Jet reconstruction at future (linear) e$^+$e$^-$ colliders

Marcel Vos
IFIC (UVEG/CSIC) Valencia

CEPC workshop
Oxford, April 2019

Based on work with Nacho García (IFIC), Philipp Roloff, Rosa Simoniello (CERN)
And ongoing studies with Jakob Beyer (DESY), Yu Kato, Junping Tian (U. Tokyo)
Acknowledging help from Gavin Salam (CERN) and Jesse Thaler (MIT)

Jet reconstruction at $e^+e^-$ colliders

Hadronic final states are important for the precision $e^+e^-$ programme
- Higgs production, arXiv:1509.02853
- Gauge boson pair production
- Top quark production, arXiv:1604.0122

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
- Jets "without the junk" (MPI, UE, pile-up)

Jet reconstruction is important
Performance goal: distinguish hadronic W and Z decays
The detectors are better suited than ever

**LC detector concepts optimized for particle flow**
- highly granular 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 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
CMS has adopted highly grangular calorimeter
Particle Flow achieves excellent response

Particle flow offers ultimate energy resolution
In theory able to achieve $\Delta E/E = 19\%/\sqrt{E}$ \textit{(theoretical limit for perfect track-cluster association)}
In practice limited to by confusion term for high energy jets: $\Delta E/E \sim 3\%$

We can achieve an excellent single-particle reponse
Jet reconstruction

In practice, the performance depends on...
- detector and particle flow response
- jet clustering (especially in complex final states)
- pile-up from bkg processes (esp. at high energy)
- neutrinos in bottom and charm jets

Detectors and particle flow are much better than at any previous experiment
Precision physics goals are very demanding of the event reconstruction
Final states are more complex (many jets, hard FSR & ISR) than at LEP

A very challenging environment for jet reconstruction!!
Jet algorithms revisited

VLC algorithm of arXiv:1607.05039

\[ d_{ij} = 2 \min (E_i^{2\beta}, E_j^{2\beta})(1 - \cos \theta_{ij})/R^2, \]
\[ d_{iB} = E_i^{2\beta} \sin^2 \gamma \theta_{iB}, \]

Two parameters (real numbers) govern the clustering order (\( \beta \)) and robustness against background (\( \gamma \)).

Recover generalized e+e- \( k_t \) for \( \gamma = 0 \)

Mimic robust longitudinally invariant algorithms with \( \gamma = 1 \)
(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
(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

Jet size and background resilience

The footprint or area of forward jets depends very strongly on the jet algorithm.

Shrinking footprint in forward region yields robust performance under background.

Extreme example: jet mass for top jets in 3 TeV collisions

Traditional $e^+e^-$ algorithms (Durham, $e^+e^- k_t$) break down
VLC is most robust, followed closely by long. inv. $k_t$

CLIC adopts VLC algorithm in many analyses
ILC (still) mostly Durham, often with two-stage clustering
Jet size and background resilience

The footprint or area of forward jets depends very strongly on the jet algorithm. Shrinking footprint in forward region yields robust performance under background.

The footprint or area of forward jets depends very strongly on the jet algorithm. Shrinking footprint in forward region yields robust performance under background.

Extreme example: jet mass for top jets in 3 TeV collisions

Traditional $e^+e^-$ algorithms (Durham, $e^+e^-$ $k_t$) break down. VLC is most robust, followed closely by long. inv. $k_t$.

CLIC adopts VLC algorithm in many analyses. ILC (still) mostly Durham, often with two-stage clustering.
A recent paper on the jet reconstruction performance of CEPC by Yongfeng Zhu and Manqi Ruan clearly shows the impact of jet clustering on the WW/ZZ separation.

Better clustering

Worse clustering

Y. Zhu, M. Ruan, EPJC 79 (2019) 274

See talk by M. Ruan in yesterday’s parallel session and by Matthias Weber this afternoon.
WW/ZZ at 1 TeV

ILD benchmark study of WW/ZZ at 1 TeV

Jakob Beyer (DESY), Y. Kato (Tokyo) and M.V.

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

Clustering leads to tails, but cores still narrow

Neutrinos affect response for bottom and charm jets

Detector response broadens cores

Background adds very pronounced tail (for Durham)
Figure-of-merit: ROC curves

Receiver-Operator-Curves: relate tagging efficiency for ZZ and WW events or true positive vs. false positive events

Figure-of-merit: quantify W/Z separation in a single number with the area-under-curve

ROC curves provide a single figure-of-merit to quantify performance: AUC = 0.5 (random) - 1 (perfect)
WW/ZZ at 1 TeV: ROC curves

- 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

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

Detector response broadens cores
AUC ~ 0.69

Background adds very pronounced tail (for Durham)
AUC ~ 0.58
**WW/ZZ at 1 TeV: jet algorithms**

<table>
<thead>
<tr>
<th></th>
<th>Durham</th>
<th>Long. invariant k_t R=1.4</th>
<th>VLC R=1.4</th>
<th>Durham on kt exclusive N=6</th>
<th>VLC R=1.4 with SoftDrop</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC particles</td>
<td>0.78</td>
<td>0.79</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>PFOs</td>
<td>0.69</td>
<td>0.73</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>+ background</td>
<td>0.59</td>
<td>0.70</td>
<td>0.72</td>
<td>0.70</td>
<td>0.71</td>
</tr>
</tbody>
</table>

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

![Plot of m_{jj} [GeV] vs. m_{jj} [GeV] with histograms of particle production and background](image)
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
Jet grooming

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

\[ e^+ e^- \rightarrow q\bar{q} \text{ (N=2)}, \]
break up into sub-jets with mass-drop filtering with \( R = R_{\text{sub}} \),
Select 3 hardest sub-jets

For fair comparison, choose
\[ R_{\text{sub}}^2 = R^2/3 \] 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
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 a lepton collider than the equivalent lepton collider algorithm?

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

**VS.**

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

Beam jets + shrinking footprint with polar angle yields increased robustness against forward-peaked $\gamma\gamma \rightarrow$ hadrons!

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

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

$e^+e^-$ distance between particles

$$d_{iB} = p_{Ti}^{2n}$$

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

$\beta$ 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)