SBN Analysis Infrastructure

Wesley Ketchum and Joseph Zennamo

SBN Oversight Board Meeting

11 Sep 2020

SBN Analysis Infrastructure

Our goal is to coordinate and address data and software infrastructure and computing resource needs across the SBN to enable SBN program's science goals

- The effort grew out of the SBN Data Management Group, as it was recognized a more comprehensive focus was needed

We work alongside the current SBN working groups and ICARUS/SBND physics groups, focusing on basic infrastructure of ***how*** to do things

- ***What*** to do is defined by the collaboration physics and analysis groups

Main focus topics

1. **Release Management**
   a. Maintain high-quality releases of SBN-specific software packages
2. **Production and Resource Management**
   a. Work with Fermilab SCD and SBN collaboration to maintain workflows & data-management and access
3. **Simulation Software Management**
   a. Develop infrastructure to support a data-driven detector simulation and maintain its consistent configuration
4. **Analysis Software Management**
   a. Develops infrastructure to process & analyze data & MC in a consistent way
5. **Beam and "Dirt" Simulation**
   a. Develop and improve beamline simulations and uncertainties

Activities of previous months

Software and Release Management Activities

Completed transition of SBN repositories to hosting on “Github” (https://github.com/SBNSoftware)

- Follows LArSoft transition to Github
- Enables better control of software updates via *pull requests*
- All SBN software accessible to collaboration members

Assessed and nearing completion of migration to common *SBN* software dependency

- Allows easier maintenance of common simulation and reconstruction algorithms and configuration, while still enabling detector-specific calibration and analysis

Production and Resource Management Activities

Completed assessment and annual presentation of necessary computing resources to Fermilab Computing Resource Scrutiny Group (FCRSG)

- Includes a comprehensive computing model evaluating needs for computing and data storage over next 3 years
- Continuing further discussion with Fermilab SCD to improve/further optimize that model

Currently participating (with DUNE) in ESNet review of global networking model and needs

Establishing data replication to offsite locations (e.g. INFN CNAF)

Baseline Computing and Data Storage Needs

Computing Needs (CPU Mhr)

	FY 2021	FY 2022	FY 2023
ICARUS	16	18	25
SBND	6	7	10
Total	***22***	***25***	***36***

Data Storage Needs (PB)

	FY 2021	FY 2022	FY 2023
ICARUS	12	15	21
SBND	1	13	23
Total	***13***	***27***	***44***

As expected, resource needs are high, but we are getting the support needed from FNAL SCD to ensure a successful physics program

Simulation and Analysis Software

Coordinating with working groups on common software improvements

- E.g. updated *GEANT4* simulation interface, optical modelling studies, data-driven cosmics simulation, etc.
- Implementing *Central Analysis Framework* (CAF) for final output files for common analysis, reducing file size and data access costs

Developing and testing improved workflows to efficiently use computing and data storage

Next steps

Preparing for Data

With SBN-FD (ICARUS) detector coming into operation soon, we are working with SBN DAQ and ICARUS software and physics groups to ensure access and timely processing of data from the detector

- Commissioning data is stored in FNAL tape-backed storage, is available on SBN computing nodes, and is searchable/indexed through SAM file catalog
- We are using common FNAL SCD tools for file access and automated processing of data as it comes from the detector

Anticipated “nearline” workflow

1. Each raw data file (“Raw”) is promptly sent from detector hall to tape-backed disk storage at FNAL central computing
    - Data at detector hall is not deleted *until* data transfer fully verified

2. Prompt processing of raw data at FNAL Fermigrid performs decompression, channel mapping, noise filtering, signal processing and region-of-interest finding, outputting a new smaller (“Reco1”) file suitable for high-level reconstruction
    - Anticipate ~350 nodes to keep-up with data-taking during operations

3. Prompt high-level reconstruction is run, to verify calibration constants and perform high-level data quality checks

4. Both “Raw” and “Reco1” files are written to archival tape at FNAL, and available to send offsite (e.g. INFN CNAF) for further processing and backup storage
    - Ideally, “Reco1” files will be used as inputs for final offline reconstruction and analysis of data, but raw data always available

Status of “nearline” data management workflow

Commissioning data is currently being transferred from SBN-FD to central storage

Automated nearline processing running on new data as it arrives, and made available for offline analysis and use

Final configurations of workflow and “Reco1” file production awaits further commissioning and analysis of data

Have previously exercised data transfers to INFN CNAF, and now making renewed effort to automate transfer of data

2021 Production Campaign

We are working with SBN working groups to prepare and execute a major simulation and data analysis campaign starting in January 2021 to achieve the following

- Reconstruct and analyze available SBN-FD (ICARUS) data with most up-to-date algorithms
- Generate simulated events corresponding to ~10 times one year's worth of neutrino interactions for both detectors to enable physics studies and analysis by next summer
- Do so in a coordinated "joint" software and data production toolkit, to prove ability to perform a joint analysis

Global Timeline

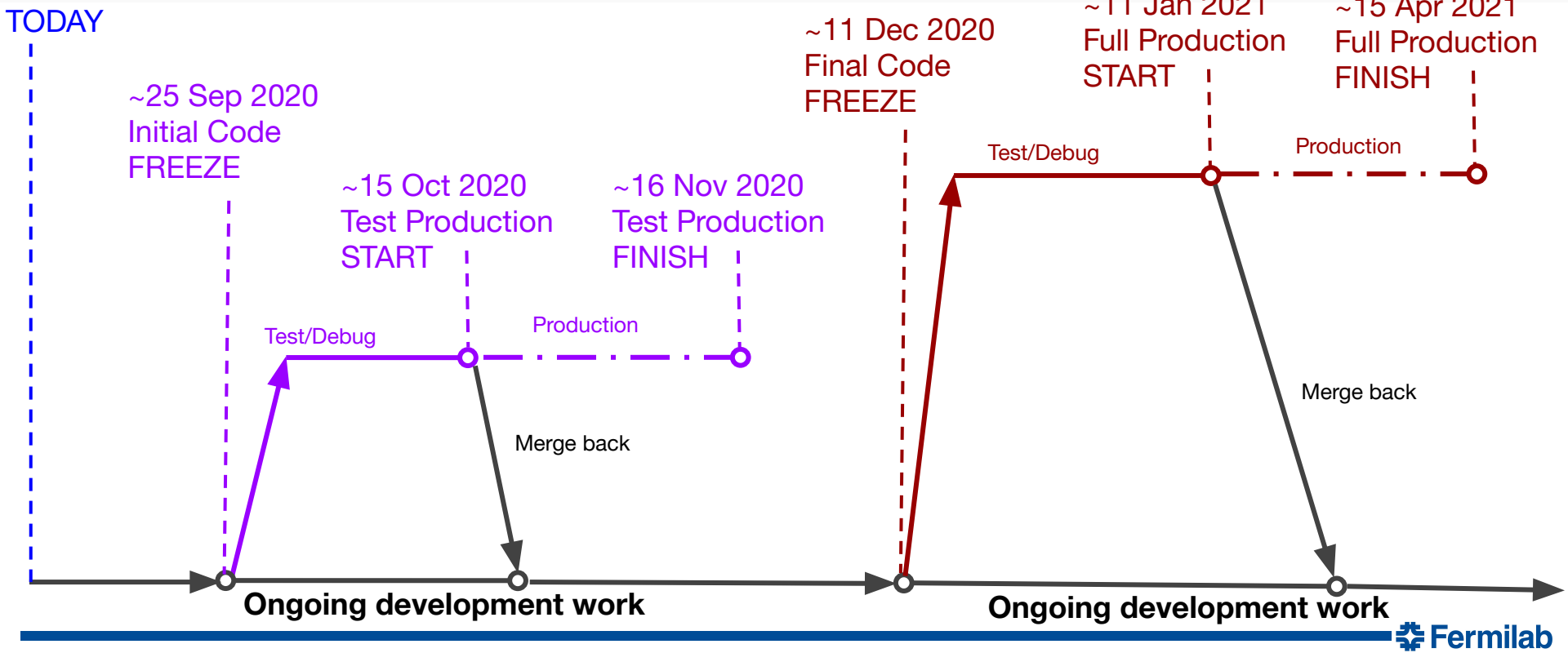
A production intended for Summer 2021 should start **January 2021**

- This allows for analyzers to fully integrate and utilize these datasets

To validate our production capabilities we aim for a small-to-mid scale production to be launched in ~October 2020 over ~1-month

- This will create necessary samples to benchmark progress towards Dec. code freeze and to serve as samples for presentations to the January PAC as necessary

Cartoon of Global Timeline



Sharing of data, software, and resources

We are working to further develop tools and documentation that ensure access to data and software across the SBN, in accordance with the *Statement of Principles for Data Sharing, Analyzing, and Publication within the SBN Program*

- We further coordinate implementation of practical details with the SBN Analysis working group, and the ICARUS and SBND Software Coordinators
- We look forward to the *Dataset and Computing Coordination Committee (D3C)* of the SBN IB and will update implementation according to any new policy

Summary

SBN Analysis Infrastructure group is actively working to develop the infrastructure to successfully execute the SBN physics program

We are preparing for the prompt production of data from the ICARUS detector, and a major simulation and reconstruction campaign to support first analysis of data from the SBN program and further development of the SBN analyses

Backup Slides

Key Assumptions

We are aiming to generate enough MC to study 1-yr worth of data taking:

ICARUS - 7,000,000 neutrinos[§], 100,000 in-time cosmics[✣]

SBND - 15,000,000 neutrinos[⋆], 1,000,000 in-time cosmics[*]

in addition we'd also process all available ICARUS physics data

Our production will have to operate within our "current" computing limitations

We will ask SCD to allocate resources based on our previously-presented requests (which was based on above numbers)

We will test our production capabilities in advance of the full production (e.g. October)

We will separate our production into three phases:

Sim+signal processing, high-level (3D) reco, & ana files

Fermilab

[§]BNB+NUMI 10x Stats [*]BNB Only 10x stats [✣]Assume modeling uses real data [⋆]4x stats

Identified Priorities (in no order)

- Migration to new LArG4
- CRT geometry updates
- Solidify the our in-time cosmic workflows
- Adopt CAF file creation into production
  - These should be developed so that they can act as the default (full sample) analysis method
- Integration of systematic packages and frameworks
- Adopt a recombination model that integrates charge and light production (e.g. ArNEST)
- Coalesce on optical modeling across both detectors
- Event overlay utilities
- 2D drift simulation
- 2D signal processing
- Highly reduced file sizes

Detailed Timeline

Goals for October:

- Migration to new LArG4 -
  Not achievable due to geo. work?
- *SBND PDS geometry updates*
- CRT geometry updates
- Solidify the in-time cosmic workflows
- Highly reduced file sizes
- Adopt CAF file creation into production
- *Op Det sim and reco integration*
- Enable the ability to run multiple optical model in each detector to cross-validate
- Test of multi-threading where implemented

Goals for January:

- Adopt a recombination model that integrates charge and light production (e.g. ArNEST)
- Coalesce on optical modeling across both detectors
- Event overlay utilities
- 2D drift simulation
- 2D signal processing (*)
- Capability to create detector simulation variation samples
- Capability to assess systematics uncertainties with CAFs