Flavour tagging performance of the New CLIC Detector

Ignacio Garcia

CLIC Detector and Physics collaboration meeting

CERN- 29th-30th August 2017

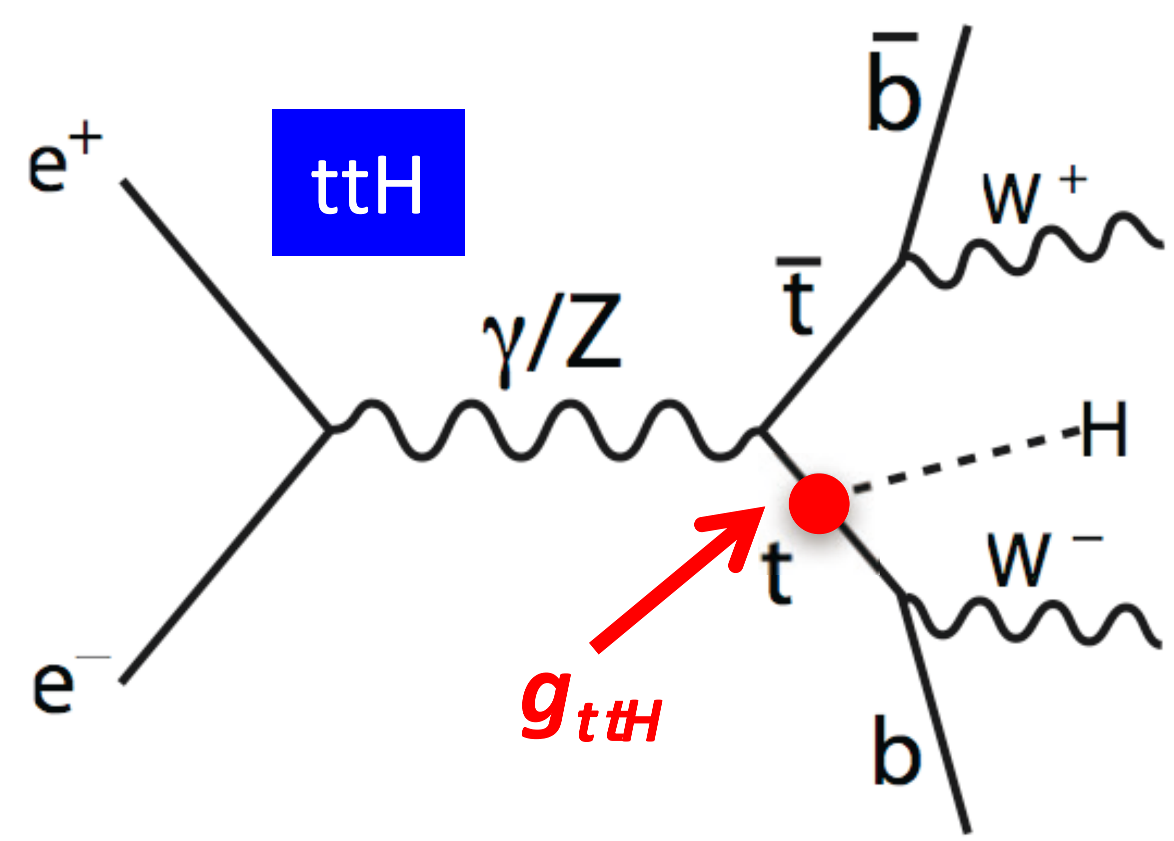
Outline

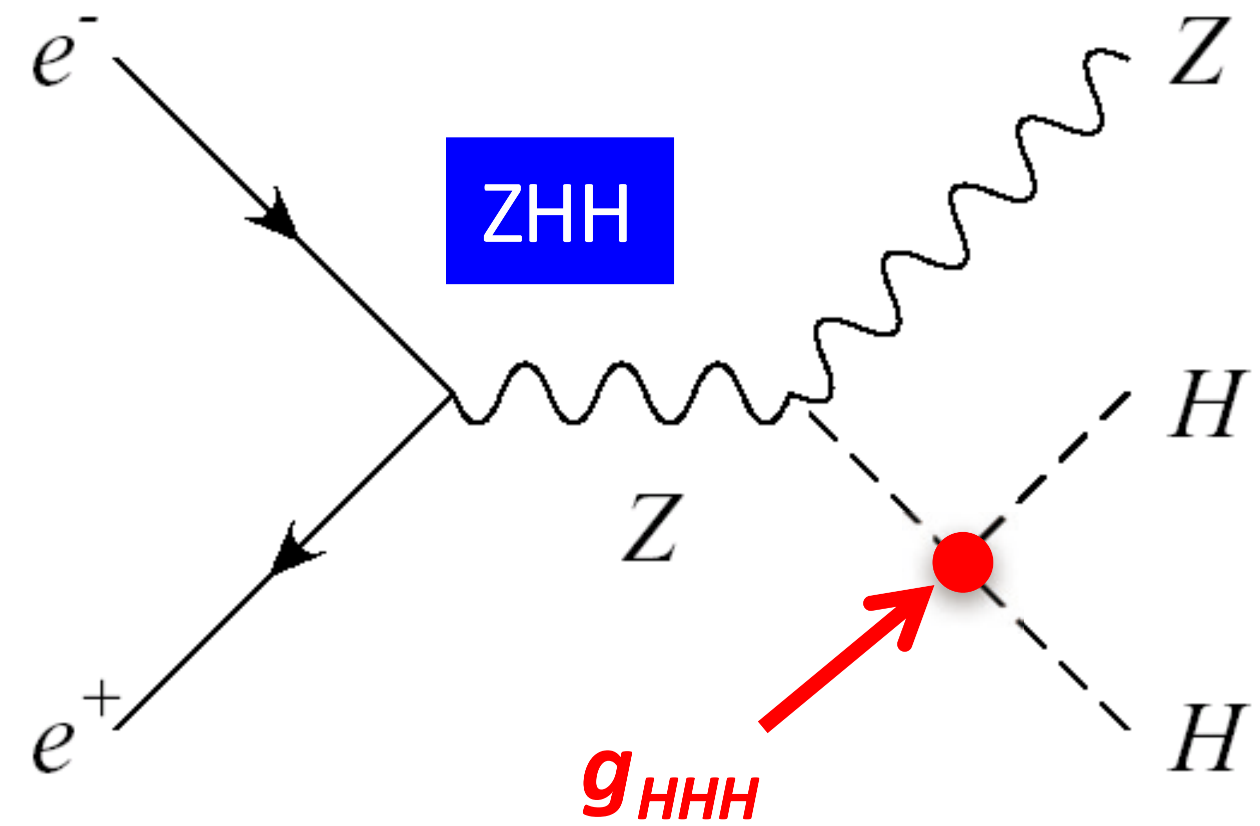
1. **Motivation**
2. **Flavour tagging features**
3. **Simulation and reconstruction**
4. **Flavour tagging performance**
    - Jet-energy dependence
    - Jet-angle dependence
    - Impact of the $\gamma\gamma \rightarrow$ hadrons background
    - Impact of the jet reconstruction
5. **Summary and future work**

Motivation

Many important CLIC benchmark processes have multiple flavour jets

- **Higgs hadronic BRs**: H → **bb**, **cc**, gg
- **Higgs self-coupling**: ZHH → qq**bbbb**
- **Top-Yukawa coupling**: ttH → **b**W**b**W**bb**





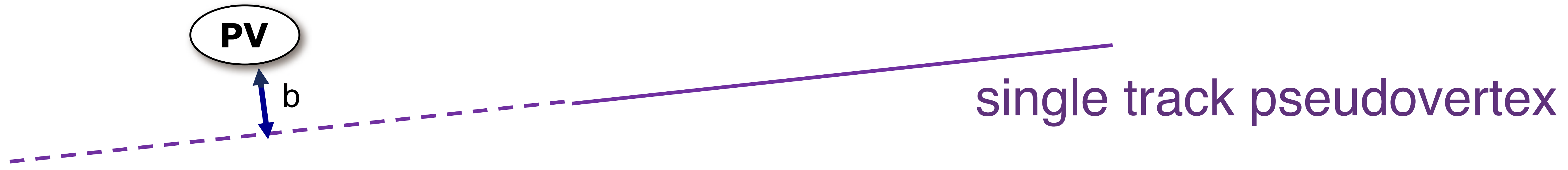
Previous CLIC detector studies have shown that a 20% change in the fake rate for light jets leads to a 6-7% effect on the precision for **H → bb** and 15% on **H → cc**

"Optimisation studies for the CLIC vertex-detector geometry", N. Alipour Tehrani
http://stacks.iop.org/1748-0221/10/i=07/a=C07001

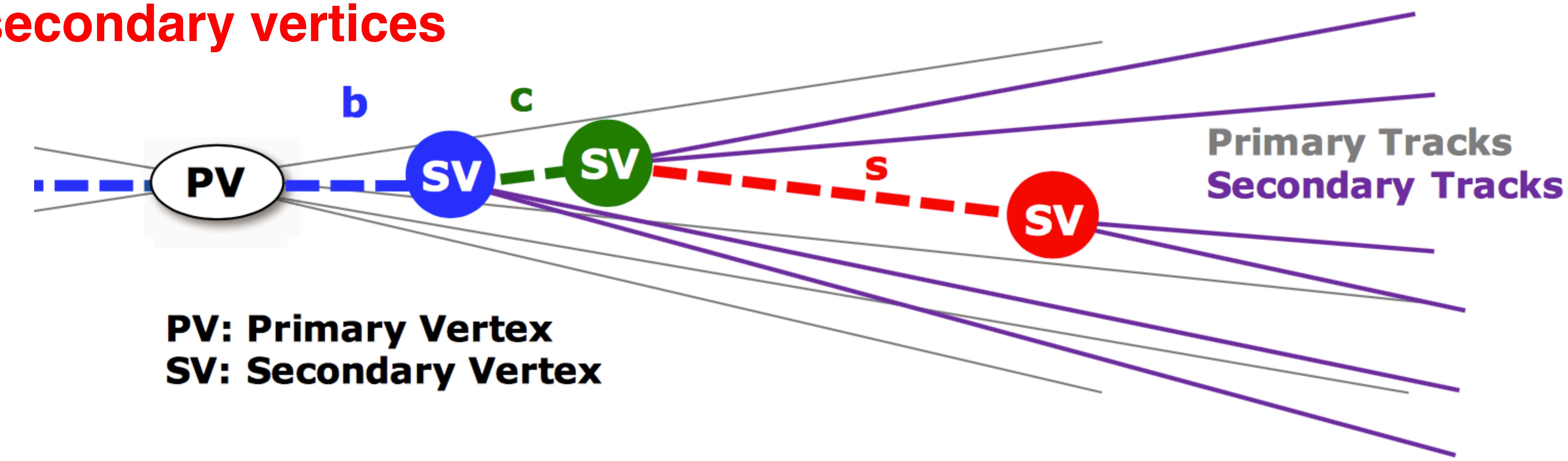
Flavour tagging performance has a large impact on final states with many **b** jets

Flavour tagging: Vertexing

- Vertex reconstruction is crucial for flavour tagging
  - Require at least two reconstructed tracks
  - Use track **impact parameter** if vertex reconstruction is not possible



- Key signature of heavy quarks → **secondary vertices**



- Lifetime of:
  - **c** hadrons: cτ ~ 80μm
  - **b** hadrons: cτ ~ 400μm

- Vertex mass is also a powerful discriminant for b/c separation
  - m_c ~ 2 GeV
  - m_b ~ 5 GeV

- Requirement: vertex detector with a great spatial resolution

Flavour tagging: Multivariate Analysis

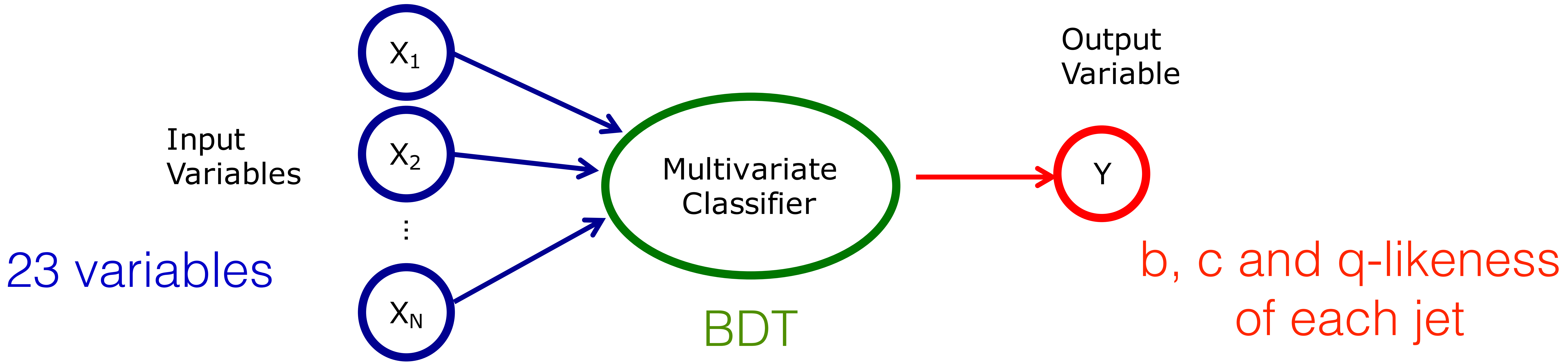
- We construct discriminating variables for each jet

- We then perform a multivariate analysis (as implemented in the **TMVA package of ROOT**):

  - To fully take advantage of the shape of the distributions, while taking into account the correlations among the variables

- We "train" the **multivariate classifier (BDT)** by using samples which we already know the "correct answer". The algorithm learns how to use the variables to arrive at the "correct answer"

  - We ensure that "training" and "testing" datasets are statistically independent when giving the results

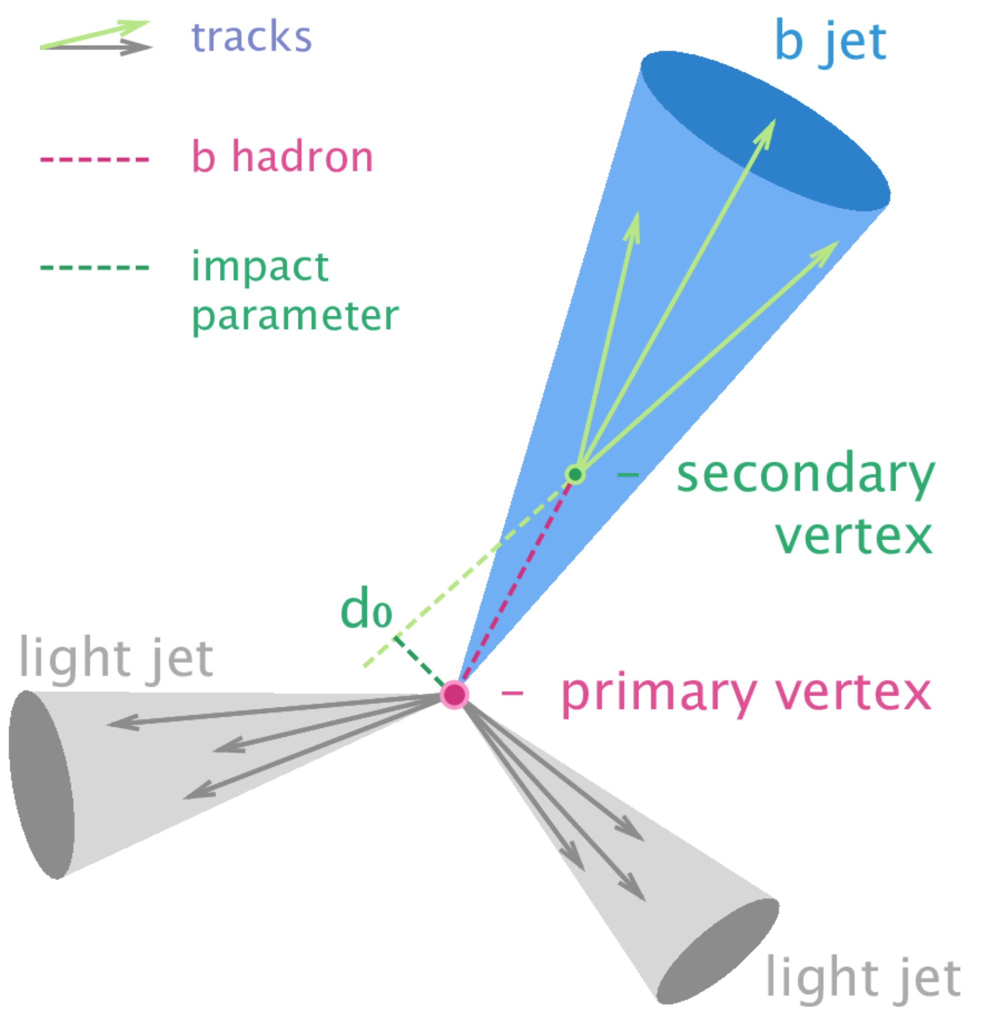


Flavour tagging: Categories

- For the training of the multivariate analysis, it is often helpful to divide the dataset into different categories. This is especially the case if we know that they will be very different.

- The dataset is divided according to the number of reconstructed secondary vertices:

Category	A	B	C	D
Number of vertices	0	1	1	2
Number of pseudovertices	0-2	0	1	0

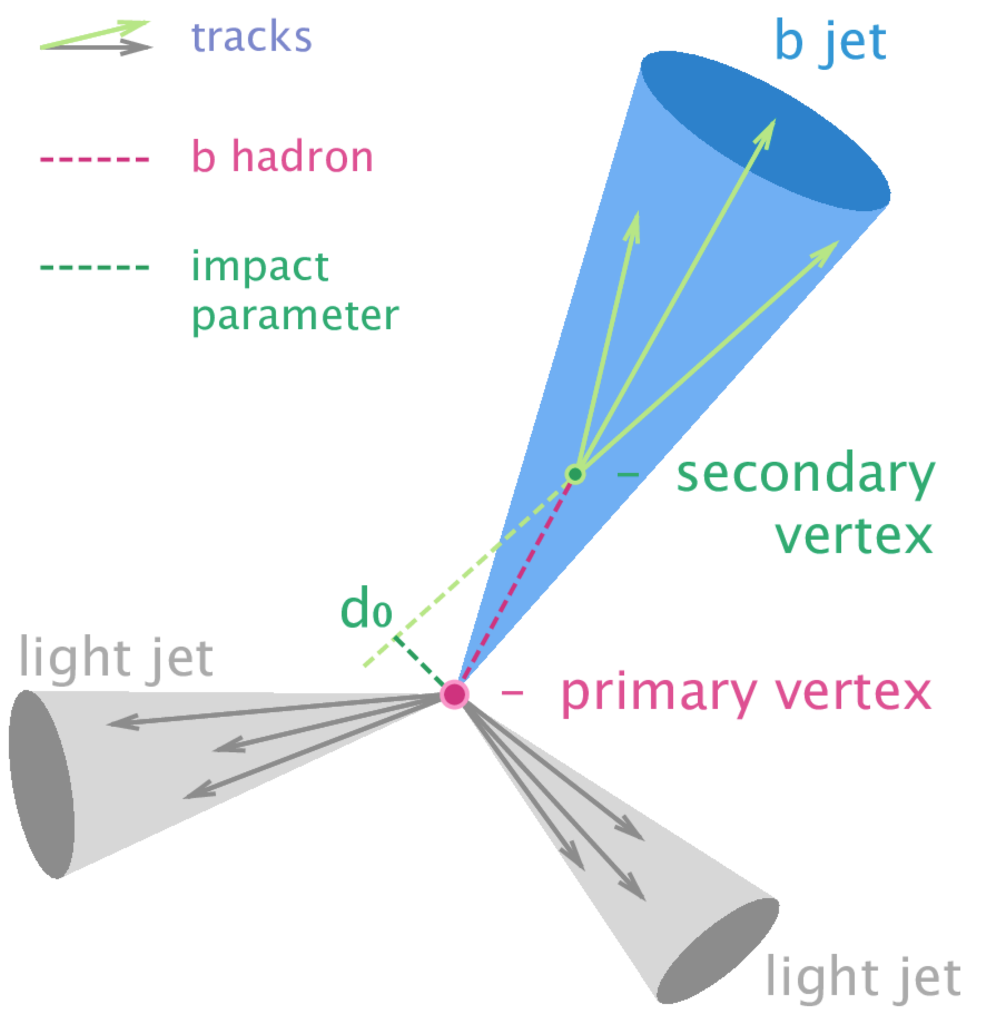


Flavour tagging: Categories

- For the training of the multivariate analysis, it is often helpful to divide the dataset into different categories. This is especially the case if we know that they will be very different.

- The dataset is divided according to the number of reconstructed secondary vertices:

Category	A	B	C	D
Number of vertices	0	1	1	2
Number of pseudovertices	0-2	0	1	0
	uds	**c**	**b**	



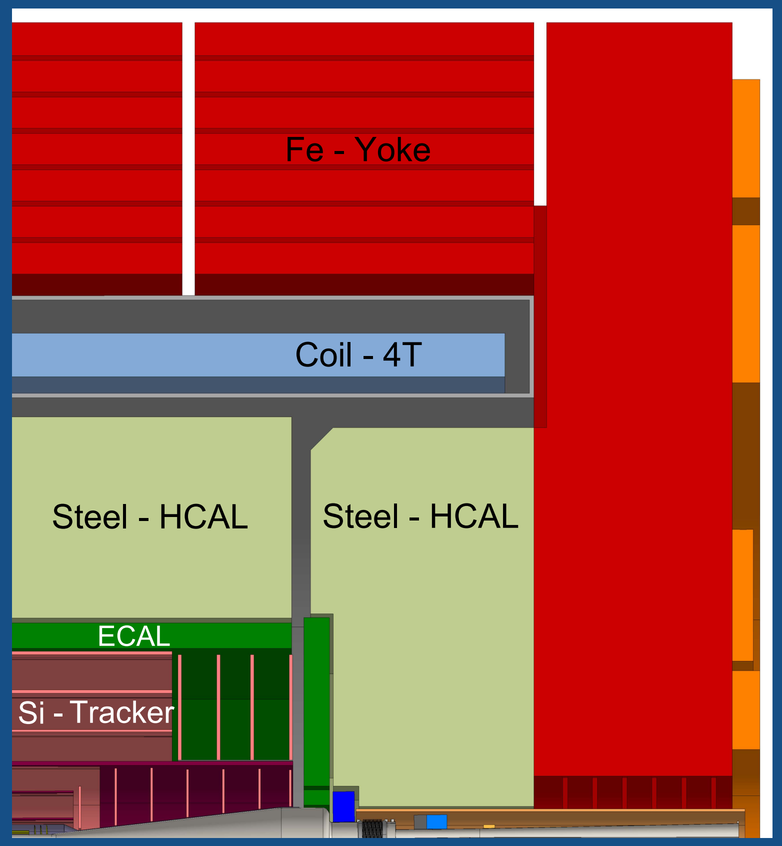
Category A: **uds** jets must be confined very well in the zero vertex category, which means a really good separation of **uds** jets from **b** and **c** jets

Category C: we can recover part of the **b** jets, which otherwise would have been grouped together in category B

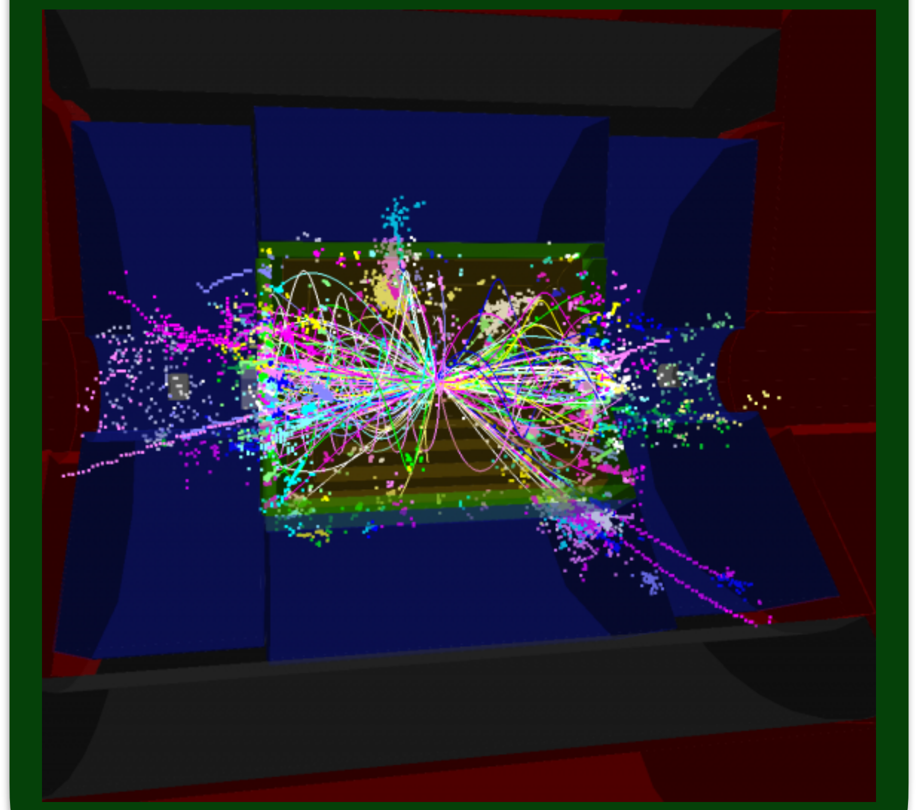
Category D: **c** and **uds** jets highly suppressed

Simulation and reconstruction

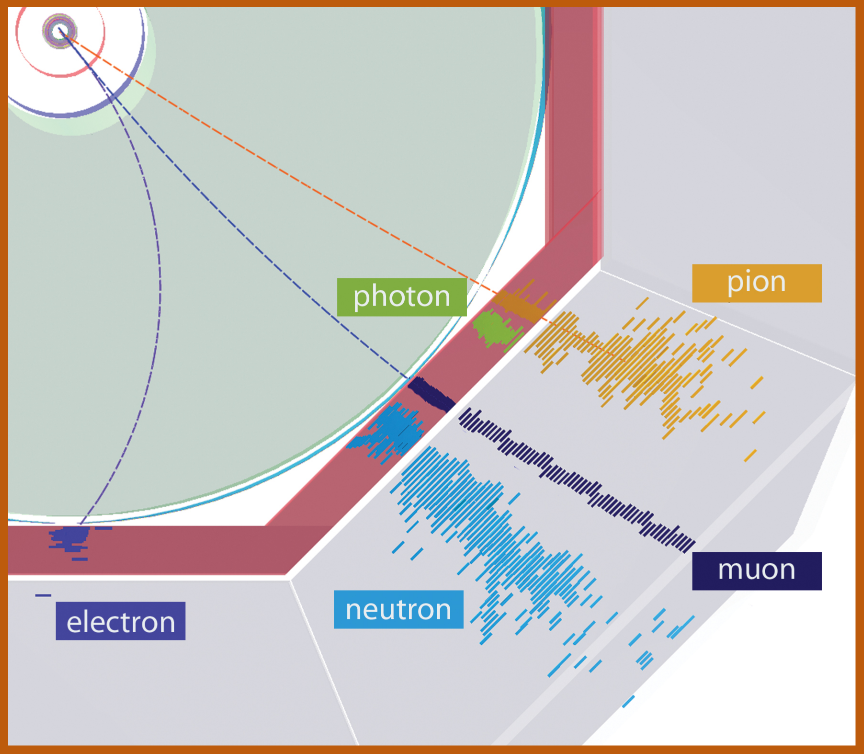
Detector Simulation



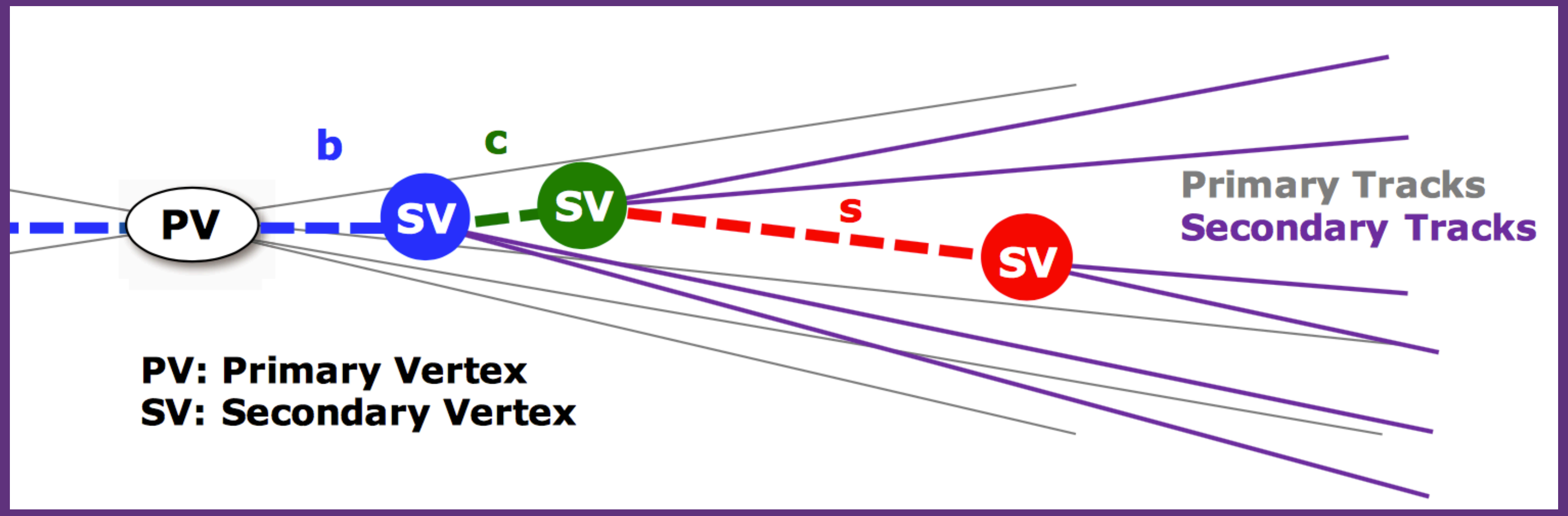
$\gamma\gamma \rightarrow$ hadrons Overlay



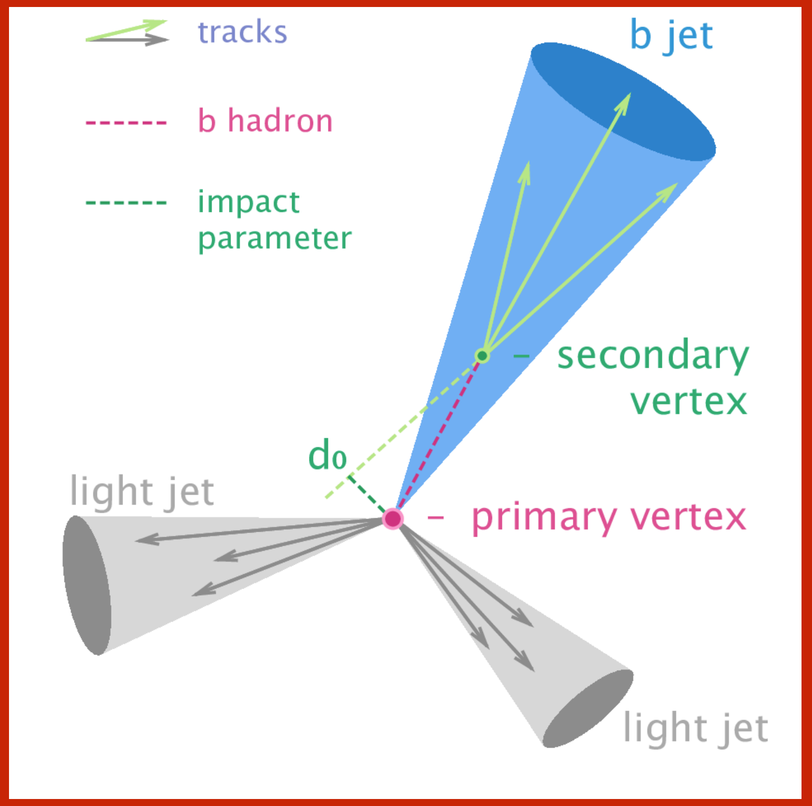
Reconstruction: Hits, Tracks, PandoraPFOs



Vertex reconstruction and jet clustering



Flavour tagging



Flavour tagging performance

- The performance of the flavour tagging have been studied for the CLIC detector model (CLIC_o3_v11) with the iLCSoft release (2017-07-12)

- Simulated and reconstructed samples:

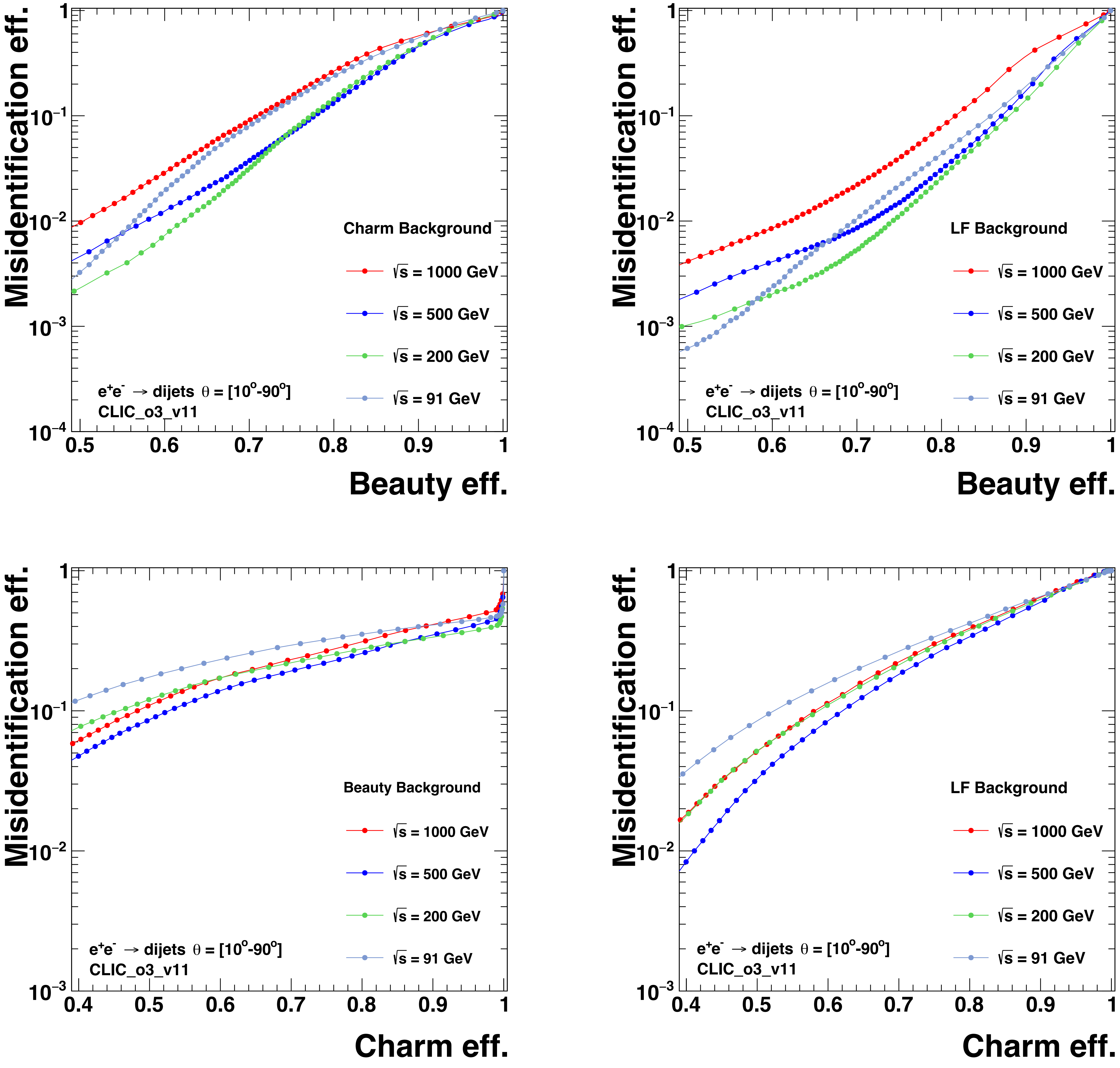
- **Dijet events $e^+e^- \rightarrow$ bb, cc, qq (q=uds)** (~160K events)
  - Different c.o.m. energies (91, 200, 500 and 1000 GeV)
  - Fixed jet angle $\theta = [10^o, 20^o, 30^o, \ldots, 90^o]$
  - No $\gamma\gamma \rightarrow$ hadrons background

- **$e^+e^- \rightarrow Z\nu\nu$ ($Z \rightarrow$ bb, cc, qq) at $\sqrt{s}$ = 350 GeV** (~190K events)
  - No $\gamma\gamma \rightarrow$ hadrons background
  - 0.0464 $\gamma\gamma \rightarrow$ hadrons / BX (Loose, Selected and Tight timing cuts available)

- The flavour tagging performance is evaluated extracting the percentage of fake rates for a b(c)-tag efficiency given (i.e. fraction of c(b) jets and uds jets that are misidentified as b(c) jets)

Jet-energy dependence

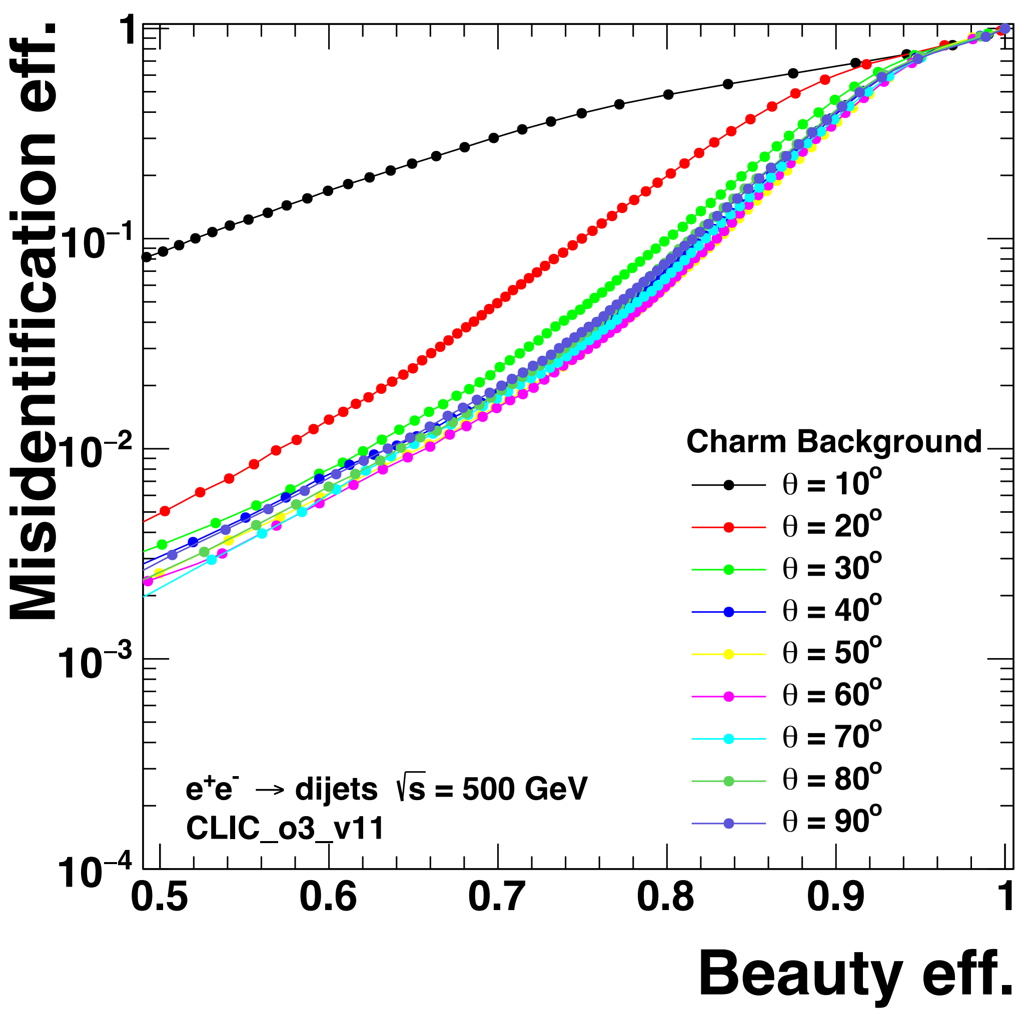
Dijet events $e^+e^- \rightarrow$ bb, cc, qq (q=uds)

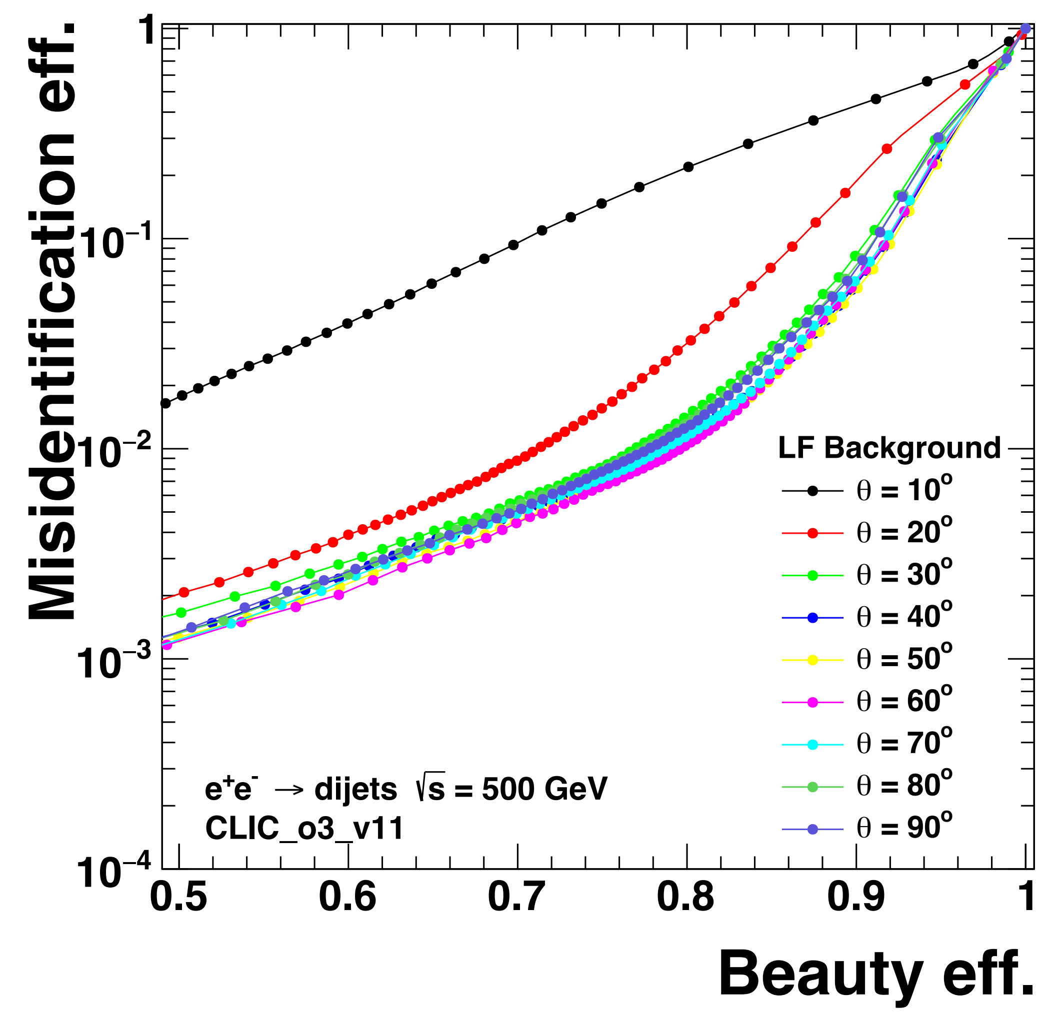


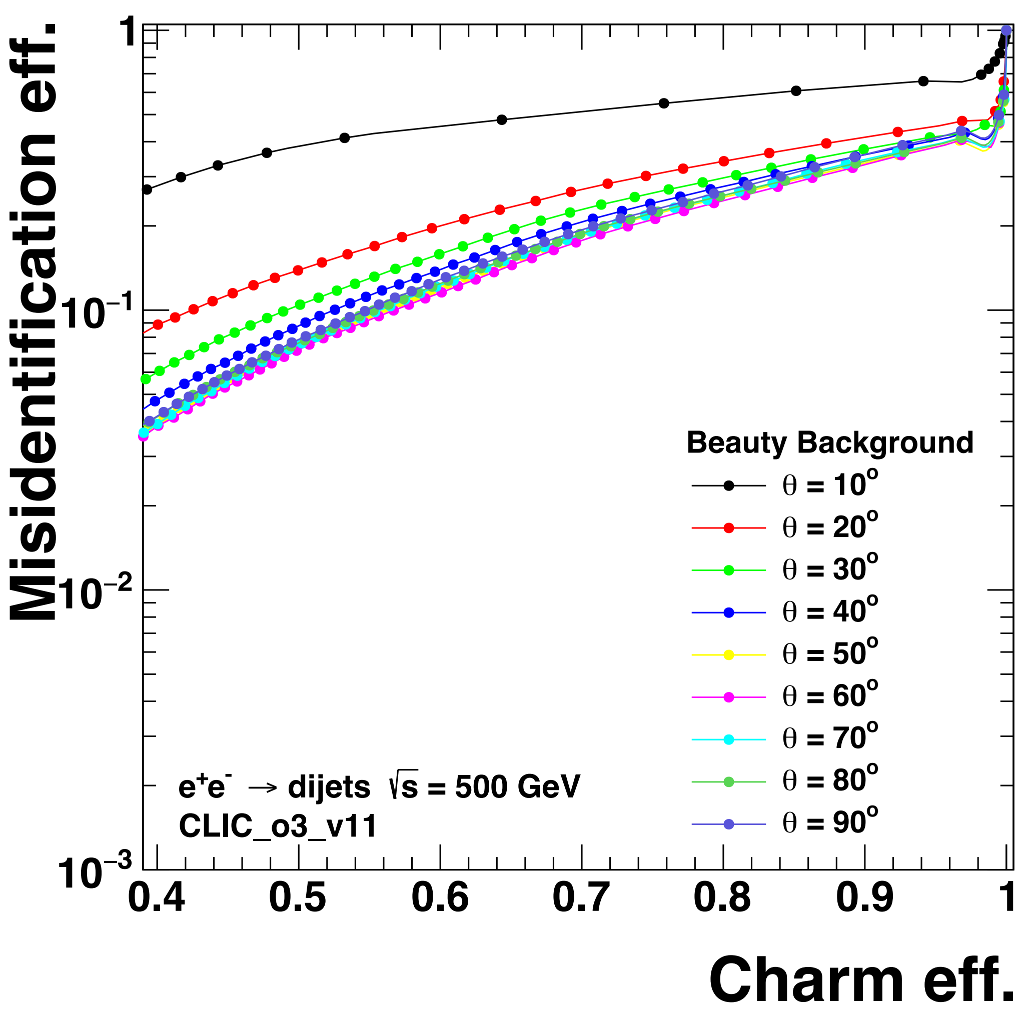
- **Mixture** of angles between **10° and 90°** for each energy
- Generally, the **b-tag** performance is **better for jets with lower energies**: low energy B hadrons have shorter decays, passing through all vertex detector layers
- The **c-tag** performance **improves considerably at 500 GeV**. At lower energies the c-quark decays close to de PV due to its shorter lifetime

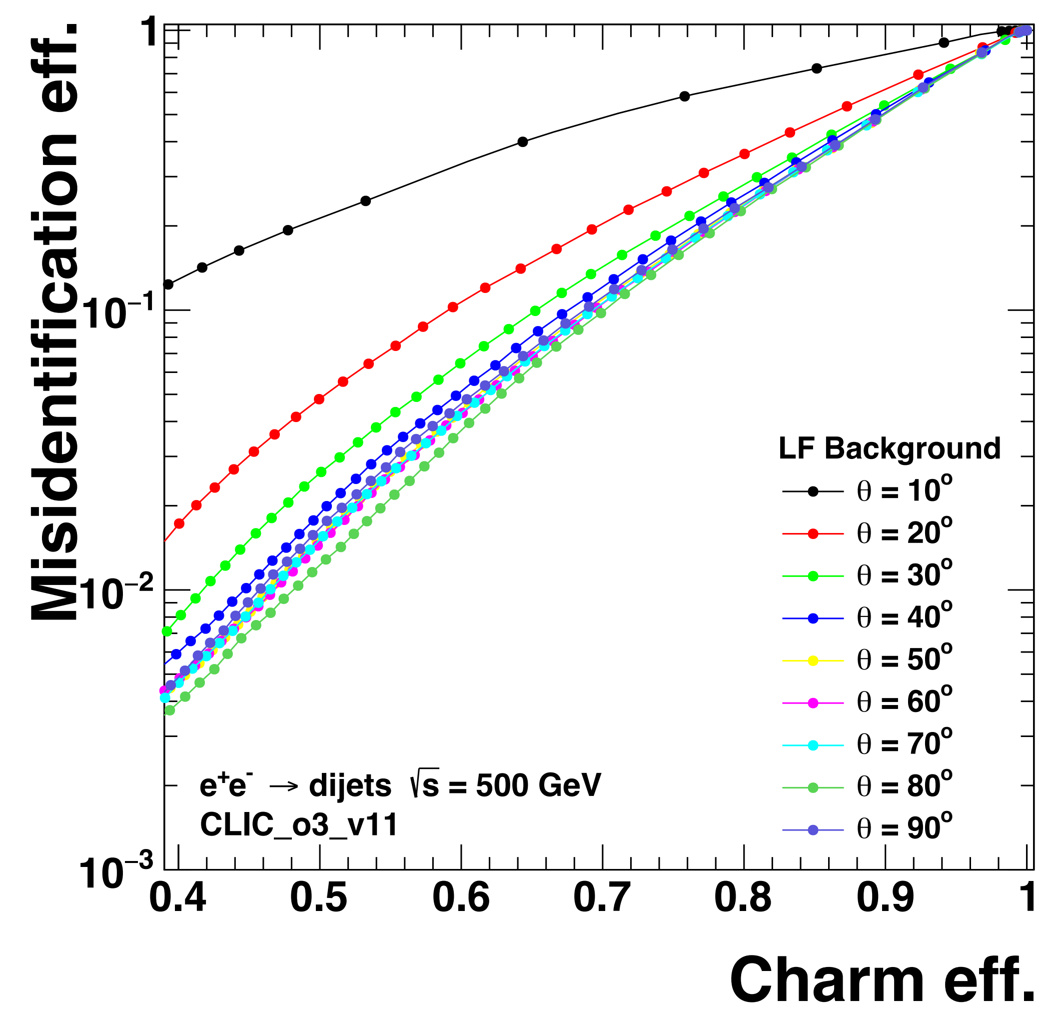
Jet-angle dependence

Dijet events $e^+e^- \rightarrow$ bb, cc, qq (q=uds)









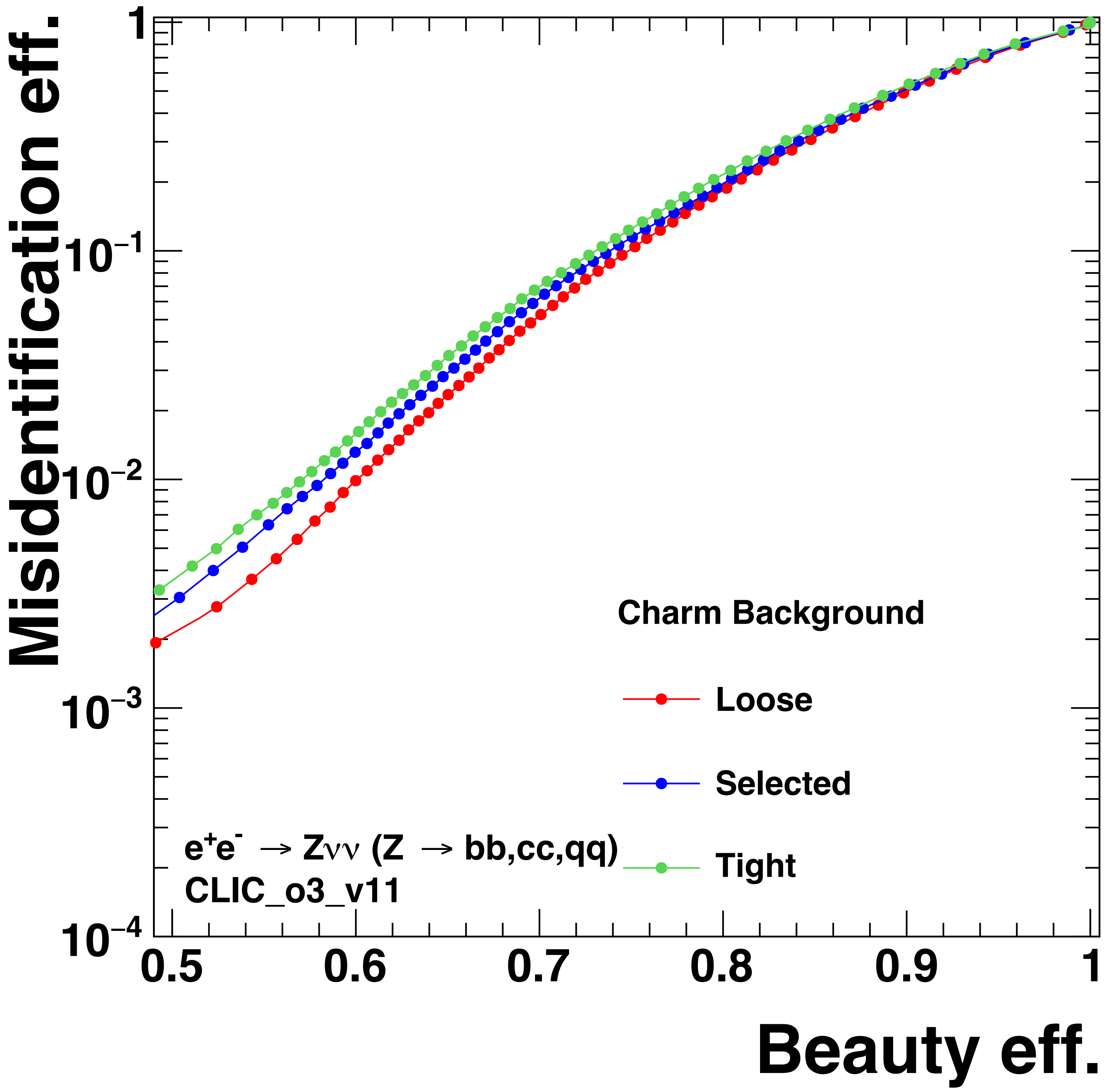
- A $\sqrt{s}$ = **500 GeV** is chosen
- A sizeable **decrease in performance** is observed in the **forward region** (~10°)
- Forward region of the vertex detector has **worse resolution** than other parts
- Some fraction of particles in jets is not reconstructed along the beam axis
- As polar angle decreases **less number of sensitive layers**

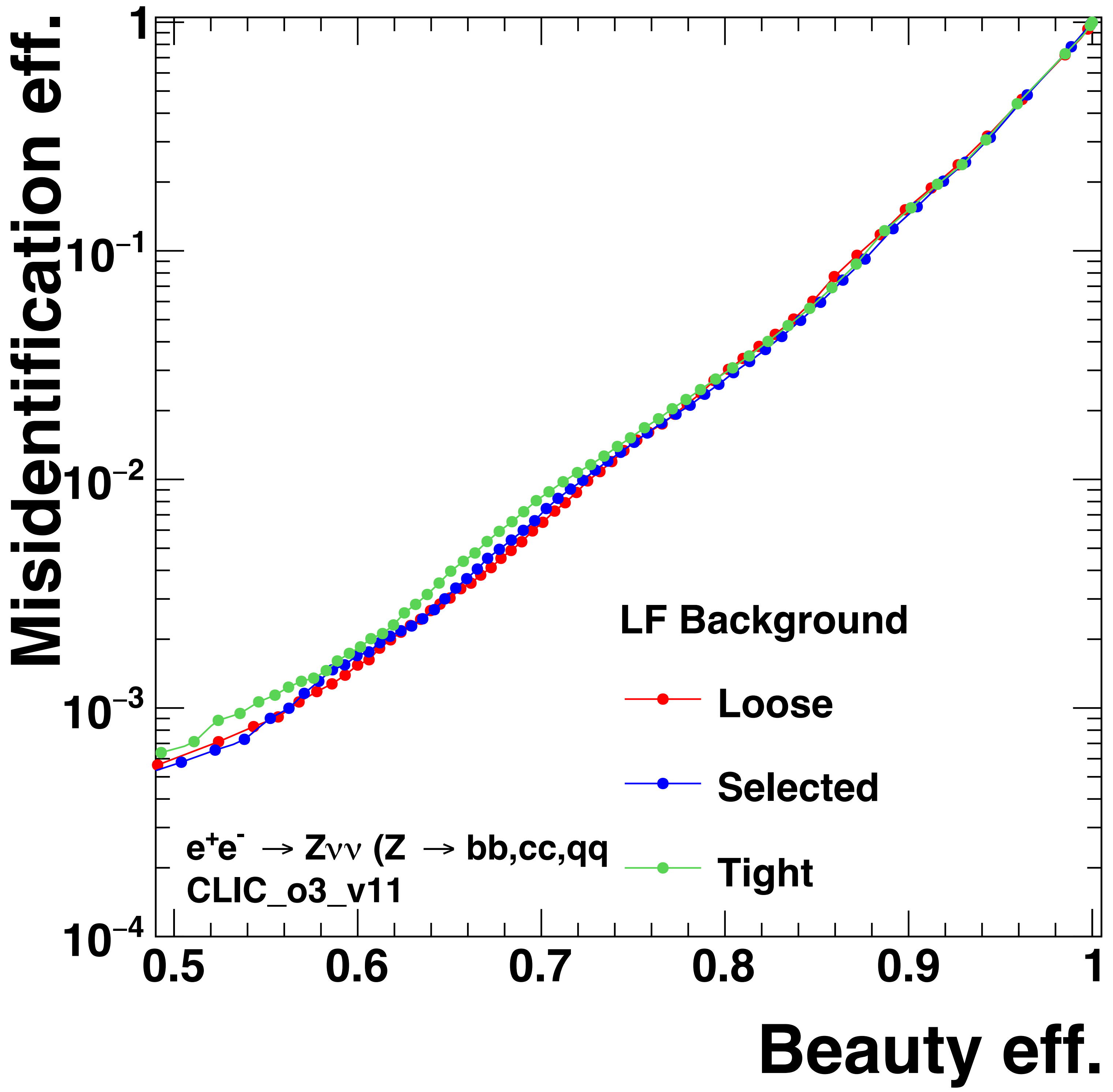
Impact of γγ → hadrons

$e^+e^- \rightarrow Z\nu\nu$ (Z → bb, cc, qq) √s = 350 GeV

0.0464 γγ → had. / BX

- Comparison of CLIC timing cuts (Loose, Selected, Tight)



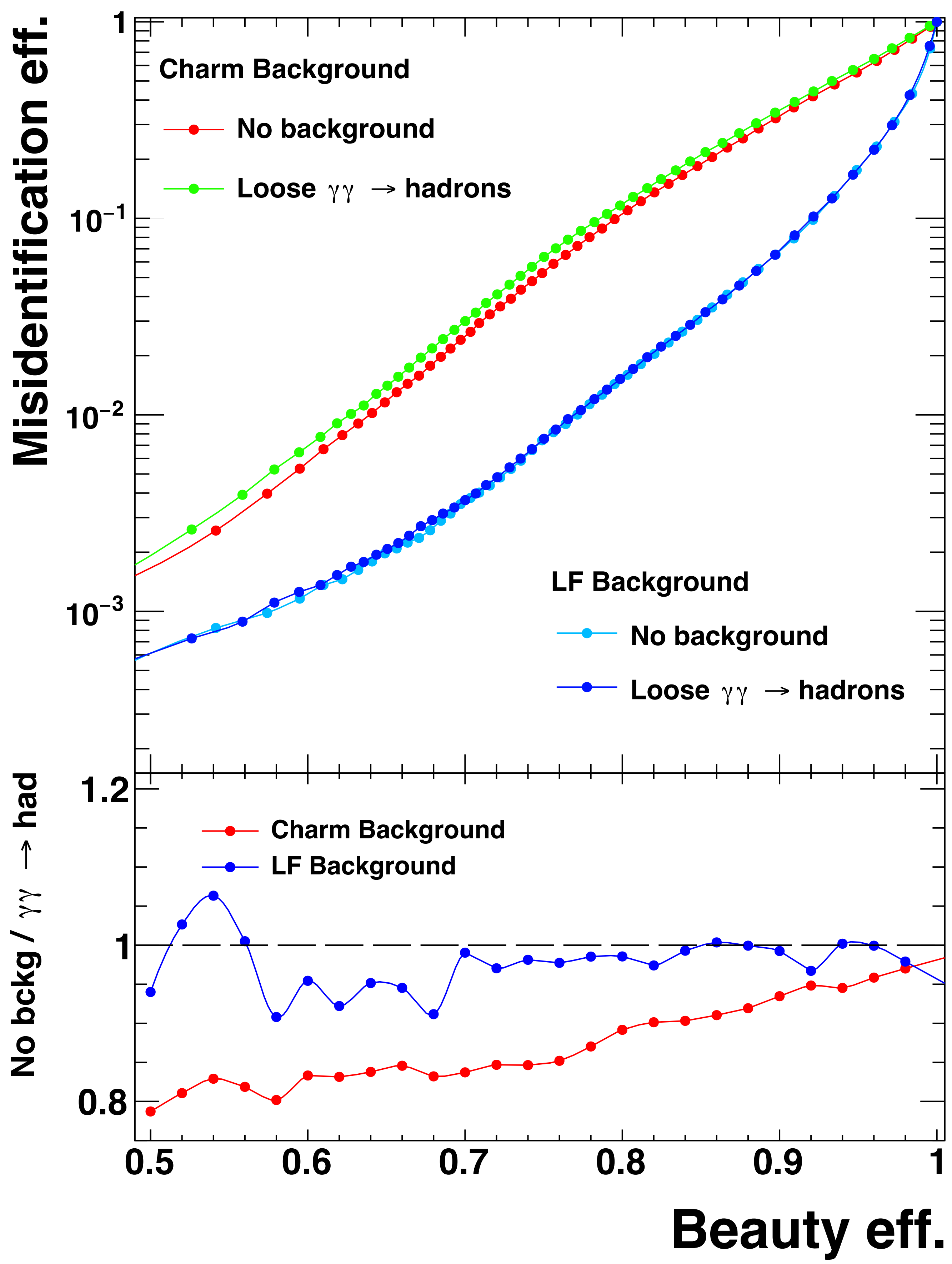


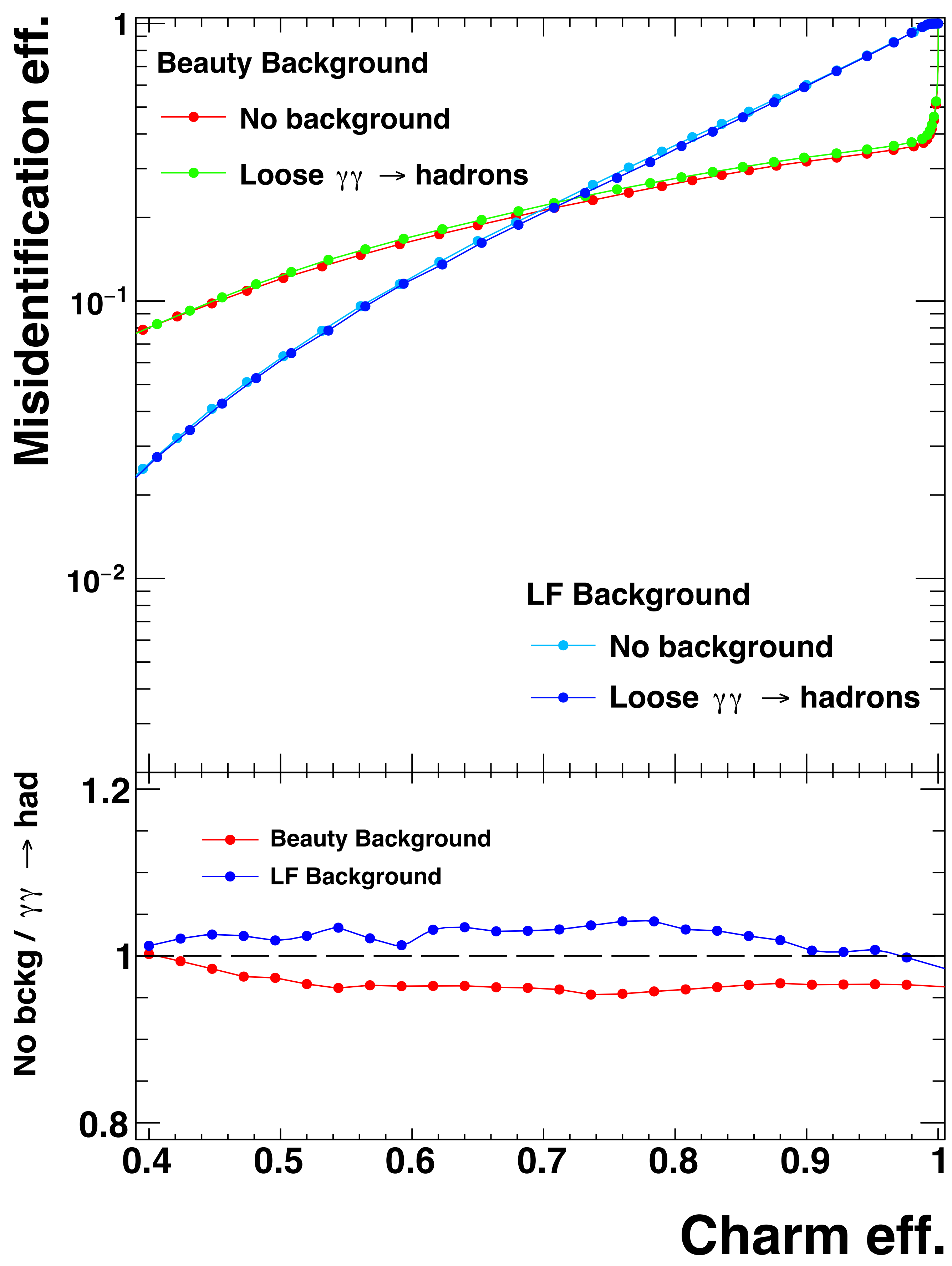
- **Loose timing cuts** seems to be the most suitable option. As expected for the γγ → hadrons background levels at the CLIC low energy stage (350-380 GeV)

Impact of γγ → hadrons

$e^+e^- \to Z\nu\nu$ (Z → bb, cc, qq) √s = 350 GeV
0.0464 γγ → had. / BX (Loose Timing cuts)

Ratio = No bckg/γγ → hadrons



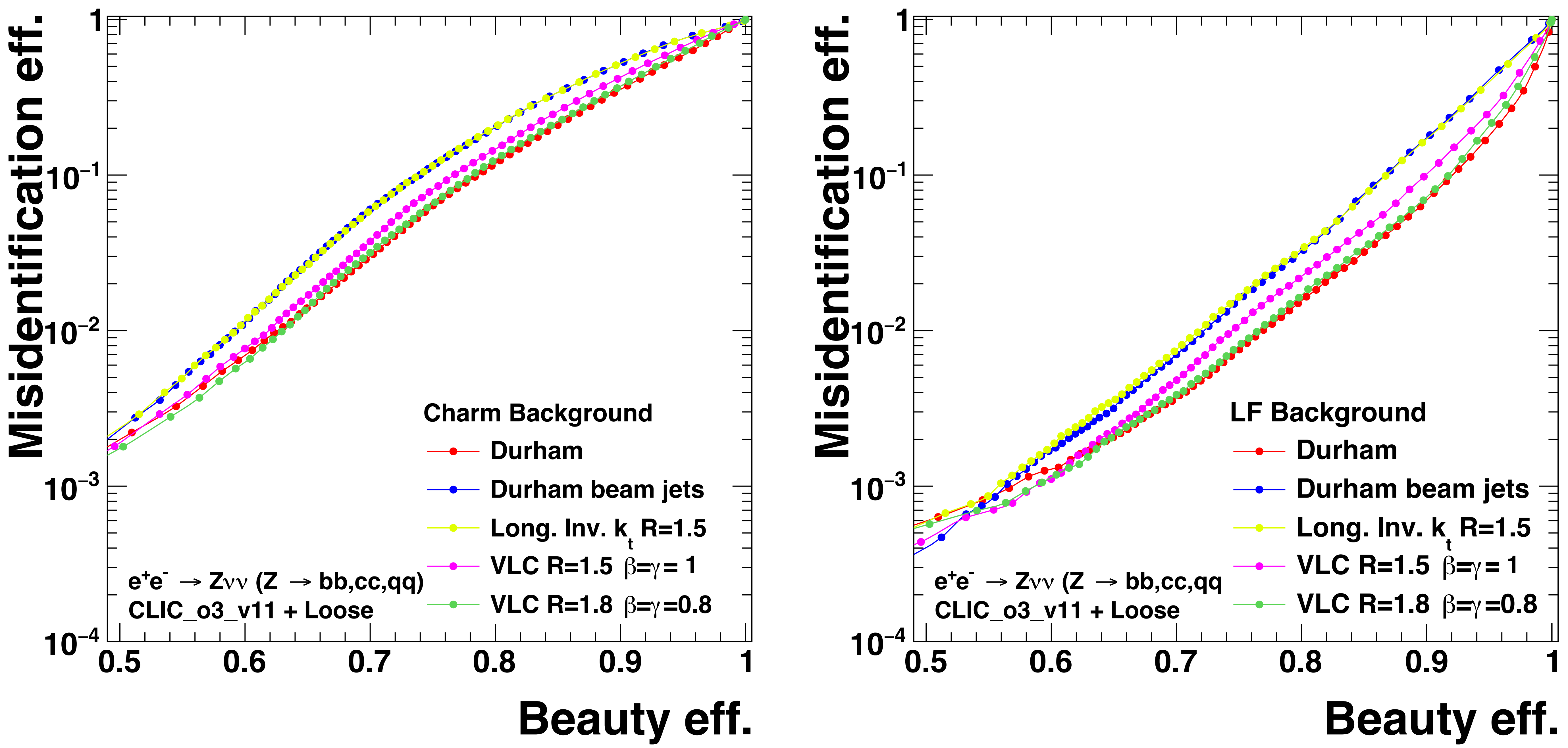


- **Durham algorithm** used for jet clustering
- **b-tagging**: The fraction of fake rates increases up to a **10% when γγ → hadrons** bckg. is overlaid, whereas decreases a 5% for LF bckg
- **c-tagging**: the impact of the background is less pronounced, variation ~5%
- **b-eff = 0.8**: 10% of misidentified c jets increases to 15% with γγ → hadrons

Jet algorithms comparison

$e^+e^- \to Z\nu\nu$ ($Z \to bb, cc, qq$) $\sqrt{s} = 350$ GeV

$0.0464\ \gamma\gamma \to$ had. / BX (Loose Timing cuts)



Several **jet algorithms** tested for comparison:

- Durham w & w/o beam jets
- Long. Inv. kt
- Valencia jet algorithm

Valencia performs better than kt for the same R value

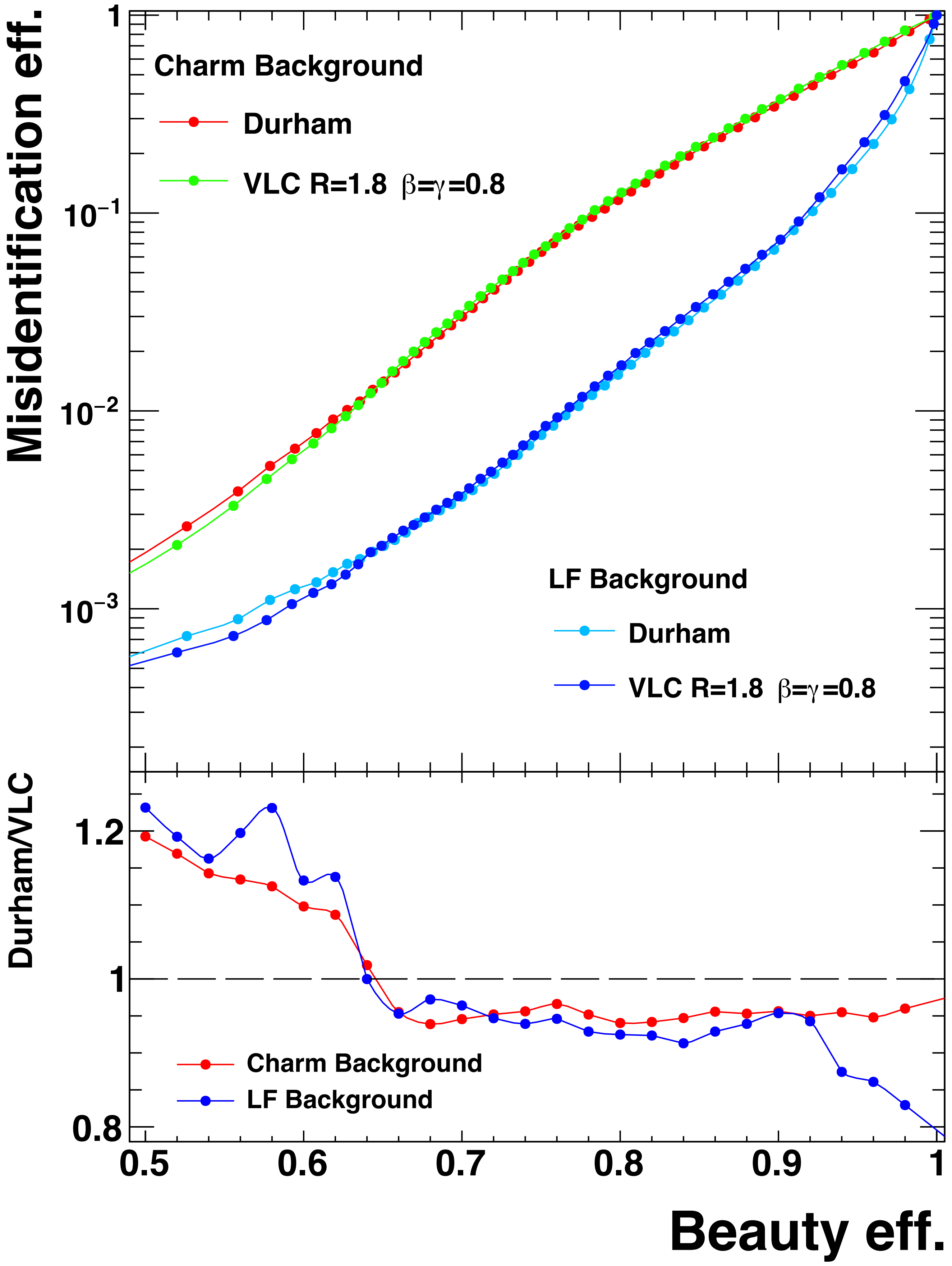
- **Best flavour tagging performance** given by the default **Durham** and **Valencia with R=1.8**
- *Lets compare both algorithms in more detail (next slide)*

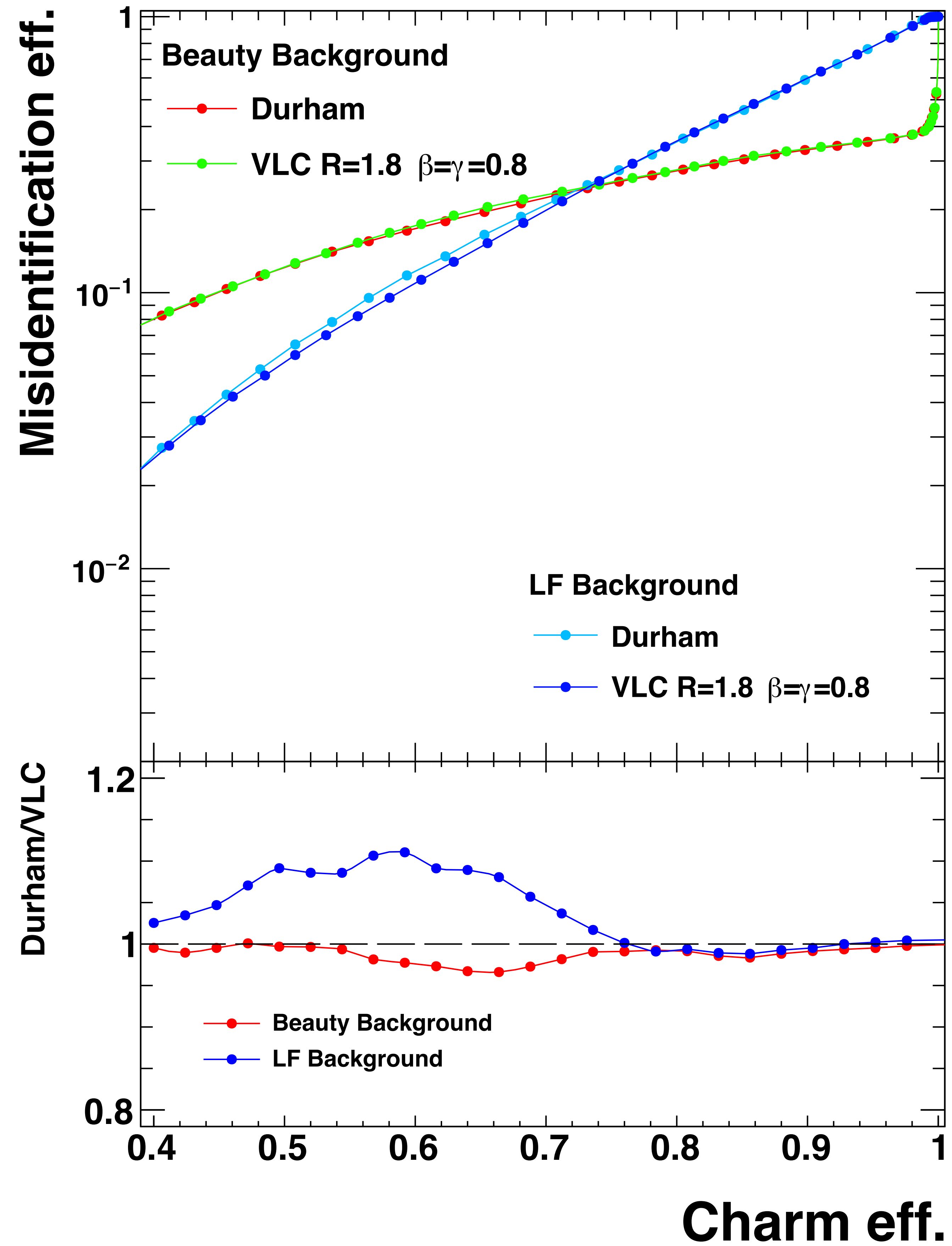
A robust jet algorithm

$e^+e^- \rightarrow Z\nu\nu$ ($Z \rightarrow$ bb, cc, qq) $\sqrt{s}$ = 350 GeV

0.0464 $\gamma\gamma \rightarrow$ had. / BX (Loose Timing cuts)

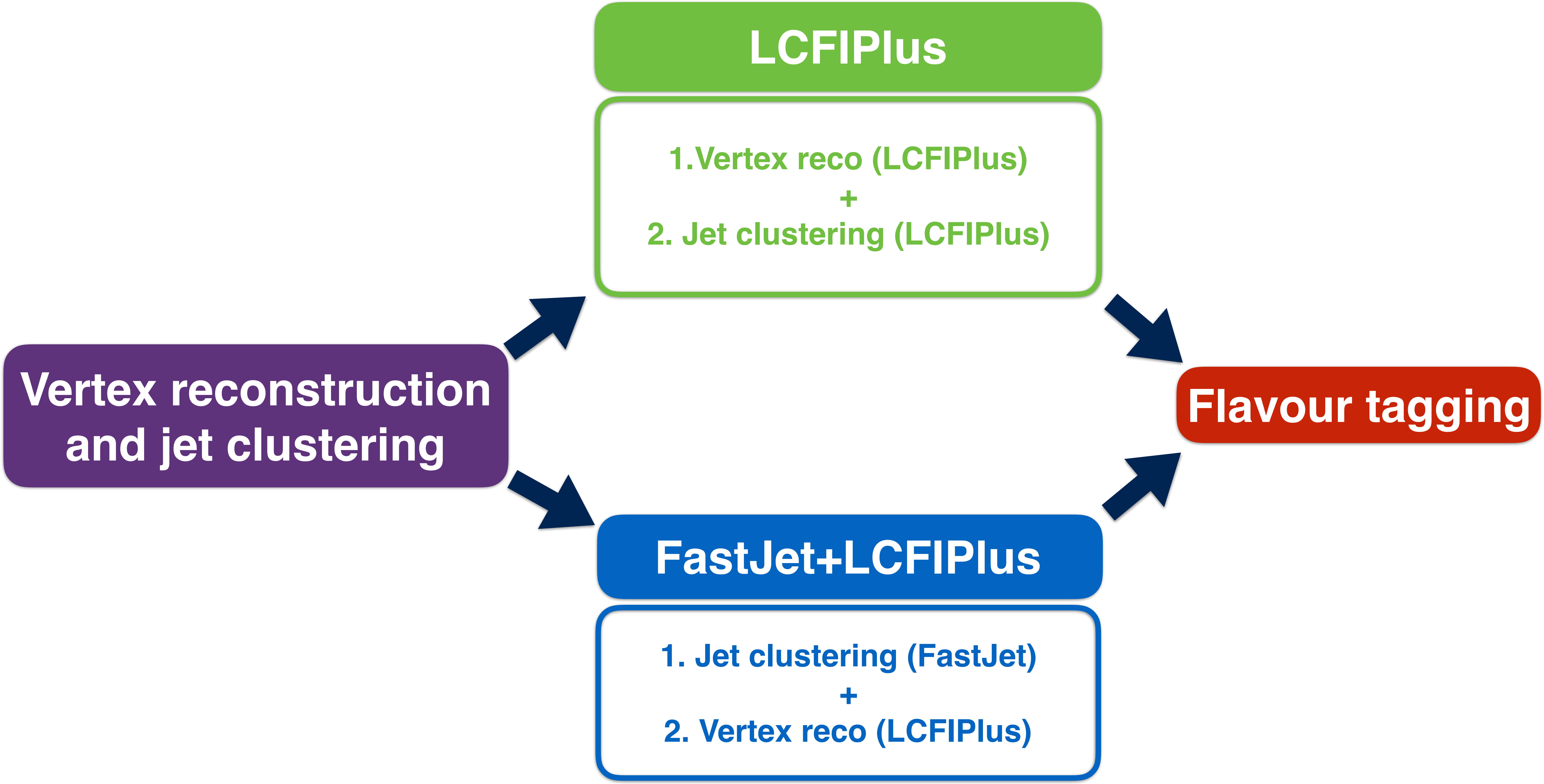
Ratio = VLC/Durham





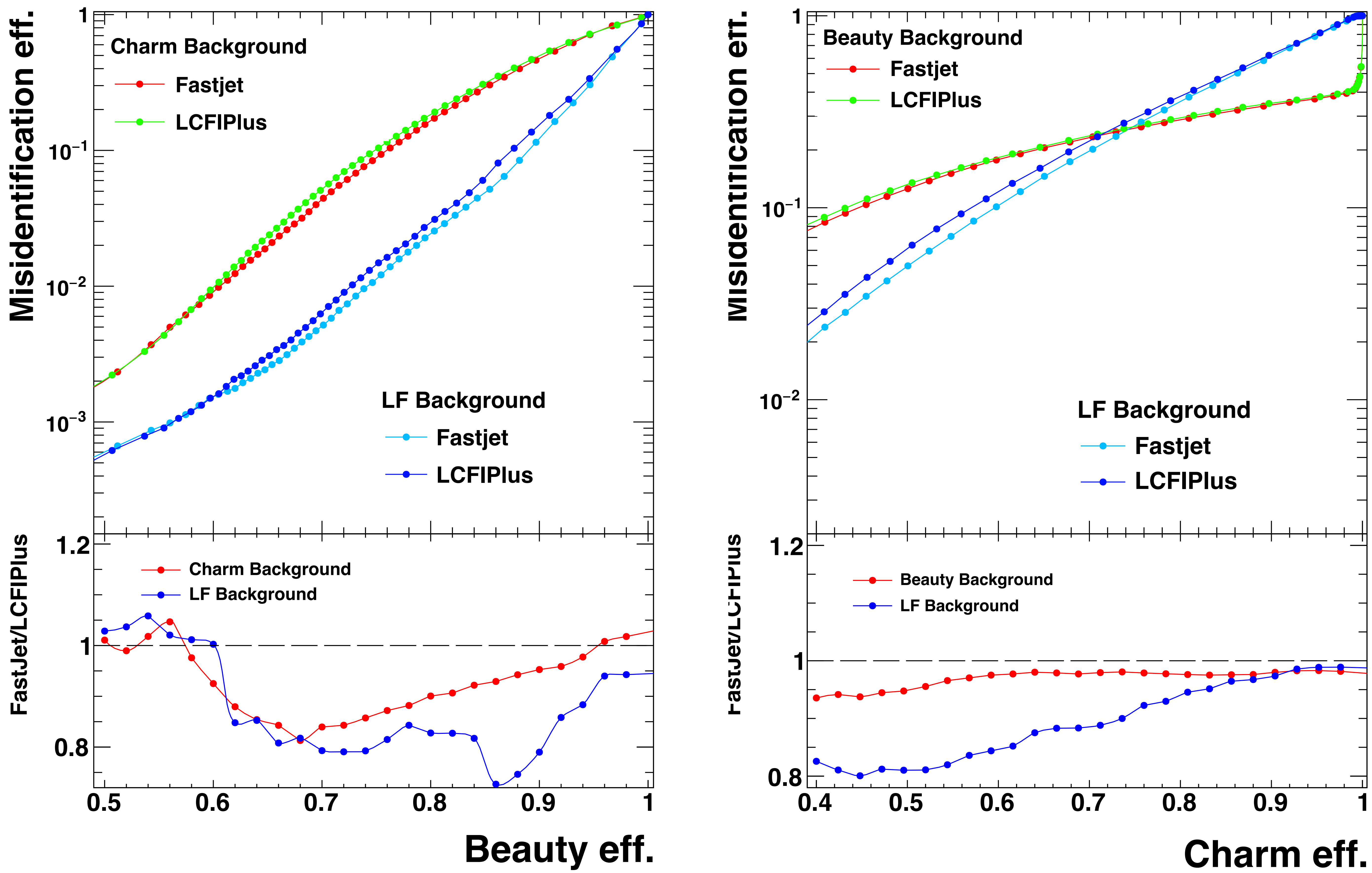
- **Durham algorithm**: All particles in the event are reconstructed into jets
- **Valencia jet algorithm**: presents more robustness against $\gamma\gamma \rightarrow$ hadrons (R, β, γ values has been chosen by optimisation scan)
- **b-tagging**: for lower values of b-eff the number of fake rates increases a 20% for Durham
- **c-tagging**: VLC algorithm allows to separate LF-jets from c-jets more efficiently

Vertex Reconstruction and jet clustering strategies



Jet clustering strategies comparison

$e^+e^- \rightarrow Z\nu\nu$ (Z → bb, cc, qq) √s = 350 GeV
0.0464 γγ → had. / BX (Loose Timing cuts)



Ratio = LCFIPlus/FastJet

- Valencia jet algorithm is used for jet clustering (R=β=γ=1)
- **b-tagging**: The LCFIPlus strategy leads to major fractions of fake rates (5-20%)
- **c-tagging**: better performance of LCFIPlus strategy at lower c-eff values
- **FastJet** strategy (jet clustering + vtx reconstruction) shows a better performance

Valencia jet algorithm reduces the number of background particles coming from γγ→hadrons, which benefits the vertex reconstruction and subsequently the flavour tagging

Summary

- Flavour tagging performance is better for lower energies and in the central region of the detector, severally degraded in the most forward region

- The impact of the $\gamma\gamma \rightarrow$hadrons on the flavour tagging performance translates into an increase of the fake rates up to 10% even using Loose timing cuts

- A robust algorithm against $\gamma\gamma \rightarrow$hadrons like Valencia jet algorithm performs slightly better than the classical Durham algorithm

- Vertex reconstruction and jet clustering strategy matters, being significantly better the FastJet + LCFIPlus strategy for b-tagging. Reduce the impact of $\gamma\gamma \rightarrow$hadrons before vertex reconstruction

- Future work:
    - Test flavour tagging performance at TeV scale (1.5TeV, 3TeV), much larger impact of $\gamma\gamma \rightarrow$hadrons expected
    - Compare the performance assuming different single point resolutions for the pixel sensor
    - Try new deep learning techniques for flavour tagging

Flavour tagging at kitchen

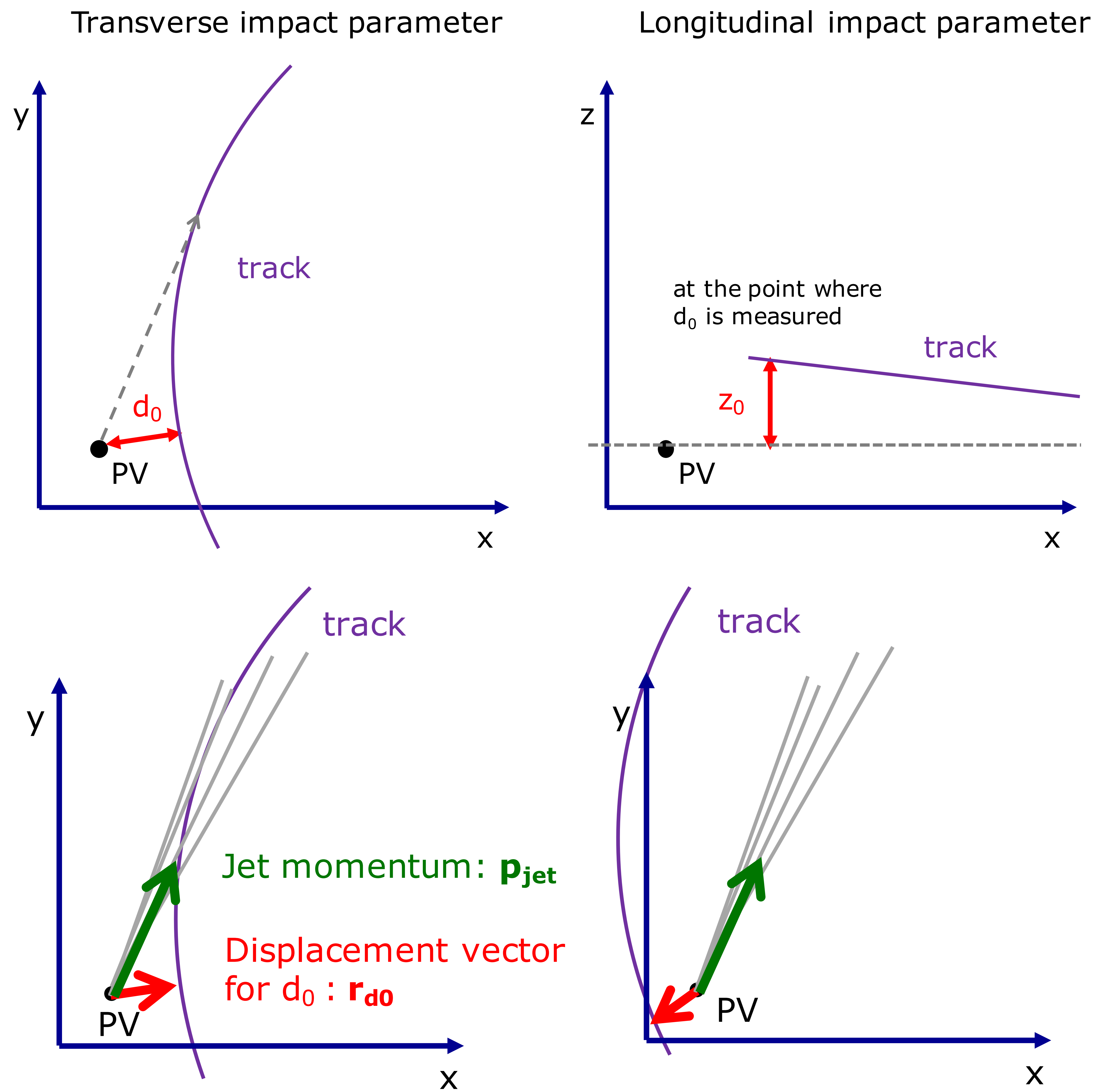
b = Coarse salt

c = Table salt

uds = Sugar

THANKS FOR YOUR ATTENTION!

Flavour tagging: Impact parameters d_0 and z_0



Positive if $(\mathbf{p_{jet}} \cdot \mathbf{r_{d0}}) > 0$

Negative if $(\mathbf{p_{jet}} \cdot \mathbf{r_{d0}}) < 0$

Impact parameter significance:
$S(d_0) = d_0/\sigma_{d0}$, $S(z_0) = z_0/\sigma_{z0}$

Uncertainty taken from track fit: σ_{d0} , σ_{z0}

Secondary decays should be in the direction of the jet

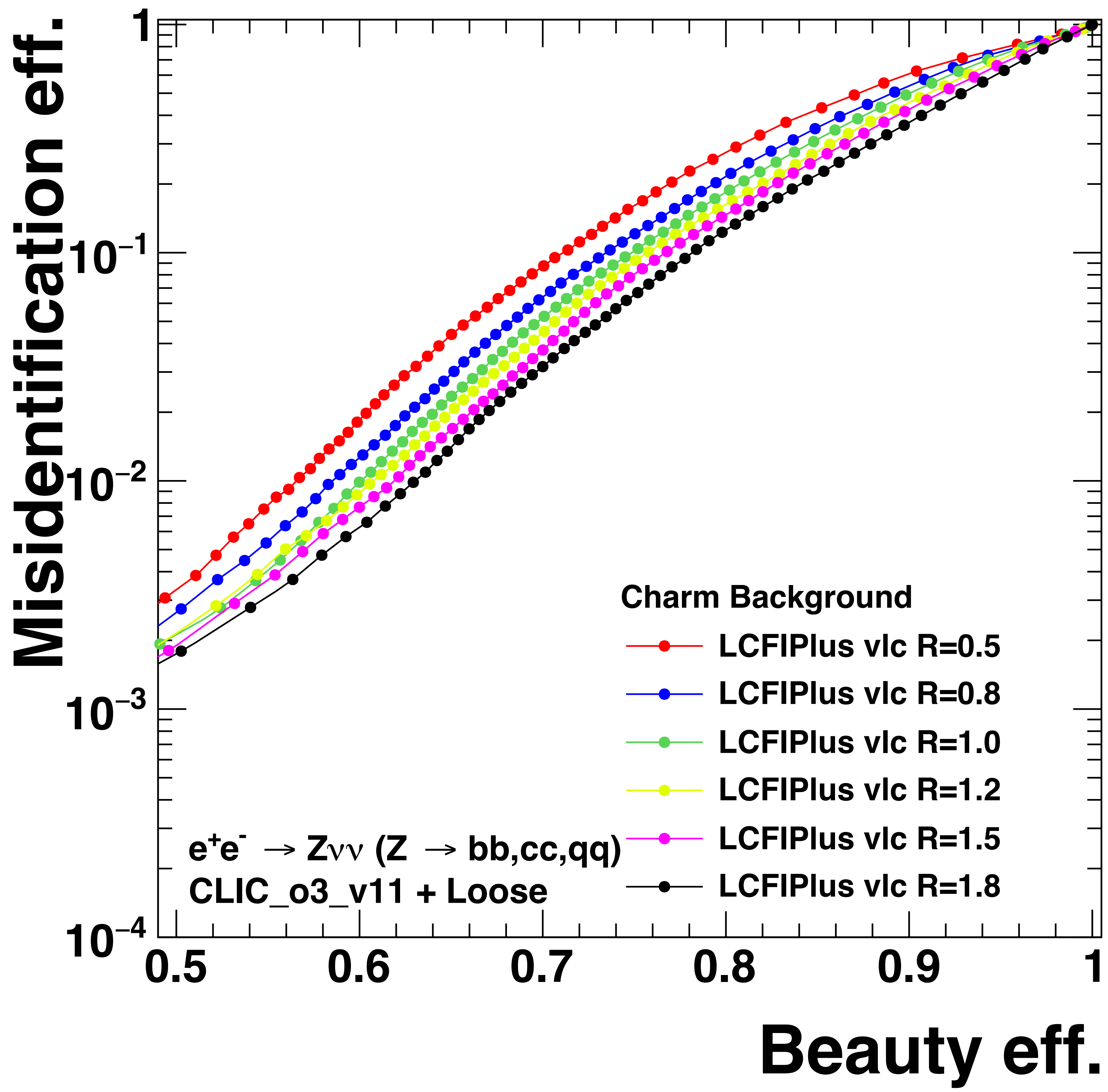
Marlin file

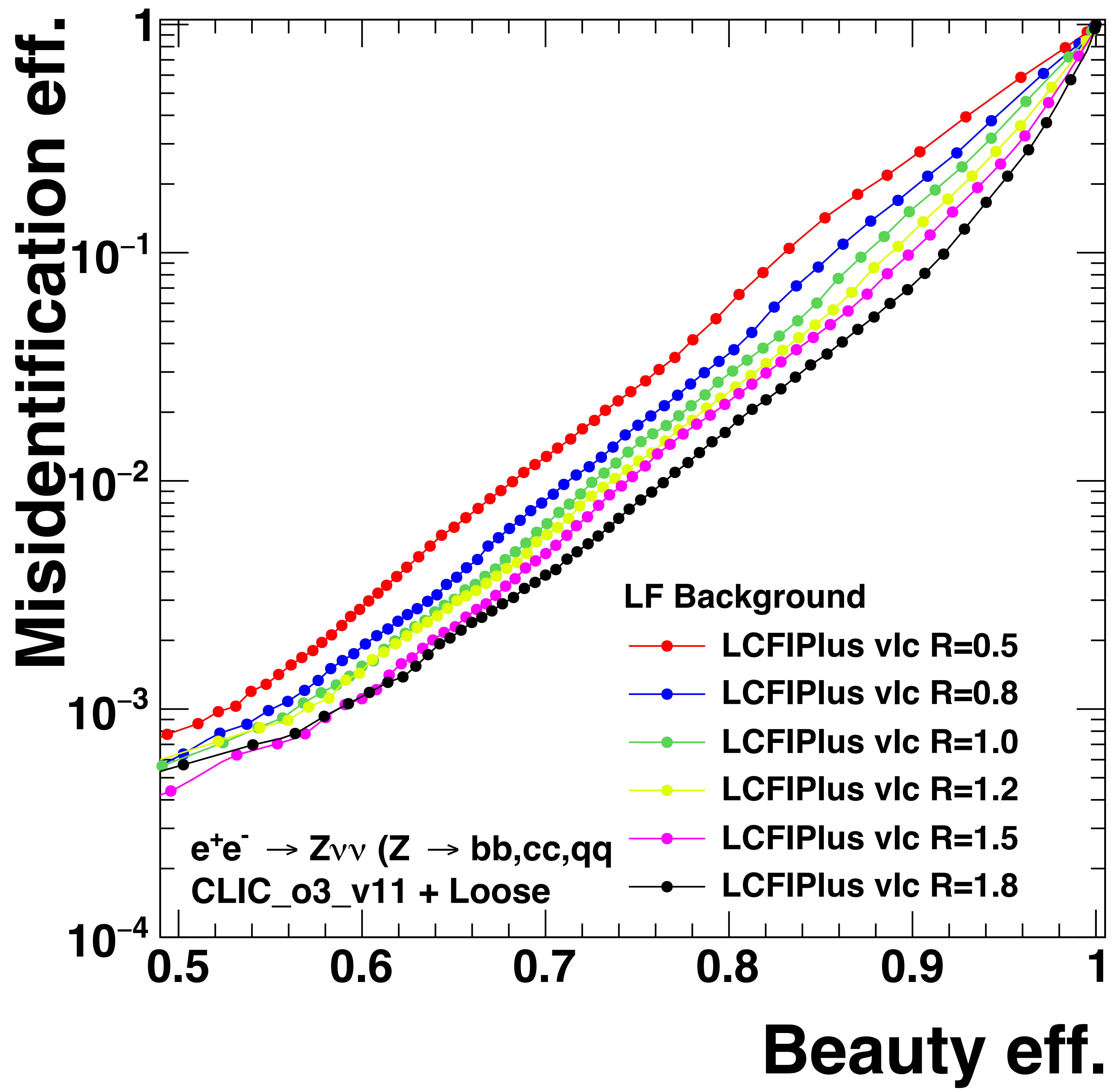
```
<!-- ========== setup  ========== -->
<processor name="MyAIDAProcessor"/>
<processor name="InitDD4hep"/>
<!-- ========== gg->hadrons background overlay  ========== -->
<processor name="MyOverlayTiming"/>
<!-- ========== digitisation  ========== -->
<processor name="VXDBarrelDigitiser"/>
<processor name="VXDEndcapDigitiser"/>
<processor name="InnerPlanarDigiProcessor"/>
<processor name="InnerEndcapPlanarDigiProcessor"/>
<processor name="OuterPlanarDigiProcessor"/>
<processor name="OuterEndcapPlanarDigiProcessor"/>

<!-- ========== tracking  ========== -->
<!-- At the moment the name of the final track collection for the MyTruthTrackFinder and MyExtrToTracker processors is the same, so that users can
    use this example to run easily both the cheater track pattern recognition (still the default for many tasks) or the real one (under final
    tests) -->
<processor name="MyTruthTrackFinder"/>
<!--<processor name="MyDDCellsAutomatonMV"/> -->   <!-- alternative to the ConformalTracking, but only in the vertex barrel region! -->
<!-- <processor name="MyConformalTracking"/> -->
<!-- <processor name="MyExtrToTracker"/> -->

<!-- === calorimeter digitization and pandora reco === -->
<processor name="MyDDCaloDigi"/>
<processor name="MyDDSimpleMuonDigi"/>
<processor name="MyDDMarlinPandora"/>
<processor name="LumiCalReco"/>
<processor name="BeamCalReco"/>
<!-- ========== monitoring  ========== -->
<processor name="MyClicEfficiencyCalculator"/>
<processor name="MyRecoMCTruthLinker"/>
<processor name="MyTrackChecker"/>
<Xprocessor name="MyHitResiduals"/> <!-- please uncomment the use of this processor only if needed -->
<!-- ========== output  ========== -->
<group name="PfoSelector" />
<!-- ========== gamma+gamma->hadrons removal =============================== -->
<Xprocessor name="MyFastJetProcessor"/>
<!-- ========== Vertex Finder ============================================== -->
<processor name="VertexFinder"/>
<!-- ========== JetClustering JetVertexRefiner FlavorTag ReadMVA  ========== -->
<Xprocessor name="jets"/>
<processor name="MyLCIOOutputProcessor"/>
```

Jet Clustering optimisation

$e^+e^- \rightarrow Z\nu\nu$ ($Z \rightarrow bb$, cc, qq) $\sqrt{s}$ = 350 GeV

0.0464 $\gamma\gamma \rightarrow$ had. / BX



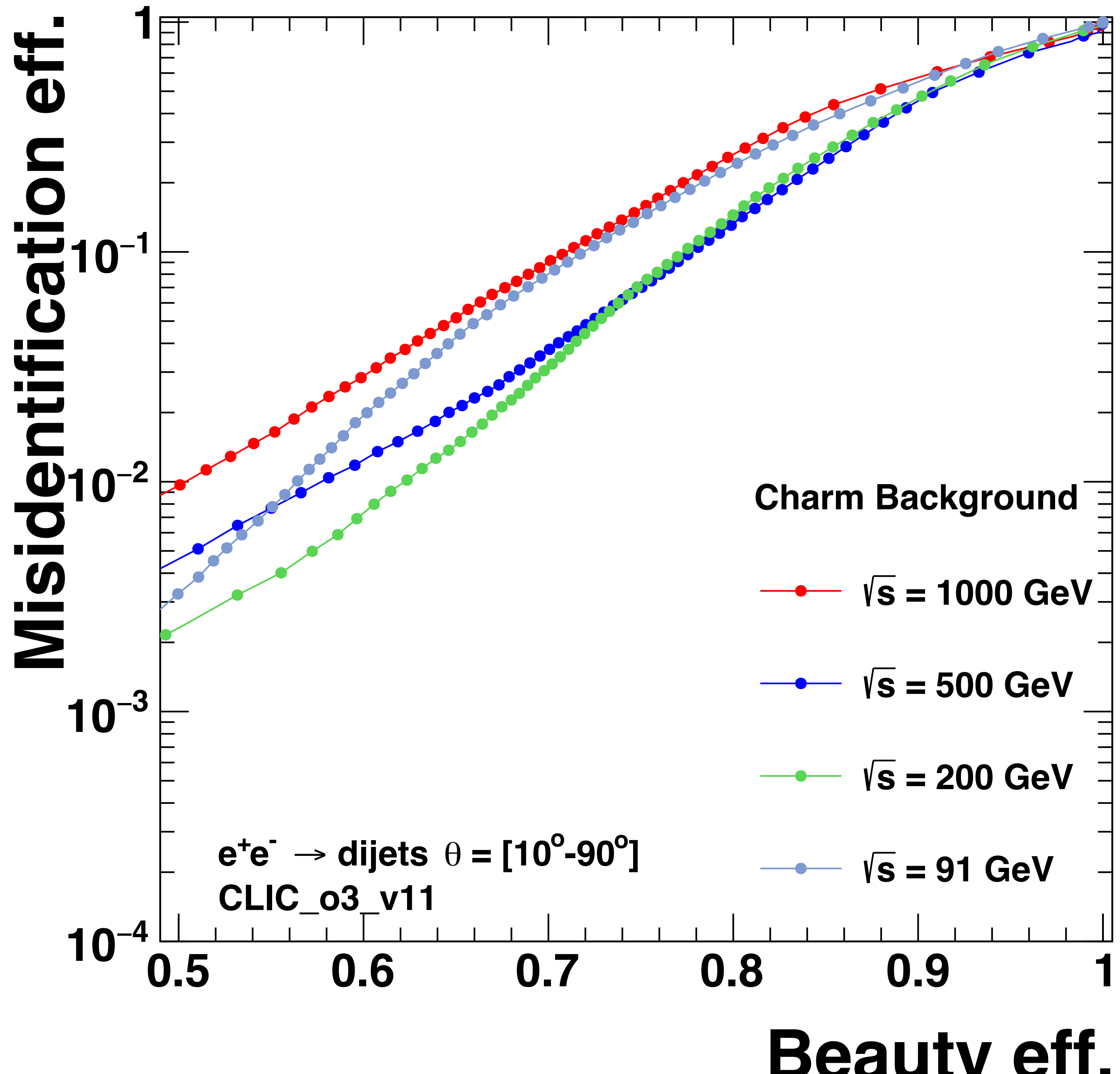


At 350 GeV, large radius (R =1.8) performs better

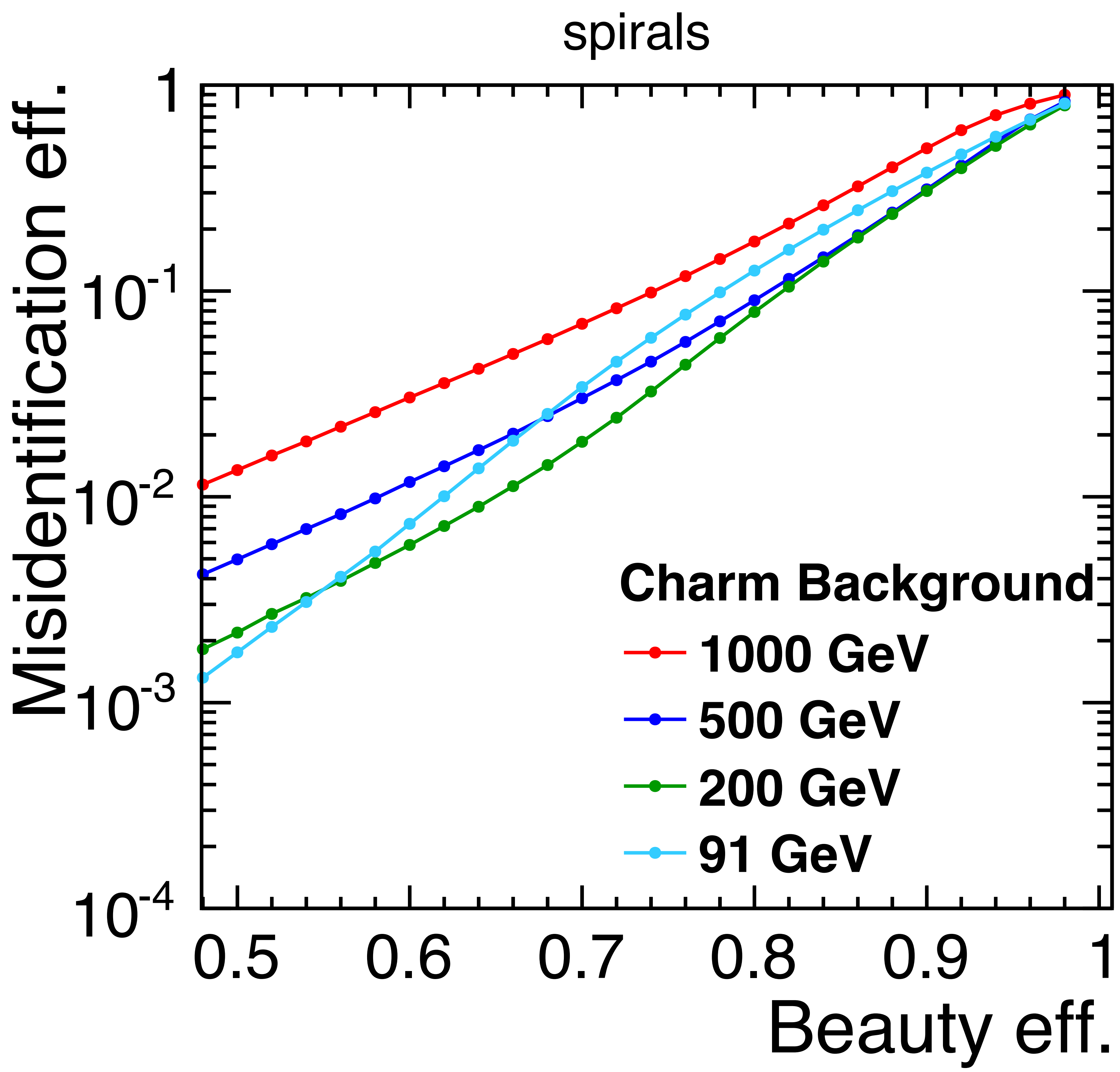
CLICdet vs CLIC_SiD: jet-energy dependence

Dijet events e⁺e⁻ → bb, cc, qq (q=uds)

NO γγ -> had. Overlaid



CLICdet



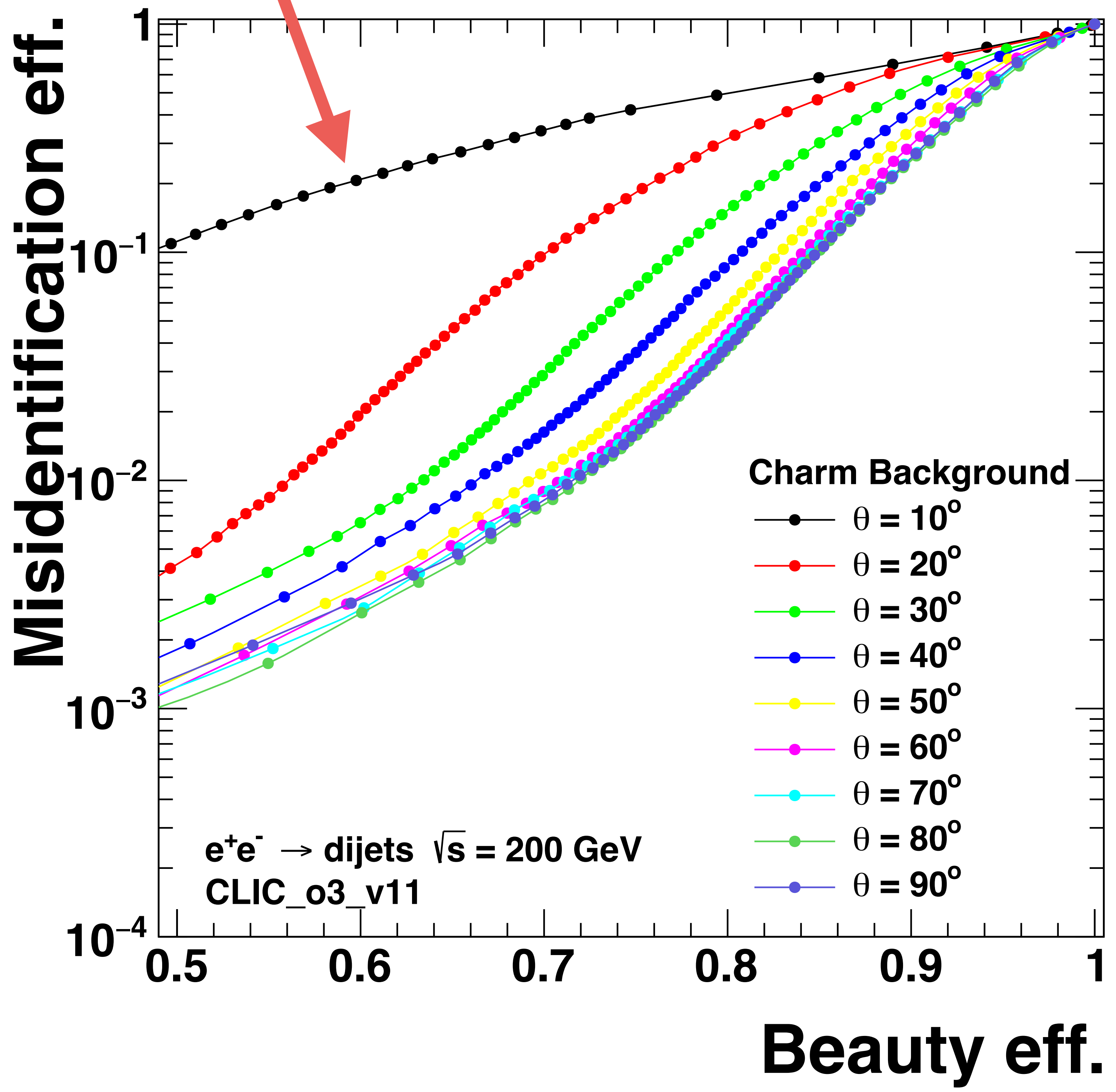
CLIC_SiD (DS)

CLICdet vs CLIC_SiD: jet-angle dependence

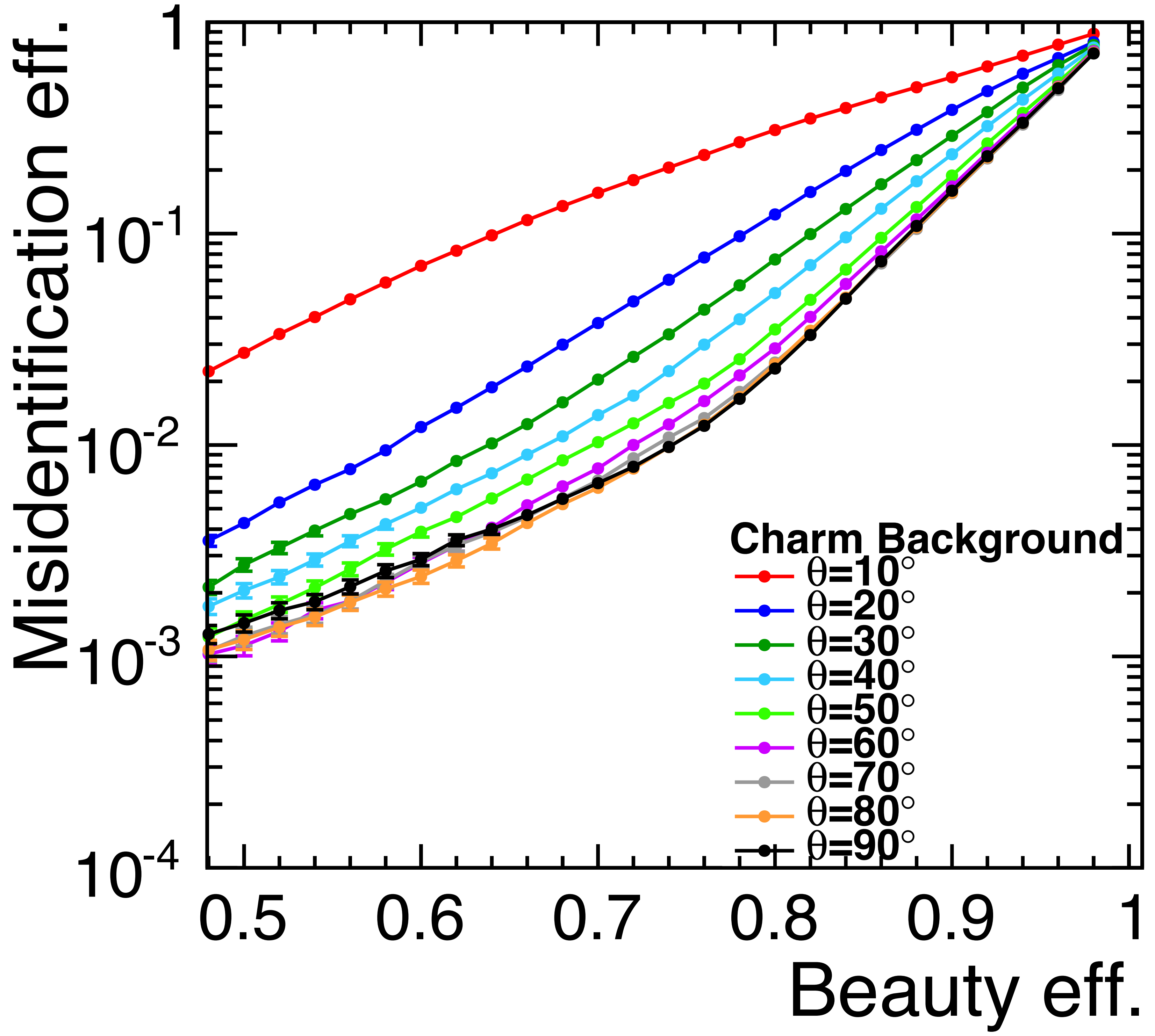
Dijet events $e^+e^- \rightarrow$ bb, cc, qq (q=uds)

NO γγ -> had. Overlaid

b-tagging performance almost an order of magnitude worse at 10°



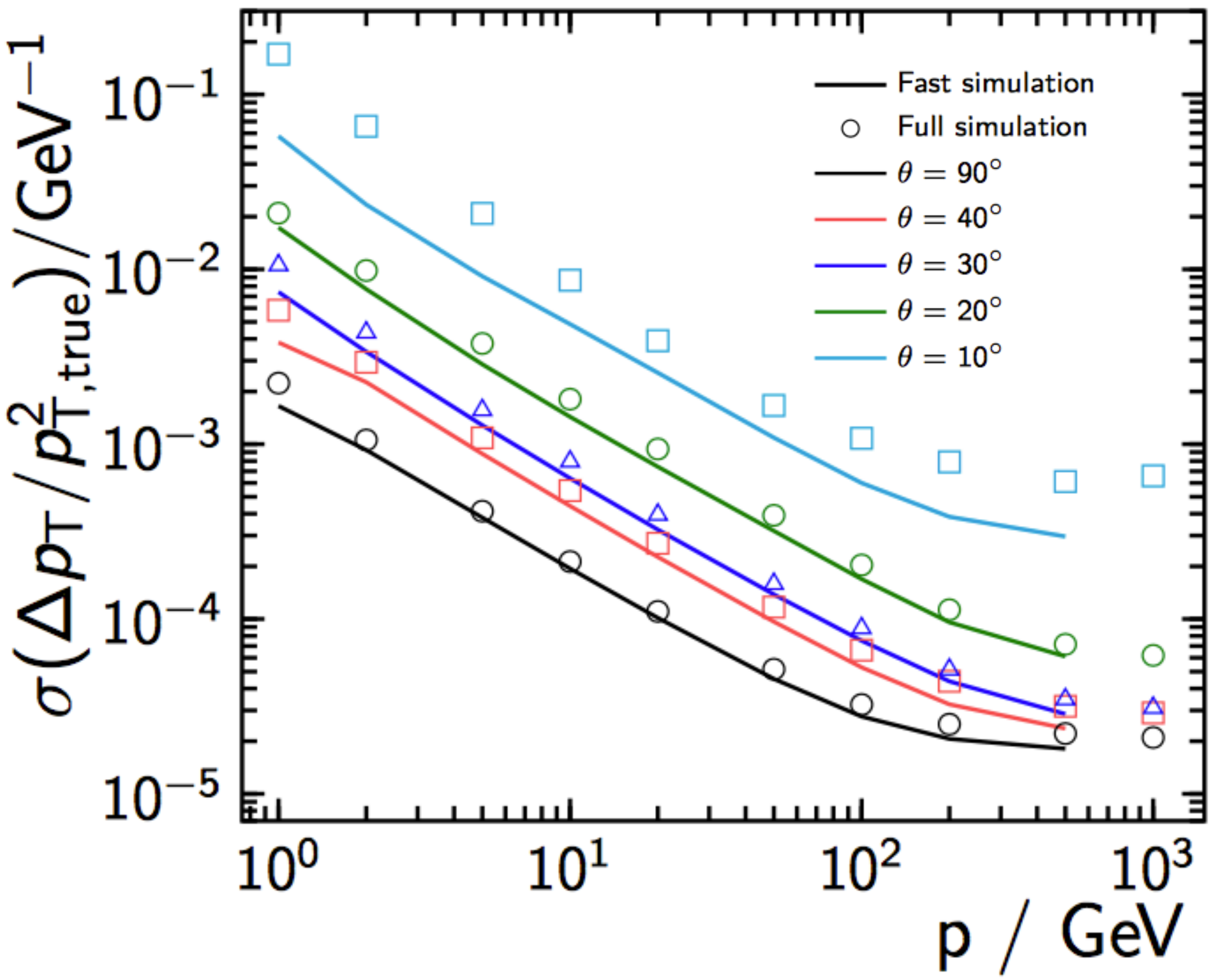
CLICdet



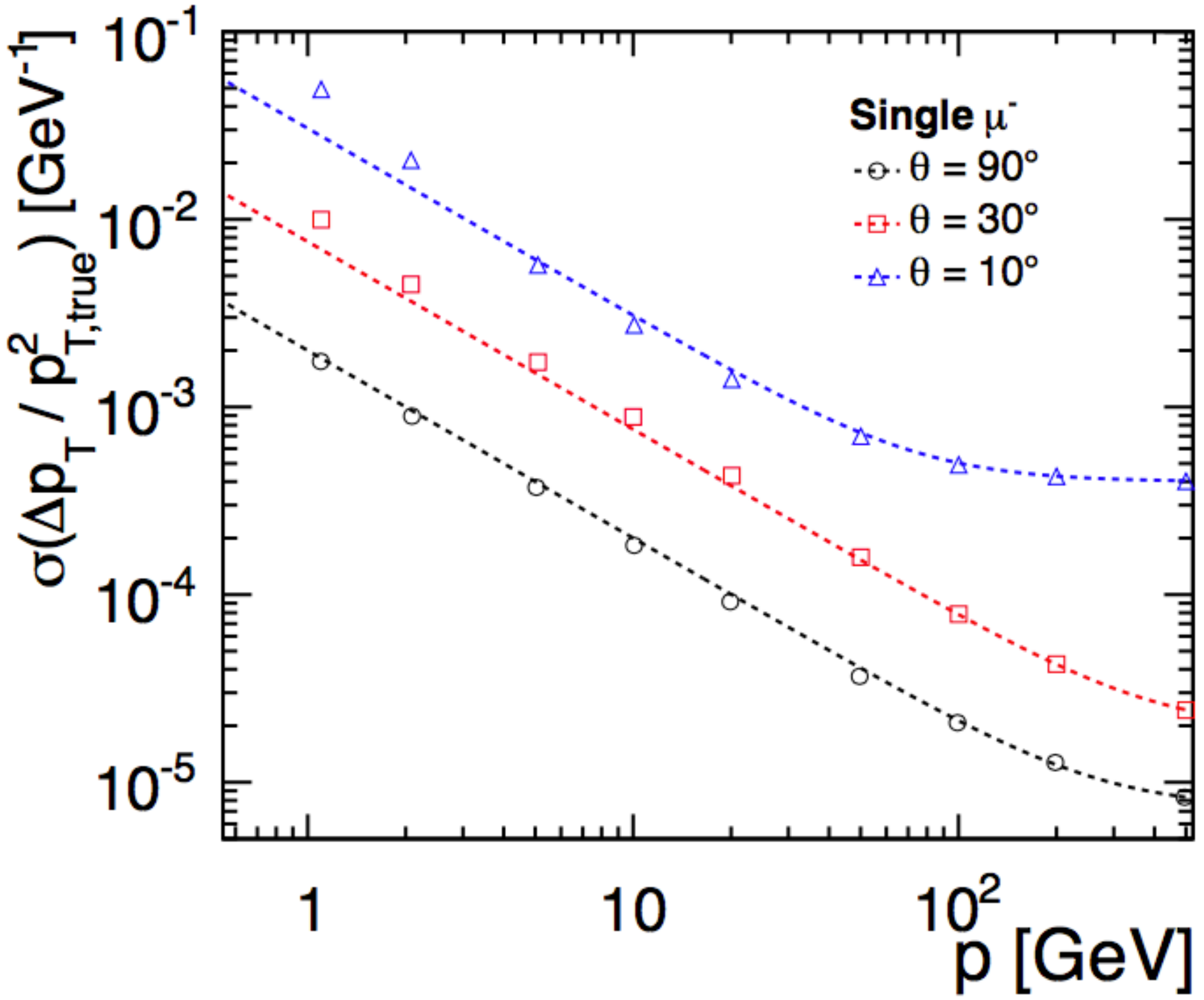
CLIC_SiD (DS)

CLICdet vs CLIC_SiD: p_T resolution

pT resolution up to a factor 4 better in CLIC_SiD for low momenta particles at 10°



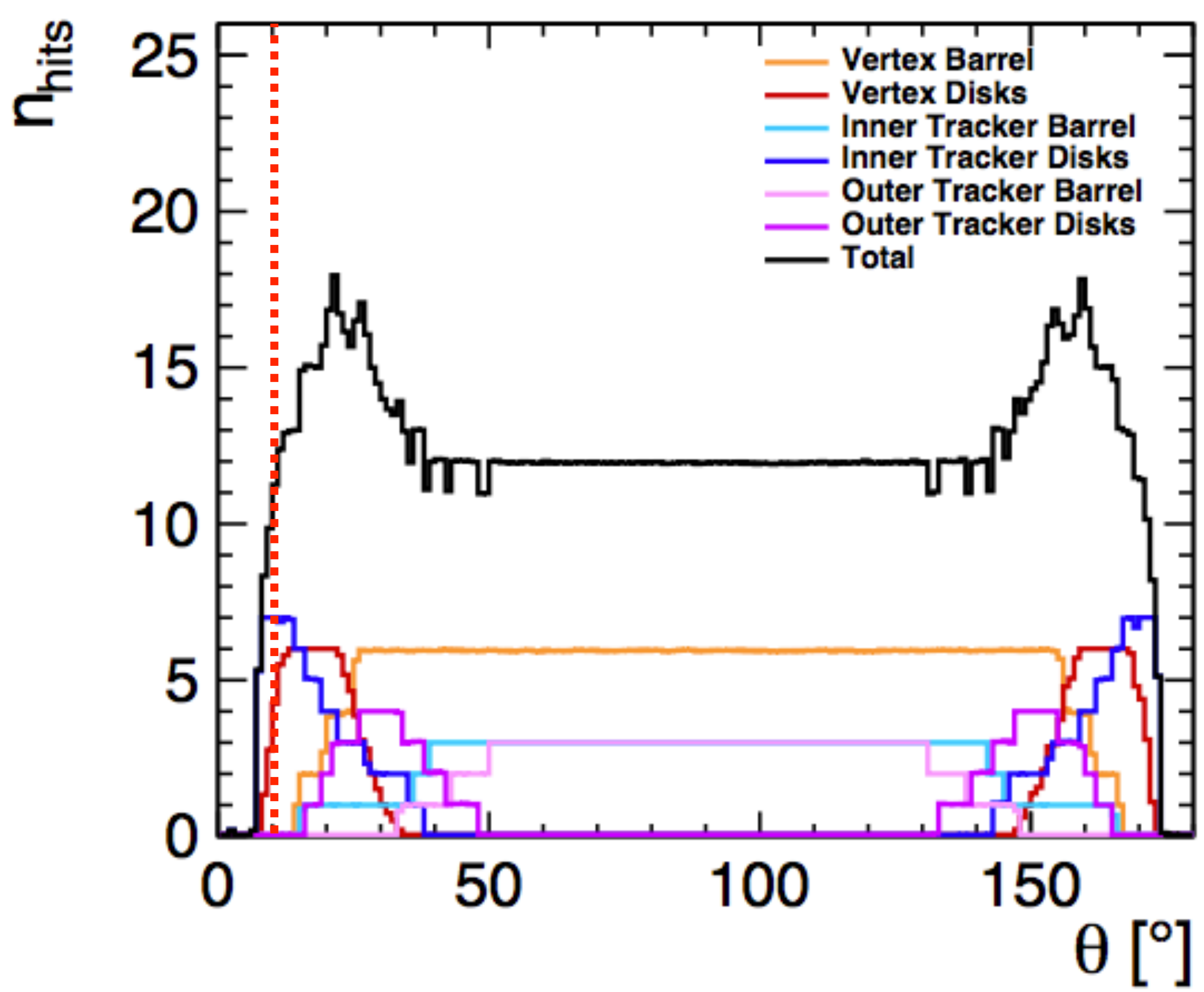
CLICdet



CLIC_SiD (DS)

n_{hits} for 10°

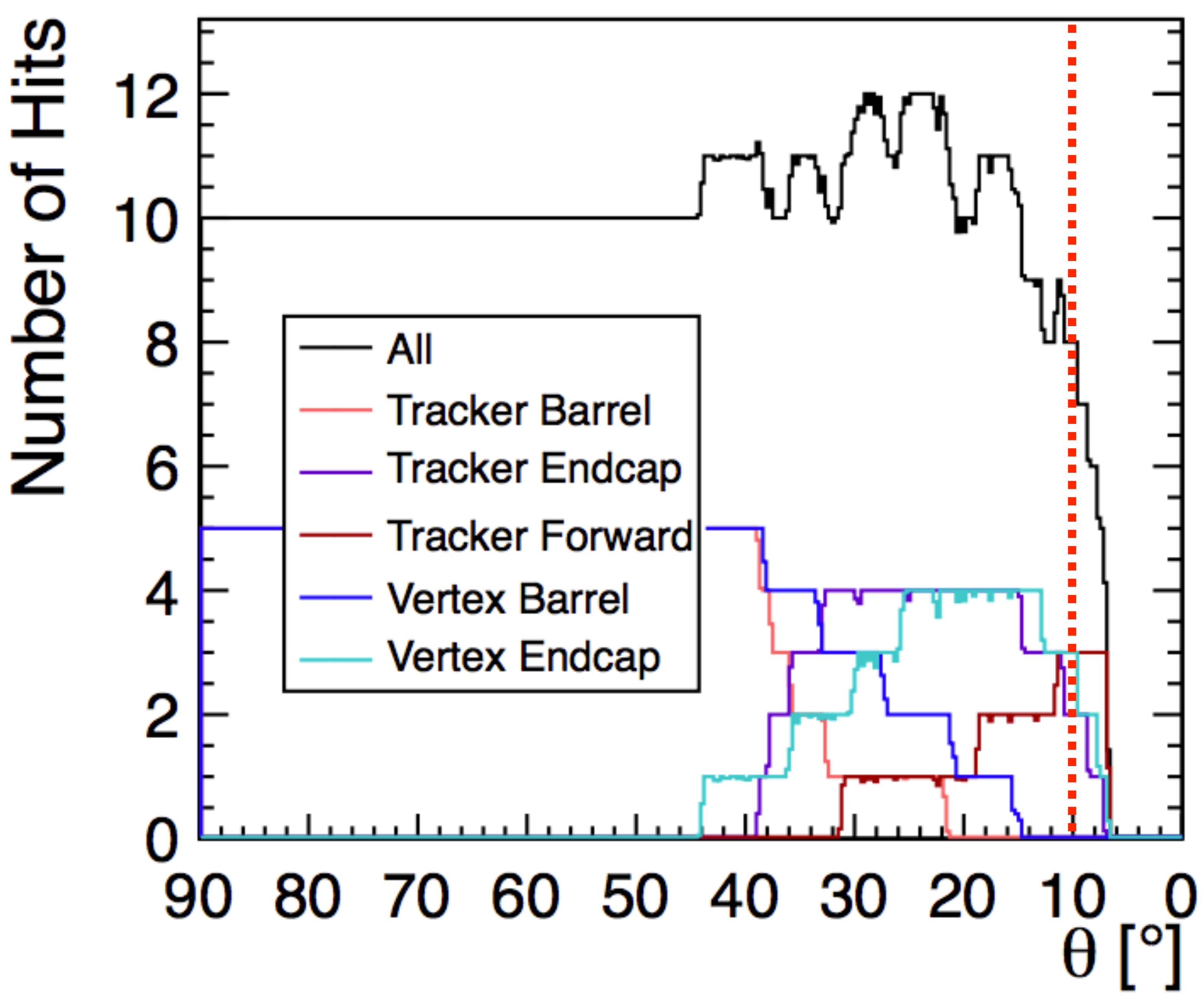
Vertex Disks: 4
Inner Tracker Disks: 7
Total: 11



CLICdet

n_{hits} for 10°

Tracker endcap: 2
Vertex endcap: 3
Tracker Forward: 3
Total: 8

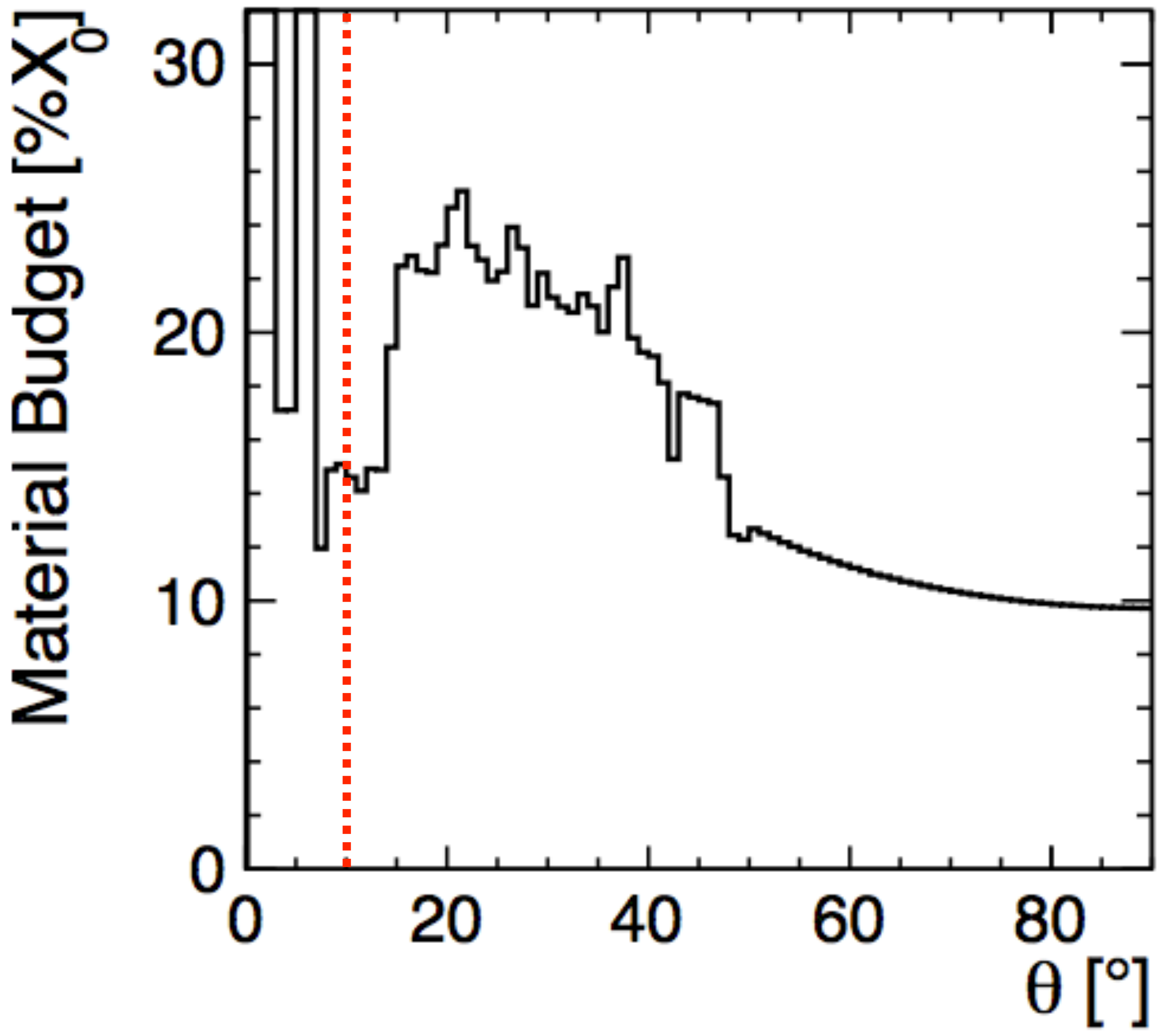


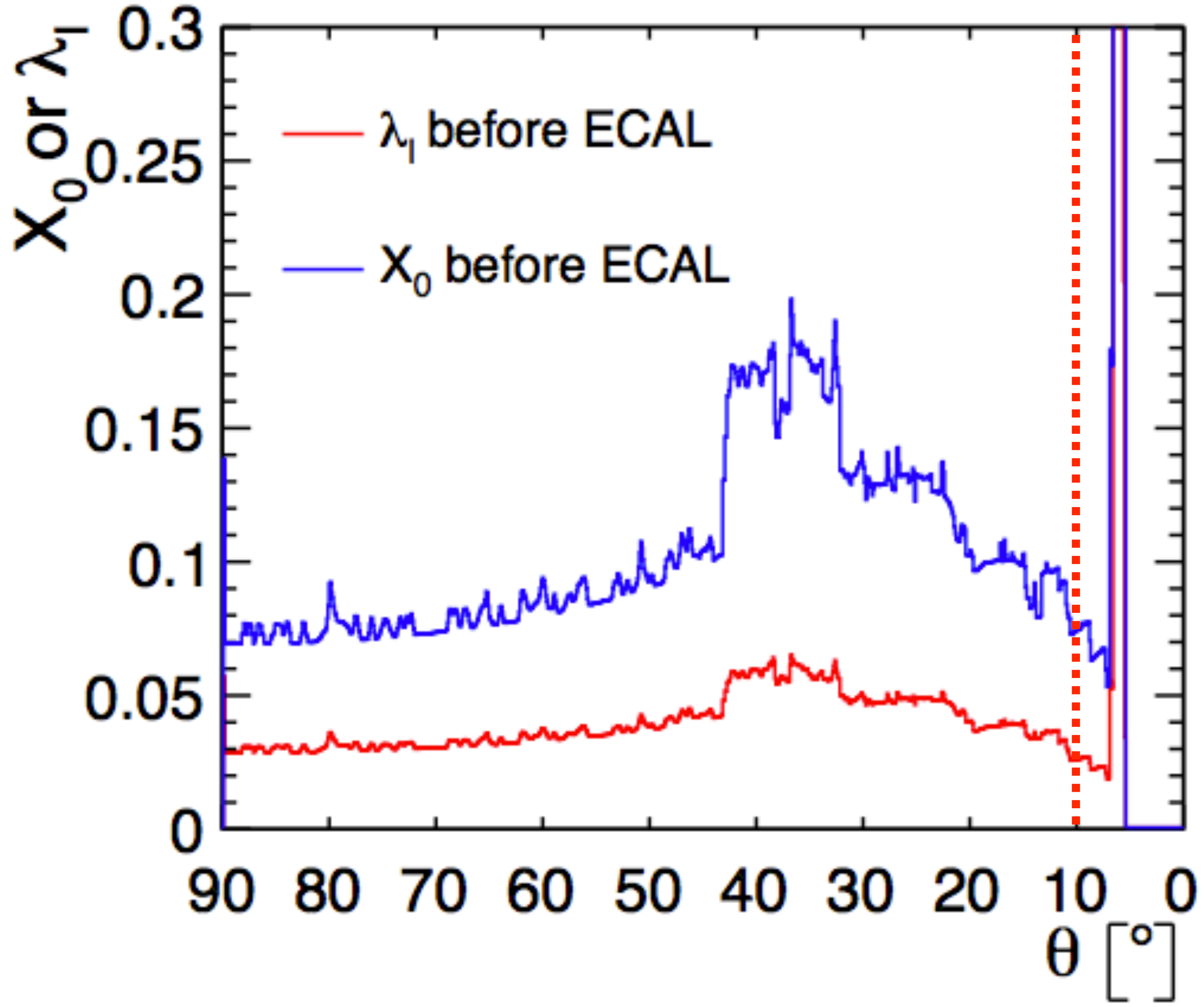
CLIC_SiD (DS)

CLICdet vs CLIC_SiD: Material Budget

$X_0(10^o) = 15\%$

$X_0(10^o) = 0{,}075$



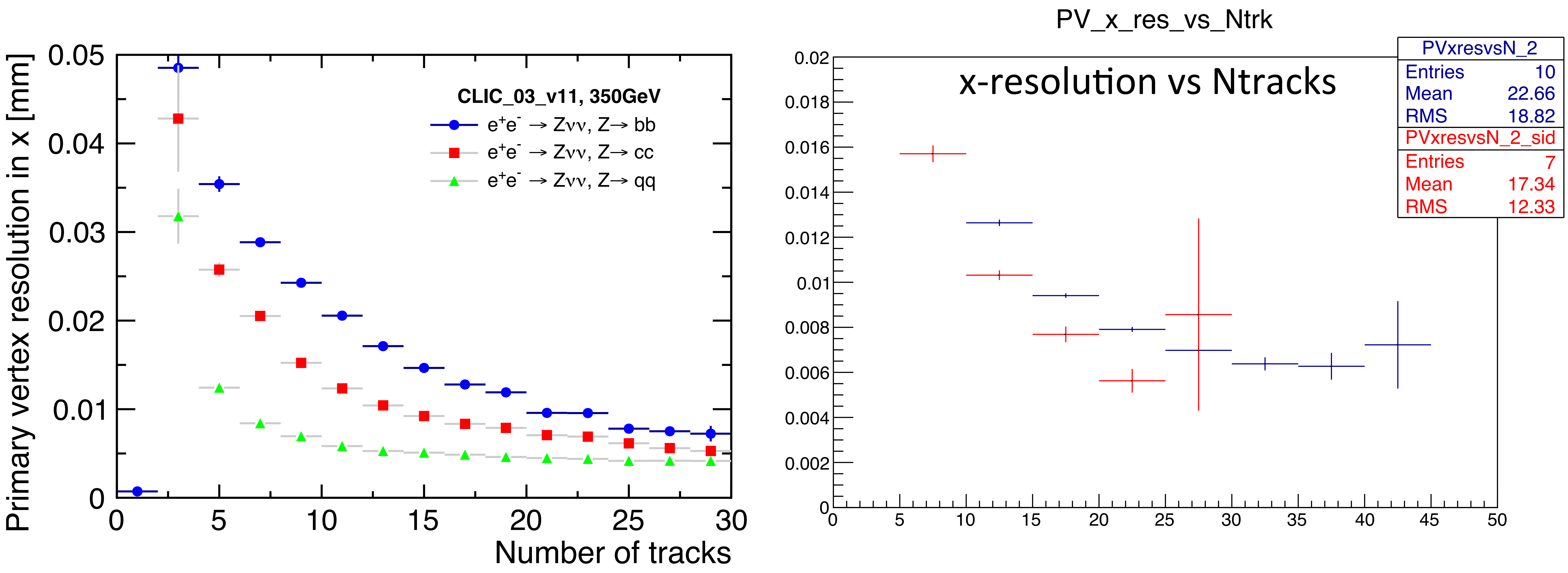


CLICdet

CLIC_SiD (DS)

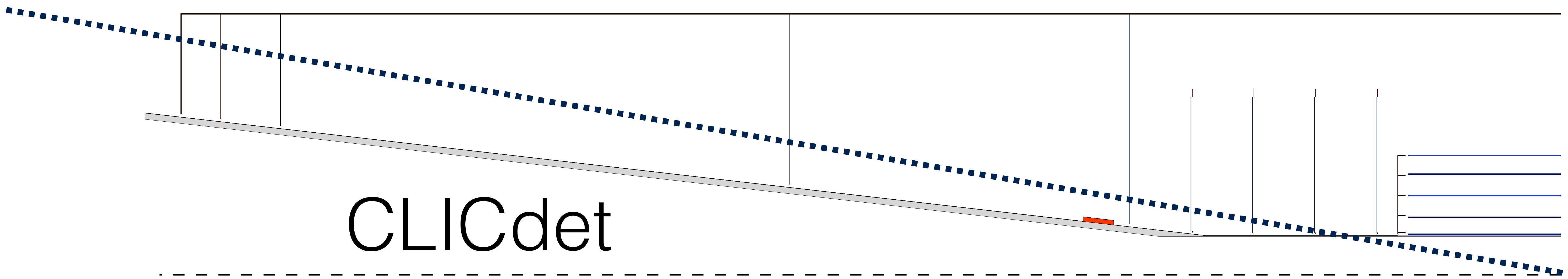
PV resolution

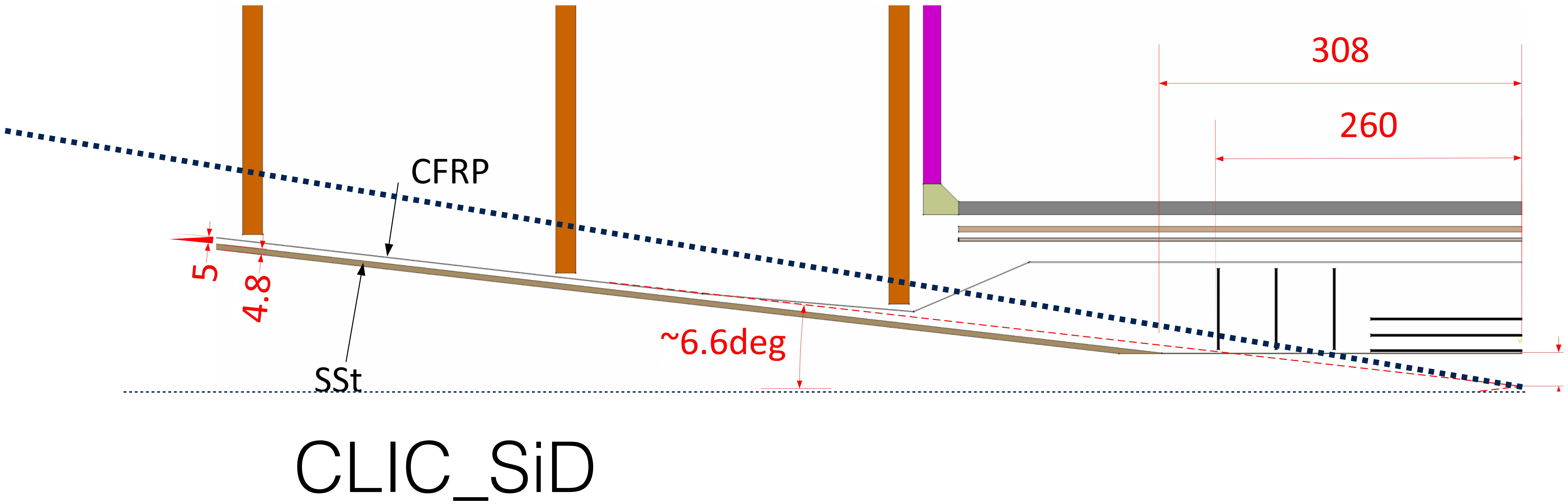
Twice better PV resolution for low number of tracks in CLIC_SiD



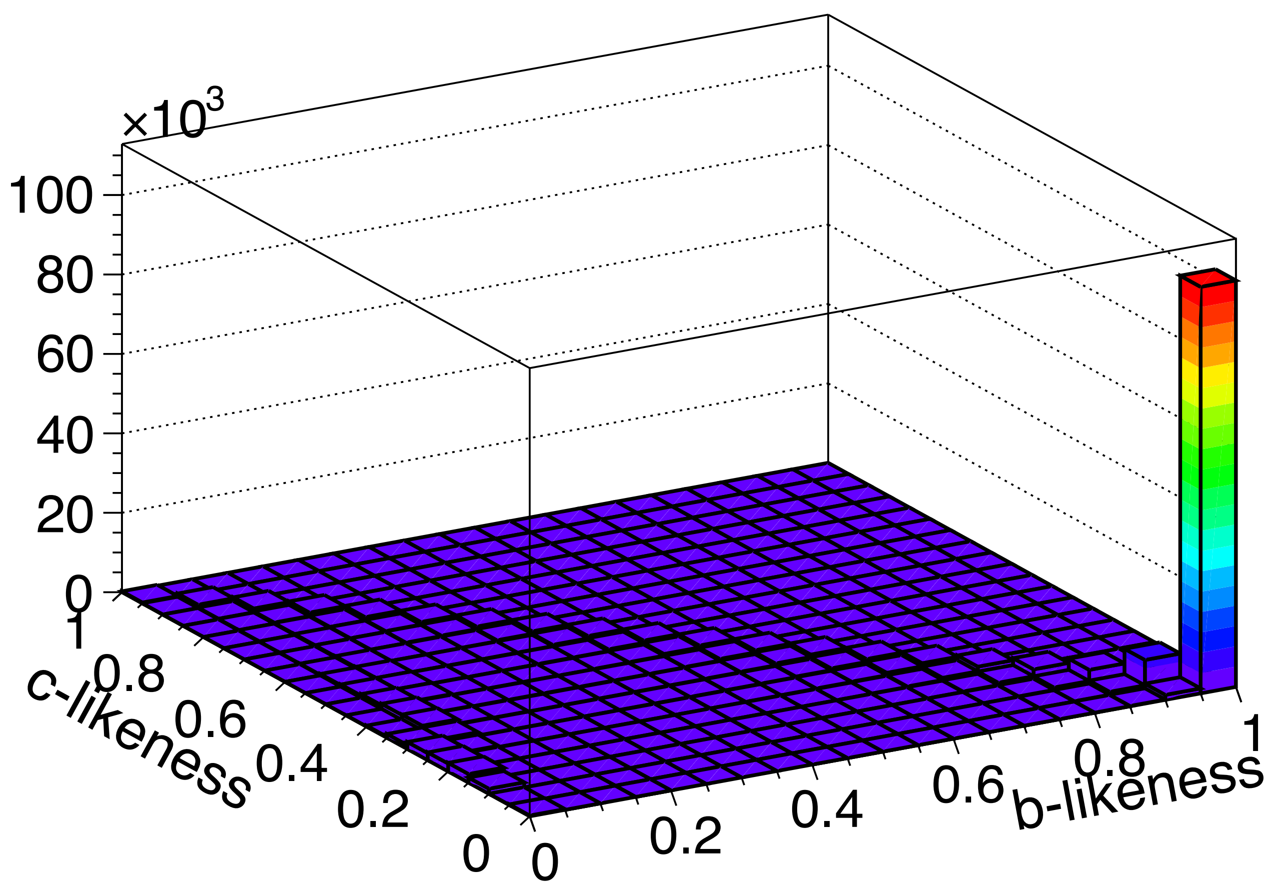
CLICdet **CLIC_SiD (DS)**

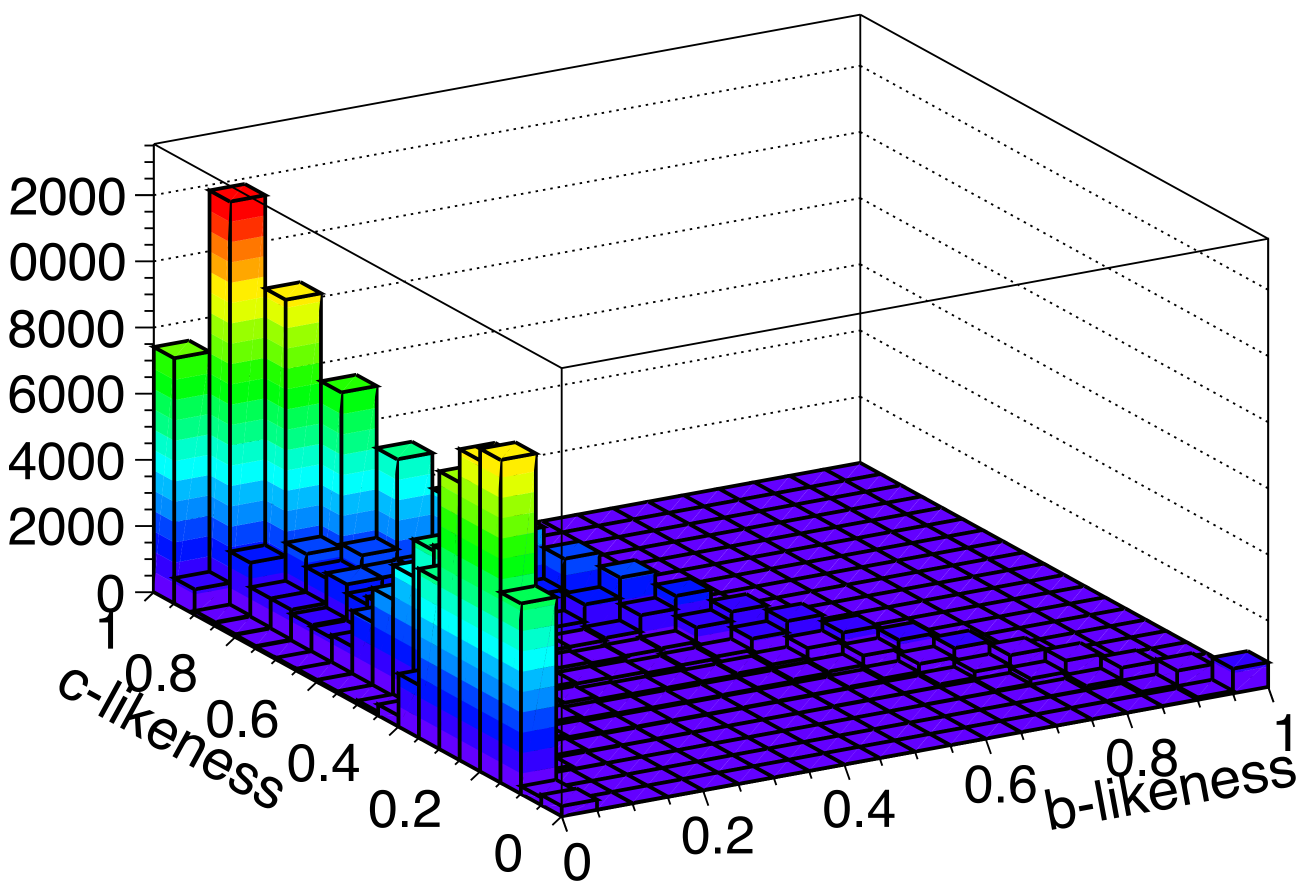
Tracker (XZ-view)

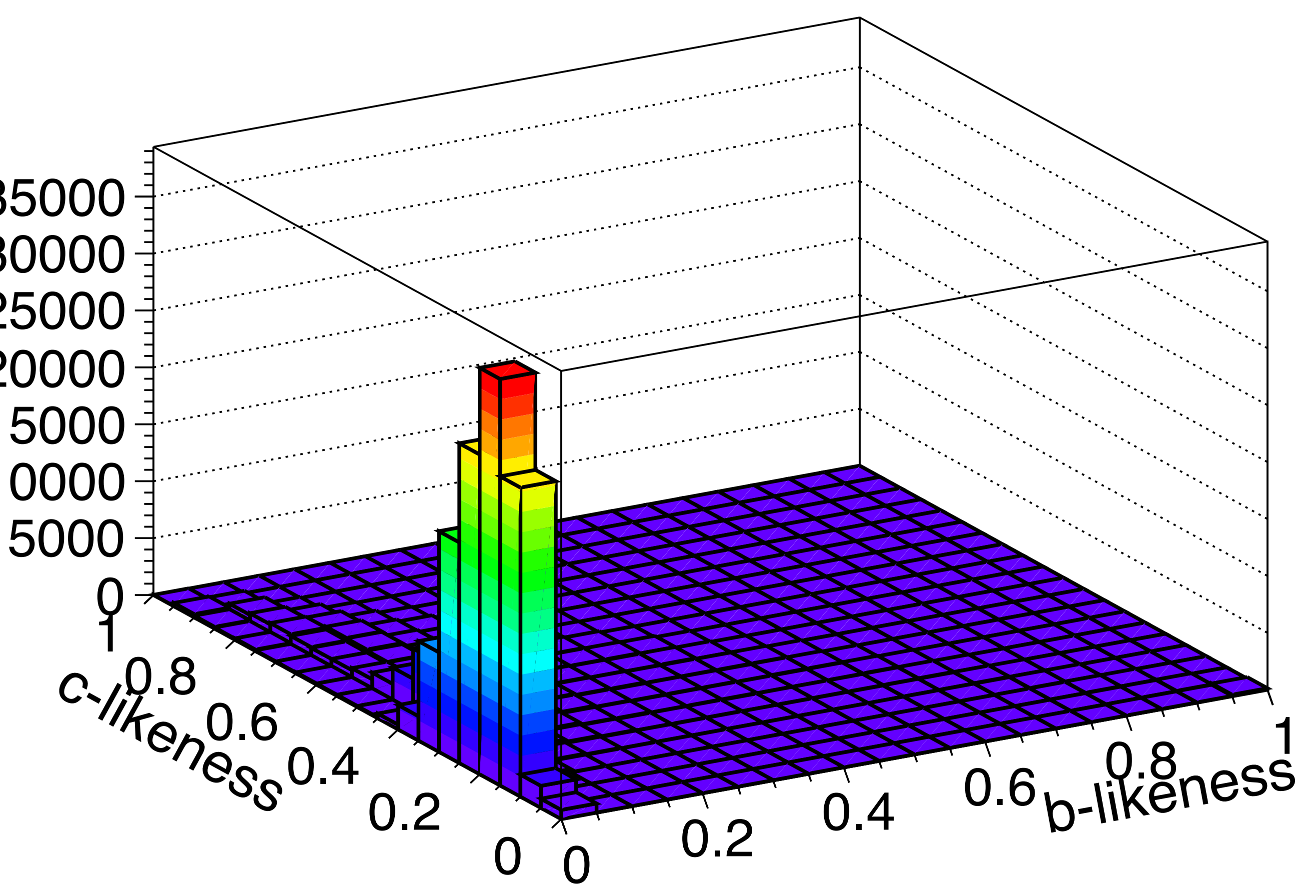




Application: Z boson hadronic decays







Z->bb **Z->cc** **Z->qq**