A hit density driven clustering algorithm for PFA study

Lei Xia
ANL-HEP
SiD calorimeter WG
**Introduction**

- **Goal:** to have a clustering algorithm that can
  - Form clusters that can closely represent single particle shower
    - Pick up as many hits as possible for a single particle
    - Distinguish different particles
  - Treat ECAL and HCAL as one detector
    - Treat cell/layer structure properly
    - Cluster doesn’t break up at boundaries
  - Adjustable parameter for PFA
- **Reality:** hadron showers have hits all over the detector
  - Impossible to pick up every hits of a shower without messing up different showers
  - Try to pick up only the central part of a shower, and deal with fragments later
  - Use hit density to drive the clustering
  - Using sidmay05 simulation, with RPC DHCal (projective geometry, barrel cell size 7.4 – 12.mm)
Clustering: hit density

Try to find a two-point density function which can reflect the closeness of two hits (relative to local cell size)

Considered hit density variation in different directions

It is a very local density function, only nearby hits contribute

So far, didn’t consider hit energy weighting
Clustering: grow a cluster

- Find a cluster seed: hit with highest density among remaining hits
- Attach nearby hits to a seed with a tight cut on hit-seed density
- Attach additional hits with a tight cut on hit-seed cluster density
  - EM hits, $D(\text{hit,cluster}) > 0.01$
  - HAD hits, $D(\text{hit,cluster}) > 0.001$
  - Grow the cluster until no hits can be attached to it
- Find next cluster seed, until run out of hits

Hit been considered

Picked up hits
## Hit efficiency: single particle

<table>
<thead>
<tr>
<th>Particle</th>
<th>ECAl hit efficiency</th>
<th>HCal hit efficiency</th>
<th>Overall hit efficiency</th>
<th>Overall energy efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photon (1GeV)</td>
<td>89%</td>
<td>43%</td>
<td>89%</td>
<td>91%</td>
</tr>
<tr>
<td>Photon (5GeV)</td>
<td>92%</td>
<td>54%</td>
<td>92%</td>
<td>96%</td>
</tr>
<tr>
<td>Photon (10GeV)</td>
<td>92%</td>
<td>61%</td>
<td>92%</td>
<td>97%</td>
</tr>
<tr>
<td>Photon (100GeV)</td>
<td>95%</td>
<td>82%</td>
<td>95%</td>
<td>&gt;99%</td>
</tr>
<tr>
<td>Pion (2 GeV)</td>
<td>78%</td>
<td>59%</td>
<td>75%</td>
<td>71%</td>
</tr>
<tr>
<td>Pion (5 GeV)</td>
<td>81%</td>
<td>70%</td>
<td>79%</td>
<td>80%</td>
</tr>
<tr>
<td>Pion (10GeV)</td>
<td>84%</td>
<td>80%</td>
<td>83%</td>
<td>85%</td>
</tr>
<tr>
<td>Pion (20GeV)</td>
<td>85%</td>
<td>87%</td>
<td>88%</td>
<td>91%</td>
</tr>
</tbody>
</table>

- Typical electron cluster energy resolution ~ 21%/sqrt(E)
- Typical pion cluster energy resolution ~ 70%/sqrt(E)
- All numbers are for one main cluster (no other fragments are included)
Hit efficiency: single particle

Very preliminary: from sidaug05_scinthcal (scint. DHCal, 7.4 – 12.mm barrel cell size)

<table>
<thead>
<tr>
<th>Particle</th>
<th>ECal hit efficiency</th>
<th>HCal hit efficiency</th>
<th>Overall hit efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pion (2 GeV)</td>
<td>78%</td>
<td>30%</td>
<td>62%</td>
</tr>
<tr>
<td>Pion (5 GeV)</td>
<td>82%</td>
<td>47%</td>
<td>66%</td>
</tr>
<tr>
<td>Pion (10 GeV)</td>
<td>86%</td>
<td>58%</td>
<td>72%</td>
</tr>
<tr>
<td>Pion (20 GeV)</td>
<td>88%</td>
<td>71%</td>
<td>78%</td>
</tr>
</tbody>
</table>

• All hadron calorimeter hits need to pass through a 0.3MeV cut
Shower separation...example

- 10 GeV photon converts to e+e- pair
- The two showers can be clearly separated
Shower separation...example

- 10 GeV $e^-$ kick out a 5 GeV photon in tracker
- With appropriate density cut, the algorithm can separate the two showers
- Old picture from SDjan03 simulation
Most of the clusters (89.7%) are pure (only one particle contributes).
For the rest 10.3% clusters
  - 55% are almost pure (more than 90% hits are from one particle)
  - The rest clusters contain merged showers, part of them are ‘trouble makers’
On average, 1.2 merged shower clusters/Z pole event
• 81.0% of the clusters are pure (only one particle contributes)
• For the rest 19.0% clusters
  – 68% are almost pure (more than 90% hits are from one particle)
• On average, 1.6 merged shower clusters/Z pole event
• Projective geometry, 7.4 – 12.mm barrel cell size
**Thing to work on...**

- Better understanding of clustering
  - Hit efficiency -> new calibration(?)
  - Optimize density cut according to PFA performance
  - What to do with merged showers?
- Cluster ID
  - e/gamma, hadron, fragment
  - Clusters that contain multiple showers