Extracting data sets from the

TrackML Particle Tracking Challenge

High Energy Physics particle tracking in CERN detectors

`https://www.kaggle.com/c/trackml-particle-identification/data`

"A data set comprises multiple independent events, where each event contains simulated measurements (essentially 3D points) of particles generated in a collision between proton bunches at the Large Hadron Collider at CERN.

The goal of the tracking machine learning challenge is to group the recorded measurements or hits for each event into tracks, sets of hits that belong to the same initial particle.

A solution must uniquely associate each hit to one track."

Downloaded 100 events (~2.6GB):

train_100_events/event000001000 … event000001099	hits.csv
	truth.csv
	particles.csv
	cells.csv

File event000001000-hits.csv

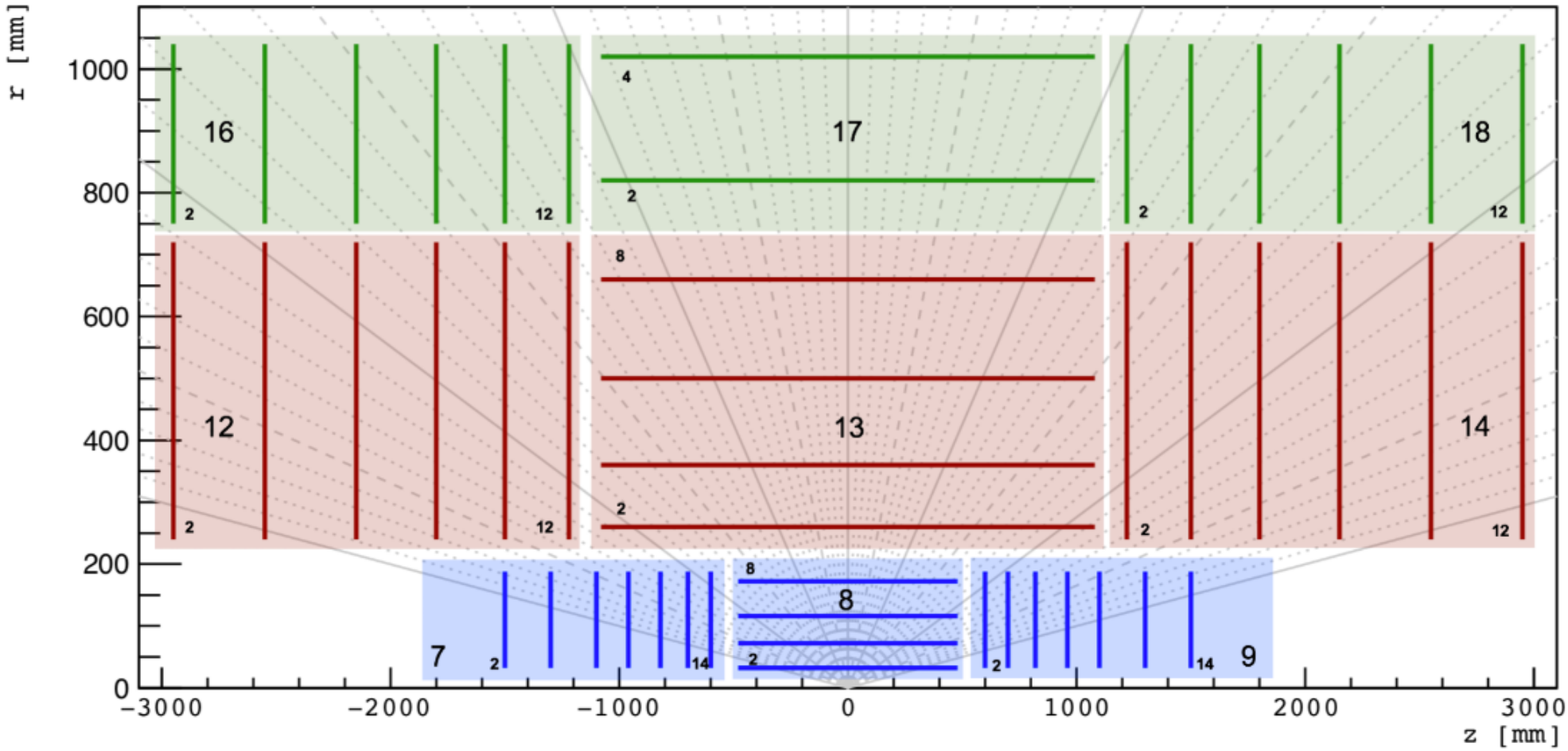
Event hits

The hits file contains the following values for each hit/entry:

- **hit_id**: numerical identifier of the hit inside the event.
- **x, y, z**: measured x, y, z position (in millimeter) of the hit in global coordinates.
- **volume_id**: numerical identifier of the detector group.
- **layer_id**: numerical identifier of the detector layer inside the group.
- **module_id**: numerical identifier of the detector module inside the layer.

	A	B	C	D	E	F	G
1	hit_id	x	y	z	volume_id	layer_id	module_id
2	1	-64.4099	-7.1637	-1502.5	7	2	1
3	2	-55.3361	0.635342	-1502.5	7	2	1
4	3	-83.8305	-1.14301	-1502.5	7	2	1
5	4	-96.1091	-8.24103	-1502.5	7	2	1
6	5	-62.6736	-9.3712	-1502.5	7	2	1
7	6	-57.0687	-8.17777	-1502.5	7	2	1
8	7	-73.8723	-2.5789	-1502.5	7	2	1
9	8	-63.8535	-10.8684	-1502.5	7	2	1
10	9	-97.2548	-10.8891	-1502.5	7	2	1
11	10	-90.2929	-3.26937	-1502.5	7	2	1
12	11	-59.183	-0.670508	-1502.5	7	2	1

The detector



The volume/layer/module id could in principle be deduced from x, y, z. They are given here to simplify detector-specific data handling.

File event000001000-truth.csv

Event truth

The truth file contains the mapping between hits and generating particles and the true particle state at each measured hit. Each entry maps one hit to one particle.

- **hit_id**: numerical identifier of the hit as defined in the hits file.
- **particle_id**: numerical identifier of the generating particle as defined in the particles file. A value of 0 means that the hit did not originate from a reconstructible particle, but e.g. from detector noise.
- **tx, ty, tz** true intersection point in global coordinates (in millimeters) between the particle trajectory and the sensitive surface.
- **tpx, tpy, tpz** true particle momentum (in GeV/c) in the global coordinate system at the intersection point. The corresponding vector is tangent to the particle trajectory at the intersection point.
- **weight** per-hit weight used for the scoring metric; total sum of weights within one event equals to one.

truth

	A	B	C	D	E	F	G	H	I
1	hit_id	particle_id	tx	ty	tz	tpx	tpy	tpz	weight
2	1	0	-64.4116	-7.16412	-1502.5	250710	-149908	-956385	0
3	2	22525763437723648	-55.3385	0.630805	-1502.5	-0.570605	0.0283904	-15.4922	9.86408E-06
4	3	0	-83.828	-1.14558	-1502.5	626295	-169767	-760877	0
5	4	297237712845406208	-96.1229	-8.23036	-1502.5	-0.225235	-0.0509684	-3.70232	8.13311E-06
6	5	418835796137607168	-62.6594	-9.37504	-1502.5	-0.281806	-0.023487	-6.57318	8.87338E-06
7	6	108087696726949888	-57.0856	-8.18971	-1502.5	-0.401129	-0.035276	-10.4669	8.13311E-06
8	7	968286151951515648	-73.8608	-2.57586	-1502.5	-0.442662	-0.0369686	-9.1301	6.96788E-06
9	8	954766419537428480	-63.8512	-10.8754	-1502.5	-0.670459	-0.0926088	-15.5407	1.13644E-05
10	9	707072769359085568	-97.2489	-10.9067	-1502.5	-0.279789	-0.0621434	-4.41292	8.13311E-06
11	10	67554956483231744	-90.2763	-3.24397	-1502.5	-0.251752	-0.0371381	-4.24922	6.45466E-06
12	11	274720539342274560	-59.1974	-0.647788	-1502.5	-0.208248	-0.018635	-5.283	8.13311E-06
13	12	0	-42.6177	-10.6661	-1502.5	-270443	83056.7	-959147	0
14	13	63061596057894912	-72.5146	2.62395	-1502.5	-0.264801	0.0295635	-5.52233	6.96788E-06

	A	B	C	D	E	F	G
1	hit_id	x	y	z	volume_id	layer_id	module_id
2	1	-64.4099	-7.1637	-1502.5	7	2	1
3	2	-55.3361	0.635342	-1502.5	7	2	1
4	3	-83.8305	-1.14301	-1502.5	7	2	1
5	4	-96.1091	-8.24103	-1502.5	7	2	1
6	5	-62.6736	-9.3712	-1502.5	7	2	1
7	6	-57.0687	-8.17777	-1502.5	7	2	1
8	7	-73.8723	-2.5789	-1502.5	7	2	1
9	8	-63.8535	-10.8684	-1502.5	7	2	1
10	9	-97.2548	-10.8891	-1502.5	7	2	1
11	10	-90.2929	-3.26937	-1502.5	7	2	1
12	11	-59.183	-0.670508	-1502.5	7	2	1

hits

Group of 10 true-hits from one particle

	A	B	C	D	E	F	G	H	I
1	hit_id	particle_id	tx	ty	tz	tpx	tpy	tpz	weight
2	2	22525763437723648	-55.3385	0.630805	-1502.5	-0.570605	0.0283904	-15.4922	9.86408E-06
3	1420	22525763437723648	-55.1544	0.621779	-1497.5	-0.569794	0.0272644	-15.4928	8.40273E-06
4	1460	22525763437723648	-47.9881	0.315591	-1302.5	-0.568883	0.0219942	-15.4929	5.66271E-06
5	3170	22525763437723648	-47.8041	0.308733	-1297.5	-0.571445	0.0208471	-15.4933	4.93204E-06
6	3234	22525763437723648	-40.6261	0.0873258	-1102.5	-0.56994	0.015274	-15.4934	4.93204E-06
7	5343	22525763437723648	-40.4422	0.0825377	-1097.5	-0.570204	0.0138721	-15.494	6.39338E-06
8	5424	22525763437723648	-35.4759	-0.0242841	-962.5	-0.569773	0.0106676	-15.494	8.76807E-06
9	5482	22525763437723648	-35.3101	-0.02716	-958	-0.571644	0.00888053	-15.4949	1.05948E-05
10	7867	22525763437723648	-35.2917	-0.027451	-957.5	-0.571617	0.00901224	-15.495	1.35174E-05
11	7951	22525763437723648	-30.3154	-0.0954825	-822.5	-0.570993	0.0062305	-15.4951	1.64401E-05

Event particles

The particles files contains the following values for each particle/entry:

- **particle_id**: numerical identifier of the particle inside the event.
- **vx, vy, vz**: initial position or vertex (in millimeters) in global coordinates.
- **px, py, pz**: initial momentum (in GeV/c) along each global axis.
- **q**: particle charge (as multiple of the absolute electron charge).
- **nhits**: number of hits generated by this particle.

File event000001000-particles.csv

	A	B	C	D	E	F	G	H	I
1	particle_id	vx	vy	vz	px	py	pz	q	nhits
2	103583272465858560	-0.00454767	0.0410508	-1.93889	0.554524	-0.231282	-14.6197	-1	12
3	103583409904812032	-0.00454767	0.0410508	-1.93889	1.037	0.0494352	-20.531	1	0
4	103583409921593345	30.2076	1.21452	-600	0.0710947	-0.200737	-0.0606917	1	6
5	103583409921597442	30.2076	1.21452	-600	0.205443	-0.0424687	-0.0526435	1	7
6	103583478624288768	-0.00454767	0.0410508	-1.93889	0.118078	0.123082	-1.44551	-1	9
7	103583478641070081	60.8909	96.1192	-962	-0.14807	0.0866795	0.213462	-1	4
8	103584028380102656	-0.00454767	0.0410508	-1.93889	0.150491	0.261247	0.414062	1	5
9	103584097099579392	-0.00454767	0.0410508	-1.93889	-0.300089	-0.263511	2.47998	-1	16
10	103584165819056128	-0.00454767	0.0410508	-1.93889	0.202925	0.340909	0.330223	1	14

	A	B	C	D	E	F	G	H	I
1	hit_id	particle_id	tx	ty	tz	tpx	tpy	tpz	weight
2	621	103583272465858560	57.5938	-21.9977	-1498	0.571584	-0.202867	-14.6159	7.62782E-06
3	637	103583272465858560	57.7503	-22.0537	-1502	0.572308	-0.206107	-14.6153	8.67994E-06
4	2227	103583272465858560	49.7895	-19.1871	-1298	0.568906	-0.206205	-14.6165	4.33997E-06
13	9256	103583272465858560	31.3302	-12.3954	-822	0.564237	-0.210088	-14.6189	1.23623E-05
14	71815	103583409921593345	22.1249	-257.835	-675.806	-0.0802422	-0.195258	-0.0611746	2.35432E-05
15	79012	103583409921593345	-41.5046	-359.546	-710.863	-0.138606	-0.156991	-0.0624846	1.08299E-05
16	79017	103583409921593345	-37.0214	-354.399	-708.844	-0.136181	-0.159431	-0.0620038	1.60094E-05
19	91856	103583409921593345	-512.671	-413.857	-875.385	-0.168112	0.116386	-0.068747	8.94642E-06
20	72107	103583409921597442	233.39	-124.251	-661.852	0.125984	-0.164272	-0.0553485	1.51401E-05
21	72113	103583409921597442	229.791	-119.644	-660.293	0.1289	-0.162165	-0.0552026	2.06456E-05
22	79060	103583409921597442	283.979	-216.166	-690.113	0.0703355	-0.193705	-0.0544624	1.1011E-05
25	91901	103583409921597442	109.483	-650.657	-829.66	-0.1833	-0.0872527	-0.0560911	6.19367E-06
26	7042	103583478624288768	60.8909	96.1192	-962	0.0650993	0.159257	-1.436	9.85932E-06
27	7064	103583478624288768	60.7112	95.6779	-958	0.063919	0.157706	-1.43709	8.2161E-06
28	9811	103583478624288768	54.2754	80.8614	-822	0.0720565	0.154851	-1.43707	4.92966E-06

truth

	A	B	C	D	E	F	G	H	I
1	particle_id	vx	vy	vz	px	py	pz	q	nhits
2	103583272465858560	-0.00454767	0.0410508	-1.93889	0.554524	-0.231282	-14.6197	-1	12
3	103583409904812032	-0.00454767	0.0410508	-1.93889	1.037	0.0494352	-20.531	1	0
4	103583409921593345	30.2076	1.21452	-600	0.0710947	-0.200737	-0.0606917	1	6
5	103583409921597442	30.2076	1.21452	-600	0.205443	-0.0424687	-0.0526435	1	7
6	103583478624288768	-0.00454767	0.0410508	-1.93889	0.118078	0.123082	-1.44551	-1	9
7	103583478641070081	60.8909	96.1192	-962	-0.14807	0.0866795	0.213462	-1	4
8	103584028380102656	-0.00454767	0.0410508	-1.93889	0.150491	0.261247	0.414062	1	5
9	103584097099579392	-0.00454767	0.0410508	-1.93889	-0.300089	-0.263511	2.47998	-1	16
10	103584165819056128	-0.00454767	0.0410508	-1.93889	0.202925	0.340909	0.330223	1	14

particles

On the UCA drive: QC/ISIMA/Projet/TrackML/event000001000/

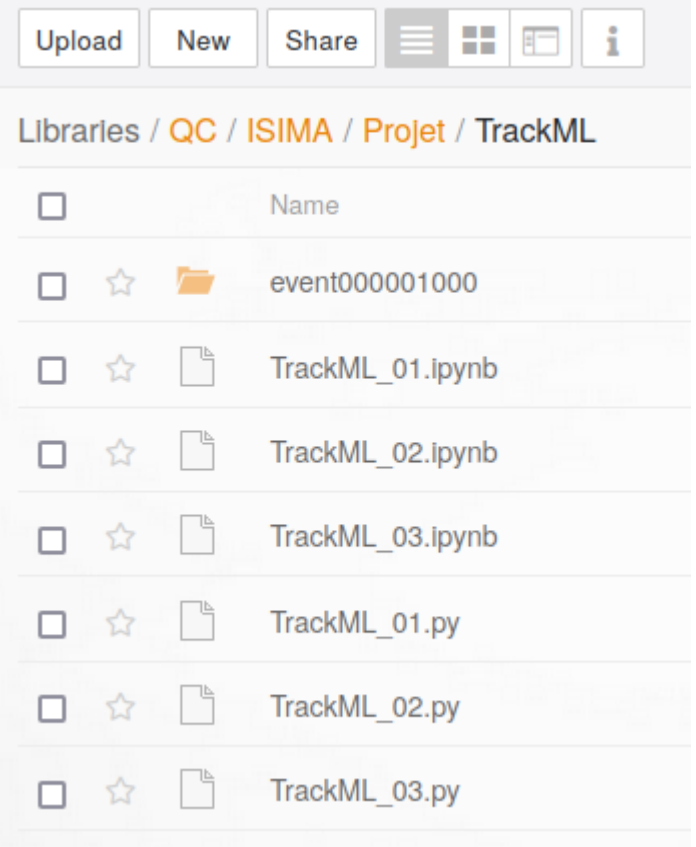
Original files:

event000001000-hits.csv

event000001000-truth.csv

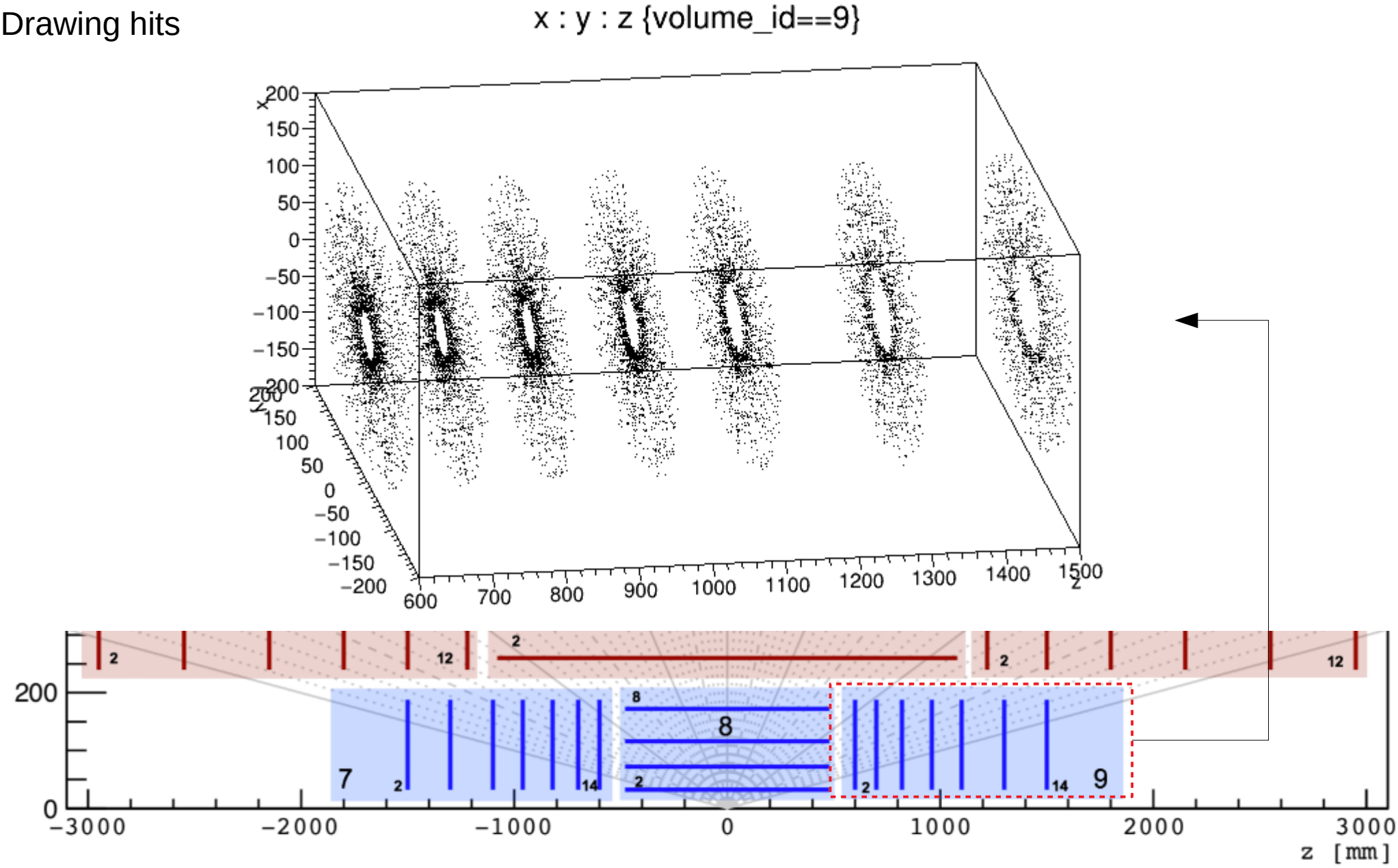
event000001000-particles.csv

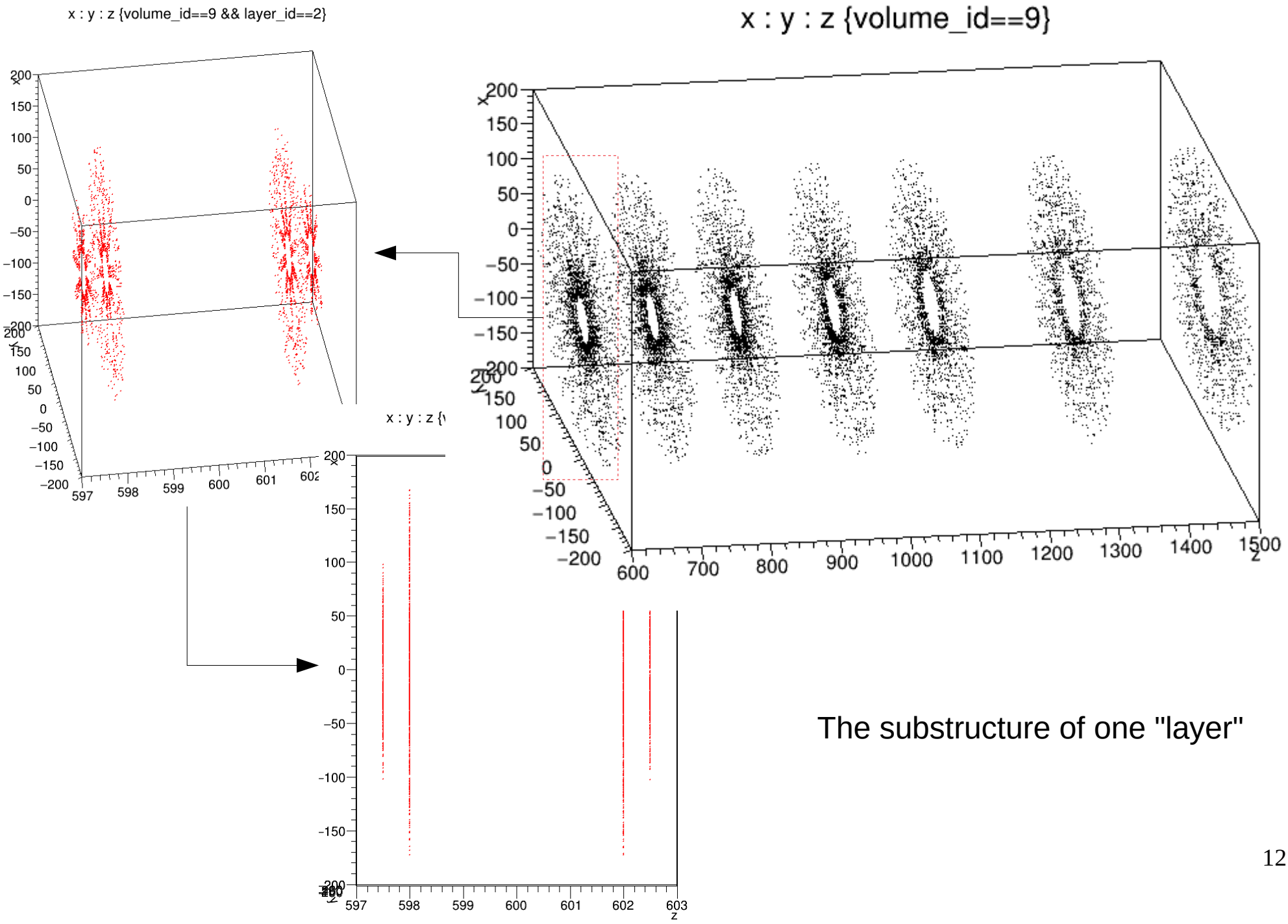
event000001000-cells.csv



hit_id = 0 ⇒ means that the "hit" is not coming from a particle; it is "detector noise"

Drawing hits




x : y : z {volume_id==9 && layer_id==2}
x : y : z {volume_id==9}
x : y : z {
The substructure of one "layer"