$\nu\bar{\nu}H$: MY FIRST STEPS WITH THE 2020 MC SAMPLE

ILD Analysis/Software Meeting

Jean-Claude Brient, Jonas Kunath (LLR)

09.12.2020

OUTLINE

$\nu\bar{\nu}H$: My first steps with the 2020 MC sample

Disclaimer :
Most differences probably originate in my incomprehension.

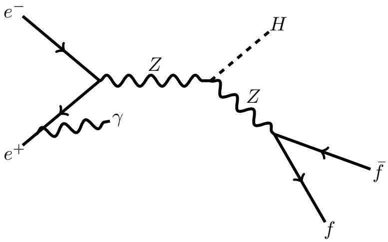
1. Research question
2. Expectations & getting started
3. Sample comparison
4. Conclusion

G_{HZZ} - WHAT CAN BE GAINED ?

RESEARCH QUESTION

Extracted from Higgsstrahlung events at a $\sqrt{s} = 250$ GeV.

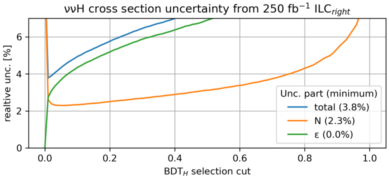
- $Z \to \mu^+\mu^-, Z \to e^+e^-$: Golden channels.
  Recoil mass method, already studied elsewhere.
- $Z \to \tau^+\tau^-$: Tagging on the τ is complicated.
  - Large τ decay opening angle (low E_τ).
  - Divers environment from the Higgs decay.
- $Z \to \nu\bar{\nu}$:
  – Significant WW-fusion contribution in $\nu\bar{\nu}H$.
  – Cannot tag event on ν.
  + Only Higgs boson (and beam overlay) present in event.
  + $6\times$ higher cross section.

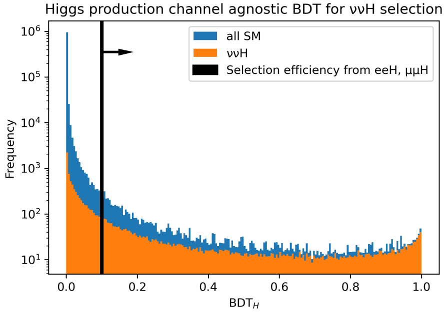


$\nu\bar{\nu}H$

RESEARCH QUESTION

- $\sigma_{\nu\bar{\nu}H}$ with contributions from both Higgsstrahlung and WW-fusion.
- Their relative size varies with the beam polarisation.
- Similar distributions for $\sqrt{s} = 250$ GeV.
- Idea : Extract the combined cross section ($\rightarrow$ production mode agnostic selection).
- WIP : Determine the benefit of this observable for G_{HZZ}, G_{HWW} from a global fit, including the correlations with e.g. $\sigma_{\text{WW-fusion}\rightarrow\nu\bar{\nu}b\bar{b}}$ (using SFitter).





EXPECTATIONS

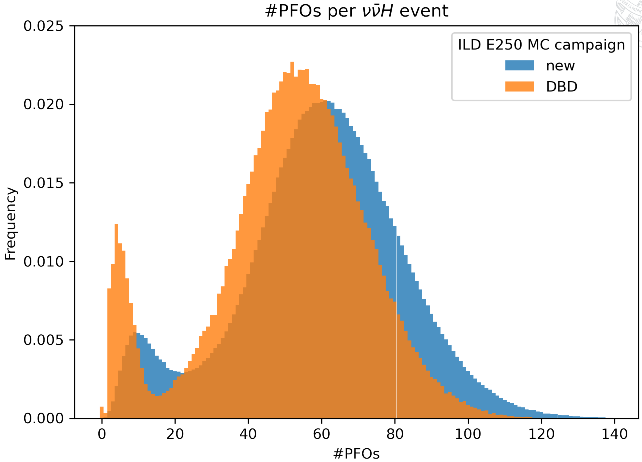
- Smoother distributions due to increases sample size ($\times 10\ \nu\bar{\nu}H$, $\times 10\ \mu^+\mu^- H$).
- Great improvement for rare (Higgs decay) modes from exclusive samples.
  - $\rightarrow$ Machine learning.
- Basically a drop-in replacement for the DBD $\sqrt{s} = 250$ GeV samples.
  - Detector (reconstruction) not altered too much.
  - Machine parameters and simulation similar.

GETTING STARTED

- Switch out file paths (e.g. at kek-cc) :
  `/group/ilc/soft/samples/mc-dbd/ild/dst-merged/250-TDR_ws/`
  $\rightarrow$ `/group/ilc/grid/storm/prod/ilc/mc-2020/ild/dst-merged/250-SetA/`
- Some changes in file naming : `Pnnh` $\rightarrow$ `Pn1n1h`, `Pn23n23h`.
- Unrealistic to store full sample locally : $\approx$50 GB for $\nu\bar{\nu}H$ alone.
- Reduced statistics for background processes up to now ?
- Start by comparing the (1D) signal distributions for my BDT input variables.

COMPARISON - # PANDORA PFOS IN A $\nu\bar{\nu}H$ EVENT

A shift towards higher values and more smeared out :
$\gamma\gamma \rightarrow$low-p_T hadron background now increased (to ≈ 1.6 events/bunch crossing).

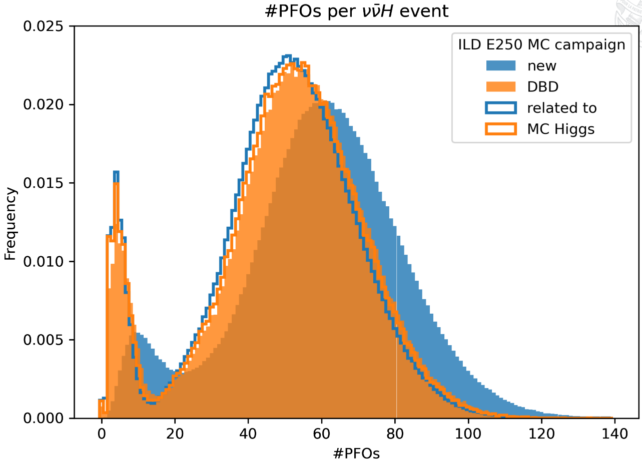


COMPARISON - # PANDORA PFOS IN A $\nu\bar{\nu}H$ EVENT

SAMPLE COMPARISON

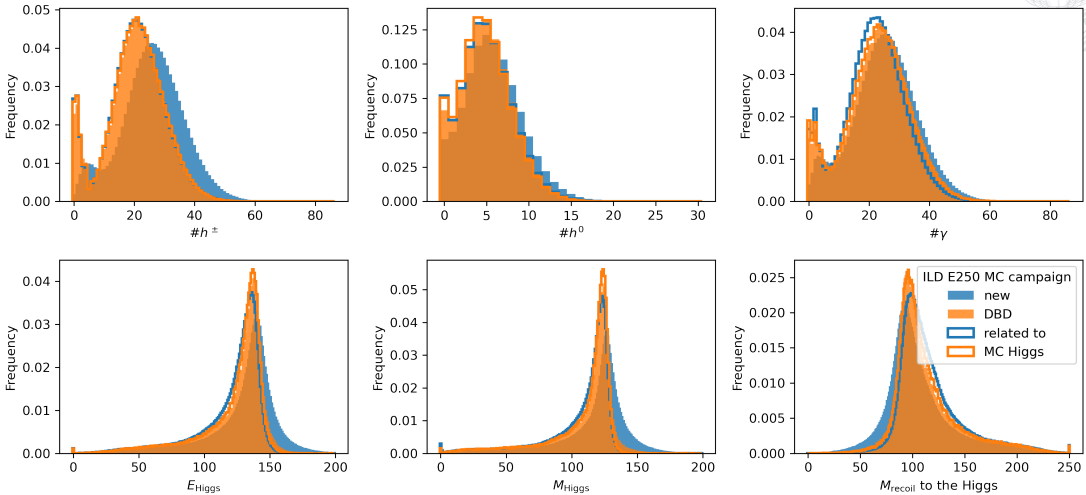
A shift towards higher values and more smeared out :
$\gamma\gamma \rightarrow$low-p_T hadron background now increased (to ≈ 1.6 events/bunch crossing).

The distributions of only the actual Higgs decay products are similar.



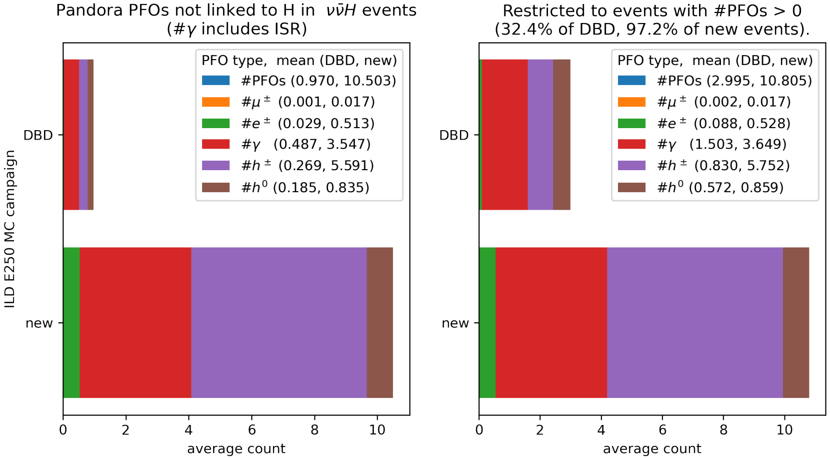
GLOBAL VARIABLES

Variable distributions in $\nu\bar{\nu}H$ event



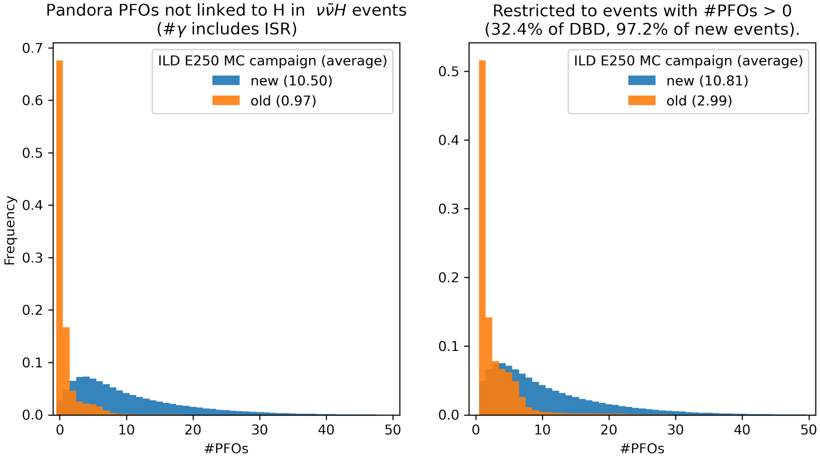
OVERLAY COMPOSITION

Frequency and size of the overlay increased, composition altered.



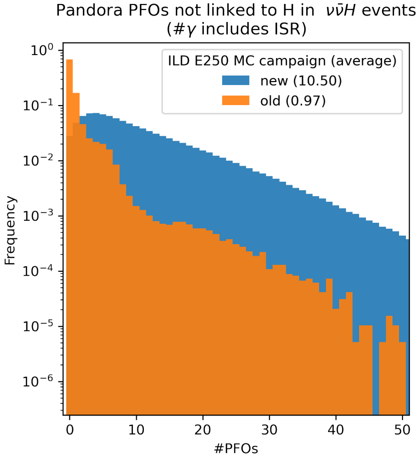
OVERLAY SHAPE

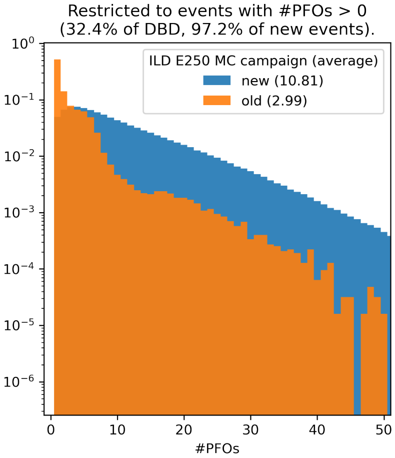
Frequency and size of the overlay increased, composition altered.



OVERLAY SHAPE - LOG SCALE

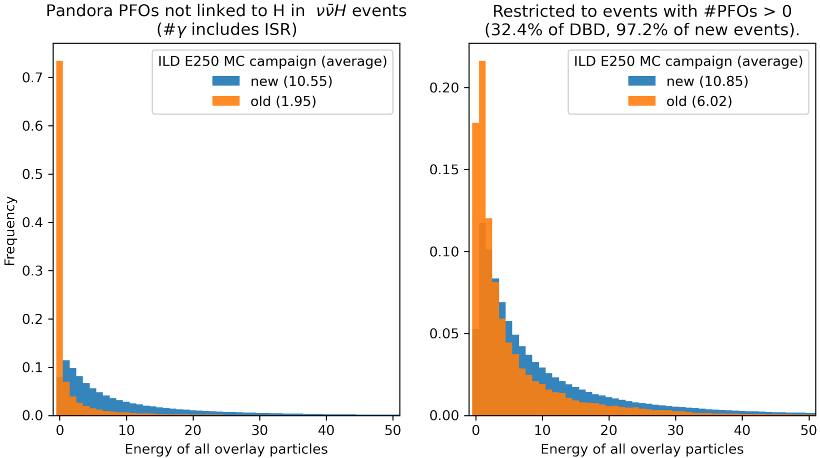
Frequency and size of the overlay increased, composition altered.





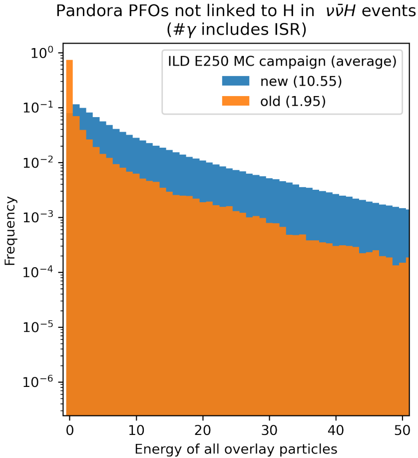
OVERLAY ENERGY

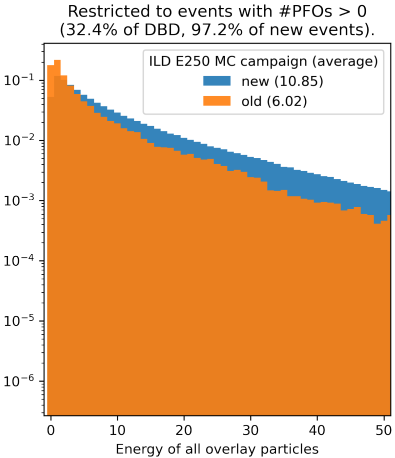
On average ≈ 10 GeV. $E_{\text{Higgs}}^{\text{mean}} \approx 130$ GeV.



OVERLAY ENERGY

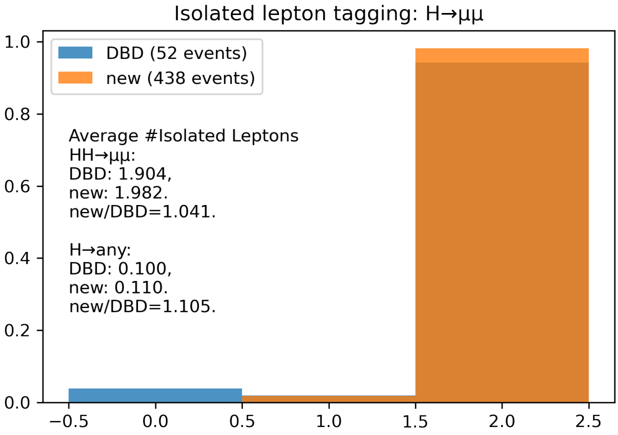
On average ≈ 10 GeV. $E_{\text{Higgs}}^{\text{mean}} \approx 130$ GeV.





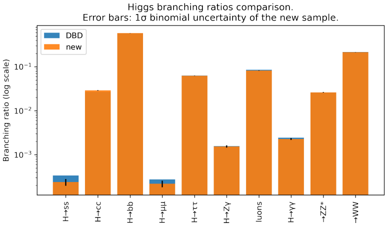
ISOLATED LEPTONS

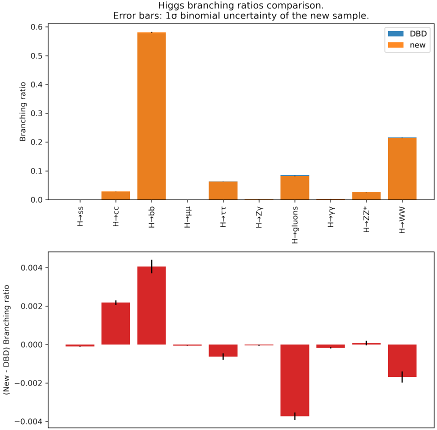
Using the IsolatedLeptonTaggingProcessor with DBD weights.



BRANCHING RATIOS

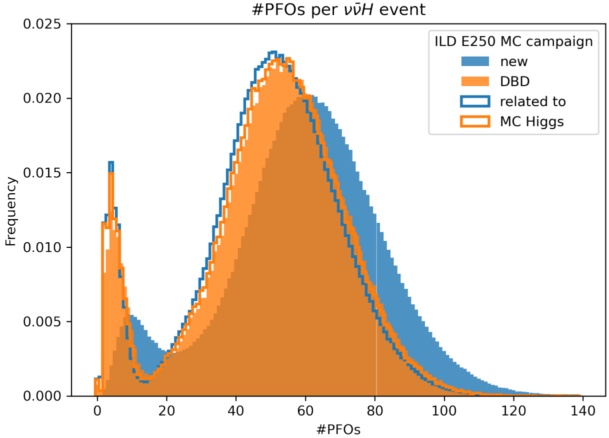
Differences (red) seem to be larger than statistical uncertainty.





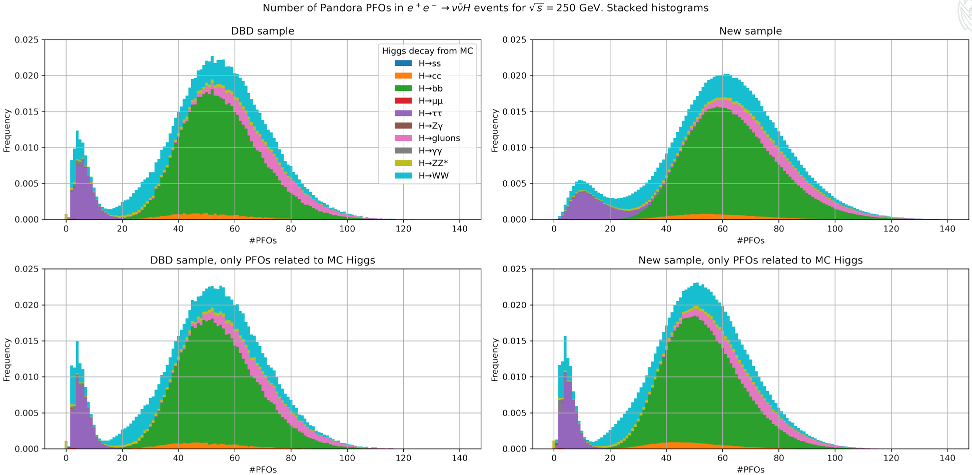
#PFOS PER HIGGS DECAY MODE

Increased overlay makes it harder to use *global* information.



#PFOS PER HIGGS DECAY MODE

Increased overlay makes it harder to use *global* information.

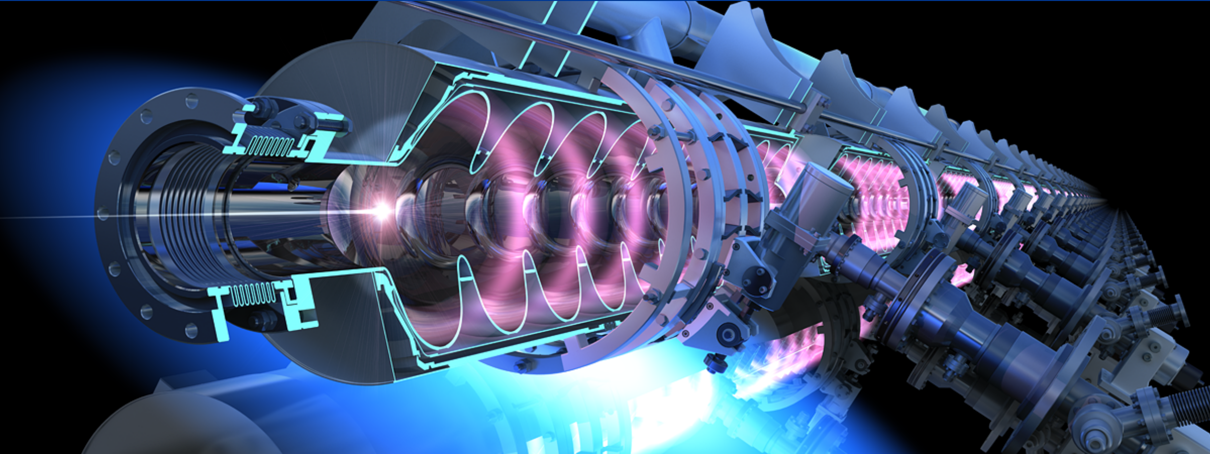


CONCLUSION

+ Smoother distributions due to increased sample size.
+ Reconstruction of the Higgs boson itself is comparable.
– Overlay should not be ignored (any more) by me :
  - Replace *global* variables by a more local version (e.g. `#PFOs` $\rightarrow$ `#PFOs in fat jet`) *or*
  - Have a customized, stricter particle definition that reduces the overlay within the analysis *and/or*
  - Adapt the selection cuts and retrain the MVA tools.

! Many thanks to all who are involved in providing the (new) samples!

4
BACK-UP



TWO TYPES OF VARIABLES

It is important to assign each observed momentum to the right parent particle.

Higgs-only variables

- E.g. M_H^{vis}, recoil to M_H^{vis}.
- But also **number of charged hadrons**.
- Differ between the Higgs decay modes.
- Same distr. for all four $Z \to l\bar{l}$ samples.
- Distributions taken from the *reference sample*.

Z-only variables

- E.g. M_{recoil}, M_Z.
- C.f. recoil mass technique.
- Independent of the Higgs boson decay (model).
- Distributions taken from Monte Carlo (MC) generated data.

TWO TYPES OF SAMPLES

Counting sample

- Count the number of events in the samples.
- Three samples are built : $Z \rightarrow e^+e^-$, $\tau^+\tau^-$ and $\nu_l\bar{\nu}_l$.
- Event selection based on both Z-only and Higgs-only variables.
- Z-only selection efficiency from MC.
- Higgs-only sel. eff. from reference sample.

Reference sample

- Extract the fraction of events passing a (Higgs-only) selection.
- Employed Higgsstrahlung events : $Z \rightarrow \mu^+\mu^-$.
- Event selection based on just the Z-only variables.
- Selection efficiency from MC.

UNCERTAINTY ON THE $\nu\bar{\nu}H$ CROSS SECTION

$$\sigma_{\nu\bar{\nu}H} = \frac{N_{\nu\bar{\nu}H}}{L} = \frac{N_{\nu\bar{\nu}H}}{BR(Z \to \nu\bar{\nu})\epsilon_Z^{\mu,e}\epsilon_H L}$$

$$\frac{\Delta\sigma_{\nu\bar{\nu}H}}{\sigma_{\nu\bar{\nu}H}} \approx \sqrt{\left(\frac{\Delta N_{\nu\bar{\nu}H}}{N_{\nu\bar{\nu}H}}\right)^2 + \left(\frac{\Delta\epsilon_H}{\epsilon_H}\right)^2}$$

$$= \sqrt{\frac{D_{\nu\bar{\nu}H}}{(N_{\nu\bar{\nu}H})^2} + \frac{D_Z^{\mu,e}}{(N_Z^{\mu,e})^2} + \frac{D_{H|Z}^{\mu,e}}{(N_{H|Z}^{\mu,e})^2} - \frac{2D_{H|Z}^{\mu,e}}{N_{H|Z}^{\mu,e}N_Z^{\mu,e}}}$$

Includes the systematic uncertainty from the selection (e.g. cut on $N_{\text{ch. hadr.}}$).
Assumption : Background distributions well known.

THE $\nu\bar{\nu}H$ BDT

Up to now trained an XGBoost BDT with the DBD MC samples. The criterium for the variables is to :

Only use Higgs boson remnants, nothing from the recoiling Z boson

+ $\delta\sigma_{H\nu\bar{\nu}} \approx 3.1\%$ achieved so far.

- $\text{N}_{\text{charged hadrons}}$, $\text{N}_{\text{neutral hadrons}}$, N_{γ}, N_{e}, N_{μ}, $\text{N}_{\text{isolated leptons}}$.
- $\text{M}_{\text{Higgs}}(=\text{M}_{\text{vis}})$, $\text{M}_{\text{Higgs-recoil}}$, $\cos(\theta_{\text{miss}})$.
- principle thrust, z-component of thrust axis, major thrust, minor thrust, sphericity, aplanarity, $\text{E}^{\text{max}}_{\text{isolated lepton}}$, $\cos(\theta_{\text{isolated lepton}})$.

and indep. of the Higgs production (WW-fusion, Higgsstrahlung).

+ $\delta\sigma_{H\nu\bar{\nu}} \approx 3.8\%$ achieved so far.

- $\text{N}_{\text{charged hadrons}}$, $\text{N}_{\text{neutral hadrons}}$, N_{γ}, N_{e}, N_{μ}, $\text{N}_{\text{isolated leptons}}$.
- $\text{M}_{\text{Higgs}}(=\text{M}_{\text{vis}})$.

THE **ILD DBD** 250 GEVDATA SET

$\sqrt{s} = 250$ GeV	cross section		N_{Gen}	
e^- polarisation	left	right	left	right
$\mu^+\mu^- H$	17.1 fb	11.0 fb	17.1k	11.0k
$e^+e^- H$	17.6 fb	11.2 fb	17.6k	11.2k
$\tau^+\tau^- H$	17.1 fb	11.0 fb	17.1k	11.0k
$\nu_l\bar{\nu}_l H$	128.6 fb	65.1 fb	0.13M	65.1k
2f_h	129.15 pb	71.27 pb	1.75M	1.43M
2f_l	46.51 pb	40.70 pb	2.63M	2.13M
4f_h	28.66 pb	966.2 fb	2.50M	0.24M
4f_l	8.89 pb	1.28 pb	2.25M	0.35M
4f_sl	31.12 pb	1.42 pb	4.43M	0.36M
$q\bar{q}H$	346.0 fb	222.0 fb	0.35M	0.22M
sum	244.86 pb	115.95 pb	14.09M	4.83M

- Large Standard Model event generation and detector simulation.
  $\rightarrow$ for Detailed Baseline Design (DBD).
- A particle flow algorithm (PandoraPFA) was run on simulated events in 2013.
- Process luminosities between ~ 20 fb^{-1} (Bhabha) and 1000 fb^{-1} (Higgs).