Measurement of $\sigma(e^+e^- \to HZ)$×Br$(H \to ZZ^\star)$ at the 250 GeV ILC

Evgeny Antonov (LPI, MEPhI), Alexey Drutskoy (LPI, MEPhI)

SW/ANA meeting, July 14, 2021

Content

- Introduction
- MC samples and analysis tools
- Event preselection procedure
- Analysis results for all channels
- Comparison of combined fit and toy MC
- Combined signal significance estimate
- Conclusions

Introduction

The width of the Higgs boson *is difficult to measure at LHC in a model-independent approach* (the uncertainty *~20% after luminosity upgrade*)

We propose to use the process $e^+e^- \to HZ$ with the subsequent decay $H \to ZZ^\star$ to measure:

$$\sigma(e^+e^- \to HZ) \times Br(H \to ZZ^\star) = C \cdot g_Z^4 / \Gamma_H$$

Constant, Error < 1% expected

Coupling HZZ Error < 0.5% expected

Higgs boson width

One of Z bosons is reconstructed in jets.

$Z \to jj \text{ or } ll, \quad Z^\star \to ll \text{ or } jj$

MC samples and analysis tools

The following sub-processes are studied:

Channel 1: $e^+e^- \to Z_1(j_1j_2)\,H, \quad H \to Z(j_3j_4)Z^\star(\ell_1\ell_2)$

Channel 2: $e^+e^- \to Z_1(j_1j_2)\,H, \quad H \to Z(\ell_1\ell_2)Z^\star(j_3j_4)$

Channel 3: $e^+e^- \to Z_1(\nu\bar{\nu})\,H, \quad H \to Z(j_1j_2)Z^\star(\ell_1\ell_2)$

Channel 4: $e^+e^- \to Z_1(\nu\bar{\nu})\,H, \quad H \to Z(\ell_1\ell_2)Z^\star(j_1j_2)$

We used the official MC data samples produced by the ILD group with:

1. Generation with Whizard 2 package
2. LCIO output format
3. Hadronization is performed by Pythia6
4. Simulation of the ILD detector (ILD_l5_o1_v02 model used)
5. ILCSoft v02-00-02 (DD4Hep, MarlinReco)
6. 100% beam polarization $\mathcal{P}_{e^-e^+} = (\mp 1.0, \pm 1.0)$
7. ISR and beam radiation processes are included
8. γγ beam induced processes are overlaid

Event preselection

Preselection tools:

1. Extraction only specific process and decay chains on *MCParticle* level.
2. Identifying two isolated lepton candidates with *IsolatedLeptonTagging*
3. ISR identification and removing procedure
4. Jet reconstruction using *FastJet* clustering tools
5. Appling a weight factors to each event to get expected number of signal or background events

$$\mathcal{P}_{e^-e^+} = (-0.8, +0.3)$$

$$W = \left[\frac{1 \pm 0.8}{2} \cdot \frac{1 \pm 0.3}{2}\right] \cdot \frac{2\ \mathrm{ab}^{-1}}{\mathcal{L}_{\mathrm{nom}}}$$

$\mathcal{L}_{\mathrm{nom}}$ - the sample nominal integrated luminosities

The basic information for all MC samples used:

Process	Integrated luminosity, ab^{-1}		Cross section, fb		Number of events	
Signal samples						
	eLpR	eRpL	eLpR	eRpL	eLpR	eRpL
$q\bar{q}H(ZZ)$	55.6	86.9	8.99	5.75	$5{\cdot}10^5$	$5{\cdot}10^5$
$\nu_e\bar{\nu}_e H(ZZ)$	316	889	1.58	0.56	$5{\cdot}10^5$	$5{\cdot}10^5$
$\nu_{\mu\tau}\bar{\nu}_{\mu\tau}H(ZZ)$	284	445	1.76	1.12	$5{\cdot}10^5$	$5{\cdot}10^5$
Background samples						
	eLpR	eRpL	eLpR	eRpL	eLpR	eRpL
$WW/ZZ(4j)$	5.00	5.32	$12.4{\cdot}10^3$	225	$62{\cdot}10^6$	$12{\cdot}10^5$
$W(2j)W(\ell\nu)$	5.00	5.77	$10.3{\cdot}10^3$	86.7	$51.4{\cdot}10^6$	$5{\cdot}10^5$
$Z(2j)Z(2\ell)$	5.06	5.00	1423	1219	$72{\cdot}10^5$	$61{\cdot}10^5$
$Z(2\nu)Z(2j) + \gamma^\star(2\ell)$	5.08	5.35	610	262	$31{\cdot}10^5$	$14{\cdot}10^5$
$2j$	5.00	5.00	$128{\cdot}10^3$	$70.4{\cdot}10^3$	$6.40{\cdot}10^8$	$3.52{\cdot}10^8$
$WW/ZZ + \gamma^\star(2\ell)$	5.82		3.44		$2{\cdot}10^4$	
$e^+e^-H(jj)$	28.3	44.9	17.7	11.1	$5{\cdot}10^5$	$5{\cdot}10^5$
$\mu^+\mu^-H(jj)$	29.4	46.0	17.0	10.9	$5{\cdot}10^5$	$5{\cdot}10^5$
$\tau^+\tau^-H(jj)$	29.5	46.1	16.9	10.8	$5{\cdot}10^5$	$5{\cdot}10^5$
$\nu_e\bar{\nu}_e H(all)$	2.3	8.3	60.4	21.5	$5{\cdot}10^5$	$5{\cdot}10^5$
$\nu_{\mu\tau}\bar{\nu}_{\mu\tau}H(all)$	7.5	11.6	67.1	42.9	$5{\cdot}10^5$	$5{\cdot}10^5$
$jjH(all)$	1.5	2.3	343.1	219.5	$5{\cdot}10^5$	$5{\cdot}10^5$

The numbers are taken from *generation logbook ELOG*

Isolated lepton identification

We using the *IsolatedLeptonTagging* processor with default set of parameters and weights to identify leptons

The detectors instrumented in the magnet yoke surrounding the muon chamber are not used in the algorithm, it results in a small decrease in efficiency

The $Z^\star$ and Z reconstruction efficiencies in the leptonic modes in the channel with four jets (two jets) are $\sim$67 % ($\sim$72 %) and $\sim$90 % ($\sim$91 %), respectively

The only events with two identified isolated leptons are kept for the following analysis. These leptons are excluded from the following jet reconstruction procedure

ISR and γγ overlay removing

γγ overlay removed using *kT jet clustering*

From arXiv:2009.04340:

ISR photon candidate is selected if it's energy E_{photon} is *greater than 10 GeV*

All charged particles in a cone with $\cos\theta_{cone} = 0.95$ around the photon are summed up

$$\boldsymbol{E_{sum} < 5\%\, E_{photon}} \rightarrow \text{ ISR photon !}$$

Jet reconstruction procedure

Valencia algorithm is used to force the remaining particles into 2 or 4 jets.

It contains 3 parameters: R - generalized jet radius, γ and β - special capture parameters in beam distance

We use β = 1 and tune R and γ with this method from ***arXiv:1607.05039***:

$$\Delta M(Z) = M_{reco}(Z) - M_{gen}(Z)$$

$$Median = Q(0.5)$$

$$IQR_{34} = \frac{Q(0.84) - Q(0.16)}{2}$$

$$RMS_{90} = \sqrt{(|M^2_{mean} - M_{mean}|)}$$

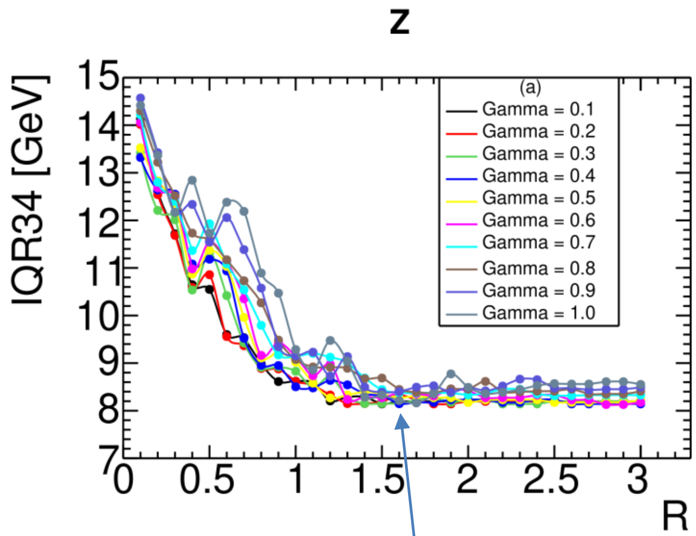
Choosing combination of minimum of IQR34, RMS90 and close to 0 median

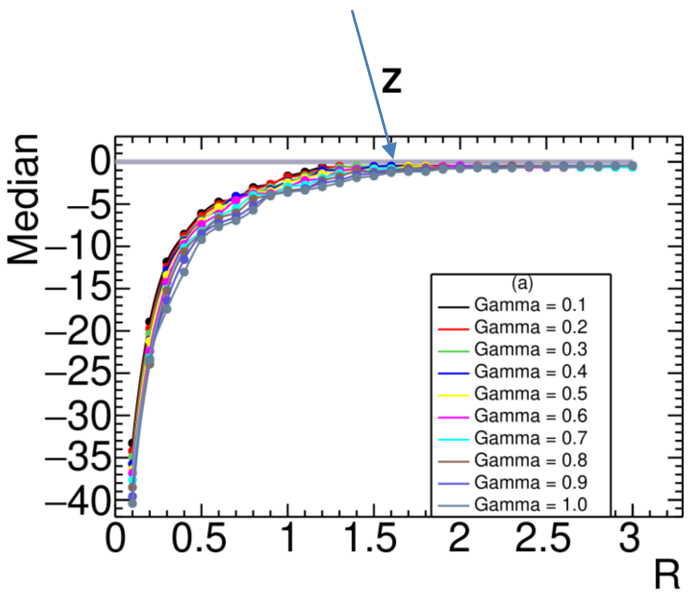
$$M_{mean} = \frac{\sum \Delta M(Z)}{N_{entries}} \qquad M^2_{mean} = \frac{\sum(\Delta M(Z))^2}{N_{entries}}$$

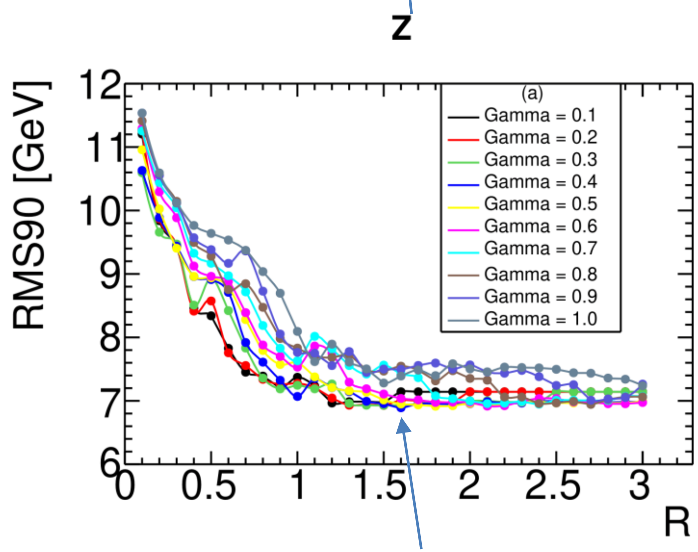
Jet reconstruction procedure

Example of jet parameters choosing procedure for $Z(jj)$

R[0.1, 3.0] and *γ[0.1, 1.0]* ranges



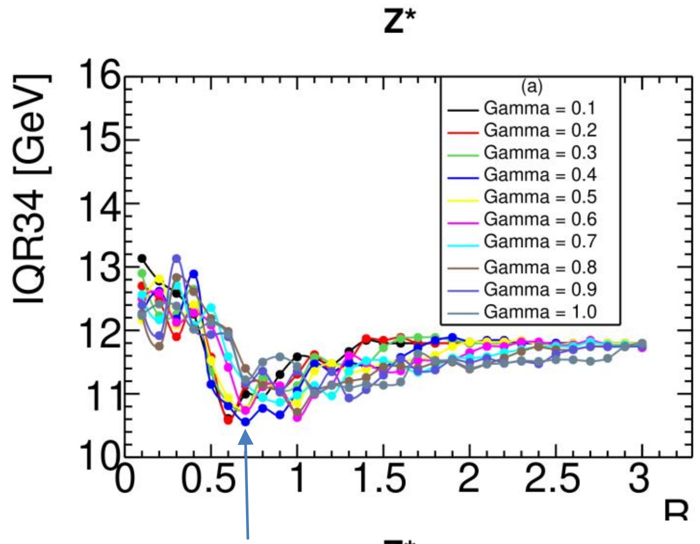


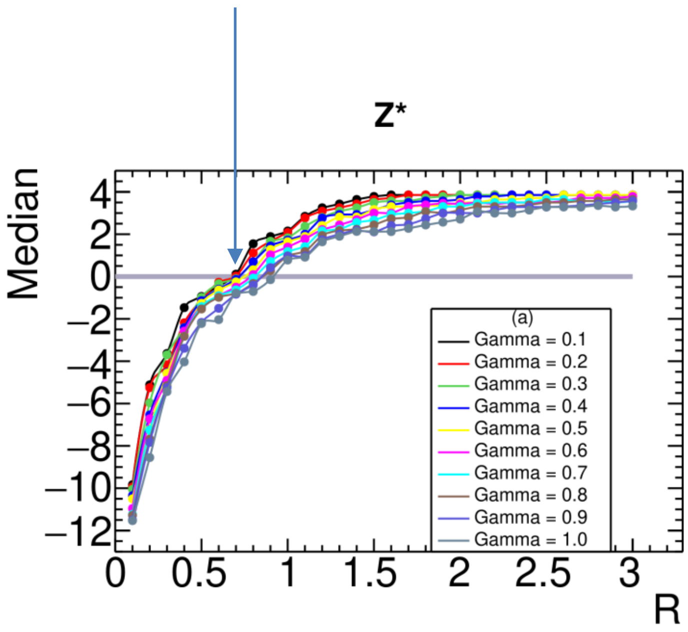


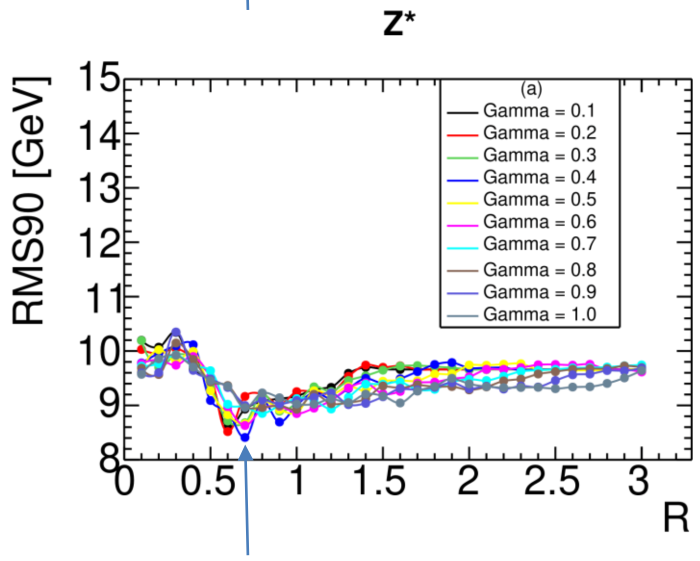
Jet reconstruction procedure

Example of jet parameters choosing procedure for $Z^{\star}(jj)$

R[0.1, 3.0] and *γ[0.1, 1.0]* ranges







Jet reconstruction procedure

The best Valencia algorithm parameters chosen for the jet reconstruction in different channels:

Valencia parameters	$Z_1(jj)$, $Z(jj)$, $Z^*(\ell\ell)$	$Z_1(jj)$, $Z(\ell\ell)$, $Z^*(jj)$	$Z_1(\nu\bar{\nu})$, $Z(jj)$, $Z^*(\ell\ell)$	$Z_1(\nu\bar{\nu})$, $Z(\ell\ell)$, $Z^\star(jj)$
β	1.0	1.0	1.0	1.0
γ	0.4	0.4	0.6	0.3
R	1.6	0.7	1.4	1.4

Preselection's results

The numbers of signal events for different final states obtained from MC samples with different polarizations before and after lepton tagging and reweighting:

Channels	$\mathcal{P}_{e^-e^+}$	MC events	Lepton tagging, events	Weight factors	Weighted number of events
$Z_1(jj)$, $Z(jj)$, $Z^*(\ell\ell)$	eLpR eRpL	23989 23845	16088 16027	$2.1 \cdot 10^{-2}$ $1.3 \cdot 10^{-3}$	338 21
$Z_1(jj)$, $Z(\ell\ell)$, $Z^*(jj)$	eLpR eRpL	23261 23132	20879 20664	$2.1 \cdot 10^{-2}$ $1.3 \cdot 10^{-3}$	439 27
$Z_1(\nu_e\bar{\nu}_e)$, $Z(jj)$, $Z^\star(\ell\ell)$	eLpR eRpL	24044 23910	17429 17259	$3.7 \cdot 10^{-3}$ $7.9 \cdot 10^{-5}$	65 2
$Z_1(\nu_e\bar{\nu}_e)$, $Z(\ell\ell)$, $Z^\star(jj)$	eLpR eRpL	23059 23096	21108 21149	$3.7 \cdot 10^{-3}$ $7.9 \cdot 10^{-5}$	79 2
$Z_1(\nu_{\mu,\tau}\bar{\nu}_{\mu,\tau})$, $Z(jj)$, $Z^\star(\ell\ell)$	eLpR eRpL	23840 23862	17103 17168	$4.1 \cdot 10^{-3}$ $1.6 \cdot 10^{-4}$	71 3
$Z_1(\nu_{\mu,\tau}\bar{\nu}_{\mu,\tau})$, $Z(\ell\ell)$, $Z^\star(jj)$	eLpR eRpL	23189 23225	21168 21246	$4.1 \cdot 10^{-3}$ $1.6 \cdot 10^{-4}$	88 4

Analysis results

1. The signal and background distributions are obtained with the weighted bin contents and uncertainties and then fitted separately.

2. The signal statistical uncertainties corresponding to the integrated luminosity of $2ab^{-1}$ are estimated using following steps:
   a) *Weighted signal and background distributions are summed*
   b) *Content of each bin is rounded to the integer number*
   c) *Poisson uncertainties for the bin contents are assumed to imitate real data*

3. The binned extended maximum likelihood fit method is applied to the obtained distributions with the fixed shapes obtained in the first step and free normalizations for the signal and background

4. The obtained results are tested using the *toy MC method*

Channel 1: $Z_1 \to jj, Z \to jj , Z^\star \to ll$

The final state of the first studied channel includes *two leptons and four jets*. To form the Z_1 and Z bosons from these four jets we calculate **χ^2** for six possible two-jet combinations:

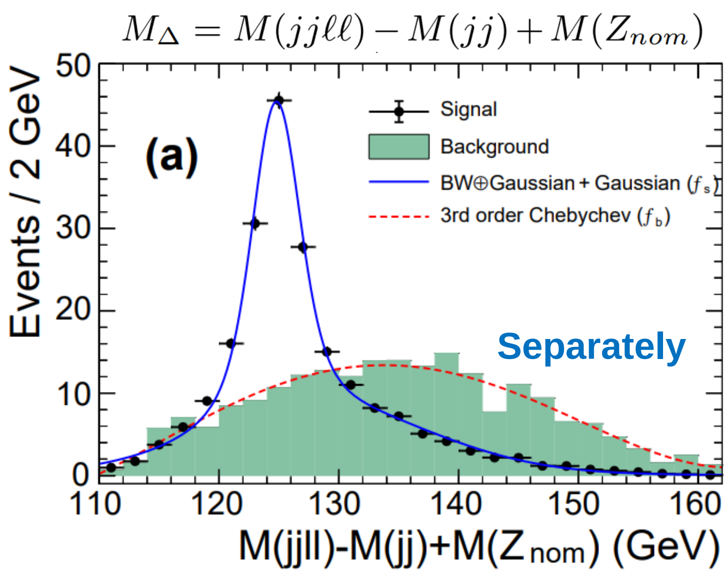
$$\chi^2 = \frac{(M(Z_1) - M(Z_{nom}))^2}{\sigma^2{}_{M_{Z_1}}} + \frac{(M(Z) - M(Z_{nom}))^2}{\sigma^2{}_{M_Z}} + \frac{(P(Z_1) - \overline{P}(Z_1))^2}{\sigma^2{}_{P_{Z_1}}} + \frac{(P(Z + Z^\star) - \overline{P}(Z_1))^2}{\sigma^2{}_{P_{Z+Z^\star}}}$$

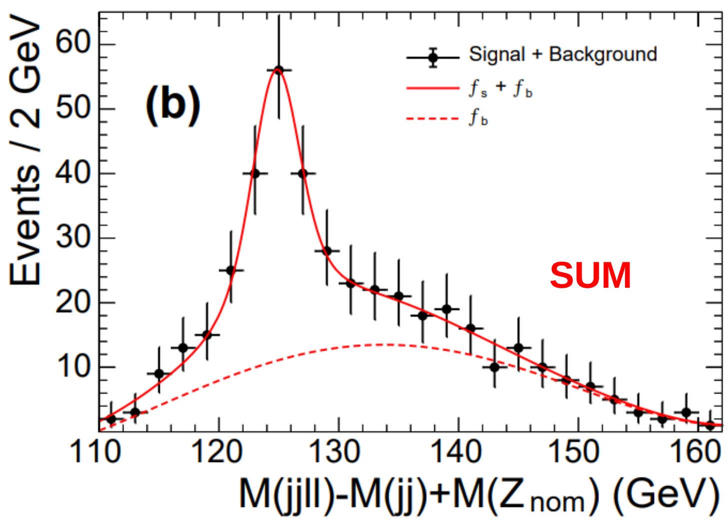
$\overline{P}(Z_1) = 60.0\ \mathrm{GeV}/c$ is the mean Z_1 momentum

$M(Z_{\mathrm{nom}}) = 91.2\ \mathrm{GeV}$

All **σ** parameters are the mean widths of corresponding mass or momentum distributions on the reconstruction level

Channel 1: $Z_1 \to jj, Z \to jj\ , Z^\star \to ll$





Significant backgrounds:

$e^+e^- \to W^+W^-\gamma^\star$ and $e^+e^- \to ZZ\gamma^\star$ - *dominant background.* Suppressed by $M(ll) > 10\ GeV/c^2$ cut

$e^+e^- \to W^+(jj)W^-(jj)$ *and* $e^+e^- \to Z(jj)Z(jj)$ - *small background (semileptonic jj decays)*. Suppressed by $P(l_{slow}) > 9\ GeV/c$ and $P(l_{fast}) < 45\ GeV/c$ cuts. The leptons from $Z^\star$ are called fast and slow depending on their momentum.

$H \to Z^\star Z^\star$ *contribution.* Suppressed by $M(jj) > 70\ GeV/c^2$

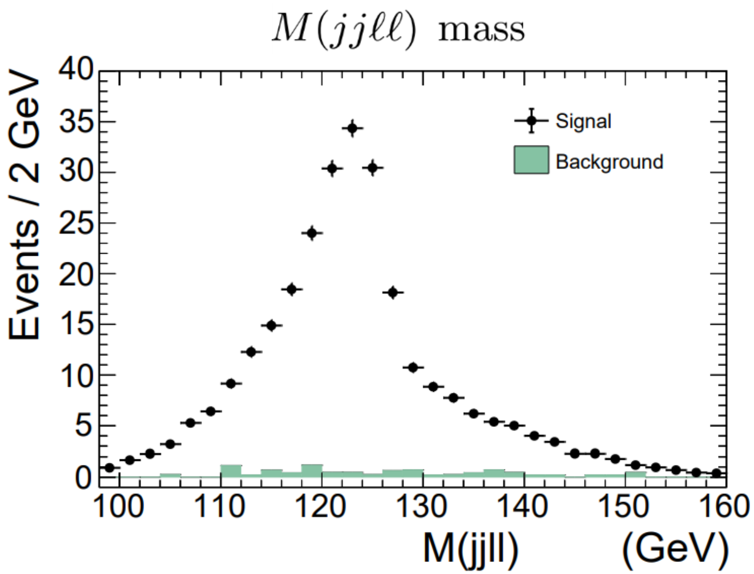
Other random backgrounds removed using $200 < M(jjjjll) < 260\ GeV/c^2$ cut

Additional wide Gaussian in fit function described residual $Z^\star Z^\star$ events and a few events due to a wrong jet matching in the χ^2 selection

The integral over all bins of the signal distribution is 204.0 events
The fit yields 202.5 ± 23.8 signal events

Statistical uncertainty is 11.69%

Channel 2: $Z_1 \to jj,\ Z \to ll,\ Z^\star \to jj$



χ^2 for six possible two-jet combinations:

$$\chi^2 = \frac{(M(Z_1) - M(Z_{nom}))^2}{\sigma^2{}_{M_{Z_1}}} + \frac{(E(Z_1) - \overline{E}(Z_1))^2}{\sigma^2{}_{E_{Z_1}}} + \frac{(P(Z_1) - \overline{P}(Z_1))^2}{\sigma^2{}_{P_{Z_1}}} + \frac{(P(Z + Z^\star) - \overline{P}(Z_1))^2}{\sigma^2{}_{P_{Z+Z^\star}}}$$

$\overline{E}(Z_1) = 110.0\ \mathrm{GeV}$ is the mean Z_1 energy

$H \to Z^\star Z^\star$ *and* *random lepton pairs* contributions are suppressed using $70 < M(ll) < 95\ GeV/c^2$ and $E(jjjjll) < 260\ GeV$ cuts.

Rejecting candidates with $M(jj) > 50\ GeV/c^2$ corresponding to the $Z^\star \to jj$ decay

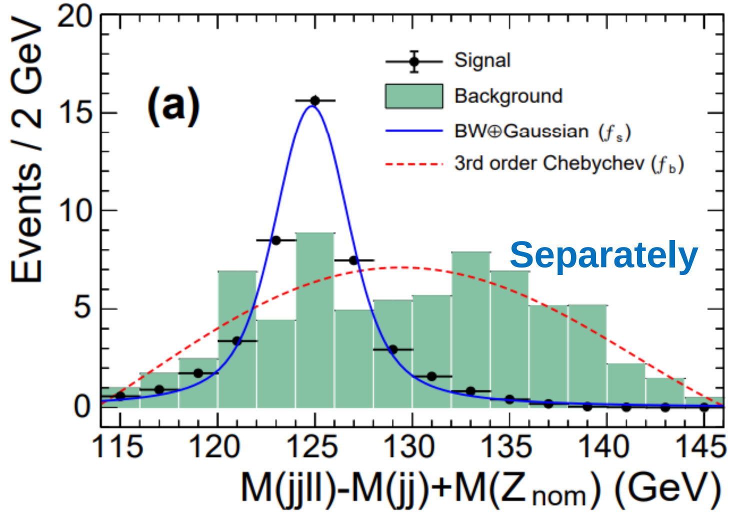
The background is very small, the integral over all bins is 10.0 events

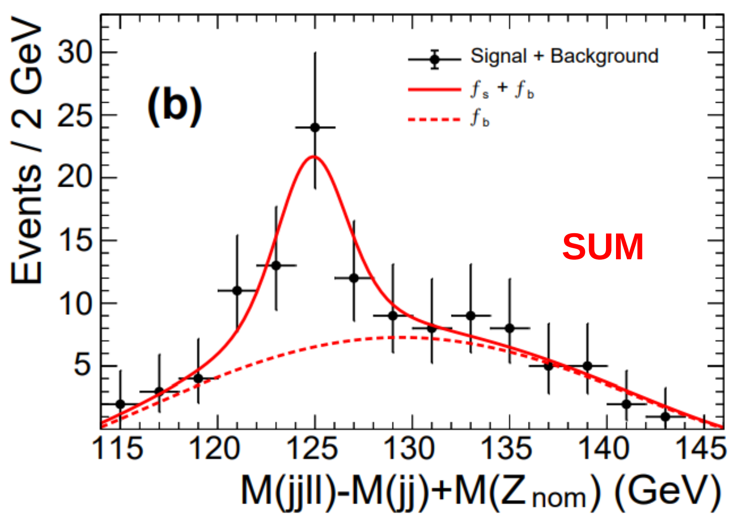
The signal mean value and uncertainty is 275.3 ± 16.9

Statistical uncertainty is 6.03%

Channel 3: $Z_1 \to \upsilon\bar{\upsilon}$, $Z \to jj$, $Z^\star \to ll$

$$M_\Delta = M(jj\ell\ell) - M(jj) + M(Z_{nom})$$





There are many background sources with large cross sections, which can contribute to this channel

Significant backgrounds:

$e^+e^- \to Z(jj)Z(\tau^+\tau^-)$ *with following leptonic decays*
$e^+e^- \to W^+(jj)W^-(l\upsilon)$

Some backgrounds:

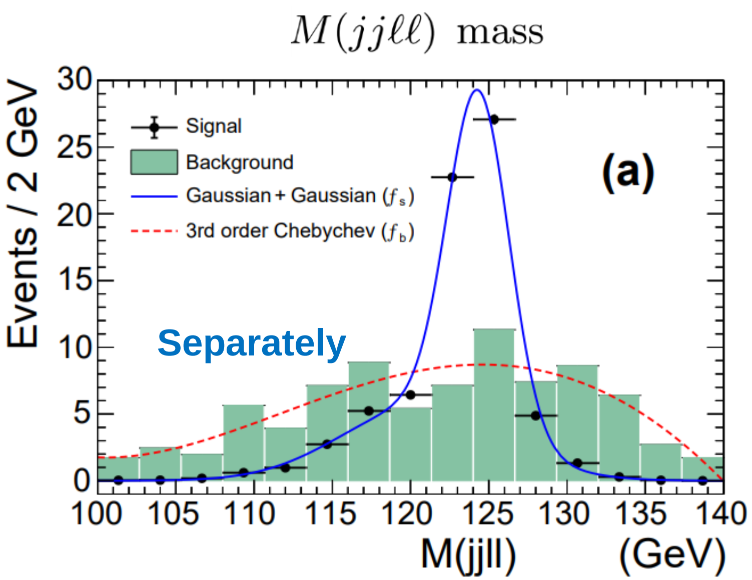
$e^+e^- \to b\bar{b}, e^+e^- \to ZH(b\bar{b})$
$H \to Z^\star Z^\star$ contribution to the signal

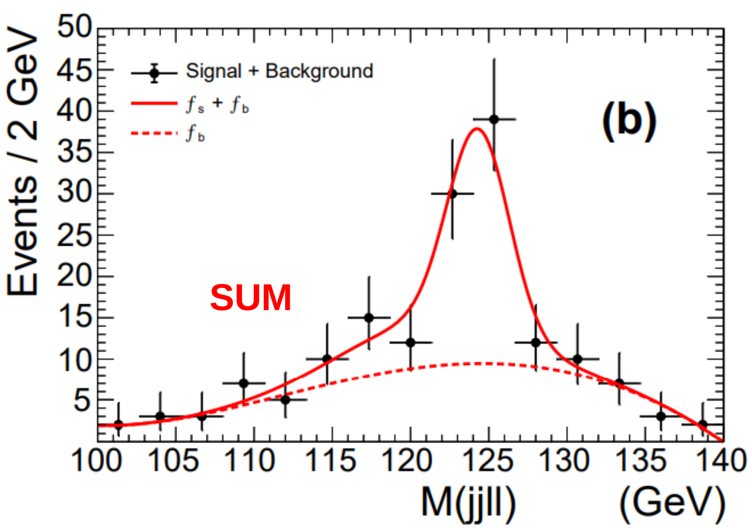
Effective cuts: $30 < P(jjll) < 75\ GeV/c$
$E(jjll) < 145\ GeV$
$|\cos(\theta_{vis})| < 0.8$ - azimuthal angle of the full system relative to the beam direction
$\Delta\varphi_{ZZ^\star} < 120^o$ - angle between the Z and Z^* boson directions
$10 < M(ll) < 40\ GeV/c^2$
$80 < M(jj) < 120\ GeV/c^2$

The fit yields 43.7 ± 10.8 signal events

Statistical uncertainty is 24.77%

Channel 4: $Z_1 \to \upsilon\bar{\upsilon}$, $Z \to ll$, $Z^\star \to jj$





The dangerous background sources are similar to the previous channel, except the $b\bar{b}$ background

Significant backgrounds:

$e^+e^- \to Z(jj)Z(\tau^+\tau^-)$ *with following leptonic decays*

$e^+e^- \to W^+(jj)W^-(l\upsilon)$

Some backgrounds:

$H \to Z^\star Z^\star$ contribution to the signal

Effective cuts: $40 < P(jjll) < 70\ GeV/c$

$E(jjll) < 145\ GeV$

$|\cos(\theta_{vis})| < 0.9$ - azimuthal angle of the full system relative to the beam direction

$\Delta\varphi_{ZZ^\star} < 140^o$ - angle between the Z and Z^* boson directions

$80 < M(ll) < 95\ GeV/c^2$

$10 < M(jj) < 40\ GeV/c^2$

$P(j_{fast}) < 40$

$P(j_{slow}) < 20$

The fit yields 70.4 ± 13.5 signal events

Statistical uncertainty is 18.59%

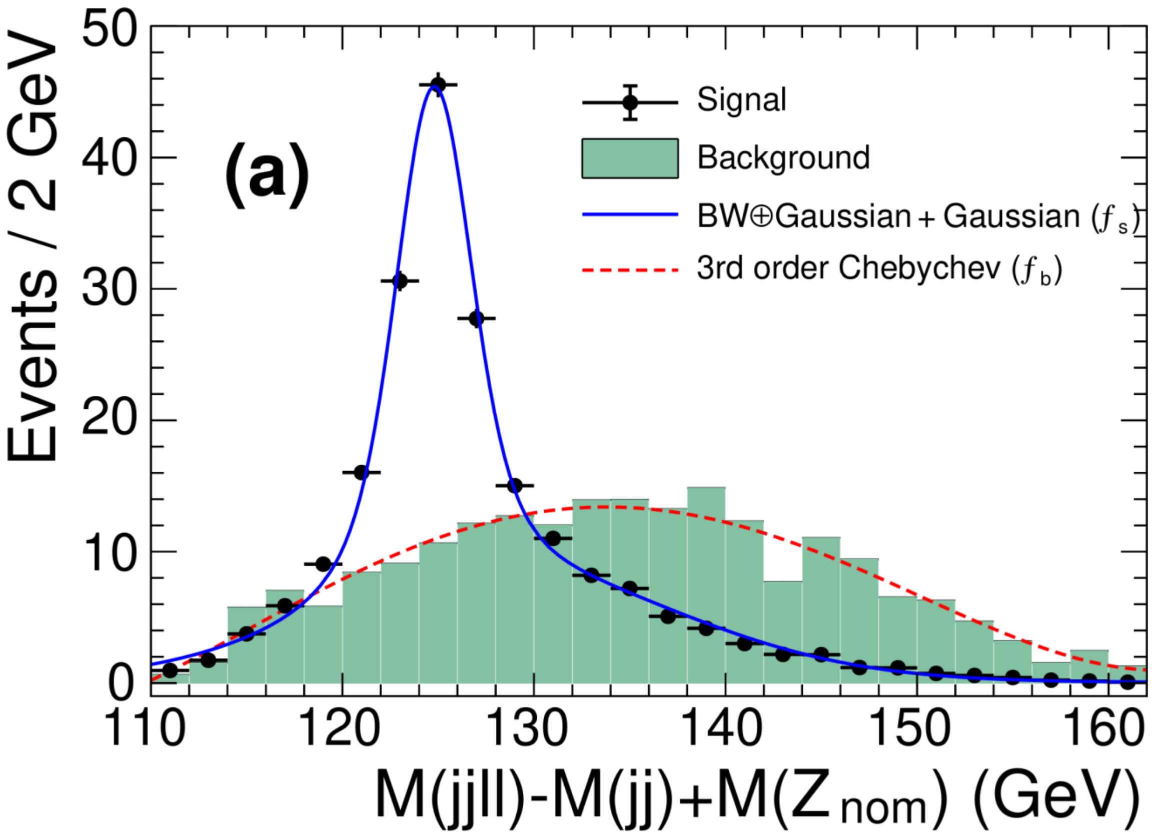
Table of cuts for all channels

	Ch 1	Ch 2	Ch 3	Ch 4
Selection	$Z_1(jj)$, $Z(jj)$, $Z^*(\ell\ell)$	$Z_1(jj)$, $Z(\ell\ell)$, $Z^*(jj)$	$Z_1(\nu\bar{\nu})$, $Z(jj)$, $Z^*(\ell\ell)$	$Z_1(\nu\bar{\nu})$, $Z(\ell\ell)$, $Z^\star(jj)$
$M(\ell\ell)$ (GeV/c^2)	[10, 40]	[70, 95]	[10, 40]	[80, 95]
$M(Z \to jj)$ (GeV/c^2)	> 70	< 50	[80, 120]	[10, 40]
$M(Z_1 \to jj)$ (GeV/c^2)	> 70	> 70		
$E(jjjj\ell\ell)$ (GeV)	[200, 260]	[200, 260]		
$E(jj\ell\ell)$ (GeV)			< 145	< 145
$P(\ell_{fast})$ (GeV/c)	< 45		< 40	
$P(\ell_{slow})$ (GeV/c)	> 9		> 8	
$P(j_{fast})$ (GeV/c)				< 40
$P(j_{slow})$ (GeV/c)				< 20
$P(jj\ell\ell)$ (GeV/c)			[30, 70]	[40, 70]
$\lvert cos\,\theta_{vis}\rvert$			< 0.8	< 0.9
$\Delta\phi_{ZZ^\star}$ (degree)			< 120	< 140

Comparison of combined fit and toy MC methods: stage 1

First stage is exactly the same for both methods.

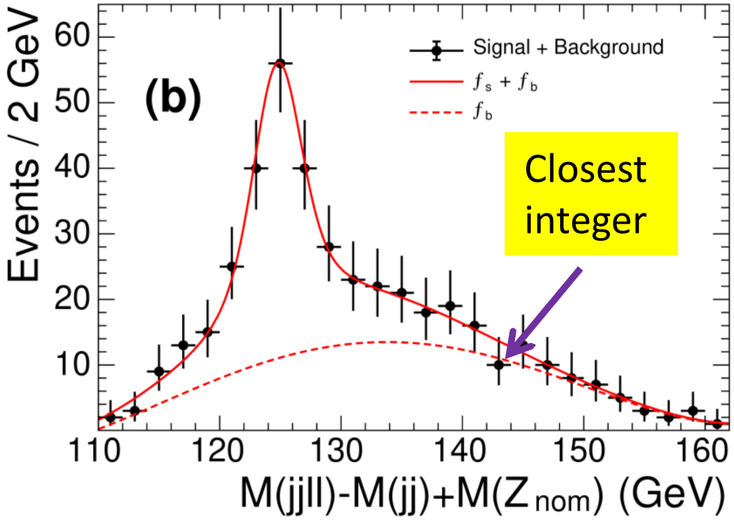
a) weighted histograms are obtained with correct uncertainties separately for signal and background distributions.
b) these histograms are fitted separately with appropriate functions
c) summary histogram for signal and background is obtained



Comparison of combined fit and toy MC methods: stage 2

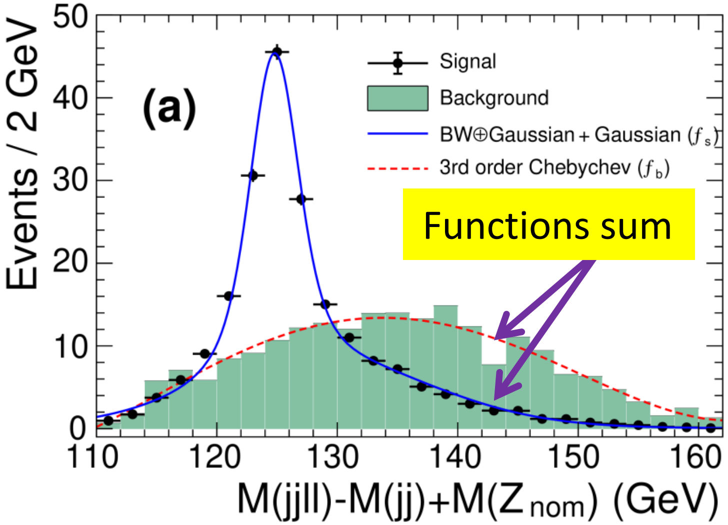
Combined fit

a) Move summary histogram bin values to closest integer number.
b) Assume Poisson uncertainties and get the best estimates from LH fit



Toy MC (RooFit)

a) Move summary histogram bin values to the functions sum value
b) Assume Poisson uncertainties and get the best estimates from LH fits



The only difference between these methods: combined fit moves bin values to closest integer value, but toy MC moves bin values to functions sum value obtained in stage 1. Then estimates for mean and uncertainty are done using the same likelihood method.

Comparison of combined fit and toy MC methods: systematics

Combined fit method is simple and elegant. Systematic uncertainty of rounding is small.

Combined fit systematics: systematic uncertainty due to rounding is $\sqrt{N_{bins}}\,/\,\sqrt{12}$ ~ 1.2 ev. (more accurately ~ 0.9 events), no additional uncertainties. Estimate is consistent, effective, gaussian, unbiased (because no close thresholds). On top of 11.7 % statistical uncertainty, we get additional 0.5 % systematic uncertainty (those can be neglected).

Toy MC systematics (RooFit method): we have to estimate quality of data description by function form. How to estimate it? We can try to use χ^2 criteria, Kolmogorov criteria, variation of functions parameters within obtained errors, maybe by eye. If description is good, this systematics is small. If not, this systematics can be large. In our case, this systematics is about of 0.5-0.6 events (about the same as in combined fit).

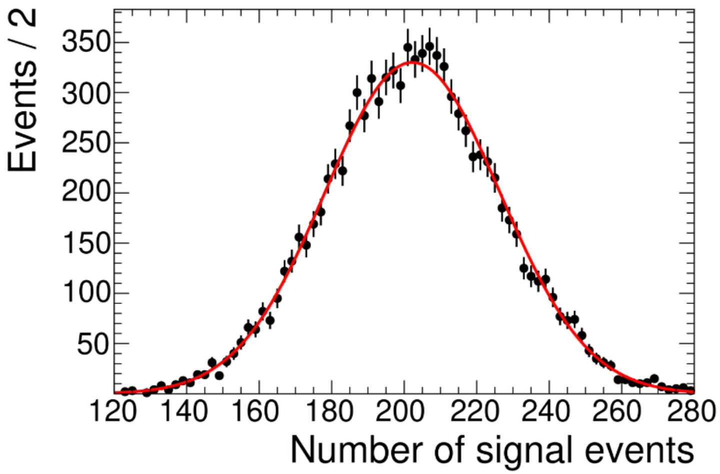
Final results:

1 channel	combined fit: N = 202.5 ± 23.8	toy MC: N = 202.3 ± 24.2
3 channel	combined fit: N = 43.7 ± 10.8	toy MC: N = 43.3 ± 11.3
4 channel	combined fit: N = 70.4 ± 13.1	toy MC: N = 69.8 ± 13.5

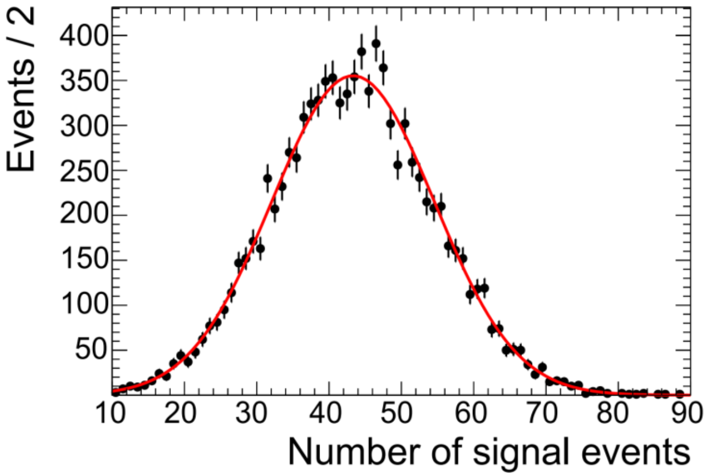
➔ Perfect agreement, as it was expected.

If combined fit method is wrong, why results agree with correct toy MC method?

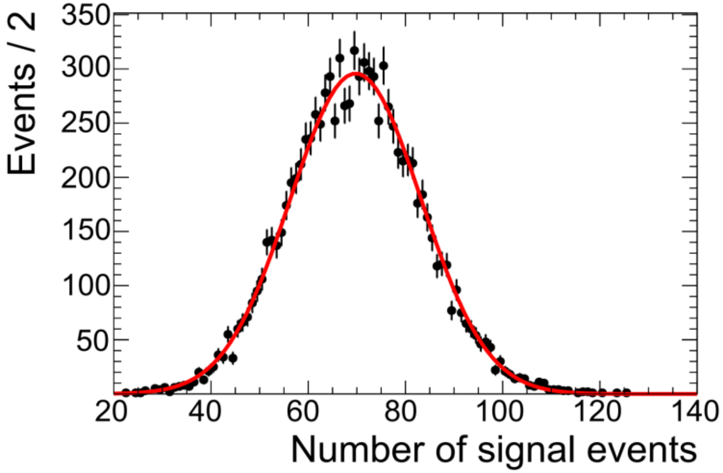
Toy MC, channel 1



Toy MC, channel 3



Toy MC, channel 4



Combined signal significance estimate

To estimate the accuracy, we calculate the combined statistical uncertainty for the four studied channels using the formula:

$$S_{\rm comb} = 1/\sqrt{\sum_{i=1}^{4} S_i^{-2}}$$

Alternatively, we assumed two data samples with the polarizations $\mathcal{P}_{e^-e^+} = (-0.8, +0.3)$, $\mathcal{P}_{e^-e^+} = (+0.8, -0.3)$ and 0.9 ab^{-1} luminosity each.

The same analysis is repeated for this data taking scheme

Number of signal events and uncertainties obtained from fits for each channel:

	$Z_1(jj)$, $Z(jj)$, $Z^*(\ell\ell)$	$Z_1(jj)$, $Z(\ell\ell)$, $Z^*(jj)$	$Z_1(\nu\bar{\nu})$, $Z(jj)$, $Z^*(\ell\ell)$	$Z_1(\nu\bar{\nu})$, $Z(\ell\ell)$, $Z^*(jj)$	Sum
		2 ab^{-1} eLpR			
Number of events	203±24	275±17	44±11	70±13	-
Uncertainty	11.69%	6.03%	24.77%	18.59%	5.04%
		0.9 ab^{-1} eLpR + 0.9 ab^{-1} eRpL			
Number of events	144±20	202±14	30±8	68±13	-
Uncertainty	13.76%	7.02%	28.01%	19.23%	5.82%

Conclusions

We studied the $e^+e^- \rightarrow HZ$ process with the subsequent $H \rightarrow ZZ^\star$ decay.

The analysis is performed assuming the integrated luminosity $2\ ab^{-1}$ collected at the e^+e^- collisions with center-of-mass energy 250 GeV and the beam polarizations $\mathcal{P}_{e^-e^+} = (-0.8, +0.3)$

Four channels are studied and the corresponding signal and background contributions are estimated using MC simulation.

Summing results obtained in the four studied channels we obtain the combined statistical uncertainty 5.04%.

This indicates, that the Higgs width can be measured using this method with an accuracy of about 5% within the model-independent approach.

The paper was significant improved after referees comments.

We would like to thank referees!

Thank you for attention