NOvA Analysis report

01/14/20

Norm Buchanan, Steven Calvez, Pengfei Ding,
Derek Doyle, Andrew Norman, Alex Sousa

COLORADO STATE
UNIVERSITY

Confidence interval estimation

- **NOvA** follows the Feldman-Cousins unified approach for **confidence interval estimation**: requires **generating and fitting many pseudo-experiments**.
- Embarassingly parallel problem well suited for **HPC platforms**.
- First implementation in 2018 at **NERSC** sped up the time to results by a **factor >50**: 30 million hours consumed with 1.1 million concurrent threads.
- Second run at NERSC in 2019 with an improved framework using **DIY**: 50 million hours consumed with 100,000s DIY blocks.
- First implementation and run in 2019 at NERSC of a new fitting technique using a **covariance matrix for the sterile search**: fit 5 times faster than 3-flavor approach, 250k hours consumed.

Confidence interval estimation: present and future

- Paper draft being written.
- NOvA will once again use NERSC to perform the 3-flavor and 3+1-flavor analyses' confidence interval estimation in **a few months**.
- Working on **improvements** to optimize the resources consumption:
  - optimized/dynamic workload distribution with DIY
  - framework optimization: Holger and Derek already essentially sped the up the procedure by a factor 2
  - modernize the fitter
  - work with NERSC to improve the performance on KNLs
- **Higher dimensional fit**: 3D or even 4D confidence intervals.

Confidence interval estimation in DUNE

- **DUNE** adopted the same analysis framework (CAFAna) as NOvA and already used our **containers** to perform their **sensitivity studies** at NERSC: easier to port improvements and optimizations to DUNE framework.

Building unfolding matrices for cross-section analyses

- **Cross-section analyses** in NOvA require **building unfolding matrices** to extract true physics parameters from observed particle distributions.
- Apply **cuts** to many variations (or universe, i.e. a random set of systematic variations) of a nominal Monte-Carlo set, composed of a large number of files (70k files, for a total of 5TB).
- Embarassingly **parallel problem** with a high memory demand.
- Framework implemented at **NERSC**: 10 times more universes obtained 200 times faster (1.5 million NERSC hours).

Building unfolding matrices for cross-section analyses: prospects

- **First neutrino cross-section analysis** to leverage **HPC resources**.
- Many cross-section analyses in NOvA could benefit from this framework.
- Possible **improvements**:
  - various optimizations (memory consumption, etc.)
  - DIY implementation
  - parallel event selection with HDF and pandana
  - move away from ROOT (BDT application, TH2, etc.)

HDF5 and event selection

- **HDFMaker** used by NOvA to produce HDF5 files.
- HDF files greatly improve Convolutional Neural Network **training speed**.
- Use **HEPnOS** and **DIY** to greatly improve NOvA's event selection speed on HPC platforms and perform complete senstivity optimization studies.