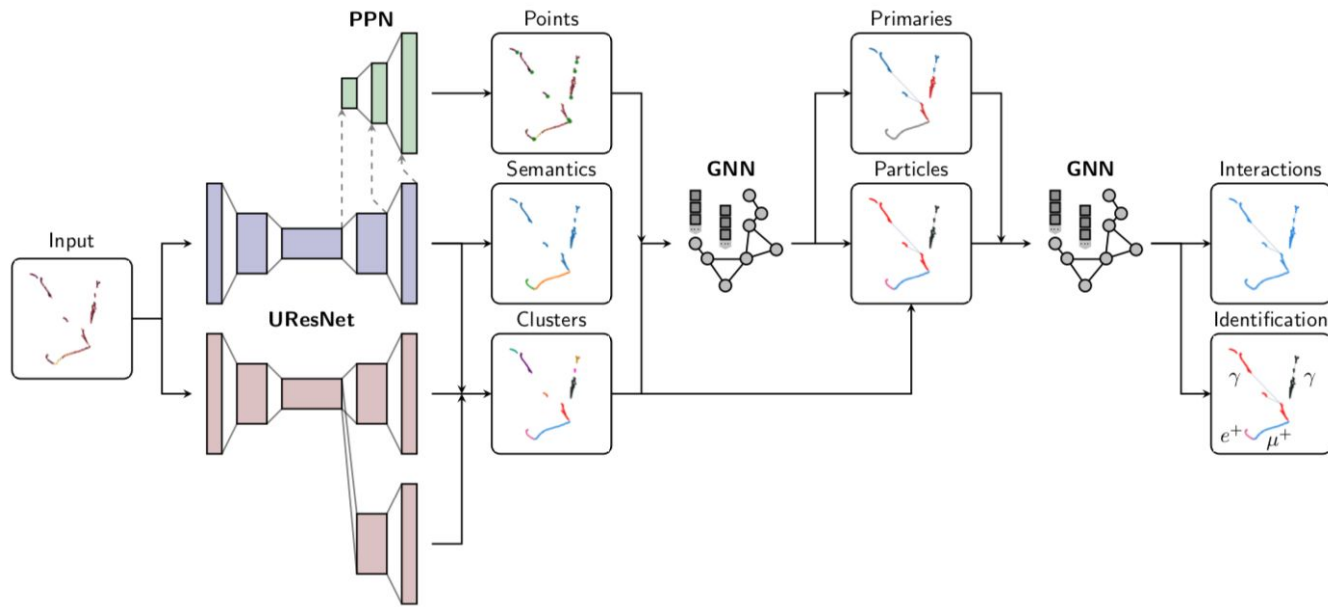


‘New’ ML-based Reconstruction Chain
~~Performance Report~~
Status Update

Kazu T. (SLAC)
on behalf of the team

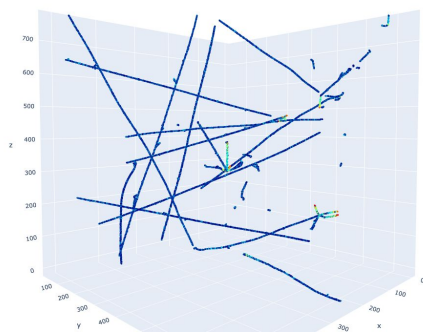
Jeremy W., Andrew M., Zach H., Yifan C., Francois D., Patrick T.



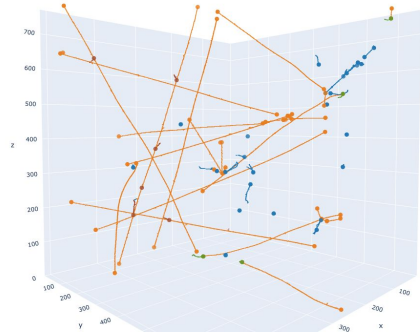
ML-based Reco Chain List of Papers

- [Full chain](#)
- [Particle/Interaction clustering](#)
- [Fragment clustering](#)
- [Particle endpoints](#)
- [Semantic segmentation](#)

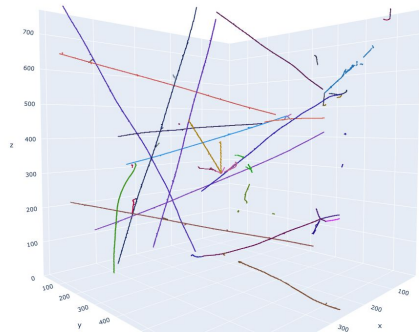
Input



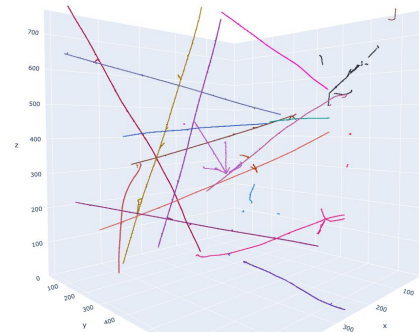
Pixel Feature



Pixel Clustering



Particle Clustering



ML-based Reconstruction: **Current Status**

- **Not ready** to take the detector simulation output.
 - a. Full chain not yet optimized for the full detector simulation output
 - Current “ML analysis” runs an old version that has known issues
 - Must move onto a new version
 - b. Current effort
 - Generating necessary info for optimization
 - Interpreting truth information at the edepsim and larndsim stages
 - Hash out details (Spill/Event ID, T0 across readout groups, etc.)
 - Some of these are done through “ML analysis” effort

ML-based Reconstruction: **What's Needed**

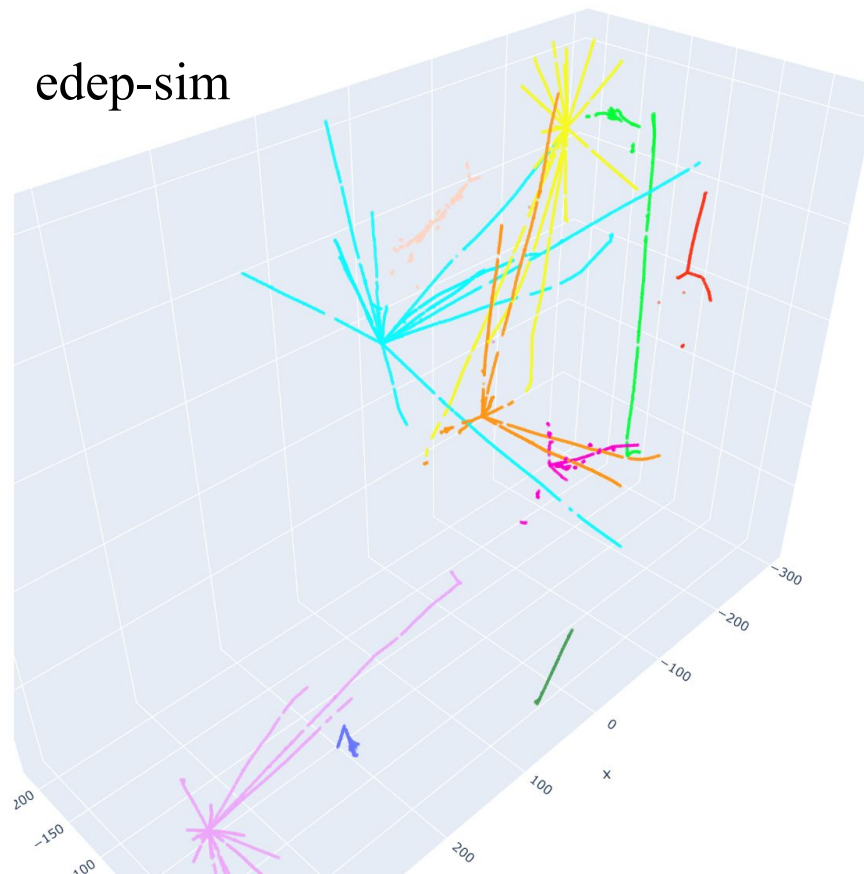
- Custom event generator + interface for edep-sim
- “Supera”: software to create necessary information for tuning ML chain
 - a. Port out a long-running code interfacing LArSoft
 - b. Interface with edep-sim
 - c. Interface with larnd-sim (much of code shared with edep-sim interface)
- Generate $O(1M)$ images each with $O(10-100)$ primary particles
 - a. Software setup to generate them
 - b. Generate
- Optimize the full chain
 - a. Baby-sit the optimization process (will need a few iterations)
 - b. Debug Supera algorithms to meet the needs
- Develop a production quality workflow for large stats
 - a. Location, storage, database, scripts

ML-based Reconstruction: **What's Needed**

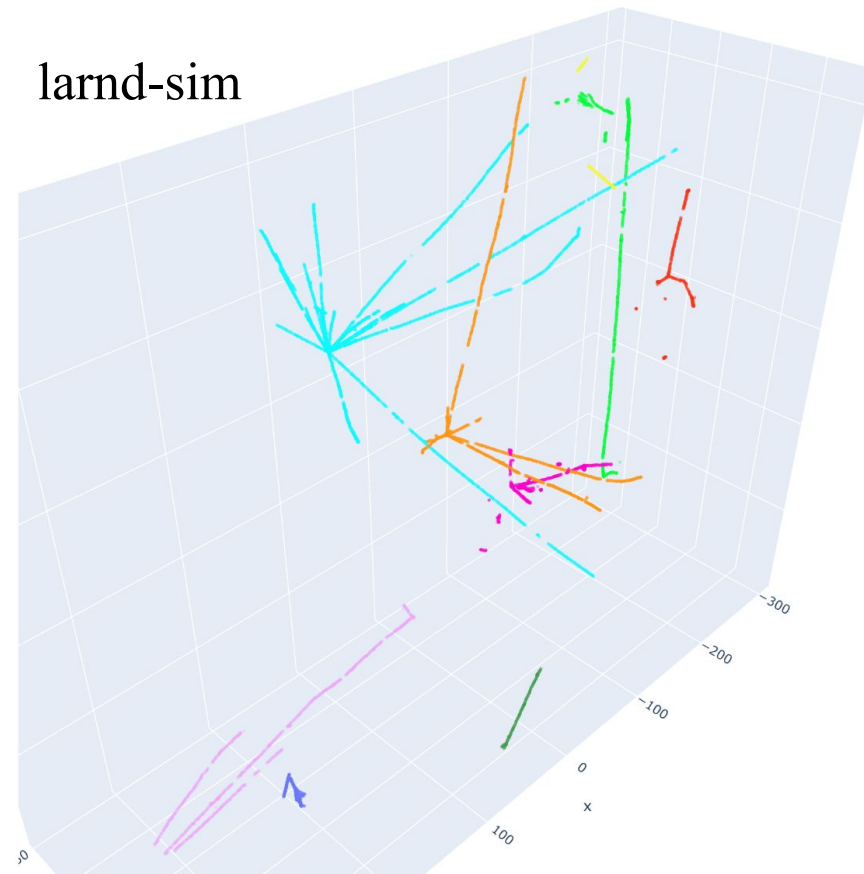
- ~~● Custom event generator + interface for edep-sim~~
- ~~● “Supera”: software to create necessary information for tuning ML chain~~
 - ~~a. Port out a long-running code interfacing LArSoft~~
 - ~~b. Interface with edep-sim~~
 - ~~c. Interface with larnd-sim (much of code shared with edep-sim interface) (almost!)~~
- Generate $O(1M)$ images each with $O(10-100)$ primary particles
 - ~~a. Software setup to generate them~~
 - b. Generate**
- Optimize the full chain
 - a. Baby-sit the optimization process (will need a few iterations)
 - b. Debug Supera algorithms to meet the needs
- Develop a production quality workflow for large stats
 - a. Location(?), storage, database, scripts

Generating “labels” from edep-sim & larnd-sim

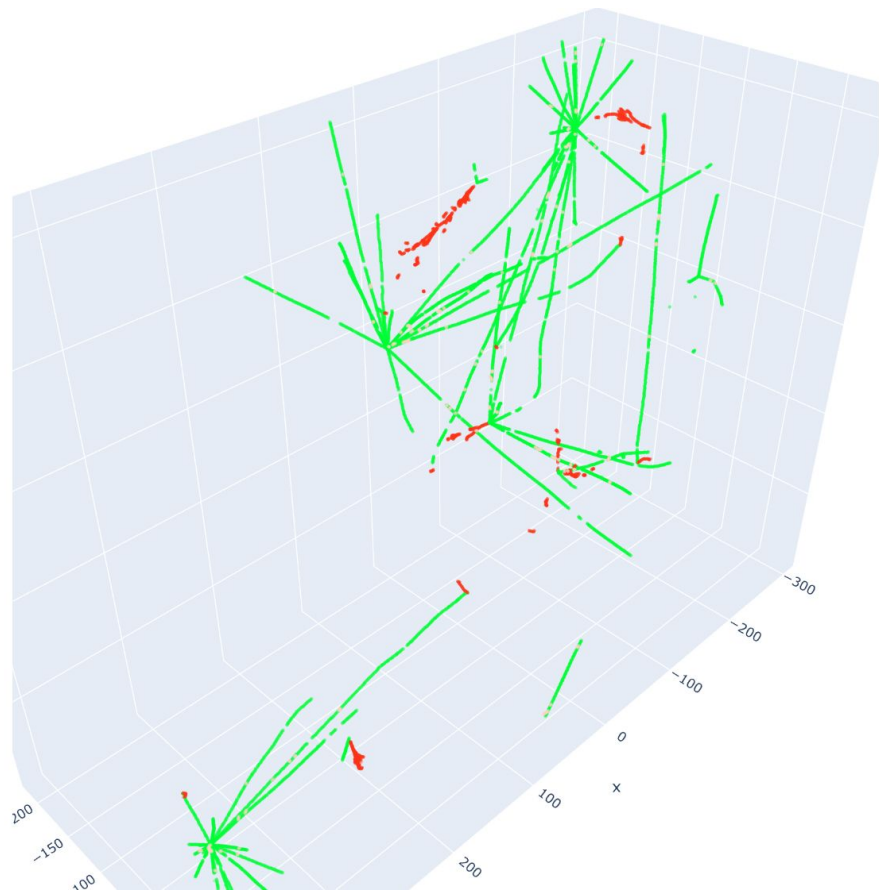
edep-sim



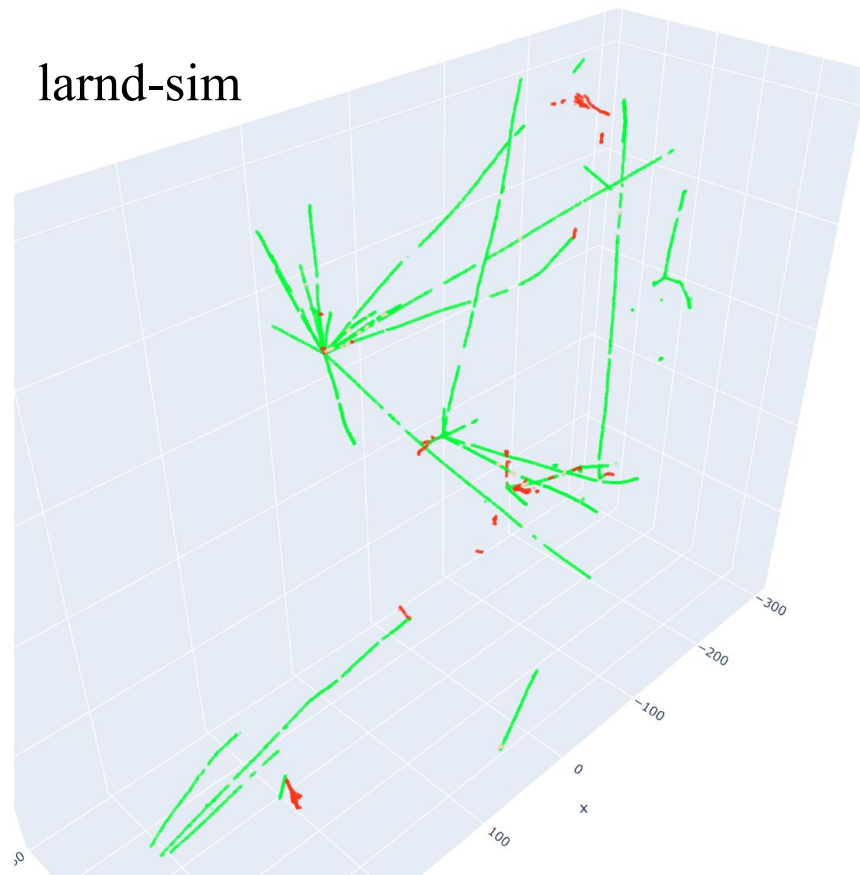
larnd-sim



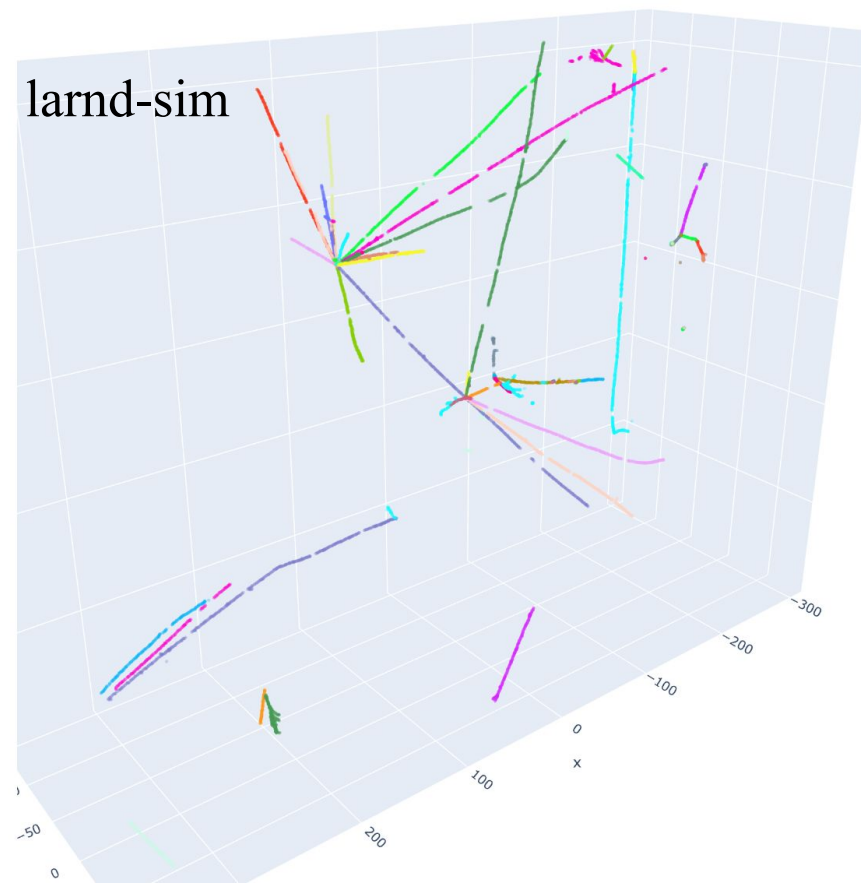
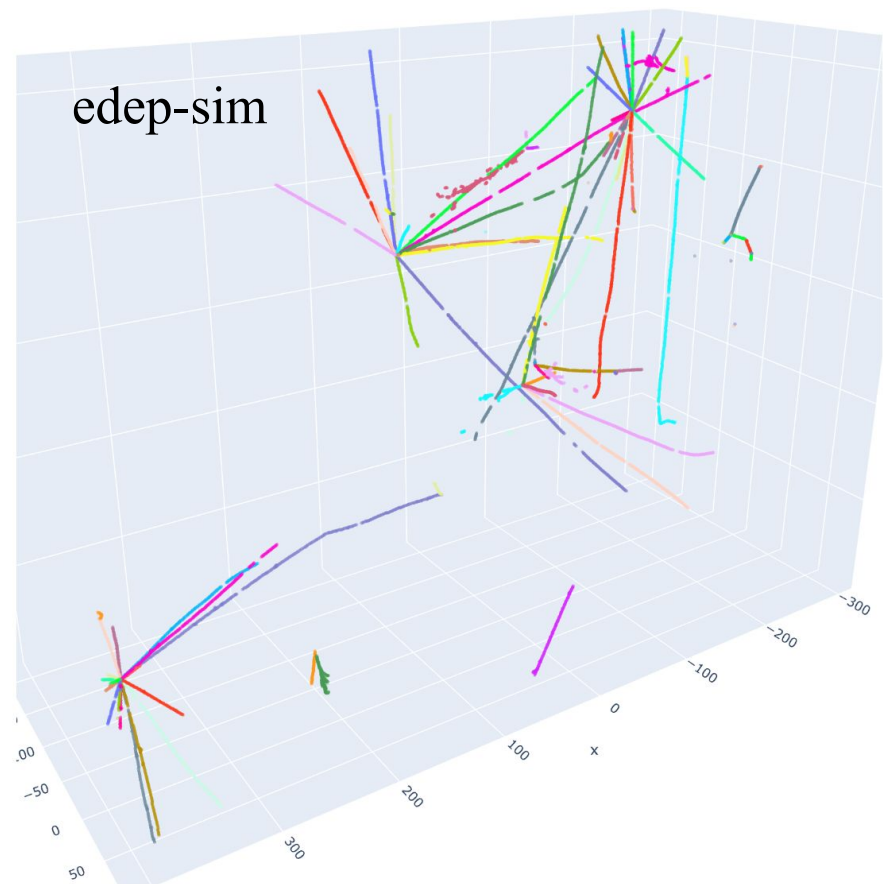
Generating “labels” from edep-sim & larnd-sim



larnd-sim



Generating “labels” from edep-sim & larnd-sim



ML-based Reconstruction: **What's Needed**

- ~~● Custom event generator + interface for edep-sim~~
- ~~● “Supera”: software to create necessary information for tuning ML chain~~
 - ~~a. Port out a long-running code interfacing LArSoft~~
 - ~~b. Interface with edep-sim~~
 - ~~c. Interface with larnd-sim (much of code shared with edep-sim interface) **(almost!)**~~
- Generate $O(1M)$ images each with $O(10-100)$ primary particles
 - ~~a. Software setup to generate them~~
 - b. Generate** ~ 1 week

The output looks qualitatively OK.
Move onto optimization/debug.
- **Optimize the full chain** ~ 1-1.5 months (actual optimization likely 1~2 weeks, ~1 month a conservative estimate for debugging)
 - a. Baby-sit the optimization process (will need a few iterations)
 - b. Debug Supera algorithms to meet the needs
- Develop a production quality workflow for large stats
 - a. Location(?), storage, database, scripts

Time scales similar to what's presented in the last ND LAr analysis meeting. “Work during holidays” just did not happen...

Summary

- Progress made albeit being slow
 - Supera prototype ready to be used (and to be debugged)
 - Shout out for Andrew, Jeremy, Zach, Yifan, Francois, and Patrick for Supera effort
 - Thanks to analyzers hashing out details of how to build/interpret detector simulation output and harnessing to Supera
- But more needed
 - Immediate next steps:
 - Generate $\sim O(1M)$ samples for optimization (~ 1 week)
 - Run optimization of the chain + debugging of Supera (1 \sim 1.5 month)
- What's coming but not figured out
 - Hash out the workflow for production
 - Handle domain shift (data/simulation discrepancies) and impact on reco
 - Short-term: commissioning analysis to improve simulation
 - Long-term: differentiable simulator + inverse imaging as a direct solution