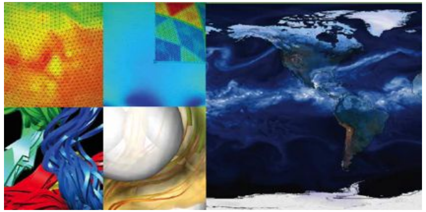
Enabling AI and Analytics on Big Science Data at NERSC

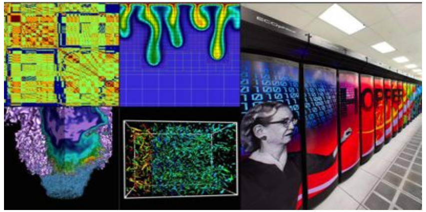
TIES Workshop at PEARC 2023

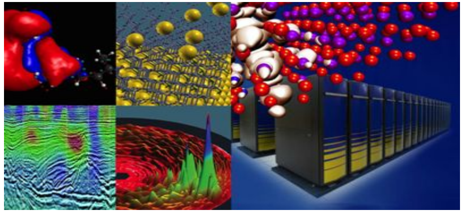
Steven Farrell
Data, AI, and Analytics Services, NERSC
Lawrence Berkeley National Laboratory
July 24, 2024

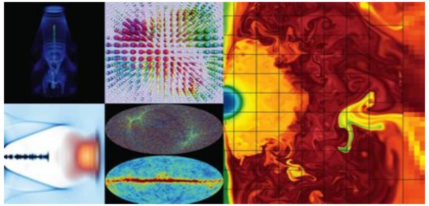
National Energy Research Scientific Computing Center

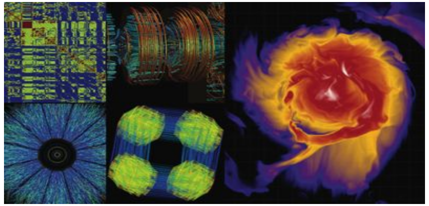
- **NERSC (at LBNL) is the *mission* High Performance Computing and Data facility for the DOE Office of Science**
  - **We deploy supercomputer systems for cutting edge simulations and data analytics at scale**
  - **8,000+ Users, 800+ Projects, ~2000 NERSC citations per year**

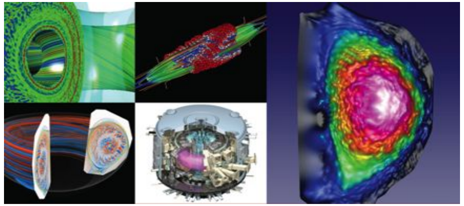

Bio Energy, Environment


Computing

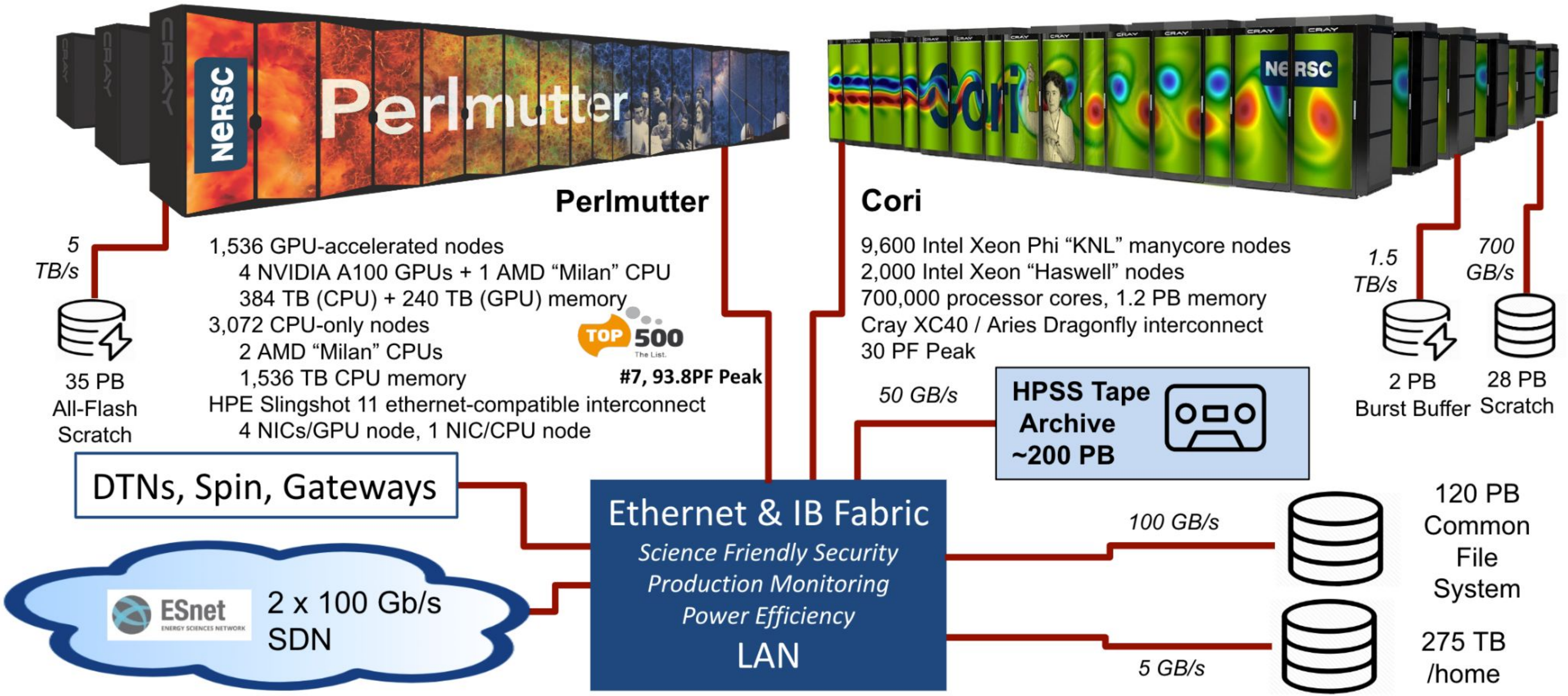

Materials, Chemistry, Geophysics


Particle Physics, Astrophysics

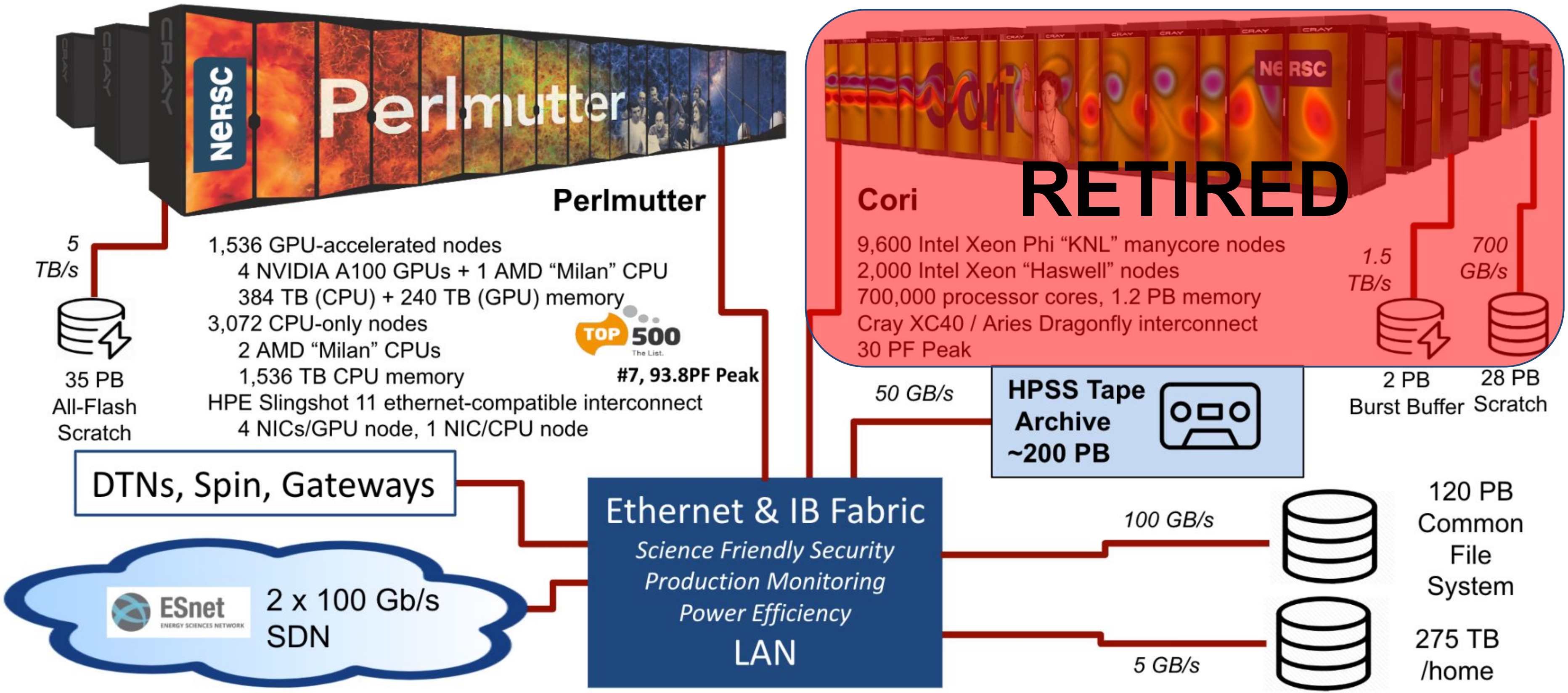

Nuclear Physics


Fusion Energy, Plasma Physics

NERSC diagram



NERSC diagram



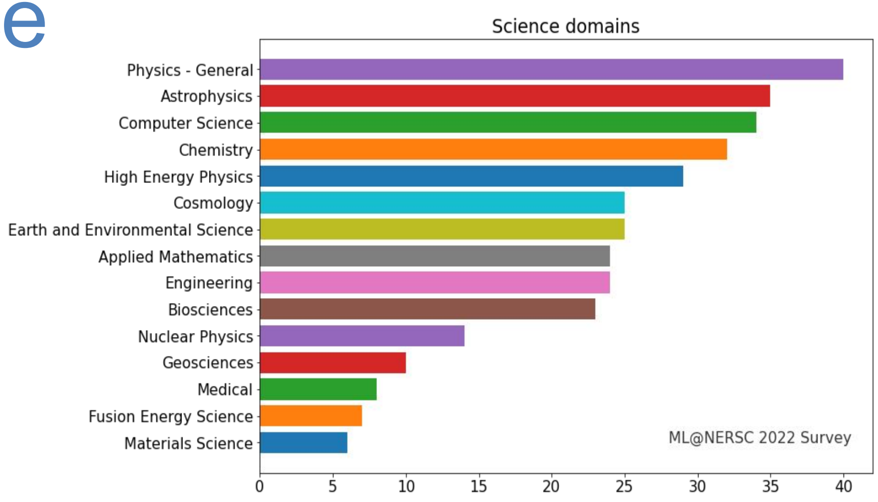
AI is transforming science

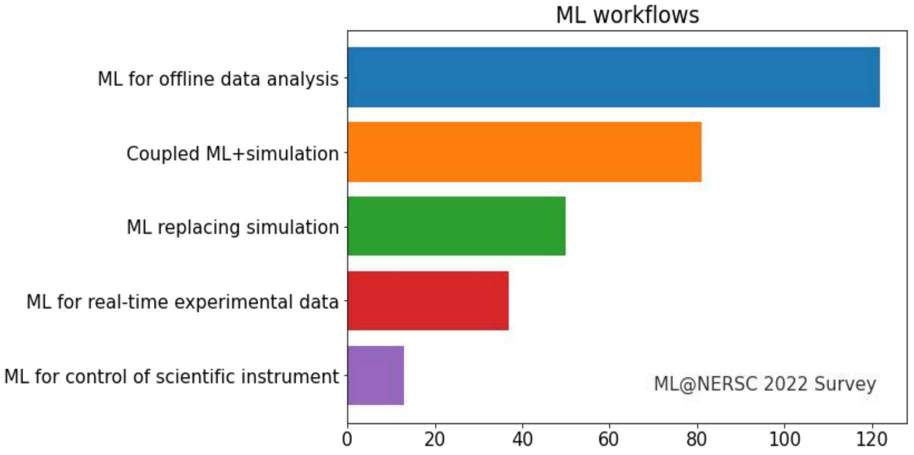
Recent AI wave in DOE science

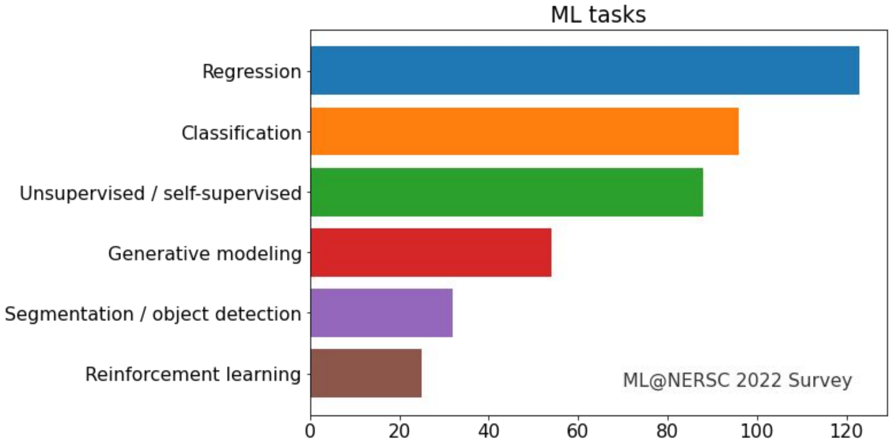
- across all domains

Some of the key application areas:

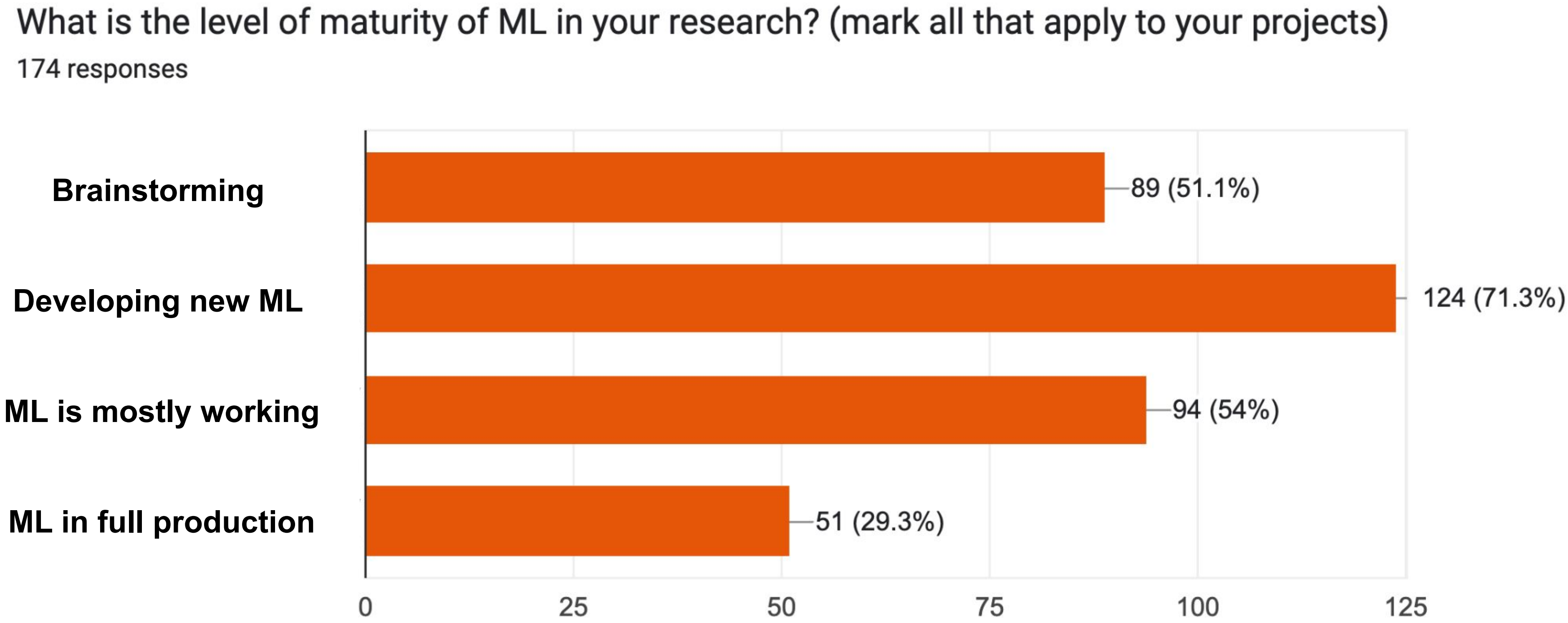
- Analysis of large datasets
- Acceleration of expensive simulations
- Control of complex experiments







AI4Science maturity

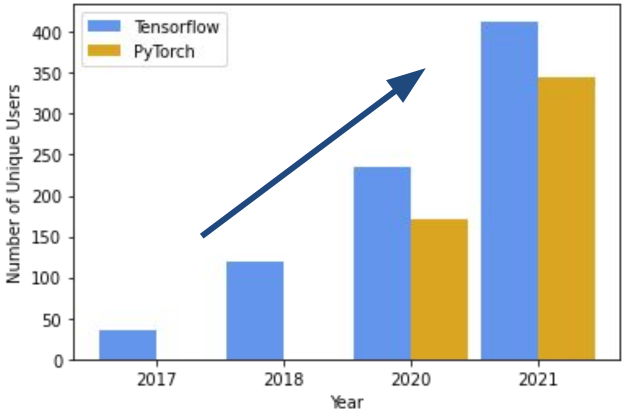


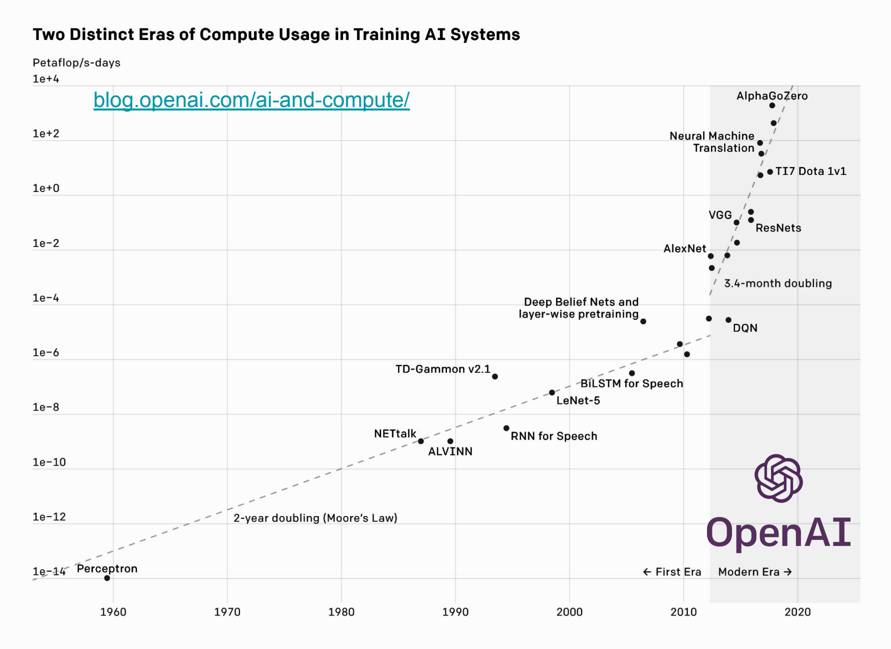
AI4Science needs HPC

Growing computational cost of training

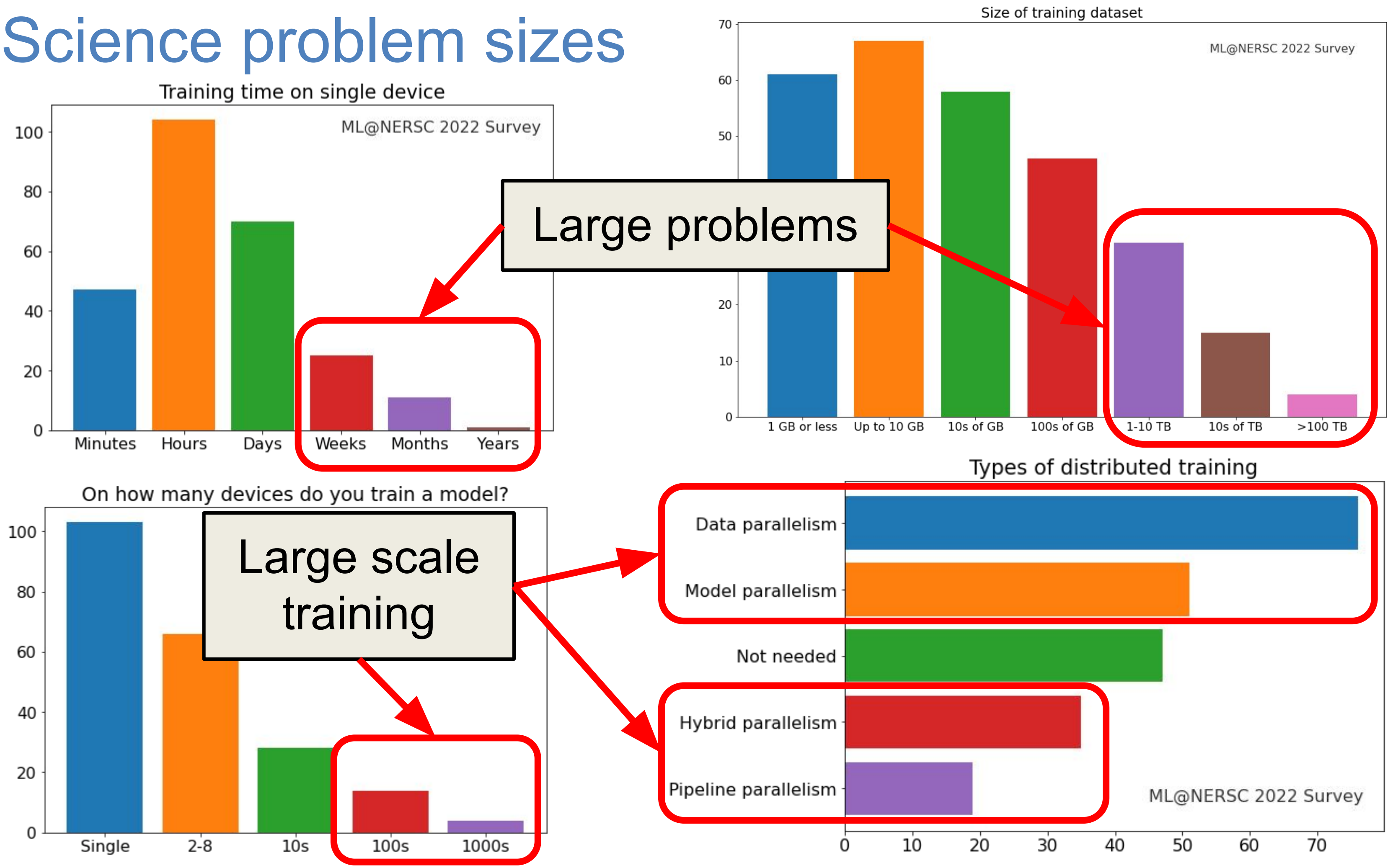
- Problems, datasets, and models getting bigger, more complex

Growing AI workload on HPC





AI4Science problem sizes



AI4Science needs HPC

HPC systems provide the capabilities needed for AI

- e.g., big foundational models for science *will* be trained on supercomputers

HPC centers play an important role beyond hardware

- provide a software ecosystem for large scale scientific workflows, tools for MLOps
- work with scientists to push on the frontiers of methods
- train scientific communities on best practices, tools
- help close the gap between large tech and smaller academic research groups
- *democratize AI for science*

NERSC AI Strategy

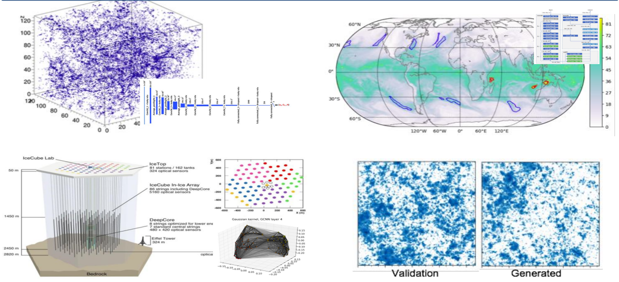
Deployment

Automation | Interactivity

Software Frameworks and Libraries

Systems w/ Accelerators

Methods and Applications



Empowerment

- ***Deploy*** optimized hardware and software systems
- ***Apply*** AI for science using cutting-edge methods
- ***Empower*** through seminars, workshops, training and schools

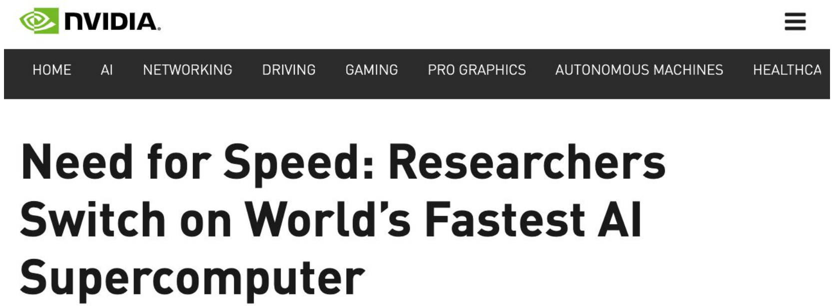
Perlmutter: A Scientific AI Supercomputer

HPE/Cray Shasta system

Phase 1 (installed in 2021):
- 12 GPU cabinets with 4x NVIDIA Ampere GPU nodes; Total >6000 GPUs
- 35 PB of All-Flash storage

Phase 2 (installed in 2022):
- 12 AMD CPU-only cabinets
- HPE/Cray Slingshot high performance network

Optimized software stack for AI
Application readiness program (NESAP)


NVIDIA

HOME AI NETWORKING DRIVING GAMING PRO GRAPHICS AUTONOMOUS MACHINES HEALTHCA

Need for Speed: Researchers Switch on World's Fastest AI Supercomputer

NERSC has a rich data ecosystem!

globus online
jupyter

data transfer and access

netCDF
HDF
mongoDB®
MySQL™

data management

R
julia

data analytics

PyTorch
scikit learn

machine learning

visit
ParaView
Parallel Visualization Application

visualization

SHIFTER
Spin

containers

GNUparallel
Parsl
papermill
FireWorks

workflows

NERSC AI software

We build and deploy optimized modules for

- Python
- PyTorch (pytorch-distributed + NCCL)
- TensorFlow (horovod + NCCL)

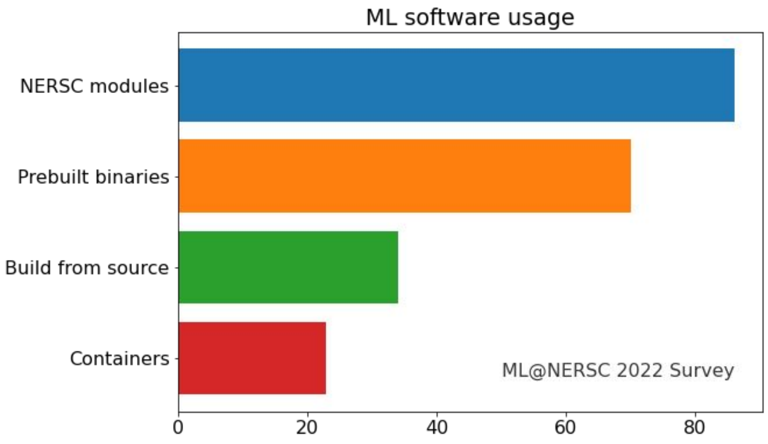
We support optimized containers

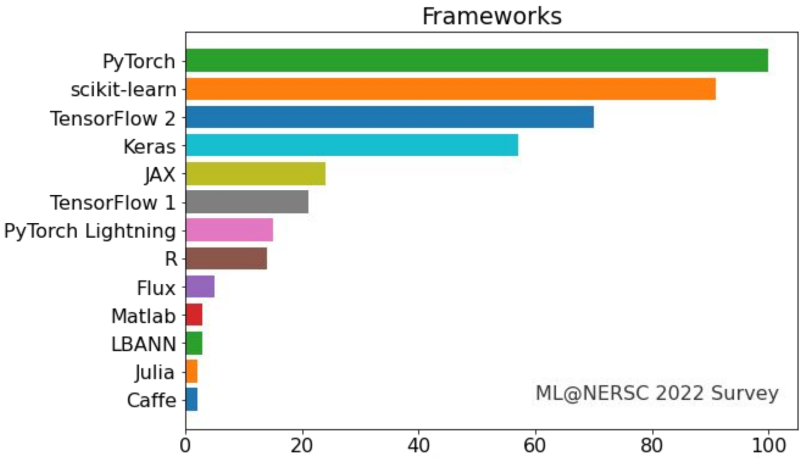
- NVIDIA's NGC DL images are most recommended
- Users can build their own

Users can use their own environments

- conda, etc.

https://docs.nersc.gov/machinelearning/





Containers at NERSC

https://github.com/NERSC/podman-hpc
https://opensource.com/article/23/1/hpc-containers-scale-using-podman

Containers are valuable to our scientific computing users

- Encapsulation, isolation, reproducibility, portability, and even scalability

NERSC supports user container workloads today via Shifter

- Developed at NERSC
- Addresses security concerns of docker (i.e. rootless) and enables scalability on HPC systems
- Users can build their images with docker, then easily convert to shifter with a simple pull command

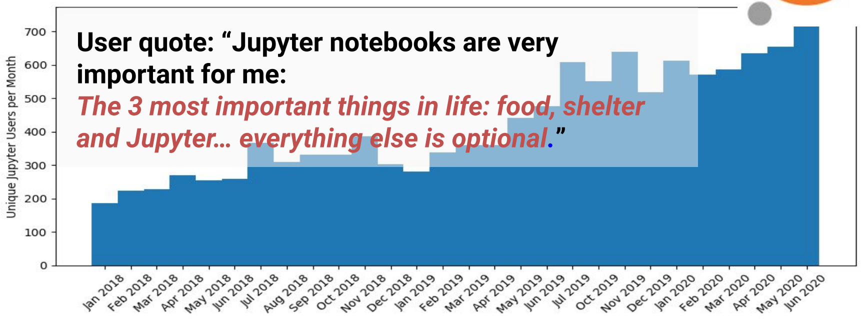
NERSC is currently moving to Podman

- All the benefits of shifter, but using OCI standard runtime
- HPC features provided via the *podman-hpc* wrapper
- Enables user builds at NERSC

Jupyter: supercharge interactive supercomputing

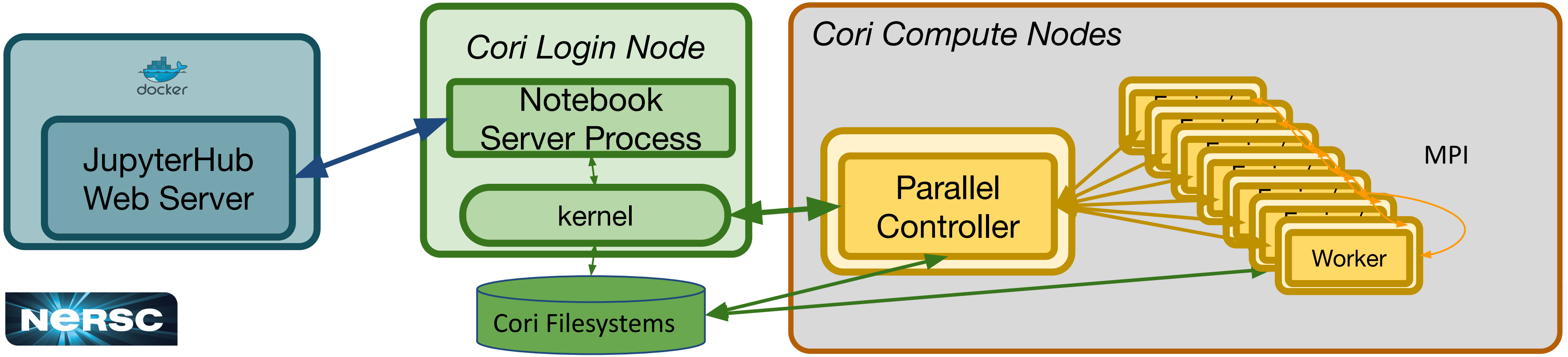
We have deployed an HPC-aware Jupyter service:

- Patterns and frameworks for connecting Jupyter with HPC
- Data Management tools in an HPC environment
- Interactive Visualization
- Reproducible Science through Containerization



Interactive supercomputing: Jupyter Notebook + HPC Workers

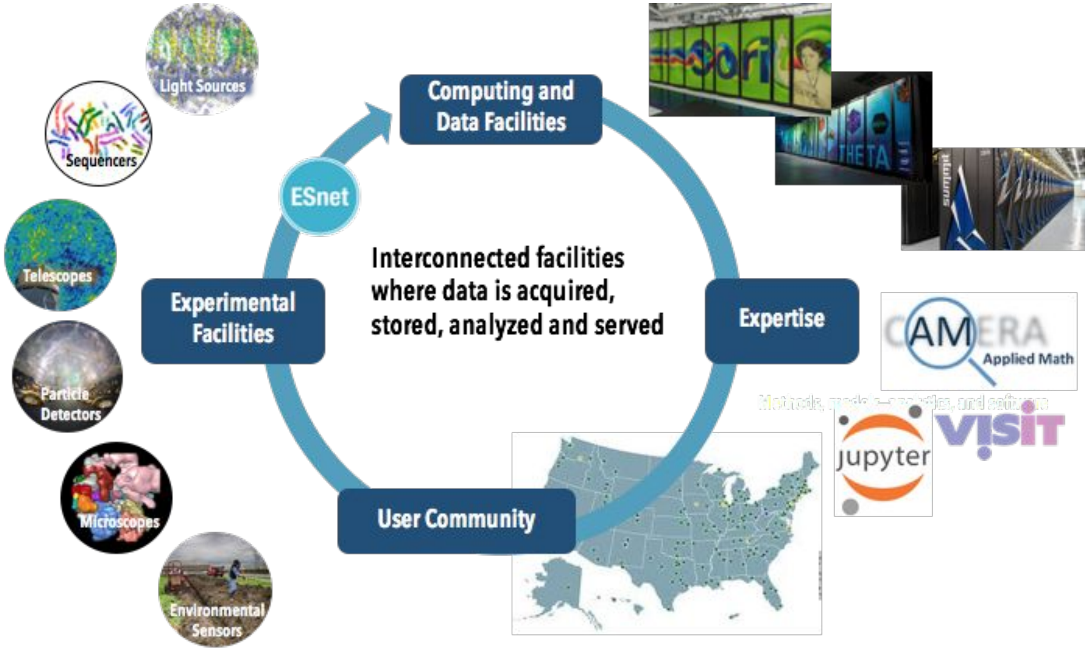
- Launch workers in a short turnaround queue
- Pull results from running HPC Jobs in realtime



The Superfacility Model: an ecosystem of connected facilities, software and expertise to enable new modes of discovery

Superfacility@LBNL: NERSC, ESnet, AMCR, & SDD working together to support experimental science

- Integrates experimental, computational, and networking facilities for reproducible science
- Enables new discoveries by coupling experimental science with large scale data analysis and simulations



NERSC ESnet ENERGY SCIENCES NETWORK CRD COMPUTATIONAL RESEARCH DIVISION

BERKELEY LAB Bringing Science Solutions to the World U.S. DEPARTMENT OF ENERGY Office of Science

Machine-readable supercomputers: the Superfacility API

Vision: all NERSC interactions are callable;
backend tools assist large or complex operations.

Endpoints currently deployed:

`/meta`	information about this Superfacility API installation
`/status`	NERSC component system health
`/account`	Get accounting information about the user's projects
`/utilities`	basic file browsing, upload and download of small files to and from NERSC
`/storage`	Transfer files between Globus endpoints.
`/compute`	Run commands and manage batch jobs on NERSC compute
`/tasks`	Get information about your pending or completed tasks
`/reservations`	submit and manage future compute reservations

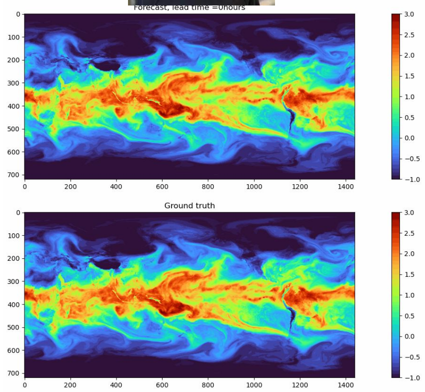
 https://api.nersc.gov/

The NESAP Learning program

Jaideep Pathak
now NVIDIA

Shashank
Subramanian
now staff

- Part of NERSC's Application Readiness Program
  - Partnerships with science and vendor teams to push on science applications
- Very successful to-date with large scale results and high-impact publications
- *New projects and postdocs starting soon!*

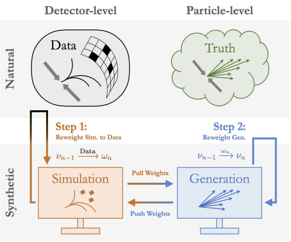


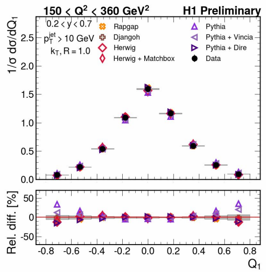
FourCastNet

Pathak et al. 2022
arXiv:2202.11214
High-resolution atmospheric forecasting. Hybrid data/ model parallel @ 4000 GPUs

First deep-learning model with skill & resolution of numerical weather prediction

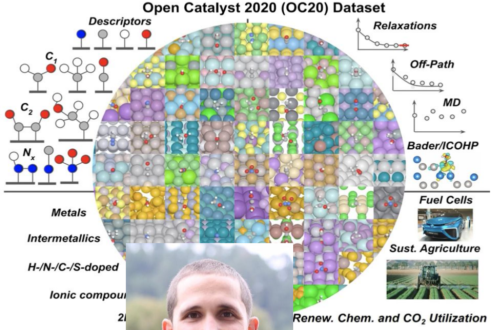
Vinicius Mikuni
NERSC Postdoc

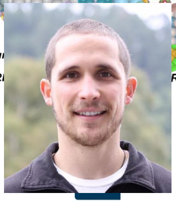




Unfolding for particle physics

H1 Collaboration ([...] Mikuni et. al.): recent press release





Brandon Wood
now Meta AI

CatalysisDL

Chanussot et al. 2021 arXiv:2010.09990
Largest catalysis dataset (OC20 and OC22); Graph-parallel NN approaches and NeurIPS 2021 + 2022 Competitions

Pre-trained models now used with DFT - e.g. FineTuna; AdsorbML

NERSC

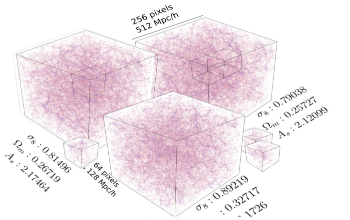
U.S. DEPARTMENT OF ENERGY | Office of Science

Optimizing AI4Science on HPC

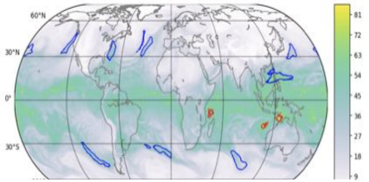
We need good *benchmarks* that represent the scientific workloads

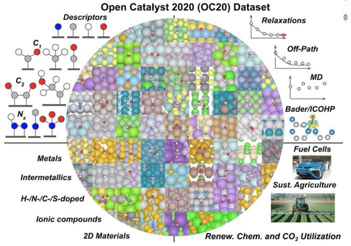
- MLPerf™ benchmarks, by the MLCommons™ organization, are the industry standard measure of ML compute performance

[MLPerf HPC](#) is thus an AI training benchmark with scientific applications designed to push on HPC systems



- CosmoFlow - 3D CNN predicting cosmological parameters
- DeepCAM - segmentation of phenomena in climate sims
- OpenCatalyst - GNN modeling atomic catalyst systems
- OpenFold (***new in ‘23***) - protein folding (AlphaFold2)





Two measurement types:

- *Time to train* a model to target accuracy
- *Throughput* (models/min) of training many models concurrently

MLPerf HPC outcomes and next steps

We've had 3 successful submission rounds so far

- *With results from several leading HPC systems across the world*
  - ANL: ThetaGPU
  - CSCS: Piz Daint
  - Dell: 32XE8545
  - Fujitsu: ABCI
  - Fujitsu+RIKEN: Fugaku
  - Helmholtz AI: HoreKa, JUWELS
  - LBNL: Cori, Perlmutter
  - NCSA: HAL
  - NVIDIA: Selene
  - TACC: Longhorn
- *Driving impressive performance and scaling improvements*
  - improvements up to 5x in most recent submission round (v2.0)
  - throughput measurements scaled up to 5,120 GPUs (Perlmutter), and 82,944 CPUs (Fugaku)
- *All submission code along with results are published and open source*
  - https://mlcommons.org/en/training-hpc-20/
  - https://github.com/mlcommons/hpc_results_v2.0

MLPerf HPC outcomes and next steps

For NERSC, participation has been extremely valuable

- Helped us shake out Perlmutter during deployment
- Enabled us to evaluate Perlmutter performance and showcase its capabilities

We have great plans for MLPerf HPC in 2023!

- Adding a new benchmark based on AlphaFold2 (OpenFold)
- Adding power measurements
- Increasing outreach, educational, and publication opportunities
- Please reach out if you are interested in learning more

Empowerment and training resources

The Deep Learning for Science School at Berkeley Lab https://dl4sci-school.lbl.gov/

- 2019 in-person lectures, demos, hands-on sessions, posters (videos, slides, code)
- 2020 summer webinar series. Recorded talks: https://dl4sci-school.lbl.gov/agenda

The Deep Learning at Scale Tutorial

- Since 2018, and with NVIDIA in 2020/21
- 2021 was first training event to use Perlmutter Phase 1 with hands-on material for distributed training
- See the SC22 material here
- Accepted again for SC23!

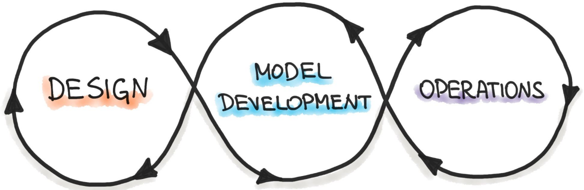
NVIDIA AI for Science Bootcamp - Aug 25-26, 2022

- View the agenda and slides

Other NERSC trainings

- New User Training, Data Day, etc.

Next steps: MLOps for science



We need tools that are easy-to-use for newcomers as well as production-grade for more mature workflows

- They should help users build, train, tune, and deploy their AI models
- Clouds and big AI enterprises all have their own interfaces for MLOps

Some cool things we're currently working on

- Distributed AI with jupyter notebooks
  - Using Ray Train + Ray Tune for distributed training and HP tuning
  - Utilities to easily spin up Ray cluster, collect metrics, show dashboard
- Distributed GPU inference serving
  - for heterogeneous CPU+GPU jobs
  - for superfacility workflows
- Platforms for automation and experiment tracking

Andrew Naylor
NERSC postdoc

RAY

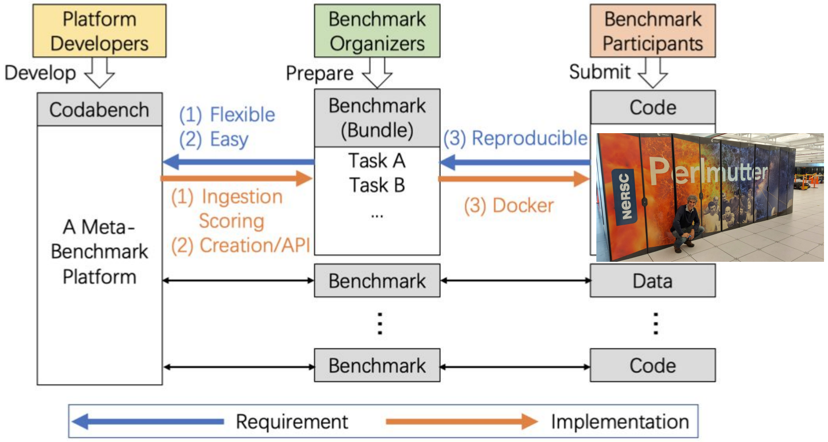
NVIDIA
TRITON INFERENCE SERVER

The FAIR Universe project

A DOE HEP project to develop an unbiased data benchmark ecosystem for physics

- Provide a large-compute-scale platform for sharing datasets, training large models, and hosting challenges and benchmarks.
- Host challenges and benchmarks focused on discovering and minimizing the effects of systematic uncertainties.

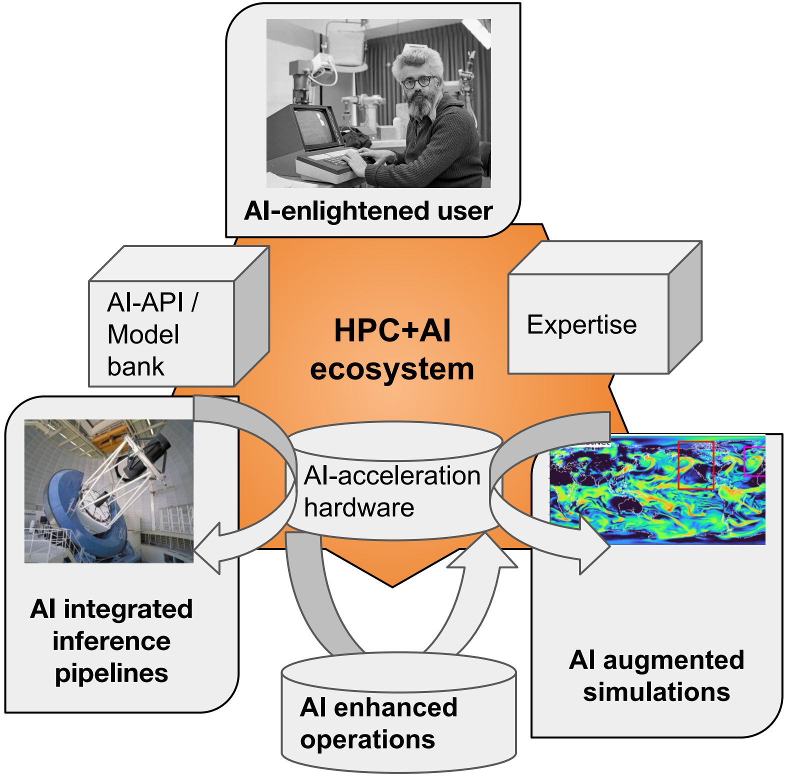
Will ultimately enable new ways of conducting open, reproducible research!



Using the CodaBench platform and extending it to interface with NERSC systems and tackle uncertainty-aware physics challenges

Closing thoughts

- HPC centers play an essential role in enabling open science
- NERSC offers world-class capabilities for scientific AI + analytics
- We would love to hear from you about what more we could be doing
- The future is looking bright for AI-enhanced scientific discovery

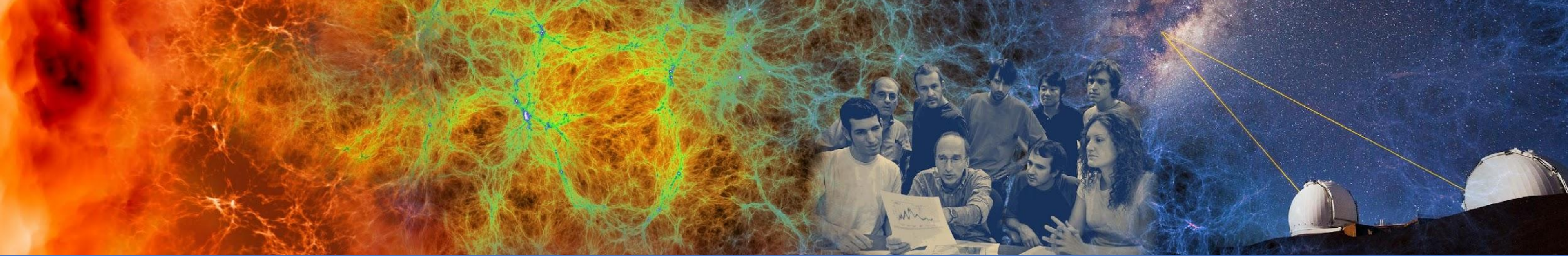


Thank you

Questions? Collaborations?

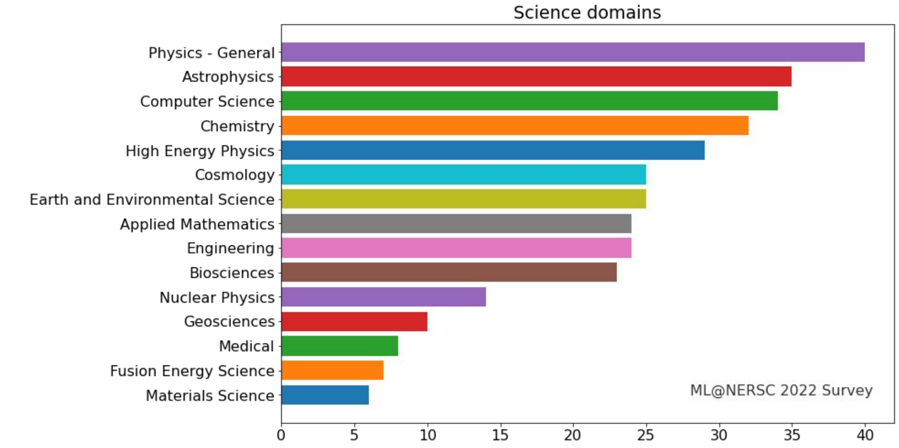
SFarrell@lbl.gov

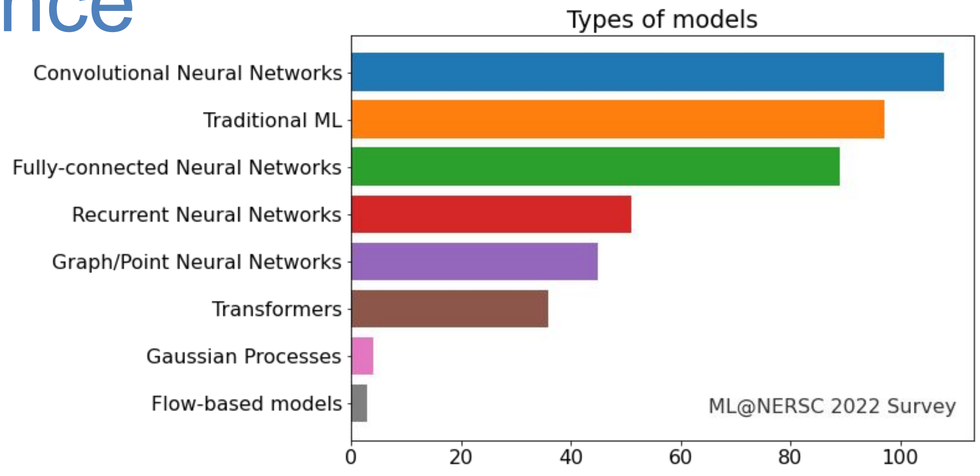
Jobs @ NERSC: https://lbl.referrals.selectminds.com/page/nersc-careers-85

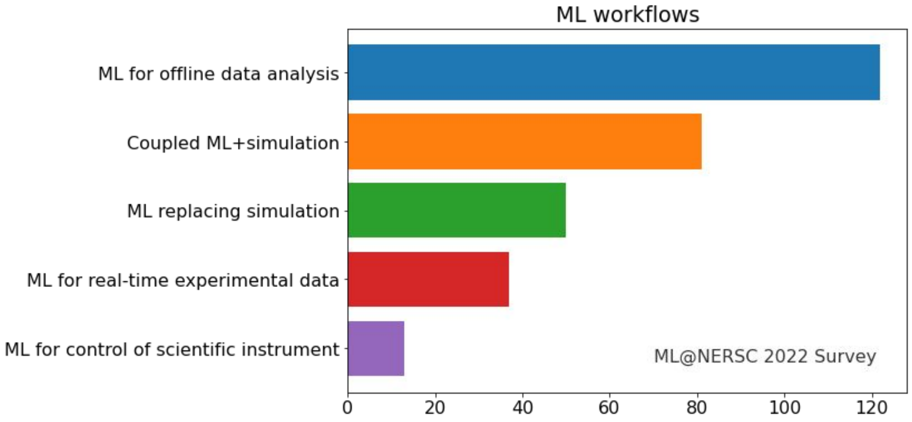


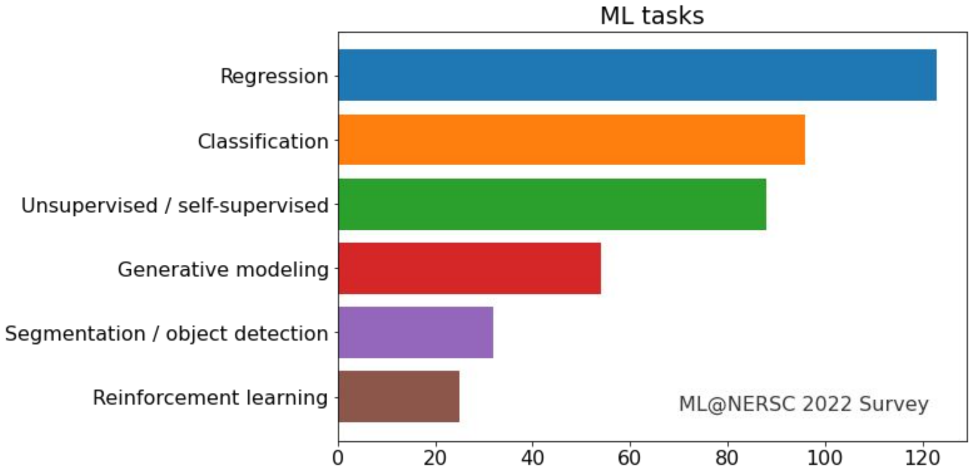
Backup

AI is transforming science









Spin: Container Services for Science

Many projects need more than HPC.

Spin is a platform for services.

Users deploy their **science gateways, workflow managers, databases, and other network services** with Docker containers.

- *Access HPC file systems and networks*
- *Use public or custom software images*
- *Orchestrate complex workflows*
- *Secure, scalable, and managed*

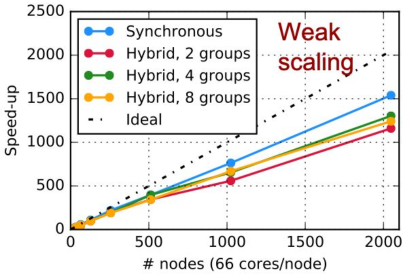
Some projects using Spin:

Project	Description	Service
Dark Energy Spectroscopic Instrument	Track and compare analyses of nightly sky surveys	science gateway
ESS-DIVE	Classify and store reusable earth sciences data	data repository
JGI Joint Genome Institute, Department of Energy	Manage production genomic workflows and data at scale	science gateway
LZ	Process real-time events for dark matter detection	workflow manager
The Materials Project materialsproject.org	Explore materials properties or build simulated materials	science gateway

Evolution of deep learning for science and *supercomputing*

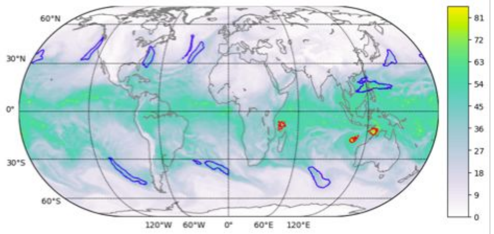
Some example projects:

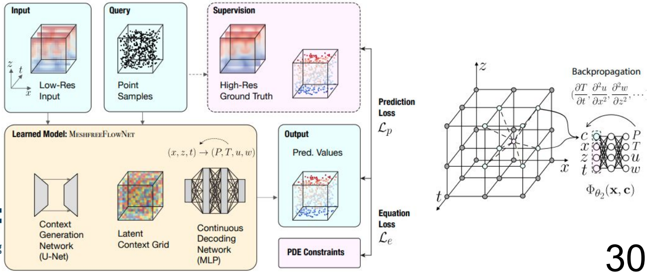
- 2017 SC17 conference [Deep learning at 15PF]
- 2018 Gordon Bell Prize Exascale DL for Climate Analytics
- 2019 Etalumis: bringing probabilistic programming to scientific simulators at scale
- 2020 SC20 MeshfreeFlowNet: a physics-constrained deep continuous space-time super-resolution framework
- 2022 FourCastNet: Accelerating Global High-Resolution Weather Forecasting using Adaptive Fourier Neural Operators



This period showed a very rapid growth in

- Available Compute
  - 15 PetaFlops in SC17 -> 'Exascale' (half-precision) in SC18
- Sophistication of models and methods
- Availability of software
  - Custom hand-rolled Caffe/MPI SC17
  - Tensorflow/Horovod and Cray DL Plugin SC18
  - Pytorch DDP SC19

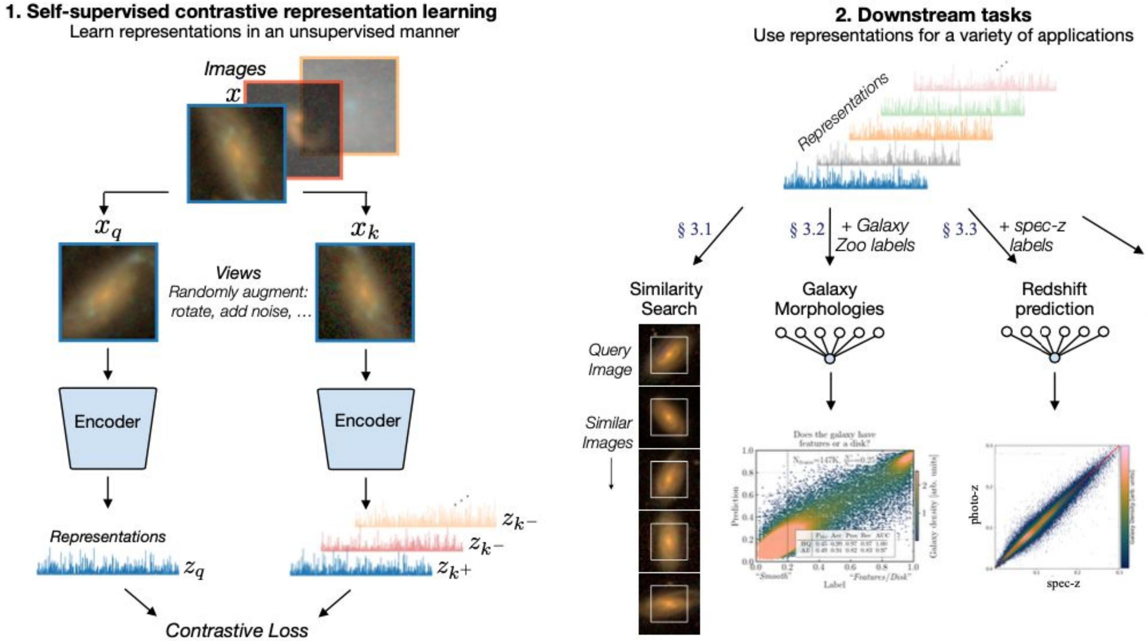


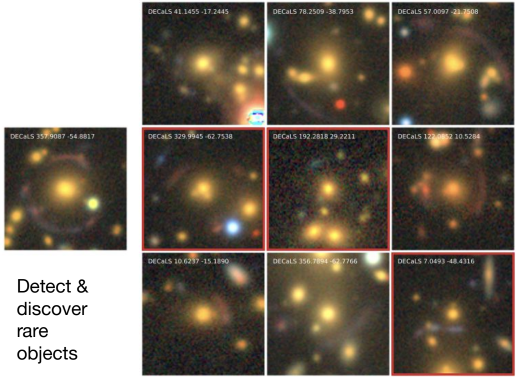


Analyze: Self-supervised sky surveys

Initial approach: Hayat et. al. (2020) arXiv:2012.13083
Strong-lens analysis: Stein et. al. (2021) arXiv:2110.00023

- Sky surveys image billions of galaxies that need to be understood
- Limited "labels", so can learn in *semi-supervised* way
- Pre-training on entire dataset on HPC, downstream task can be on laptop/edge
- Recently used to find > 1000 previously undiscovered strong-lens candidates





Peter Harrington
NERSC ML Engineer

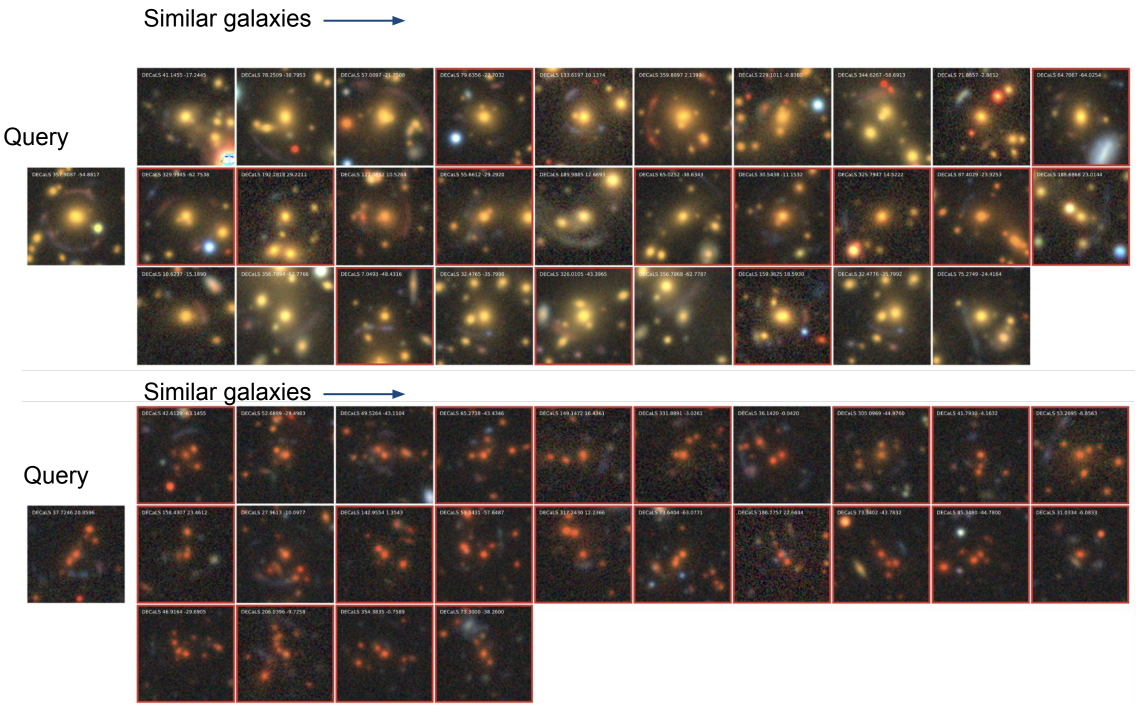
Similarity search

- Given just a **single example**, instantly search for similar objects.

- Discover **new lenses or other phenomena** given just a few queries

Direction for future deep learning for science:

- **Community can benefit from multipurpose models trained on large-scale computing**

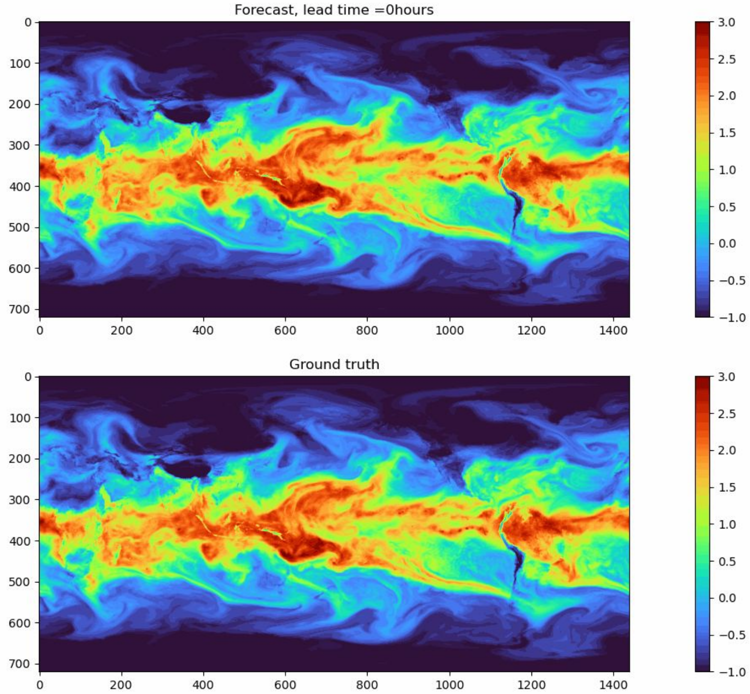


Try it out yourself:
share.streamlit.io/georgestein/galaxy_search

Accelerate: Data-driven atmospheric modeling

Pathak et al. 2022
arXiv:2202.11214

- Data-driven modeling of atmospheric flows using a state-of-the-art transformer-based “Fourier Neural Operator”
- Collaboration with NVIDIA, Caltech and others
- Forecasts global weather at 0.25° resolution
  - Order of magnitude greater resolution than state-of-the-art deep learning models
  - Forecasts wind speeds, precipitation and water vapor close to the skill of numerical weather prediction models up to 8 days
  - Produces a 24hr 100-member ensemble forecast in 7 seconds on a Perlmutter GPU node
  - Traditional NWP: 5 mins on *thousands of CPU nodes* for equivalent ensemble



Data-driven forecast of an atmospheric river

Jaideep Pathak
former NERSC
Postdoc now NVIDIA

Peter Harrington
NERSC ML
Engineer

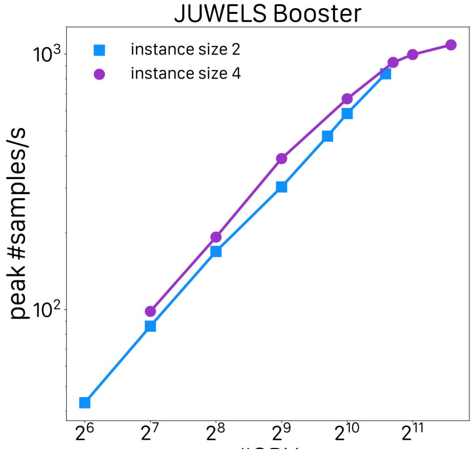
Shashank Subramanian
NERSC Postdoc

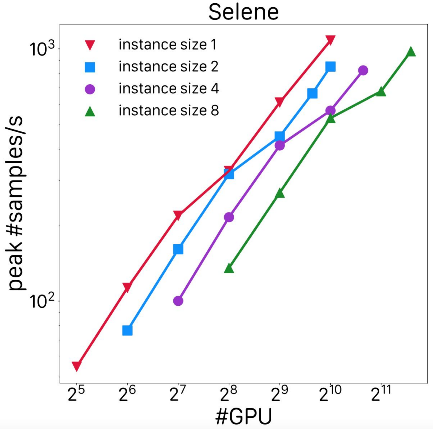
33

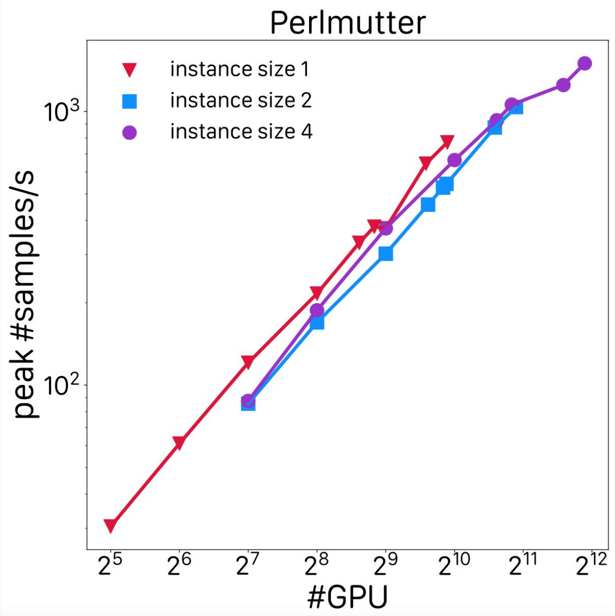
FourCastNet: Large-compute scaling

Pathak et al. 2022
arXiv:2202.11214
Kurth et al. 2022
arXiv:2208.05419

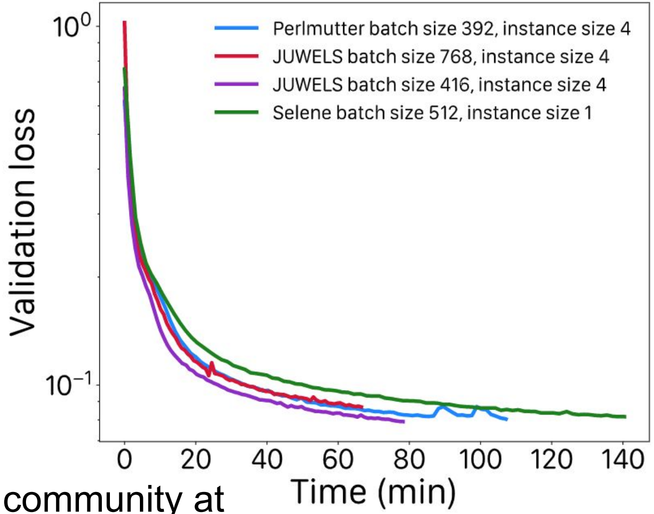
Scales to e.g. 3808 GPUs on Perlmutter with model parallel on 4-gpus







Train large models on ~1hr timescales compared to 40 hrs on 32 nodes or >~45days on a single GPU :



Model and weights made available to community at
https://github.com/NVlabs/FourCastNet

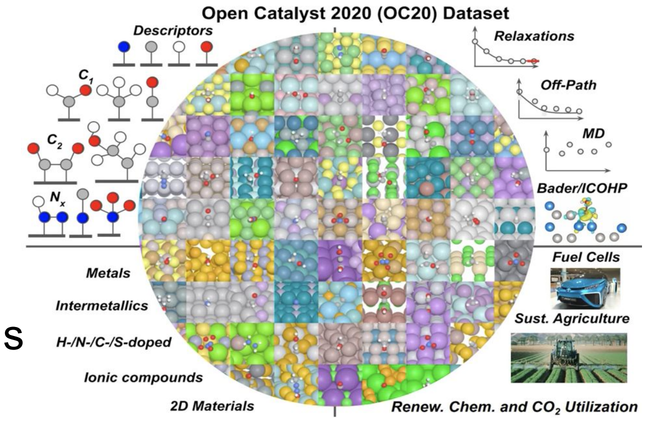
BERKELEY LAB Bringing Science Solutions to the World

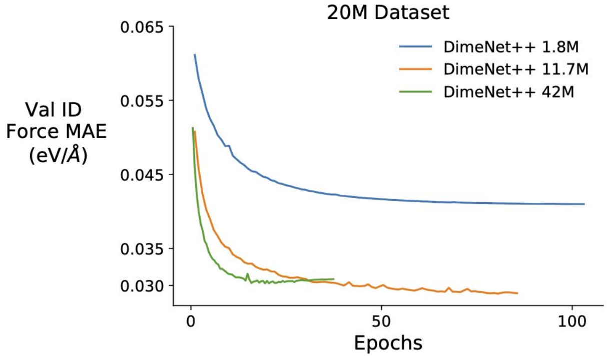
U.S. DEPARTMENT OF ENERGY Office of Science

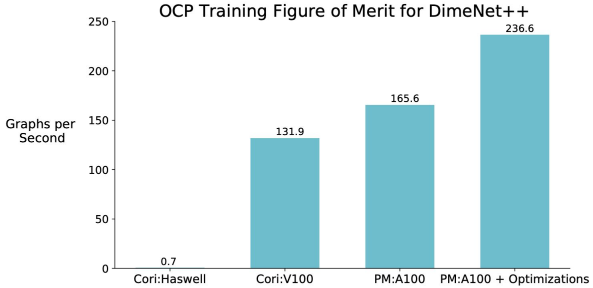
Automate: discovering new catalysts

https://opencatalystproject.org/
Chanussot et al. 2021 arXiv:2010.09990

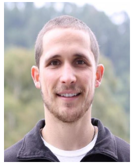
- GraphNNs to accelerate catalyst discovery for energy storage and climate change mitigation
- Collaboration with CMU and Facebook/Meta
- Largest catalysis datasets to date (OC20 and OC22)
  - Challenges in NeurIPS 2021 and 22
- Perlmutter helps push to larger better performing models
- Exploiting Graph-parallel NN approaches







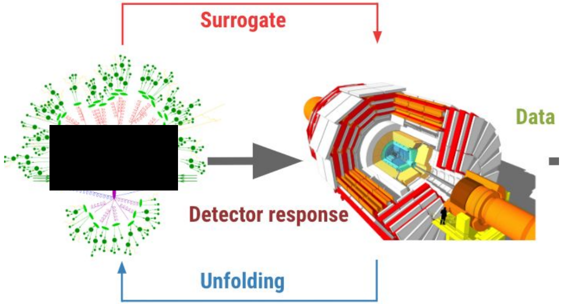
Performance comparison of Perlmutter (PM) with Cori CPU and GPU nodes.
Optimizations carried out in collaboration with NVIDIA DevTechs



Brandon Wood
NERSC Postdoc now
Meta AI

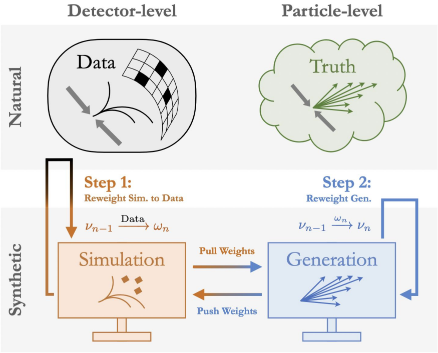
- Public pre-trained models on OC20 now used by CMU group for 90% faster relaxation fine-tuned by DFT in active learning framework https://github.com/ulissigroup/finetuna

Unfolding for particle physics

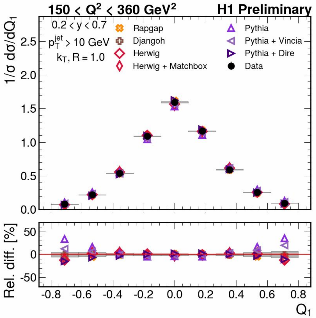


H1 Collaboration ([...] Mikuni et. al. 2002 [Phys. Rev. Lett. 128, 132002](), [2022 Deep Inelastic Scattering (DIS) Conference](). and recent [press release]()

- "Unfolding" of fundamental particle interactions from observation in complex building-size experiments
- Collaboration with LBL Physics Division and H1 Collaboration
- Combines novel iterative ML approach [OmniFold]() with GraphNN to extracts new physics insights
- Uses Perlmutter for 1000s of bootstrapping and UQ runs each using 128 GPUs for training

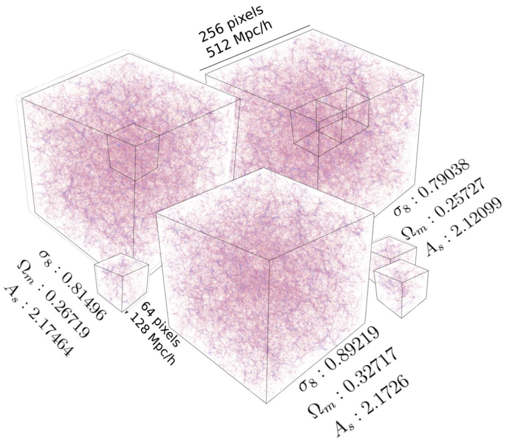


- **Other projects to replace full detector simulation** (expensive and not easily scalable)
  - **Using ML surrogate models** incorporating **diffusion generative models** for the first time in particle physics
  - **More info here:** [arXiv:2206.11898]()



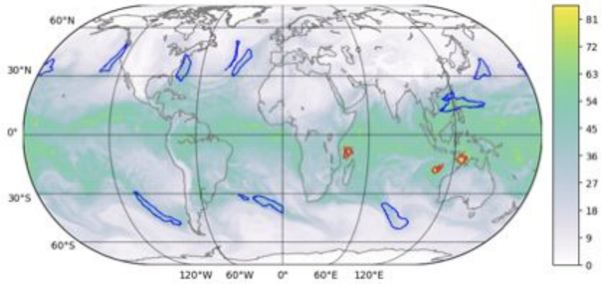
Vinicius Mikuni
NERSC Postdoc

MLPerf HPC Benchmarks



CosmoFlow

- 3D CNN regression on cosmology simulations
- <u>Originally published at SC18</u>
- Target: MAE < 0.124
- Data shape: (128, 128, 128, 4), total size 10.2 TB



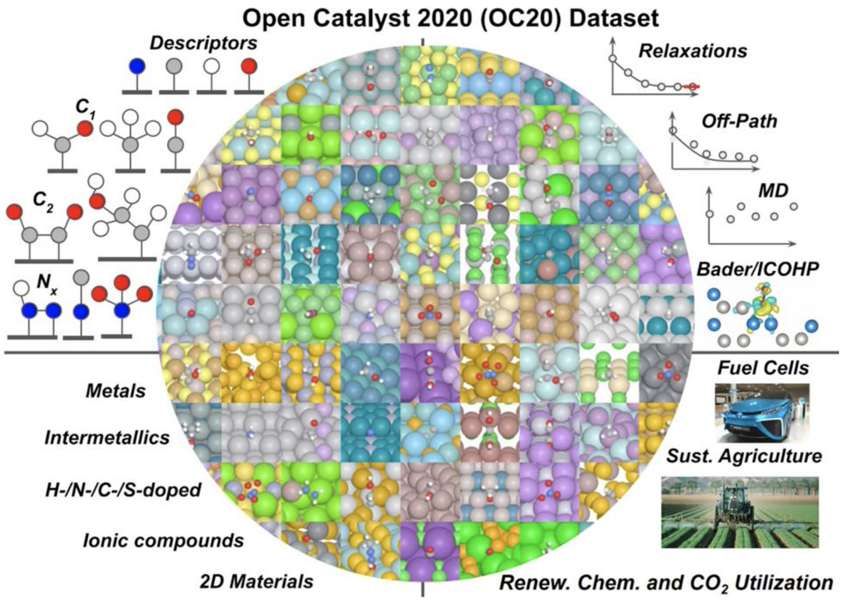
DeepCAM

- 2D CNN segmentation, identifying weather phenomena in climate simulations
- <u>2018 GB prize paper</u>
- Target: IOU > 0.82
- Data shape: (768, 1152, 16), total size 8.8 TB

MLPerf HPC Benchmarks

OpenCatalyst

- GNN predicting energy and forces in atomic catalyst systems (material surface + molecule)
- Dataset: Open Catalyst 2020 (OC20), variable system size, 300GB total size
- Reference model: DimeNet++, 1.8M parameters
- Target: forces MAE < 0.036



https://opencatalystproject.org/