement (taking into account the experimental sources of uncertainty) with the cross section dependency on α_S which includes a rectangular prior on Q^2 . Using TOP++ with the NNPDF2.1 PDF set [12] it is determined that $\alpha_S(m_Z) = 0.1178^{+0.0046}_{-0.0040}$ [13] in good agreement with the world average. The dependency of the cross section on m_t has been parametrized for the dilepton channel measurement as: $\sigma_{t\bar{t}}/\sigma_{t\bar{t}}(m_t = 172.5) = 1.00 - 0.008 \cdot (m_t - 172.5) - 0.000137 \cdot (m_t - 172.5)^2$. At the time this manuscript measurement of m_t from the latest $\sigma_{t\bar{t}}$ is on-going and will update the preliminary result reported in [14].

2.3. Differential $t\bar{t}$ measurements

With the large statistics collected it has been possible to perform measurements of the differential $t\bar{t}$ production cross section: $1/\sigma_{t\bar{t}} \cdot d\sigma_{t\bar{t}}/dx$ where x is a kinematical variable of the $t\bar{t}$ system,

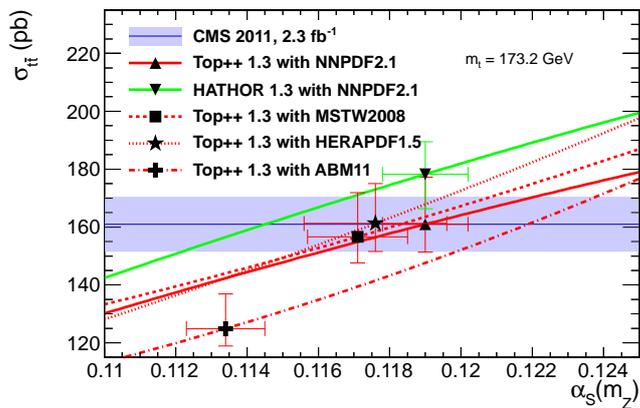


Figure 2. Evolution of the $t\bar{t}$ cross section as function of α_s for different calculations at approx. NNLO and different PDF sets.

or of each top quark in the event [15] or related to the jet activity [16, 17] or E_T^{miss} [18] in the event.

Measurements of $\sigma_{t\bar{t}}$ as function of the top/ $t\bar{t}$ system variables require the reconstruction of the kinematics of the event. In the lepton+jets channel, where all the degrees of freedom are specified it is possible to use a constrained kinematics fit to choose the best combination of leptons and jets to reconstruct the $t\bar{t}$ system. In the dilepton channel this is no longer the case owing to the presence of at least two neutrinos in the final state which lead to an unspecified degree of freedom in the kinematics. In this case the expected distribution for the neutrino energies (from simulation) is used to rank possible solutions in the $100 < m_t < 300$ GeV range. Solutions with b -tagged jets are prioritised. After the reconstruction of the kinematics an unfolding procedure is used in order to estimate the particle level kinematics. The unfolding procedure takes not only into account the intrinsic efficiencies and the resolution of the physics objects, but also possible contributions from combinatorial mis-assignment of these objects in the reconstruction of the $t\bar{t}$ system. A regularised unfolding method has been used for this purpose where for each distribution a response matrix, computed based on simulated events, is used to map the migrations and efficiencies. The generalised inverse of the matrix is obtained after minimising a χ^2 and it is furthermore used to obtain the unfolded distributions. The results are observed to be in good agreement among the different channels analysed and also with the SM predictions at both NLO (obtained with the POWHEG [19] and MC@NLO [20] generators) and approx. NNLO where calculations are available. Figure 3 (*left*) shows the good agreement obtained with the approx. NLO prediction for the case of the p_T of the top quark.

Besides measuring differentially the properties of the $t\bar{t}$ system one can learn about perturbative QCD (pQCD) by studying the properties of the “ $t\bar{t}$ environment”: jet activity, E_T^{miss} are two examples. In particular it is crucial to address how well does pQCD model the data and to identify its main dependencies and uncertainties related to the Q^2 scale, ME-PS matching thresholds, hadronisation and fragmentation effects. These are crucial characterisations on the road to the search for new physics effects underlying the $t\bar{t}$ sample as well as to establish the $t\bar{t}H$ vertex where at least 4 jets are expected in the case $H \rightarrow b\bar{b}$. Precision measurements such as m_t have, in general, a dependency on these parameters as well.

An interesting technique is employed to obtain pure distributions for $t\bar{t}+N$ jets without having to resort to an unfolding procedure. The distribution of the $\sqrt{\chi^2}$ obtained after the fit for the reconstructed kinematics in the lepton+jets channel is analysed using distribution functions derived from simulation for different jet multiplicities. After subtracting the background contribution and comparing to different generators it is observed that both MADGRAPH+PYTHIA and POWHEG+PYTHIA yield a similar level of agreement with the data,

while MC@NLO+HERWIG undershoots higher jet multiplicities. Different MADGRAPH settings with varied Q^2 and ME-PS thresholds (by factors of 2 and 1/2 with respect to the nominal values) are also used to compare with the variations observed in data. The so called jet gap fraction (fraction of events with no jet above a given p_T threshold) is particular sensitive to these variations and it is shown in Fig. 3, *right*. It displays not only the preference for a higher Q^2 value than the one used by the nominal simulation in CMS but also the fact that the variations considered to evaluate the systematic uncertainties pertaining these parameters are in good agreement with what is observed in data. It is important to conclude referring that in these studies the uncertainties ranges usually from 3% at low multiplicities to 20% at high multiplicities and are dominated by jet energy scale and $t\bar{t}$ modelling parameters.

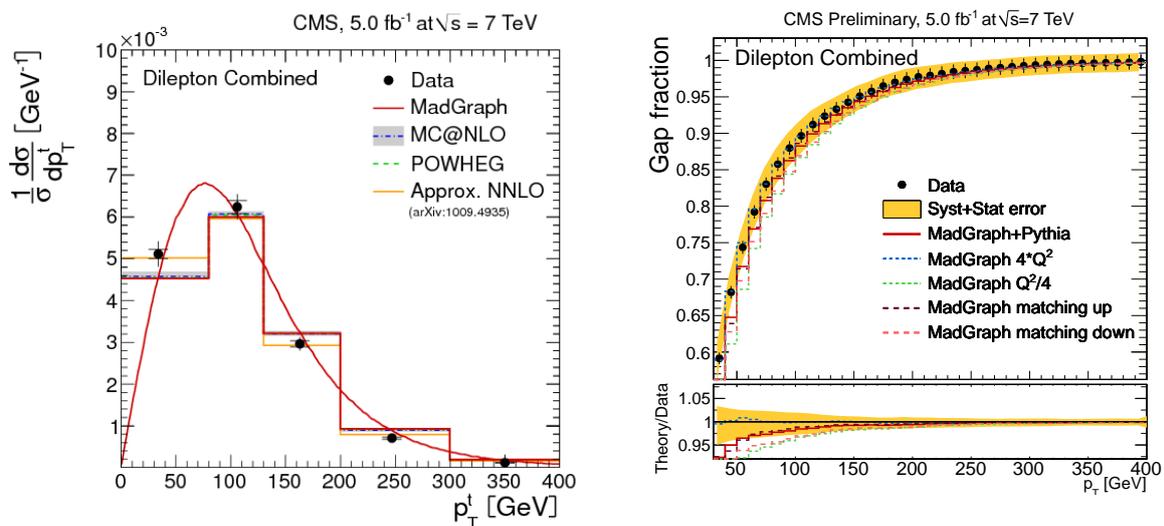


Figure 3. *Left:* Normalised differential $\sigma_{t\bar{t}}$ in the dilepton channel as function of the p_T of the top quark. The inner (outer) bars represent the statistical (total) uncertainty. *Right:* Measured gap fraction as a function of the additional jet p_T compared to different MADGRAPH settings. The shaded band represents to total uncertainty (stat+sys).

3. Top quark mass

At CMS the top quark mass has been measured in the lepton+jets [21], dilepton [22, 23] and full hadronic [24] channels. The results obtained in each channel are found to be in agreement with each other and with the current world average. After the combination of the different measurements, using a Best Linear Unbiased Method technique [25], CMS measures $m_t = 173.4 \pm 0.4_{\text{stat}} \pm 0.9_{\text{sys}}$. The total uncertainty attained deserves to be explored in more detail, in particular the measurement in the lepton+jets channel which drives the final result.

The measurement in the lepton+jets channel is performed simultaneously with the in-situ calibration of the jet energy scale using the $W \rightarrow qq'$ reconstructed decay. A kinematics fitter is used to evaluate all possible combinations of jets available for the reconstruction of the $t\bar{t}$ decays. Each single combination is assigned with a probability of the goodness-of-fit based on the χ^2 value: $P_{\text{gof}} = e^{-\chi^2/2}$. To increase the purity and also the fraction of assignments which are found to be correct events with at least two b -tagged jets and $P_{\text{gof}} > 0.5$ are required. This tight requirements depletes the original statistics by 71% but increases the purity (90% \rightarrow 96%) and the fraction of correct assignments (13% \rightarrow 44%). It's the latter number which is relevant

to attain an optimal resolution in the reconstruction of m_t . The method proceeds with the construction of an ideogram where each event is assigned a probability based on P_{gof} and on the probability that each assignment was originated after one of three possibilities: correct or wrong assignment of the original kinematics and unmatched assignment (either due to the rejection of one jet or due to the fact the the event is from residual backgrounds). The per-event probabilities are then multiplied to build a likelihood which is function of m_t and the jet energy scale (JES). Both m_t and JES are left to float freely when maximising the following likelihood:

$$\mathcal{L}(m_t, \text{JES}) \propto \prod_{\text{events}} \left[\sum_{i=1}^n cP_{\text{gof}}(i) P \left(m_{t,i}^{\text{fit}}, m_{W,i}^{\text{reco}} | m_t, \text{JES} \right) \right]^{w_{\text{event}}}$$

The method is calibrated based on simulated events with both different m_t and JES scenarios which are used to assess the bias and the statistical coverage of the fit. Small corrections to the final measurement (<0.5 GeV) are estimated based on this studies.

The systematic uncertainties affecting the measurement can be grouped in two main groups: experimental and theoretical sources. The experimental-type uncertainties comprehend jet energy scale and resolution effects, b -tagging, E_T^{miss} , pileup, lepton energy scale. Although the jet energy scale is calibrated in-situ only an average scale is derived and it pertains specifically to light flavoured jets. Consequently a flavour-specific uncertainty and a residual p_T and η -dependency have to be propagated to the measurement of m_t by using simulated pseudo-experiments. These two aspects result as the dominant sources of uncertainty to the measurement. Theory uncertainties are related to the modelling of the $t\bar{t}$ signal its production environment. These are also evaluated using pseudo-experiments with alternative parameter settings of the generator which reflect the different sources of uncertainties. Both Q^2 and ME-PS matching scale uncertainties are estimated using $t\bar{t}$ samples where these parameters are varied independently by a factor of 2 or 1/2. These variations will reflect not only in the $t\bar{t}$ kinematics but also in different ISR/FSR evolution. As previously exposed in Section 2.3 we expect these variations to constitute an accurate envelope of the observed data with respect to the uncertainty in the description of the hadronic activity in $t\bar{t}$ events. Underlying event-related uncertainties are evaluated using different variations of the so-called PERUGIA11 tune [26]. This UE tune has been adopted by both CMS and ATLAS collaborations. The PERUGIA11 variations cover different alternative scenarios for FSR and hadronisation, ISR and primordial k_T underlying event, beam remnants, colour reconnections and energy scaling. Of particular interest is the PERUGIANOCR where the description of data is done without resorting to the modelling of colour reconnection and which is used to conservatively estimate the effect on m_t reconstruction. Colour reconnection is the second main source of systematic uncertainty in the measurement of m_t . Table 1 summarises the different sources of uncertainty affecting the measurement of m_t in the lepton+jets channel.

Measurements in other $t\bar{t}$ channels have larger uncertainties with respect to the measurement described above. Most uncertainties are induced by instrumental effects: in the dilepton channel no in-situ jet energy scale calibration is performed and there is a strong dependency on the E_T^{miss} estimate which is used to constrain the kinematics of the escaping neutrinos. In the full-hadronic channel the uncertainty is larger due to resolution effects stemming from the high number of combinatorial jet assignments and due to the non-negligible multijets background contamination ($\approx 60\%$). Figure 4 (*left*) shows the reconstructed top quark mass spectrum in the fully hadronic channel.

An interesting exception with respect to “canonical” measurements of m_t is the analysis of kinematics endpoints using the dilepton channel. This method is not based on a simulation-based calibration and uses only kinematics distributions reconstructed in data to fit both the spectrum and the endpoints using LO calculations. Fig. 4 (*right*) shows an example of a fit to

Table 1. Summary of the main systematic uncertainties affecting the measurement of m_t in the lepton+jets channel.

Type	Source	Δm_t [GeV]
Method	Calibration	0.06
Instrumental	b jet energy scale	0.61
	p_T and η -dependent jet energy scale	0.28
	Jet energy resolution	0.23
	Lepton energy scale	0.28
	E_T^{miss}	0.06
	b -tagging efficiency	0.12
	Pileup	0.07
	Non- $t\bar{t}$ background	0.13
Theory	PDF	0.07
	$Q^2(\mu_r, \mu_f)$	0.24
	ME-PS matching scale	0.18
	Underlying event	0.15
	Color reconnection	0.54
Total		0.98

the lepton-jet invariant mass spectrum. The measurement makes further use of other variables which factorize event by event the boost of the $t\bar{t}$ system and which are suited to analyse events with symmetric three body decays [27] such as the ones expected in several BSM scenarios. In the endpoint-analysis the systematic uncertainties sources are partially orthogonal to the ones affecting the other measurements and in the future are expected to improve further the final combination for the measurement of m_t . In the future the study of alternative methods which factorize specific systematic uncertainties such as jet energy scale or colour reconnection will hopefully contribute to a better definition and measurement of m_t .

The measurement of the difference between m_t and $m_{\bar{t}}$ has also been performed by using ℓ^+ +jets and ℓ^- +jets events. The result is agreement with the SM prediction and finds no evidence for any source of CPT violation in the top quark sector: $\Delta m_t = -0.44 \pm 0.46_{\text{stat}} \pm 0.27_{\text{syst}}$ GeV. It's important to notice that as a difference is taken most of the systematic uncertainties cancel out yielding a statistically dominated precise measurement of this quantity [28].

4. Summary

In this presentation we have focused on specific precision measurements in the top quark sector such as the production cross section of $t\bar{t}$ events and the measurement of m_t using the CMS detector. All measurements are found to be overall consistent with the SM predictions and have attained uncertainties comparable or surpassing the current theoretical predictions. In particular the precision in the measurement of m_t opens the window to more detailed studies with a high statistics sample which will hopefully lead to a better understanding of this quantity and of its interplay with the EWK symmetry breaking mechanism.

References

- [1] J. Beringer *et al.* [Particle Data Group Collaboration], Phys. Rev. D **86** (2012) 010001.
- [2] S. Chatrchyan *et al.* [CMS Collaboration], JINST **3** (2008) S08004.
- [3] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, JHEP **1106** (2011) 128 [arXiv:1106.0522 [hep-ph]].

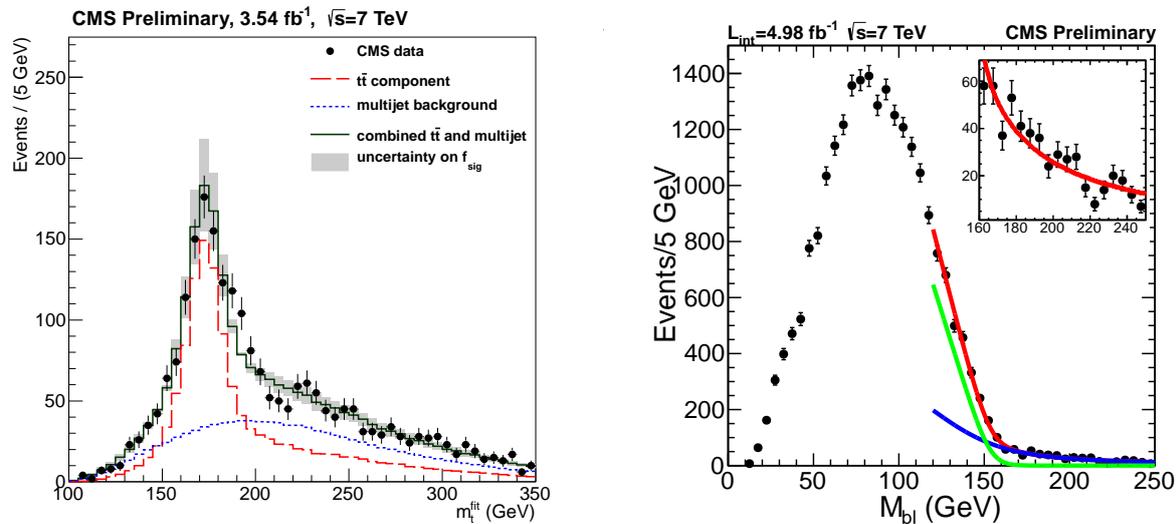


Figure 4. *Left:* Reconstructed m_t from a kinematic fit in the fully hadronic final state. The uncertainty band indicates the uncertainty from the signal fraction. *Right:* Result of the fit to the lepton-jet invariant mass spectrum and its endpoint using dilepton events. The inset shows a zoom of the tail region in M_{bl} spectrum to illustrate the level of agreement between the background shape and the data points.

- [4] T. Sjostrand, L. Lonnblad, S. Mrenna and P. Z. Skands, hep-ph/0308153.
- [5] Z. Was, Nucl. Phys. Proc. Suppl. **98** (2001) 96 [hep-ph/0011305].
- [6] S. Chatrchyan *et al.* [CMS Collaboration], JHEP **1211** (2012) 067 arXiv:1208.2671 [hep-ex]
- [7] CMS collaboration, CMS-PAS-PFT-10-002 (2010)
- [8] N. Kidonakis, Phys. Rev. D **82** (2010) 114030 arXiv:1009.4935 [hep-ph]
- [9] V. Ahrens, A. Ferroglia, M. Neubert, B. D. Pecjak and L. L. Yang, JHEP **1009** (2010) 097 [arXiv:1003.5827 [hep-ph]].
- [10] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer and M. Wiedermann, Comput. Phys. Commun. **182** (2011) 1034 arXiv:1007.1327 [hep-ph]
- [11] CMS collaboration, CMS-PAS-TOP-12-007 (2012)
- [12] M. Czakon and A. Mitov, arXiv:1112.5675 [hep-ph]
- [13] CMS collaboration, CMS-PAS-TOP-12-022 (2012)
- [14] CMS collaboration, CMS-PAS-TOP-11-008 (2011)
- [15] S. Chatrchyan *et al.* [CMS Collaboration], arXiv:1211.2220 [hep-ex]
- [16] CMS collaboration, CMS-PAS-TOP-12-018 (2012)
- [17] CMS collaboration, CMS-PAS-TOP-12-023 (2012)
- [18] CMS collaboration, CMS-PAS-TOP-12-019 (2012)
- [19] S. Alioli, P. Nason, C. Oleari and E. Re, JHEP **1006** (2010) 043 [arXiv:1002.2581 [hep-ph]].
- [20] S. Frixione and B. R. Webber, JHEP **0206** (2002) 029 [hep-ph/0204244].
- [21] S. Chatrchyan *et al.* [CMS Collaboration], arXiv:1209.2319 [hep-ex].
- [22] S. Chatrchyan *et al.* [CMS Collaboration], Eur. Phys. J. C **72** (2012) 2202 [arXiv:1209.2393 [hep-ex]].
- [23] CMS collaboration, CMS-PAS-TOP-11-027 (2012)
- [24] CMS collaboration, CMS-PAS-TOP-11-17 (2012)
- [25] CMS collaboration, CMS-PAS-TOP-11-18 (2012)
- [26] P. Z. Skands, Phys. Rev. D **82** (2010) 074018 [arXiv:1005.3457 [hep-ph]].
- [27] K. T. Matchev and M. Park, Phys. Rev. Lett. **107** (2011) 061801 [arXiv:0910.1584 [hep-ph]].
- [28] S. Chatrchyan *et al.* [CMS Collaboration], JHEP **1206** (2012) 109 [arXiv:1204.2807 [hep-ex]].