12th SiD Optimization meeting
29 October 2014

# Jet reconstruction at Linear Colliders

**I. García*, M. Vos**
IFIC (CSIC/UVEG)
Valencia, Spain

With help from Gavin Salam (CERN)

# Introduction

- Jet reconstruction is a crucial step in the data analysis and it has to be adapted to the conditions of the machine
- **Future high-energy lepton colliders** present an environment that **differs** in several important respects from that encountered **at the Z-pole**
- *Do we need to rethink jet reconstruction? which algorithms are most suitable?*

# A brief history of sequential recombination algorithms

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

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

**n=0**: Cambridge-Aachen
**n=1**: Longitudinally invariant $k_t$
**n=-1**: Anti-$k_t$ (LHC default)

Feed back into $e^+e^-$ algorithms

Generalised $e^+e^-$ $k_t$ algorithm

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

Moretti, Lonblad, Sjostrand, JHEP9808 (1998)
Catani, Dokshitzer, Webber, Phys.Lett. B285 (1992)
Catani, Dokshitzer, Seymour, Webber, Nucl.Phys. B406 (1994)
Ellis, Soper, Phys.Rev. D48 (1993)
All algorithms available in FastJet

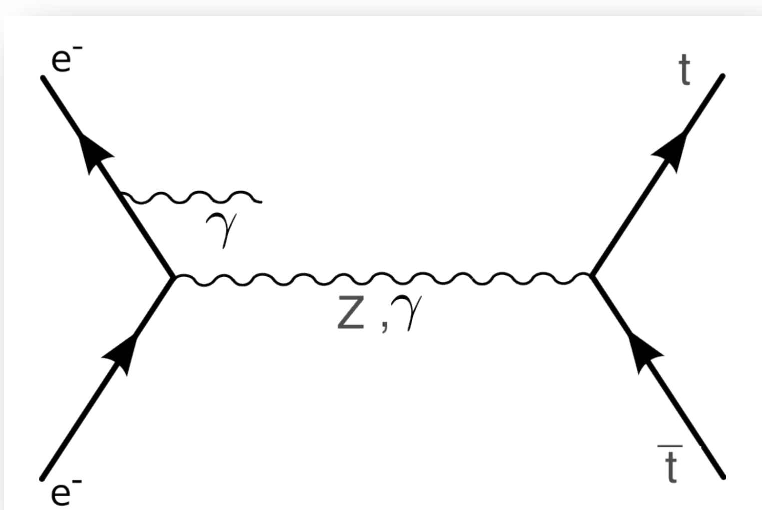
Time to rethink $e^+e^-$ algorithms!!

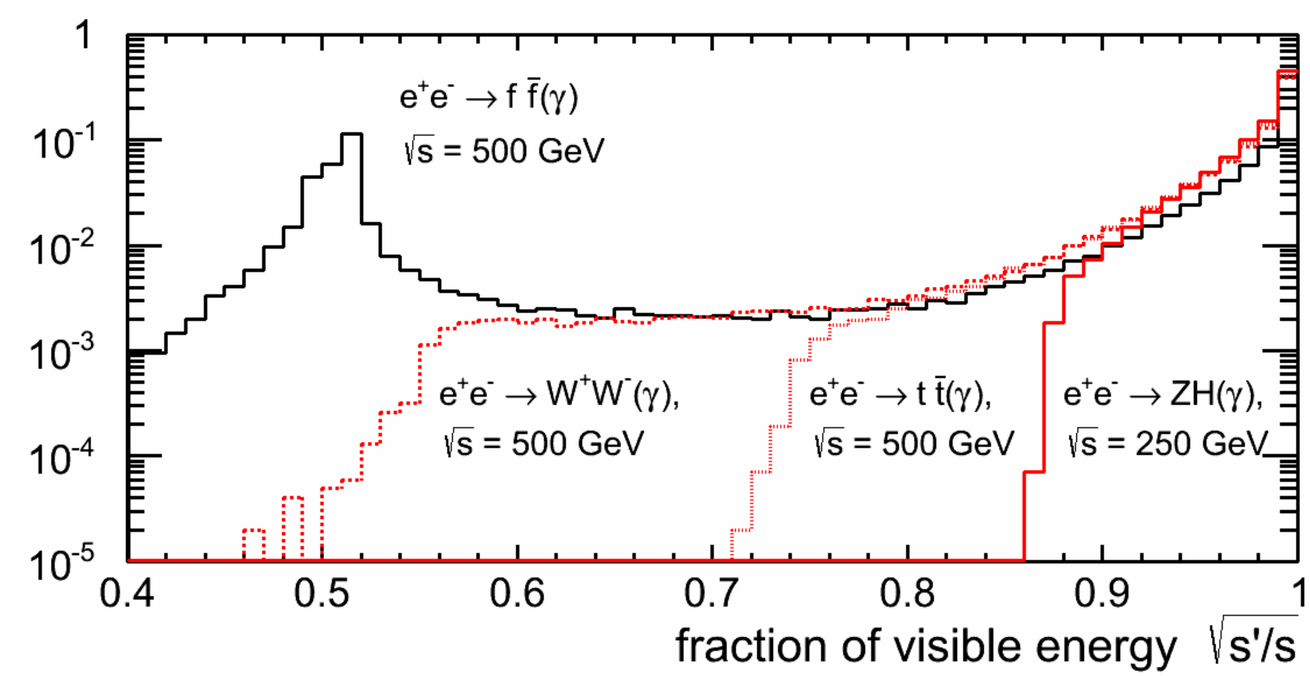
# Boost invariance at hadron colliders

- At **hadron colliders the partons** that participate in the hard process generally **carry different fractions of the initial** hadron **energy.**
- The final state acquires a substantial **Lorentz boost** along the beam axis.
  - LHC di-jets: $\beta_z \sim 1$
  - LHC tt: $\beta_z \sim 0.5$
- Replace the [**energy, polar angle**] basis by [**transverse momentum, rapidity**]

# Boost invariance at lepton colliders

- Photons emitted by the incoming beam particles (**Initial State Radiation**) can carry away a significant fractions of the nominal center-of-mass energy
- For $e^+e^- \rightarrow Z/\gamma^* \rightarrow f\bar{f}$ process, with $m_f < M_Z/2$ → large fraction of events tends to ***return*** to the **Z-pole**
- However for **most interesting processes** at a future lepton collider ISR plays a much **less important role**
- At **lepton colliders ISR** leads to **a minor boost**
- **The basis [E,θ]** is the most natural choice



# Background levels at future LC

- The **pile-up** at the **LHC** is a serious challenge that has led to a large body of work on mitigation and correction methods
- **LEP** or **SLC presented effectively negligible background**
- The $\gamma\gamma$—> **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]
- Using hadron collider algorithms can reduce these problems [CLIC CDR]

# The Valencia 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**

$$d_{iB} = p_T^{2\beta}$$

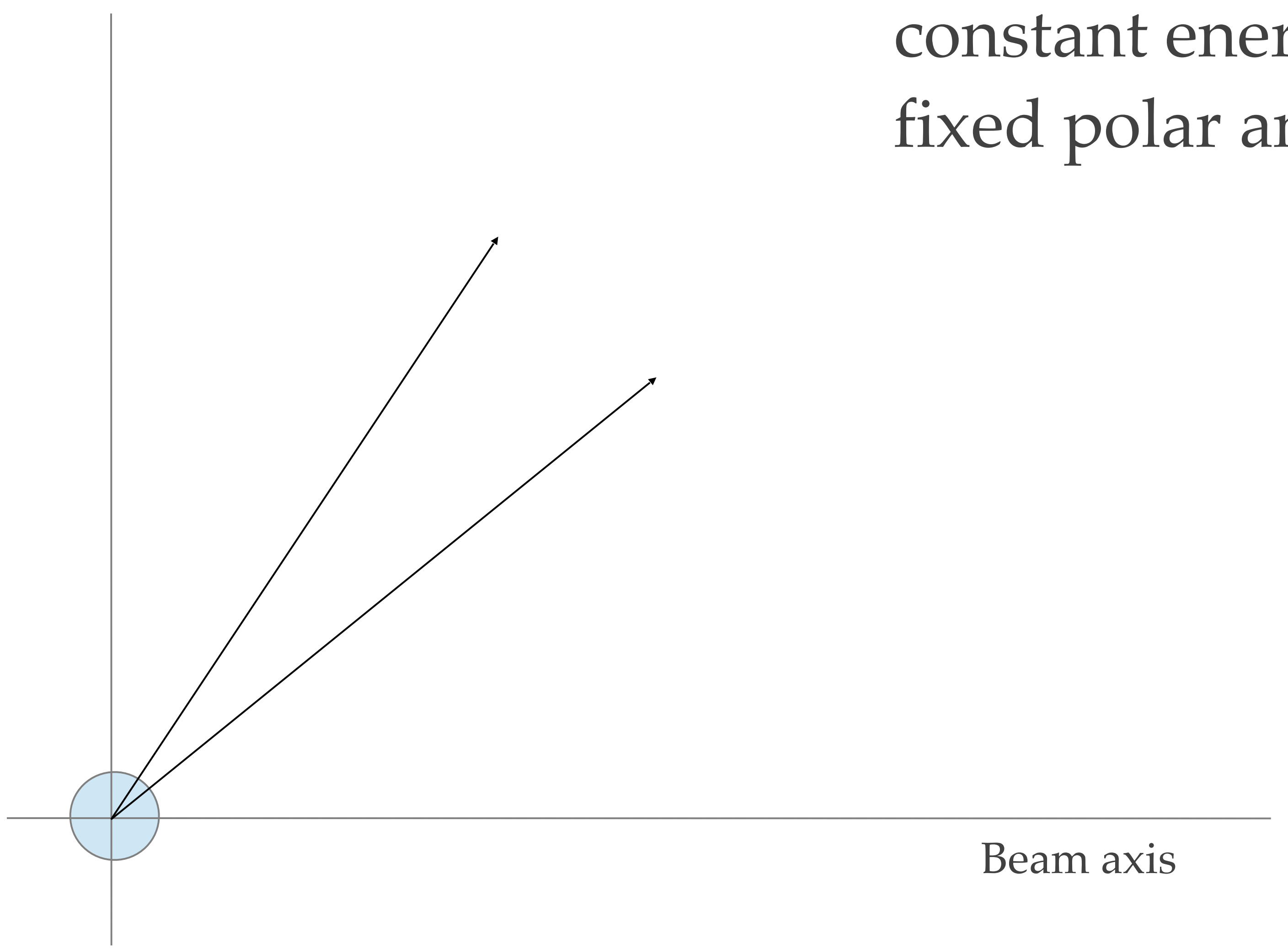
The exponent β allows to ***tune* the background rejection** level

The algorithm has been implemented as a **plugin for the *FastJet*** package and is available in fjcontrib

https://fastjet.hepforge.org/trac/browser/contrib/contribs/ValenciaJetAlgorithm

# Comparison of the distance criteria

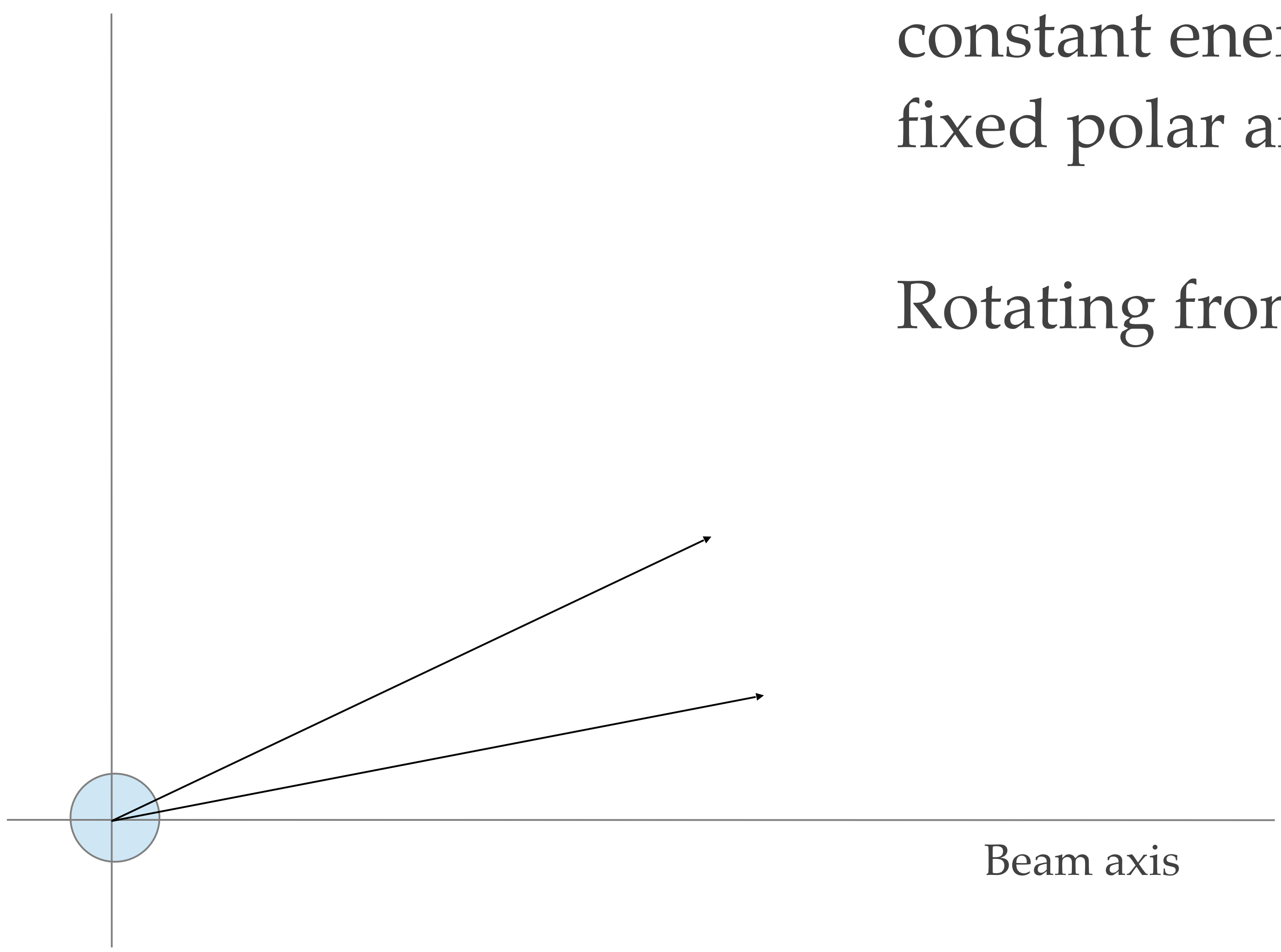
**Two test particles** with
constant energy (E = 1 GeV) and
fixed polar angle separation (100 mrad)



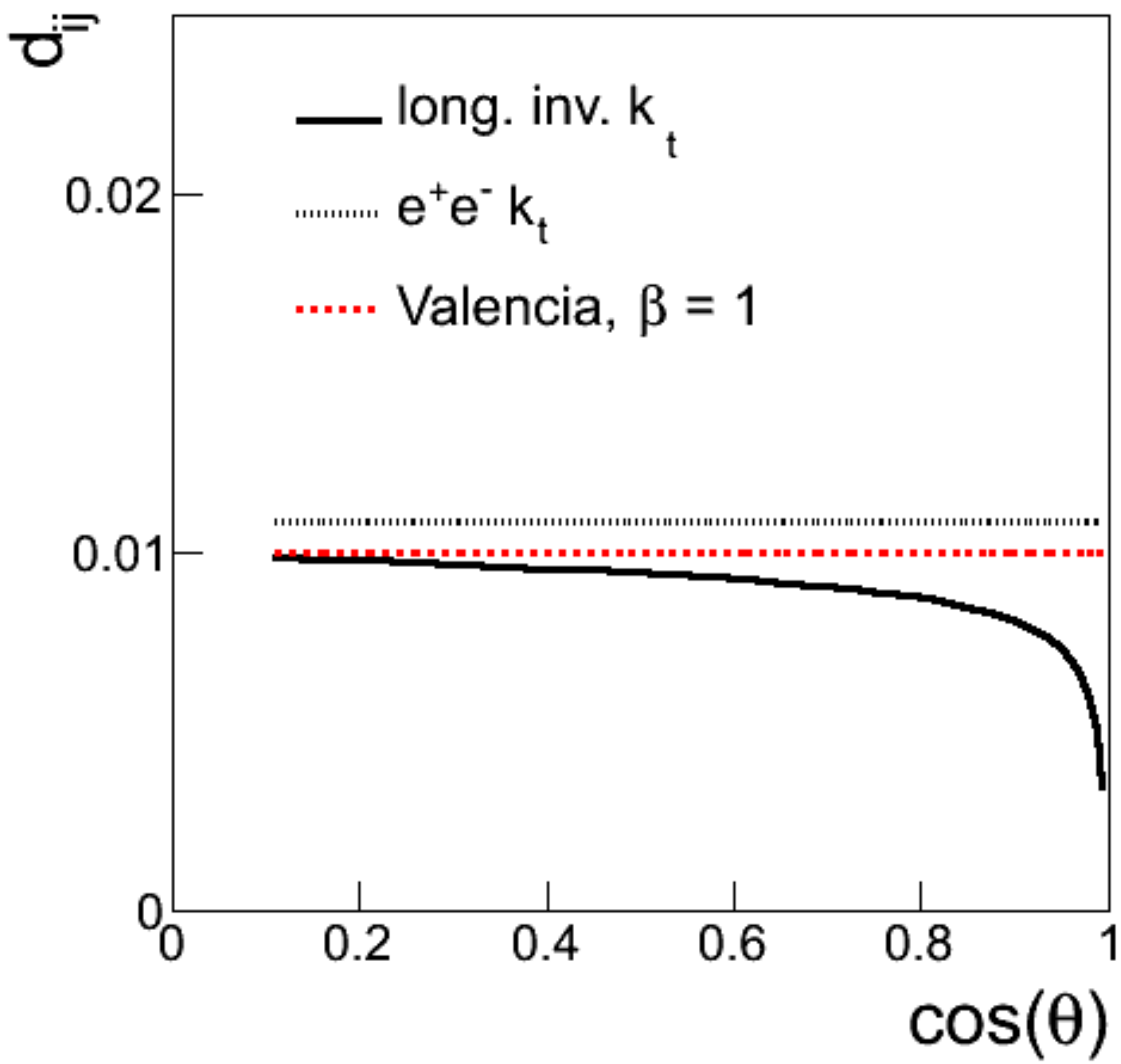
# Comparison of the distance criteria

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

Rotating from central to forward region



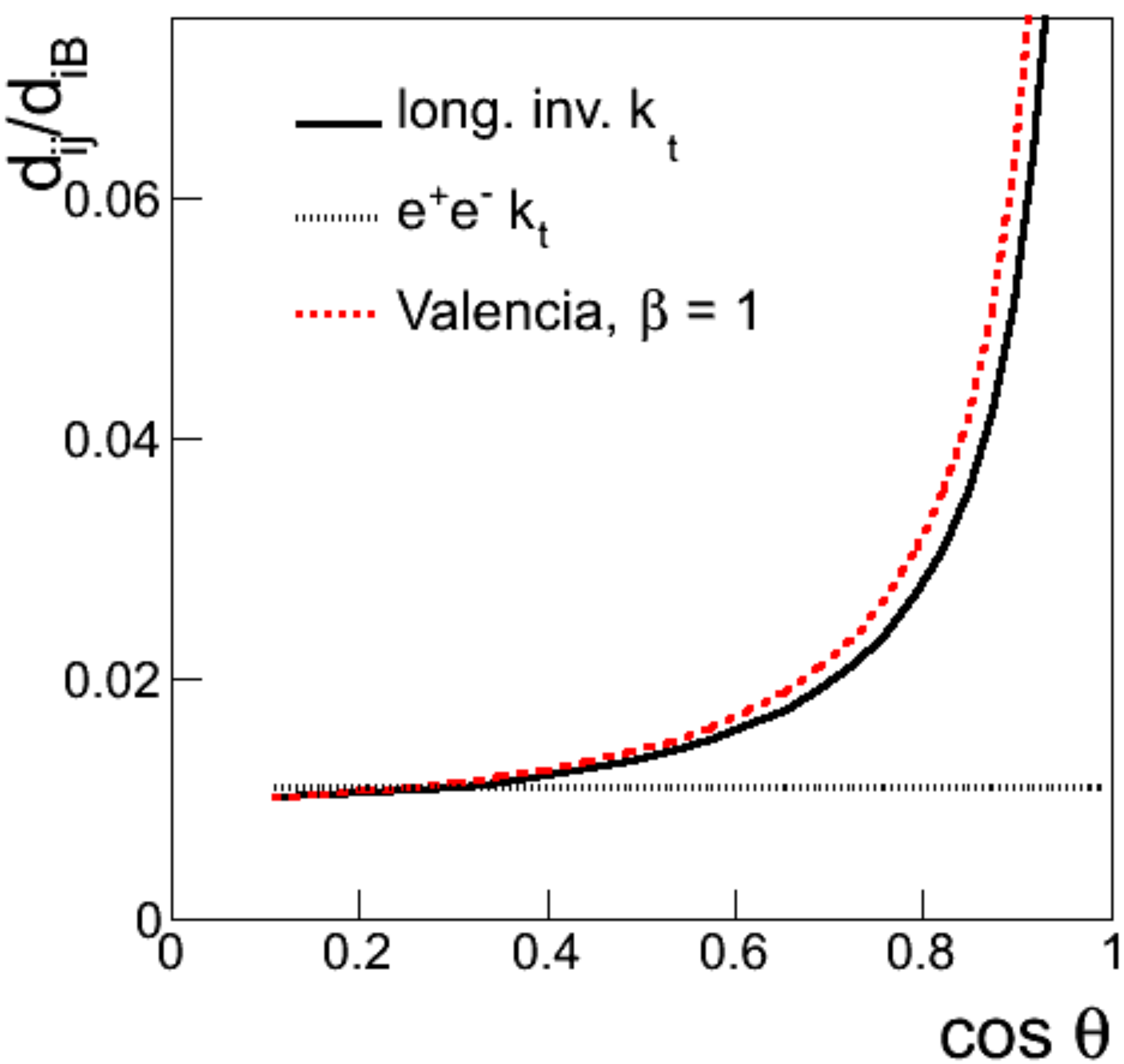
# Comparison of the distance criteria




As the two-particle system rotates into the forward region, the distance $d_{ij}$ of longitudinally invariant $k_t$ decreases
($\Delta\eta$ increases, $p_T$ decreases faster)

Traditional $e^+e^-$ algorithms and Valencia have constant $d_{ij}$

# Comparison of the distance criteria

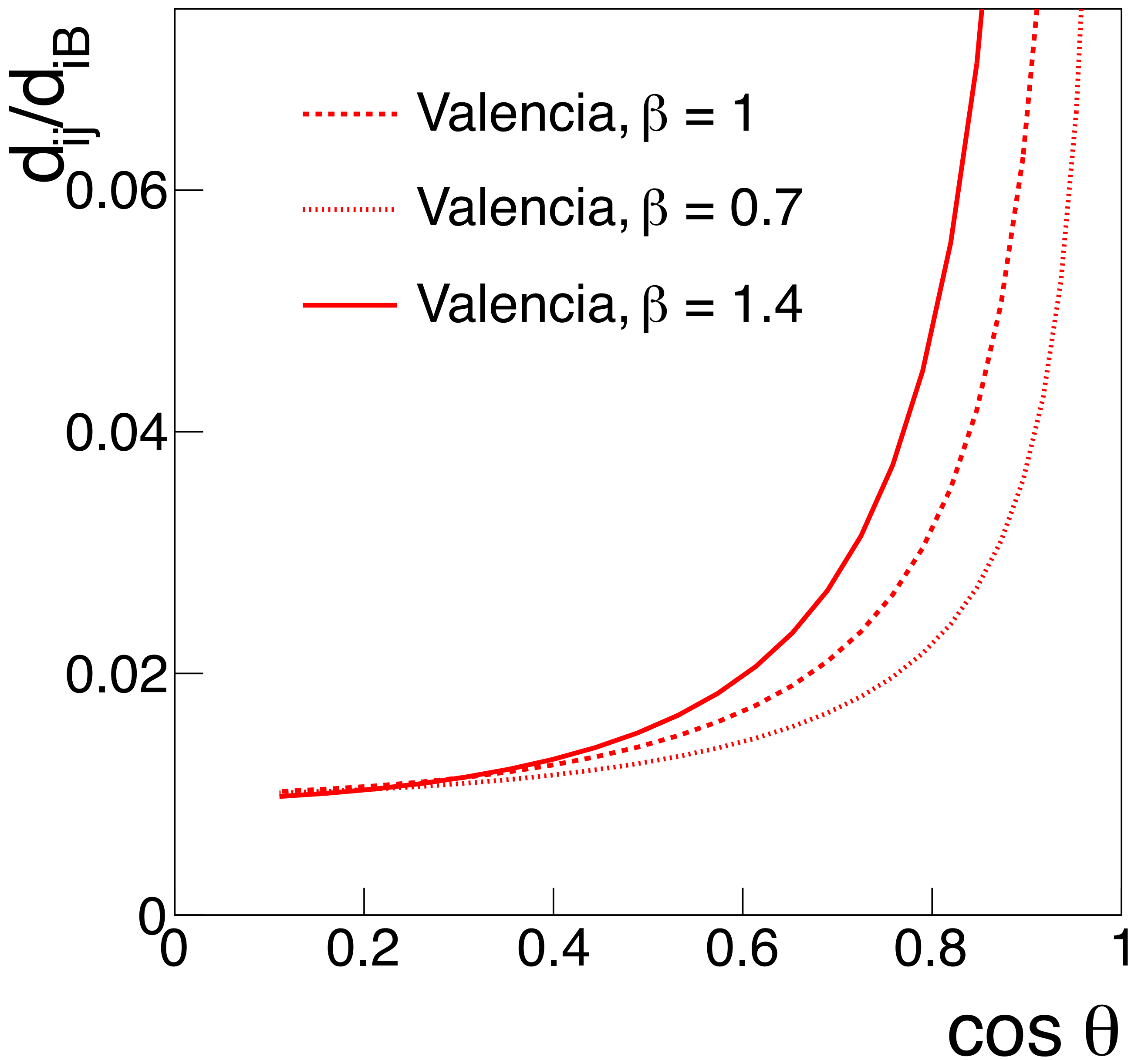



The ratio of the inter-particle distance and the beam distance: $d_{ij}/d_{iB}$ drives the robustness to (forward) background: the decision to assign the particle to final-state or beam jets depends on this ratio (and R)

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

Valencia with β=1 is similar (by design) to long. inv. $k_t$

# 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

Valencia with β=1 is similar (by design) to long. inv. $k_t$

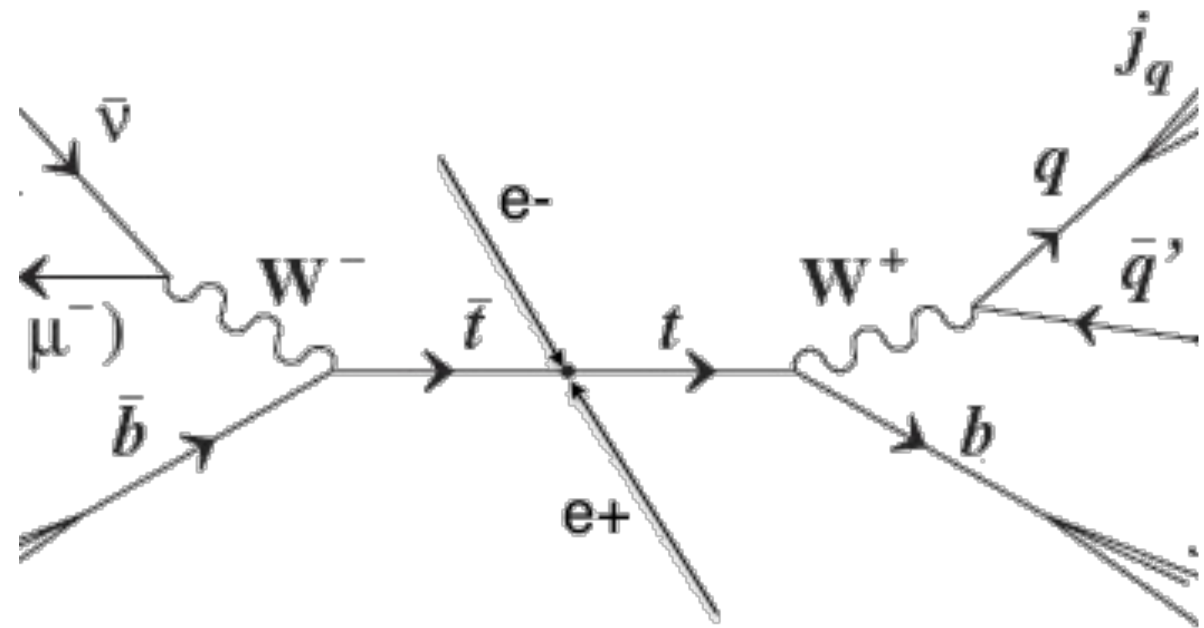
# Jet reconstruction performance

IFIC/LAL study of ILC lepton+jets tt @ 500 GeV, [arXiv:1307.8102]

Event Generation
*Whizard 1.95*

Reconstruct Particle Flow objects using PANDORA

Reconstruct jets (exclusive, n=4)




DBD Samples

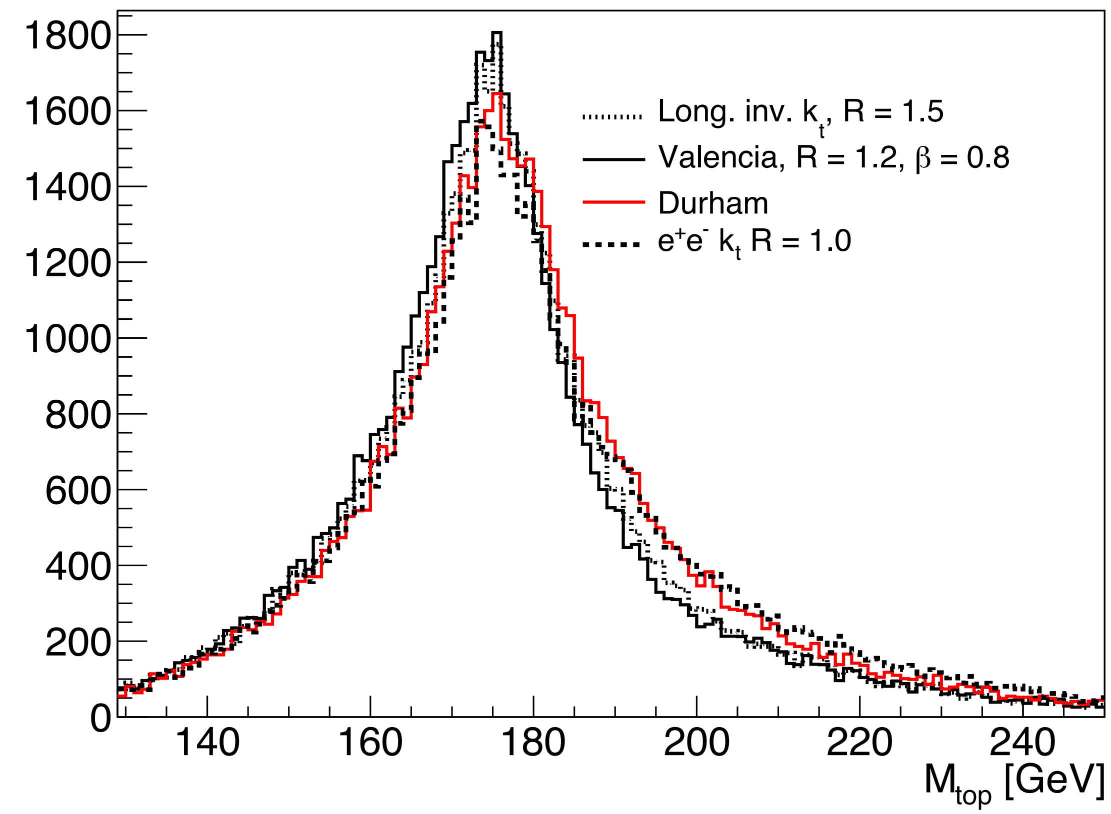
The signal is reconstructed by choosing the combination of $b$ quark jet and W boson that minimises the following equation

$$d^2 = \left(\frac{m_{cand.} - m_t}{\sigma_{m_t}}\right)^2 + \left(\frac{E_{cand.} - E_{beam}}{\sigma_{E_{cand.}}}\right)^2 + \left(\frac{p_b^* - 68}{\sigma_{p_b^*}}\right)^2 + \left(\frac{cos\theta_{bW} - 0.23}{\sigma_{cos\theta_{bW}}}\right)^2$$

# tt →(bjj)(blν)

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
- **Valencia** 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 γγ -> hadrons, longitudinally invariant $k_t$ and Valencia 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 |
| Valencia | 21.4 | 18.0 | 18.8 | 18.2 | 20.0 |

Four-jet system ($E_{4j}$)

Hadronic W candidate ($E_W$, $m_W$)

Hadronic top candidate ($E_t$, $m_t$)

**Durham and e+e- $k_t$ significantly degraded.**

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

# Jet reconstruction performance

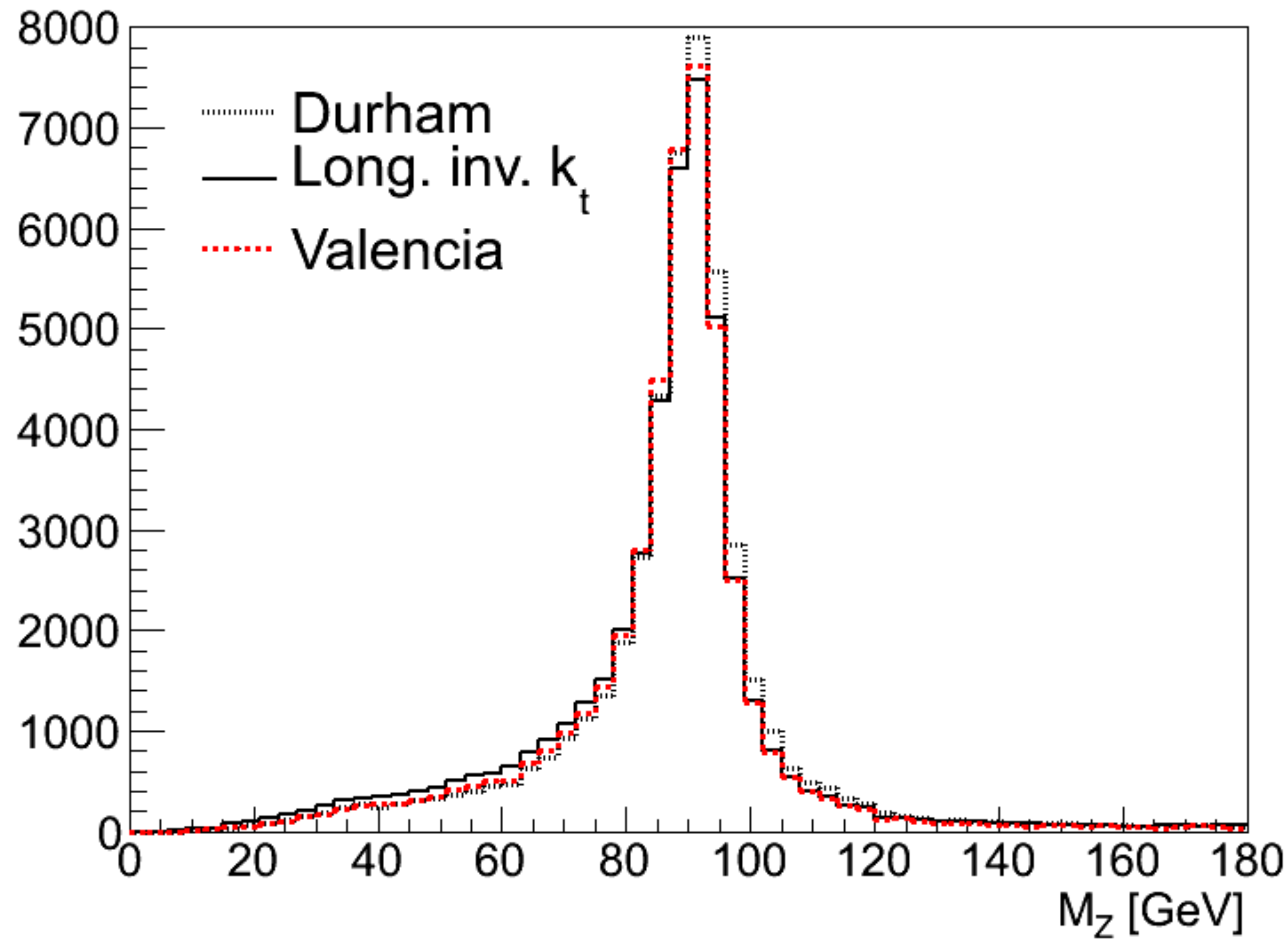
CLIC di-boson (ZZ) production @ 500 GeV

Reconstruct Particle Flow objects using PANDORA

Reconstruct jets (exclusive, n=4)

Form Z boson candidates, selecting best jet pairs

Chosen to facilitate comparison with Marshall&Thomson, CLIC CDR




**No background: it doesn't really matter which algorithm you pick**

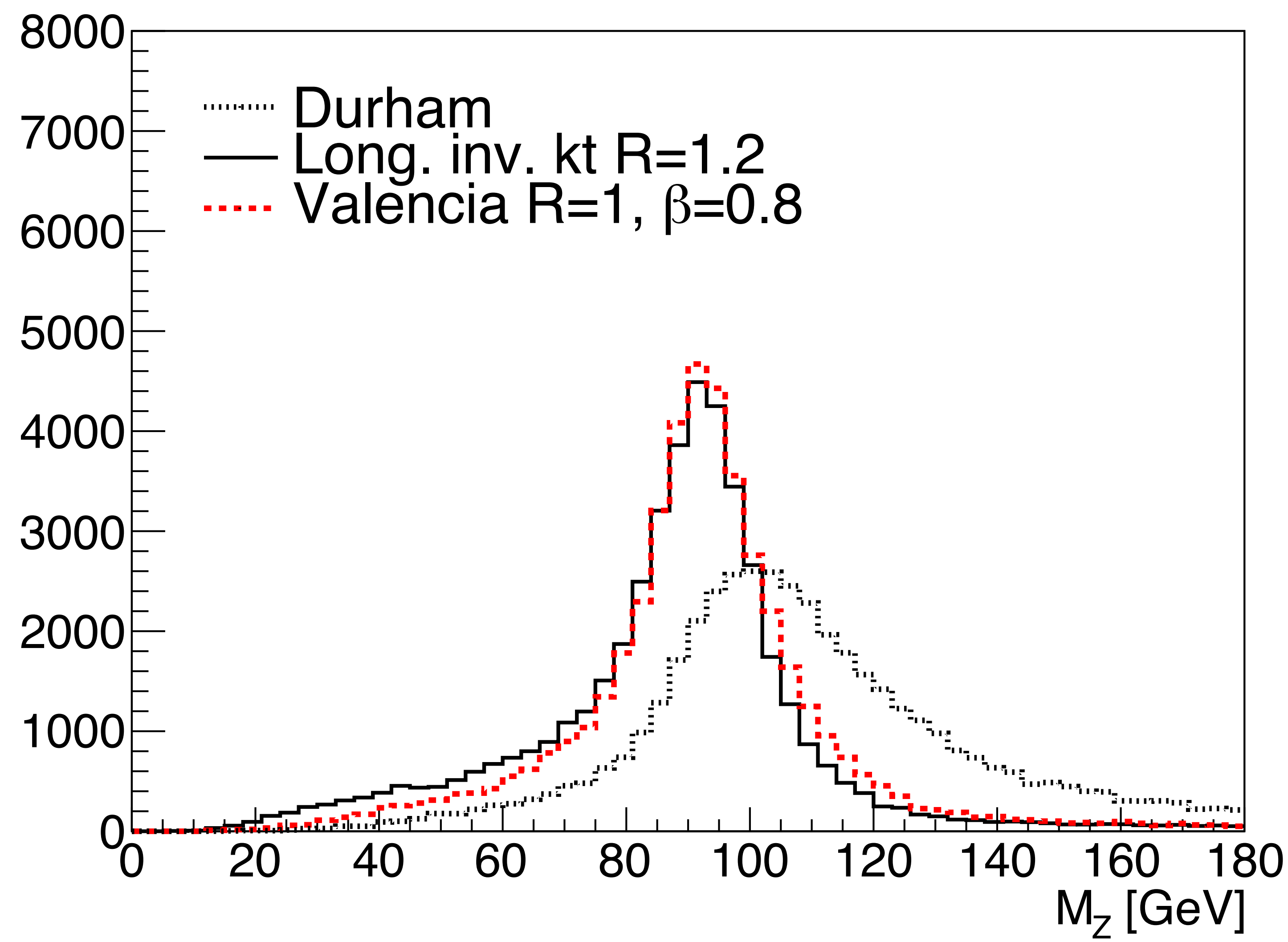
# Jet reconstruction performance

CLIC di-boson (ZZ) production @ 500 GeV
+ 300 BX of γγ → hadrons

Reconstruct Particle Flow objects using PANDORA
+ quality and timing cuts

Reconstruct jets (exclusive, n=4)

Form Z boson candidates, selecting best jet pairs




**Nominal background: Durham is severely affected, longitudinally invariant $k_t$ and Valencia OK**

# Jet reconstruction performance

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

| $\sqrt{s} = 500$ GeV, no background overlay | | | |
|---|---|---|---|
| [ GeV ] | $m_Z$ | $\sigma_Z$ | $RMS_{90}$ |
| Durham | 90.6 | 5.4 | 13.8 |
| long. inv. $k_t$ | 90.4 | 5.3 | 14.3 |
| Valencia | 90.3 | 5.2 | 12.5 |
| $\sqrt{s} = 500$ GeV, 0.3 $\gamma\gamma \rightarrow$ *hadrons* events/BX | | | |
| [ GeV ] | $m_Z$ | $\sigma_Z$ | $RMS_{90}$ |
| Durham | 101.1 | 13.6 | 28.8 |
| long. inv. $k_t$ | 92.0 | 9.0 | 17.2 |
| Valencia | 92.5 | 9.2 | 16.2 |

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

# Boosted top quarks

CLIC 3 TeV ($e^+e^- \to tt$)

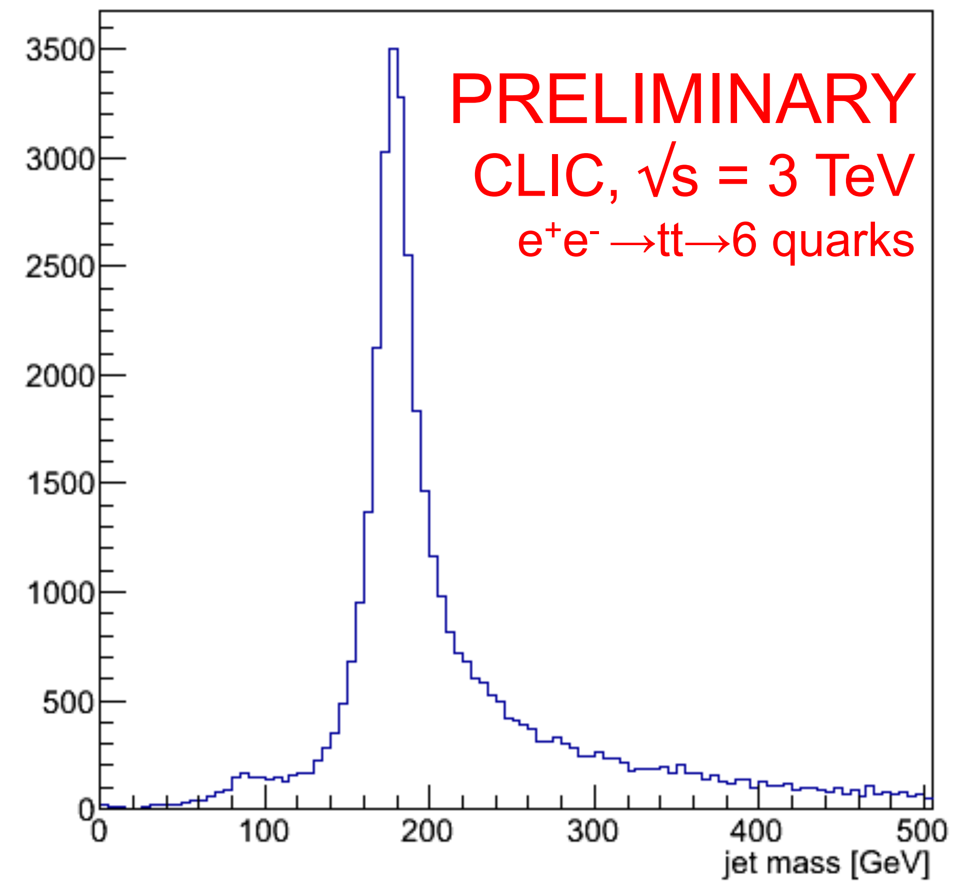
Without $\gamma\gamma \to$ hadrons background

CLIC-ILD detector simulation
PANDORA PFA

Valencia $e^+e^-$ jet algorithm ($N_j$ =2, R=1, b=1)
Could have picked long. inv. $k_t$ with R=0.8-1.2

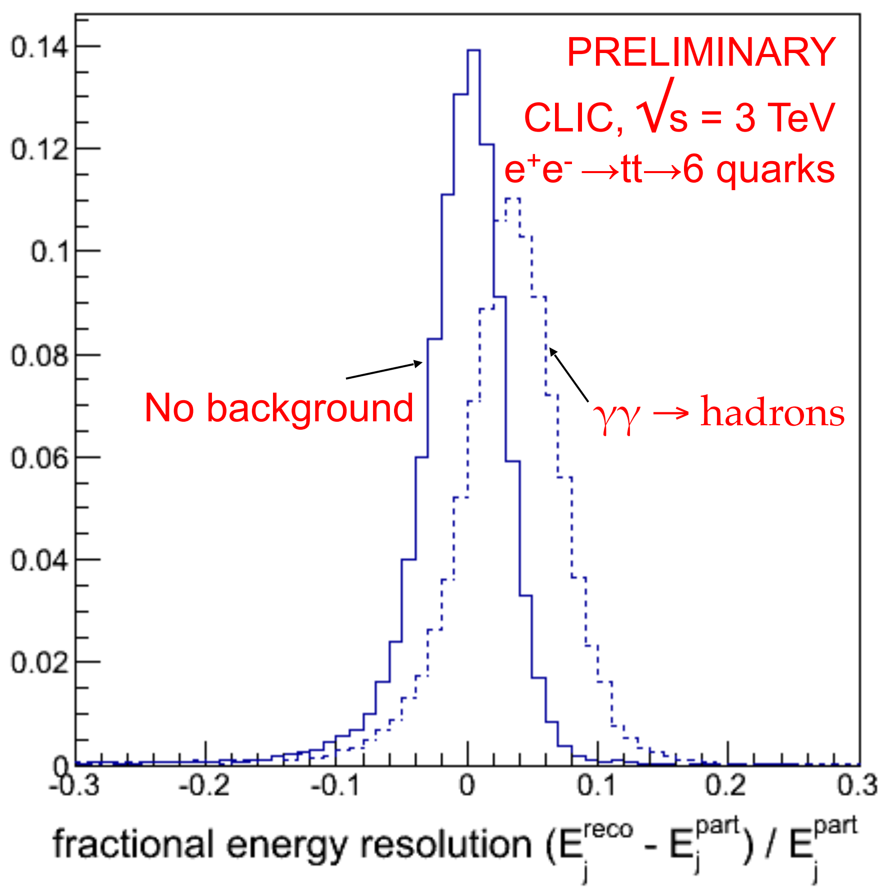
**Detector performance for boosted hadronic top jets (E~1200 GeV)**

- Energy resolution (RMS90) = 2.4%
- Jet mass resolution (RMS90) = 3.2%




Note: resolution considers reconstructed energy versus stable particle jets; relative to the actual top parton the energy resolution is 5% and the width of the mass peak ~7%

# Boosted top quarks




CLIC 3 TeV $e^+e^- \to tt$

Adding $\gamma\gamma \to$ hadrons background

CLIC-ILD detector simulation

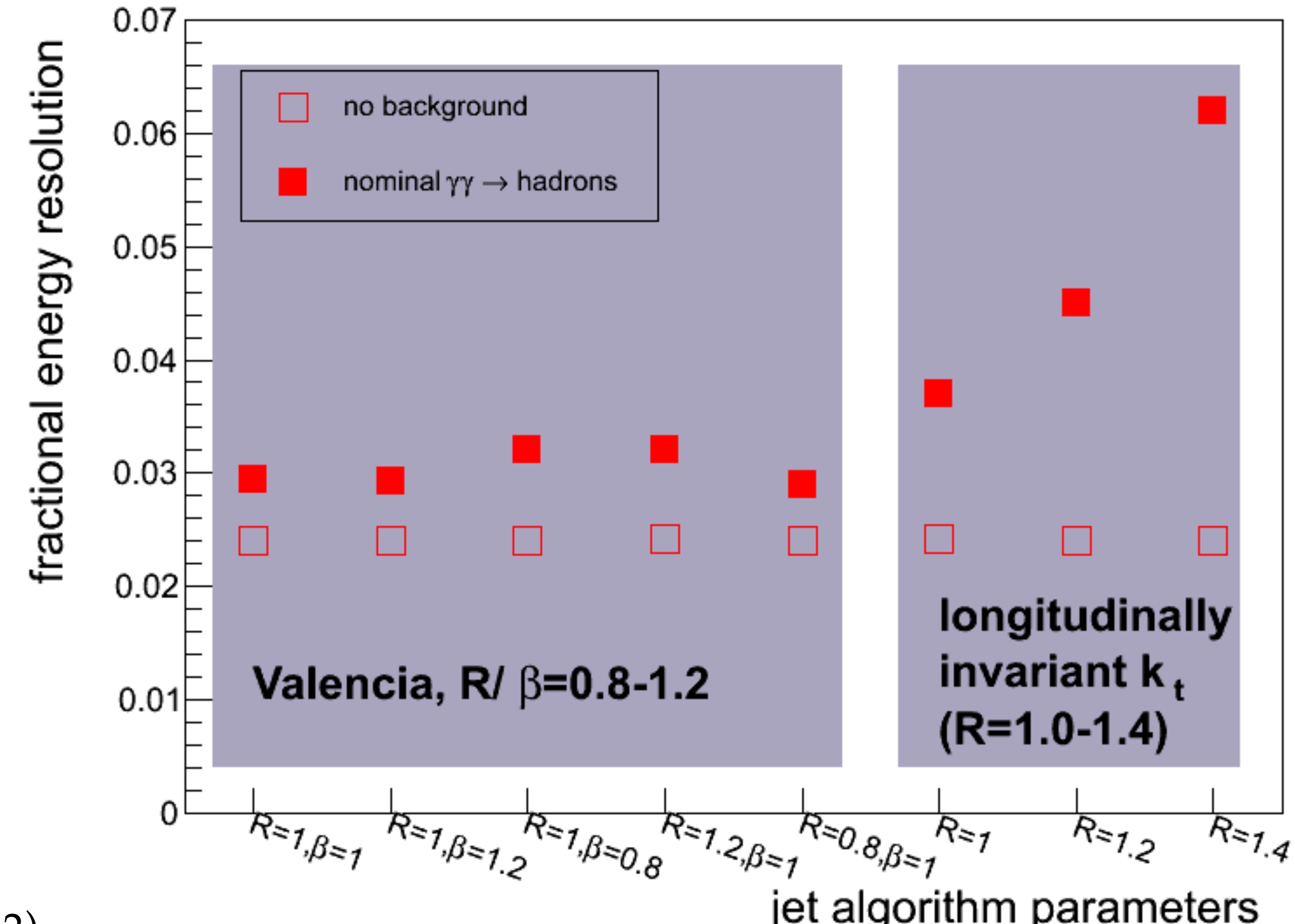
PANDORA PFA + quality and timing cuts

Valencia $e^+e^-$ jet algorithm ($N_j$ =2, R=1, b=1.2)

Significantly better now than long. inv. $k_t$ with R=0.8-1.2

**Background has impact on fat jets:**

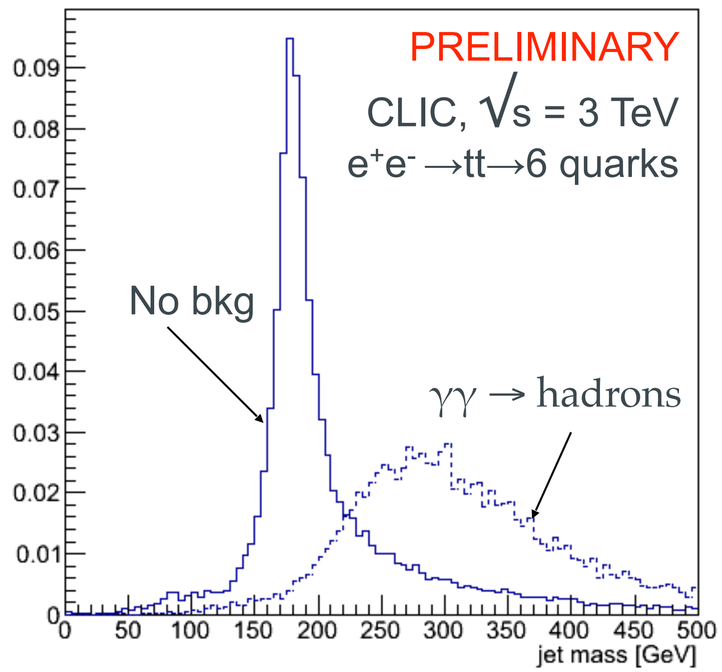
Energy resolution degraded 2.4% → 2.9%



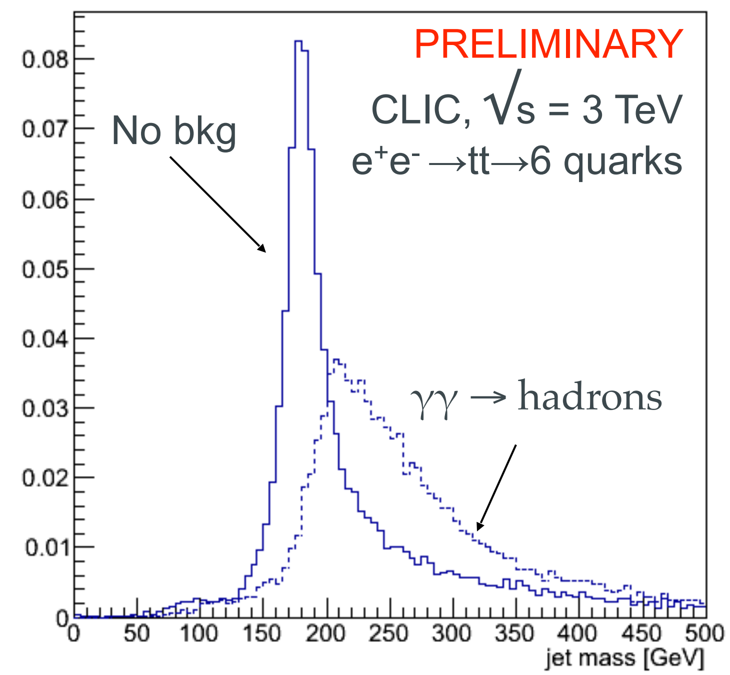

Note: particle jets used to determine resolution do not contain particles from $\gamma\gamma \to$ hadrons

# Boosted top quarks

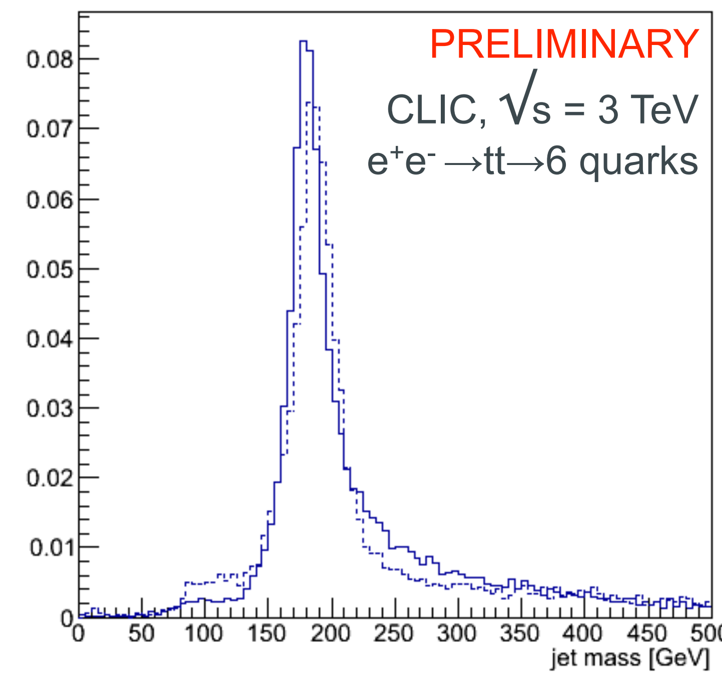
Longitudinally invariant $k_t$ (R=1)




Valencia (R=1, β=1)




Valencia trimming




With $\gamma\gamma \rightarrow$ hadrons background

**Background has a profound impact on fat jet substructure:**
Raw jet mass resolution badly degraded (from dream 3.2% to nightmare 16%)

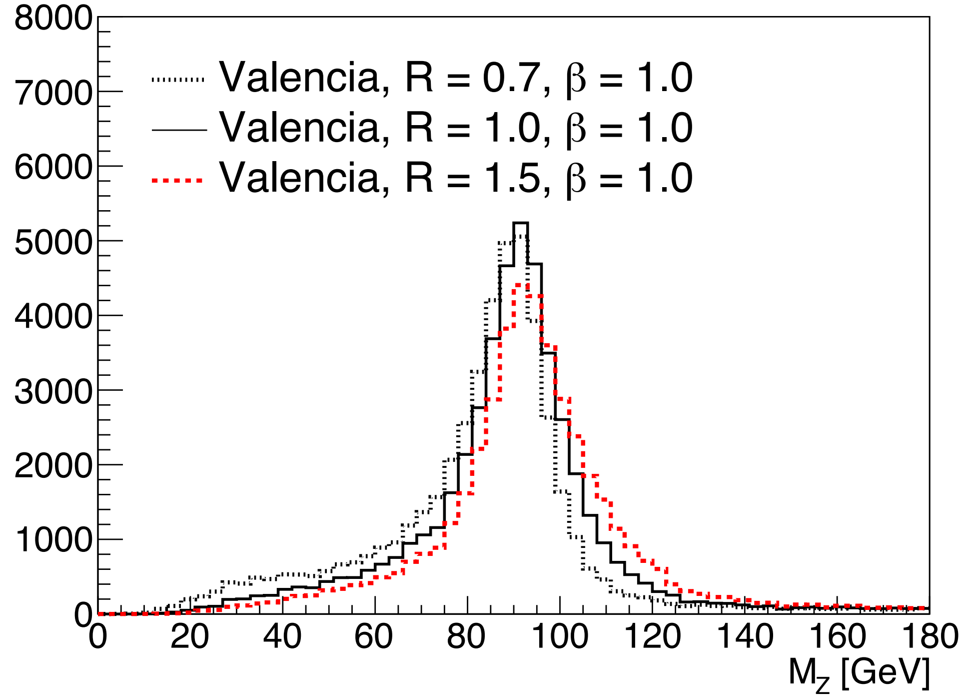
Preliminary: grooming jets restores jet mass resolution to ~4%
Results correspond to a primitive $e^+e^-$ variant of trimming based on 3+3 Valencia R=0.2 jets
→ optimisation needed

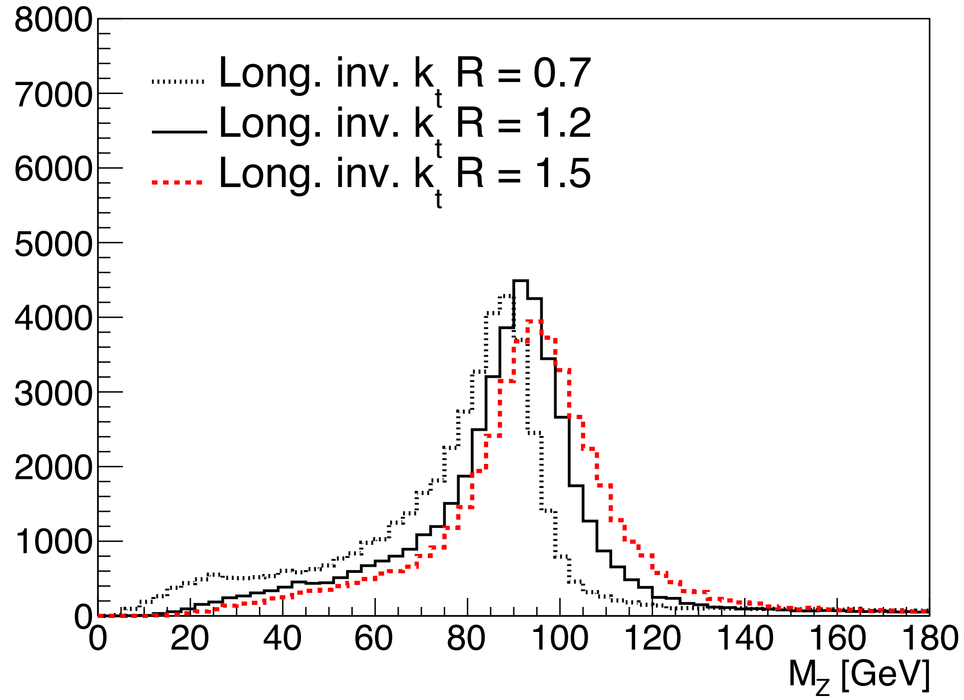
# Conclusions

- $\gamma\gamma \to$ **hadrons bkg. forces us to rethink $e^+e^-$ algorithms** because **old e+e- algorithms** are severally **degraded**
- **The Valencia** jet algorithm retains the natural inter-particle **distance criterion for $e^+e^-$** collisions **and** offers **robust performance** in the presence of the $\gamma\gamma \to$ **hadrons background** levels expected at lepton colliders
- Shown to work on several benchmark analyses. Pre-print out on the arXiv since last week: *Boronat, Garcia,Vos*, **A new jet reconstruction algorithm for lepton colliders, arXiv:1404.4294**
- Do try this at home! https://fastjet.hepforge.org/trac/browser/contrib/contribs/ValenciaJetAlgorithm
- Contact me if help is needed: Ignacio.Garcia@ific.uv.es

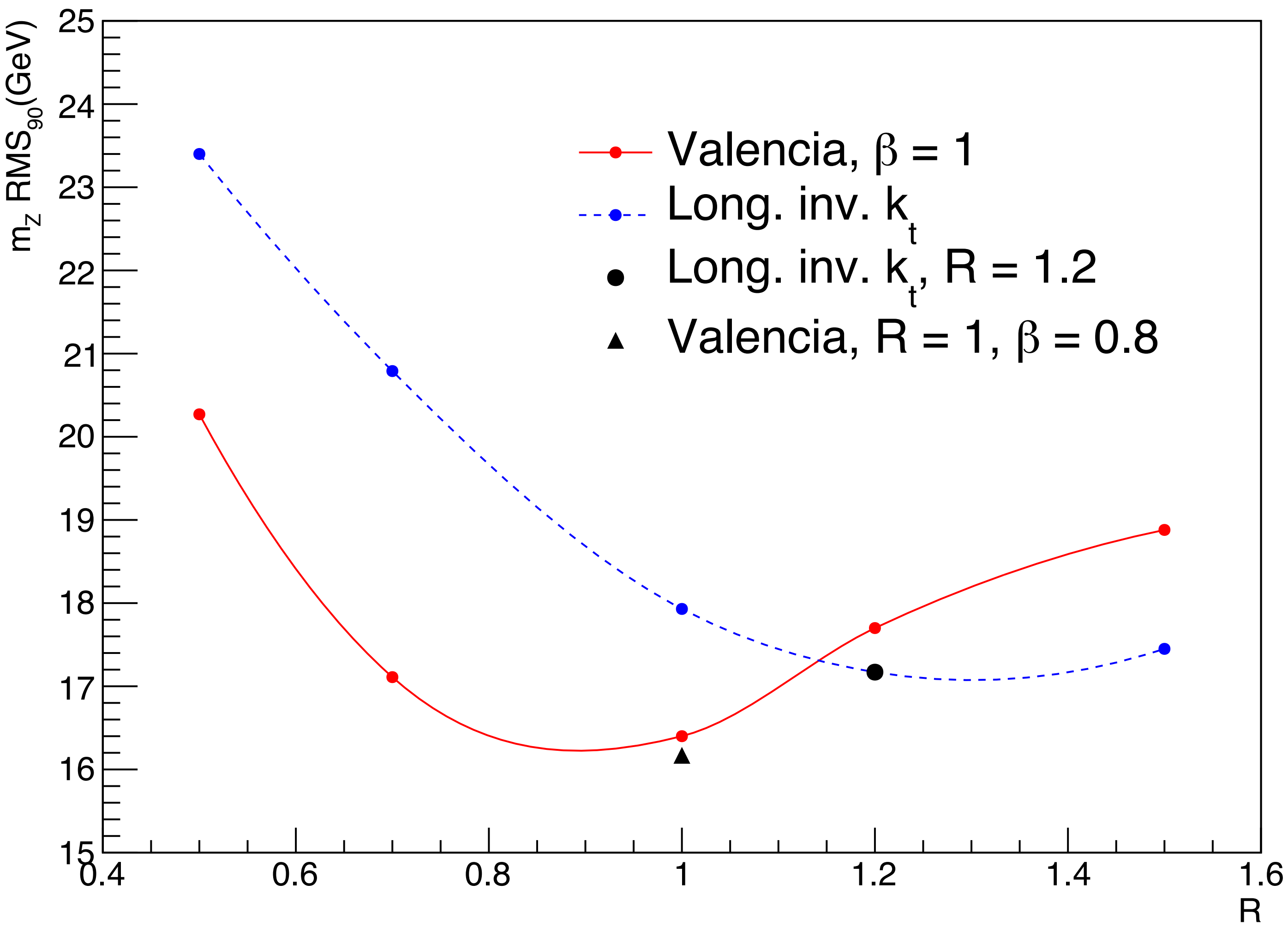
# BACK-UP SLIDES

# Algorithm parameters optimisation: R scan

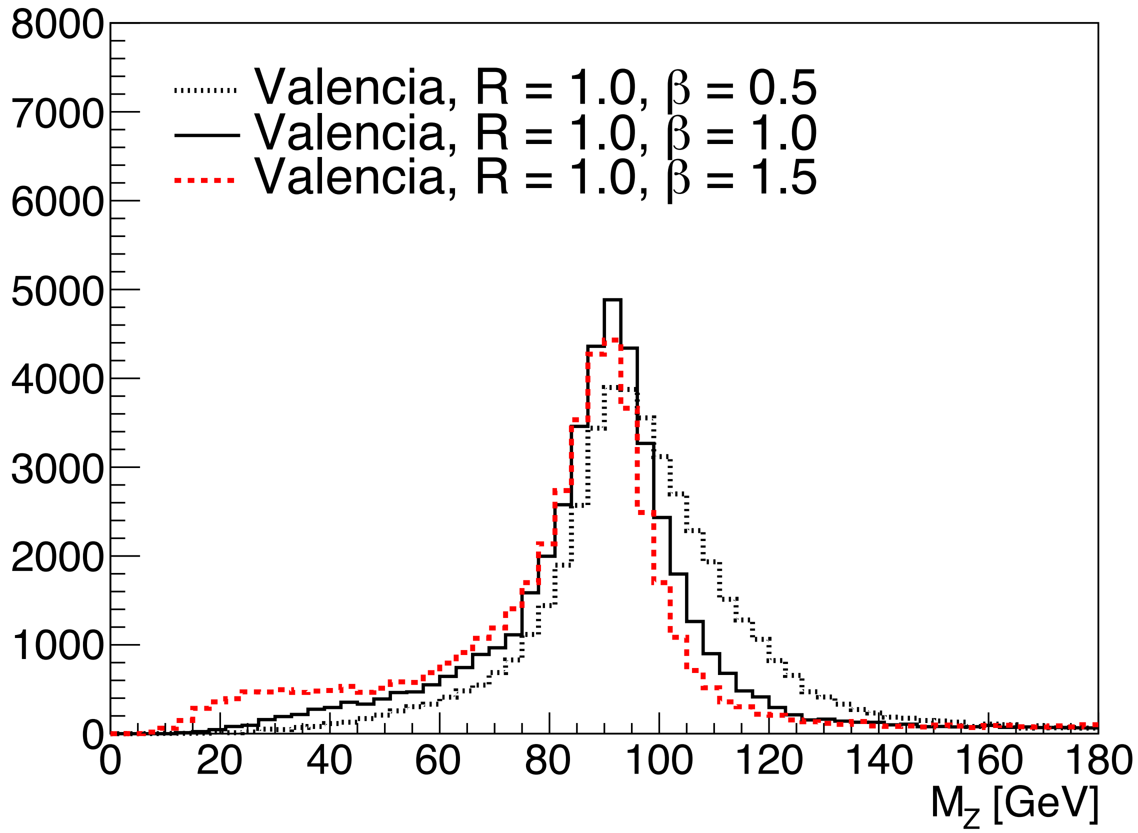


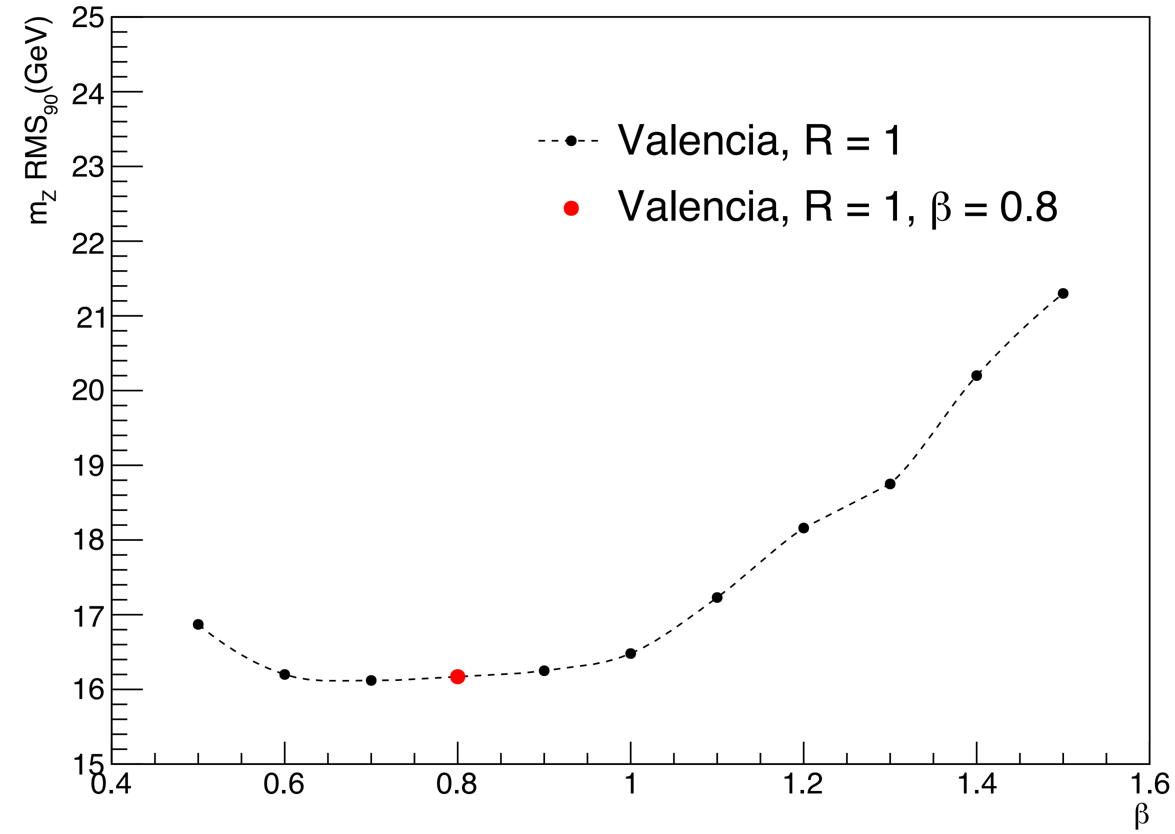




The choice of parameters corresponds to the optimal setting determined in a scan over a broad range of parameters.



# Algorithm parameters optimisation: β scan



# IR-safety

**A key requirement to jet algorithms**

From Salam & Soyez, JHEP 0705 (2007)
An infinitely soft particle cannot lead to a new (hard) [jet] being found...
...it makes no sense for the structure of multi-hundred GeV jets to change radically just because hadronisation, the underlying event or pileup threw a 1 GeV particle in between them.

The sequential recombination structure underlying the Valencia algorithm is generally thought to be intrinsically safe

A large number of **standard IR-safety tests** were performed on the FastJet plugin. **All succeeded.**