NOvA Feldman-Cousins status

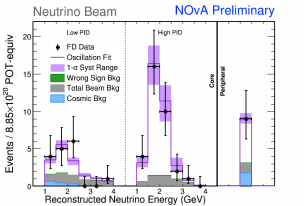
01/20/21

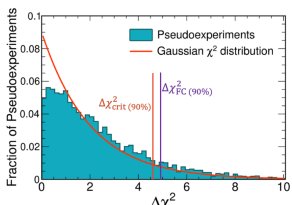
Norm Buchanan, Steven Calvez, Derek Doyle, Andrew Norman, Tom Peterka, Holger Schulz, Alex Sousa, Tarak Thakore

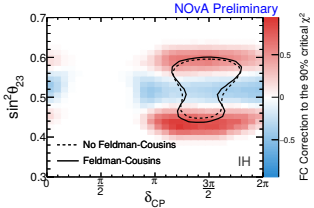
COLORADO STATE UNIVERSITY

Introduction

- Use **test statistic** $\Delta\chi^2$ to estimate the agreement between **data** and **predictions** made at a particular point of the **parameter space**.
- Because of **low statistics** and **physical boundaries**, we cannot derive a **significance** analytically (Wilks' theorem not satistified).
- Instead, build **empirical** $\Delta\chi^2$ **distributions** across the parameter space by generating and fitting **millions of pseudoexperiments**: Feldman-Cousins unified approach.







HPC implementation

- **Computationally intensive** Frequentist approach, well suited for **HPC**.
- Since the first **FC campaign at NERSC** in **2018**, we have been improving the framework **every year**.

2020 improvements: better DIY implementation (Tarak & Steven)

- The **number of pseudoexperiments** needed at each point of the parameter space depends on the significance: use Gaussian approximated significance to **estimate the number of pse** needed in each bin to reach a desired **statistical accuracy**.
- Modified the **DIY implementation**, from a fixed size 3D block space to a 1D block space with a custom number of blocks mapped to the parameter space.
- Translates to a **40% CPU hours saving**.

2020 improvements: standalone containers (Derek)

- Our **containers** are too large for NERSC's registry (close to 20GB) and we experienced bad performances when **reading** a few GB-sized input files from the burst buffer or the scratch area with $\mathbf{10^5}$ **ranks**.
- Solution was to make **squashfs archives** and submit a ticket to NERSC to turn them into regular **shifter images**.

2020 improvements: osc. calculators (Derek, Holger, Tarak & Steven)

- A significant fraction of the **compute time** is dedicated to **making predictions** and **solving oscillation probability equations**.
- Replaced ROOT with the linear algebra library **Eigen** for **faster predictions** and **oscillation probability calculations**.
- Implemented an **approximated oscillation calculator DMP**.

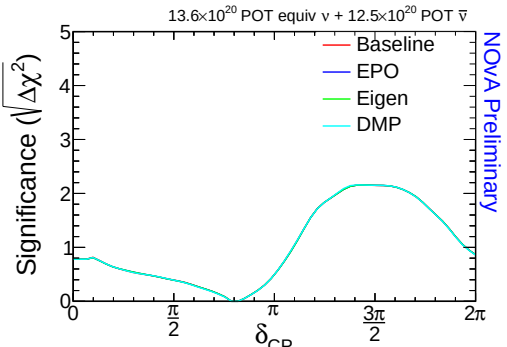
2020 improvements: osc. calculators (Derek, Holger, Tarak & Steven)

- Implemented a **reproducible seeds** scheme for **validation**.
- Ran **benchmark and validation jobs**:
  - baseline: ROOT based predictions and standard calculator
  - EPO: Eigen predictions only
  - Eigen: predictions and calculator
  - DMP: approximated calculator

	EPO	Eigen	DMP
Speed-up	13%	28%	28%

2020 improvements: osc. calculators (Derek, Holger, Tarak & Steven)

- Variety of **validation studies performed**.
- **No visible differences** in a slice.
- **Deviations with Eigen** were shown to be **smaller** than our current FC-related statistical uncertainty.
- For FC purpose, **accepted by NOvA**. Eigen is now default in CAFAna for NOvA and DUNE.



Conclusion

- Successful FC campaign at **NERSC** allowing **NOvA's latest results** to be shown this summer.
- Between the FC2020 campaign for the summer conferences, the validation studies, and the studies for the paper, the **Intensity Frontier allocation** has become the **8th largest consumer** of NERSC machine hours in 2020 (120M hours).

- Almost a **factor 2 improvement** compared to last year.
- A general **NOvA paper** about the **FC procedure** and a **technical paper** about the **HPC implementation** are in preparation.

- Join us in the **breakout sessions** to discuss ideas for further **improvements**:
  - MPI broadcasting with DIY to decrease memory usage and double the number of ranks per node.
  - OMP parallelization to fit one pseudoexperiment at different seeds at the same time.
  - Faster and more stable minimizer.

Conclusion

- Successful FC campaign at **NERSC** allowing **NOvA's latest results** to be shown this summer.
- Between the FC2020 campaign for the summer conferences, the validation studies, and the studies for the paper, the **Intensity Frontier allocation** has become the **8th largest consumer** of NERSC machine hours in 2020 (120M hours).

- Almost a **factor 2 improvement** compared to last year.
- A general **NOvA paper** about the **FC procedure** and a **technical paper** about the **HPC implementation** are in preparation.

- Join us in the **breakout sessions** to discuss ideas for further **improvements**:
  - MPI broadcasting with DIY to decrease memory usage and double the number of ranks per node.
  - OMP parallelization to fit one pseudoexperiment at different seeds at the same time.
  - Faster and more stable minimizer.

Conclusion

- Successful FC campaign at **NERSC** allowing **NOvA's latest results** to be shown this summer.
- Between the FC2020 campaign for the summer conferences, the validation studies, and the studies for the paper, the **Intensity Frontier allocation** has become the **8th largest consumer** of NERSC machine hours in 2020 (120M hours).

- Almost a **factor 2 improvement** compared to last year.
- A general **NOvA paper** about the **FC procedure** and a **technical paper** about the **HPC implementation** are in preparation.

- Join us in the **breakout sessions** to discuss ideas for further **improvements**:
  - MPI broadcasting with DIY to decrease memory usage and double the number of ranks per node.
  - OMP parallelization to fit one pseudoexperiment at different seeds at the same time.
  - Faster and more stable minimizer.