## Particle Flow Calorimetry and PandoraPFA

Mark Thomson University of Cambridge



#### <u>This Talk:</u>

#### • Why the ILC

- **❷** ILC Physics ↔ ILC Calorimetry
- **The Particle Flow Paradigm**
- Calorimetry in the ILC Detector Concepts
- **• PFA and Detector Design**
- PandoraPFA Particle Flow Algorithm (PFA)
- Performance and Detector Optimisation Studies
- S Future Development

#### Occursion

## • Why the ILC ?

The LHC and ILC provide a complimentary approach to studying the physics of EWSB and beyond

### The LHC

- **★** Will open the door to new physics !
- **★** Pushes the energy frontier with proton-proton collisions at 14 TeV
  - qq, qg and gg collisions in the energy range 0.5-5 TeV

### The ILC

**★** A different approach:

very high precision as opposed to very high energy

- ★ Electron-positron collisions in the energy range 0.1-1 TeV
- ★ Very clean final states + high resolution detectors
  - very precise measurements (as at LEP)
  - → detailed understanding of new physics + tight
    - constraints on theory (as at LEP)

The case for having both the LHC and ILC very well studied:

e.g. "Physics Interplay of the LHC and ILC", G. Weiglein et al., Phys. Rept. 426 (2006) 47-358

## $e^+ e^- \equiv precision$

## ★Electron-positron colliders provide clean environment for precision physics



\*At electron-positron the final state corresponds to the underlying physics interaction, e.g. above see  $H \rightarrow b\overline{b}$  and  $Z \rightarrow \mu^+\mu^-$  and nothing else...

## ❷ ILC Physics ↔ Calorimetry

### **ILC PHYSICS:**

#### **Precision Studies/Measurements**

- ★ Higgs sector
- **★** SUSY particle spectrum (if there)
- **★** SM particles (e.g. W-boson, top)
- ★ and much more...

### **Physics characterised by:**

High Multiplicity final states
 often 6/8 jets

### **\***Small cross-sections

e.g. σ(e<sup>+</sup>e<sup>-</sup>→ZHH) = 0.3 fb



 Require High Luminosity – i.e. the ILC
 Detector optimized for precision measurements in difficult multi-jet environment

## **Compare with LEP**



\*Backgrounds dominate 'interesting' physics
 \*Kinematic fitting much less useful: Beamsstrahlung + final states with > 1 neutrino

- Physics performance depends critically on the detector performance (not true at LEP)
- \* Places stringent requirements on the ILC detector

★Of the ILC goals the most challenging is (probably) that of jet energy resolution:

**★**So why is this important ?

## **Calorimetry at the ILC**

Jet energy resolution:

Best at LEP (ALEPH):  $\sigma_{E}/E = 0.6(1+|\cos\theta_{Jet}|)/\sqrt{E(GeV)}$ 

#### ILC GOAL: $\sigma_{E}/E = 0.3/\sqrt{E(GeV)}$

THIS IS HARD !

#### **\*** Jet energy resolution directly impacts physics sensitivity



Reconstruction of two di-jet masses allows discrimination of WW and ZZ final states Often-quoted Example:

If the Higgs mechanism is not responsible for EWSB then QGC processes important e<sup>+</sup>e<sup>-</sup>→<sub>VV</sub>WW→<sub>VV</sub>qqqq, e<sup>+</sup>e<sup>-</sup>→<sub>VV</sub>ZZ→<sub>VV</sub>qqqq



★ EQUALLY applicable to any final states where want to separate
 W→qq and Z→qq !

## **Calorimetry Goals**

\* Aim for jet energy resolution giving di-jet mass resolution similar to Gauge boson widths \* For a pair of jets have:  $m^2 = m_1^2 + m_2^2 + 2E_1E_2(1 - \beta_1\beta_2\cos\theta_{12})$ \* For di-jet mass resolution of order  $\Gamma_{W/Z}$   $\frac{\sigma_m}{m} \approx \frac{2.5}{91.2} \approx \frac{2.1}{80.3} \approx 0.027$   $rac{\sigma_{E_j}/E_j < 3.8\%$ + term due to  $\theta_{12}$  uncertainty \* Assuming a single jet energy resolution of normal form

	$\sigma_E/E = \alpha(E)/\sqrt{E(\text{GeV})}$
	$\sigma_m/m \approx \alpha(E_j)/\sqrt{E_{jj}(\text{GeV})}$
⇒	$\alpha(E_j) < 0.027 \sqrt{E_{jj}(\text{GeV})}$

E <sub>jj</sub> /GeV	$\alpha(E_{jj})$
100	< 27 %
200	< 38 %

★ Typical di-jet energies at ILC (100-300 GeV) suggests jet energy resolution goal of  $\sigma_E/E < 0.30/\sqrt{E_{jj}(\text{GeV})}$ 



Very hard (may not be possible) to achieve this with a traditional approach to calorimetry

Limited by typical HCAL resolution of  $> 50\%/\sqrt{E(GeV)}$ 



a new approach to calorimetry

## **B** The Particle Flow Paradigm

#### ★ In a typical jet :

- 60 % of jet energy in charged hadrons
- + 30 % in photons (mainly from  $\pi^0 o \gamma\gamma$  )
- + 10 % in neutral hadrons (mainly  $n \mbox{ and } K_L$  )
- **★** Traditional calorimetric approach:
  - Measure all components of jet energy in ECAL/HCAL !
  - + ~70 % of energy measured in HCAL:  $\sigma_{\rm E}/{\rm E} \approx 60\,\%/\sqrt{{\rm E}({\rm GeV})}$
  - Intrinsically "poor" HCAL resolution limits jet energy resolution





- **★** Particle Flow Calorimetry paradigm:
  - charged particles measured in tracker (essentially perfectly)
  - Photons in ECAL:  $\sigma_{\rm E}/{\rm E} < 20\,\%/\sqrt{{\rm E}({\rm GeV})}$
  - Neutral hadrons (ONLY) in HCAL
  - Only 10 % of jet energy from HCAL ⇒ much improved resolution

## **Particle Flow Calorimetry**

#### Hardware:

★Need to be able to resolve energy deposits from different particles
 → Highly granular detectors (as studied in CALICE)





#### Software:

**\***Need to be able to identify energy deposits from each individual particle !

Sophisticated reconstruction software



**\*** Particle Flow Calorimetry = HARDWARE + SOFTWARE

## **Particle Flow Reconstruction (PFA)**

#### **Reconstruction of a Particle Flow Calorimeter:**

- **\*** Avoid double counting of energy from same particle
- **\*** Separate energy deposits from different particles



If these hits are clustered together with these, lose energy deposit from this neutral hadron (now part of track particle) and ruin energy measurement for this jet.

Level of mistakes, "confusion", determines jet energy resolution <u>not</u> the intrinsic calorimetric performance of ECAL/HCAL

### sounds easy....

# PFA performance depends on detailed reconstruction Relatively new, still developing ideas (not just software) Studies need to be based on a sophisticated detector simulations

## **4** The ILC Detector Concepts

#### **ILC Detector Concepts:**

- ILC Detector Design work centred around 3 detector "concepts"
- ★ Partial "technical designs" ~ 2010 ?
- \* 2 main concepts "optimised" for PFA Calorimetry SiD, ILD

#### SiD : Silicon Detector





Δ

**ILD** : International Large Detector



#### **★**Major ongoing simulation study to define baseline ILD parameters



Central Tracker and ECAL

	SiD	LDC	GLD
Tracker	Silicon	TPC	TPC
ECAL	SiW	SiW	Pb/Scint

SiD + LDC + GLD all designed for PFA Calorimetry !

**★** also "4<sup>th</sup>" concept designed for more "traditional" approach to calorimetry

## e.g. LDC/SiD Calorimetry

#### ECAL and HCAL inside coil





#### ECAL: silicon-tungsten (SiW) calorimeter:

- Tungsten : X<sub>0</sub> /λ<sub>had</sub> = 1/25, R<sub>Moliere</sub> ~ 9mm (gaps between Tungsten increase effective R<sub>Moliere</sub>)
- Lateral segmentation: ~1cm<sup>2</sup> matched to R<sub>Moliere</sub>
- Longitudinal segmentation: 30 layers (24  $X_0$ , 0.9 $\lambda_{had}$ )
- Typical resolution:  $\sigma_{E}/E = 0.15/\sqrt{E(GeV)}$

#### Very high longitudinal and transverse segmentation

## **Calorimeter Reconstruction**

- High granularity calorimeters <u>very different</u> to previous detectors (except LEP lumi. calorimeters)
- \* "Tracking calorimeter" requires a new approach to ECAL/HCAL reconstruction



**+PARTICLE FLOW** 

- **ILC calorimetric performance = HARDWARE + SOFTWARE** 
  - **Performance will depend on the software algorithm**

Nightmare from point of view of detector design/optimisation

## **• PFA and ILC detector design ?**



Z

Z

#### For example:

#### **★** Wish to compare performance of say LDC and SiD detector concepts

e.g. tt event in LDC





- **\*** However performance = DETECTOR + SOFTWARE
- ★ Non-trivial to separate the two effects
- **\*** NEED REALISTIC SIMULATION **+** REALISTIC <u>RECONSTRUCTION</u> !
  - can't use fast simulation etc.

For design of ILC Calorimetry : need <u>realistic</u> reconstruction chain ~10 years before start of ILC !!!

(ideally before start of LHC)

(even more challenging: the software has to work for multiple detector design parameters)

> We already have a first version.... (many vital contributions from DESY)

## **Reconstruction Framework (C++)**

#### **★** Everything exists – level of sophistication ~LEP experiment



## **6** PandoraPFA



Need sophisticated Particle Flow reconstruction before it is possible to start full detector design studies



New paradigm – nobody really knows how to approach this



So where are we now ?

Significant effort (~4 groups developing PFA reconstruction worldwide)

### For this talk concentrate on: PandoraPFA



This is still work-in-Progress – currently it gives the best performance



Then discuss some first detector optimisation studies

## **PandoraPFA Overview**

★ ECAL/HCAL reconstruction and PFA performed in a single algorithm

- **★** Keep things fairly generic algorithm
  - applicable to multiple detector concepts
- **★** Use tracking information to help ECAL/HCAL clustering
- **★** This is a fairly sophisticated algorithm : 10<sup>4</sup> lines of code

#### Six Main Stages:

- i. Preparation
- ii. Loose clustering in ECAL and HCAL
- iii. Topological linking of clearly associated clusters
- iv. Courser grouping of clusters
- v. Iterative reclustering
- vi. Photon Identification/Recovery
- vii. Fragment removal
- viii. Formation of final Particle Flow Objects (reconstructed particles)

## i) Tracking

★The use of optimal use of tracking information in PFA is essential

- ★ Non trivial for looping tracks (even in a TPC)
- **★** Matching of tracks to endcap clusters is non-trivial
- **★** Use of track information is a major part of PandoraPFA
- **★** Big effort to use as many tracks in the event as possible
  - helps particularly for lower energy jets
  - motivation I : better energy resolution
  - motivation II : correct measurement of direction

★ TPC-oriented: take advantage of pattern recognition capability (the algorithm would need modification for Si tracker)



## e.g. Kinks

#### **★** Kink finding extends to "loopers"



#### **★** Can give a measure of missing energy



- Consider physics hypothesis, e.g.  $K^{\pm} \rightarrow \mu^{\pm} v$ 
  - Use Helix fits to start and end of tracks to reconstruct missing particle e.g. v
  - Can then reconstruct primary mass
  - If consistent with hypothesis, e.g. m<sub>K</sub> use primary track for PFO four-momentum

PandoraPFA reconstructs (some) neutrinos !

## ii) ECAL/HCAL Clustering

- **★** Start at inner layers and work outward
- ★ Tracks can be used to "seed" clusters
- **★** Associate hits with existing Clusters
- ★ If no association made form new Cluster
- **\*** Simple cone based algorithm



## iii) Topological Cluster Association

#### +By design, clustering errs on side of caution i.e. clusters tend to be split

Philosophy: easier to put things together than split them up
 Clusters are then associated together in two stages:

- 1) Tight cluster association clear topologies
- 2) Loose cluster association fix what's been missed

### 🔆 <u>Photon ID</u>

**\***Photon ID plays important role

**\*Simple** "cut-based" photon ID applied to all clusters

Clusters tagged as photons are immune from association procedure – just left alone



#### **★** Clusters associated using a number of topological rules

#### **Clear Associations:**

 Join clusters which are clearly associated making use of high granularity + tracking capability: very few mistakes



27

## **Example : MIP segments**

\*Look at clusters which are consistent with having tracks segments and project backwards/forward (defined using local straight-line fits to hits tagged as MIP-like)



Apply tight matching criteria on basis of projected track [NB: + track quality i.e. chi2]

**★**Here, association based on "tracking" in calorimeters

## iv) Cluster Association Part II

- Have made very clear cluster associations
- Now try "cruder" association strategies
- BUT first associate tracks to clusters (temporary association)
- Use track/cluster energies to "veto" associations, e.g.



**Provides some protection against obvious mistakes** 



## v) Iterative Reclustering

★ Upto this point, in most cases performance is good – but some difficult cases...



★ At some point hit the limit of "pure" particle flow

• just can't resolve neutral hadron in hadronic shower







Change clustering parameters until cluster splits and get sensible track-cluster match

**NOTE:** NOT FULL PFA as clustering driven by track momentum

This is <u>very</u> important for higher energy jets

### **Iterative Reclustering Strategies**

#### ① Cluster splitting **Reapply entire clustering algorithm 18 GeV** to hits in "dubious" cluster. Iteratively 30 GeV reduce cone angle until cluster splits to give acceptable energy match to track 10 GeV Track \* + plug in alternative clustering algorithms **2** Cluster merging with splitting 38 GeV Look for clusters to add to a track to get sensible energy association. If necessary iteratively split up clusters 12 GeV to get good match. **③** Track association ambiguities

In dense environment may have multiple tracks matched to same cluster. Apply above techniques to get ok energy match.

#### **4** "Nuclear Option"

★ If none of above works – kill track and rely on clusters alone

## vi) Photon ID/Recovery

#### Use simple cut-based photon ID in the early (CPU intensive) stages of PandoraPFA

★ In the final stages, use improved photon ID based on the expected EM longitudinal profile for cluster energy E<sub>0</sub>

$$\Delta E = E_0 \frac{(t/2)^{a-1} e^{-t/2}}{\Gamma(a)} \Delta t \qquad a = 1.25 + \frac{1}{2} \ln E_0 / E_c$$

\* Convert cluster into energy depositions per radiation length (use cluster to determine the layer spacing, i.e. geometry indep.)



- Shower Profile fixed by cluster energy
- But fit for best shower start, s
- Normalise areas to unity and calc.

$$f = \sum_{i} |o_i - e_i|$$

- Gives a measure of fractional
- disagreement in obs/exp profiles
- Use f and s to ID photons

## **Photon Recovery**

- **★** With cone clustering algorithm, photons close to early showering charged hadrons can be merged into a single cluster.
- **★** Use longitudinal + transverse profile to recover these
- **★** Essentially, for each cluster associated with a track:
  - project ECAL hits onto plane perpendicular to radial vector to point where track intersects ECAL
  - search for peaks...



Use profiles to "dig out" photons overlapping with hadronic clusters:

- Also look for photons where only a single peak is found
- Implemented by looking at longitudinal profile of "shower"



Only allowed if it results in acceptable track-cluster energy consistency...

NOTE: in PandoraPFA, photon identification is an "iterative", rather than one-off process: different levels of sophistication applied at different stages of algorithm

## viii) Fragment removal : basic idea

**★** Look for "evidence" that a cluster is associated with another



- ★ Convert to a numerical evidence score E
- ★ Compare to another score "required evidence" for matching, R, based on change in E/p chi-squared, location in ECAL/HCAL etc.
- If E > R then clusters are merged
- ★ Rather ad hoc but works well but works well

## Putting it all together...



## Performance / Detector Studies





Mark Thomson

### **Performance (LDC00)**

rms90	PandoraPFA v02-01		
E <sub>JET</sub>	$\sigma_{\rm E}/{\rm E} = \frac{\alpha}{\sqrt{E_{\rm jj}}}$  cosθ <0.7	σ <sub>Ε</sub> /Ε <sub>j</sub>	
45 GeV	0.235	3.5 %	
100 GeV	0.306	3.1 %	
180 GeV	0.427	3.2 %	
250 GeV	0.565	3.6 %	

NOTE: studies based on ILD detector concept are "work-in-progress"

Tesla TDR detector modelFull simulationFull reconstruction

### In simulation

- **★** Particle flow can achieve ILC goal of  $\sigma_{\rm E}/E_{\rm i}$  < 3.8 %
- ★ For lower energy jets Particle Flow gives unprecedented levels of performance, e.g. @ 45 GeV : 3.5% c.f. ~10% (ALEPH)
- **\star** "Calorimetric" performance ( $\alpha$ ) degrades for higher energy jets
- ★ + current code is not perfect can do better

#### **PARTICLE FLOW CALORIMETRY WORKS !**

## **Detector Optimisation : How ?**

\* Optimise detector based on physics performance vs cost
 \* First step is to parameterise physics performance vs R, B, L, ...



★ Then fold in cost to motivate baseline ILD design

<u>BUT:</u> non-trivial:

- **★** PFA reconstruction is complex
- **★** Tracking plays a major role in own right and in PFA reconstruction

★ Flavour tagging essential to many physics studies Beware: short-cuts...

Solid study of ILC detector performance/optimisation requires a full detector simulation and REALISTIC reconstruction

Beware: software...

Interpretation: observing effects of detector + imperfect software

### e.g. HCAL Depth and Transverse segmentation

- **\*** Investigated HCAL Depth (interaction lengths)
  - Generated Z→uds events with a large HCAL (63 layers)
    - approx 7  $\lambda_{I}$
  - In PandoraPFA introduced a configuration variable to truncate the HCAL to arbitrary depth
  - Takes account of hexadecagonal geometry



#### e.g. change HCAL tile size $1x1 \rightarrow 10x10 \text{ mm}^2$



## **Radius vs Field**



design from point of view of Particle Flow Calorimetry

- **\*Need to extend to physics sensitivity**
- **\***Need to match this with detector cost model
- \*In very near future should have a much better idea of the parameters of a cost-performance optimized ILC detector

## 8 Random Comments

#### **Deficiencies:**

- ★ PandoraPFA has evolved solely with the aim of improving performance ... never overly concerned with niceties...
- ★ Very little has been optimised:
  - Photon ID good be better
  - Photon Recovery crude
  - Fragment Removal <u>very crude</u>

### Plenty of room for improvement

#### Why does PandoraPFA work reasonably well?

- **★** PFA = much more than clustering
  - basic clustering algorithm developed in about a week shortly after Snowmass – essentially unchanged
- **★** Lessons learnt in developing code:
  - advantages in having a single "coherent" approach
  - always concentrated on optimising jet E performance, not photonID efficiencies etc.
  - extreme care with all stages avoid unnecessary mistakes
  - great care needed in track/cluster matching
  - use of track momentum cluster energy to spot to PFA errors absolutely vital
  - PFA reconstruction is an iterative process, use more sophisticated techniques as knowledge of event improves

## The Downside

- ★ PandoraPFA evolved; only initial aspects designed (necessary as didn't know approach when starting out)
- ★ Consequently PandoraPFA is
  - Large ~15000 lines of (partially commented) C++
  - Not particularly transparent
  - Increasingly difficult to maintain/develop
  - Probably slower than necessary



## The Future ?

- **★** Now have a much better idea of the general approach
- ★ Could contemplate "starting from scratch"
  - Improved structure (i.e. some)
  - Improved interfaces, e.g. clustering, to allow easier development of alternative algorithms
  - Increase "expert" base for future development
- **\*** BUT would require significant effort
- **★** Would be helped validation through comparison to existing code
- **★** Worth the effort...?

## **9** Conclusions

- **★** Great deal of effort (worldwide) in the design of the ILC detectors
- \* Centred around 3 "detector concept" groups: ILD (GLD+LDC), SiD + 4<sup>th</sup>
- ★ Widely believed that calorimetry and, in particular, jet energy resolution drives detector design
- **★** Also believed that it is likely that PFA is the key to achieving ILC goal

### **PFA IS HARD – BUT POTENTIALLY IMPORTANT !**

- **★** Calorimetry at the ILC = HARDWARE + SOFTWARE (new paradigm)
- ★ It is difficult to disentangle detector/algorithm....
- ★ Can only address question with "realistic algorithms"
  - i.e. serious reconstruction 10+ years before ILC turn-on
- ★ With PandoraPFA have reached the ILC "goal" (for Z →uds events)
- **★** More importantly, getting close to being able to address real issues:
  - What is optimal PFA detector size/B-field, etc.

#### FINAL COMMENT:

**★**GLD, LDC, SiD calorimetry "designed" for PFA

- Needed to demonstrate this actually makes sense
  until relatively recently not completely proven !
- **\***Now try and understand ultimate reach of PFA calorimetry

## The End