Study of b- and c- jets identification for Higgs coupling measurement at muon collider

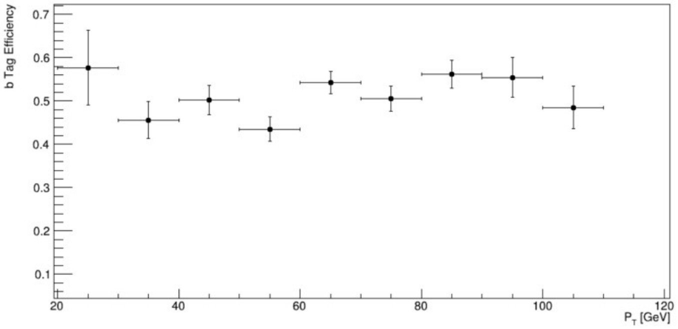
Giacomo Da Molin, Donatella Lucchesi, Lorenzo Sestini,
Luca Giambastiani, Laura Buonincontri

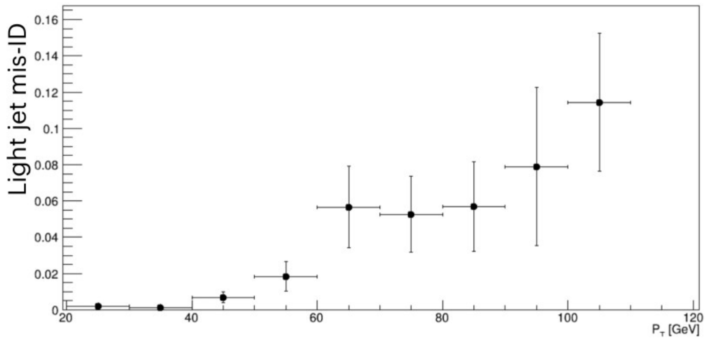
Objective

From Lorenzo Sestini's talk (at 1.5 TeV E cm)

Secondary vertex reconstruction

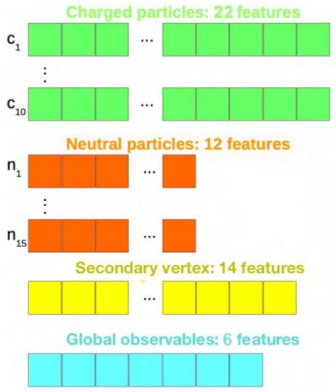
As of now, b-tagging is performed by looking for a SV inside the jet cone.





Improve this by using DNN to perform both b- and c-tagging, then exploit it for H→ bb and H→ cc.

LHCb one (from Padova INFN LHCb group) as starting point, but tuning required.



The starting DNN uses 420 variables to classify hadronic jets according to their flavour: b vs c vs q. In our application we will need to reduce the number of variables and select only the most significant.

Steps

- Preliminary study on tagging efficiencies and importance of variables to feed to the DNN
- Finish up the DNN, measure eff. and mistag
- Study b and c jets separation
- Extraction of $H \rightarrow bb$ and $H \rightarrow cc$ yields at 3 TeV.

Preliminary studies

DISCLAIMER:
Everything in this presentation is a work in progress

Datset used:

- 1000 $H \rightarrow b\bar{b}$ at 1.5 TeV (changing several of the vertexing parameters and the kT radius of the jet)

- 1000 $H \rightarrow bb$ at 3 TeV; for the mistag also studied 200 $\mu\mu \rightarrow ll$ at 3 TeV

- 1000 $H \rightarrow bb$ at 3 TeV but with BIB*!
- 600 $\mu\mu \rightarrow ll$, 1600 $\mu\mu \rightarrow cc$, 1500 $\mu\mu \rightarrow bb$ with BIB*!

* = BIB at 1.5 TeV, for now only few events

Preliminary studies

- Secondary Vertex tagging efficiencies and variables
- Invariant Mass Plots: see how well the Higgs is reconstructed
- DNN variables Plots (only at 3TeV with BIB)

Tagging and Truth Matching

Tagging is performed by requiring at least a SV in a jet cone.

Such control is performed by requiring ΔR between the jet axis and the SV position (from the PV) to be smaller than the cone radial dimension (usually 0.5).

SV built with “BuildUpVertexCollectionTuple” processor. In these presentation only plots with its default parameters (see Backup), but some other configurations were tried.

Tagging and Truth Matching

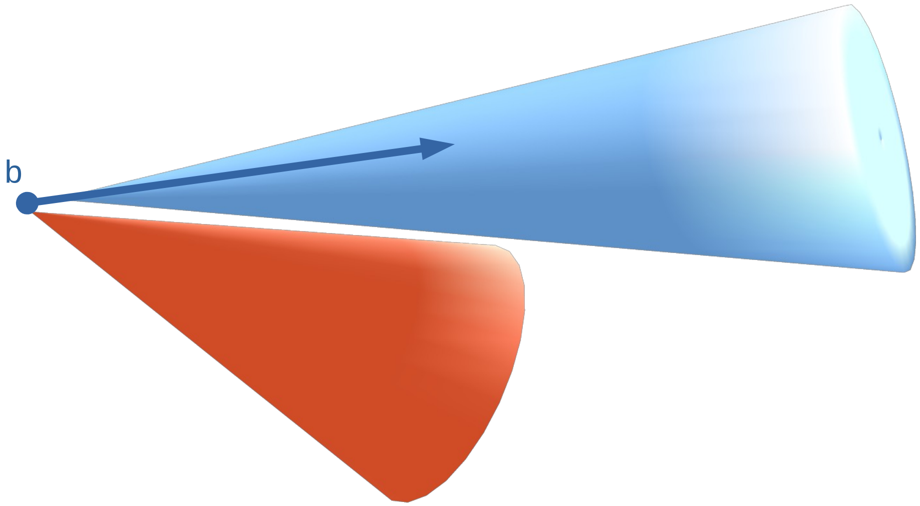
Lots of jets not coming from b even in $\rightarrow$bb samples due to gluonic jets or fake jets from BIB:

To study efficiencies we need Truth Matching:

Use MC b-quark momenta direction to distinguish b-jets (blue in figure) from other kind of jets (ex. Red one coming from a gluon).
A jet is considered truth-matched to a b if the b momenta direction lies inside the jet, i.e. it has

$$\Delta R < 0.5$$

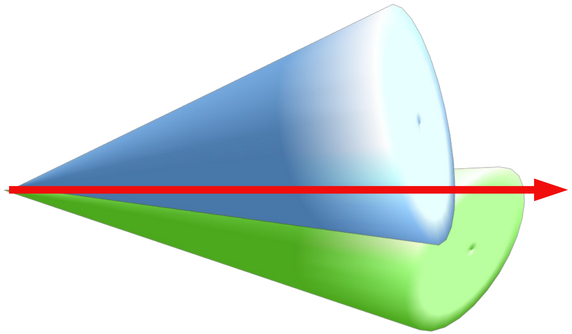
between jet axis and b momenta



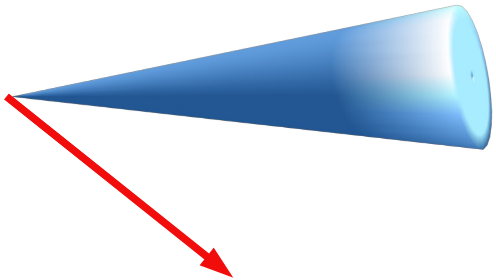
Tagging and Truth Matching

Pitfalls:

- Close jets may both contain b direction:

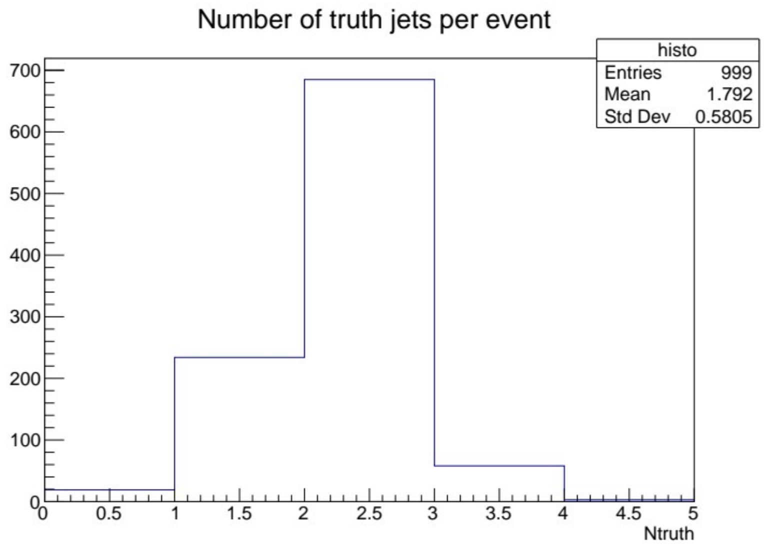


- Momenta of b may not be inside any jet:
  change direction and hadronize after gluon emission, badly reconstructed jet (or missed)

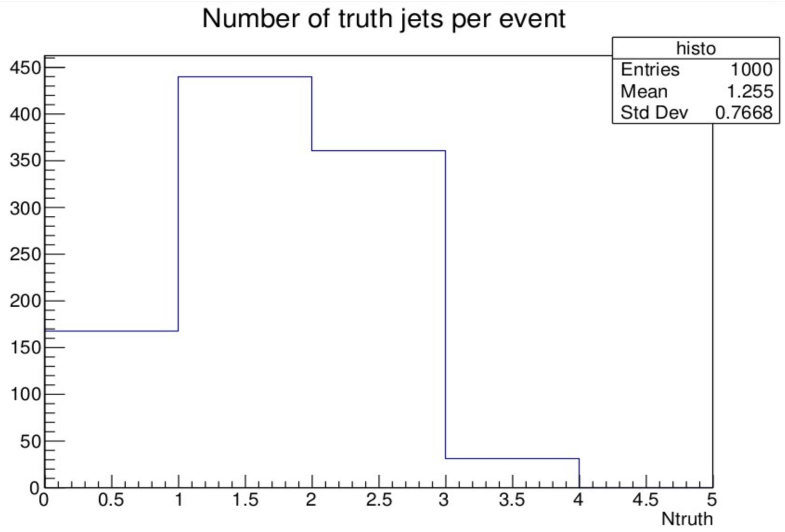


Tagging and Truth Matching

H→bb 3 Tev, no BIB



H→bb 3 Tev, with BIB



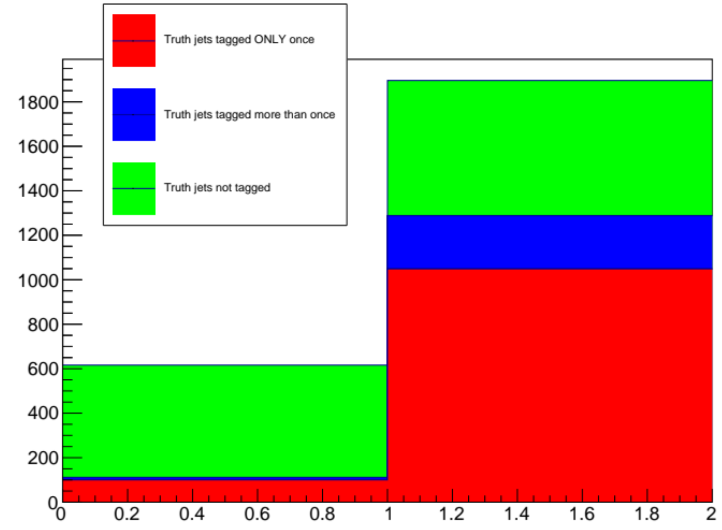
The BIB makes jet reconstruction much more difficult: less jets coming from the products of μμ are reconstructed, and also they are reconstructed with less precision. The bad reconstruction of the jet means it's less likely to match with the b-quark momenta direction, as can be seen in the histograms.

NOTATION: from now on we're gonna call the jets truth matched with a b quark simply "truth jets".

$$Efficiency = \frac{truth-and-tagged}{truth}$$

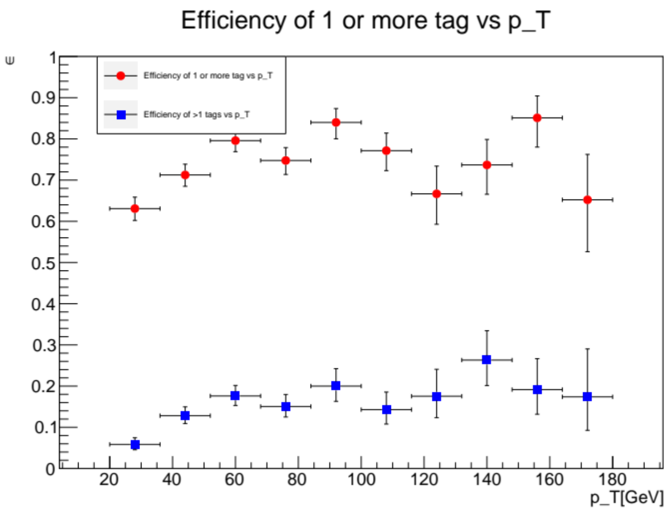
$$Mistag = \frac{NOTtruth-and-tagged}{NOTtruth}$$

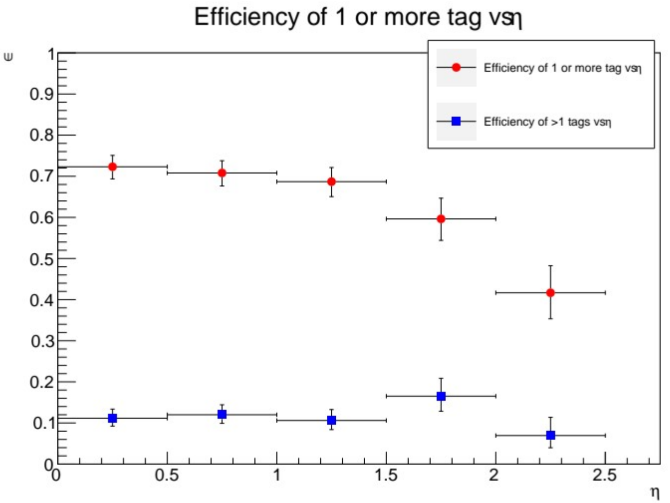
Efficiencies: H→bb 1.5 TeV, no BIB



Efficiency with ONLY one tag (1 SV in the jet): 0.553±0.011
Efficiency with more tags (2 or more SV in the jet): 0.126±0.008
Efficiency with AT LEAST one tag: 0.679±0.011

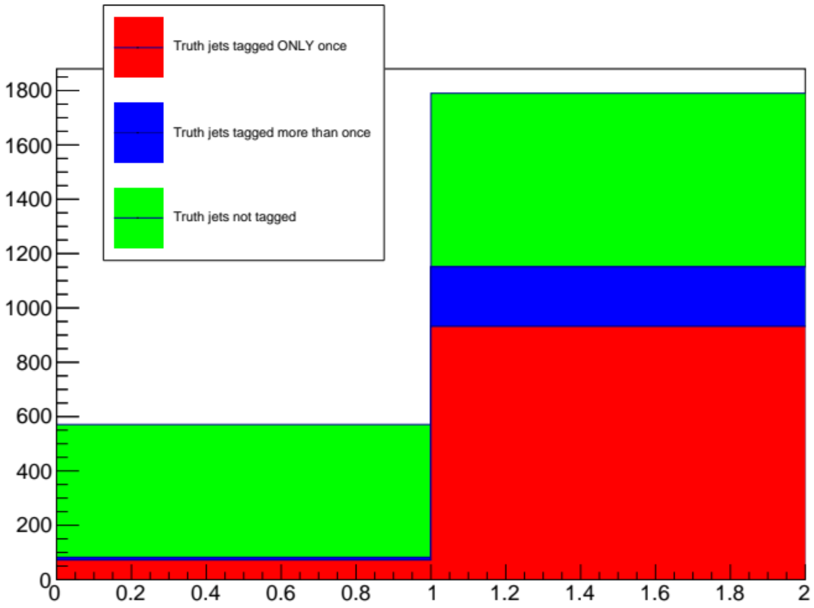
The mistag with ONLY one tag: 0.0527±0.0088 and with more tags: 0.0058±0.0030





Efficiencies vs other variables were plotted.
This is with default configuration for vertices. Others were studied

Efficiencies H→bb 3 TeV, no BIB



AT 1.5 TeV
Efficiency with ONLY one tag: 0.553±0.011
Efficiency with more tags: 0.126±0.008
Efficiency with AT LEAST one tag: 0.679±0.011

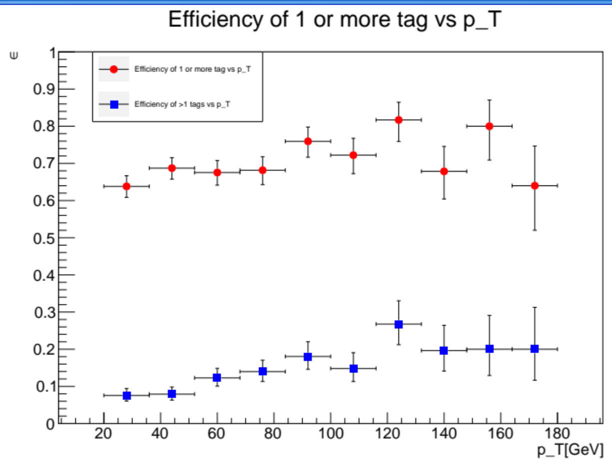
The mistag with ONLY one tag: 0.053±0.009 , with more tags: 0.006±0.003

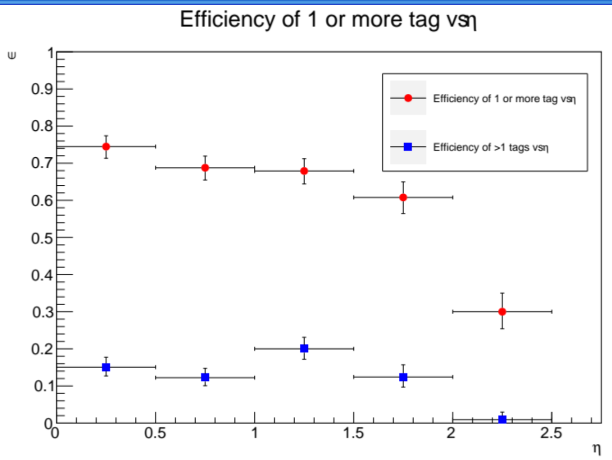
AT 3 TeV
Efficiency with ONLY one tag: 0.521±0.012
Efficieny with more tags: 0.122±0.008
While efficiency with AT LEAST one tag: 0.643±0.011

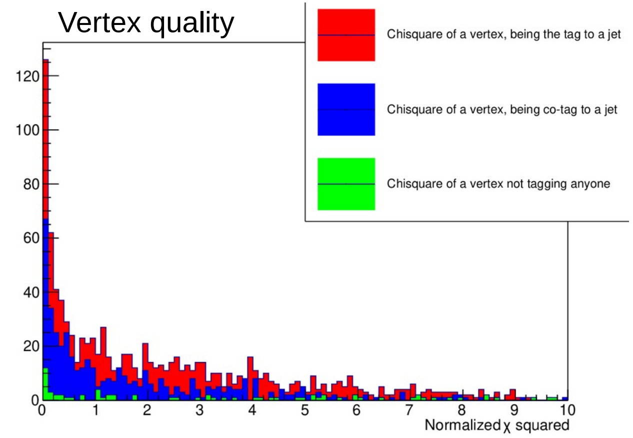
The mistag with ONLY one tag: 0.040±0.008, with more tags: 0.006±0.003

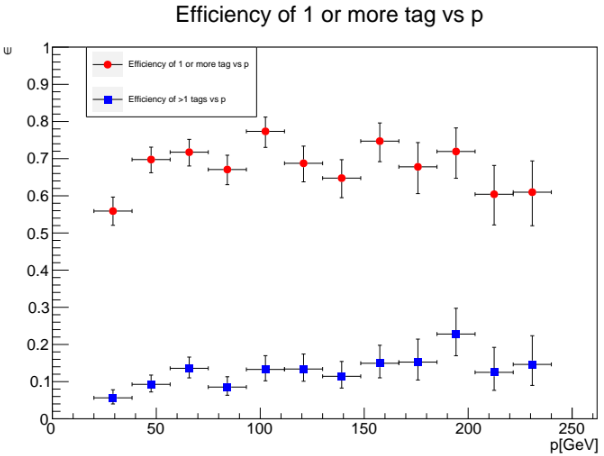
A small drop in both efficiency and mistag

Efficiencies H→bb 3 TeV, no BIB



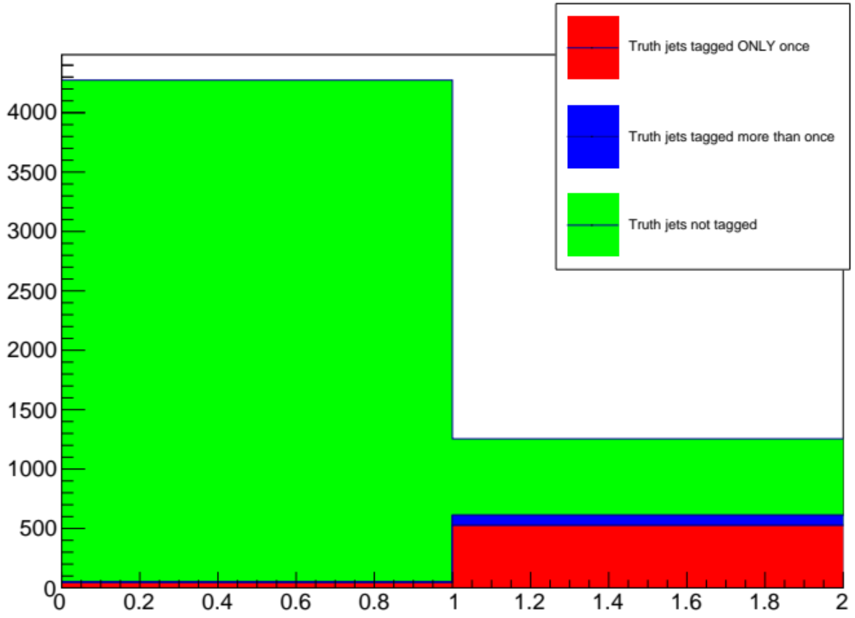






A study on μμ→ll at 3 TeV without BIB has been performed to better determine mistag, but low statistics. Their results appear compatible with the mistag in the previous slides

Efficiencies H → bb 3TeV, with BIB



AT 3 TeV, NO BIB
Efficiency with ONLY one tag: 0.521±0.012
Efficieny with more tags: 0.122±0.008
While efficiency with AT LEAST one tag: 0.643±0.011

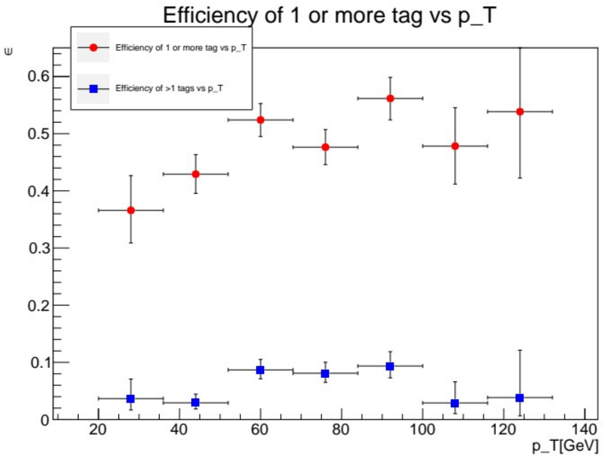
The mistag with ONLY one tag: 0.040±0.008 , with more tags: 0.006±0.003

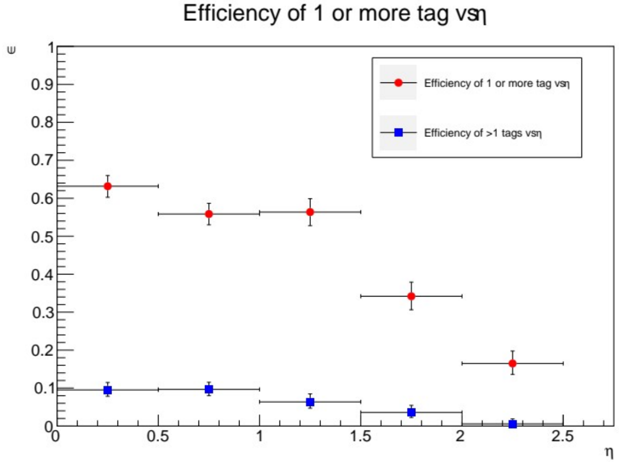
NOW, AT 3 TeV, with BIB
Efficiency with ONLY one tag: 0.419±0.014
Efficiency with more tags: 0.068±0.007
While efficiency with AT LEAST one tag: 0.487±0.014

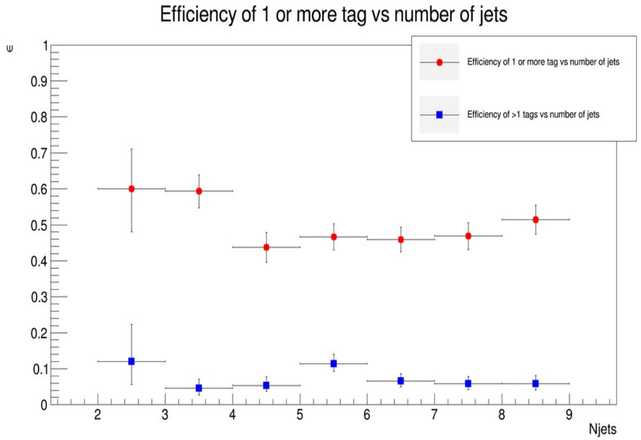
The mistag with ONLY one tag: 0.036±0.003 , with more tags: 0.004±0.001

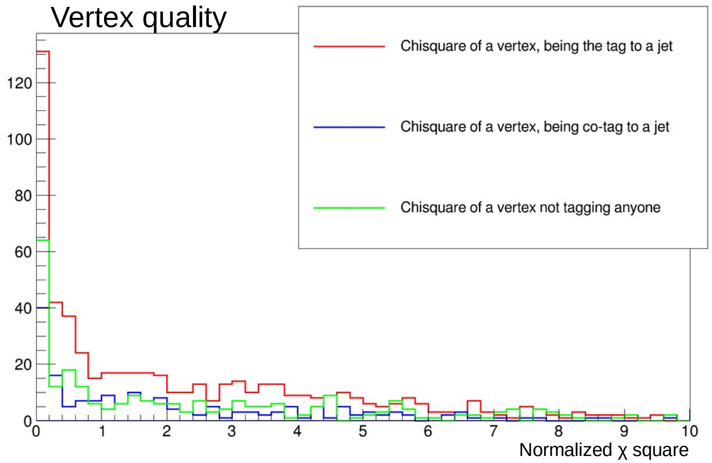
Considerable drop in efficiency! But no increase in the mistag

Efficiencies H → bb 3TeV, with BIB







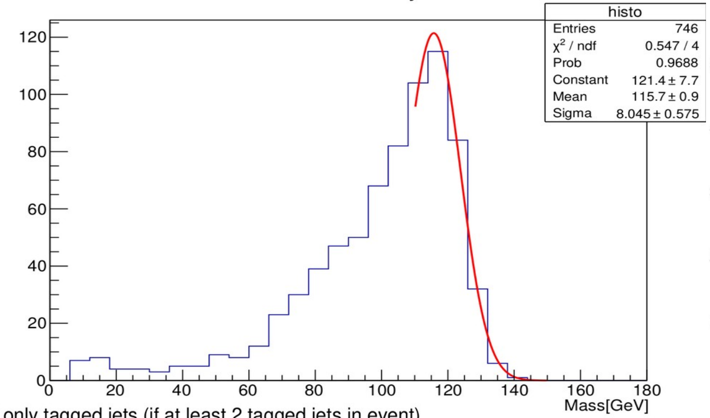


Reconstructing Higgs invariant masses H → bb 3 TeV

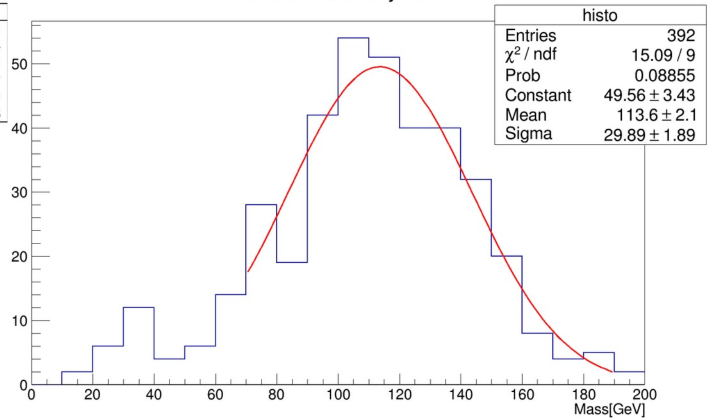
First column No BIB
Second with BIB

Truth →

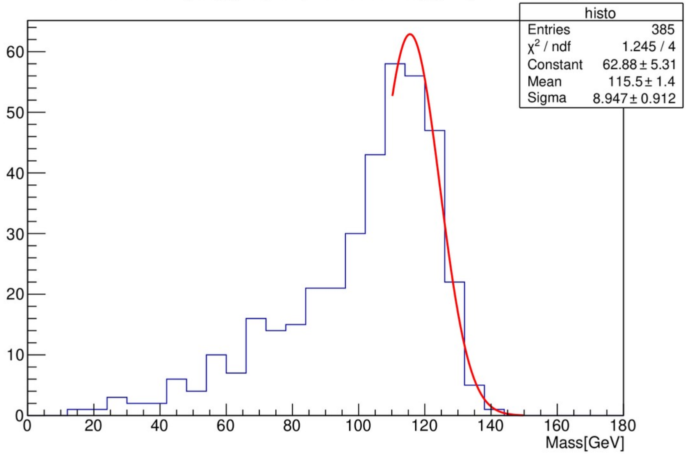
Mass of true b jets



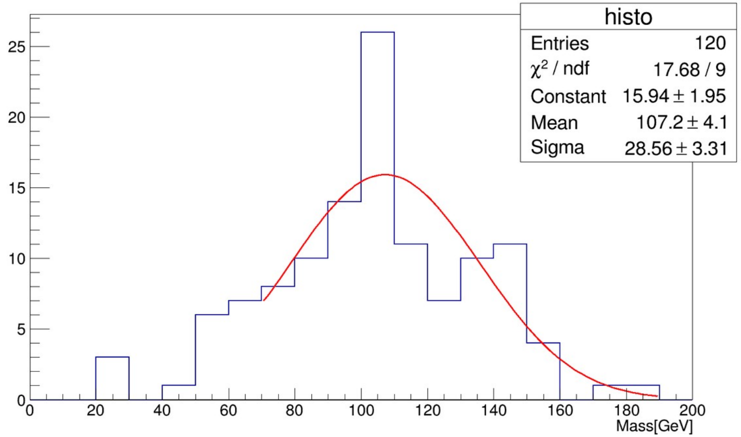
Mass of true b jets



Mass of only tagged jets (if at least 2 tagged jets in event)



Mass of only tagged jets (if at least 2 tagged jets in event)



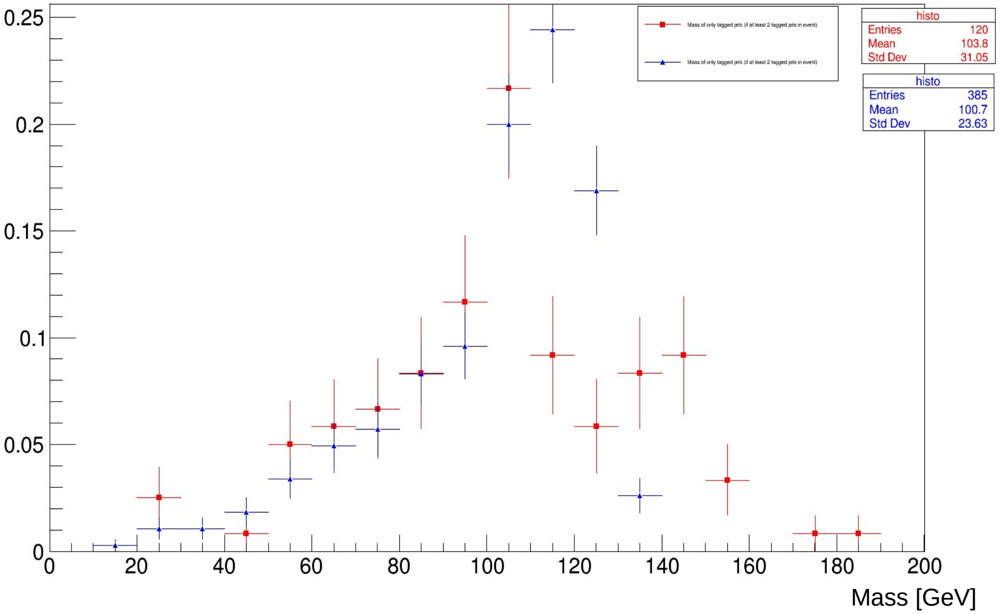
← Tag

Reconstructing Higgs invariant masses H → bb 3 TeV

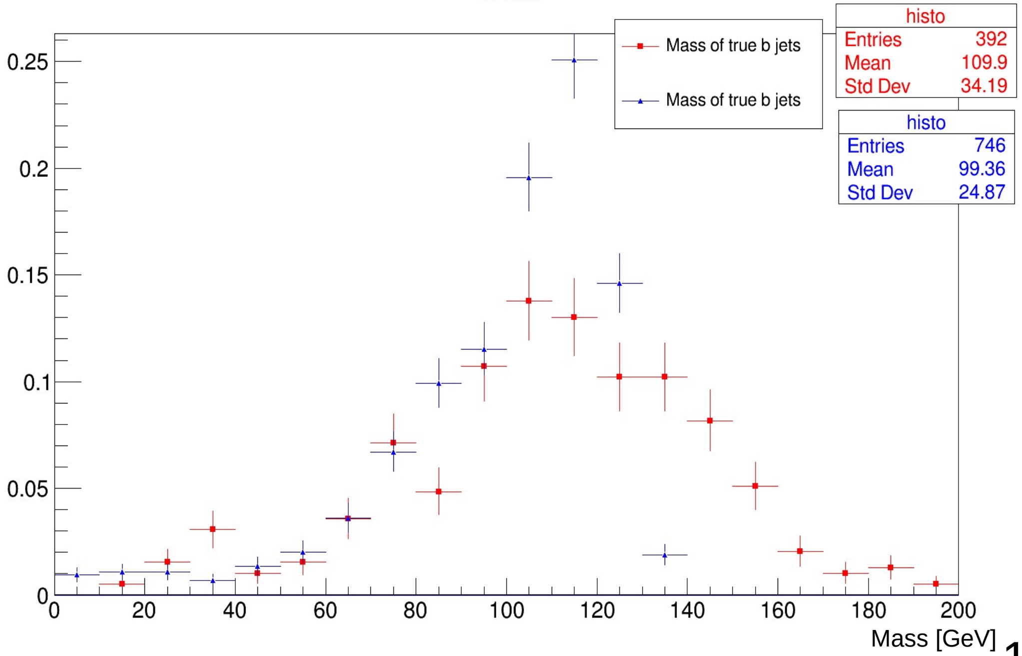
Red points: BIB

Blue points:No BIB

Tagged



Truth matched



DNN variables:

Implementation in LCFI environment by Luca Giambastiani

3 classes DNN: b-jets, c-jets, not heavy (light jets and BIB fake jets)

6 Jet variables:

Pt,E,Pz,M, number of charged particles (up to 7), number of neutral particles (up to 13)

9 Secondary vertex variables:

Transv flight distance, Z, lifetime, number of tracks, mass, corrected mass, ΔR, PT_{SV}/PT_{jet} absolute value of the sum of charges of the tracks

Note: right now only 1 SV for jet is saved: if there are more in same jet the one with biggest PT is chosen.

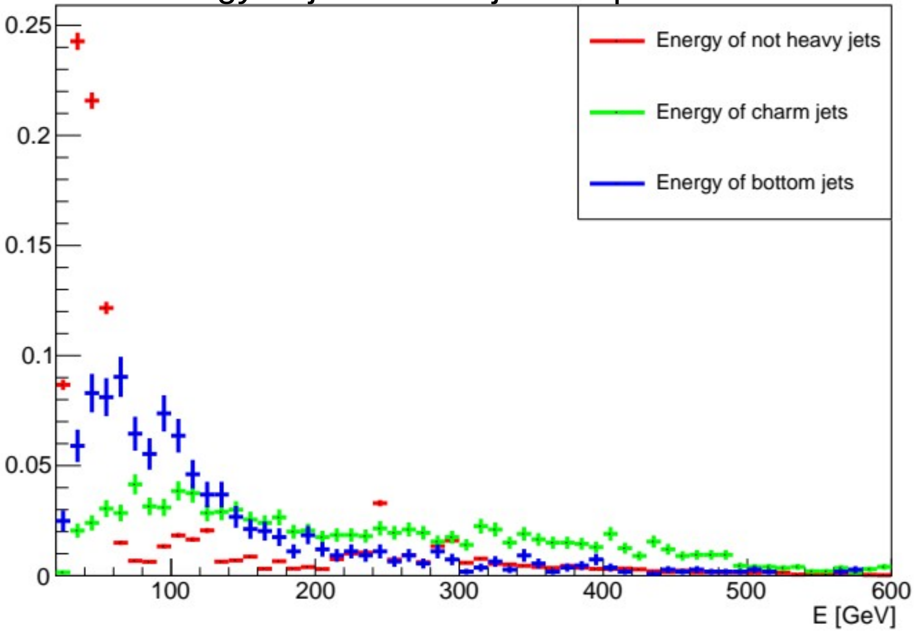
Still to implement in my plots:
Single charged particle variables: E,Pt,Pz,Q/P,D0,Ω,Z0,tan(λ)
Single neutral particle variables: E,Pt,Pz

DNN variables: Jet Energy

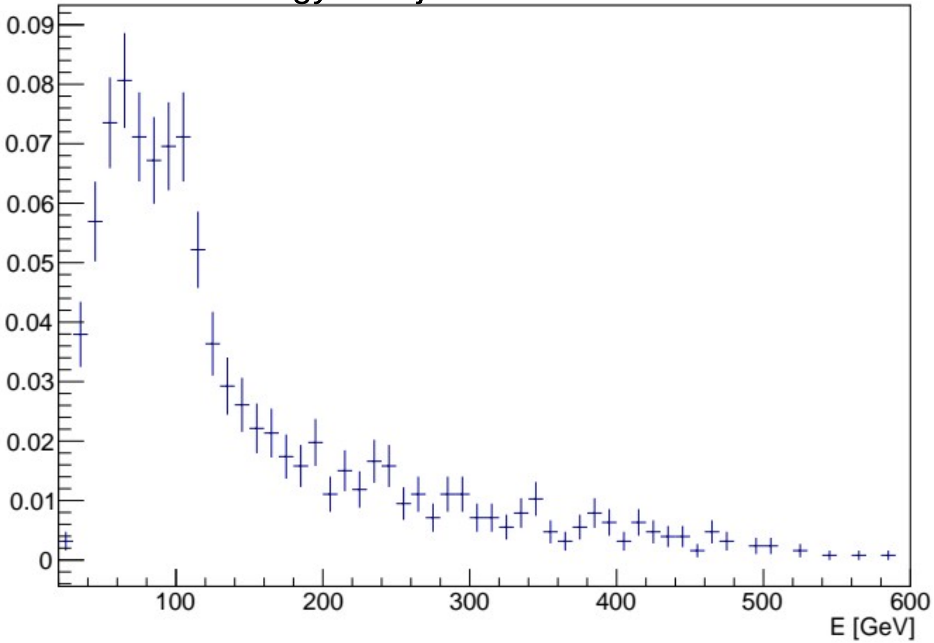
Working with 1085 b-jets, 2001 c-jets and 15509 others.
Working with 484 Sv in b-jets, 576 Sv in c-jets and 732 Sv in others.
In the Higgs sector from 5064 jets, a total of 1265 b-jets were found, of which 633 contained Sv.

In this and the following slides, we're gonna now show the normalized distributions of some of these variables.

Energy of jets from dijet samples

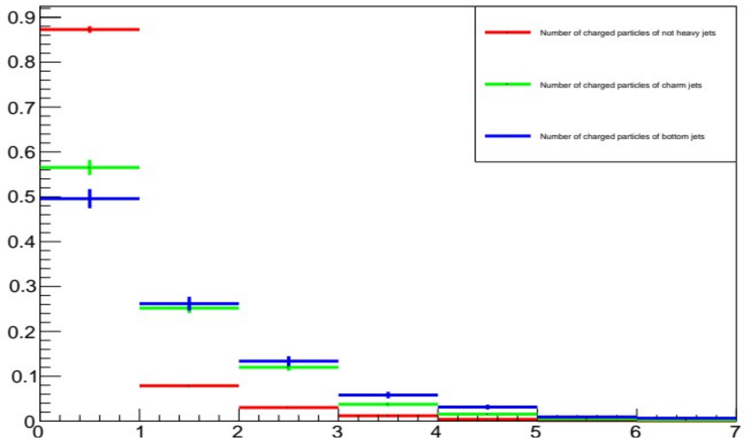


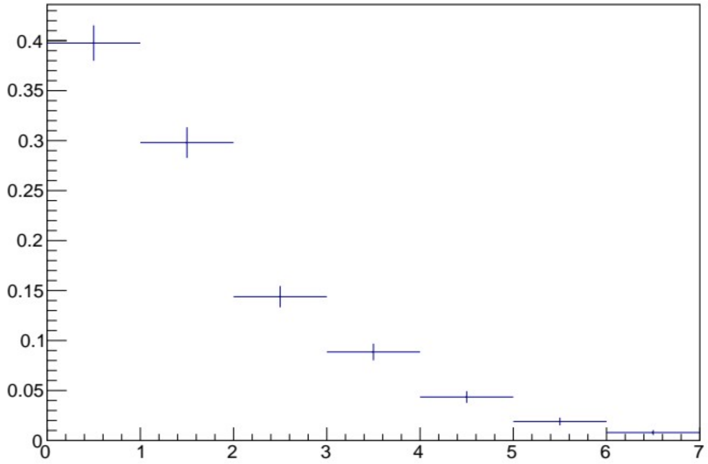
Energy of b-jets from H→ bb



DNN variables: jet number of particles

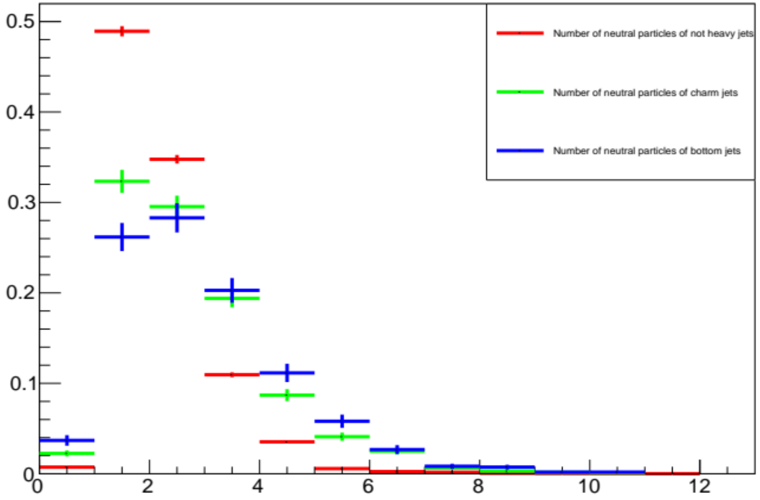
Number of Charged Particles in dijets

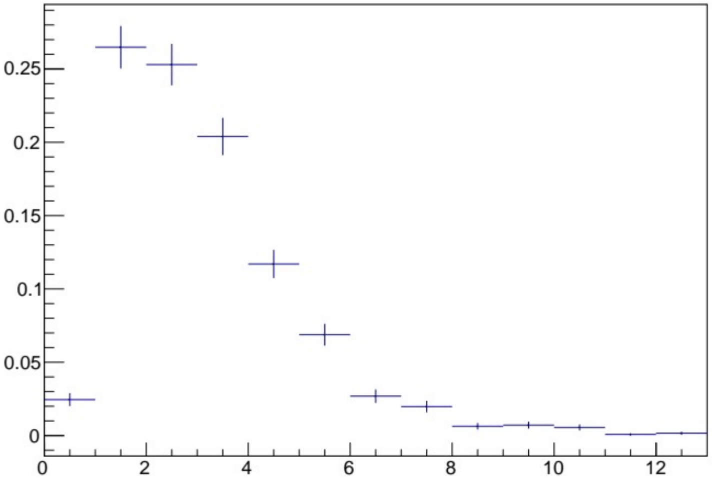




Number of Charged Particles in b-jets from H $\rightarrow$ bb

Number of Neutral Particles in dijets

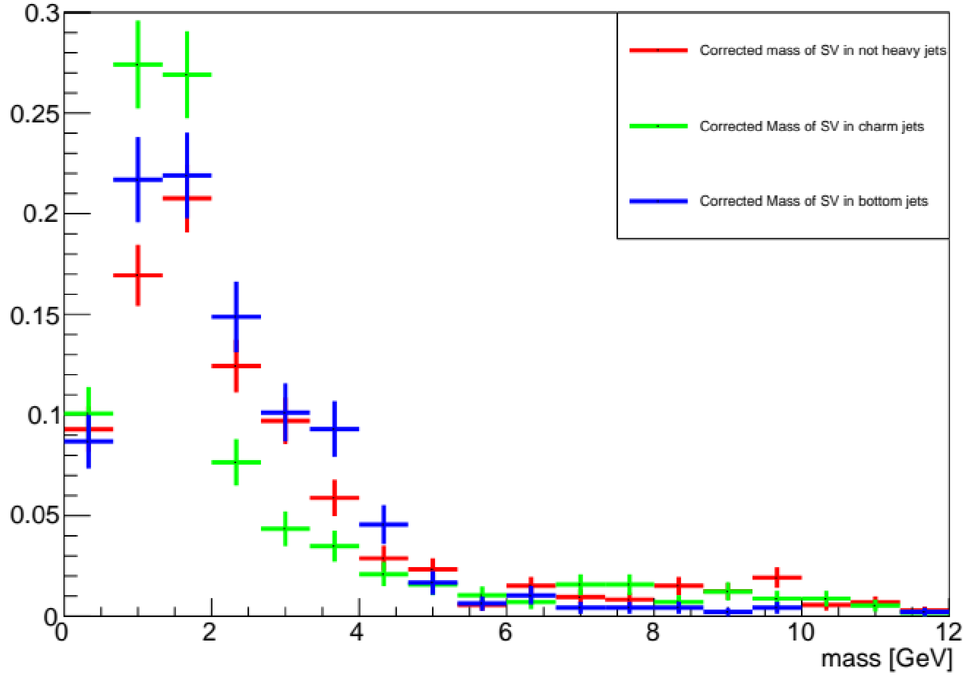




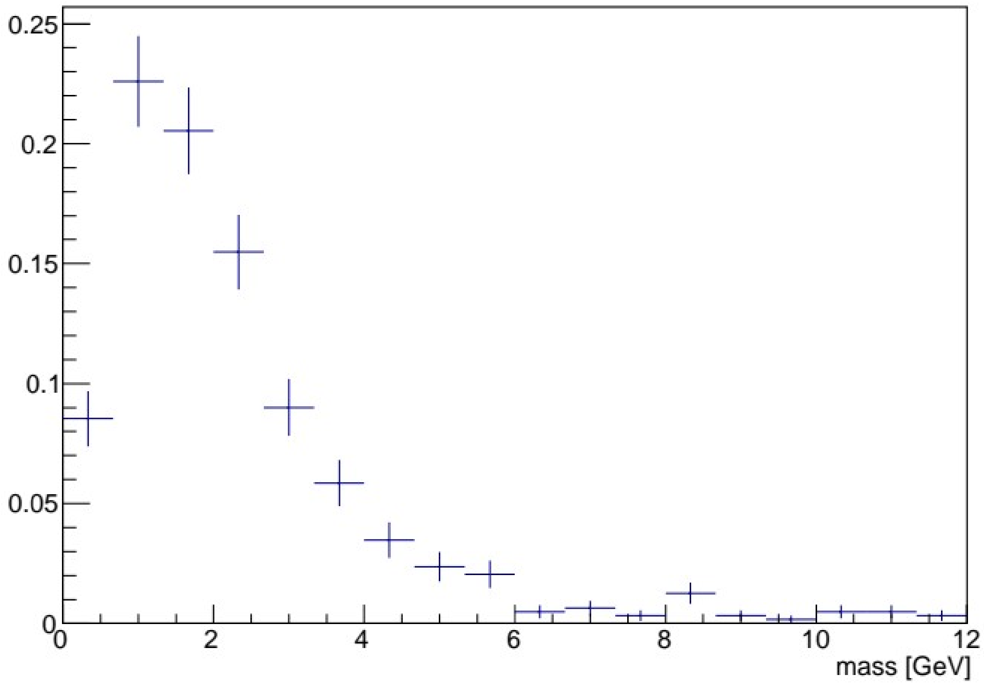
Number of Neutral Particles in b-jets from H $\rightarrow$ bb

DNN variables: SV corrected Mass

Corrected mass of jets from dijet samples

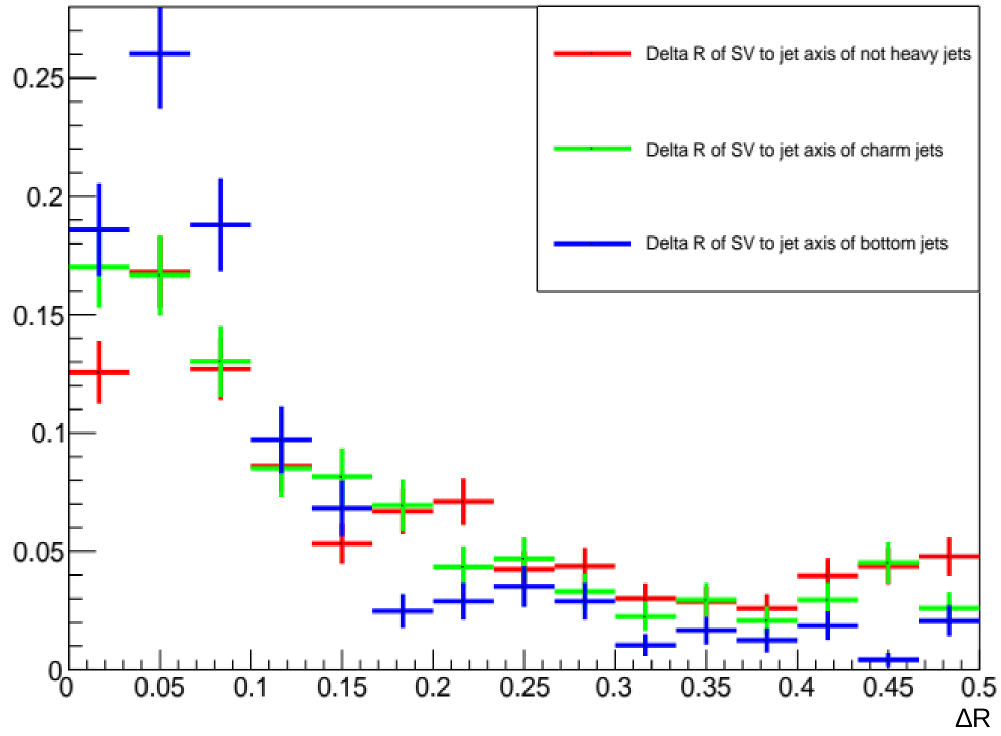


Corrected mass of b-jets from H → bb

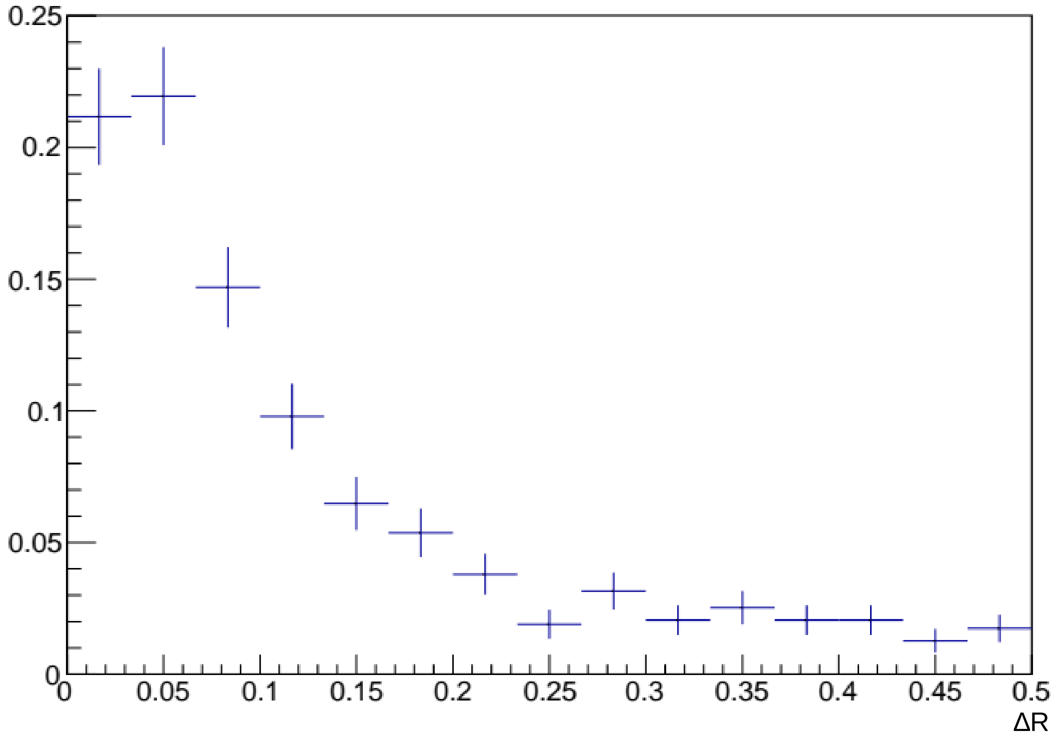


DNN variables: ΔR of SV with jet axis

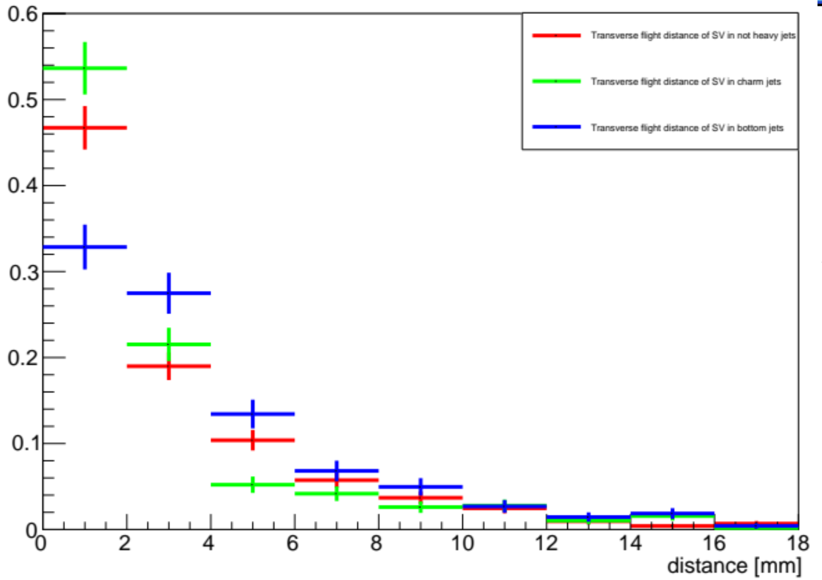
ΔR of SV in jets from dijet samples



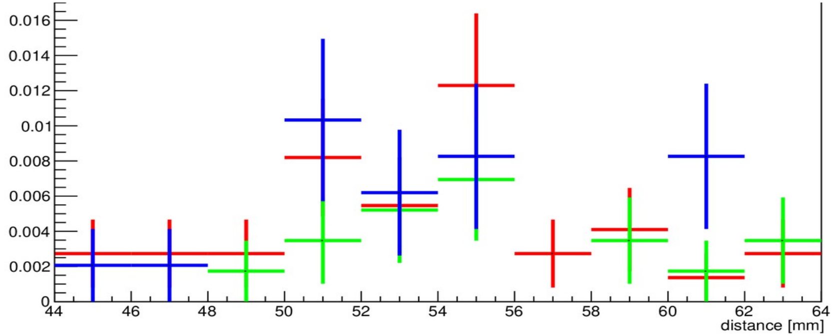
ΔR of SV in b-jets from H→bb



DNN variables: SV transverse flight distance

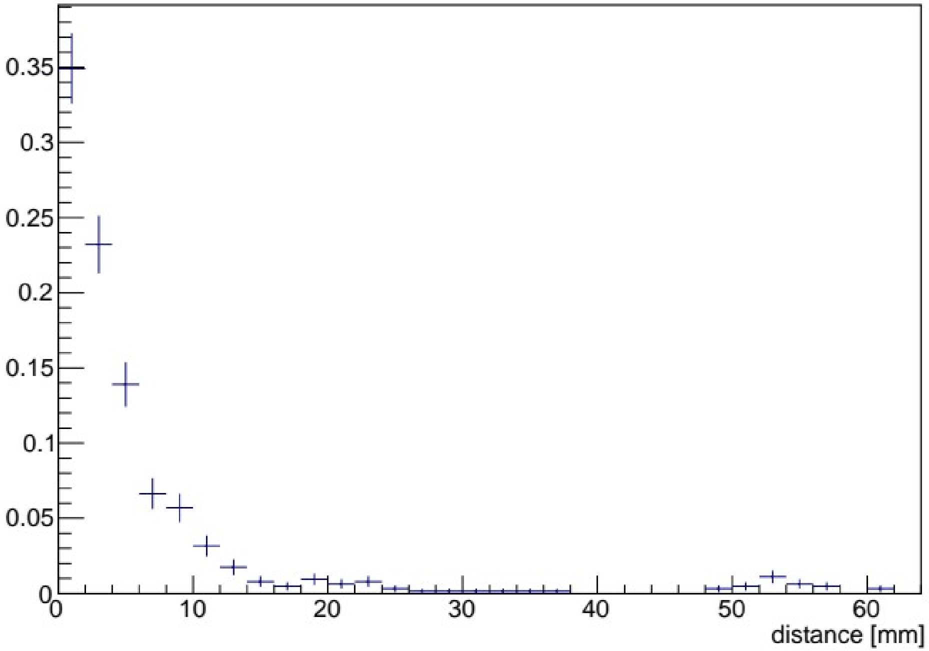


← Transverse flight distance of SV in dijets, zoomed at small distances



← Transverse flight distance of SV in dijets, zoomed at small small peak at high distances

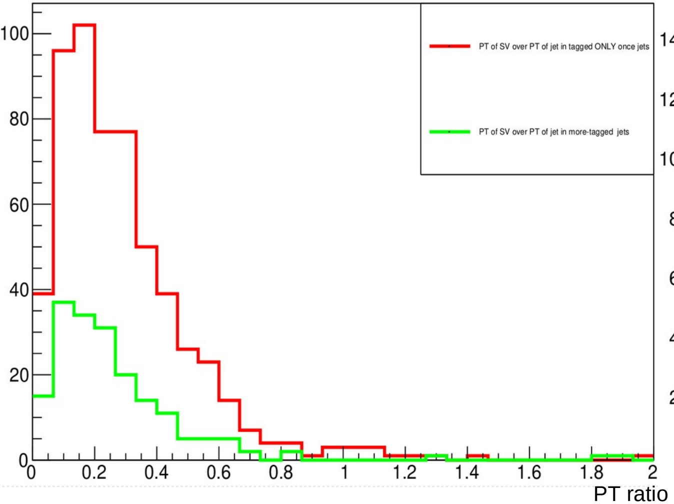
Transverse flight distance of SV in b-jets from H→bb



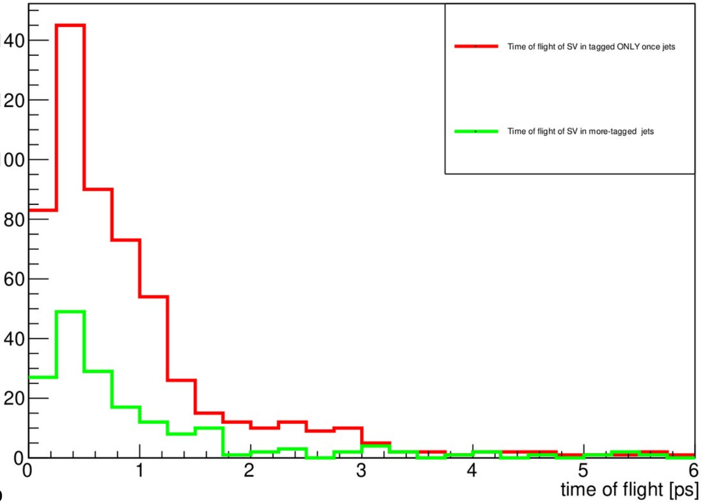
DNN variables for SV of Jets tagged by 1 SV vs tagged by more SV

For this test, all SVs in the same jet are saved.
Pretty much consistent distributions across the board.

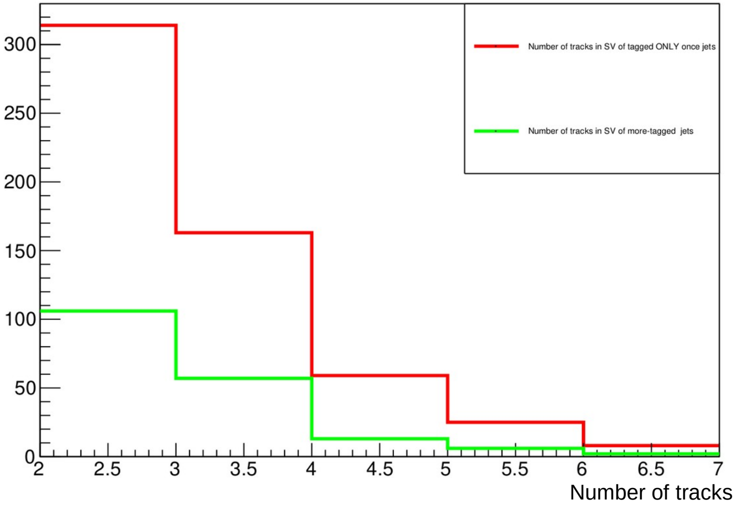
Pt of SV over PT of jet



Time of flight of SV



Number of tracks in SV



Next steps:

- **SV-Tagging efficiencies:** further investigation of optimal jet and SV configuration
- **Invariant Mass:** study of invariant mass distribution of Z and H samples, to see if Z resonance cover H peak
- **DNN:** study single particle variables; provide more heavy jet statistics

That’s all folks!

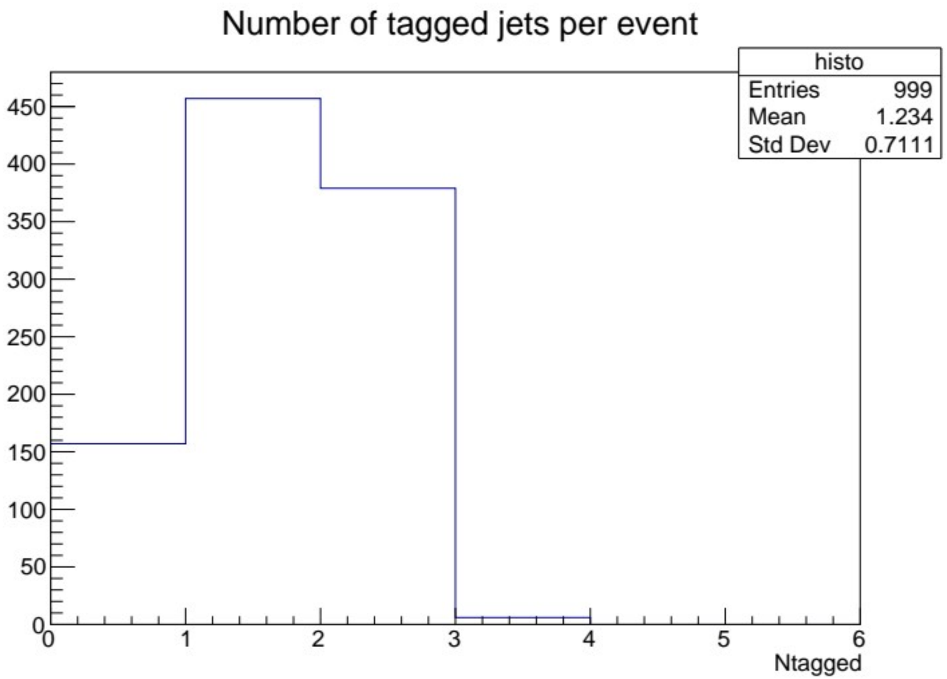
Thank you for your attention!

Or is it?

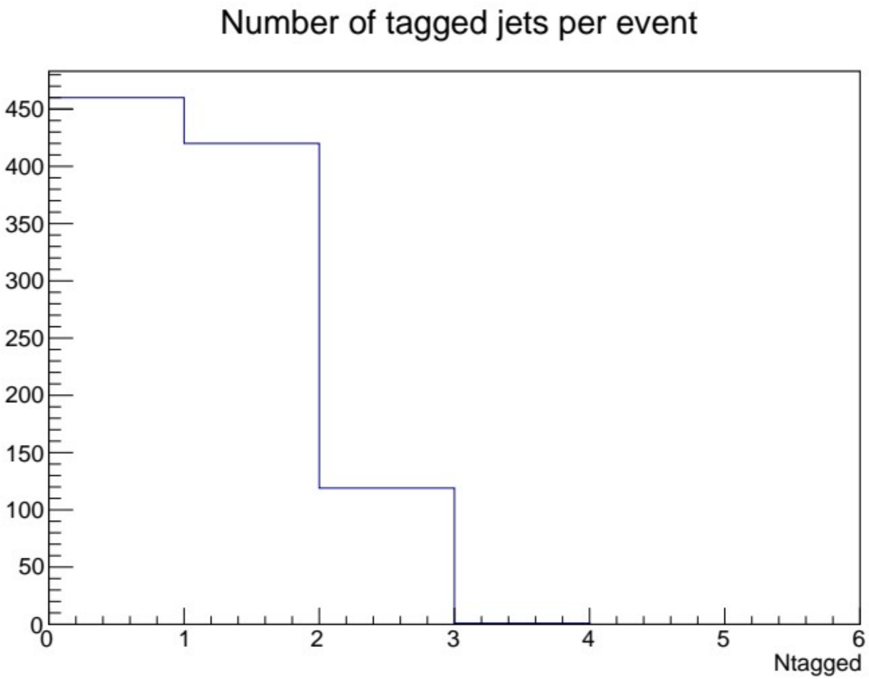
Backup

Ntagged: H → bb at 3 TeV

No BIB



with BIB



Default tracking settings:

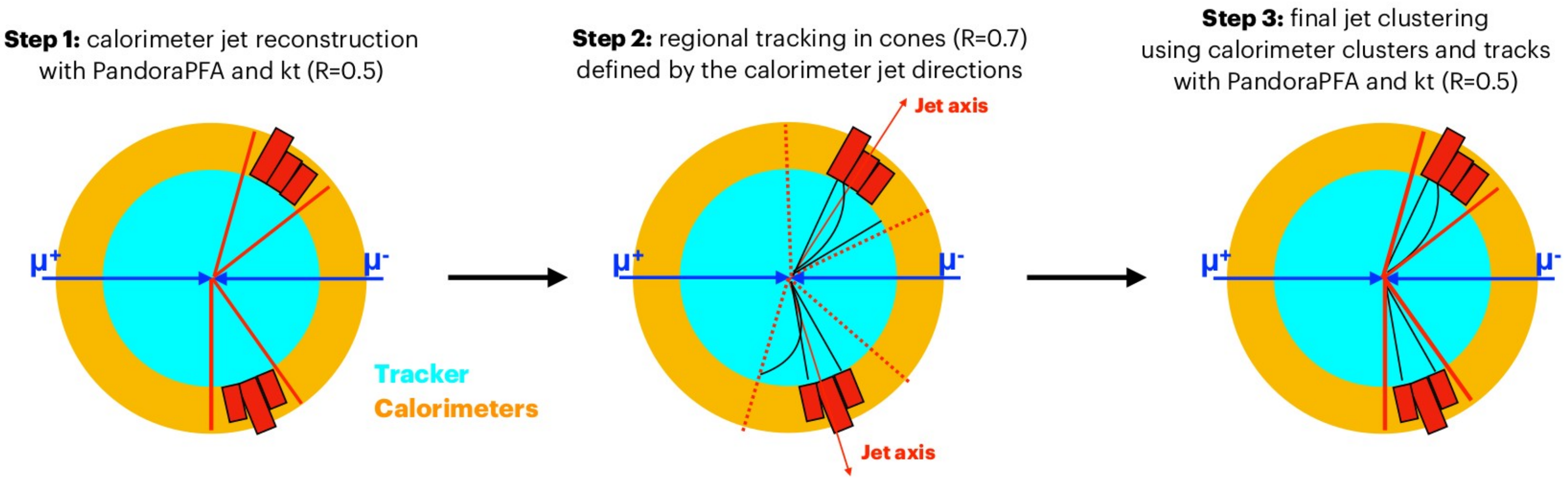
- <!-- parameters for secondary vertex finder -->
- <parameter name="BuildUpVertex.TrackMaxD0" type="double" value="5" />
- <parameter name="BuildUpVertex.TrackMaxZ0" type="double" value="5" />
- <parameter name="BuildUpVertex.TrackMinD0Z0Sig" type="double" value="2" />
- <parameter name="BuildUpVertex.TrackMinPt" type="double" value="0.8" />
- <parameter name="BuildUpVertex.TrackMaxD0Err" type="double" value="1" />
- <parameter name="BuildUpVertex.TrackMaxZ0Err" type="double" value="1" />
- <parameter name="BuildUpVertex.TrackMinVxdFtdHits" type="int" value="4" />
- <parameter name="BuildUpVertex.PrimaryChi2Threshold" type="double" value="10." />
- <parameter name="BuildUpVertex.SecondaryChi2Threshold" type="double" value="5." />
- <parameter name="BuildUpVertex.MassThreshold" type="double" value="10." />
- <parameter name="BuildUpVertex.MinDistFromIP" type="double" value="0.3" />
- <parameter name="BuildUpVertex.MaxChi2ForDistOrder" type="double" value="1.0" />
- <parameter name="BuildUpVertex.AssocIPTracks" type="int" value="1" />
- <parameter name="BuildUpVertex.AssocIPTracksMinDist" type="double" value="0." />
- <parameter name="BuildUpVertex.AssocIPTracksChi2RatioSecToPri" type="double" value="2.0" />
- <parameter name="BuildUpVertex.UseV0Selection" type="int" value="1" />
- <!-- AVF -->
- <parameter name="BuildUpVertex.UseAVF" type="boolean" value="true" />
- <parameter name="BuildUpVertex.AVFTemperature" type="double" value="1.0" />

Secondary vertex tracks requirements

Jet Reconstruction

Full jet reconstruction algorithm

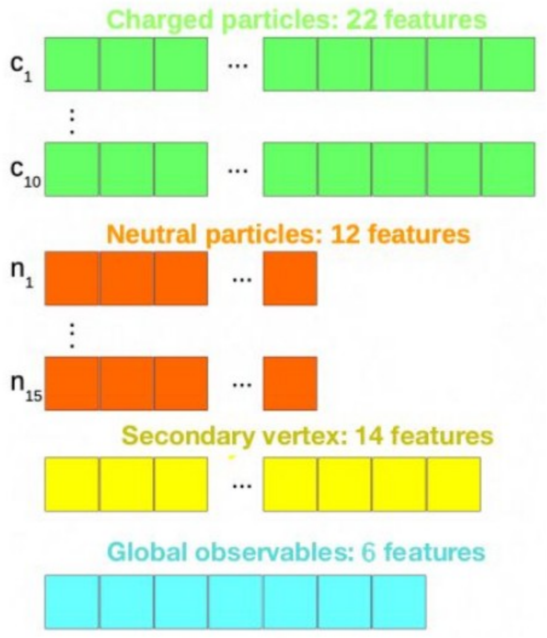
- **In order to reduce the tracking combinatorial problem, a regional tracking strategy is employed.**
- Motivated by the Physics, **it is also useful to save a lot of CPU time**.



DNN of Padova's INFN LHCb group

Rationale

- Purpose: **Classification of hadronic jets** according to their flavour
  - **b vs c vs q**
  - q includes light quarks (u, d, s) and gluons
- Used a Deep Neural Network (**DNN**). Input variables related to jet:
  - 22 variables x 10 charged particles
  - 12 variables x 15 neutral particles
  - 14 variables related to SV in the jet
  - 6 jet global variables
  - Static number of variables: if a particle or a vertex is missing its variables are set to 0



Possibly more plots