Top identification and top pair production at CLIC

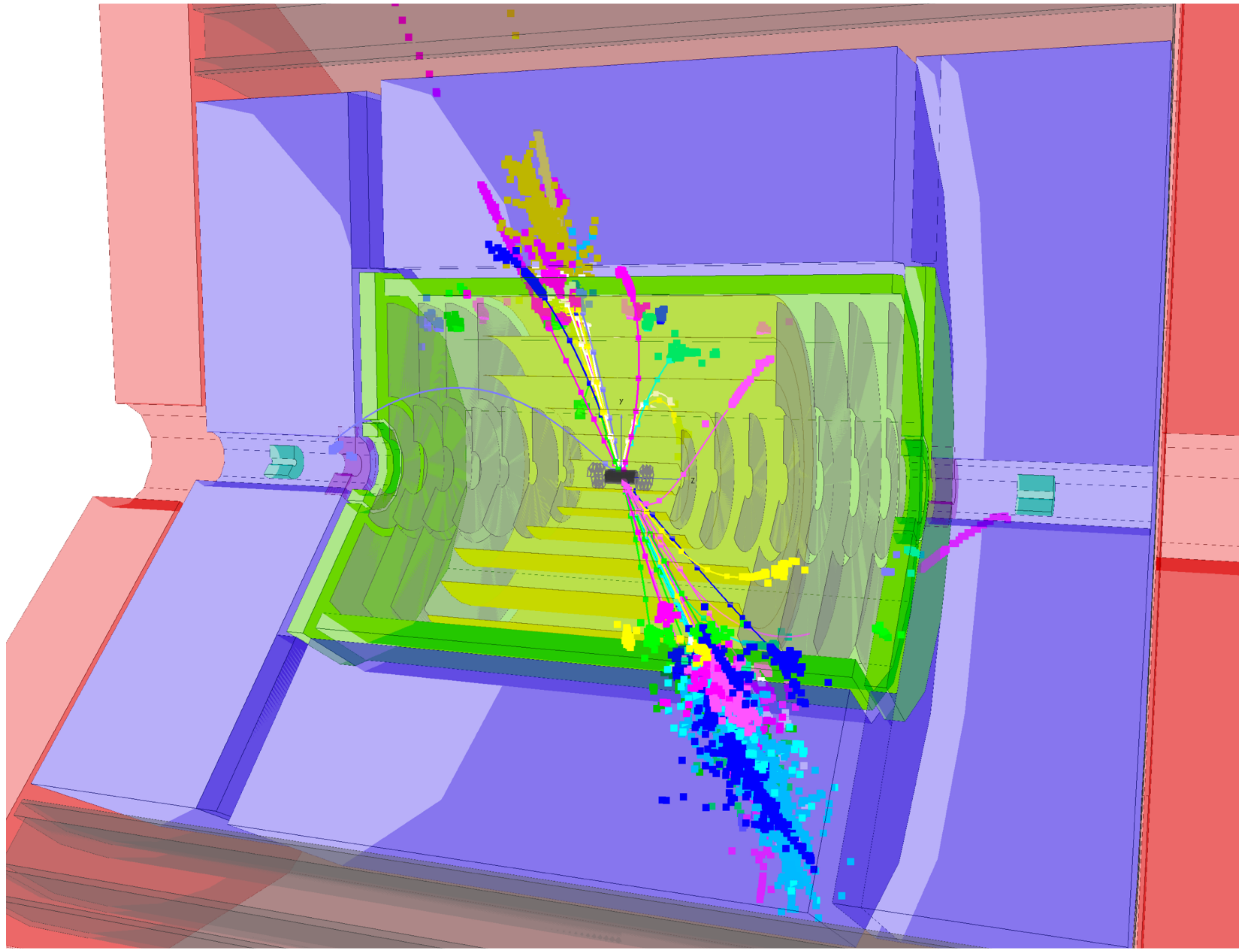
Workshop on top physics at the LC 2018,
June 4-6, 2018

On behalf of the CLICdp Collaboration

Philipp Roloff, Rickard Ström (CERN)
Martín Perelló Roselló, Ignacio Garcia Garcia, Marcel Vos (IFIC - U. Valencia/CSIC)
Nigel Watson, Alasdair Winter (University of Birmingham)

Outline

- Top pair production at CLIC
- General analysis strategy
- Analysis at 380 GeV
- Analyses above 1 TeV
  - Top identification above 1 TeV
- Results



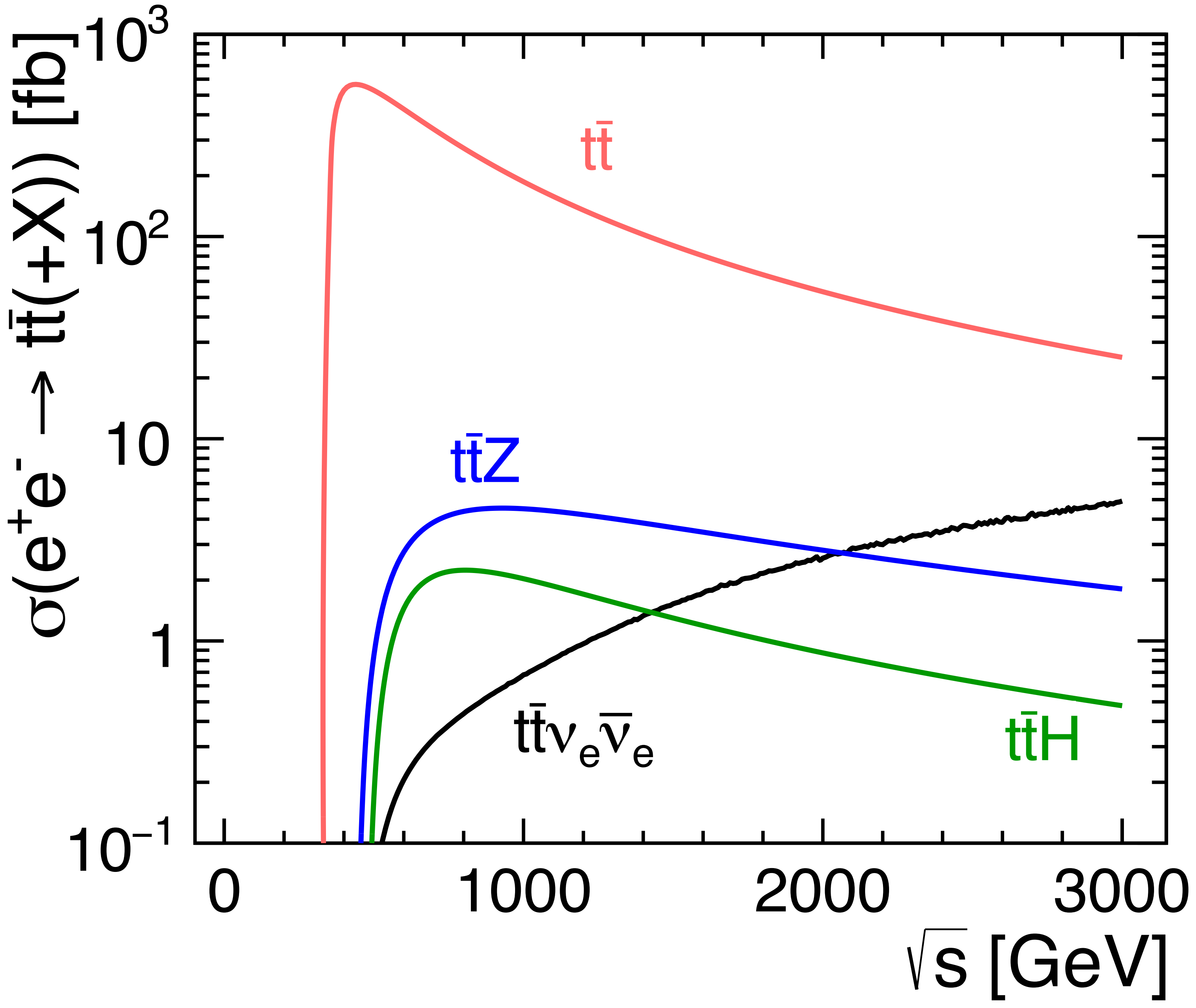
$e^+e^- \rightarrow tt \rightarrow qqqq\mu\nu_\mu$ at 3 TeV in **CLICdet**
CLICdet described in **CLICdp-Note-2017-001**

Top quark production at CLIC

- CLIC staging baseline: CERN-2016-004
- Top quarks are produced at all stages of CLIC (threshold and continuum)
- Dedicated threshold scan at 350 GeV (talk by Frank Simon tomorrow); top mass determination above threshold by ISR (talk by Pablo Gomis Lopez tomorrow)
- Top pair production up to 3 TeV; **this talk** + talk on EFT interpretation by Martín Perelló Roselló (later today)
- FCNC top decays at 380 GeV; talk by Filip Zarnecki (tomorrow)
- Associated production (ttH, VBF) > 1 TeV, talk by Yixuan Zhang on ttH (next)

- **Top physics overview paper in collaboration review (to be published 2018)**

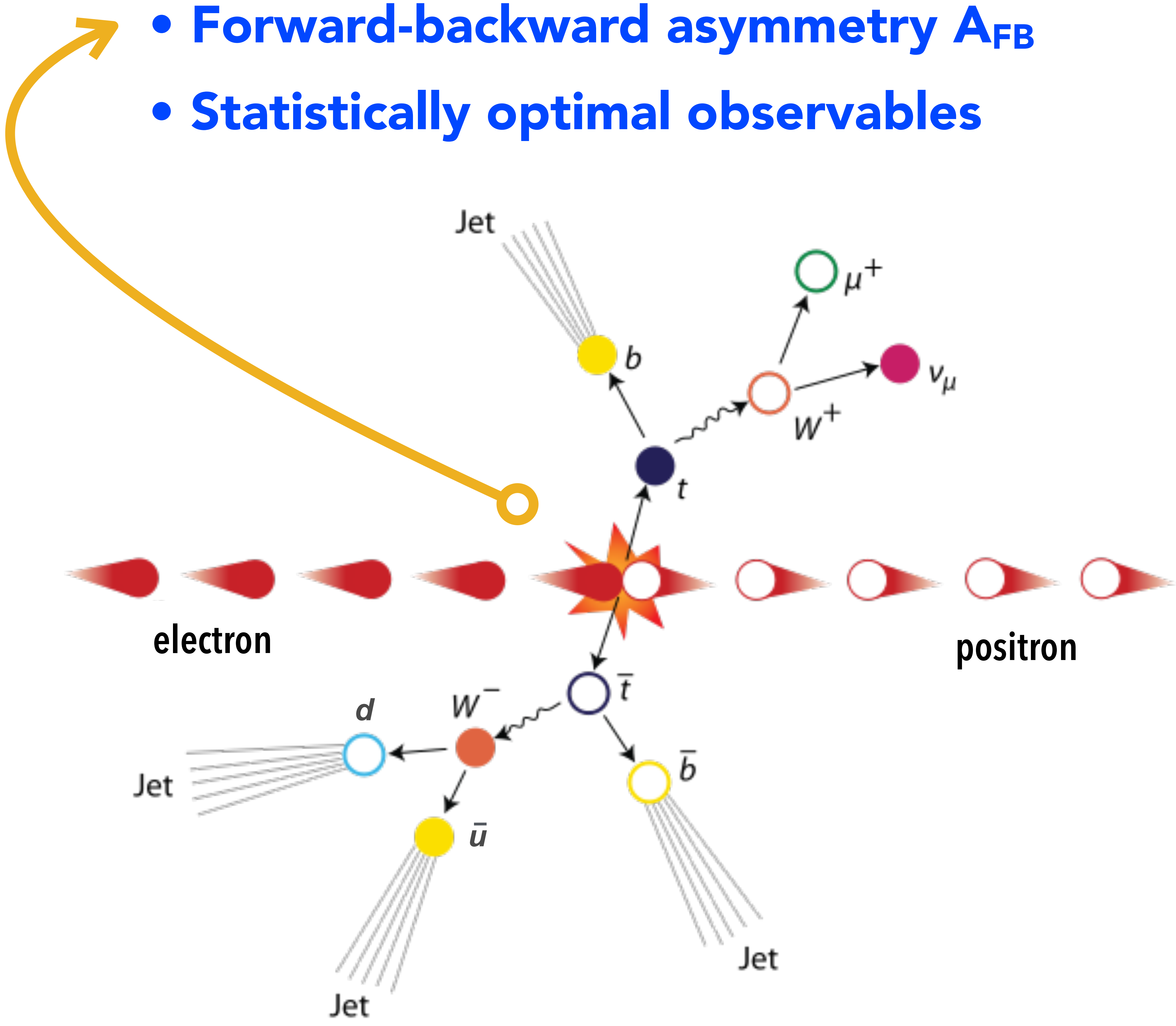
Stage	$\sqrt{s}$ [GeV]	$\mathscr{L}_{\text{int.}}$ [fb^{-1}]	L [km]
1	380	500	11,4
	350	100	
2	1500	1500	29,0
3	3000	3000	50,1



Top pair production at CLIC

- Top mass and couplings to Z and γ are among the main focuses of the initial stage of CLIC
- **New**: Extended top coupling study to higher energy stages by using methods developed for boosted topologies
- Precision measurements on anomalous electroweak couplings yield sensitivity to new physics at scales well beyond the direct reach
- Reconstruction methods also useful for other BSM searches

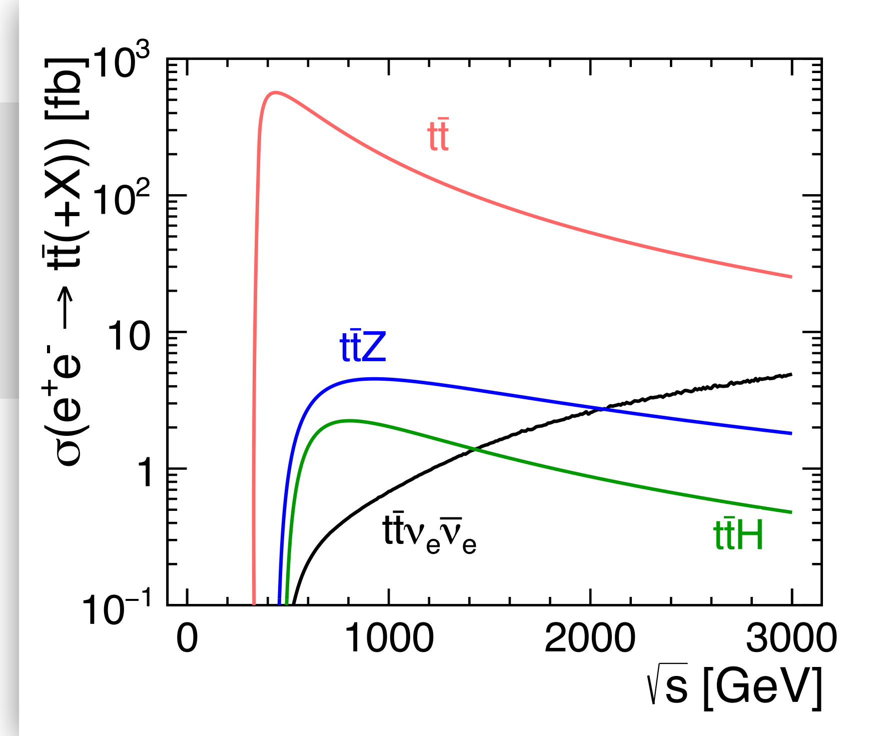
- **Cross-section σ**
- **Forward-backward asymmetry A_{FB}**
- **Statistically optimal observables**



Top pair production at CLIC

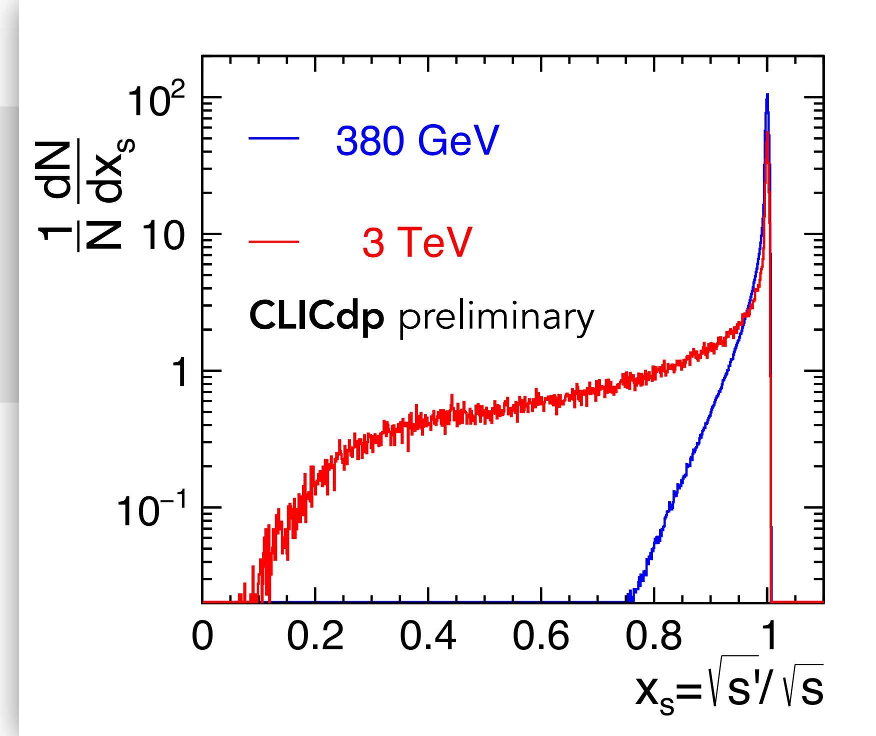
Cross section $e^+e^- \rightarrow tt$

@LO incl. ISR, unpolarised



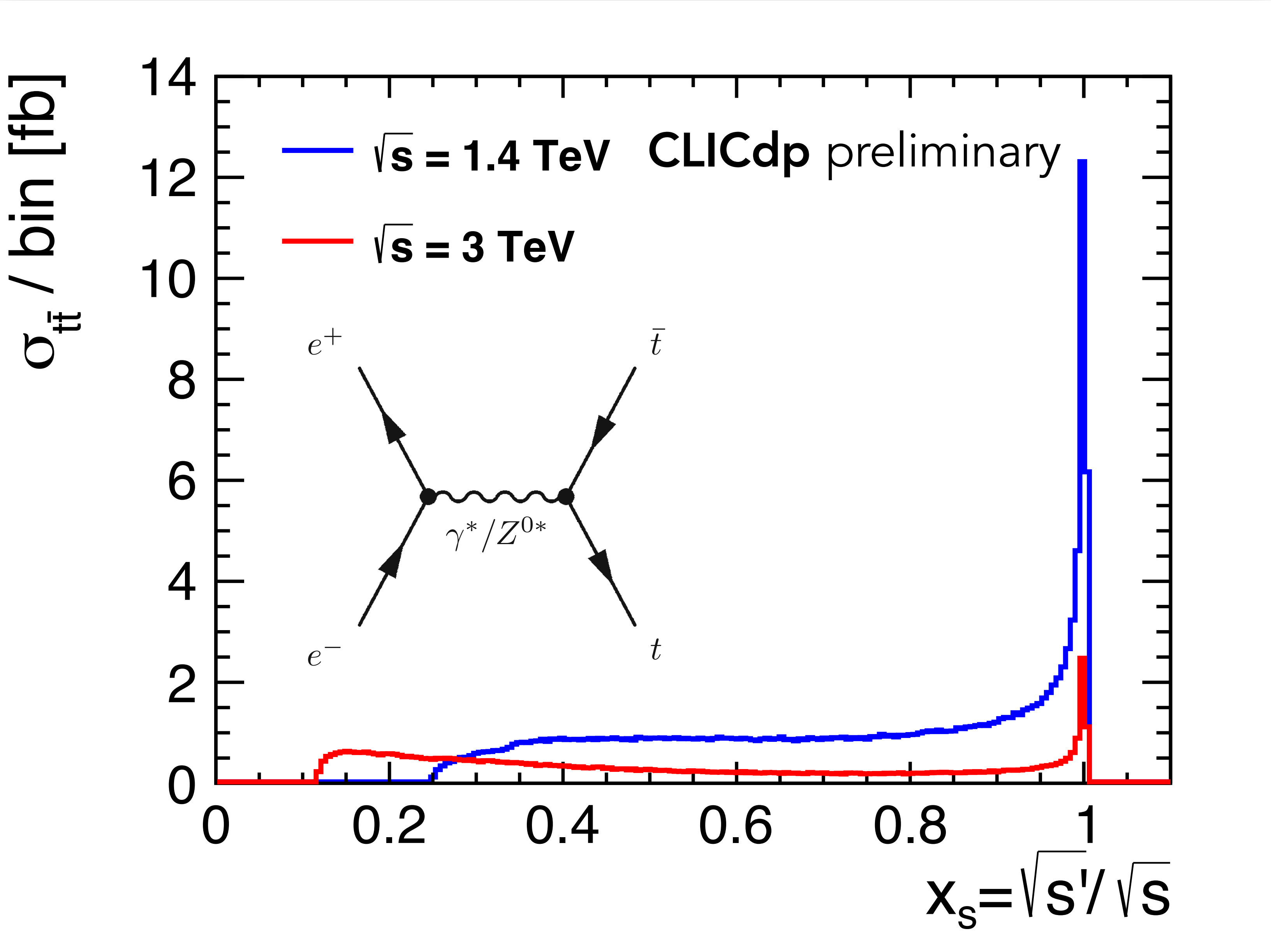
X

CLIC luminosity spectrum

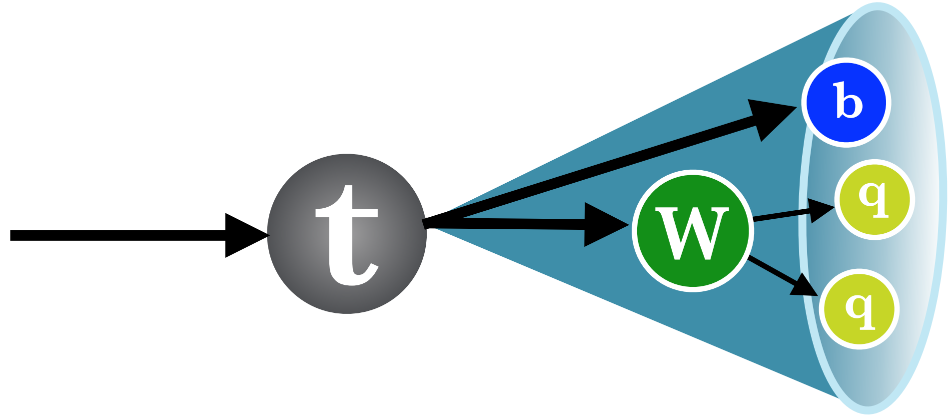


=

Effective centre-of-mass energy √s′



- Operation at √s > 1TeV leads to more highly collimated jet environments
- **Resolved strategy**
  - Production close to threshold
  - Combine jets, b-tagging, etc.
- **Boosted strategy (large R-jets/fat-jets)**
  - Standard identification techniques challenging since tracks are very close to each other and the W decay products are not isolated from each other or b-jet
  - b-tagging alone no longer viable
  - Instead tag tops by identifying prongy jet sub-structure

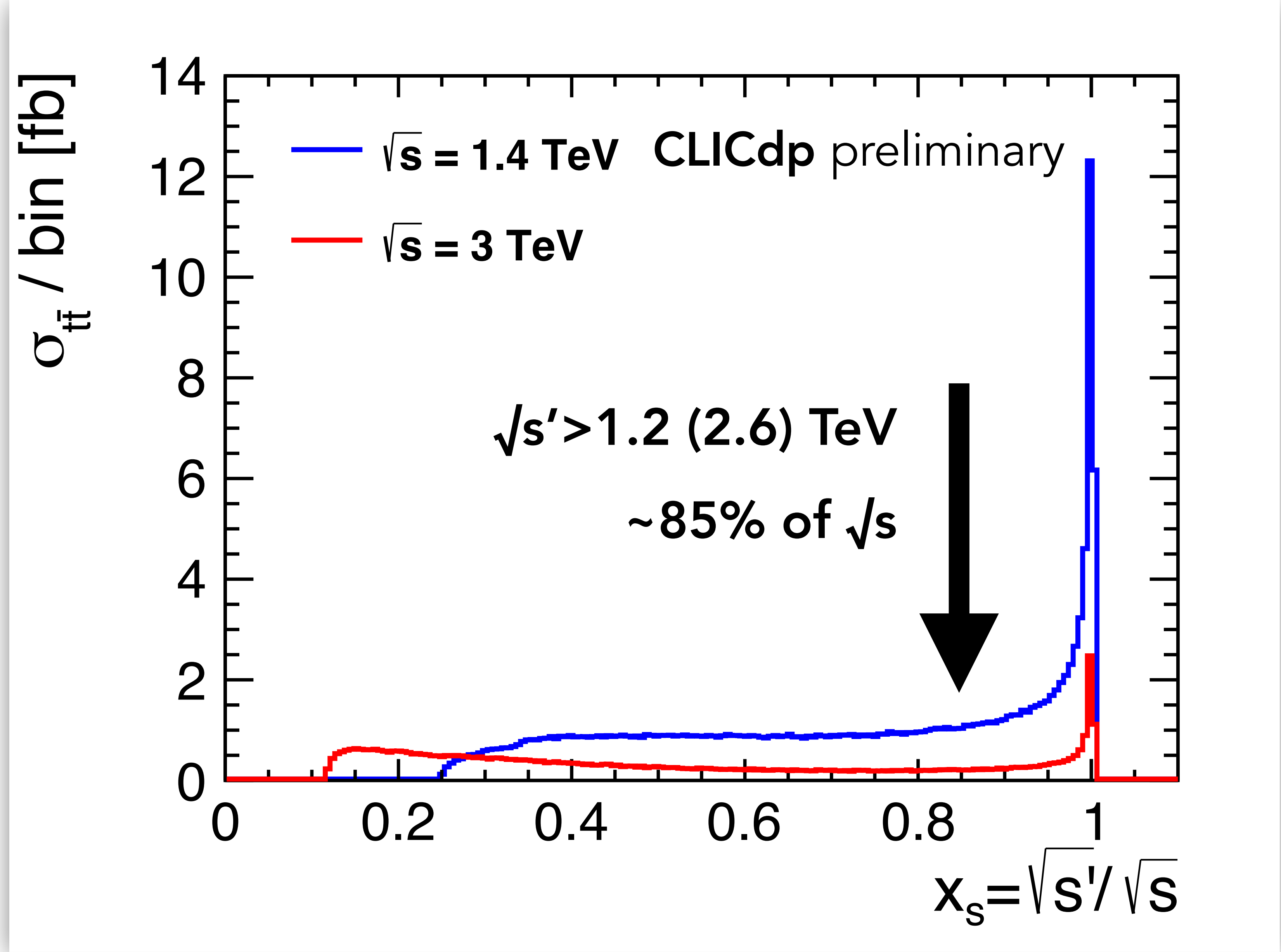


Fully merged top quark

General analysis strategy

- **Semi-leptonic ttbar events** (tt → qqqlν, l=e,μ)
- Use lepton charge to reconstruct the charge of the top/anti-top
- Six-fermion final states are generally dominated by ttbar but have a sizeable contribution from single-top and triple gauge boson production above 1 TeV
- Backgrounds considered: fully-hadronic ttbar, fully-leptonic ttbar, non-ttbar qqqlν, di-jet, WW-fusion, ZZ-fusion
- Top pair production studied at the nominal collision energies as 380 GeV, 1.4 TeV, and 3 TeV
- In addition, radiative events are studied at $\sqrt{s}$ = 1.4 TeV
- Studies done for CLIC_ILD using full simulation/reconstruction

Effective centre-of-mass energy $\sqrt{s'}$

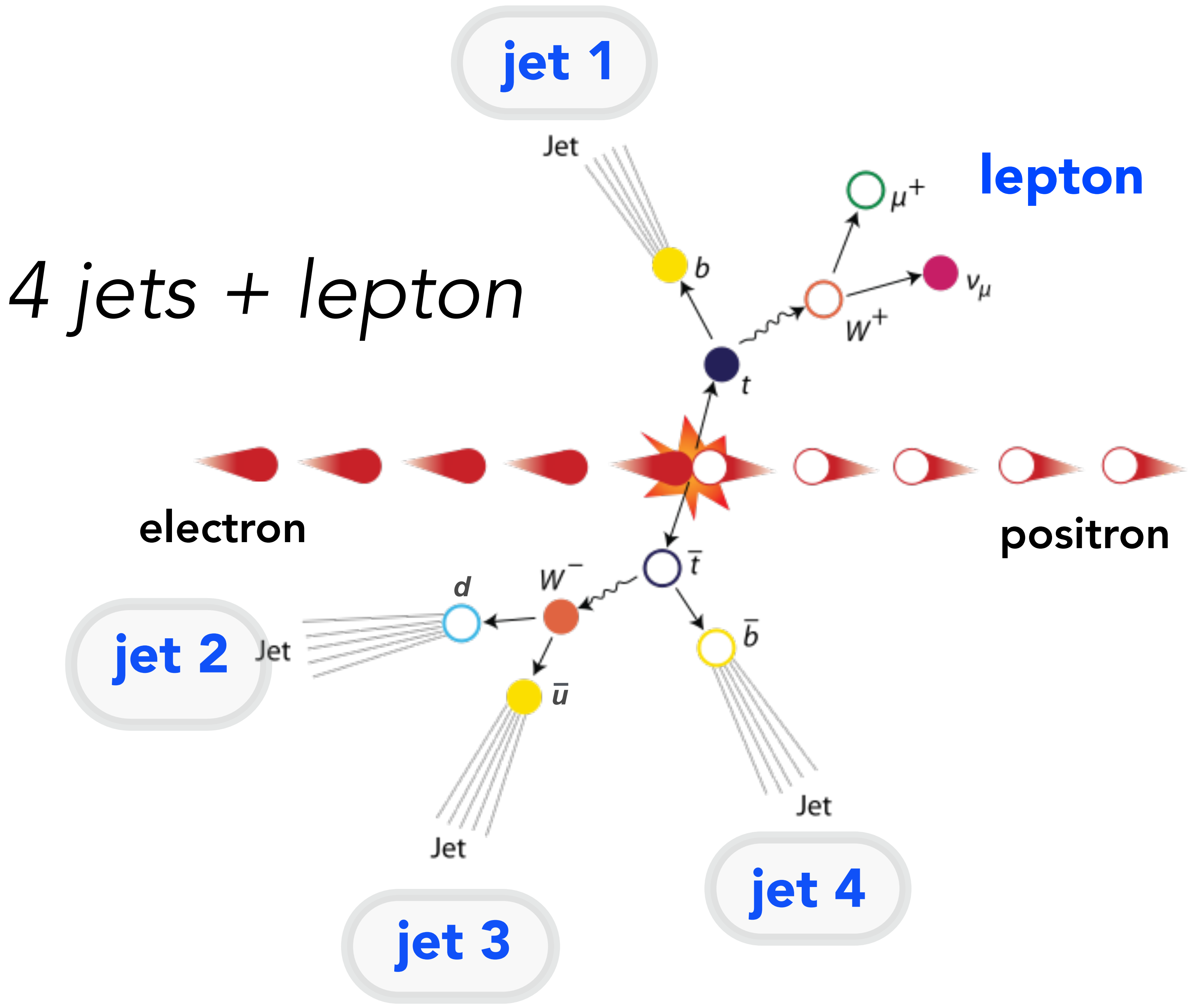


$\sqrt{s'}$>1.2 (2.6) TeV

~85% of $\sqrt{s}$

Top pair production at 380 GeV

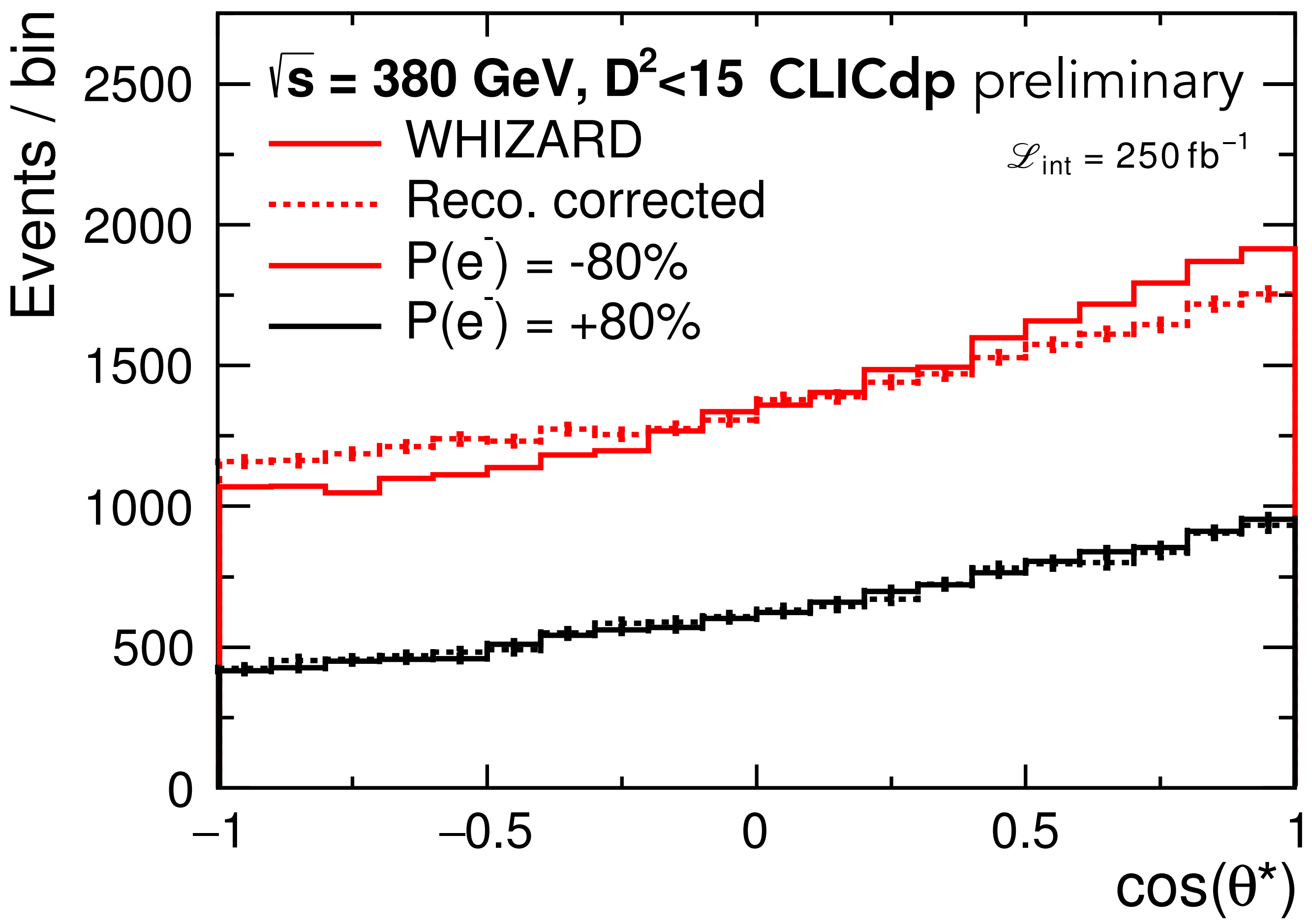
- Top-quark candidates are formed by combining jets
- Large-R possible due to clean environment, PFO clustered into **4 exclusive jets** using the VLC algorithm (Eur. Phys. J. C78 (2016) 144)
- Analysis is based on that developed in Eur. Phys. J. C75 (2015) 512
- Further details are available in CERN-THESIS-2016-214 by I. G. García

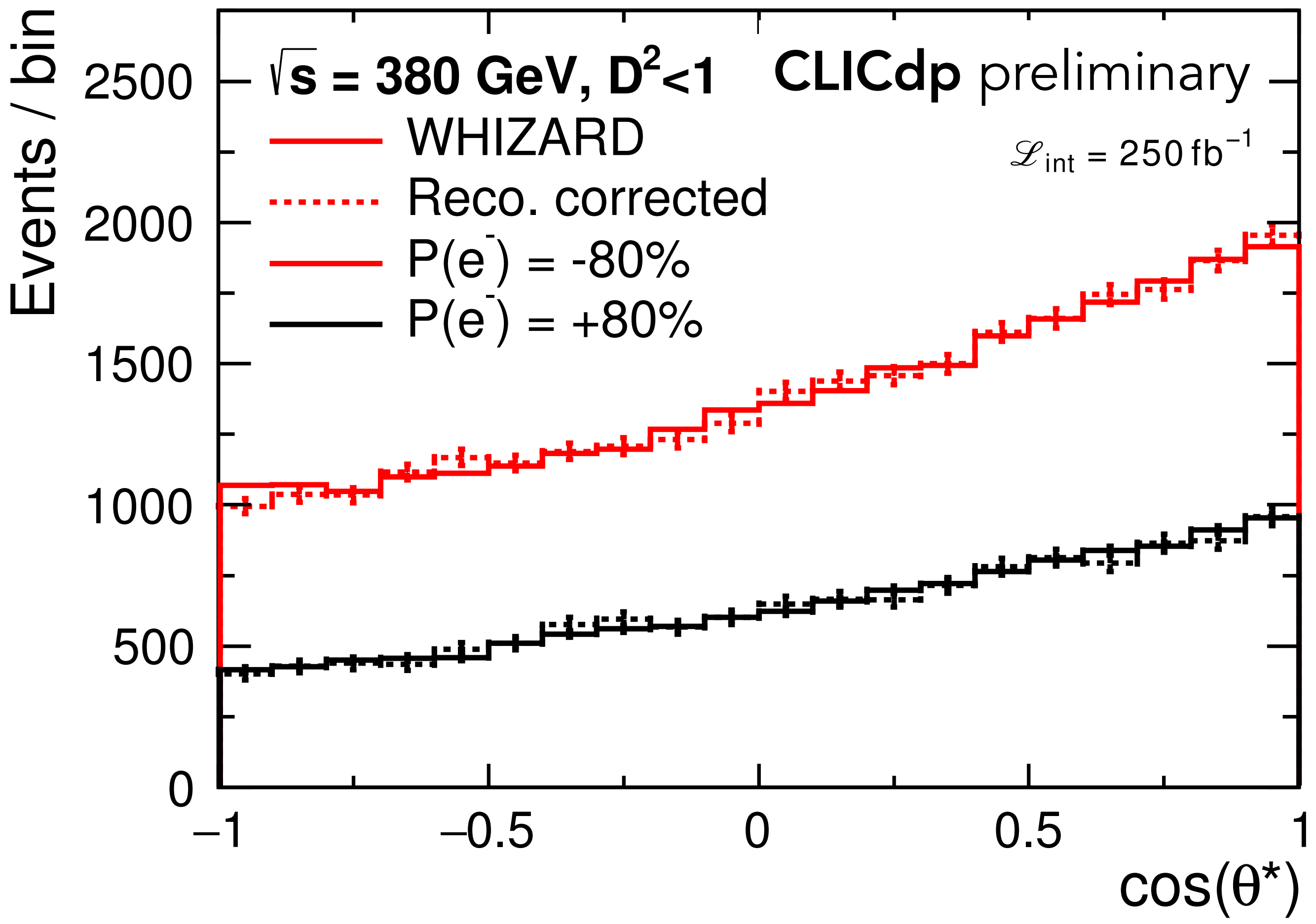

4 jets + lepton

Top pair production at 380 GeV

- **Event selection:**
  - 1 isolated lepton (electron or muon)
  - 2 b-tagged jets
  - Each of the non-b-tagged jets: $E_{\mathrm{jet}} > 15$ GeV
  - Top-quark candidates are formed by merging each b-tagged jet with the two remaining non-b-tagged jets (best candidate selected)
  - Mass cuts on the resulting system: $m_t \in [100,250]$ GeV, $m_W \in [40,190]$ GeV
  - Minimising D^2 – reduce the effect of migrations in the top polar angle distribution for A_{FB} measurement in $P(e^-) = -80\%$

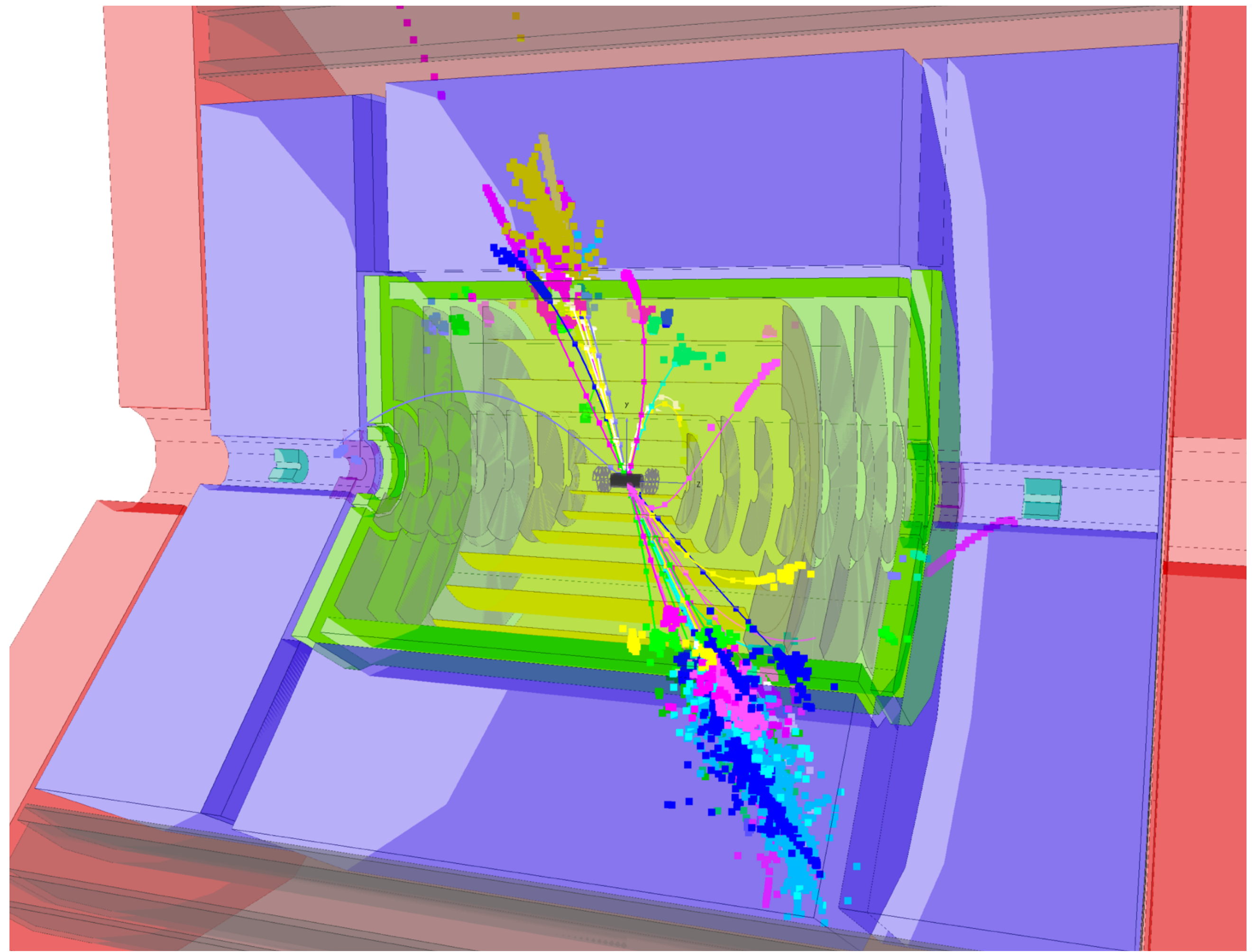
$$D^2 = \left(\frac{\gamma_{\mathrm{t}} - \langle\gamma_{\mathrm{t}}\rangle}{\sigma_{\gamma_{\mathrm{t}}}}\right)^2 + \left(\frac{E^*_{\mathrm{b}} - 68\,\mathrm{GeV}}{\sigma_{E^*_{\mathrm{b}}}}\right)^2 + \left(\frac{\cos\theta_{\mathrm{bW}} - \langle\cos\theta_{\mathrm{bW}}\rangle}{\sigma_{\cos\theta_{\mathrm{bW}}}}\right)^2$$





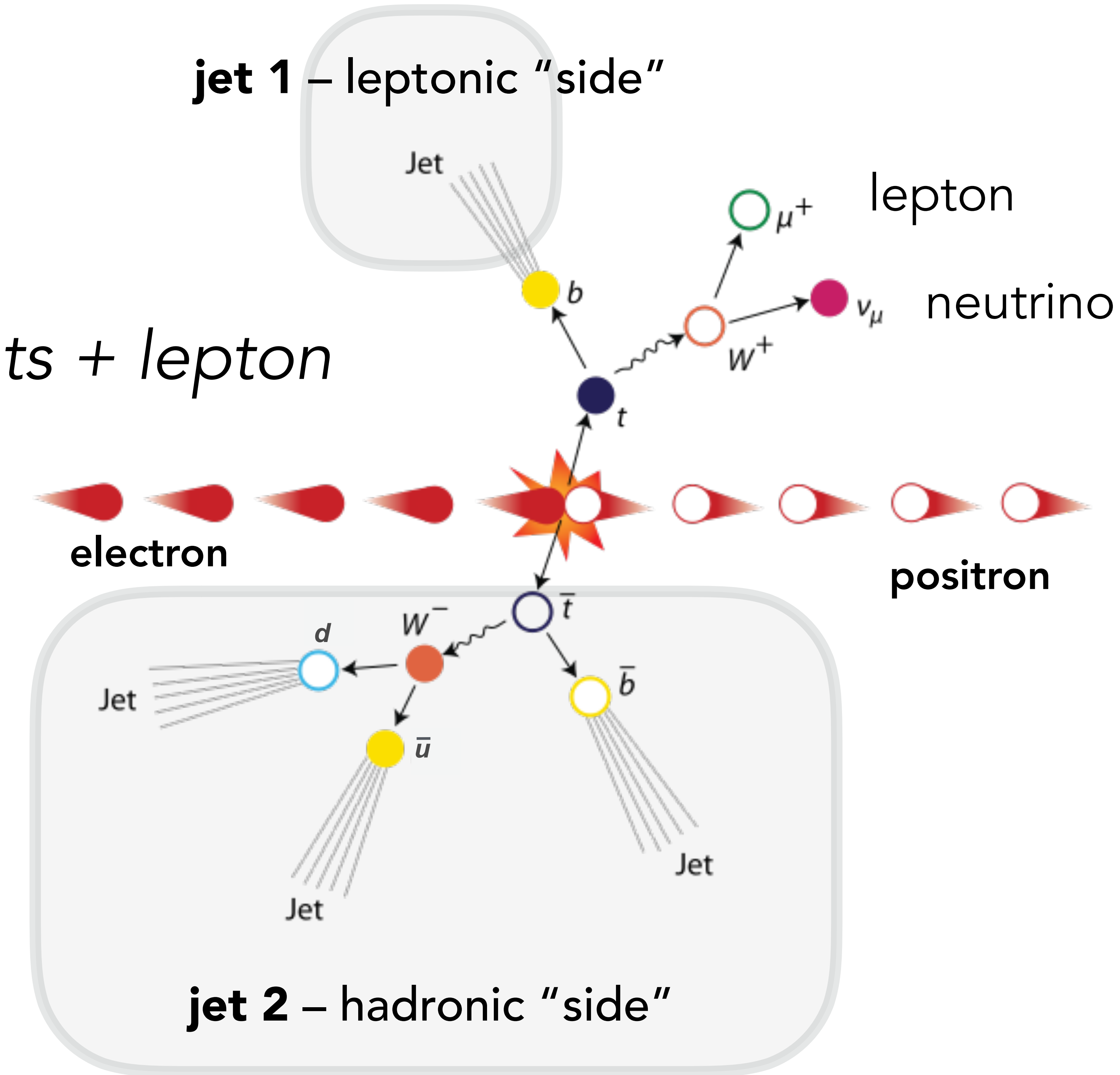
Top pair production above 1 TeV

$e^+e^- \rightarrow tt \rightarrow qqqq\mu\nu_\mu$
at 3 TeV **CLICdet**



- Event topology very different from that of top quarks produced close to the threshold that have an isotropic topology; the boosted tops are more easily distinguishable from each other → less migrations
- Large-R possible due to clean environment, jet clustering on pre-trimmed PFO collection; **2 exclusive jets** using the VLC algorithm with R=1.4(1.5), β=1.0, γ=1.0 (VLC14, VLC15)

2 large-R jets + lepton



Top identification above 1 TeV

- Reconstruction of top quark in a large-R jet with identification of substructure
- Studied ~10 years for the LHC
- New and active effort for CLIC

- **Strategy I:**

  “Top tagging – parsing jet substructure”

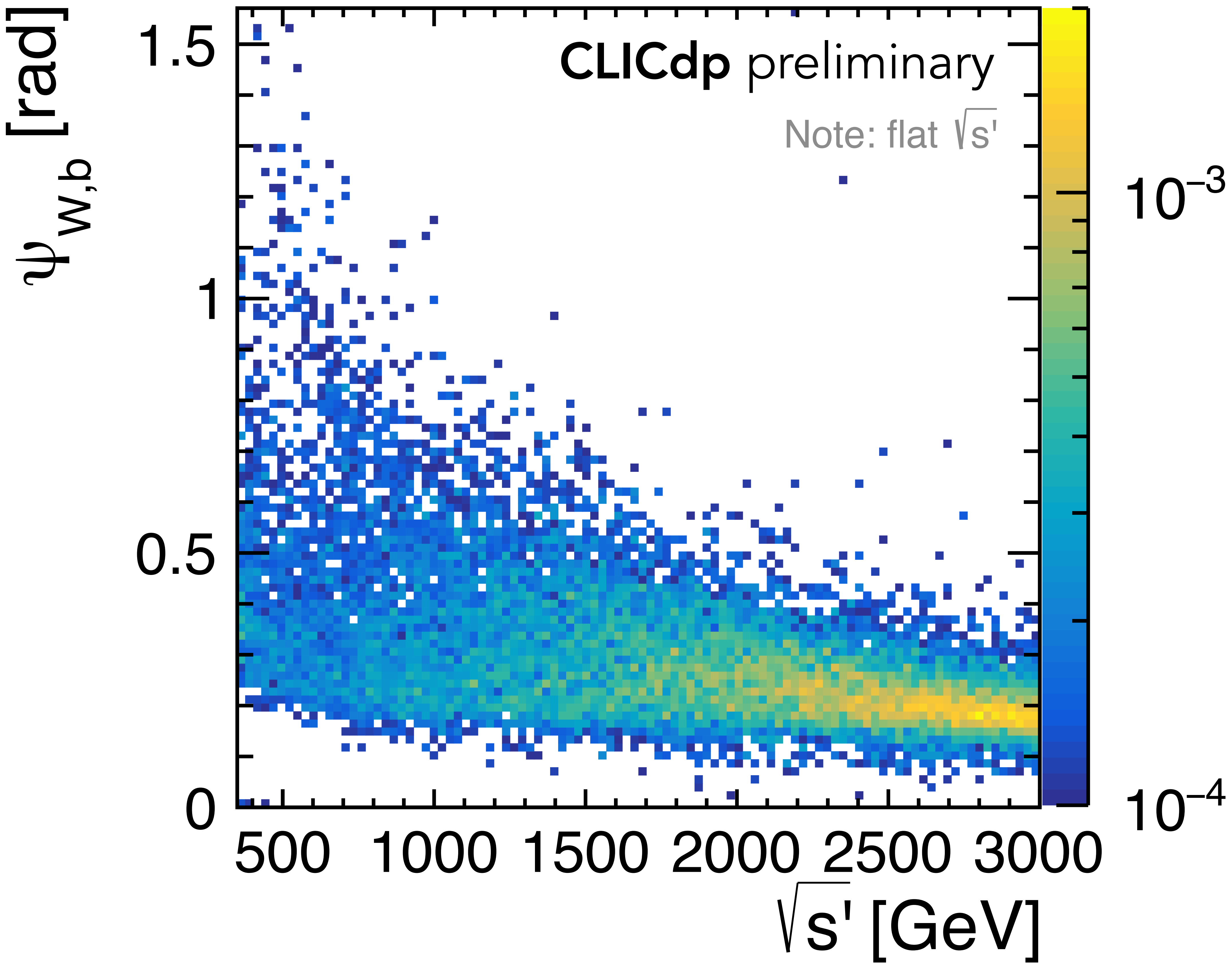
  → described here; applied for the analysis at nominal collision energy, √s, of 1.4 and 3 TeV

- **Strategy II:**

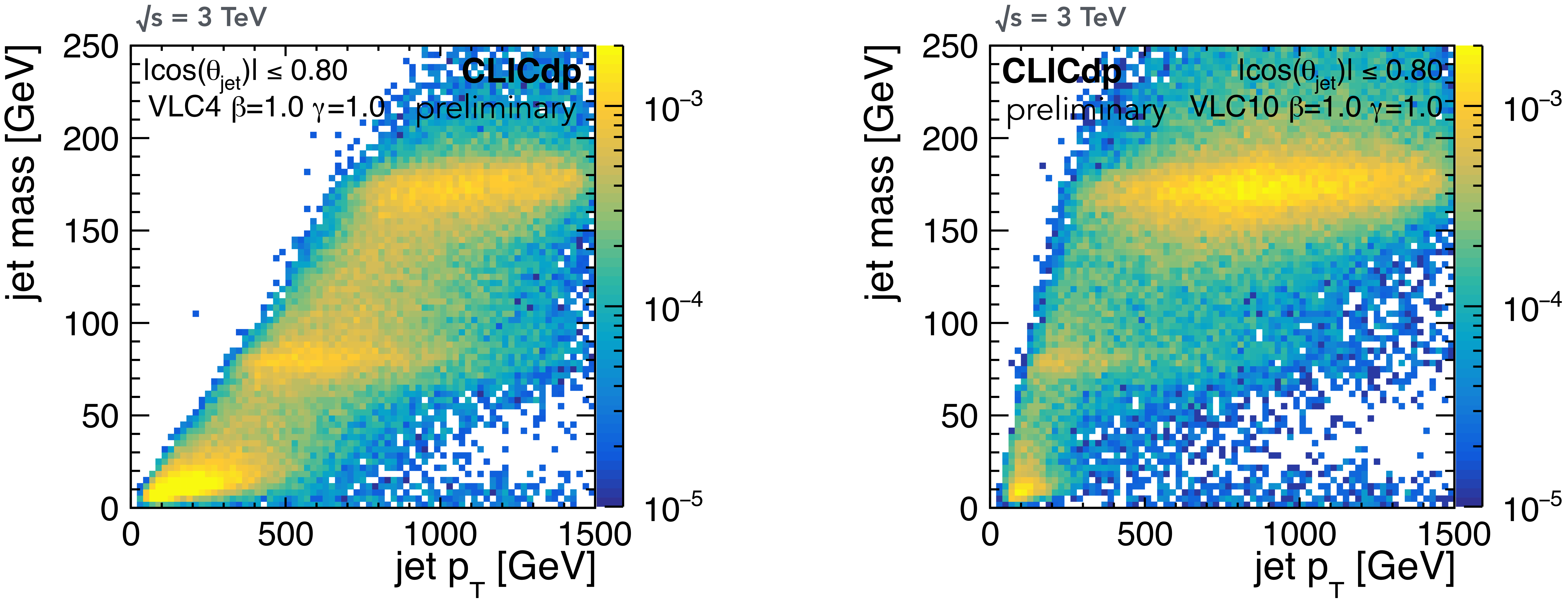
  “MVA-based – using substructure variables”

  → applied for the analysis of radiative events at √s = 1.4 TeV (three intervals below √s)

Angle between top decay products: W and b



Jet clustering optimisation



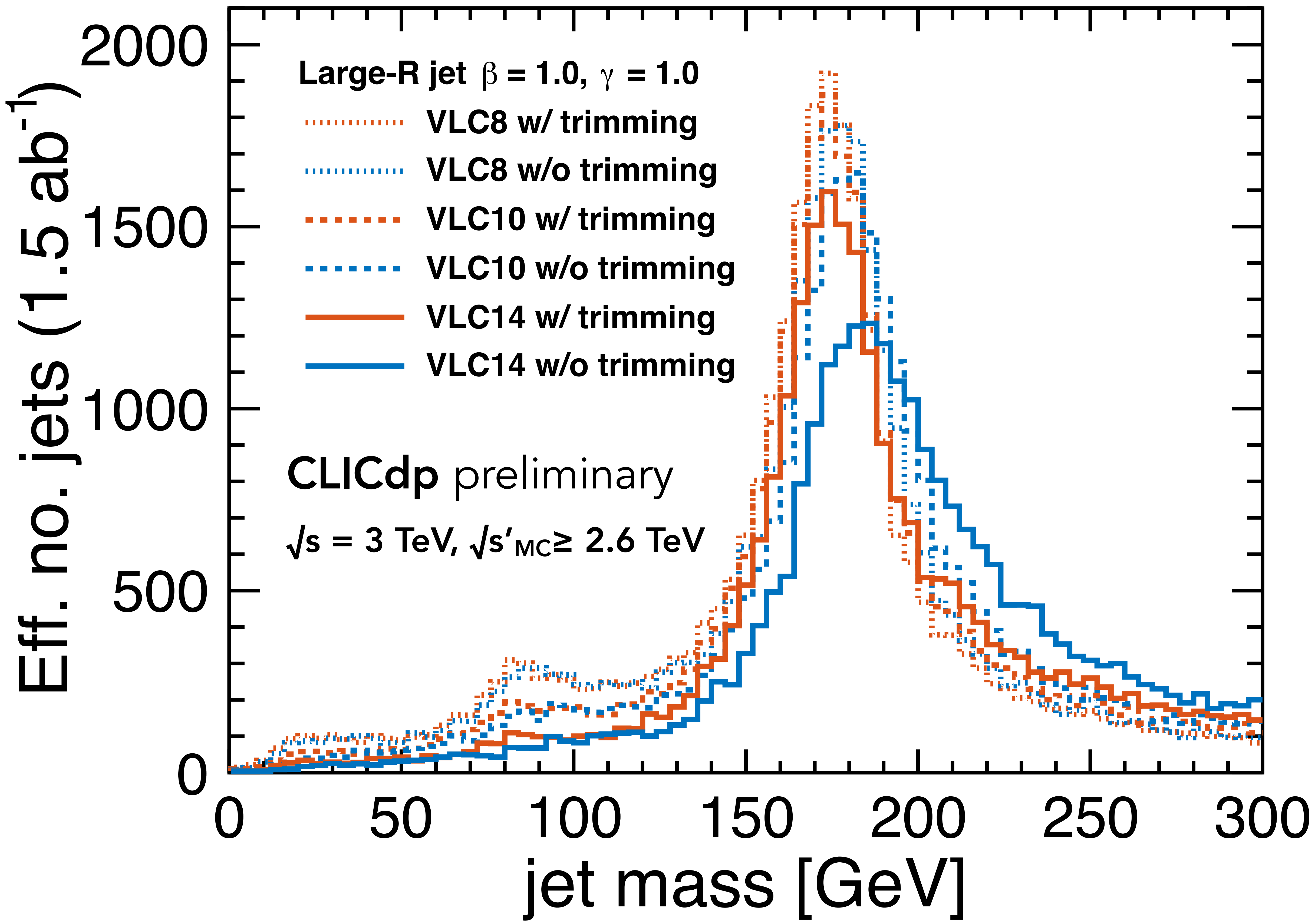
- Tendency of higher √s′ events to be more collimated
- Large jet radius needed

-see CLICdp-Note-2017-007 by Sascha Dreyer (for WW/ZZ: optimal R~0.8)

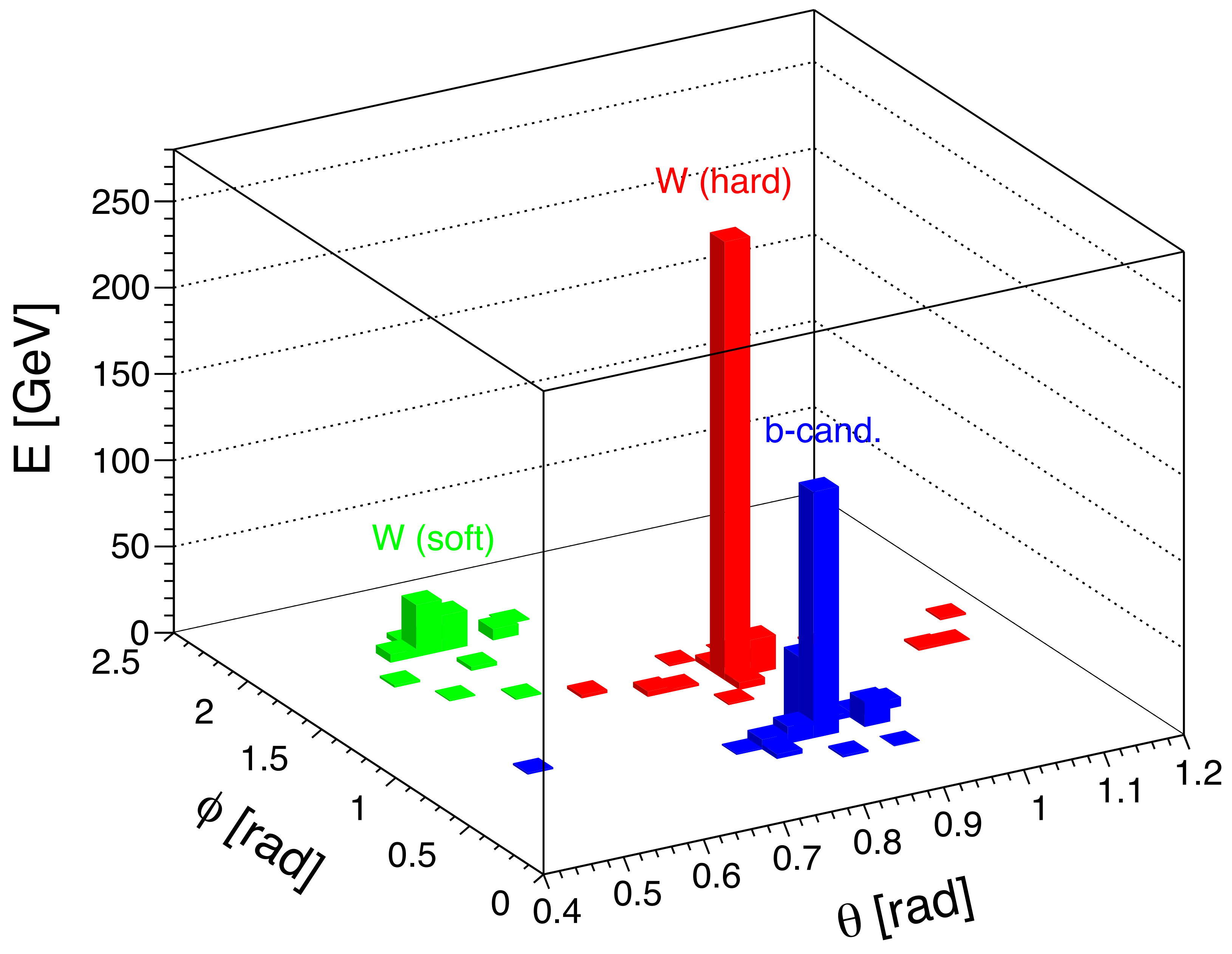
-adding b-quark → R=1.4(1.5) at 1.4 TeV (radiative), and R=1.0 at 3 TeV

Jet clustering optimisation

- Jet parameter optimisation done on fully-hadronic ttbar events (qqqqqq)
- VLC algorithm with 2 exclusive jets
  - The γ parameter - evolution of beam jet area with polar angle → 1.0
  - The β parameter - changes the clustering order → 1.0
- Pre-trimming of PFO collection useful; complementary way to reduce the impact from beamstrahlung
  - Pre-clustering into micro-jets (R=0.4); inclusive clustering with minimum p_T threshold (E_{th}=5 GeV)
  - Used for results at nominal collision energies: 1.4 TeV and 3 TeV



Strategy I: Top tagging



- The top-tagging is based on the Johns-Hopkins top tagger as implemented in FastJet (DOI: 10.1103/PhysRevLett.101.142001)
  - Iteratively decluster jet to search for true three- or four-prongy structures (de-clustering steered by two parameters: δ_R, δ_P)
  - Mass cuts on the resulting system:

    $m_t \in [120,230]$ GeV, $m_W \in [50,110]$ GeV
- Since the background level is generally low at a lepton collider, the top tagger was optimised for high signal efficiency: qqqqqq (fully-hadronic ttbar sample). Previous studies show that MVA are powerful in removing remaining background.
- Optimisation: minimise four-quark background to provide **70% integrated signal efficiency for jet $p_T \in [500,1500]$ GeV and $\cos(\theta_{jet}) \leq 0.8$**
- Corresponding background efficiency: 4.4% (8.8%) for jets from the four-quark (di-jet) events

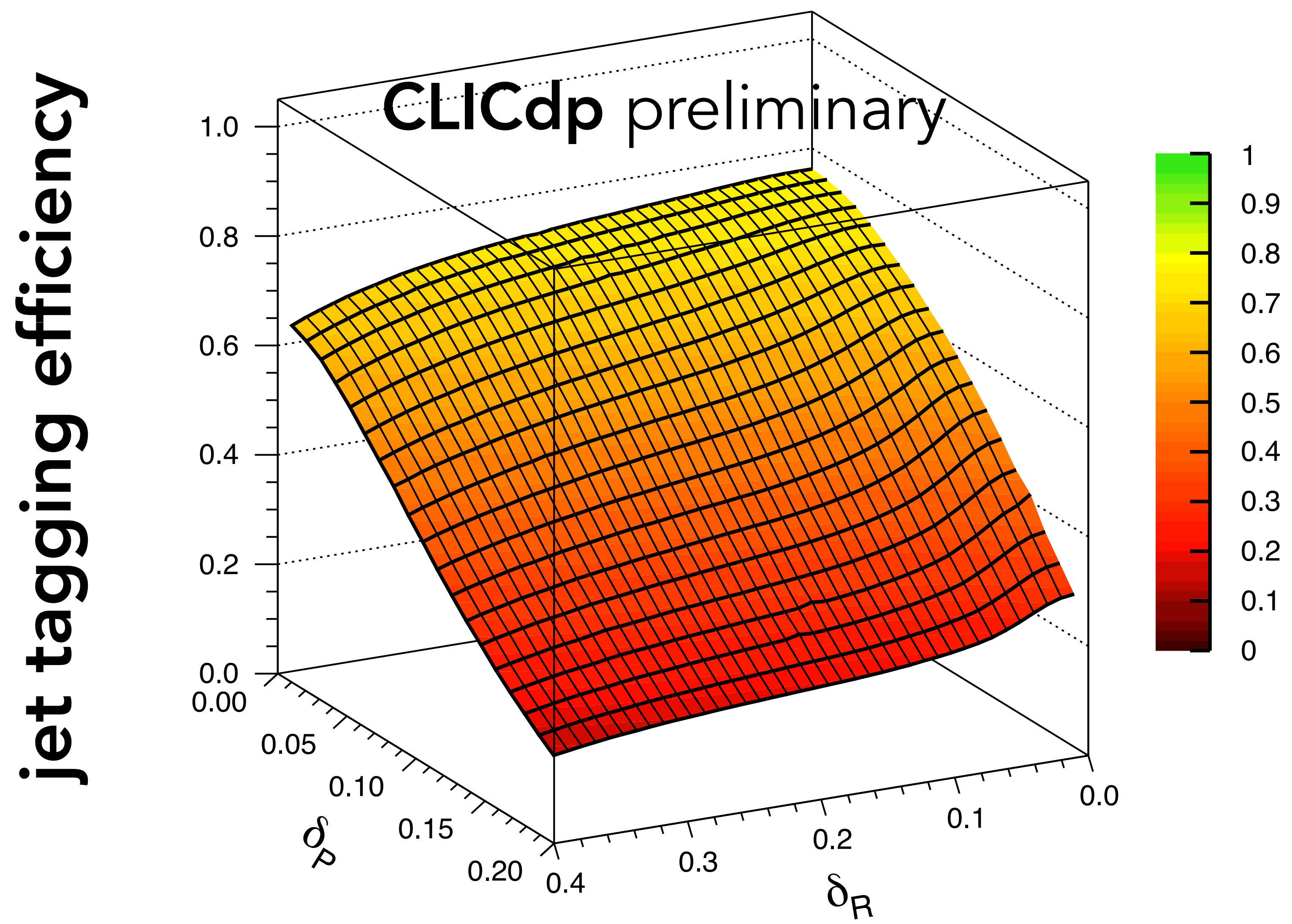
Top tagger optimisation at 1.4 TeV

- VLC14 (β=1.0, γ=1.0), 2 exclusive, 0.00≤|cos(theta)|≤0.80
- $50 \leq m_W \leq 110$, $120 \leq m_{top} \leq 230$ (loose)
- Hadronic tops optimised against hadronic W/Z background
- Optimal parameters: δR = 0.25, δP = 0.03

Optimisation: minimise four-quark background to provide **70% integrated signal efficiency for jet $p_T \in [500,1500]$ GeV and $\cos(\theta_{jet}) \leq 0.8$**

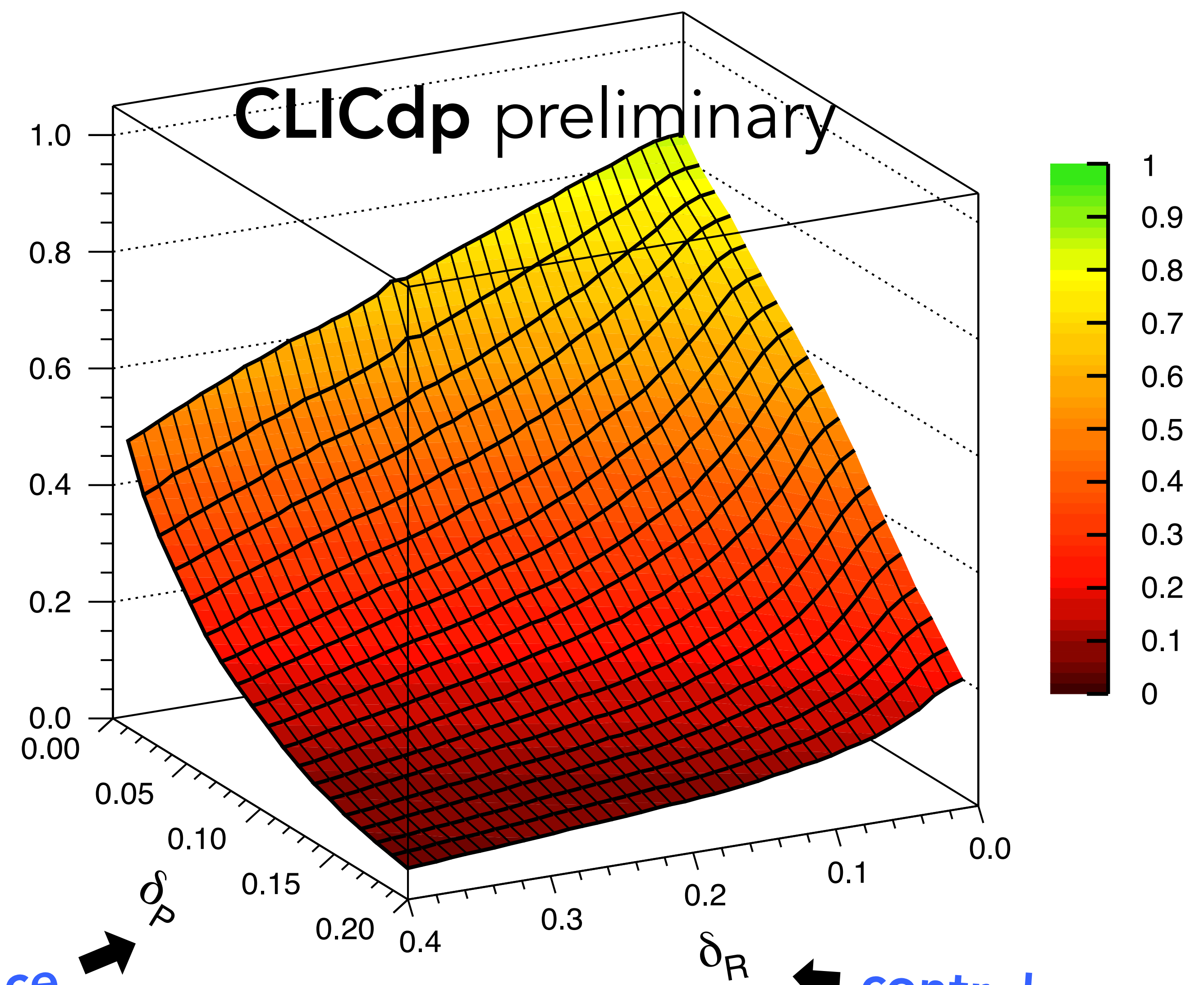
Signal

$e^+e^- \rightarrow tt \rightarrow qqqqqq$

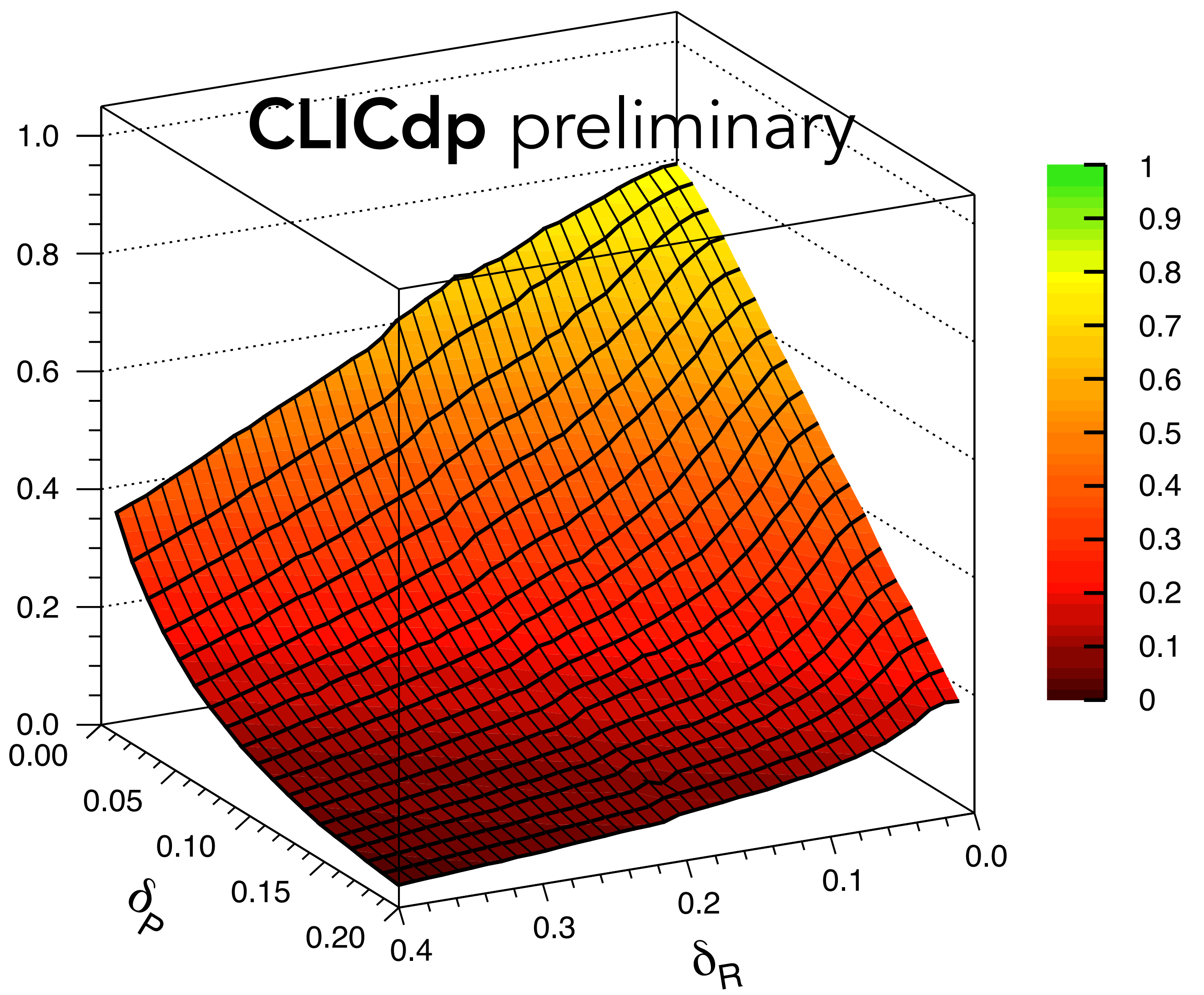


Background

$e^+e^- \rightarrow qqqq$ (10 x)



$e^+e^- \rightarrow qq$ (5 x)



controls soft-split acceptance ➡

⬅ controls spacial acceptance

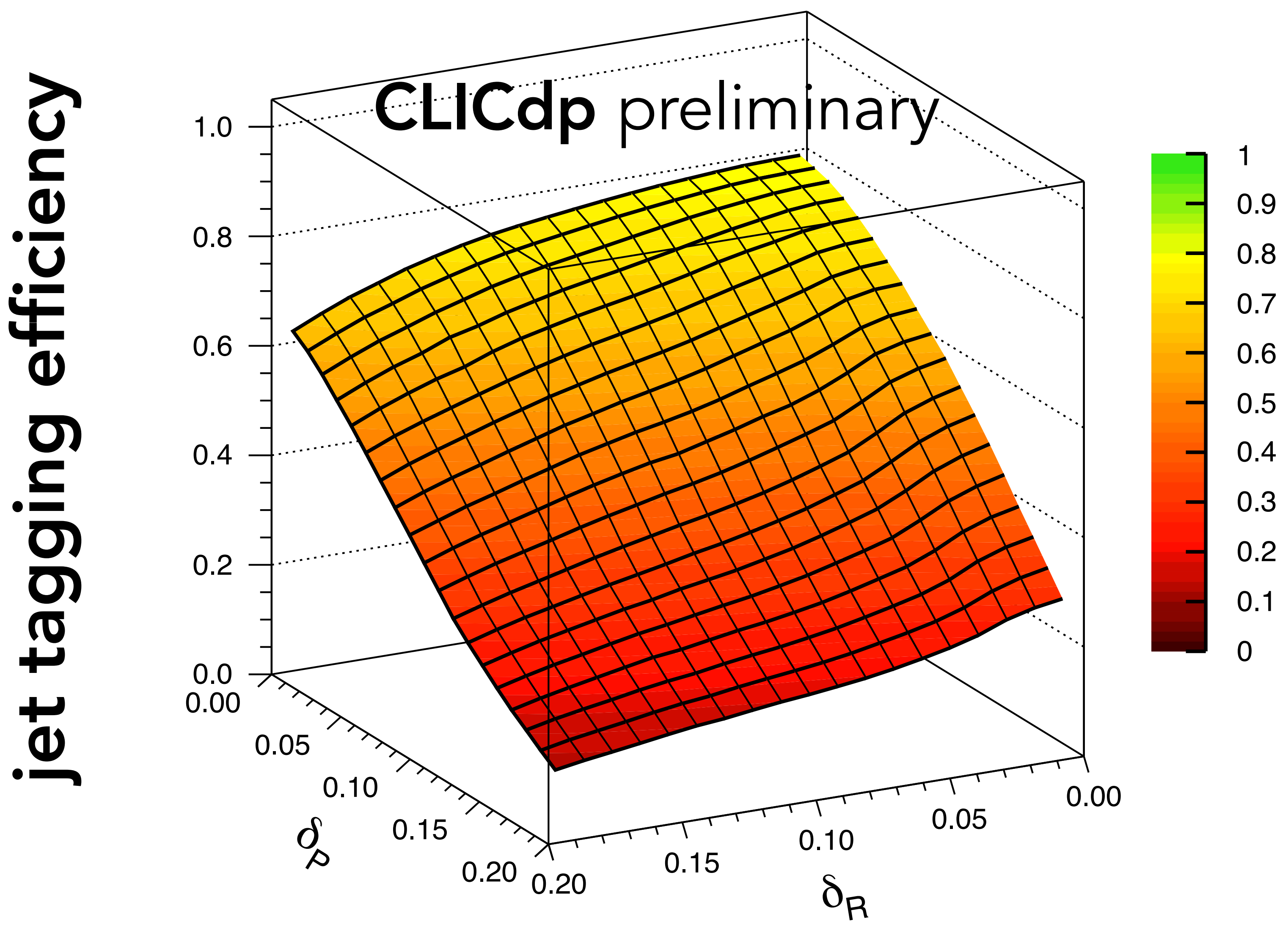
Top tagger optimisation at 3 TeV

- VLC10 (β=1.0, γ=1.0), 2 exclusive, 0.00≤|cos(theta)|≤0.80
- $50 \leq m_W \leq 110$, $120 \leq m_{top} \leq 230$ (loose)
- Hadronic tops optimised against hadronic W/Z background
- Optimal parameters: $\delta R = 0.11$, $\delta P = 0.03$

Optimisation: minimise four-quark background to provide **70% integrated signal efficiency for jet $p_T \in [500,1500]$ GeV and $\cos(\theta_{jet}) \leq 0.8$**

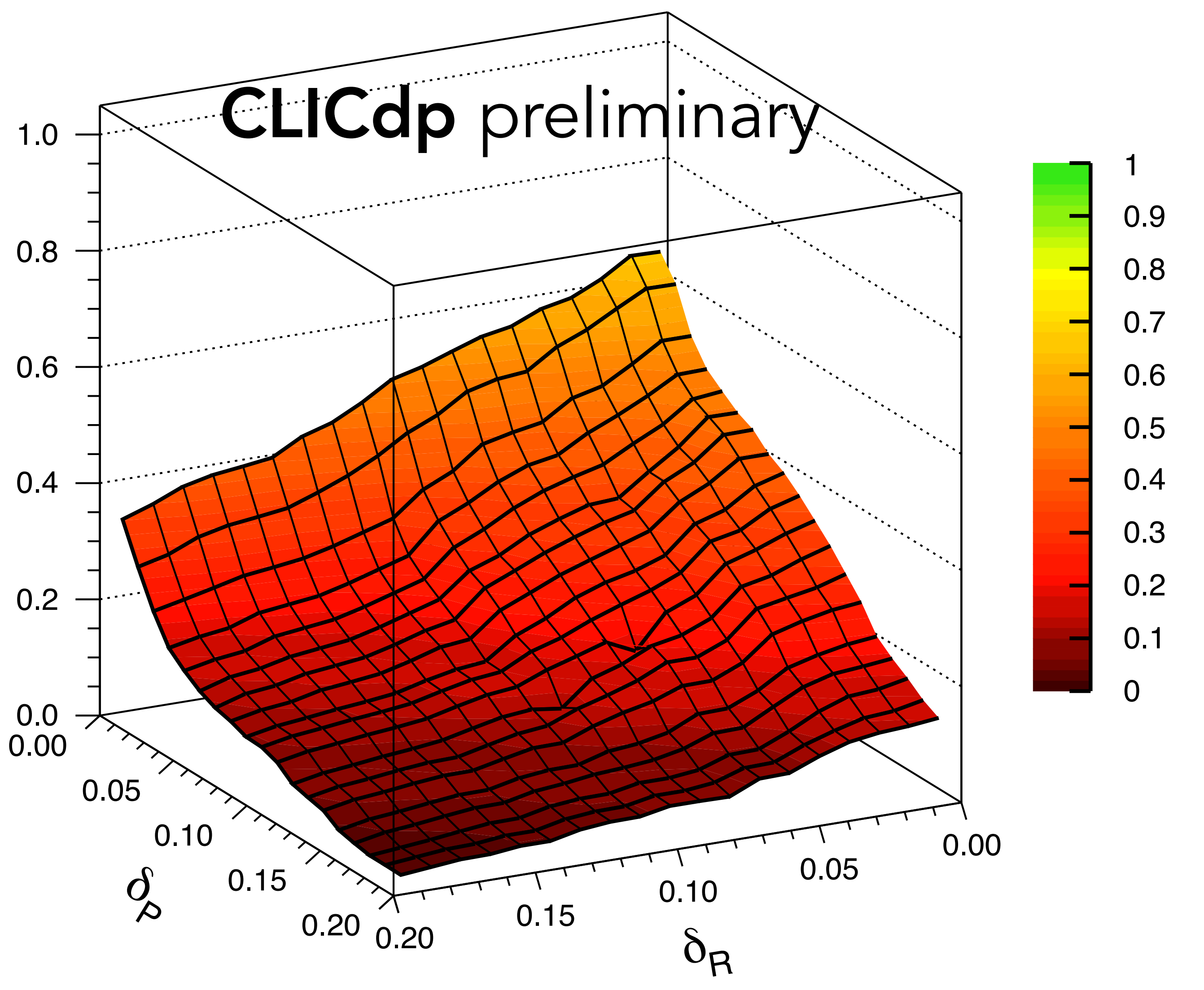
Signal

$e^+e^- \to tt \to qqqqqq$

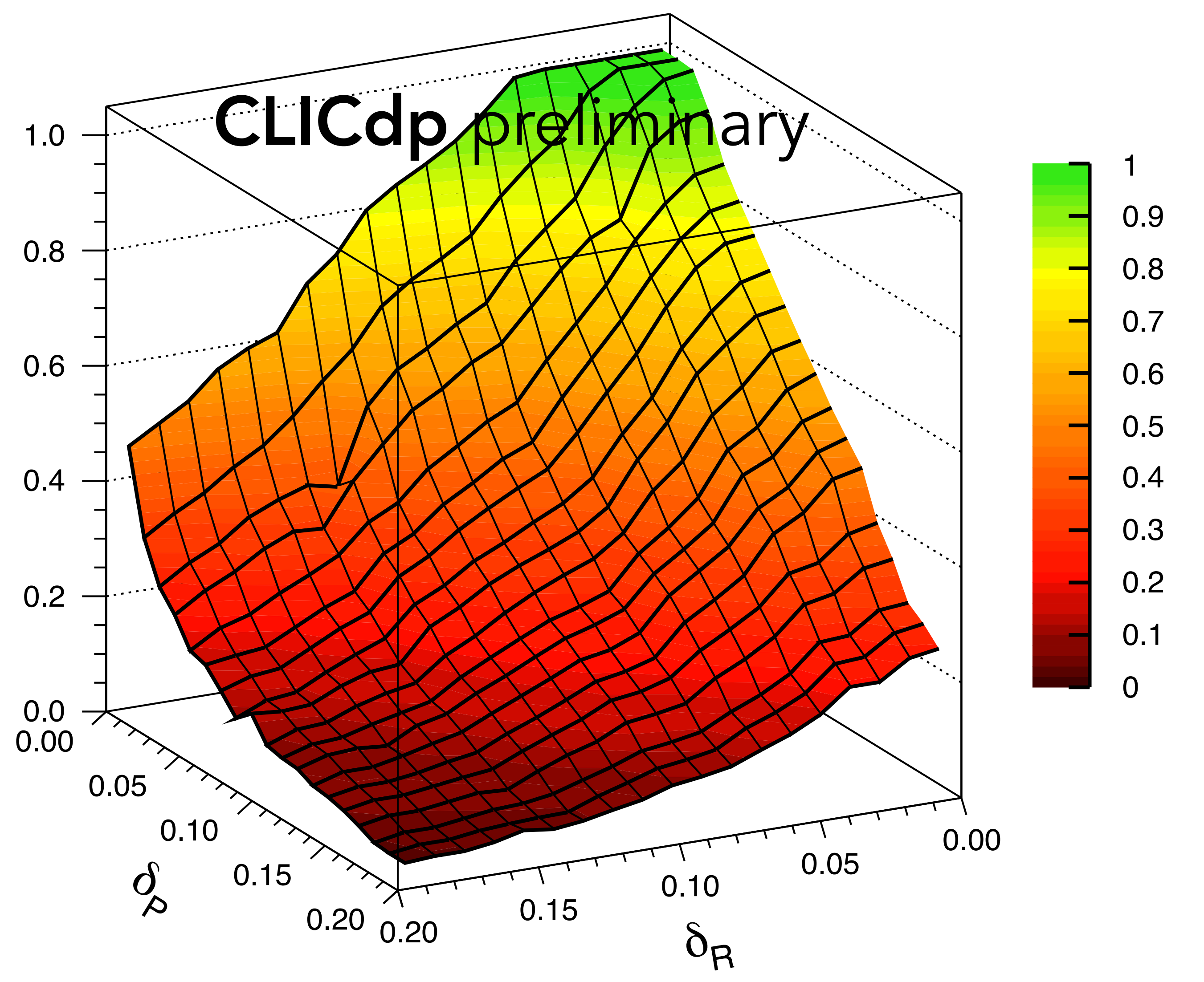


Background

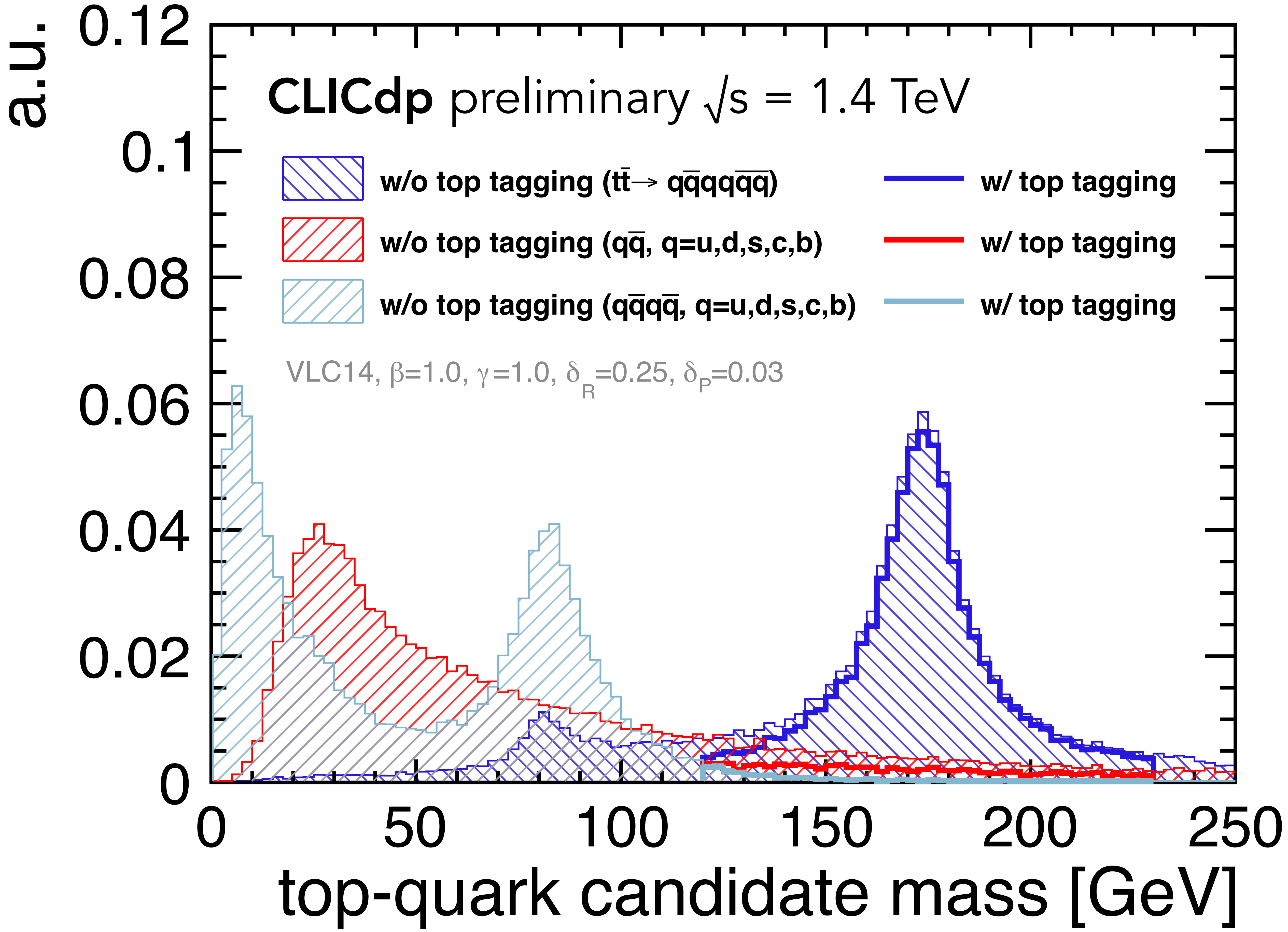
$e^+e^- \to qqqq$ (10 x)

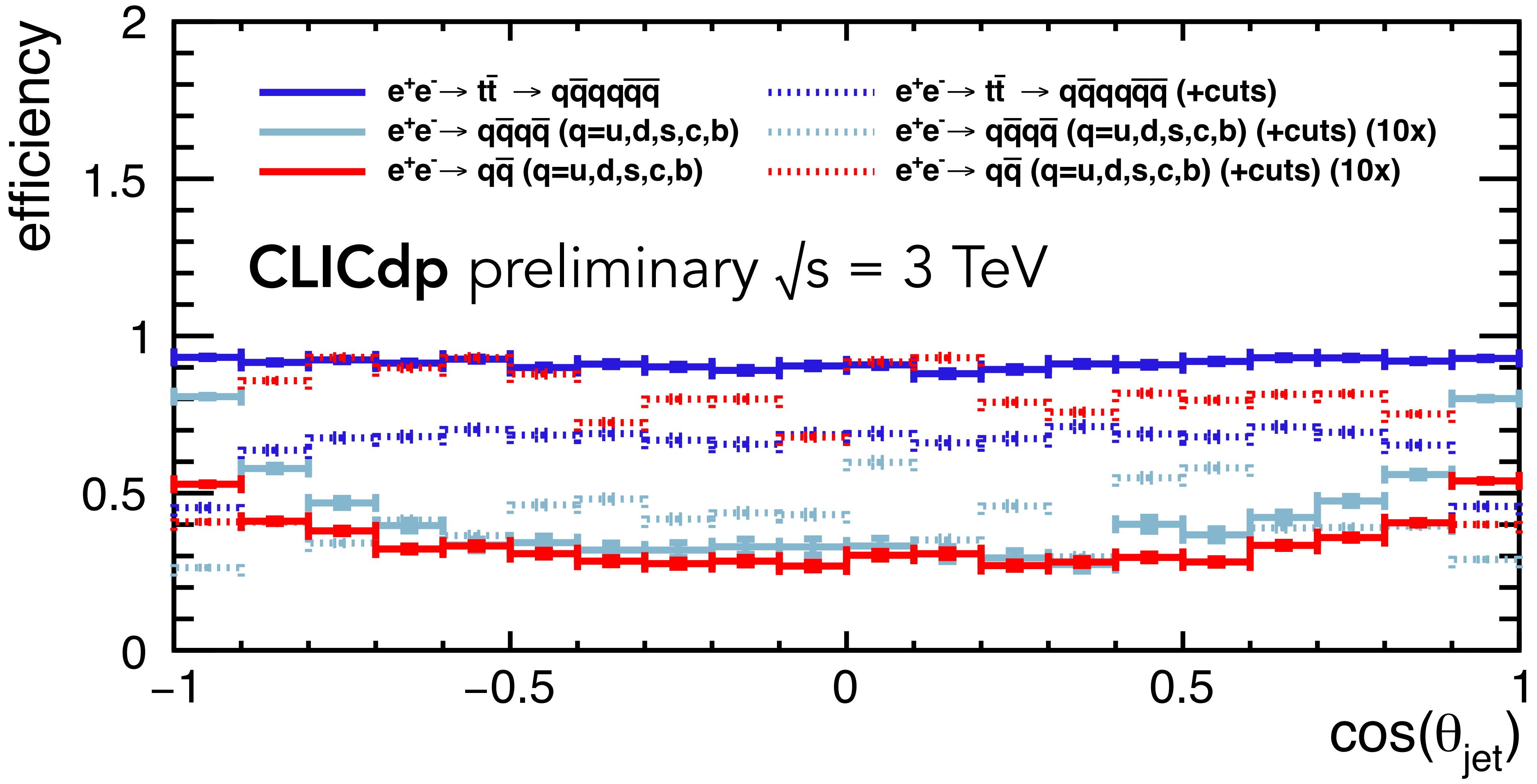


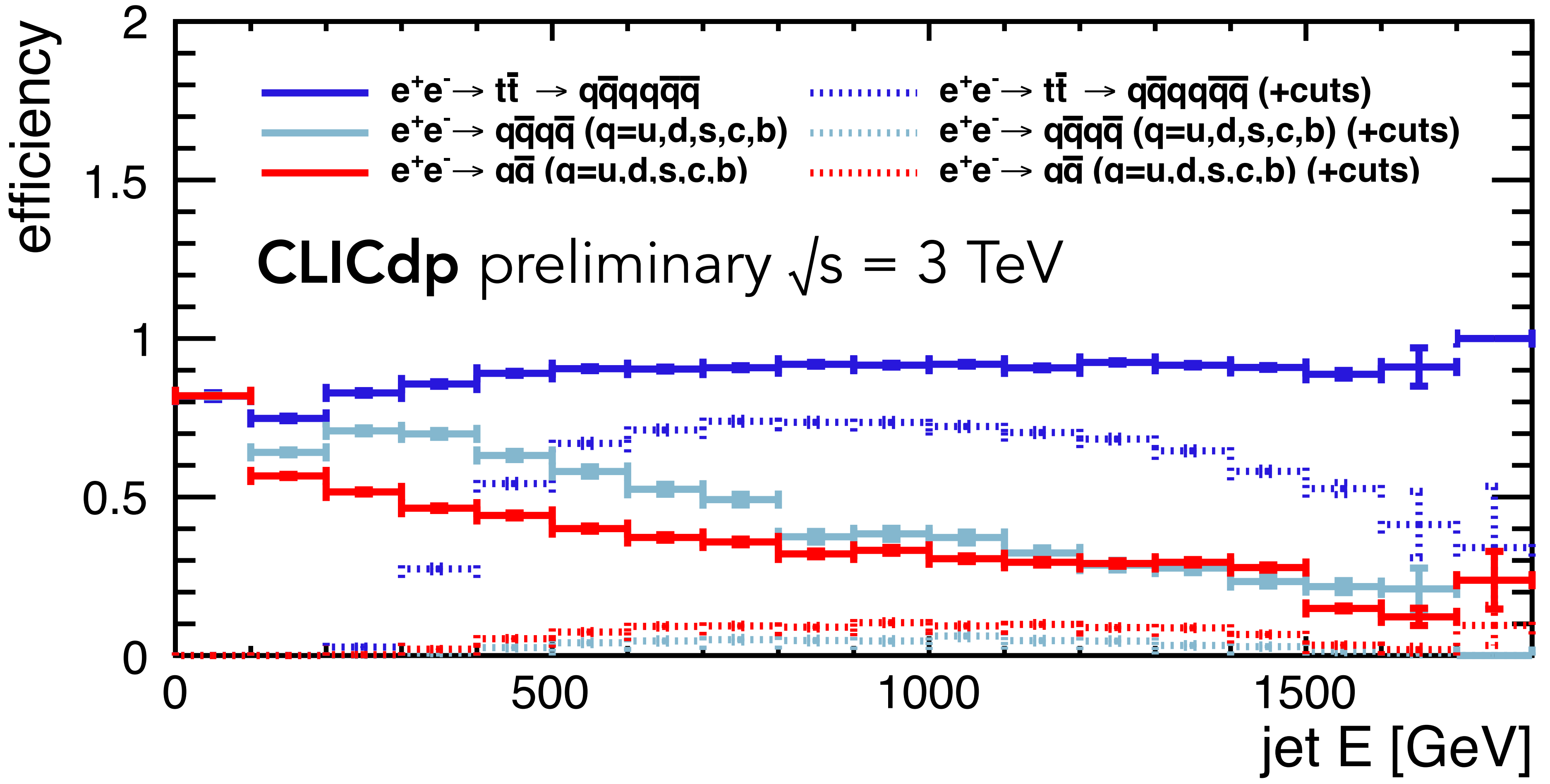
$e^+e^- \to qq$ (5 x)



Top tagging performance





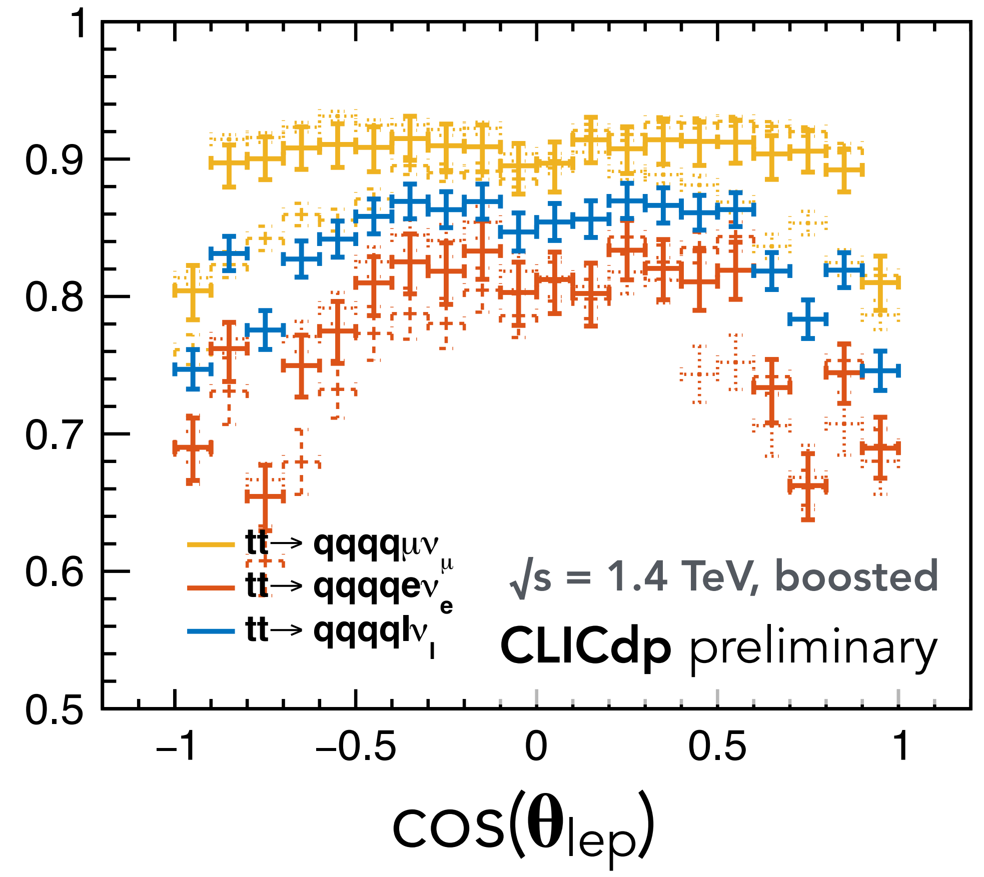


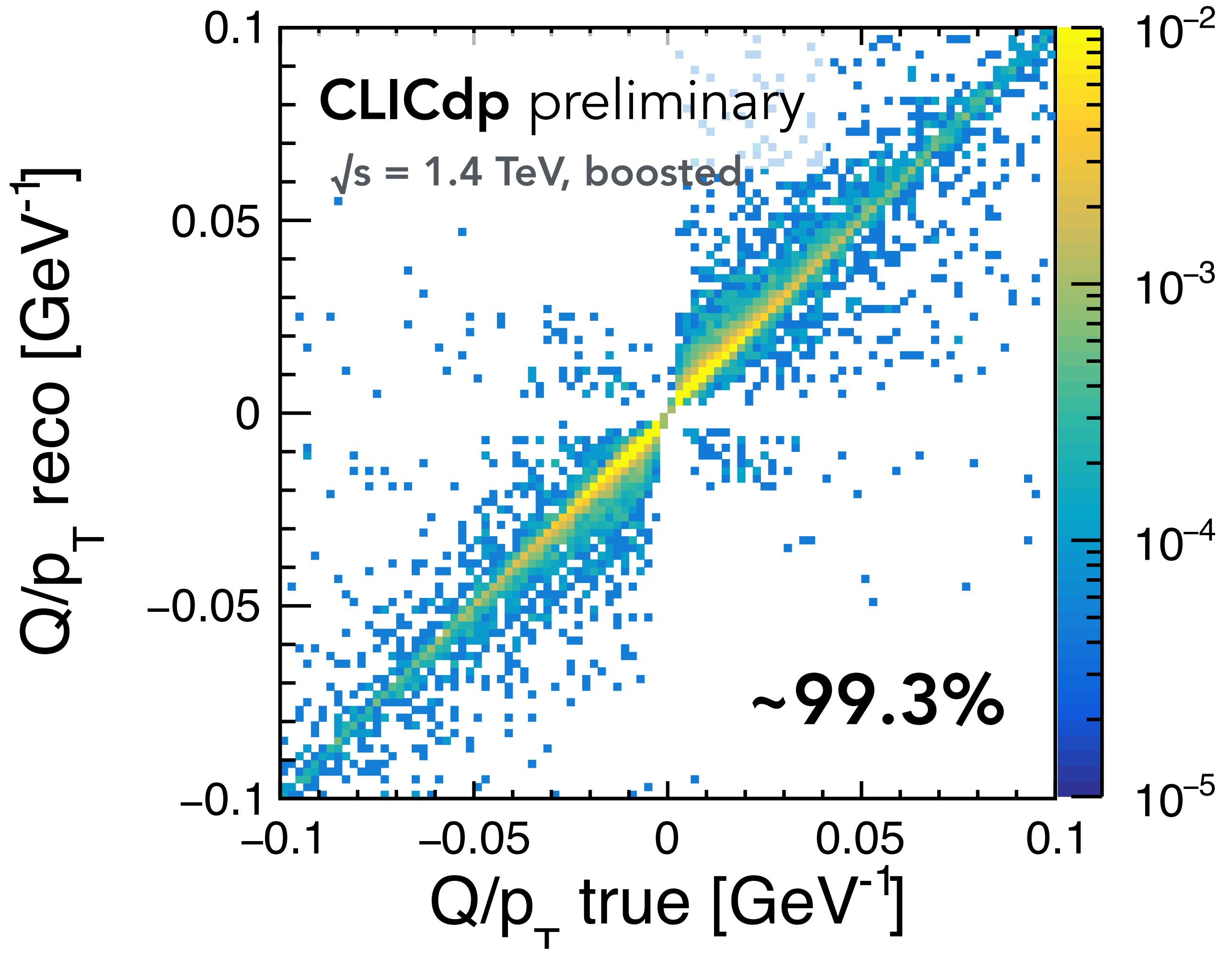
Top pair production at $\sqrt{s}$ = 1.4 (3) TeV

- **Event selection using top-quark identification strategy I – "Top tagging – parsing jet substructure"**
  - 1 isolated lepton (electron or muon)
  - 1 top-tagged large-R jet (hadronic top-quark candidate)
  - No isolated high-energy photons
  - Effective collision energy $\sqrt{s'}$ is reconstructed; cut applied at $\sqrt{s'} \geq 1.2$ (2.6) TeV (matching cut is applied at MC true level)
  - Multivariate analysis exploiting kinematics of both the hadronically and leptonically decaying top quark (including the detailed output from the top-tagger), event missing p_T, visible energy and event shape, lepton kinematics, flavour tagging, jet splitting scales, and substructure variables

Isolated lepton

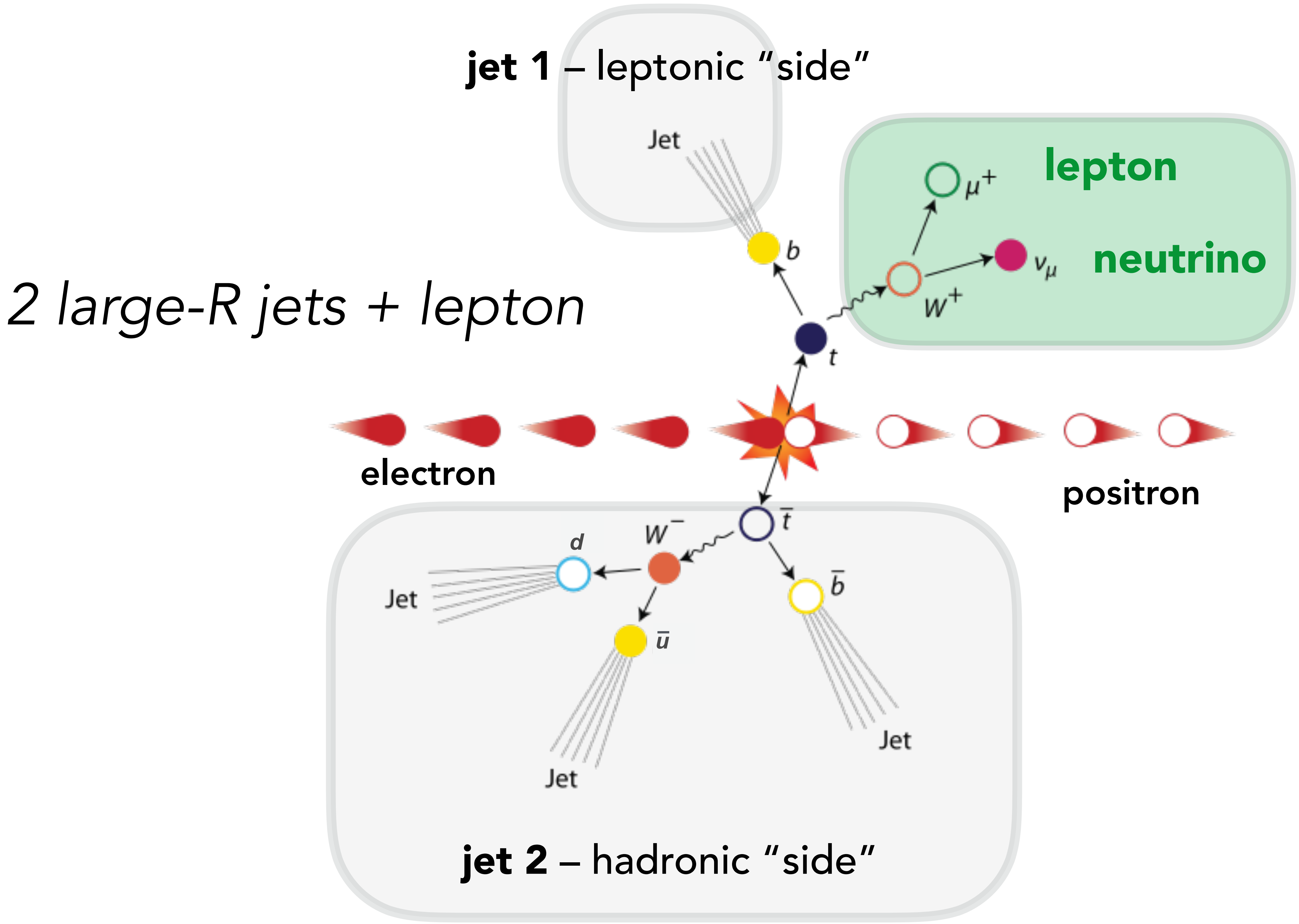
Isolated lepton efficiency vs. lepton polar angle





Effective collision energy √s’ reconstruction

- The effective collision energy √s’ is reconstructed to be able to study the electroweak production at/close-to the nominal collision energy
- Reconstruction:
  - In addition to the neutrino, missing energy/momentum also appear in the forward direction (ISR, beamstrahlung)
  - Assumption: neutrino $p_T = \sqrt{(p_x^2+p_y^2)}$ (MET)
  - The lepton+neutrino system constrained to m_W
  - The equation is quadratic in p_z, and has no solution if X is imaginary. In such cases the MET is scaled to provide a real solution (X=0)
  - We select the solution closest to m_{top} when combined with one of the large-R jets

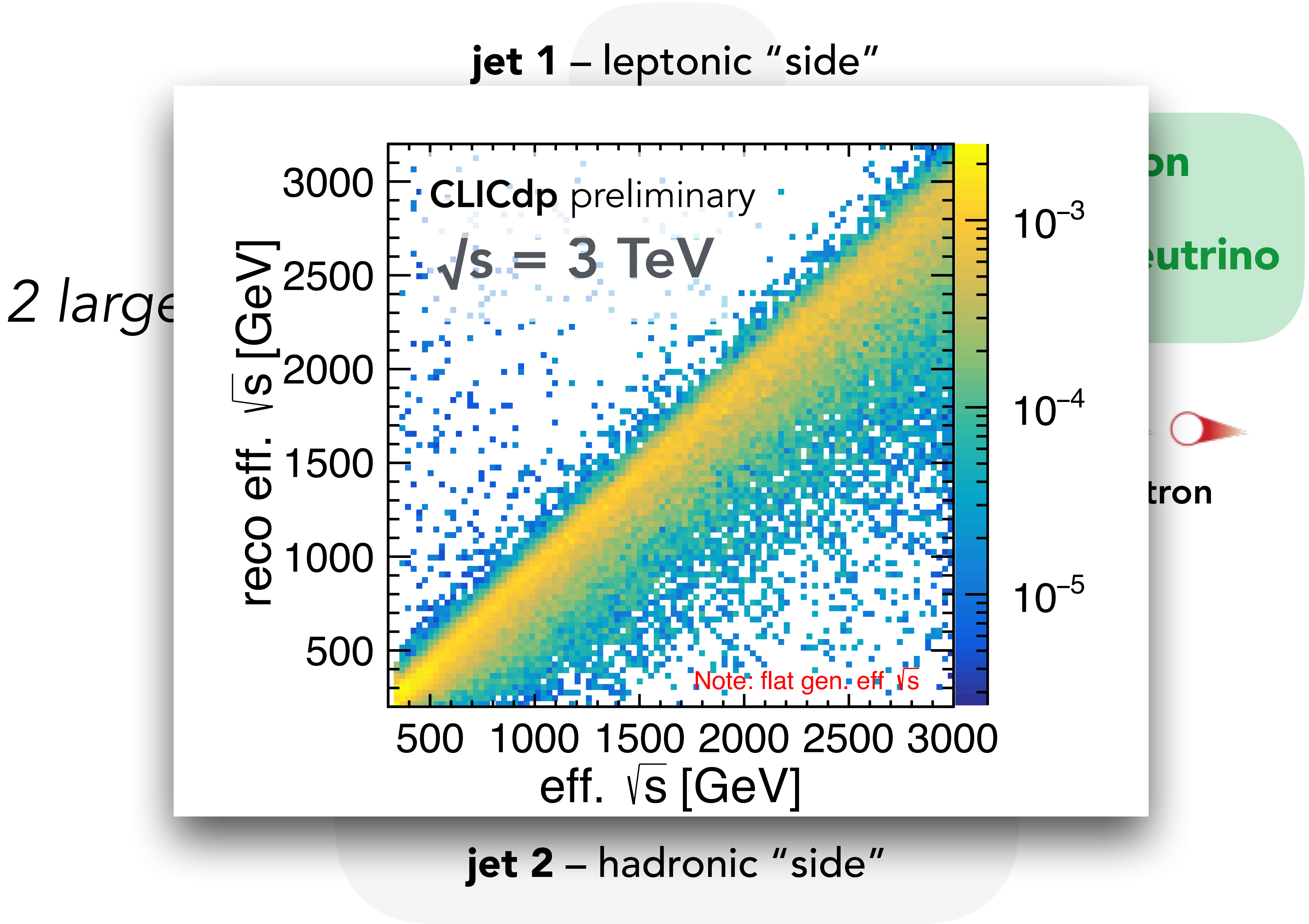


$$p_{\nu,z} = \frac{1}{2(p_{l,z}^2 - E_l^2)}\left(p_{l,z}m_l^2 - p_{l,z}M_W^2 - 2p_{l,x}p_{l,z}p_{\nu,x} - 2p_{l,y}p_{l,z}p_{\nu,y} + X\right)$$

$$X = \sqrt{E_l^2\left[\left(M_W^2 - m_l^2 + 2(p_{l,x}\cdot p_{\nu,x} + p_{l,y}\cdot p_{\nu,y})\right)^2 + 4\mathcal{E}_T^2(-E_l^2 + p_{l,z}^2)\right]}$$

Effective collision energy √s’ reconstruction

- The effective collision energy √s’ is reconstructed to be able to study the electroweak production at/close-to the nominal collision energy
- Reconstruction:
  - In addition to the neutrino, missing energy/momentum also appear in the forward direction (ISR, beamstrahlung)
  - Assumption: neutrino $p_T = \sqrt{(p_x^2+p_y^2)}$ (MET)
  - The lepton+neutrino system constrained to m_W
  - The equation is quadratic in p_z, and has no solution if X is imaginary. In such cases the MET is scaled to provide a real solution (X=0)
  - We select the solution closest to m_{top} when combined with one of the large-R jets



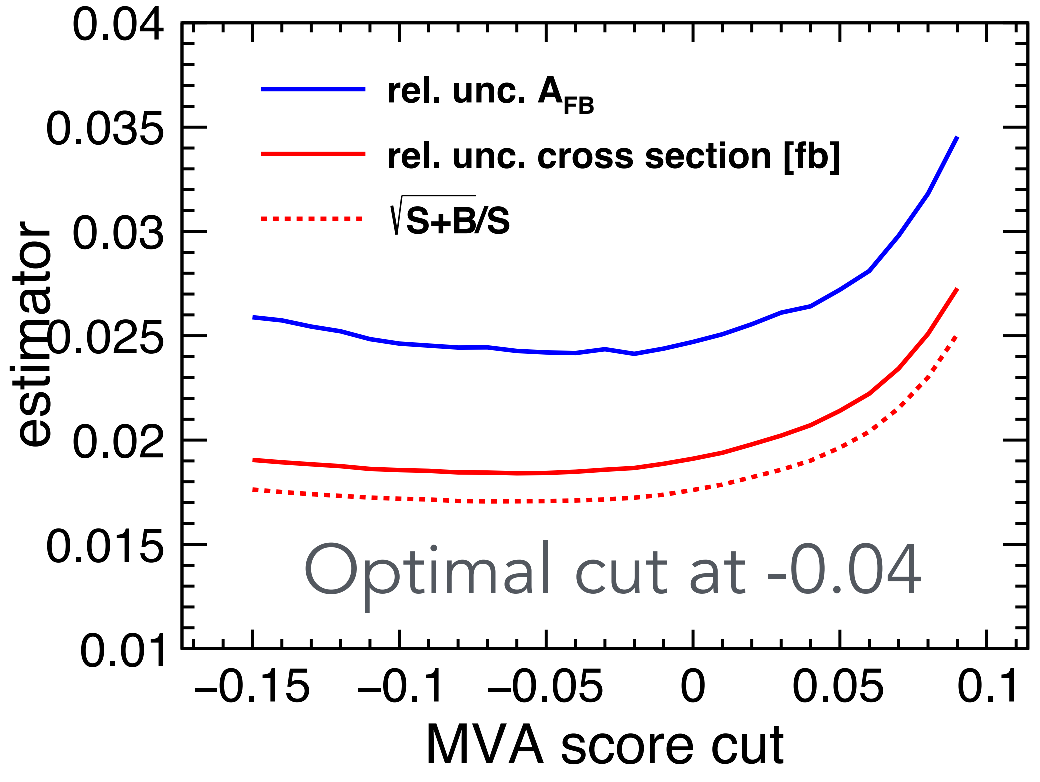
$$p_{\nu,z} = \frac{1}{2(p_{l,z}^2 - E_l^2)}\left(p_{l,z}m_l^2 - p_{l,z}M_W^2 - 2p_{l,x}p_{l,z}p_{\nu,x} - 2p_{l,y}p_{l,z}p_{\nu,y} + X\right)$$

$$X = \sqrt{E_l^2\left[\left(M_W^2 - m_l^2 + 2(p_{l,x}\cdot p_{\nu,x} + p_{l,y}\cdot p_{\nu,y})\right)^2 + 4\mathcal{E}_T^2(-E_l^2 + p_{l,z}^2)\right]}$$

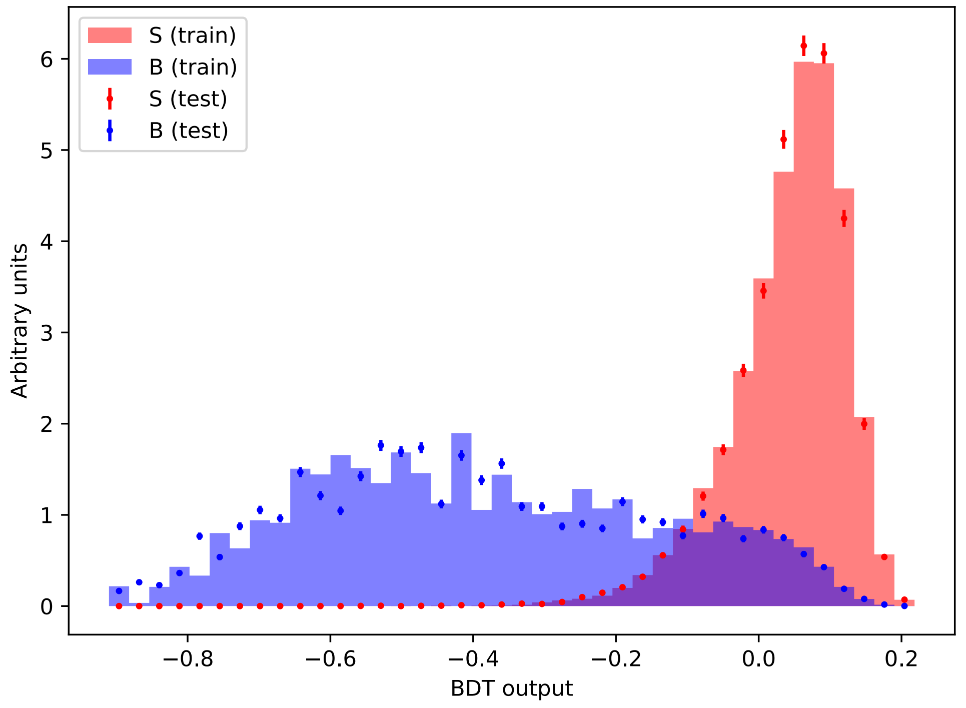
Multivariate analysis (MVA)

- Background rejection performance limited with single MVA, likely due to the large variety of backgrounds (some mimicking signal well and others not)
- Found that dedicated MVAs targeting certain backgrounds separately was useful
- Pre-MVA scores are fed into final MVA
- Using 2→1 approach
  - MVA1 focus on qq++ events (qq, qqvv, qqll, qqlv, qqlvlv)
  - MVA2 focus qqqq, qqqqqq
  - MVA3 focus on all backgrounds
- Each MVA uses the 20 most important variables and the parameters of the algorithm are tuned to reduce overtraining
- Examples of important variables: score from MVA1/MVA2, missing p_T, event shape variables, top mass/energy, NSubjettiness τ_3/τ_2, sum of b-tags, y_{34}, mass of leptonic top
- MVA cut optimisation done on final observables: cross section and A_{FB}

$\sqrt{s}$ = 1.4 TeV P(e^-) = -80%



$\sqrt{s}$ = 1.4 TeV P(e^-) = -80%



Event selection summary

Event selection summary for the analysis of events at nominal collision energy 380 GeV, 1.4 TeV and 3 TeV

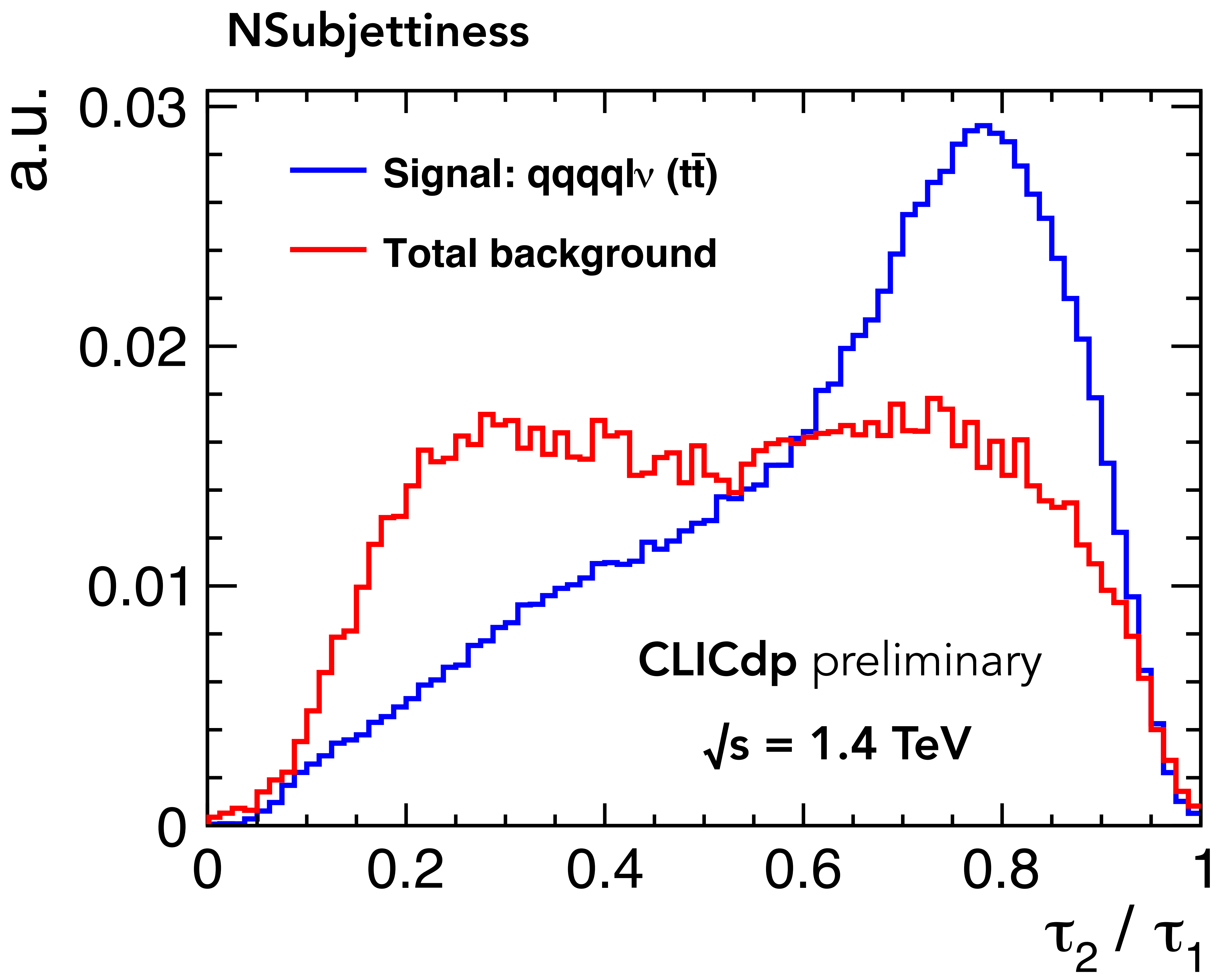
Dataset		Signal $e^+e^- \to t\bar{t} \to qqqql\nu$ ($l=e,\mu$)	
$\sqrt{s}$	$P(e^-)$	(Reconstructed events)	Signal purity
380 GeV	−80 %	25540	85 %
	+80%	12687	85 %
1.4 TeV	−80 %	5051	68 %
	+80%	2854	72 %
3 TeV	−80 %	1711	60 %
	+80%	1038	66 %

Analysis of radiative events at 1.4 TeV

- **Event selection using top-quark identification strategy II - "MVA-based – using substructure variables"**
  - 1 isolated lepton (electron or muon)
  - Associate large-R jet with highest energy with the hadronically decaying top
  - Pre-cuts to remove background
  - Quality-cuts to reduce the effect of migrations
  - Effective collision energy $\sqrt{s'}$ is reconstructed using a kinematic fit; cuts applied to define three regions-of-interest (matching cuts are applied at MC true level)
  - Multivariate analysis exploiting for example kinematics of both the hadronically and leptonically decaying top quark, visible p_T/E, lepton kinematics, flavour tagging, event shape, jet splitting scales, and substructure variables

Analysis of radiative events at 1.4 TeV

- **Substructure variables:**
  - Jet multiplicity (number of PFO within large-R jet)
  - NSubjettiness (metric for determining the number of subjets within fat-jet) (J.Thaler, K. Van Tilburg, arXiv:1011.2268)
  - Angular relations (relative angles between subjet pairs, identifies forced splitting)



Event selection summary

Event selection summary for the analysis of radiative events at nominal collision energy 1.4 TeV

Dataset		Signal $e^+e^- \to tt \to qqqql\nu$ (l=e,μ)	
√s = 1.4 TeV	$P(e^-)$	(Reconstructed events)	Signal purity
$0.40 \leq \sqrt{s'} \leq 0.90$ TeV	−80 %	449	44 %
	+80%	300	50 %
$0.90 \leq \sqrt{s'} \leq 1.2$ TeV	−80 %	2352	36 %
	+80%	1116	36 %
$1.2 \leq \sqrt{s'} \leq 1.4$ TeV	−80 %	4349	58 %
	+80%	2244	58 %

Results

$$\frac{d\sigma}{d(\cos(\theta^*))} = \sigma_1(1+\cos(\theta^*))^2 + \sigma_2(1-\cos(\theta^*))^2 + \sigma_3(1-\cos^2(\theta^*))$$

- Equation describes the differential cross section for the top quarks in the ttbar centre-of-mass system
- At tree level the three terms can be related to the cross sections for producing top-quark pairs with different helicity combinations in the final state

- Note that anti-tops are added with inverted sign
- The cross section and asymmetry are **extracted from fit**
- Detector efficiency included
- Statistical uncertainty from background included

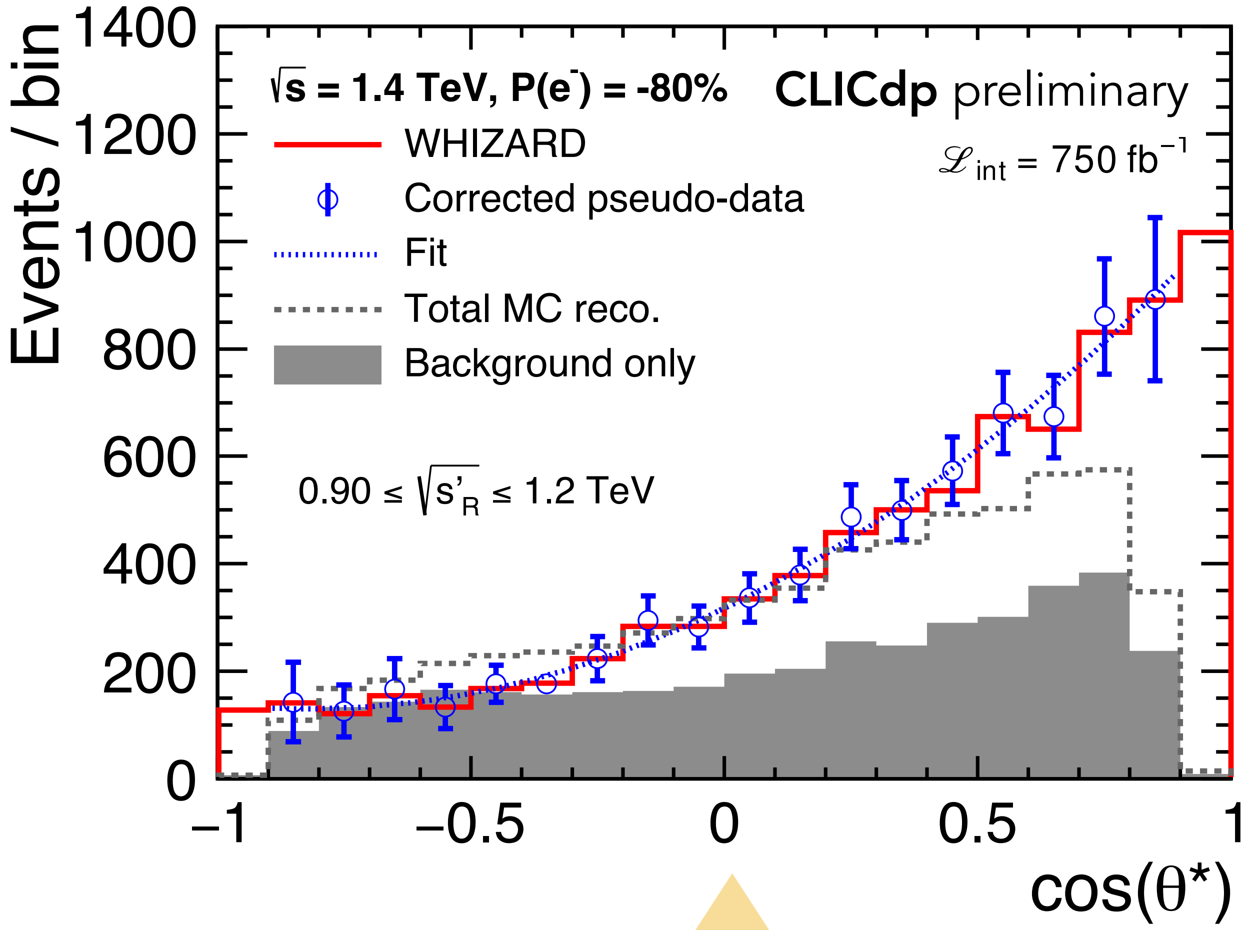
Observable definitions:

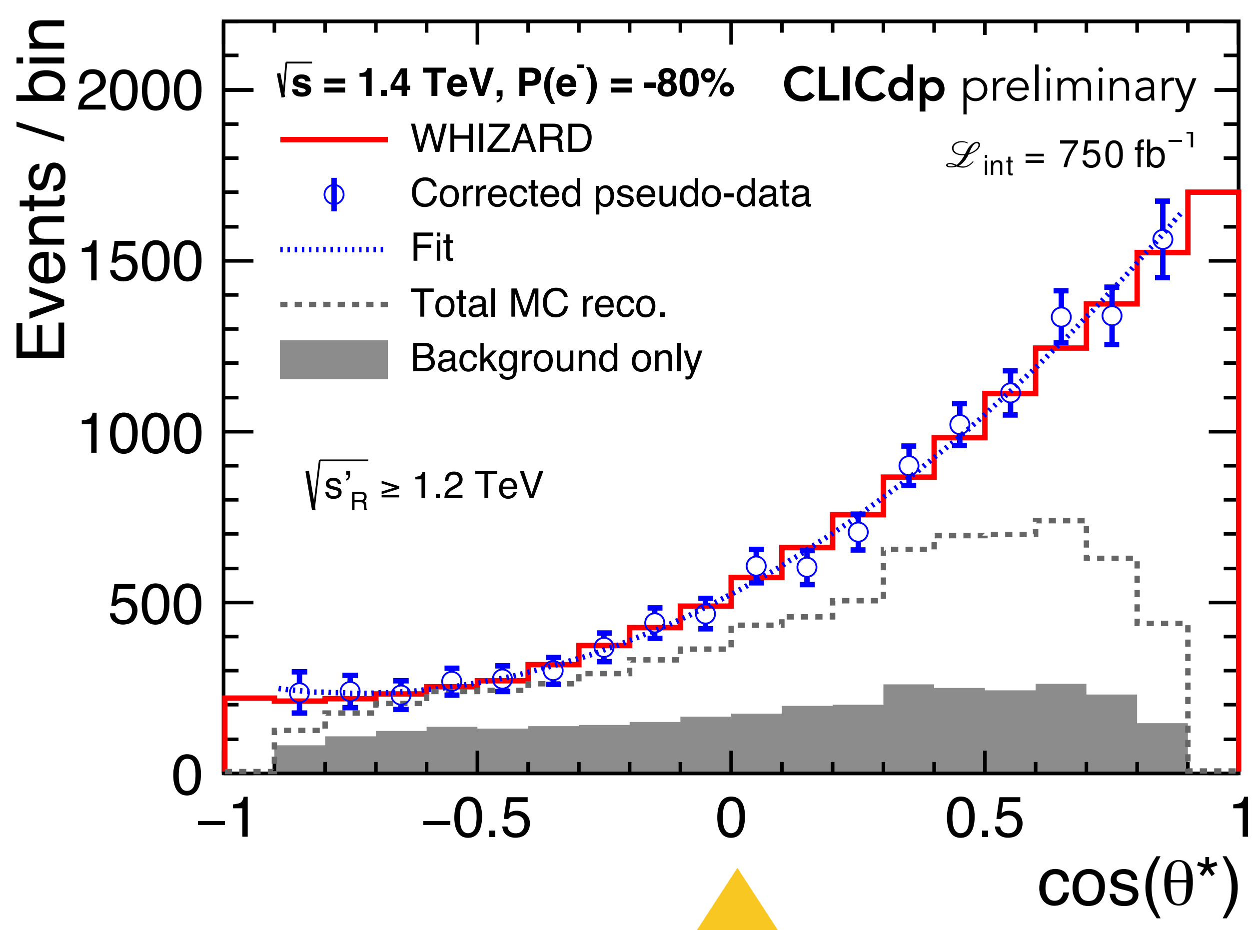
$$\sigma_{\mathrm{t\bar{t}}} = \sigma_{\mathrm{F}} + \sigma_{\mathrm{B}} = (4/3)(2\,\sigma_1 + 2\,\sigma_2 + \sigma_3)$$

$$A_{\mathrm{FB}} = \frac{\sigma_{\mathrm{F}} - \sigma_{\mathrm{B}}}{\sigma_{\mathrm{F}} + \sigma_{\mathrm{B}}} = \frac{1}{\sigma_{\mathrm{t\bar{t}}}}\,2\,(\sigma_1 - \sigma_2)$$

Results – radiative events at 1.4 TeV

- Results from the analysis of radiative events at 1.4 TeV using **strategy II**

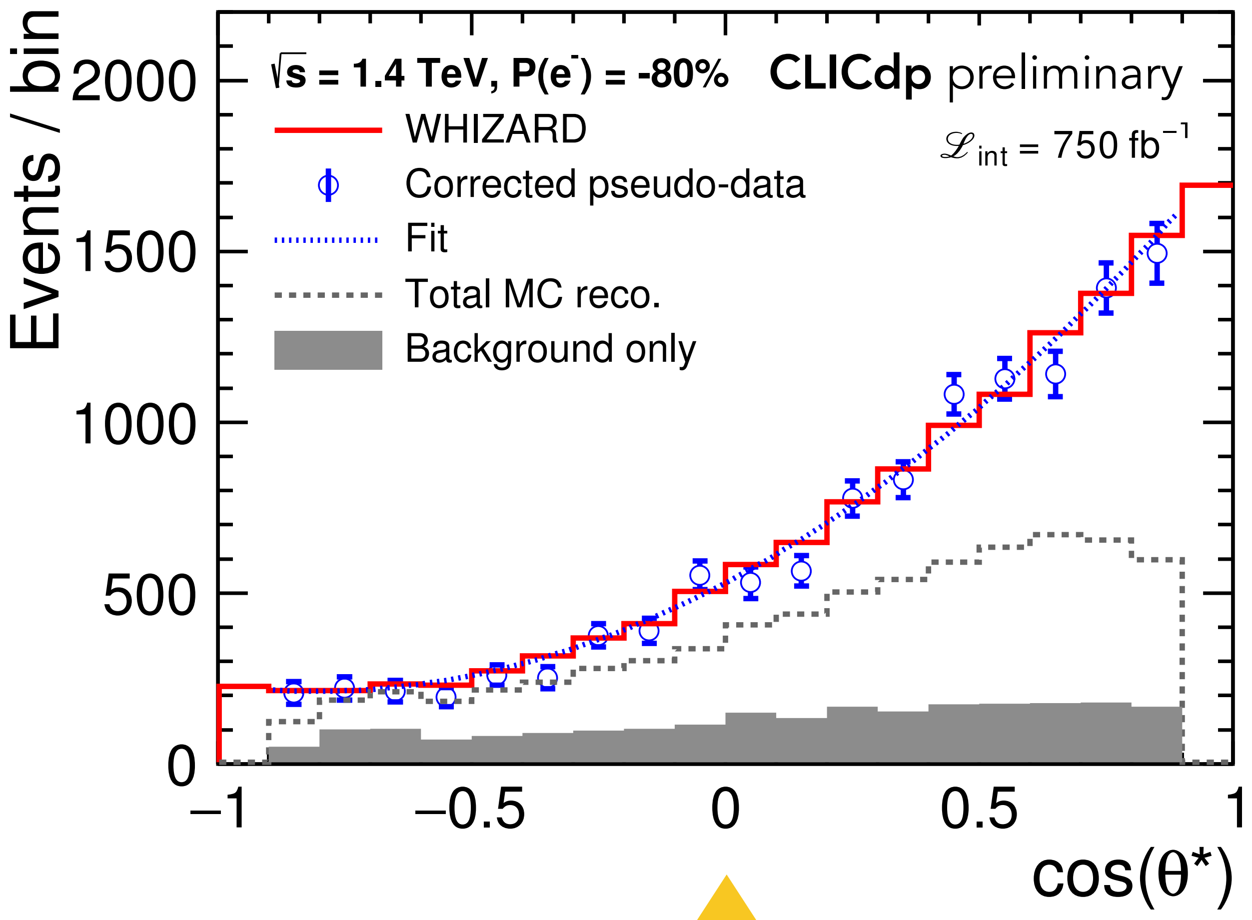


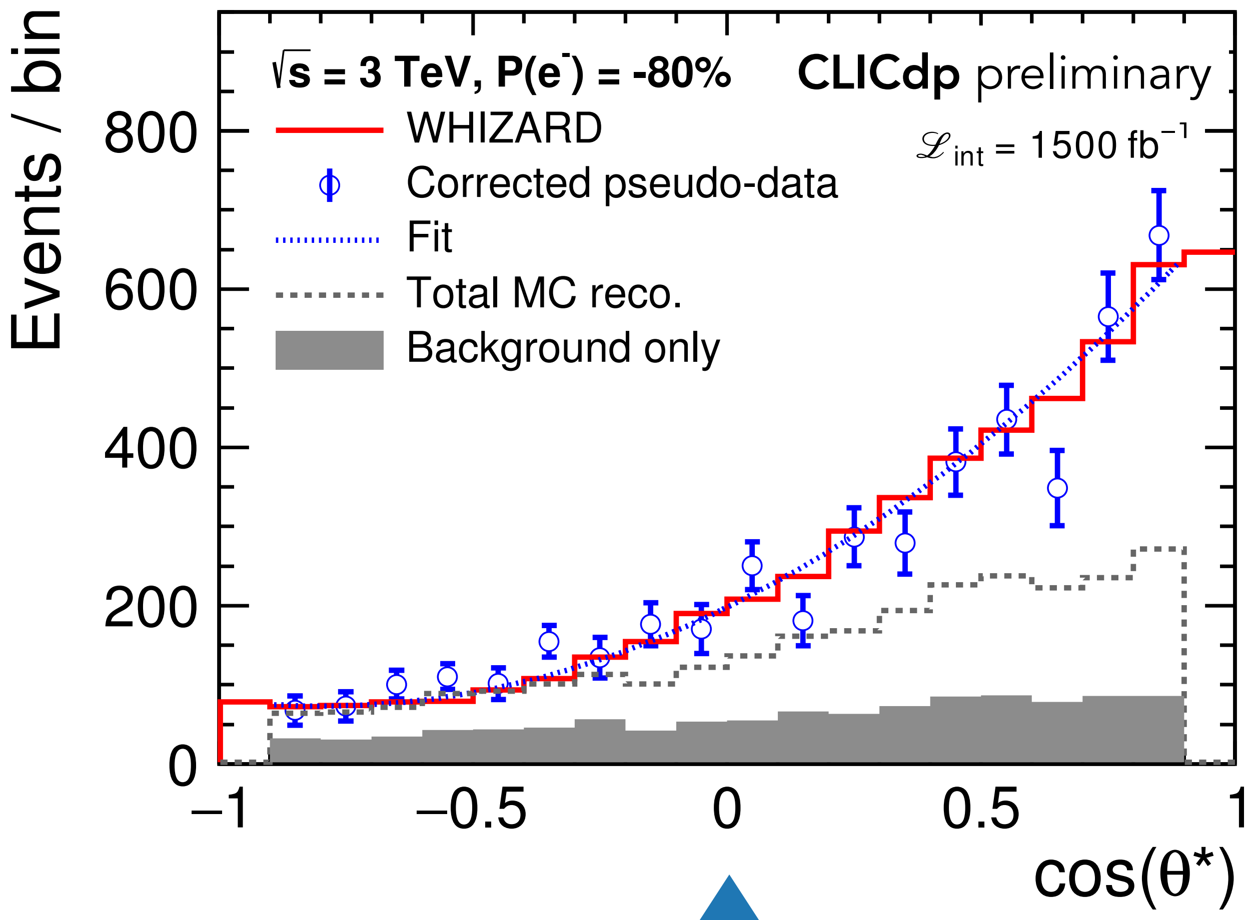


CLICdp preliminary	$0.40 \leq \sqrt{s'} \leq 0.90$ TeV		$0.90 \leq \sqrt{s'} \leq 1.2$ TeV		$1.2 \leq \sqrt{s'} \leq 1.4$ TeV	
P(e$^-$)	−80 %	+80%	−80 %	+80%	−80 %	+80%
A_{FB} Reco	0.458 ± 0.081 (18%)	0.514 ± 0.105 (20%)	0.546 ± 0.034 (6.2%)	0.588 ± 0.045 (7.7%)	0.562 ± 0.018 (3.2%)	0.621± 0.024 (3.7%)
σ Reco [fb]	16.56 ± 1.31 (7.9%)	8.63 ± 0.83 (9.6%)	11.01 ± 0.38 (3.5%)	5.87 ± 0.29 (4.9%)	18.41 ± 0.37 (2.0%)	9.84 ± 0.28 (2.8%)

Results at nominal collision energy

- Results from the analysis at nominal collision energy 1.4 and 3 TeV using **strategy I**

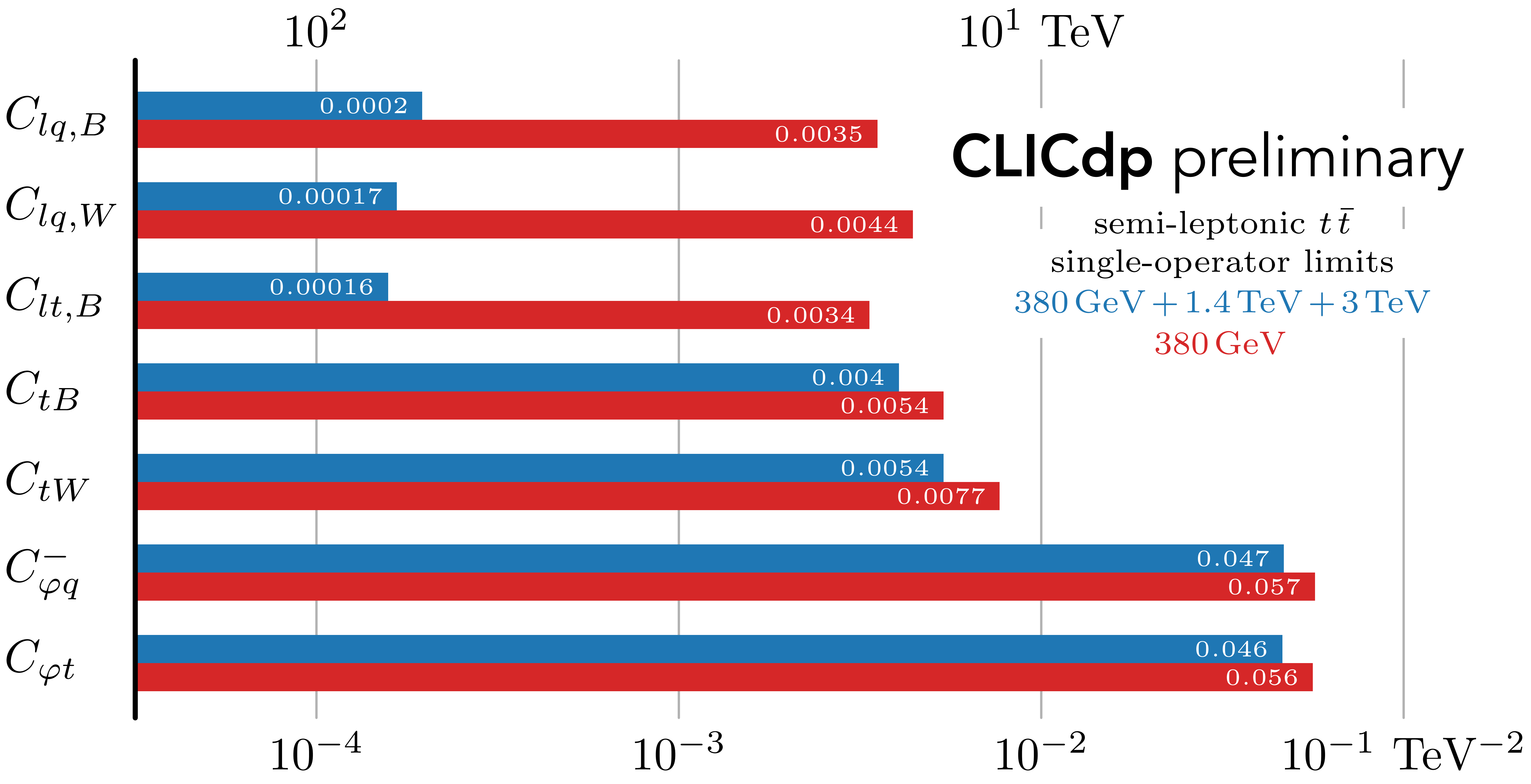




CLICdp preliminary	$\sqrt{s}$ = 380 GeV		$\sqrt{s}$ = 1.4 TeV		$\sqrt{s}$ = 3 TeV	
P(e⁻)	−80 %	+80%	−80 %	+80%	−80 %	+80%
A_{FB} Reco	0.1761 ± 0.0094 (5.3%)	0.207 ± 0.0084 (4.1%)	0.567 ± 0.014 (2.5%)	0.620 ± 0.016 (2.6%)	0.596 ± 0.022 (3.7%)	0.645 ± 0.028 (4.3%)
σ Reco [fb]	161.00 ± 1.09 (0.68%)	75.97 ± 0.73 (0.96%)	18.44 ± 0.34 (1.8%)	9.84 ± 0.23 (2.3%)	3.52 ± 0.12 (3.4%)	1.91 ± 0.08 (4.2%)

Results at nominal collision energy

- Adding the multi-TeV analyses to the EFT fit leads to large improvement
- BSM effect described through effective dim-6 operators:



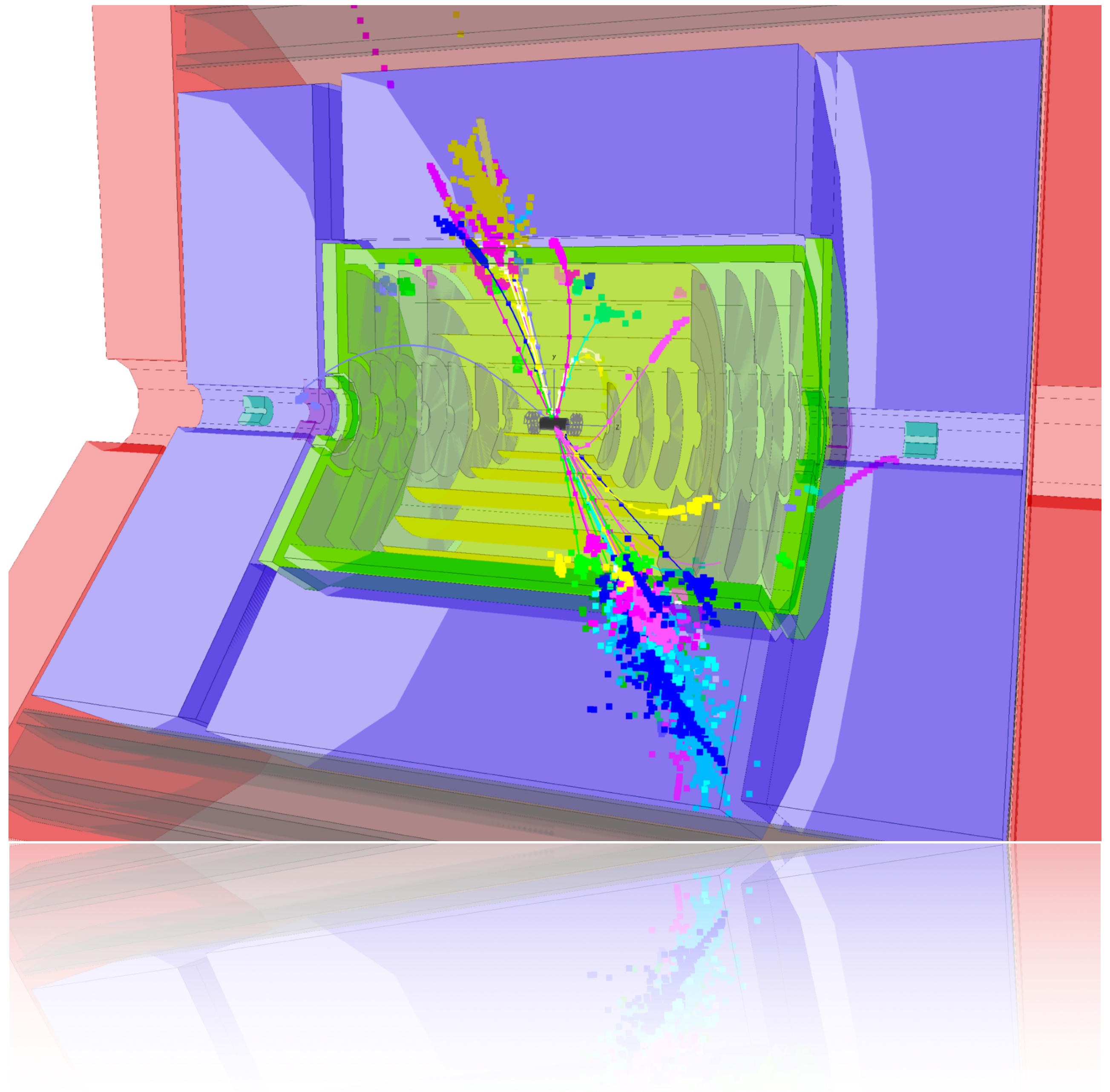
$$\mathscr{L}_{\text{EFT}} = \mathscr{L}_{\text{SM}} + \sum_i C_i Q_i$$

Wilson coefficients

See more details in talk "EFT fits of top physics at ILC and CLIC" by Martín Perelló Roselló (later today)

Summary

- **Top pair production** is studied at all stages of **CLIC**
- Results for production cross section and forward-backward asymmetry were presented in this talk and are interpreted in a top-philic EFT approach (talk by Martín Perelló Roselló later today)
- **Multi-TeV** analyses leads to **significant improvements** for 4-fermion "contact operators" while 2-fermion "vertex operators" are best measured at 380 GeV
- **Top physics overview paper in CLICdp collaboration review**
  - See also talk by Filip Zarnecki (tomorrow) on **FCNC top decays** at 380 GeV
  - Talk by Yixuan Zhang (next) on **ttH**
  - Top threshold scan, talk by Frank Simon tomorrow
  - Top mass determination above threshold, talk by Pablo Gomis Lopez tomorrow



Top identification and top pair production at CLIC

Workshop on top physics at the LC 2018,
June 4-6, 2018

On behalf of the CLICdp Collaboration

Philipp Roloff, Rickard Ström (CERN)
Martín Perelló Roselló, Ignacio Garcia Garcia, Marcel Vos (IFIC - U. Valencia/CSIC)
Nigel Watson, Alasdair Winter (University of Birmingham)

Backup slides

Top tagger algorithm

based on the Johns-Hopkins top tagger DOI: 10.1103/PhysRevLett.101.142001

1) PFO objects are clustered into jets of size R (large jet)
   - Iteratively merge 4-vector pairs with closest $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ until $\Delta R < R$

2) Iteratively decluster each resulting jet (reversing each step in the jet clustering) to search for subjets
   - Split into two parts, reject softest if …………………………………………………… $\frac{p_T^{\text{subjet}}}{p_T^{\text{jet}}} < \delta_p$
   - Declustering continues on the harder object until:

Both subjets are harder than $p_T^{\text{jet}} \cdot \delta_p$ ✓

Both subjets are too close $|\Delta\eta| + |\Delta\phi| < \delta_r$ ✗

Both subjets are softer than $p_T^{\text{jet}} \cdot \delta_p$ ✗

3) If an original jet declusters into two subjets - step 2 is repeated on those subjets
   - Results in 1 (original jet), 2, 3, or 4 (additional soft gluon emission) subjets

$\left\{ \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \end{array} \right.$
1 ✗
2 ✗
3 ✓
4 ✓

4) Kinematic cuts

✗ = irreducible