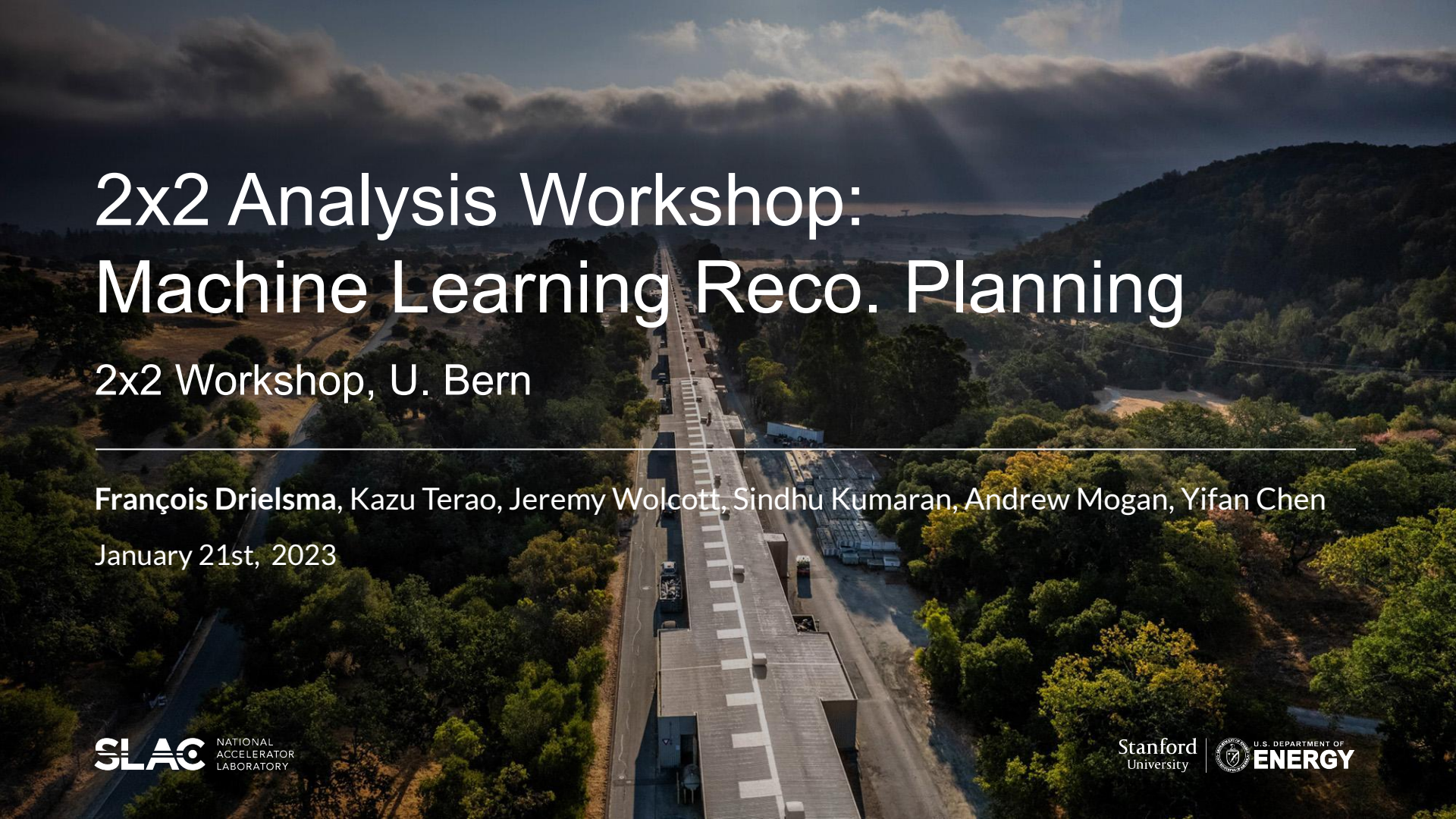


2x2 Analysis Workshop: Machine Learning Reco. Planning

2x2 Workshop, U. Bern

François Drielsma, Kazu Terao, Jeremy Wolcott, Sindhu Kumaran, Andrew Mogan, Yifan Chen

January 21st, 2023

SLAC NATIONAL ACCELERATOR LABORATORY

Stanford University | U.S. DEPARTMENT OF ENERGY

2x2 ML Reconstruction Planning

Requirements

Do you have what you need for your part of your work? If not, what are you missing?

- A set of software containers to run the simulation and reconstruction pipeline
  - A container to run GENIE + another for sim+reco (the latter exists)
- Software (perhaps just a set of scripts) to run a production (SLAC/LBNL)
  - Early version available from Jeremy
- Update to latest lartpc_mlreco3d
- GPU computing resources (“more”, likely with V100/A100)
  - Need to evaluate/request GPU-hours
- Update to using the flow
- Human resources to develop above + the core work (ML chain optimization)
- Need up-to-date geometry for sample production (for everyone)

2x2 ML Reconstruction Planning

Human resources

Do you have manpower? If not, what are some specific areas where people could jump in? What previous knowledge is needed?

- Areas:
  - Development of a production workflow (again, just a set of scripts to start, but perhaps we want a file database? Need a pipeline generating CAF)
  - Generation of simulation samples
  - Help analyzers to utilize the reconstruction output
- Necessary previous knowledge
  - Rough ideas on how a production works (for the first sub-bullet)
  - How to run a simulation chain (for all areas)
  - How the reconstruction works and what are the outputs (for the last bullet)

2x2 ML Reconstruction Planning

Human resources

Do you have manpower? If not, what are some specific areas where people could jump in? What previous knowledge is needed?

Enough resources in the immediate future:

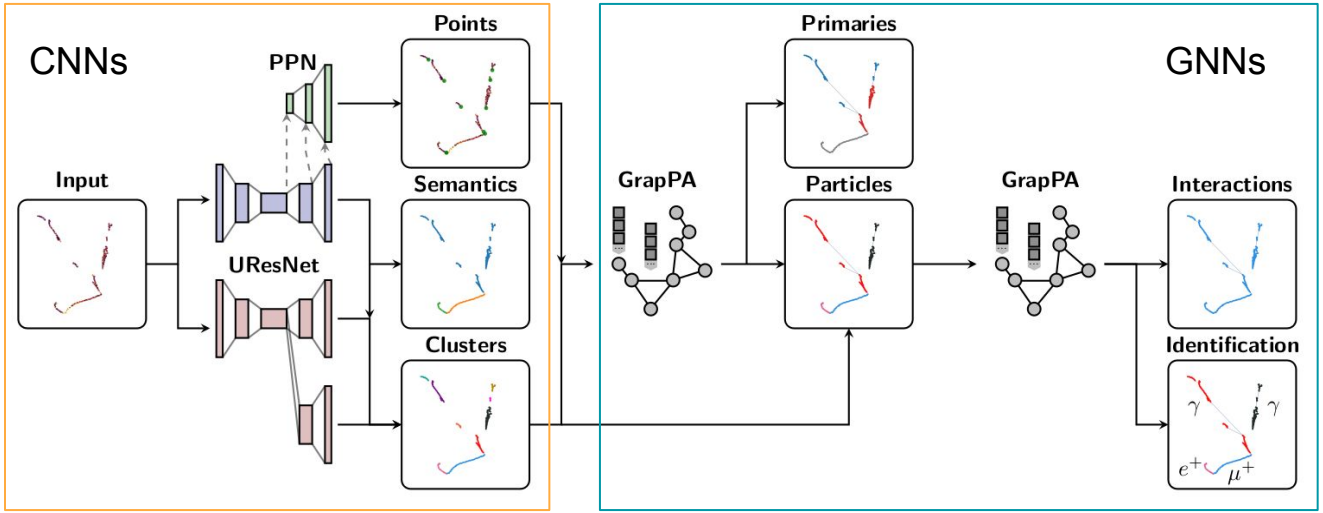
- **Jeremy**: supervision + educate
- **Sindhu**: learn to train/optimize reco. chian
- **Andrew**: larnd-sim label making (until now), TBD
- **Jessie**: integrate minerva reconstruction output in the reco. chain
- **Orgho/Gurkan**: production (learn to train/optimize reco. chain)
- **Francois**: solve reco. related issues + educate
- **Kazu**: file/label making + train/optimize
- **YiFan**: interface Supera with 2x2-flow output
- **Patrick**: light modeling/reconstruction

2x2 ML Reconstruction Planning

Concerns

Any concerns that came out of the workshop?

- No concerns per se but new features needed:
  - Integrate Minerva objects into the reconstruction
  - Deal with mismatched pixel pitch in Module 2



2x2 ML Reconstruction Planning

Timeline

What is feasible to have by this timeline and what is not?

- *Feb. 28: simulation of 10 weeks of data*
- *Apr. 30: CAF produced (i.e. reco output available)*

First path (no minerva) seems feasible. From Kazu's talk:

- Training sample production: ~1 week
- Reconstruction chain optimization: 1.5 months
  - Training + debugging

What about Minerva?

- Status of the reconstruction? Prerequisite
- Need a new data representation to interface with Minerva reco. output