“The Rise of NRP AI/ML Computing Across Diverse Disciplines”

Invited Presentation

Fifth National Research Platform Conference (5NRP)

UC San Diego

March 21, 2024

Dr. Larry Smarr

Founding Director Emeritus, California Institute for Telecommunications and Information Technology;

Distinguished Professor Emeritus, Dept. of Computer Science and Engineering

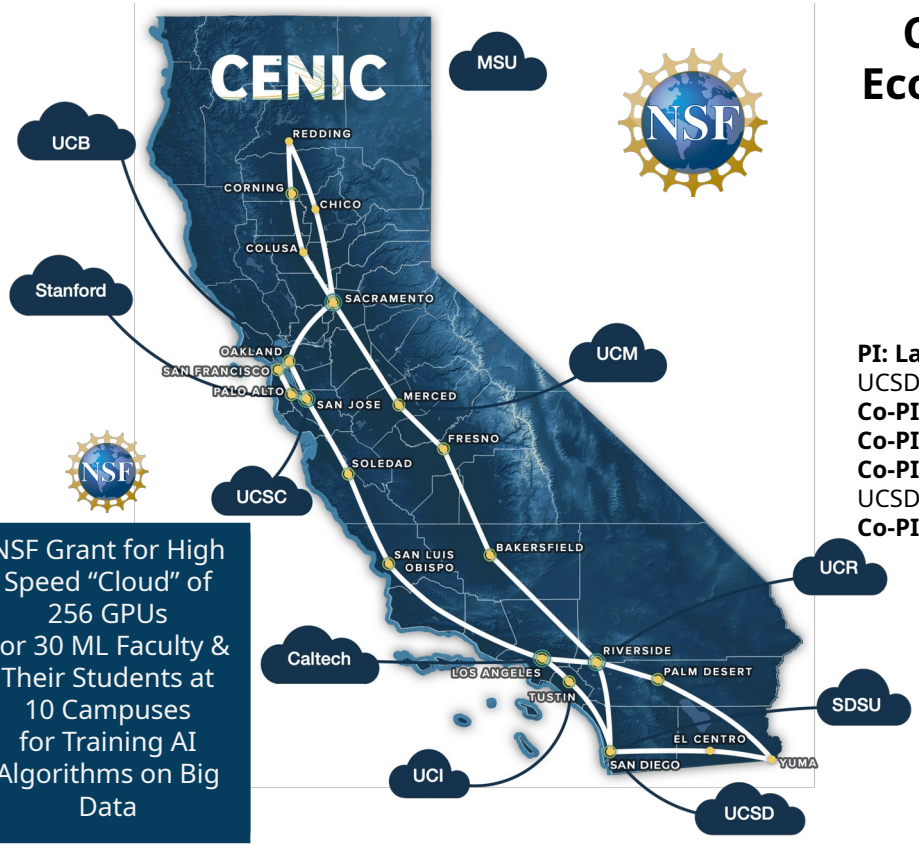
Jacobs School of Engineering, UCSD

http://lsmarr.calit2.net

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

2017-2020: NSF CHASE-CI Grant Adds a Machine Learning Layer Built on Top of the Pacific Research Platform



NSF Grant for High Speed "Cloud" of 256 GPUs For 30 ML Faculty & Their Students at 10 Campuses for Training AI Algorithms on Big Data

CI-New: Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI)

For the Period September 1, 2017 – August 21, 2020

SUBMITTED – January 18, 2017

PI: Larry Smarr, Professor of Computer Science and Engineering, Director Calit2, UCSD
Co-PI: Tajana Rosing, Professor of Computer Science and Engineering, UCSD
Co-PI: Ken Kreutz-Delgado, Professor of Electrical and Computer Engineering, UCSD
Co-PI: Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center, UCSD
Co-PI: Tom DeFanti, Research Scientist, Calit2, UCSD

JACOBS SCHOOL OF ENGINEERING
SDSC UC San Diego QUALCOMM INSTITUTE

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

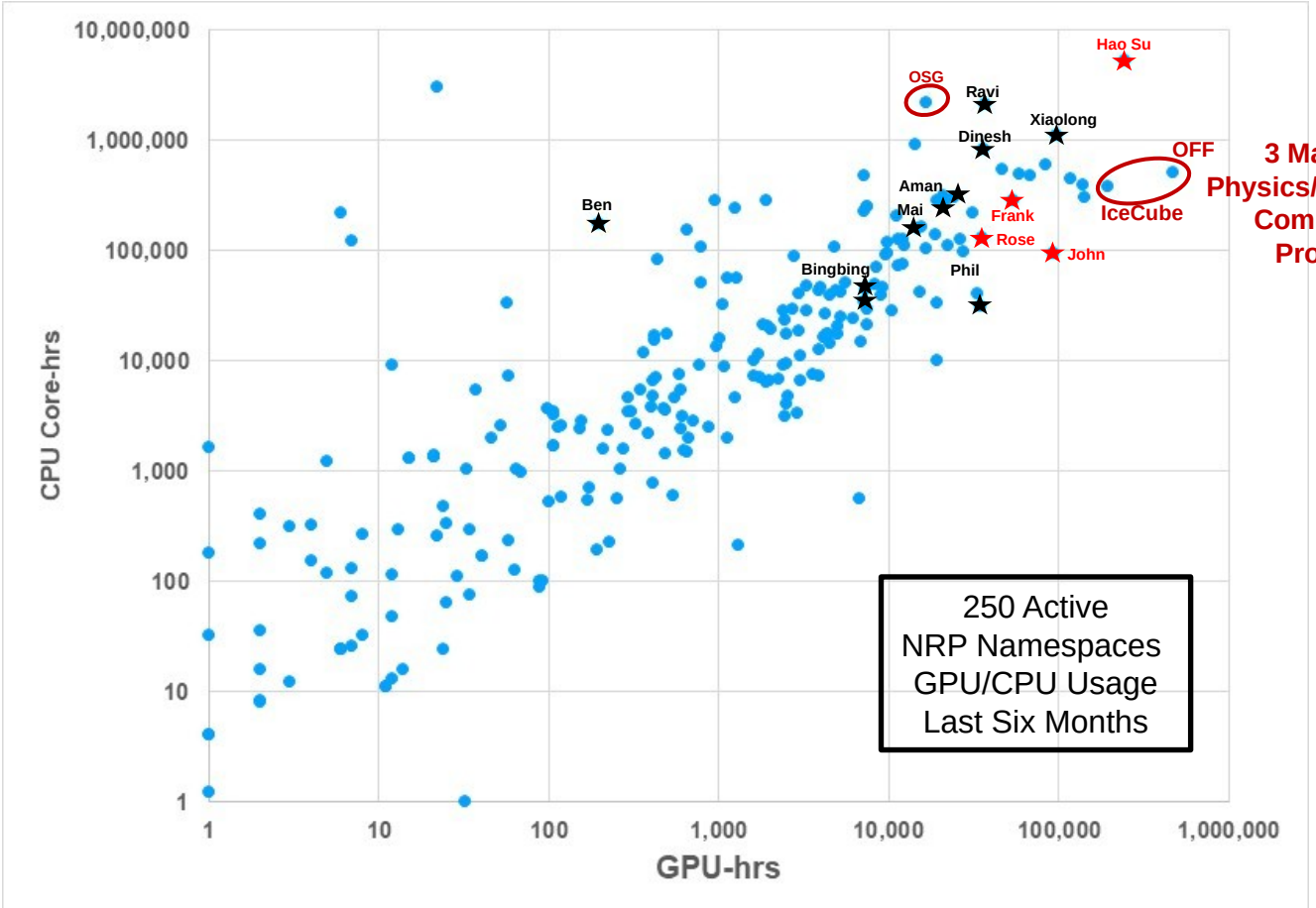
2023: The New Pacific Research Platform Video Shown at 4NRP Highlighted 3 Applications, But Had No Mention of AI/ML



Pacific Research Platform Video:
https://nationalresearchplatform.org/media/pacific-research-platform-video/

2024: By 5NRP
Almost All NRP Namespaces Use AI/ML

5NRP
Speakers:
- ★ Weds/Thurs
- ★ My Talk



3 Massive Physics/Chemistry Community Projects

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

The Rise of Machine Learning ML Applications on the NRP

NRP's Nautilus Cyberinfrastructure Supports a Wide Array of AI/ML Algorithms

Nautilus was Designed to Support Research in 6 Broadly Defined Families of Information Extraction and Pattern Recognition Algorithms that are Commonly Used in AI/ML Research:

1) **Deep Neural Network (DNN) and Recurrent Neural Network (RNN) algorithms including layered networks:**
   - **Convolutional layers (CNNs),**
   - **Generative adversarial networks (GANs), &**
   - **Transformer Neural Networks (e.g., LLMs).**
2) **Reinforcement Learning (RL) and inverse-RL algorithms & related Markov decision process (MDP) algorithms.**
3) **Variational Autoencoder (VAE) and Markov Chain Monte Carlo (MCMC) stochastic sampling**
4) **Support Vector Machine (SVM) algorithms and various ensemble ML algorithms**
5) **Sparse Signal Processing (SSP) algorithms, including Sparse Bayesian Learning (SBL)**
6) **Latent Variable (LVA) Algorithms for source separation algorithms.**

NRP NATIONAL RESEARCH PLATFORM

Source: CHASE-CI Proposal

NRP Nautilus

Today's Over 1000 Nautilus Namespaces Have Utilized Many of These Algorithms

The Great Majority of Nautilus AI/ML Namespaces are Using Some Form of NNs or RL

- **For NNs PyTorch, TensorFlow, and Keras are the preferred (in that order) open-source deep learning (DL) frameworks used on Nautilus.**
- **Our AI/ML researchers use different subtypes of DNNs, including:**
  - **Deep Belief Networks (DBN),**
  - **Quantum NNs (QNN),**
  - **Graph NNs (GNNs) and**
  - **Long Short-Term Memory (LSTM) RNNs-specifically designed to handle sequential data, such as time series, speech, and text.**
- **Nautilus namespaces use RL and inverse-RL algorithms in many areas of dynamic decision-making, robotics, and human/robotic transfer learning.**

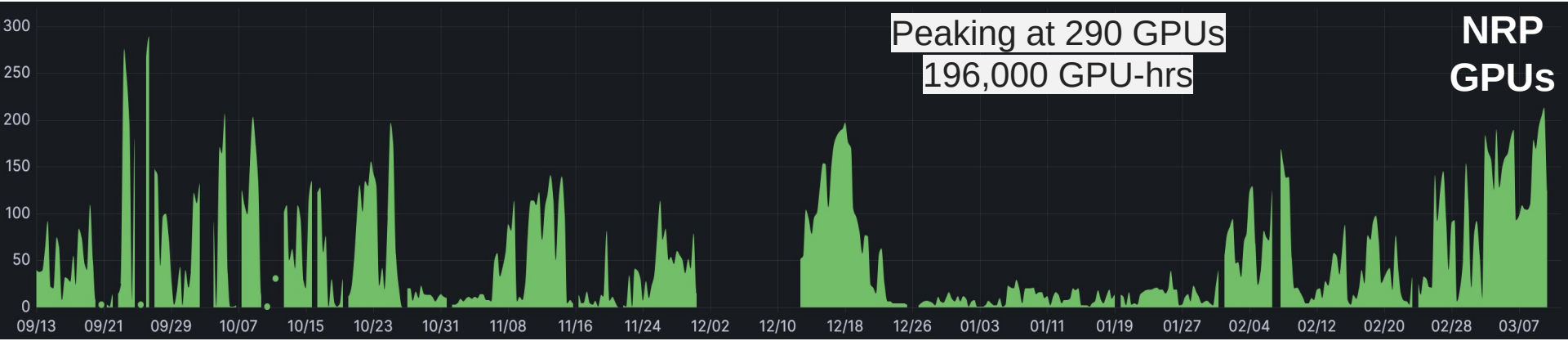
Nautilus Namespaces with Descriptions:
https://portal.nrp-nautilus.io/namespaces-g

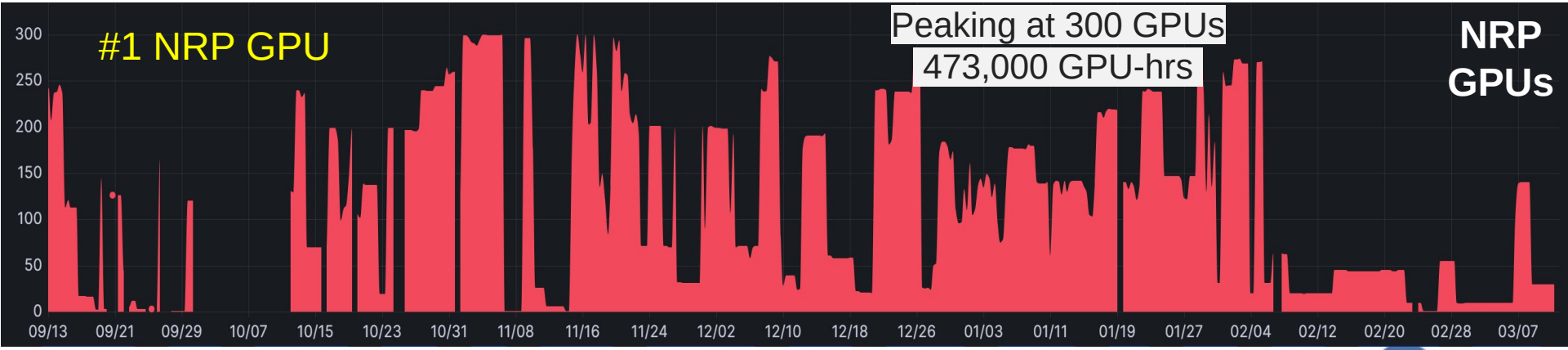
NRP's Largest GPU-Consuming AI/ML Researchers Point to the Rapid Growth of Transformer NNs

Transformer NNs Have Become the Default Architecture for Applications Involving Images, Sound, or ext

- A growing number of NRP namespaces are using Transformer-based Large Language Models (LLMs), such as GPT, LLaMa, and BERT in Natural Language Processing (NLP), or Vision Language Models, such as CLIP and ViT, for image understanding research.
- Also popular are Generative models, such as GANs and Diffusion models, which are prevalent in data synthesis, such as for text to image generation, like Stable Diffusion.
- Finally, we see many namespaces working in fields such as Learning for Dynamics and Control (L4DC), Computer Vision (CV), and Trustworthy ML.

Namespace OpenForceField Surpasses Namespace osg-icecube GPU Usage Over Last 6 Months





NRP NATIONAL RESEARCH PLATFORM

Namespaces osg-icecube, openforcefield

NRP Nautilus

OpenForceField Uses OPEN Software, OPEN Data, OPEN Science and PRP to Generate Quantum Chemistry Datasets for Druglike Molecules



OPEN SOFTWARE

Automated infrastructure enables rapid experimentation with minimum human intervention



OPEN DATA

Access to large, high quality experimental and quantum chemical data facilities easy curation of balanced train / test sets



OPEN SCIENCE

Exploring new force field science:

hypothesis –

build software –

train - test - iterate

is now almost routine

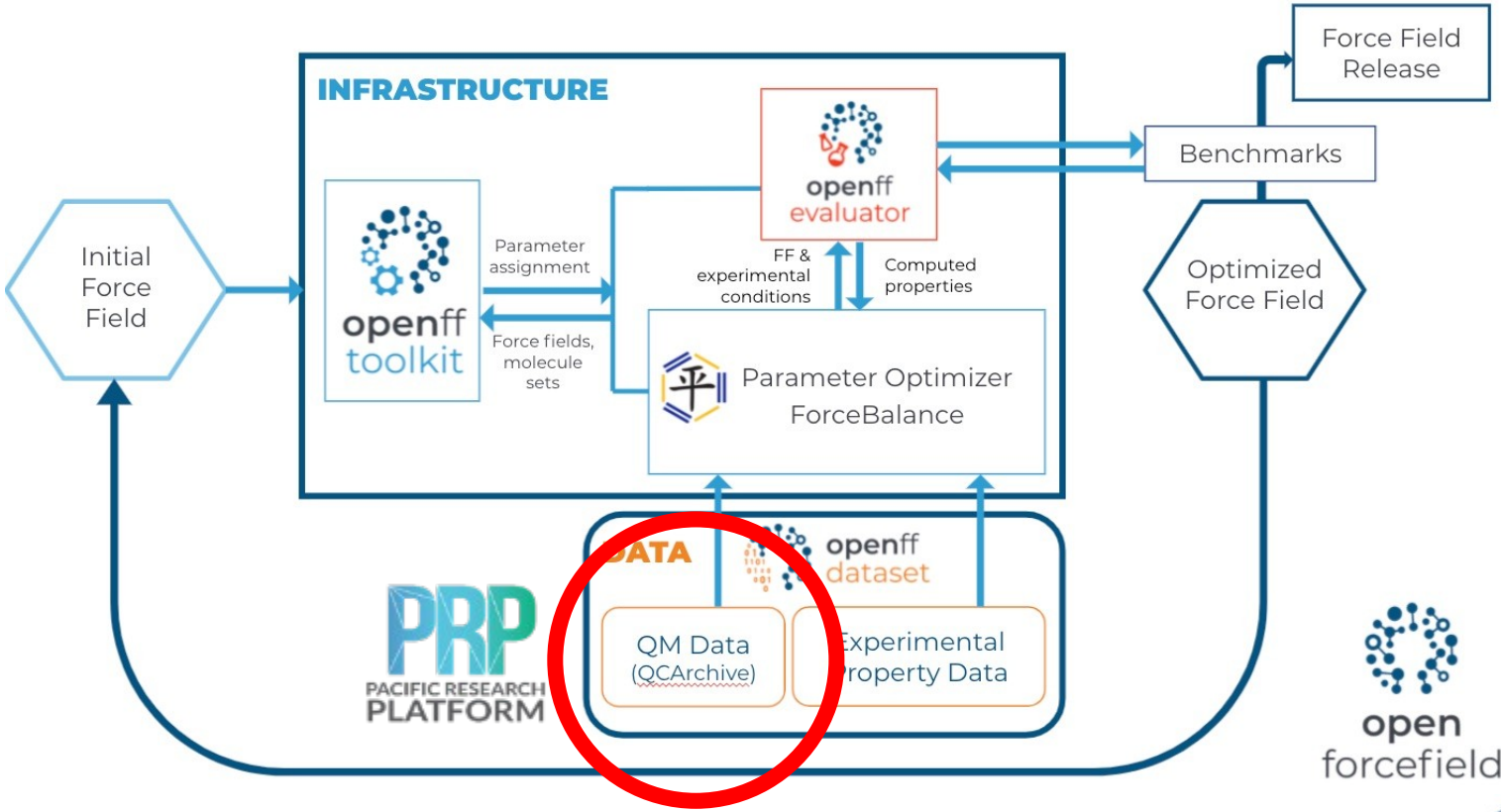
OFF Open-Source Models are Used in Drug Discovery, Including in the COVID-19 Computing on Folding@Home.

www.openforcefield.

NRP NATIONAL RESEARCH PLATFORM

open forcefield

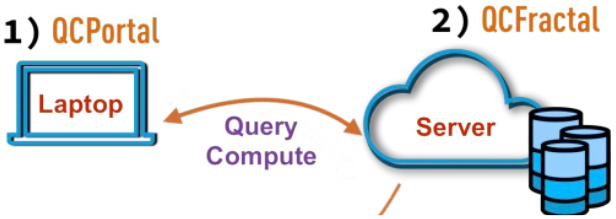
OFF Runs Quantum Mechanical Computations on Many Molecules to Determine Their Optimized Force Fields



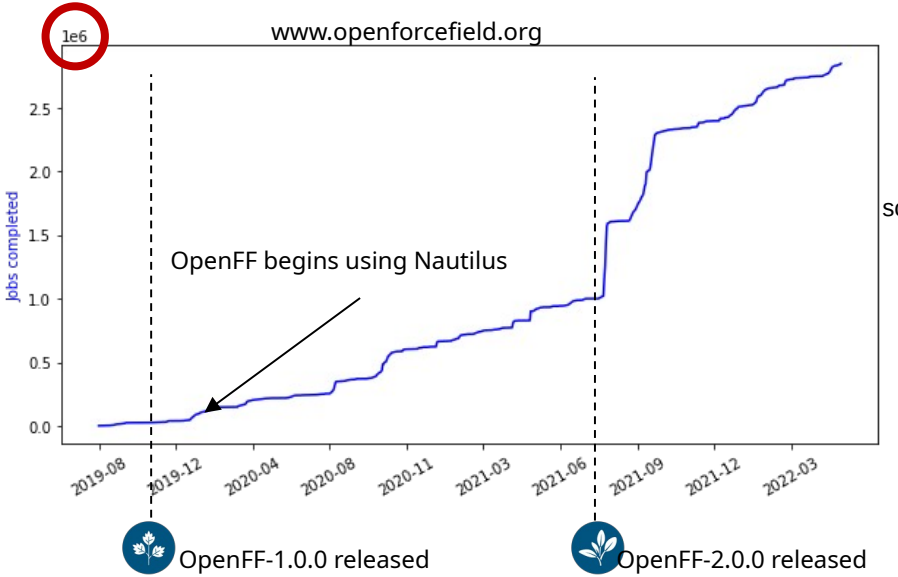
PRP is Capable of Running Millions of Quantum Chemistry Workloads

QCArchive
A MolSSI Project
VIRGINIA TECH





We run "workers" that pull down QC jobs for computation from a central project queue. These jobs require between minutes and hours, and results are uploaded to the central, public QCArchive server. Workers are deployed from Docker images and scheduled on PRP's Kubernetes system. Due to the short job duration, these deployments can still be effective if interrupted every few hours.



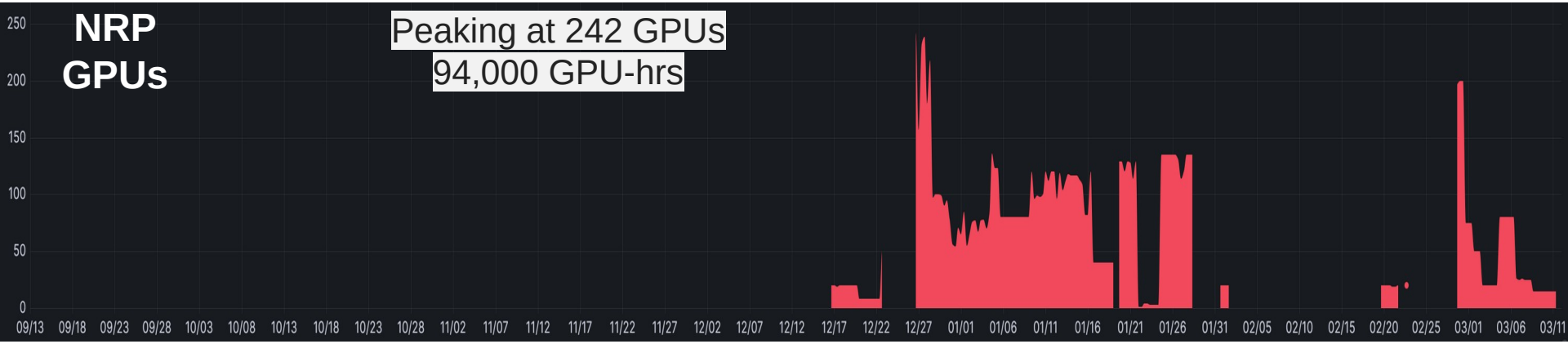
50% of OFF compute is run on Nautilus.

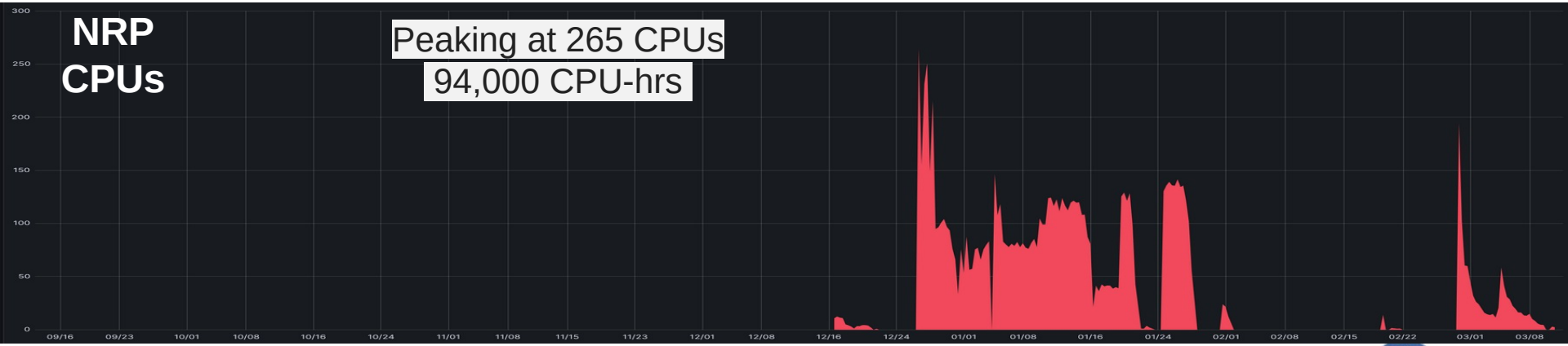
PRP PACIFIC RESEARCH PLATFORM

NRP NATIONAL RESEARCH PLATFORM

open forcefield

John Chodera, Memorial Sloan-Kettering Cancer Center
GPU/CPU Usage Per Day, Last 6 Months



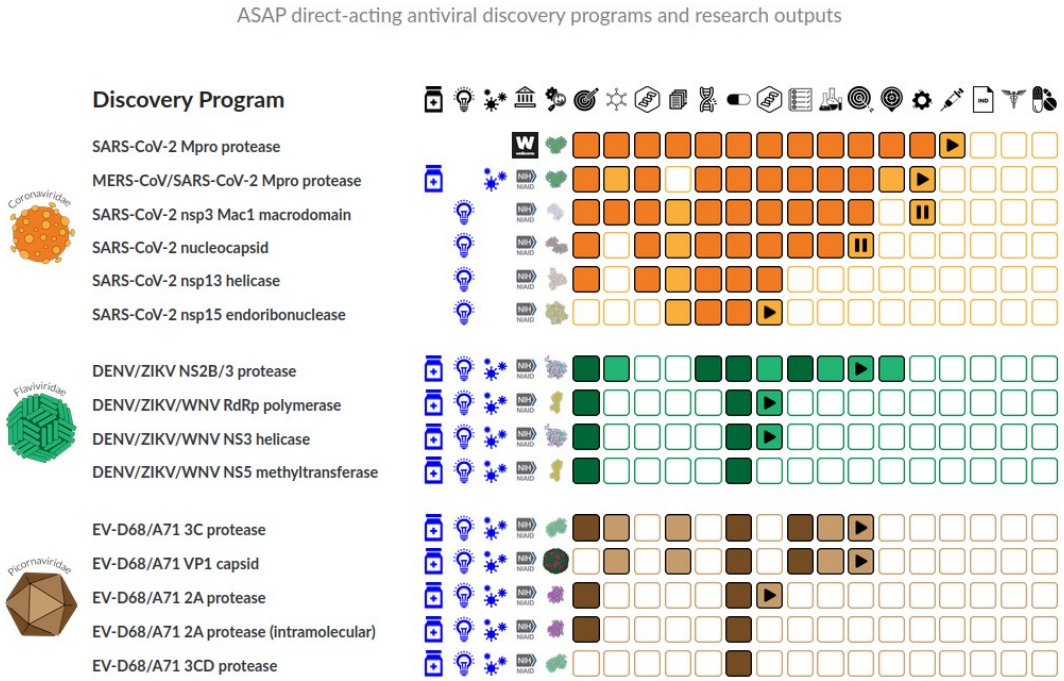


NRP NATIONAL RESEARCH PLATFORM

Namespace choderalab

NRP Nautilus

The AI-driven Structure-Enabled Antiviral Platform (ASAP) is a $68M NIH-Funded Open Science Drug Discovery Effort



ASAP uses AI/ML and computational chemistry to accelerate structure-based, open science antiviral drug discovery and deliver oral antivirals for pandemics with the goal of global, equitable, and affordable access.

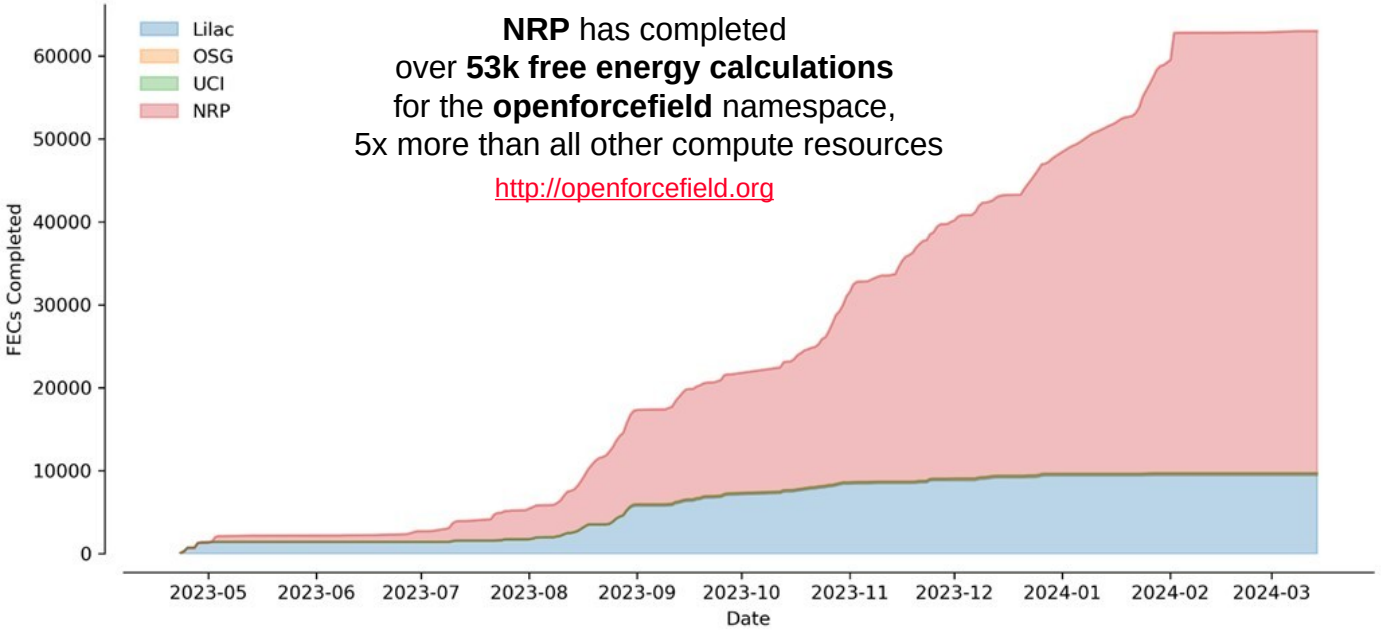
ASAP
Accessible Antivirals to Prevent Pandemics

https://asapdiscovery.org/

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

OpenFF Uses Alchemiscale On NRP To Assess Progress In Improving The Accuracy Of Biomolecular Forcefields Against Current Best-Practice Methods



NRP has completed
over **53k free energy calculations**
for the **openforcefield** namespace,
5x more than all other compute resources

http://openforcefield.org

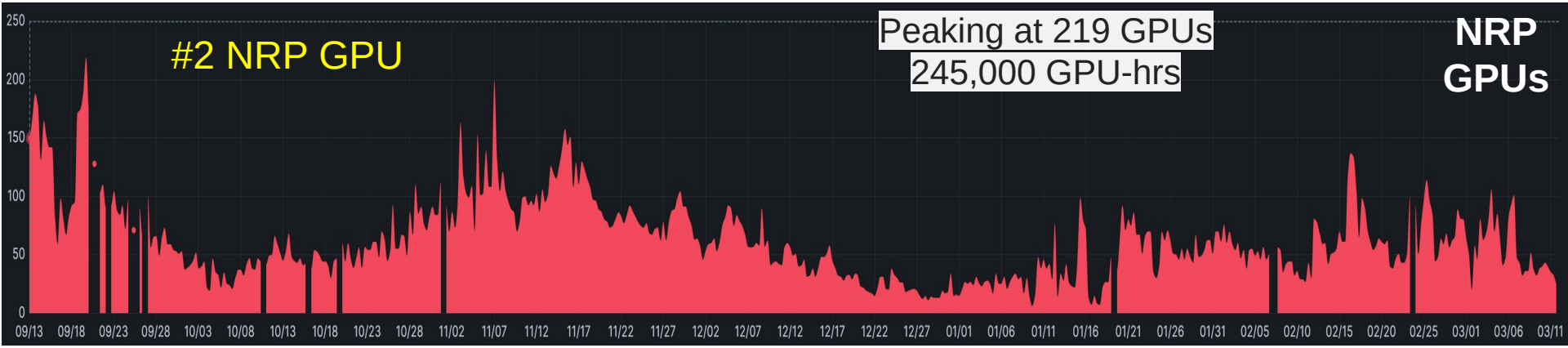
ALCHEMISCALE

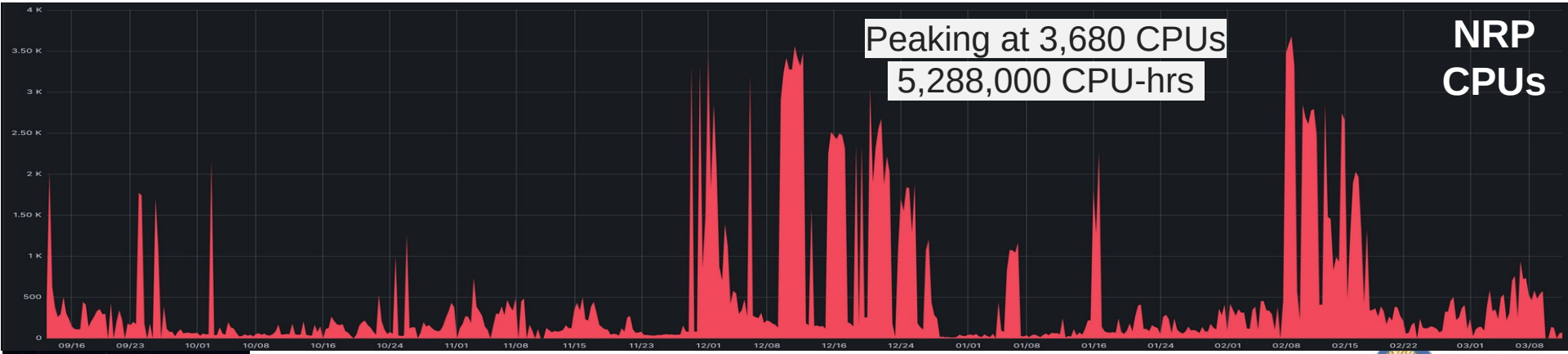
https://docs.alchemiscale.org/

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

Hao Su, UC San Diego
GPU/CPU Usage Per Day, Last 6 Months





NRP NATIONAL RESEARCH PLATFORM

Namespace ucsd-haosulab

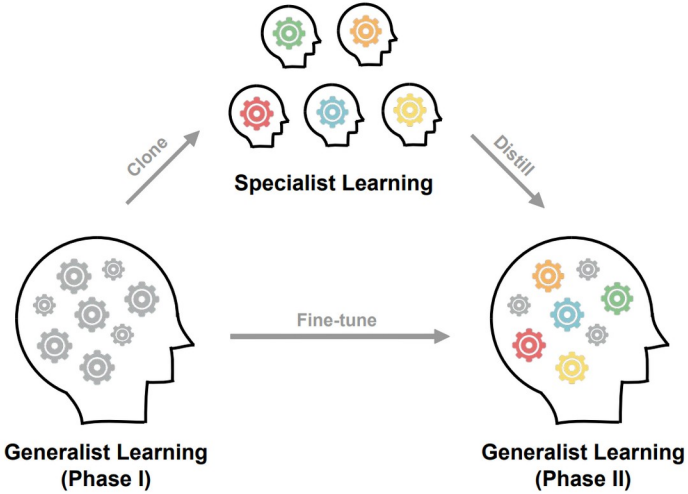
NRP Nautilus

A Major Project in UCSD's Hao Su Lab is Large-Scale Robot Learning

- **We Build A Digital Twin of The Real World in Virtual Reality (VR) For Object Manipulation**
- **Agents Evolve In VR**
  - **Specialists (Neural Nets) Learn Specific Skills by Trial and Error**
  - **Generalists (Neural Nets) Distill Knowledge to Solve Arbitrary Tasks**

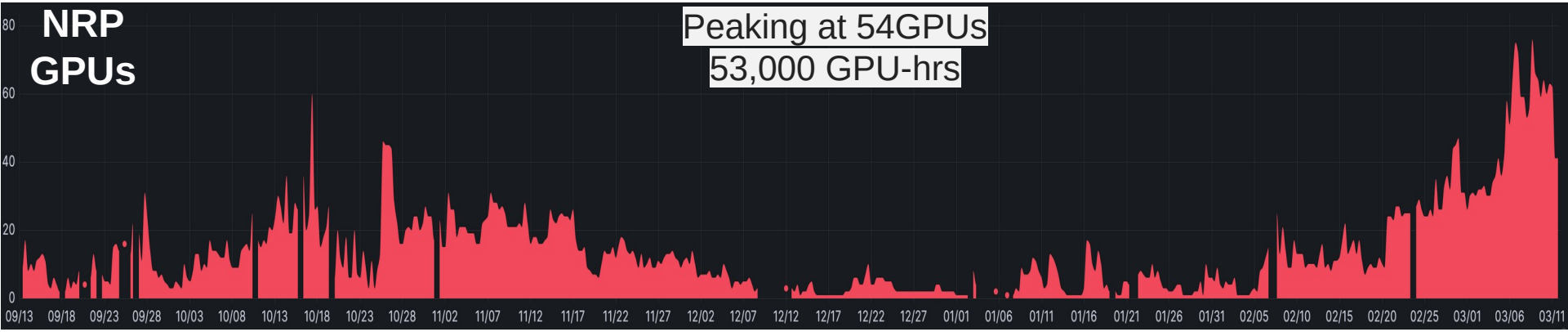
NRP

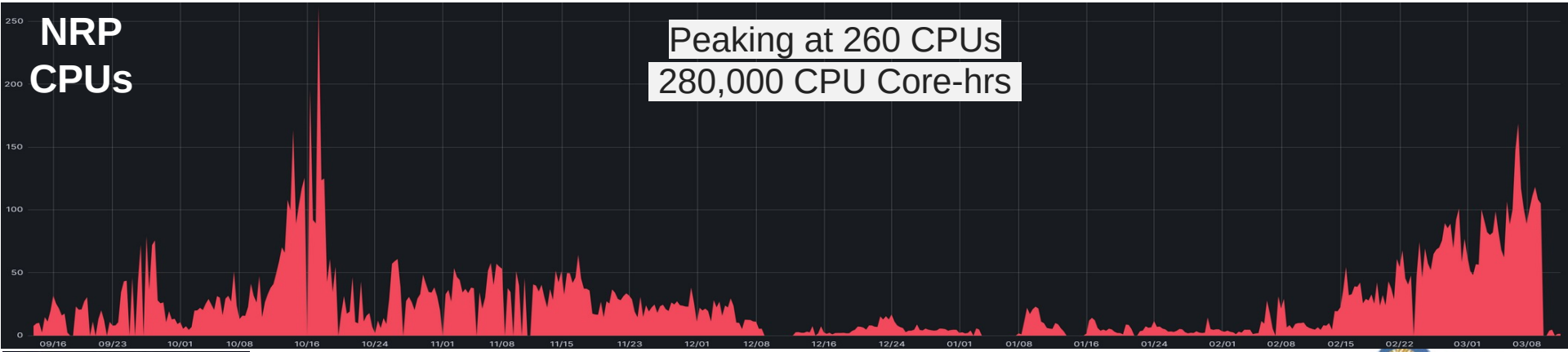
- On **Nautilus**:
  - *Hundreds* of specialists have been trained
  - Each specialist is trained in *millions* of environment variants
  - ~10,000 GPU hours per run



Source: Prof. Hao Su, UCSD

Frank Wuerthwein & Javier Duarte, UC San Diego
GPU/CPU Usage Per Day, Last 6 Months





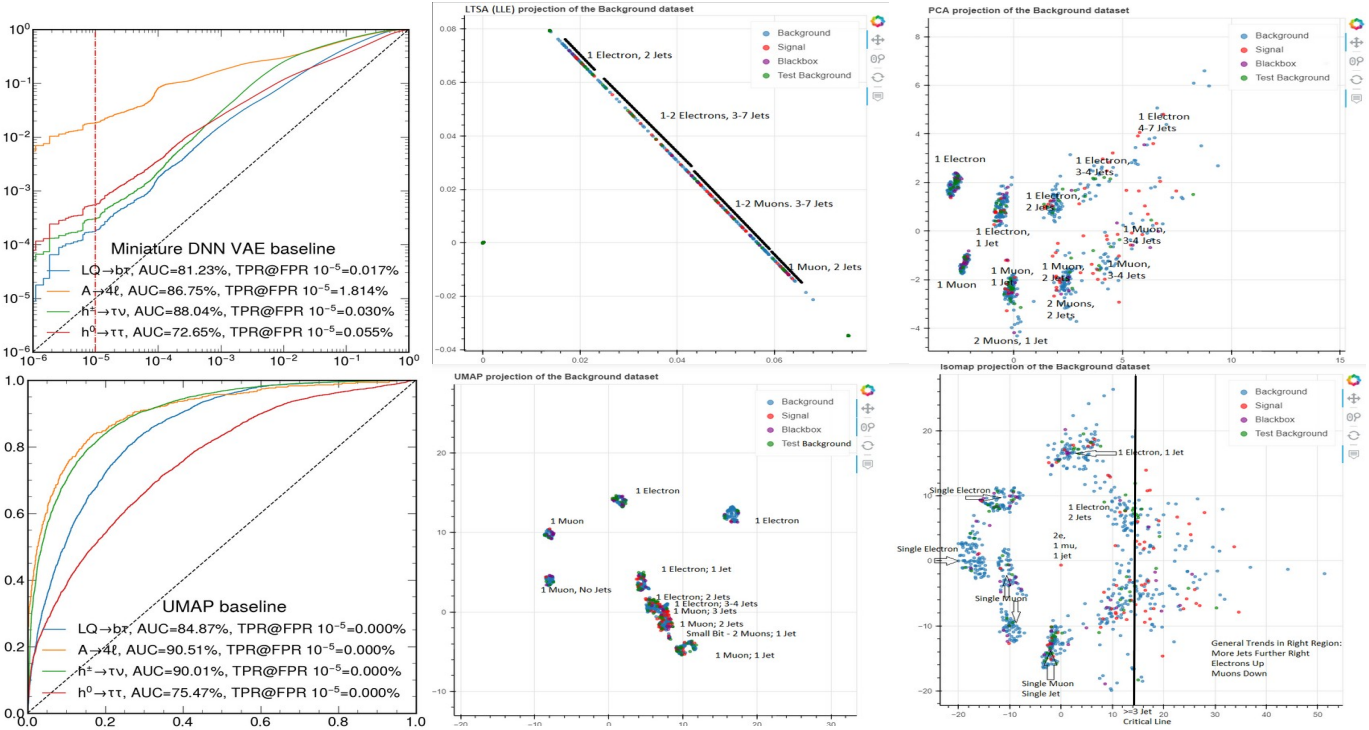
NRP NATIONAL RESEARCH PLATFORM

Namespace cms-ml

NRP Nautilus

Self-Supervised Learning (SSL) for Jet Tagging

Exploring alternative dimension reduction based ML architectures & algorithms for unsupervised anomaly detection using public CMS dataset as a benchmark



Benchmark variational autoencoder vs. UMAP, LDLE, t-SNE …

Rohan Sachdeva, Melissa Quinnan

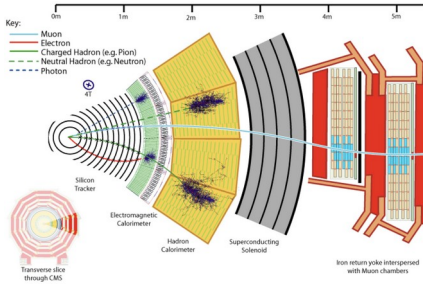
NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

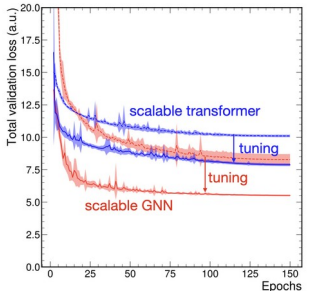
Machine-Learned Particle-Flow Reconstruction

We study *Scalable neural networks* for event reconstruction at current and future colliders.

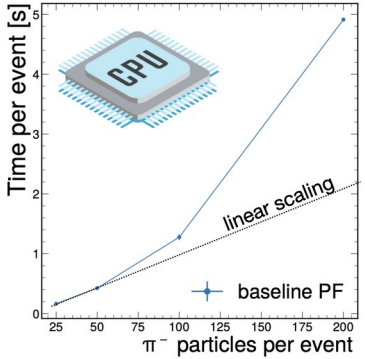
Current and future multilayered detectors need complex data reconstruction → particle flow algorithm



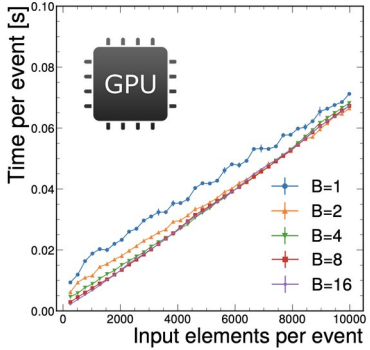
Model hypertuning achieves better than SOTA performance!



Baseline (untuned) algo runs **only on CPU**, scales ~quadratically, **runtime per event is in seconds**.



ML model **scales linearly**, runs in **milliseconds per event** on a consumer 8GB GPU



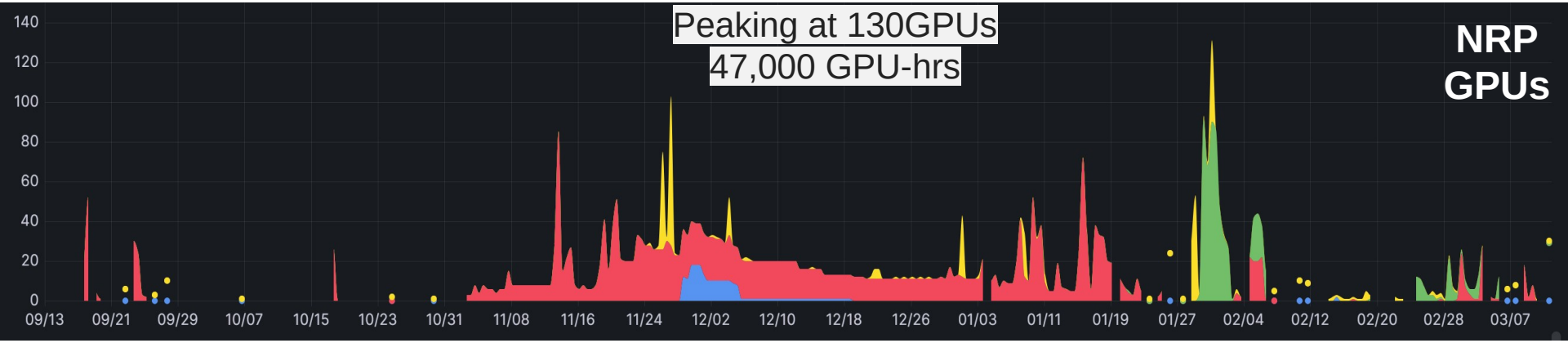


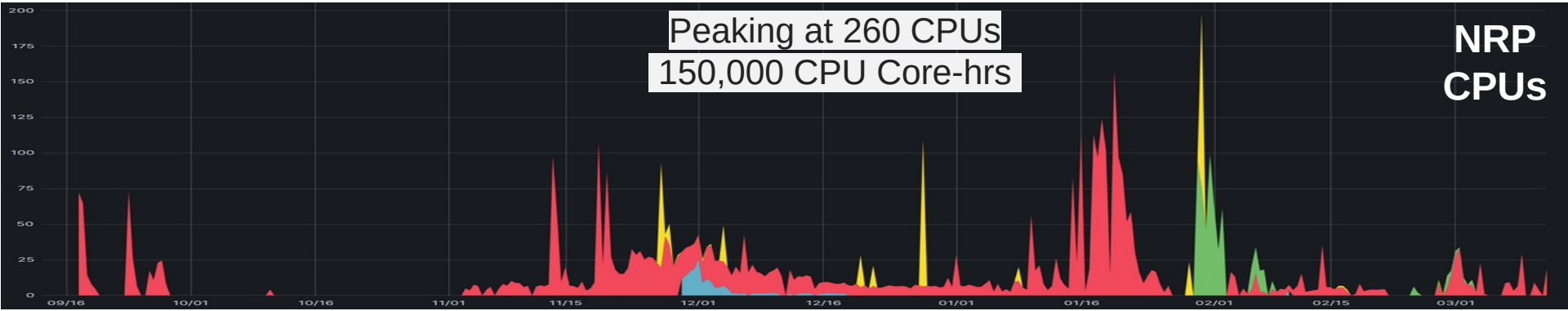
Farouk Mokhtar

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

Rose Yu, UC San Diego
GPU/CPU Usage Per Day, Last 6 Months





Namespaces deep-forecast, deep-point-process, spatiotemperal decision, climate-ml

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

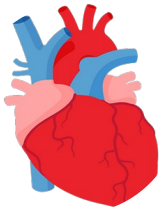
Physics-Guided AI for Large-Scale Spatiotemporal Data: Learning Spatiotemporal Dynamics



climate science



mechanical engineering



biomedical engineering

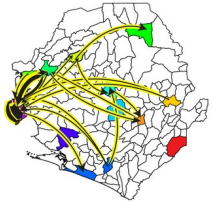


transportation

Spatiotemporal Dynamics



sports analytics



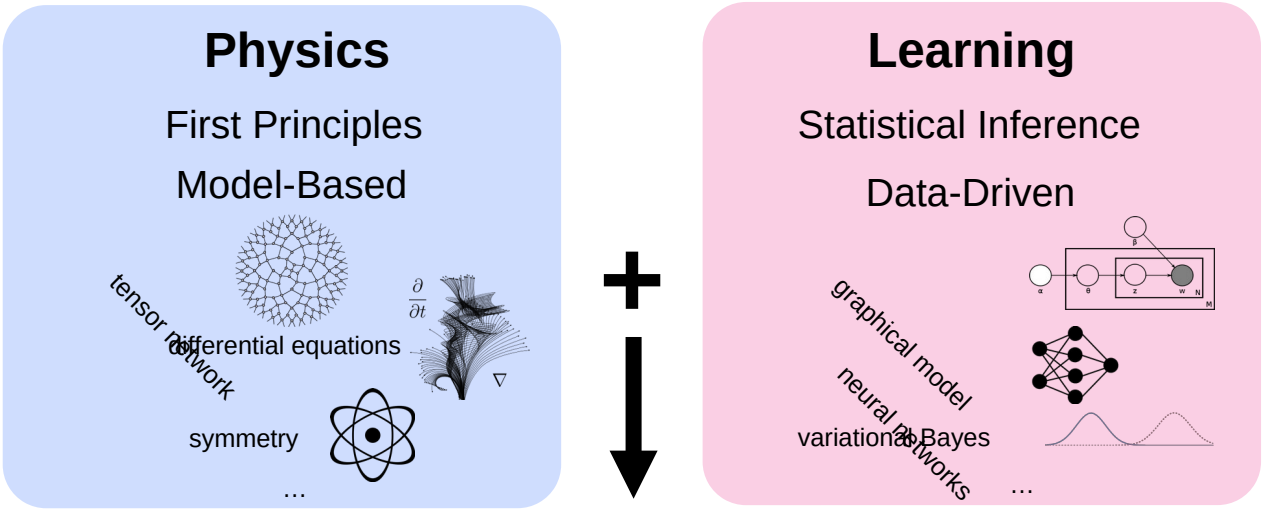
public health



quantum chemistry

Rose Yu
UC San Diego

Physics-Guided AI



Encode Inductive Bias

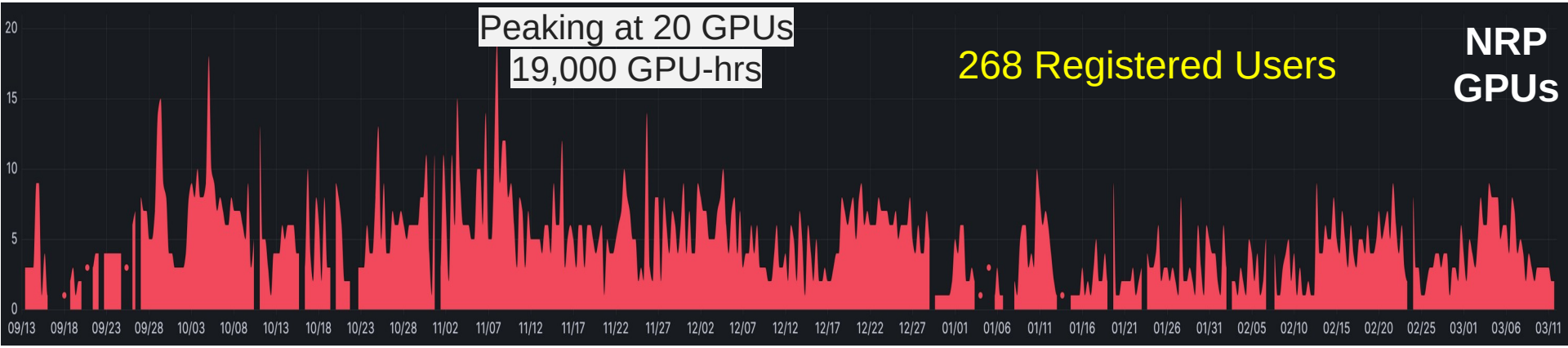
Improve Generalization

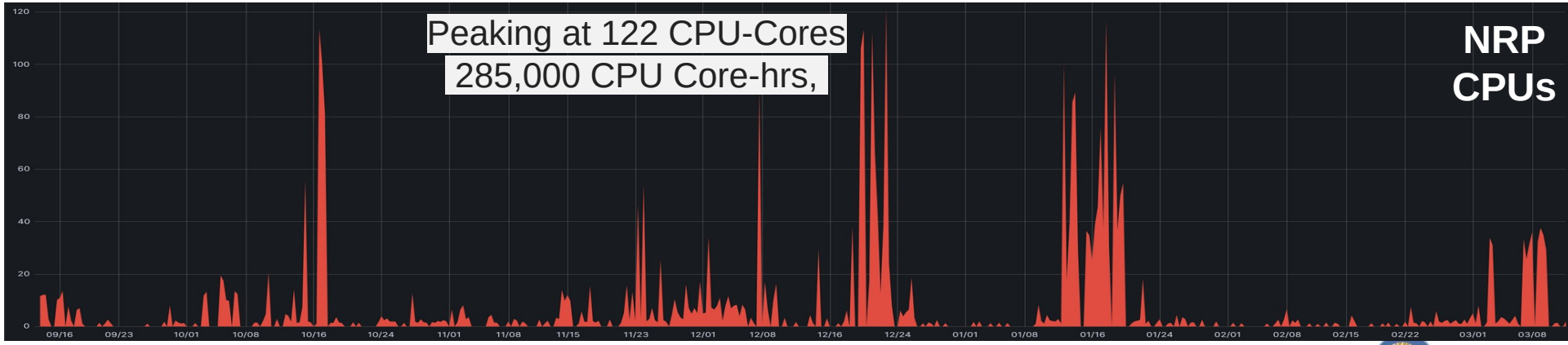
Reduce Sample Complexity

Increase **Trust** in AI

Source: Rose Yu
UC San Diego

Larry Smarr, UC San Diego
GPU/CPU Usage Per Day, Last 6 Months





NRP NATIONAL RESEARCH PLATFORM

Namespace jupyterlab

NRP Nautilus

California State University San Bernardino is an Excellent Example of How to Help Your CSU Faculty and Students To Use NRP



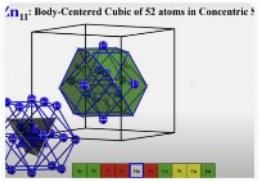
Their Campus HPC Program
Enabled CSUSB Faculty & Students
to Use More NRP GPU-Hours
In the Last 12 Months
Than 8 of the 10 UC Campuses!

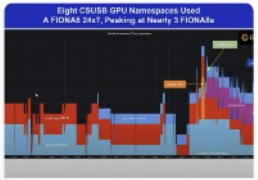


JupyterHub, NRP



ACCESS


Projects

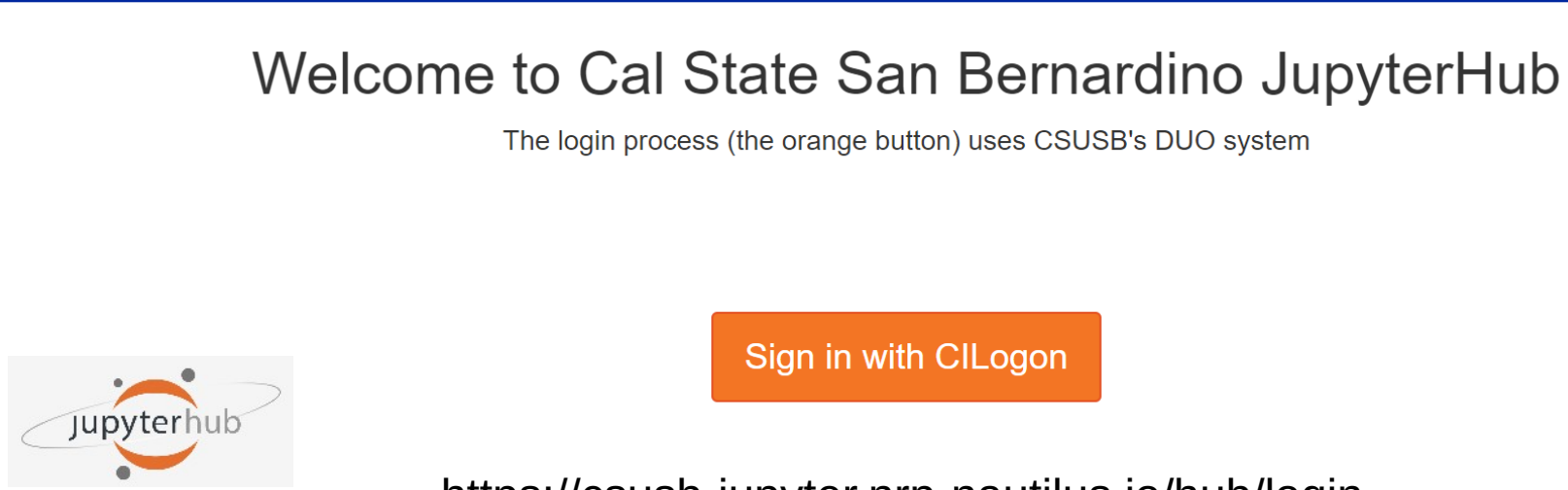

2023 Faculty Showcase

www.csusb.edu/academic-technologies-innovation/xreal-lab-and-high-performance-computing/high-performance-computing

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

A Key Reason CSUSB Has The Largest CSU Nautilus Usage: They Installed and Publicized the JupyterHub “Easy Button”



https://csusb-jupyter.nrp-nautilus.io/hub/login

Over 450 Total Users!

Slide Adapted from Prof. Youngsu Kim

Department of Mathematics
Faculty & Staff

NRP NATIONAL RESEARCH PLATFORM

CSUSB WE DEFINE THE Future

NRP Nautilus

CSUSB Provides Human and On-Line Support For Faculty, Students, and Staff to Easily Use JupyterHub to Access NRP

www.csusb.edu/faculty-center-for-excellence/idat/high-performance-computing/jupyterhub-nrp

CSUSB JupyterHub

- Accessible to anyone with a CoyoteID
- Each user is provided with 50 GB of dedicated storage, expandable upon request
- Resources include large RAM and storage and multiple CPUs and GPUs
- Pre-installed software includes Linux desktop, Python, Matlab, R, Scipy, Pyspark, Tensorflow, a data science package, and SageMath
- Other software can be installed upon request
- Can be used for research and classes

CSUSB JupyterHub for faculty research

- Dedicated to faculty research
- Contact Youngsu Kim or Dung Vu to access this server



National Research Platform

- National Research Platform (NRP) Nautilus Cluster
- Petabyte of distributed Ceph storage
- CPU and GPU computing distributed across 150+ nodes
- Our JupyterHubs are hosted on the PRP server

NRP NATIONAL RESEARCH PLATFORM

CSUSB WE DEFINE THE Future

NRP Nautilus

We are Entering
a New Reality of Ambient AI/ML



3 Weeks Ago Outside Atkinson Hall

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus