FACULTY OF PHYSICS UNIVERSITY OF BELGRADE

Dimitrije Maletić

SOME ASPECTS OF THE BACKGROUND REDUCTION OF THE TWO PHOTON HIGGS DECAY CHANNEL IN THE CMS DETECTOR

- Ph.D. Thesis -

(Shorter version in English)

Belgrade, January 2009.

The work in this thesis was done in the Group for elementary particle physics in Laboratory 010 of Institute of Nuclear Sciences "Vinča", and within a Science project 141038, supported by the Ministry of Sience and Technological Development of Republic of Serbia.

I want to thank my mentor, prof. dr Peter Adzić and the CMS Belgrade group for providing me the opportunity to work on this thesis.

I owe gratitude to collegues Dragoslav Jovanović, Predrag Milenović, Jovan Puzović and Miloš Đorđević for useful sugestions and help in my work.

I thank my wife Slavica for understanding and help and my son Stefan for continous inspiration.

Dimitrije Maletić

Abstract

The main subject of my thesis is in connection with the research of properties and phenomena of two photon Higgs decay acording to the Standard Model (SM). The main goal was a development of most optimal method of background reduction of this decay channel. To achieve this goal, it was necessary to do researchs and complete several research tasks. These tasks are:

- gaining a good knowledge of CMS ECAL detector, with emphasis on its barrel part and Preshower detector. Also, there was a need to have a good knowledge of CMS simulation and reconstruction softwares like: OSCAR, ORCA and CMSSW.
- application of gained knowledge to create stand alone Geant4 detector simulation for test beam eksperiement of Preshower prototype in 2004. and to compare simulation and eksperimental results.
- to use knowledge of Preshower detector to work on modification and improvement of geometry description in CMSSW.
- as a developer of a Preshower geometry, to help with simulation for combined test beam of ECAL, HCAL and Preshower detector in 2007.
- To have a good knowledge of previous analysis of signal detection in two photon Higgs decay.
- To impelment previous knowledge about separation of neutral pions and photons form analysis for L3 and CMS ekperiments into algorithm based on Artificial Neural Network (ANN) simulations.
- There was a need to implement this algorithm into ORCA, and later in CMSSW, then to test it against previous algorithms and analysis of two photon Higgs decay in CMS.

This thesis shows that the simulation results developed for Preshower prototype test during 2004 are in excellent agreement with the experimental results.

It is also shown that the use of algorithm for separation of neutral pions and photons based on information from ECAL and Preshower detector give better results than previous methods. For a signal of unconverted and isolated photon candidates the gained reduction of neutral pions from photon signal can vary from 29% to 75% depending on method used for unconverted photon selection, particle's transverse momentum and direction. Algorithm test on $H\rightarrow\gamma\gamma$ and γ +jet events showed that the background from γ +jet events can be reduced by 60% leaving 90% of signal photons from two photon Higgs decay.

Finally, it should be mentioned that the developed algorithm was implemented in CMS reconstruction software – CMSSW and that it is continuously updating and adapting for official analysis of two photon Higgs decays on CMS detector.

Contents

1. Introduction	5.
2. Preshower prototype test beam 2004	6.
2.1. Geant4 Preshower prototype simulation	
2.2. Results of comparison of Preshower detector	
experimental and simulation data with incident muons	6.
2.3. Results of comparison of Preshower detector	
experimental and simulation data with incident electrons and pions	7.
3. Algorithmic XML geometry description of the Preshower	<i>9</i> .
4. Algorithm for separation of unconverted and isolated	
neutral pions and photons	
4.1. ANN training and performance	11.
4.2. Trying to train the ANN from data	14.
5. The use of photon background reduction in	
official Higgs to two photons decay analysis	15.
6. Conclusion	18.
Literature	19.

1. Introduction

The work on improvement of two photon Higgs decay background reduction consists of two parts: work on Preshower detector and work on algorithm for separation of neutral pions and photons in CMS software.

The part of work on Preshower detector starts in the middle of 2004. by participating in Preshower prototype test beam setup activities. Most important contribution is a development of stand alone Geant4 detector simulation. Results of this simulation and comparison with experimental results made possible the calibration of Preshower detector.

The work on Preshower was continued with the implementation of new detector geometry in CMS reconstruction software – OSCAR (Object oriented Simulation for Cms Analysis and Reconstruction).

It should be mentioned that there was a need of modification of Preshower detector during the years because the actual technical design of Preshower was also changed, and also its porting from ORCA to CMSSW. Next task was to use the algorithmic descrytion in XML (eXtensible Markup Language) description code. Algorithmic XML description code allowed an execution of external C++ algorithms. After the optimization of the use of algorithmic XML Preshower geometry description, this geometry description was not changed in CMSSW.

The second part of work on two photon Higgs decay background reduction was the work on development of algorithm for separation of neutral pions and photons which started also in 2004. The current status of this work is that the developed algorithm is implemented in CMSSW and it is adapting for use by official analysis of two photon Higgs decay.

At the beginning of algorithm development there was a need to find optimal neutral pion discriminating variables and work on Artificial Neural Network simulators. Algorithms were developed and implemented first in ORCA and later also in CMSSW.

It is probable that there will be a need for further improvement of this algorithm, until there is a possibility o final improvements using real data from CMS detector during LHC runs. Until then, algorithm had a last test using real data form a Cosmic tests - CRUZET (Cosmic Run at Zero Tesla) in 2008.

2. Preshower prototype test beam 2004

2.1. Geant4 Preshower prototype simulation

With a development of stand alone Geant4 Preshower prototype simulation it was important to precisely define detector geometry. Since the experimental data came from electronics which deforms a signal shape it was necessary to deform in a similarly way the clean signal from a detector simulation by adding noise or simulate pedestal cuts etc. Only simulation date processed in this way could be compared with the experimental data.

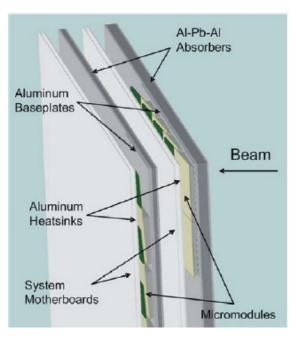


Image 1: Visualization of geometry in simulation of Preshower prototype.

2.2. Results of comparison of Preshower detector experimental and simulation data with incident muons

Image 2 shows the simulation results of totally deposited energy for sensors in both planes, X – first and Y the second plane. The peak around 0.09 MeV corresponds to 1 mip (minimal ionisating particle), and it can be noticed that there is another peak around 2 mip which can be explained by delta electrons created on the path of muons passing through the same silicon sensor as a incident muon. Existing Landau tail at higher energies is a consequence of bremsstrahlung [1].

Absolute energy calibration of silicon sensors are:

Sensor in X plane 0.0887 ± 0.0001 : 49.01 ± 0.10 ADC channels; sensor in Y plane 0.0871 ± 0.0001 : 43.75 ± 0.11 ADC channels.

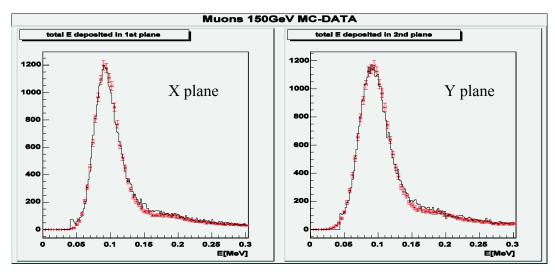


Image 2: Results of simulation with noise added compared with experimental results.

To make a simulation more realistic the noise was added to each silicon strip. This noise has a Gauss distribution with a width of 5.39 ADC channels for X sensor and 5.9 ADC channels for Y sensor which corresponds to mean noise of 0.0098 MeV for sensor X, and 0.0117 MeV for sensor Y.

In image 2 it can be noticed that when the calibration is applied i.e. when ADC channels are converted into energy in MeVs, there is an excellent agreement of simulation with experimental results [1].

2.3. Results of comparison of Preshower detector experimental and simulation data with incident electrons and pions

Incident pions behave like mips. They pass through Preshower without any interactions even in the most dense material of lead absorber. Deposited energies are described using Landau distribution with most probable energy of 91 keV for 320 μ m thick silicon sensor. Only a small percentage of events have a hadronic interaction with Preshower material. In those events there is a deposited amount of energy in silicon sensors.

Incident electrons initiate a creation of a shower in one of the lead absorbers. With the increase of energy of incident particle the shower has a wider spread resulting in larger energy deposition in silicon sensors (image 3). In image 4 the mean energy deposition is shown in dependence of momentum of incident electron for both Preshower planes.

The difference in energy deposition can be used to separate electrons and pions.

Image 5a. shows the good agreement of experimental results and a simulation both for electrons and pions.

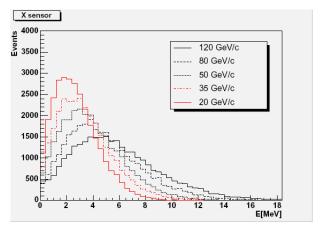


Image 3: Deposited energy for incident electron with a momentum from 20 to 120 GeV.

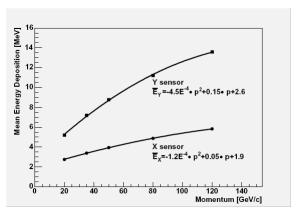


Image 4: Mean deposted energy in sensors X and Y depending on electron momentum.

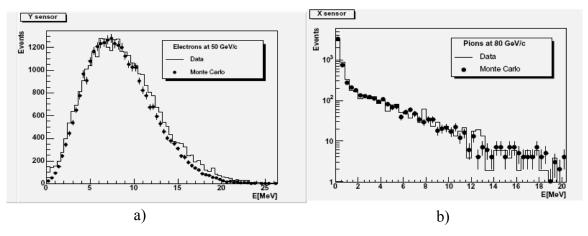


Image 5: Mean deposited energy of incident electron a) and pion b).

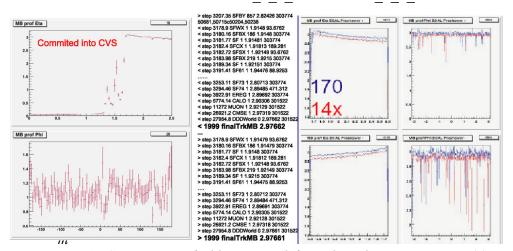
3. Algorithmic XML geometry description of the Preshower

A detector geometry can be described in many ways (C++ code, converted technical drawings etc.), and in CMS simulation programs OSCAR before, and CMSSW today it is described by using an algorithmic XML (eXtended Markup Language) code. Update and improvement of geometry description consisted of implementations of improvements in technical design of Preshower detector as well as the optimization of geometry description implementations [2,3].

Summarized results on implementation and optimization of algorithmic XML Preshower geometry description in CMSSW is practically a difference of geometry descriptions of old and new versions CMSSW_1_4_0 and CMSSW_1_7_0 respectively:

- The description had to be as close as possible to the installed Preshower;
- Geometry was based on technical drawings of parts of Preshower;
- Inner and outer ring were added;
- New shape of lead absorber was added:
- The thickness and size of silicon sensors were changed;
- about 4300 micromodules are organized in 12 types of "ladders" in order to improve the efficiency of the simulation;
- silicon sensors were realistically positioned.

The static description contained in one description code all the definitions of all geometrically described parts of the detector – *solids*, all logical parts describing all the properties of parts of a detector except its geometry, and physical copies. This code had 860.000 characters in 26.500 lines, which new geometry reduces 36 times. Static code contained also the positions of 4300 physical copies of micromodules, which in the new geometry should be calculated by an iterative way. Results of the implementation of first algorithmic geometry gave perfect agreement of old static and new algorithmic geometry (image 6a). Image 6b shows the difference of geometries with all mentioned changes i.e. the geometries in CMSSW versions CMSSW 1 4 0 μ CMSSW 1 7 0.



age 6: Comparison of old static and first algorithmic geometries (a). Comparison after all the changes in Preshower geometry i.e. comparison of geometries in CMSSW 1 4 0 and CMSSW 1 7 0 (6).

The goal for reorganization of micromodules into "ladders" is to lower the number of iterations when going through copies of micromodules in a search for impact points of

incident particles. There are 4300 copies of micromodules of Preshower and half of them were in each iterative process. After the reorganization (Image 7c), there were 12 types of "ladders" containing several micromodules. There are only two common one ladder types: "type0" and "type1". If, for instance, we have one incident particle then the number of iteration within the search for a impact micromodule is like: 2 (+,- part of CMS – two Preshowers) + 2 (common types of ladders) + 110 (copies of ladders) + 10 (micromodules in a ladder) which gives the maximum of 124 iteration, comparing to 2152 iteration with old geometry.

Te new geometry description is, naturally, useful for visualization of the Preshower detector. This geometrical organization is also more realistic from a design point of view. For instance it is easy to select all ladders of the same type at once or to see which ladders will give signal to which motherboards etc. (Image 7a).

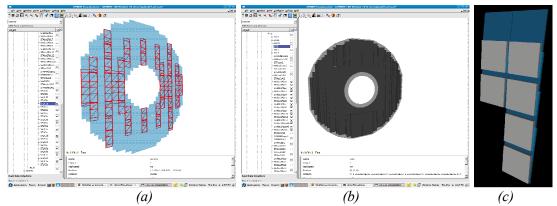


Image 7: A possibility to visualize ladders of a same type in new algorithmic geometry implemented in CMSSW_1_7_0. Visualized by IGUANA-e in CMSSW (a). Optimized shape of lead absorbers (b). Micromodules organized in one type of ladder (c).

Also, there was a need to implement a new shape of lead absorber. This new shape was developed to avoid partial covering of ECAL endcap crystals by lead absorbers in Preshower. Image 7b shows optimized lead absorber visualized using IGUANA program as part of CMSSW.

4. Algorithm for separation of unconverted and isolated neutral pions and photons

CMS "Higgs to two gamma group" had, besides standard way of separation neutral pions and photons by application of cuts on kinematical variables of photon candidates, initiated a research of new way of separation using of the Artificial Neural Network simulators. This method was developed within CMS "Egamma" group. Work on separation was divided in two parts. The physicists who focus on separation of neutral pions and unconverted photon candidates are members of CMS Belgrade group (INS Vinča and Faculty of Physics) and group from Institute "Democritos", Athens. Other groups are focused on converted photon candidates.

A significant background to the $H\to\gamma\gamma$ process originates from π^0 's in jets, which fake single isolated photons. Thus, an electromagnetic cluster may be due to an incident photon, or a pair of closely-spaced photons from the decay of a π^0 inside a jet. The lateral shape of the energy deposit in the Electromagnetic calorimeter (ECAL) can be used to distinguish a π^0 from an unconverted and isolated photon especially at low energies where the distance between two photons from the π^0 decay is larger.

For the ECAL barrel a method developed for L3 [4] was adapted and it is based on analysis of the patterns of energy deposited in the ECAL by electromagnetically showering particles.

4.1. ANN training and performance

For the ANN training, the set of γ +Jet events with isolated π^0 s inside the jet forced from the generated level and with various π^0 s E_T ranges: 15-25,25-35,35-45,45-55,55-65 GeV data sample was used (called "GammaJetIsoPi0"), along with the $H \rightarrow \gamma \gamma$ events all produced under the CSA07 official production. An equal number of photon and π^0 s has been also selected after HTL2Photon and Isolation conditions from the same data sample. The number of events in some cases (especially in the Preshower region) is not sufficient to be used for an efficient training of the neural network. In order to overcome the above limitation the neural network has been initialized with the weights obtained from a training with single particles events (40K per particle type per E_T bin) produced with an older version of the CMS reconstruction code where the conversion information was taken from the generation level. After the initialization the network has been retrained for each E_T range using half of photons/ π^0 s for training and half for testing.

Images 8 and 9 left, show overall response of ANN for photons in the ECAL barrel with photon energies of 20 GeV and 60 GeV and for both unconverted photon candidates selection methods "NTr κ =0" μ "R9>0.93" respectively. Images on the right show efficiency and discrimination for these events. Better performance can be noticed for lower energies and "R9>0.93" selection method in the ECAL barrel.

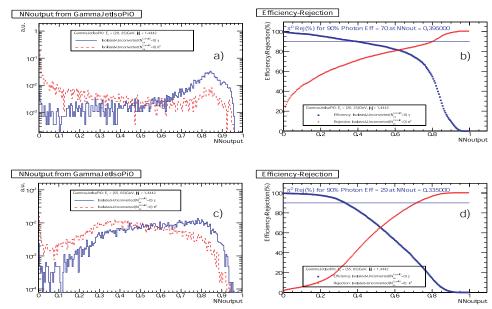


Image 8: (a) NNoutput for isolated, unconverted ("NTrk = 0" method) photons (solid line) and π^0 s (dashed line) of the Barrel ECAL region and of ET 20 GeV, (b) π^0 rejection (red triangle) for 90% photon efficiency (blue box) vs. NNoutput for isolated, unconverted particles of E_T 20 GeV in the Barrel ECAL region. (c) NNoutput for the isolated, unconverted photons (solid line) and π^0 s (dashed line) for the Barrel ECAL region and for ET 60 GeV, (d) π^0 rejection (red triangle) for 90% photon efficiency (blue box) vs NNoutput for isolated, unconverted particles of E_T 60 GeV in the Barrel ECAL region. The plot produced using "GammaJetIsoPi0" events.

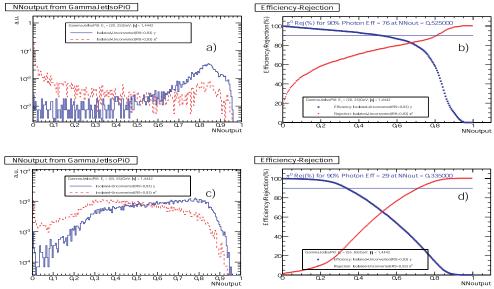


Image 9: NNoutput for isolated, unconverted "R9>0.93" method) photons (solid line) and π^0 s (dashed line) of the Barrel ECAL region and of ET 20 GeV, (b) π^0 rejection (red triangle) for 90% photon efficiency (blue box) vs. NNoutput for isolated, unconverted particles of E_T 20 GeV in the Barrel ECAL region. (c) NNoutput for the isolated, unconverted photons (solid line) and π^0 s (dashed line) for the Barrel ECAL region and for ET 60 GeV, (d) π^0 rejection (red triangle) for 90% photon efficiency (blue box) vs. NNoutput for isolated, unconverted particles of E_T 60 GeV in the Barrel ECAL region. The plot produced using "GammaJetIsoPi0" events.

The use of "NTp κ =0" method for selection of unconverted photon candidates shows that for 90% photon efficiency π^0 reduction can be as high as 70% (Tables 1 and 2). Reduction significantly degrades to 29% for E_T = 60 GeV. In tables 1 and 2 the reduction for 90% photon efficiency in the ECAL barrel and endcap is shown.

E _T (GeV)	Редукција π^0 за 90% еф. фот.		
Lar(Ocv)	"NTrk=0"	"R9>0.93"	
20-25	70.0±1.4	69.9±1.4	
25-35	62.3±0.9	57.6±0.9	
35-45	43.4±1.0	44.9±1.0	
45-55	33.2±0.7	36.0±0.7	
55-65	29.3±0.6	30.5±0.6	

$E_T(GeV)$	Редукција π^0 за 90% еф. фот.		
	"NTrk=0"		
20-25	64.1±2.0		
25-35	62.1±1.3		
35-45	57.1±1.2		
45-55	54.4±1.2		
55-65	51.8±1.0		

Tables 1 and 2: Reduction of π^0 s in barrel a) and endcap b).

A better π^0 rejection performance is observed in most of the E_T in the ECAL Barrel region for the R9 > 0.93 method while this is reversed in the Preshower region.

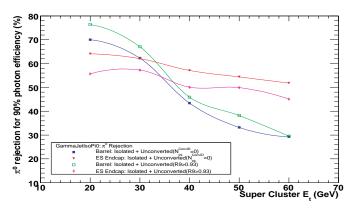
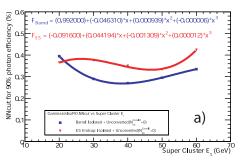


Image 10: π^0 reduction depending on Super Cluster E_T . The rejections is shown for both unconverted photon candidate selection methods and for ECAL barrel and endcap.

In figure 2.40 we summarize the π^0 rejection for 90% photon efficiency vs particle's super cluster E_T . Filled boxes represent the π^0 rejection in the ECAL Barrel region for isolated and unconverted particles using "NTrk = 0" method. Filled triangles represent the π^0 rejection in the Endcap Preshower (ES) region for isolated and unconverted particles using "NTrk = 0" method. Open boxes represent the π^0 rejection in the ECAL Barrel region for isolated and unconverted particles using "R9 > 0.93" method. Open diamonds represent the π^0 rejection in the Endcap Preshower (ES) region for isolated and unconverted particles using "R9 > 0.93" method.

In figure 2.41 we plot the NNcut vs. E_T . NNcut is the value of the NNoutput for which we have 90% photon efficiency and is produced using the efficiency/rejection plots that indicatively are shown in figures 2.38(b,d) and 2.39(b,d). In figure 2.41(a) the NNcut = $f(E_T)$ for "NTrk = 0" selection of unconverted photon candidates method is shown for both Barrel and ECAL Endcap Preshower (ES) region. Similarly, in figure 2.41 (b) the NNcut = $f(E_T)$ for "R9 > 0.93" method is shown.



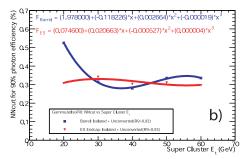


Image 11.: $NNcut = f(E_T)$ for barrel and endcap and for $,N_{trk}^{ConvID} = 0$ "method (a) and ,R9 > 0.93" method (b).

4.2. Trying to train the ANN from data

It is very challenging to find a way to train the ANN from real data samples, when they become available. This implies finding a source as pure as possible of photons and $\pi 0s$. Taking into account the stability of the ANN in contaminating samples (image 12), we propose the following method for trying to train the ANN from real data:

- Select events after applying the HLT2Photon trigger;
- apply of $E_T > 20$ GeV cut in both photon candidates;
- apply isolation to both photon candidates;
- select unconverted photon candidates with NTrk=0 (image 8) или R9>0.93 (image 9) method;
- exclude events with [110 GeV < M $\gamma\gamma$ < 150GeV] since in this mass region the Higgs signal is supposed to be;
- select a sample enriched in photons putting a cut in higher NNoutput values and a sample enriched in neutral π^0 s putting a cut in lower NNoutput values.

As an example we applied the method to CSA07 officially produced γ +Jet envents - "GammaJet" events. In figure 2.46 the NNoutput is plotted after the above cuts and when "R9>0.93" method used for selecting unconverted particles. Figure 2.45(a) refers to ECAL Barrel region: black solid line all photon candidates, green dashed line candidates matched to real photons, red dot line candidates matched to real π^0 s while (c) refers to the Endcap ECAL ES region with the same color and style notation as in (a). In (b) and (d) the corresponding purity of photons (blue boxes) and π^0 s (red triangles) is calculated from figures 2.46(a) and (c) respectively. From these last plots an 80% - 90% pure photon sample can be selected for 0.7 < NNoutput< 0.9 while a π^0 sample of the same purity with NNoutput< 0.1. The distributions 2.45(a), (c) are normalized to L = 1fb⁻¹ Thus, the stability of the ANN to contamination, can justify that real data selected with the procedure described above could be used to train the ANN in an iterative procedure.

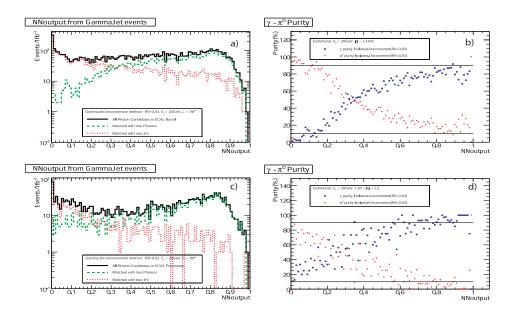


Image 12: NNoutput of photon candidates from "GammaJet" events after HLT2Photon trigger, $E_T > 20$ GeV cut in both photon candidates, isolation and selection of unconverted particles using "R9>0.93" method. The photon candidates with 110GeV/ $c^2 < M\gamma\gamma < 150$ GeV/ c^2 ere not taken into account since in this mass region the Higgs signal is supposed to be. (a) ECAL Barrel region: black solid line all photon candidates, green dashed line candidates matched to real photons, red dot line candidates matched to real π^0 s, (b) the purity of photons (blue boxes) and π^0 s (red triangles) calculated from (a),(c)Endcap ECAL ES region: black solid line all photon candidates, green dashed line candidates matched to real photons, red dot line candidates matched to real π^0 s,(d) the purity of photons (blue boxes) and π^0 s (red triangles) calculated from (c). The distributions (a), (c) are normalized to L = 1fb⁻¹

5. The use of photon background reduction in official Higgs to two photons decay analysis

The performance of algorithm for photon background reduction was tested on few channels of Higgs creation and on both: irreducible background, coming from direct two photon production, and reducible background coming from γ +jet and multi jet events [5]. This preliminary analysis of use of photon background reduction uses only algorithm for unconverted photon candidates already implemented in CMSSW, while the part of algorithm for converted photon candidates still was not migrated from ORCA to CMSSW.

Three ways of algorithm implementations were tested in cut based analysis:

- A) For **unconverted** photon candidates passing the cut analysis (Et cut: $E_{T_1}>40$ GeV μ $E_{T_2}>35$ GeV) the cut on NNoutput >0.1 value was applied.
- **B)** For **all** photon candidates which don't pass cut based analysis, the weaker cuts on E_T are applied: $E_{T1} > 20$ GeV μ $E_{T2} > 20$ GeV the cut on NNoutput > 0.1 was applied.
 - C) both, A) and B) simultaneously.

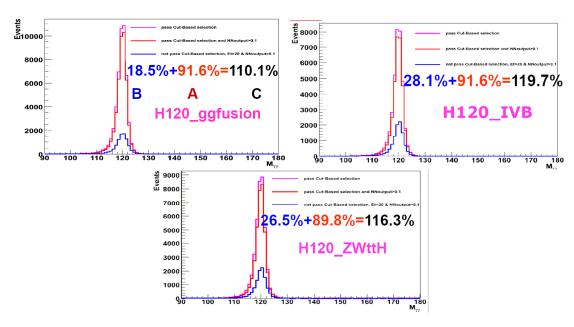


Image 13: A,B and C way of implementation of photon background reduction algorithm in cut based analysis for Higgs to two photon decay channel.

Results using a way B give an increase of signal by 20% with events with photon candidates with lower transverse energy. The result of application of way C show overall increase of two-photon signal by 11.5%. Image 13 shows an increase of two-photon signal for various (SM) Higgs creation channels and summed result of all Higgs creation channels gives mentioned signal increase of 11.5%.

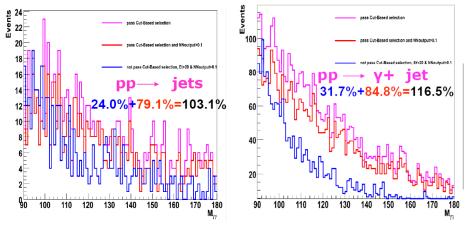


Image 14: A, B and C way of application of reduction algorithm in cut based analysis for reducible background of two-photon Higgs decay signal.

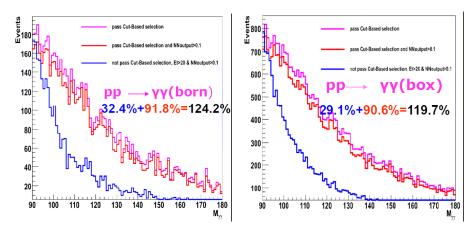


Image 15: A, B and C way of application of reduction algorithm in cut based analysis for irreducible background of two-photon Higgs decay signal.

All three ways were, naturally, tested for reducible (Image 14) and reducible background (Image 15) in two photon candidate's energy sum interval {119 GeV, 121 GeV}. Table 3 shows the change of signal significancy (S/\sqrt{B}) defined as:

$$(N+N_{add})_{sig}/\sqrt{(N+N_{add})_{bg}}$$
 (2.16)

for four different type of background, depending on a way of reduction algorithm application. With B and C way it can be noticed the increase of significancy, and the biggest increase can be seen with background coming from multiple jets.

	$(N+N_{add})_{sig}/\sqrt{(N+N_{add})_{bg}}$					
	pp→jets	pp→ γ+jet	$pp \rightarrow \gamma \gamma (born)$	$pp \rightarrow \gamma \gamma (box)$		
В	1.08	1.05	1.04	1.06		
C	1.10	1.03	1.00	1.02		

Table 3: The change of significancy depending on a way of application of background algorithm in cut based analysis.

These three tests represent the First (not the optimal) attempt for implementation of photon reduction algorithm into analysis based on cats, and as a result they give small but not insignificant change of two photon signal significance (S/\sqrt{B}). Inclusion of converted photon candidates will certainly increase the significance. Work on inclusion of complete discrimination algorithm (both unconverted and converted photon candidates) is still ongoing.

6. Conclusion

The main subject of my thesis is in connection with the research of properties and phenomena of two photon Higgs decay acording to the Standard Model (SM).

This thesis shows that the simulation results developed for Preshower prototype test during 2004 are in excellent agreement with the experimental results.

It is also shown that the use of algorithm for separation of neutral pions and photons based on information from ECAL and Preshower detector give better results than previous methods.

Finally, it should be mentioned that the developed algorithm was implemented in CMS reconstruction software – CMSSW and that it is continuously updating and adapting for official analysis of two photon Higgs decays on CMS detector.

Literature

- [1] D.Barney, W.Bialas, P.Kokkas, N.Manthos, D.Maletic, I.Papadopoulos, A.Peisert, S.Reynaud, P.Vichoudis. *Detection of muons at 150 GeV/c with a CMS Preshower Prototype*. CMS NOTE-2006/046, 2006.
- [2] Geometry/EcalCommonData/src/esalgo.xml
- [3] Geometry/EcalCommonData/src/DDEcalPreshowerAlgo.cc
- [4] D. Kirkby , " π^0/γ Discrimination in L3 Using Shower-Shape Analysis", L3 Note 1816
- [5] Z.Zhang, D, Maletic et al. $H\rightarrow gg$: Results on CMSSW validation of cut-based analysis + 1st look at use of π^0 discriminator Egamma group 21.04.2008.