Jet reconstruction – challenges and opportunities

Marcel Vos

IFIC (UVEG/CSIC) Valencia

LCWS19, U. Tohoku, 29 october 2019

Jet reconstruction performance

A precise reconstruction of hadronic final states is crucial for the ILC

Reconstruction is affected by multiple issues:

- PF response; how well can we reconstruct single particle energy?

- neutrinos;   can we improve the b/c jet response?

- background; can we distinguish the hard scatter from $\gamma\gamma \rightarrow$ hadrons?

- clustering;   does the algorithm associate particles to the right jet?

The scientific return of most analyses is affected (to varying degrees) by all these sources of confusion . I will focus on the latter two.

Particle Flow

Particle flow with highly granular calorimeters offers "ultimate" single particle response
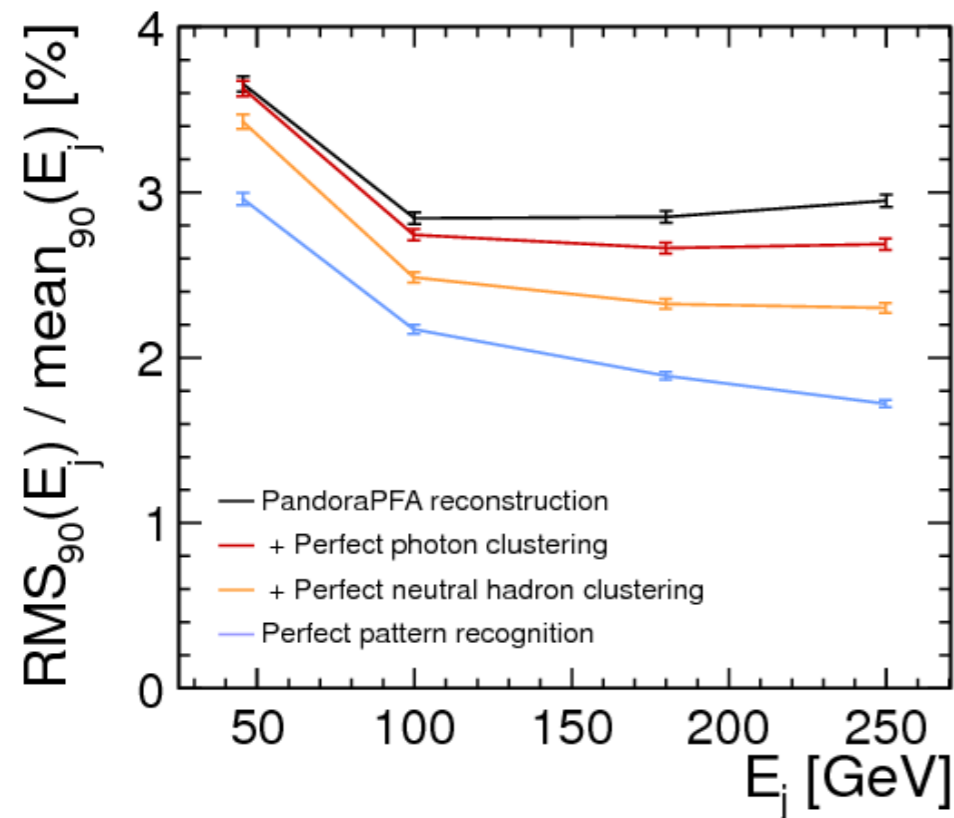
Di-jet events, energy resolution for "jets" inferred from total visible energy
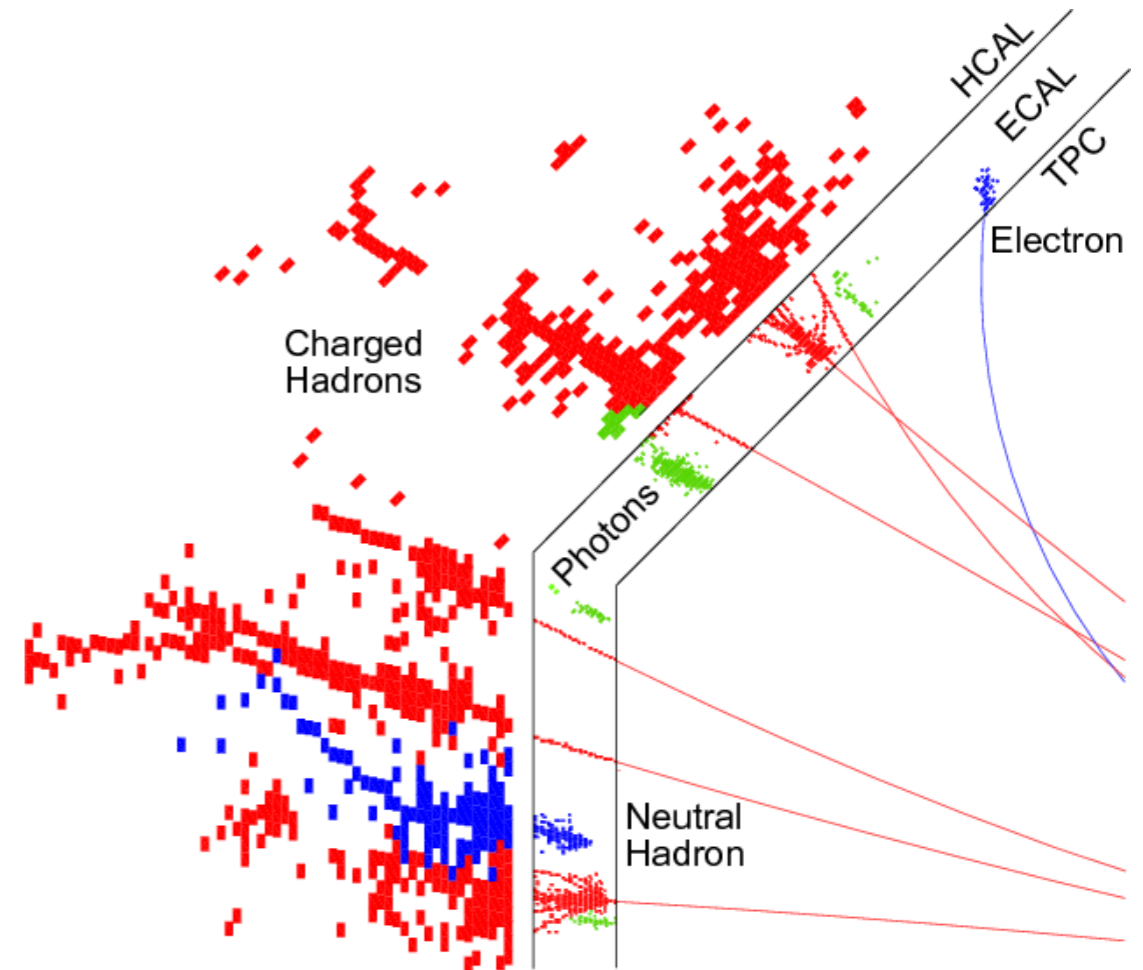
The jet energy resolution is excellent in very simple final states
→ in practice we're somewhat limited by confusion term at high energy: $\Delta E/E \sim 3\%$
Most analyses that we care about present a more complex situation

Jet clustering

Everyone uses sequential recombination algorithms

Standard approach at lepton colliders:

Exclusive* clustering with the k_t (Durham**) algorithm

Standard approach at the LHC:

Inclusive clustering with anti-kt and a small radius parameter

(*) Inclusive jet clustering with anti-k_t yields better resolution in some cases, but has not been shown to improve the overall performance of the analysis

(**) Background levels force to adapt lepton collider e+e- algorithms

Jet algorithms and jet area

The background energy that is clustered into the jet – and the effect on jet parameters - is proportional to the catchment area of the jet
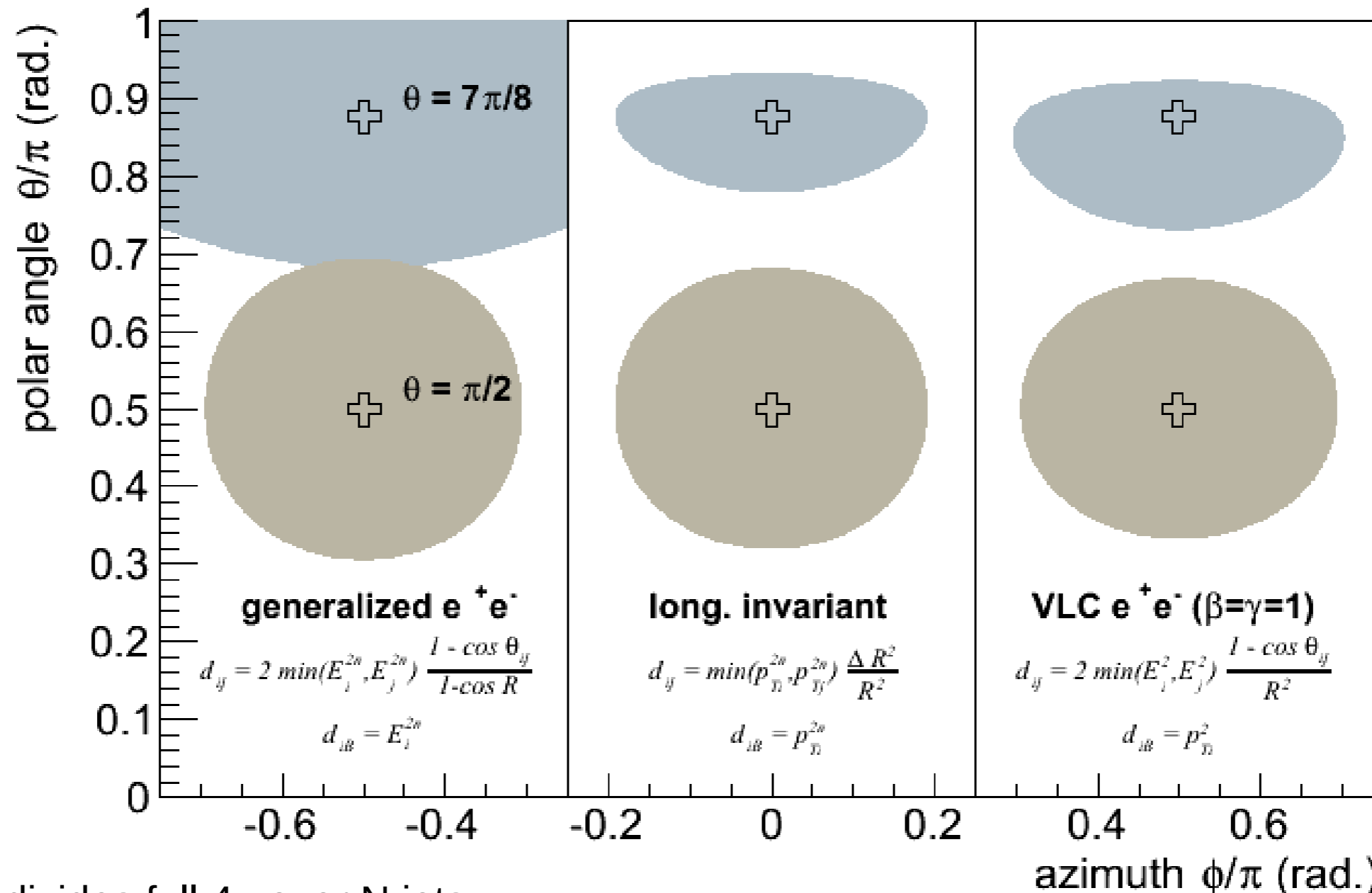
- Durham divides full 4π over N jets
- Algorithms with beam jets have a definite size – given by radius parameter R
- Algorithms with small footprint for forward jets (longitudinally invariant k_t, VLC) are robust

Jet algorithm space
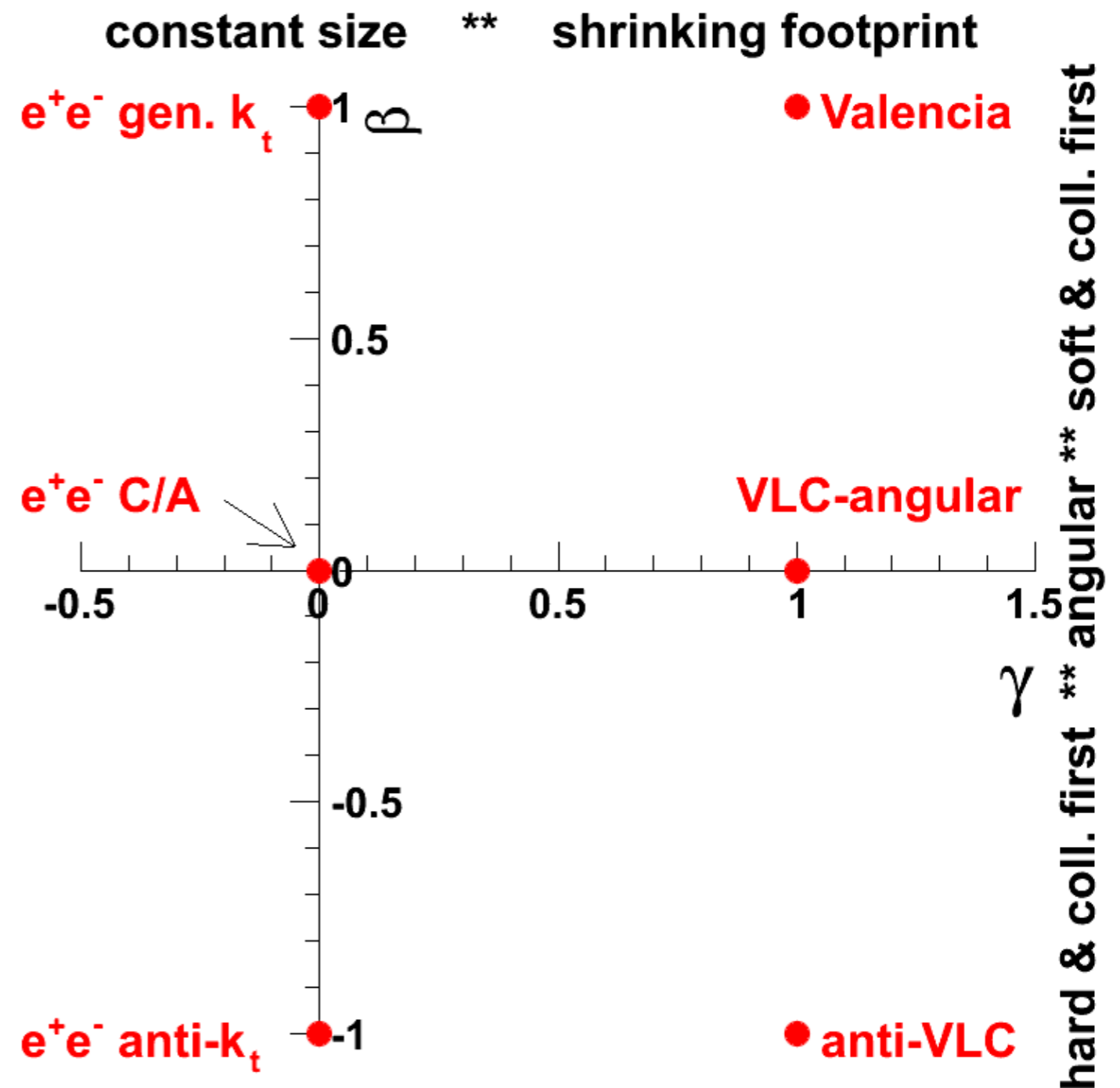
VLC algorithm of arXiv:1607.05039

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

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

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

Recover generalized e^+e^- kt for γ=0

Mimic robust longitudinally invariant algorithms with γ=1

constant size **** **shrinking footprint**

e^+e^- **gen. k**$_t$

Valencia

e^+e^- **C/A**

VLC-angular

e^+e^- **anti-k**$_t$

anti-VLC

hard & coll. first ** angular ** soft & coll. first

γ

β

Check out fjcontrib 1.040 or later if you're using FastJet. Thanks to F. Zarnecki for check fastjet and LCFI codes.

Jet grooming algorithms

Grooming techniques remove soft contamination from the jet so as to improve the jet substructure resolution.

Groomed jets have reduced effective area (see arXiv:1803.06991) and hence improve the resilience against pile-up and underlying event

Grooming is part of the standard procedure for large-R jets at the LHC and is used in CLIC boosted top reconstruction (arXiv:1807.02441)

Soft drop algorithm (Larkoski, Marzani, Soyez, Thaler, arXiv:1402.2657)

$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$

Large-R jet is decomposed and softer constituent removed

More amenable to calculations than first generation of algorithms (trimming, pruning, etc.)
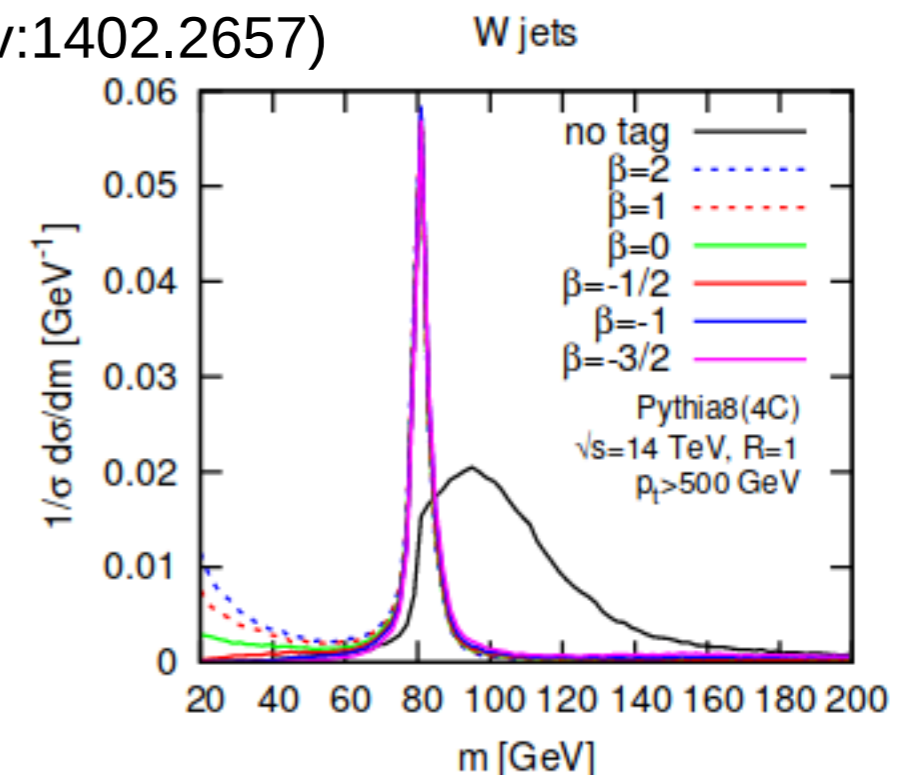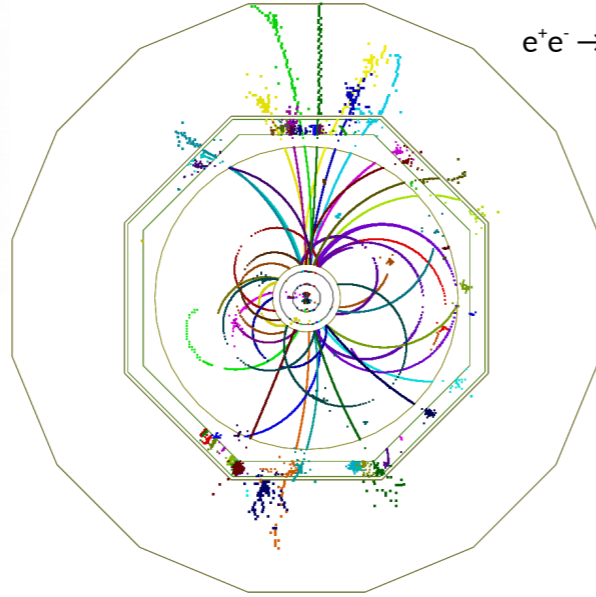
Multi-jet final states

$e^+e^- \rightarrow Zh, Z \rightarrow \mu^+\mu^-, h \rightarrow b\bar{b}$

High-energy linear colliders
- starting with 250 GeV -
open up a can of worms
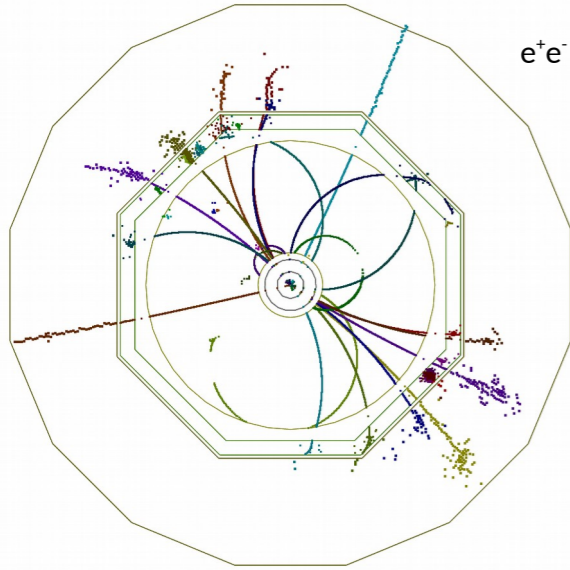
$e^+e^- \rightarrow Zh, Z \rightarrow q\bar{q}, h \rightarrow b\bar{b}$
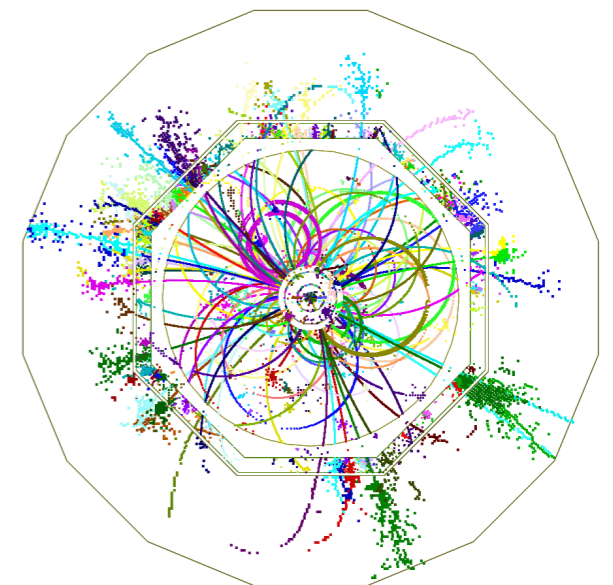
$e^+e^- \rightarrow t\bar{t}$, fully hadronic, 500 GeV
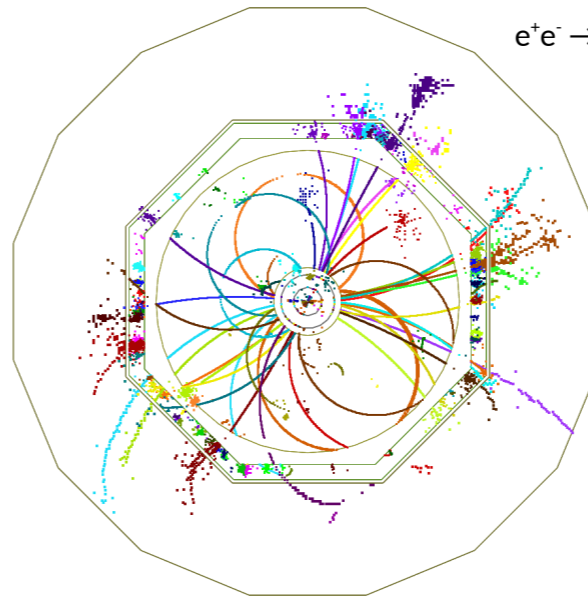
Two-jet topologies are easy
Four-jet topologies, not quite so easy
Six- and 8-jet topologies are ~ impossible

$e^+e^- \rightarrow t\bar{t}h, h \rightarrow b\bar{b}$, 8 jets, > 550 GeV

Complex final states

In complex final states jet clustering limits the performance

In multi-jet final states with multiple scales k_t will sometimes give the wrong answer

The impact on the mass resolution can be very sizable

CLIC mass resolution for Higgs boson candidates ~22% in di-Higgs boson production at 3 TeV
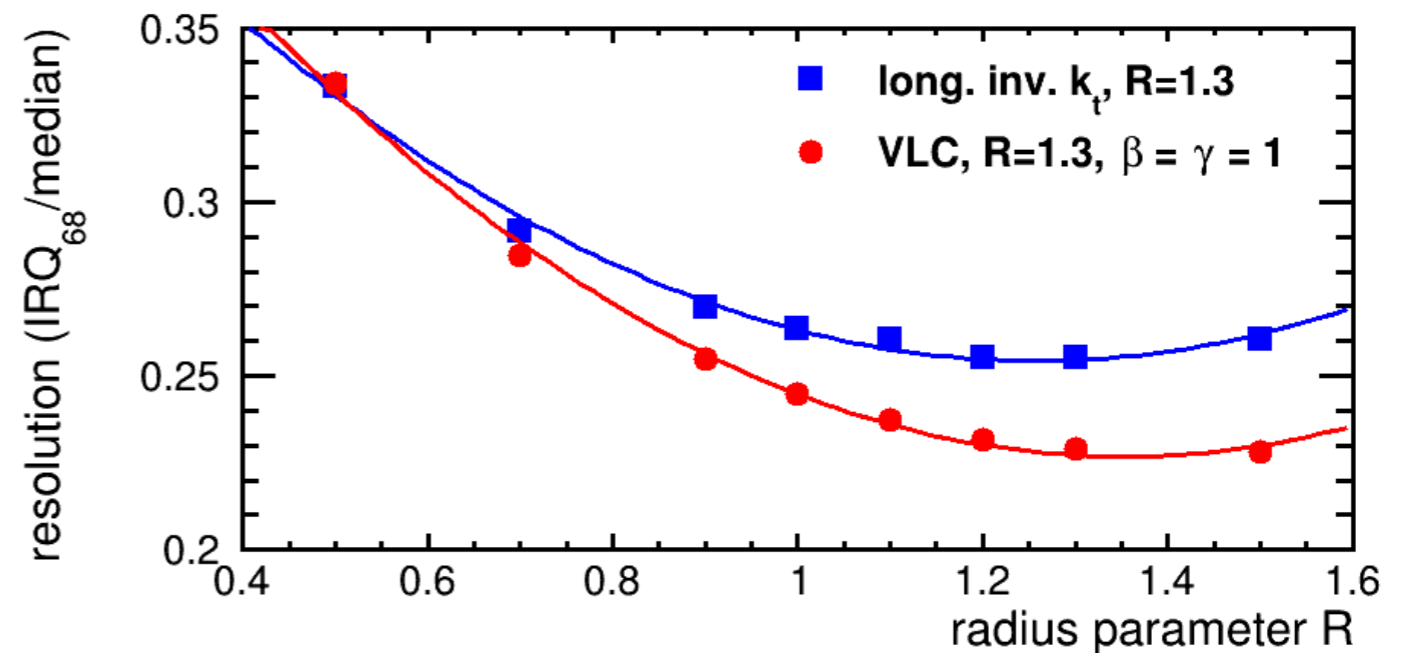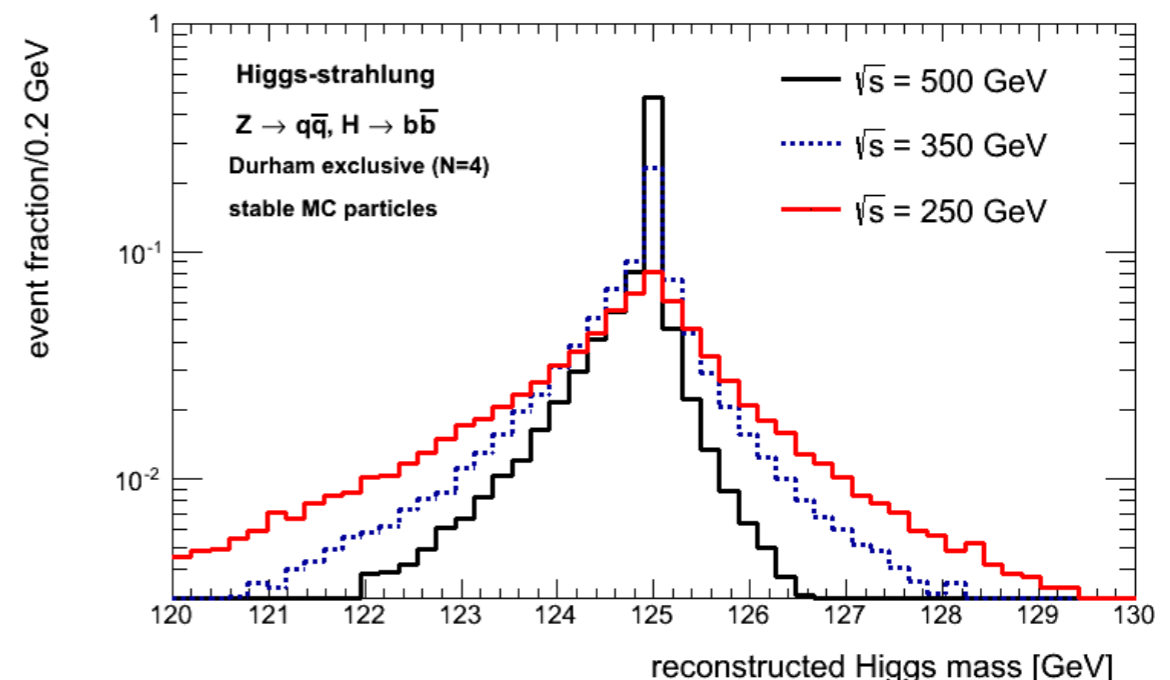

Notorious examples:

tt, t → cH (Zarnecki)
ttH (Price & Strube)
ZHH (J. Tian, M. Weber)
WW/ZZ (J. Beyer)
H → invisible (Y. Kato)

Particle-level jet reconstruction in ZH production: tails in reconstructed energy entirely due to "confusion" in clustering

https://agenda.linearcollider.org/event/7760/contributions/40910/attachments/32767/49845/JetRec_ILD2018.pdf
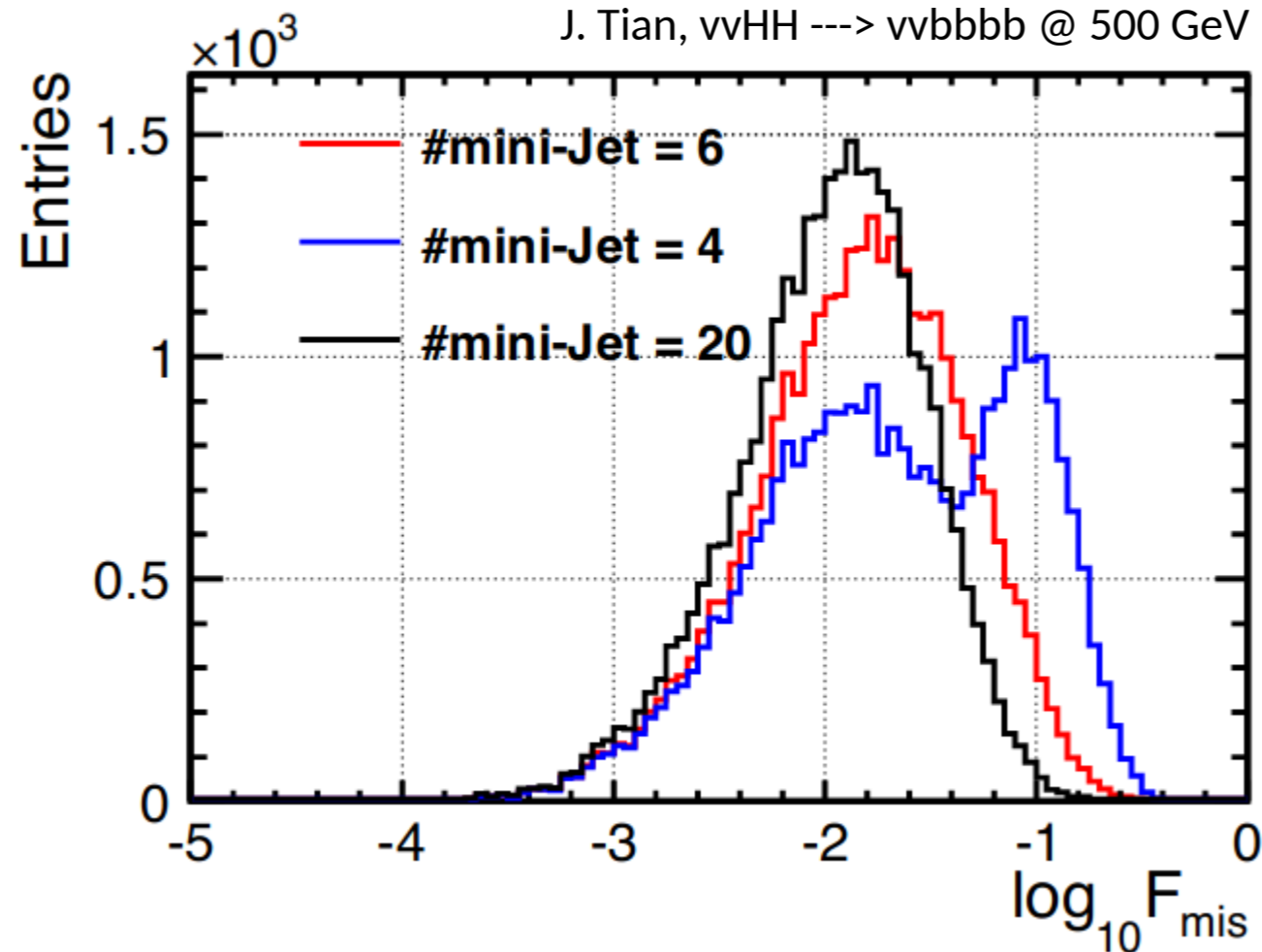
Complex final states

In multi-jet final states with multiple scales k_t will sometimes give the wrong answer

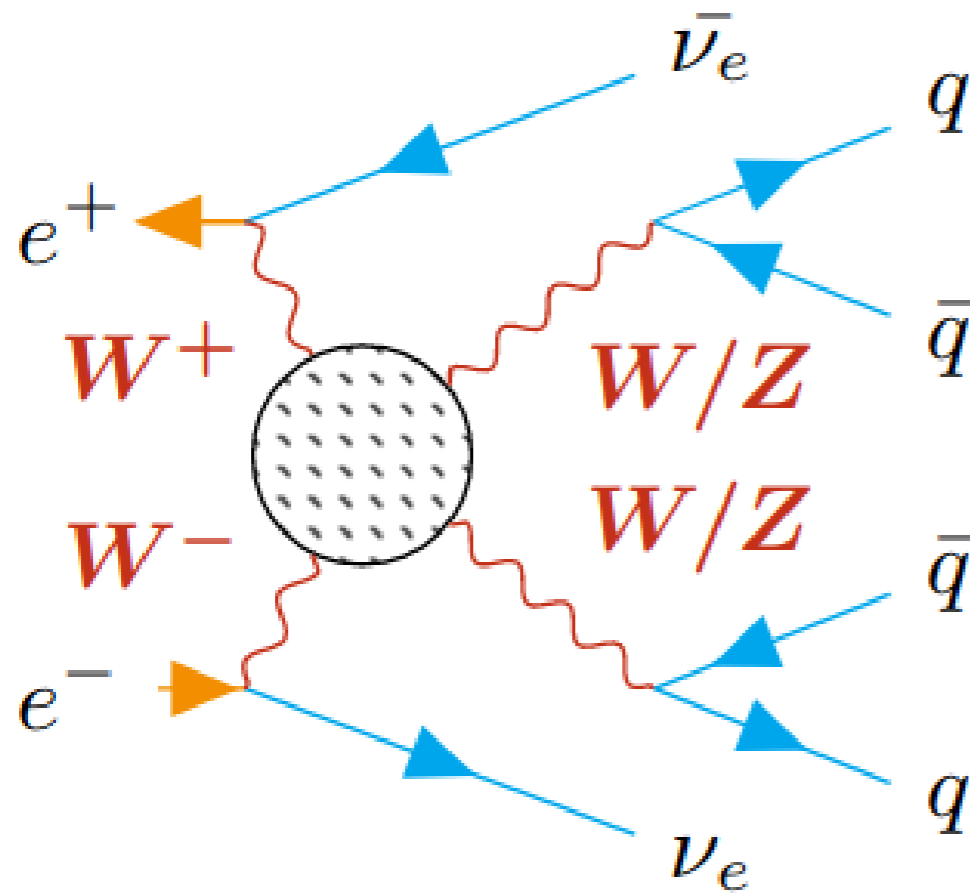Can quantify this by tracking stable particles back to the colour singlet using MC information
F_{miss} = fraction of wrongly associated energy

Often the problem originates in the last clustering steps (a hard gluon emitted from a more energetic singlet occupies one of the jets and then forces a wrong merger down the line)

J. Tian, vvHH ---> vvbbbb @ 500 GeV

WW/ZZ at 1 TeV

ILD benchmark study of WW/ZZ at 1 TeV
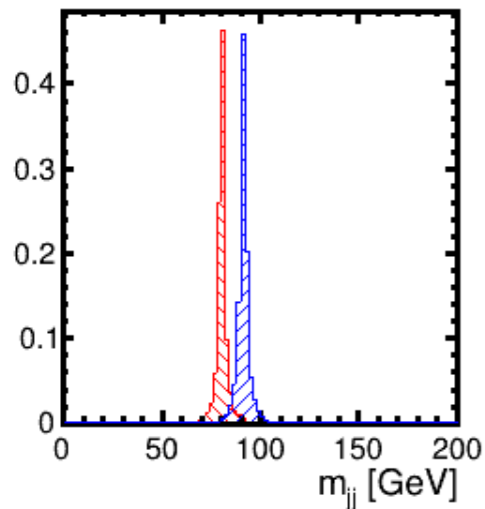
Jakob Beyer (DESY)

Full ILD simulation with Pandora PFA
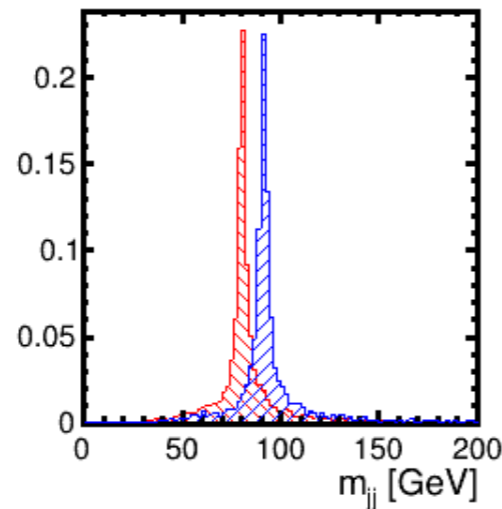
MC truth selection to isolate pure WW and ZZ samples
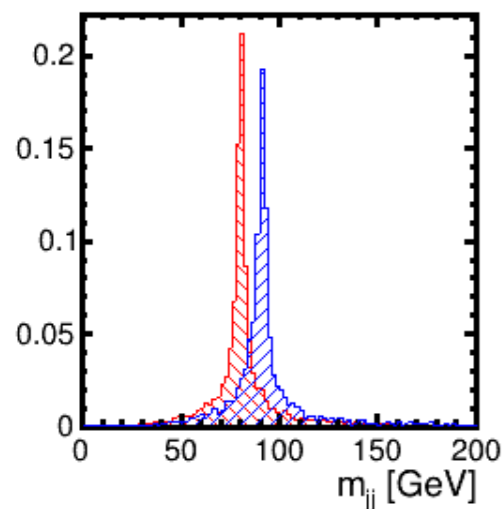
Jet clustering with Durham, exclusive N=4
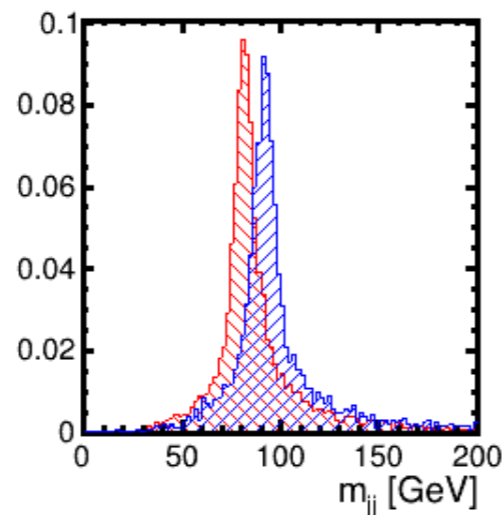
mass separation

True W and Z mass

Stable signal MC
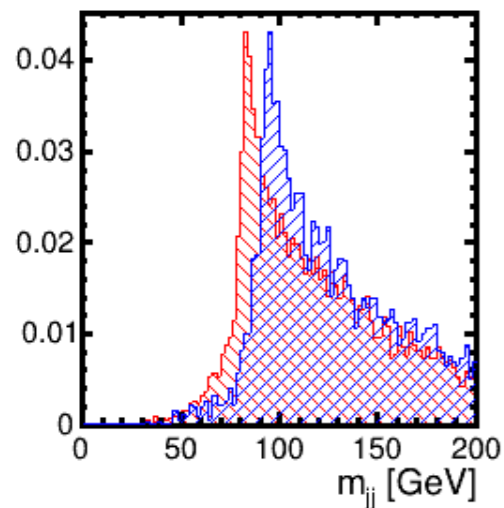particles (including ν)

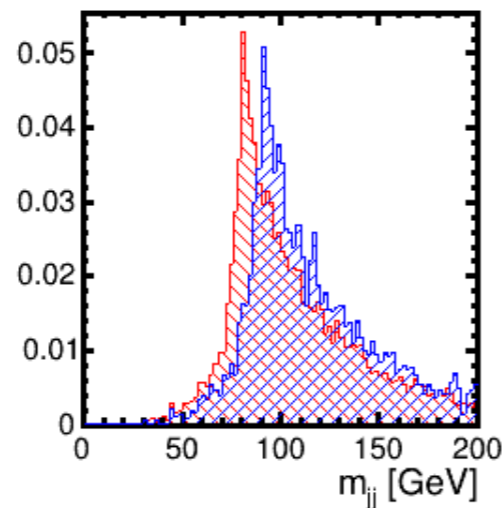Stable signal MC
particles (excluding ν)

PFOs dominated by
signal particles

Stable MC particles
(including background)

PFO – final result
(including background)

Clustering leads to tails, but cores still narrow

Neutrinos affect response for bottom and charm jets
Calibrate? Tag semi-leptonic decays? Exclusive
neutrino reconstruction? → Jakob Beyer

Detector response broadens cores

Background adds very pronounced tail (for Durham)

WW/ZZ at 1 TeV: ROC curves

True W and Z mass

Stable signal MC
particles (including ν)

Stable signal MC
particles (excluding ν)

PFOs dominated by
signal particles

Stable MC particles
(including background)

PFO – final result
(including background)

**Receiver-Operator-Curves (true positive vs. false positive)**
ROC curves provide a single figure-of-merit to quantify performance: Area Under Curve = 0.5 (random) - 1 (perfect)

- red/blue lines: integrate distributions of slide 13
- grey reference lines: Gaussian JER 2,4,6,8,10%
- dashed reference: Gaussian JER fitted to distribution

Clustering leads to tails, but cores still narrow
AUC ~ 0.78

Particle Flow objects broaden cores
AUC ~ 0.69

Background adds very pronounced tail
AUC ~ 0.58
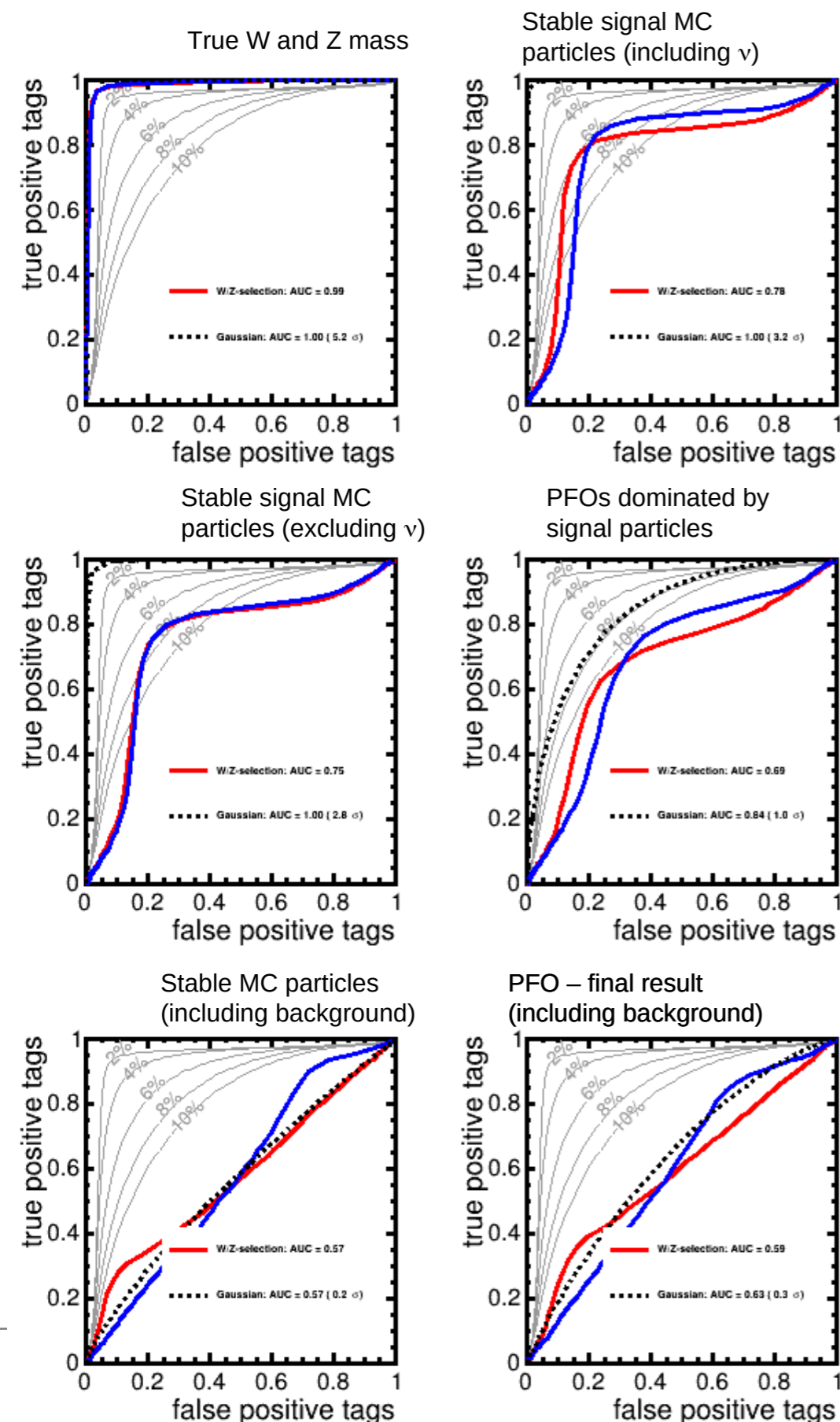
WW/ZZ at 1 TeV: jet algorithms

	Durham	Longitudinally invariant k_t R=1.4	VLC R=1.4	Durham on kt exlusive N=6	VLC R=1.4 with SoftDrop
Clustering	0.78	0.79	0.78	0.78	0.78
PFOs	0.69	0.73	0.72	0.72	0.72
background	0.59	0.70	0.72	0.70	0.71

Robust algorithms yield big jump in performance with background
        --- problem solved ----

Clustering essentially identical for all k_t algorithms

    --- a real limitation of kt distance ----

Addressing the challenge:
    - find better clustering algorithm (Masakazu Kurata)
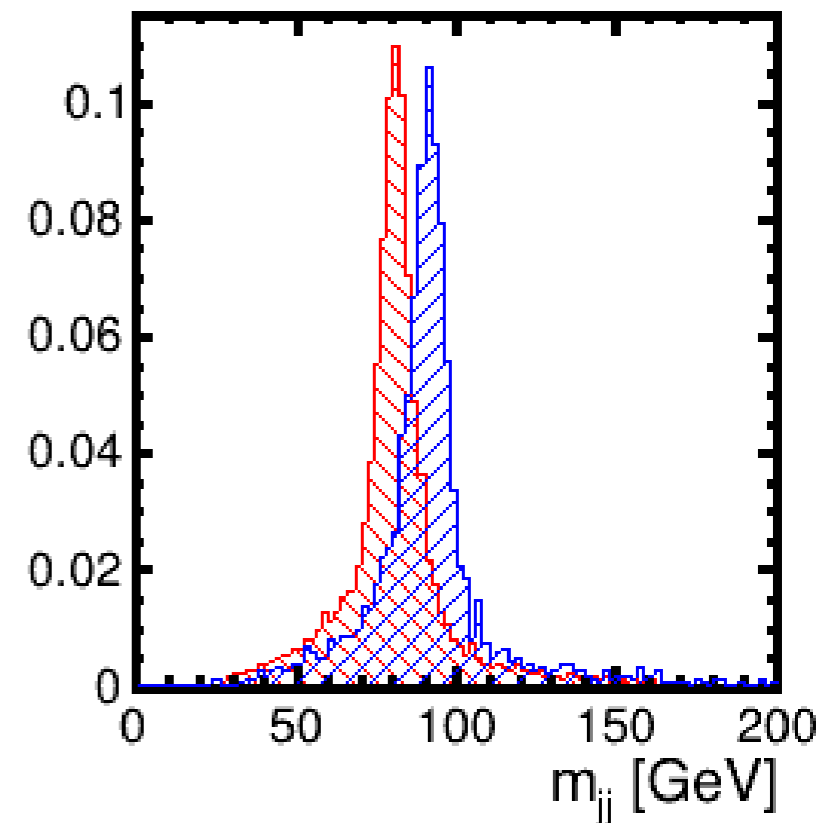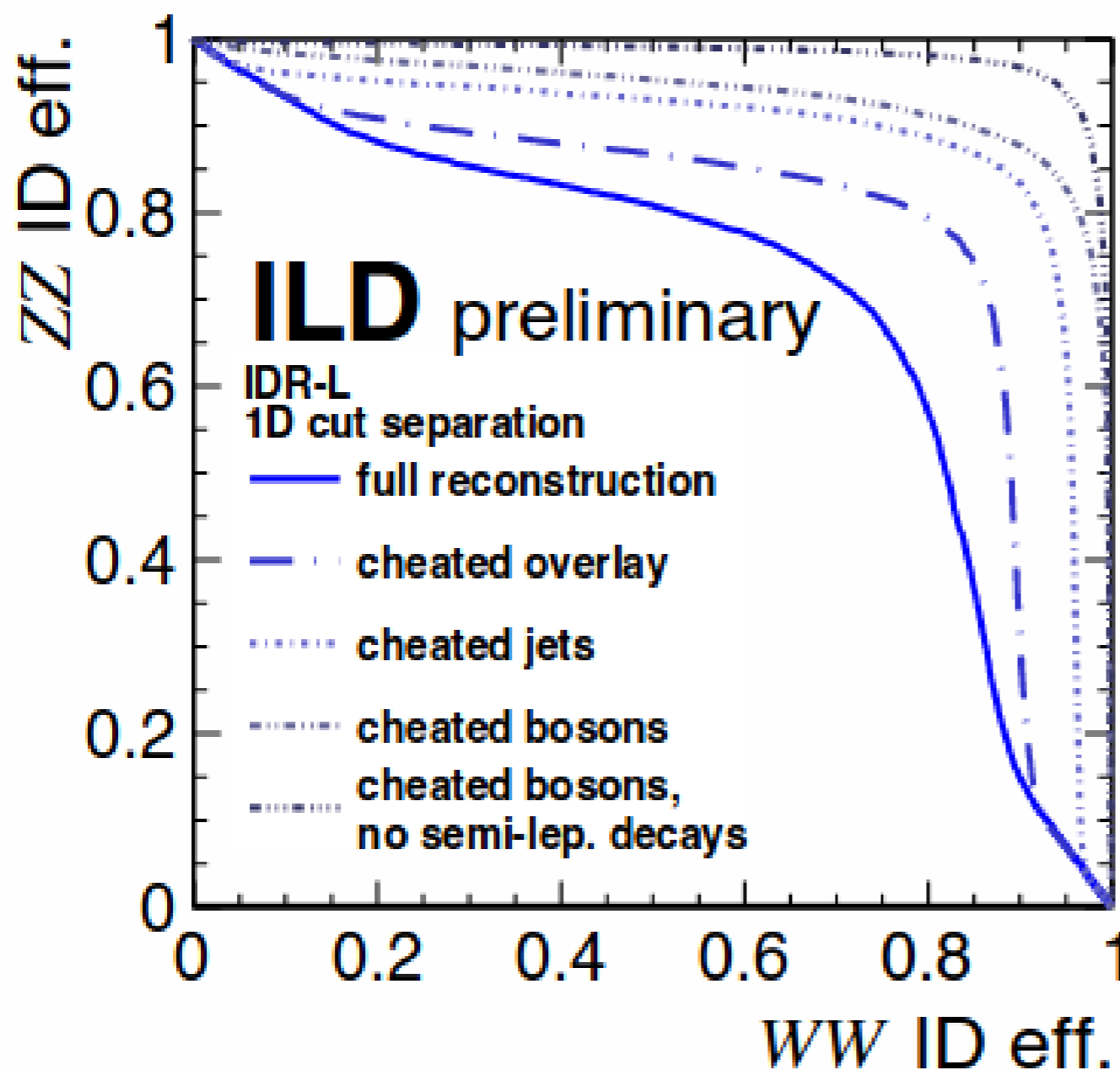   - use high-level information to fix things (Shogo Kajiwara)

Figure-of-merit:ROC curves

ILD preliminary

IDR-L
1D cut separation

— full reconstruction
—·— cheated overlay
······ cheated jets
—··— cheated bosons
— — cheated bosons, no semi-lep. decays

WW ID eff.

ZZ ID eff.

Jakob Beyer, Jenny List, ILD-PHYS-PUB-2019-005

TAKE-HOME MESSAGES:

Relate analysis result to jet reconstruction performance, breaking down into different sources of confusion

Receiver-Operator-Curve (ROC) offers complete specification of performance

Area-Under-Curve (AUC) offers simple figure of merit

Jet clustering

A CEPC paper by Yongfeng Zhu and Manqi Ruan confirms the impact of jet clustering on the WW/ZZ separation

They separate the poorly clustered events from the well-reconstructed events using "truth" information
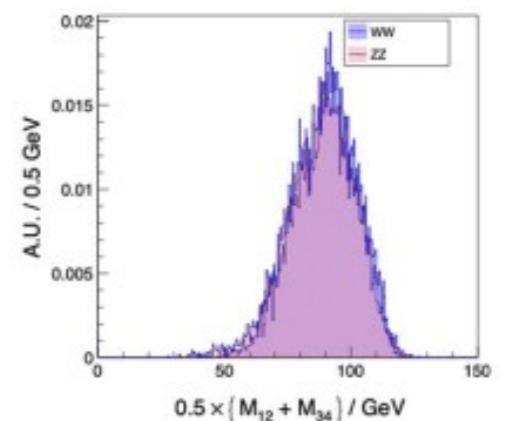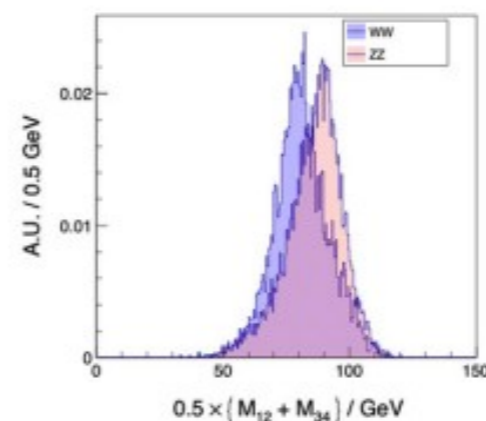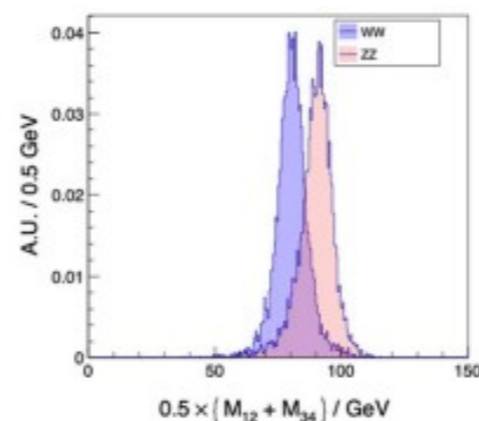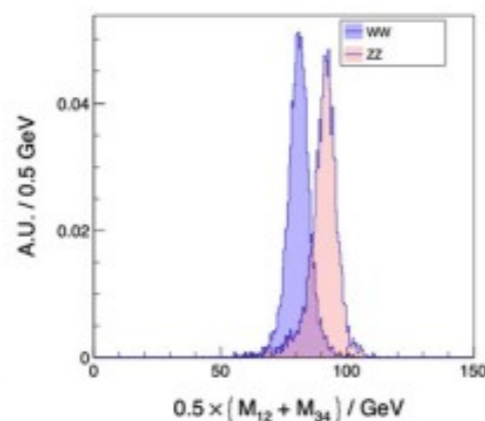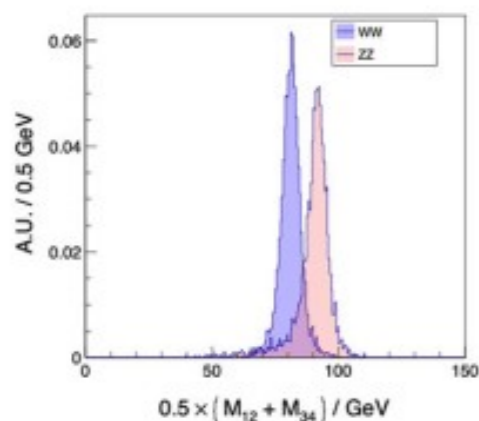
Identify badly clustered events from the clustering history?
- use size of d_{nn+1} estimators to identify marginal decisions
- group "good" and "bad" events in separate categories
- teach a machine to
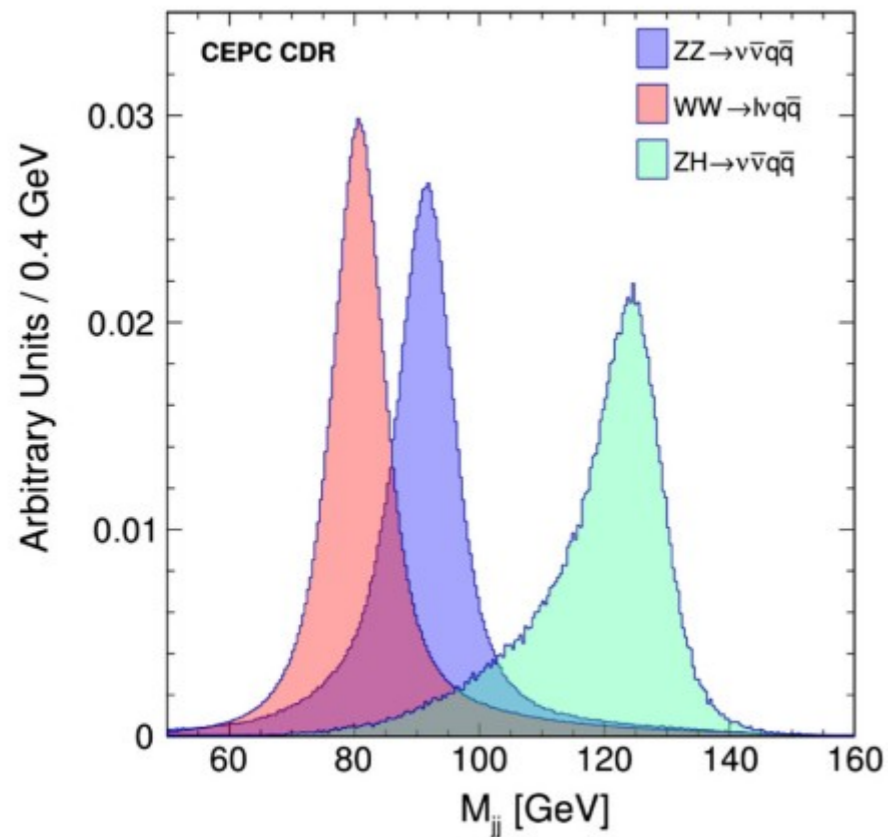
← better clustering   worse clustering →

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

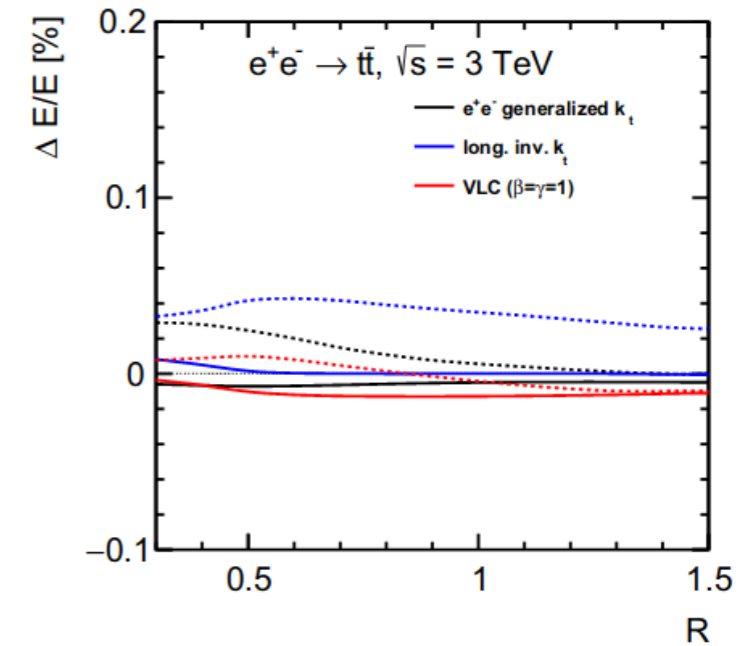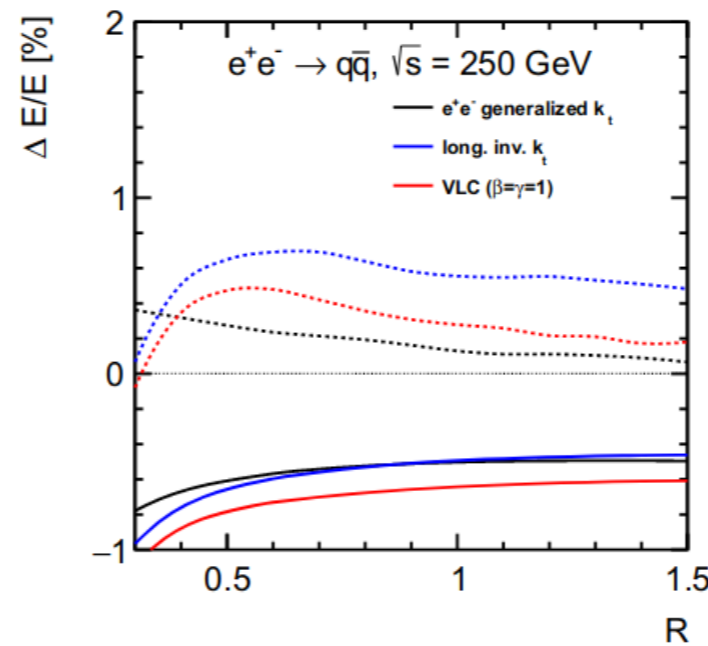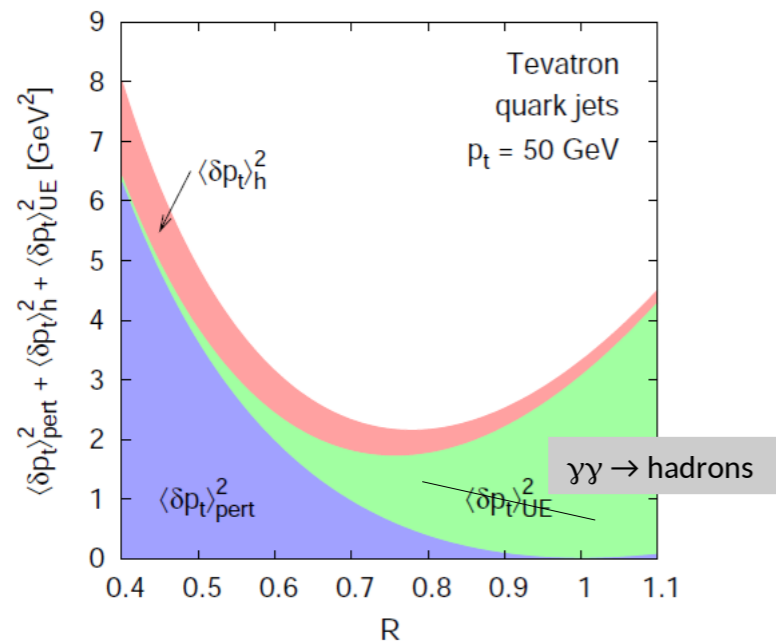# (non-) perturbative corrections

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

EPJC 78 (2018) 2 144, arXiv:1607.05039

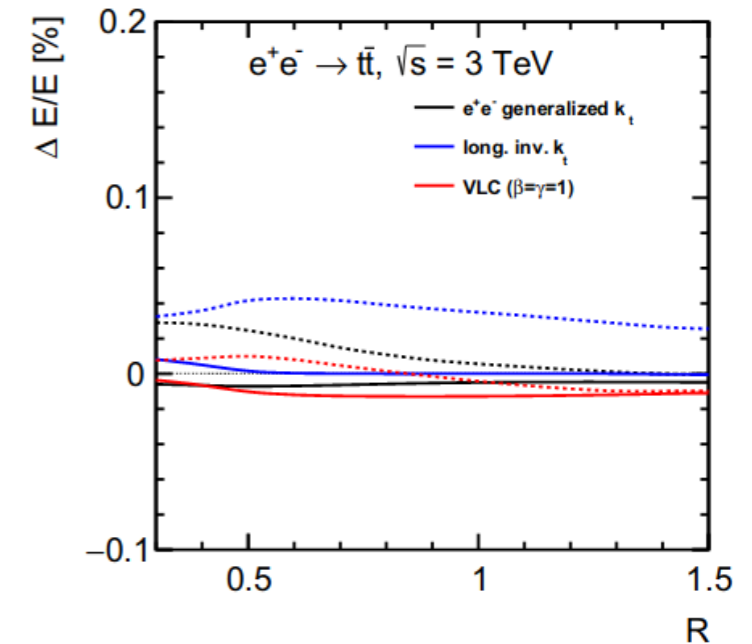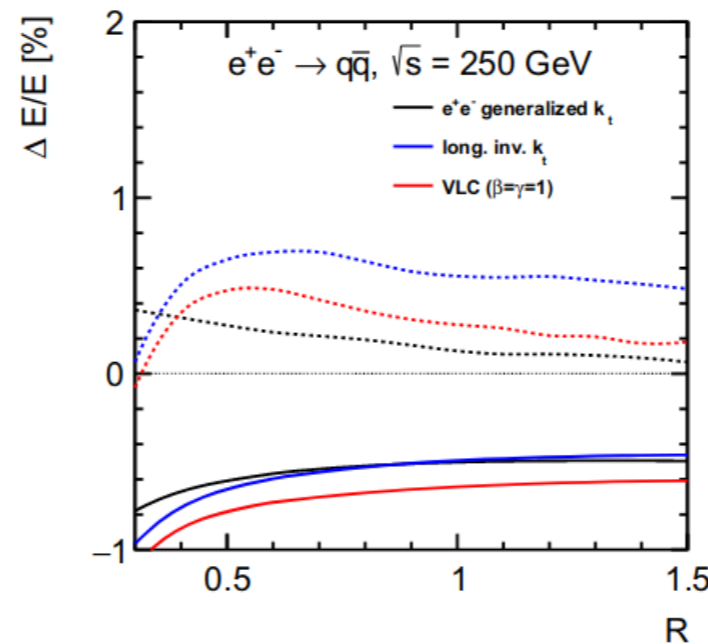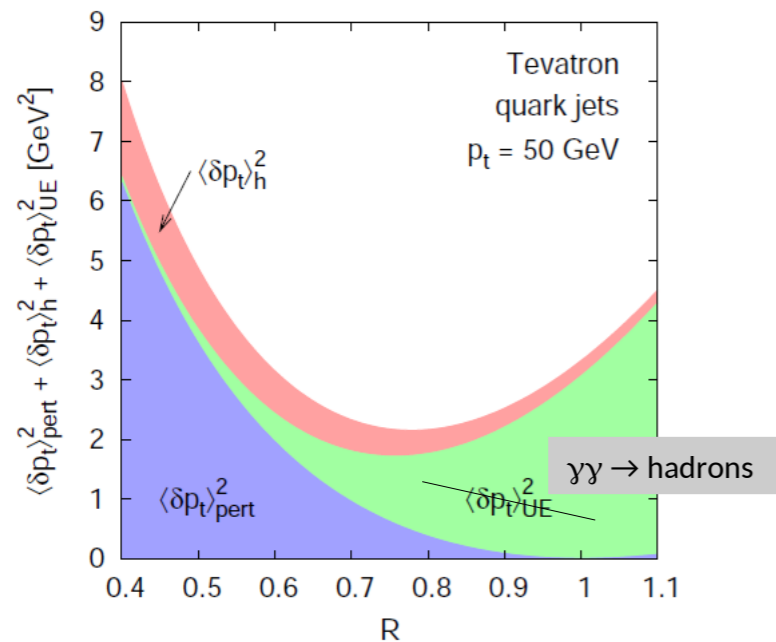# (non-) perturbative corrections

Uncertainties in jet response are an important source of systematics

Jet area and footprint determine affect uncertainty in energy response:

- background contribution scales with R^2
- (non-) perturbative corrections decrease with increasing R and $\sqrt{s}$
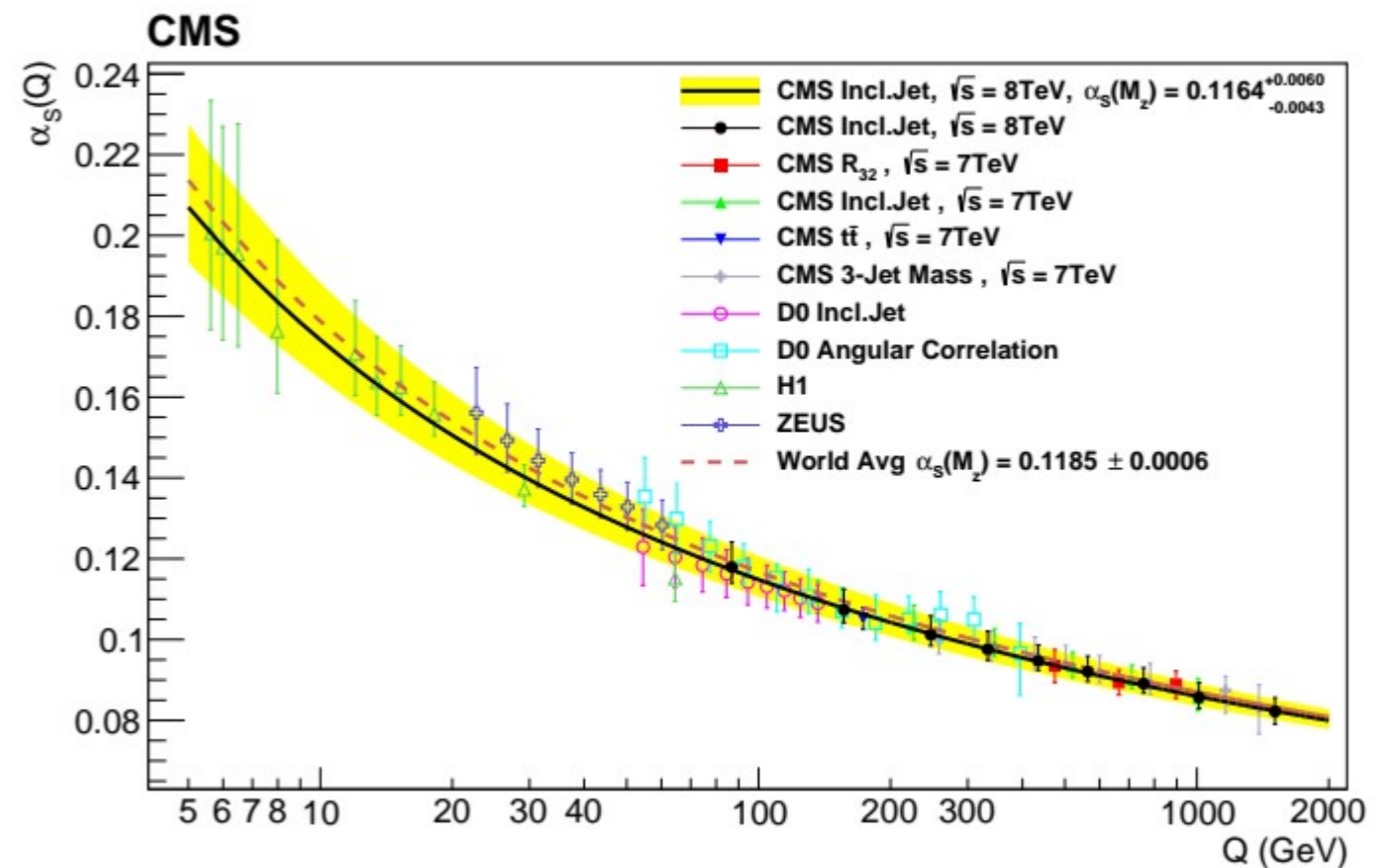
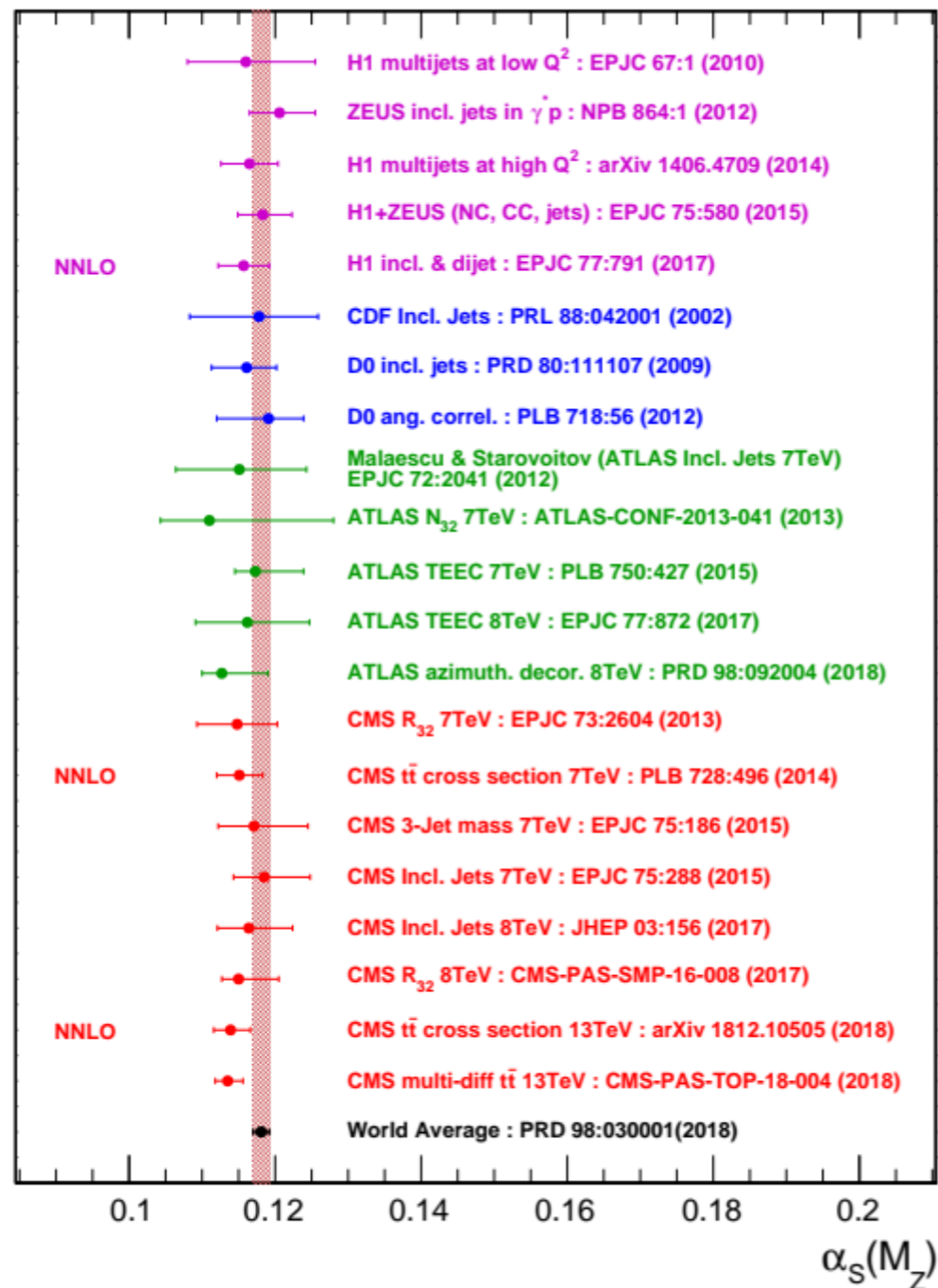Dasgupta, Magnea, Salam, JHEP0802 (2008) 055

EPJC 78 (2018) 2 144, arXiv:1607.05039

Non-perturbative corrections: note the order of magnitude decrease in

non-perturbative corrections and α_s

Energy frontier colliders do not provide competitive measurements of the strong coupling constant a low scale

Still: reference is $\alpha_s (m_Z)$

Opportunity in LC: high-scale α_s

QCD starts to "feel" new, massive coloured states once the energy is high enough. A precise measurement at the highest energy yields stringent and quite model-independent bounds.
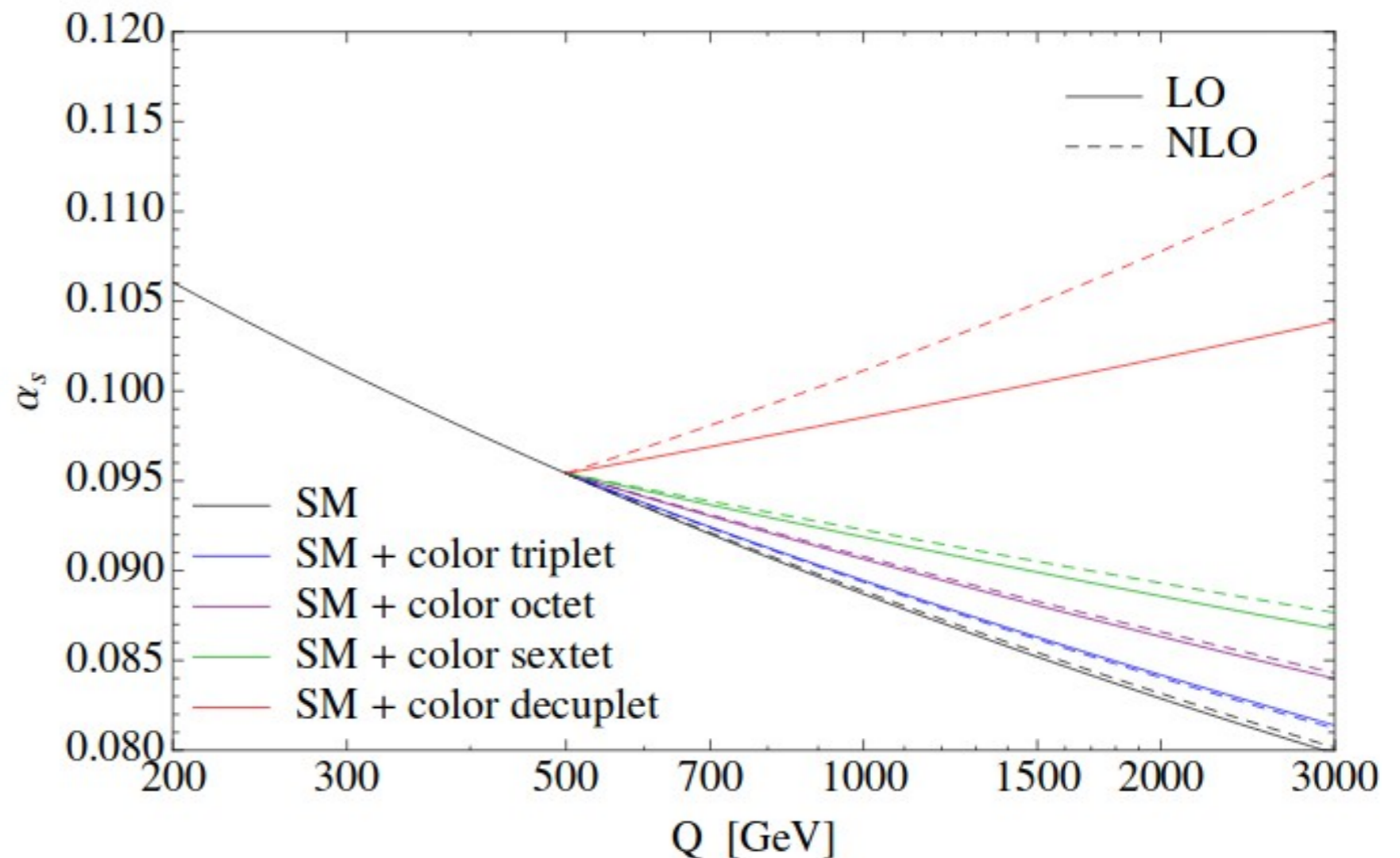
Figure from Becciolini et al., arXiv:1403.0741

See also, Berger et al, 2004

TESLA QCD, hep-ph/0308094

Summary

Jet clustering performance is key for the success of a linear collider

It is important that we understand which effects limit a given analysis – particle response, neutrinos and beam energy spread, background, and jet clustering, as each requires a different solution

Precision measurements may be limited by systematic uncertainties in modelling of non-perturbative (hadronization) corrections – we need to develop a method to estimate their size (and reduce it)