

Higgs to Invisible: MVA Second Pass

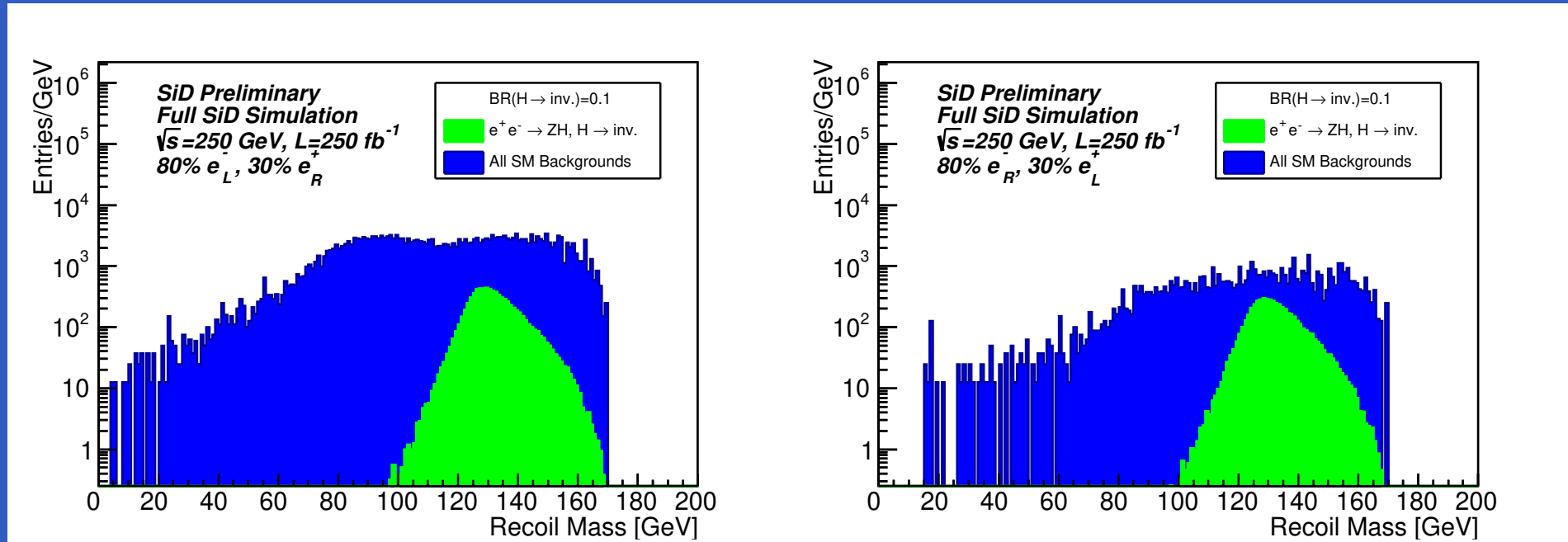


Chris Potter

University of Oregon

[Recall SiD Optimization, 17 Feb 2021] Full SiD Simulation

Signal is green with $BR(H \rightarrow \text{inv.})=0.10$, the Barklow DBD all_SM_background is blue.



Above, recoil mass after full signal selection. Below, cutflow yields and significances.

Requirement (Full)	$S(\text{LR})$	$B(\text{LR})$	$\frac{S}{\sqrt{S+B}}$	$S(\text{RL})$	$B(\text{RL})$	$\frac{S}{\sqrt{S+B}}$
$20 \leq p_T^{\text{vis}} \leq 70 \text{ GeV}$	1.25e+04	7.71e+06	4.48	8.84e+03	1.07e+06	8.53
$75 \leq m_{\text{vis}} \leq 105 \text{ GeV}$	1.16e+04	1.79e+06	8.63	8.21e+03	3.14e+05	14.5
$N_{\text{jet}} = 2$	1.16e+04	1.79e+06	8.63	8.21e+03	3.14e+05	14.5
$-0.9 \leq \cos \theta_{jj} \leq -0.2$	1.08e+04	8.68e+05	11.5	7.65e+03	1.78e+05	17.7
$110 \leq m_{\text{recoil}} \leq 150$	1.03e+04	3.6e+05	17	7.33e+03	8.39e+04	24.2

Full simulation (ILCSOFT v02-00-02, SID o2_v3) scaled from 250fb^{-1} to 900fb^{-1} .

[Recall SiD Optimization, 17 Feb 2021] Conclusions

- From work previously shown here...
 - ◆ The Tokyo/ILD study arXiv:2002.12048 is current ILC best for invisible Higgs decay.
 - ◆ This study appears to neglect a major background, $3f e\gamma \rightarrow eZ, \nu W$.
 - ◆ A simple SiD cut/count analysis neglecting $3f$ background gives consistent yields.
- From new work shown today here...
 - ◆ A rough first pass at MVA suggests sizeable significance enhancements.
 - ◆ Limitations of this first coarse look:
 - Some important input variables may have been overlooked.
 - All background categories were thrown at TMVA for training of one MVA.
 - The training sample `a11_SM_background` is statistics limited.
 - MLP and BDT structure and training parameters were not optimized.
- Work for the future....
 - ◆ Develop high statistics samples for each background $eZ, \nu W, WW, ZZ, Z\nu\nu$.
 - ◆ Develop separate MVA for each background with customized structure/training.
 - ◆ Combine MVA from each background in combined likelihood or MVA.

Work Since SiD Optimization, 17 Feb 2021

- Two additional inputs have been added to the BDT/MLP: y_{43} and y_{32} , the jet distance parameter cut value required to go from four to three jets, and three to two jets.
- Separate BDT/MLP are now trained on each background category:
 $eZ, \nu W, WW, ZZ, Z\nu\nu, We\nu$.
- High statistics samples are now used for training and testing the MLP/BDT: all available Whizard files for these backgrounds from the DBD study have been fully SiD simulated and used in the training and testing.
- BDT/MLP structure and training parameters in TMVA have been tuned, though results are robust against tuning away from the defaults.
- The BDT consistently outperform the MLP in all background categories, though the performance difference is small.
- The BDT outputs from each background category have been used as inputs into a new MVA for final use in the analysis: Likelihood, BDT and MVA yield equal performance.
- The Likelihood MVA using outputs from all twelve BDT outputs as inputs has been incorporated into the analysis as a standalone C++ generated by TMVA.
- Results running the full analysis chain plus the standalone MVA are consistent with expectations from the TMVA analysis: the background is decimated.

Combined MVA: 12 inputs are BDT outputs

