Data format

by
Vincent Boudry and Henri Videau
LLR, Institut Polytechnique de Paris
25/11/2024

This is a deliverable in the ANRs

Purpose

Define a format for data extracted from the simulation such that all the needed information to perform particle flow development and analysis is recorded:

- data simulating real data
- information needed to validate the methods and analyses developed.

For the time being this involves

- Monte Carlo track information,
- data from the calorimetric system and specifically Ecal and Hcal in the form of space-time localised energy deposits by Geant4 (sub-hits) together with the species of particles which have generated the sub-hits and the reference to the hardware cells they belong to,
- calorimeter cells data (hits) with information on the cells hardware and properties provisionally derived from the linked sub-hits.
- Set (Silicon External Tracker) data

More raw tracker information has to be added at some point, (TPC may be not), but also reconstructed tracks.

Notice that the hits full information, specifically energy and time, has to be elaborated from the associated sub-hits information by a digitisation process aware of the hardware specificities.

A key point is to ensure the possibility to visualise these data in the best way to understand them.

Content of the object sets (or collections)

As we have understood the power of time and of time-space structure,
as we intend to make use of this 4D structure
in the process of reconstructing calorimetric objects
we try to exhibit 4-vectors in the format wherever it has a meaning.

The format, currently for these four pieces:

event*.nikp: (nikp stands for kin_particle)

"{index:d} {pdg:d} {parent:d} {status:d} {prim_rank:d} {charge:.1f}",	
"{start_t:.5e} {start_x:.5e} {start_y:.5e} {start_z:.5e}",	track start point 4D vector
"{energy:.5e} {mom_x:.5e} {mom_y:.5e} {mom_z:.5e}",	track energy-momentum 4D vector
"{end_t:.5e} {end_x:.5e} {end_y:.5e} {end_z:.5e}",	track 4D end point

calo*.subhit:

"{primary_pdg:d} {secondary_pdg:d} {energy:.4e} {length:.2e}",	
"{time:.5e} {pos_x:.5e} {pos_y:.5e} {pos_z:.5e}",	4D position
"{hit_id:d},	link to hit it belongs to

Numerical precision for time and space recording

The time is provided in ns, the space coordinates in mm
They are in a 5e format which provides a precision of 10^{-5} of the value, the time being of the order of 10ns its precision is about 30 micrometres,
while the distances being at the level of 3m their precision is about 30 µm, close to the time precision.

calo*.hits:

"{system:d} {stave:d} {module:d} {x:d} {y:d} {tower:d} {layer:d} {wafer:d}",	hardware info
"{energy:.4e} {first_time:.5e} {pos_x:.5e} {pos_y:.5e} {pos_z:.5e}",	energy hit 4D position
"{first_pdg:d}",	link to particle list

set*.hits

"{system:d} {side:d} {layer:d} {module:d} {sensor:d} ",	hardware info
"{EDep:.4e} {time:.5e} {pos_x:.5e} {pos_y:.5e} {pos_z:.5e}",	deposited energy, hit 4D position
"{pdg:d} {mcp:d}",	link to particle list, link to tracks

Value of “system” for the different pieces

Ecal Barrel	20
Ecal Ring	21
Ecal Endcaps	29
Hcal Barrel	22
Hcal Ring	23
Hcal Endcaps	30

Note: calo*.hits contains Ecal and Hcal together in the same format but distinguished by the variable « system » which designates a piece of hardware, see table on the right, more systems can be added like Ecal (or Hcal) ring.

The Monte-Carlo tracks (event*.nikp or NIKP in HV data model)

This set of objects contains strictly the result of the Géant4 simulation for selected tracks in the detector

event*.nikp:	nikp stands for kin_particle
"{index:d} {pdg:d} {parent:d} {status:d} {prim_rank:d} {charge:.1f}",	
"{start_t:.5e} {start_x:.5e} {start_y:.5e} {start_z:.5e}",	track start point 4D vector
"{energy:.5e} {mom_x:.5e} {mom_y:.5e} {mom_z:.5e}",	track energy-momentum 4D vector
"{end_t:.5e} {end_x:.5e} {end_y:.5e} {end_z:.5e}",	track end point 4D vector

Meaning

index	Géant4 track index	
pdg	→ PTCL	contains the particle data group particle code, ban be transformed into a direct relationship to the PDG particle set
parent	→ NIKP	contains the Géant4 mother particle index, auto-relationship
status	INTE	not yet implemented, should contain information about creation and decay of the particle
prim_rank	INTE	? Not yet implemented
Charge	INTE	charge of the particle
Start_P (Start_P.t Start_P.x Start_P.y Start_P.z)		4D vector providing time and position of the start of the track
En_Mo (En_Mo.t En_Mo.x En_Mo.y En_Mo.z)		4D vector providing the energy and momentum of the track
End_P (End_P.t End_P.x End_P.y End_P.z)		4D vector providing time and position of the track end-point the end-point is where the track decays or interacts with the detector

The sub-hits (calo*.subhit or SBHT in HV data model)

This set of objects contains the direct result of the Géant4 simulation of the energy deposits by the tracks in the sensitive media

```
calo*.subhit:
"{primary_pdg:d} {secondary_pdg:d} {energy:.4e} {length:.2e}",
"{time:.5e} {pos_x:.5e} {pos_y:.5e} {pos_z:.5e}",        position 4D vector
"{hit_id:d},                                              link to hit
```

ppdg	→ PTCL primary	contains the particle data group code of the primary particle, transformed into a direct relationship to the PDG particle set
spdg	→ PTCL secondary	contains the particle data group particle code of the secondary (actually depositing energy) particle
energy		deposited energy (in some unit) along the length of the step
length		length of the step along which the energy has been deposited
P (P.t P.x P.y P.z)		4D position (time and space) of the middle of the step along which the energy has been deposited, may be modified
hit_id	→ HITS cell	link to which cell (hit) belongs the sub-hit

Some changes on this information have been proposed (see Ecal_sub-hits) like
using the 4D vector at the start of the step or better using the step start and step end: may be implemented
Beware at the steps crossing cells limits, for this purpose the maximum step length has been set to 250µ instead of a value larger than the thickness of the wafers.
If we detect sub-hits deposited by Géant4 in a suite they could be linked together providing more information, a pis-aller.

Detailed content of the object sets

The hits or cells (calo.hits (or EKQL in HV data model))

This set of objects contains hardware information about the cell including its 4D position
it is linked by a reciprocal relationship to the sub-hits
The energy is proposed as well as a pdg code but those have to be redone from the sub-hits using a proper digitisation process
The set contains all the fired cells of the calorimetric system, for example Ecal (barrel, end-caps, ring) and Hcal, distinguished via the "system" variable

For the hardware info see drawings next slide

calo.hits:

"{system:d} {stave:d} {module:d} {x:d} {y:d} {tower:d} {layer:d} {wafer:d}",	hardware info
"{energy:.4e} {first_time:.5e} {pos_x:.5e} {pos_y:.5e} {pos_z:.5e}",	energy, hit 4D position
"{first_pdg:d} {mcp:d}",	link to particle list, provided by the MC but reevaluated in the digitisation software process
	link to MC tracks NIKP

syst		mechanical piece of the detector, Ecal barrel, Hcal endcap +,… (partial list of codes in the red frame)
stave		stave number within the system
mod		module number within the stave
ix		cell x index within the wafer
jx		cell y index within the wafer
tower		part of a module
layer		sensitive layer number
wafer		what does it mean outside of silicon detectors?
energy		energy recorded in the cell, has to be redefined in the digitisation
P P.t P.x P.y P.z		4D position, centre of the cell but time has to be redefined in the digitisation
pdg	→ PTCL	PDG code of the particle depositing energy, link to the pdg particle set
mcp	→ NIKP	link to the particle list

Value of syst for the different pieces

Ecal Barrel	20
Ecal Ring	21
Ecal Endcaps	29
Hcal Barrel	22
Hcal Ring	23
Hcal Endcaps	30

Three of these pieces of information have to be properly defined: energy, cell time and pdg code
energy and cell time are elaborated in the digitisation process to mimic the hardware, important in the 4D approach
The pdg code has no strong interest except for figuring out the shower structure

The hits or strips of the Silicon External Tracker (SET), NSET in the HV data model
The SET is made, to my knowledge, of two layers of silicon strips (along the detector axis), draw it but the Monte Carlo provides points and this does not correspond to any hardware implementation
It stays between the external wall of the TPC and the Ecal barrel, see Page 10.

tracker*.hits

"{system:d} {side:d} {layer:d} {module:d} {sensor:d} ",	hardware info, we should try to know more but the points are good enough fo the purpose of checking the calorimetre pattern
"{EDep:.4e} {time:.5e} {pos_x:.5e} {pos_y:.5e} {pos_z:.5e}",	deposited energy in the strip, 4D position
"{pdg:d} {mcp:d}",	pdg of the depositing particle, link to the NIKP track

syst		mechanical piece of the detector, Ecal barrel, Hcal endcap +,...
side		stave number within the system
layer		sensitive layer #, 1 or 2 or 0, 1 ?
mod		module number
sensor		silicon sensor #
energy		energy recorded in the cell (strip)
P P.t P.x P.y P.z		4D position
pdg	→ PTCL	PDG code of the particle depositing energy, link to the pdg particle set
Mcp	→ NIKP	link to track list

SET dimensions
SET_th 20.
SET_Tpc_clear 10.
SET_Ec_clear 5.
SET_iR Ec_BA_iR SET_Ec_clear – SET_th –
SET_l Ec_BA_Z

Example from 50 GeV tau$^-$

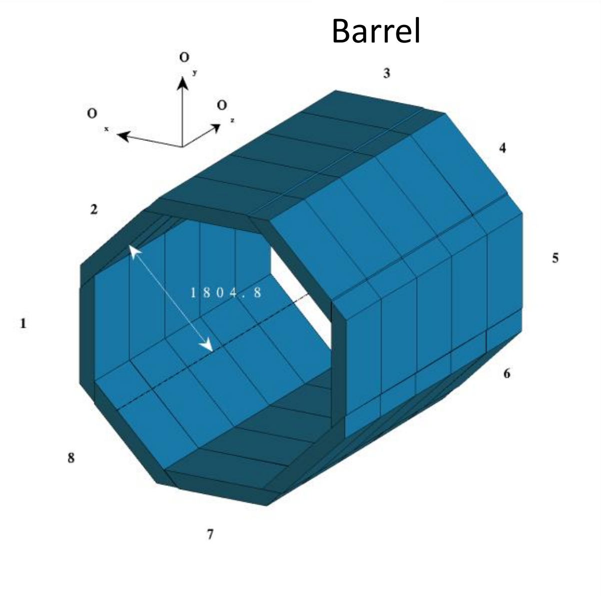
SET event 00

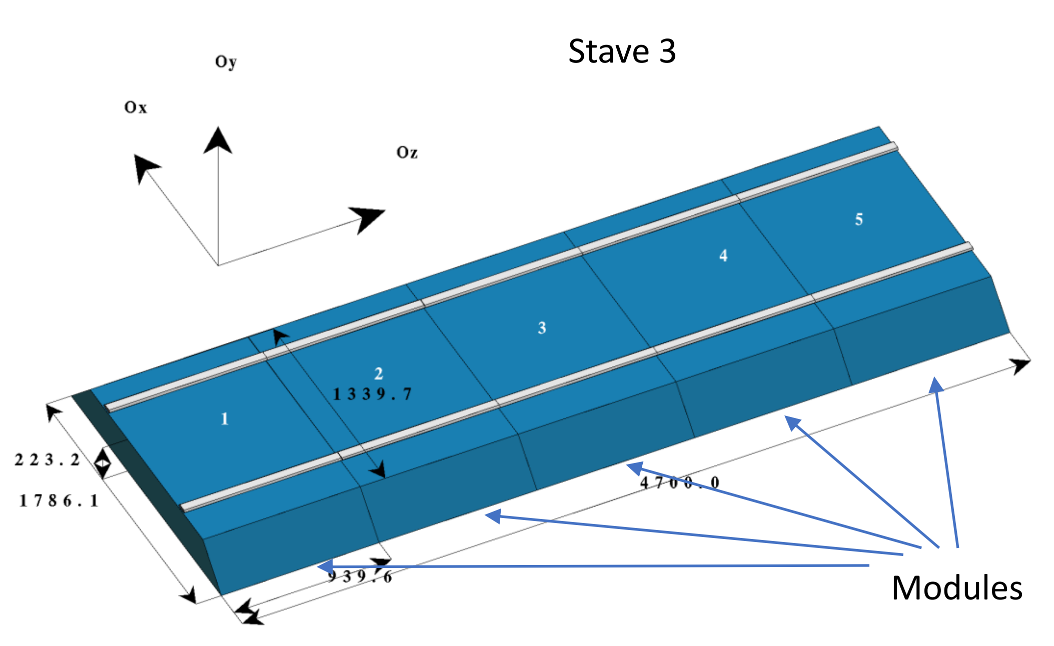
Syst	side	layer	mod	sensor	energy	P				pdg	mcp
5	0	1	12	33	4.1942e-05	1.70858e+01	-1.77520e+03	-2.25938e+01	7.81824e+02	22	4
5	0	1	12	33	5.0216e-05	1.70861e+01	-1.77520e+03	-2.26065e+01	7.81840e+02	22	4
5	0	0	8	26	9.8362e-05	6.13017e+00	-7.55129e+02	1.61108e+03	1.63170e+02	-211	1
5	0	1	8	26	5.6777e-05	6.13902e+00	-7.57121e+02	1.61282e+03	1.63335e+02	-211	1

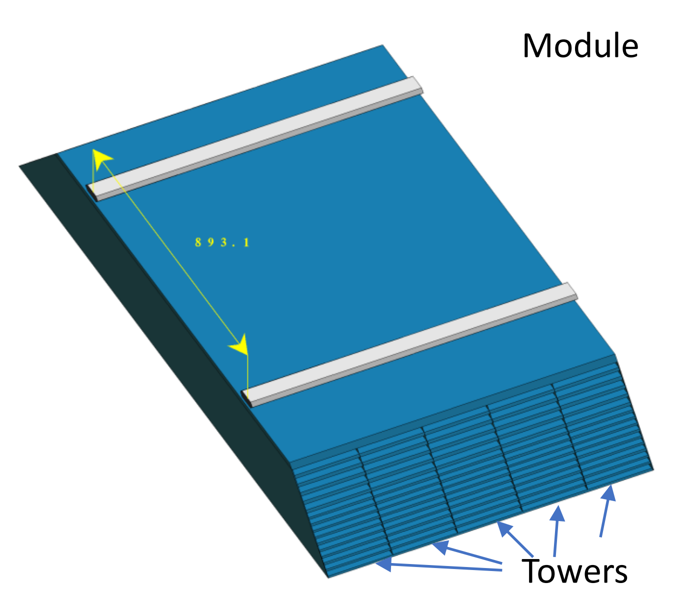
mcp refers properly to NIKP

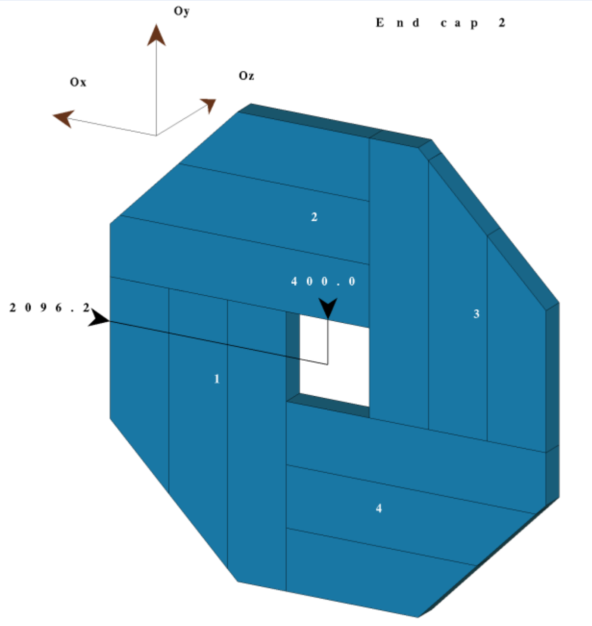
NIKP
0 15 -1 0 0 -1.0 0.00000e+00 0.00000e+00 0.00000e+00 5.00000e+01 5.00337e+01 4.84369e-01 4.99996e+01 1.50933e-01 2.30795e-04 6.69775e-04 6.91435e-02 5.00002e+01
1 -211 0 0 0 -1.0 2.30795e-04 6.69775e-04 6.91435e-02 5.00002e+01 2.33525e+00 -6.17278e-02 2.32597e+00 1.41289e-01 0.00000e+00 -9.65049e+02 1.76703e+03 1.82664e+02
2 16 0 0 0 0.0 2.30795e-04 6.69775e-04 6.91435e-02 5.00002e+01 2.19982e+01 4.49568e-01 2.19831e+01 -6.79232e-01 0.00000e+00 3.06758e+02 1.50000e+04 -4.13466e+02
3 111 0 0 0 0.0 2.30795e-04 6.69775e-04 6.91435e-02 5.00002e+01 2.57003e+01 9.64569e-02 2.56905e+01 6.88876e-01 2.74485e-04 7.18934e-04 8.22364e-02 5.00006e+01
4 22 3 0 0 0.0 2.74485e-04 7.18934e-04 8.22364e-02 5.00006e+01 2.15602e+01 3.38631e-02 2.15520e+01 5.93648e-01 1.24860e+01 2.84786e+00 1.81212e+03 9.99132e+01
5 22 3 0 0 0.0 2.74485e-04 7.18934e-04 8.22364e-02 5.00006e+01 4.14012e+00 6.25938e-02 4.13855e+00 9.52277e-02 0.00000e+00 2.74375e+01 1.81413e+03 9.17418e+01
6 11 4 0 0 -1.0 1.24860e+01 -1.54773e+03 8.55874e+02 -2.56378e+02 1.73508e-03 -1.37381e-03 -5.82774e-04 -7.22777e-04 0.00000e+00 -1.54865e+03 8.55004e+02 -2.56842e+02
7 11 4 0 0 -1.0 2.00470e+01 -4.00317e+01 -1.69616e+03 2.24007e+03 2.18249e-03 -1.22493e-04 -1.88436e-03 9.67645e-04 0.00000e+00 -3.93051e+01 -1.69821e+03 2.24230e+03
8 -11 4 0 0 1.0 6.63791e+00 -1.20107e+02 1.76559e+03 1.14391e+02 2.14737e-03 -1.85207e-03 -6.98765e-04 6.56996e-04 0.00000e+00 -1.21206e+02 1.76564e+03 1.16035e+02
9 11 4 0 0 -1.0 7.80768e+00 4.24612e+01 1.76903e+03 -4.12081e+01 2.25406e-03 1.83268e-04 -7.85916e-04 -2.04167e-03 0.00000e+00 4.38286e+01 1.76883e+03 -4.33129e+01/

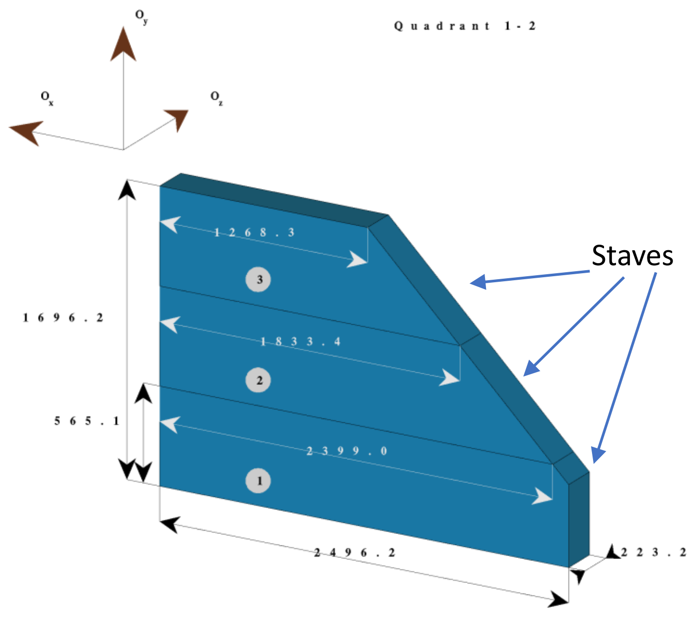
Ecal hardware pieces from the large ILD model

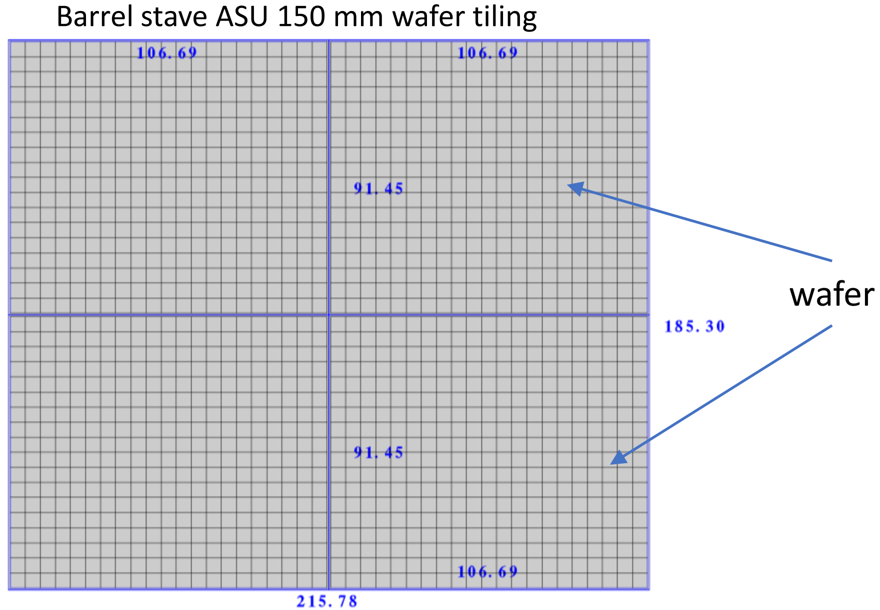












When we try to reconstruct a particle shower from the hits the particle has left behind
we need information linking these hits (or even sub-hits)
to the particle entering the calorimeter system
to understand the performances of our shower reconstruction in terms of
showers identification, separation and energy estimate.

This is provided as a link in the hit set or collection.

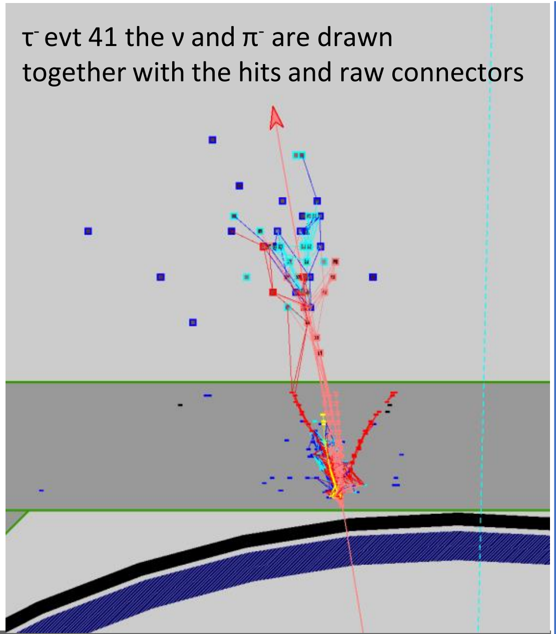
A not so marginal point remains,
the possibility of having a hit linked to more than one shower (particle), in the frequent case of showers overlap.
Then a link of the sub-hit to the particle would be better suited. Does it exist?

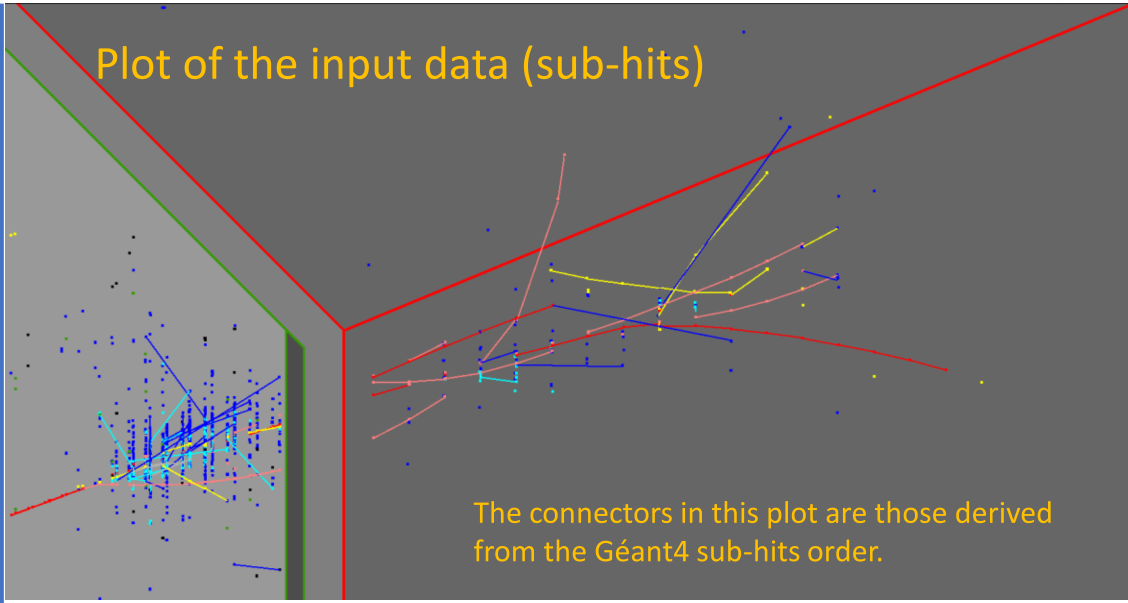
A link between a shower and its start and the corresponding SET points may be constructed.
Then we have also an indirect (through SET) way to check properly the validity of the reconstruction.

Nevertheless it could be interesting to follow the shower development from any sub-hit to the next ones
or rather a link between a sub-hit and its antecedent.

The sub-hit order in the Geant4 collection keeps part of the information.
See E&H page 22 or top right plot of page 10.

This has been implemented and tested on 50 GeV π^+ and τ^- by displaying the raw information and reconstructed calorimetric objects





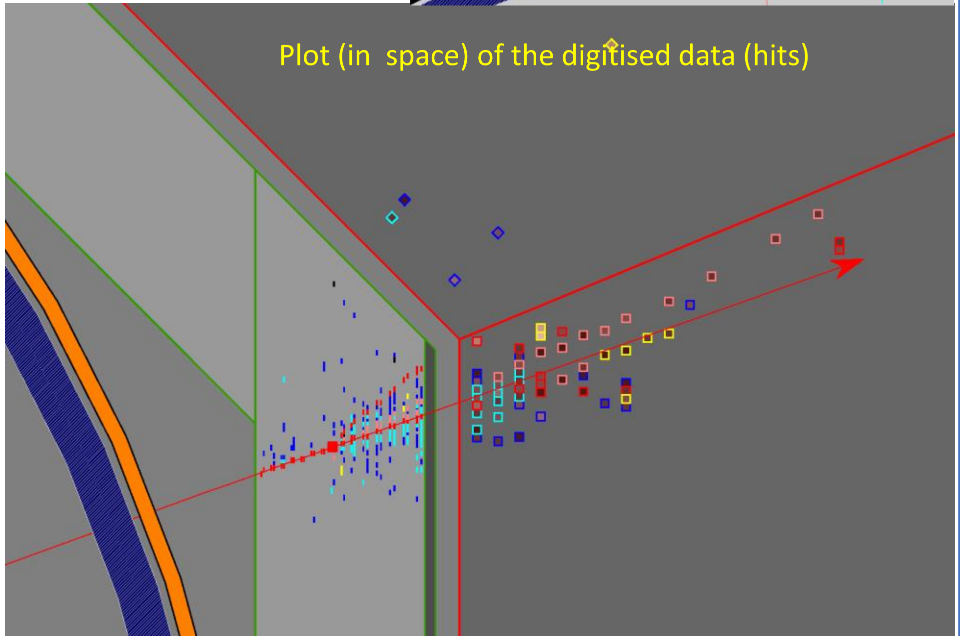
π^+ Evt01_xys (above)

The modules of Ecal and Hcal are drawn
The sub-hits are drawn and coloured according to their pdg.
On top of them are drawn the segments between sub-hits adjacent in Géant4 and verifying causality.
That way we observe in the Hcal some tracks (mostly hadrons) from the shower. In the gas of the Hcal there is very little fog.

π^+ Evt03_xyz (below)

To understand this plot
refer to the « Viewing_event » presentation

The modules of Ecal and Hcal are drawn as well as the outer wall of the TPC and the SET
The hits are drawn with their shape but the size is linked to the energy as well as the area colour when the borders are coloured according to their pdg.



End

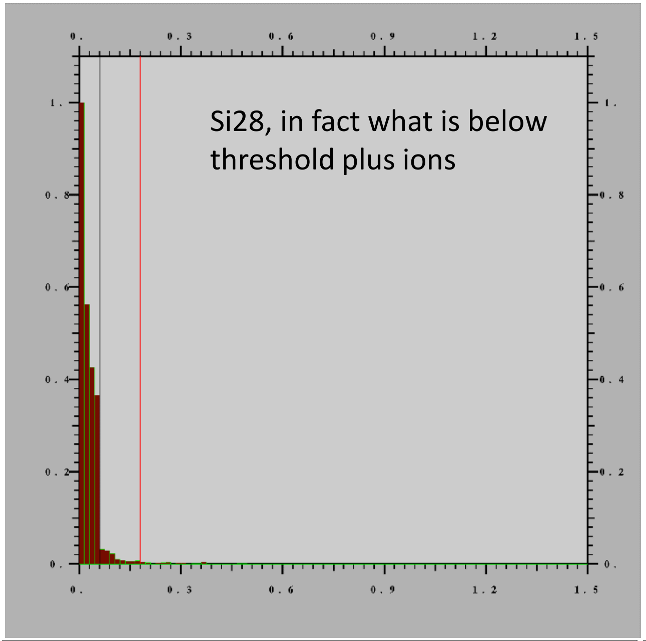
Henri Videau

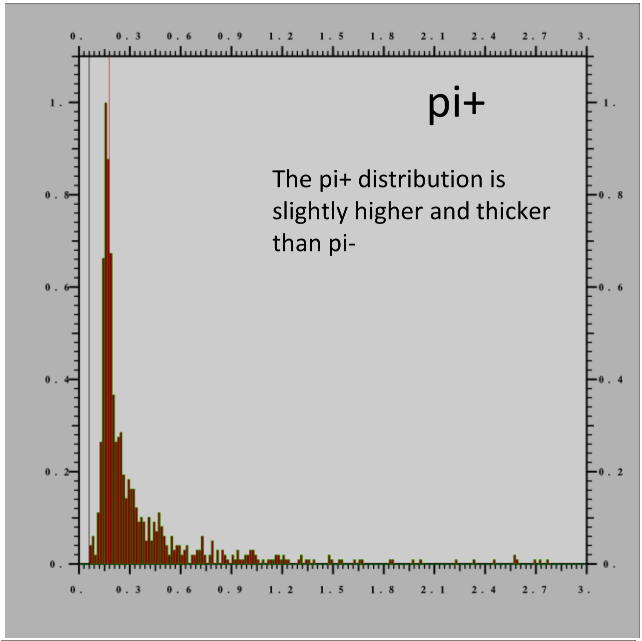
v5calo_250um_90x90_2cmx2cm+5cm_tau_50 — Ecal deposited energy spectra

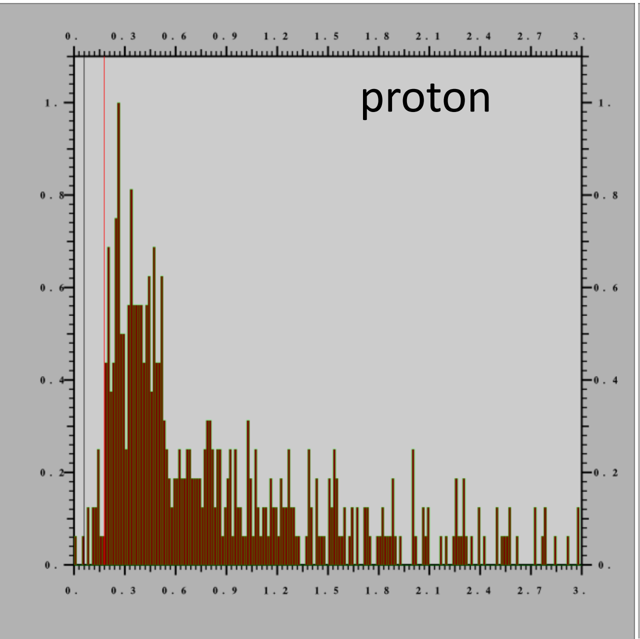
We study the energy distributions for hits deposited by
50 GeV tau showers in Ecal and Hcal, in the new format
A first point is to calibrate the hcal hit energies not in order
to get a correct shower energy but simply to be able
to compare the ecal and hcal distributions.
We get the hcal minion level from the pion energy distributions

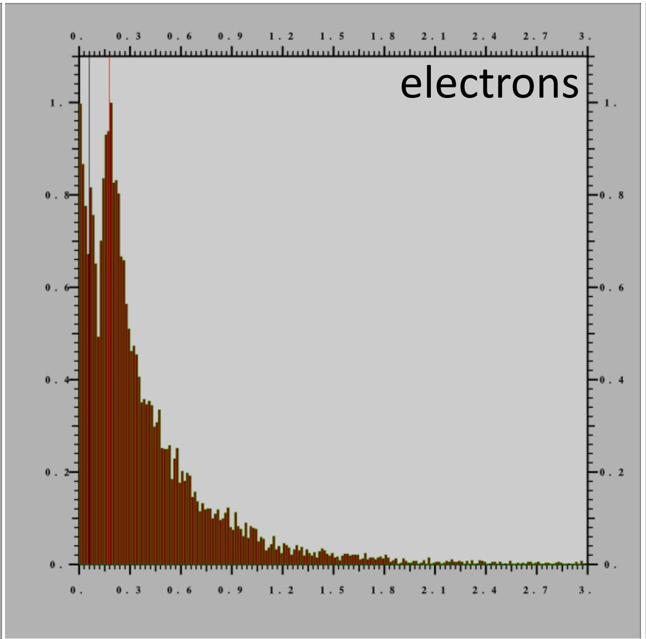
Population per species in Ecal	
Kaon -	40
Proton	543
Kaon+	26
Mu+	8
E+	7337
Pi+	809
E-	11776
Mu-	172
pi-	1449
other	3250

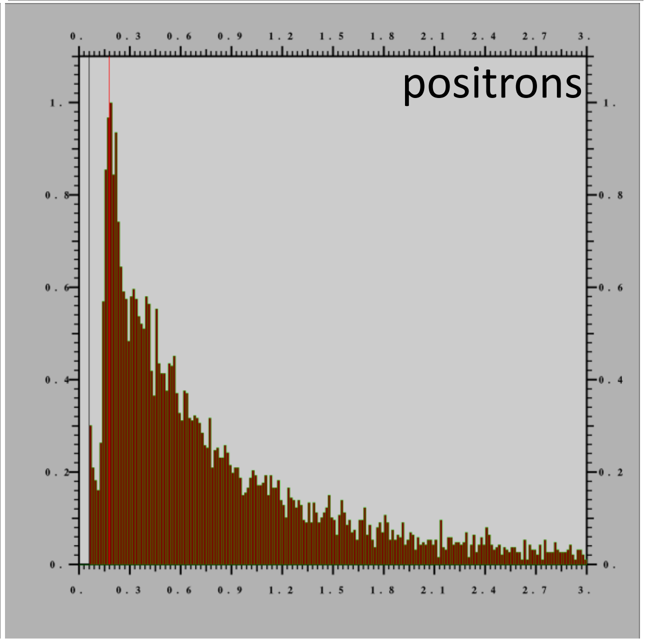
Number of events 50, number of cells 30534 including
25410 in the ecal barrel and 4973 in the hcal barrel
the remaining ones 151 are mostly piece 29, ecal end-caps,
and a number 21, ecal ring.
We study first ecal barrel then hcal barrel
plotting the energy distribution for protons, e+, pi+, e-, pi-, other.

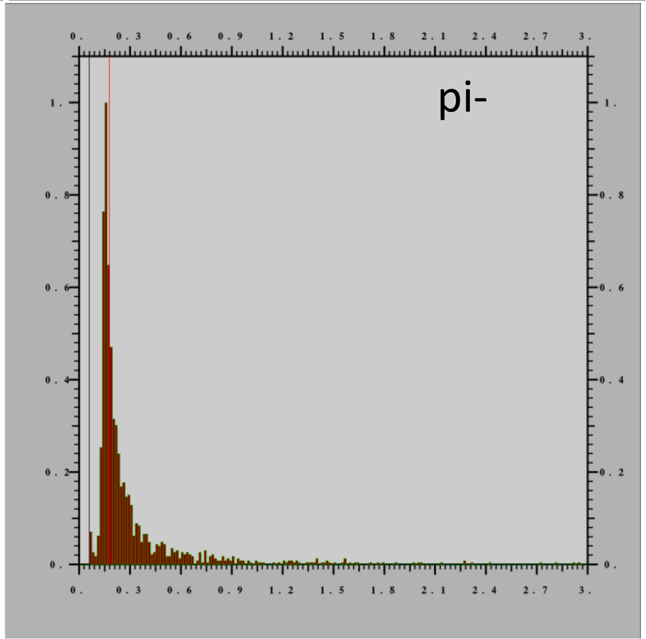




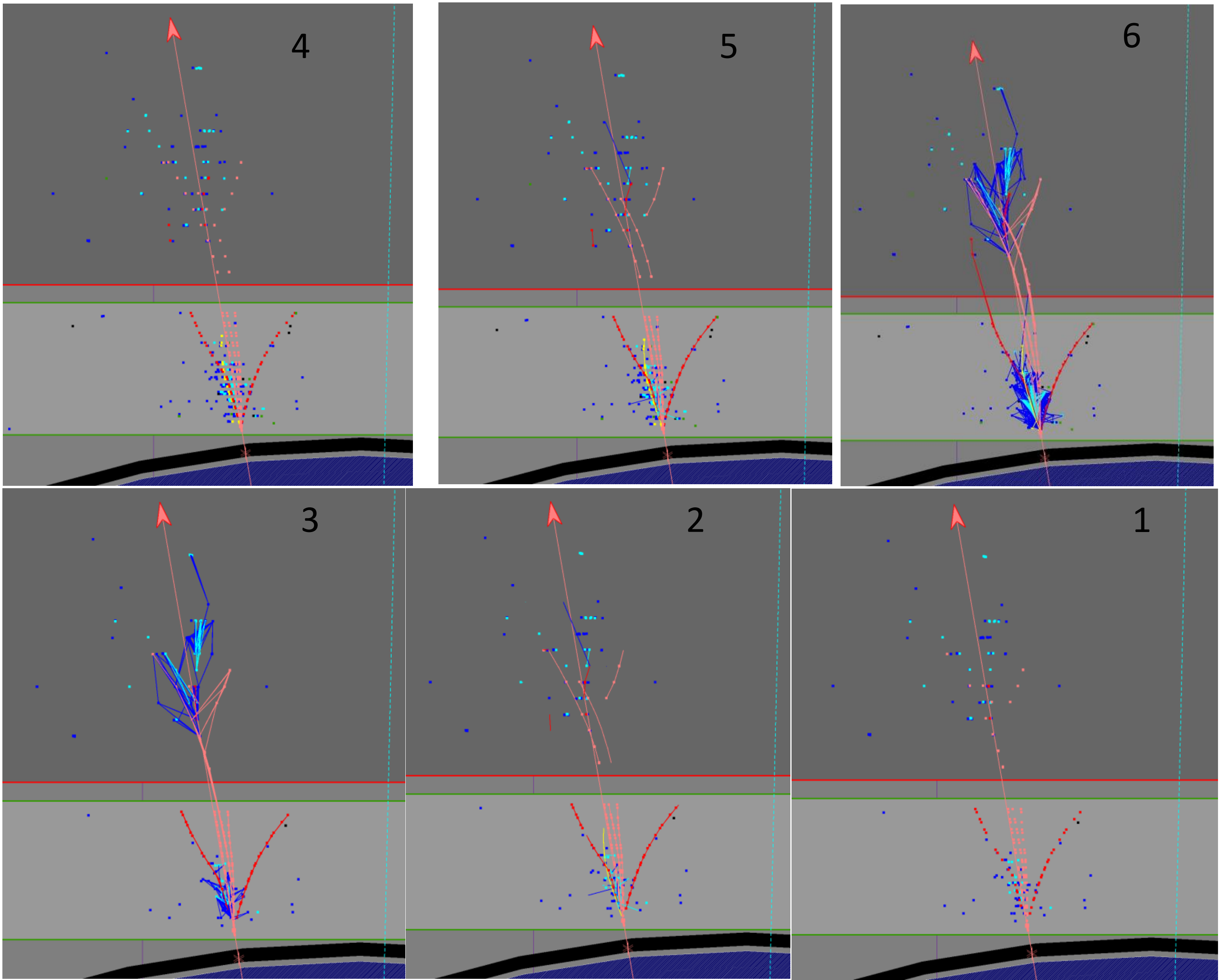








Etude de la coupure autour du minion sur l'événement v5calo_250um_90x90_2cmx2cm_tau-_50/ evt41



All	e^-	e^+	p	n	π^+	π^-	Si
SBHT # after cleaning (ecal)							
733	413	115	21	9	61	104	10
After cell energy between 0.6 minion and 1.5 minion							
227	58	17	0	0	36	63	1
Reduction factors							
0.31	0.14	0.15	0	0	0.59	0.61	0.1

After imposing the cell energy deposit to be a minion electrons and positrons are deeply reduced, all the p and n are gone, almost all ions, only 40% of pions.

Faire avec EKQL au lieu de SBHT

In the upper row left we draw string-cleaned ecal and hcal sub-hits (4), in the centre the SHCL géant4 connectors which exhibit well the hadronic tracks especially a proton (5). On the right (6) we draw the arbor4D connectors some of which cross the pieces.

Then we take hits in minion energy range: 0. – 0. for the ecal and 0.06-0.36 for the hcal (gas)
We consider their sub-hits, and do a string cleaning.
We plot the cleaned sub-hits (1)
We build géant4 connectors SHCL from all SBHT
and draw them on top of 1 (2),
then we build from the cleaned sub-hits arbor4D connectors SCCL
and draw them on top of 1 (3).

Sub-hits, the ultimate
Deposited energy,an essential tool of the pattern