SLIDES-21-029-SCD

Fermilab U.S. DEPARTMENT OF ENERGY Office of Science

Grid-based minimization at scale: Feldman-Cousins corrections for light sterile neutrino search

Guangun Ge[(2)], Georgia Karagiorgi[(2)], Wesley Ketchum[(1)], James Kowalkskowski[(1)], Marc Paterno[(1)], Mark Ross-Lonergan[(2)], Holger Schulz[(3)], Saba Sehrish[(1)], **Marianette Wospakrik**[(1)]

[(1)]Fermi National Laboratory [(2)] Columbia University [(3)] University of Durham

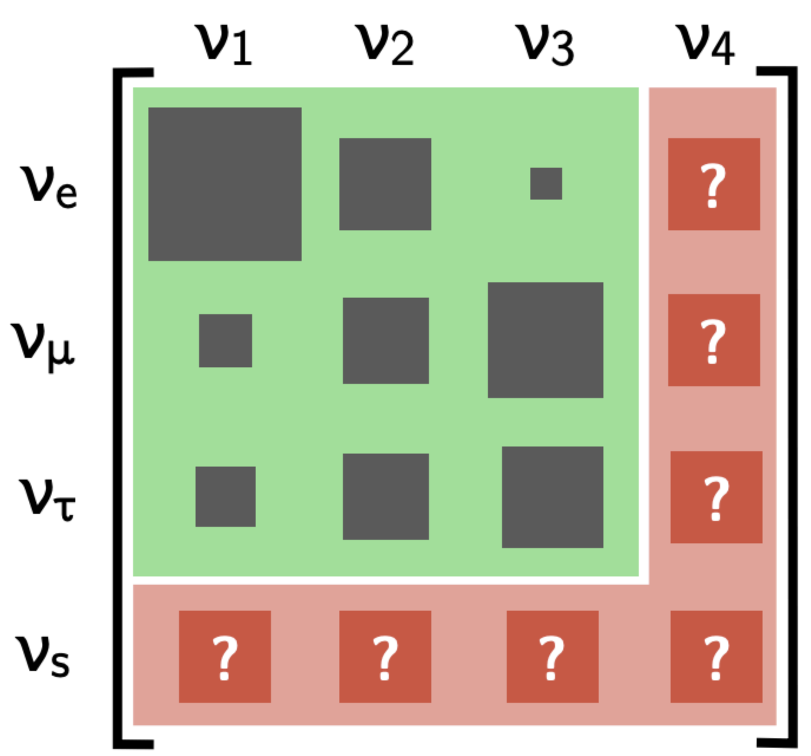
25th International Conference on Computing in High-Energy and Nuclear Physics

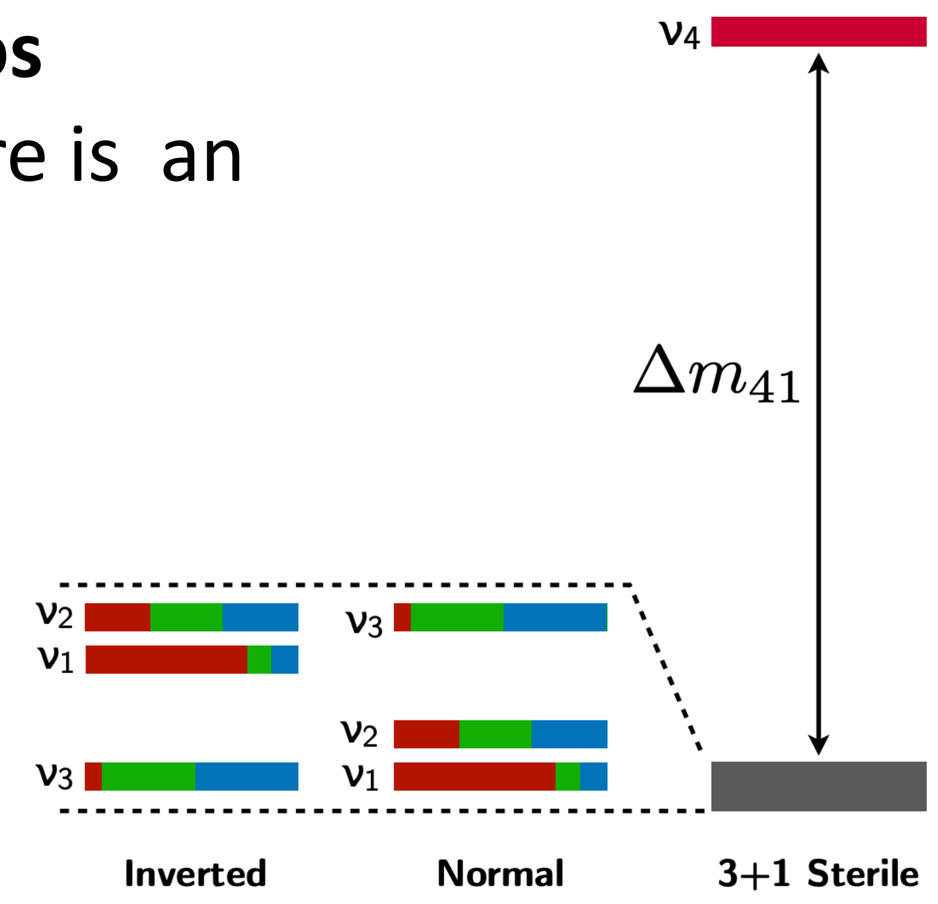
19 May 2021

Physics Problem: Light Sterile Neutrino Search

- Neutrino oscillation is a well-established phenomenon in particle physics
  - » can be used to probe the existence of additional neutrino species beyond the three known flavors, called **sterile neutrinos**
- Minimal extension to the Three-Neutrino Paradigm is if there is an additional sterile neutrino (3+1 paradigm)

$U =$





$$P_{\alpha\to\beta} = sin(2\theta)^2 sin\left(\frac{\Delta m^2 L}{4E}\right)^2$$

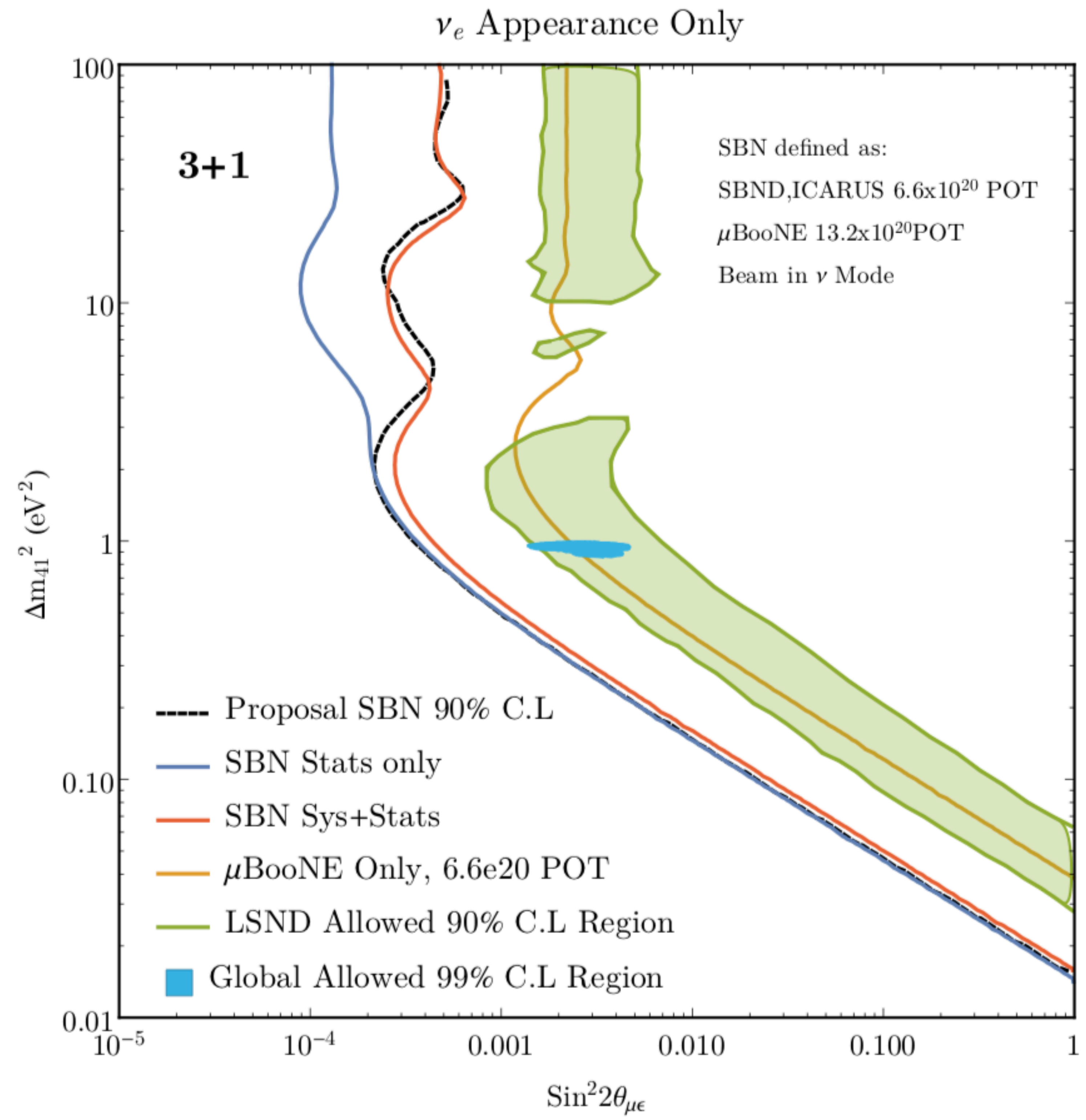
Δm^2: frequency of oscillation

$\sin^2(2\theta)$: amplitude of oscillation

Only depends on 2 parameters!

Oscillation probability via this mixing:

$$P_{\alpha\beta} = 4|U_{\alpha 4}|^2|U_{\beta 4}|^2 \sin^2\left(1.27\frac{\Delta m^2_{41} L}{E}\right)$$



Phys. Rev . D 96, 055001 (2017)

$\boxed{\frac{L}{E} \sim \frac{m}{MeV}}$: short-baseline

Fermilab

Statistical Coverage for Sterile Neutrino Oscillation Parameter

Compute a χ^2 surface in the (Δm^2_{41}, $\sin^2(2\theta)$) oscillation parameter plane:

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta) = \sum_{i,j} \left[N_i^{null} - N_i^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right] (E_{ij})^{-1} \left[N_j^{null} - N_j^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right]$$

Statistical Coverage for Neutrino Oscillation Parameter

Compute a χ^2 surface in the ($\Delta\, m^2_{41}$, $\sin^2(2\theta)$) oscillation parameter plane:

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta) = \sum_{i,j} \left[N_i^{null} - N_i^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right] (E_{ij})^{-1} \left[N_j^{null} - N_j^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right]$$

predicted number of events generated after oscillation

expected number of events assuming no oscillation (null hypothesis)

scaling

$$P_{\alpha\to\beta} = sin(2\theta)^2 sin\left(\frac{\Delta m^2 L}{4E}\right)^2$$

Statistical Coverage for Neutrino Oscillation Parameter

Compute a χ^2 surface in the ($\Delta\, m_{41}^2$, $\sin^2(2\theta)$) oscillation parameter plane:

A symmetric, positive definite covariance matrix describing correlated systematics

$$\chi^2(\Delta m_{41}^2, \sin^2 2\theta) = \sum_{i,j} [N_i^{null} - N_i^{osc}(\Delta m_{41}^2, \sin^2 2\theta)] (E_{ij})^{-1} [N_j^{null} - N_j^{osc}(\Delta m_{41}^2, \sin^2 2\theta)]$$

predicted number of events generated after oscillation

expected number of events assuming no oscillation (null hypothesis)

scaling

$$P_{\alpha\rightarrow\beta} = sin(2\theta)^2 sin\left(\frac{\Delta m^2 L}{4E}\right)^2$$

Fermilab

Statistical Coverage for Neutrino Oscillation Parameter

Compute a χ^2 surface in the ($\Delta\, m^2_{41}$, $\sin^2(2\theta)$) oscillation parameter plane:

A symmetric, positive definite covariance matrix describing correlated systematics

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta) = \sum_{i,j} \left[N_i^{null} - N_i^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right] (E_{ij})^{-1} \left[N_j^{null} - N_j^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right]$$

predicted number of events generated after oscillation

expected number of events assuming no oscillation (null hypothesis)

scaling

$$P_{\alpha\rightarrow\beta} = sin(2\theta)^2 sin\left(\frac{\Delta m^2 L}{4E}\right)^2$$

1. Test statistic to exclude other parameter values: $\Delta\chi^2 = \chi^2(\Delta m^2_{41}, sin^2\theta) - \chi^2(\Delta m^2_{41}, sin^2\theta)_{best}$

2. Once the $\Delta\chi^2$ distribution is known, we can calculate the p-values at each point in oscillation parameter plane to determine confidence intervals.

Statistical Coverage for Neutrino Oscillation Parameter

Compute a χ^2 surface in the ($\Delta\, m^2_{41}$, $\sin^2(2\theta)$) oscillation parameter plane:

A symmetric, positive definite covariance matrix describing correlated systematics

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta) = \sum_{i,j} [N_i^{null} - N_i^{osc}(\Delta m^2_{41}, \sin^2 2\theta)]\,(E_{ij})^{-1}\,[N_j^{null} - N_j^{osc}(\Delta m^2_{41}, \sin^2 2\theta)]$$

predicted number of events generated after oscillation

expected number of events assuming no oscillation (null hypothesis)

scaling

$$P_{\alpha\to\beta} = sin(2\theta)^2 sin\left(\frac{\Delta m^2 L}{4E}\right)^2$$

1. Test statistic to exclude other parameter values: $\Delta\chi^2 = \chi^2(\Delta m^2_{41}, sin^2\theta) - \chi^2(\Delta m^2_{41}, sin^2\theta)_{best}$

2. Once the $\Delta\chi^2$ distribution is known, we can calculate the p-values at each point in oscillation parameter plane to determine confidence intervals.

Wilk's theorem assume that $\sigma \approx (\Delta\chi^2)$ if (1) large statistics; (2) parameters are far from physical boundary

Fermilab

Statistical Coverage for Neutrino Oscillation Parameter

Compute a χ^2 surface in the ($\Delta\, m^2_{41}$, $\sin^2(2\theta)$) oscillation parameter plane:

A symmetric, positive definite covariance matrix describing correlated systematics

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta) = \sum_{i,j} \left[N_i^{null} - N_i^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right] (E_{ij})^{-1} \left[N_j^{null} - N_j^{osc}(\Delta m^2_{41}, \sin^2 2\theta)\right]$$

predicted number of events generated after oscillation

expected number of events assuming no oscillation (null hypothesis)

scaling

$$P_{\alpha\to\beta} = sin(2\theta)^2 sin\left(\frac{\Delta m^2 L}{4E}\right)^2$$

1. Test statistic to exclude other parameter values: $\Delta\chi^2 = \chi^2(\Delta m^2_{41}, sin^2\theta) - \chi^2(\Delta m^2_{41}, sin^2\theta)_{best}$

2. Once the $\Delta\chi^2$ distribution is known, we can calculate the p-values at each point in oscillation parameter plane to determine confidence intervals.

Wilk's theorem assume that $\sigma \approx (\Delta\chi^2)$ if (1) large statistics; (2) parameters are far from physical boundary

<u>**NOT SUFFICIENT FOR THE PROBLEM!**</u>

⇒ statistical inaccuracies and serious under-coverage for values of the oscillation parameters that are below the sensitivity of the experiment

Feldman-Cousins Procedure

Near Detector — Neutrino energy spectrum

Neutrino energy spectrum

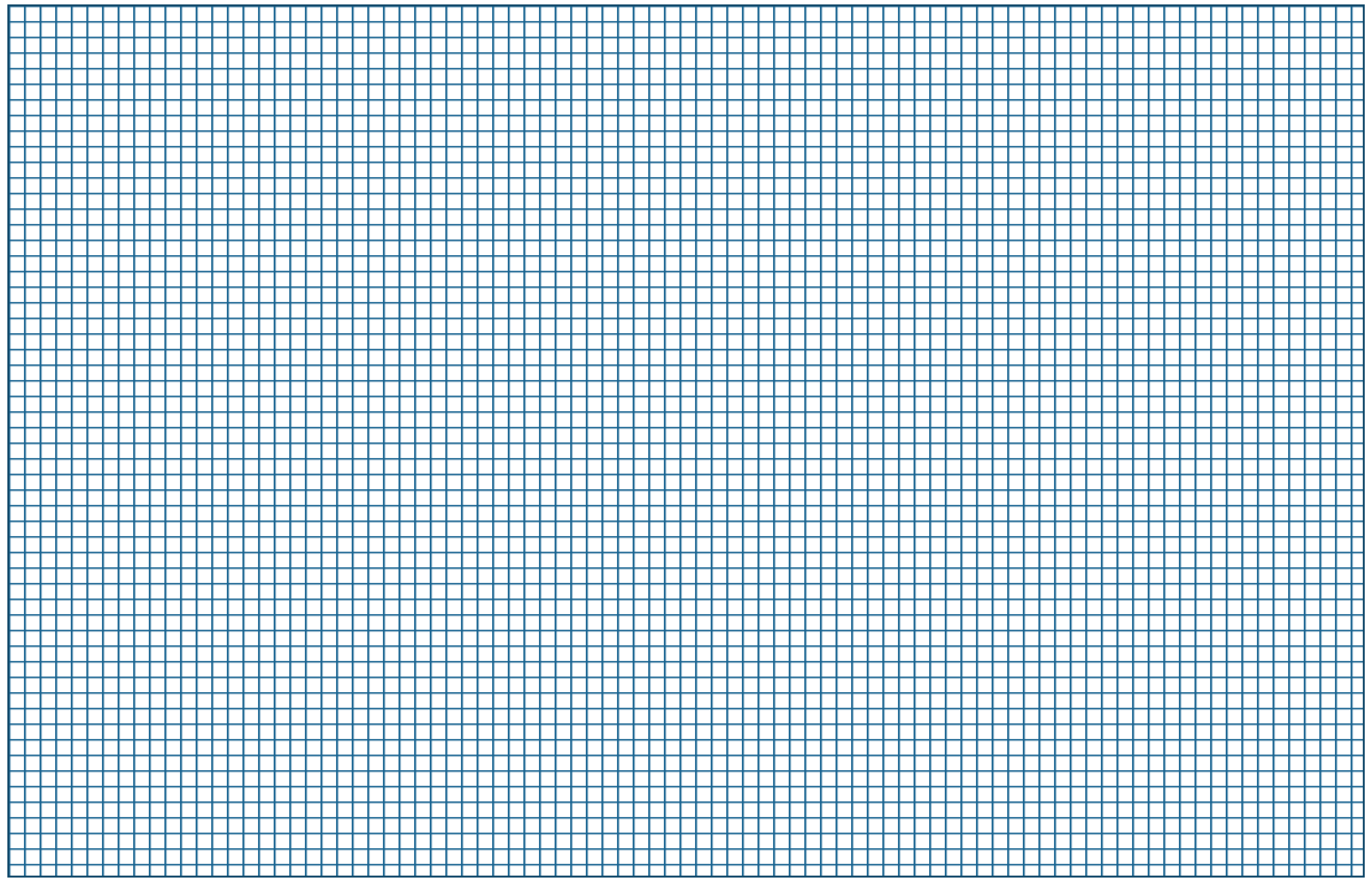
Far Detector — Neutrino energy spectrum

Covariance Matrix

Input: TH1D Histograms

Perform χ^2 calculation at each **oscillation parameter space point (Ngridpoint)** where we statistically fluctuate the oscillated prediction to generate **N number of toy experiments (Nuniv)** – Feldman Cousins procedure

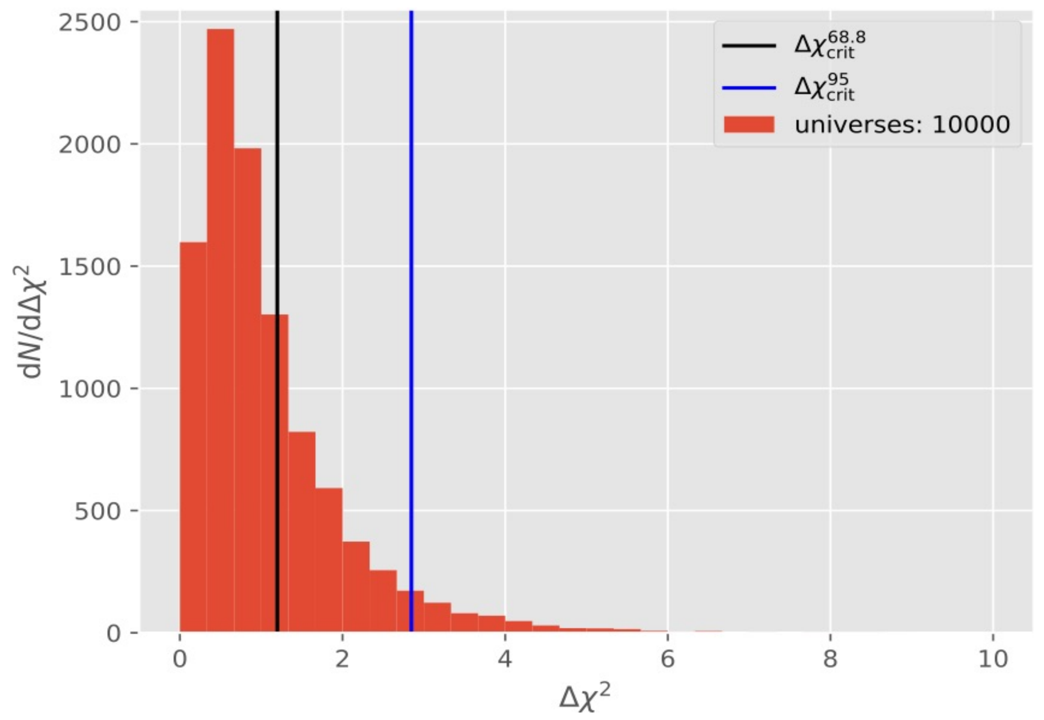
Δm^2



$\sin^2 2\theta$

Computationally expensive procedure
but
highly parallel problem!

Output

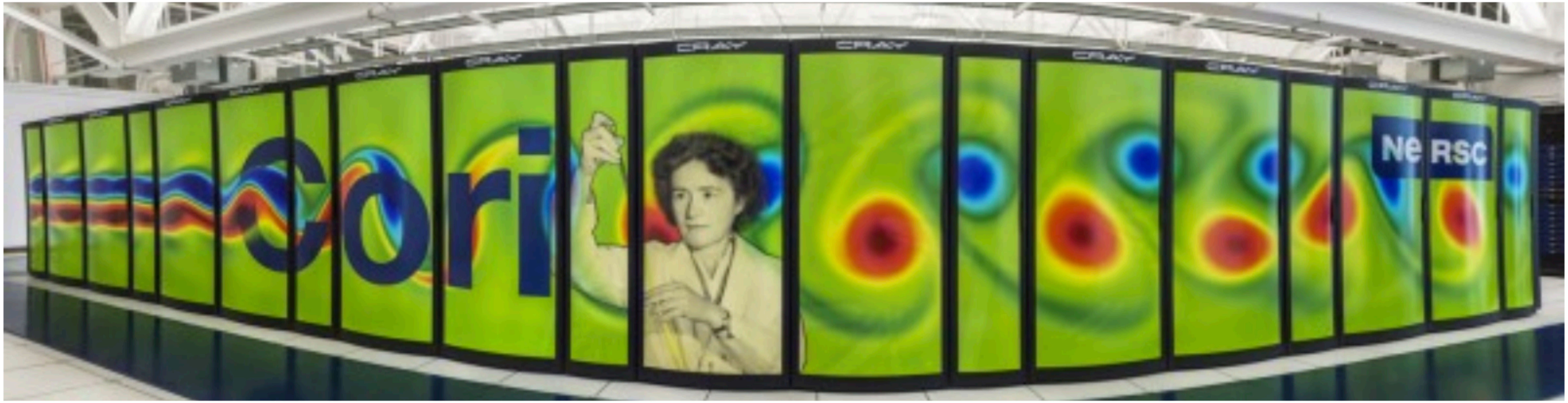


ROOT Ntuples

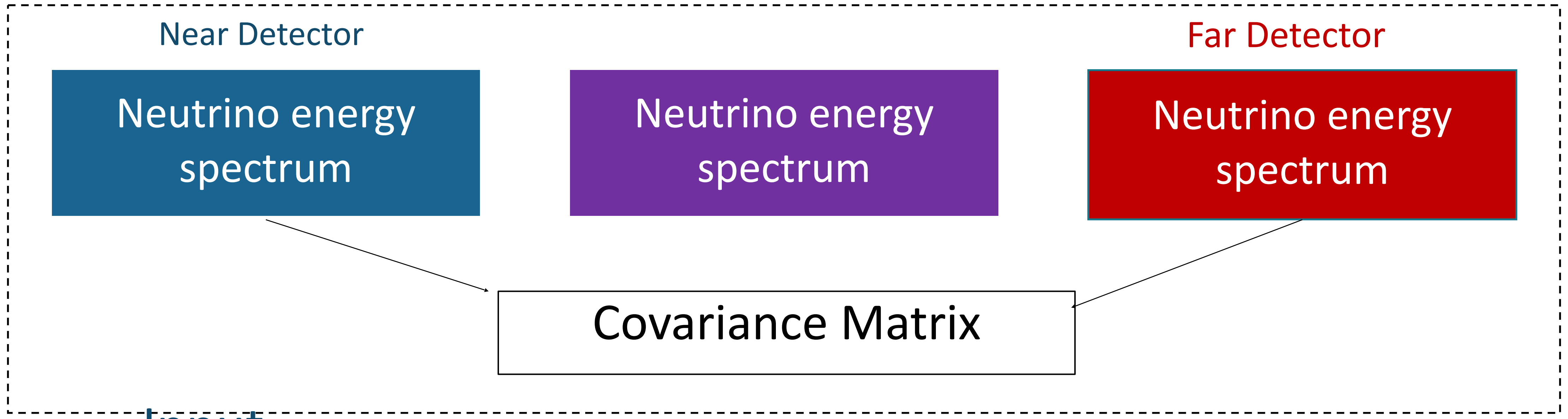
n-dimensional contour plot

NERSC - National Energy Research Scientific Computing
Lawrence Berkeley National Laboratory

Machine	Cori phase 1 (Haswell)	Cori phase 2 (KNL)
CPU	Intel Xeon E5-2698 v3	Intel Xeon Phi 7250
Clockspeed	2.3 GHz	1.4 GHz
Cores per node	32	68
Ranks per core	2	4



Accelerating Feldman Cousins Procedure on HPC

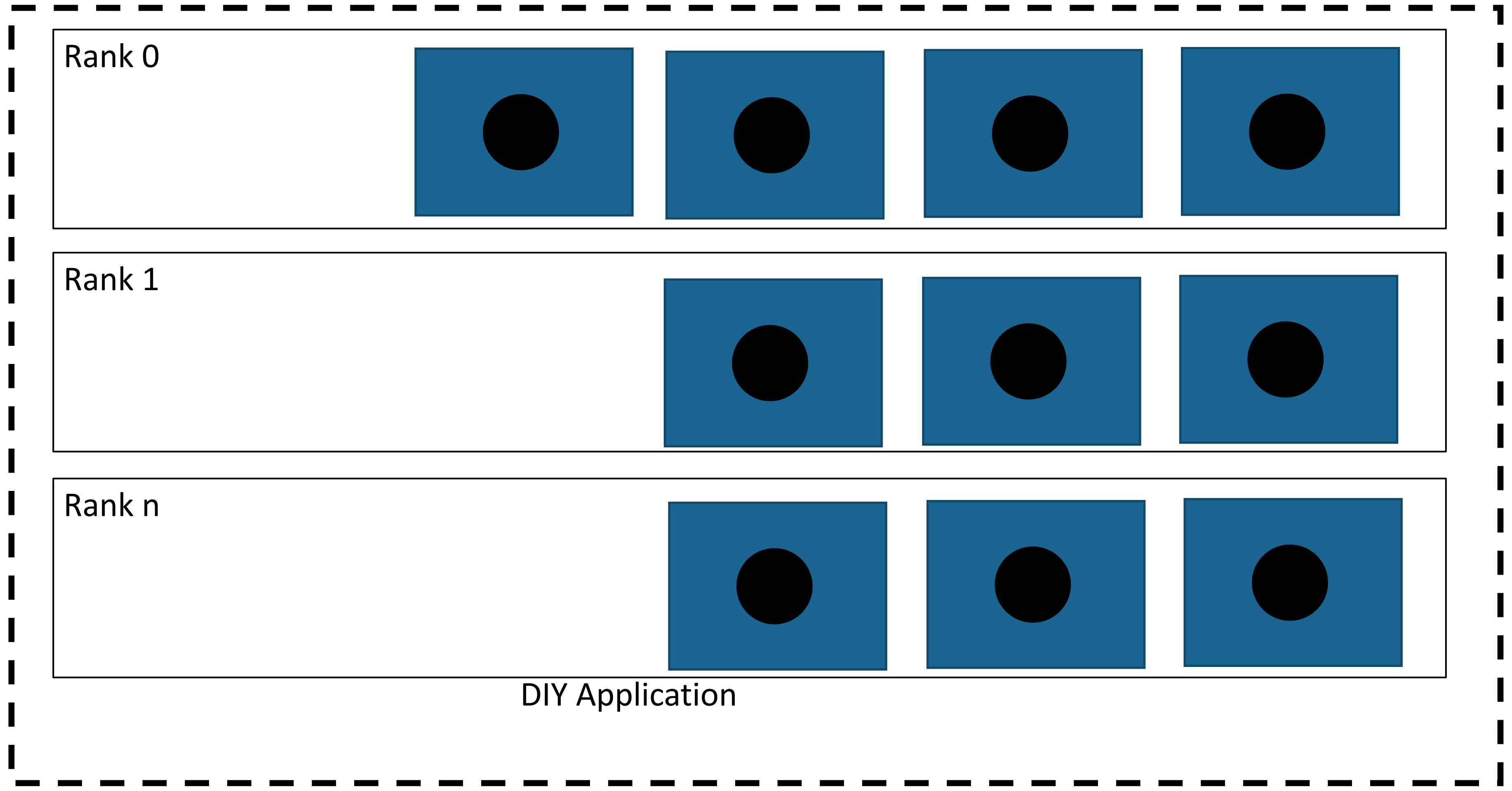


Input

restructured the codes to read the input files only once using MPI

MPI to distribute the work load to all ranks through memory

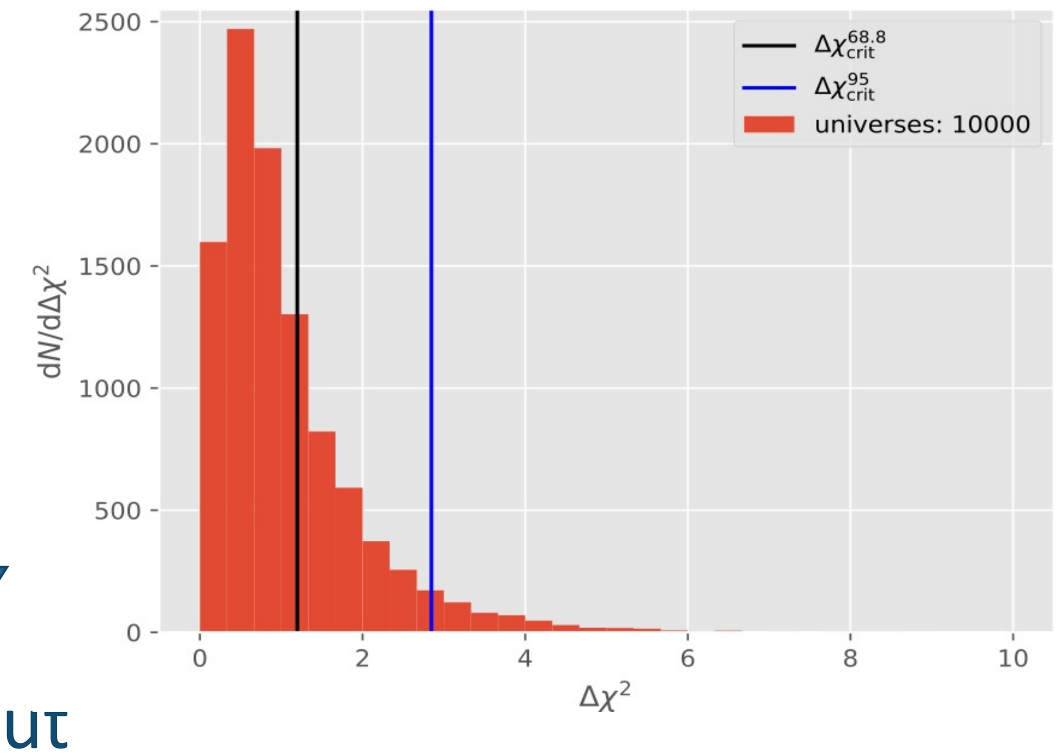
*MPI: Message Passing Interface



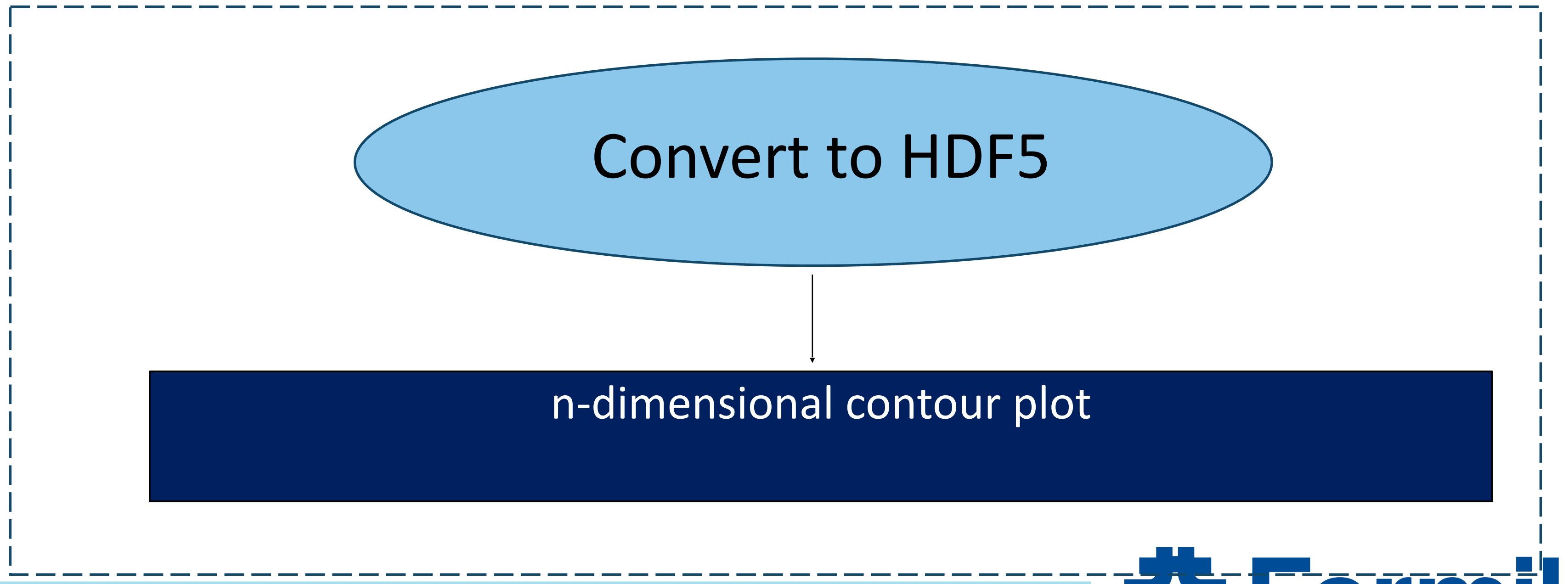
replace ROOT histogram objects with types Eigen3 in linear algebra and vectorized calculation to compute χ^2

» Code is no longer memory-limited, only CPU-limited, allowing up to 350x speedups from original implementation

Perform χ^2 calculation at each **oscillation parameter space point (Ngridpoint) per toy experiments (Nuniv) per rank in a node** – Feldman Cousins procedure



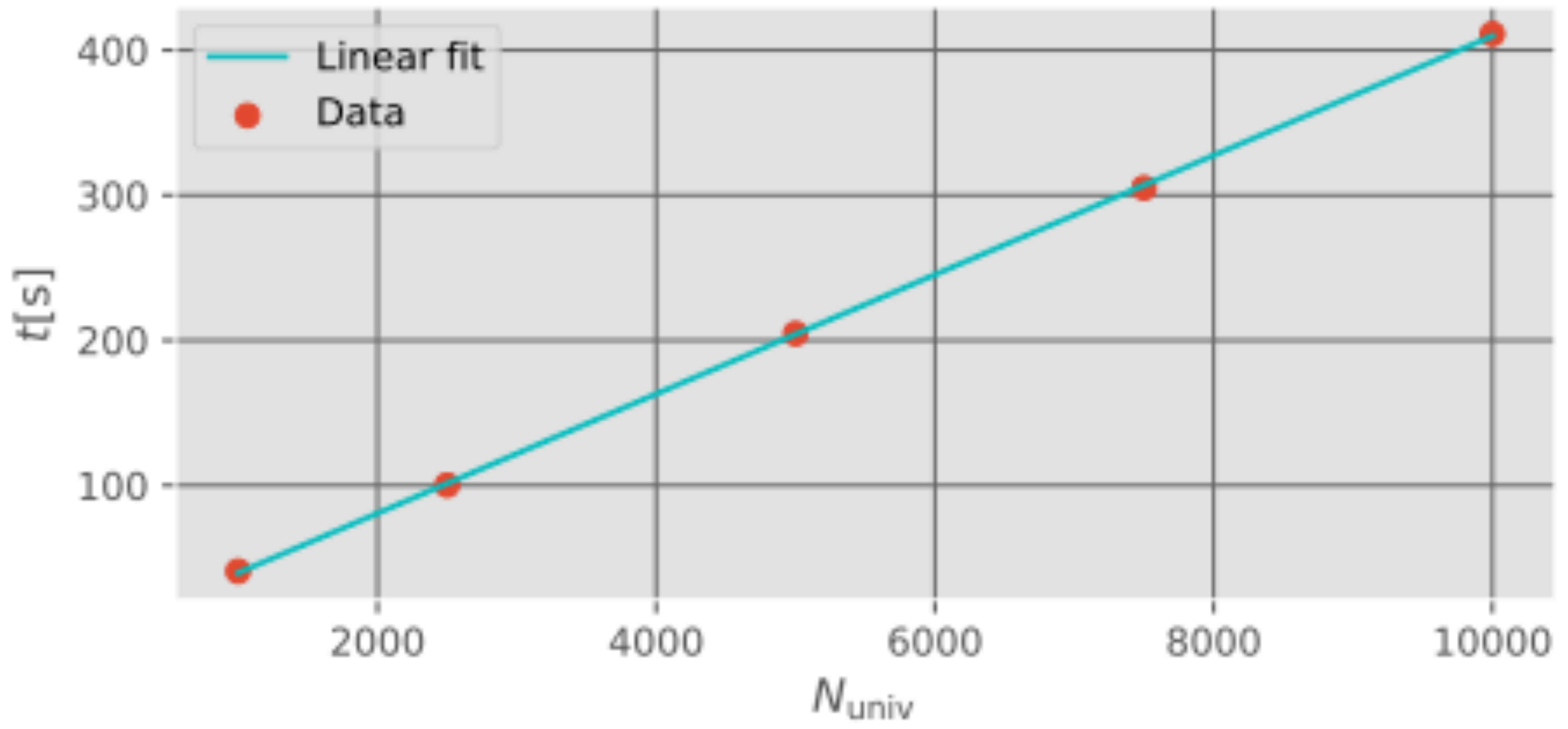
Output

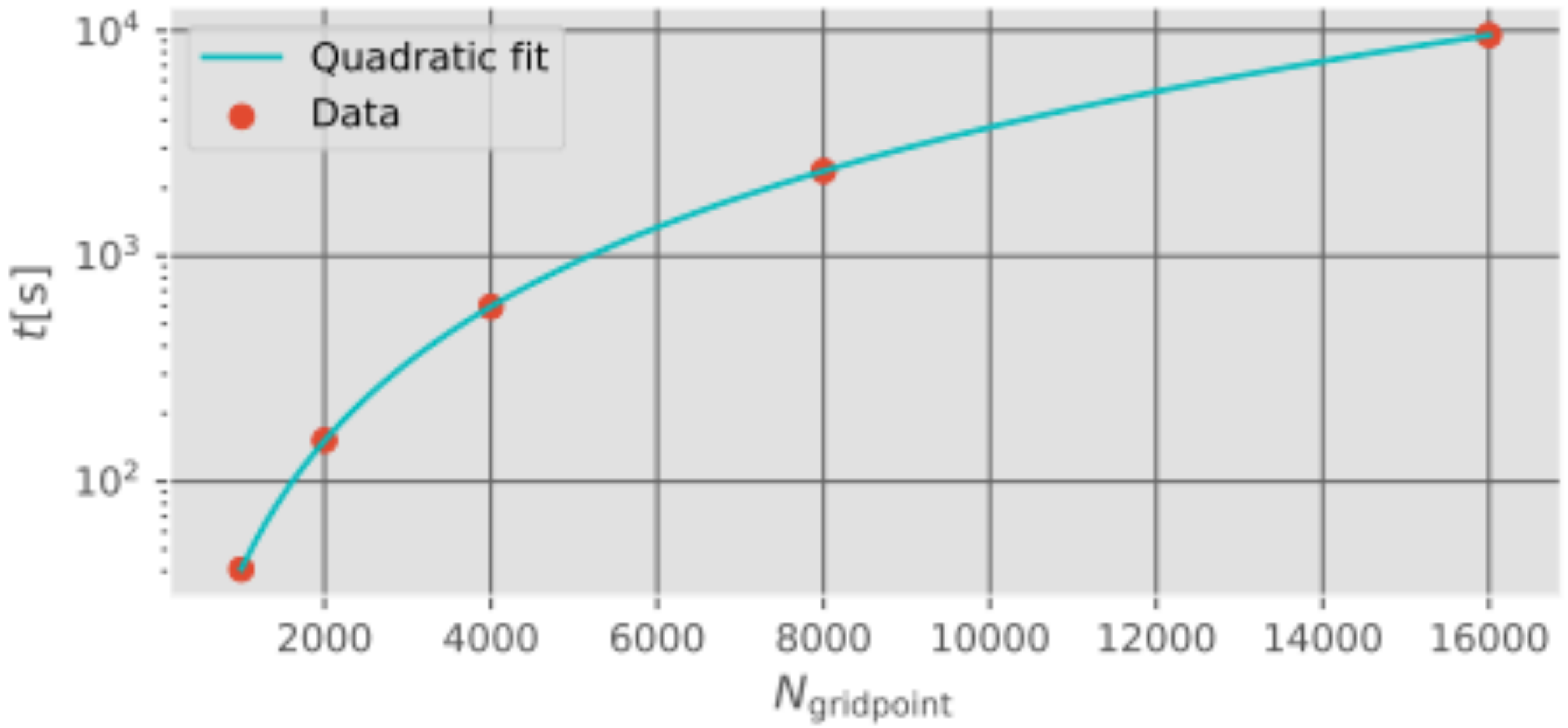


Fermilab

Performance and Scaling

Scalability is important for parallel computing to be efficient





- Fixed number of **Haswell nodes** (and ranks)
- Fix ***Ngridpoint*** and varied ***Nuniv*** independently
- The data is fit **linearly** and show no visible deviation from linear scaling

- Fix the ***Nuniv*** and compute the program execution time as function of ***Ngridpoint***
- The data is fit **quadratically** and shows that there is no visible deviation from the quadratic scaling

Fermilab

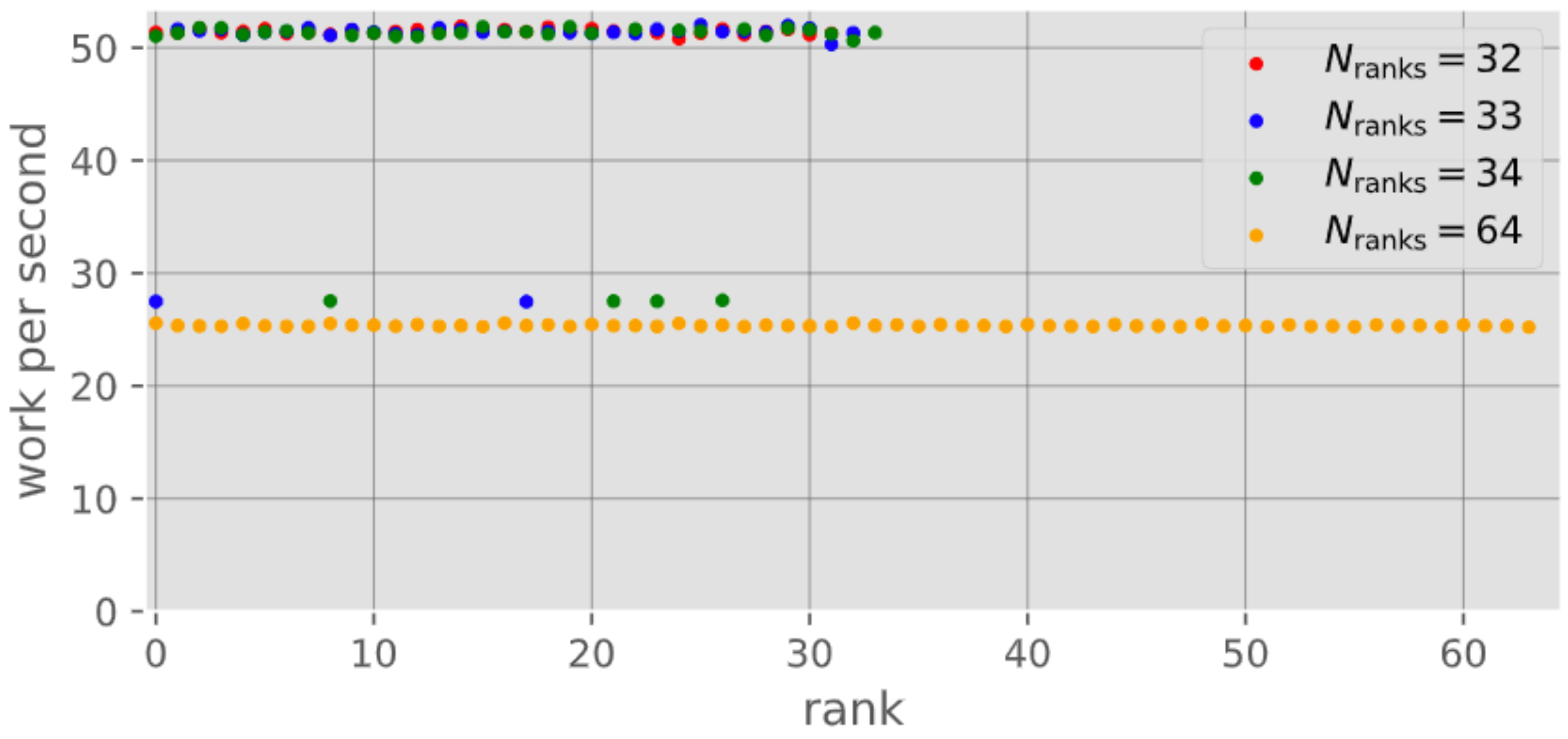
Performance and Scaling

Single-node Scaling

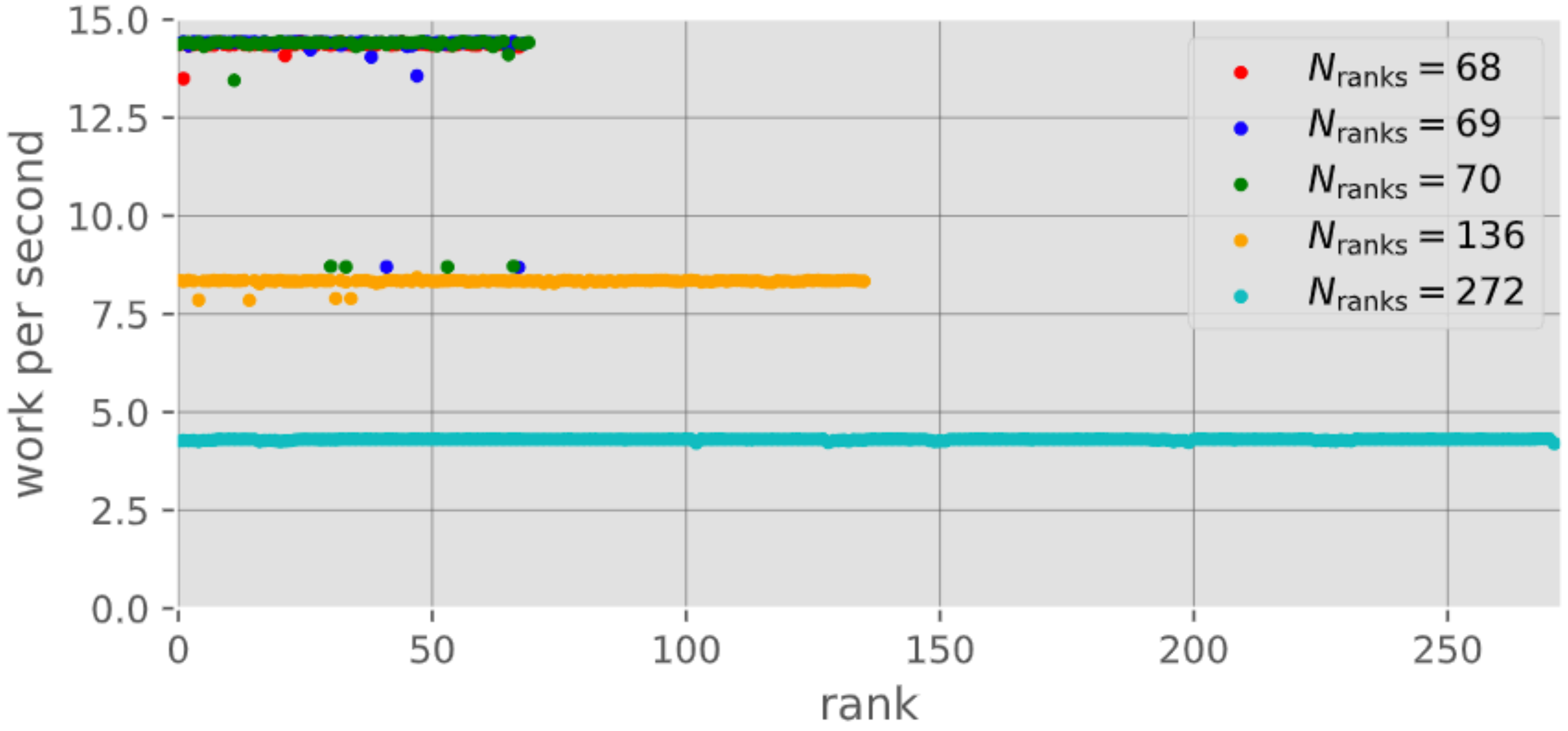
- **Goal**: assess benefits gained from using multiple MPI ranks:
  - Select a fixed size problem as function of ***Ngridpoint*** and ***Nuniv***
  - Distribute work among the ranks as evenly as possible
  - Record time spent by the slowest rank (tmax) in the main part of the program for each rank separately.
- **Haswell** attains the best single node scaling when using 32 ranks.
  - No significant improvement in execution time when multi-threading is enabled
  - Oversubscribing 1 and 2 cores in the 33 ranks and 34 ranks respectively, leads to overall worse performance
- Similar performance on **KNL** when using non-integer multiples of the number of ranks (68) but improvements at higher number of ranks

Ranks	68	136	272
t_{max}[s]	87	75	70
Gain w.r.t. 68 ranks		16%	24%

Haswell Node



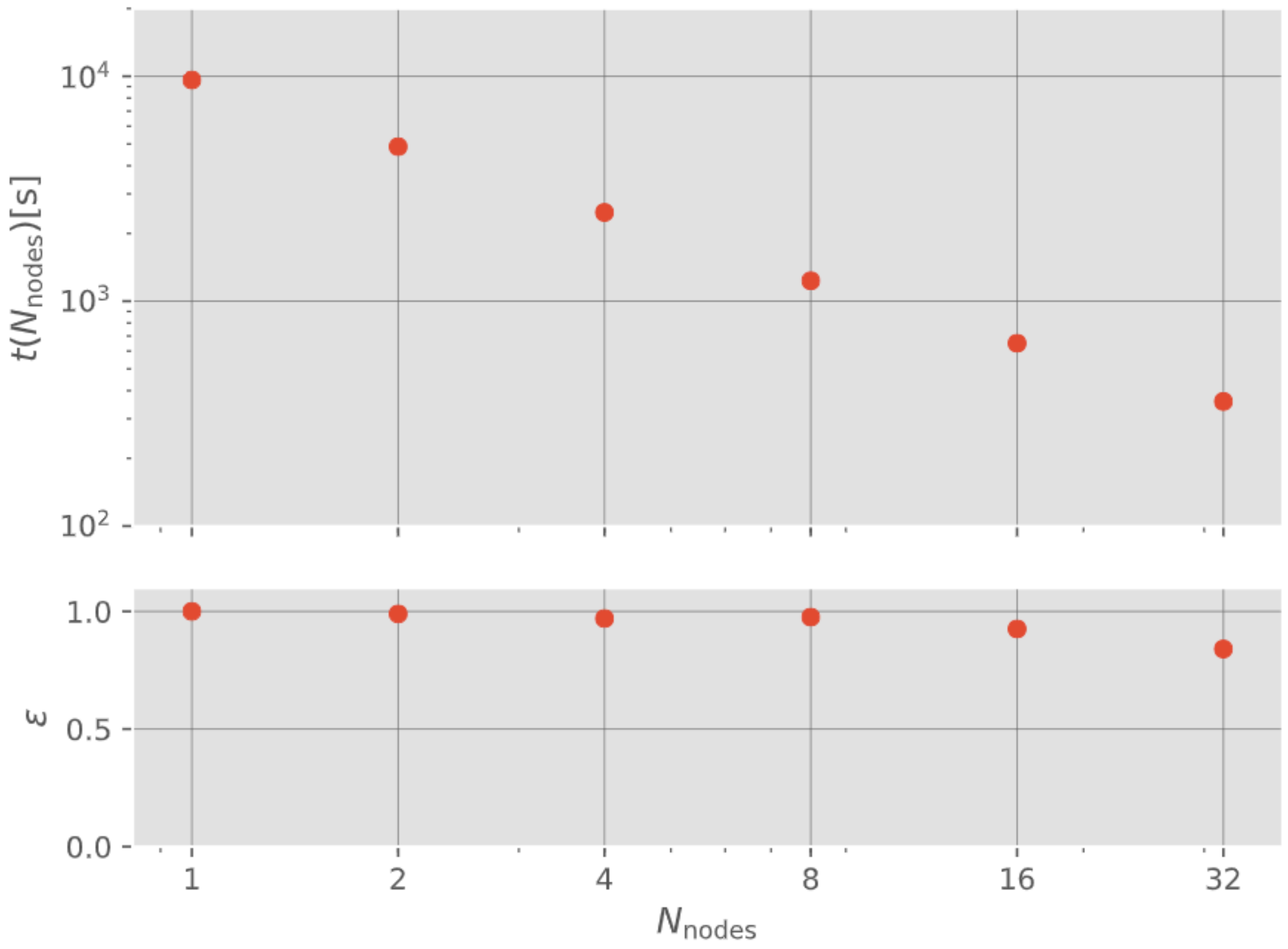
KNL Node



Performance and Scaling
Multi-node Scaling

- Understand scalability of multi node systems to gauge potential performance gain as more computing resources are added
  - Define reasonable large problem of fixed size.
  - Measure the time it takes to complete the main part of the program as a function of the number of used **Haswell nodes** (N_{nodes}).
  - Efficiency of the program execution:

$$\varepsilon = N_{\text{nodes}} \cdot \frac{t(N_{\text{nodes}})}{t(1)},$$



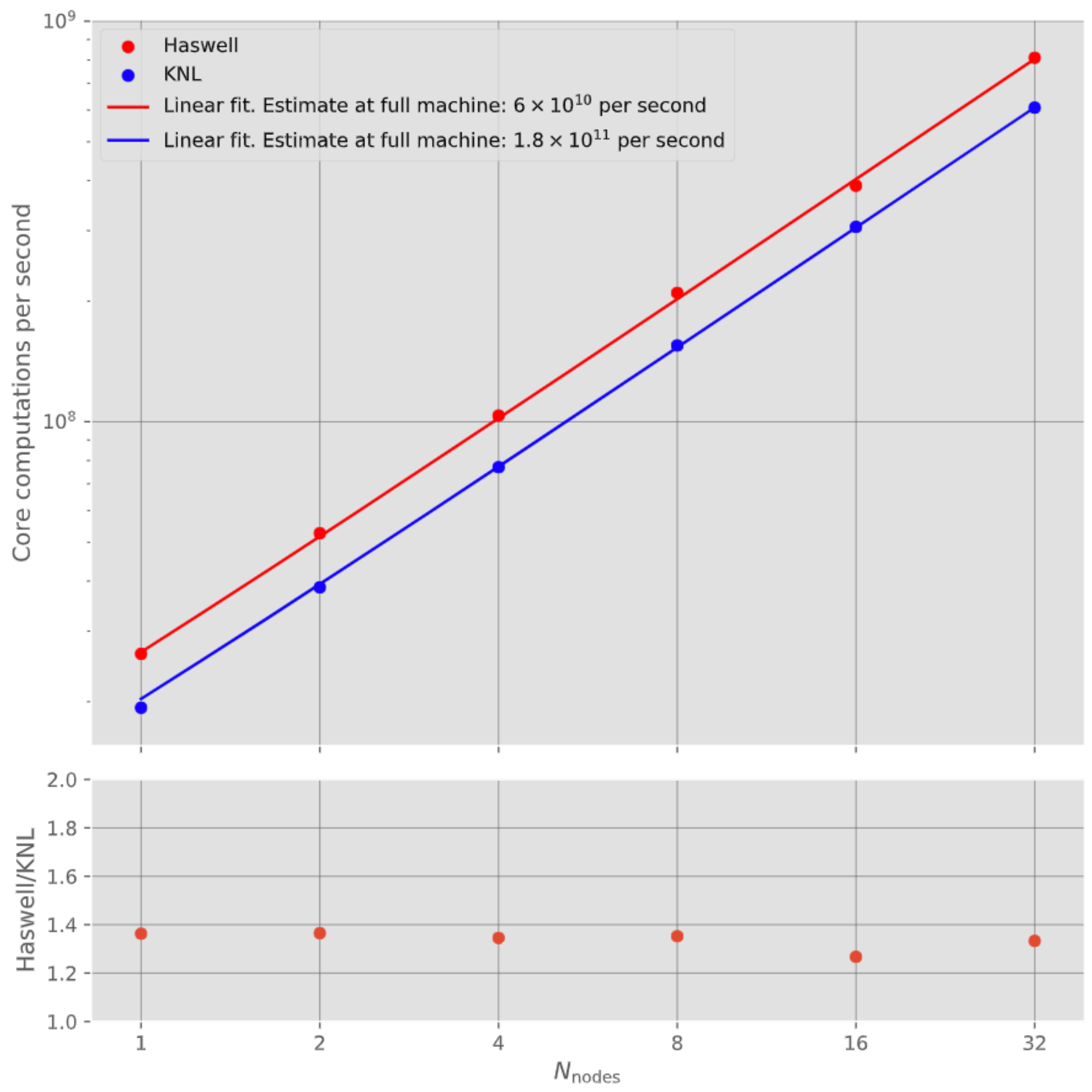
32 ranks per node used in this study

Fermilab

Performance and Scaling

Estimation of what is possible

- Estimate an upper limit of of core computations if the whole machine were available:
  - **Cori phase 1 (Haswell):** estimate an upper boundary of $6\text{x}10^{10}$ s^{-1} from a linear extrapolation to the entirety of 2388 nodes.
  - **Cori phase 2 (KNL):** estimate an upper boundary of $1.8\text{x}10^{11}$ s^{-1} when using all 9688 nodes of Cori phase 2 (KNL).
- The linear scaling assumption is optimistic. Should be interpreted as an upper limit of the performance of the program.



Fermilab

Conclusions

- Low statistics and physical boundaries require an empirical approach to calculating a significance
- Adapted a grid-search-based fitting application that calculates Feldman-Cousins corrections to run efficiently on current state-of-the-art processor architectures and systems available at HPC facilities (NERSC computing center).
- Transformed a memory limited serial-execution program into an MPI parallel application that scales up to available compute power of a computing facility.
- Achieved node-level and thread-level parallelism through DIY, and accomplish significant performance improvements by restructuring algorithms to use Eigen3 for matrix multiplications and array manipulations.
- This work was performed in the context of a specific application (SBNFit) designed for short-baseline neutrino oscillation experiments, however, the techniques employed can provide similar benefits in terms of performance and design for broader neutrino oscillation physics program.
- <u>**What's next?**</u>
    - Explore alternative techniques to grid scan procedure, such as utilizing optimizers and approximating discrete binned data into to a continuous function using the Multivariate Functional Approximation (MFA) model (doi: 10.1109/LDAV.2018.8739195.) as well as exploring different linear algebra libraries, such as Kokkos.
        - » Allows to move towards probing higher dimensions in oscillation parameter