

SBN Computing Calculations Overview

Joseph Zennamo and Wes Ketchum

SBN Oversight Board Meeting

11 December 2020

Computing Model

The Fermilab computing divisions require an annual review of expected computing resource usage for all major experiments

Presented to the “Fermilab Computing Resource Scrutiny Group” (FCRSG)

Templates for the presentation are provided, with generally coarse questions of the expected CPU and data storage needs, alongside requests of service and support

In this year’s review (May 2020), we presented a combined SBN computing needs, and worked with SBND, ICARUS, and SBN groups to develop a comprehensive model to inform these numbers

We also made available detector-specific numbers upon request

Many of the inputs may change given new information from detector commissioning and production tests, so numbers should be taken as estimates and will be updated soon

Inputs to the computing model: data

For data, the main inputs (from experiments' calculations) are

- Expected final trigger rates from beam sources
 - For ICARUS, this is ~ 0.15 Hz for the BNB, and also ~ 0.17 Hz for NuMI
 - For SBND, this is 0.9 Hz for the BNB
- Additional trigger rates of data for calibration and cosmics background modeling
 - ICARUS: an additional 0.5 Hz rate is assumed
 - SBND: an additional 1.69 Hz rate is assumed,
larger due to need for cosmic data to generate simulated event
- Expected data sizes, including lossless compression
 - ICARUS: ~ 143 MB/event is assumed
 - SBND: ~ 37.5 MB/event is assumed
- Planned for higher trigger rates during commissioning to allow a gradual implementation of the trigger systems
 - For both experiments, assume a 3 month period with average 3.33 Hz total data rate (5 Hz first month, 3 Hz second month, 2 Hz third month)

Inputs to the computing model: simulation

For simulation, the main inputs (from experiments' calculations) are

- The expected number of neutrino events per POT and POT collected per year to get total expected neutrino events in data
 - For ICARUS: $2e5$ BNB events and $4e5$ NuMI events per year
 - For SBND: $1e6$ BNB events per year
- Estimated the simulation statistics based on these data statistics
 - For both detectors, assumed 10x statistics for BNB, 5x stats for ICARUS NuMI
- Included an estimates for the number and statistics of additional systematic variation samples
 - For both detectors, assume 5 such datasets with x1 data statistics
- The number of additional simulation events needed (dominated by cosmic simulations)
 - Once data-taking starts, assume $1e5$ yearly for each detector
 - Prior to data-taking, assume $1e6$

Inputs to the computing model: CPU usage

For both simulation and data, on average for each event we need to know (based on previous productions)

- Time to process for simulation stage
 - ICARUS: 240 s for events with (simulated) cosmics, 50 s for neutrinos only
 - SBND: 420 s for events with (simulated) cosmics, 30 s for neutrinos only
 - *SBND due to longer readout window configuration, which is being further evaluated for necessity for upcoming productions*
- Time to process for signal-processing stage
 - ICARUS: 150 s per event
 - SBND: 30 s per event
- Time to process for reconstruction stage
 - ICARUS: 100 s per event
 - SBND 10 s per event
 - *SBND numbers by contrast here may be too optimistic in the longer run*

Inputs to the computing model: disk usage

For both simulation and data, on average for each event we need to know (based on previous productions)

- Data size after simulation stage
 - ICARUS: ~165 MB/event
 - SBND: ~57.5 MB/event
- Data size after signal-processing stage
 - Assuming raw waveforms are dropped for most events
 - ICARUS: 60 MB/event
 - SBND: 50 MB/event
- Data size after reconstruction stage
 - ICARUS: 80 MB/event
 - SBND: 60 MB/event

Note: assumed additional data volume of analysis-level files is small relative to these numbers

Assumptions of the computing model

- All collected data after commissioning is saved to permanent tape storage
 - A backup copy at off-site location is assumed for critical (beam) data
- Two major production campaigns per year
 - Only the final reconstruction output saved to tape for most events
 - While it is possible to save simulated truth and waveform data to tape, we found it advantageous in the computing model to re-simulate ‘from scratch’ twice yearly, rather than save and restore from tape
 - Each production campaign assumed to take between 60 and 90 days
- Collected detector data is “immediately” processed through full reconstruction for calibration purposes
 - Goal to make signal processing stable as early as possible, so signal-processing output of this ‘keep-up’ processing is used as input for later reconstruction campaigns
- Reconstructed data from previous campaigns is ‘retired’ (overwritten on tape) after two years to keep total data footprint lower
 - Raw data is never retired
 - Analysis level files will be maintained permanently as part of data retention strategy
 - Plan for staged retirement of most commissioning data as well

ICARUS: Data Events Collected

	FY 2021	FY 2022	FY 2023	FY 2024
<i>Commissioning Data</i>	26.3e6			
BNB Triggered	2.37e6	3.55e6	3.55e6	3.55e6
NuMI Triggered	2.68e6	4.02e6	4.026e6	4.028e6
Other	11.8e6	15.8e6	15.8e6	15.8e6
Total “Physics” Data	16.9e6	23.3e6	23.3e6	23.3e6

ICARUS: Total Events in Major Campaigns

	FY 2021	FY 2022	FY 2023	FY 2024
Total Data (BNB, NuMI, Cosmics Bkg.)	11.8e6	23.7e6	35.5e6	47.3e6
BNB Simulated	2e6	4e6	6e6	8e6
NuMI Simulated	2e6	4e6	6e6	8e6
Other Simulated (incl. systematics)	3e6	6e6	9e6	12e6
Total Simulated	7e6	14e6	21e6	28e6

SBND: Data Events Collected

	FY 2021	FY 2022	FY 2023	FY 2024
<i>Commissioning Data</i>		26.3e6		
BNB Triggered		14.2e6	21.3e6	21.3e6
Other		40.0e6	53.3e6	53.3e6
Total “Physics” Data		54.2e6	74.6e6	74.6e6

SBND: Total Events in Major Campaigns

	FY 2021	FY 2022	FY 2023	FY 2024
Total Data (BNB, Cosmics Bkg.)	0	39.0e6	78.1e6	117e6
BNB Simulated	10e6	20e6	40e6	60e6
Other Simulated (incl. systematics)	6e6	10e6	20e6	30e6
Total Simulated	16e6	30e6	60e6	90e6

Total CPU Usage (MCPU-Hr)

	FY 2021	FY 2022	FY 2023	FY 2024
ICARUS Data	12.5	11.7	15.1	18.6
ICARUS Simulation	3.5	6.8	10.2	13.5
ICARUS Total	16.0	18.5	25.3	32.1
SBND Data	0	4.3	5.4	7.1
SBND Simulation	5.6	2.4	4.7	7.1
SBND Total	5.6	6.7	10.3	14.1
<i>SBN Total</i>	<i>21.9</i>	<i>25.1</i>	<i>35.6</i>	<i>46.2</i>

Total Data Usage (PB)

	FY 2021	FY 2022	FY 2023	FY 2024
ICARUS Raw Data	6.2	7.6	10.9	14.2
ICARUS Reco Data	4.4	4.2	5.1	4.7
ICARUS Simulation	1.1	2.8	4.5	6.2
ICARUS Total	11.8	14.6	20.5	25.0
SBND Raw Data	0	3.1	5.4	8.3
SBND Reco Data	0	5.7	8.9	11.2
SBND Simulation	1.0	4.1	9.0	14.4
SBND Total	1.0	12.9	23.3	33.9
<i>SBN Total</i>	<i>12.7</i>	<i>27.5</i>	<i>43.9</i>	<i>58.9</i>

Final Notes

These are estimates under the inputs and assumptions presented, not determinations of detector trigger configuration, SBN program planning, or specific physics analysis needs

We are aiming for improvements in many areas to reduce need, including:

- More efficient algorithms and data storage

- Use of multithreading and heterogeneous computing (including at High-Performance Computing centers)