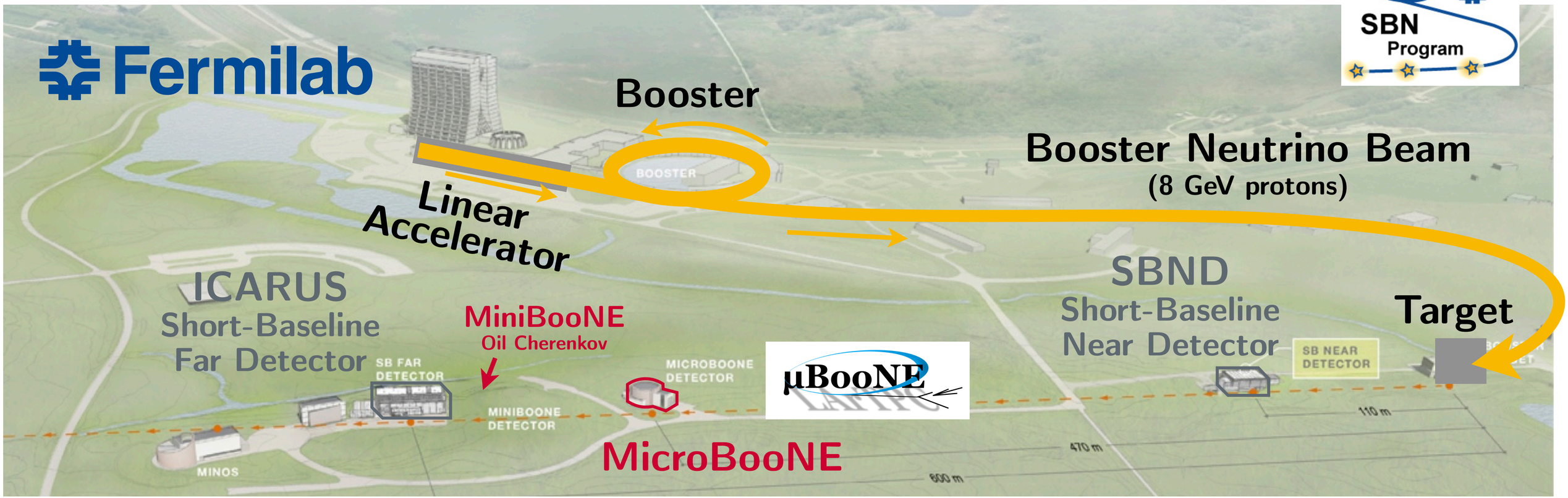
# Framework needs for the SBN experiments

Steven Gardiner

Fermilab Frameworks Workshop, 6 June 2023

# The Short-Baseline Neutrino (SBN) Program




- Add **ICARUS** and **SBND** to MicroBooNE to search for sterile neutrino oscillations with unprecedented precision
  - Further develop liquid argon time projection chamber (LArTPC) technology for DUNE
  - Measure interaction cross-sections, test various new physics scenarios

# Some framework-relevant challenges for ICARUS and SBND


- ICARUS is big (476-tonne active mass, cf. 85 tonnes for MicroBooNE)
  - Step towards DUNE scale, performance considerations (Geant4 memory, etc.)

- SBND is close to the Booster Neutrino Beam target
  - High statistics
  - Analysis infrastructure needs to consider non-negligible pileup

- Continue to support the broadest possible physics program
  - Oscillations, cross-sections, BSM, …
  - Multiple reconstruction paradigms

# Support for multiple neutrino event generators

- GENIE v3.4.0 is the primary neutrino interaction simulation for SBN
  - AR23_20i_00_000 model set shared with DUNE
  - Alternative model sets configurable via genie_xsec ups product

- LArSoft + sbncode linked to specific GENIE version
  - Changing GENIE requires rebuilding nearly the full stack → not ideal
  - MicroBooNE patch builds of GENIE were cumbersome

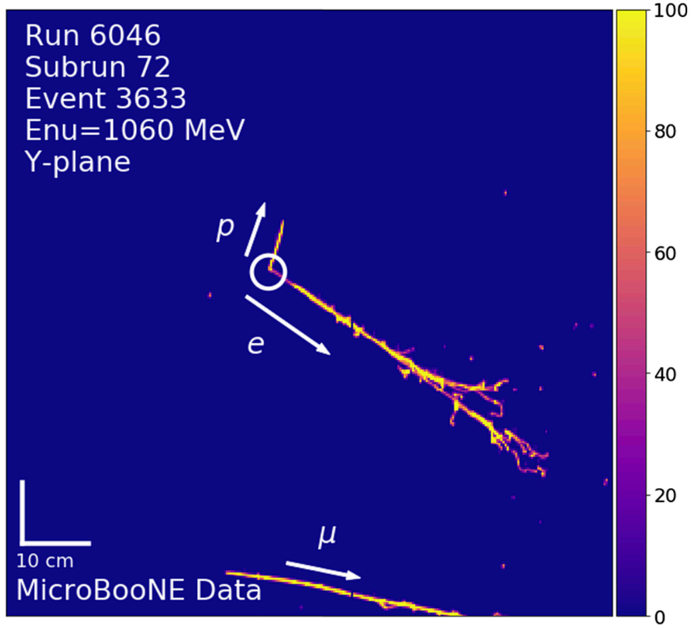
- Systematics studies benefit greatly from multi-generator production
  - Currently handled by experiments in an ad hoc way, direct support in the framework would be better
  - Full solution also requires generator community effort (e.g., moving to HepMC3 as standardized event format)

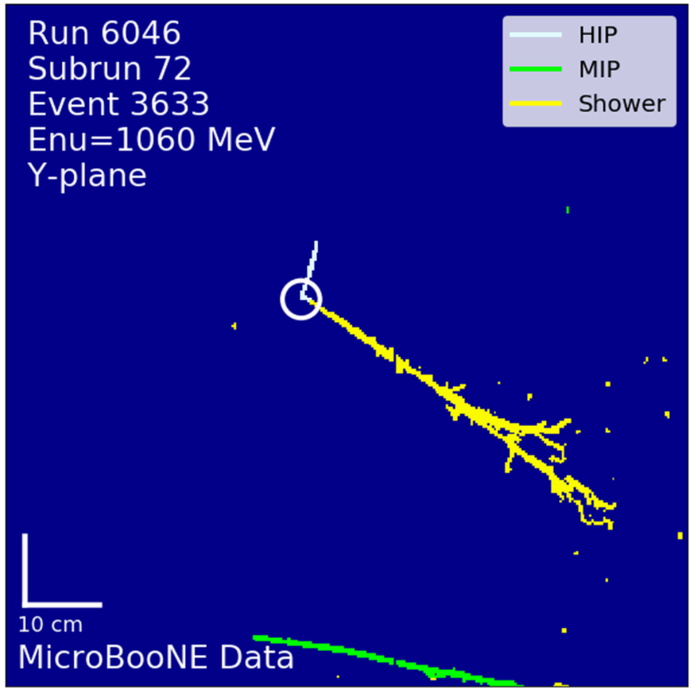
# Detector modularity

- ICARUS has four TPCs → step towards DUNE scale
- Managing relevant information has subtleties beyond assigning a TPC index
- Run signal processing per-TPC or unified?
  - 4 hit finder instances, etc.
  - Performance and organizational considerations
- Nominally identical components have practical differences
  - e.g., electron lifetime varies between TPCs
- General issue for framework development: How best to handle different detector pieces while avoiding unnecessary duplication?

# Machine learning for LArTPC event reconstruction

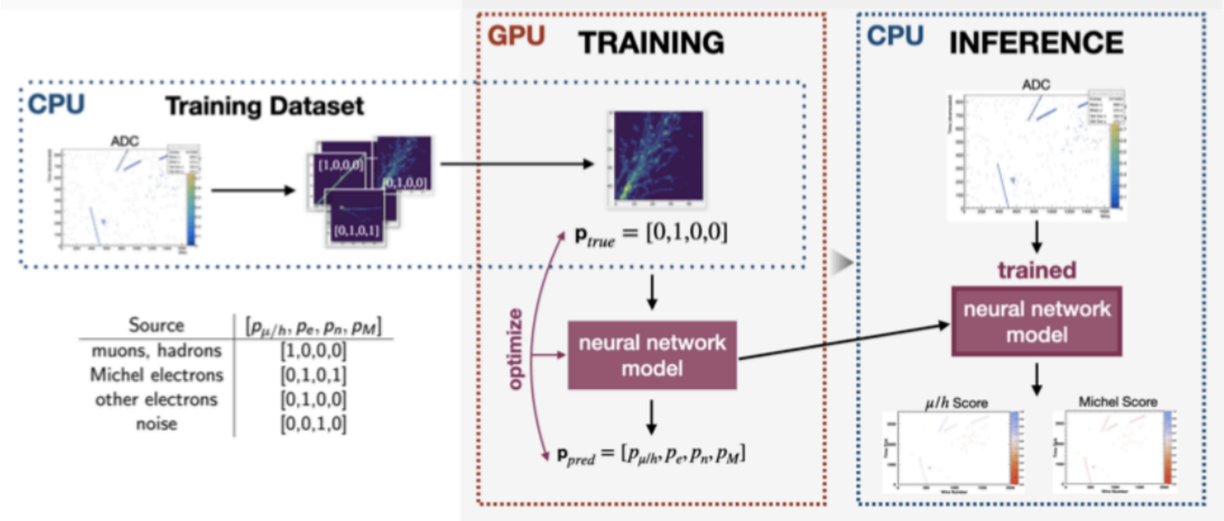
- Active area of research, multiple success stories
- SparseSSNet usage in MicroBooNE: **Phys. Rev. D 103, 052012 (2021)**
  - Key enabling technology for CCQE-based low-energy excess search: **Phys. Rev. D 105, 112003 (2022)**
- High interest in SBN to leverage ML in new analyses
- Future framework development should plan to provide support



# ML in SBND

- Tools from other experiments (DUNE, …) are being integrated into LArSoft



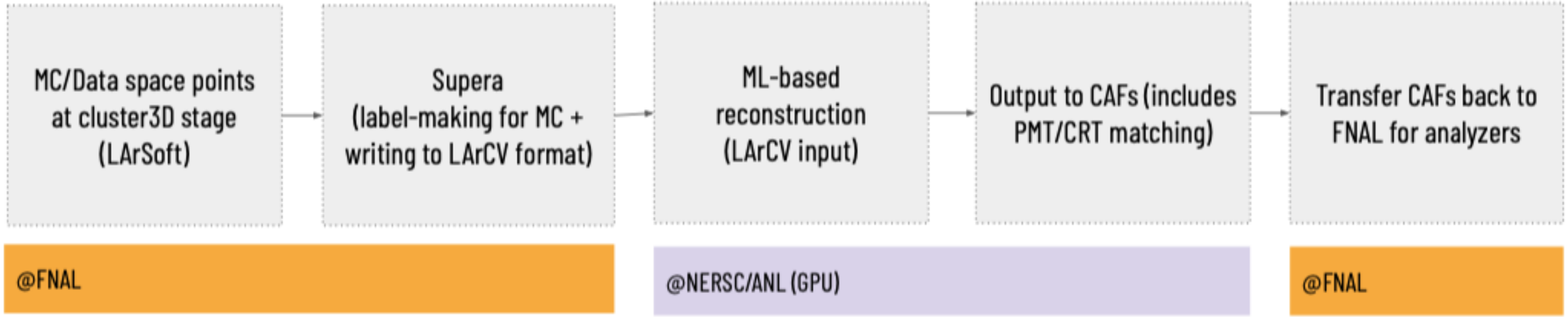

| Source | $[p_{\mu/h}, p_e, p_n, p_M]$ |
|---|---|
| muons, hadrons | [1,0,0,0] |
| Michel electrons | [0,1,0,1] |
| other electrons | [0,1,0,0] |
| noise | [0,0,1,0] |

- Example: CNN for hit classification
  - Use local information to distinguish Michel electron hits from others (muons/hadrons, noise, other electrons)
  - Training uses GPUs at Fermilab EAF
  - Inference can run in standard LArSoft-based CPU workflow
  - Easy integration into CAF files, etc.

- Use of this and similar ML tools in production is a work in progress

# ML in ICARUS

- Full ML-based reconstruction chain from 3D space points → interaction + particle identification
  - See **slides** from FNAL AI Infrastructure Planning Mini-Workshop
- LArCV files (with converted LArSoft spacepoints) transferred to off-site GPU HPC cluster (NERSC, ANL)
  - Inference outputs provided as analysis-level files (CAFs)



# ML Summary

- The two experiments are pursuing very different paths for ML usage so far
  - SBND: Integrate tools into LArSoft, run inference on CPUs with minimal workflow adjustments
  - ICARUS: offload prior LArSoft results to external HPC clusters, transfer final results back for analysis

- Combining strengths of each workflow into something more unified could be an optimal path forward
  - Many details to be worked out, of course

- Future resource needs for ML in SBN unclear at present, but development continues

# General HPC workflows

- Also of interest for LArTPC reconstruction and analysis beyond ML techniques
- See Giuseppe Cerati's slides for some discussion of ongoing work
- Potential for greater performance on many fronts
- However, fully refactoring simulation + reconstruction to be HPC-friendly would be a “huge lift”
  - Expect more targeted improvements for now (e.g., ICARUS 1D signal processing)

# Bringing data products “back to art”

- DAQ/simulation + much of the reconstruction happens in the usual art pathway
- Some workflows rely on exporting art data products to non-art environments
  - e.g., the ICARUS ML reconstruction strategy
- Sharing the results with all analyzers via seamless merging would be ideal
  - e.g., importing a collection of recob::Shower objects from an external source
- Potentially doable already, but perhaps could be simplified
- ICARUS ML effort trying something similar downstream
  - Import into CAF files rather than artroot files
  - Still available to all analyzers at a later stage

# User experience during analysis development

- Framework machinery makes ample use of “deep” C++ features (related to templates, etc.)
  - Powerful, and often “the right tool for the job”

- Can represent a challenge for analyzers who are novice programmers
  - Many physicists learn their C++ “on the streets”
  - Some opportunities for direct training in our community, these are valuable!

- Interoperability with Python can mitigate a lot of the challenges
  - Seems to be preferred by many junior collaborators these days
  - Jupyter notebooks, etc.

- The Python interface to gallery can be part of the solution
  - Actively used in SBN for small analyses and “under the hood” for a popular event display
  - Continued support is welcome

# Job submission

- Perhaps outside of the nominal “framework” scope, but important infrastructure
- POMS recommended over the project.py script from larbatch
  - More optimal use of resources, etc.
- POMS is not LArSoft-specific
  - Lower-level approach can present difficulties for the average LArSoft user
- Perhaps worth considering a higher-level layer to make a POMS workflow more friendly to newcomers

# Revisiting the “ntuple maker” strategy

- LArSoft-based production workflows process artroot files
  - Manipulation of data products requires ROOT dictionaries
  - Relatively bulky files
- Downstream analysis work typically uses slimmed files of some kind
  - Avoid grid jobs, prestaging lots of data + MC when unnecessary
- MicroBooNE has several such formats, each developed by a distinct group
  - “ntuple maker” code used to dump data products from artroot format
- SBN primarily uses CAFAna format (shared by NOvA, DUNE)
  - Still requires a dedicated “CAF maker” processing stage
- Could be worth considering alternatives
  - Parallel “plain ntuple” stream for data products?

# Conclusion

- Computing infrastructure, including Fermilab-led software framework development, plays a crucial role for the success of SBN
  - Deliver physics from our LArTPCs and pave the path to DUNE

- SBN framework needs exist on multiple fronts, including but not limited to
  - Interaction simulations
  - ML-based reconstruction
  - General HPC workflows
  - Analyzer-level user experience

- Thanks for your continuing support of our experimental program. We look forward to collaborating on new solutions!