PROCEEDINGS OF SCIENCE

$t\bar{t}$ production at CMS

R.Chierici[*]

CERN, CH-1211 Geneve, Switzerland.

With about 10 millions top-pair event per year at low luminosity, the LHC will effectively be a top-factory. This large amount of data will allow a very precise determination of the properties of the top quark. Studying top-pair total and differential cross-sections will allow important tests on perturbative QCD predictions and will help in the proton PDF constraining. Associated ttH production will be an important complementary discovery channel for the Higgs boson at low masses and will allow a direct measurement of the top Yukawa coupling. Moreover, the top sector is an excellent place where new physics could reveal itself, given the closeness of the top mass to the electroweak scale. The first studies in CMS concerning top-pair events' selection, from trigger setup to full reconstruction analysis, will be presented. A brief review of the measurements connected to top-pair production at the LHC within the Standard Model (SM) and in non-SM scenarios will be given, with emphasis on the ones which have been studied in more details in CMS.

Top2006, Coimbra, Portugal

PoS(TOP2006)021

---

[*] Speaker.

`http://pos.sissa.it/`

1. Introduction

The top quark was discovered about ten years ago, but still so little is known about it. Its large mass gives unique features for the investigation of the electroweak symmetry breaking and for probing physics beyond the Standard Model (SM).

LHC will be the ideal machine for investigating the characteristic of the heaviest quark and its role in the SM; with an NLO production cross-section of about 830 pb, 2 top-pair per second are expected and more than 10 million per year at low luminosity conditions of $2\times10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The importance of studying in detail the properties of top-pair events is three-fold:

- **Precision physics**: from the production and decay mechanisms of the top quark in p-p collisions it is possible to perform precision measures of fundamental quantities like the top quantum numbers, but also to infer measurements of the gluon density function in the proton. The precise determination of the top mass, in particular, has a key role in the constraining of the SM. Indeed the Higgs boson mass, last unmeasured parameter of the SM, is related to the other parameters of the model via loop corrections. Since m_t is still the one with the largest experimental uncertainty, an improvement of its measurement will greatly reduce the allowed range for the Higgs boson mass and eventually discriminate between the SM and the presence of new physics.
- **New physics**: thanks to the enormous value of m_t, compared to other masses in the theory, the coupling of the top to the Higgs is the largest and makes it the perfect place where new physics could manifest itself, both in production and in decay mechanisms.
- **Tool for calibration**: top-pair events typically contain large hadronic activity and always present b-jets in them. These characteristics make them the ideal place for performing calibration of the b and light jets energy scale, by fully reconstructing the event or the W in them, and for the calibration of the b-tagging algorithm.

Ultimately one should not forget that understanding standard physics, and in particular top physics, is essential whan looking for new signals and for any claim of a discovery of new physics. In this paper I will not focus on the use of top events for calibration, which is treated in detail elsewhere [1], but will try to give an overview of how CMS is getting ready to reconstruct and select top-pair events, and the use we intend to make of them.

The first preliminary studies presented in the next pages are based on signal and background events generated with the Toprex [2] and PYTHIA [3] generators, with full detector simulation realized in GEANT 4. The parameter settings for the description of the initial state, hadronization, fragmentation, the minimum bias and underlying event contribution in the generation were taken from [4].

2. Top pair production and trigger

Top pair production at the LHC happens mainly via gluon fusion. Final states result from the decay of two tops (with Branching Ratio (BR) 100% into Wb) and are briefly called fully hadronic, with both Ws decaying into quarks (BR $\approx$ 46%), semileptonic, with one W into leptons and the second one into quarks (BR $\approx$ 44%), and fully leptonic when both Ws decay into leptons (BR $\approx$ 10%). Reconstructing top decays requires therefore an excellent energy flow for the determination of the lepton and jet energies and a good resolution on the missing energy, an excellent lepton identification and the ability to identify b-jets.

The first issue on efficiency at a hadronic machine like the LHC, where collisions are dominated by QCD processes, is to set in place a triggering scheme able to efficiently select as much as possible of a wide range of standard physics final states. The inclusive triggers at CMS include, among others, single isolated electron, muon and di-lepton triggers, many-jet triggers, jet and missing E_{T} trigger and also an isolated tau trigger [5]. If these triggers (especially the leptonic ones) efficiently cover the inclusive production of top-pair events in the leptonic channels, this is not the case for the fully hadronic channel, contaminated by a huge component of QCD.

In this last case the trigger selection is based on the inclusive jets trigger, which considers multi-jets with different E_{T} thresholds depending on the number of jets (up to four). The thresholds are lowered with repect to reference [5] in order to increase the efficiency on the $t\bar{t}$ signal, and a special inclusive fast b-jet tagging is used as well in the trigger scheme. The fast b-jet trigger is based either on a fast pixel track and vertex reconstruction, requiring two tracks with impact parameter significance exceeding 2σ on the two most energetic jets, or on a regional full track reconstruction of the two most energetic jets requiring 3 tracks with impact parameter significance exceeding 2.5σ. The b-jet stream significantly improves the efficiency of the inclusive jets stream for fully hadronic final states up to 15%.

3. Event selection

3.1 Leptonic channel

The selection of events in this channel requires, after the single lepton and di-lepton trigger selection, the presence of just two oppositely charged leptons with $E_{\mathrm{T}} > 20$ GeV and $|\eta| <2.4$ for muons or $|\eta| <2.5$ for electrons.

Furthermore, electrons are required to have a ratio between the energy in the hadronic and electromagnetic calorimeter below 0.05, and the ratio between the energy in the electromagnetic calorimeter and the track momentum in the range $(0.8, 3)$. The reconstruction efficiency is very good for both muons (more than 97%) and electrons (about 90%).

An electron is considered isolated if the total uncorrected E_{T} not associated to the electron itself and in a cone around its track of opening angle $\Delta\mathrm{R} \leq 0.3$, is less than 30% of the lepton E_{T}. Here and in the following ΔR is defined as $\sqrt{\Delta\phi^2 + \Delta\eta^2}$. In a similar way a muon is considered isolated if the sum of the p_{T} of all the tracks present in a cone of $\Delta\mathrm{R} \leq 0.3$ minus the p_{T} of the μ itself, is less than 2 GeV.

Candidate events must have missing $E_{\mathrm{T}} > 40$ GeV. The analysis requires at least two jets with uncorrected $E_{\mathrm{T}} > 20$ GeV detected within $|\eta| <2.5$, where a jet is defined as a fixed-cone cluster with a cone size of $\mathrm{R} < 0.5$. Electrons faking jets produced by electrons are discarded before applying the previous selection by removing those jets which have a large electromagnetic component within $\Delta\mathrm{R} = 0.2$ and the ratio between the electromagnetic energy and the total energy is above 0.75. Using these selection cuts, the efficiency at generator level is about 20 % and a similar value is obtained at reconstruction level.

After removal of cosmic-ray muons by timing cuts and photon-conversion electrons in the electron selection, the dominant backgrounds to di-lepton $t\bar{t}$ events comes from di-boson (WW, WZ, and ZZ) + jets production, top pair events into semileptonic or from tau decays producing leptons, and from Drell-Yan ($Z/\gamma^\star \rightarrow \ell^+\ell^-$) and W$\rightarrow \ell\nu$ + jet production, where a jet is falsely reconstructed as a lepton candidate.

In a $t\bar{t}$ event two genuine jets arise from the hadronisation of b quarks. Thus, b-tagging techniques are used to further suppress backgrounds in which no jets from bs are present. The b-tagging is based on the explicit reconstruction of a secondary vertex in a jet. A variable combining its properties, among which the mass of the charged particles associated to the vertex and the distance between the position of the secondary and primary vertices, is computed for all the jets in an event. The last step in the selection of signal events is then based on the use of the b-tagging variable defined in that way. Candidate events must have the two selected jets as moderately b-tagged. After this selection an efficiency of 5% is obtained, with a very high rejection of all the backgrounds considered at the level of 10^{-3} or below. A S/B value of 5 is obtained, the main background being the one arising from the di-lepton channel in which at least one of the W decays into $\tau\nu_\tau$ and the τ decays leptonically.

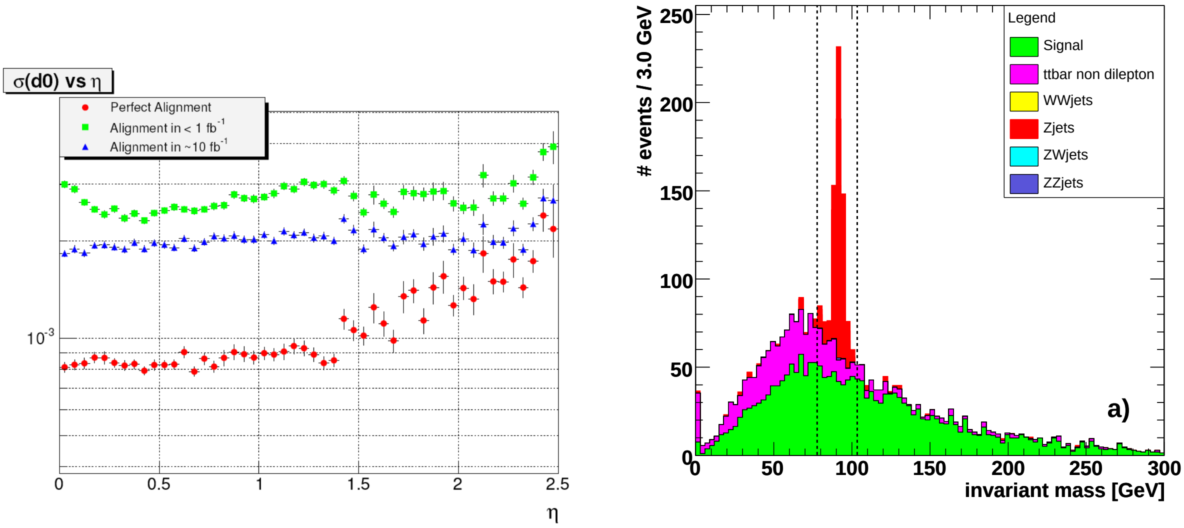

Figure 1: Impact parameter resolution as a function of η for two misalignment scenario compared to the ideal case (left). Di-lepton invariant mass for signal and backgrounds after all the other cuts have been applied and in the case of no b-tagging applied (right).

For the cleanness of the final state, the selection of top pair events in the di-lepton channel has a big importance in the first period of data taking, for a proper commissioning of the detector and the physics tools. The analysis was therefore applied to the detector conditions in the first 1/fb of integrated luminosity, to check what will be the effect on the selection performance.

In the first period of data taking the degradation of the detector, especially for what concerns top pair event reconstruction, will be dominated by misalignment effects, which could affect the b-tagging and the track reconstruction. The degradation should be small as shown, for instance, in figure 1 for the resolution on the impact parameter, since the pixels system is expected to be already aligned at the level of about 10 μm. For the di-lepton channel the ability of selecting top-pair events was checked even against the hypothesis of absence of b-tagging. In this case the major background comes from Z events, which can anyway be removed by cutting on the invariant mass of the two leptons, as shown in figure 1, with the possibility of achieving a signal to background ratio better than 3.

3.2 Semileptonic channel

The selection performance in the semileptonic top-pair channel was studied in detail with the decay chain $t\bar{t} \rightarrow b\bar{b}WW \rightarrow b\bar{b}q\bar{q}\mu\nu_\mu$. The inclusive single muon trigger is applied on the simulated samples. At least one muon is then required within the tracker acceptance of $|\eta| < 2.4$.

The jets in the final state are reconstructed with the iterative cone algorithm using an opening angle of ΔR=0.5. Energy depositions associated with the identified lepton are removed for jet clustering. Seeds for the cones were selected from all calorimetric towers with energy above a threshold, function of their pseudo-rapidity in order to reduce the contribution from the underlying event.

A simple pre-selection was applied on the event requiring at least four jets with pseudo-rapidities in the range of the tracker, $|\eta| < 2.4$, and a raw E_T above 10 GeV. The jets must have a flight direction inside the tracker acceptance to allow for a proper performance of the b-tagging algorithm. In order to discriminate between jets originating from the heavy b-quarks compared to the light quarks, a b-tag probability was constructed using a likelihood ratio technique from the b-tag discriminant variable. Both the discriminant, applied to different jet types, and the purity-efficiency curve are shown in figure 2.

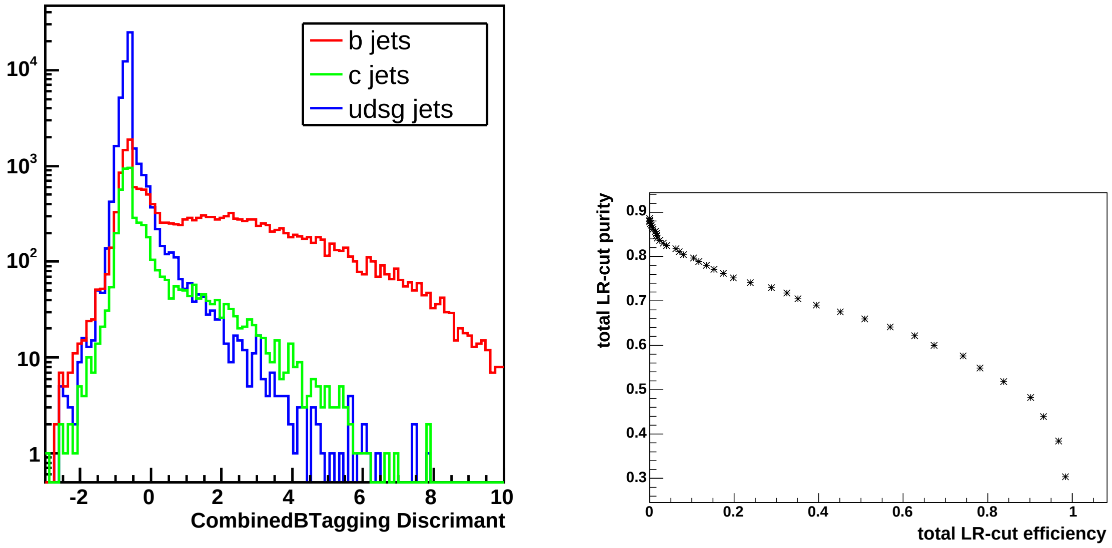

Figure 2: b-tagging discriminant for different jet flavours (left, arbitrary units) and resulting purity-efficiency plot after a cut on a discriminant variable based on it (right).

The event is required to have exactly four jets with p_{T} above 30 GeV/c. Of these jets, exactly two must have a b-tag probability larger than 60%, the remaining two jets must have a probability value of less than 30% . The two anti-b-tagged jets are assigned to the W boson decay, resulting in an efficiency of 80% for choosing the correct jet combination. It is also required that the cones of these four jets do not overlap in (η,ϕ) by requiring $\Delta \mathrm{R}_{jet-jet} > 1.0$. The reconstructed hard lepton has to have a transverse momentum p_{T} exceeding 20 GeV/c. Of the two possible b-jet/top-quark associations, at least one should yield a mass of the hadronically decaying top quark of less than 350 GeV/c^2 for the event to be selected. Applying all cuts a signal-to-noise ratio of 3.8 is obtained. Due to the limited size of the W+jets background simulation samples, one has to take into account a significant uncertainty in the estimate of the signal-to-noise ratio of about 0.5. Most of the remaining background events are $\mathrm{t\bar{t}}$ events with a μ from t→W→ $\tau \rightarrow \mu$ in the final state. These events have, for the hadronic part, the same properties of the signal events, and should therefore not be considered as a real background for the top quark mass measurements. A more refined event selection was studied to improve even more the purity of the selected sample. This can be achieved by a mass cuts after the full kinematical reconstruction of the event, where the jet combination is assigned via a likelihood ratio technique which takes into account the reconstructed masses in the event and the b-tagging probabilities.

3.3 Fully hadronic channel

The fully hadronic final state, characterized by the nominal six-jets topology from the top hadronic decays $\mathrm{t\bar{t}} \rightarrow WW\mathrm{b\bar{b}} \rightarrow \mathrm{q\bar{q}q\bar{q}b\bar{b}}$, has the largest branching fraction and kinematics that can be fully reconstructed. However, this channel is affected by a large background from QCD multi-jet production, which makes the isolation of the signal rather challenging.

As described in section 2 the trigger requires either a multi-jet condition or a b-tagged jet among the two highest-E_{T} jets. After the trigger selection the QCD rate is reduced to 20 Hz, the signal efficiency is 16.7% and the signal to background ratio, amounts to 1/160.

The $\mathrm{t\bar{t}}$ fully hadronic efficiency (normalized to the trigger efficiency) and the QCD rate are shown in figure 3 as a function of the jet transverse energy for different values of the minimum number of jets considered.

The first step of the selection requires a topology of $6 \leq \mathrm{N_{jet}} \leq 8$, consistent with the considered process, and taking into account possible additional jets from final state radiation. For a jet to be counted, it must be in the central region of the detector ($|\eta| <$2.4) and its transverse energy must be greater than 30 GeV. To improve the signal to background ratio, cuts on the following distributions are taken into account:

- centrality: fraction of the hard scatter energy going in the transverse plane $\sum E_{\mathrm{T}}/\sqrt{\hat{s}} \geq 0.72$. Here, $\hat{s} = (\sum E)^2 - (\sum P_{\mathrm{z}})^2$, where the sums here and in the following run over the reconstructed jets.
- non-leading jet total transverse energy obtained removing the two most energetic jets: $\sum E_{\mathrm{T}}$-E_{T} (1) -E_{T} (2) ≥ 164 GeV.
- $\frac{3}{2}Q_1 \geq 0.024$ where Q_1 is smallest of the three normalized eigenvalues of the sphericity tensor $M_{ab} = \sum_j P_{ja}P_{jb}$ called aplanarity.

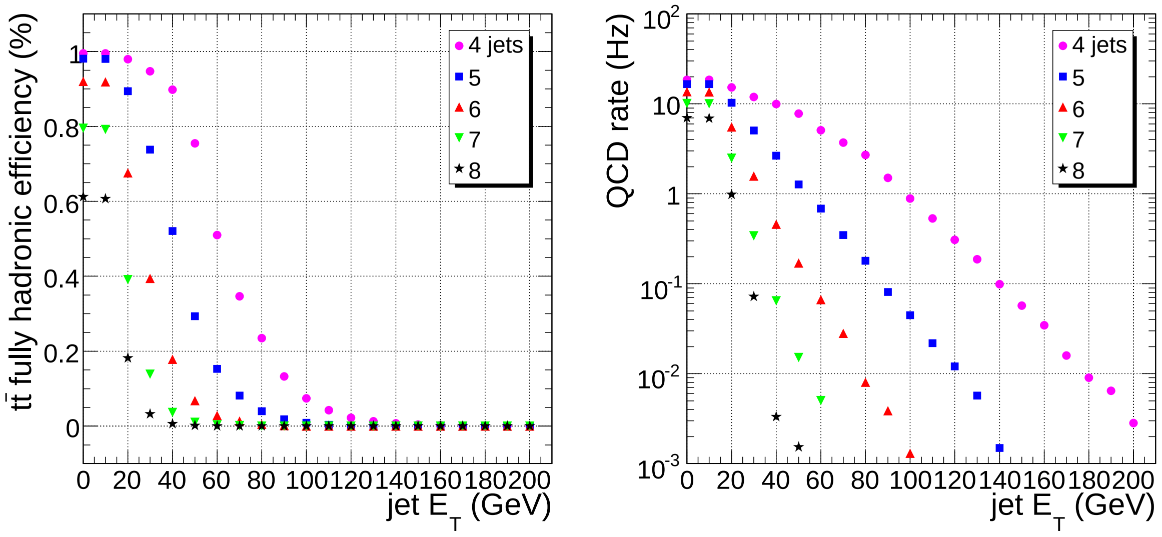

Figure 3: The $t\bar{t}$ fully hadronic efficiency (left) and the QCD rate (right) as a function of jet transverse energy for different values of the minimum number of jets considered, after the trigger selection.

After the kinematic selection a b-tagging is applied to the surviving samples of $t\bar{t}$ all-hadronic and QCD events. Selection criteria of at least one b-jet and of two b-jets are considered, for which a final signal to background ratio is 1/8 (1/4) respectively for one (two) b-tag samples, for a signal efficiency of 3% (1.9%) relative to the fully-hadronic $t\bar{t}$ sample.

4. Systematics

At the LHC the measurement of total and differential top cross-sections will soon be dominated by systematic uncertainties, the precise determination of which is currently under study. Nonetheless, it is interesting to briefly mention the main categories with a prospect for the expected contribution of the corresponding error.

- Luminosity: the error on the luminosity determination will enter in any absolute cross-section measurement. The luminosity is determined by the number of interactions per bunch crossing and by measuring the total p-p elastic cross-section and current estimates believe that a relative error of about 5% should be very well in reach at the LHC.
- Theoretical sources: they reflect our poor understanding of the reality of a p-p collision and the imperfect way in which it is implemented in our Monte Carlo (MC) simulation. It is difficult to assess today the magnitude of the systematics effects due to modelling, but the sources which are expected to be particularly relevant are the description of the quarks and gluons p.d.f.s in the proton, the underlying event and the minimum bias, and the description of the radiation and fragmentation. Temporary recipes, based on MC studies, for determining the size of these uncertainties in the analyses are provided in [4] and will be adopted to quantify these errors in the absence of data. When the data will be avilable many more

constraints and MC tuning, especially on QCD data, will help in reduce quite significantly these uncertainties.

- Experimental sources: they come from the uncertainties on the measured variables on which the analyses are performed. Particularly relevant are the knowledge of the jet and lepton energy scale and resolution, and the level of knowledge of the b-tagging efficiency and fake rate. Preliminary studies [1] show that the jet energy calibration using external samples like Z(γ)+jet or the W mass constraint in $t\bar{t}$ events can bring the uncertainty on the jet energy scale to a level of a few percent and that the use of top events can also assure an internal calibration of the b-tagging at the level of 4-5%.

All these estimations still have to be translated into proper errors on the measured cross-sections.

5. The importance of top-pair production

Top-pair production, in terms of rates, total and differential cross-sections, can be an invaluable tool for performing precision physics at the LHC or constraining new physics. Among the most important standard physics measurements at the LHC there are the top mass and the top quantum numbers (spin and charge), which are treated in detail in other contributions to this conference. Moreover, top pair production can be used also for indirect measurements of great interest:

- $|V_{tb}|$, via the ratio of doubly b-tagged to singly b-tagged events, directly related to the CKM matrix element when the SM is assumed. This measurement can nicely complement the measure of $|V_{tb}|$ from single top production [6].
- constraint of the gluon p.d.f. in the proton, by using the angular differential distributions of the decay products, directly related to the fraction of momentum taken by the originary partons [7].
- measurement of top-pair production in association with Z or γ, that would allow the first direct information of top couplings to neutral bosons.

There is also an high interest in searching for signals of new physics in the top sector, and they can manifest themselves directly in the top-pair production and decay mechanisms:

- Production: there are many theories (Minimal Supersymmetric SM (MSSM), technicolour, Strong ElectroWeak Symmetry Breaking (SEWSB)) predicting resonances in the tt mass which can be investigated by a differential cross-section measurement. Moreover, in the MSSM, gluino decays can bring to a production of events with two tops in the final state which, for R-parity conserving scenarios, will provide unique signatures with high missing E_T.
- Decay: change in the top decay BR can happen in many MSSM scenarios, for instance via decays to charged Higgs bosons $t\to H^+ b\to csb$, $\tau\nu b$, or via flavour changing neutral currents decays $t\to Zq$, γq, gq, that can be enhanced by factors up to 10^4 in the MSSM and in theories like SEWSB, making them visible at the LHC. These productions can be probed by measuring the decay BR from top-pair production.

- Loop corrections: non-SM loop corrections are anticipated to be important, introducing relative variations up to 10% or more in certain regions of the MSSM parameter space. In the absence of direct signals of supersymmetry, the presence of new physics can therefore also be inferred by a precise measurement of the cross-section.
- Associated production: top-pair production can be accompanied by SM/MSSM Higgs at relatively low mass scales thans to the large top Yukawa coupling, making it also an important discovery channel for standard Higgs bosons of masses up to 130 GeV.

As an example I will discuss the associated ttH production investigated in CMS. This is a very challenging channel [8] because of the high number of jets in the final state which makes the reconstruction difficult and because of the extremely high cross-section of the main backgrounds, represented by top-pair production in association with other jets. Events are selected if there is an isolated electron or muon with p_T above 10 GeV$/c$ in the tracking acceptance and at least six jets in the event with E_T larger than 20 GeV and with $|\eta| < 2.5$. Four out of these six jets are requested to be b-tagged and the jet assignment to the top or Higgs decays is performed via an event likelihood function which also account for the b-tagging. Figure 4 shows the encouraging result for the mass reconstruction where, under the peak, for an efficiency between 1 and 2% for the signal, a value of S/B of 0.6 can be reached for a value of the input Higgs boson mass of 115 GeV$/c^2$ and for 30/fb of integrated luminosity. The result can also be translated into limits in the MSSM plane for a light supersymmetric Higgs boson, properly rescaling the expected cross-sections. The resulting limits in the (tanβ,m$_A$) plane are shown in figure 4. The discovery fails at low and high values of masses because the expected cross-section is very much reduced.

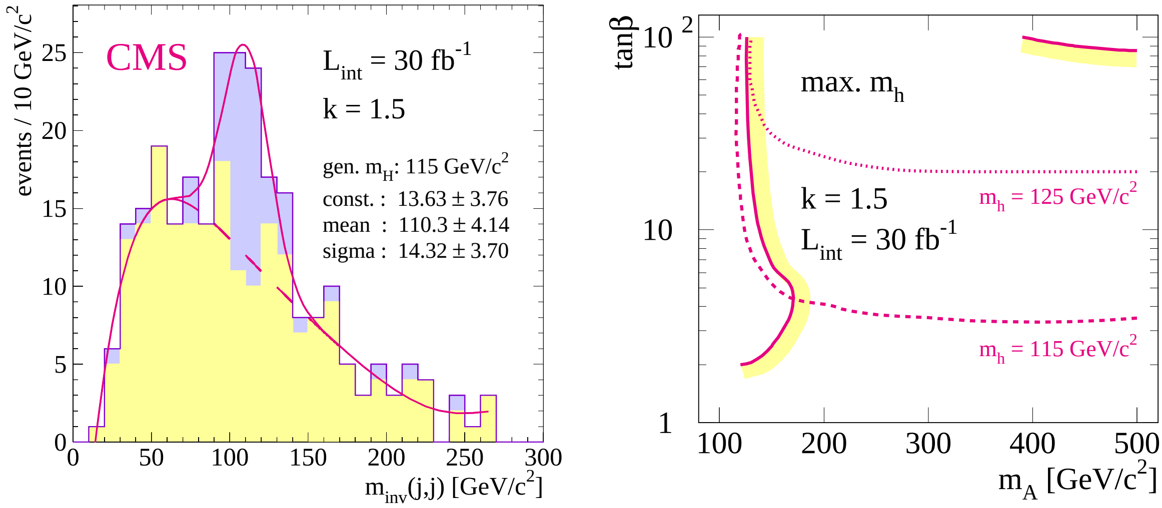

Figure 4: Reconstructed invariant mass in the ttH analysis for signal and backgrounds and m$_H$=115 GeV$/c$ 2 (left) and the 5 σ discovery contours in the MSSM plane (right). The dotted and dashed lines are isomasses for m$_h$=125 GeV$/c^2$ and m$_h$=115 GeV$/c^2$, respectively. Both plots are for an integrated luminosity of 30/fb.

It should be noted, however, that this study is very preliminary and should be repeated with the description of backgrounds with more realistic generators including also leading higher orders contributions.

6. Conclusions and outlook

Top physics at the LHC will be extremely challenging and fascinating. The study of top-pair production will certainly allow precision measurements that will drastically improve our current knowledge about the SM, will represent an invaluable tool for understanding the detector and calibrating the reconstruction, and will likely be the window where new physics should manifest itself.

CMS is getting ready for this kind of physics by tuning the trigger selections, by studying and refining the potential analyses and by starting to estimate the systematic effects on the measurements and to understand what is the best use of the data to constrain them. Particular effort is also put into the first data taking scenario, when it will be important to have ready analyses able to select top-pair events for a fast feedback on the detector operation and on possible spectacular evidence for new physics.

References

[1] J. Heyninck, these Proceedings.

[2] S.R. Slabospitsky, Comput. Phys. Commun. 148 (2002) 87.

[3] T. Sjöstrand et al., Comput. Phys. Commun. 135 (2001) 238.

[4] P. Bartalini, R. Chierici, A. deRoeck, *Guidelines for the estimation of theoretical uncertainties at the LHC*, CMS NOTE 2005/013.

[5] CMS Collaboration, *DAQ Tecnical Design Report*, CERN/LHCC/2002-26.

[6] A. Giammanco, these Proceedings

[7] ATLAS Collaboration, *ATLAS detector and physics performance*, Vol. II, CERN/LHCC/99-15.

[8] D. Denegri, V. Drollinger, T. Müller, *Searching for Higgs bosons in association with top quark pairs in the H→bb decay mode*, CMS NOTE 2001/054.