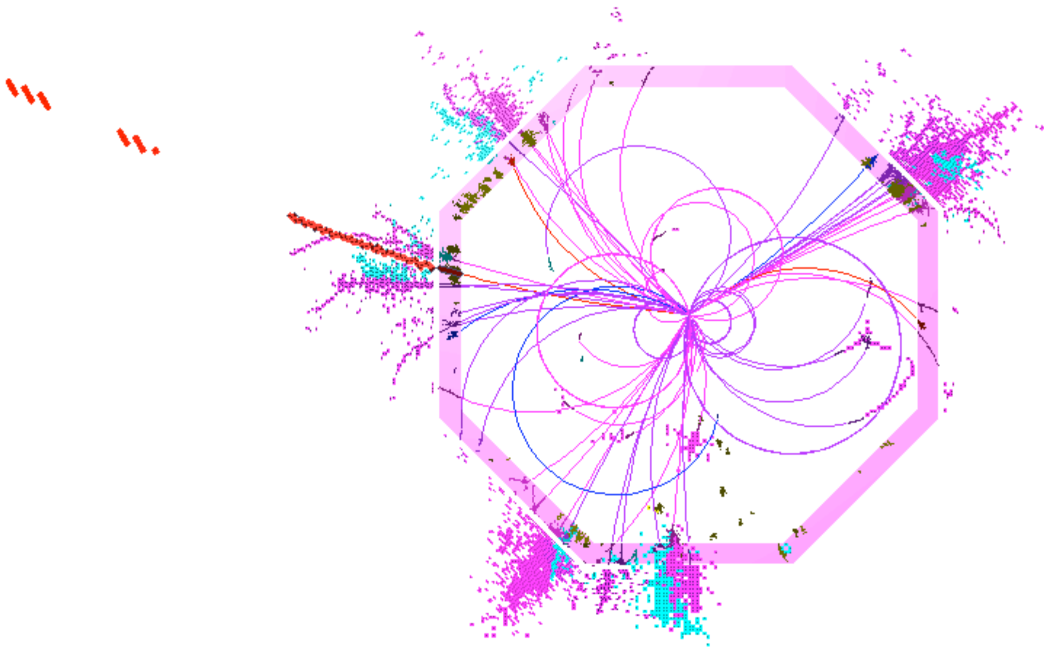


Pandora Particle Flow Algorithm

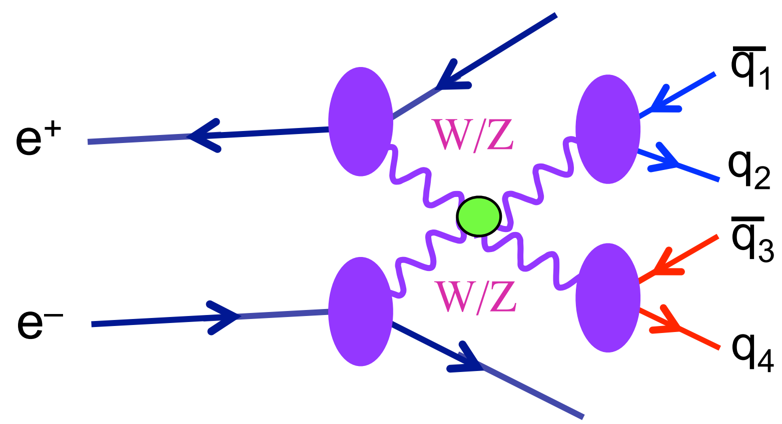
J.S. Marshall, M.A. Thomson, University of Cambridge
CHEF2013, Paris, 24th April 2013

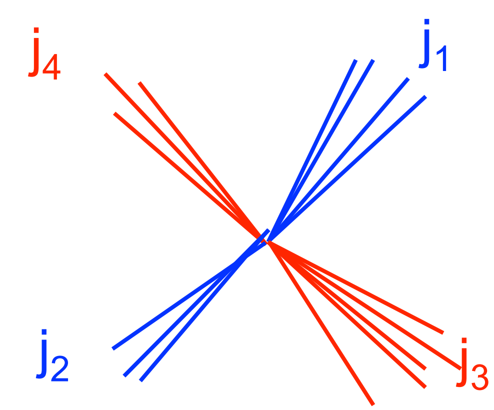
Overview

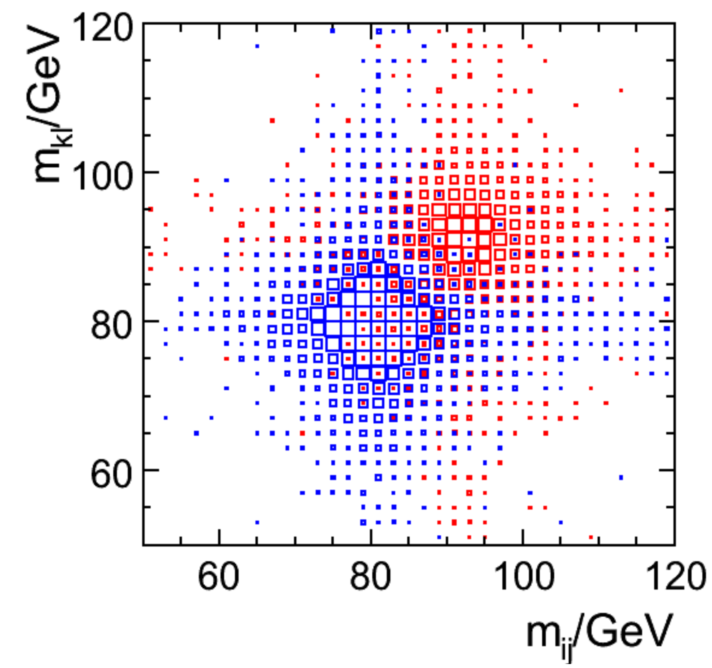
1. Calorimetry Goals at a Linear Collider
2. Fine Granularity Particle Flow Calorimetry
3. Pandora Particle Flow Algorithm
4. Particle Flow Performance at ILC
5. Particle Flow Performance at CLIC
6. Summary

LC Calorimetry Goals

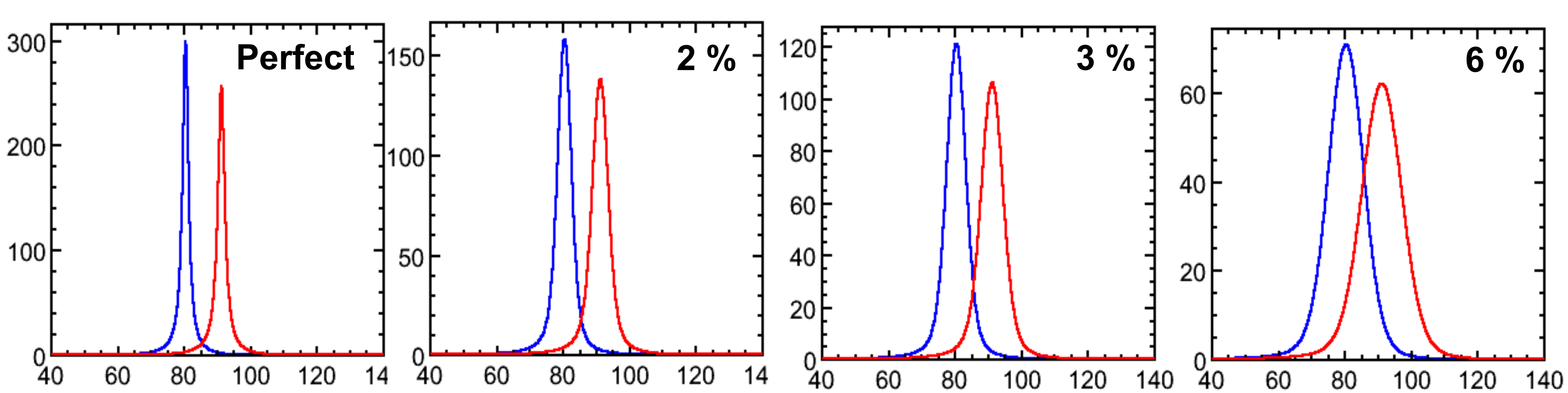
- Jet energy resolution requirements depend on physics…
- Likely to be primarily interested in di-jet mass resolution.
- Strong desire to separate W/Z hadronic decays.





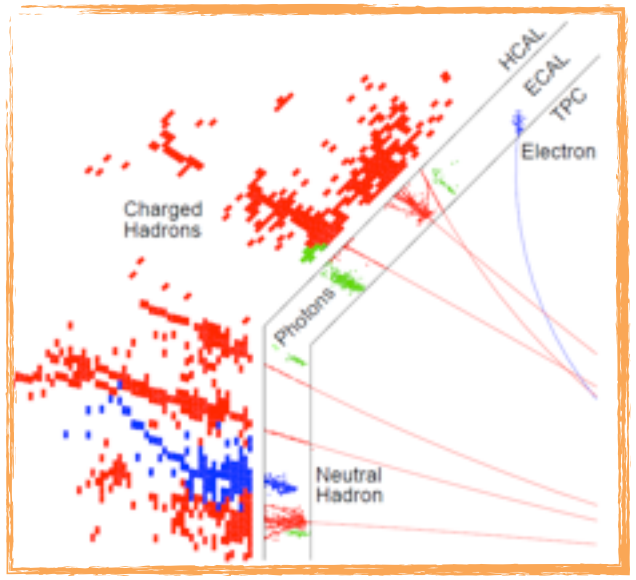


- 3-4% jet energy resolution gives decent 2.6-2.3σ W/Z separation.
- Sets a reasonable choice for LC jet energy minimal goal ~3.5%.
- For W/Z separation, not much further gain; limited by natural widths.



W/Z sep:
$(m_Z - m_W) / \sigma_m$

Jet E res.	W/Z sep
Perfect	3.1 σ
2%	2.9 σ
3%	2.6 σ
4%	2.3 σ
5%	2.0 σ
10%	1.1 σ

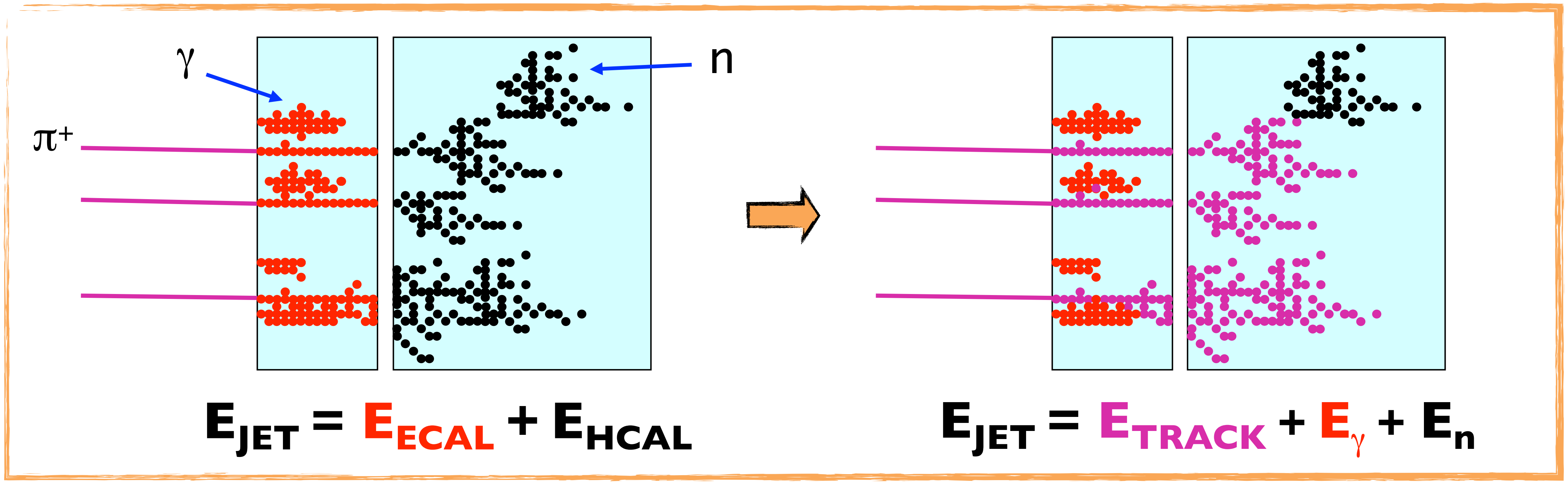


In a typical jet:

- 60 % of jet energy in charged hadrons
- 30 % in photons (mainly from $\pi^0 \rightarrow \gamma\gamma$))
- 10 % in neutral hadrons (mainly n and K_L)

Traditional calorimetric approach:

- Measure all components of jet energy in ECAL/HCAL
- Approximately 70% of energy measured in HCAL: $\sigma_E/E \approx 60\% / \sqrt{E(GeV)}$



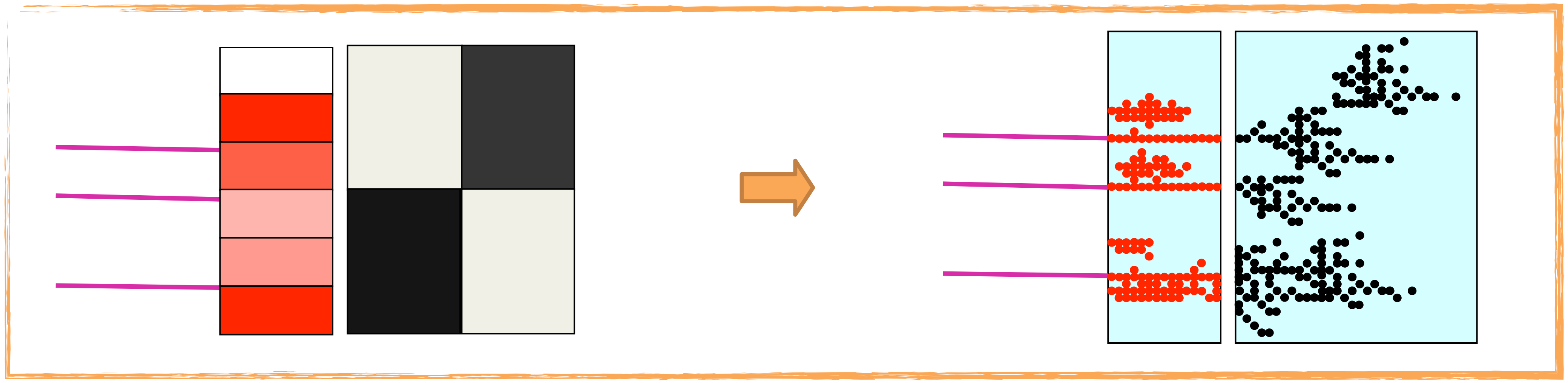
Fine granularity Particle Flow Calorimetry: reconstruct individual particles.

- Charged particle momentum measured in tracker (essentially perfectly)
- Photon energies measured in ECAL: $\sigma_E/E < 20\% / \sqrt{E(GeV)}$
- Only neutral hadron energies (10% of jet energy) measured in HCAL: much improved resolution.

Fine Granularity Particle Flow

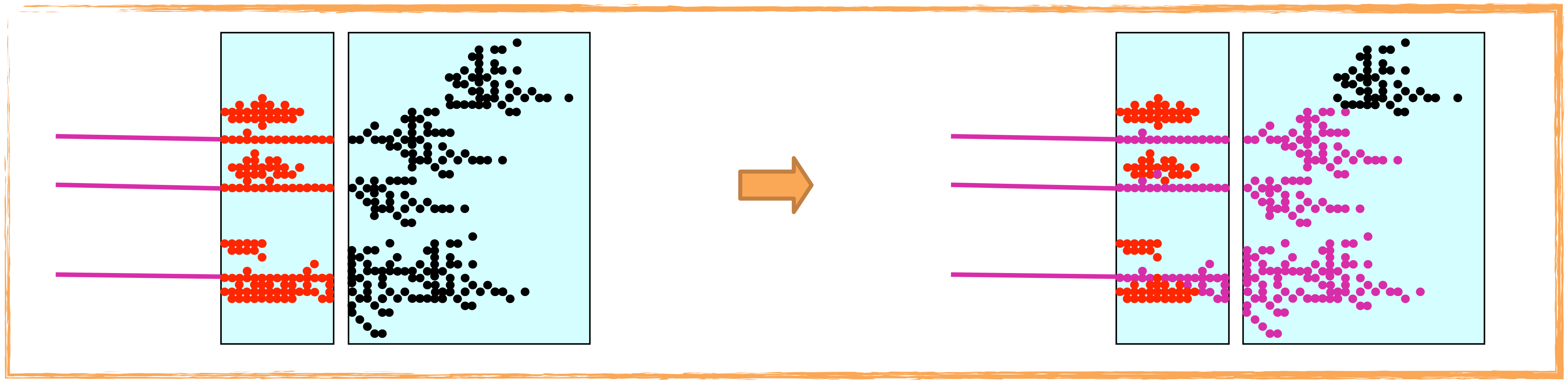
Hardware: need to be able to resolve energy deposits from different particles.

- Require highly granular detectors (as studied by CALICE).



Software: need to be able to identify energy deposits from each individual particle.

- Require sophisticated reconstruction software to deal with complex events, containing many hits.



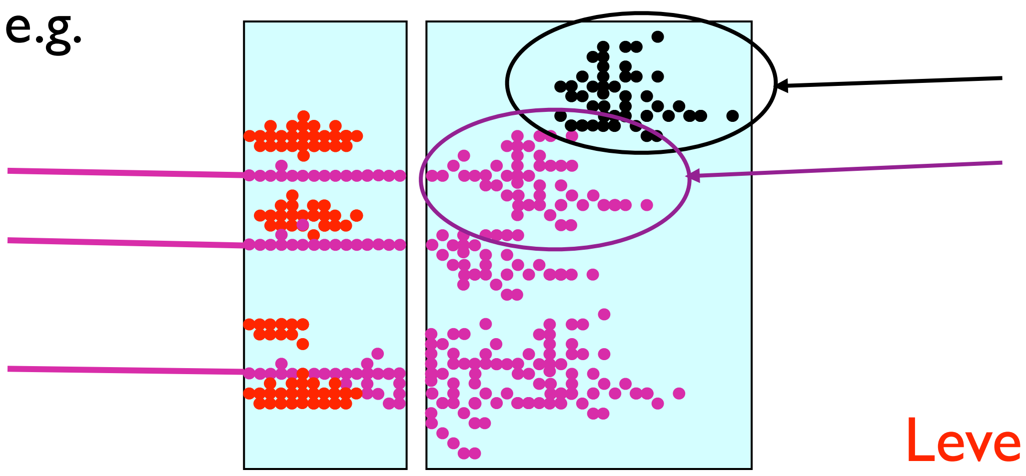
Particle Flow Calorimetry = HARDWARE + SOFTWARE

Particle Flow Algorithms

The challenge for fine granularity particle flow algorithms:

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

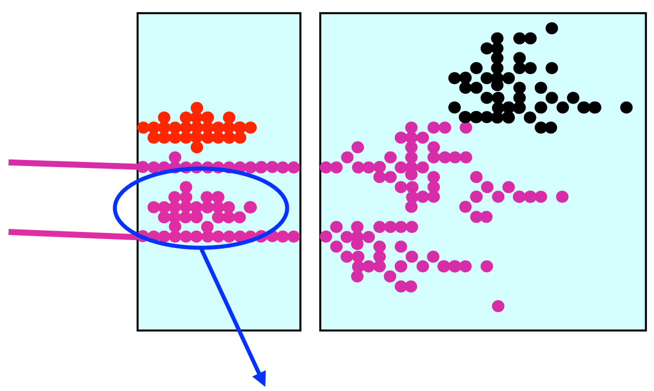
e.g.



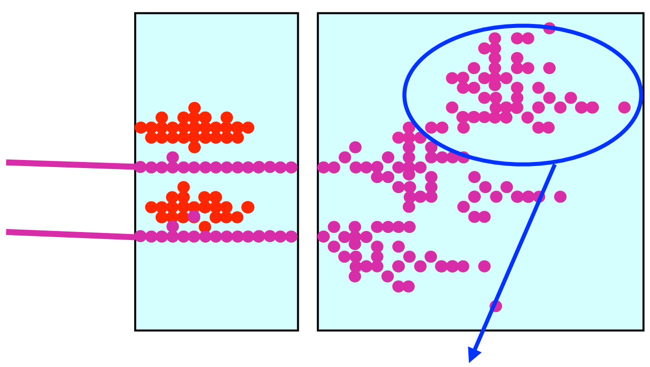
If these hits are clustered together with these, lose energy deposit from this neutral hadron (now part of track particle) and ruin energy measurement for this jet.

Level of mistakes, "confusion", determines jet energy resolution, not intrinsic calorimetric performance

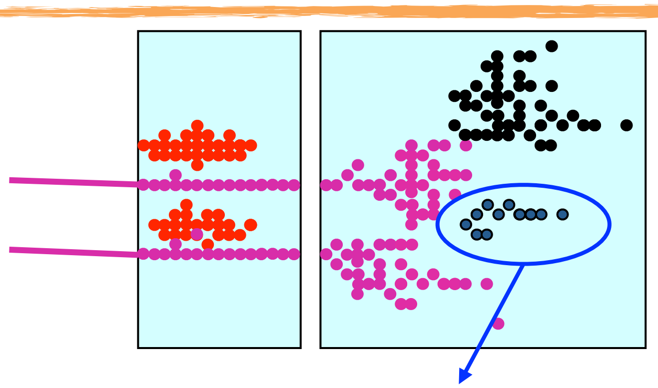
Three basic types of confusion:



Failure to resolve photons



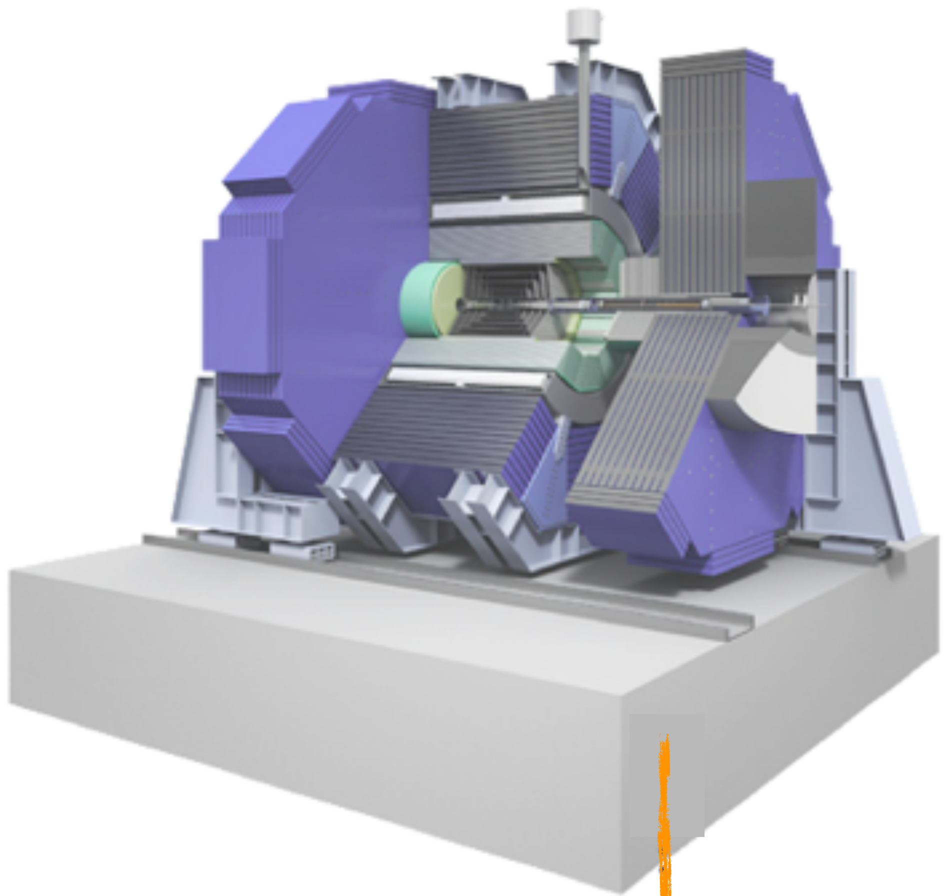
Failure to resolve neutral hadrons



Reconstruct fragments as separate neutral hadrons

LC Detector Concepts

- Fine granularity Particle Flow must be studied in context of whole detector. Need detailed GEANT4 simulations of potential detector designs, e.g. ILC detector concepts:

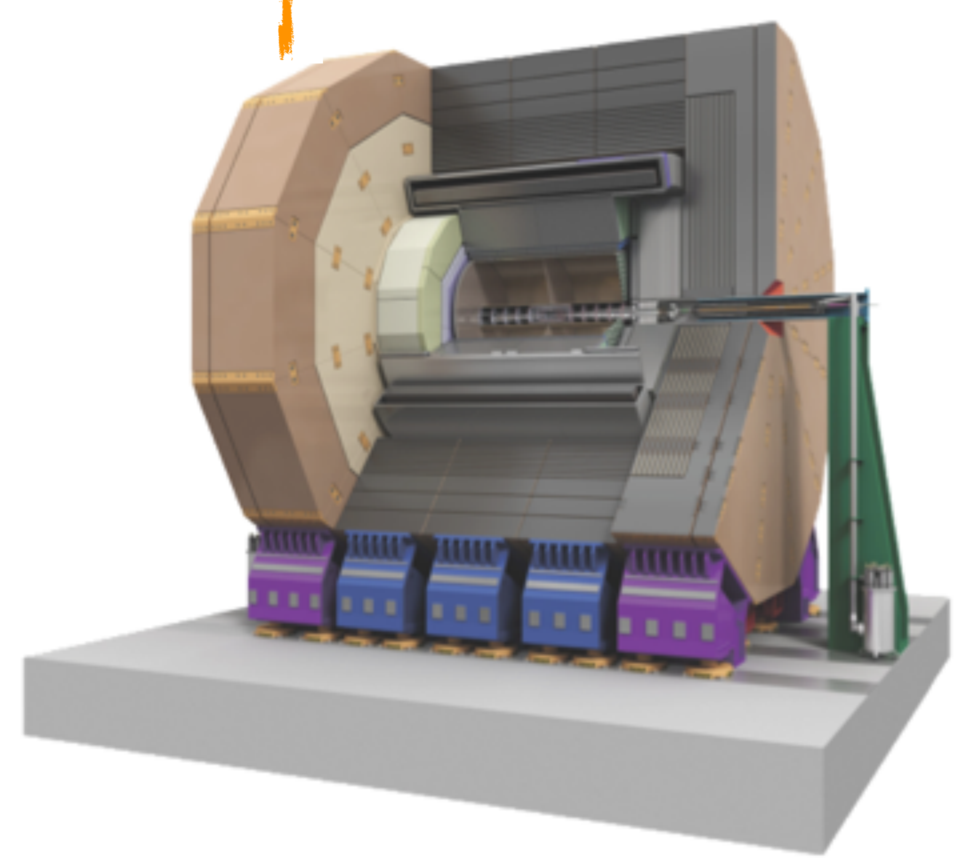


SiD (Silicon Detector)

“Small” : tracker radius 1.2m
B-field : 5 T
Tracker : Silicon (5 layers)
Calorimetry : fine granularity particle flow
ECAL + HCAL inside large solenoid

“Large” : tracker radius 1.8m
B-field : 3.5 T
Tracker : TPC (220 layers)
Calorimetry : fine granularity particle flow
ECAL + HCAL inside large solenoid

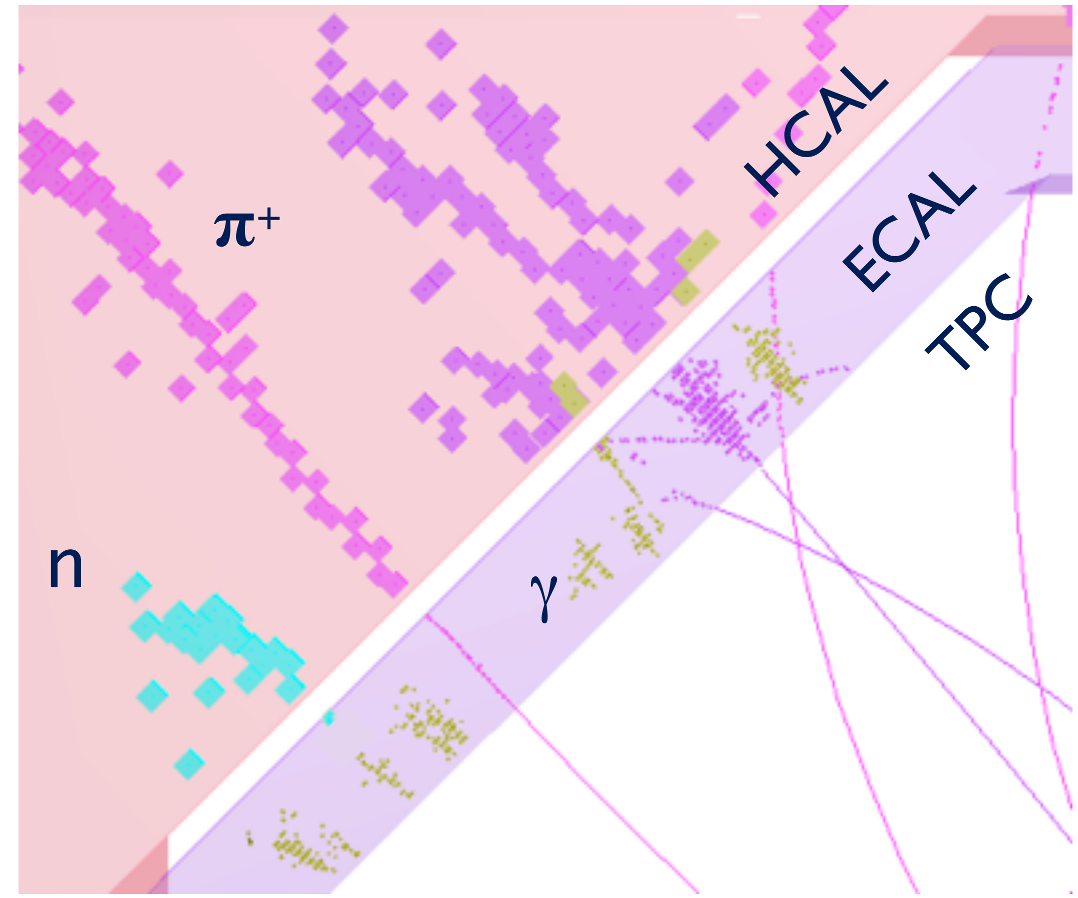
ILD (International Large Detector)



Calorimeter Design

ECAL requirements:

- Minimise transverse spread of EM showers:
  - Small Molière radius & transverse segmentation
- Longitudinally separate EM/Hadronic showers:
  - Large ratio λ_I/X_0
- Identification of EM showers
  - Longitudinal segmentation.



HCAL requirements:

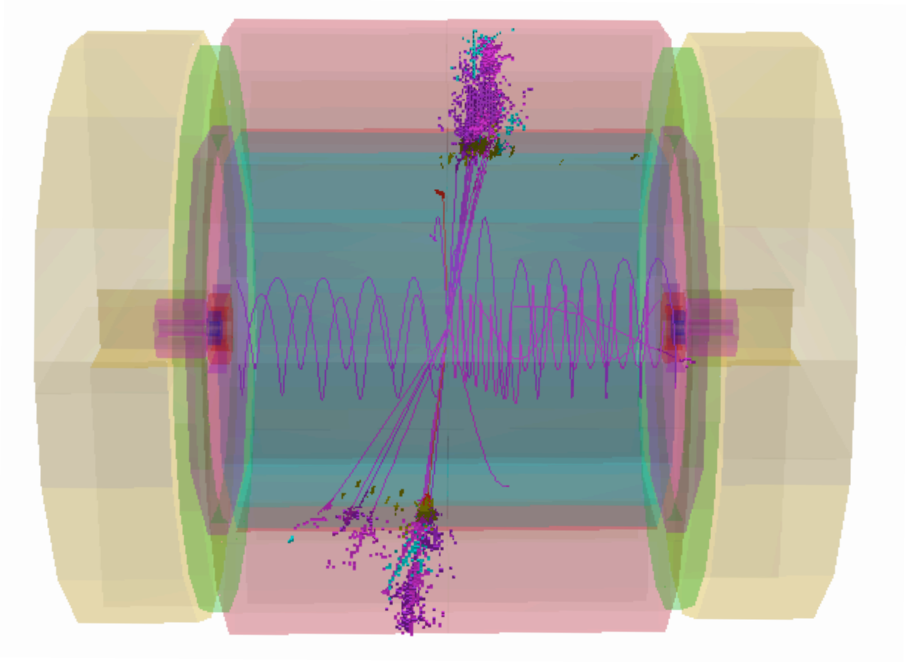
- Fully contain hadronic showers:
  - Small λ_I
- Resolve hadronic shower structure:
  - Longitudinal and transverse segmentation
- HCAL will be rather large:
  - Cost and structural properties important

Suitable absorber materials:

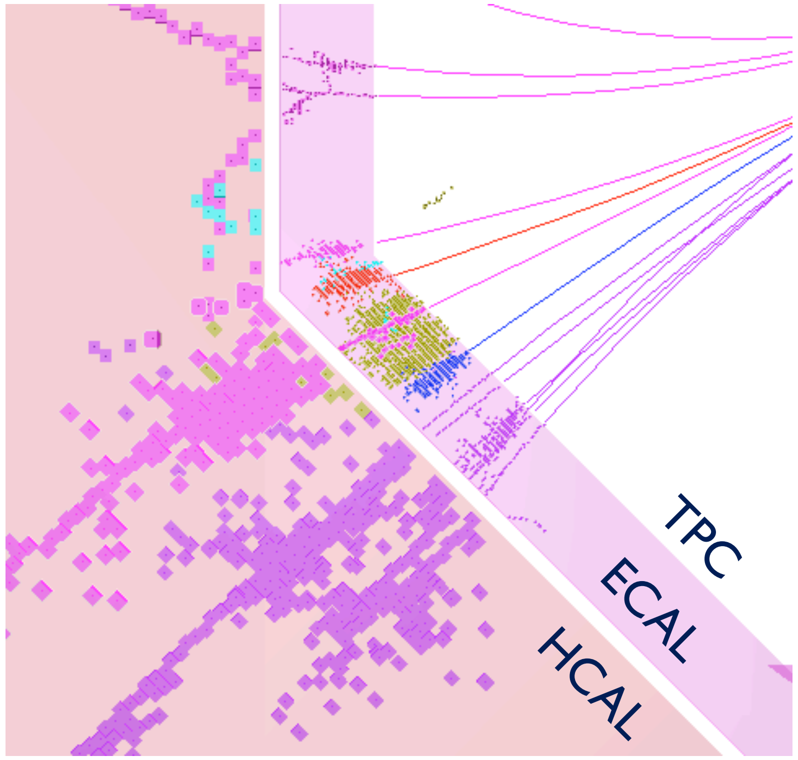
Material	X_0/cm	ρ_M/cm	λ_I/cm	λ_I/X_0
Fe	1.76	1.69	16.8	9.5
Cu	1.43	1.52	15.1	10.6
W	0.35	0.93	9.6	27.4
Pb	0.56	1.00	17.1	30.5

- Fine granularity particle flow calorimetry lives or dies on quality of reconstruction of particles.
- Require high-performance software, in terms of:
  - Algorithmic sophistication, with reliable implementation.
  - CPU/memory usage; these are complex events with many hits.

- Almost all ILC/CLIC studies use code developed with Pandora C++ Software Development Kit.
- Consists of a framework library with carefully designed Application Programming Interfaces.
- Used to implement highly sophisticated particle flow reconstruction algs for LC-style detectors.
- Flexible, reusable with other pat-rec problems.



Typical topologies of simulated 250GeV jets in ILD_o1_v05

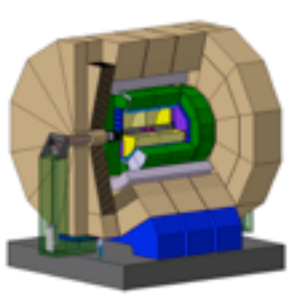


Pandora Particle Flow Reconstruction

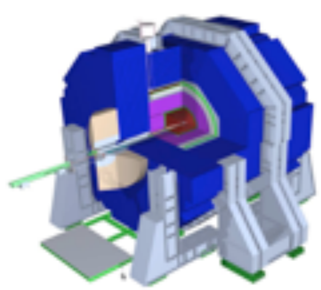
Pandora App

Provides: self-describing tracks and hits

Receives: final particles



ILD/Marlin



SiD/org.lcsim

Pandora Content Libraries

Reusable algs bundled in e.g. FineGranularity and LAr libraries.
Algs registered via app.

APIs

Pandora SDK

Provides: reco. object definitions, manager classes, interface classes, helper functions

Runs registered algs and performs event memory management

Aim: keep algs simple!

APIs

Pandora Algs

Perform particle flow reconstruction.

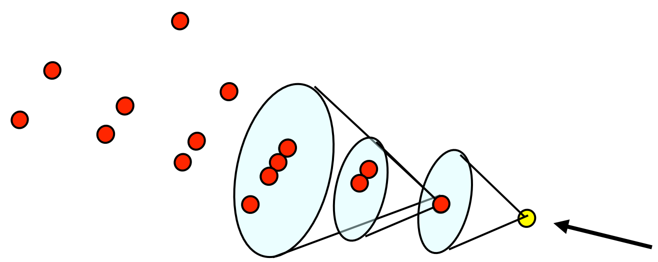
XML-configured.

Use APIs to access or manipulate reco. objects.

Physics-driven code, using SDK services.

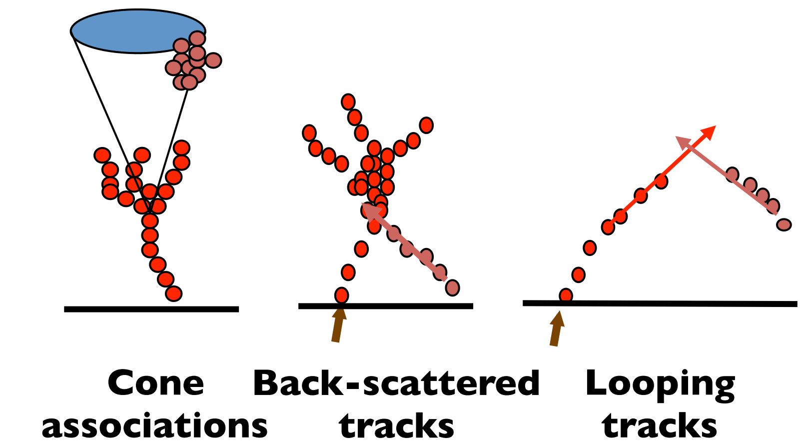
Structure ensures code is efficient, flexible and easy to maintain

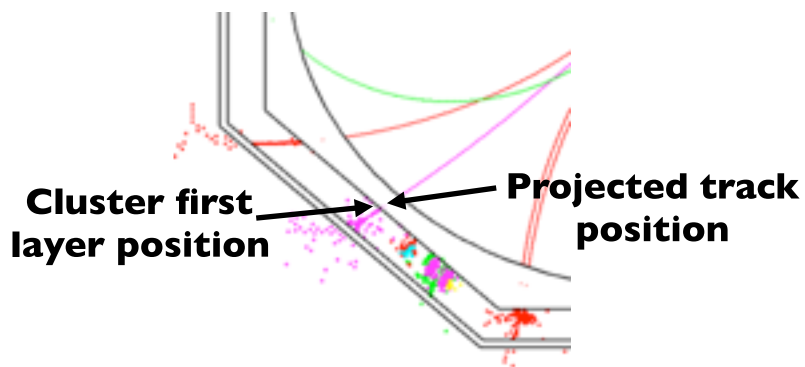
Fine-Granularity Algorithms



ConeClustering Algorithm

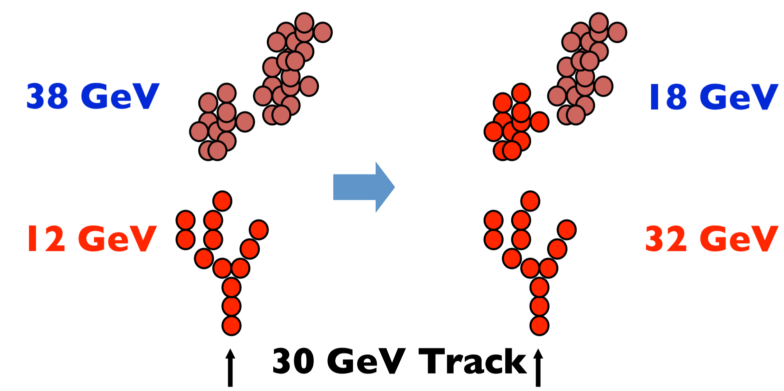
Topological Association Algorithms

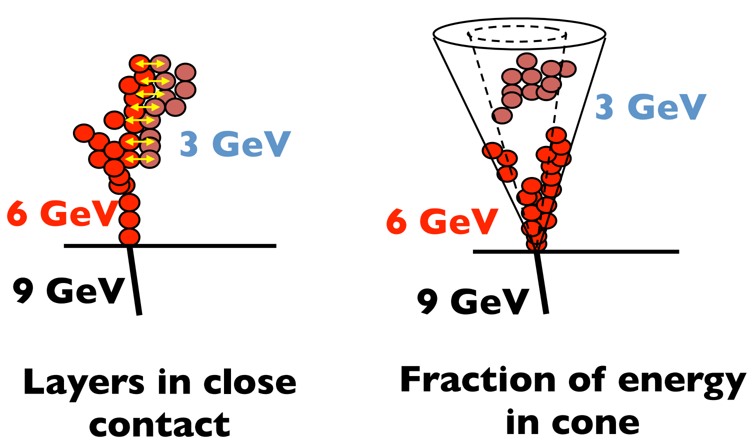




Track-Cluster Association Algorithms

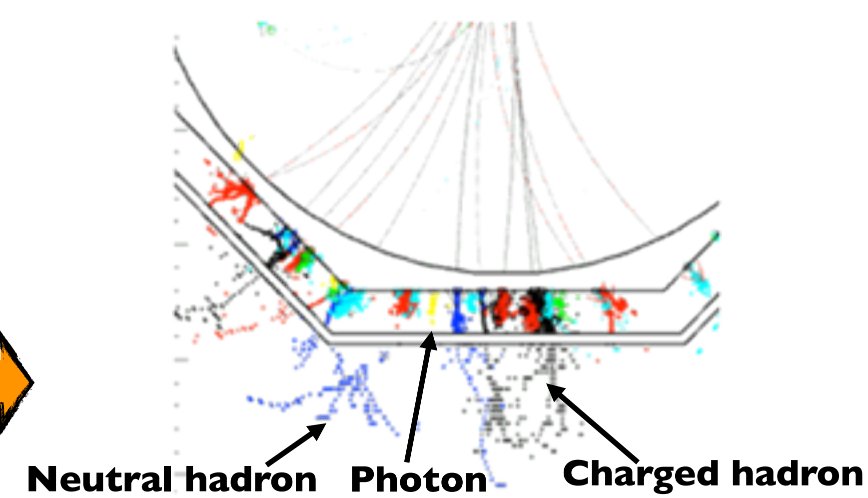
Reclustering Algorithms





Fragment Removal Algorithms

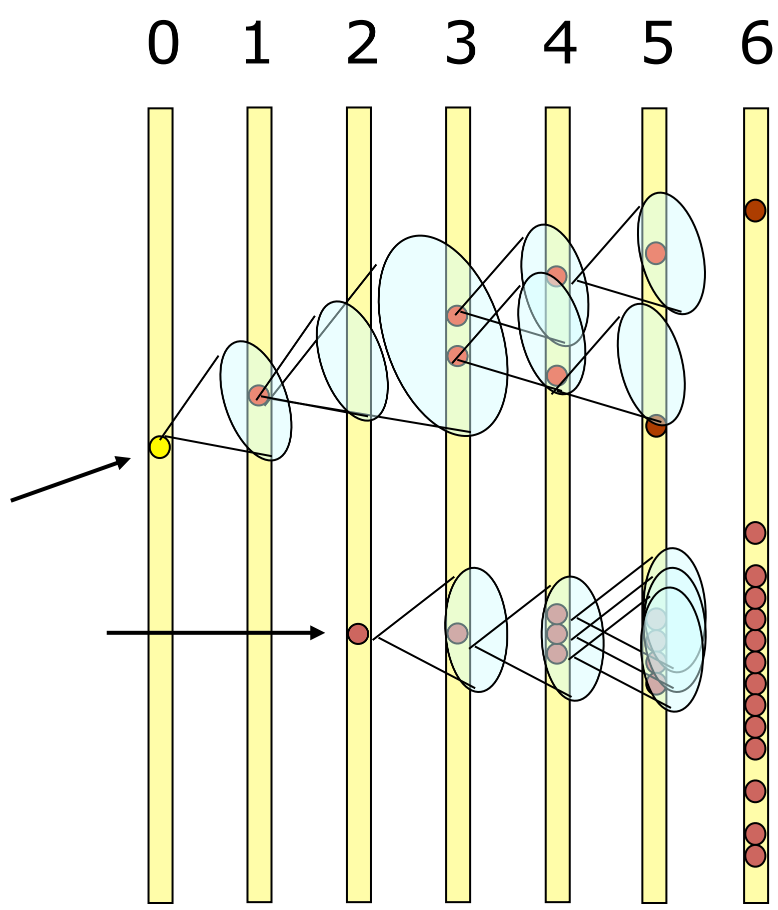
PFO Construction Algorithms



Clustering and Topological Association

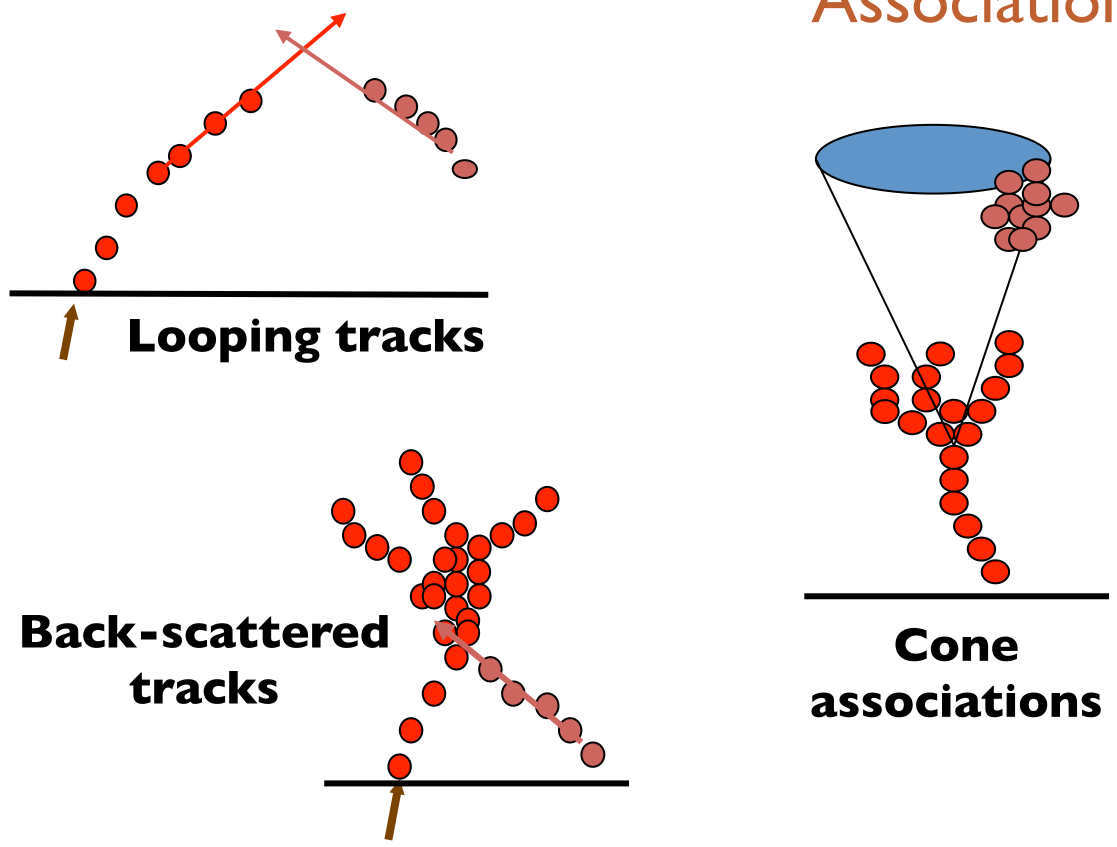
Pandora reconstruction philosophy: "It's easier to put things together than to split them up"

Clustering



- Cone based clustering configured to create clusters that are fragments of single particles, rather than merging deposits from separate particles.

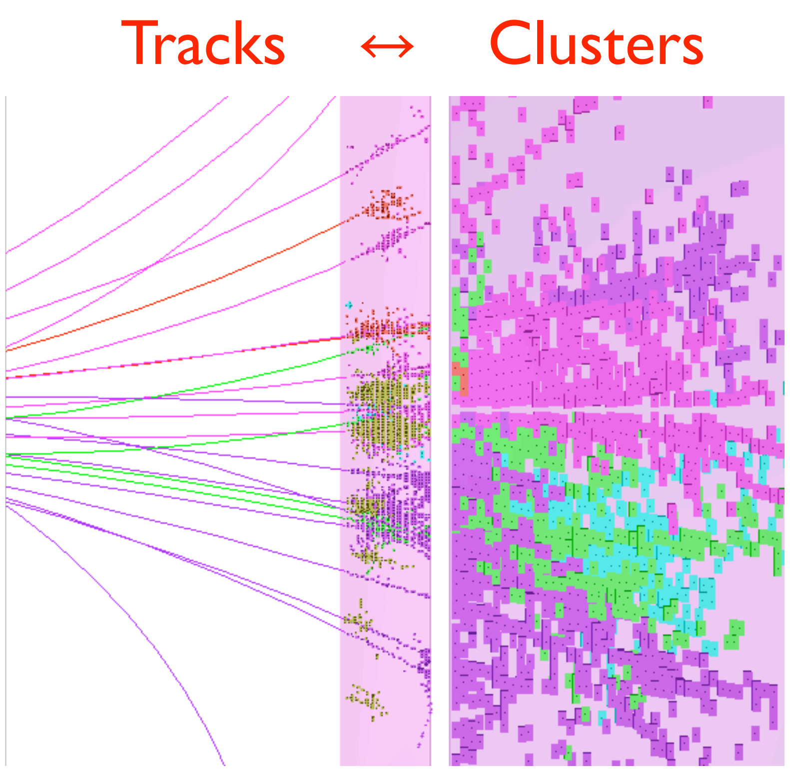
Association

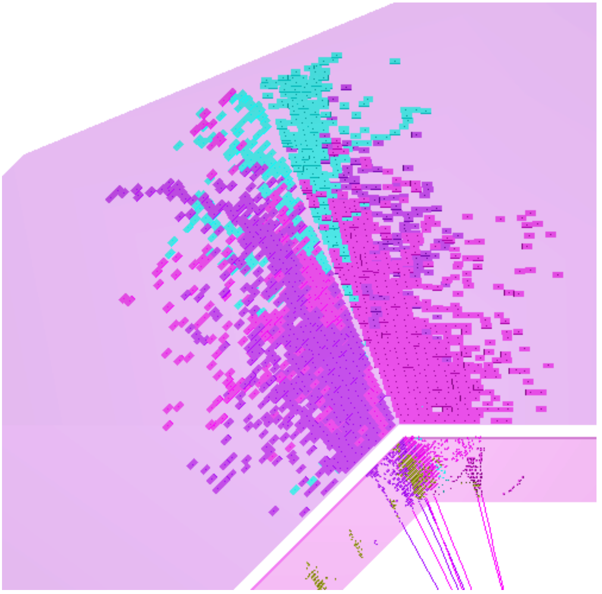


- Fragments merged together by series of algs, each following clear topological rules.
- Fine granularity and tracking capabilities of detector exploited to merge clusters that are clearly associated. Few mistakes made.

Track-Cluster Association

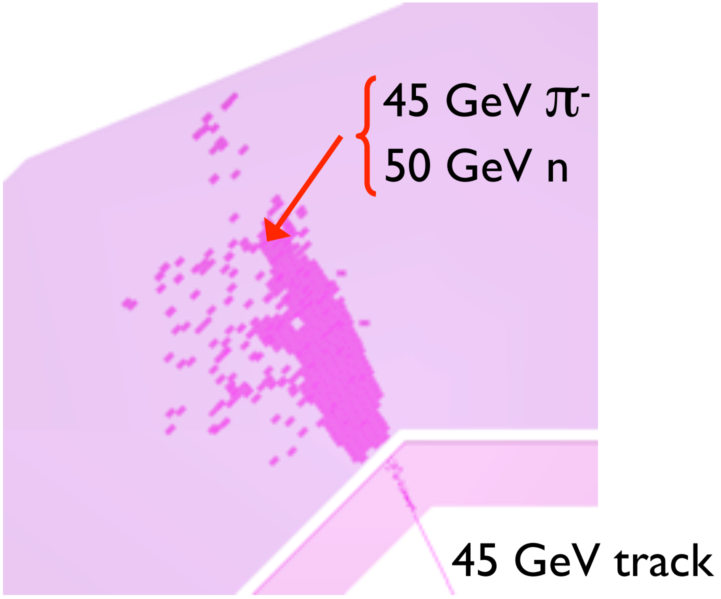
- Track-cluster association algs match cluster positions and directions with helix-projected track states at calorimeter.
- In very high-density jets, reach limit of "pure" particle flow: can't cleanly resolve neutral hadrons in hadronic showers.
- Identify pattern-recognition problems by looking for significant discrepancies between cluster E and track p.
- Choose to recluster: alter clustering parameters or change alg entirely until cluster splits and consistent E/p achieved.





After topological association

Compare E/p values to find problems

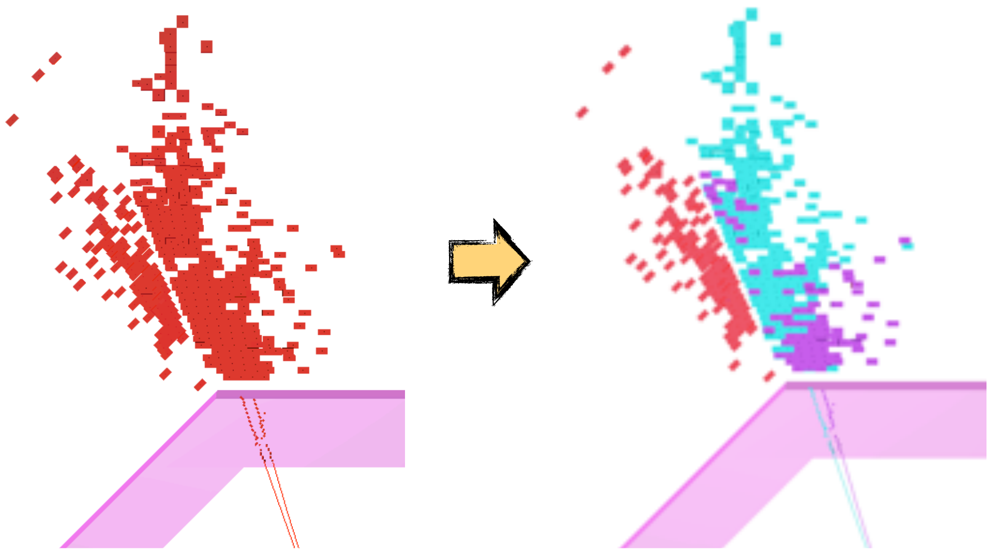


Find n absorbed into π^- cluster

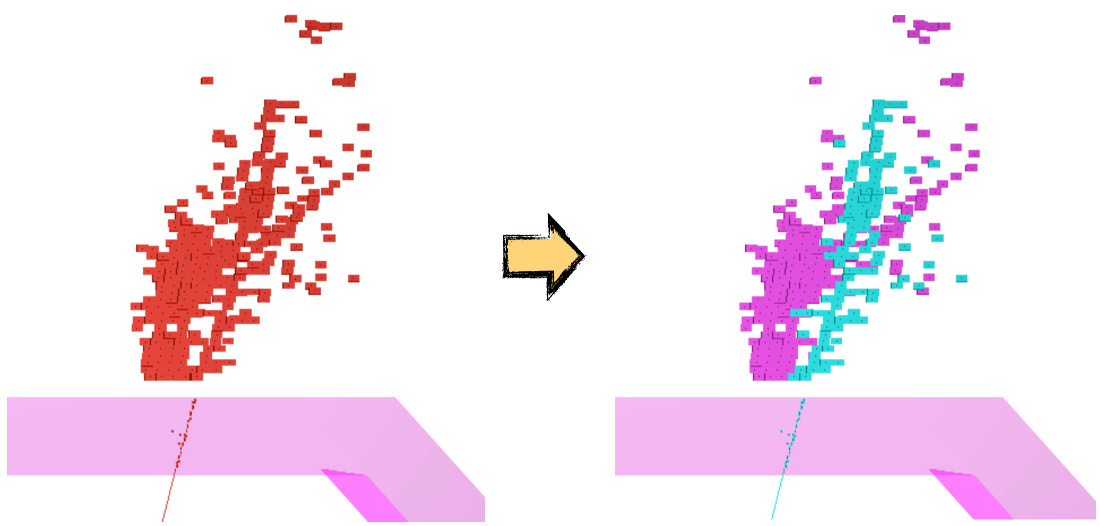
e.g. 45GeV track associated to 95GeV cluster:

identify and address clustering problem

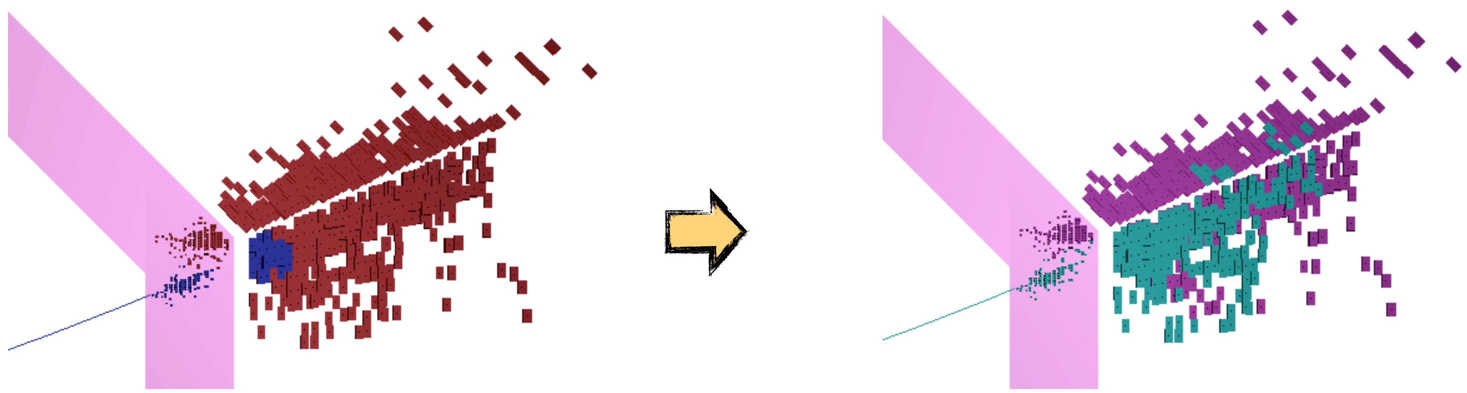
Reclustering Strategies



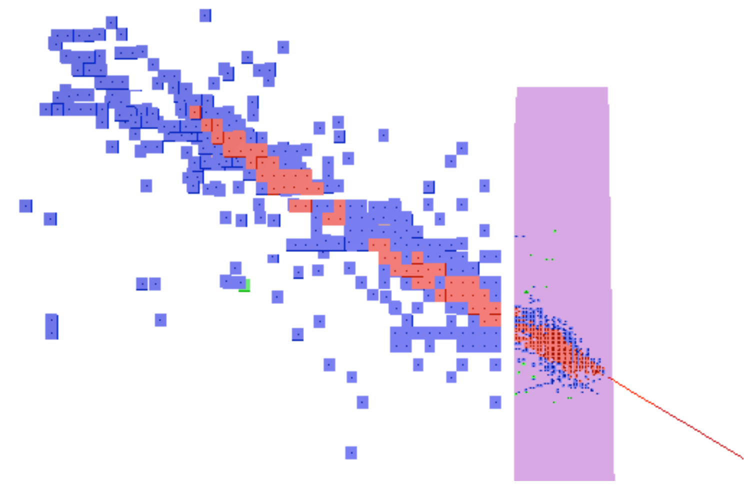
1. Multiple tracks associated to single cluster – split cluster.



2. Cluster energy much greater than track momentum – split cluster.



3. Track momentum much greater than cluster energy – bring in nearby clusters and reconfigure.

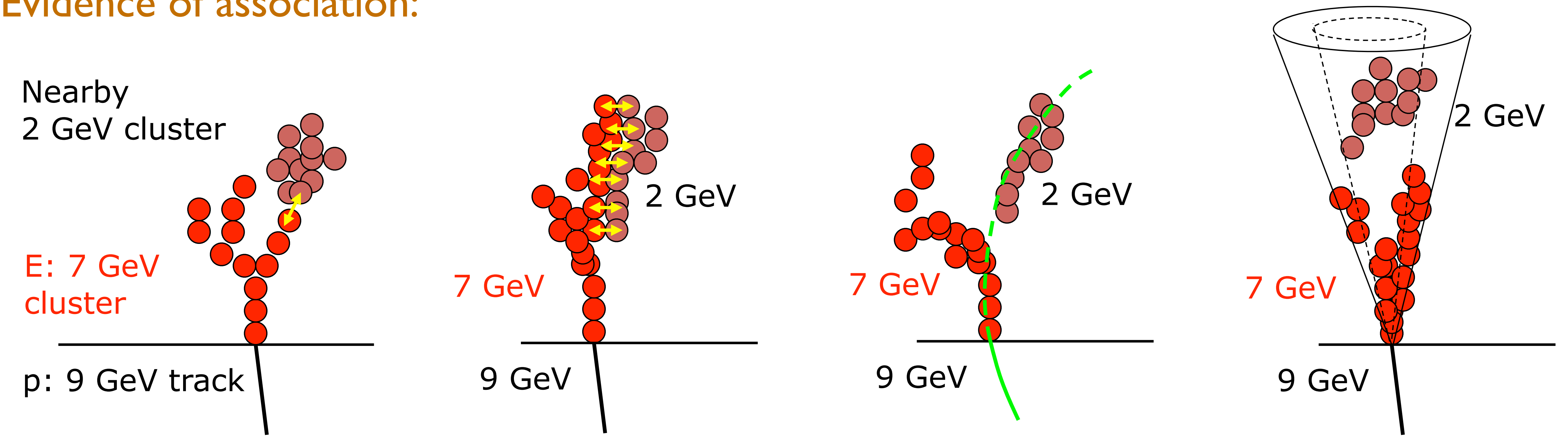


4. If, and only if, no E/p match emerges, can force track-cluster consistency ⇒ energy flow.

Fragment Removal

- Fragment removal algs aim to remove neutral clusters (those without track-associations) that are really fragments of charged (track-associated) clusters.
- Algs look for evidence of association between nearby clusters, merging the clusters together. In order to merge clusters, the change must bring about a satisfactory change in E/p χ^2.

Evidence of association:



Small distance of closest approach

Multiple layers in close contact

Small distance to track extrapolation

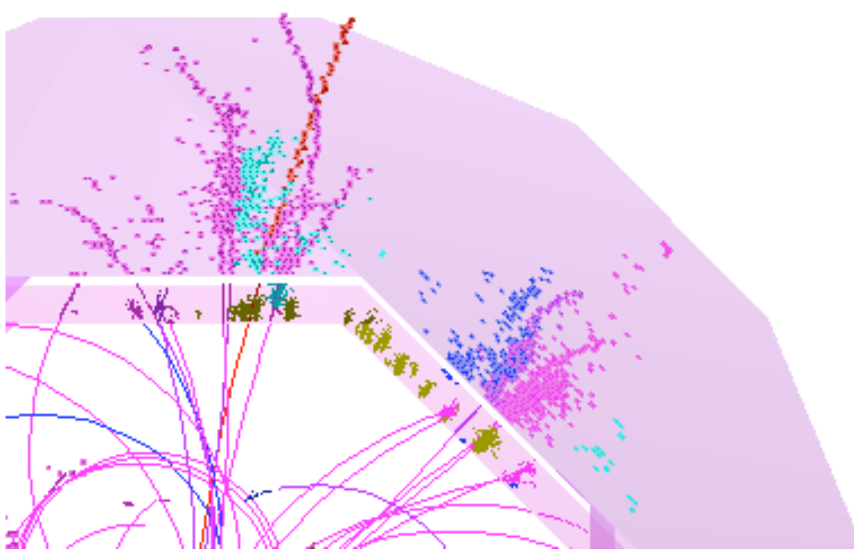
Large fraction of energy in cone

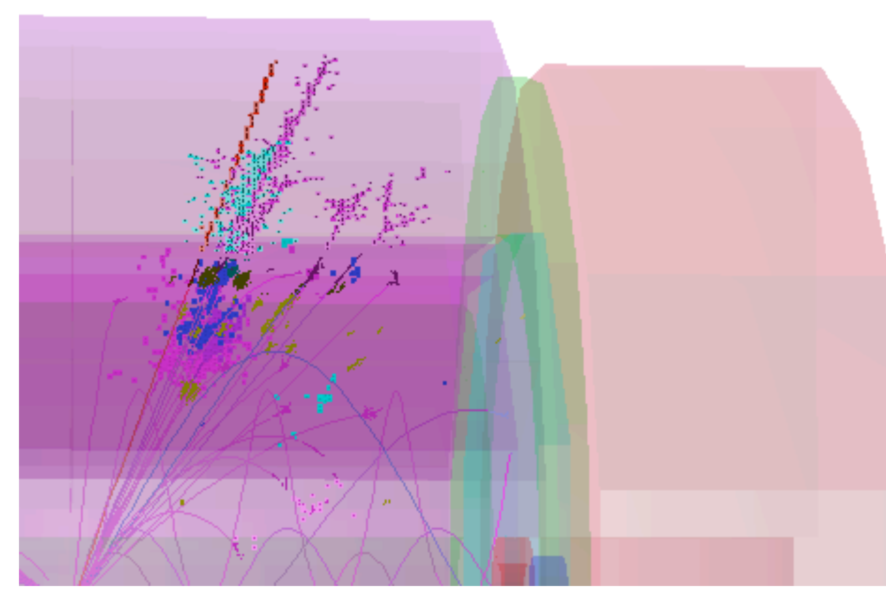
Particle Identification

- Particle ID is crucial for many physics analyses, and photon ID is vital for reconstruction of jet energies in non-compensating calorimeters. Currently available: charged lepton and photon ID.
- Some algs can perform dedicated reconstruction of specific particle types before standard reconstruction. Removal these particles from the event then helps to reduce confusion.

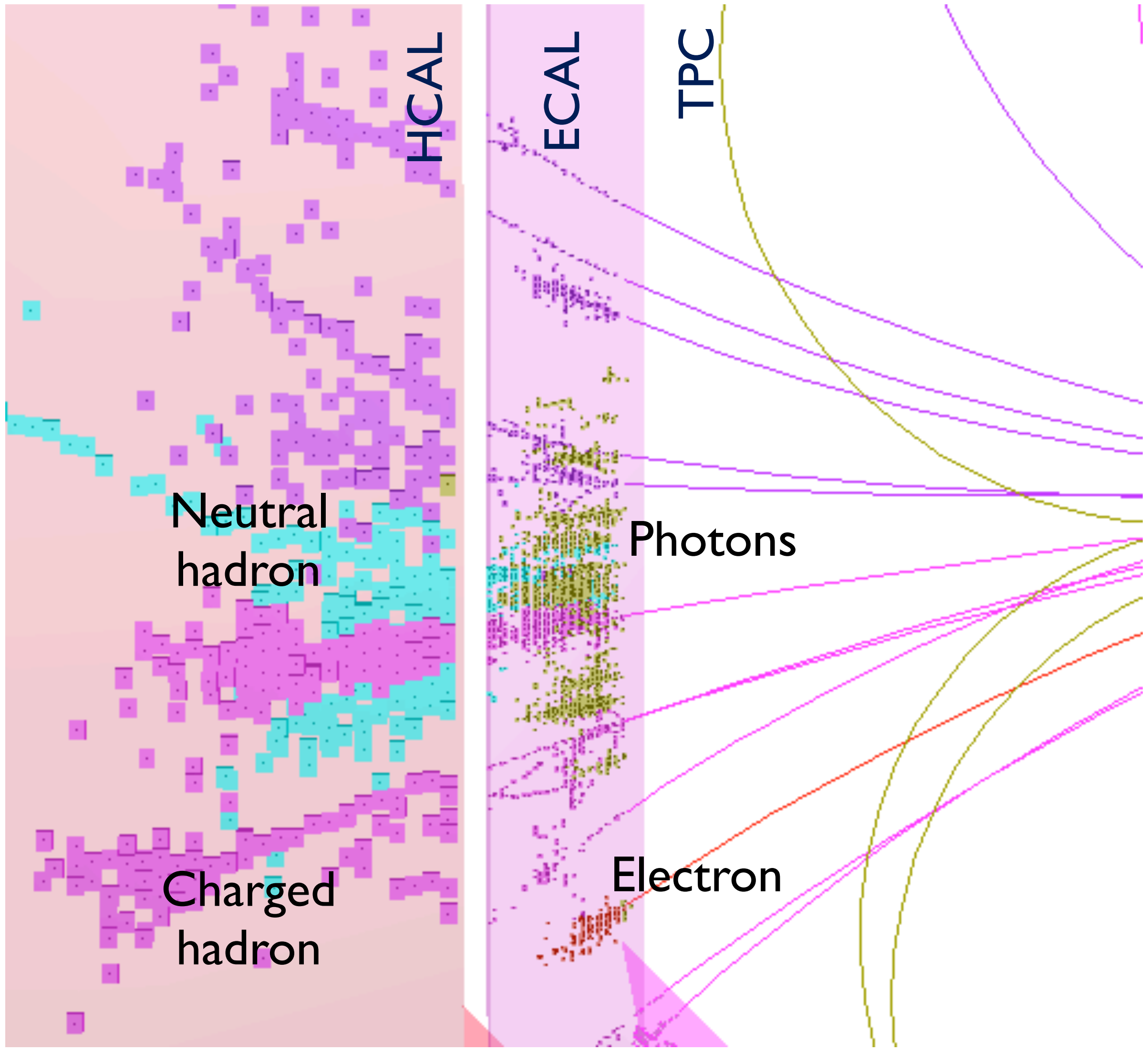
e.g. dedicated muon alg.

1. **Cluster hits in muon yoke**
2. **Associate to inner detector track**
3. **"Swim" through calorimeter**





Typical 250GeV Jet in ILD_o1_v05:



Particle flow objects (PFOs) built from tracks and (associated) clusters using set of simple rules:

- Obtain list of reconstructed particles, with energies and particle ID.
- Calorimeter energy resolution not critical – most energy from tracks.
- Level of mistakes in building particles dominates jet energy resolution.
- Proceed by building jets and studying physics performance.

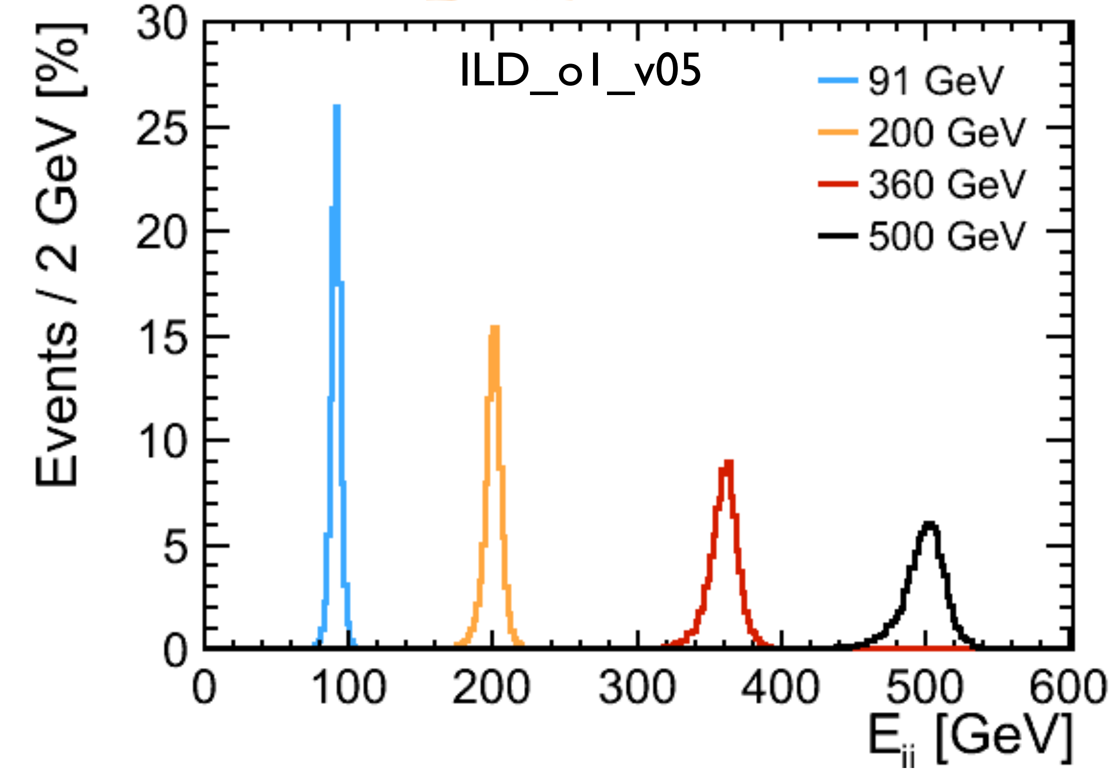
Can now assess performance of fine granularity particle flow using simulation...

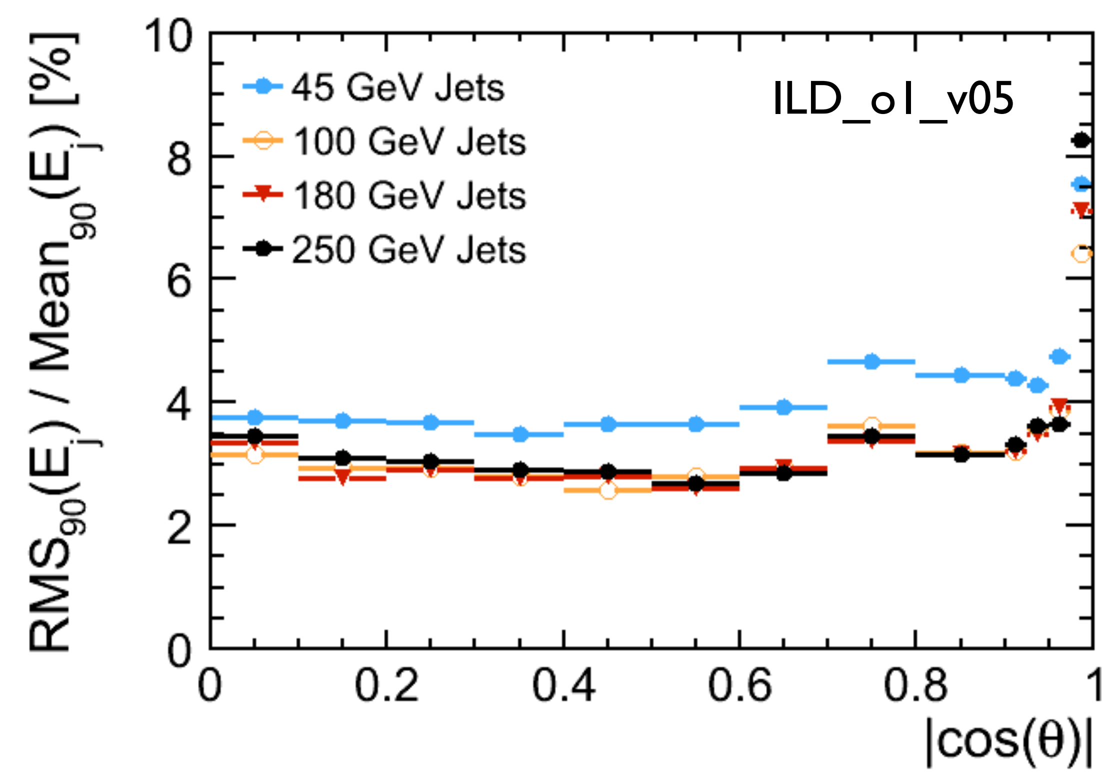
Jet Energy Resolution: ILC

- Recall motivation for fine granularity particle flow:
  - Jet energy resolution: $\sigma_E/E < 3.5\%$
- Benchmark performance using jet energy resolution in Z decays to light quarks.
- Use total energy to avoid complications of jet finding and no backgrounds included.
- Current performance, full GEANT4 simulations:

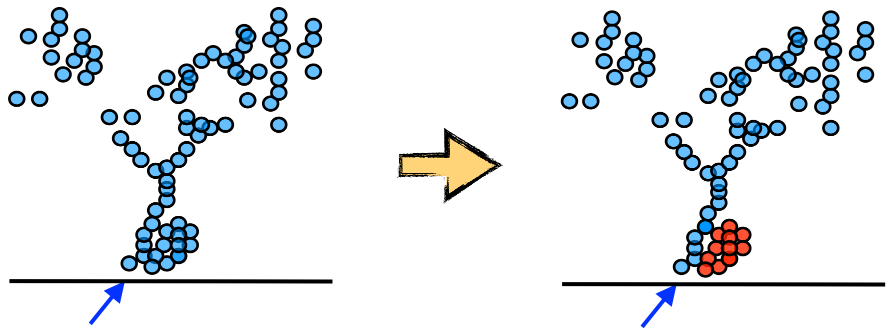
E_j	$RMS_{90}(E_j)$ / $mean_{90}(E_j)$
45 GeV	3.7%
100 GeV	2.8%
180 GeV	2.9%
250 GeV	2.9%

$$\frac{RMS_{90}(E_j)}{mean_{90}(E_j)} = \frac{RMS_{90}(E_{jj})}{mean_{90}(E_{jj})}\sqrt{2}$$



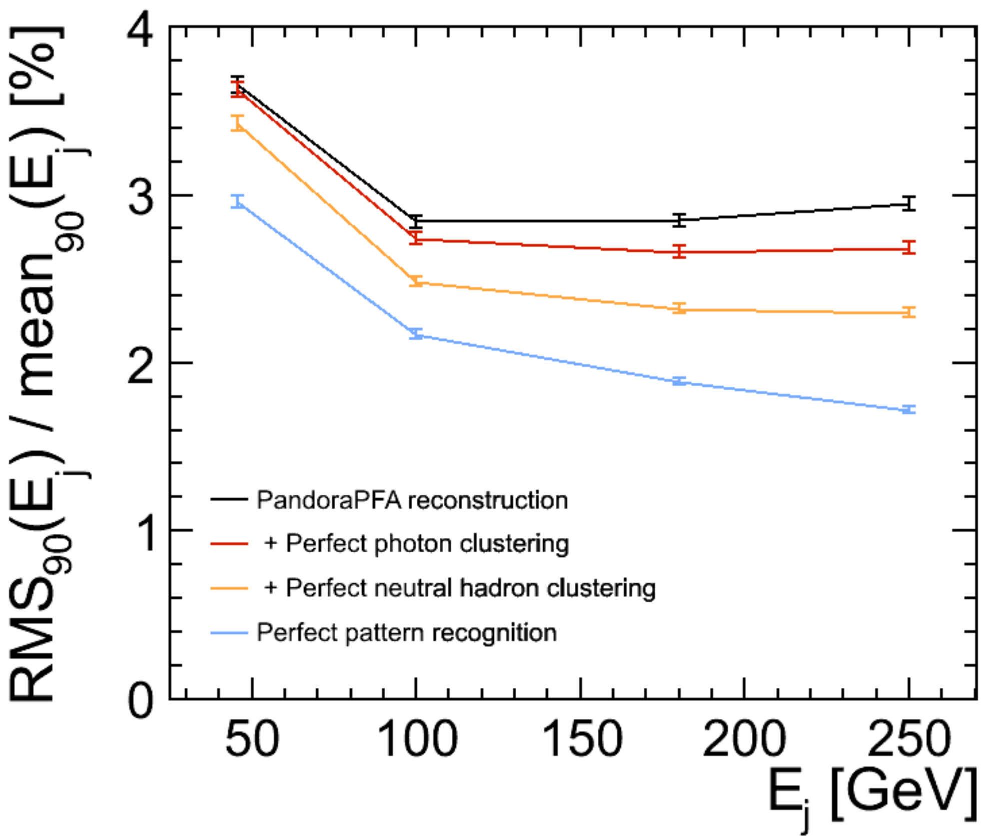


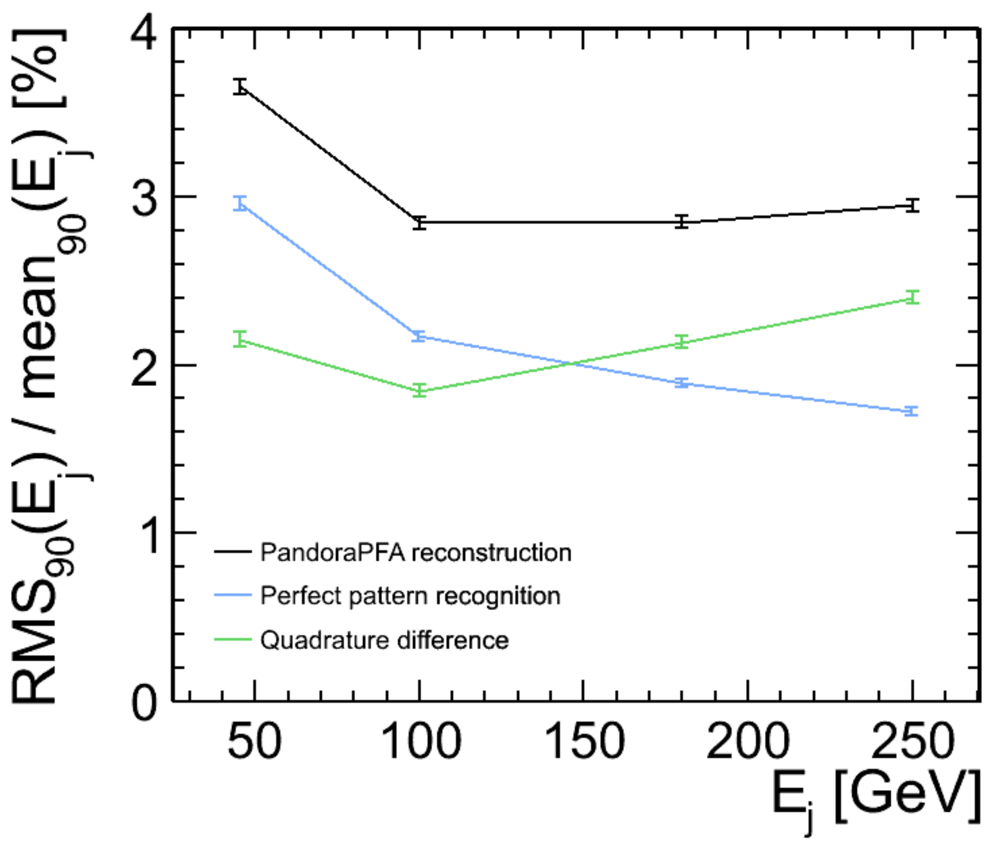
- Switch some standard algs with MC cheating versions to understand resolution:



e.g. Perfect photon reconstruction

- Main performance driver varies with energy:
  - Low energy jets: resolution
  - High energy jets: confusion
  - Cross-over between 100 and 180 GeV
  - Very high energy: leakage will be important





Simulation of Hadronic Showers

- Know that modelling of hadronic showers is far from perfect, so can we believe PFA results?
- Previously compared PandoraPFA/ILD performance using 5 very different GEANT4 physics lists:

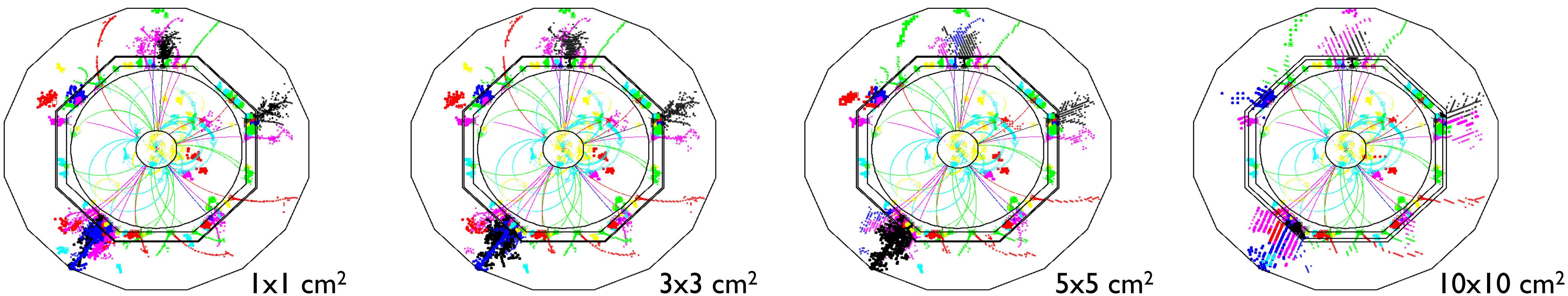
Older results, 2010

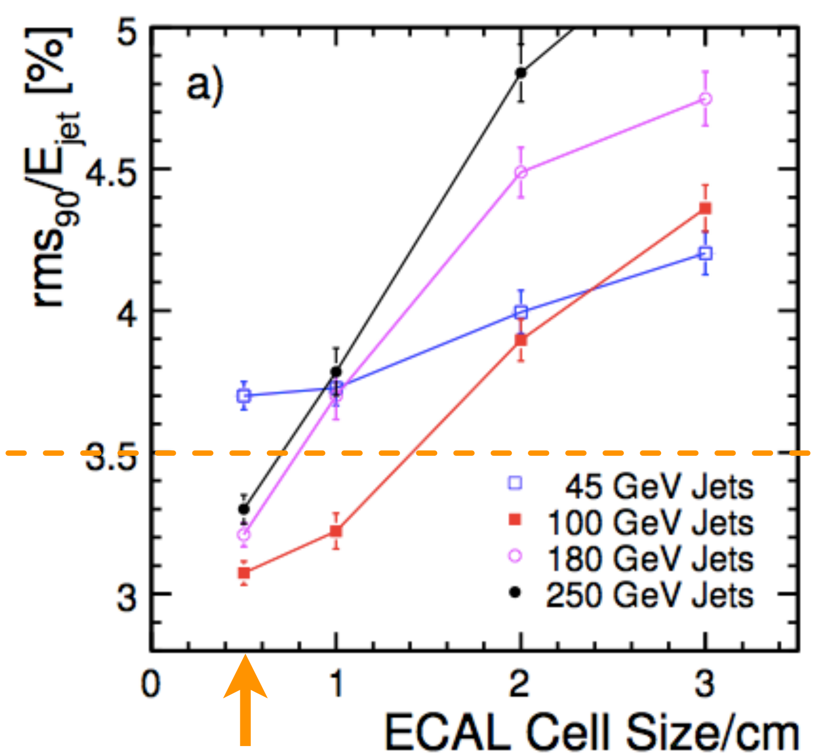
Physics List	Jet Energy Resolution			
	45 GeV	100 GeV	180 GeV	250 GeV
LCPhys	3.74 %	2.92 %	3.00 %	3.11 %
QGSP_BERT	3.52 %	2.95 %	2.98 %	3.25 %
QGS_BIC	3.51 %	2.89 %	3.12 %	3.20 %
FTFP_BERT	3.68 %	3.10 %	3.24 %	3.26 %
LHEP	3.87 %	3.15 %	3.16 %	3.08 %
rms	4.2 %	3.9 %	3.5 %	2.5 %

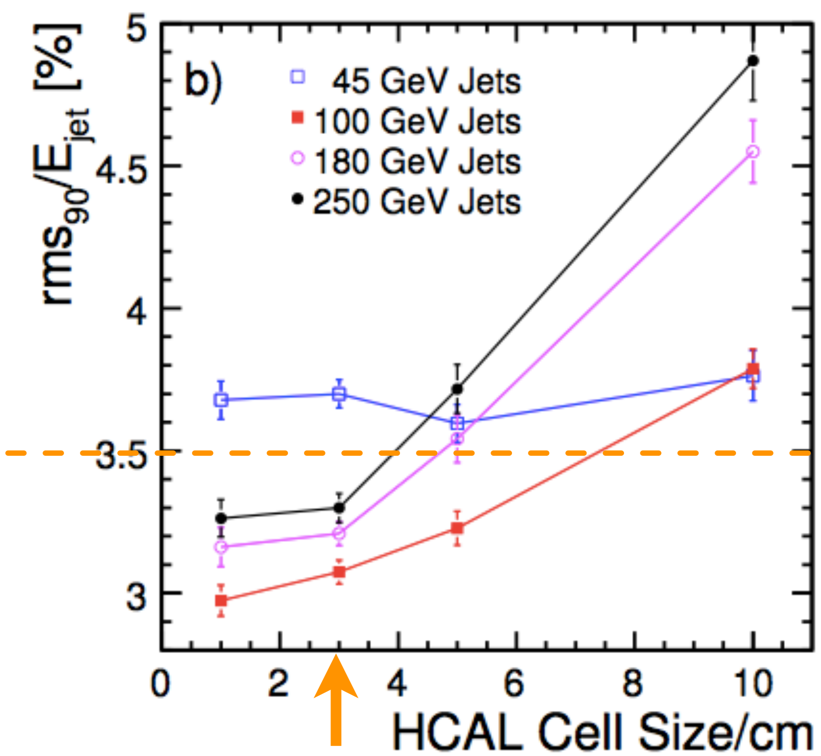
- Only a weak dependence < 5% (on the total resolution, not just the hadronic confusion term)

Study suggests Particle Flow is rather robust to modelling of hadronic showers

- To assess granularity requirements, vary ECAL Si pixel size and HCAL tile size in ILD, then examine jet energy resolutions obtained with particle flow reconstruction. e.g. HCAL tiles:







LC Goal

Older results, 2010

CLIC Background Suppression

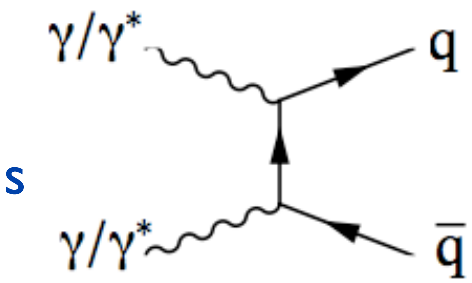
1. CLIC 3TeV input to reconstruction:

- ECAL 10ns → PandoraPFA
- Fe HCAL EndCap 10ns → PandoraPFA
- W HCAL Barrel 100ns → PandoraPFA
- 10ns of tracks → PandoraPFA

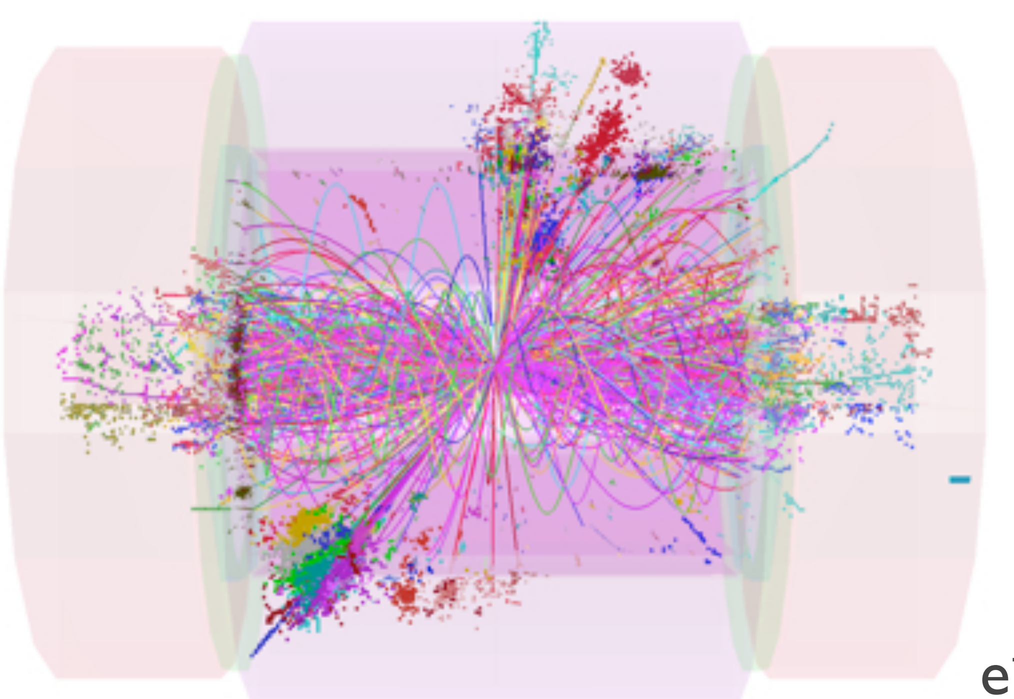
From ILC to CLIC

- $\sqrt{s}$=3TeV: detector occupancies increase and particle flow more difficult.
- Increase in beam-induced backgrounds, with a bunch spacing of only 0.5 ns.

Pile-up of "mini-jets":
20 BXs = 10ns of $\gamma\gamma \rightarrow$ hadrons



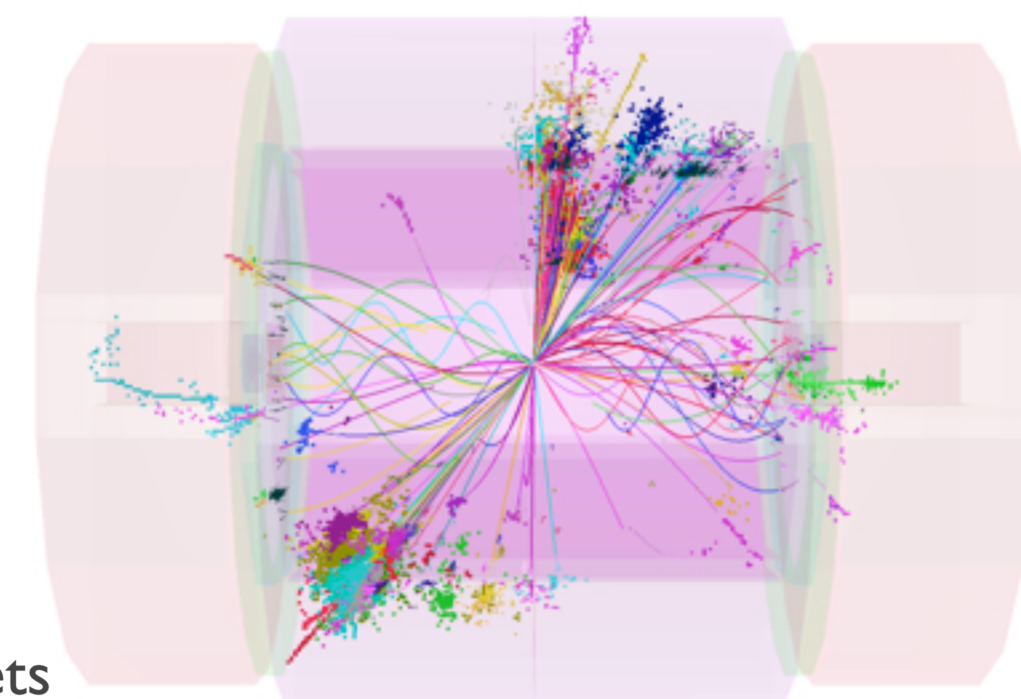
2. Reconstructed particles, bkg energy 1.2TeV:



Apply timing and p_T cuts to reject background PFOs

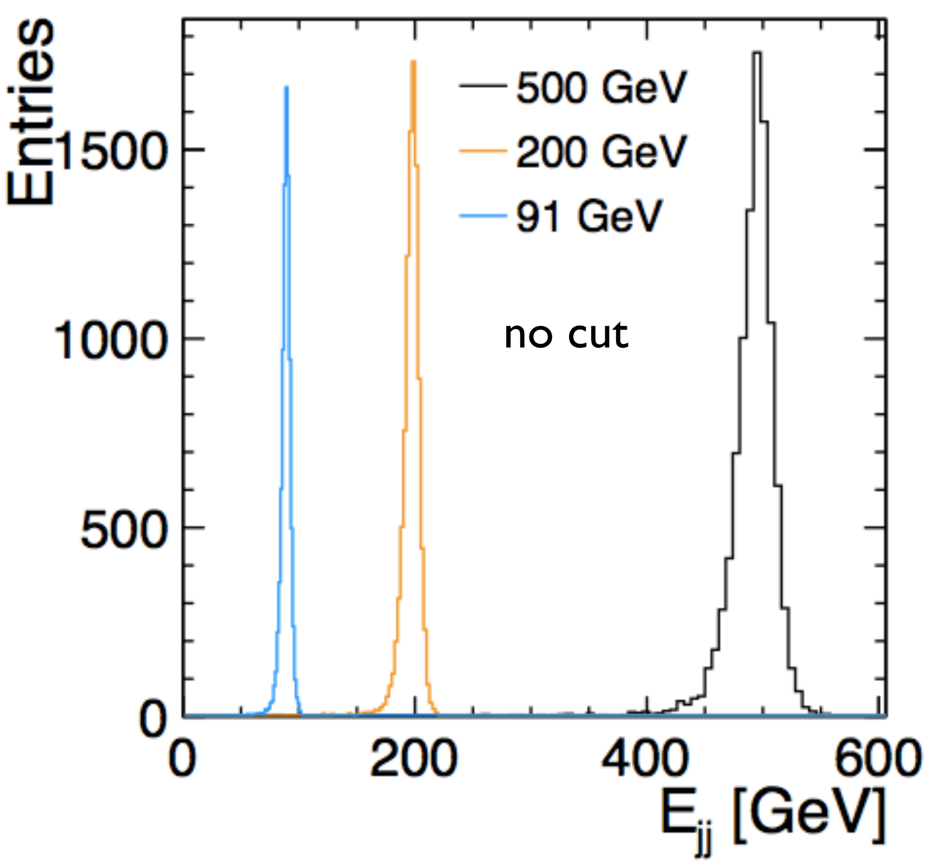
$e^+e^- \rightarrow H^+H^- \rightarrow t\bar{b}b\bar{t} \rightarrow 8$ jets

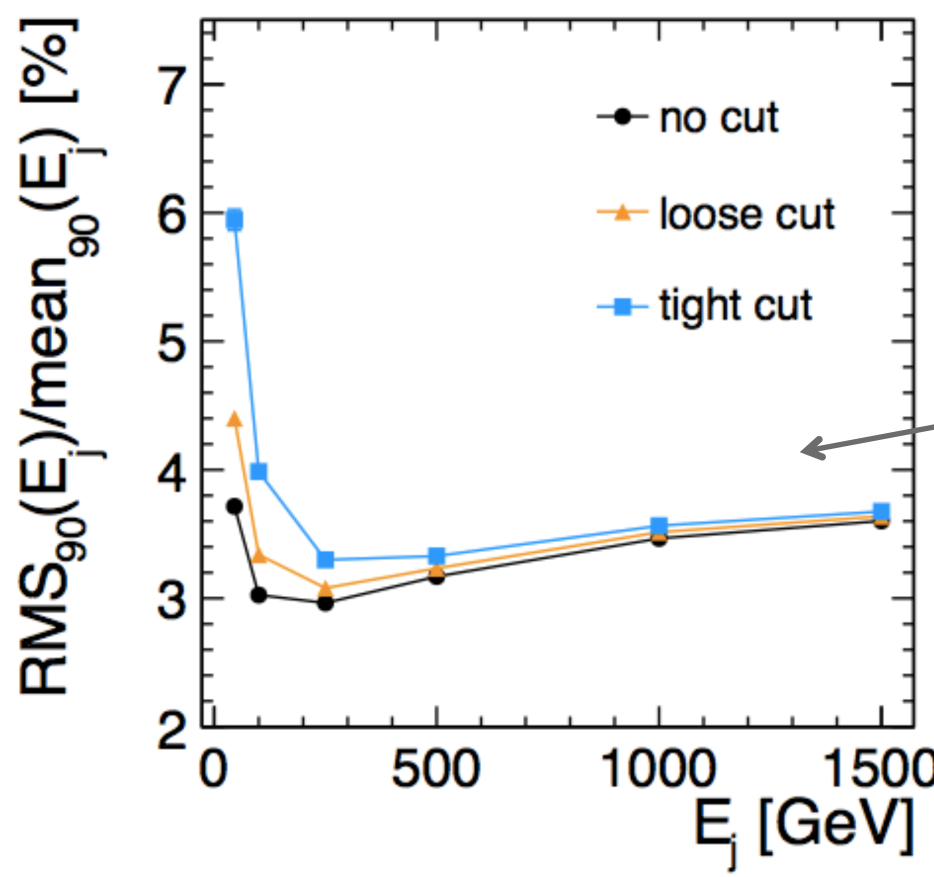
3. Selected particles, bkg energy 85GeV:



Jet Energy Resolution: CLIC

- To assess jet energy resolution, and impact of PFO selection cuts, use samples of Z decays to light quarks without any overlaid backgrounds. Consider jet energies in range 45-1500GeV.
- At low energies, PFO selection cuts have significant impact on jet energy resolution. At higher jet energies, the jet energy reconstruction performance is basically unaffected by the cuts.





Jets in barrel region only, $|\cos(\theta)| < 0.7$

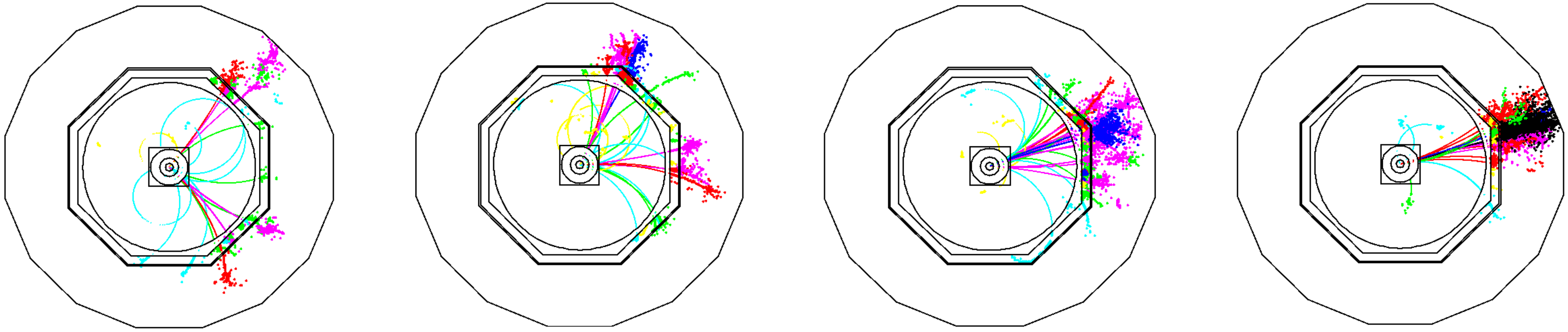
No cuts: resolution better than ~3.7% for all energies considered

CLIC_ILD_CDR

W/Z Reconstruction

- Return to an important aim of fine granularity particle flow calorimetry and examine ability to separate W/Z hadronic decays via di-jet invariant mass reconstruction at CLIC.
- On-shell W/Z decay topology depends on energy, so obtain "mono-jet" topology at high energies:
  - Particle multiplicity does not change
  - Boost means higher particle density

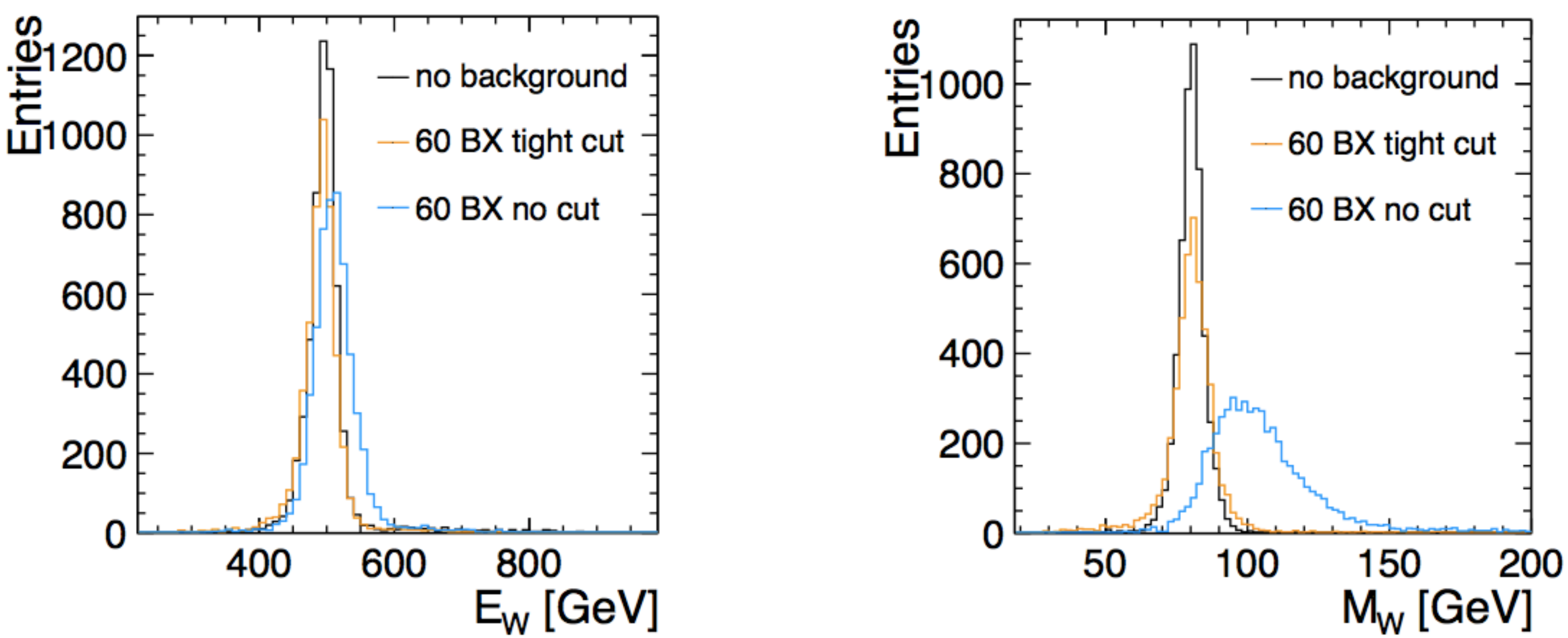
  } more confusion!



125 GeV Z | 250 GeV Z | 500 GeV Z | 1 TeV Z

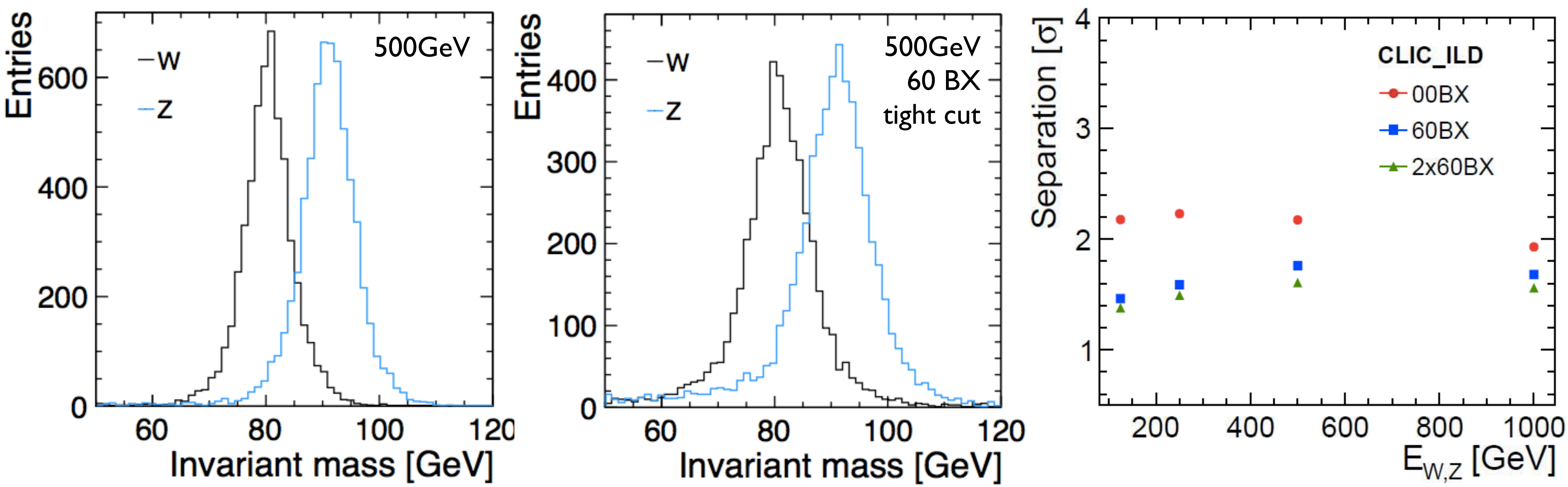
W Reconstruction

- W samples provided by $e^+e^- \rightarrow WW \rightarrow \mu\nu qq$ events in energy range 125-1000GeV. Used full GEANT4 simulation, PandoraPFA reconstruction and considered different levels of background.
- Additional reconstruction and selection procedures: removal of muon, removal of neutral fragments from background, jet reconstruction (`kt` algorithm) and jet angular selection cuts.



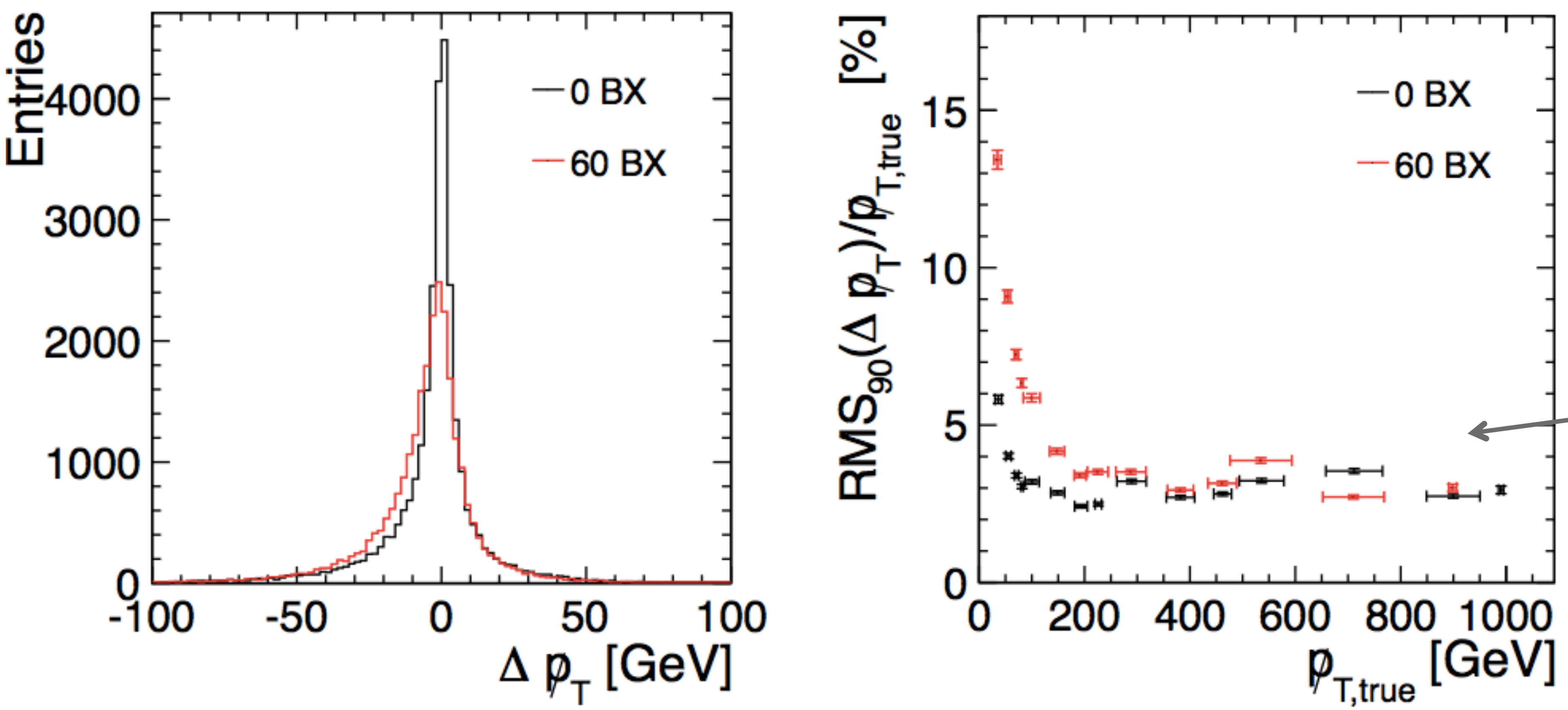
W and Z Separation

- The di-jet mass distributions obtained from the $e^+e^- \rightarrow WW \rightarrow \mu\nu qq$ event samples were then compared with those obtained from $e^+e^- \rightarrow ZZ \rightarrow \nu\nu qq$ event samples.
- Without background a 2σ separation is maintained for W/Z energies between 125-1000GeV. The separation is reduced to about 1.7σ when 60BX of $\gamma\gamma \rightarrow$ hadrons background is included.



Missing p_T Resolution

- Reconstruction of missing momentum important in many physics analyses. The missing p_T resolution was quantified using $e^+e^- \rightarrow ZZ \rightarrow \nu\nu qq$ samples.
- Missing p_T was calculated from the vector sum of the momenta of all particles in the reconstructed jets. This was compared to the generated missing p_T of the two neutrinos.



Missing p_T can be measured with an asymptotic precision of ~3% for true missing p_T>100GeV

CLIC Gaugino Pair Production

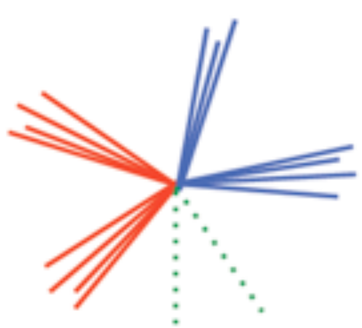
- Have also demonstrated power of fine granularity particle flow in physics analyses, e.g.

$m(\tilde{\chi}_1^0) = 340\,\text{GeV} \quad m(\tilde{\chi}_2^0), m(\tilde{\chi}_1^+) \approx 643\,\text{GeV}$

- Pair production and decay:

$$e^+e^- \rightarrow \tilde{\chi}_1^+\tilde{\chi}_1^- \rightarrow \tilde{\chi}_1^0\tilde{\chi}_1^0 W^+W^-$$
$$e^+e^- \rightarrow \tilde{\chi}_2^0\tilde{\chi}_2^0 \rightarrow hh\,\tilde{\chi}_1^0\tilde{\chi}_1^0 \quad 82\,\%$$
$$e^+e^- \rightarrow \tilde{\chi}_2^0\tilde{\chi}_2^0 \rightarrow Zh\,\tilde{\chi}_1^0\tilde{\chi}_1^0 \quad 17\,\%$$

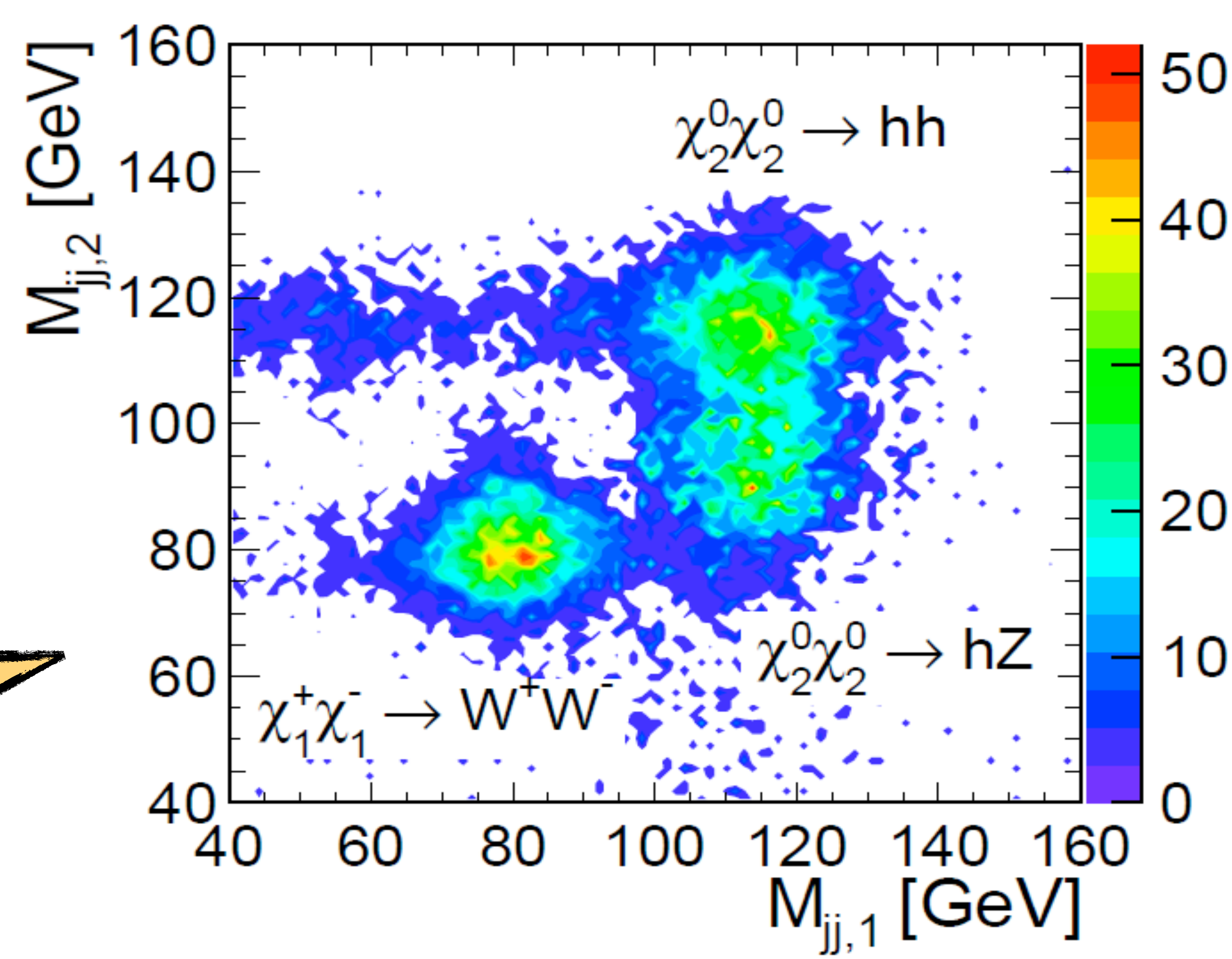
- Largest BR decay has same topology for all final states:



- Separate using di-jet invariant masses:

$$m(\tilde{\chi}_1^\pm) \;:\; \pm 7\,\text{GeV}$$
$$m(\tilde{\chi}_2^0) \;:\; \pm 10\,\text{GeV}$$

Full Simulation with background



Summary

- Fine Granularity Particle Flow Calorimetry is the baseline for the detector at the ILC or CLIC:
  - Such a detector can be built (at a cost).
  - Would provide unprecedented performance.

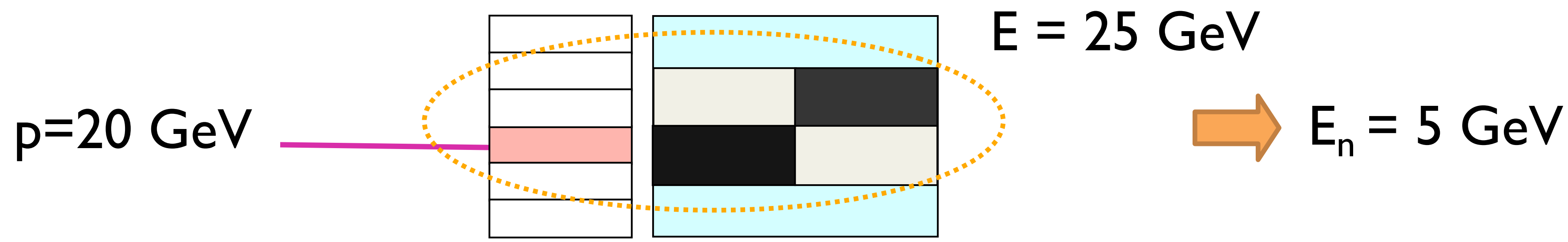
- Pandora Fine Granularity Particle Flow Algorithms:
  - Provide proof of principle over wide range of energies and physics processes.
  - Excellent performance from $\sqrt{s}$ = 500 GeV to $\sqrt{s}$ = 3 TeV.

- Pandora SDK, and many Pandora algorithms, sufficiently generic to be used elsewhere.

Backup Slides

Energy Flow, Particle Flow

- The idea behind particle flow calorimetry is not new, and a similar idea was used by ALEPH:
  - ENERGY FLOW algorithm removes ECAL deposits from identified electrons/photons, leaving (mostly) charged and neutral hadrons.
  - Coarse HCAL granularity means neutral hadrons can only be identified as significant excesses of energy. Neutral hadron energy obtained by subtraction: $E_n = E_{calo} - p_{track}$



$\sigma_E/E \sim 10\%$ jet E resolution for 45 GeV jets

NIM A360:481-506, 1995

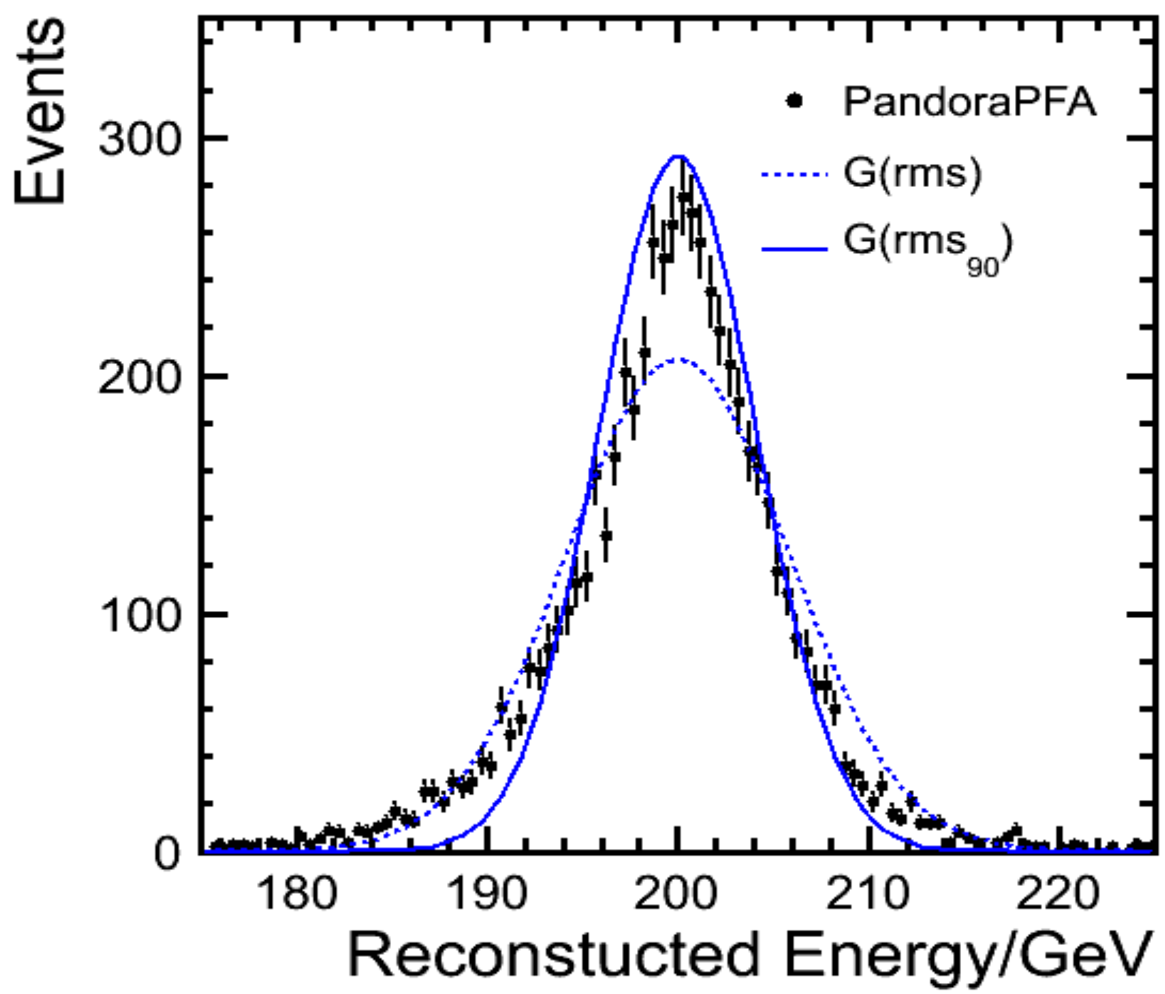
- Similar approach used by a number of other collider experiments.
- FINE GRANULARITY PARTICLE FLOW significantly extends this approach:
  - Now directly reconstruct neutral hadrons.
  - Potentially much better performance.
  - But need highly granular calorimeters and sophisticated software.

PFA Performance

- Particle Flow reconstruction inherently non-Gaussian, so resolution presented in terms of rms_{90}
  - Defined as "rms in smallest region containing 90% of events"
  - Introduced to reduce sensitivity to tails in a well defined manner
- For a true Gaussian distribution, $rms_{90} = 0.79\sigma$
- However, this can be highly misleading:
  - Distributions almost always have tails
  - Gaussian usually means fit to some region
  - $G(rms_{90})$ larger than central peak from PFA
- MC studies to determine equivalent statistical power indicate that:

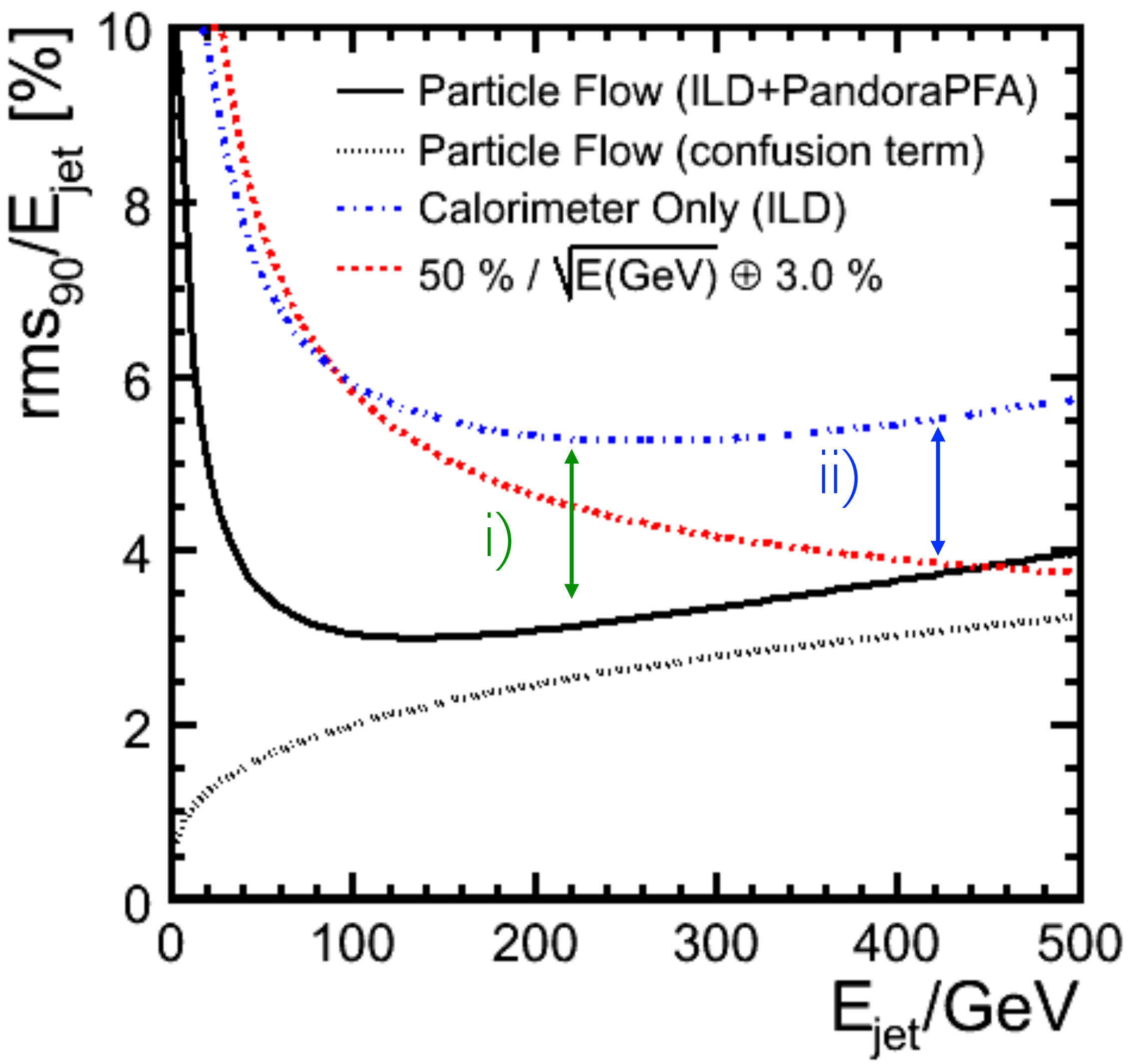
$$rms_{90} \approx 0.9\sigma_{Gaus}$$

- Now use rms_{90} as a sensible convention, but does not mean PFA produces particularly large tails.



PFA vs. Conventional Calorimetry

- ILD/SiD intended for PFA, but also good conventional calorimeters:
  - ECAL ~15%/√E
  - HCAL ~55%/√E



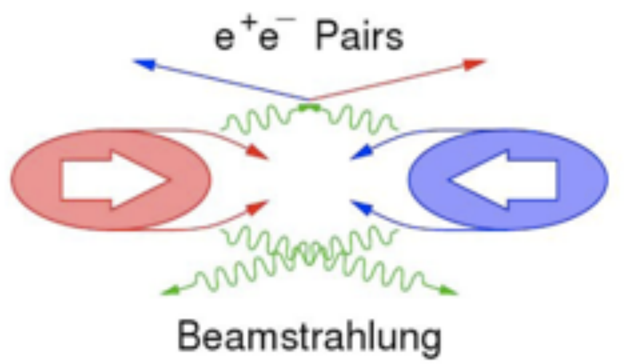
i) PFA always wins over purely calorimetric approach
ii) Effect of leakage clear at high energies

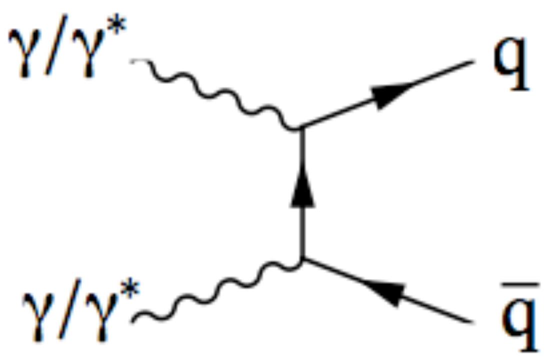
CLIC Machine Environment

	LEP 2	ILC 0.5 TeV	CLIC 0.5 TeV	CLIC 3 TeV
L [cm^{-2}s^{-1}]	5×10^{31}	2×10^{34}	2×10^{34}	6×10^{34}
BX/train	4	2670	350	312
BX sep	247 ns	369 ns	0.5 ns	0.5 ns
Rep. rate	50 kHz	5 Hz	50 Hz	50 Hz
L/BX [cm^{-2}]	2.5×10^{26}	1.5×10^{30}	1.1×10^{30}	3.8×10^{30}
$\gamma\gamma\rightarrow$X / BX	neg.	0.2	0.2	3.2
σ_x/σ_y	240 / 4 mm	600 / 6 nm	200 / 2 nm	40 / 1 nm

Drives timing Requirements for CLIC detector

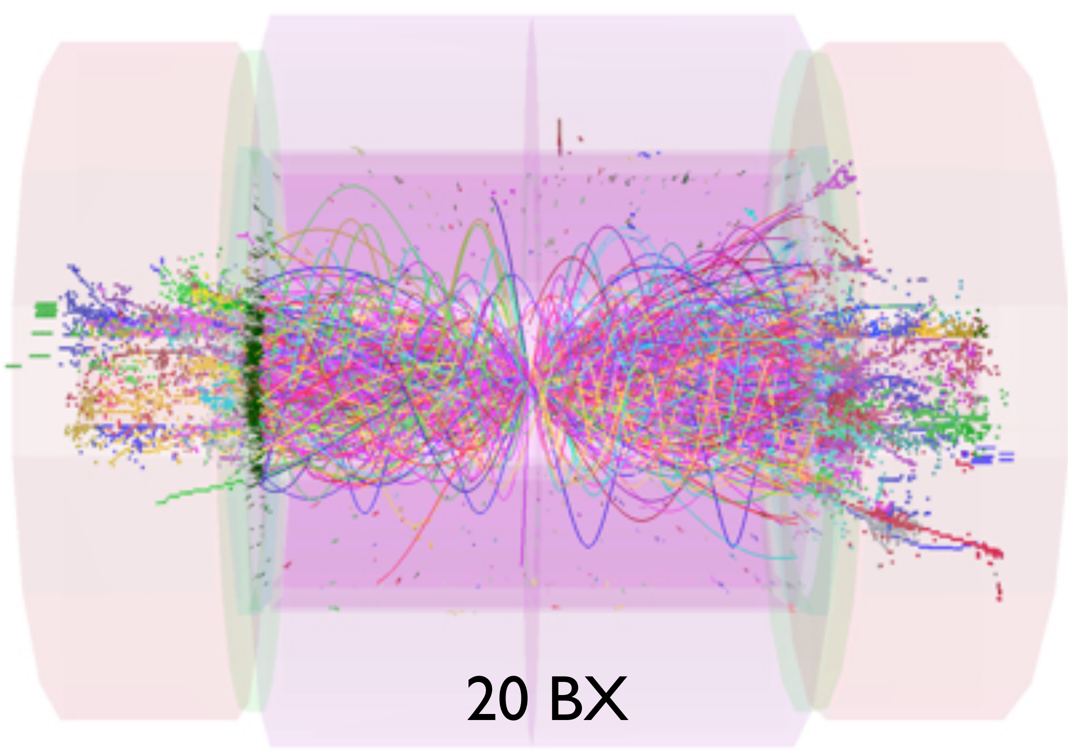
- Beam-related background:
  - Small beam-profile at IP leads to very high E-field
    - Beamstrahlung
    - Pair-background
  - Interactions of real and virtual photons
    - $\gamma\gamma\rightarrow$hadrons "mini jets"
    - Integrate over multiple BXs of $\gamma\gamma\rightarrow$hadrons
    - 19TeV visible energy per 156ns bunch train

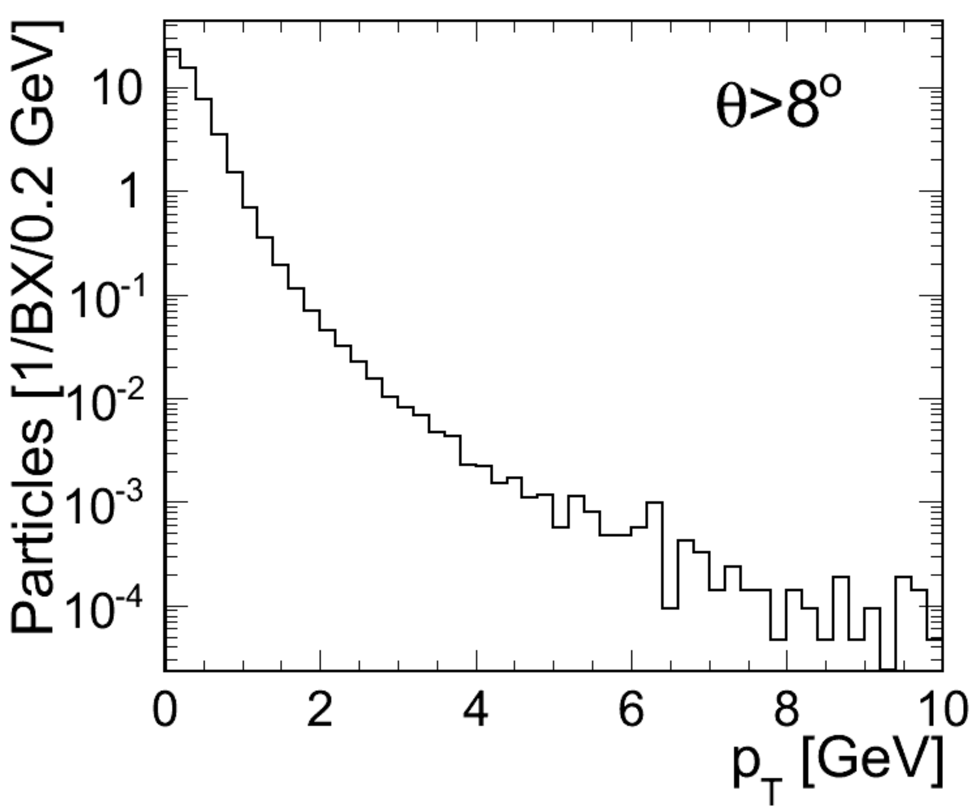




CLIC Backgrounds

- Pair background largely affects very low angle region.
- Background in calorimeters and central tracker dominated by $\gamma\gamma \rightarrow$ hadrons "mini-jets".
- At 3 TeV, average 3.2 events per BX (approximately 5 tracks per event).
- For entire bunch-train (312 BXs):
  - 5000 tracks (mean momentum 1.5 GeV) giving total track momentum : 7.3 TeV
  - Total calorimetric energy (ECAL + HCAL) : 19 TeV
- Largely low p_T particles, but an irreducible background.

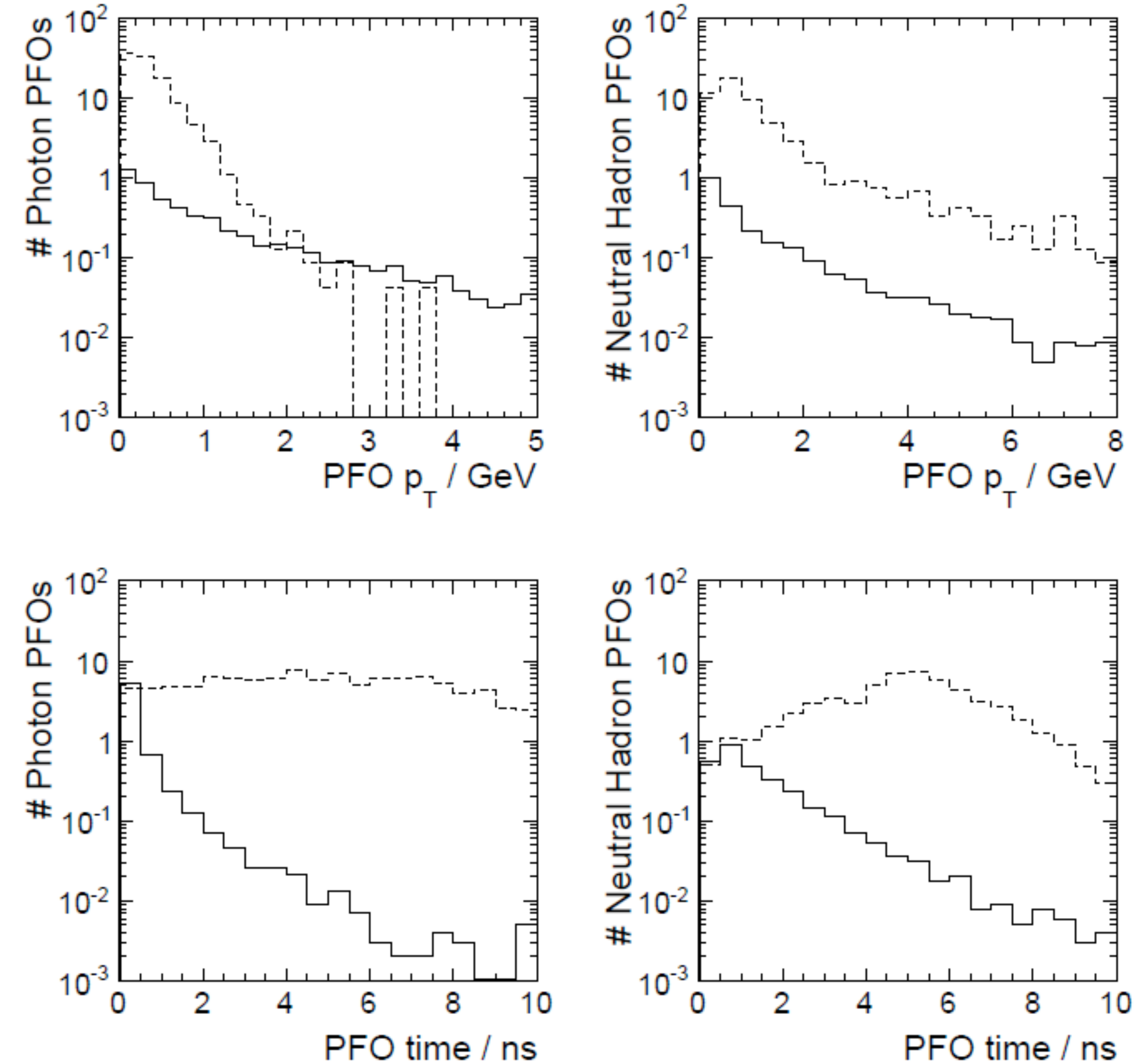




CLIC PFO Selection

- Pandora algs cluster energy in detector into individual particles, which can be identified as background or from underlying interaction.
- Cannot place timing cuts on individual hits prior to reconstruction, but can cut on timing and p_T properties of reconstructed PFOs.
- PFOs from physics event have range of p_T values and times close to t_0.

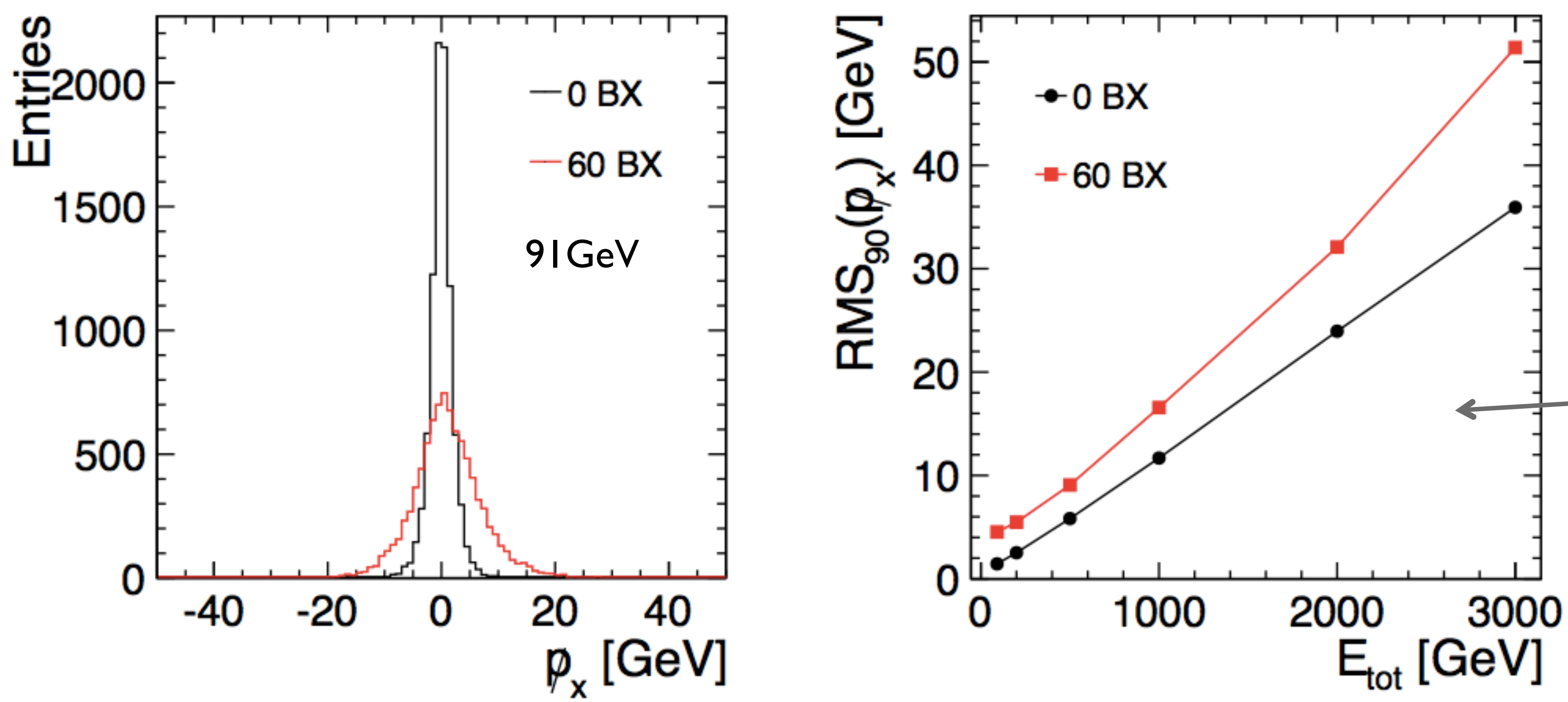
Cut	$\gamma\gamma \rightarrow$ hadrons Energy (GeV)	500 GeV di-jet Energy (GeV)	500 GeV di-jet Energy loss
No cut	1210	500.2	0%
Loose	235	498.8	0.3%
Default	175	498.0	0.5%
Tight	85	496.1	0.8%
$p_T > 3.0\,\mathrm{GeV}$	160	454.2	9.2%



Solid histograms show distributions for $ZZ \rightarrow qq\nu\nu$ events at $\sqrt{s}$=3TeV, whilst dashed histograms are for pile-up from $\gamma\gamma \rightarrow$ hadrons

Fake Missing Momentum

- Fake missing momentum can result from limitations in detector coverage and from failed reconstruction of particle momenta. Quantified using samples of Z decays to light quarks.
- Examine distribution of single component (e.g. x-component) of the fake missing momentum with and without background. Resolution then obtained by calculating RMS_{90} of distribution.



Level of fake missing momentum (in one coordinate) is approx. 1-2% of event energy