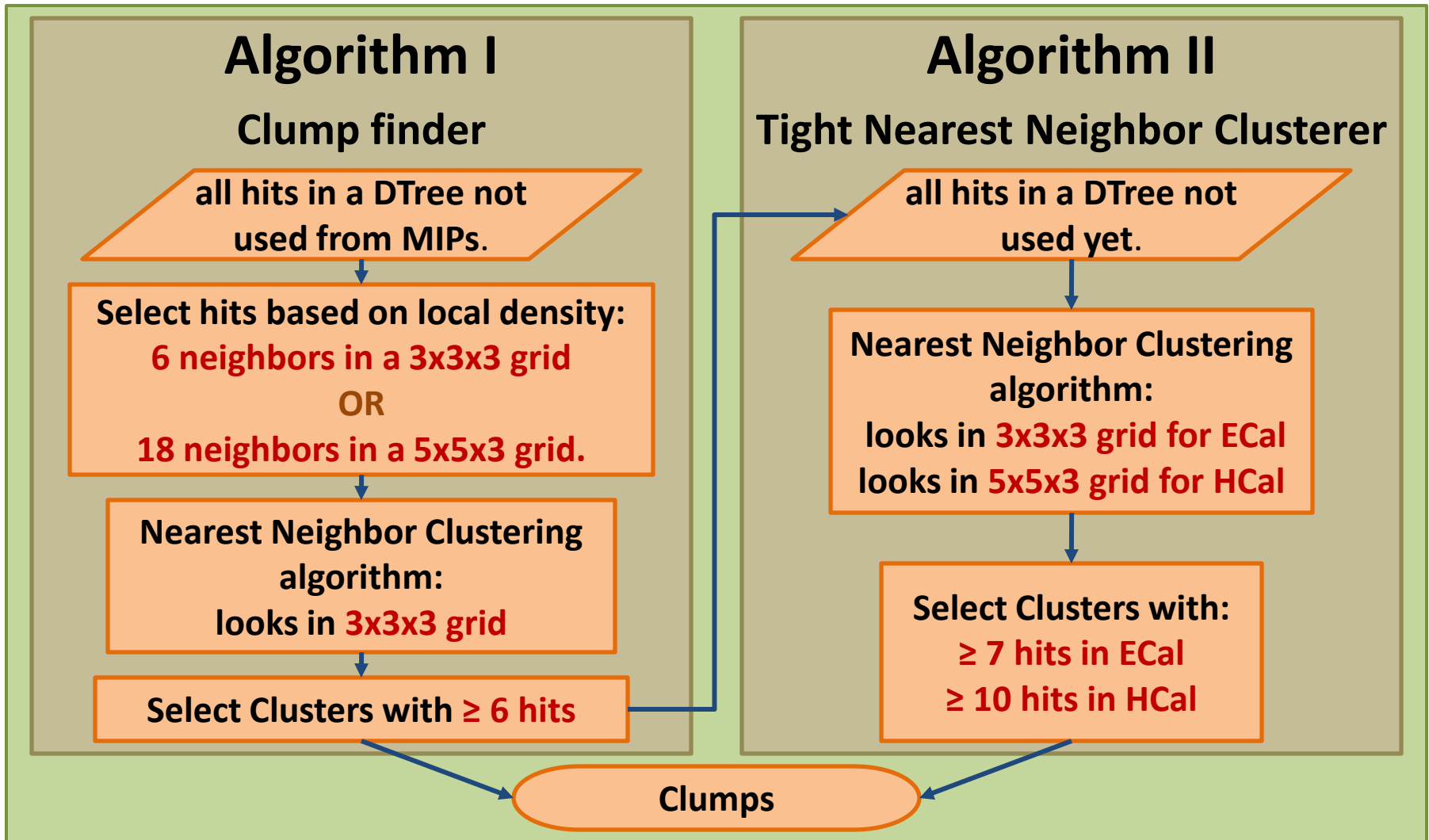


# DTree Sub-Clustering: Ideas for Clump reconstruction

R. Cassell, M. Charles, G. Halladjian,  
U. Mallik, R. Zaidan

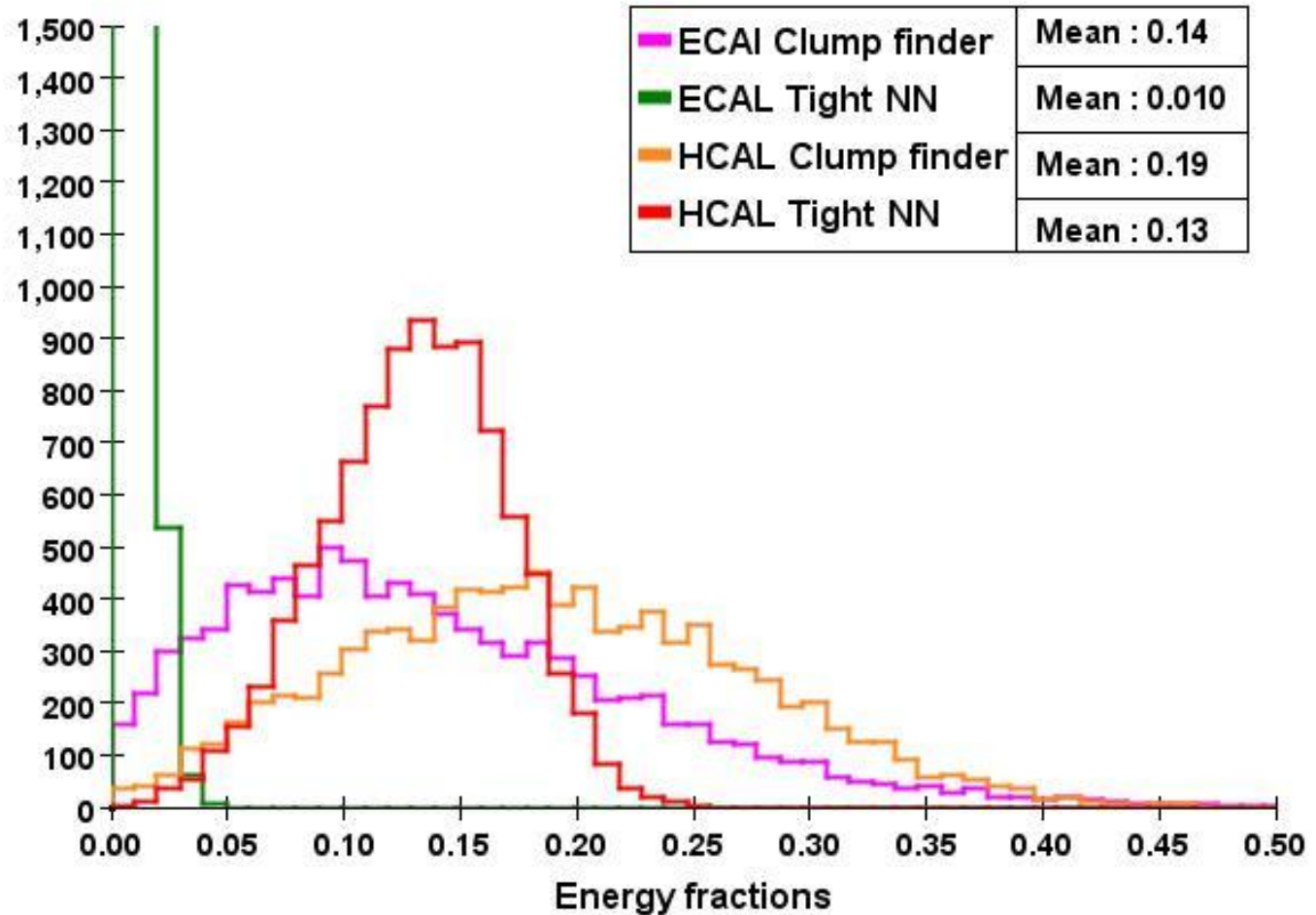


# Clump algorithms in use now



# Clumps Energy Fractions

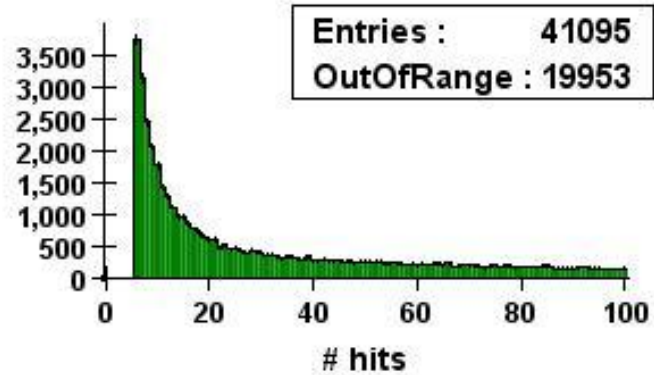
Clumps	Fractions	
Ecal: Alg I	32%	93%
Ecal: Alg II		7%
HCal: Alg I	68%	59%
HCal: Alg II		41%



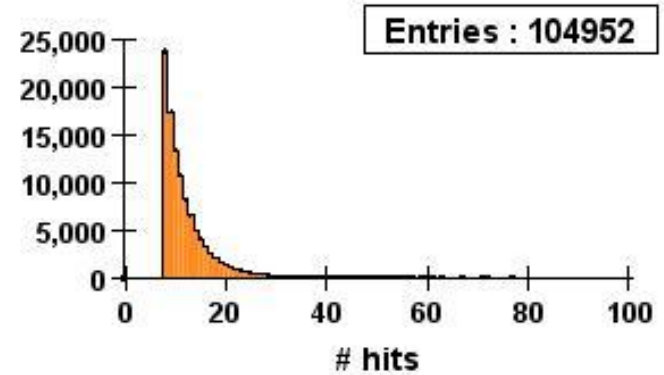
# Clumps hit multiplicities

The first Clump finder tend to reconstruct less but larger clumps.

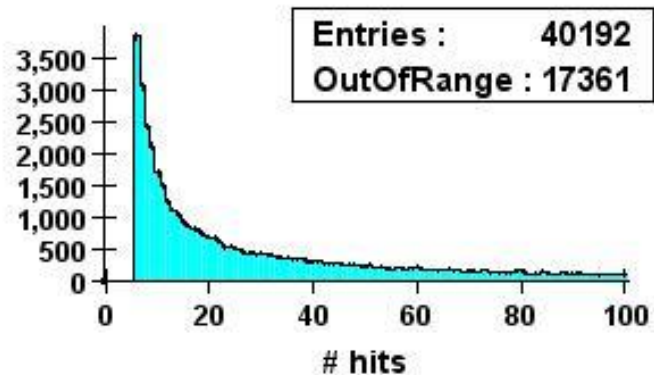
Clump finder in ECal



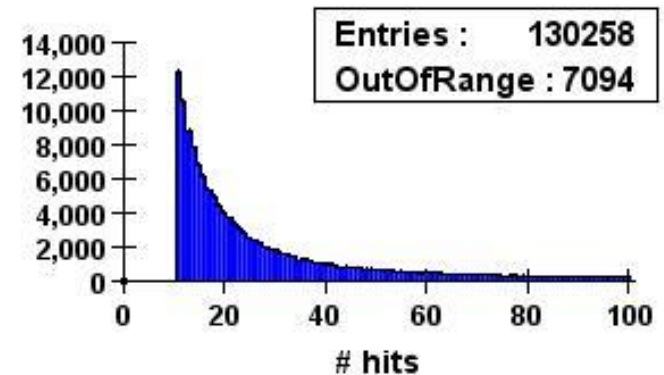
Tight NN in ECal



Clump finder in HCal



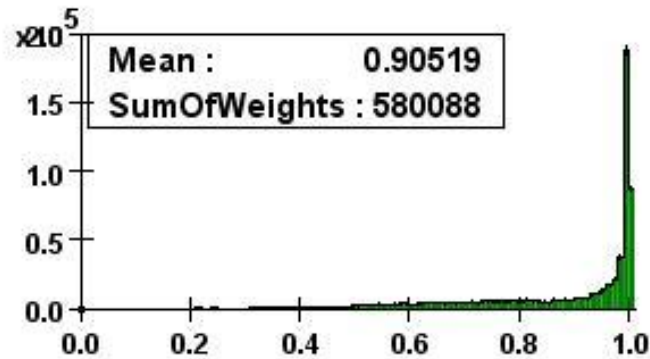
Tight NN in HCal



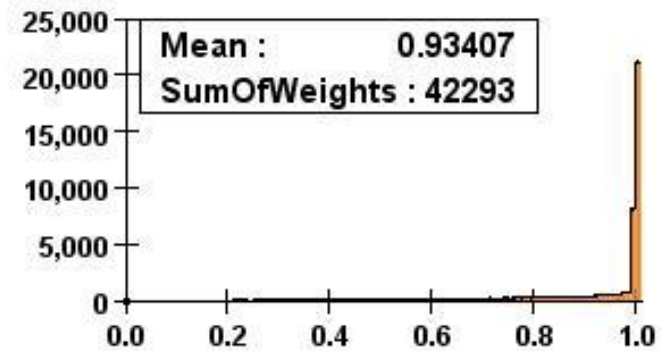
# Clumps purities

Clumps	Purity
ECal: Alg I	<b>90.5%</b>
ECal: Alg II	<b>93.4%</b>
HCal: Alg I	<b>80.7%</b>
HCal: Alg II	<b>78.6%</b>

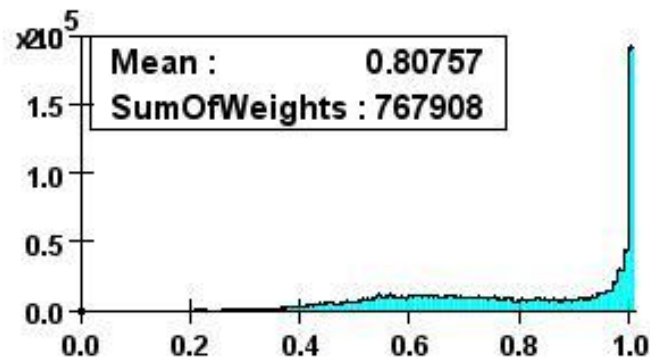
Clump finder in ECal



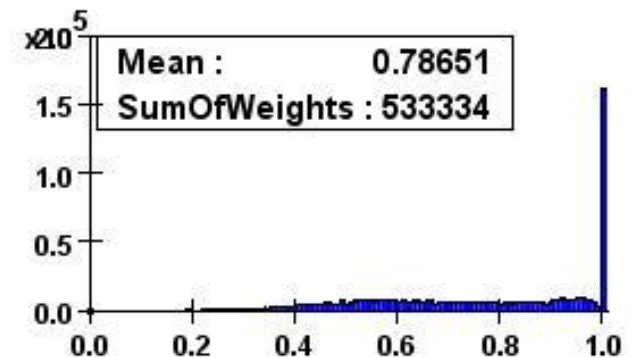
Tight NN in ECal



Clump finder in HCal



Tight NN in HCal

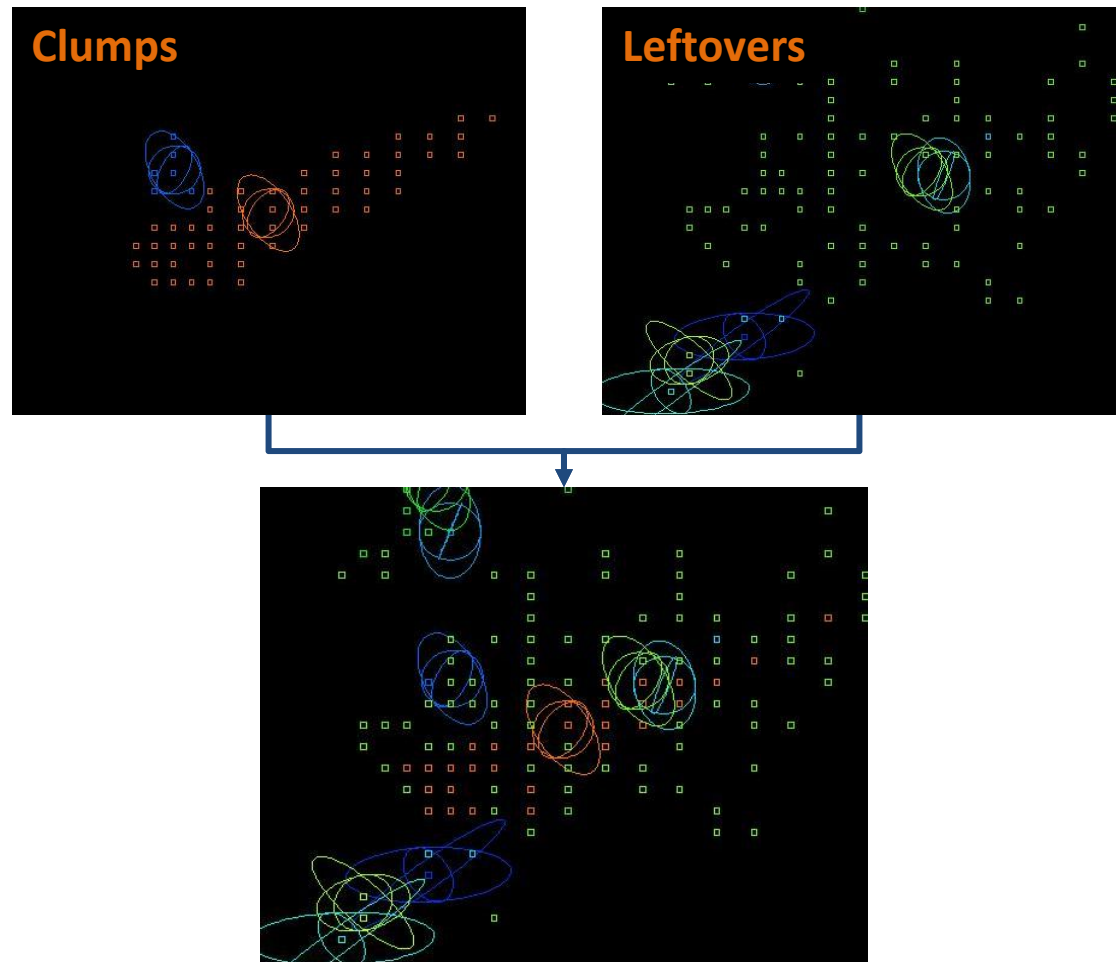


# Leftover hits

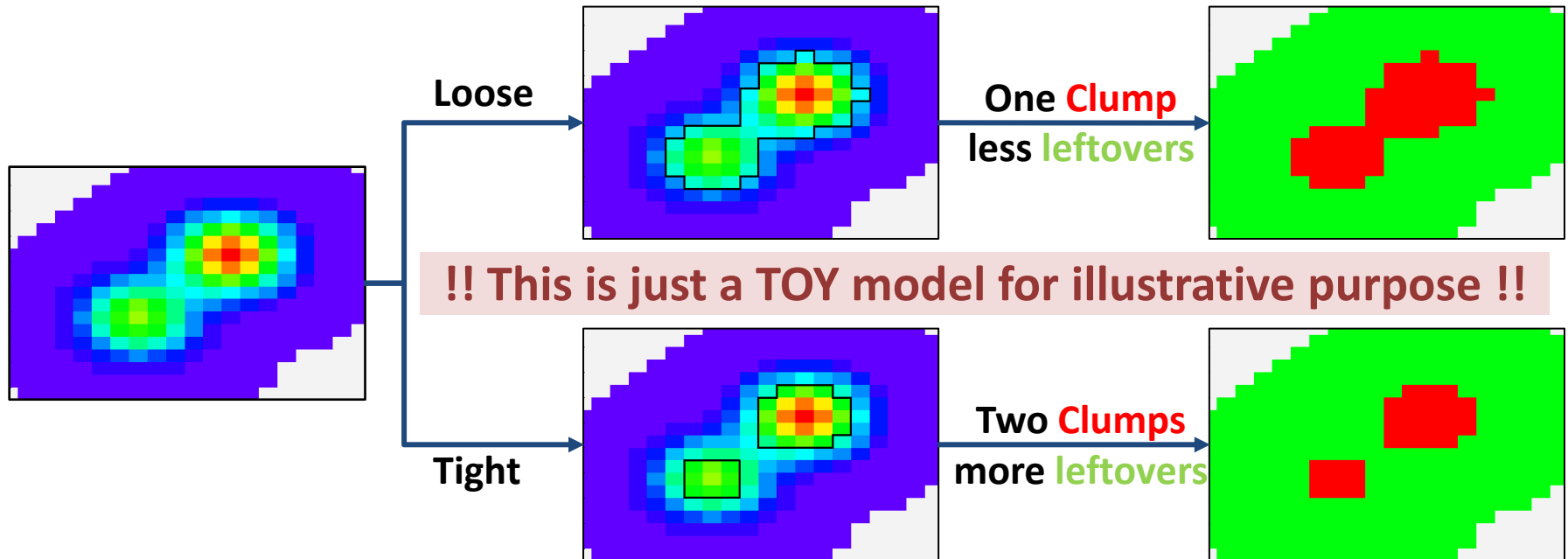
**Leftovers take often the shape of an empty shell around Clumps...**

Rejecting low density hits allows better definition of position and shape for better linking.

The sharing algorithm should take care of correctly assigning the leftover energy:  
**Does it still work in an “overlap” scenario??**



# Using density/energy gradient



Optimization possibilities with the current algorithm are limited:

- One tunable parameter: we lose information in both loose and tight scenarios!
- The actual clustering algorithm is “color-blind”: the density information is only used for hit pre-selection.

**A simple Nearest Neighbour algorithm may not be suitable for overlap scenarios.**



# Alternative Clustering algorithm

## *k*-means clustering algorithm

A - Define a “distance”  
and a “mean”

Distance: a **metric** that  
tells how likely a hit  
belongs to a cluster.

Mean: a **procedure** to  
“group” hits together  
→ **Build Clusters.**

B - Define *k* seeds

Seeds: an **initial set of clusters.** The **number  
of seeds** determine the **number of Clumps.**

C - Assign each hit to “closest” seed

D - Update seeds by  
“averaging” over assigned hits

Yes  
Assignment  
changed?

No

Done

Many algorithms exist in the  
market.

I like this one because:

- it is simple.
- it leaves the flexibility for  
“physics” input:

**Physics would go into (A) and (B).**

# Next steps

- Simple starting point:
  - **Leftovers:** Lower the density threshold below which a hit is called a “leftover”:
    - To be optimized between the performance of the linking algorithm and that of the sharing algorithm.
  - **Seeds:** Define the seeds by finding local high-density maxima.
  - **Metric:** Define “distance” as the geometrical distance.
- Ideas if needed (need studies and optimization):
  - Use energy/density gradient and local direction in the definition of the “distance”.
  - Use measured energy instead of hit density in the ECal.
  - Include a “Split and Merge” step in the algorithm, either at the end or at each iteration, to avoid fixing the number of Clumps at the initial step.