Argonne NATIONAL LABORATORY | 75 1946–2021

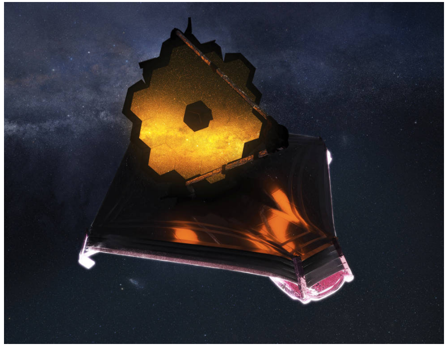
AI4PHYSICS

ELIU HUERTA
Lead for Translational AI
Data Science and Learning Division, Argonne National Laboratory
Department of Computer Science, The University of Chicago

ALCF Intro to AI-driven Science in Supercomputers
Argonne, 1 November 2022

AI FOR SCIENCE
Why

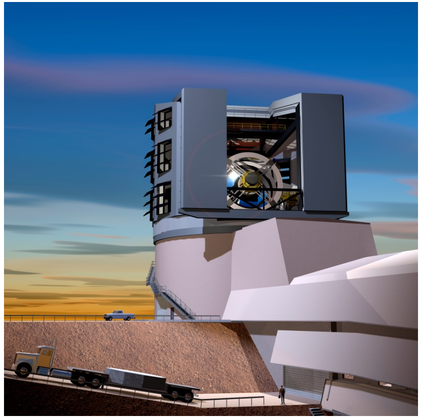
© CERN



© JWST

© LIGO

© SKA



© Rubin Observatory

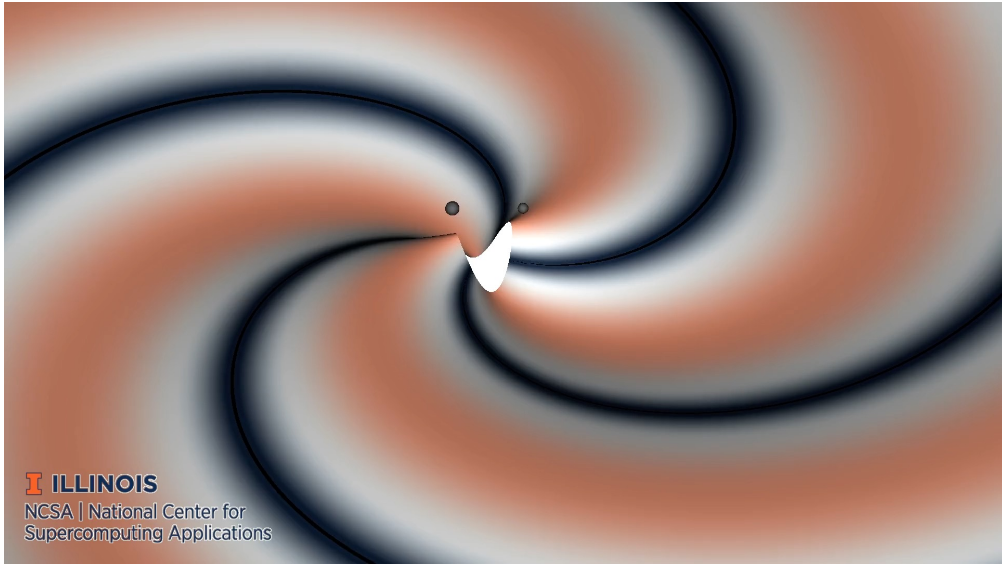
SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

What

Challenges

High velocity datasets

High dimensional parameter space



SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

What

Challenges

Signal processing tools are compute-intensive and poorly scalable

Need to go beyond dedicated supercomputing clusters

IEEE International Conference on e-Science and Grid Computing

BOSS-LDG: A Novel Computational Framework that Brings Together Blue Waters, Open Science Grid, Shifter and the LIGO Data Grid to Accelerate Gravitational Wave Discovery

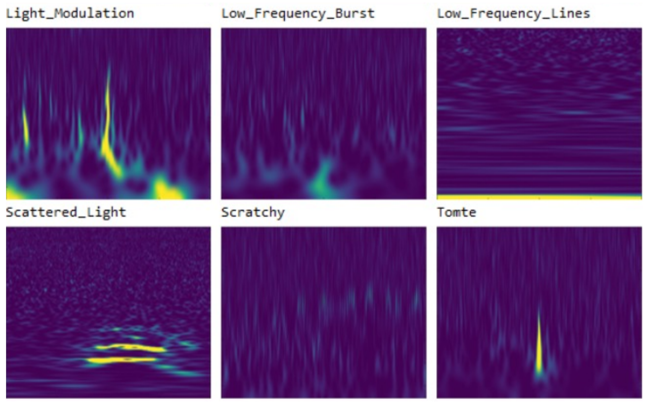
E. A. Huerta[1], Roland Haas[1], Edgar Fajardo[2], Daniel S. Katz[1],
Stuart Anderson[3], Peter Couvares[3], Josh Willis[4], Timothy Bouvet[1]
Jeremy Enos[1], William T. C. Kramer[1], Hon Wai Leong[1] and David Wheeler[1]

[1]NCSA, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
{elihu, rhaas, dskatz, tbouvet, jenos, wtkramer, hwleong, dwheeler}@illinois.edu
[2]University of California, San Diego, La Jolla, California 92093, USA
emfajard@ucsd.edu
[3]LIGO, California Institute of Technology, Pasadena, California 91125, USA
{anderson, peter.couvares}@ligo.caltech.edu
[4]Abilene Christian University, Abilene, Texas 79699, USA
josh.willis@acu.edu

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

How

Grand challenge: identify weak signals embedded in large backgrounds, experimental noise is non-Gaussian and non-stationary



© Gravity Spy Project

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

How

Break down key challenges, and be relentless in addressing them thoroughly

What are the limitations and strengths of state-of-practice algorithms?

Awareness: similar challenges in other disciplines? what can we learn and translate into new domains?

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

How

30 December 2016

MIT Technology Review

Featured Topics Newsletters Events Podcasts Sign in Subscribe

The Best of the Physics arXiv (week ending January 14, 2017)

This week's most thought-provoking papers from the Physics arXiv.

Deep neural networks to enable real-time multimessenger astrophysics

Daniel George and E. A. Huerta

Phys. Rev. D **97**, 044039 – Published 26 February 2018

Novel approach

learn from simulated data, bypass the use of large banks of modeled waveforms; search for signals with a single GPU or mobile phone faster than real-time

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

How

8 November 2017

PHYS ORG

Home / Physics / General Physics

JANUARY 26, 2018

Scientists pioneer use of deep learning for real-time gravitational wave discovery

by University of Illinois at Urbana-Champaign

ELSEVIER

Physics Letters B
Volume 778, 10 March 2018, Pages 64-70

Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data

Daniel George [a, b], E.A. Huerta [b]

Novel approach

learn from real data, bypass the use of large banks of modeled waveforms; search for signals with a single GPU or mobile phone faster than real-time

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Size the problem

Proof of concept

2D (masses of objects)

Training set: 40k signals

Resources: 1 GPU, 3 hrs of training

Enhanced approach

4D (masses and spins of objects)

Training set: 30M signals

Resources: 1 GPU, 1 month of training

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Disrupt again

Convergence of AI and supercomputing

© ORNL

ELSEVIER

Physics Letters B
Volume 808, 10 September 2020, 135628

Physics-inspired deep learning to characterize the signal manifold of quasi-circular, spinning, non-precessing binary black hole mergers

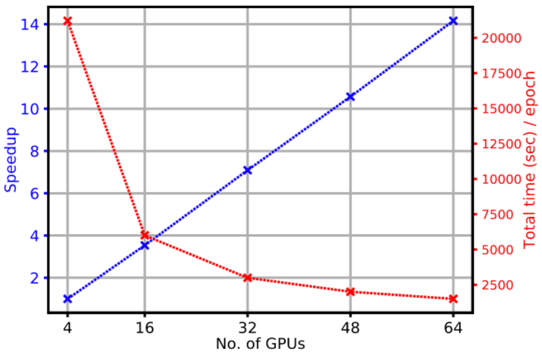
Asad Khan [a, b], E.A. Huerta [a, b, c], Arnav Das [a, d]

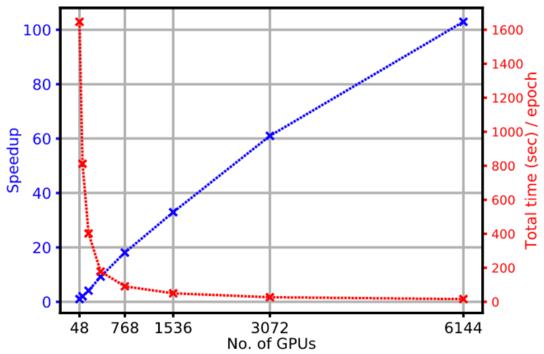
Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.physletb.2020.135628   Get rights and content

Introduce domain knowledge in AI models, harness high performance computing, reduce time-to-insight from months to hours!





SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Disrupt again

Convergence of AI and supercomputing

ELSEVIER

Physics Letters B
Volume 812, 10 January 2021, 136029

Deep learning ensemble for real-time gravitational wave detection of spinning binary black hole mergers

Wei Wei [a, b, c], Asad Khan [a, b, c], E.A. Huerta [a, b, c, d, e], Xiaobo Huang [a, b, f], Minyang Tian [a, b, c]

Show more

Share Cite

https://doi.org/10.1016/j.physletb.2020.136029

Get rights and content

Open access

4D signal manifold

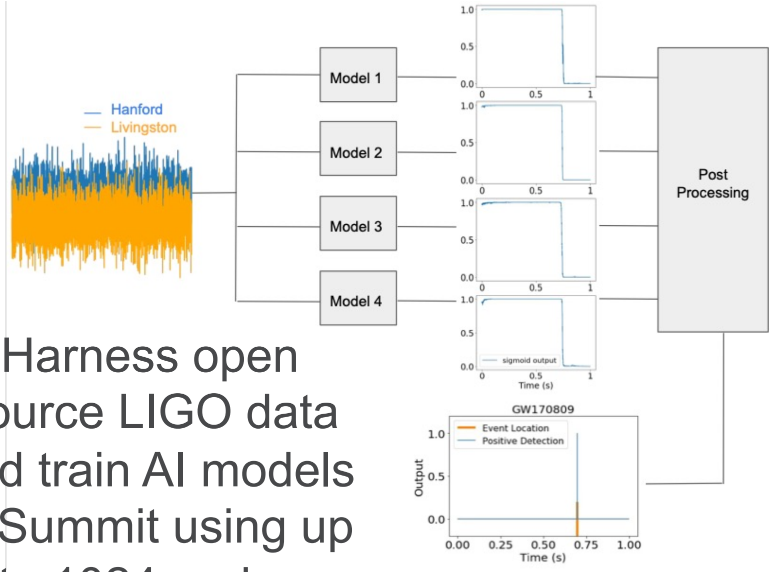
Processes real data faster than real time with 4 NVIDIA V100 GPUs

1 misclassification for every 2.7 days of searched data!

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Production scale approach

Convergence of AI and supercomputing



Harness open source LIGO data and train AI models in Summit using up to 1024 nodes

Optimize AI ensemble for inference, containerize and deploy on
Data and Learning Hub for Science (DLHub)

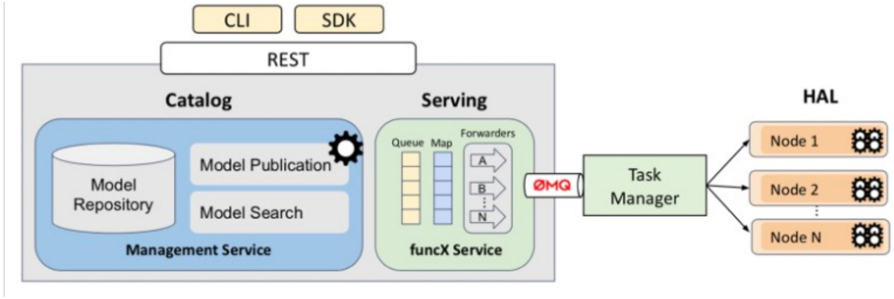
© ORNL

DLHub

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Production scale approach

Convergence of AI and supercomputing



funcX

Leverage **ALCF/JLSE PetrelKube** for model containerization and workflow management

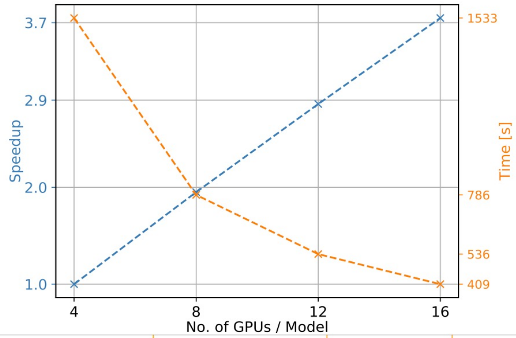
SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

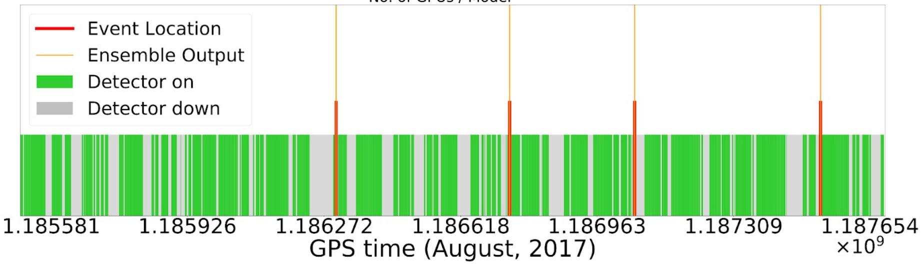
Production scale approach

Convergence of AI and supercomputing

Outcome:
one month's worth of advanced LIGO data processed in 7 minutes

all binary black holes detected with zero misclassifications





SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

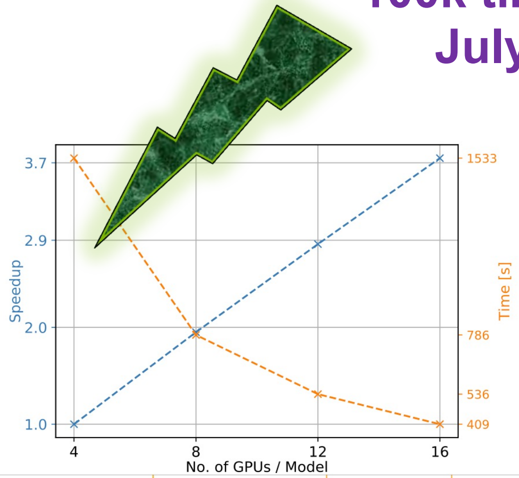
Production scale approach

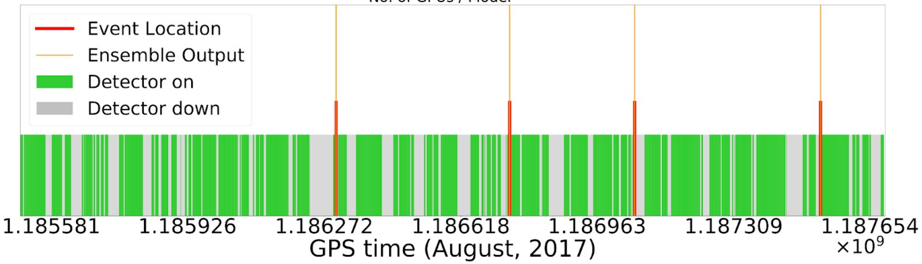
These models have been invoked over 100k times since July 2021!

Convergence of AI and supercomputing

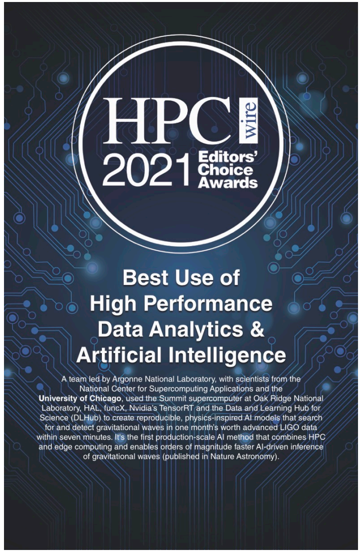
Outcome:
one month’s worth of advanced LIGO data processed in 7 minutes

all binary black holes detected with zero misclassifications





IMPACT



Contributor | Nature Astronomy

BEHIND THE PAPER

From Disruption to Sustained Innovation: Artificial Intelligence for Gravitational Wave Astrophysics

Eliu Huerta

Lead for Translational AI, Argonne National Laboratory

Follow

Published Jul 06, 2021

Article | Published: 05 July 2021

Accelerated, scalable and reproducible AI-driven gravitational wave detection

E. A. Huerta, Asad Khan, Xiaobo Huang, Minyang Tian, Maksim Levental, Ryan Chard, Wei Wei, Maeve Heflin, Daniel S. Katz, Volodymyr Kindratenko, Dawei Mu, Ben Blaiszik & Ian Foster

Nature Astronomy **5**, 1062–1068 (2021) | Cite this article

840 Accesses | **11** Citations | **206** Altmetric | Metrics

MIT Technology Review

Featured Topics Newsletters Events Podcasts Sign in Subscribe

SPACE

3 space science questions that computing is helping to answer

Astronomers are using AI, supercomputing, and the cloud to tackle the universe's biggest mysteries.

By Tatyana Woodall October 27, 2021

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Go the extra mile

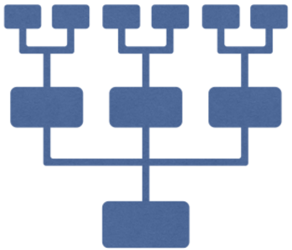
End-to-end GPU-accelerated computing for gravitational wave detection

Distributed training in Summit

Suite of models trained using 192 NVIDIA V100 GPUs within 2 hours

Identical AI architecture but random weight initialization

Model selection



AI models are selected by testing their classification accuracy on long advanced LIGO strain data segments

Inference optimization in HAL NVIDIA DGX

4 AI models are optimized for accelerated inference with NVIDIA TensorRT

TensorRT models retain accuracy of original AI models and provide 3-fold speedup

Accelerated inference in ThetaGPU

TensorRT AI ensemble processes an entire month of advanced LIGO data in 50 seconds by distributing inference over 160 NVIDIA A100 Tensor Core GPUs

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Go the extra mile

frontiers in Artificial Intelligence | Big Data and AI in High Energy Physics

Inference-Optimized AI and High Performance Computing for Gravitational Wave Detection at Scale

Pranshu Chaturvedi[1,2,3*], Asad Khan[1,3,4], Minyang Tian[3,4], E. A. Huerta[1,4,5] and Huihuo Zheng[6]

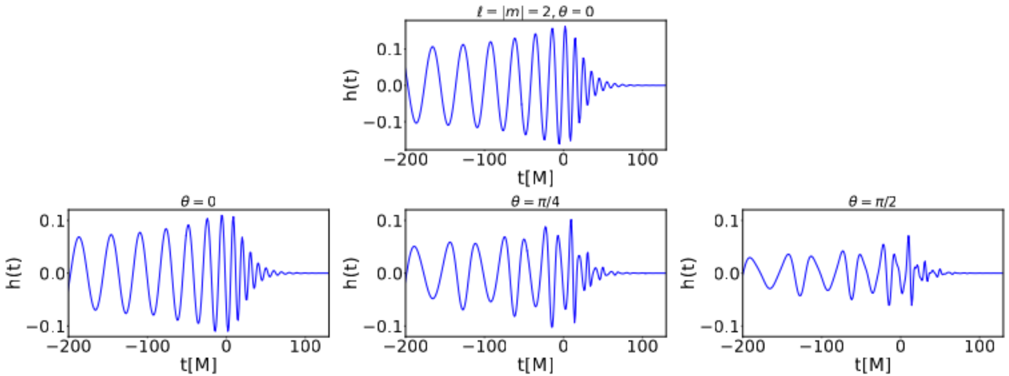
[1]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL, United States
[2]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[3]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[4]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[5]Department of Computer Science, University of Chicago, Chicago, IL, United States
[6]Leadership Computing Facility, Argonne National Laboratory, Lemont, IL, United States

AI-inference for gravitational waves **53,000X faster than real-time**

Using a synthetically enhanced 5 yr-long advanced LIGO dataset, AI ensemble **identified known gravitational wave sources** and reported **one misclassification for every month of searched data**

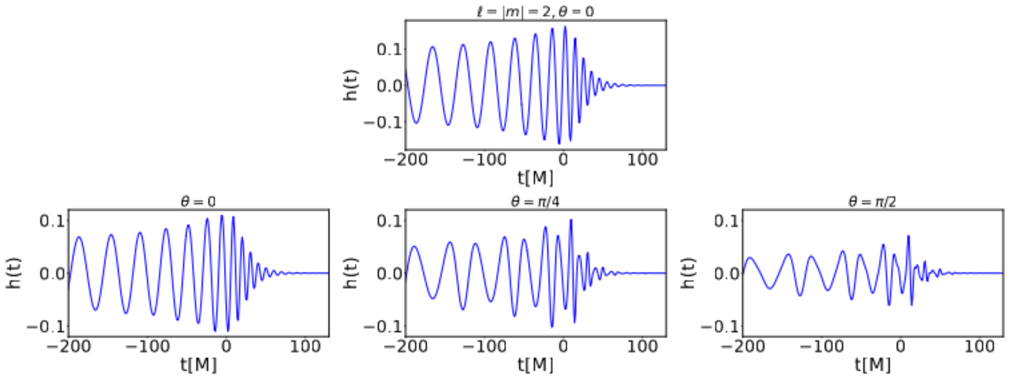
GRAVITATIONAL WAVE REGRESSION

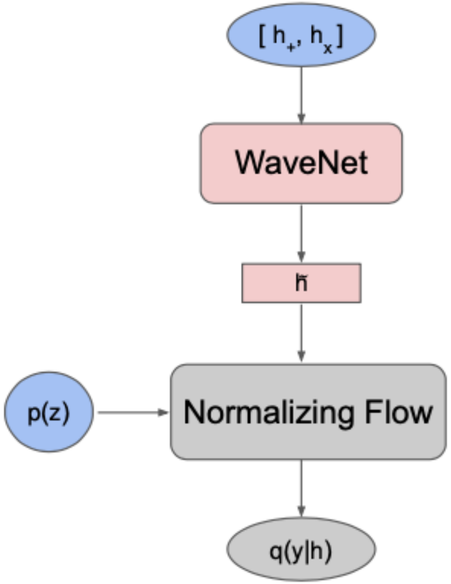
High dimensional signal manifolds



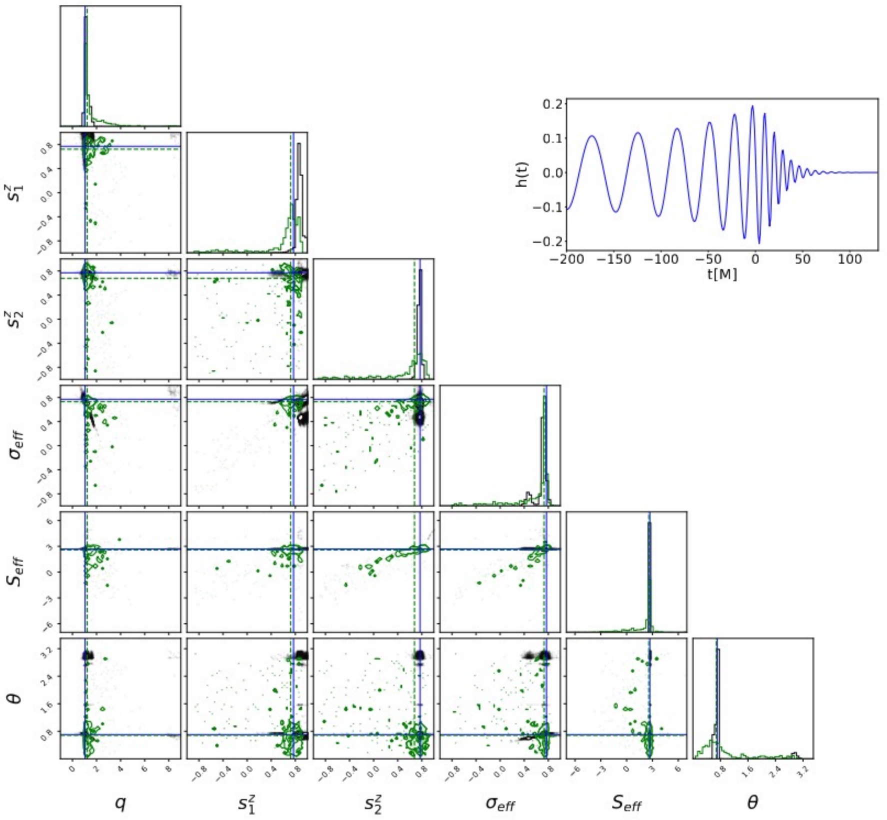
GRAVITATIONAL WAVE REGRESSION

High dimensional signal manifolds





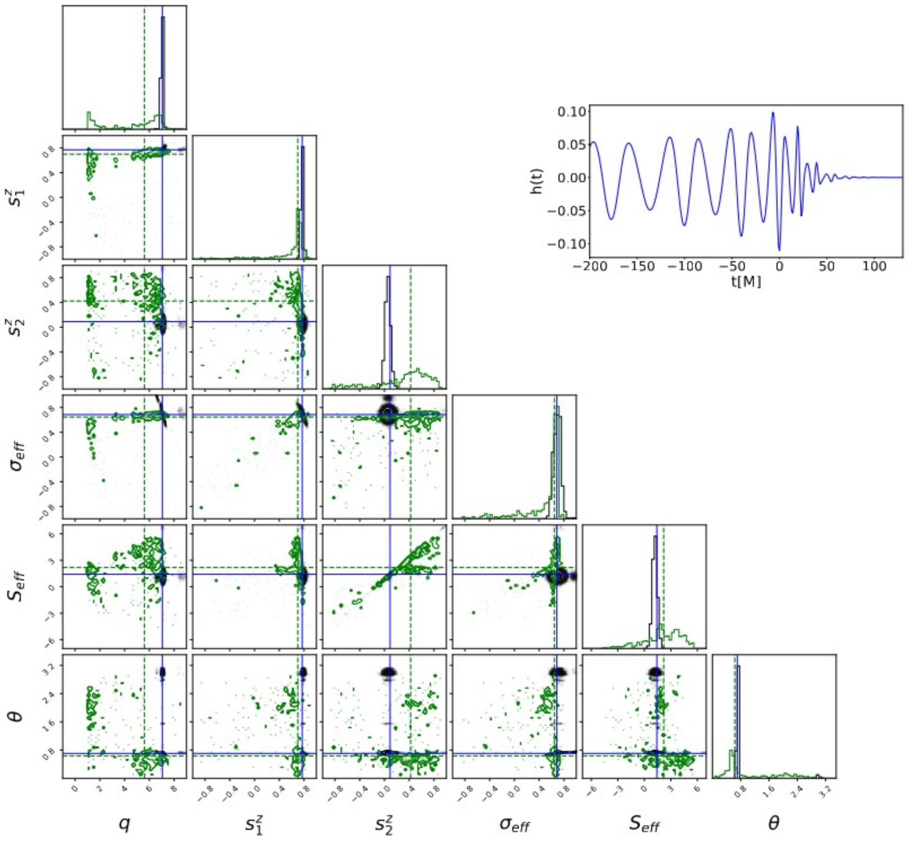
GRAVITATIONAL WAVE REGRESSION



AI posterior distributions (in black), **PyCBC** Inference results (in green), and **ground truth** values (in blue) for an equal mass-ratio binary black hole

AI histograms show the distribution of 100, 000 samples drawn from the posterior.

GRAVITATIONAL WAVE REGRESSION



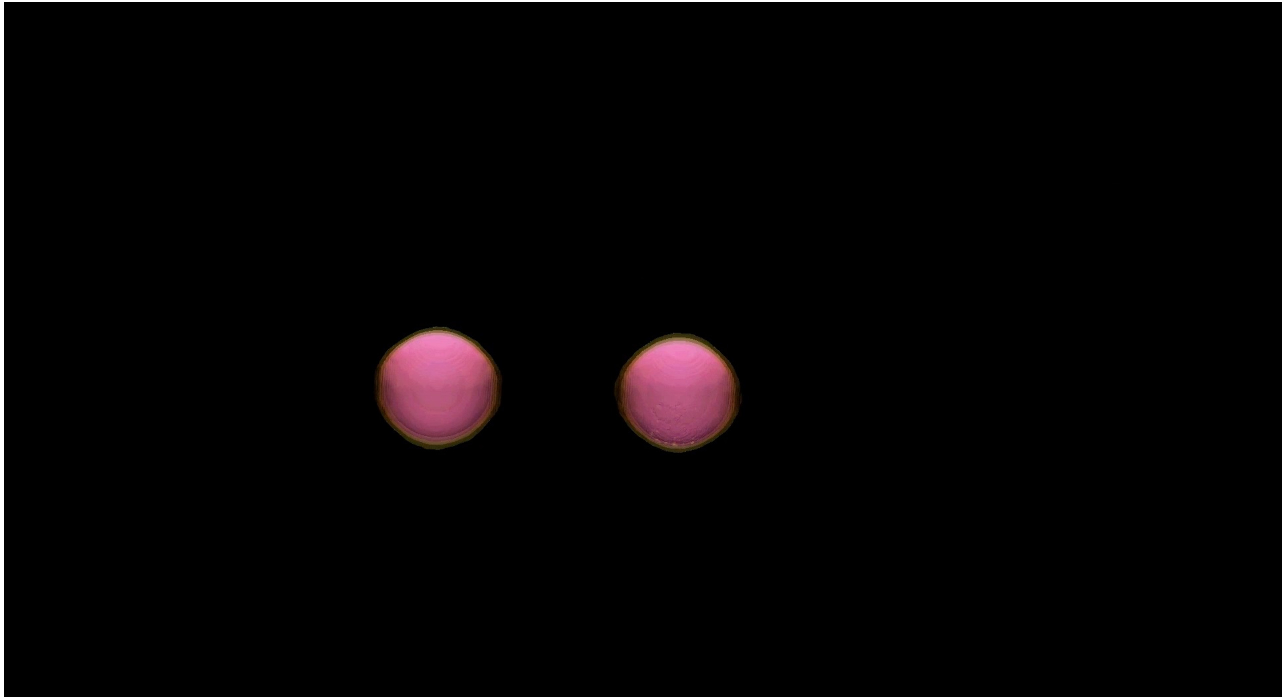
AI posterior distributions (in black), **PyCBC** Inference results (in green), and **ground truth** values (in blue) for an equal mass-ratio binary black hole

AI histograms show the distribution of 100, 000 samples drawn from the posterior.

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Multimessenger sources

Let's turn our attention to compact binary mergers that may emit gravitational, electromagnetic and astro-particle counterparts

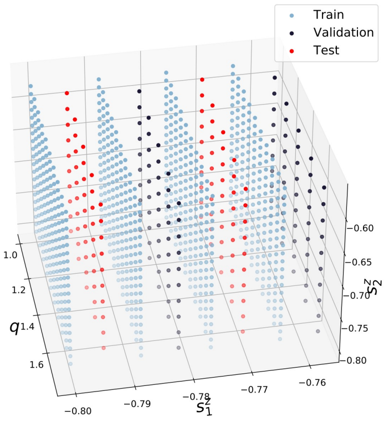


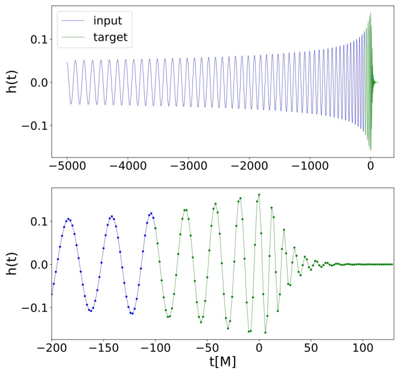
SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Learn physics, forecast non-linear dynamics and dive deep into interpretable AI

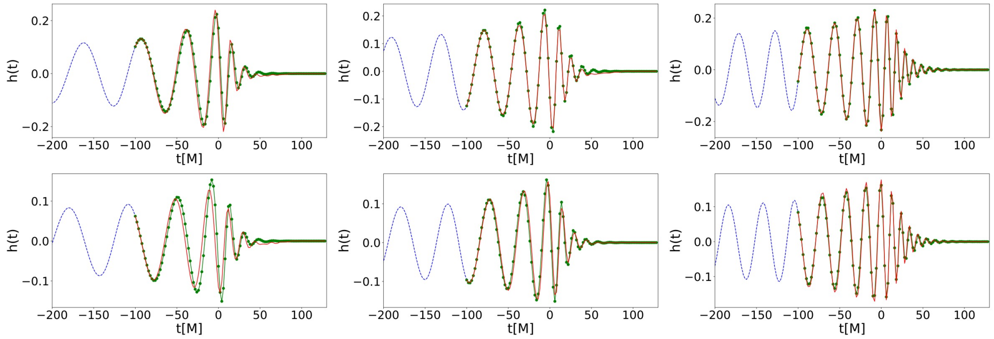
Interpretable AI forecasting for numerical relativity waveforms of quasicircular, spinning, nonprecessing binary black hole mergers

Asad Khan, E. A. Huerta, and Huihuo Zheng
Phys. Rev. D **105**, 024024 – Published 6 January 2022





$$q = 4.24; \quad s_1^z = s_2^z = \{-0.7, 0, 0.7\}$$



$$q = 6.80; \quad s_1^z = s_2^z = \{-0.7, 0, 0.7\}$$

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

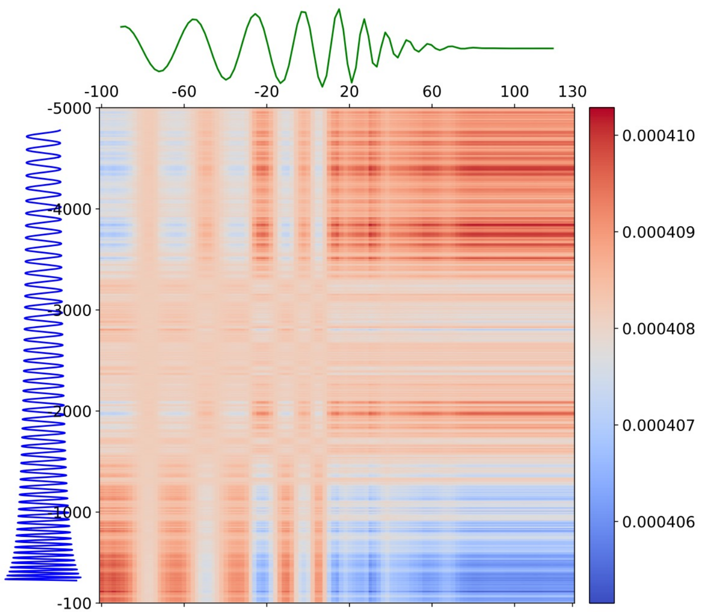
Interpretable AI forecasting for numerical relativity waveforms of quasicircular, spinning, nonprecessing binary black hole mergers

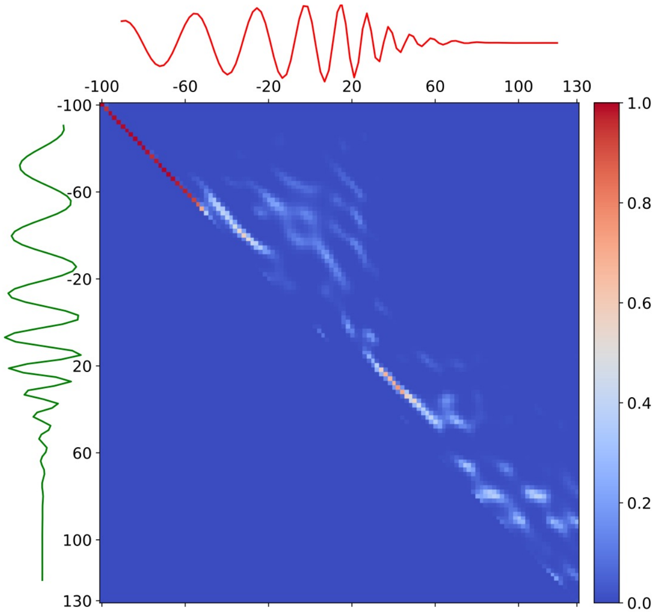
Asad Khan, E. A. Huerta, and Huihuo Zheng

Phys. Rev. D **105**, 024024 – Published 6 January 2022

Learn physics, forecast non-linear dynamics and dive deep into interpretable AI

https://khanx169.github.io/gw_forecasting/interactive_results.html





SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

AI surrogates

Why

Physical processes can be naturally described using partial differential equations (PDEs)

Numerical solvers have been developed to solve complex PDEs with supercomputing platforms

Multi-scale and multi-physics phenomena challenge this paradigm

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

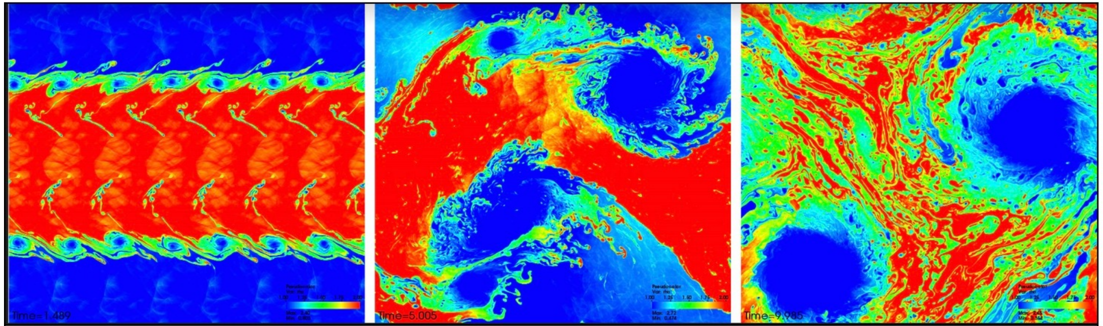
AI surrogates

Artificial neural network subgrid models of 2D compressible magnetohydrodynamic turbulence

Shawn G. Rosofsky and E. A. Huerta
Phys. Rev. D **101**, 084024 – Published 9 April 2020

Artificial Intelligence on XSEDE Systems Is Key to Speeding Simulations of Neutron Star Mergers

By Ken Chiacchia, Pittsburgh Supercomputing Center



The intense magnetic fields accompanying movement of matter from neutron-stars past each other causes increasingly complicated turbulence that is computationally expensive with standard simulation methods. In this time series, a deep learning AI provides a simulation of this process at a fraction of the computing time.

Shawn Rosofsky

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

AI surrogates
Physics informed neural operators

Shawn Rosofsky

$$\frac{\partial(\eta)}{\partial t} + \frac{\partial(\eta u)}{\partial x} + \frac{\partial(\eta v)}{\partial y} = 0\,,$$

$$\frac{\partial(\eta u)}{\partial t} + \frac{\partial}{\partial x}\left(\eta u^2 + \frac{1}{2} g \eta^2\right) + \frac{\partial(\eta u v)}{\partial y} = \nu\left(u_{xx} + u_{yy}\right)\,,$$

$$\frac{\partial(\eta v)}{\partial t} + \frac{\partial(\eta u v)}{\partial x} + \frac{\partial}{\partial y}\left(\eta v^2 + \frac{1}{2} g \eta^2\right) = \nu\left(v_{xx} + v_{yy}\right)\,,$$

$$\text{with} \quad \eta(x, y, 0) = \eta_0(x, y),\; u(x, y, 0) = 0,\; v(x, y, 0) = 0, \qquad x, y \in [0, 1),\; t \in [0, 1]$$

SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Physics informed neural operators

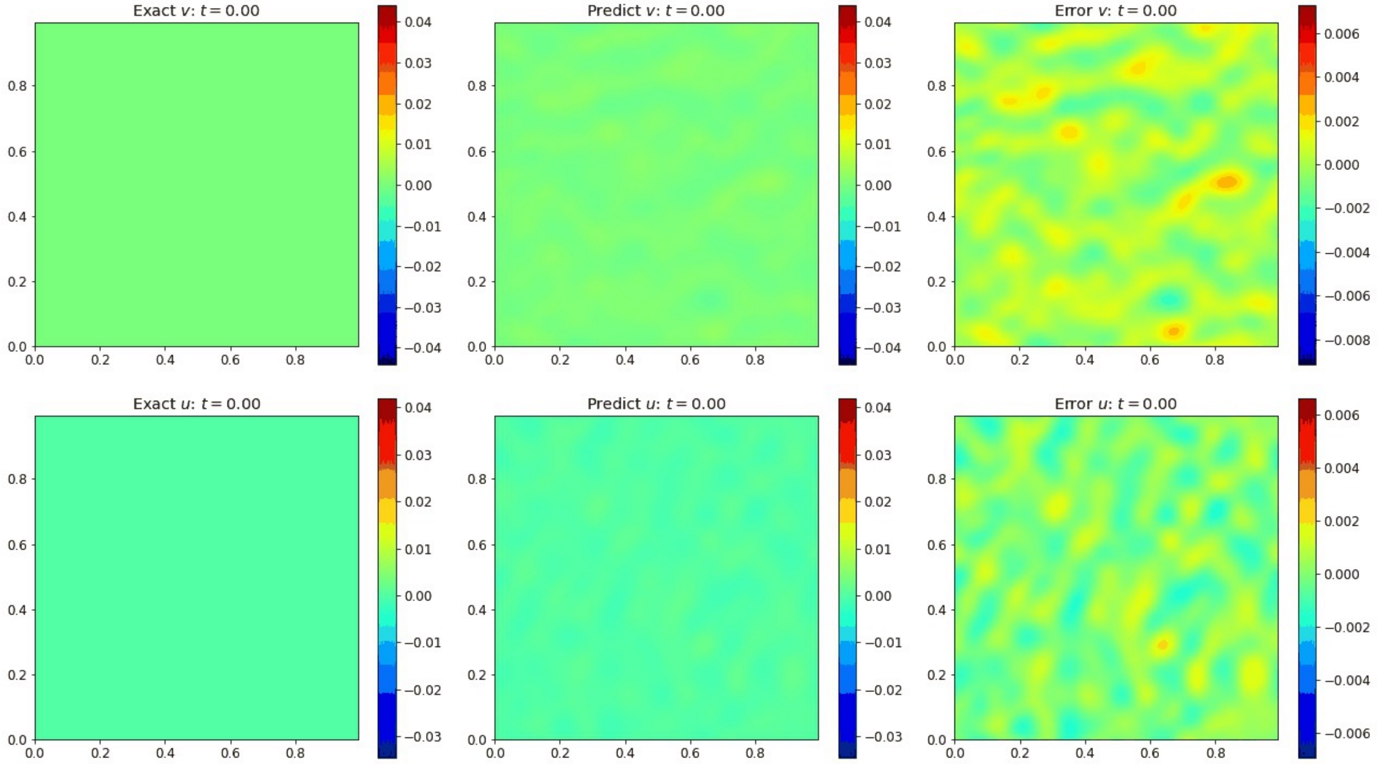
arXiv > physics > arXiv:2203.12634

Physics > Computational Physics

[Submitted on 23 Mar 2022]

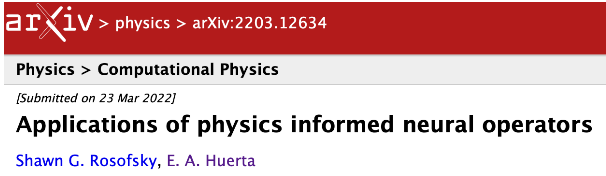
Applications of physics informed neural operators

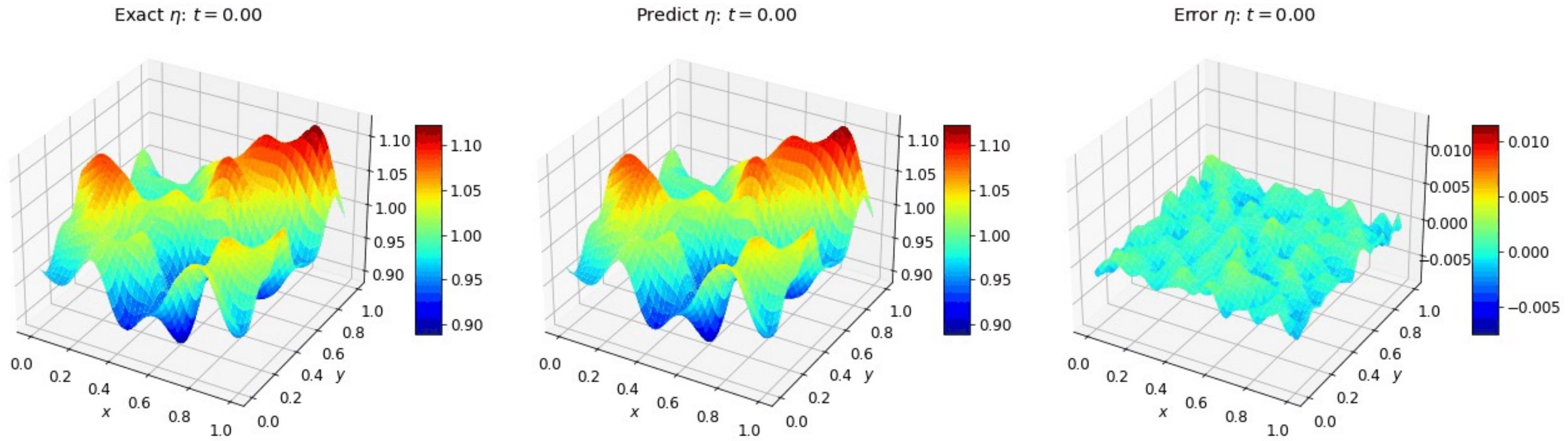
Shawn G. Rosofsky, E. A. Huerta



SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

Physics informed neural operators





DYNAMIC AI

DLHub+funcX: reproducible, scalable and accelerated AI-discovery at the edge

Summit
ThetaGPU, AURORA
…
Reduce time-to-insight with HPC platforms

Deploy dynamic AI models in DLHub

Edge Distributed Computing TensorRT …

Major upgrade of AI models

Active/Transfer/Reinforcement

Burst training

REFERENCES

Gravitational Wave Data Analysis | Machine Learning

https://iphysresearch.github.io/Survey4GWML/

AI-ready datasets

Innovative computing

FAIR, interpretable, physics-inspired, accelerated AI models

Data fusion & new modes of data-driven discovery & smart cyberinfrastructure

ACKNOWLEDGEMENTS

This material is based upon work supported by Laboratory Directed Research and Development (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357

This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725

We acknowledge support from NSF OAC-1931561, OAC-1934757, OAC-2004894, NVIDIA and IBM

Argonne
NATIONAL LABORATORY

75
1946–2021