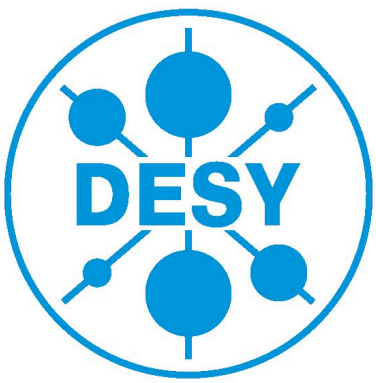

Signal and background studies for MSSM H $\rightarrow b\bar{b}$

DESY Summer Student project report

Andrii Terliuk, Taras Shevchenko National University of Kyiv, Ukraine

September 12, 2011

Abstract

This is report of my Summer Student project at DESY in 2011, all tasks were done during July 19-September 8 under guidance of Rainer Mankel in the Higgs group at DESY/CMS. Here You can find information about Minimal Supersymmetry Standard Model and Higgs boson as a part of it. Method of reconstruction MSSM $H \rightarrow b\bar{b}$ is shown in this work, combinatorial background and mass resolution were calculated using it. Small shift of peak was found and partially was explained by neutrinos from B-hadrons semileptonic decays. Big part of this report covers b-taggind, its efficiency and mis-tag rates. Also principle to chose discriminant and its level are here. $t\bar{t}$ background was estimated and compared with first 501 pb^{-1} of data collected by CMS experiment at center of mass system energy $\sqrt{s} = 7$ TeV.

Contents

1 Introduction

Results of my summer student project during DESY Summer Student Program are presented here. This work was done under the guidance of Rainer Mankel (DESY/CMS Higgs group). I worked on Higgs analysis for CMS experiment at LHC. My job was connected with the search of Higgs boson in the Minimal Supersymmetry Standard Model, where Higgs production is in association with b-jets. The $H \to b\bar{b}$ may be accessible due to the $\tan\beta$ enhancing. Higgs boson production more to be produced in association with jets from bottom flavor. Some aspects of the signal shape will be described here like mass resolution study, mass shift connected with neutrinos from B-hadron semileptonic decays and propper jet energy corrections for b-jets. Background study includes B-tagging efficiency study. Here efficiency and mis-tag rates were calculated. Background from $t\bar{t}$ process was estimated and compared with first 501 pb^{-1} real data obtained by CMS experiment at the LHC.

2 Theory

2.1 The Higgs boson in the Standard Model

Standard Model (Figure 1) is a very successful theory for explanation of microcosm behavior. It includes fermions - leptons, quarks and bosons - force carriers. There are 3 families of leptons: e, μ, τ with corresponding neutrinos ν_e, ν_μ and ν_τ. There are 3 generation of quarks, each include 2 two quarks. 1st generation - u(up), d(down), 2nd generation - s(strange), c(charm) and 3rd generation - b(bottom), t(top). Bosons carries the elementary forces: electromagnetic (EM), electroweak and strong. Carriers of the electroweak interaction are γ, $W^{\pm}$, Z. The strong interaction is carried by 8 types of gluons.

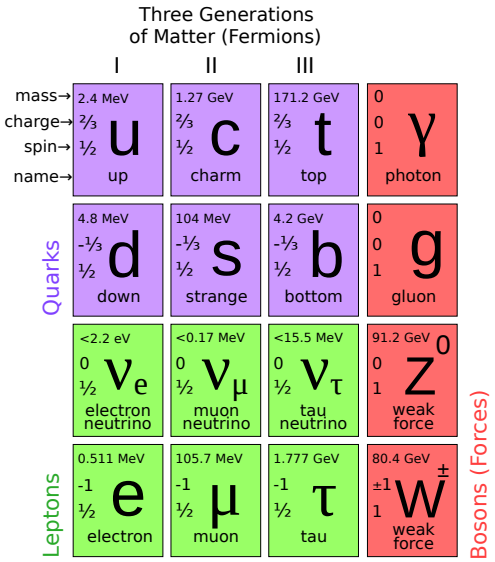

Figure 1: Particles at Standard Model

But in this model all particles are massless, so the theory is incomplete without a special mechanism of electroweak symmetry breaking, what needs to be included to complete the Theory. The most popular mechanism was proposed by Peter Higgs in 1964.

The Higgs mechanism includes the Higgs doublet with a specially shaped potential for spontaneous symmetry breaking. This potential is shown in Figure 2. It has a non-zero vacuum expectation value and generates a scalar field on a physical state. Only one symmetrical point in this potential is unstable(meta-stable), hence symmetry will be spontaneously broken.

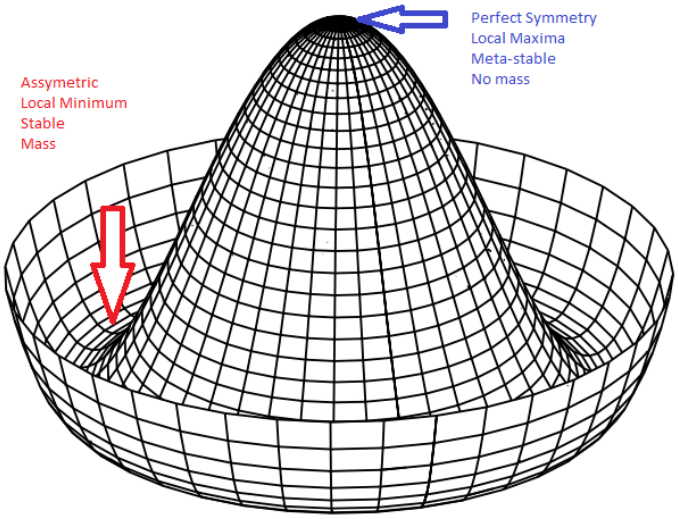

Figure 2: Potential for Higgs mechanism

2.2 Minimal Supersymmetric Standard Model

One direction of physics beyond Standard Model is Supersymmetry(SUSY). In supersymmetric theories each fermion has corresponding boson, and each boson has a corresponding fermion.

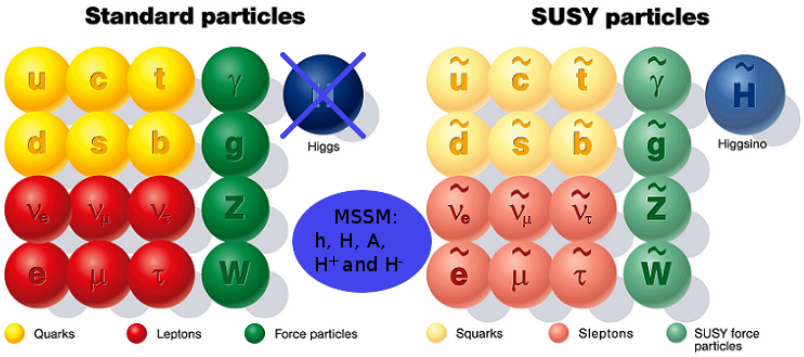

Figure 3: Standard model and SUSY particles

The Higgs sector is extended to Standard Model to the MSSM. This theory includes not one Higgs boson, but it have 5 physical states after electroweak symmetry breaking. Two of these Higgs bosons are CP-even - h and H, one CP odd neutral A, and 2 charged $H^{\pm}$. In Figure 3 particles of the Standard and Supersymmetry models are shown.

There are several ways to produce a MSSM Higgs boson: gluon fusion and associated production with b-jets. These processes include Yukawa coupling at level $g \sim tan\beta$

and cross-section depends like $\sigma \sim tan^2\beta$. In popular scenarios $tan\beta$ is at the level of 20. Therefore, the cross section is enhanced by factor of 400 in comparision to the SM, decay to $b\bar{b}$ and associative production with jets from bottom flavor are very probable. Feynman diagrams for processes are shown at Figure 4.

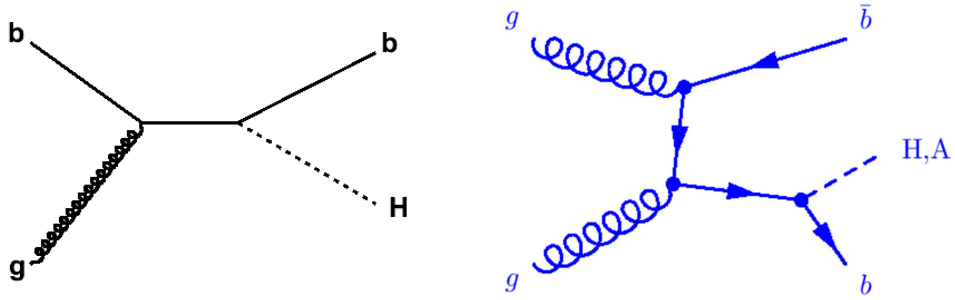

Figure 4: Higgs boson production

In the final state we have 4 jets from bottom quark, but the experimental signature to find $H \to b\bar{b}$ decay is at least 3 b-jets in final state, due to low P_T of 4th jet from b.

3 Experiment

3.1 The Large Hadron Collider

At this moment, Larger Hadron Collider is the biggest particle accelerator all over the World. It was built to answer the most fundamental questions of physics and understand better the deepest laws nature.

The main purposes and questions for the LHC are:

1. Is the Higgs mechanism of mass generating due spontaneous electroweak symmetry breaking correct?
2. Is supersymmetry realized in nature, do all the particles have corresponding partners?
3. Are there extra dimensions as predicted by string theory?
4. What is the dark matter?

The LHC lies in a tunnel of 27 km in circumference, as deep as 175 meters beneath the Franco-Swiss border near city of Geneva. The machine is designed to collide protons at an energy of 7 TeV per beam (center of mass energy $\sqrt{s} = 14$ TeV). Now it is being operated at an half-energy 3.5 TeV per beam($\sqrt{s} = 7$ TeV).

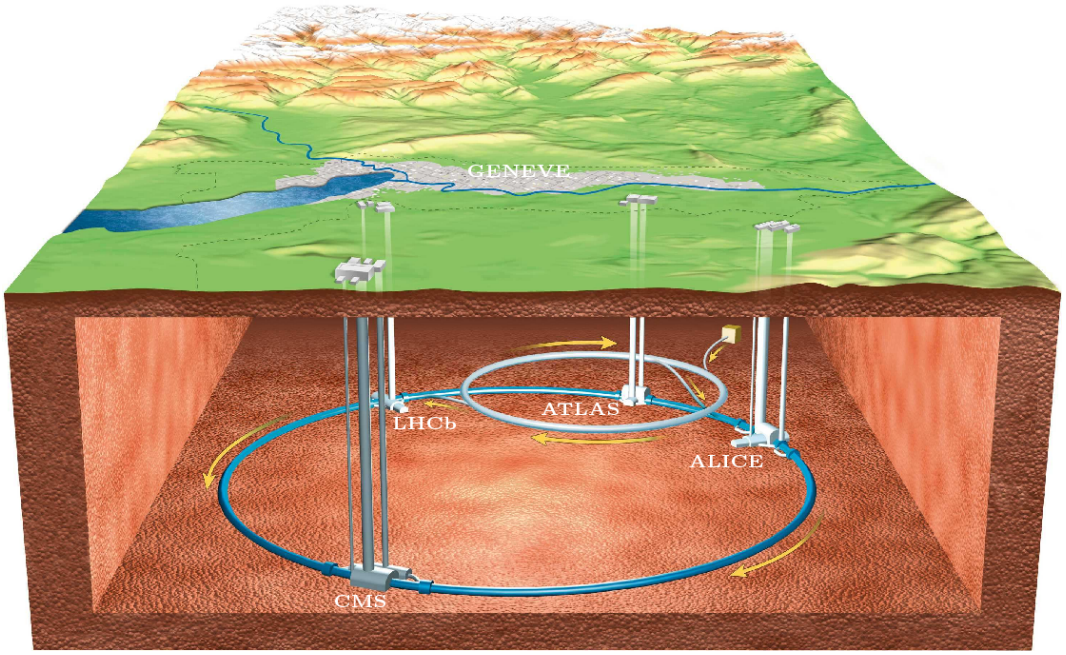

Figure 5: Experiments at the LHC

There are 4 main experiments at LHC. Two of them, The Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS), are multipurpose detectors. The

Large Hadron Collider beauty(LHCb) experiment was built to measure parameters of CP violation. A Large Ion Collider Experiment(ALICE) was built to understand better the properties of a quark-gluon plasma. A schematical overview of the LHC and experiments is shown in Figure 5.

3.2 The CMS detector

The compact muon solenoid is one of the two multipurpose detectors at the LHC. It have size of 15×15×21 meters and a weight near 12,500 tons. It has a barrel and end caps design and consists of several detector layers. An overview of detector is shown in Figure 6.

The first layer, the nearest to interaction point, is the tracker for measuring tracks from charged particles with high precision. It consist of 13 layers(14 in the end caps). The innermost three layers have pixels and consist in total of 66 millions of 100×150 μm silicon pixels. The next four layers consist of 10 cm ×180 μm strips, and the remaining layers are made from 25 cm × 180 μm strips. There are 9.6 million silicon strip channels in total.

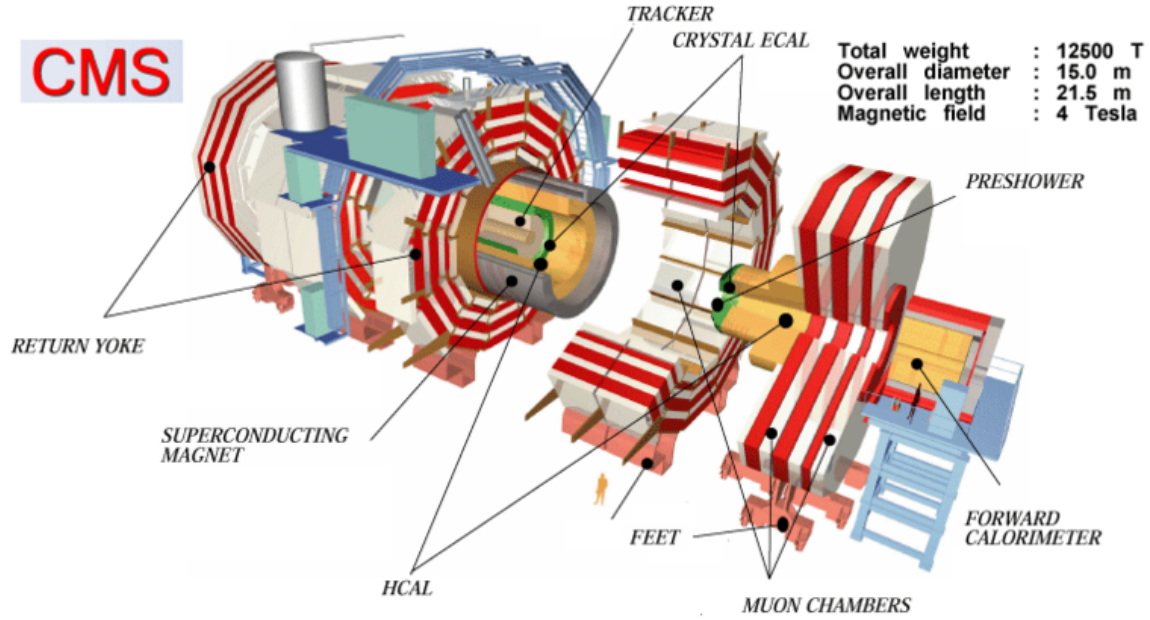

Figure 6: CMS detector

The second layer of the detector is the Electromagnetic Calorimeter(ECAL) and designed to make energy measurement of photons and electrons with high accuracy. It consists of crystals of lead tungstate, $PbWO_4$. This is an extremely dense but optically clear material, ideal for stopping of high energy particles. The crystals have a front size of 22×22 mm and a depth of 230 mm. Each crystal is backed by silicon avalanche

photo-diodes for readout. The barrel region consists of 61,200 crystals, with a further 7,324 in each of the end caps.

The next layer is the Hadronic Calorimeter(HCAL) and its purpose is both to measure the energy of individual hadrons produced in each event, and to be as near to hermetic around the interaction region as possible to allow events with missing energy to be identified. The HCAL consists of layers of dense material (brass or steel) interleaved with tiles of plastic scintillators, read out via wavelength-shifting fibers by hybrid photo-diodes. This combination was determined to allow the maximum amount of absorbing material inside of the magnet coil.

The fourth layer of CMS is the large solenoid magnet for creating a strong magnetic field. This allows the charge/mass ratio of particles to be determined from the curved track that they follow in the magnetic field. It is 13 m long and 6 m in diameter, and its superconducting niobium-titanium coils are designed to produce a 4 T magnetic field. The inductance of the magnet is 14 H and the nominal current for 4 T is 19,500 A, giving a total stored energy of 2.66 GJ, equivalent to about half-a-tonne of TNT.

The outer layer of detector is the muon system and return yoke. It was built to identify muons and measure their momenta, CMS uses three types of detector: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The DTs are used for precise trajectory measurements in the central barrel region, while the CSCs are used in the end caps. The RPCs provide a fast signal when a muon passes through the muon detector, and are installed in both the barrel and the end caps.

4 Analysis

4.1 Algorithm and expected background

For selecting candidates for the MSSM Higgs boson we need a suitable algorithm. As was shown before, we need 3 b-jets as experimental signature. So, the event must contain at least 3 jets with next criteria:

1. All jets $|\eta| < 2.5$. This corresponds to the acceptance for B-tagging.
2. 1st jet $P_T > 46$ GeV.
3. 2nd jet $P_T > 38$ GeV.
4. 3rd jet $P_T > 15$ GeV.

The reconstructed mass is calculated from the sum of 4-vectors momenta for the first two leading jets.

For the analysis of this decay channel we need to understand its signal shape and combinatorial background after reconstruction. We expect high background from QCD multi-jet production, which can contain 3 b-jets in the final state. Another possibility is 2 b-jets production with a 3rd jet is from charm or light flavor production. To suppress this background we need to recognize B-hadron decays in the jet - good B-tagging is essential.

At LHC energies we need also to understand the background from $t\bar{t}$ events, which is much higher than at the Tevatron.

4.2 Signal shape and mass resolution

For understanding the signal shape and resolution, a Monte Carlo(MC) signal sample with with integrated luminosity 9.823 fb^{-1} and generated mass of Higgs boson $m_H = 120$ GeV was used.

The "natural" spectrum and shape of signal are shown on Figure 7. Using the MC generator information we know, which jets are from the Higgs decay, thus we can studu combinatorial background. In the right plot, green color means true Higgs(two jets are matched by MC generator to Higgs boson), red curve - Higgs fitting of "true Higgs" mass, light blue color - when only one jet is matched to Higgs and dark blue - when 2 jets are not matched. The latest two are combinatorial background. The peak is clearly visible, and the combinatorial background is not high.

For finding the mass resolution, the difference between reconstructed mass and generated Higgs mass at this event was plotted, which shown on Figure 8.

From Gaussian fitting of the "true Higgs"(red curve at right histogram at Fig. 7) mass spectrum, we see that center of the peak is $m_{Hrec} = 111.3$ GeV, from fitting on Fig. 8 we can estimate the width of the peak $\sigma_m = 12.7$ GeV. From this information we can estimate the mass resolution for this mass:

$$\frac{\sigma_m}{m_{Hrec}} \approx \frac{12.7GeV}{111.3GeV} \approx 11.4\%$$

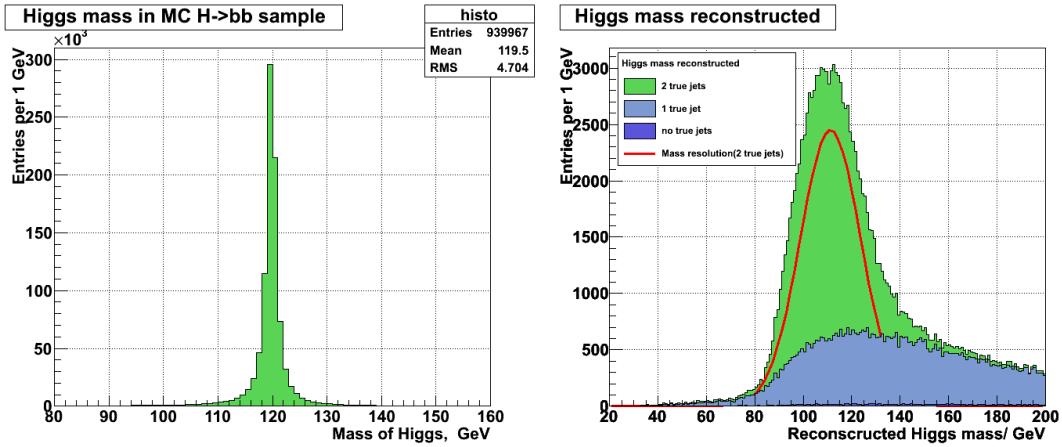

Figure 7: Mass of generated Higgs before and after reconstruction

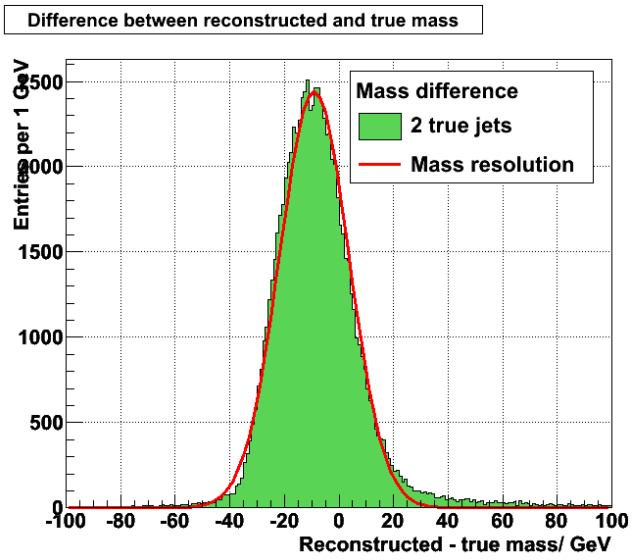

Figure 8: Difference between reconstructed and generated Higgs mass

This value is in agreement whith the resolution estimated from the jet energy and momentum resolution.

4.3 Mass shift and neutrinos

From fitting of Figure 8, a mass shift at level -9 GeV was found. One possible reasons is neutrinos from semi-leptonic decays of B-hadrons in the jet. Because neutrinos are invisible for the detector, the part of the energy is lost. The effect of neutrinos was studied by selecting jets with associated muons. For matching of muon to jet, the

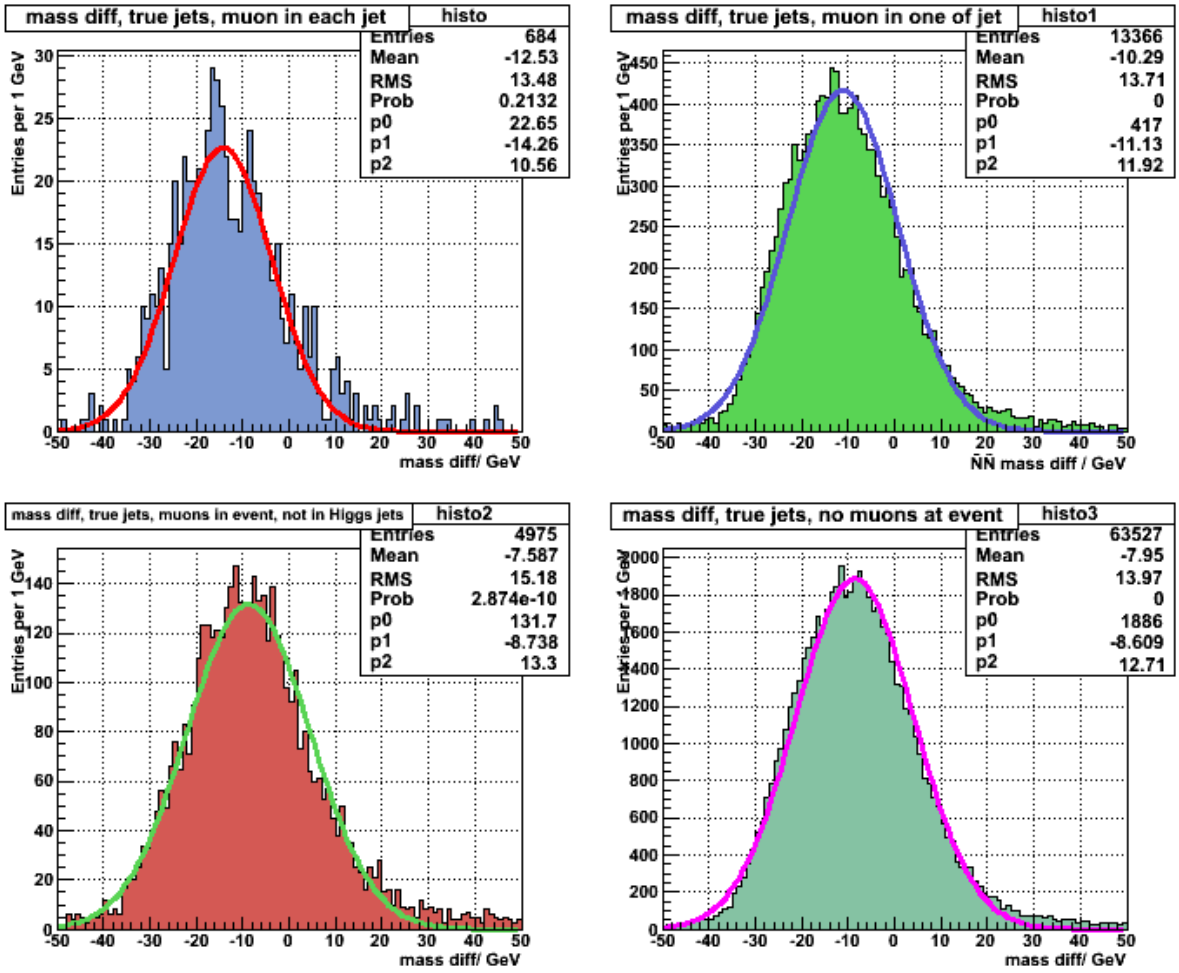

Figure 9: Mass shift

standard ΔR distance was used. The muon is matched to the jet if:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.5$$

where $\Delta\phi$ is the polar angle difference between jet direction and muon direction and $\Delta\eta$ the difference in pseudo-rapidity.

In Table 1 the mass bias is shown for different muon position.

Hence, neutrinos from semileptonic decays can explain approx. 3 GeV of shift per jet with a neutrino. Another possible reason for a mass shift is the jet energy correction, what was not specific for b-jets, but the study of this effect, what was not a goal of this work.

Muon position	Mass shift
Muon in each jet	-14.3 GeV
Muon in one jet	-11.1 GeV
Muon in event	-8.7 GeV
No muons	-8.6 GeV

Table 1: Mass shift for different muon positions

4.4 B-tagging

As shown before, we need to suppress background from light flavors, so we need to cut off light flavor jets and keep bottom jets. Bottom hadrons have large lifetime, this fact allows to determine b-jets using impact parameters of shifted tracks from B-hadron decays. Due to the high mass of the b-quark, ∼5 GeV, the jets also have high multiplicity of displaced tracks.

4.4.1 Track counting method

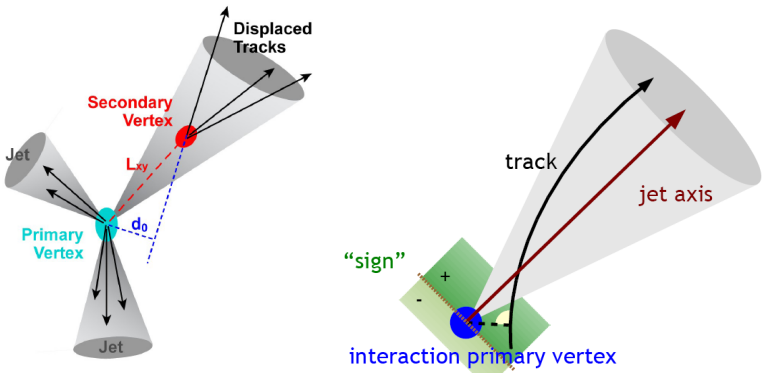

Figure 10: Illustration of TCHE(TCHP) method

Track counting is one of methods to recognize b-jets. There are 2 main variants: TCHE(Track Counting High Efficiency) uses the 2nd displaced track with highest impact parameter significance:

$$TCHE = \frac{d_{3D}}{\sigma_{3D}},$$

for TCHP(Track Counting High Purity) one uses 3rd track:

$$TCHP = \frac{d_{3D}}{\sigma_{3D}}$$

where d_{3D} - 3-dimensional impact parameter, σ_{3D} - uncertainty of 3D IP.

Determination of impact parameter and sign of discriminant are shown in Figure 10.

4.4.2 B-tag efficiency

Understanding of efficiency is needed for good B-tagging. Choosing the true b-jets from Monte Carlo it's possible to know it. Used Pythia QCD Multi-jet MC sample.

For determination efficiency of B-tagging the fillowing formula is used:

$$\epsilon = \frac{N_{b-tagged}}{N_{b-all}}$$

where $N_{b-tagged}$ - number of tagged b-jets, N_{b-all} - number of all b-jets.

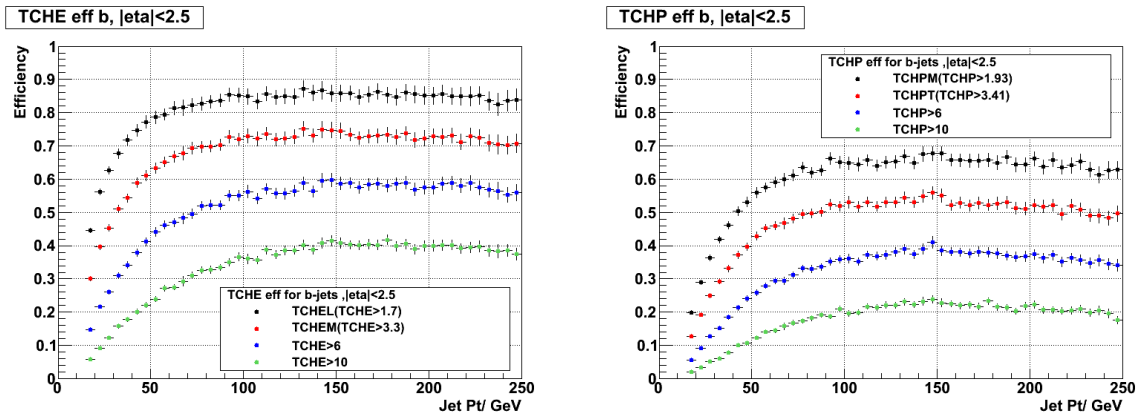

Figure 11: TCHE and TCHP efficiencies

Efficiencies shown on Figure 11, left part for TCHE and right for TCHP. The TCHE(TCHP) b-tagging efficiency decreases with higher threshold of discriminant. The efficiency for TCHE is higher than for TCHP for the same level of the discriminant.

The result are comparable with the official CMS b-tagging group paper(BTV-11-001).

4.4.3 B-tag mis-tag rate

It is important to know for th background study, how many jets from light flavor and charm flavor are selected by b-tagging. For this we can plot the "efficiency" for light flavor and charm jets, in the way like it was shown at Figure 11. Mis-tag levels for different flavor and discriminants are shown in Figure 12.

TCHE is plotted in the left column at Figure 12, TCHE - at right part. Light flavors are shown in the top half, charm - at bottom. From this plot we can conclude, that the mis-tag rate rises with increasing jet P_T. The efficiency for non b-jets decreases with higher values of the TCHE and TCHP discriminant, so higher discriminants have a better purity.

Also it is very important that jets from charm flavor have much higher mis-tag efficiency due similarly high mean lifetime.

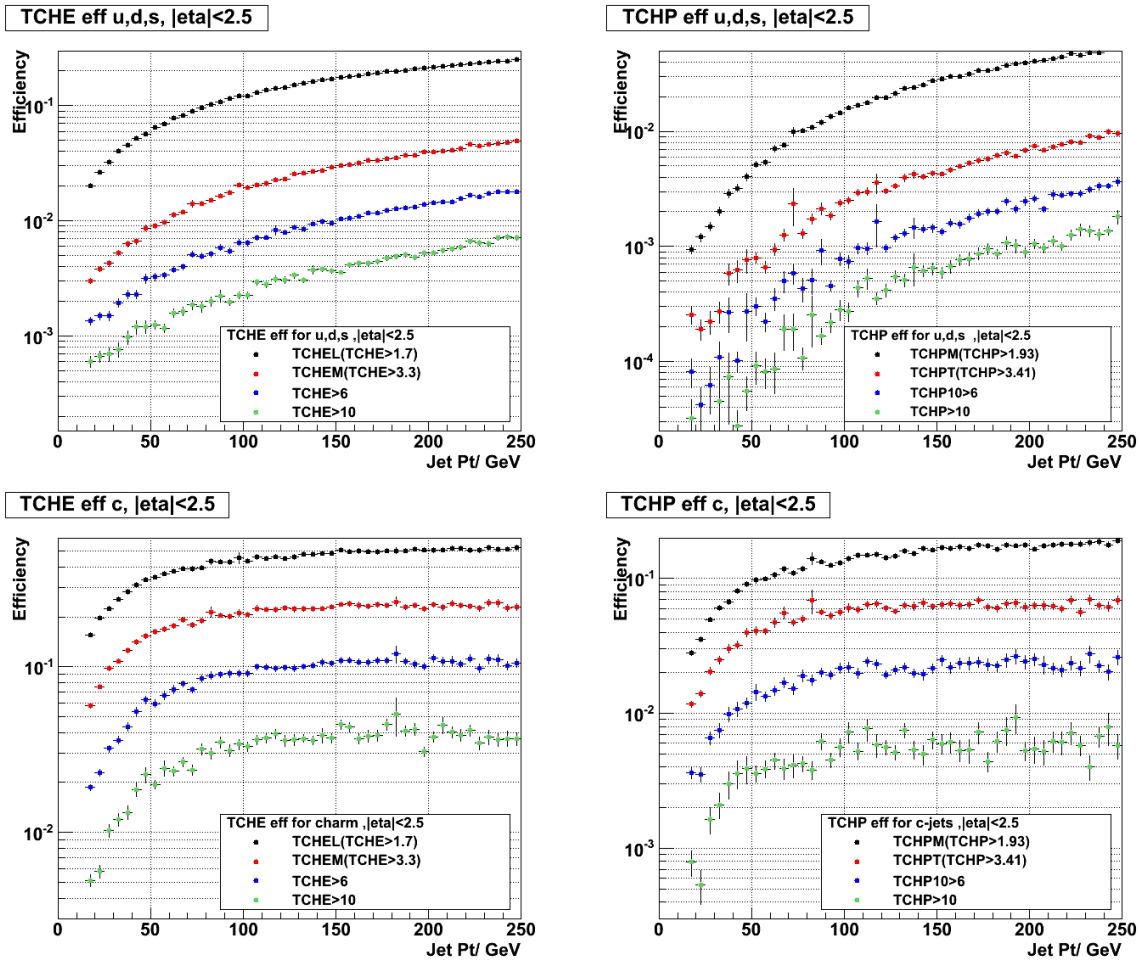

Figure 12: TCHE and TCHP mis-tag rates for light flavor and charm jets

4.4.4 Working point study

Due to the high background from QCD we need to choose which discriminant and at what level is suitable. For this purpose we need to understand how the mis-tag rate depends on the efficiency for different methods and threshold levels. This kind of plot is shown on Figure 13, the top plot shows this function for TCHE method and bottom - for TCHP. For these plots overall efficiency and mis-tag rate for jets are shown with $p_T > 50$ GeV.

From these plots we can conclude, that higher purity threshold level have worse efficiency, but better purity. So, choice of working point is a compromise between this two parameters. For example, look at two working point: with efficiency at level 80 % (high efficiency working point) and with mis-tag probability of 10^{-3}(high purity working point).

For THCE at 80 % efficiency we have a mis-tag probability at a level of 10 %, but for

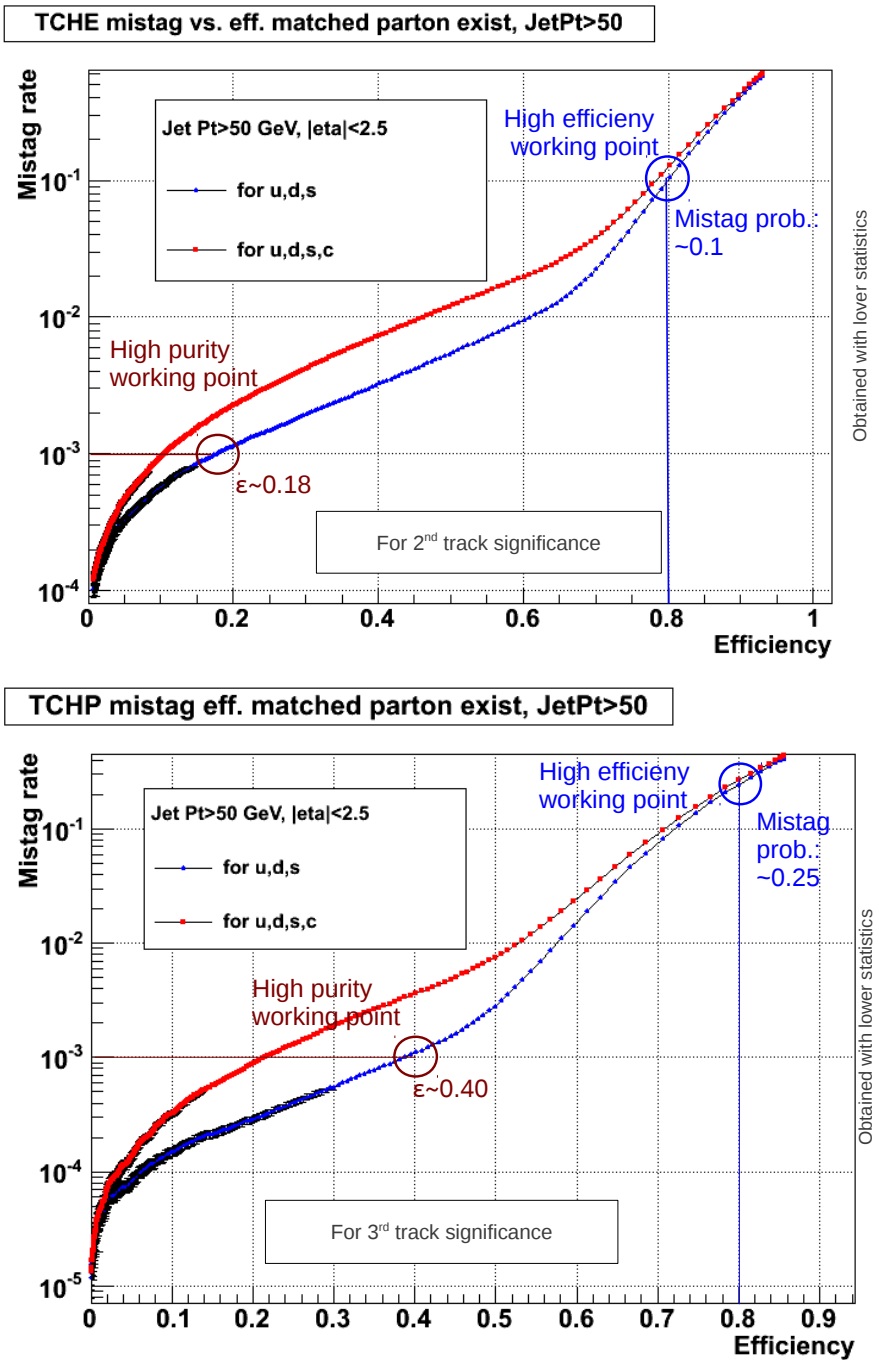

Figure 13: TCHE and TCHP mis-tag rate vs. efficiency

TCHP it is 25 %. But then have a look into high purity points: for TCHE efficiency 18 %, but for TCHP it is 40 %. So, we can explain more about different discriminants usage. In the high efficiency regime, usage of TCHE is better, because it gives better purity, otherwise for better purity usage of TCHP is preferable - it gives better efficiency.

As a result from here, for $H \to b\bar{b}$ analysis we need to use the high purity regime, so we need at least TCHP>3.41(TCHP Tight) for a better signal to background ratio.

4.5 $t\bar{t}$ background

Because of much more higher center of mass system $\sqrt{s}$ energy at LHC we need to consider and look into one more background, into $H \to b\bar{b}$ events. This process also produces b-jets in the final state.

For this purpose we can use Monte Carlo for this process and compare it with expected signal, with weighting with respect to generated luminosity. This comparison is shown on Figure 14. For selecting b-jets b-tagging was used at level of TCHPT(which means TCHP>3.41).

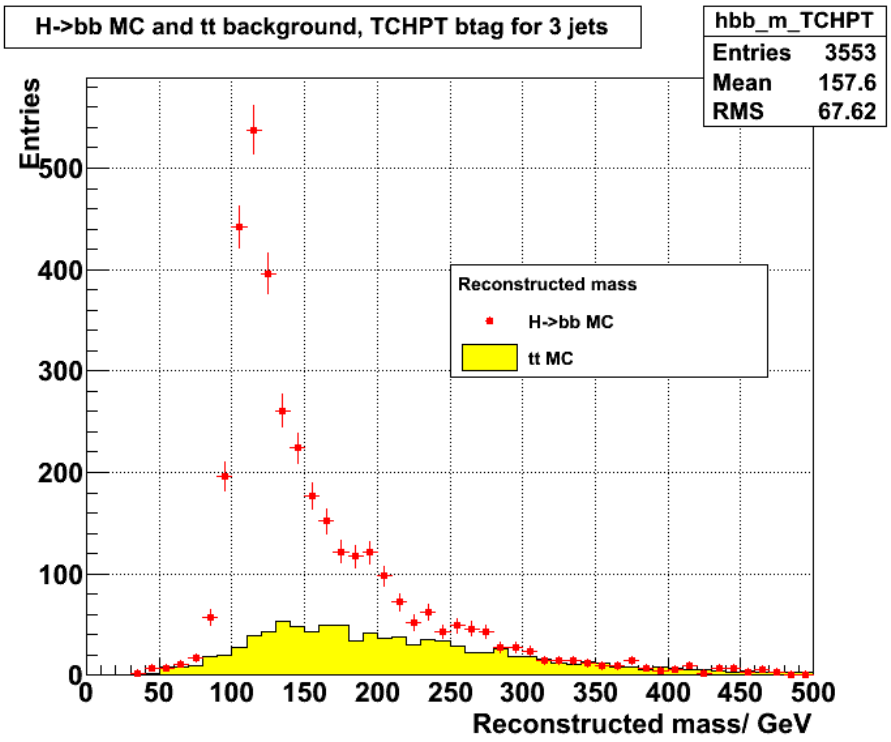

Figure 14: $H \to b\bar{b}$ in comparison with $t\bar{t}$ background

From Figure 14 we can make the conclusion, that background from the $t\bar{t}$ process is much more wider and lower, and at this point of analysis it can be neglected.

Finaly we compare with real data. Figure 15 shows comparison between first 501 pb^{-1}, signal sample MC and $t\bar{t}$ background. "HLT_CentralJet46 BTagIP3D_CentralJet38_ BTagIP3D" trigger was used for signal sample and real data. It is looking for events with 1st jet energy greater than 46 GeV, second - greater than 38 GeV, what were b-tagged. At this level TCHPT b-tagging was required for all 3 jets.

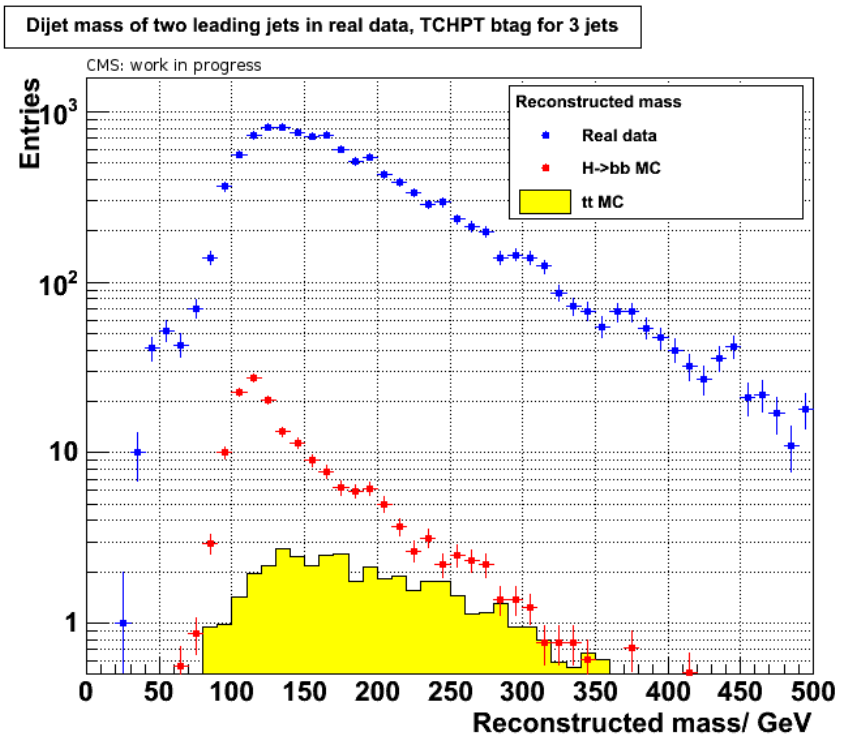

Figure 15: $H \to b\bar{b}$ in comparison with $t\bar{t}$ background and real datas

The bump in the range of 50 GeV corresponds to gluon splitting and can be subtracted with making a cut on the ΔR distance between two jets. If apply cut at the level of 1, this peak disappears. Almost all background corresponded is to QCD multi-jet production. So, this background needs very careful optimization and treatment to subtract it.

5 Summary

During my Summer Student Project I have worked in the context of MSSM Higgs in decay channel $H \to b\bar{b}$.

The signal shape of reconstruction was studied in this channel and the resolution found to be in the range expected from the jet energy resolution. Also the combinatorial background from selecting method is not high and it does not disturb the peak. For mass resolution we have a number around 11 %. A slight mass shift was found and partially described by neutrinos from B-hadron semileptonic decays, which can explain -3 GeV of bias. Another reason is the b-jets are wider and need specific jet energy corrections.

At b-tagging studies efficiency and mis-tag rates was calculated. From this data was calculated mis-tag vs. efficiency plot and level and type of b-tagging discriminant was chosen for better purity. This leads to the decision to use TCHP discriminant cut at least for a threshold level of 3.41.

$t\bar{t}$ background was studied and compared with the expected signal and real data, so $t\bar{t}$ production gives us a wide and small background, which currently moment it can be neglected in this analysis. Most of the background is from QCD multijet production, and it needs very careful optimizations and subtraction of this kind of processes.

References

[1] CMS official note: BTV-11-001

[2] Top Quark Physics. Flavor Physics from the Tevatron to the LHC. 18 Aug 2011, DESY Summer Student Program, Hamburgs *Georg Steinbruck*

[3] Higgs Searches at CMS, CMS Hamburg/DESY Meeting, 24 August 2011 *Alexei Raspereza*

[4] Higgs Theory, Lepton-Photon 2011 *Abdelhak Djouadi (LPT Orsay)*

[5] http://www.mpi-hd.mpg.de/lhcb/index.php?id=lhc

[6] http://en.wikipedia.org/wiki/LHC

[7] http://cms.web.cern.ch/

[8] http://en.wikipedia.org/wiki/Compact_Muon_Solenoid/