Identification of leptons inside jets at future Higgs factories

DPG Spring Meeting

Leonhard Reichenbach, Yasser Radkhorrami, Jenny List

Heidelberg (virtual), 22.03.2022

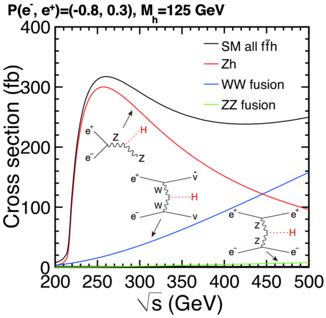
HELMHOLTZ

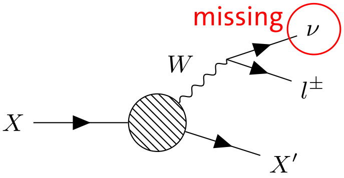
FTX

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

Introduction





Goal:

> precision Higgs measurements

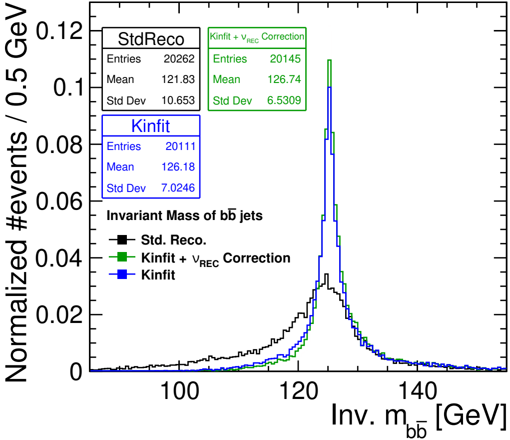
Conditions at $\sqrt{s} = 250\,\text{GeV}$:

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

But:

> $H \to b\bar{b}$ most common decay
> semi-leptonic decays (SLD) $b/c \to \ell\nu X$ in 2/3 of those events ($\ell = e, \mu$)
> missing neutrino energy $\Rightarrow$ bad for precision

Motivation: Neutrino Correction



Previous work by Yasser Radkhorrami:

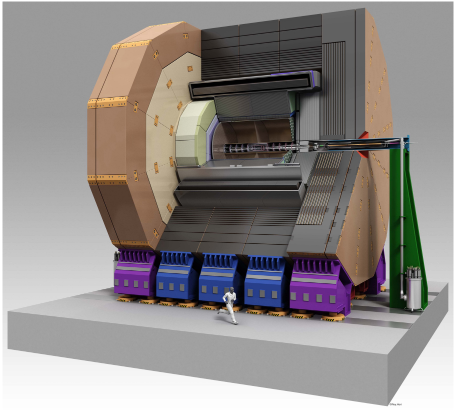
- reconstruct p_ν from decay kinematics
- huge improvement of invariant Higgs mass in combination with a kinematic fit
- last missing piece: SLD lepton identification!

Details in a recent talk (agenda.linearcollider.org/event/9514)

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 large scale Monte Carlo production and full (Geant4) simulation
- 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 $ZH \to \mu^+\mu^- b\bar{b}$ 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 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

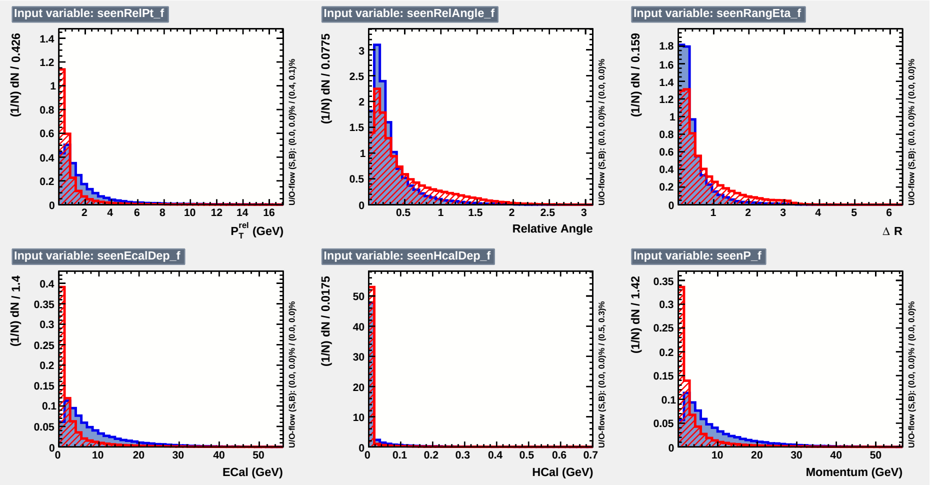
else: background

selection impact

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

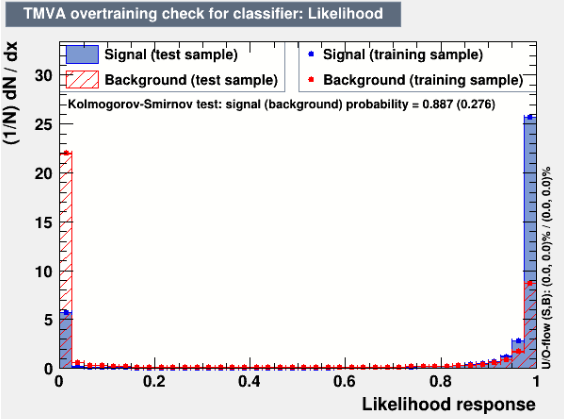
Additional selection of pions substantially improves muon pre-selection efficiency but degrades classifier

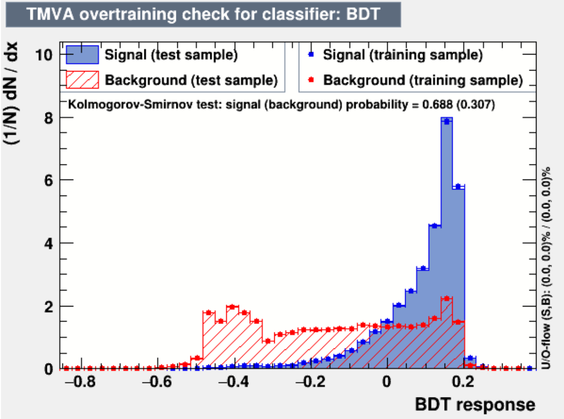
e Input Distributions



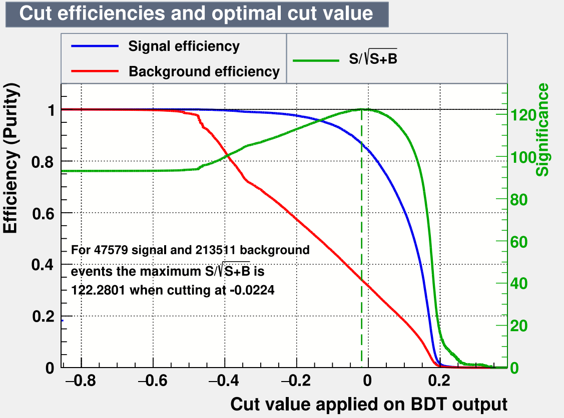
Additionally: impact parameters & significances, E/P

e Classifier Responses

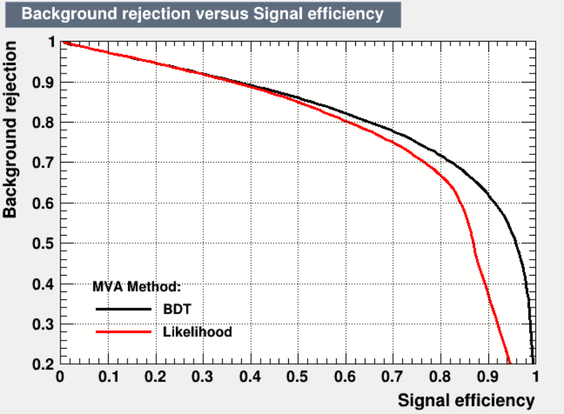




e Results

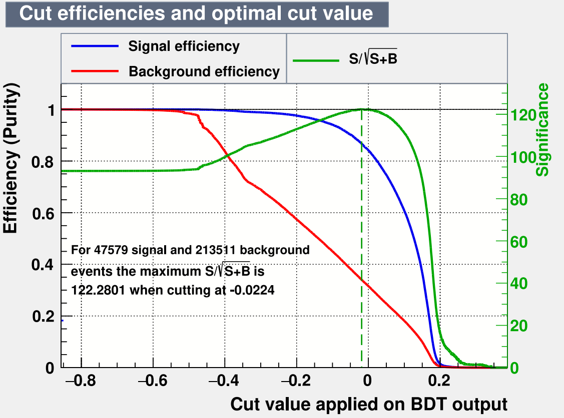


Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
S/√S+B
For 47579 signal and 213511 background
events the maximum S/√S+B is
122.2801 when cutting at -0.0224
Efficiency (Purity)
Significance
Cut value applied on BDT output
1
0.8
0.6
0.4
0.2
0
120
100
80
60
40
20
0
–0.8
–0.6
–0.4
–0.2
0
0.2



Background rejection versus Signal efficiency
MVA Method:
BDT
Likelihood
Background rejection
Signal efficiency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

e Results



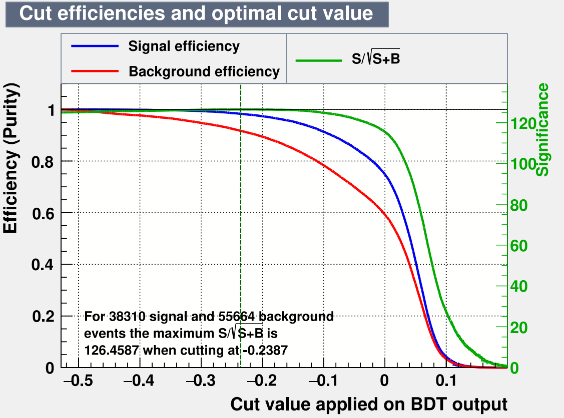
Results at "optimal" cut value:

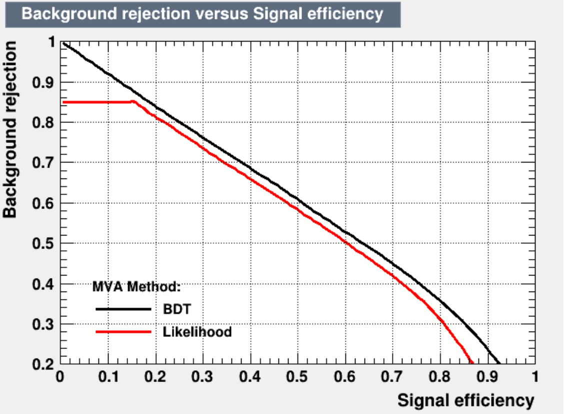
true/BDT	pos	neg
pos	41491	6088
neg	73641	139870

- classification efficiency: $\varepsilon_{\text{class}} = 87.2\%$
- total efficiency: $\varepsilon = \varepsilon_{\text{pre}} * \varepsilon_{\text{class}} = 78.3\%$
- background rejection: $1 - \varepsilon_{\text{bkg}} = 65.5\%$
- purity: 36% (18.2% in the sample)

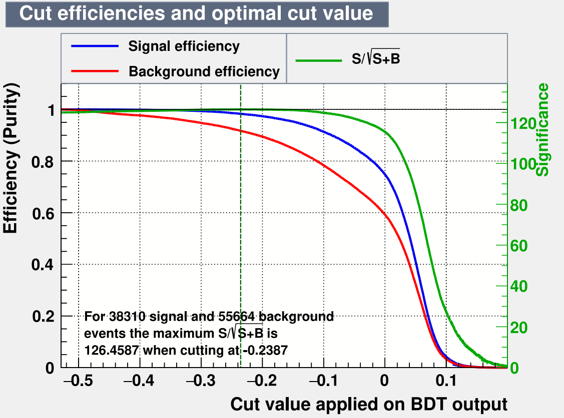
lots of room for improvement

μ Results


Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
S/√S+B
Efficiency (Purity)
Significance
For 38310 signal and 55664 background events the maximum S/√S+B is 126.4587 when cutting at -0.2387
Cut value applied on BDT output


Background rejection versus Signal efficiency
Background rejection
MVA Method:
BDT
Likelihood
Signal efficiency

μ Results



Results at “optimal” cut value:

true/BDT	pos	neg
pos	37690	620
neg	51139	4525

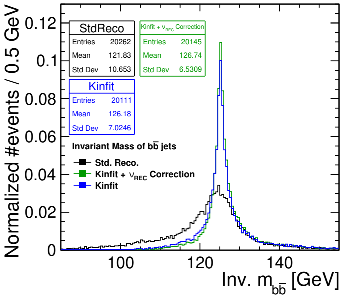
- classification efficiency: $\varepsilon_{\text{class}} = 98.4\%$
- total efficiency: $\varepsilon = 68.3\%$
- background rejection: $1 - \varepsilon_{\text{bkg}} = 8.1\%$
- purity: 42% (40.8% in the sample)

Conclusions

> The presented approach identifies
  - SLD electrons with 78.3% efficiency and 36% purity
  - SLD muons with 68.3% efficiency and 42% purity

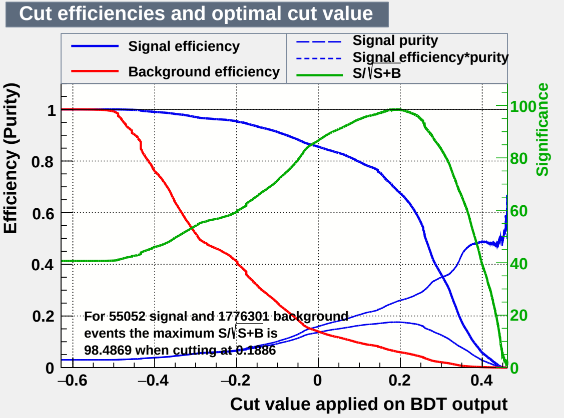
> However, much room for improvement
  - find better discriminating observables
  - fine-tune classifiers
  - incorporate pions to increase SLD muon efficiency

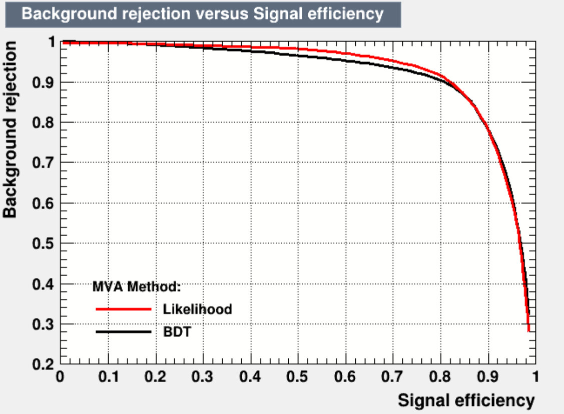
> Future: evaluate the classifier in the neutrino correction



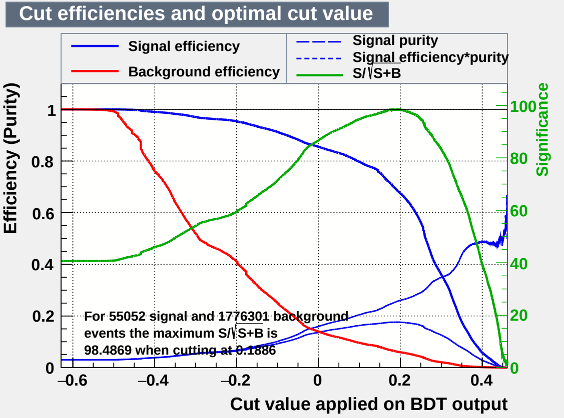
Backup

$\mu + \pi$ Results





$\mu + \pi$ Results



Results at "optimal" cut value:

true/BDT	pos	neg
pos	38133	16919
neg	111782	1664519

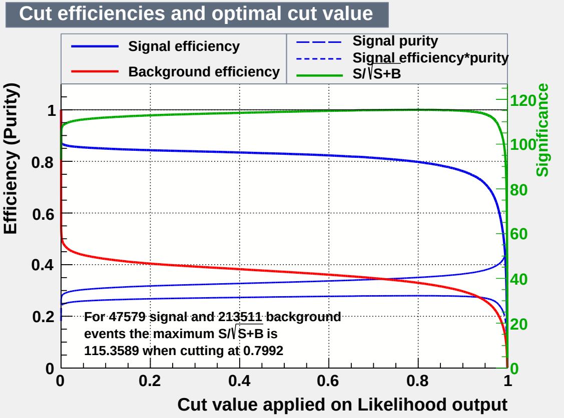
> classification efficiency: $\varepsilon_{\text{class}} = 69.3\%$

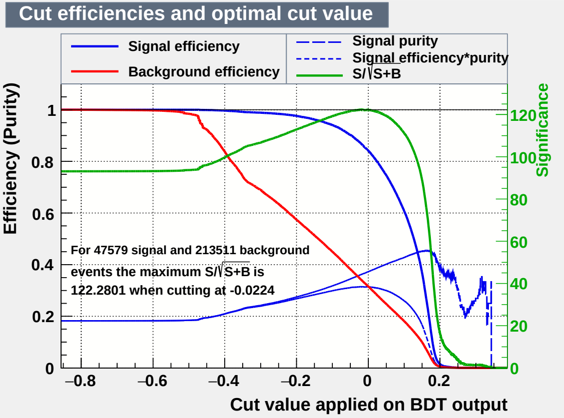
> total efficiency: $\varepsilon = 69\%$

> background rejection: $1 - \varepsilon_{\text{bkg}} = 93.7\%$

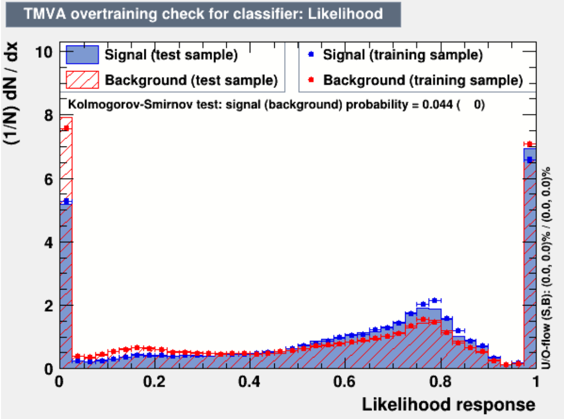
> purity: 25% (3% in the sample)

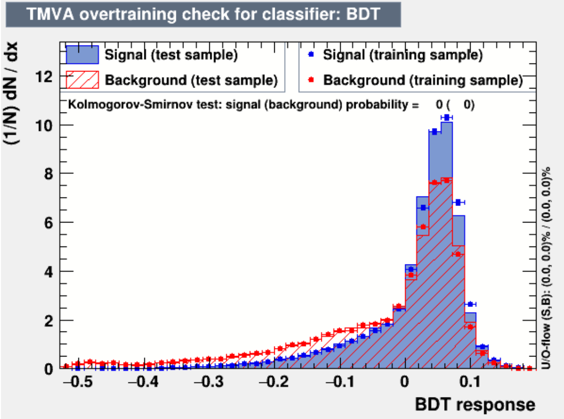
e Cut Efficiencies



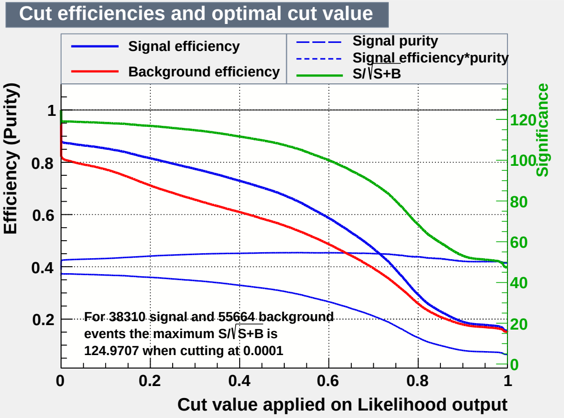


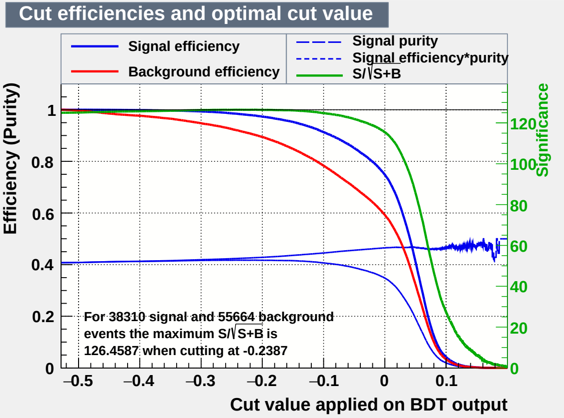
μ Classifier Responses



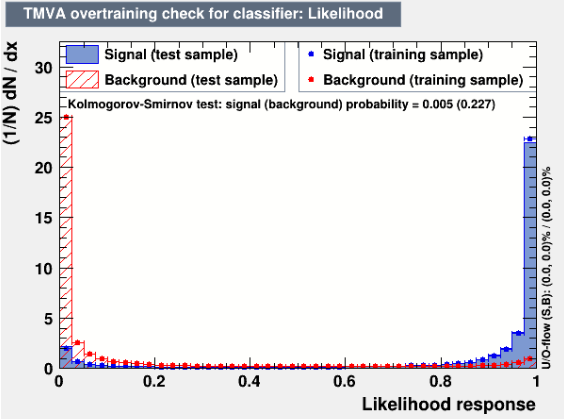


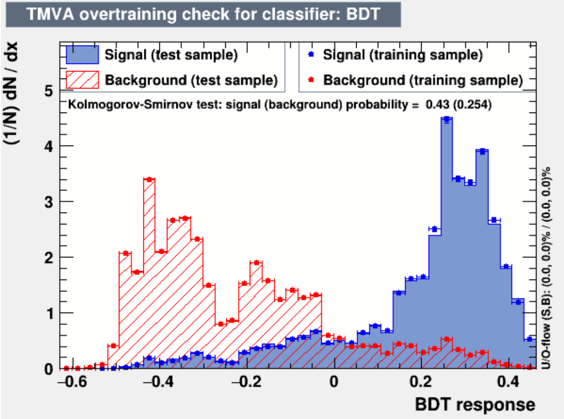
μ Cut Efficiencies



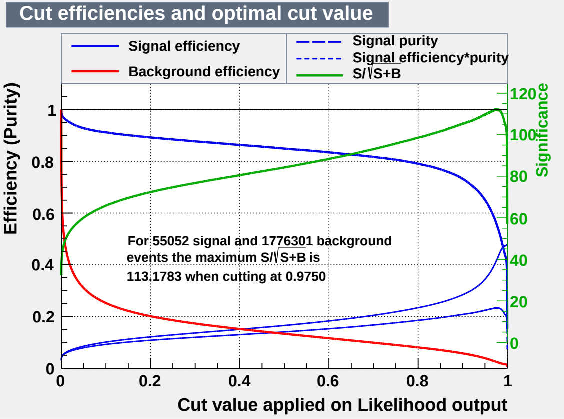


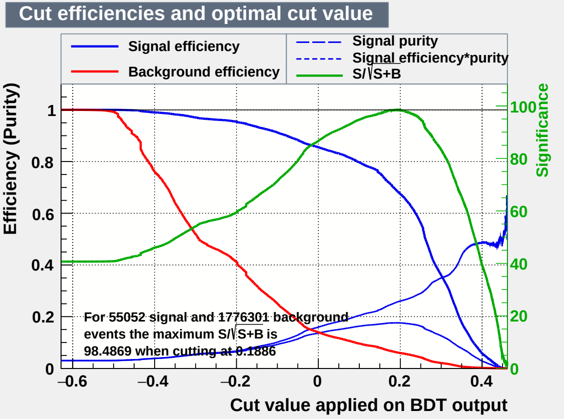
$\mu + \pi$ Classifier Responses





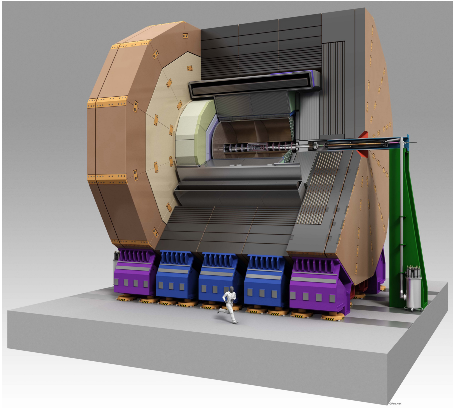
$\mu + \pi$ Cut Efficiencies





International Large Detector (ILD)

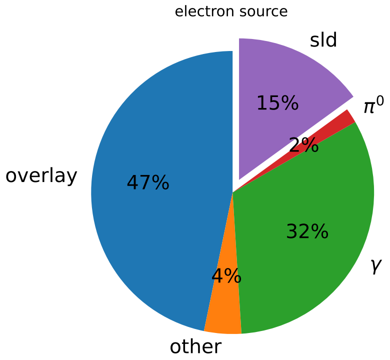
A ParticleFlow Detector

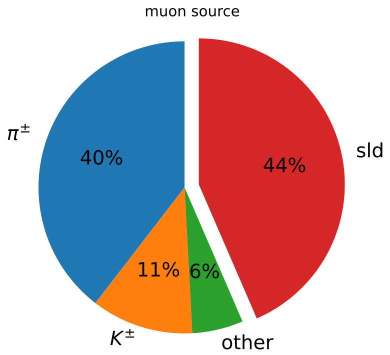


Detector concept for the International Linear Collider

- momentum resolution:
  $\sigma_{\frac{1}{p_T}} \sim 2{\times}10^{-5}\,\mathrm{GeV}^{-1}$
- impact parameter:
  $\sigma_{d_0} \sim 5\,\mu\mathrm{m}$
- jet energy resolution:
  $\frac{\sigma_{E_j}}{E_j} \sim 3 - 4\%$
- Optimized for Particle Flow

Backgrounds





jet PFO source/parent from MC data