IL NUOVO CIMENTO Vol. ?, N. ? ?

The $t\bar{t}H$ analyses at the LHC

G. Aad(1) and J. Steggemann(2) for the ATLAS and CMS collaborations.

(1) *Centre de Physique des Particules de Marseille, CNRS/IN2P3 and Université de la Méditerranée Aix-Marseille II*
(2) *III. Physikalisches Institut A, RWTH Aachen University*

Summary. — Detecting the presence of a light Higgs boson at the LHC is very difficult. For this reason, different experimental signatures will have to be combined to ensure its discovery. Among them we study the associate production with a pair of top quarks and the subsequent Higgs boson decay into b quark pairs, the dominant decay mode for $m_H \lesssim 135\,\mathrm{GeV/c^2}$. This channel allows an accurate estimation of the top quark Yukawa coupling within the Standard Model. We present several observability studies of the $t\bar{t}H(\to b\bar{b})$ channel with the ATLAS and the CMS detectors. In addition, the decay modes $H \to WW$ and $H \to \gamma\gamma$ in $t\bar{t}H$ processes have been investigated and are briefly reported at the end of this paper.

PACS `14.80.Bn` – Top quarks.
PACS `14.65.Ha` – Standard-model Higgs bosons.

1. – Top-associated Higgs boson production at the LHC

The Large Hadron Collider (LHC), the new proton-proton collider at CERN, is scheduled to start operations at a centre-of-mass energy $\sqrt{s} = 10\,\mathrm{TeV}$ in 2008. The design energy of $\sqrt{s} = 14\,\mathrm{TeV}$ will be reached in 2009. In this phase, the LHC is expected to run at an instantaneous luminosity of $10^{33}\,\mathrm{cm^{-2}s^{-1}}$, and each of the two general-purpose experiments, ATLAS and CMS, will accumulate an integrated luminosity of about $30\,\mathrm{fb^{-1}}$ within 3 years of data-taking.

At the LHC, the Higgs boson will be produced mainly by gluon-gluon fusion through a top loop, this process is commonly referred to as "direct production". The second most important production modes are $q\bar{q}H$ via vector boson fusion, and associated production modes, *e.g.* WH, ZH, and $t\bar{t}H$. Despite having 2 to 3 orders of magnitude lower cross sections, the latters present more distinct final state signatures which can lead to a better suppression of the physics background.

The dominant decay mode for a light Higgs boson, as favoured by the electroweak precision data [1], is $H \to b\bar{b}$. The overwhelming background from QCD $b\bar{b}$ and multi-jet production makes it impossible to detect this final state in the Higgs boson direct production. Therefore, the main focus of this report is the observability of the $H \to$

$b\bar{b}$ when the Higgs boson is produced in association with a top-antitop quark pair for both ATLAS and CMS [2] experiments. In addition to the possible contribution to the discovery of a low-mass Higgs boson, this channel allows an accurate estimation of the top quark Yukawa coupling. The CMS study discussed here considers three different final states with respect to the number of leptons (0, 1, 2), while the ATLAS study concentrates on the final state with one lepton, and the analysis carried out with three different techniques. Sect. **2** presents the topology of the $t\bar{t}H(H \to b\bar{b})$ channel. In Sect. **3**, the generated signal and background samples are discussed in detail. Sects. **4** to **6** contain the analysis details as well as the results for the different decay topologies. Finally, an ATLAS study of the $H \to WW$ mode and a CMS study of the $H \to \gamma\gamma$ channel are discussed in Sects. **7** and **8**, respectively.

2. – The $t\bar{t}H(H \to b\bar{b})$ channel

Since within the Standard Model top quarks decay about 100% to a W boson and a b quark, the $t\bar{t}H(H \to b\bar{b})$ channel exhibits a striking signature due to the presence of four b-jets in the event. Different final states are categorised according to the number of leptons and light-flavoured jets from the decay of the two W bosons. One can distinguish three configurations: no lepton and 8 jets (full-hadronic), 1 lepton and 6 jets (lepton-plus-jets), and 2 leptons plus 4 jets (dileptonic).

The correct identification of b-jets is extremely important to reject background events. The main physical background is $t\bar{t}$ production in association with two or more extra jets. When the extra jets are b-jets, the signature is exactly like the signal.

Both ATLAS and CMS follow a similar general strategy. First, the $t\bar{t}$ system is reconstructed, then the remaining two b-jets yield the reconstructed Higgs boson mass distribution. Resolving the jet combinatorics and thereby finding the jets originating from the Higgs boson is crucial for all analyses, especially for the full-hadronic and the lepton-plus-jets channels. In addition, small uncertainties on the jet energy scale corrections (JES) and on the (b quark) jet resolutions are necessary to be able to see a signal peak on top of the background $b\bar{b}$ mass distribution.

3. – Event Generation and Detector Simulation

The ATLAS and CMS studies use different Monte Carlo generators for the signal and background processes. Both consider the background coming from $t\bar{t}$-plus-jets production to be the most important background. The contributions from other background processes are found to be negligible, especially W-plus-jets and QCD multijet background, except for the fully hadronic channel. Here, the CMS study estimates the QCD multijet background with PYTHIA [3], which is not expected to well model events with many partons in the final state. For this reason this particular background will have to be estimated directly from data. Table I lists the generators and the values for the cross sections used for each process.

Due to the small signal cross section, the studies require a high amount of recorded data. The ATLAS studies are performed assuming an integrated luminosity of $30\,\mathrm{fb}^{-1}$, while CMS assumes $60\,\mathrm{fb}^{-1}$.

The CMS studies also include the pile-up effects expected from an instantaneous luminosity of $10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. For both experiments, the full detector simulation is used.

Table I. – *Monte Carlo generators and simulated cross sections used by the ATLAS and CMS analyses. All generators, except MC@NLO which uses Herwig [7], are interfaced to PYTHIA for the simulation of the initial and final state radiation, hadronisation, and further decay.*

	ATLAS		CMS	
Sample	Generator	σ(pb)	Generator	σ(pb)
$t\bar{t}H$	Pythia	0.537 LO	CompHEP[5]	0.664 NLO
$t\bar{t}b\bar{b}$ QCD	AcerMC[4]	8.7 LO	CompHEP	3.28 LO
$t\bar{t}b\bar{b}$ EW	AcerMC	0.94 LO	CompHEP	0.65 LO
inclusive $t\bar{t}$	MC@NLO[6]	833 NLO+NLL	-	
inclusive $t\bar{t}$+1jet	-		Alpgen[8]	371 LO
QCD	-		Pythia	4.9×10^5 LO

4. – The lepton-plus-jets channel

For both experiments the focus is on the lepton-plus-jets channel. This final state consists of one lepton with high transverse momentum (p_T), typically used for triggering, four b-jets, at least two light-flavoured jets, and missing transverse energy due to the presence of a neutrino. The decay of τ leptons is not explicitly considered. A first preselection step, requiring at least one lepton and at least six jets, is used to reduce non-top backgrounds. The additional requirement of four b-tagged jets is used to significantly suppress $t\bar{t}$-plus-light-jets events. The following step is a full reconstruction of the $t\bar{t}$ system. The two remaining b-jets are then combined to form a Higgs boson candidate.

4·1. *Triggers and preselection.* – Neither ATLAS nor CMS developed a dedicated $t\bar{t}H$ trigger. A general trigger, requiring one isolated high p_T electron or muon, yields a reasonable trigger efficiency ($\approx 65\%$).

For the offline reconstruction, one high p_T lepton (electron or muon) is required. For ATLAS, this lepton is required to pass identification, acceptance, and isolation cuts. CMS cuts on a likelihood discriminant formed by combining several related observables. Finally a veto on the presence of a second lepton is applied. ATLAS demands at least 6 jets with $p_T > 20\,\mathrm{GeV/c}$ and a pseudorapidity $|\eta| < 5$. For CMS, events with 6 or 7 reconstructed jets and $p_T > 20\,\mathrm{GeV/c}$ and $|\eta| < 3$ are accepted.

In ATLAS, b-jets are identified using a multivariate tagger which uses track information, such as the track impact parameter, and properties of the secondary vertices [9]. CMS as well uses a combined secondary vertex b-jet tagging algorithm. Two extensions have been developed in the course of this analysis, the explicit reconstruction of tertiary vertices, and the inclusion of a soft lepton tagging algorithm. The choice of the b-tagging criterion in terms of b-jet efficiency ϵ_b and background rejection has to fulfil two requirements: On the one hand, it is crucial to have a high light-jet rejection to control the $t\bar{t}$-plus-jets events, the dominant background after the previous selection cuts. On the other hand, the event selection efficiency is proportional to ϵ_b^4. Therefore it is crucial to maintain a high b-tagging efficiency. *E.g.* in ATLAS, a b-tagging efficiency higher than 65% is usually selected. This allows the reduction of the $t\bar{t}$-plus-jets background by a factor of 50.

In the CMS study, a two-fold approach is applied. First, four "loosely" b-tagged jets are required. Then, the b-taggging weights are combined into a b-tagging likelihood. Two choices of the cut on this discriminator variable are made. The first one maximises the signal significance $S/\sqrt{B}$ and defines the "loose" working point of the analysis. The other one maximises the signal purity S/B, yielding the "tight" working point.

ATLAS has the same approach for one of the reconstruction techniques discussed in Sect. **4**˙2, the constrained fit analysis. Regarding the cut-based and likelihood analyses, the other two reconstruction techniques used by ATLAS, tight cuts are directly applied on the 4 jets. The effective efficiencies of both methods are comparable.

4˙2. *Reconstruction.* – The identification of the two b-jets produced by the Higgs boson decay first requires a full reconstruction of the $t\bar{t}$ system to recognise the two b-jets associated to the top decay chain. The remaining two b-jets define the Higgs system. For a full kinematic reconstruction of the leptonically decaying top quark, the longitudinal momentum of the neutrino needs to be estimated. This is done by imposing a W mass constraint on the invariant mass of the reconstructed lepton and the neutrino candidate from the missing transverse energy.

To be able to solve the jet combinatorics and reconstruct the $t\bar{t}$ system, ATLAS uses 3 different approaches:

- The cut-based analysis uses a χ^2 minimisation with constraints on both top quark masses. All combinations with $|m_{top_{reco}} - m_{top}| > 25$ GeV/c^2 or $|m_{W_{reco}} - m_W| > 25$ GeV/c^2 are excluded. The combination that minimises the χ^2 is chosen.

- The pairing likelihood analysis exploits 6 kinematic variables of the $t\bar{t}$ system. The combination that maximises the resulting likelihood output is kept. To increase the purity, a final selection cut is applied. Fig. 1 shows the invariant mass of both reconstructed top quarks using the likelihood technique.

- The constrained fit analysis uses the jet resolution information and fits the jet p_T and the E_{Tmiss} with W and top quark mass constraints. The χ^2 of the fit is combined with kinematic and b-tagging variables to build a multi-dimensional likelihood. The combination that maximises the likelihood output is kept. A final selection likelihood is used for additional separation between the signal and the physical background.

In the CMS analysis, several strategies have been studied. All yield comparable efficiencies of $\sim 30\%$. For the results, a combined event likelihood is used, including kinematic variables, b-tagging variables, and the top quark and W boson masses.

The Higgs invariant mass distributions for the ATLAS likelihood analysis and the CMS study are shown in Figs. 2 and 3. The intrinsic jet resolution accounts for the broadness of the central peak, whereas the combinatorial background, *i.e.* events where at least one of the b-jets is wrongly assigned to the Higgs boson, significantly reduces the overall Higgs boson mass resolution. The Higgs purity, *i.e* events where both b-jets are correctly assigned to the Higgs boson with respect to all reconstructed events, is about 30% for both experiments. Figs. 2 and 3 show the invariant mass of the Higgs boson candidates for both signal and background events. Both for the CMS and ATLAS studies, the signal shape does not offer a clear separation from background. To be able to extract the $t\bar{t}H$ signal, hence a good knowledge of the background shape and normalisation is necessary.

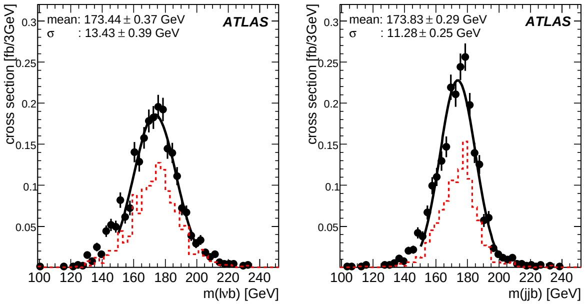

Fig. 1. – On the left-hand (right-hand) side, the invariant mass distribution of the reconstructed leptonic (hadronic) top quark candidates is shown for the configuration with the maximum likelihood. The dotted histograms indicate solutions with a correct b-jet assigned to the considered top quark.

5. – All-hadronic and dileptonic channel

The CMS study also includes the all-hadronic and the dileptonic channels. For both channels, an orthogonal event selection is applied to allow for a straightforward combination of the results with the lepton-plus-jets channel.

In the dileptonic channel, two well-identified leptons passing lepton-likelihood cuts are required as well as a missing transverse energy of at least 40 GeV and 4-7 jets, of which 3-4 have to be identified as b-jets. A counting experiment is carried out in the end, *i.e.* no explicit reconstruction of the Higgs boson mass is performed.

Due to the presence of eight jets in the final state, the all-hadronic analysis places emphasis on the choice of the jet reconstruction algorithm. A detailed study indicates that a cone size of ΔR =0.4 gives the best results. The final analysis makes use of a

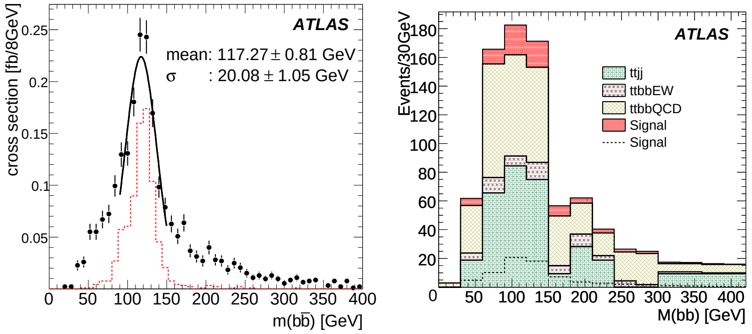

Fig. 2. – Invariant mass of the reconstructed Higgs boson candidates for the ATLAS study. On the left, only signal events are considered and the dotted histogram indicates the correct combinations. On the right, the signal is shown on top of the background distribution. Both plots are normalised to an integrated luminosity of 30 fb^{-1}.

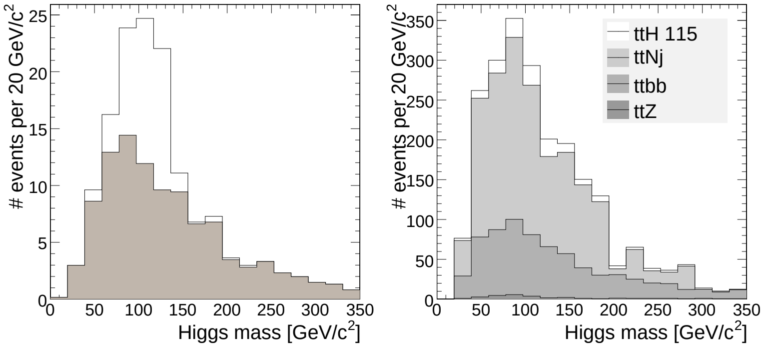

Fig. 3. – Invariant mass of the reconstructed Higgs boson candidates for the CMS study. On the left, only signal events are considered where the combinatorial background is shaded grey. On the right, the signal distirbution (white) is shown on top of the background distribution. The numbers of events are normalised to an integrated luminosity of $60\,\mathrm{fb}^{-1}$.

number of kinematic variables, and a mass χ^2 to reduce the combinatorial background. The cuts are optimised to give a high significance $S/\sqrt{B}$ while keeping a high purity S/B.

6. – Results for the $t\bar{t}H(H \to b\bar{b})$ channel

Table II lists the systematic uncertainties for the different ATLAS analyses and the different channels in the CMS analysis. For both experiments the main uncertainties come from the b-tagging efficiency, the jet energy scale, and the jet energy resolution, and have comparable sizes.

Table II. – *Combined systematic uncertainties on selection efficiencies for the ATLAS and CMS analyses for the $t\bar{t}H(H \to b\bar{b})$ channel.*

	Cut-based		Pairing likelihood		Constrained fit	
ATLAS	signal	background	signal	background	signal	background
	±18%	±22%	±20%	±25%	±19%	±28%
	semileptonic		All-hadronic		Dileptonic	
CMS	signal	background	signal	background	signal	background
	±22%	±34%	±20%	±27%	±11%	±18%

In table III, the significances for the different analyses are shown. In all analyses a counting experiment is performed and the significance is given in terms of $S/\sqrt{B}$ and $S/\sqrt{B + \Delta B^2}$, where ΔB is the uncertainty of the background yield estimation.

In the CMS study, the lepton-plus-jets analysis gives the highest statistical significance. The combined significance of all three analyses is 3.32 for the "loose" working points. The significances with systematics included are given for the "tight" working points: due to the high uncertainties on the background yields, the highest significance is obtained when the selection is optimised for high purity S/B. The final combined significance expected from the CMS analysis, after adding the systematic uncertainties, is 0.41 for $60\,\text{fb}^{-1}$ of data. To increase the signal purity, ATLAS uses a final mass window cut of $30\,\text{GeV/c}^2$ around the Higgs mass before computing the significance. The best significance is 2.18 and is given by the constrained fit analysis for $30\,\text{fb}^{-1}$. After adding the systematic uncertainties, the cut-based analysis gives the best significance of 0.49.

Both studies indicate clearly that the identification of the $t\bar{t}H(H \to b\bar{b})$ signal at the LHC will be very challenging. The measurement of the background level and shape directly from real data control samples will be crucial to reduce the systematic uncertainties and to reach a much better signal significance. In parallel, higher order calculations help to reduce theoretical uncertainties on background cross sections. The next-to-leading order calculation is already available for the $t\bar{t}$-plus-one-jet [10] and the $q\bar{q} \to t\bar{t}b\bar{b}$ [11] processes.

TABLE III. – *Expected significances for the ATLAS and CMS studies at $m_H = 120$ GeV/c^2. For CMS, the signal-to-background ratio and the significance are given for the "loose" working point, the significance with systematics for the "tight" working point.*

	S/B	$S/\sqrt{B}$	$S/\sqrt{B+\Delta B^2}$
ATLAS cut-based (30 fb^{-1})	0.11	1.82	0.49
ATLAS likelihood (30 fb^{-1})	0.10	1.95	0.40
ATLAS constrained fit (30 fb^{-1})	0.12	2.18	0.43
CMS semileptonic (60 fb^{-1})	0.053	2.5	0.29
CMS all-hadronic (60 fb^{-1})	0.015	2.4	0.22
CMS dileptonic (60 fb^{-1})	0.018	1.4	0.27
CMS combined (60 fb^{-1})	-	3.32	0.41

Since the presented analyses will only be performed after several years of data-taking, many improvements relying on information extracted from real data are expected. On the one hand, the shape of background distributions can be extracted from data. This will open several possibilities, *e.g.* an analysis which combines the shapes of the distributions of a number of variables using a multivariate classifier instead of doing a counting experiment, or a side-band analysis. On the other hand, improvements of algorithms, *e.g.* b-tagging and jet reconstruction, are foreseen, or already at hand.

7. – The $t\bar{t}H(H \to WW)$ channel

The $t\bar{t}H(H \to WW)$ channel is expected to give valuable information about the properties of the Higgs boson. One can consider several final state configurations depending on the decay of each of the four W bosons. The ATLAS experiment has studied this channel, combining the final states with two (2L) and three leptons (3L), where leptons are either electrons or muons. The main background for both channels is $t\bar{t}$X production.

An isolated high p_T lepton trigger is enough to ensure high efficiency for the signal, more than 80% for the 2L channel and more than 90% for the 3L analysis. For the 2L (3L) channel, 2 (3) isolated high p_T leptons are required to pass identification cuts. For the 2L channel, the 2 leptons are required to have the same charge. All events must have also at least 6 (4) jets for the 2L (3L) channel and must pass a Z-veto

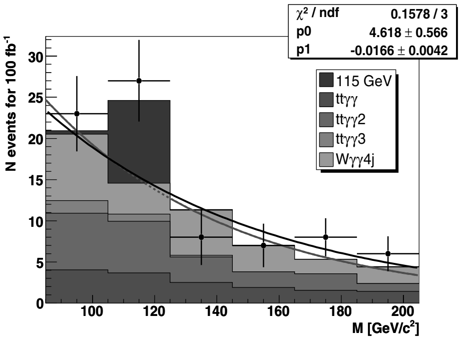

Fig. 4. – Background estimation from sideband fit in the reconstructed di-photon invariant mass spectrum for a Higgs boson with $m_H = 115$ GeV and 100 fb^{-1} of data (CMS experiment).

($75 < m_{\ell\ell} < 100\,\text{GeV/c}^2$). This selection yields a S/B of 1.85/10.3 for the 2L analysis and 0.82/3.4 for the 3L analysis. This analysis is very challenging and the background has currently large uncertainties and an ongoing analysis is trying to develop some strategies in order to better estimate it.

8. – The $t\bar{t}H(H \to \gamma\gamma)$ channel

The $t\bar{t}H(H \to \gamma\gamma)$ channel exhibits an even more distinct signature in contrast to the decay of the Higgs boson into a $b\bar{b}$ pair, offering a high intrinsic background rejection. A measurement of this production channel will allow a precise determination of the top quark Yukawa coupling. Due to the low production cross section times branching ratio, this channel will most probably be interesting for a high amount of integrated luminosity.

The CMS experiment studied this channel in the Physics Technical Design Report [12]. The analysis exploits the semileptonic final state. All irreducible backgrounds are considered, *i.e.* top quark pair production with two additional photons, and the production of $W\gamma\gamma$ with four additional jets. Additional vetoes against reducible backgrounds which could not be simulated with sufficient statistics are applied.

In the selection step, two high p_T isolated photons passing several quality criteria are required. In addition, an isolated lepton and at least four high p_T jets, of which at least one is identified as a b-jet, need to be present. Fig. 4 shows how the background can be estimated from sidebands by fitting an exponential function to the measured di-photon invariant mass spectrum.

The presented study yields a signal-to-background fraction of 4:1. This indicates that the Higgs boson can be observable in the $t\bar{t}H(H \to \gamma\gamma)$ channel in excess of 3 σ with 100 fb^{-1} of data.

9. – Summary

The ATLAS and the CMS experiments have performed studies of $t\bar{t}H$ production at the LHC. This channel allows an accurate measurement of the top quark Yukawa coupling, and might contribute to the discovery of the Higgs boson.

The presented ATLAS and CMS analyses of the $t\bar{t}H(H \to b\bar{b})$ channel, important for low Higgs boson masses $\lesssim 135\,\text{GeV/c}^2$, yield comparable results. Both perform a complete kinematic reconstruction of the event using advanced analysis methods, necessary to separate the signal events from background. The results indicate that with the present analysis strategies, the observation of the Higgs boson in this channel will only be possible if the size of the systematic uncertainties can be reduced compared to the present estimations, *e.g.* by developing techniques for the extraction of background shapes and normalisation from data.

In addition, studies of the $t\bar{t}H(H \to WW)$ and the $t\bar{t}H(H \to \gamma\gamma)$ channels have been shown. The ATLAS $t\bar{t}H(H \to WW)$ analysis results in a measurement of the production cross section given 30 fb^{-1} and a Higgs boson mass of 160 $\mathrm{GeV/c}^2$. The CMS analysis of the $t\bar{t}H(H \to \gamma\gamma)$ channel indicates that a light Higgs boson can be observed in this channel with 100 fb^{-1} of data.

* * *

This work was in part supported by the Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen, the Bundesministerium für Bildung und Forschung (BMBF), the Deutsche Forschungsgemeinschaft (DFG), and by the Marie Curie International Reintegration Grant MIRG-CT-2005-21700. G.A. thanks the conveners of the ATLAS Higgs Working Group for useful discussions and the members of the ATLAS ttH(bb) sub-group in which the ATLAS analysis was conducted.

REFERENCES

[1] The LEP Electroweak Working Group, http://lepewwg.web.cern.ch/LEPEWWG/.

[2] D. Benedetti *et al.*, *J. Phys. G*, **34** (2007) N221.

[3] T. Sjostrand, S. Mrenna and P. Skands, *JHEP*, **0605** (2006) 26.

[4] B. Kersevan and E. Richter-Was, *Comp. Phys. Comm*, **149** (2003) 142 (and updates).

[5] M. Dubinin *et al.*, *INP-MSU*, **41** (1998) 542

[6] S. Frixione and B.R. Webber, *JHEP*, **0206** (2002) 029 (and updates).

[7] G. Corcella *et al.*, *JHEP*, **0101** (2001) 010.

[8] M.L. Mangano *et al.*, *JHEP*, **0307** (2003) 001 (and updates).

[9] S. Corréard *et al.*, *b*-tagging performance, ATL-PHYS-2004-006.

[10] S. Dittmaier, P. Uwer and S. Weinzierl, arXiv:0807.1223 [hep-ph] Preprint, 2008.

[11] A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, arXiv:0807.1248 [hep-ph] Preprint, 2008.

[12] G. L. Bayatian *et al.* [CMS Collaboration], *J. Phys. G*, **34** (2007) 995.