

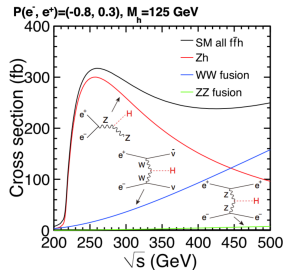
Identification of leptons inside jets at future Higgs factories

DPG Rehearsals

Leonhard Reichenbach

Hamburg, 14.03.2022

Introduction



Goal:

> precision Higgs measurements

Conditions at $\sqrt{s} = 250$ GeV:

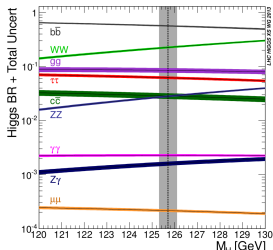
> $e^+e^- \rightarrow ZH$ has the highest Higgs production cross section

> $H \rightarrow b\bar{b}$ most common decay

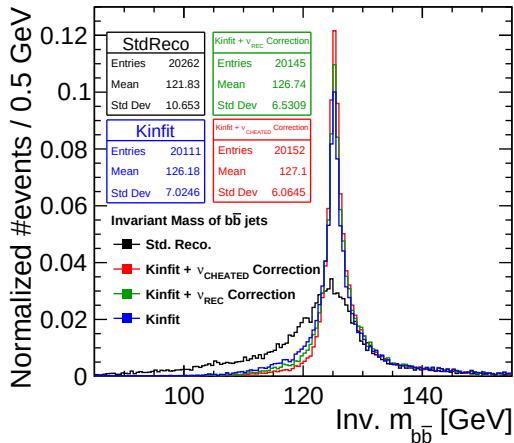
But:

> semi-leptonic decays (SLD) $b/c \rightarrow \ell\nu X$ in 2/3 of those events ($\ell = e, \mu$)

> missing neutrino energy $\Rightarrow$ bad for precision



Motivation: Neutrino Correction



Previous work by Yasser Radkhorrani:

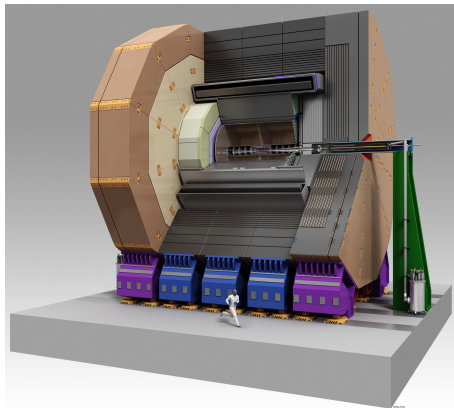
- > reconstruct p_ν from decay kinematics
- > huge improvement of invariant Higgs mass in combination with a kinematic fit
- > still depends on Monte Carlo truth information to identify SLD leptons

Details in a [recent talk](#)

Technical Details

Analysis done in the context of a mature particle flow detector concept, the International Large Detector (ILD) at the International Linear Collider (ILC)

- > data from full (Geant4) simulation
- > after standard ILD reconstruction including creation of Pandora ParticleFlow Objects (PFOs)
- > select $ZH \rightarrow \mu^+ \mu^- b\bar{b}$ events on Monte Carlo level
- > muons from Z decay isolated
- > remaining PFOs clustered into 2 jets



Approach

Goal:

- > identify e/μ from semi-leptonic b/c decays in $ZH \rightarrow \mu^+ \mu^- b\bar{b}$ at 250 GeV

Approach:

- > don't reinvent the wheel, use existing PID to select muons and electrons
- > build a classifier to decide if a selected e/μ is from a SLD

proof of concept:

- > select PFOs based on rudimentary Pandora PID information
- > use jet PFO observables as classifier input
- > ROOT TMVA based classifiers
 - boosted decision trees (BDT)
 - naive Bayes classifier (Likelihood)
- > out-of-the-box performance, not yet optimized

Signal and Backgrounds

sld ℓ candidate selection:

- > iterate over charged PFOs from the jets
- > consider PFOs ID'd by Pandora as ℓ
- > only keep PFOs not from primary vertex (ν correction constraint)

signal if:

- > PFO has a corresponding MC ℓ which comes from a semi-leptonic b, c decay

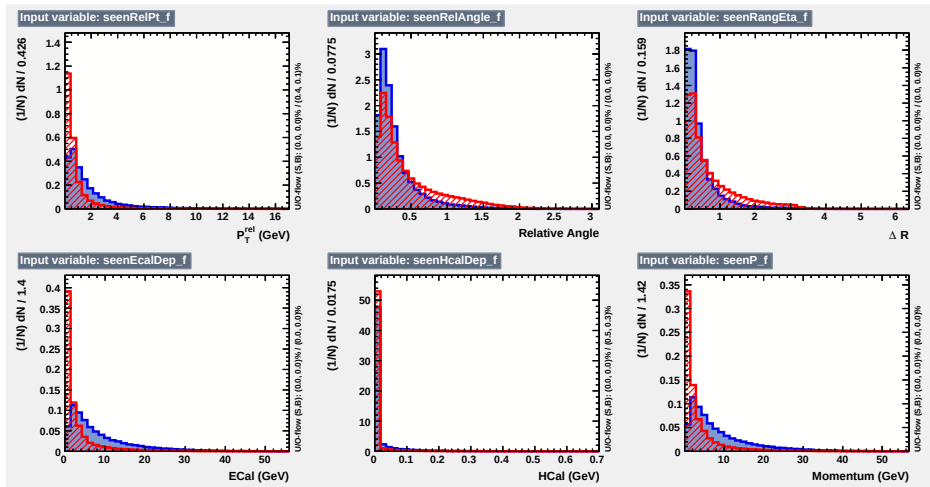
else: background

selection impact

	efficiency (ε_{pre})	purity
e	89.8%	18.2%
μ	69.4%	40.8%
$\mu + \pi$	99.7%	3%

Additional selection of pions substantially improves muon pre-selection efficiency but degrades classifier

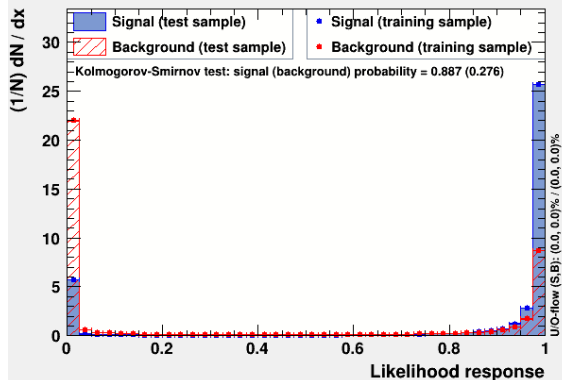
e Input Distributions



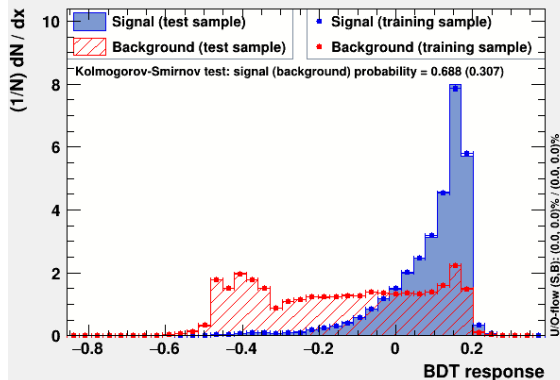
Additionally: impact parameters & significances, E/P

e Classifier Responses

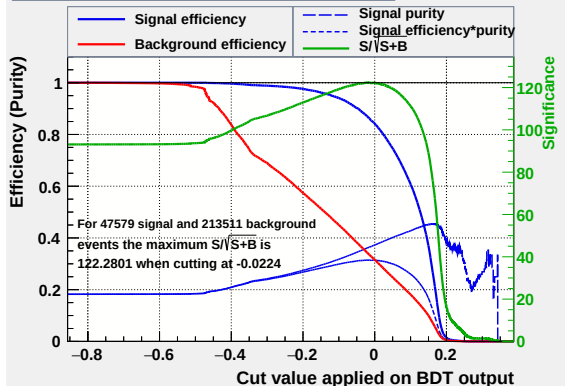
TMVA overtraining check for classifier: Likelihood



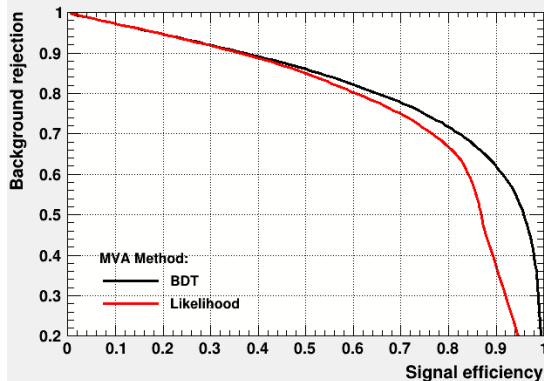
TMVA overtraining check for classifier: BDT



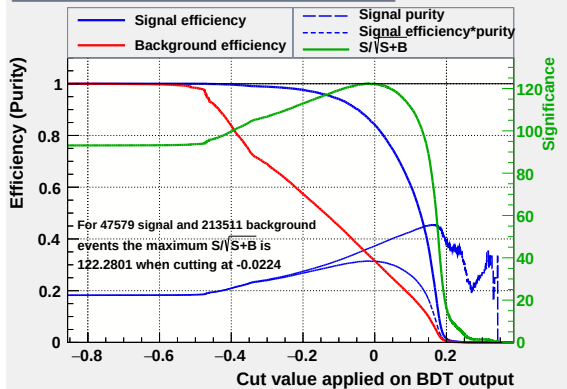
Cut efficiencies and optimal cut value



Background rejection versus Signal efficiency



Cut efficiencies and optimal cut value



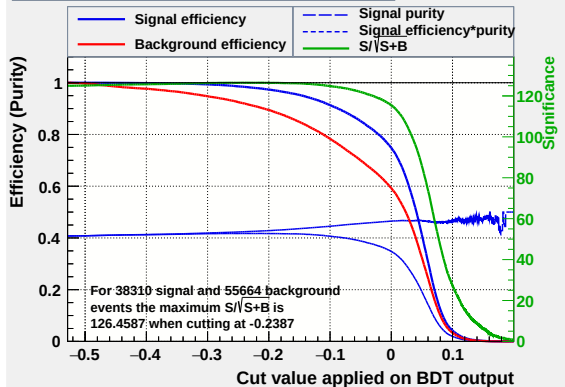
Results at “optimal” cut value:

true/BDT	pos	neg
pos	41491	6088
neg	73641	139870

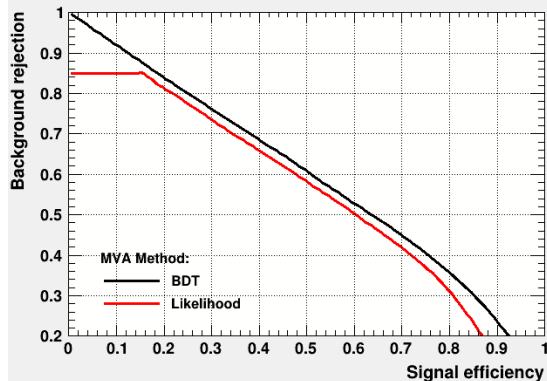
- > classification efficiency: $\epsilon_{\text{class}} = 87.2\%$
- > total efficiency: $\epsilon = \epsilon_{\text{pre}} * \epsilon_{\text{class}} = 78.3\%$
- > background efficiency: $\epsilon_{\text{bkg}} = 34.5\%$
- > purity: 36% (18.2% in the sample)

lots of room for improvement

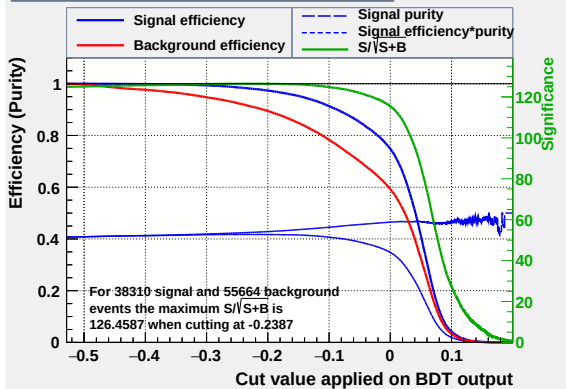
Cut efficiencies and optimal cut value



Background rejection versus Signal efficiency



Cut efficiencies and optimal cut value



Results at “optimal” cut value:

true/BDT	pos	neg
pos	37690	620
neg	51139	4525

- > classification efficiency: $\varepsilon_{\text{class}} = 98.4\%$
- > total efficiency: $\varepsilon = 68.3\%$
- > background efficiency: $\varepsilon_{\text{bkg}} = 91.9\%$
- > purity: 42% (40.8% in the sample)

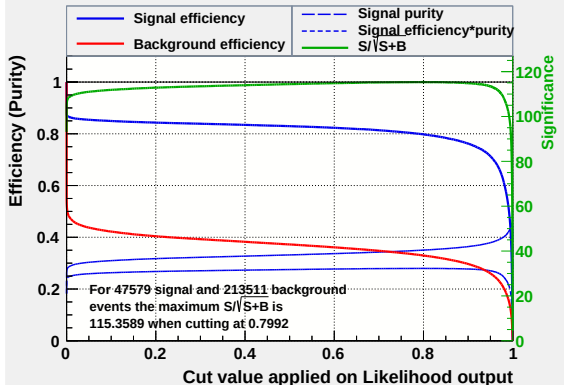
Conclusions

- > ID of SLD leptons using Pandora PID in combination with an out-of-the-box BDT on basic jet PFO observables is possible
- > However, much room for improvement
 - find better discriminating observables
 - fine-tune classifier parameters
 - incorporate pions to increase SLD muon efficiency
 - find out how much purity is needed
- > Future: implement the classifier to be usable in the neutrino correction

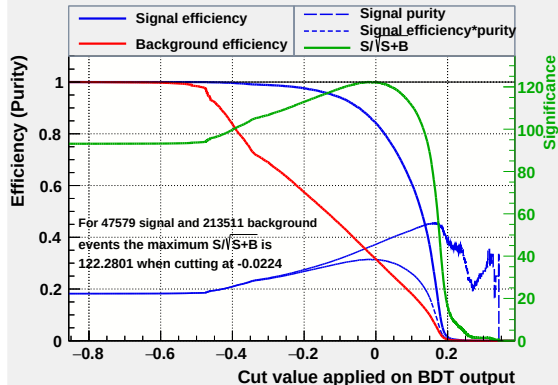
Backup

e Cut Efficiencies

Cut efficiencies and optimal cut value

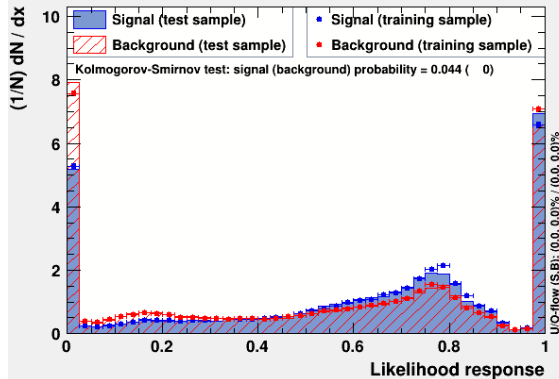


Cut efficiencies and optimal cut value

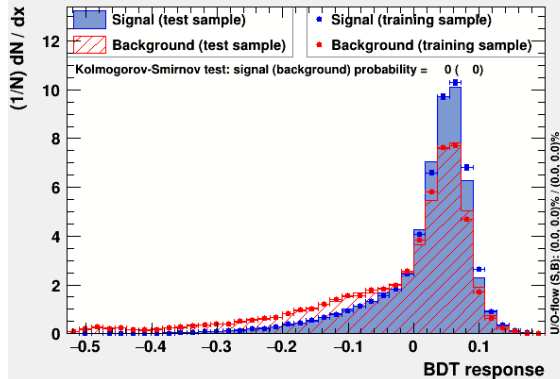


μ Classifier Responses

TMVA overtraining check for classifier: Likelihood

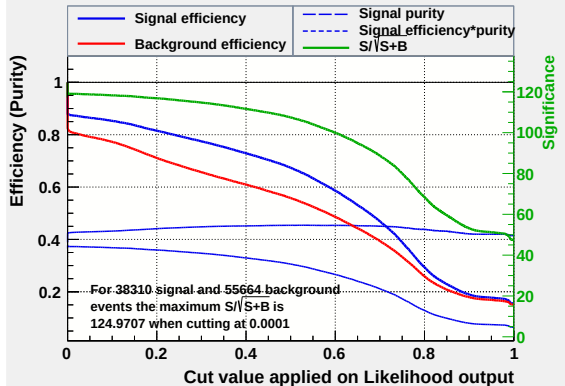


TMVA overtraining check for classifier: BDT

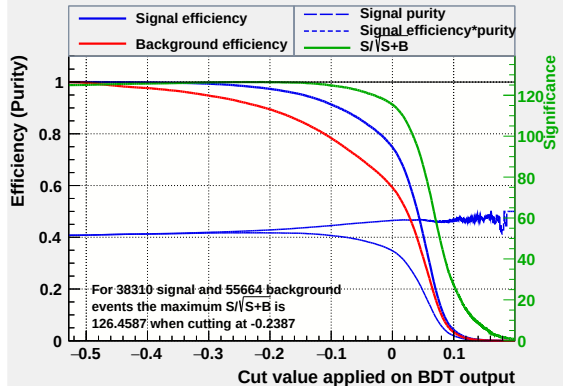


μ Cut Efficiencies

Cut efficiencies and optimal cut value

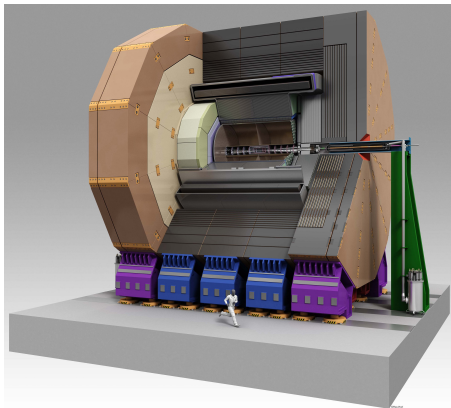


Cut efficiencies and optimal cut value



International Large Detector (ILD)

A ParticleFlow Detector



Detector concept for the International Linear Collider

- > momentum resolution:

$$\sigma_{\frac{1}{p_T}} \sim 2 \times 10^{-5} \text{ GeV}^{-1}$$

- > impact parameter:

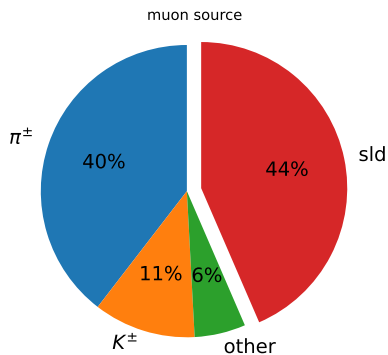
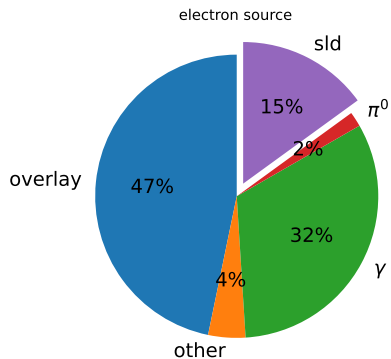
$$\sigma_{d_0} \sim 5 \mu\text{m}$$

- > jet energy resolution:

$$\frac{\sigma_{E_j}}{E_j} \sim 3 - 4\%$$

- > Optimized for Particle Flow

Backgrounds



jet PFO source/parent from MC data