Analysis of the Higgsstrahlung process ($e^+e^- \rightarrow Z(qq)H$) in the hadronic decay mode for the context of the ILC with the SDHCAL.

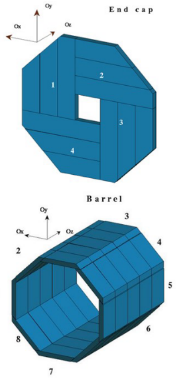
Héctor García Cabrera

CIEMAT

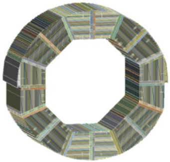
Setup

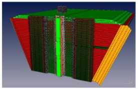
- Analysis of the Higgsstrahlung process in the context of the ILC.
- The detector model used is the ILD *l5_o2_v02*. Large model with the SDHCAL in the Tesla geometry and the SiWECal. List of all models.

SiWECAL



SDHCAL





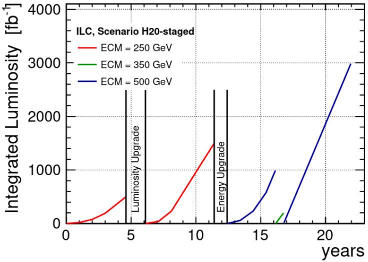
Running scenarios

CIEMAT

- Original scenarios in the proposal of 2015: ILC Operating Scenarios

	$\int \mathscr{L} dt$ [fb^{-1}]			
$\sqrt{s}$	G-20	H-20	I-20	Snow
250 GeV	500	2000	500	1150
350 GeV	200	200	1700	200
500 GeV	5000	4000	4000	1600

- H-20 scenario consolidated in the report to Snowmass 2021: ILC report to Snowmass 2021



MC Production

- At the moment of the request the request the Snowmass 2015 scenario was used as reference. L = 1150 fb -> 10.5 years of operation, which equals to $\sim$ 8.5 years in the H-20 scenario due to the early luminosity upgrade.
- The beam polarization values of operation in all scenarios are $P_{e^-}(80\%)$ and $P_{e^+}(30\%)$ creating the following proportions:

$$eL = 90\% \quad eR = 10\%$$
$$pL = 35\% \quad pR = 65\%$$

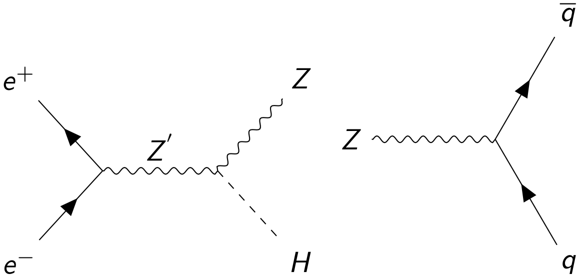
- The Snowmass 2015 scenario has the polarization mixture of: $(-,+) = 67.5\%$, $(+,-) = 22.5\%$, $(-,-) = 5\%$ and $(+,+) = 5\%$. Producing the following polarization factors for the total samples:

$$eLpR = 42.175\% \quad eLpL = 25.825\%$$
$$eRpL = 17.425\% \quad eRpR = 14.575\%$$

- H-20 polarization mixture: $(-,+) = 45\%$, $(+,-) = 45\%$, $(-,-) = 5\%$ and $(+,+) = 5\%$. Easily adapted from the Snowmass 2021 by weighting the histograms.

Signal

The signal studied in this analysis is the Higgsstrahlung process with the Z decaying hadronically:

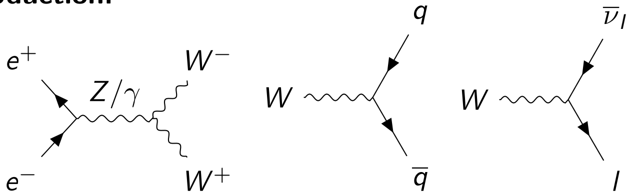


Also the MC request includes all the Higgs decay modes to make per channel tests: $H \longrightarrow b\overline{b}$, $H \longrightarrow c\overline{c}$, $H \longrightarrow gg$, $H \longrightarrow WW^*$, $H \longrightarrow ZZ^*$, $H \longrightarrow \mu\overline{\mu}$, $H \longrightarrow \tau\overline{\tau}$, $H \longrightarrow \gamma\gamma$ and $H \longrightarrow Z\gamma$

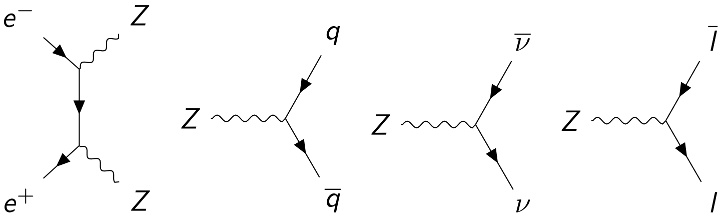
Backgrounds I

The background includes all processes with two or more jets in the final state.

W pair production.



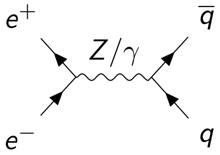
Z pair production.



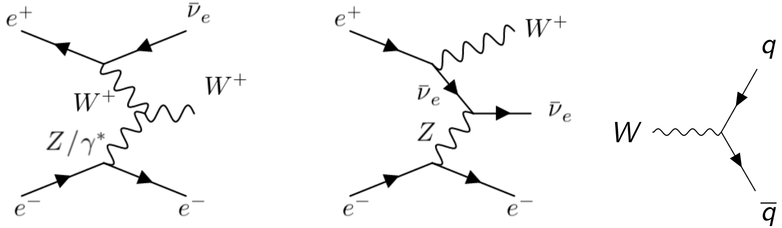
Backgrounds II

The background includes all processes with two or more jets in the final state.

Z hadronic decay.



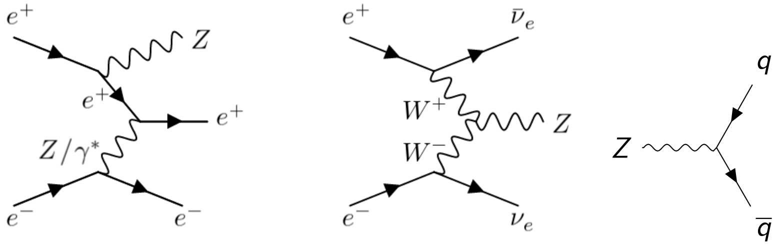
Single W production (+ charge conjugation).



Backgrounds III

The background includes all processes with two or more jets in the final state.

Single Z production (+ charge conjugation).



Signal request

CIEMAT

Process	eLpR (58.5%)	eRpL (3.5%)	eLpL (31.5%)	eRpR (6.5%)	Size
	σ_{Full} (fb) / NEvts / ReqEvts	σ_{Full} (fb) / NEvts / ReqEvts	σ_{Full} (fb) / NEvts / ReqEvts	σ_{Full} / NEvts / ReqEvts	
Pqqh	3.4303023e+02 / 211049	2.1948615e+02 / 8834	-	-	28.5 GB
Pqqh_aa	7.78678622e-01 / 479 / 10k	4.98233561e-01 / 20 / -	-	-	1.3 GB
Pqqh_az	5.24836252e-01 / 322 / 10k	3.35813810e-01 / 13 / -	-	-	1.3 GB
Pqqh_bb	1.99643594e+02 / 122830 / 100k	1.27740939e+02 / 5141 / 10k	-	-	14.3 GB
Pqqh_cc	9.91357365e+00 / 6099 / 10k	6.34314973e+00 / 255 / 10k	-	-	2.6 GB
Pqqh_e2e2	7.47805901e-02 / 46 / -	4.78479807e-02 / 1 / -	-	-	-
Pqqh_e3e3	2.15079954e+01 / 13232 / 10k	1.37617816e+01 / 553 / 10k	-	-	2.6 GB
Pqqh_gg	2.80941758e+01 / 17284 / 10k	1.79759157e+01 / 723 / 10k	-	-	2.6 GB
Pqqh_ww	7.34084692e+01 / 45164 / 50k	4.69700361e+01 / 1890 / 10k	-	-	7.8 GB
Pqqh_zz	8.98739203e+00 / 5529 / 10k	5.75053713e+00 / 231 / 10k	-	-	2.6 GB
Total					64 GB

Table 1: Full crossections σ_{Full}, number of events from the Snowmass scenario ($NEvts = L * \sigma_{Full} * f_{Pol}$; L = 1150 fb and f_{Pol} being the respective polarization factor), and the requested number of events per event and polarization. The last column is the estimated size of the total requested sample.

Background request

Process	eLpR (58.5%)	eRpL (3.5%)	eLpL (31.5%)	eRpR (6.5%)	Size
	σ_{Full} (fb) / NEvts / ReqEvts	σ_{Full} (fb) / NEvts / ReqEvts	σ_{Full} (fb) / NEvts / ReqEvts	σ_{Full} / NEvts / ReqEvts	
P2f_z_h	1.27965530e+05 / 78730792 / 1M	7.04167430e+04 / 2834273 / 1M	-	-	260 GB
P4f_sw_sl	1.02640160e+04 / 6314935 / 1M	8.66961490e+01 / 3489 / 10k	1.90531440e+02 / 69020 / 100k	1.90637490e+02 / 14250 / 20k	147 GB
P4f_sze_sl	1.42330980e+03 / 875691 / 1M	1.21939670e+03 / 49080 / 50k	1.15583340e+03 / 418700 / 400k	1.15720060e+03 / 86500 / 100k	202 GB
P4f_sznu_sl	4.53869760e+02 / 279243 / 300k	1.31219580e+02 / 5281 / 10k	-	-	40 GB
P4f_ww_h	1.48664200e+04 / 9146564 / 1M	1.36821530e+02 / 5507 / 10k	-	-	131 GB
P4f_ww_sl	1.87791450e+04 / 11553868 / 1M	1.73468290e+02 / 6982 / 10k	-	-	131 GB
P4f_zz_h	1.40506000e+03 / 864463 / 1M	6.06709780e+02 / 24420 / 30k	-	-	134 GB
P4f_zznu_sl	6.09878860e+02 / 375227 / 400k	2.61567200e+02 / 10528 / 10k	-	-	54 GB
P4f_zzorww_h	1.23892920e+04 / 7622511 / 1M	2.25568680e+02 / 9079 / 10k	-	-	131 GB
P4f_zz_sl	8.38079490e+02 / 515628 / 500k	4.66816440e+02 / 18789 / 20k	-	-	68 GB
Total					1.3 TB

Table 2: Full crossections σ_{Full}, number of events from the Snowmass scenario ($NEvts = L * \sigma_{Full} * f_{Pol}$; L = 1150 fb and f_{Pol} being the respective polarization factor), and the requested number of events per event and polarization. The last column is the estimated size of the total requested sample.

Final request

Process Name	Process ID	Polarization	Requested Events
Pqqh	I402011	eLpR	211049
Pqqh	I402012	eRpL	8834
P2f_z_h	I500010	eLpR	1000000 (1M)
P2f_z_h	I500012	eRpL	1000000 (1M)
P4f_sw_sl	I500105	eLpL	100000 (100k)
P4f_sw_sl	I500106	eLpR	1000000 (1M)
P4f_sw_sl	I500108	eRpL	10000 (10k)
P4f_sw_sl	I500107	eRpR	20000 (20k)
P4f_sze_sl	I500101	eLpL	400000 (400k)
P4f_sze_sl	I500102	eLpR	1000000 (1M)
P4f_sze_sl	I500104	eRpL	50000 (50k)
P4f_sze_sl	I500103	eRpR	100000 (100k)
P4f_sznu_sl	I500110	eLpR	300000 (300k)
P4f_sznu_sl	I500112	eRpL	10000 (10k)
P4f_ww_h	I500066	eLpR	1000000 (1M)
P4f_ww_h	I500068	eRpL	10000 (10k)
P4f_ww_sl	I500082	eLpR	1000000 (1M)
P4f_ww_sl	I500084	eRpL	10000 (10k)
P4f_zz_h	I500062	eLpR	1000000 (1M)
P4f_zz_h	I500064	eRpL	30000 (30k)
P4f_zznu_sl	I500078	eLpR	400000 (400k)
P4f_zznu_sl	I500080	eRpL	10000 (10k)
P4f_zzorww_h	I500070	eLpR	1000000 (1M)
P4f_zzorww_h	I500072	eRpL	10000 (10k)
P4f_zz_sl	I500074	eLpR	500000 (500k)
P4f_zz_sl	I500076	eRpL	20000 (20k)
Pqqh_aa	I402214	eLpR	10000 (10k)
Pqqh_az	I402215	eLpR	10000 (10k)
Pqqh_bb	I402209	eLpR	100000 (100k)
Pqqh_bb	I402218	eRpL	10000 (10k)
Pqqh_cc	I402210	eLpR	10000 (10k)
Pqqh_cc	I402219	eRpL	10000 (10k)
Pqqh_e3e3	I402216	eLpR	10000 (10k)
Pqqh_e3e3	I402225	eRpL	10000 (10k)
Pqqh_gg	I402211	eLpR	10000 (10k)
Pqqh_gg	I402220	eRpL	10000 (10k)
Pqqh_ww	I402212	eLpR	50000 (50k)
Pqqh_ww	I402221	eRpL	10000 (10k)
Pqqh_zz	I402213	eLpR	10000 (10k)
Pqqh_zz	I402222	eRpL	10000 (10k)

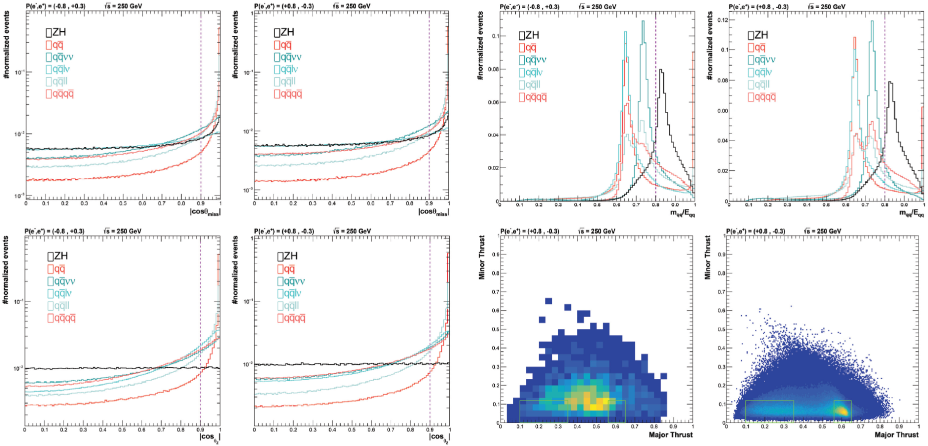
Event selection

- The samples include the simulation, reconstruction, digitizacion and overlay.
- The next step is to reconstruct the jets using FastJet.
- An event is selected if it has a dijet compatible with the Z mass.
- The parameters from FastJet are optimized by maximizing the percentage of well reconstructed jets. ($Z_{Tag} > 0.9$ and $Z_{Purity} > 0.9$).

$$\{j_1, j_2\} = argmin|m_{j1,j2} - m_Z| \;\; ; \;\; j_1, j_2 \in N_{Jets} \; ; j_1 \neq j_2$$

$$Z_{Tag} = \frac{E_Z^{j1,j2}}{E_Z^{Total}} \;\; ; \;\; Z_{Purity} = \frac{E_Z^{j1,j2}}{E^{j1,j2}}$$

Background rejection



M_Z and M_{Recoil}

Taking into account the incident angle of the two beams of 14mrad and $\sqrt{s} = 250 GeV$ the computations of the recoil mass are the following:

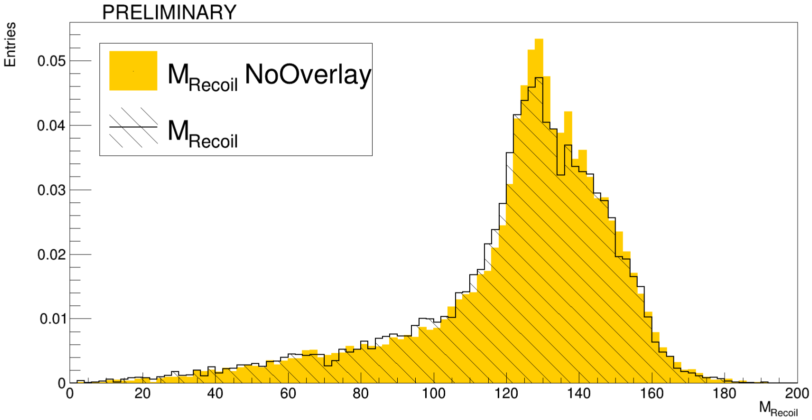
$$E_{q\overline{q}} = E_q + E_{\overline{q}} \;\; ; \;\; \overrightarrow{p_{q\overline{q}}} = \overrightarrow{p_q} + \overrightarrow{p_{\overline{q}}} \;\; ; \;\; M_Z = \sqrt{E_{q\overline{q}}^2 - \overrightarrow{p_{q\overline{q}}}^2}$$

$$M_{Recoil}^2 = (2E + E_{q\overline{q}})^2 - ((2p_x + p_{q\overline{q}_x})^2 + p_{q\overline{q}_y}^2 + p_{q\overline{q}_z}^2)$$

$$2E = 250.0061252 GeV \;\; ; \;\; 2p_x = 1.7500286 GeV$$

Recoil mass ee_genkt

With this algorithm the optimal parameters have been found to be R = 1.0 and YCut = 0.004 and an impact of less than 0.3 % from overlay in the Z reconstruction.



Cross section

$$\sigma_{ZH} = \frac{N_{Total} - N_B}{BR(Z \to q\overline{q}) \cdot \epsilon \cdot L}$$

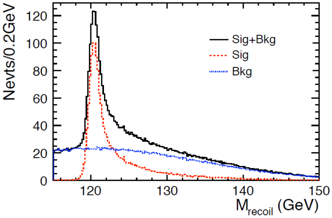
The error of the cross section is computed as:

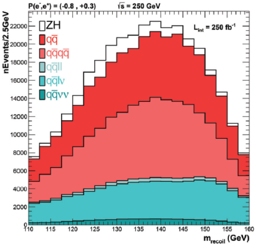
$$\Delta\sigma = \left(\frac{\Delta N_{Total}}{\epsilon L}\right)_{stat} \oplus \left(\frac{\Delta N_B}{\epsilon L} \oplus \sigma\frac{\Delta\epsilon}{\epsilon} \oplus \sigma\frac{\Delta L}{L}\right)_{syst} \; ; \; \frac{\Delta g_{HZZ}}{g_{HZZ}} = \frac{\Delta\sigma_{ZH}}{2\sigma_{ZH}}$$

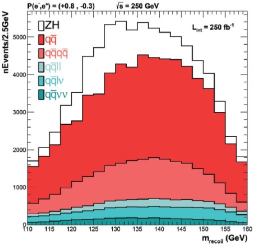
- $\Delta N_{Total} = \Delta N_B = 0$ if no fit to the data is performed Question: ¿How can the fit be made in the hadronic channel? Next slide.
- $\Delta\epsilon$ is the greatest challange since it involves the error propagation from the background rejection process.
- $\Delta L/L \sim 0.1\%$ from studies of the ILC and it is negligible compared to the others.

Results from other studies

$$F_M(x;\ \mathrm{M_H}, N_S) = N_S \cdot F_S(x;\ \mathrm{M_H}) + N_B \cdot F_B(x)$$







BACKUP

Open questions

- Processes uncertain if they apply as background: *aa_2f* ; *aa_4f* ; *ae_3f* ; *ae_5f*

Setup

- The latest MC production *sv02-02-01*, set 250 GeV, reconstructed with the SDHCAL *rv02-02-01*.mILD_l5_o2_v02 has now included the overlay events.
- Number of available events 50k in the DST-Merged files analyzed.
- Polarization eLpR (analysis of other polarizations is planned in the future).
- Beam crossing angle correction applied.

The objective of this preliminary analysis is to understand the effect of the newly included overlay into the signal events.

Z identification (previous results)

- We run over all possible jet combinatorial possibilities.
- Then we choose the pair that give us the closest mass to the Z known mass.
- Two quality variables are defined:

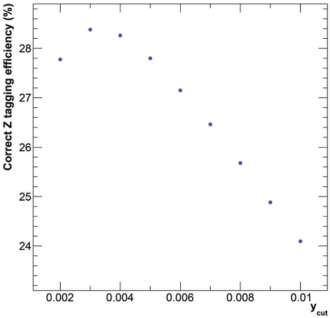
$$Z_{Tag} = \frac{E_Z^{j_1,j_2}}{E_Z^{Total}} \quad Z_{Purity} = \frac{E_Z^{j_1,j_2}}{E^{j_1,j_2}}$$

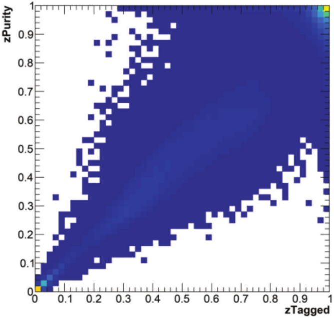
where $E_Z^{j_1,j_2}$: MC-truth di-jet energy from the Z, E_Z^{Total}: MC-truth Z energy and E^{j_1,j_2}: MC-truth total di-jet energy.

- The $E_Z^{j_1,j_2}$ is computed through a loop over all the PFOs. The associated MC particle energy is weighted using the RecoMCTrughLink excluding the particle with a Higgs as a parent (pdgId = 25).

Z identification (previous results)

The algorithm used in this study was *ee_kt_algorithm* within fast-jet with ExclusiveYCut strategy. An optimization of y_{cut} has to be done in order to maximize the di-jet reconstruction selection efficiency. A di-jet is considered efficient if $Z_{Tag} > 0.9$ and $Z_{Purity} > 0.9$.





The optimal value in this previous study was $y_{cut} = 0.003$.

FastJet algorithms

We want to find the set of parameters from the jet clustering algorithms from which the impact of the overlay is minimal. The two algorithms studied are:

- *Generalized k_t algorithm for e^+e^- collisions (ee_genkt) in the YCut exclusive mode.*

$$d_{ij} = min(E_i^{2p}, E_j^{2p})\frac{1 - cos(\theta_{ij})}{(1 - cos(R))} \quad ; \quad d_{iB} = E_i^{2p}$$

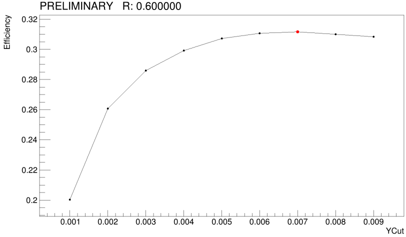
- *k_t algorithm for e^+e^- collisions (ee_kt) in the YCut exclusive mode.*

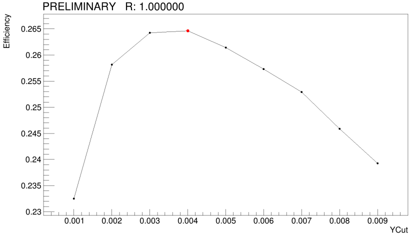
$$d_{ij} = 2min(E_i^2, E_j^2)(1 - cos(\theta_{ij}))$$

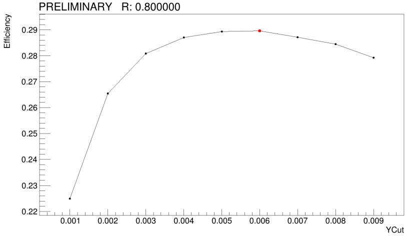
Parameters: ee_genkt -> **YCut**, **R** and **P** ; ee_kt -> **YCut**

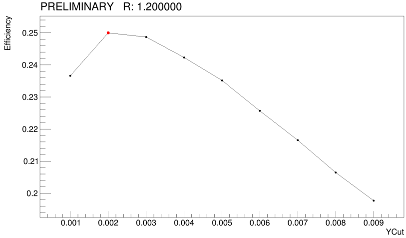
YCut scans ee_genkt

YCut scans like the ones in Slide 4 are performed for different values of **R** and **P** for the ee_genkt algorithm and another one for the ee_kt algorithm.



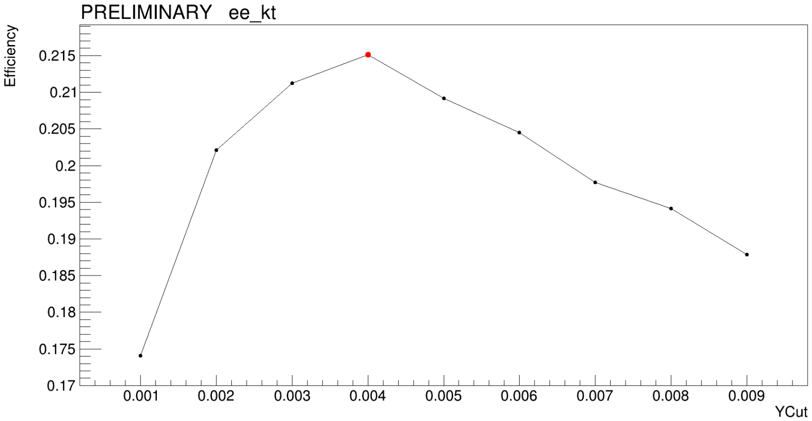






YCut scans ee_kt

YCut scans like the ones in Slide 4 are performed for different values of **R** and **P** for the ee_genkt algorithm and another one for the ee_kt algorithm.

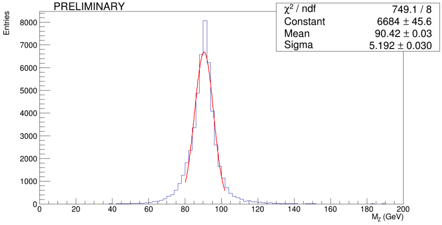


Quality variables

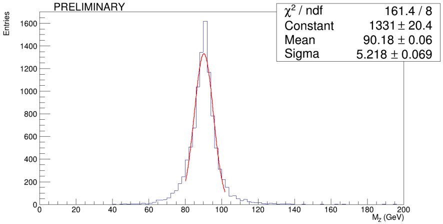
From the studied parameters R = 0.6 gives the highest efficiency. However this does not mean that gives the best jet reconstruction. The quality variables computed are:

- Z_{Mass} differences (%). The Z mass is reconstructed, following the previous procedure, with and without overlay. This distribution is fitted to a gaus function and then two differences are computed:
  $ZDiff = |M_Z - M_Z^{NoOverlay}|/M_Z$
  $OverlayDiff = |M_Z^{Overlay} - M_Z^{NoOverlay}|/M_Z^{NoOverlay}$

R = 0.6



R = 0.6 No Overlay

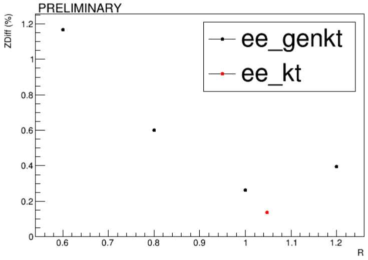


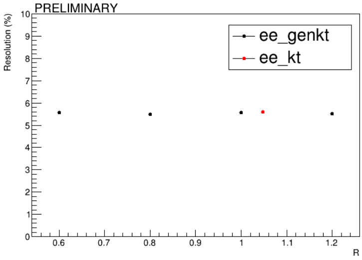
Quality variables

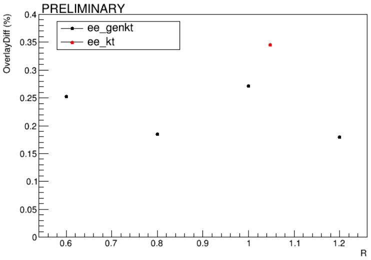
From the studied parameters R = 0.6 gives the highest efficiency. However this does not mean that gives the best jet reconstruction. The quality variables computed are:

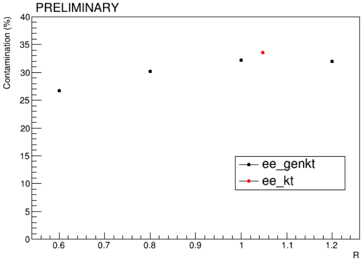
- Z_{Mass} differences (%). The Z mass is reconstructed, following the previous procedure, with and without overlay. This distribution is fitted to a gaus function and then two differences are computed:
  $ZDiff = |M_Z - M_Z^{NoOverlay}|/M_Z$
  $OverlayDiff = |M_Z^{Overlay} - M_Z^{NoOverlay}|/M_Z^{NoOverlay}$
- Z_{Mass} "resolution". (σ/M_Z) from the previous fit.
- Jet Contamination (%). The mean percentage of jets that have some overlay contribution.

Quality variables



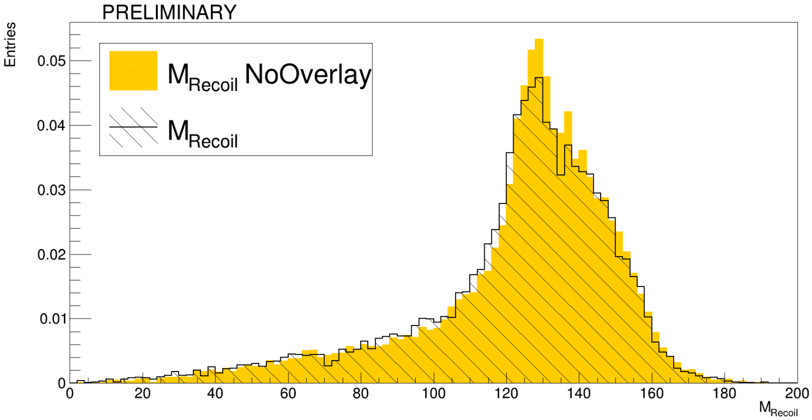






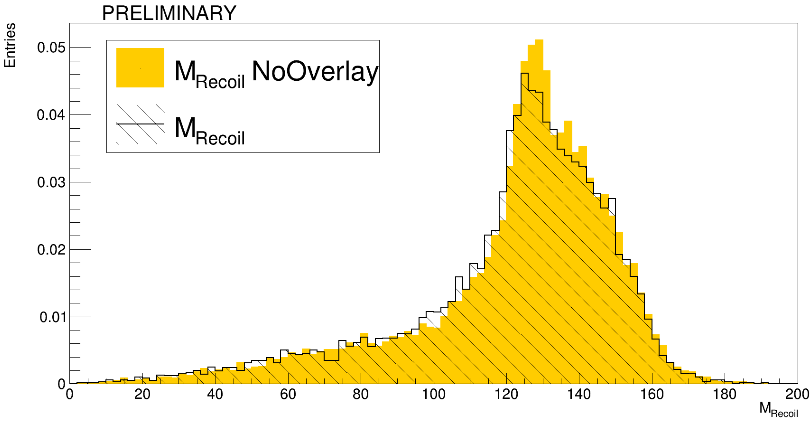
Recoil mass ee_genkt

With this algorithm the optimal parameters have been found to be R = 1.0 and YCut = 0.004 and an impact of less than 0.3 % from overlay in the Z reconstruction.



Recoil mass ee_kt

With this algorithm the optimal parameter is YCut = 0.004 and an impact of 0.35 % from overlay in the Z reconstruction.



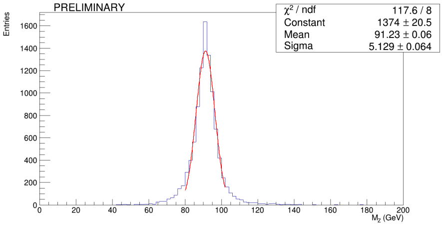
Conclusion

It have been shown the set of parameters that minimizes the overlay effect while efficiently reconstructing the Z mass and showing healthy recoil mass distributions.

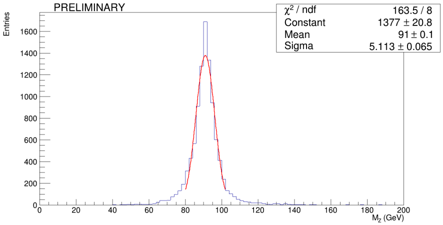
A slightly better shape in the recoil mass is obtained from the ee_kt method that indicates that it is the algorithm that should be used.

M_Z R = 1.0

Overlay

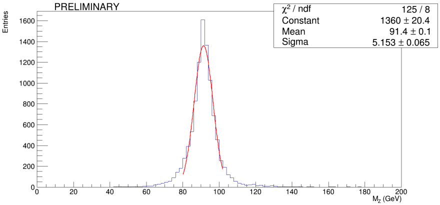


No Overlay



M_Z ee_kt

Overlay



No Overlay