



Production Pipeline Aurora

Wirecell Meeting

Avinay Bhat
09/22/2026

SBND Gen-2 production on Aurora

- Production framework: SBN-ALCF sbn_parsl
- Target production release: sbndcode v10_14_02_03
- Full chain demonstrated successfully:
 - GEN → G4 → DetSim → Reco1 → Reco2 → flat CAF
- Initial Parsl test:
 - 100 GEN attempts
 - 30 accepted physics events
 - all 30 propagated successfully through CAF
- Successful production exercised:
 - standard SBND Gen-2 FCLs
 - Aurora container environment
 - Parsl task dependencies
 - intermediate ROOT outputs
 - LArCV / calibration outputs
 - flat CAF generation
- Critical Aurora container issues were identified and documented
 - bind-mount requirements are release-dependent
 - v10_14_02_03 and v10_14_02_05 require different handling

Comparison against Thomas Wester's production

- Used the same 20 Thomas G4 files
- 88 total physics events
- Same downstream release:
 - sbndcode v10_14_02_05
- Same standard:
 - DetSim
 - Reco1
 - Reco2
- Our Parsl chain additionally produced flat CAFs
- Agreement88 / 88 events reproduced
- Identical persisted schemas:
 - DetSim: 59 / 59 branches
 - Reco1: 73 / 73 branches
 - Reco2: 231 / 231 branches
- File sizes agree within approximately 0.2%
- All deterministic G4 / truth pass-through products are bit-identical
- No missing products or unexpected structural differences

Readiness for a 1M-Event Production Campaign

- DetSim remains the dominant CPU stage
- Typical process memory remains modest:
 - peak approximately 6–7 GiB
- Reco1 output: approximately 28 MiB/event
 - about 27 TiB for 1M events
- Reco2 output: approximately 31 MiB/event
 - about 30 TiB for 1M events
- Flat CAF: approximately 5 MiB/event
 - about 5 TiB for 1M events
- Keeping all major intermediate stages would require $O(100 \text{ TB})$ of storage
- Production strategy
 - Restore the official v10_14_02_03 Gen-2 configuration
 - Freeze the Parsl code, TOML, FCLs and container setup
 - Run a staged scale-up:
 - small production validation
 - $O(10k)$ events
 - $O(100k)$ events
 - then 1M
 - At each step verify:
 - success / retry rates
 - events/hour/node
 - memory
 - storage growth
 - metadata integrity
 - CAF completeness

Back Up Slides