Jet reconstruction at future lepton colliders

Marcel Vos
IFIC (UVEG/CSIC) Valencia

Parton radiation and fragmentation
CERN, November 2016

Based on work with Nacho García (IFIC), Philipp Roloff, Rosa Simoniello (CERN)
Acknowledging help from Gavin Salam (CERN) and Jesse Thaler (MIT)

PLB750 (2015) 95-99, arXiv:1404.4294
ArXiv:1607.05039

Jets in e^+e^- colliders

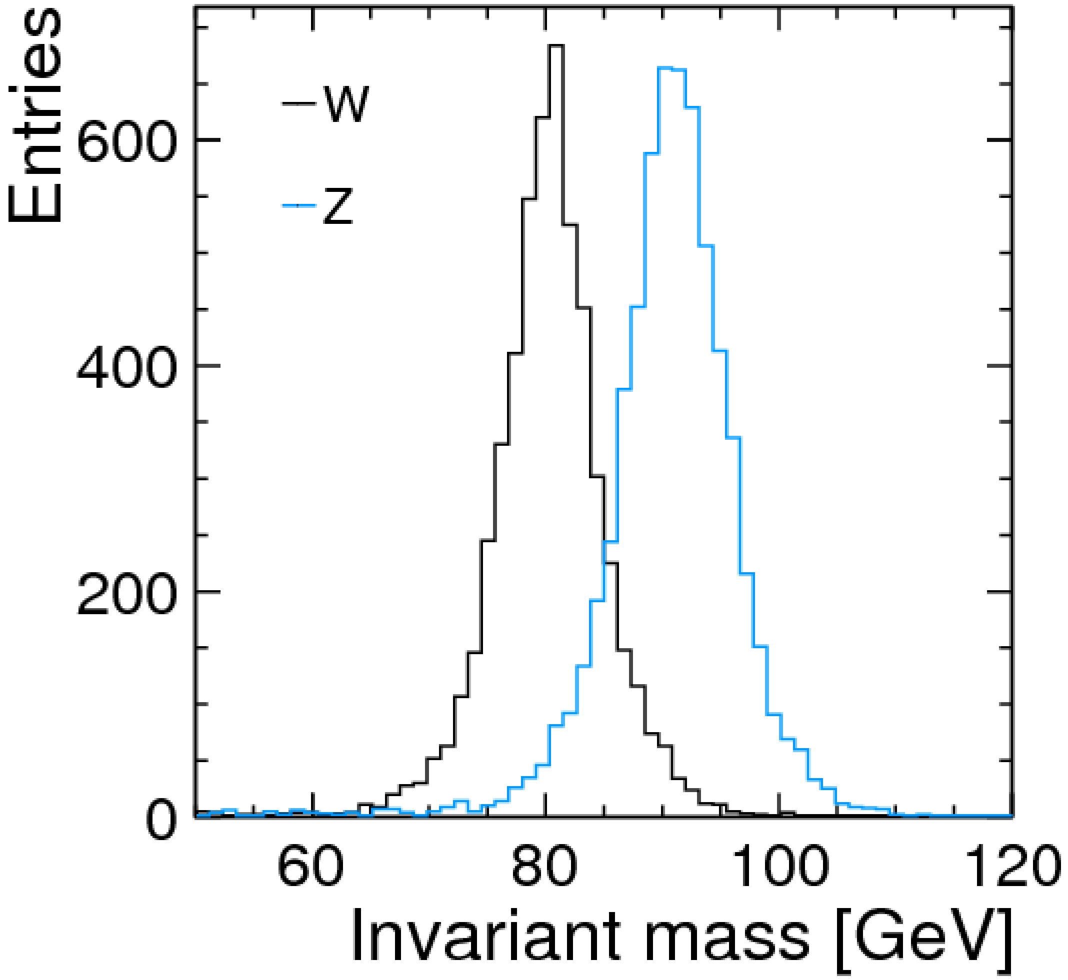
Hadronic final states are important for the precision e^+e^- programme

- *Higgs production, arXiv:1509.02853*
- *Top quark production, arXiv:1604.0122*
- *Gauge boson pair production*

Lepton colliders are for PS + fragmentation what DIS is for PDFs

- Controlled and calculable initial state

- Reference samples of q/g/b/W/Z/H/t jets "without the junk" (MPI, UE, pile-up)

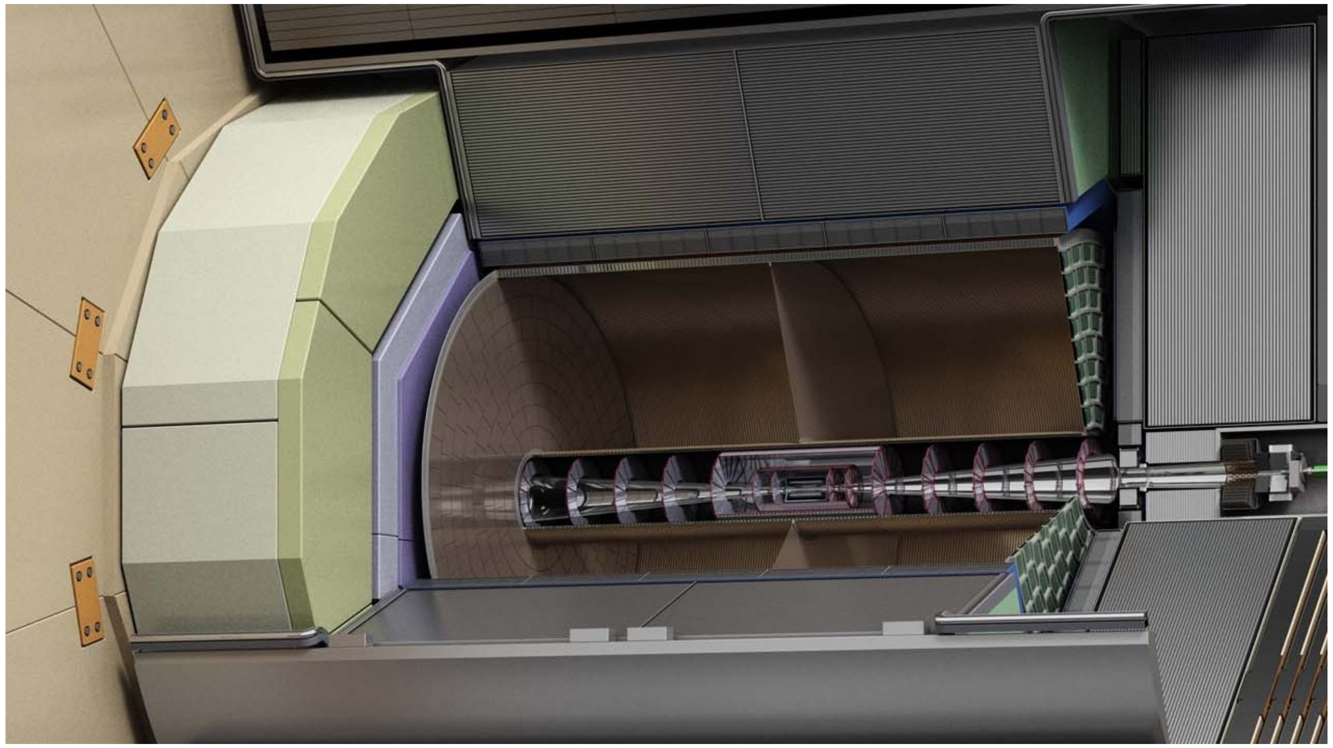


Jet reconstruction is important

Performance goal: distinguish hadronic W and Z decays

Marcel Vos (marcel.vos@ific.uv.es)

Detectors



LC detector concepts optimized for particle flow

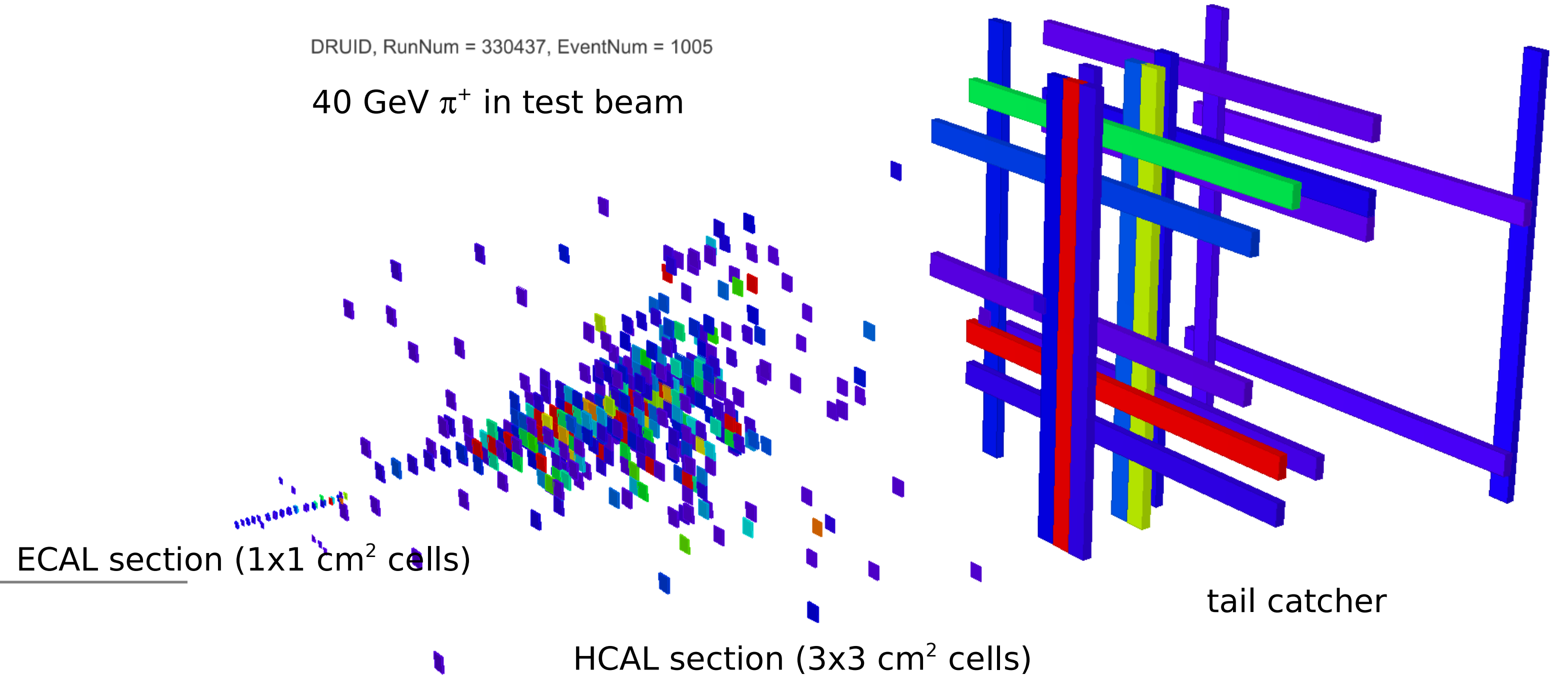
- highly granlular calorimeter*
- 4-5 Tesla solenoid
- state-of-the art low-mass tracking system
- precision vertexing

For details:
CLIC CDR, arXiv:1202.5940
ILC TDR, arXiv:1306.6329

Detailed Geant4 model and adequate reconstruction software allow for realistic estimates of performance. This includes beam energy spectra and “pile-up” from background processes.

Not (entirely) science fiction:

The **CALICE** R&D collaboration has constructed and tested ultra-granular SiW EM calorimeters and a 1 m^3 prototype ScW hadronic calorimeter

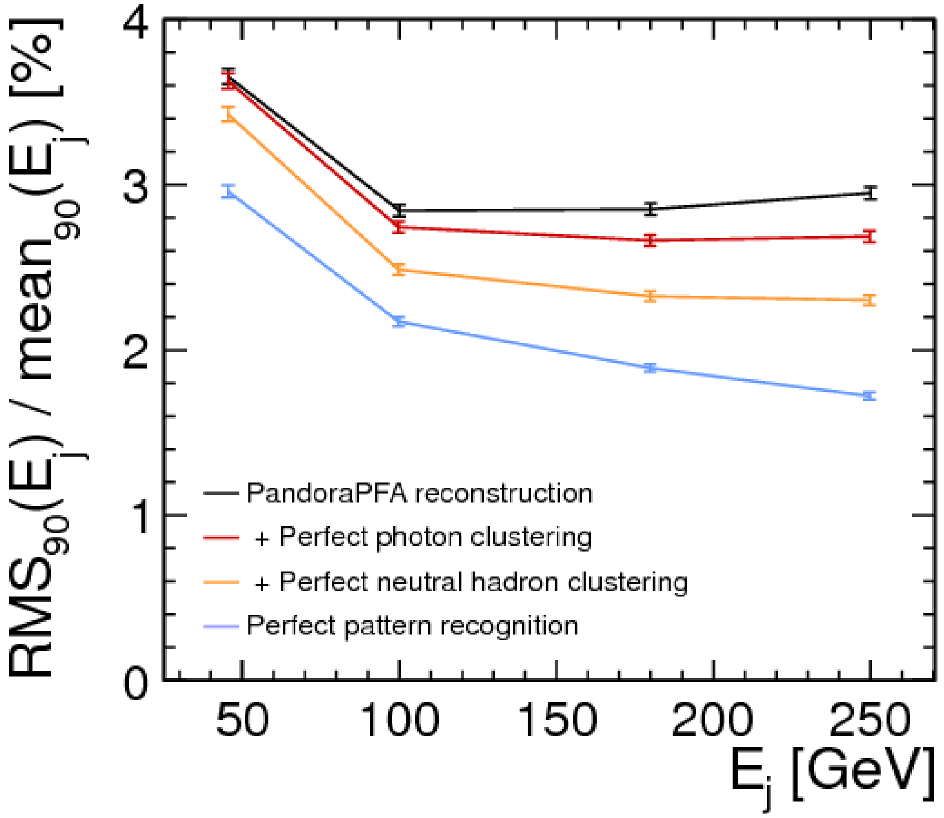


Particle Flow

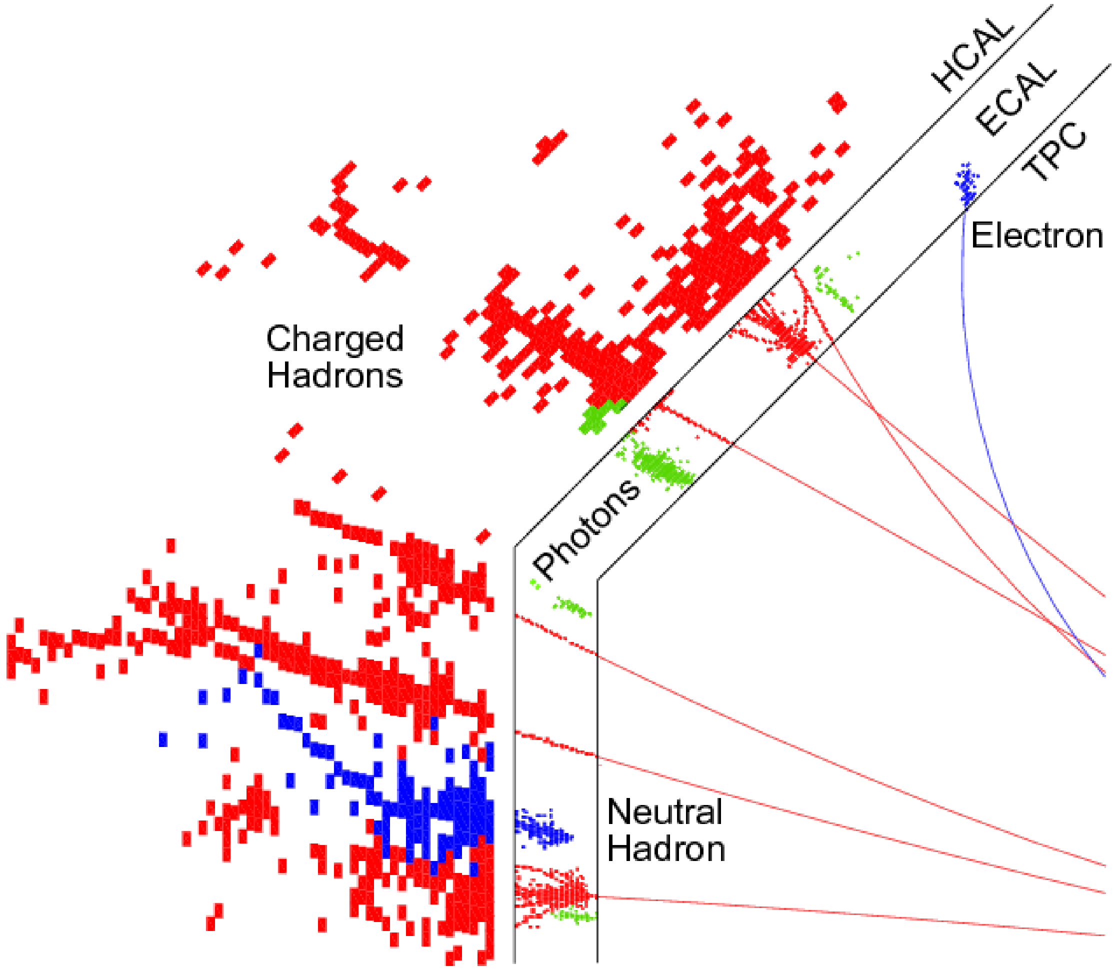
Particle flow offers “ultimate” detector performance

In theory able to achieve $\Delta E/E = 19\%/\sqrt{E}$ *(theoretical limit for perfect track-cluster association)*

In practice limited to by confusion term for high energy jets: $\Delta E/E \sim 3\%$



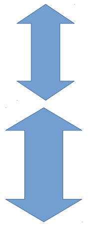
Di-jet events, energy resolution for “jets” inferred from total visible energy



Jet reconstruction must match excellent single-particle reponse

Jet reconstruction

In complex final states jet clustering may limit the performance



Detector level (Particle Flow objects)

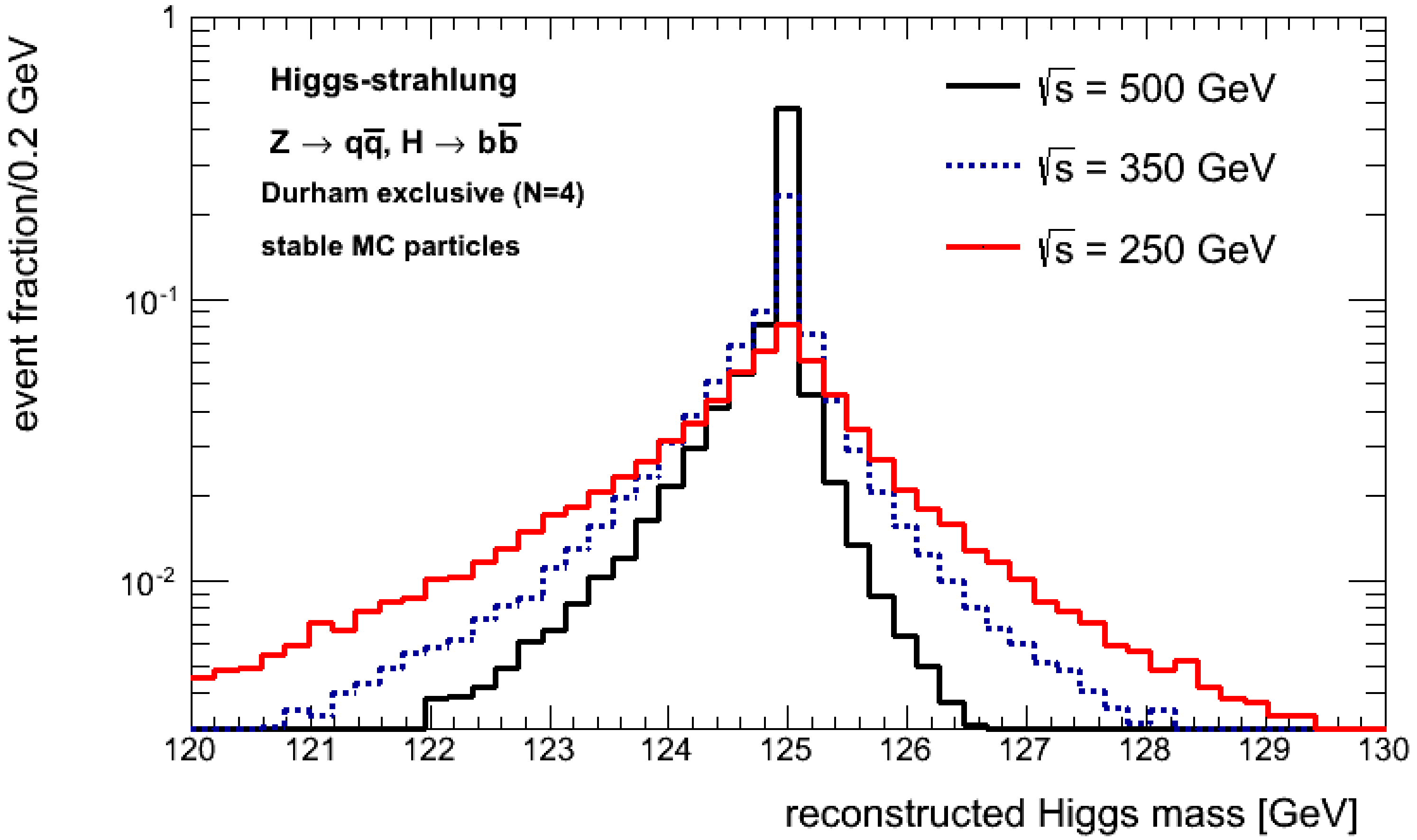
Particle level (stable MC particles)

Parton level (W, Z, Higgs or top mass)

Detector limitations...

Llimitations
of jet algorithms...

Particle-level jet reconstruction: non-zero resolution due to “confusion” in clustering



Exclusive clustering

Exclusive clustering assumes that N hardest splits correspond to N final state partons

At high-energy colliders phase space for hard emissions opens up

An example: top quark pair production

K_t distance between top quark decay products governed by t and W mass

→ $\sqrt{d_{qq}}$ roughly constant

Distance of QCD final state radiation dqg increases with available phase space

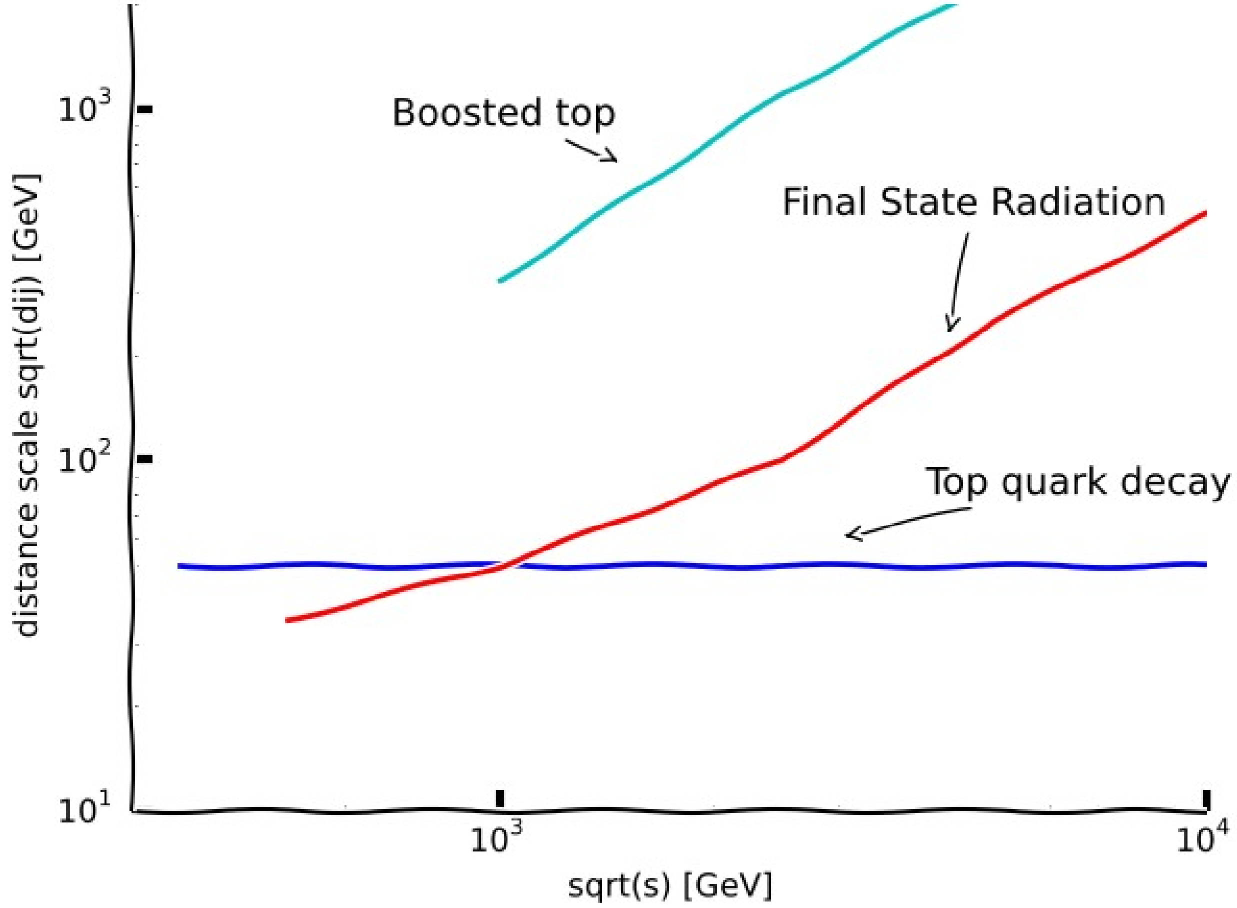
→ $\sqrt{d_{qg}}$ proportional to $\alpha_s\sqrt{s}$

N=6 exclusive clustering at $\sqrt{s} \sim 2$ m

N=2 exclusive clustering at $\sqrt{s} \sim 1$ Te

→ $\sqrt{d_{tg}}$ is approx. $\sqrt{s}/2$

Radiation from top quarks threatens N=6 exclusive clustering at high energy, but N=2 clustering takes over right in time



"Evolution of average distance scales"

Inclusive vs. exclusive

Not-so-easy: ZHH production

Hierarchy in energies increases with √s

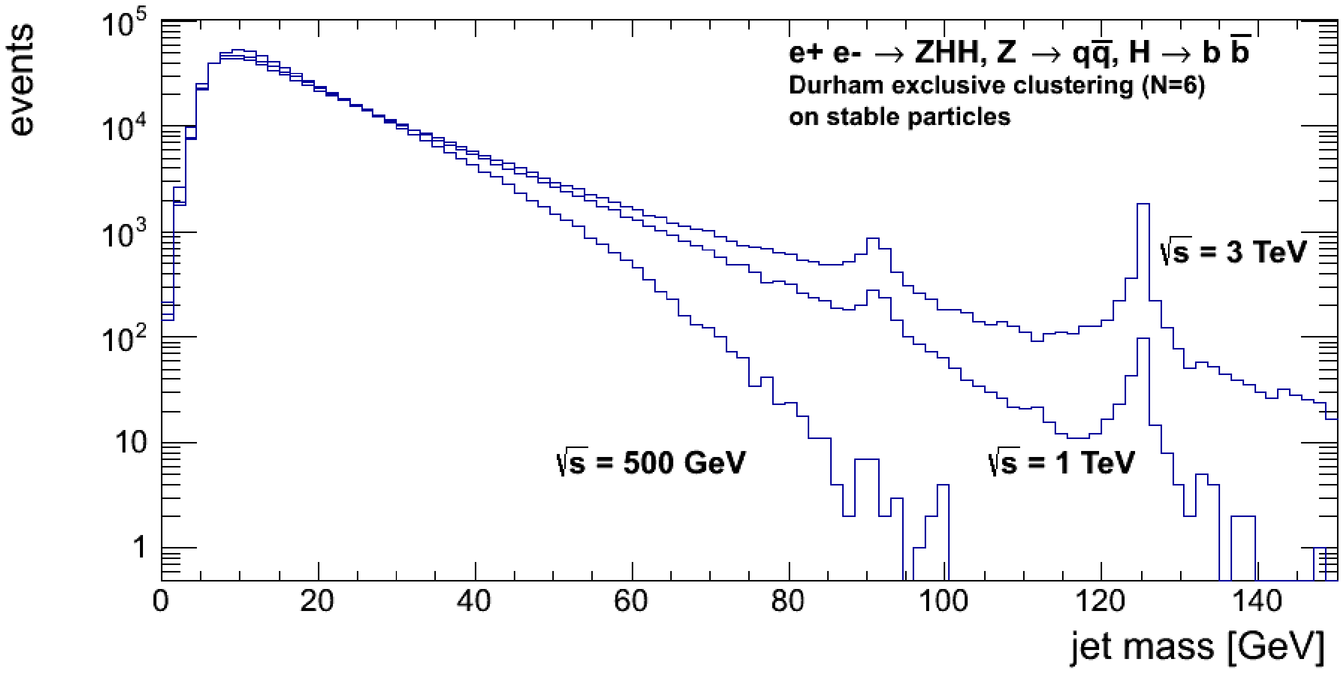
When qg split is harder than Z → q$\bar{q}$ or H → b$\bar{b}$:

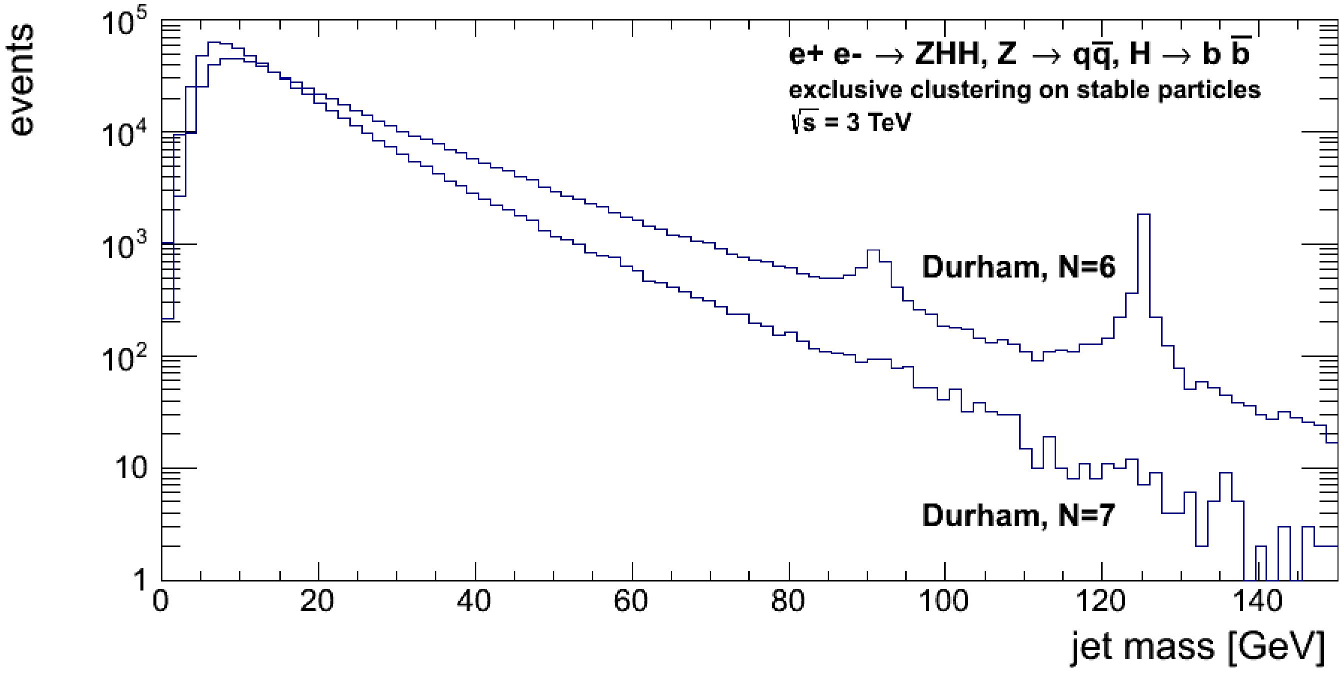
Z → q$\bar{q}$ or H → b$\bar{b}$ erroneously merged

Failure mode is indeed observed in a small fraction (few %) of events at high energy (1-3 TeV)

Allow a seventh split? (N=7)

Detailed analysis carried out by Junping Tian (KEK)





Marcel Vos (marcel.vos@ific.uv.es)

Lepton collider backgrounds

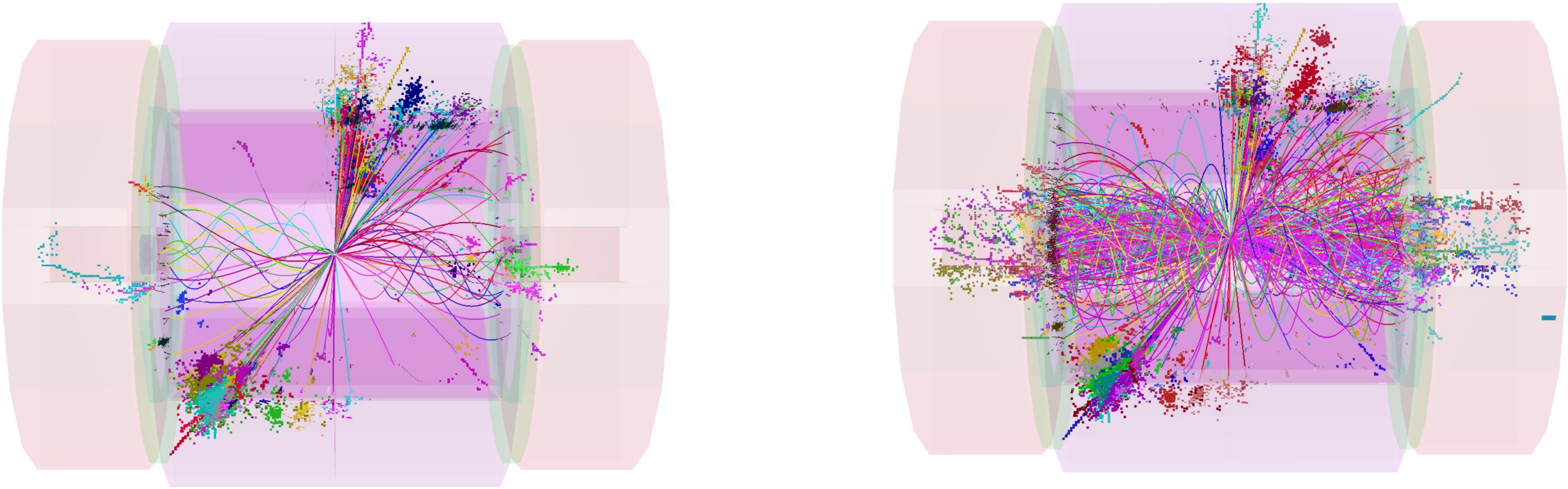
Lepton colliders offer a relatively clean environment (compared to the LHC), but not quite to the level of LEP or SLC. We cannot ignore background:

- Incoherent pair production

  very soft: relevant for vertex detector and forward systems
- $\gamma\gamma \rightarrow$ hadrons production

  particles reach central detectors and affect jet reconstruction
- Synchrotron radiation (?)

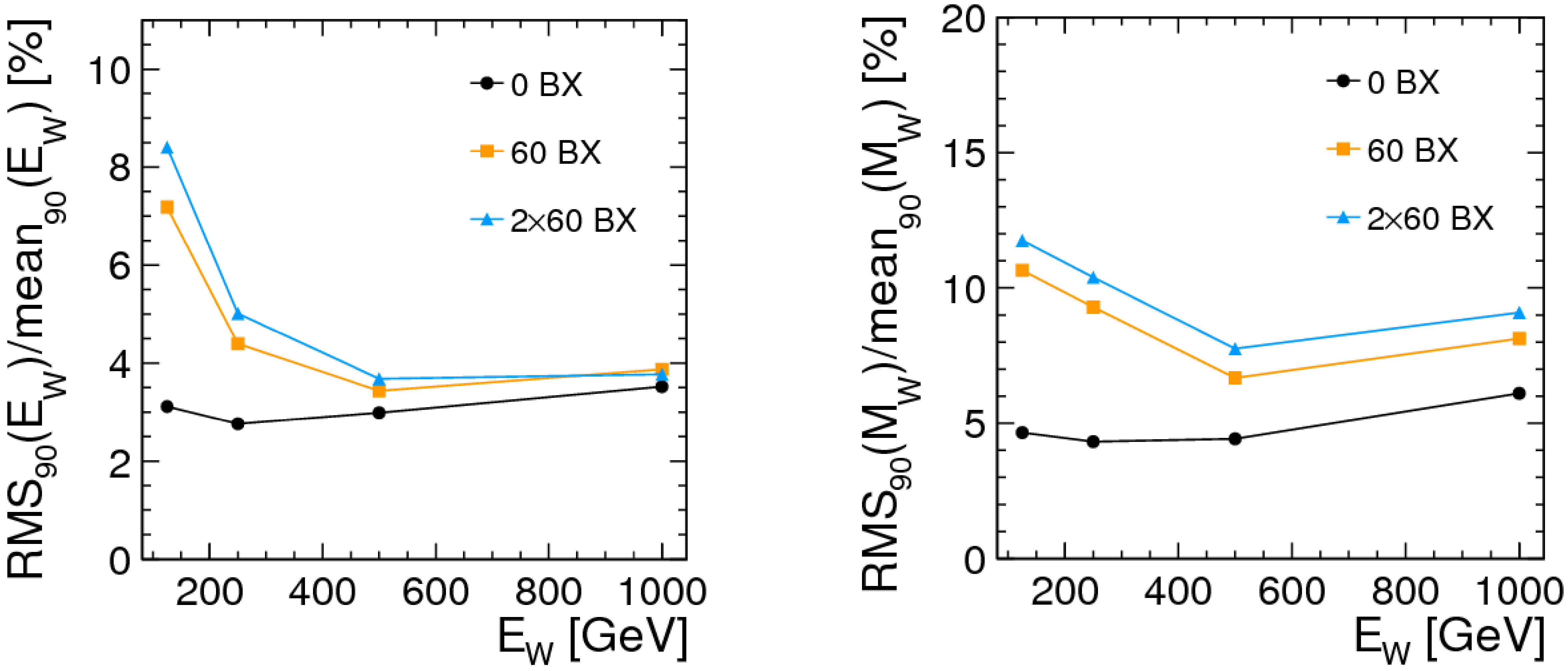
Important at high energy and luminosity/BX



Example: a CLIC bunch train worth of $\gamma\gamma \rightarrow$ hadrons superposed on a physics event. If all CLIC3TeV detector systems integrate over 10 ns (=20BX), background deposits 1.2 TeV of energy in the calorimeter systems.

Impact of background

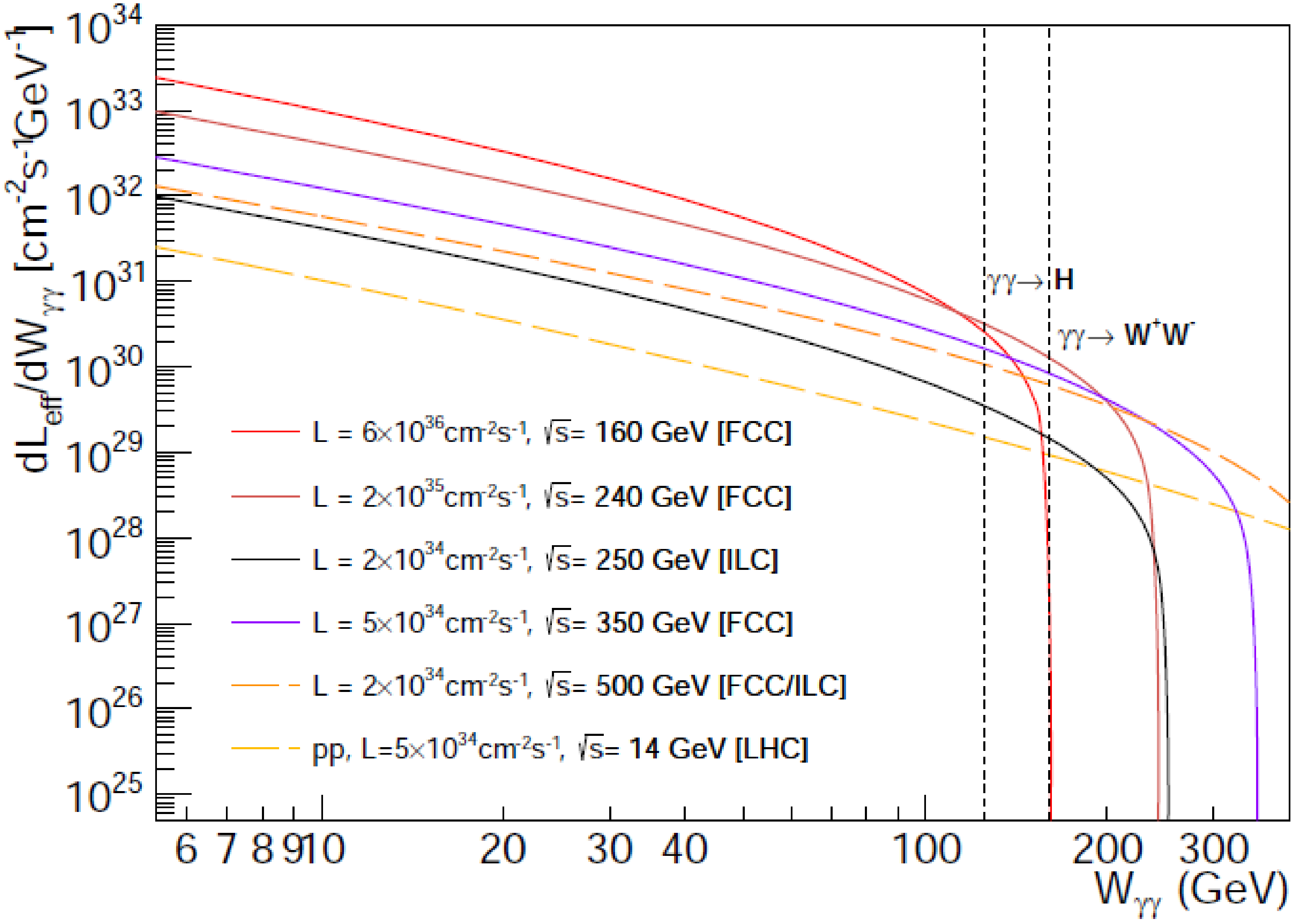
$e^+e^- \rightarrow W^+W^- \rightarrow l\nu\, q\bar{q}$ events at CLIC with W energies of 100, 250, 500 and 1000 GeV
Overlay 60 (120) BX worth of $\gamma\gamma \rightarrow$ hadrons, select in-time reconstructed particles, remove lepton
Reconstruct long. inv. k_t jets exclusively (N=2, R=0.7)



Energy resolution at high energy is not too badly affected,
mass resolution suffers everywhere

[CLIC CDR, Marshall, Münnich & Thomson, arXiv:1209.4039], non-negligible even for ILC physics [many studies, arXiv:1307.8102]

$\gamma\gamma \to$ hadrons



Use CLIC case as a stress test for jet reconstruction;
If it works there, it's good for ILC too.
FCCee has much smaller $\gamma\gamma \to$ hadrons background still.

Summary, so far

Jet reconstruction at future lepton collider projects is more challenging than for previous generation of e^+e^- colliders:

- Better detectors force jet reconstruction algorithms to step up
- More complex multi-jet final states with hierarchy of scales
- Non-negligible background

Revisit jet reconstruction

Not discussed further: XCone jets, arXiv:1508.01516

Jet reconstruction algorithms

Lepton colliders:

Hadron colliders:

Distance based on E, angle

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

Good old Durham (n=1, e^+e^- k_t) and
Cambrige/Aachen (n=0)

Generalized lepton collider aglorithms:

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

Introduce beam jets → jet size R

The Valencia algorithm

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

Hadron colliders:
longitudinally invariant
algorithms based on p_T, ΔR

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

N=1,	longitudinally invariant kt
N=0,	Cambridge/Aachen
N=-1,	anti-kt

Valencia jets, PLB750 (2015) 95-99

Maintain distance base on energy and angle
Choose beam distance for robust performance

Jet algorithm space

VLC algorithm of arXiv:1607.05039

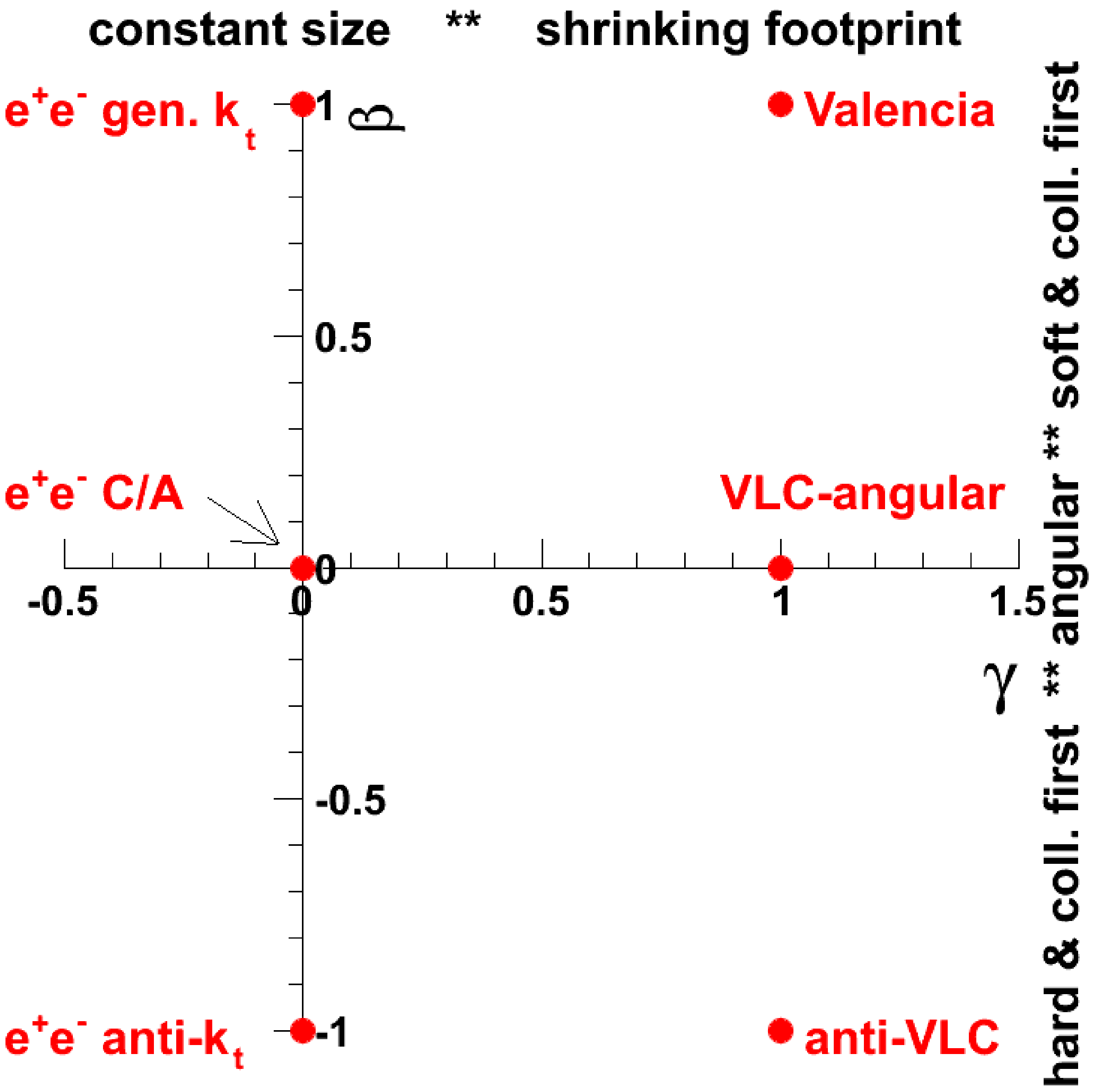
$$d_{ij} = 2\min(E_i^{2\beta}, E_j^{2\beta})(1-\cos\theta_{ij})/R^2,$$

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

Two parameters (real numbers) govern the clustering order (b) and robustness against background (g)

Recover generalized e+e- kt for g=0

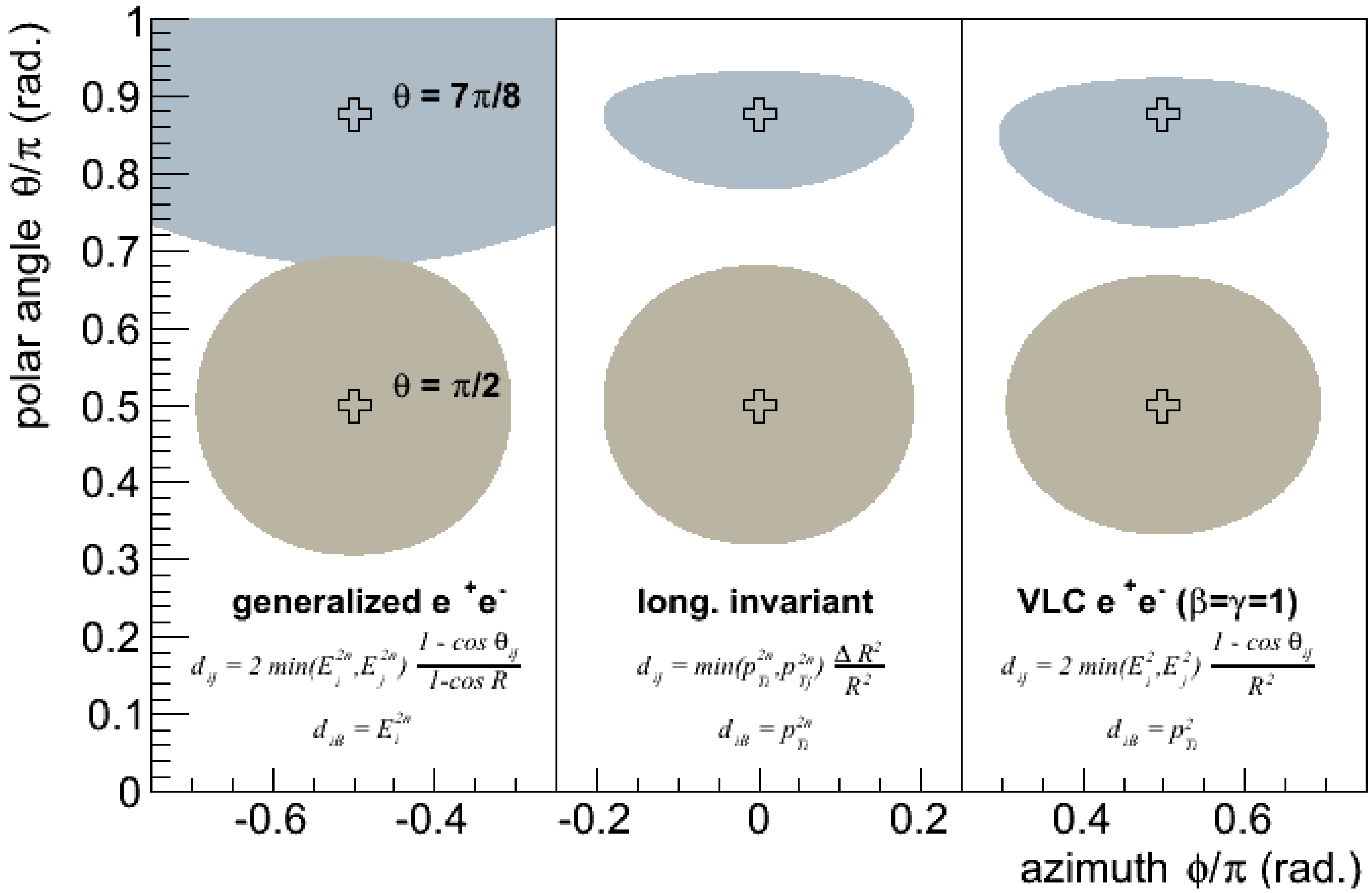
Mimic longitudinally invariant algorithms with g=1



Marcel Vos (marcel.vos@ific.uv.es)

ϕ extension blows up in this projection (cf. Antartica on a map)

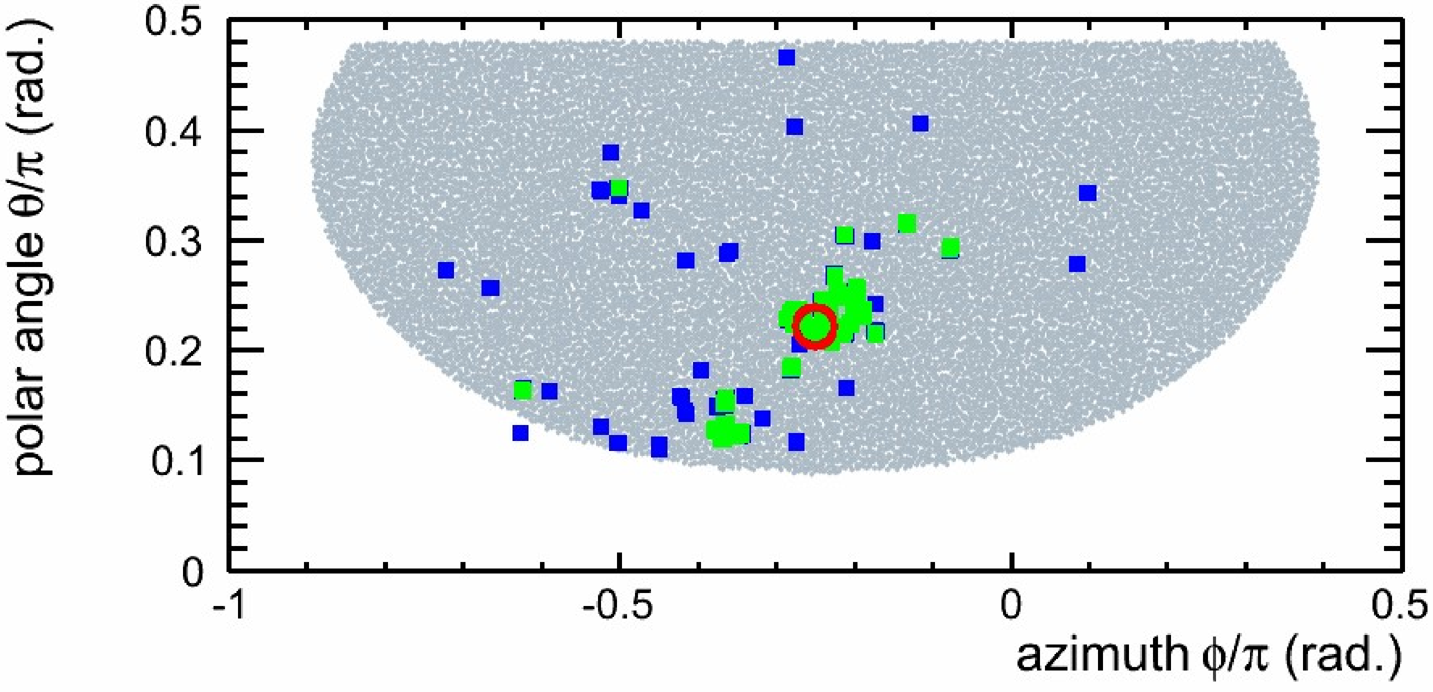
Jet sizes



Circular jets in (y,ϕ) space
asymmetric ellipses in (θ,ϕ)

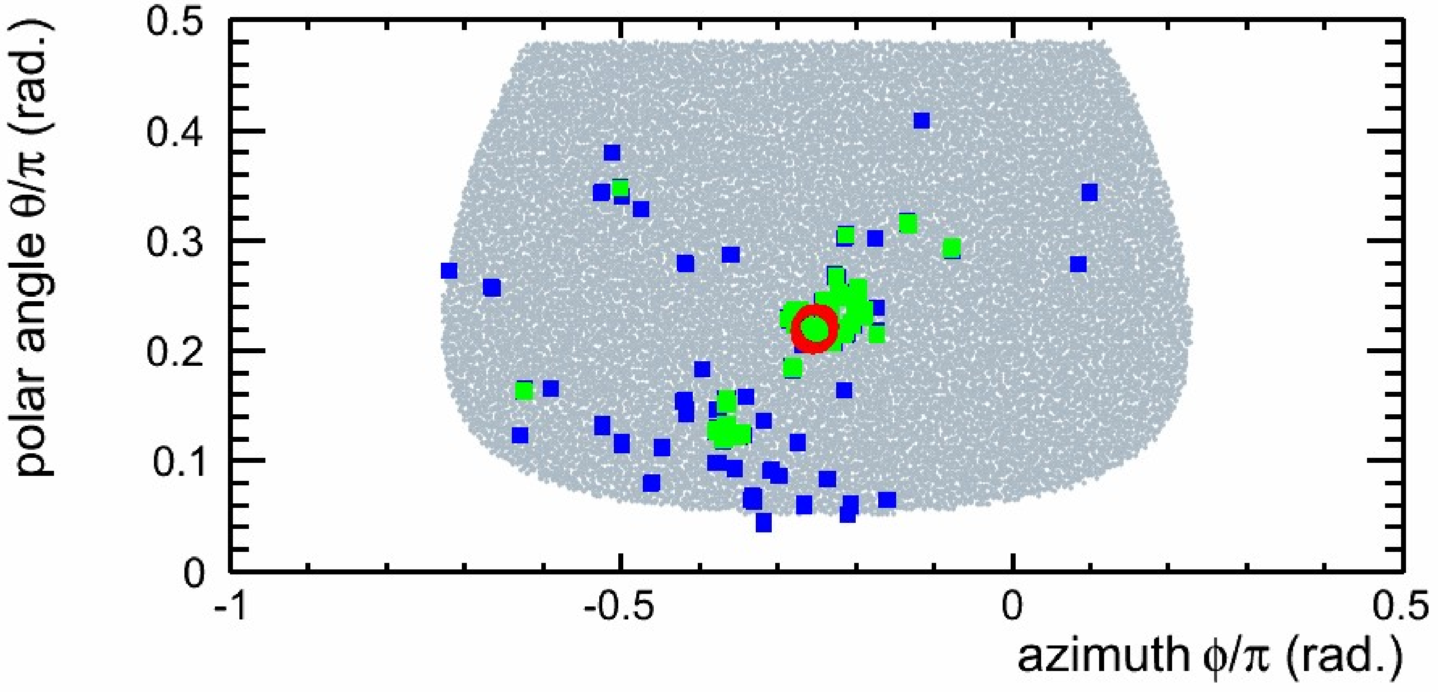
The footprint or area of jets depends on the jet algorithm
Three algorithms that yield a similar, circular area in the central detector produce very different jets in the forward region

Jet shapes and barycenters



VLC (with β=γ=1):
- ellipse in theta-phi plane
- barycenter close to edge

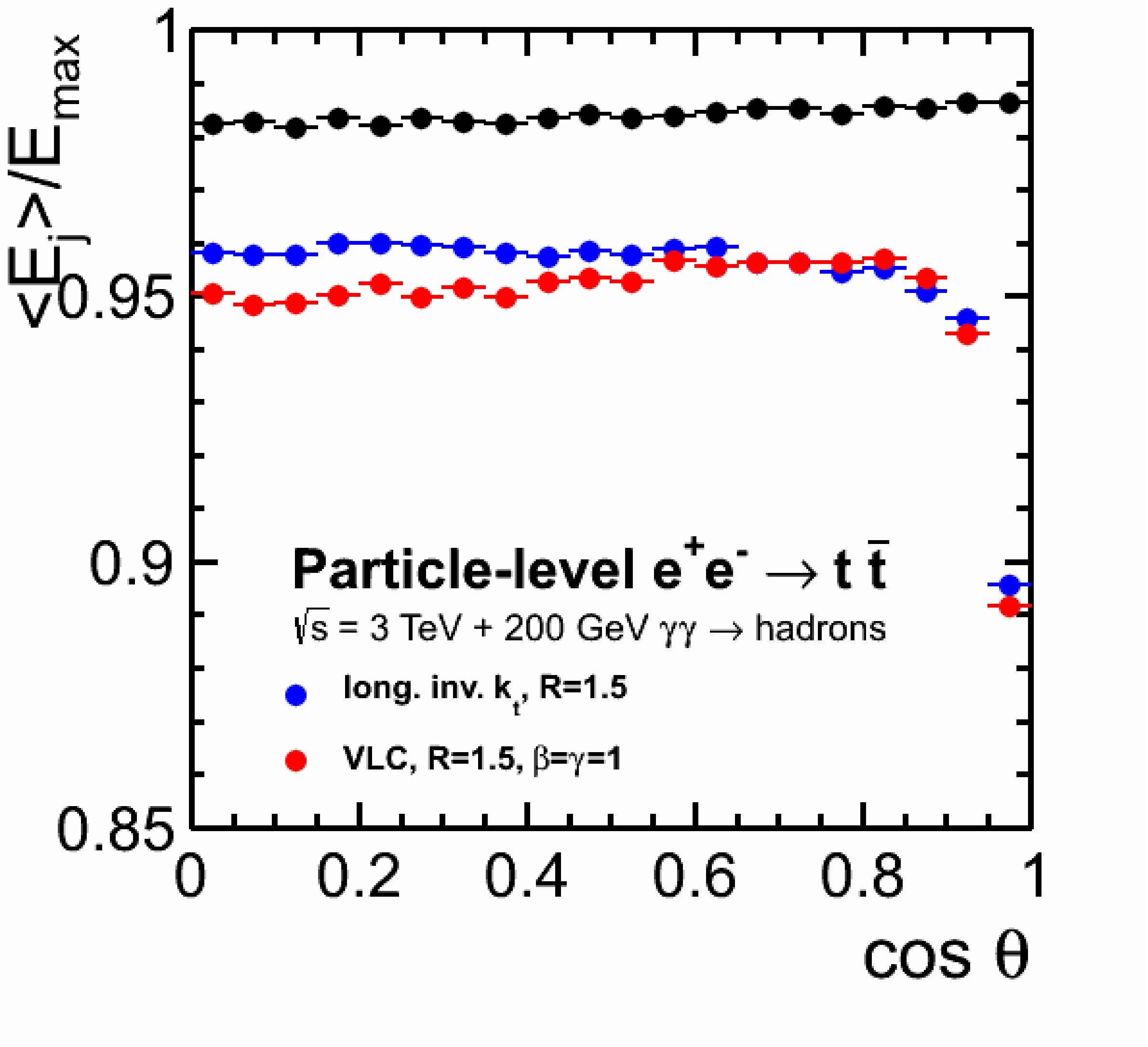
Long. Invariant k_t:
- pear shape in theta-phi plane
- extends further forward
- scoops

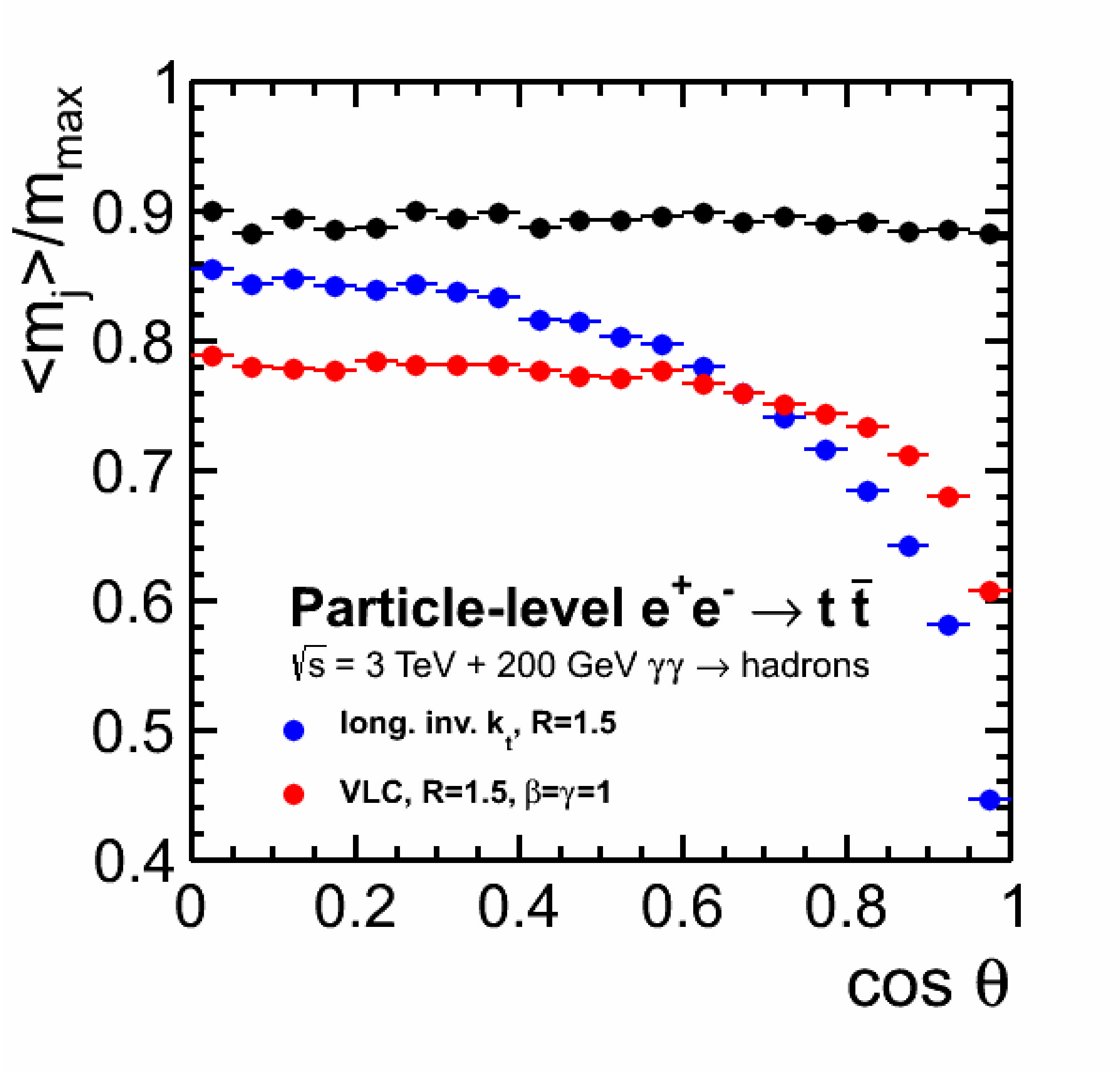


Marcel Vos (marcel.vos@ific.uv.es)

Impact on response

Toy study: e+e- → tt @ 3 TeV
MadGraph5.2.2 + Pythia8.180 + FastJet 3.0.6
Stable particles → no detector, no beam pipe, no ISR
Exclusive reconstruction into two jets
Reference (max.) energy/mass from Durham

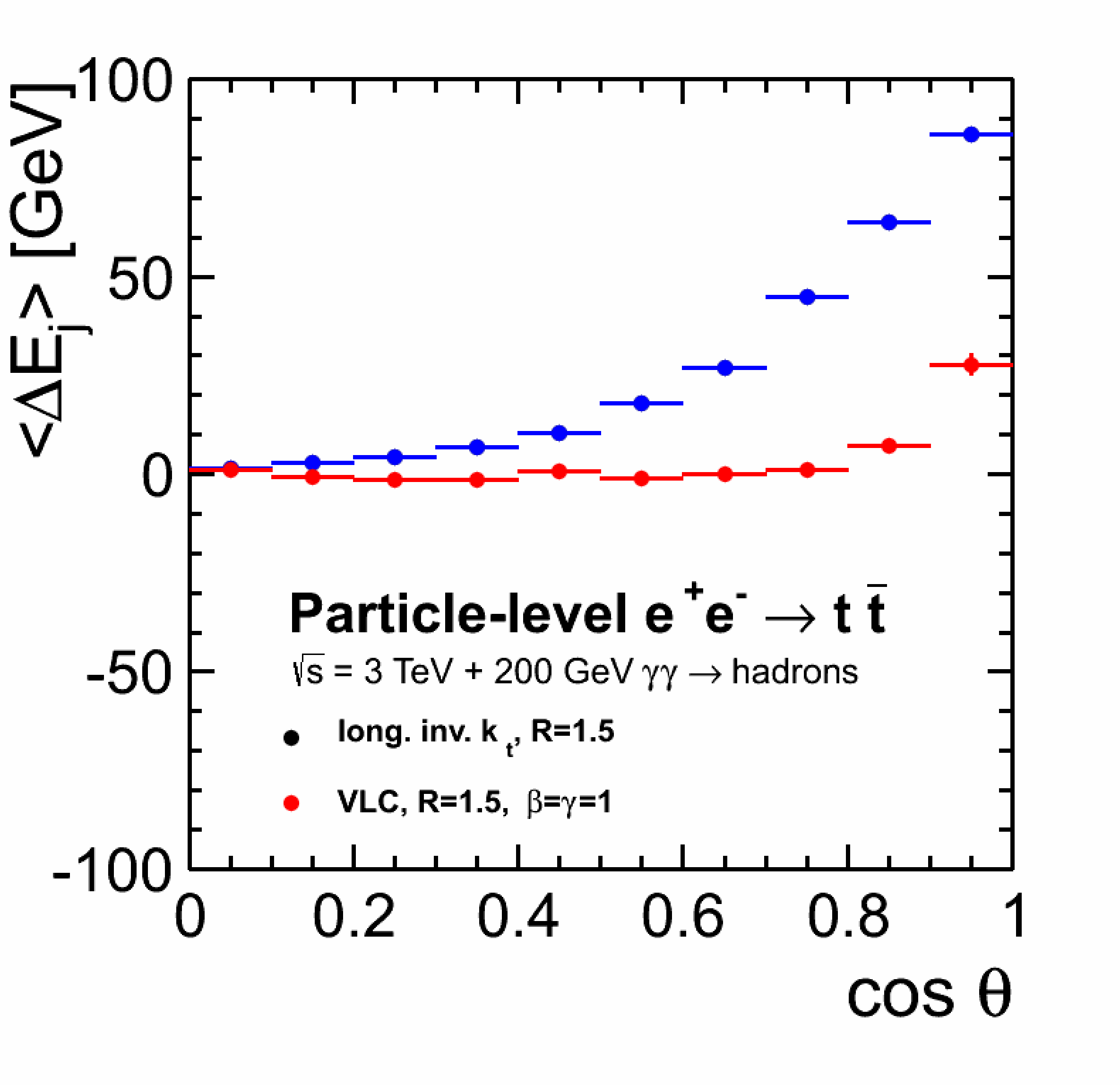


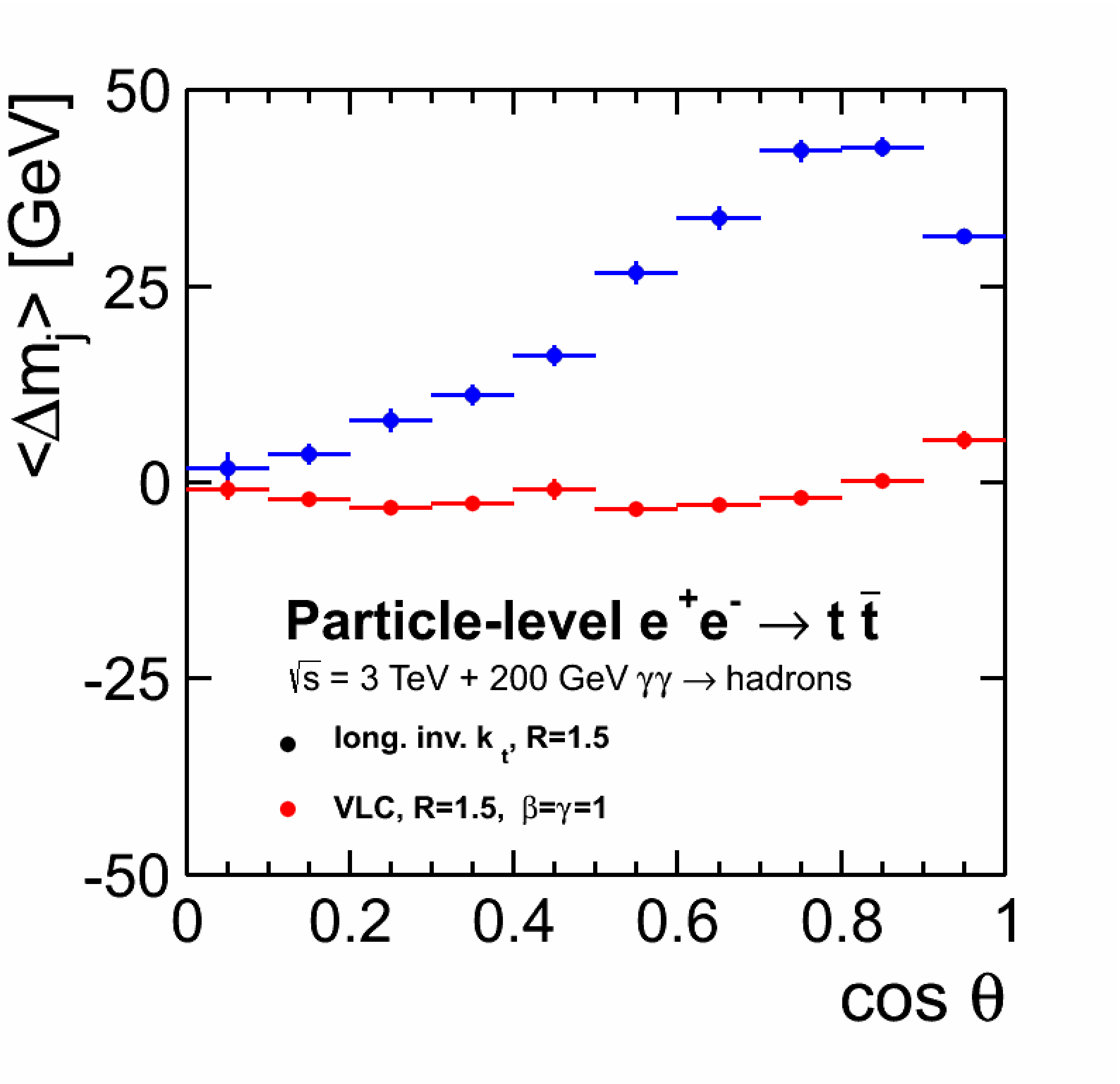


VLC mass response is much more stable

Background resilience

Now add background (random 200 GeV, forward-peaked) and register difference in jet energy and mass versus polar angle

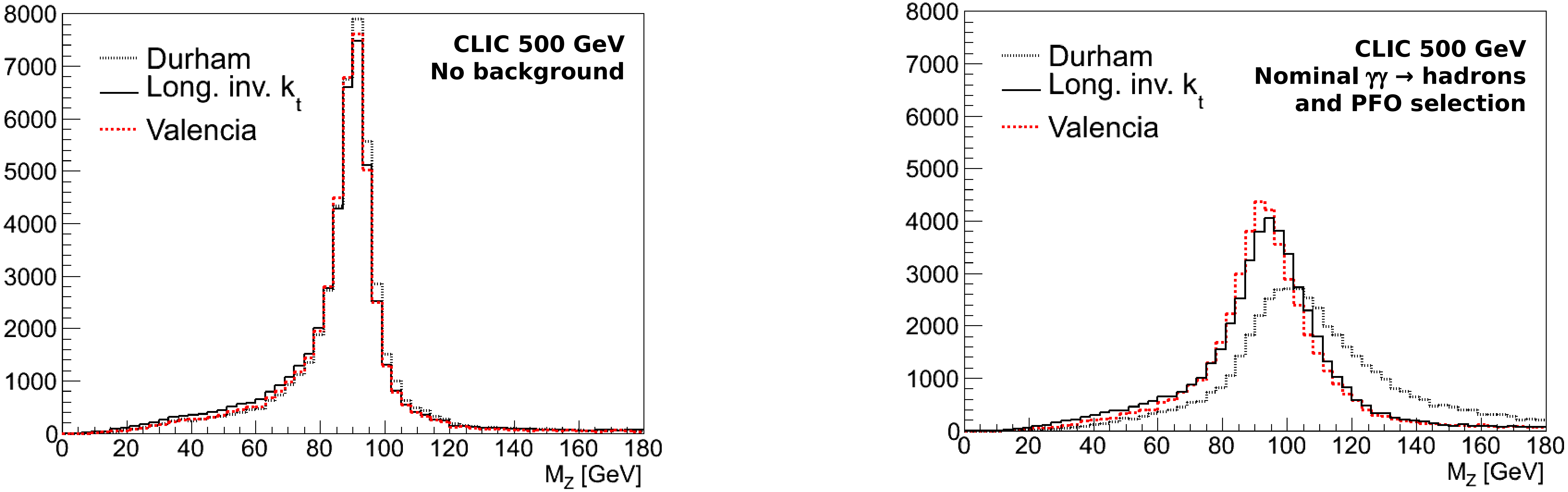




Longitudinally invariant kt much more affected, even in not-so-forward region
Sco

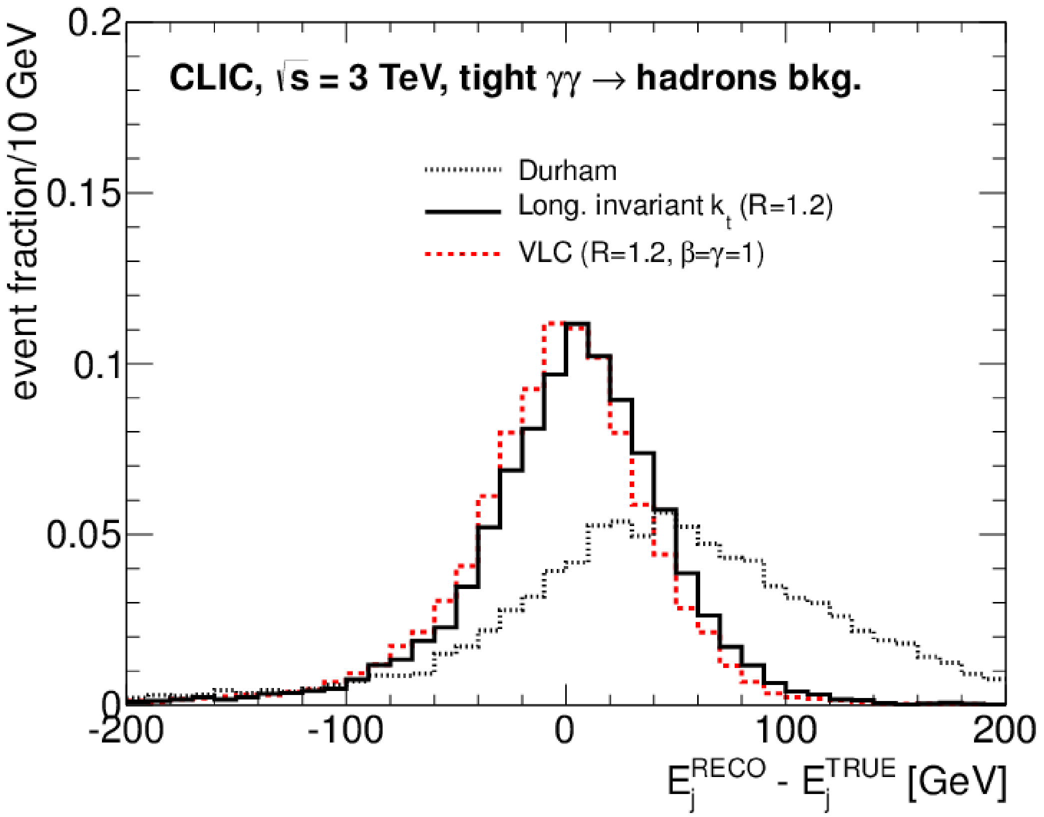
Benchmark: $t\bar{t}$ production

500 GeV: Jet energy reconstruction with nominal background much less degraded with algorithms with shrinking footprint (long. Invariant algorithms, Valencia) than e^+e^- algorithms



ZZ → $q\bar{q}q\bar{q}$ events at CLIC with $\sqrt{s}$ = 500 GeV. Remove forward, or off-shell Z
Reconstruct exactly 4 jets, with optimized R (=1.2 for longitudinally invariant k_t, 1.0 for Valencia)
Find best pairs and report di-jet mass for background-free and nominal background

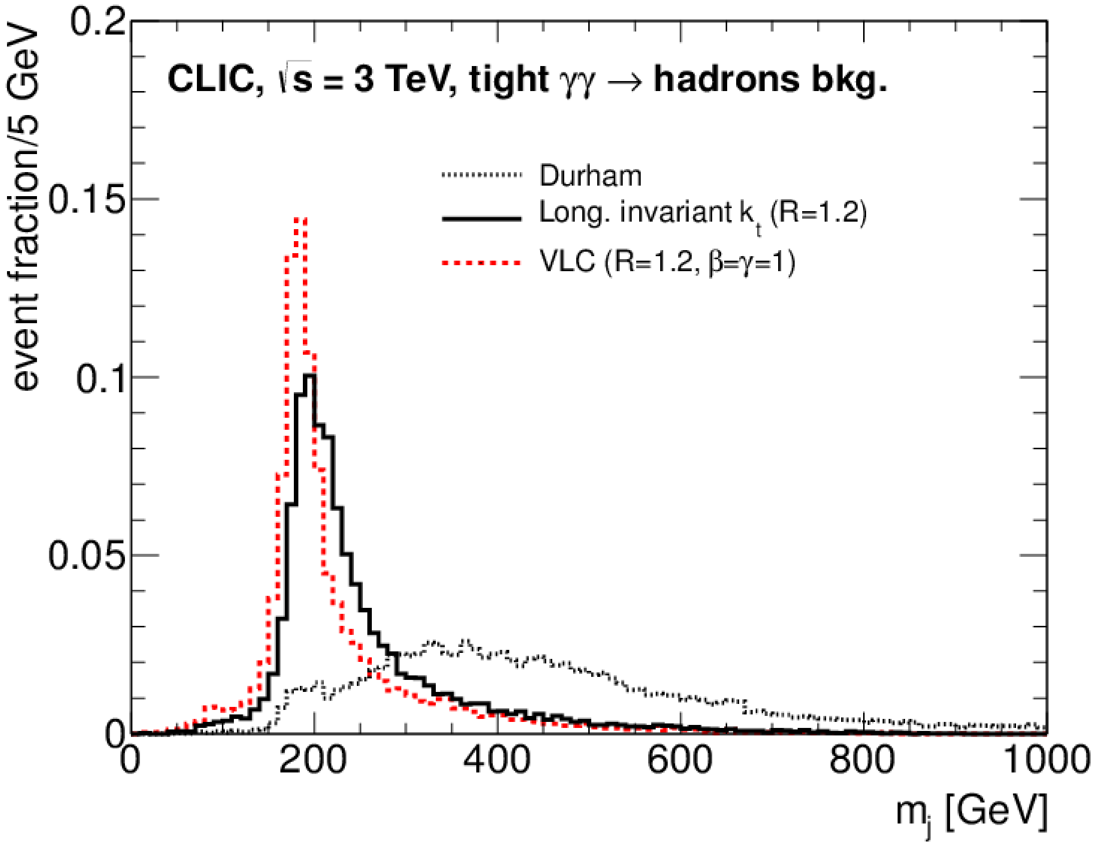
Benchmark: $t\bar{t}$ production



Classical lepton collider algorithms cannot cope with background

VLC algorithm is more robust than hadron collider algorithms

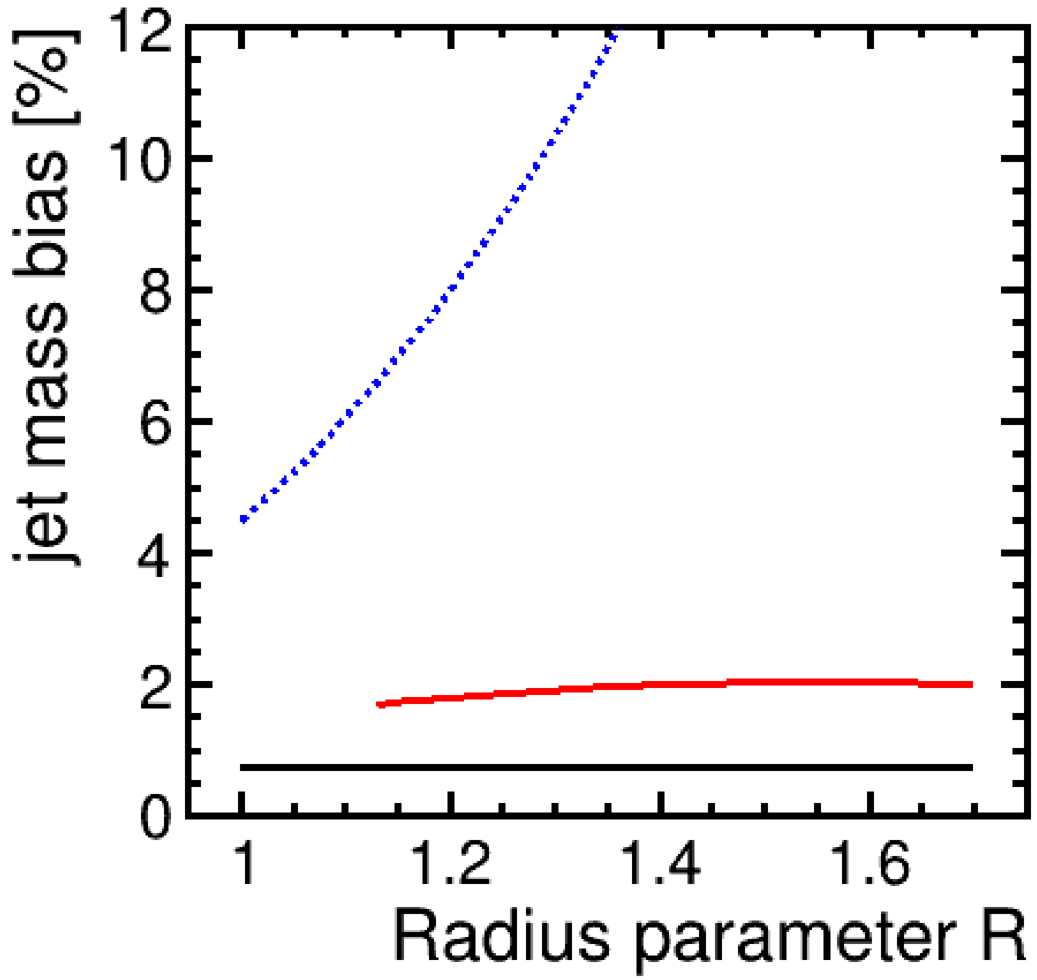
Some improvement in energy resolution and a strong effect on jet mass

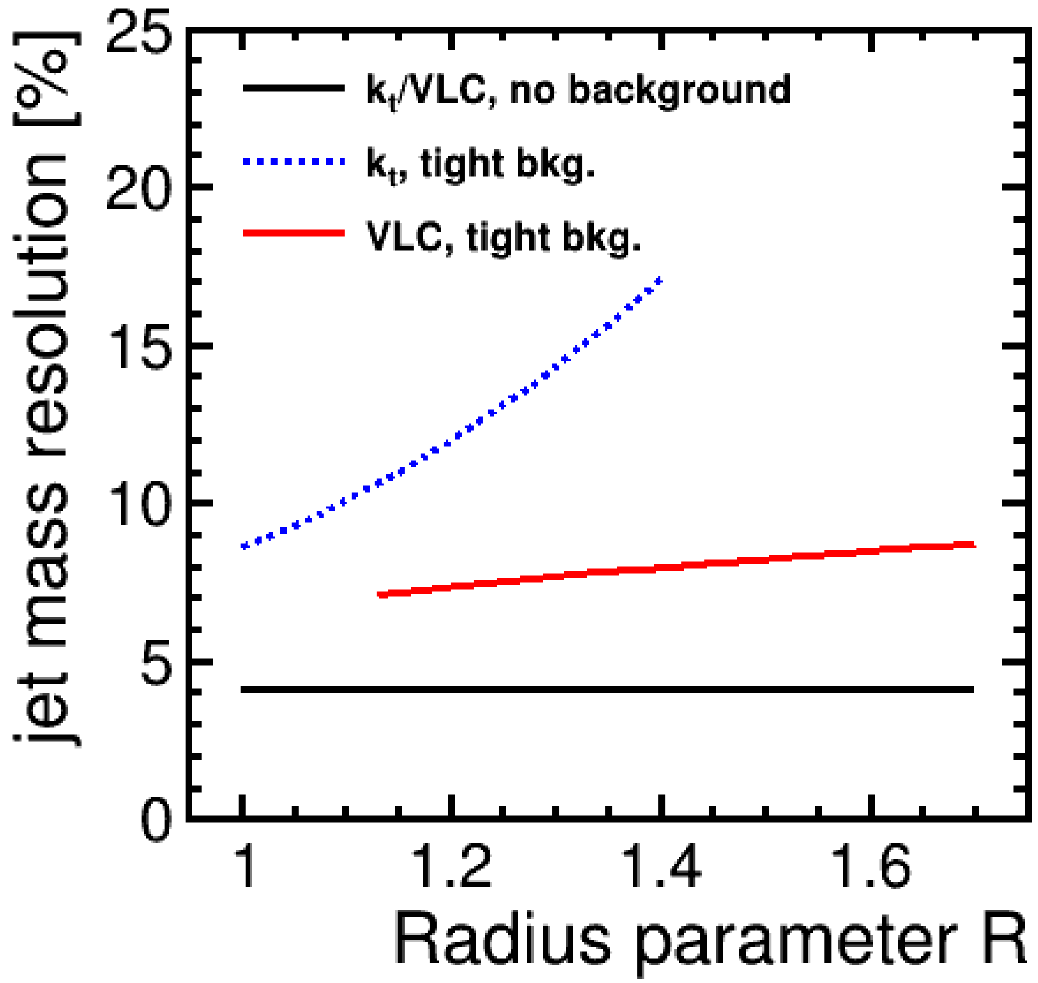


Marcel Vos (marcel.vos@ific.uv.es)

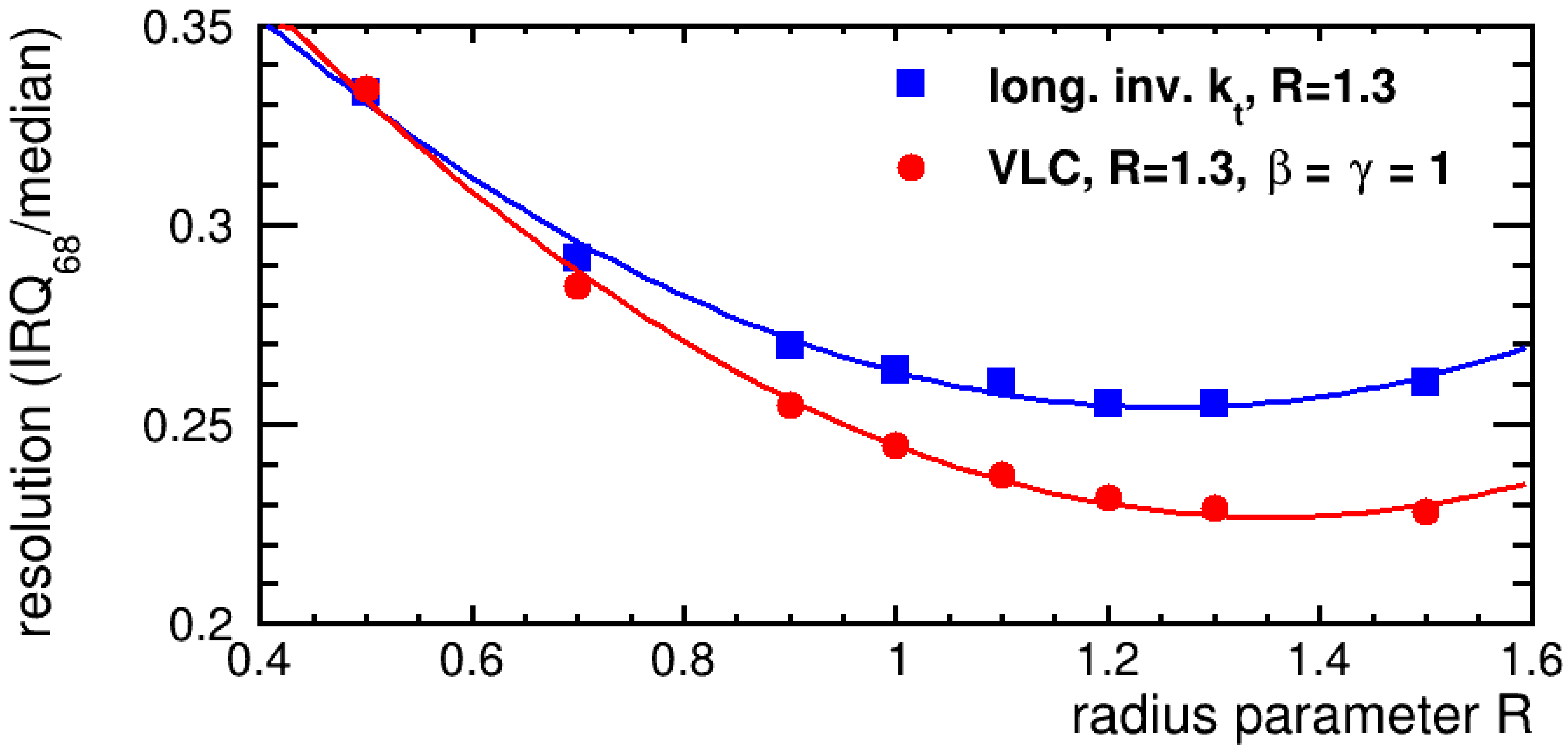
Benchmark: $t\bar{t}$ production

3 TeV: VLC outperforms longitudinally invariant algorithms





Benchmark: di-Higgs production



Di-Higgs production at 3 TeV: VLC outperforms longitudinally invariant algorithms

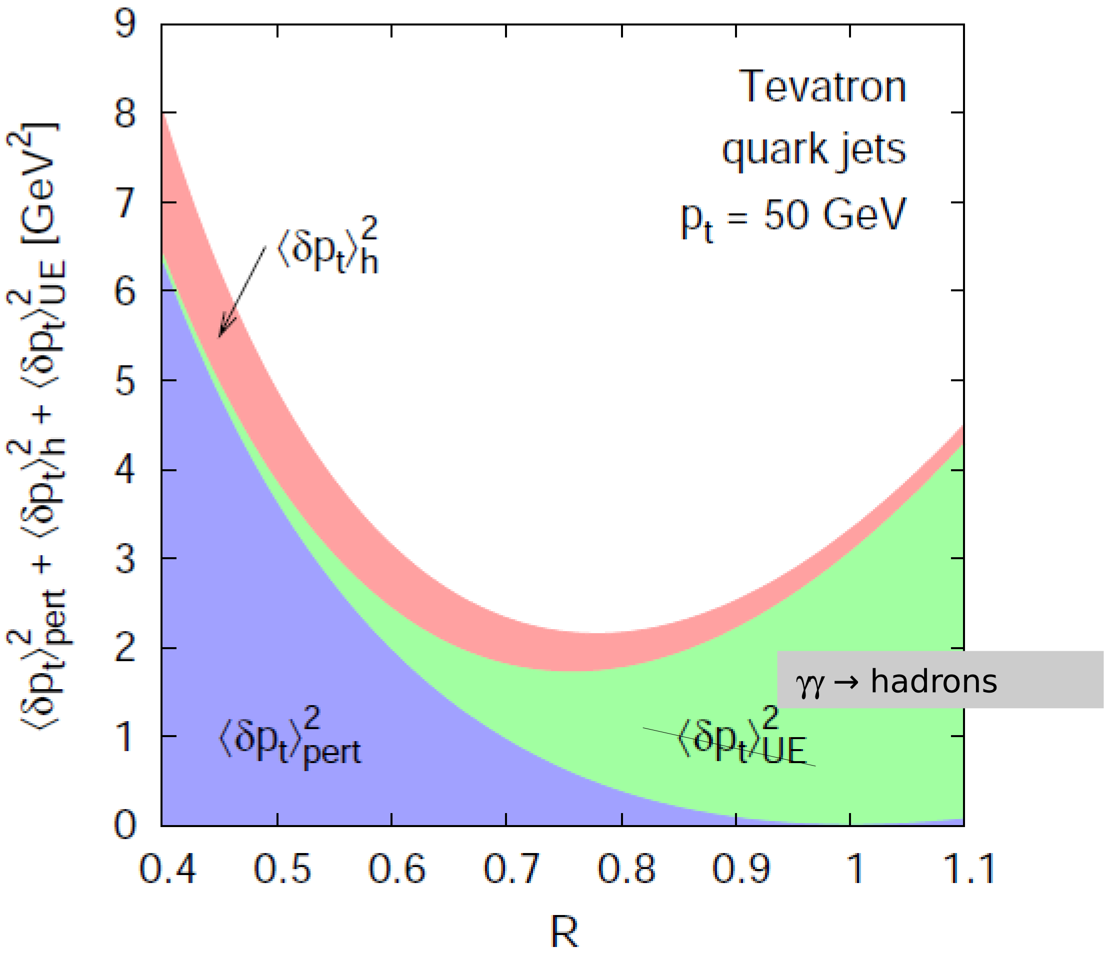
(non-) perturbative corrections

Uncertainties in jet response is important source of systematics

Jet area and footprint determine energy response:

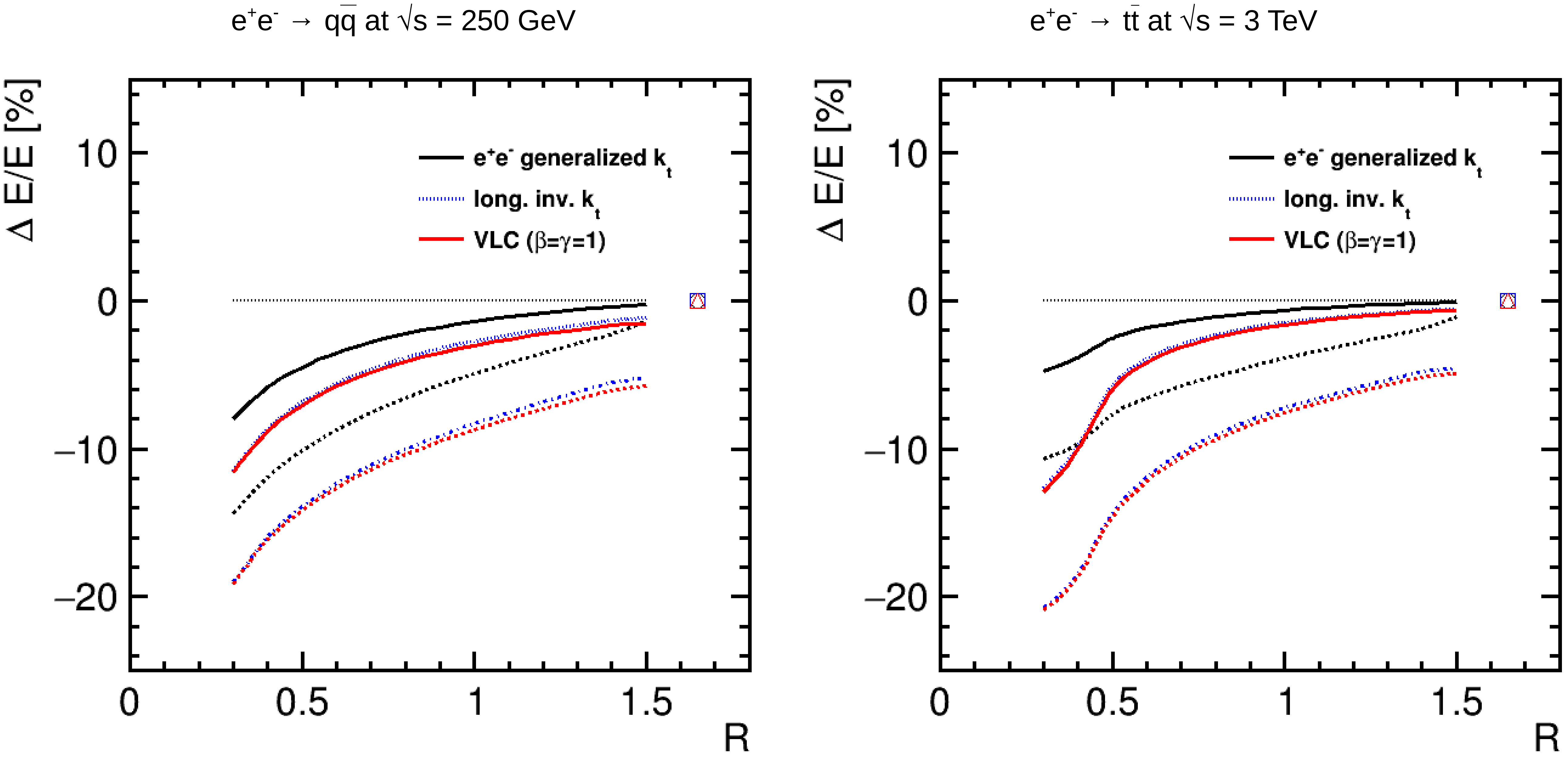
- (non-) perturbative corrections decrease with increasing R
- background contribution scales with R^2

Dasgupta, Magnea, Salam, JHEP0802 (2008) 055



Marcel Vos (marcel.vos@ific.uv.es)

Perturbative corrections

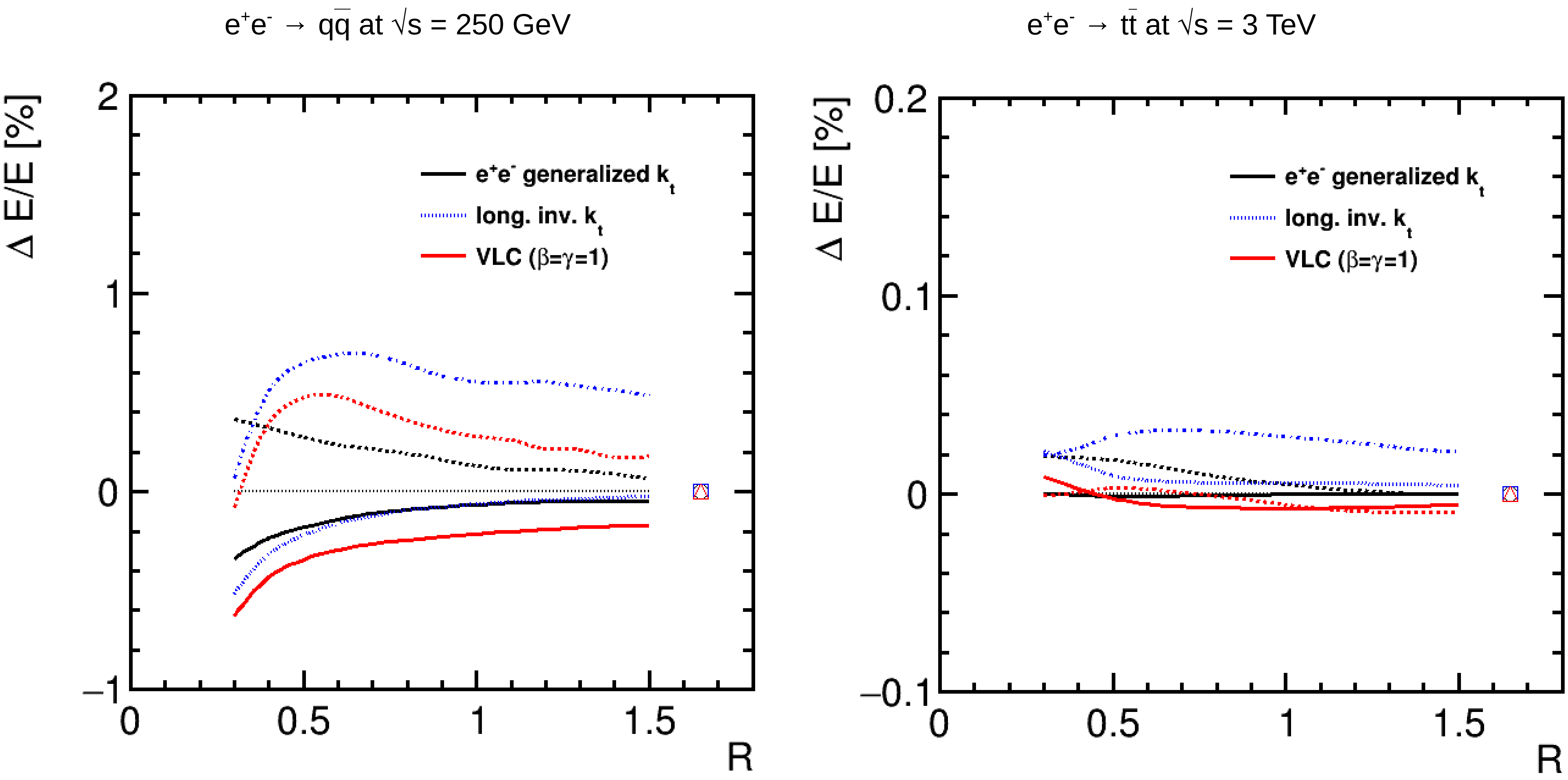


mean

median

Algorithm with largest footprint has the smallest correction
Skewed distributions: mean ≠ median
VLC and long. Invariant k_t virtually dentical

Non-perturbative corrections



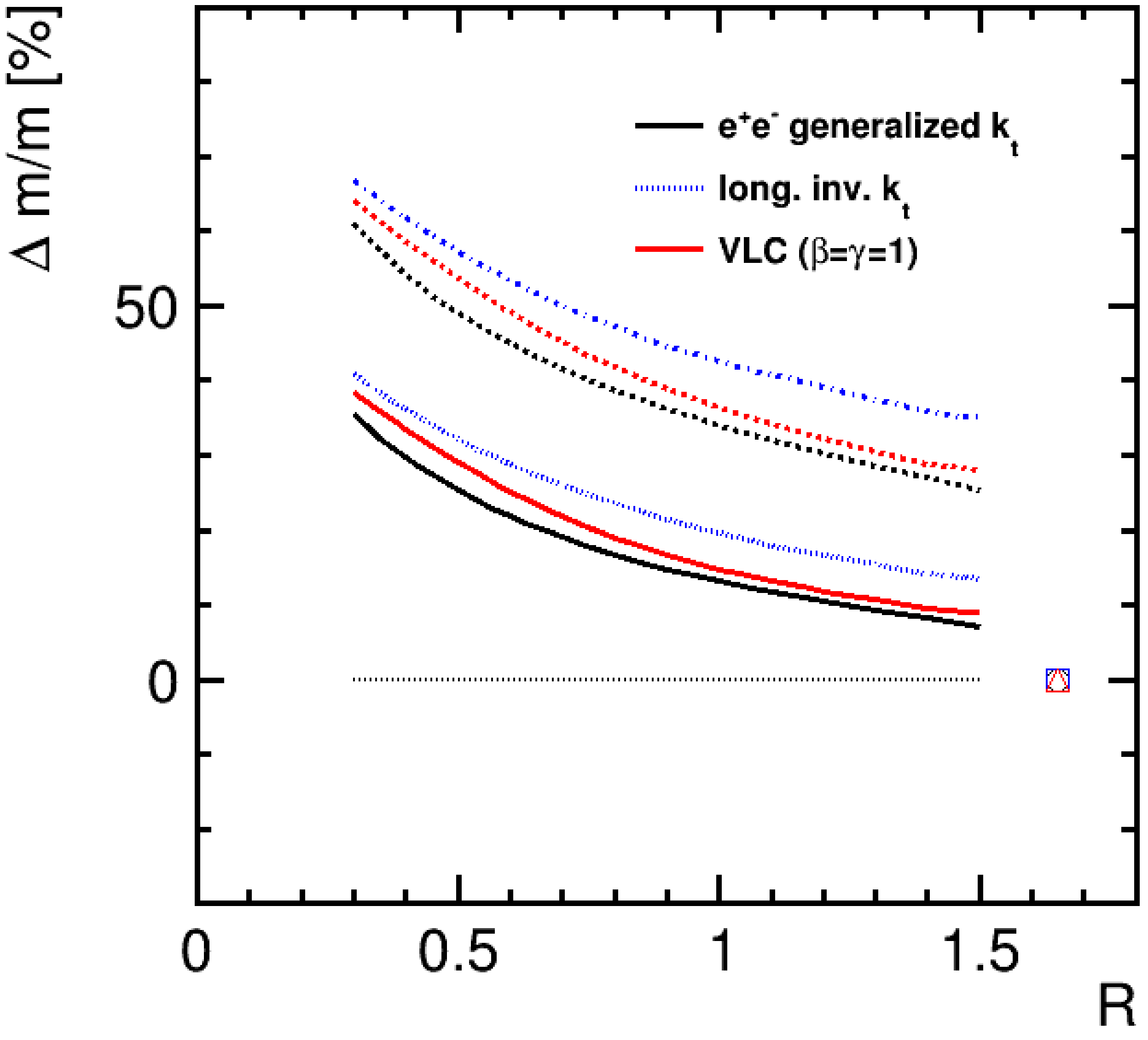
Algorithm with largest footprint has the smallest correction

VLC and long. invariant k_t no longer identical

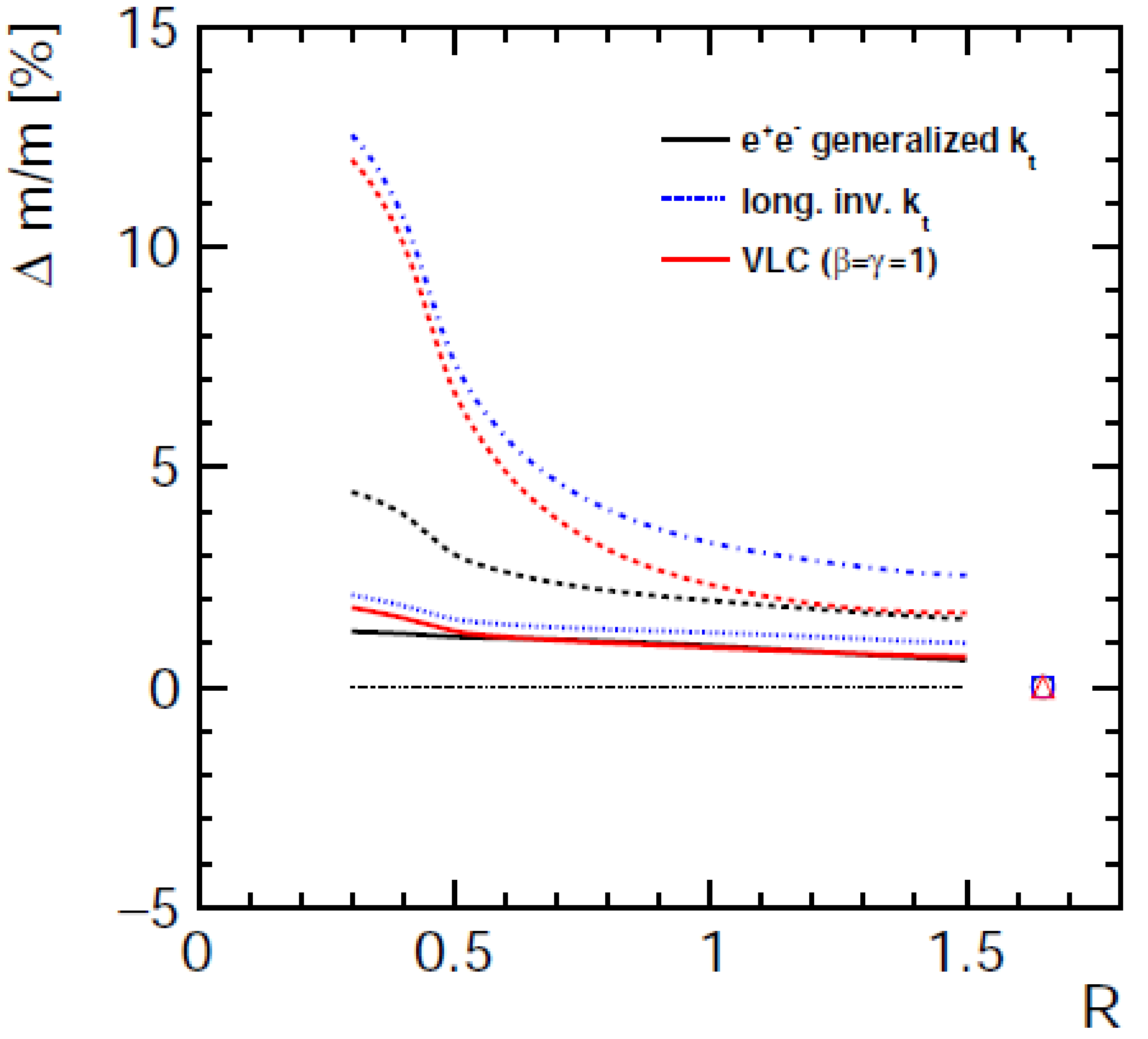
Few per mil effect at 250 GeV, 10^{-4} at 3 TeV

Non-perturbative corrections – jet mass

$e^+e^- \rightarrow q\bar{q}$ at $\sqrt{s}$ = 250 GeV



$e^+e^- \rightarrow t\bar{t}$ at $\sqrt{s}$ = 3 TeV



— mean

┈ median

Corrections to jet mass much larger than to energy
VLC much closer to generalized e^+e^-

Summary

Future lepton colliders:

- an opportunity to understand the process from parton to jet.
- a challenge to jet reconstruction (better detectors, complex final states, enhanced phase space, background, tighter control over systematics)

Traditional lepton collider algorithms fail to cope with the background level expected at future linear (circular?) colliders

Longitudinally invariant algorithms work well... and we understand why

Refurbished e^+e^- algorithms can be better still:

VLC is currently the most robust algorithm on the market

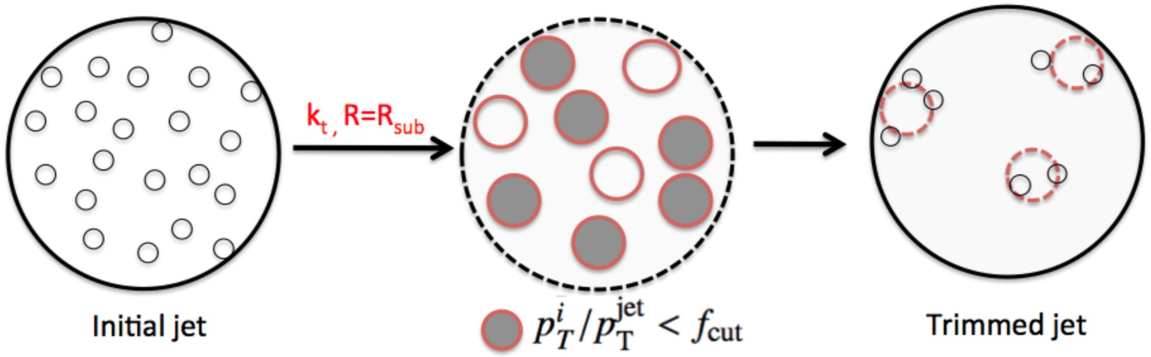
Non-perturbative corrections are less important than at LEP, but non-trivial differences between algorithms merit further study

Marcel Vos (marcel.vos@ific.uv.es)

Jet grooming

Jet grooming

One of the main recipees at the LHC to deal with pile-up contamination of large-area jets

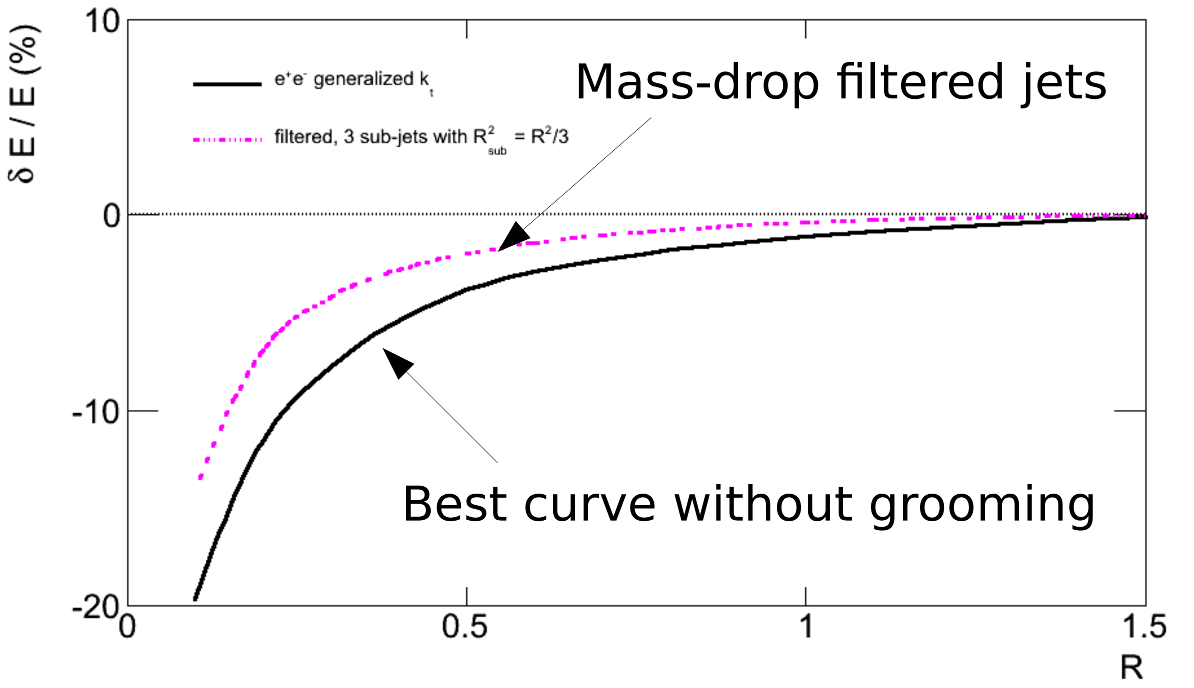


e$^+$e$^-$ grooming

Reconstruct exclusive Durham jets in e$^+$e$^-$ → $q\bar{q}$ (N=2), break up into sub-jets with mass-drop filtering with R = R$_{sub}$,

Select 3 hardest sub-jets

For fair comparison, choose R$^2_{sub}$ = R^2/3 so that area of 3 sub-jets adds up to same area



Mass-drop filtered jets

Best curve without grooming

Grooming reduces perturbative corrections for a given jet area

→ better energy response

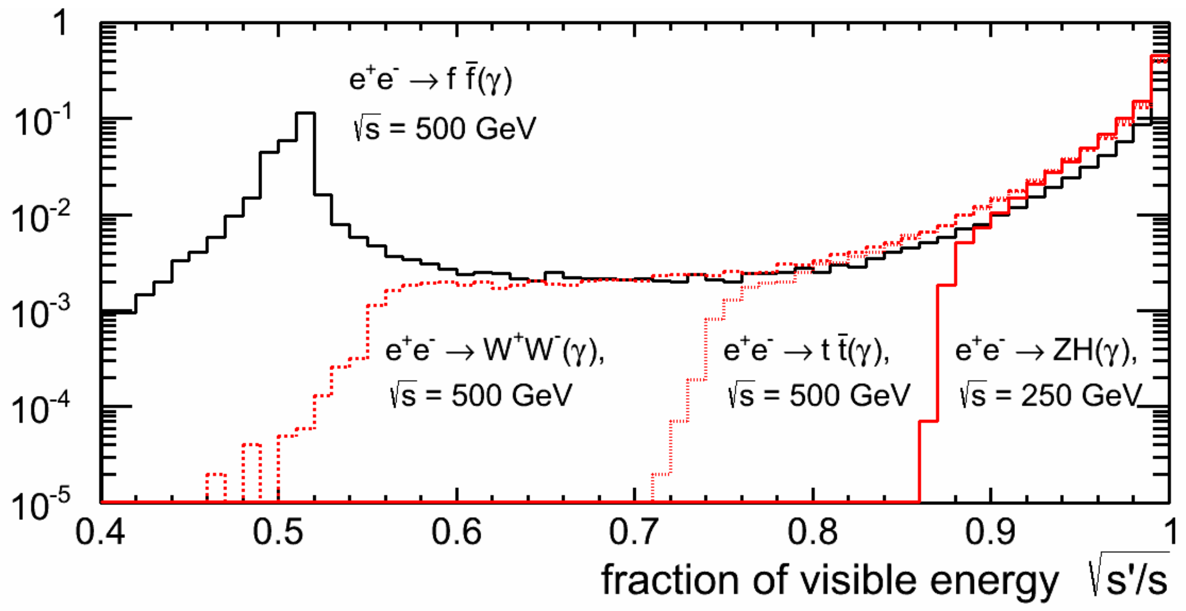
→ less exposure to background

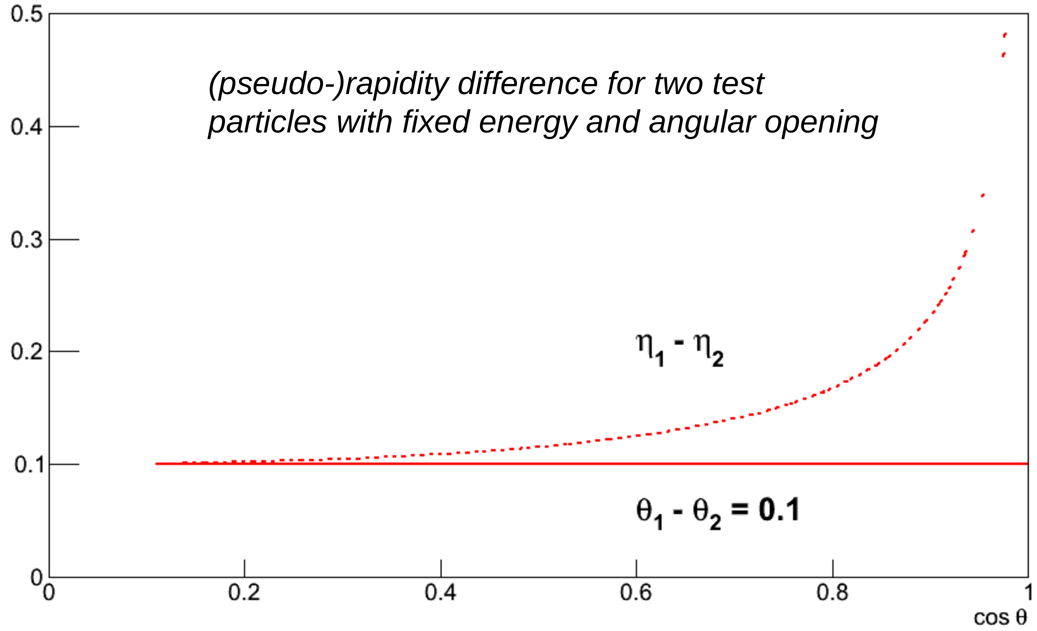
Large improvement! Deserves further study!

Jet reconstruction

Do we need/want longitudinal invariance?

No. ISR and beamstrahlung lead to some boost, but in most interesting processes, it's very small.





Should we use rapidity instead of polar angle?

No. It's potentially harmful. The rapidity difference is a poor measure of angular separation in collisions that are at rest in the laboratory.

Jet reconstruction

Does a hadron collider algorithm work better at at lepton collider than the equivalent lepton collider algorithm?

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

vs.

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

Beam jets + shrinking footprint with polar angle yields increased robustness against forward-peaked γγ → hadrons!

The Valencia algorithm is an attempt to get the best of both worlds (with a twist):

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

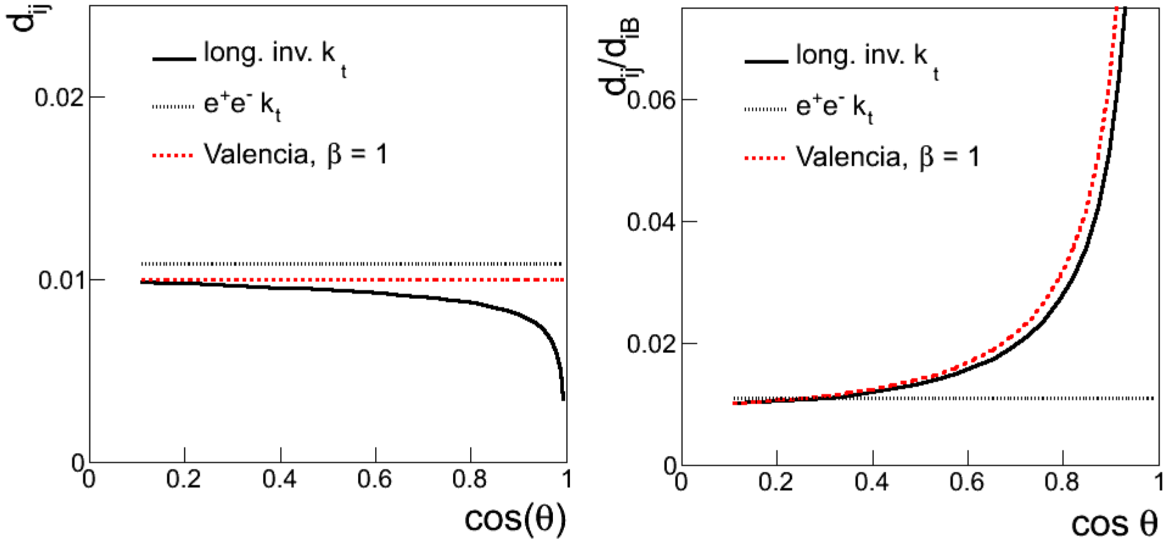
e^+e^- distance between particles

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

beam distance to mimic d_{ij}/d_{iB} behaviour

β to tweak background rejection

Boronat, Garcia, MV, arXiv:1404.4294, fjcontrib/trunk



Two test particles with constant energy (E = 1 GeV)

and fixed polar angle separation (100 mrad)