Search for the neutral Supersymmetric Higgs boson with the CMS experiment at LHC

Antonio Branca for the CMS Collaboration

University of Padova - Department of Physics and Astronomy and INFN, Via Marzolo 8, 35131 Padova - Italy

E-mail: antonio.branca@pd.infn.it

Abstract. The search of neutral Higgs boson in two different decay channels, a pair of τ leptons and a pair of b-quarks, is presented. The search is performed in the context of the Minimal Supersymmetric Standard Model (MSSM), using data collected by the CMS experiment at the LHC during 2011-12 at a center-of-mass energy of 7-8 TeV. In the τ decay four independent pair final states, where one or both τs decay leptonically, are studied. In the $b\bar{b}$ channel we look for final states with two high energetic jets plus a third jet, exploiting the production in association with b-quarks to enhance the signal purity in the selections of events. A description of the analyses strategies adopted for the two channels is provided. The results are presented and interpreted in the m_h^{max} scenario of the MSSM framework.

1. Instability of the Standard Model Higgs mass

The Standard Model (SM) Higgs boson suffers a theoretical problem, known as the *Hierarchy problem*, causing instability of its mass [1]. The Higgs boson mass receives enormous quantum corrections from the virtual effects of every particle that couples, directly or indirectly, to the Higgs field; drawing the value of the mass away from the experimental results. Direct searches at LEP e^+e^- and Tevatron $p\bar{p}$ colliders have led, respectively, to a lower-mass bound of $m_h > 114.4\ GeV$ [2], and to an exclusion in the range $162 - 166\ GeV$ at 95% of confidence level (CL) [3]. Indirect constraints from precision measurements favour the mass range $m_h < 158\ GeV$ at 95% of CL [4]. More recently, searches performed at the LHC [5] put more stringent limits on the SM Higgs particle mass [6] and eventually a boson with the signatures compatible with a SM Higgs boson has been observed with a mass around 125 GeV [7, 8].
The Minimal Supersymmetric Standard Model (MSSM) [9] provide a solution to the hierarchy problem. Because of the new symmetry between fermions and bosons, the supersymmetry, the number of particles is doubled in the MSSM with respect to the SM. Also the new particles gives enormous quantum corrections to the Higgs mass, but they have sign such that cancel the SM divergencies.

2. MSSM Higgs boson sector

The Higgs boson sector of the MSSM is more complex than the SM one. Electroweak (EWK) anomaly cancellation requires two doublets of complex scalars, corresponding to eight degree of freedom. These turn into five Higgs boson after EWK symmetry breaking through the Higgs

mechanism: 2 neutral CP even, h and H, 1 neutral CP odd, A, 2 charged, $H^{\pm}$. Their masses depend only by two parameters of the MSSM model, the A mass M_A and $tan\beta$, at tree-level. Important radiative corrections, depending on other MSSM parameters, affect the Higgs masses. In the experimental searches these parameters are fixed defining the so called benchmark scenarios. The benchmark scenario adopted for the searches presented here is the m_h^{max} [10].
It can be shown that if M_A is not too large and $tan\beta$ is large, the couplings of the neutral Higgs bosons to down-type fermions are enhanced. Grand unification theories (GUTs) prefer large values of $tan\beta$. Moreover, two out of the three neutral Higgs boson states are degenerate in mass for large values of $tan\beta$. This contributes to enhance a possible experimental signal. Thus it is experimentally important to search for signals produced by channels where the Higgs boson decays into a τ-leptons or b-quarks pair.

3. MSSM neutral Higgs boson production at LHC and branching ratios

The main production mechanisms of the neutral Higgs bosons at LHC are the gluonn-gluon fusion and the production in association with b-quarks. The cross section for these mechanisms is shows in Figure 1 for pp collisions at a center of mass energy of 7 TeV. The distributions

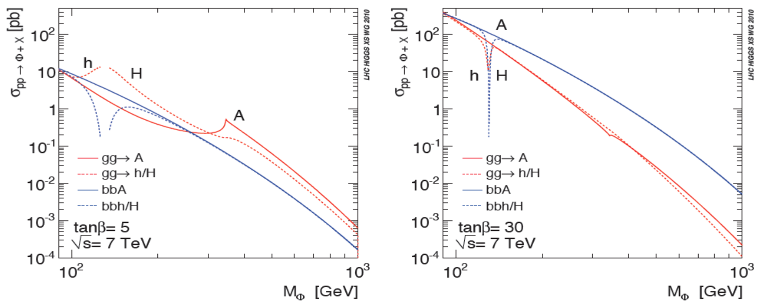
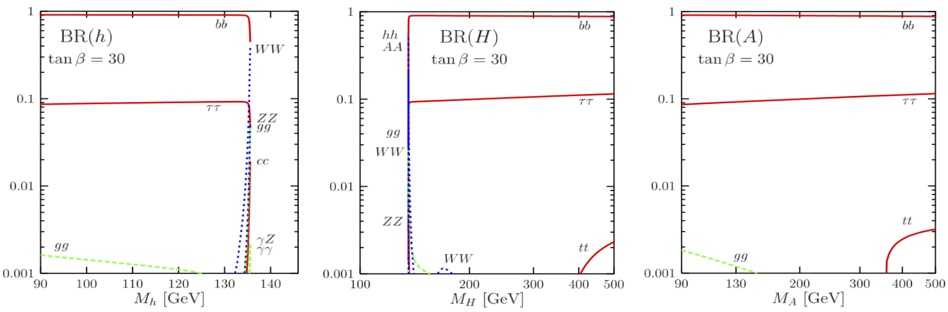


Figure 1. Neutral MSSM Higgs production cross sections at the LHC at the center of mass energy of 7 TeV for gluon fusion and the associated production with bottom quarks, including QCD corrections. The values are shown as a function of the h, H, A masses and for $tan\beta = 5$ (left) and $tan\beta = 30$ (right). The m_h^{max} scenario is considered.

show that the production in association with b-quarks is dominant for large values of $tan\beta$. Specifically, cross sections between $O(200) - O(0.01)$ pb are obtained at $tan\beta = 30$. Figure 2 shows the branching ratios of the neutral Higgs bosons into SM particles for $tan\beta = 30$. The branching ratio values are around 90% for the b-quarks pair decay and around 10% for the τ pair decay.
The decay channels in τ-leptons and b-quarks pairs of the MSSM Higgs boson have been explored by the analyses performed by the Collaboration of the CMS experiment [11] at the LHC.

4. $\tau\tau$ channel search

In this analysis [12] the channel where a neutral Higgs boson decays into a τ pair, producing two energetic opposite charge leptons in the final state, is considered. The analyzed data sample is made of 4.9 fb^{-1} and 12.1 fb^{-1} at pp center of mass energy of 7 TeV and 8 TeV, respectively, recorded by the CMS experiment. Specific High Level Trigger (HLT) paths, requiring a combination of electron, muon and tau particles, are implemented to select the final states from the LHC pp collisions compatible with the expected final state. Four final states of the $\tau\tau$

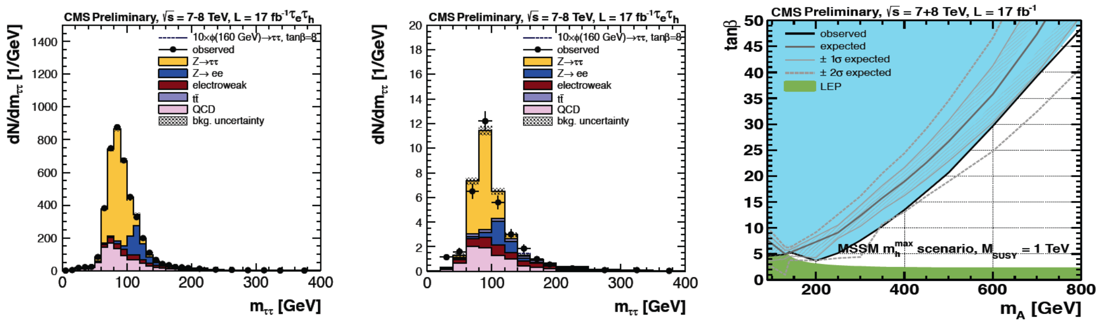


(a) h branching ratios. (b) H branching ratios. (c) A branching ratios.

Figure 2. The decay branching ratios of the neutral Higgs bosons as a function of their masses for $tan\beta = 30$. The values are obtained in the m_h^{max} scenario.

pair are considered for the analysis: $e\tau_h$, $\mu\tau_h$, $e\mu$ and $\mu\mu$. Here τ_h stands for the hadronic decay of the τ lepton. The recorded data sample is further reduced by selections requiring the presence of two well reconstructed high p_T opposite charge leptons, matching one of the four final states, in the event. The data sample is split into two categories: a sample where at least one jet is identified as coming from b-quark (*b-tagged* jet) with transverse momentum of $p_T > 20\ GeV$ and no more than one jet with $p_T > 30\ GeV$ is present (*B-Tag* category); a sample where are not present b-tagged jets with $p_T > 20\ GeV$ (*No B-Tag* category). This separation is done in order to enhance the sensitivity to the production in association with b-quarks (*B-Tag* category) and to the production through gluon-gluon fusion (*No B-Tag* category).
The largest background comes from $Z \to \tau\tau$ events, which is irreducible, followed by the reducible QCD multi-jets. Other reducible backgrounds come from EWK processes: $W + jets$, $t\bar{t}$, WW, ZZ, WZ and $Z \to e^+e^-/\mu^+\mu^-$. The reduction of the reducible backgrounds is achieved by different cuts. The QCD multi-jet is reduced by requiring low charged/neutral particle activity around the selected leptons (lepton isolation). For $e\tau_h$ and $\mu\tau_h$ the $W + jets$ is reduced requiring the transverse mass between the p_T of the electron or muon and the missing transverse energy to be less than 40 GeV. For $e\mu$ and $\mu\mu$ a topological requirement is applied to reduce the $W + jets$ and $t\bar{t}$ backgrounds. In the τ decay the visible (from the detected particles) and the invisible (from neutrinos) momenta have about the same direction. This is not true, in average, for fake τs. This observation leads to the construction of a topological variable sensible to the visible and invisible momenta directions in the $\tau\tau$ final states. This is used as discrimination variable for signal and background events. To reduce $Z \to e^+e^-/\mu^+\mu^-$, events with more than one electron or muon with $p_T > 25\ GeV$ are excluded for the $e\tau_h$ and $\mu\tau_h$ final states; for $e\tau_h$ events an additional requirement on transverse missing energy is performed. Different techniques are exploited to estimate the backgrounds that pass all the selections. The irreducible $Z \to \tau\tau$ background is estimated in shape using a $Z \to \mu\mu$ data sample, where the muon is interpreted as tau lepton (*embedding* technique). This way, a better estimation of the event migrations between the two data sample categories is achieved. Also the normalization is obtained from $Z \to \mu\mu$ events. For the QCD multi-jet in the $e\mu$ final state, the shape is obtained from same-sign events, whereas the normalization is given by $e\mu$ events with fake-electron selections weighted by the efficiency of jet misidentification as electron (ϵ_{fake}). For the $e\tau_h$ and $\mu\tau_h$ final states, QCD multi-jet background come from random combination of a reconstructed electron or muon and a τ. Thus the ratio of same-sign and opposite-sign events is close to one. Same-sign events give the background prediction for both shape and normalization. Monte Carlo (MC) simulated W events are used to predict the shape for the $W + jets$ background. The normalization is given by a W enriched control sample from data.

MC simulation is also used for $t\bar{t}$ shape prediction. The measured cross section for $t\bar{t}$ events is used for the normalization. Finally, the $Z \to e^+e^-/\mu^+\mu^-$ backgrounds are predicted using data, where the fraction of leptons misidentified as τ_h is measured from the same events exploiting the Tag and Probe technique [13].
The signal extraction is performed by a binned maximum likelihood fit to the full reconstructed invariant mass of the $\tau\tau$ final states. A $\tau\tau$ invariant mass resolution of about 20% is achieved. In Figure 3 is shown the $\tau\tau$ invariant mass for the $e\tau_h$ final state. Since no evidence of a Higgs



(a) No B-tag sample fit. (b) B-Tag sample fit. (c) 95% CL limits.

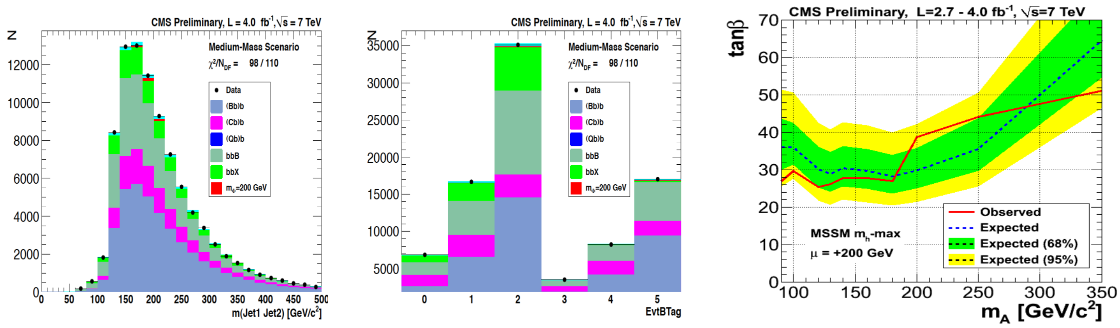
Figure 3. (a-b) $e\tau_h$ invariant mass for the *No B-Tag* and *B-Tag* categories. The data (black dots) are fitted with the background predictions (full histograms) within the systematic uncertainties (grey dashed band). (c) 95% CL exclusion contours in the $(M_A, tan\beta)$ parameter space of the MSSM model in the m_h^{max} benchmark scenario. The black line is the observed limit in data. The grey line is the median expected limit for the background only hypothesis. The dashed grey band (dotted grey line band) is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median. The [CELESTE] area is the excluded area corresponding to the observed limit. The green area is the excluded area by the LEP experiment.

boson signal is found, 95% CL limits on the cross section times the branching ratio in τ pair are calculated with the CL_s method [14]. These limits are converted in exclusion plots in the $(M_A, tan\beta)$ parameter space, shown in Figure 3. The analysis excludes values of $tan\beta$ as low as 4 at $M_A = 200\ GeV$. The limits calculation take into account the statistics and systematic uncertainties, that contribute to worsen the sensitivity of the analysis. Different systematic uncertainties have been identified. They affect the event yield, the largest contribution coming from b-tagging efficiency (10%) and background normalization ($10-30\%$), and the $\tau\tau$ mass shape, where contributions come from τ_h (33%), muon (1%) and electron ($1-2.5\%$) energy scale and resolution. Theoretical systematic uncertainties range from 20% to 25%.

5. $b\bar{b}$ channel search: all hadronic

In this analysis [15] the search is performed in the channel where the Higgs boson is produced in association with b-quarks and decays into a b-quarks pair, producing up to four jets in the final state. In average, the two jets from the Higgs decay are high p_T jets, while the associated jets are softer. The analyzed data sample is made of a total of 6.7 fb^{-1} at pp center of mass energy of 7 TeV, selected with specific HLT paths requiring 2 or 3 jets, with 2 of them b-tagged. The final analyzed data sample is filtered requiring 3 well reconstructed high p_T b-tagged jets in the tracker region of the CMS detector. The final state contains only jets, and it is said all hadronic. The analysis is performed in two different Higgs mass regions: *low Higgs mass* ($M_\Phi < 180\ GeV$), where data are collected with HLT paths with low p_T thresholds, and

high Higgs mass ($M_\Phi \geq 180\ GeV$), where data are collected with HLT paths with higher p_T thresholds.
The dominant background comes from heavy-flavour multi-jet QCD, made of 3 real b-jets or 2 real b-jets plus 1 jet originating from a light parton or c-quark. Other backgrounds are EWK $t\bar{t}$ and $Z + jets$, but they are negligible as shown by MC studies.
The Higgs boson search is performed in the two dimensional distribution of the two leading b-tagged jets invariant mass (M_{12}), and an event b-tagging variable condensing the b-jet properties of the 3 b-tagged jets in the event. Data driven background templates are built for the background prediction. Since about the 98% of the events in the data sample of 3 b-tagged jets is made of two real b-jets (from MC studies), the data sample used for the background modeling is made of 2 b-tagged jets (b) plus 1 jet (x) where the b-tagging requirement is not applied (*untagged* jet). Three categories are defined depending on the rank of the x jet p_T in this sample: bbx, bxb and xbb. The untagged jet x can originate from partons with three possible flavours: light ($u\ d\ s\ g$), charm (c) and beauty (b). Thus, a total of 9 templates have to be built for the background modeling. These are reduced to 5, since there are templates with similar shapes. The templates are obtained by weighting the x jet with the probability for the jet to be b-tagged (from MC simulation), to model the M_{12} variable, and the probability to have different event b-tagging variable values (from $t\bar{t}$ MC simulated events). The 2 b-tagged data sample used for background modeling contains $2-3\%$ of the events where one or both the b-tagged jets are not originated by b-quarks (from MC simulation). This can have slight effects on the M_{12} shape of the templates. The contribution is evaluated from non-bb events from mistag rate measurements [16]. The effect is small and is subtracted from the three categories.
The signal extraction is done by performing a binned least-squares fit in the M_{12} and event b-tagging variable space of the 3 b-tagged data sample. The linear combination of the background templates and signal template, from MC simulation, is used for the fit. Only the background template shapes are predicted by the data-driven method, the normalization is obtained from the fit. Figure 4 shows the projection of the fit in the M_{12} and event b-tagging variables. Since



(a) M_{12} fit projection.

(b) Event b-tagging variable fit projection.

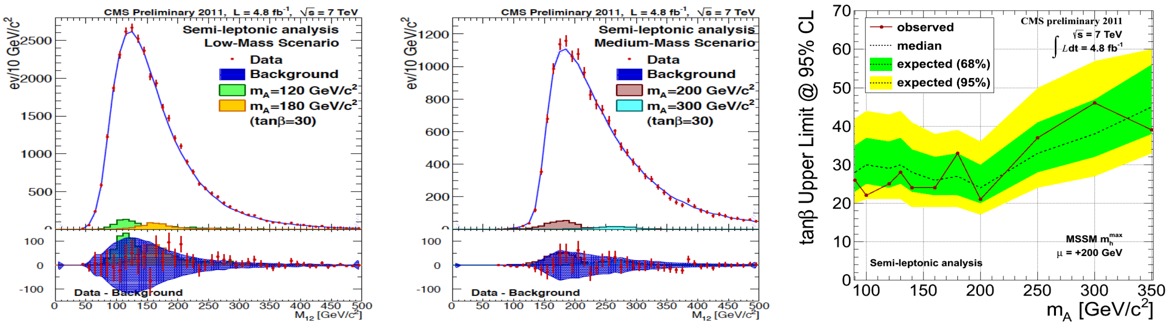
(c) 95% CL limits.

Figure 4. (a-b) Results of the fit in data (black dots) using the background and signal ($M_\Phi = 200\ GeV$) templates (full distributions). The projection of the fit in each variable is shown in the plots. (c) 95% CL exclusion limits in the ($M_A, tan\beta$) parameter space of the MSSM model in the m_h^{max} benchmark scenario. The red line is the observed limit in data. The blue dotted line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

no evidence of a Higgs boson signal is found, 95% CL limits on the cross section times the branching ratio in b pair are calculated with the CL_s method. These limits are converted in exclusion plots in the $(M_A, tan\beta)$ parameter space, shown in Figure 4. The analysis excludes values of $tan\beta$ as low as 25 at $M_A = 120\ GeV$. The limits calculation take into account the statistics and systematic uncertainties, that contribute to worsen the sensitivity of the analysis. Different systematic uncertainties have been identified. They affect the signal yield and shape, the largest contribution coming from b-tagging efficiency $(10 - 13\%)$, and signal yield only, the largest contribution coming from the online b-tagging efficiency (32%). Theoretical systematic uncertainties range from 3% to 28%.

6. $b\bar{b}$ channel search: semi leptonic

This analysis [17] exploits the same channel used in the all hadronic analysis to perform the Higgs boson search. A data sample of a total of 4.7 fb^{-1} at pp center of mass energy of 7 TeV are analyzed, selected by HLT paths requiring one muon, one or two jets with one or two of them b-tagged. The final analyzed data sample is filtered requiring 3 well reconstructed high p_T b-tagged jets in the tracker region of the CMS detector and a muon inside one of the two leading jets. The semi leptonic decay of the b-quark is exploited in this analysis.
The semi leptonic analysis is affected by the same backgrounds of the all hadronic analysis, with the QCD multi-jet the dominant source and the $t\bar{t}$ and $Z + jets$ negligible.
The Higgs boson search is performed in the invariant mass of the two leading b-tagged jets M_{12}. Two data-driven techniques are used to estimate shape and normalization of the QCD background: *B-Tagging matrix* and *nearest-neighbour* methods. The *B-Tagging matrix* uses a data sample of 2 b-tagged jets plus a third jet where the b-tagging requirement is not applied (bbj sample). The prediction for the distribution of the relevant kinematic variables in the 3 b-tagged jets sample (bbb sample), is obtained by weighting each bbj event with the probability for the third jet j to be b-tagged. This probability is built with the efficiencies to b-tag a jet originating from a x-flavour parton (from MC simulation) and the fractions of jets originated by that x-flavour parton (from data). To avoid a bias in the measure of the fraction of jets originated by b-quarks due to a possible Higgs boson signal, control and signal regions are defined from a discriminator variable built using QCD and signal simulated events. The discriminator is defined for two different Higgs mass regions: *low Higgs mass* ($M_\Phi < 200\ GeV$) and *high Higgs mass* ($M_\Phi \geq 200\ GeV$). The *nearest-neighbour* method use a data samle of 1 b-tagged jet plus a second and a third jets where the b-tagging requirement is not applied (bjj sample). A sample of similar events to those in the signal region are selected from a large training sample obtained in the control region. The probability for a bjj event to have 3 b-tagged jets is obtained from the ration of bbb and bjj events in the sample of similar events. This probability is used to predict the relevant kinematic distributions in the bbb sample. The two data-driven methods are independent, thus the two background templates are combined.
The Higgs boson signal search is done performing a binned likelihood fit of the M_{12} variable in the 3 b-tagged data sample, shown in Figure 5. The combined background template from the two data-driven methods and the signal templates from MC simulation are used in the fit. Since no evidence of a Higgs boson signal is found, 95% CL limits on the cross section times the branching ratio in b pair are calculated with the CL_s method. These limits are converted in exclusion plots in the $(M_A, tan\beta)$ parameter space, shown in Figure 5. The analysis excludes values of $tan\beta$ as low as 20 at $M_A = 200\ GeV$. The limits calculation take into account the statistics and systematic uncertainties, that contribute to worsen the sensitivity of the analysis. Different systematic uncertainties have been identified. They affect the signal yield, the largest contribution coming from b-tagging efficiency (12%), and the background normalization (5%). Theoretical systematic uncertainties range from 3% to 28%.



(a) Low Higgs mass region. (b) High Higgs mass region. (c) 95% CL limits.

Figure 5. (a-b) Results of the fit in data (red dots) using the combined background from the two data-driven methods (blue line) and signal templates (full distributions) for the low and high Higgs mass regions. In the lower section of the distributions is shown the ratio of data and background prediction (red dots) compared to the uncertainties (blue band). (c) 95% CL exclusion limits in the $(M_A, tan\beta)$ parameter space of the MSSM model in the m_h^{max} benchmark scenario. The red line is the observed limit in data. The black dotted line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median.

7. All hadronic and semi leptonic combination

The same Higgs boson production mechanism and decay channels are exploited by the all hadronic and semi leptonic analyses. The sensitivities reached by the two analyses are similar and a combination of the results is therefore performed to gain sensitivity [18]. All the systematic uncertainties have been considered for the calculation. The background systematic uncertainties are all uncorrelated, the common signal systematic uncertainties are 100% correlated and the analyses specific signal systematic uncertainties are uncorrelated. The small overlap of about 2.5% between the data samples of the two analyses has been removed for the combination. Figure 6 shows the result of the combined limits. As it can be seen the overall sensitivity is

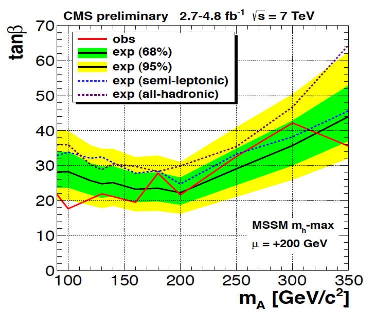


Figure 6. 95% CL combined exclusion limits in the $(M_A, tan\beta)$ parameter space of the MSSM model in the m_h^{max} benchmark scenario. The red line is the observed limit in data. The black line is the median expected limit for the background only hypothesis. The green (yellow) band is the range that is expected to contain 68% (95%) of all the observed limit excursions from the median. The limits are compared to the expected limits from the all hadronic (magenta dotted line) and semi leptonic (blue dotted line) analyses.

improved with respect to the sensitivity of the single analyses.

8. Conclusions

The search for a neutral MSSM Higgs boson with the CMS experiment at the LHC has been reported for two decay channels: $\tau\tau$ and $b\bar{b}$. The data show no significant excess with respect to the expected background. Limits at 95% of CL are then calculated and interpreted as exclusion limits on $tan\beta$ as a function of M_A in the m_h^{max} of the MSSM theory. In the $\tau\tau$ final state the analysis excluded $tan\beta$ as low as 4 for $M_A = 200\ GeV$. In the $b\bar{b}$ final state the analyses excluded $tan\beta \sim 20-40$ in the full M_A range up to $M_A = 350\ GeV$. These results extend the area in the $(M_A, tan\beta)$ parameter space excluded by the Tevatron experiments [19, 20, 21]. Both the $\tau\tau$ and $b\bar{b}$ analyses are important since they test couplings to different particles of the Higgs boson.

References

[1] Steven Weinberg 1976 Implications of dynamical symmetry breaking *Phys. Rev.* D **13** pp. 974-996.
[2] R. Barate et al. 2003 Search for the standard model Higgs boson at LEP *Phys. Lett.* B **565** pp. 61-75.
[3] T. Aaltonen et al. 2010 Combination of Tevatron searches for the standard model Higgs boson in the W+W- decay mode *Phys. Rev. Lett.* **104** p. 061802.
[4] S. Schael et al. 2006 Precision electroweak measurements on the Z resonance *Phys. Rept.* **427** pp. 257-454.
[5] (ed.) Evans Lyndon and (ed.) Bryant Philip 2008 LHC Machine *JINST* **3** S08001.
[6] CMS Collaboration 2012 Combination of SM, SM4, FP Higgs boson searches *CMS-PAS-HIG-12-008*.
[7] CMS Collaboration 2012 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC *Phys. Lett.* B **716** pp. 30-61.
[8] ATLAS Collaboration 2012 Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC *Phys. Lett.* B **716** pp. 1-29.
[9] I. J. R. Aitchison 2005 Supersymmetry and the MSSM: An Elementary introduction *hep-ph/0505105*.
[10] Marcela S. Carena et al. 2003 Suggestions for benchmark scenarios for MSSM Higgs boson searches at hadron colliders *Eur. Phys. J.* C **26** pp. 601-607.
[11] CMS Collaboration 2008 The CMS experiment at the CERN LHC *JINST* **3** S08004.
[12] CMS Collaboration 2012 Search for neutral Higgs bosons decaying to τ pairs in pp collisions at $\sqrt{s} = 7$ TeV *Phys.Lett.* B **713**, 68 [arXiv:1202.4083 [hep-ex]].
[13] CMS Collaboration 2011 Measurement of Inclusive W and Z Cross Sections in pp Collisions $\sqrt{s} = 7\ TeV$ *JHEP* **1110** 132.
[14] A. L. Read 2000 Modified frequentist analysis of search results (the CLs method) *CERN-OPEN-2000-205*.
[15] CMS Collaboration 2012 MSSM Higgs production in association with b quarks - all hadronic *CMS-PAS-HIG-12-026*.
[16] CMS Collaboration 2011 b-Jet Identification in the CMS Experiment *CMS-PAS-BTV-11-004*.
[17] CMS Collaboration 2012 Search for SuperSymmetric Higgs boson states decaying into $b\bar{b}$ and produced in association with b-quarks in events collected by semi-leptonic triggers in pp collisions at $\sqrt{s} = 7\ TeV$ *CMS-PAS-HIG-12-027*.
[18] CMS Collaboration 2013 Search for a Higgs boson decaying into a b-quark pair and produced in association with b quarks in proton-proton collisions at 7 TeV *arXiv:1302.2892* [hep-ex].
[19] Tevatron New Phenomena & Higgs Working Group Collaboration 2010 Combined CDF and D0 Upper Limits on MSSM Higgs Boson Production in tau-tau Final States with up to 2.2 fb-1 *arXiv:1003.3363* [hep-ex].
[20] CDF Collaboration 2012 Search for Higgs Bosons Produced in Association with b-quarks *Phys. Rev.* D **85**, 032005 [arXiv:1106.4782 [hep-ex]].
[21] D0 Collaboration 2011 Search for neutral Higgs bosons in the multi-b-jet topology in 5.2fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV *Phys. Lett.* B **698**, 97 [arXiv:1011.1931 [hep-ex]].