

Label cells to build training datasets

- For some Blobs in imaging step, the blob can be big.
 - Intrinsic ambiguity
 - Provide more information like topology to ML to reduce the ambiguity
- Blobs is defined by strips overlap, contains several cells.
- For each cell, label the cell is real or not based on simulation.
 - A Cell here may not be strictly a single cell, limitation of the computation, divide each blobs according to requirement: Coarse or fine
- Use labeled cells from Blobs to build training datasets
- Using protodune-vd simulation.
- Compare vd-simulation imaging and true energy deposit from simulation.

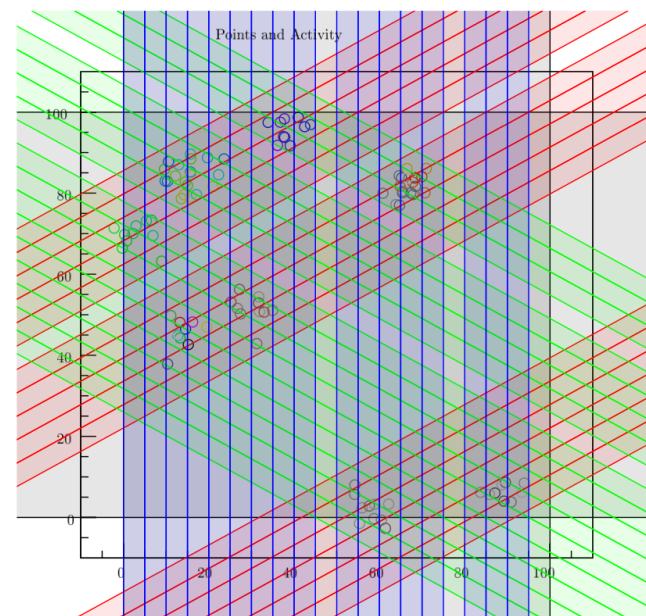
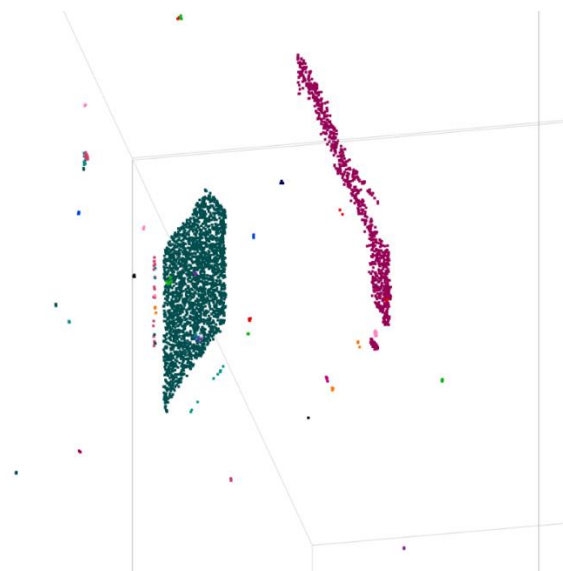
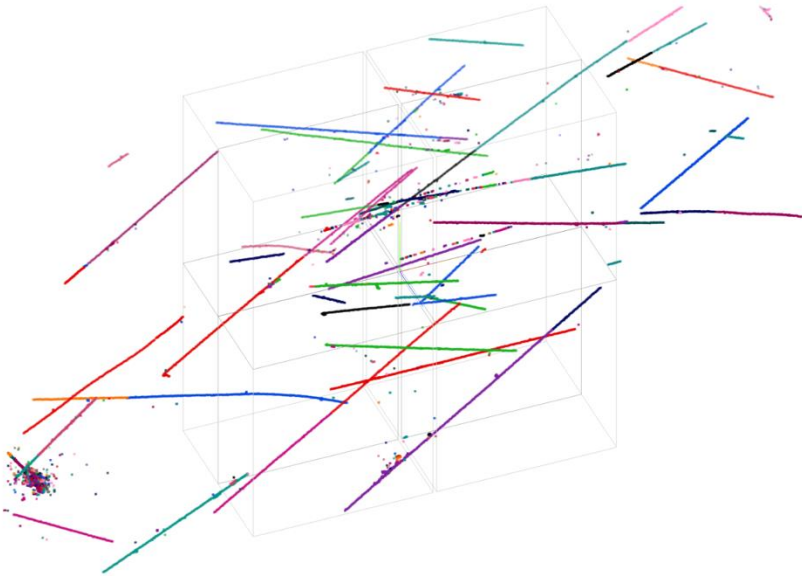


Figure 2: Depiction of the location of generated “electrons” in a toy simulation and the resulting activity arrays which colored based on the ray grid layer the span.

Imaging vs true energy deposit

- vd-simulation imaging:

<https://www.phy.bnl.gov/twister/bee/set/ac7fd044-31f0-4c35-8cc6-645f615c29ba/event/0/>



Next thoughts:

- Make sure these two x-axis is consistent
- In each slice, cut Blob to cells
 - Merge a cell to its neighbor if it is too small
- Label if the cell is true or not

- simulation true energy deposit:
 - `sim::SimEnergyDeposit`
 - `cellTree_module.cc`

<https://www.phy.bnl.gov/twister/bee/set/ac7a6634-e6c4-4325-938c-21a3bafd00ff/event/0/>



Thanks Xin, Wenqiang, Jay, Chao, Hokyeong for all the discussion, information and suggestions