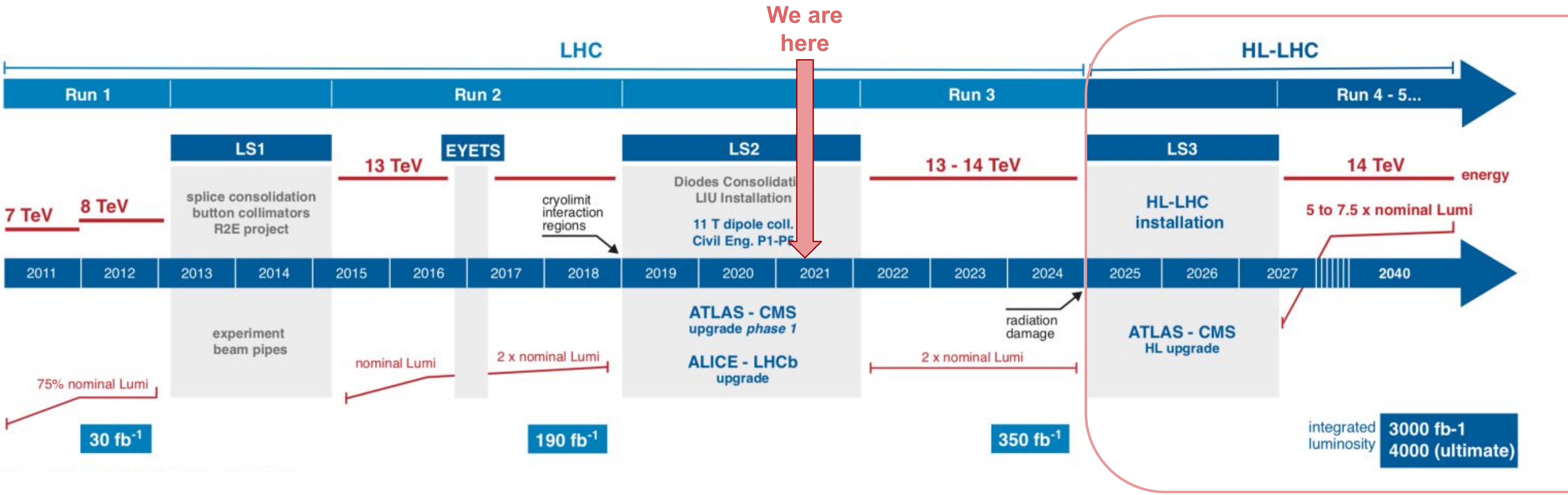
Novel reconstruction algorithms for an imaging calorimeter for HL-LHC

Erica Brondolin
on behalf of the CMS Collaboration

2021 International Workshop on Future Linear Colliders, 16th March 2021

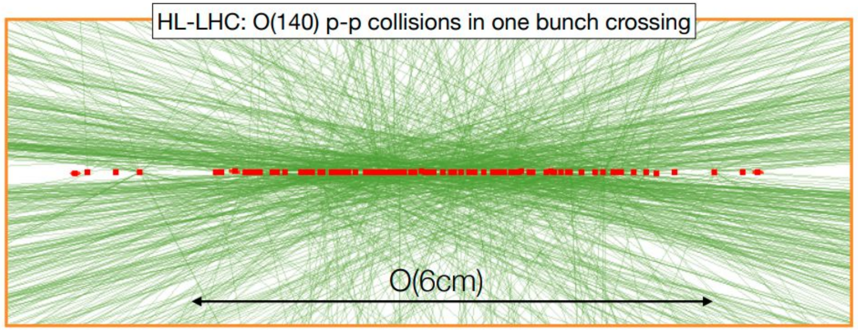
LHC roadmap

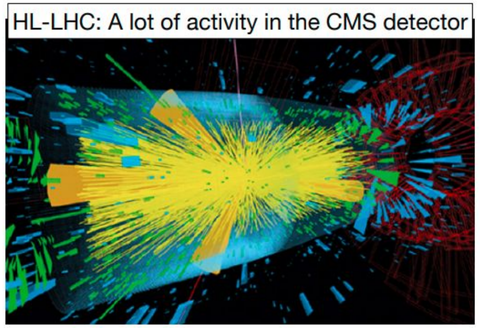


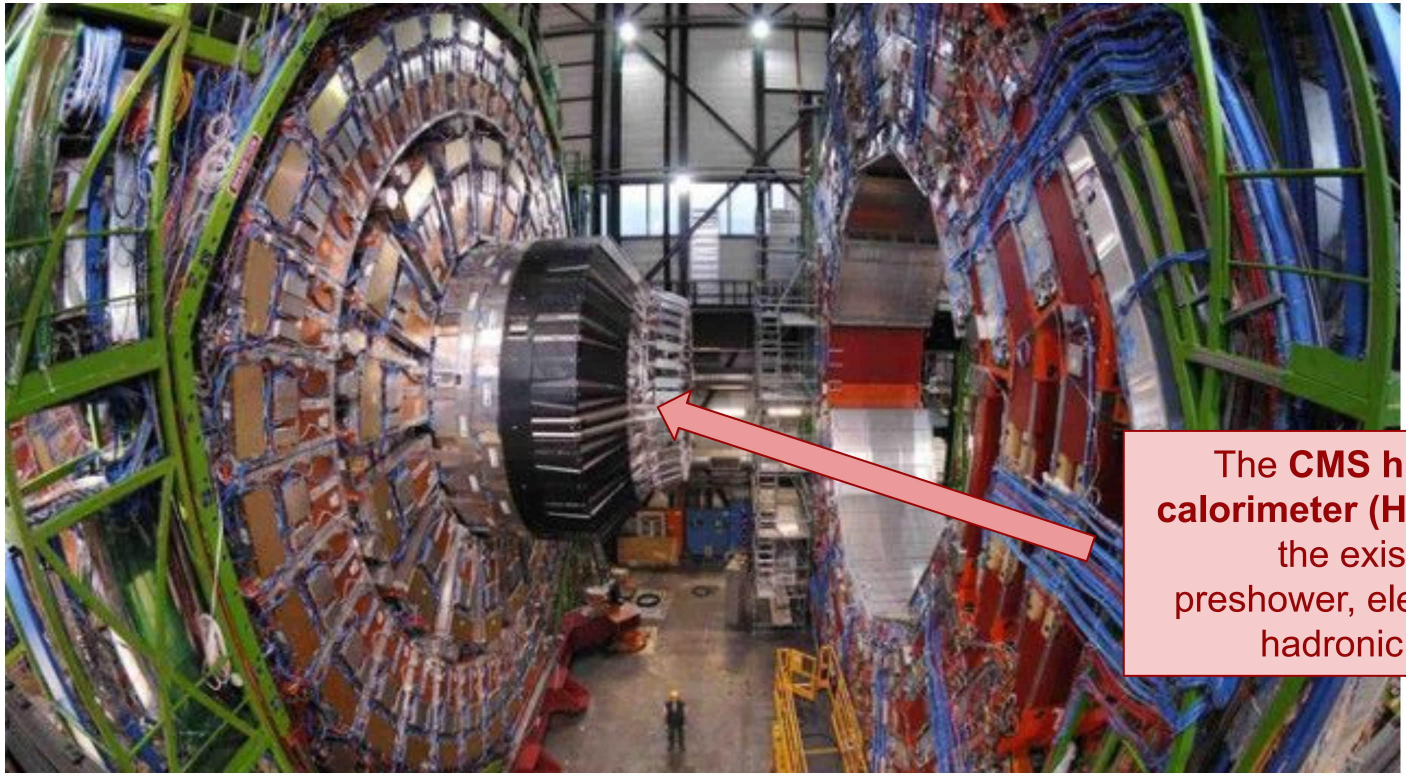
- HL-LHC
  - A significant increase in the instantaneous luminosity (5 - 7.5x)
    - Great opportunity for sensitive SM tests and BSM **searches**
    - **Pileup**: up to ~200 collisions / bunch crossing (vs. Run-2 LHC about 40)
    - **Radiation level**: 1 year at HL-LHC ~ 10 years at LHC

→ Installation of the Phase-2 upgrades of the CMS and ATLAS experiments

CMS HL-LHC upgrades







The **CMS high granularity calorimeter (HGCAL)** will replace the existing endcap preshower, electromagnetic and hadronic calorimeters

- Existing CMS designed for ~500 fb^{-1}
  - Radiation damage for many subsystems

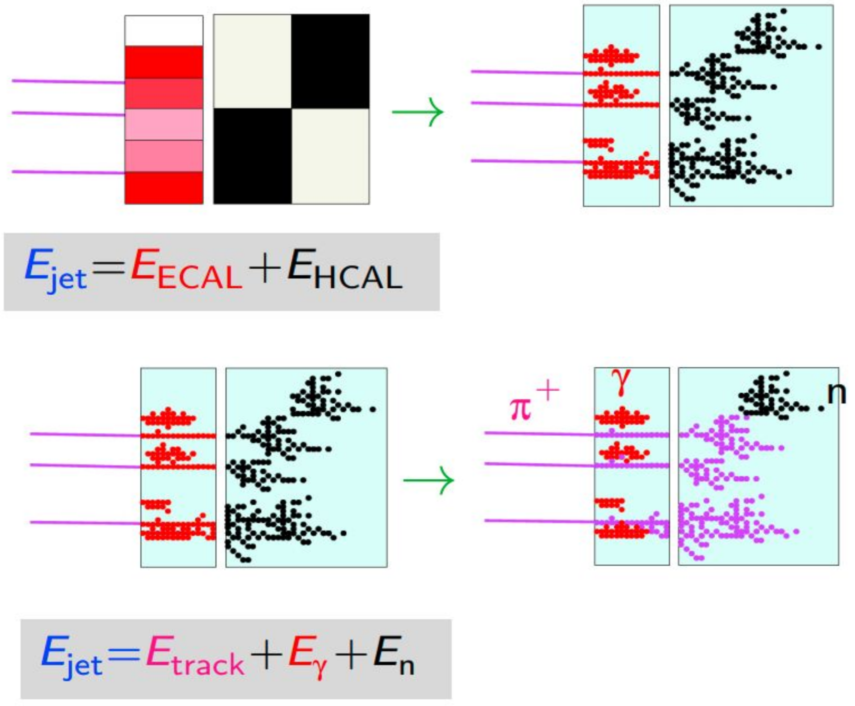
Phase-2 upgrade is needed:

- Cope with the harsh conditions @ HL-LHC (increased PU, radiation levels, …)
- Fully explore the HL-LHC potential
  - Very rich physics program ahead
  - Forward region particularly important

The case for high granularity

- CMS currently follows the **Particle Flow** paradigm
  Average jet composition:
  60% charged particles
  30% photons
  10% neutral hadrons

  Detector w/ the best information:
  → Tracker
  → ECAL
  → HCAL
- Main **goal**:
  - Accurate reconstruction of each particle (within a jet as well isolated)
- Main ingredients:
  - *Hardware*
    Assignment of energy deposits to tracks:
    Granularity is important!
  - *Software*
    Sophisticated reco software to identify energy deposits from each individual particle

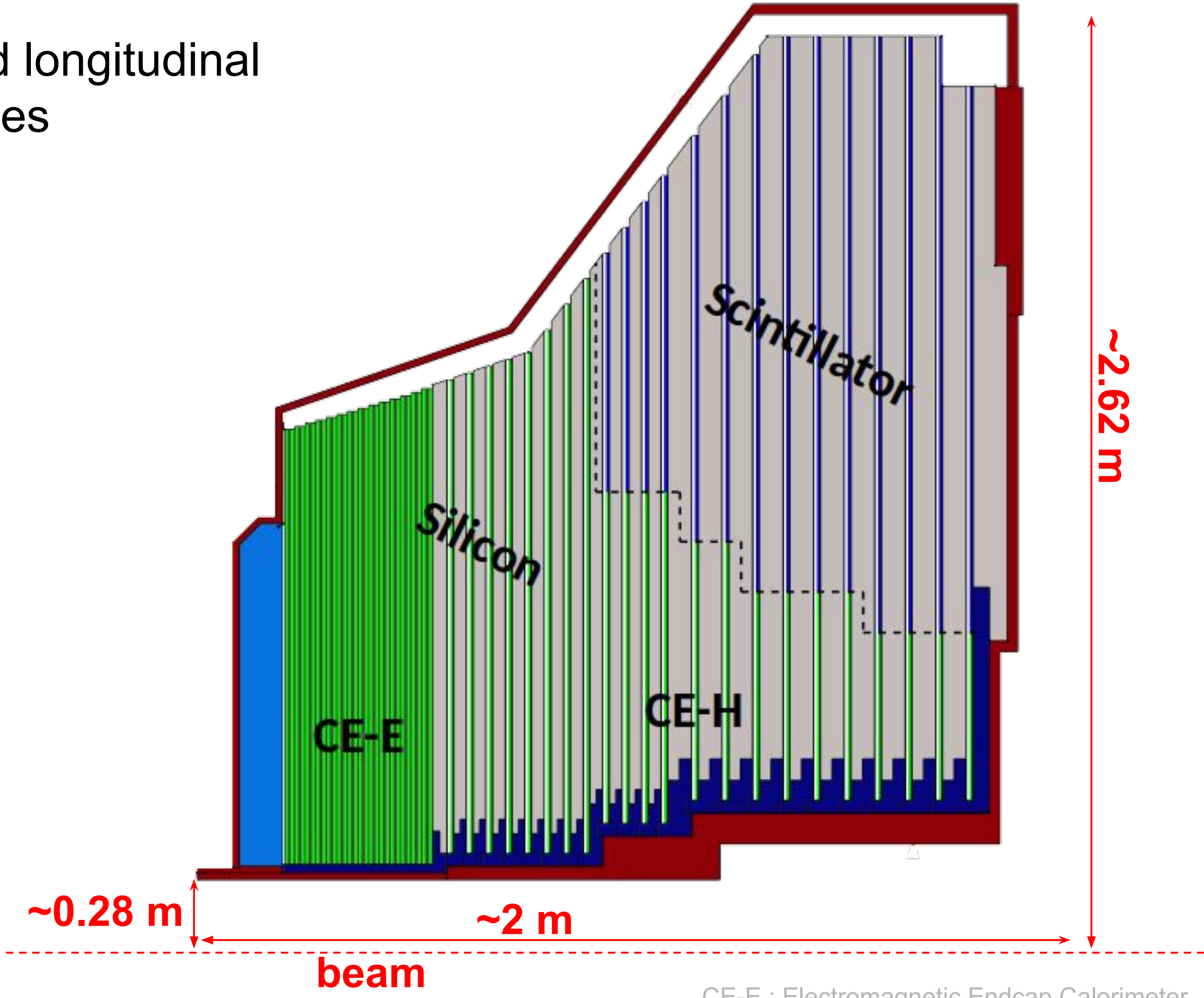


CMS High Granularity Calorimeter

- Imaging calorimeter with very fine lateral and longitudinal segmentation, and precision timing capabilities
- Covering $1.5 < \eta < 3.0$

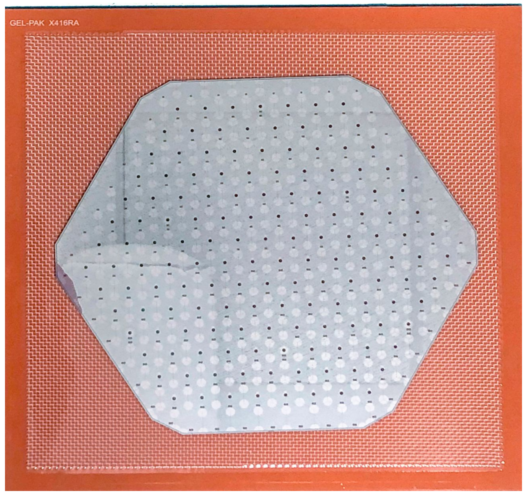
Both endcaps	Silicon	Scintillators
Area	**~620 m^2**	**~400 m^2**
Channel size	0.5 - 1 cm^2	4 - 30 cm^2
#Modules	~30'000	~4'000
#Channels	**~6 M**	240 k
Op. temp.	-30 °C	-30 °C

Ref.

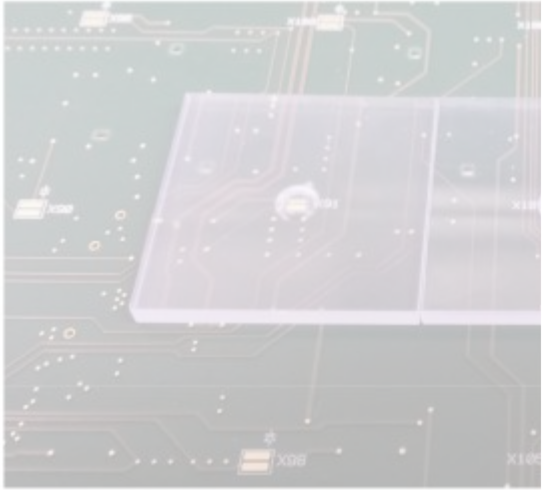


CE-E : Electromagnetic Endcap Calorimeter
CE-H : Hadronic Endcap Calorimeter

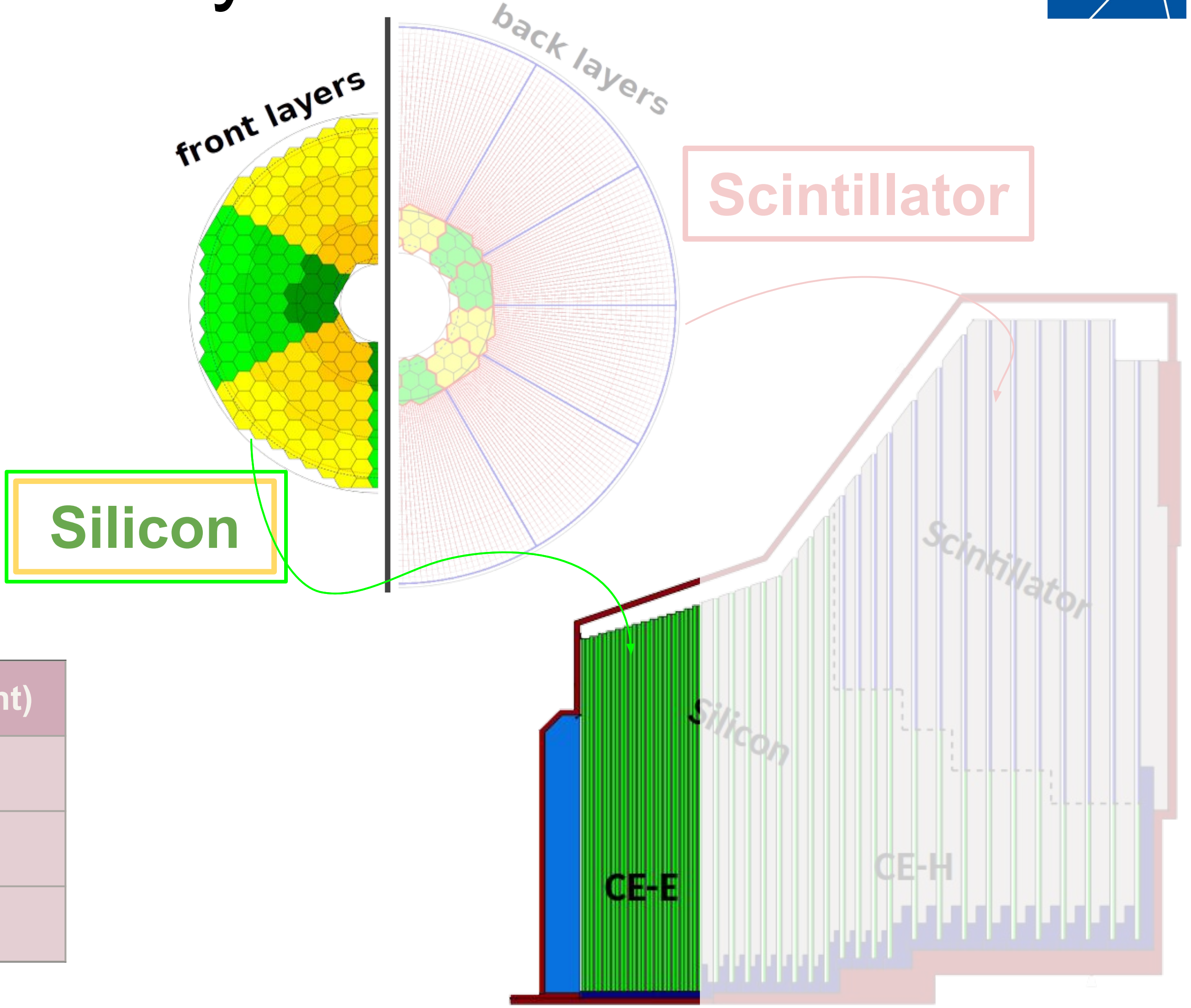
CMS High Granularity Calorimeter



8'' silicon sensor



Scintillator tile + SiPM

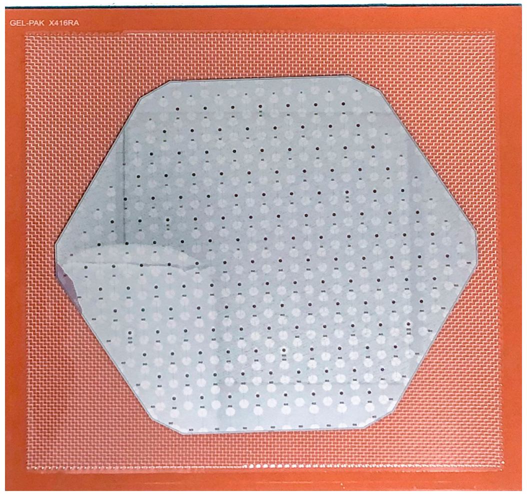


Per endcap	CE-E	CE-H (Si)	CE-H (Si+Scint)
Absorber	Pb & CuW & Cu	Stainless steel	
Depth size	25 X_0 & 1.3 λ	8.5 λ	
Layers	28	8	14

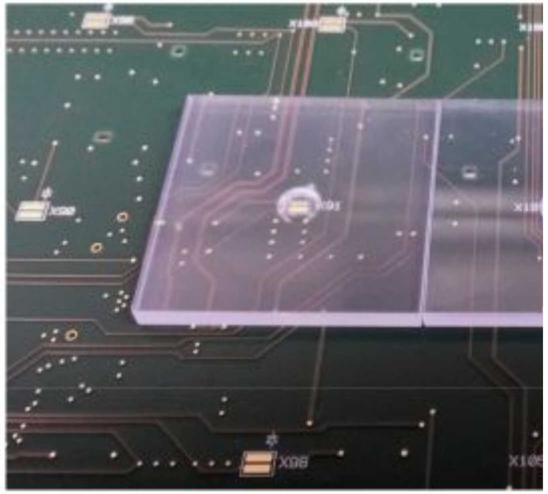
Ref.

CE-E : Electromagnetic Endcap Calorimeter
CE-H : Hadronic Endcap Calorimeter

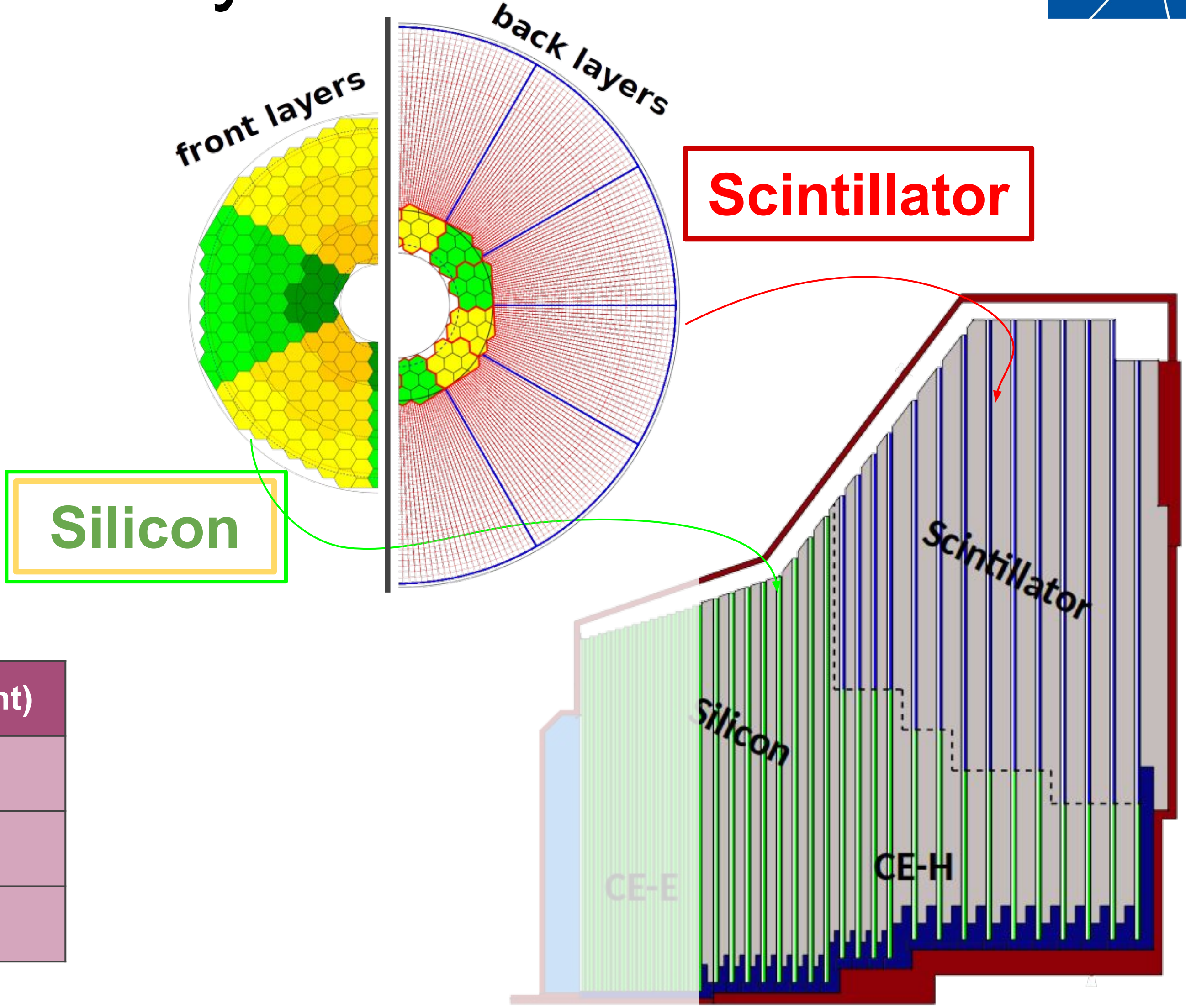
CMS High Granularity Calorimeter



8'' silicon sensor



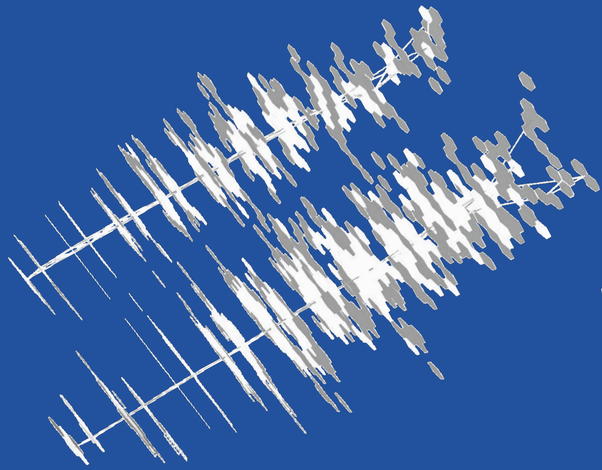
Scintillator tile + SiPM



Per endcap	CE-E	CE-H (Si)	CE-H (Si+Scint)
Absorber	Pb & CuW & Cu	Stainless steel	
Depth size	25 X_0 & 1.3 λ	8.5 λ	
Layers	28	8	14

[Ref.]

CE-E : Electromagnetic Endcap Calorimeter
CE-H : Hadronic Endcap Calorimeter



Shower reconstruction using TICL

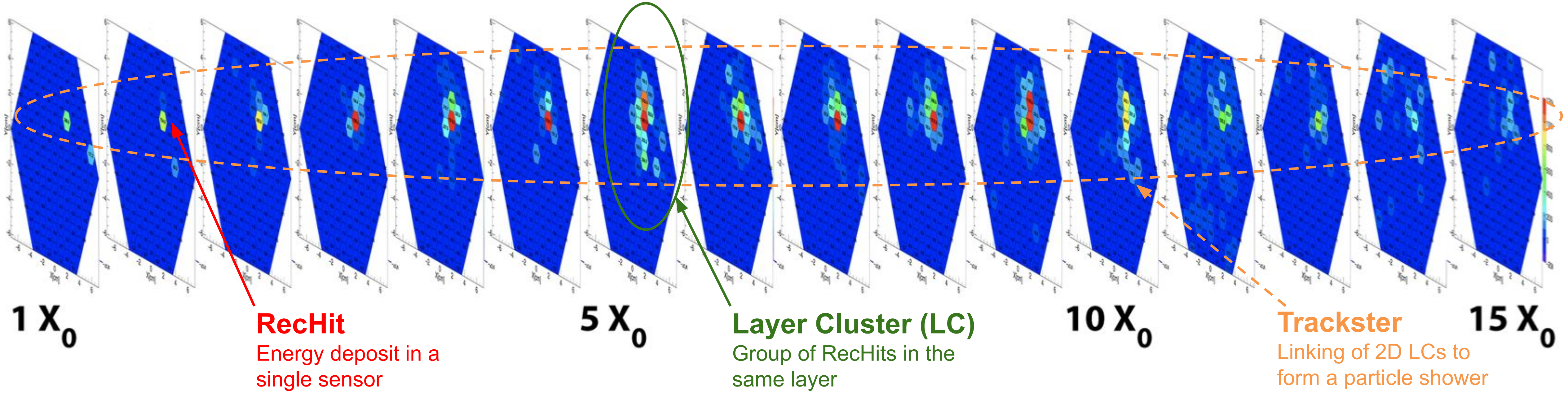
Reconstruction in HGCAL

- Particle shower reconstruction in **high-granularity calorimeter** is very interesting and crucial task in high-density environments
  - Typical situation at HL-LHC → Many showers tend to **overlap**
  - Standard reconstruction algorithms using combinatorics are expected to fail due to **memory/timing explosion**
  - **Fertile ground** for new techniques and algorithms: clustering, machine learning, graph theory, and modern computer architectures
    - Planned and designed, taking into account the information from the tracking system and timing detectors
- Development can profit from experience with CMS Particle Flow techniques
- New flexible framework can be re-used in other (future) experiments using high-granularity calorimeters

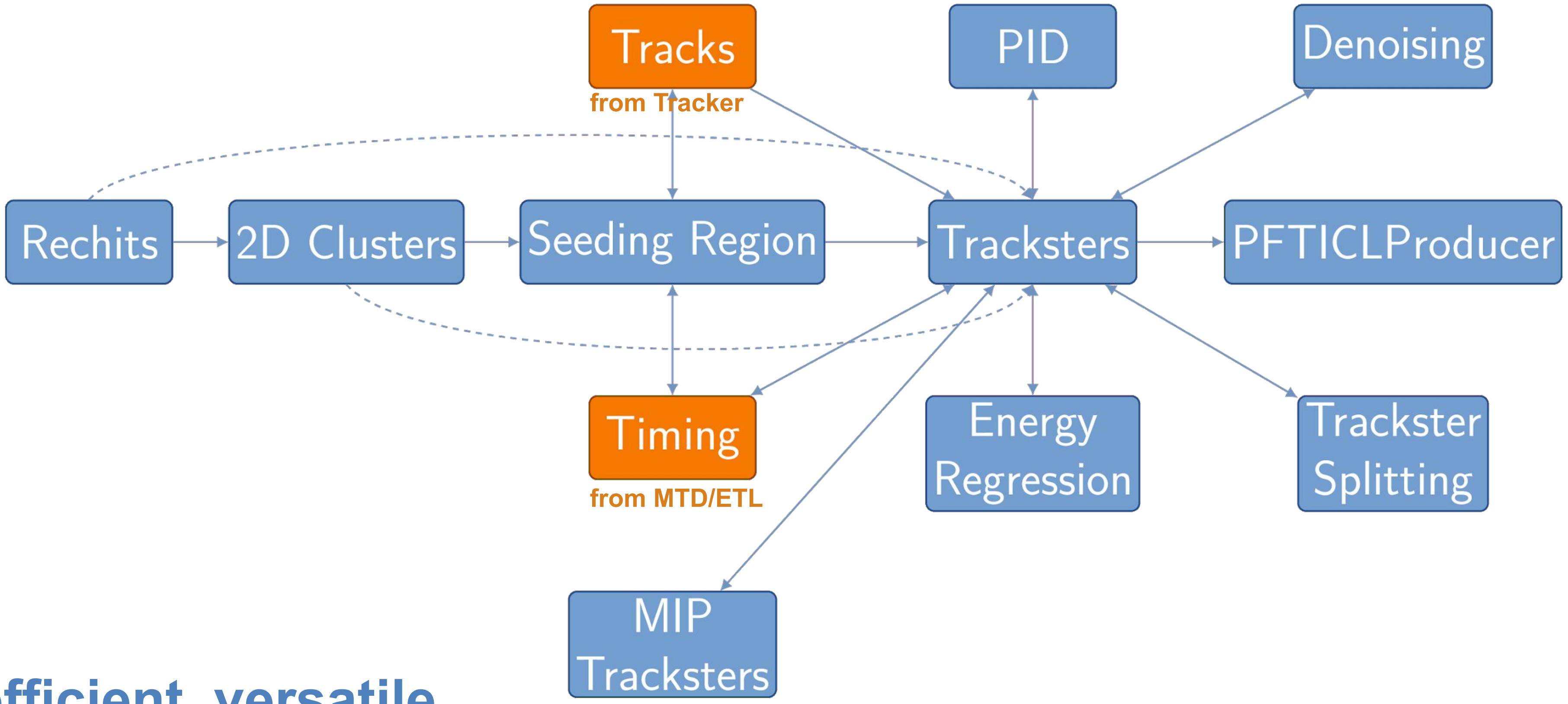
What is TICL?

- TICL (**The Iterative Clustering**) is a modular framework integrated and under development in CMS software (CMSSW)
- Main **purpose**: processing calo 5D rechits (x, y, z, t, E) and returning particle properties and probabilities
- In a nutshell: grouping 2D **Layer Clusters** into 3D clusters (**Tracksters**) **iteratively** to reconstruct different particle species

[Electron with E=32 GeV; Test beam 2016 @FNAL]



TICL modules/components

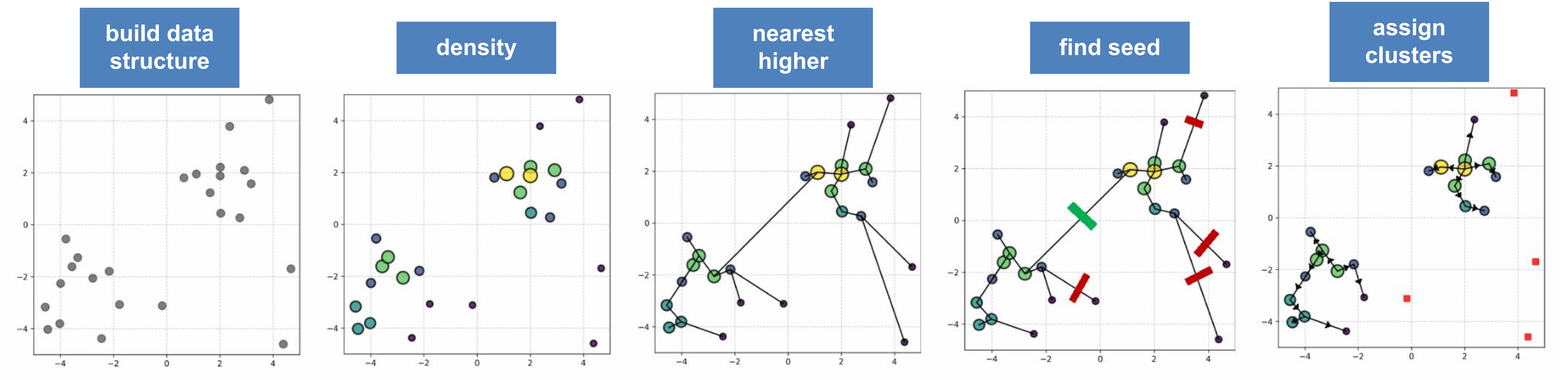


Flexible, efficient, versatile

- Algorithms are designed as swappable plugins, with heterogeneous architectures / portability in mind
- Skip and/or change modules easily
- Explore new algorithms or techniques (e.g. Machine Learning) with plug-in on top
- Mostly geometry independent

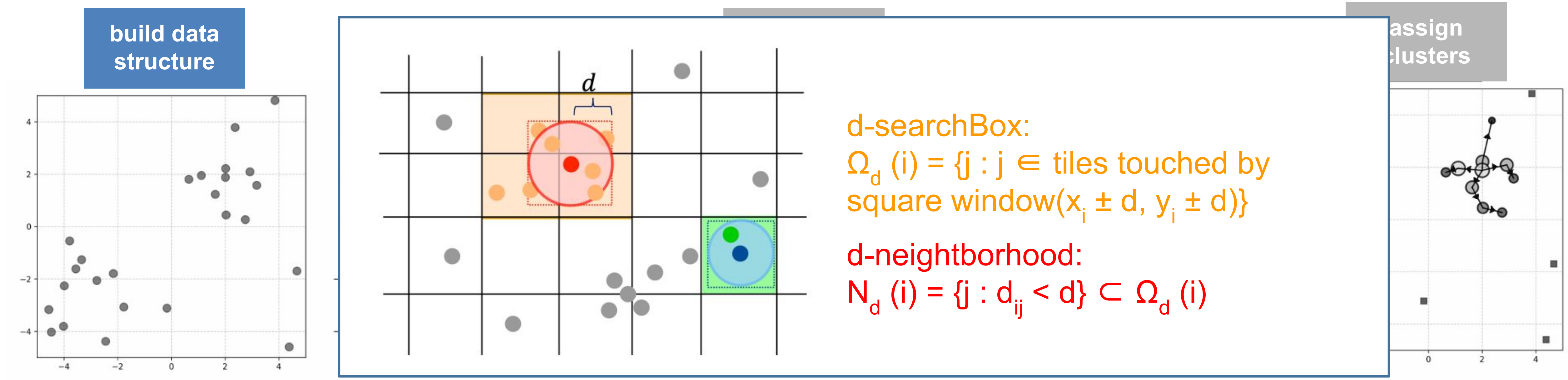
2D Clusters with CLUE

- CLUE (**CLUstering by Energy**) is an algorithm inspired by “Clustering by fast search and find of density peaks” (Ref.)
- Main characteristic:
  - *Energy density* - rather than individual cell energy - used to define ranking, seeding threshold, etc…
- GPU-friendly, i.e. suitable for the upcoming era of heterogeneous computing in HEP



Cluster Algorithm: CLUE

- Querying neighborhood is a frequent operation in density-based clustering → **fast!**
- Build **Fixed-Grid Spatial Index** for hits on each layer (η, φ space)
  - Grid tiles are small compared to the size of HGCAL layer
  - Each tile in the grid hosts indices of hits inside it and has a fixed length of memory to store the hosted indices
- To find the neighborhood hits $N_d(i)$ of *i*-hit, we only need to loop over hits in $\Omega_d\,(i)$



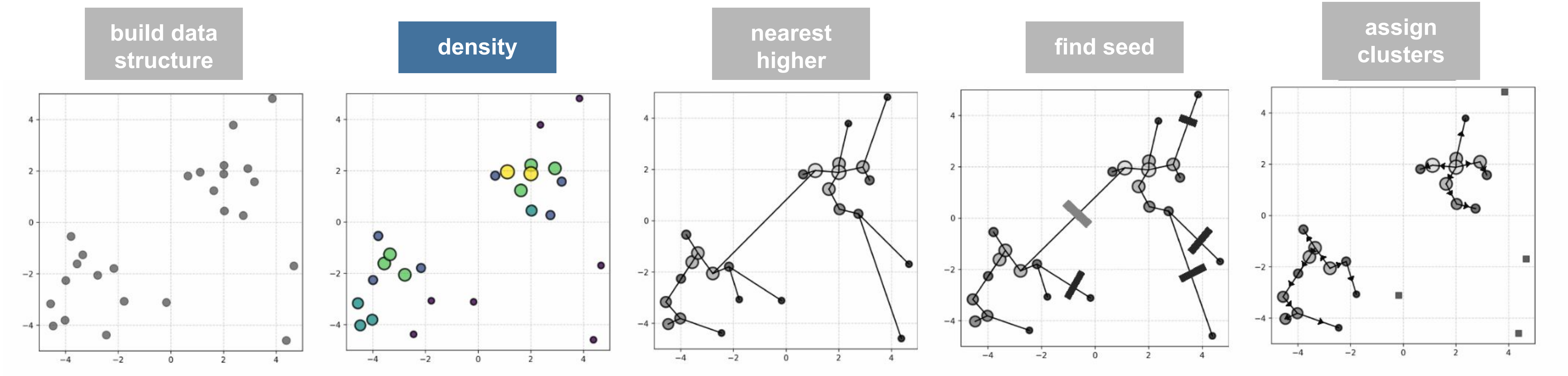
Cluster Algorithm: CLUE

ρ_c, d_c, δ_c , δ_0 are <u>tunable parameters</u> chosen with purity vs fake studies

- Calculate local energy density (ρ) in a distance (d_c)
  - Each hit *j* weighted by the deposited energy (E_j)
  - For each hit, calculate ρ_i
  - Individual d_c values considered in HGCAL Silicon and Scintillator sections

convolution kernel

$$\rho_i = \sum_{j \in N_d(i)} E_j \times f(d_{ij}) \, ; \quad f(d_{ij}) = \begin{cases} 1, \text{if } i = j \\ k, \text{if } 0 < d_{ij} \le d_c \\ 0, \text{if } d_{ij} > d_c \end{cases}$$



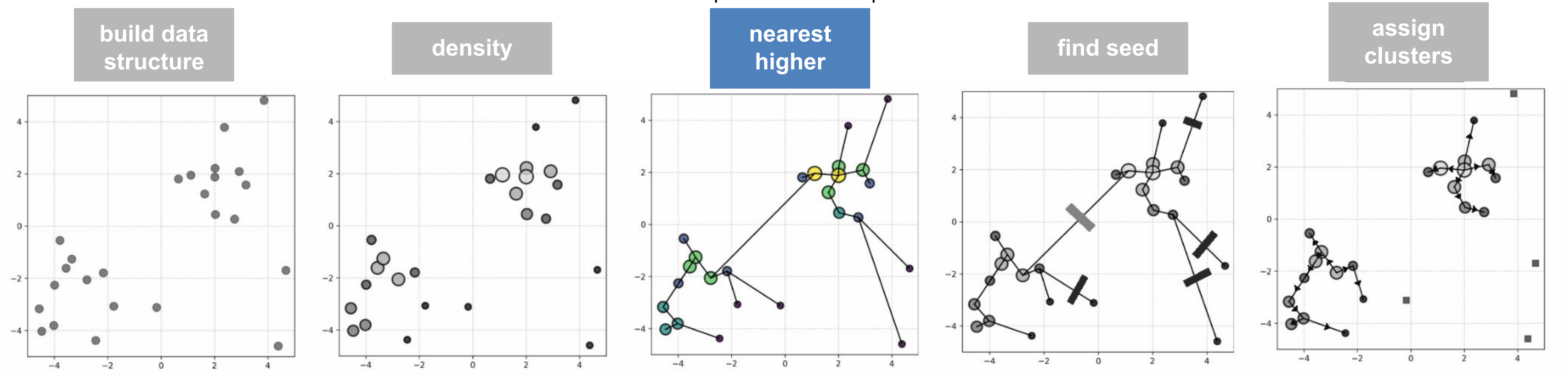
Cluster Algorithm: CLUE

ρ_c, d_c, δ_c , δ_0 are tunable parameters chosen with purity vs fake studies

- Calculate “Nearest-Higher” hit within $N_{dm}(i)$
  - Define $d_m = \max(\delta_o, \delta_c)$, δ_o and δ_c parameters for outlier demotion and seed promotion
  - Find the closest hit with higher local energy density, nh_i

$$nh_i = \begin{cases} argmin_{j \in \hat{N}_{d_m}(i)} d_{ij}, \text{ if } |\hat{N}_{d_m}| \neq 0, \hat{N}_{d_m}(i) = \{j : j \in N_{d_m}(i), \rho_j > \rho_i\} \\ -1, \text{otherwise} \end{cases}$$

  - Calculate the separation distance $\delta_i = \mathrm{dist}(i, nh_i)$



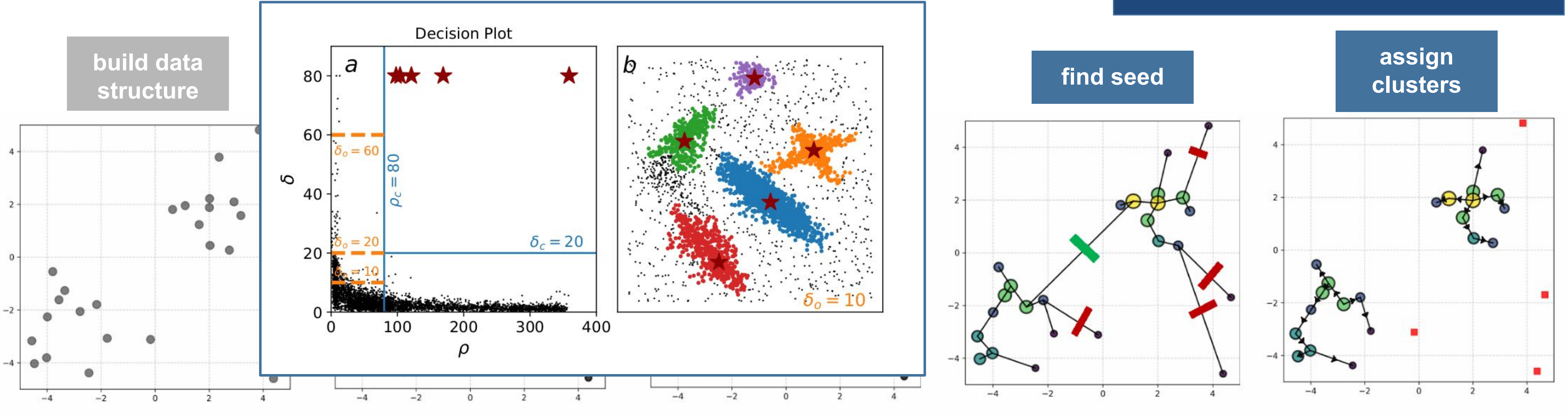
Cluster Algorithm: CLUE

ρ_c, d_c, δ_c , δ_0 are <u>tunable parameters</u> chosen with purity vs fake studies

- Promote as **seed** if $\rho_i > \rho_c$, $\delta_i > \delta_c$
- Demote as **outlier** if $\rho_i < \rho_c$, $\delta_i > \delta_o$
- Assign unique, progressive clusterID to each cluster
  - **Followers** are defined and associated to their closest seed

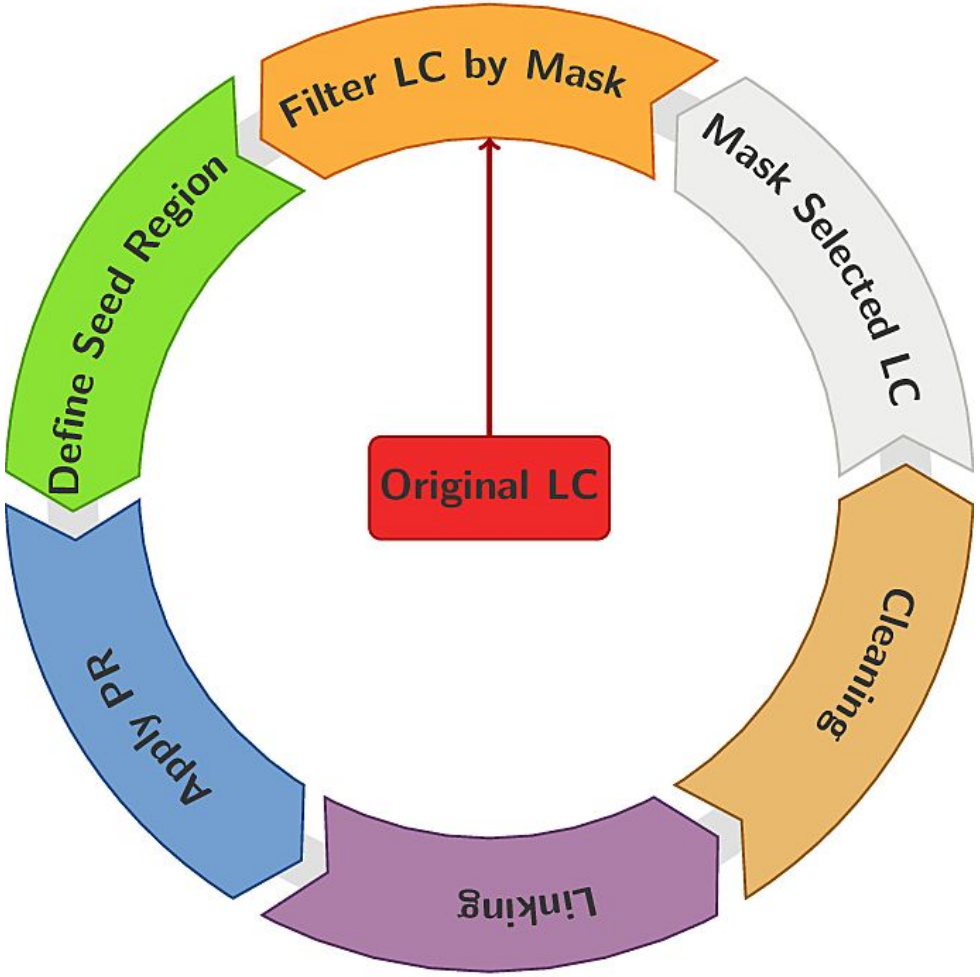
→ Rock solid against noise
→ Clustering almost all energy
→ Tested successfully also on test beam data

Example of decision plot



A TICL iteration

- Input: **Layer Clusters** built by CLUE
  Typical PU event: O(10^4) Layers Clusters/event out of O(10^5) RecHits
- **Pattern Recognition** is restricted to available Layer Clusters inside a **Seeding Region**
    - Track- or Global seeded
    - Currently using Cellular Automaton
- **Linking**, **Cleaning and Classification** to assign ID with a given *probability* may exploit Machine Learning
- Layer Clusters belonging to already identified physics objects are **masked-out** to reduce combinatorics
- Use timing information, where available



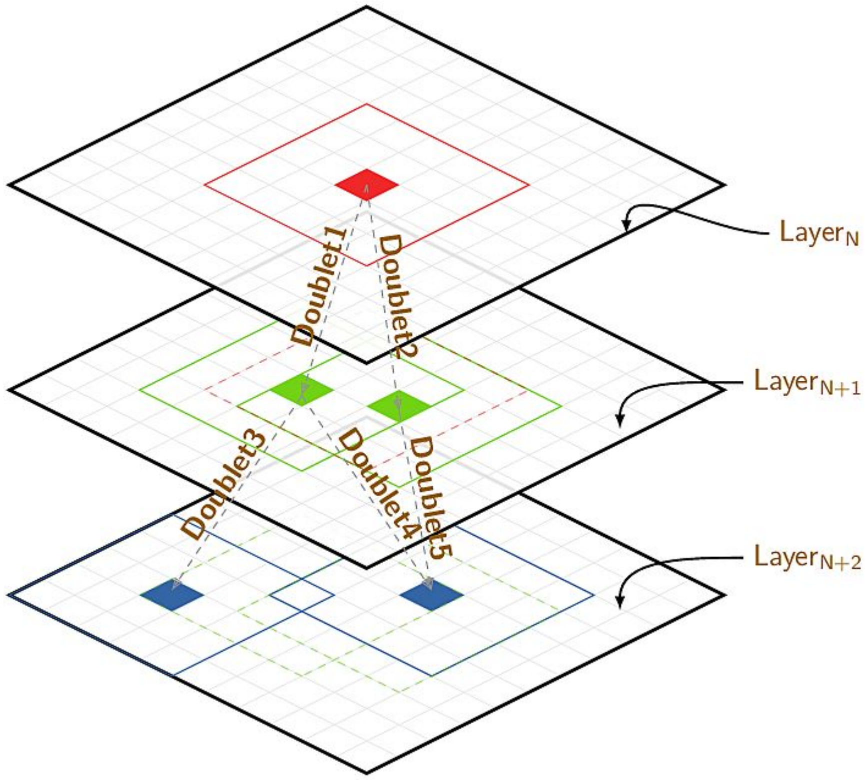
TICL Pattern Recognition

Algorithm: Cellular Automaton

1. **Doublet creation:**
   - For each 2D layer cluster in a $Layer_N$, open a search window in $Layer_{N+1}$
     - Search uses a 2D histogram in (η, φ)

       - Bin size is 0.05
       (i.e. ~ 70 mm at η = 1.6 and 20 mm at η = 2.8)

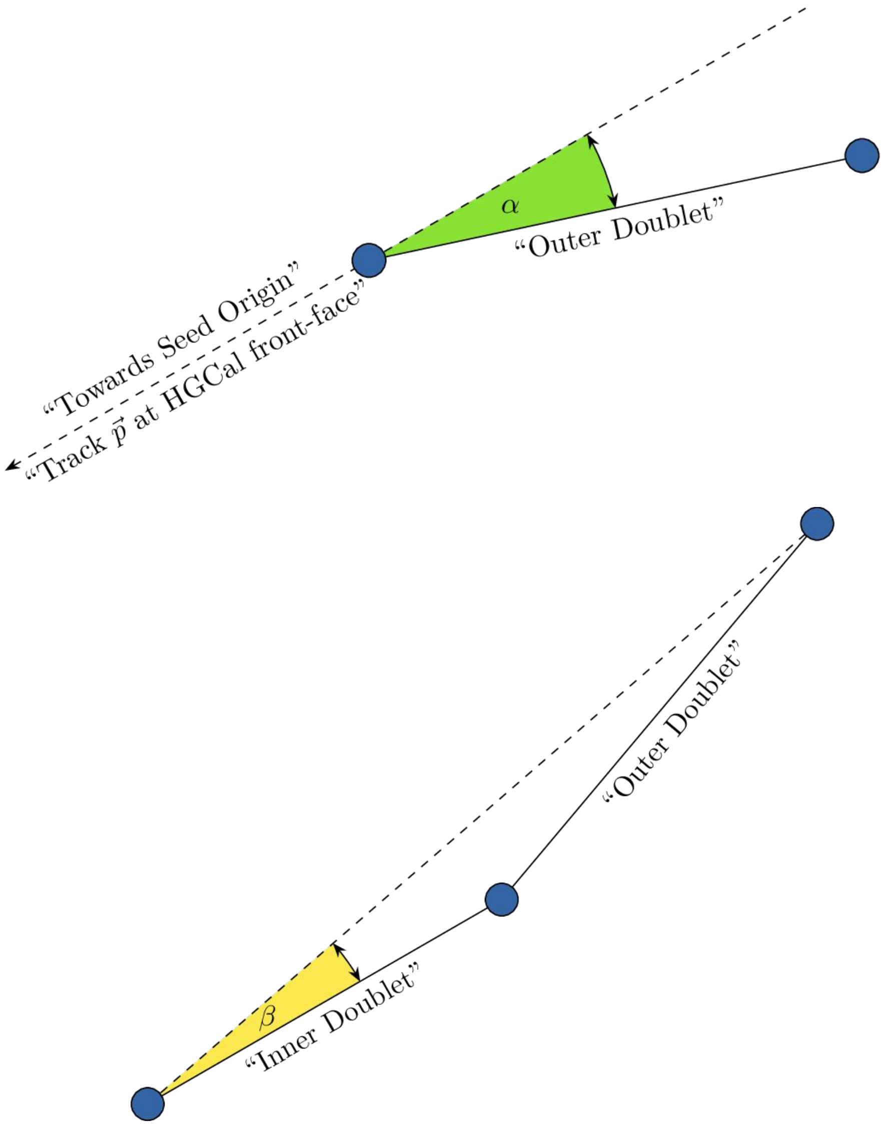
       - Search window is (3 × 3) [(5 × 5)] in the region
       η < 2.1 [η ≥ 2.1] bins centred on the bin in which the layer cluster sits
   - A layer cluster in this search window will make a doublet with the original layer cluster
     - Timing information is used to select compatible LCs



TICL Pattern Recognition

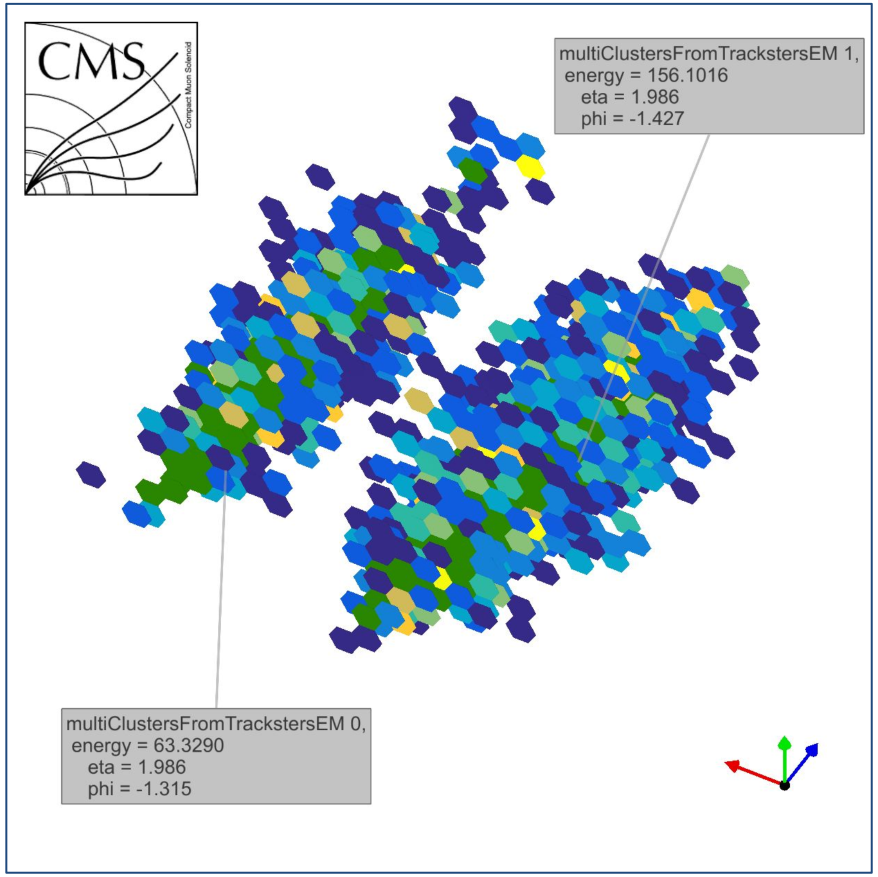
Algorithm: Cellular Automaton

1. **Doublet creation**
2. **Doublets linking**
   - If passing requirements on:
     - Direction of each doublet wrt the vertex (or wrt a track direction if this is a track seeded iteration)
     - Angle between the doublets
3. **Tracksters creation**
   - Collect the whole structure as a single trackster
   - Additional requirements on:
     - #(missing layers), timing info, etc.



TICL iterative scheme

- The following iterations have been implemented:
  - **track-seeded** (e.g. electrons, charged hadrons)
  - **electromagnetic** (e.g. electrons, photons)
  - **hadronic** (e.g. neutral hadrons)
  - **MIP-like** (e.g. muons)
    - This will require some modification to the CLUE algorithm in order to have high efficiency also for single cell (anti)-clusters.
    - Work on-going

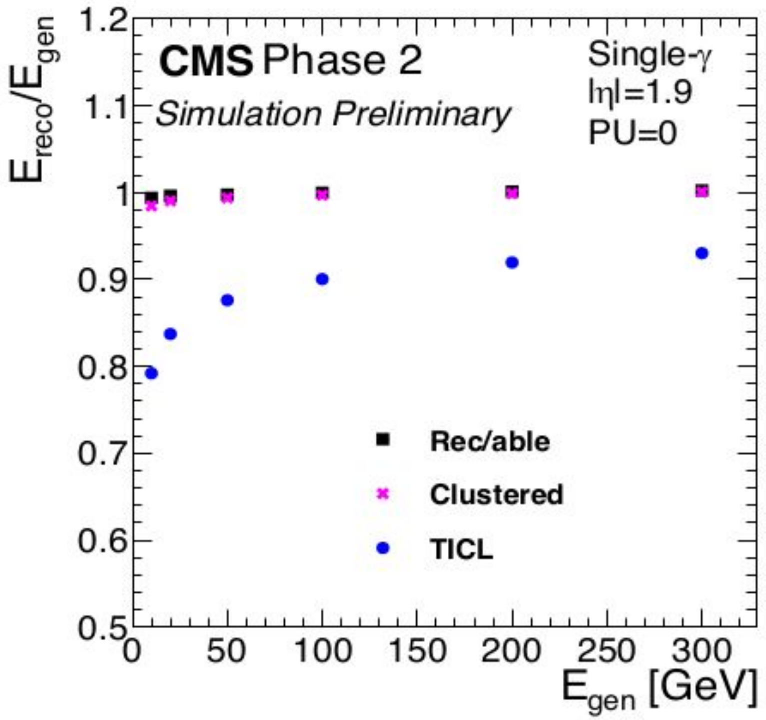


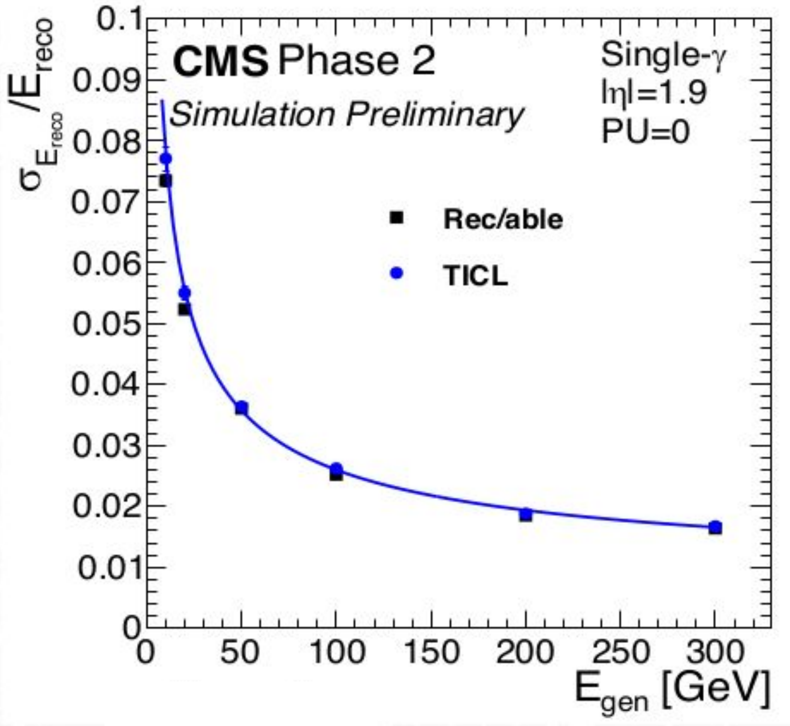
Example of two close-by reconstructed tracksters

Performance

Electromagnetic showers

- TICL has been tested on electromagnetic showers
- Collaboration with CMS objects groups for a full blown electron reconstruction with bremsstrahlung recovery
- Preliminary results are very encouraging





Rec/able: Sum of energy of all RecHits produced by the generated particle

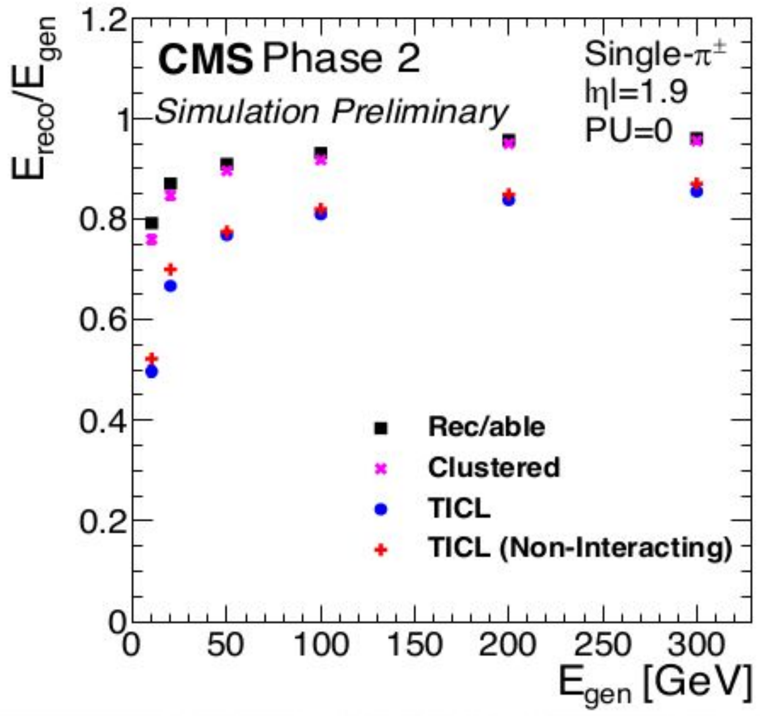
CLUE: Sum of the energy in the 2D clusters returned by the CLUE algorithm

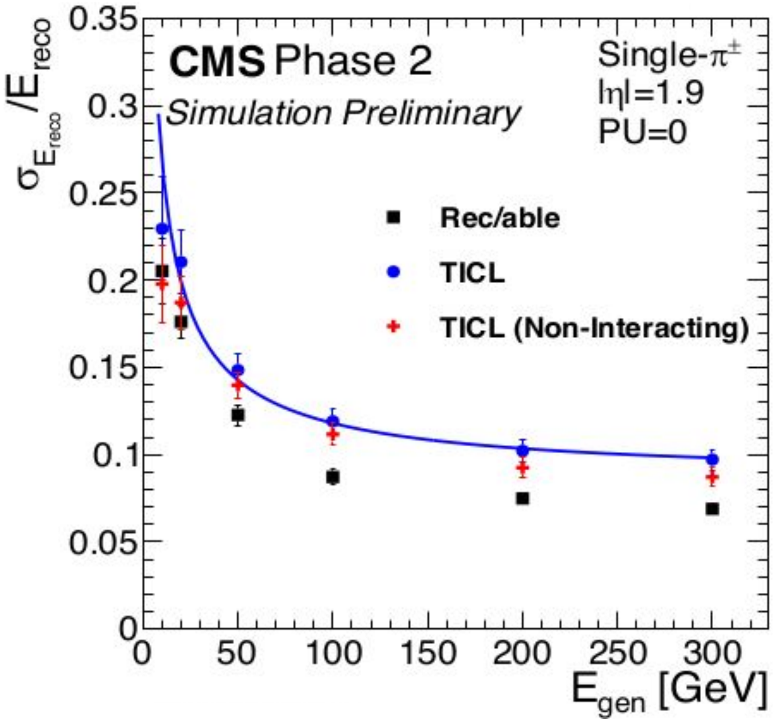
TICL: sum of the energy in the shower reconstructed using TICL

	Stochastic	Constant
Rec/able	23%	0.9%
TICL	25%	0.9%

Hadronic showers

- Hadronic showers vary in shape, and often resemble a branching tree of sub-structures
- Current approach: reconstruct the whole shower as a single trackster
- Preliminary results are very promising





Rec/able: Sum of energy of all RecHits produced by the generated particle

CLUE: Sum of the energy in the 2D clusters returned by the CLUE algorithm

TICL: sum of the energy in the shower reconstructed using TICL

TICL (non-interacting): sum of the energy in the shower reconstructed using TICL produced by pions that did not interacted in the tracker volume in front of HGCAL

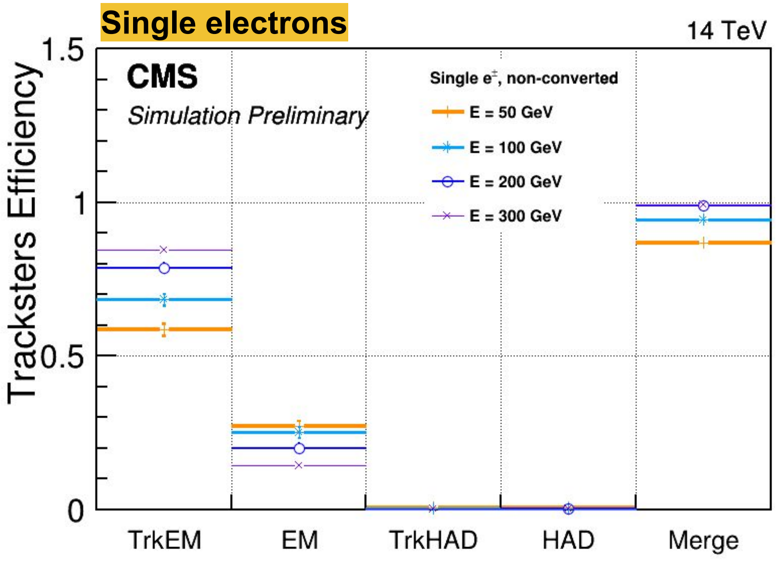
	Stochastic	Constant
Rec/able	72%	5.5%
TICL	80%	8.7%
TICL (non int)	72%	8.0%

Latest developments

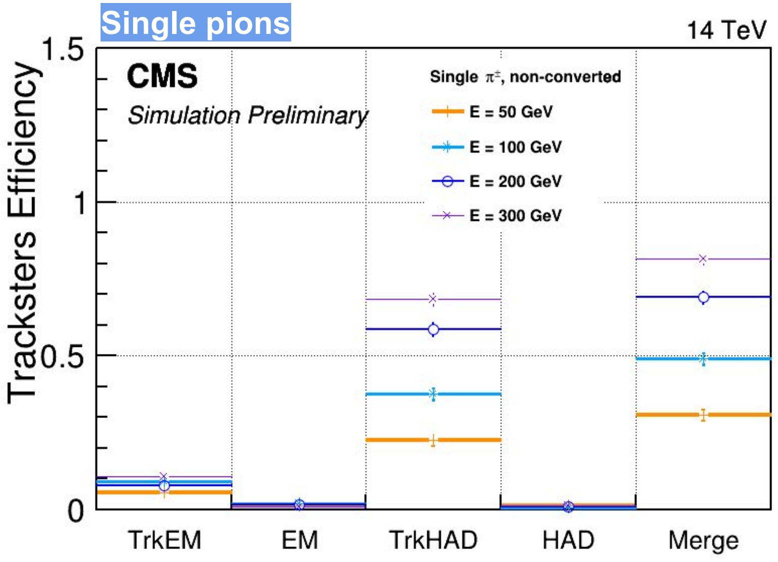
- Main recent changes on following topics:
  - Ordering of iterations:
    - Optimise order of the iterations
    - New iteration targeting electromagnetic showers seeded from a reconstructed track
  - Updates in Trackster reconstruction:
    - Reduce the impact of the high pile-up environment
  - Particle-Flow interpretation:
    - Each single trackster evaluated as a PF candidate according to the TICL iteration

- **Short-term goals achieved**
  - ✓ Enhance performance on single particles
  - ✓ Make it more robust against PU
    - in particular EM reconstruction

Single electrons



Single pions



Outlook & Conclusions

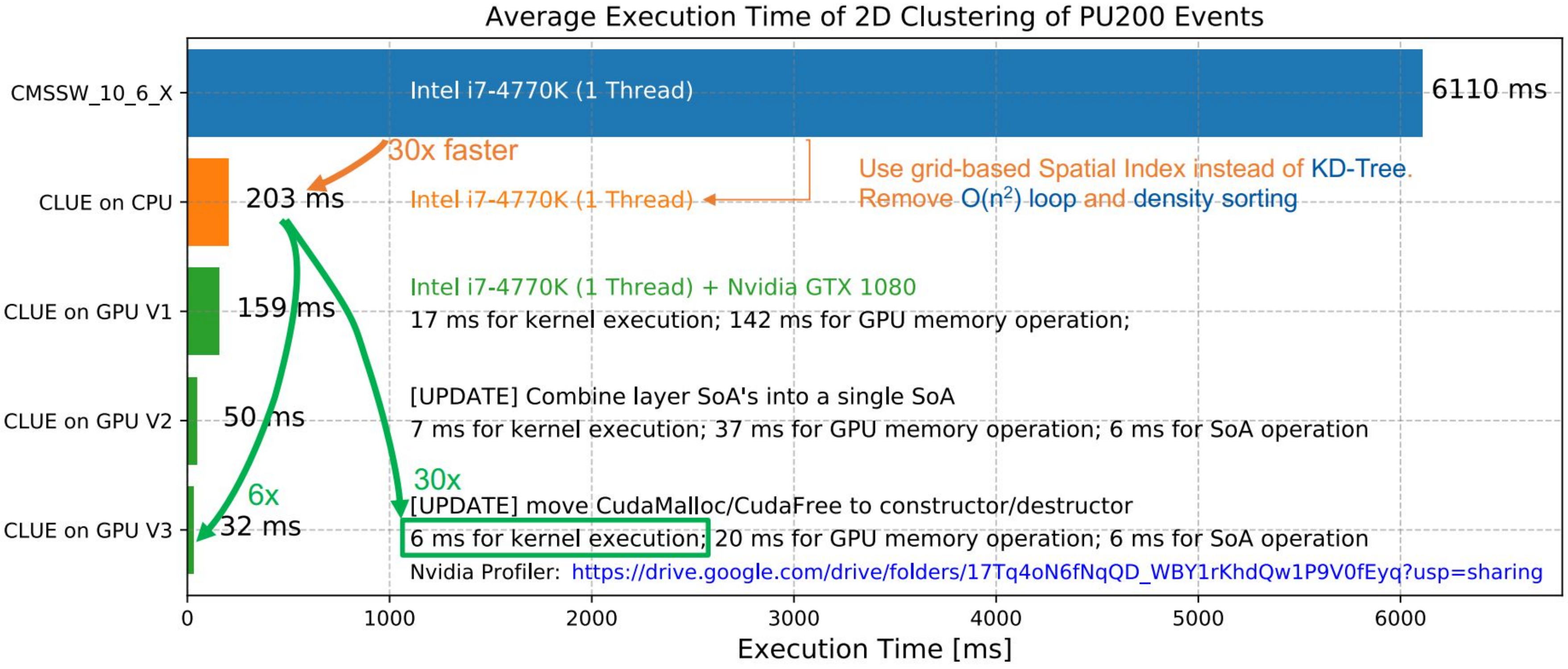
Conclusions

- The reconstruction in **High Granularity Calorimeters** at future colliders (HL-LHC and beyond) has many unprecedented challenges
  - A “tracking” device with high hit multiplicity and precise time information
- A **novel reconstruction algorithm** for imaging calorimeters has been developed in CMS
  - Follows a modular/factorized approach necessary to **exploit** the full potential of such detectors and **understand** the results
  - Fertile ground for the exploitation of **neural networks** and other novel algorithms
  - Was developed with **parallelism** in mind (CLUE in backup)
- First **results** are extremely encouraging (single particle with PU=0)
- **Next steps:**
  - Systematic studies on thresholds/cuts per-iteration
  - Improve implementation of PF-objects interpretation
  - Look into possibility of locally purify our tracksters from PU contribution
- Making it a **common library** would benefit experiments at future colliders and contribute to TICL’s improvement by using it with different topologies/geometries

Backup

Heterogeneous Reconstruction

- The complexity of the detector and the extrapolated computing needs of CMS pose challenges from the computation point of view
- Heterogeneous approaches are actively being explored and implemented
- 2D clustering has been fully ported and validated on GPUs (Ref.)



CMSSW
CLUE on CPU
CLUE on GPU

One order of magnitude in speed-up achieved with identical results

&

Similar speed-up in the CPU version

TICL-PF interpretation

- Very first draft
- All tracksters reconstructed in EM iter labelled as **photons**
- All tracksters reconstructed in HAD iter labelled as **kaons long**
- Tracksters from the **TrkEM iter** :
  - Check if there is another trackster coming from **TrkHAD iter** that is seeded by the very same track. If so, merge tracksters and the result injected as a **charged hadron**
  - Search the one most compatible (in η-Φ and p_T space) with the seeding track → labelled as **electron**
  - Remaining additional trackster are promoted as **photons**
- Trackster from the **TrkHAD iter** are promoted as charged hadrons
- All general tracks (high purity, p_T > 1) which have not been used are promoted as **charged hadrons**