UNIVERSITY
of
GLASGOW

Department of Physics & Astronomy
Experimental Particle Physics Group
Kelvin Building, University of Glasgow,
Glasgow, G12 8QQ, Scotland
Telephone: +44 (0)141 339 8855 Fax: +44 (0)141 330 5881

GLAS-PPE/2005-02
9th March 2005

Signal "spot" and mass window cut for the low mass SM Higgs

A. Cheplakov, R. StDenis, A. S. Thompson

University of Glasgow, Glasgow, G12 8QQ, Scotland

Abstract

An attempt to improve signal significance for the Standard Model low mass Higgs boson produced in $WH(H \to b\bar{b})$ channel in ATLAS experiment [1] is described in the Note. The most significant background processes, WZ, W+jets and $t\bar{t}$, are considered. The approach exploits the condition for the minimal angular separation in two-particle decays.

The analysis was not aimed to get the final number of the Higgs over the background, but to examine some new kinematical cuts and check the potential improvements.

ATLAS experiment Note ATL-PHYS-PUB-2005-004

1 Introduction

The resonant WZ ($Z \to b\bar{b}$) and continuum backgrounds from W+jets and $t\bar{t}$ dominate Higgs signal from the WH ($H \to b\bar{b}$) over the whole range of the invariant jet+jet mass of two tagged b-jets [1]. The technique is proposed to find the range of some kinematical variables (signal "spot") where signal events are more "visible" over the huge background rates. It may help to suppress the continuum background and separate signal from another "spots" of the resonant processes.

For a particle decay $0 \to 1+2$ the equations $\vec{p_\circ} = \vec{p_1} + \vec{p_2}$ and $E_\circ = E_1 + E_2$ give us the expression for the value of $cos\theta_{1,2}$, where $\theta_{1,2}$ is the opening angle between the particle 1 and 2 (these are two b-jets in our analysis):

$$cos\ \theta_{1,2} = \frac{E_1 \cdot E_2 - (m_\circ^2 - m_1^2 - m_2^2)/2}{|\vec{p_1}| \cdot |\vec{p_2}|}$$

or

$$cos\ \theta_{1,2} = \frac{E_1 \cdot (E_\circ - E_1) - (m_\circ^2 - m_1^2 - m_2^2)/2}{\sqrt{E_1^2 - m_1^2} \cdot \sqrt{(E_\circ - E_1)^2 - m_2^2}}.$$

If a heavy particle is decaying into the light ones

$$m_\circ \gg m_1, m_2 \tag{1}$$

then

$$cos\ \theta_{1,2} = 1 - \frac{m_\circ^2}{2 \cdot E_1 \cdot (E_\circ - E_1)}$$

and we get a well known condition for the minimum angular separation in terms of $cos\ \theta_{1,2}$:

$$cos\ \theta_{1,2} \leq 1 - \frac{2m_\circ^2}{E_\circ^2} \tag{2}$$

or (equivalently),

$$sin(\theta_{1,2}/2) \geq \frac{m_\circ}{E_\circ}. \tag{3}$$

The requirement (1) can be applied to the Higgs decay into the b-quarks. For the signal channel the decay products (b-jets) have to concentrate around the threshold values (2) or (3), determined by the Higgs mass $m_\circ$. This is not the case for the continuum background and thus can be used for the background suppression.

One may also consider the 2-dimensional plot of $cos\ \theta_{1,2}$ versus $2/E_\circ^2$ or $sin(\theta_{1,2}/2)$ versus $1/E_\circ$ and analyse signal and background events taken from the area of the signal "spot". The range of the "spot" in these plots could be shrunk according to the value of the required signal acceptance (for high purity sample), and the difference in masses of Z and H bosons may be used for Z suppression by varying the value of $m_\circ$ as a cut parameter. Due to an additional dimension (sum of jet energies) the signal "spot" approach could be more effective for signal extraction than the commonly used inclusive kinematical cuts (e.g. mass window cut).

2 Monte Carlo events generation

The Monte Carlo generator PYTHIA 6.157 [2] and the CTEQ5L [3] set of the parton structure functions have been used in the analysis. The events were simulated taking into account the response of the ATLAS detector by means of the fast simulation package ATLFAST 2.53 [4].

To estimate better the advantages of the proposal we have followed in general the procedure described in details in papers [5], [6] and [7]. In particular, we used the "standard" (within the ATLAS experiment) criteria for the event selection [5]:

- the threshold of 20 GeV/c for p_T of isolated electron and muon, both with $|\ \eta\ | < 2.5$, events with a second lepton within the same interval of pseudorapidity and with $p_T > 6$ GeV/c were vetoed;
- 15 GeV/c threshold to the transverse momenta used for the jet reconstruction in the fixed cone size of $\Delta R_{cone} = 0.6$;

- only two jets ($b-tagged$ jets) in the event are allowed, events with additional jets of any type in the pseudorapidity interval $\mid \eta \mid < 5$ and with $p_T > 30$ GeV/c were vetoed.

The jets reconstruction was performed in ATLFAST by means of the fixed cone algorithm with default settings and $\Delta R_{cone} = 0.6$. Several parameterizations for the jet energy calibration have been developed in the ATLFAST framework (see [8], [9], [10]). We have attempted to calibrate the low p_T jets more accurately and included a new set of *p.d.f.* (CTEQ5L). More details about the new parameterization may be found in [11].

2.1 Signal and background processes

Some PYTHIA parameters used in the event generation, the corresponding production cross sections for each process, the Monte Carlo statistics and the number of events which passed the selection criteria are presented in Table 1. At this stage the p_T-dependent b-tagging efficiency from the atlfast-b package [8] was applied.

Table 1: *PYTHIA parameters, corresponding production cross sections and Monte Carlo statistics for signal and background processes.*

Process (Higgs mass)	PYTHIA parameters	Cross section [pb]	#events generated	#events selected
WH (90)	MSUB(26)=1	0.53	10M	764K
WH (100)	H→ $b\bar{b}$, W→ $e\nu/\mu\nu$	0.39	10M	812K
WH (110)		0.28	10M	847K
WH (120)		0.20	10M	874K
WH (130)		0.12	10M	897K
WH (140)		0.07	10M	913K
WH (150)		0.03	10M	927K
WZ	MSUB(23)=1	0.88	100M	6.4M
	Z→ $b\bar{b}$, W→ $e\nu/\mu\nu$			
$t\bar{t}$	MSEL=6	168.1	200M	2.6M
	$t \to Wb, W_1 \to e\nu/\mu\nu, W_2 \to hadrons$			
W+jets	MSUB(16)=1, MSUB(31)=1			
	W→ $e\nu/\mu\nu$			
	CKIN(3,4)=(10,30)	13305	962M	153K
	CKIN(3,4)=(30,50)	3097	500M	155K
	CKIN(3,4)=(50,100)	1722	393M	168K
	CKIN(3,4)=(100,200)	322	295M	163K
	CKIN(3,4)=(200,–1)	33	200M	133K

The samples of events presented in the last column of the Table were studied, and we assume 100% acceptances for all the processes from this starting point. In Table 2 the corresponding cumulative acceptances are shown together with the number of events expected for the integrated luminosity of 30 fb^{-1} (in Nov'2009?). A significant background suppression is required to extract the Higgs signal.

In the earlier papers [6], [7] in order to estimate the signal significance the mass window around the peak position was applied to the invariant mass spectra of two b-tagged jets. The number of selected events within (100±20) GeV/c^2 for WH(100) and within (120±24) GeV/c^2 for WH(120) process and backgrounds are presented in Table 2 for the reference analysis and for our study. An acceptable agreement within ∼12% is observed for almost all processes, except WZ and $t\bar{t}$. The difference for WZ could be due to modification of the jet reconstruction parameters: we used the value of $\Delta R_{cone} = 0.6$ and got a smaller width of the resonant peak. A

Table 2: *A cumulative acceptance of signal and background channels and event numbers expected in ATLAS for the integrated luminosity of 30 fb^{-1}. Lepton identification (90%) and p_T-dependent b-tagging efficiencies are assumed. The last two columns compare the number of events expected in this study and in the reference analysis ([5],[7]) within the mass windows* $m_H = (100 \pm 20)$ *GeV/c^2 for WH(100) and (120±24) GeV/c^2 for WH(120) and background processes.*

Channel	Trigger lepton	Two b-jets	Jet veto	#events expected	within mass window	Reference analysis
WH (90)	61%	10%	8%	1103		
WH (100)	62%	10%	8%	856	678	605
WH (110)	63%	11%	8%	648		
WH (120)	64%	12%	9%	464	364	325
WH (130)	65%	12%	9%	298		
WH (140)	66%	13%	9%	163		
WH (150)	67%	14%	10%	70		
WZ	57%	8%	6%	1535	262	325
$t\bar{t}$	26%	6%	0.6%	59878	14967	10500
W+jets	56%	0.03%	0.02%	108099	14653	13100

new set of *p.d.f.* could also explain the difference for $t\bar{t}$ background. We note that we have about 20% more background events than in the reference analysis, so one may expect more "pessimistic" results for the Higgs signal significance. More detailed comparison with the mass window cut will be presented in the last section of the paper.

3 Search for the optimal kinematical cuts

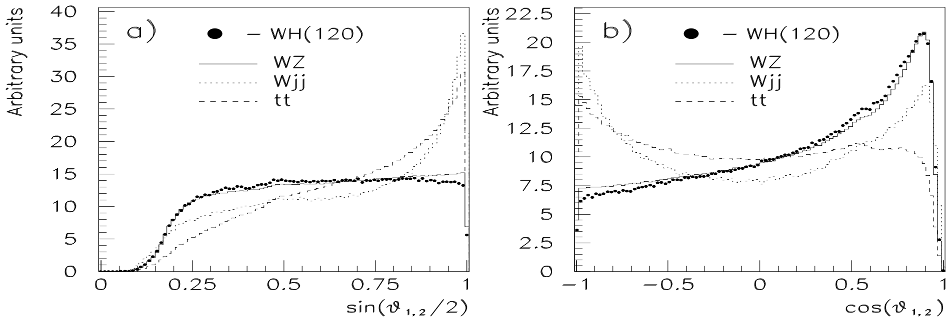

Figure 1: *Distributions on the value of (a) sin($\theta_{1,2}/2$) and (b) cos($\theta_{1,2}$) for b-tagged jets produced in the WH (m_H=120 GeV/c^2), WZ, W+jets and $t\bar{t}$ channels.*

It is interesting to compare the kinematics for signal and background processes in terms of the variables presented in (2) and (3). In Fig.1 the distributions on the value of $\sin(\theta_{1,2}/2)$ and $\cos(\theta_{1,2})$ are shown for pairs of b-tagged jets selected in the signal (m$_H$=120 GeV/c^2) and background channels. For better presentation the spectra are normalized to the same value. Some cut values for sine (e.g. $\sin(\theta_{1,2}/2) < 0.9$) or cosine (e.g. $\cos(\theta_{1,2}) > -0.5$) of the opening angle could be extracted from these spectra in order to decrease the W+jets and $t\bar{t}$ backgrounds, but not for the resonant WZ process, which is very close to the signal spectrum. In the corresponding contour plots, shown in Fig. 2 for the illustration, the event rates are presented for WH (with different masses m_H=100, 120 and 140 GeV/c^2), WZ, W+jets and $t\bar{t}$ processes. The top part of the figure demonstrates distributions on the value of $\sin(\theta_{1,2}/2)$ versus $1/(E_1 + E_2)$ (SIN-plots), and the bottom part shows distributions on the value of $\cos(\theta_{1,2})$ versus $2/(E_1 + E_2)^2$ (COS-plots), where $\theta_{1,2}$ is the opening angle and E_1, E_2 are the energies of two

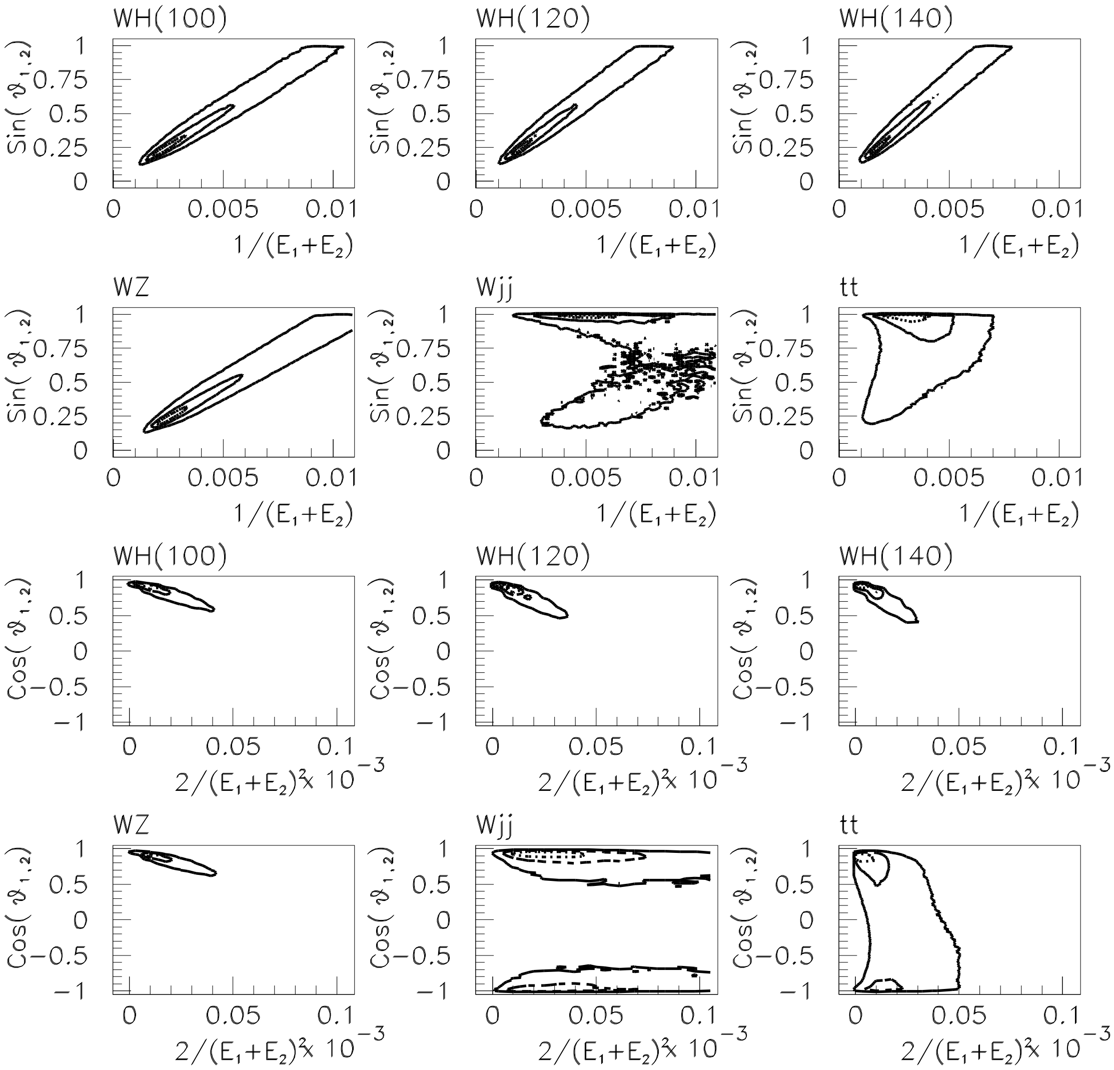

Figure 2: *Top part of the figure shows distributions on the value of sin($\theta_{1,2}/2$) versus* $1/(E_1 + E_2)$ *(in* $(GeV/c^2)^{-1}$*), and the bottom part presents distributions on the value of cos($\theta_{1,2}$) versus* $2/(E_1 + E_2)^2$ *(in* $(GeV/c^2)^{-2}$*), where* $\theta_{1,2}$ *is an opening angle and* E_1, E_2 *are the energies of two b-tagged jets produced in the signal WH process for different Higgs masses and in the background WZ, W+jets and* $t\bar{t}$ *channels. The most occupied areas are outlined: the dotted line corresponds to 20%, dashed line - 50% and solid line - 80% of the histogram integral contents.*

b-tagged jets. The most populated areas are outlined: the dotted line corresponds to 20%, dashed line - 50% and solid line - 80% of the histogram integrated contents. The plots are quite different for the resonant and continuum channels. We also note the contour displacement for the resonant decays for various masses. These facts we shall use for the further optimization of the kinematical cuts.

3.1 An attempt to separate WH and WZ processes

Although the m_H and m_Z masses are close, we used the $m_\circ$ mass parameter in (2) and (3) as a slope value for the cut-lines (either $\sin(\theta_{1,2}/2) = m_\circ/(E_1 + E_2)$ or $\cos(\theta_{1,2}) = 1 - 2\, m_\circ^2/(E_1 + E_2)^2$) to separate the areas where H or Z decays predominate.

As an example, the comparison of the SIN- and the COS-plots for the b-jets produced in WH (m_H=120 GeV/c^2) and WZ processes are presented in Fig. 3 a) and Fig. 3 b) respectively. The contour plots show the event rate levels for signal (WH, thick lines) and background (WZ, thin lines) processes for 20%, 50% and 80% level of the histogram contents. The cut-lines for $m_\circ$ = 110 GeV/c^2 are superimposed. For this particular value of $m_\circ$ we would have about 72% of signal WH(120) events and below 18% of background events from WZ, W$b\bar{b}$ and

W+jets. Fig. 3 c), d) show the acceptance values for various Higgs masses (100, 120 and 140 GeV/c^2) and background processes as a function of parameter $m_\circ$.

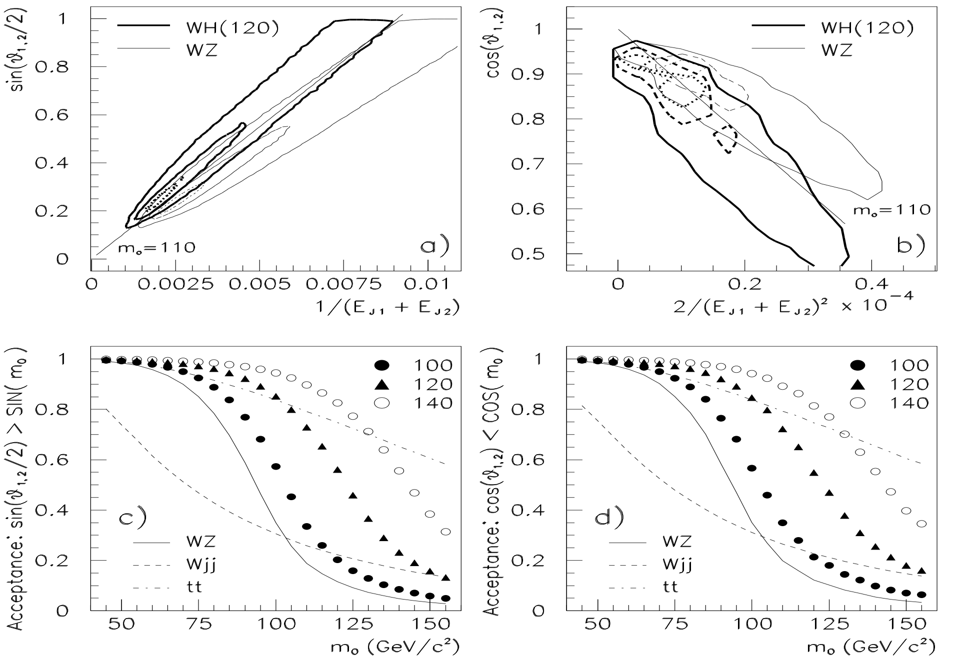

Figure 3: *Contour plots for WH(120) (wide lines) and WZ (thin lines) event rates for (a) SIN- and (b) COS-plots. The cut-line correspondent to the slope parameter $m_\circ$=110 GeV/c^2. Two bottom figures show signal and background acceptance as a function of the slope $m_\circ$. The results are presented for different Higgs masses for (c) SIN- and (d) COS-plots.*

In Fig. 4 the invariant mass spectra of two b-tagged jets expected in 3 years in ATLAS (corresponding to an integrated luminosity of 30 fb^{-1}) for WZ and WH are presented for different values of m_H and various values of $m_\circ$. It is seen that $m_\circ$ could be used as an effective cut-parameter for suppression of the irreducible WZ resonant background. By losing an acceptable number of signal events one can significantly improve the "purity" of the selected sample. It also allows the suppression of the considerable contribution from the W+jets processes. This parameter is not effective for $t\bar{t}$ background and for WH/WZ separation with low mass m_H close to m$_Z$.

3.2 Signal acceptance level as a cut parameter

As seen from Fig. 3 the use of a cut on $sin(\theta_{1,2}/2)$ decreases the contribution from background processes. But they still remain significant due to the high production cross sections.

The continuum background almost completely occupies the plots in Fig. 2. But it is possible to decrease the relative contribution of the background by preferentially selecting the events from the area occupied by Higgs decays, i.e. applying a 2-dimensional cut. The area selection could be performed by excluding the less occupied cells from the SIN- or COS-plots for the Higgs decays until the required signal acceptance value is obtained.

In Fig. 5 a)-c) the remaining fractions of various background events estimated from the SIN-plots, are presented as a function of the signal acceptance for different Higgs masses - 100, 120 and 140 GeV/c^2. In Fig. 5d) the fraction of background events for a sum of the background processes is shown as a function of signal acceptance. It is seen that with no loss of signal events one may remove a considerable fraction of the potential background (up to 50-60% for W+jets and 70-80% for $t\bar{t}$, depending on m_H). The background suppression could be more significant if the selected area of kinematical variables is shrunk due to the required value of the signal acceptance (for 80% of signal acceptance one may remove 30-80% of the resonant and 70-85% of continuum background, depending on m_H). In general, we have about 25% of the initial background rate for a sample of 80% of the signal events and this depends weakly on the Higgs mass.

Summarizing the possible applications of kinematical cuts for the background suppression:

- the value of $m_\circ$ parameter within (80÷130) GeV/c^2 is effective against the WZ resonant background for

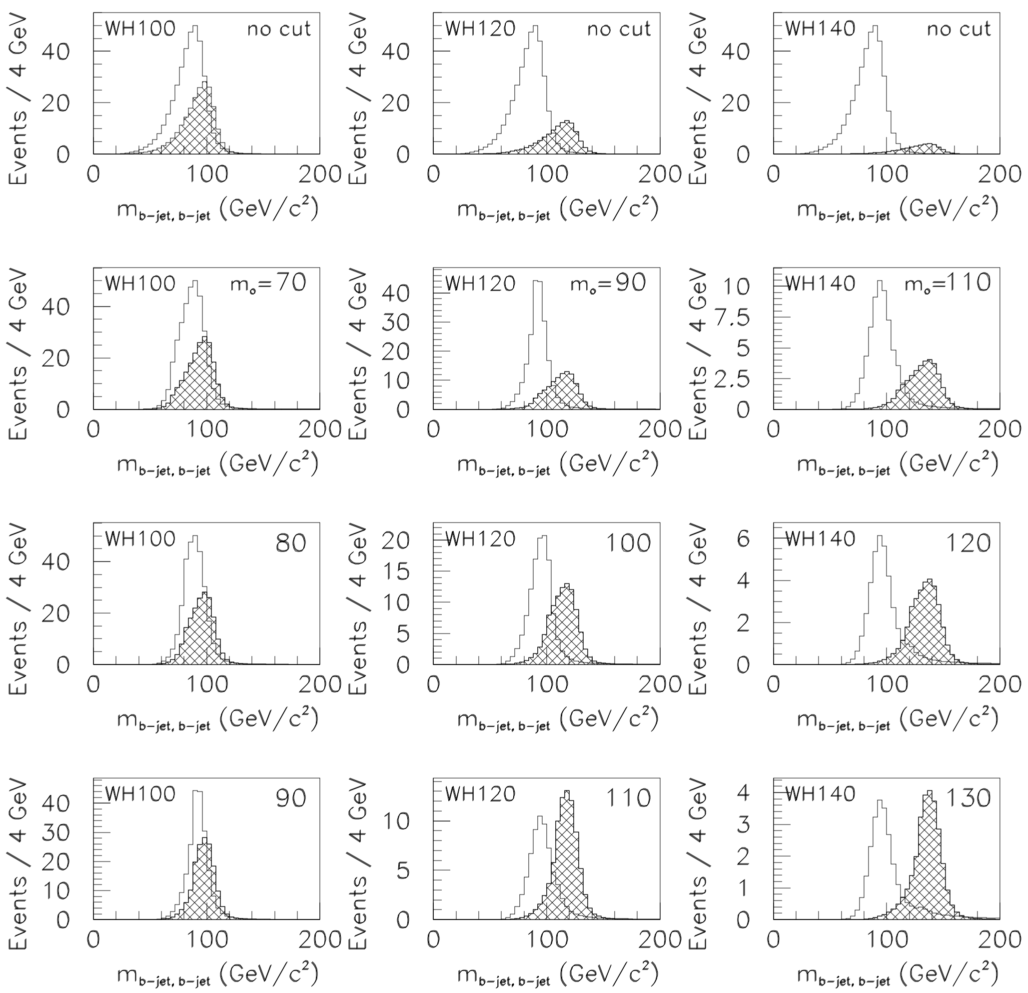

Figure 4: *Comparison of the $M_{b\bar{b}}$ spectra for WZ (open) and WH (hatched area) processes at integrated luminosity of 30 fb^{-1}. The Higgs masses of 100, 120 and 140 GeV/c^2 have been used. The results are shown for different values of the slope parameter $m_\circ$ (see (2)).*

m_H >100 GeV/c^2;

- the cut $sin(\theta_{1,2}/2) < 0.9$ could remove some fraction of W+jets processes, but is not effective for WZ and $t\bar{t}$ events;

- the signal "spot" in the plot of kinematical variables $sin(\theta_{1,2}/2)$ versus $1/(E_1+E_2)$, which one may outline based on the required value of the signal acceptance, is an effective way to suppress various background processes to get a "pure" sample of signal events.

3.3 An optimal mass window cut

The actual values for the cut parameters can be optimized in order to get the highest signal significance. Unlike for the 1-dimensional mass window cut, we have collected signal and background events from the 2-dim "spot" in the SIN-plot: the less occupied cells have being excluded from the signal plot until the required acceptance is obtained. This allowed us to outline the "valid" region of the plot and collect signal and background events only from that area (so, S/$\sqrt{B}$ seems to be applicable to estimate signal significance). Several sets of two b-jet events have been obtained for different values of the acceptance cut from 0.9 to 0.1. Then the corresponding invariant mass spectra for each set of events have been analysed using the varying mass window cut.

An example of the invariant mass spectra for signal and background processes is presented in Fig. 6 for the cut values of 0.75, 0.6 and 0.3 together with the initial spectra (no acceptance cut). The spectra are normalized

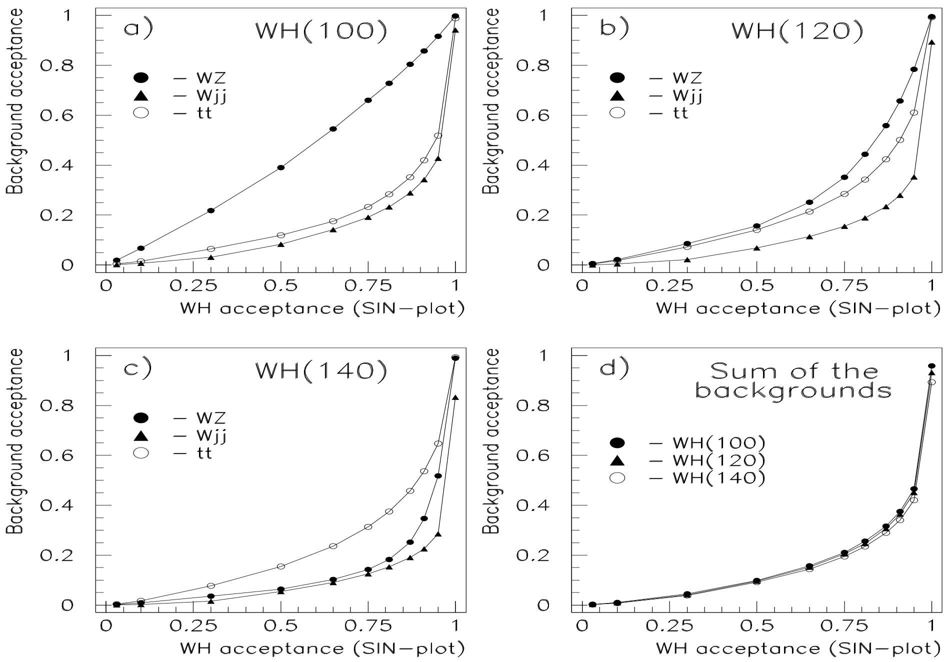

Figure 5: *Fraction of remaining background events for the areas of kinematical variables* $sin(\theta_{1,2}/2)$ *versus* $1/(E_1 + E_2)$ *(SIN-plot) defined by the required level of the WH signal rates as a function of the signal acceptance. The results are presented for various background processes and for sum of WZ, W+jets and* $t\bar{t}$ *backgrounds for different Higgs masses (100, 120 and 140 GeV/*c^2*).*

to the expected luminosity of 30 fb^{-1}. Lepton identification (90%) and b-tagging efficiencies are taken into account. Although the number of background events are reduced significantly due to the signal acceptance cut, the background level at the region of a signal peak remains high. One can see from Fig. 6 a) that the peak position of the signal distribution has moved slightly to higher masses for a small value of the acceptance cut, but this effect is not significant for the acceptance values higher than 0.5. The background spectra are more affected by the event selection: *ttbar* events tend to peak at the same area of the invariant mass (see Fig. 6 d), but the dominating contribution from Wjj processes (Fig. 6 c) remains much wider.

Uncertainties in reconstruction of the jet energy and direction will result in spreading of the signal "spot" and reduction of the number of signal events for the same values of the acceptance cut. This effect will be more significant when a smaller cone size will be used in the reconstruction algorithm. We have used an "optimal" value of 0.7 [11], giving 10% smaller width for mass of the top, and we expect that the spread of the background spectra will be more significant due to the systematics. Nevertheless, a better knowledge of the background shape is required to fit the experimental spectra and extract signal of the Higgs boson.

In Fig. 7 a) the invariant mass spectra for the initial sample of signal (m_H=120 GeV/c^2) and background events are shown by solid line, and those obtained for the 75%-acceptance cut are presented by the dashed line.

We have studied the invariant mass spectra in more details. The size of a mass window cut for the analysis of the invariant mass spectra could be varying based on the required fraction of the remained signal events. We have varied the fraction from 0.1 to 1.0 and for each value the mass window around the peak position was selected by excluding the less populated bins at the tails of the mass spectra so that the remaining part of the spectra contains the required fraction of the signal events. The same mass window was used then to count for the number of background events. The results for this "variable" cut are presented in Fig. 7 b)–d) as a function of the accumulated signal acceptance value, which is the product of two cuts - signal acceptance cut used to get a set of events (signal "spot") and the fraction of signal events inside the varying mass window.

Fig. 7 b) illustrates how effective the varying mass window could be used for the background suppression. It shows the suppression efficiency as a function of the accumulated signal acceptance. The solid line represents results for the initial mass spectra and dots shows efficiency obtained for the invariant mass spectra for the events taken from the signal "spots" constrained by the shown acceptance cut values. It is seen that the results

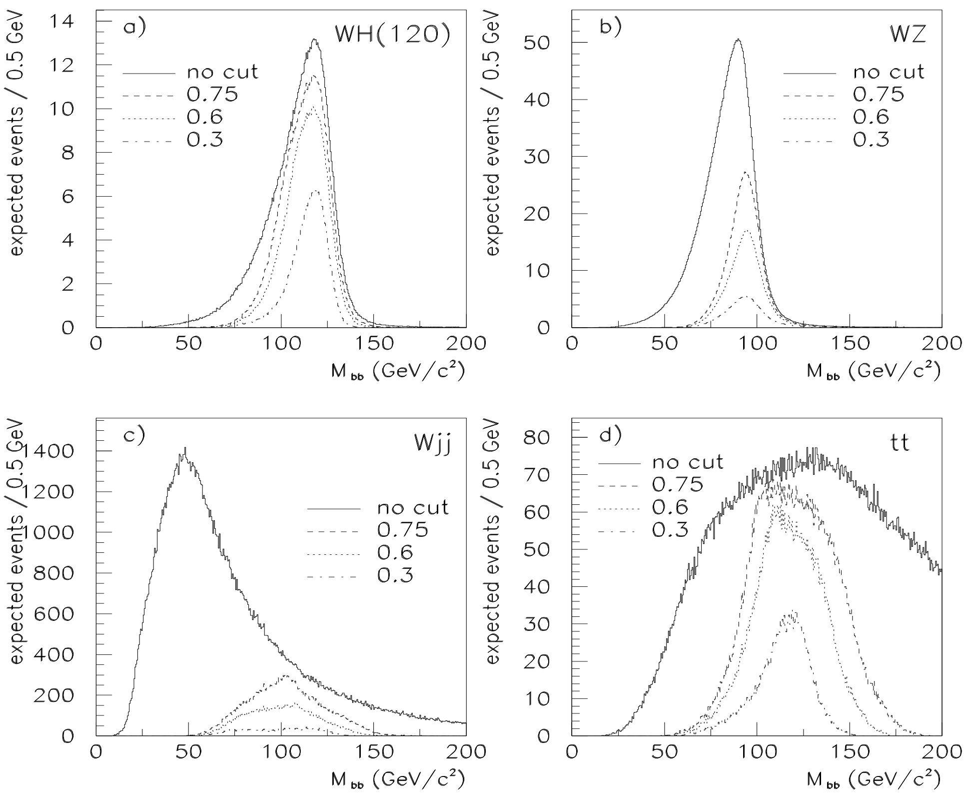

Figure 6: *The invariant mass spectra of two b-tagged jets for WH and background events for Higgs mass of m_H= 120 GeV/c^2. Initial sample of events (solid lines) is compared with those selected by the acceptance cut parameter of 0.75, 0.6 and 0.3. Spectra are shown for (a) WH, (b) WZ, (c) Wjj and (d) $t\bar{t}$ processes.*

of the "signal spot approach" are similar to those for the variable mass window cut for high values of signal acceptance, but the method reveals some advantages for a smaller signal fraction of ~30% and below. It is more evident from the Fig. 7 c) where the signal to background ratio is presented as a function of the accumulated signal acceptance. For m$_H$=120 GeV/c^2 the signal "spot" allows us to get signal to background ratio about 6% (for 30%-set of signal events) and higher when combining the signal acceptance cut and the varying mass window. For m$_H$=100 GeV/c^2 the ratio S/B is about 8% for signal acceptance below 30%.

The signal "spot" approach and variable window cut again give similar results for signal significance for high values of signal acceptance as shown in Fig. 7 d). The maximal value is $\sim$ 3.0 if no initial cut is applied and it is by 8% higher if the events are selected from the "spot" for the signal acceptance cut of 0.75. We also note a 50% gain when comparing this result with the symmetrical mass window cut from the previous analysis [7]: the estimated signal significance was about 2.0 – 2.1 for m$_H$=120 GeV/c^2 for the mass window of (120±24) GeV/c^2.

This analysis was performed for various Higgs mass values from 90 to 150 GeV/c^2. The results for the signal significance are presented in Fig. 8, as a function of Higgs mass, by black dots. The triangles show the signal significance expected earlier in [7] and [5] for m$_H$=100 and 120 GeV/c^2. It is seen that the signal "spot" approach together with the variable mass window cut gives better values.

4 Conclusions

The results presented in the Note show an additional possibility of the background suppression and improvement of the signal significance for the Standard Model low mass Higgs produced in WH (H→ $b\bar{b}$) process in ATLAS.

An opening angle and energy of jets from the Higgs decay have been used to develop a signal "spot" approach for the event selection. The mass parameter m$_\circ$ in (2) and (3) can be applied for the resonant (WZ) background suppression, whereas the region of "valid" kinematical parameters defined for the specific value of the signal acceptance (a signal "spot") can be used for the suppression of the continuum background (like W+jets and

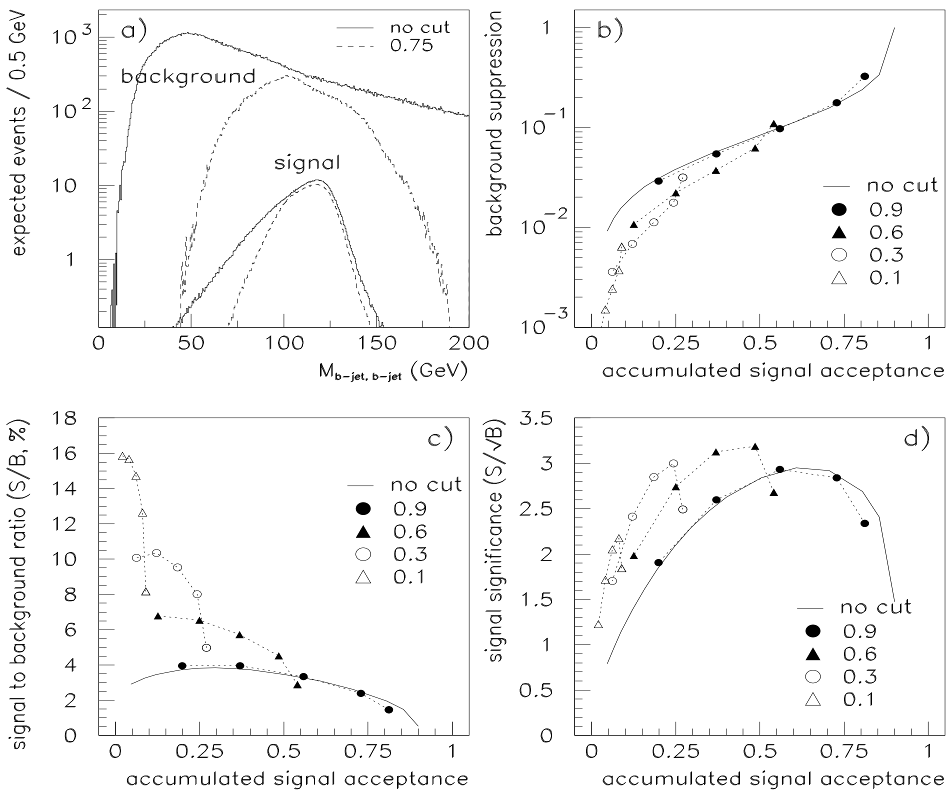

Figure 7: *(a) The invariant mass spectra of two b-tagged jets for WH and background events for Higgs mass of m_H= 120 GeV/c^2. Initial sample of events (solid lines) is compared with the one selected by the acceptance cut parameter of 0.75. (b) The background suppression as a function of the accumulated signal acceptance. The results for different signal "spots" are shown by the dots and compared with the results obtained for the "unbiased" sample. (c) The signal to background ratio and (d) signal significance as a function of the accumulated signal acceptance. The solid line represents the results for the "unbiased" sample, and the dots show the results for the events taken from the various signal "spots".*

$t\bar{t}$).

The asymmetrical mass window cut for the jet+jet invariant mass spectra was optimized for ~70-80% of the total number of signal events. The signal significance measured as S/$\sqrt{B}$ has been found to be ~5.9 for m$_H$=100 GeV/c^2 and ~3.0 for m$_H$=120 GeV/c^2. We do not expect that the missed single-top background could significantly change these values. Also when compare with the previous results one has to note that we have used another version of PYTHIA with different set of p.d.f. and a cone size for the jet reconstruction [11].

When combined with a signal "spot" the optimal mass window cut can provide a higher "purity" of the selected sample of events. This could be useful for the Higgs spin study.

Acknowledgements
We would like to thank Daniel Froidevaux for kind support and patient discussions.

References

[1] ATLAS Collaboration, Technical Proposal, CERN/LHCC/94-43 (1994);

[2] T. Sjöstrand, CERN preprint CERN-TH.7111/93 and CERN-TH.7112/93;

[3] CTEQ Collaboration $hep-ph-9903282$;

[4] E. Richter-Was et al., ATLAS Note ATL-PHYS-98-131 (1998);

[5] ATLAS Collaboration, ATLAS Detector and Physics Performance Technical Design Report, CERN/LHCC/99-14, ATLAS TDR 14 (1999);

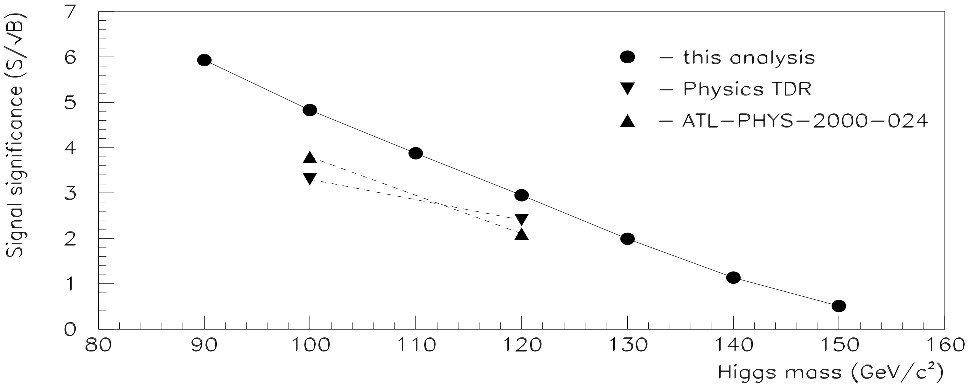

Figure 8: *The signal significance as a function of the Higgs mass. Black dots show the results of our analysis. The data from [7] and[5] are presented by triangles. Lepton identification (90%) and b-tagging efficiencies are included.*

[6] D. Froidevaux and E. Richter-Was, ATLAS Note ATL-PHYS-94-043 (1994); Z.Phys.**C67** (1995) 213

[7] E. Richter-Was, ATLAS Note ATL-PHYS-2000-024 (2000);

[8] see http://atlasinfo.cern.ch/Atlas/GROUPS/PHYSICS/HIGGS/Atlfast.html

[9] P.Brückman and E. Richter-Was, ATLAS Note ATL-COM-PHYS-2002-025 (2002);

[10] M. David et al., ATLAS Note ATL-PHYS-2002-007 (2002);

[11] A. Cheplakov, R. StDenis, A. S. Thompson, ATLAS Note ATL-PHYS-2003-010 (2003)