Jet reconstruction – challenges

Marcel Vos
IFIC (UVEG/CSIC) Valencia

ILD analysis meeting, 16/01/2019

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

Jet reconstruction performance

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

Reconstruction is affected by several issues:

- PF response; how well can we reconstruct single particle energy?
- background; can we distinguish the hard scatter from pile-up?
- clustering; does the algorithm associate particles to the right jet?

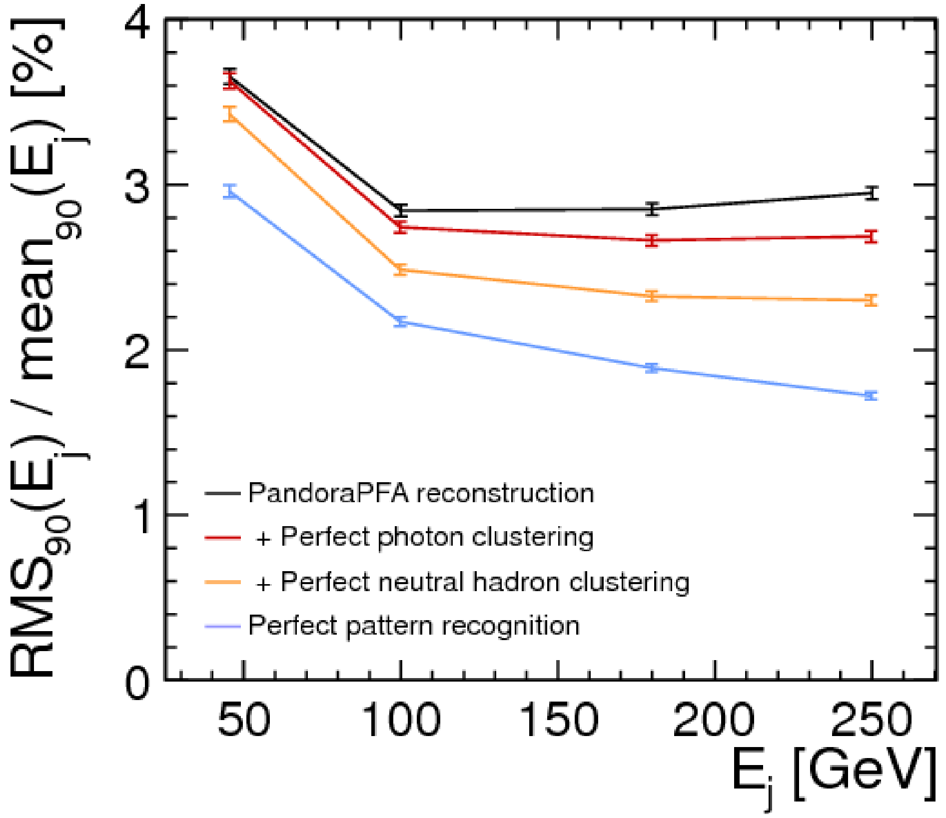
The final result of most analyses is affected (to varying degrees) by all these three sources of confusion

Particle Flow

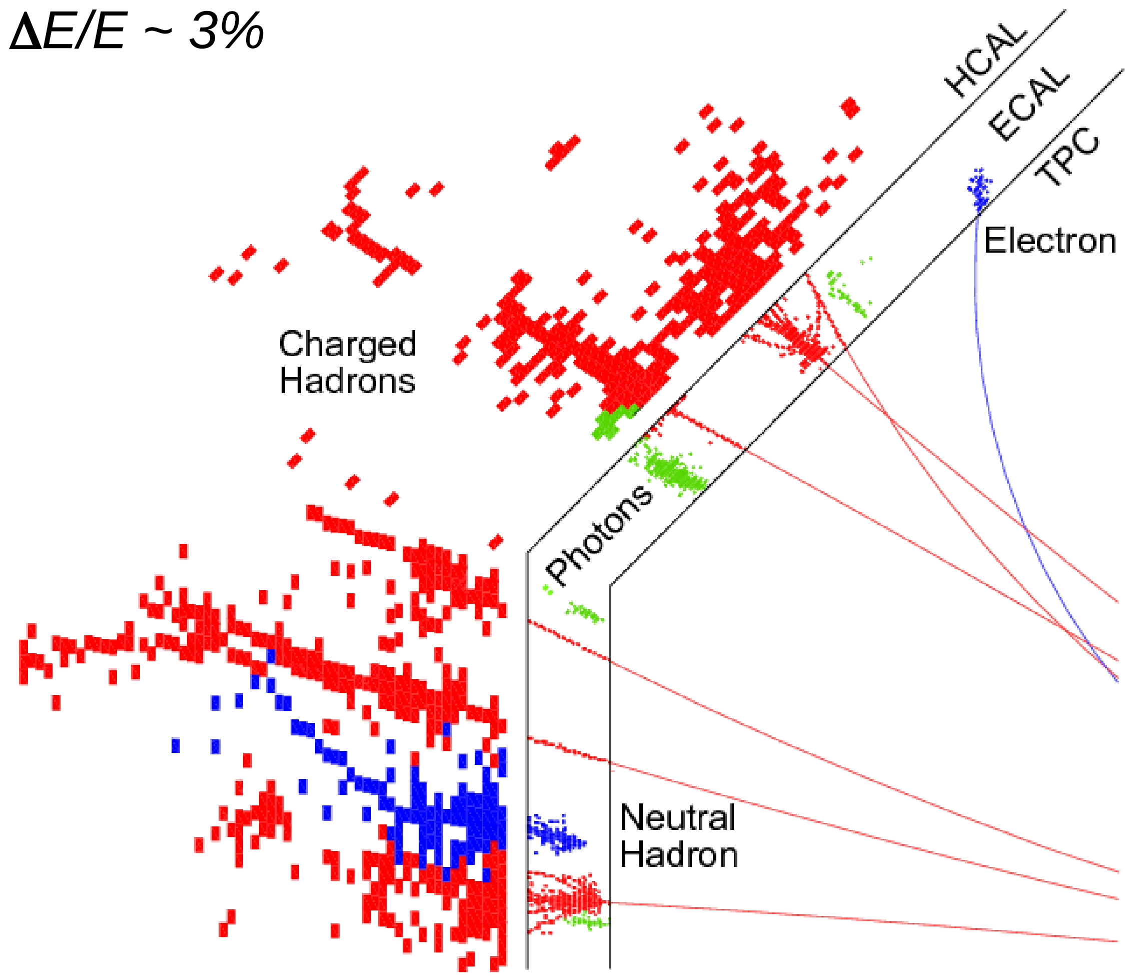
See presentations by Niall Thomas, Graham Wilson, Remi Eté,Matias Weber

Particle flow offers “ultimate” single particle response

→ in practice we’re somewhat limited by confusion term: *ΔE/E ~ 3%*



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



The jet energy resolution is measured on very simple final states, to minimize the effect of jet clustering and thus show off our PFA detectors

That’s a legitimate approach – we need to benchmark the detectors and develop the PFA – but most analyses present a more complex situation

Background and jet area

Background processes such as $\gamma\gamma \rightarrow$ hadrons produce additional energy flow

→ must be corrected for comparison to theory

→ sharply peaked in forward and backward directions

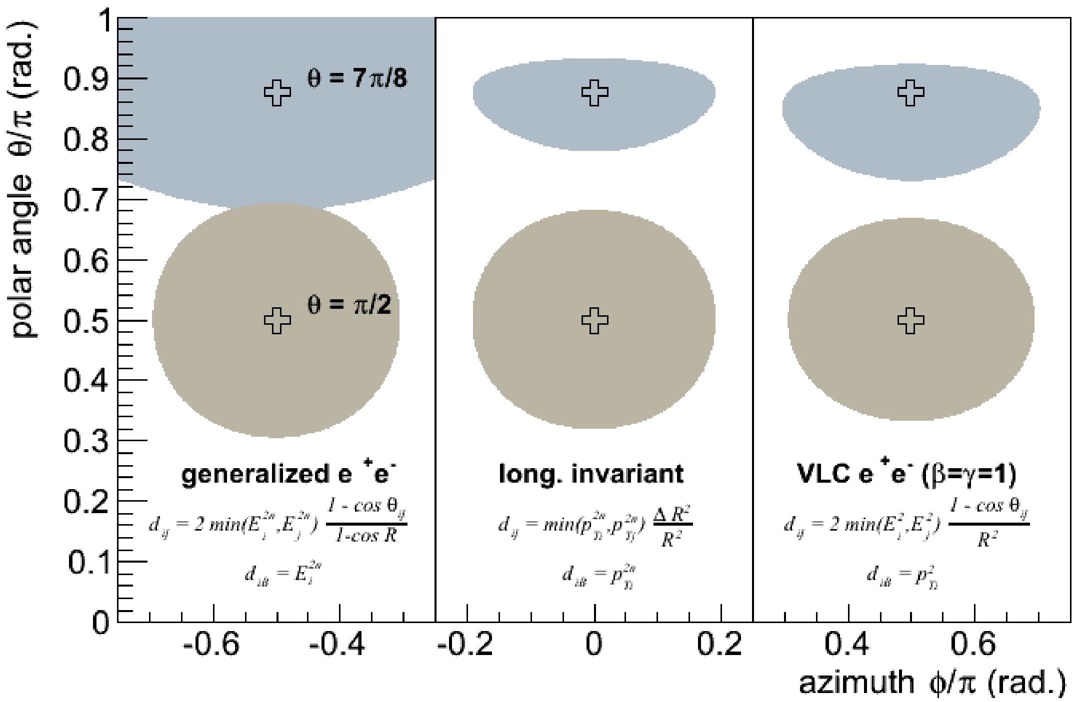
The background energy that is clustered into the jet – and the effect on jet parameters - is proportional to the catchment area of the jet

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

Groomed jets have reduced effective area (see arXiv:1803.06991)



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...

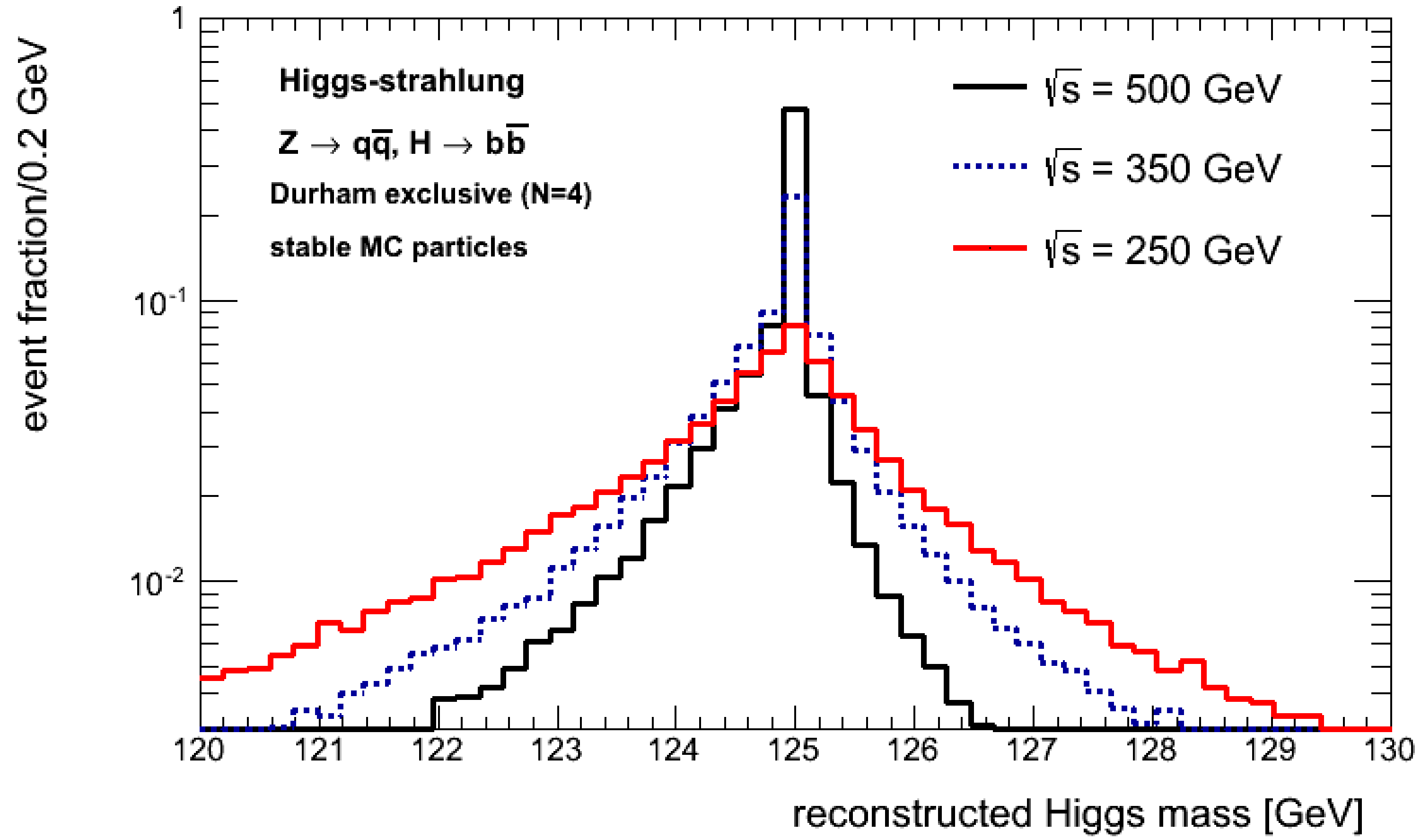
Limitations of jet algorithms...

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

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

Notorious examples:

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



Jet performance studies

Study the jet reconstruction performance in a number of benchmark channels

Separate the effect of the different sources of “confusion”

Evaluate a large number of performance estimators and relate them to the high-level analysis outcome

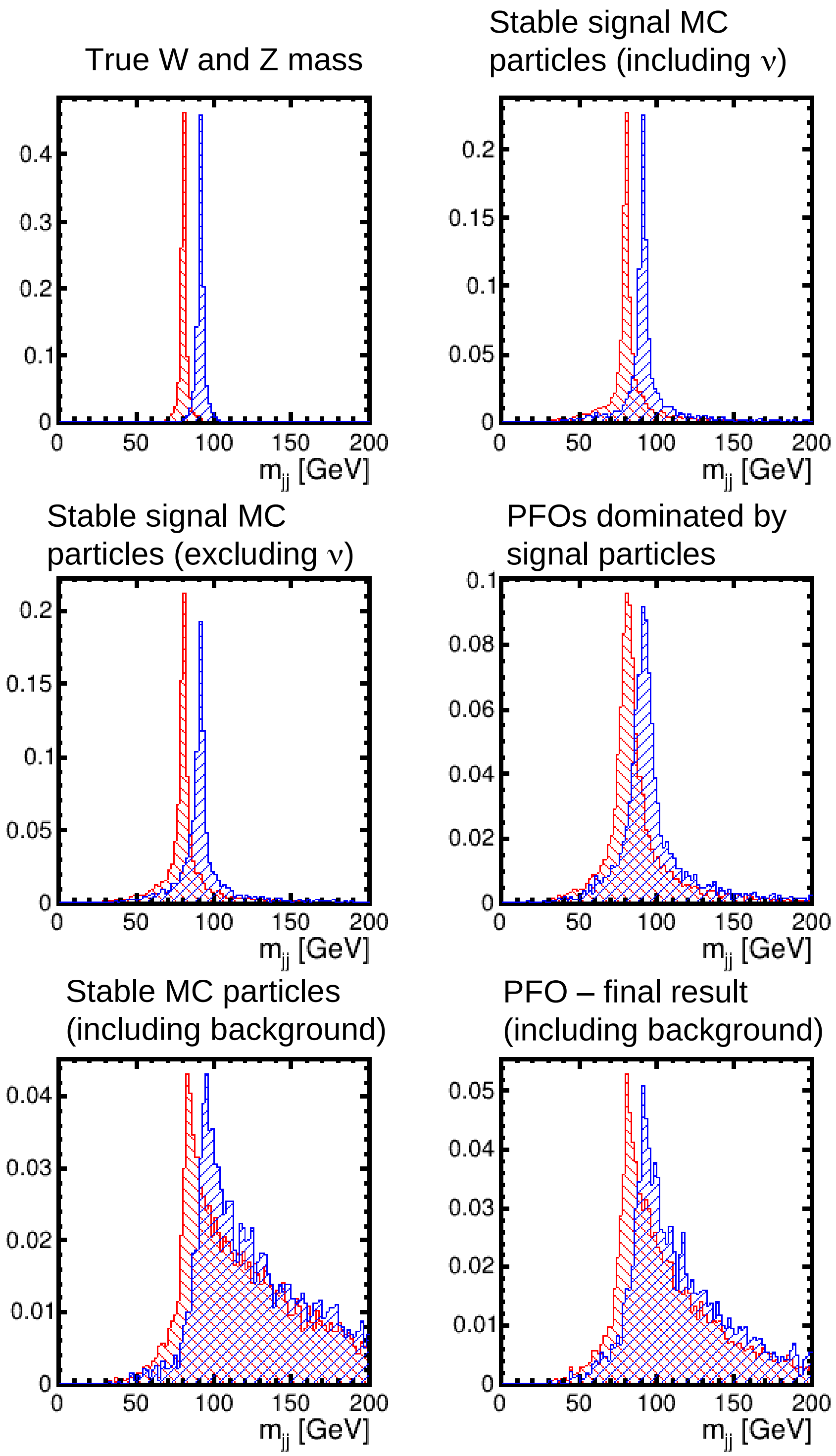
Scan over jet reconstruction algorithms – find choice that maximizes performance

See: CLIC jet paper arXiv:1607.05039

Work off a standard format (LCTuple, following prescription by Jenny List) provided by analysis teams

Benchmark channels so far:
- WW/ZZ production at 1 TeV (analysis by Jakob Beyer and Jenny List, DESY)
- ZH recoil mass analysis at 250 GeV (analysis by Yu Kato and Junping Tian, U. Tokyo)
- ….

WW/ZZ at 1 TeV: mass separation



MC truth selection to isolate pure WW and ZZ samples

Jet clustering with Durham, exclusive N=4

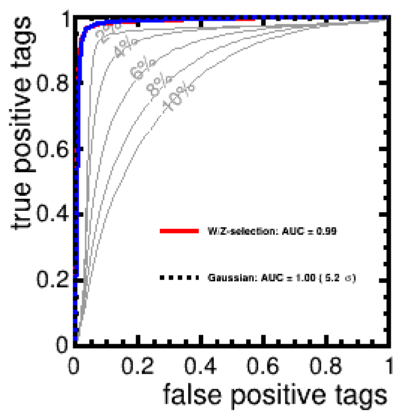
Clustering leads to tails, but cores still narrow

Particle Flow objects broaden cores

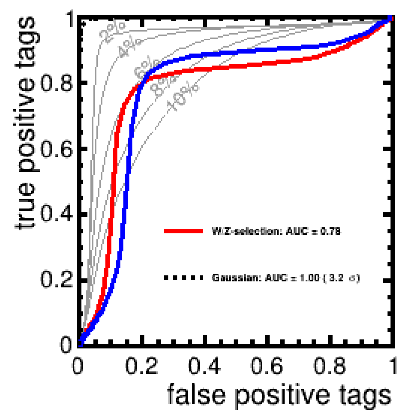
Background adds very pronounced tail

WW/ZZ at 1 TeV: ROC curves

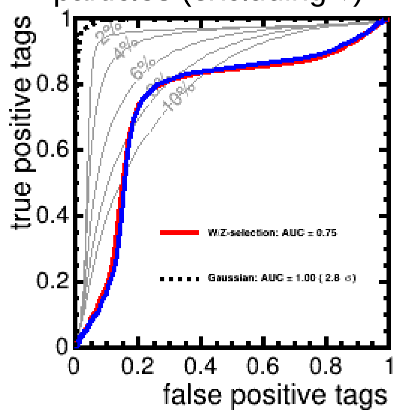
True W and Z mass



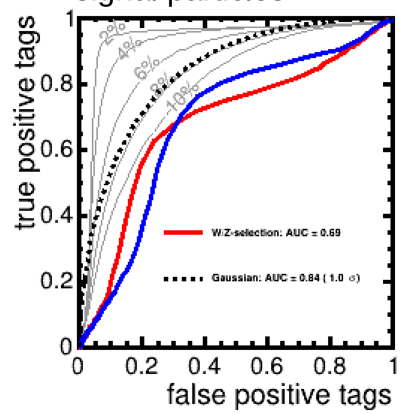
Stable signal MC particles (including ν)



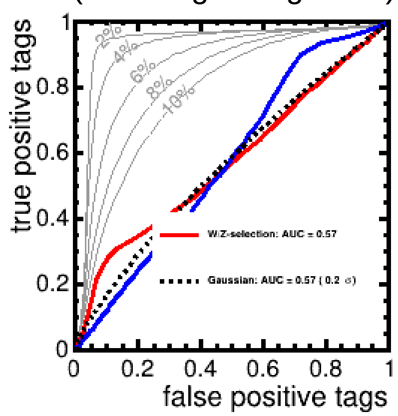
Stable signal MC particles (excluding ν)



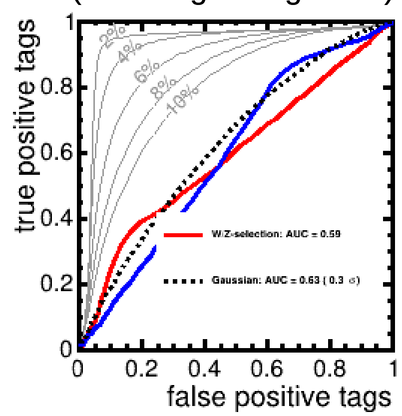
PFOs dominated by signal particles



Stable MC particles (including background)



PFO – final result (including background)



Receiver-Operator-Curves (true positive vs. false positive)
- red/blue lines: integrate distributions of slide 7
- grey reference lines: Gaussian JER 2,4,6,8,10%
- dashed reference: Gaussian JER fitted to distribution

Quantify W/Z separation with Area-under-curve
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

Jet algorithm space

VLC algorithm of arXiv:1607.05039

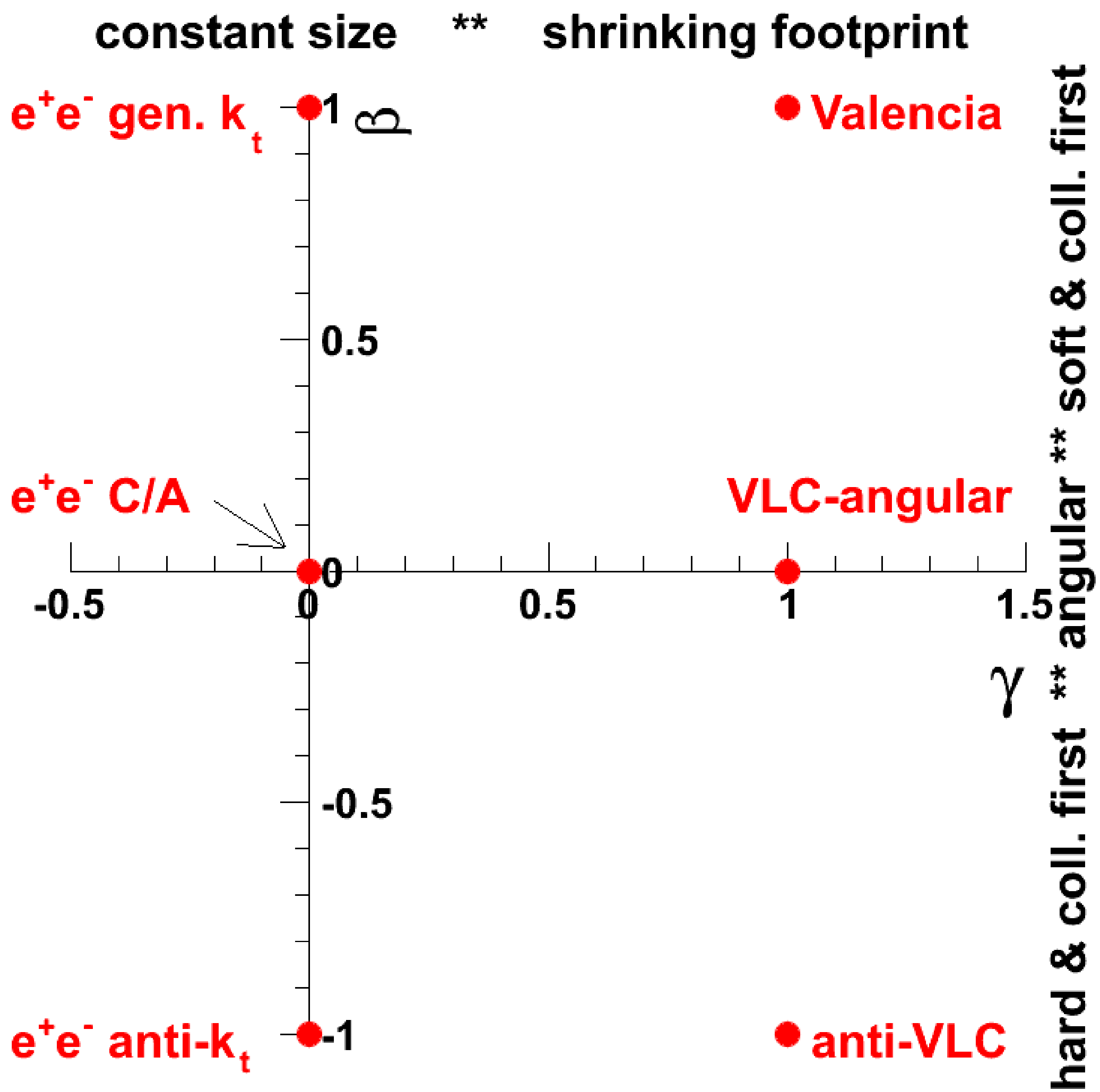
$$d_{ij} = 2\min(E_i^{2\beta}, E_j^{2\beta})(1-\cos\theta_{ij})/R^2,$$

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

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

Recover generalized e+e- kt for γ=0

Mimic robust longitudinally invariant algorithms with γ=1



Check out fjcontrib 1.040 or later if you're using FastJet

Grooming algorithms

Grooming techniques remove soft contamination from the jet so as to improve the jet substructure resolution and improve the resilience against pile-up and underlying event

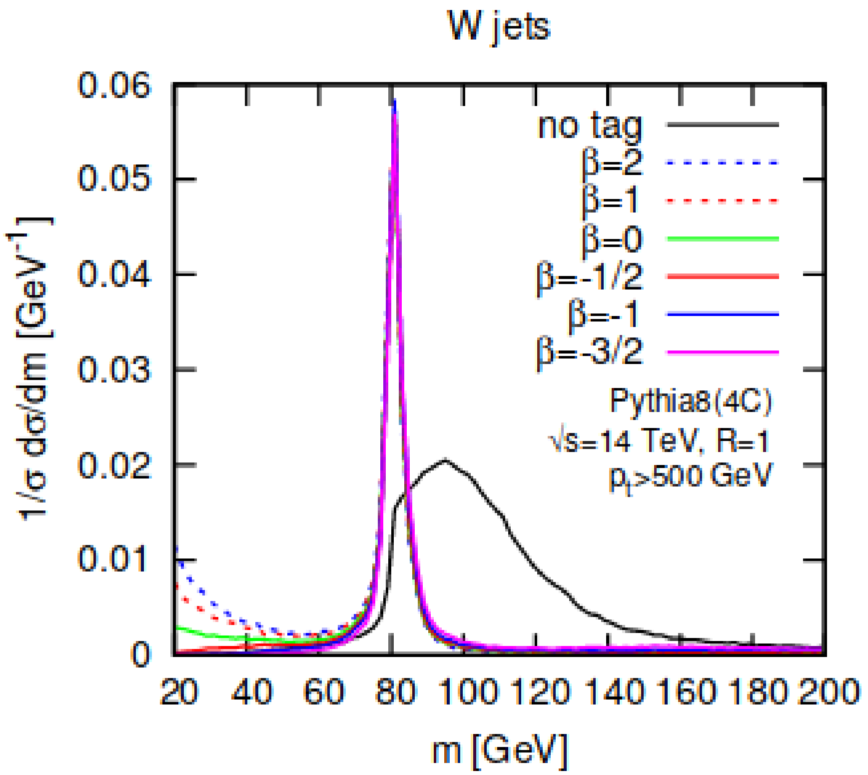
Grooming is part of the standard procedure for large-R jets at the LHC

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 trimming, pruning, etc.



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

WW/ZZ at 1 TeV: jet algorithms

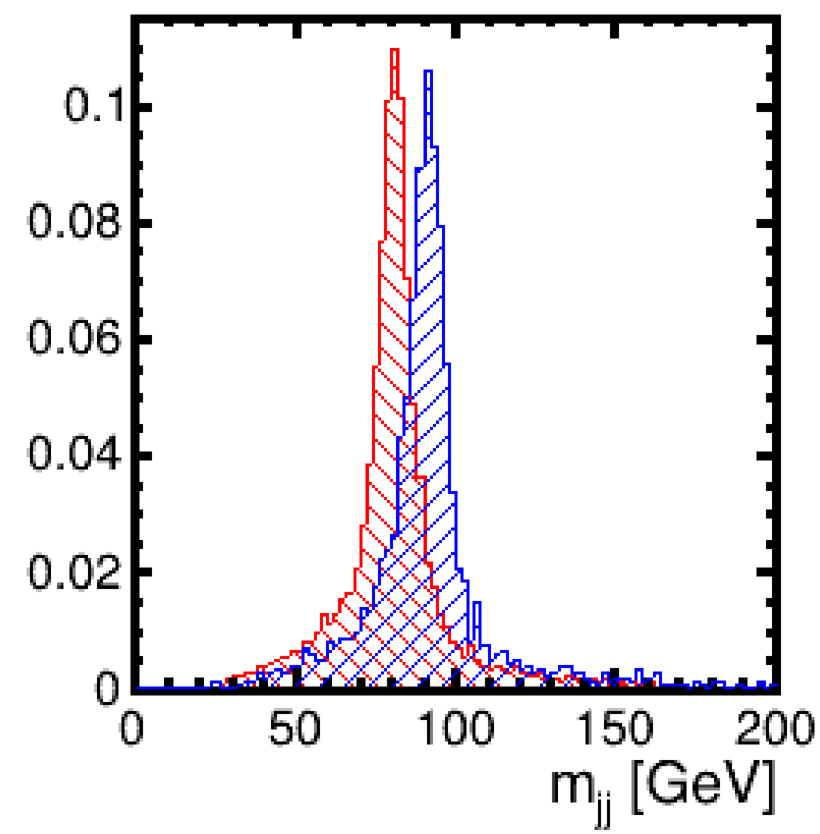
	Durham	Longitudinally invariant k_t R=1.4	VLC R=1.4	Durham on kt exlusive N=6	VLC R=1.4 with SoftDrop
Clustering	0.78	0.79	0.78	0.78	0.78
PFOs	0.69	0.73	0.72	0.72	0.72
background	0.59	0.70	0.72	0.70	0.71

Clustering essentially identical for all k_t algorithms

Robust algorithms yield big jump in performance with background

Only slight differences between the robust options

Possibly some further gain by optimizing parameters



WW/ZZ at 1 TeV: jet energy resolution

Estimate jet energy resolution by comparing reconstructed jets to stable particle jets
- the LHC way, reduces impact of clustering

Include background particles in reference jets
- reduce also the effect of background

Estimate all Graham Wilson's statistical estimators using his library

VLC R=1.4 on PFOs, reference = all stable signal particles (excl. background)

RMS90	:	6.6%
IQR68/2	:	6.0%
IQR95/4	:	8.4%

VLC R=1.4 on PFOs, reference = all stable MC particles (incl. background)

RMS90	:	6.1%
IQR68/2.	:	5.6%
IQR95/4.	:	6.4%

RMS90 is much larger than in simple reference analysis, where clustering and background are negligible

Interquantile range increases more strongly from IQR68 to IQR95 than it would for a Gaussian

Cannot recover reference performance even when background and clustering are largely taken into account

ZH recoil mass analysis

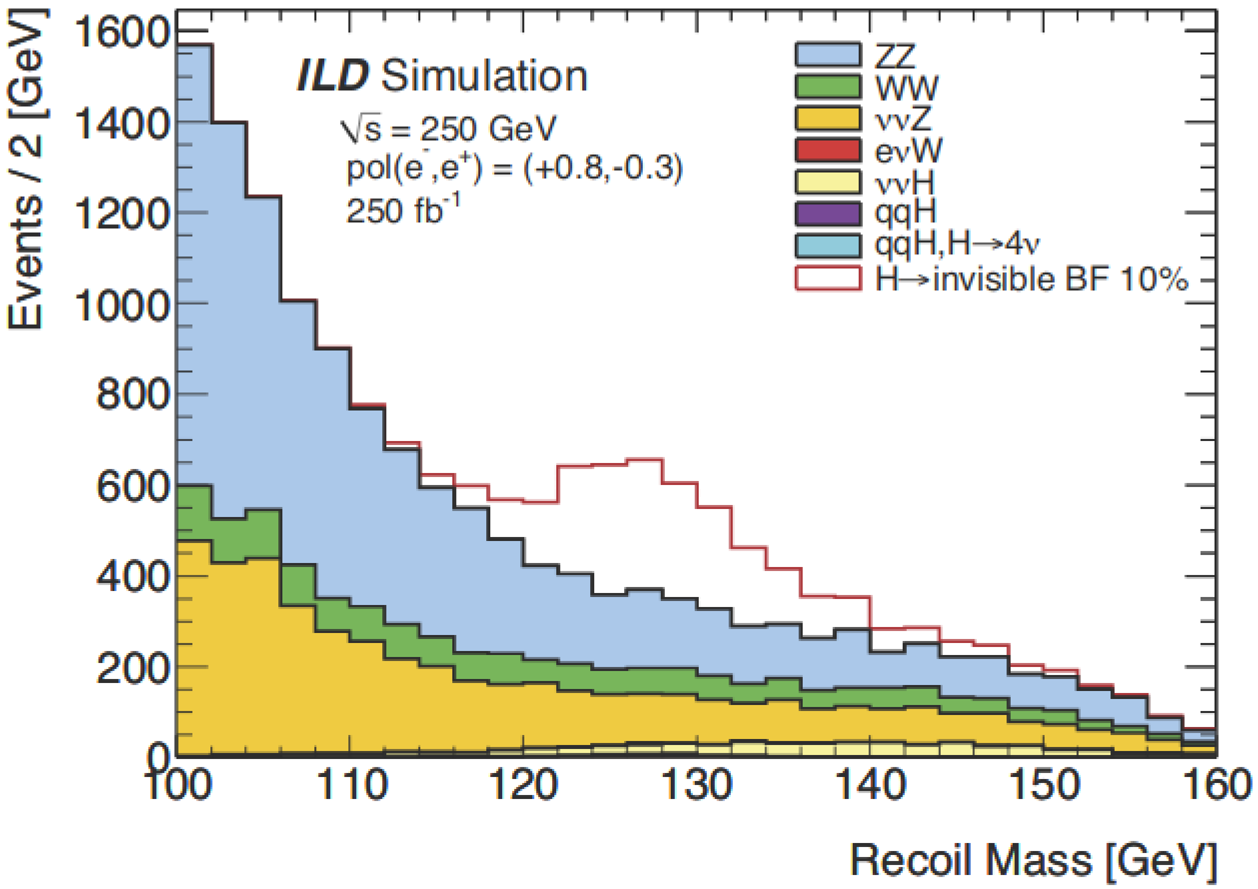
Recoil mass at 250 GeV (Yu Kato, Junping Tian)

Signal = $q\bar{q}H$, with $H \rightarrow gg$, $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$

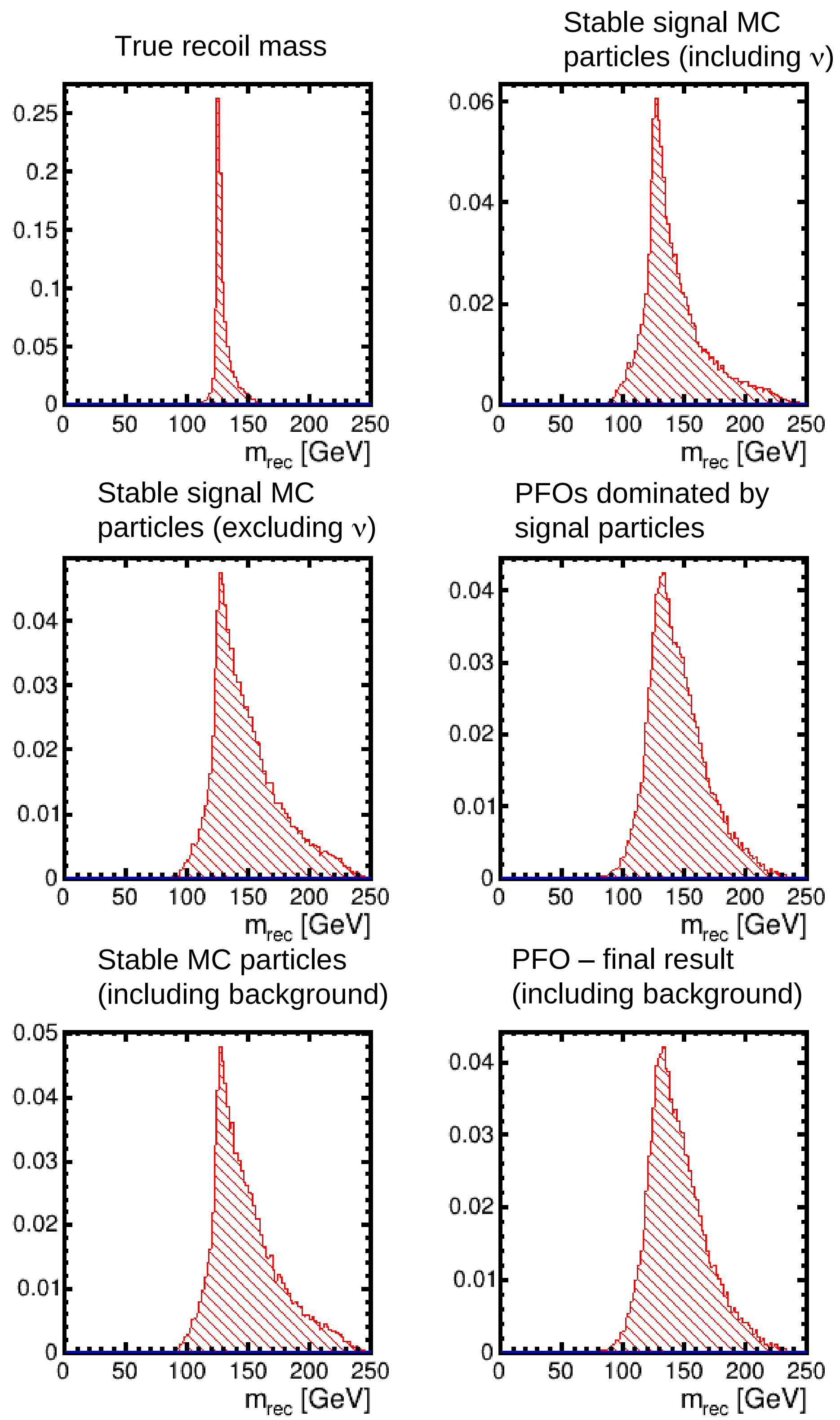
Background = ZZ, WW, etc.

Exclusive jet clustering with N=4

Calculate recoil mass with two jets



Recoil mass analysis: first results



First attempt at recoil mass analysis

VLC jets with R=1.4

True recoil mass distribution ~OK

Clustering has profound effect

Background is less important at 250 GeV

Include backgrounds and develop figure-of-merit in analogy to WW/ZZ analysis….

Summary

Some progress towards a common study of jet performance in several important benchmark analyses – thanks to help of Jakob Beyer and Yu Kato

Much work to do…

- cheated jets?
- figure-of-merit?
- further benchmarks?