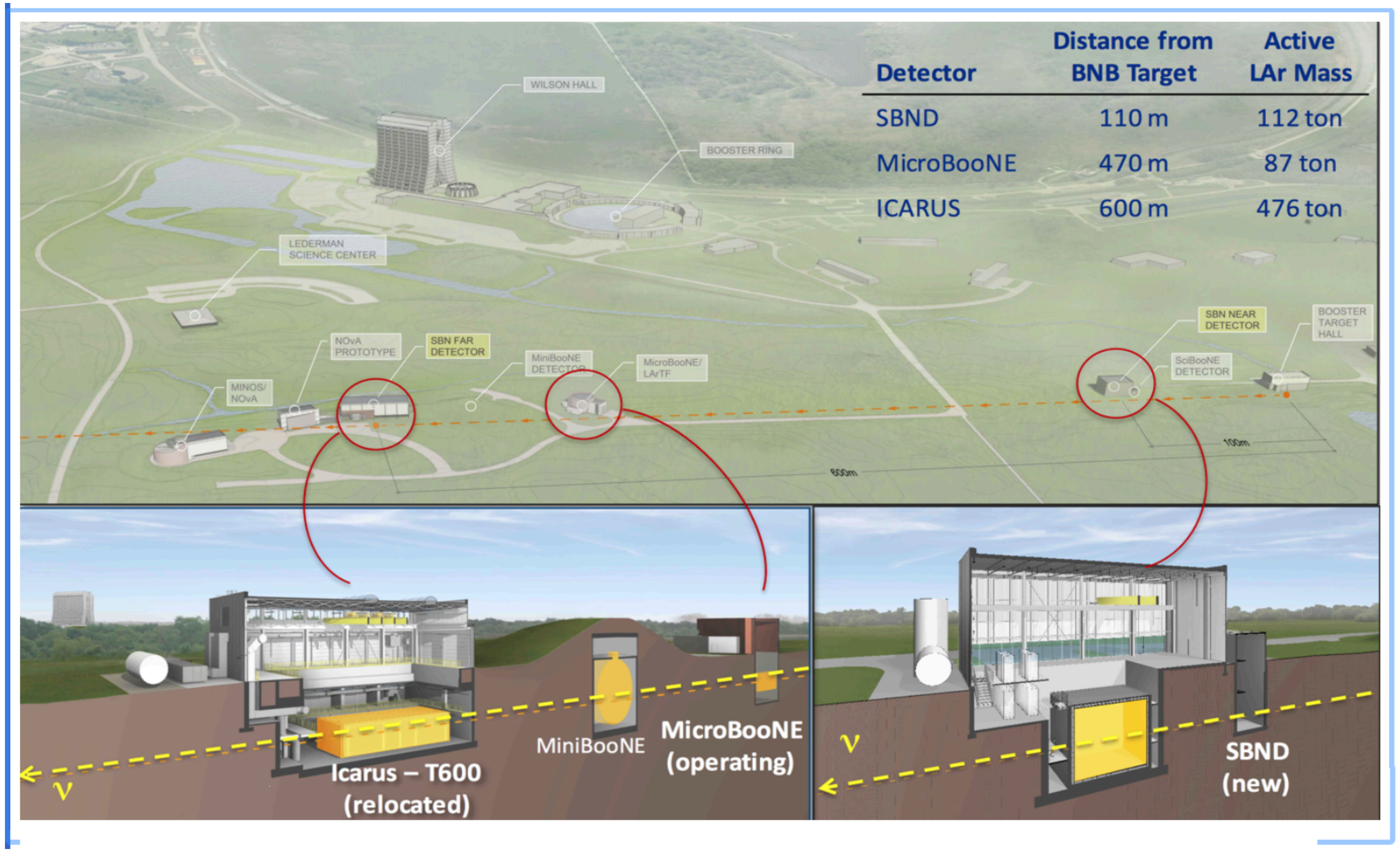
SBNFit Update

H. Schultz & M.Wospakrik

SBN Program



- **SBN**: **3 liquid argon detectors** located in the **Fermilab Booster Neutrino Beamline**.
  - Study existence of **light sterile neutrino(s)** driving oscillations at $\Delta m^2 \sim 1eV^2$
- Each detector shares the **same neutrino flux** & **argon cross-sections** -> measurement is **highly correlated**.
- Exploit the **correlations** to **reduce systematic uncertainties** by multiple side-by-side channel fits for the neutrino oscillation analysis.

What is SBNfit

- Framework designed to perform simultaneous fits across data from multiple, correlated distributions.
- Developed by MicroBooNE/SBN collaborators at Columbia University (Georgia Karagiorgi, Mark Ross –Lonergan, Guanqun, Davio Cianci, et al.)

Multi-mode	Multi-detector	Multi-channel
• Neutrino/anti-neutrino • BNB/Numi Beam	• SBN: • SBND+MicroBooNE+ICARUS • MiniBooNE+MicroBooNE • SBN+DUNE	• 1 electron + 1 proton • 1 muon + 1 proton • 1 gamma + 1 proton

Allows for combined fitting of ***arbitrarily large*** number of modes, detectors and channels simultaneously, fully accounting for systematic correlations.

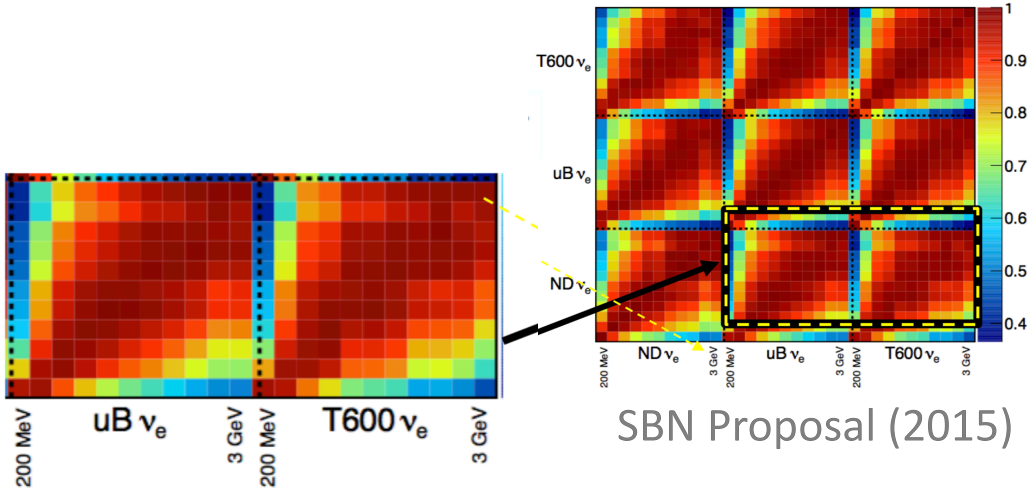
SBNfit Framework

- ROOT-based framework
- Use covariance-matrix based approach
- Chi2 Test Statistics:

$$\chi^2(\Delta m^2_{i1}, U_{\alpha i}, \phi_{ij}) = \sum_{k=1}^{M}\sum_{l=1}^{M}\left[N_k^{\text{null}} - N_k^{\text{osc}}(\Delta m^2_{i1}, U_{\alpha i}, \phi_{ij})\right] E^{-1}_{kl} \left[N_l^{\text{null}} - N_l^{\text{osc}}(\Delta m^2_{i1}, U_{\alpha i}, \phi_{ij})\right]$$

3 scenarios:
3 parameters: 3N+1
7 parameters: 3N+2
12 parameters: 3N+3

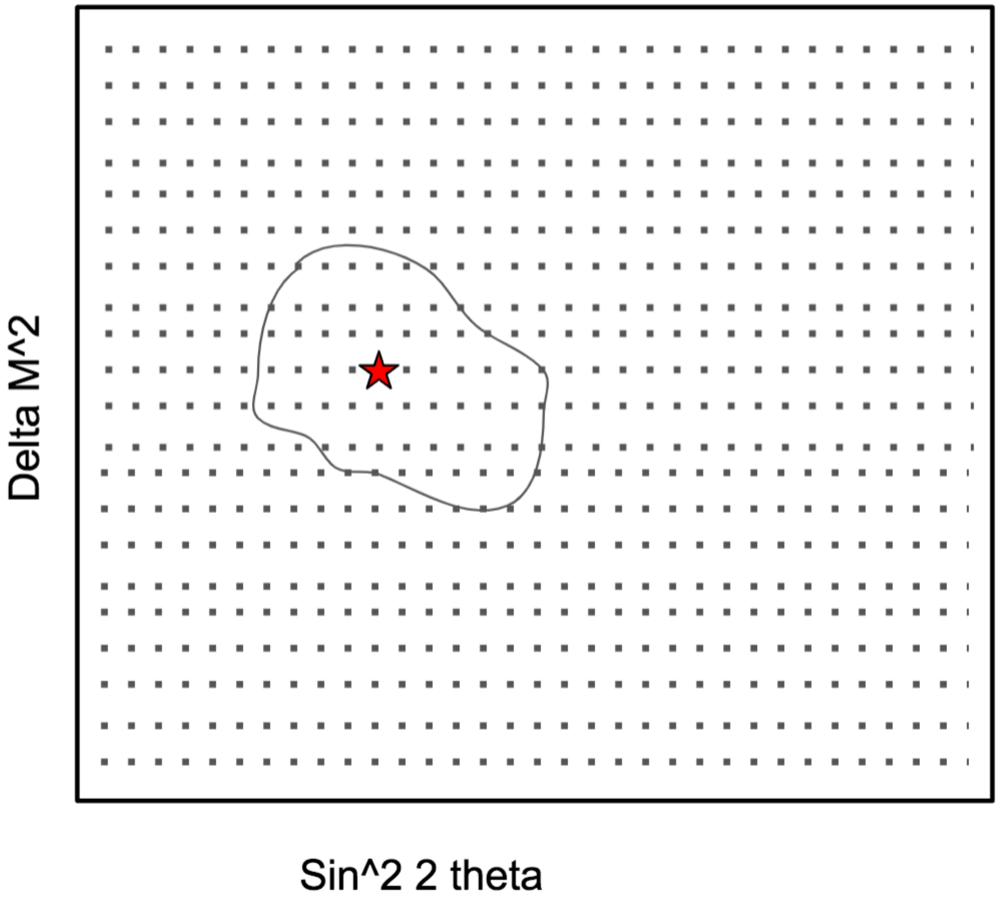
3 Detector ν_e Background Systematics Correlation Matrix
(flux and cross section systematics)



SBN Proposal (2015)

SBNfit (No FC Correction)

Figure from M. Ross-Lonergan

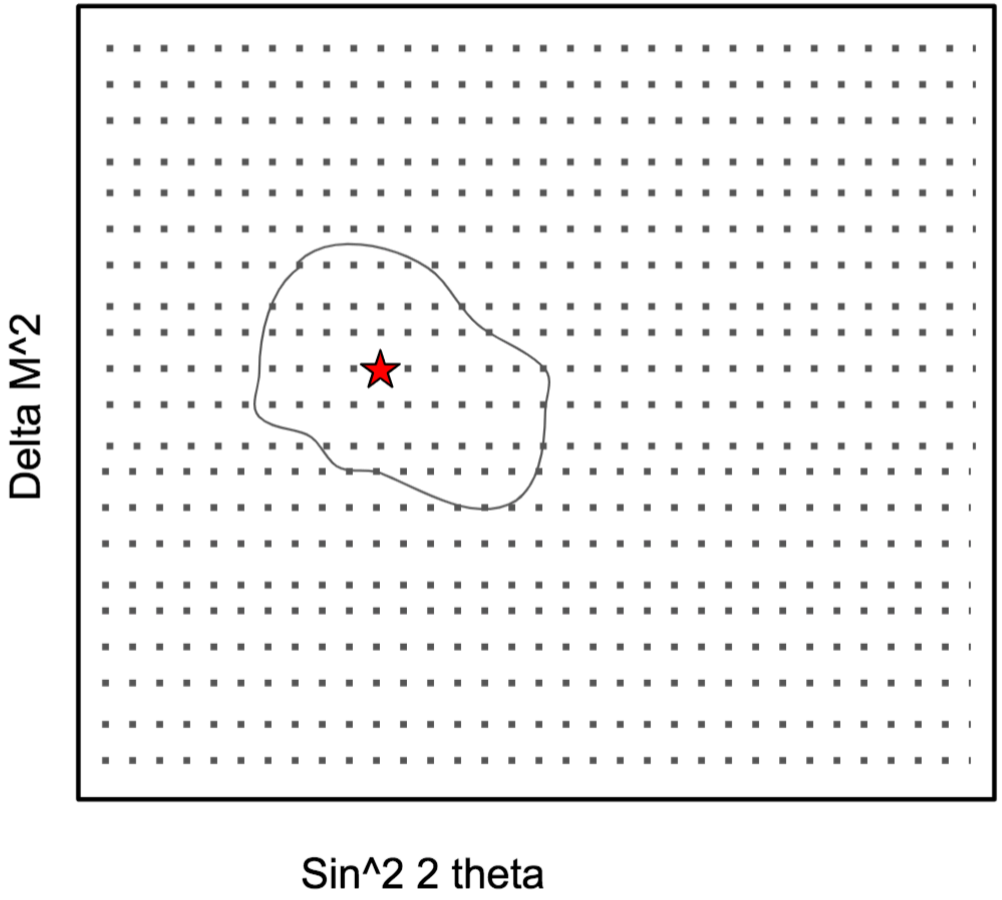


Log likelihood ratio with Gaussian approach

1. Given observed data spectrum D
2. Find the grid point with the minimum χ^2
3. Calculate $\Delta\chi^2$ between at each grid point
4. Assuming delta $\Delta\chi^2$ is distributed as distribution with 2 d.o.f, draw a contour of allowed regions around the Best Fit point.

SBNfit (with FC correction)

Figure from M. Ross-Lonergan

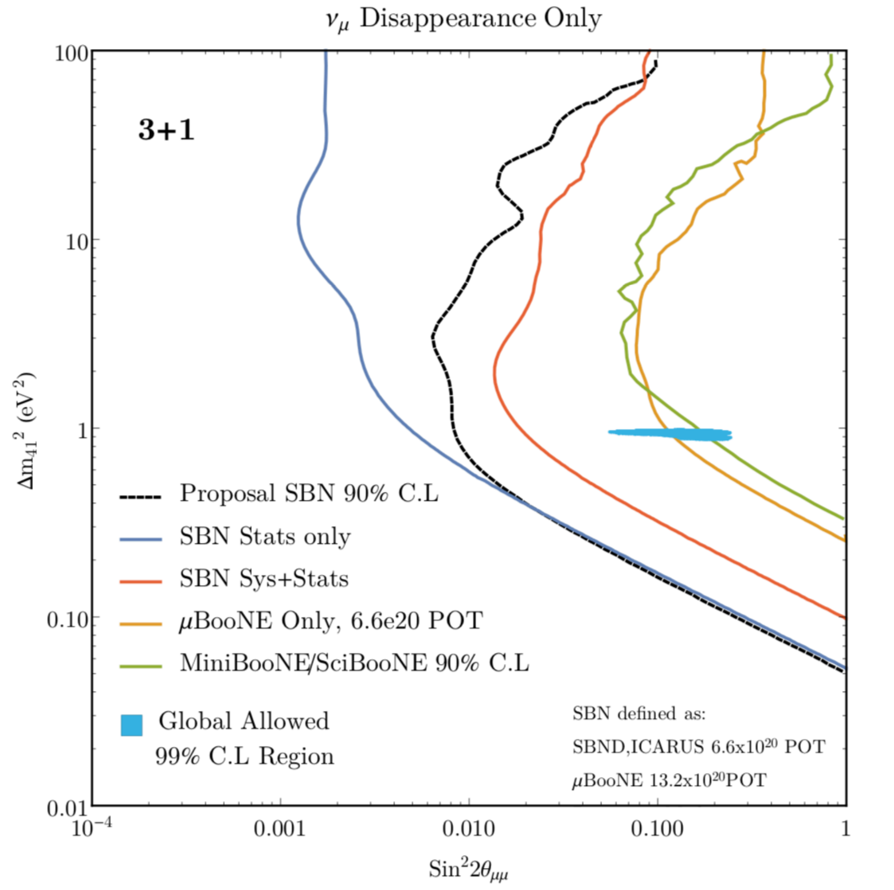


Ideally, 100*20*20, but we perform the test with 100*100*100

Log likelihood ratio with Gaussian approach

1. Given observed data spectrum D
2. Find the grid point with the minimum χ^2
3. Calculate $\Delta\chi^2$ at each grid point
4. Calculate exactly the value of that 90% of experiments would be in by generating pseudo experiments, rather than by assuming gaussianity, wilks theorem and 2 d.o.f.
5. Do this for 3σ (10^4 pseudo experiments/"universes") and 5σ ($\sim 10^8$ pseudo experiments/"universes")

SBNfit

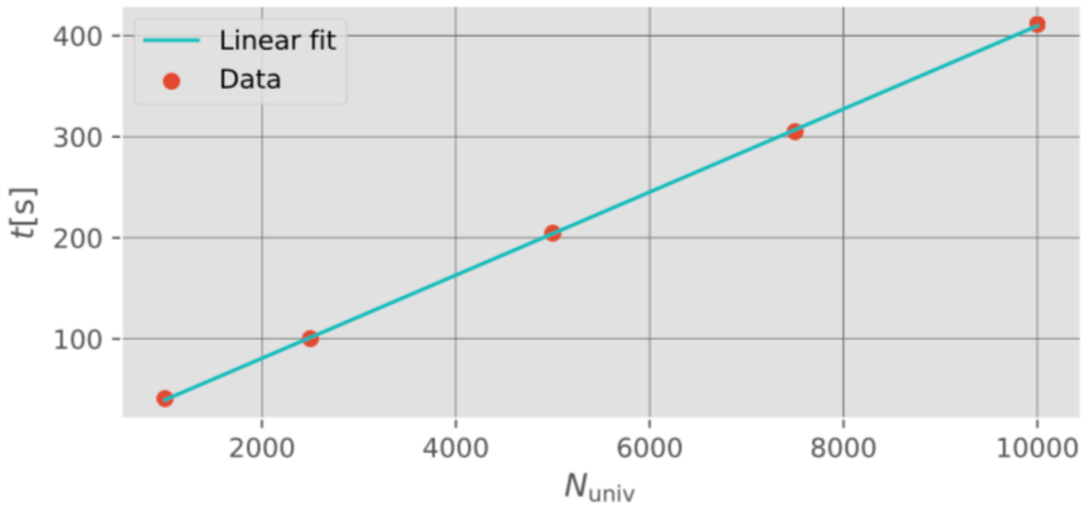


D. Cianci et al., **Phys. Rev . D** 96, 055001 (2017)

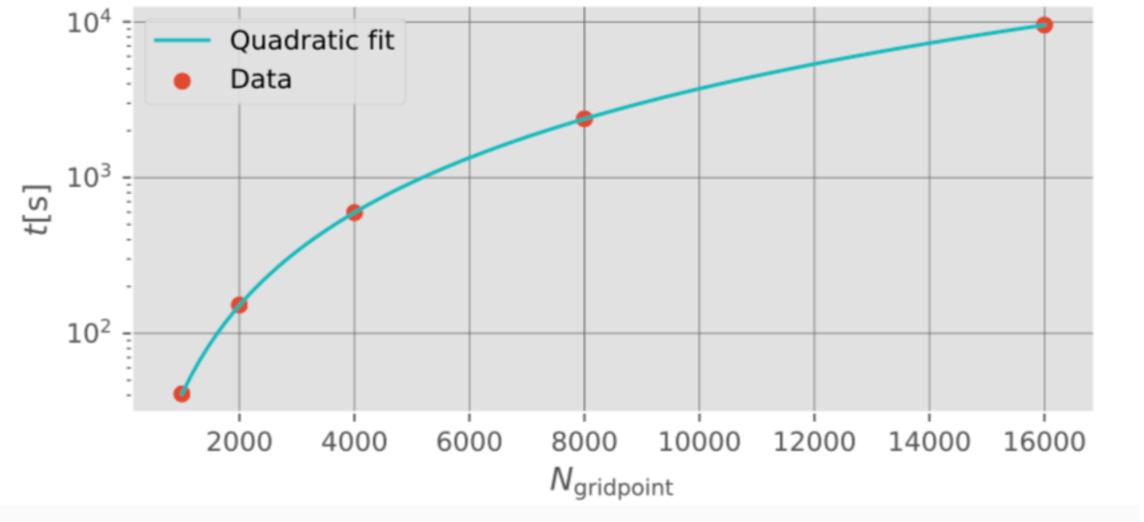
Improvements

- **Read the input files only once** and use collective MPI operations to copy data to all ranks trough memory.
- **Replace ROOT TH1D objects with types Eigen3** in matrix-vector multiplications and array computation of the χ^2
    - **A gain of factor of 2**
    - No longer memory-limited, only CPU-limited
- Use of **HDF5** as the **output format**
- Code is now **350x faster** when run on **single-core**

Scaling with N universes & N grid points

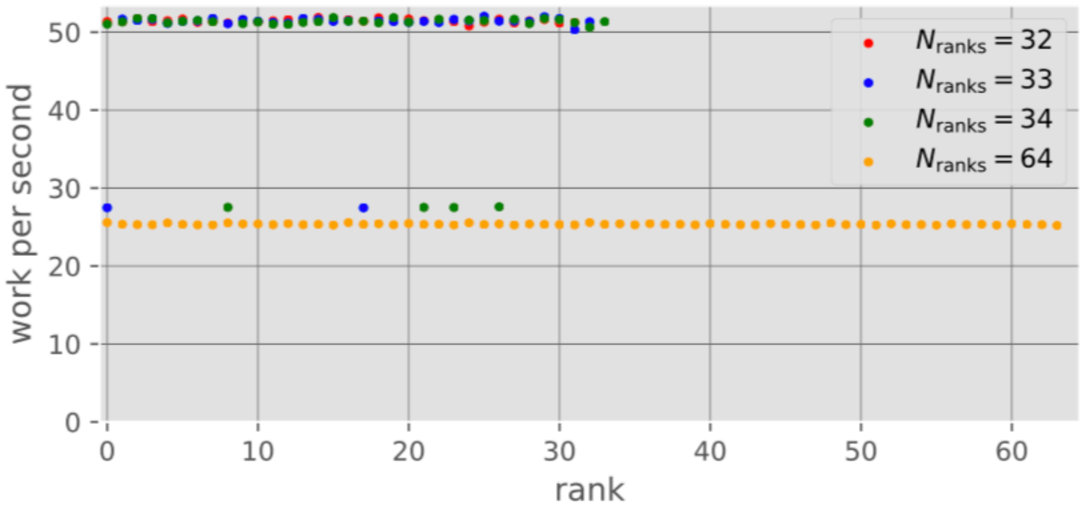


Scaling of the program run time with the number of universes, demonstrating a linear dependence on Nuniv

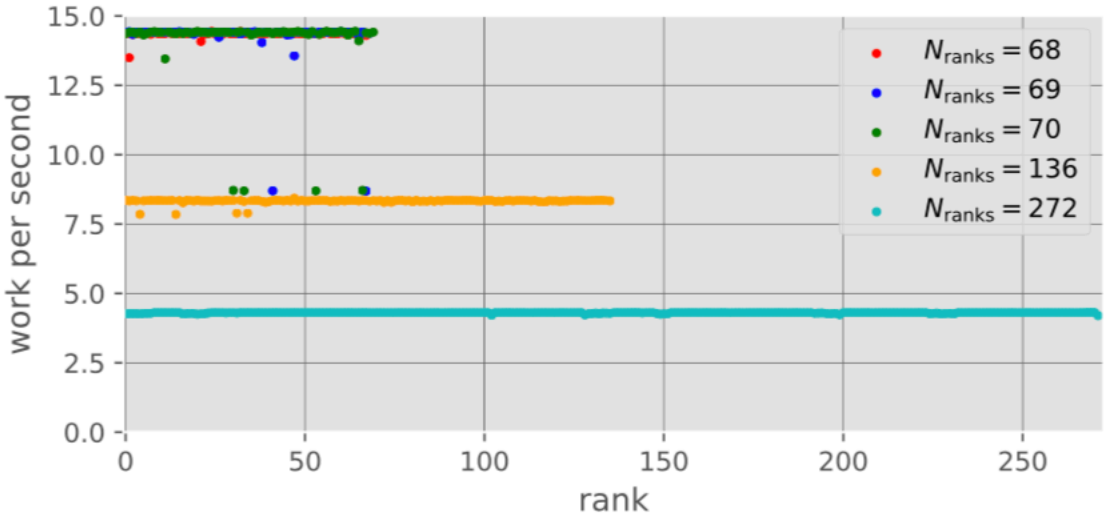


Scaling of the program run time with the number of grid points, demonstrating a quadratic dependence on Nuniv.

Scaling on single node

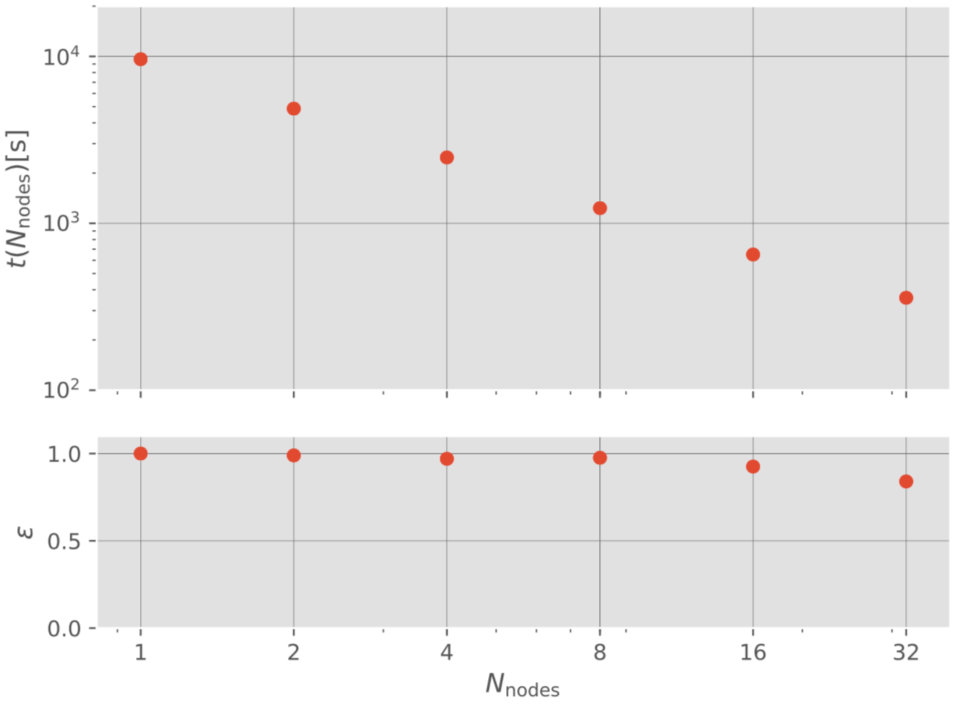


Measurement of the single Haswell node performance for a fixed problem size.



Measurement of the single KNL node performance for a fixed problem size

Scaling on Multi-node



Strong scaling measurements on Haswell nodes.
Generally good scaling up to the point where the amount of work per rank becomes relatively small.

	$N_{\text{univ}} = 10^4$ (3σ)	$N_{\text{univ}} = 10^8$ (5σ)
Cori phase 1 (Haswell)	7.2×10^5	7.2×10^3
Cori phase 2 (KNL)	1.2×10^6	1.2×10^4

Table 3: Upper boundaries on grid sizes that can be processed when running *a full day* on all of Cori phase 1/2.

Plan for Publication

- **Technical paper** that highlights the improvement gained from the parallelization and its scalability at HPC has been prepared.
- **Physics paper:**
  - Gauging interests from collaboration on a broader-scope SBN sensitivity/physics paper [in discussion], e.g.
    - Perform the non-FC method vs FC method. Compare the sensitivity outcomes for the two methods (slide 7).
    - Testing the effects of different parameterizations of systematics, study biases and model dependencies of the current analysis.

Summary & Outlook

- A lot of progress and improvements since summer 2019.
- Obtain parallelism through DIY
  - significant performance improvements by restructuring algorithms to use **Eigen3** for **matrix multiplications and array manipulations**
- Deliver the parallelization for the 3N+1 scenario
  - https://github.com/mayawos/whipping_star/tree/holger
- Find alternate approach to the grid scanning for higher dimensionality (7 parameters for 3N+2 → 3.2 mil grid points!).
- Look into replacing Eigen3 with kokkos and test the performance.
- Plan to incorporate nuisance parameters approach