UNIVERSITÀ DI PAVIA
Dipartimento di Fisica

Coloring the jets at LHC for Xbb tagger improvement

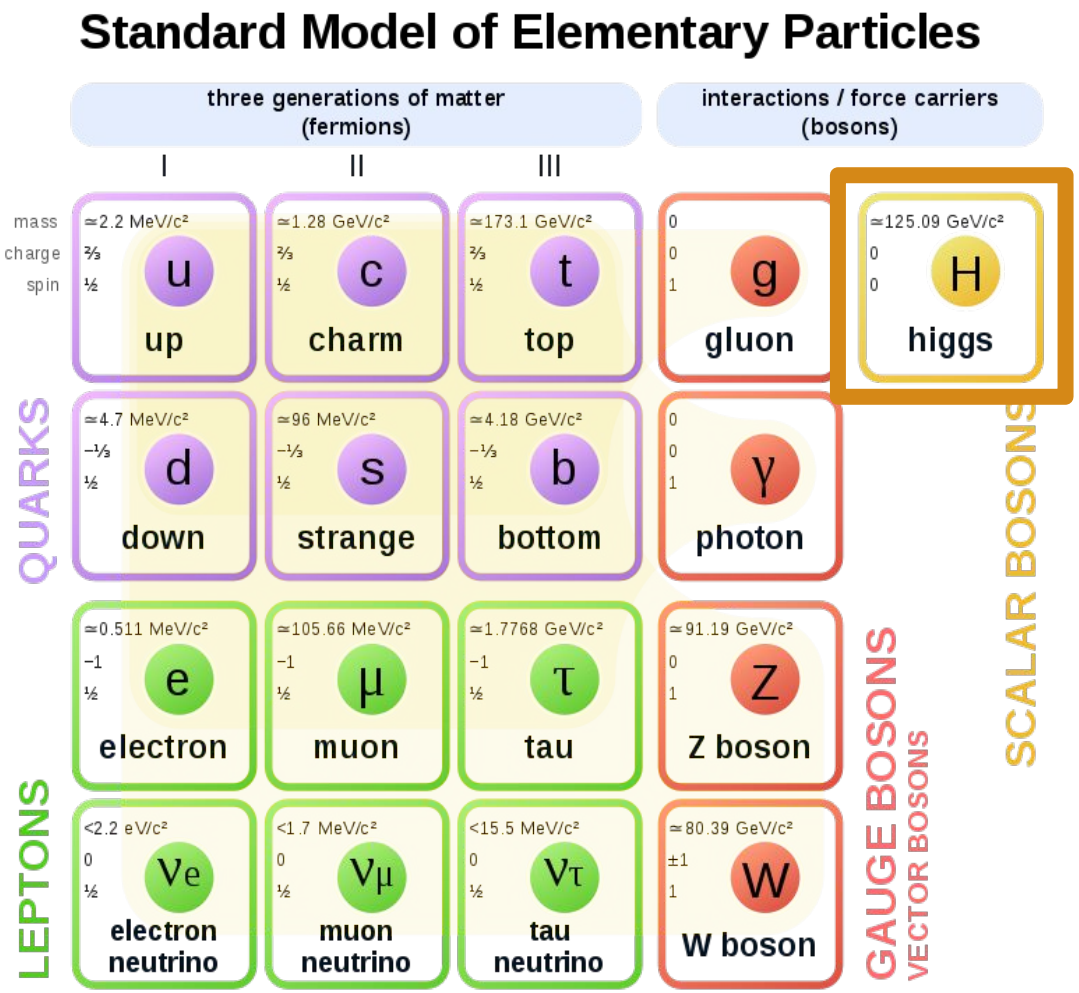
Giulia Manco

XX FRASCATI SUMMER SCHOOL "BRUNO TOUSCHEK"
in Nuclear, Subnuclear and Astroparticle Physics

ATLAS EXPERIMENT

INFN
Istituto Nazionale di Fisica Nucleare

Introduction: Higgs Boson



10 years from the discovery!

Introduction: Higgs Boson

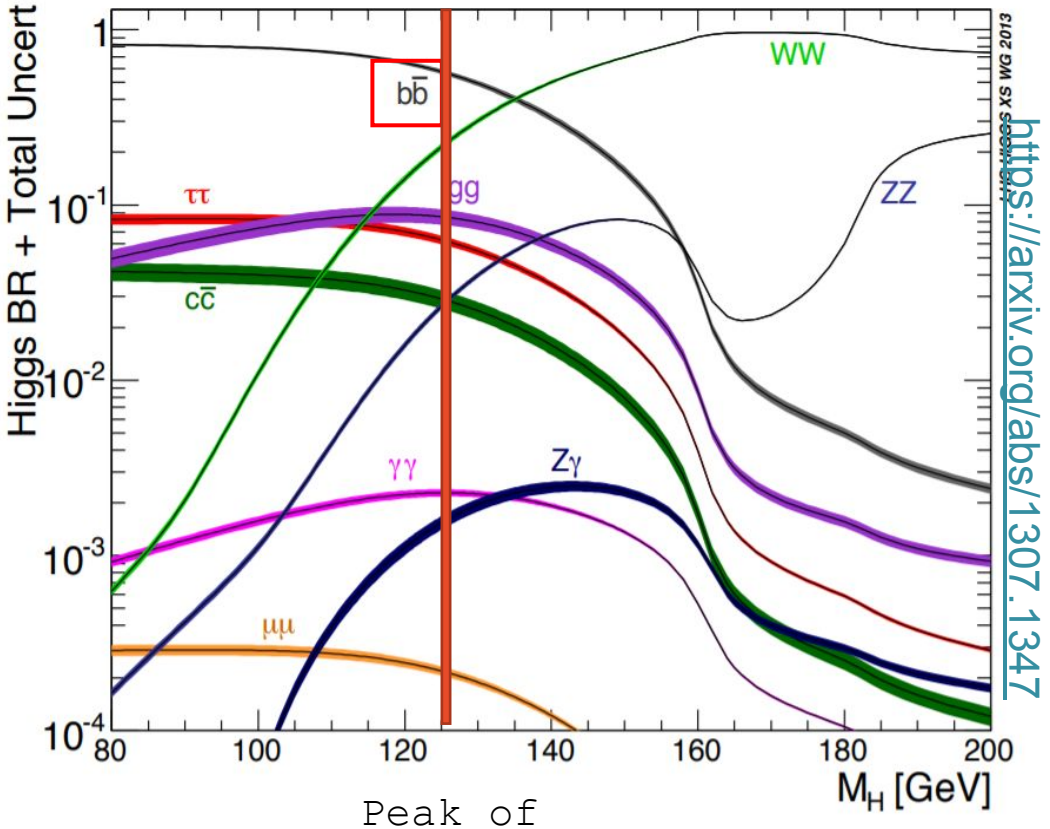
Why do we like him/her/them?

- **Strength of interaction** related to **mass particles**
  - The only interaction that **distinguishes** between the fermion generations
  - Important to study interactions with **third and second generation** fermions
- Interacts also with itself (vacuum)

Higgs bb decay

$$H \rightarrow b\bar{b}$$

PRO
Max probability
Study **heavy** fermions



https://arxiv.org/abs/1307.1347

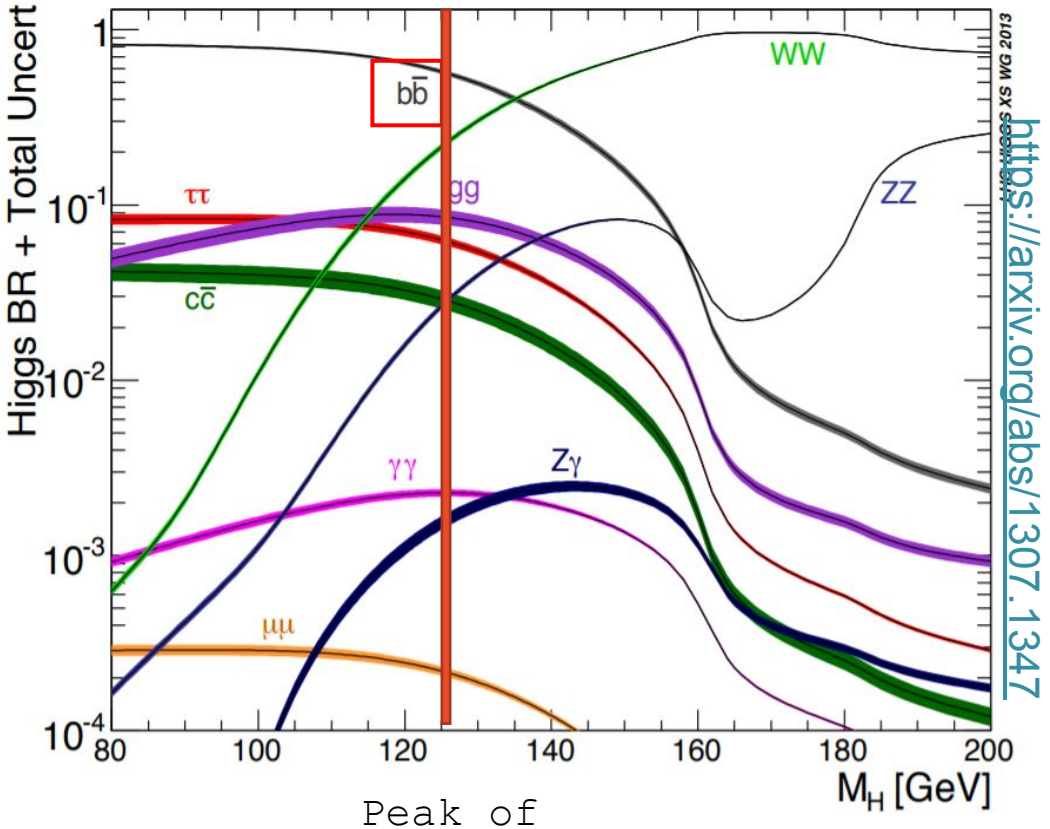
Peak of Higgs mass

Higgs bb decay

$$H \rightarrow b\bar{b}$$

PRO
Max probability
Study **heavy** fermions

CONTRO
b jets identification



https://arxiv.org/abs/1307.1347

Peak of Higgs mass

Higgs bb decay

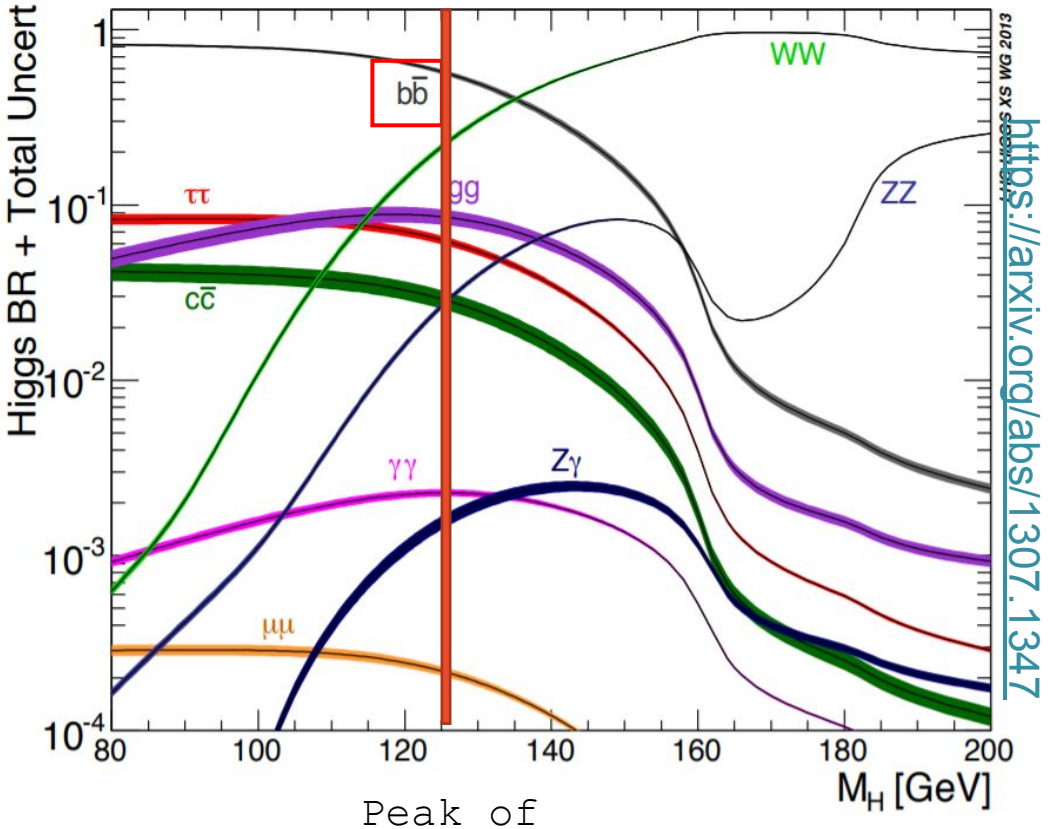
$$H \rightarrow b\bar{b}$$

PRO
Max probability
Study **heavy** fermions

CONTRO
b jets identification
Enormous background from QCD (**g → bb**)

Need to discriminate **signal** and **background**

Higgs **gluon**

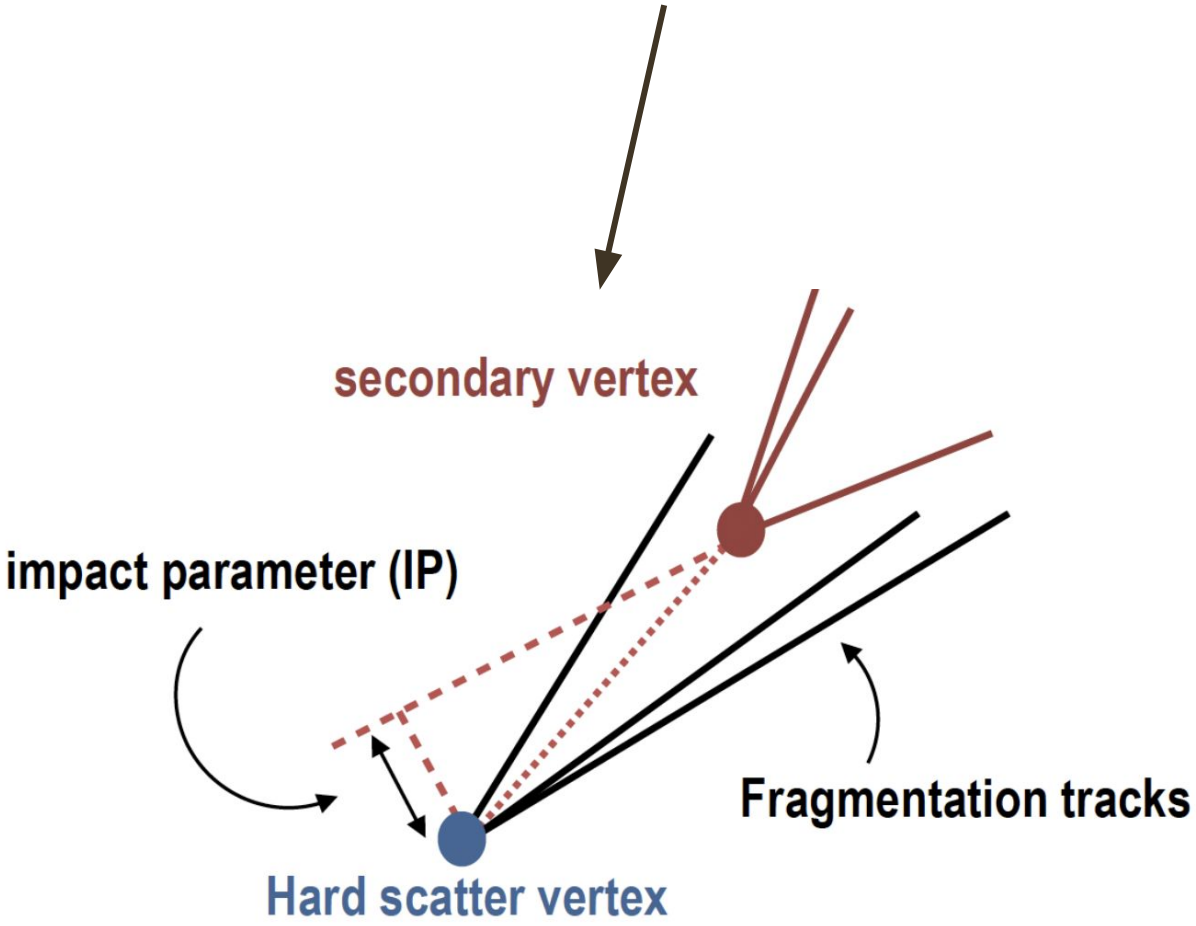


https://arxiv.org/abs/1307.1347

Peak of Higgs mass

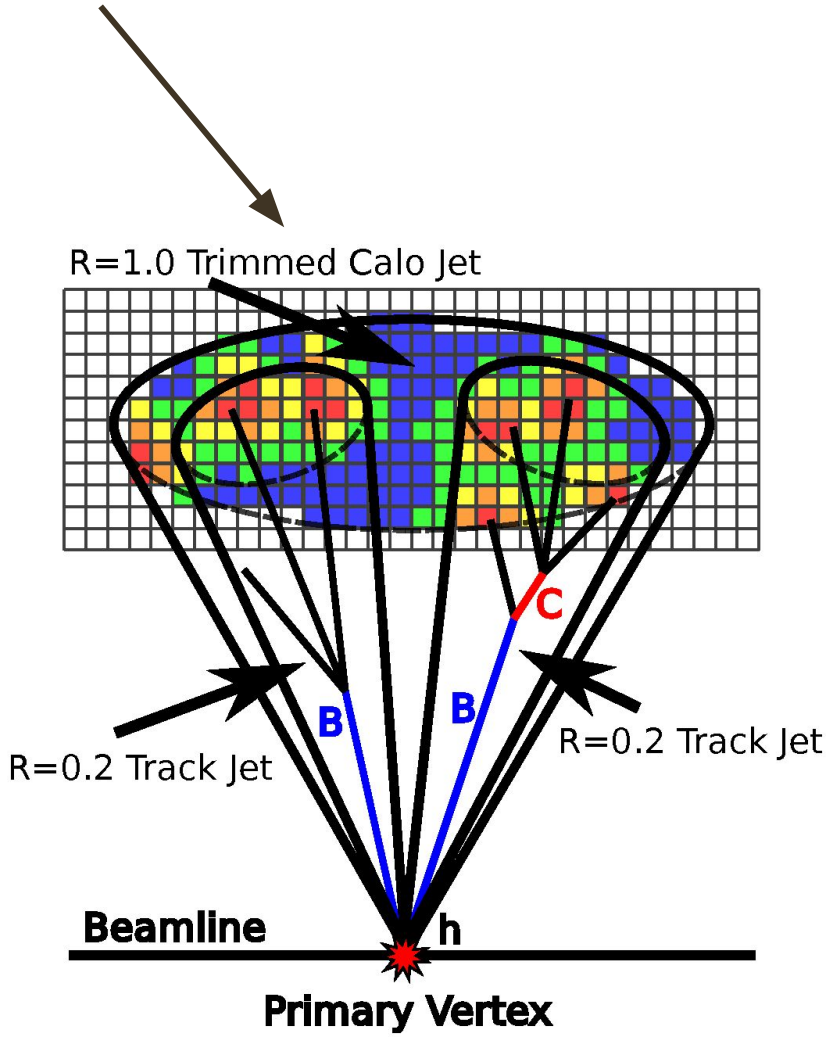
Xbb tagger

b tagging + jet substructures in boosted H->bb topologies in ATLAS



Xbb tagger

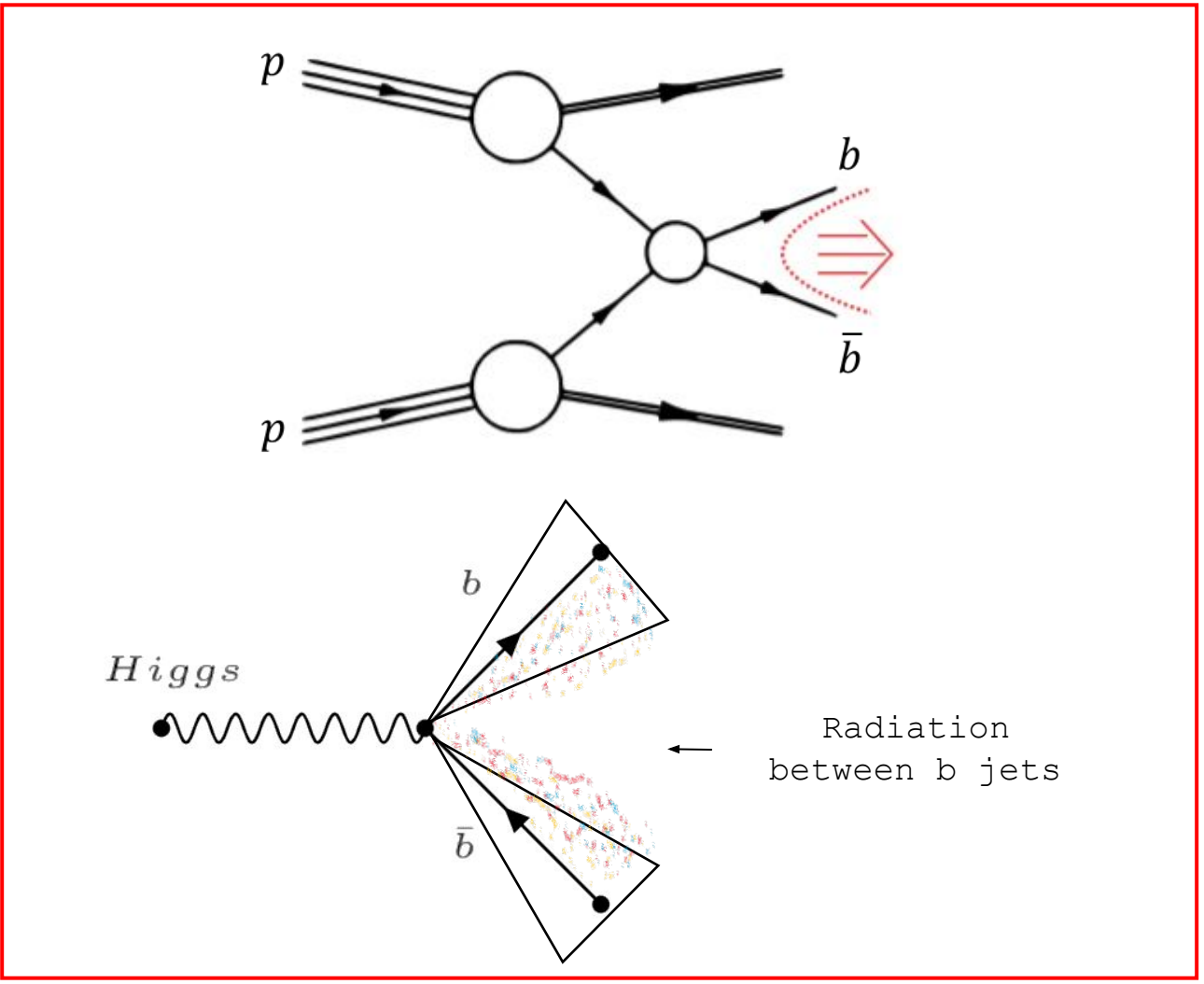
b tagging + **jet substructures** in **boosted** H->bb topologies in ATLAS



COLOR CONNECTIONS

COLOR CONNECTIONS

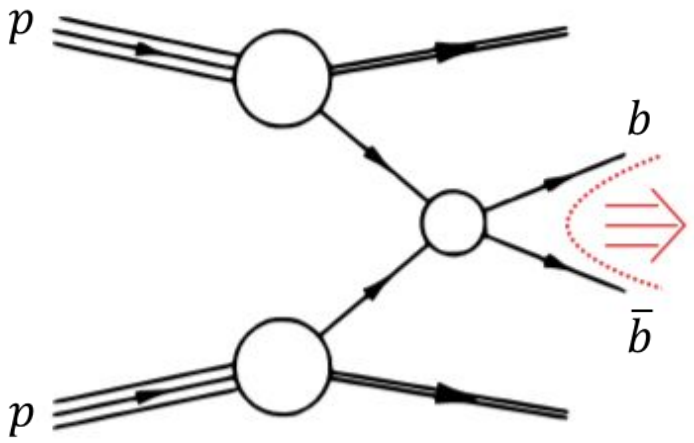
SIGNAL

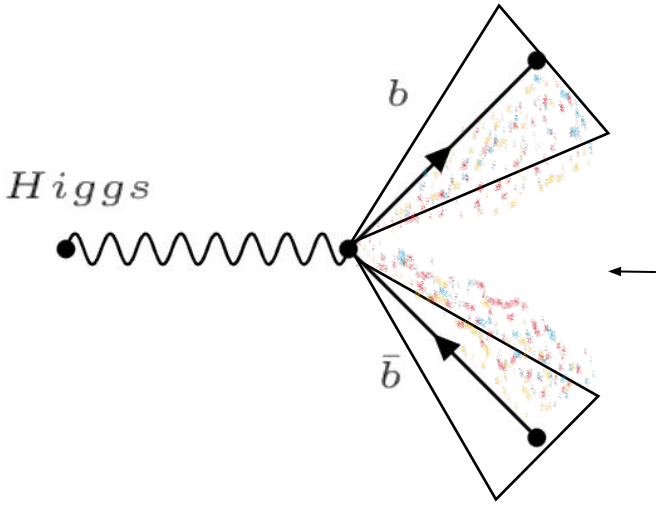


https://arxiv.org/pdf/1001.5027.pdf

COLOR CONNECTIONS

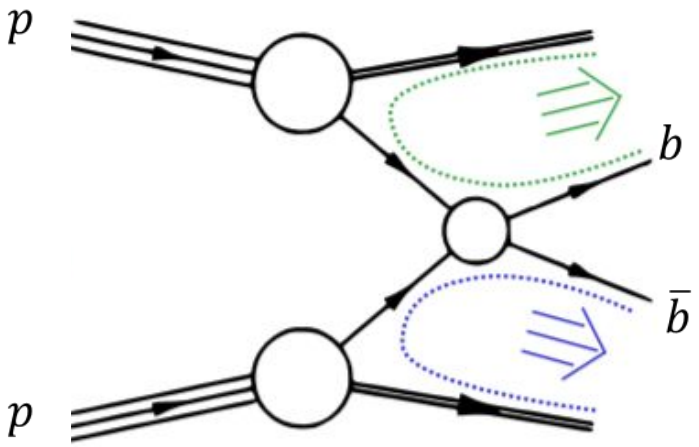
SIGNAL

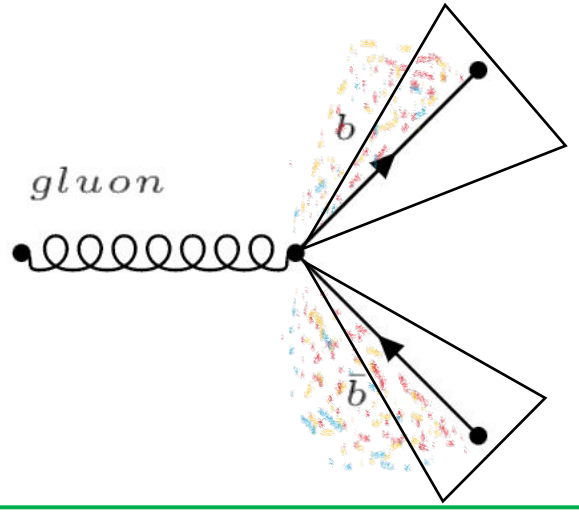




Radiation between b jets

BACKGROUND

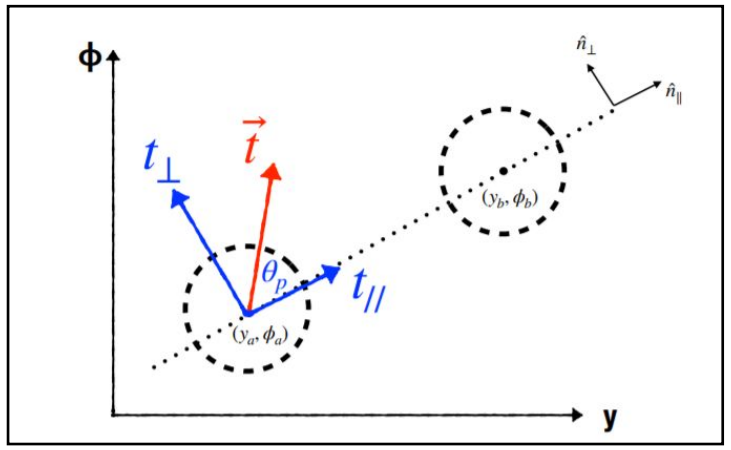




Radiation outside the b jets

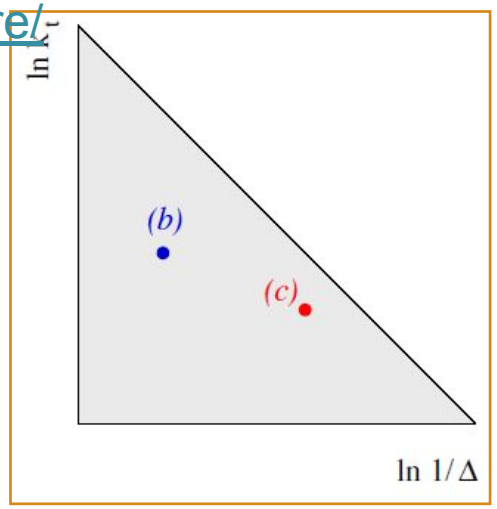
Color sensitive variables

Variables sensitive to color flow



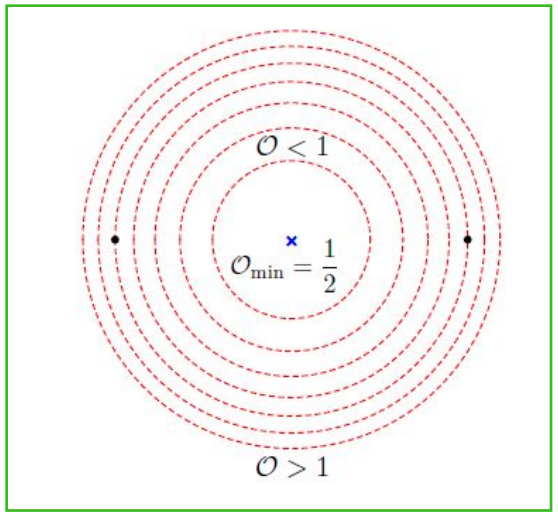
PULL VECTOR

https://inspirehep.net/literature/1764711



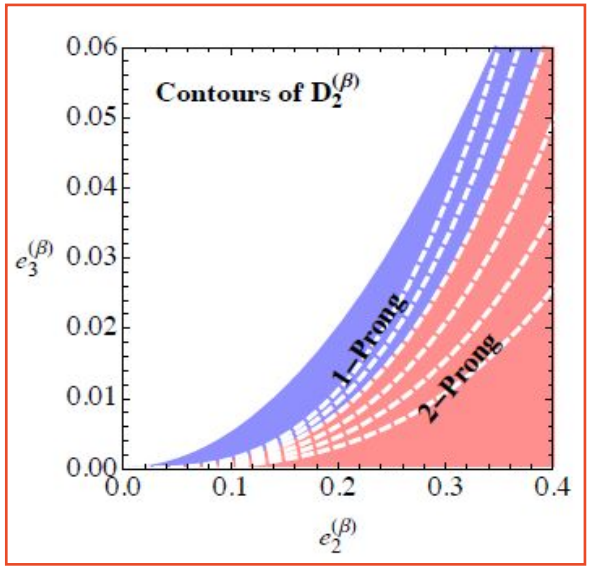
LUND PLANE

https://arxiv.org/abs/1807.04758



COLOR RING

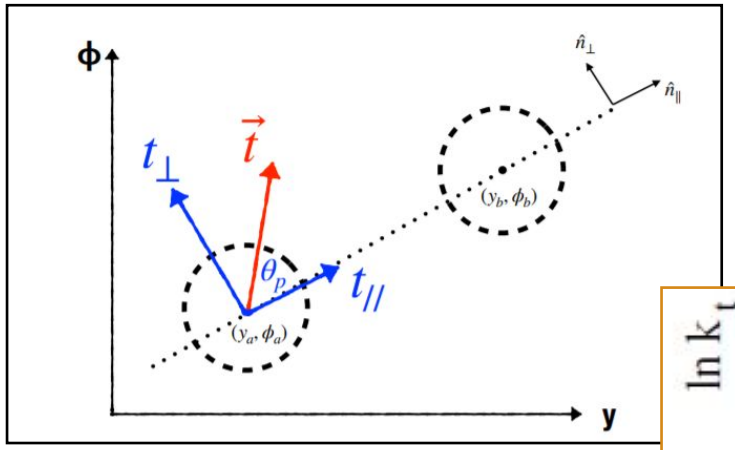
https://arxiv.org/abs/2006.10480



D_2

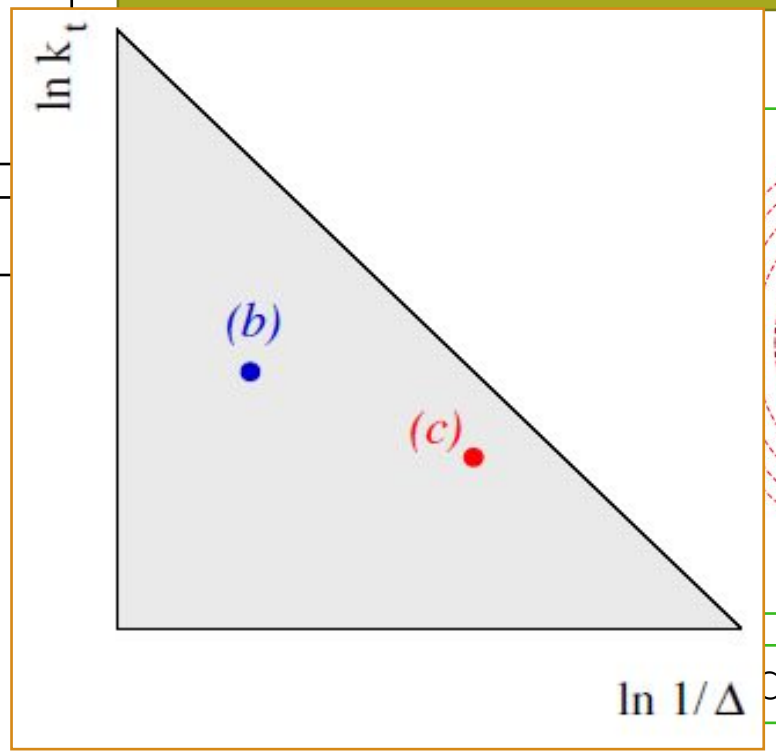
https://arxiv.org/abs/1409.6298

Color sensitive variables

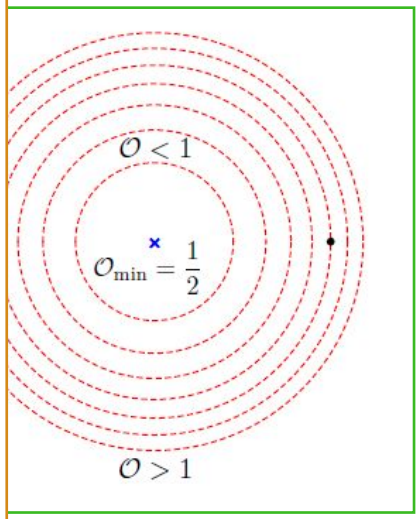


PULL VECTOR

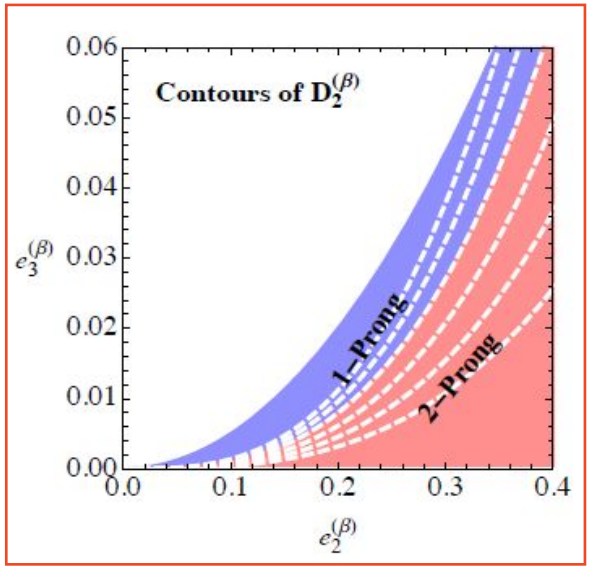
Variables sensitive to color flow



LUND PLANE

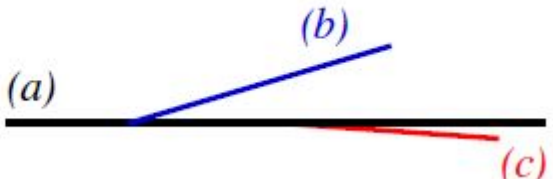


COLOR RING



D_2

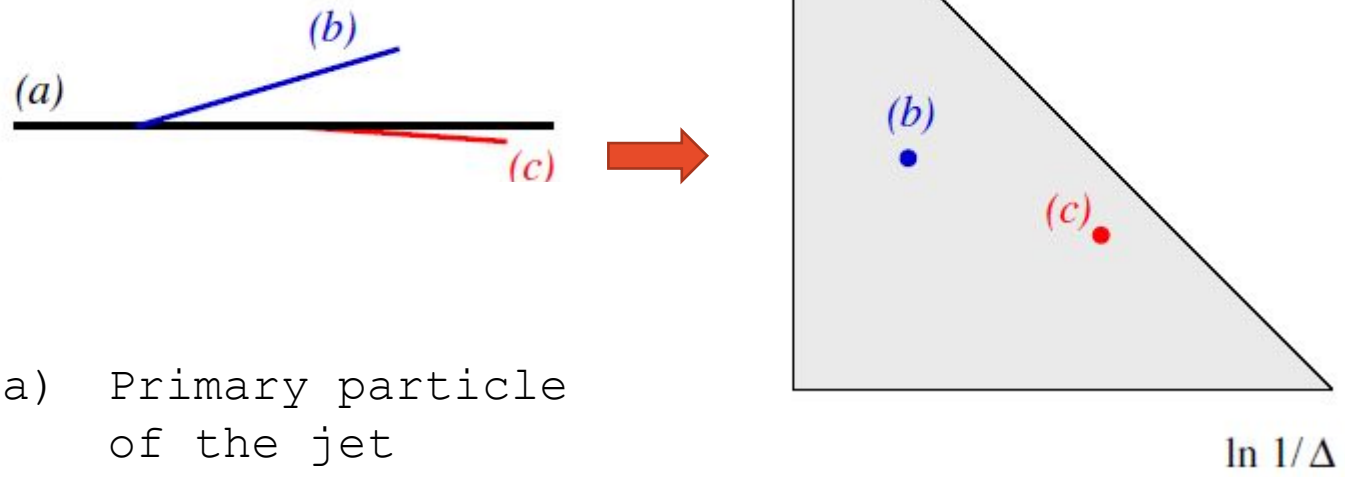
color sensitive variables:



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

DECLUSTERING

color sensitive variables:

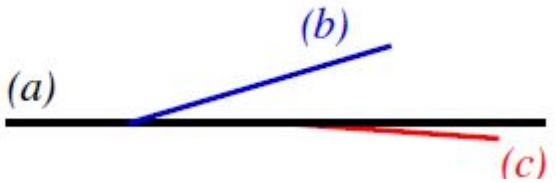


(a) Primary particle of the jet
(b) Emission
(c) Emission

DECLUSTERING

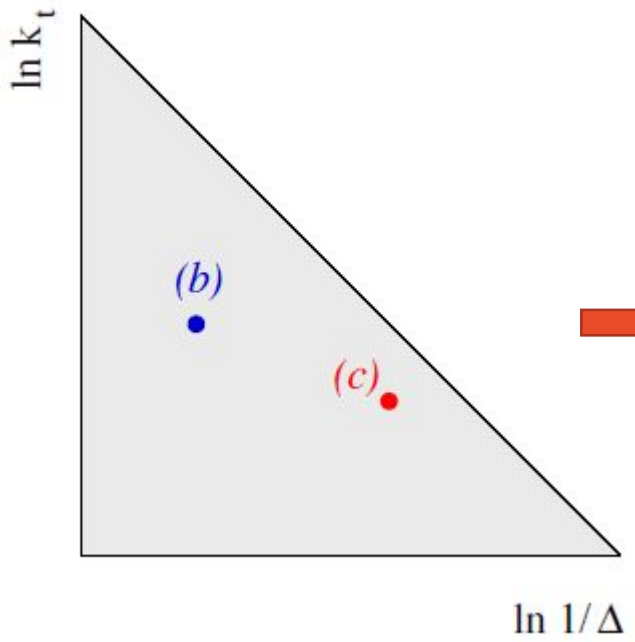
LUND PLANE

color sensitive variables:

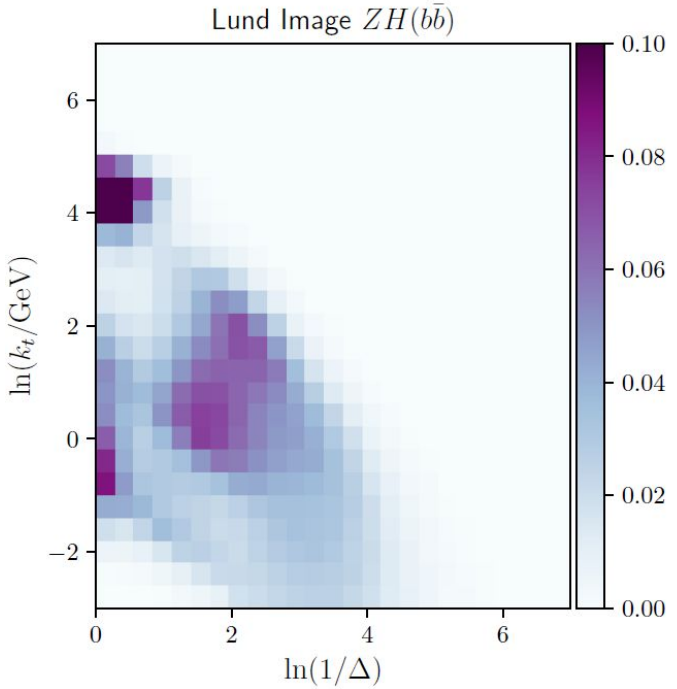


(a) Primary particle of the jet
(b) Emission
(c) Emission

DECLUSTERING

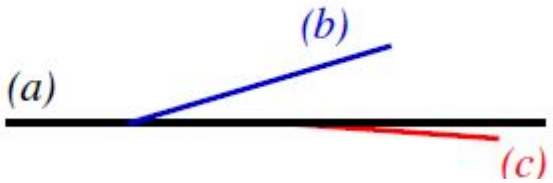


LUND PLANE



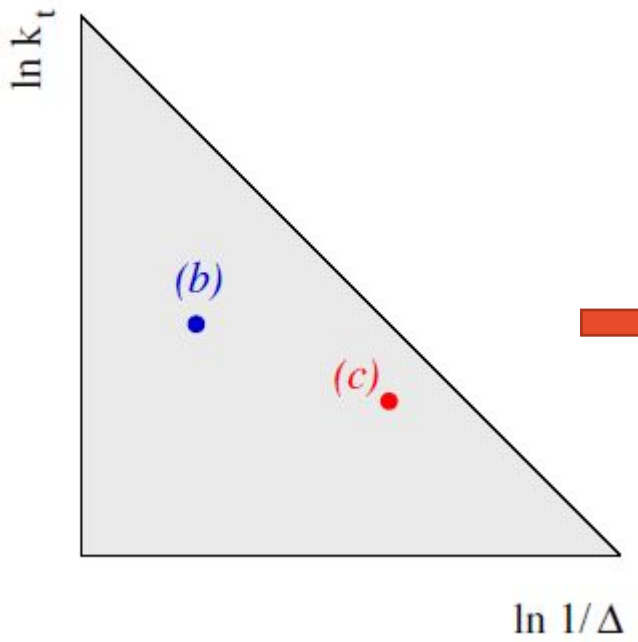
2D histogram
LUND PLANE IMAGE

color sensitive variables:

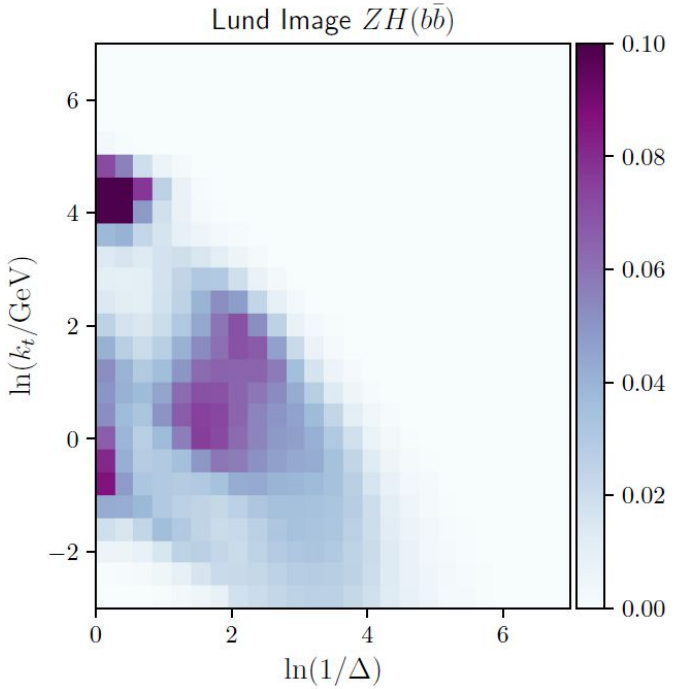


(a) Primary particle of the jet
(b) Emission
(c) Emission

DECLUSTERING



LUND PLANE



2D histogram
LUND PLANE IMAGE

CONVOLUTIONAL NEURAL NETWORK

Phenomenolgical 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.

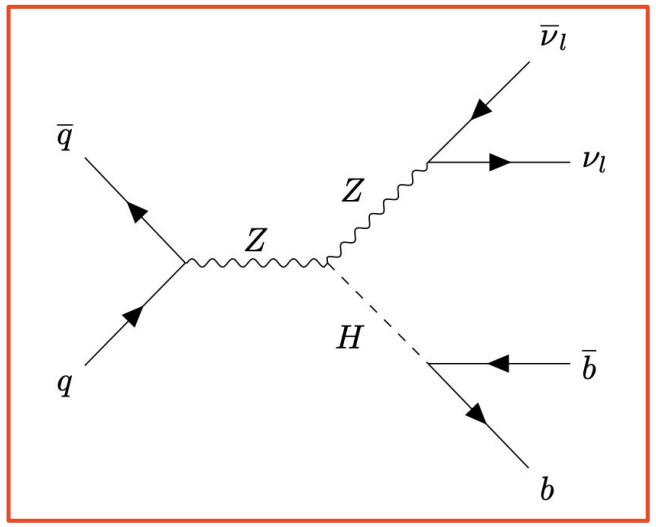
Comments:	12 pages, 5 figures, 5 tables. v2 matches published version
Subjects:	**High Energy Physics - Phenomenology (hep-ph)**; High Energy Physics - Experiment (hep-ex)
Report number:	ZU-TH 50/21
Cite as:	arXiv:2112.09650 **[hep-ph]**
	(or arXiv:2112.09650v2 **[hep-ph]** for this version)
	https://doi.org/10.48550/arXiv.2112.09650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

Submission history

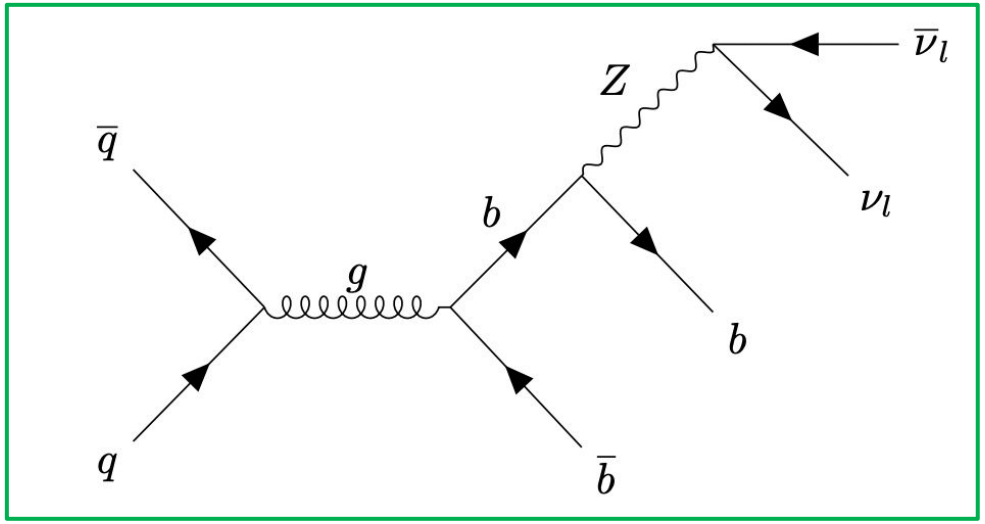
From: Giovanni Stagnitto [view email]
[v1] Fri, 17 Dec 2021 17:51:33 UTC (1,102 KB)
[v2] Wed, 1 Jun 2022 07:47:57 UTC (1,102 KB)

Phenomenolgical studies

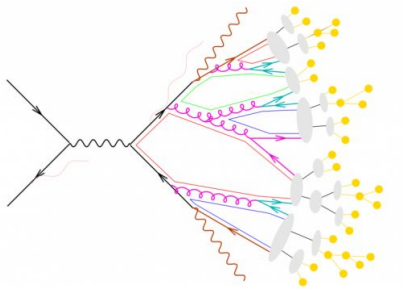
Simulation of signal and background



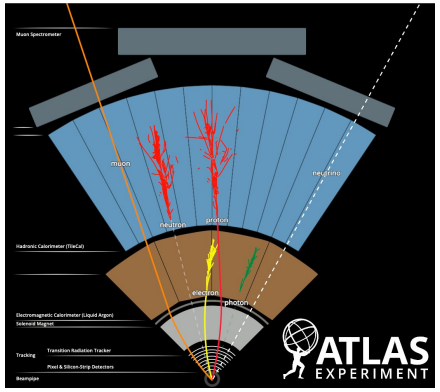
SIGNAL



BACKGROUND



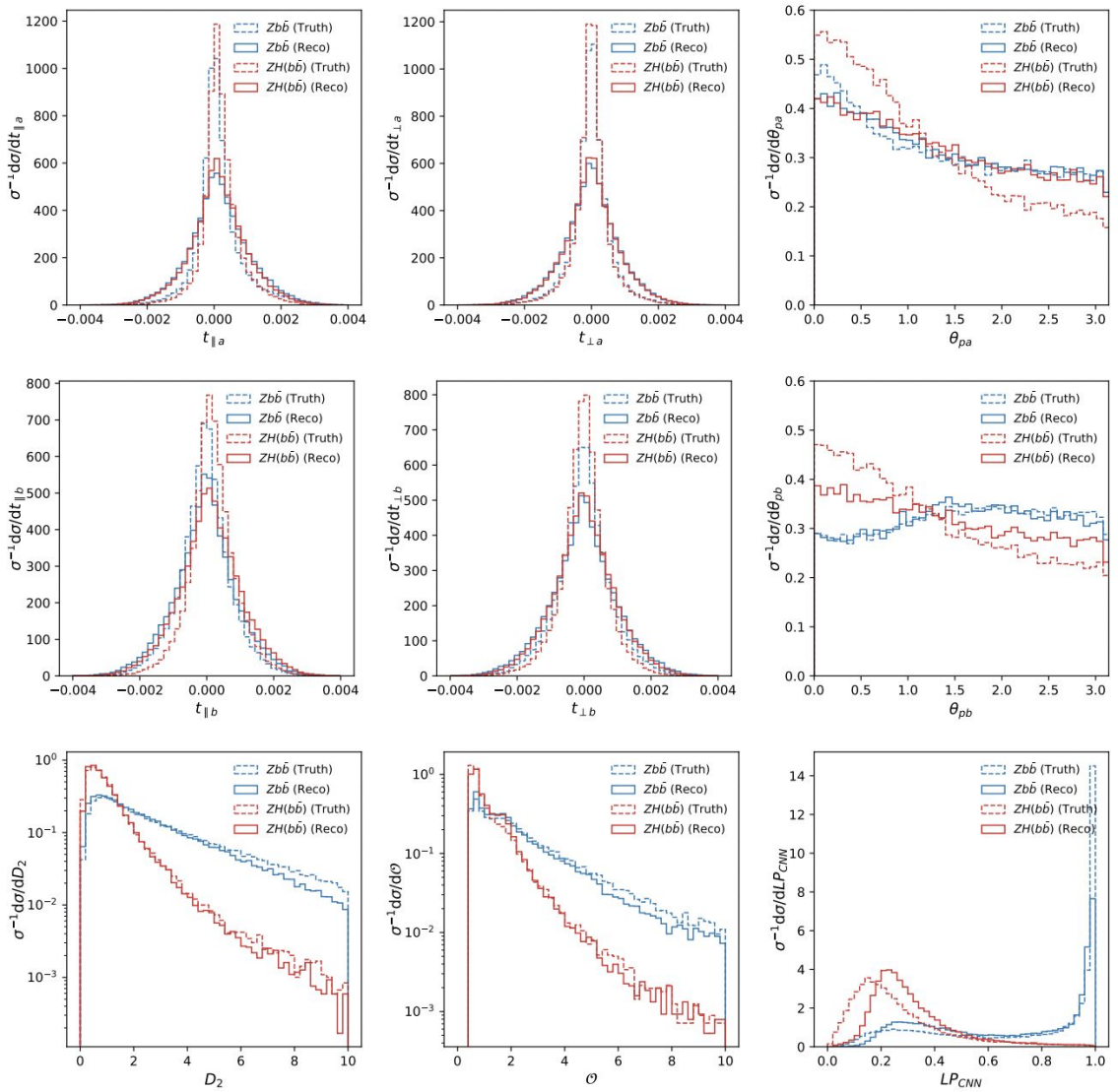
Parton level + shower
MadGraph + Pythia



Detector response: Delphes

Phenomenolgical studies

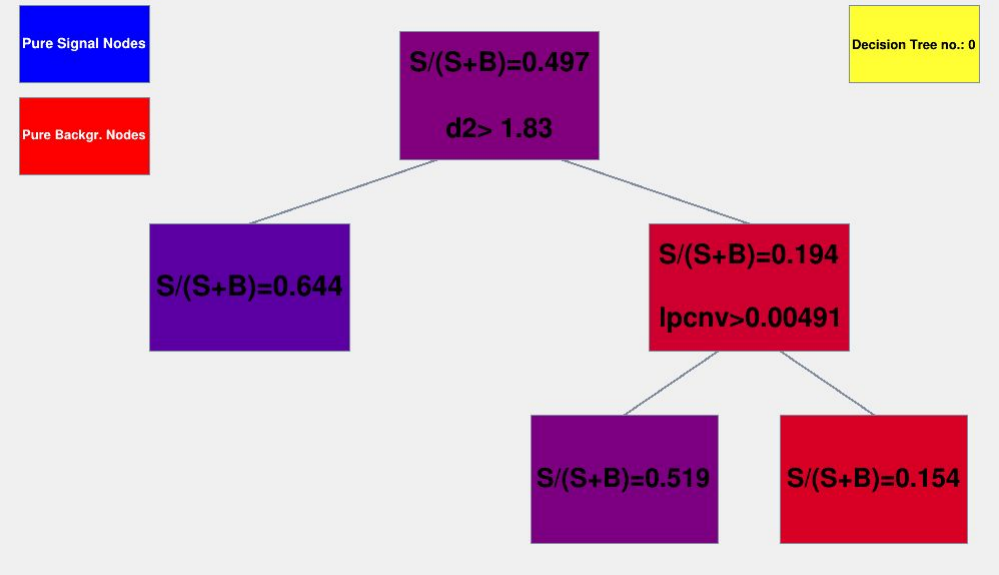
Variable extraction



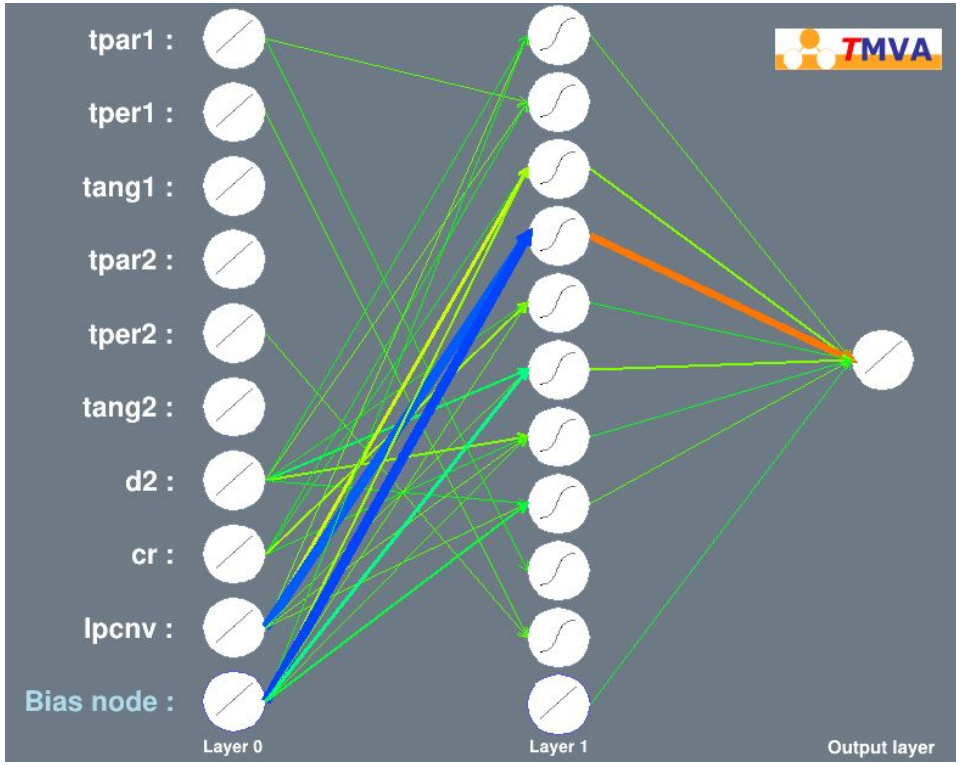
Phenomenolgical studies

Machine Learning Algorithms

BOOSTED DECISION TREE (BDT)



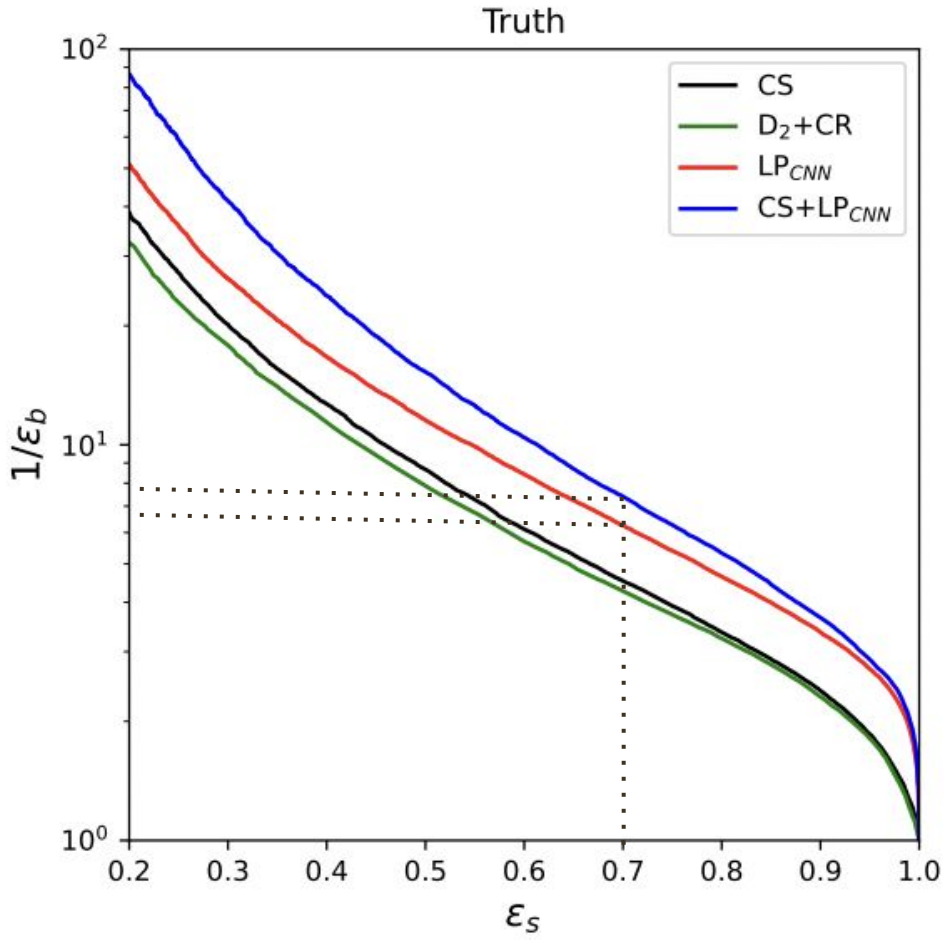
MULTILAYER PERCEPTRON (MLP)



Phenomenolgical studies

Results

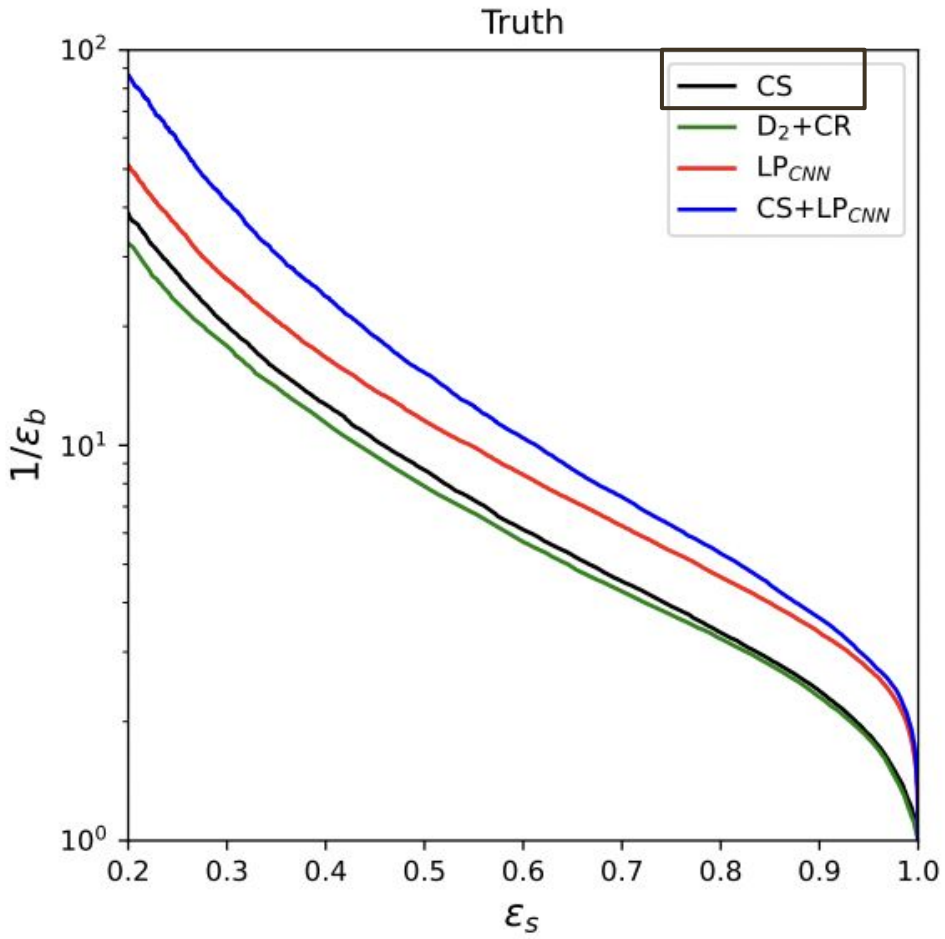
ROC(BDT)

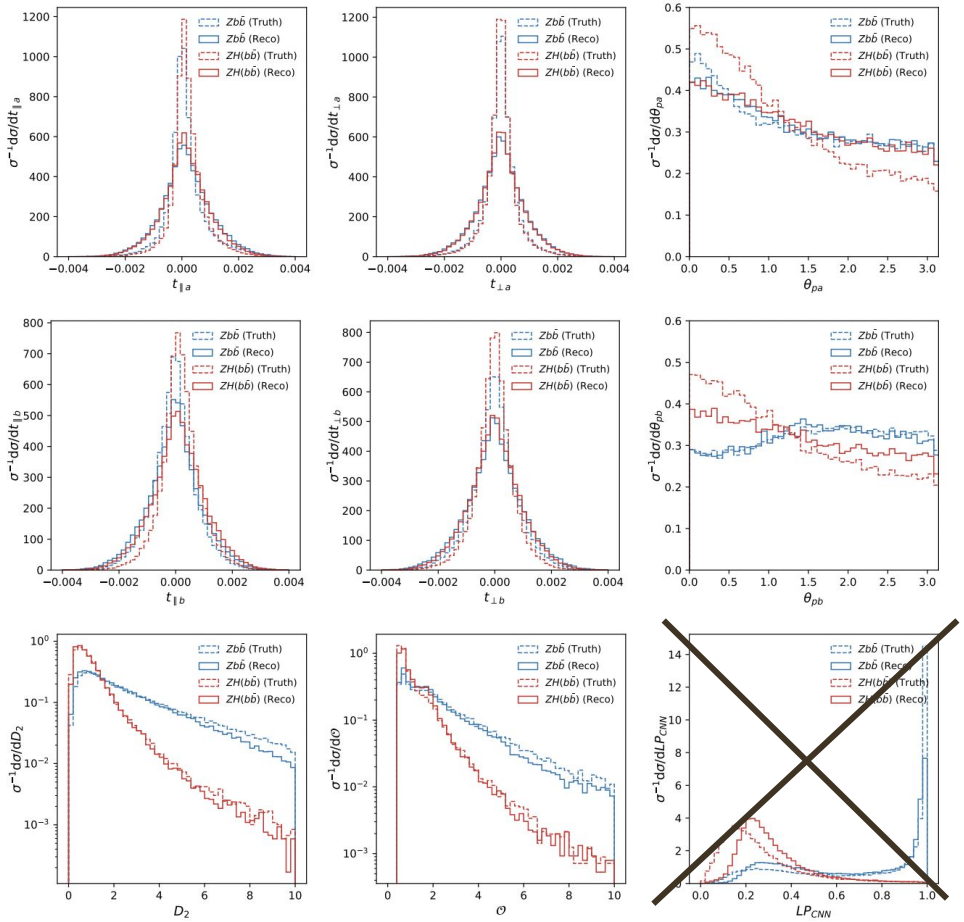


Phenomenolgical studies

Results

ROC

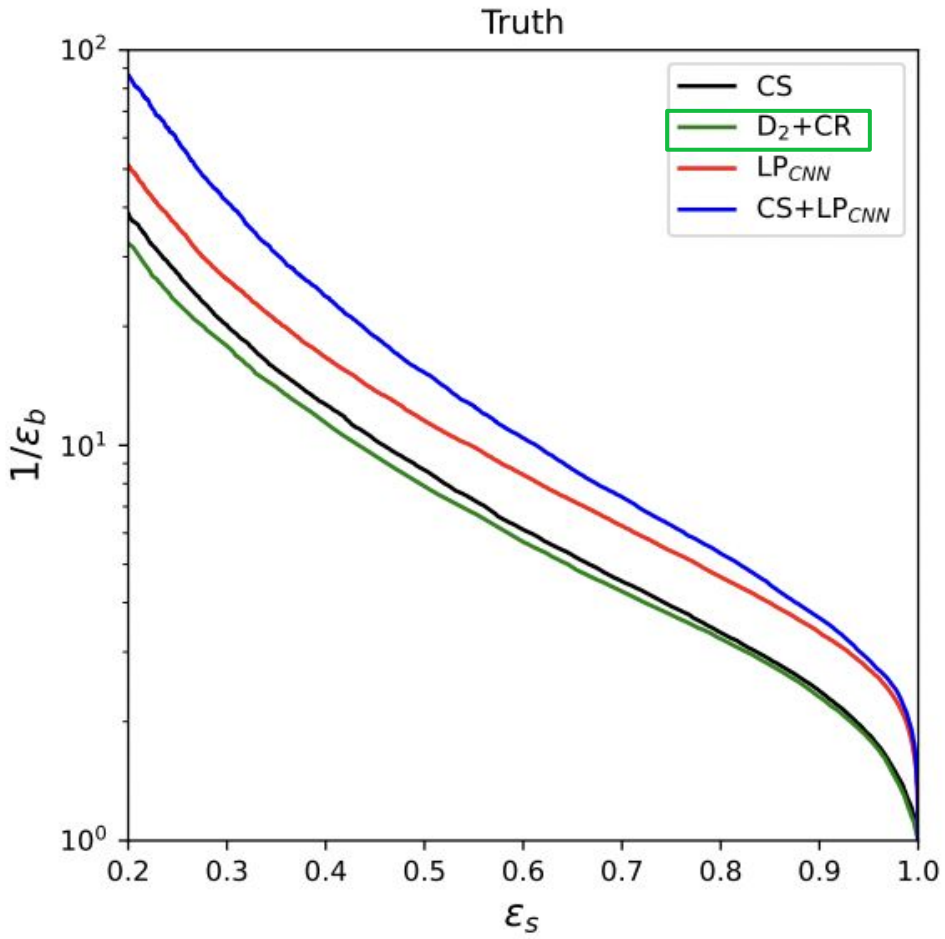


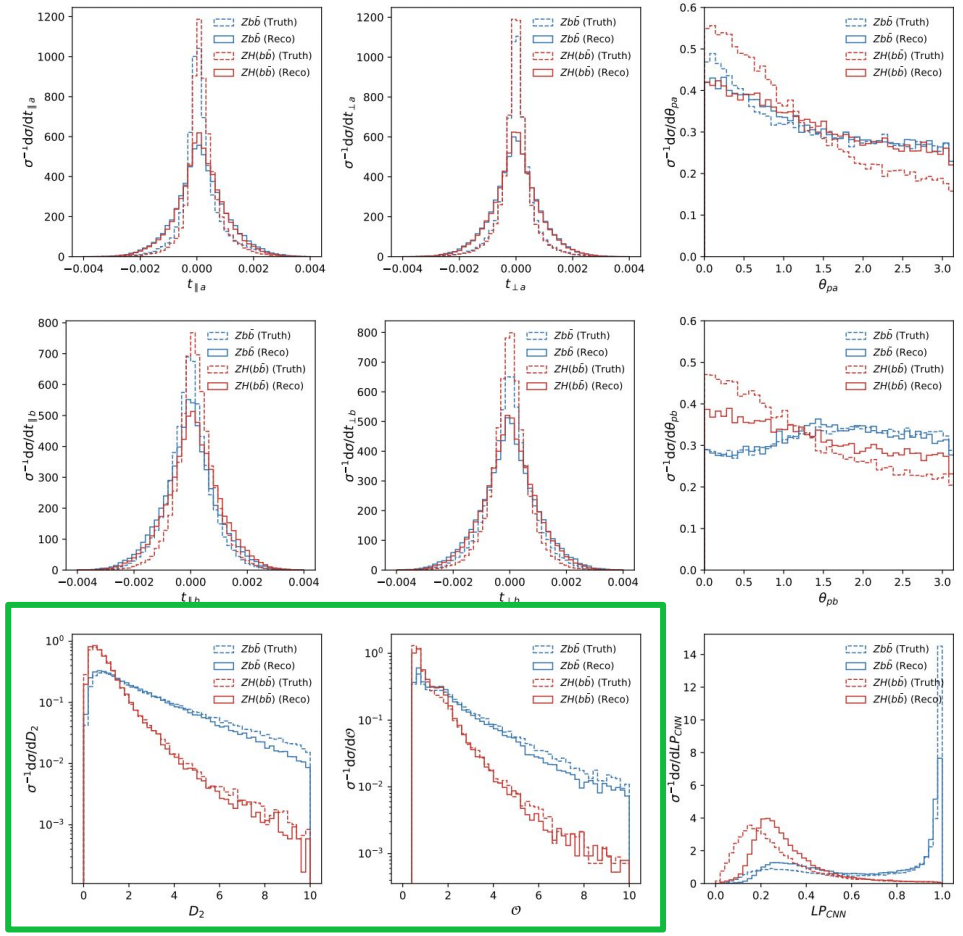


Phenomenolgical studies

Results

ROC

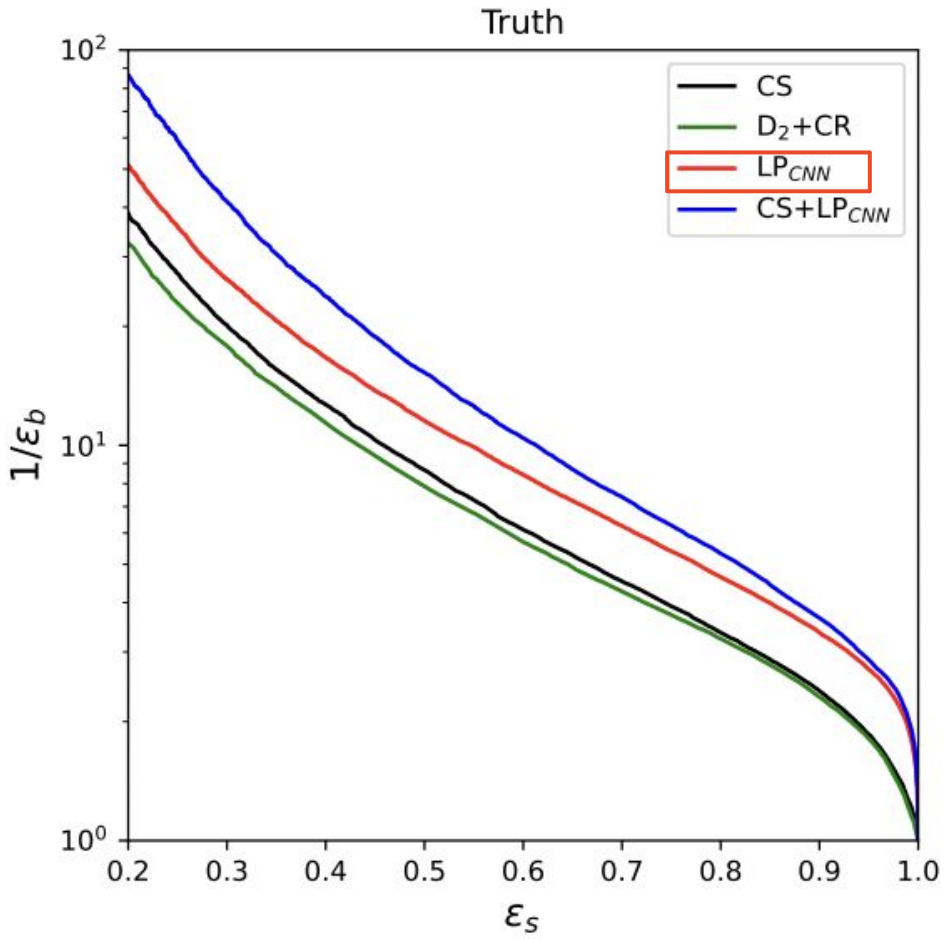


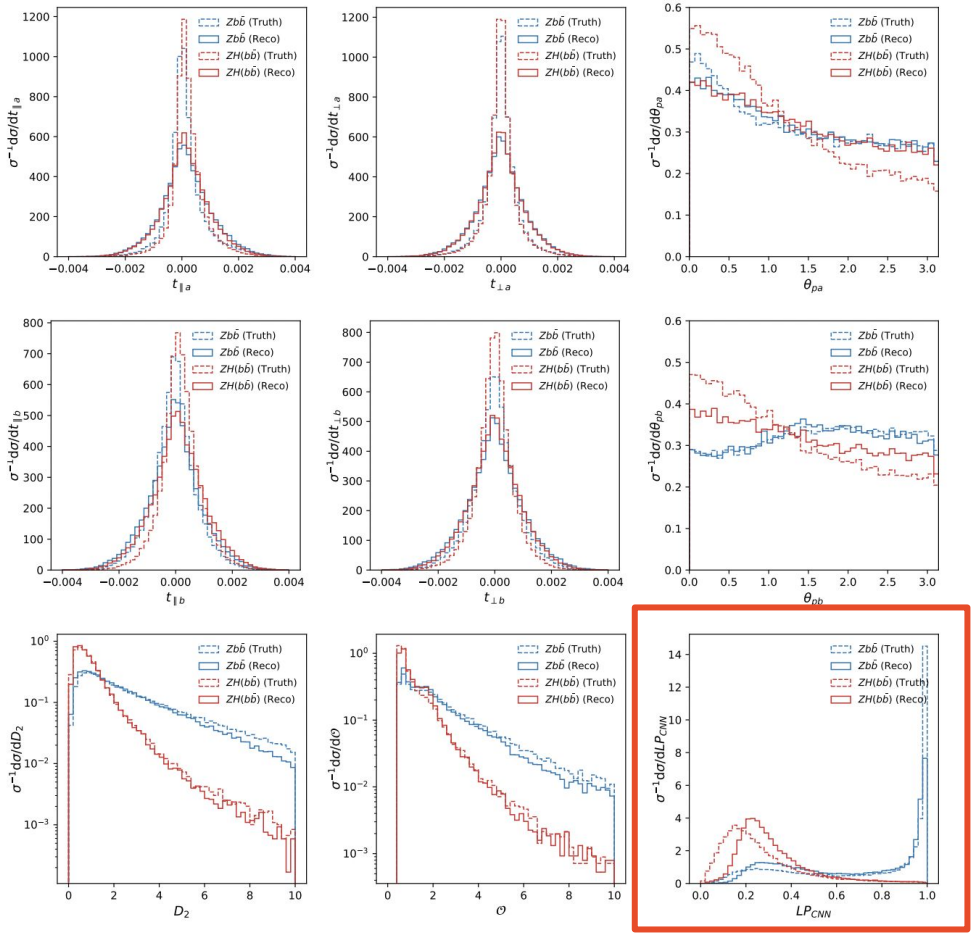


Phenomenolgical studies

Results

ROC

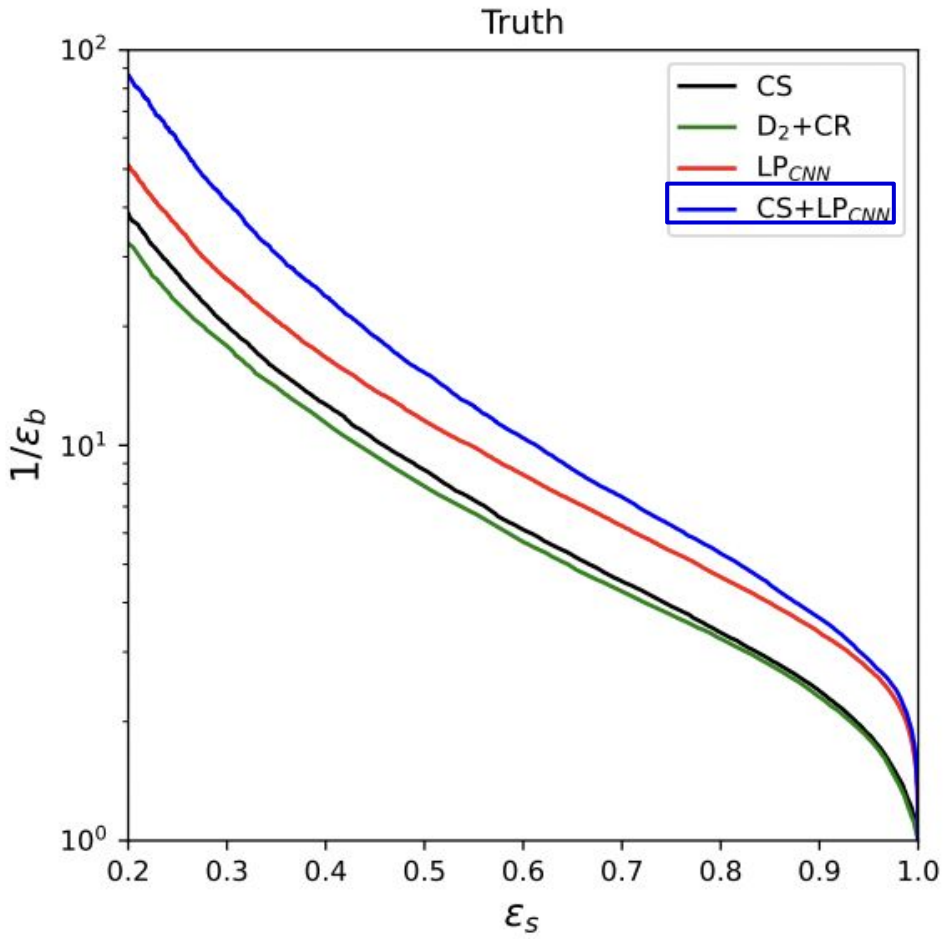


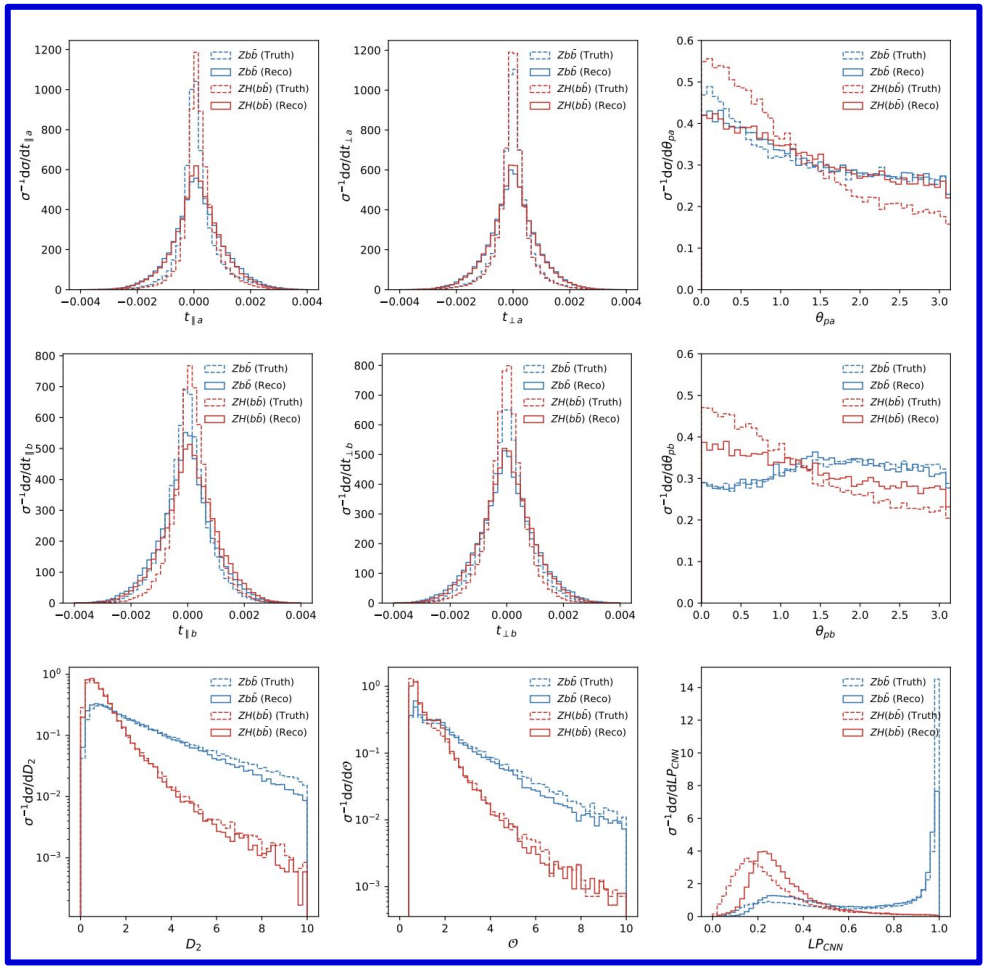


Phenomenolgical studies

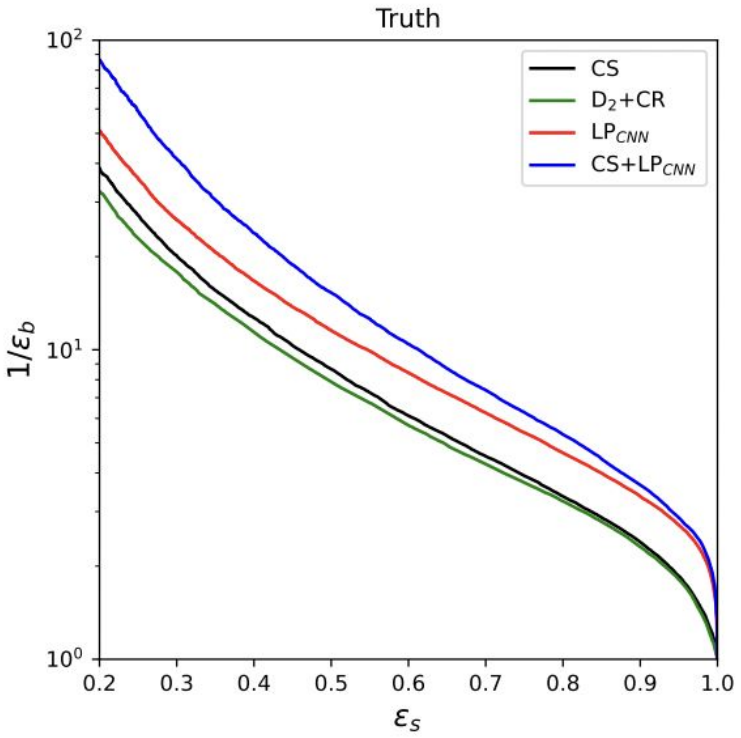
Results

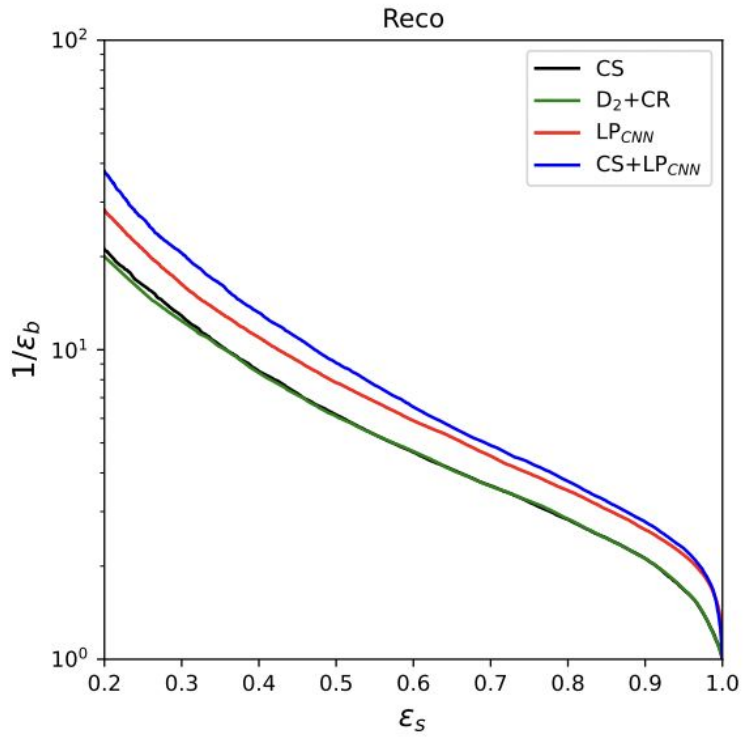
ROC





Phenomenolgical studies





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

Observable Ranking				
	Truth		Reco	
Rank	Obs.	Importance	Obs.	Importance
1	$\mathrm{LP_{CNN}}$	6.6×10^{-1}	$\mathrm{LP_{CNN}}$	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}

Conclusions

- Higgs boson celebrates 10 years, but still we have a lot to discover.
- Higgs in two b quarks is the most probable decay but it has a large **QCD background.**
- **Xbb tagger** combines b tagging and 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.
- Now we are working in the Xbb tagger framework, testing the performances with **ATLAS simulations.**

Special thanks to:

Daniela Rebuzzi

Alberto Rescia

Giovanni Stagnitto

Luca Cavallini

Charanjit Khosa

Fabrizio Parodi

Andrea Coccaro

Simone Marzani

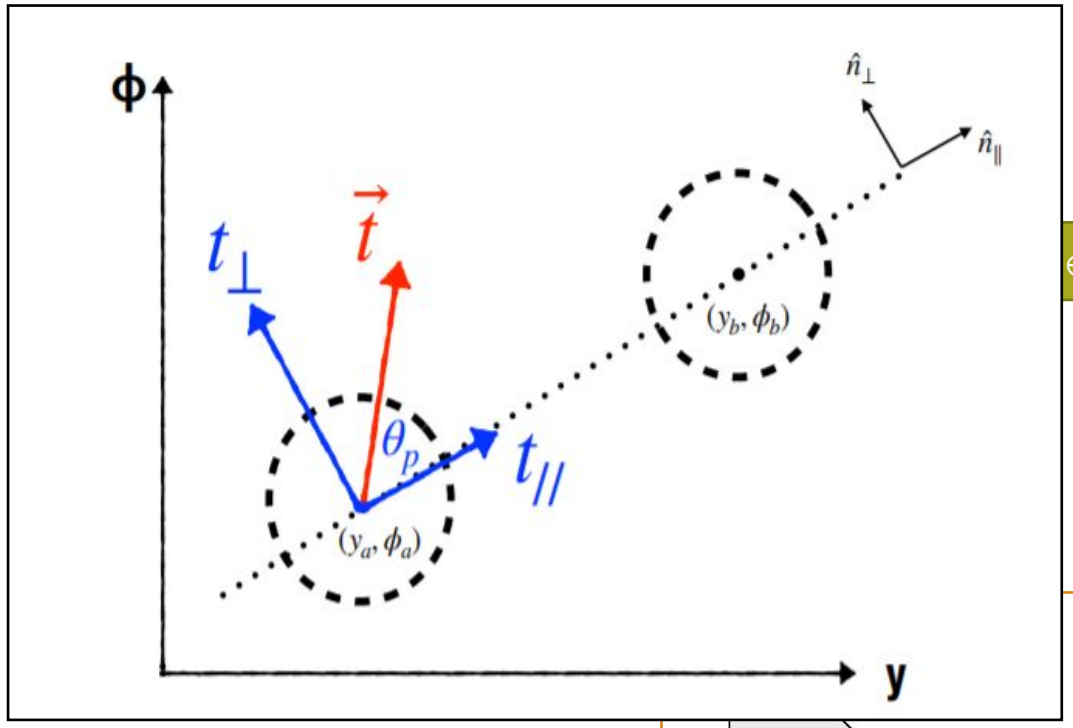
Carlo Schiavi

Federico Sforza

Thank you!

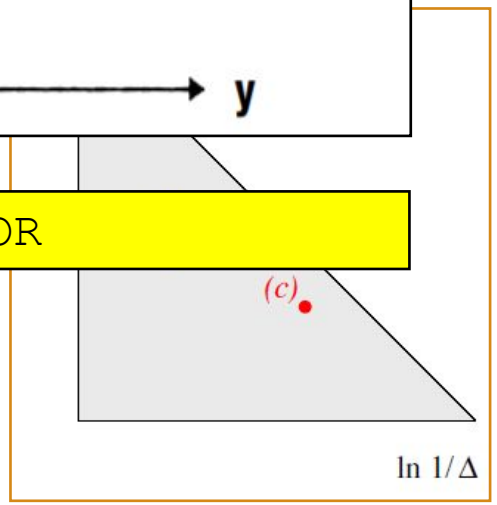
BACKUP

color sensitive variables

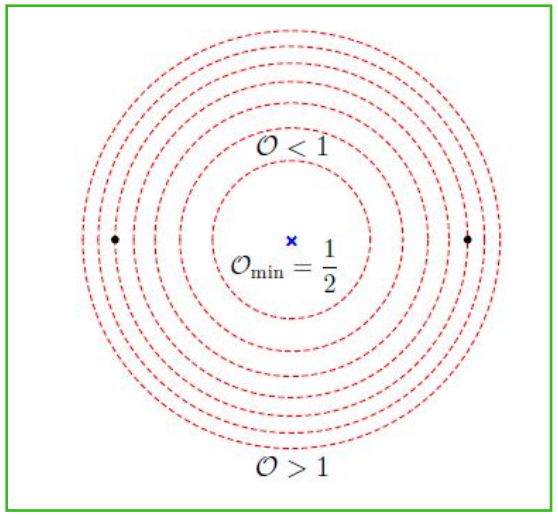


PULL VECTOR

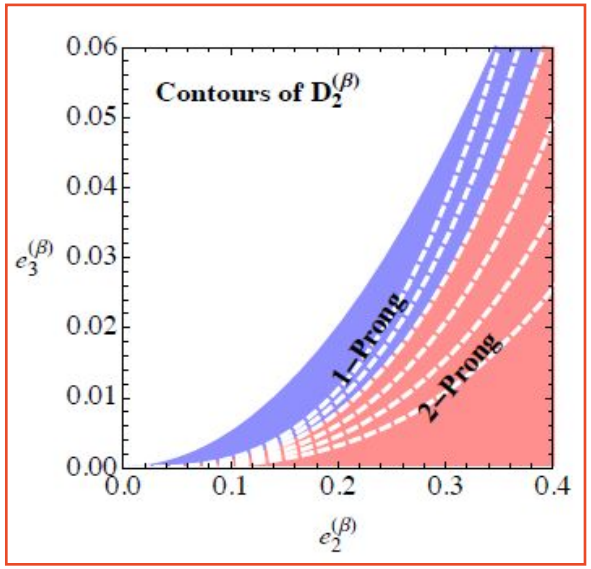
es sensitive to color flow



LUND PLANE

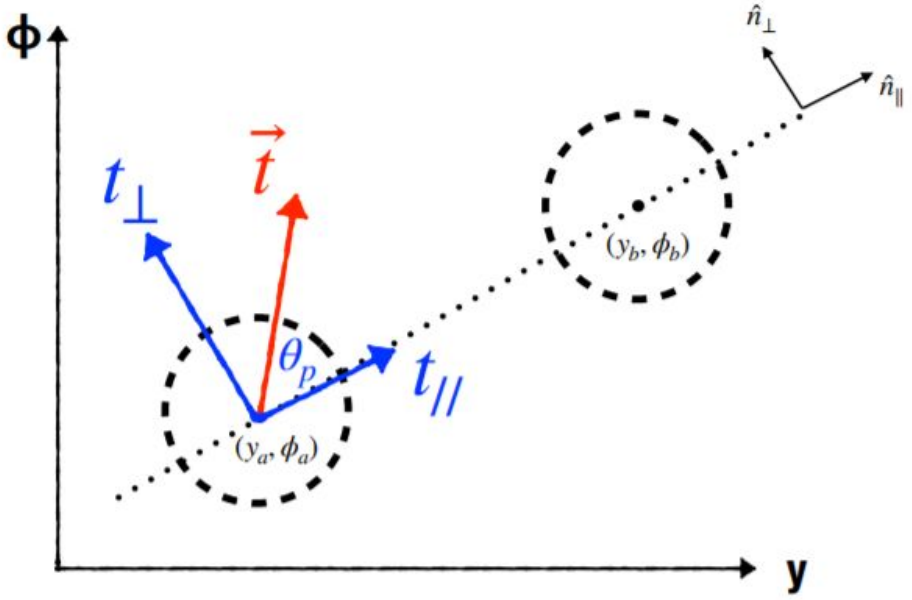


COLOR RING

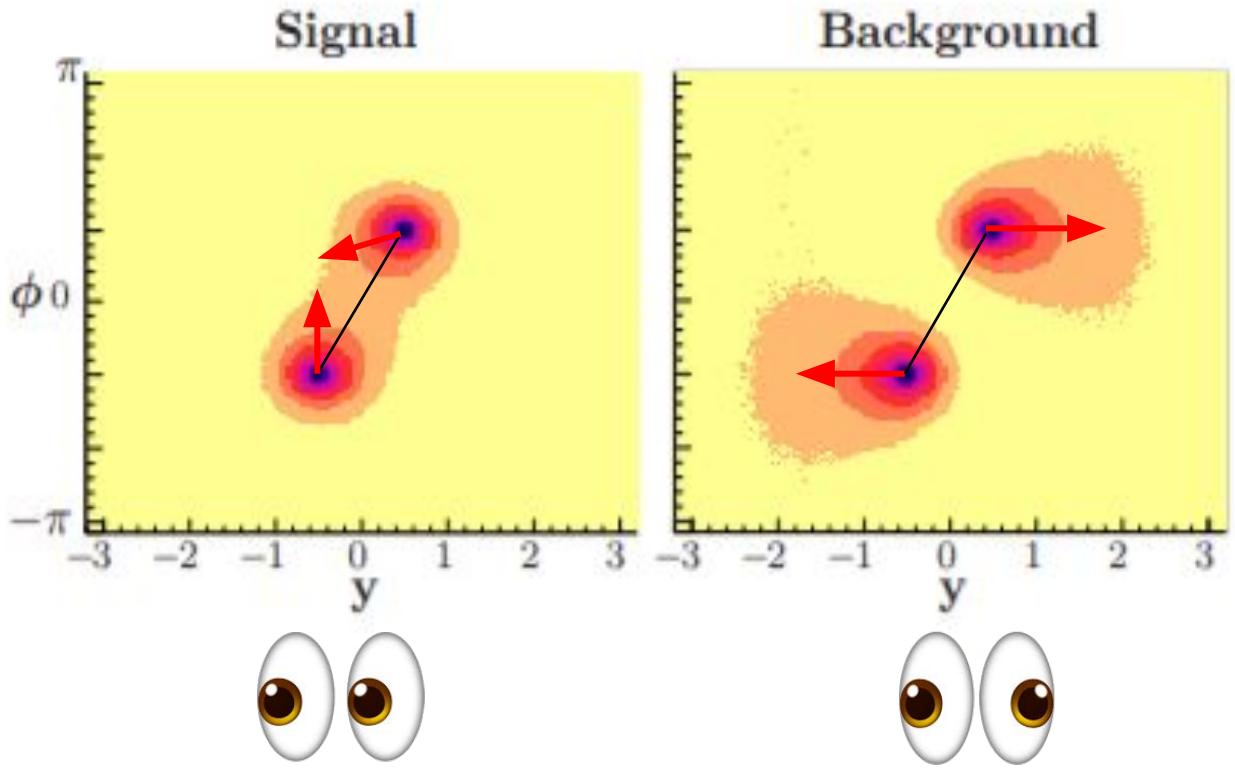


D_2

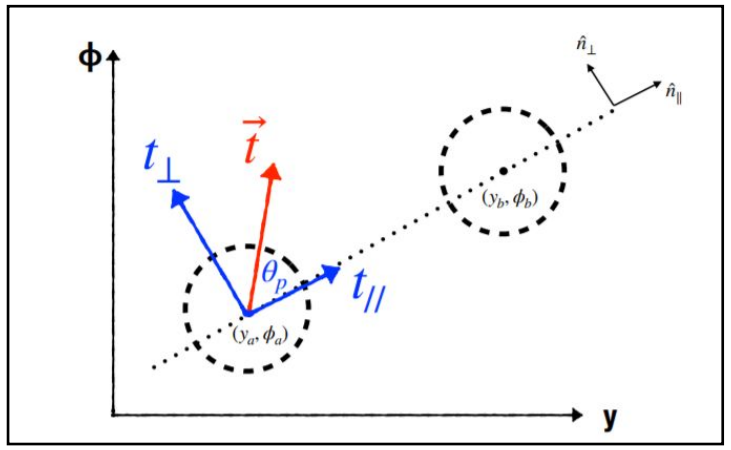
color sensitive variables:



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

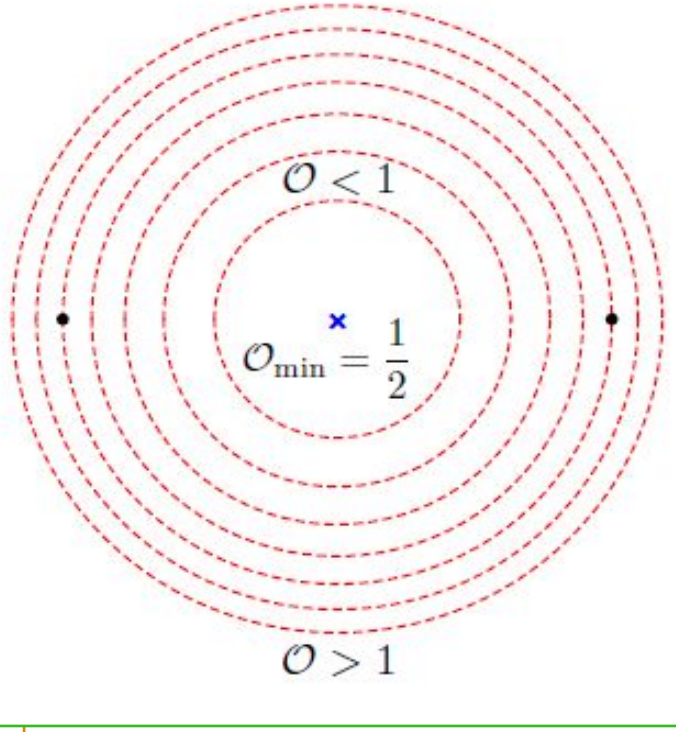


color sensitive variables

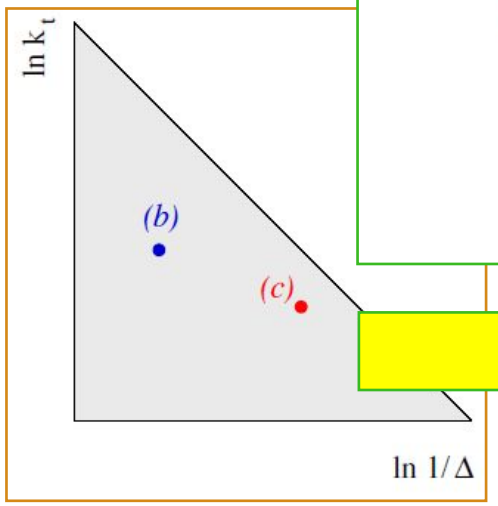


PULL VECTOR

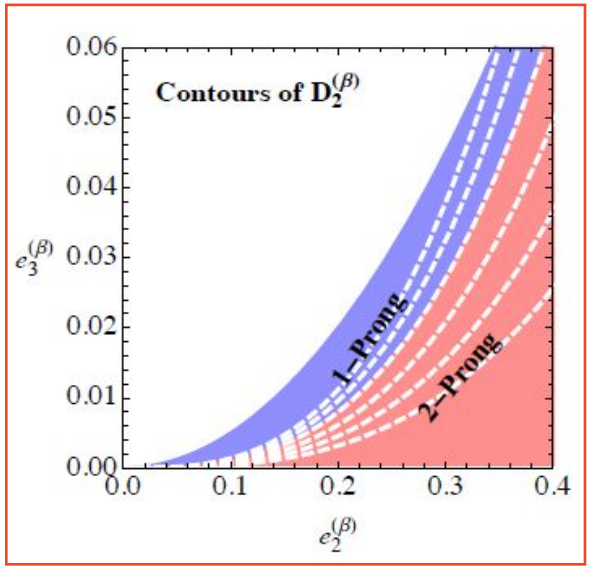
Varia



COLOR RING

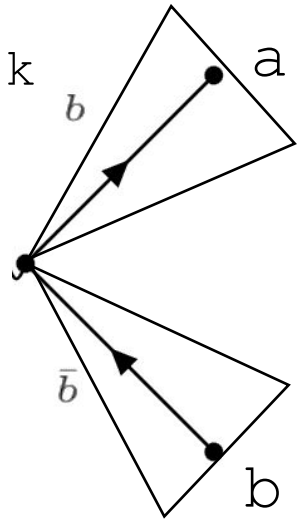


LUND PLANE

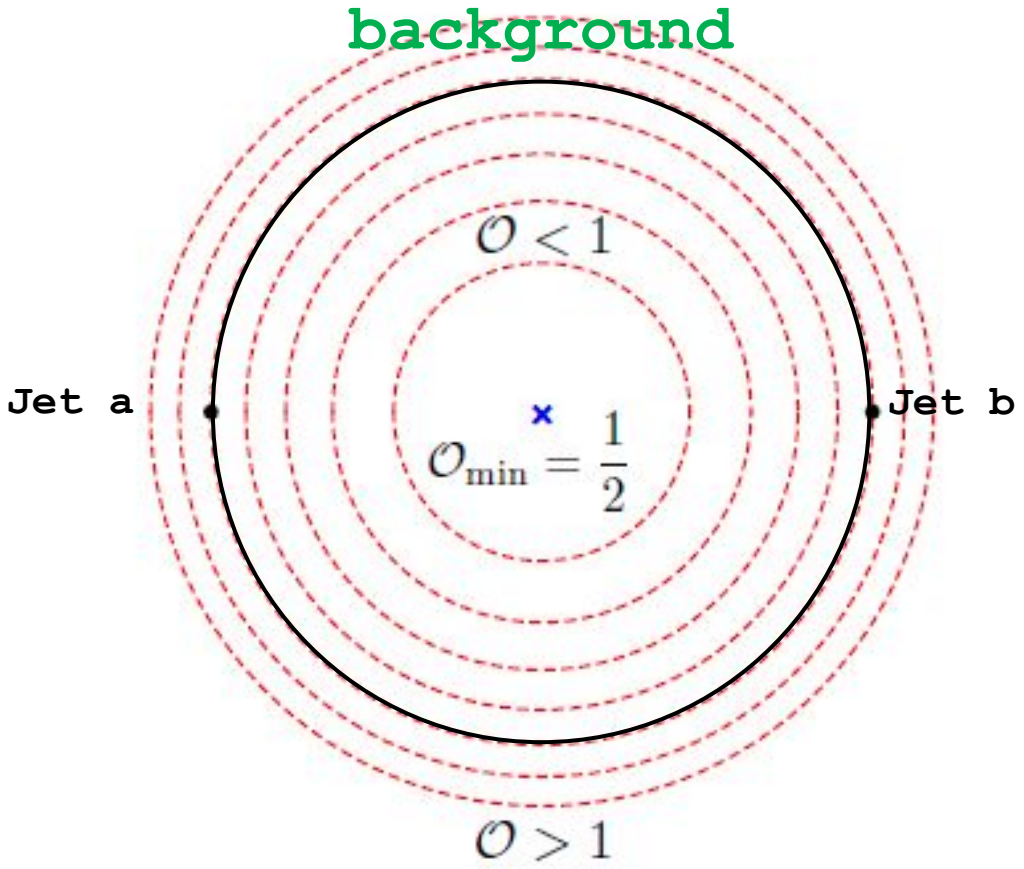


D_2

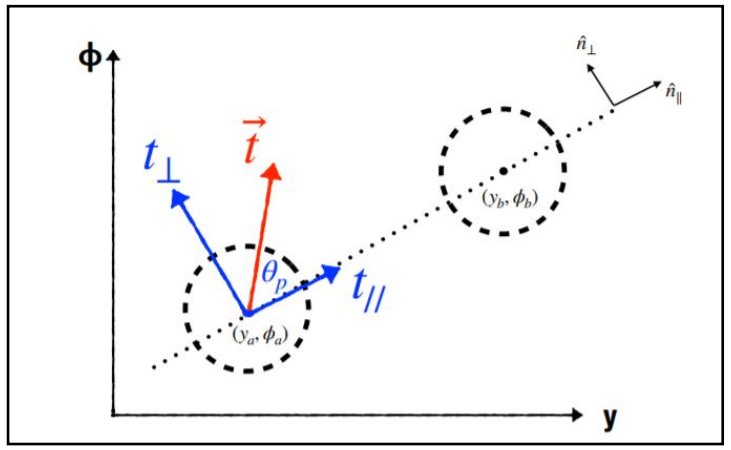
color sensitive variables: jet



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

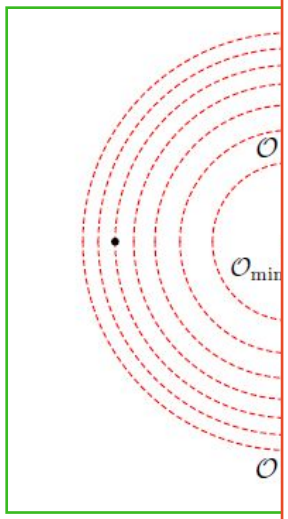


color sensitive variables

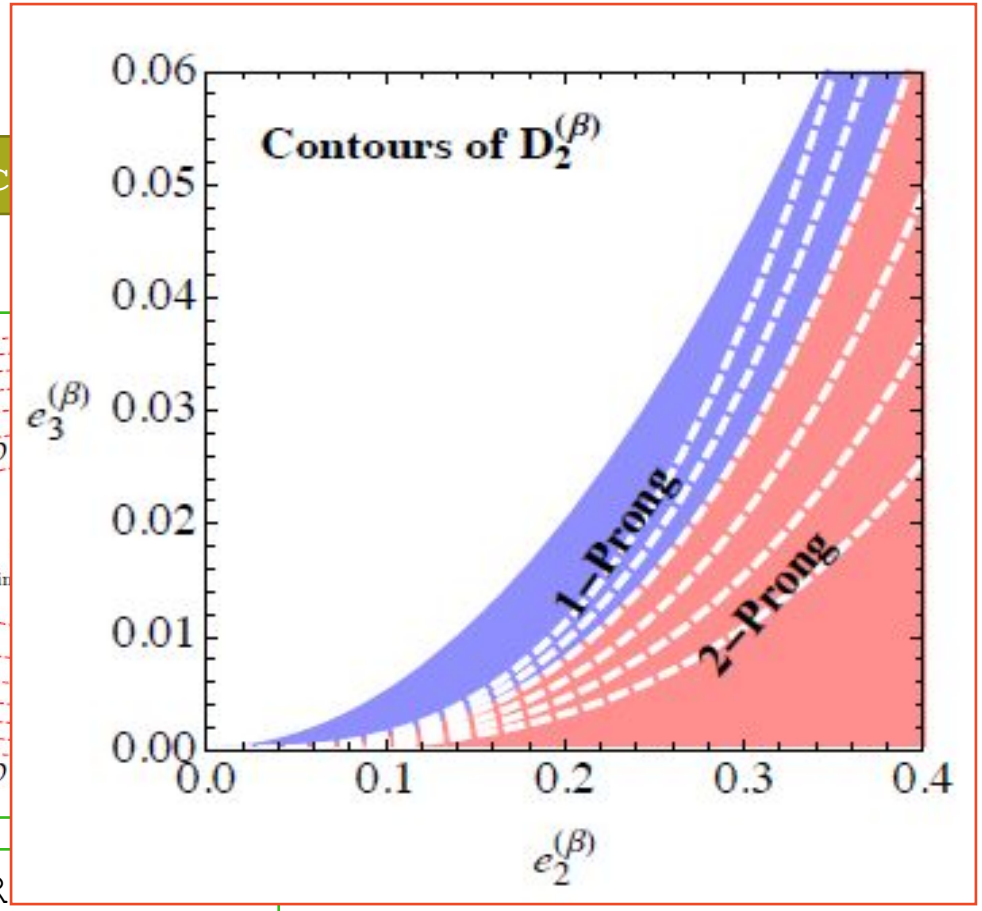


PULL VECTOR

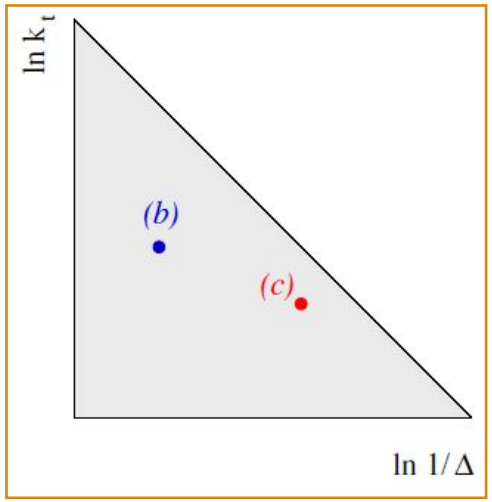
Variables sensitive to c



COLOR

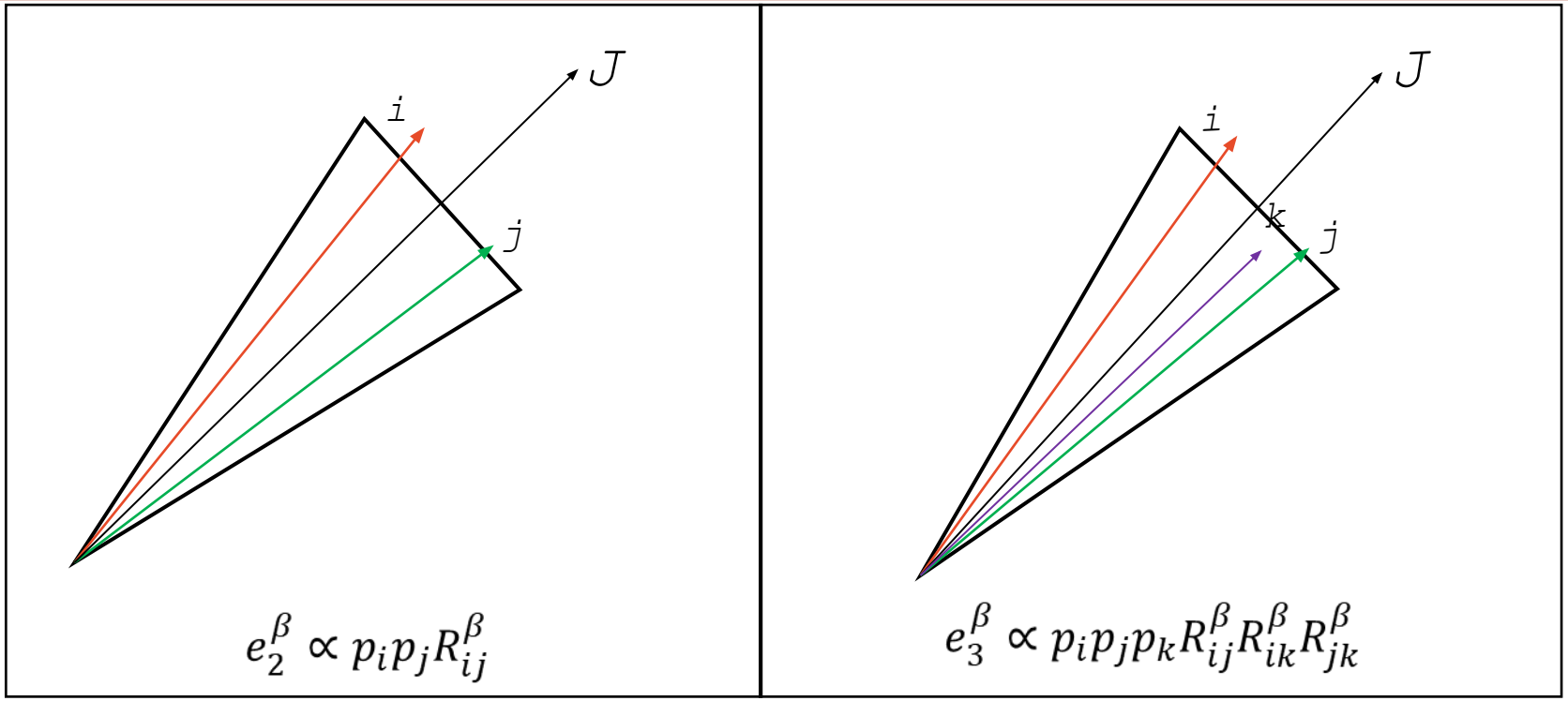


D_2



LUND PLANE

color sensitive variables: d2



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

SIGNAL

Small value of D_2

BACKGROUND

Large value of D_2

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