UNIVERSITÀ DI PAVIA
Dipartimento di Fisica

Color sensitive variables for Higgs->bb tagger

Giulia Manco, Alberto Rescia, Daniela Rebuzzi, Fabrizio Parodi, Giovanni Stagnitto, Simone Marzani, Andrea Coccaro, Charanjit K. Khosa, Luca Cavallini

LFC22: Strong interaction from QCD to new strong dynamics at LHC and Future Colliders

INFN
Istituto Nazionale di Fisica Nucleare

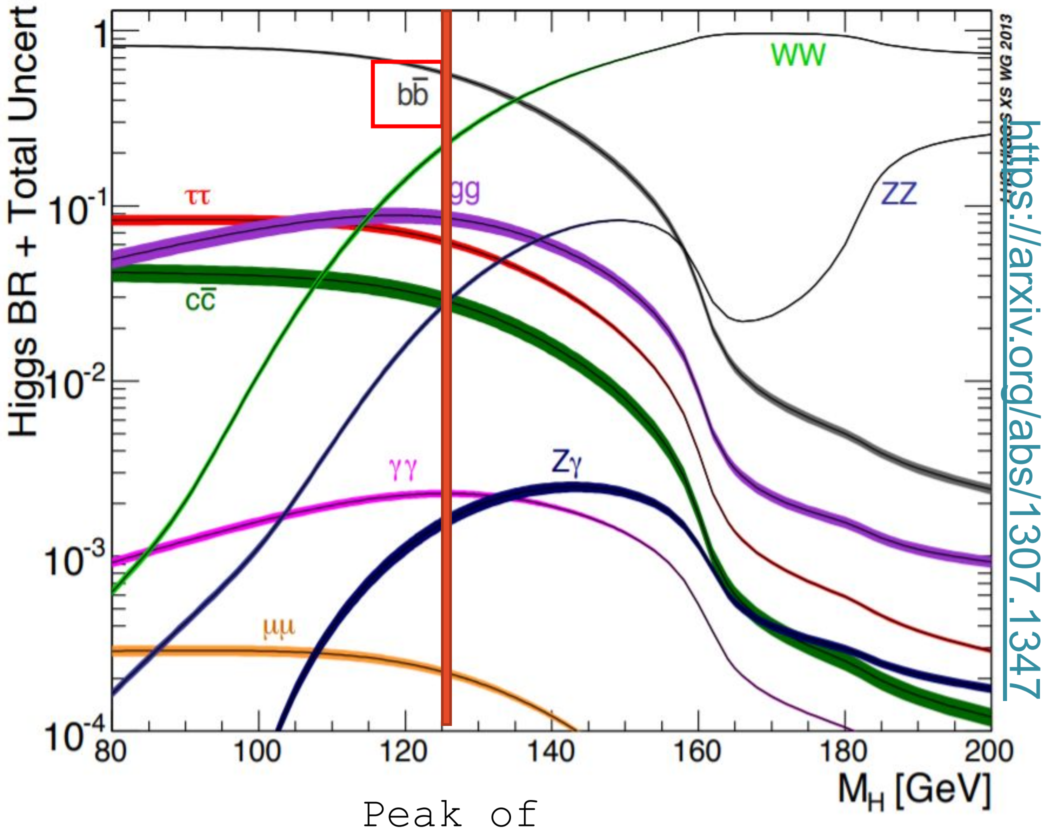
Xbb tagger

$$H \to b\bar{b}$$

PRO
Max probability
Study **third fermion generation coupling**

CONTRO
Irreducible QCD background (**g** → **bb**)

Xbb tagger for discrimination

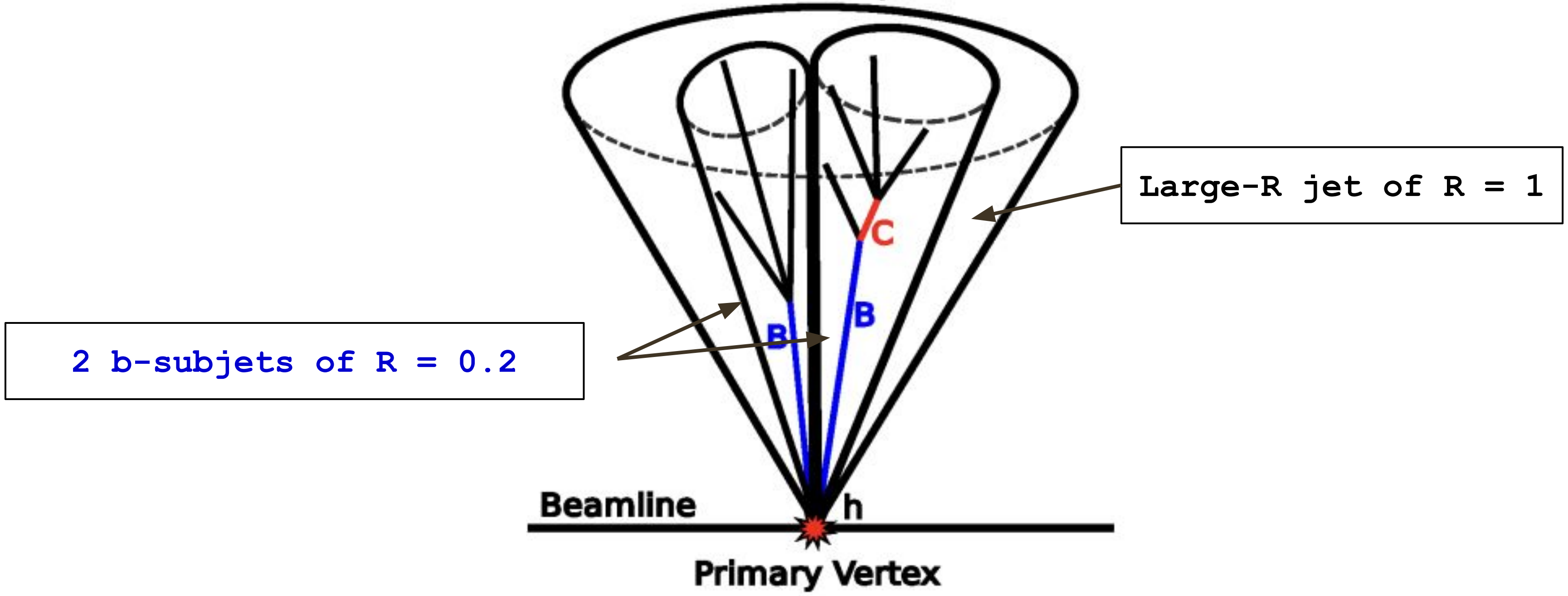


Peak of Higgs mass

https://arxiv.org/abs/1307.1347

Xbb tagger

Algorithms act to perform reconstruction in **boosted Higgs → bb** topologies using **jet substructure** information



In this presentation:

Phenomenological studies with color sensitive variables for Xbb tagger

https://arxiv.org/abs/2112.09650

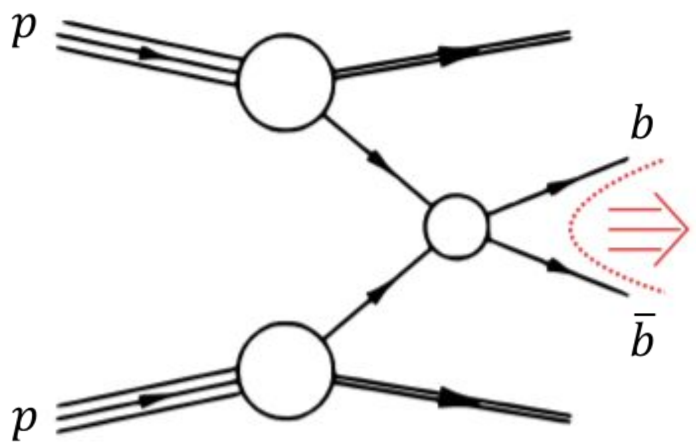
Tagging the Higgs boson decay to bottom quarks with colour-sensitive observables and the Lund jet plane

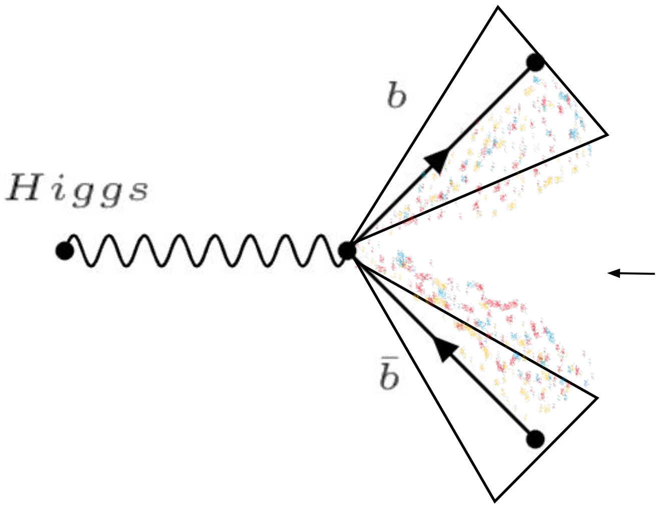
Luca Cavallini, Andrea Coccaro, Charanjit K. Khosa, Giulia Manco, Simone Marzani, Fabrizio Parodi, Daniela Rebuzzi, Alberto Rescia, Giovanni Stagnitto

We study the problem of distinguishing b-jets stemming from the decay of a colour singlet, such as the Higgs boson, from those originating from the abundant QCD background. In particular, as a case study, we focus on associate production of a vector boson and a Higgs boson decaying into a pair of b-jets, which has been recently observed at the LHC. We consider the combination of several theory-driven observables proposed in the literature, together with Lund jet plane images, in order to design an original Hbb tagger. The observables are combined by means of standard machine learning algorithms, which are trained on events obtained with fast detector simulation techniques. We find that the combination of high-level single-variable observables with the Lund jet plane provides an excellent discrimination performance. We also study the dependence of the tagger on the invariant mass of the decaying particles, in order to assess the extension to a generic Xbb tagger.

Color connections

SIGNAL



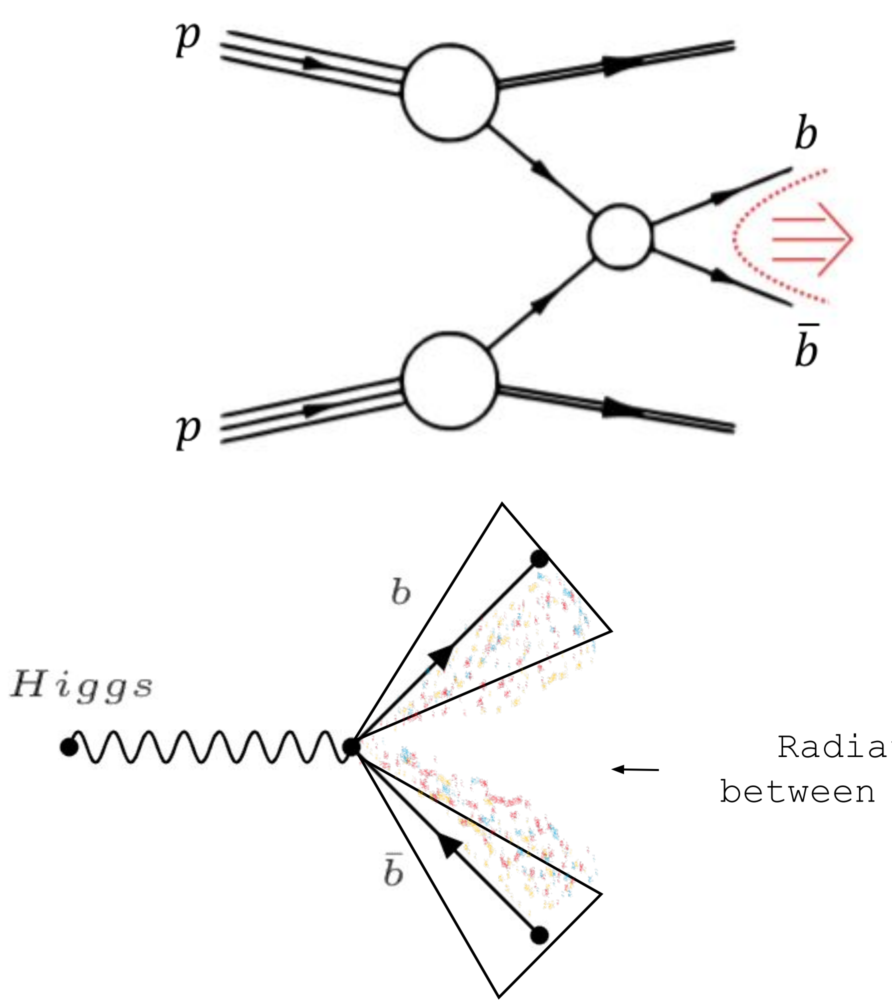


Radiation
between b-jets

https://arxiv.org/pdf/1001.5027.pdf

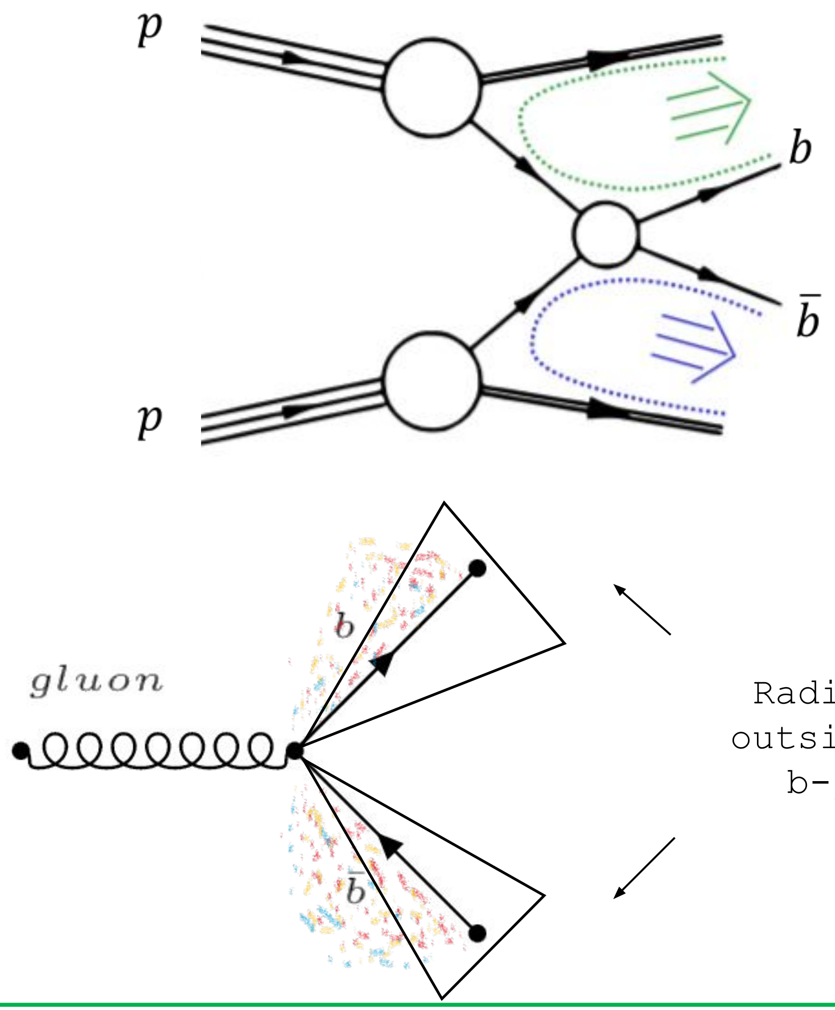
Color connections

SIGNAL



Radiation
between b-jets

BACKGROUND

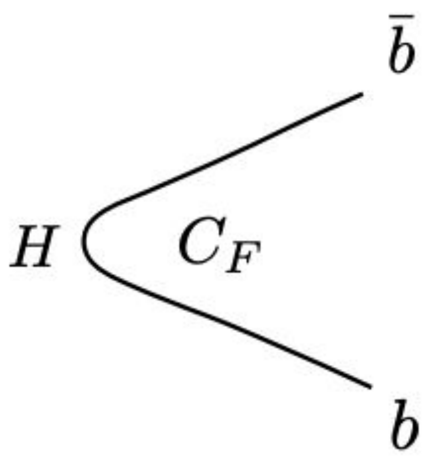


Radiation
outside the
b-jets

Color connections

Signal: H→bb + single low-energy gluon

$$|\mathcal{M}_S|^2 = C_F \frac{n_a \cdot n_b}{(n_a \cdot k)(n_b \cdot k)}$$

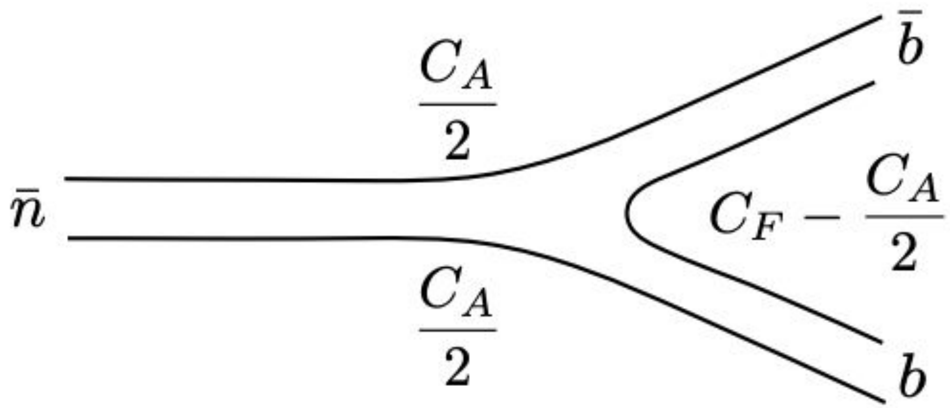


Direction of color flow for the **singlet** configuration.
Emission **inside** the two b-jets

Background: g→bb + single low-energy gluon

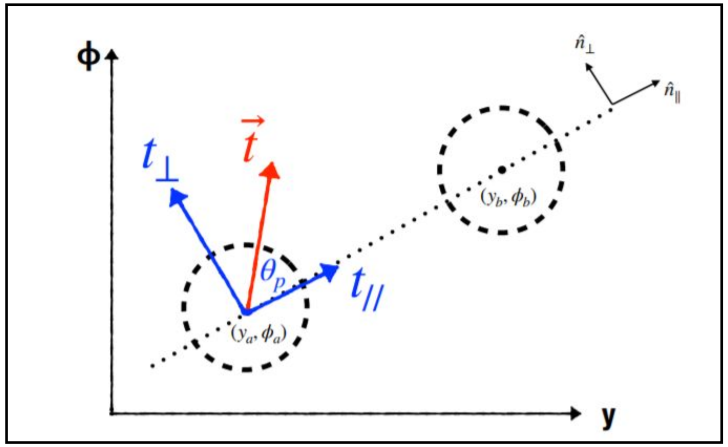
$$|\mathcal{M}_\mathcal{B}|^2 = C_\mathcal{B} \frac{n_a \cdot n_b}{(n_a \cdot k)(n_b \cdot k)} + \widetilde{C}_\mathcal{B} \left(\frac{n_a \cdot \bar{n}}{(n_a \cdot k)(\bar{n} \cdot k)} + \frac{n_b \cdot \bar{n}}{(n_b \cdot k)(\bar{n} \cdot k)} \right)$$

$C_\mathcal{B} = C_F - C_A/2$ $\quad\quad$ $\widetilde{C}_\mathcal{B} = C_A/2$



Direction of color flow for the **octet** configuration in the collinear limit.
Favored emission **outside** the two b-jets

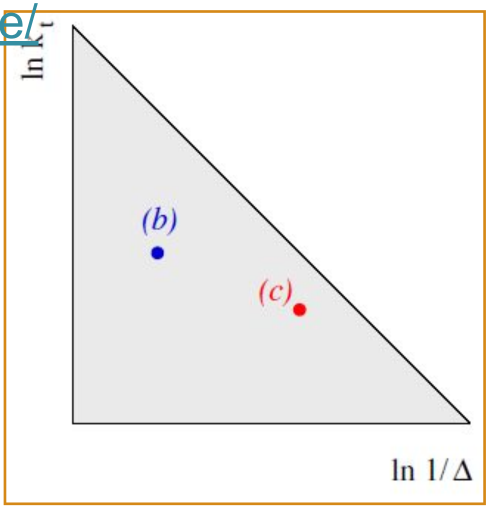
Color sensitive variables



PULL VECTOR

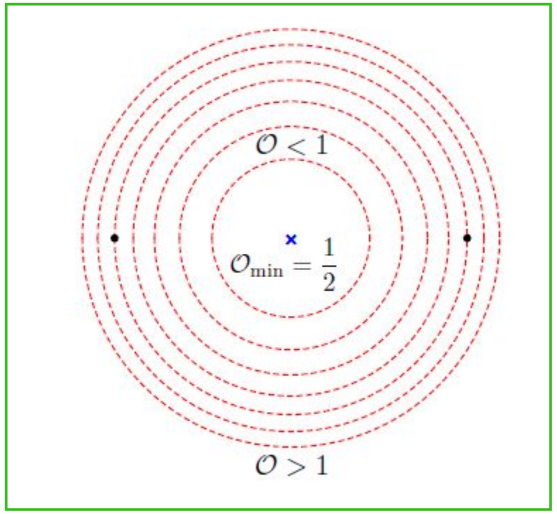
https://inspirehep.net/literature/1764711

Variables sensitive to color flow



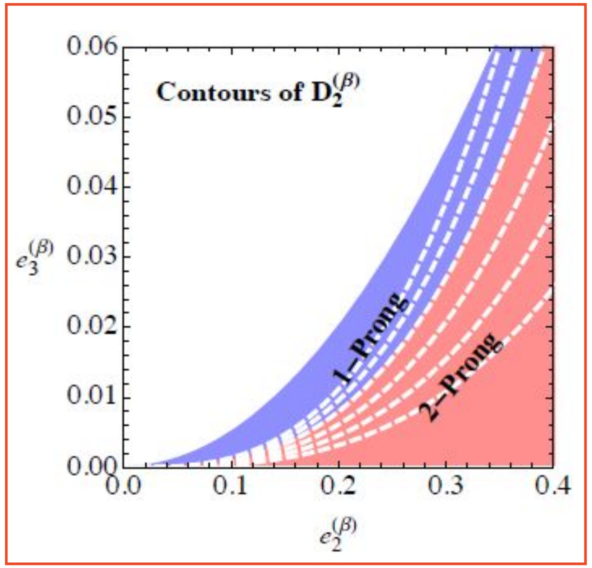
LUND PLANE

https://arxiv.org/abs/1807.04758



COLOR RING

https://arxiv.org/abs/2006.10480

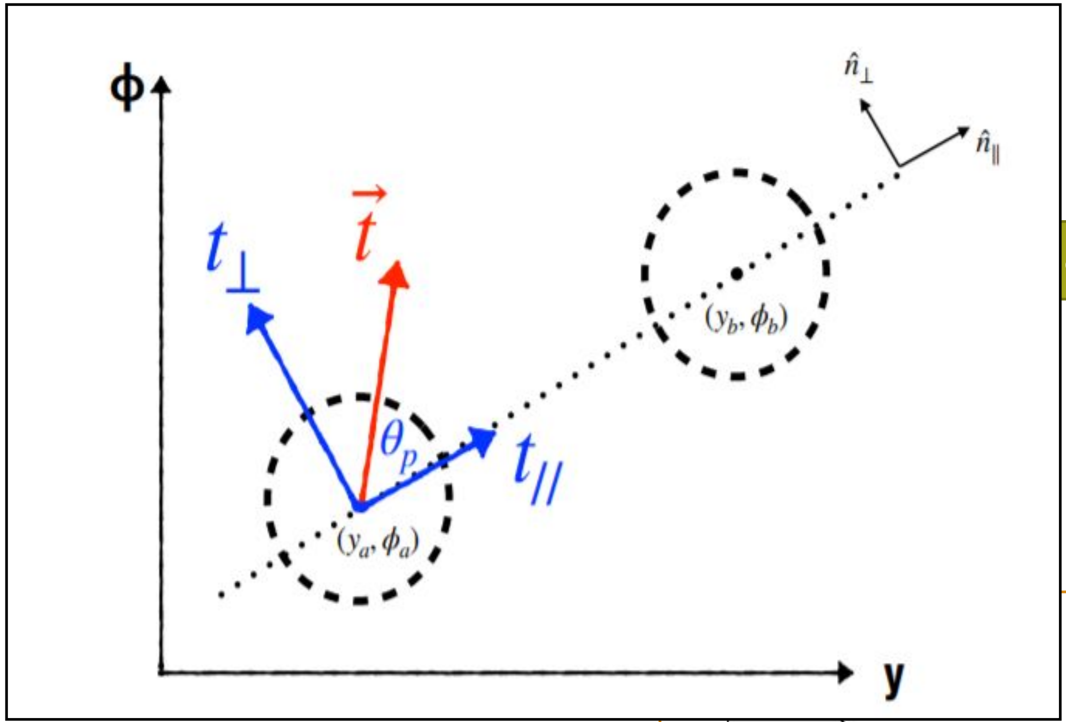


D_2

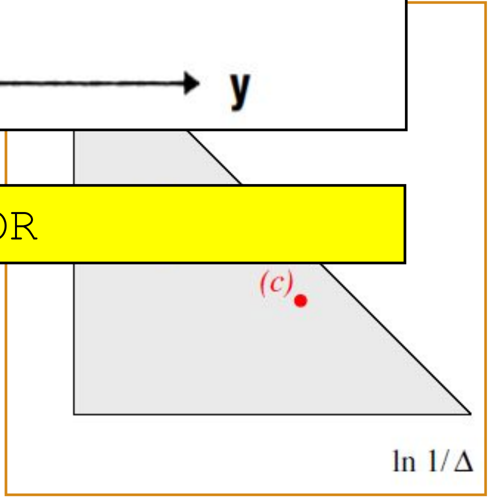
https://arxiv.org/abs/1409.6298

color sensitive variables

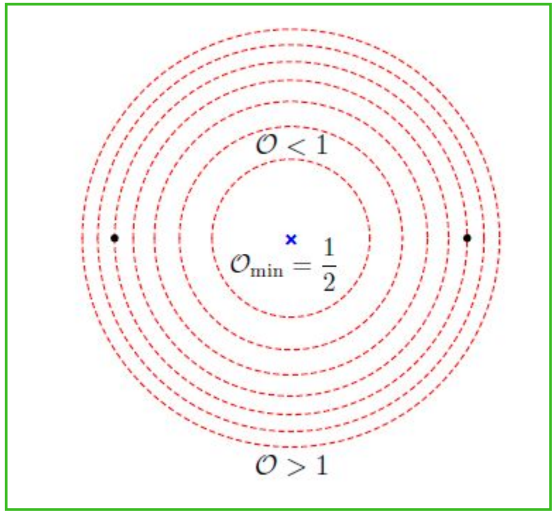
es sensitive to color flow



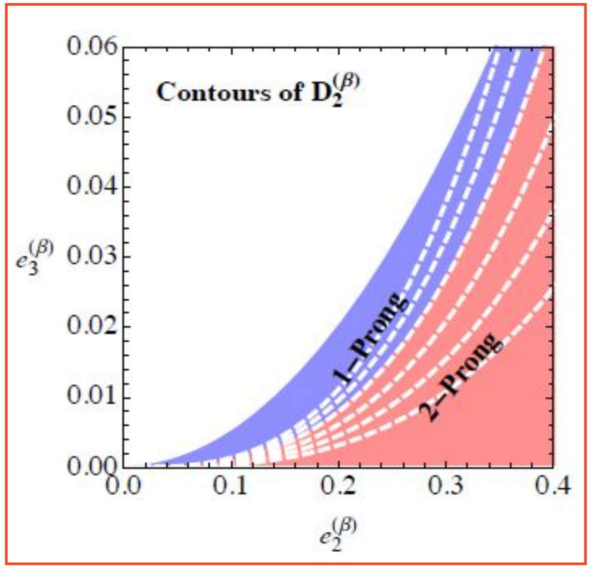
PULL VECTOR



LUND PLANE

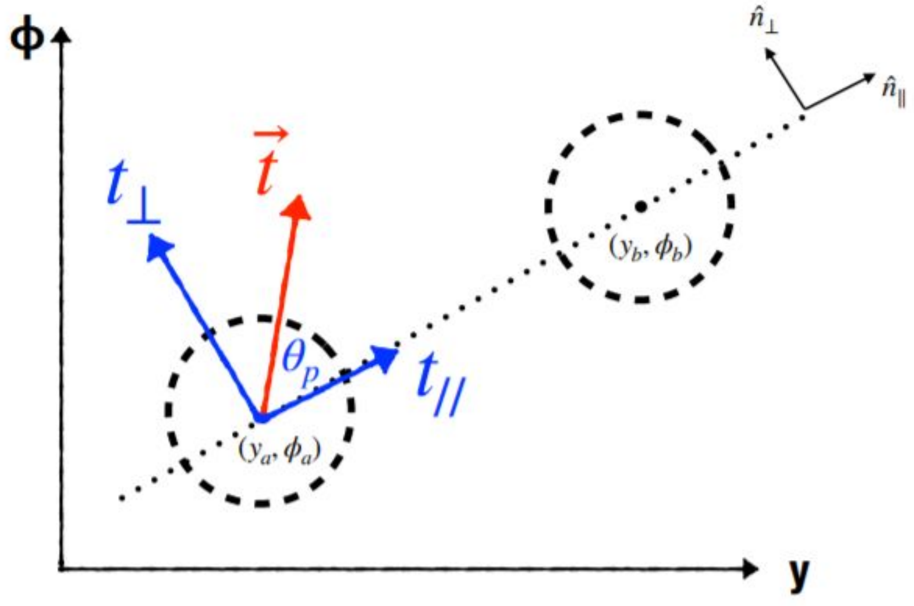


COLOR RING

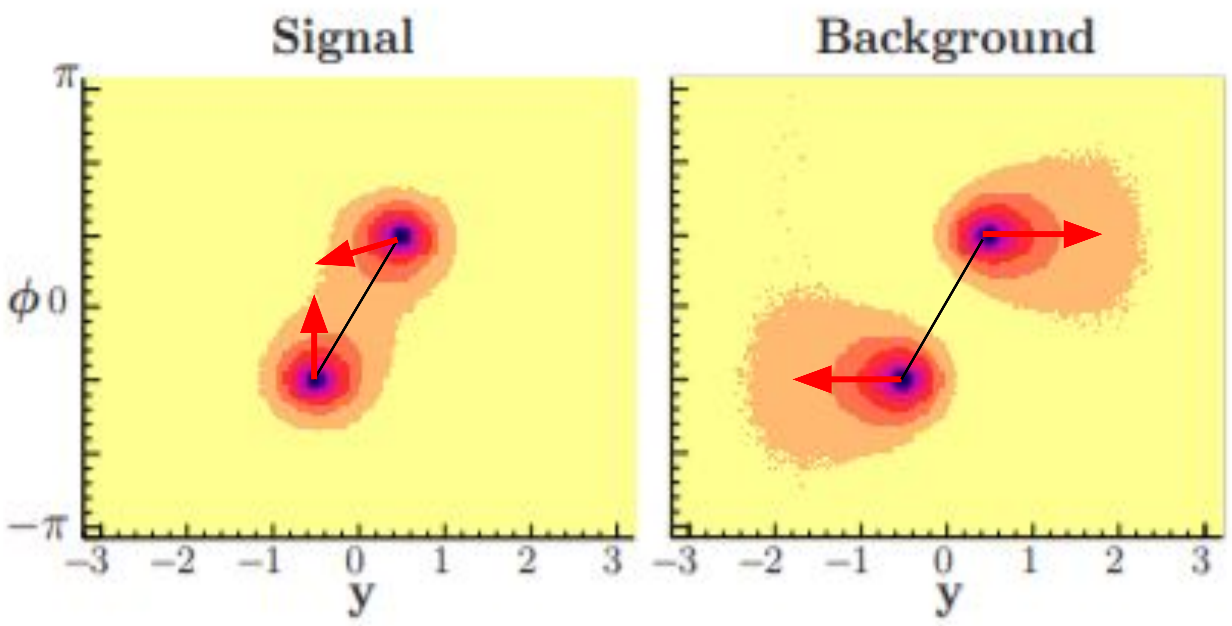


D_2

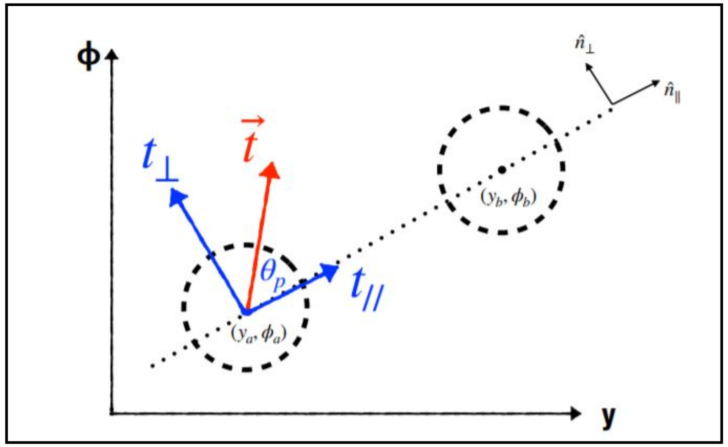
color sensitive variables:



a jet: b **hardest** jet
b jet: the other b-jet

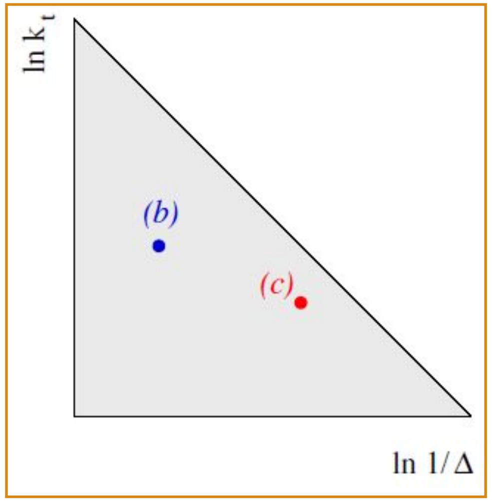


color sensitive variables

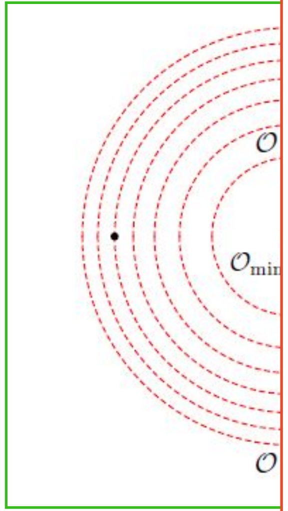


PULL VECTOR

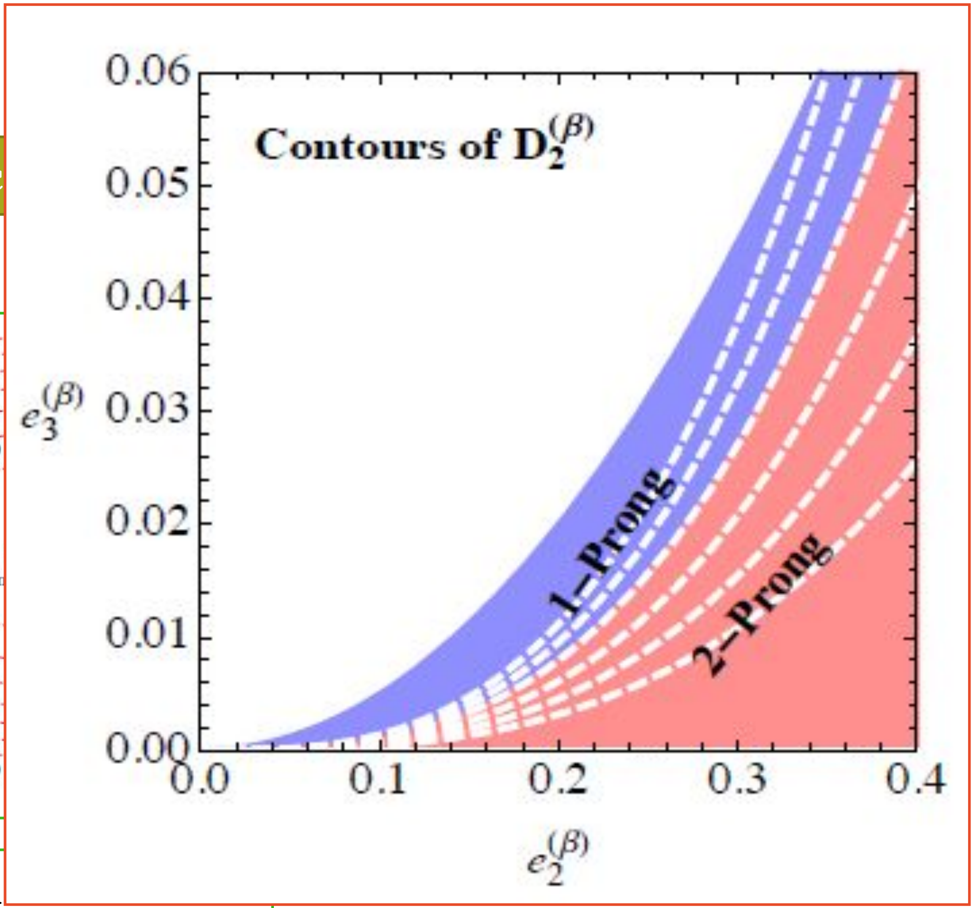
Variables sensitive to c



LUND PLANE

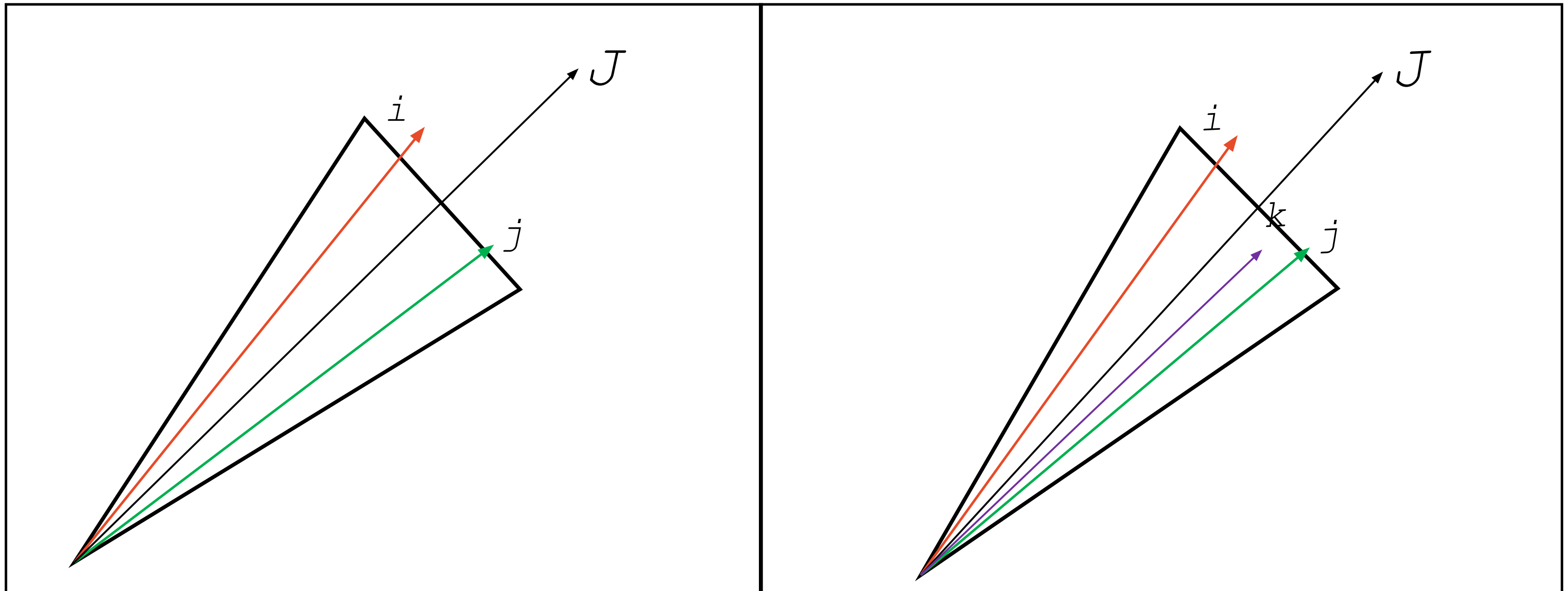


COLOR



D_2

color sensitive variables: d2



$e_2^\beta \propto p_i p_j R_{ij}^\beta$

$e_3^\beta \propto p_i p_j p_k R_{ij}^\beta R_{ik}^\beta R_{jk}^\beta$

$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}$$

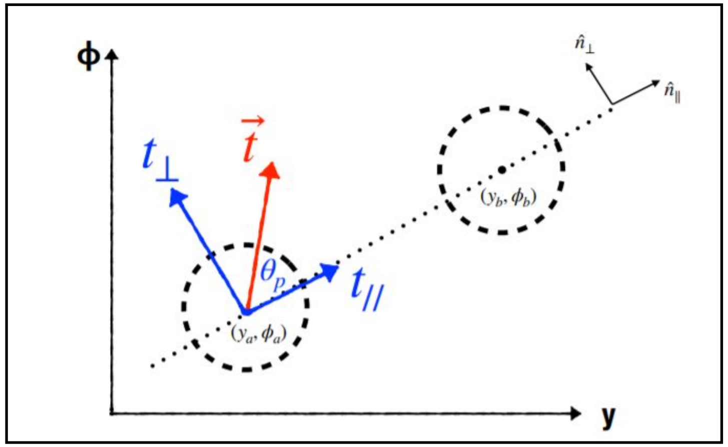
SIGNAL

Small value of D_2

BACKGROUND

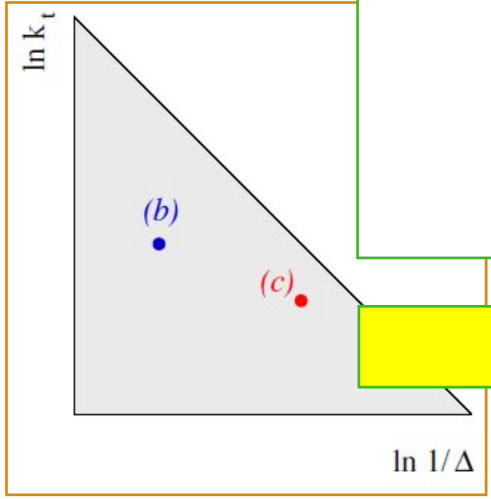
Large value of D_2

Color sensitive variables

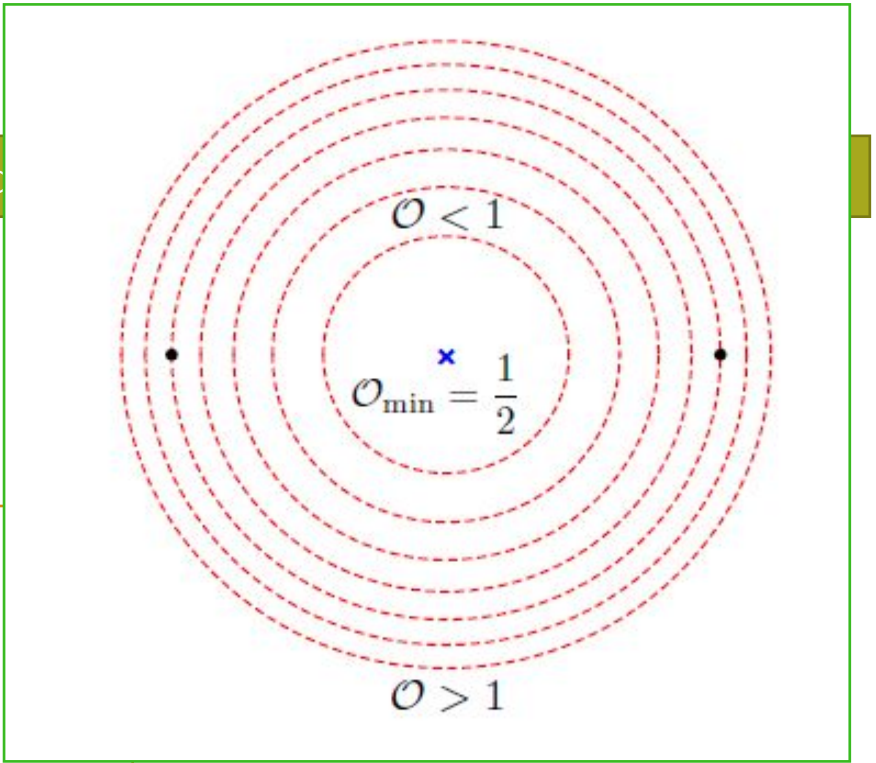


PULL VECTOR

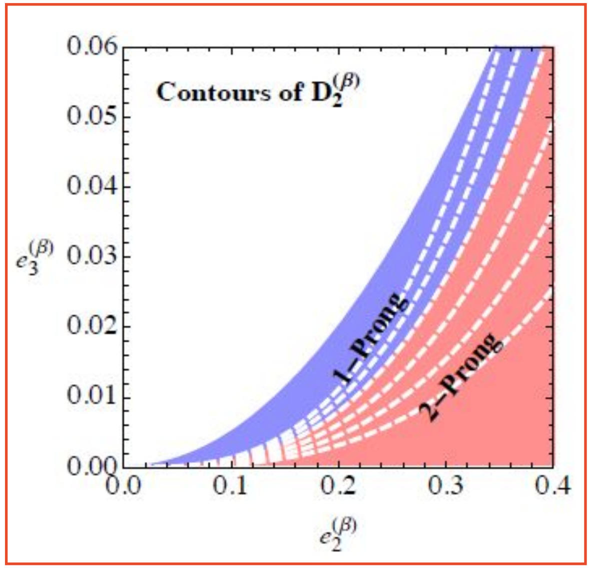
Variab



LUND PLANE



COLOR RING

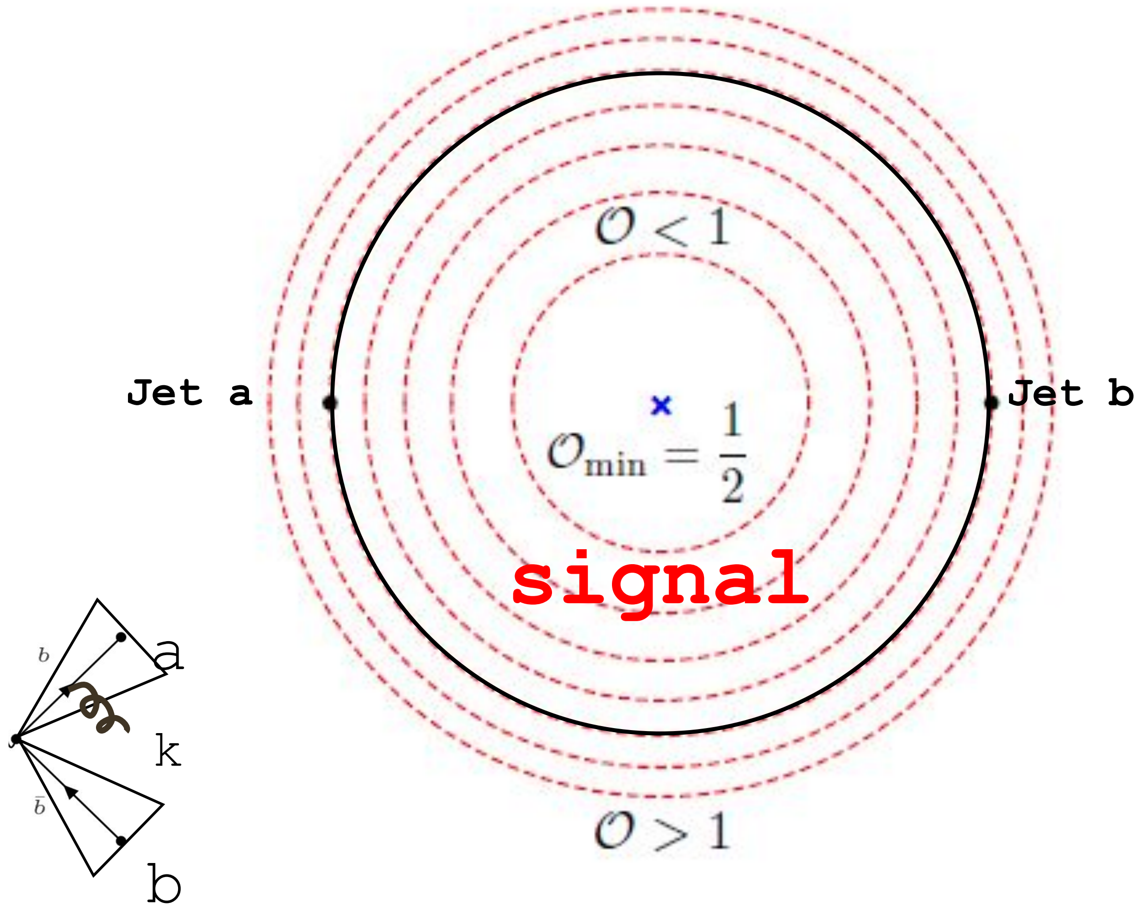


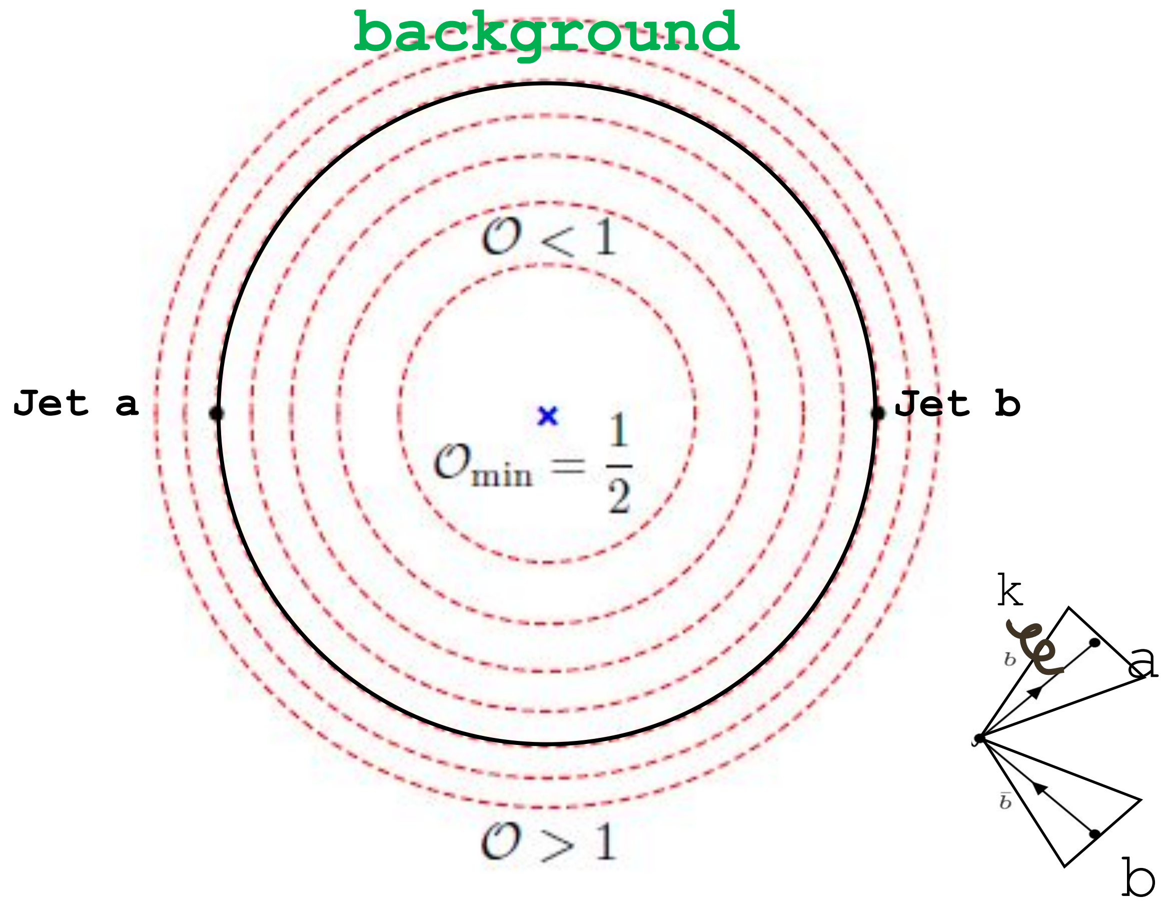
D_2

Jet Color Ring

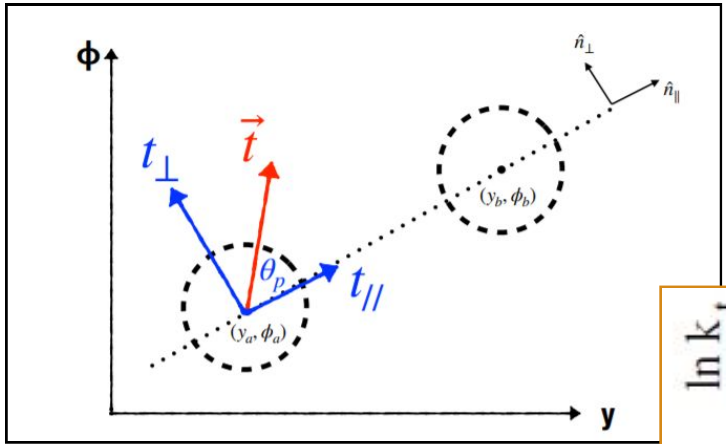
$$\mathcal{O} = \frac{|\mathcal{M}_B|^2}{|\mathcal{M}_S|^2} = \frac{\theta_{ak}^2 + \theta_{bk}^2}{\theta_{ab}^2}$$

Ratio of squared matrix elements in the **boosted limit**



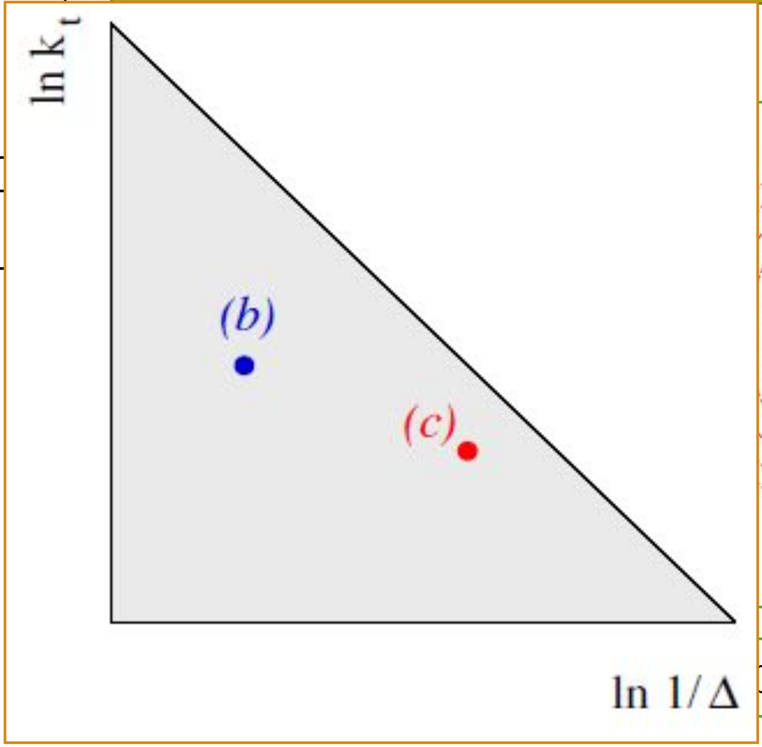


Color sensitive variables

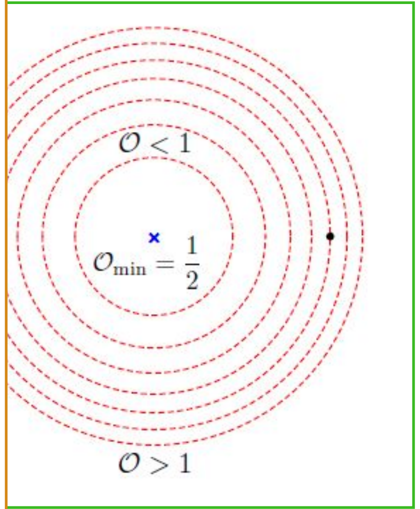


PULL VECTOR

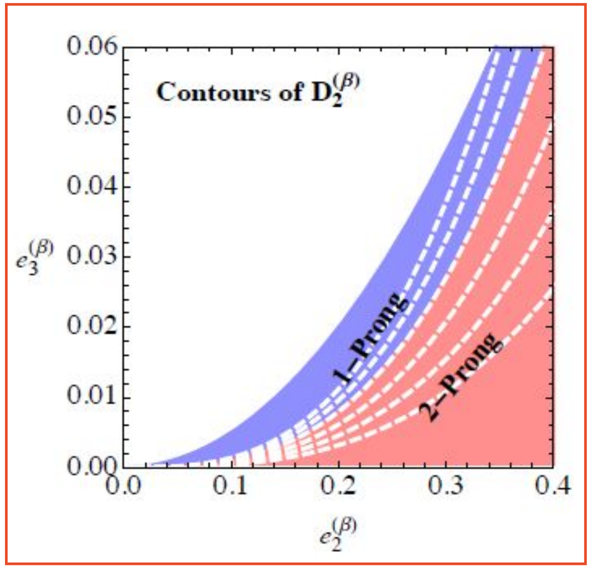
Variables sensitive to color flow



LUND PLANE

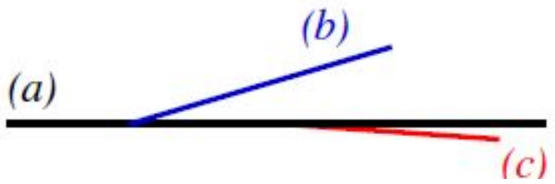


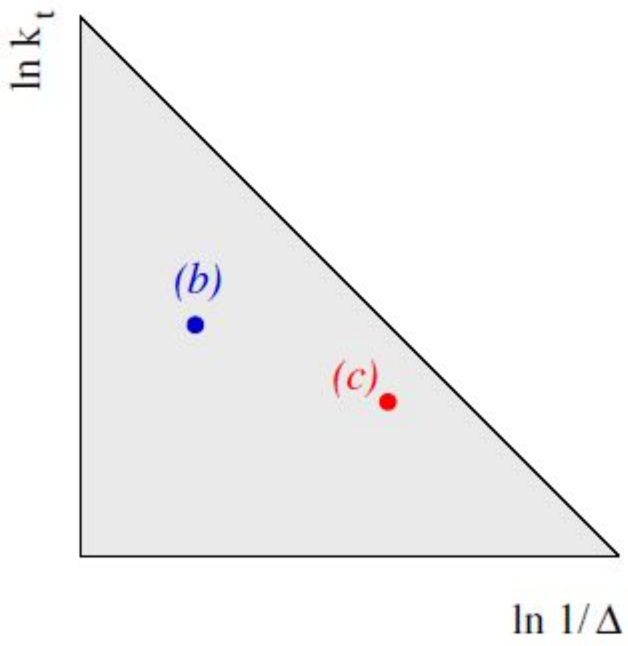
COLOR RING



D_2

Jet Lund Plane Image





(a) Primary particle of the jet
(b) Emission
(c) Emission

DECLUSTERING

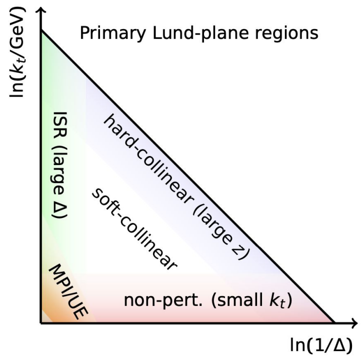
LUND PLANE

1. Jet are re-clustered with **Cambridge-Aachen** (C/A) algorithm, (particles closest in rapidity and and azimuth φ and recombines them into a "pseudojet")

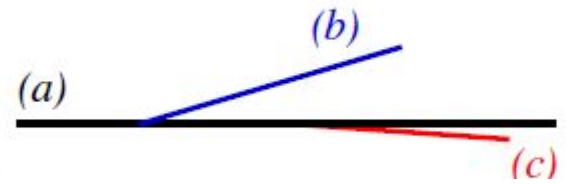
2. Decluster the jet to produce two **pseudojets**, pa and pb, labelled such that $p_{t,a} > p_{tb}$, where b is the emission and $p_{t,a} + p_{tb}$ is the jet

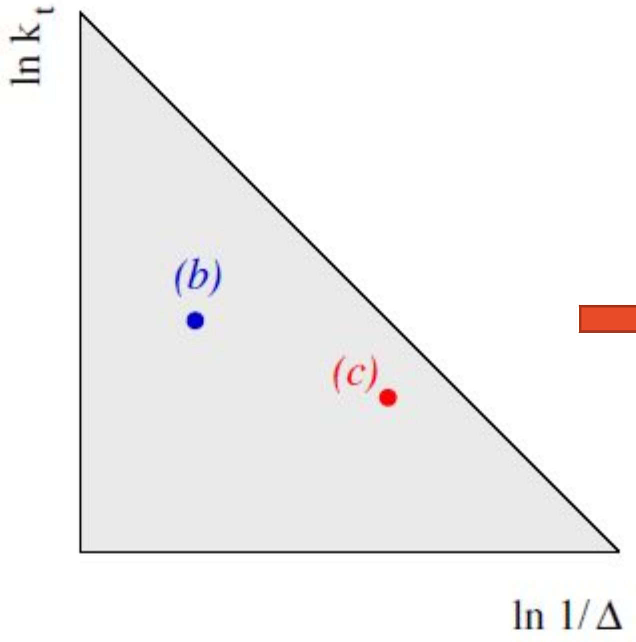
3. Construct Lund Plane variables

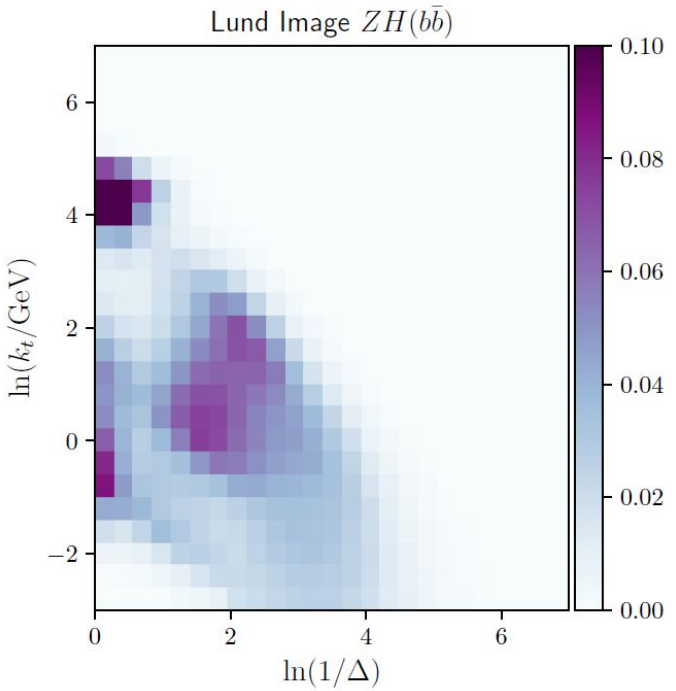
$$\Delta \equiv \Delta_{ab}, \quad k_t \equiv p_{tb}\Delta_{ab}$$



Jet Lund Plane Image







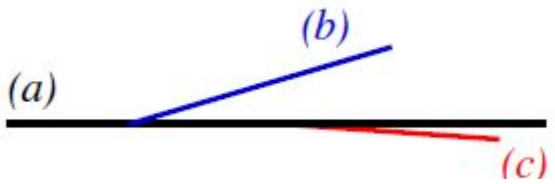
(a) Primary particle of the jet
(b) Emission
(c) Emission

DECLUSTERING

LUND PLANE

2D histogram
LUND PLANE IMAGE

Jet Lund Plane Image

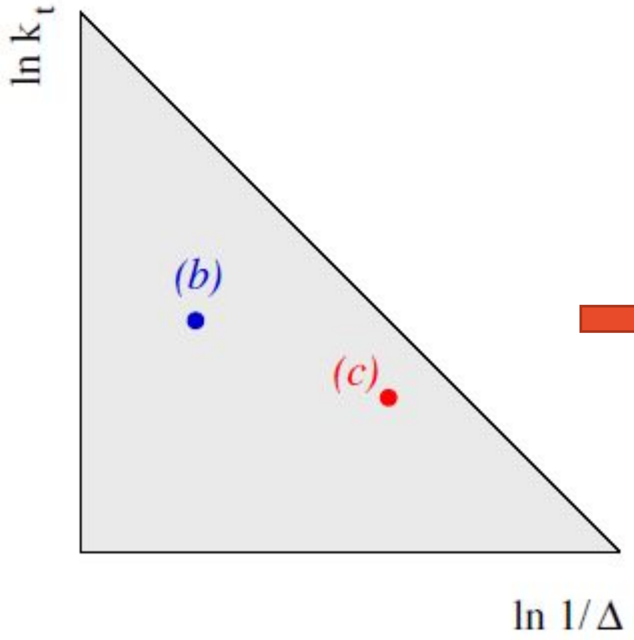


(a) Primary particle of the jet

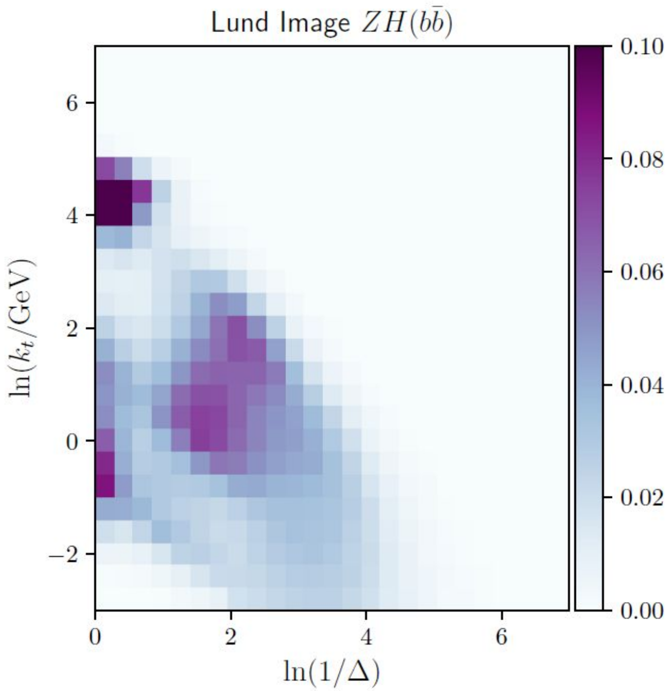
(b) Emission

(c) Emission

DECLUSTERING



LUND PLANE



2D histogram

LUND PLANE IMAGE

CONVOLUTIONAL NEURAL NETWORK

Phenomenological studies

Tagging the Higgs boson decay to bottom quarks with colour-sensitive observables and the Lund jet plane

Tagging the Higgs boson decay to bottom quarks with colour-sensitive observables and the Lund jet plane

Luca Cavallini, Andrea Coccaro, Charanjit K. Khosa, Giulia Manco, Simone Marzani, Fabrizio Parodi, Daniela Rebuzzi, Alberto Rescia, Giovanni Stagnitto

We study the problem of distinguishing b-jets stemming from the decay of a colour singlet, such as the Higgs boson, from those originating from the abundant QCD background. In particular, as a case study, we focus on associate production of a vector boson and a Higgs boson decaying into a pair of b-jets, which has been recently observed at the LHC. We consider the combination of several theory-driven observables proposed in the literature, together with Lund jet plane images, in order to design an original Hbb tagger. The observables are combined by means of standard machine learning algorithms, which are trained on events obtained with fast detector simulation techniques. We find that the combination of high-level single-variable observables with the Lund jet plane provides an excellent discrimination performance. We also study the dependence of the tagger on the invariant mass of the decaying particles, in order to assess the extension to a generic Xbb tagger.

Comments:	12 pages, 5 figures, 5 tables. v2 matches published version
Subjects:	**High Energy Physics - Phenomenology (hep-ph)**; High Energy Physics - Experiment (hep-ex)
Report number:	ZU-TH 50/21
Cite as:	arXiv:2112.09650 **[hep-ph]**
	(or arXiv:2112.09650v2 **[hep-ph]** for this version)
	https://doi.org/10.48550/arXiv.2112.09650
Journal reference:	Eur.Phys.J.C 82 (2022) 5, 493
Related DOI:	https://doi.org/10.1140/epjc/s10052-022-10447-1

Submission history

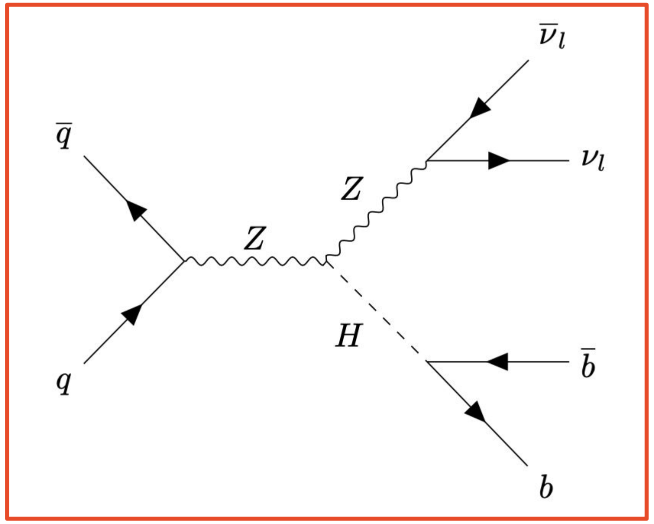
From: Giovanni Stagnitto [view email]

[v1] Fri, 17 Dec 2021 17:51:33 UTC (1,102 KB)

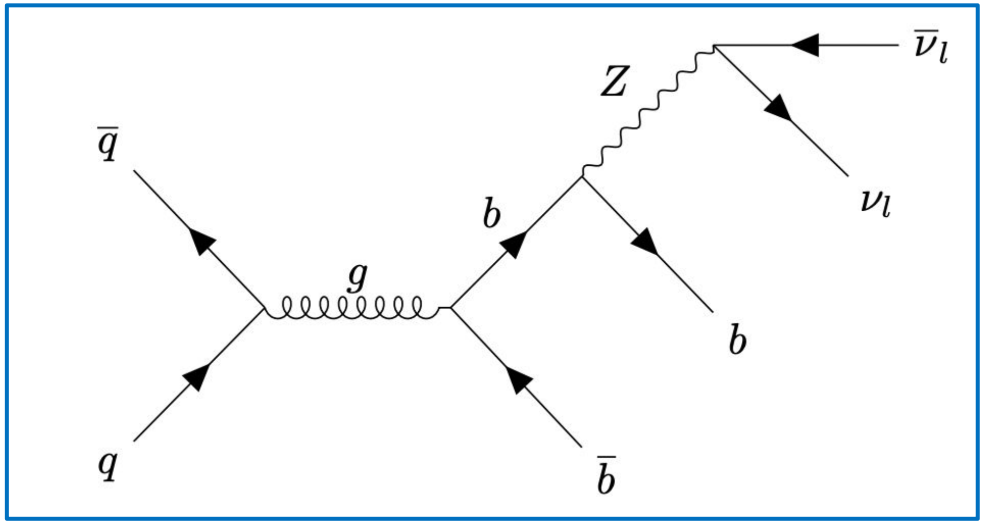
[v2] Wed, 1 Jun 2022 07:47:57 UTC (1,102 KB)

Simulation

Simulation of signal and background



SIGNAL



BACKGROUND

- Generate 300k signal, 4M background events in **MG5 AMC@NLO v2.8.3.2**
  - Require p_T>200 GeV for ν pairs
- Shower in **PYTHIA v8.305**
- Simulate detector effects with **DELPHES v3.5.0** using modified **ATLAS card**

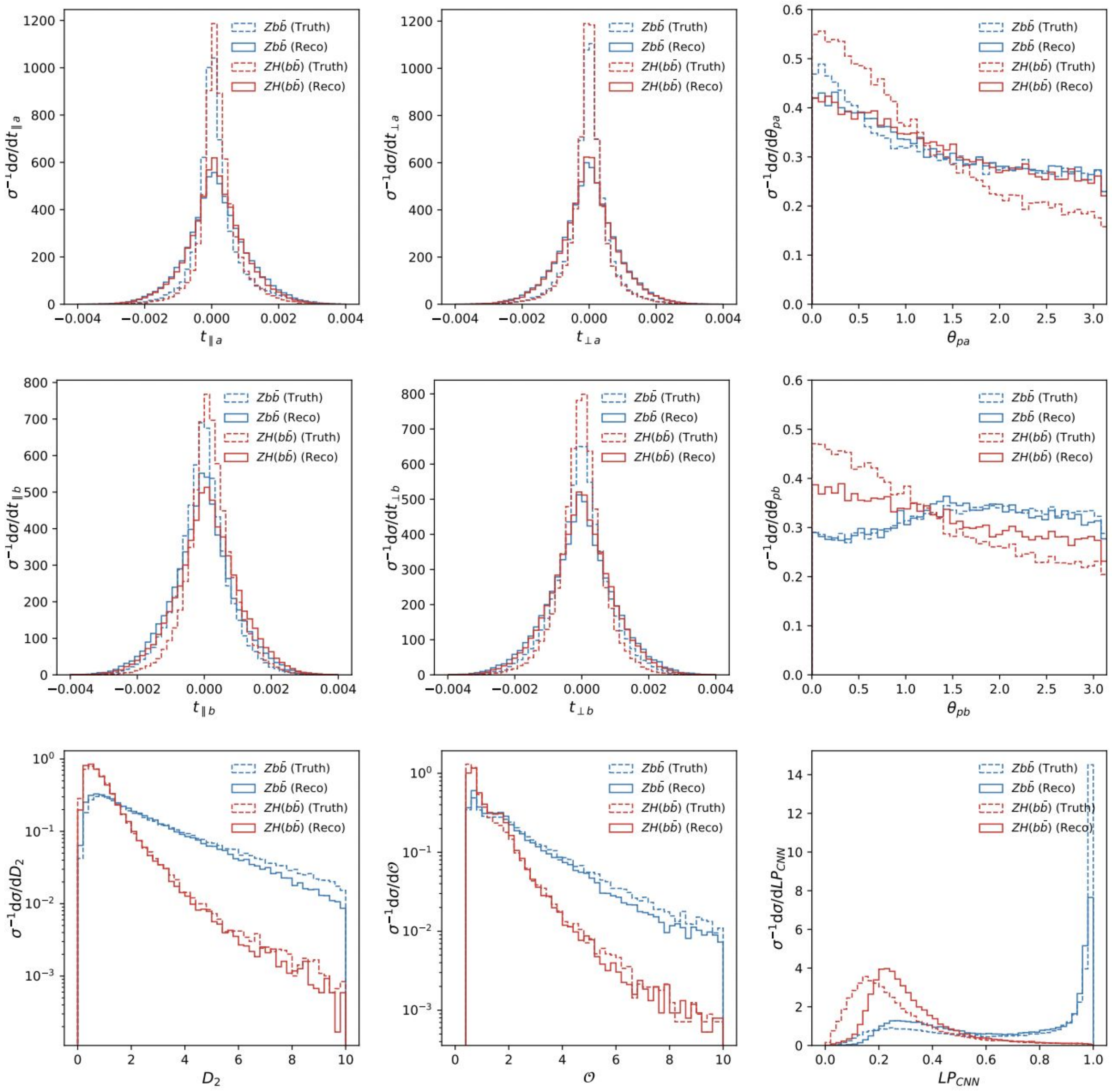
Analysis

- **Truth:** all generated particles removing neutrinos with $p_T > 0.5$ into jets
  - Fastjet jet reconstruction with anti-kt algorithm
  - Jet p_T min = 0 GeV
- **Reco:** ECal and HCal towers and muons:
  - Calo jet Fastjet reconstructed with anti-kt algorithm
  - Jet p_T min = 5 GeV
- **Large jet : R = 1** and select the hardest with:
  - $p_T > 250$ GeV
  - $|y| < 1.5$
- Exactly **2 sub-jets** of **R = 0.2:**
  - $p_T > 10$GeV
  - $\Delta R = 0.8$ from the large jet
- Flavor association for b tagging:
  - b-parton with $p_T > 5$ GeV
  - $\Delta R = 0.2$
  - b-parton $\eta_{MAX} = 2.5$

Events Passed

	Truth	Reco
Signal	20%	17%
Background	1.6%	1.3%

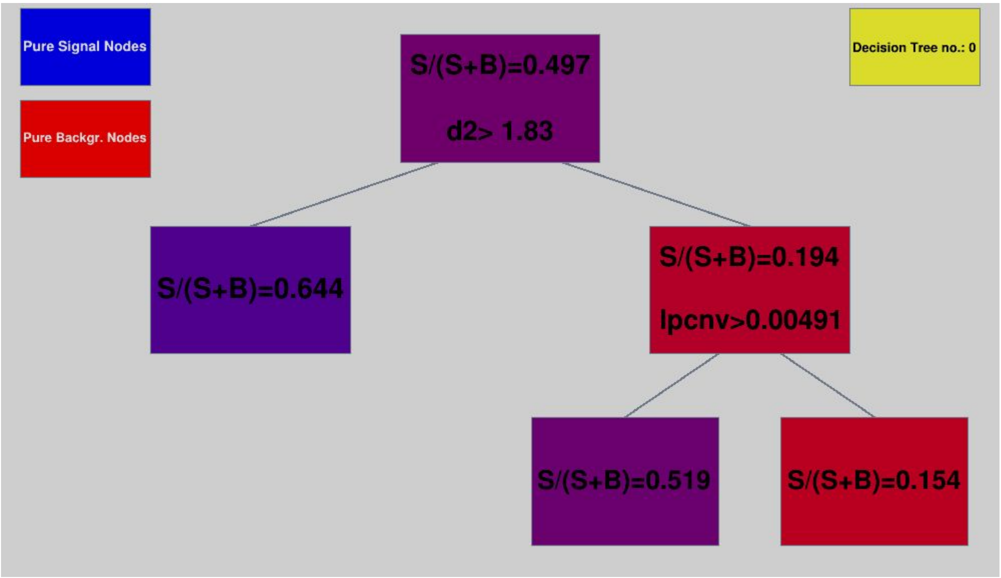
Distributions



Good discrimination in particular for D2, Color Ring and Lund Plane

Machine Learning Algorithm

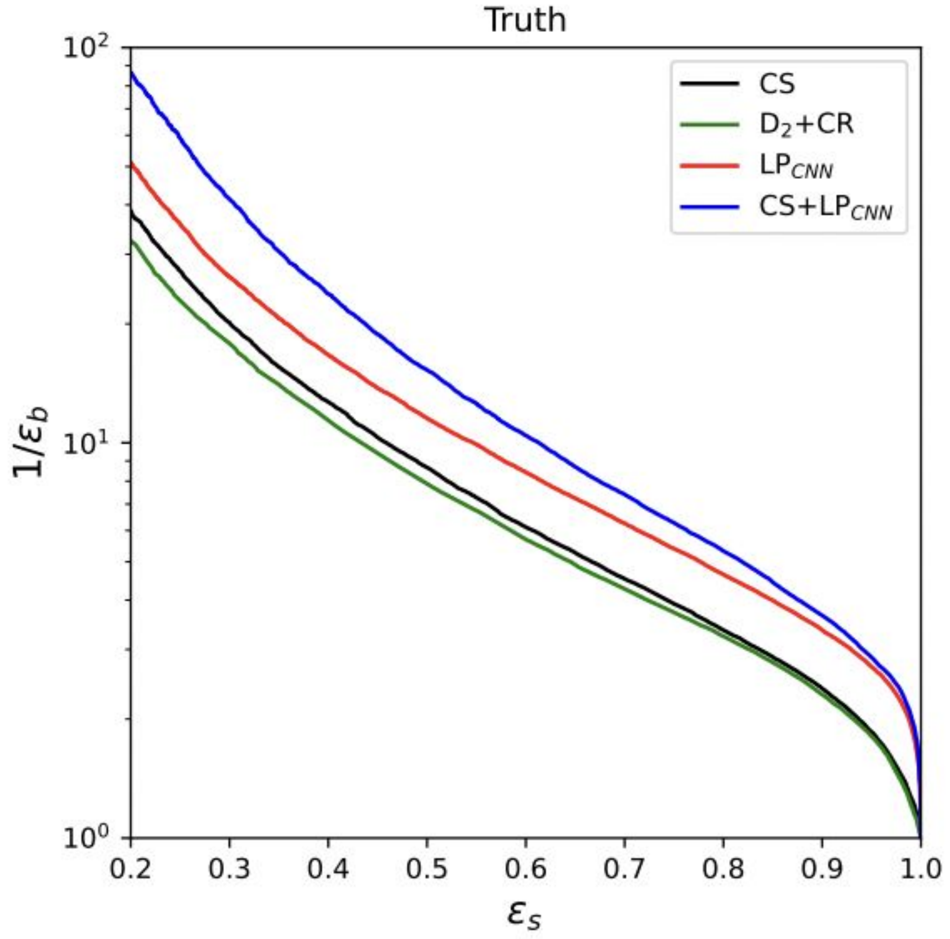
BOOSTED DECISION TREE (BDT)



Parameters	Value
No. of Trees	100
Max Depth	3
MinNodeSize	2.5%
Boost Type	AdaBoost
Train/Test	50/50
No. of Cuts	200
Downsampling	No

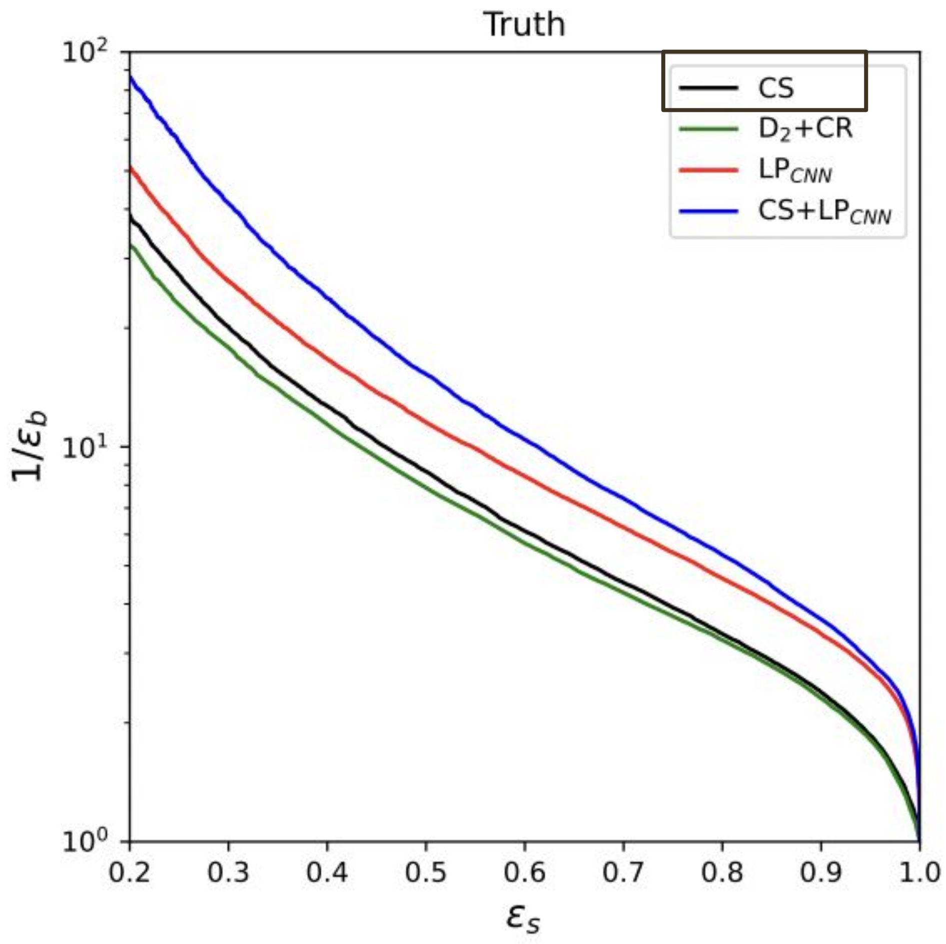
Results

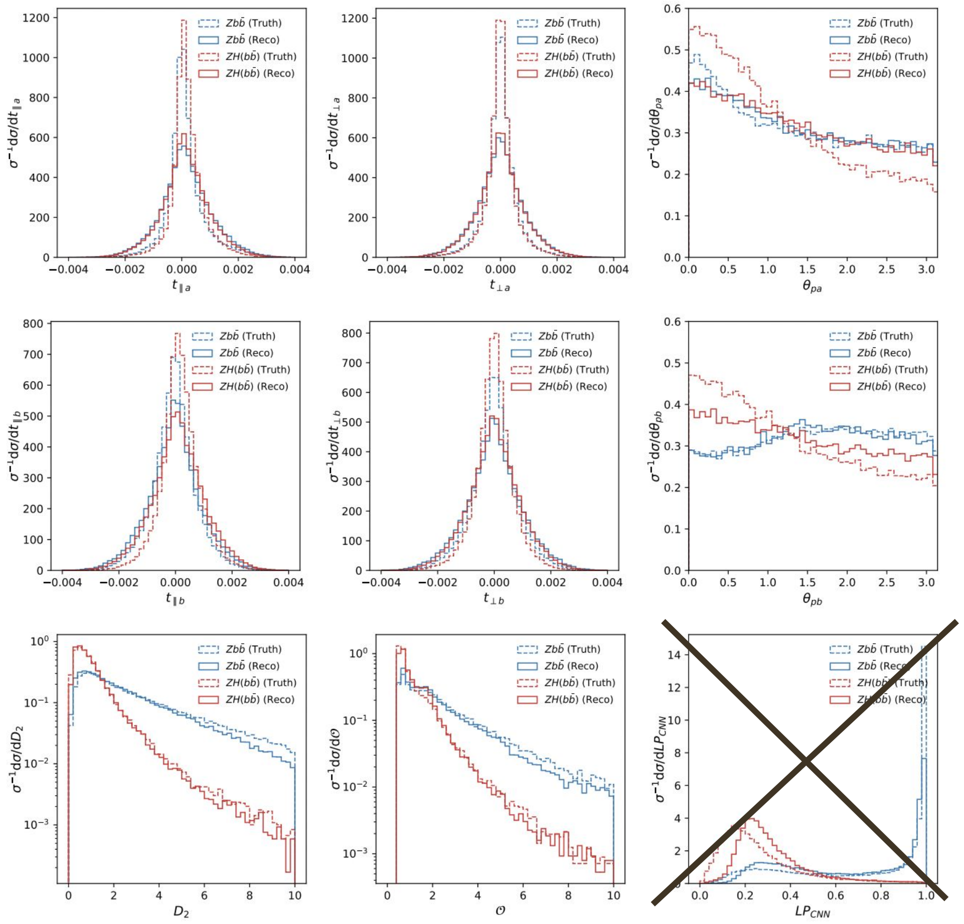
ROC (BDT)



Results

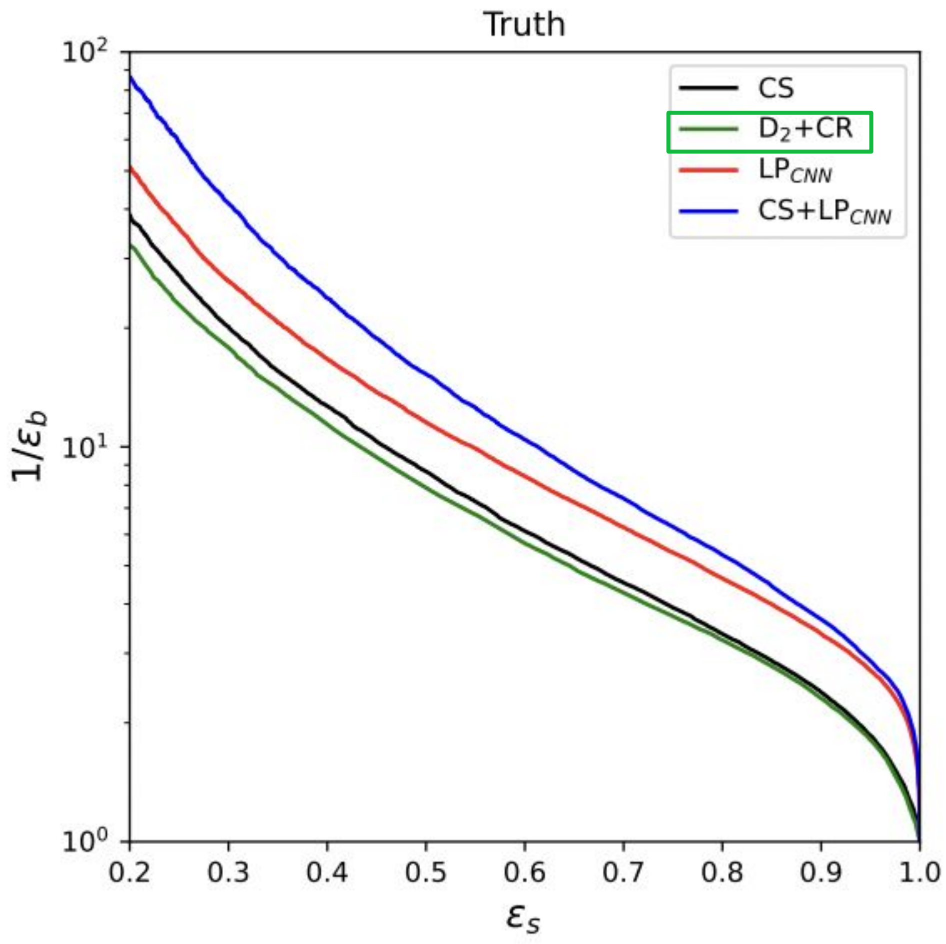
ROC


Truth
CS
D2+CR
LPCNN
CS+LPCNN
1/εb
εS

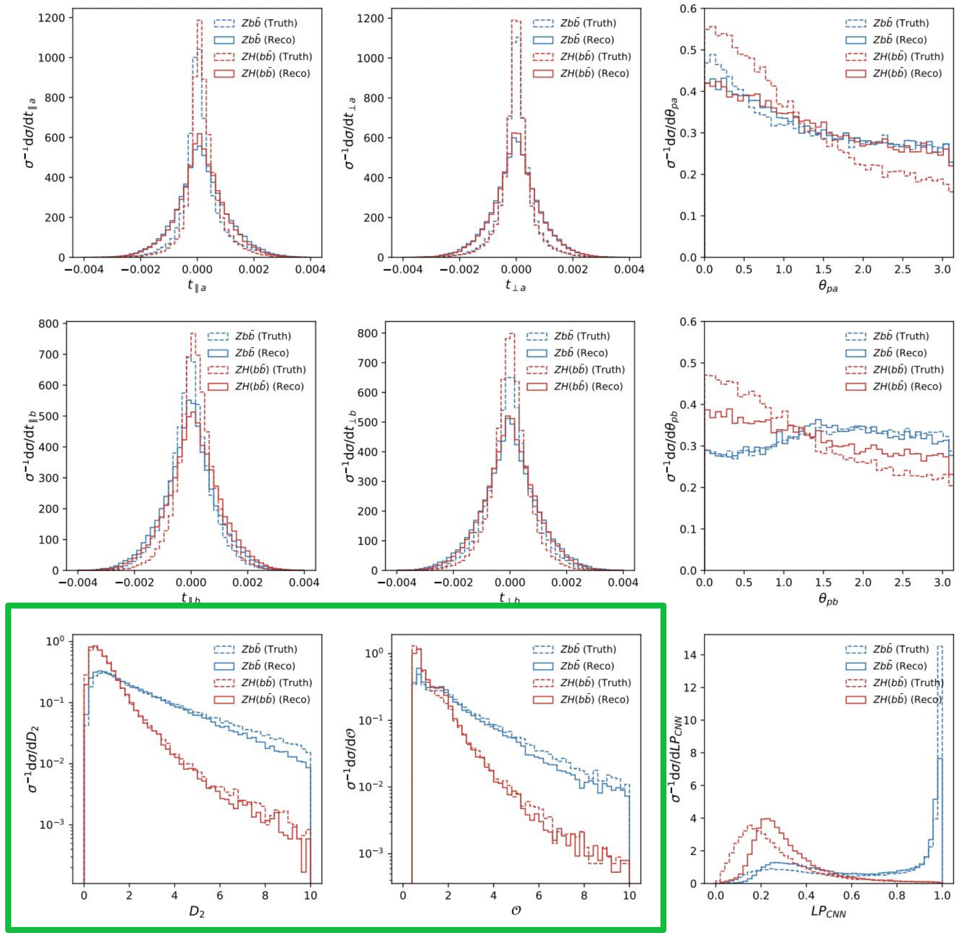

Zbb̄ (Truth)
Zbb̄ (Reco)
ZH(bb̄) (Truth)
ZH(bb̄) (Reco)
D2
LPCNN

Results

ROC



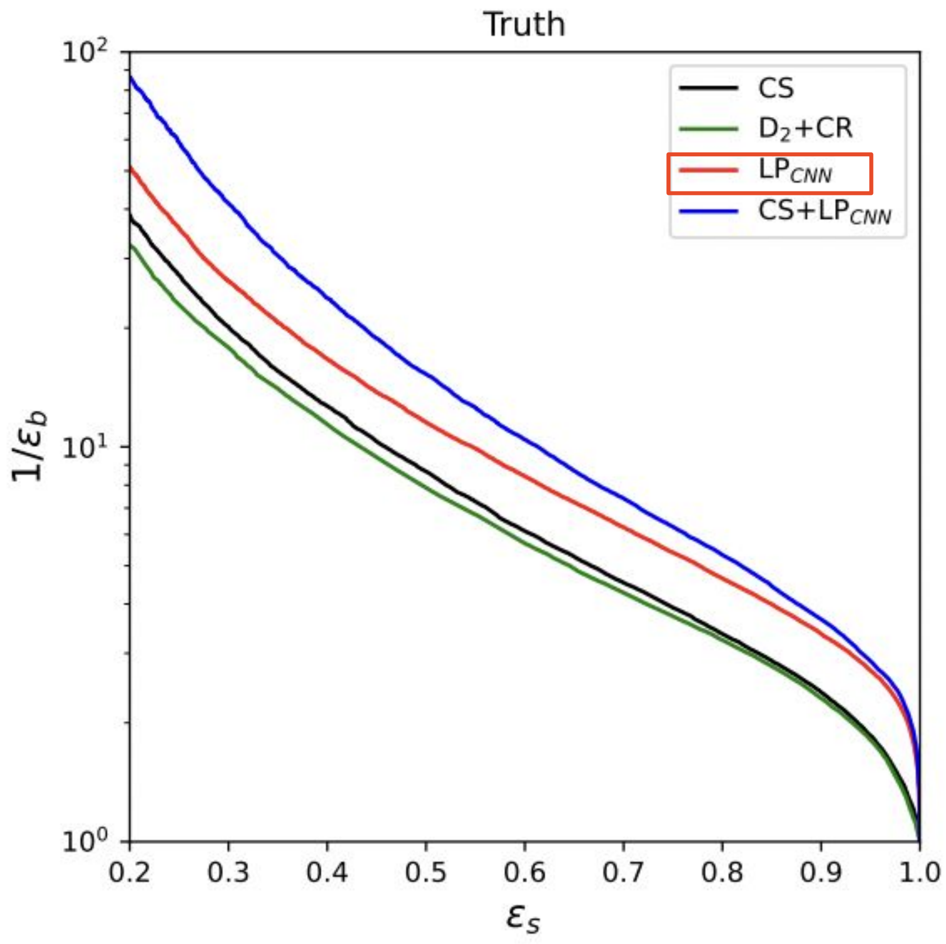
Truth
CS
D_2+CR
LP_{CNN}
CS+LP_{CNN}
$1/\varepsilon_b$
ε_S



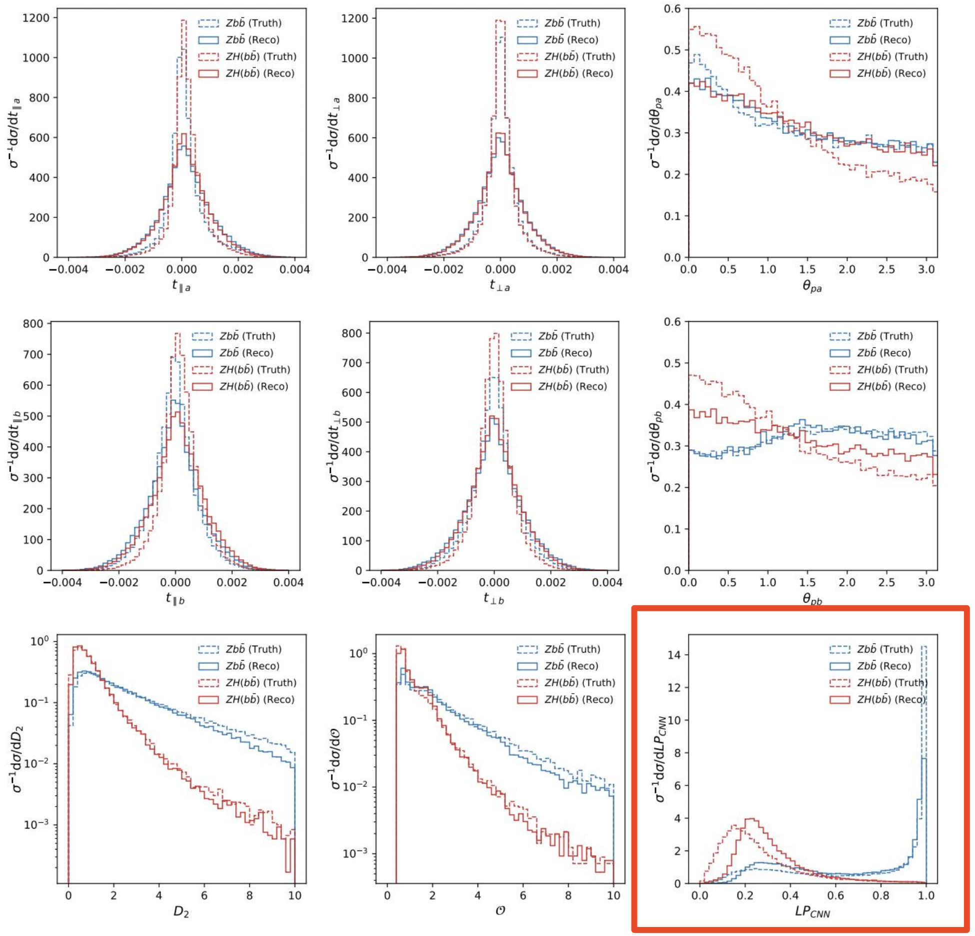
$Zb\bar{b}$ (Truth)
$Zb\bar{b}$ (Reco)
$ZH(b\bar{b})$ (Truth)
$ZH(b\bar{b})$ (Reco)
$t_{\parallel a}$
$t_{\perp a}$
θ_{pa}
θ_{pb}
D_2
LP_{CNN}

Results

ROC



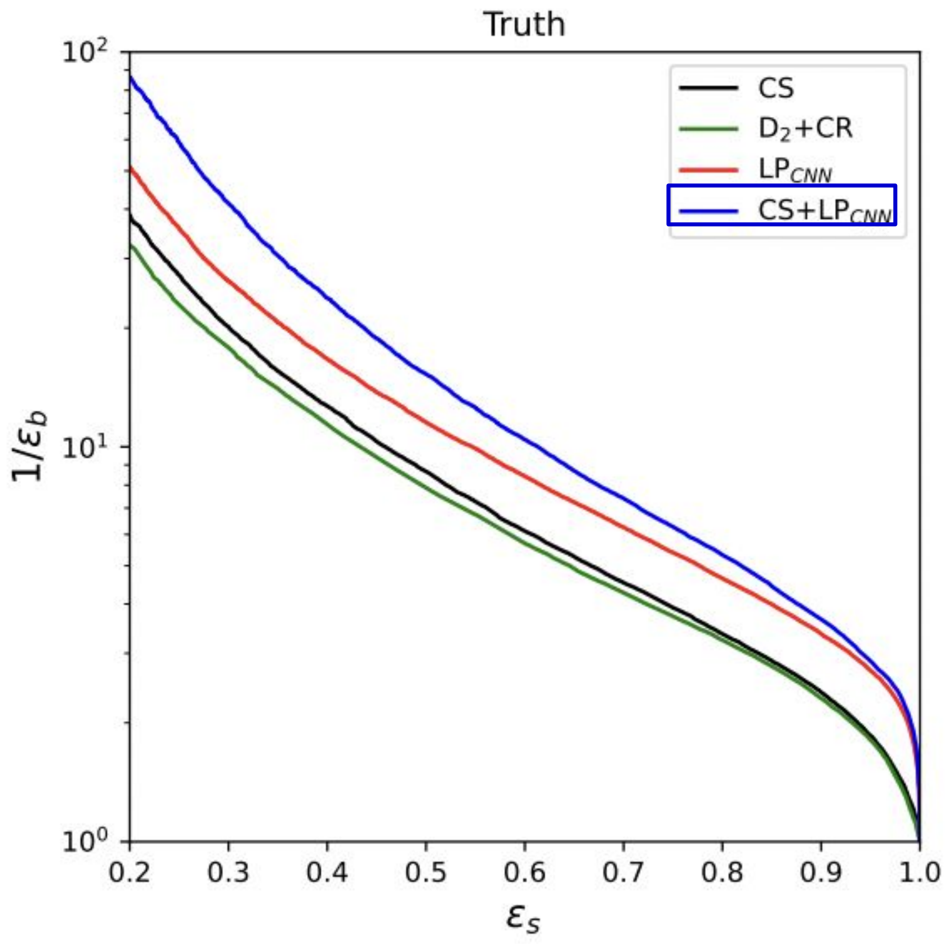
Truth
CS
D2+CR
LPCNN
CS+LPCNN
1/εb
εs



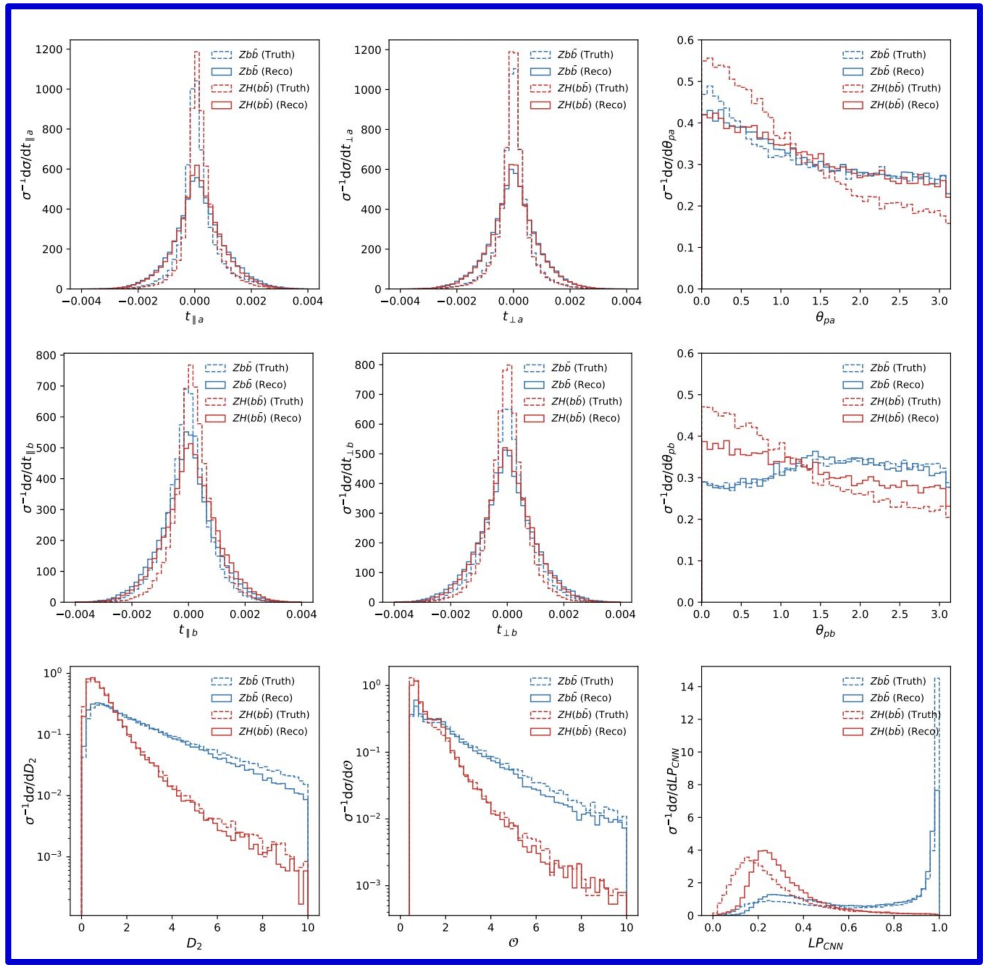
Zbb̄ (Truth)
Zbb̄ (Reco)
ZH(bb̄) (Truth)
ZH(bb̄) (Reco)
D2
LPCNN

Results

ROC

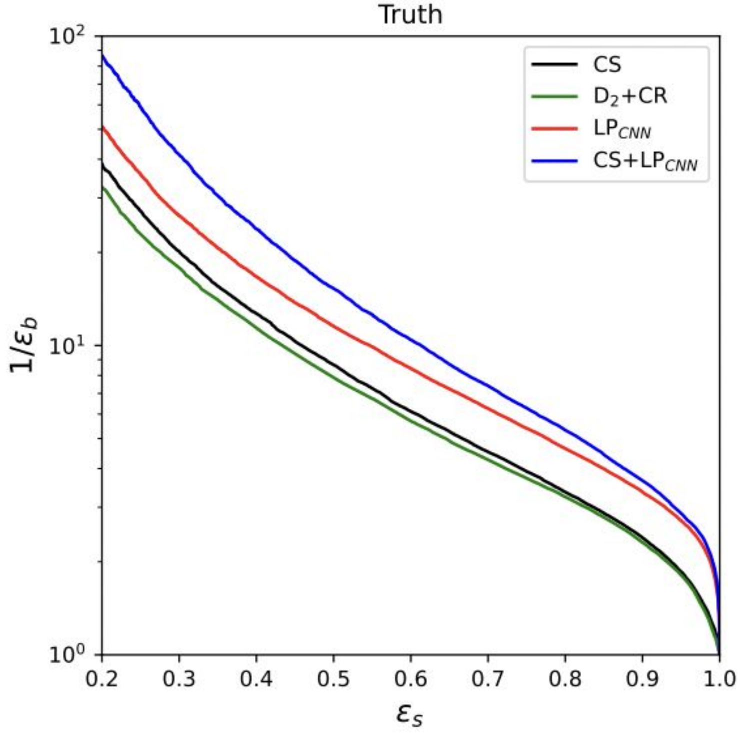


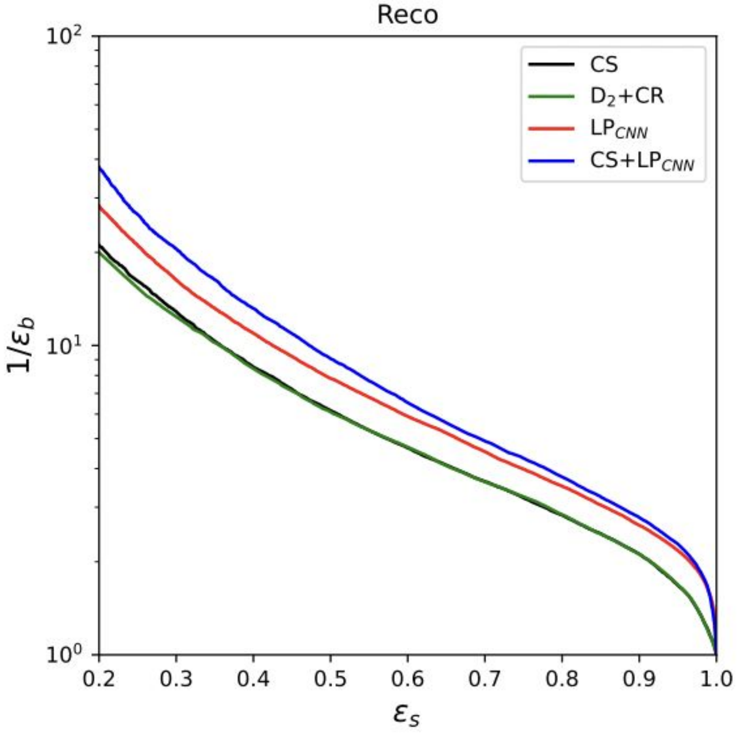
Truth
CS
D_2+CR
LP_{CNN}
CS+LP_{CNN}
$1/\varepsilon_b$
ε_S



$Zb\bar{b}$ (Truth)
$Zb\bar{b}$ (Reco)
$ZH(b\bar{b})$ (Truth)
$ZH(b\bar{b})$ (Reco)
$t_{\parallel a}$
$t_{\perp a}$
θ_{pa}
$t_{\parallel b}$
$t_{\perp b}$
θ_{pb}
D_2
LP_{CNN}

Results



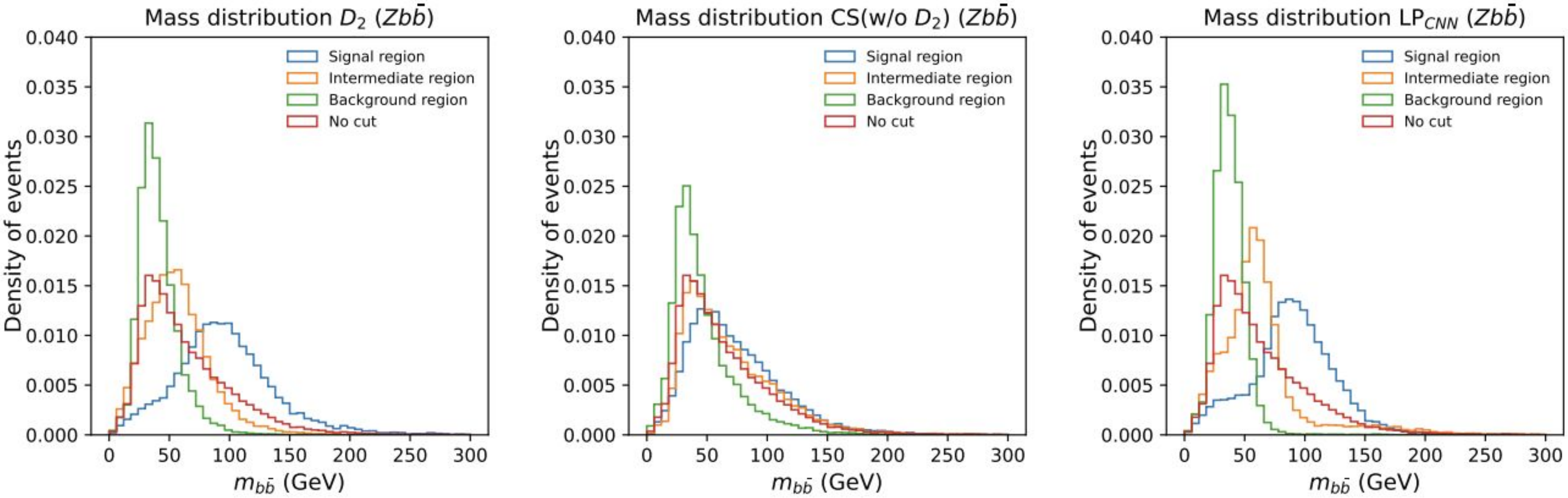


	AUC - Test Sample	
	Truth	Reco
CS observables	0.826	0.788
D_2+CR	0.817	0.787
LP$_{\text{CNN}}$	0.876	0.828
CS + LP$_{\text{CNN}}$	0.893	0.846

Observable Ranking				
	Truth		Reco	
Rank	Obs.	Importance	Obs.	Importance
1	LP$_{\text{CNN}}$	6.6×10^{-1}	LP$_{\text{CNN}}$	4.8×10^{-1}
2	D_2	1.4×10^{-1}	$\mathcal{O}$	1.0×10^{-1}
3	$\mathcal{O}$	5.7×10^{-2}	D_2	9.3×10^{-2}
4	θ_{pb}	3.0×10^{-2}	θ_{pb}	7.0×10^{-2}
5	θ_{pa}	2.9×10^{-2}	θ_{pa}	6.5×10^{-2}
6	$t_{\parallel b}$	2.6×10^{-2}	$t_{\perp b}$	6.0×10^{-2}
7	$t_{\parallel a}$	2.4×10^{-2}	$t_{\parallel a}$	4.5×10^{-2}
8	$t_{\perp b}$	1.9×10^{-2}	$t_{\perp a}$	4.3×10^{-2}
9	$t_{\perp a}$	1.0×10^{-3}	$t_{\parallel a}$	3.3×10^{-2}

Invariant Mass Independence

- In order to apply the tagger in other contexts, it should be independent on the invariant mass of the b-subjets
- In figure mass distributions for different regions, which correspond to different **cuts on discriminant BDT variable**
- **D2** and **Lund Plane** are mass correlated
- Removing D2 or Lund Plane causes worse performance in discrimination

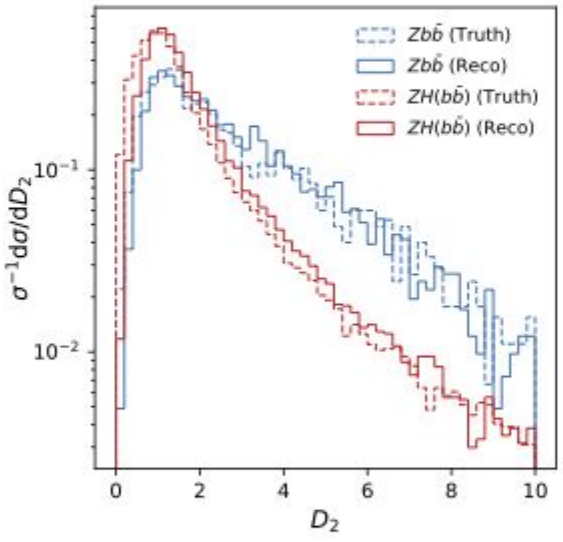


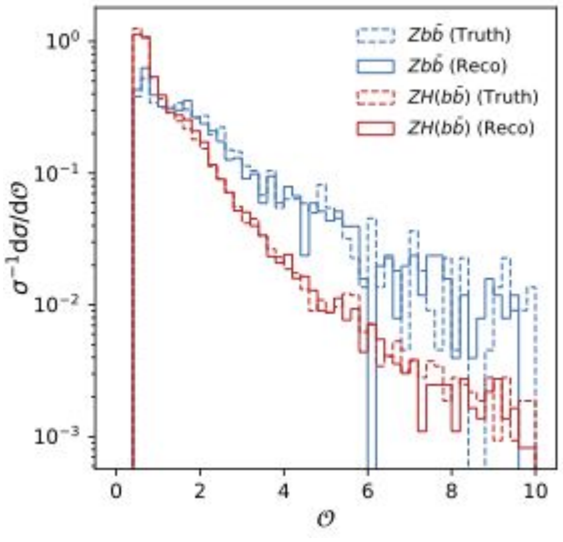
Studies at FCC-hh regime (A. Rescia, G. Manco)

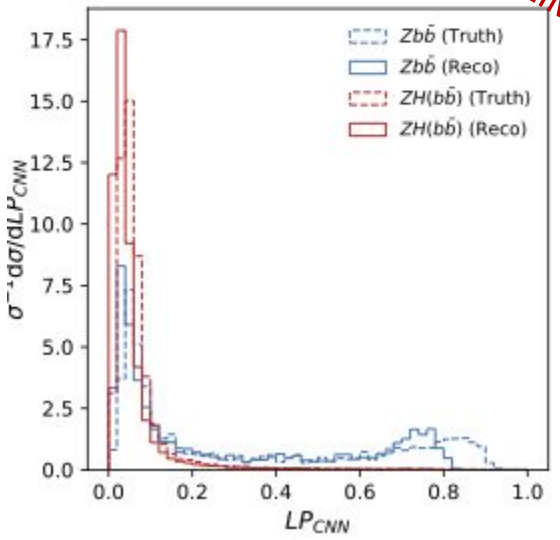
- Exercise the same analysis at FCC-hh regime :
  - Generate 200k events of background and 100k of signal from Madgraph v2.9.11, with energy of CM 100 TeV
  - Shower in PYTHIA v.8.235 and simulate effects with Delphes v3.5.0 using FCC-hh card modified

- Same analysis cuts, different rapidity cut of large-R jet ($|\eta| < 5$)
- Here only preliminary studies at low statistics, the idea is to see the behavior of color sensitive variables at higher energy and FCC-hh fast detector simulation

PRELIMINARY



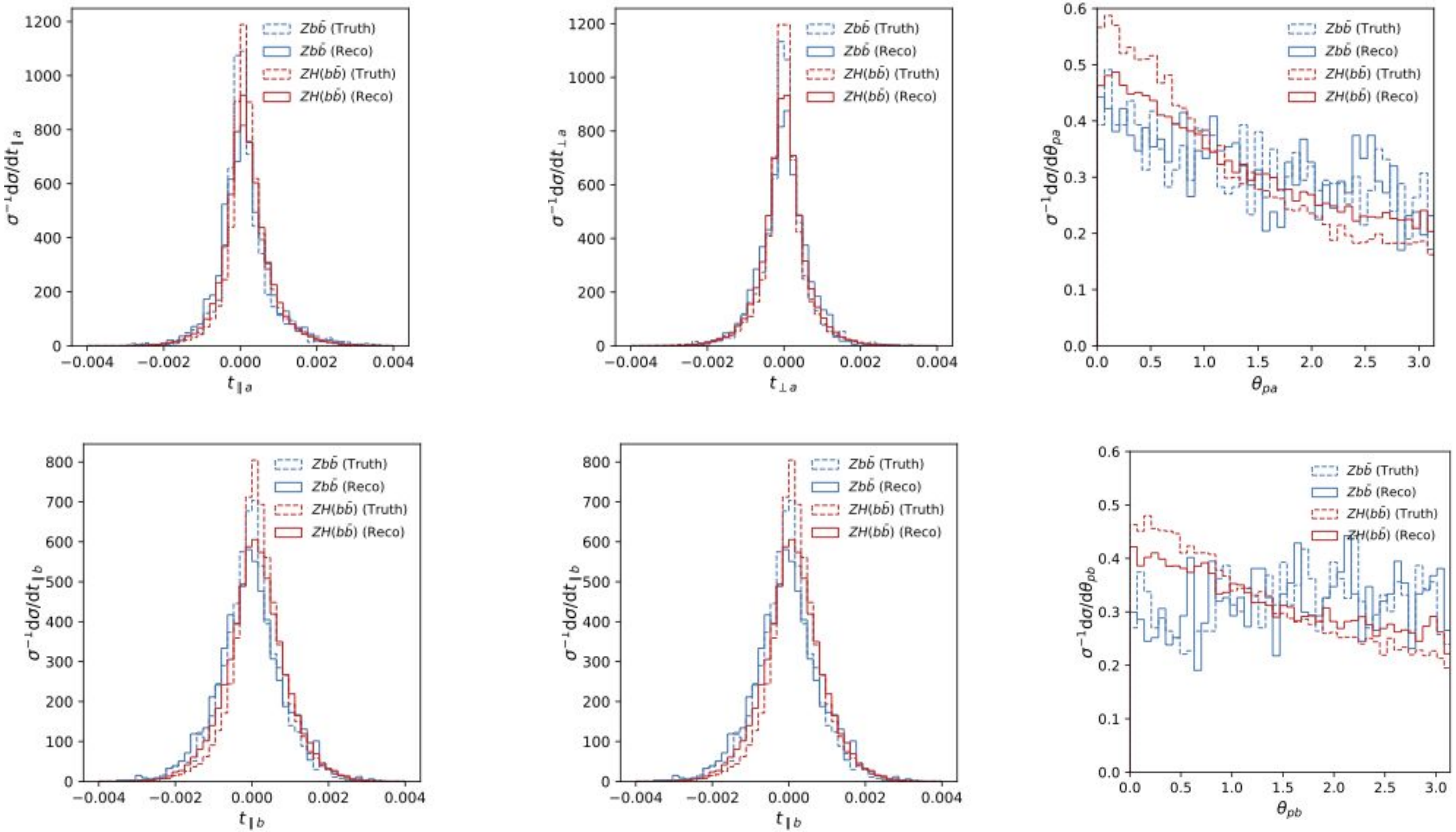




fluctuations due to low background statistics

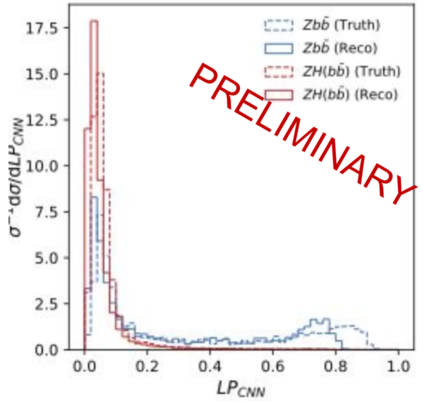
Studies at FCC-hh regime (A. Rescia, G. Manco)

PRELIMINARY

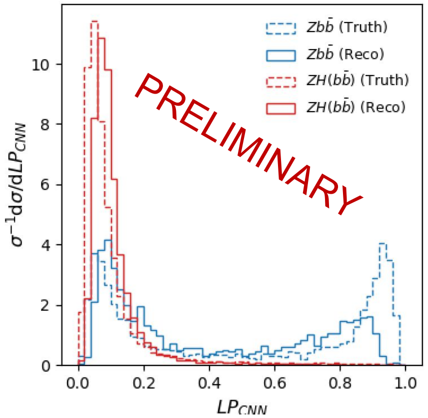


Studies at FCC-hh regime (A. Rescia, G. Manco)

- Observations:
  - Distributions show similar behavior as 13 TeV analysis
  - We expect similar performances
  - Lund Plane variable shows a different behavior with flatter background (probably due to CNN sensibility to statistical fluctuations)
- Next:
  - See the behavior at higher statistics
- Perspectives:
  - improved xbb tagger techniques at LHC can be useful in the future colliders
  - discrimination between singlet and octect of color is helpful in Higgs and BSM studies at FCC-hh (see next slides)

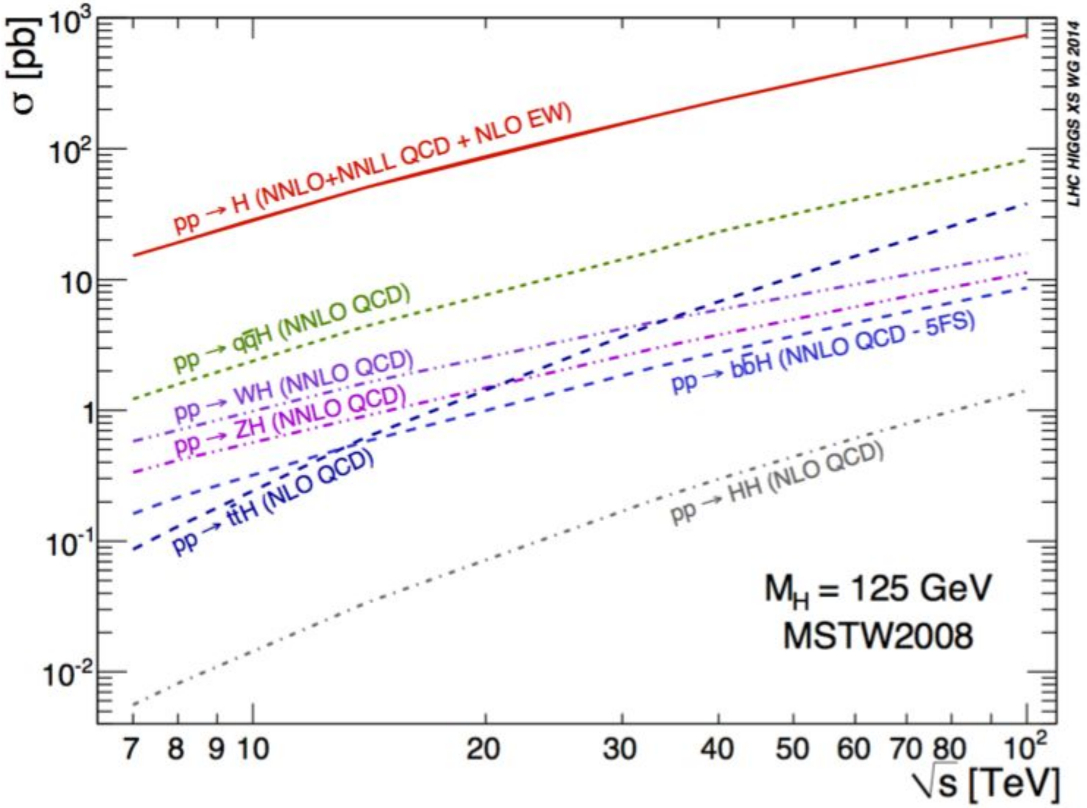


100 TeV



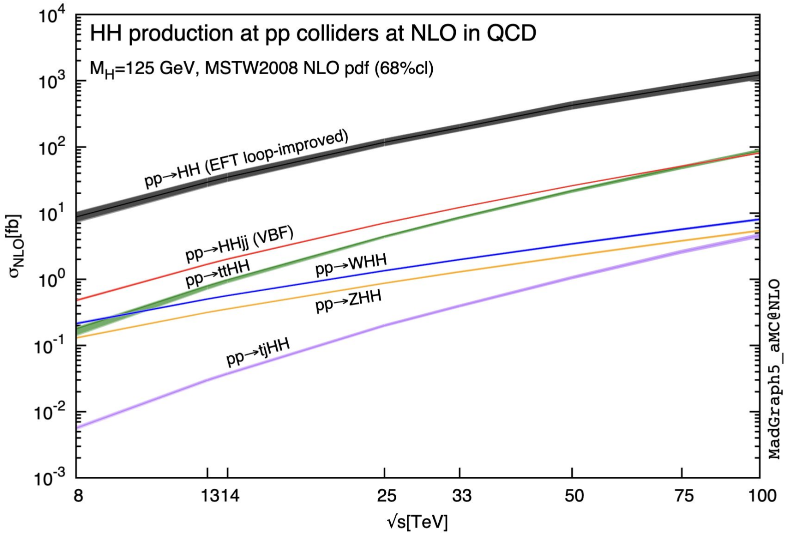
13 TeV (same statistic as 100 TeV)

Xbb tagger: FCC-hh Higgs



https://arxiv.org/pdf/1910.11775.pdf

- Next colliders need to investigate EW sector, which is still puzzling: for example the Δm_H quantum correction of Higgs mass (naturalness problem)
- LHC produced 8 million Higgs bosons, HL-LHC will improve by a factor 20, improving uncertainties of Higgs couplings by factor 5-10
- Higgs self-coupling and nature of EWSB will remain unknown even after HL-LHC and FCC-ee (indirect only)
- At higher energies many Higgs processes are most probable

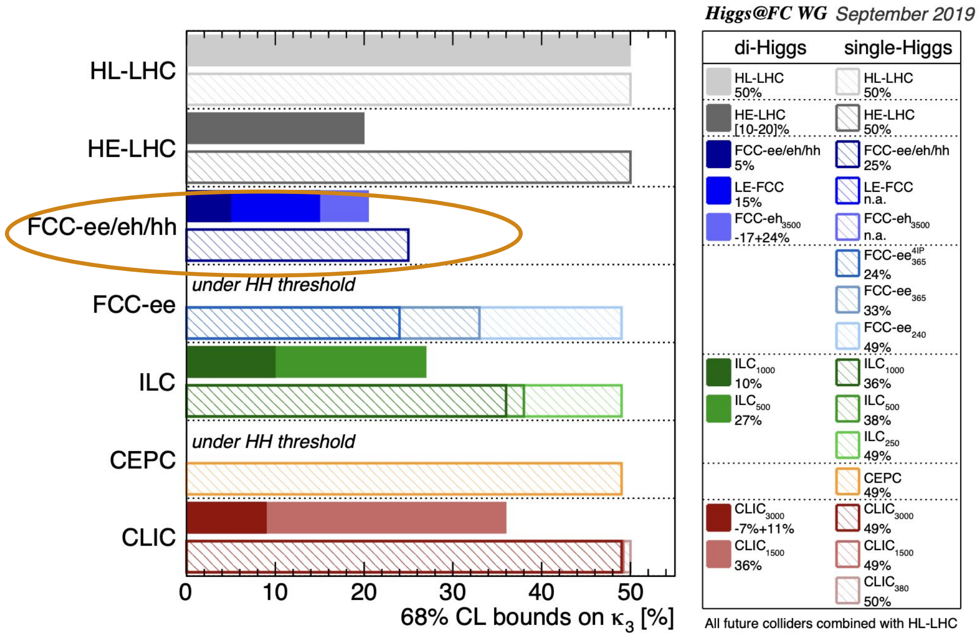


Xbb tagger: FCC-hh SM processes

- Di-Higgs production allow comparison of SM λ parameters of the Higgs potential

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda_{3H} v H^3 + \frac{1}{4}\lambda_{4H} H^4$$

- FCC-hh will improve the sensitivity on κ_3 (= λ_3/λ_3^{SM}) from 50% of HL-LHC to 5% of FCC-hh



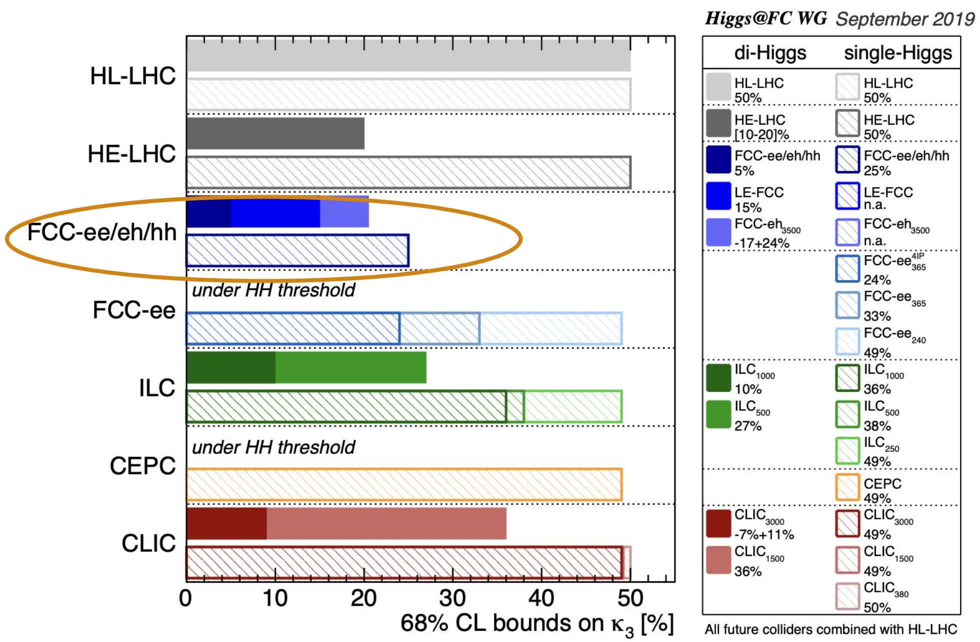
https://arxiv.org/pdf/1910.11775.pdf

Xbb tagger: FCC-hh SM processes

- Di-Higgs production allow comparison of SM λ parameters of the Higgs potential

$$V(H) = \frac{1}{2} m_H^2 H^2 + \lambda_{3H} v H^3 + \frac{1}{4}\lambda_{4H} H^4$$

- FCC-hh will improve the sensitivity on κ_3 (= λ_3/λ_3^{SM}) from 50% of HL-LHC to 5% of FCC-hh



https://arxiv.org/pdf/1910.11775.pdf

Di-Higgs bb irreducible background from QCD

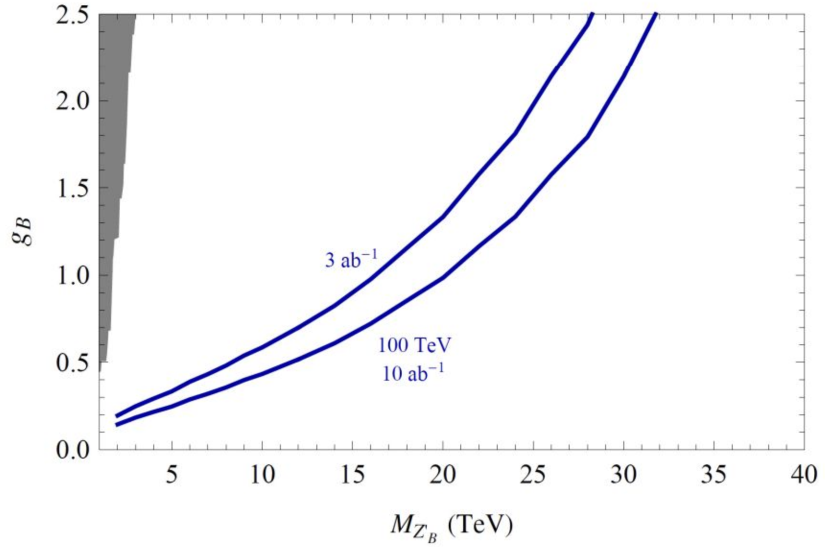
channel	σ(100 TeV) (fb)
$\mathbf{hh} \to (b\bar{b})(\ell^+\ell^-\gamma)$	0.21
$\mathbf{b\bar{b}Z}\gamma \to b\bar{b}(\ell^+\ell^-)\gamma$, $p_{T,b} > 30$ GeV	26.00×10^3
$\mathbf{t\bar{t}}\gamma \to (L^+ b\nu_L l)(L^- \bar{b}\bar{\nu}_L)\gamma$	7.94×10^3
$\mathbf{hZ}\gamma \to (b\bar{b})(\ell^+\ell^-)\gamma$	1.72
$\mathbf{b\bar{b}Z} \to b\bar{b}(\ell^+\ell^-)$+mis-tagged γ , $p_{T,b} > 30$ GeV	107.36×10^3
$\mathbf{t\bar{t}} \to (\ell^+ b\nu_\ell)(\ell'^- \bar{b}\bar{\nu}_{\ell'})$+mis-tagged γ, cuts as in Eq. 1	25.08×10^3

channel	BR	σ(14 TeV) (fb)	σ(100 TeV) (fb)
$hh \to (b\bar{b})(ZZ) \to (b\bar{b})(\ell^+\ell^-\ell'^+\ell'^-)$	0.016%	0.006	0.26
$hh \to (b\bar{b})(Z\gamma) \to (b\bar{b})(\ell^+\ell^-\gamma)$	0.013%	0.005	0.21
$hh \to (b\bar{b})(W^+W^-) \to (b\bar{b})(\ell^+\ell'^- + \not{E})$	1.658%	0.667	27.16
$hh \to (b\bar{b})(\tau^+\tau^-) \to (b\bar{b})(\ell^+\ell'^- + \not{E})$	0.893%	0.360	14.63
$hh \to (b\bar{b})(\mu^+\mu^-)$	0.025%	0.010	0.42
$hh \to (b\bar{b})(\gamma\gamma)$	0.263%	0.106	4.31

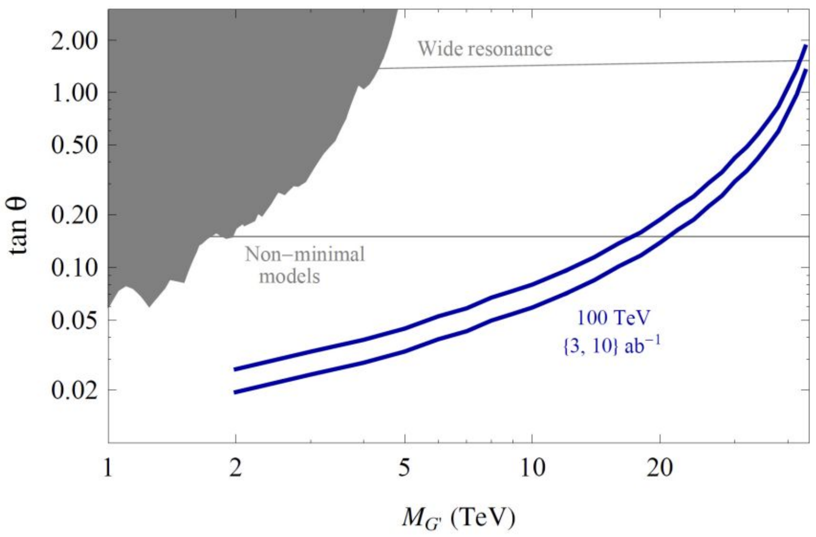
https://arxiv.org/pdf/1504.04621.pdf

channel	σ(100 TeV) (fb)
$\mathbf{hh} \to (b\bar{b})(W^+W^-) \to (b\bar{b})(\ell'^+\nu_{\ell'}\ell^-\bar{\nu}_\ell)$	27.16
$\mathbf{hh} \to (b\bar{b})(\tau^+\tau^-) \to (b\bar{b})(\ell'^+\nu_{\ell'}\bar{\nu}_\tau\ell^-\bar{\nu}_\ell\nu_\tau)$	14.63
$\mathbf{t\bar{t}} \to (\ell^+ b\nu_\ell)(\ell'^- \bar{b}\bar{\nu}_{\ell'})$, cuts as in Eq. 1	25.08×10^3
$\mathbf{b\bar{b}Z} \to b\bar{b}(\ell^+\ell^-)$, $p_{T,b} > 30$ GeV	107.36×10^3
$\mathbf{ZZ} \to b\bar{b}(\ell^+\ell^-)$	356.0
$\mathbf{hZ} \to b\bar{b}(\ell^+\ell^-)$	99.79
$\mathbf{b\bar{b}h} \to b\bar{b}(\ell^+\ell^-)$, $p_{T,b} > 30$ GeV	26.81
$\mathbf{b\bar{b}W^\pm} \to b\bar{b}(\ell^\pm\nu_\ell)$, $p_{T,b} > 30$ GeV +mis-tagged ℓ	1032.6
$\ell^+\ell^-$ + jets $\to (\ell^+\ell^-)$ + mis-tagged $b\bar{b}$	2.14×10^3

Xbb tagger:FCC-hh BSM resonances



$$\mathcal{L} \supset \frac{g_B}{6} Z'_{B\mu} \bar{q} \gamma^\mu q$$



$$\mathcal{L} \supset g_s \tan\theta \bar{q} \gamma^\mu T^a G'^a_\mu q$$

- Dijet resonances in the final states (i.e. 2 b-jets)
- Xbb tagger in Z'_B color singlet and G′ color octect vector resonances respect QCD background or to discriminate to each other
- Z'_B dijet resonance predicted in models with gauged baryon number
- Coloron G′ arises in extended $SU(3)_C$ color models as a heavy cousin of the SM gluon, and also couples universally to quarks with a coupling gs tan θ

https://arxiv.org/pdf/1308.1077.pdf

Conclusions

- Higgs in two b quarks is the most probable decay but it has a large **QCD background**
- **Xbb tagger** uses jet substructures for Hbb boosted topologies
- Our work is using **color flow variables** to perform the separation between signal and background
- ML techniques show **good result** in discrimination, with ROC around 0.89
- This approach can be useful for future colliders in SM (i.e. Di-Higgs) or BSM (i.e. Dijet resonances) processes

BACKUP

ML parameters

BDT

Parameters	Value
No. of Trees	100
Max Depth	3
MinNodeSize	2.5%
Boost Type	AdaBoost
Train/Test	50/50
No. of Cuts	200
Downsampling	No

CNN

Parameter	Value
N_1 Conv2D	30
N_2 Conv2D	30
Dropouts	– (0.3)
N_3 Conv2D	30
Dropouts	- (0.3)
N_4 Conv2D	10
Dropouts	- (0.1)
Flat Layer	150
Epochs	30
Batch Size	800