**CLIC Workshop 2015 (26-30 January 2015) CERN**

# Jet reconstruction at Linear Colliders

M.Boronat, J.Fuster, **I. García***, E.Ros, M.Vos

Valencia
Spain

IFIC INSTITUT DE FÍSICA CORPUSCULAR
VNIVERSITAT DE VALÈNCIA
CSIC

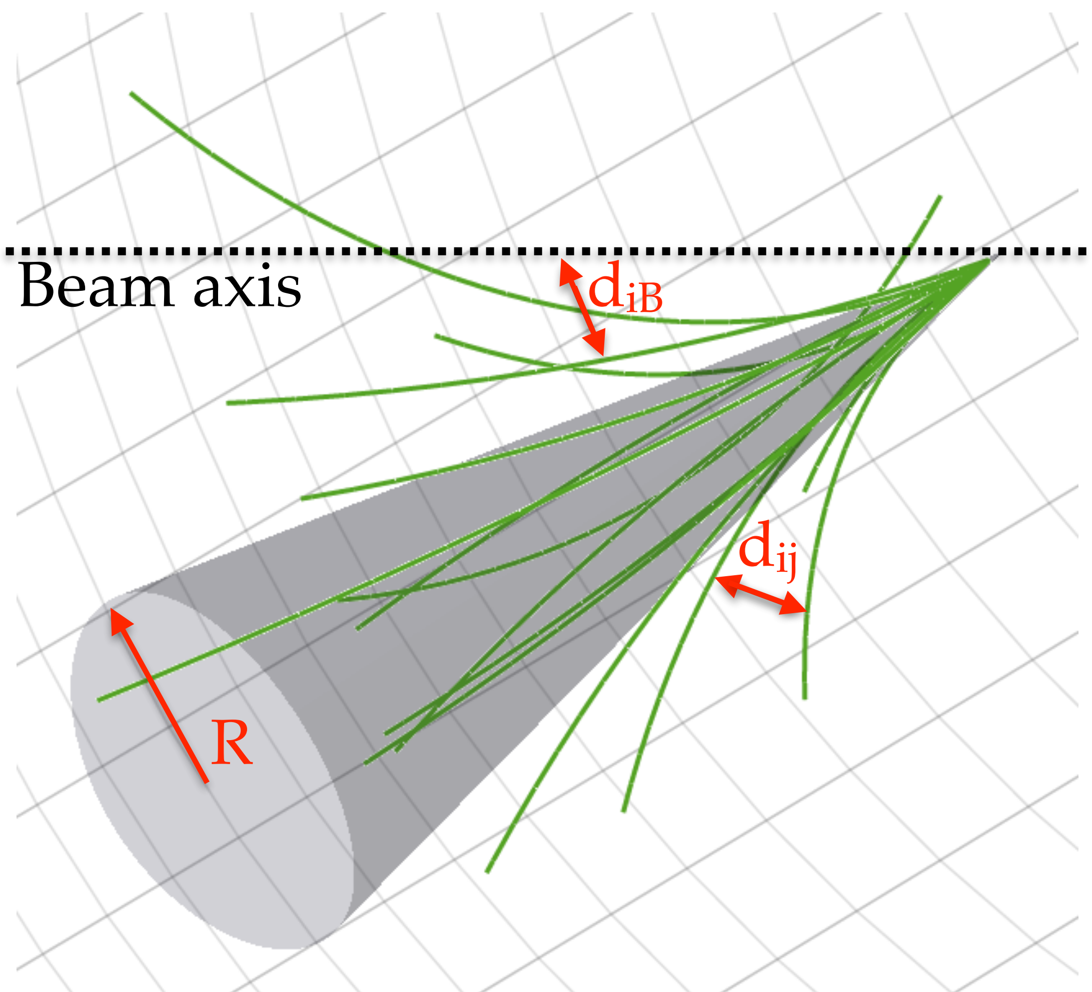
With thanks to Gavin Salam, André Sailer, Jesse Thaler

# Outline

1. **Jet algorithms**
    1. Sequential recombination jet algorithms
2. **Challenges for jet reconstruction at high energy $e^+e^-$ colliders**
    1. $\gamma\gamma \rightarrow$ hadrons
    2. jet multiplicity
    3. ISR
3. **A robust jet algorithm for $e^+e^-$ colliders (VLC algorithm)**
4. **Distance criteria and jet energy corrections**
5. **Monte Carlo simulation and realistic benchmark analysis**
    1. *tt* at ILC@500 GeV
    2. *ZZ* at CLIC@500 GeV
    3. Boosted tops at CLIC@3TeV
6. **Conclusions**

# Jet algorithms

A **jet reconstruction algorithm** is a mathematical tool used by experiments at colliders to **cluster the collimated sprays of particles** (quarks, gluons in the final state)




**Jet algorithm**

- $d_{ij}$: inter-particle distance
- $d_{iB}$: beam distance
- R: Radius parameter

# Sequential recombination algorithms

## Lepton colliders [E,θ]

**JADE 1980s**

$$y_{ij} = \frac{E_i^2, E_j^2}{Q^2}(1 - \cos\theta_{ij})$$

Experience on e⁺e⁻ data at Z-pole

**Durham or e⁺e⁻ $k_t$ algorithm**

(LEP and SLC)

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

Adapt to hadron colliders (cope with the boost along the beam axis)

## Hadron colliders [$p_T$,ΔR]

$$d_{ij} = min(p_{Ti}^{2n}, p_{Tj}^{2n})\Delta R_{ij}^{2n}/R^{2n}$$

$$d_{iB} = p_{Ti}^{2n}$$

**n=0**: Cambridge-Aachen

**n=1**: Longitudinally invariant $k_t$

**n=-1**: Anti-$k_t$ (LHC default)

Include beam distance in e⁺e⁻ algorithms

**Generalised e⁺e⁻ $k_t$ algorithm**

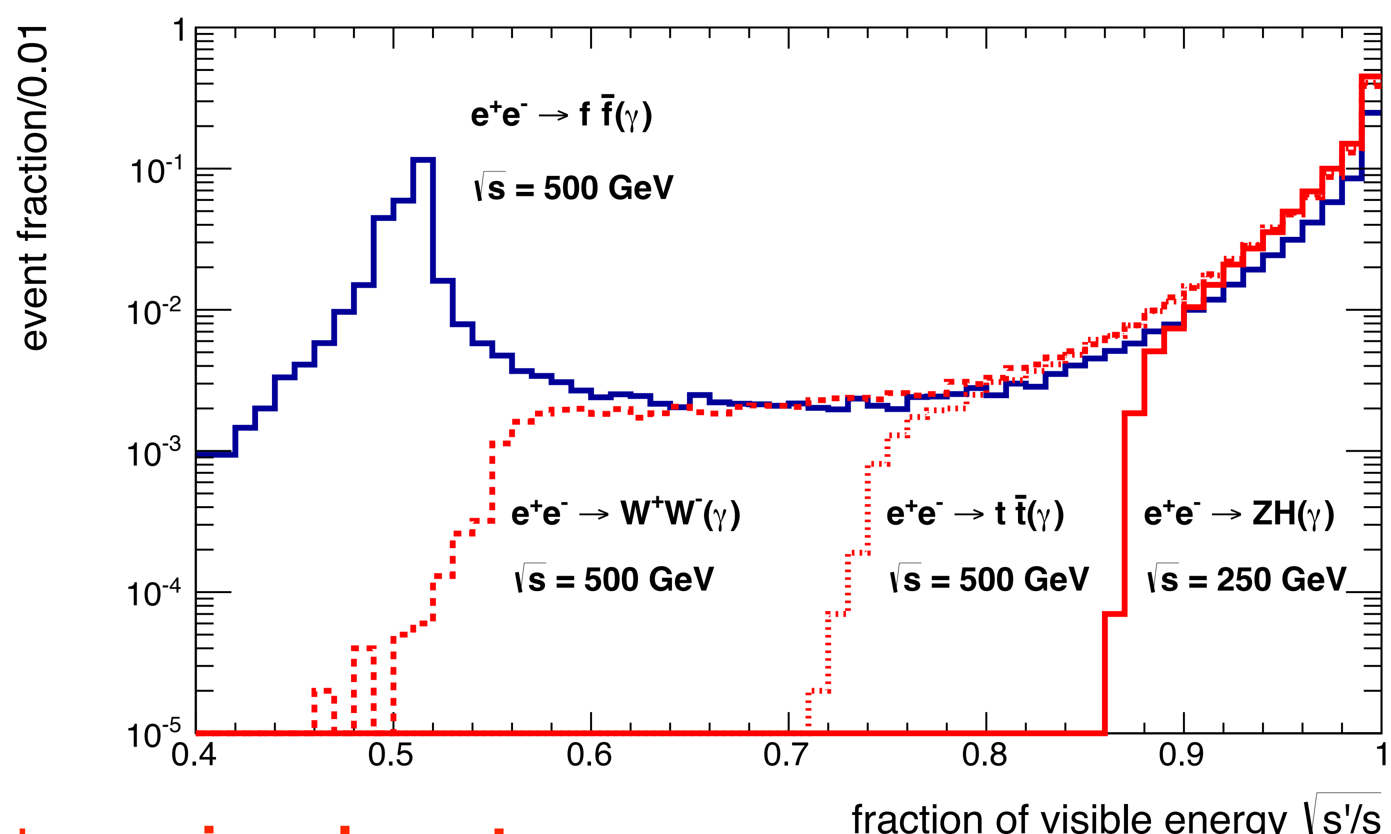
$$d_{ij} = min(E_i^2, E_j^2)(1 - \cos\theta_{ij})/(1 - \cos R)$$

$$d_{iB} = E_i^2$$

**Time to rethink e⁺e⁻ algorithms!!**

# Initial State Radiation (ISR)

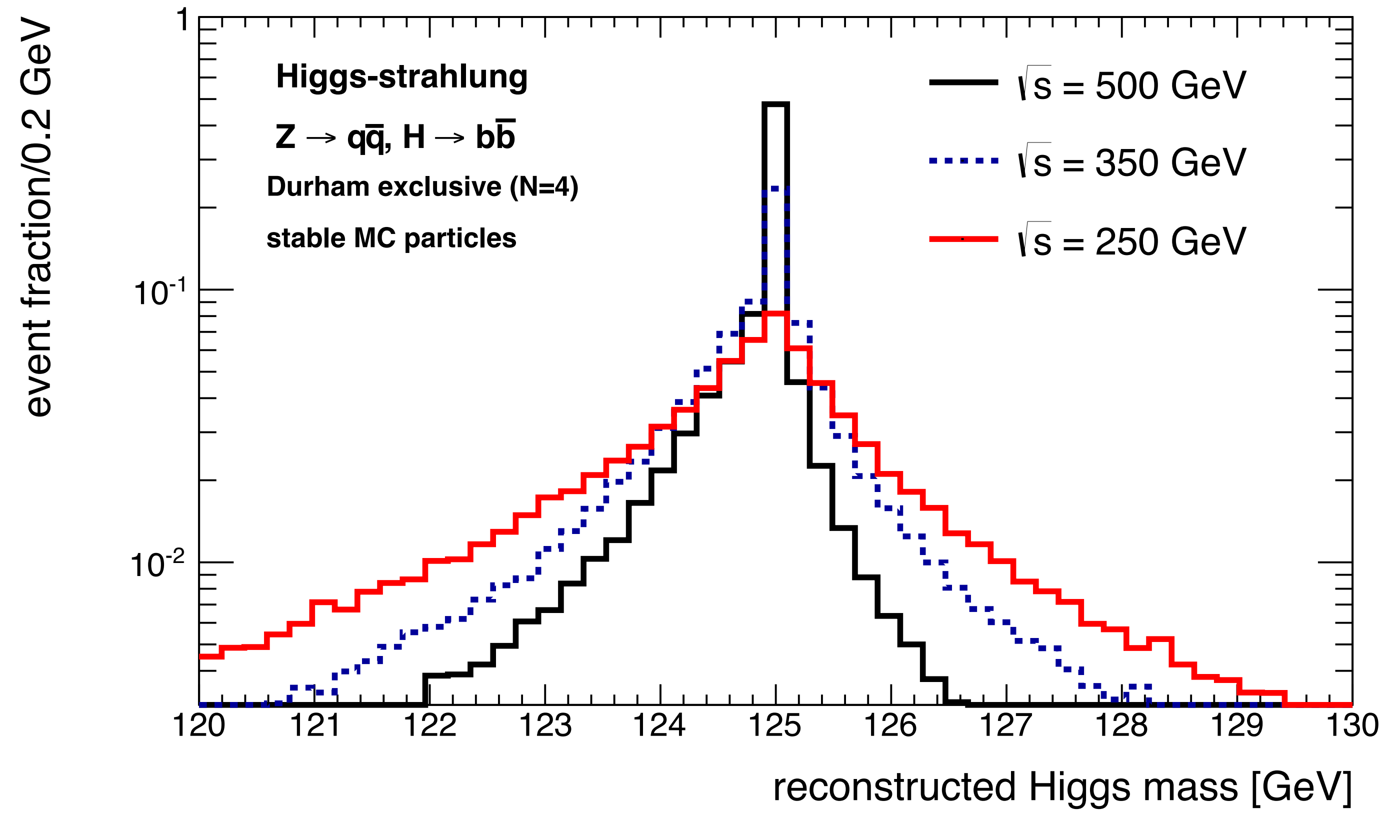
- Photons emitted by the incoming beam particles (**ISR**) can carry away a significant fractions of the nominal center-of-mass energy (*lower-energy lepton colliders, Z-pole*)
- However for **most interesting processes** at a future lepton collider ISR plays a much **less important role** ($\beta z = v_z/c$ is smaller than 0.1)
- Compared to hadron colliders such boost is very small
  - LHC di-jets: $\beta_z \sim 1$
  - LHC tt: $\beta_z \sim 0.5$




- At **lepton colliders ISR** leads to **a minor boost**

# Jet multiplicity

- For the **first time**, lepton colliders offer the possibility to study **multi-jet final states**
- $e^+e^- \rightarrow ZH$ ($2\rightarrow4$) (Higgs-strahlung)
- $e^+e^- \rightarrow tt$ // $e^+e^- \rightarrow ZHH$ ($2\rightarrow6$)
- $e^+e^- \rightarrow ttH$ ($2\rightarrow8$)




**Final state particles from the MadGraph + Pythia** Monte Carlo generator are clustered with the Durham algorithm (exclusive clustering with N = 4)

Even **without detector or backgrounds, the width is non-negligible**

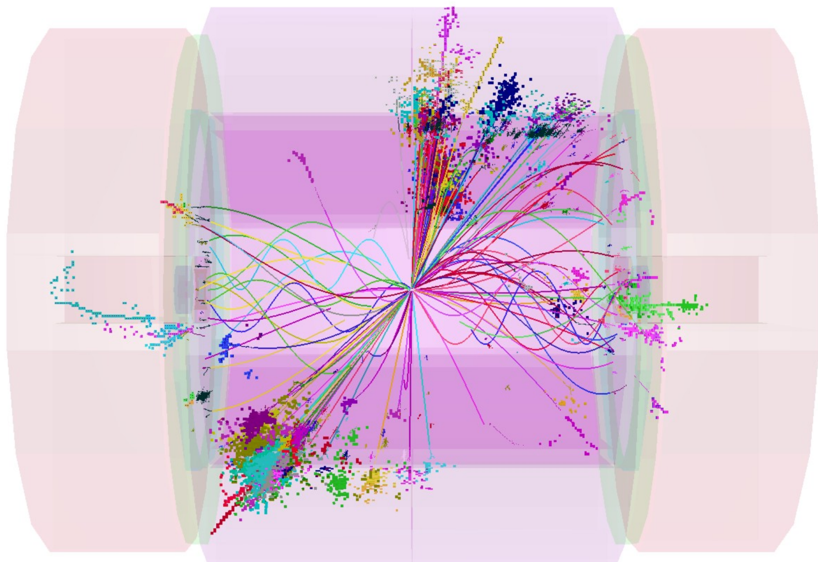
The effect is **reduced with center-of-mass energy → Greater boost, cleaner separation**

# $\gamma\gamma \to$ hadrons

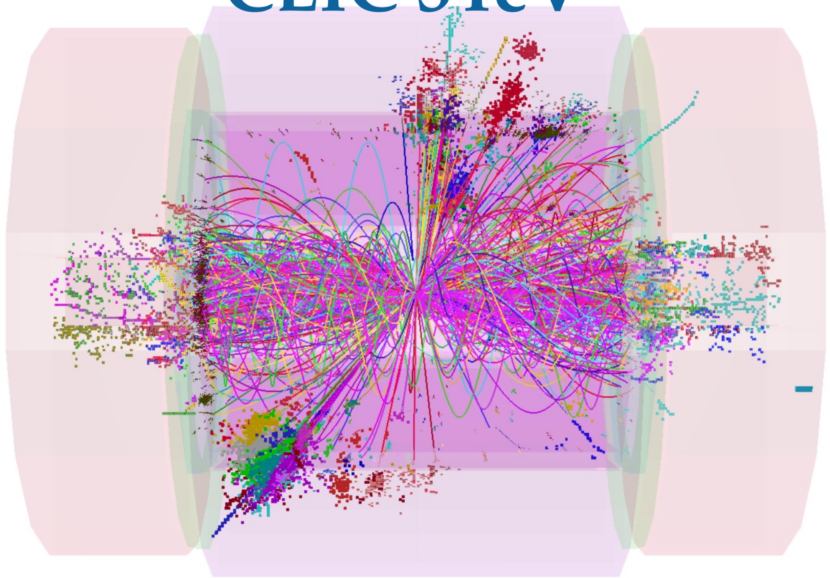
**$\gamma\gamma \to$ hadrons:**

1. Strongly peaked in the **forward region (90%)**
2. Background **scales with instantaneous luminosity** → Much larger at 3TeV than at 500GeV
3. Its impact depends on the **bunch structure** and **detector read-out speed**
   - → *ILC, 1300 bunches spaced by 500 ns*
   - → *CLIC, 312 bunches spaced by 0.5 ns*

**CLIC 500GeV**



**CLIC 3TeV**



The $\gamma\gamma \to$ **hadrons** background at **CLIC** has **strong** impact on jet reconstruction performance [CLIC CDR, Marshall & Thomson, arXiv:1308.4537]

Less pronounced, but **non-negligible** impact on **ILC** physics [many studies, arXiv:1307.8102]

**Use CLIC case to take jet reconstruction to the limit; if it works there, it's good for ILC too**

Several proposals to mitigate its effect is the use of hadron algorithms like long. inv. $k_t$ algorithm, but **is it the best we can do??**

# The VLC jet algorithm

A new clustering jet reconstruction algorithm that combines the good features of lepton collider algorithms, in particular the **Durham-like distance criterion;**

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

with the **robustness against background of** the longitudinally invariant **$k_t$ algorithm**

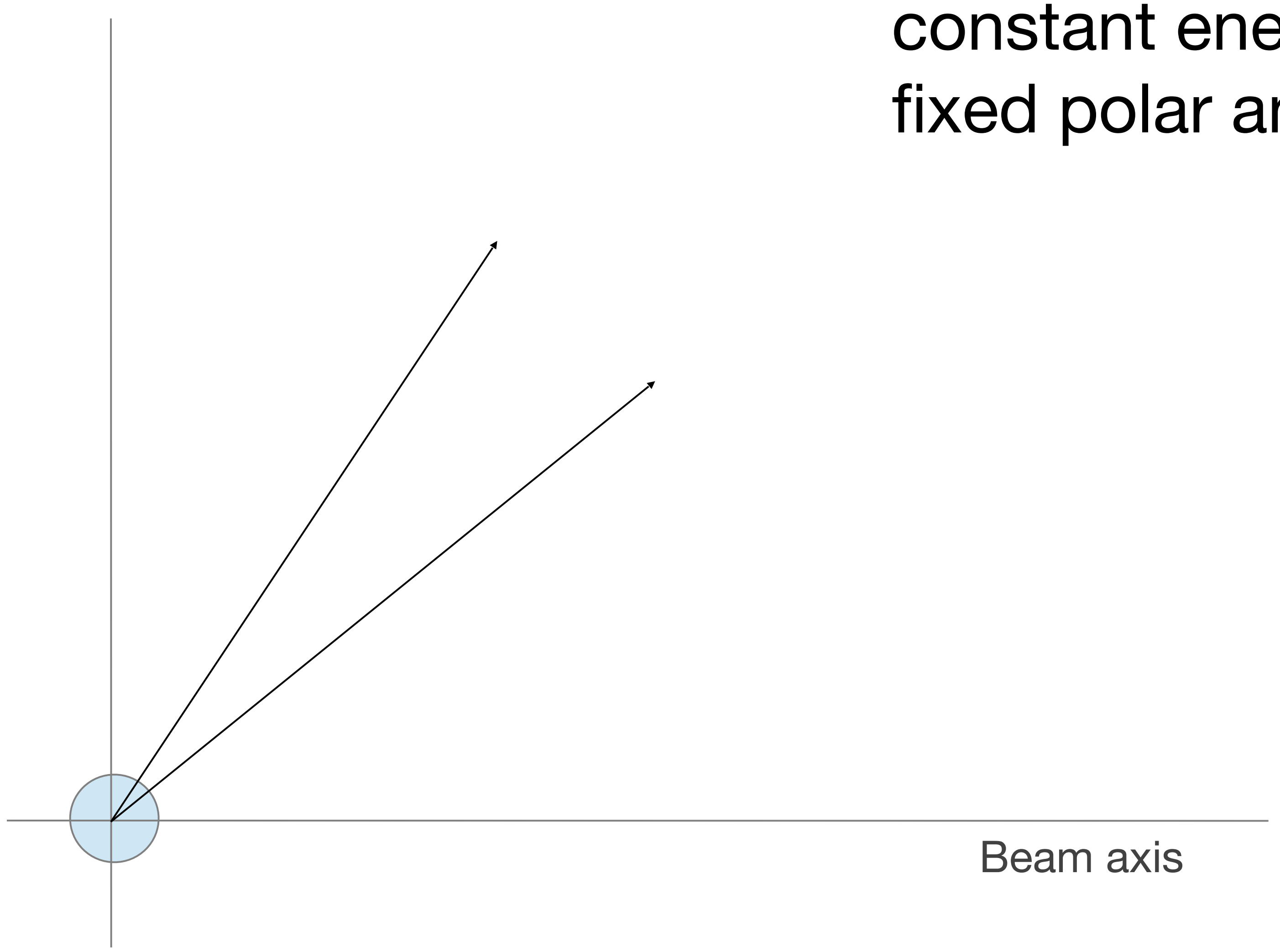
$$d_{iB} = E^{2\beta} \sin^{2\gamma}\theta_{iB}$$

The $\gamma$ parameter governs the evolution of the jet area with polar angle and $\beta$ allows to ***change the clustering order.***

*In the default settings the two exponents β and γ are equal. For **β=γ=1** the expression simplifies to $\boldsymbol{d_{iB} = E^2 \sin^2 \theta_{iB} = p^2_{ti}}$

# Comparison of the distance criteria

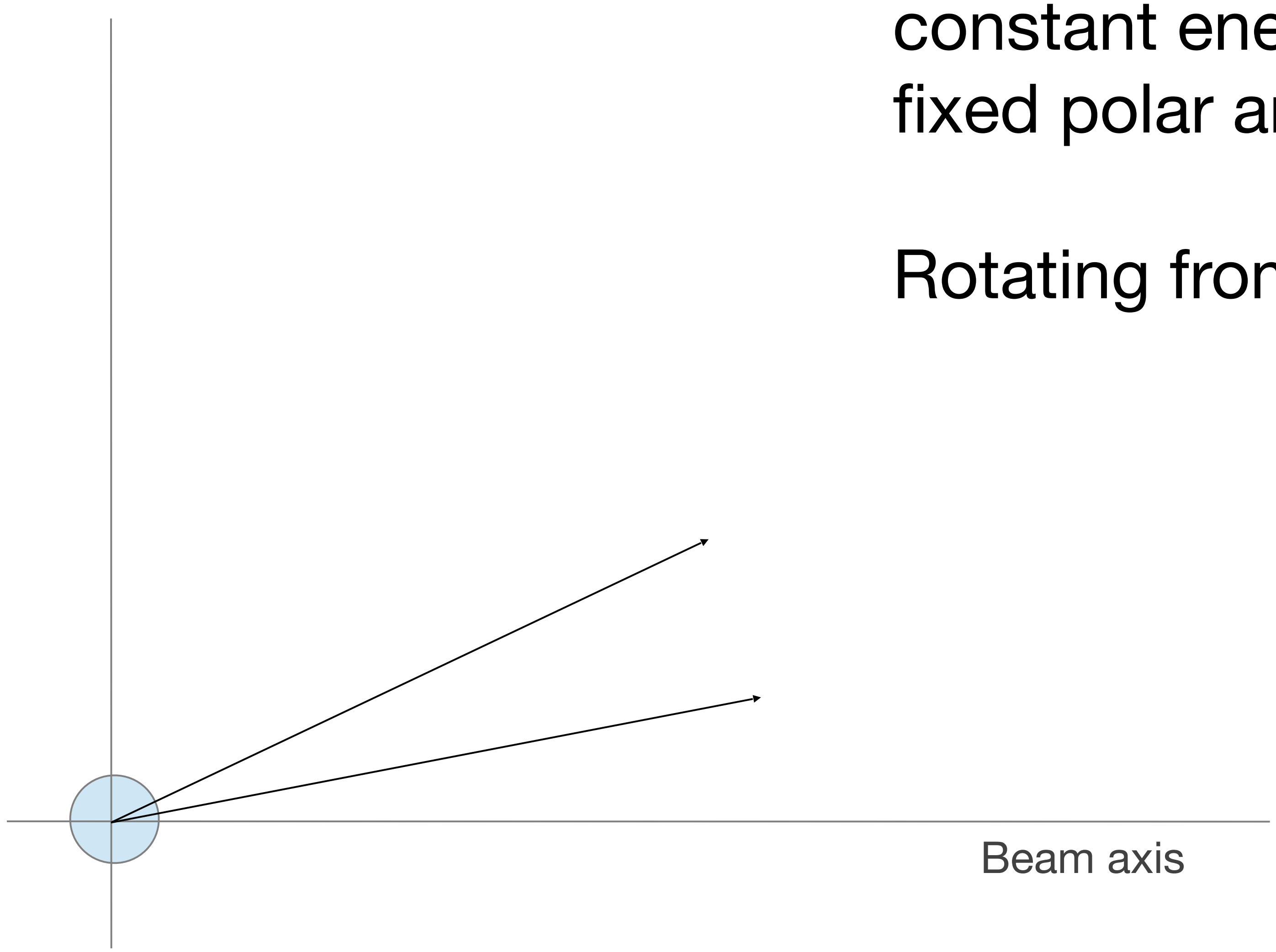
**Two test particles** with
constant energy (E = 1 GeV) and
fixed polar angle separation (100 mrad)



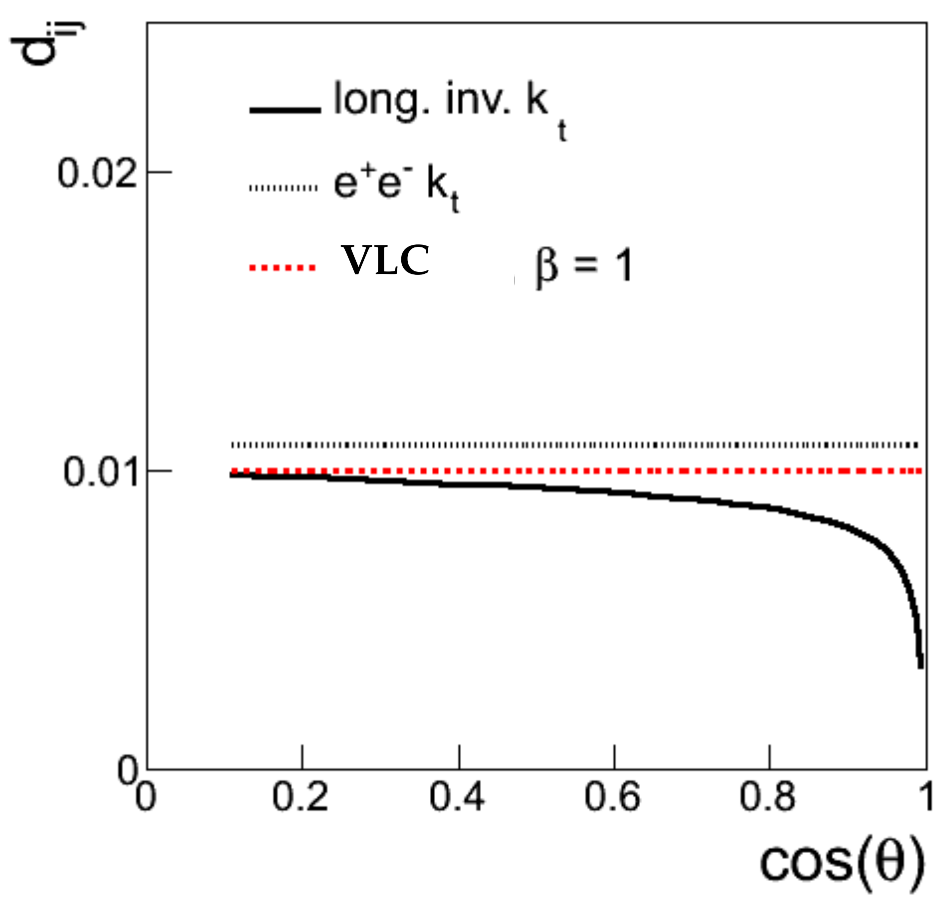
# Comparison of the distance criteria

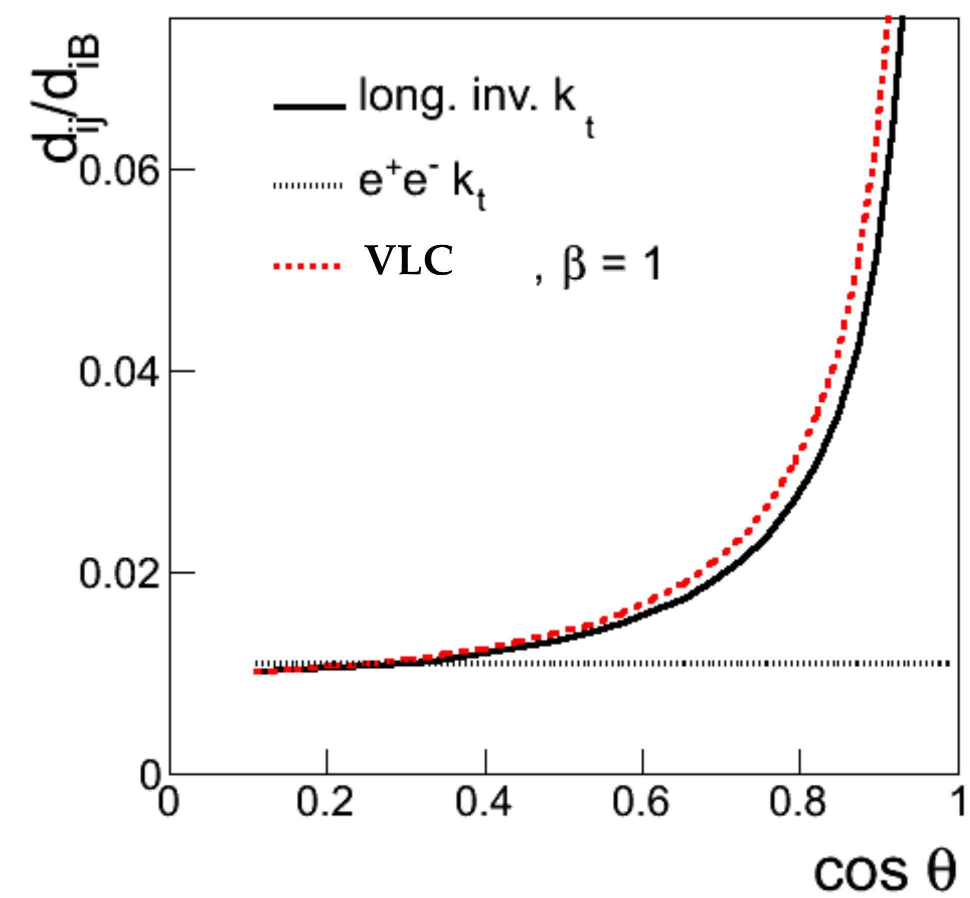
**Two test particles** with
constant energy (E = 1 GeV) and
fixed polar angle separation (100 mrad)

Rotating from central to forward region



# Comparison of the distance criteria



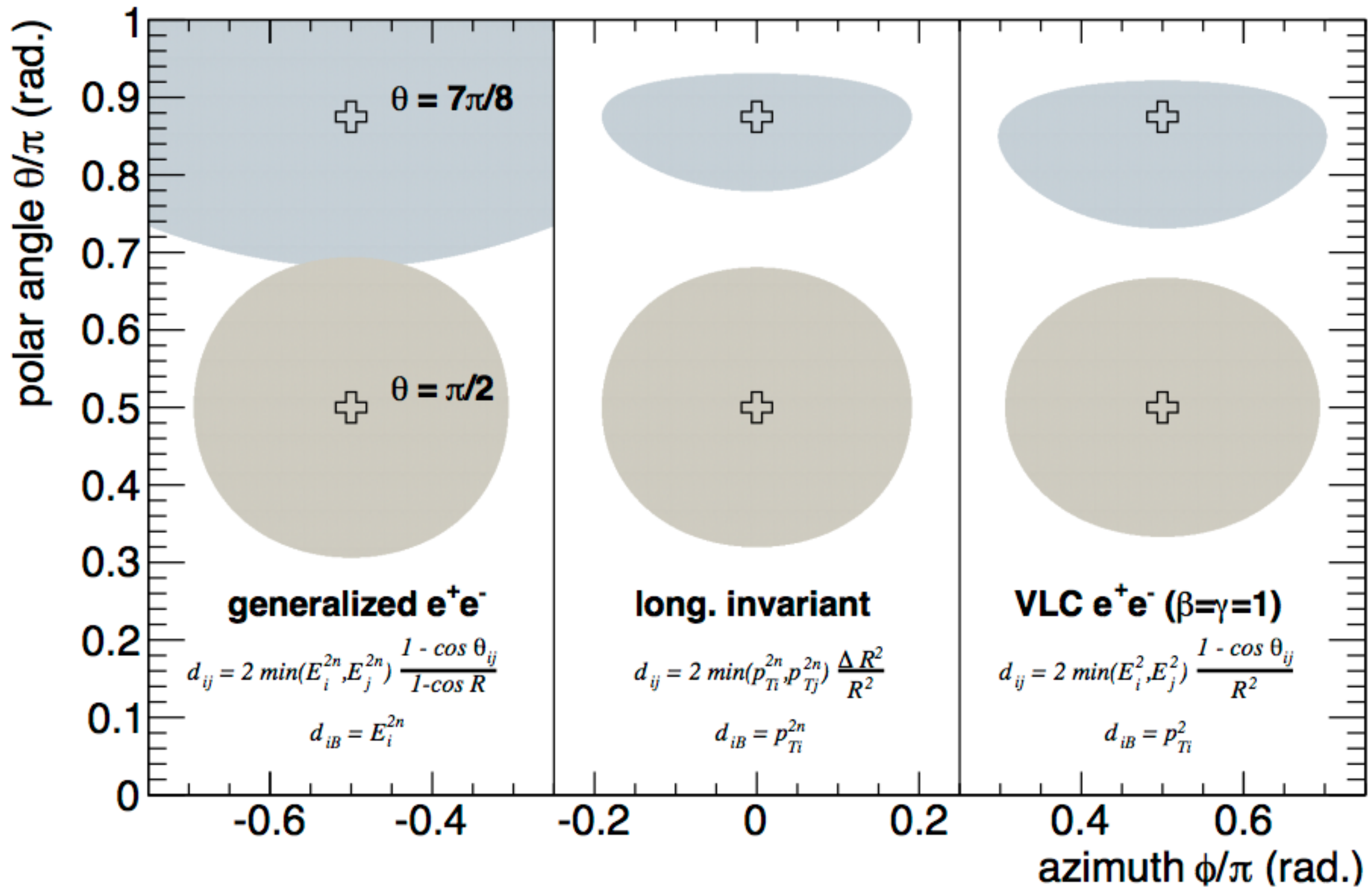




The ratio of the inter-particle distance and the beam distance: $d_{ij}/d_{iB}$ drives the robustness to (forward) background: the decision to assign the particle to final-state or beam jets depends on this ratio (and R)

Long. inv. $k_t$'s robustness is indeed due to its increasing $d_{ij}/d_{iB}$ ratio

VLC with $\beta$=1 is similar (by design) to long. inv. $k_t$

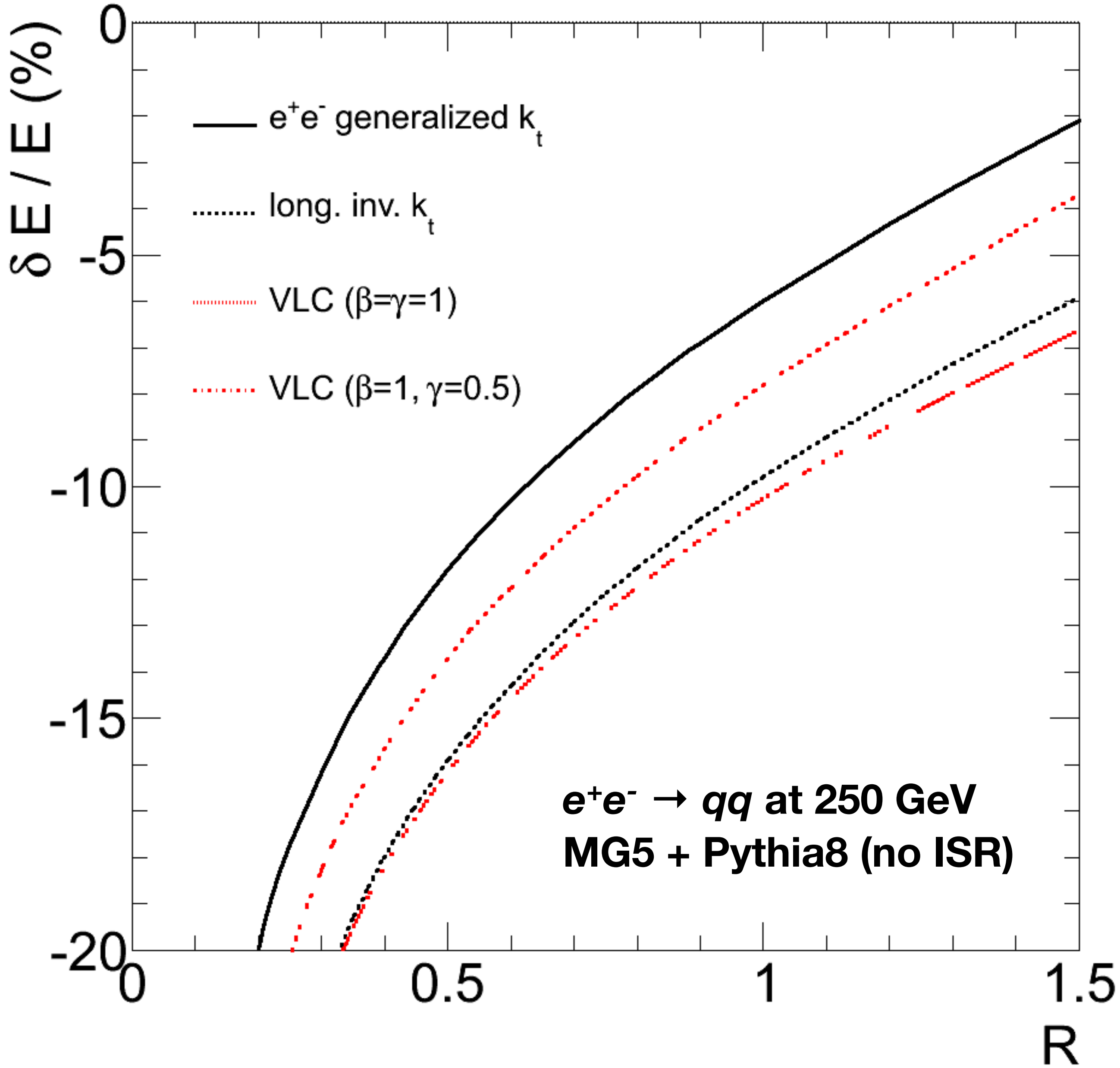
# Comparison of the jet sizes




**The footprint or area of jets depends on the jet algorithm**

Three algorithms that yield a similar, circular area in the central detector produce very different jets in the forward region → *crucial feature for the fight against* $\gamma\gamma$ → hadrons

# Jet energy corrections



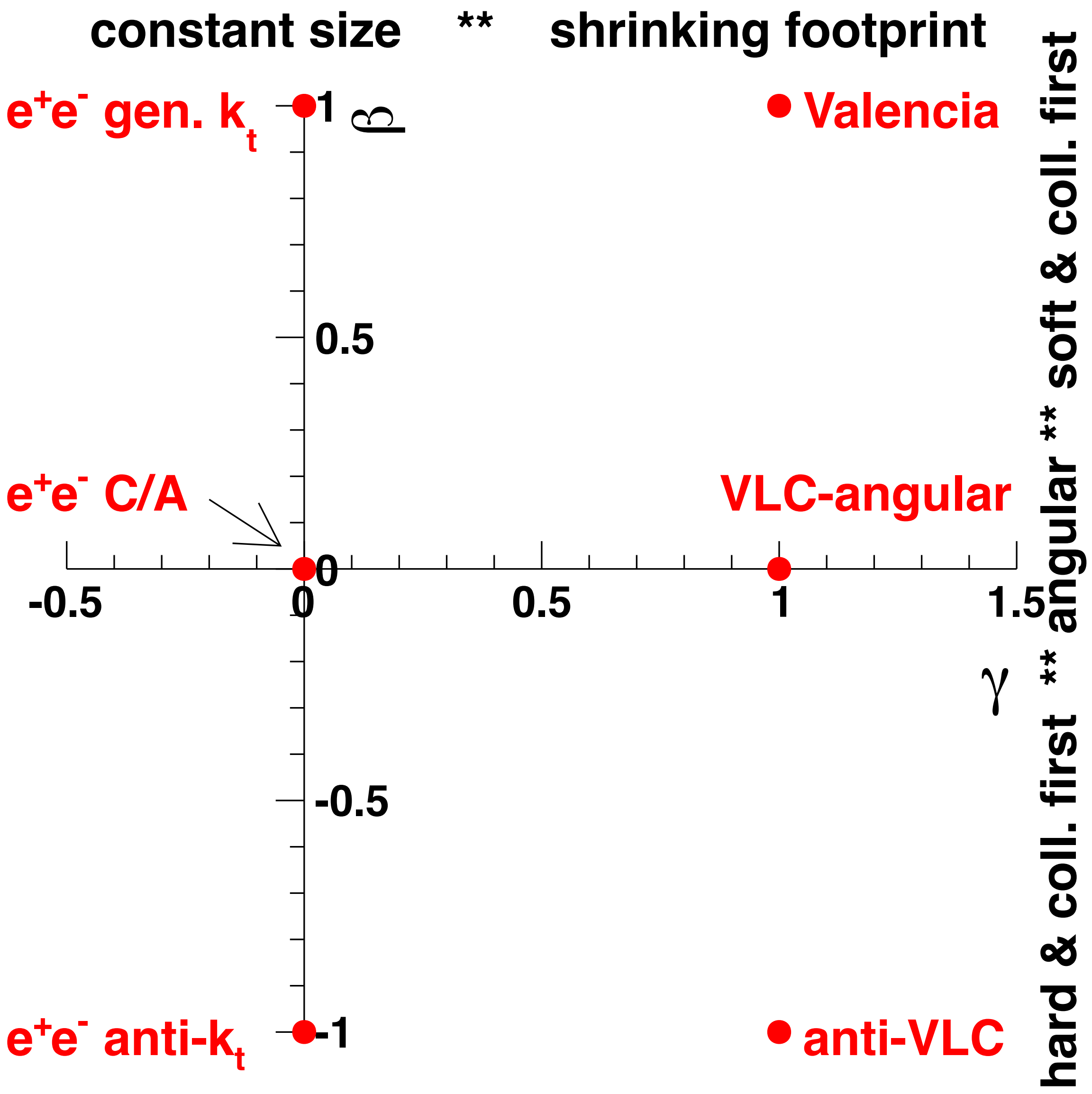

- **Perturbative corrections grow as** jet size shrinks (**1/R**)
- **VLC** with **$\beta$=$\gamma$=1** follows **long. inv. $k_t$**
- Dialling $\gamma$ **< 1** leads to **reduced corrections**

- **Non-perturbative corrections** and **difference between NLO vs LO** corrections show the **same trends as perturbative corrections.**
- Very **small corrections** (typically **%-level for R>0.3**)

**Greater mitigation of the background (shrinking jets) leads to larger corrections.**

**VLC algorithm allows tune $\gamma$ parameter depending on the situation.**

# The (β,γ) space




LOCATION OF EXISTING ALGORITHMS IN THE (β,γ) SPACE

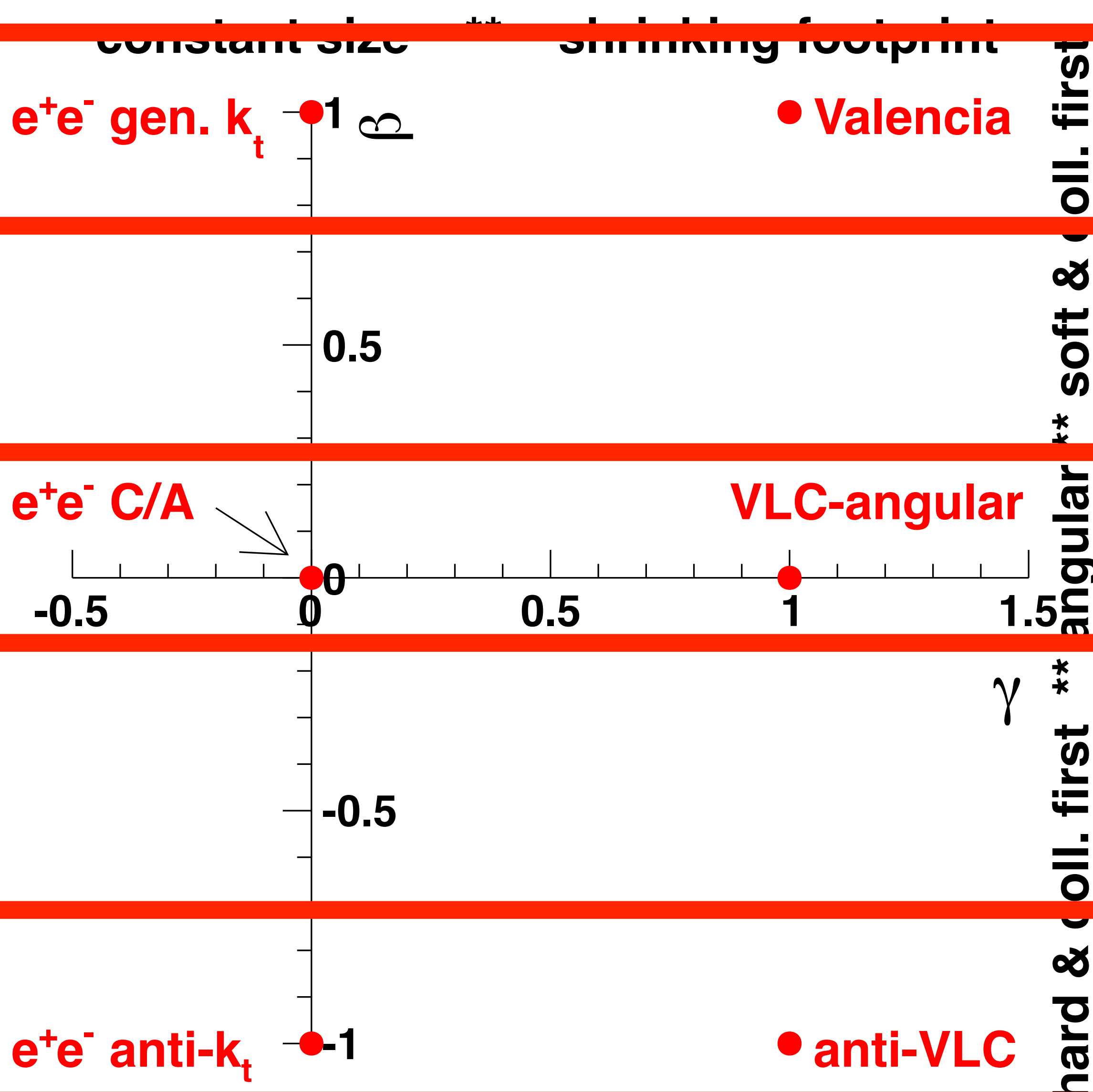
β = clustering order

γ = jet area with polar angle

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

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

# The ($\beta,\gamma$) space



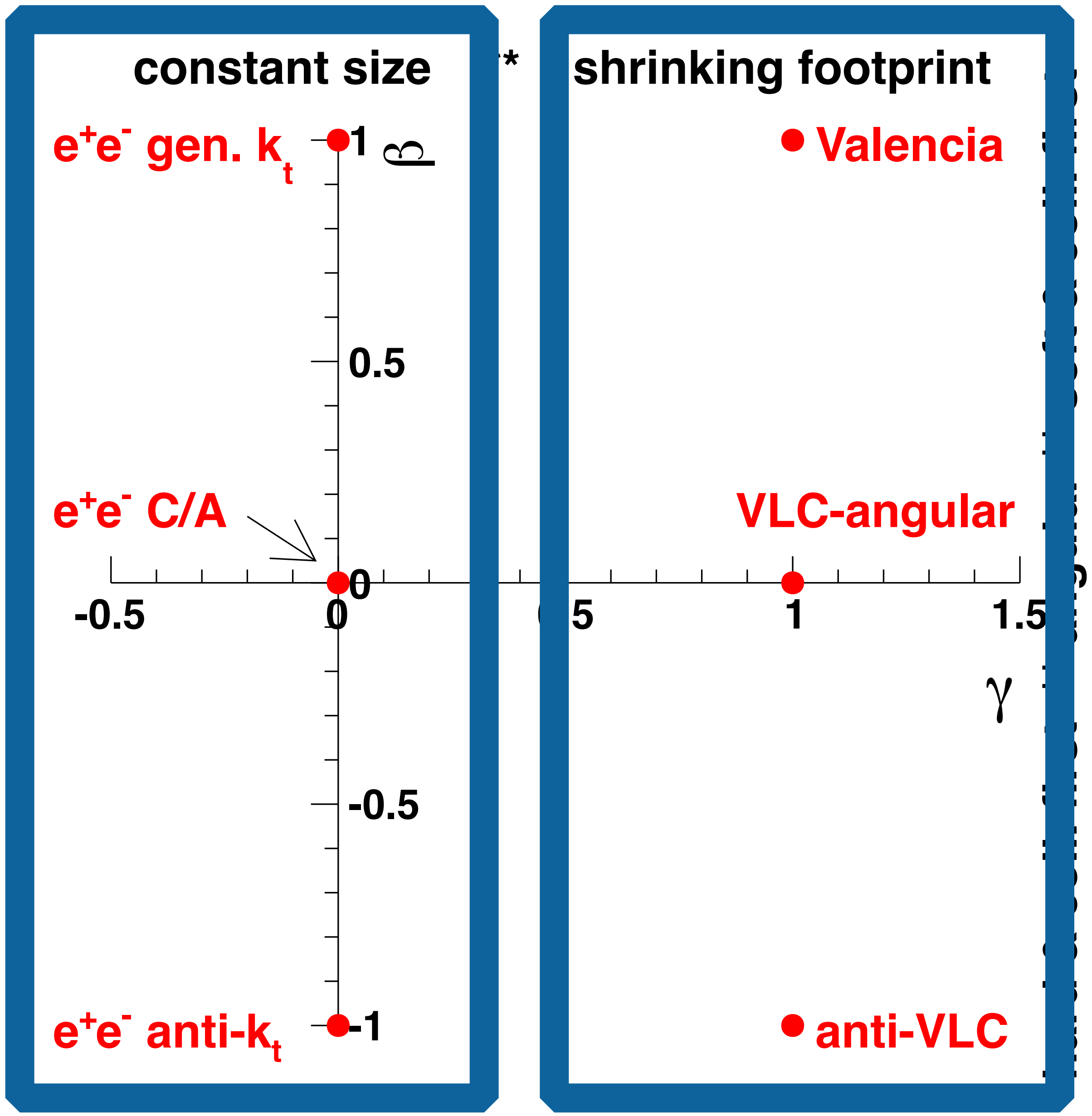

$\beta = 1$ soft, collinear radiation

$\beta = 0$ angular clustering

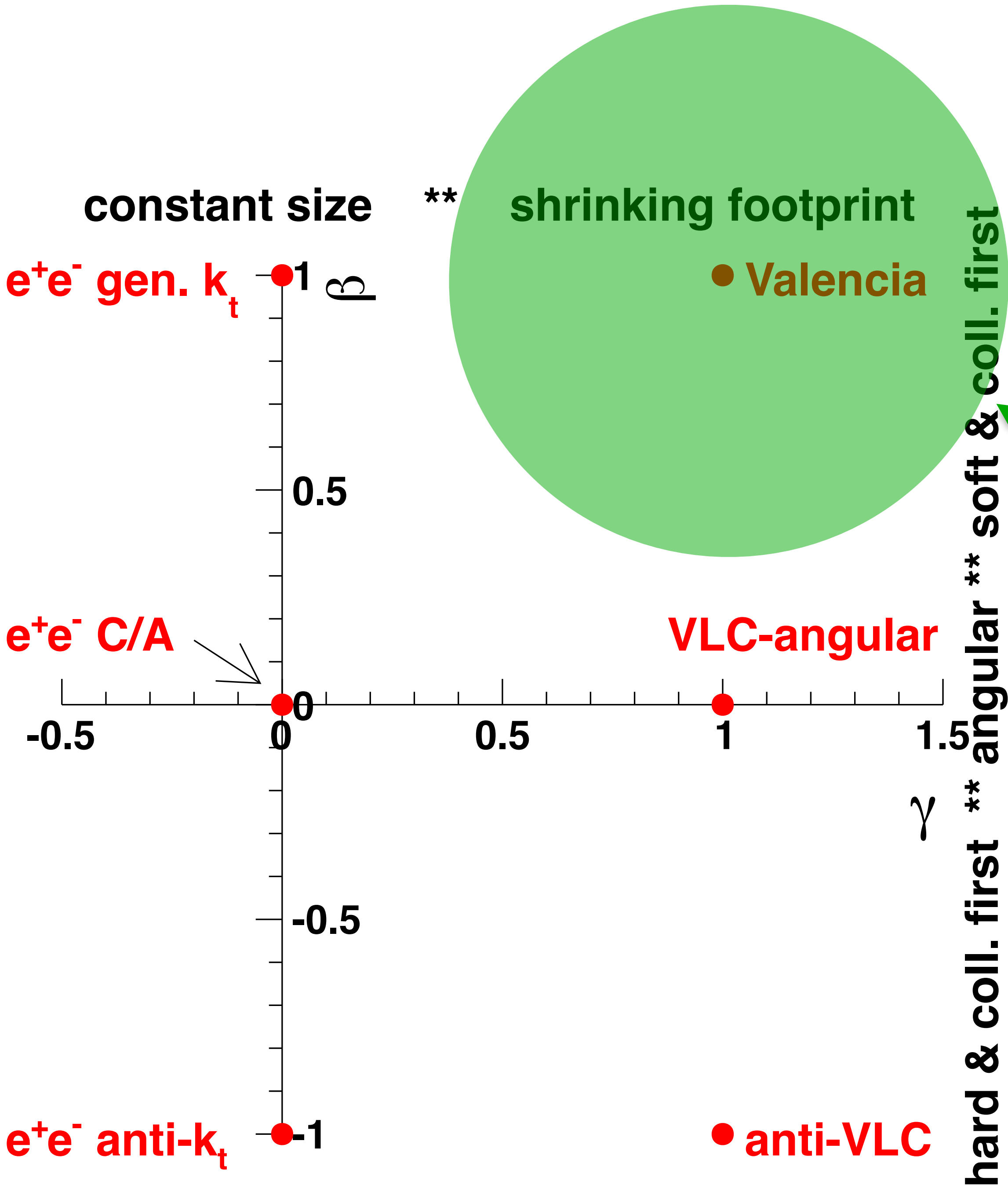
$\beta = -1$ hard, collinear radiation

# The (β,γ) space

$\gamma = 0$ $\gamma \geq 1$



# The (β,γ) space




**We only work close to the default configuration of the VLC algorithm**

$$\beta \simeq \gamma \simeq 1$$

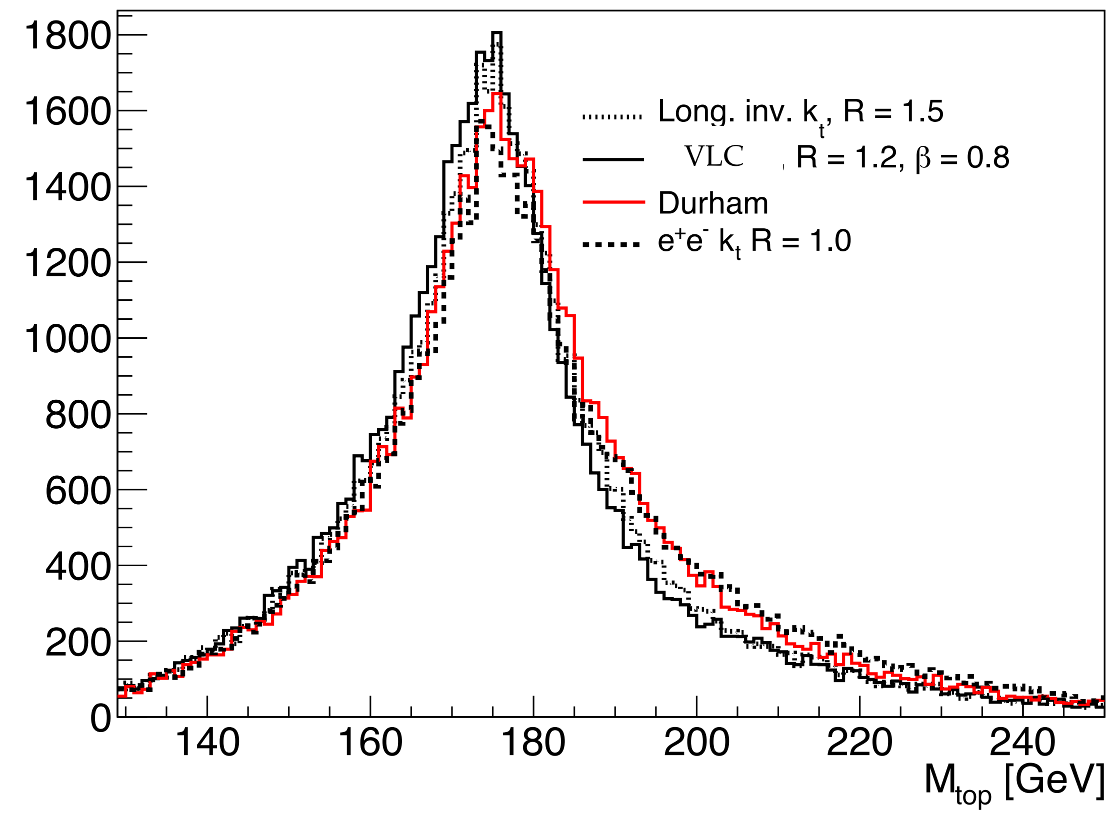
# ILC realistic benchmark

**IFIC/LAL study of ILC lepton+jets $t\bar{t}$**
@ 500 GeV, [arXiv:1307.8102]

$t\bar{t} \rightarrow b\bar{b}j_1j_2l\nu$

We consider four jet reconstruction algorithms

- **Durham** algorithm
- **Generic e+e− $k_t$** algorithm with beam jets with R = 1
- **Longitudinally invariant $k_t$** algorithm with R = 1.5
- **VLC** algorithm with R =1.2 and β = 0.8.




The choice of parameters corresponds to the optimal setting determined in a scan over a broad range of parameters.

**Durham is affected by $\gamma\gamma$ -> hadrons, longitudinally invariant $k_t$ and VLC OK**

# Resolution on jets reconstruction

**Degradation** of all jet-related measurements due to $\gamma\gamma \rightarrow$ **hadrons background**

| RMS$_{90}$ [GeV] | $E_{4j}$ | $E_W$ | $m_W$ | $E_t$ | $m_t$ |
|---|---|---|---|---|---|
| Durham | 23.2 | 19.6 | 20.3 | 19.5 | 21.4 |
| $e^+e^-$ $k_t$ | 25.6 | 20.8 | 21.6 | 20.5 | 22.8 |
| long. inv. $k_t$ | 21.7 | 18.4 | 18.9 | 18.4 | 20.1 |
| VLC | 21.4 | 18.0 | 18.8 | 18.2 | 20.0 |

Four-jet system

Hadronic W candidate

Hadronic top candidate

**Durham and $e^+e^-$ $k_t$ are degraded**

**Long. inv. $k_t$ algorithm and VLC offer better reconstruction for all hadronic observables**

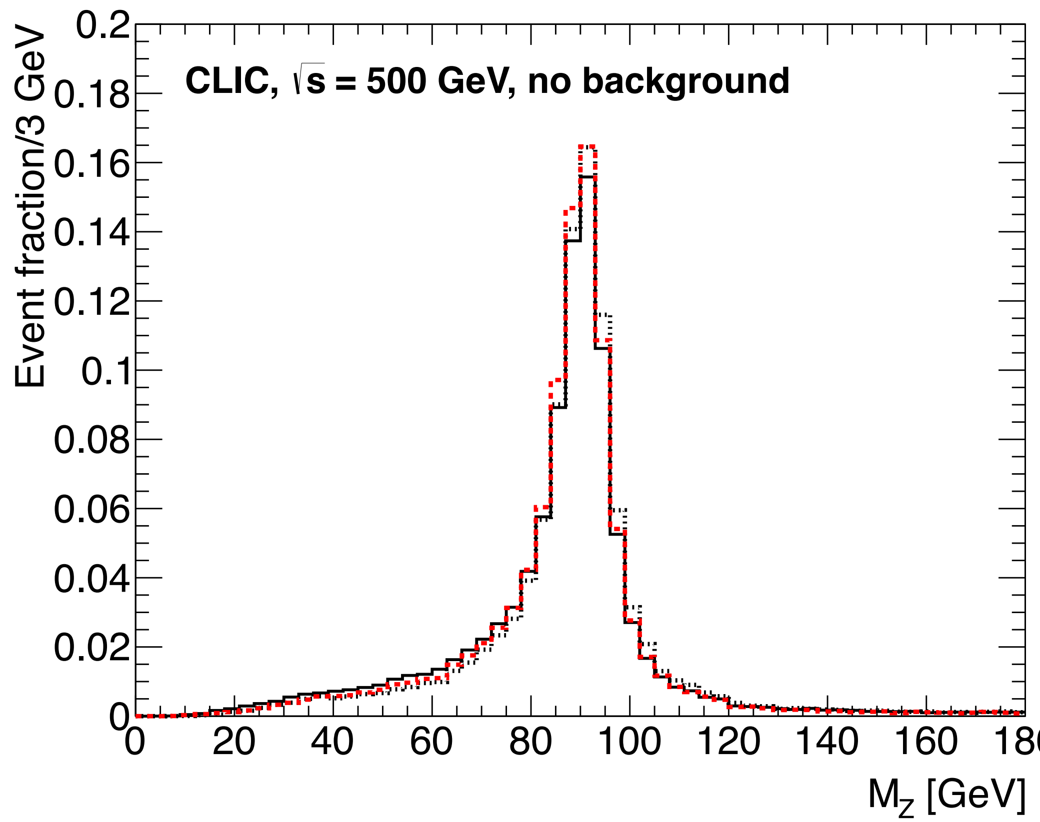
# CLIC realistic benchmark including background

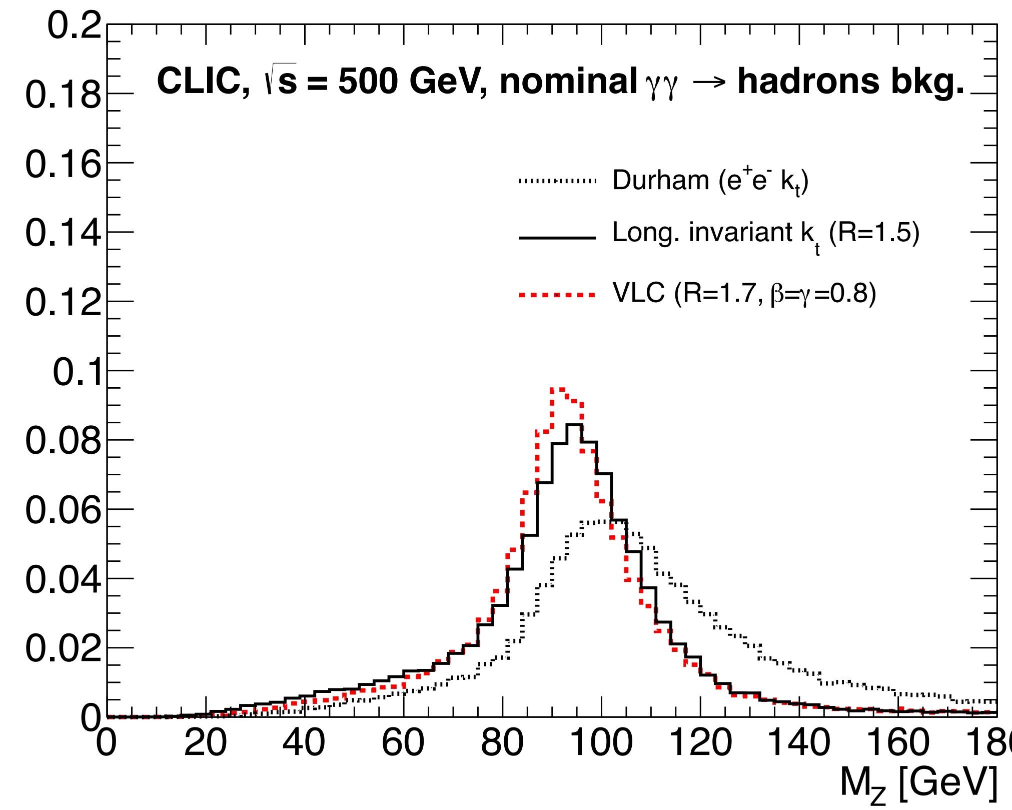
**CLIC di-boson (ZZ) production @ 500 GeV**

Reconstruct Particle Flow objects using PANDORA

Reconstruct jets (exclusive, n=4) and form Z boson candidates, selecting best jet pairs

$$\chi^2 = \frac{(E_{Z1} - E_{Z2})^2}{(250\ \mathrm{GeV})^2} + \frac{(m_{Z1} - m_{Z2})^2}{(91\ \mathrm{GeV})^2} + \frac{\angle(Z_1, Z_2)}{(\pi)^2}$$




**Jet energy reconstruction with nominal background much less degraded with algorithms with shrinking footprint (long. Invariant algorithms, VLC) than $e^+e^-$ algorithms (CLIC, high energy)**

# Jet reconstruction performance

The previous results in numbers: central value, width of the Z-boson mass peak and $\mathrm{RMS}_{90}$

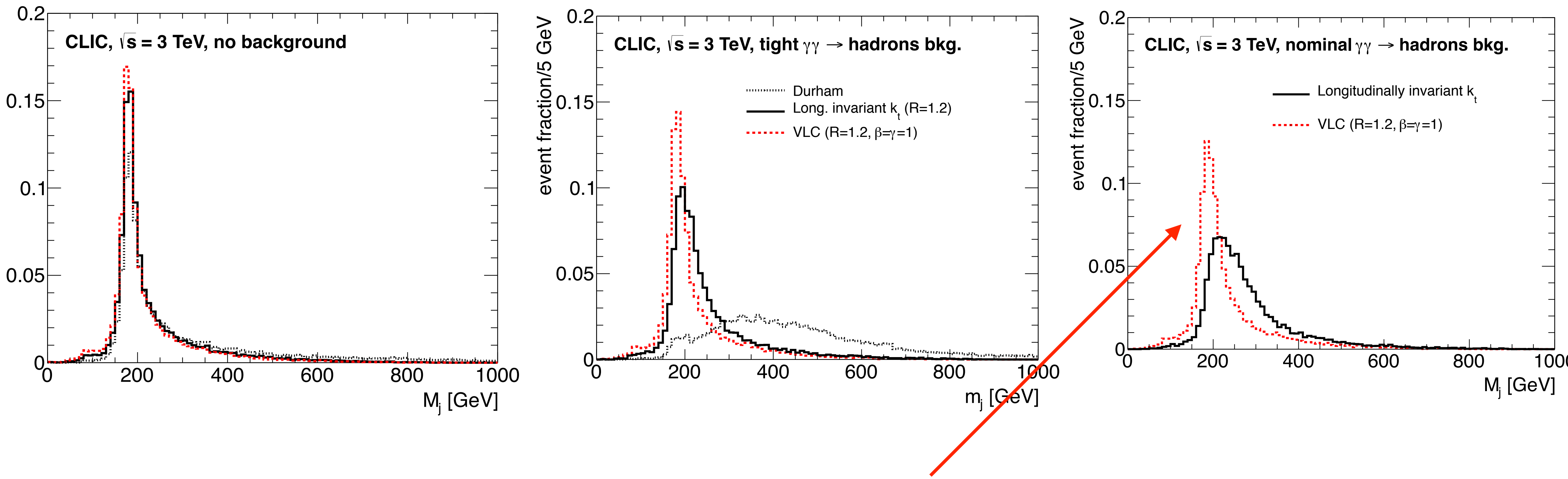
| CLIC, $\sqrt{s}$ = 500 GeV, no background overlay | | | |
|---|---|---|---|
| [ GeV ] | $m_Z$ | $\sigma_Z$ | $\mathrm{RMS}_{90}$ |
| Durham | 90.6 | 5.4 | 13.8 |
| long. inv. $k_t$ | 90.4 | 5.3 | 14.3 |
| VLC ($\beta = \gamma = 1$) | 90.3 | 5.2 | 12.5 |
| **CLIC, $\sqrt{s}$ = 500 GeV, nominal PFO selection** | | | |
| [ GeV ] | $m_Z$ | $\sigma_Z$ | $\mathrm{RMS}_{90}$ |
| Durham | 101.1 | 13.6 | 28.8 |
| long. inv. $k_t$ | 92.0 | 9.0 | 17.2 |
| VLC ($\beta = \gamma = 1$) | 92.5 | 9.2 | 16.2 |

**$e^+e^-$ style algorithm can compete with hadron collider algorithm**

# Boosted tops at CLIC@3TeV

**CLIC tt production @3TeV**

$e^+e^- \rightarrow t\bar{t} \rightarrow b\bar{b}q\bar{q}'q''\bar{q}'''$ (fully hadronic decay)




**The VLC algorithm performs significantly better than the classical algorithms, including longitudinally invariant $k_t$ .**

# Boosted tops at CLIC 3TeV

| CLIC, $\sqrt{s} = 3$ TeV, no background overlay | | | |
|---|---|---|---|
| RMS$_{90}$ [ % ] | $E_j$ (top) | $E_j$ (truth) | $m_j$ |
| Durham | 5.8 | 3.7 | 12 |
| generic $e^+e^- k_t$ | 6.2 | 2.7 | 4.5 |
| long. inv. $k_t$ | 6.1 | 2.4 | 3.4 |
| VLC | 5.9 | 2.4 | 3.4 |
| CLIC, $\sqrt{s} = 3$ GeV, tight PFO selection | | | |
| RMS$_{90}$ [ % ] | $E_j$ (top) | $E_j$ (truth) | $m_j$ |
| Durham | 7.2 | 5.6 | 44 |
| generic $e^+e^- k_t$ | 6.8 | 3.4 | 15 |
| long. inv. $k_t$ | 6.1 | 2.6 | 9.9 |
| VLC | 6.0 | 2.6 | 6.8 |
| $\sqrt{s} = 3$ GeV, nominal $\gamma\gamma \to had.$ selection | | | |
| RMS$_{90}$ [ % ] | $E_j$ (top) | $E_j$ (truth) | $m_j$ |
| long. inv. $k_t$ | 6.3 | 2.7 | 16 |
| VLC | 6.1 | 2.6 | 9.8 |

**At higher energy including the $\gamma\gamma$->hadrons background, VLC algorithm offers even better resolution than the hadron collider algorithm long. inv. $k_t$**

# Conclusions

- $\gamma\gamma \rightarrow$ **hadrons bkg. forces us to rethink $e^+e^-$ algorithms** because **old e+e- algorithms** are severally **degraded**
- **The VLC** jet algorithm retains the natural inter-particle **distance criterion for $e^+e^-$** collisions **and** offers **robust performance** in the presence of the $\gamma\gamma \rightarrow$ **hadrons background** levels expected at lepton colliders
- Shown to work on several benchmark analyses.
- In the **most challenging environment the VLC algorithm has significantly better background rejection** than the longitudinally invariant $k_t$ algorithm.
- Pre-print out on the arXiv:
  - *Boronat, Fuster, Garcia, Ros,Vos*, **A robust jet reconstruction algorithm for high-energy lepton colliders, arXiv:1404.4294**
  - https://fastjet.hepforge.org/trac/browser/contrib/contribs/ValenciaPlugin

Thank you for your attention