Jet reconstruction – challenges

Marcel Vos
IFIC (UVEG/CSIC) Valencia

ILD top/QCD topical meeting, 22 Nov. 2017

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

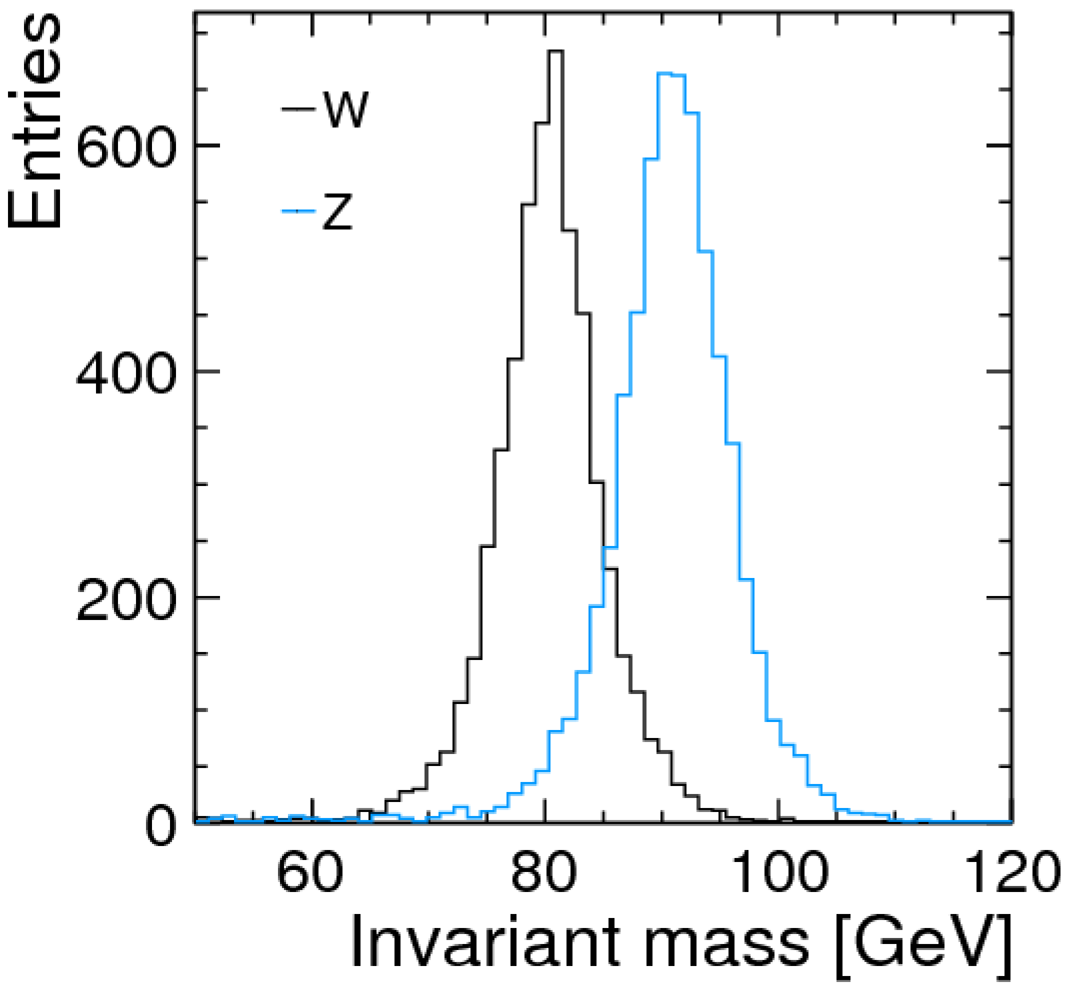
Hadronic final states are important for the precision e^+e^- programme

- *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

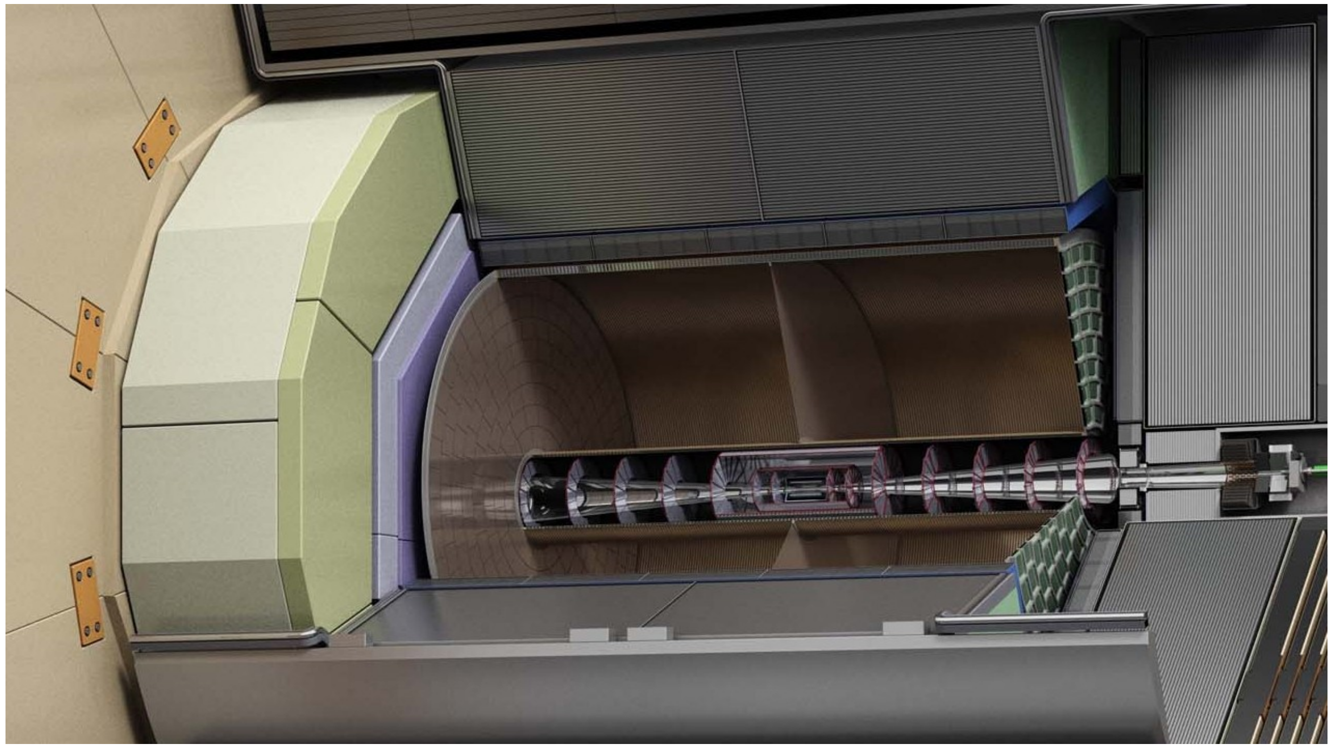
Performance goal: distinguish hadronic W and Z decays

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

Detectors

LC detector concepts optimized for particle flow

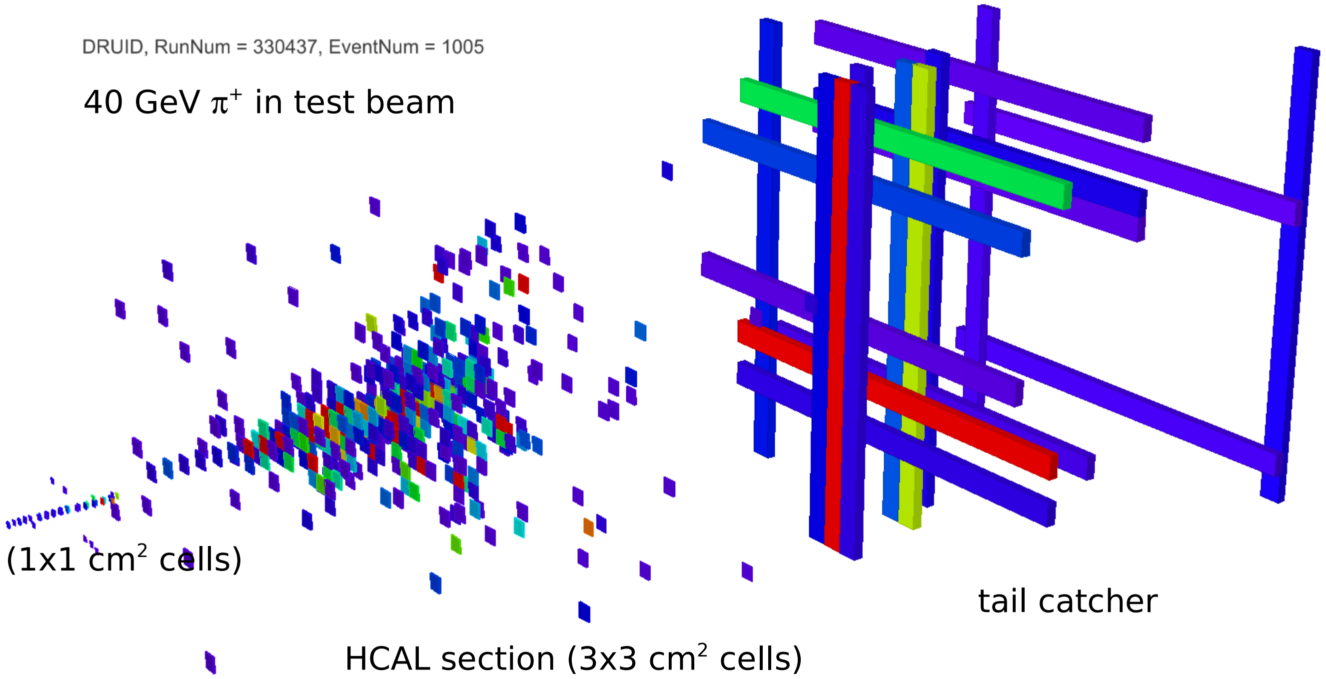
- highly granular calorimeter
- 3-5 Tesla solenoid
- state-of-the art low-mass tracking system
- precision vertexing



Particle flow with highly granular calorimeter and ultra-transparent is expected to yield the best single-particle response ever achieved

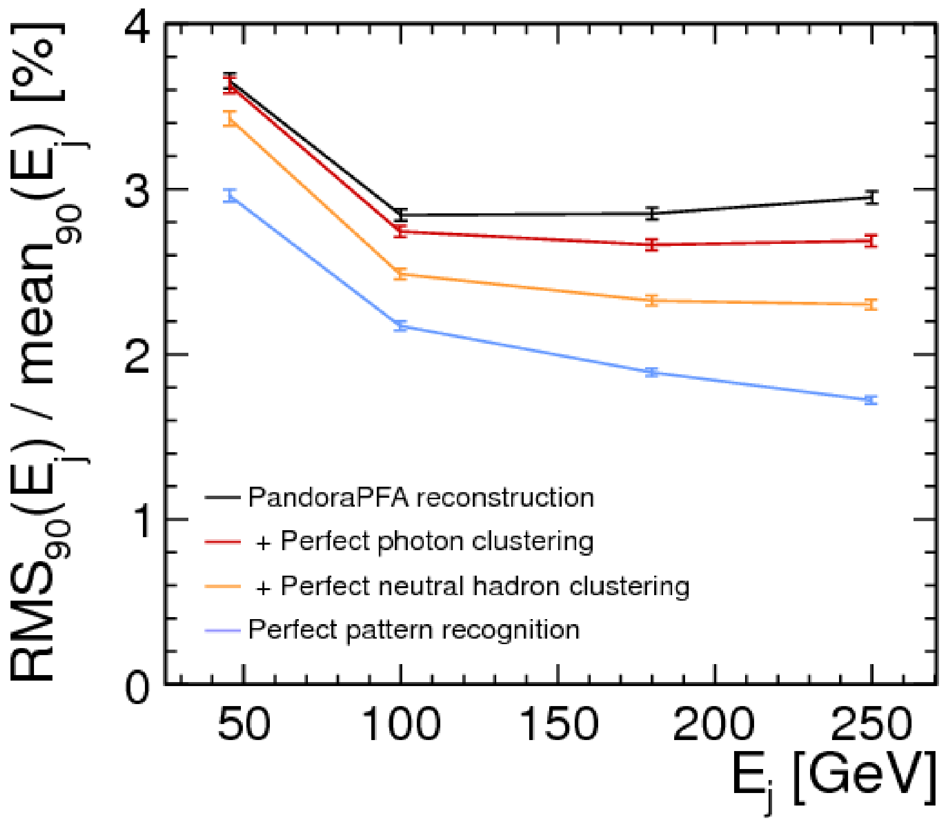
Not (entirely) science fiction:

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

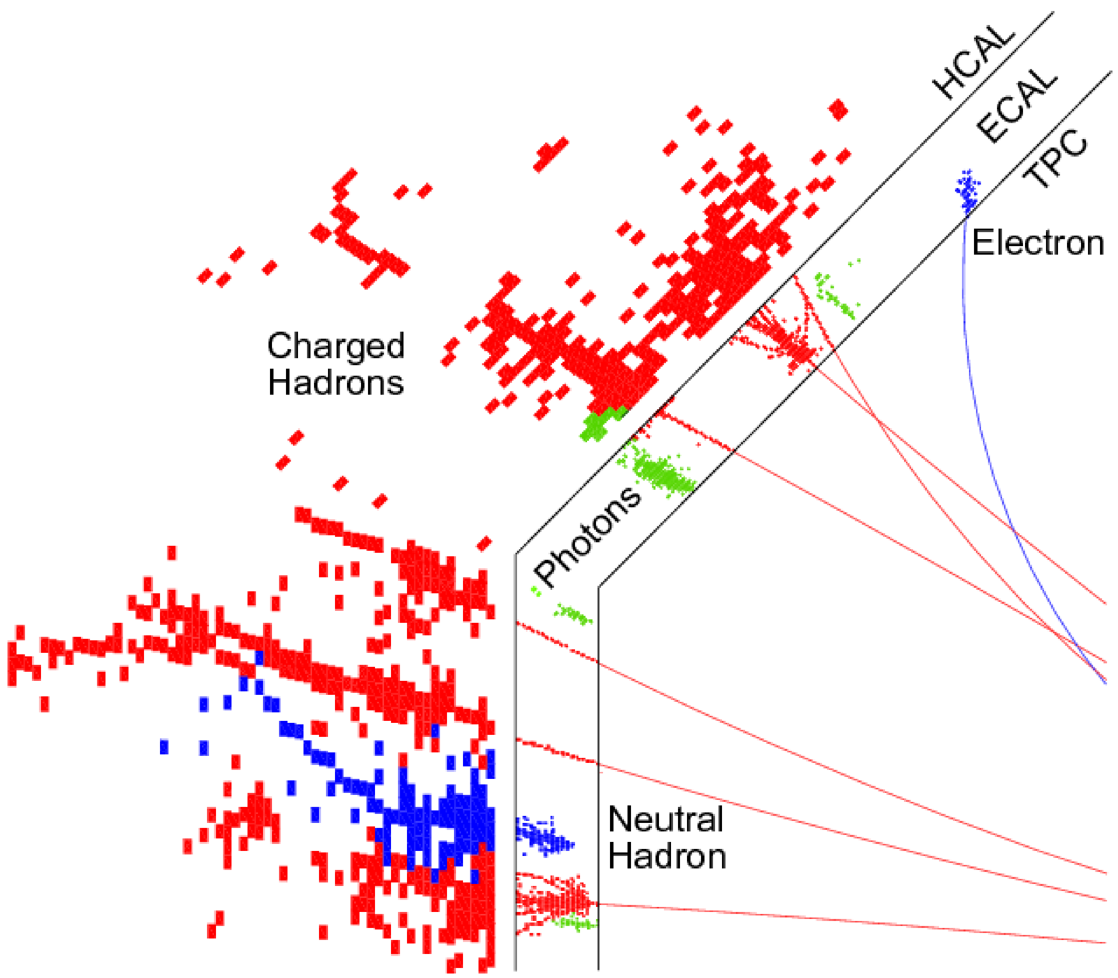


Particle Flow

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: $\Delta E/E \sim 3\%$
This performance is measured and detectors are benchmarked on very simple final states, or even without jet clustering



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 limits the performance

Detector level (Particle Flow objects)

Particle level (stable MC particles)

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

Detector limitations...

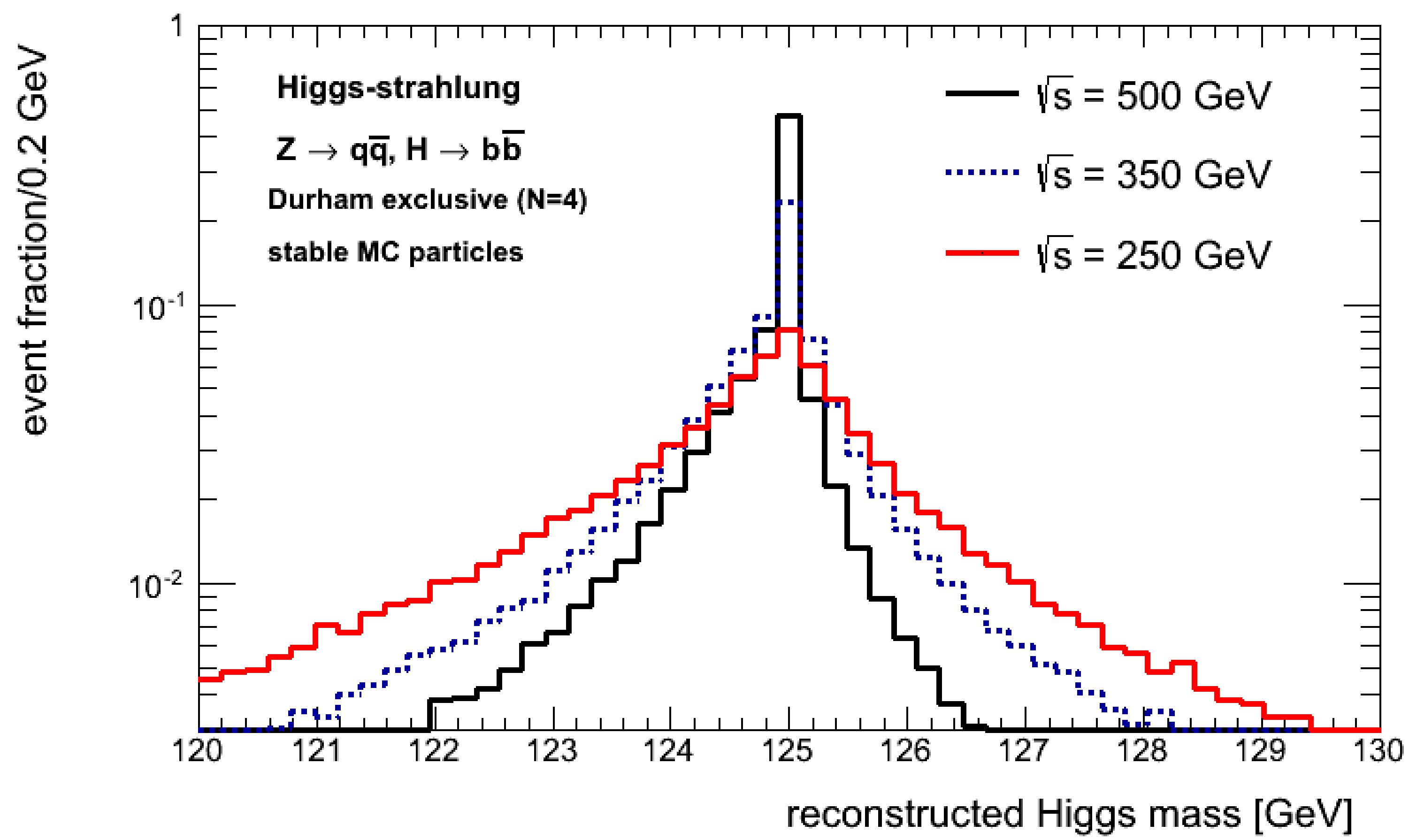
Limitations
of jet algorithms...

Jets reconstructed on truth particles are a crucial tool to separate detector and clustering

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

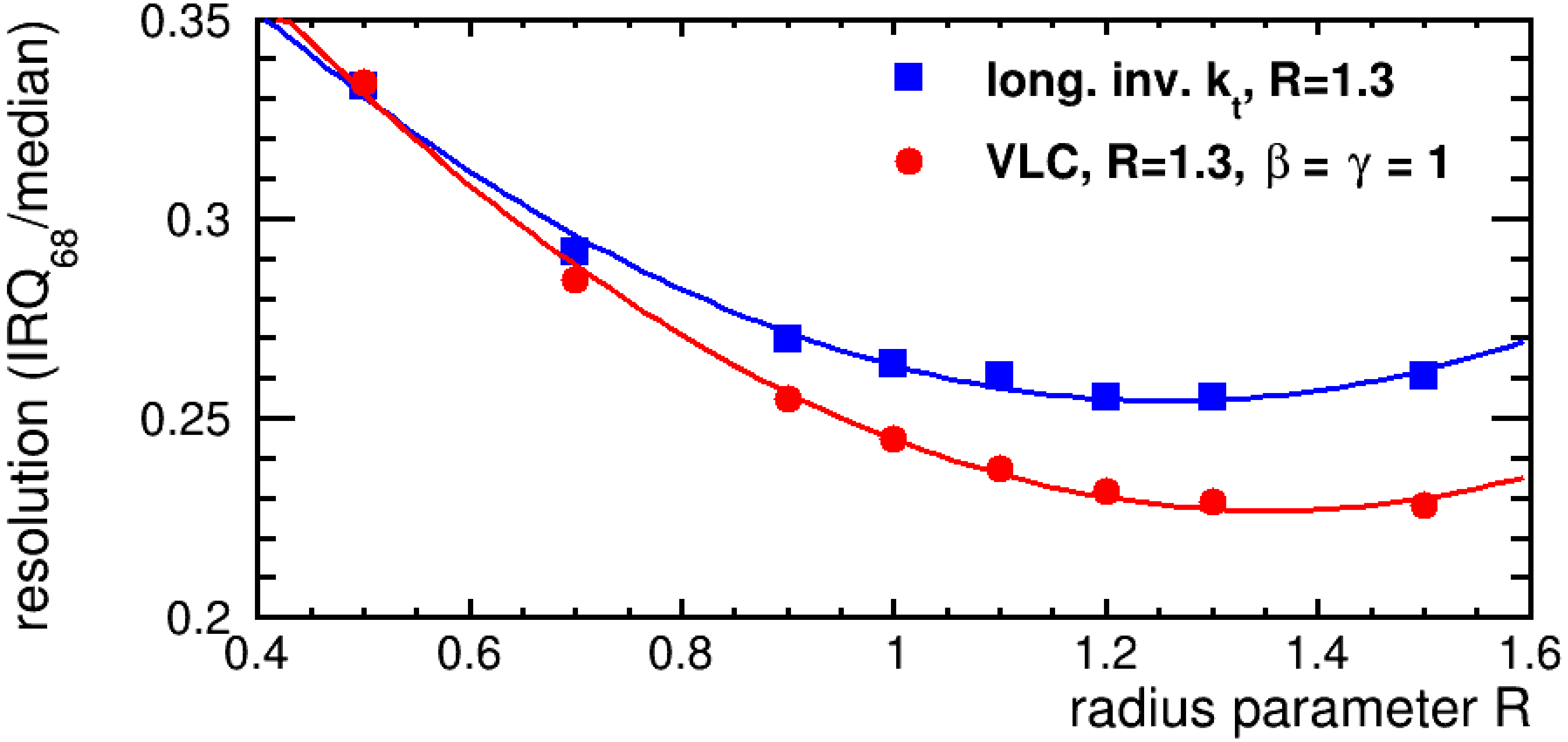
Notorious examples:

tt, t → cH (Zarnecki)
ttH (Price & Strube)
ZHH (Junping Tian et al.)



Problem: simple distance criterion is insufficient, teach jet algorithms more QCD (shower deconstruction, fuzzy jets, machine learning workshop at LBNL in December)

Benchmark: di-Higgs production



Di-Higgs production at CLIC at 3 TeV: 20-25% mass resolution for the Higgs boson candidates

Inclusive or exclusive clustering

Exclusive clustering: N hardest splits correspond to N final state partons

- Only possible for sequential jet algorithms with a positive exponent (i.e. no anti-kt)
- Basic assumption breaks down in complex final states with a hierarchy of scales

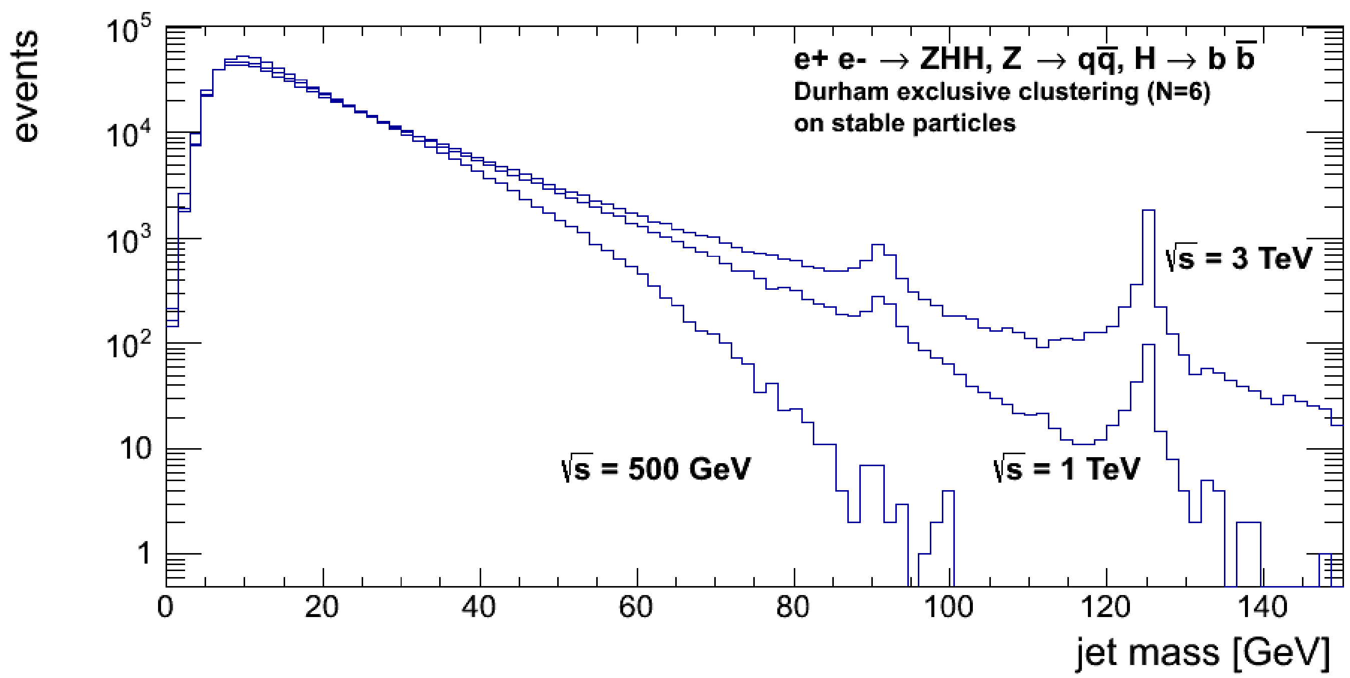
Still: exclusive clustering found to offer superior performance in ~all benchmark analyses

Explore anti-kt-like exclusive clustering: XCone jets, arXiv:1508.01516

ZHH production at very high energy

Hierarchy in energies increases with √s

Failures in clustering show up at high energy

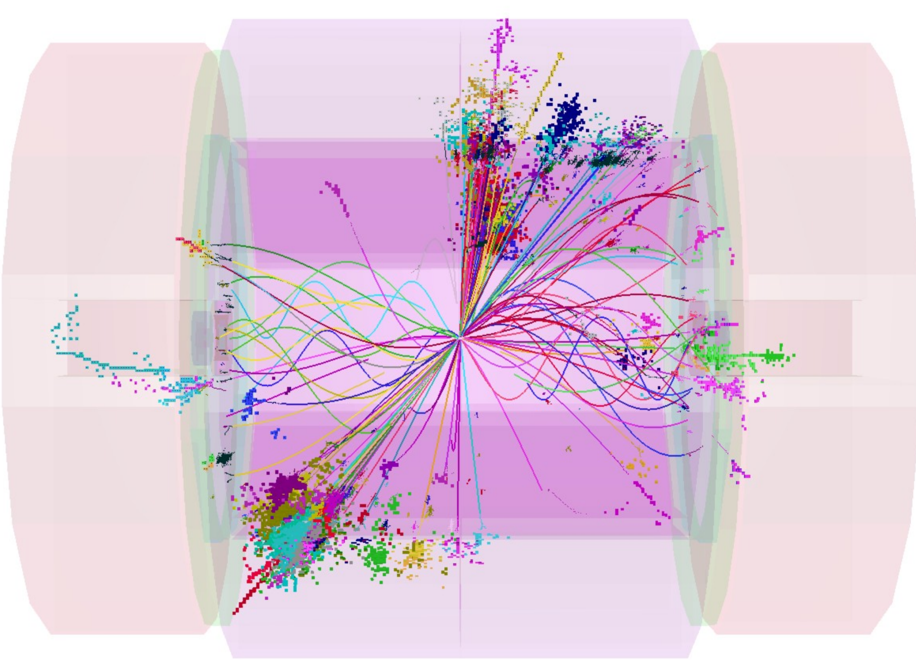


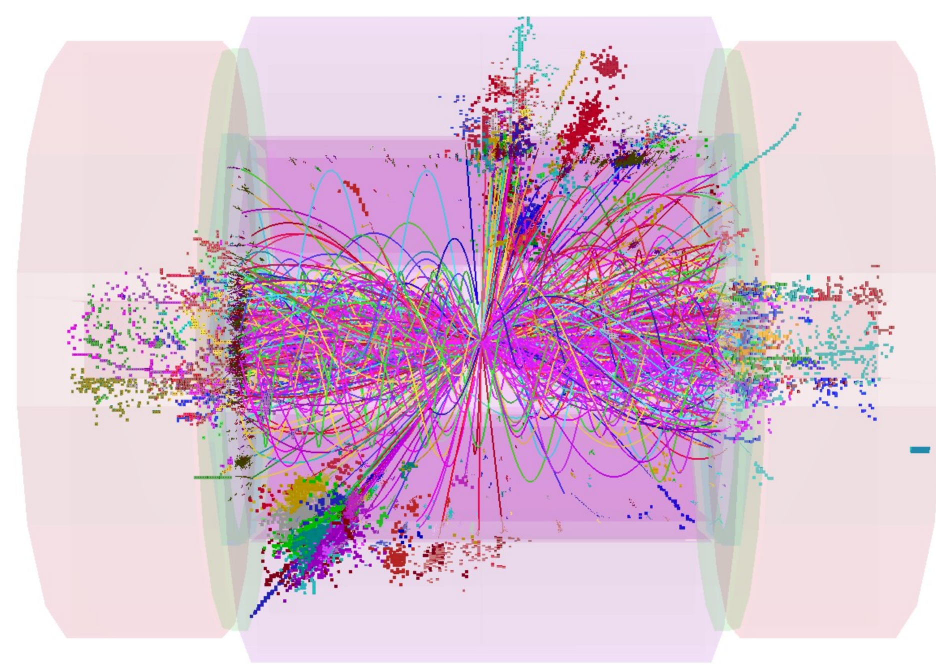
Lepton collider backgrounds

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

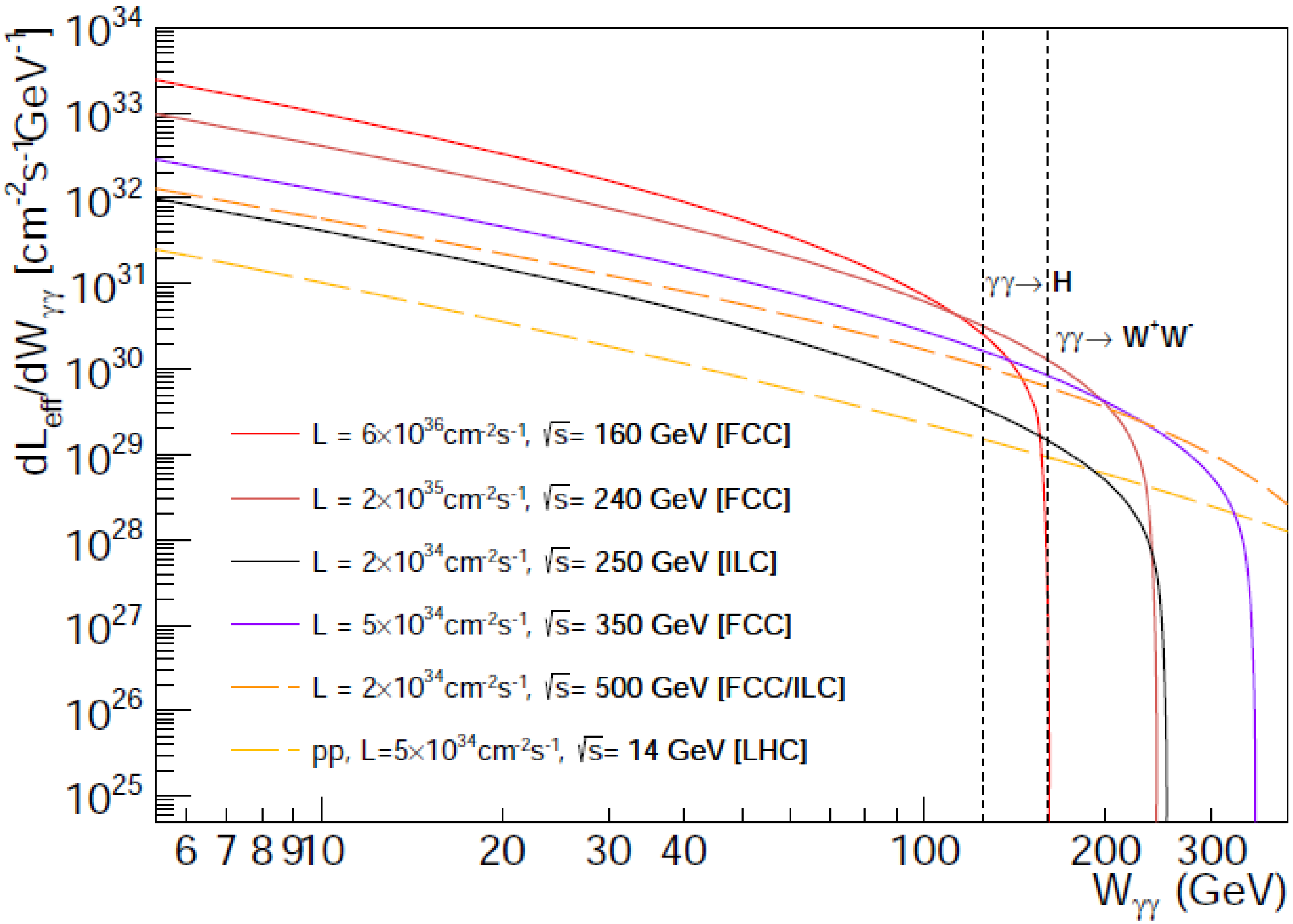
} Important at high energy and luminosity/BX





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.

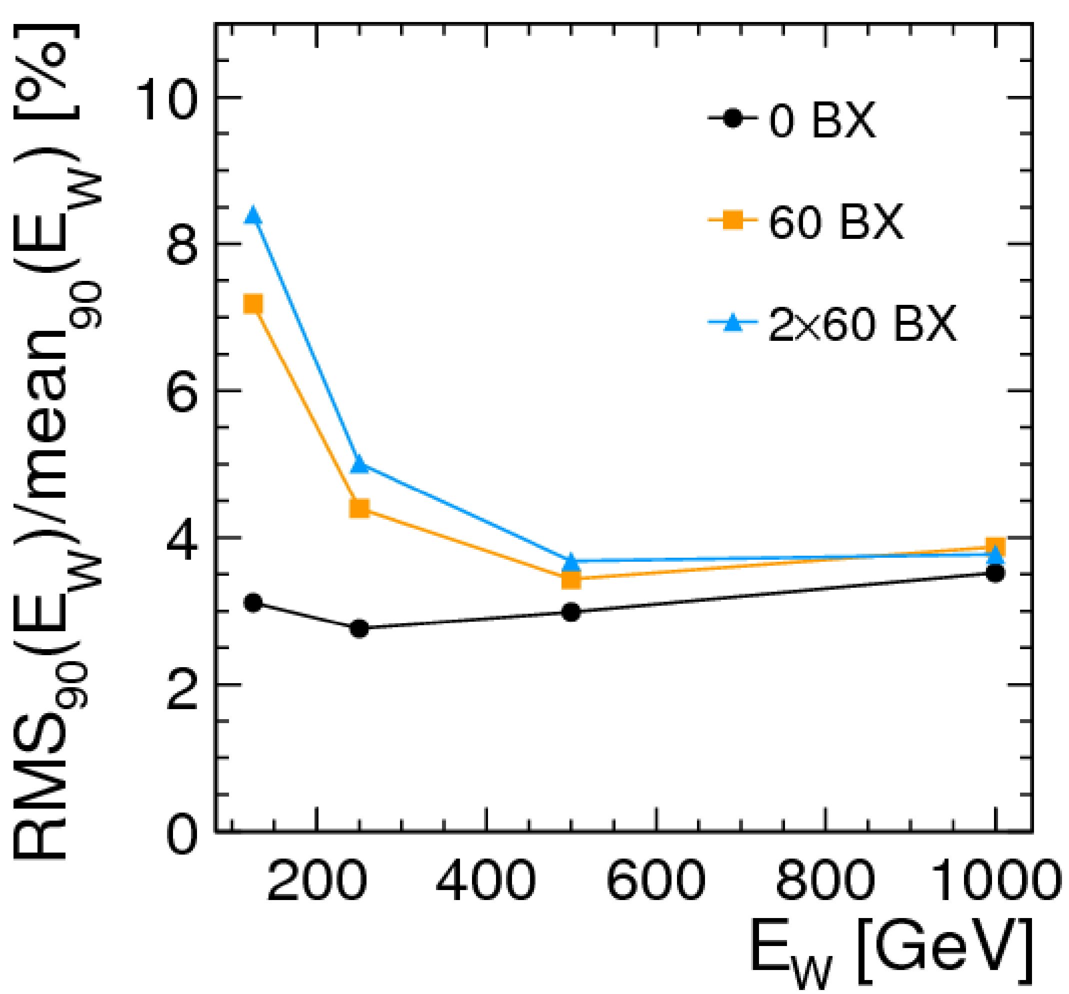
$\gamma\gamma \rightarrow$ hadrons

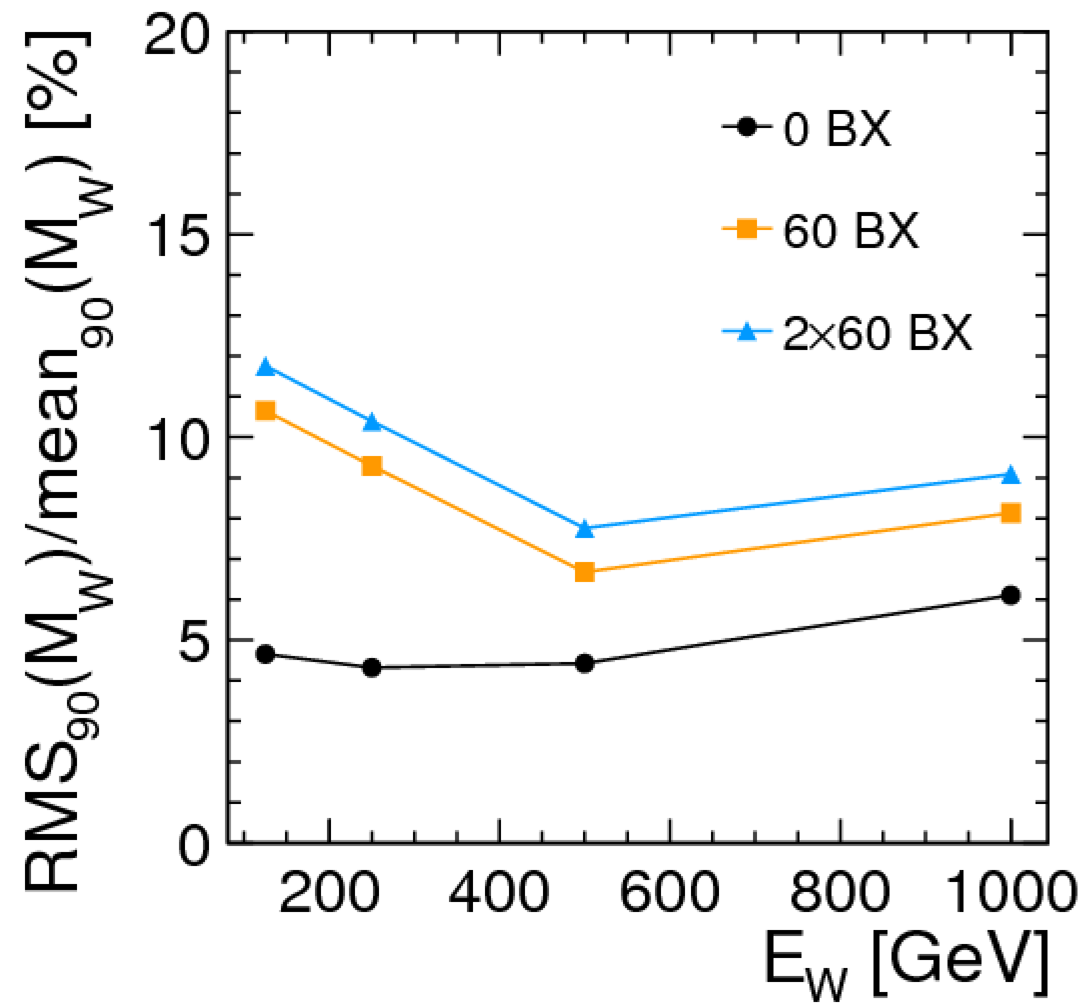


Use CLIC case as a stress test for jet reconstruction;
If it works there, it's good for ILC too.
FCCee has much smaller $\gamma\gamma \rightarrow$ hadrons background still.

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)



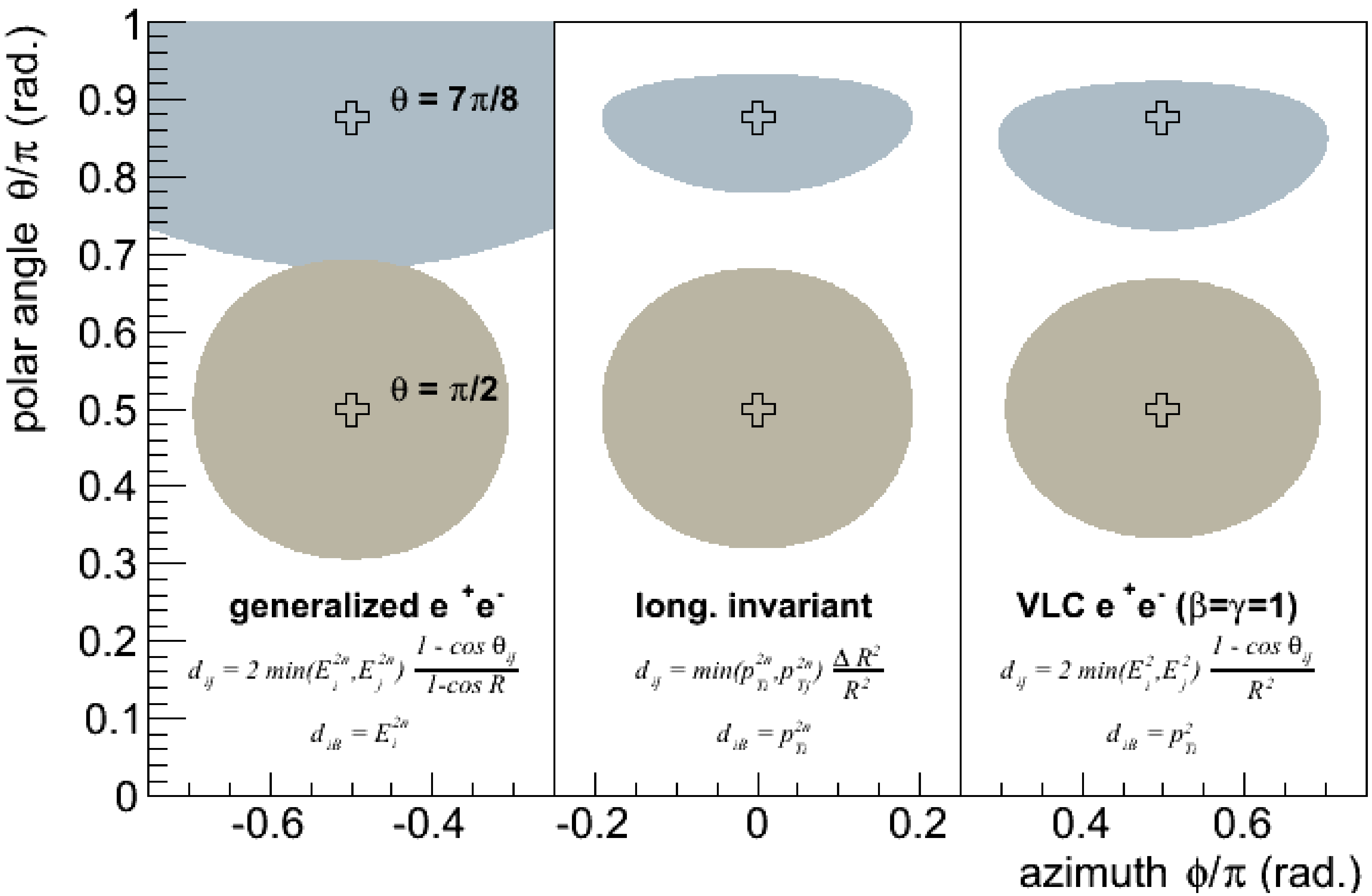


Energy resolution at high energy is not too badly affected,
mass resolution suffers everywhere

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

Jet area

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



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 algorithm space

VLC algorithm of arXiv:1607.05039

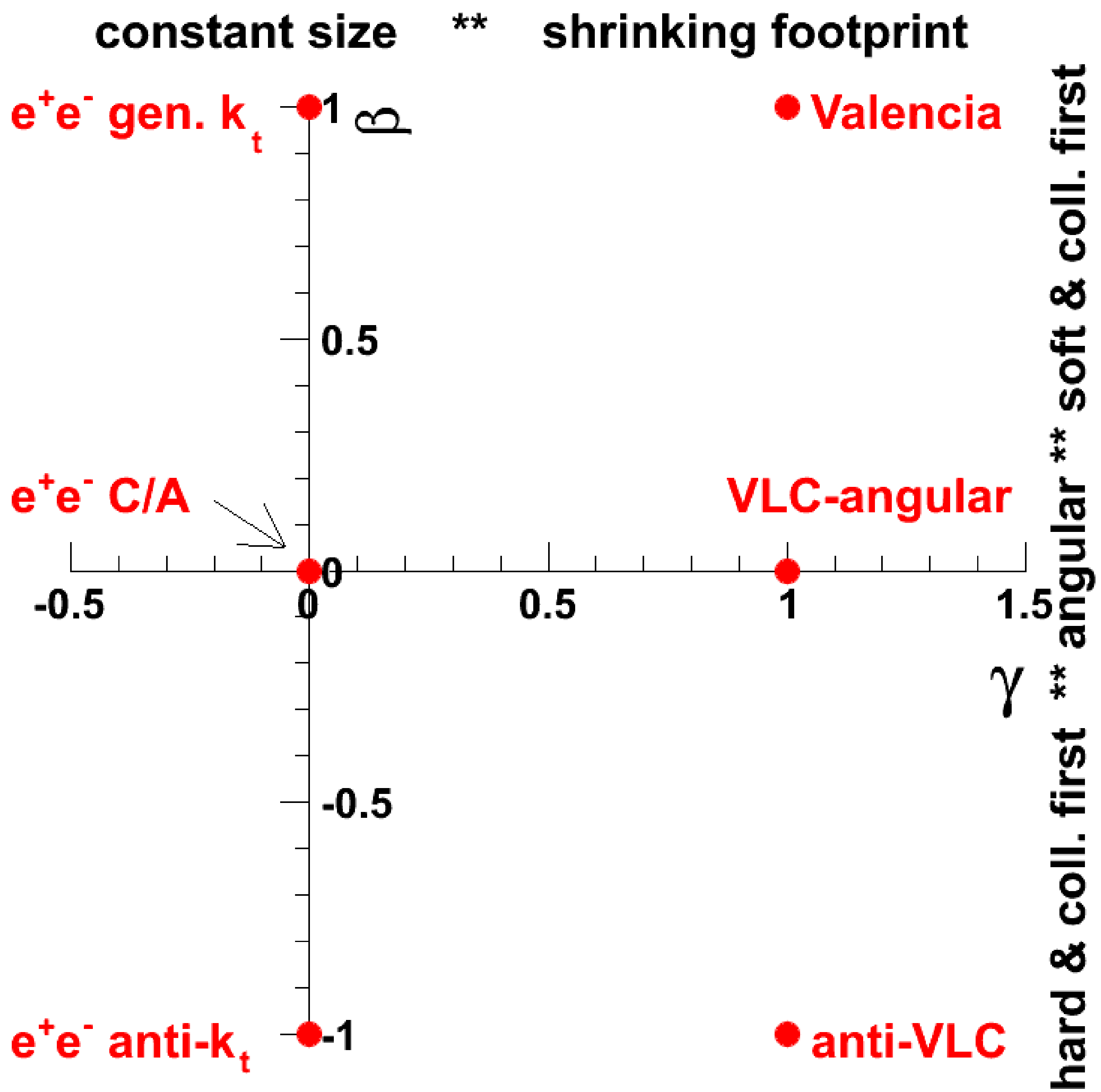
$$d_{ij} = 2 \min(E_i^{2\beta}, E_j^{2\beta})(1 - \cos\theta_{ij})/R^2,$$

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

Two parameters (real numbers) govern the clustering order (β) and robustness against background (γ)

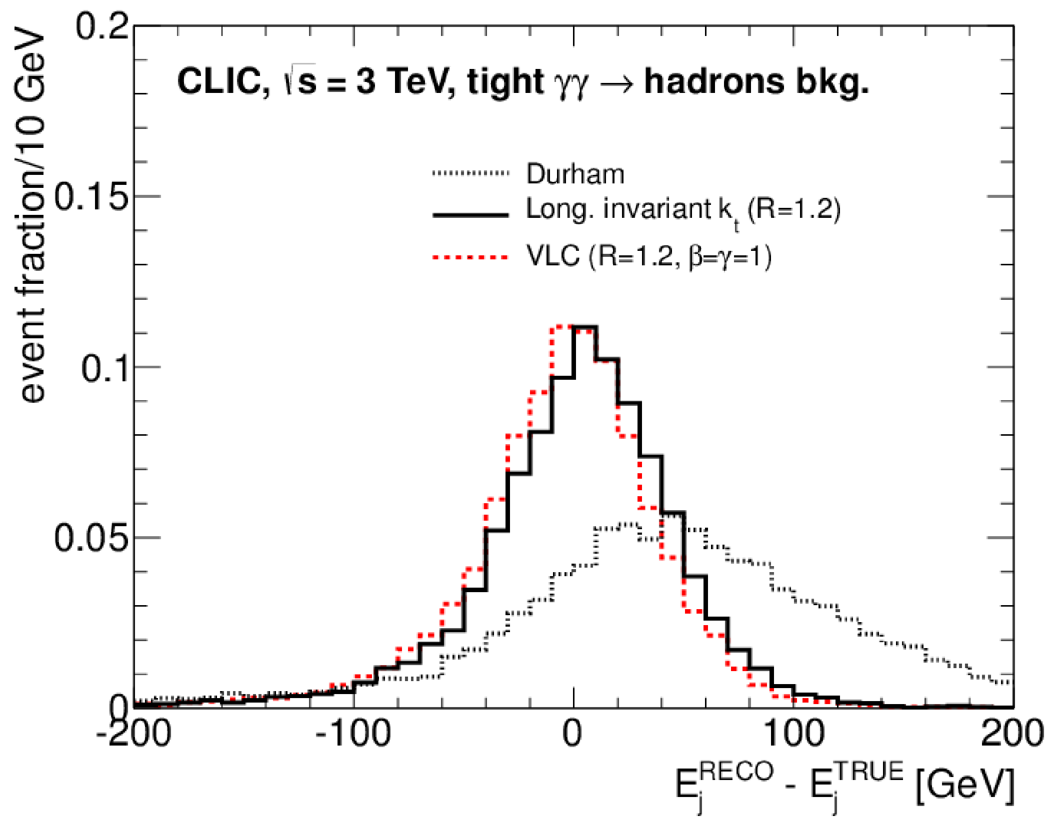
Recover generalized e+e- kt for γ=0

Mimic longitudinally invariant algorithms with γ=1



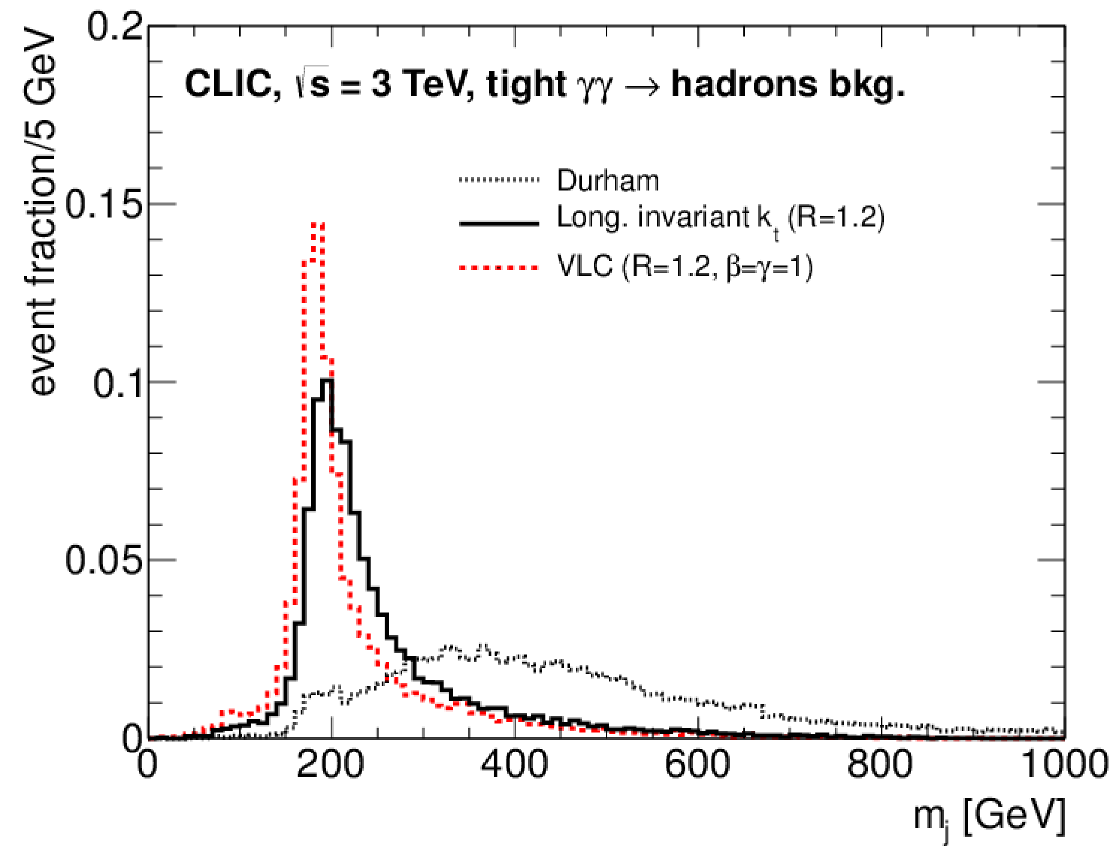
Marcel Vos (marcel.vos@ific.uv.es)

Benchmark: $t\bar{t}$ production



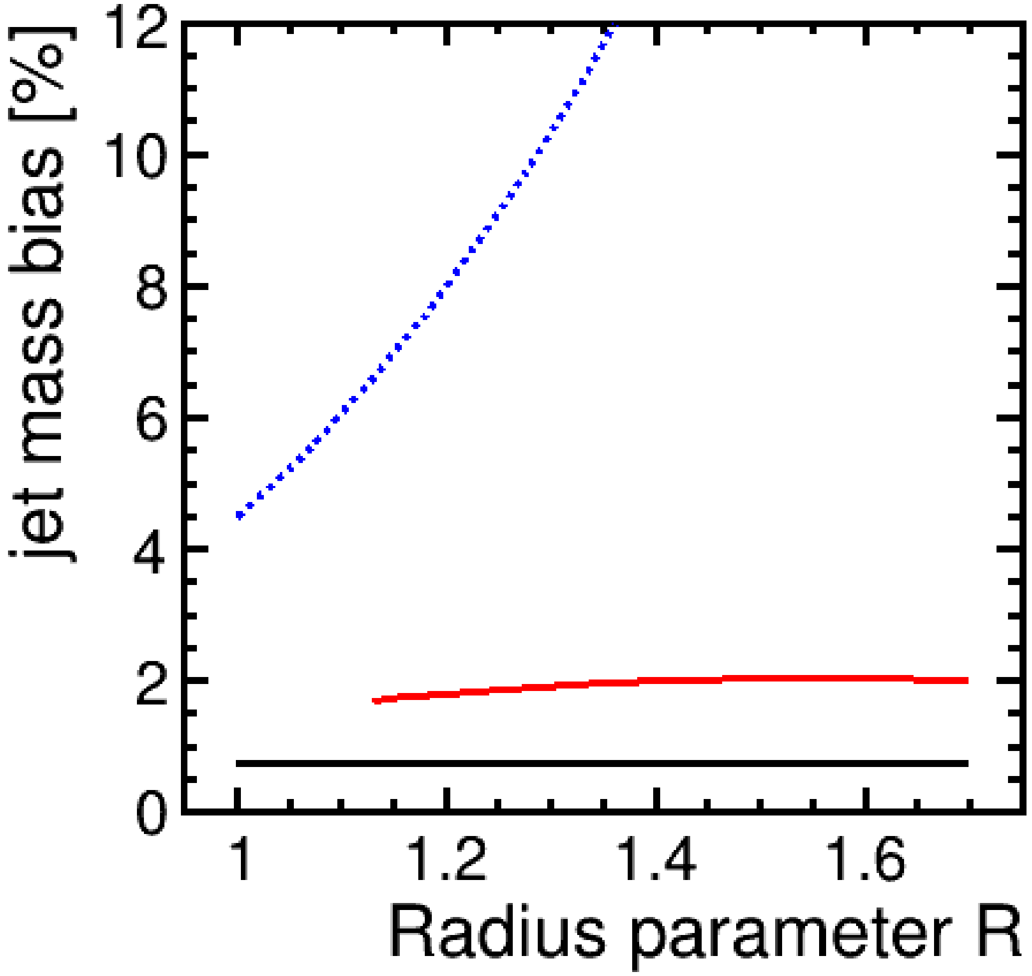
Confirming once more:
classical lepton collider algorithms cannot cope with background

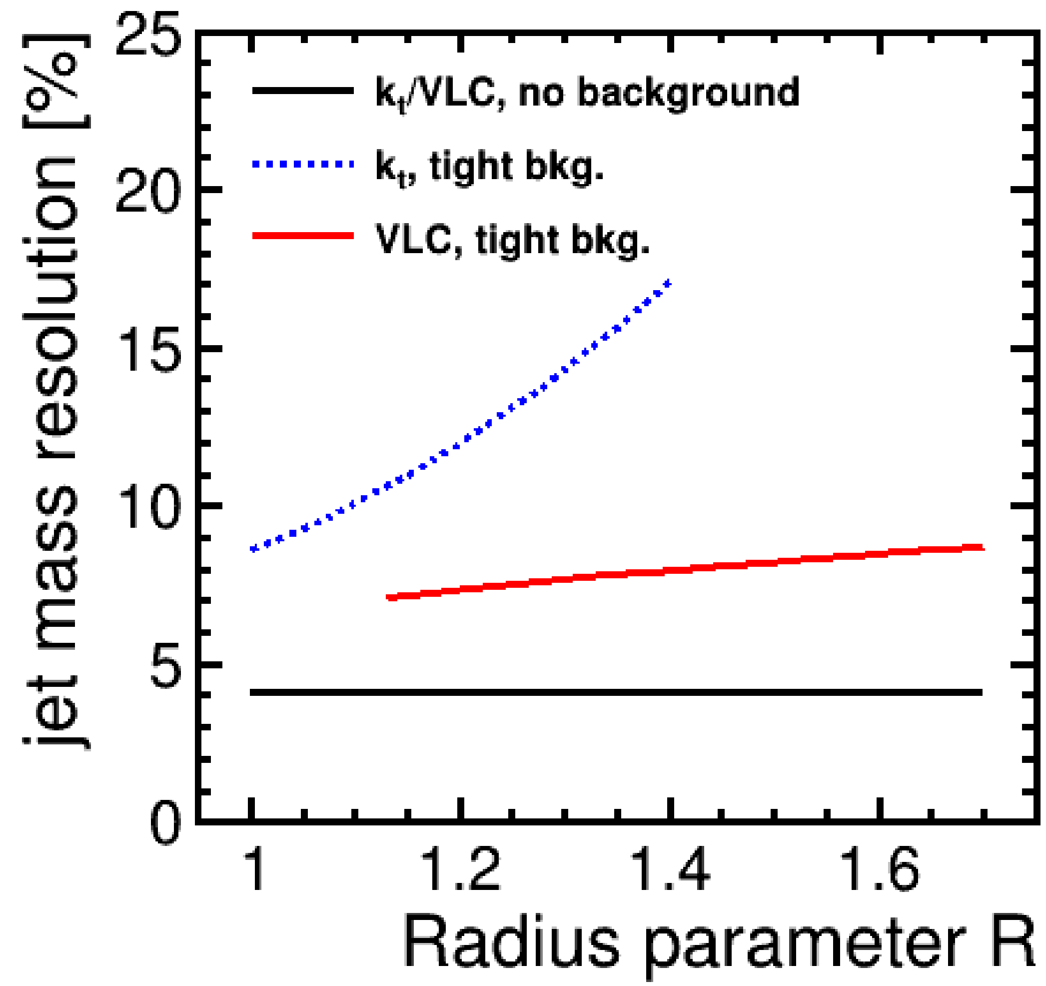
Subtle differences between alternatives
VLC algorithm is more robust than hadron collider algorithms



Benchmark: $t\bar{t}$ production

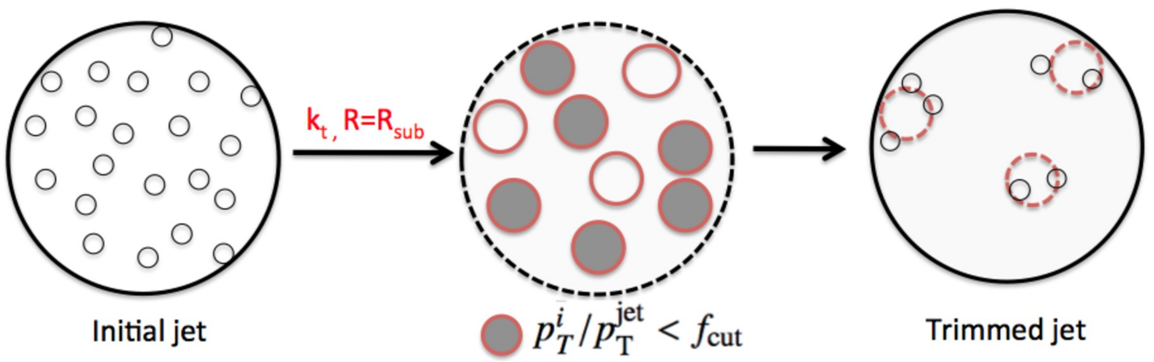
Lessons from 3 TeV analysis (if it works there, it will work everywhere):
Classical Durham algorithm is off the chart
VLC outperforms longitudinally invariant algorithms.
We do not fully recover the no-background performance.





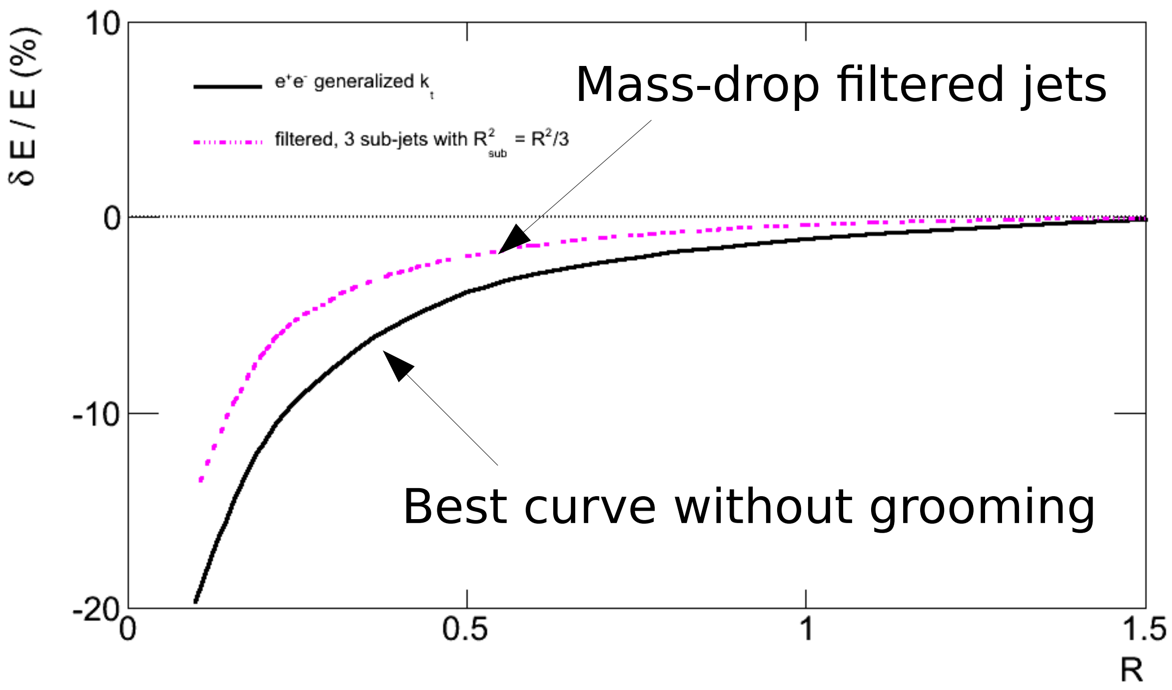
Jet grooming

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 $e^+e^- \rightarrow q\bar{q}$ (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



Grooming reduces perturbative corrections for a given jet area
→ better energy response
→ less exposure to background

An e+e- variant of reclustered jets (roughly equivalent to trimming) is adopted by CLIC for highly boosted top quark reconstruction (I. García, R. Ström)

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

(non-) perturbative corrections

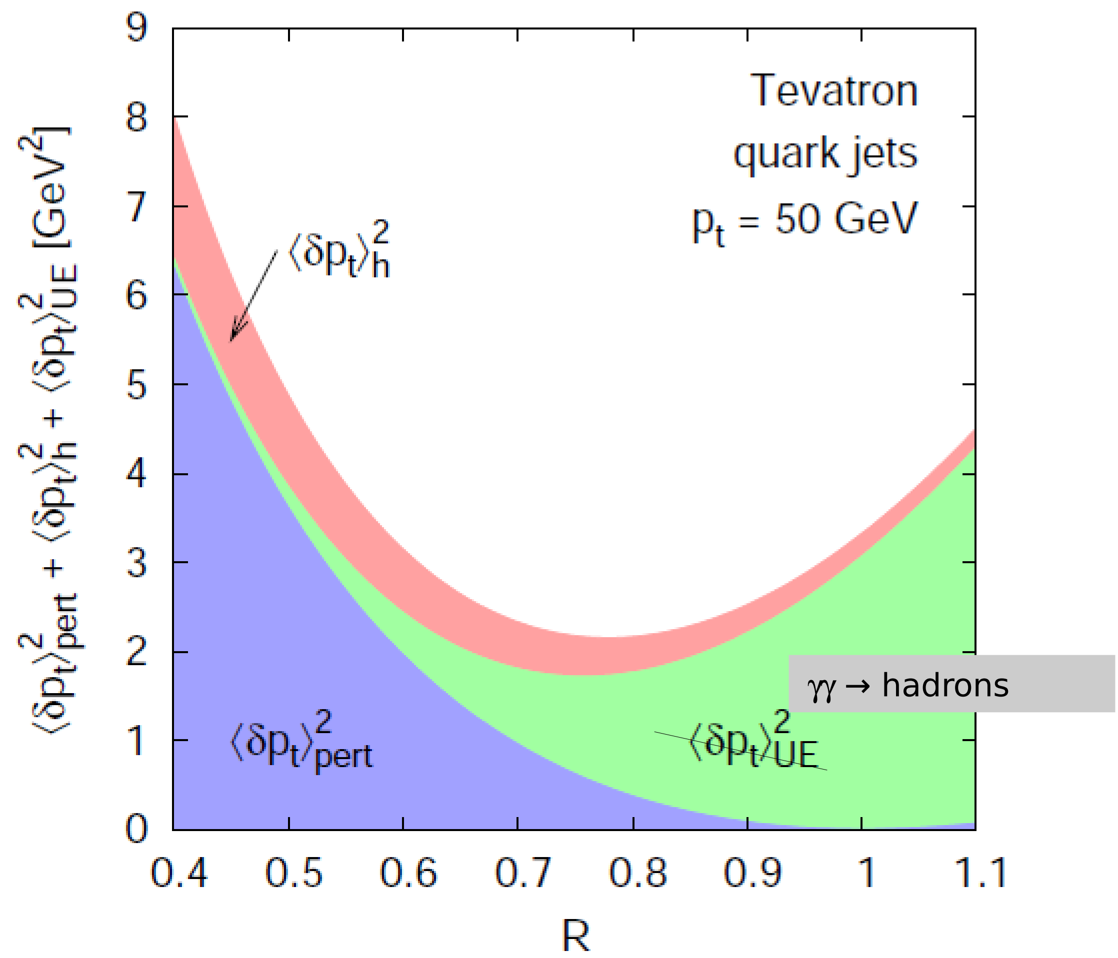
Uncertainties in jet response are an important source of systematics

Jet area and footprint determine energy response:

– (non-) perturbative corrections decrease with increasing R

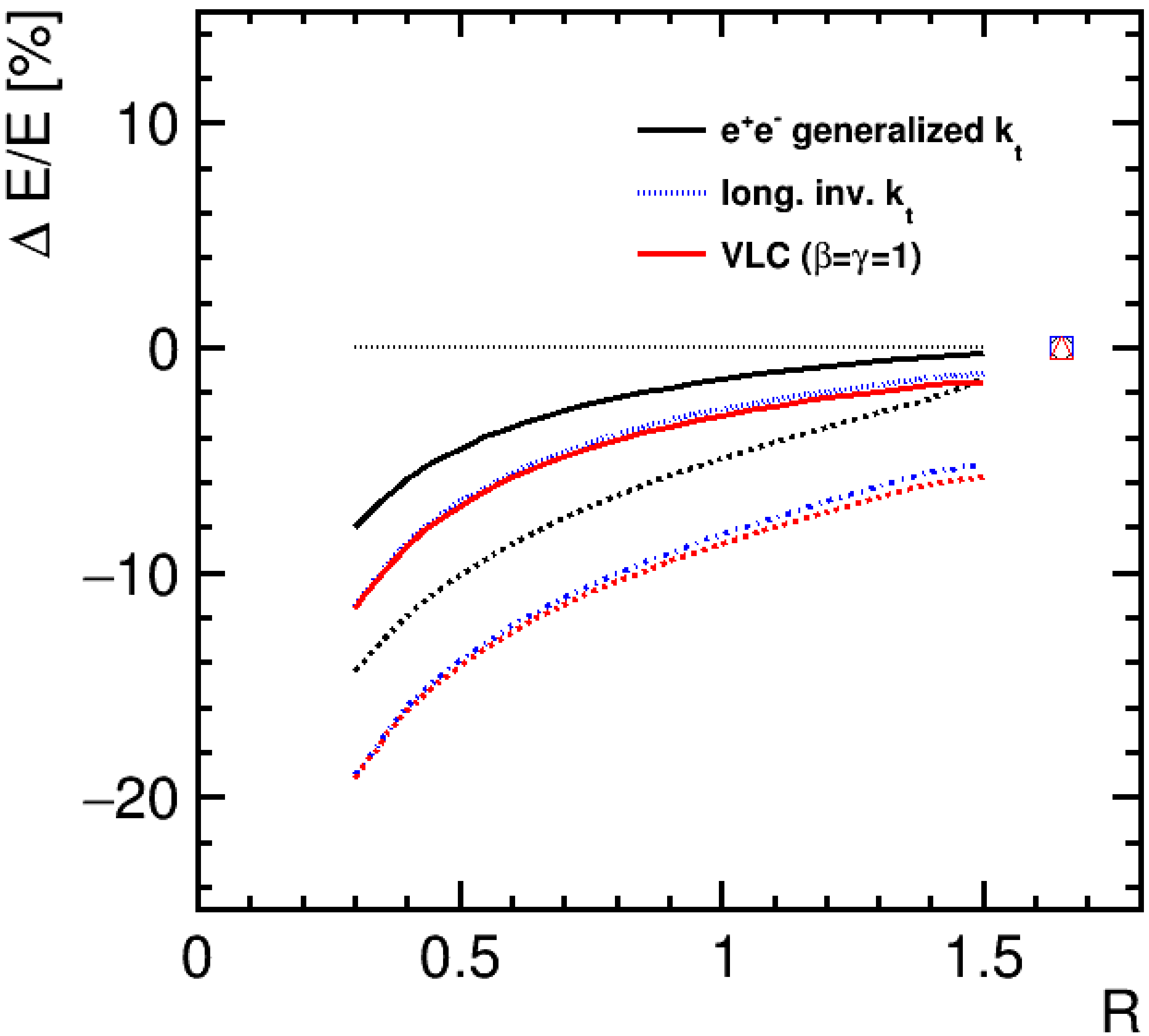
– background contribution scales with R^2

Dasgupta, Magnea, Salam, JHEP0802 (2008) 055

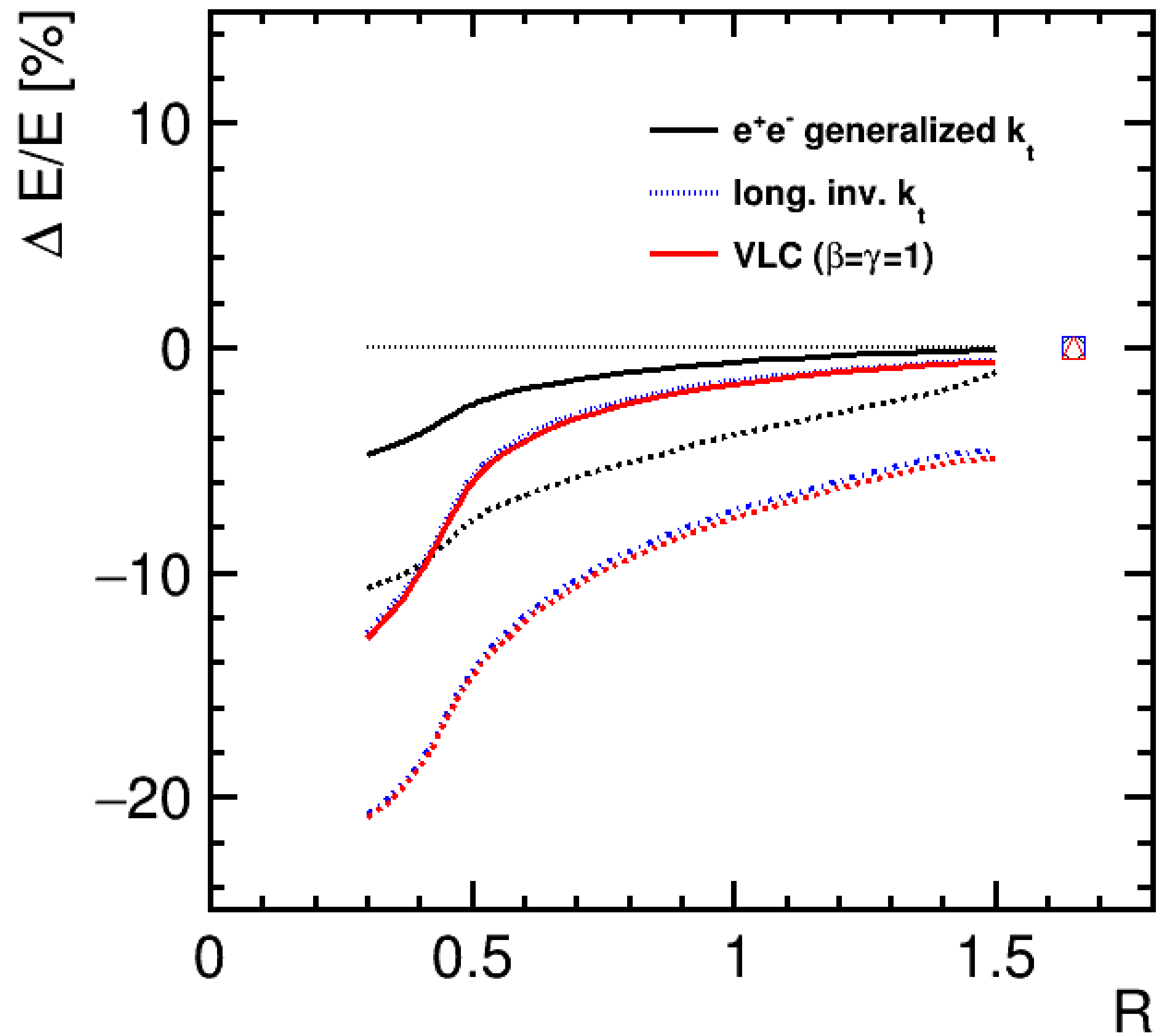


Perturbative corrections

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



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



mean

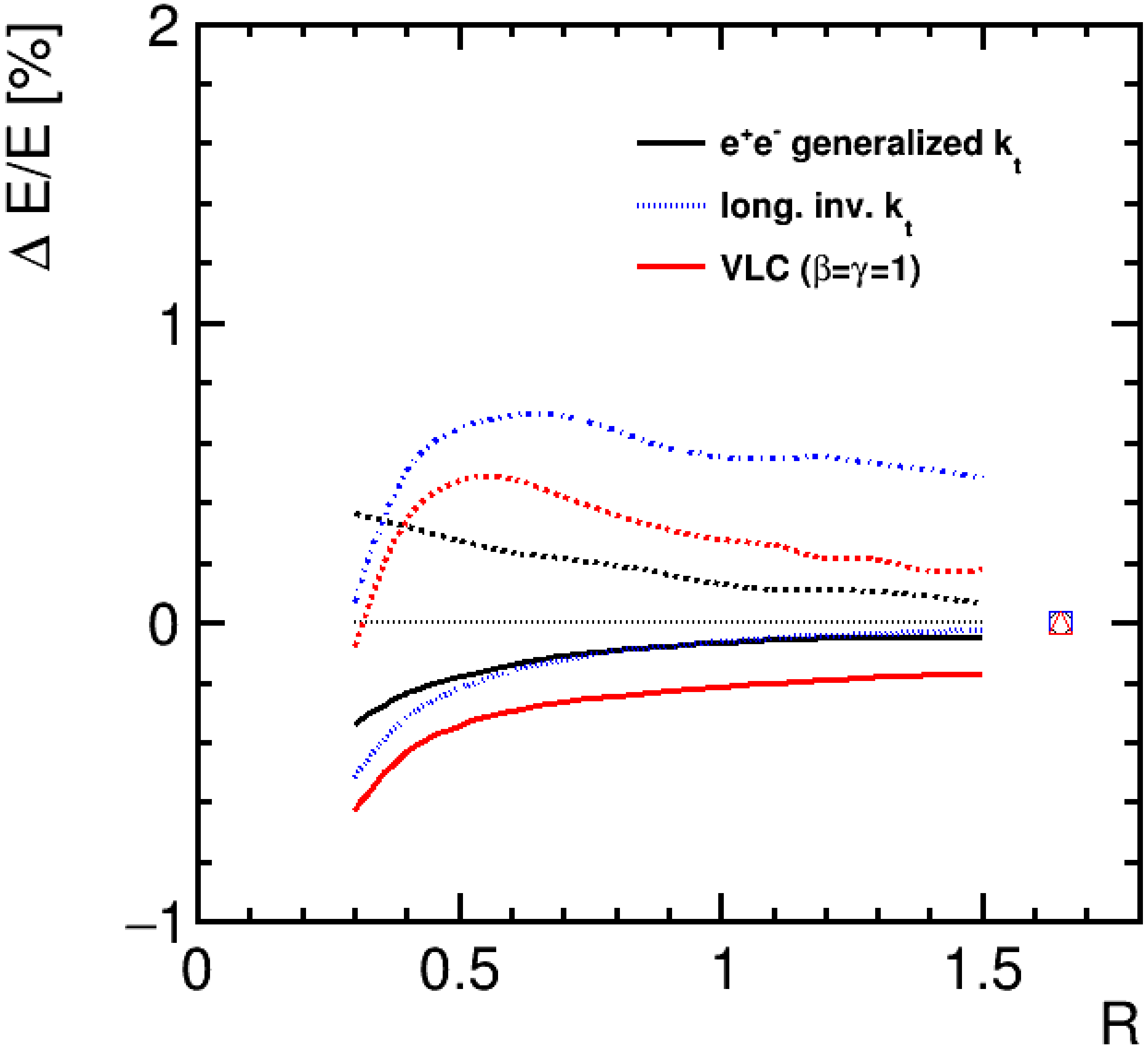
median

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

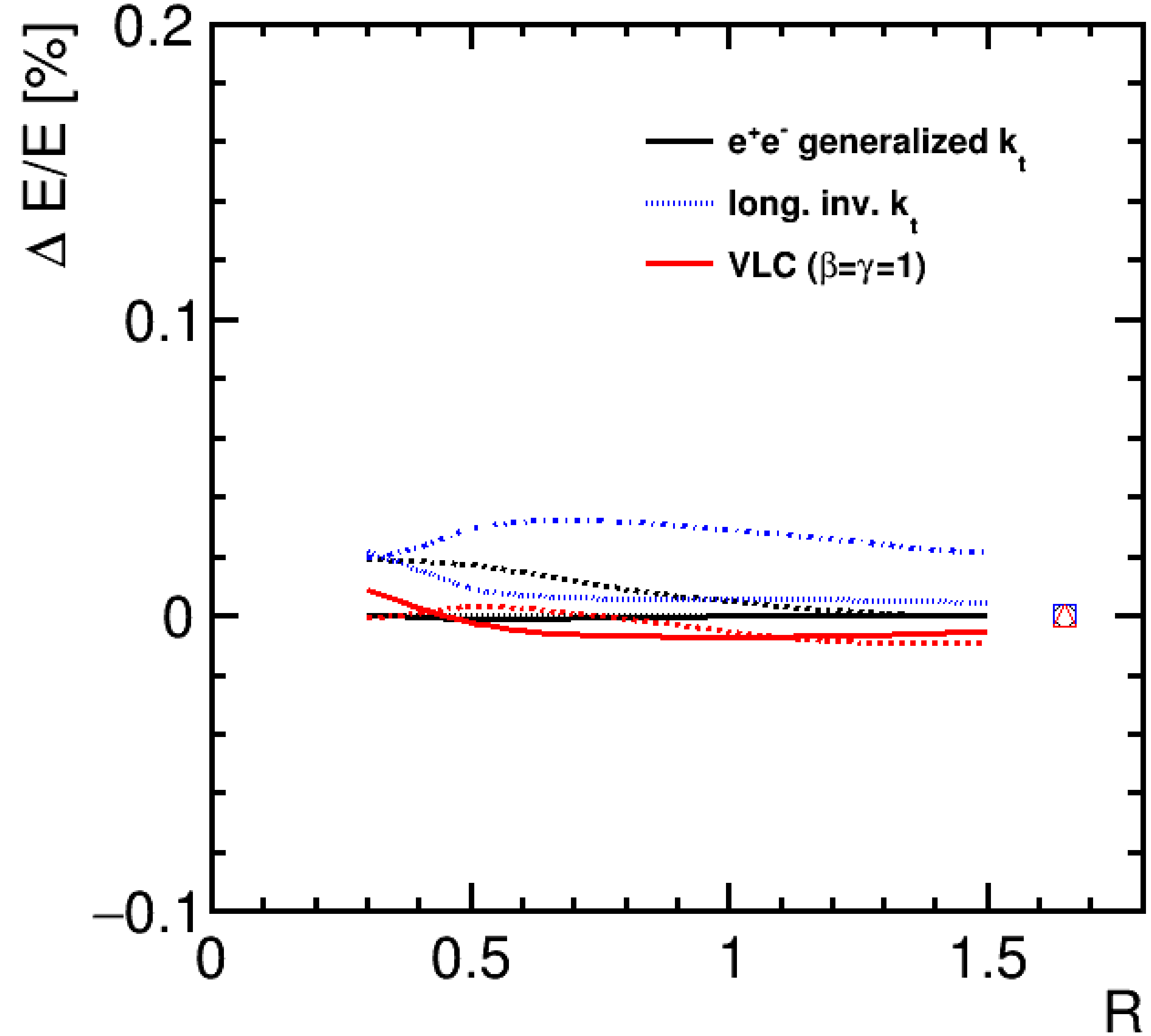
Marcel Vos (marcel.vos@ific.uv.es)

Non-perturbative corrections

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



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



mean

median

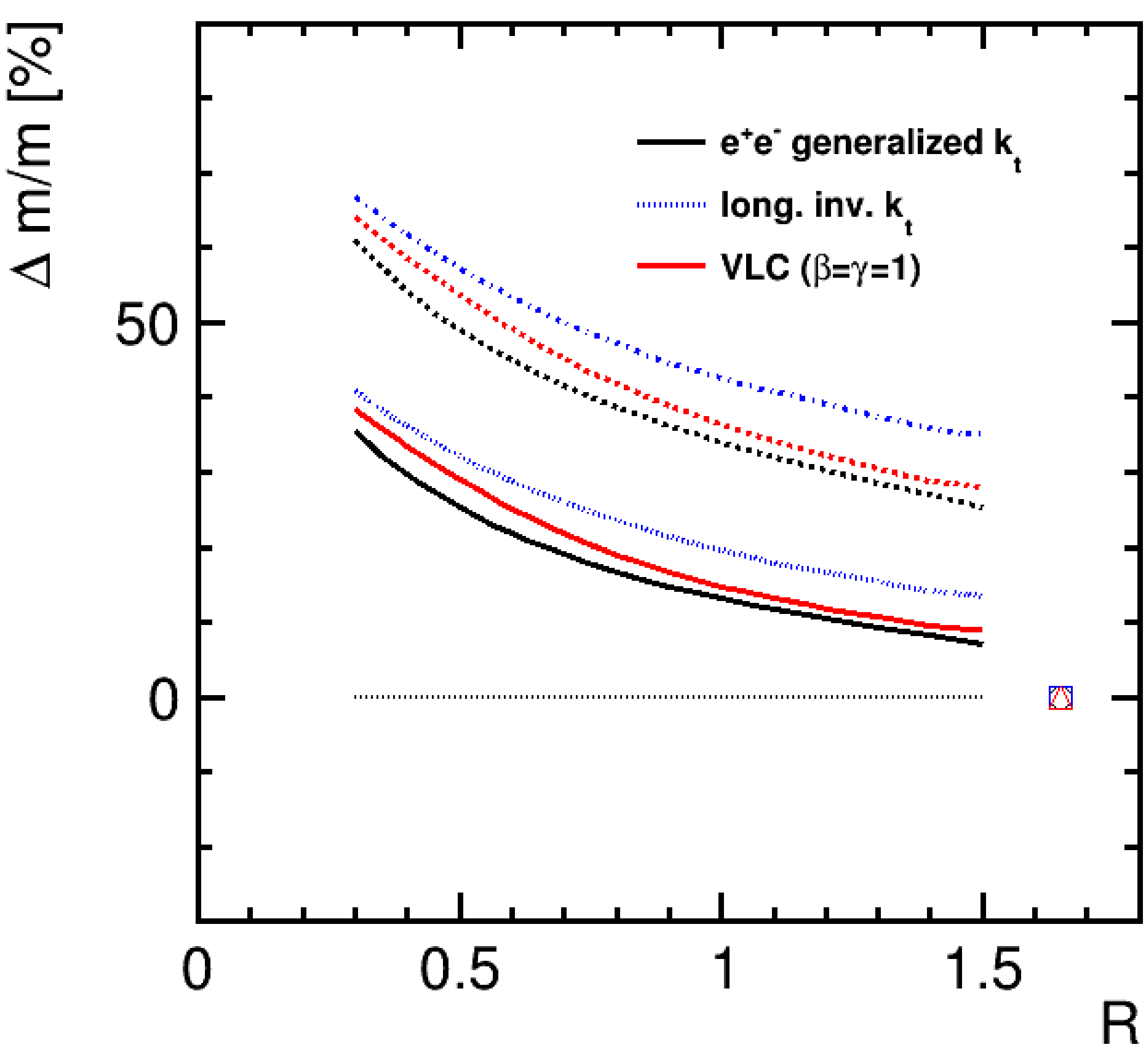
Algorithm with largest footprint has the smallest correction

VLC and long. invariant k_t no longer identical

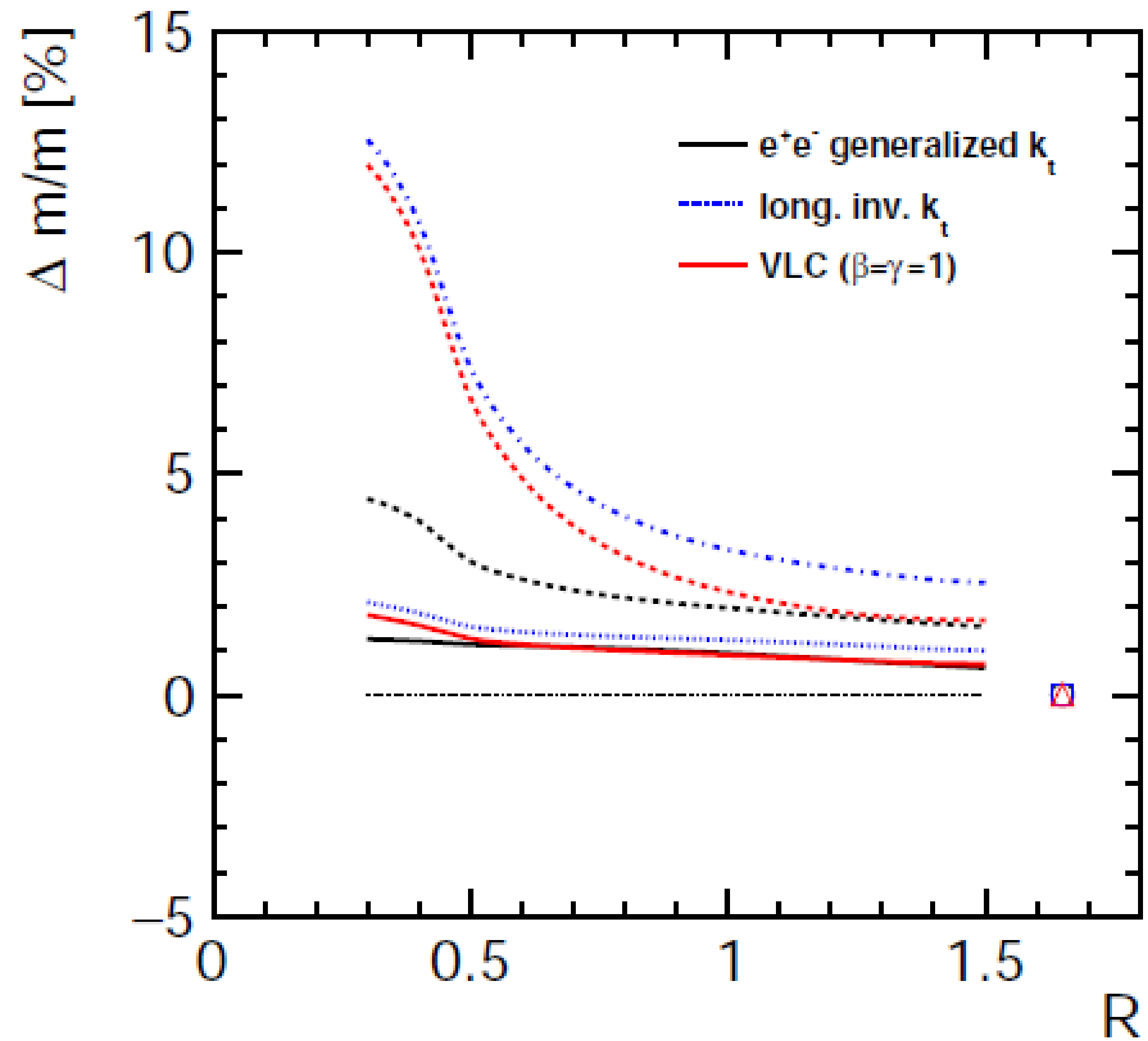
Few per mil effect at 250 GeV, 10^{-4} at 3 TeV

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^-

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

Challenges

Challenges are formidable:

- we cannot hide behind the calorimeters any more!
  (need to step up the effort on jet clustering)
- jet clustering limits physics output in many analyses
  (very little progress)
- some pressure on exclusive reconstruction scheme
  (but still unchallenged)
- background forces to introduce beam jets and control jet area
  (some new ideas, with non-trivial implications)

Need more effort to address this crucial issue