Particle Flow Algorithms

Daniel Jeans, KEK

IAS Program on High Energy Physics
HKUST
Hong Kong
January, 2018

introduction and motivation

bias towards e+ e- collisions

general features of detectors and reconstruction

for optimal application of Particle Flow

a specific example

A particle interaction produces a collection of
  "stable" charged and neutral particles of different types:
    electrons, muons, photons, neutrinos, various hadrons

to get maximal information about the underlying interactions,
  and the physics processes producing them,
    ideally identify and measure **each** of these particles

this is the basic philosophy between
  particle flow

to do this in an optimal way needs
  a **detector** with dedicated design
    and
  reconstruction algorithms
    to analyse the data they collect

in high energy e+ e- collisions,
interesting processes often include one or more
W, Z, or H bosons

these bosons decay preferentially to quarks, which
produce **hadronic jets**:
collimated collection of hadrons, photons

→ a **majority** of interesting final states include **hadronic jets**

at high energy e+ e- colliders,
many measurements will be limited by statistics

to make best use of such a facility,
should make use of the
dominant hadronic component of final states
as well as possible

(this argument does not really apply to hadron colliders:
hadronic background processes are very large)

this leads to a focus on the **di-jet mass resolution**:
the ability to distinguish hadronic decays
(typically into a pair of jets)
of W, Z, H, …

Past and present collider experiments do not typically
have sufficiently good **Jet Energy Resolution**
to achieve this

The most promising avenue seems to be the use
of **Particle Flow** techniques
to measure **hadronic jets**

a W, Z, or H boson usually decays to a quark – pair

which hadronises to a collection of hadrons

mostly pions, Kaons, protons, …

some of which decay further

particularly neutral pions → 2 photons

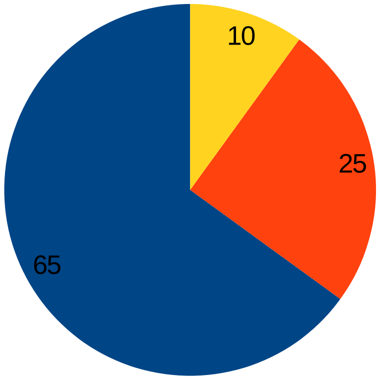
on average, the energy of a hadronic jet is

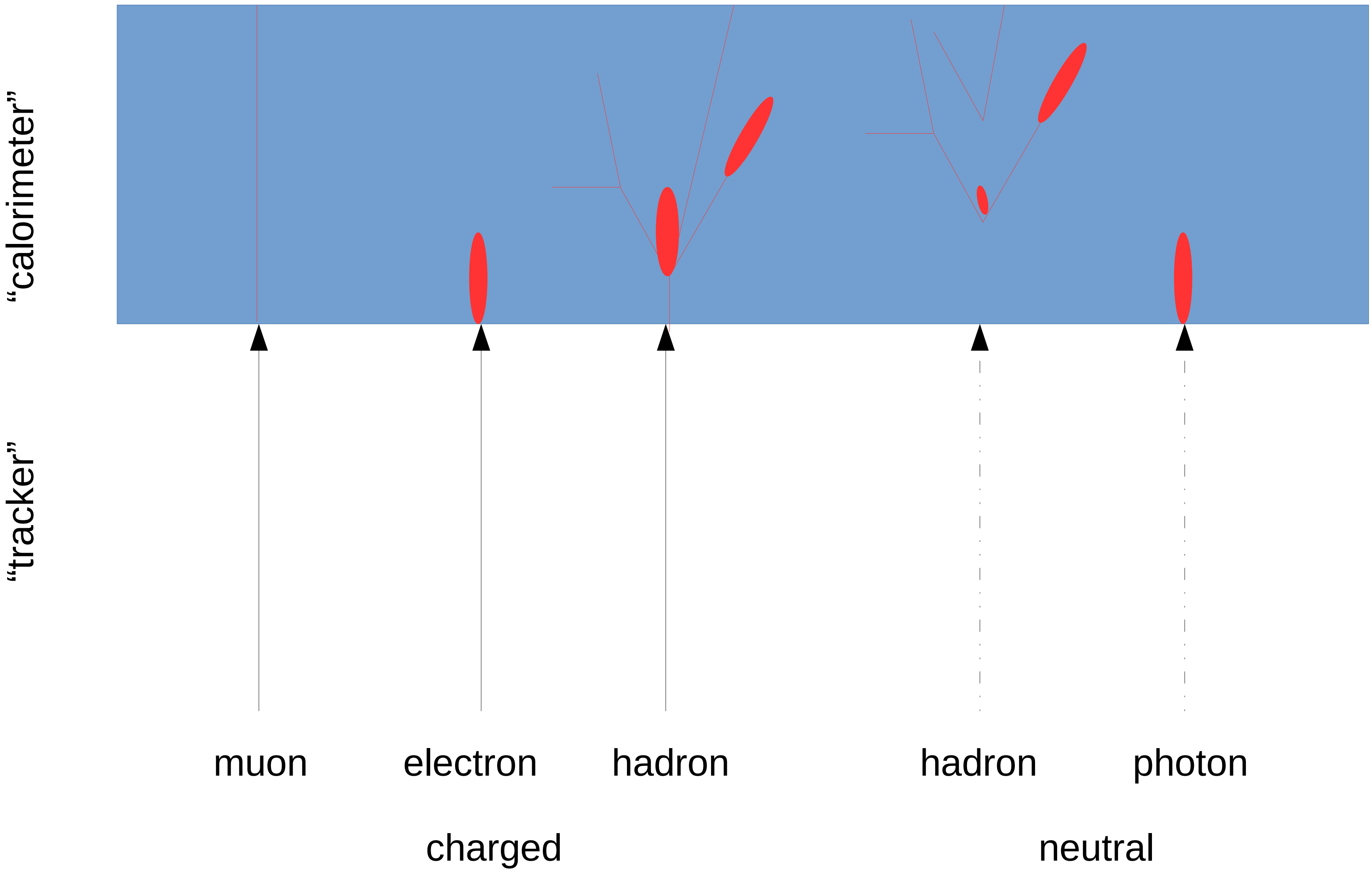
~ 65% **charged hadrons** (mostly pions)

~ 25% **photons** (mostly from pi0 decays)

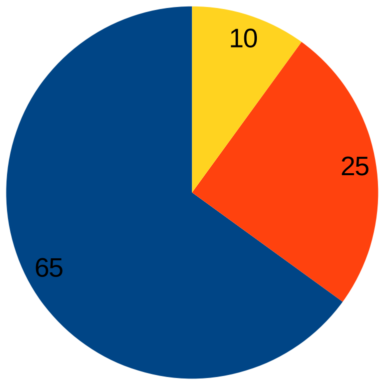
~ 10% **neutral hadrons**

these fractions fluctuate wildly from jet-to-jet




"calorimeter"
"tracker"
muon
electron
hadron
hadron
photon
charged
neutral

~ 65% **charged hadrons** (mostly pions)
~ 25% **photons** (mostly from pi0 decays)
~ 10% **neutral hadrons**



A traditional approach uses
calorimeter systems to measure jet energy
→ 75% of jet energy measured by hadron calorimeter

The calorimetric energy measurement of hadrons
suffers from large fluctuations
→ typical single hadron energy resolution
$\sigma/E \sim 100\% / \mathrm{sqrt}(E)$

However, charged hadron component is also
measured by the much more precise tracking detectors

If Particle Flow can be achieved within hadronic jets,
and each final state particle is measured individually,
→ potential to improve jet energy measurement

Steps were taken in this direction in the LEP detectors:

energy flow

e.g. ALEPH, reconstructed energy in e+e- → qq @ 91 GeV

"traditional approach" – calorimeter only:

jet energy resolution σ/E ~ 120% / sqrt(E)

"energy flow"

– identify electrons, photons, muons, V0s, ... in tracker+calorimeters

– assume remainder of calorimeter energy is deposited by charged and neutral hadrons

if E_{calo} >> $E_{trk,}$ associate extra calorimeter energy to neutral hadron

if E_{calo} ~ $E_{trk,}$ use track energy as estimator of total energy

neutral hadrons **inferred** from **energy imbalance**

susceptible to large fluctuations

in single particle energy response

"energy flow" resolution σ/E ~ 59% / sqrt(E)

→ factor 2 improvement!

ALEPH collab,
Performance of the ALEPH detector at LEP,
Nucl.Instrum.Meth. A360 (1995) 481

Energy Flow → **Particle** Flow

infer presence of neutral hadrons from trk-calo imbalance → explicitly reconstruct neutral hadrons

if we could perfectly identify each particle in a “representative” detector,

charged energy ~ 65% on average

measured by **tracker** (<<1% resolution for typical momenta)

photon energy ~ 25% on average

measured in **E-CAL** $\sigma/E \sim 15\ \% / \mathrm{sqrt}(E)$ for sampling calorimeter

neutral hadrons ~ 10% on average

measured in (E+**H)-CAL** $\sigma/E \sim 50\text{-}100\ \% / \mathrm{sqrt}(E)$

could get **jet energy resolution** $\sigma/E \sim$ **14%** / sqrt(E)

In practice, resolution limited by **confusion**

→ imperfect identification of charged and neutral calorimeter energy deposits

→ leads to over– or under– estimation of energy

→ worsens energy resolution, particularly for highly collimated, high energy jets in which showers are larger, and closer together

implications for detector design

The aim of Particle Flow is to make less use
of the relatively imprecise calorimetric measurement

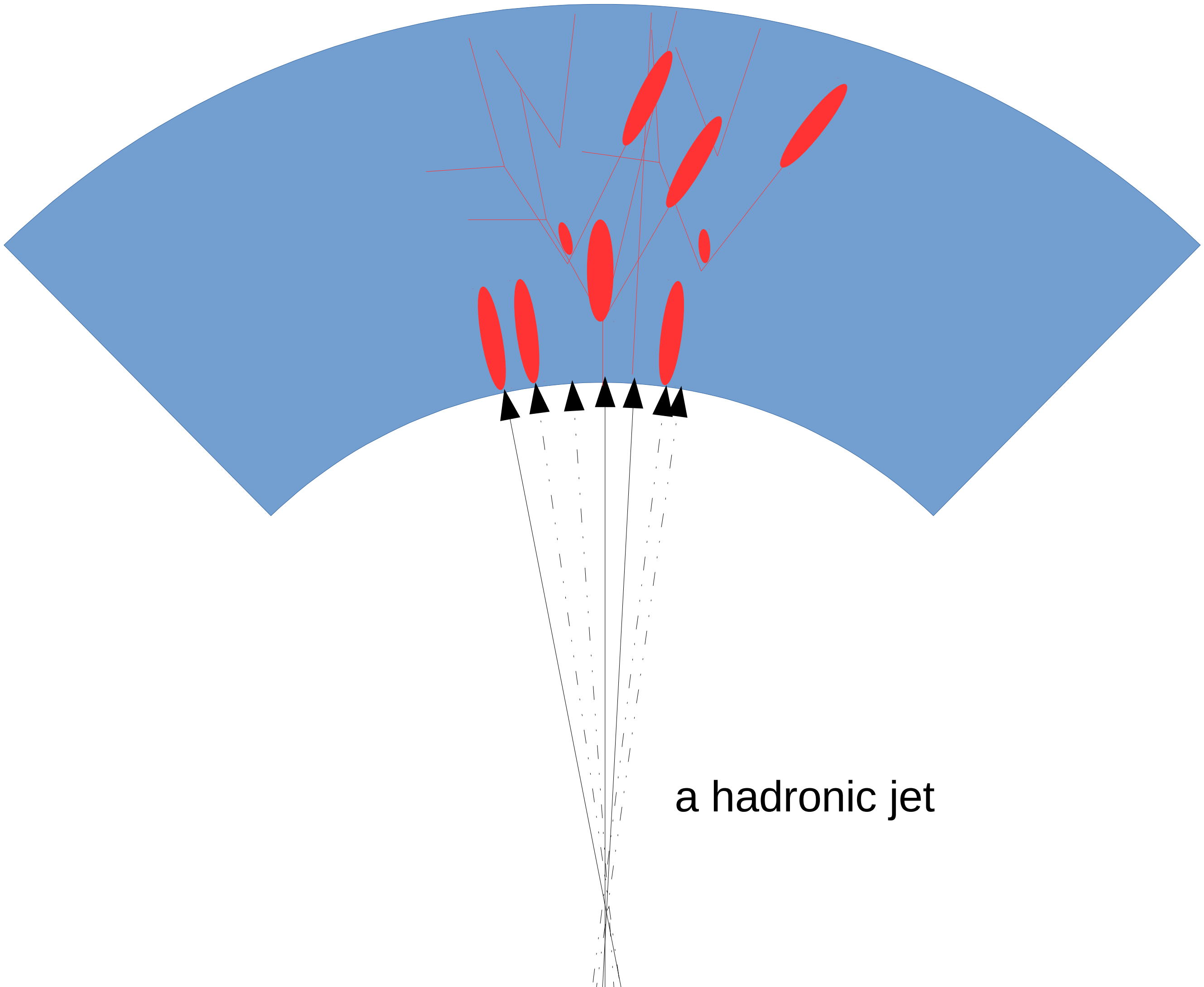
→ single particle energy resolution is not
our only figure of merit

→ reduction of “confusion” is more important

→ choose materials to give small particle showers,
easier to separate from neighbours

→ highly segmented readout
to allow separation of nearby showers
(in 3 dimensions, and
possibly also using energy, time information)

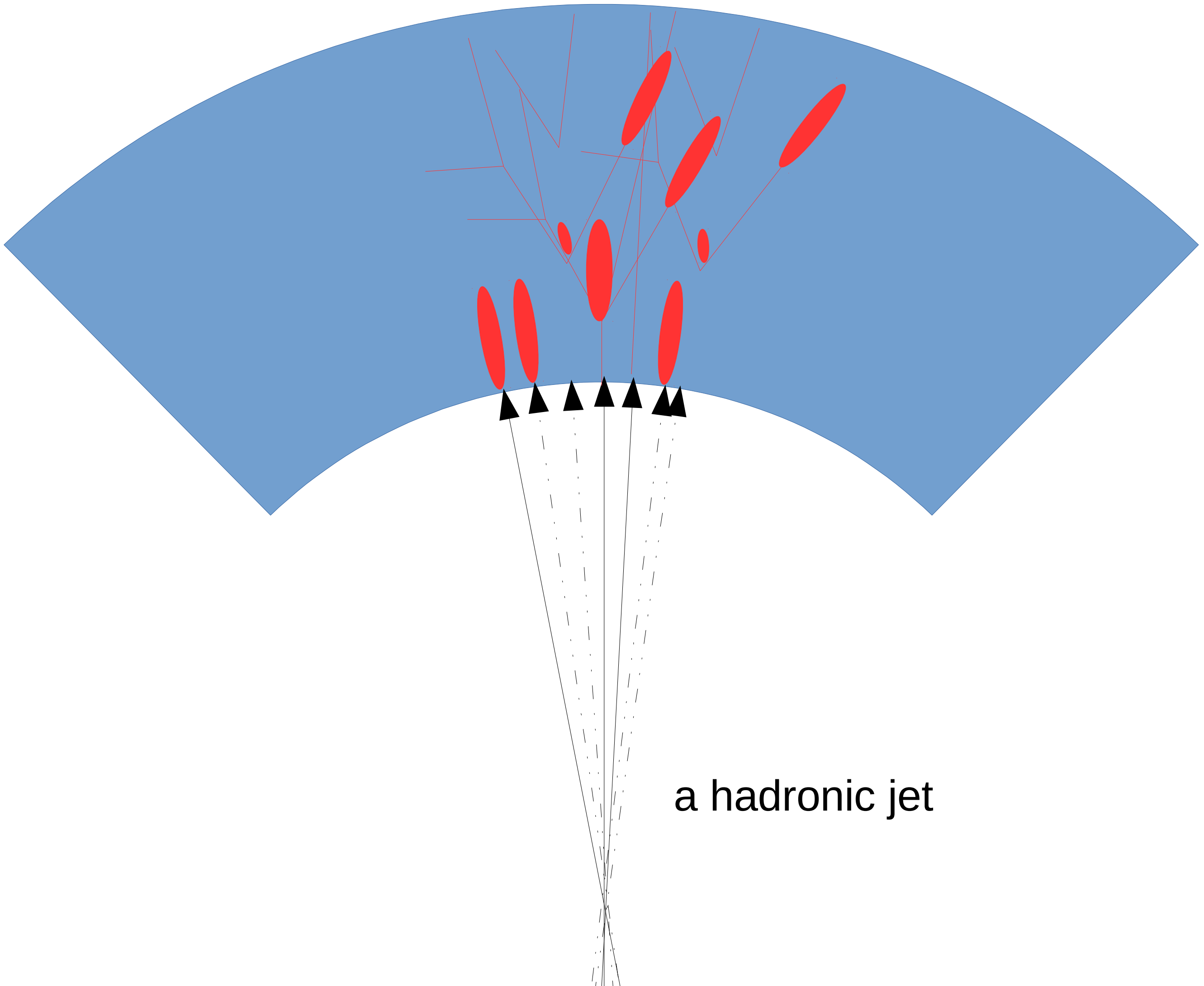
high density calorimeters,
made of small X0 and Moliere radius materials (e.g. tungsten)

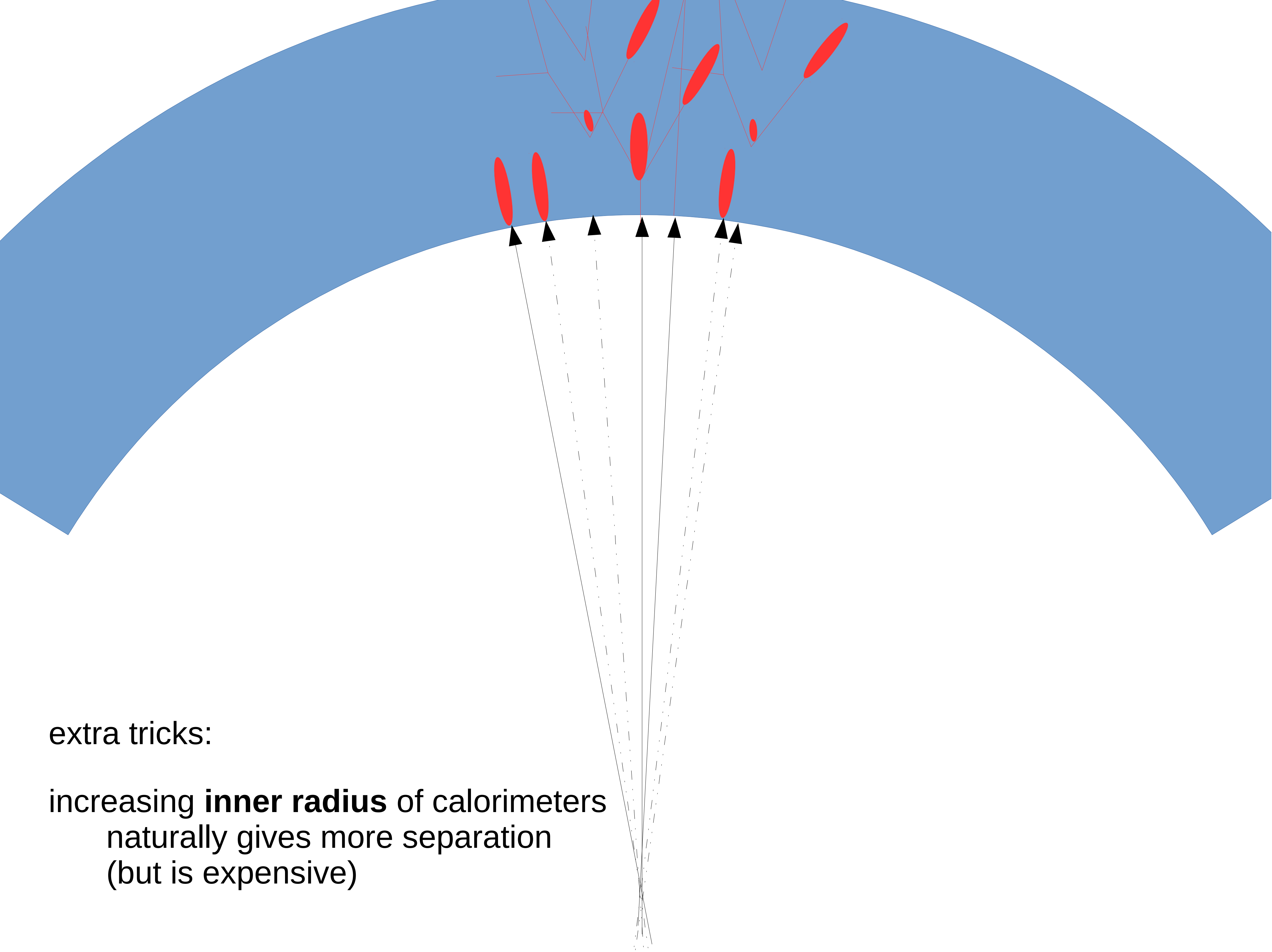


segmentation:

sampling calorimeter has natural segmentation in one dimension
readout layers can be segmented to give 3-d segmentation

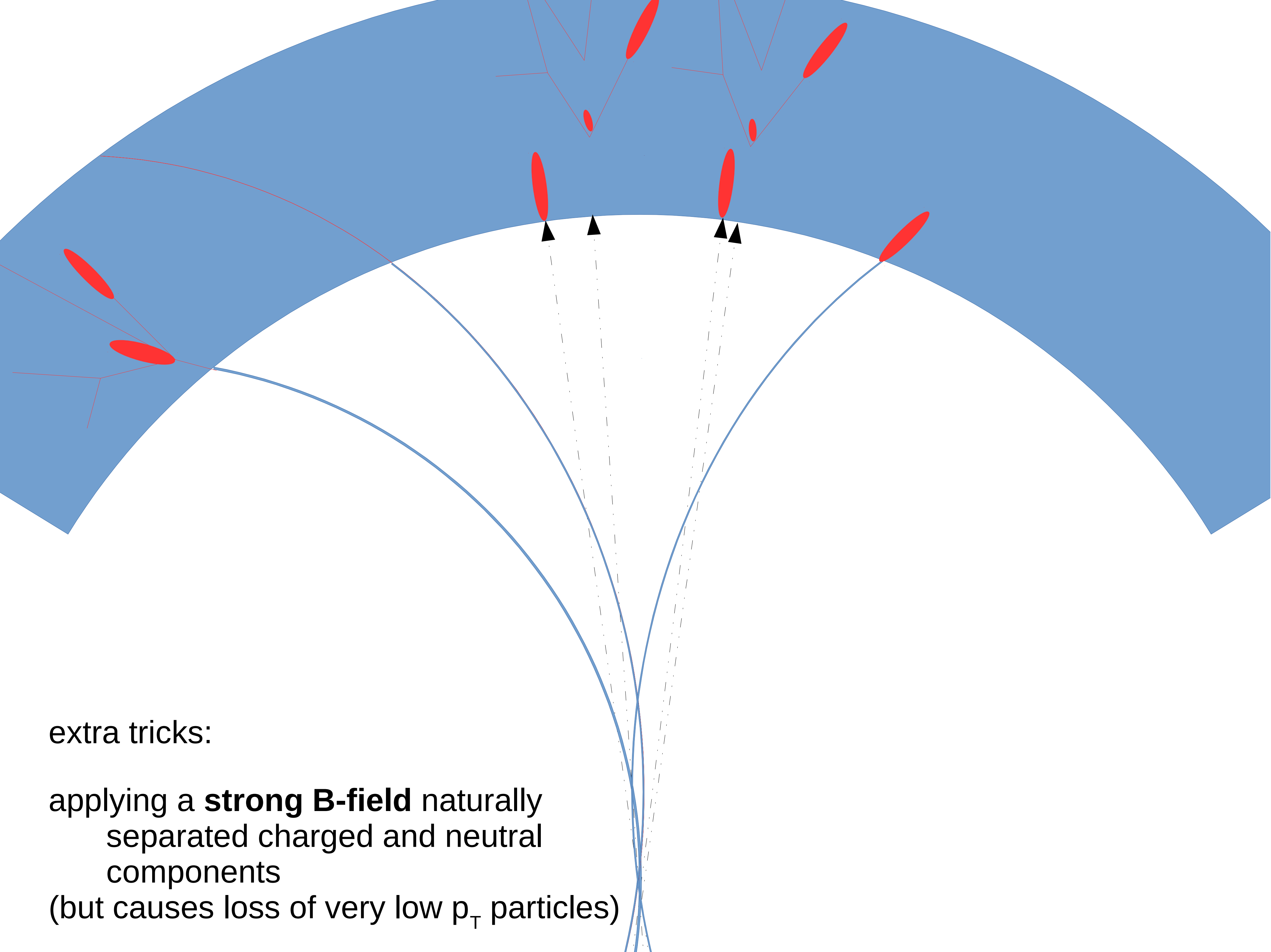
transverse segmentation: scale set by Moliere radius
longitudinal segmentation: determines the single particle energy resolution





extra tricks:

increasing **inner radius** of calorimeters
naturally gives more separation
(but is expensive)



extra tricks:

applying a **strong B-field** naturally
separated charged and neutral
components
(but causes loss of very low p_T particles)

key to particle flow:

identify calorimeter deposits of neutral particles

(mostly) by **topological** means

i.e. without comparing calorimeter energy to track momentum

benefits from calorimeter

→ with high **spatial granularity**

→ see sub-structure of showers (particularly hadronic)

helps to separate nearby particle showers

→ with **high density,**

ECAL: small radiation length X_0 , Moliere radius

large hadron interaction length λ: separate EM and HAD showers

HCAL: small λ in HCAL

→ compact particle showers, reduce overlap

→ at **large distance** between IP and calorimeter

→ allow particles to drift apart

→ in strong **magnetic field**

→ sweep charged particles away from neutrals

→ as little as possible **material** in front of calorimeters:

especially hadronic interactions can cause much confusion

sophisticated **pattern-recognition software**

→ **Particle Flow Algorithms**

Particle Flow **Detectors** (many more details in later talks)

ILD, SiD @ ILC, CLIC detector
CEPC's detector
CMS HGCAL upgrade

Particle Flow **Algorithms**

PandoraPFA

M. Thomson,
Particle Flow Calorimetry and the PandoraPFA Algorithm,
NIM A611 (2009) 25-40

J. Marshall, M Thomson,
The Pandora Software Development Kit for Pattern Recognition,
Eur.Phys.J. C75 (2015) no.9, 439

GARLIC

D. Jeans, J-C. Brient, M. Reinhard,
GARLIC: GAmma Reconstruction at a LInear Collider experiment,
JINST 7 (2012) P06003

ArborPFA

M. Ruan, H. Videau,
Arbor, a new approach of the Particle Flow Algorithm
arXiv:1403.4784

CMS PFA

CMS collaboration
Particle-flow reconstruction and global event description with the CMS detector
JINST 12 (2017) no.10, P10003

generic outline of particle flow algorithms

- identify “easy” components of showers

    MIP-like **calorimeter tracks** from muons:
        extrapolate identified tracks through calorimeter

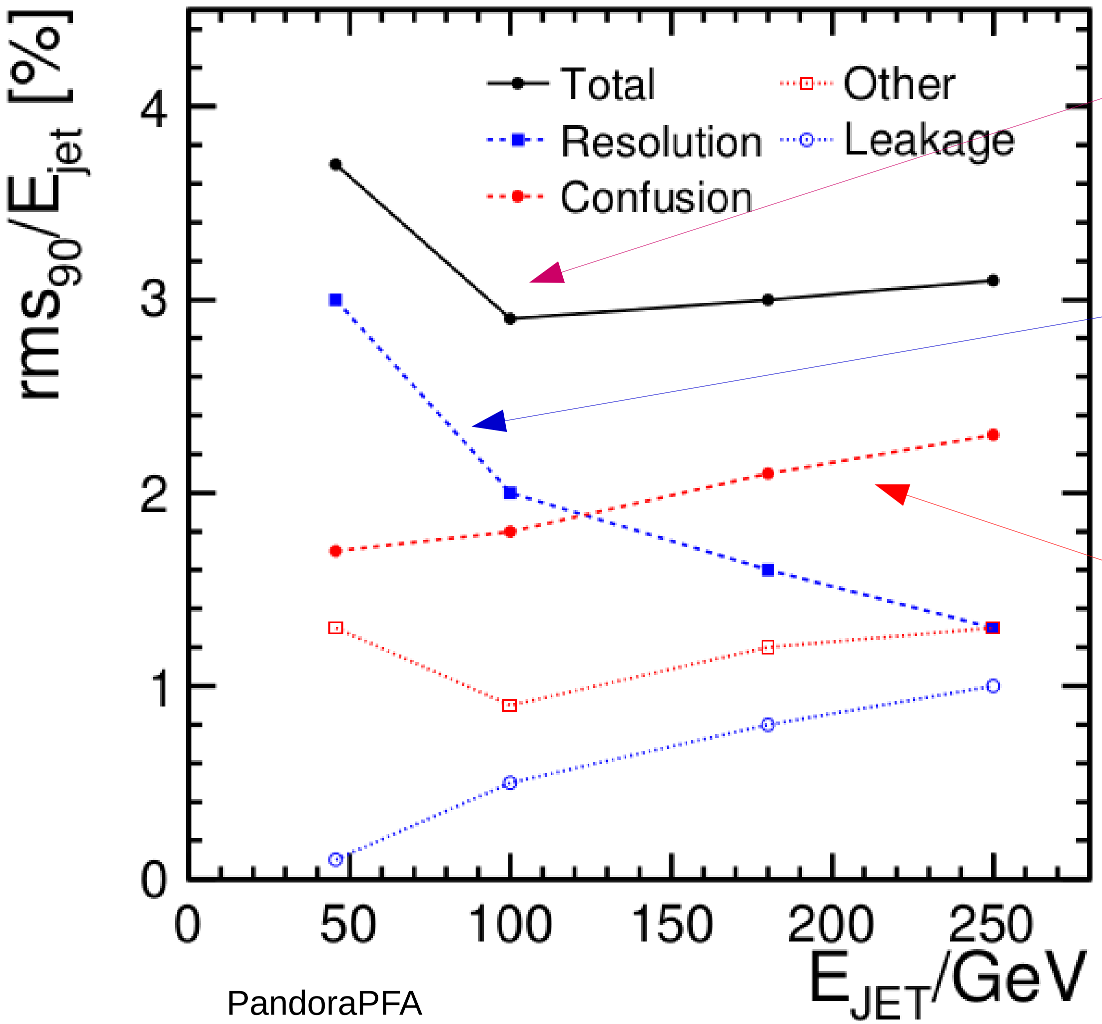
    electrons: **charged track** ending with characteristic **E-M shower**
        → complication: bremsstrahlung

    photons: characteristic **EM shower** without track

- treat remaining tracks and calorimeter deposits
    small-scale local (possibly directional) **clustering** in calorimeters
    combination of clusters with each other and tracks
        creating **“particle-flow”** objects

    typically based mostly on topology
        use energy compatibility to check for errors
            reclustering → iterative approaches

PandoraPFA jet energy resolution



PandoraPFA
NIM A611 (2009) 25-40

at 100 GeV,
$\sigma/E \sim 30\% / \mathrm{sqrt}(E)$

at lower jet energies <100 GeV,
energy resolution largely limited
by intrinsic **calorimetric** performance

at higher energy, jets have
more particles, higher boost
→ smaller distance between particles
→ more overlap between
calorimeter showers
→ pattern recognition
becomes more challenging
→ **confusion**

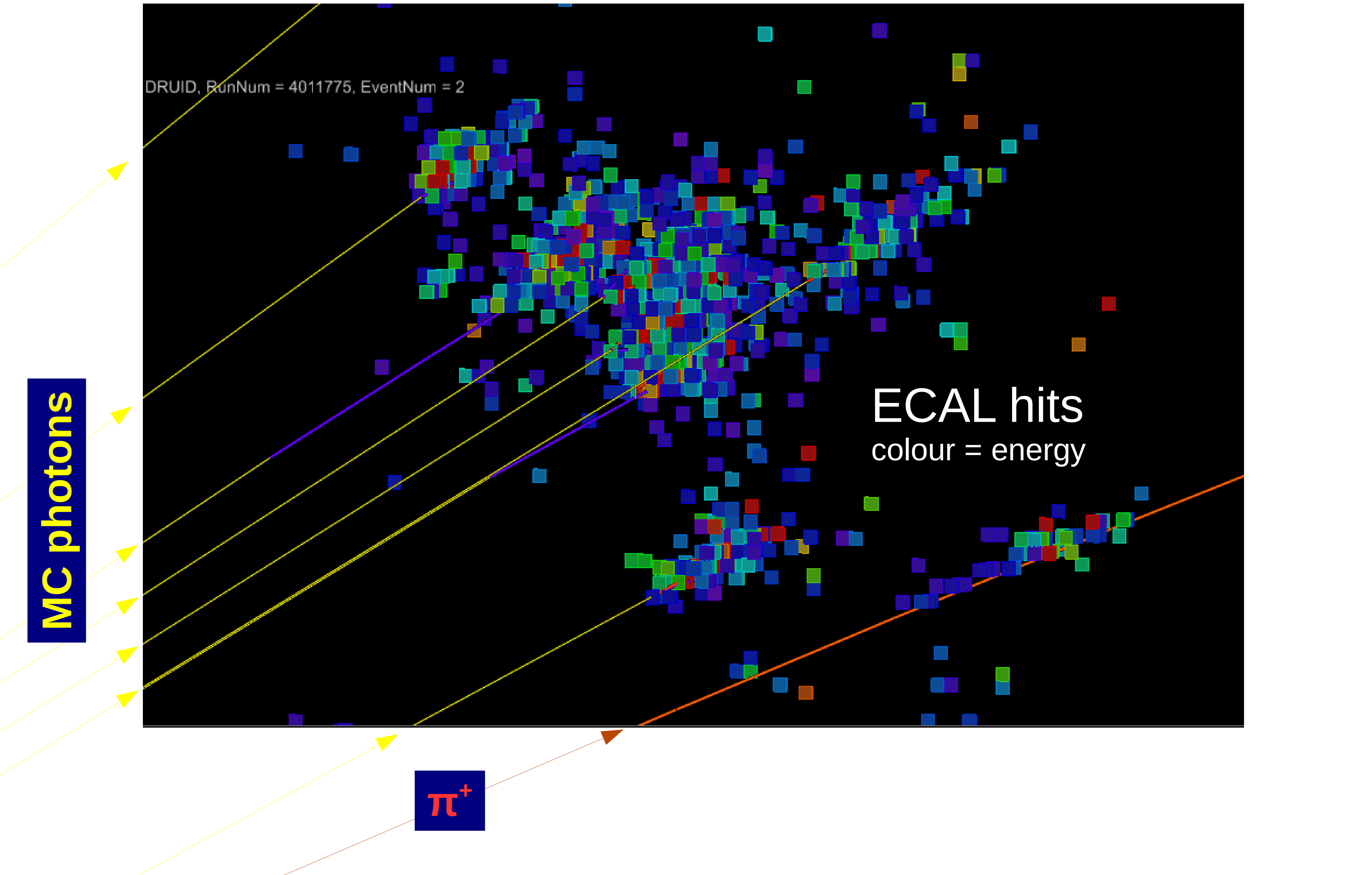
GARLIC
Gamma Reconstruction at a Linear Collider

Photon identification in hadronic jets
in a highly segmented calorimeter

not a full Particle Flow algorithm,
rather one element of a full reconstruction

D. Jeans, J-C. Brient, M. Reinhard,
GARLIC: GAmma Reconstruction at a LInear Collider experiment,
JINST 7 (2012) P06003

Dense portion of 250 GeV jet in ILD SiW ECAL



GARLIC
Gamma Reconstruction at a Linear Collider

Photon identification in hadronic jets
in a highly segmented calorimeter

Algorithm

Track veto
Remove hits close to extrapolated tracks

Seed finding
Identify cluster seeds in first part of ECAL

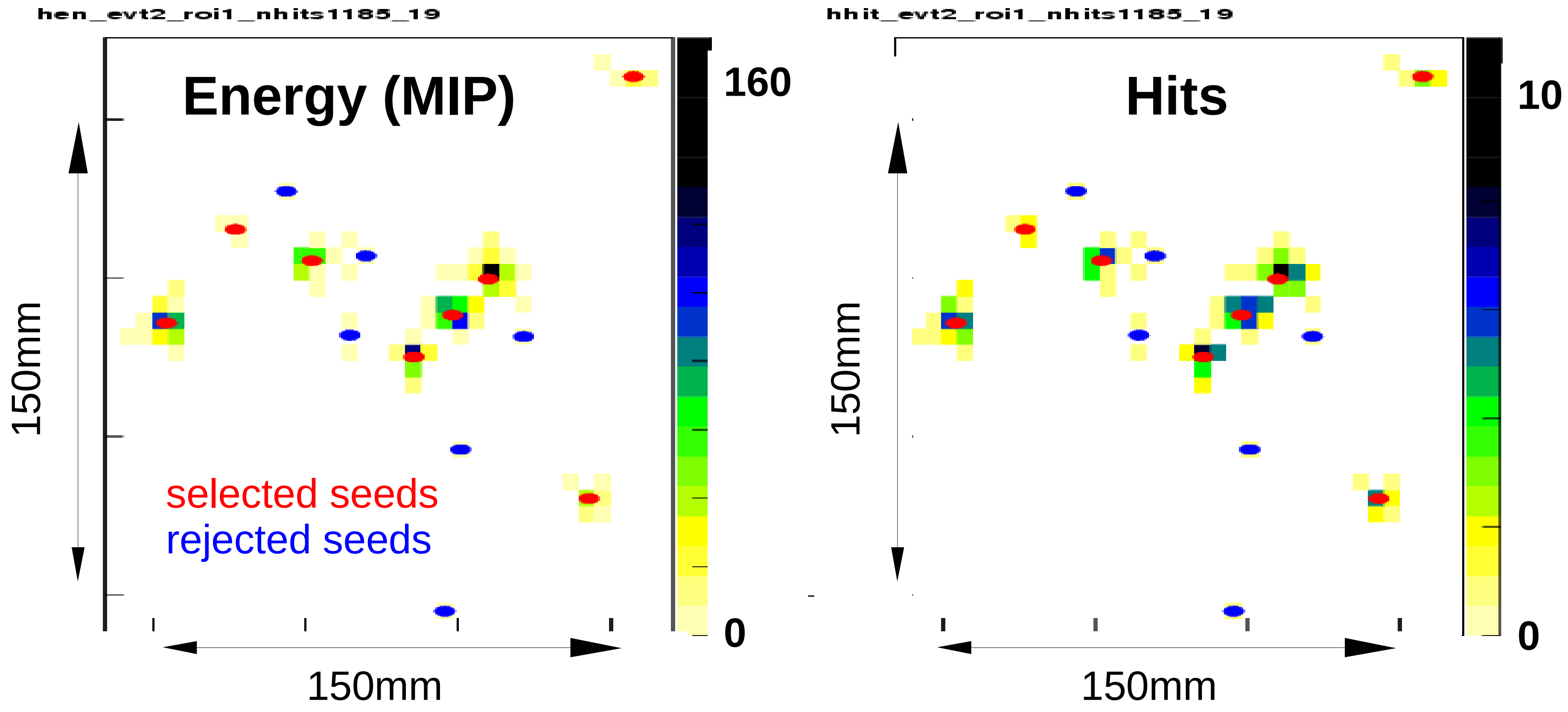
Core building
Build dense core of EM shower

Final clustering
Add nearby hits: “halo” around the core

Neural Network identification
Decide if cluster is photon-like

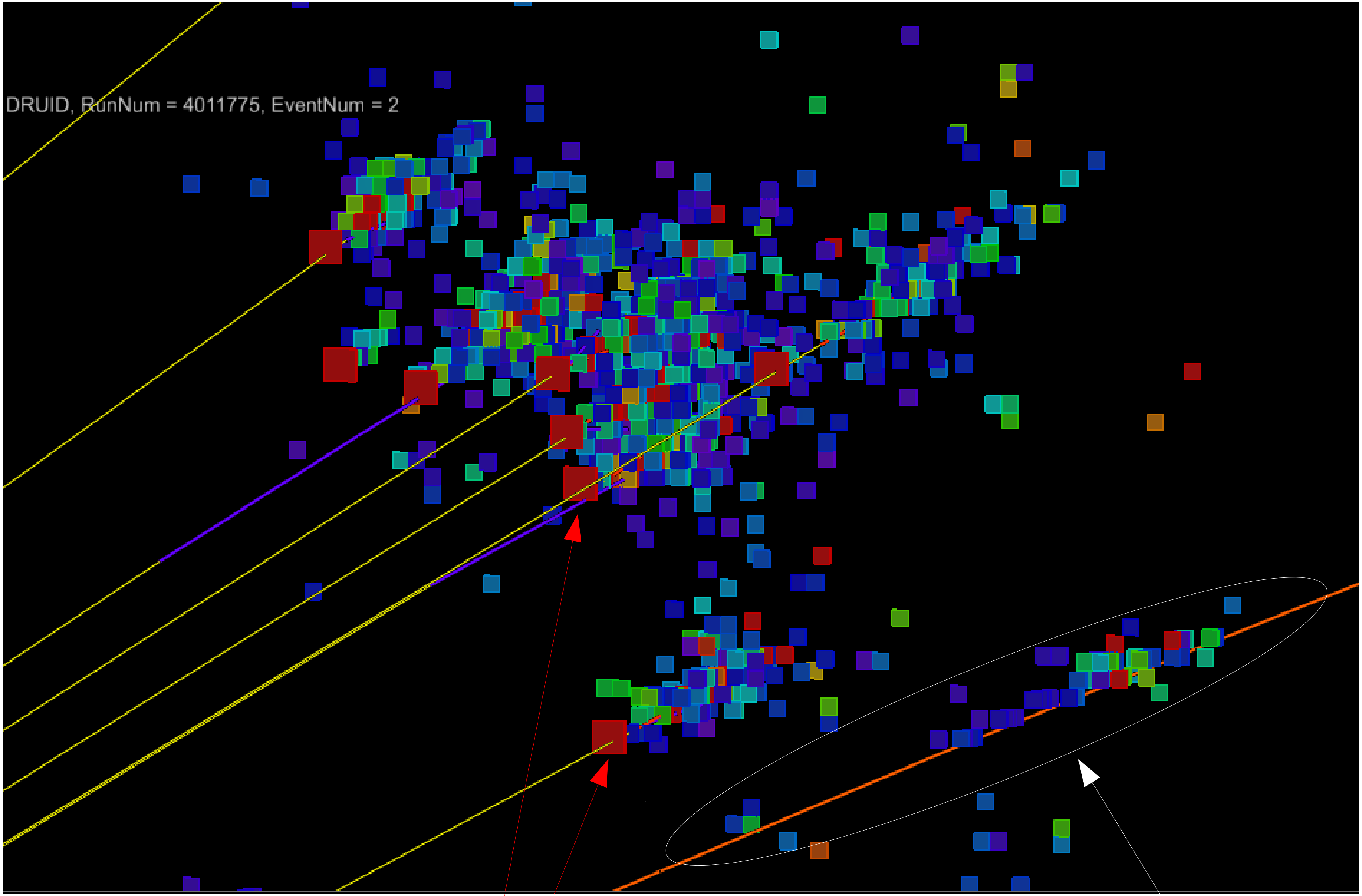
Project hits from first ~half of ECAL onto front face

simple nearest neighbour clustering
cluster seed candidates
requirements on energy and number of hits



→ defines a set of cluster **SEEDS**

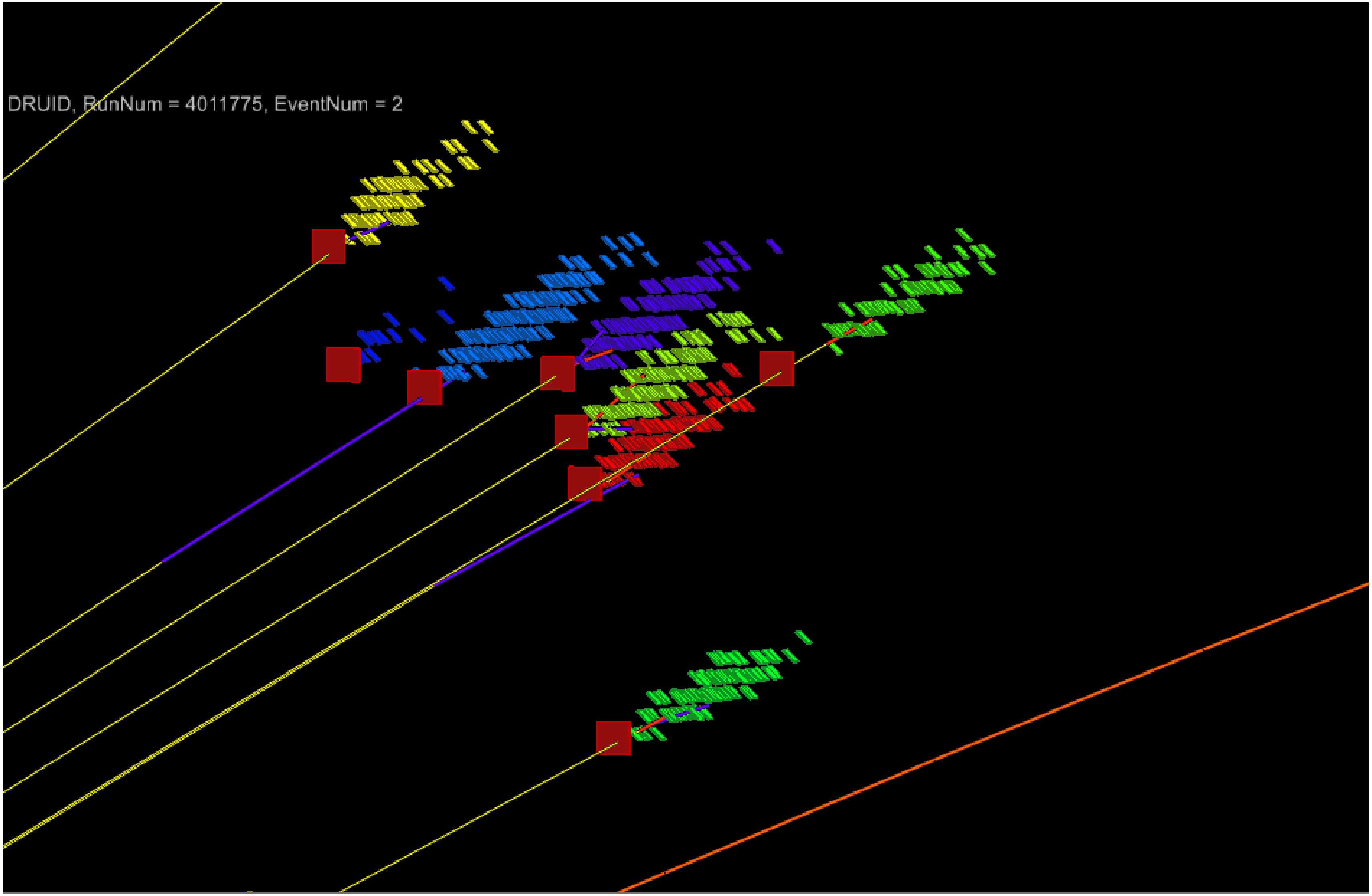
Track veto, seed finding



Cluster seeds

Remove hits near track extrapolation

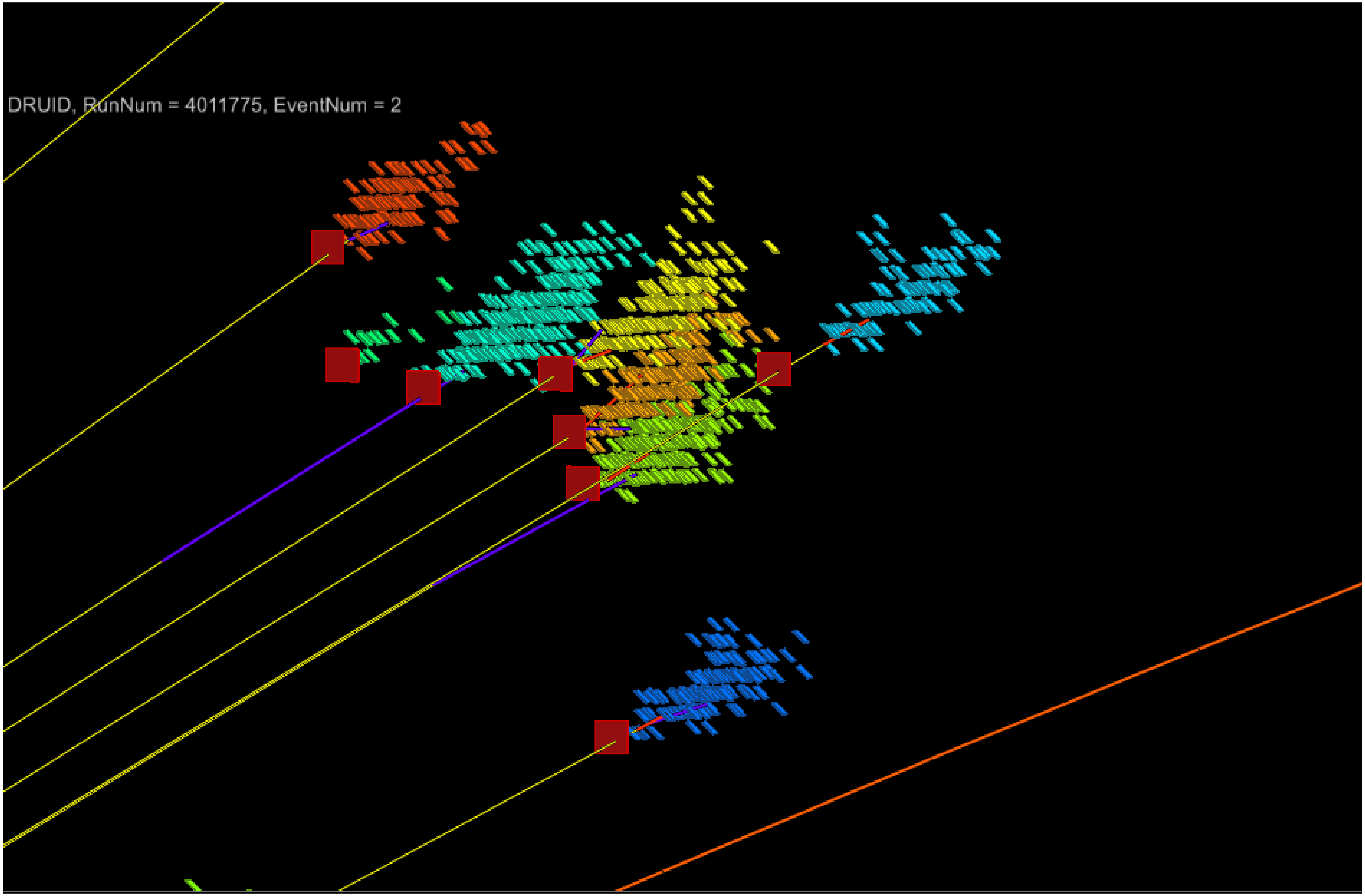
Core building : radius ~ cell size
project seed positions into ECAL



High energy core of shower is << Molière radius

Final clustering : radius ~ Molière radius

Add nearby hits to shower cores

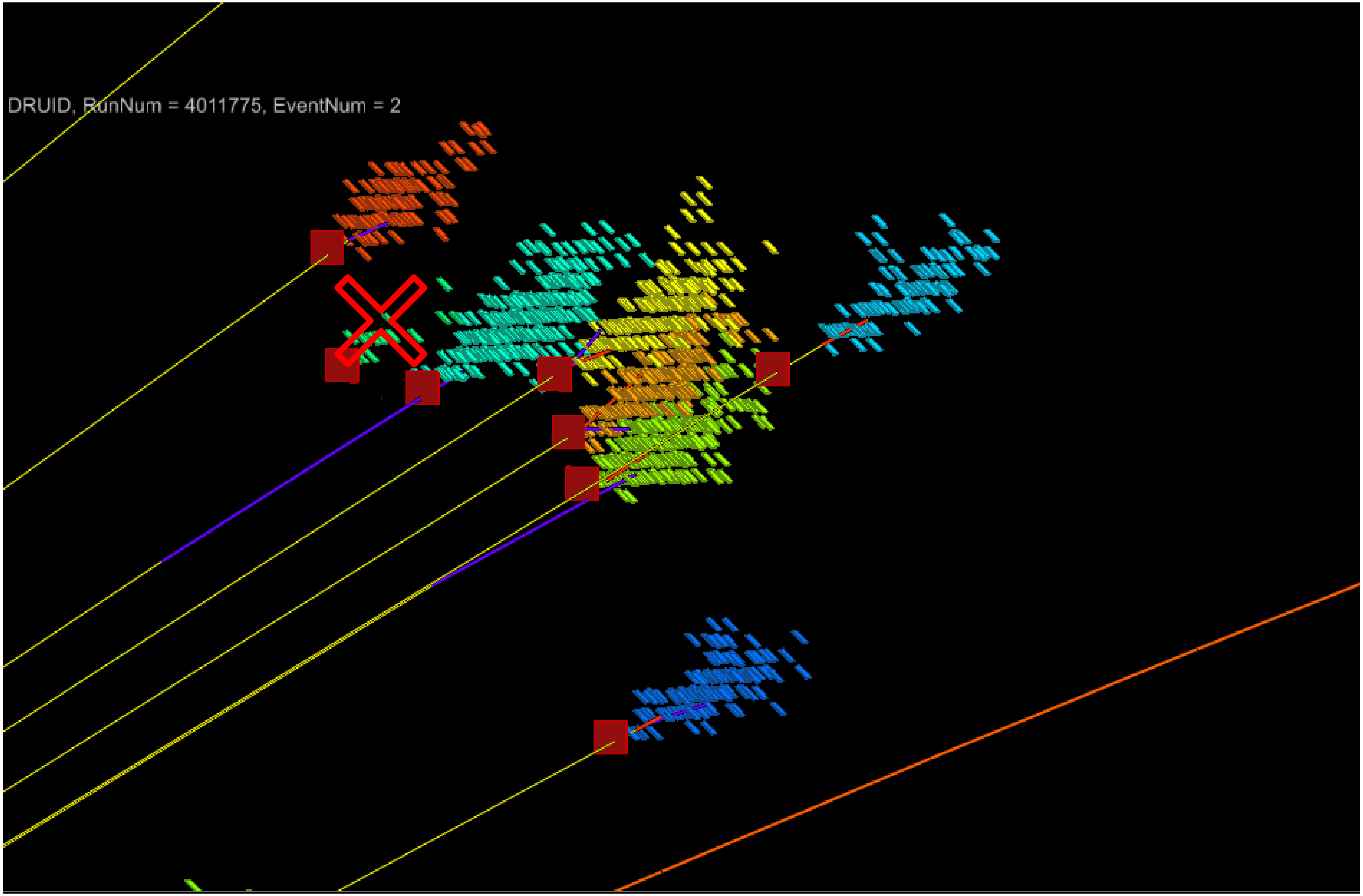


Collect large majority of shower hits, and almost all of the energy

Loosely restrict window size to prevent “eating” nearby showers

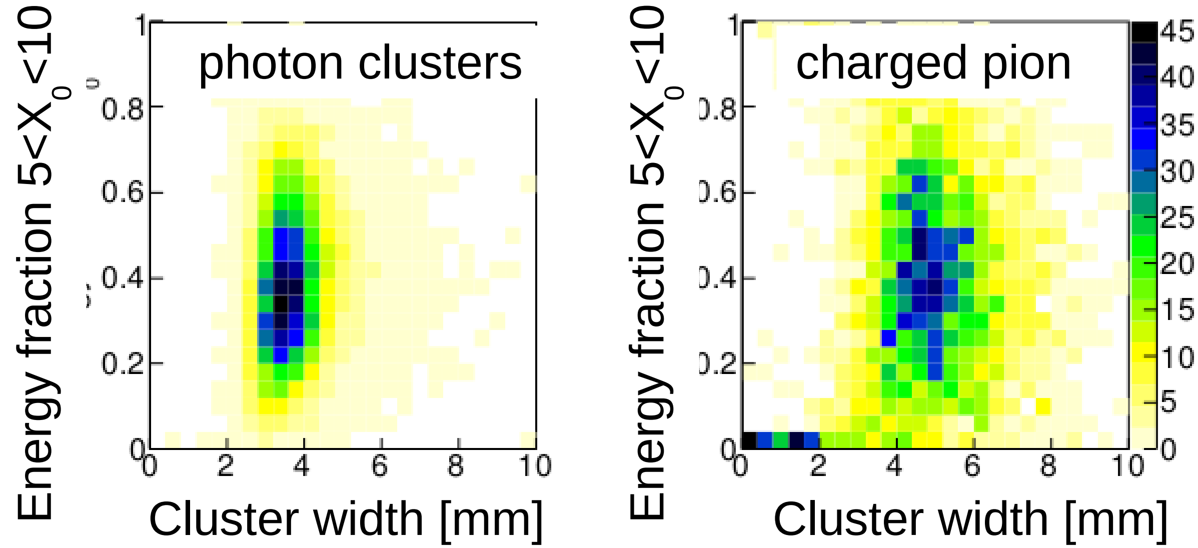
Neural Network-based selection:

reject clusters which don't look like photon showers



Example of some NN input variables

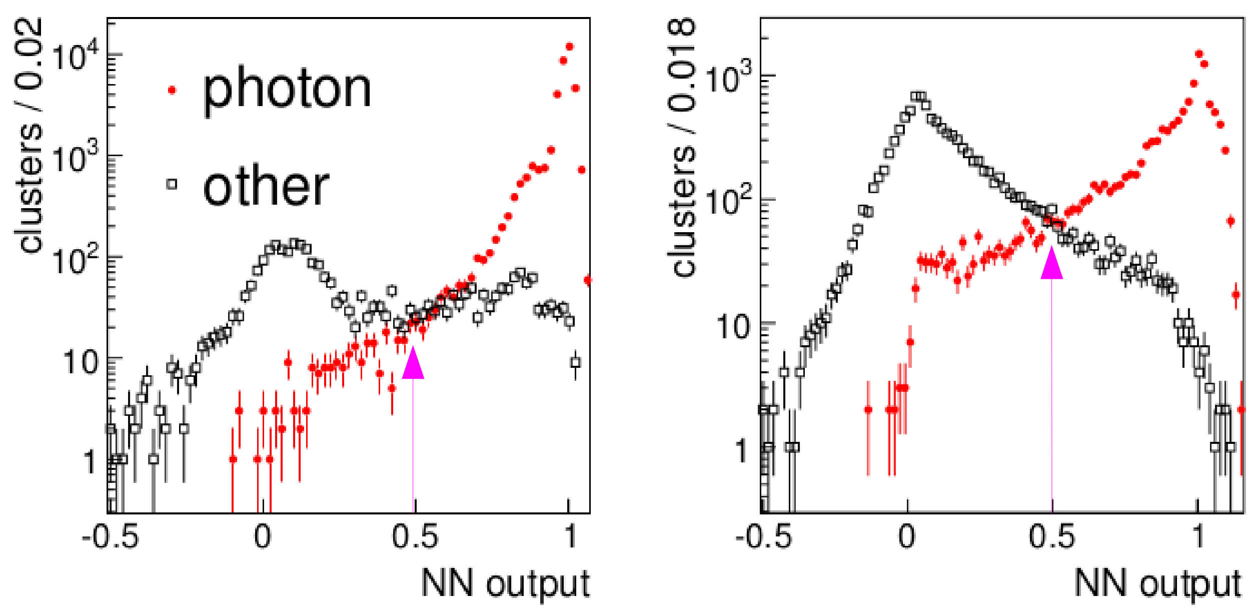
1->3 GeV clusters reconstructed in 500 GeV 4-quark events



Example of NN output 1->3 GeV clusters

Clusters far from track

clusters near to track



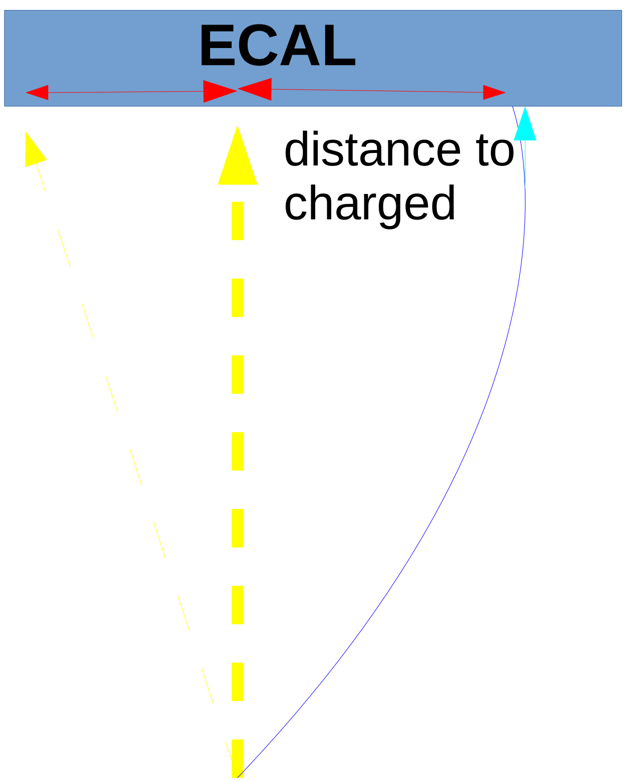
Rather clean separation possible

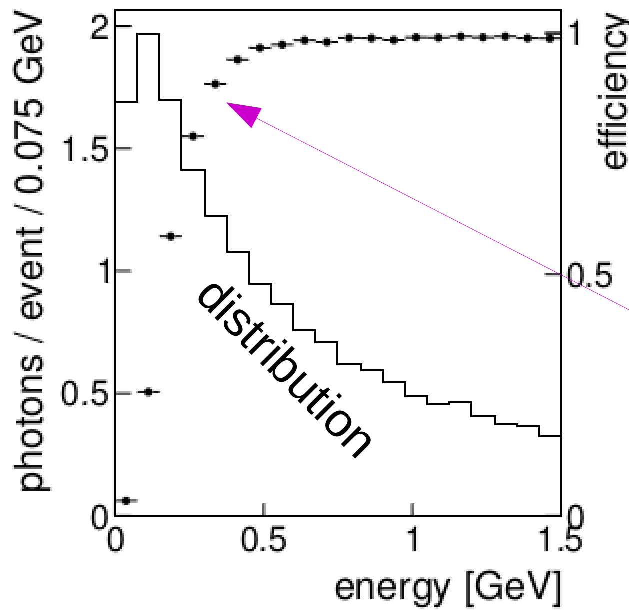
GARLIC Performance

Estimated in jet events: 4-quark events at 500 GeV centre-of-mass energy

GARLIC Performance in 4-quark events at 500 GeV

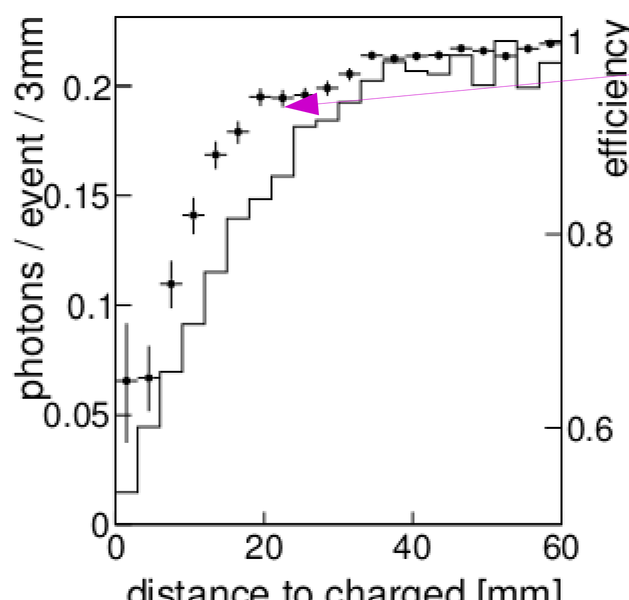
Photon distributions and GARLIC efficiency in 500 GeV e+e- -> 4 light quark events





inefficiencies at:

low energy (<500 MeV)



close to charged track
< 2 cm
disentangle hadronic shower from photon

Otherwise, efficiency to correctly identify photon energy is ~99%

Conclusions

individually measuring every particle produced in particle collisions
will maximise available information for analysis
→ **Particle Flow**

Particularly useful for Jet Energy Measurement
→ significant improvements with respect to past approaches
around a factor 2: 60% → 30% / Sqrt (E)

→ essential to maximally exploit high energy e+ e- collisions
identify and use hadronic W, Z, H decays

several dedicated detector designs and algorithms
are being actively developed

real-world examples being used, particularly at CMS/LHC

we look forward to building and using such detectors at e+ e- colliders !