Optimization of preselection for bb/cc/ss studies at ILC250 (2000 fb^{-1}) with the ILD

Jesús Pedro Márquez Hernández
Adrian Irles Quiles
IFIC – CSIC – UV

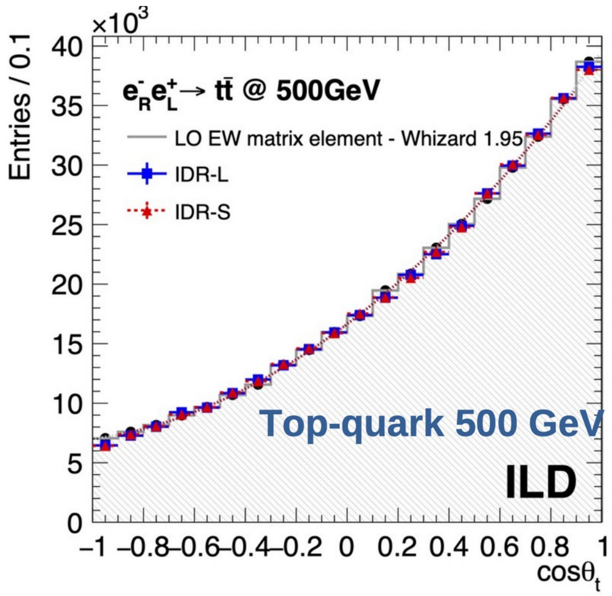
16/06/2021

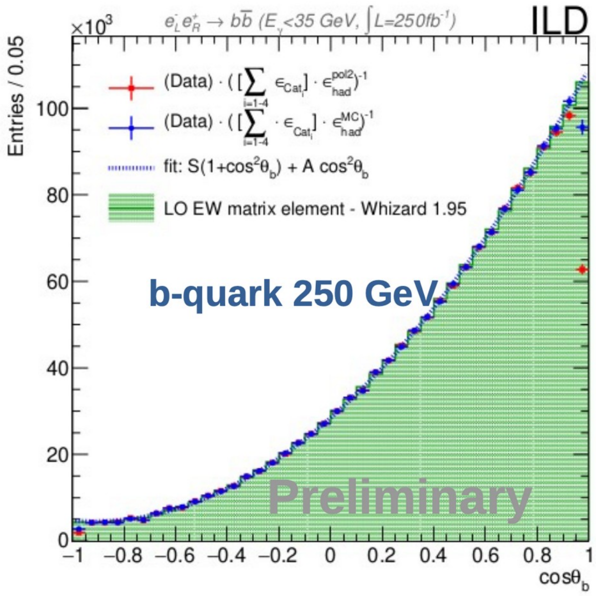
Realistic studies with full simulation

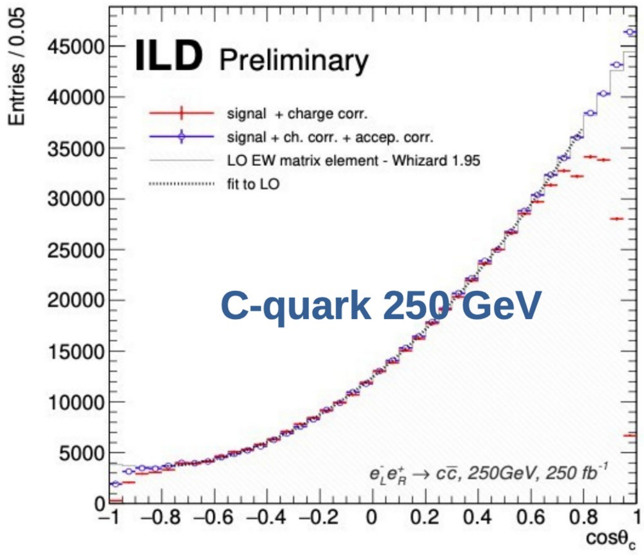
ILD Note
(Okugawa et al)

Preliminary Results
Presented in several conferences
(Irles, Poeschl, Richard)
Draft for publication in progress

Preliminary Results
Presented in LCWS19 (and others)
(Irles et al)
More detailed analysis in progress







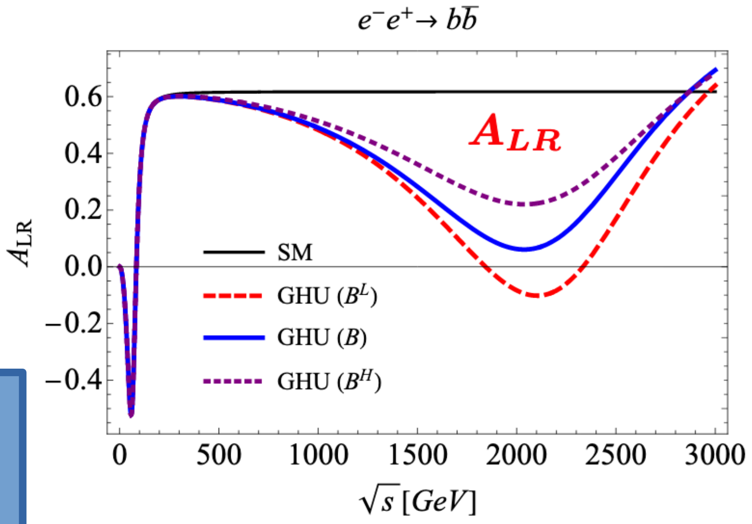
The study presented here defines the preselection for all new ss/cc/bb studies at √s = 250 GeV
and will define the starting point for √s=500 GeV samples

ss-quark study going on
(Okugawa)

Global strategy

Research team:
Instituto de Física Corpuscular (IFIC)
and
Irène Joliot-Curie Lab (IJCLab)

Potential for studying BSM physics
(e.g. Hosotani's GHU Model)



Funatsu et al (arXiv:2006.02157v3)

For a deep and complete study of the quark EW couplings (SM and BSM) we need to study different flavors and run at different energies (√s=91, 250, 500 GeV)

My study: Introduction

Data analysis of ILC simulated data:

- Final goal: Studying b-quark EW couplings.
  - Observables: R_b^{cont}, $\frac{d\sigma}{d\cos\theta_b}$ and $A_{FB}^{b\bar{b}}$
- Data processed with the VLC algorithm:
  - Pfos are ordered in 2 jets
    - For the signal we expect the two jets in back-to-back topology (but not for the background)
- Most of the background is radiative return
  - And most of the data is background! (x3 for $e^-_L e^+_R$ and x6 for $e^-_R e^+_L$)
- Event preselection procedure:
  - Cut in radiative returns
  - Cut in background from pair of heavy bosons

We need a preselection with homogeneous efficiency in the volume of the detector and with minimal flavor dependence: to avoid modeling uncertainties (see, for example, Irles talk)

Results before 2020 (I)

Results from the DBD samples:

- Cuts:
  - K_{reco} < 35 GeV & m_{2jets} > 130 GeV
  - N pfos > 5
  - Photo vetoing (100, 70)
  - y_{23} < 0.015
  - $m_{j1} + m_{j2}$ < 100 GeV
- Efficiencies:

Efficiency (%)			
$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
71.5	71.1	70.7	0.7

Table 1: Total efficiency of the preselection for the different quark flavors and radiative return (DBD samples)

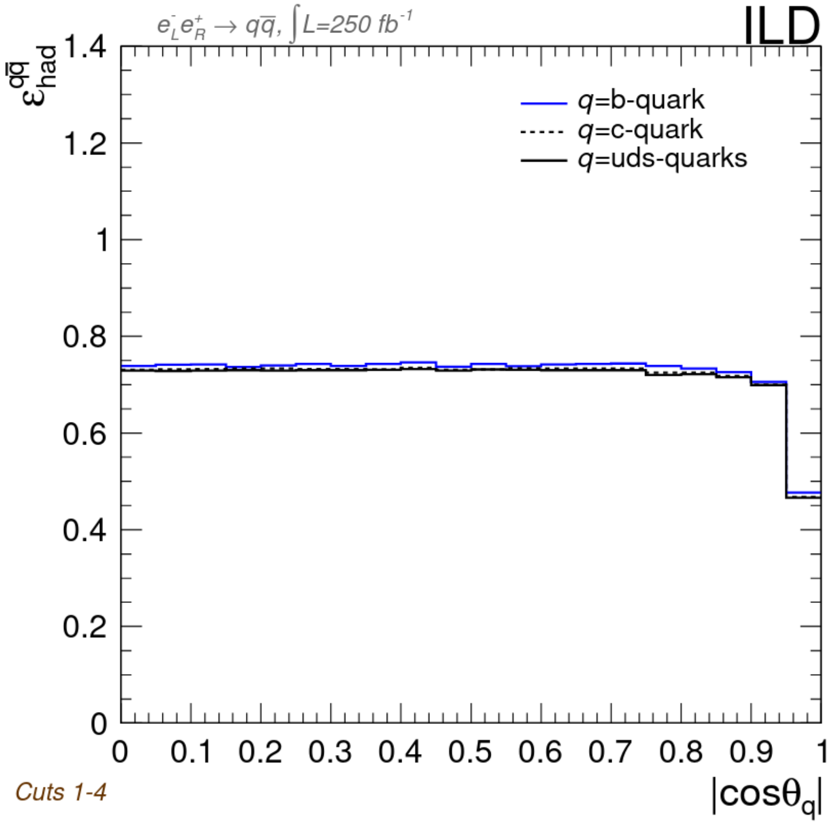


Fig. 1: Efficiency of the preselection for the different quark flavors vs the angular distribution of the two jet system (DBD samples)

Results before 2020 (I)

- Aim:
  - Obtaining maximal, uniform and flavor non-dependent efficiency
  - Lowest background possible

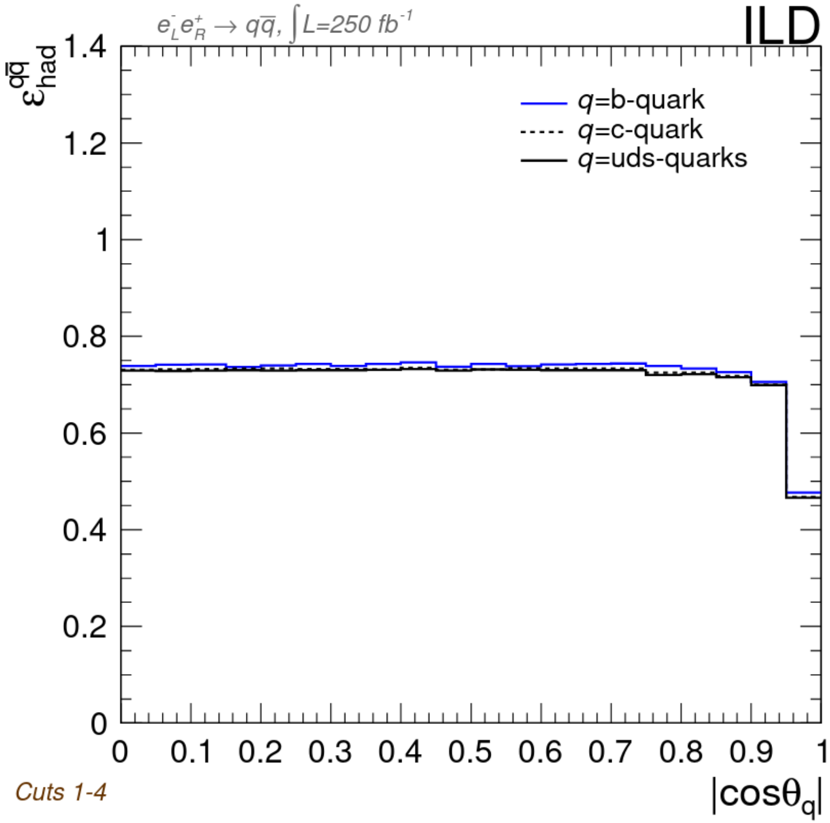


Fig. 1: Efficiency of the preselection for the different quark flavors vs the angular distribution of the two jet system (DBD samples)

New samples

Results from the new samples applying DBD designed cuts:

- Cuts:
  - $K_{reco} < 35$ GeV & $m_{2jets} > 130$ GeV
  - N pfos > 5
  - Photon veto
  - $y_{23} < 0.015$
  - $m_{j1} + m_{j2} < 100$ GeV
- Efficiencies:

Efficiency (%)			
$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
71.5	69.9	68.4	0.7

Table 2: Total efficiency of the preselection for the different quark flavors and radiative return (new samples applying DBD designed cuts)

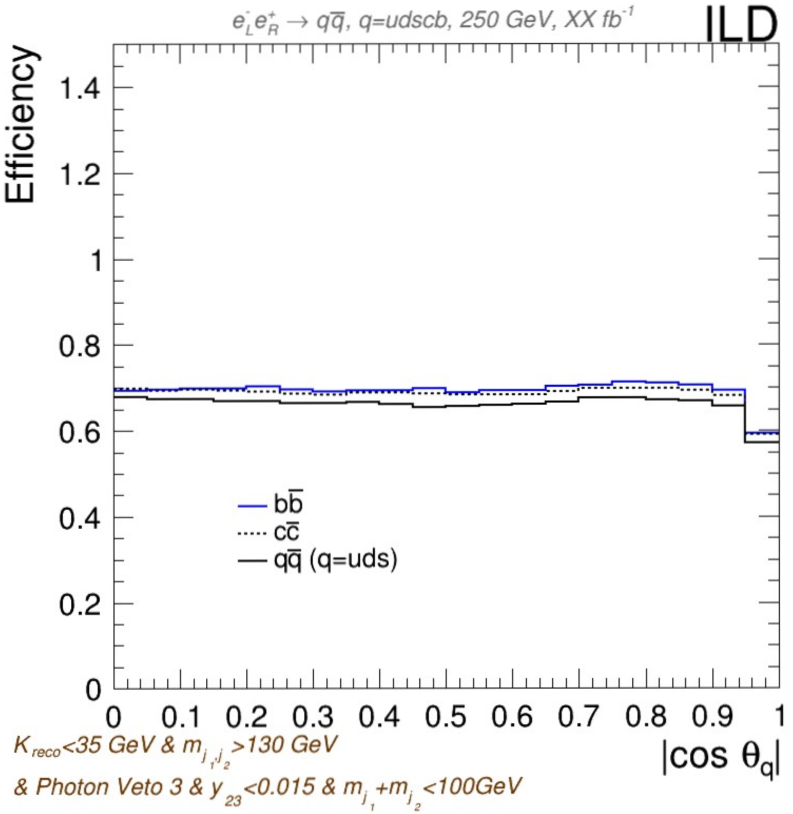


Fig. 2: Efficiency of the preselection for the different quark flavors vs the angular distribution of the two jet system (new samples applying DBD designed cuts)

New samples

Results from the new samples applying DBD designed cuts::

- Cuts:
  - K_{reco} < 35 GeV & m_{2jets} > 130 GeV
  - N pfos > 5
  - Photon veto
  - y_{23} < 0.015
  - $m_{j1} + m_{j2}$ < 100 GeV
- Efficiencies:

This preselection creates differences between flavors

E[illegible]%)			
$b\bar{b}$	c[illegible]	$q\bar{q}(uds)$	ISR
71.5	69.9	68.4	0.7

Table 2: Total efficiency of the preselection for the different quark flavors and radiative return (new samples applying DBD designed cuts)

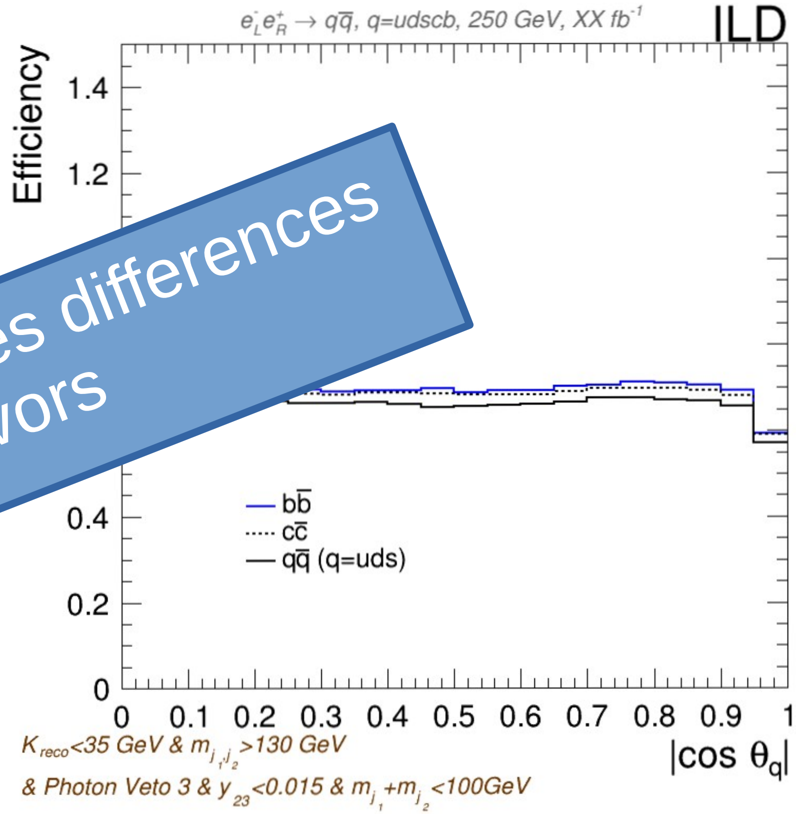


Fig. 2: Efficiency of the preselection for the different quark flavors vs the angular distribution of the two jet system (new samples applying DBD designed cuts)

New samples (II)

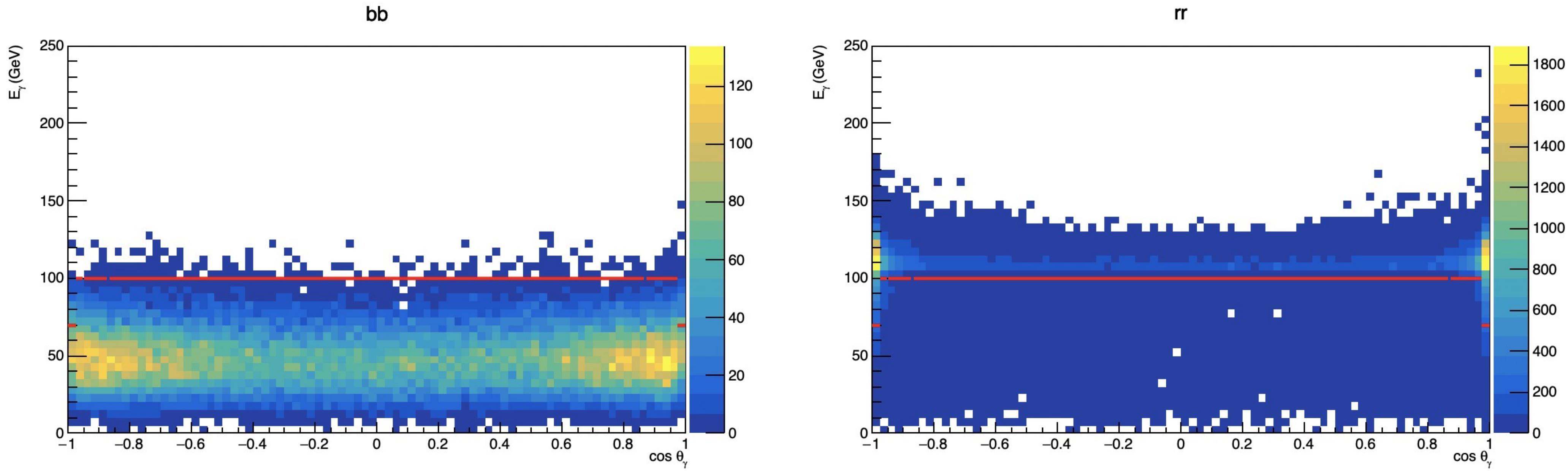


Fig. 3: Two dimensional maps used for the photon vetoing; impact in different quark flavors (new samples). Left plot is for b-quark events while the right plot is for ISR. The y-axis is the energy (GeV) of the pfos identified as photons by the jet reconstruction algorithm and the x-axis is cos(θ) of those pfos.

IFIC INSTITUT DE FÍSICA CORPUSCULAR

New samples (III)

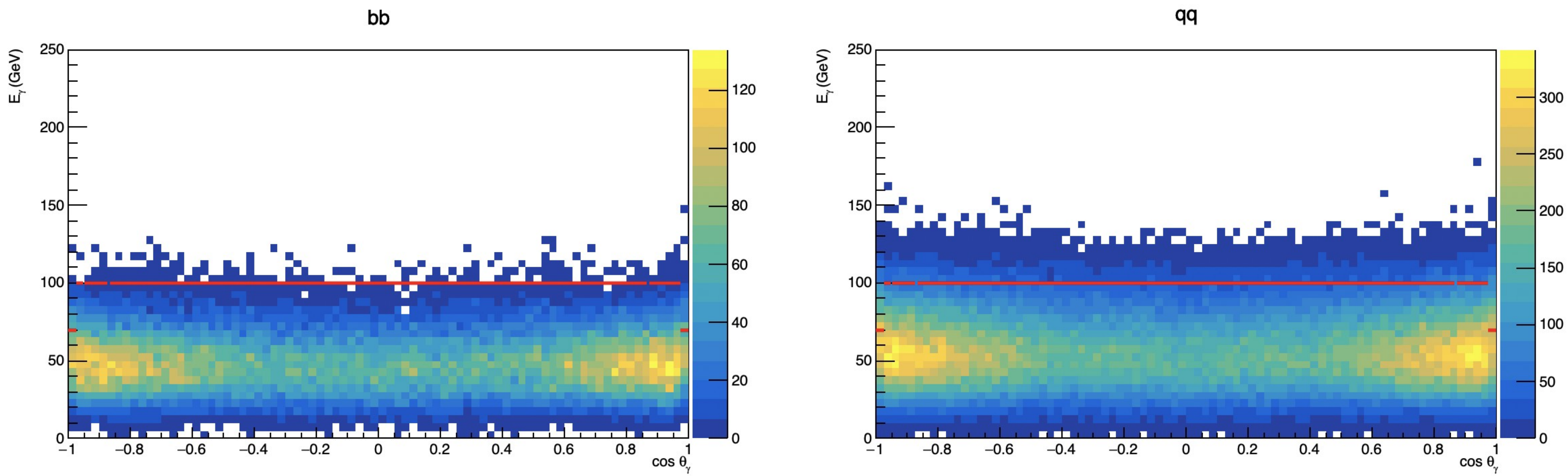


Fig. 4: Two dimensional maps used for the photon vetoing; impact in different quark flavors (new samples). Left plot is for b-quark events while the right plot is for uds-quarks. The y-axis is the energy (GeV) of the pfos identified as photons by the jet reconstruction algorithm and the x-axis is cos(θ) of those pfos.

If we try to re-balance the selection efficiencies by leveling up these energies levels then the rejection efficiency for the radiative return goes all the way up to ~2%

Next step:
Re-optimization of the cuts

Re-optimization of the cuts: K_{reco}

- K_{reco} is a good estimator of E_γ:
  - Definition of acolinearity:

$$\sin\Psi_{acol} = \frac{\vec{p_{j_1}} \times \vec{p_{j_2}}}{|\vec{p_{j_1}}| \cdot |\vec{p_{j_1}}|}$$

  - Momentum of the collinear photon in the ultrarrelativistic limit ($m_{jets} << p_{jets}$):

$$|\vec{k}| \approx K_{reco} = \frac{250\,\mathrm{GeV} \cdot \sin\Psi_{acol}}{\sin\Psi_{acol} + sin\theta_1 + sin\theta_2}$$

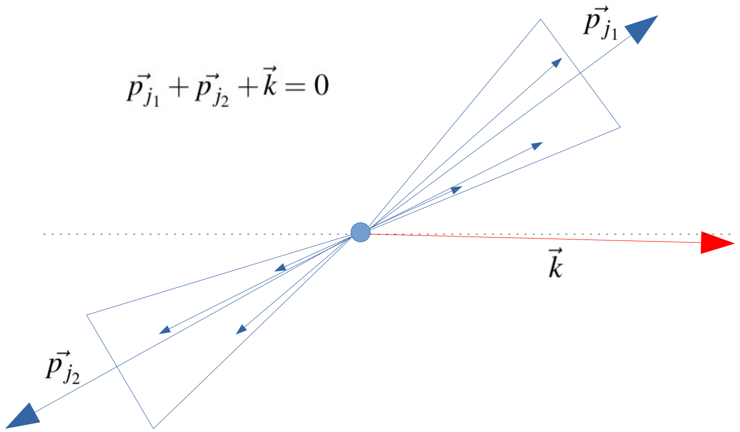


Fig. 5: Kinematics of a two jets system reconstruction with ISR

Re-optimization of the cuts: K_{reco} (II)

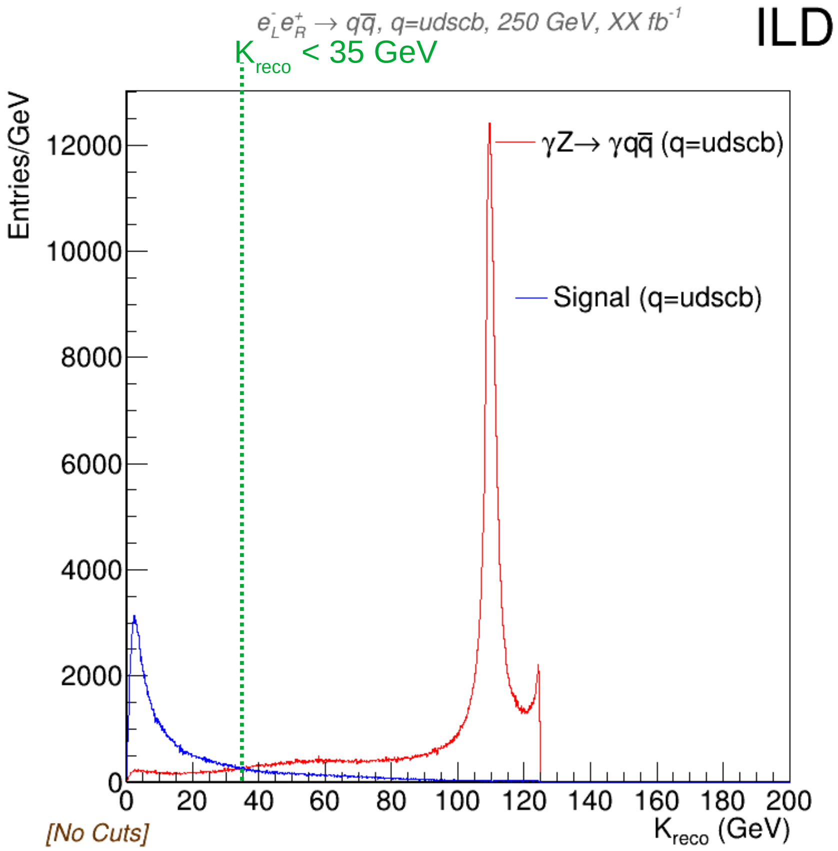


Fig. 6: Distribution of events vs K_{reco} . The green vertical lines represent the different cuts tried for K_{reco}

We set the cut to K_{reco} < 35 GeV

IFIC INSTITUT DE FÍSICA CORPUSCULAR

Re-optimization of the cuts: m_{2jets}

m_{2jets} is the sum of the masses of both jets

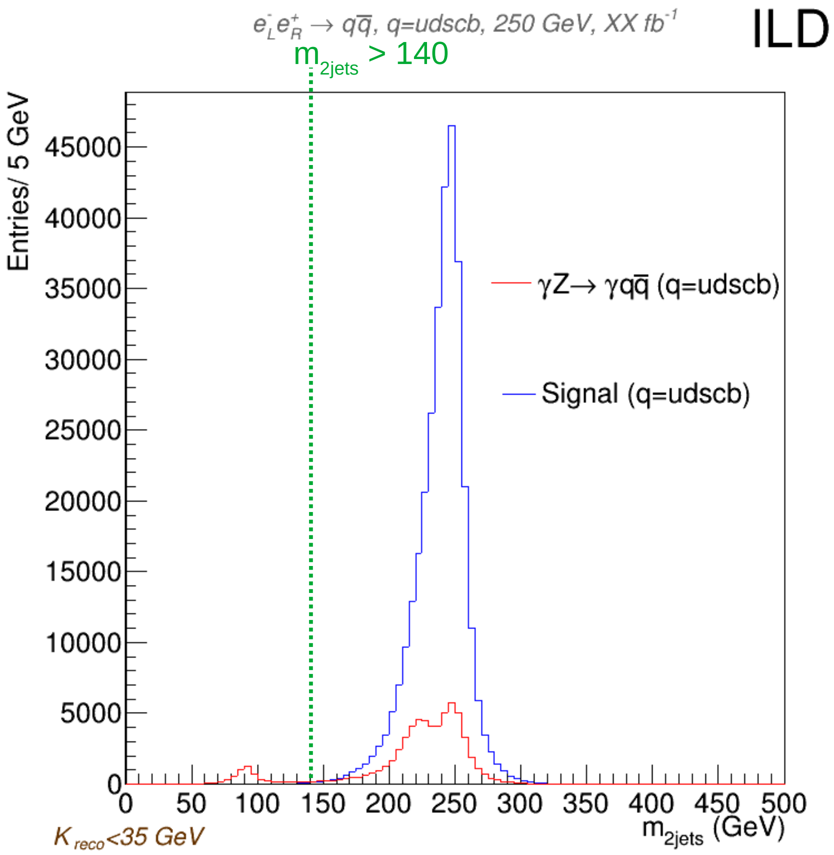
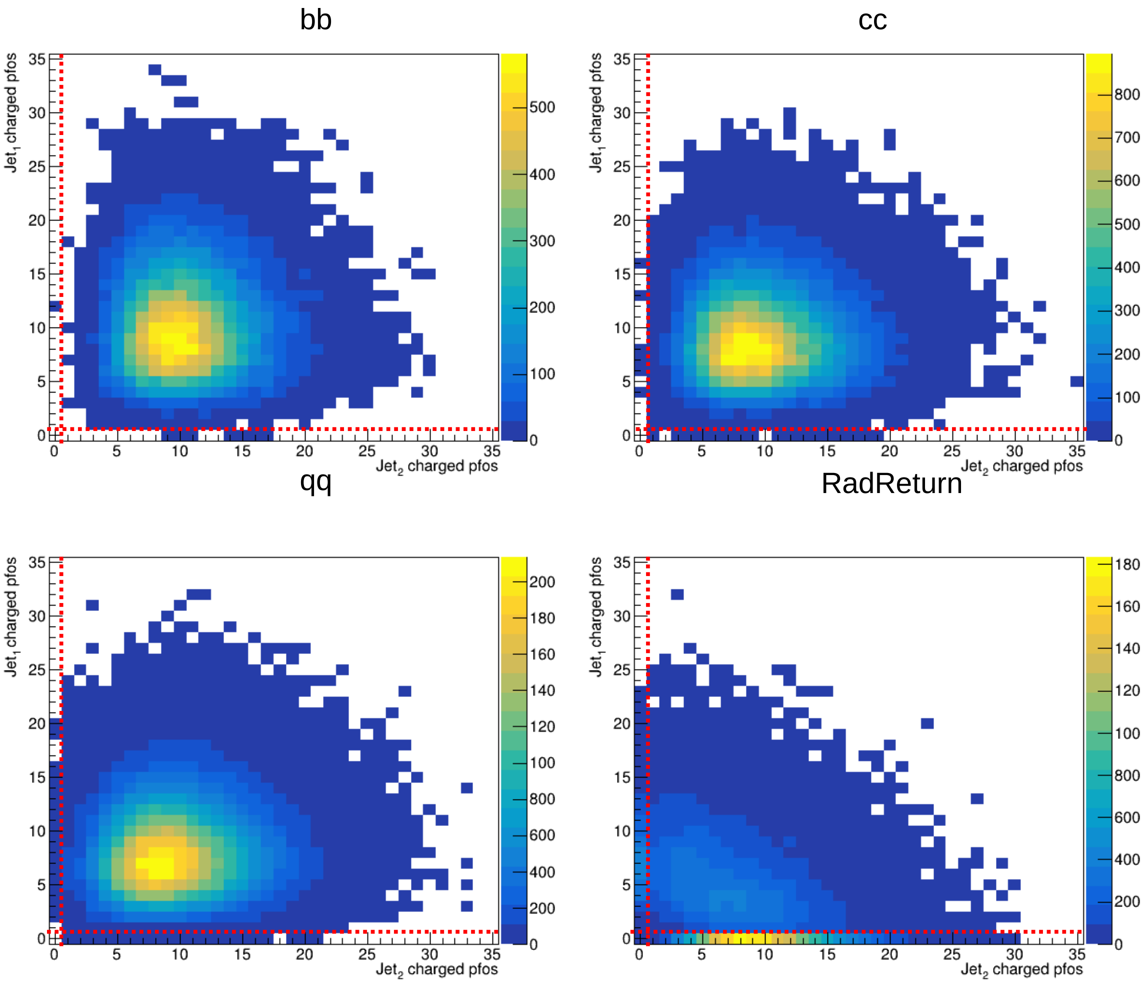


Fig. 7: Distribution of events vs m_{2jets}. The green vertical line represents the selected cut for m_{2jets}

We set the cut to m_{2jets} > 140 GeV

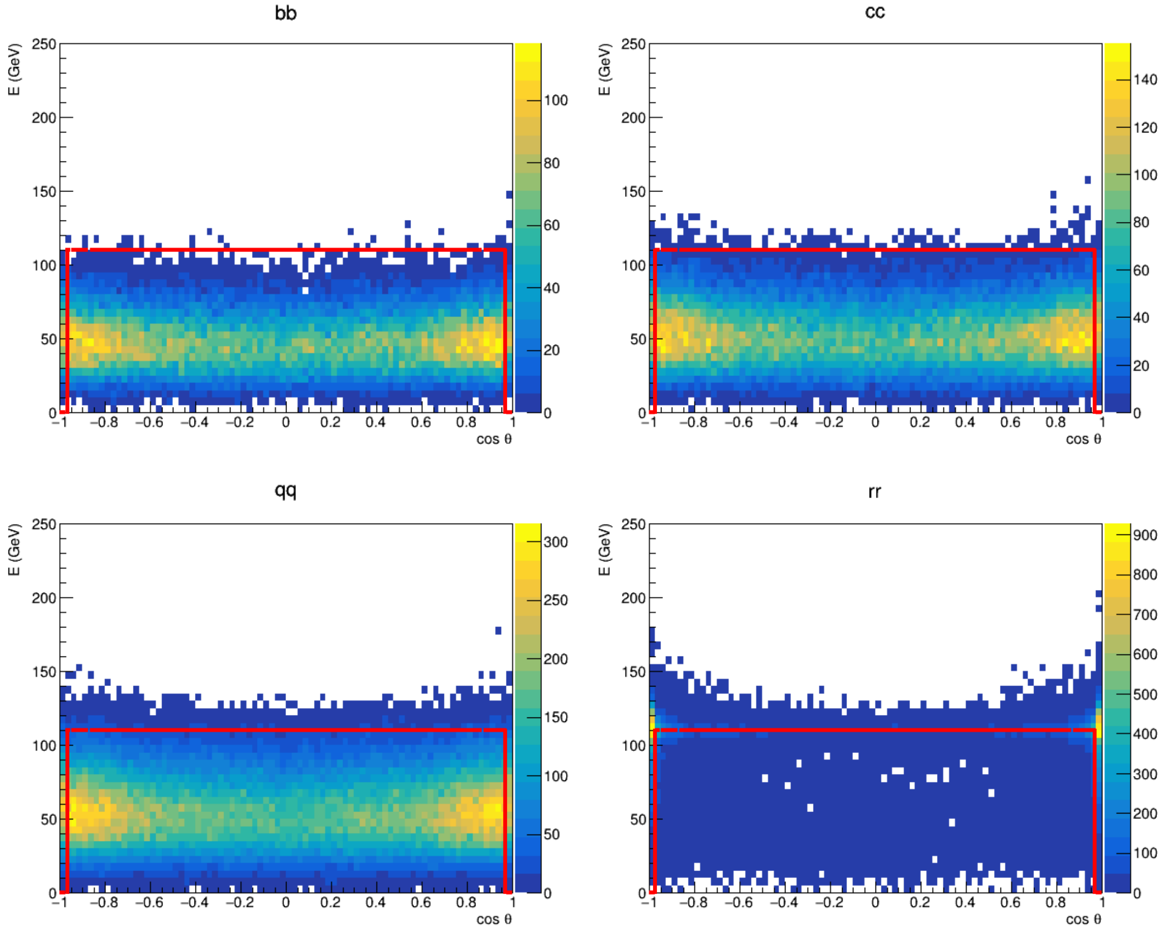
Re-optimization of the cuts: Charged PFOs



At least 1 charged pfo in each jet

Fig. 8: Selected cut for the number of charged pfos associated to each jet

Re-optimization of the cuts: Photon veto



E > 110 GeV
and
|cosθ|<0.97

Fig. 9: Selected cut for the angular distribution of energy (photon vetoing)

Re-optimization of the cuts: Overview

Previous approach:

- Cuts:
  - K_{reco} < 35 GeV & m_{2jets} > 130 GeV
  - N pfos > 5
  - Photon veto (DBD cuts)
  - y_{23} < 0.015
  - $m_{j1} + m_{j2}$ < 100 GeV
- Efficiencies:

Efficiency (%)			
$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
71.5	69.9	68.4	0.7

Table 3: Total efficiency of the preselection for the different quark flavors and radiative return with the previous cut selection (new samples)

New approach:

- Cuts:
  - K_{reco} < 35 GeV
  - m_{2jets} > 140 GeV
  - Charged pfos > 1
  - Photon veto (previous slide)
  - Y_{23} < 0.015 (we kept this unchanged)
- Efficiencies:

Efficiency (%)			
$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
68.4	68.4	68.0	1.2

Table 4: Total efficiency of the preselection for the different quark flavors and radiative return with the new cut selection (new samples)

Better but not good enough :/

After optimizing the cuts, we tried to optimize the performance of the jet reconstruction algorithm

Re-optimization of the cuts: VLC algorithm

- Definitions

$$d_{ij} = 2\min(E_i^{2\beta}, E_j^{2\beta})(1-\cos\theta_{ij})/R^2,$$

$$d_{iB} = E_i^{2\beta}\sin^{2\gamma}\theta_{iB},$$

- We played with the parameters R and γ:
  - R: (0.50, 0.96, 1.00, 1.05, 1.17, 1.28, 1.36, 1.68)
  - γ : (0.0, 0.5, 1.0)

- We kept β=1

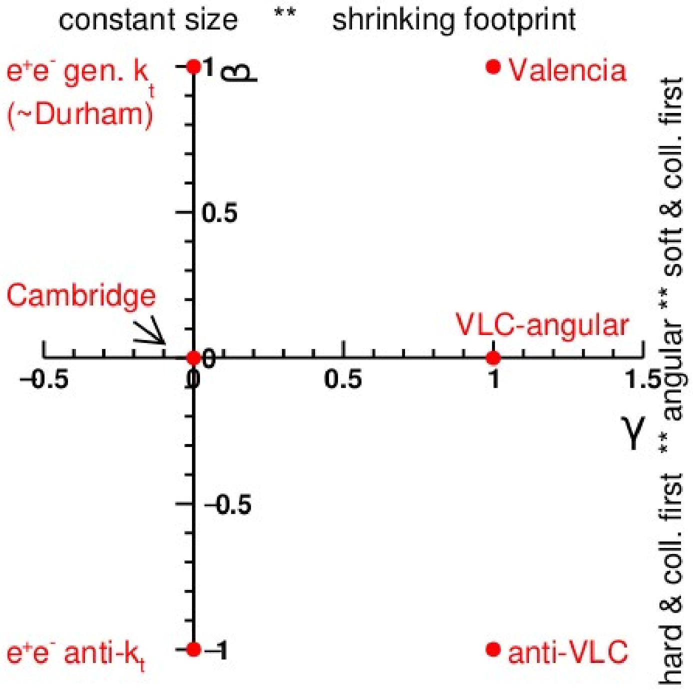


Fig. 10: Diagram of the parameter space spanned by exponents γ and β. [1]

[1] M. Boronat et al. (arXiv:1607.05039)

Re-optimization of the cuts: VLC algorithm – R (I)

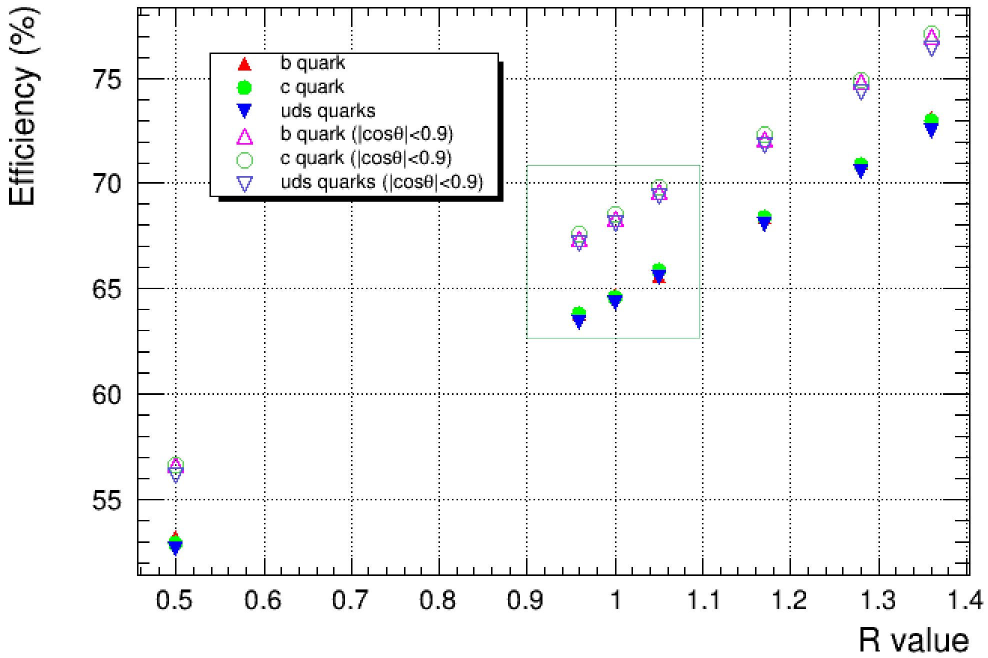


Fig. 11: Selection efficiencies for different quark flavors and R values (with and without angular restriction)

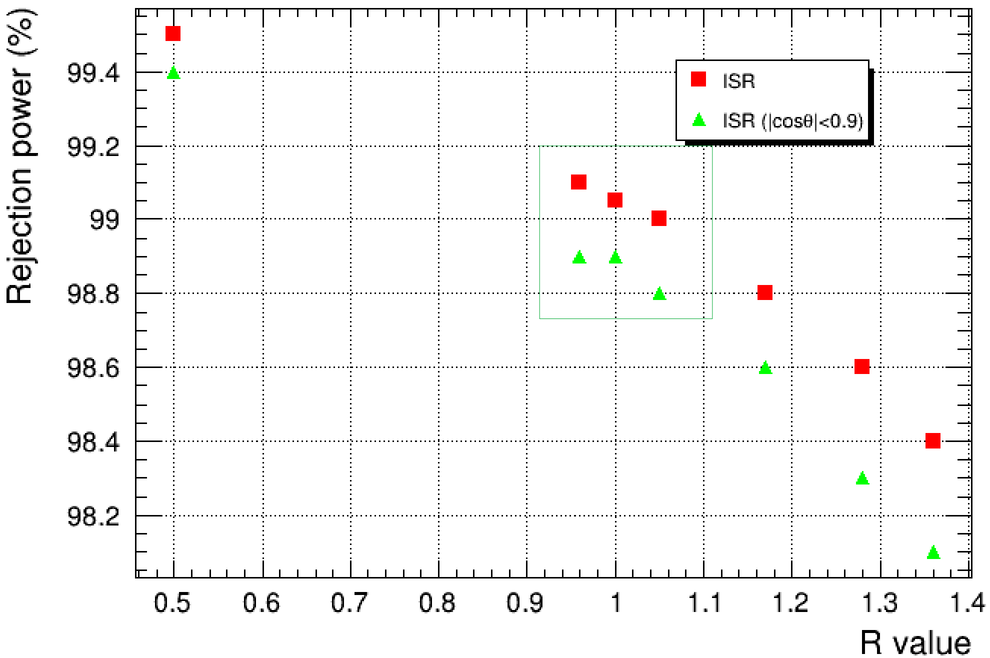


Fig. 12: Rejection power for different R values (with and without angular restriction)

Re-optimization of the cuts: VLC algorithm – R (II)

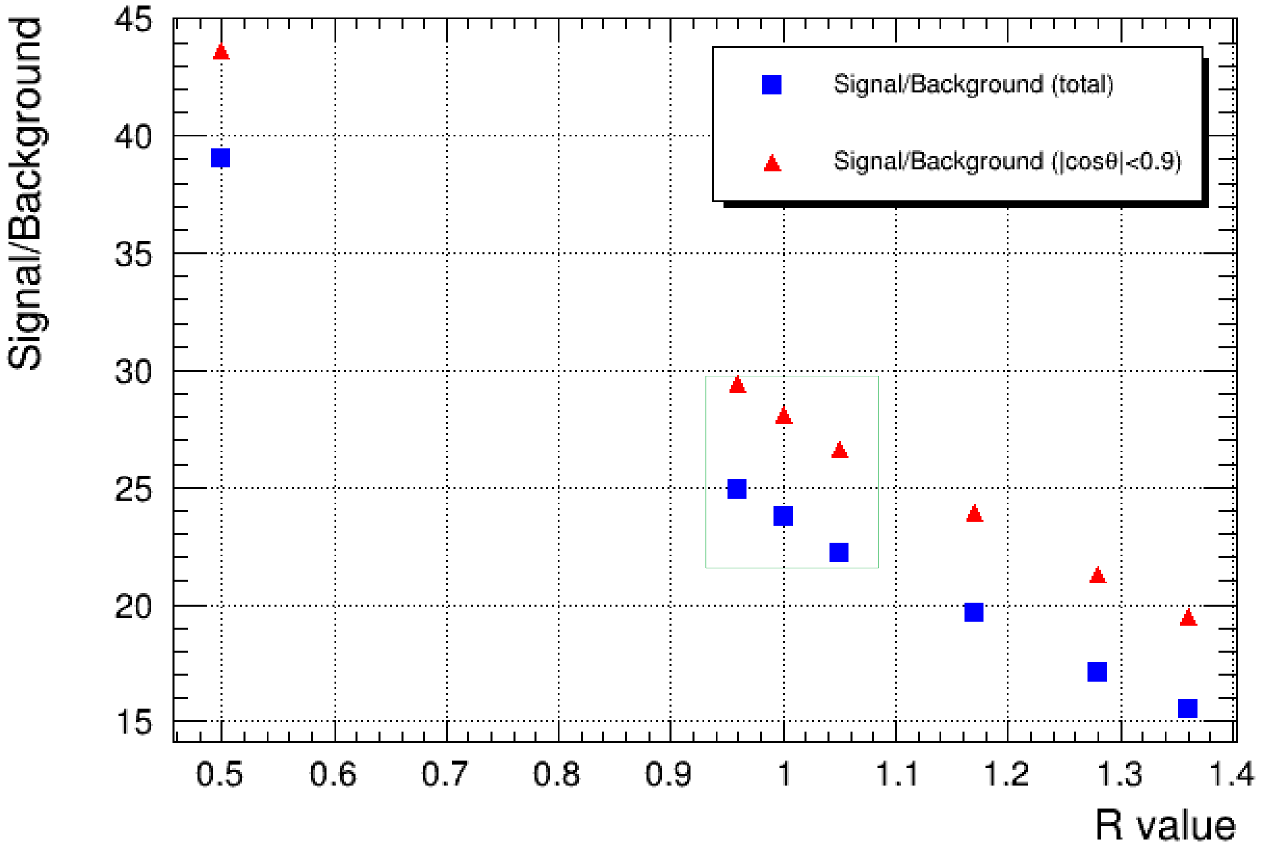


Fig. 13: Signal/Background for different values of R (with and without angular restriction)

Re-optimization of the cuts: VLC algorithm – γ

- Impact in the very forward/backward regions:

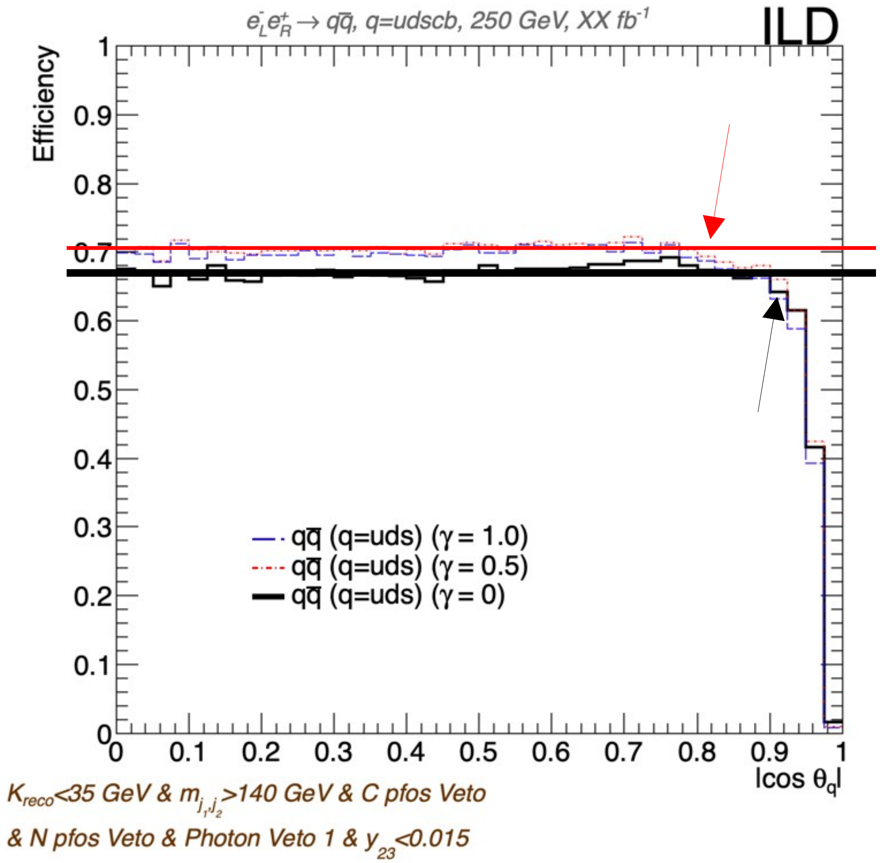


Fig. 14: Efficiency of the preselection for light quarks (uds) vs the angular distribution of the two jet system. Plots with R=1 and γ=0.0, 0.5 and 1.0

Smaller γ provide slightly smaller global efficiencies but more homogeneous efficiencies in most of the detector volume (|cos(θ)|<0.9)

The best parameter choices are R=1 and γ=0.

Conclusions

- A new optimization of the preselection is provided. This is to be used in ss/cc/bb analysis at √s=250GeV and will be the starting point for √s=500GeV
- At higher energies, the VLC algorithm may play a more important role (larger beam background expected)
- Chosen configuration:

Cuts:

- K_{reco} < 35 GeV
- m_{2jets} > 140 GeV
- Charged N pfos
- Photon veto
- Y_{23} < 0.015

VLC Algorithm parameters:

- R = 1.0
- γ = 0.0
- β = 1.0

	Efficiencies (%)				
R	$b\bar{b}$	$c\bar{c}$	$q\bar{q}$ (uds)	ISR	S/B
1.0	64.7	64.6	64.3	0.9	23.7
	68.3	68.5	68.1	1.1	28.1

← |cosθ| < 0.9

Table 5: Total efficiency of the preselection for the different quark flavors and radiative return for the chosen configuration (γ=0). The second row is for |cosθ| < 0.9

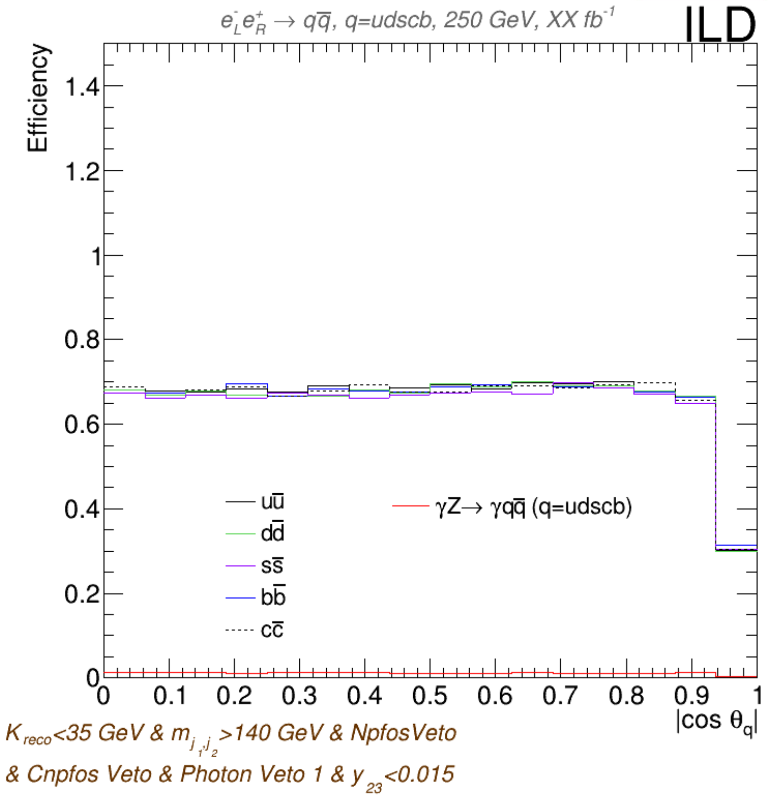


Fig. 15: Efficiency of the preselection for the different quark flavors vs the angular distribution of the two jet system (new samples, final configuration)

IFIC INSTITUT DE FÍSICA CORPUSCULAR

END OF THE PRESENTATION

THANKS FOR YOUR ATTENTION

Jesús P. Márquez Hernández

BACK-UP SLIDES

Re-optimization of the cuts: Why do we need equal efficiencies?

- Impact in the Double Tag method for b-tagging:
  - Observables:
    - Ratio of q-flavored quarks:

$$R_q^{cont.}(|cos\theta_q|) = \frac{\sigma^{cont.}_{e^- e^+ \to q\bar{q}}(|\cos\theta|)}{\sigma^{cont.}_{had.}(|\cos\theta|)}$$

For a specific q flavor (numerator)

All possible q flavors (denominator)

Having different selection efficiencies for each quark flavor will lead to a biased computation of this observable

This bias affects the b-tagging process

- Fraction of jets tagged as b-quark (first tag):

$$f_1 = \varepsilon_b R_b^{cont.} + \varepsilon_c R_c^{cont.} + \varepsilon_{uds}(1 - R_b^{cont.} - R_c^{cont.}) + F_1^{bkg}(\varepsilon_c, \varepsilon_b, \varepsilon_{uds}, BKG)$$

- Fraction of preselected events in which both jets are tagged as b-quark (second tag):

$$f_2 = \varepsilon_b^2(1+\rho_b) R_b^{cont.} + \varepsilon_c^2 R_c^{cont.} + \varepsilon_{uds}^2(1 - R_b^{cont.} - R_c^{cont.}) + F_2^{bkg}(\varepsilon_c^2, \varepsilon_b^2, \varepsilon_{uds}^2, BKG)$$

Re-optimization of the cuts: Leveling up the photon veto

Previous approach:

- Cuts:
  - K_{reco} < 35 GeV & m_{2jets} > 130 GeV
  - N pfos > 5
  - Photon vetoing
  - y_{23} < 0.015
  - $m_{j1} + m_{j2}$ < 100 GeV
- Efficiencies:

Efficiency (%)			
$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
71.5	69.9	68.4	0.7

Table E1: Total efficiency of the preselection for the different quark flavors and radiative return with the previous cut selection (new samples)

Re-balancing the photon veto:

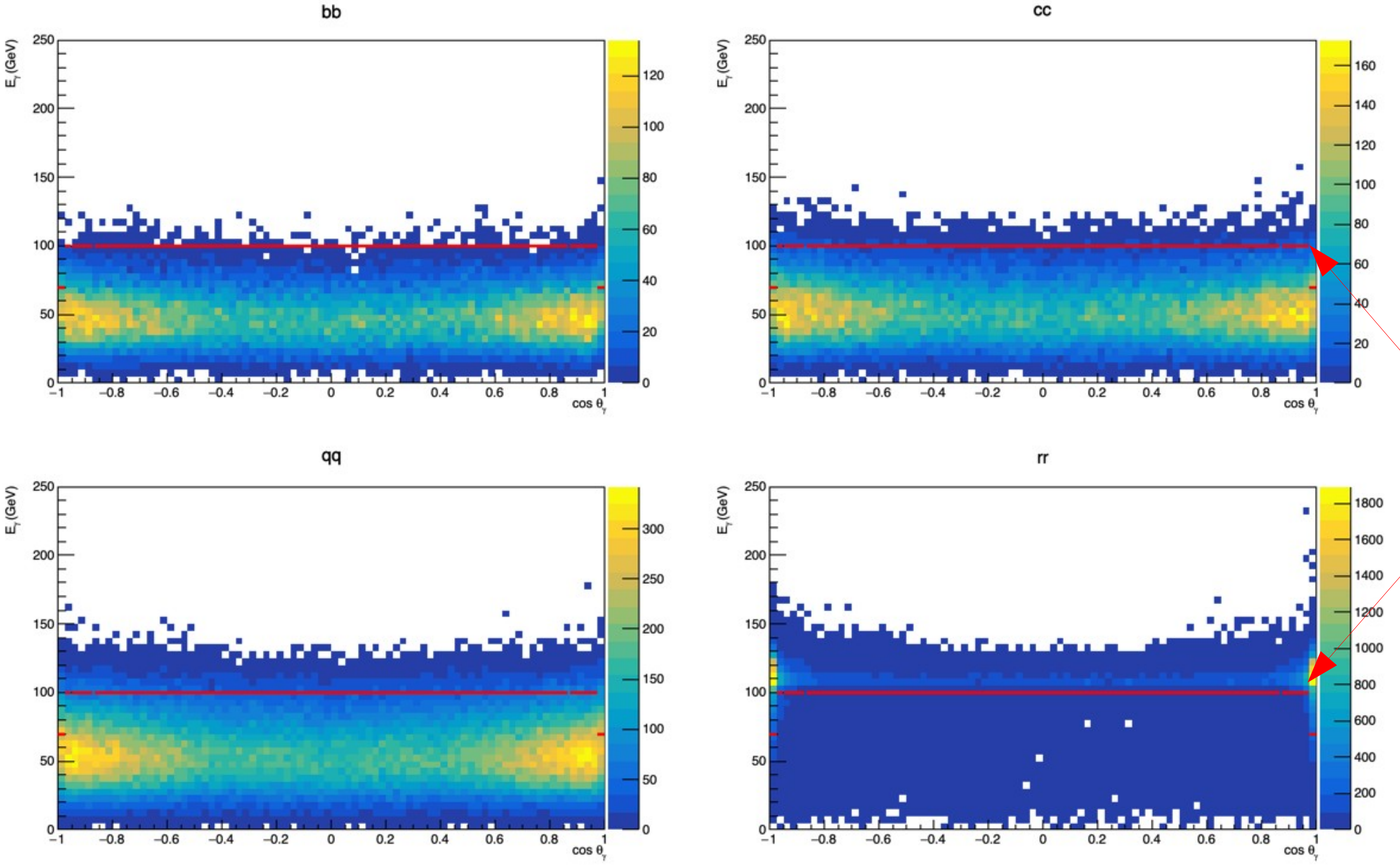
- Cuts:
  - K_{reco} < 35 GeV & m_{2jets} > 130 GeV
  - N pfos > 5
  - Photon vetoing (leveling up)
  - y_{23} < 0.015
  - $m_{j1} + m_{j2}$ < 100 GeV
- Efficiencies:

Efficiency (%)			
$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
71.2	71.1	71.0	1.9

Table E2: Total efficiency of the preselection for the different quark flavors and radiative return with the new cut selection (new samples)

IFIC INSTITUT DE FÍSICA CORPUSCULAR

New samples (Photon veto)



If we try to re-balance the selection efficiencies by leveling up these energies levels then the rejection efficiency for the radiative return goes up to ~2%

Fig. E1: Two dimensional maps used for the photon vetoing; impact in different quark flavors (new samples). The y-axis is the energy (GeV) of the pfos identified as photons by the jet reconstruction algorithm and the x-axis is cos(θ) of those pfos.

IFIC INSTITUT DE FÍSICA CORPUSCULAR

Re-optimization of the cuts: N. PFOs

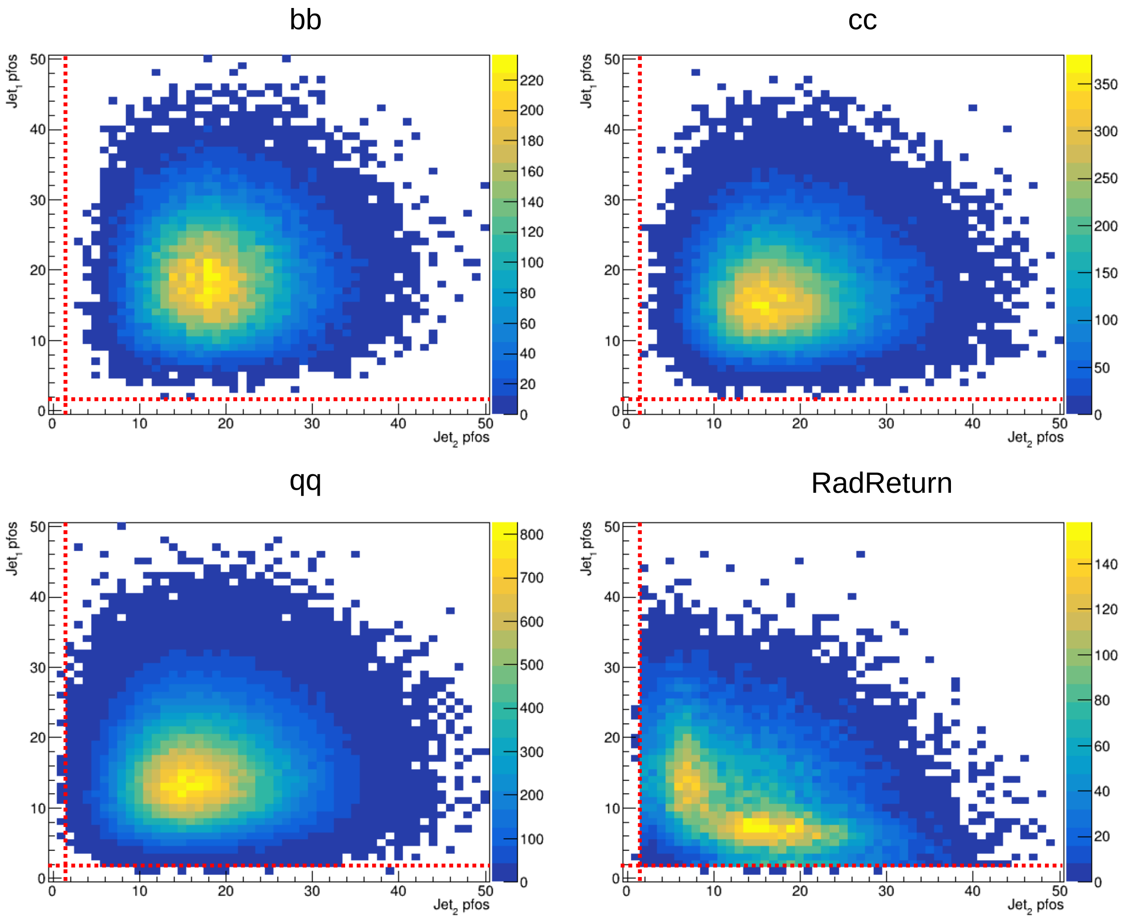


Fig. E2: Selected cut for the number of total pfos associated to each jet

At least 2 pfos in each jet

Re-optimization of the cuts: VLC algorithm – γ (E. I)

	Efficiency (%)											
	$\gamma = 0.0$				$\gamma = 0.5$				$\gamma = 1.0$			
R	$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR	$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR	$b\bar{b}$	$c\bar{c}$	$q\bar{q}(uds)$	ISR
0.96	63.8	63.8	63.4	0.9	68.0	67.7	67.2	0.9	68.1	67.8	67.2	0.9
	67.4	67.6	67.2	1.1	71.0	70.9	70.3	1.1	70.6	70.4	69.7	1.0
1.00	64.7	64.6	64.3	0.9	68.8	68.5	67.9	1.0	68.8	68.4	67.8	0.9
	68.3	68.5	68.1	1.1	71.8	71.7	71.1	1.1	71.3	71.0	70.4	1.0
1.05	66.0	65.9	65.5	1.0	69.8	69.5	68.9	1.0	69.7	69.4	68.8	1.0
	69.6	69.8	69.4	1.2	72.9	72.7	72.2	1.2	72.2	72.0	71.4	1.1

Table E3: Total efficiency of the preselection for the different quark flavors and radiative return for different values of R and γ. The second row of each set is constrained to the barrel region (for |cosθ| < 0.9)

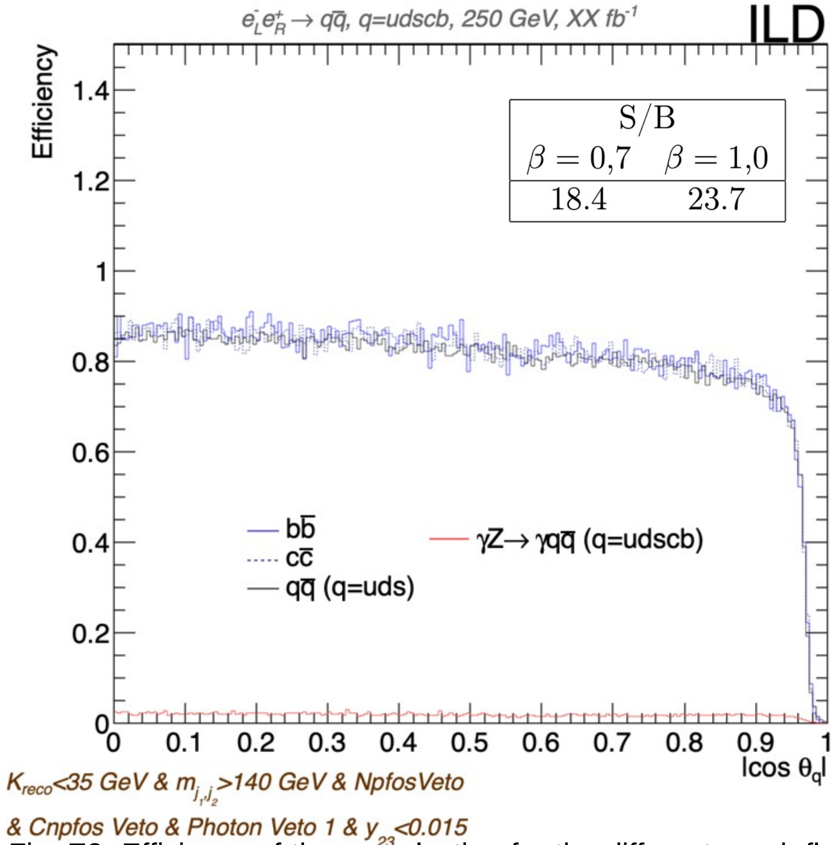
The best parameter choices are R= 0.96 or 1 (in global efficiency estimation) and γ=0 (due to the loss of efficiency in the forward/backward regions for other values)

	Signal/Background		
R	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 1.0$
0.96	24.9	25.2	26.6
	29.4	27.6	27.8
1.00	23.7	24.5	25.7
	28.1	26.9	27.1
1.05	22.2	23.4	24.5
	26.6	25.9	26.0

Table E4: S/B for the different quark flavors and radiative return for different values of R and γ. The second row of each set is constrained to the barrel region (for |cosθ| < 0.9)

IFIC INSTITUT DE FÍSICA CORPUSCULAR

Re-optimization of the cuts: VLC algorithm – β (E. I)



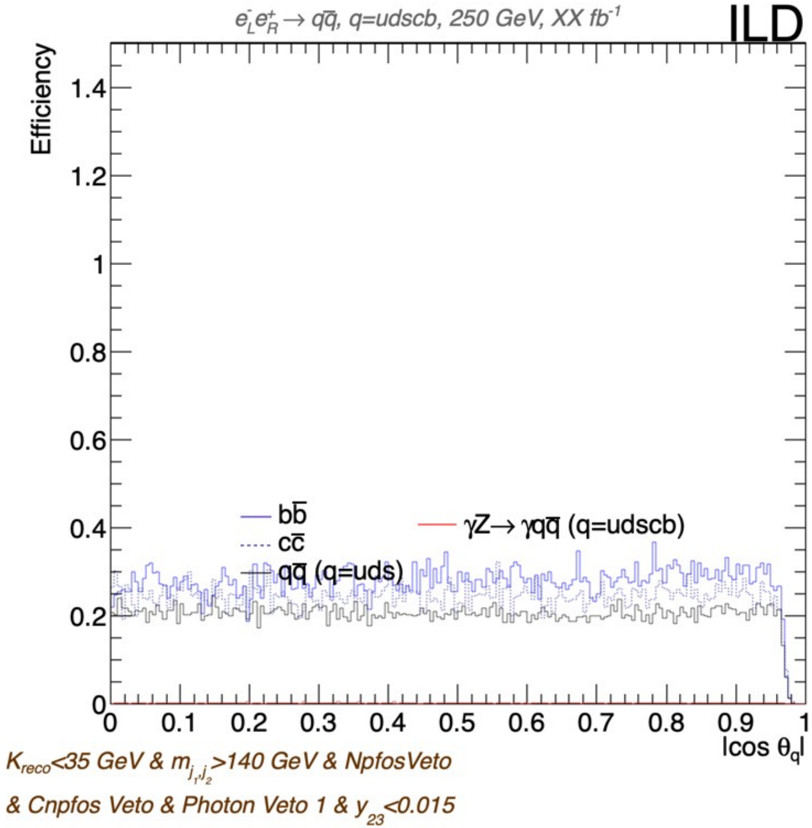


Fig. E3: Efficiency of the preselection for the different quark flavors (b,c, uds) vs the angular distribution of the two jet system (new samples, final configuration) for different β values: 0.7 (left) and 1.4 (right)

Lower values of β result in a lower S/B due to leakage of ISR while higher values leads to lower efficiencies and differences between quark flavors. <u>We decided to keep β=1</u>

Preselection for all quark flavors (u,d,s,c,b)

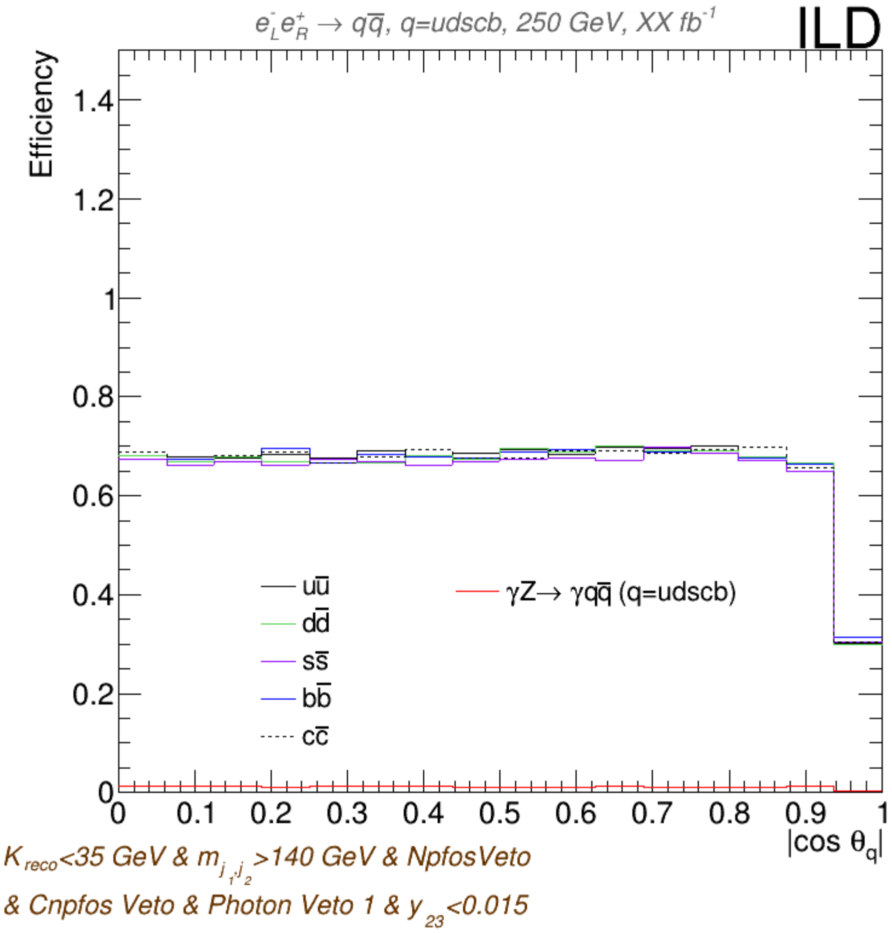


Fig. E4: Efficiency of the preselection for the different quark flavors (u,d,s,c,b) vs the angular distribution of the two jet system (new samples, final configuration)

IFIC INSTITUT DE FÍSICA CORPUSCULAR

Efficiencies after each cut

Cuts:

1) K_{reco} < 35 GeV
2) m_{2jets} > 140 GeV
3) Charged N pfos
4) Photon veto
5) Y_{23} < 0.015

Cut	Efficiencies (%) $b\bar{b}$	$c\bar{c}$	$q\bar{q}$ (uds)	ISR	S/B
1	81.2	80.8	81.0	5.7	4.9
2	80.9	80.7	81.0	5.2	5.4
3	80.9	80.7	81.0	5.2	8.4
4	77.9	77.3	76.5	1.5	18.4
5	64.7	64.6	64.3	0.9	23.7

Table E5: Selection efficiency, rejection power and Signal/Background ratio after each cut

Cut	Efficiencies (%) $b\bar{b}$	$c\bar{c}$	$q\bar{q}$ (uds)	ISR	S/B
1	83.3	83.0	83.0	5.8	6.8
2	83.1	82.9	83.0	5.1	7.6
3	83.0	82.9	82.8	2.6	15.0
4	82.9	82.5	81.6	1.8	21.1
5	68.3	68.5	68.1	1.1	28.1

Table E6: Selection efficiency, rejection power and Signal/Background ratio after each cut. Data constrained to $|\cos\theta| < 0.9$

B-tagging

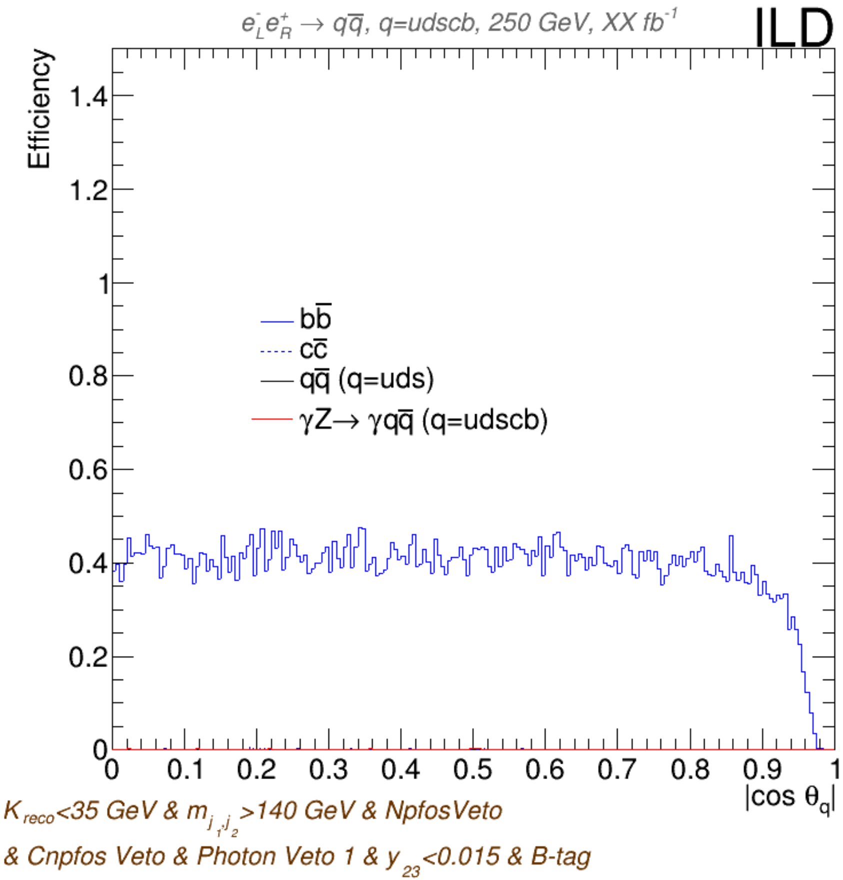


Fig. E5: Efficiency of the preselection for different quark flavors and ISR vs the angular distribution of the two jet system after applying b-tagging

Efficiencies (%)				
$b\bar{b}$	$c\bar{c}$	$q\bar{q}$ (uds)	ISR	S/B
37.51	0.02	0.00	0.06	35.34
40.80	0.02	0.00	0.08	40.49

Table E7: Total efficiency of the preselection for the different quark flavors and radiative return after applying b-tagging. The second row is constrained to |cosθ| < 0.9

Durham algorithm

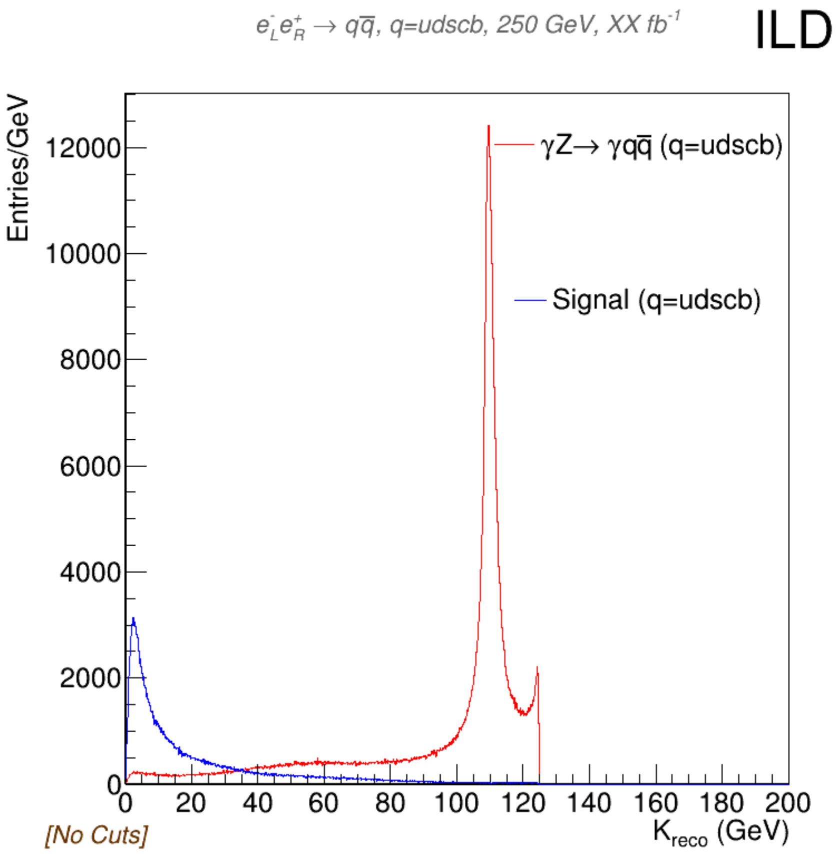


Fig. E6: Distribution of events vs K_{reco} for the reconstructed data with the Durham algorithm (left) and the VLC algorithm (right)

Durham algorithm (II)

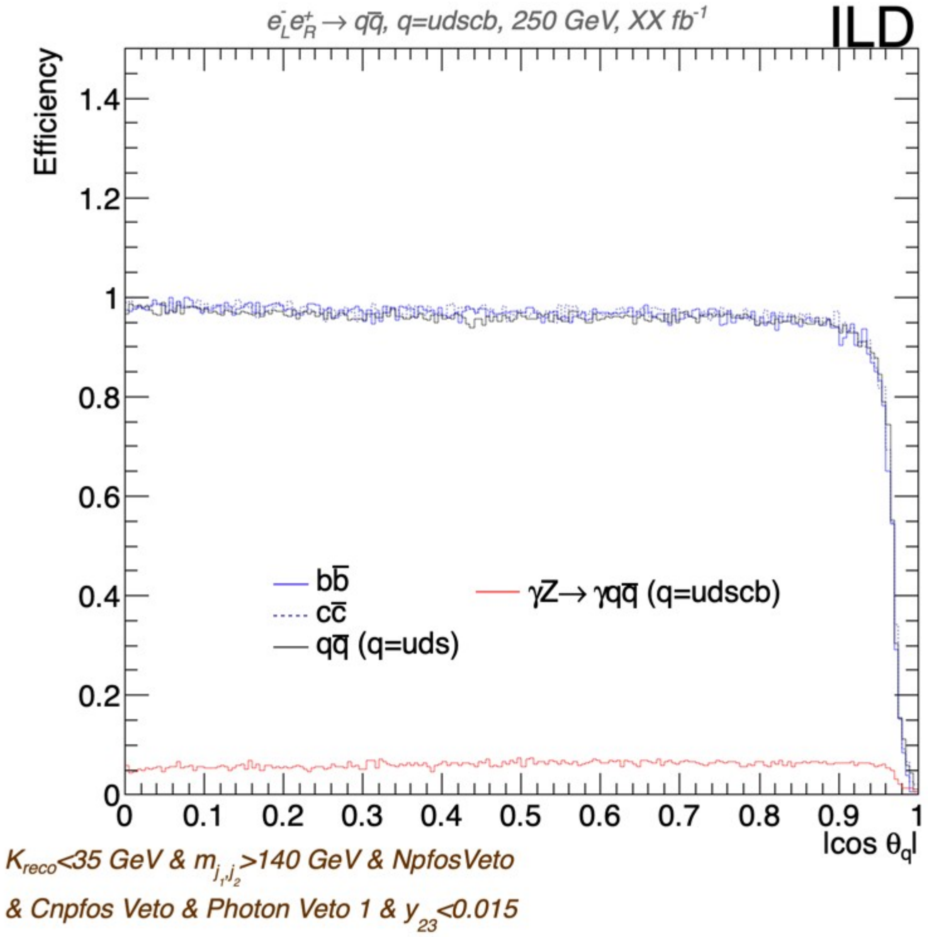


Fig. E7: Efficiency of the preselection for the different quark flavors (b,c,uds) vs the angular distribution of the two jet system (new samples, final configuration)

Efficiencies (%)				
$b\bar{b}$	$c\bar{c}$	$q\bar{q}$ (uds)	ISR	S/B
91.5	91.8	91.1	5.3	6.1
96.7	97.0	96.0	6.2	7.4

Table E8: Selection efficiency, rejection power and Signal/Background ratio after performing all cuts with the data reconstructed using the Durham algorithm