High-Throughput Data Processing at FRIB Using ESnet

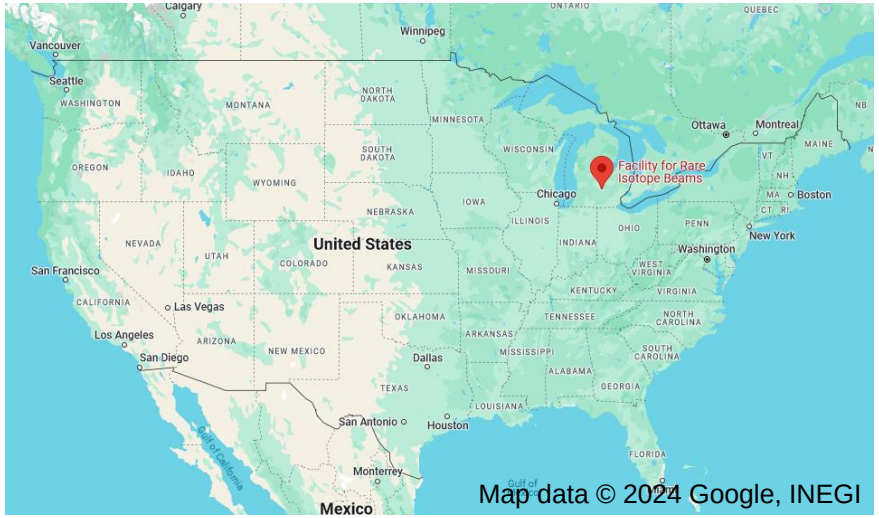
Aaron Chester
Data Acquisition Scientist
chester@frib.msu.edu

25 April 2024

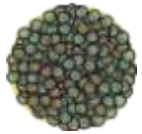
This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics and used resources of the Facility for Rare Isotope Beams (FRIB) Operations, which is a DOE Office of Science User Facility under Award Number DE-SC0023633.

The Facility for Rare Isotope Beams (FRIB)

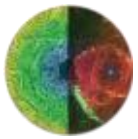
- FRIB is a scientific user facility funded by the US Department of Energy Office of Science (DOE-SC), Michigan State University (MSU), and the State of Michigan
  - Open to researchers from around the world
  - User Organization: 1800 members (125 colleges and universities, 13 national labs, 53 countries)
  - Experimental program began in May, 2022
- Key feature of FRIB: high-power LINAC, 400 kW at 200 MeV/u for ^{238}U
  - Beams from oxygen to uranium
- FRIB also provides stopped and low-energy re-accelerated beams (< 12 MeV/u)
- **Access to atomic nuclei across the nuclear chart at a range of energies**



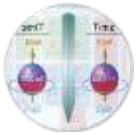
The FRIB Experimental Program



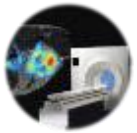
- Nuclear structure
  - How does subatomic matter arrange itself and how does it evolve?
  - What combinations of neutrons and protons form bound atomic nuclei?



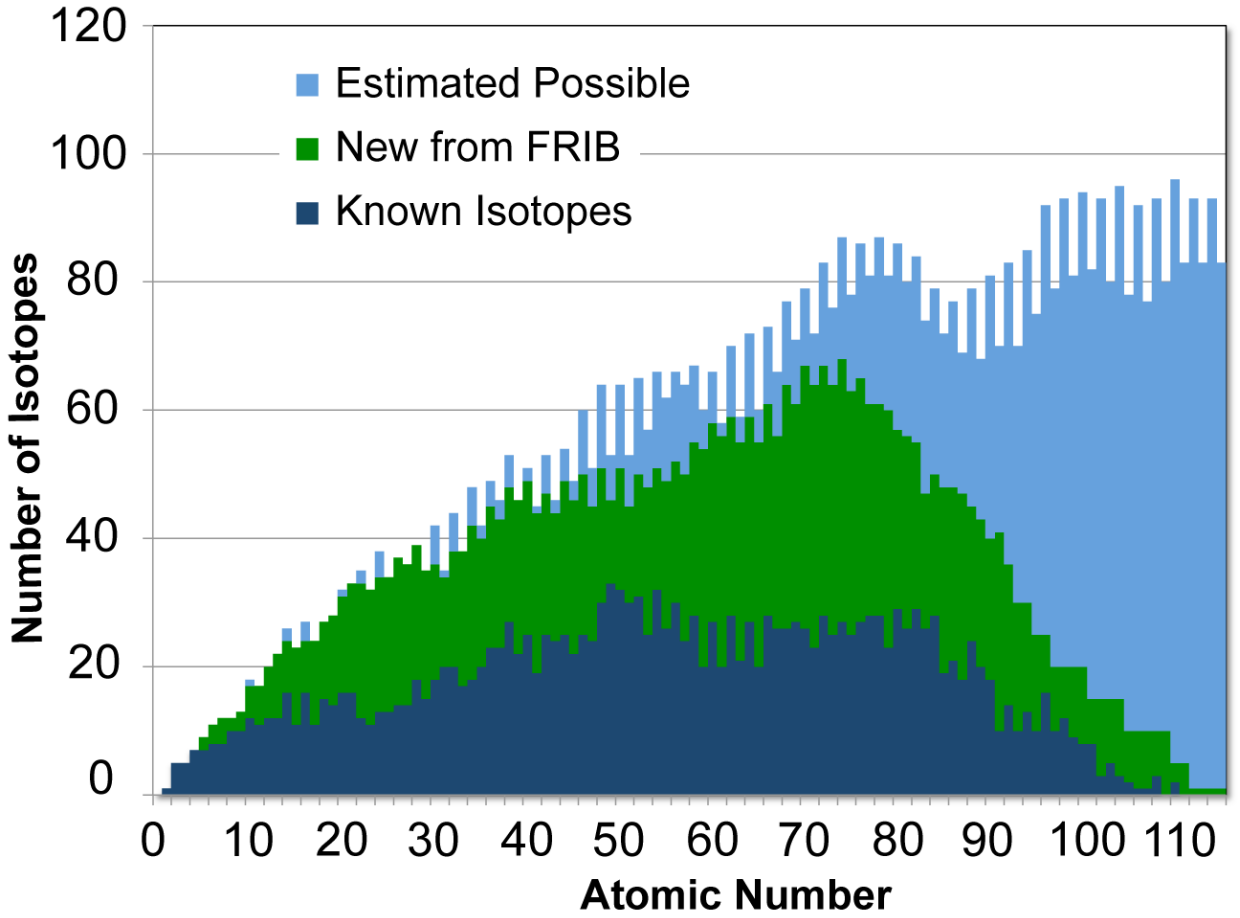
- Nuclear astrophysics
  - How are the chemical elements created?
  - What is the nature of matter at extreme temperatures and densities?



- Fundamental symmetries
  - Why is there more matter than antimatter?
  - Are neutrinos their own antiparticles?



- Societal applications and benefits
  - Medicine, energy, material sciences, environment, workforce development, etc.

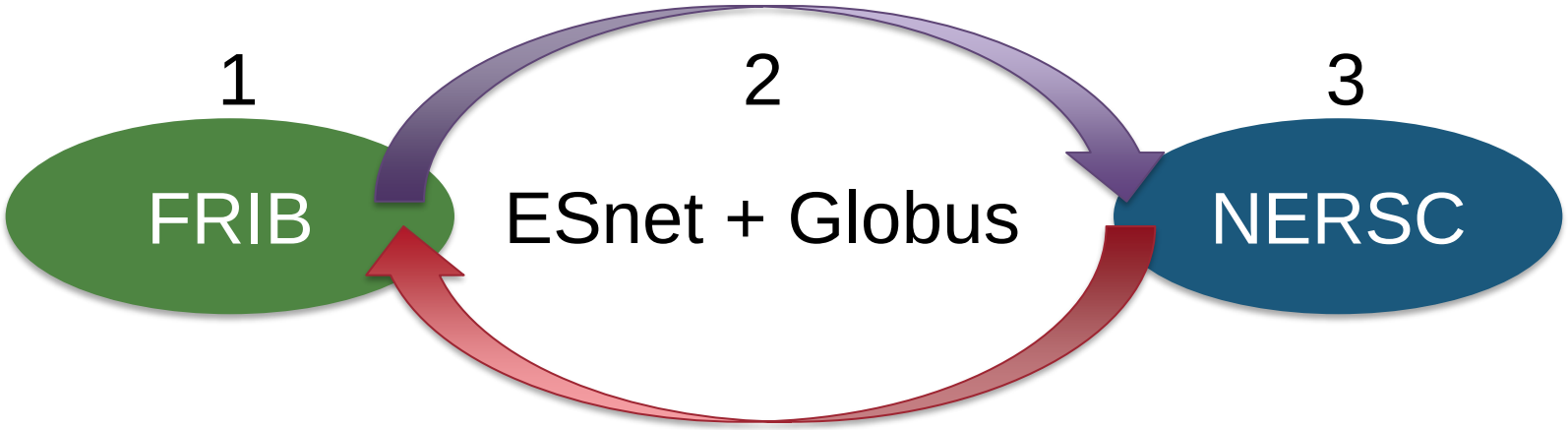


Erler et al., Nature **486** (2012) 509-512.

How to Make the FRIB Experimental Program a Success?

- Detector development: higher rate, improved resolution, …
  - Adoption of new technologies
- Upgrade and develop new data acquisition (DAQ) systems
  - Extensibility, flexibility, portability (containers)
  - Library of solutions for different detector and hardware types
  - Contributions from users (GitHub)
  - Collaboration with other laboratories
- Adopt new computing solutions to reduce time to enable discoveries and improve decision-making during experiments
  - Local computing cluster
  - High-speed network connecting local storage to high-performance computing (HPC) centers

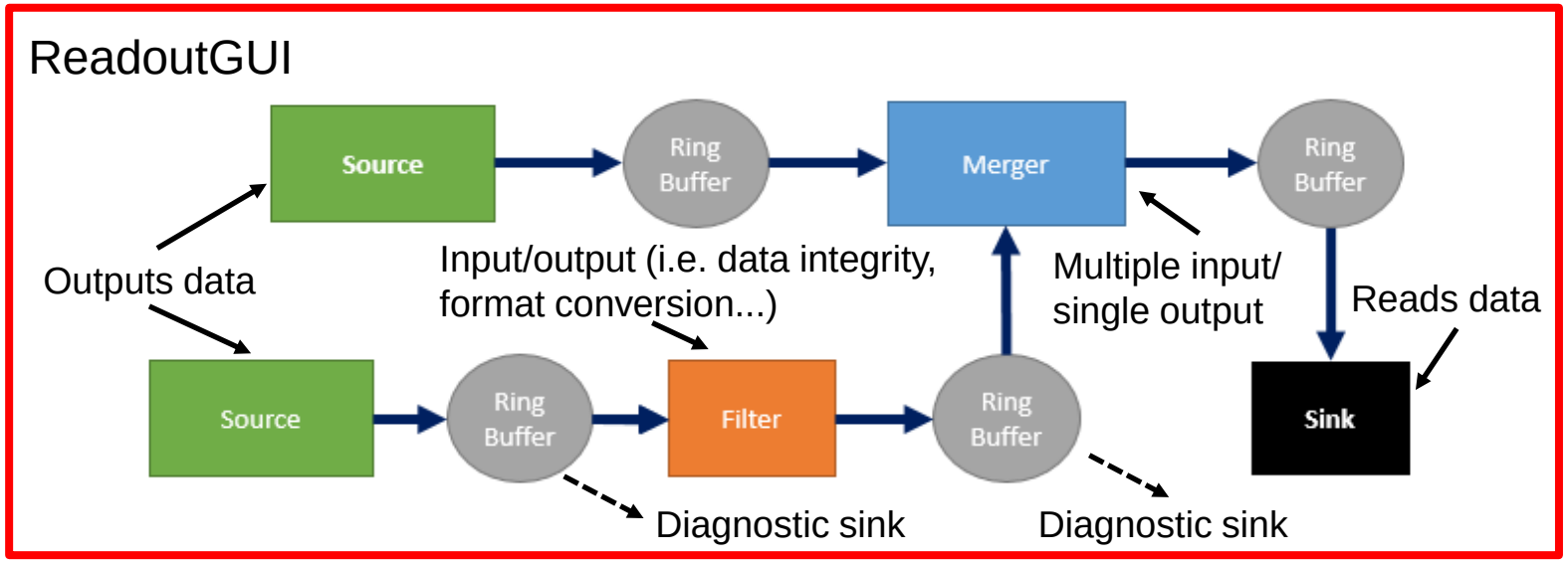
A High-Throughput Data Processing Pipeline for FRIB



- Leverage DOE's Energy Sciences Network (ESnet) for high-speed data transfer and the National Energy Research Scientific Computing Center's (NERSC) Perlmutter supercomputer in Berkeley, CA for data processing
- Components of the system
  1. Data acquisition
  2. Data transfer and pipeline management: ESnet and Globus Flows
  3. Parallel processing at NERSC (or other HPC facility)
- Experiment-motivated development starting in 2019

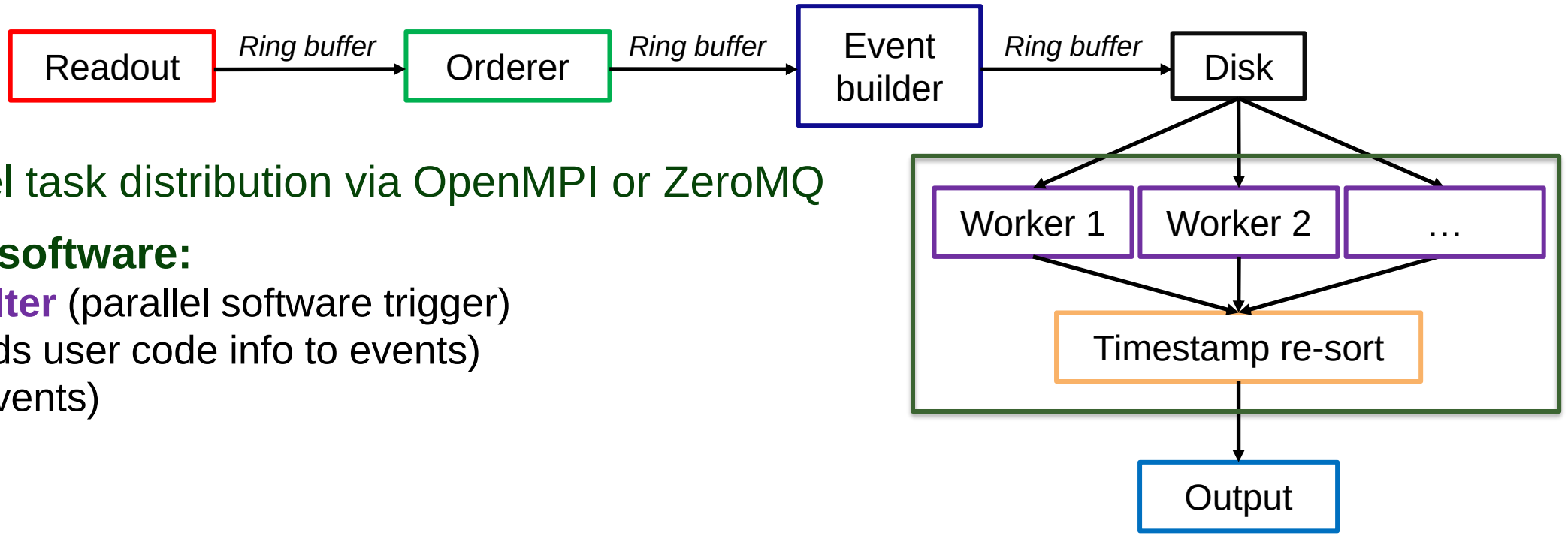
1: FRIBDAQ Readout Architecture

- Software suite (C++, Tcl, Python) that provides a flexible and extensible framework for handling the data flow produced by nuclear physics experiments
- Support for a wide variety of electronics: VME, CAMAC, CAEN and XIA digitizers, GET, …
- Manage the data stream by breaking it into smaller pieces
- Event-building based on timestamps
- Flexible trigger configuration
- ReadoutGUI is the “conductor” of the system
  - Pipeline management, data recording, consolidation of data to output, …



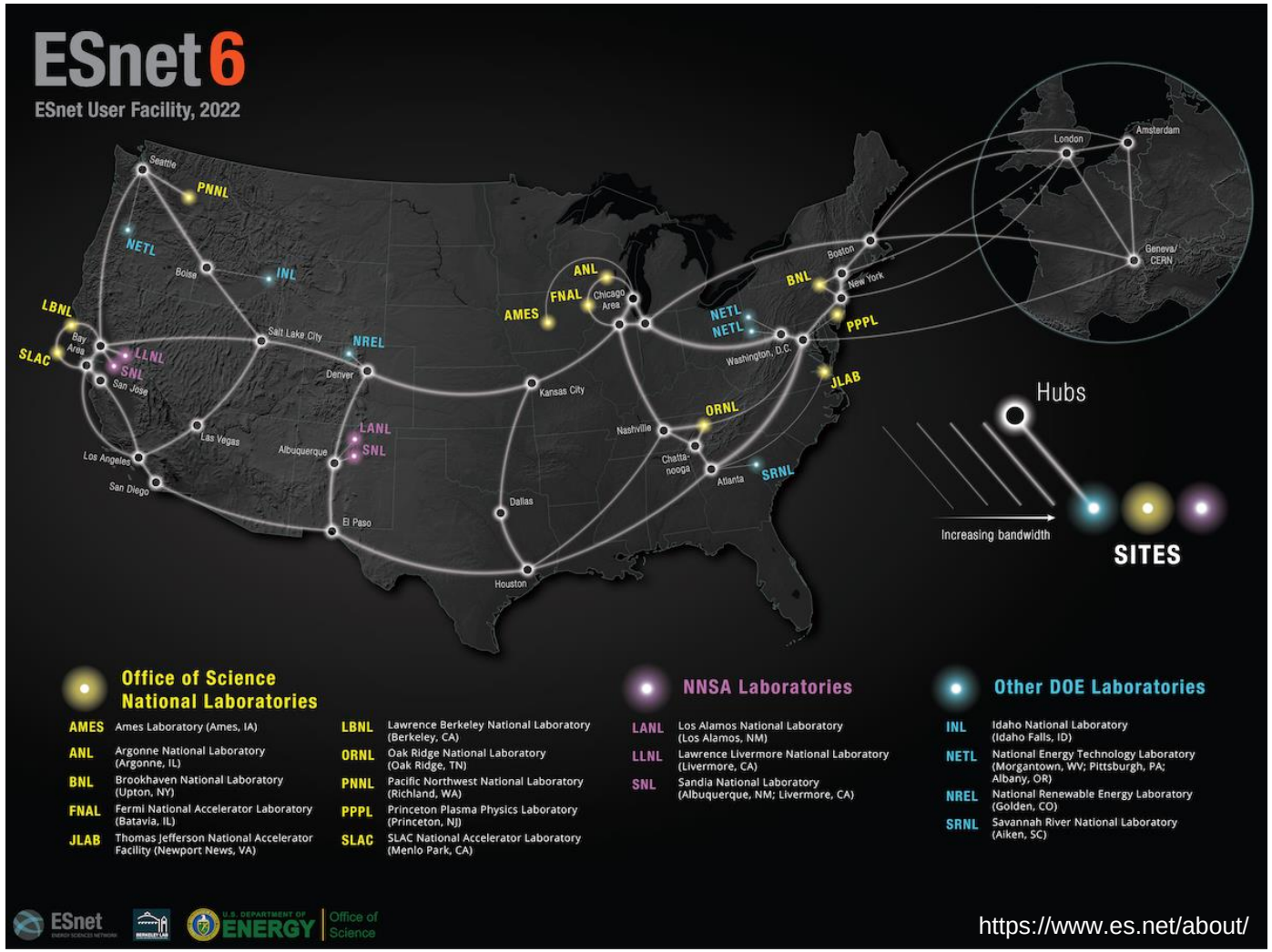
1: FRIBDAQ Upgrades Enable High-Rate Data Taking

- Process-level pipeline parallelism for readout
  - Separate readout and timestamp sorting
- Zero-copy whenever possible
- Common operations (e.g., fit an ADC trace) implemented via plugin libraries



- Support parallel task distribution via OpenMPI or ZeroMQ
- **Event-editing software:**
  - **Classifier + Filter** (parallel software trigger)
  - **Appender** (adds user code info to events)
  - **Editor** (edits events)

2: ESnet Infrastructure at FRIB

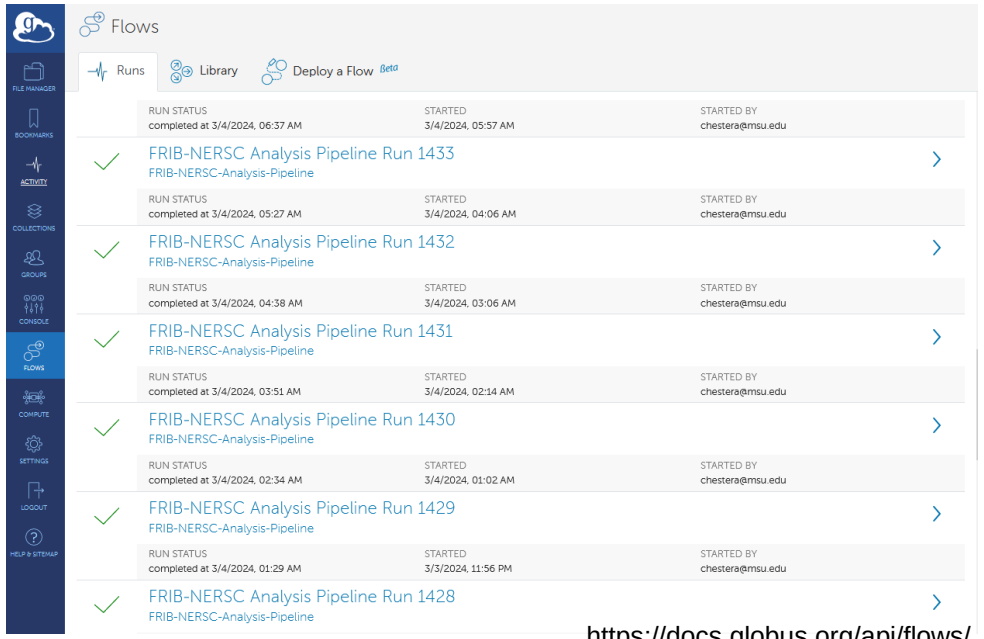


- ESnet: high-performance, unclassified network funded by DOE-SC
- Multiple 100 Gbps optical channels totaling 46 Tbps bandwidth
- Infrastructure at FRIB
  - 2x 100 Gbps connections to ESnet managed by DOE
  - FRIB "Science DMZ"[1]: provides storage and network connection between FRIB experiments and ESnet

[1]"Demilitarized zone"

2: Automated Data Processing With Globus Flows

- "Secure, managed automation of complex workflows at scale"
  - Common tasks implemented via hosted "action providers" (e.g. data transfer)
  - Globus Compute provides a "function as a service" (FaaS) platform for remote execution of user code
  - Error handling
- Action providers and other operations can be assembled into a workflow
- Python-based SDKs[1] for Globus and Globus Compute
  - Register applications, configure inputs, run, manage results
- Web-based monitoring
- Data processing workflow:
  1. Acquire data at FRIB
  2. Transfer raw data to NERSC via ESnet
  3. Process data remotely at NERSC
  4. Transfer processed data back to FRIB via ESnet



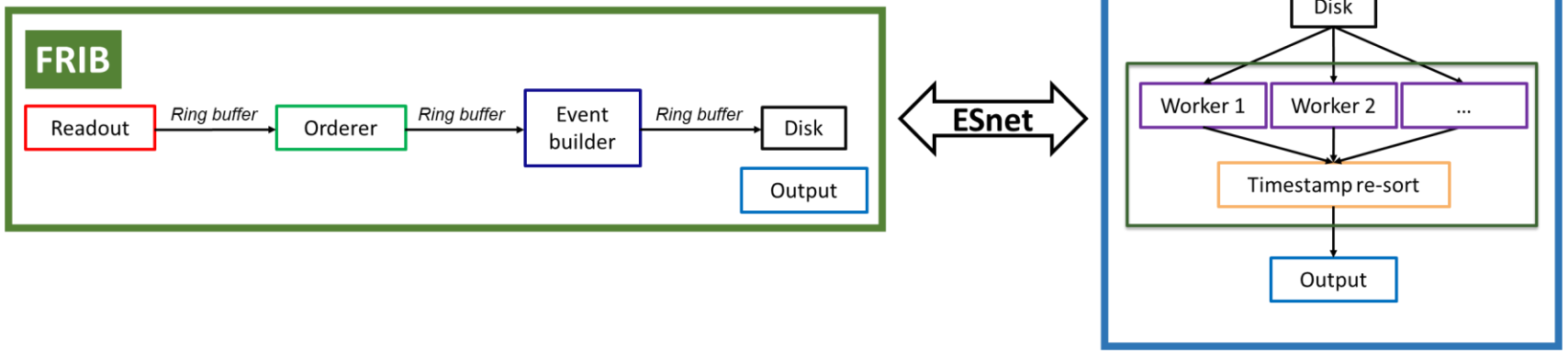
https://docs.globus.org/api/flows/

[1]Software development kit

3: Data Processing at NERSC

- NERSC is the “primary scientific computing facility for DOE-SC”
- It houses the Perlmutter supercomputer:
  - Over 1700 GPU and 3000 CPU nodes
  - All-flash, 35 PB Lustre scratch file system for fast I/O
  - #12 in the TOP500 as of November 2023
  - Job scheduling via Slurm
  - Container support using Shifter
- Globus Compute FaaS
  - Containers allow us to reconstruct the FRIB runtime environment on Perlmutter
- Editor software: fit ADC traces, store fit results
  - Search for rare events, characterize their properties

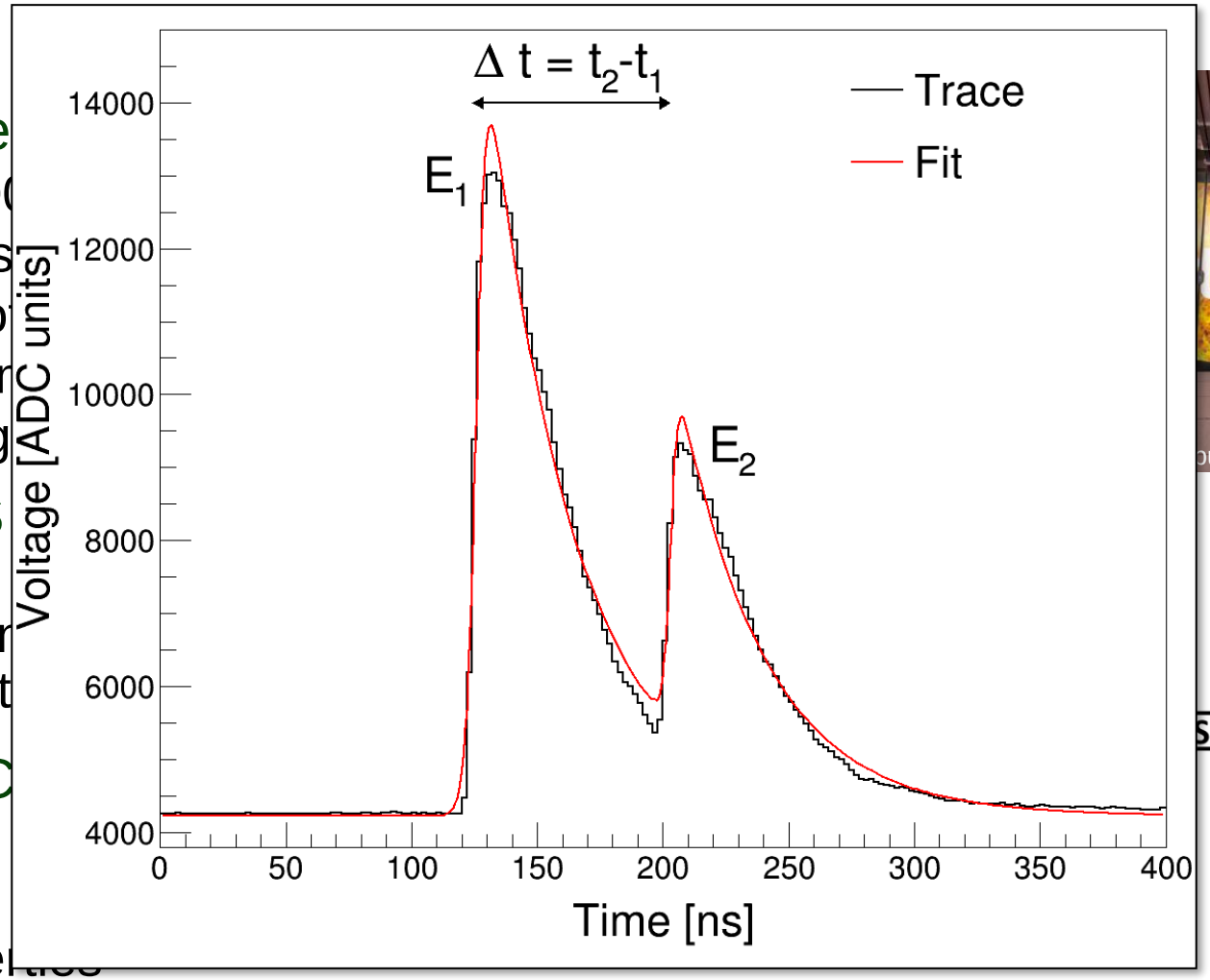
https://www.nersc.gov/news-publications/galleries/nersc-systems-photos/



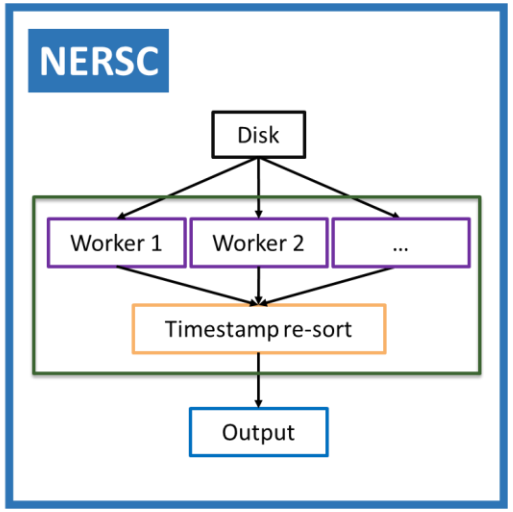
https://www.nersc.gov/about/; A. Chester et al., Phys. Rev. C **104** (2021) 054314; A. Chester et al., Phys. Rev. C **105** (2022) 024319

3: Data Processing at NERSC

- NERSC is the “primary
- It houses the Perlmutte
  - Over 1700 GPU and 30
  - All-flash, 35 PB Lustre s
  - #12 in the TOP500 as o
  - Job scheduling via Slurr
  - Container support using
- Globus Compute FaaS
  - Containers allow us to reconstruct the FRIB rur environment on Perlmut
- Editor software: fit ADC traces, store fit results
  - Search for rare events, characterize their prope



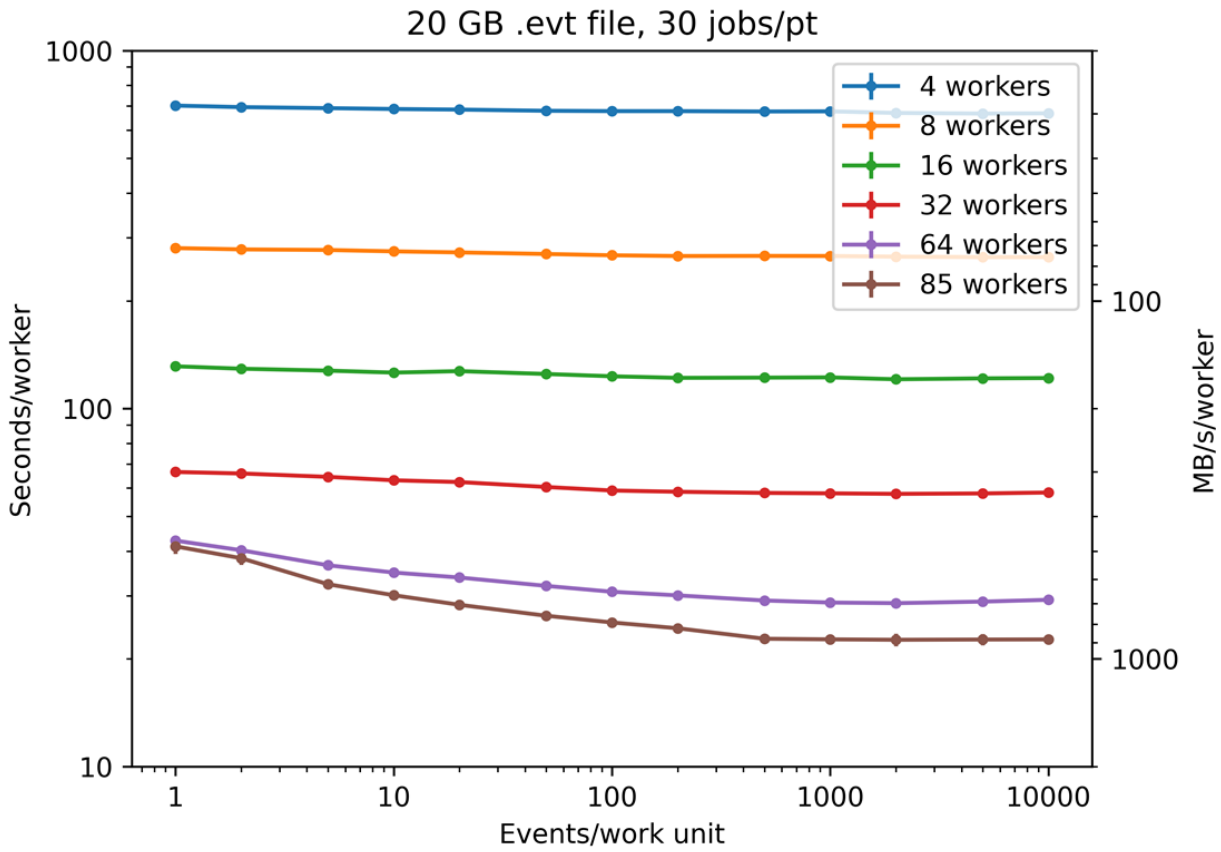
Snet



https://www.nersc.gov/about/; A. Chester et al., Phys. Rev. C **104** (2021) 054314; A. Chester et al., Phys. Rev. C **105** (2022) 024319

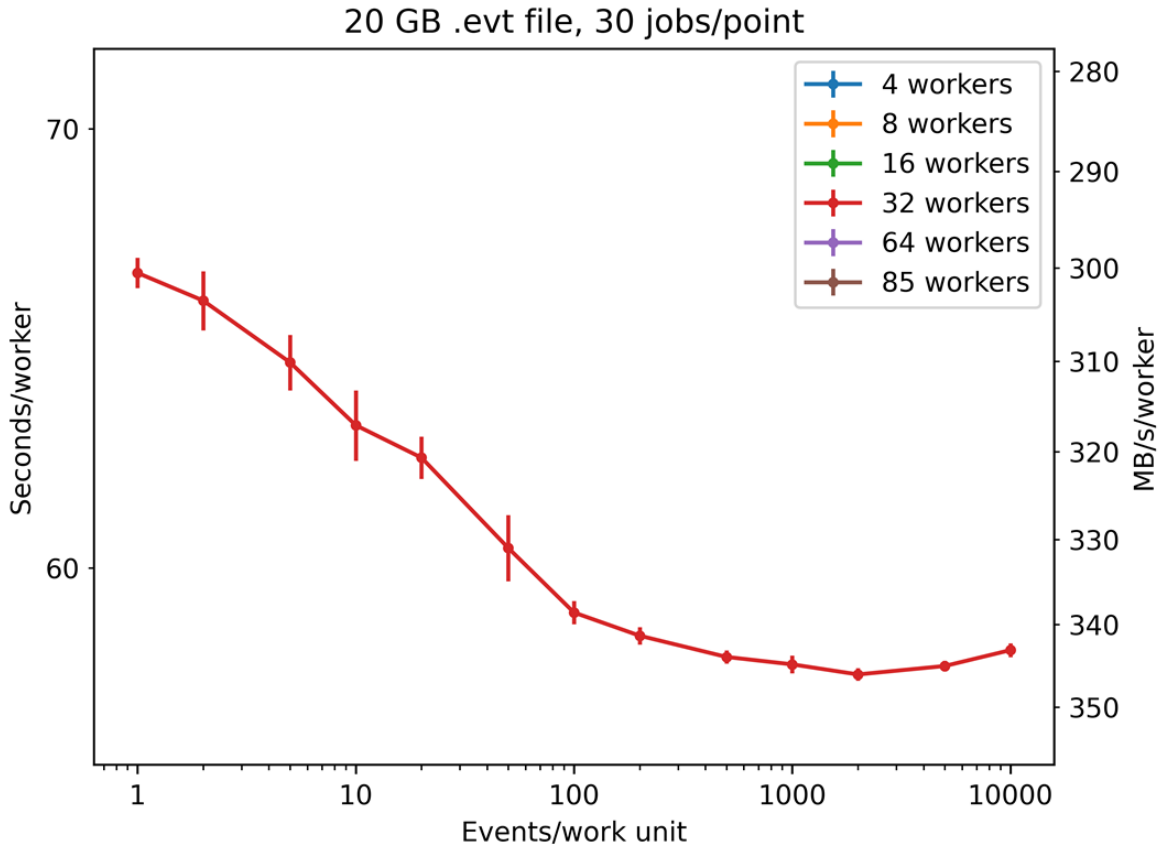
3: Data Processing at NERSC: Proof of Principle

- Free parameters:
  1. Number of MPI workers
  2. Number of events per work unit passed to each worker
- Tested using production data from the first FRIB experiment
  - Fit 2.7M traces (~1.4 GB trace data) from a single channel
  - Trace template detector response model
- Running conditions at NERSC:
  - I/O from node-accessible scratch space
  - Request entire node, allocation at the mercy of the Slurm scheduler
  - Nodes are equivalent



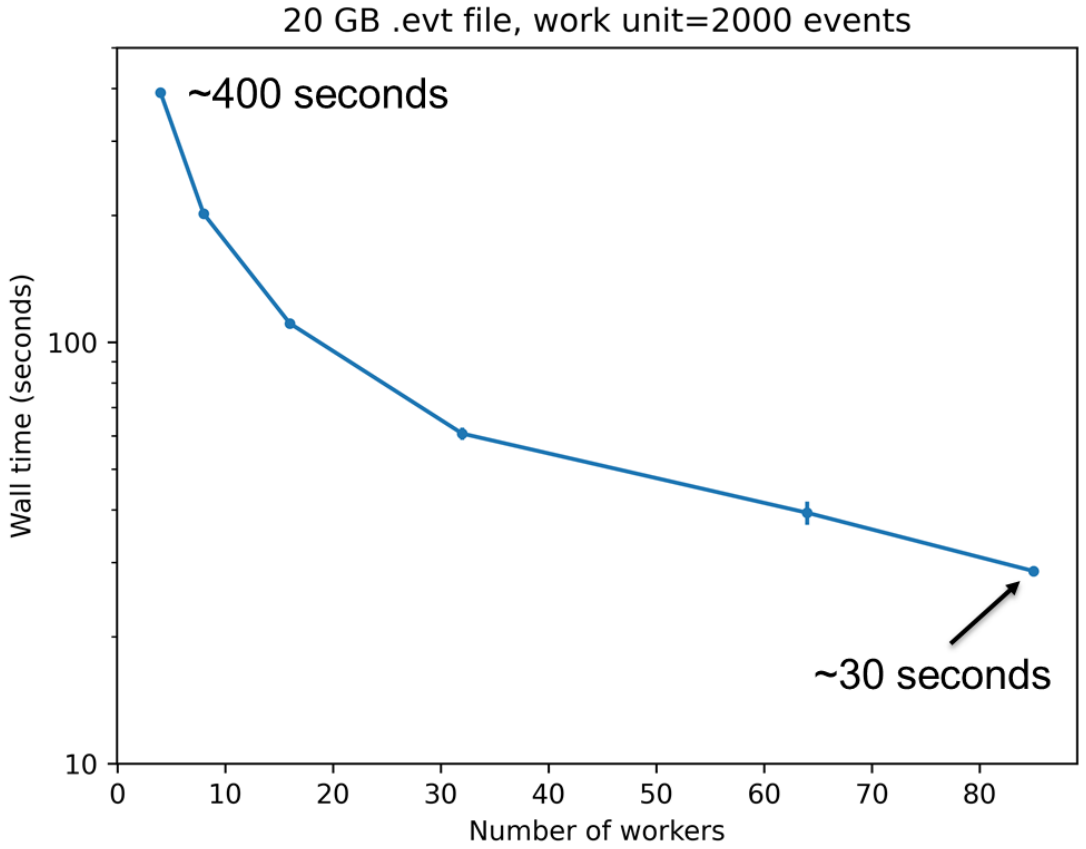
3: Data Processing at NERSC: Proof of Principle

- Free parameters:
  1. Number of MPI workers
  2. Number of events per work unit passed to each worker
- Tested using production data from the first FRIB experiment
  - Fit 2.7M traces (~1.4 GB trace data) from a single channel
  - Trace template detector response model
- Running conditions at NERSC:
  - I/O from node-accessible scratch space
  - Request entire node, allocation at the mercy of the Slurm scheduler
  - Nodes are equivalent

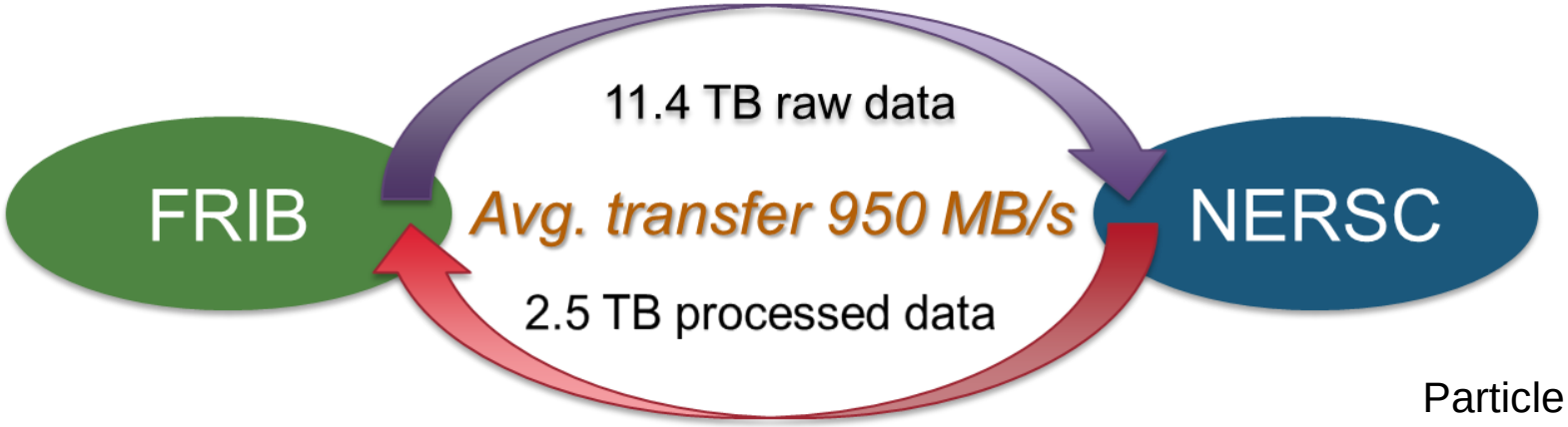


3: Data Processing at NERSC: Proof of Principle

- Free parameters:
  1. Number of MPI workers
  2. Number of events per work unit passed to each worker
- Tested using production data from the first FRIB experiment
  - Fit 2.7M traces (~1.4 GB trace data) from a single channel
  - Trace template detector response model
- Running conditions at NERSC:
  - I/O from node-accessible scratch space
  - Request entire node, allocation at the mercy of the Slurm scheduler
  - Nodes are equivalent



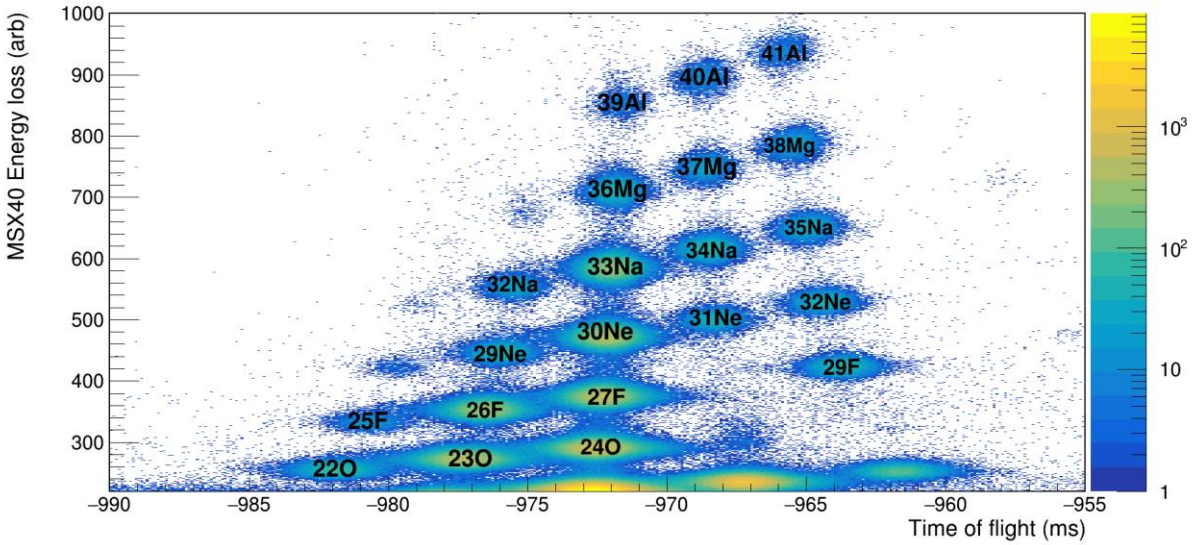
3: Automated Processing Pipeline: In Production



- Experiment ran 140 hours 28 Feb. – 4 Mar. 2024
- 80% reduction in data size
- One hour of recorded data is fully processed in ~10 minutes[1]
- Simultaneous utilization of 3800 CPU cores in realtime queue with startup latency ~few seconds
- Total of 718 compute hours used

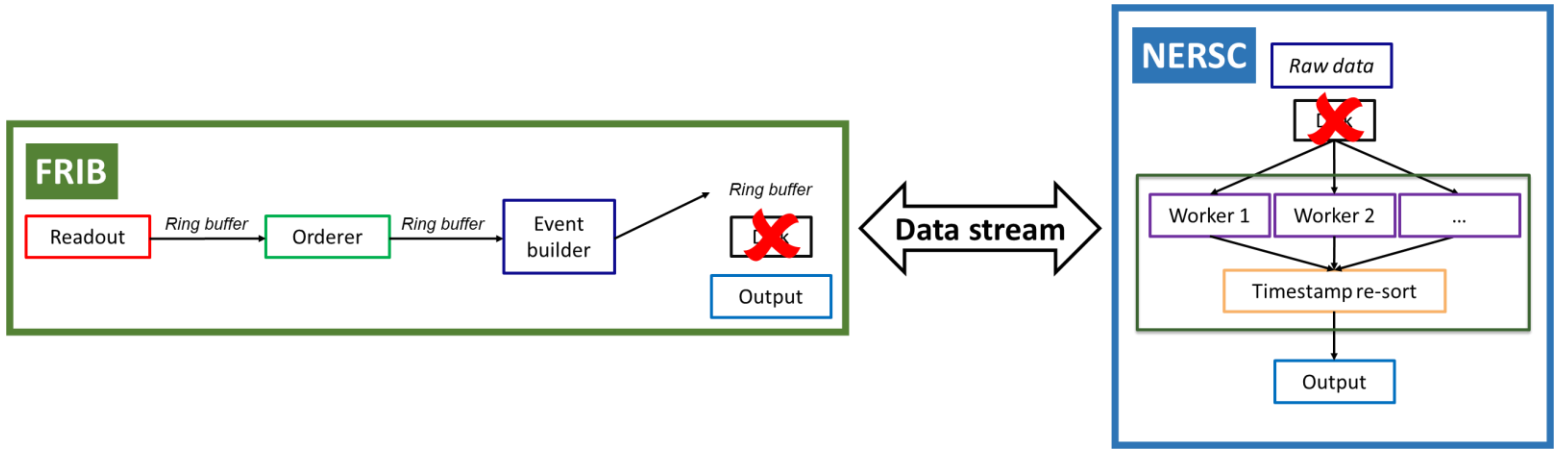
[1]Assuming code optimized for parallel processing

Particle identification plot generated in near real time from data processed at NERSC (courtesy of R. Lubna)



Conclusions

- Nearline analysis is possible using FRIBDAQ parallel analysis framework
- High-speed data transfer between local storage and HPC facility realized using ESnet
- Successfully automated the processing pipeline using Globus Flows
- Globus Compute FaaS platform was used to run FRIBDAQ code remotely at NERSC
- **Important take-home message: improvements to DAQ and data processing software driven by user's needs!**
- Future work:
  - Streaming readout to HPC facilities
    - » Talks by M. Goodrich, M. Battaglieri
  - Improvements to supported software
    - » Modeling detector response
    - » Machine learning classification
    - » Alternative parallel analysis frameworks
    - » …



Acknowledgements

Giordano Cerizza, Genie Jhang, Ron Fox, Simon Giraud, Sean Liddick, Rebeka Lubna, Paul Reece, Tom Rockwell, Bashir Sadeghi, Jerry Vasquez

Heather Crawford, Mario Cromaz

Eli Dart, Jason Zurawski

Debbie Bard, Nicholas Tyler, NERSC support

This material is based upon work supported by the DOE Office of Science, Office of Nuclear Physics and used resources of the FRIB Operations, which is a DOE Office of Science User Facility under Award Number DE-SC0023633.