AFB studies at 500 GeV (update)

ILD Top/HF group meeting 08/04/22

A. Irles*, J. P. Márquez*

*Orsay/Tohoku/Valencia HQ@ILC Team

AITANA group at IFIC-CSIC/UV

Recapitulation

- We presented the different cuts for the preselection of the signals (back-up slides).

- We studied the dependence of the A_{FB} (at Monte-Carlo level) for different K_{ISR} values:
  - This showed that the cut in K_{reco} is safe.
  - We fixed a value of K_{ISR}=50 GeV. (back-up slides).

- We checked the b-tag and c-tag setting used at 250 GeV and applied it to this 500 GeV samples, and selected a cut for each (back-up slides).

- We were about to try a re-training of the tagging and compare the performance.
  - This presentation. ***Work in progress***

Retraining procedure

- Training algorithms:
  - TrackNtuple and TrackProb.C:
    - Prepare the data files to be used in flavor tagging.
    - Necessary with different geometry (different vertex detector configuration).
  - MakeNtuple:
    - Prepare training data.
    - Needs a “vertexing” file (output of TrackNtuple+TrackProb.C).
  - Train:
    - Run TMVA to train the flavor tagging BDT and produce the weights to b/c-tag our data.
    - Needs MakeNtuple output.

Retraining procedure

- Changing vtx file is non necessary but we tried 3 options to check:
  - 2f250, already available.
  - VvH250, already available.
  - 2q500, obtained with TrackNtuple.
- We use 2q500 samples (qqbar production at 500 GeV) for $e^-_L e^+_R$ and $e^-_R e^+_L$:
  - Weights from $e^-_L e^+_R$ to tag $e^-_R e^+_L$
  - Weights from $e^-_R e^+_L$:to tag $e^-_L e^+_R$
- Used default configuration of the BDT in train.xml:
  - NTrees=1000:BoostType=Grad:Shrinkage=0.10:UseBaggedBoost:BaggedSampleFraction=0.50:nCuts=20:MaxDepth=6
- Once we obtained the weights we processed the samples with the same $Q\overline{Q}$processor that we used before.
- We encounter (and fixed) few problems while doing it (next slides).

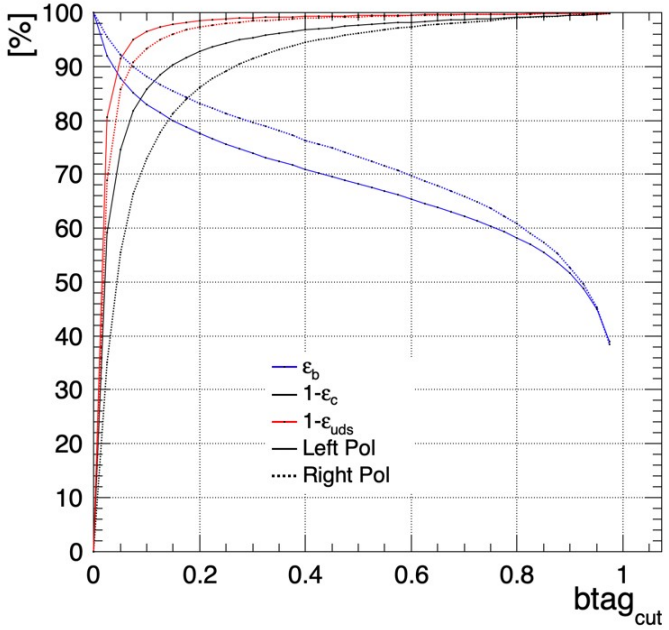
Retraining procedure (issues)

- Aida issue: The steering files seem to be incompatible with Aida.
- Fatal error in the code NaN or +-inf (~700 events out of 100K):
  - Solution (thanks to Ryo Yonamine):
    - Adding “!Tmath::IsNaN(parameter_that_glitches)” in FlavorTag.CategoryPreselection
- “Warming” in the code:
  - Deprecated option UseBaggedGrad changed to UseBaggedBoost.
    - Changed. Doesn’t seem to have an impact.
- *Right now*, reproducing the previous weights with the same samples, to check if the entire process its working as expected.

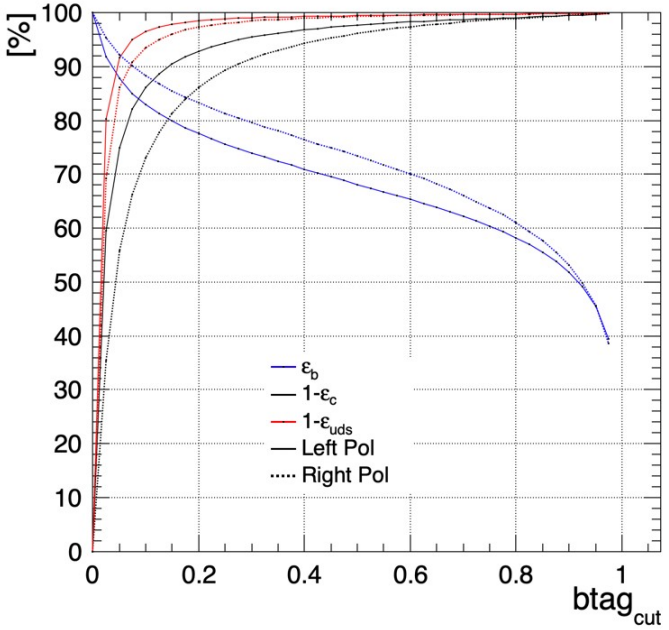
NOTE: The results in the next slides are in a very early stage.
WORK IN PROGRESS

Results for b-tag

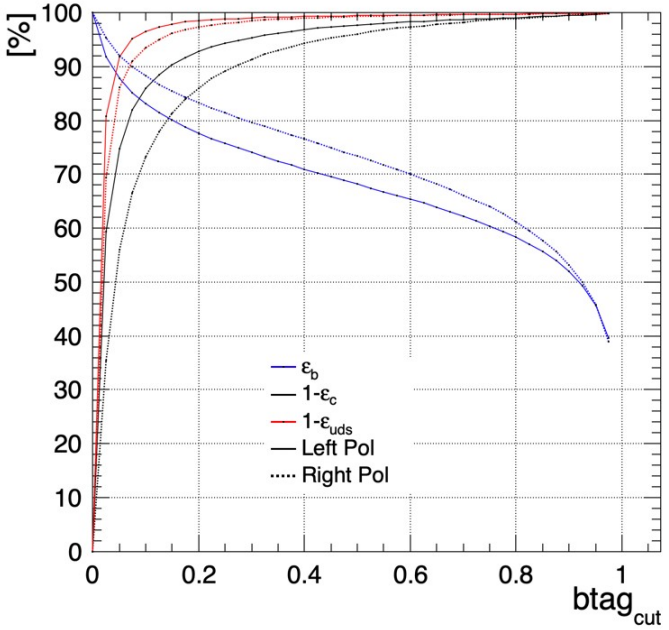
“2q500”



“2f250”



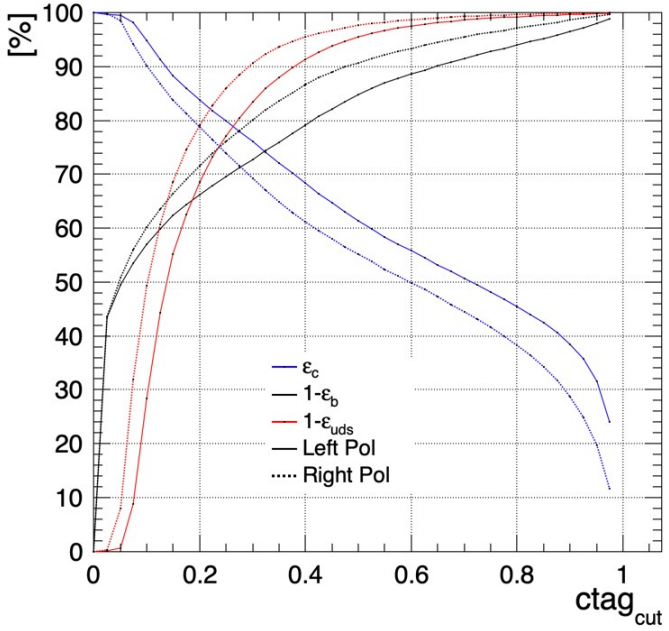
“vvH”



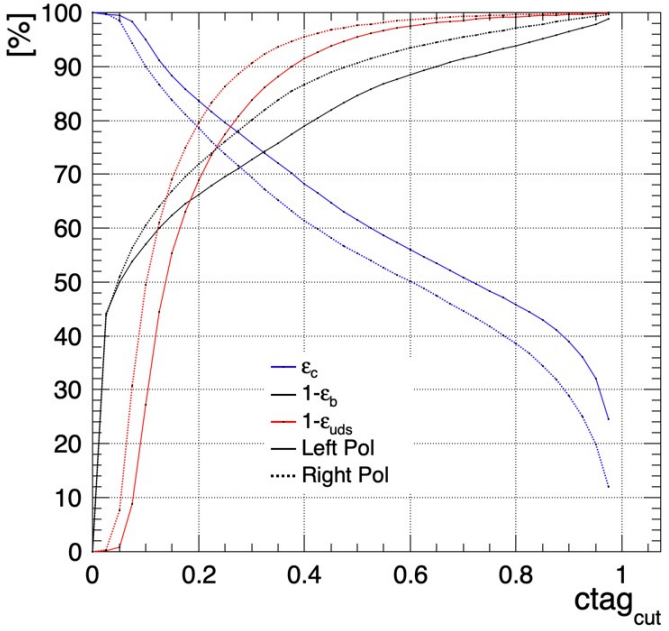
As we expected, the 3 different vtx files reproduce the same result.

Results for c-tag

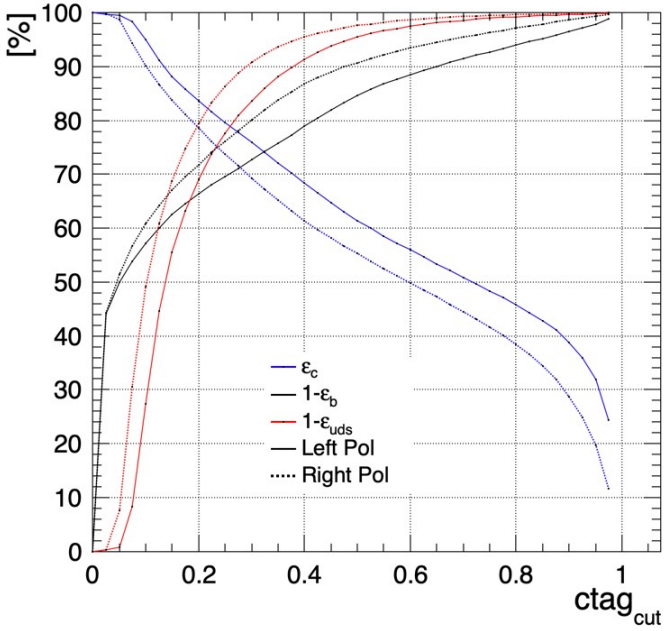
“2q500”



“2f250”



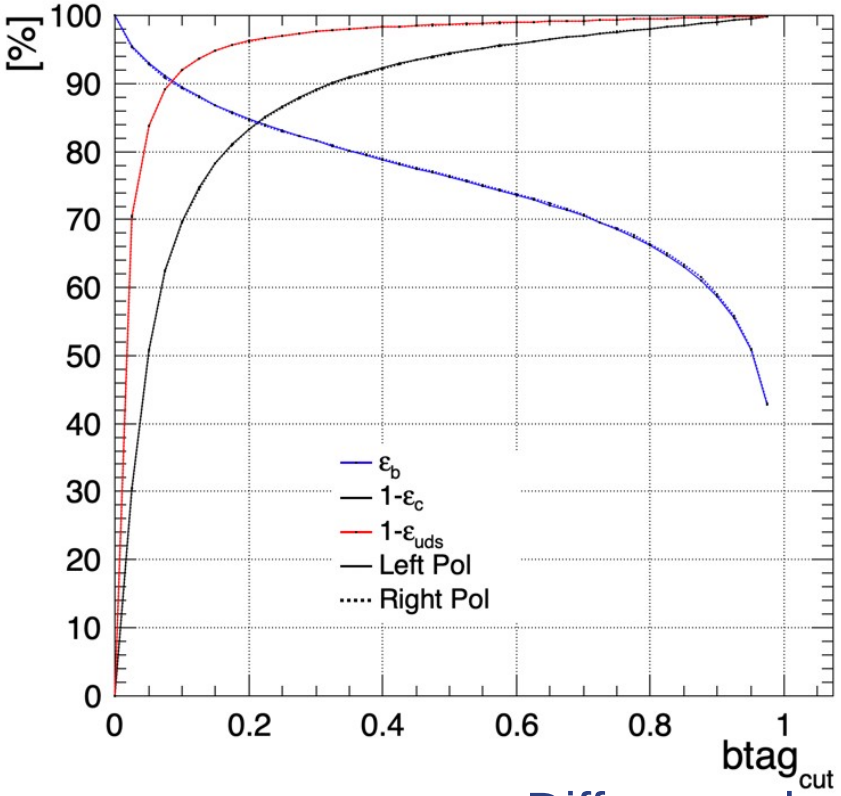
“vvH”



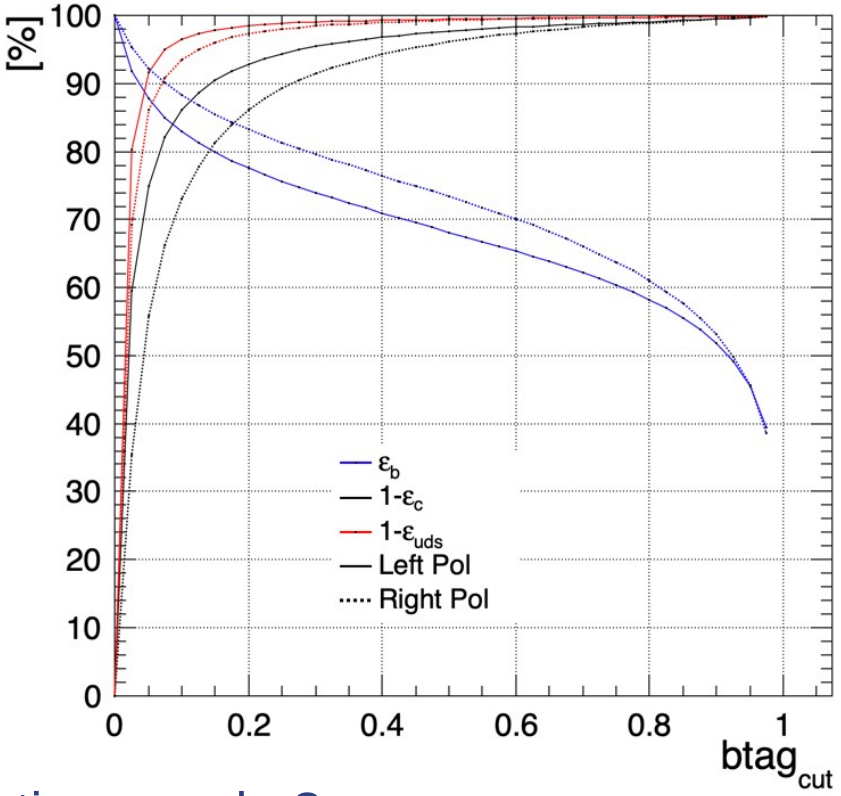
As we expected, the 3 different vtx files reproduce the same result.
From now on, I will only use the 2f250 one to compare with the previous tagging

Comparison with the previous tagging (b)

Previous tagging



New tagging

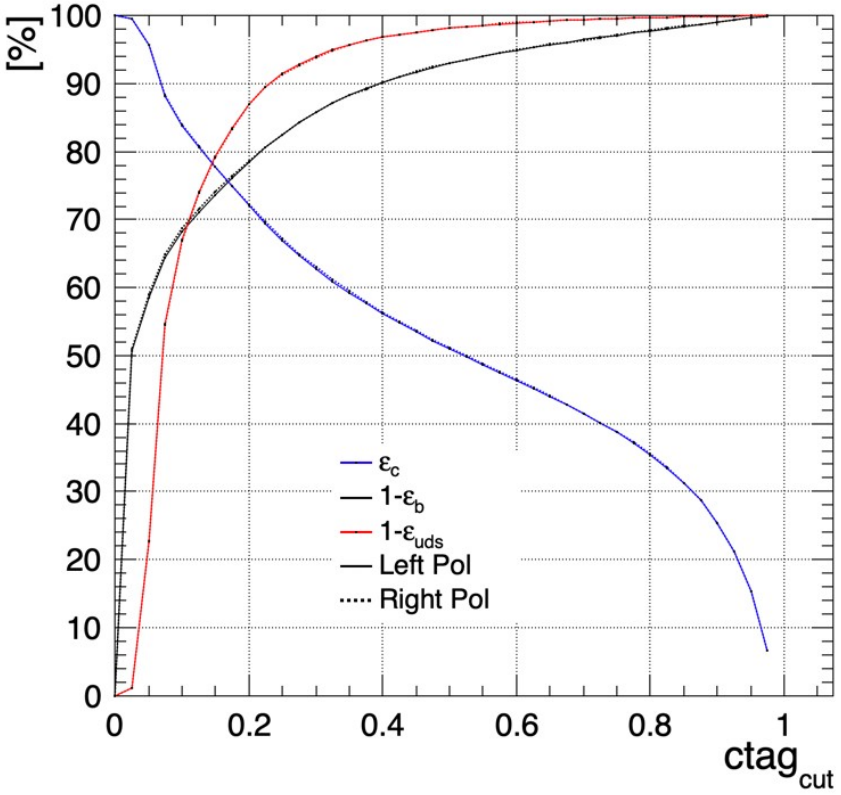


Difference between polarizations… why?

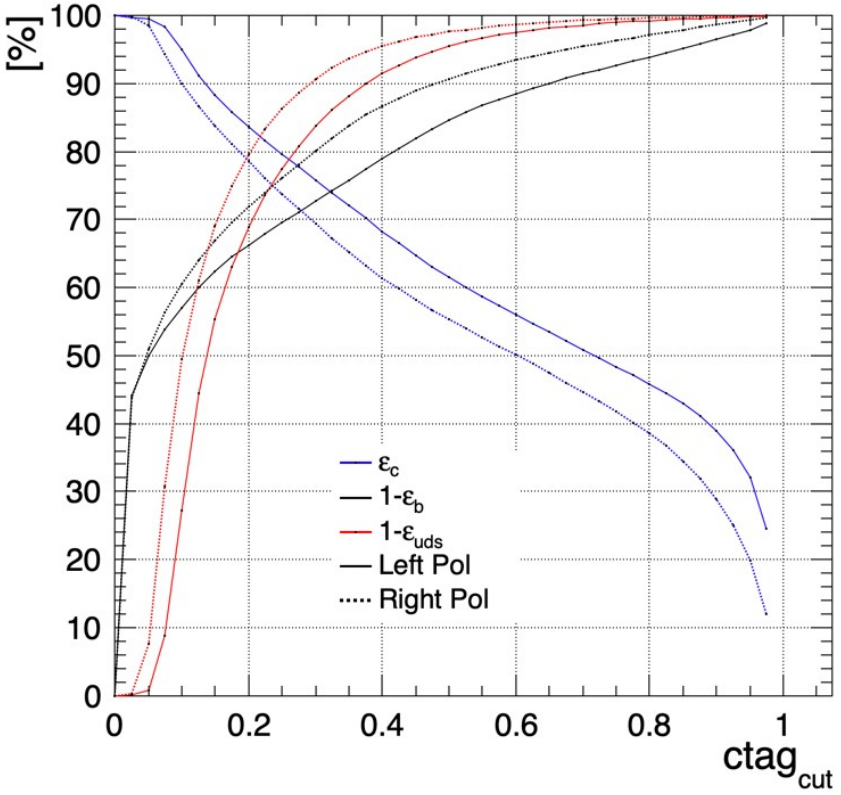
The balance between btag$_{cut}$ and selection/rejection efficiency is shifted

Comparison with the previous tagging (c)

Previous tagging



New tagging



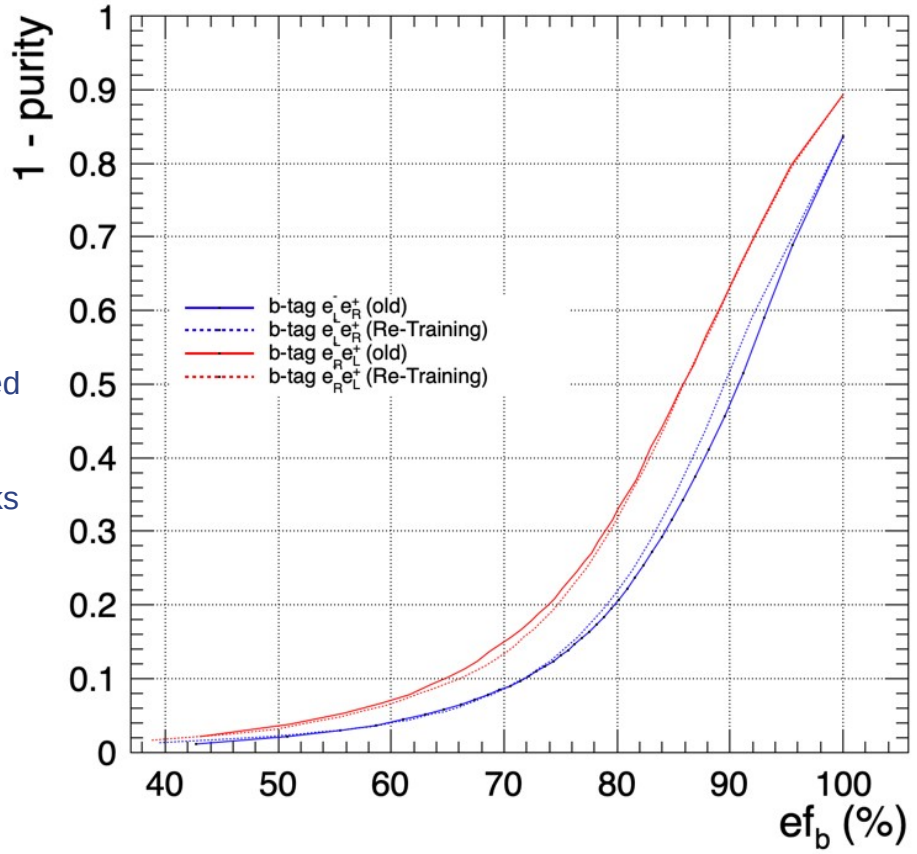
Same effects. But beside this, are the new weights better or worse?

Comparison with the previous tagging (b)

$Purity_f = N_{f,tagged}/N_{T,tagged}$

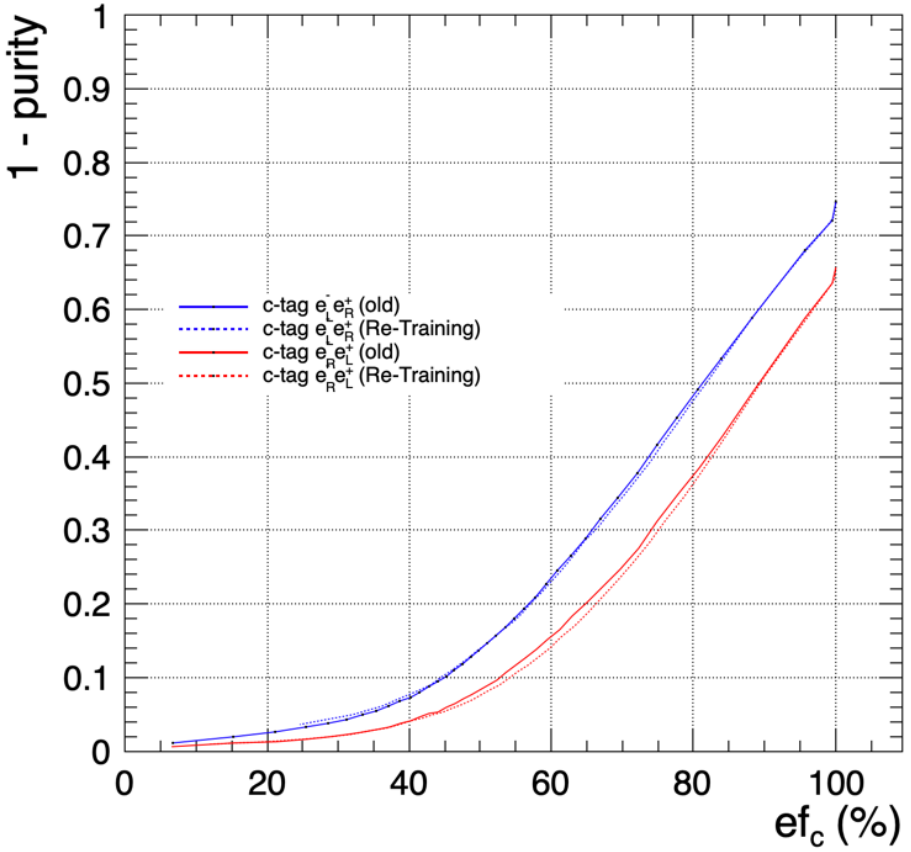
$N_{f,tagged}$ = Events of flavor f properly tagged as f quarks

$N_{T,tagged}$ = All the events tagged as f quarks



The performance is very similar

Comparison with the previous tagging (c)



Again, very similar

On going work

- *Right now*, reproducing the previous weights with the same samples, to check if the entire process its working as expected.
  - If it is working well. Check if there's difference in polarization for those weights.
    - If there is, check which polarization were used to train the previous tag (since we only used 1 file for the samples of both polarizations).
      - And… why is there a difference?
    - If not: why do we have it with the 500 GeV weights?
- To do:
  - Digging in the code
    - Be aware of any new "warming" message.
    - Check if the configuration used for the bdt could be improved.
    - Check it the **methods** in the bdt could be improved.

Back-Up slides

Final preselection ($e_L p_R$)

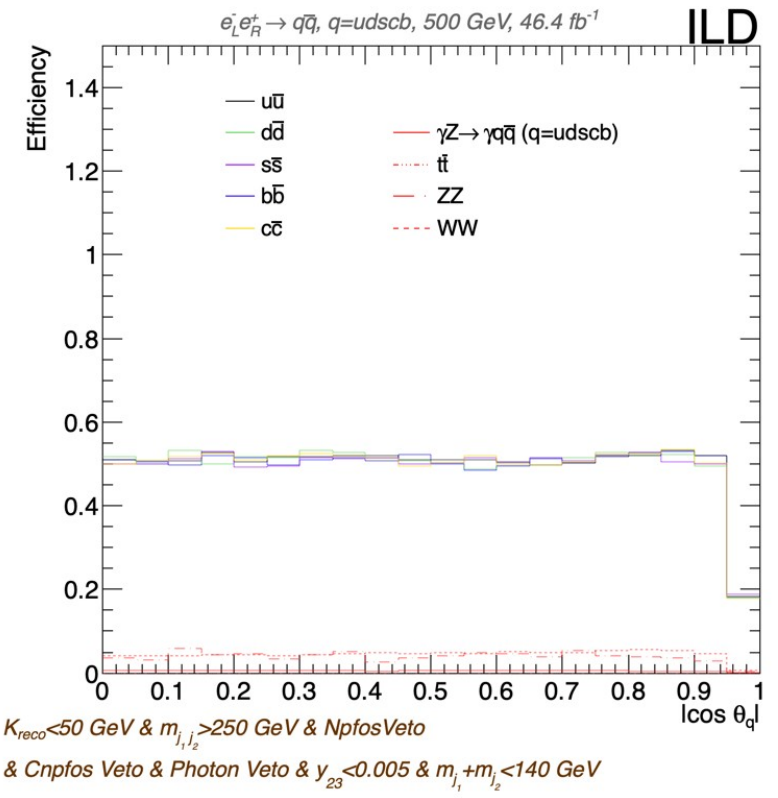
Cuts:

- K_{reco} < 50 GeV
- m_{2jets} > 250 GeV
- Charged N pfos > 0.5
- Neutral N pfos > 3.5
- Photon veto
- y_{23} < 0.005
- $m_{j1}+m_{j2}$ < 140 GeV

VLC Algorithm parameters:

- R = 1.0
- γ = 0.0
- β = 1.0

	Efficiency (%)			Background/Signal			
	$b\bar{b}$	$c\bar{c}$	$q\bar{q}$	ISR	WW	ZZ	$t\bar{t}$
No cut	100	100	100	3.50	1.06	0.09	0.10
+ Cut 1	74.9	74.7	74.7	0.76	0.77	0.06	0.01
+ Cut 2	74.8	74.6	74.7	0.74	0.77	0.06	9e-03
+ Cut 3	74.8	74.5	74.3	0.16	0.77	0.06	9e-03
+ Cut 4	74.7	74.5	74.1	0.11	0.77	0.06	9e-03
+ Cut 5	72.1	71.7	71.1	0.05	0.58	0.05	9e-03
+ Cut 6	49.6	49.7	49.6	0.03	0.09	0.01	1e-04
+ Cut 7	48.6	48.7	48.7	0.02	0.06	5e-03	5e-06



Final preselection ($e_R p_L$)

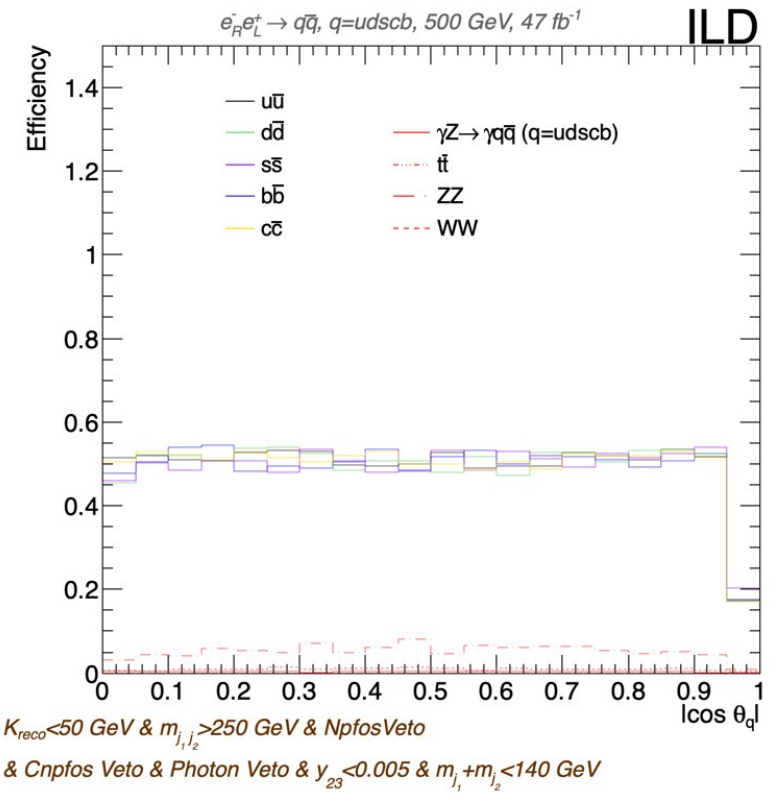
Cuts:

- K_{reco} < 50 GeV
- m_{2jets} > 250 GeV
- Charged N pfos > 0.5
- Neutral N pfos > 3.5
- Photon veto
- y_{23} < 0.005
- $m_{j1}+m_{j2}$ < 140 GeV

VLC Algorithm parameters:

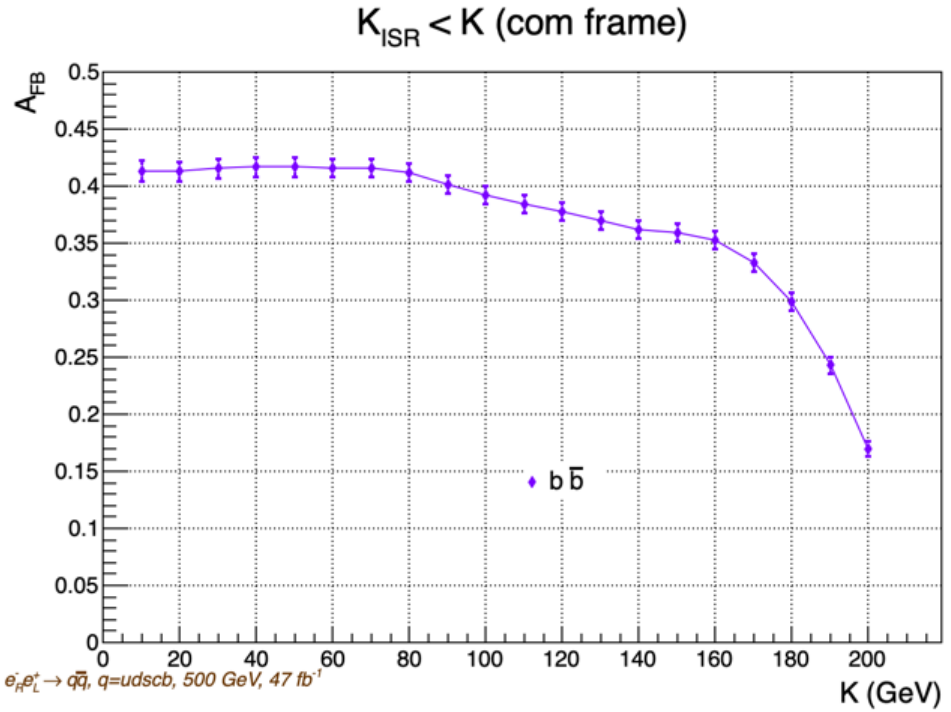
- R = 1.0
- γ = 0.0
- β = 1.0

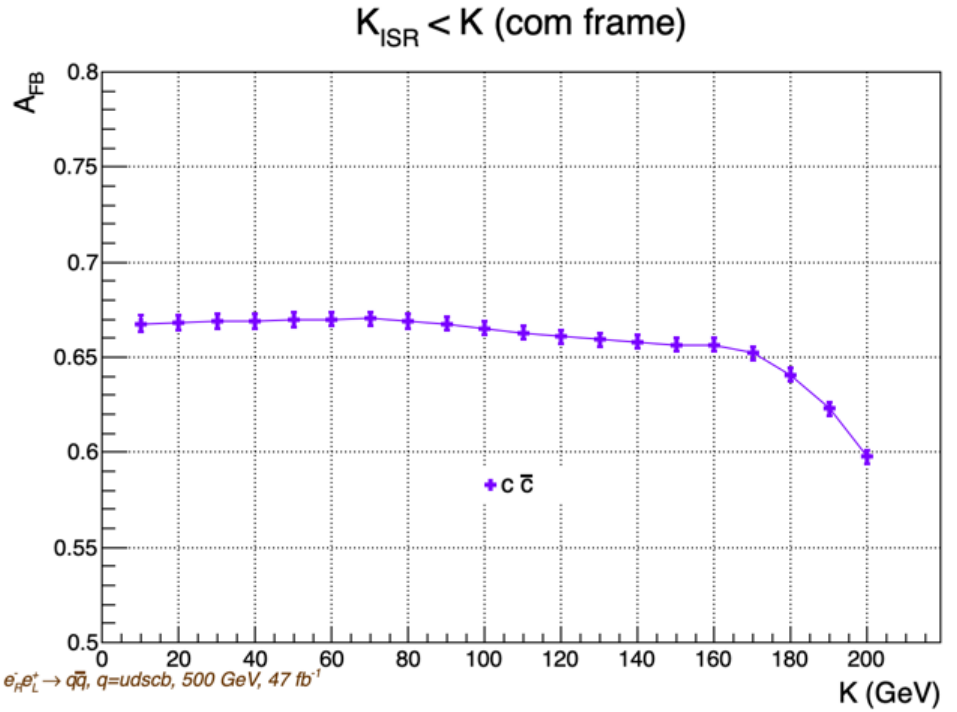
	Efficiency (%)			Background/Signal			
	$b\bar{b}$	$c\bar{c}$	$q\bar{q}$	ISR	WW	ZZ	$t\bar{t}$
No cut	100	100	100	6.51	0.01	0.11	0.10
+ Cut 1	74.6	74.6	75.0	1.45	0.01	0.07	0.01
+ Cut 2	74.5	74.5	75.0	1.43	0.01	0.07	0.01
+ Cut 3	74.5	74.4	74.7	0.26	0.01	0.07	0.01
+ Cut 4	74.5	74.4	74.5	0.18	0.01	0.07	0.01
+ Cut 5	71.9	71.7	71.5	0.07	0.01	0.06	0.01
+ Cut 6	49.5	49.6	49.6	0.03	5e-04	0.01	9e-05
+ Cut 7	48.5	48.8	58.7	0.03	3e-04	8e-03	3e-06



Zoom in K_{ISR} ($e_R p_L$)

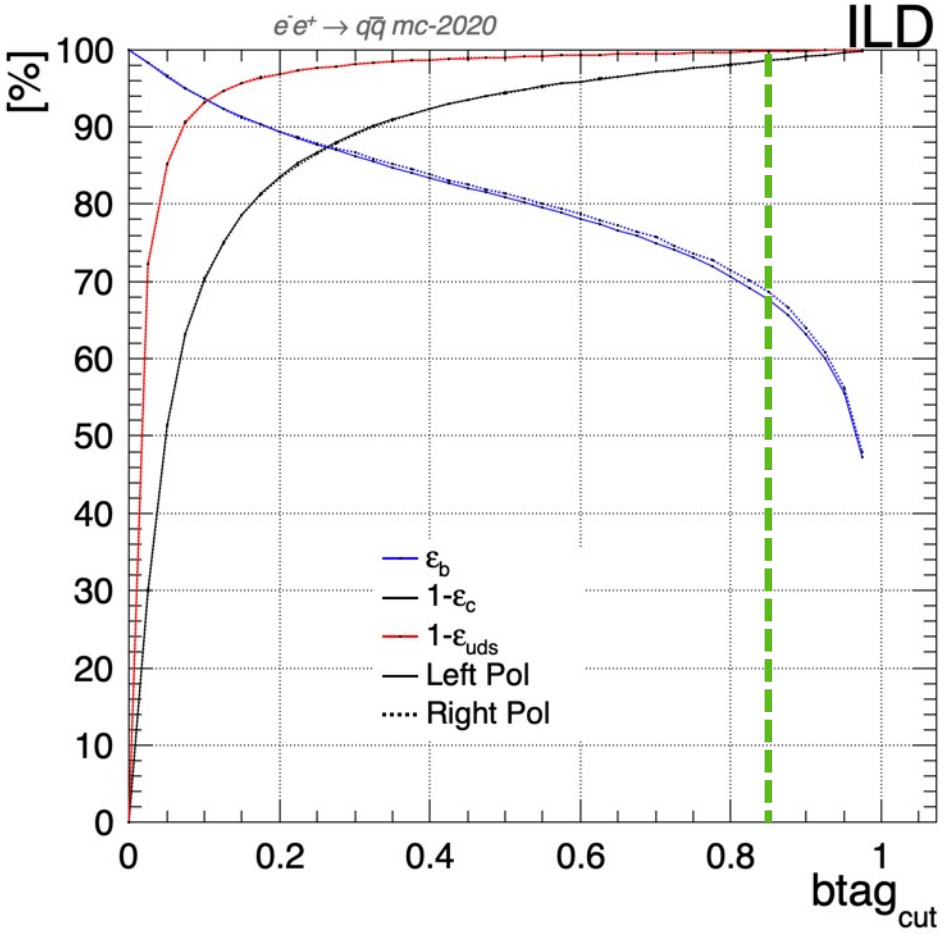
The signals change drastically when K_{ISR}>80 GeV



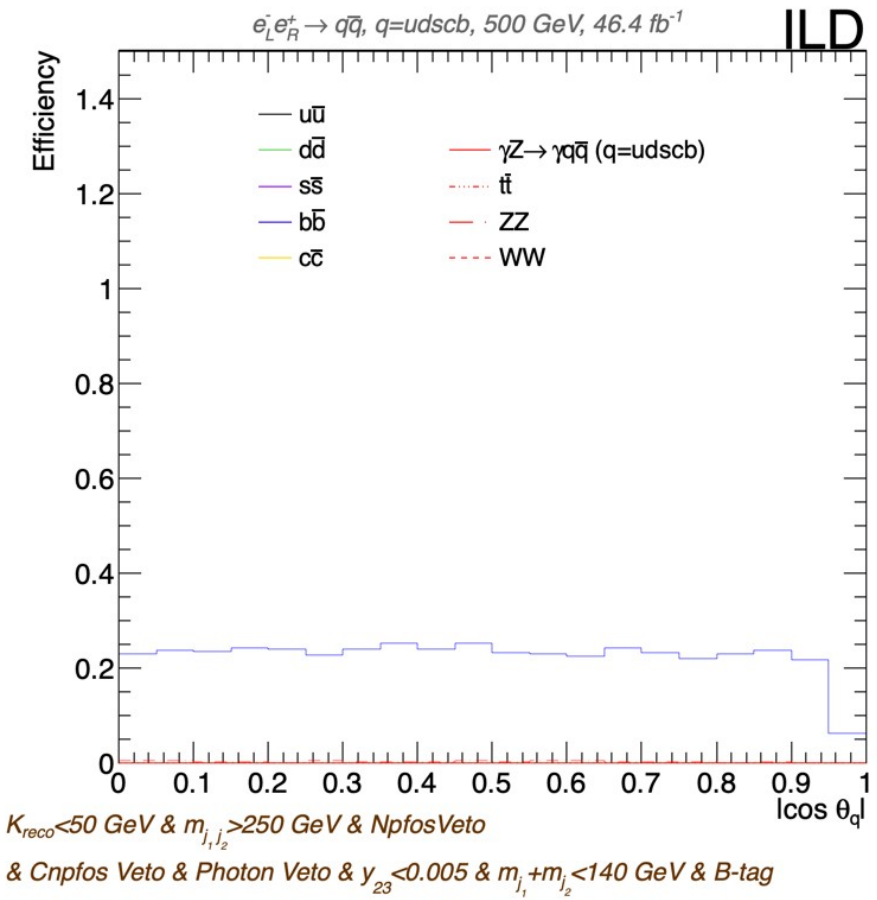


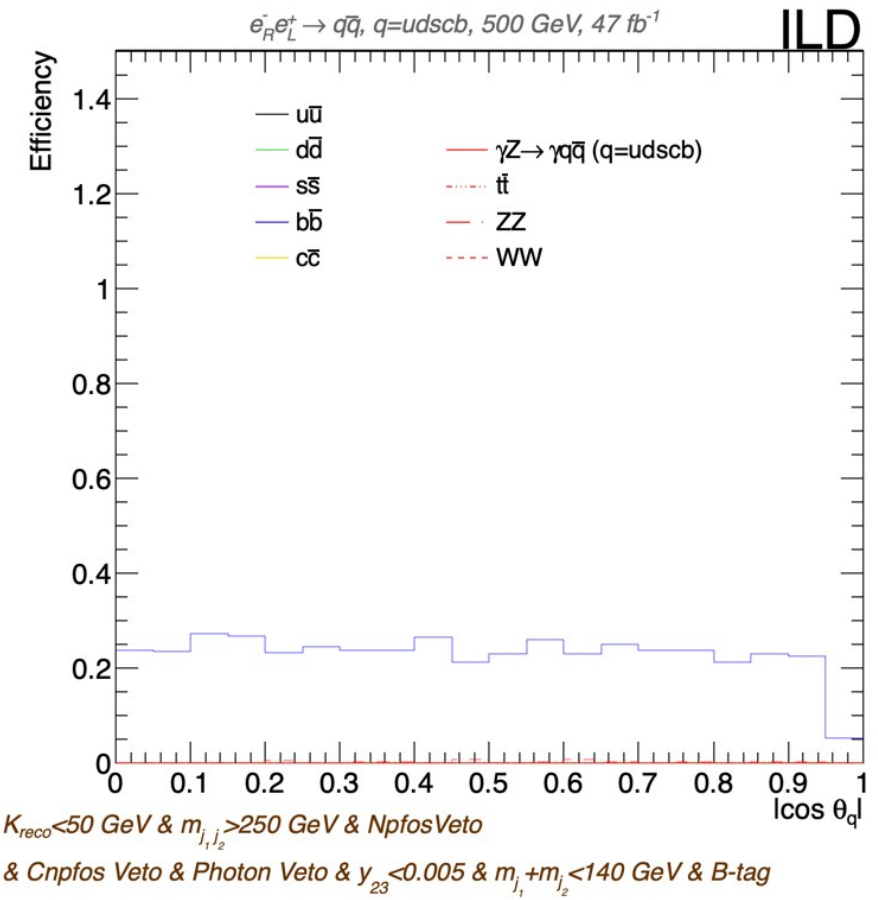
Safe region from 30 to 70 GeV, we fixed the limit at K_{ISR}=50 GeV

Cut in b-tag


$e^-e^+ \to q\bar{q}$ mc-2020
ILD
[%]
100
90
80
70
60
50
40
30
20
10
0
0
0.2
0.4
0.6
0.8
1
$btag_{cut}$
ε_b
$1-\varepsilon_c$
$1-\varepsilon_{uds}$
Left Pol
Right Pol

Cut in b-tag

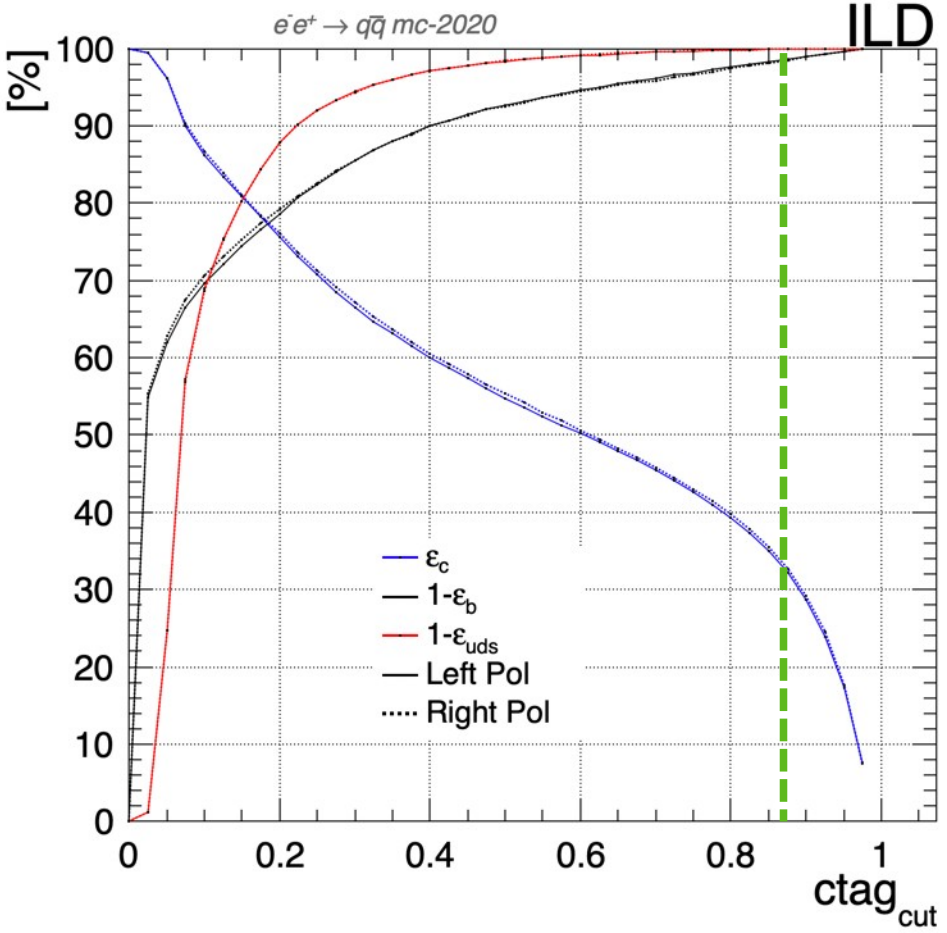

$e^-_L e^+_R \to q\bar{q}$, q=udscb, 500 GeV, 46.4 fb^{-1}
ILD
Efficiency
uū
dd̄
ss̄
bb̄
cc̄
γZ→ γqq̄ (q=udscb)
tt̄
ZZ
WW
|cos θ$_q$|
K_{reco}<50 GeV & $m_{j_1 j_2}$>250 GeV & NpfosVeto
& Cnpfos Veto & Photon Veto & y_{23}<0.005 & $m_{j_1}+m_{j_2}$<140 GeV & B-tag


$e^-_R e^+_L \to q\bar{q}$, q=udscb, 500 GeV, 47 fb^{-1}
ILD
Efficiency
uū
dd̄
ss̄
bb̄
cc̄
γZ→ γqq̄ (q=udscb)
tt̄
ZZ
WW
|cos θ$_q$|
K_{reco}<50 GeV & $m_{j_1 j_2}$>250 GeV & NpfosVeto
& Cnpfos Veto & Photon Veto & y_{23}<0.005 & $m_{j_1}+m_{j_2}$<140 GeV & B-tag

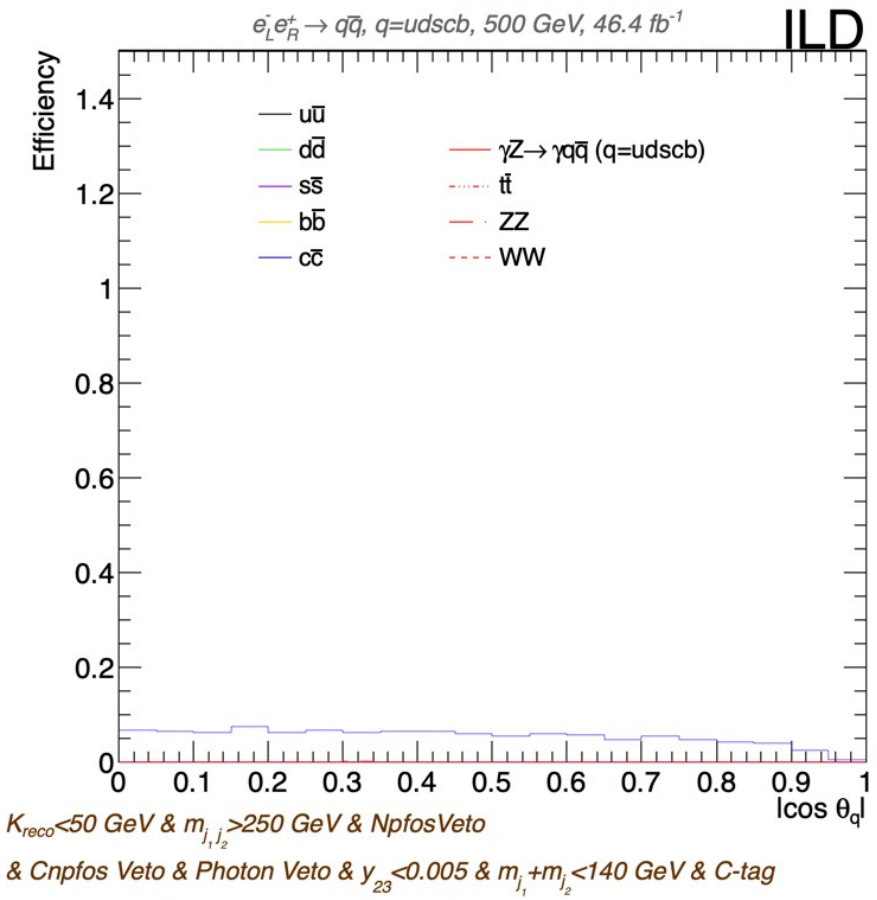
Cut in b-tag

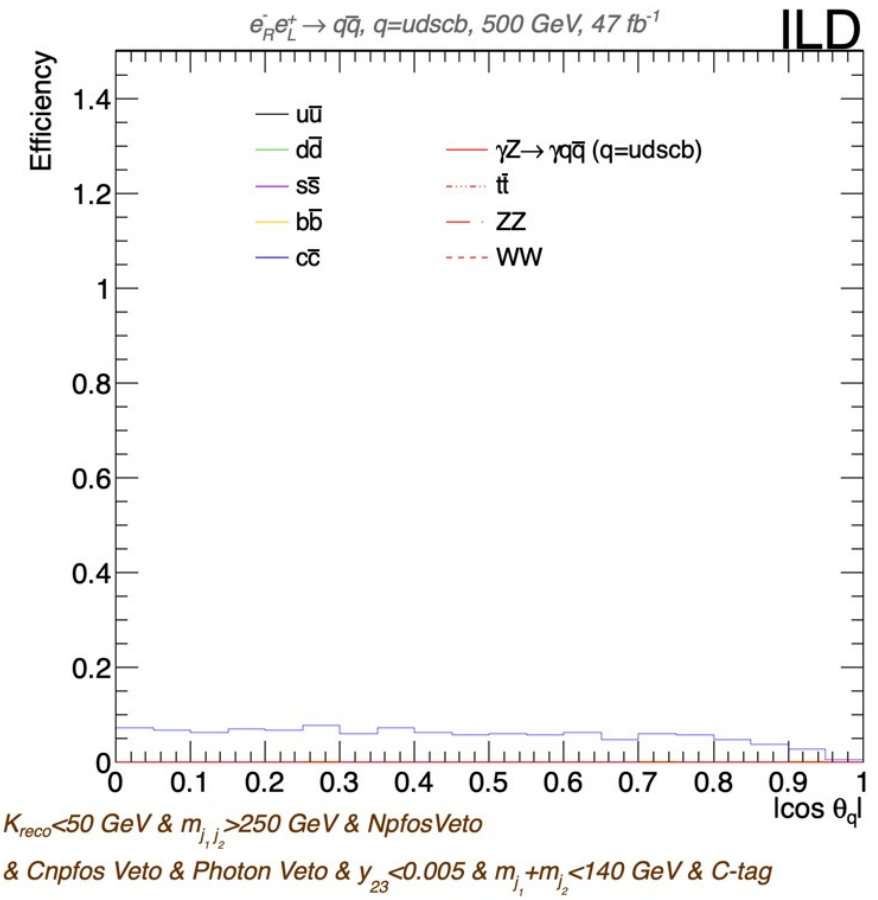
	Efficiency (%)			Background/Signal			
	$b\bar{b}$	$c\bar{c}$	$q\bar{q}$	ISR	WW	ZZ	$t\bar{t}$
$e_L p_R$	22.1	0.01	0	0.02	8e-05	3e-03	6e-06
$e_R p_L$	22.4	0.01	2e-03	0.02	0	6e-03	0

Cut in c-tag


$e^-e^+ \to q\bar{q}$ mc-2020
ILD
[%]
100
90
80
70
60
50
40
30
20
10
0
0
0.2
0.4
0.6
0.8
1
$ctag_{cut}$
ε_c
$1-\varepsilon_b$
$1-\varepsilon_{uds}$
Left Pol
Right Pol

Cut in c-tag


$e^-_L e^+_R \to q\bar{q}$, q=udscb, 500 GeV, 46.4 fb^{-1}
ILD
Efficiency
uū
dd̄
ss̄
bb̄
cc̄
γZ→ γqq̄ (q=udscb)
tt̄
ZZ
WW
|cos θq|
K_{reco}<50 GeV & $m_{j_1 j_2}$>250 GeV & NpfosVeto
& Cnpfos Veto & Photon Veto & y_{23}<0.005 & $m_{j_1}+m_{j_2}$<140 GeV & C-tag


$e^-_R e^+_L \to q\bar{q}$, q=udscb, 500 GeV, 47 fb^{-1}
ILD
Efficiency
uū
dd̄
ss̄
bb̄
cc̄
γZ→ γqq̄ (q=udscb)
tt̄
ZZ
WW
|cos θq|
K_{reco}<50 GeV & $m_{j_1 j_2}$>250 GeV & NpfosVeto
& Cnpfos Veto & Photon Veto & y_{23}<0.005 & $m_{j_1}+m_{j_2}$<140 GeV & C-tag

Cut in c-tag

	Efficiency (%)			Background/Signal			
	$b\bar{b}$	$c\bar{c}$	$q\bar{q}$	ISR	WW	ZZ	$t\bar{t}$
$e_L p_R$	0.01	5.0	0	0.02	2e-04	6e-04	0
$e_R p_L$	0.03	5.3	0	0.02	0	3e-04	0